Identification and modification of amine dehydrogenases by genomic and structural approaches for biocatalytic amine synthesis

Laurine Ducrot

▶ **To cite this version:**

Laurine Ducrot. Identification and modification of amine dehydrogenases by genomic and structural approaches for biocatalytic amine synthesis. Catalysis. Université Paris-Saclay, 2022. English. NNT : 2022UPASF079 . tel-04023477

HAL Id: tel-04023477
https://theses.hal.science/tel-04023477

Submitted on 10 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

université
PARIS-SACLAY

Identification and modification of amine dehydrogenases by genomic and structural approaches for the biocatalytic amine synthesis

Identification et modification d'amine déshydrogénases par approches génomique et structurale pour la synthèse biocatalysée d'amines

Thèse de doctorat de l'Université Paris-Saclay

École doctorale n°571, Sciences Chimiques : Molécules, Matériaux, Instrumentation et Biosystèmes (2MIB)
Spécialité de doctorat : chimie
Graduate School : Chimie. Référent : Université d'Évry-Val-d'Essonne

Thèse préparée dans l'unité de recherche **Génomique Métabolique** (Université Paris-Saclay, Univ Evry, CNRS, CEA) sous la direction du **Dr. Carine VERGNE-VAXELAIRE**, Directrice de Recherche CEA.

Thèse soutenue à Paris-Saclay, le 6 décembre 2022 par

Laurine DUCROT

Composition du Jury

Florent ALLAIS Professeur des Universités, Agro ParisTech	Président
Frédéric BEISSON Directeur de recherche CNRS, CEA Cadarache	Rapporteur
Bettina NESTL Cadre scientifique, Innophore	Rapportrice
Juliette MARTIN Cadre scientifique, Protéus SA	Examinatrice
Erwan POUPON Professeur des Universités, Université Paris-Saclay	Examinateur

THÈSE DE DOCTORAT

NNT : 2022UPASF079

ACKNOWLEDGMENTS

First, I would like to express my sincere gratitude to Marcel Salanoubat, Patrick Wincker, Véronique de Berardinis, the former and present leading team at the Genoscope, for offering me the opportunity to pursue my PhD project in such great conditions.

I would like to thank the reviewers of my jury, Dr. Bettina Nestl and Dr. Frédéric Beisson, for accepting to review this manuscript, and with them, the rest of my jury, Prof. Florent Allais, Dr. Juliette Martin and Prof. Erwan Poupon for showing interest in my work and for the future great discussions that will come out of it.

I would like to express my gratitude to Prof. Gideon Grogan, for our joint effort on the nat-AmDHs family and for your great advice on the structural analysis of proteins. I missed several opportunities to meet you in person but luckily you also accepted to be part of my jury as our guest, and I thank you for that. Of course, I would to thank Megan Bennett for her efforts on the resolution of new nat-AmDH structures.

While I am still writing in English I would like to thank the team of Andreas Bommarius, Bettina Bommarius and especially Adam Caparco for giving me the opportunity to be part of their work on the AmDHs stability.

Je me permets de continuer en français pour tout d'abord exprimer mes plus vifs remerciements à ma directrice de thèse, Dr. Carine Vergne-Vaxelaire, l'élément moteur du projet AmDH. Merci de m'avoir fait confiance, de m'avoir laissé mon autonomie et d'avoir était un énorme soutien. Tu as toujours su avoir les mots exacts pour me remotiver quand il le fallait. Merci également pour tous tes conseils et pour toutes nos discussions sur le casse-tête scientifique que sont parfois les AmDHs.

Je voudrais bien sûr remercier Pr. Anne Zaparucha, le cœur de cette petite équipe de biocatalyse, pour son soutien, ses conseils et ses traits d'humour pinceant que j'apprécie beaucoup. Je remercie bien entendu avec elle toute l'équipe de biocatalyse; Aurélie, pour ses conseils et sa disponibilité dans n'importe quelle situation critique (souvent liée aux machines) et Chloé, pour son optimisme, son soutien sans fail, sa créativité et ses horaires décalés. Tu as toujours été une source d'inspiration pour moi et cela même au-delà du scientifique. Un grand merci au reste de l'équipe Océane, Josemarco, Dipesh, et Alex pour leur présence et leurs conseils, mais aussi à Sacha pour sa participation au projet.

Je tiens à remercier l'équipe de la plateforme de clonages et de criblages; Jean-Louis (appelez le Dieu), qui m'a énormément appris au cours de ces trois années et qui est toujours force de propositions face à mes problèmes de protéines insolubles, mais aussi Adrien et Virginie. Merci également à l'équipe de la plateforme FPLC; Alain pour tes précieux conseils sur mes problèmes de paramètres cinétiques et bien sûr Peggy, pour ton travail acharné dans la production de nos protéines. Merci à Pierre-Loïc et Delphine pour leur aide précieuse dans le développement de méthodes analytiques et pour m'avoir énormément appris dans l'utilisation des GC.

Un grand merci également à l'équipe de Métabolisme Synthétique pour leur persévérance dans ce projet; Madeleine Bouzon, Volker Döring, Valérie Delmas pour ta patience et ta pédagogie, et Ivan Dubois, pour tes petites updates sur l'évolution des AmDHs (« ça va pas fort en ce moment », « elle est en train de repartir, c'est pas mal ! »).

Je remercie également toute l'équipe du projet MODAMDH, David Vallenet, Mark Stam, Raphaël Meheust et Eric Pelletier, pour m'avoir donné l'opportunité de découvrir un domaine qui m'était totalement inconnu. Je ne pourrais jamais assez remercier Eddy qui a été plus qu'un soutien scientifique dans ma thèse. Merci pour tes conseils précieux pour résoudre mes nombreux problèmes de modélisation, docking et autres dynamiques moléculaires. Je remercie aussi ton don naturel à me faire rire, la semaine passée à Biocat a été une réelle bouffée d'air frais pendant ma fin de thèse.

Un grand remerciement revient à Dr. Gwenaëlle André-Leroux pour son implication dans le projet et surtout pour avoir pris le temps de m'apprendre tout ce dont j'avais besoin en dynamique moléculaire, ce qui n'était pas une mince affaire. Je te remercie pour ta gentillesse et ton soutien sans pareil.

Mes remerciements vont également à Dr. Bogdan Iorga pour avoir accepté de faire partie de mon comité de suivi de thèse et pour m'avoir prodigué tous ses conseils, grandement bénéfiques à ce projet.

J'aimerais adresser un remerciement particulier à la super équipe d'organisation du CBSO 2022, Chloé, Célestin, Cédric et Juan, avec qui j'ai partagé l'une des expériences les plus valorisantes de ma thèse. Je vous remercie de votre implication dans ce succès et je suis prête pour toute reconversion professionnelle dans l'évenementiel avec vous. Bien sûr, je n'oublie pas de remercier le bureau du CBSO,

Pr. Laurence Hecquet, Dr. Véronique Alphand et Dr. Bastien Doumèche mais aussi Egon Heuson et Carine pour leur soutien et leurs conseils tout au long de cette aventure.

J'adresse également un grand merci à toutes les personnes que j'ai pu côtoyer durant ces trois années : Ivan, Jean-Louis et Magali pour le café du mercredi, Mathieu qui a toujours été d'une gentillesse incroyable, Mélanie, entre personnes qui finissent trop tard, Anne, Isabelle, Nathalie, Christophe, Nadia, Sébastien, Laurence, Franck, Aline, Bruno, Christine, Agnès, et enfin Catherine, pour ton aide précieuse dans toutes nos démarches administratives.

Un immense merci à l'ensemble des doctorants, post-docs, nouveaux ingénieurs que j'ai eu la chance de rencontrer : Chloé L., Déborah, Oriane D., Julie, Romuald, Eddy, Brieuc, Océane, Clément, Tom, Océane, Marie, Nina, Chloé B., Marion, Oriane M., William, Paul, Jérôme, Aysis, Antoine, Aude, Dipesh, Adelme et Kevin. Du fond du cœur, j'aimerais adresser un remerciement tout particulier à Déborah pour nos moments de soutien dans nos galères mutuelles, pour m'avoir fait rire avec ses histoires farfelues pendant nos pauses café et pour m'avoir supporté durant ces derniers mois.

Enfin, un immense merci à mes amis qui ont contribué à rendre ces trois années si spéciales : les copines de galère de thèse, Aurélie, Naïma et Vivi, bien sûr Cécile parce que You are my Destiny, Marie, Célie, Jeanne, Kaiwei, Lénaïg, Majda, Valentin, Abder, mes colocs, en particulier Victoria et Kim pour ton magnifique travail sur la couverture, et enfin Hiro, Aqil, et Tsuchi pour notre futur voyage.

Mes derniers remerciements vont bien entendu à ma petite mais très chouette famille, mon père, ma sœur et mes grand-parents et bien sûr mon petit chien Helios. Ils ne comprennent toujours pas ce que je fais mais ont été mon havre de paix et mon soutien moral durant ces trois années.

TABLE OF CONTENTS

ACKNOWLEDGMENTS 3

ABBREVIATIONS 13

LIST OF FIGURES 18

LIST OF TABLES 22

LIST OF SCHEMES 23

LIST OF APPENDIXES 25

INTRODUCTION 31

CHAPTER I: INTRODUCTION TO STRATEGIES FOR DISCOVERY AND IMPROVEMENT OF ENZYMES, STATE-OF-THE-ART OF THE BIOCATALYTIC SYNTHESIS OF AMINES 35

I. Strategies for research and improvement of biocatalysts 35

I.1 Discovery of novel enzymes 35
I.1.1 Activity-based strategies 36
I.1.2 Sequence-based strategies 36
I.1.2.1 Sequence retrieval 36
I.1.2.1.1 PCR-based strategies 36
I.1.2.1.2 Sequence similarity-based strategies 37
I.1.2.1.3 Key motif-based strategies 37
I.1.2.1.4 Profile-sequence-based strategies 37
I.1.2.2 Retrieved sequences analysis and partitioning 38
I.1.2.2.1 Phylogenetics 38
I.1.2.2.2 Sequence Similarity Network 38
I.1.2.2.3 Genomic context 39
I.1.3 3D structure-based strategies 40
I.1.3.1 Protein modeling 40
I.1.3.2 Constellation-guided strategies 41
I.1.3.3 Machine Learning/Computationally-aided strategies 41

I.2 Enzyme improvement 41
I.2.1 Random mutagenesis strategies 42
I.2.2 Focused mutagenesis strategies 42
I.2.3 Diversification by homologous recombination 43
I.2.4 Computationally-assisted strategies 44
I.2.5 I*n vivo* evolution 45
I.2.6 Introduction of non-canonical amino acids 45

I.3 De novo enzyme design and artificial enzyme 46
I.3.1 Enzyme design 46
I.3.2 Artificial Metalloenzymes 46
I.3.3 Bioinspired chemistry 47

I.4 Conclusion on the state-of-the-art on enzyme discovery, improvement and design 47

II. Biocatalytic synthesis of amines 47

II.1 Amine compounds 47

II.2 Biocatalysts for amine synthesis 49
II.2.1 Monoamine oxidases 50
II.2.2 Pictet-Spenglerases 52
II.2.3 Imine reductases and reductive aminases 53
II.2.3.1 First imine reductases catalyzing the reduction of imines 54
II.2.3.2 Discovery of the first reductive aminase 56
II.2.3.3 *Asp*RedAm homologs and further mutational work 58
II.2.3.4 Imine reductases and reductive aminases collections 60
II.2.3.5 Imine reductase improvement through machine-directed evolution 63
II.2.3.6 Imine reductases identified from metagenomic data 63
II.2.3.7 Specific case of EneIRED 64
II.2.3.8 Summarized work on the discovery and engineering of imine reductases and reductive aminases 65
II.2.4 Engineered amino acid dehydrogenases 69
II.2.4.1 Leucine dehydrogenases 70
II.2.4.2 Phenylalanine dehydrogenases 71
II.2.4.3 Transposition of mutations K68S/D261L into other leucine and phenylalanine dehydrogenases 72
II.2.4.4 Additional protein engineering work on K68/D261 73
II.2.4.5 Additional protein engineering for substrate scope expansion 74
II.2.4.6 Additional protein engineereing to reach amine dehydrogenase activity 78
II.2.4.7 Glutamate dehydrogenase 78
II.2.4.8 *L*-lysine dehydrogenase 79
II.2.4.9 Summarized work on the discovery and engineering of amine dehydrogenases from amino acid dehydrogenases 81
II.2.5 Engineered opine dehydrogenases 85
II.2.6 Natural amine dehydrogenases 87
II.2.6.1 First genome mining to find native reductive amination activity 87
II.2.6.2 First set of native amine dehydrogenases 89
II.2.6.3 Expansion of the family by exploring metagenomic databases 94
II.2.6.4 Summarized work on the discovery and engineering of native amine dehydrogenases 95

II.3 Conclusion on the state-of-the art on biocatalysts for amine synthesis 98

III. Conclusion, objectives and strategy 98

CHAPTER II: CHARACTERIZATION AND COMPREHENSION OF NATIVE AMINE DEHYDROGENASES 101

I. Structural description 101

I.1 Global structure 101
I.1.1 Description 101
I.1.2 Diversity within the reported native amine dehydrogenases 103

II. Comprehension of the role of the essential residues in the active site 105

II.1 P3 for catalytic activity 107

II.2 P12 for active site closing 108

II.3 P17 for correct positioning of P3 109

II.4 P10 and P18 for substrate positioning and stabilization 110

II.5 P6 for cofactor anchoring 113

II.6 Conclusion on the essential positions of the nat-AmDHs active site....113

III. Stability of nat-AmDHs....113

III.1 Stability of nat-AmDHs monitored though specific activity assays....114

III.2 Correlation between experimental results and amine dehydrogenases structural features....117

IV. Further characterization of the native amine dehydrogenases carbonyl substrate spectrum....118

IV.1 Small prochiral (hydroxy)ketones....118
IV.1.1 Screening and hits characterization....119
IV.1.2 Correlation with *in silico* analysis....120

IV.2 Other short aliphatic and aromatic carbonyl compounds....124
IV.2.1.1 Monitoring of the amine formation....124
IV.2.1.2 Monitoring of the alcohol formation....126

V. Opening to the reductive sulfidation activity....127

VI. Conclusion....130

CHAPTER III: IMPROVEMENT OF NATIVE AMINE DEHYDROGENASES....131

I. Expansion of the native amine dehydrogenases substrate scope through structural exploration and protein engineering....131

I.1 *In silico* analysis....131

I.2 Specific activity of *Cfus*AmDH variants towards longer aldehydes....134

I.3 Conversion assays with NAD(P)-recycling system....135

I.4 Transposition of the positive mutations into other native amine dehydrogenases....136
I.4.1 Mutagenesis....136
I.4.2 Screening analytical development....137
I.4.3 Screening results on carbonyl compounds harboring a chain length of at least six carbon atoms....140
I.4.4 Screening results on short aliphatic and aromatic carbonyl compounds....143
I.4.5 Complementary results on ketones with the carbonyl function on the third carbon atom of the chain....144
I.4.6 Biochemical characterizations of *Cfus*AmDH-W145A....145
I.4.6.1 Kinetic parameters....145
I.4.6.2 Thermostability....147
I.4.6.3 Tolerance to substrate loading....147
I.4.6.4 Utilisation of *Cfus*AmDH-W145A in preparative scale reactions....148
I.4.7 Structural characterization of *Cfus*AmDH-W145A....149
I.4.7.1 Cristallization.... Erreur ! Signet non défini.
I.4.7.2 Molecular dynamics....153
I.4.7.2.1 Overall dynamics....153
I.4.7.2.2 Affinity of CfusAmDH-W145A for larger substrates....155
I.4.7.2.3 Coordination sphere of the mutated site....156
I.4.7.2.4 Adenosine binding site....158

II. *In vivo* directed evolution of amine dehydrogenases....160

II.1 Principle of *in vivo* directed evolution using the automatic devices available at the Genoscope....160

II.1.1 Chemostat regime 161
II.1.2 Turbidostat regime 161
II.1.3 Medium swap regime 161

II.2 Selection screen based on NH_3 release 162
II.2.1 Context of the *in vivo* evolution of amine dehydrogenases 162
II.2.2 Evolution system using *D*-ornithine as starting substrate 164
II.2.3 Evolution system using 4-aminopentanoate as starting substrate 166
II.2.3.1.1 Application for the evolution of AmDH5 and TtherAmDH 166
II.2.3.1.2 Application for the evolution of AmDH4-I80T/P224S/E296G 167

II.3 Selection screen based on NADPH release 169
II.3.1 Application for the evolution of *Micro*AmDH 169
II.3.2 Application for the evolution of *Asp*RedAm 170

III. Design of native amine dehydrogenases chimeras 175

IV. Conclusion 178

CHAPTER IV: INTENSIVE RESEARCH FOR NATIVE AMINE DEHYDROGENASES AMONG (META)GENOMIC DIVERSITY 181

I. First selection of native amine dehydrogenases with larger active sites 181

II. Selection of other native amine dehydrogenases from the MODAMDH project 184

II.1 Datasets construction and clustering 184

II.2 Selection strategy using distant homology 185

II.3 Experimental validation of the extended family of native amine dehydrogenases 186

II.4 Selection strategy using active site analogy 187

II.5 Selection of native amine dehydrogenases with novel features 189
II.5.1 Selection of enzymes for accommodation of carbonyl substrates harboring a chain length of at least six carbon atoms 190
II.5.2 Selection of enzymes for accommodation of primary or secondary amines 191
II.5.3 Selection of enzymes for accommodation of ketones harboring a carbonyl function on the third carbon atom of the chain 194
II.5.4 Selection of enzymes to reach (*R*)-stereoselectivity 196
II.5.5 Final selection of key novel nat-AmDHs 200

III. Conclusion 201

CONCLUSION AND PERSPECTIVES 203

EXPERIMENTAL SECTION 207

I. Generals 207

I.1 Chemicals 207

I.2 Equipments ..... **Erreur ! Signet non défini.**

I.3 Softwares 208

II. Enzyme production 209

II.1 General method for enzyme cloning 209
II.2 General method for site-directed mutagenesis 209
II.3 General method for enzyme induction 209
II.4 General method for small-scale purification of enzymes 210
II.5 General method for large-scale purification of enzymes 210
II.6 Enzyme production for stability experiments 210
II.7 Enzyme production protocols tested for insoluble enzymes 211
II.7.1 Insoluble enzyme production using ArcticExpress (DE3) competent cells 211
II.7.2 Insoluble enzyme production using an autoinduction protocol 211

III. Spectrophotometric assays 212
III.1 General method for enzymatic activities through spectrophotometric assays 212
III.2 Stability assays of a panel of native amine dehydrogenases through specific activity measurements 212
III.3 Enzymatic activity of *Cfus*AmDH mutants with enlarged active sites 213
III.4 Kinetic parameters of *Cfus*AmDH-W145A 213
III.5 Thermostability of *Cfus*AmDH and *Cfus*AmDH-W145A 213
III.6 Enzymatic activity in the oxidative deamination direction 213
III.6.1 Enzymatic activity of AmDH4 and AmDH4-I80T/P224S/E296G towards 4-aminopentanoate 214
III.6.2 Enzymatic activities of *Micro*AmDH towards a panel of key amine substrates 214

IV. Conversion assays 214
IV.1 General method for biocatalytic reactions 214
IV.2 Conversion monitoring using liquid chromatography 215
IV.2.1 General method using UHPLC-UV/MS 215
IV.2.2 Derivatization with benzoyl chloride and UHPLC-UV/MS analytical method 215
IV.2.3 *Cfus*AmDH, *Cfus*AmDH-F140A and *Cfus*AmDH-W145A tolerance to DMSO 215
IV.2.4 Monitoring of 4-phenylbutan-2-one conversion and estimation of *ees* after derivatization with FDAA 215
IV.2.5 Derivatization with FDAA for monitoring of the conversion of ketones harboring a carbonyl function on the third carbon atom of the chain 216
IV.3 Conversion monitoring using gas chromatography 216
IV.3.1 Conversion monitoring using GC-FID 216
IV.3.2 Monitoring of reductive sulfidation using GC-HS-MS 217
IV.3.2.1.1 Reductive sulfidation reactions 217
IV.3.2.1.2 Analytical method for reductive sulfidation monitoring 218
IV.4 Semi-preparative scale biocatalytic synthesis 218
IV.4.1 *Cfus*AmDH-W145A tolerance to higher substrate loading 218
IV.4.2 Semi-preparative scale reactions and work up using *Cfus*AmDH-W145A 219

V. *In vivo* directed evolutions 219
V.1 General method for gene subcloning 220

V.2 General method for gene cloning 220
V.3 General method for growth assays in liquid medium 220
V.4 General method for bioscreen assays 221
VI. Structural characterizations 221
VI.1 Crystallization and structural resolution of *Cfus*AmDH-W145A 221
VI.2 Bioinformatic analyses 222
VI.2.1 Protein modeling 222
VI.2.2 Molecular dockings 222
VI.2.2.1 General method for molecular dockings 222
VI.2.2.2 Molecular docking of short chiral amines 222
VI.2.2.3 Molecular docking of pentan-1-amine to decan-1-amine into *Cfus*AmDH, *Cfus*AmDH-F140A and *Cfus*AmDH-W145A 223
VI.2.2.4 Molecular docking of ethylamine and methylamine into *Cfus*AmDH and *Cfus*AmDH-L177A 223
VI.2.2.5 Molecular docking of (*2R/S*)-pentan-2-amine into MGYP000357504158, GUT_GENOME186969_00552 and MGYP000037226974 223
VI.2.3 Molecular dynamics 224
VI.2.3.1.1 Molecular dynamic simulations 224
VI.2.3.1.2 Molecular dynamic outputs 225
BIBLIOGRAPHY 227
APPENDIXES 245

ABBREVIATIONS

(*S*)-MOIPA	(2*S*)-1-methoxypropan-2-amine (**185**)
2,4-DAPDH	2,4-diaminopentanoate dehydrogenase
2A4OP	2-amino-4-oxopentanoic acid
2C ketone	Ketone with the carbonyl moiety on the second carbon atom of the carbon chain (forming a methyl group on one side)
3C ketone	Ketone with the carbonyl moiety on the third carbon atom of the carbon chain (forming an ethyl group on one side)
3D	Three dimensional
4-AP	Commercial racemic 4-aminopentanoate used in Chapter III, II.2.3
4OP-AmDH	First set of native AmDHs active towards 4-oxopentanoic acid (**152**)
AADH	Amino Acid Dehydrogenase
ADO	Assembly of Designed Oligonucleotides
AmDH	Amine Dehydrogenase
ANR	French National Research Agency
API	Active Pharmaceutical Ingredient
ASMC	Active Site Modeling and Clustering
ASMC1	Restricted set of nat-AmDHs
ASMC2	Enlarged set of nat-AmDHs
ATP	Adenosine triphosphate
AUC	Analytical UltraCentrifugation
BASF	Baden Aniline and Soda Factory
BLASTP	Basic Local Alignment Search Tool Program
BVMO	Baeyer-Villiger Monooxygenase
BzCl	Benzoyl chloride (**221**)
CAST	Combinatorial Active Site Saturation Test
CEA	French Alternative Energies and Atomic Energy Commission
COMPASS	Comparison of Multiple Protein Alignments with Assessment of Statistical Significance
CPD	Computational Protein Design
C-ter	C-terminal
CX ketone	Ketone chain with X carbon atoms
DIAMOND	Double Index Alignment of Next Generation Sequencing Data
DMS	Deep Mutational Scanning

DMSO	Dimethyl sulfoxide (**204**)
DNA	Desoxyribonucleic acid
Dnase	Desoxyribonuclease
dNTP	Deoxynucleotide triphosphate
DoE	Design-of-Experiments
DSF	Differential Scanning Fluorimetry
eDNA	Environmental desoxyribonucleic acid
ee	Enantiomeric excess
EFI-EST	Enzyme Function Iterative-Enzyme Similarity Tool
EM	Electron Microscopy
epPCR	Error-prone Polymerase Chain Reaction
eq.	Equivalent
ERED	Ene-reductase
EtOAc	Ethyl acetate
FDAA	1-fluoro-2-4-dinitrophenyl-5-*L*-alanine (**227**)
FDH	Formate dehydrogenase
FRESCO	Framework for Rapid Enzyme Stabilization Computational Libraries
FRISM	Focused Rational Iterative Site-Specific Mutagenesis
G1	Group 1 from ASMC1
G2	Group 2 from ASMC1
G3	Group 3 from ASMC1
G4	Group 4 from ASMC1
G5	Group 5 from ASMC1
GC	Gas Chromatography
GC-FID	Gas Chromatography equipped with a Flame Ionization Detector
GC-HS-MS	Gas Chromatography equipped with a HeadSpace Module and Mass Detection
GC-MS	Gas Chromatography equipped with a Mass detection
GDH	Glucose Dehydrogenase
GDH-105	Glucose Dehydrogenase from *Candida boidinii*
GFP	Green Flurorescent Protein
GluDH	Glutamate dehydrogenase
GM3	Automated evolution devices GENEMAT (version GM3), technology developed by the CEA in collaboration with ALTAR
GSK	GlaxoSmithKline
GSK_IRXX	IRED collection from GSK reported by Roiban *et al.* (2017)

GST	Glutathione *S*-transferase
HEP	Heptan-1-amine (**213**) ligand in Chapter III, I.5.5.2
HMM	Hidden Markov Model
idX/covX	X% identity over X% of the sequence length
IGC	Integrated Gene Catalog
INRAE	National Research Institute for Agriculture, Food and the Environment
IPTG	Isopropyl β-*D*-1-thiogalactopyranoside (**250**)
IRED	Imine reductase
ISM	Iterative Saturation Mutagenesis
I-TASSER	Iterative Threading ASSEmbly Refinement
ITCHY	Incremental Truncation for the Creation of Hybrid
k_{cat}	Catalytic Constant (turnover number)
K_M	Michaelis-Menten constant
L2BMS	Laboratory of Bioremediation, Biocatalysis and Synthetic Metabolism
LABGeM	Laboratory of Bioinformatics Analyses for Genomics and Metabolism
LAMA	Local Alignment of Multiple-Alignments
LB	Luria Broth
LCOB	Laboratory of Chemistry and Biocatalysis
LeuDH	Leucine Dehydrogenase
LGBM	Laboratory of Genomics and Biochemistry of Metabolism
LIC	Ligation Independent Cloning
LOD	Limit Of Detection
LysEDH	Lysine Dehydrogenase
MAGE	Multiplex Automated Genome Engineering
MaIAGE	Applied Mathematics and Computer Science, from Genome to Environment
MAO	Monoamine oxidase
MATOU	Marine Atlas of Tara Oceans Unigenes
MD	Molecular Dynamic(s)
MTBE	Methyl tert-butyl ether
METdb	Marine Eukaryotes Transcriptomes
MMS	Microfluidic Modulation Spectroscopy
MS	Mass
w/v	Ratio weight/volume
NAD	NAD^+ ligand in Chapter III, I.5.5.2

NAD^+	Nicotinamide adenine nucleotide (oxidized form)
NADH	Nicotinamide adenine nucleotide (reduced form)
NADP	$NADP^+$ ligand in Chapter III, I.5.5.2
$NADP^+$	Nicotinamide adenine nucleotide phosphate (oxidized form)
NADPH	Nicotinamide adenine nucleotide phosphate (reduced form)
nat-AmDH	native Amine Dehydrogenase
NCBI	National Center for Biotechnology Information
NExT	Nucleotide Exchange and Excision Technology
NMR	Nuclear Magnetic Resonance
NTA	Nitriloacetic acid
N-ter	N-terminal
OA	Ornithine Aminomutase from *Clostridium Sticklandii*
OD	Optical Density
OmniSEC	Gel permeation and size exclusion chromatography
OM-RGC	Ocean Microbial Reference Gene Catalog
OpDH	Opine Dehydrogenase
P5A	Mutation of the residue located at P5 position of nat-AmDHs in Ala
P8A	Mutation of the residue located at P8 position of nat-AmDHs in Ala
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PEN	Pentan-1-amine (**211**) ligand in Chapter III, I.5.5.2
Pfizer_IRXX	IRED collection from Pfizer reported by France *et al.* (2018)
PheDH	Phenylalanine Dehydrogenase
PLP	Pyridoxal phosphate (**243**)
PROSS	Protein Repair One Stop Shop
RACHITT	Random Chimeragenesis on Transient Templates
RCSB	Research Collaboratory for Structural Bioinformatics
REAP	Reconstructed Evolutionary Adaptative Path methodology
RedAm	Reductive Aminase
RGN	Recurrent Geometric Network
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
Roche_IRXX	IRED collection from Roche reported by Wetzl *et al.* (2016)
rpm	Rotation per minute

RT	Room temperature
SDM	Site-Directed Mutagenesis
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SeSaM	Sequence Saturation Mutagenesis
SHIPREC	Sequence Homology Independent Protein Recombination
SPM	Short Path Map
SSN	Sequence Similarity Network
StEP	Staggered Extension Process
STY	Space Time Yield
SUMO	Small Ubiquitin-like Modifier
T_{50}^{30}	Temperature at which enzyme activity is reduced to 50% after a 30-min heat treatment
TA	Transaminase
TB	Terrific Broth
T_m	Melting temperature
TON	Turnover number
tRNA	Transfer ribonucleic acid
TTN	Total Turnover Number
UHPLC	Ultra-High Performance Liquid Chromatography
UHPLC-MS	Ultra-High Performance Liquid Chromatography equipped with Mass detection
UHPLC-UV	Ultra-High Performance Liquid Chromatography equipped with a Ultraviolet detection
UMR	Unité Mixte de Recherche
UV	Ultra Violet
v/v	Ratio volume/volume
WT	Wild-type

Amino acids are indicated using their three-letters denomination (Ala, Glu, Phe, etc.) except when corresponding to specific residues in a protein sequence for which they are indicated as their single-letter denomination followed by the residue position (K68, E108, etc.). Mutations are indicated as AXB with A, the native amino acid, X, its position in the protein sequence and B, the amino acid after the mutation.

The structure and numbering of the molecules mentioned in Chapter I, II, III and IV are given in Appendix 1.

LIST OF FIGURES

Figure 1. Schematic research steps in biocatalysis from enzyme discovery to product recovering 35
Figure 2. Simplified phylogenetic tree depicting the potential of this approach to reach ancestral enzymes and new modern enzymes. 38
Figure 3. Sequence Similarity Networks of the nitroreductase superfamily 39
Figure 4. Global structure and active site of *Asp*RedAm. 57
Figure 5. Active Site of *At*RedAm in complex with redox-inactive NADPH4, cyclohexanone and allylamine 58
Figure 6. Logo representation of summary of key residues and mutation work done on selected IREDs with reductive aminase activity 66
Figure 7. Substrate specificity of all the engineered IREDs described in this part. 69
Figure 8. Structure of LeuDH from *Bacillus stearothermophilus* 70
Figure 9. Structure of *Ec*GluDH from *E. coli*-K12 79
Figure 10. Model of the ε-deaminating *L*-lysine dehydrogenase from *Geobacillus stearothermophilus*. 80
Figure 11. Logo representation of the mutational work done on LeuDHs and PheDHs 82
Figure 12. Substrate specificity of all the engineered AADHs and ε-deaminating *L*-lysine dehydrogenase described in this part 84
Figure 13. Global structure and active site of CENDH 86
Figure 14. Global structure and active site of AmDH4 89
Figure 15. Global structure and active site of *Cfus*AmDH 91
Figure 16. ASMC1 clusterization. 93
Figure 17. Logo representation of the pocket diversity found in the nat-AmDHs family and mutation work done on nat-AmDHs 95
Figure 18. Substrate specificity of the nat-AmDH family. 97
Figure 19. Structure of *Cfus*AmDH 102
Figure 20. Structure of MATOUAmDH2 in complex with $NADP^+$ 103
Figure 21. Sequence identity matrix of the reported nat-AmDHs 104
Figure 22. Variable helix in nat-AmDHs.. 105
Figure 23. *Cfus*AmDH active site 106
Figure 24. Logo representation of the conservation pattern of P1-P21 positions among ASMC1 G3 and G4 groups. 106
Figure 25. Influence of mutations at P3 in *Cfus*AmDH. 107
Figure 26. Influence of mutations at P12 in *Cfus*AmDH and MATOUAmDH2 108

Figure 27. Influence of mutations at P17 in *Cfus*AmDH and MATOUAmDH2.....109
Figure 28. Influence of mutations at P10 in *Cfus*AmDH and *Msme*AmDH 111
Figure 29. Influence of additional mutations at P10 in *Cfus*AmDH and double mutations at P10 and P18 in *Cfus*AmDH and *Msme*AmDH.....112
Figure 30. Relative remaining activity of IGCAmDH5, *Acol*AmDH, IGCAmDH1, *Chat*AmDH, *Msme*AmDH, *Sgor*AmDH, *Cfus*AmDH and MATOUAmDH2 towards **170** after different storage conditions.....116
Figure 31. Putty representation of *Cfus*AmDH, *Msme*AmDH and MATOUAmDH2.....117
Figure 32. Docking simulation of (2*S*)- and (2*R*)- short amines 121
Figure 33. *Msme*AmDH RX structure in complex with $NADP^+$ cofactor with docked structures of **182** and **193**.....121
Figure 34. Docking of (2*S*)-1-hydroxypropan-2-amine and (2*R*)-1-hydroxypropan-2-amine in *Micro*AmDH and alignment with *Cfus*AmDH, *Msme*AmDH and MATOUAmDH2.....122
Figure 35. Docking of (2*S*)-1-hydroxybutan-2-amine and (2*R*)-1-hydroxybutan-2-amine in *Cfus*AmDH and *Msme*AmDH 124
Figure 36. Thiol and alcohol formation from cyclopentancarbaldehyde in presence of Na_2S and *Cfus*AmDH.....128
Figure 37. Thiol and alcohol formation from hexanal in presence of Na_2S and MATOUAmDH2. 129
Figure 38. Active site of *Cfus*AmDH, *Cfus*AmDH-F140A and *Cfus*AmDH-W145A with docked structures of pentan-1-amine to decan-1-amine.....132
Figure 39. Results of of pentan-1-amine to decan-1-amine dockings in *Cfus*AmDH, *Cfus*AmDH-F140A, *Cfus*AmDH-W145A.....133
Figure 40. Conversion rates obtained from heptanal with *Cfus*AmDH, *Cfus*AmDH-F140A and *Cfus*AmDH-W145A with various ratios of DMSO.....136
Figure 41. Conversion of heptanal with *Cfus*AmDH estimated from UHPLC-UV, UHPLC-MS and GC-FID detection methods 139
Figure 42. Conversions obtained for WT enzymes and mutants towards substrates **207**, **217**-**220**, **43**, **128 67**, **135** and **36**. 141
Figure 43. Conversion of 3C ketones with *Micro*AmDH, *Msme*AmDH, *Porti*AmDH, MATOUAmDH2 and their P5A and P8A respective mutants.....144
Figure 44. Thermostability of *Cfus*AmDH and *Cfus*AmDH-W145A at 25°C, 30°C and 40°C 147
Figure 45. Substrate loading influence on conversion of heptanal and octan-2-one with *Cfus*AmDH-W145A.....148
Figure 46. Structural alignment of *Cfus*AmDH, *in silico* mutated *Cfus*AmDH-W145A and *Cfus*AmDH-W145A structure obtained from crystallization.....150

Figure 47. Active site of *Cfus*AmDH-W145A 3D-structure in complex with $NADP^+$/pentan-1-amine and NAD^+ 151
Figure 48. Pentan-1-amine positioning when co-cristallized in *Cfus*AmDH-W145A 3D-structure or docked in modelled *Cfus*AmDH-W145A. 152
Figure 49. Cofactor binding site of *Cfus*AmDH-W145A 3D-structure in complex with $NADP^+$/pentan-1-amine (**211**) and NAD^+. 152
Figure 50. *Cfus*AmDH flexible sections 154
Figure 51. Evolution of G14-G174 distance as a function of time 155
Figure 52. Interaction energies of Protein-Ligand calculated as the average of the total simulation. 155
Figure 53. Close-view of segments 4 and 5 of the RMSF analysis of WT and M in complex with NAD or NADP and the amine products PEN or HEP 156
Figure 54. Close-view of Q141 in WT and M at different time of the dynamic 157
Figure 55. Interaction energies of Protein-Cofactor calculated as the average of the total simulation. 158
Figure 56. Distances between the adenosine moiety of the cofactor and the main residues of its binding site 158
Figure 57. Interaction energies calculated between atoms of the adenosine moiety and (A) M and (B) WT 159
Figure 58. First selection system for the *in vivo* directed evolution of AmDH5. 163
Figure 59. Second selection system for the *in vivo* directed evolution of AmDH5 and *Ther*AmDH 164
Figure 60. Specific activity of AmDH4-I80T/P224S/E296G towards **152** and 4-AP in both ways of reaction. 168
Figure 61. G5780 growth on a **244**-containing medium medium with a range of **54** concentrations. 172
Figure 62. First attempt of G5780 evolution in continuous cultivation in the GM3 device 172
Figure 63. Second attempt of G5780 evolution in continuous cultivation in the GM3 device. 174
Figure 64. Structure alignment at 0° and 180° of *Cfus*AmDH, *Micro*AmDH, MATOUAmDH2 on AmDH4 176
Figure 65. Multiple sequence alignment of the template 1 (AmDH4), the templates 2 (*Cfus*AmDH, *Msme*AmDH and MATOUAmDH2) and the corresponding chimeras AmDH4_X and AmDH4_X_P17all (X being the enzyme used as template 2). 177
Figure 66. ASMC2 clusterization. 185
Figure 67. Phylogenetic tree of (A) the extended reference nat-AmDH set (id80/cov80) and (B) the extended reference nat-AmDH set with addition of the sequences retrieved from the distant homology approach 186
Figure 68. Atoms and distances considered for the design of minimum constellations in *Cfus*AmDH. 188

Figure 69. *Cfus*AmDH active site with docking of cyclohexanone and ammonia. 191
Figure 70. Docking of methylamine and ethylamine amine substrates in *Cfus*AmDH and *Cfus*AmDH-L177A. 192
Figure 71. Second pocket in nat-AmDHs 195
Figure 72. Examples of nat-AmDHs with potential catalytic residue at P5′ or P20. 198
Figure 73. Docking of (2*S*)-pentan-2-amine and (2*R*)-pentan-2-amine in MGYP000357504158 (MGY-158) GUT_GENOME186969_00552 (GUT-552) and MGYP000037226974 (MGY-974) with potential catalytic residues at P5′ or P20. 199

LIST OF TABLES

Table 1. Summary of key residues and mutation work done on selected IREDs with reductive aminase activity 67
Table 2. Summary of mutational work done on LeuDHs and PheDHs 83
Table 3. List of main residues defining the catalytic pocket in WT nat-AmDHs and the corresponding mutated residues in the engineered enzyme. 96
Table 4. Screening of nat-AmDHs towards furfural, cyclopentancarbaldehyde, benzaldehyde, acetophenone and hexan-3-one. 126
Table 5. Specific activities of *Cfus*AmDH and *Cfus*AmDH-W145A towards pentanal to octanal 134
Table 6. Results in amine formation formation for the screening of nat-AmDHs WT and mutants towards **57**, **199**, **31**, **39** and **200**. 143
Table 7. Kinetic parameters of *Cfus*AmDH-W145A. 146
Table 8. Summary of liquid cells growth and bioscreen results of the different strains created in the three source of nitrogen discussed in this part. 165
Table 9. Summary of liquid cells growth results of the different strains created in presence of **153** and 4-AP as nitrogen source.. 167
Table 10. Liquid cells growth results of the different strains containing AmDH4 and AmDH4-I80T/P224S/E296G in presence of 4-AP as nitrogen source 168
Table 11. Growth of G5313 and G5641 in presence of gluconate and/or amine substrate 170
Table 12. Growth of G5313, G5647 and G5648 in presence of gluconate and/or *N*-methylcyclohexylamine 171
Table 13. Comparison of P1-P21 residues of AmDH4, *Cfus*AmDH, *Micro*AmDH and MATOUAMDH2. 175
Table 14. P1-P21 residues of the nat-AmDHs selected within the ASMC1 set to accommodate large carbonyl substrates. 182
Table 15. Atoms and flexibility of amino acid type for the constellations design 188
Table 16. P1-P21 residues of the selected native AmDHs with a larger active site for carbonyl substrates 190
Table 17. P1-P21 residues of the selected nat-AmDHs to accommodate larger amine substrates 193
Table 18. P1-P21 residues of the selected nat-AmDHs to accommodate 3C ketones 196
Table 19. P1-P21 residues of the selected nat-AmDHs to switch the catalytic residue position 200
Table 20. Gradients used for analyses in GC-FID. 217

LIST OF SCHEMES

Scheme 1. Structure of some pharmaceuticals and agrochemicals containing an amine function. 48
Scheme 2. Reductive amination reaction 49
Scheme 3. Monoamine oxidase-mediated deracemization process of chiral amines 50
Scheme 4. Synthesis of chiral amine intermediates using MAO-N D11 mutant towards the synthesis of Levocetirizine® and Solifenacin® 51
Scheme 5. MAO-N 401-mediated desymmetrization for the synthesis of a key intermediate in the synthesis of Boceprevir 52
Scheme 6. Strictosidine synthase-mediated Pictet-Spenglerase last step reaction in the total synthesis of L-strictosidine 53
Scheme 7. General reactions catalyzed by IREDs and RedAms 53
Scheme 8. Example of reaction catalyzed by (*S*)-IRED from *Streptomyces* sp. GF3546 54
Scheme 9. Some reductive amination reactions of ketones using the IRED collection at Roche 55
Scheme 10. Some reductive amination reactions catalyzed by *Asp*RedAm from *Aspergillus oryzae* using low amine:ketone ratios 57
Scheme 11. Mechanism proposed for the fungal RedAm-catalyzed reductive amination of a ketone by small amines 59
Scheme 12. Synthesis of (2*R*)-(phenylamino)-(1*R*)-methylcyclopentan-1-ol by reductive amination using IR007-143 mutant 61
Scheme 13. Synthesis of lysine-specific demethylase-1-inhibitor GSK2879552 precursor by reductive amination using mutant 'M3' of IRED GSK_IR46 acquired by directed evolution 62
Scheme 14. Synthesis of Rotigotine by reductive amination using pIR-221 64
Scheme 15. Conjugate reduction and reductive amination catalyzed by EneIRED on a preparative scale reaction 65
Scheme 16. General reactions catalyzed by WT and engineered LeuDHs and PheDHs discussed in this part 69
Scheme 17. Some reductive amination reactions catalyzed by engineered AADHs with mutations on positions K68/N261 73
Scheme 18. Some reductive amination reactions catalyzed by engineered AADHs with mutations inside and outside the catalytic pocket 76
Scheme 19. Some reductive amination reactions catalyzed by engineered AADHs obtained after several rounds of mutagenesis 78
Scheme 20. Reactions catalyzed by WT and engineered *Ec*GluDH 79

Scheme 21. General reaction catalyzed by LysEDH and the engineered LysEDH 80
Scheme 22. Example of reductive amination reaction catalyzed by the engineered ε-deaminating *L*-lysine dehydrogenase 81
Scheme 23. General reaction catalyzed by OpDHs and engineered CENDH 85
Scheme 24. General reactions catalyzed by nat-AmDHs 87
Scheme 25. Example of reductive amination reaction catalyzed by AmDH4. 88
Scheme 26. Examples of reductive amination reactions catalyzed by nat-AmDHs 90
Scheme 27. Hypothetical mechanism of nat-AmDHs 92
Scheme 28. Example of reductive amination reaction catalyzed by IGCAmDH1 obtained from metagenomic data 94
Scheme 29. Semi-preparative scale synthesis of (2*S*)-1-methoxypropan-2-amine and (2*S*)-butan-2-amine catalyzed by *Msme*AmDH 119
Scheme 30. Expected reductive sulfidation of cyclopentancarbaldehyde in presence of Na_2S and *Cfus*AmDH 127
Scheme 31. Semi-preparative scale synthesis of heptan-1-amine and (2*S*)-octan-2-amine 149
Scheme 32. Selection system for the *in vivo* directed evolution of nat-AmDH based on NADPH release. 169
Scheme 33. Selection system for the *in vivo* directed evolution of *Asp*RedAm based on NADPH release upon deamination of *N*-methylcyclohexylamine. 170
Scheme 34. Schematic view of a catalytic glutamate located on the opposite side of P3 in the active site 197

LIST OF APPENDIXES

Appendix 1. List of molecules mentioned in the manuscript 245

Appendix 2. (Chapter I, I.1.3.1) Algorithms for protein fold prediction 250

Appendix 3. (Chapter I, II.2.6.1) Genome mining approach of the Cloning and Screening Platform (Genoscope) 251

Appendix 4. (Chapter I, II.2.6.2) ASMC pipeline 251

Appendix 5. (Chapter II, II) Comparison of the P1–P21 positions of *Msme*AmDH (G3) and *Cfus*AmDH (G4) active site homologs from ASMC analysis (Mayol *et al*.,2019). 252

Appendix 6. (ChapterII, II.3) Structural alignment of *Cfus*AmDH and AmDH4 of comparison of P17 positions and surroundings. 256

Appendix 7. (Chapter II, IV.1.1) Conversions and *ee* results obtained with *Cfus*AmDH, *Msme*AmDH, MATOUAmDH2 and *Micro*AmDH towards short chiral amines 257

Appendix 8. (Chapter II, IV.2) Enzymes and substrates annotations for GC-FID and UHPLC-UV chromatograms from the screening of nat-AmDHs towards a range of short aliphatic and aromatic aldehydes and ketones. 258

Appendix 9. (Chapter II, IV.2.1.1 and Chapter III, I.4.4) GC-FID chromatograms for reaction of furfural (**57**) with various nat-AmDHs and mutants. 259

Appendix 10. (Chapter II, IV.2.1.1 and Chapter III, I.4.4) GC-FID chromatograms for reaction of cyclopentancarbaldehyde (**199**) with various nat-AmDHs and mutants 262

Appendix 11. (Chapter II, IV.2.1.1 and Chapter III, I.4.4) GC-FID chromatograms for reaction of benzaldehyde (**31**) with various nat-AmDHs and mutants 265

Appendix 12. (Chapter II, IV.2.1.1 and Chapter III, I.4.4) GC-FID chromatograms for reaction of acetophenone (**39**) with various nat-AmDHs and mutants 267

Appendix 13. (Chapter II, IV.2.1.1 and Chapter III, I.4.4) GC-FID chromatograms for reaction of hexan-3-one (**200**) with various nat-AmDHs and mutants 269

Appendix 14. (Chapter II, IV.2.1.1 and Chapter III, I.4.4) UHPLC-UV chromatograms for reaction of acetophenone (**39**) with various nat-AmDHs and mutants 272

Appendix 15. (Chapter II IV.2.1.1 and Chapter III, I.4.4) UHPLC-UV chromatograms for reaction of hexan-3-one (**200**) with various nat-AmDHs 273

Appendix 16. (Chapter II, IV.2.1.2 and Chapter III, I.4.4) Results in alcohol formation for the screening of nat-AmDHs WT (described in Chapter II, IV.2.1.2) and mutants (described in Chapter III, I.4.4) towards furfural (**57**), cyclopentancarbaldehyde (**199**), benzaldehyde (**31**), acetophenone (**39**) and hexan-3-one (**200**) 274

Appendix 17. (Chapter II, V) GC-HS-MS chromatograms of (A) cyclopentancarbaldehyde (**199**) (15.01 min), (B) cyclopentylmethanethiol (**205**) (17.12 min), (C) cyclopentylmethanol (**206**) (16.32 min) of reaction with *Cfus*AmDH in presence of 500 mM Na_2S. 275
Appendix 18. (Chapter II, V) GC-HS-MS chromatograms of (A) hexanal (**207**) (14.04 min), (B) hexane-1-thiol (**208**) (16.10 min), (C) hexane-1-ol (**209**) (15.43 min) of reaction with MATOUAmDH2 in presence of 250 mM Na_2S.....276
Appendix 19. (Chapter II, V) Thiol and alcohol formation from cyclohexanone (**45**) in presence of Na_2S and *Msme*AmDH. Area resulted from the transformation of **45** in presence or absence of *Msme*AmDH and 0, 250 and 500 mM Na_2S. (A) cyclohexanethiol formed, (B) cyclohexanol formed and (C) remaining aldehyde. 277
Appendix 20. (Chapter II, V) GC-HS-MS chromatograms of (A) **45** (16.36 min), (B) cyclohexanethiol (16.89 min), (C) cyclohexanol (15.97 min) of reaction with *Msme*AmDH in presence of 500 mM Na_2S. 278
Appendix 21. (Chapter III, I.3) Effect of DMSO and temperature on conversion rates towards heptanal (**217**). 279
Appendix 22. (Chapter II, IV.2) Enzymes and substrates annotations for GC-FID and UHPLC-UV chromatograms from the screening of nat-AmDHs and mutants towards a range of carbonyl-containing compouds 279
Appendix 23. (Chapter III, I.4.2) Calibration curves from the screening with (A) hexanal (**207**), (B) heptanal (**217**), (C) furfural (**57**), (D) octanal (**218**), (E) nonanal (**219**) and (F) decanal (**220**).....280
Appendix 24. (Chapter III, I.4.3) Results in amine formation for the screening of nat-AmDHs WT and mutants towards hexanal (**207**), heptanal (**217**), octanal (**218**), nonanal (**219**), decanal (**220**), hexan-2-one (**43**), heptan-2-one (**128**), octan-2-one (**67**), nonan-2-one (**135**) and 4-phenylbutan-2-one (**36**) 281
Appendix 25. (Chapter III, I.4.3) Results in alcohol formation for the screening of nat-AmDHs WT and mutants towards hexanal (**207**), heptanal (**217**), octanal (**218**), nonanal (**219**), decanal (**220**), hexan-2-one (**43**), heptan-2-one (**128**), octan-2-one (**67**), nonan-2-one (**135**) and 4-phenylbutan-2-one (**36**).282
Appendix 26. (Chapter III, I.4.3) GC-FID chromatograms for reaction of hexanal (**207**) with various nat-AmDHs and mutants 283
Appendix 27. (Chapter III, I.4.3) GC-FID chromatograms for reaction of heptanal (**217**) with various nat-AmDHs and mutants 286
Appendix 28. (Chapter III, I.4.3) GC-FID chromatograms for reaction of octanal (**218**) with various nat-AmDHs and mutants. 289
Appendix 29. (Chapter III, I.4.3) GC-FID chromatograms for reaction of nonanal (**219**) with various nat-AmDHs and mutants 292

Appendix 30. (Chapter III, I.4.3) GC-FID chromatograms for reaction of decanal (**220**) with various nat-AmDHs and mutants 295
Appendix 31. (Chapter III, I.4.3) GC-FID chromatograms for reaction of hexan-2-one (**43**) with various nat-AmDHs and mutants. 298
Appendix 32. (Chapter III, I.4.3) GC-FID chromatograms for reaction of heptan-2-one (**128**) with various nat-AmDHs and mutants. 301
Appendix 33. (Chapter III, I.4.3) GC-FID chromatograms for reaction of octan-2-one (**67**) with various nat-AmDHs and mutants. 304
Appendix 34. (Chapter III, I.4.3) GC-FID chromatograms for reaction of nonan-2-one (**135**) with various nat-AmDHs and mutants. 307
Appendix 35. (Chapter III, I.4.3) UHPLC-UV chromatograms for reaction of 4-phenylbutan-2-one (**39**) with various nat-AmDHs and mutants for conversion and *ee* estimation 310
Appendix 36. (Chapter III, I.4.3) UHPLC-UV chromatograms for reaction of hexan-2-one (**43**) with various nat-AmDHs and mutants for *ee* estimation 312
Appendix 37. (Chapter III, I.4.3) UHPLC-UV chromatograms for reaction of heptan-2-one (**128**) with various nat-AmDHs and mutants for *ee* estimation. 313
Appendix 38. (Chapter III, I.4.3) UHPLC-UV chromatograms for reaction of octan-2-one (**67**) with various nat-AmDHs and mutants for *ee* estimation 314
Appendix 39. (Chapter III, I.4.3) UHPLC-UV chromatograms for reaction of nonan-2-one (**135**) with various nat-AmDHs and mutants for *ee* estimation. 315
Appendix 40. (Chapter III, I.4.6.1) Michaelis-Menten plots for the determination of *Cfus*AmDH-W145A kinetic parameters in the system **217**/NADPH/NH_3. 316
Appendix 41. (Chapter III, I.4.6.1) Michaelis-Menten plots for the determination of *Cfus*AmDH-W145A kinetic parameters in the system **217**/NADH/NH_3. 316
Appendix 42. (Chapter III, I.4.6.1) Michaelis-Menten plots for the determination of *Cfus*AmDH-W145A kinetic parameters in the system **175**/NADPH/NH_3. 317
Appendix 43. (Chapter III, I.4.6.1) Michaelis-Menten plots for the determination of *Cfus*AmDH-W145A kinetic parameters in the system **175**/NADH/NH_3. 318
Appendix 44. (Chapter III, I.4.6.4) UHPLC-UV chromatograms chromatograms of (2*S*)-octan-2-amine (**231**) recovered from the semi-preparative scale reaction with *Cfus*AmDH-W145A. 319
Appendix 45. (Chapter III, I.4.6.4) UHPLC-UV chromatograms of heptan-1-amine (**213**) recovered from the semi-preparative scale reaction with *Cfus*AmDH-W145A 319
Appendix 46. (Chapter III, I.4.6.4) ^{1}H NMR spectrum of **231** recovered from the preparative scale reaction with *Cfus*AmDH-W145A. 320

Appendix 47. (Chapter III, I.4.6.4) ^{13}C NMR spectrum of **231** recovered from the preparative scale reaction with *Cfus*AmDH-W145A. 321
Appendix 48. (Chapter III, I.4.7.1) Data Collection and Refinement Statistics for *Cfus*AmDH-W145A in complex with NAD^+ and $NADP^+$ with pentan-1-amine (**211**) 322
Appendix 49. (Chapter III, I.4.7.2.1) Open-to-closed event observed in *Cfus*AmDH-W145A MD simulations 323
Appendix 50. (Chapter III, I.4.7.2.3) 3D representation of F140-G154 spine helix in *Cfus*AmDH 324
Appendix 51. (Chapter III, I.4.7.2.3) Distance between Q141 and cofactor along the time of the simulation 325
Appendix 52. (Chapter III, I.4.7.2.4) Distance between the adenosine moiety of the cofactor and its binding site. 326
Appendix 53. (Chapter III, I.4.7.2.4) 3D representation of the adenosine binding site of NADP and residues implied in its mobility. 326
Appendix 54. (Chapter III, I.4.7.2.4) Close-view of NADP adenosine interactions with protein residues along the simulation (WT_NADP_HEP_1) 327
Appendix 55. (Chapter III, II.1) GM3 culture vessels during the culture phase and the whashing/rinse phase 328
Appendix 56. (Chapter III, II.1) Principle of cultivation in (A) chemostat regime, (B) turbidostat regime and (C) medium swap regime. 328
Appendix 57. (Chapter III, II.2 and II.3) *E. coli* strains and constructions used for the *in vivo* evolution of nat-AmDH/RedAm activity 329
Appendix 58. (Chapter III, II) Maps of pSP100 and pVDM18 vectors 330
Appendix 59. (Chapter III, II.2 and II.3) Composition of the culture medium used in cells growth assays in liquid medium, bioscreen assays and evolution in the GM3 331
Appendix 60. (Chapter III, II.2.3.1.1) Growth assays using G4480 strain in presence of NH_4Cl and/or 4-AP. NH_4Cl eq. 2% corresponds to the addition of 4-AP 10mM if contaminated with eq. 2% v/v of NH_4Cl. 331
Appendix 61. (Chapter III, II.2.3.1.1) Monitoring of 4-AP contained in the culture medium over time in presence of G4480. A calibration curve containing various concentrations of 4-AP enabled the estimation of 4-AP concentration. 331
Appendix 62. (Chapter III, II.3) Schematic representation of the NADPH auxotrophic *E. coli* 332
Appendix 63. (Chapter III, II.3.1) Specific activity of *Micro*AmDH towards a panel of short amines in the oxidative deamination direction 332
Appendix 64. (Chapter III, II.3.2) Growth curves of G5313, G5647 and G5648 on minimum medium containing gluconate (**244**) and/or *N*-methylcyclohexylamine (N-MeCYH, **54**) as the NADPH source 333

Appendix 65. (Chapter III, II.3.2) Growth curves of G5780 and two isolates sampled from the first evolution of G5780 in the GM3 on minimum medium containing **244** and/or **54** (N-meCYH) and/or **45** (CYH) 334
Appendix 66. (Chapter III, III) Structural alignment of *Cfus*AmDH, *Micro*AmDH, MATOUAmDH2 (template 2) on AmDH4 (template 1) focusing on the top of the spine helix V134-I150. 335
Appendix 67. (Chapter III, III) Structural alignment of *Cfus*AmDH, *Micro*AmDH, MATOUAmDH2 (template 2) on AmDH4 (template 1) focusing on the linker segment G191-G199. AmDH4, *Cfus*AmDH, *Micro*AmDH and MATOUAmDH2 are colored in grey, blue, green and yellow, respectively 335
Appendix 68. (Chapter III, III) SDS-PAGE gels of the chimeras designed between AmDH4 and *Cfus*AmDH, *Micro*AmDH and MATOUAmDH2. M 336
Appendix 69. (Chapter IV, I) Sequences of the nat-AmDHs selected within ASMC1 for accomodation of larger carbonyl substrates. 336
Appendix 70. (Chapter III, III) Primers used for the cloning and production of the nat-AmDHs selected within ASMC1 for accomodation of larger carbonyl substrates 337
Appendix 71. (Chapter III, I) SDS-PAGE gels of A0A1H8XY75 overexpressed in BL21 C+ or ArcticExpress. 337
Appendix 72. (Chapter IV, II) (Meta)genomic databases used for nat-AmDH family expansion 338
Appendix 73. (Chapter III, II.3) Preliminary results of the screening carried out to describe the extended nat-AmDH family. (A) Phylogenetic tree of the extended set with id95/cov90 threshold (B) Activity results labelled as circles of proportional size according to the activity measured towards cyclohexanone (**45**), hexanal (**207**), butan-2-one (**134**), hexan-3-one (**200**), furfural (**57**), benzaldehyde (**31**) and (2*R*)-2-amine-4-oxopentanoic acid (**166**)..... 339
Appendix 74. (Chapter IV, II.5.5) Example of the design of a N-ter truncated enzyme (MGYP000882735950, in a black frame) selected within the ASMC2. 339
Appendix 75. Summary of the PhD work in French..... 339

INTRODUCTION

In order to address the global challenges that our society is currently facing, including climate change and environmental degradation, governments and scientific boards often set up new regulations and list of goals to ensure a more sustainable future for all. More addressed to the chemical industries, the twelve principles of Green Chemistry set limits on the "design of chemical products and processes to reduce or eliminate the use and generation of hazardous substances" (Anastas & Eghbali, 2010; Horváth & Anastas, 2007). More recently, the United Nations adopted The Sustainable Development Goals as 17 interlinked goals to be achieved by 2030 (United Nations website). In a comment in *Nature Catalysis*, Kristala L. J. Prather discussed the contribution of engineered enzymes and organisms and more generally biocatalysis to address seven of these goals including "Responsible consumption and production", "Life below water" or "Affordable and clean energy" (Prather, 2020). Biocatalysis, *i.e.* the use of enzymes for synthetic purposes, offers many benefits as compared to conventional chemistry. One can cite their capacity to operate at nearly ambient temperature and pressure conditions in aqueous environments and to prevent the excess use of unsustainable transition metals and complexes or hazardous solvents. Enzymatic systems can also provide high reactive selectivity and specificity that enable to decrease the number of steps in the process.

Biocatalysis is already involved into industrial synthetic pathways of key molecules whether in material [acrylamide (**1**)], pharmaceutical [Sitagliptin (**2**)], fine chemical [Vanillin (**3**)], fragrance [Nootkatone (**4**)] or even food industries [Aspartame (**5**)] (Bell *et al.*, 2021). The introduction of enzymatic processes in the chemical industry offers the possibility to completely rethink the synthetic pathways currently designed around the constraints of conventional chemistry (Devine *et al.*, 2018).

The amine function stands as a recurrent moiety in pharmaceuticals, agrochemicals as well as in fine and bulk chemicals (Chapter I, II.1). Hence, the reductive amination of carbonyl-containing compounds appeared in the top 5 of the most aspirational reactions for the industries according to Constable *et al.* in 2007 (Constable *et al.*, 2007) and represented 5.3% of all reactions of the medicinal chemist's toolbox according to Roughley and Jordan in 2011 (Roughley & Jordan, 2011). Therefore, it was considered important to find greener processes to perform such a significant reaction, for example employing enzymes such as amine dehydrogenases (AmDHs) or other catalytically similar enzyme families as described in Chapter I.

The AmDH project started in 2015 in the former Laboratory of Organic Chemistry and Biocatalysis (LCOB) now renamed as the Team of Biocatalysis within the Laboratory of Bioremediation, Biocatalysis and Synthetic Metabolism (L2BMS). This team aims to discover new biocatalysts, mainly among genomic and metagenomic resources, characterize them to demonstrate their potential for synthetic applications. A focus is done on enzymatic reactions challenging in conventional chemistry and enzymes for which none or only few are reported. The philosophy is to actively participate at the expansion of the biocatalysis toolbox for chemistry, while taking advantage of the specific expertise of the Unité Mixte de Recherche (UMR) Genomics Metabolics. At the time, only two native enzymes were described with an AmDH activity but without associated gene and very poor enantioselectivity (Itoh *et al.*, 2013; Wang & Fang, 2013; Wang 2016) and Bommarius and co-workers described the first engineering works on Amino Acid Dehydrogenases (AADHs) to create proper AmDH activity (Chapter I, II.2.4.1 and II.2.4.2). Within this context, the team bet on the discovery of native sequences within biodiversity that would give access to new templates for the synthesis of various amines. As deeply described in Chapter I, II.2.6, the first sets of native AmDHs (nat-AmDHs) were reported during the PhD projects of Ombeline Mayol and Adam Caparco but these lacked notably in active site diversity that translated into a rather limited substrate scope. The general goal of my PhD project was then to provide this missing diversity through structure-guided approaches and either protein engineering or biodiversity mining.

Chapter I first introduces the approaches in enzyme discovery, evolution or design followed by a comprehensive but not exhaustive description of the biocatalytic solutions for reductive amination, that have greatly expanded during the last two decades. This chapter enables to put our strategy into the context of what is currently done for the biocatalytic amine synthesis.

A deeper structural analysis and partial mechanism "resolution" of the native AmDH family are provided in Chapter II along with a complementary characterization of their biocatalytic potential in the reductive amination of key carbonyl substrates. A very preliminary study of the reductive sulfidation activity of some AmDHs is also described at the end of this Chapter.

Several approaches of evolution were undertaken to improve the AmDH activity with a focus put on the expansion of their substrate scope. Described in Chapter III, it comprises site-directed mutagenesis on two identified hot spots for active site enlargement, *in vivo* evolution using the automated devices available at the Genoscope and eventually the attempts to form an AmDH chimera that would combine both (thermo)stability and activity on carbonyl substrate free from any functional requirement.

Using all these acquired structural and biochemical knowledge on the native AmDHs, a deep exploration of genomic and metagenomic resources and selection of novel native sequences are described in Chapter IV. Ultimately, this could result in the discovery of AmDHs natively bearing target features of enlarged active site to accept a broader range of both carbonyl and amine substrate as well as reverse stereoselectivity.

Except when indicated, I was actively involved in all the studies and experiments described in this manuscript. All the experimental protocols are detailed in the Experimental section.

CHAPTER I

Introduction to strategies for discovery and improvement of enzymes, state-of-the-art of the biocatalytic synthesis of amines

I. Strategies for research and improvement of biocatalysts

The research in biocatalysis has always been driven by the necessity to find increasingly efficient biocatalysts to conduct given chemical reactions and be competitive against conventional chemistry in a challenging global economic context. From the enzyme discovery to the application in large-scale synthesis of the product of interest, multiple steps require the involvement of diverse expertise (Figure 1) (Wiltschi *et al.*, 2020). In this part, the light is shed on the enzyme discovery, selection and engineering steps.

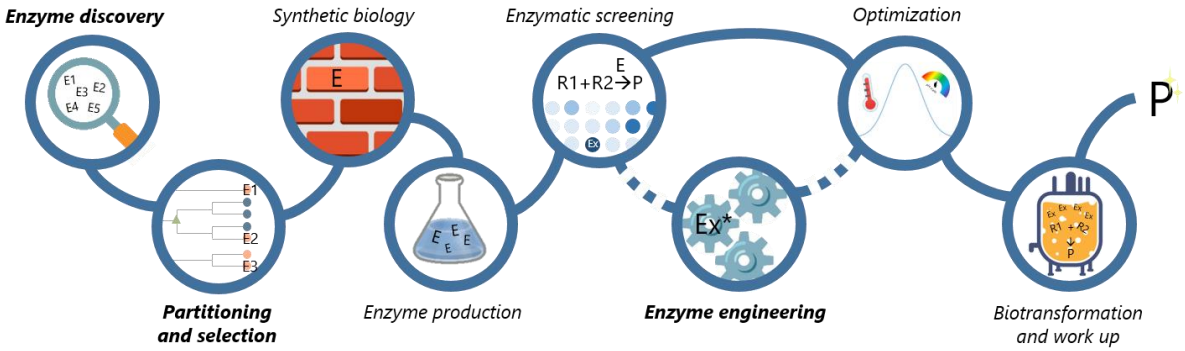


Figure 1. Schematic research steps in biocatalysis from enzyme discovery to product recovering. The steps labeled in bold are described in this manuscript. E = enzyme; R1 = reactant 1; R2 = reactant 2; P = product. Inspired by Wiltschi *et al.*, 2020.

The number of approaches has multiplied over the years and can be sorted in three categories: the discovery of native enzymes within the biodiversity, the improvement of the already reported sequences and the *de novo* enzyme design. These are described in this part as a non-exhaustive list of approaches and techniques reported thus far.

I.1 Discovery of novel enzymes

With the increasing sequencing capacities, new enzyme sequences can be retrieved in a growing number of resources from the fields of genomics (Zaparucha *et al.*, 2018) to metagenomics (Handelsman,

2004), metatranscriptomics (Warnecke *et al.*, 2009), (meta)proteomics (Verberkmoes *et al.*, 2009) or metabolomics (Dunn *et al.*, 2013).

I.1.1 Activity-based strategies

Traditionally, enzymes were isolated after enrichment selection of microorganisms by growth under limiting conditions followed by sequential purification steps. First being time-consuming and not highly successful, it was greatly simplified by approaches based on enzyme overexpression thanks to the development of recombinant desoxyribonucleic acid (DNA) techniques. To access the enzymes of the larger part of non-cultivable organisms, the complete genomic DNA of complex environments has led to the generation of metagenomes libraries. The latter can be screened by activity-based assays that rely on the identification of clones that exhibit a particular phenotype *e.g.* antibiotic resistance, colorimetric or fluorometric features. Despite being limited to reactions that can be screened rapidly with sometimes the need of specific high-throughput screening equipment, the activity-guided screening is still one of the most effective methods for sequence-independent enzyme discovery (Robinson *et al.*, 2021). This is also permitted thanks to the sequencing of entire metagenome libraries that can be carried out in increasingly reasonable price and timelines (Behrens *et al.*, 2011). It has been extensively used for the expansion of industrially relevant enzyme families such as lipases or cellulases (Berini *et al.*, 2017).

I.1.2 Sequence-based strategies

I.1.2.1 Sequence retrieval

I.1.2.1.1 PCR-based strategies

The polymerase chain reaction (PCR) is a method to amplify specific DNA segments. PCR-based research strategies enable the retrieval of proteins of interest using degenerate primers for the amplification of conserved gene domains, particularly from environmental DNA (eDNA). These are highly sensitive, relatively inexpensive and can be used in high-throughput screenings when using pooling and deconvolution steps (Hrvatin & Piel, 2007; Owen *et al.*, 2013). However, the method itself relies on the conservation of DNA patterns in target sequences and are not adapted to detect novel enzyme folds (Robinson *et al.*, 2021) and amino acid sequences with low homology compared to activity-based strategies. These methods are still highly applicable in metagenome exploration and were, for example, used for the discovery of new laccases from marine samples. The primers were designed and degenerated to target the conserved copper-binding site region of laccases (Fang *et al.*, 2011).

I.1.2.1.2 Sequence similarity-based strategies

Unlike the previously described method, one of the most popular discovery strategies relies on the overall sequence similarity of target sequences with reference enzymes exhibiting the activity of interest. The exploration is conducted within public genomic databases (UniprotKB, GenBank, etc.) by feeding an alignment tool such as Basic Alignment Search Tool (BLAST) (Altschul *et al.*, 1990; Boratyn *et al.*, 2013) or Double Index Alignment of Next Generation Sequencing Data (DIAMOND) (Buchfink *et al.*, 2015) with reference enzyme sequences as queries for pairwise protein sequence alignment. The sequences retrieved can then be partitioned according to their sequence similarity, phylogeny, genomic context or structural features to afford a more comprehensive view of the family and help the selection of representative sequences of hypothesized iso-functional groups. This strategy is particularly efficient to find enzymes with similar catalytic activity with the query sequences but displaying higher stability, switching in stereoselectivity or complementary substrate scopes (Pyser *et al.*, 2021; Zaparucha *et al.*, 2018). Among a wide variety of examples, one can note the work done on the nitrilase family (O. Kaplan *et al.*, 2006; Ondřej Kaplan *et al.*, 2013; O'Reilly & Turner, 2003; Vergne-Vaxelaire *et al.*, 2013; Salwan *et al.*, 2021). This strategy is the one used by the cloning and screening platform of biocatalytic activities of the UMR Genomics Metabolics and has used for the discovery of AmDHs among biodiversity described in Chapter I.

I.1.2.1.3 Key motif-based strategies

Enzyme families gather enzyme sequences sharing characteristic domains or conserved sequences. InterPro database provides such information with a functional analysis of proteins based on a classification into motif families, domain and important sites prediction including Hidden Markov Model (HMM). Compared with the sequence similarity strategies, these shared protein signatures can be used for the discovery of new enzymes harboring the essential motifs/residues of the (super)family they belong but with higher dissimilarity with the already known enzymes (Zaparucha *et al.*, 2018). The approach can target either a cofactor binding site, binding residues or the active site residues. One can cite the seminal work on the identification of novel Baeyer-Villiger monooxygenases (BVMOs) based on the protein sequence motif [FXGXXXHXXXW(DP)] reported by Fraaije *et al.* (Fraaije *et al.*, 2005).

I.1.2.1.4 Profile-sequence-based strategies

Multiple sequence alignment can be used to determine a sequence profile that would attest the amino acid frequencies. The definition of a sequence profile has led to a great improvement in the sensitivity of the profile-sequence comparison methods over sequence-sequence ones for the protein homology detection. Subsequently, several programs have been developed based on profile-profile comparisons to enable an even greater sensitivity in homology recognition [Local Alignment of Multiple-

Alignments (LAMA), PROF_SIM, Comparison of Multiple Protein Alignments with Assessment of Statistical Significance (COMPASS)] (Söding, 2005). The HMM profiles consider the amino acid frequencies, as other simple sequence profiles, but also include position-specific probabilities for insertions and deletions along the alignment. This search strategy is particularly useful for the identification of more remote homologs as illustrated by the discovery of new glycosyltransferase genes in genomes combining HMM profiles and fold recognition strategies (Hansen *et al.*, 2010; Robinson *et al.*, 2021). Inherently to its encoding, HMM search (Söding, 2005) is much quicker than a traditional pairwise sequence comparison, even if DIAMOND BLASTP (Buchfink *et al.*, 2015) has recently provided much quicker BLAST-type approaches.

I.1.2.2 Retrieved sequences analysis and partitioning

I.1.2.2.1 Phylogenetics

Phylogenetic trees are commonly used tools to represent the relationship between homologous proteins and the changes that occurred during their evolution (Pyser *et al.*, 2021; Rokas, 2011) (Figure 2). The analysis of the tree can lead to the discovery of uncharacterized enzymes gathered in a distinct phylogenetic clades compared to the already characterized catalysts and are therefore likely to feature new functions, substrate scope, etc. The phylogeny of an enzyme family can also allow the researchers to "move back in time" to identify and reconstruct ancestral protein sequences that are often postulated to afford higher (thermo)stability and substrate promiscuity (Pyser *et al.*, 2021; Siddiq *et al.*, 2017; Wheeler *et al.*, 2016).

Figure 2. Simplified phylogenetic tree depicting the potential of this approach to reach ancestral enzymes and new modern enzymes. Adapted from Pyser *et al.*, 2021.

I.1.2.2.2 Sequence Similarity Network

Sequence similarity network (SSN) tool was first described in 2003 to overcome the limit of phylogenetic representation for large protein families (Shannon et al., 2003; Pyser et al., 2021). It allows the visualization of relationships among large sets of protein sequences in two steps: an all-by-all BLAST alignment followed by the sequences filtering into clusters using web-based tool Enzyme Function

Iterative-Enzyme Similarity Tool (EFI-EST) (Gerlt *et al.*, 2015) and the free-to-download software Cytoscape (Shannon *et al.*, 2003). The nodes represent a protein sequence or group of sequences that are connected by edges corresponding to a user-defined similarity threshold (Figure 3). This approach is particularly useful when looking for new enzymes from families with a limited number of characterized proteins as illustrated by the work of Fisher *et al.* for the identification of novel flavin-dependent halogenases (Fisher *et al.*, 2019) and Copp *et al.* with the nitroreductase superfamily (Copp *et al.*, 2018) (Figure 3).

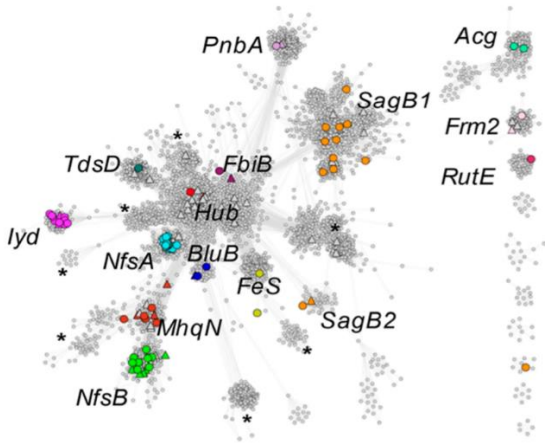


Figure 3. Sequence Similarity Networks of the nitroreductase superfamily. The network displays 5995 nodes gathering the sequences sharing > 50% identity. Edges correspond to an E-value threshold of 1 x 10^{-22}. Nodes containing characterized biochemical or physiological functions are colored and labeled. Extracted from Copp *et al.*, 2018.

I.1.2.2.3 Genomic context

The investigation of the gene context is a less popular approach but remains highly attractive especially for natural product biosynthesis. Indeed, the knowledge of the genetic neighborhood of a target gene can provide useful information about the enzyme natural substrates, cofactors and activity. Zhao *et al.* reported the structure- and genomic context-based prediction of the substrate scope of several structurally characterized enzymes with unknown functions (Zhao *et al.*, 2013). The correct assignation of functions to proteins remains one of the greatest challenges that will require tremendous efforts in experimental characterization and computational research (Pyser *et al.*, 2021). Analysis of genomic context can drastically help in this task, especially in identification of metabolic pathways. Such analyses can also be highly effective to search for a particular activity by looking in the neighborhood of postulated former or later activities. Tools to classify proteins of a family based on their conserved genomic contexts have been developed; one can cite NetSyn created by the Laboratory of Bioinformatics Analyses for Genomics and Metabolism (LABGeM) of the UMR Genomics Metabolics.

I.1.3 3D structure-based strategies

I.1.3.1 Protein modeling

The three dimensional (3D) structure-guided methods need to access 3D structures of both query and target sequences. Only a few years ago, this was mainly permitted through structure resolution and subsequent homology modeling.

One of the most popular methods for protein structure resolution consists in the controlled precipitation of a concentrated protein solution in a saturating medium containing precipitating agents to form crystals of an adequate size for diffraction measurements using X-ray crystallography. However, it remains a time-consuming and empirical step with rather poor possibilities of sequence-based rationalization (McPherson & Gavira, 2014). On a lesser extent, protein structure can also be obtained using Nuclear Magnetic Resonance (NMR) spectroscopy or Electronic Microscopy (EM) (Jonic *et al.*, 2009; Takashima, 2006). In 2019, X-ray crystallography and NMR spectroscopy accounted for 89% and 8% of the total structures from the Protein Database (PDB), respectively, but Cryo-EM is becoming an increasingly attractive method (Schirò *et al.*, 2020). According to the Research Collaboratory for Structural Bioinformatics (RCSB) PDB, among the released PDB structures of 2021, 74% were obtained from X-ray crystallography, 3% from NMR and 23% from EM (RCSB PDB website).

If some validated crystallographic data are available for a given enzyme and all the more for a family, it is possible to rapidly generate a 3D model by homology with the template using, for example, the free-to-access SwissModel website (Waterhouse *et al.*, 2018) or MODELLER software (Fiser *et al.*, 2000; Fiser *et al.*, 2003; Martí-Renom *et al.*, 2000; Sali *et al.*, 1993; Webb *et al.*, 2016).

Prediction of 3D-structure *ab initio* could also be carried out but with low accuracy, thanks to the Iterative Threading Assembly Refinement (I-TASSER) (Yang *et al.*, 2015), an integrated platform based on the sequence-to-structure-to-function paradigm, or Rosetta, a fully automated structure prediction server (Ovchinnikov *et al.*, 2018). In 2021, DeepMind developed a novel protein structure prediction algorithm, AlphaFold2, which outperformed any other tool reaching near experimental accuracy (Jumper *et al.*, 2021). The code is now open source and an easy-to-use version is available under the name ColabFold for accelerated enzyme structure prediction compared to the full AlphaFold2 package (Mirdita *et al.*, 2022). All the protein sequences reported in UniprotKB are now supplied with their predicted 3D structure. Since the success of AlphaFold2 algorithms, many other ones has been released (Appendix 2). Another predictive algorithm was described by AlQuraishi and relies on Recurrent Geometric Networks (RGN) to predict protein structures in milliseconds with as high accuracies (AlQuraishi, 2019; Chowdhury *et al.*, 2022).

Thanks to these models and/or structures, different strategies can be undertaken to search for enzymes based on specific part of the enzyme or on the conservation of the overall structures and folds, such as PDBeFold (Krissinel & Henrick, 2004). The following paragraph details briefly only the more recent ones used for biocatalysis purpose.

I.1.3.2 Constellation-guided strategies

This strategy consists in the determination of an active site configuration that allows the reaction to proceed and its screening into databases. Some algorithms were reported in this way such as TESS and Jess (Barker & Thornton, 2003; Wallace *et al.*, 1997). Later on, Steinkellner *et al.* optimized the approach by considering only a minimum active site required to perform the desired catalytic reaction among the characterized enzymes. The nature of the selected residues as well as their position to one another are fixed and screened among the available datasets of 3D structures. Ultimately, this can result in the discovery of new proteins with similar catalytic function but completely different folds. As a proof-of-concept of their bioinformatic method, Steinkellner *et al.* predicted a promiscuous ene-reductase activity in two enzymes with comparable activity and binding modes as the Old Yellow Enzymes but with an inverted stereopreference (Steinkellner *et al.*, 2014). The patented Catalophore^{TM} strategy has led to the spin-off company Innophore which provides solutions to rapidly identify enzymes with desired properties (Innophore website). However one limit is the requirement of a good comprehension of the reaction mechanism as well as the role of the active site residues. This strategy was used during my PhD project by Dr. Eddy Elisée within the MODAMDH project and is described in Chapter IV, II.4.

I.1.3.3 Machine Learning/Computationally-aided strategies

In recent years, machine learning established as the spreadhead for the resolution of the relationship between protein sequences, structures and functions. AlphaFold2 algorithms already provided great advances in protein structure prediction and >35 machine learning-based methods have been reported for protein function prediction (Bonetta & Valentino, 2020; Robinson *et al.*, 2021). Although considered as a powerful method, one of the major disadvantages of machine learning is its extreme hunger for quantitative and qualitative data that brings the researchers community to share their data on public databases.

I.2 Enzyme improvement

As mentioned in the introduction of this chapter, another strategy to enrich the catalog of enzymes is to modify the corresponding gene sequence to enhance a specific required feature. Various approaches have been developed and the pioneer researcher in this field is Pr. Frances Arnold who received the Nobel Prize of Chemistry 2018 for her achievement in conducting the first directed evolution

of enzymes in 1993 (K. Chen & Arnold, 1993).

I.2.1 Random mutagenesis strategies

Random mutagenesis strategies are useful when lacking of structural data or prior directed evolution on hypothetically similar enzymes. Several techniques are now well described to randomly damage DNA (Packer & Liu, 2015):

- The traditionally used chemical or physical mutagenesis using alkylating or deaminating compounds, base analogs or ultraviolet irradiation.
- The use of mutator strains containing deactivated proofreading and repair enzymes to increase the rate of errors during DNA replication. To avoid deleterious mutagenesis on the host genome, Ravikumar *et al.* developed an orthogonal *in vivo* DNA replication system that only mutate the target DNA (Ravikumar *et al.*, 2014).
- Error-prone PCR (epPCR) that relies on the low fidelity of DNA polymerases when used under non optimal conditions (high magnesium concentrations, manganese supplementation, use of deoxynucleotide triphosphate (dNTP) analogs, etc.) during the PCR amplification of the gene of interest. Some processes, such as the fusion of the target gene with a Green Fluorescent Protein (GFP) reporter can lead to neutral drift libraries devoid of deleterious mutations for protein expression (Gupta & Tawfik, 2008).
- Sequence saturation mutagenesis (SeSaM) is one of the most common saturation mutagenesis technique (Wong *et al.*, 2004). Developed by Wong *et al.*, it relies on the insertion of the universal base deoxyinosine along the target gene to reduce the imbalance in the nucleotide mutational bias obtained using epPCR.

These techniques are rather easy and not expensive to set up for the molecular biology steps but result in the generation of large libraries of mutants (10^5 - 10^8 variants) that requires (ultra) high-throughput screening methods and equipment to discriminate the beneficial mutations (Bell *et al.*, 2021; Packer & Liu, 2015).

I.2.2 Focused mutagenesis strategies

When prior structural or biochemical data are available within an enzyme family, it is possible to rationalize the mutagenesis work by targeting only a limited number of residues in the sequence. This allows to drastically reduce the screening effort to 10^2 – 10^4 variants but requires substantial prior *in silico* studies with, for example, molecular docking to understand the substrate or cofactor positioning and identify the hot spots. Among the various techniques of focused mutagenesis strategies, one can site the site saturation mutagenesis (SSM) that uses degenerate codons NNN, NNK or NNS to randomize

the mutation on selected hot spots into the 20 natural amino acids (Bell *et al.*, 2021; Packer & Liu, 2015). As the number of mutants exponentially increases with the number of randomized sites, a good strategy is to only introduce specific amino acid substitutions based on structural rationalization or phylogenetic analyses of homologous sequences. Combinatorial active-site saturation test (CAST), introduced by Reetz *et al.* in 2005, consists in the systematic design and screening of active site saturation libraries carried out in several rounds to reach a synergistic effect of the combined beneficial mutations (Reetz *et al.*, 2005). Another variant method of multi step focused mutagenesis strategies is the Iterative Saturation Mutagenesis (ISM) that comprises repeated cycles of multisite saturation mutagenesis and selection of the best variants (activity, substrate scope, thermostability, solvent tolerance, etc.). At each round, the substituted sites are different until reaching the required improvement of the biocatalyst (Acevedo-Rocha *et al.*, 2014). The group of Reetz and co-workers coupled an ISM strategy with the monitoring of the B factor for variant selection at each round to increase the thermostability of a lipase (Reetz *et al.*, 2006).

Various other methodologies have been developed combining the previously described strategies with integration of extensive prior studies to keep reducing the number of variants to screen such as Focused Rational Iterative Site-Specific Mutagenesis (FRISM) that integrates molecular docking, consideration of consensus data with homologous proteins, predicted B-factors, etc. (Li *et al.*, 2020).

Eventually, site directed mutagenesis (SDM) is the most controlled mutagenesis strategy as it allows the modification of a target residue into one specific amino acid leading to only one mutant (Carter, 1986). The screening libraries are therefore considerably enlighten but the rate of success depends upon a good comprehension of the protein structure.

I.2.3 Diversification by homologous recombination

The naturally occurring homologous recombination is a process that can be mimicked to access beneficial coupling of mutations through DNA reassortment of a diverse panel of functional genes (Packer & Liu, 2015).

Among the fragmentation-based recombination methods, one of the most popular is the DNA shuffling, first described by Stemmer *et al.* in 1994 (W. P. Stemmer, 1994; W. P. C. Stemmer, 1994). It enables the genes fragmentation using a DNA ligase and the random annihilation of the fragments in a primer-free PCR reaction. Coco described the related method of Random Chimeragenesis on Transient Templates (RACHITT) combining the same desoxyribonuclease (DNase)-mediated genes fragmentation with a reassembly protocol using an uracil-containing scaffold for temporary fragments annealing (Coco,

2001). Derived newer methods provide a better control of the rearrangement such as the Nucleotide Exchange and Excision Technology (NExT) (Müller *et al.*, 2005).

Homologous recombination can also be performed without any gene fragmentation but by working on the PCR protocol. For instance, the Staggered Extension Process (StEP) prematurely interrupts the elongation step by heat denaturation to force the incomplete extension products to anneal with another template to create a recombination (Zhao *et al.*, 1998). The assembly PCR, also known as assembly of designed oligonucleotides (ADO) or synthetic shuffling, uses the multiplication of mutation-containing PCR primers that overlap during the extension step until reaching a full-length recombinated gene (Stemmer *et al.*, 1995).

In contrast to the previously described techniques, some recombination methods do not rely on sequence homology between the recombinant genes and enable the shuffling of distinct genes, especially for protein families with similar function but divergent sequences. It is the case for the Sequence Homology Independent Protein Recombination (SHIPREC) (Udit *et al.*, 2003), the Incremental Truncation for the Creation of HYbrid (ITCHY) method based on the incremental truncation and fusion of two distinct genes (Ostermeier *et al.*, 1999), or the non-homologous random recombination (NRR) to introduce substantial rearrangements of domain topology within functional proteins (Bittker *et al.*, 2004; Packer & Liu, 2015).

I.2.4 Computationally-assisted strategies

Many tools are now available to guide the mutagenesis work for enzyme improvement. To support the need for enzymes with improved activity or substrate promiscuity, one can cite the package of webservers developed by Damborsky and co-workers. This includes HotSpot Wizard for the design of smart variant libraries (Pavelka *et al.*, 2009), CAVER (Jurcik *et al.*, 2018) and CaverDock (Vavra *et al.*, 2019) for the analysis of tunnel dynamics and identification of the residues forming tunnel bottlenecks. Similarly, Fleishman and co-workers developed the webserver FuncLib for automatic design of multipoint mutations for enzyme activity improvement targeting the active site based on phylogenetic analysis and Rosetta design calculations (Khersonsky *et al.*, 2018). The reconstruction of ancestral sequences can also guide the mutagenesis work for substrate scope expansion as illustrated by Chen *et al.* on DNA polymerases using Reconstructed Evolutionary Adaptative Path methodology (REAP) (Chen *et al.*, 2010). More recently, Osuna and co-workers described the use of protein conformation landscapes to generate a Short Path Map (SPM) tool to predict distal and active site mutations (Bell *et al.*, 2021; Osuna, 2021).

To address the problem of protein stability, one of the earliest methodologies was the consensus approach using the information provided by alignments with mesophilic enzyme sequences to allow

targeted and rapid design of thermostabilized and functional variants as illustrated by the work of Lehmann *et al.* on hydrolases (Lehmann *et al.*, 2000). Nowadays, a large panel of tools based on theoretical calculations and predictions are available such as FireProt for the design of thermostable multiple-points mutants (Musil *et al.*, 2017) and FireProt^{ASR} for reconstruction of ancestral sequences (Musil *et al.*, 2020). Damborsky and co-workers also provide SoluProt tool for the prediction of enzyme solubility (Hon *et al.*, 2021). Among others, one can cite the Rosetta algorithms-based tools such as Protein Repair One Stop Shop (PROSS) for the design of proteins with increased bacterial expression and stability (Goldenzweig *et al.*, 2016) or the Framework for Rapid Enzyme Stabilization Computational Libraries (FRESCO) (Wijma *et al.*, 2014).

The homologous recombination experiments described in the previous part can also be computationally guided as illustrated by the work of Arnold and co-workers who used the SCHEMA algorithm to identify the fragments of β-lactamases to recombine without disturbing the integrity of the protein fold (Voigt *et al.*, 2002).

I.2.5 I*n vivo* evolution

The *in vivo* evolution of biocatalysts offer ultra-high-throughput screening platforms (10^6 - 10^{10} variants) applicable to the systems that can relate the catalytic improvement with cell viability. They were traditionally carried out through serial transfers of cultures into simple flasks or gel plates but can now be implemented into automated devices that will maintain the continuous evolution of the cells (Barrick & Lenski, 2013). The set up of such evolution systems is extremely sensible to the stringency of the selection pressure and the ability of the host organism to evolve through alternative mechanisms of survival (Bell *et al.*, 2021). Hence, they may take years to optimize as it is illustrated in this manuscript with the attempt of *in vivo* evolution of enzymes performing the reductive amination of carbonyl substrates. Nonetheless, some key successes were reported such as the evolution of pyrrolysyl-transfer ribonucleic acid (tRNA) synthetases for the genetic incorporation of non-canonical amino acids into protein sequences (Wang *et al.*, 2001). This *in vivo* evolution could be coupled to Multiplex Automated Genome Engineering (MAGE) to target, as in *in vitro* focused mutagenesis strategies, specific part of sequences and so enhancing the success of *in vivo* evolution (Wang *et al.*, 2009). MAGE allows simultaneous mutagenesis of multiple targeted sites in bacterial genomes using short nucleotides.

I.2.6 Introduction of non-canonical amino acids

Although the mutational possibilities within one protein sequence can lead to gigantic mutant libraries, these are still "restricted" to the 20 canonical amino acids bearing a limited number of functional groups. Thus, the use of an extended amino acid alphabet [selenocysteine (**6**), pyrrolisine (**7**), etc.] or the

introduction of post-translational modifications, such as phosphorylation, sulfonation or glucosylation, currently represent exciting new possibilities for biocatalysis. The strategies to incorporate such mutations into unnatural amino acid analogs are not deeply described in this manuscript (Bell *et al.*, 2021; Kirk *et al.*, 2018; Ravikumar *et al.*, 2015). One notable example is the work of Hoes *et al.* on a lipase from *Thermoanaerobacter thermohydrosulfiricus* who achieved a 10-fold increasing in enzymatic activity through the replacement of Met residues along the enzyme sequence with the norleucine (**8**) (Hoesl *et al.*, 2011). The latter exhibiting a greater hydrophobicity than the natural Met, this could enhance the interactions with the hydrophobic lipids.

I.3 De novo enzyme design and artificial enzyme

I.3.1 Enzyme design

Diverse scales of enzyme designs have been reported to access catalytic function that have not been found in natural protein scaffolds. The creation of new enzymatic function on demand in protein harboring a different fold can be performed using structure-based computational protein design (CPD) algorithms. These have been developed for the identification of enzyme sequences that can adopt a specific 3D-fold and display desired properties (stability, binding affinity, specificity, etc.). Applied to the computational enzyme design, the steps consist in the modeling of a minimum active site of side chains required for the reaction and that could accommodate a high reaction intermediate, such as a transition state of the molecule of interest, the selection of the scaffold to contain it, the energy optimization of the created models and the ranking of the resulting designs. The challenges remain on the ability of CDP algorithms to create a binding pocket to accommodate the desired substrate, introduce the appropriate catalytic machinery to perform the reaction of interest while maintaining the overall structure integrity of the template fold (St-Jacques *et al.*, 2018). A key reported success was the *de novo* design of Kemp eliminases for the proton elimination in the 5-nitrobenzisoxazole (**9**) (Röthlisberger *et al.*, 2008; Zanghellini, 2014). One can also cite the *de novo* enzyme Syn-F4 which catalyzes the hydrolysis of ferric enterobactin (**10**) *in vitro* and *in vivo* (Donnelly *et al.*, 2018).

I.3.2 Artificial Metalloenzymes

The design of artificial metalloenzymes is at the interface of biocatalysis and metal catalysis to insert novel functionalities in protein and overcome the low efficiency of enzymes towards non-natural substrates. Comparably to the abovementioned *de novo* enzyme design, this approach involves the incorporation of a catalytically active metal (complex) in a protein scaffold (Wiltschi *et al.*, 2020). This can be performed through the direct insertion of an inorganic metal ion, the supramolecular anchoring of a metal cofactor or the covalent binding of a metal cofactor protein. In 2016, Okamoto *et al.* set up a

regeneration system of NADH mimics using an artificial metalloenzyme designed from streptavidin variant templates (Okamoto *et al.*, 2016).

I.3.3 Bioinspired chemistry

Reversely to the artificial metalloenzyme design that incorporate a desired catalytic site into a protein template, some researchers try to mimic the enzyme active site to form a catalytic system with the same activity. Prof. Fontecave is one of the French leaders of such a bioinspired chemistry. With his team, they used hydrogenases to design metal-free catalytic nanomaterials for H_2 production (Le Goff *et al.*, 2009).

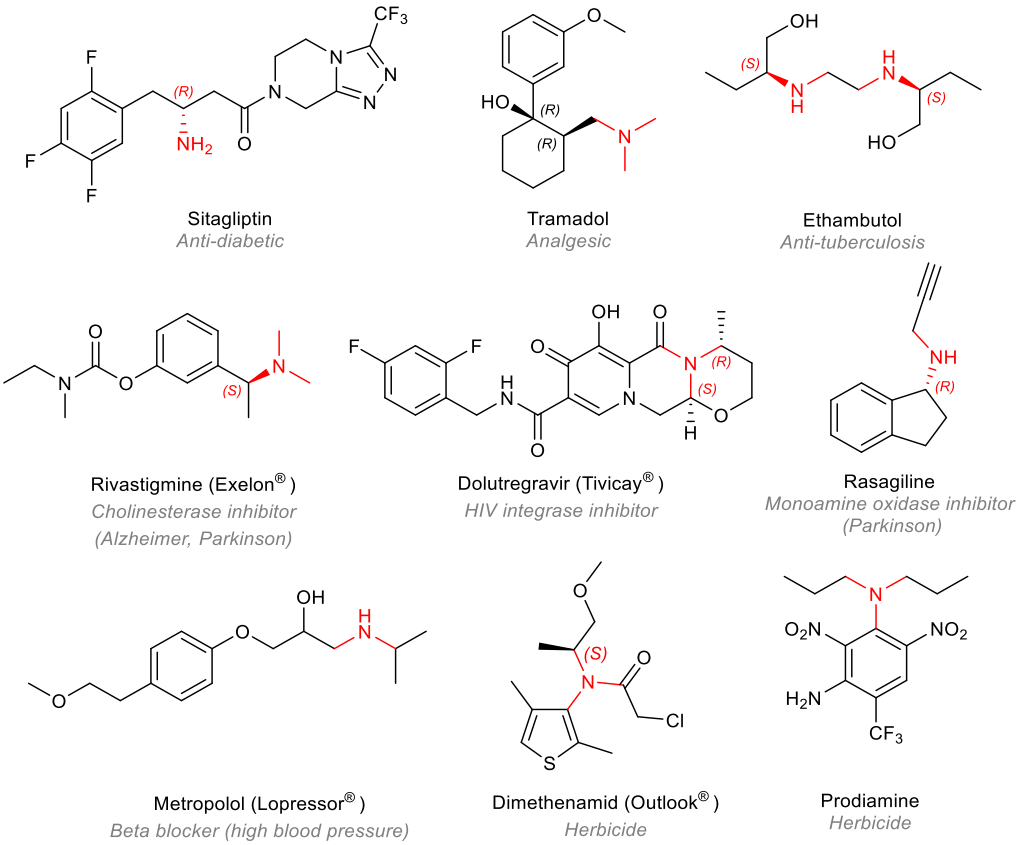
I.4 Conclusion on the state-of-the-art on enzyme discovery, improvement and design

The three approaches of enzyme discovery, improvement and design, briefly described in this part, comprise a growing number of techniques and variants to have access to increasingly efficient biocatalysts for synthetic applications. First relying on the screening of enormous enzyme or variant libraries, they progressively integrated computational strategies by taking advantage on the great advances in structural prediction and machine learning processes. The next part focuses on the biocatalytic solutions available for amine synthesis found thanks to some of the abovementioned approaches.

II. Biocatalytic synthesis of amines

II.1 Amine compounds

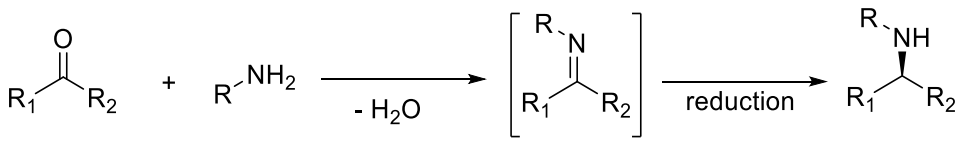
Amines are ubiquitous in natural compounds as it occurs in amino acids, proteins, living tissues, etc. In this paragraph and in this work, we focused on amine as a nitrogen atom having adjacent sp3 carbon or hydrogens, so it excludes amino acids and derivatives. With a high propensity to form hydrogen bonds, they are key intermediates in many fields of chemical industry such as, among others, pharmaceuticals, crop protection molecules, detergents, lubricants or polymers. In the drug industry, they are estimated to be present in more than 90% of current top-selling or newly approved small molecule drugs (Scheme 1). Chiral amines, restricted in this manuscript to "α-chiral amine", represent the major part but the overall market for achiral and racemic amines represent the largest volumes and values (Breuer *et al.*, 2004; Nugent & El-Shazly, 2010; Wu *et al.*, 2021). In the polymer market, they are key monomers for the synthesis of polyamides, polyureas or polyepoxydes, highly valuable for applications in automotive, aerospace, building or health (Froidevaux *et al.*, 2016).



Scheme 1. Structure of some pharmaceuticals and agrochemicals containing an amine function.

In conventional organic chemistry, many routes are available for the synthesis of amine compounds and are mainly based on nucleophilic additions or substitutions and reductive reactions. Among the most typical ways, we can highlight the nucleophilic substitution of alkyl halides (Br, I or Cl), the reduction of nitrogen-containing functional groups (amides, nitriles, nitro, azides) or the reductive amination of carbonyl compounds with primary, secondary amines or ammonia. Except for the latter that is our reaction of interest for which we intend to find biocatalytic solutions, the chemical synthesis of amines is not described in this manuscript. The reductive amination is one of the most versatile and common approaches for amine synthesis from a carbonyl-containing substrate (Scheme 2). The iminium intermediate formed is reduced by a reducing agent which can also introduce the stereochemistry in the case of prochiral ketones. Methods are described with ammonia (**11**), for example in presence of Ti complexes, for the synthesis of primary amines but generally, the amine source is a substituted amine RNH_2 to form secondary amines and the obtention of the corresponding primary amine requires the cleavage of the R moiety. The most popular reducing agents comprise borohydride complexes and hydrogenation with hydrogen gas. The stereoselective amine formation requires the use of transition

metal (Rh, Ru, Ir and Ti) coupled with chiral ligands or organocatalysis, thus giving rise to multi-step processes (Oleg *et al.*, 2019; Ghislieri & Turner, 2014).



Scheme 2. Reductive amination reaction.

The conventional approaches of amine synthesis require the use or go along with the formation of toxic and/or dangerous compounds, proceed in large fractions of unsustainable organic solvents and function at high temperature and pressure conditions. In a global context of energetic transition, it was therefore necessary to set up more eco-friendly and efficient processes for amine synthesis. A good alternative was found in catalysis, including in biocatalysis, a field that has been on the rise during the last few decades. In the case of chiral amine synthesis, biocatalysis is of particular interest as it generally provides higher enantiomeric purity than the conventional chemical processes.

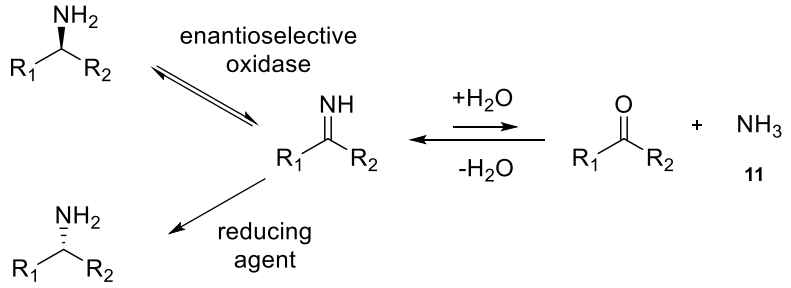
II.2 Biocatalysts for amine synthesis

Many biocatalysts are already reported to catalyze the amine formation from five out of the six enzymes classes: hydrolases (EC 3.), transferases (EC 2.), isomerases (EC 5.), lyases (EC 4.) and oxidoreductases (EC 1.) (Hall, 2021; Mutti & Knaus, 2021). Lipases (EC 3.1.1.3.) naturally catalyze the hydrolysis of ester into the corresponding alcohol and acid during the lipolysis and are used for kinetic resolution of racemic amines. Baden Aniline and Soda Factory (BASF) has optimized a lipase-catalyzed route to optically active amines that can be run at a scale of several thousand tons and led to the synthesis of many chiral amines (ChiPros®). Aminomutases (EC 5.4.3.) reversibly exchange a hydrogen bond with a neighbor amine function. Ammonia lyases (EC 4.3.1.25/26) catalyze the reversible hydroamination of ammonia (**11**) on a *trans* double bond located at the α position of an acid to form the corresponding amino acid. The most studied class of enzymes for amine formation, the transaminases (TA) (EC 2.6.1.X.), can reversibly aminate a ketone acceptor with the transfer of an amine function from a donor to the ketone "acceptor" (Gomm & O'Reilly, 2018). They represent one of the most well known success stories of biocatalysis usage in industry with the TA-mediated large scale synthesis of the anti-diabetic Sitagliptin (**2**) (Savile *et al.*, 2010; Zawodny & Montgomery, 2022). Other examples of all of these biocatalysts are already used or enough performant for the synthesis of natural compounds as well as pharmaceuticals, intermediates (Zawodny & Montgomery, 2022) and bulk chemicals (Groger, 2019). These alternatives enable to design new synthetic pathways for the implementation of more efficient and safer processes in the chemical industry.

In this part, the light is shed on the oxidoreductases that represent a key class of enzymes for the amine synthesis and comprise the amine dehydrogenases. Therefore, many reviews relate the latest advances in this field. Among others, one can cite Cheng *et al.*(2020), Cosgrove *et al.* (2018), Patil *et al.* (2018), Vidal *et al.* (2017), and one from our group (Ducrot *et al.*, 2021). The parts III.5.2, III.5.3 and III.5.4 have been largely inspired by this review published in *Advanced Synthesis & Catalysis* that I have written extensively in collaboration with Prof. Grogan and his PhD student, Megan Bennett. Yet, these parts have been updated with the latest developments of these families. More details were provided, especially on the global shape and active site structure of these enzymes, compared to the non oxidoreductase enzyme families as they represent the only examples of NAD(P)H-dependent enzymes that enable the asymmetric reductive amination of carbonyl, using an amine donor.

II.2.1 Monoamine oxidases

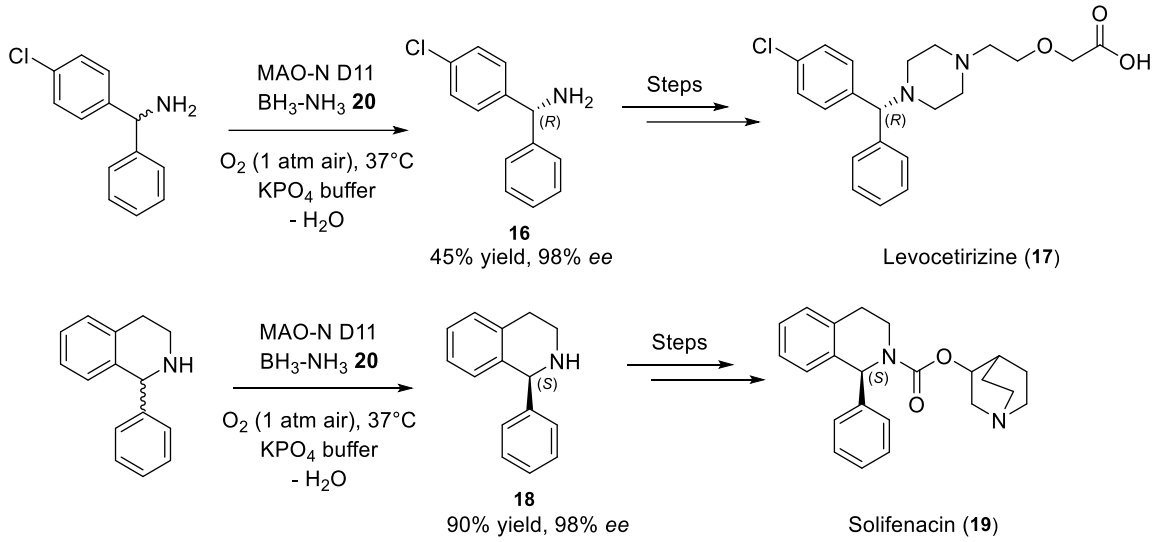
Monoamine oxidases (MAOs) (EC 1.4.3.4) are a class of enzymes catalyzing the oxidation of an amine function into the corresponding imine using molecular oxygen along with the generation of the hydrogen peroxide H_2O_2 (**12**) (Duan *et al.*, 2019). They have been classified into two groups, the Cu/TOPA-dependent MAOs (Type I) and the flavin-dependent MAOs (Type II). Originally, the type I was not used in biocatalysis as the imine formed stayed covalently bound to the protein. In 2002, Alexeeva *et al.* were the first to report the use of MAO-N from *Aspergillus Niger* in a new deracemization method. The stereoinversion of one enantiomer to the other was achieved by repeated cycles of MAO-N-mediated oxidation to the imine and its subsequent non selective reduction into the amine until reaching a majority of the desired enantiomer (Scheme 3) [Alexeeva *et al.*, 2002; Sablin *et al.*, 1998; Schilling & Lerch, 1995 (A); Schilling & Lerch, 1995 (B)].



Scheme 3. Monoamine oxidase-mediated deracemization process of chiral amines. Adapted from Alexeeva *et al.*, 2002.

This method was applied on the deracemization of 1-phenyl-1-ethylamine (**13**) for which the wild-type (WT) MAO-N displayed a very low activity for the oxidation of either (*R*) or (*S*)-enantiomer. The high-throughput screening of 150,000 variants, based upon the release of **12** and the use of a peroxidase and

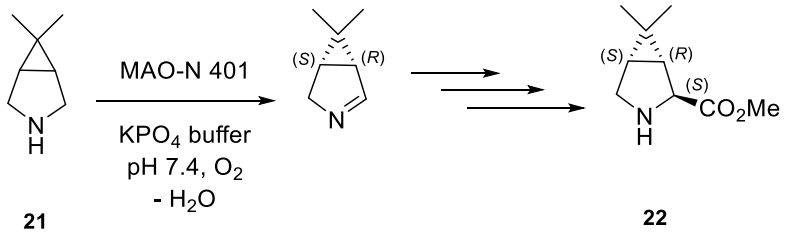
a highly colored product upon oxidation, led to the top single mutation N336S affording a 10-fold higher activity and 80-fold higher selectivity for the (*S*)-enantiomer as compared to the WT MAO-N. The use of MAO-N mutant in the deracemization of (1*S*)-1-phenylethan-1-amine (**14**), coupled with the ammonia-borane (**20**) complex as the reducing agent, led to the synthesis of (1*R*)-1-phenylethan-1-amine (**15**) in 77% analytical yield and 93% enantiomeric excess (*ee*). Thereafter, an enormous effort on further directed evolution of MAO-N was done to expand its substrate scope. This implied additional rounds of random mutagenesis using mutator strain (Carr *et al.*, 2005; Dunsmore *et al.*, 2006), saturation mutagenesis libraries targeting residues of the active site pocket and channel (Ghislieri *et al.*, 2013), iterative saturation mutagenesis using NDT codon degenerated (Li *et al.*, 2014), epPCR (Li *et al.*, 2012) and homology model-guided mutagenesis with recombination of positive mutants and family shuffling (Li *et al.*, 2016). Along with the number of directed evolution steps, substituents with increasing hindrance on both part of the amine could be accepted by MAO-N mutants. Therefore, it enabled the access to key building blocks such as 1-(4-chlorophenyl)-(1*R*)-phenylmethylamine (**16**) [Levocetirizine® (**17**)] in 45% yield and 97% *ee*, and (1*R*)-1-phenyl-1,2,3,4-tetrahydroisoquinoline (**18**) [(Solifenacin® (**19**)] in 90% yield and 98% *ee* in the MAO-N D11-mediated deracemization process using the reducing agent **20** (Scheme 4) (Ghislieri *et al.*, 2013).



Scheme 4. Synthesis of chiral amine intermediates using MAO-N D11 mutant towards the synthesis of Levocetirizine® (**17**) and Solifenacin® (**19**). Adapted from Ghislieri *et al.,* 2013.

Another striking reward for this mutagenesis effort was published by Li *et al.* with the mutant MAO-N 401 that was successfully used for the oxidative desymmetrization of 6,6-dimethyl-3-azabicyclo[3.1.0]hexane (**21**) in the large-scale production of (1*R*,2*S*,5*S*)-6,6-dimethyl-3-azabicyclo[3.1.0]hexane-2-carbonitrile (**22**), a key intermediate in the synthesis of Boceprevir (**23**)

(Victrelis®), in an overall 56% yield with > 99% *ee* (Scheme 5) (Li *et al.*, 2012; Zawodny & Montgomery, 2022). Other (*S*)-monoamine oxidases have been studied, such as the cyclohexylamine oxidase CHAO from *Brevibacterium oxydans* IH-35A and its variants (Leisch *et al.*, 2012; Li *et al.*, 2014).



Scheme 5. MAO-N 401-mediated desymmetrization for the synthesis of a key intermediate in the synthesis of Boceprevir (**23**). Adapted from Zawodny & Montgomery, 2022.

As no (*R*)-selective monoamine oxidase had been reported yet, in 2014, Yasukawa *et al.* engineered a flavin-dependent porcine kidney *D*-amino acid oxidase. After several round of saturating mutagenesis, the double variant Y228L/R283G displayed the highest activity towards **15** oxidation (18.2 U mg^{-1}) and further mutagenesis work was done based on the crystal structure obtained from this variant to reach a larger panel of (*R*)-amines (PDB: 3WGT) (Yasukawa *et al.*, 2018; Yasukawa *et al.*, 2014).

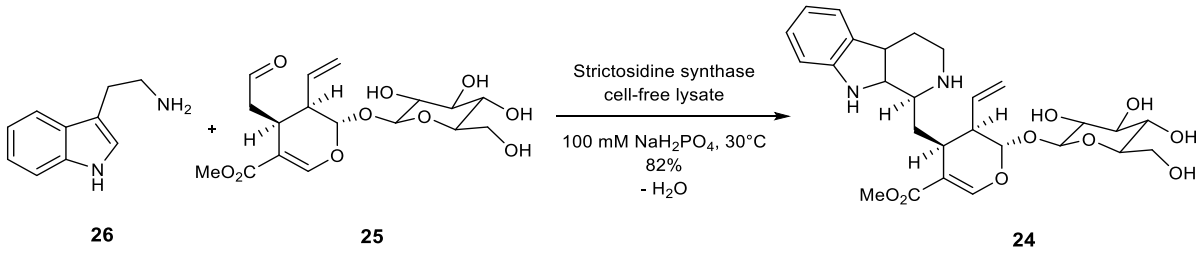
Another (*R*)-selective amine oxidase was developed by Heath *et al.* from 6-hydroxy-*D*-nicotinamide oxidase using CASTing approach for directed evolution (Heath *et al.*, 2014).

MAOs represent a good alternative for the resolution of chiral amines. Nevertheless, the amine must be introduced in previous steps. Thus, these enzymes are not appropriate for the formation of the amine function, unlike the following types of enzymes which meet the objective of my doctoral work that was focused on the amine formation from a carbonyl function, namely the reductive amination. Thus, more details of the state of the art of the following enzymes performing reductive amination are provided (Afanasyev *et al.*, 2019).

II.2.2 Pictet-Spenglerases

The Pictet-Spenglerases were named out the reaction of a β-arylethylamine condensation with a carbonyl and the subsequent ring closure discovered by Amé Pictet and Theodor Spengler in 1911. Their usage in biocatalysis is limited due to a narrow substrate scope and issues regarding enzyme isolation and stability (Roddan *et al.*, 2020). Nonetheless, a strictosidine synthase was recently used in the 10-step synthesis of L-strictosidine (**24**), a natural intermediate in the biosynthesis of numerous pharmaceuticals (Anthony *et al.*, 2021; Zawodny & Montgomery, 2022). Applied as a cell free lysate in

the final step, the enzyme could convert L-secologanin (**25**) and tryptamine (**26**) into the target **24** in 82% yield on a mg-scale (Scheme 6).



26 **25** **24**

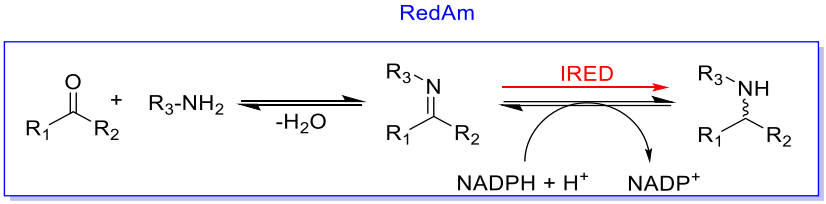
Scheme 6. Strictosidine synthase-mediated Pictet-Spenglerase last step reaction in the total synthesis of L-strictosidine (**24**). Adapted from Zawodny & Montgomery, 2022.

The other notable subclass of the Pictet-Spenglerase family is the norcoclaurine synthases which are involved in the biosynthesis of tetrahydroisoquinoline (**27**) by catalyzing the stereoselective C-C coupling of dopamine (**28**) and 4-hydroxyphenylacetaldehyde (**29**) to form (*S*)-norcoclaurine (**30**) (Sangster *et al.*, 2022). Due to the absence of known Pictet-Spenglerase active towards benzaldehyde (**31**) and its substituted derivatives, Eger *et al.* developed a strictosidine synthase from *Rauvolfia serpentina* (*Rs*STR) to enable their condensation with tryptamine (**26**), using beneficial mutations between different WT backbones (Eger *et al.*, 2020).

II.2.3 Imine reductases and reductive aminases

This paragraph of the review was mainly written by Megan Bennett and Prof. Gideon Grogan but I updated it with the latest developments reported since 2020.

Imine reductases (IREDs) are NADPH-dependent oxidoreductases that have been shown to catalyze the asymmetric reduction of pre-formed prochiral imines (Grogan & Turner, 2016; Mangas-Sanchez *et al.*, 2017). IREDs represent a good starting point for biocatalytic reductive amination as the imine reduction step constitutes the second half of the reductive amination reaction (Scheme 7) and a subset of these enzymes, reductive aminases (RedAms), have been shown to also catalyze the imine formation reaction.

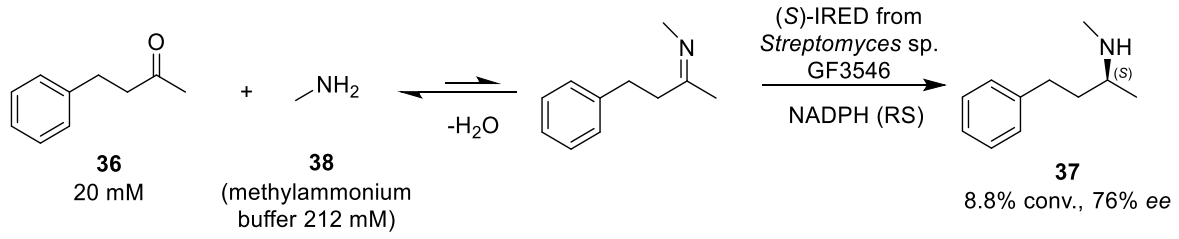


Scheme 7. General reactions catalyzed by IREDs and RedAms. IREDs catalyze the reduction of imines preformed in solution to give optically active amines. RedAms catalyze the formation of the imine from a carbonyl substrate and

amine donor and also the reduction of the imine. Extracted from Ducrot *et al.*, 2020.

II.2.3.1 First imine reductases catalyzing the reduction of imines

Stereocomplementary IREDs from *Streptomyces* spp. GF3587 and GF3546, with (*R*)- and (*S*)-selectivity respectively for the reduction of 5-methyl-3,4-dihydro-*2H*-pyrrole (**32**), were the first examples of IREDs used for a synthetic purpose (Mitsukura *et al.*, 2013; Mitsukura *et al.*, 2011). This stable imine was more convenient to study than non cyclic imines which are easily hydrolyzed in aqueous media. Other IREDs have been mentioned in different biosynthetic pathways as the reduction of C=N bond constitutes a physiological reaction leading to various metabolites such as folate (**33**) (Posner *et al.*, 1996), the siderophore yersiniabactin (**34**) (Meneely & Lamb, 2012) or morphine (**35**) (Winzer *et al.*, 2015). Nevertheless, no synthetic application was described due to their very narrow substrate scope restricted to these metabolites (Mangas-Sanchez *et al.*, 2017). In 2014, Huber *et al.* observed, in the presence of a large excess of an amine donor [10^3 equivalents (eq.) of methylammonium buffer at pH 9.5] and high loadings of the (*S*)-IRED from *Streptomyces* sp. GF3546 (Uniprot ID: M4ZS15), 8.8% conversions of 4-phenylbutan-2-one (**36**) to the secondary amine (2*S*)-(methylamino)-4-phenylbutane (**37**) with 76% *ee* (Scheme 8) (Huber *et al.*, 2014). The excess of amine donor and high pH were thought to favor the formation of an intermediate imine that would be recruited from solution and reduced from one prochiral face by the enzyme.

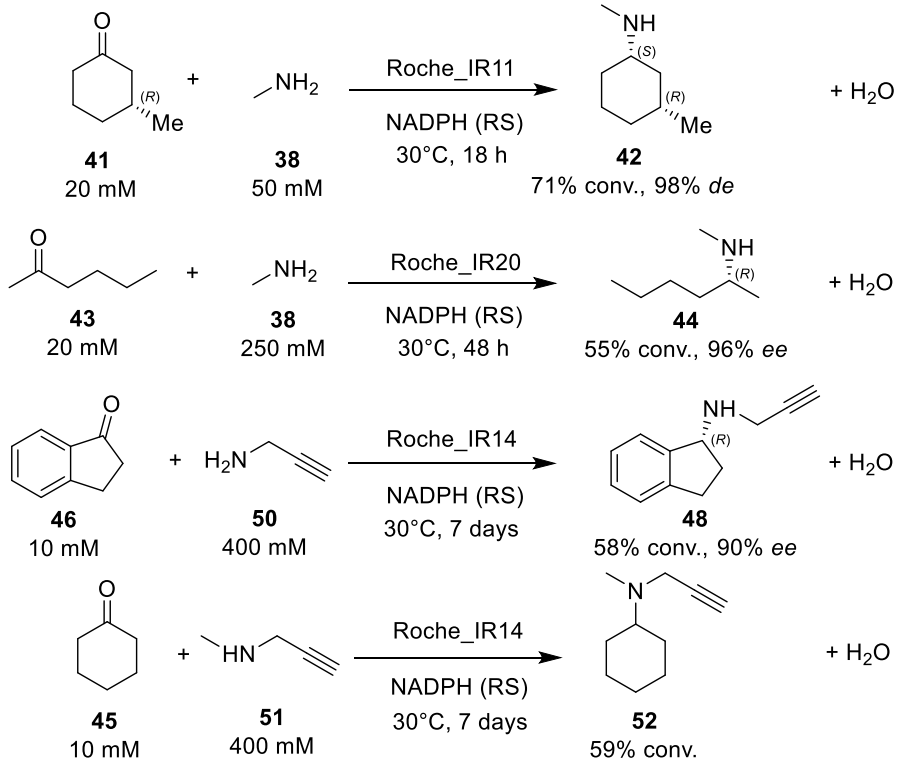


Scheme 8. Example of reaction catalyzed by (*S*)-IRED from *Streptomyces* sp. GF3546. Adapted from Ducrot *et al.*, 2020.

The discovery of new IREDs and the exploration of their substrate scope gave rise to further observations of reductive amination reactions. (*R*)-IRED-Sr (Uniprot ID: D2B7Z8), from *Streptosporangium roseum*, was screened for carbonyl reaction scope as well as activity towards different amine donors in reductive amination reactions (Scheller *et al.*, 2014; Scheller *et al.*, 2015). In the presence of (*R*)-IRED-Sr, benzaldehyde (**31**) was transformed to amine products using either ammonia (**11**) or methylamine (**38**), with conversions of up to 73% for the reaction with 50 eq. of **38**. Moreover, incubation of acetophenone (**39**) with **38** gave 39% conversion to the (*R*)-amine product (1*S*)-*N*-methyl-1-phenylethanamine (**40**) with 87% *ee* using 50 eq. of amine and at pH 9.0, each of which was thought to favor the formation of

the intermediate imine in solution. Indeed, NMR studies failed to show accumulation of an imine intermediate, suggesting that (*R*)-IRED-Sr, was remarkably efficient in withdrawing the imine from solution for reduction.

The application of IREDs to the asymmetric reductive amination of ketones was expanded by Wetzl *et al.* at Roche, who screened a library of 28 IREDs (named here Roche_IRXX) for activity with various aromatic, aliphatic and cyclic ketones, for the production of either secondary or primary amines (Wetzl *et al.*, 2015). It was found that all of the IREDs displayed activity towards at least one ketone, with higher conversions typically achieved with **38** as the amine donor, when supplied at 12.5 molar eq. and at pH 9.3. The best examples were carried out on 100 mg scale, with the IRED Roche_IR11 (Uniprot ID: F4F8G5) converting (3*R*)-3-methylcyclohexanone (**41**) and **38** to the (1*S*, 3*R*)-3-methylcyclohexan-1-amine (**42**) product in 71% yield with 98% *de*, and hexan-2-one (**43**) to the (*2R*)-*N*-methylhexan-2-amine (**44**) in 55% yield with 96% *ee* (Scheme 9).



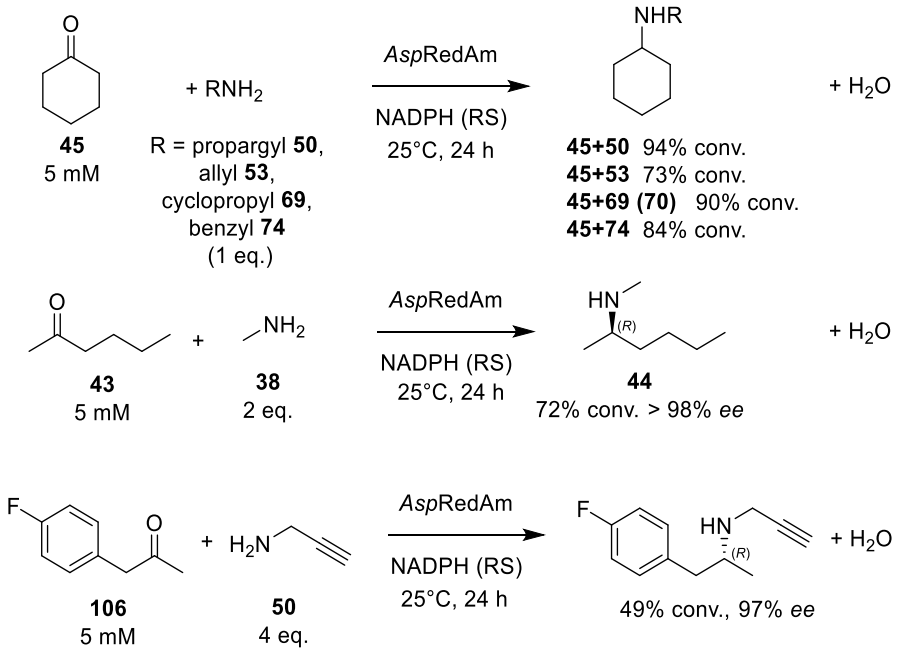
Scheme 9. Some reductive amination reactions of ketones using the IRED collection at Roche. RS = recycling system. Adapted from Ducrot *et al.*, 2020.

Again, NMR studies did not reveal significant formation of the imine intermediate in solution. Matzel *et al.* subsequently screened the Roche IRED enzymes and others for additional reductive amination reactions using cyclohexanone (**45**) and indanone (**46**), among other ketone substrates, and

short alkyl amine donors, including secondary amines such as pyrrolidine (**47**) (Matzel *et al.*, 2017). The enzymes Roche_IR14 (Uniprot ID: H6RB67) and IR-Sip (Uniprot ID: L1KNB7) were applied to the synthesis of the (*R*)- and (*S*)- enantiomers respectively of the anti-Parkinson's agent rasagiline (**48-49**) from **46** and propargylamine (**50**), the latter supplied at 40 eq. at pH 9.5 (Scheme 9). In this way, (*R*)- and (*S*)-rasagiline (**48-49**) were produced in 58% and 81% yields with 90% and 83% *ee* respectively. Additionally, Roche_IR14 enabled the reductive amination of **45** with 3-methylamino-1-propyne (**51**) to form the tertiary amine *N*-methyl-*N*-(prop-2-yn-1-yl)cyclohexanamine (**52**) in 59% yield (Scheme 9). These advances were significant in establishing the utility of IREDs in asymmetric reductive aminations, however, the reactions were still disadvantaged by their dependence upon a large molar excess of amine to favor the non-catalyzed imine formation over its hydrolysis in solution and so creating the substrate for IRED-catalyzed reduction.

II.2.3.2 Discovery of the first reductive aminase

In 2017, an oxidoreductase of the IRED family from the fungus *Aspergillus oryzae*, *Asp*RedAm (Uniprot: Q2TW47), was reported to be the first IRED enabling the reductive amination of ketones with amine partners supplied at, or near to, molar equivalence and at neutral pH (Aleku *et al.*, 2017). This suggested that this enzyme was both catalyzing imine formation, as well as imine reduction within the active site and therefore constituted a true biocatalyst for enzymatic reductive amination reaction. In contrast to the previously identified IREDs, the activity of *Asp*RedAm was largely unaffected by pH between values of 7.0 and 9.0, indicating that catalysis was not dependent upon the ambient concentration of imine formed in solution by abiotic means. *Asp*RedAm catalyzed the reductive amination of alicyclic and aliphatic carbonyls of various chain lengths using small amine donors with typically higher conversion rates when compared to previously studied IREDs. In some cases, notably with **45** and small amines such as allylamine (**53**) and propargylamine (**50**), up to 94% conversion to the amine product was achieved even at equimolar concentrations of the amine (Scheme 10). The reductive amination of 50 mM **45** with two eq. of **38** was carried out on a 100 mg scale using 0.1 mg mL^{-1} *Asp*RedAm to give 75% isolated yield of the *N*-methylcyclohexylamine (**54**) product. Kinetic studies using **45** and **38** suggested an ordered sequential Ter bi mechanism in which, following the binding of reduced cofactor, ketone and amine were bound sequentially, as had previously been demonstrated for the *N*-methyl-*L*-amino acid dehydrogenase from *Pseudomonas putida* (Mihara *et al.*, 2005). A structure of *Asp*RedAm in complex with the amine product (*R*)-rasagiline (**48**) revealed that the enzymes displayed an overall dimer fold similar to that of known IREDs (Figure 4).



Scheme 10. Some reductive amination reactions catalyzed by *Asp*RedAm from *Aspergillus oryzae* using low amine:ketone ratios. RS = recycling system. Adapted from Ducrot *et al.*, 2020.

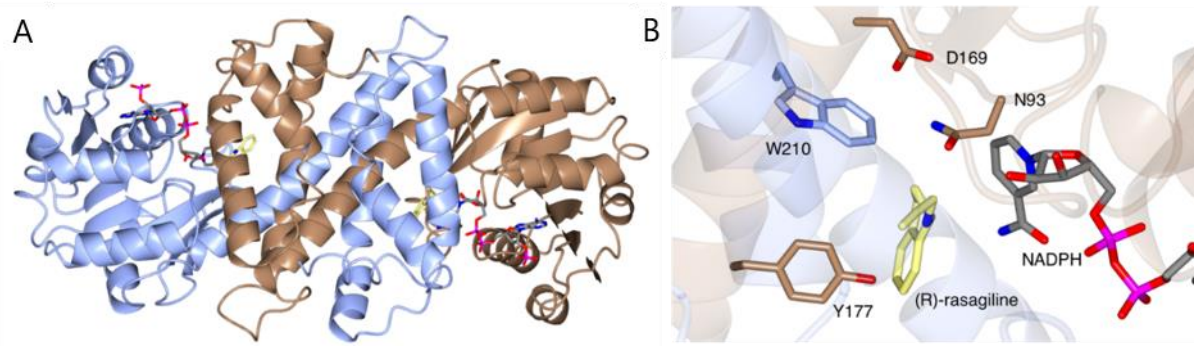


Figure 4. Global structure and active site of *Asp*RedAm. (A) Structure of *Asp*RedAm (PDB: 56GS) with monomers A and B shown in ribbon format in brown and blue respectively (Aleku *et al.*, 2017). B) Active site of *Asp*RedAm in complex with (*R*)-rasagiline (**48**). Side chains from monomers A and B are shown in brown and yellow, respectively. NADPH and **48** are shown in grey and yellow, respectively.

Each monomer in the dimer features a N-terminal (N-ter) Rossmann fold domain connected to a C-ter (C-ter) helical bundle through a long inter-domain helix. A reciprocal domain sharing results in the active site being formed at the interface between the N-ter domain of one monomer and the C-ter domain of its neighbor. The structure in complex with **48**, revealed active site residues in close contact with the ligand (Figure 4), and these observations suggested mutations that might alter the activity of the enzyme. SDM highlighted the importance of Y177 and D169 in the reductive amination abilities of *Asp*RedAm. The mutation of another active site residue, W210, to Ala, resulted in a mutant W210A that

displayed a switch in enantioselectivity for the reductive amination of 4-phenyl-butan-2-one (**36**) with allylamine (**53**) from 30% *ee* (2*R*)-4-phenylbutan-2-amine (**55**) for the WT to 90% (2*S*)-4-phenylbutan-2-amine (**56**). In a recent work, Yang *et al.* used a structure-guided semi-rational engineering on *Asp*RedAm to improve its catalytic performances towards biobased furans (Yang *et al.*, 2021). The top mutant, obtained upon W210F mutation, was able to catalyze the reductive amination of a variety of furan derivatives, such as furfural (**57**) and 5-hydroxymethylfurfural (**58**), with diverse amines yielding good to excellent conversions (51 - > 99%), The preparative-scale synthesis of *N*-propylfurfurylamine (**59**) was achieved with 98% isolated yield (93% purity) leading to a 3200 total turnover number (TTN), five-fold higher than those reported in chemical processes.

II.2.3.3 *Asp*RedAm homologs and further mutational work

Further mechanistic insight into the reductive amination catalyzed by fungal RedAms was provided by structural and mutational studies on the closely related enzyme *At*RedAm from *Aspergillus terreus* (Uniprot ID: Q0CCT3) (Sharma *et al.*, 2018). A crystal complex of *At*RedAm with **45** and **53**, achieved through co-crystallization with the redox inactive cofactor analog $NADPH_4$, suggested recognition of the carbonyl group by Y183 through a water molecule, and the amine nitrogen by D175 (Figure 5).

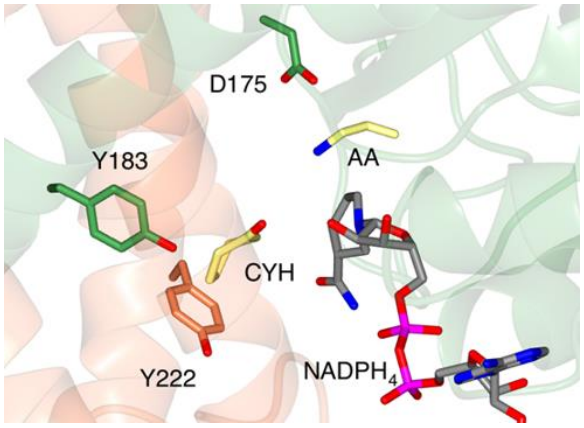
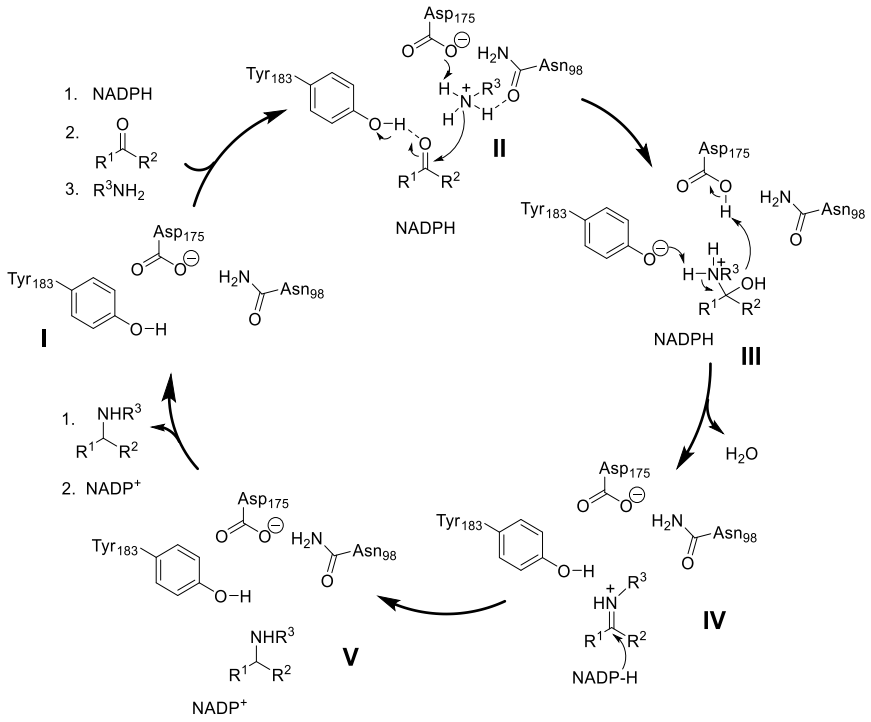


Figure 5. Active Site of *At*RedAm (PDB: 6H7P) in complex with redox-inactive NADPH4, cyclohexanone (**45**, CYH) and allylamine (**53**, AA) (Sharma *et al.*, 2018). It was proposed that the carbonyl of **45** is activated by the side chain of Y183 and that **53** is then activated for attack at the electrophilic carbon of **45**.

A mechanism was proposed in which, following sequential binding of the reduced cofactor and the ketone, the amine would be deprotonated by D175 and positioned for nucleophile attack to form hemiaminal species *en route* to the imine intermediate, which would then be reduced from one prochiral face to deliver the amine product (Scheme 11).



Scheme 11. Mechanism proposed for the fungal RedAm-catalyzed reductive amination of a ketone by small amines. Residue numbers correspond to those of *At*RedAm. Extracted from Ducrot *et al.*, 2020, adapted from Sharma *et al.,* 2018.

The Met residue (M244 in *At*RedAm) participates in the closure of the active site required for iminium stabilization and stereoselectivity through a more constrained cavity. The amine substrate spectrum of *At*RedAm was also expanded though the mutation I123A which gave a variant displaying higher specific activities towards small amines such as **38** or **53** but also towards larger ones such as hexan-1-amine (**60**), pyrrolidine (**47**) and even aniline (**61**) which was not transformed with the WT enzyme. Fungal RedAms such as *Asp*RedAm and *At*RedAm have more recently been applied to the transformation of fluoroarylketones such as 2-fluoroacetophenone (**62**) and 2,2-difluoroacetopheone (**63**) with **11** and small aliphatic amines to give optically active fluoroamine products (González-Martínez *et al.*, 2020). Hence *Ad*RedAm from *Ajellomyces dermatitidis* (Uniprot ID: C5GTJ9) was used for the reductive amination of **62** with **38** to give the (1*S*)-(2-fluoro-1-phenylethyl)-methan-1-amine (**64**) in 44% isolated yield and with 96% *ee*. Additionally, Finnigan *et al.* reported the use of *Ad*Redam co-immobilized with a glucose dehydrogenase (GDH) in the development of a model-based continuous flow reaction system for reductive amination. After the improvements of the conditions of cofactor and amine donor **53** concentrations and reaction residence time, the system could convert hydrocinnamaldehyde (**65**) into **66** in a 56% isolated yield after 2 h and a Space Time Yield (STY) of 103 g L^{-1} h^{-1} (Finnigan *et al.*, 2020).

Another fungal RedAm homolog, *Nf*RedAm from *Neosartorya fumigatus* (Uniprot ID: Q4WDZ8), also displayed exceptional activity in the reductive amination of ketones with **11** to give primary amines

(Mangas-Sanchez *et al.*, 2020). *Nf*RedAm catalyzed the amination of hexan-2-one (**43**) with **11** to give (2*R*)-*N*-methylhexan-2-amine (**44**) in a STY of 8.1 g L^{-1} h^{-1}.

With identified key roles for some residues, mutation work has been carried out on these RedAms to expand the ketone and amine scope. In the case of *At*RedAm, space was created at the rear of the active site through the mutation of Y222 to Ala leading to the acceptance of longer chain aliphatic ketones. In this way the conversion in the amination of octan-2-one (**67**) with **53** was increased from 4% to 53% (Sharma *et al.*, 2018). The equivalent mutation H215A in *Ad*RedAm even enabled higher conversion of decan-2-one (**68**). The identified position I123 in *At*RedAm proved to be a key position to extend amine range, as evidenced by the increased specific activities of the mutant I123A for both small (**11**, **38**) and bulkier amines [(hexan-1-amine (**60**), pyrrolidine (**47**), aniline (**61**)] (Sharma *et al.*, 2018).

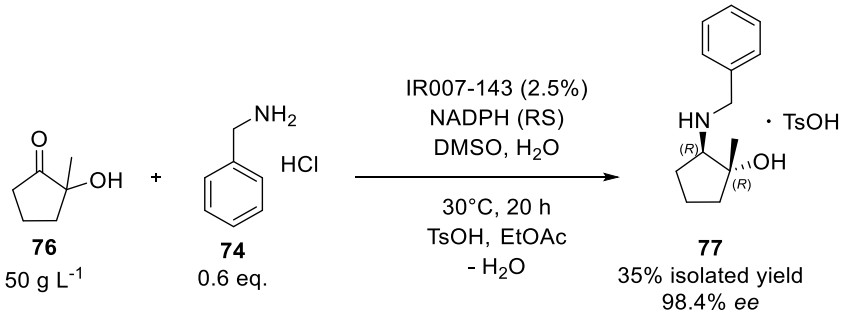
In a recent work, Li *et al.* used a structure-guided genome mining approach reported a set of new RedAms selected for their highly conserved substrate-binding and catalysis motif alignment with *Asp*RedAm (Li *et al.*, 2021). The activity of the five resulting enzymes (*Rs*RedAm, *Sim*RedAm, *Kf*RedAm, *Mt*RedAm and *Amt*RedAm) was tested for the transformation of **45** into a panel of primary and secondary amines and compared to *Asp*RedAm performances in the same conditions. Notable increase in activity compared to *Asp*RedAm towards **45** was obtained with *Rs*RedAm displaying 6.2 and 2.5-fold higher activity when supplied with **38** and cyclopropylamine (**69**), respectively. Eventually, *Mt*RedAm achieved a 100 mg scale production of *N*-cyclopropylcyclohexylamine (**70**) with a STY of 64.2 g L^{-1} d^{-1}. Conversion assays for the synthesis of 1-(4-fluorophenyl)propan-2-amine (**71**) in presence of **11** resulted in 49% conversion and 75% *ee* for the (*S*)-enantiomer (**72**) with *Kf*RedAm while *Nf*RedAm and *Nfis*RedAm, the only reported RedAm for this substrate, gave 76% and 29% conversion in 40% and 18% *ee* for the opposite (*R*)-enantiomer (**73**), respectively (Mangas-Sanchez *et al.*, 2020).

II.2.3.4 Imine reductases and reductive aminases collections

The description of reductive aminase activity in fungal enzymes such as *Asp*RedAm and *At*RedAm has prompted further screening of enzyme libraries for similar activity, in some cases by industrial groups. Roiban *et al.* at GlaxoSmithKline (GSK) screened 85 IRED homologs (named here GSK_IRXX) for reductive amination activity at pH 7.0 using amine:ketone ratios of 1:1 and including especially arylamines as substrates (Roiban *et al.*, 2017). Several candidate enzymes (particularly GSK_IR01 and GSK_IR49) catalyzed the coupling of **45** with both aniline (**61**) and benzylamine (**74**) with analytical conversions up to 99% and 74% respectively, extending the known substrate range of enzymatic reductive amination reactions.

A further screen of a library of 48 IRED-related sequences (corresponding enzymes named here Pfizer_IRXX) from bacteria was performed by France and colleagues at Pfizer, for enzymes that enable the reductive amination at low amine:ketone ratios (France *et al.*, 2018). Interestingly, several homologs were described to catalyze the coupling of **45** with pyrrolidine (**47**) to form the tertiary product 1-cyclohexylpyrrolidine (**75**) in up to 99% conversion. The enzymes proved amenable to the catalysis of scaled-up reactions, with Pfizer_IR44 (Uniprot ID: X6HB51) catalyzing this same reaction on a 625 mg scale to give **75** in 71% isolated yield. Notably, none of these IREDs gave significant conversions with **11**, although Pfizer_IR66 afforded 20% conversion to product.

In an IRED collection from Pfizer, IR007 was used and improved by multiple iterations of site saturation mutagenesis on 98 positions by Duan *et al.* to access a key intermediate to PF-06873600, a cyclin-dependent kinase 2/4/6 inhibitor, after reductive amination of the corresponding hydroxyketone (**76**) with benzylamine (**74**) (Duan *et al.*, 2022). Eventually, IR007-143 could achieve 43% conversion (35% isolated yield), 98.4% *ee* with 50 g L^{-1} substrate concentration and 2.5 wt % enzyme loading after 20 h at 30°C (Scheme 12).



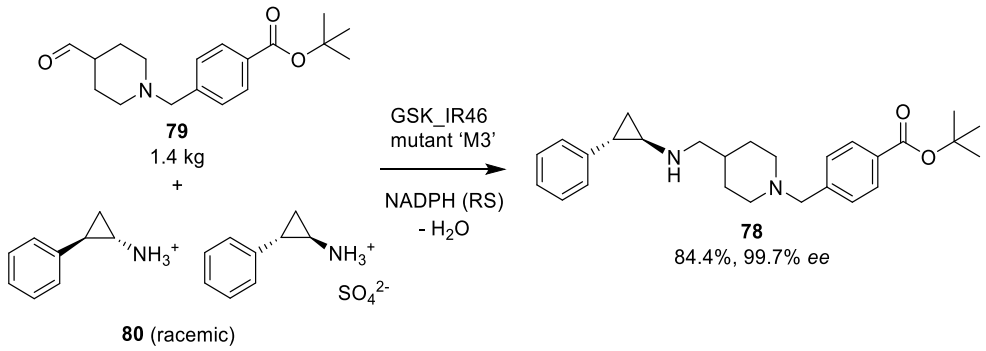
Scheme 12. Synthesis of (2*R*)-(phenylamino)-(1*R*)-methylcyclopentan-1-ol (**77**) by reductive amination using IR007-143 mutant. Adapted from Duan *et al.*, 2022.

Further work from researchers at Johnson-Matthey reported a study of 95 IRED sequences (corresponding enzymes named here JM_IRXX) including homologs from bacteria, fungi and plants (Montgomery *et al.*, 2020). The study unearthed further enzymes, such as JM_IR23, with improved activity for the reductive amination of **45** with only two eq. of aniline (**61**). Homology modeling led to the identification of residues potentially interacting with substrate/product. However, mutation work or *in silico* analysis (residues 170, 176, 180, 210 and 244, *Asp*RedAm numbering) did not reveal critical determinants for activity towards specific substrates, except for a possible role of residue 244 in catalysis and selectivity following the slight increased conversion obtained with **61** with the mutant F246L (equivalent to position 244 of *Asp*RedAm) in JM_IR48.

The industrial screening have led to more substantial studies of the scalable applications of IRED homologs for reductive amination reactions. Dominguez *et al.* at Johnson Matthey used a Design-of-

Experiments (DoE) approach to improve the reductive amination of **45** with cyclopropylamine (**69**) to give a process that gave a 90% isolated yield of product at substrate concentration of 750 mM and volumetric productivity of 12.9 g L^{-1} h^{-1} (Bornadel *et al.*, 2019). Workers at GSK reported the application of GSK_IR46 from *Saccharothrix espanaensis* (Uniprot ID: K0K1B1) to the synthesis of the lysine-specific demethylase-1 inhibitor GSK2879552 precursor (**78**) from aldehyde (**79**) and racemic *trans*-2-phenylcyclopropanamine mandelic acid salt (**80**) (Scheme 13) (Schober *et al.*, 2019).

Using the WT enzyme, a 43% isolated yield of the product with 99.9% *ee* was obtained in a 5 g reaction. Significantly, this process was then optimized through three rounds of *in vitro* directed evolution of the GSK_IR46 enzyme for process suitability and efficiency. The third round of evolution led to the variant M3 with a specific activity 13-fold improved over the WT GSK_IR46 and with a melting temperature (*T*m) of 70ºC versus 41ºC for the WT. This variant was applied to a 20 L scale biotransformation yielding 1.4 kg of the product in 84.4% isolated yield and with 99.7% *ee* (Scheme 13).



Scheme 13. Synthesis of lysine-specific demethylase-1-inhibitor GSK2879552 precursor (**78**) by reductive amination using mutant 'M3' of IRED GSK_IR46 acquired by directed evolution. The amine substrate 80 is resolved during the enzymatic process. Adapted from Ducrot *et al.*, 2020.

Additionally, Zhan *et al.* used a rational cavity design and a CAST strategy to find an IRED mutant applicable for the synthesis of the building block compounds azacycloalkylamines [precursor of Linagliptin (**81**), Alogliptin (**82**), etc.] (Zhang *et al.*, 2022). IR-G36 was selected as the scaffold for mutations after a screening of a panel of 86 IREDs (IR-G01-85 and *Asp*RedAm) against *N*-Boc-3-piperidinone (**83**) with the use of benzylamine (**74**) as the co-amine substrate. The final variants, IR-G36-M5 (N121T/S260F/F207I/M238A/M264H/A267H/L197I/N271K/ H189A/G145K) was able to perform the conversion of a range of large and complex azacycloalkylamine derivatives in industrial conditions of substrate loading (up to 47 g L^{-1}), conversions up to > 99% and > 99% *ee* towards the formation of (*R*)-enantiomers as compared to the 8% conversion 78% *ee* obtained with the WT enzyme in the initial selection screening.

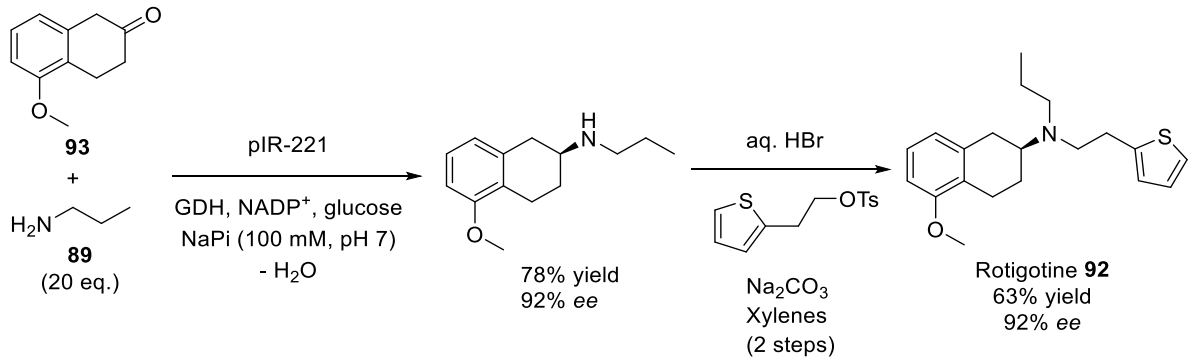
II.2.3.5 Imine reductase improvement through machine-directed evolution

In a recent work, Ma *et. al* wanted to compare a machine-directed evolution approach to more conventional strategies such as deep mutational scanning (DMS) and epPCR (Ma *et al.*, 2021). By using a genome mining strategy relying on HMM profiles, they selected the WT IRED (IRED-88) that performed the best in the reductive amination of the Cbz-protected 3-oxopyrrolidine (**84**) with **38** and was still challenging for the improvement of both catalytic efficiency and enantioselectivity (70% conversion, 30% *ee*). The results obtained from the 10 000+ mutants of the DMS and epPCR approaches served as a training set for machine-directed evolution and a structure guided machine-directed evolution, the latter focusing on a set of top active mutants according to the location of their mutated sites whether in the active site, the neighboring of the active site and distal to the active site. Eventually, all these approaches led to the design of at least one beneficial variant displaying high conversion and high (*R*)-enantioselectivity from a highly reduced set of variants to test (less than 100 variants vs thousands), with the structure-guided machine-directed evolution. However, machine learning processes will always be in a need of experimental data on these thousands of variants to predict the beneficial mutations to assay in such a low number of attempts.

II.2.3.6 Imine reductases identified from metagenomic data

In 2021, Marshall *et al.* reported a set of over 300 new IREDs from metagenomic data obtained from terrestrial and coastal sites across the United Kingdom (Marshall *et al.*, 2021). Their work also reports a high-throughput screening with a colorimetric assay based on the oxidation of a target amine catalyzed by the putative IRED. The NAD(P)H generated was detected after the reduction of the tetrazolium salt (**85**) by a NAD(P)H dehydrogenase that subsequently led to the formation of a detectable red/pink formazan (**86**) dye. This easy-to-go system enabled to screen 36 amine substrates including carbonyl-corresponding important molecular scaffolds such as acetophenone (**39**), α- and β-keto ester derivatives. The hits obtained from this assay were then tested in the way of the desired amine formation. Despite no data were published with **11** as amine donor, impressive results were achieved in this work and are so detailed here. One can mention pIR-23 that could catalyze the reductive coupling of hydrocinnamaldehyde (**65**) and dimethylamine (**87**) with 86% conversion. Eventually, crude cell-free extracts of the enantiocomplementary pIR-117, pIR-356 and pIR-355 were used in the preparative-scale reductive amination of ethyl acetoacetate (**88**) with a panel of primary amine donors [cyclopropylamine (**69**), propylamine (**89**), allylamine (**53**), propargylamine (**50**)] reaching good to excellent yields (31 - > 99%) and > 99 % *ee*. All the 384 IREDs were also screened for the conversion of 2-tetralone (**90**) and 3-chromanone (**91**) and derivatives found in a large variety of bioactive molecules and Active Pharmaceutical Ingredients (APIs). This work eventually led to the successful chemo-enzymatic synthesis

of the therapeutic ingredients for Parkinson's disease Rotigotine (**92**) in only 3 steps (Citoler *et al.*, 2021; Poewe *et al.*, 2007). This synthesis includes the pIR-221-mediated reductive amination of 5-methoxy-2-tetralone (**93**) with propylamine (**89**) (20 eq.), with a final yield of 63% and 92% *ee* (Scheme 14), as compared to the only reported five-step synthesis announced by the authors.



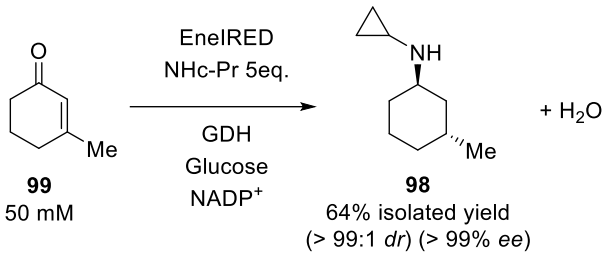
Scheme 14. Synthesis of Rotigotine (**92**) by reductive amination using pIR-221. Adapted from Citoler *et al.* ,2021.

Very recently, Aleku *et al*. used pIR-23 combined with the carboxylic acid reductase from *S. rugosus* (SrCAR) in a one-pot two-steps system for the transformation of 100 mg of cinnamic acid (**94**) to the *N*-cinnamylcyclopropanamine (**95**), *N*-cinnamylpropargylamine (**96**) and *N*-cinnamylpyrrolidine (**97**) in 58%, 62% and 48% isolated yield, respectively, using 2.5 eq. of amine with NADPH and adenosine triphosphate (ATP)-recycling systems (Aleku *et al.*, 2022).

The pIR-259, coupled with a *D*-fructose-6-phosphate-aldolase, have been recently used in a cascade reaction to access amino diols and amino polyols from an aldehyde, a hydroxyl ketone and an amine, resulting in good to excellent conversion yields (85-94%) (Ford *et al.*, 2022).

II.2.3.7 Specific case of EneIRED

After their work on a one-pot process combining an IRED with an ene-reductase (ERED) to reduce both C=C and C=N bonds of cyclic α,β-unsaturated imines (Thorpe *et al.*, 2022), Thorpe *et al.* aimed at finding biocatalysts that could directly catalyze both of these steps. Among the 389-membered set of (meta)genomic IREDs (Marshall *et al.*, 2021), 44 enzymes were found able to perform the two reductions, especially the renamed EneIRED (Thorpe *et al.*, 2022; Chen *et al.*, 2022). Interestingly, the ratio of products obtained from solely conjugate reduction or reductive amination or both can largely vary from one substrate to another but the cyclic enones seemed to be enfavored for the complete two steps. The applicability of the EneIRED for this double reduction was assessed via preparative scale syntheses, notably the 110 mg scale formation of amine (**98**) obtained in 64% isolated yield starting from 50 mM of 3-methyl-2-cyclohexen-1-one (**99**) and 5 eq. of cyclopropylamine (**69**) (Scheme 15). Mechanistic and structural studies led to the proposition of the overall catalytic cycle not detailed here.



Scheme 15. Conjugate reduction and reductive amination catalyzed by EneIRED on a preparative scale reaction. The *ee* was only reported in the initial screening. Adapted from Thorpe *et al.*, 2022.

II.2.3.8 Summarized work on the discovery and engineering of imine reductases and reductive aminases

Each of these studies illustrated the value of wide-range screening of homologous sequences for identifying biocatalysts with broader substrate scope, but also efficient enzymes with active sites different to those of reference enzymes (*Asp*RedAm and *At*RedAm in this case). Figure 6 and Table 1 summarize the documented active-site residues of the studied RedAm members and their conservation among this recently identified family. Table 1 has been updated with key example sequences reported since the publication of the review in 2020 (Ducrot *et al.*, 2020). However, due to a too high number of new sequences within the many IREDs/RedAms libraries published that could not be all included in the study, Figure 6 has not been updated. In fact, the logo representation gives a statistical overview and it would not be relevant to update it with only a few examples. The logo representation in Figure 6 still illustrates the conservation patterns of some of the discussed residues, the mutated residues located in the first or second layer of the active site. With a majority of mutations into short residues such as Ala or Leu, it well represents the will to enlarge the catalytic pocket for larger substrates accommodation. The mutation work was sometimes done in residues that are already present in the biodiversity, such as for 217 and 244 positions. The mutation work done on GSK_IR36 by Schober *et al.* and on IR-007-143 by Duan *et al.* led to mutants with modifications all along the enzyme sequence that enabled to not only increase the conversion for their targeted substrate but also to enhance their applicability for synthesis (*e.g.* higher substrate tolerance, lower enzyme loading required, usability as lyophilized enzyme, thermoactivity, etc.). Figure 7 illustrates the main carbonyl and amine substrates described in IRED/RedAms publications. It highlights the aim to use biodiversity mining and protein engineering to reach increasingly complex carbonyl and amine substrates. As also depicted in these Figures, there is still room left for further engineering work to increase the biocatalytic potential of IREDs and RedAms.

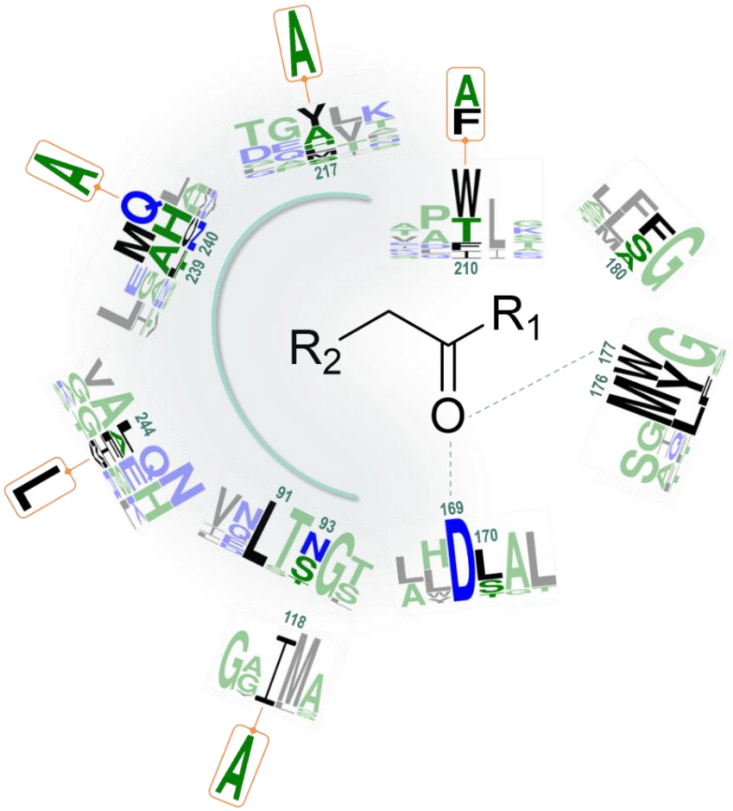


Figure 6. Logo representation of summary of key residues and mutation work done on selected IREDs with reductive aminase activity *i.e.* the ability to catalyze secondary amine formation from equimolar or low eq. of ketone and amine. This representation gives the conservation patterns of main residues of these enzymes defining the catalytic pocket and targeted for mutations inside or outside the active site. The two neighboring residues before and after these residues are included as part of the logos to illustrate their environment. The logo of the variants produced for extending substrate scope are placed above the corresponding WT residue and are highlighted in orange. Mutated residues of the variants M1-M3 of GSK_IR46 used by Schober *et al.* have not been included for clarity. The sequence numbering used is that of *Asp*RedAm. The color code is based on hydrophobicity (low hydrophobicity: blue, middle hydrophobicity: green, high hydrophobicity: black). Logos were generated using the webservice WebLogo3. Extracted from Ducrot *et al.* ,2020.

Table 1. Summary of key residues and mutation work done on selected IREDs with reductive aminase activity *i.e.* the ability to catalyze secondary amine formation from equimolar or low eq. of ketone and amine. List of main residues defining the catalytic pocket and targeted for mutations inside or outside the active site in WT selected IREDs/RedAms (black) and the corresponding mutated residues in the engineered enzymes (orange). Mutated residues of the variants M1-M3 of GSK_IR46, used by Schober *et al.*, IR-007-143 by Duan *et al.* and IR-G36 variant by Zhang *et al.* have not been included for clarity. Adapted from Ducrot *et al.*, 2020.

Enzyme name	NCBI accession number / Uniprot ID	Selected residues (*Asp* RedAm numbering)												
		91	93	118	169	170	176	177	180	210	217	239	240	244
Asp RedAm	Q2TW47	L	N	I	D	L	M	Y	F	W	Y	M	Q	V
Asp RedAm										A				
Asp RedAm													A	
Asp RedAm										F				
At RedAm	Q0CCT3	L	N	I	D	L	M	Y	F	W	Y	M	Q	L
At RedAm				A										
At RedAm											A			
Ad RedAm	C5GTJ9	L	N	I	D	L	M	Y	F	W	H	M	Q	A
Ad RedAm											A			
Nf RedAm	Q4WDZ8	L	N	I	D	L	M	Y	F	W	Y	M	Q	I
Nfis RedAm	A1D0J0	L	N	I	D	L	M	Y	F	W	Y	M	Q	I
JM_IR23	A0A1L9AVJ5	L	S	I	D	S	L	W	M	T	A	A	H	L
JM_IR48	WP_051455530.1	L	S	I	D	T	M	W	L	F	A	A	H	F
JM_IR48														L
Pfizer_IR44	X6HB51	L	S	I	D	S	M	W	L	T	L	A	H	M
GSK_IR01		L	S	I	D	S	M	W	L	T	A	A	H	F
GSK_IR46		L	S	I	D	L	L	W	L	W	S	L	N	L
GSK_IR49		L	T	I	D	L	L	F	M	I	M	T	C	A
IR-G36		L	N	I	D	L	M	Y	F	W	S	M	Q	F
IRED-88		L	S	I	N	T	M	Y	L	W	S	M	N	L
IRED-88-2		L	S	I	N	T	M	Y	L	W	S	M	N	L
pIR9		L	N	I	D	L	M	Y	F	W	Y	M	Q	I
pIR16		L	N	I	D	L	M	Y	F	W	Y	M	Q	V
pIR23		L	S	I	D	S	L	W	M	A	A	-	-	-
pIR117		L	S	I	D	S	M	W	L	A	L	P	H	F
pIR351		L	S	I	D	A	M	W	M	W	F	V	Q	I
pIR355		L	N	I	D	L	M	Y	F	W	H	M	Q	A
pIR361		L	T	I	D	M	M	W	L	M	Y	V	H	M
pIR221		H	T	I	K	M	I	A	V	W	F	W	M	L
Rs RedAm	WP_077989474.1	L	N	I	D	L	M	Y	F	W	E	M	Q	L
Sim RedAm	WP_018011194.1	L	N	I	D	L	M	Y	F	W	T	M	Q	L
Kf RedAm	D2PU58	L	N	I	D	L	M	Y	F	W	T	M	Q	F
Mt RedAm	WP_095494073.1	L	N	I	D	I	M	Y	F	W	W	M	Q	F
Amt RedAm	WP_027931121.1	L	N	I	D	I	M	Y	F	W	L	M	Q	F
EneIRED		M	T	I	Y	L	Y	F	F	I	G	V	H	Q

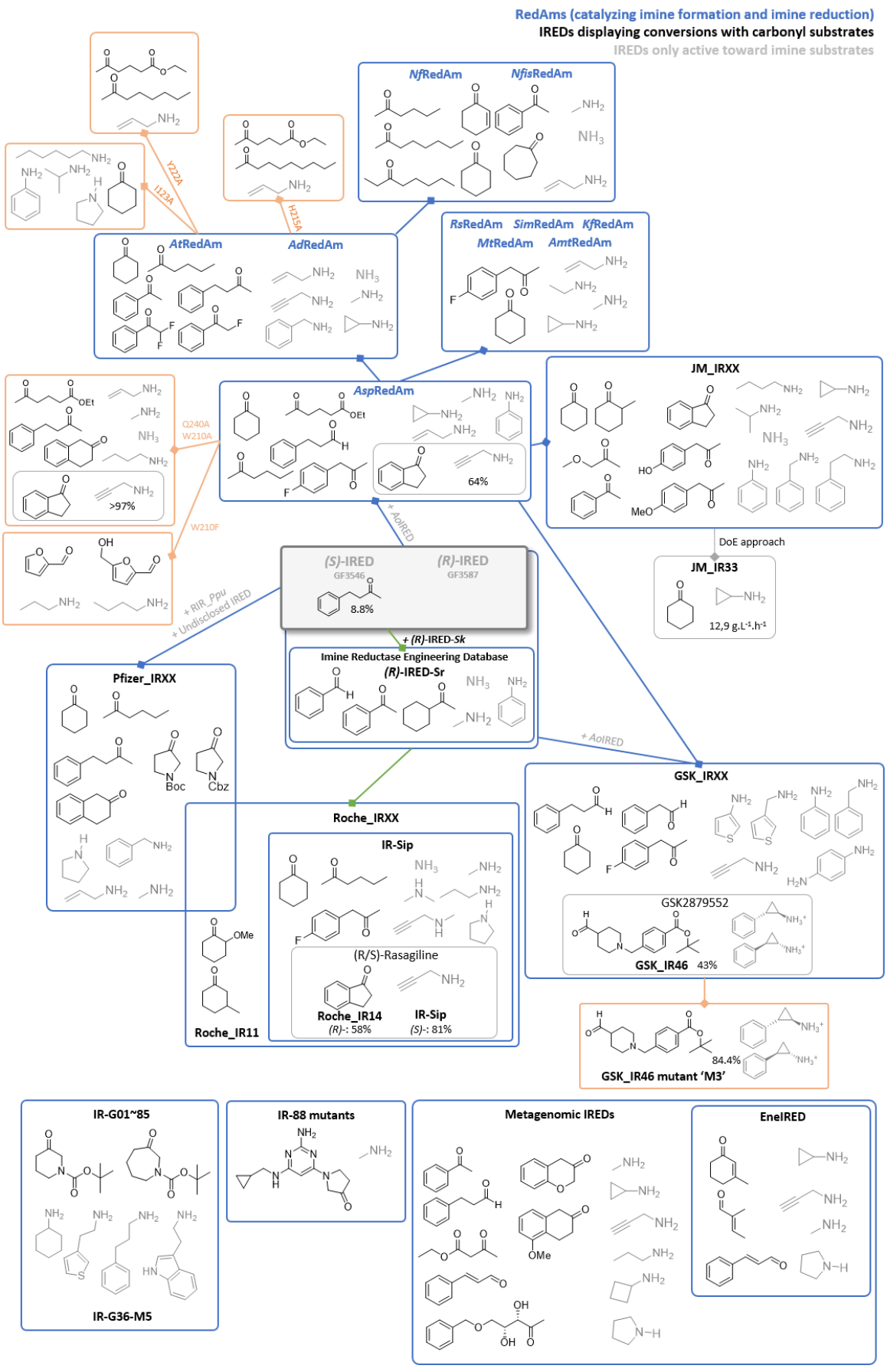
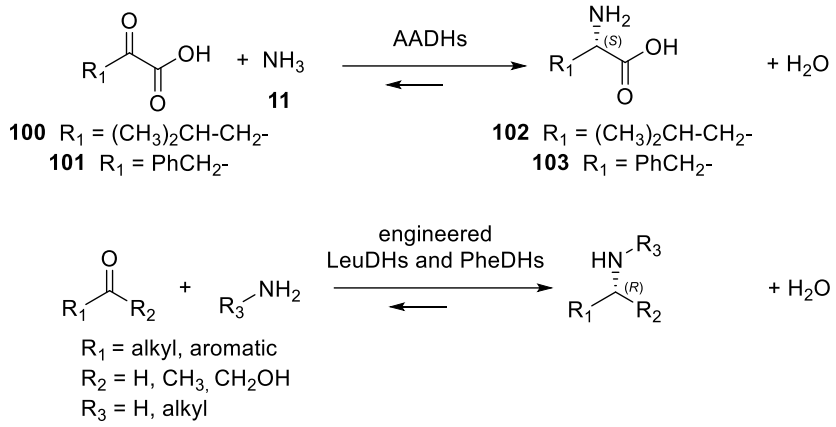

RedAms (catalyzing imine formation and imine reduction)
IREDs displaying conversions with carbonyl substrates
IREDs only active toward imine substrates
NfRedAm
NfisRedAm
RsRedAm
SimRedAm
KfRedAm
MtRedAm
AmtRedAm
AtRedAm
AdRedAm
AspRedAm
Y222A
I123A
H215A
Q240A
W210A
W210F
64%
>97%
JM_IRXX
DoE approach
JM_IR33
12,9 g.L-1.h-1
+ AoIRED
(S)-IRED
GF3546
(R)-IRED
GF3587
8.8%
+ (R)-IRED-Sk
Imine Reductase Engineering Database
(R)-IRED-Sr
+ RIR_Ppu
+ Undisclosed IRED
Pfizer_IRXX
Boc
Cbz
GSK_IRXX
GSK2879552
GSK_IR46 43%
84.4%
GSK_IR46 mutant 'M3'
Roche_IRXX
IR-Sip
(R/S)-Rasagiline
Roche_IR14
(R)-: 58%
IR-Sip
(S)-: 81%
Roche_IR11
IR-G01~85
IR-G36-M5
IR-88 mutants
Metagenomic IREDs
EneIRED

Figure 7. Substrate specificity of all the engineered IREDs described in this part. The substrates associated to each enzyme or group of enzymes were selected as the ones displaying the highest activities or conversion rates or considered as "new" substrates compared to the substrate scope of the whole family. Carbonyl and amine substrates are drawn in black and grey, respectively. The blue lines correspond to an exploration of the biodiversity. The orange lines correspond to an engineering work inside or outside the catalytic pocket. Conversion rates were given only for some examples targeting molecules of industrial interest. For the sake of clarity, the numbering of the molecules is not indicated. Adapted from Ducrot *et al.*, 2020.

II.2.4 Engineered amino acid dehydrogenases

Ubiquitous in nature with central roles in amino acid metabolism, AADHs (E.C. 1.4.1.X) catalyze the reversible NAD(P)-dependent oxidative deamination of amino acids into the corresponding ketoacids with release of **11** (Scheme 16). Due to a pressing demand for enzymes capable of catalyzing the reductive amination of carbonyl compounds into chiral primary amines (Roughley & Jordan, 2011), these AADHs appeared to be promising templates to obtain such biocatalysts by removing the acid function requirement of the substrate in the reverse reaction. In 2012, the group of Bommarius designed the first enzyme with such a potential by using leucine dehydrogenases (LeuDH, E.C. 1.4.1.9) and phenylalanine dehydrogenases (PheDH, E.C. 1.4.1.20), as platforms for engineering. These well-described WT enzymes catalyze the reversible reductive amination of 4-methyl-2-oxopentanoic acid (**100**) and 2-oxo-3-phenylpropanoic acid (**101**) into (*S*)-amino acid products **102** and **103**, respectively, using NAD(P)H as the hydride donor (Scheme 16). As the LeuDH from *Bacillus stearothermophilus* (Uniprot ID: P13154) and the PheDH from *Bacillus badius* (Uniprot ID: Q59224) were used as the first templates for AADHs engineering into AmDHs, their sequence numbering are used as references in this part.



Scheme 16. General reactions catalyzed by WT and engineered LeuDHs and PheDHs discussed in this part.

II.2.4.1 Leucine dehydrogenases

Extensive characterizations of AADHs including mechanistic investigations have been carried out and numerous X-ray crystallographic structures have been presented since 1992, including those of glutamate dehydrogenase (E.C. 1.4.1.2) from *Clostridium symbiosum* (PDB: 1AUP) and LeuDH from *Bacillus sphaericus* (PDB: 1LEH) (Baker *et al.*, 1995). The monomer structure of LeuDH is composed of two domains separated by the cofactor binding area, which gather and close in the presence of substrate to enable the hydride transfer. The main residue involved in the mechanism, K80 (LeuDH from *Bacillus stearothermophilus*), ensures the stabilization of the transformed carbonyl and acts as a base *via* its side chain (Sharma *et al.*, 2017). Figure 8 illustrates the example of the 3D structure of LeuDH from *Bacillus stearothermophilus*.

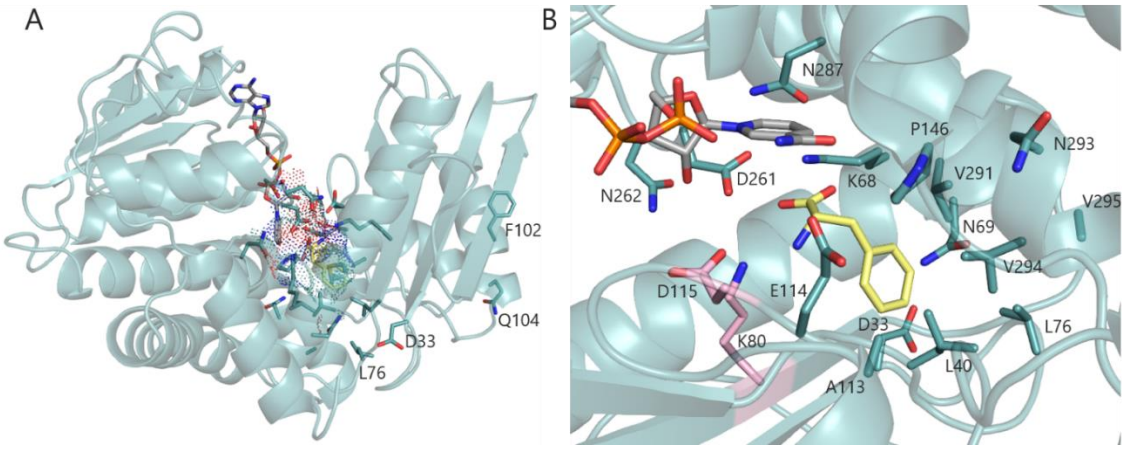


Figure 8. Structure of LeuDH from *Bacillus stearothermophilus* (model using 1C1D of PheDH from *Rhodococcus* sp. M4 as template). (A) Structure of LeuDH with the key positions targeted for mutagenesis located outside the catalytic pocket. The active site volume is highlighted in dots. (B) Active site of LeuDH in complex with *L*-phenylalanine (**103**). Side chains from the two catalytic residues K80 and D115 are shown in pink. Side chains of all the positions mutated in the AADHs included in this review are shown in blue. NADPH and **103** are shown in grey and yellow respectively.

Thanks to these published 3D structures and detailed mechanistic studies, SDM targeting first the residues involved in the coordination of the vicinal carboxylate group, was carried out in order to remove their specificity for α-keto acids and so create a proper AmDH activity towards unfunctionalized ketones (Scheme 16). The model of LeuDH from *Bacillus stearothermophilus* served as reference for *in silico* analysis and prediction of residues to target. This model was made using the already solved structure of PheDH from *Rhodococcus sp*. M4 (PDB: 1C1D; Uniprot ID: Q59771) (Vanhooke *et al.*, 1999) as template for the cofactor and substrates positioning inside the catalytic pocket. The residues equivalent to K80 and D115 described as essential for the catalytic mechanism, were excluded from the mutation libraries. A first run of site-directed evolution of this LeuDH with SSM at position K68 combined or not with other positions gave successful results and created the first 'true' AmDH activity. Named, L-

AmDH, this engineered biocatalyst harbors the two main important synergistic mutations K68S/D261L or K68M/D261V and active, among others, towards **45** and methyl-isobutyl ketone (**104**) (Abrahamson *et al.*, 2012).

This work was the starting point for all the following research for enzymes performing reductive amination (in the chemical meaning, so excluding TAs). Compared to IREDS, AmDHs use preferentially, as their native counterparts, the ammonia **11** as amine source. Giving rise to primary amines, they were so complementary to IREDs allowing the very preferential formation of secondary and tertiary amines.

II.2.4.2 Phenylalanine dehydrogenases

PheDHs catalyze the NADH-dependent interconversion of **103** and pyruvate (**105**). The 3D-structure of PheDH from *Rhodococcus* sp. M4 by Vanhooke *et al.* revealed a high similarity with LeuDH, consisting in two domains separated by a large cleft (Vanhooke *et al.*, 1999; Brunhuber *et al.*, 2000). The transposition of the beneficial mutations made in L-AmDH into the relative homolog PheDH from *Bacillus badius* led to the creation of the double variant K77M/N276V (Abrahamson *et al.*, 2013). Using the restricted and efficient codon DDK to reduce the screening effort, a high-throughput screening of the diverse two-site PheDH K77DDK/N276DDK library led to the top candidate K77S/N276L (F-AmDH). In addition to displaying higher specific activities (4 U mg^{-1} at 25°C), this variant was capable of aminating 1-(4-fluorophenyl)propan-2-one (**106**) to (2*R*)-1-(4-fluorophenyl)propan-2-amine (**73**) with 93.8% conversion and other aromatic ketones such as 1-phenoxypropan-2-one (**107**) (Scheme 17) (Knaus *et al.*, 2017). Due to the high K_M for **11** (around 500 mM) presumably due to a decrease in the activation of the carbonyl devoid of the carboxylic acid moiety, a high concentration of ammonia buffer (2 M NH_4Cl/NH_4OH pH 9.6) was required with all the engineered AADHs to achieve high activities and conversion rates. The K_M for **11** in native AADHs are already rather high, *e.g.* 75 mM for the LeuDH from *Geobacillus stearothermophilus*. However, the reasonable price of **11** compared to that of primary or secondary amines make it less an issue than with IREDs or RedAms.

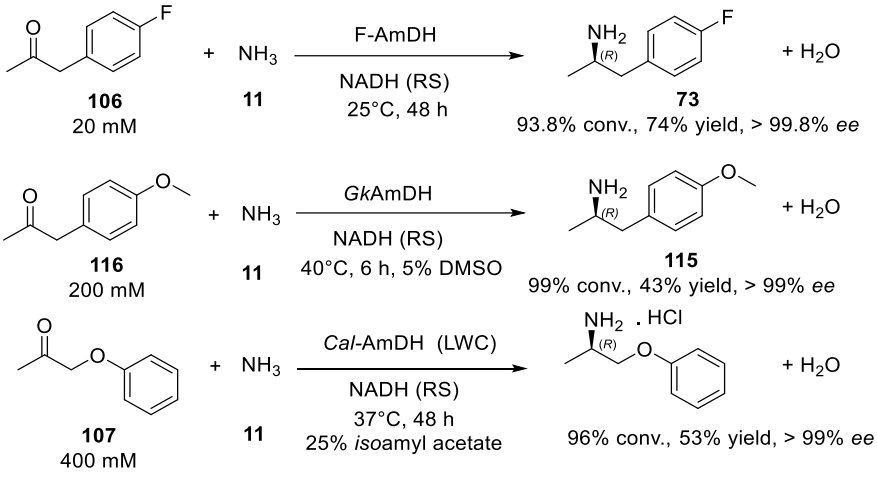
Both L-AmDH and F-AmDH served as parental enzymes for domain shuffling to generate the chimera cFL1-AmDH, a combination of the F-AmDH N-ter segment (1-149) and the L-AmDH C-ter segment (140-366). While maintaining the substrate scope of F-AmDH and the same stereoselectivity, cFL1-AmDH was also active towards benzylic ketones such as acetophenone (**39**) and 1-tetralone (**108**) and displayed an optimum temperature (>70°C) more than 20°C higher than that of F-AmDH (Bommarius *et al.*, 2014; Knaus *et al.*, 2017).

II.2.4.3 Transposition of mutations K68S/D261L into other leucine and phenylalanine dehydrogenases

Other protein engineering experiments have been carried out since this pioneering work, opening the door to 'reductive aminase' reactions for biocatalysis. The first double mutation K68S/D261L identified (Abrahamson *et al.*, 2012) has subsequently been applied to homologous sequences, supplemented or not with other mutations, to give further AmDHs based on AADH scaffolds. With regard to the mutants directly resulting from the transposition of these mutations on other sequences of AADHs, in 2015, Chen *et al.* used the mutant K77S/N270L of LeuDH from *Exiguobacterium sibiricum* (Uniprot ID: B1YLR3), *Es*LeuDH-DM, to convert secondary alcohols to amines (2*R*)-3-methylbutan-2-amine (**109**), (2*R*)-3-methylpentan-2-amine (**110**) and 3-methylcyclohexan-1-amine (**111**) by coupling it with an ADH from *Streptomyces coelicolor* in a redox-neutral cascade (Chen *et al.*, 2015). Adopting the same strategy with two other LeuDHs from *Lysinibacillus fusiformis* (*Lf*LeuAADH, Uniprot ID: D7WWB4) and *Bacillus sphaericus* (Uniprot ID: Q76GS2), they published, in 2018, *Lf*AmDH and *Bsp*AmDH (K68S/N261L variants) that could catalyze the reductive amination of simple aliphatic ketones such as methyl-isobutylketone (**104**), pentan-2-one (**112**) and hexan-2-one (**43**) (Chen *et al.*, 2018). In 2020, Liu *et al.* have applied equivalent double mutations to the selected sequence from *Geobacillus kaustophilus* (Uniprot ID: Q5KYC0) (Liu *et al.*, 2020). The PheDH mutant, *Gk*AmDH harboring the mutations K78S/N276L, displayed the desired activity towards various ketones including chiral vicinal keto alcohols 1-hydroxy-1-phenylpropan-2-amine (**113**) and 1-hydroxy-3-phenylpropan-2-one (**114**), with **11** or primary amines. In a 100 mL scale up reaction, (2*R*)-1-(4-methoxyphenyl)propan-2-amine (**115**) was isolated with 43% yield based on a > 99% conversion from the corresponding ketone **116** within 6 h, corresponding to a calculated space time yield (STY) of 130.9 g L^{-1} d^{-1} (Scheme 17).

Recently, *Gk*AmDH was implied in a chemoenzymatic cascade for the direct conversion of propargylethers to chiral amines (Chang *et al.*, 2022). The first hydration step to the ketone was carried out by an Au/carbene complex followed by the enantioselective amine formation. This chemo-/biocatalytic system, supplemented with formate (**117**)/Formate dehydrogenase (FDH)-NADH recycling system, was demonstrated efficient for the synthesis of a wide range of aryloxy and alkoxy chiral amines with 82-97% yield and 96-99% *ee*. Also in 2020, Sun and co-workers used L-AmDH and F-AmDH as templates for BLAST data mining to find new AADH templates to mutate and obtain AmDHs active towards α and β-hydroxyketones [Wang *et al.*, 2020 (B)]. The direct transposition of K68S/N261L L-AmDH equivalent mutations resulted in five new AmDHs: *Gs*AmDH from *Geobacillus stearothermophilus* [National Center for Biotechnology Information (NCBI) Accession number: AAA22570], *Bs*AmDH from *Bacillus stearothermophilus* (NCBI Accession number: WP_000171355), *Ls*AmDH from *Lysinibacillus sphaericus* CBAM5 (Uniprot: W7S058), *Sp*AmDH from *Sporosarcina psychrophila* (Uniprot: I0IJU1) and

*Ti*AmDH from *Thermoactinomyces intermedius* (Uniprot: Q60030). When used as wet cells (0.1 g mL^{-1}) in 0.5 mL total volume at 30°C for 24 h, they could achieve 19-99 % conversion of 1-hydroxybutan-2-one (**118**) into the corresponding (2*S*)-1-hydroxybutan-2-amine (**119**) with >99 % *ee*, (*S*)-selectivity being due to a reversal of Cahn-Ingold-Prelog priority. *Gs*AmDH activity was further characterized towards a range of aliphatic and aromatic α and β-hydroxyketones.



Scheme 17. Some reductive amination reactions catalyzed by engineered AADHs with mutations on positions K68/N261 (numbering of LeuDH from *Bacillus stearothermophilus*). LWC = lyophilized whole cells; RS = recycling system.

II.2.4.4 Additional protein engineering work on K68/D261

In addition to these mutations equivalent to K68S/N261L in other AADH sequences, some publications have reported an additional evolution work starting from these two key positions. In 2015, Ye *et al.* used NKK degenerate codons for SSM of the K66/N262 variant (respectively K68/D261 in *Bacillus stearothermophilus*) of the distant homolog PheDH of *Rhodococcus* sp. M4 (Uniprot ID: Q59771), to create an enzyme with a different substrate scope (Ye *et al.*, 2015). The best variant, DM_pheDH K66Q/N262L, could already convert aromatic ketones such as phenylpropan-2-one (**120**) and 4-phenylbutan-2-one (**36**), but additional mutations were also introduced to further enhance its activity. Pushpanath *et al.* screened a double-saturation (K68X/N266X) mutant library on the PheDH from the thermophile *Caldalkalibacillus thermarum* (Uniprot ID: F5L9G2) (Pushpanath *et al.*, 2017). The starting AADH, with a higher native thermostability than any other studied AADHs, was used as a scaffold to engineer a stable AmDH and displayed a moderate sequence identity with the originally engineered F-AmDH (63%) with only one change in the active site (S303A). The best variant, *Cal*-AmDH, harbored the same mutations as L-AmDH (K68S/N266L). It displayed good activity against some bulky aromatic ketones such as 1-(4-methylphenyl)propan-2-one (**121**) and 1-(2-methoxyphenyl)propan-2-one (**20**)

but also showed a T_m of 83.5°C, 27°C higher than the F-AmDH. A biphasic system with 25% *iso*-amyl acetate (**123**) enabled the conversion of 400 mM of 1-phenoxypropan-2-one (**107**) with a STY of 60 g L^{-1} d^{-1} (Scheme 17).

In 2019, Chen *et al.* investigated beneficial mutations on these two positions using saturation mutagenesis in *Lf*LeuAADH from *Lysinibacillus fusiformis* (Chen *et al.*, 2019). In addition to *Lf*AmDH (K68S/N261L), they obtained an equivalent variant (K68C/N261L) and the additional AmDH-M_0 (K68T/N261L), which showed a preference for aliphatic α-hydroxy ketones substrates such as 1-hydroxy-3-methylbutan-2-one (**124**) or 1-hydroxypentan-2-one (**125**) over methyl ketones. Such residues enable the same type of interactions with the α-hydroxyl group as the ones present in the WT with the α-carboxylate group. The change in the orientation of the hydroxyl group in Thr compared to Ser may have helped to create favorable interactions with the hydroxyl group of the ketone substrate, thereby improving recognition of these substrates. However, further studies would be required to validate this hypothesis not mentioned by the authors.

II.2.4.5 Additional protein engineering for substrate scope expansion

Protein engineering experiments have also been carried out on other positions. Indeed, even though the first rounds of mutations on the two key hotspots enabled the creation of AmDH activity, in most cases, the newly created AmDHs were only active towards simple or short ketone substrates. For this reason, many teams have focused their efforts on enhancing the substrate scope of these engineered enzymes by mutating other positions inside or outside the catalytic pocket, while always retaining the two key mutations on K68 and N261.

Thanks to a considerable effort in protein engineering on LeuDH form *Bacillus stearothermophilus*, Bommarius and co-workers found the best reductive amination activity towards several ketones with the quadruple mutant K68S/E114V/N261L/V291C (Abrahamson *et al.*, 2012). The added E114V and V291C mutations allowed them to directly enlarge the pocket size and to illustrate that position V291 was important for substrate binding in LeuDH (Kataoka *et al.*, 2003). The mutation of N262, located next to the mutated N261L, into Leu in cFL1-AmDH (cFL2-AmDH N270L/N271L) increased the apparent k_{cat} value for 1-(4-fluorophenyl)propan-2-one (**106**) (Bommarius *et al.*, 2014).

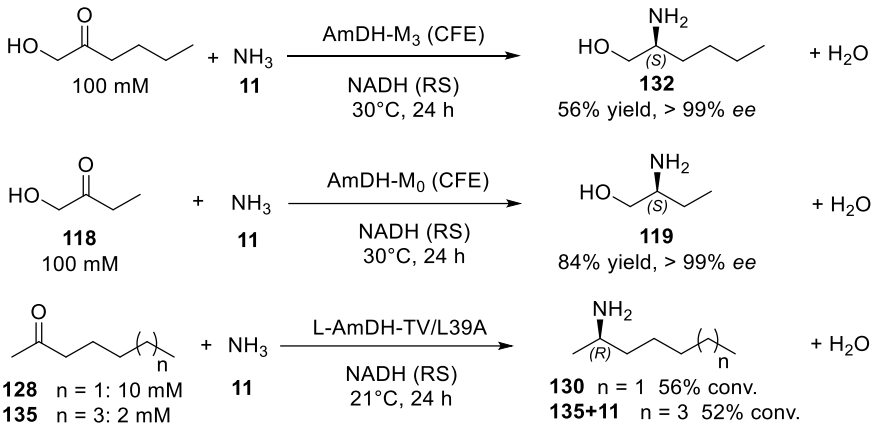
Starting from their engineered DM_pheDH, Ye *et al.* targeted the residue S149, located at the bottom of the active site pocket (Ye *et al.*, 2015). The mutation S149G, which reduced the side chain length of the amino acid by ~2Å, enabled the enlargement of the binding pocket entrance and improved the ability of the enzyme, TM_pheDH, to convert 1-phenylpropan-2-one (**120**) into (2*R*)-1-phenylpropan-2-amine (**126**) and 4-phenylbutan-2-one (**36**) into (2*R*)-4-phenylbutan-2-amine (**55**) with

99% *ee*. The latter is a building block used for the synthesis of the pharmaceutical Dilevalol (**127**) (Knaus *et al.*, 2017).

A second round of mutations also appeared in the work of Chen *et al.*, who focused on the two positions A113 and T134 (*Lf*AmDH numbering) to enlarge the cavity on the opposite side of the catalytic and substrate-binding residues (Chen *et al.*, 2018). The single mutation A113G was applied to their previously engineered *Es*LeuDH-DM (*Es*AmDH), *Lf*AmDH and *Bsp*AmDH enzymes. This allowed them to enhance the activity towards medium-sized aliphatic ketones such as hexan-2-one (**43**) but also to accommodate longer ones like heptan-2-one (**128**) or bulkier ones like 5-methylhexan-2-one (**129**). The neighboring residues of A113 (T112, E114, T134, V294) were then mutated into Ala, and Gly for the positive mutants. The double mutations A113G/T134A and A113G/T134G led to a 43 $Å^3$ increased pocket volume suitable for reductive amination of **128**, confirmed by a nearly total conversion in (2*R*)-heptan-2-amine (**130**) in less than 12 h for these two mutants *Lf*AmDH-M2 and *Lf*AmDH-M3, respectively. The same strategy was applied for AmDH-M_0 (K68T/N261L) and the quadruple variant AmDH-M_3 (K68T/N261L/A113G/T134G) displayed the highest activity towards longer α-hydroxy ketones up to 1-hydroxyheptan-2-one (**131**). These enzymes were used for the synthesis of (2*S*)-1-hydroxyhexan-2-amine (**132**) at 100 mM in 56% yield and > 99% *ee* but also the short precursor of ethambutol (**133**), (2*S*)-1-hydroxybutan-2-amine (**119**) in 84% yield and >99% *ee* (Scheme 18). It is worth noting that for all the enzymes engineered by Chen *et al.*, this second round of mutations led to a decreased activity toward the shortest substrates like pentan-2-one (**112**) or 1-hydroxypentan-2-one (**125**). Very recently, Lu *et al.* carried out a SDM on *Es*AmDH targeting a list of positions comprising, among others, L52 (equivalent to T43 in L-AmDH), never described before and located in the second layer of the active site with its side chain pointing towards the opposite side (Lu *et al.*, 2022). Notably, the top mutants *Es*AmDH-T143C (T134 in L-AmDH) and *Es*AmDH-L52S exhibited 2.06 and 1.55 mU mg^{-1} against **112** compared to 0.71 mU mg^{-1} reached with *Es*AmDH.

Franklin *et al.* conducted separate additional sets of mutations starting from the original L-AmDH (K68S/E114V/N261L/V291C) published in 2012 (Abrahamson *et al.*, 2012; Franklin *et al.*, 2020). The first set was selected to increase activity and stability and focused on two residues further away from the active site, D32 and F101, previously shown to be promising candidates. (David *et al.*, 2004) L-AmDH-TV (D32A/F101S/C290V) displayed an average of 2.5-fold higher activity towards aliphatic ketones **112** and **43** and butan-2-one (**134**) that were already accepted by L-AmDH. To expand the substrate scope and accommodate larger substrates, a third round of mutations was performed on L-AmDH-TV considering the two previously described mutations A112G and T133A/G (L-AmDH-TV numbering) but also L39A/G, all pointing towards the substrate binding site (Chen *et al.*, 2018; Kataoka *et al.*, 2003). The

single mutant L39A enabled to double the activity towards **43** and moderate activity towards octan-2-one (**67**). The mutation L39G permitted the accommodation of ketones as large as nonan-2-one (**135**) and decan-2-one (**68**), with a lower activity for the shorter substrates **134** and **112**. The benefit of these mutations was exemplified by the conversion of 56% and 52% of **128** and **135**, respectively, with L-AmDH-TV/L39A in 24 h with a formate (**117**)/FDH NADH recycling system (Scheme 18). The addition of A112G and T133G further enhanced the activity towards long aliphatic ketones even with methyl substituents such as 6-methylheptan-2-one (**136**) and 6-methyloctan-2-one (**137**).



Scheme 18. Some reductive amination reactions catalyzed by engineered AADHs with mutations inside and outside the catalytic pocket. CFE = crude cell-free extract; RS = recycling system. Adapted from Ducrot *et al.*, 2020.

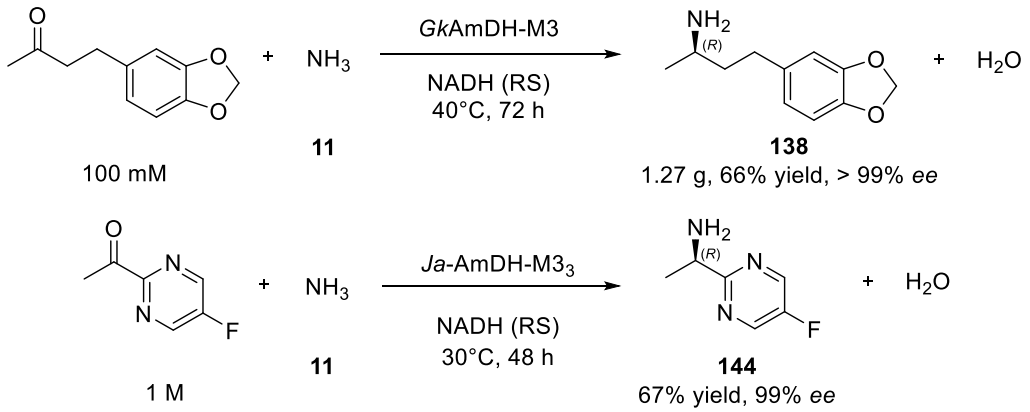
The substrate-binding pocket of *Gk*AmDH was also enlarged through semi-rational mutagenesis, combining epPCR with several steps of SSM of determined residues within the first or close-second layer of the active site (Wang *et al.*, 2021). The best mutants, *Gk*AmDH-M3 (V144A/V309A/A310G) and *Gk*AmDH-M8 (V144A/V309A/A310G/S156T/Q308A /C79N/F86M) enabled the gram-scale production of the two pharmaceutical building blocks, (2*R*)-4-phenylbutan-2-amine (**55**) and 4-(1,3-benzodioxol-5-yl)-(2*R*)-butan-2-amine (**138**), precursor of medroxalol (**139**), with up to 85% and 66% isolated yield, respectively (Scheme 19). Molecular dynamic (MD) analyses provided insights into the effects of the key mutations on the substrate specificity.

Starting from the two discovered *Gs*AmDH and *Sp*AmDH (see Chapter I, II.2.4.3), Sun and co-workers reported in two distinct publications, an extensive knowledge and structural-guided mutagenesis approach to improve their activity towards α- and β-hydroxyketones (Tong *et al.*, 2021). To do so, they used the work already performed by other teams in other AmDHs templates and a combination of CAST, SDSM and ISM strategies. Notably, *Sp*AmDH top mutant (K68S/N261L/I111F/E114V/V294C) was used as cell lysates to transform 200 mM of 1-hydroxybutan-2-one (**118**) into the corresponding (*S*)-amine (**119**) in 91% conversion and >99% *ee* within 18 h at 30°C.

Similarly, *Gs*AmDH best variant (K68S/D261L/E114V/T134C/P146V/H187D/V294C) achieved 99% conversion and 90% isolated yield of 100 mM 4-hydroxybutan-2-one (**140**) giving (3*R*)-1-hydroxybutan-3-amine (**141**) with >99% *ee* within 36 h (Ming *et al.*, 2022).

Recently, Mu *et al.* described an iterative Ala scanning-base multiple site protein engineering on L-*Bc*AmDH, previously constructed with the introduction of the double mutation K70S/N263L in a LeuDH from *Bacillus cereus* (Mu *et al.*, 2021). In addition to the already described A115G mutation (L-*Bc*AmDH-M1) (Chen *et al.*, 2018), eight mutated spots were identified as the non-Ala residues located within 6 Å of G115 and the hindered phenyl group of one of the targeted substrate **36** docked in the pocket. Each step of mutation allowed them to find the top pocket conformation to fit increasingly hindered substrates: L-*Bc*AmDH-M3-1 (A115G/T136A/L42A) towards 1-(4-methylphenyl)propan-2-one (**121**), L-*Bc*AmDH-M4 (A115G/T136A/L42A/V296A) towards **36** and 4-phenylpentan-2-one (**142**) and eventually L-*Bc*AmDH-M5 (A115G/T136A/L42A/V296A/V293A) towards 4-phenylhexan-2-one (**143**). Interestingly, this large number of mutations into Ala in the same area did not lead to any major instability as these three mutants displayed a T_{50}^{30} (temperature at which enzyme activity is reduced to 50% after a 30-min heat treatment) of 68, 56 and 57°C, respectively, as compared to 61°C for the WT enzyme.

LeuDH from *Bacillus cereus* was also used by Kong *et al.* as template for a BLASTP homologous sequence searching against NCBI database (Kong *et al.*, 2022). After the introduction of the typical double mutation (K68S/N261L) into a LeuDH from *Jeotgalicoccus aerolatus*, *Ja*-AmDH was further mutated until exhibiting activities and good conversion potential towards α-(hetero)arylketones, building blocks of many pharmaceuticals. The mutated spots correspond to the already described E113 (Abrahamson *et al.*, 2012), V291 (Abrahamson *et al.*, 2012, Kataoka *et al.*, 2003), V294 and L39 (Franklin *et al.*, 2020). For instance, *Ja*-AmDH-M3$_3$ enabled the synthesis of the building block (1*S*)-1-(5-fluoropyrimidin-2-yl)ethan-1-amine (**144**) [AZD 1480 (**145**)] in 67% yield and 99% *ee* (1 M substrate loading, Turnover number (TON): 15075) (Scheme 19). Using the same strategy as done by Bommarius and co-workers in 2014, Li *et al.* replaced the cofactor-binding domain of F-AmDH from *Bacillus badius* by that of L-*Bc*AmDH obtained from the LeuDH from *Bacillus cereus* to increase its cofactor affinity (Li *et al.*, 2022). The chimera cFLF-AmDH (F-AmDH 1-156, L-*Bc*AmDH 147-346 and F-AmDH 360-380) displayed a two-fold higher affinity for NADH, a 4.4-fold higher corresponding catalytic efficiency in presence of **112** and unexpected increased thermal stability and substrate scope as compared to the parent F-AmDH.



Scheme 19. Some reductive amination reactions catalyzed by engineered AADHs obtained after several rounds of mutagenesis. RS = recycling system.

II.2.4.6 Additional protein engineering to reach amine dehydrogenase activity

In 2020, Jiang *et al*. succeeded in changing the substrate specificity of PheDH from *Bacillus halodurans* (Uniprot ID: Q9KG94) without the common double mutation K77S/N275L (PheDH *Bacillus halodurans* numbering) but targeting, as in cFL2-AmDH, the adjacent residue N276 and E113, located further away from the cavity (Jiang *et al.*, 2020). With a combination of single and double mutations at these two positions, the PheDH variants, named *Bh*AmDHs, not only displayed an improved activity for the natural substrates, *L*-phenylalanine (**103**) (oxidative deamination) and 2-oxo-3-phenylpropanoic acid (**101**) (reductive amination), but also towards some aliphatic ketones without a carboxylic group at the α-position, which were not substrates of the WT enzyme. For example, the double variant PheDH-E113D/N276L displayed a good activity (1.31 U mg^{-1}) towards 5-methyl-2,3-hexanedione (**146**) and PheDH-N276L towards 3-methylcyclohexanone (**147**) (1.67 U mg^{-1}). These results highlight the effect of mutations surrounding identified key residues or those positioned even further away.

II.2.4.7 Glutamate dehydrogenase

In a recent work, Zhou *et al.* achieved to give a proper AmDH activity to a glutamate dehydrogenase (GluDH, EC 1.4.1.2) from *E. coli*-K12 (*Ec*GDH). The comparison of its crystal structure [Uniprot ID: P00370; PDB: 4BHT (Oliveira *et al.*, 2016)] to that of other GDHs enzyme-ligand complex structure enabled to identify two substrate-binding residues, K116 and N348 (Figure 9) (Zhou *et al.*, 2022). *Ec*GDH consists in a homohexameric structure with each subunit being formed by a N-ter domain that mediates the hexamer assembly and a Rossmann fold C-ter domain for NAD(P) binding (Oliveira *et al*., 2016).

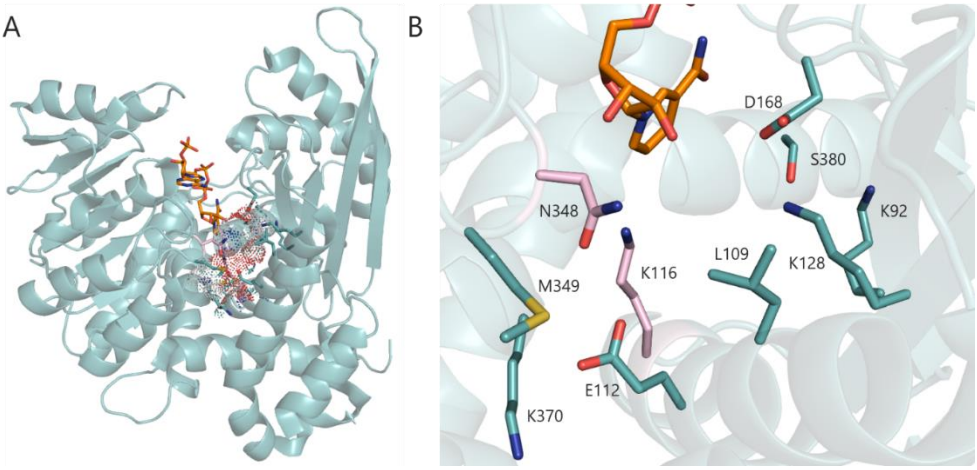
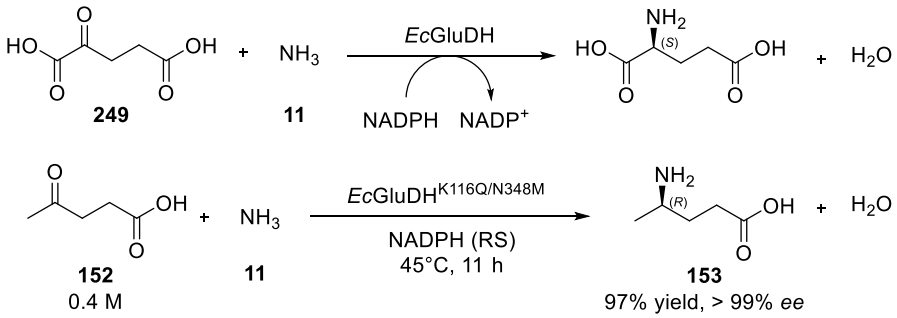


Figure 9. Structure of *Ec*GluDH from *E. coli*-K12 (PDB: 4BHT) (A) Structure of *Ec*GluDH with highlighting of the active site volume (B) Detail of the residues forming the catalytic pocket and the key positions described in this part, K116 and N348, are shown in pink. $NADP^+$ cofactor is colored in orange.

A combinatorial saturation mutagenesis library and an activity screening towards the targeted substrate 4-oxopentanoic acid (**152**) led to the discovery of the top mutant *Ec*GluDH$^{K116Q/N348M}$ that was eventually able to convert 0.4 M **152** to the (*R*)-aminopentanoic acid (**153**) with 97% yield (> 99% *ee*) in 11 h (Scheme 20).

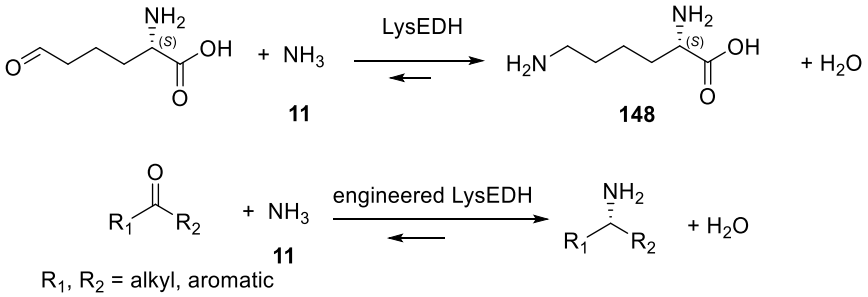


Scheme 20. Reactions catalyzed by WT and engineered *Ec*GluDH. Adapted from Zhou *et al.*, 2022.

II.2.4.8 *L*-lysine dehydrogenase

Unlike previously described enzymes engineered from very similar scaffolds of α-AADHs, Tseliou *et al.* focused on the ε-deaminating *L*-lysine dehydrogenase (LysEDH) from *Geobacillus stearothermophilus* (Uniprot ID: Q9AJC6) [Tseliou *et al.*, 2019(A)]. In this case, the reaction occurs at the terminal amino group of *L*-lysine (**148**) instead of the α-amine moiety, which was shown to be non-essential for the enzyme activity. The objectives were to remove the essential nature of the α-carboxyl group for recognition by the WT enzyme, which prevents any activity towards substrates devoid of this

function, such as heptan-2-one (**128**), acetophenone (**39**) or 4-chromanone (**149**) (Scheme 21). Based on computational studies of a LysEDH homology model, created using different templates with both NADH cofactor and ligand within the active site, several residues were identified as potential targets for mutagenesis (V130, A131, V172, F173, H181, Y238, T240 and R242) (Figure 10).



Scheme 21. General reaction catalyzed by LysEDH and the engineered LysEDH.

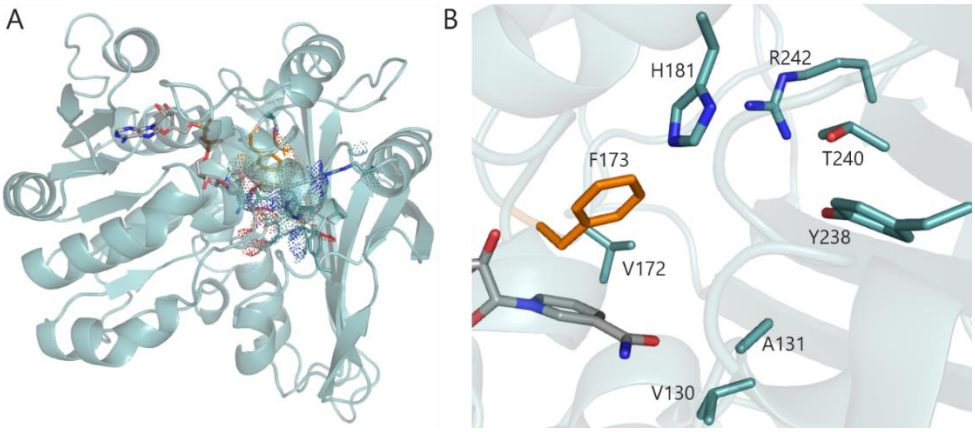
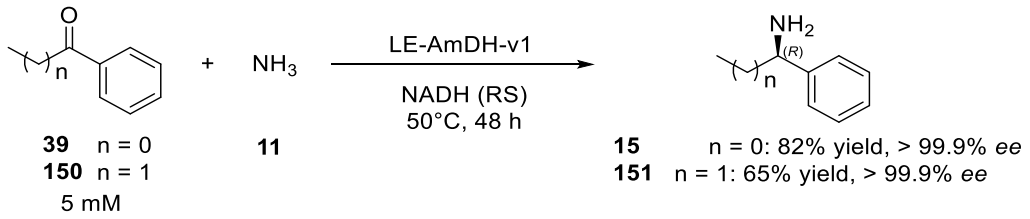


Figure 10. Model of the ε-deaminating *L*-lysine dehydrogenase from *Geobacillus stearothermophilus* (model using 1E5Q of saccharopine dehydrogenase from *Magnaporthe oryzae*). (A) Structure of LysEDH with highlighting of the active site volume (B) Focus on the key positions of the active site. F173 is highlighted in orange.

Among them, only the mutation of F173 enabled the conversion of medium-sized aliphatic ketones and, above all, the targeted bulky aromatic ketones such as **39** or 1-tetralone (**108**). The bulky residue F173 is located at the opposite side of the hydrophilic cavity which accommodates the α-amino moiety of **148** and serves to provide both the hydrophobic property of the cavity and the orientation of the substrate side chain. Its mutation into Ala enabled the enlargement of the hydrophobic binding pocket by reducing the size of the residue by 4.3 Å. Remarkably, LE-AmDH-v1 (F173A) displayed a high stability (T_m: 69°C) and a reduced product inhibition compared to other engineered AADHs, thus leading to high biocatalytic performance. The reductive amination of 5 mM **39** and 1-phenylpropan-1-one (**150**)

was performed on a 600 mg scale to give respectively 82% and 65% isolated yield of the (*R*)-amine products **15** and **151** with >99.9% *ee* (Scheme 22).



Scheme 22. Example of reductive amination reaction catalyzed by the engineered ε-deaminating *L*-lysine dehydrogenase.

II.2.4.9 Summarized work on the discovery and engineering of amine dehydrogenases from amino acid dehydrogenases

Figure 11 and Table 2 illustrate the mutational work carried out on a diverse panel of AADHs throughout the years. Overall, we can clearly see that the number of found beneficial sites has expanded since 2020 (Ducrot *et al.*, 2021). The first results obtained by Bommarius and co-workers have largely been used and retrieved later on, with improvements by further mutational efforts. The approaches targeting the residues located in the first layer to obtain AmDHs from AADHs with various substrate scope led to mostly the same beneficial residues. Figure 12 illustrates the main carbonyl substrates described in engineered AADHs publications. It highlights the aim to use biodiversity mining and protein engineering to reach increasingly complex carbonyl substrates. However, after having explored all the possibilities in the first layer, the more recent engineering works are now targeting an increasing number of residues located in the second layer of the active site to have access to a higher diversity of AmDH scaffolds.

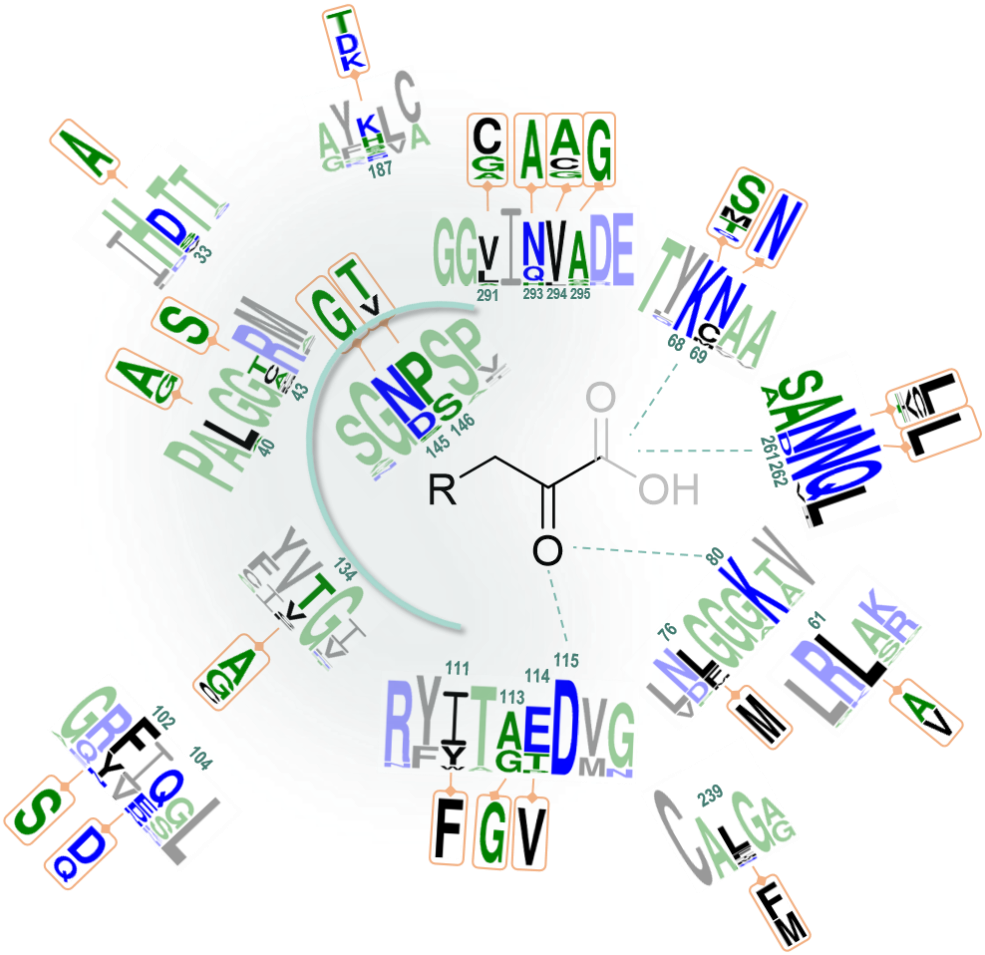


Figure 11. Logo representation of the mutational work done on LeuDHs and PheDHs. This representation shows the conservation patterns of all the residues targeted for mutation in the different studies focusing on engineered AADHs. The two residues described as essential for the catalytic mechanism (K80 and D115) are also indicated. The two neighboring residues located before and after the key residues in the protein sequence are also included to illustrate their environment. The sequence numbering used is that of LeuDH from *Bacillus stearothermophilus*. The logo of the variants produced for each of them are highlighted in orange. The color code is based on hydrophobicity (low hydrophobicity: blue, middle hydrophobicity: green, high hydrophobicity: black). Logo was generated using the webservice WebLogo3. Adapted from Ducrot *et al.*, 2020.

Table 2. Summary of mutational work done on LeuDHs and PheDHs. List of residues targeted for mutations in WT AADHs (black) and the corresponding mutated residues in the different engineered enzymes (orange). For the sake clarity, the work on *Gs*AmDH and *Sp*AmDH from Sun and co-workers has been restricted to the top mutations. Adapted from Ducrot *et al.*, 2020.

Enzyme name	Uniprot ID / NCBI accession number	Selected residues (LeuDH from *bacillus stearothermophilus* numbering)																														
		33	40	43	61	68	69	76	80	102	104	111	113	114	115	134	145	146	187	239	261	262	291	293	294	295
LeuDH (*b.s* .)	P13154	D	L	T	L	K	N	L	K	F	Q	I	A	E	D	T	N	P	H	L	D	N	V	N	V	A
K68M/E114V/N261V/V291C						M								V							V		C			
K68S/E114V/N261I/V291C						S								V							I		C			
L-AmDH						S								V							L		C			
L-AmDH_D32A			A			S								V							L		C			
L-AmDH_D32A/F101S			A			S				S				V							L		C			
L-AmDHTV			A			S				S				V							L					
L-AmDHTV_L39A			A	A		S				S				V							L					
L-AmDHTV_L39G			A	G		S				S				V							L					
L-AmDHTV_A112G			A			S				S			G	V							L					
L-AmDHTV_A112G/T133G			A			S				S			G	V		G					L					
L-AmDHTV_L39A/A112G			A	A		S				S			G	V							L					
L-AmDHTV_L39A/A112G/T133G			A	A		S				S			G	V		G					L					
PheDH (*b.d* .)	Q59224	D	L	C	L	K	C	F	K	F	D	Y	G	T	D	V	D	S	K	F	N	N	L	N	V	A
Q59224_K77M/N276V						M															V					
Q59224_K77S/N276L						S															L					
cFL1AmDH*		D	L	C	L	S	C	F	K	F	D	Y	G	T	D	V	D	S	H	L	L	N	C	N	V	A
cFL2AmDH_N270L/N271L																						L				
LeuDH (*r.sp.*)	Q59771	S	A	T	L	K	M	M	K	N	D	W	G	P	D	F	S	S	S	M	N	N	A	H	L	V
DM_PheDH						Q															C					
TM_PheDH						Q											G				C					
PheDH (*c.t.*)	F5L9G2	N	L	C	L	K	C	F	K	F	Q	Y	G	T	D	V	D	S	K	I	N	N	L	Q	V	S
K68S/N266L						S															L					
***Es* LeuDH**	B1YLR3	D	L	L	L	K	N	L	K	Y	Q	I	A	E	D	T	N	P	N	L	N	N	V	N	V	A
Es AmDH						S															L					
Es AmDH_A112G						S							G								L					
Es AmDH_A112G/T134A						S							G			A					L					
Es AmDH_A112G/T134G						S							G			G					L					
Es AmDH_L52S				S		S															L					
Es AmDH_T143C						S										C					L					
***Lf* LeuDH**	D7WWB4	D	L	A	L	K	N	L	K	F	Q	I	A	E	D	T	N	P	K	L	N	N	V	N	V	A
Lf AmDH						S															L					
Lf AmDH-M1						S							G								L					
Lf AmDH-M2						S							G			A					L					
Lf AmDH-M3						S					Q		G			G					L					
***Bsp* LeuDH**	Q76GS2	D	L	A	L	K	N	L	K	F	Q	I	A	E	D	T	N	P	K	L	N	N	V	N	V	A
Bsp AmDH_DM						S															L					
Bsp AmDH_A122G						S							G								L					
Bsp AmDH_A122G/T143A						S							G			A					L					
Bsp AmDH_A122G/T143G						S							G			G					L					
AmDH-M0						T															L					
AmDH-M1						T							G								L					
AmDH-M2						T							G			A					L					
AmDH-M3						T							G			G					L					
***Bc* -LeuDH**		D	L	T	L	K	N	L	K	Y	Q	I	A	E	D	T	N	P	H	L	N	N	V	N	V	A
L-*Bc* AmDH						S															L					
L-*Bc* AmDH-M1						S							G								L					
L-*Bc* AmDH-M2						S							G			A					L					
L-*Bc* AmDH-M3-1			A			S							G			A					L					
L-*Bc* AmDH-M3-2						S							G			A					L			A		
L-*Bc* AmDH-M4			A			S							G			A					L			A		
L-*Bc* AmDH-M5			A			S							G			A					L		A	A		
Ja-LeuDH		D	L	I	L	K	N	L	K	Y	N	I	A	E	D	C	N	P	S	L	N	N	V	N	V	A
Ja -AmDH-M1						S															L					
Ja -AmDH-M2						S								V							L					
Ja -AmDH-M3$_1$						S								V							L		C			
Ja -AmDH-M3$_2$						S								V							L		A			
Ja -AmDH-M3$_3$						S								V							L		G			
Ja -AmDH-M4						S								V							L		G		G	
Ja -AmDH-M5			G			S								V							L		G			
PheDH (*g.k* .)	Q5KYC0	S	L	C	L	K	C	F	K	F	E	Y	G	T	D	V	D	S	K	F	N	N	L	Q	V	A
K78S/N276L						S															L					
K78S/N276C						S															C					
K78S/N276T						S															T					
K78T/N276L						T															L					
K78T/N276C						T															C					
Gk AmDH_V144A						S										A					L					
Gk AmDH_V144A/A310G						S										A					L					G
Gk AmDH_V144A/V309A						S										A					L			A		
Gk AmDH_V144A/V309A/A310G						S										A					L			A		G
Gk AmDH_V144A/V309A/A310G/S156T/Q308A						S										A		T			L			A	A	G
Gk AmDH_V144A/V309A/A310G/S156T/Q308A/C79N/F86M						S	N	M								A		T			L			A	A	G
PheDH (*b.h.*)	Q9KG94	D	L	C	L	K	C	F	K	F	E	Y	G	T	D	G	D	S	K	Q	N	N	L	Q	V	A
Q9KG94_N276L						S															L	L				
Q9KG94_E113D						S					D										L					
Q9KG94_E113D/N276L						S					D										L	L				
***Gs* LeuDH**	AAA22570	D	L	T	L	K	N	L	K	F	Q	I	A	E	D	T	N	P	H	L	D	Q	V	N	V	A
Gs AmDH						S															L					
Gs AmDH-mh174						S								V		C		V	D		L				C	
***Sp* LeuDH**	I0IJU1	D	L	T	L	K	N	L	K	Y	E	I	A	E	D	T	N	P	A	L	N	N	V	N	V	A
Sp AmDH-wh84						S						F		V							L				C	

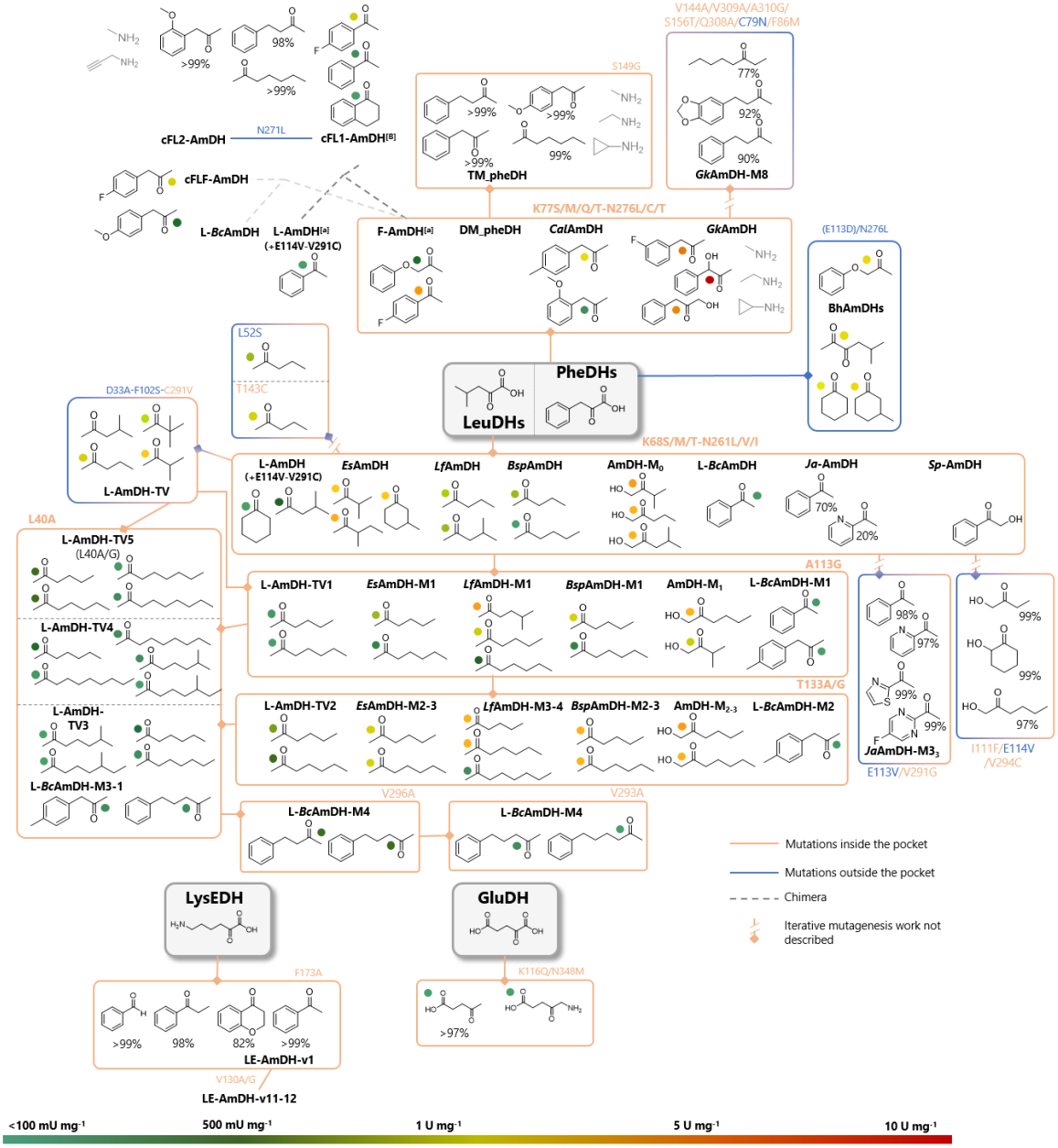
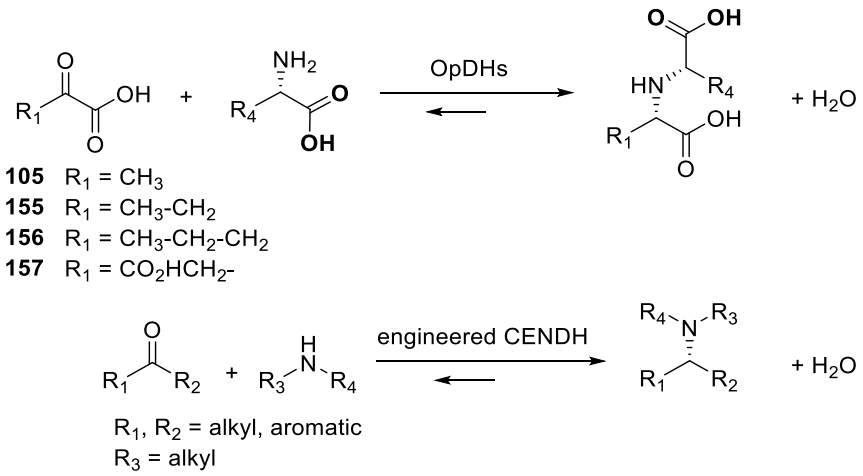


Figure 12. Substrate specificity of all the engineered AADHs and ε-deaminating *L*-lysine dehydrogenase described in this part. The substrates associated to each variant were selected as the ones displaying the highest activities or considered as "new" substrates compared to the substrate scope of the whole family. Carbonyl and amine substrates are drawn in black and grey, respectively.. If not mentioned, the amine substrate is ammonia (**11**). The orange and blue lines correspond to a set of mutations respectively inside or outside the catalytic pocket. The grey dashed lines correspond to the creation of a chimera from two parental enzymes. The specific activities indicated in this Figure are given as information but not to be compared as all the assays were not done with the exact same conditions. The following ones correspond to a range of typical conditions: 10-80 mM carbonyl substrate, 0.1-0.2 mM NAD(P)H cofactor, 1-6 M ammonium-based buffer pH 8.5-9.6, 25-30°C. In the case of enzymes without activity data, the substrates showing the highest conversion rates for one experimental condition were selected For the sake of clarity, the numbering of the molecules is not indicated. [a] Activity assays done with lower **11** concentrations 0.225-0.5 M that could not allow to reach the saturation of **11**. [b] Activity assays done at 60°C, conversion tests at 30°C. The mutations written in bold correspond to the L-AmDH or F-AmDH numbering while the other mutations refer to the numbering of the targeted enzymes.

II.2.5 Engineered opine dehydrogenases

Opine dehydrogenases (OpDHs) are a class of oxidoreductases that catalyze the reductive amination of α-keto acids, not with **11** like AADHs, but with α-amino acids, using NADH as cofactor (Scheme 23). The resulting products, opines, have very different physiological roles, depending on whether they are formed in invertebrates or in plants (Springer Handbook of Enzymes, 2005). In invertebrates, their formation ensures the consumption of pyruvate (**105**) and NADH keeping glycolysis running upon anaerobic exercise; in plants they ensure the growth of the parasitic *Agrobacteria* responsible for crown gall disease of plants after genomic insertion of the OpDH coding region into the host plant (Watanabe *et al.*, 2015).

The OpDH from *Arthrobacter* sp. strain 1C (Uniprot ID: Q44297), named CENDH, is one of the most characterized OpDHs (Asano *et al.* 1989), together with that of the great scallop *Pecten maximus* (Smits *et al.*, 2008). In addition to NAD^+-dependent oxidation of opines of plant origin, such as methiopine (**154**), reductive aminations are also performed by these enzymes, with the (*R*)-configuration at the newly generated stereogenic center of the opine products. They preferentially bind pyruvate (**105**) as electrophile and, to a lesser extent, 2-oxobutanoate (**155**), 2-oxopentanoate (**156**) and oxaloacetate (**157**). In addition to its activity towards (*S*)-methionine (**158**) and (*S*)-phenylalanine (**103**), ODH was found to be even more active towards *L*-norvaline (**159**), hence its name *N*-(1-*D*-carboxyethyl)-*L*-norvaline dehydrogenase (Scheme 23).



Scheme 23. General reaction catalyzed by OpDHs and engineered CENDH.

The resolution of its structure [PDB: 1BG6 apo structure (Britton *et al.*, 1998)] combined with a mechanistic study of the already resolved OpDH from *Pecten maximus* [PDB: 3C7C holo structure with *L*-arginine (**160**) (Smits *et al.*, 2008); 3C7D holo structure with **105** (Smits *et al.*, 2008)] provided

preliminary insights into its mechanism (Figure 13). Like other AADHs, these enzymes possess a two-domain structure with a cleft at the interface holding the cofactor. Despite the identification of key residues binding **105** (Q118, H212 and E142 in the case of OpDH from *Pecten maximus*) and some NMR evidence of sequential order of substrate binding (Smits *et al.*, 2010), no further description of the mechanism has been reported.

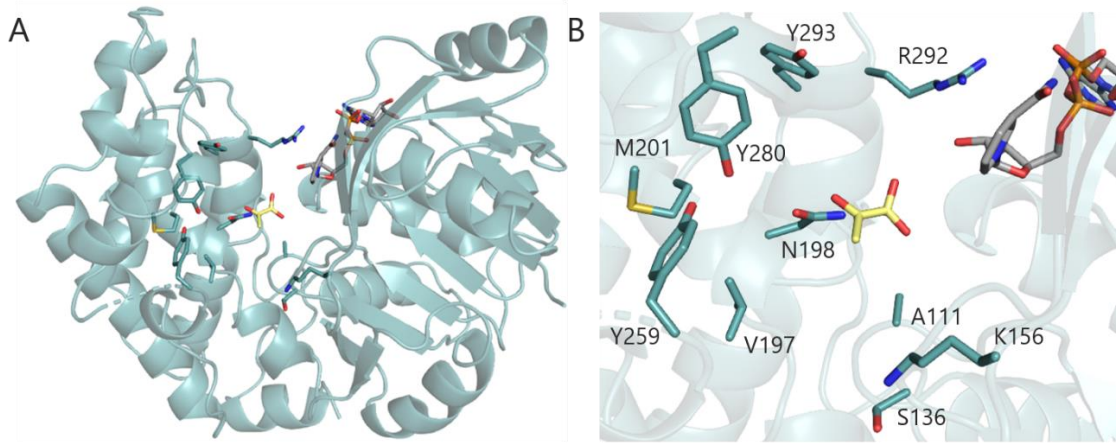
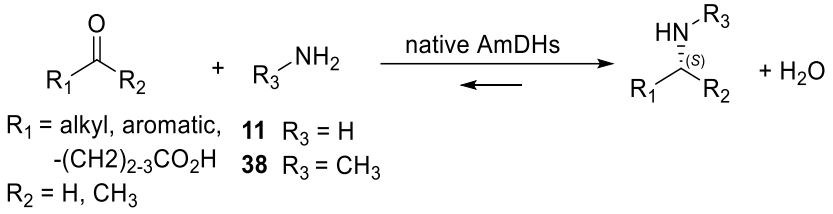


Figure 13. Global structure and active site of CENDH. (A) Structure of CENDH (PDB: 1BG6) in ribbon format (B) Detail of the key positions targeted for mutagenesis by Codexis, shown with carbon atoms in blue. The position of NADPH and **105**, respectively in grey and yellow, were obtained by 3D-alignment with the structure of OpDH from *Pecten maximus* (PDB: 3C7D).

Considering its relaxed substrate scope, Codexis used CENDH as a starting template to evolve it into a catalyst able to synthesize tertiary amines (Chen, 2018). The variants, obtained after an extensive effort on directed evolution experiments using the protein engineering technology CodeEvolver®, were not only more active towards the two native substrates of CENDH, **105** and **159**, but were also able to catalyze the coupling of a range of ketone substrates, such as **45**, cyclopentanone (**161**) or 5-methoxytetralone (**162**) (Scheme 23). A variety of amine partners was accepted, notably butan-1-amine (**163**), and transformed to products with high *ee*. The mutations that could positively impact the activity towards a wide range of substrates mostly targeted residues located in the active site cleft. For example, the multiple variant A111M/K156T/N198H/Y259M/Y280L/R292V/Y293H was ranked first for the conversion of **45** with **163** into *N*-butyl-*N*-cyclohexylamine (**164**). To our knowledge, no other work regarding the engineering of these enzymes for reductive amination has been published. More structural studies are required to expand their biocatalytic potential, having in mind their native activity towards various type of amines and the actual mechanistic knowledge of AADHs. Other types of OpDHs, such as saccharopine DHs, may also be valuable templates for engineering. In view of the significant differences between their native substrates and the targeted substrates, the evolution of OpDHs constitutes a considerable piece of enzyme engineering.

II.2.6 Natural amine dehydrogenases

To meet the demand in synthetically useful AmDHs, an exploration of protein biodiversity was considered a good strategy to find WT AmDHs with various substrate scopes and new scaffolds to provide other templates than AADHs for protein engineering (Scheme 24). Some examples of this activity occurring in nature were published at the beginning of the 2000s by Itoh *et al.*, who reported an amine dehydrogenase from *Streptomyces virginiae* that was active towards a broad range of ketone and aldehyde substrates but displayed low enantioselectivity for most of these substrates (Itoh *et al.*, 2000). The corresponding gene was not identified which prevented further work. Later on, Wang *et al.* also identified an AmDH activity using whole cells of *Pseudomonas kilonensis* (Wang & Fang, 2013; Wang, 2016). In the context of a collaboration between the biocatalysis team of the UMR and S. Wang in 2018, gene identification within this strain using comparative genomics was not successful and the AmDH activity of this strain could not be confirmed.

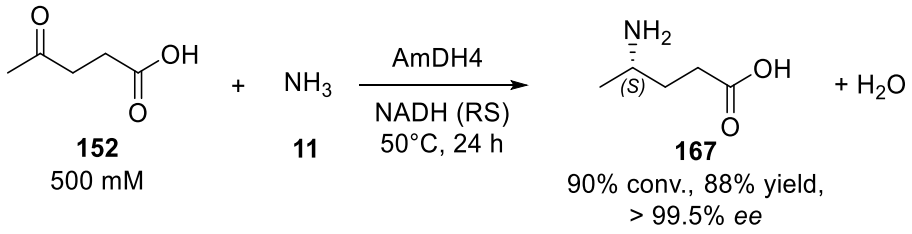


Scheme 24. General reactions catalyzed by native AmDHs.

II.2.6.1 First genome mining to find native reductive amination activity

In 2015, the biocatalysis team of the research unit decided to search for WT AmDHs by genome mining. The strategy was first to compile an inventory of all the natural NAD(P)H-dependent reductive amination enzymes using **11** in the biological genomic databases and then exclude the reactions on α- or β-ketoacids (*i.e.* α- or β-AADHs). This led to the sequence of (2*R*,4*S*)-2,4-diaminopentanoate dehydrogenase (2,4-DAPDH) from *Clostridium sticklandii* (Uniprot ID: E3PY99) (Mayol *et al.*, 2016). This enzyme catalyzes the oxidative deamination of (2*R*,4*S*)-2,4-diaminopentanoic acid (**165**) and was shown, some years ago in our unit, to be involved in the ornithine degradation pathway (Fonknechten *et al.*, 2009). The corresponding sequence was used as a reference known to perform the amination of (2*R*)-2-amino-4-oxopentanoic acid (**166**) without any close functionalization unlike AADHs, in order to identify homologs with proper AmDH activity. This genome mining approach was carried out with the Cloning and Screening Platform of the research unit, led by Veronique de Berardinis, to build a collection of candidate enzymes. The particularity of the pipeline used by the Platform, shown in Appendix 3, is the integration of several criteria such as the availability of the strain in the in-house strains library

comprising around 1,200 strains. Among the collected and selected enzymes, 6 of them (AmDH2, AmDH4, AmDH5 being the originally considered 2,4-DAPDH, and AmDH7-9) were found active towards 4-oxopentanoic acid (**152**) (4OP-AmDHs), in addition to their native substrate **166**. This substrate was interesting to consider as the α-amine function of the α-aminoacid part was not mandatory to obtain activity, thus approaching an AmDH type substrate without functional prerequisite in the structure. The NADH-dependent AmDH4, from the thermophilic bacteria *Petrotoga mobilis* (Uniprot ID: A9BHL2), was further investigated, including a study of its biocatalytic performance. High substrate loading tolerance enabled the isolation of (4*S*)-4-aminopentanoic acid (**167**) in 88% yield with a significant STY of 53 g $L^{-1}d^{-1}$ (Scheme 25). The very good *ee* for the (*S*)-enantiomer (> 99%) was in accordance with the high preference of 2,4-DAPDH AmDH5 for **165** over (2*R*, 4*R*)-2,4-diaminopentanoic acid (**168**) (Fonknechten *et al.*, 2009).



Scheme 25. Example of reductive amination reaction catalyzed by AmDH4. RS = recycling system.

In collaboration with the group of Prof. Gideon Grogan from the University of York, the resolution of its structure (PDB: 6G1H) gave some insights into its mechanism (Mayol *et al.*, 2019). E102 was found to be the main catalytic residue, while R161, N163 and H264 were observed to bind the carboxylate moiety of **165** and **167** through electrostatic interactions (Figure 14). Site directed mutagenesis of E102 into Ala resulted in a mutant with which it proved impossible to reach **11** saturation in kinetics studies, underlying the importance of this residue for both the fixation and activation of **11**. To alter the substrate specificity and enable the conversion of ketone substrates devoid of the carboxylic moiety, these polar residues were mutated into non-polar ones. Among the 8 variants produced, N135V/N163V/R161M/H264L mutant, named here AmDH4_M1, displayed the highest activity with 105 mU mg^{-1} towards **112**. The main mutation responsible for this switch of substrate scope seemed to be R161M but no further mutation work has been done yet to further explore this site. This result was the first step in the identification of other type of AmDH easily accessible from biodiversity by minor protein engineering. With the objective to increase AmDH4 activity towards **152**, easily accessible material from biobased lignocellulosic waste, Cai *et al* carried out directed evolution combining several rounds of epPCR and DNA shuffling (Cai *et al.*, 2020). Eventually, the top mutant obtained (AmDH4-I80T/P224S/E296G) exhibited higher solubility, 18-fold higher catalytic efficiency towards **152** compared

to the WT enzyme. AmDH4 mutant was able to aminate 0.5 M **152** in the corresponding amine **167** at 40°C yielding > 97% conversion and > 99% *ee*.

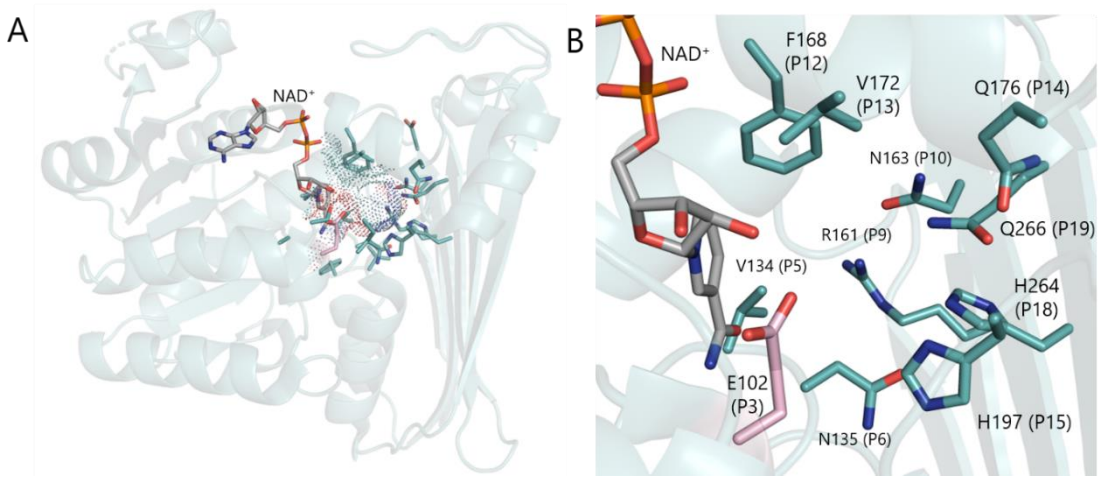


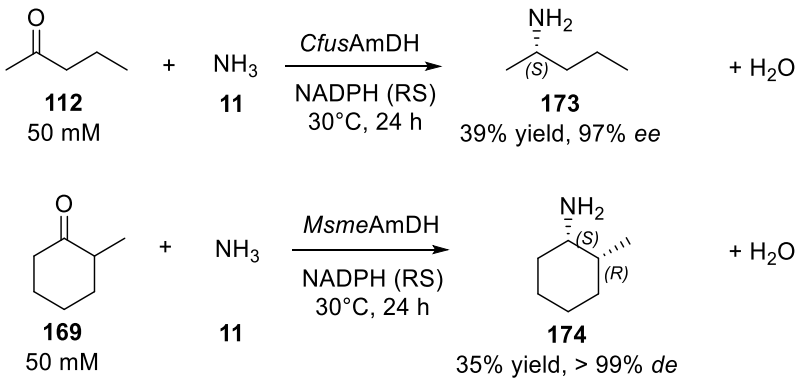
Figure 14. Global structure and active site of AmDH4. (A) Structure of AmDH4 (PDB: 6G1M), the volume of the active site is represented with the dot surface of the residues constituting the first layer of the pocket (B) Active site of AmDH4 with a focus on the P1-P21 positions (blue), shown in cylinder format, defining the cavity with NAD^+ (grey). The catalytic residue E102 (P3) is highlighted in pink. For clarity, P1-2-4-6-7-8-11-16-17-20-21 are omitted. P1-P21 numbering explanations are given in Chapter I, II.2.6.2.

Interestingly, 4OP-AmDHs do not share significant sequence homology with the previously described engineered AADHs nor RedAms, demonstrating that enzymes catalyzing reductive amination are distributed among varied families. Moreover, 4OP-AmDHs are (*S*)-selective enzymes, thus providing the opposite amine enantiomer compared to previously described enzymes. For example, AmDH4 is a very good complementary enzyme to the (*R*)-selective *Ec*GluDH$^{K116Q/N348M}$ mentioned in Chapter I, II.2.4.7, despite being non-homolog enzymes (Zhou *et al.*, 2022). Therefore, 4OP-AmDHs constituted a real platform to study in addition to the RedAms and engineered AADHs.

II.2.6.2 First set of native amine dehydrogenases

To further explore AmDH biodiversity and reveal more distant homologs with activity towards unfunctionalized ketones, 4OP-AmDHs were used in a second sequence-driven approach by our group. This iterative approach successfully led to the identification and characterization of six AmDHs [*Msme*AmDH (Uniprot ID: A0A8B4R7U8, formerly A0A0D6I8P6), *Cfus*AmDH (Uniprot ID: S9Q235), *Micro*AmDH (Uniprot ID: C3UMY1), *Apau*AmDH (Uniprot ID: E3CZE3), *Myco*AmDH (Uniprot ID: A0A101AWU7) and *Mvac*AmDH (Uniprot ID: K0UKT5)] with similar substrate spectra, with a preference for cycloalkanones such as **45** and 2-methylcyclohexanone (**169**), and aliphatic aldehydes such as isobutyraldehyde (**170**) and 2-methylbutanal (**171**). *Msme*AmDH displayed also significant activity towards 3-hydroxybutan-2-one (**172**). This interesting activity towards (hydroxy)ketones was studied in

more details during my PhD (see Chapter II, IV.1) (Ducrot *et al.*, 2021). No or only traces of activity towards the initial substrates **166** and **152** was detected with these enzymes. With low sequence identity (< 30%) with 2,4-DAPDH and 4OP-AmDHs, and no apparent common biological role, these enzymes are members of the first identified group of nat-AmDHs (Mayol *et al.*, 2019). As indicated by biochemical characterization and substrate spectrum, these enzymes are slightly more efficient with NADPH than NADH even if this specificity seems to be substrate-dependent. As for engineered AADHs, the high K_M of **11** required the use of a high concentration of ammonia buffer, usually in the range of 1 – 2 M, at pH 8.0 – 9.5. Biocatalytic synthesis performed on 100 mg scale afforded (2*S*)-pentan-2-amine (**173**) or (1*S*, 2*R*)-2-methylcyclohexylamine (**174**) from the equivalent carbonyl compounds in 39% and 35% yield respectively, corresponding to a STY of 2.4-2.6 g $L^{-1}d^{-1}$ (Scheme 26).



Scheme 26. Examples of reductive amination reactions catalyzed by nat-AmDHs. RS = recycling system.

The structural resolution of *Cfus*AmDH (PDB: 6IAU, Figure 15) and *Msme*AmDH (PDB: 6IAQ) showed that their whole structure is highly similar to that of AmDH4 (Figure 14).

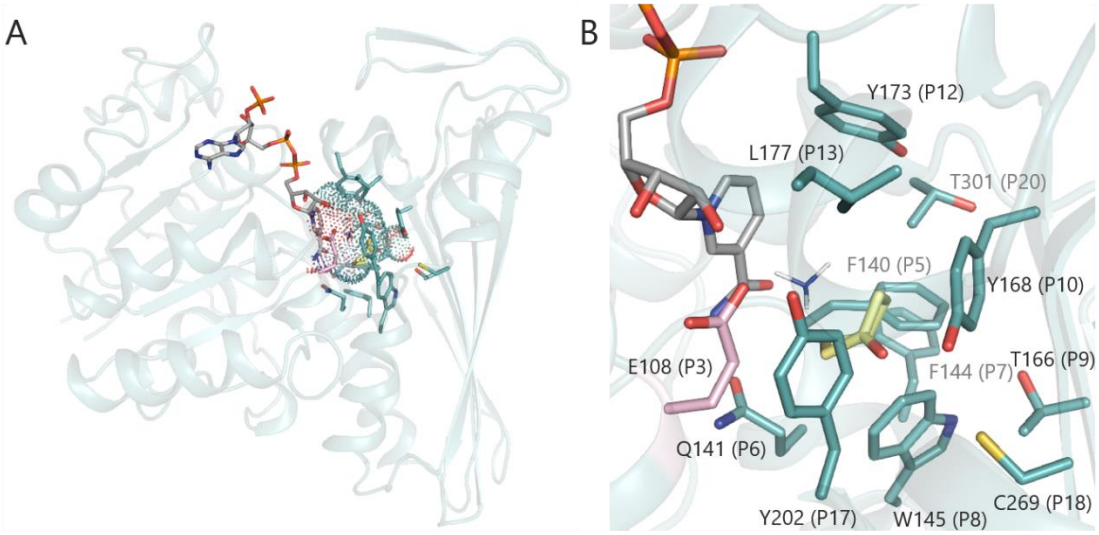
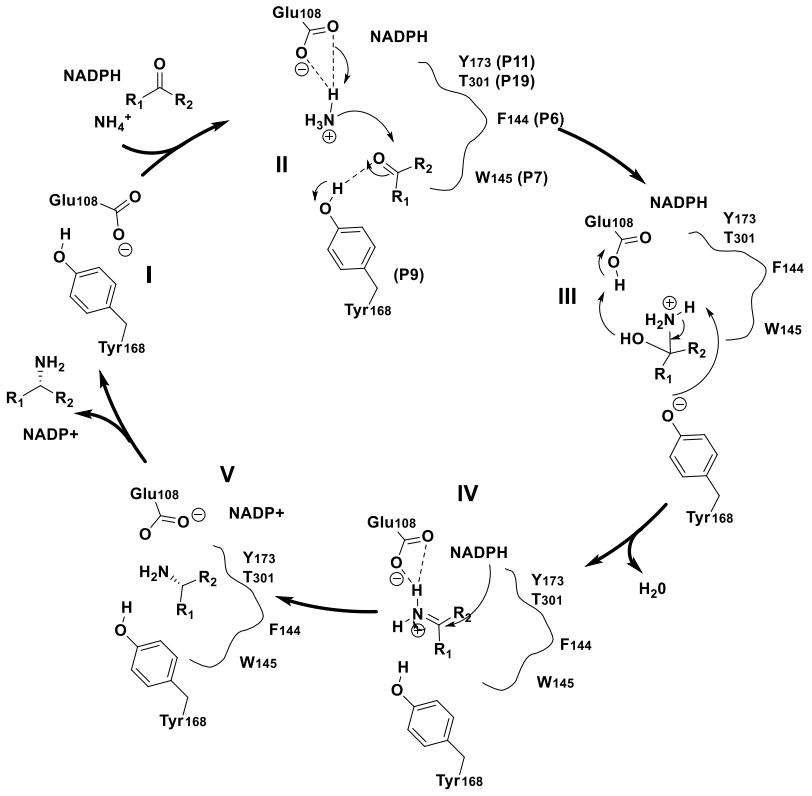


Figure 15. Global structure and active site of *Cfus*AmDH. (A) Structure of *Cfus*AmDH (PDB: 6IAU), the volume of the active site is represented with the dot surface of the residues constituting the first layer of the pocket (B) Active site of *Cfus*AmDH with a focus on the P1-P21 positions (blue), shown in cylinder format, defining the cavity with NADPH (grey), **45** (yellow) and **11**. The catalytic residue E108 (P3) is highlighted in pink. For clarity, P1-2-4-11-14-15-16-19-21 are omitted. P1-P21 numbering explanations are given later on in this paragraph.

Similarly to AmDH4, the hypothetical mechanism involved the conserved glutamate activation **11**, which in turn would attack the electrophilic carbon of the carbonyl hypothetically secured by the conserved hydrogen bond donors Tyr/Trp within the active site (II). The resulting carbinolamine (III) would be dehydrated to lead to the amine product (V) after reduction by the hydride of the nicotinamide cofactor (IV) on the *re*-face in the case of prochiral substrate (Scheme 27).



Scheme 27. Hypothetical mechanism of nat-AmDHs. The proposed mechanism is showed for *Cfus*AmDH and **45** as carbonyl substrate.

Using the solved structures of AmDH4, a classification of the nat-AmDH family based on the spatial conservation of the residues defining the active site pocket (positions initially named P1 to P20) was built according to the ASMC (Active Site Modeling and Clustering) pipeline, described in Appendix 4, set up in our unit and already successfully applied to other enzymes (Bastard *et al.*, 2014; de Melo-Minardi *et al.*, 2010). A mechanistic and structural role has been assigned to several of these residues (Mayol *et al*., 2019) and was further studied in Chapter II, II. The so called ASMC1 classification partitioned the members of this family into the following groups: group 2 (G2) included 2,4-DAPDHs such as AmDH4, group 3 (G3) *Msme*AmDH and *Micro*AmDH, group 4 (G4) *Cfus*AmDH, group 1 (G1) enzymes with yet unidentified substrate but presumably active towards substrates bearing a carboxylic moiety, and group 5 (G5) enzymes probably without any imine formation capability due to the absence of the key Glu in P3 (Figure 16). Nevertheless, the G5 enzymes were classified in this family due to their sequence homology (Mayol *et al*., 2019).

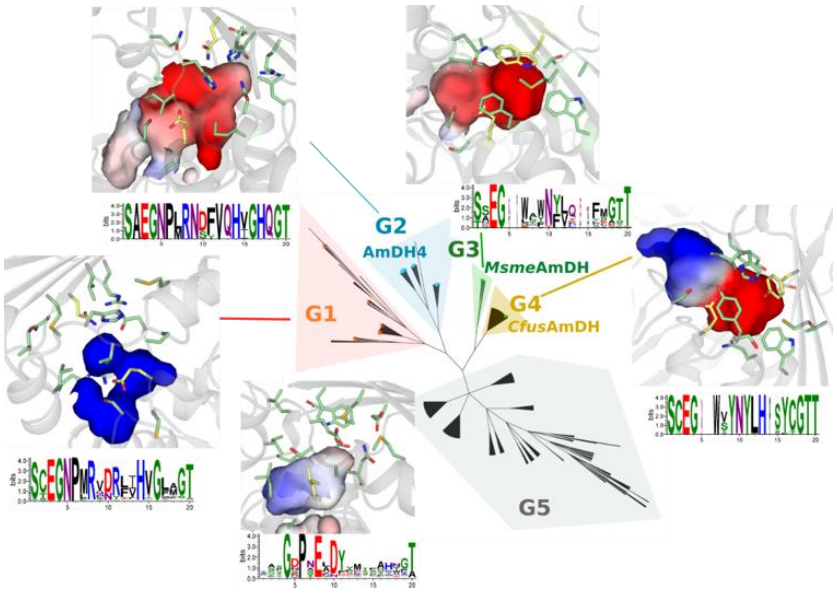
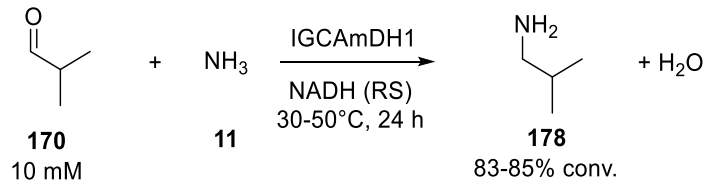


Figure 16. ASMC1 clusterization. This hierarchical tree of active sites was generated by the ASMC pipeline and includes 2,011 non-redundant proteins (redundancy was removed at 100% of similarity). Each group of active site homologs (G1-G5) is represented by a logo of P1-P20 positions and a representative structure in which P1-P20 residues are displayed in stick format. In each logo, the height of the letters reflects conservation of the active site residues while the color code used refers to the chemical properties of the corresponding residue [green: polar residues, black: hydrophobic residues, red: acidic residues, blue: basic residues and purple: neutral residue (at physiological pH)]. Adapted from Mayol *et al.*, 2019.

A focus has been done on the enzymes active towards non Ω-keto-α-amino acids as the one in G2 group. The goal being to better understand the potential of nat-AmDHs to aminate carbonyl substrates with no functional requirement, the light has been shed on *Cfus*AmDH and *Msme*AmDH P1-P20 homologs, namely G3 and G4 groups. Also, the enzymes belonging to G5 group did not share any particular P1-P20 consensus and was therefore not considered. During the deeper analysis of the active site pockets of G3 and G4 members during the first year of my PhD work, a new position was added between P4 and P5 (see Chapter III, I.1). This spatial position is occupied by hindered residues in *Cfus*AmDH and *Msme*AmDH but by a short one without any suspected role in AmDH4; this position was therefore not considered in the first ASMC classification based on AmDH4. The updated P1-P21 numbering will be used in the rest of this manuscript. As exemplified by the structures of AmDH4, *Msme*AmDH and *Cfus*AmDH, some variability in P1-P21 can be observed among the family (Appendix 5). For example, other residues than Tyr at position P10 (Y168, *Cfus*AmDH numbering) hypothesized to bind the ketone, can be found and all the same for position P12 (F169, *Msme*AmDH numbering) thought to be essential for closing the active site during catalysis.

II.2.6.3 Expansion of the family by exploring metagenomic databases

In 2020, thanks to a collaboration between our team and the one of A. Bommarius from Georgia Tech Institute, Caparco *et al.* identified other members of this family among metagenomics databases, hitherto unexplored for their biocatalytic resources. These databases comprise IGC (Integrated Gene Catalog) sourced from the human microbiome, OM-RGC (Ocean Microbial Reference Gene Catalog) and MATOUv1 (Marine Atlas of Tara Oceans Unigenes) from marine samples resulting from the worldwide expedition Tara Oceans whose Genoscope was the main sequencing center (Caparco *et al.*, 2020). Some of the selected enzymes and, in particular, MATOUAmDH2 from an eukaryotic organism, displayed the highest native activities recorded to date among nat-AmDHs towards aliphatic aldehydes, such as isobutyraldehyde (**170**) (11.1 U mg^{-1}) and pentanal (**175**) (0.6 U mg^{-1}). Above all, this enzyme with low homology with already characterized nat-AmDHs (< 34%) also showed high activity towards 1,2-cyclohexadione (**176**) and slight activity towards bulkier substrates such as norcamphor (**177**) or benzaldehyde (**31**), not substrates of previously identified nat-AmDHs. The structure of this enzyme and the 17 other identified ones were modeled using *Cfus*AmDH and *Msme*AmDH as templates. The noticeable differences in terms of P1-P21 positions can explain their different cofactor specificities (NADH-dependent enzymes, excepting MATOUAmDH2) and slightly different substrate scope compared to previously reported nat-AmDHs. Further studies including resolution of some 3D-structures and protein engineering were required to better link structure and activity of these recently discovered enzymes which have a synthetic potential, as exemplified by the 85% conversion of 10 mM **170** into isobutyrylamine (**178**) by IGCAmDH1 at 30°C or even 50°C (Scheme 28). Additional results have been obtained during my PhD and are detailed in this manuscript, notably the resolution of MATOUAmDH2 which have provided interesting insights (see Chapter II, I.1.1 and Chapter IV, II.5.2) (Bennett *et al.*, 2022).



Scheme 28. Example of reductive amination reaction catalyzed by IGCAmDH1 obtained from metagenomic data. RS = recycling system.

II.2.6.4 Summarized work on the discovery and engineering of native amine dehydrogenases

Figure 17 and Table 3 illustrate the diversity of P1-P21 residues in the nat-AmDH family through a logo representation and a table of the conservation patterns of the main residues defining the catalytic pocket of the characterized members of this superfamily. The very restricted protein engineering work carried out on AmDH4 before the start of my PhD project, is also illustrated (Table 3).

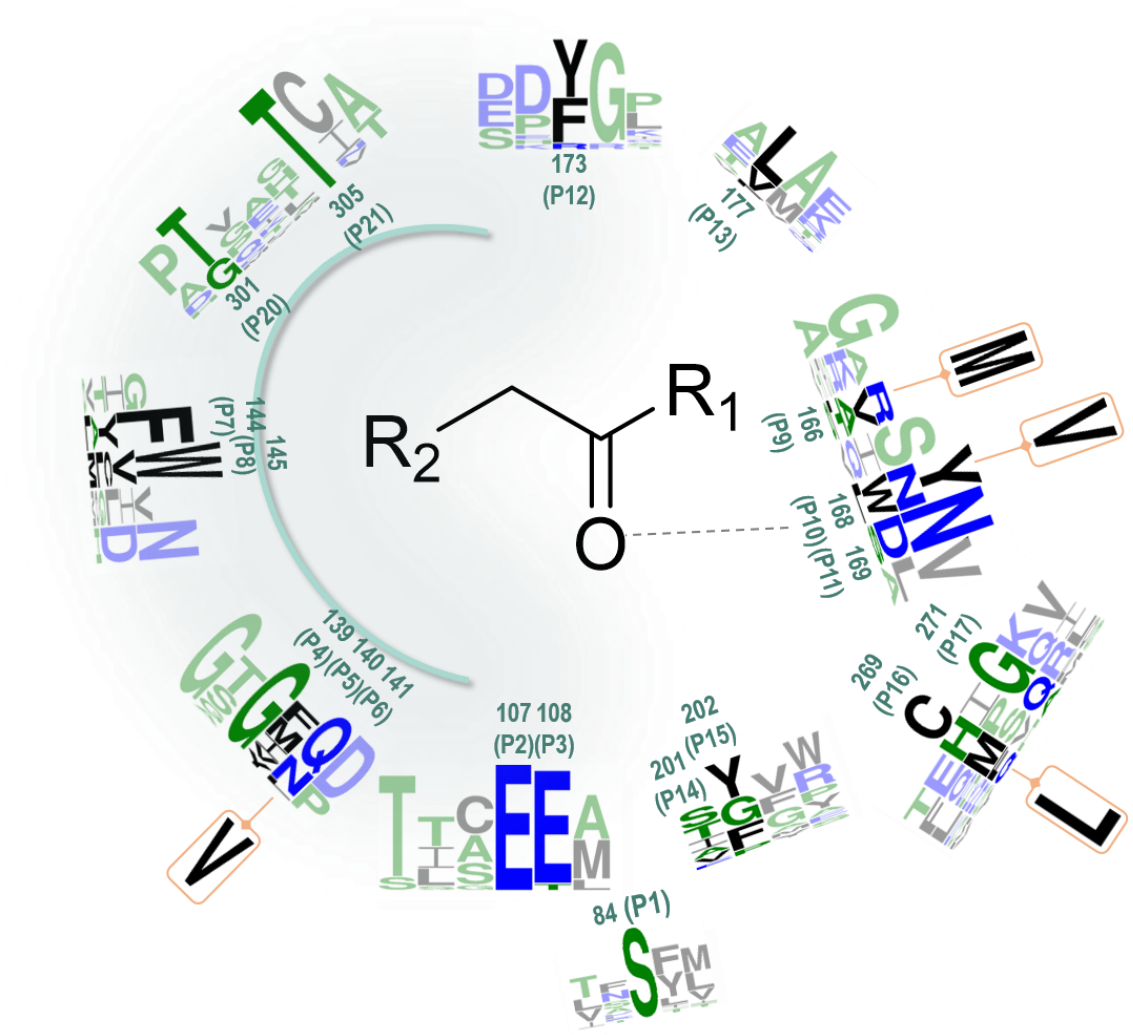


Figure 17. Logo representation of the pocket diversity found in the nat-AmDHs family and mutation work done on AmDH. This representation shows the conservation patterns of the main residues defining the catalytic pocket of characterized native AmDHs (AmDH2, AmDH4-5, AmDH7-9, *Msme*AmDH, *Cfus*AmDH, *Micro*AmDH, *Apau*AmDH, *Myco*AmDH, *Mvac*AmDH, *Chat*AmDH, *Sgor*AmDH, *Acol*AmDH, IGCAmDH1, IGCAmDH5 and MATOUAmDH2). For clarity, positions P13 and P14 have been removed from this Figure. The two neighboring residues before and after these residues are included as part of the logos to illustrate their environment. The mutated amino acids in AmDH4 engineering are placed above the corresponding WT residue and are highlighted in orange. The sequence numbering used is that of *Cfus*AmDH. The color code is based on hydrophobicity (low hydrophobicity: blue, middle hydrophobicity: green, high hydrophobicity: black). Logos were generated using the webservice WebLogo3. Extracted from Ducrot *et al.*, 2020.

Table 3. List of main residues defining the catalytic pocket in WT nat-AmDHs (black) and the corresponding mutated residues in the engineered enzyme (orange). AmDH4 mutants produced to attest the role of specific residues of the active site are not given in the table.

Enzyme name	Uniprot ID	Selected residues (*Cfus* AmDH numbering)																			
		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P16	P17	P18	P19	P20	P21
		84	107	108	139	140	141	144	145	166	168	169	173	177	201	202	269	271	301	305
AmDH2	A7HNJ8	S	A	E	G	V	N	F	V	R	N	D	F	V	I	G	H	Q	G	T
AmDH4	A9BHL2	S	A	E	G	V	N	F	V	R	N	D	F	V	I	G	H	Q	G	T
AmDH4_M1							V			M	V						L			
AmDH5	E3PY99	S	A	E	G	I	N	L	I	R	N	S	F	V	V	G	H	Q	G	T
AmDH7	D2Z5Z0	S	A	E	G	I	N	L	M	R	N	S	F	V	V	G	H	Q	G	T
AmDH8	D7DCF5	S	C	E	G	I	N	F	L	R	L	D	R	F	V	G	F	A	G	T
AmDH9	D9RY14	S	A	E	G	I	N	F	V	R	N	D	F	V	I	G	H	Q	G	T
Msme AmDH	A0A0D6I8P6	S	S	E	G	Y	Q	F	W	A	W	N	F	L	T	F	M	G	T	T
Cfus AmDH	S9Q235	S	G	E	G	F	Q	F	W	T	Y	N	Y	L	S	Y	C	G	T	T
Micro AmDH	C3UMY1	S	S	E	G	F	Q	F	W	S	W	N	Y	L	T	F	M	G	T	T
Mvac AmDH	K0UKT5	S	S	E	G	F	Q	F	W	A	F	N	F	L	T	F	M	G	T	T
Apau AmDH	E3CZE3	S	C	E	G	Y	Q	F	W	T	Y	N	Y	L	S	Y	C	G	T	T
Myco AmDH	A0A101AWU7	S	S	E	G	F	Q	F	W	A	W	N	F	L	T	F	M	G	T	T
Sgor AmDH	A8AV93	S	C	E	G	M	Q	Y	W	V	Y	N	Y	L	S	Y	C	G	T	T
Chat AmDH	D3AE01	S	C	E	G	M	Q	Y	W	V	Y	N	Y	L	A	Y	C	G	T	T
Acol AmDH	B0P810	S	C	E	G	M	Q	F	W	V	Y	N	Y	L	A	Y	C	G	T	T
MATOUAmDH2		S	A	E	G	F	L	A	C	L	Y	N	Y	L	S	Y	Q	G	T	T
IGCAmDH1		S	C	E	G	M	Q	F	W	V	Y	N	Y	L	A	Y	C	G	T	T
IGCAmDH5		S	C	E	G	M	Q	Y	W	V	Y	N	Y	L	S	Y	C	G	T	T

Figure 18 highlights the main substrates of native AmDHs discovered throughout the different (meta)genomic minings. As the 4OP-AmDHs, these other nat-AmDHs preferentially formed the (*S*)-amines, in contrast to engineered AADHs and LE-AmDH-v1, each of which giving the (*R*)-enantiomer for the reaction with **11** (following Cahn-Ingold-Prelog priority rules in the case of unfunctionalized carbonyl substrates). The latter is the amine source preferentially accepted by these native enzymes, but secondary activity towards **38** was also recorded for *Msme*AmDH, *Micro*AmDH, *Cfus*AmDH, *Mvac*AmDH and *Apau*AmDH towards **170** and/or **45**. The comparison with their activity in presence of **11** is difficult as the assays were done using with both NADH and NADPH whether used separately or as a mix. However, it is interesting to note that the activity towards **45** in presence of **38** greatly decreased for all enzymes (*e.g.* *Micro*AmDH activity of 90.4 mU mg^{-1} with **38** and a mix NADH/NADPH while being 614.5 mU mg^{-1} and 388.5 mU mg^{-1} with **11** in presence of NADH and NADPH, respectively. With **170**, the influence of the use of **38** as amine source was enzyme-dependent for the enzymes tested (Mayol *et al.*, 2019). Interesting new insights on a residue identification to expand the amine substrate scope was permitted thanks to MATOUAmDH2 structure and subsequent protein engineering work carried out by the team of Prof. Grogan and are described in Chapter IV, II.5.2.

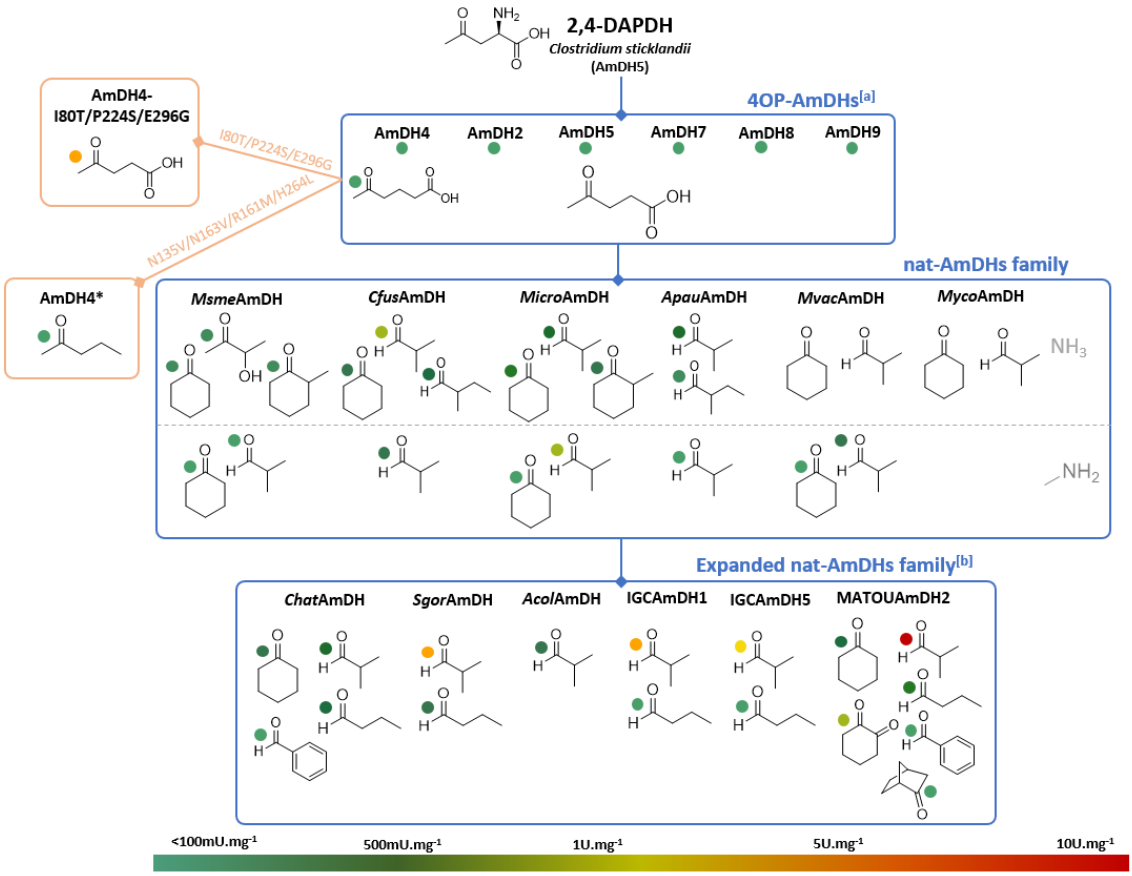


Figure 18. Substrate specificity of the nat-AmDH family. The substrates associated to each enzyme were selected as the ones displaying the highest activities or considered as "new" substrates compared to the substrate scope of the whole family. If not mentioned, the amine substrate is **11**. The blue and orange lines correspond to biodiversity mining and mutation work, respectively. The specific activities are given for information but not to be compared because all the assays were not done with the exact same reaction conditions. The following ones correspond to a range of typical conditions: 10 mM carbonyl substrate, 0.2 mM NAD(P)H cofactor, 1-2 M ammonium-based buffer pH 8-9.5, 20-25°C. For the sake of clarity, the numbering of the molecules is not indicated. [a] Activity assays done with 100 mM carbonyl substrate and 0.3 M ammonium-based buffer [b] Activity assays performed at 50°C. Adapted from Ducrot *et al.*, 2020.

II.3 Conclusion on the state-of-the art on biocatalysts for amine synthesis

Paragraph III summarizes the considerable effort exert in the discovery of diverse enzyme templates for reductive amination reaction. The main families, namely AmDHs from engineering on AADHs, native AmDHs and RedAms/IREDs complement each other reaching a diversity of carbonyl and amine substrate scopes and enantioselectivities. Although they are still less described and applied in synthesis than the TAs, they provide a good alternative to this family that can suffer from waste induced by the sacrificial amine donor and the unfavored thermodynamic equilibrium. However, they still require further improvements to reach competitive broad substrate spectrums and catalytic efficiency for application in industrial synthesis.

III. Conclusion, objectives and strategy

A vast diversity of approaches are currently used and investigated to find always more efficient biocatalysts applicable in synthesis. These encompass either to find new native sequences among biodiversity, for example by taking advantage of the source microorganism features and natural environment, to improve the already known sequences through random to target engineering or to design a biocatalyst on demand to satisfy a specific functionality in synthesis. The richness of techniques available always provides to the researchers one method that will suit their needs and constraints in terms of biochemical or structural data available, screening capacities or bioinformatic expertise.

The amine moiety being ubiquitous in synthetic compounds either in pharmaceuticals, agrochemicals or general fine chemicals, the direct reductive amination is a highly desirable reaction for the chemists and for researchers in biocatalysis to apply it to greener processes. From the pioneer work of Bommarius and co-workers in 2012 with the engineering of AADHs, the discovery of native AmDH sequences in our group to that of RedAms and IREDs by the group of Prof. Nicholas Turner, this portfolio of enzymes has greatly expanded thanks to a huge effort of biodiversity mining and engineering work to provide a large number of enzyme templates for reductive amination. Some of them were already applied for the synthesis of compounds of interest such as Rotigotine (**92**) or lysine-specific demethylase-1 inhibitor GSK2879552. About the native AmDH family, the previous work of Ombeline Mayol in our laboratory enabled the discovery and characterization of the first AmDH native sequences that was completed with the involvement of Adam Caparco for the expansion of the family through metagenomic mining. A collaboration with the team of Prof. Grogan resulted in the first crystallographic data for some nat-AmDHs but no deep study was carried out to further understand the nat-AmDHs structure and mechanism or try to improve them through engineering. In line with the objectives of the

research unit, the biocatalysis team focused on the first steps of biocatalyst discovery and improvement Thus, we aim at giving access to new templates that could be used by other groups for deeper catalytic improvement and process optimization for application in synthesis. In this context, the objectives my PhD project were as follows:

- Gain more insights into the nat-AmDHs from a structural point of view but also through biochemical characterizations of their stability, substrate scope, stereoselectivity, cofactor preference, etc.
- Provide first improvements of the nat-AmDHs through protein engineering or *in vivo* evolution to expand their substrate scope, increase their stability or modify their cofactor preference
- Identify novel nat-AmDHs in the biodiversity with native target features of enlarged substrate scope or stereoselectivity

The results corresponding to the research work carried out to achieve these objectives are discussed in the Chapter II, III and IV, respectively.

CHAPTER II

Characterization and comprehension of native amine dehydrogenases

With the aim to find new nat-AmDHs that could perform the desired reaction with a larger carbonyl or amine substrate scope, exhibiting a different cofactor specificity or even the opposite enantioselectivity, it was important to further characterize the set of nat-AmDHs we already have in hands. Taking advantage of the 3D crystallographic structures available and the possibility to generate homology models, we provide in this chapter a structural description of the nat-AmDHs. The goal was to complete the first descriptions published by Mayol *et al.* (2019) and Caparco *et al.* (2020) by studying the global structures and active sites of the already described nat-AmDHs and the members of the G3 and G4 groups of the ASMC1 classification (see Chapter I, II.2.6.2, Figure 16) (Mayol *et al.*, 2019). Our knowledge on the active site and mechanism was completed with a mutational work that aimed at providing further explanations of the role of some of the P1-P21 residues. Following on from the collaboration established with the team of Prof. Bommarius in the discovery of nat-AmDHs in metagenomic data (see Chapter I, II.2.6.3), we investigated the stability of some nat-AmDHs to complete the only experimental data reported yet on melting temperatures obtained by Adam Caparco (Caparco *et al.*, 2020). The study of their carbonyl substrate scope was also widened with a particular interest for the short (hydroxy)ketones and phenyl substituted aldehydes, still not previously studied. Eventually, a very preliminary work is described to estimate the potential of the nat-AmDHs to perform reductive sulfidation in presence of a sulfur donor instead of **11**.

I. Structural description

I.1 Global structure

I.1.1 Description

As mentioned in Chapter I, II.2.6.2, 3D-structures of AmDH4 (PDB: 6G1H, 6G1M), *Cfus*AmDH (PDB: 6IAU) and *Msme*AmDH (PDB: 6IAQ) were solved before the beginning of my PhD project, by our collaborators, the Laboratory of Structural and Applied Enzymology of Prof. Gideon Grogan from the University of York. During my thesis, the structure of MATOUAmDH2 (PDB: 7ZBO, 7R09) was obtained

by his PhD student, Megan Bennett (Bennett *et al.*, 2022). I had the opportunity to exchange with her and Pr. Grogan, as I was studying the corresponding model of this enzyme. Her main objectives were to solve the structure of MATOUAmDH2 and modify it by targeted protein engineering to gain more insights into its particular substrate scope. Our results were interesting to share as they were beneficial for each of our research goals. I gained a first expertise into the field of enzyme modeling and 3D analysis thanks to a training that I attended at the start of my project ["Modélisation 3D des protéines", taught by Véronique Martin and Dr. Gwenaëlle André-Leroux from the Applied Mathematics and Computer Science, from Genome to Environment (MaIAGE) Research Unit at the National Research Institute for Agriculture, Food and the Environment (INRAE)]. AmDH4 structure was obtained in complex with NAD^+ (PDB: 6G1H) and in complex with $NADP^+$ and $PO4^-$ (PDB: 6G1M). *Cfus*AmDH was crystallized in complex with $NADP^+$ and cyclohexylamine (**179**) while *Msme*AmDH was obtained only in complex with $NADP^+$. All the crystallized structures revealed a dimeric structure similar to meso-diaminopimelate dehydrogenases, with a well-conserved Rossmann fold in the N-ter domain, but significant differences in the C-ter β-sheet domain (Figure 19). The two monomers are linked thanks to the interaction of an α-helix (F140-G154) located in the N-ter and a β-strand (D291-R297) of the C-ter. The $NADP^+$ cofactor was bound within the cleft between the two domains in each monomer.

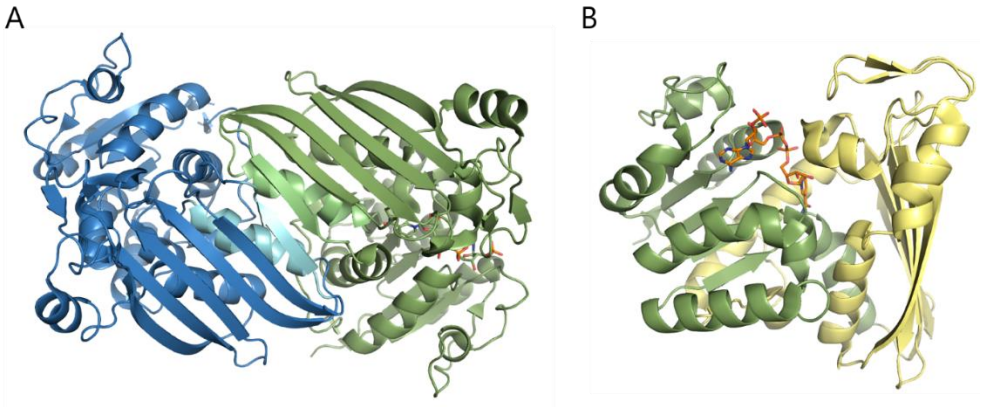


Figure 19. Structure of *Cfus*AmDH. (A) Dimeric structure of *Cfus*AmDH (PDB: 6IAU). Chain A and B are represented in blue and green, respectively. The α-helix and the β-strand that form the interactive segments between the two monomers are represented in light blue and light green, respectively. (B) Monomeric structure of *Cfus*AmDH (PDB: 6IAU, chain B) in complex with $NADP^+$ (orange). N-ter and C-ter are highlighted in green and yellow, respectively.

The superposition of the two monomers of AmDH4 revealed that the structure can adopt two conformations, open and closed. While the N-ter folding remains unchanged, the helical segment N172-S189 and the loop T237-G259 are highly displaced between the open and the closed conformation. For example, in AmDH4, for the helical segment and the loop, a distance of 9.1 Å and 10.3 Å, respectively, can be observed between a C-atom of the backbone in chain A (open) and the same atom in chain C (closed). MATOUAmDH2, obtained as two dimers, also displaying open and closed conformations,

seems to adopt a more opened conformation than any other nat-AmDH structure. Indeed, a shift of only 4.4 Å and 6.6 Å of the helical segment and the loop, respectively, can be measured between the more opened and the closest conformation (Figure 20). In nat-AmDHs, this movement could enable the active site closing to bring the key residues and the substrates in closer proximity to start the reaction. For instance, in MATOUAmDH2 open form, E111 (P3) is 8.9Å and 11.6 Å away from Y171 (P10) and Y176 (P12), respectively, while in its closed form, P3 is at a distance of 7.2Å and 7.4 Å of these same residues [distances between E111 (OE2) and Y171 (CZ) and Y176 (CB)]. Notably, this displacement is accompanied by a torsion of the β-strand V227-T237 (Figure 20B).

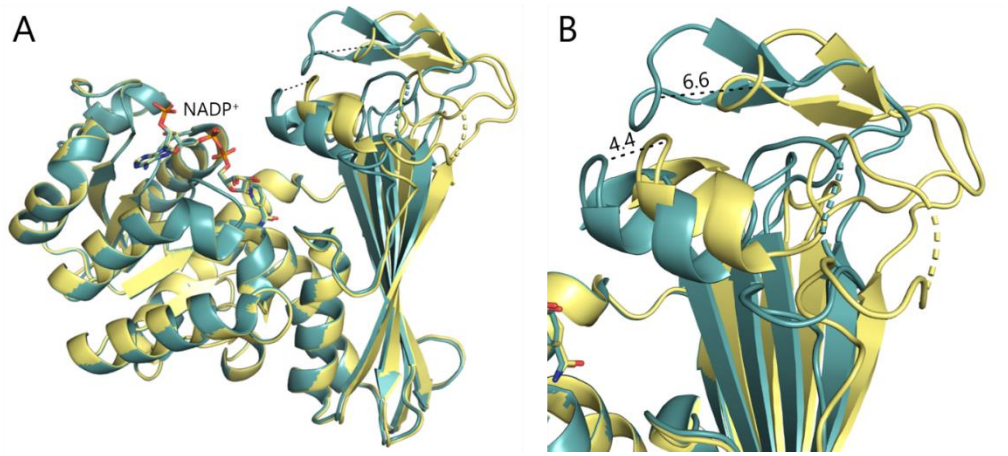


Figure 20. Structure of MATOUAmDH2 in complex with $NADP^+$ (PDB: 7ZBO) with alignment of chains C and D. (A) Overall view of the open chain (chain D, in yellow) and the closest chain (chain C, in blue) (B) View focused on the highly flexible sections described here. Distances are given in Angstroms (Å). RMSD (MATOUAmDH2 chain C – chain D) = 0.237Å

We recently gained an additional understanding of these enzymes through MD simulations carried out for the study of the substrate spectrum expansion described in Chapter III, I.4.7.2.

I.1.2 Diversity within the reported native amine dehydrogenases

The nat-AmDHs with validated activity available at the start of the project shares between 19.6% and 89.9% identity over the whole length of the sequence (Figure 21). When considering the diversity of nat-AmDHs described by the ASMC analysis (Mayol *et al.*, 2019) and particularly *Cfus*AmDH and *Msme*AmDH P1-P21 homologs (G3 group, 28 members and G4 group, 263 members, respectively) (Figure 16) the global structure is conserved with a N-ter Rossmann fold domain and C-terminal domain comprising a seven-stranded β-sheet and the cofactor $NAD(P)^+$ bound within the cleft between the two domains.

	AmDH4	*Tther* AmDH	MATOUAmDH1_design	MATOUAmDH2_design	OM-RGCAmDH2	*Rgna* AmDH	*Dfai* AmDH	*Cfus* AmDH	*Sgor* AmDH	*Apau* AmDH	IGCAmDH2	IGCAmDH6	IGCAmDH4	IGCAmDH5	IGCAmDH3	*Crypt* AmDH	IGCAmDH1	*Chat* AmDH	*Acol* AmDH	OM-RGCAmDH1	*Micro* AmDH	*Porti* AmDH	*Myco* AmDH	*Msme* AmDH	*Mvac* AmDH
AmDH4	100	53.6	24.1	22.2	22.6	23.3	24.2	22.4	23.3	22.4	23.0	25.1	24.7	25.2	23.6	26.9	25.7	23.6	24.5	24.2	25.4	24.2	25.1	26.1	24.6
Tther AmDH	53.6	100	22.0	20.8	28.6	24.5	25.2	25.2	26.0	26.7	26.5	30.6	30.9	28.9	28.1	28.6	27.1	27.8	28.8	29.1	30.4	28.9	28.6	30.8	30.5
MATOUAmDH1_design	24.1	22.0	100	52.7	21.2	32.3	34.1	32.4	30.2	32.0	32.9	33.7	34.0	33.8	30.2	35.9	33.7	32.6	33.4	29.1	27.8	29.4	29.4	29.9	30.1
MATOUAmDH2_design	22.2	20.8	52.7	100	19.6	31.8	34.8	32.9	28.8	33.4	33.2	33.9	31.9	34.3	30.2	34.8	32.4	32.3	33.3	28.1	31.3	32.3	30.2	32.6	32.3
OM-RGCAmDH2	22.6	28.6	21.2	19.6	100	25.8	29.1	27.4	29.4	28.8	27.0	27.6	28.8	27.4	27.6	26.1	27.0	27.6	27.3	32.9	34.2	32.8	32.1	33.0	31.8
Rgna AmDH	23.3	24.5	32.3	31.8	25.8	100	59.8	49.3	44.6	53.5	51.8	47.2	45.7	45.9	49.0	48.7	49.7	51.0	49.1	34.5	35.1	37.8	35.4	35.9	34.8
Dfai AmDH	24.2	25.2	34.1	34.8	29.1	59.8	100	49.0	45.5	59.1	51.9	49.6	49.0	52.7	50.9	50.9	52.3	51.2	51.0	36.9	35.8	38.1	37.0	40.1	38.9
Cfus AmDH	22.4	25.2	32.4	32.9	27.4	49.3	49.0	100	40.6	58.6	49.7	37.9	39.9	45.3	46.9	43.8	45.9	47.1	46.2	33.9	37.9	37.1	37.3	37.1	37.5
Sgor AmDH	23.3	26.0	30.2	28.8	29.4	44.6	45.5	40.6	100	52.2	56.7	53.7	58.6	52.9	58.3	54.7	55.8	59.6	56.0	36.8	40.4	40.6	37.4	38.5	38.6
Apau AmDH	22.4	26.7	32.0	33.4	28.8	53.5	59.1	58.6	52.2	100	65.6	50.4	52.3	57.4	61.0	57.5	61.6	60.4	59.5	40.7	38.1	41.2	38.4	38.5	38.2
IGCAmDH2	23.0	26.5	32.9	33.2	27.0	51.8	51.9	49.7	56.7	65.6	100	57.4	58.6	59.2	65.7	63.5	65.0	66.0	63.9	36.7	38.5	39.8	35.9	39.7	39.8
IGCAmDH6	25.1	30.6	33.7	33.9	27.6	47.2	49.6	37.9	53.7	50.4	57.4	100	66.7	64.5	62.9	65.8	64.6	67.3	65.2	39.8	38.7	40.3	39.6	39.4	37.6
IGCAmDH4	24.7	30.9	34.0	31.9	28.8	45.7	49.0	39.9	58.6	52.3	58.6	66.7	100	62.9	60.7	68.0	65.7	69.6	67.0	39.0	39.1	41.2	38.4	39.7	39.4
IGCAmDH5	25.2	28.9	33.8	34.3	27.4	45.9	52.7	45.3	52.9	57.4	59.2	64.5	62.9	100	67.0	69.7	68.5	71.7	67.4	38.2	40.6	40.7	39.7	41.6	39.3
IGCAmDH3	23.6	28.1	30.2	30.2	27.6	49.0	50.9	46.9	58.3	61.0	65.7	62.9	60.7	67.0	100	65.4	70.4	73.6	67.9	34.4	41.5	40.3	38.0	38.5	36.9
Crypt AmDH	26.9	28.6	35.9	34.8	26.1	48.7	50.9	43.8	54.7	57.5	63.5	65.8	68.0	69.7	65.4	100	73.9	77.7	73.3	37.9	37.1	41.5	39.1	42.1	39.5
IGCAmDH1	25.7	27.1	33.7	32.4	27.0	49.7	52.3	45.9	55.8	61.6	65.0	64.6	65.7	68.5	70.4	73.9	100	80.1	75.4	36.5	38.3	41.5	38.5	38.2	37.1
Chat AmDH	23.6	27.8	32.6	32.3	27.6	51.0	51.2	47.1	59.6	60.4	66.0	67.3	69.6	71.7	73.6	77.7	80.1	100	82.0	37.9	41.2	41.5	37.4	38.8	37.4
Acol AmDH	24.5	28.8	33.4	33.3	27.3	49.1	51.0	46.2	56.0	59.5	63.9	65.2	67.0	67.4	67.9	73.3	75.4	82.0	100	38.9	40.8	40.6	37.8	39.1	37.1
OM-RGCAmDH1	24.2	29.1	29.1	28.1	32.9	34.5	36.9	33.9	36.8	40.7	36.7	39.8	39.0	38.2	34.4	37.9	36.5	37.9	38.9	100	41.5	44.1	42.4	43.2	42.6
Micro AmDH	25.4	30.4	27.8	31.3	34.2	35.1	35.8	37.9	40.4	38.1	38.5	38.7	39.1	40.6	41.5	37.1	38.3	41.2	40.8	41.5	100	64.1	63.3	66.0	64.8
Porti AmDH	24.2	28.9	29.4	32.3	32.8	37.8	38.1	37.1	40.6	41.2	39.8	40.3	41.2	40.7	40.3	41.5	41.5	41.5	40.6	44.1	64.1	100	69.7	68.0	66.5
Myco AmDH	25.1	28.6	29.4	30.2	32.1	35.4	37.0	37.3	37.4	38.4	35.9	39.6	38.4	39.7	38.0	39.1	38.5	37.4	37.8	42.4	63.3	69.7	100	74.9	74.6
Msme AmDH	26.1	30.8	29.9	32.6	33.0	35.9	40.1	37.1	38.5	38.5	39.7	39.4	39.7	41.6	38.5	42.1	38.2	38.8	39.1	43.2	66.0	68.0	74.9	100	89.9
Mvac AmDH	24.6	30.5	30.1	32.3	31.8	34.8	38.9	37.5	38.6	38.2	39.8	37.6	39.4	39.3	36.9	39.5	37.1	37.4	37.1	42.6	64.8	66.5	74.6	89.9	100

Figure 21. Sequence identity matrix of the reported nat-AmDHs. Sequence identity is highlighted in shades from white and light grey for low values to orange and red for high values. This matrix was done using Clustal Omega webservice (Goujon *et al.*, 2010; McWilliam *et al.*, 2013; Sievers *et al.*, 2011).

The comparison of the 3D-structures of *Cfus*AmDH, *Msme*AmDH and MATOUAmDH2, and the models generated for the proteins with no 3D-structure available revealed only few structural differences. Notably, the segment D186-S210 (*Cfus*AmDH numbering), located between the helical section of the closing loop (N142-S189) and Y202 of the active site, adopt generally a helical configuration (3-4 turns) with a linker segment to the Y202-L214 helix but in some cases, such as in MATOUAmDH2, *Rgna*AmDH and *Dfai*AmDH, this section is 6-to-16-residues longer and cannot follow this particular 3D arrangement (Figure 22). It might have an impact on the overall enzyme folding or flexibility that could be related to the protein stability but no further analysis was done. However, it is important to have in mind that such comparisons are made using models, which are highly dependent upon the template used to generate them. This is all the more the case when the sequence homology is low. Now that artificial intelligence systems, such as AlphaFold2, were proven highly accurate, this structural prediction should also be used to remove the bias of the template in the case of too low sequence identity and allow us to compare structures without having resolved 3D-structure.

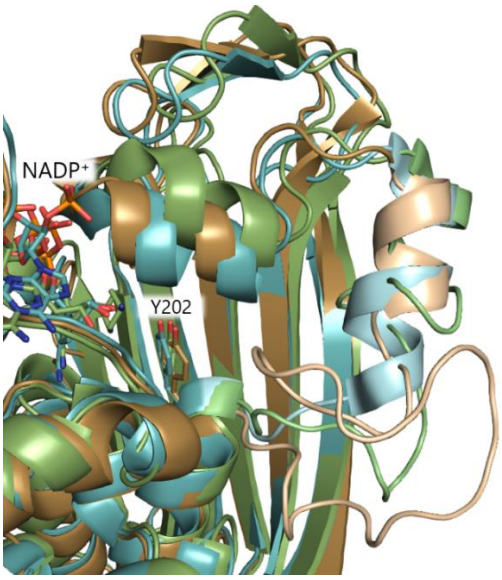


Figure 22. Variable helix in nat-AmDHs. Alignment of *Cfus*AmDH (blue) and MATOUAmDH2 (green) and *Rgna*AmDH (brown). The segments D186-S201 (*Cfus*AmDH numbering) are highlighted in light blue, light green and light brown in *Cfus*AmDH, MATOUAmDH2 and *Rgna*AmDH, respectively. Root Mean Square Deviation (RMSD) (MATOUAmDH2-*Cfus*AmDH) = 1.376 Å ; RMSD (*Rgna*AmDH-*Cfus*AmDH) = 0.898 Å.

II. Comprehension of the role of the essential residues in the active site

Among all the residues that form the active site, only P3 (E108 in *Cfus*AmDH), was already evidenced to have a key role for both fixation and activation of NH_3, with a drastic drop of activity measured with the corresponding Ala mutant (Mayol *et al.*, 2019). Before my work, other positions had already been hypothesized to (un)directly take part in the enzymatic mechanism. These first assumptions were based on the occurrence of the residue in the nat-AmDH sequences, their positioning inside the active site in the crystal structures, their distance to the crystallized or modelized product or substrate and the knowledge of the global mechanism of a reductive amination. P12 (Y173) may enable to close the active site, P17 (Y202) to anchor the P3 catalytic residue in a favorable position towards NH_3 and the cofactor and P10 (Y168) to ensure the well positioning of the substrate in the active site (Figure 23). For clarity, the active site is generally observed from the point of view shown in Figure 23 where the active site part constituted by, among others, L177 (P13) and Y173 (P12) is called the active site ceiling while the part formed by Q141 (P6) or W145 (P8) is called the active site floor.

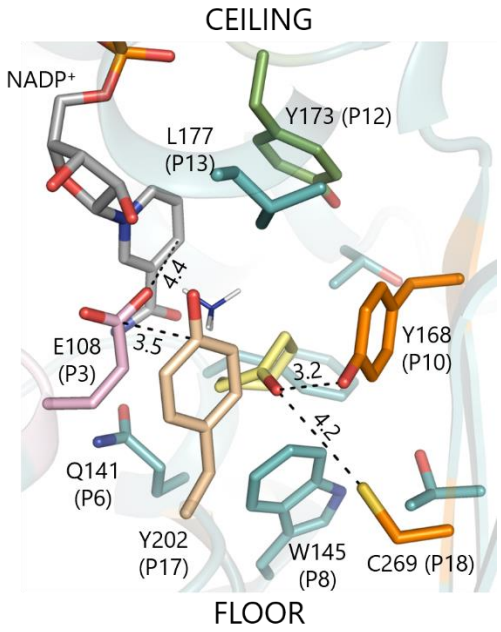


Figure 23. *Cfus*AmDH active site. The positions studied in this part are highlighted as follows: P3 (E108) in pink, P12 (Y173) in green, P17 (Y202) in beige, P10 and P18 (Y168 and C269, respectively) in orange and other residues forming the active site in blue. $NADP^+$ and **45** are represented in grey and yellow, respectively. Distances are given in Angstroms (Å).

We carried out the same strategy by mutating specific residues into unfunctional or smaller residues and test their activity to check their key role in the reaction. The mutations to applied were determined either based on pure structural considerations or on the occurrence of some residues in the biodiversity. The "biodiversity" considered here is the set of G3 and G4 groups, gathering *Cfus*AmDH and *Msme*AmDH P1-P21 homologs obtained from the first ASMC (Mayol *et al.*, 2019). The diversity of P1-P21 in G3 and G4 is given in Figure 24 and Appendix 5. However, it has to be noted that the diversity within the same ASMC group is, by definition, limited as the classification is done based on position homology. This work was first to highlight the role of some active site residues but was also used to determine a minimum active site required for the structural biodiversity exploration to find structural analogs, detailed in Chapter IV, II.4.

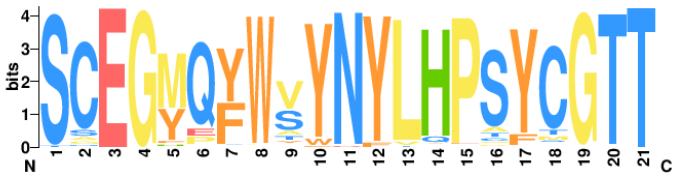


Figure 24. Logo representation of the conservation pattern of P1-P21 positions among ASMC1 G3 and G4 groups. The color code used refers to the polarity and charge of the corresponding residue [blue: polar residues, yellow: hydrophobic residues, orange: aromatic residue, red: negatively charged residues, and green: positively charged residues (charges at physiological pH)]. The logo was generated using the webservice WebLogo3.

II.1 P3 for catalytic activity

In the nat-AmDHs, the catalytic position P3 is mainly occupied by a Glu residue positioned with OE2 atom at around 4.5 Å of the cofactor C4N atom allowing the hydride release for the imine reduction. The only known exception is OM-RGCAmDH2 (Caparco *et al.*, 2020) which harbors a shorter Asp residue at this position. However, this enzyme could only be produced at really low expression rate thus preventing any further investigation. Bearing the same chemical function but on a shorter side chain than Glu, Asp residue could fix NH_3 but maybe closer to the active site floor, and therefore influence the enantioselectivity. To demonstrate the possibility of having an Asp at P3 the mutant *Cfus*AmDH-E108D was produced by targeted protein engineering, overexpression in *E. coli* (see Experimental section, II.2, II.3 and II.4) and purification on a Ni-nitrolo acetic acid (NTA) column thanks to a hexahistidine sequence tag (His-tag) placed at the N-ter. The mutant was assayed towards a range of carbonyl substrates [isobutyraldehyde (**170**), pentanal (**175**), 1-hydroxypropan-2-one (**180**), cyclohexanone (**45**), pentan-2-one (**112**), pentan-3-one (**181**)] in presence of either NADH or NADPH (Figure 25). The common spectrophotometric assay, detailed in the Experimental section, III.1, is based on the monitoring at 340 nm of the decrease concentration of the reduced form of NAD(P)H (see Chapter I, II.2.6.2, Scheme 27). Unless when written or indicated with error bars, all the specific activities described in Chapter II, II, were calculated from one replicate.

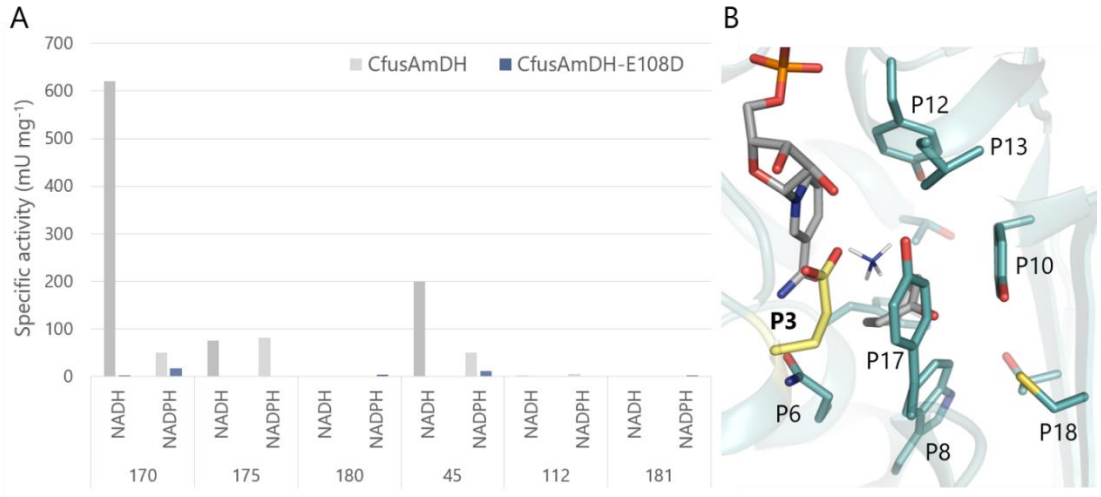


Figure 25. Influence of mutations at P3 in *Cfus*AmDH. (A) Specific activity (mU mg^{-1}) of *Cfus*AmDH and its P3 variant. The reaction conditions were as follows: final volume 100 µL, 10 mM carbonyl substrate, 1 M NH_4HCO_2 buffer pH 8.5, 0.2 mM NADPH or NADH, 0.03-0.15 mg mL^{-1} purified enzyme, room temperature (RT). (B) *Cfus*AmDH active site with P3 highlighted in yellow.

*Cfus*AmDH was inactivated upon E108D mutation towards a range of typical substrates, proving that, in *Cfus*AmDH-like structure, an Asp residue at P3 is not correctly positioned to fix NH_3 in the active site. This may be due to a longer distance found between the O of the Asp carboxylate and the C4N of

the cofactor, estimated > 5 Å in PyMol. The residual activity obtained towards **170** and **45** in presence of NADPH should be confirmed by doing the reaction in duplicates. Nevertheless, this result did not mean that in other structures with a possible different cofactor positioning, an Asp residue could not be appropriate for P3 function. This example represents the drawback to use a reference enzyme to estimate if some positions of the active site can or cannot be occupied by one particular residue. Such an Asp residue responsible of NH_3 activation has been reported for example in *meso*-diaminopimelate dehydrogenase (D92 in PDB: 3WBF) (Liu *et al.*, 2014).

II.2 P12 for active site closing

P12 position is located in the active site ceiling in the closing helical segment N172-S189 (*Cfus*AmDH numbering) mentioned in Chapter II, I.1.1 (Figure 23). The nat-AmDHs diversity among G3/G4 also displays very little variability with mainly Phe and Tyr residues except in OM-RGCAmDH1, OM-RGCAmDH2 and MATOUAmDH1, three enzymes poorly expressed in *E. coli*, where it is occupied by Ala, Leu and Asn, respectively. The approach was, first, to remove the steric hindrance of Phe/Tyr residues by modification in Ala in *Cfus*AmDH, *Msme*AmDH and MATOUAmDH2. The three mutants were produced as described for P3 mutants and detailed in Experimental section, II. The purity of *Msme*AmDH-F169A was not sufficient to be included in the specific activity assay performed on *Cfus*AmDH and MATOUAmDH2 WT and mutants towards **170**, **45** and **175** (Figure 26).

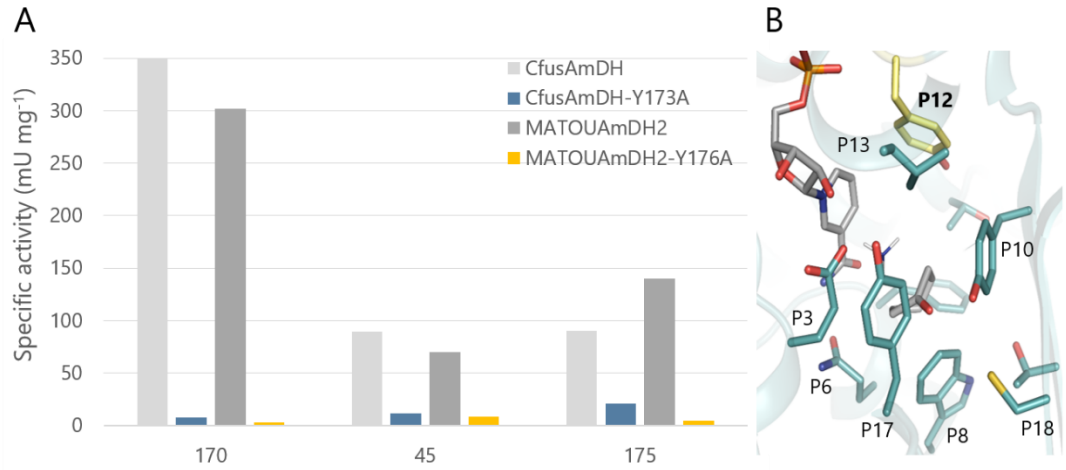


Figure 26. Influence of mutations at P12 in *Cfus*AmDH and MATOUAmDH2. (A) Specific activity (mU mg^{-1}) of *Cfus*AmDH, MATOUAmDH2 and P12 variants. The reaction conditions were as follows: final volume 100 µL, 10 mM carbonyl substrate, 1 M NH_4HCO_2 buffer pH 8.5, 0.1 mM NADPH, 0.1 mM NADH, 0.05-0.22 mg mL^{-1} purified enzyme, RT. (B) *Cfus*AmDH active site with P12 highlighted in yellow.

As expected, the reduction of the residue size at P12 from Tyr to Ala led to almost complete loss of activity. This study was completed by two additional mutations on *Cfus*AmDH, Y173L and Y173N to test the potential of residues found in OM-RGCAmDH2 and MATOUAmDH1. *Cfus*AmDH-Y173L and

*Cfus*AmDH-Y173N were also inactive, even towards substrates that are more specific to MATOUAmDH1 [methylisobutylketone (**104**), 4-oxopentanoate (**152**) or benzaldehyde (**31**)], the enzyme harboring an Asn at this position (data not show). Altogether, these observations support the hypothesis of the importance of the recurrent Phe/Tyr residue at P12 position for substrate positioning and pocket closing in structures similar to *Cfus*AmDH. The movement inducing the pocket closing/opening is more described in Chapter III, I.4.7.2.1.

II.3 P17 for correct positioning of P3

As P3 and P12, P17 is a position that does not display high variability of residues in G3 and G4 groups. The majority of the known enzymes harbors a hindered Tyr or Phe residues, parallel to P3, that were hypothesized to help the correct positioning of the catalytic residue through anion-π interactions (Lucas *et al.*, 2016) and steric hindrance (Figure 23). Some exceptions, such as OM-RGCAmDH1, OM-RGCAmDH2 and a set of uncharacterized nat-AmDHs (A0A1N7N2Q9, A0A1X4NNW3, A0A1I3IZQ4, L9XSF2, A0A1L1VXV4, A0A076NBW6) harbor Pro, Gly, Val, Leu and Ser at this position with similar 3D-arrangement than in *Cfus*AmDH or *Msme*AmDH when looking at their predicted structure from AlphaFold2 available on Uniprot. In AmDH4, the linker section between the helixes T181-S191 and G199-G212 (D186-S210 and Y202-G215 in *Cfus*AmDH, Figure 22) takes a highly different 3D-arrangement as compared to nat-AmDHs of G3 and G4, with a longer section G192-I198 located in the P17 area without bearing any big aromatic residue (Appendix 6). In the reference enzymes, *Cfus*AmDH, *Msme*AmDH and MATOUAmDH2, P17 was mutated into Ala. The purity of *Msme*AmDH-F199A was not sufficient to be included in the specific activity assays performed on *Cfus*AmDH and MATOUAmDH2 WT and P17 mutants towards **170**, **45** and **175** (Figure 27).

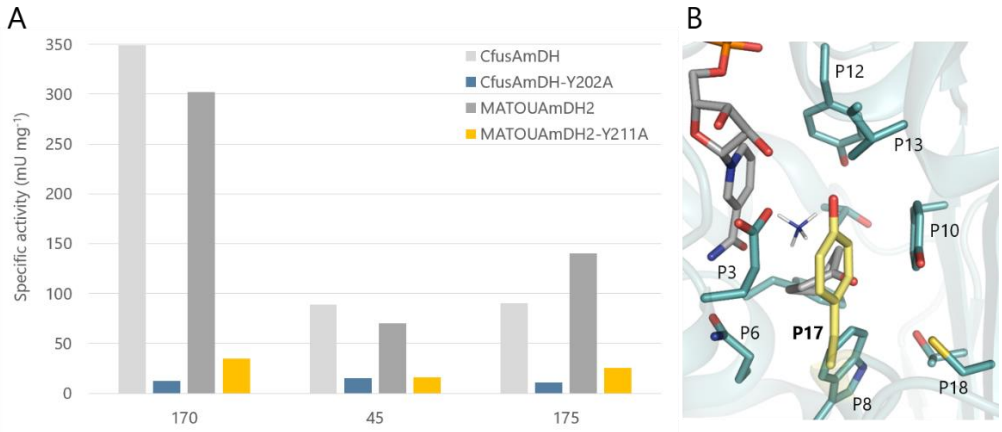


Figure 27. Influence of mutations at P17 in *Cfus*AmDH and MATOUAmDH2. (A) Specific activity (mU mg^{-1}) of *Cfus*AmDH and MATOUAmDH2 P17-variants. The reaction conditions were as follows: final volume 100 µL, 10 mM carbonyl substrate, 1 M NH_4HCO_2 buffer pH 8.5, 0.1 mM NADPH, 0.1 mM NADH, 0.05-0.22 mg mL^{-1} purified enzyme, RT. (B) *Cfus*AmDH active site with P17 highlighted in yellow.

The activity towards all the substrates tested decreased substantially compared to the WT enzymes but a residual activity could be detected especially with MATOUAmDH2-Y176A. This observation confirmed the importance of the presence of hindered residues or hindered sections, as found in AmDH4, for P3 positioning. However, even if Glu of P3 is not anchored in a good orientation by Phe or Tyr, it can still statistically adopt the optimum positioning to start the reaction but on a much lesser extent regarding the energy supplied to the system.

II.4 P10 and P18 for substrate positioning and stabilization

Understanding the substrate positioning and stabilization system in the pocket could help to gain more insights into the nat-AmDHs substrate spectrum and their (*S*)-stereoselectivity. The docking of **45** and **11** inside the active site of *Cfus*AmDH and *Msme*AmDH structures complexed with $NADP^+$ revealed a particular orientation of **45** carbonyl moiety towards the P10 position occupied by Y168 and W164 in *Cfus*AmDH and *Msme*AmDH, respectively. More precisely, the C=O moiety seems to point towards the functionalized moiety of P10 residues, the hydroxyl in Y168 with a distance between the two O-atoms of 3.2 Å and the N-atom of the indole moiety of W164 with a O-N distance of 2.9 Å. The residues at P10 were then hypothesized to take part on the substrate stabilization in the active site in a position favorable for the reaction. The diversity among nat-AmDHs is limited to bulky Trp, Tyr and Phe but some enzymes also harbor small and/or unfunctionalized residues such as Ile, Thr or Val (OM-RGCAmDH1, OM-RGCAmDH2 and L9XSF2, respectively). To a lesser degree, the P18 position could also be implied in this mechanism as a second layer key residue. This position is highly conserved among nat-AmDHs with mainly Cys and Met residues but also Ser and Thr in the nat-AmDHs set (Appendix 5), the distance between the carbonyl moiety and the hydroxyl or the thiol of the short Cys/Ser/Thr residues make it impossible to consider any stabilizing hydrogen bond. *Cfus*AmDH and *Msme*AmDH were mutated at those two positions in order to verify the role of P10 and P18 residues in the enzyme mechanism. First, the mutation Y168 (*Cfus*AmDH) and W164 (*Msme*AmDH) in Phe mimicked the loss of the stabilizing hydroxyl without removing the hindrance. On the other hand, the mutation into Thr led to the opposite feature by maintaining the hydroxyl function while removing the hindrance. The specific activity of the corresponding variants *Cfus*AmDH-Y168F/T and *Msme*AmDH-W164F/T was assayed towards a range of carbonyl substrates **170**, **45**, **175**, **112**, **181** in presence of whether NADH or NADPH cofactor (Figure 28).

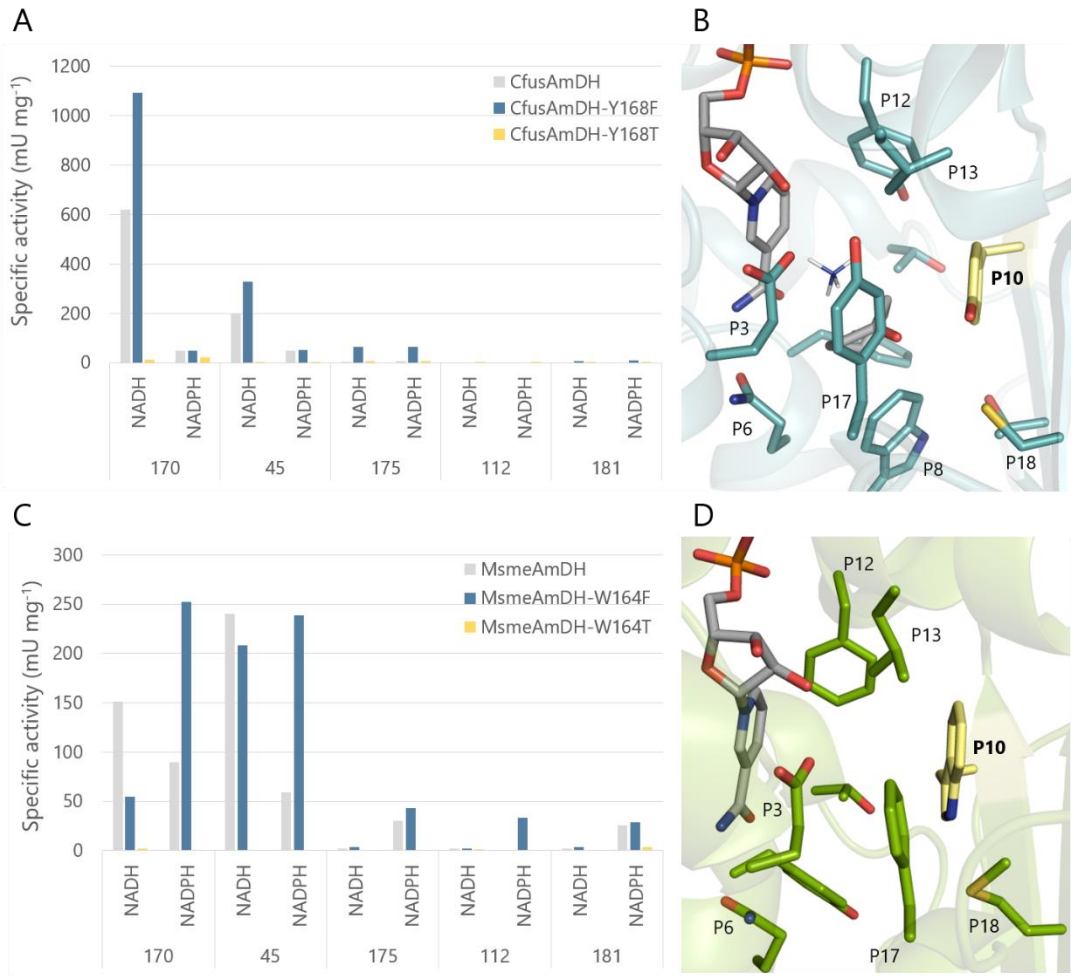


Figure 28. Influence of mutations at P10 in *Cfus*AmDH and *Msme*AmDH. (A)-(C) Specific activity (mU mg^{-1}) of *Cfus*AmDH and *Msme*AmDH P10-variants, respectively. The reaction conditions were as follows: final volume 100 µL, 10 mM carbonyl substrate, 1 M NH_4HCO_2 buffer pH 8.5-8.8, 0.2 mM NAD(P)H, 0.02-0.3 mg mL^{-1} purified enzyme, RT. (B)-(D) *Cfus*AmDH and *Msme*AmDH active sites with P10 highlighted in yellow.

Rather than leading to a drop in activity with the mutation P10F, the mutants *Cfus*AmDH-Y168F and *Msme*AmDH-W164F displayed similar or higher activity than the WT enzymes depending on the cofactor used. With *Cfus*AmDH-Y168F, the activity was estimated 1.7-to-1.8-fold higher towards **170** and **45**, respectively, only in presence of NADH, while being rather similar to the WT with NADPH. On the other hand, *Msme*AmDH-W164F displayed 2.8-to-4.1-fold higher activity towards **170** and **45**, respectively, in presence of NADPH, but 1.2-to-2.7-fold lower activity towards the same substrates in presence of NADH. The cofactor-dependent influence of the mutations P9F is hard to explain as the cofactor is located on the opposite side of the mutated residue in the active site. MD simulations could help to attest a change in cofactor and substrate positioning in the pocket in presence or absence of the mutation. Regardless of this aspect, the present results does not confirm the necessity of a functionalized residue at P10 position to stabilize the substrate. The steric hindrance of the residue seems to be of higher importance as the mutants P10T, *Cfus*AmDH-Y168T and *Msme*AmDH-W164T, were inactive. We

further investigated this hypothesis by mutating Y168 of *Cfus*AmDH in amino acids of decreasing steric hindrance (His, Ile, Leu) completed with the double mutation Y168F/C269A in *Cfus*AmDH and W164F/M266A in *Msme*AmDH to attest the role of P18 residues (Figure 29). All the mutants were then assayed for specific activity towards **170**, **45** and **175** in a presence of a mix of NADH/NADPH to avoid the cofactor dependency.

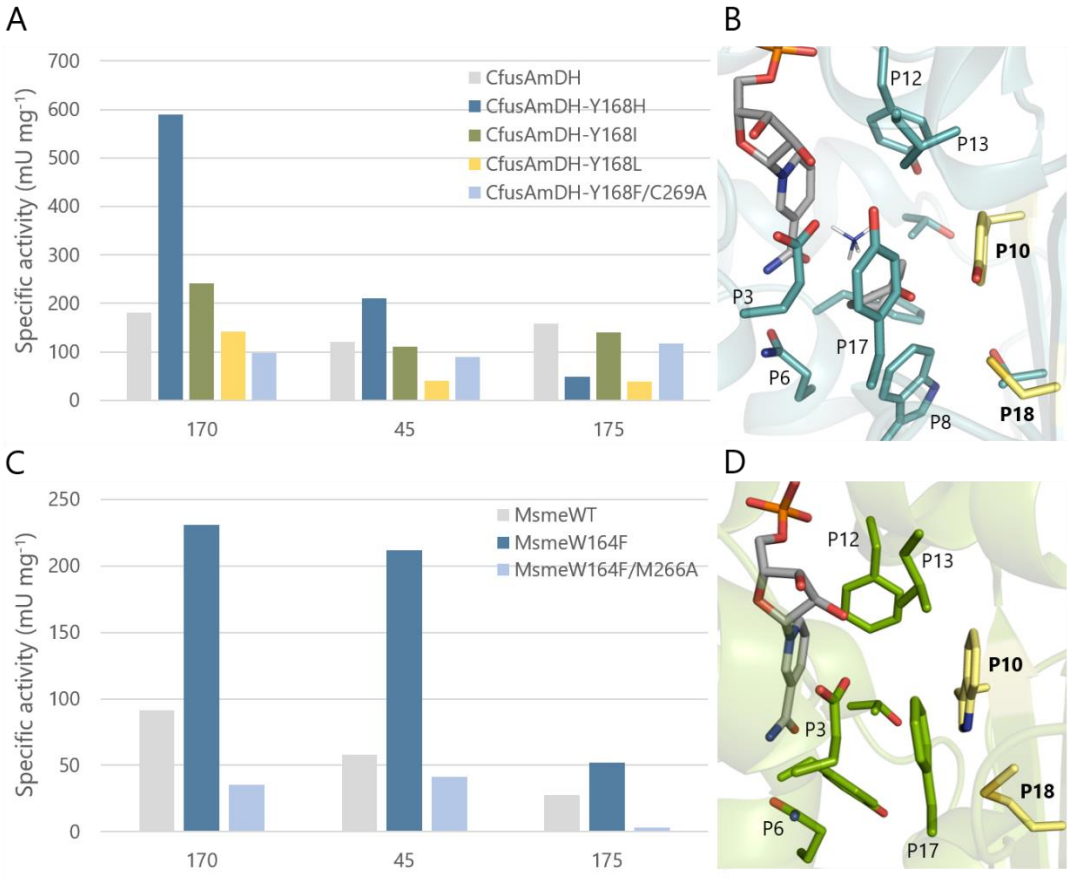


Figure 29. Influence of additional mutations at P10 in *Cfus*AmDH and double mutations at P10 and P18 in *Cfus*AmDH and *Msme*AmDH. (A)-(C) Specific activity (mU mg^{-1}) of *Cfus*AmDH and *Msme*AmDH P10- and P18-variants, respectively. The reaction conditions were as follows: final volume 100 µL, 10 mM carbonyl substrate, 1 M NH_4HCO_2 buffer pH 8.5-8.8, 0.2 mM NAD(P)H, 0.02-0.3 mg mL^{-1} purified enzyme, RT. (B)-(D) P10 and P18 highlighting (yellow) in *Cfus*AmDH and *Msme*AmDH, respectively. *Cfus*AmDH and *Msme*AmDH active sites with P10 and P18 highlighted in yellow.

*Cfus*AmDH-Y168F obtained from a previous batch of mutants production was inactive even after storage at -80°C and was therefore not included in this second study. As observed with *Cfus*AmDH-Y168F, *Cfus*AmDH-Y168H were found more active than the WT (3.3 and 1.7-fold higher towards **170** and **45**, respectively). Interestingly, the decreasing steric hindrance of the residues at P10 in *Cfus*AmDH and the loss of aromatic side chain via Ile and Leu mutations retained a substantial activity towards **170** and **45**. As already observed with *Cfus*AmDH-Y168F, *Cfus*AmDH-Y168H were also found more active than

the WT (3.3 and 1.7-fold higher towards **170** and **45**, respectively). The double mutation P10F/P18A also led to a decreasing activity as compared to the single mutant P10F, in the case of *Msme*AmDH, but also to the WT. This observation seems to be in favor of the hypothesis about the role of P18 residues in the active site second layer, maybe for P10 positioning, but this mutation can also cause a left space in this area or even general destabilization of the enzyme folding. Due to a lack of time, the biocatalytic potential of *Cfus*AmDH-Y168H and -Y168F could not be more deeply studied but it would be interesting to do so later, in view of the notable increased activities of these mutants.

II.5 P6 for cofactor anchoring

MD simulations detailed in Chapter III, I.4.7.2.3 revealed strong interactions between Q141 of *Cfus*AmDH (P6) and the amide moiety of the cofactor. Hydrogen bonds between Q141 carboxylate and N-atom of the amide function of the cofactor might help the nicotinamide moiety anchoring inside the active site. This complements the already strong positioning of the rest of the molecule via the Rossmann fold domain. The set of nat-AmDHs shows a rather high diversity of amino acids at this position with mainly Gln and Glu but also Pro, Asn, Tyr and even residues without potential of hydrogen bond formation such as Phe, Ile and Leu. This hypothesis was not supported by any mutational work.

II.6 Conclusion on the essential positions of the nat-AmDHs active site

This study revealed the great importance of the presence of specific residues or specific features at some positions in the first layer of the active site. It confirmed the essential role of P3 in the mechanism but nuanced that of P10. It could be completed with kinetic parameters of all the mutants to have more insights into the effect of the mutations on the mechanism or on the affinity or catalytic efficiency. But it was not the main goal here so these time-consuming experiments were not carried out. Also, only the residues of the first layer were more deeply studied but one can note that some residues belonging to the second layer might also play a key role in the reaction but this is much more difficult to anticipate by just looking at static structures. Eventually, it has to be noted that the mutation can negatively affect one enzyme as the reference enzymes used here but can be beneficial in other enzymes harboring a similar 3D folding but with differences in the amino acids surrounding the mutational site.

III. Stability of nat-AmDHs

In a collaboration with the team of Prof. Andreas Bommarius from Georgia Tech Institute (Atlanta, USA), we investigated the stability of a set of nat-AmDHs. The aim of their study was to compare stability and oligomeric state of nat-AmDHs and other AmDHs derived from AADHs they published ten years ago (L-AmDH, F-AmDH and cFL1-AmDH described in Chapter I, II.2.4.1 and II.2.4.2). This work

includes Differential Scanning Fluorimetry (DSF), gel permeation and Size Exclusion Chromatography (OmniSEC), Analytical UltraCentrifugation (AUC), to access the sediment coefficient, and Microfluidic Modulation Spectroscopy (MMS). Adam Caparco and Bettina S. Bommarius, from this team, performed all these characterizations and we were mainly involved in the common stability assays of nat-AmDHs by monitoring their specific activity over time in different storage conditions. This key aspect of nat-AmDHs stability is also of great importance to help the selection for biocatalytic applications.

III.1 Stability of nat-AmDHs monitored though specific activity assays

The set of nat-AmDHs studied comprises *Cfus*AmDH, *Msme*AmDH, MATOUAmDH2, IGCAmDH1, IGCAmDH5, *Sgor*AmDH, *Acol*AmDH and *Chat*AmDH, characterized in previous publications (Caparco *et al.*, 2020; Mayol *et al.*, 2019) and known to be well overexpressed in *E. coli*. The storage parameters included the type of buffer (potassium phosphate or ammonium formate, pH) to understand the influence of ammonia (2 M) in the enzymes stability, the enzyme concentration (0.5 or 1 mg mL^{-1}) and the temperature (4°C, 25°C and 37°C). Separate batches of each enzymes were stored for 6 days in these conditions and their specific activity towards **170** was regularly monitored by spectrophotometry assay (see Experimental section, III.2). The results are reported in Figure 30. First, it can be noted that there is no correlation between the enzyme concentration and stability. However the concentration tested are relatively low and the observation could be different at higher concentrations. Only IGCAmDH1 and *Msme*AmDH seems to be sensitive to the presence of **11** as their remaining activity over time depends on the storage buffer with a preference for the ammonia-free one over AF at the same pH and temperature. For example, *Msme*AmDH activity remained at 82.4% of the T_0 activity after 6 days in potassium phosphate buffer while it dropped to 44.1% when stored in ammonium formate buffer, for the same enzyme concentration (1 mg mL^{-1}) and temperature (4°C). Expectedly, the key parameter for enzyme stability is the temperature. In all cases, in potassium phosphate buffer, all the tested enzymes showed remaining activity > 80% after 6 days at 4°C. Interestingly, at the higher temperatures, 25°C or 37°C, we observed three different profiles. One group of enzymes exhibited good stability at either 25 or 37°C such as IGCAmDH1 (>70% after 6days), IGCAmDH5 (> 80%) and *Acol*AmDH with no loss of activity observed after 6days. Some others showed high stability at 25°C but progressive loss of activity at 37°C. It is the case for *Msme*AmDH [14.3% after 6 days with half life ($t_{1/2}$) around 2.5 days] and *Chat*AmDH (34.4% after 6 days with $t_{1/2}$ around 2.5 days). Finally, *Cfus*AmDH, *Sgor*AmDH and MATOUAmDH2 were more sensitive to the temperature and displayed correct to good stability at 25°C (65.8 - 84.9% relative stability after 6days) but complete or almost complete loss of activity after only one day at 37°C. The particular instability of MATOUAmDH2 and *Sgor*AmDH at higher temperatures is an issue as these enzymes are more active at 40 - 50°C (Caparco *et al.*, 2020). At this time, we did not

have the results obtained with the AmDHs engineered from AADHs to establish a first comparison between these two types of structures. Regarding what is already published, one can notice that *Gk*AmDH-M0 exhibited almost complete loss of activity after 4 days at 40°C ($t_{1/2}$ around 1.5 days) while the complementary mutants *Gk*AmDH-M3 and *Gk*AmDH-M8 displayed $t_{1/2}$ of around 18 and 8 days, respectively (Wang *et al.*, 2021). Also, *Asp*RedAm was shown to precipitate after only 1 h at 40°C (Aleku *et al.*, 2017).

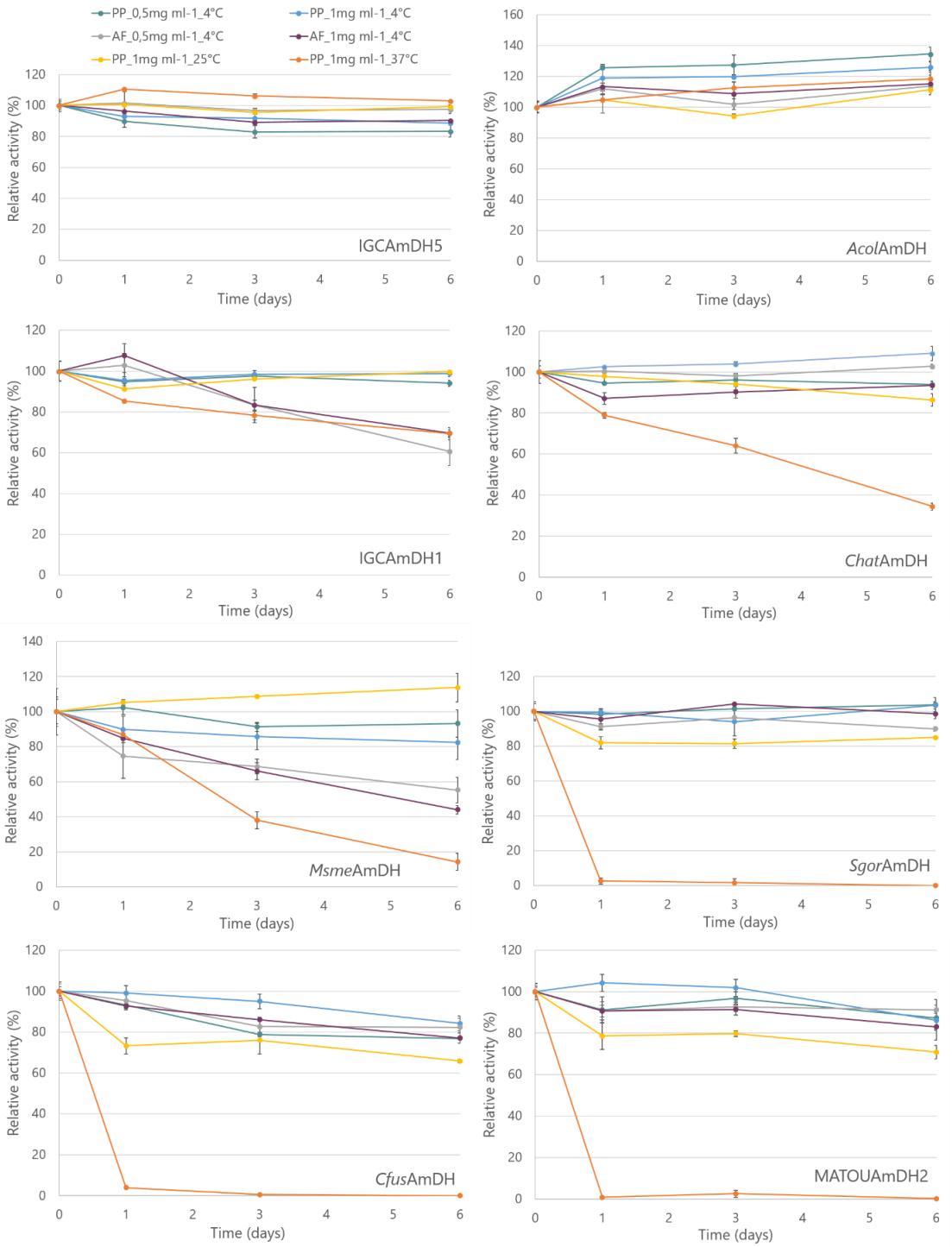


Figure 30. Relative remaining activity of IGCAmDH5, *Acol*AmDH, IGCAmDH1, *Chat*AmDH, *Msme*AmDH, *Sgor*AmDH, *Cfus*AmDH and MATOUAmDH2 towards **170** after different storage conditions. 100% alludes to the enzyme specific activity measured the day of the purification (IGCAmDH5: 619.0-661.2 mU mg^{-1}; *Acol*AmDH: 219.0-274.3 mU mg^{-1}; IGCAmDH1 349.7-525.5 mU mg^{-1}; *Chat*AmDH: 801.5-858.0 mU mg^{-1}; *Msme*AmDH: 59.0-147.2 mU mg^{-1}; *Sgor*AmDH: 790.0-986.7 mU mg^{-1}; *Cfus*AmDH: 3.6-3.8 U mg^{-1}; MATOUAmDH2: 3.4-4.0 U mg^{-1}). Reaction conditions were as follows: final volume 100 µL, 10 mM **170**, 2 M NH_4HCO_2 buffer pH 8.5, 0.2 mM NADH (NADPH for MATOUAmDH2), 0.001-0.12 mg mL^{-1} purified enzyme, 40°C. Errors represent standard deviations of two independent experiments with two different enzymes aliquots stored in the same conditions. PP = potassium phosphate buffer; AF = ammonium formate buffer.

III.2 Correlation between experimental results and amine dehydrogenases structural features

These observations and the results of the biophysical characterizations, especially enzymes melting temperature in presence of different concentration of **11** done at Georgia Tech (not shown), were correlated to *in silico* physico-chemical parameters. This rationalization approach, mainly done by Dr. Carine Vergne-Vaxelaire, is not detailed here. The most relevant structural parameters to explain the enzyme stability were the internal and external hydrophobicity, the number of salt bridges, the number of hydrogen bonds (Wang *et al.*, 2018) and the occurrence of some key residues such as Arg, Gly, Pro or Cys that can impact enzyme flexibility. The putty representation in PyMol enables the visualization of the B-factor along the enzyme structure from apparent rigidity with low B-factor represented in blue and thin segments to apparent flexibility with high B-factor represented in red and large segments. This representation cannot be used to compare the overall flexibility of one enzyme to another but to visualize the most flexible segments within one structure. Interestingly, *Msme*AmDH, *Cfus*AmDH and MATOUAmDH2 crystallographic structures display at least one highly flexible section compared to the rest of structure that could induce their poor stability at higher temperature. However, this prediction was not visible in the corresponding AlphaFold2 models thus preventing any comparison with the other enzymes IGCAmDH5, *Acol*AmDH, IGCAmDH1 and *Sgor*AmDH for which no crystallographic structures are available (Figure 31).

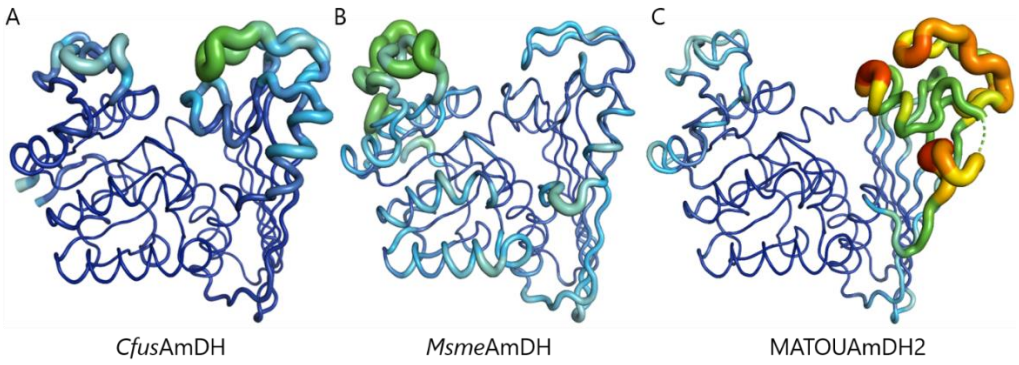


Figure 31. Putty representation of (A) *Cfus*AmDH (PDB: 6IAU, chain D), (B) *Msme*AmDH (PDB: 6IAQ, chain B) and (C) MATOUAmDH2 (PDB: 7R09, chain D).

The known nat-AmDHs are more likely to be used as templates for further exploration or mutagenesis work to improve their biocatalytic synthetic potential, including their stability, than for direct usage in synthesis. However, this study was important to gain more insights into their intrinsic stability in given conditions even if deeper analyses could be carried out with experts in this field. Furthermore, the correlation of structural features with experimental stability is still a research area at an early stage, so the accumulation of such data is required to progress in this field. All things considered,

these observations could be of great help to discriminate one sequence to another in future selection of nat-AmDHs candidates. This sub-project was useful for the whole understanding of this enzyme family and has been included in a draft of publication written by Dr. Adam Caparco.

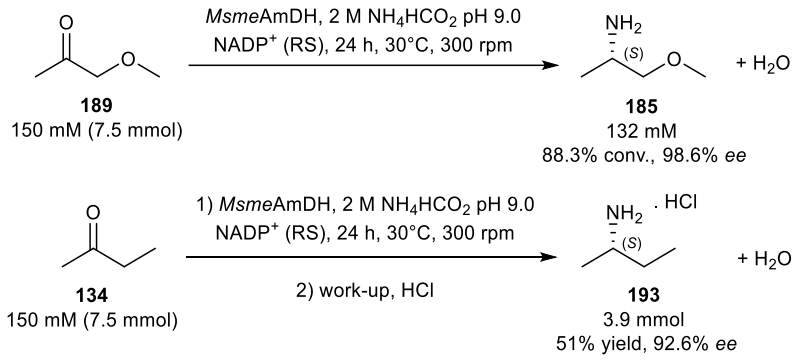
IV. Further characterization of the native amine dehydrogenases carbonyl substrate spectrum

IV.1 Small prochiral (hydroxy)ketones

Many research aim at making enzymes suitable for the transformation of hindered molecules to demonstrate their wide applicability. Nevertheless, small targets have also their importance as precursors of pharmaceuticals and agrochemicals, such as (2*R*)-butan-2-amine (**182**) in the herbicide Bromacil (**183**) and in the drug candidate XL888 (**184**) or (2*S*)-1-methoxypropan-2-amine (**185**) in the herbicide Outlook® (**186**) (BASF). The ChiPros™ process of BASF, mentioned in Chapter I, II.2, is highly efficient for the kinetic resolution of alkyl amines such as (*2S*)-nonan-2-amine (**187**) with E values >1000 but much less in the case of small alkyl amines with $E_{butan\text{-}2\text{-}amine}$ = 8, for example. The works published in 2019 and 2020 reported their substrate spectrum but was obviously not exhaustive (Mayol *et al*., 2019; Caparco *et al*. 2020). As already mentioned, their top tested substrates were mainly **45** and **170**, but some of them displayed activity towards small alkyl linear ketones such as **112** (5.3 – 51.2 mU mg^{-1}) or hydroxy derivatives such as 3-hydroxybutan-2-one (**172**) (43.1 – 324.4 mU mg^{-1}) with high (*S*)-enantioselectivity. These preliminary results prompted us to consider that high enantioselectivity could be achieved for other small chiral amines such as the building blocks listed above. Hence, we decided to complete the already reported nat-AmDHs activity towards short prochiral ketones (Mayol *et al*., 2019) and study in more details the biocatalytic potential of some of these enzymes (*Cfus*AmDH, *Msme*AmDH, *Micro*AmDH, and MATOUAmDH2) for the synthesis of small chiral alkyl amines and small amino alcohols. These results were published in 2021 in Frontiers in Catalysis (Ducrot *et al.*, 2021). The substrate screening and deeper characterizations (substrate loading and preparative scale reaction) were carried out by Dr. Carine Vergne-Vaxelaire and are therefore only summarized in this part. My main contribution to this study was on the *in silico* discussion to provide insights into the stereoselectivity of the enzymes considered. The contents of the parts written below are largely inspired by the results and discussion written in the publication.

IV.1.1 Screening and hits characterization

The four enzymes were tested towards butan-2-one (**134**), pentan-2-one (**112**), 3-methylbutan-2-one (**188**), 1-methoxypropan-2-one (**189**), 1-hydroxypropan-2-one (**180**), 1-hydroxybutan-3-one (**190**), 2-hydroxybutan-3-one (**191**) and 1-hydroxybutan-2-one (**118**) at 10 and 50 mM by monitoring the conversions and *ee* of the corresponding amines after a 24 h [30°C, 400 rotation per minute (rpm)] reaction with the cofactor recycling system. The screening results are given in Appendix 7 for later *in silico* discussion. Interestingly, with all enzymes, higher conversions could be achieved for the formation of the functionalized small amines, such as (*2S*)-1-methoxypropan-2-amine (**185**) [(*S*)-MOIPA, Outlook®, produced at 5,000 T year^{-1} by ChiPros™] and (*3S*)-2-hydroxybutan-3-amine (**192**), rather than the plain (*2S*)-butan-2-amine (**193**) and (*2S*)-pentan-2-amine (**173**). *Msme*AmDH was selected for further characterization and notably, substrate-loading tolerance assay. It could convert up to 150 mM of **189** and **134** with 83.5% and 58.3% analytical conversion respectively after 48 h at 30°C without suffering major decrease in conversion (80.7% and 73.8% with 10 mM substrate loading, respectively). Nevertheless, any inhibition study was carried out. These conditions were used to run 50 mL-scale reactions to access 6.6 mmol (98.6% *ee*) of **185** (88.3% analytical yield) and 3.9 mmol (92.6% *ee*) of **193** (51% isolated yield) (Scheme 29). The higher conversions obtained, compared to reactions at 100 µL-scale could be due to different parameters such as a lower rotation speed of the 50 mL-scale reaction that could be beneficial for the enzyme stability. This point was not studied in detail.



Scheme 29. Semi-preparative scale synthesis of (2*S*)-1-methoxypropan-2-amine (**185**) and (2*S*)-butan-2-amine (**193**) catalyzed by *Msme*AmDH. Adapted from Ducrot *et al.* 2021.

This work completes the biocatalytic toolbox for the synthesis of such short amines as the other RedAms and AmDHs engineered from AADHs were mainly studied towards increasingly large substrates and featuring the opposite (*R*)-enantioselectivity. Among other examples, the tested nat-AmDHs are complementary to the engineered AmDHs described by Sun and co-workers [Ming *et al.*, 2022; Tong *et al.*, 2021; H. Wang *et al.*, 2020 (B)]. Starting from **190**, *Cfus*AmDH could reach 65.7% conversion from 10 mM substrate and 99.1% *ee* for the formation of (*3S*)-1-hydroxybutan-3-amine (**194**) while *Gs*AmDH-

wh18 (K68S/D261L) could afford 45% conversion from 30 mM substrate and > 99% *ee* for the (*3R*)-1-hydroxybutan-3-amine (**195**) [Ming *et al.*, 2022; H. Wang *et al.*, 2020 (B)]. *Gs*AmDH-wh18 was further engineered through several round of mutagenesis to reach 99% analytical conversion and 99% *ee* with this same substrate (Ming *et al.*, 2022).

The nat-AmDHs tested, and especially *Msme*AmDH, are also complementary to TAs such as the (*S*)-selective HewT and the (*R*)-selective *RTA-X43 used by Heckmann *et al.* for continuous flow synthesis of (2*S*)-butan-2-amine (**193**) and (2*R*)-butan-2-amine (**182**), respectively (Heckmann *et al.*, 2021). When used as free enzymes, HewT and *Msme*AmDH reached similar conversions but *Msme*AmDH could afford clearly higher *ee* with 91% at 150 mM compared to 45% at 100 mM with HewT.

IV.1.2 Correlation with *in silico* analysis

Molecular docking of some tested amines in *Cfus*AmDH and *Msme*AmDH crystallographic in complex with their cofactor $NADP^+$ and in *Micro*AmDH and MATOUAmDH2 homology models were carried out using Autodock software (Morris *et al.*, 2009). *Micro*AmDH homology model was generated using the closed form of *Msme*AmDH as a template. At this time, as the closed structure of MATOUAmDH2 was still not available, we used a model formerly made based on *Cfus*AmDH closed structure. Among the fifty docking conformations obtained from the docking simulations (see Experimental section, VI.2.2.2), the ratio of "correct" conformations harboring (*S*)- or (*R*)-stereochemistry were analyzed (Figure 32). As the final step of the reductive amination mechanism is the transfer of the hydride from C4N of the NAD(P)H nicotinamide ring to the C-atom of the iminium intermediate to give the oxidized $NAD(P)^+$ and the amine product, only conformations having the correct orientation of the provided hydrogen towards the cofactor were considered as "correct". For the butan-2-amine ligands **182** and **193**, Figure 32 shows a majority of (*S*)-conformations **193** generated with the correct orientation in *Cfus*AmDH, *Msme*AmDH and MATOUAmDH2, respectively 16, 30 and 35 out of 50 poses while the (*R*)-conformations **182** were present in only 2, 1 and 2 correct poses out of 50. On the other hand, for *Micro*AmDH, the difference is much lower with 15 out of 50 poses of (*S*)-conformations **193** and 12 out of 50 poses of (*R*)-conformations **182**. These *in silico* observations correlate with the experimental data showing good *ee* values for the first three enzymes (88.9-98.1%) while *Micro*AmDH reached only 69.1% *ee* with this substrate. The different (*R*)/(*S*) ratios are difficult to explain as the side chain orientation of the docked amines in their (*R*)- and (*S*)- conformations, are almost identical with no particular interaction or steric hindrance that could discriminate one or the other (Figure 33). No clear correlation between residues surrounding the substrate, or residues of the second layer could be established with the docking observations. Possibly, the features of the docking algorithm, such as the assumption of rigid protein and ligand, can obscure the reasons for these differences in enantioselectivity.

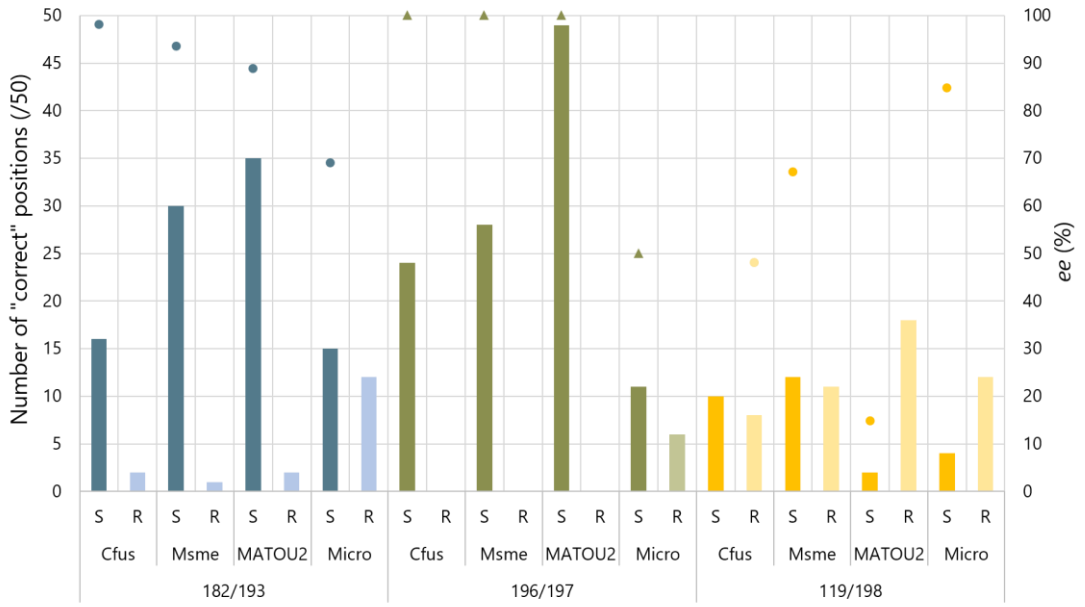


Figure 32. Docking simulation of (2*S*)- and (2*R*)- short amines. The docking was performed separately on (2*S*)-butan-2-amine (**193**), (2*R*)-butan-2-amine (**182**), (2*S*)-1-hydroxypropan-2-amine (**196**), (2*R*)-1-hydroxypropan-2-amine (**197**), (2*S*)-1-hydroxybutan-2-amine (**119**) and (2*R*)-1-hydroxybutan-2-amine (**198**) in *Cfus*AmDH (Cfus), *Msme*AmDH (Msme), MATOUAmDH2 (MATOU2) and *Micro*AmDH (Micro). Bars represent the number of "correct" positions obtained out of 50 generated conformations while the scatter plot represents the *ee* obtained from the corresponding reaction (circles: calculated experimental data, triangles: estimated experimental data). *ee* values are indicated for the enantiomer formed in majority. For **196** and **197** ligand, *ee* values could not be determined but **197** was only found in traces with *Cfus*AmDH, *Msme*AmDH and MATOUAmDH2 (*ee* values arbitrarily indicated at 100%) but was obtained in larger proportion with *Micro*AmDH (*ee* arbitrarily indicated at 50%). Adapted from Ducrot *et al*. 2021.

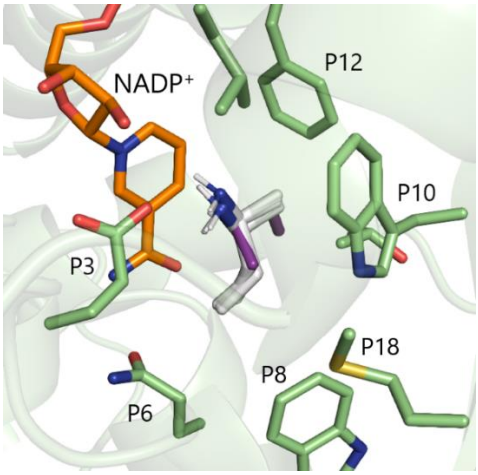


Figure 33. *Msme*AmDH RX structure (PDB: 6IAQ) in complex with NADP$^+$ cofactor with docked structures of **182** and **193**. NADP$^+$ is highlighted in orange, **193**, in white and **182**, in purple. All the **193** docked structures belong to the twenty conformations displaying the lowest binding energy. Adapted from Ducrot *et al*. 2021.

The *In silico* analysis was also in accordance with *in vitro* results for the 1-hydroxypropan-2-amine ligands **196** and **197** as only the (*S*)-enantiomer **196** was generated in the correct pose from the docking and experimentally produced with *Cfus*AmDH, *Msme*AmDH and MATOUAmDH2. On the contrary, the (*R*)-enantiomer **197** was formed in reaction catalyzed by *Micro*AmDH, without being quantified, and represented 35% of the correct conformations generated from the docking (11/50 for (*S*)-conformation **196** against 6/50 for (*R*)-conformation **197**). This different enantioselectivity and (*R*)-/(*S*)- ratios obtained with *Micro*AmDH with 1-hydroxypropan-2-amine **196** and **197** can be due to a coordination by hydrogen bonding of the hydroxyl groups of the ligand and the residue Y169 (P12), which can favor the positioning of the ketone for the formation of the corresponding (*R*)-amine **197** (Figure 34). *Cfus*AmDH and MATOUAmDH2 also harbor a Tyr residue at P12, Y173 and Y176, respectively, but with a slightly different position that moves the hydroxyl group further away from the ligand (Δ = 0.4 Å between Y169 of *Micro*AmDH and Y173 of *Cfus*AmDH, and Δ = 0.8 Å between Y169 of *Micro*AmDH and Y176 of MATOUAmDH2).

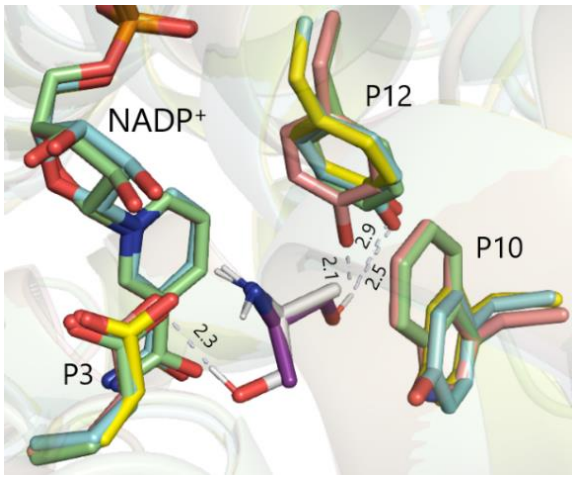


Figure 34. Docking of (2*S*)-1-hydroxypropan-2-amine (**196**) and (2*R*)-1-hydroxypropan-2-amine (**197**) in *Micro*AmDH and alignment with *Cfus*AmDH, *Msme*AmDH and MATOUAmDH2. Structural alignment of *Cfus*AmDH RX structure (PDB: 6IAU; blue), *Micro*AmDH homology model (PDB template: 6IAQ; pink) and MATOUAmDH2 homology model (PDB template: 6IAU; yellow) to *Msme*AmDH RX structure (PDB: 6IAQ; green). The corresponding $NADP^+$ cofactors are shown in the same color as the enzyme to which it belongs. Docked structures of **196** and **197** in *Micro*AmDH are respectively shown in white and purple. The distance between the H-atom of the hydroxyl group of **196** and the O-atom of the carboxylate group of E104 (P3) in *Micro*AmDH, and between the H-atom of the hydroxyl group of **197** and the O-atom of the hydroxyl group of Y169, Y173 and Y176 (P11), respectively of *Micro*AmDH, *Cfus*AmDH and MATOUAmDH2, are given in Angstroms (Å). RMSD (*Cfus*AmDH-*Msme*AmDH) = 0.987 Å; RMSD (*Micro*AmDH-*Msme*AmDH) = 0.254 Å; RMSD (MATOUAmDH2-*Msme*AmDH) = 0.984 Å. Adapted from Ducrot *et al.* 2021.

In addition, the more hindered W164 (P10) in *Micro*AmDH compared to Y168 and Y171 in *Cfus*AmDH and MATOUAmDH2 can also impart less flexibility to the alkylhydroxyl chain. In the docking poses, the latter is more oriented towards Y169 in *Micro*AmDH and its hydroxyl group, increasing the

access to (*R*)-configurations **197**. In *Msme*AmDH, P12 is occupied by a Phe that cannot provide a hydrogen bond with the substrate to favor the formation of the (*R*)-amine **197**. In that case, only the (*S*)-enantiomer **196** can be docked with the hydroxyl group oriented towards the carboxylate function of Glu at P3, with which a hydrogen bond can occur (Figure 35). However, we must nuance these observations as the homology model generation of *Micro*AmDH was based on *Msme*AmDH, the template sharing the highest sequence identity, but that harbors a Phe at P12 instead of a Tyr that may be behind the different side chain positioning. Mutations at these positions could give more weight to these hypotheses but were not performed due to a lack of time.

Eventually, for 1-hydoxybutan-2-amine **119** and **198** ligands, overall, the *in silico* analysis revealed much more (*R*)-poses **198** with all the enzymes compared to the previous ligands, which correlates with the low *ee* values experimentally obtained for this amine. For example, in *Msme*AmDH, approximately the same number of (*S*)- **119** and (*R*)- **198** correct poses were obtained, and the *ee* value for **119** was experimentally 67%. Nevertheless, the higher *ee*s obtained for *Micro*AmDH and *Msme*AmDH compared to MATOUAmDH2 and *Cfus*AmDH are not directly related to the respective ratios of correct **198** and **119** conformations in this modeling study. For instance, in MATOUAmDH2, the ratio of (*R*)/(*S*) poses was higher than with *Cfus*AmDH, whereas *Cfus*AmDH facilitated the formation of the (*R*)-enantiomer **198** with 48% *ee*, and MATOUAmDH2 the formation of (*S*)-enantiomer **119** with 15% *ee*. For *Cfus*AmDH and *Msme*AmDH, this ratio was similar (respectively 8/10 and 11/12), whereas *Msme*AmDH displayed a clear preference for the formation of (*S*)- **119** (*ee* = 67%). Structurally, many (*R*)-poses seemed stabilized with the hydroxyl group of Y173 in *Cfus*AmDH (Figure 35A), thus explaining the greater rate of (*R*)-amines **198** formation compared to other substrates, even with low *ee* value ($ee_{(R)}$ = 48.1%). The more hindered ethyl moiety of **197**, compared to **198**, might "push" the rotation of the substrate and make the hydroxyl group closer to the Tyr even in *Cfus*AmDH (2 Å) and MATOUAmDH2 (2.9 Å) while still keeping the chiral center close to the $NADP^+$ hydride carbon. It is worth noting that for MATOUAmDH2 and *Msme*AmDH, the different conformations displayed much more variability of positions. The presence of other residues (P5, P8, P9, P16, P17) surrounding these key positions may enable other interactions and positioning of the key residues potentially less visible in these models (Figure 35B).

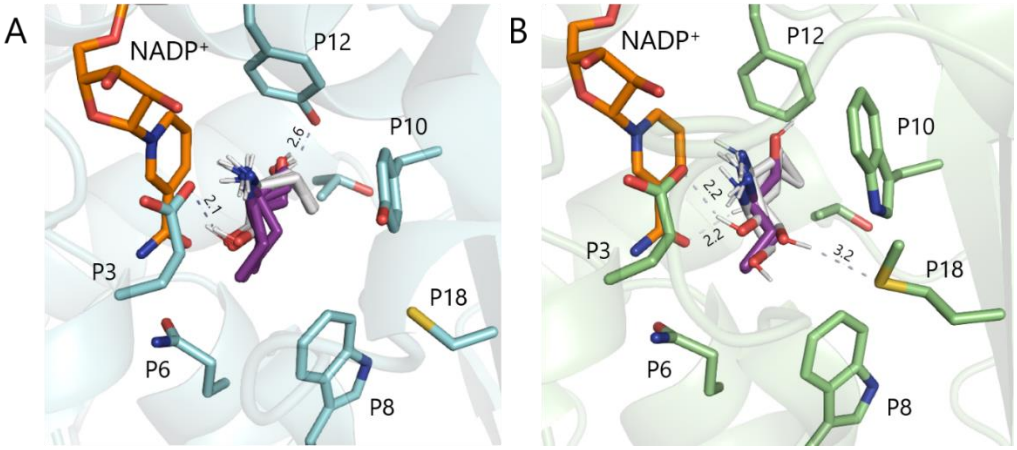


Figure 35. Docking of (2*S*)-1-hydroxybutan-2-amine (**119**) and (2*R*)-1-hydroxybutan-2-amine (**198**) in *Cfus*AmDH and *Msme*AmDH. (A) *Cfus*AmDH RX structure (PDB: 6IAU) and $NADP^+$ cofactor, in orange, with docked structures of **119**, in white, and **198**, in purple; (B) *Msme*AmDH RX structure (PDB: 6IAQ) and $NADP^+$ cofactor, in orange, with docked structures of **119**, in white, and **198**, in purple. Distances are given in Angstroms (Å).

The structurally-based preliminary explanation of the slight differences in (*S*)-enantioselectivities observed experimentally for **134**, **112**, **188**, **189**, **180**, **190** and **191**, together with the important variability for **118**, helps to better understand the enzyme-substrate interactions within the catalytic pockets of these four enzymes. It also highlitghted the role of several residues in the good accommodation of hydroxyketones, that might result in such good conversions. It would be interesting to perform some deeper computational studies, including MD simulations, as done by Tseliou *et al.*, to explain the substrate-dependent stereo-switchable selectivity in the case of the mutant LE-AmDH-v1 [Tseliou *et al.*, 2019(A)]. *In silico* studies with the iminium intermediates or carbinolamines, instead of the amine products, may also provide additional explanations.

The study of the nat-AmDHs substrate scope for the short prochiral ketones revealed a higher variability of stereoselectivity that was limited to (*S*)-amine formation in the previous reports. These observations, correlated to structural features, could help the search of nat-AmDHs providing (*R*)-amines with good *ee*.

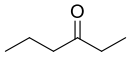
IV.2 Other short aliphatic and aromatic carbonyl compounds

IV.2.1.1 Monitoring of the amine formation

As for the small aliphatic/hydroxy linear ketones, the substrate screening, published before my work in 2019 and 2020, was lacking some key carbonyl substrates. We decided to test other aromatic carbonyls and linear ketones and focused on the following: furfural (**57**), cyclopentancarbaldehyde (**199**), benzaldehyde (**31**), acetophenone (**39**) and hexan-3-one (**200**). These substrates were tested during the screening assay carried out with the nat-AmDHs mutants and described in Chapter III, I.4.3, but in this part, the results focus on the nine corresponding WT nat-AmDHs (*Cfus*AmDH, *Msme*AmDH, *Micro*AmDH,

*Apau*AmDH, MATOUAmDH2, *Chat*AmDH, IGCAmDH1, *Rgna*AmDH, *Porti*AmDH). The selected nat-AmDHs were only scarcely tested, or even not tested on these substrates in previous studies. It was therefore beneficial to test them to complete the nat-AmDHs substrate spectrum. After incubation at 25°C during 24 h (Experimental section, IV.1), the amine formed from the reductive amination with cofactor recycling system were extracted in ethyl acetate (**201**) (EtOAc) and monitored using Gas Chromatography equipped with a Flame Ionization Detector (GC-FID) (Table 4). In case of the prochiral ketones **39** and **200**, the *ee*s were estimated using Ultra-High Performance Liquid Chromatography equipped with Ultraviolet detection (UHPLC-UV) analysis after derivatization with 1-fluoro-2-4-dinitrophenyl-5-L-alanine (FDAA, **227**). The analytical development carried out to perform this screening is also more described in Chapter III, I.4.2 and Experimental section, IV.3.1 but the corresponding GC-FID and UHPLC-UV chromatograms are given in Appendix 9 to Appendix 15 according to Appendix 8. The results confirmed a good activity, shared by a large panel of the enzymes tested, towards short aldehyde bearing (aromatic) cyclic substituents such as **57** (41.1 - 56.0% conversion), **199** (37.2 - 83.5%) and **31** on a lesser extent (9.7 - 79.0%). With the addition of a methyl group in **39**, the conversions dropped to a maximum of 13.9% obtained with *Micro*AmDH for the formation of (1*S*)-1-phenylethan-1-amine (**14**) with 99.1% *ee*. This might be due to a too constrained planar structure of the aromatic cycle that cannot make the additional methyl to fit correctly in the small pocket of these nat-AmDHs. Interestingly, only a restricted group of nat-AmDHs could convert **200**: *Msme*AmDH (39.1%), *Micro*AmDH (51.2%), both only displaying 6.8 and 28.1 mU mg^{-1} specific activity in Mayol *et al.*, 2019, *Porti*AmDH (46.5%) and on a lesser extent, MATOUAmDH2 (12.4%) and *Cfus*AmDH (5.8%). Due to a higher symmetry of the molecule, the *ee* obtained are much more variable. When *Micro*AmDH and *Msme*AmDH provided the hypothesized (*S*)-**202** in 88.4-87.6% *ee*, *Porti*AmDH and *Cfus*AmDH gave the same enantiomer but with only 13.9-29.1% *ee* and eventually MATOUAmDH2 gave the opposite (*R*)-**203** in only 38.2% *ee*. The differences in the active site shape or key residues of these five enzymes are of particular interest to consider altering the typical (*S*)-enantioselectivity of nat-AmDHs, whether by doing mutagenesis or looking for homologs in the biodiversity. To the best of our knowledge, these nat-AmDHs are the only enzymes performing reductive amination yet reported to provide **202**. *Asp*RedAm and mutants were well characterized by Li and co-workers towards furan derivatives but in presence of primary amines (Jia *et al.*, 2021; Yang *et al.*, 2021). LE-AmDH-v1, reported by Tseliou *et al.*, outperformed the nat-AmDHs in these conditions reaching 99% conversion from both 10 and 50 mM of **31** while *Cfus*AmDH and *Rgna*AmDH reached 79.0% and 76.5%, respectively [Tseliou *et al.*, 2019 (A)].

Table 4. Screening of nat-AmDHs towards furfural (**57**), cyclopentancarbaldehyde (**199**), benzaldehyde (**31**), acetophenone (**39**) and hexan-3-one (**200**). The cells highlighted in shades of green gives the analytical conversion (%) obtained for the synthesis of the corresponding amines and the cells colored in shades of yellow gives the corresponding *ee* for the formation of the (*S*)-amine (%). For hexan-3-amine (**202**-**203**), the commercial standard of pure (*S*)-enantiomer (**202**) or (*R*)-enantiomer (**203**) was not available in our laboratory but the peak obtained in majority was assumed to be the (*S*)-enantiomer **202**. L (Left peak) was assumed to correspond to the (*R*)-enantiomer **203**. Reaction conditions were as follows: final volume 100 µL, 10 mM substrate (**31** and **39** diluted in dimethyl sulfoxide (**204**) (DMSO) final 5% volume/volume (v/v)), 2 M NH_4HCO_2 buffer, pH 8.5, 0.2 mM $NAD(P)^+$, 1.2 eq. glucose (**222**), 3 U mL^{-1} GDH-105, 0.5 mg mL^{-1} purified enzyme, 25°C, 24 h, 400 rpm. conv.: analytical conversion; *n.t.*: not tested; *n.d.*: not detected.

		Chat AmDH	IGCAmDH1	*Rgna* AmDH	*Apau* AmDH	MATOUAmDH2	*Cfus* AmDH	*Msme* AmDH	*Porti* AmDH	*Micro* AmDH
	conv. (%)	76.3	62.7	68.2	77.3	79.7	83.5	78.9	44.5	37.2
	conv. (%)	41.1	44.6	47.1	43.7	49.2	51.3	56.0	49.4	48.3
	conv. (%)	9.7	14.7	76.5	22.2	56.9	79.0	37.2	23.5	14.5
	conv. (%)	*n.t.*	*n.t.*	*n.t.*	<1	<1	<1	1.9	7.4	13.9
	ee (%)							96.5	97.0	99.1
	conv. (%)	*n.d.*	*n.d.*	*n.d.*	<1	12.4	5.8	39.1	46.5	51.2
	ee (%)					38.2	29.1	88.4	13.9	87.6

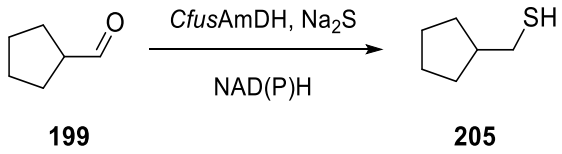
IV.2.1.2 Monitoring of the alcohol formation

GC-FID analytical method was not only largely used to monitor the amine formation but could also enable the monitoring of the alcohol. Indeed, some substrates could adopt a specific positioning in the active site that would favor the direct reduction of the carbonyl into alcohol. This aspect has been rarely studied within the nat-AmDHs and could be highly enzyme and substrate-dependent. Therefore, it should be integrated as a routine study especially for the characterization of novel enzymes or substrates. For the screening described in Chapter III, I.4.3, such a monitoring has been performed. Among the WT nat-AmDHs, the alcohol formation never reached higher than 1.2% obtained with *Chat*AmDH/**31** (9.7% conversion in amine) and *Porti*AmDH/**200** (46.5% conversion in amine) enzyme/substrates couples, corresponding to a ration of amine:alcohol formation of 8:1 and 33:1, respectively (Appendix 16). This work is currently being completed by Dr. Carine Vergne-Vaxelaire and Aurélie Fossey-Jouenne. It describes the ability of some nat-AmDHs and mutants to catalyze the direct reduction into alcohol in presence or absence of **11** and will be submitted for a publication soon. Some

in silico analyses will aim at giving more insights into the substate-enzyme-dependent results.

V. Opening to the reductive sulfidation activity

Within the L2BMS, another research focus is bioremediation and, more particularly, the study of the degradation of the well-known pesticide chlordecone by bacterial communities. Largely implied in this project, Pierre-Loïc Saaidi's team of the Laboratory of Genomics and Biochemistry of Metabolism (LGBM) of the same Research Unit, previously identified a reductive sulfidation activity on this compound by the sulfate-respiring bacterium *Desulfovibrio* sp.86 (Della-Negra *et al.*, 2020). More recently, they completed this study by comparing the microbial and chemical reductive sulfidation of a range of carbonyl compounds through experimental and theoretical approaches (Della-Negra *et al.*, 2021). We decided to combine their expertise in performing and monitoring reductive sulfidation reactions with our knowledge on enzymatic catalysis to carry out very preliminary tests to evaluate the potential of nat-AmDHs to catalyze the formation of thiols from carbonyls when supplied with a sulfur entity as nucleophile instead of **11**. To the best of our knowledge, no enzymatic alternative was ever described for the reductive sulfidation. The selection of carbonyl substrates to test the nat-AmDHs-mediated reductive sulfidation was based upon the set of compounds tested by Della-Negra *et al.* in the aforementioned publication (Della-Negra *et al.*, 2021). A large number of these compounds belong to nat-AmDHs substrate scope such as isobutyraldehyde (**170**), (substituted) cyclohexanone (**45**), furfural (**57**) or cyclopentancarbaldehyde (**199**). Considering the high capacity of *Cfus*AmDH in the reductive amination of **199**, described in Chapter II, IV.2.1.1, this enzyme-substrate couple was selected for preliminary tests to study a panel of reaction conditions in presence of Na_2S as the sulfur donor (Scheme 30).



Scheme 30. Expected reductive sulfidation of cyclopentancarbaldehyde (**199**) in presence of Na_2S and *Cfus*AmDH.

Considering the high required concentration of **11** for reductive amination, the first key parameter to look at was Na_2S concentration to provide enough sulfur donor but avoid enzyme inhibition. The activity was monitored by Gas Chromatography equipped with a HeadSpace Module and Mass Detection (GC-HS-MS) through conversion after 24 h at 30°C (Experimental section, IV.3.2) but we rapidly removed the GDH/**222** recycling system normally used and provided NAD(P)H at equimolar

concentration to avoid any additional problem of GDH inhibition. The thiol formation monitoring in blank reactions without *Cfus*AmDH revealed a spontaneous reductive sulfidation of **199** highly dependent upon Na_2S concentration. It was therefore not possible to estimate the biocatalytic thiol formation based on the calibration curves, we remained on the Mass (MS) response area comparison of the targeted peaks in the reaction with *Cfus*AmDH and in blanks without the enzyme (Figure 36).

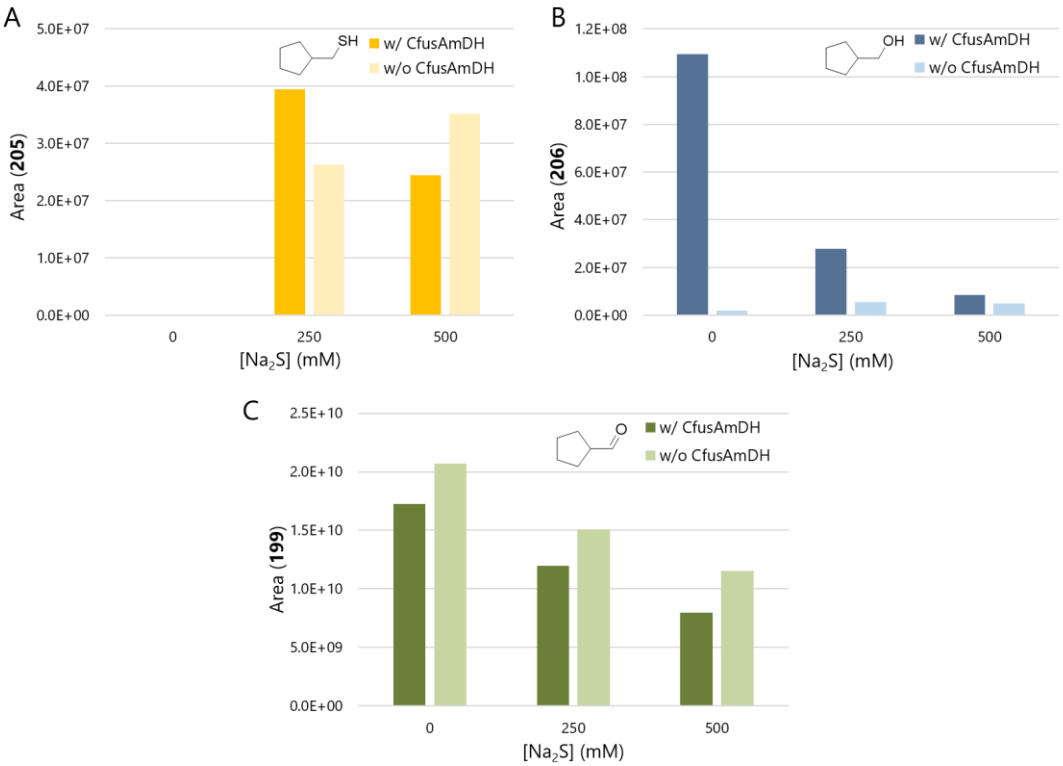


Figure 36. Thiol and alcohol formation from cyclopentancarbaldehyde (**199**) in presence of Na_2S and *Cfus*AmDH. Area resulted from the transformation of **199** in presence or absence of *Cfus*AmDH and 0, 250 and 500 mM Na_2S. MS response areas of (A) cyclopentylmethanethiol (**205**) formed (B) cyclopentylmethanol (**206**) formed and (C) remaining **199**. The reaction conditions were as follows: final volume 100 µL, 10 mM **199**, 0-500 mM Na_2S, 100 mM potassium phosphate buffer pH 9, 5 mM NADPH, 5 mM NADH, 0.5 mg mL^{-1} purified *Cfus*AmDH, 30°C, 24 h, 400 rpm. GC-HS-MS chromatograms for peaks identification are given in Appendix 17.

The results revealed a similar proportion of thiol formed in the blanks and in the nat-AmDH-containing reaction and confirmed the high tendency of spontaneous reductive sulfidation of this compound in the tested conditions. Interestingly, in absence of Na_2S, *Cfus*AmDH activity for direct reduction into cyclopentylmethanol (**206**) was much higher and drastically decreased in presence of 250 and 500 mM Na_2S. This can be typical from an enzyme inhibition or to the inference of Na_2S in the pocket that does not allow the correct substrate positioning for direct reduction. Another peak was detected and could correspond to a sulfur-containing compound that was not clearly identified. As it was a very preliminary work, the reactions were not run in duplicates but it would have been better to do so to confirm the slight differences observed between the blanks and the reactions. The same

procedure was carried out for other enzyme-substrate couples: *Msme*AmDH/**45**, MATOUAmDH2/hexanal (**207**) and *Micro*AmDH/pentan-3-one (**181**). Some substrates were less likely to proceed to spontaneous reductive sulfidation according to Della-Negra *et al.* assays (Della-Negra *et al.*, 2021). Among them, only MATOUAmDH2 exhibited higher hexan-1-thiol (**208**) formation than the corresponding blank at 250 mM Na_2S (Figure 37 and Appendix 18). To confirm this observation, we ran again the same reaction comparing a larger panel of Na_2S concentrations. The thiol formation in presence of the enzyme was estimated two-to-six-fold higher than without the enzyme thus confirming the catalytic role of MATOUAmDH2 in the reductive sulfidation of **207**.

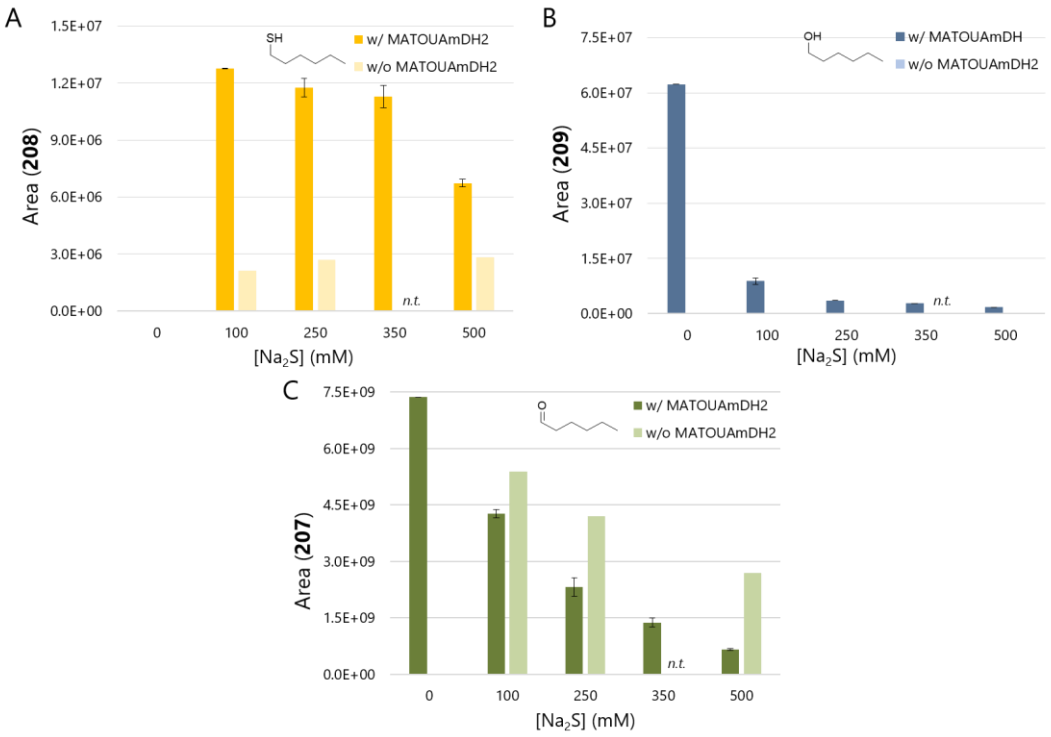


Figure 37. Thiol and alcohol formation from hexanal (**207**) in presence of Na_2S and MATOUAmDH2. Area resulted from the transformation of **207** in presence or absence of MATOUAmDH2 and 0 to 500 mM Na_2S. MS response areas of (A) hexan-1-thiol (**208**) formed, (B) hexan-1-ol (**209**) formed and (C) remaining **207**. The reaction conditions were as follows: final volume 100 µL, 10 mM **207**, 0-500 mM Na_2S, 100 mM potassium phosphate buffer pH 9, 5 mM NADPH, 5 mM NADH, 0.5 mg mL^{-1} purified MATOUAmDH2, 30°C, 24 h, 400 rpm. GC-HS-MS chromatograms for peaks identification are given in Appendix 18. *n.t.*: not tested.

However, the high instability of thiol in the reaction mixture make it impossible to quantify the thiol formation as the calibration curves appeared highly different when prepared the day of the reaction starting or the day of the injection, 24 h later. Dr. Oriane Della Negra carried out additional tests with MATOUAmDH2/**207** couple using lower Na_2S concentrations and also tested MATOUAmDH2/**57** and reached the same conclusions. The results with *Msme*AmDH/**45** are given in Appendix 19 and Appendix 20. As for the couple *Micro*AmDH/**200**, no particular peak stood out from the background and we did

not have the corresponding sulfur standard to be sure of the retention time to target a peak of low intensity. No additional study was done until this was written but in the future, it would be interesting to estimate the nat-AmDH activity for reductive sulfidation using spectrophotometry that enables a direct comparison of the nat-AmDH-mediated reaction and the spontaneous one and the removal of thiols instability issue. On the other hand, it will not enable to discriminate thiol to alcohol formation. Also, other sulfur donor than Na_2S should be considered such as H_2S bubbling or sodium thiosulfate (**210**). The decreasing of the pH from 9 to 7-8 could also enfavor the presence of the H_2S of $Na_2S/HS^-/H_2S$ system which could be more accepted in the active site. Eventually, the carboxylate-mediated fixation of NH_3 for the reductive amination might not be adapted for HS^- fixation in the tested conditions and one option could be to mutate a nat-AmDH or find native enzymes with similar scaffold but a positive (Arg, Lys) or neutral residues at this position.

VI. Conclusion

The deeper structural analysis on whether nat-AmDHs crystallographic structures or homology models gave us a view of the active site and global fold diversity found in the family, particularly into the G3-G4 groups of ASMC1 classification. This guided the engineering work performed on key reference nat-AmDHs to highlight the importance of some key positions of the active site. The stability study of a panel of enzymes in different storage conditions revealed an enzyme-dependent influence of the temperature. The correlation of such experimental data with structural features was not an easy task but could give additional hints for future enzymes selection. Also, without knowing the native substrate of the G3-G4 enzymes, we enriched our knowledge of their promiscuous substrate scope and confirmed their large preference for short carbon chains (up to C5) with activities easily comparable to that of reported RedAms, engineered AmDHs or even TAs. This work also confirmed the need for protein engineering to have access to amine with carbon chains larger than C5 which is the main topic of Chapter III. Eventually, a preliminary study was carried out to estimate the potential of some nat-AmDHs to catalyze the reductive sulfidation of carbonyl substrates in presence of the sulfur donor Na_2S. Although MATOUAmDH2 seemed to afford a catalytic thiol formation in the midst of the high rate of spontaneously formed thiol, further experiments are needed to confirm this observation.

CHAPTER III

Improvement of native amine dehydrogenases

The deeper characterization of nat-AmDHs, described in Chapter II, revealed a great capability of these enzymes to transform key carbonyl substrates that were not reported yet. Nevertheless, except MATOUAmDH2 harboring a slightly wider substrate scope thanks to a larger active site volume, no other nat-AmDH could accept substrates with a main chain bearing more than five C-atoms such as hexanal (**207**) or hexan-2-one (**43**). The *in silico* analysis of the active site enabled the identification of key spots to overcome this hurdle through directed mutagenesis. In addition, two other approaches for improvement of nat-AmDHs are described in this chapter. An approach of *in vivo* directed evolution was set up to adapt some nat-AmDHs and a RedAm and enhance their activity towards selected substrates that are not accepted by the native enzymes. Eventually, some preliminary tests were carried out to create a chimera that would merge the thermostability and the substrate scope of their parental enzymes.

I. Expansion of the native amine dehydrogenases substrate scope through structural exploration and protein engineering

This part has been largely inspired by the paper recently published in ChemCatChem for which I wrote the entire initial draft and participated actively to its improvement and to the submission (Ducrot et al., 2022). In this work, the crystallization studies were carried out in collaboration with the team of Prof. Grogan (University of York), mainly by him and his PhD student Megan Bennett. The MD studies were done in collaboration with Dr. Gwenaëlle André-Leroux from MaIAGE unit at INRAE and with the great help of Dr. Eddy Elisée from LABGeM of our Research Unit.

I.1 *In silico* analysis

By first looking at the structures of *Cfus*AmDH (Uniprot: S9Q235, PDB: 6IAU), *Msme*AmDH (Uniprot: A0A8B4R7U8, PDB: 6IAQ) and the models of their 291 homologs already generated for the ASMC1 (Mayol *et al.*, 2019), we hypothesized that the restricted substrate scope of nat-AmDHs may be due to a very small active site. Indeed, it is mainly built by hindered aromatic amino acids located on the first layer, including Tyr or Phe at P5, P12 and P17, Tyr, Phe or Trp at P10 and Trp at P8 (Figure 23 and Figure 38). The accommodation of the *n*-alkyl chain as the already reported butan-2-one (**134**) (Ducrot

et al., 2021) drove the hypothesis that the system formed by the positions P5 (F140 in *Cfus*AmDH) and P8 (W145 in *Cfus*AmDH) represents a good target for mutations to enlarge the pocket and extend the substrate scope to longer *n*-alkyl ketones and aldehydes ≥ C6 (Figure 38). As depicted in Figure 24 and Appendix 5, discerning examples from biodiversity display smaller residues at these positions, such as Ala at P5 in A0A1N7N2Q9 and A0A1X4NNW3, Cys at P8 in MATOUAmDH2, Gly at P8 in OM-RGCAmDH2, A0A078BAL0, A0A078AU13 and A0A077ZQK8, and Leu at P8 in A0A076NBW6. The occurrence in close nat-AmDHs homologs of smaller amino acids at positions P5 and P8, while harboring identical or similar neighbors, suggests that mutations at these positions into smaller residues should not be detrimental neither for enzyme folding nor for activity. This hypothesis still needs to be nuanced by the heterologous expression in *E. coli*. The effect of a small residue at P8 for broader substrate scope was supported by MATOUAmDH2 (P8 = Cys) activity in presence of bulkier substrates such as norcamphor (**177**) and analysis of MATOUAmDH2 RX structure (PDB: 7R09, 7ZBO) (Bennett *et al.*, 2022; Caparco *et al.*, 2020). The crystallographic structure of *Cfus*AmDH (PDB: 6IAU) was used as a reference for *in silico* mutations and subsequent docking of ligands with longer *n*-alkyl chains. As the native enzyme showed a preference for aldehydes over ketones (Mayol *et al.*, 2019), the docking was done with *n*-terminal primary amine products, from pentan-1-amine (**211**), hexan-1-amine (**212**), heptan-1-amine (**213**), octan-1-amine (**214**), nonan-1-amine (**215**), decan-1-amine (**216**) (Experimental section VI.2.2.3).

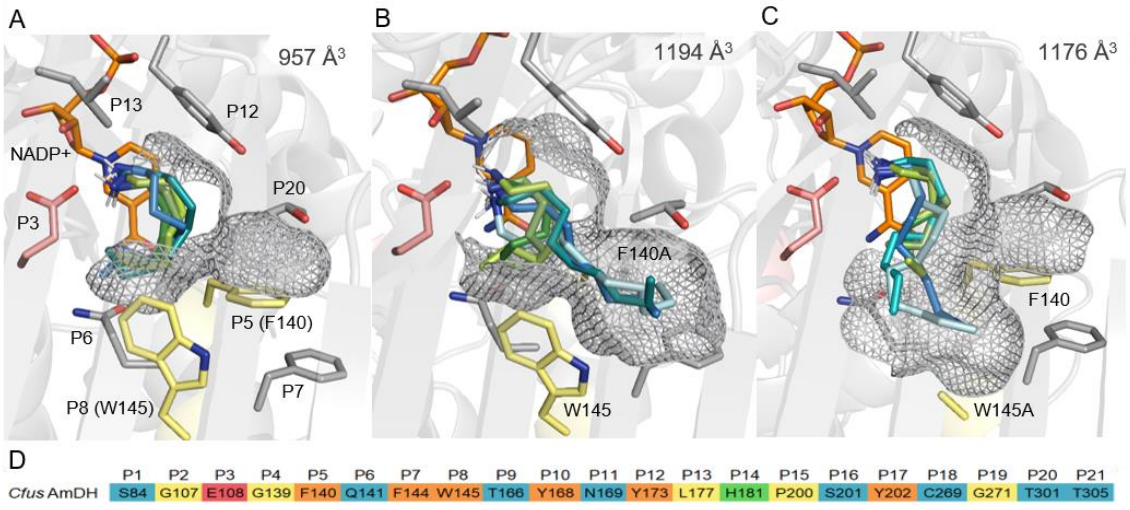


Figure 38. Active site of *Cfus*AmDH (A) (RX structure; PDB: 6IAU), *Cfus*AmDH-F140A (B) and *Cfus*AmDH-W145A (C) with docked structures of pentan-1-amine (**211**) to heptan-1-amine (**213**) in shades from dark to light green and octan-1-amine (**214**) to decan-1-amine (**216**) [nonan-1-amine (**215**) for *Cfus*AmDH] in shades from black to light blue. $NADP^+$ is represented in orange, the catalytic residue E108 (P3) is in pink, F/A140 (P5) and W/A145 (P8) are in yellow and some residues that form the first layer of the pocket are in grey. For sake of clarity, only the most energetically favorable pose is represented. The pocket enlargement in the two mutants is shown using the grey surface representation. The active site volumes are given in $Å^3$. (D) P1-P21 residues in *Cfus*AmDH. The color code used refers to the polarity and charge of the corresponding residue [blue: polar residues, yellow: hydrophobic residues, orange: aromatic residues, red: negatively charged residues, and green: positively charged residues (charges at physiological pH)].

As depicted in Figure 38, the mutations F140A and W145A increased the active site volume from 957 Å^3 in *Cfus*AmDH to 1194 Å^3 and 1176 Å^3, respectively. These were calculated using CAVER Web tool (Stourac *et al.*, 2019). For *Cfus*AmDH, the small chains **211** and the longest amine formed experimentally **212**, were obtained in the highest number of energetically viable poses inside the active site with top pose binding energy of -5.5 and -5.9 kJ mol^{-1}, respectively (Figure 39).

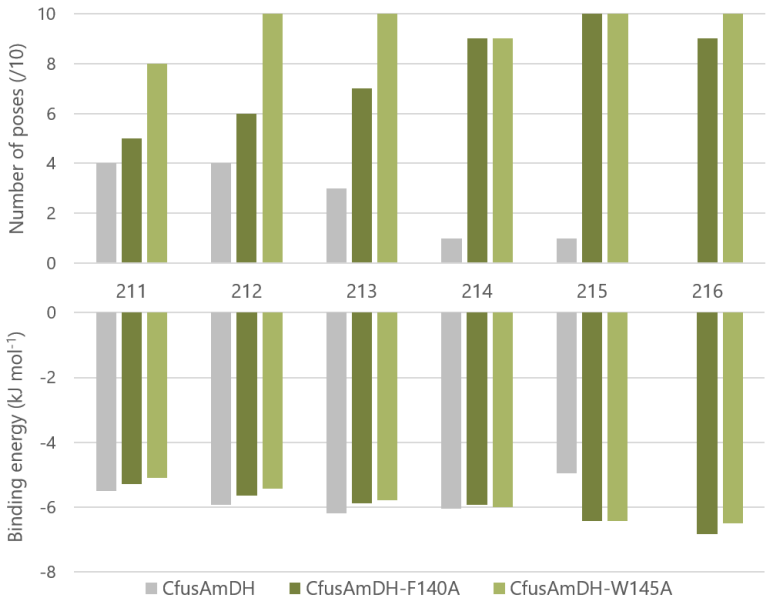


Figure 39. Results of pentan-1-amine (**211**) to decan-1-amine (**216**) dockings in *Cfus*AmDH, *Cfus*AmDH-F140A, *Cfus*AmDH-W145A. (A) Number of poses docked inside the active site out of the 10 poses generated (B) Lowest binding energies of the ligands docked, given in kJ mol^{-1}.

Fewer than three poses of the ten were obtained with larger ligands up to **215** and in a single conformation with higher binding energy (-5.0 kJ mol^{-1}) that appeared highly constrained within the active site. In *Cfus*AmDH-F140A and *Cfus*AmDH-W145A mutants, the pocket could easily accommodate longer *n*-alkylamines up to **216** (longer chains have not been tested) with nine and ten poses inside the pocket out of ten generated poses, and with binding energies even lower than for *Cfus*AmDH/**211** (-6.8 and -6.5 kJ mol^{-1} for *Cfus*AmDH-F140A and *Cfus*AmDH-W145A respectively) (Figure 39). Docked amines occupied the space freed up by the small Ala residue and led to linear conformations in *Cfus*AmDH-F140A (Figure 38B) and to slightly curved conformations in *Cfus*AmDH-W145A (Figure 38C), both directed towards the active site floor. The structure of nat-AmDHs has been observed in both open and closed conformations, the open conformation maintaining the active site pocket accessible to the solvent (Bennett *et al.*, 2022; Mayol *et al.*, 2019). Thus we can hypothesize that the accessibility of larger compounds such as heptanal (**217**), octanal (**218**), nonanal (**219**) or decanal (**220**) should not be a limitation compared to enzymes displaying a restrictive access tunnel (Riziotis *et al.*, 2022). Further *in*

silico experiments such as MD simulations studying how the substrates access the pocket could complement this hypothesis.

I.2 Specific activity of *Cfus*AmDH variants towards longer aldehydes

Based on *in silico* results, positions F140 and W145 in *Cfus*AmDH were mutated into Ala. The mutations were introduced using SDM kits on the corresponding nat-AmDH gene-containing-pET22b(+) plasmid, first in a small batch and then in a larger batch (see Experimental sections, II.2, II.3, II.4 and II.5). In small batches, the two mutated proteins were obtained in similar production rate as the WT enzyme thus suggesting that the mutations did not altered enzyme expressability or folding. However, it was observed a slight protein degradation of *Cfus*AmDH-F140A at the temperature of the spectrophotometric assay (30°C). Thus, only the specific activity of *Cfus*AmDH-W145A was measured towards 10 mM of pentanal (**175**) to octanal (**218**) in presence of 5% DMSO to avoid substrate insolubility issue (Table 5) (Experimental section, III.3). *Cfus*AmDH-F140A apparent instability correlated with a total loss of activity in large scall production when using the general method. The modification of the lysis step protocol from sonication to BugBuster®-mediated lysis and changes in the buffer recipe enabled the activity recovery (see Experimental section II.5).

Table 5. Specific activities (mU mg^{-1}) of *Cfus*AmDH and *Cfus*AmDH-W145A towards pentanal (**175**) to octanal (**218**). Reaction conditions: final volume 100 µL, 10 mM aldehyde substrate (200 mM in DMSO), 2 M NH_4HCO_2 buffer, pH 8.5, 0.2 mM NAD(P)H, 0.03-0.25 mg mL^{-1} purified enzyme, 30°C. Errors represent standard deviations of two or three independent experiments.

	Specific activities (mU mg^{-1})							
	pentanal (**175**)		hexanal (**207**)		heptanal (**217**)		octanal (**218**)	
	NADH	NADPH	NADH	NADPH	NADH	NADPH	NADH	NADPH
Cfus AmDH	73 ± 7	188 ± 2	12 ± 1	19 ± 1	17 ± 5	-	-	-
Cfus AmDH-W145A	16 ± 2	62 ± 1	195 ± 7	401 ± 2	405 ± 10	371 ± 11	197 ± 12	483 ± 23

As expected from the docking experiments, *Cfus*AmDH displayed very low or no activity towards substrates larger than **175** (< 20 mU mg^{-1}). On the contrary, *Cfus*AmDH-W145A was clearly active, displaying activities of 401, 405 and 483 mU mg^{-1} towards **207**, **217** and **218**, respectively. Interestingly, this mutant was less active than the WT enzyme towards the smaller substrate **175** as it was already reported for some engineered AADHs (Chen *et al.*, 2019; Chen *et al.*, 2018; Franklin *et al.*, 2020). This might be due to an unsuitable positioning of the shorter substrates in the pocket with active spatial arrangement mainly induced by steric hindrance along with the hypothesized weak stabilization by interaction with Y168 (P10) (see Chapter II, II.4). The specific activities of *Cfus*AmDH-W145A towards aldehydes **207**, **217** and **218** were significantly higher than the only other reductive amination activities reported for the same substrates with *Es*LeuDH-DM (320 *vs* 400 mU mg^{-1} for **207**, 190 *vs* 371 mU mg^{-1}

for **217** and 90 *vs* 483 mU mg^{-1} for **218** (Löwe *et al.*, 2018). Even if it could not be correctly estimated due to protein instability, *Cfus*AmDH-F140A was clearly active towards **175**, **207**, **217** and **218**.

Another interesting observation can be done on the substrate-dependent cofactor preference of the mutants. With **175**, both *Cfus*AmDH and *Cfus*AmDH-W145A seemed to prefer NADPH over NADH with ratios of 2.5:1 and 4:1, respectively. However, this ratio decreased to 2:1 with **207** and was even reversed with **217** with a slight preference for NADH (specific activities of 405 and 371 mU mg^{-1} with NADH and NADPH, respectively). For the longer substrate **218**, NADPH appeared to be preferred. This phenomenon was more pronounced for the W145A mutant while activities remained very low for the WT enzyme. It was already shown, in preliminary experiments, that the specificity of *Cfus*AmDH for NADH or NADPH was substrate-dependent for different classes of carbonyl substrates (Mayol *et al.*, 2019). Here, clear differences can be also observed among *n*-alkylketones (Table 5).

As Gly was also found at P8 of nat-AmDHs (OM-RGCAmDH2, A0A078BAL0, A0A078AU13 and A0A077ZQK8) and in order to further probe the catalytic potential of mutations at positions P5 (F140) and P8 (W145), substitutions to Gly were also investigated. The goal was to reach even better accommodation of the longer substrates nonanal (**219**) and decanal (**220**) as obtained by Franklin *et al.* with higher activity of L-AmDH_TV_L39G towards octan-2-one (**67**), nonan-2-one (**135**) and decan-2-one (**68**) than L39A mutant (226.7 vs 62.7 mU mg^{-1} towards **67**, for example (Franklin *et al.*, 2020). Mutants *Cfus*AmDH-F140G and *Cfus*AmDH-W145G were produced and tested towards **217**-**219** similarly to Ala mutants. Insolubility even at 10 mM with 5% DMSO prevented accurate specific activity measurements towards longer substrates such as **220**, but qualitative analysis showed that the mutation W145G did not lead to notably better activities towards longer substrates **218**-**219** compared to W145A and even resulted in total inactivity in the case of F140G (data not shown). Using higher ratio of DMSO cosolvent up to 10% did not help to obtain sufficiently stable UV signal to quantify the activity towards **219**. Thus, this mutation was not further investigated.

I.3 Conversion assays with NAD(P)-recycling system

The potential of the mutants F140A and W145A for biocatalytic reactions was confirmed through the quantification of heptanal (**217**) conversion to heptan-1-amine (**213**). Reactions containing 10 mM substrate were incubated for 24 h at 25°C with a GDH from *Candida boidinii* (GDH-105)/**222** cofactor recycling system and 1 to 20% (v/v) of DMSO as co-solvent. After derivatization with benzoyl chloride

(BzCl, **221**), the amine formed was monitored using UHPLC-UV at 250 nm (Figure 40) (Experimental section IV.2.3).

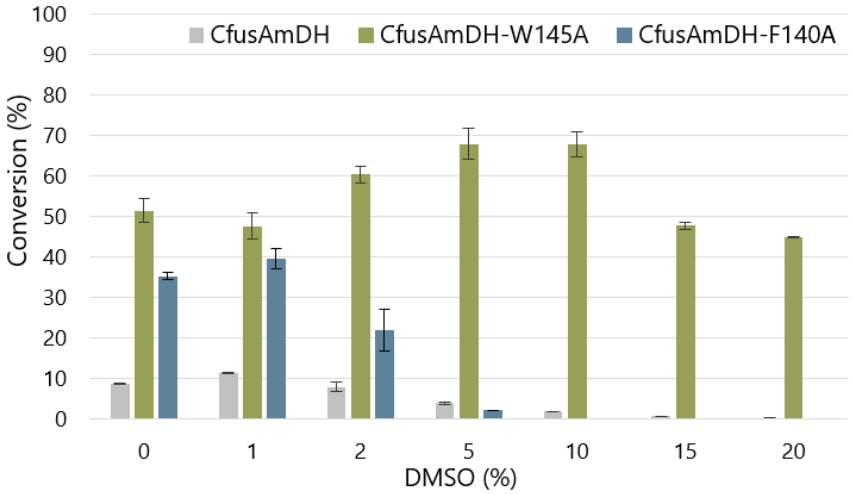


Figure 40. Conversion rates obtained from heptanal (**217**) with *Cfus*AmDH, *Cfus*AmDH-F140A and *Cfus*AmDH-W145A with various ratios of DMSO. Conversion rate obtained for *Cfus*AmDH, *Cfus*AmDH-F140A and *Cfus*AmDH-W145A reported in this paper towards heptanal (**217**). Reaction conditions: final volume 100 µL, 10 mM **217** (final ratio of DMSO 1-20%), 2 M NH_4HCO_2 buffer, pH 8.5, 0.2 mM $NAD(P)^+$, 1.2 eq. **222**, 3 U mL^{-1} GDH-105, 0.5 mg mL^{-1} purified enzyme, 25°C, 24 h, 400 rpm. The temperature was set at 25°C after a preliminary study that displayed lower conversions at 30°C (Appendix 21). The monitoring of heptan-1-amine (**213**) was performed by UHPLC-UV after derivatization with BzCl. Error bars represent standard deviations of two independent experiments.

Mutants W145A and F140A were able to catalyze the conversion of 39.6 – 68.0% of **217**, respectively, in contrast to the WT enzyme which gave < 10% conversion. As for the WT enzyme, mutant F140A was more affected by higher DMSO concentrations, as the conversion dropped from 39.6% at 1% DMSO to 2.2% at 5% DMSO, compared to mutant W145A that conserved the same range of conversion in presence of 1-20% DMSO. With the mutant W145A, an increasing concentration of DMSO led to higher conversions from 47.7% in 1% DMSO to 68.0% at 10% v/v, and a good conversion of 45.0% was still observed at 20% v/v DMSO. Based on these results, 5% v/v of DMSO was considered as an appropriate compromise for further experiments with these mutants.

I.4 Transposition of the positive mutations into other native amine dehydrogenases

I.4.1 Mutagenesis

In order to investigate the transposability of the mutational pattern defined above, the mutations corresponding to F140A and W145A in *Cfus*AmDH were replicated in nine other nat-AmDHs (*Msme*AmDH; *Micro*AmDH; *Apau*AmDH; MATOUAmDH1; MATOUAmDH2; *Porti*AmDH; *Rgna*AmDH; *Chat*AmDH; IGCAmDH1). Mutations into Ala were preferred over Gly to avoid unstable and/or inactive protein as observed in *Cfus*AmDH upon F140G mutation, although it could have helped reaching activity

towards even larger substrates in other templates. Some of the selected nat-AmDHs were already active towards a broader range of substrates compared to the rest of the family, namely MATOUAmDH1, MATOUAmDH2 (Caparco *et al.*, 2020; Bennett *et al.* 2022) and *Micro*AmDH, the latter displaying low activity towards 3C ketones such as pentan-3-one (**181**) and hexan-3-one (**200**) (192.0 mU mg^{-1} and 28.1 mU mg^{-1}, respectively) (Mayol *et al.*, 2019). *Porti*AmDH and *Rgna*AmDH were part of the nat-AmDHs found among metagenomic databases but were not extensively characterized as purified enzymes (Caparco *et al.*, 2020). Among the nine selected enzymes and their 18 mutants produced using the same protocol as described for *Cfus*AmDH mutants by overexpression in *E. coli* and purification on Ni-NTA columns (see Experimental section II.2, II.3 and II.4), only MATOUAmDH1 and its two mutants could not be produced with a sufficient yield to be used in activity screening. Mutagenesis primers and SDS-PAGE gels can be found in the Supporting Information of the publication (Ducrot *et al.*, 2022). All variants were assayed and compared to their respective WT enzyme against a range of carbonyl substrates, including the aldehydes hexanal (**207**), heptanal (**217**), octanal (**218**), nonanal (**219**) and decanal (**220**), and the ketones from hexan-2-one (**43**), heptan-2-one (**128**), octan-2-one (**67**) and nonan-2-one (**135**), as well as the commonly tested 4-phenylbutan-2-one (**36**). The five substrates cyclopentancarbaldehyde (**199**), furfural (**57**), acetophenone (**39**), benzaldehyde (**31**) and hexan-3-one (**200**) were also part of this screening but the results were not reported in the publication as the mutants were not significantly more active than the WT enzymes towards these substrates. These results using WT enzymes are described in Chapter II, IV.2. The summary of substrates and mutants tested are reported in Appendix 22.

I.4.2 Screening analytical development

Besides monitoring the amine formed, we also wanted to monitor the alcohol formation. The mutation could result in a different positioning of the substrate, closer to the C4N of the cofactor, and thus lead to a different rate of direct reduction compared to what was previously obtained (see Chapter II, IV.2.1.2). Notably, Jia *et al.* recently reported a two-fold improvement in promiscuous ADH activity of *Asp*RedAm harboring a single mutation N93A (Jia *et al.*, 2021). Previous monitoring by UHPLC-UV after BzCl derivatization prevented any sensitive detection of many targeted non-UV alcohols and water-compatible derivatization method for alcohols are not appropriate for an efficient screening. Ultra-High Performance Liquid Chromatography equipped with Mass detection (UHPLC-MS) could have been used to reach the expected sensitivity (0.01- 0.05 mM), but the moderate to high volatility of listed carbonyl compounds and alcohols tend to prefer GC as the more appropriate analytical method. By the way, GC-FID or Gas Chromatography coupled with Mass detection (GC-MS) were largely reported for biocatalytic research on reductive amination or transamination of similar carbonyl compounds [Kollipara et al., 2022; Marshall et al., 2021; Mu et al., 2021; V. Tseliou et al., 2019 (A)]. In the meantime, the carbonyl substrates

monitoring was considered a more satisfying method, not based only on product formation as with UHPLC-UV BzCl protocol. Following advices obtained by Aurélie Fossey-Jouenne from GC columns suppliers, we decided to test the TG-35MS AMINE column, harboring a diphenyl/dimethyl polysiloxane 35/65 phase for better retention of amines, while still being suitable for alcohols and ketones. For practical organization in the laboratory and ease of use, GC-FID was preferred over GC-MS. The first test and optimizations were performed using the system octanal (**218**)/octan-1-ol (**223**)/octan-1-amine (**214**), after extraction of the amine, the alcohol and aldehyde from buffered aqueous media, using EtOAc as extracting solvent. When injected alone, the amine standard detection showed a very good linearity depending only on amine concentration but, surprisingly, it was not the case in presence of the aldehyde. Thus, with the help of Dr. Pierre-Loïc Saaidi, we hypothesized the formation of the corresponding imine.

The imine formation was confirmed using increased running time methods with the detection of a peak whose retention time depended upon the aldehyde or amine chain length and whose area depended on the proportion of both entities. We first tried to inhibit this imine formation by changing the extraction solvent [methyl tert-butyl ether (**224**, MTBE), CH_2Cl_2, toluene (**225**), isooctane (**227**)], increasing the injection temperature up to the column limit of 250°C, adding 2% or 10% CH_3OH to compete with the amine and prevent imine formation in the injector or even add 0.5 mM *p*-toluenesulfonic acid (**227**) to "reduce" the amine nucleophily that could favor the imine formation. However, none of these methods reduced the imine formation (data not shown). Therefore, we decided to estimate the amine formation by adding both amine and imine areas in reaction mixtures and corresponding calibration curves. We confirmed the viability of this method first by establishing calibration curves made in the same conditions in presence of amine and aldehyde and we attested the linearity of the estimated "amine" area (see some examples of calibration curves from the screening in Appendix 23). Then, by comparing the conversion obtained using GC-FID or UHPLC-UV/MS detection on the transformation of heptanal (**217**) using *Cfus*AmDH-W145A, we definitely confirmed this analytical protocol. This test also served to compare the use of whether NAD^+, $NADP^+$ or a mix of both with *Cfus*AmDH (Figure 41). The results revealed a slightly lower conversion estimated with GC-FID compared to UHPLC-UV and even more with UHPLC-MS. However, the conversions obtained in GC-FID are given with high error bars that include the standard deviations of two peaks instead of one. We estimated that the results obtained with GC-FID were still not abnormal and could be used for the screening when considering the large benefit of monitoring the alcohol formation. During the screening, detailed in the following part, we observed that the imine formation occurred only with aldehyde substrates. To the best of our knowledge, such an analytical issue was not reported in publications relating GC-monitoring of amine formation from aldehydes. Either the authors did not mention this analytical detail, or the treatment or running conditions used here favored the imine formation.

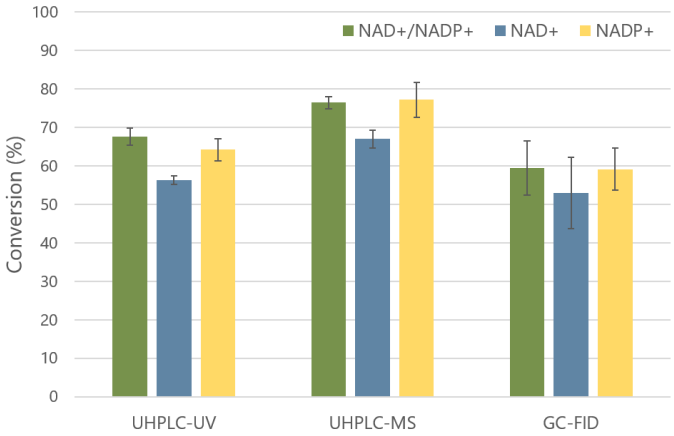


Figure 41. Conversion of heptanal (**217**) with *Cfus*AmDH estimated from UHPLC-UV, UHPLC-MS and GC-FID detection methods. Reaction conditions: final volume 100 µL, 10 mM substrate (final ratio of DMSO 5%), 2 M NH_4HCO_2 buffer, pH 8.5, 0,2 mM $NADP^+$ or 0,2 mM NAD^+ or 0.1 mM $NADP^+$/0.1 mM NAD^+, 1.2 eq. **222**, 3 U mL^{-1} GDH-105, 0.5 mg mL^{-1} purified enzyme, 25°C, 24 h, 400 rpm. Error bars represent standard deviations of two independent experiments.

Regarding the cofactor, *Cfus*AmDH delivered an average of 1.1-fold higher conversion rates when supplemented in $NADP^+$ compared to NAD^+, although the error bars with $NADP^+$ appeared slightly higher. When supplemented in a mixture of both oxidized cofactors, the resulting conversion was similar to that obtained with $NADP^+$ thus meaning that the enzyme is sufficiently supplied in its apparent "preferred" cofactor for this reaction. Also, is has to be noted that GDH-105 used for cofactor recycling is more active towards NAD^+ than $NADP^+$ thus making more difficult to estimate the cofactor preference of one enzyme using such an assay. Anyway, the screening was carried out in presence of both cofactors as the tested WT enzymes were already reported with different cofactor preference and we previously showed that this feature could also be substrate-dependent.

The extracting protocol and injection methods used for each substrate have been optimized for each threesome carbonyl/amine/alcohol and are detailed in Experimental section, IV.1 and IV.3.1. Only the reactions supplied in 4-phenylbutan-2-one (**36**) were monitored using UHPLC-UV due to a boiling point higher than the limit temperature of TG-35MS AMINE column. The detection was done on the compound derivatized with FDAA to combine conversion and *ee* estimations (Experimental section, IV.2.4). The activity screening has not been carried out in duplicates considering the high number of reactions and analytical GC injections. All the preliminary experiments, including those regarding the imine issue and the extracting protocol optimizations, led to satisfying reproductible results (3-9% error). It also has to be noted that the protocol used from reaction conditions to extraction was optimized only for *Cfus*AmDH-W145A and the threesome heptanal (**217**)/heptan-1-amine (**213**)/heptan-1-ol (**228**). This could not be well-adapted for other enzyme templates and substrates. Only the GC-FID running method

was optimized for each substrate to have clear separated peaks of each substrate and product in the shortest possible time (Table 20, Experimental section, IV.3.1).

I.4.3 Screening results on carbonyl compounds harboring a chain length of at least six carbon atoms

The outcome of the screening are given in Figure 42, Appendix 24 and Appendix 25. All the corresponding GC-FID or UHPLC-UV chromatograms are given in Appendix 25 to Appendix 39 according to Appendix 22. To ensure accurate results, cautions have been taken during the long series of GC injections, notably regarding evaporation in the vials and the calibration points by performing injections of the calibration points before and after the various series. It has to be noted that in "real" reaction mixtures, the imine peak was detected for hexanal (**207**), heptanal (**217**), furfural (**57**), cyclopentancarbaldehyde (**199**) and benzaldehyde (**31**) and both the amine and imine peaks were added as described above (Appendix 23A, B and C). However, for **218-220**, the imine peak was not detected even with elongated analytical methods and the calibration curves based on the amine peak were still not linear, revealing the bias brought by imine formation (Appendix 28, Appendix 29 and Appendix 30). The conversions were therefore estimated using the most adapted polynomial model in each case on the calibration curves considering only the peak corresponding to the amine. This estimation is of course not ideal mainly due to the presence of alcohol in the calibration curves and potentially in the reaction mixtures that "breaks" the direct relationship between amine, aldehyde and imine ratios in the mixture. The most notable improvements were seen for the aldehydes hexanal (**207**) to nonanal (**219**) with moderate to high conversions (20 - > 99 %) by mutant enzymes whereas most of the WT enzymes, such as *Chat*AmDH, IGCAmDH1, *Rgna*AmDH, *Cfus*AmDH and *Apau*AmDH, could not convert substrates larger than **207**. For example, *Rgna*AmDH converted 12.5% of **207** whereas its mutants Y150A and W155A enabled conversions of 95.2% and 85.6%, respectively. *Rgna*AmDH-Y150A was one of the best performing P5A mutants with 44.6% conversion of **219**. *Chat*AmDH and IGCAmDH1 were barely active as WT enzymes towards aldehydes **207** and **217**-**219**, but the P8 mutations W166A and W144A achieved moderate conversions of 26.2% and 34.6%, respectively. Similarly, for *Porti*AmDH, for which the WT enzyme converted 60.8% of **207** and 7.9% of **217**, only mutant W140A exhibited significant conversions (22.0 - 23.8%) of longer chain substrates up to **219**. An even higher conversion of > 99.5% was obtained with **207** and *Cfus*AmDH-W145A. As expected, the natively larger active site of MATOUAmDH2 enabled 72.2% conversion of **207** and longer substrates up to **219** with 20.5%. In this case, the mutations of P5 and P8 residues (-F143A and -C148A, respectively) did not noticeably increase the conversion for all the other aldehydes. Interestingly, only *Rgna*AmDH-Y150A, MATOUAmDH2-F143A, *Cfus*AmDH-F140A and *Apau*AmDH-Y136A mutated in P5 displayed notable conversions of **220** (7.8 – 11.5%), suggesting that this mutation led to improved accommodation of longer carbon chain. Such reported conversion data

towards aldehydes **207** and **217**-**220** with the direct use of **11** as the amine donor are of importance, as the only other example in the literature is that of *Es*LeuDH-DM (Löwe *et al.*, 2018). This mutant was introduced in a formate dehydrogenases-based system for *in situ* cofactor regeneration using algae technology to provide formate **117** from CO_2. In these conditions, it could achieved 99% formation of **212** but the longer **214** was only obtained as traces. Some conversion data on **207** and **217** by enzymes performing reductive amination have been reported, but only with methylamine (**38**) and propargylamine (**50**) as amine donor [Aleku *et al.*, 2019; Tseliou *et al.*, 2019 (B)].

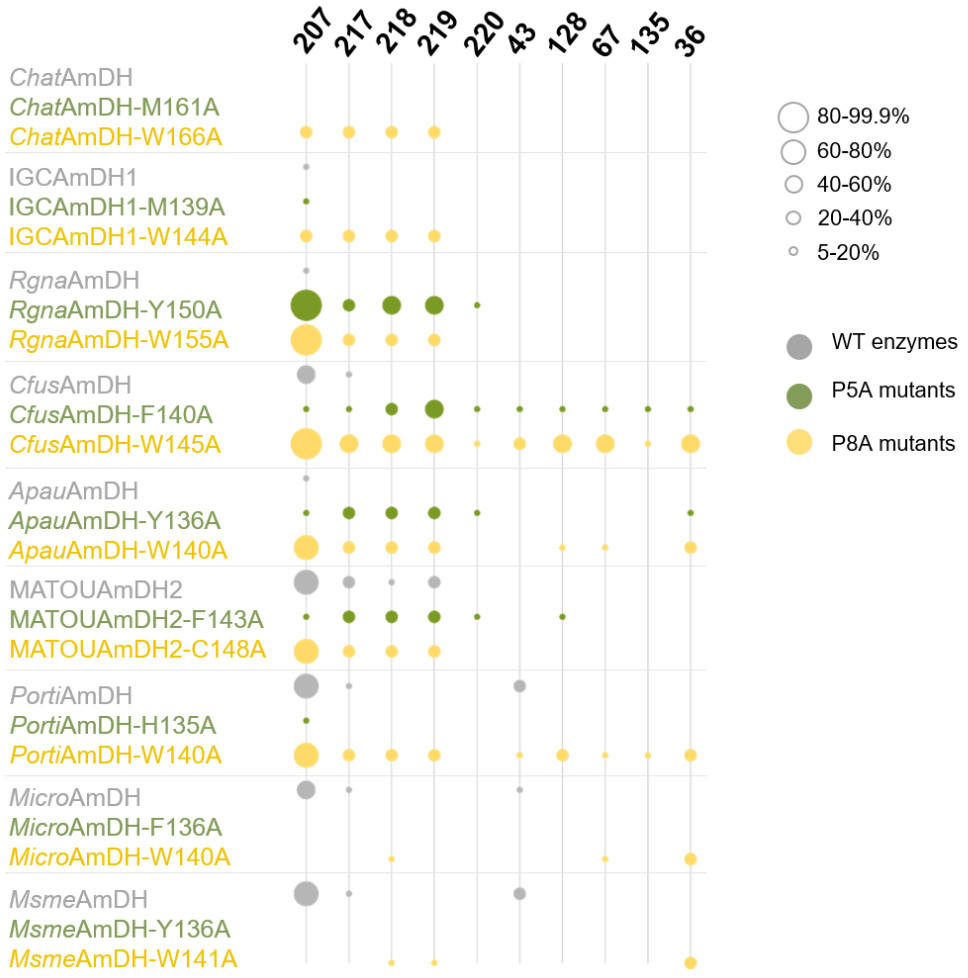


Figure 42. Conversions obtained for WT enzymes and mutants towards substrates **207**, **217**-**220**, **43**, **128 67**, **135** and **36**. Reaction conditions: final volume 100 µL, 10 mM substrate (DMSO 5% v/v except for **207** and **43** solutions prepared in water), 2 M NH_4HCO_2 buffer, pH 8.5, 0.2 mM $NAD(P)^+$, 1.2 eq. **222**, 3 U mL^{-1} GDH-105, 0.5 mg mL^{-1} purified enzyme, 25°C, 24 h, 400 rpm. The monitoring of the remaining aldehydes or ketones, the amines formed and the corresponding alcohols was performed by GC-FID after extraction in EtOAc. The monitoring of 4-phenylbutan-2-amine **55**-**56** was carried out by UHPLC-UV after derivatization with FDAA (Appendix 35).

For the *n*-alkylketone substrates from hexan-2-one (**43**) to nonan-2-one (**135**), most of the WT enzymes displayed very low or no conversion even towards the small **43**, except for *Porti*AmDH and *Micro*AmDH, with 26.5% and 22.5% conversion of this substrate with 79.5% and 94.6% *ee*, respectively, in favor of the (2*S*)-hexan-2-amine (**229**). Among the mutants, *Cfus*AmDH-F140A, *Cfus*AmDH-W145A, *Apau*AmDH-W145A and *Porti*AmDH-W140A gained activity towards ketones from **43** to **135**. *Cfus*AmDH-W145A gave the highest conversions with 39.7% of **43** (99.3% *ee* for **229**), 53.7% of **128** [99.7% *ee* for (2*S*)-heptan-2-amine (**230**)], 40.3% of **67** [99.5% *ee* for (2*S*)-octan-2-amine (**231**)] and 16.0% of **135** [> 99.9% *ee* (2*S*)-nonan-2-amine (**187**)]. In most cases, the (*S*)-enantioselectivity of the WT enzymes was conserved. Notably, the P5 mutation in *Porti*AmDH (*Porti*AmDH-W140A) led to a slightly lower conversion (26.5% for WT against 18.2% for the mutant) but with a noticeable improved stereoselectivity with *ee* from 79.4% to 99.5%. Thanks to these mutations, we have now in hand alternative nat-AmDH templates for the reductive amination of aliphatic ketones, complementary to both AmDHs engineered from α- and ε-AADHs and RedAms, which already displayed high efficiency and stereoselectivity towards such kinds of carbonyl substrate, but giving the opposite (*R*)-enantiomer. For instance, with *Gk*AmDH from *Geobacillus kaustophilus*, Liu *et al.* reported 99% and 98% conversion (>99% *ee*) for the formation of (2*R*)-hexan-2-amine (**232**) and (2*R*)-heptan-2-amine (**130**), respectively, but the conversion dropped to < 5% with **67** despite high enzyme loadings (1 - 3 mg mL^{-1}) (Liu *et al.*, 2020). Importantly, *Cfus*AmDH-W145A, and to a lesser extent *Porti*AmDH-W140A, constitute interesting alternatives to *Nf*RedAm, the only enzyme reported thus far that gives the (*S*)-alkylamines with 41% formation of (2*S*)-decan-2-amine (**233**) with 84% *ee* (Mangas-Sanchez *et al.*, 2020). Finally, *Cfus*AmDH-F140A, *Cfus*AmDH-W145A, *Apau*AmDH-W141A, *Porti*AmDH-W140A and *Micro*AmDH-W141A were active towards 4-phenylbutan-2-one (**36**), a substrate not observed previously to be (*S*)-aminated by nat-AmDHs and converted to (2*R*)-4-phenylbutan-2-one (**55**) with engineered enzymes [Aleku *et al.*, 2017; Franklin *et al.*, 2020; Knaus *et al.*, 2017; Mu *et al.*, 2021; Pushpanath *et al.*, 2017; Sharma *et al.*, 2018; Tseliou *et al.*, 2019 (A); Tseliou *et al.*, 2019 (B); Wang *et al.*, 2021; Ye *et al.*, 2015]. Higher conversions were again obtained with *Cfus*AmDH-W145A, with 52.5% conversion into (2*S*)-4-phenylbutan-2-one (**56**) (99.6% *ee*). It is worth noting that for *Micro*AmDH and *Msme*AmDH, the P8 mutation of the residue was only beneficial for **36** transformation. For all assays, the alcohols were formed in very small amount or as traces, without exceeding the 4.0% reached with IGCAmDH1-W144A for the reduction of octanal (**218**) into octan-1-ol (**223**) (Appendix 25). These results confirmed, as previously reported, that nat-AmDHs do not significantly catalyze the reduction of carbonyls to alcohols in presence of excess of **11** and that this behavior is not altered by the P5 and P8 mutations (Jongkind *et al.*, 2022; Mayol *et al.*, 2019).

I.4.4 Screening results on short aliphatic and aromatic carbonyl compounds

This part completes the results already described in Chapter II, IV.2 on the activity of the WT nat-AmDHs with that of the corresponding mutants obtained towards furfural (**57**), cyclopentancarbaldehyde (**199**), benzaldehyde (**31**), acetophenone (**39**) and hexan-3-one (**200**) (Table 6).

Table 6. Results in amine formation for the screening of nat-AmDHs WT and mutants towards **57**, **199**, **31**, **39** and **200**. The cells highlighted in shades of green give the conversion rate (%) in the corresponding amines and the cells colored in shades of yellow gives the corresponding *ee* in the case of the formation of chiral amine (%). If not indicated, the *ee* is given for the (*S*)-amine. For hexan-3-amine **202** and **203**, the commercial standard of pure (*S*)-**202** or (*R*)-**203** was not available in our laboratory but the peak obtained in majority was assumed to be the **202**. [a] L (Left peak) was assumed to correspond to the **203**. conv.: analytical conversion; *n.t.*: not tested; *n.d.*: not detected

Substrate		*Chat* AmDH			IGCAmDH1			*Rgna* AmDH			*Apau* AmDH			MATOUAmDH2			*Cfus* AmDH			*Msme* AmDH			*Porti* AmDH			*Micro* AmDH		
		WT	M161A	W166A	WT	M139A	W144A	WT	Y150A	W155A	WT	Y136A	W141A	WT	F143A	C148A	WT	F140A	W145A	WT	Y136A	W141A	WT	H135A	W140A	WT	F136A	W141A
57	conv. (%)	76.3	46.1	<1	62.7	64.1	<1	68.2	82.7	3.2	77.3	61.2	1.5	79.7	72.5	72.4	83.5	88.1	13.3	78.9	1.2	<1	44.5	30.2	4.1	37.2	<1	<1
199	conv. (%)	41.1	34.8	2.1	44.6	32.0	1.3	47.1	48.4	25.8	43.7	30.1	12.3	49.2	51.9	54.3	51.3	50.1	18.6	56.0	9.2	21.8	49.4	53.4	53.9	48.3	52	48
31	conv. (%)	9.7	3.1	2.4	14.7	6.7	3.7	76.5	62.9	2.6	22.2	14.1	1.7	56.9	<1	55.3	79.0	65.5	14.8	37.2	*n.t.*	*n.t.*	23.5	27.2	35.0	14.5	*n.t.*	*n.t.*
39	conv. (%)	*n.t.*	*n.t.*	*n.t.*	*n.t.*	*n.t.*	*n.t.*	*n.t.*	*n.t.*	*n.t.*	<1	<1	<1	<1	<1	<1	<1	<1	2.2	1.9	*n.d.*	<1	7.4	<1	2.6	13.9	*n.d.*	<1
	ee (%)																		96.9	96.5			97.0		97.3	99.1		
200	conv. (%)	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	<1	*n.d.*	*n.d.*	12.4	1.4	9.0	5.8	1.1	1.7	39.1	*n.d.*	<1	46.5	5.2	6.7	51.2	<1	4.4
	ee[a] (%)													38.2 (L)	31.2 (L)	12.1 (L)	29.1	77.9	91.3	88.4			13.9	59.3	98.0	87.6		>99.9

These substrates clearly belong to the substrate scope of the WT enzymes. The mutations of residues at P5 and P8 in Ala, rarely achieved to increase the conversion and most often led to similar or even decreased activity. The only notable exception was observed for *Rgna*AmDH-Y150A whose mutation significantly improved the conversion of **199** (82.7% vs 68.2% for the WT enzyme). Except **200**, the steric hindrance of the substrates tested here are on the same order of magnitude as the "small" cyclic substrates already well identified as good substrates for WT enzymes. These results are therefore logical. This general observation correlates with a possible detrimental effect of such mutations on the activity towards shorter substrates as already mentioned before and observed in other engineering work on nat-AmDHs (Chen *et al.*, 2019; Chen *et al.*, 2018; Franklin *et al.*, 2020). On the other hand, when considering MATOUAmDH2, *Cfus*AmDH, *Porti*AmDH and *Micro*AmDH activity towards **200**, the mutations P5A and P8A highly decreased the conversion but greatly improved the *ee* for the hypothesized **202** formation. This effect is much more visible with P8A mutants than P5A ones, thus suggesting that P8A mutation can afford a clearer orientation of the substrate whatever its C-atom chain length to favor the (*S*)-amine formation. For instance, *Porti*AmDH, -H135A and -W140A mutants exhibited 46.5%, 5.2% and 6.7% conversion towards **200**, respectively, but 13.9%, 59.3% and 98.0% *ee* for **202**, respectively. As for MATOUAmDH2, -F143A and -C148A, the conversion decreased from 12.4%

to 1.4% and 9.0% and the *ee*, hypothesized to be for the reverse **203** decreased from 38.2% to 31.2% and 12.1%, suggesting again a higher proportion of (*S*)-amine formed. These analyzes of *ee* are however to be moderated because they are made at quite low conversions, and are so less accurate, even if detected amount were above the Limit Of Detection (LOD).

I.4.5 Complementary results on ketones with the carbonyl function on the third carbon atom of the chain

In light of the encouraging conversions obtained with *Porti*AmDH, *Msme*AmDH, *Micro*AmDH and, on a lesser extent, MATOUAmDH2 towards **200**, we decided to test these enzymes, and their respective mutants P5A and P8A towards longer 3C ketones such as heptan-3-one (**234**) and octan-3-one (**235**). As already mentioned in Chapter I, IV.2, having in hands some notable conversions data towards 3C ketones with nat-AmDHs was of large interest as such substrates were rarely reported to be enzymatically aminated. After 24 h of incubation at 25°C the amines formed were directly derivatized with FDAA to both estimate the quantity of amine formed and the *ee* (Experimental section, IV.2.5). As the amine standards of heptan-3-amine (**236**) and octan-3-amine (**237**) were not available in our laboratory, only an estimated conversion is given based on a calibration curve done on hexan-3-amine (**238**). The hits obtained should be validated with the standard amines ordered recently (Figure 43).

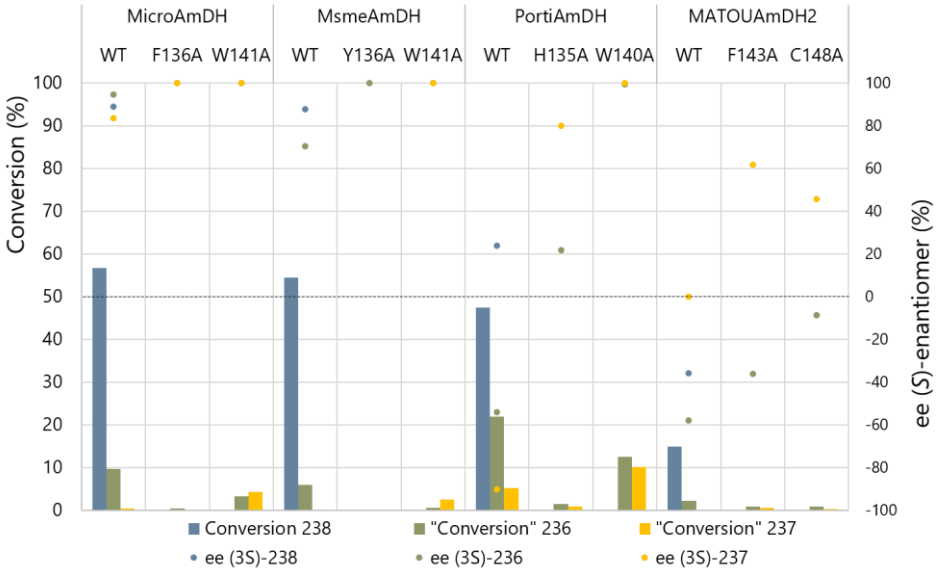


Figure 43. Conversion of 3C ketones with *Micro*AmDH, *Msme*AmDH, *Porti*AmDH, MATOUAmDH2 and their P5A and P8A respective mutants. Reaction conditions: final volume 100 µL, 10 mM substrate (DMSO 5% v/v), 2 M NH_4HCO_2 buffer, pH 8.5, 0.2 mM NAD(P)$^+$, 1.2 eq. **222**, 3 U mL^{-1} GDH-105, 0.5 mg mL^{-1} purified enzyme, 25°C, 24 h, 400 rpm. Bars represent the conversion (%) in hexan-3-amine (**238**) and estimated conversion in heptan-3-amine (**236**) and octan-3-amine (**237**). The scatter plots represent the corresponding *ee* (%) for the hypothesized (*S*)-amines. The black dashed line gives the threshold for 0% *ee* for the (S)-enantiomer, corresponding to an equimolar formation of (*S*)- and (*R*)-. Negative *ee* values show an excess in (*R*)-amines.

The conversion of **200** confirmed the hits obtained from the screening with a great potential of *Micro*AmDH, *Msme*AmDH and *Porti*AmDH to accept 3C ketones, with once again no activity of the respective mutants towards this C6 substrate. With the addition of only one C-atom in the substrate chain, the conversions dropped from 47.4 - 56.7% to estimated 6 - 21.9%, the maximum being observed with *Porti*AmDH (21.9%) and *Porti*AmDH-W140A (12.9%). The mutation started being beneficial with **235**, equivalent to a **207** when looking at the number of C-atom in the longest chain, with a 1.9-fold higher estimated conversion obtained with *Porti*AmDH-W140A than with the WT enzyme (10.1% vs 5.2%). Very low conversion rates (< 5%) were observed with *Micro*AmDH and *Msme*AmDH mutants towards those two substrates. This is in agreement with the screening results where none of these mutants was particularly more active than the WT enzymes towards longer substrates chains either towards aldehydes or 2C ketones. 3C ketones with more than six C-atoms in the longest chain (in total > C9 ketones) would have been interesting to test, especially to attest P5 and P8 mutation effects. That has not been performed yet for solubility issues and by lack of time. Interestingly, we reached the same observation of higher proportion of (*S*)-amine formed when the reaction is catalyzed by the mutants instead of the WT enzyme, although these are still associated with low conversions. *Porti*AmDH provided the (3*R*)-octan-3-amine (**239**) with 91.2% *ee* while *Porti*AmDH-W140A formed the opposite (3*S*)-octan-3-amine (**240**) in an enantiomerically pure form (detection limit not determined). All things considered, *Porti*AmDH and its mutants represent good systems that should be studied to rationalize the acceptance of 3C ketones and even the nat-AmDHs stereoselectivity (see Chapter IV, II.5.3). To the best of our knowledge, these enzymes and mutants are the only one yet reported to provide the (*S*)-enantiomer of such 3C ketones, which remained poorly accepted substrates. *Nf*RedAm and *Nfis*RedAm were the only enzyme described to form 90% and 52% of the **239** but with low 40% and 58% *ee* (Mangas-Sanchez *et al.*, 2020).

I.4.6 Biochemical characterizations of *Cfus*AmDH-W145A

I.4.6.1 Kinetic parameters

Kinetic parameters were determined to further characterize one of the most promising mutants *Cfus*AmDH-W145A arising from the conversion screening, and more specifically to gain insights on the observed change in cofactor specificity dependent upon the aldehyde substrate (see Chapter III, I.2). The following four operating conditions were assayed, all by spectrophotometric monitoring the UV maximum absorbance of the nicotinamide cofactor at 340 nm in micro-cuvettes: heptanal (**217**)/NADPH/NH_3, **217**/NADH/NH_3, pentanal (**175**)/NADPH/NH_3, **175**/NADH/NH_3 (Table 2 and see Experimental section, III.4). Plots are reported in Appendix 40-Appendix 43, respectively.

Table 7. Kinetic parameters of *Cfus*AmDH-W145A. Kinetic parameters were measured in a 100 µL final volume in NH_4HCO_2 buffer (0.05-4 M, pH 8.5), 0.1-12 mM of **217**, 0.5-120 mM of **175**, 0.01-0.9 mM of NADPH and 0.05-0.9 mM of NADH. k_{cat} are given in s^{-1}, K_M and K_i in mM and k_{cat}/K_M in s^{-1} mM^{-1}. [a] K_M values calculated with Lineweaver-Burk model. [b] Apparent kinetic parameters as the NADH concentration used could not reach saturation. [c] Apparent kinetic parameters as the calculation was made on a limited number of points due to inhibition profile that does not fit any inhibition equation.

	k_{cat}	K_M	K_i	k_{cat}/K_M
217	0.74 ± 0.03	0.29 ± 0.04		2.56
NADPH	0.78 ± 0.02	0.21 ± 0.02		3.71
NH_3	0.72 ± 0.02	196.20 ± 25.50		3.62
217	0.59 ± 0.02[b]	1.03 ± 0.15[b]		0.57[b]
NADH	1.78 ± 0.17	3.00 ± 0.77[a]		0.59
NH_3	0.52 ± 0.01[b]	126.40 ± 13.30[b]		3.98[b]
175	0.20 ± 0.05	12.51 ± 4.60	41.34 ± 15.73	0.02
NADPH	0.10 ± 0.01	0.07 ± 0.01		1.4
NH_3	0.10 ± 0.01[c]	319.80 ± 52.00[c]		0.31[c]
175	0.03 ± 0.01[b,c]	3.95 ± 0.43[b,c]		0.01[b,c]
NADH	0.07 ± 0.01	0.74 ± 0.05[a]		0.09
NH_3	0.03 ± 0.01[b,c]	191.00 ± 33.90[b,c]		0.17[b,c]

As expected, whether using NADPH or NADH, *Cfus*AmDH-W145A displayed 57 to 128-fold higher (apparent) catalytic efficiency for **217** than for **175**, with k_{cat}/K_M with NADPH cofactor of 2.56 and 0.02 s^{-1} mM^{-1}, respectively. This is mainly due to the K_M being 43 and four-fold higher for **175** than **217** with NADPH and NADH, respectively, that can attest a difference of protein affinity for the two substrates. Surprisingly, compared to the specific activities presented in Chapter III, III.1, the K_M of the cofactor were 11-to-14-fold lower for NADPH over NADH with **175** and **217**. This difference may be due to the different NAD(P)H concentrations used, optimized at 0.6 mM to reach saturation compared to 0.2 mM when running specific activities but this effect has not been checked. Interestingly, the poor affinity of *Cfus*AmDH-W145A for **175** was accompanied by an inhibition effect with a K_i estimated at 41.34 mM with NADPH (Appendix 42 and Appendix 43). This also led to an inhibition over 1.5 M NH_4HCO_2 buffer, which was not observed with **217** at even higher concentrations (Appendix 40-Appendix 41). No explanation was found for this behavior but complementary analyses confirmed that it was neither due to pH, ionic strength of the reaction mixture nor product inhibition (data not shown). It is worth noting that the mutation did not impact significantly the affinities for NH_3 in any of the assayed conditions; the K_M of 98.9 mM being previously reported for the WT enzyme (Mayol *et al.*, 2019) and K_M of 100 – 300 mM being common for nat-AmDHs. The **11** fixation by P3 was so not altered by P5 and P8 mutations,

which was important to check as P3 is the catalytic residue. Only the k_{cat} and affinity of longer aldehydes such as **217** were beneficially impacted; no data could be obtained with the WT enzyme that was poorly active towards **217**. To our knowledge, no kinetic parameters have been reported for such aldehydes and any enzymes performing reductive amination that would enable any comparison.

I.4.6.2 Thermostability

Despite the occurrence of small amino acid residues at P8 in native homologs, one cannot exclude that W145A mutation could have an impact on the structure flexibility and stability of *Cfus*AmDH. Thus, we compared the thermostability of *Cfus*AmDH and *Cfus*AmDH-W145A at 25°C, 30°C and 40°C through activity assays (Figure 44 and see Experimental section, III.5). The mutant appeared to be slightly less thermostable than the WT enzyme but still displayed good $t_{1/2}$ of 350 h and 200 h at 25°C and 30°C, respectively, compared to > 600 h and 480 h for *Cfus*AmDH. *Cfus*AmDH-W145A experimental features (stability at the optimum temperature range, DMSO tolerance) (Figure 40) make this mutant a good candidate with respect to envisaged synthetic applications.

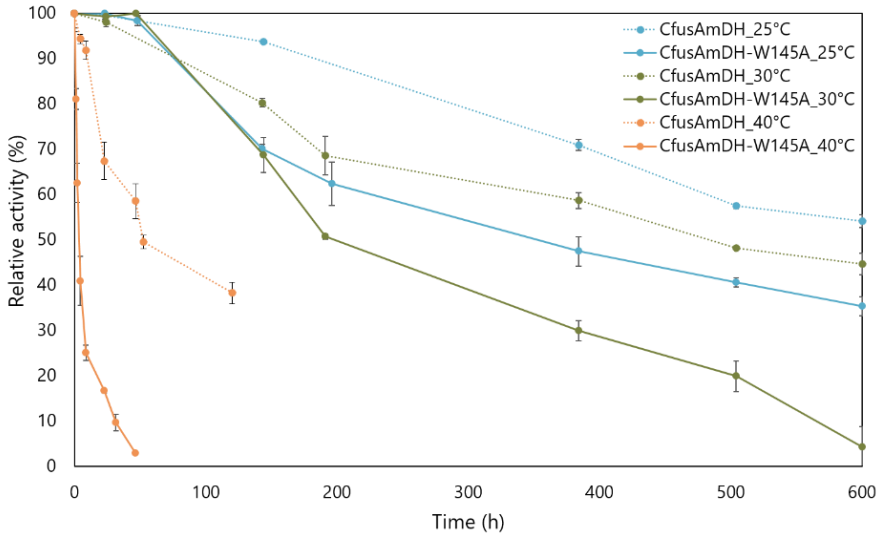


Figure 44. Thermostability of *Cfus*AmDH and *Cfus*AmDH-W145A at 25°C, 30°C and 40°C. Specific activity assays were carried out at different time points using the following reaction conditions: 10 mM pentanal (**175**) with *Cfus*AmDH and heptanal (**217**) with *Cfus*AmDH-W145A, 2 M NH_4HCO_2 buffer, pH 8.5, 0.2 mM NADPH, 0.05-0.12 mg mL^{-1} purified enzyme, 30°C. The maximum activity of *Cfus*AmDH towards **175** was 219.2 mU mg^{-1} and the maximum activity of *Cfus*AmDH-W145A towards **217** was 256.3 mU mg^{-1}. Errors represent standard deviations of two or three independent experiments.

I.4.6.3 Tolerance to substrate loading

For synthetic purposes, the substrate tolerance of *Cfus*AmDH-W145A was studied with increasing loadings of heptanal (**217**) and octan-2-one (**67**) up to 200 mM and various enzyme loadings for reaction times of 24 h and 48 h (Figure 45 and see Experimental section, IV.4.1). During the

experiments for kinetic parameters determination, no substrate inhibition was observed but it was only studied up to 20 mM substrate. The results obtained after derivatization with BzCl and UHPLC-UV analysis showed that *Cfus*AmDH-W145A still catalyzed the formation of more amine up to 100 and 200 mM of **217** and **67**, respectively, than at 10 mM. Up to these respective concentrations, the same trend was observed for both substrates; a decrease in conversion but still 20.6% (**217**) and 15.0% (**67**) at 48 h with 0.5 mg mL^{-1} of purified enzyme (Figure 45). Higher substrate loading appeared to be possible with this mutant especially with the linear ketone. The enzyme loading did not drastically enhance the conversion, thus at 1 mg mL^{-1}, conversions of **217** and **67** at 100 mM were 24.2% and 29.2% at 48 h, respectively (20.6% and 24.3% with 0.5 mg mL^{-1}).

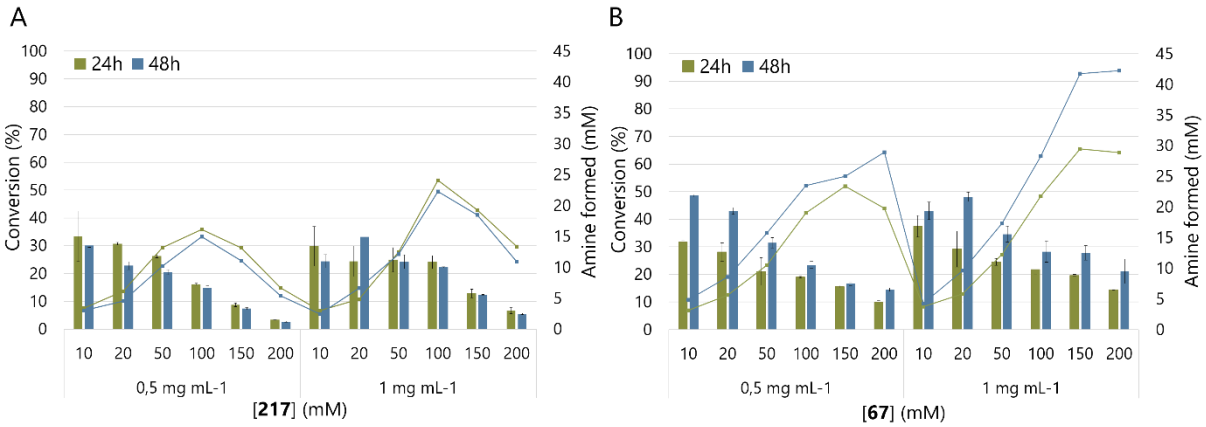
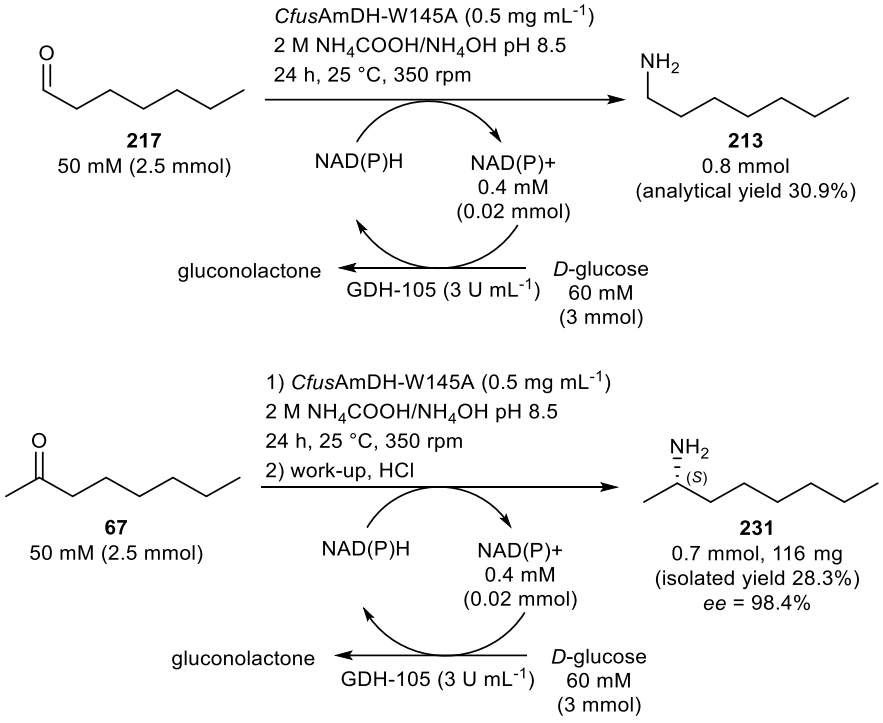


Figure 45. Substrate loading influence on conversion of heptanal (**217**) and octan-2-one (**67**) with *Cfus*AmDH-W145A. Conversion rate (%, bars) and corresponding molarity of amine formed (mM, scatter dots) with *Cfus*AmDH-W145A towards (A) **217** and (B) **67** both after 24 and 48 h at 25°C, 400 rpm.

I.4.6.4 Utilization of *Cfus*AmDH-W145A in preparative scale reactions

Semi-preparative reactions were performed to demonstrate the viability of *Cfus*AmDH-W145A for synthetic applications. Using the previous results, the reactions were carried out on a 50 mL scale with 50 mM of **67** and **217** and 0.5 mg mL^{-1} enzyme loading for 24 h to find a compromise between conversion, quantity of amine formed, reaction time and amount of purified enzyme. Higher conversions should be reached with higher enzyme loadings such as 1 mg mL^{-1} studied above, but the scale up deep optimization remains an integral part of research and development in itself that was not our goal here. After liquid work up, 116 mg of the hydrochloride salt (*2S*)-octan-2-amine (**231**) were recovered (32.2% analytical conversion, 28.3% isolated yield, *ee* 98.4%) (Scheme 31 and Appendix 44). ^{1}H and ^{13}C NMR analyses were in accordance with the commercial standard (Appendix 46 and Appendix 47). However, the heptan-1-amine (**213**) product, obtained with 30.9% conversion (Scheme 31 and Appendix 45), could not be isolated and recovered using the same protocol (Experimental section, IV.4.2). The washing and extraction steps resulted in good yields but even after amine salification using 1.8 eq. HCl in ether, the

evaporation led to a very poor yield of a brownish wet solid that could not be further dried. Preliminary tests on standard **213** in a solution mimicking the final reactive mixture [2M ammonium formate buffer, 30 mM **217**, 20 mM **213**, 40 mM glucose (**222**), 20 mM gluconolactone (**241**)] revealed a high loss of the amine during the LiCl washing, previously adjusted for the reaction work up but still resulted in the formation of a dried white power. We did not have enough enzyme materials to conduct the reaction from **217** again at this scale and decided not to produce a new batch of purified enzyme. The reductive amination of **67** led to STY of 2.32 g L^{-1} d^{-1} and TON of 2112 and that of **217** were estimated from the conversion to STY of 2.33 g L^{-1} d^{-1} and TN of 2322. In comparison, Mangas-Sanchez *et al*. reported on the similar substrate, hexan-2-one (**43**), a STY of 8.16 g L^{-1} d^{-1} and a TON of 3250 that decreased in presence of DMSO (Mangas-Sanchez *et al*., 2020). Even though these values are slightly higher than those calculated for preparative scale reactions with *Cfus*AmDH-W145A, they remain in the same order of magnitude.



Scheme 31. Semi-preparative scale synthesis of heptan-1-amine (**213**) and (2*S*)-octan-2-amine (**231**).

I.4.7 Structural characterization of *Cfus*AmDH-W145A

I.4.7.1 Crystallization

In order to both explore the structural effect of the mutation and the reasons of the altered activity of *Cfus*AmDH-W145A and its cofactor specificity, discussed in Chapter III I.2, its structure was solved using X-ray crystallography (Experimental section, VI.1). Two structures were determined; one in

complex with NAD^+ (PDB: 7QZN) and the other with $NADP^+$ (PDB: 7QZL) and pentan-1-amine (**211**), to resolutions of 1.64 Å and 1.50 Å, respectively (Appendix 48). Unfortunately, no structure could be obtained in a reasonable time for *Cfus*AmDH-W145A in complex with cofactor and heptan-1-amine (**213**). Overall, the structures obtained superimposed well with the WT structure in complex with $NADP^+$ and cyclohexylamine (**179**) (6IAU, Mayol *et al.*, 2019), with a RMSD < 0.361 Å over > 277 C-alpha carbons, each representing 'closed' conformations of the nat-AmDH as previously defined (Mayol *et al.*, 2019). The structures verify that the W145A mutation has created the extra active site space with the pocket volume estimated at 1349 $Å^3$ in *Cfus*AmDH-W145A in complex with NAD^+ compared to 957 $Å^3$ in *Cfus*AmDH. Compared to the *in silico* structure mutation done as preliminary experiment (see Chapter III, I.1), only a shift of 0.9 Å was observed in T166 (P9) hydroxyl side chains allowed by the absence of W145 bulky side chain (Figure 46). This shows the limits of using *in silico* mutagenesis tools in PyMol that does not enable to visualize the subsequent liberty given to the neighbor residues of a mutation. An additional step of energy minimization of the *in silico* mutated model might have helped to get closer to the structural reality. On the other hand, protein crystallization is also like taking a screen shot of the protein structure at a given moment even if it gives more relevant hints on the protein structure.

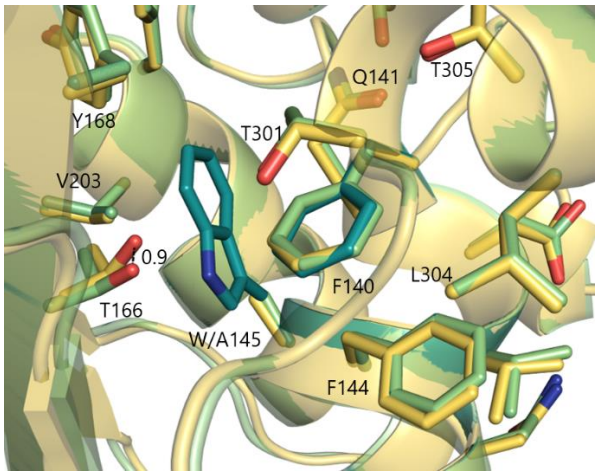


Figure 46. Structural alignment of *Cfus*AmDH (PDB: 6IAU, chain B; blue), *in silico* mutated *Cfus*AmDH-W145A (green) and *Cfus*AmDH-W145A structure obtained from crystallization (PDB: 7QZL, chain A; yellow). The distance is given in Angstroms (Å).

Notable differences were observed between the $NADP^+$- and NAD^+-mutant structures both in the active site and in the cofactor phosphate binding site. In the NAD^+ complex, the most striking difference locates in the active site with the side-chain rotation of 110-115° for Y202 (P17). In subunit A, this side chain could be modelled in two conformational rotamers at 50% occupancy each. In subunit B, only the rotated form was observed. This rotation led to the loss of the hydrogen bond between the E108 carboxylate moiety and the Y202 hydroxyl group. The W145A mutation has left space for Y202 side chain to rotate and take the space previously occupied by the indole ring of W145 (Figure 47A). In the

NADP$^+$ complex with **211**, the Y202 side chain is held in the WT position in both subunits, but the alkyl chain of the amine ligand extends into the space that was made by the W145A mutation (Figure 47B). This behavior was not obviously visible in the *in silico* mutated protein and it was therefore interesting to underline from the structures.

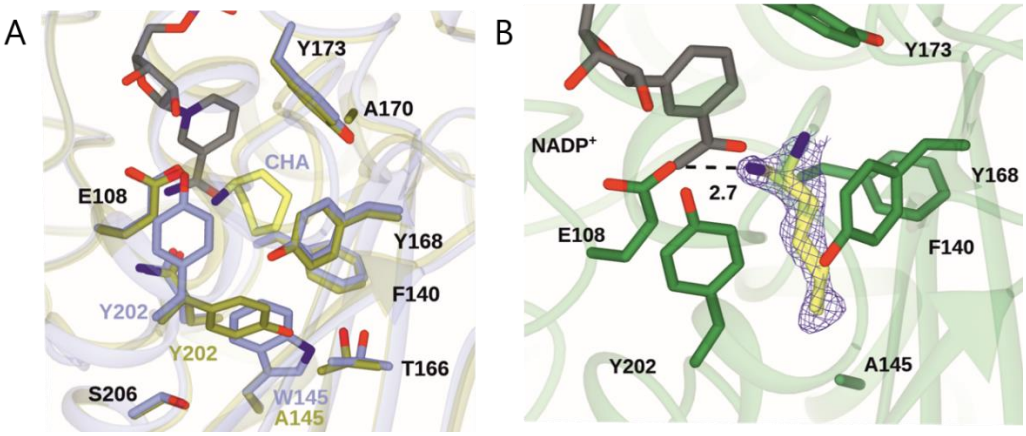


Figure 47. Active site of *Cfus*AmDH-W145A 3D-structure in complex with NADP$^+$/pentan-1-amine (**211**) and NAD$^+$. (A) Active sites of *Cfus*AmDH/NADP$^+$ complex with cyclohexylamine (**179**) (CHA, 6IAU, carbon atoms in blue) and *Cfus*AmDH-W145A/NAD$^+$ (carbon atoms in gold), showing rotation of Y202 side chain to fill space vacated by W154A mutation; (B) Active site of *Cfus*AmDH-W145A/NADP$^+$ complex with **211**, modelled in two conformations, each at 50% occupancy. Electron density corresponds to the F_o-F_c (omit) map at a level of 3σ obtained prior to building and refinement of the ligand. Interaction between the ligand N-atom and E108 side chain is shown as a black dashed line with the distance in Angstroms (Å).

Each of these results suggests that the space made by the mutation can result in flexible substrate specificity of this mutant through the combined effects of the mutation and the rotation of Y202. Compared to *in silico* docking experiments, **211** harbors a similar orientation of the carbon chain in the active site but with a different positioning in the pocket. Notably, a shift of the N-atom position from 0.7 to 2.2 Å can be observed (Figure 48).

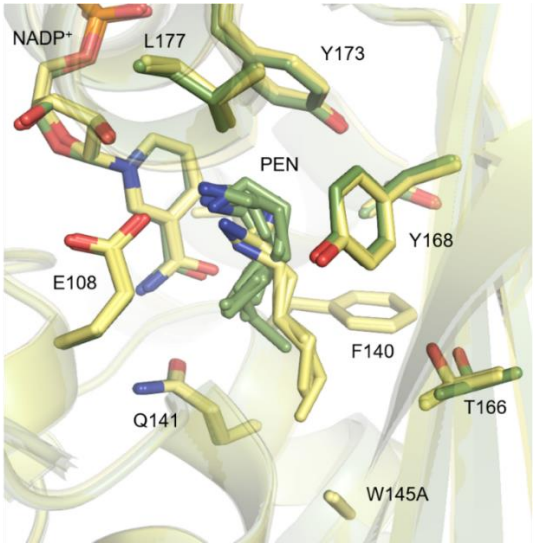


Figure 48. Pentan-1-amine (**211**) positioning when co-crystallized in *Cfus*AmDH-W145A 3D-structure or docked in modelled *Cfus*AmDH-W145A. Superimposition of active sites of *Cfus*AmDH-W145A/NADP$^+$ complexed to **211** (PDB: 7QZL, chain A and D, carbon atoms in yellow) and *Cfus*AmDH with *in silico* W145A mutation and the 10 most energetically favorable docked conformations of **211** (carbon atoms in green).

In terms of cofactor binding, the NADP$^+$ ribose 2′ phosphate in the W145A-NADP$^+$ complex makes interactions with the side chains of H37, N38 and a number of water molecules, while the 3′ OH interacts with D36 and R41, as observed for *Cfus*AmDH (Figure 49A) (Mayol *et al.*, 2019). When the NADP$^+$ phosphate is replaced by the simple 2′ hydroxyl in the NAD$^+$ complex, this group only makes direct contact with the side chain of H37, which moves slightly towards the cofactor (Figure 49B).

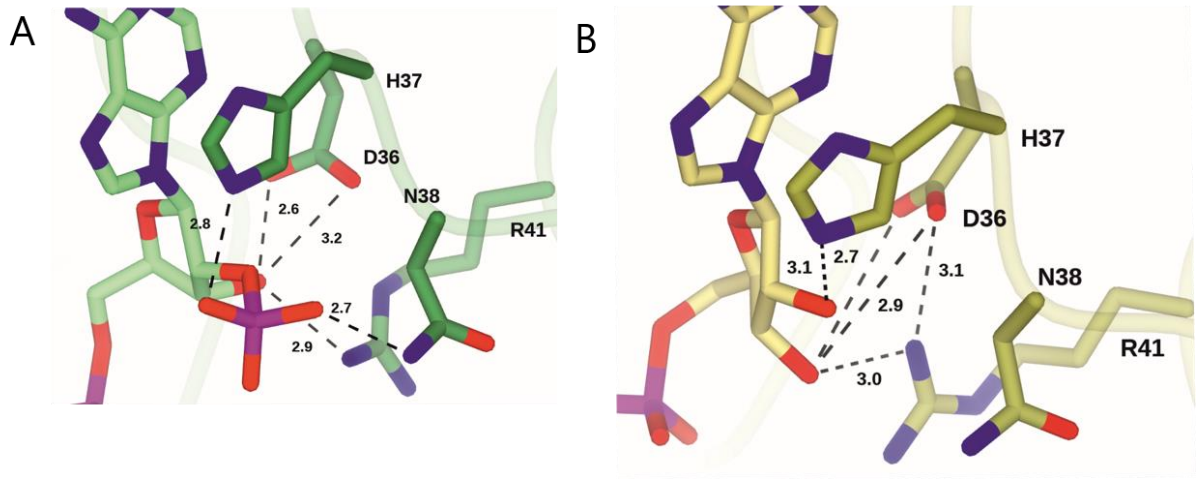


Figure 49. Cofactor binding site of *Cfus*AmDH-W145A 3D-structure in complex with NADP$^+$/pentan-1-amine (**211**) and NAD$^+$. (A) Cofactor binding pocket of *Cfus*AmDH-W145A/NADP$^+$ complex (B) Cofactor binding pocket of *Cfus*AmDH-W145A/NAD$^+$. The ionic interaction between D36 and R41 side chains in the NAD$^+$ complex is absent in the NADP$^+$ complex. Selected interactions are indicated with black dashed lines with the distance in Angstroms (Å).

In addition, the side chain of R41 displays a different orientation that introduces an ionic interaction of the NH1 atom with the side chain of D36 that is absent in the $NADP^+$ complex. The reduced number of interactions in the NAD^+-complex can perhaps reflect the overall preference of *Cfus*AmDH-W145A for $NADP^+$ over NAD^+ of 11:1.

A comparison of the active sites of $NADP^+$ and NAD^+ complexes with W145A does not offer a molecular clue for the apparent different cofactor specificities observed when using different substrates. However, the flexibility of the active site, enhanced by W145A substitution and evidenced by Y202 rotation, suggests a plasticity in substrate recognition that may give rise to variable results with different substrates under different conditions.

I.4.7.2 Molecular dynamics

In addition to X-ray crystallography, MD simulations were performed to understand the influence of W145A mutation on protein flexibility and specificity towards cofactors and substrates [Rajakumara *et al*., 2022; Wang *et al*., 2021; Wang *et al*., 2020 (A)]. As this expertise was not mastered in the biocatalytic team, I carried out this study in close collaboration with Gwenaëlle André-Leroux from MaIAGE unit at INRAE and with the help of Dr. Eddy Elisée from LABGeM (Genoscope). Accordingly, the study was carried out using GROMACS software (Abraham *et al*., 2015; Berendsen *et al*., 1995; Hess *et al*., 2008; Lindahl *et al*., 2001; Pronk *et al*., 2013; Van Der Spoel *et al*., 2005; Bekker *et al*., 1993; Lemkul, 2018) on an extensive set of Enzyme-Cofactor-Ligand systems where both *Cfus*AmDH-W145A (M) or *Cfus*AmDH (WT) were considered, in complex with NAD^+ (NAD) or $NADP^+$ (NADP) cofactors, and **211** ligand (PEN) or **213** ligand (HEP) ligands (Experimental section, VI.2.3.1.1). Exclusively in this paragraph, WT refers to wild-type *Cfus*AmDH only.

The analysis of all MD simulations leads to the following observations that deal with *(1)* the overall dynamics of the protein, *(2)* the affinity of M for larger substrates, *(3)* the mutated site W145A and its coordination sphere and *(4)* the adenosine binding site (Experimental section, VI.2.3.1.2).

I.4.7.2.1 Overall dynamics

M did not suffer from any major unfolding over the time of simulation, which confirms that the mutation W145A did not impact the global 3D fold of the protein as already evidenced by both the structural and experimental studies. The enzyme normal modes were neither affected by the mutation. The closed-to-open event, already discussed in Chapter II, I.1.1, and previously observed for the nat-AmDH family in AmDH4 (PDB: 6G1M) (Mayol *et al*., 2019) and more recently for MATOUAmDH2 (PDB: 7ZBO; 7R09) (Bennett *et al*., 2022), was observed for all the systems (Appendix 49). This occurrence was

revealed thanks to the Root Mean Square Fluctuation (RMSF), measured in Å for each full length protein (Figure 50).

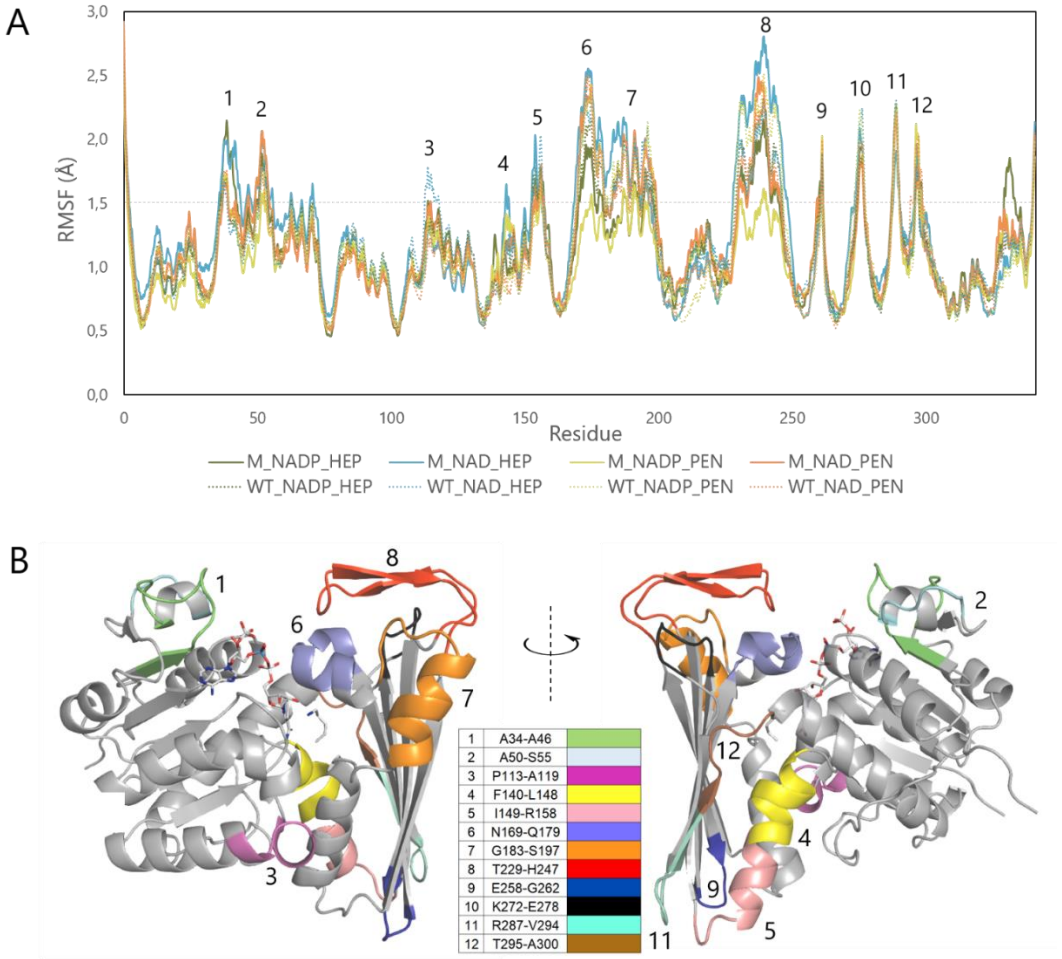


Figure 50. *Cfus*AmDH flexible sections. (A) RMSF analysis (B) 3D representation at 0° and 180° of WT monomer in complex with NAD and PEN. Segments 1-12 display a backbone fluctuation beyond 1.5 Å in RMSF analysis and are highlighted with a gradient of colors reported in the inserted table.

Particularly, this parameter identified segments N169-Q179 (6), G183-S197 (7) and T229-H247 (8) as highly flexible (Figure 50). Monitoring the G14-G174 distance, where residues G14 and G174 are facing each other and located on each side of the aperture, reveals that all the systems converge towards a so-called "relaxed" conformation opened at 14 Å (Figure 51 and Appendix 49). Interestingly, in M systems, a higher velocity of opening is observed when the protein is complexed to $NADP^+$ as compared to NAD^+. Indeed, with $NADP^+$ the systems reached the so-called "relaxed conformation within the first 10 ns, while NAD^+ systems took 30 and 50 ns with HEP and PEN ligands, respectively. Therefore, in contrast to WT complexes, the nature of the cofactor could influence the velocity of the loop opening in M complexes and so the k_{cat} of the system. It is also interesting to note that the opening event correlates with a torsion of the β-sheet domain located at the C-ter, and induces a substantial 11.9 Å shift of residue

Y173 at the active site. Such an amplitude was measured between the closest and widest conformations in the example of M_NADP_PEN_1 simulation (Appendix 49B).

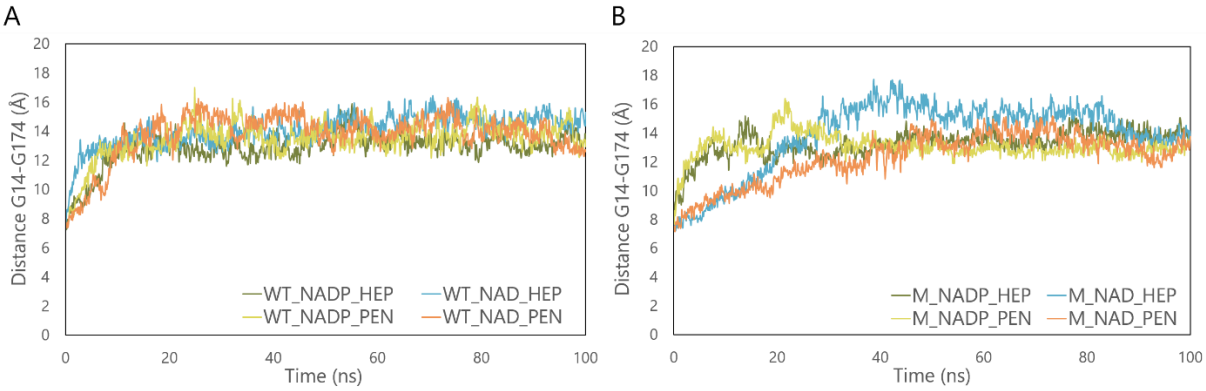


Figure 51. Evolution of G14-G174 distance (Å) as a function of time (ns) (A) WT in complex with NAD^+ (NAD) or $NADP^+$ (NADP) and PEN or HEP (B) Same curves with M. The curves are averaged onto four independent simulations starting with a random velocity seed.

I.4.7.2.2 Affinity of CfusAmDH-W145A for larger substrates

The Enzyme-Ligand interaction energies confirmed a better adaptation of HEP ligand for the mutated pocket, with a global 1.2-fold lower energy values as compared to WT, for both $NADP^+$ and NAD^+ (Figure 52). However, the calculated interaction energies for M_PEN systems, which were similar or lower when compared to those for WT_PEN, do not correlate with experimental data as M displays a loss of specific activity towards PEN.

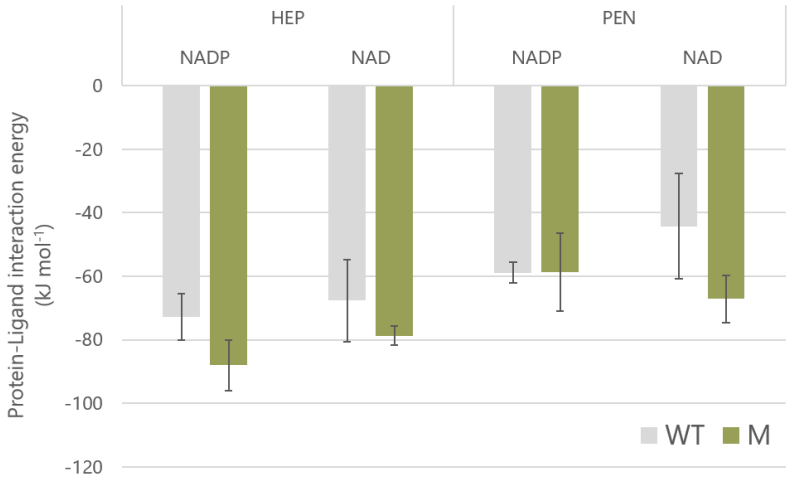


Figure 52. Interaction energies of Protein-Ligand calculated as the average of the total simulation. These energies, given in kJ mol^{-1}, were calculated by addition of short-range Coulombic interaction and Lennard-Jones interaction energies. Error bars represent the standard deviation of the four independent simulations. The grey and green bars correspond to WT and M respectively.

I.4.7.2.3 Coordination sphere of the mutated site

A fourteen-residue long helix, located between F140 and G154 of the β-sheet domain, and harboring the mutation site, is encompassed within the flexible segments 4 and 5 (Figure 50, Figure 53, Appendix 49 and Appendix 50). Notably, this helix is positioned within the core of the protein, where residues F140, F144, L148 and L152 protrude and align together. W145 creates a slight distortion to fit into the active site but this could be energetically compensated by the π-stacking between F140 and F144 (Appendix 50). All those features could act as a structural hydrophobic spine that maintains integrity of the system while allowing its flexibility. This section is also largely involved in the homodimerization.

MD suggests that this spine helix remains rather unchanged over the opening-closing cycles despite the particularly tensed twist at F140-G146, mainly due to the presence of W145 in WT. Consistently, W145A substitution could relax this tension and provide elasticity to this helix, as suggested by the RMSF around D142-G146 in the systems M_NAD_HEP and M_NADP_PEN, respectively calculated on average 0.21 Å and 0.39 Å higher than in the corresponding WT systems (Figure 53). In the other M systems, this parameter is more similar to the corresponding WT systems. Interestingly, the profile of M_NADP_HEP displays a higher flexibility than the other systems in the section S138-F140 located at the edge of the helix, close to the nicotinamide moiety of the $NADP^+$ cofactor. The rest of the helix I149-A154 gained flexibility especially in the system M_NAD_HEP, evidencing that the substrate participates to the protein flexibility during the course of catalysis.

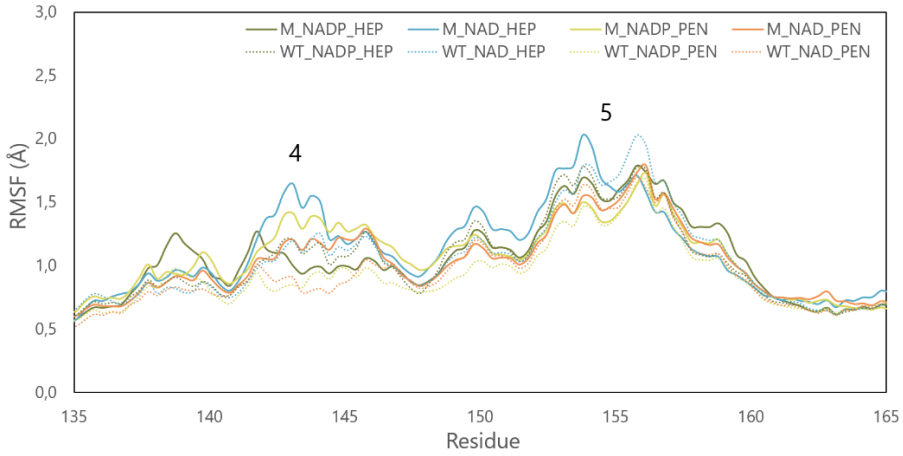


Figure 53. Close-view of segments 4 and 5 of the RMSF analysis of WT and M in complex with NAD or NADP and the amine products PEN or HEP. RMSF is given in Angstroms (Å).

Also in the coordination sphere of the mutated site, and particularly in the catalytic area, Q141 plays a key role in maintaining the nicotinamide moiety of the cofactor through hydrogen bonds between its oxygen and the amide (Figure 54 and Appendix 50). In WT, when Q141 loses its interaction

with NAD(P), it remains quite stable in its initial position while in M, Q141 becomes more mobile (Figure 54 and Appendix 51). Therefore, as the mutation has an impact on the fixation sphere of the cofactor through its alteration at Q141, it may affect the affinity of the protein towards its cofactor.

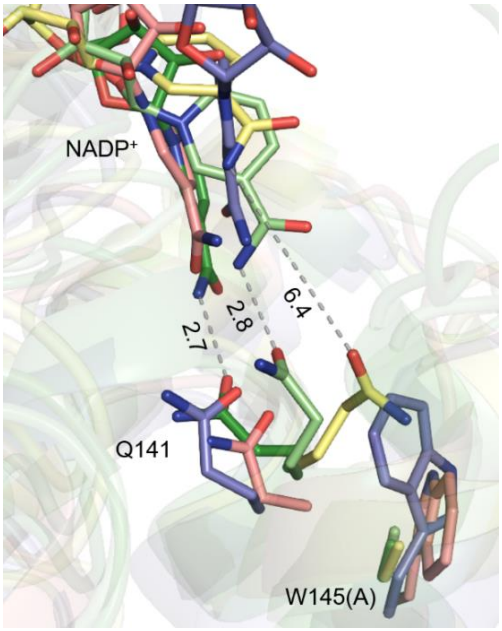


Figure 54. Close-view of Q141 in WT and M at different time of the dynamic. M_NAD_HEP_2 frames at 0.1, 16.2 and 18.5 ns are represented in dark green, light green and yellow, respectively. WT_NAD_HEP_4 at 29.0 ns and WT_NAD_HEP_1 at 53.2 ns are represented in purple and pink, respectively. Q141, A/W145 and NAD^+ (NAD) are shown in stick format.

This correlates with the average of 1.1-fold higher Enzyme-Cofactor interaction energies between WT and M, whichever cofactor and ligand were considered (Figure 55). Interestingly, Gln is the most conserved residue at P6 position among the nat-AmDH family, except for MATOUAmDH1 and 2 (Ile and Leu respectively), *Rgna*AmDH and 21 *Cfus*AmDH P1-P21 homologs (Pro), AmDH4 (N) and IGCAmDH6 and 23 *Cfus*AmDH homologs (Glu), A0A138ZYM0 (Ser), A0A078BAL0 (Tyr), A0A077ZQK8 and A0A078AU13 (Phe) which suggests different dynamic behaviors depending both on cofactors and ligands (Appendix 5).

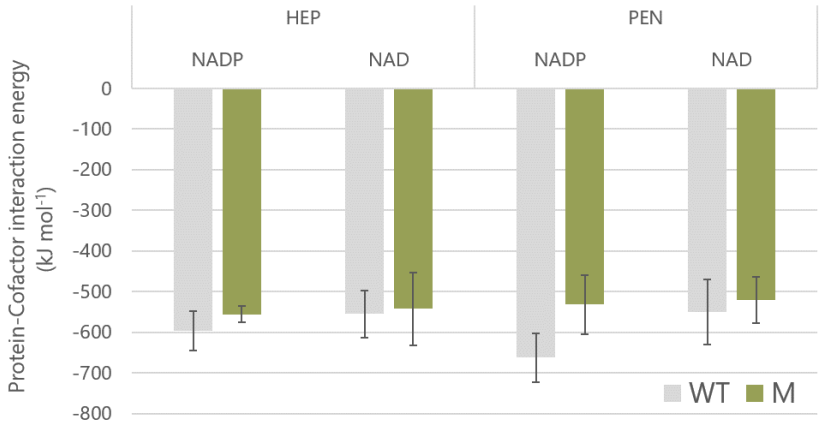


Figure 55. Interaction energies of Protein-Cofactor calculated as the average of the total simulation. These energies, given in kJ mol^{-1}, were calculated by addition of short-range Coulombic interaction and Lennard-Jones interaction energy. Error bars represent the standard deviation of the four independent simulations. The grey and green bars correspond to WT and M respectively.

I.4.7.2.4 Adenosine binding site

Considering the equilibrated systems and frames at 0 ns, the distances registered between H37 and adenine, as well as between D36 and ribose revealed a closer proximity of the binding site with the non-phosphorylated cofactor (Figure 56).

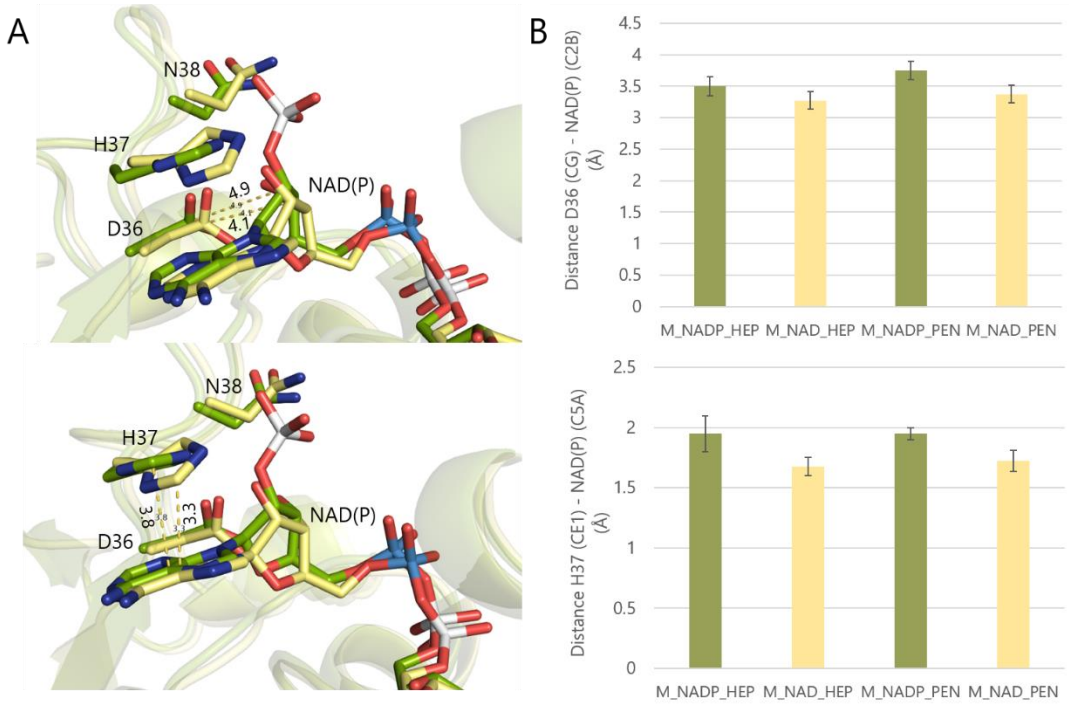


Figure 56. Distances between the adenosine moiety of the cofactor and the main residues of its binding site (A) Superposition on NAD^+ and $NADP^+$ in the adenosine binding site. D36, H37 and N38 are shown in cylinder format. (B) Comparison of the initial distance at t=0 ns between D36 (CG) and the ribose (C2B) (above) and H37 (CE1) and

adenosine (C5A) (below). The distances are given in Å. RMSD (M_NADP_HEP_1 – M_NAD_HEP_1) = 0.304 Å.

The steric hindrance brought about by PO_3^- towards N38 seems to induce a shift of 0.25 Å in the ribose positioning. In the case of the shorter substrate PEN, which does not occupy the entire mutated catalytic pocket, the global larger space left over might play a role. Either way, the additional PO_3^--N38 interaction appeared to increase the anchoring of the phosphorylated cofactor in the binding site, hence the affinity of the cofactor for the protein, as measured with an average of 1.73 and 2.28-fold better interaction energies at 0 ns compared to NAD systems, between the adenosine moiety and the protein in M and WT systems, respectively. This tendency remained over the whole course of the MD (Figure 57).

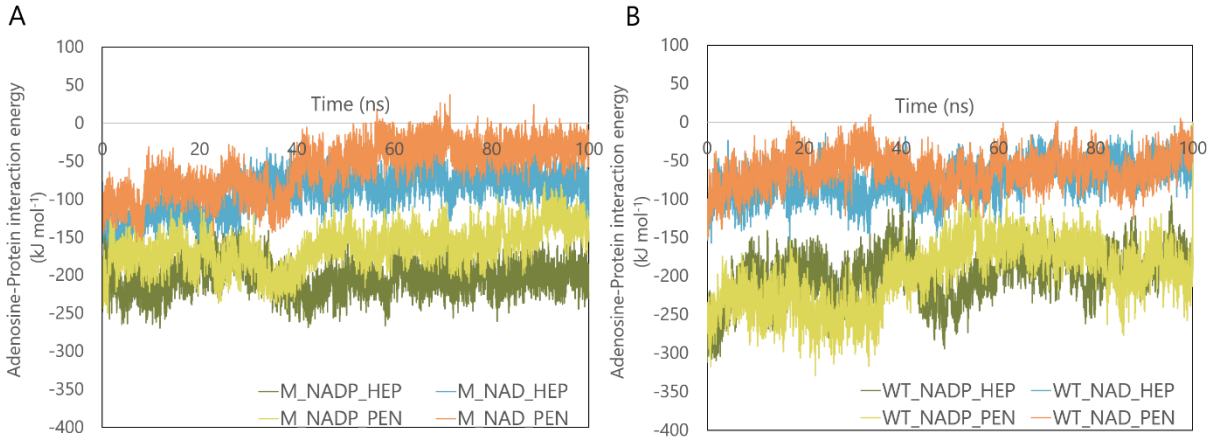


Figure 57. Interaction energies calculated between atoms of the adenosine moiety (C4B; C3B; C2B; O4B; C1B; O2B; N9A; N9A; N7A; C4A; C5A; N3A; C2A; N1A; C6A; N6A; O3B; ± P2B; ± O1X; ± O2X; ± O3X) and (A) M and (B) WT. These energies, given in kJ mol^{-1}, were calculated by addition of short-range Coulombic and Lennard-Jones interaction energies.

This is in accordance with the *in vitro* experiments displaying ten-fold lower K_M for M_NADP as compared to M_NAD systems (see Chapter II, I.4.6.1). Reversely, when tracing these distances along the simulation path, molecular explanations were difficult to rationalize for a higher mobility of the adenosine ring calculated in case of M_PEN systems as compared to M_HEP systems (Appendix 52). Several other residues can induce the mobility of the adenosine of both NAD and NADP (Appendix 53). Notably, a double planar π-stacking between H37, adenine and R83 can be seen in 81.3% of the simulations (68.8% with M and 93.8% with WT), with R83 also acting as the transporter of the adenine outside of its initial binding site, directing it towards the other half of the protein to interact with S175 and K180 for example (Appendix 54).

In addition to providing some answers to the initial questions justifying this study, MD has made it possible to highlight the overall behavior of these proteins and key residues to be considered during

future studies. Other dynamic studies should be undertaken to further improve knowledge of these proteins, in particular the entry and exit of substrates and the corresponding key hotspot residues.

II. *In vivo* directed evolution of amine dehydrogenases

Another strategy of nat-AmDHs improvement consisting in *in vivo* directed evolution was set up in collaboration with Dr. Madeleine Bouzon-Bloch and Dr. Volker Döring along with the strong participation of Valérie Delmas, Ivan Dubois and Nathalie Vega-Czarny from the team of Synthetic Metabolism from L2BMS.

The *in vivo* evolution takes benefit of the process of natural selection that results in the appearance of spontaneous mutations in a bacterial community. This population grows in given conditions and is put under selection pressure where only the cells containing beneficial mutations are able to develop. It is even possible to direct the evolution by introducing a constraint in the culture medium and force the bacteria to express a specific feature to adapt to this new environment. With the aim of improving an enzymatic activity, the approach requires to genetically modify a strain to make it dependent to this activity, hence the strain needs to express the enzyme to grow. One of the main advantages of such an approach is that it does not require prior knowledge on the enzyme structure or mechanism. Also, there is no need to produce and/or test *in vitro* thousands of variants as with protein engineering. The major drawback is that the cells can evolve by a different way than a modification inducing a "naturally" engineered enzyme, such as modification of its expression or the awakening of a metabolic pathway, even if preliminary engineered precautions can be made to avoid it.

II.1 Principle of *in vivo* directed evolution using the automatic devices available at the Genoscope

In vivo directed evolution have been used since decades in non-automated systems through serial transfers into simple flasks or gel plates (Barrick & Lenski, 2013). However, such experiments are extremely time consuming and automatic devices have been developed to maintain a continuous evolution of the cells. At the Genoscope, the automated evolution devices GENEMAT (version GM3), renamed as GM3 in this manuscript, is a technology developed by the French Alternative Energies and Atomic Energy Commission (CEA) in collaboration with ALTAR. GM3 is a fluidic self-cleaning cultivation device that enables the maintenance of cells cultures over an extended period of time while preventing biofilm formation (Mutzel & Marliere, 2004). Traditionally, preliminary tests are carried out to attest the feasibility of the selection pressure to be used for evolution. These are done using strains containing the gene to be involved in a plasmidic construction but eventually, the gene is introduced in the strain

chromosome before inoculation in the device to avoid any possible rejection during the evolution. Two growth vessels are alternatively used for culture growth and sterilization with NaOH 5 M (Appendix 55). The dilutions with fresh medium are controlled by an automated sluice gate system. The bacterial growth is continuously monitored by the measurement of the culture transmittance (1/absorbance) and culture samples are regularly taken out and backuped at -80°C. Genomic sequencing enables to deduce the mutations that were beneficial for the evolution. The adaptable parameters that can impact the culture growth are the temperature, the stirring provided by an sterile air flow, oxygen concentration, recurrence of washing cycles and, depending on the regime used, the transmittance threshold, the number of dilutions, etc. Three regimes are possible with this system: chemostat, turbidostat and medium swap (Appendix 56).

II.1.1 Chemostat regime

In a chemostat regime, cells are grown in a constant volume of medium, regularly diluted by a fresh medium inflow while a corresponding volume of old medium is discarded (Appendix 56). The growth medium provides all nutrients necessary for the cells growth in excess except the limiting one(s) (Monod, 1950; Novick & Szilard, 1950). If the bacteria assimilate this limiting substrate(s) more rapidly, its growth speeds up between two dilutions resulting in a progressive decreasing of the transmittance. The evolved bacterial population is therefore more efficient to assimilate the target substrate(s).

II.1.2 Turbidostat regime

Unlike the chemostat, the thermostat regime is not limited by a nutrient in the growth medium (Appendix 56). The dilutions do not occur at fixed intervals but are triggered by the culture growth. The culture is supplied in fresh medium when the cell density exceeds a transmittance threshold fixed by the operator. The increasing of the dilutions frequency indicates that the bacterial population is adapting to the given medium. This regime is widely used to enhance the generation time of a strain that can already grow in a given medium.

II.1.3 Medium swap regime

This last regime has been conceived for complicated evolutions that are not possible with chemostat or turbidostat regimes, typically, when the non-adapted strain displays very low initial activity required to overcome the selection pressure (Appendix 56). In medium swap regime, the culture is diluted at fixed intervals of time by alternation of two media. The relaxing or permissive medium enables the growth of the non-adapted cells while the stressing medium is the target for the selection of the culture in which the non-adapted cells cannot grow. A turbidostat regime in the relaxing medium generally precedes a medium swap regime to condition the strains and fix its generation time. This latter

information is used to define the transmittance threshold that triggers the dilution medium. If the transmittance measured before the dilution is above the threshold, the culture is supplied in relaxing medium while if it is below the threshold, the culture is diluted with stressing medium. Hence, the culture adaptation is visible when the frequency of stressing medium selection for the dilutions increases. When the selected stressing medium represents a too big gap, intermediate evolutions with increasing proportion of stressful medium are carried out. For instance, this was done by the synthetic metabolism team of L2BMS for the adaptation of the methylotrophic *Methylobacterium extorquens* AM1 to make it accept high CH_3OH concentrations (Belkhelfa *et al*., 2019).

Within the scope of the AmDH project, *in vivo* directed evolution using GM3 has been attempted sporadically for more than six years with the goal to provide a proof of concept that such a strategy can also improve the reductive aminase activity of some nat-AmDHs. Nevertheless, GM3, and *in vivo* directed evolution in general, are much more designed to improve strains or pathways but not enzymes for isolated biocatalytic applications. To do so, the strategy aims at obtaining mutations within the selected enzyme to afford higher activity towards specific substrates. In the team, there was an only one-time attempt for the improvement of an α-ketoglutarate dependent dioxygenase. In this case, the enzyme was part of the metabolic pathway to be modified and not a heterologous enzyme providing a non-metabolic activity for *E. coli* as for nat-AmDHs (Souterre, 2017). Some experiments have been done before my involvement in this project to test this strategy for nat-AmDHs. Based on these preliminary tests, part of my PhD work was dedicated to validate or not the possibility to modify nat-AmDH activity by *in vivo* evolution using GM3. Two different approaches have been undertaken that are described in the following paragraphs.

II.2 Selection screen based on NH_3 release

II.2.1 Context of the *in vivo* evolution of amine dehydrogenases

The first phases of this project were conceived and carried out during the PhD project of Ombeline Mayol and need to be detailed here to understand the following steps that we focused on during the time of my thesis project. Within the scope of nat-AmDHs activity improvement through *in vivo* directed evolution, the *E. coli* strain, must be dependent of the nat-AmDH activity to survive. Therefore, when cultivated in the selective medium, the cells were forced to express an improved version of a nat-AmDH to grow. Considering the reverse reaction of oxidative deamination of an amine substrate, the first selection pressure was based on the release of **11** that becomes the only nitrogen source of the medium, required for cells growth (Figure 58). This goes with the hypothesis that the improved activity in the oxidation way induces an improved activity also in the reverse targeted reductive way. The first objective was to improve the nat-AmDH activity towards (4*S*)-4-oxopentanoate (**152**) using

AmDH5 from *Clostridium Sticklandii* (PDB: E3PY99), a non-thermophilic enzyme that can be active at 37°C, ideal for cells growth (Mayol *et al.*, 2016). Besides, AmDH5 exhibits a high oxidative deamination activity, as it is its native reaction direction in *Clostridium Sticklandii* metabolism. This choice was also guided considering that the substrates and products of 4OP-AmDHs are hydrophilic thanks to their carboxylic moiety, which therefore minimizes the problems related to the crossing through the cell membrane. The gene coding for AmDH5 was introduced in the genome of an *E. coli* K12 strain deleted of three genes coding for two TAs (puuE and gapT) and one diaminoacide oxidase (dadA). These enzymes could be mutated during the evolution process and accept (4*S*)-4-aminopentanoate (**167**), thus preventing AmDH5 to evolve. First, it was considered to cultivate this strain in a medium swap regime with (2*R*,4*S*)-2,4-diaminopentanoate (**165**), AmDH5 natural substrate, as the relaxing medium, and the target **167**, as the stressing medium (Figure 58). In both media, the amine substrates was the only nitrogen supply. Deamination activity of AmDH5 towards **165** was reported by Fonknechten *et al.* with v_{max} = 51.6 s^{-1} (Fonknechten *et al.*, 2009). It would have been better to also record it that of **167** at an early stage of the project to design the strategies with more experimental bases.

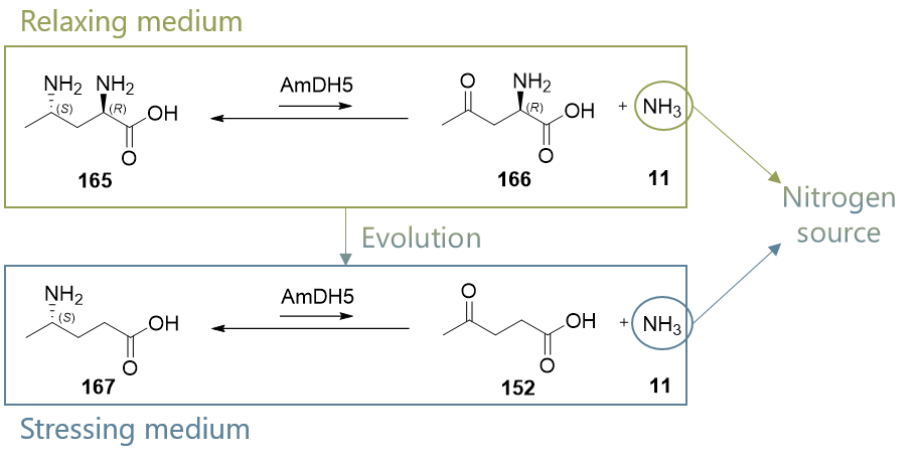


Figure 58. First selection system for the *in vivo* directed evolution of AmDH5.

The constructed strain G4488, containing the gene coding for AmDH5 in its chromosome (Appendix 57), was not able to grow even in the relaxing medium, the evolution was designed in two steps. A medium swap regime with NH_4Cl/**165** as relaxing/stressing medium to adapt G4488 to **165** was followed by a second step with **165**/**167** as relaxing/stressing medium for evolution towards **167**. Unfortunately, due to the high cost and difficulties to synthesize a sufficient amount of **165**, the evolution could not be long enough to confirm an evolution of AmDH5 and the project was put on stand-by (Mayol, 2019).

All the strains genotype and constructions mentioned in this part are summarized in Appendix 57.

II.2.2 Evolution system using *D*-ornithine as starting substrate

At the beginning of my PhD project, the selection system had been redesigned to include the Ornithine Aminomutase from *Clostridium Sticklandii* (OA) to produce **165** *in situ* from *D*-ornithine (**242**) (Figure 59). The goal was first to improve the activity of the enzyme towards **167** and then reach activity towards carbonyl substrates devoid of any carboxylic moiety. The latter being more ambitious due to the poor deamination activity obtained with any nat-AmDH towards these substrates but the cells do not require high production of NH_3 to grow so it was still considered accessible. *Tther*AmDH was added as another nat-AmDH that could be evolved due to its structural similarity with the set of 4OP-AmDHs (Mayol *et al.*, 2016, Caparco *et al.*, 2020). It was also decided to first attempt with enzymes in plasmidic constructions and then to move to chromosomic constructions in case of success. General protocols used for gene cloning and subcloning are given in Experimental section, V.1 and V.2. Beforehand, G4480 was cultivated in GM3 devices for 3 months in medium containing a not sufficient supply in NH_4Cl (2 mM) and in **242** (10 mM) to check that no enzyme would wake up during an evolution, become able to deaminate **242** and so provide nitrogen source.

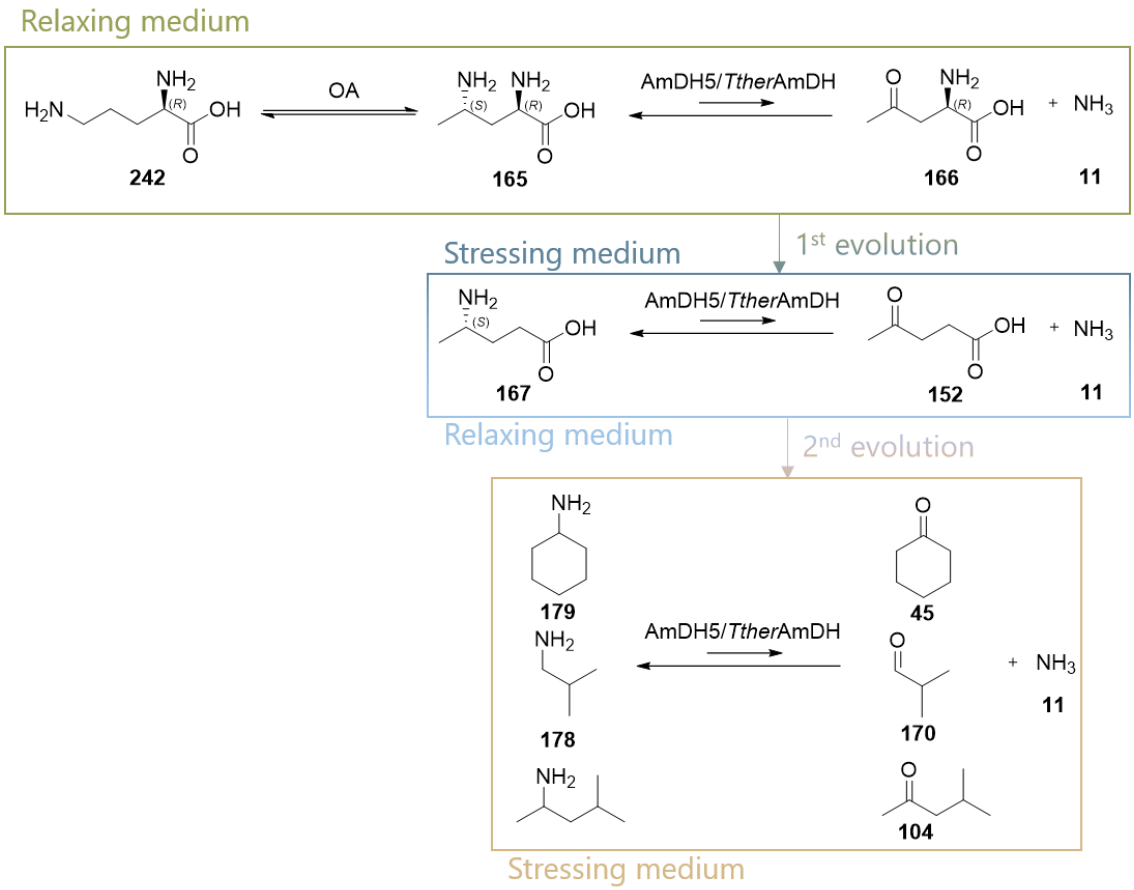


Figure 59. Second selection system for the *in vivo* directed evolution of AmDH5 and *Tther*AmDH.

With the great help of Valérie Delmas from Synthetic Metabolism team, the genes coding for AmDH5, *Tther*AmDH and OA, comprising two subunits oraS-oraE, were separately introduced in two different plasmids: pSP100 (high copy number 20-700) and pVDM18 (low copy number 5-15) (Appendix 58). These vectors are not inducible but constitutive to constantly afford the required expression level. These constructions were then transformed in *E. coli* K12 strains whom genetic context is now well known. The results of the liquid cells growth assays, monitored through Optical Density (OD) measurement in medium containing the relaxing NH_4Cl, the intermediate **165** or the target **242** as the only nitrogen source are given in Table 8 (Experimental section, V.3 and V.4 for general methods for liquid cells growth and bioscreen assays). Some of these culture growths were also monitored through bioscreen to gain more insights into the dynamic profile of the growth (lag phase, growth rate, OD of the stationary phase, etc.).

Table 8. Summary of liquid cells growth and bioscreen results of the different strains created in the three source of nitrogen discussed in this part. These results merge several assays to give a general tendency obtained from these strains simplified in a code where "++" corresponds to a growth similar as in NH_4Cl-containing medium, "+" to still substantial but highly decreased growth, "-" to no growth. Growth conditions: MSN^- medium, glucose (**222**) 0.2%, [NH_4Cl]=5 mM, ±[**165**]=1-5-10 mM, ±[**242**]=5 mM, 37°, 150-180 rpm. The composition of all the medium used in the paragraph is detailed in Appendix 59. [a] Inhibition is observed with addition of *D*-ornithine (**242**). *n.t.* not tested

Entry	Strain	Gene	Vector	Nitrogen source		
				NH_4Cl	**165**	**242**
1	G4480	/	/	++[a]	-	-
2	G4265	AmDH5	pSP100	++	++	*n.t.*
3	G5376	AmDH5	pVDM18	++	+	*n.t.*
4	G5393	*Tther* AmDH	pSP100	++	++	*n.t.*
5	G5378	*Tther* AmDH	pVDM18	++	-	*n.t.*
6	G5392	OA	pSP100	++[a]	*n.t.*	-
7	G5377	OA	pVDM18	++[a]	*n.t.*	-
8	G5379	AmDH5 OA	pSP100 pVDM18	++[a]	+	-
9	G5394	AmDH5 OA	pVDM18 pSP100	++[a]	*n.t.*	-
10	G5492	AmDH5 OA	pSP100	++	+	-
11	G5490	AmDH5 OA	chromosome + mutation in kdgk promotor pSP100	++	+	-

Unlike with G4488 containing the gene of AmDH5 in the chromosome, we could observe cells growth in **165**-containing medium with both AmDH5 and *Tther*AmDH as plasmidic constructions (entry 2, 3, 4, 5 in Table 8). The slightly more favorable growth of strains containing AmDH5 (G4265 and G5376) rather than *Tther*AmDH (G5393 and G5378) in presence of **165** drove the selection of AmDH5 for the double transformations with OA (G5379 and G5394) to proceed to the whole reaction cascade from its natural substrate **242**. While the strains containing AmDH5 can grow on **165**-containing medium, the

addition of the OA did not enable to observe cells growth on **242**-containing medium when OA was introduced as an additional plasmidic construction or in a double cloning with AmDH5 gene in pSP100 (entry 8, 9, 10 in Table 8). This can be due to a poor expression level of OA in both vectors or to an inactivity in this medium. Indeed, OA normally works under anaerobic conditions in presence of vitamin B12 and pyridoxal phosphate (PLP, **243**) but the heterologous expression has already been successfully carried out in *E. coli* BL21 (Fonknechten *et al*., 2009). Even with **243** and vitamin B12 supplementation and under more strict anaerobic conditions, no significant growth could be observed. A range of **242** concentrations was tested from 5 mM to 20 mM but even such higher concentrations did not enable to improve the cells growth. With the hypothesis of a bad folding of OA in *E. coli* K12 cells, we also transformed AmDH5-pVDM18+OA-pSP100 and AmDH5-pSP100+OA-pVDM18 plasmidic constructions in BL21 [Novagen HT96 BL21 (DE3) competent cells], instead of K12, to afford a different genetic context that could enable a higher expression rate of the two enzymes, but without success. Another hypothesis to explain these failures was the high ratio of AGA codon in OA nucleic sequence that is not well adapted for expression in *E. coli*. However, as mentioned above, the heterologous expression of OA was already done successfully in BL21 strains without prior codon optimization of the gene (Fonknechten *et al*., 2009).

Another approach was to use the strain containing AmDH5 in the chromosome (G4488 already discussed earlier) and to employ the pORTMAGE technique to introduce a mutation in the kdgK promoter that was shown to enhance the expression level of the gene downstream (Bonde *et al*., 2014; Hermes *et al*., 1989; Nyerges *et al*., 2018; Nyerges *et al*., 2016; Sawitzke *et al*., 2011; Wang *et al*., 2009). This was supposed to overcome the limit of having a single copy of the gene in the chromosome. The pSP100 vector with the gene coding for OA was transformed in this strain. The resulting construction could still grow on a **165**-containing medium but not in presence of **242** as the only nitrogen source (entre 12 in Table 8).

II.2.3 Evolution system using 4-aminopentanoate as starting substrate

II.2.3.1.1 Application for the evolution of AmDH5 and Tther*AmDH*

Given the failures to get a system that is efficient enough to enable the bacterial growth on a medium solely containing **242** as the nitrogen source, we reconsidered the use of **167** directly. This compound, available at a more reasonable price than **165**, was not assimilated in the first strains containing AmDH5 in its chromosome (G4488) but the new constructions might give different results. The strains G5378, G5393, G5469, G5471 were cultivated on a medium containing either the racemic form **153+167** (4-AP) or the (*S*)-enantiomer **167** and compared to the empty G4480. A result summary of the culture growths is given in Table 9.

Table 9. Summary of liquid cells growth results of the different strains created in presence of **153** and 4-AP as nitrogen source. * Except G42465, all the strains mentioned harbor this excision. These results merge several assays to give a general tendency obtained from these strains simplified in a code where "++" corresponds to a growth similar as in NH_4Cl-containing medium, "+" to still substantial but highly decreased growth, "-" to no growth. Growth conditions: MAN^- medium, **222** 0.2%, ±[NH_4Cl]=5 mM, ±[**167**±**153**]=5-10-20 mM, 37°C, 150-180 rpm. G5471 is the same construction as G42465 but devoid of the kan and cat resistance cassettes.

Strain	Gene	Construction	Nitrogen source	
			NH_4Cl	**167**±**153**
G4480	/	/	++	+
G5471*	AmDH5	pSP100	++	+
G5393	*Tther* AmDH	pSP100	++	+
G5378	*Tther* AmDH	pVDM18	++	+
G5469	AmDH5	chromosome with mutation in the kdgK promotor	++	+

Surprisingly, all the strains, including G4480, displayed a low but clear growth depending upon **167**±**153** concentration. Contradicting previous results, some complementary tests have been carried out. They revealed neither bacterial contamination in the **167**±**153**. However, a contamination of ammonia **11** in the 4-AP commercial solution and/or in the synthesized **167** was hypothesized. We could not provide an exact ratio of **11** but estimated it at less than 2%. Subsequent liquid growth tests in presence of this corresponding contamination showed clear growth at equivalent concentration of NH_4Cl but the different associations of substrates did not give clear results to rule out this hypothesis (Appendix 60). To have more accurate data on 4-AP assimilation, G4480 was cultivated on a (4-AP)-containing MSN^- medium and sampled regularly. The monitoring of 4-AP by UHPLC-MS after centrifugation and derivatization with BzCl revealed no clear reduction of this compound concentration in the culture medium, thus meaning no or less than 0.05 mM assimilation by G4480 (Appendix 61). However, we do not know if the sensitivity of the test was sufficient enough to explain the slight growth of the strain; traces of **11** in all the components being difficult to check.

II.2.3.1.2 Application for the evolution of AmDH4-I80T/P224S/E296G

In 2020, Cai *et al*. reported the mutant AmDH4-I80T/P224S/E296G resulting in a 6.7-fold higher specific activity and 18-fold higher catalytic efficiency towards 4-oxopentanoate (**152**) compared to the native enzyme (Cai *et al*., 2020). This mutant was produced and tested for its specific activity towards **152** and 4-AP in both ways of reductive amination and oxidative deamination, respectively (Figure 60, Experimental section III.1 and III.6.1).

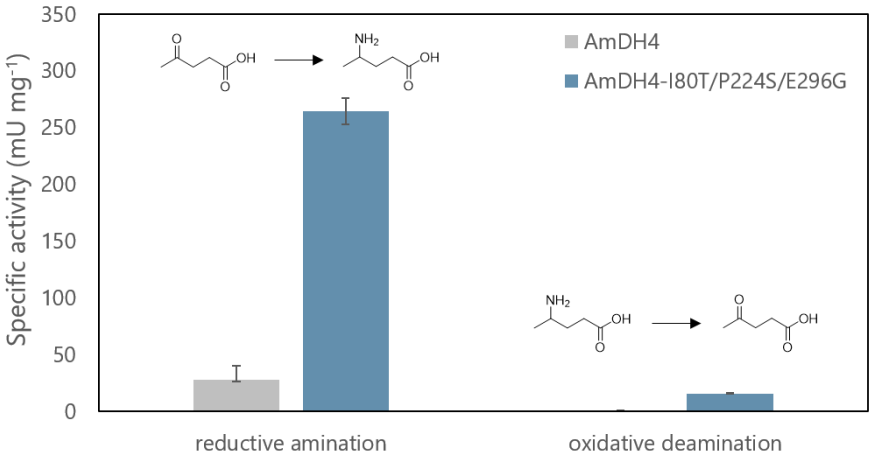


Figure 60. Specific activity (mU mg^{-1}) of AmDH4-I80T/P224S/E296G towards **152** and 4-AP in both ways of reaction. Reaction conditions for the reductive amination were as follows: final volume 100 µL, 10 mM **152**, 1 M NH_4HCO_2 buffer pH 8.5, 0.4 mM NAD(P)H, 0.1-0.3 mg mL^{-1} purified enzyme, 30°C. Reaction conditions for the oxidative deamination were as follows: final volume 100 µL, 10 mM 4-AP, 50 mM Tris HCl buffer pH 9.0, 0.4 mM NAD(P)H, 0.2-0.5 mg mL^{-1} purified enzyme, 30°C. Error bars represent the standard deviation of two experiments.

As expected, AmDH4-I80T/P224S/E296G exhibited 9.5-fold higher activity for the reductive amination of **152** than the WT (265 vs 28 mU mg^{-1}). On the other hand, its specific activity for the oxidative deamination of 4-AP remained low (16 mU mg^{-1}) but significantly higher than the native enzyme estimated at <1 mU mg^{-1}. These results encouraged us to clone the gene of AmDH4 mutant in a pSP100 vector and transform the construction into *E. coli* K12 strain G4480 for cells culture growth on (4-AP)-containing medium. However, once again, the growth rate was similar to that of the empty G4480 meaning that there was no improvement in 4-AP assimilation to produce **11**, required for cells growth (Table 10).

Table 10. Liquid cells growth results of the different strains containing AmDH4 and AmDH4-I80T/P224S/E296G in presence of 4-AP as nitrogen source. These results merge several assays to give a general tendency obtained from these strains simplified in a code where "++" corresponds to a growth similar as in NH_4Cl-containing medium, "+" to still substantial but highly decreased growth, "-" to no growth. Growth conditions: MSN^- medium, **222** 0.2%, ±[NH_4Cl]=10 mM, ±[4-AP]=10 mM, 37°C, 150-180 rpm.

Strain	Gene	Vector	Nitrogen source		
			/	NH_4Cl	4-AP
G4480	/	/	-	++	+
G5618	AmDH4	pSP100	-	++	+
G5592	AmDH4-I80T/P224S/E296G	pSP100	-	++	+

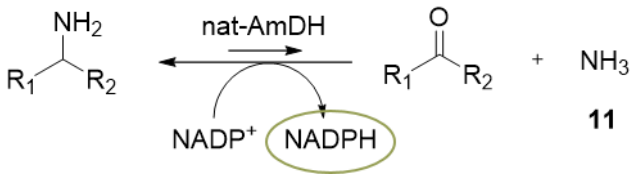
As side project for understanding, G4480 was cultivated for evolution in the GM3 in order to determine the origin of its ability to grow in presence of 4-AP as the only nitrogen source. A culture of this strain was inoculated and cultivated in turbidostat regime in permissive medium (MS base in absence of NH_4Cl

and in presence of 10 mM 4-AP and 0.2% **222**) at 37°C. The culture was not able to densify even with changes in the initial DO of the inoculation or the dilution threshold. These observations did not enable any conclusion regarding the growth of G4480 in such a medium.

The many attempts on using the selection pressure based on release of **11** for cells growth proved it too stringent for the poor deamination activity of the 4OP-AmDHs. The gap of activity between **165** and 4-AP appeared to be too important. For a proof of concept, it could still be interesting to carry out this strategy with substrates with closer affinities towards the selected enzyme. Therefore, this system was abandoned and our work focused on another selection strategy.

II.3 Selection screen based on NADPH release

In a collaboration with Arren Bar-Even from the Max Planck Institute, the team of Synthetic Metabolism of the Genoscope reported the construction of a NADPH-auxotroph *E. coli* strain (Bouzon *et al.*, 2021; Lindner *et al.*, 2018). This strain genome is deleted with all the genes corresponding to a NADPH-producing reaction except that of a 6-phosphogluconate dehydrogenase (Appendix 62). Accordingly, gluconate (**244**) works as the NADPH source when supplied in the medium and in its absence, the strain becomes a biosensor to reveal the ability of a given enzyme to produce NADPH. In our case, it can be used as a selection pressure for nat-AmDH activity evolution as the cells will be able to grow only if the reaction occurs in the oxidative deamination reaction to produce NADPH from $NADP^+$ (Scheme 32). This system is supposed to be less demanding than the previous one in terms of NADPH amount required for cells growth.



Scheme 32. Selection system for the *in vivo* directed evolution of nat-AmDH based on NADPH release.

In this part, the strain used, G5313, contains the additional deletions ΔmaeA+Δlpd (Bouzon *et al.*, 2021).

II.3.1 Application for the evolution of *Micro*AmDH

Among the nat-AmDHs active towards simple ketones in the reductive amination way and with clear preference for the phosphorylated nicotinamide cofactor, *Micro*AmDH was already available in a pVDM18 vector construction that was transformed in G5313 strains (G5641). We first measured *Micro*AmDH for oxidative deamination activity towards simple and cheap amine substrates that could

be used for the evolution (Appendix 63). The top three substrates, 3-methylbutan-2-amine (**245**), cyclohexylamine (**179**) and pentan-3-amine (**246**) displayed rather low specific activity of 20.5, 17.4 and 15.8 mU mg^{-1}, respectively. G5641 was grown on a MA medium supplemented in 10 mM glycerol (**247**), 10 mM acetate (**248**) and 5 mM α-ketoglutarate (**249**) and in presence or absence of 10 mM **244** and/or 10 mM amine substrate (Table 11).

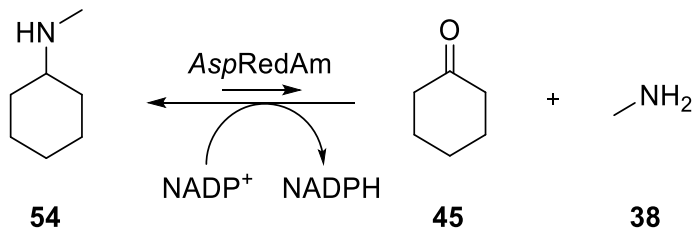
Table 11. Growth of G5313 and G5641 in presence of gluconate (**244**) and/or amine substrate. Growth conditions: MA medium, 10 mM **247**, 10 mM **248** and 5 mM **249**, ±10 mM **244** and ±10 mM amine substrate **245**, **179** or **246**, 30°C, 150-180 rpm. These results merge several assays to give a general tendency obtained from these strains simplified in a code where "++" corresponds to a growth similar as in **244**-containing medium, "+" to still substantial but highly decreased growth, "-" to no growth.

Strain	Gene	Construction	NADPH source							
			/		**246**		**179**		**245**	
			/	**244**	/	**244**	/	**244**	/	**244**
G5313	/	/	-	++	-	++	-	++	-	++
G5641	*Micro* AmDH	pVDM18	-	++	-	++	-	++	-	++

No growth was observed when the strain was supplied with only the amine substrates in absence of **244**. On the other hand, no inhibition of the amine tested was observed. Hence, even in these conditions, the nat-AmDHs are definitely not enough active in the required way of reaction to start the evolution in the GM3 which still requires traces of growth to be used.

II.3.2 Application for the evolution of *Asp*RedAm

We decided to search for another enzyme reported in the literature as much more active in the oxidative deamination direction of a widely available and cheap amine. Among all the IREDs, RedAms and AmDHs engineered from AADHs available at this time, *Asp*RedAm was particularly well described for this reaction towards a range of secondary amines (Aleku *et al.*, 2018). Thus, we started a collaboration with Prof. Nicholas Turner and one of his PhD student, James Marshall to attempt the *in vivo* evolution on *Asp*RedAm. The substrate selected was the *N*-methylcyclohexylamine (**54**) due to its affordability and its reasonable size required to enter the cells membrane (Scheme 33).



Scheme 33. Selection system for the *in vivo* directed evolution of *Asp*RedAm based on NADPH release upon deamination of *N*-methylcyclohexylamine (**54**).

The cloning of *Asp*RedAm into pSP100 and pVDM18 was carried out by Valérie Delmas, and the constructions were subsequently transformed into G5313 (G5648 and G5647, respectively). Several cells growth assays and bioscreens, revealed no particular substrate inhibition but also no growth of these strains on **54**-containing medium in absence of **244** (Table 12 and Appendix 64). Notably, the *Asp*RedAm cloned in the high copy number vector pSP100 led to a lower growth suggesting an inhibition induced by *Asp*RedAm expression.

Table 12. Growth of G5313, G5647 and G5648 in presence of gluconate and/or *N*-methylcyclohexylamine. Growth conditions: MA medium, 10 mM **247**, 10 mM **248** and 5 mM **249**, ± 1-10 mM **244** and ± 10-60 mM **54**, 30°C, 150-180 rpm. These results merge several assays to give a general tendency obtained from these strains simplified in a code where "++" corresponds to a growth similar as in gluconate-containing medium, "+" to still substantial but highly decreased growth, "-" to no growth.

Strain	Gene	Vector	NADPH source								
			54	/	/	/	10 mM	30 mM	60 mM	10 mM	30 mM	60 mM
			244	/	1 mM	10 mM	/	/	/	10 mM	10 mM	10 mM
G5313	/	/		-	+	++	-	-	-	++	++	++
G5647	*Asp* RedAm	pVDM18		-	+	++	-	-	-	++	++	++
G5648	*Asp* RedAm	pSP100		-	+	+	-	-	-	+	+	+

Even though there was no growth observed in the sole presence of the target amine substrate, we attempted an evolution in the GM3 in a medium swap regime. The relaxing medium included both **244** and **54** while the stressing medium only contained **54** as NADPH source. To avoid the risk of a plasmid rejection throughout the evolution, the gene coding for *Asp*RedAm was integrated in G5313 chromosom by Valérie Delmas (G5780). She also attempted to add the already known C→T mutation in the kdgK promotor to enhance the *Asp*RedAm expression level but without success even after 4 cycles in an intermediate strain (G5782) that does not harbor all the deletions present in G5780. Culture cells growth of this strain on a medium containing 5 mM **244** and 0 to 60 mM **54** revealed an inhibition especially at 60 mM **54** and slightly visible with a lag phase at only 30 mM (Figure 61). Working with up to 60 mM amine concentration was tested to potentially provide the cell with more NADPH at a specific time.

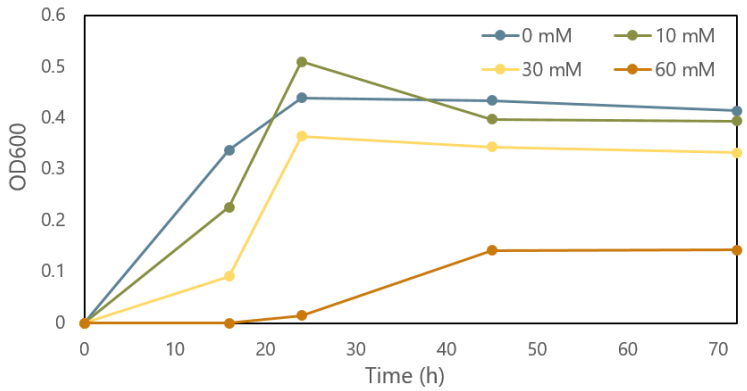


Figure 61. G5780 growth on a **244**-containing medium with a range of **54** concentrations. Growth conditions: MA medium, **247** 10 mM, **248** 10 mM and **249** 5 mM, ±**244** 5 mM and ±**54** 0-60 mM, 30°C, 180 rpm.

A first evolution of G5780 was attempted in the following conditions simultaneously in two different chambers:

- Turbidostat in relaxing medium (MA base, 10 mM **247**, 10 mM **248**, 5 mM **249**, 5 mM **244**, 30 mM **54**), at 30°C until reaching 3h32 and 4h37 of generation time.
- Medium swap regime with relaxing medium (MA base, 10 mM **247**, 10 mM **248**, 5 mM **249**, 30 mM **54**) and stressing medium (relaxing medium devoid of **244**) at 30°C with a 5h25 fixed generation time.

For the sake of clarity, the graph representing the evolution of G5780 in the abovementioned conditions is given in Figure 62 for only one chamber, estimated as the most relevant for interpretation.

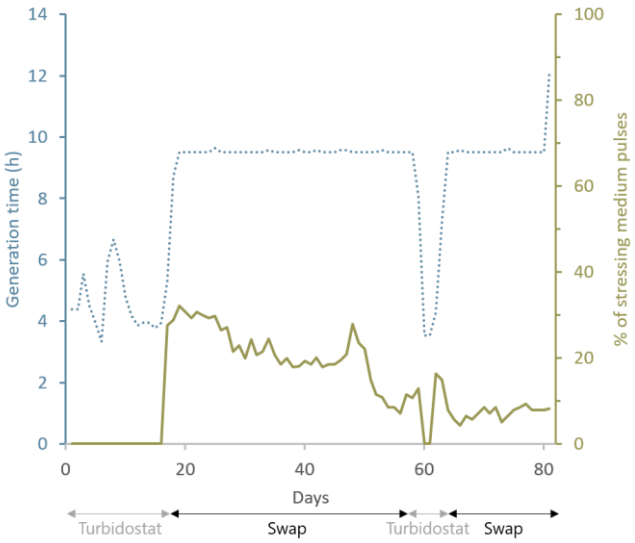


Figure 62. First attempt of G5780 evolution in continuous cultivation in the GM3 device. The graph shows the generation time, given in hours, and the ratio of stressing medium pulses. A second phase of turbidostat regime was carried out after around 60 h to try to densify the culture. The graph was automatically generated by a program designed by Laurent Gaillon.

Even in permissive medium, the strain behavior was not a typical one with very few crowding potential of the strain that can be due to an accumulation of a toxicity, potentially from cyclohexanone (**45**) release. Indeed, growth assays in presence of the reaction products **45** and methylamine (**38**) in the medium revealed a toxicity of **45** from 3 mM (data not shown) which was not the case in presence of the same concentrations of **38**. In swap regime, the ratio of stressing medium pulses was decreasing everyday meaning that the culture was not adapting to this medium.

Isolates sampled from the two chambers at the end of this first stay in the GM3 were grown on medium containing various concentrations of **244**, **54** and **45** in bioscreen assays given in Appendix 65. In these assays, **54** caused growth inhibition at 30 mM compared to what has been seen previously with only a lag phase even with the already tested G5780. Such inconsistencies between bioscreen results and behavior in GM3 do not have explanation yet. Also, the presence of **45**, even at 10 mM, negatively impacted the growth rate and the final OD. Thus, both the amine substrate and the ketone product seem to be detrimental for the evolution. On the other hand, increasing **244** concentration greatly improved the growth rate and the final OD reached by the original G5780 and the two isolates. The relaxing and stressing medium composition were reconsidered for higher amount of carbon source to avoid too limiting conditions for the strain: complementary tests displayed an improved growth with 10 mM **244** and 25 mM **247** instead of 5 mM and 10 mM, respectively, used for the first evolution. A second evolution is still running now in the following conditions:

- Turbidostat in relaxing medium (MA base, 25 mM **247**, 10 mM **248**, 5 mM **249**, 10 mM **244**, 10 mM **54**), one at 30°C and the other one at 37°C until reaching 2h20 generation time.
- Medium swap regime with relaxing medium (MA base, 25 mM **247**, 10 mM **248**, 5 mM **249**, 10 mM **244**, 10 mM **54**) and stressing medium (relaxing medium devoid of **244**) at 30 and 37°C with a generation time fixed at 4h45.

For the sake of clarity, the graph representing the evolution of G5780 in the abovementioned conditions is given in Figure 63 for only one chamber, estimated as the most relevant for interpretation.

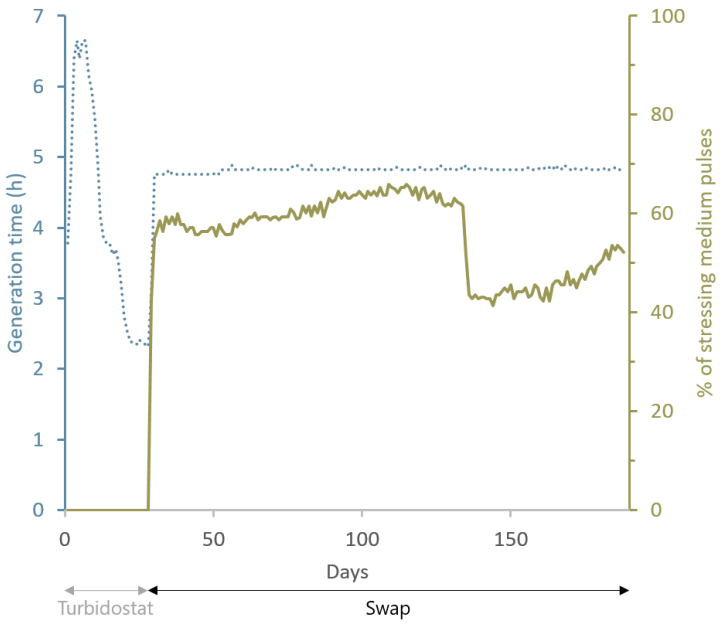


Figure 63. Second attempt of G5780 evolution in continuous cultivation in the GM3 device. The graph shows the generation time, given in hours, and the ratio of stressing medium pulses. The drop in the ratio of stressing medium pulses at 135 h occurred after a medium renewal. The graph was automatically generated by a program designed by Laurent Gaillon.

Compared to the previous evolution, the changes in **244**, **247** and **54** concentrations in both the relaxing and stressing medium drastically enhanced the generation time to 2h20 and led to higher ratio of stressing medium pulses. However, no particular event, especially in the number of pulses required by the culture, yet revealed any evolution of *Asp*RedAm even after more than 150 days of culture. The continuous culture is still running and is now slowly recovering from the drop observed at 135 days that occurred after a renewal of both mediums. The reason of this drop is still unexplained.

Despite the absence of positive results obtained, this strategy of improvement of nat-AmDH activity based on *in vivo* evolution remains interesting to look at. More efforts have perhaps to be directed towards NADPH-enzymes known to display high activity in the oxidative way such as Alcohol Dehydrogenases (ADH), with substrates already well studied and with interesting applications. These preliminary experiments revealed that such evolution of a heterogeneous enzyme not involved in the metabolism is a huge challenge This has to be a project in itself and not a side-project as during my PhD work and this despite the high involvement of the whole synthetic metabolism team.

III. Design of native amine dehydrogenases chimeras

From previous works on the nat-AmDH project during the thesis of Ombeline Mayol, we had experimental validated data of two different sets of homologous enzymes. One comprises the 4OP-AmDHs, and notably AmDH4, that are mostly thermostable and thermoactive but active towards a range of substrates with the requirement of the terminal carboxylic moiety. The second set encompasses nat-AmDHs that display a range of activities towards substrates without functional restriction but exhibiting their maximum activity around room temperature and rather poor thermostability (see Chapter II, III.1), while bearing the same global 3D fold. We attempted to merge their two best features to form an enzyme that would be both active towards simple ketones, thermostable and thermoactive. Thus, taking into account that all nat-AmDHs display a highly similar 3D structure, we envisioned a chimera gathering the global sequence and structure of AmDH4 (template 1) but bearing the active site of nat-AmDHs, namely *Cfus*AmDH, *Micro*AmDH and MATOUAmDH2 (template 2) (Table 13). Based on the hypothetical mechanism and analysis of their active sites, we already knew that the two different type of substrate scope was due to a notable different active site. The term "chimera" is not very well adapted as it is not a whole section of one enzyme fused with the all section of another one as it was done, among others, by Bommarius and co-workers between L-AmDH and F-AmDH, but we decided to keep this term here. The strategy was not to mutate isolately P1-P21 residues but whole segments to also mimic the neighboring residues found in the enzymes used as template 2. Chimeric enzymes designed for similar goals have already been reported, such as BVMOs (Van Beek *et al.*, 2012) or nitrilases (Zhang *et al.*, 2020).

Table 13. Comparison of P1-P21 residues of AmDH4, *Cfus*AmDH, *Micro*AmDH and MATOUAMDH2. The color code used refers to the polarity and charge of the corresponding residue [blue: polar residues, yellow: hydrophobic residues, orange: aromatic residues, red: negatively charged residues, and green: positively charged residues (charges at physiological pH)].

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21
AmDH4	S78	A101	E102	G133	V134	N135	F138	V139	R161	N163	D164	F168	V172	Q176	H197	I198	G199	H264	Q266	G299	T303
Cfus AmDH	S84	G107	E108	G139	F140	Q141	F144	W145	T166	Y168	N169	Y173	L177	H181	P200	S201	Y202	C269	G271	T301	T305
Micro AmDH	S80	S103	E104	G135	F136	Q137	F140	W141	S162	W164	N165	Y169	L173	Q177	P196	T197	F198	M265	G267	T297	T301
MATOUAmDH2	S87	A110	E111	G142	F143	L144	A147	C148	L169	Y171	N172	Y176	L180	H184	K209	S210	Y211	Q278	G280	T312	T316

The method used to create these chimeras was as follows:

- Deeply study each P1-P21 positions and neighbor residues between template 1 (AmDH4) and template 2 (*Cfus*AmDH, *Micro*AmDH and MATOUAmDH2) (Table 13)
- Check the structural alignment of these positions, the neighbor and surrounding residues (Figure 64)
- Mutate some P1-P21 residues and neighbors in template 1 to detect hindrance incompatibility with the direct environment and verify the feasibility of the mutations

- Determine the segments to mutate in AmDH4 and subsequently create the homology models of the chimeras using AmDH4 (PDB: 6G1M) as template

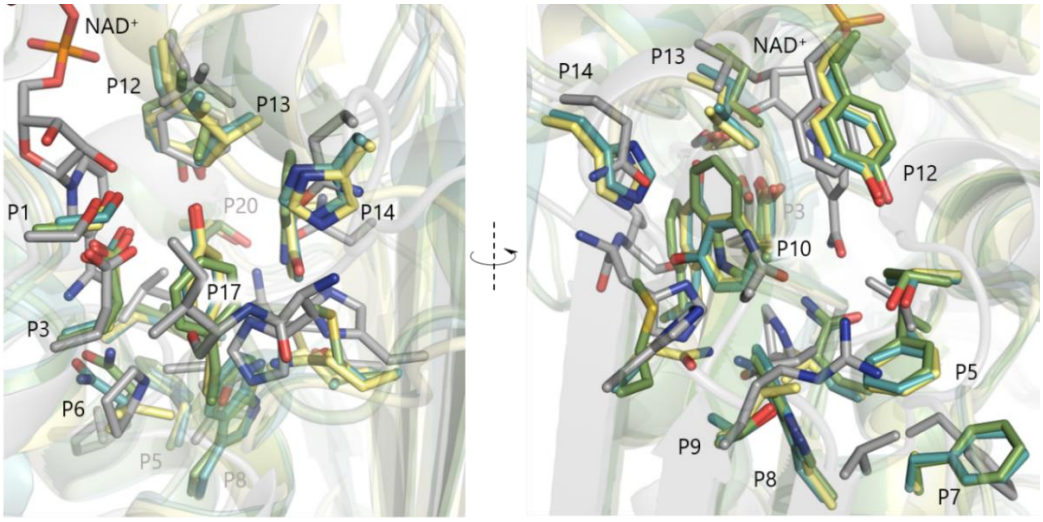


Figure 64. Structure alignment at 0° and 180° of *Cfus*AmDH (PDB: 6IAU), *Micro*AmDH (model), MATOUAmDH2 (model) on AmDH4 (PDB: 6G1M). AmDH4, *Cfus*AmDH, *Micro*AmDH and MATOUAmDH2 are colored in grey, blue, green and yellow, respectively. Some P1-P21 are represented in stick format. RMSD (AmDH4-*Cfus*AmDH) = 1.936 Å ; RMSD (AmDH4-*Micro*AmDH) = 1.442 Å ; RMSD (AmDH4-MATOUAmDH2) = 1.906 Å.

Two sections were particularly difficult to design, namely the top of the spine helix V134-I150 mentioned in Chapter III, I.4.7.2.3 (F140-G154 in *Cfus*AmDH, Appendix 66) and the linker sections G191-G199 between the helix P182-S191 and P17 residue mentioned in Chapter II, I.1.2 (D186-S210 in *Cfus*AmDH, Appendix 67). In the first segment, the helix in AmDH4 includes an additional residue that increase the diameter of the helix, thus not well defined in AmDH4 structure. Accordingly, we decided to completely modify the segment G133-T142 in AmDH4 by the corresponding sequence of the template 2 (G139-N147 in *Cfus*AmDH, G135-N143 in *Micro*AmDH and G142-G150 in MATOUAmDH2). In AmDH4, the whole second section G191-G199 enables the correct positioning of the catalytic residues P3 when it is only guaranteed by the hindered residues in P17 in the nat-AmDHs. Thus, for each system, and in addition to all the other segment modifications, two chimeras were generated, one with only the modification of P15-16-17 in AmDH4, replaced by the residues found in the template 2 (named AmDH4_X, X being the enzyme used as template 2), and another one with the modification of the whole section S191-G199 to mimic the folding found in the nat-AmDHs (named AmDH4_X_P17all). The alignment of the six chimeras generated and the mother enzymes is given in Figure 65.

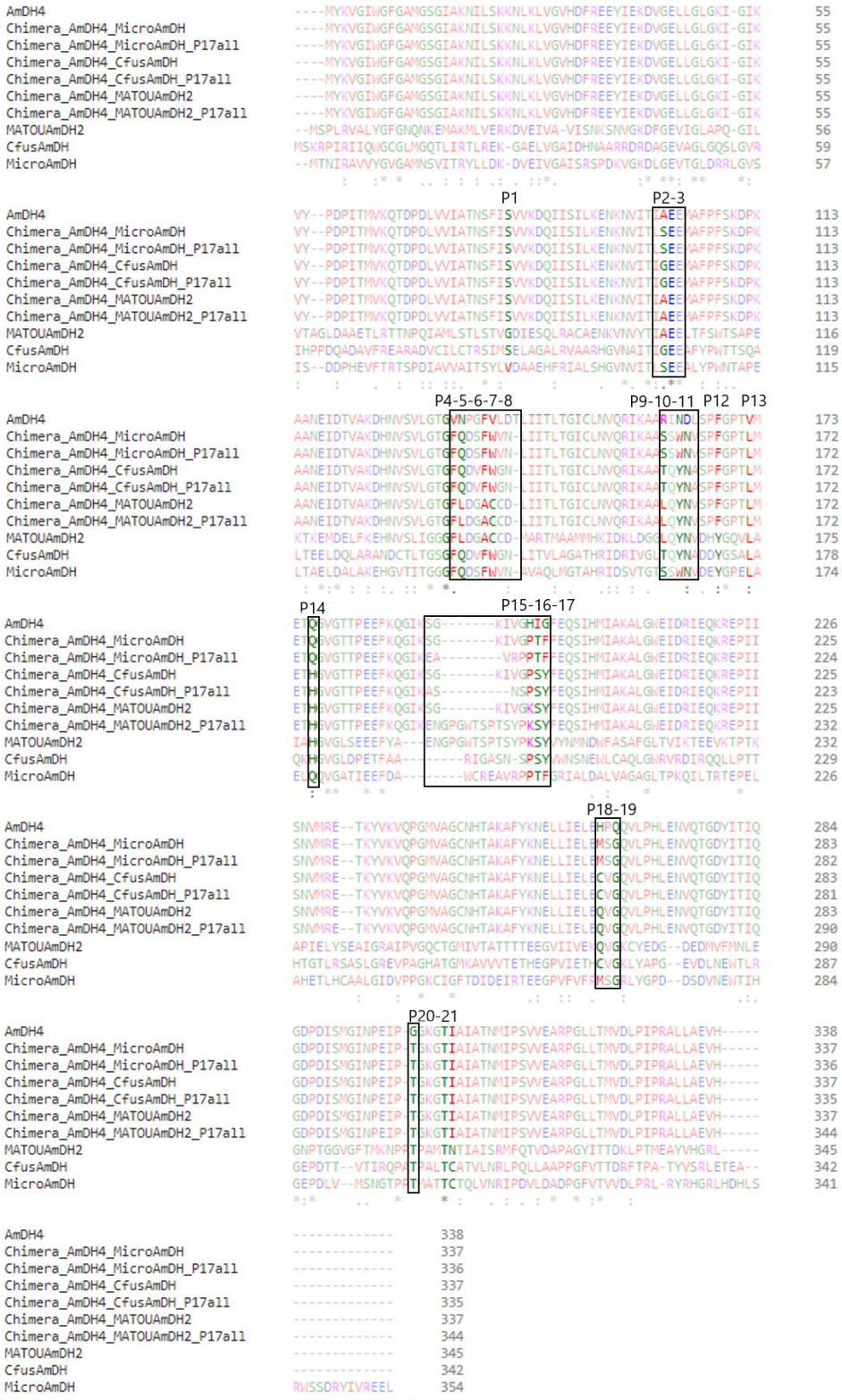


Figure 65. Multiple sequence alignment of the template 1 (AmDH4), the templates 2 (*Cfus*AmDH, *Msme*AmDH and MATOUAmDH2) and the corresponding chimeras AmDH4_X and AmDH4_X_P17all (X being the enzyme used as template 2). The frames indicate the mutated segments in AmDH4. The alignment was done using Clustal Omega web service.

The synthetic genes were purchased for each chimera and cloned into a pET22b(+) vector. Among the six chimeras overexpressed in *E. coli*, AmDH4_*Cfus*AmDH and AmDH4_*Micro*AmDH were well expressed and purified in a good yield but were not active towards some typical nat-AmDHs substrates such as isobutyraldehyde (**170**) cyclohexanone (**45**) and pentan-2-one (**112**), AmDH4_*Cfus*AmDH_P17all and AmDH4_*Micro*AmDH_P17all were well expressed but much more difficult to purify. This observation may highlight the importance of the linker section conservation to ensure the correct folding of AmDH4. Eventually, the two chimeras generated from AmDH4 and MATOUAmDH2 were well expressed but insoluble. The SDS-PAGE gels of these productions are given in Appendix 68. Complementary tests were done to use the insoluble fractions containing the enzymes for biocatalytic reactions but no conversion were observed after 24 h at 30 or 50°C with any chimera while AmDH4 displayed traces of **170** conversion after 24 h at 50°C (data not detailed). The monitoring of the chimeras activity towards typical substrates of AmDH4 [(2*R*)-2-amino-4-oxopentanoic acid (**166**) or 4-oxopentanoic acid (**152**)] has not been done but we could expect a drastic decrease of activity due to the removal of Arg161, important for the accommodation of the carboxylic moiety in the active site.

Giving the priority to other part of the project, no further investigation was carried out from this time. A collaboration with Gwenaëlle André-Leroux, who was already involved in the MD study of *Cfus*AmDH-W145A, is underway to redesign a new set of chimeras with the same objective. A good strategy could be the one done by Bommarius and co-workers to obtain cFL-AmDH (Bommarius *et al.*, 2014).

IV. Conclusion

Until now, only new members of the rather recent nat-AmDH family were reported and no specific work was carried out to improve their activity or expand their substrate scope. The only engineering work was performed on AmDH4 both by our team and by Cai *et al.* (Cai *et al.*, 2020) but nothing on the true nat-AmDHs set. Thus, we had many opportunities and space left to improve their activity, substrate scope and stability using various approaches. Thanks to active and very good collaborations, this set of enzymes has now been well investigated. First, the *in silico* analysis of the crystallographic structures and models available allowed us to expand the substrate spectrum of a set of nat-AmDHs to make them accept larger carbonyl compounds. Notably, the strong point of these mutants was to enable the formation of *n*-terminal aliphatic amine and (*S*)-amines from 2C and 3C ketones, the latter being rarely reported in other enzyme families performing reductive amination. An improvement of activity by *in vivo* directed evolution, which is uncommon in biocatalysis, was also attempted but difficult to implement due to a poor activity of our set of nat-AmDHs in the oxidative direction. The evolution of the promising

*Asp*RedAm is still underway in the GM3 and further work on other oxidoreductases will still be highly interesting to try. Finally, the very ambitious project of creating a chimera that would combine the thermostability and thermoactivity of AmDH4 and the substrate spectrum of nat-AmDHs did not succeed for now but the objective is worth testing new designs with the help of experts in enzyme structure analysis. This goal could also be achieved through deeper protein engineering work on AmDH4 or homologs, regardless of any consideration of fusion with the active site of other nat-AmDHs. Although the objectives of stability improvement or switch of cofactor preference has not been reached or even taken up due to a lack of time, all these new data complete our knowledge on the nat-AmDHs structure, mechanism and biocatalytic potential. Combined with the use of bioinformatics tools and databases, these findings have been essential to find other nat-AmDHs with interesting features among biodiversity. This point is described in the following Chapter IV.

CHAPTER IV

Intensive research for native amine dehydrogenases among (meta)genomic diversity

One of the main objectives of my PhD work was to deeply explore the nat-AmDH family to identify some candidates with key features, quite novel compared to the ones already characterized (substrate scope, enantioselectivity, cofactor preference). From SDM studies described in the previous chapters for active site understanding and improvement, some target residues have been identified to play a key role in the reaction or, for example, in the accommodation of a limited range of substrates. All these studies were the starting point for further exploration of the biodiversity and are described in this chapter. Two subsets of new nat-AmDHs selection have been generated and described here. One came from the exploration of the restricted set of reference nat-AmDHs (ASMC1, Figure 16) while the second selection was done in within the ASMC2 set built during the MODAMDH project, briefly reported here.

I. First selection of native amine dehydrogenases with larger active sites

The restricted set of nat-AmDHs, classified in ASMC1 (Figure 16) comprised 2,011 members. Six of them, harboring a short residue at P8 position within G3 and G4 groups mentioned in Chapter III, I.1, were selected for heterologous expression in *E. coli* to be tested towards substrates similar to that accepted by the P5A and P8A mutants. None could be found with a short residue at P5 within the groups considered. Among the selection, A0A078BAL0 and A0A078AU13 were found in *Stylonychia lemnae*, an eukaryotic organism (Table 14 and Appendix 5).

Table 14. P1-P21 residues of the native nat-AmDHs selected within the ASMC1 set to accommodate large carbonyl substrates. *Cfus*AmDH and *Msme*AmDH are given as references. Colors refer to the polarity and charge of the corresponding residue [blue: polar residues, yellow: hydrophobic residues, orange: aromatic residues, red: negatively charged residues, and green: positively charged residues (charges at physiological pH)].

Enzyme	Organism	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21
Msme AmDH	*Mycolicibacterium smegmatis*	S80	S103	E104	G135	Y136	Q137	F140	W141	A162	W164	N165	F169	L173	Q177	P197	T198	F199	M266	G268	T298	T302
Cfus AmDH	*Cystobacter fuscus* DSM 2262	S84	G107	E108	G139	F140	Q141	F144	W145	T166	Y168	N169	Y173	L177	H181	P200	S201	Y202	C269	G271	T301	T305
A0A1Q9SC61	*Pseudonocardia* sp. CNS-004	S80	G103	E104	G135	F136	Q137	F140	L141	E162	F164	N165	Y169	V173	Q177	P197	S198	F199	L266	G268	T298	T302
A0A1H8XY75	*Amycolatopsis saalfeldensis*	S78	G101	E102	G133	F134	Q135	F138	L139	Q160	F162	N163	Y167	V171	Q175	P195	G196	F197	L264	G266	T296	T300
A0A1I3IZQ4	*Natronobacterium gregoryi*	S79	S102	E103	G134	Y135	Q136	F139	E140	N161	Y163	N164	Y168	V172	F176	E196	S197	L198	E265	A267	V297	T301
A0A078BAL0	*Stylonychia lemnae* (Ciliate)	S78	N101	E102	G133	Y134	Y135	L138	G139	I160	Y162	N163	Y167	V167	N172	P192	T193	Y194	A262	S264	T296	T300
A0A078AU13	*Stylonychia lemnae* (Ciliate)	S78	N101	E102	G133	Y134	F135	L138	G139	V160	F162	T163	F167	V168	D173	I193	A194	Y195	A263	S265	T297	T301
A0A076NBW6	*Fischerella* sp. ATCC 43239	S84	A107	E108	G139	F140	Q141	F144	L145	I166	Y168	N169	Y173	V177	Y181	F201	P202	S203	V269	G271	T301	T305

These enzymes were first overexpressed in *E. coli* BL21 C+ using the typical in-house induction protocol (Experimental section, II.1, II.3 and II.4, Appendix 69 and Appendix 70 report the sequence of the proteins selected and the primers used, respectively) but only very low amounts of even not pure batches were obtained because of their apparent high insolubility. Except A0A1I3IZQ4 that was barely expressed, all the others were rather well induced but could not be recovered during the lysis step due to inclusion bodies formation. Because of the very low amount of His-tag protein extracted from the lysate, the purification step led to the co-elution of a high amount of undesirable *E. coli* proteins. Changes in the induction protocol or the use of other techniques did not help producing these enzymes in drastically higher amounts. All the protocols tested are described in Experimental section, II.7. For example, we tested using a lower isopropyl ß-D-1-thiogalactopyranoside (IPTG, **250**) concentration, 0.1 and 0.2 mM instead of 0.5 mM to slow down the induction, enable a better enzyme folding and thus avoid the formation of inclusion bodies. With the same strategy, we also used ArcticExpress competent cells®, that can grow at lower temperature, and tested them at 12°C and 15°C using either Luria Broth (LB) or Terrific Broth (TB) medium. The use of such strains to induce the reference *Cfus*AmDH resulted in a good overexpression and purification even if a lower amount of enzyme could be recovered compared to overexpression in BL21 C+. In fact, the low temperature used for the induction produced a lower amount of biomass. For the enzyme candidates of Table 14, the use of ArticExpress® helped producing A0A078BAL0, A0A076NBW6 and A0A1H8XY75 in a slightly higher amount and better purification rate. For example, for the same volume of elution buffer used to recover the purified enzyme, A0A1H8XY75 was obtained in 0.28 mg mL^{-1} using BL21 C+ and 0.53 mg mL^{-1} using ArticExpress. The Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) also showed a slightly more intense band for the overexpressed enzyme compared to the other protein inherently expressed by the *E. coli* strain (Appendix 71). The desalting step to remove imidazole (**251**) in the final buffer was also removed to avoid losing more enzyme material despite no clear knowledge of the effect of keeping **251** in the enzyme storage buffer. An autoinduction protocol, that was used once for the production of metagenomic nat-AmDHs and suggested by Dr. Adam Caparco (Caparco *et al.*, 2020), was tested but

the enzyme were not even expressed. The two most promising enzymes for production, A0A076NBW6 and A0A1H8XY75, were also overexpressed with His-Tag removal or switch from N-ter to C-ter position and, separately, with the addition of highly soluble tags $(Lys)_5$, $(Arg)_5$ and $(Asp)_5$ that were supposed to facilitate the enzyme solubility (Paraskevopoulo *et al.*, 2018). No improvement was observed with any of these clonings whether in terms of amount produced or purity.

Despite having only very small amount of enzymes, the activity were assayed mainly towards isobutyraldehyde (**170**) and pentanal (**175**), and some with hexanal (**207**). No activity could be detected by spectrophotometry monitoring NAD(P)H consumption. Nevertheless, the background activity due to the poor purity of the batches made it hard to quantify the activity induced by the overexpressed enzyme. The low amount of enzymes available made it impossible to test them in conversion assays in presence of the cofactor recycling system and for amine formation monitoring using UHPLC. Following the work done on *Cfus*AmDH mutants at P5 and P8, we can now criticize this choice of substrates. In fact, the loss of activity subsequently observed with *Cfus*AmDH mutants towards shorter substrates, discussed in Chapter III, I.2 may be the reason of the inactivity of theses enzymes selected for their larger active site. No further investigation was conducted since then on these batches.

Due to these failures to produce the selected nat-AmDHs, we decided to focus on P5 and P8 nat-AmDH mutants to expand the substrate scope. With more time, it would have been interesting to attempt the production of some candidates with fusion tag of hypersoluble protein protocols [*i.e.* Small Ubiquitin-like Modifier (SUMO), Glutathione *S*-transferase (GST), etc.], different lysis protocols [i.*e.* urea (**252**) or guanidine (**253**)] or, in the case of the eukaryotic proteins A0A078BAL0 and A0A078AU13, to use yeast expression systems such as *Saccharomyces cerevisiae* or *Pichia Pastoris*.

Very recently, A0A078BAL0 and A0A078AU13 were produced in addition to other nat-AmDHs, selected from the ASMC2 set, to describe the substrate scope of the enlarged nat-AmDH family (see Chapter IV, II.3). The lysates production, performed in 96-well plates, using 0.5 mM IPTG led to similarly low enzyme recovering but a second production, carried out using 1 mM IPTG, led to significantly higher amount of A0A078BAL0 cell free lysate that was found active towards **45** with either **11** or **38** as amine sources but also towards hexan-3-one (**200**). Another alternative would have been to use PROSS webservice, described by Goldenzweig *et al.*, to identify and introduce targeted mutation in the sequences to enhance their expressability in *E. coli* as well as their stability (Goldenzweig *et al.*, 2016).

II. Selection of other native amine dehydrogenases from the MODAMDH project

The MODAMDH project was introduced by Dr. Carine Vergne-Vaxelaire and David Vallenet (LABGeM, Genoscope), financed by the French National Research Agency (ANR) (ANR-19-CE07-0007) and mainly carried out by Dr. Eddy Elisée, a post doctoral student on this project. The first objective was to widen the landscape of protein sequences catalyzing reductive amination by finding enzymes with high diversity, either distant homologs or structural analogs. This diversity can translate into both active site or the overall fold of the enzymes. This broad survey was achieved by screening the biodiversity collected in large genomic and metagenomic public and in-house databases, which represents around 2.6 billion sequences (Appendix 72). The purpose and steps of the MODAMDH project are briefly described in this chapter for the sake of clarity but this study was mainly carried out by Dr. Eddy Elisée (Chapter IV, II.1-II.4). Thus, the focus is done on the selection of nat-AmDHs harboring specific features within the extended set built during the project.

II.1 Datasets construction and clustering

They first extended the sequence diversity of the reference nat-AmDH family from 2,011 (ASMC1) to 17,039 sequences, after removal of the redundancy, by screening the aforementioned databases using HMM profiles built from the ASMC1 groups (Figure 16). This extended set was studied through phylogenetic trees but also through a new hierarchical active sites classification, named ASMC2 (Figure 66) according to the pipeline described in Appendix 4. First it was submitted to MODELLER (Sali & Blundell, 1993) to build 3D models for use in the ASMC analysis. The modeling was only done for the proteins sharing at least 23% of sequence identity with one of the crystallographic structures used as templates (AmDH4, *Msme*AmDH and *Cfus*AmDH), which represented 9,886 models. The template was automatically the one sharing the highest sequence identity with the target. Hence, the ASMC method was applied to these models using *Cfus*AmDH pocket as reference to find the corresponding P1-P21 in each of the 9,886 models. Based on the P1-P21 profiles, all the models were clusterized to give the classification presented in Figure 66.

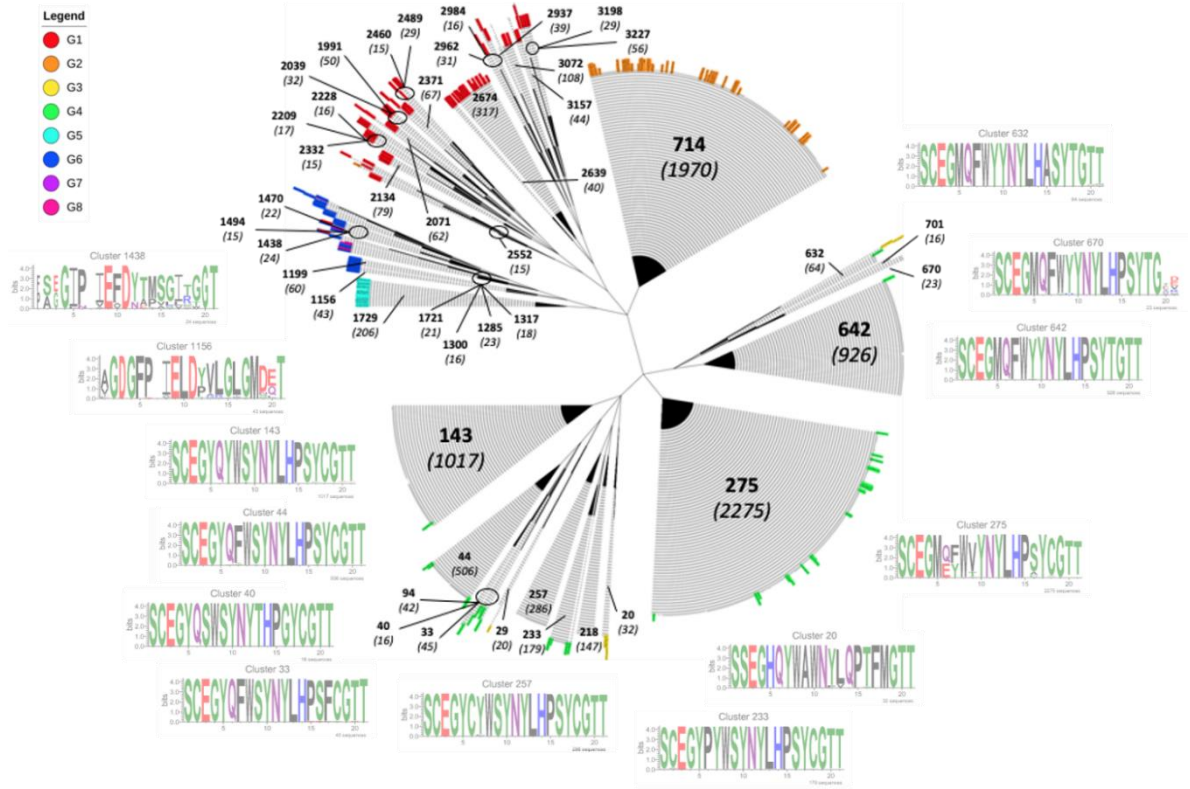


Figure 66. ASMC2 clusterization. For sake of clarity, only the clusters containing more than 15 members are represented. For better tracking, the colored labels refer to the previous G1-G8 groups from ASMC1. Black bold numbers correspond to the cluster numbering and number in brackets to the cluster size. The logos are related to the mainly studied clusters in Chapter IV, II.5. The logo represents the conservation of the active site residues and the color code refers to the chemical properties of the corresponding residue [green: polar residues, black: hydrophobic residues, red: acidic residues, blue: basic residues and purple: neutral residue (at physiological pH)].

In addition to this reference family, the sequences to be screened among the 2.6 billion available were restricted to enzymes bearing an NAD(P)-binding Rossmann fold domain, as we only targeted NADPH-dependent enzymes performing reductive amination. This was carried out using the specific SCOP Superfamily entry SSF51735 and eventually led to 20,315,745 sequences. These sequences were also clusterized and each of the 104,686 resulting subclusters was associated to an HMM profile.

II.2 Selection strategy using distant homology

Using HMM-HMM comparison (Söding, 2005), HMM profiles of the reference family were screened against the 104,686 HMM profiles of the NAD(P)-dependent families using the hhblits software (Gabler *et al.*, 2020; Zimmermann *et al.*, 2018). 440 new sequences, displaying less than 80% identity over 80% of their sequence length (id80/cov80), were retrieved with this method but most of them were fragments, thus resulting in an only few number of valuable candidates. Moreover, when comparing the phylogenetic trees (Figure 67), one can note that the addition of these new sequences did not create a new branch. This suggests that we already have a complete picture of the nat-AmDH family with the

reference family of 17,039 sequences. Only the protein METDB-00128-1-DN9853 from the Marine Eukaryotes Transcriptomes (METdb) database was selected to be experimentally screened and has been added to the set of enzymes described in Chapter IV, II.3.

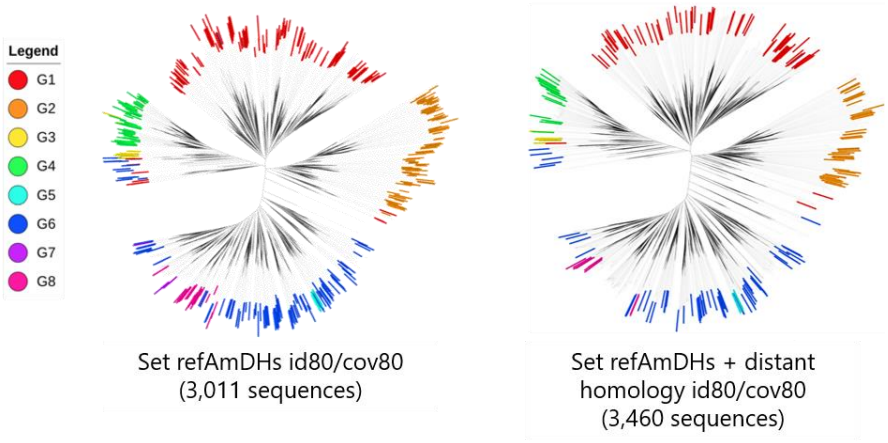


Figure 67. Phylogenetic tree of (A) the extended reference nat-AmDH set (id80/cov80) and (B) the extended reference nat-AmDH set with addition of the sequences retrieved from the distant homology approach. 3,114 sequences were identified and added to the extended set of 27,282 reference nat-AmDHs (without removal of the redundancy). This new set of 30,396 sequences was then pooled at id80/cov80 to finally obtain 3,460 non-redundant sequences. The color code refers to the G1-G8 active site homology groups.

II.3 Experimental validation of the extended family of native amine dehydrogenases

In a first phase of nat-AmDH family description, Jean-Louis Petit (LGBM) selected one or several representative member(s) of each cluster of the AMSC2 id95/cov90 (7,039 sequences) to cover the tree diversity. A total of 122 nat-AmDHs were selected and overexpressed in *E. coli* by Adrien Debard and Virginie Pellouin, from the Cloning and Screening Platform. The enzymes successfully produced were then screened by Virginie Pellouin and Dr. Carine Vergne-Vaxelaire against a range of typical nat-AmDHs substrates [(2*R*)-2-amino-4-oxopentanoate (**166**), cyclohexanone (**45**), butan-2-one (**134**), 1-methoxypropan-2-one (**189**), furfural (**57**)], but also some less identified substrates [hexanal (**207**), hexan-2-one (**43**), hexan-3-one (**200**), benzaldehyde (**31**), 5-hydroxylmethylfurfural (**58**)], monitoring their conversion by UHPLC-UV and/or NAD(P)H-spectrophotometry assay. For enzymes of G1, with still no identified substrate for reductive amination, ten other compounds were hypothesized from *in silico* analysis and tested, mainly in the oxidative deamination direction due to the non availability of the corresponding ketones (not detailed here). These putative amines harbor a carboxylic group like in **165** as their active site display a conserved Arg at the position equivalent to R161 in AmDH4 (P9). The partially analyzed results are presented in Appendix 73.

The screening methods and quantitative results are not detailed but the main conclusions were as follows:

- The substrate(s) of enzymes from G1 group have still not be determined despite the screening of the ten putative substrates. Some enzymes were active towards **166**, meaning their behavior is more similar to that of G2 enzymes. Their automatic classification in G1 has not been analyzed.
- Enzymes from G2 were all active towards **166**, indicating their potential role in ornithine degradation pathway, as for AmDH4, also included in this group.
- An important part of enzymes from G3 and G4 were active towards their expected substrate such as **45**, **57** or **134**. Some of them were also active towards **166** but with lower conversions than for G2 enzymes.
- The enzymes METDB-00128-1-DN9853 (G3-G4), resulting from the screening for distant homologs, was a good hit with formation of cyclohexylamine (**179**), (*2S*)-butan-2-amine (**193**), hexan-3-amine (**238**), furfurylamine (**254**), benzylamine (**74**).
- A0A646KJR1 (G3) displayed a good activity towards hexan-3-one (**200**) and was therefore used for an *in silico* analysis aiming at finding other enzymes active towards 3C ketones (Chapter IV, II.5.3).

II.4 Selection strategy using active site analogy

In addition to the HMM-HMM comparison strategy, a more structure-based approach was attempted with the design of a nat-AmDH minimum active site constellation to recover analogs (Steinkellner *et al.*, 2014). Such a design consists in the determination of a minimum active site that is sufficient to perform the required reaction. It considers the nature of the amino acid and the distances between each selected residue to find enzymes structures harboring the same 3D-configuration. I contributed to the design based on the experimental results obtained from the protein engineering work detailed in Chapter II, II.

The first designs were based on *Cfus*AmDH and *Msme*AmDH available 3D-structures (PDB: 6IAU and 6IAQ, respectively) and on MATOUAmDH2 preliminary 3D-structure provided by the team of Prof. Grogan. In view of the results of the SDM strategy performed to determine the importance of some specific residues (see Chapter II, II), the minimum active sites comprised the residues summarized in Figure 68 and Table 15 (Constellation 1). The first constellations were built with a range of distances between each considered atoms in *Cfus*AmDH, *Msme*AmDH and MATOUAmDH2. The screening would be done in the available PDB structures bearing NAD(P)-cofactor but also in the AlphaFold structures database, comprising over 900,000 of mostly apo structures at this time (Varadi *et al.*, 2021). Thus, the

set considered could include whether closed or open conformations of the active site. Therefore, considering P12 in the minimum active site was not relevant as this amino acids is highly mobile through the enzyme open-closed cycles. The C4N atom of the cofactor also could not be considered as the AlphaFold structures does not include an atom of the cofactor. We proposed three other minimum active sites replacing P12 and NAD(P)$^+$ by more spatially stable and conserved residues interacting with the nicotinamide moiety of the cofactor. Notably, P6 was already mentioned in Chapter III, I.4.7.2.3 to bind with the amide substituent of the nicotinamide ring and is rather well conserved among the family. Also, P21 and a position named here P1' (I106 in *Cfus*AmDH), located behind the ring and well-conserved in the family, were added to keep considering the cofactor through its nicotinamide binding site. The new designs are detailed in Figure 68 and Table 15 (Constellations 2-3 and 4). The constellations 2 and 3 differ only from the flexibility of amino acids considered for P21.

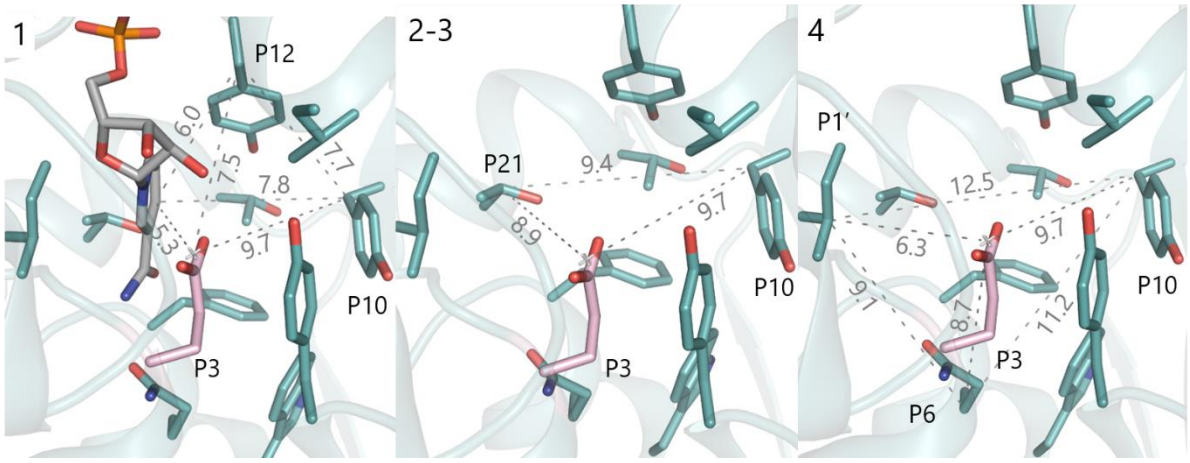


Figure 68. Atoms and distances considered for the design of minimum constellations in *Cfus*AmDH. Catalytic residue P3 and NADP$^+$ are highlighted in pink and grey, respectively. Distances are given in Angstroms (Å).

Table 15. Atoms and flexibility of amino acid type for the constellations design. "-" means that all amino acids were allowed at this position.

Const.	Position	Atom	Type
1	P3	pseudoatom OE1-OE2	Glu
	P10	CB	Tyr/Phe/Trp
	P12	CB	Tyr/Phe/Trp
	NAD(P)	C4N	
2	P3	pseudoatom OE1-OE2	Glu
	P10	CB	Tyr/Phe/Trp
	P21	CB	Thr
3	P3	pseudoatom OE1-OE2	Glu
	P10	CB	Tyr/Phe/Trp
	P21	CB	-
4	P1'	CB	Ile/Thr/Leu
	P3	pseudoatom OE1-OE2	Glu
	P6	CB	Gln/Cys/Ile/Leu/Ser
	P10	CB	Tyr/Phe/Trp

The screening included the following steps: each Glu residue were detected and a pocket search was subsequently carried out using Fpocket software (Barber *et al.*, 1996; Le Guilloux *et al.*, 2009; Liang *et al.*, 1998; Schmidtke *et al.*, 2010). If a pocket was found, the constellation was screened using YASARA (Krieger *et al.*, 2002). Other criteria were considered such as angles between the selected atoms and the accessibility of the Glu, to remove the cases of Glu buried outside of the detected pocket. It resulted in 96 hits from the AlphaFold dataset and 98 NAD(P)-containing PDB structures, including the nat-AmDHs reference structures, but after closer investigation of these hits, also carried out by Dr. Eddy Elisée, none of them were considered viable for reductive amination. Within the MODAMDH project, all the models that could be generated from the 20 M sequences of NAD(P)-proteins retrieved from biodiversity should have been screened to find the designed constellations but the time project did not allow to do this step.

II.5 Selection of native amine dehydrogenases with novel features

As the two bioinformatic approaches to find distant homologs or active site analogs did not result in hits satisfying the expected criteria (enzyme of a new phylogenetic branch, enzyme with another type of fold, etc.), the search for nat-AmDHs with interesting features was fully done within the extended set of reference nat-AmDHs clusterized in the AMSC2 (17,039 sequences at id100/cov100 without redundancy).

I carried out a structure-guided exploration by searching for nat-AmDHs candidates with particular P1-P21 residues that could lead to some wanted features. The exploration set considered all the extended nat-AmDH family clusterized in ASMC2 id100/cov100, which represented 9,763 models from *Cfus*AmDH as the template for P1-P21 recognition (Figure 66). For the first round of selection, conducted with Dr. Eddy Elisée and Dr. Carine Vergne-Vaxelaire, we particularly focused on the clusters corresponding to the G3 and G4 P1-P21 profiles, namely enzyme active towards carbonyl substrates without functional requirement. Then, I sorted all the candidate enzymes into different objectives that could be fulfilled according to their specific structural features and decided to mainly focus on three of them: *(1)* increased active site volume to accommodate longer aldehydes or 2C ketones, *(2)* increased active site volume to accommodate more complex amine substrates than **11**, *(3)* particular active sites to accommodate 3C ketones and *(4)* switch from a typical (*S*)- to a (*R*)-stereoselectivity. The next parts describe the method, positions considered and potential docking simulations carried out to refine the selection.

Due to first production failure with ASMC1 candidate enzymes, we decided to use the solubility prediction tool SoluProt to have more insights into the expressability and solubility of candidates based

only on their sequence information (Hon *et al.*, 2021). Even though SoluProt provides only 58.6% accuracy, the highest to date, this parameter was useful, for example, to discriminate two enzymes with similar active site features but different overall sequences and predicted solubility.

II.5.1 Selection of enzymes for accommodation of carbonyl substrates harboring a chain length of at least six carbon atoms

Making the link with the P5A and P8A mutants (Chapter III, I) and the attempts of production of nat-AmDHs from ASMC1 bearing similar features (Chapter IV, I), another round of nat-AmDHs selection was carried out from the ASMC2 clusters with the same objective. The main targeted positions were again P5 and P8, but also P7 located in the second layer, in π-stacking interactions with P5, and mainly occupied by hindered Tyr or Phe. We hypothesized that the presence of a shorter amino acid at this position could impact P5 (Tyr/Phe/Met/…) side chain positioning to the outside of the active site and thus, leaving space for larger substrates. Table 16 summarizes the enzymes selected for this objective and their corresponding P1-P21.

Table 16. P1-P21 residues of the selected nat-AmDHs with a larger active site for carbonyl substrates. *Cfus*AmDH is given as a reference and P5, P7 and P8 are highlighted as the target positions for this objective. Colors refer to the polarity and charge of the corresponding residue [blue: polar residues, yellow: hydrophobic residues, orange: aromatic residues, red: negatively charged residues, and green: positively charged residues (charges at physiological pH)]. The theoretical solubility was estimated using SoluProt web service (score > 0.5, yellow to green, corresponds to a soluble protein while a score < 0.5, yellow to red, corresponds to an insoluble protein).

	Cluster (Group)	P1	P2	P3	P4	**P5**	P6	**P7**	**P8**	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	Solubility
Cfus AmDH	275 (G3/G4)	S	C	E	G	M	Q	F	W	V	Y	N	Y	L	H	P	S	Y	C	G	T	T	0.501
A0A138ZYM0	233 (G3/G4)	S	A	E	G	Y	S	A	W	S	Y	N	Y	L	H	P	S	Y	T	G	T	T	0.497
MGYP000165741795	257 (G3/G4)	S	C	E	G	Y	C	Y	W	S	Y	N	Y	L	H	E	M	K	C	G	T	T	0.724
MGYP001184784288	670 (G3/G4)	S	C	E	G	M	Q	F	G	Y	Y	N	Y	L	H	P	S	Y	T	G	-	D	0.601
MGYP000563817680	642 (G3/G4)	S	C	E	G	V	Q	F	W	Y	Y	N	Y	L	H	P	S	Y	T	G	G	T	0.319
MGYP000963671166	642 (G3/G4)	S	C	E	G	M	Q	-	-	Y	Y	N	Y	L	H	P	S	Y	T	G	T	T	0.757
GUT_GENOME220312_01244	642 (G3/G4)	S	C	E	G	I	Q	F	W	Y	Y	N	Y	L	H	P	S	Y	T	G	T	T	0.594
MGYP001470669209	20 (G3/G4)	S	S	E	G	H	Q	Y	W	A	W	N	F	V	Q	P	T	F	M	G	T	T	0.515
METDB-02	275 (G3/G4)	S	A	E	G	V	Q	L	W	I	Y	N	Y	L	H	P	A	Y	M	G	T	T	0.534
METDB-03	275 (G3/G4)	S	A	E	G	A	Q	L	W	V	Y	N	Y	L	H	P	A	Y	M	G	T	T	0.535
GUT_GENOME000603_01215	275 (G3/G4)	A	C	E	G	A	Q	G	W	V	Y	N	Y	L	H	P	C	Y	C	G	T	T	0.717
MGYP000417998329	275 (G3/G4)	S	C	E	H	S	K	I	T	V	Y	N	Y	L	H	P	S	Y	C	G	T	T	0.488
MGYP000037226974	275 (G3/G4)	S	C	E	G	E	D	L	G	V	Y	N	Y	L	H	P	S	Y	C	G	T	T	0.602
MGYP000528433141	275 (G3/G4)	S	C	E	G	A	N	Q	Y	A	Y	N	Y	L	H	P	S	Y	C	G	T	T	0.637
MGYP000288071185	275 (G3/G4)	S	C	E	G	K	E	F	W	V	Y	N	Y	L	H	P	S	Y	C	G	T	T	0.322
MGYP000273355962	257 (G3/G4)	S	C	E	G	Y	C	L	W	S	Y	N	Y	L	H	P	S	Y	C	G	T	T	0.558
IGC-14	44 (G3/G4)	S	C	E	G	Y	Q	A	W	S	Y	N	Y	L	H	P	S	Y	C	G	T	T	0.625
MGYP001051592001	44 (G3/G4)	S	C	E	G	Y	Q	S	W	S	Y	N	Y	L	H	P	S	Y	C	G	T	T	0.470

The 17 selected sequences were purchased as synthetic genes for production by overexpression in *E. coli* and will be screened against a range of aliphatic aldehydes, 2C and 3C ketones and towards 4-phenylbutan-2-one (**36**), which was only accepted by the P5A and P8A mutants (Figure 42 and Appendix 24).

II.5.2 Selection of enzymes for accommodation of primary or secondary amines

In the search for other mutable residues to alter the nat-AmDHs substrate scope, P13 (L177 in *Cfus*AmDH) was investigated. This residue is located in the active site ceiling (Figure 23 and Figure 69), close to the hindered P12 residue. If we refer to *Cfus*AmDH model with **45** and **11** docked in the active site by Karine Bastard, L177 is the direct first layer of the active site in **11** side. Its mutation to a smaller residue could improve the accommodation of more complex primary amines as amine donor instead of the simple **11** or **38**. Even though the nat-AmDH family was more described with **11**, some nat-AmDHs were already reported to accept **38**, especially *Micro*AmDH, *Mvac*AmDH, *Cfus*AmDH and *Msme*AmDH, displaying 851.3, 262.6, 257.7 and 144.5 mU mg^{-1} specific activity, respectively, towards **170** in presence of **38** (Mayol *et al.*, 2019).

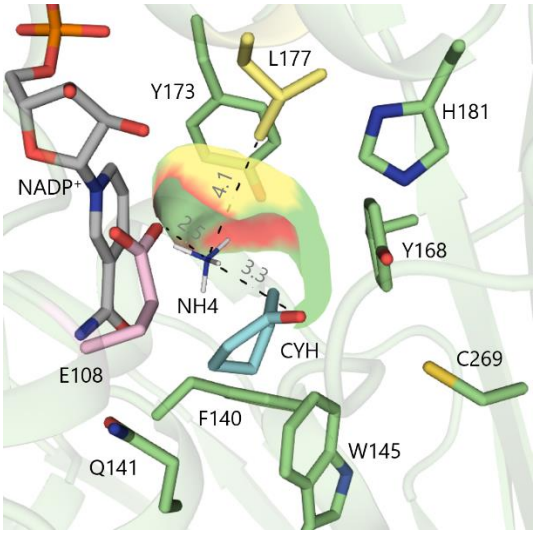


Figure 69. *Cfus*AmDH active site with docking of **45** and **11**. The surface of the upper part of the catalytic pocket is represented. L177 is highlighted in yellow and some of the P1-P21 positions are represented in green. $NADP^+$ and **45** (CYH) are colored in grey and blue, respectively. Distances are given in Angstroms (Å).

Among the restricted set of nat-AmDHs, this position shows very little diversity with mainly Leu and some exceptions such as Val in AmDH4 and other *Msme*AmDH P1-P21 homologs and Thr in OM-RGCAmDH2 and other *Cfus*AmDH P1-P21 homologs. In some cases, such as A0A078BAL0, A0A077ZQK8, and A0A078AU13, the automatic identification of a residue at this position by the ASMC pipeline even failed, meaning that no residue was found close enough to the P13 residue of *Cfus*AmDH template (Appendix 5). Within the expanded set, Leu remains in majority with also some clusters displaying Thr (*e.g.* cluster 40, 1438), Val (*e.g.* cluster 20), Ala (*e.g.* cluster 1438) or even Arg (*e.g.* cluster 1156) (Figure 66). To study the potential of targeting this residue to find nat-AmDHs active with primary amines, the docking of the amine substrate **38** and ethylamine (**255**) was carried out in *Cfus*AmDH-L177A vs WT

active site already containing the docked **45** (Figure 70). The parameters used for the docking are described in Experimental section, VI.2.2.4.

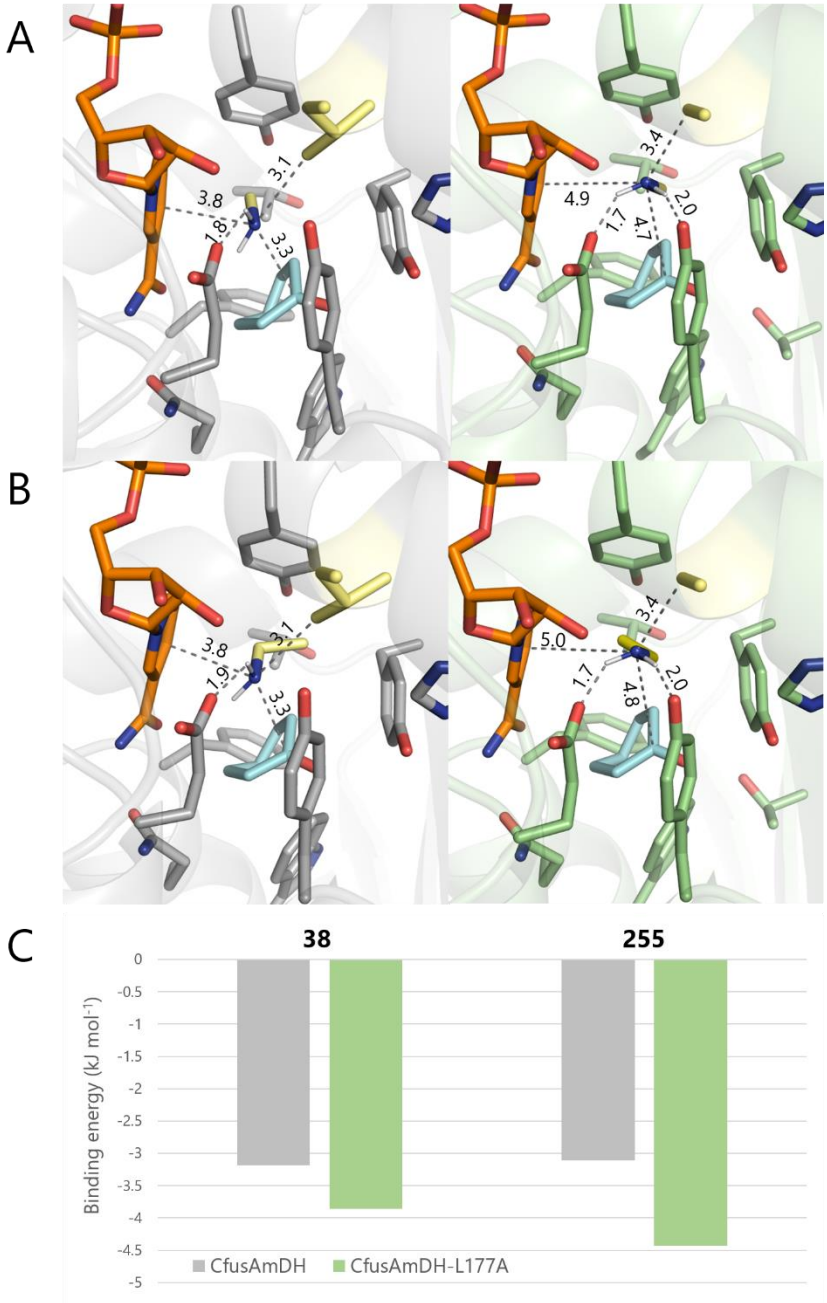


Figure 70. Docking of **38** and **255** amine substrates in *Cfus*AmDH and *Cfus*AmDH-L177A. The model used, made by Karine Bastard, contained the docked substrates **45** and **11**. The latter was removed to introduce primary amines: (A) **38** and (B) **255**. (C) shows the average binding energies (kJ mol^{-1}) associated to the 10 poses generated. In all cases, the configurations display very little variation. *Cfus*AmDH and *Cfus*AmDH-L177A are colored in grey and green, respectively. L/A177 and the amine pose with the lowest binding energy are highlighted in yellow. The distance are given in Angstroms (Å).

Interestingly, the two amine substrates can be docked in both the WT and the mutant, but in different conformations while still retaining a similar distance with P3. In *Cfus*AmDH, **38** and **255** were docked close to the nicotinamide ring (3.8 Å). With the space left by the mutation in *Cfus*AmDH-L177A, the amine substrates were systematically docked closer to the pocket ceiling, towards Y202 (P17) with one hydrogen pointing towards the hydroxyl moiety at 2.0 Å distance. This could contribute to the lowest binding energies obtained in the mutant than in the WT enzymes (Figure 70C). This positioning pushed them further away from **45** but the docking should have been done giving flexibility to the already docked **45** or by docking the two substrates at the same time for more relevancy.

This study was recently completed with the corresponding mutation in MATOUAmDH2 (L180A) carried out by Megan Bennett (Bennett *et al.*, 2022). The mutant was much more active than the WT enzyme to catalyze the reductive amination of **45** with **38** achieving 41% in 24 h at 25°C instead of 8%.

These results drove the selection of enzymes bearing short residues at P13 position among the ASMC2 clustering. Table 17 summarizes the selected enzymes for this objective.

Table 17. P1-P21 residues of the selected nat-AmDHs to accommodate larger amine substrates. *Cfus*AmDH is given as a reference and P13 is highlighted as the target position for this objective. The color code refers to the polarity and charge of the corresponding residue [blue: polar residues, yellow: hydrophobic residues, orange: aromatic residues, red: negatively charged residues, and green: positively charged residues (charges at physiological pH)]. The theoretical solubility was estimated using SoluProt web service (score > 0.5, yellow to green, corresponds to a soluble protein while a score < 0.5, yellow to red, corresponds to an insoluble protein).

	Cluster (Group)	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	**P13**	P14	P15	P16	P17	P18	P19	P20	P21	Solubility
Cfus AmDH	275 (G3/G4)	S	C	E	G	M	Q	F	W	V	Y	N	Y	L	H	P	S	Y	C	G	T	T	0.501
MGYP001470669209	20 (G3/G4)	S	S	E	G	H	Q	Y	W	A	W	N	F	V	Q	P	T	F	M	G	T	T	0.515
A0A365ZD63	20 (G3/G4)	S	S	E	G	H	Q	Y	W	A	W	N	Y	V	Q	P	S	F	M	G	T	T	0.398
A0A2G6MY80	40 (G3/G4)	S	C	E	G	Y	Q	S	W	S	Y	N	Y	T	H	P	G	Y	C	G	T	T	0.481
A0A2G6Q2D5	40 (G3/G4)	S	C	E	G	Y	Q	S	W	S	Y	N	Y	T	H	P	G	Y	C	G	T	T	0.637
A0A2R3MYQ8	40 (G3/G4)	S	C	E	G	Y	Q	S	W	S	Y	N	Y	T	H	P	G	Y	C	G	T	T	0.530
A0A3P2AE89	40 (G3/G4)	S	C	E	G	Y	Q	S	W	S	Y	N	Y	T	H	P	G	Y	C	G	T	T	0.600
A0A4R9C3Q3	40 (G3/G4)	S	C	E	G	Y	Q	S	W	S	Y	N	Y	T	H	P	G	Y	C	G	T	T	0.668
MGYP000142934144	40 (G3/G4)	S	C	E	G	Y	Q	S	W	S	Y	N	Y	T	H	P	G	Y	C	G	T	T	0.548
MGYP000879705604	40 (G3/G4)	S	C	E	G	Y	Q	S	W	S	Y	N	Y	T	H	P	G	Y	C	G	T	T	0.574
MGYP000893205724	40 (G3/G4)	S	C	E	G	Y	Q	S	W	S	Y	N	Y	T	H	P	G	Y	C	G	T	T	0.521
MGYP000996099441	40 (G3/G4)	S	C	E	G	Y	Q	S	W	S	Y	N	Y	T	H	P	G	Y	C	G	T	T	0.535
MGYP000619964244	275 (G3/G4)	S	C	E	G	M	Q	Y	W	V	Y	N	Y	V	H	P	S	Y	C	G	T	T	0.502
IGC-32	143 (G3/G4)	Y	C	E	G	F	Q	Y	W	S	Y	D	C	I	H	P	S	F	T	G	A	T	0.291
MGYP000230715187	143 (G3/G4)	S	C	E	G	Y	Q	Y	W	S	Y	D	C	I	H	P	S	F	T	G	A	T	0.360
A0A1A2V6I1	1438 (G6)	-	S	E	G	I	P	-	I	E	F	D	Y	A	P	S	G	I	T	G	G	T	0.270
A0A1A3HJ65	1156 (G6)	V	G	D	G	F	P	-	I	E	L	D	Y	A	L	G	L	G	M	S	L	T	0.282
MGYP001064279418	1156 (G6)	I	G	D	G	F	P	-	V	E	L	D	Y	A	M	A	K	A	M	D	L	T	0.287
MGYP001136017405	1375 (G6)	S	T	D	G	I	Q	I	W	A	A	N	F	A	A	P	T	P	F	V	G	T	0.294

These 18 enzymes will be screened in conversion assays against typical carbonyl substrates such as **45** and **57** with both **11** and **38**. The hits will be further tested with longer or more complex primary amines such as **255** or cyclopropylamine (**69**).

II.5.3 Selection of enzymes for accommodation of ketones harboring a carbonyl function on the third carbon atom of the chain

Along with the different assays and screenings carried out, some nat-AmDHs, either native or mutants, featured an ability to accept 3C ketones. These compounds harbors an ethyl group instead of the methyl or simple hydrogen found in 2C ketones and aldehydes, respectively, the major substrate classes of the nat-AmDHs family. As already mentioned in Chapter III, I.4.5, only a few examples of RedAms can facilitate the reductive amination of such compounds giving the (*R*)-enantiomer with very low *ee* (Mangas-Sanchez *et al*., 2020). Thus, it would be a real advantage to structurally rationalize the ability of some nat-AmDHs to accommodate these substrates and find enzymes with a still rarely described substrate spectrum. We searched into *Msme*AmDH, *Micro*AmDH, *Porti*AmDH and A0A646KJR1 in particular, a structural clue that could explain 3C ketones accommodation. Indeed, *Msme*AmDH, *Micro*AmDH, *Porti*AmDH could afford 39.1%, 51.2% and 46.5% conversion of hexan-3-one (**200**), respectively (Table 6). A0A646KJR1 was not characterized as purified enzyme but could already catalyze a substantial amount of **200** when used as cell lysates in the descriptive 96-well plate screening described in Chapter IV, II.3. The models of *Micro*AmDH and *Porti*AmDH were generated by homology with *Msme*AmDH (66% and 68% sequence identity, respectively) using Swiss Model (Bertoni *et al.*, 2017; Bienert *et al*., 2016; Guex *et al.*, 2009; Studer *et al.*, 2019; Waterhouse *et al*., 2018). As for A0A646KJR1, which displays a maximum of 41.7% sequence identity with either *Cfus*AmDH or *Msme*AmDH, the homology model generated was not reliable and it was visible with a shift in P17 position in the F204-I215 helix (D186-S210 in *Cfus*AmDH). Thus, its model was generated using ColabFold, the accessible tool using AlphaFold2 pipeline (Jumper *et al*., 2021; Mirdita *et al.*, 2022).

The presence of a second pocket below the catalytic residue P3 was particularly visible in these three enzymes and absent in other nat-AmDHs, such as *Cfus*AmDH and MATOUAmDH2 that do not accept 3C ketones or on a much lesser extent (Figure 71). The additional space available on the other side of the pocket, opposite to the previously described P5 and P8, might facilitate the accommodation of the ethyl moiety of such compounds. Four amino acids form the second pocket, namely, in *Msme*AmDH, L107 (PA), Q137 (P6), N202 (PB) and V203 (PC). Its formation is not fully understood as, in *Cfus*AmDH, there seems to be a similar space left by these positions compared to *Micro*AmDH, *Porti*AmDH, *Msme*AmDH or A0A646KJR1, but the connection with the main pocket does not occur due to a slightly shifted positions of P8 and P17 which block the entrance (Figure 71A).

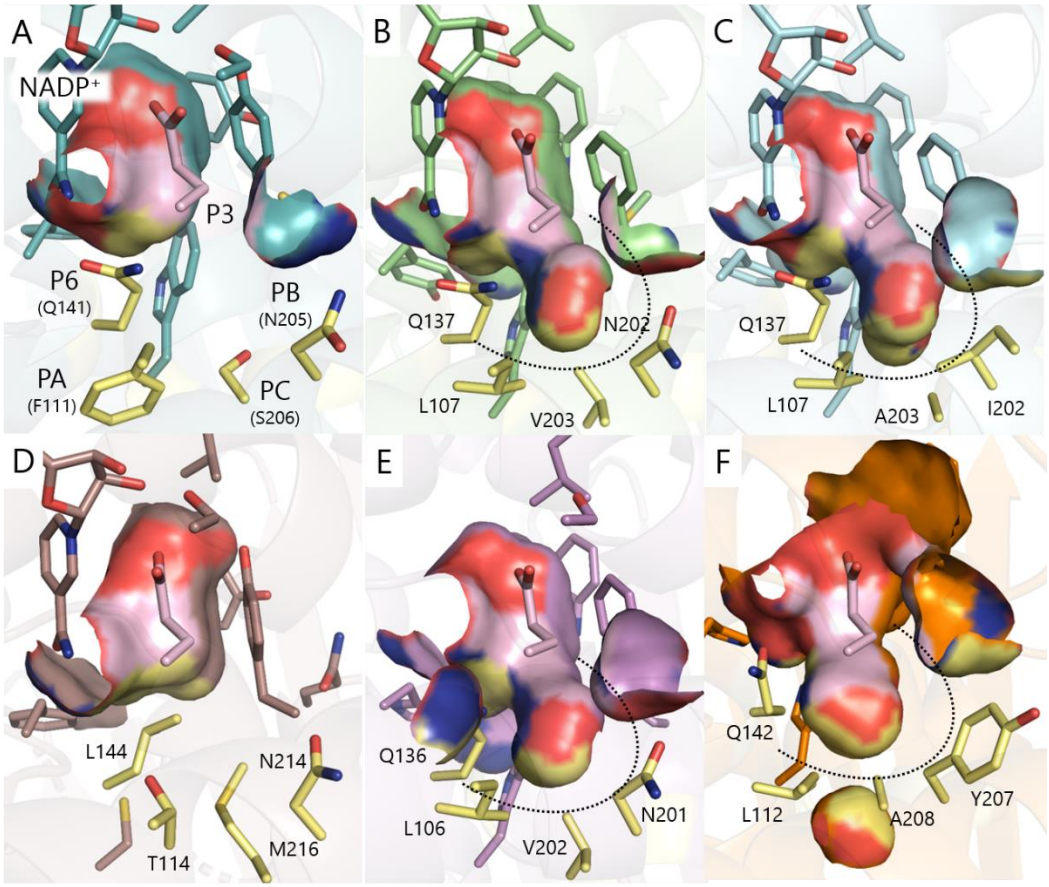


Figure 71. Second pocket in nat-AmDHs. (A) *Cfus*AmDH (PDB: 6IAU) (B) *Msme*AmDH (PDB: 6IAQ) (C) *Micro*AmDH (Model generated using Swiss Model and 6IAQ-chainA as template) (D) MATOUAmDH2 (PDB: 7R09) (E) *Porti*AmDH (Model generated using Swiss Model and 6IAQ-chainA as template) and (F) A0A646KJR1 (ColabFold model). P3 is colored in pink. PA, P6, PB and PC are highlighted in yellow. The pocket is represented with the surface mode on PA, PB, PC, P3, P5, P6, P8 and P13. RMSD (*Msme*AmDH-*Cfus*AmDH) = 0.987 Å; RMSD (*Micro*AmDH-*Cfus*AmDH) = 0.878 Å; RMSD (MATOUAmDH2-*Cfus*AmDH) = 1.376 Å; RMSD (*Porti*AmDH-*Cfus*AmDH) = 0.960 Å; RMSD (*Porti*AmDH-*Cfus*AmDH) = 1.012 Å.

Theoretically, this second pocket can also lead to variations in the stereochemistry if the longer substrate chain can be positioned towards P5-P8 or towards the second pocket, thus not presenting the same face of the carbonyl bond. Experimentally, the screening of *Porti*AmDH P5A and P8A mutants resulted in the major formation of (*2R*)-octan-2-amine (**256**) and (*2R*)-4-phenylbutan-2-amine (**55**) (49.3% and 49.6% *ee*, respectively). The accuracy of these *ee*s must be nuanced in view of the low conversions discussed (3.4% and 2.4%, respectively) and therefore the low UV areas taken into account. Accordingly, we selected 15 enzymes bearing at least one short residue at the positions PA, P6, PB or PC. Even though the bottleneck of the second pocket formation is still not clear, having enough space left in this area can reveal enzymes with still unreached high potential for 3C ketones with different sizes and substitutions. We also added in this selection some active site homologs of *Msme*AmDH, *Micro*AmDH, *Porti*AmDH and A0A646KJR1, namely part of members of the clusters 14 (5 members,

*Micro*AmDH and *Msme*AmDH cluster), 20 (33 members, *Porti*AmDH cluster) and 4 (6 members, A0A646KJR1 cluster). The clusters 4 and 14 does not appear in the ASMC2 tree as they comprise less than 15 members. All the enzymes selected for this objective are summarized in Table 18.

Table 18. P1-P21 residues of the selected nat-AmDHs to accommodate 3C ketones. *Cfus*AmDH is given as a reference and PA, P6, PB and PC are highlighted as the target positions for this objective. The color code used refers to the polarity and charge of the corresponding residue [blue: polar residues, yellow: hydrophobic residues, orange: aromatic residues, red: negatively charged residues, and green: positively charged residues (charges at physiological pH)]. The theoretical solubility was estimated using SoluProt web service (score > 0.5, yellow to green, corresponds to a soluble protein while a score < 0.5, yellow to red, corresponds to an insoluble protein).

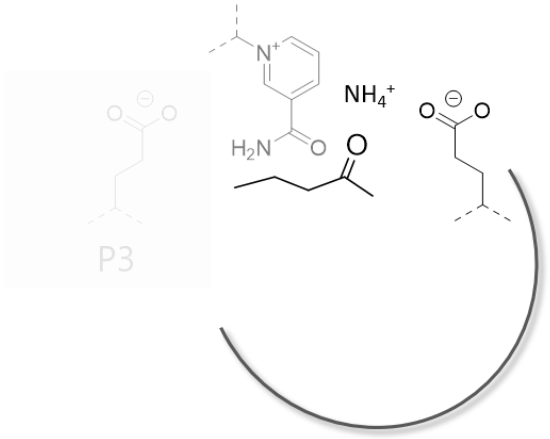
	Cluster (Group)	P1	P2	P3	**PA**	P4	P5	**P6**	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	**PB**	**PC**	P18	P19	P20	P21	Solubility
Cfus AmDH	275 (G3/G4)	S	C	E	F	G	M	Q	F	W	V	Y	N	Y	L	H	P	S	Y	N	S	C	G	T	T	0.501
A0A1X1SL76	14 (G3/G4)	S	S	E	L	G	Y	Q	Y	W	A	W	N	F	L	Q	P	T	F	I	A	M	G	T	T	0.486
A0A2U9PR62	14 (G3/G4)	S	S	E	L	G	Y	Q	Y	W	A	W	N	F	L	Q	P	T	F	I	A	M	G	T	T	0.253
I7GAY8	14 (G3/G4)	S	S	E	L	G	Y	Q	F	W	A	W	N	F	L	Q	P	T	F	N	V	M	G	T	T	0.317
MGYP000173707123	20 (G3/G4)	S	S	E	L	G	H	Q	Y	W	A	W	N	Y	L	Q	P	T	F	N	V	M	G	T	T	0.531
MGYP000211951848	20 (G3/G4)	S	S	E	L	G	H	Q	Y	W	A	W	N	Y	L	Q	P	T	F	N	V	M	G	T	T	0.452
MGYP000529792485	20 (G3/G4)	S	S	E	L	G	H	Q	Y	W	A	W	N	Y	L	Q	P	T	F	N	V	M	G	T	T	0.518
MGYP000689866863	20 (G3/G4)	S	S	E	L	G	H	Q	Y	W	A	W	N	Y	L	Q	P	T	F	N	V	M	G	T	T	0.469
MGYP001209562846	20 (G3/G4)	S	S	E	L	G	H	Q	Y	W	A	W	N	Y	L	Q	P	T	F	N	V	M	G	T	T	0.541
A0A519DQZ7	20 (G3/G4)	S	S	E	M	G	H	Q	Y	W	A	W	N	F	L	Q	P	T	F	N	A	M	G	T	T	0.53
MGYP001470669209	20 (G3/G4)	S	S	E	L	G	H	Q	Y	W	A	W	N	F	V	Q	P	T	F	N	V	M	G	T	T	0.515
A0A365ZD63	20 (G3/G4)	S	S	E	L	G	H	Q	Y	W	A	W	N	Y	V	Q	P	S	F	N	V	M	G	T	T	0.398
A0A229HGK2	4 (G3/G4)	T	A	E	L	G	H	Q	Y	W	L	W	N	F	V	Q	P	T	F	T	S	S	G	T	T	0.499
A0A1C6LG02	4 (G3/G4)	T	A	E	L	G	H	Q	Y	W	L	W	N	F	L	Q	P	S	F	Y	S	S	G	T	T	0.275
A0A1Q4UXH9	4 (G3/G4)	T	A	E	L	G	H	Q	Y	W	L	W	N	F	L	Q	P	S	F	Y	S	S	G	T	T	0.273
MGYP000970333896	233 (G3/G4)	S	C	E	L	G	Y	P	Y	W	S	Y	N	Y	L	H	P	S	Y	N	S	C	G	T	T	0.596
A0A138ZYM0	233 (G3/G4)	S	A	E	T	G	Y	S	A	W	S	Y	N	Y	L	H	P	S	Y	S	A	T	G	T	T	0.497
MGYP001097556939	257 (G3/G4)	S	C	E	L	G	Y	C	Y	W	S	Y	N	Y	L	H	P	S	Y	N	A	T	G	T	T	0.569
GUT_GENOME133859_01404	257 (G3/G4)	S	C	E	L	G	Y	C	Y	W	S	Y	N	Y	L	H	S	G	K	-	-	C	G	T	T	0.598
GUT_GENOME103710_02937	275 (G3/G4)	S	C	D	L	G	F	P	A	Y	A	Y	N	Y	L	H	A	I	P	N	A	S	G	N	I	0.499
MGYP001042153616	275 (G3/G4)	S	C	D	L	G	F	P	A	Y	A	Y	N	Y	L	H	A	I	P	N	A	S	G	N	I	0.49
A0A4S3B2N2	275 (G3/G4)	S	C	E	I	G	M	Q	F	W	V	Y	N	Y	L	H	P	S	Y	N	A	C	G	T	T	0.603
MGYP000075503774	275 (G3/G4)	S	C	E	T	G	M	E	F	W	V	Y	N	Y	L	H	P	S	Y	N	A	T	G	T	T	0.592
MGYP000827707224	275 (G3/G4)	S	C	E	T	G	M	E	F	W	V	Y	N	Y	L	H	P	S	Y	N	A	T	G	T	T	0.479
GUT_GENOME149041_01575	275 (G3/G4)	S	C	E	T	G	M	G	F	W	V	Y	N	Y	L	H	P	S	Y	N	S	C	G	T	T	0.551
MGYP001048001229	275 (G3/G4)	S	C	E	T	G	M	D	F	W	V	Y	N	Y	L	H	P	S	Y	N	S	C	G	T	T	0.548
GUT_GENOME088561_01464	642 (G3/G4)	S	C	E	V	G	M	Q	F	W	Y	Y	N	Y	L	H	P	S	Y	N	A	T	G	T	T	0.606
MGYP000457035951	642 (G3/G4)	S	C	E	V	G	M	Q	F	W	Y	Y	N	Y	L	H	P	S	Y	N	A	T	G	T	N	0.607
GUT_GENOME149657_00029	642 (G3/G4)	S	C	E	I	G	M	L	F	W	Y	Y	N	Y	L	H	P	S	Y	N	A	T	G	T	T	0.639
MGYP001082061073	642 (G3/G4)	S	C	E	T	G	M	Q	F	W	Y	Y	N	Y	L	H	P	S	Y	N	A	T	G	T	T	0.666

The 29 selected enzymes will be tested towards a range of aliphatic aldehydes, 2C and 3C ketones but also towards the particularly rigid acetophenone (**39**) which was only accepted by *Porti*AmDH, *Msme*AmDH and *Micro*AmDH so far (Table 4).

II.5.4 Selection of enzymes to reach (*R*)-stereoselectivity

The nat-AmDH family was mainly described to provide a (*S*)-stereoselectivity, except for particular substrates or enzymes such as *Porti*AmDH which achieved 47.4% conversion of hexan-3-one (**200**) and 23.9% *ee* for the suspected (*3S*)-hexan-3-amine (**202**) and a conversion of 21.9% of heptan-3-one (**234**) and 54.2% *ee* for the suspected (*3R*)-heptan-3-amine (**257**) (Chapter III, I.4.5). Regardless, it would be interesting to have a structural explanation for this stereoselectivity and expand the family with new members providing the reverse one, thus providing nat-AmDH templates with on demand

characteristics. This study requires to work on the symmetry of the cavity even though there is still no confirmation of the ketone positioning resting from 3D structures. Also, the docking results are not easy to correlate with the experimentally obtained stereoselectivity. Theoretically, the reversal of ceiling and floor (Figure 23) symmetry seems difficult to reach in our case because it would require a big enlargement of the pocket ceiling that the overall nat-AmDHs folding and dynamic could not afford. However, the "right to left" symmetry can occur with the second pocket described in the previous part. Another possibility is to move the catalytic residue to the other side of the pocket at equidistance with the nicotinamide ring of the cofactor to still enable the reaction (Scheme 34).



Scheme 34. Schematic view of a catalytic glutamate located on the opposite side of P3 in the active site. As an example, pentan-2-one (**112**), ammonia (**11**) and a part of the nicotinamide cofactor are also represented.

In nat-AmDHs, P20 or P5 are at a correct distance to the nicotinamide ring to play the role of "new" catalytic position if occupied by negative Glu or Asp. Three examples of nat-AmDHs found in ASMC2 are given in Figure 72. Among all the structures considered, only one displayed a Glu or Asp at an equivalent position of P5, namely MGYP000037226974, which harbors an additional residue between P5 and P6, named P5' here, as what is found in AmDH4 (Chapter III, III). The other candidate enzymes harbor a Glu or Asp residue at position P20.

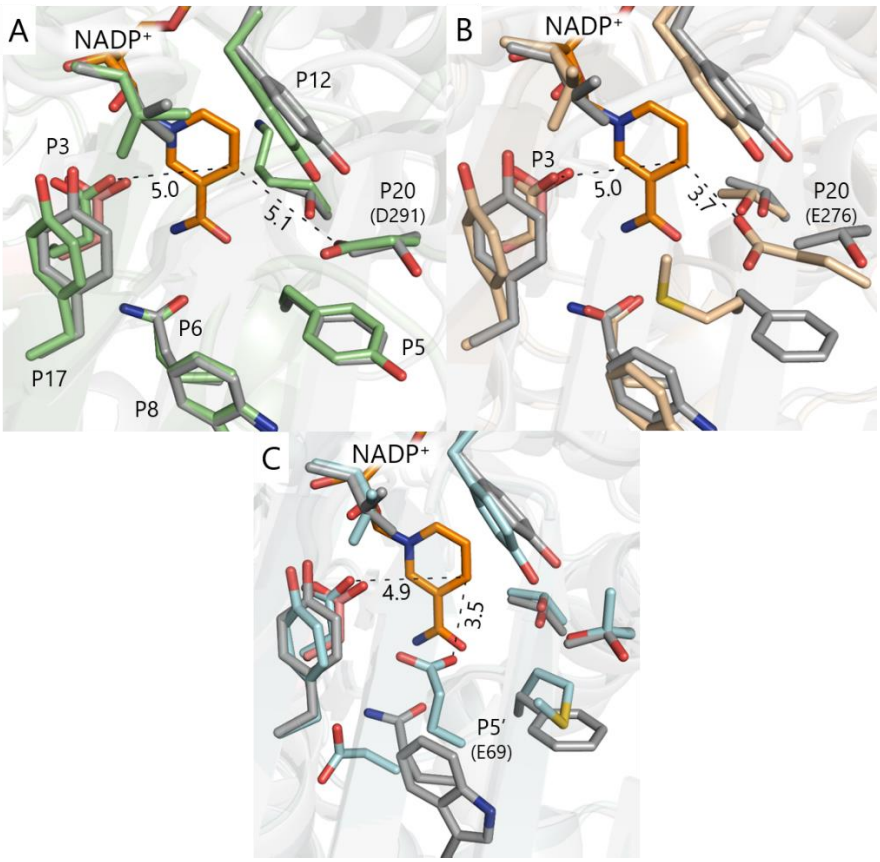


Figure 72. Examples of nat-AmDHs with potential catalytic residue at P5′ or P20. (A) MGYP000357504158 (green) with Asp at P20 (B) GUT_GENOME186969_00552 (beige) with Glu at P20 and (C) MGYP000037226974 (blue) with Glu between P5 and P6 (P5′) aligned with *Cfus*AmDH (PDB: 6IAU), shown in grey. P3 and NAD(P)$^+$ are highlighted in pink and orange, respectively. RMSD (MGYP000357504158-*Cfus*AmDH) = 0.403 Å; RMSD (GUT_GENOME186969_00552-*Cfus*AmDH) = 0.431 Å; RMSD (MGYP000037226974-*Cfus*AmDH) = 0.377 Å.

An Asp at P20 seems to be the most favorable system as the distance D291 (OD1)-NADP(C4N) is similar to E106 (OE2)-NADP(C4N) (5.1 and 5.0 Å, respectively). In the other systems, the longer Glu side chain bring the carboxyl moiety at less than 4 Å of C4N that could be unsuitable for **11** to position and for the reaction to proceed. However, these observations are based on homology models generated using *Cfus*AmDH as template with which they shares only 41.5 - 44.5% sequence identity that might bias the side chain positioning. These models may not be reliable enough to trust the side chain orientation. All the systems Glu at P5′ or P20 or Asp at P20 were considered for the selection. It is clear that the presence of the conserved Glu at P3 indicates that the selected enzymes are more likely to use this residue for **11** fixation and activation than the additional one at P5′ or P20. Thus, if the first hits display a substantial (*R*)-amine formation, it could be interesting to remove Glu at P3 to potentially achieve full conversion in (*R*)-amine. Complementary docking experiments of (2*S*)-pentan-2-amine (**173**) and (2*R*)-pentan-2-amine (**258**) were carried out in the three examples abovementioned to estimate the amine positioning towards the original P3 or P20/P5′ and try to predict the stereoselectivity of the reaction for

this substrate. As already explained in Chapter II, IV.1.2 for the *in silico* study on the short (hydroxyl)amines, the number of poses bearing the correct amine moiety orientation were considered and related to their binding energy (Figure 73) (Experimental section, VI.2.2.5).

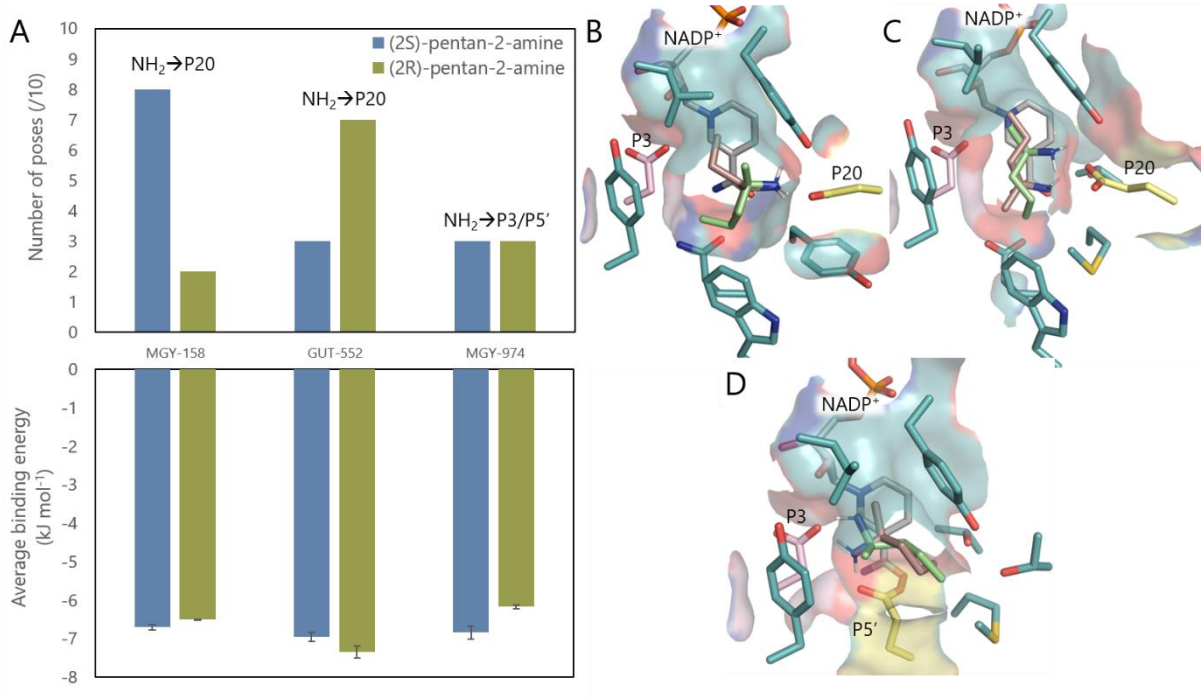


Figure 73. Docking of (2*S*)-pentan-2-amine (**173**) and (2*R*)-pentan-2-amine (**258**) in MGYP000357504158 (MGY-158) GUT_GENOME186969_00552 (GUT-552) and MGYP000037226974 (MGY-974) with potential catalytic residues at P5′ or P20. (A) gives the number of "correct" poses and the average binding energies (kJ mol^{-1}) associated. The bars represent the standard deviation compared to the average. 3D-representation of the top pose of the docked amines in (B) MGY-158 (B) GUT-552 and (C) MGY-974.

Interestingly, in MGY-158 and GUT-552, all the poses, either for **173** or **258**, were oriented towards P20 even in the presence of Glu at P3. Also, MGY-158 seemed to better accommodate the (*S*)-amine with 8 poses vs 2 for the (*R*)-amine, and slightly lower binding energy with -6.7 kJ mol^{-1} for the (*S*)-amine and -6.5 kJ mol^{-1} for the (*R*)-amine. This might be due to the side chain positioning of Asp or Glu at P20 presenting two O-atoms, thus leading to more favorable interactions with the amine moiety than the single O-atom provided by Glu at P3. On the contrary, GUT-552 favored the (*R*)-amine accommodation with 7 poses (-7.3 kJ mol^{-1}) vs 2 (-6.9 kJ mol^{-1}) for the (*S*)-amine. In MGY-974, both (*S*)- and (*R*)-amines were oriented with the amine moiety towards P3 and/or P5′. In comparison, the docking of (2*R*/*S*)-butan-2-amine (**182** and **193**) in *Cfus*AmDH, *Msme*AmDH or *Micro*AmDH, described in Chapter II, IV.1.2, led to -5.3 - -6.1 kJ mol^{-1} binding energies, then supposedly slightly less favorable.

The enzyme selection for this objective is given in Table 19.

Table 19. P1-P21 residues of the selected nat-AmDHs to switch the catalytic residue position. *Cfus*AmDH is given as a reference and P5' and P20 are highlighted as the target positions for this objective. Colors refer to the polarity and charge of the corresponding residue [blue: polar residues, yellow: hydrophobic residues, orange: aromatic residues, red: negatively charged residues, and green: positively charged residues (charges at physiological pH)]. The theoretical solubility was estimated using SoluProt web service (score > 0.5, yellow to green, corresponds to a soluble protein while a score < 0.5, yellow to red, corresponds to an insoluble protein).

	Cluster (Group)	P1	P2	P3	P4	P5	**P5'**	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	**P20**	P21	Solubility
Cfus AmDH	275 (G3/G4)	S	C	E	G	M		Q	F	W	V	Y	N	Y	L	H	P	S	Y	C	G	T	T	0.501
MGYP000457694760	257 (G3/G4)	S	C	E	G	Y		C	Y	W	S	Y	N	Y	L	H	E	M	K	C	G	E	K	0.612
MGYP000036139749	670 (G3/G4)	S	C	E	G	M		Q	F	W	Y	Y	N	Y	L	H	P	S	Y	T	G	E	E	0.448
MGYP000792632202	632 (G3/G4)	S	C	E	G	M		Q	F	W	Y	Y	N	Y	L	H	A	S	Y	T	G	E	Y	0.675
MGYP000852990308	143 (G3/G4)	S	C	E	G	Y		Q	Y	W	S	Y	N	Y	L	H	P	S	Y	C	G	E	T	0.695
MGYP000784286641	143 (G3/G4)	S	C	E	G	Y		Q	Y	W	S	Y	N	Y	L	H	P	S	Y	C	G	E	V	0.742
MGYP000357504158	143 (G3/G4)	S	C	E	G	Y		Q	Y	W	S	Y	N	Y	L	H	P	S	Y	C	G	D	K	0.694
MGYP000833766750	143 (G3/G4)	S	C	E	G	Y		Q	Y	W	S	Y	N	Y	L	H	P	S	Y	C	G	D	W	0.755
MGYP000769455001	143 (G3/G4)	S	C	E	G	H		Q	Y	W	S	Y	N	Y	L	H	P	S	Y	C	G	E	Y	0.698
MGYP000823225546	143 (G3/G4)	S	C	E	G	Y		Q	Y	W	S	Y	N	Y	L	H	P	S	Y	C	G	D	E	0.719
MGYP000737964612	143 (G3/G4)	S	C	E	G	Y		Q	Y	W	S	Y	N	Y	L	H	A	S	V	Q	Q	D	M	0.651
MGYP000712401467	143 (G3/G4)	S	C	E	G	Y		Q	Y	W	T	Y	N	Y	L	H	R	N	K	-	-	E	I	0.72
MGYP001123837121	143 (G3/G4)	S	C	E	G	Y		Q	Y	W	S	Y	N	Y	L	H	A	R	Q	L	-	E	L	0.753
MGYP000192187441	143 (G3/G4)	S	C	E	G	Y		Q	Y	W	S	Y	N	Y	L	H	V	A	S	Y	E	E	I	0.733
IGC-105	143 (G3/G4)	S	C	E	G	Y		Q	Y	W	S	Y	N	Y	L	H	V	A	S	Y	D	E	I	0.764
MGYP001172620759	275 (G3/G4)	S	C	E	G	M		Q	F	W	T	Y	N	Y	L	H	P	A	Y	C	G	E	D	0.389
MGYP000785448156	275 (G3/G4)	S	C	E	G	M		Q	F	W	T	Y	N	Y	L	H	P	A	Y	C	G	E	P	0.491
MGYP001296945020	275 (G3/G4)	S	C	E	G	M		Q	Y	W	V	Y	N	Y	L	H	P	S	Y	C	G	D	E	0.524
MGYP000619909836	275 (G3/G4)	S	C	E	G	M		Q	Y	W	V	Y	N	Y	L	H	P	S	Y	C	G	D	M	0.504
MGYP000802697525	275 (G3/G4)	S	C	E	G	M		E	F	W	V	Y	N	Y	L	H	P	S	Y	C	G	E	Y	0.712
MGYP000895013343	275 (G3/G4)	S	C	E	G	M		E	F	W	V	Y	N	Y	L	H	P	S	Y	C	G	E	S	0.653
MGYP000080283555	275 (G3/G4)	S	C	E	G	M		E	F	W	V	Y	N	Y	L	H	P	S	Y	C	G	E	R	0.36
MGYP001020886770	275 (G3/G4)	S	C	E	G	M		E	F	W	V	Y	N	Y	L	H	P	C	Y	C	G	D	C	0.662
GUT_GENOME248805_00009	275 (G3/G4)	S	C	E	G	M		E	F	W	V	Y	N	Y	L	H	P	S	Y	C	G	D	C	0.495
MGYP000116736805	275 (G3/G4)	S	C	E	G	M		E	F	W	V	Y	N	Y	L	H	P	S	Y	C	G	D	G	0.701
MGYP000683436341	275 (G3/G4)	S	C	E	G	M		E	F	W	V	Y	N	Y	L	H	P	S	Y	D	-	E	T	0.495
GUT_GENOME186969_00552	275 (G3/G4)	S	C	E	G	M		E	F	W	V	Y	N	Y	L	H	P	S	Y	C	G	E	T	0.355
MGYP000585659698	275 (G3/G4)	S	C	E	G	M		E	F	W	V	Y	N	Y	L	H	P	S	Y	C	G	D	T	0.318
MGYP000037226974	275 (G3/G4)	S	C	E	G	M	E	D	L	G	V	Y	N	Y	L	H	P	S	Y	C	G	T	T	0.602
MGYP000882735950	33 (G3/G4)	S	C	E	G	Y		Q	F	W	S	Y	N	Y	L	H	P	S	F	C	G	D	T	0.521
MGYP000274952423	44 (G3/G4)	S	C	E	G	Y		Q	F	W	S	Y	N	Y	L	H	P	S	Y	C	G	E	Y	0.636

The 30 selected enzymes will be assayed against a range of typical nat-AmDHs substrates such as pentan-2-one (**112**) or cyclohexanone (**45**) also with methyl substituents to bring the chirality and eventually against the functionalized 1-hydroxybutan-3-one (**190**) that could benefit from the charged Glu or Asp at P5' or P20, even though the switch of stereochemistry will not occur in this case.

II.5.5 Final selection of key novel nat-AmDHs

A total of 92 enzymes were selected for the four objectives. This set also includes two sequences from MATOU database that was not included in the extended set of reference nat-AmDHs at the time of the AMSC2 running. Among the 92 sequences, 29 had to be redesigned due to a truncated N-ter or C-ter compared to the conserved nat-AmDH folding. This is one of the limitations of working with metagenomic databases that sometimes hold incomplete sequence data. As it was previously done to design MATOUAmDH1 and MATOUAmDH2 (Caparco *et al.*, 2020), the missing sections were designed from a consensus segment found in the complete sequences sharing the highest sequence identity with the target (Appendix 74). In some cases, when the identity was particularly high with one sequence, the

missing sections were just copied from the template. Experimentally, we purchased the corresponding synthetic genes and primers necessary to amplify and produce both the designed and the short version of the enzymes. The enzyme production, from cloning to cell lysates retrieval, was done by Adrien Debard and the successfully produced enzymes will be screened soon.

III. Conclusion

Within the scope of the MODAMDH project, the extension of the set of reference nat-AmDHs from 2,011 to 17,039 sequences gave access to a higher diversity of templates to explore. From the 92 selected enzymes, only nine were present in the initial ASMC1. The results of the screening will inform us on the viability of the structure-guided selection done here, even if issues with their production will perhaps prevent such conclusions. In all cases, this will give more give us some starting clues to consider for protein engineering. The selection of enzyme candidates for the discussed features, for the moment scarcely or not observed in the nat-AmDH family, has been done manually, even if based on the P1-P21 results from the automatic pipeline ASMC. Such a work would benefit to be done automatically to avoid missing some interesting enzymes, for example for the *in silico* detection of the second pocket for activity towards 3C ketones. Obviously, MD can also support some hypotheses and help the selection. This Chapter also demonstrated the high interconnections between the different studies brought up along the previous chapters. In fact, the accumulation of structural and experimental data about the currently known nat-AmDHs and mutants was essential for the selection.

CONCLUSION AND PERSPECTIVES

The work done during the three years of my PhD project mostly answered to the objectives identified for the nat-AmDH family structural study, exploration and diversification exposed in Chapter I, III.

Over the structural analysis of some either characterized or still uncharacterized members, the family was shown to exhibit a high conservation of the global fold with a characteristic open-to-closed dynamic. Some key hotspot segments displaying higher diversity in their 3D arrangement were identified and could be interesting to investigate as they could impact the enzyme stability, expressability or dynamic velocity. The analysis of the amino acid occurrence at some positions among the family combined with a SDM strategy enabled to further understand the nat-AmDH mechanism, that is still only hypothesized. P3 was already known to be the catalytic residue for ammonia (**11**) activation but we highlighted the importance of having bulky amino acids at P12 for active site closing and substrate positioning, at least in the analogs of the reference *Cfus*AmDH. P10 was previously hypothesized to ensure the substrate stabilization within the active site mainly thanks to a H-donor hindered residue. Yet, our study revealed it could accept increasingly shorter and unfunctionalized amino acids without suffering complete deactivation of *Cfus*AmDH. This position would be particularly interesting to investigate in other nat-AmDH scaffolds for substrate scope expansion.

In a collaboration with the team of Prof. Bommarius, we studied the stability of a set of nat-AmDHs through the monitoring of their specific activity under various storage conditions. The temperature appeared to be the critical parameter but two nat-AmDHs, IGCAmDH5 and *Acol*AmDH, stood out with practically no activity loss after 6 days at 37°C. The relationship between experimental stability and structural features could not be completely rationalized. It could be interesting to pursue this study with experts in protein stability, but this sub-project is, for the moment, not a priority for the team. The writing of the related publication is in progress, mainly in the hands of the collaborators who led this study.

Our knowledge on the nat-AmDHs substrate promiscuity, was enriched by the study of their biocatalytic potential towards short prochiral ketones bearing methoxy and hydroxy moieties [butan-2-one (**134**), 1-methoxypropan-2-one (**189**), etc.], short aldehydes with aliphatic and aromatic cyclic substituents [furfural (**57**), benzaldehyde (**31**), etc.] but also 3C aliphatic ketones [hexan-3-one (**200**)]. Some nat-AmDHs appeared to be quite efficient towards such substrates, thus complementing the

biocatalytic toolbox of enzymes for reductive amination, rather limited for these substrates. Molecular docking experiments, carried out with the amines formed from the reductive amination of key short prochiral ketones, gave first insights into the characteristic (*S*)-enantioselectivity of the nat-AmDH family. In the future, more downstream processes for the synthesis of some of these key amines would be interesting to carry out in collaboration with other teams. Some project discussions in this direction are underway in the team. A very preliminary study was also performed to estimate the nat-AmDHs potential for reductive sulfidation of carbonyl compounds in presence of the sulfur donor Na_2S. The catalytical thiol formation was difficult to attest due to a high proportion of spontaneous reaction in the conditions tested. Some alternatives were proposed to improve the thiol formation and monitoring. Clear proofs of nat-AmDH-mediated sulfidation need to be obtained to potentially lead to a sub-project on this secondary activity. This may be carried out internally in the team in the coming months.

One of the main objectives was to diversify the nat-AmDHs substrate scope either by protein engineering or biodiversity mining. Inspired by the diversity found within the family and supported by molecular docking experiments, we identified P5 and P8 positions as hotspots for active site expansion. These positions were first mutated in shorter Ala on the reference enzyme *Cfus*AmDH and then transposed into nine other nat-AmDHs displaying similar scaffolds but slightly different P1-P21 residues in the active site. Ultimately, thirteen of these mutants were more active than the native enzymes towards C6 to C10 aldehydes or 2C ketones under non-optimized reaction conditions. A structural characterization of the top mutant *Cfus*AmDH-W145A was done thanks to our collaboration with the team of Prof. Grogan (University of York) for the crystallization and structural resolution and with Gwenaëlle André-Leroux (INRAE) for MD simulations. The combination of both approaches gave more insights into the enzyme global dynamics and brought partial explanation for the altered mutant substrates and cofactor specificity of the mutant. An extended biodiversity mining, within the MODAMDH project, has also contributed to fulfill these objectives of nat-AmDH substrate scope diversification, in addition to the amine substrate scope and the reversal of the typical (*S*)-enantioselectivity. All structural and experimental data collected thus far by Dr. Eddy Elisée were used to refine the selection of nat-AmDH sequences retrieved from genomic and metagenomic databases to reach these goals. The 92 enzymes selected are now being produced as cell lysates to test their activity towards key substrates. The MODAMDH strategy and some of the experimental results obtained are planned to be submitted to a journal such as *Nature Catalysis* early 2023. Depending on the identification of some nat-AmDHs with notable alternative features compared to the already characterized ones, a synthetic application could be carried out by an upcoming student.

Another strategy of *in vivo* evolution of enzymes catalyzing the reductive amination of carbonyl compounds was attempted taking advantage of the automated continuous evolution devices available at the Genoscope. The first evolution system was based on the ammonia (**11**) released from the oxidative deamination of amines by nat-AmDHs to supply the evolving strains in nitrogen. The activity of nat-AmDHs in the deamination way was probably too limited to provide enough nitrogen needed for cells growth. The second less demanding evolution system utilizes the same way of reaction catalyzed by *Asp*RedAm to release the NADPH required for the growth of an in-house NADPH-autotrophic strain. The evolution is still underway but no remarkable event has yet been noticed during the evolution that could allude an enhancement of *Asp*RedAm activity. Such an *in vivo* strategy for improvement of heterologous AmDH activity in *E. Coli* would be still challenging to lead and may require a full-time work on it. Indeed, it will presumably need to combine *in vivo* mutagenesis and evolution in GM3 to manage to unlock the critical points, in particular sufficient initial activity to allow sufficient growth to start the evolution.

My PhD work was part of the multiple sub-projects that are jointly conducted within the AmDH project to fully exploit the potential of these enzymes. Notably, these were combined with an ene-reductase in a bi-enzymatic cascade to access chiral amines from unsaturated ketones and aldehydes. This work arose from a collaboration with the team of Caroline E. Paul (T.U. Delft) (Jongkind *et al.*, 2022). Another PhD student, Josermarco Mendoza Avila is currently supervised by Dr. Carine Vergne-Vaxelaire and Prof. Francesco Mutti (University of Amsterdam) in collaboration with Dr. Louis Mouterde (URD ABI, AgroParisTech) to modify *E. coli* cells for the use of AmDHs as whole-cell biocatalysts for amine synthesis. Indeed, compared to the general use of enzymes as cell lysates or purified enzymes generally done within the L2BMS laboratory, the use of whole-cells have several advantages of lower production cost, higher enzyme stability and usability in synthesis with easier operational work up processes.

Another improvement of the nat-AmDHs that was proposed is a switch of cofactor preference from NADPH to NADH, for which various recycling systems are available, thus being more attractive for the industries. Preliminary studies were carried out by Dr. Eddy Elisée to determine the sequence profile associated to the cofactor binding site and that could help to predict the cofactor preference from an nat-AmDH protein sequence. Also, Dr. Carine Vergne-Vaxelaire deeply looked at the diversity of residues found at the cofactor binding site within the nat-AmDH family to correlate structural features with experimental behavior with the provided cofactors. Both of these studies could help to rationalize the selection of novel enzymes with native preference for NADH or to force the switch through protein engineering.

Thanks to this research work, various advances on nat-AmDH enzymes have been achieved. Some of the results open the door to many other scientific questions that would be interesting to address, as mentioned at the end of each of the concluding paragraphs.

All the works conducted within the AmDH project highlight the pluridisciplinary approaches that are necessary in biocatalysis but also in scientific research, in general. The involvement of people with different backgrounds and expertise including bioinformaticians, chemists and biologists, is the only alternative to address the current challenges and build a more sustainable chemical industry.

EXPERIMENTAL SECTION

I. Generals

I.1 Chemicals

All the chemicals were purchased from commercial suppliers (Merck®/Sigma Aldrich®, Enamine®, Thermo Fisher Scientific®, Alfa Aesar®, VWR®, Combi-Blocks®) and used without additional purification. (4*S*)-4-aminopentanoic (**167**) acid and (2*R*,4*S*)-2,4-diaminopentanoic acid (**165**) were previously synthesized in our laboratory. Glucose dehydrogenase GDH-105 (GDH) was given by Codexis®. Ammonium formate buffer were produced from substances purchased from Sigma Aldrich® and adjusted to the desired pH value with ammonium hydroxide 30%. Synthetic genes and primers used for cloning or SDM were purchased from Twist Bioscience®.

I.2 Equipment

Molecular biology experiments were performed using the following equipment: Invitrogen Qubit® fluorometer for DNA titration, QIAcube (QIAGEN®) for plasmid DNA Miniprep and BioRad MicroPulser™ for cells electroporation.

Enzymes production were carried out using the following equipment: Jouan incubator for cells growth on gel plates at 37°C, Innova 4300 Incubator Shaker (New Brunswick Scientific®) for cells growth in liquid medium at 37°C, INFORS HT Multitron Standard for cells growth at given temperature and enzymes induction at 20°C, SONICS VibraCell™ equipped with a Tapered Microtip 3 mm conical microprobe for sonication. Invitrogen Novex Minicell and Xcell Sure Lock for SDS-page gels and Spectro max 384 Plus (Molecular Devices) for enzymes titration using Bradford technic.

Bioscreens were performed using Bioscreen C (Thermo Scientific Labsystems).

Biocatalytic reactions were performed under agitation using Eppendorf Thermomixer C. UHPLC analyses were performed on a UHPLC U3000 RS 1034 bar system (Thermo Fisher Scientific®) equipped with a DAD3000 diode array detector and a MSQ Plus™ Single Quadrupole Mass Spectrometer using a Kinetex® F5 (Phenomenex®) column (100 × 2.1 mm; 1.7 µm). GC-FID analyses were performed on a Gas Chromatograph Trace 1300 (Thermo Fisher Scientific®) equipped with an autosampler injector AI/AS1310, a Flame Ionization Detector (FID) and a H_2 Generator Alliance (FDGSi). Samples were injected on a TG-35MS AMINE (Thermo Fischer Scientific®) column (30 m; 0,25 mm; 1 µm). Spectrophometric assays were recorded on a Safas UVMC2 (Safas®) thermostated with a refrigerated/heating circulator Polystat36 (Thermo Fisher Scientific®) or CORIO CD-200F cryostat (JULABO GmbH®) and using

microcells high-precision cell quartz with 6 or 10 mm light path (Hellma Analytics). GC-HS-MS analyses were performed on a Trace 1300 coupled to an ISQ 7000 VPI single quadrupole mass spectrometer (Thermo Fisher Scientific®). The instrument was equipped with a 30 m × 0.25 mm × 0.25 µm DB-624-UI column (Agilent J&W®), a split/splitless injector, and an automatic sampler TriPlus RSH coupled to a Headspace tool. NMR spectra were recorded on a Bruker (Bruker) 600 MHz spectrometer (Evry University, France) for ^{1}H and ^{13}C experiments. Chemical shifts (expressed in ppm) of ^{1}H spectra were referenced to the solvent peaks δ(H) = 4.65 for D_2O.

The diffraction of *Cfus*AmDH-W145A crystals was tested using a Rigaku Micromax-007HF fitted with Osmic multilayer optics and a MARRESEARCH MAR345 imaging plate detector.

I.3 Softwares

Multiple sequence alignment (MSA) and sequence identity matrices were performed using Clustal Omega webservice. P1-P21 conservation logo were generated using WebLogo webservice.

Enzymes structural analyses were performed using the following softwares: SwissModel webservice (Waterhouse *et al*., 2018), MODELLER software (Fiser *et al*., 2000; Fiser *et al*., 2003; Martí-Renom *et al*., 2000; Sali *et al*., 1993; Webb *et al*., 2016) for the generation of homology models, ColabFold webservice (Mirdita *et al*., 2022) for the generation of template-free models, PyMol Schrödinger LLC for 3D visualization, Autodock for molecular docking after ligand PDB file generation using the free demo of Corina Molecular webservice, CAVER WEB webservice (Stourac *et al*., 2019) for pocket volume calculation. MD simulations were performed using GROMACS software (version 2021.3) (Abraham *et al*., 2015; Berendsen *et al*., 1995; Hess *et al*., 2008; Lindahl *et al*., 2001; Pronk *et al*., 2013; Van Der Spoel *et al*., 2005; Bekker *et al*., 1993). Normal modes were generated using elNémo webservice [Suhre *et al*., 2004 (A), Suhre *et al*., 2004 (B)].

Enzyme solubility predictions were performed using SOLUPROT 1.0 webservice (Hon *et al*., 2021).

Molecular biology experiments were supported by the following softwares: ExPasy Translate Tool (Gasteiger *et al*., 2003) for nucleotidic sequence translation into proteic sequence, OligoCalc (Kibbe, 2007) and GC content Calculator for primers design and BioEdit software (Hall, 1999) for multiple sequence alignment and sequencing data analyses. Enzyme titration were performed using SoftMax Pro 5.4 (Molecular Devices).

Bioscreens were performed using BioScreener software (Growth Curves, USA).

Spectrophotometric assays were recorded on SAFAS SP2000v7 software. UHPLC and GC-FID assays were recorded and analyzed using Chromeleon software. GC-HS-MS assays were recorded using XCalibur software. Sigma Plot software was used for curve fitting and calculation of kinetic parameters.

ChemDraw was used as a drawing tool for molecules and reactions schemes.

II. Enzyme production

II.1 General method for enzyme cloning

The primers were designed with addition of specific extension for cloning into our in-house plasmid pET22b(+) (Novagen) modified for ligation independent cloning (LIC). The forward primers also harbor a His-tag after the methionine for enzyme purification. Gene amplifications were performed using Q5 high fidelity DNA polymerase (QIAGEN®) following the instructions of the supplier.

The purified PCR product was then incubated with T4 DNA polymerase in presence of dGTP (25°C, 20 min) for preparation of the LIC insert. After ligation into the LIC-treated pET22b(+) plasmid, the construction was inserted into NovaBlue(DE3) Competent Cells. A colony PCR using Dream Taq DNA polymerase in presence of primers designed from the vector sequence on both sides of the cloning site, enabled to pick up the clones containing the insert for growth on LB medium and subsequent MiniPrep and sequence checking. PCR mixtures and thermal cycles were done following the instructions of the supplier.

II.2 General method for site-directed mutagenesis

One and multiple SDM were carried out using QuikChange II XL Site-directed and QuikChange Multi-Site directed Mutagenesis Kits, respectively, from Agilent® following the protocols provided by the supplier. Primers for site-directed mutagenesis were design with the appropriate change in nucleotidic sequence to introduce the mutation into the AmDH gene-containing-pET22b(+) construction. The triplet chosen was the one displaying the highest usage frequency in *E. coli*.

II.3 General method for enzyme induction

The verified constructions were then transformed into in-house *E. coli* BL21 C+ competent cells for induction. These were grown on Terrific Broth (TB) medium containing 0.5 M sorbitol, 5 mM betaine and 100 g mL^{-1} carbenicillin at 37°C until reaching an OD_{600} of 0.8-1.2. IPTG was added at 0.5 mM final concentration to start the protein induction and the cells were further grown overnight at 20°C. The culture medium was split into 50 mL culture medium aliquots. After centrifugation, the pellet was stored at -80°C for at least 4 h to facilitate cells membrane breakage. The frozen pellets were then resuspended

in 3 mL lysis buffer (50 mM potassium phosphate buffer, pH 7.5, 50 mM NaCl, 10% glycerol) containing 1 mM Pefabloc®SC and 5 µL Lysonase TM bioprocessing reagent (Novagen®), agitated for 30 min at RT and sonicated using Ultrasonic Processor. After centrifugation, the cell-free extract was recovered and stored at -80°C.

II.4 General method for small-scale purification of enzymes

After cloning or SDM and enzyme induction detailed in Experimental section, II.1; II.2 and II.3, respectively. Small-scale purifications were carried out using nickel affinity chromatography by loading cell-free extract onto a Ni-NTA column (QIAGEN) according to the supplied protocol. The washing buffer contained 50 mM potassium phosphate buffer (pH 7.5), 50 mM NaCl, 10 % glycerol and 30 mM imidazole. The elution buffer contained 50 mM potassium phosphate buffer (pH 8.5), 50 mM NaCl, 10 % glycerol and 250 mM imidazole. The eluted fractions were desalted using Amicon® Ultra-4 10K (Merck Millipore®) by a three cycles of desalting buffer loading (50 mM potassium phosphate buffer pH 7.5, 50 mM NaCl, 10% glycerol) and centrifugations. Protein concentration of the purified fractions were measured by the Bradford method with bovine serum albumin standard (Bio-Rad®). The purified fractions were also analyzed by SDS-PAGEs using Invitrogen NuPAGE system. The purified enzymes were stored at -80°C.

II.5 General method for large-scale purification of enzymes

Large-scale purifications were conducted by Peggy Sirvain (LGBM, Genoscope) from a 500 mL culture by nickel affinity chromatography in tandem with gel filtration (Hi Load 16/600 Superdex 200pg) as described elsewhere (Perchat *et al.*, 2018). The storage buffer was 50 mM phosphate pH 7.5, 50 mM NaCl, 10% glycerol and 1 mM DTT. Large-scale purification of *Cfus*AmDH-F140A was conducted using the same protocol except for the mechanical sonication changed into a chemical lysis using Bug Buster® Protein Extraction Reagent. The final ratio of glycerol has also been increased to 30% due to inactivity and/or instability of this mutant when produced from the general protocol used for other nat-AmDHs. Protein concentrations were determined by the Bradford method with bovine serum albumin as the standard. The samples were analyzed by SDS-PAGEs using the Invitrogen NuPAGE system. The purified proteins were stored at −80°C.

II.6 Enzyme production for stability experiments

For the protein stability experiments detailed in Chapter II, III, *Cfus*AmDH, *Msme*AmDH, MATOUAmDH2, IGCAmDH1, IGCAmDH5, *Sgor*AmDH, *Chat*AmDH and *Acol*AmDH were produced using ArcticExpress (DE3) Competent Cells (Agilent Technologies®). The transformation step was done according to the manufacturer's instructions. The transformed cells were grown on Terrific Broth (TB)

medium containing 0.5 M sorbitol, 5 mM betaine, 100 g mL^{-1} carbenicillin, 20 g mL^{-1} Gentamycin at 30°C until reaching an OD_{600} of 0.2-0.4. IPTG was added at 0.2 mM final concentration to start the protein induction and the cells were further grown for 24 h at 18°C. The lysis and purification steps were carried out as detailed in Experimental section, II.4 whether in potassium phosphate buffer (pH 8, 100 mM) or ammonium formate (pH 8, 200 mM) in absence of NaCl and glycerol not to influence protein stability.

II.7 Enzyme production protocols tested for insoluble enzymes

The selected enzymes A0A1Q9SC61, A0A1H8XY75, A0A1I3IZQ4, A0A078BAL0, A0A078AU13, A0A076NBW6 could not be produced in sufficient yield and purity using the general protocol described in II.3 and II.4 because of their high insolubility in these conditions. Therefore, alternatives protocols were tested to increase their solubility.

As described in Chapter IV, I, the usual protocol was first slightly changed with modification of the IPTG concentration from 0.5 mM to 0.1 and 0.2 mM.

II.7.1 Insoluble enzyme production using ArcticExpress (DE3) competent cells

A0A1Q9SC61, A0A1H8XY75, A0A1I3IZQ4, A0A078BAL0, A0A078AU13, A0A076NBW6 were produced using ArcticExpress (DE3) Competent Cells (Agilent Technologies®). The transformation step was done according to the manufacturer's instructions. Following their instructions, the transformed cells were grown on Terrific Broth (TB) medium containing 0.5 M sorbitol, 5 mM betaine or LB medium, at 30°C until reaching an OD_{600} of 0.8-1.1. IPTG was added at 0.5 mM final concentration to start the protein induction and the cells were further grown for 24 h at 12 °C under agitation at 150 rpm. The lysis and purification steps were carried out as detailed in Experimental section, II.4. Some tests were also carried out by removing the final desalting step during the purification.

Another variant protocol, advised by our collaborators from Georgia Tech, was as follows: the transformed cells were grown on TB medium containing 0.5 M sorbitol, 5 mM betaine, 100 g mL^{-1} carbenicillin, 20 g mL^{-1} Gentamycin at 30°C until reaching an OD_{600} of 0.2-0.4. IPTG was added at 0.2 mM final concentration to start the protein induction and the cells were further grown for 24 h at 15°C under agitation at 150 rpm. The lysis and purification steps were carried out as detailed in Experimental section, II.4. Some tests were also carried out by removing the final desalting step during the purification.

II.7.2 Insoluble enzyme production using an autoinduction protocol

This protocol was recommended by our collaborators from Georgia Tech and applied in attempts for A0A1H8XY75 production. The transformed BL21 C+ cells were grown in TB medium containing 0.5 M sorbitol, 5 mM betaine, 100 g mL^{-1} carbenicillin at 30°C and the protein induction was directly started by

addition of lactose 2 g L^{-1} and glucose 0.5 g L^{-1}. The cells were further grown for 19 h at 30°C under agitation at 150 rpm. The lysis and purification steps were carried out as detailed in Experimental section, II.4.

III. Spectrophotometric assays

III.1 General method for enzymatic activities through spectrophotometric assays

Enzymatic activities were determined using spectrophotometry through NAD(P)H-monitoring at 340 nm. All the reactions were conducted from duplicate (when indicated with error bars) at room temperature in spectrophotometric cells (10 mm light path) in a final reaction volume of 100 µL. To a mixture of ammonium formate buffer (1-2 M NH_4HCO_2/NH_4OH, pH 8.5-9), carbonyl-containing compound (10 mM) and 0.2 – 0.4 mM NAD(P)H) were added. The cell was placed in the spectrophotometer for slope recording in absence of enzyme for at least 30s and the reaction was initiated with addition of an appropriate amount of purified enzyme (0.02 - 0.5 mg mL^{-1}). The initial slope determined the specific activity of the enzyme according to Beer–Lambert's law and the molar absorptivity of β-NAD(P)H (ε = 6220 M^{-1} cm^{-1}) after subtraction of the slope obtained under the same conditions without enzyme.

III.2 Stability assays of a panel of native amine dehydrogenases through specific activity measurements

After enzyme incubation under the reported storage conditions, enzymatic activities were determined using spectrophotometry through NAD(P)H-monitoring at 340 nm. A reaction mixture of isobutyraldehyde (**170**) (10 mM) in ammonium formate buffer (2 M, pH 8.5) was preincubated at 40°C for 1 min. After addition of 0.2 mM NADH cofactor (NADPH for MATOUAmDH2), and transfer in a thermostated spectrophotometric cell (10 mm light path) (Vf = 100 µL). The cell was placed in the spectrophotometer thermostated at 40°C for slope recording in absence of enzyme for at least 30s and the reaction was initiated with addition of an appropriate amount of purified enzyme (0.001–0.12 mg mL^{-1}). The initial slope determined the specific activity of the enzyme according to Beer-Lambert's law and the molar absorptivity of β-NAD(P)H (ε =6,220 M^{-1} cm^{-1}) after subtraction of the slope obtained under the same conditions without enzyme. The uncertainties are those generated out of two to three experiments.

III.3 Enzymatic activity of *Cfus*AmDH mutants with enlarged active sites

Enzymatic activities were determined using spectrophotometry through NAD(P)H-monitoring at 340 nm. All the reactions were conducted from duplicate at 30°C in a spectrophotometric cell (6 mm light path) in a final reaction volume of 100 µL. To a mixture of ammonium formate buffer (2 M NH_4HCO_2/NH_4OH, pH 8.5), carbonyl-containing compound (10 mM final concentration from a stock solution at 200 mM in DMSO) was added, the mixture was vortexed and incubated at 30°C for 1 min, then NADH or NADPH (0.2 mM) was added, the mixture was vortexed again and introduced in the pre-heated cell. The cell was placed in the spectrophotometer thermostated at 30°C for slope recording in absence of enzyme for at least 30s and the reaction was initiated with addition of an appropriate amount of purified enzyme (0.03 - 0.25 mg mL^{-1}). The initial slope determined the specific activity of the enzyme according to Beer–Lambert's law and the molar absorptivity of β-NAD(P)H (ε = 6220 M^{-1} cm^{-1}) after subtraction of the slope obtained under the same conditions without enzyme. The uncertainties are those generated out of two to three experiments.

III.4 Kinetic parameters of *Cfus*AmDH-W145A

Kinetic parameters of *Cfus*AmDH-W145A for the carbonyl-containing compound, the cofactor and ammonia (**11**) were determined using spectrophotometry through NAD(P)H-monitoring at 340 nm. All the reactions were performed in ammonium formate buffer pH 8.5 in a final volume of 100 µL at 30°C, in spectrophotometer cell with optical paths of 0.6 cm and the procedure was as described in Experimental section, III.3. Initial rates of the reaction were measured with various concentrations of substrate and saturated concentrations of the other substrates. Data were fitted to the Michaelis and Menten equation or to the equation provided in case of inhibition by excess of substrate, using Sigma Plot software. The uncertainties are those generated out of two to three experiments. Plots are provided in Appendix 40-Appendix 43.

III.5 Thermostability of *Cfus*AmDH and *Cfus*AmDH-W145A

Separate batches of an appropriate amount of enzymes were incubated at 25, 30 and 40°C for up to 600 h and the remaining activity was regularly measured as described in Experimental section, III.3.

III.6 Enzymatic activity in the oxidative deamination direction

To estimate the enzymatic activity of AmDH4 and AmDH4-I80T/P224S/E296G towards 4-AP and that of *Micro*AmDH towards a panel of small amines, spectrophotometry assays were conducted through NAD(P)H-monitoring at 340 nm.

III.6.1 Enzymatic activity of AmDH4 and AmDH4-I80T/P224S/E296G towards 4-aminopentanoate

Enzymatic activities were determined using spectrophotometry through NAD(P)H-monitoring at 340 nm. The reactions were conducted in duplicate at 30°C in spectrophotometric cell (10 mm light path) in a final reaction volume of 100 µL. To a mixture of Tris HCl buffer (50 mM, pH 9), 4-AP (10 mM) and 0.4 mM $NAD(P)^+$ were added. The cell was placed in the spectrophotometer thermostated at 30°C for slope recording in absence of enzyme for at least 30s and the reaction was initiated with addition of an appropriate amount of purified enzyme (0.2 - 0.5 mg mL^{-1}). The initial slope determined the specific activity of the enzyme according to Beer–Lambert's law and the molar absorptivity of β-NAD(P)H (ε = 6220 M^{-1} cm^{-1}) after subtraction of the slope obtained under the same conditions without enzyme. The uncertainties are those generated out of two to three experiments.

III.6.2 Enzymatic activities of *Micro*AmDH towards a panel of key amine substrates

Enzymatic activities were determined using spectrophotometry through NAD(P)H-monitoring at 340 nm. All reactions were conducted in duplicate at 37°C in spectrophotometric cell (10 mm light path) in a final reaction volume of 100 µL. A mixture of Tris HCl buffer (50 mM, pH 9), amine compound (10 mM) was first incubated at 37°C for 1 min and supplemented with 0.4 mM $NADP^+$. The cell was placed in the spectrophotometer thermostated at 37°C for slope recording in absence of enzyme for at least 30s and the reaction was initiated with addition of an appropriate amount of purified enzyme (0.2 - 0.4 mg mL^{-1}). The initial slope measured at 340 nm determined the specific activity of the enzyme according to Beer–Lambert's law and the molar absorptivity of β-NAD(P)H (ε = 6220 M^{-1} cm^{-1}) after subtraction of the slope obtained under the same conditions without enzyme. The uncertainties are those generated out of two to three experiments.

IV. Conversion assays

IV.1 General method for biocatalytic reactions

To a reaction mixture (100 µL in 500 µL Eppendorf tubes) containing 10 mM carbonyl-containing substrate (generally prepared in water but in DMSO in the case of long insoluble substrates), 0.2 – 0.4 mM $NAD(P)^+$, 3 U mL^{-1} GDH-105, 1.2 eq. glucose (**222**) in 1 - 2 M NH_4HCO_2/NH_4OH buffer (pH 8.5) was added 0.5 mU mg^{-1} of purified enzymes. Calibration points were prepared using various concentrations of the targeted amine in mixture containing 2 M NH_4HCO_2/NH_4OH buffer (pH 8.5) and an appropriate volume of desalting buffer (50 mM potassium phosphate buffer pH 7.5, 50 mM NaCl, 10% glycerol). The reaction mixtures and the calibration points were let at 25- 30°C for 24 - 48 h under agitation at 400

rpm. The reaction conditions specific to each assay are described in the legend of the Figures displaying the corresponding results.

IV.2 Conversion monitoring using liquid chromatography

IV.2.1 General method using UHPLC-UV/MS

IV.2.2 Derivatization with benzoyl chloride and UHPLC-UV/MS analytical method

The derivatization with BzCl were carried out according to the following protocol: to 20 µL of reaction mixture were added 10 µL of a 1 M Na_2CO_3/$NaHCO_3$ aqueous solution pH 10, 40 µL of H_2O, and 30 µL of a 50 mM BzCl solution in CH_3CN. The mixture was vortexed for 30 s and then quenched with addition of 20 µL of a 1 M HCl aqueous solution and 30 µL of a 1/1 solution of H_2O/CH_3CN. After filtration (0.22 µm), the mixture was analyzed by UHPLC-UV (eluent CH_3CN/H_2O 0,1% formic acid with a linear gradient 20/80 during 1 min, then 20/80 to 70/30 in 3 min (hold 2 min), then 70/30 to 20/80 in 1 min (hold 1 min) and a re-equilibration time of 2 min; flow 0.5 mL min^{-1}; temperature 25°C; injection volume 3 µL; UV detection at λ = 250 nm); MS detection: polarity +eV, cone 75V, ionization method: ESI. Except when indicated, only the chromatograms from UV detection were considered for the estimations of the conversion.

IV.2.3 *Cfus*AmDH, *Cfus*AmDH-F140A and *Cfus*AmDH-W145A tolerance to DMSO

The conversion rates obtained for *Cfus*AmDH, *Cfus*AmDH-F140A and *Cfus*AmDH-W145A towards heptanal (**217**) in presence of a range of DMSO ratios (Figure 40) were carried out according to the following protocol: final volume 100 µL, 10 mM substrate (prepared as solutions of 50-1000 mM in DMSO for a final v/v ratio of 1-20%), 2 M NH_4HCO_2 buffer, pH 8.5, 0.2 mM $NAD(P)^+$, 1.2 eq. glucose, 3 U mL^{-1} GDH-105, 0.5 mg mL^{-1} purified enzyme, 25°C, 24 h. The temperature was set at 25°C after a preliminary study that displayed lower conversions at 30°C (Appendix 21). The monitoring of heptan-1-amine (**213**) was performed by UHPLC-UV after derivatization with BzCl. The uncertainties are those generated out of two to three experiments.

IV.2.4 Monitoring of 4-phenylbutan-2-one conversion and estimation of *ee*s after derivatization with FDAA

The conversion of 4-phenylbutan-2-one (**36**) and the hits obtained with prochiral ketones from the screening in GC-FID (Chapter II, IV.2 and Chapter II, I.4.3) were analyzed for enantiomeric excess of the amine formed by derivatization with 1-fluoro-2-4-dinitrophenyl-5-*L*-alanine amide (FDAA) in a 96-well plate. To 20 µL of the reaction mixture were added 8 µL $NaHCO_3$ 1M (pH 8) and 20 µL FDAA 15 mM (dissolved in Acetone/ethanol 1:1) and the mix was let at 55°C for 2 h. The reaction was then quenched

by addition of 4 µL HCl 2 M. After addition of 100 µL CH_3OH/H_2O 1:1, the samples were filtrated (0.22 µm) and analyzed by UHPLC-UV (eluent CH_3OH/H_2O 0,1% formic acid with a linear gradient 40/60 during 1 min, then 40/60 to 85/15 in 2 min (hold 3 min), then 85/15 to 40/60 in 1 min and a re-equilibration time of 2 min; flow 0.3 mL min^{-1}; temperature 25°C; injection volume 3 µL; UV detection at λ = 340 nm).

IV.2.5 Derivatization with FDAA for monitoring of the conversion of ketones harboring a carbonyl function on the third carbon atom of the chain

The conversion rates of hexan-3-one (**200**), heptan-3-one (**234**) and octan-3-amine (**235**) using *Msme*AmDH, *Porti*AmDH, *Micro*AmDH, MATOUAmDH2 and their corresponding P5A and P8A mutants (Chapter III, I.4.5) were analyzed for enantiomeric excess of the amine formed by derivatization with FDAA in a 500 µL Eppendorf tubes. To 20 µL of the reaction mixture were added 8 µL $NaHCO_3$ 1M (pH 8) and 20 µL FDAA 15 mM (dissolved in acetone/ethanol 1:1) and the mix was let at 55°C for 2 h. The reaction was then quenched by addition of 4 µL HCl 2 M. After addition of 30 µL CH_3OH/H_2O 1:1, the samples were filtrated (0.22 µm) and analyzed by UHPLC-UV (eluent CH_3OH/H_2O 0,1% formic acid with a linear gradient 40/60 during 1 min, then 40/60 to 85/15 in 2 min (hold 3 min), then 85/15 to 40/60 in 1 min and a re-equilibration time of 2 min; flow 0.3 mL min^{-1}; temperature 25°C; injection volume 3 µL; UV detection at λ = 340 nm).

IV.3 Conversion monitoring using gas chromatography

IV.3.1 Conversion monitoring using GC-FID

*Cfus*AmDH, *Msme*AmDH, *Micro*AmDH, *Apau*AmDH, MATOUAmDH2, *Chat*AmDH, IGCAmDH1, *Rgna*AmDH, *Porti*AmDH and their corresponding P5A and P8A mutants (Chapter II, IV.2 and Chapter III, I.4.3) were screened for conversion of a range of carbonyl substrates as follows: to a reaction mixture (100 µL in 500 µL Eppendorf tubes) containing 10 mM carbonyl-containing substrate (prepared as solutions in DMSO at 200mM for a 5% final ratio of DMSO), 0.2 mM NAD^+, 0.2mM $NADP^+$, 3 U. mL^{-1} GDH-105, 1.2 eq. glucose (**222**) in 2 M NH_4HCO_2/NH_4OH buffer (pH 8.5) was added 0.5 mU mg^{-1} of purified enzymes. For each substrate, a blank mixture was prepared in the same manner but lacking the enzyme replaced by an appropriate volume of desalting buffer. The calibration points were prepared using various concentrations of the corresponding amine product (0, 3, 6 and 10 mM) in mixture containing 2 M NH_4HCO_2/NH_4OH buffer (pH 8.5) and an appropriate volume of desalting buffer. The reaction mixtures and the calibration points were let at 25°C for 24 h under agitation at 400 rpm. The mixtures were extracted in EtOAc as follows: to a mix of 20µL NaOH 10 M and 80 µL of the sample was added 100 µL EtOAc, after 5 s vortexing, a volume of 70 µL of the organic layer was recovered (steps repeated twice). The combined organic layers were then analyzed by GC-FID. Samples were injected on

a TG-35MS AMINE (Thermo Fischer Scientific®) column (30 m; 0,25 mm; 1 µm) using the following parameters: injector temperature 220°C, split flow 20, split ratio 10, flame 250°C, H_2 flow 30 mL min^{-1}.

The gradients used for each substrate are given in Table 20.

Table 20. Gradients used for analyses in GC-FID.

57	85°C (hold 0.5 min); 9.73°C min^{-1} up to 100°C; 1°C min^{-1} up to 105°C; 20°C min^{-1} up to 200°C (hold 5 min)
199	85°C (hold 0.5 min); 10°C min^{-1} up to 99°C (hold 2min); 10°C min^{-1} up to 110°C; 20°C min^{-1} up to 200°C (hold 5 min)
31	95°C (hold 0.5 min); 20°C min^{-1} up to 200°C (hold 12min)
39	95°C (hold 0.5 min); 20°C min^{-1} up to 200°C (hold 5min)
200	82°C (hold 0.5 min); 2°C min^{-1} up to 87°C; 30°C min^{-1} up to 200°C (hold 5 min)
207	70°C (hold 0.5 min); 9.73°C min^{-1} up to 80°C (hold 5 min); 10°C min^{-1} up to 130°C; 30°C min^{-1} up to 200°C (hold 5 min)
217	85°C (hold 0.5 min); 10°C min^{-1} up to 110°C; 20°C min^{-1} up to 200°C (hold 5 min)
218	80°C (hold 0.5 min); 10°C min^{-1} up to 140°C; 20°C min^{-1} up to 200°C (hold 8.5 min)
219	85°C (hold 0.5 min); 20°C min^{-1} up to 200°C (hold 10 min)
220	85°C (hold 0.5 min); 20°C min^{-1} up to 200°C (hold 10 min); 20°C min^{-1} up to 220°C (hold 5 min)
43	80°C (hold 0.5 min); 2°C min^{-1} up to 90°C; 30°C min^{-1} up to 300°C (hold 5 min)
128	80°C (hold 0.5 min); 9.73°C min^{-1} up to 90°C; 10°C min^{-1} up to 130°C; 30°C min^{-1} up to 200°C (hold 5 min)
67	80°C (hold 0.5 min); 9.94°C min^{-1} up to 115°C (hold 2 min); 20°C min^{-1} up to 200°C (hold 5 min)
135	80°C (hold 0.5 min); 10°C min^{-1} up to 200°C (hold 5 min)

IV.3.2 Monitoring of reductive sulfidation using GC-HS-MS

IV.3.2.1.1 Reductive sulfidation reactions

*Cfus*AmDH, *Msme*AmDH, *Micro*AmDH and MATOUAmDH2 were tested for reductive sulfidation of cyclopentancarbaldehyde (**199**), cyclohexanone (**45**), hexanal (**207**) and hexan-3-one (**200**), respectively. To a reaction mixture (100 µL in 500 µL Eppendorf tubes) containing 10 mM carbonyl-containing substrate, 0, 250 or 500 mM Na_2S, 10 mM NAD(P)H in 100 mM K_2HPO_4/KH_2PO_4 buffer (pH 9) was added 0.5 mU mg^{-1} of purified enzymes. Na_2S was prepared as a 4 M solution and the pH was

adjusted to 8.8-9 with addition of HCl fuming 37%. The change in Na_2S molarity in the solution, generally up to 3-3.5 M was considered for the addition in the reactive mixture. For each substrate-enzyme couple and each Na_2S concentration, a blank mixture was prepared in the same manner but lacking the enzyme. No calibration curve was prepared and the analyses were done by comparison of the response area between the reaction mixture and the corresponding blank reaction. The reaction mixtures and the calibration points were let at 30°C for 24 h under agitation at 400 rpm and then analyzed using GC-HS-MS.

IV.3.2.1.2 Analytical method for reductive sulfidation monitoring

The vials were incubated for 5 min at 50°C and sampled with a syringe at 50°C. After the incubation time, 1 mL of the headspace gas was injected each time with a filling speed of 10 mL min^{-1}, an injection speed of 10 mL min^{-1}, and a penetration speed of 10 mm s^{-1}. The splitless mode was applied. The temperatures of the injection and transfer lines were set at 150°C and 280°C. The carrier gas was helium at a constant flow rate 0.5 mL min^{-1}. The GC program started at 30°C (hold time 6 min), continued with 15°C min^{-1} up to 130°C (hold time 0.5 min) and followed by 7°C min^{-1} up to 250°C (hold time 10 min). For mass spectrometry (MS) analyses, the following standard working conditions were applied: electronic impact ionization; positive mode detection; ion source at 220°C; detector voltage, 70 eV; full scan mode, m/z 50−300 (scan time 0.20 s).

IV.4 Semi-preparative scale biocatalytic synthesis

IV.4.1 *Cfus*AmDH-W145A tolerance to higher substrate loading

The tolerance of *Cfus*AmDH-W145A for higher concentrations of heptanal (**217**) and octan-2-one (**67**) was analyzed at 10, 20, 50, 100, 150 and 200 mM at 0.5 and 1 mg mL^{-1} enzyme loading. To a reaction mixture (100 µL in 500 µL Eppendorf tubes) containing 10 - 200 mM **217** or **67** (prepared as different stock solutions in DMSO for a final ratio of DMSO 5%), 0.2 mM NAD^+, 0.2 mM $NADP^+$, 3 U mL^{-1} GDH-105, 1.2 eq. glucose (**222**), in 2 M NH_4HCO_2/NH_4OH buffer (pH 8.5) was added 0.5 - 1.0 mg mL^{-1} of purified enzymes. The calibration points were prepared using various concentrations of the corresponding amines (0, 5, 10 and 20 mM) in mixture containing 2 M NH_4HCO_2/NH_4OH buffer (pH 8.5) and an appropriate volume of desalting buffer (the volume of enzyme solution used in the reaction mixture). The reaction mixtures starting from 50, 100, 150 and 200 mM of substrate were diluted with 2 M NH_4HCO_2/NH_4OH buffer (pH 8.5) up to eq. 20 mM in case of full conversion. The amine formation was monitored after 24 and 48 h of agitation at 400 rpm, 25°C, by UHPLC-UV after derivatization with BzCl as detailed in Experimental section, IV.2.2.

IV.4.2 Semi-preparative scale reactions and work up using *Cfus*AmDH-W145A

In a 50 mL-Greiner tube equipped with a screw cap was poured **67** or **217** (2.5 mL of a 1 M stock solution in DMSO, final ratio of DMSO 5%, 2.5 mmol), distilled water (28.6 mL), ammonium formate pH 8.5 (2 mL of a 10 M stock solution), glucose (**222**) (3 mL of a 1 M stock solution, 3 mmol), $NADP^+$ (1 mL of a 10 mM, 10 µmol), NAD^+ (1 mL of a 10 mM, 10 µmol), GDH-105 (1.5 mL of a 100 U mL^{-1} stock solution, 150 U) and purified *Cfus*AmDH-W145A (2.4 mL of a 10.5 mg mL^{-1} stock solution, 25.2 mg). The reactions were shaken at 25°C, 350 rpm. After 24 h, the conversions were monitored using UHPLC-UV after BzCl derivatization of 20 µL sample as detailed in Experimental section, IV.2.2 (Appendix 44 and Appendix 45). The reactions were acidified with HCl fuming >37% solution to pH <2 and the products were washed twice with MTBE to remove the remaining **67** and **217** (2 x 15 mL). The aqueous layers were then basified to pH 12 with 10 M KOH solution and the products were extracted with MTBE (5 x 20 mL). The combined layers were then washed with 10 mL 5% weight/volume (w/v) LiCl solution to remove the remaining DMSO. This washing with aqueous phase also allowed the removal of the remaining ammonia (**11**). The organic layers were dried ($MgSO_4$) and filtrated before the addition of a solution of 2 M HCl in diethylether (1.8 eq. of the estimated **231** and **213** produced, deduced from the analytical conversion). Solvent were removed by evaporation under reduced pressure to afford **231** as monohydrochloric salt (116 mg, 28% isolated yield, white solid) but only a brownish wet solid of **213**. *Ee* of 98.4% was obtained by UHPLC-UV after FDAA derivatization of the obtained amine and commercial **231** at 0.025 – 0.05 mM (Appendix 44). NMR analyses were in accordance with the commercial standard. ^{1}H NMR (600MHz) δ 3.22 (m, 1H), 1.48 (m, 2H), 1.27-1.14 (m, 8H; d, J = 6.24 Hz, 3H), 0.74 (t, J = 6.90 Hz, 3H). ^{13}C NMR (150 MHz, D_2O) δ 47.9, 34.0, 30.8, 28.0, 24.5, 21.9, 17.7, 13.4 (Appendix 46 and Appendix 47).

V. *In vivo* directed evolutions

The detailed protocol for each molecular biology experiments is not given for the sake of clarity and also because they were not all performed by myself. Only the general method is given. For Chapter III, II.2, describing the selection screen based on NH_3 release, general methods for gene subcloning, gene cloning and liquid cells growth assays are given. For Chapter III, II.3, describing the selection screen based on NADPH release, the molecular biology steps were all carried out by Valérie Delmas and are therefore not detailed here. Only the general method for bioscreens assays given in Appendix 64 and Appendix 65, also performed by her, are detailed for the sake of clarity.

V.1 General method for gene subcloning

The DNA construction containing the gene of interest was first digested using PacI-NotI restriction enzymes, the digested fragment was purified on QIAGEN DNA purification kit, titrated using QuBit protocol and device and ligated to a (PacI-NotI)-digested vector (1/5 ratio of vector/fragment ; 5-50 ng vector). After dialysis purification, the fragment was transformed into DH10B commercial competent cells (Thermo Scientific™) or in-house batches through electroporation. The transformed cells were spread onto LB agar plates containing the appropriate antibiotic . pSP100 and pVDM18 vectors contain a carbenicillin and chloramphenicol resistance cassette, respectively (Appendix 58). The medium was supplemented with 80 µg mL^{-1} 5-bromo-4-chloro-3-indolyl-beta-*D*-galactopyranoside (X-gal) and 1 mM IPTG for a blue/white colony screening. Colony PCR using Dream Taq kit was done on at least eight positive clones to verify the length of the inserted fragment. The positive clones were then cultivated in antibiotic-containing LB medium before DNA extraction using Miniprep kit and sequencing. The batches with correct sequence were pooled before transformation of the appropriate *E. coli* K12 strain.

The gene coding for AmDH5 was subcloned from pSP100 to pVDM18. oraS-oraE operons coding for OA were subcloned from pVDM18 to pSP100.

V.2 General method for gene cloning

Only the first steps differ from the subcloning protocol. A PCR is performed to amplify the gene of interest using forward and reverse primers containing PacI and NotI restrictions sites. oraS-oraE operons coding for OA was retrieved from *Clostridium sticklandii* genomic DNA and the gene coding for *Tther*AmDH from a construction made in pET22b(+). A *DpnI* digestion step was added when the gene was amplified from a pET22b(+) plasmidic construction. After digestion of the amplified fragment using *PacI* and *NotI* restriction enzyme, the protocol proceeds to the purification, titration and ligation steps as detailed in Experimental section, VI.1.

V.3 General method for growth assays in liquid medium

A general method for growth assays in liquid medium leading to the result tables indicating "++", "+" and "-" is as follows: 1-2 mL precultures were grown overnight in the indicated medium devoid of amine substrate. Precultures were diluted in growth medium to reach an OD_{600} of 0.01. The cultures were then incubated at the indicated temperature and 150-180 rpm for 24 - 72 h. OD_{600} was measured at regular intervals.

V.4 General method for bioscreen assays

For growth curve experiments, shown in Appendix 64 and Appendix 65, a Bioscreen C (Thermo Scientific Labsystems) was used consisting of a thermostatic incubator and a culture growth-monitoring device (OD reader). Precultures in MA medium supplemented with 10 mM glycerol (**247**), 10 mM acetate (**248**) and 5 mM α-ketoglutarate (**249**), ± 5-10 mM gluconate (**244**), ± 10-30 mM cyclohexanone (**45**) were cultivating at 30°C for 48 h; 200 µL aliquots of the cell suspensions in growth medium (final OD_{600} = 0.01) were distributed into honeycomb 100-wells plates. The plates were incubated at 30°C under continuous agitation. Bacterial growth was followed by recording OD_{600} every 15 min for 72 h. Each experiment was performed in technical triplicate. The average curves are shown on Appendix 64 and Appendix 65.

VI. Structural characterizations

VI.1 Crystallization and structural resolution of *Cfus*AmDH-W145A

Pure *Cfus*AmDH-W145A was concentrated to 10 mg mL^{-1} and complexed with either 10 mM NAD^+ or $NADP^+$. Protein was subjected to crystallization trials using commercially available screens in 96-well plate format in which 300 nL drops, comprising 150 nL protein solution and 150 nL of precipitant solution were employed. The best NAD^+ crystals were obtained using 0.1 M Tris-HCl buffer 8.5 with 0.2 M $MgCl_2.6H_2O$, 8% (w/v) PEG 20,000 and 8% (v/v) PEG 500 MME. The best $NADP^+$ crystals were obtained using 0.1 M Bis-Tris buffer pH 5.5 with 25% (w/v) PEG 3,350. These crystals were incubated with a solution of the mother liquor containing 10 mM **211** and incubated for 30 min prior to fishing. Crystals were flash-cooled from the crystallization drops in liquid nitrogen and tested for diffraction using a Rigaku Micromax-007HF fitted with Osmic multilayer optics and a MARRESEARCH MAR345 imaging plate detector. Those crystals displaying diffraction of > 3 Å resolution were retained for data collection at the synchrotron. Datasets for crystals of *Cfus*AmDH-W145A in complex with NAD^+ or $NADP^+$ and **211** were collected on beamline I03 at the Diamond Light Source Synchrotron in Oxford, U.K. Data, which were collected to 1.64 and 1.50 Å respectively, were processed and integrated using XDS (Kabsch, 2010) and scaled using SCALA (Evans, 2011) as part of the Xia2 processing system (Winter, 2010). Data collection statistics are given in Appendix 48. The structures were solved with the program MOLREP (Vagin & Teplyakov, 1997), using a monomer of the WT *Cfus*AmDH (PDB: 6IAU, Mayol *et al.*, 2019) as a model. Each solution contained two molecules in the asymmetric unit. The structures were built and refined using iterative cycles within the programs COOT (Emsley & Cowtan, 2004) and REFMAC (Murshudov *et al.*, 1997), respectively. Following building of the protein and water molecules in the $NADP^+$ structure, clear density was observed adjacent to the nicotinamide ring of $NADP^+$. This was successfully modelled

and refined as the added ligand **211**. The NAD^+ and $NADP^+$ complexes were refined to R_{cryst}/R_{free} values of 17.8/21.4 and 16.4/18.8% respectively. The structures were validated upon deposition within the Protein DataBank (PDB). Refinement statistics are presented in Appendix 48. Coordinates for the *Cfus*AmDH-W145A structures in complex with NAD^+ or $NADP^+$ and **211** have been deposited in the PDB with the accession codes 7QZN and 7QZL, respectively.

VI.2 Bioinformatic analyses

VI.2.1 Protein modeling

The crystallographic structures used were provided by the team of Prof. Grogan from protocols details elsewhere (Mayol *et al.*, 2019) or above for *Cfus*AmDH-W145A. The homology models were either generated by homology as part of the ASMC pipeline using MODELLER or using SwissModel webservice with automatic parameters and selection of the template displaying the highest ranking score. The mutants were modelled from the solved RX structure or models using the Wizard mutagenesis tool on PyMol and rotamer selection. When indicated, some models were generated using ColabFold webservice using automatic parameters or took from Uniprot providing AlphaFold2 predicted structures.

VI.2.2 Molecular dockings

VI.2.2.1 General method for molecular dockings

The ligand PDB files were generated using CORINA Molecular Online Tool. With AutoDockTool (Morris *et al.*, 2009), the docking simulations were performed on rigid structures, with no flexibility given to any catalytic pocket residue. The number of Genetic Algorithm (GA) runs were fixed at 10 or 50 using the Lamarckian GA (4.2). The 10 or 50 ligand conformations obtained were then analyzed on PyMol Schrödinger LLC.

VI.2.2.2 Molecular docking of short chiral amines

In order to explain the ratio of *R*- and *S*- short amines shown in Chapter II, IV.1.2. (*2R/S*)-butan-2-amine (**182** and **193**), (*2R/S*)-1-hydroxypropan-2-amine (**197** and **196**) and (*2R/S*)-1-hydroxybutan-2-amine (**198** and **119**) were docked into *Cfus*AmDH and *Msme*AmDH crystallographic structures (PDB: 6IAU and 6IAQ, respectively), and *Micro*AmDH and MATOUAmDH2 homology models obtained using *Msme*AmDH and *Cfus*AmDH RX structures as templates, respectively. The box was centered in the active site with the following size: 46, 46, 46 (AutoDock parameters). Among the 50 conformations generated, only the poses with the C-NH_2 bond oriented away from the cofactor were listed. The analyses were

based on the ratio of "correct" poses of whether the (*R*)- or (*S*)-amine and on their respective binding energies calculated by AutoDock and given in kJ mol^{-1}.

VI.2.2.3 Molecular docking of pentan-1-amine to decan-1-amine into *Cfus*AmDH, *Cfus*AmDH-F140A and *Cfus*AmDH-W145A

To estimate the potential of mutations into Ala at P5 and P8 positions in *Cfus*AmDH, the dockings of pentan-1-amine (**211**), hexan-1-amine (**212**), heptan-1-amine (**213**), octan-1-amine (**214**), nonan-1-amine (**215**) and decan-1-amine (**216**) were performed into *Cfus*AmDH crystallographic structure (PDB: 6IAQ) and the corresponding structure with *in silico* F140A and W145A mutations. The *in silico* mutations were done using Wizard Mutagenesis PyMol tool giving only one possible conformation in the case of a mutation into Ala. The box was centered in the native or enlarged active sites with the following size: 62, 62, 62 (AutoDock parameters). Among the 10 generated poses for each docking, the analyses were based on the number of poses inside the active site, with the amine moiety in the area of the cofactor C4N atom and P3, and their corresponding binding energies calculated by AutoDock and given in kJ mol^{-1}.

VI.2.2.4 Molecular docking of ethylamine and methylamine into *Cfus*AmDH and *Cfus*AmDH-L177A

To estimate the potential of the mutation of L177 into Ala in *Cfus*AmDH to accommodate larger amine substrates, ethylamine (**255**) and methylamine (**38**) were docked into the *Cfus*AmDH model previously docked with cyclohexanone (**45**) and ammonia (**11**) by Karine Bastard, and the corresponding L177A variant. **11** was removed from this model and the *in silico* mutagenesis of L177 into Ala was performed using the Wizard Mutagenesis Tool of PyMol. The box was centered on the upper part of the active site, in the remaining space between the docked **45** and the active site edges, with the following size: 34, 34, 34 (AutoDock parameters). Among the 10 generated poses, the analyses were based on the orientation of the amine moiety, the conservation of their position and the binding energies calculated by AutoDock and given in kJ mol^{-1}.

VI.2.2.5 Molecular docking of (*2R/S*)-pentan-2-amine into MGYP000357504158, GUT_GENOME186969_00552 and MGYP000037226974

To estimate the influence of the presence of another negatively charged amino acid in the pocket to move the fixation site of **11** and alter the typical (*S*)-enantioselectivity of the nat-AmDHs, both enantiomers of pentan-2-amine [(*S*)-**173** and (*R*)-**258**] were docked into MGYP000357504158, GUT_GENOME186969_00552 and MGYP000037226974. Homology models performed using MODELLER during the ASMC pipeline were used for this docking. In GUT_GENOME186969_00552, M136 (P5) was mutated into Met using Wizard Mutagenesis tool on PyMol to artificially move its side chain aside from

the center of the active site. The selection of the final conformation still considered the steric clashes with neighbor residues, the statistical conformation adopted by Met residues and the conformation found in other nat-AmDHs bearing a Met residue at P5 (IGCAmDH1/2/3/4/6, AcolAmDH, etc., Appendix 5). The box was centered into the active site with the following size: 54, 48, 64 (AutoDock parameters). Among the 10 generated poses, the analyses were based on the number of (*R*)- or (*S*)-amines accommodated in the pocket, the amine moiety orientation in the active site and their respective binding energies calculated by AutoDock and given in kJ mol^{-1}.

VI.2.3 Molecular dynamics

VI.2.3.1.1 Molecular dynamic simulations

MD simulations were executed using the GROMACS software (version 2021. 3) (Abraham *et al.*, 2015; Berendsen *et al.*, 1995; Hess *et al.*, 2008; Lindahl *et al.*, 2001; Pronk *et al.*, 2013; Van Der Spoel *et al.*, 2005; Bekker *et al.*, 1993), with CHARMM36 force field, following the GROMACS tutorial on protein/ligand complex (Lemkul, 2018). In order to generate MD simulations starting from fully equilibrated structures, we first considered every system composed of Enzyme-Cofactor-Ligand, where Enzyme is either *Cfus*AmDH (WT) or mutant *Cfus*AmDH-W145A (M), Cofactor is either NAD^+ (NAD) or $NADP^+$ (NADP) and eventually, Ligand is either pentan-1-amine (**211**, PEN) or heptan-1-amine (**213**, HEP). Each system was placed in a dodecahedron box, filled with TIP3P waters to model water solvation, then counter-ions were added to reach system neutralization. The borders of the simulation box were cut off at 10 Å from the surface of the protein. A first step of energy minimization was carried out through steepest descent algorithm, to relax the eventual steric constraints. Then each system followed a two-step protocol equilibration, with position restraints applied to the protein, co-factor and substrate heavy atoms. The first phase ran a 100 ps NVT (constant volume) simulation followed by a second 100 ps NPT (constant pressure) simulation to maintain pressure isotropically at 1.0 bar. All production steps were carried out under NPT conditions with 2 fs step size. The LINCS algorithm was used to constrain the lengths of hydrogen containing bonds and the waters were restrained using the SETTLE algorithm. Van der Waals forces were treated using a 12 Å cut-off. Long-range electrostatic forces were treated using the Particle Mesh Ewald method. A number of four independent MD simulations (with random initial velocities at NVT equilibration) were executed per system. Each production contains a trajectory of 100 ns, with a snapshot saved every 100 ps, thus resulting in 1000 frames (counting the starting structure) per simulation. These frames were submitted to analysis and the systems were compared using the average of all the four simulations performed. Visualization and figures of protein structures were generated using PyMol Schrödinger LLC.

VI.2.3.1.2 Molecular dynamic outputs

The RMSF analysis was done using the rmsf module on protein backbone in each simulation. The distances were calculated using the distance module between G14 (CA) and G174 (CA), H37 (CE1) and NA(D)P$^+$ (C5A), D36 (CG) and NA(D)P$^+$ (C2B) and N38 (CG) and NADP$^+$ (P2B) and between Q141 (OE1) and NAD(P)$^+$ (N7N). The interaction energies were calculated after a rerun of a 10 ns simulation and by extraction of the energy using the energy module for Protein-Cofactor, Protein –Ligand and Protein-Adenosine interactions. The Adenosine moiety was defined with the following atoms: C4B; C3B; C2B; O4B; C1B; O2B; N9A; N9A; N7A; C4A; C5A; N3A; C2A; N1A; C6A; N6A; O3B; ± P2B; ± O1X; ± O2X; ± O3X. The last four atoms were only present in NADP.

BIBLIOGRAPHY

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX, 1-2*, 19-25. doi:10.1016/j.softx.2015.06.001

Abrahamson, M. J., Vazquez-Figueroa, E., Woodall, N. B., Moore, J. C., & Bommarius, A. S. (2012). Development of an amine dehydrogenase for synthesis of chiral amines. *Angew Chem Int Ed Engl, 51*(16), 3969-3972. doi:10.1002/anie.201107813

Abrahamson, M. J., Wong, J. W., & Bommarius, A. S. (2013). The Evolution of an Amine Dehydrogenase Biocatalyst for the Asymmetric Production of Chiral Amines. *Advanced Synthesis & Catalysis, 355*(9), 1780-1786. doi:10.1002/adsc.201201030

Acevedo-Rocha, C. G., Hoebenreich, S., & Reetz, M. T. (2014). Iterative saturation mutagenesis: a powerful approach to engineer proteins by systematically simulating Darwinian evolution. *Methods Mol Biol, 1179*, 103-128. doi:10.1007/978-1-4939-1053-3_7

Afanasyev, O. I., Kuchuk, E., Usanov, D. L., & Chusov, D. (2019). Reductive Amination in the Synthesis of Pharmaceuticals. *Chemical Reviews, 119*(23), 11857-11911. doi:10.1021/acs.chemrev.9b00383

Afanasyev, O. I., Kuchuk, E., Usanov, D. L., & Chusov, D. (2019). Reductive Amination in the Synthesis of Pharmaceuticals. *Chem Rev*. doi:10.1021/acs.chemrev.9b00383

Aleku, G. A., France, S. P., Man, H., Mangas-Sanchez, J., Montgomery, S. L., Sharma, M., Leipold, F., Hussain, S., Grogan, G., Turner, N. J. (2017). A reductive aminase from *Aspergillus oryzae*. *Nat Chem, 9*, 961-969. doi:10.1038/nchem.2782

Aleku G., Mangas-Sanchez, J., Citoler, J., France, S. P., Montgomery, S. L., Heath, R. S., Thompson, M. P., Turner, N. J. (2018). Kinetice resolution and deracemization of racemic amines using a reductive aminase. ChemCatChem, *10*(3), 515-519. doi:10.1002/cctc.201701484

Aleku, G. A., Titchiner, G. R., Roberts, G. W., Derrington, S. R., Marshall, J. R., Hollfelder, F., Turner, N. J., Leys, D. (2022). Enzymatic *N*-Allylation of Primary and Secondary Amines Using Renewable Cinnamic Acids Enabled by Bacterial Reductive Aminases. *ACS Sustainable Chemistry & Engineering*, *10*(20), 6794-6806. doi:10.1021/acssuschemeng.2c01180

Alexeeva, M., Enright, A., Dawson, M. J., Mahmoudian, M., & Turner, N. J. (2002). Deracemization of α-Methylbenzylamine Using an Enzyme Obtained by *In Vitro* Evolution. *Angewandte Chemie International Edition, 41*(17), 3177-3180. doi:10.1002/1521-3773(20020902)41:17<3177::AID-ANIE3177>3.0.CO;2-P

AlQuraishi, M. (2019). End-to-End Differentiable Learning of Protein Structure. *Cell Syst, 8*(4), 292-301.e293. doi:10.1016/j.cels.2019.03.006

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology, 215*(3), 403-410. doi:10.1016/S0022-2836(05)80360-2

Anastas, P., & Eghbali, N. (2010). Green Chemistry: Principles and Practice. *Chemical Society Reviews, 39*(1), 301-312. doi:10.1039/B918763B

Anthony, S. M., Tona, V., Zou, Y., Morrill, L. A., Billingsley, J. M., Lim, M., Tang, Y., Huk, K. N., Garg, N. K. (2021). Total Synthesis of (−)-Strictosidine and Interception of Aryne Natural Product Derivatives "Strictosidyne" and "Strictosamidyne". *Journal of the American Chemical Society, 143*(19), 7471-7479. doi:10.1021/jacs.1c02004

Asano, Y., Yamaguchi, K., & Kondo, K. (1989). A new NAD^+-dependent opine dehydrogenase from Arthrobacter sp. strain 1C. *J Bacteriol, 171*(8), 4466-4471. doi:10.1128/jb.171.8.4466-4471.1989

Baker, P. J., Turnbull, A. P., Sedelnikova, S. E., Stillman, T. J., & Rice, D. W. (1995). A role for quaternary structure in the substrate specificity of leucine dehydrogenase. *Structure, 3*(7), 693-705. doi:10.1016/s0969-2126(01)00204-0

Barber, C. B., Dobkin, D. P., & Huhdanpaa, H. (1996). The quickhull algorithm for convex hulls. *ACM Trans. Math. Softw., 22*(4), 469–483. doi:10.1145/235815.235821

Barker, J. A., & Thornton, J. M. (2003). An algorithm for constraint-based structural template matching: application to 3D templates with statistical analysis. *Bioinformatics, 19*(13), 1644-1649. doi:10.1093/bioinformatics/btg226

Barrick, J. E., & Lenski, R. E. (2013). Genome dynamics during experimental evolution. *Nature Reviews Genetics, 14*(12), 827-839. doi:10.1038/nrg3564

Bastard, K., Smith, A. A. T., Vergne-Vaxelaire, C., Perret, A., Zaparucha, A., De Melo-Minardi, R., Mariage, A., Boutard, M., Debard, A., Lechaplais, C., Pelle, C., Pellouin, V., Perchat, N., Petit, J. –L., Kreimeyer, A., Medigue, C., Weissenbach, J., Artiguenave, F., De Berardinis, V., Vallenet, D., Salanoubat, M. (2014). Revealing the hidden functional diversity of an enzyme family. *Nat Chem Biol, 10*(1), 42-49. doi:10.1038/nchembio.1387

Behrens, G. A., Hummel, A., Padhi, S. K., Schätzle, S., & Bornscheuer, U. T. (2011). Discovery and Protein Engineering of Biocatalysts for Organic Synthesis. *Advanced Synthesis & Catalysis, 353*(13), 2191-2215. doi:10.1002/adsc.201100446

Bekker, H.; Berendsen, H. J. C.; Dijkstra, E. J.; Archterop, S.; Van Drunen, R.; Van Der Spoel, D.; Sijbers, A.; Keegstra, H., GROMACS: A parallel computer for molecular dynamics simulations. Phys. Comput. 1993, *92*, 252-256.

Belkhelfa, S., Roche, D., Dubois, I., Berger, A., Delmas, V. A., Cattolico, L., Perret, A., Labadie, K., Perdereau, A. C., Darii, E., Pateau, E., De Berardinis, V., Salanoubat, M., Bouzon, M., Döring, V. (2019). Continuous Culture Adaptation of Methylobacterium extorquens AM1 and TK 0001 to Very High Methanol Concentrations. *Frontiers in Microbiology, 10*. doi:10.3389/fmicb.2019.01313

Bell, E. L., Finnigan, W., France, S. P., Green, A. P., Hayes, M. A., Hepworth, L. J., .Lovelock, S. L., Niikura, H., Osuna, S., Romero, E., Ryan, K. S., Turner, N. J., Flitsch, S. L. (2021). Biocatalysis. *Nature Reviews 1*(46). doi:10.1038/s43586-021-00044-z

Bennett, M., Ducrot, L., Vergne-Vaxelaire, C., & Grogan, G. (2022). Structure and Mutation of the Native Amine Dehydrogenase MATOUAmDH2. *Chembiochem*, *23*(10), e202200136. doi:10.1002/cbic.202200136

Berendsen, H. J. C., van der Spoel, D., & van Drunen, R. (1995). GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun., 91*(1), 43-56. doi:10.1016/0010-4655(95)00042-E

Berini, F., Casciello, C., Marcone, G. L., & Marinelli, F. (2017). Metagenomics: novel enzymes from non-culturable microbes. *FEMS Microbiology Letters, 364*(21). doi:10.1093/femsle/fnx211

Bertoni, M., Kiefer, F., Biasini, M., Bordoli, L., & Schwede, T. (2017). Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Scientific Reports, 7*(1), 10480. doi:10.1038/s41598-017-09654-8

Bienert, S., Waterhouse, A., de Beer, Tjaart A. P., Tauriello, G., Studer, G., Bordoli, L., & Schwede, T. (2016). The SWISS-MODEL Repository—new features and functionality. *Nucleic Acids Research, 45*(D1), D313-D319. doi:10.1093/nar/gkw1132

Bittker, J. A., Le, B. V., Liu, J. M., & Liu, D. R. (2004). Directed evolution of protein enzymes using nonhomologous random recombination. *Proceedings of the National Academy of Sciences, 101*(18), 7011-7016. doi:doi:10.1073/pnas.0402202101

Bommarius, B. R., Schurmann, M., & Bommarius, A. S. (2014). A novel chimeric amine dehydrogenase shows altered substrate specificity compared to its parent enzymes. *Chemical Communications, 50*(95), 14953-14955. doi:10.1039/c4cc06527a

Bonde, M. T., Klausen, M. S., Anderson, M. V., Wallin, A. I. N., Wang, H. H., & Sommer, M. O. A. (2014). MODEST: a web-based design tool for oligonucleotide-mediated genome engineering and recombineering. *Nucleic Acids Research, 42*(W1), W408-W415. doi:10.1093/nar/gku428

Bonetta, R., & Valentino, G. (2020). Machine learning techniques for protein function prediction. *Proteins: Structure, Function, and Bioinformatics, 88*(3), 397-413. doi:10.1002/prot.25832

Boratyn, G. M., Camacho, C., Cooper, P. S., Coulouris, G., Fong, A., Ma, N., Madden, T. L., Matten, W. T., McGinnis, S. D., Merezhuk, Y., Raytselis, Y., Sayers, E. W., Tao, T., Ye, J., Zaretskaya, I. (2013). BLAST:

a more efficient report with usability improvements. *Nucleic Acids Research, 41*(W1), W29-W33. doi:10.1093/nar/gkt282

Bornadel, A., Bisagni, S., Pushpanath, A., Montgomery, S. L., Turner, N. J., & Dominguez, B. (2019). Technical Considerations for Scale-Up of Imine Reductase Catalyzed Reductive Amination: A Case Study. *Organic Process Research & Development*. doi:10.1021/acs.oprd.9b00123

Bouzon, M., Döring, V., Dubois, I., Berger, A., Stoffel, G. M. M., Ramirez, L. C., Meyer, S., Fouré, M., Roche, D., Perret, A., Erb, T. J., Bar-Even, A., Lindner, S. N. (2021). Emergence of $NADP^+$-reducing enzymes in *Escherichia coli* central metabolism via adaptive evolution. *mBio*, *12*(4), e00329-21. doi:10.1128/mBio.00329-21

Breuer, M., Ditrich, K., Habicher, T., Hauer, B., Keßeler, M., Stürmer, R., & Zelinski, T. (2004). Industrial Methods for the Production of Optically Active Intermediates. *Angewandte Chemie International Edition, 43*(7), 788-824. doi:https://doi.org/10.1002/anie.200300599

Britton, K. L., Asano, Y., & Rice, D. W. (1998). Crystal structure and active site location of *N*-(1-*D*-carboxylethyl)-*L*-norvaline dehydrogenase. *Nat Struct Biol, 5*(7), 593-601. doi:10.1038/854

Brunhuber, N. M. W., Thoden, J. B., Blanchard, J. S., & Vanhooke, J. L. (2000). Rhodococcus l-Phenylalanine Dehydrogenase: Kinetics, Mechanism, and Structural Basis for Catalytic Specifity. *Biochemistry, 39*(31), 9174-9187. doi:10.1021/bi000494c

Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods, 12*(1), 59-60. doi:10.1038/nmeth.3176

Cai, R.-F., Liu, L., Chen, F.-F., Li, A., Xu, J.-H., & Zheng, G.-W. (2020). Reductive Amination of Biobased Levulinic Acid to Unnatural Chiral γ-Amino Acid Using an Engineered Amine Dehydrogenase. *ACS Sustainable Chemistry & Engineering, 8*(46), 17054-17061. doi:10.1021/acssuschemeng.0c04647

Caparco, A. A., Pelletier, E., Petit, J. L., Jouenne, A., Bommarius, B. R., de Berardinis, V., Zaparucha, A., Champion, J. A., Bommarius, A. S., Vergne-Vaxelaire, C. (2020). Metagenomic Mining for Amine Dehydrogenase Discovery. *Advanced Synthesis & Catalysis, 362*(12), 2427-2436. doi:10.1002/adsc.202000094

Carr, R., Alexeeva, M., Dawson, M. J., Gotor-Fernández, V., Humphrey, C. E., & Turner, N. J. (2005). Directed evolution of an amine oxidase for the preparative deracemisation of cyclic secondary amines. *Chembiochem, 6*(4), 637-639. doi:10.1002/cbic.200400329

Carter, P. (1986). Site-directed mutagenesis. *Biochem J, 237*(1), 1-7. doi:10.1042/bj2370001

Chang, F., Wang, C., Chen, Q., Zhang, Y., & Liu, G. (2022). A Chemoenzymatic Cascade Combining a Hydration Catalyst with an Amine Dehydrogenase: Synthesis of Chiral Amines. *Angewandte Chemie International Edition, 61*(10), e202114809. doi:10.1002/anie.202114809

Chen, F.-F., Liu, Y.-Y., Zheng, G.-W., & Xu, J.-H. (2015). Asymmetric Amination of Secondary Alcohols by using a Redox-Neutral Two-Enzyme Cascade. *ChemCatChem, 7*(23), 3838-3841. doi:10.1002/cctc.201500785

Chen, F.-F., Xu, J.-H., & Zheng, G.-W. (2022). Multifunctional Biocatalysis: An Unusual Imine Reductase. *Engineering Microbiology*, *2*(2), 100023. doi:10.1016/j.engmic.2022.100023

Chen, F., Cosgrove, S. C., Birmingham, W. R., Mangas-Sanchez, J., Citoler, J., Thompson, M., Zheng, G.-W., Xu, J.-H., Turner, N. J. (2019). Enantioselective Synthesis of Chiral Vicinal Amino Alcohols Using Amine Dehydrogenases. *ACS Catalysis*, *9*(12), 11813-11818. doi:10.1021/acscatal.9b03889

Chen, F., Gaucher, E. A., Leal, N. A., Hutter, D., Havemann, S. A., Govindarajan, S., Ortlund, E. A., Benner, S. A. (2010). Reconstructed evolutionary adaptive paths give polymerases accepting reversible terminators for sequencing and SNP detection. *Proceedings of the National Academy of Sciences, 107*(5), 1948-1953. doi:10.1073/pnas.0908463107

Chen, F., Zheng, G.-W., Liu, L., Li, H., Chen, Q., li, F.-L., Li, C.-X., Xu, J.-H. (2018). Reshaping the Active Pocket of Amine Dehydrogenases for Asymmetric Synthesis of Bulky Aliphatic Amines. *ACS Catalysis, 8*(3), 2622-2628. doi:10.1021/acscatal.7b04135

Chen, H., Moore, J., Collier, S. J., Smith, D., Nazor, J., Hughes, G., Janey, J., Huisman, G., Novick, S., Agard, N., Alvizo, O., Cope, G., Yeo, W.-L., Sukurmaran, J., Ng, S. (2013). Engineered imine reductases

and methods for the the reductive amination of ketone and amine compounds. US Patent No.: 20130302859.

Chen, K., & Arnold, F. H. (1993). Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proceedings of the National Academy of Sciences, 90*(12), 5618-5622. doi:10.1073/pnas.90.12.5618

Cheng, F., Li, Q., Li, H., & Xue, Y. (2020). NAD(P)H-dependent oxidoreductases for synthesis of chiral amines by asymmetric reductive amination of ketones. *Sheng Wu Gong Cheng Xue Bao, 36*(9), 1794-1816. doi:10.13345/j.cjb.190582

Chowdury, R., Bouatta, N., Biswas, S., Floristean, C., Kharkare, A., Roye, K., Rochereau, C., Ahdritz, G., Zhang, J., Church, G. M., Sorger, P. K., AlQuraishi, M. (2022). Single-sequence protein structure prediction using a language model and deep learning. *Nature Biotechnology*. doi: 10.1038/s41587-022-01432-w

Citoler, J., Harawa, V., Marshall, J. R., Bevinakatti, H., Finnigan, J. D., Charnock, S. J., & Turner, N. J. (2021). Synthesis of Pharmaceutically Relevant 2-Aminotetralin and 3-Aminochroman Derivatives via Enzymatic Reductive Amination. *Angewandte Chemie International Edition, 60*(46), 24456-24460. doi:10.1002/anie.202110321

Coco, W. M. (2003). RACHITT. In *Directed Evolution Library Creation*, 111-127. Humana Press.

Constable, D. J. C., Dunn, P. J., Hayler, J. D., Humphrey, G. R., Leazer, J. J. L., Linderman, R. J., Lorenz, K., Manley, J., Pearlman, B. A., Well, A., Zaks, A., Zhang, T. Y. (2007). Key green chemistry research areas—a perspective from pharmaceutical manufacturers. *Green Chemistry, 9*(5), 411-420. doi:10.1039/B703488C

Copp, J. N., Akiva, E., Babbitt, P. C., & Tokuriki, N. (2018). Revealing Unexplored Sequence-Function Space Using Sequence Similarity Networks. *Biochemistry, 57*(31), 4651-4662. doi:10.1021/acs.biochem.8b00473

Cosgrove, S. C., Brzezniak, A., France, S. P., Ramsden, J. I., Mangas-Sanchez, J., Montgomery, S. L., Heath, R. S., Turner, N. J. (2018). Imine Reductases, Reductive Aminases, and Amine Oxidases for the Synthesis of Chiral Amines: Discovery, Characterization, and Synthetic Applications. In *Methods Enzymology*, 608, 131-149. Scrutton, Nigel. doi:10.1016/bs.mie.2018.04.022

Daniel González-Martínez, A. C., Mahima Sharma, Marina García-Ramos,, & Iván Lavandera, V. G.-F., and Gideon Grogan. (2020). Asymmetric Synthesis of Primary and Secondary β-Fluoroarylamines using Reductive Aminases from Fungi. *Chem. Cat. Chem Communications, 12*(9), 2421-2425. doi:10.1002/cctc.201901999

David, R. J., Ling, H., Martin, M., Scott, N. (2004). Mutants of enzymes and methods for their use. US Patent No : 20040115691.

de Melo-Minardi, R. C., Bastard, K., & Artiguenave, F. (2010). Identification of subfamily-specific sites based on active sites modeling and clustering. *Bioinformatics, 26*(24), 3075-3082. doi:10.1093/bioinformatics/btq595

Della-Negra, O., Chaussonnerie, S., Fonknechten, N., Barbance, A., Muselet, D., Martin, D. E., Fouteau, S. E., Fischer, C., Saaidi, P.-L., Le Paslier, D. (2020). Transformation of the recalcitrant pesticide chlordecone by *Desulfovibrio* sp.86 with a switch from ring-opening dechlorination to reductive sulfidation activity. *Scientific Reports, 10*(1), 13545. doi:10.1038/s41598-020-70124-9

Della-Negra, O., Le Cacher de Bonneville, B., Chaussonnerie, S., Le Paslier, D., Frison, G., & Saaidi, P.-L. (2021). Microbiological versus Chemical Reductive Sulfidation: An Experimental and Theoretical Study. *ACS OMEGA, 6*(11), 7512-7523. doi:10.1021/acsomega.0c06041

Devine, P. N., Howard, R. M., Kumar, R., Thompson, M. P., Truppo, M. D., Turner, N. J. (2018). Extending the application o biocatalysis to meet the challenges of drug development. *Nature Reviews*, *2*(12), 409-421. doi:10.1038/s41570-018-0055-1

Donnelly, A. E., Murphy, G. S., Digianantonio, K. M., & Hecht, M. H. (2018). A *de novo* enzyme catalyzes a life-sustaining reaction in *Escherichia coli*. *Nature Chemical Biology, 14*(3), 253-255. doi:10.1038/nchembio.2550

Duan, J., Li, B., Qin, Y., Dong, Y., Ren, J., & Li, G. (2019). Recent progress in directed evolution of stereoselective monoamine oxidases. *Bioresources and Bioprocessing, 6*(1), 37. doi:10.1186/s40643-019-0272-6

Duan, S., Widlicka, D. W., Burns, M. P., Kumar, R., Hotham, I., Desrosiers, J.-N., Bowles, P., Jones, K. N., Nicholson, L. D., Buetti-Weekly, M. T., Han, L., Steflik, J., Hansen, E., Hayward, C. M., Strohmeyer, H., Monfette, S., Sutton, S. C., Morris, C. (2022). Application of Biocatalytic Reductive Amination for the Synthesis of a Key Intermediate to a CDK 2/4/6 Inhibitor. *Organic Process Research & Development, 26*(3), 879-890. doi:10.1021/acs.oprd.1c00255

Ducrot, L., Bennett, M., André-Leroux, G., Elisée, E., Marynberg, S., Fossey-Jouenne, A., Zaparucha, A., Grogan, G., Vergne-Vaxelaire, C. (2022). Expanding the substrate scope of native amine dehydrogenases through in silico structural exploration and targeted protein engineering. ChemCatChem, e202200880. doi:10.1002/cctc.202200880

Ducrot, L., Bennett, M., Caparco, A. A., Champion, J. A., Bommarius, A. S., Zaparucha, A., Grogan, G., Vergne-Vaxelaire, C. (2021). Biocatalytic Reductive Amination by Native Amine Dehydrogenases to Access Short Chiral Alkyl Amines and Amino Alcohols. *Frontiers in Catalysis, 1*. doi:10.3389/fctls.2021.781284

Ducrot, L., Bennett, M., Grogan, G., & Vergne-Vaxelaire, C. (2020). NAD(P)H-Dependent Enzymes for Reductive Amination: Active Site Description and Carbonyl-Containing Compound Spectrum. *Advanced Synthesis and Catalalysis, 363*(2), 328-351. doi:10.1002/adsc.202000870

Dunn, W. B., Erban, A., Weber, R. J. M., Creek, D. J., Brown, M., Breitling, R., Hankemeier, T., Goodacre, R., Neumann, Kopka, J., Viant, M. R. (2013). Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics, 9*(1), 44-66. doi:10.1007/s11306-012-0434-4

Dunsmore, C. J., Carr, R., Fleming, T., & Turner, N. J. (2006). A chemo-enzymatic route to enantiomerically pure cyclic tertiary amines. *Journal of the American Chemical Society, 128*(7), 2224-2225. doi:10.1021/ja058536d

Eger, E., Schrittwieser, J. H., Wetzl, D., Iding, H., Kuhn, B., & Kroutil, W. (2020). Asymmetric Biocatalytic Synthesis of 1-Aryltetrahydro-β-carbolines Enabled by "Substrate Walking". *Chemistry – A European Journal, 26*(69), 16281-16285. doi:doi.org/10.1002/chem.202004449

Emsley, P., & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallographica D, 60*(12 Part 1), 2126-2132. doi:10.1107/S0907444904019158

Evans, P. (2011). An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallographica D, 67*(4), 282-292. doi:10.1107/S090744491003982X

Fang, Z., Li, T., Wang, Q., Zhang, X., Peng, H., Fang, W. Hong, Y., Ge, H., Xiao, Y. (2011). A bacterial laccase from marine microbial metagenome exhibiting chloride tolerance and dye decolorization ability. *Applied Microbiology and Biotechnology, 89*(4), 1103-1110. doi:10.1007/s00253-010-2934-3

Fiser, A., Do, R. K., & Sali, A. (2000). Modeling of loops in protein structures. *Protein Science, 9*(9), 1753-1773. doi:10.1110/ps.9.9.1753

Fiser, A., & Šali, A. (2003). Modeller: Generation and Refinement of Homology-Based Protein Structure Models. In *Methods in Enzymology*, *374*, 461-491. Academic Press. doi:10.1016/S0076-6879(03)74020-8

Fisher, B. F., Snodgrass, H. M., Jones, K. A., Andorfer, M. C., & Lewis, J. C. (2019). Site-Selective C–H Halogenation Using Flavin-Dependent Halogenases Identified via Family-Wide Activity Profiling. *ACS Central Science, 5*(11), 1844-1856. doi:10.1021/acscentsci.9b00835

Fonknechten, N., Perret, A., Perchat, N., Tricot, S., Lechaplais, C., Vallenet, D., Vergne, C., Zaparucha, C., Le Paslier, D., Weissenback, J., Salanoubat, M. (2009). A Conserved Gene Cluster Rules Anaerobic Oxidative Degradation of *L*- Ornithine. *Journal of Bacteriology, 191*(9), 3162-3167. doi:10.1128/JB.01777-08

Ford, G. J., Baldwin, C. R., Bradshaw Allen, R. T., Marshall, J. R., Mattey, A. P., Turner, N. J., Clapés, P., Flitsch, S. L. (2022). Three-component stereoselective enzymatic synthesis of amino diols and amino-polyols. *ChemRxiv*. doi:10.26434/chemrxiv-2022-61fhr

Fraaije, M. W., Wu, J., Heuts, D. P., van Hellemond, E. W., Spelberg, J. H., & Janssen, D. B. (2005). Discovery of a thermostable Baeyer-Villiger monooxygenase by genome mining. *Applied Microbiology and Biotechnology, 66*(4), 393-400. doi:10.1007/s00253-004-1749-5

France, S. P., Howard, R. M., Steflik, J., Weise, N. J., Mangas-Sanchez, J., Montgomery, S. L., Crook, R., Kumar, R., Turner, N. J. (2018). Identification of Novel Bacterial Members of the Imine Reductase Enzyme Family that Perform Reductive Amination. *ChemCatChem, 10*(3), 510-514. doi:10.1002/cctc.201701408

Franklin, R. D., Mount, C. J., Bommarius, B. R., & Bommarius, A. S. (2020). Separate sets of mutations enhance activity and substrate scope of amine dehydrogenase. *ChemCatChem*, *12*(9), 2436-2439. doi:10.1002/cctc.201902364

Froidevaux, V., Negrell, C., Caillol, S., Pascault, J.-P., & Boutevin, B. (2016). Biobased Amines: From Synthesis to Polymers; Present and Future. *Chemical Reviews, 116*(22), 14181-14224. doi:10.1021/acs.chemrev.6b00486

Gabler, F., Nam, S.-Z., Till, S., Mirdita, M., Steinegger, M., Söding, J., Lupas, A. N., Alva, V. (2020). Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. *Current Protocols in Bioinformatics, 72*(1), e108. doi:10.1002/cpbi.108

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., Bairoch, A. (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, *31*, 3784-3788. doi:10.1093/nar/gkg563

Gerlt, J. A., Bouvier, J. T., Davidson, D. B., Imker, H. J., Sadkhin, B., Slater, D. R., & Whalen, K. L. (2015). Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochimica et Biophysca Acta, 1854*(8), 1019-1037. doi:10.1016/j.bbapap.2015.04.015

Ghislieri, D., Green, A. P., Pontini, M., Willies, S. C., Rowles, I., Frank, A., Grogan, G., Turner, N. J. (2013). Engineering an Enantioselective Amine Oxidase for the Synthesis of Pharmaceutical Building Blocks and Alkaloid Natural Products. *Journal of the American Chemical Society, 135*(29), 10863-10869. doi:10.1021/ja4051235

Ghislieri, D., & Turner, N. J. (2014). Biocatalytic Approaches to the Synthesis of Enantiomerically Pure Chiral Amines. *Topics in Catalysis, 57*(5), 284-300. doi:10.1007/s11244-013-0184-1

Goldenzweig, A., Goldsmith, M., Hill, S. E., Gertman, O., Laurino, P., Ashani, Y., Dym, O., Unger, T., Albeck, S., Prilusky, J., Lieberman, R. L., Aharoni, A., Silman, I., Sussman, J. L., Tawfik, D. S, Fleishman, S. J. (2016). Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability. *Molecular Cell, 63*(2), 337-346. doi:10.1016/j.molcel.2016.06.012

Gomm, A., & O'Reilly, E. (2018). Transaminases for chiral amine synthesis. *Current Opinion in Chemical Biology, 43*, 106-112. doi:10.1016/j.cbpa.2017.12.007

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., & Lopez, R. (2010). A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Research, 38*(Web Server issue), W695-699. doi:10.1093/nar/gkq313

Grogan, G., & Turner, N. J. (2016). InspIRED by Nature: NADPH-Dependent Imine Reductases (IREDs) as Catalysts for the Preparation of Chiral Amines. *Chemistry–A European Journal, 22*(6), 1900-1907. doi:10.1002/chem.201503954

Groger, H. (2019). Biocatalytic concepts for synthesizing amine bulk chemicals: recent approaches towards linear and cyclic aliphatic primary amines and omega-substituted derivatives thereof. *Applied Microbiology and Biotechnology, 103*(1), 83-95. doi:10.1007/s00253-018-9452-0

Guex, N., Peitsch, M. C., & Schwede, T. (2009). Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis, 30*(S1), S162-S173. doi:10.1002/elps.200900140

Gupta, R. D., & Tawfik, D. S. (2008). Directed enzyme evolution via small and effective neutral drift libraries. *Nature Methods, 5*(11), 939-942. doi:10.1038/nmeth.1262

Hall, M. (2021). Enzymatic strategies for asymmetric synthesis. *RSC Chemical Biology,* (4). doi:10.1039/D1CB00080B

Hall, T.A. (1999). BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, *41*, 95-98.

Handelsman, J. (2004). Metagenomics: Application of genomics to uncultured microorganisms. *Microbiology and Molecular Biology Reviews, 68*(4), 669-685. doi:10.1128/MMBR.68.4.669-685.2004

Hansen, S. F., Bettler, E., Rinnan, Å., Engelsen, S. B., & Breton, C. (2010). Exploring genomes for glycosyltransferases. *Molecular BioSystems, 6*(10), 1773-1781. doi:10.1039/C000238K

Heath, R. S., Pontini, M., Bechi, B., & Turner, N. J. (2014). Development of an R-Selective Amine Oxidase with Broad Substrate Specificity and High Enantioselectivity. *ChemCatChem, 6*(4), 996-1002. doi:10.1002/cctc.201301008

Heckmann, C. M., Dominguez, B., & Paradisi, F. (2021). Enantio-Complementary Continuous-Flow Synthesis of 2-Aminobutane Using Covalently Immobilized Transaminases. *ACS Sustainable Chemistry & Engineering, 9*(11), 4122-4129. doi:10.1021/acssuschemeng.0c09075

Hermes, J. D., Parekh, S. M., Blacklow, S. C., Koster, H., & Knowles, J. R. (1989). A reliable method for random mutagenesis: the generation of mutant libraries using spiked oligodeoxyribonucleotide primers. *Gene, 84*(1), 143-151. doi:10.1016/0378-1119(89)90148-0

Hess, B., Kutzner, C., van der Spoel, D., & Lindahl, E. (2008). GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation, 4*(3), 435-447. doi:10.1021/ct700301q

Hoesl, M. G., Acevedo-Rocha, C. G., Nehring, S., Royter, M., Wolschner, C., Wiltschi, B., Budisa, N., Antranikian, G. (2011). Lipase Congeners Designed by Genetic Code Engineering. *ChemCatChem, 3*(1), 213-221. doi:10.1002/cctc.201000253

Hon, J., Marusiak, M., Martinek, T., Kunka, A., Zendulka, J., Bednar, D., & Damborsky, J. (2021). SoluProt: prediction of soluble protein expression in Escherichia coli. *Bioinformatics, 37*(1), 23-28. doi:10.1093/bioinformatics/btaa1102

Horváth, I. T., & Anastas, P. T. (2007). Innovations and Green Chemistry. *Chemical Reviews, 107*(6), 2169-2173. doi:10.1021/cr078380v

Hrvatin, S., & Piel, J. (2007). Rapid isolation of rare clones from highly complex DNA libraries by PCR analysis of liquid gel pools. *Journal of Microbiological Methods, 68*(2), 434-436. doi:10.1016/j.mimet.2006.09.009

Huber, T., Schneider, L., Präg, A., Gerhardt, S., Einsle, O., & Müller, M. (2014). Direct Reductive Amination of Ketones: Structure and Activity of *S*-Selective Imine Reductases from *Streptomyces*. *ChemCatChem, 6*(8), 2248-2252. doi:10.1002/cctc.201402218

Innophore website can be found under https://innophore.com/

Itoh, N., Yachi, C., & Kudome, T. (2000). Determining a novel NAD^+-dependent amine dehydrogenase with a broad substrate range from *Streptomyces virginiae* IFO 12827: purification and characterization. *Journal of Molecular Catalysis B: Enzymatic, 10*(1), 281-290. doi:10.1016/S1381-1177(00)00111-9

Jia, H.-Y., Yang, Z.-Y., Chen, Q., Zong, M.-H., & Li, N. (2021). Engineering Promiscuous Alcohol Dehydrogenase Activity of a Reductive Aminase AspRedAm for Selective Reduction of Biobased Furans. *Frontiers in Chemistry, 9*. doi:10.3389/fchem.2021.610091

Jongkind, E. P. J., Fossey-Jouenne, A., Mayol, O., Zaparucha, A., Vergne-Vaxelaire, C., & Paul, C. E. (2022). Synthesis of Chiral Amines via a Bi-Enzymatic Cascade Using an Ene-Reductase and Amine Dehydrogenase. *ChemCatChem, 14*(2), e202101576. doi:10.1002/cctc.202101576

Jonic, S., & Vénien-Bryan, C. (2009). Protein structure determination by electron cryo-microscopy. *Current Opinion in Pharmacology, 9*(5), 636-642. doi:10.1016/j.coph.2009.04.006

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., . . . Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature, 596*(7873), 583-589. doi:10.1038/s41586-021-03819-2

Jurcik, A., Bednar, D., Byska, J., Marques, S. M., Furmanova, K., Daniel, L., Kokkonen, P., Brezovsky, J., Strnad, O., Stourac, J., Pavelka, A., Manal, M., Damborsky, J., Kozlikova, B. (2018). CAVER Analyst

2.0: analysis and visualization of channels and tunnels in protein structures and molecular dynamics trajectories. *Bioinformatics, 34*(20), 3586-3588. doi:10.1093/bioinformatics/bty386

Kabsch, W. (2010). XDS. *Acta Crystallogrica D, 66*(Pt 2), 125-132. doi:10.1107/S0907444909047337

Kaplan, O., Vejvoda, V., Charvátová-Pisvejcová, A., & Martínková, L. (2006). Hyperinduction of nitrilases in filamentous fungi. *Journal of Industrial Microbiology and Biotechnology, 33*(11), 891-896. doi:10.1007/s10295-006-0161-9

Kaplan, O., Veselá, A. B., Petříčková, A., Pasquarelli, F., Pičmanová, M., Rinágelová, A., Bhalla, T., C., Pátek, M., Martínková, L. (2013). A Comparative Study of Nitrilases Identified by Genome Mining. *Molecular Biotechnology, 54*(3), 996-1003. doi:10.1007/s12033-013-9656-6

Kataoka, K., & Tanizawa, K. (2003). Alteration of substrate specificity of leucine dehydrogenase by site-directed mutagenesis. *Journal of Molecular Catalysis B: Enzymatic, 23*(2), 299-309. doi:10.1016/S1381-1177(03)00093-6

Khersonsky, O., Lipsh, R., Avizemer, Z., Ashani, Y., Goldsmith, M., Leader, H., Dym, O., Rogotner, S., Trudeau, D. L., Prilusky, J., Amengual-Rigo, P., Guallar, V., Tawfik, D. S., Fleishman, S. J. (2018). Automated Design of Efficient and Functionally Diverse Enzyme Repertoires. *Mol Cell, 72*(1), 178-186.e175. doi:10.1016/j.molcel.2018.08.033

Kibbe, W. A. (2007). OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Research*, *35*, W43-W46. doi:10.1093/nar/gkm234

Kirk, J., Huber, T., & Jackson, C. J. (2018). Chapter 6 Modulating Enzyme Activity via Incorporation of Non-canonical Amino Acids *Modern Biocatalysis: Advances Towards Synthetic Biological Systems*,153-177. The Royal Society of Chemistry. doi:10.1039/9781788010450-00153

Knaus, T., Böhmer, W., & Mutti, F. G. (2017). Amine dehydrogenases: efficient biocatalysts for the reductive amination of carbonyl compounds. *Green Chemistry, 19*(2), 453-463. doi:10.1039/C6GC01987K

Kollipara, M., Matzel, P., Sowa, M., Brott, S., Bornscheuer, U., & Höhne, M. (2022). Characterization of proteins from the 3N5M family reveals an operationally stable amine transaminase. *Applied Microbiology and Biotechnology*. doi:10.1007/s00253-022-12071-1

Kong, W., Liu, Y., Huang, C., Zhou, L., Gao, J., Turner, N. J., & Jiang, Y. (2022). Direct Asymmetric Reductive Amination of Alkyl (Hetero)Aryl Ketones by an Engineered Amine Dehydrogenase. *Angewandte Chemie International Edition, 61*(21), e202202264. doi:10.1002/anie.202202264

Krieger, E., Koraimann, G., & Vriend, G. (2002). Increasing the precision of comparative models with YASARA NOVA--a self-parameterizing force field. *Proteins, 47*(3), 393-402. doi:10.1002/prot.10104

Krissinel, E., & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallographica Section D, 60*(12 Part 1), 2256-2268. doi:10.1107/S0907444904026460

Lemkul, J. A. (2018). From proteins to perturbated hamiltonians: a suite of tutorials for the GROMACS-2018 molecular simulation package, v1.0. Living Journal of Computtational Molecular Science, *1*(1), 5068. http://www.mdtutorials.com/gmx/

Le Goff, A., Artero, V., Jousselme, B., Tran, P. D., Guillet, N., Métayé, R., Fihri, A., Palacin, S., Fontecave, M. (2009). From Hydrogenases to Noble Metal–Free Catalytic Nanomaterials for H_2 Production and Uptake. *Science, 326*(5958), 1384-1387. doi:10.1126/science.1179773

Le Guilloux, V., Schmidtke, P., & Tuffery, P. (2009). Fpocket: An open source platform for ligand pocket detection. *BMC Bioinformatics, 10*(1), 168. doi:10.1186/1471-2105-10-168

Lehmann, M., Pasamontes, L., Lassen, S. F., & Wyss, M. (2000). The consensus concept for thermostability engineering of proteins. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology, 1543*(2), 408-415. doi:10.1016/S0167-4838(00)00238-7

LeischHannes, GrosseStephan, IwakiHiroaki, HasegawaYoshie, & C.K., L. (2012). Cyclohexylamine oxidase as a useful biocatalyst for the kinetic resolution and dereacemization of amines. *Canadian Journal of Chemistry, 90*(1), 39-45. doi:10.1139/v11-086

Li, B.-B., Zhang, J., Chen, F.-F., Chen, Q., Xu, J.-H., & Zheng, G.-W. (2021). Direct reductive amincation of ketones with amines by reductive aminases. *Green Synthesis and Catalysis, 2*(4), 345-349. doi:10.1016/j.gresc.2021.08.005

Li, D., Wu, Q., & Reetz, M. T. (2020). Focused rational iterative site-specific mutagenesis (FRISM). *Methods in Enzymology, 643*, 225-242. doi:10.1016/bs.mie.2020.04.055

Li, G., Ren, J., Yao, P., Duan, Y., Zhang, H., Wu, Q., Feng, J., Lau, P. C., Zhu, D. (2014). Deracemization of 2-Methyl-1,2,3,4-Tetrahydroquinoline Using Mutant Cyclohexylamine Oxidase Obtained by Iterative Saturation Mutagenesis. *ACS Catalysis, 4*(3), 903-908. doi:10.1021/cs401065n

Li, G., Yao, P., Cong, P., Ren, J., Wang, L., Feng, J., Lau, P. C. K., Wu, Q., Zhu, D. (2016). New recombinant cyclohexylamine oxidase variants for deracemization of secondary amines by orthogonally assaying designed mutants with structurally diverse substrates. *Scientific Reports, 6*(1), 24973. doi:10.1038/srep24973

Li, J., Mu, X., Wu, T., & Xu, Y. (2022). High coenzyme affinity chimeric amine dehydrogenase based on domain engineering. *Bioresources and Bioprocessing, 9*(1), 33. doi:10.1186/s40643-022-00528-0

Li, T., Liang, J., Ambrogelly, A., Brennan, T., Gloor, G., Huisman, G., Lalonde, J., Lekhal, A., Mijts, B., Muley, S., Newman, L., Tobin, M., Wong, G., Zaks, A., Zhang, X. (2012). Efficient, Chemoenzymatic Process for Manufacture of the Boceprevir Bicyclic [3.1.0]Proline Intermediate Based on Amine Oxidase-Catalyzed Desymmetrization. *Journal of the American Chemical Society, 134*(14), 6467-6472. doi:10.1021/ja3010495

Liang, J., Edelsbrunner, H., & Woodward, C. (1998). Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Science, 7*(9), 1884-1897. doi:10.1002/pro.5560070905

Lindahl, E., Hess, B., & van der Spoel, D. (2001). GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling, 7*(8), 306-317. doi:10.1007/s008940100045

Lindner, S. N., Ramirez, L. C., Krüsemann, J. L., Yishai, O., Belkhelfa, S., He, H., Bouzon, M., Döring, V., Bar-Even, A. (2018). NADPH-auxotrophic *E. coli*: a sensor strain for testing *in vivo* regeneration of NADPH. *ACS Synthetic Biology, 7*(12), 2742-2749. doi:10.1021/acssynbio.8b00313

Liu, L., Wang, D.-H., Chen, F.-F., Zhang, Z.-J., Chen, Q., Xu, J.-H., Wang, Z.-L., Zheng, G.-W. (2020). Development of an engineered thermostable amine dehydrogenase for the synthesis of structurally diverse chiral amines. *Catalysis Science & Technology, 10*(8), 2353-2358. doi:10.1039/D0CY00071J

Liu, W., Li, Z., Huang, C.-H., Guo, R.-T., Zhao, L., Zhang, D., Chan, X., Wu, Q., Zhu, D. (2014). Structural and Mutational Studies on the Unusual Substrate Specificity of meso-Diaminopimelate Dehydrogenase from Symbiobacterium thermophilum. *ChemBiochem, 15*(2), 217-222. doi:10.1002/cbic.201300691

Löwe, J., Siewert, A., Scholpp, A.-C., Wobbe, L., & Gröger, H. (2018). Providing reducing power by microalgal photosynthesis: a novel perspective towards sustainable biocatalytic production of bulk chemicals exemplified for aliphatic amines. *Scientific Reports, 8*(1), 10436. doi:10.1038/s41598-018-28755-6

Lu, J., Wang, Z., Jiang, Y., Sun, Z., & Luo, W. (2022). Modification of the substrate specificity of leucine dehydrogenase by site-directed mutagenesis based on biocomputing strategies. *Systems Microbiology and Biomanufacturing*. doi:10.1007/s43393-022-00116-5

Lucas, X., Bauzá, A., Frontera, A., & Quiñonero, D. (2016). A thorough anion–π interaction study in biomolecules: on the importance of cooperativity effects. *Chemical Science, 7*(2), 1038-1050. doi:10.1039/C5SC01386K

Ma, E. J., Siirola, E., Moore, C., Kummer, A., Stoeckli, M., Faller, M., . . . Snajdrova, R. (2021). Machine-Directed Evolution of an Imine Reductase for Activity and Stereoselectivity. *ACS Catalysis, 11*(20), 12433-12445. doi:10.1021/acscatal.1c02786

Mangas-Sanchez, J., France, S. P., Montgomery, S. L., Aleku, G. A., Man, H., Sharma, M., Ramsden, J., I., Grogan, G. Turner, N. J. (2017). Imine reductases (IREDs). *Curr Opin Chem Biol, 37*, 19-25. doi:10.1016/j.cbpa.2016.11.022

Mangas-Sanchez, J., Sharma, M., Cosgrove, S. C., Ramsden, J. I., Marshall, J. R., Thorpe, T. W., Palmer, R. B., Grogan, G., Turner, N. J. (2020). Asymmetric synthesis of primary amines catalyzed by thermotolerant fungal reductive aminases. *Chemical Science, 11*(19), 5052-5057. doi:10.1039/D0SC02253E

Marshall, J. R., Yao, P., Montgomery, S. L., Finnigan, J. D., Thorpe, T. W., Palmer, R. B., Mangas-Sanchez, J., Duncan, R. A. M., Heath, R. S., Graham, K. M., Cook, D. J., Charnock, S. J. Turner, N. J. (2021). Screening and characterization of a diverse panel of metagenomic imine reductases for biocatalytic reductive amination. *Nature Chemistry, 13*, 140-148. doi:10.1038/s41557-020-00606-w

Martí-Renom, M. A., Stuart, A. C., Fiser, A., Sánchez, R., Melo, F., & Sali, A. (2000). Comparative protein structure modeling of genes and genomes. *Annual Review of Biophysics and Biomolecular Structure, 29*, 291-325. doi:10.1146/annurev.biophys.29.1.291

Matzel, P., Gand, M., & Höhne, M. (2017). One-step asymmetric synthesis of (*R*)- and (*S*)-rasagiline by reductive amination applying imine reductases. *Green Chemistry, 19*(2), 385-389. doi:10.1039/C6GC03023H

Mayol, O. (2018). Amine déshydrogenases pour la synthèse biocatalysée d'amines chirales : recherche, application et évolution. UMR 8030 Genomics Metabolics, Evry.

Mayol, O., Bastard, K., Beloti, L., Frese, A., Turkenburg, J. P., Petit, J.-L., Mariage, A., Debard, A., Pellouin, V., Perret, A., De Berardinis, V. Zaparucha, A., Grogan, G., Vergne-Vaxelaire, C. (2019). A family of native amine dehydrogenases for the asymmetric reductive amination of ketones. *Nature Catalysis*, *2*, 324-333. doi:10.1038/s41929-019-0249-z

Mayol, O., David, S., Darii, E., Debard, A., Mariage, A., Pellouin, V., Petit, J.-L., Salanoubat, M., De Berardinis, V., Zaparucha, A., Vergne-Vaxelaire, C. (2016). Asymmetric reductive amination by a wild-type amine dehydrogenase from the thermophilic bacteria Petrotoga mobilis. *Catalysis Science & Technology, 6*(20), 7421-7428. doi:10.1039/c6cy01625a

McPherson, A., & Gavira, J. A. (2014). Introduction to protein crystallization. *Acta Crystallographica Section F, 70*(1), 2-20. doi:10.1107/S2053230X13033141

McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y. M., Buso, N., Cowley, A. P., Lopez, R. (2013). Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Research, 41*(W1), W597-W600. doi:10.1093/nar/gkt376

Meneely, K. M., & Lamb, A. L. (2012). Two Structures of a Thiazolinyl Imine Reductase from Yersinia enterocolitica Provide Insight into Catalysis and Binding to the Nonribosomal Peptide Synthetase Module of HMWP1. *Biochemistry, 51*(44), 9002-9013. doi:10.1021/bi3011016

Mihara, H., Muramatsu, H., Kakutani, R., Yasuda, M., Ueda, M., Kurihara, T., & Esaki, N. (2005). *N*-methyl-*L*-amino acid dehydrogenase from *Pseudomonas putida*. A novel member of an unusual NAD(P)-dependent oxidoreductase superfamily. *FEBS Journal, 272*(5), 1117-1123. doi:10.1111/j.1742-4658.2004.04541.x

Ming, H., Yuan, B., Qu, G., & Sun, Z. (2022). Engineering the activity of Amine Dehydrogenase in the Asymmetric Reductive Amination of Hydroxyl Ketones. *Catalysis Science & Technology*. doi:10.1039/D2CY00391K

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., & Steinegger, M. (2022). ColabFold - Making protein folding accessible to all. *Nature Methods*, 19, 679-682. doi: 10.1038/s41592-022-01488-1

Mitsukura, K., Kuramoto, T., Yoshida, T., Kimoto, N., Yamamoto, H., & Nagasawa, T. (2013). A NADPH-dependent (*S*)-imine reductase (SIR) from *Streptomyces* sp. GF3546 for asymmetric synthesis of optically active amines: purification, characterization, gene cloning, and expression. *Applied Microbiology and Biotechnology, 97*(18), 8079-8086. doi:10.1007/s00253-012-4629-4

Mitsukura, K., Suzuki, M., Shinoda, S., Kuramoto, T., Yoshida, T., Nagasawa, T. (2011). Purification and Characterization of a Novel (*R*)-Imine Reductase from *Streptomyces* sp. GF3587. *Bioscience, Biotechnology, and Biochemistry, 75*(9), 1778-1782. doi:10.1271/bbb.110303

Monod, J. (1950). La technique de culture continue: theorie et applications. *Selected Papers in Molecular Biology by Jacques Monod, 79*, 390-410.

Montgomery, S. L., Pushpanath, A., Heath, R. S., Marshall, J. R., Klemstein, U., Galman, J. L., Woodlock, D., Bisagni, S., Taylor, C. J., Mangas-Sanchez, J., Ramsden, J. I., Dominguez, B. Turner, N. J. (2020). Characterization of imine reductases in reductive amination for the exploration of structure-activity relationships. *Science Advances, 6*(21), eaay9320. doi:10.1126/sciadv.aay9320

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry, 30*(16), 2785-2791. doi:10.1002/jcc.21256

Mu, X., Wu, T., Mao, Y., Zhao, Y., Xu, Y., & Nie, Y. (2021). Iterative alanine scanning mutagenesis confers aromatic ketone specificity and activity of *L*-amine dehydrogenases. *ChemCatChem, 13*(24), 5243-5253. doi:10.1002/cctc.202101558

Müller, K. M., Stebel, S. C., Knall, S., Zipf, G., Bernauer, H. S., & Arndt, K. M. (2005). Nucleotide exchange and excision technology (NExT) DNA shuffling: a robust method for DNA fragmentation and directed evolution. *Nucleic Acids Research, 33*(13), e117-e117. doi:10.1093/nar/gni116

Murshudov, G. N., Vagin, A. A., & Dodson, E. J. (1997). Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallogrica D, 53*(3), 240-255. doi:10.1107/S0907444996012255

Musil, M., Khan, R. T., Beier, A., Stourac, J., Konegger, H., Damborsky, J., & Bednar, D. (2020). FireProtASR: A Web Server for Fully Automated Ancestral Sequence Reconstruction. *Briefings in Bioinformatics, 22*(4). doi:10.1093/bib/bbaa337

Musil, M., Stourac, J., Bendl, J., Brezovsky, J., Prokop, Z., Zendulka, J., Martinek, T., Bednar, D., Damborsky, J. (2017). FireProt: web server for automated design of thermostable proteins. *Nucleic Acids Research, 45*(W1), W393-W399. doi:10.1093/nar/gkx285

Mutti, F. G., & Knaus, T. (2021). Enzymes Applied to the Synthesis of Amines *Biocatalysis for Practitioners*, 143-180. De Gonzalo, G., Lavandera, I.

Mutzel, R. and Marliere, P. (2004). Method and device for selecting accelerated proliferation of living cells in suspension. US Patent No.: 6686194B1

Novick, A., & Szilard, L. (1950). Description of the chemostat. *Science, 112*(2920), 715-716. doi:10.1126/science.112.2920.715

Nugent, T. C., & El-Shazly, M. (2010). Chiral amine synthesis–recent developments and trends for enamide reduction, reductive amination, and imine reduction. *Advanced Synthesis & Catalysis, 352*(5), 753-819. doi:10.1002/adsc/200900719

Nyerges, Á., Csörgő, B., Draskovits, G., Kintses, B., Szili, P., Ferenc, G., Révész, T., Ari, E., Nagy, I., Bálint, B., Vásárhelyi, B. M., Bihari, P., Száamel, M., Balogh, D., Papp, H., Kalapis, D., Papp, B., Pál, C. (2018). Directed evolution of multiple genomic loci allows the prediction of antibiotic resistance. *Proceedings of the National Academy of Sciences, 115*(25), E5726-E5735. doi:10.1073/pnas.1801646115

Nyerges, Á., Csörgő, B., Nagy, I., Bálint, B., Bihari, P., Lázár, V., Apjok, G., Umenhoffr, K., Bogos, B., Posfai, G., Pál, C. (2016). A highly precise and portable genome engineering method allows comparison of mutational effects across bacterial species. *Proceedings of the National Academy of Sciences, 113*(9), 2502-2507. doi:10.1073/pnas.1520040113

O'Reilly, C., & Turner, P. D. (2003). The nitrilase family of CN hydrolysing enzymes – a comparative study. *Journal of Applied Microbiology, 95*(6), 1161-1174. doi:10.1046/j.1365-2672.2003.02123.x

Okamoto, Y., Köhler, V., Paul, C. E., Hollmann, F., & Ward, T. R. (2016). Efficient In Situ Regeneration of NADH Mimics by an Artificial Metalloenzyme. *ACS Catalysis, 6*(6), 3553-3557. doi:10.1021/acscatal.6b00258

Oliveira, T., Sharkey, M. A., Engel, P. C., & Khan, A. R. (2016). Crystal structure of a chimaeric bacterial glutamate dehydrogenase. *Acta crystallographica. Section F, Structural biology communications, 72*(Pt 6), 462-466. doi:10.1107/S2053230X16007305

Ostermeier, M., Shim, J. H., & Benkovic, S. J. (1999). A combinatorial approach to hybrid enzymes independent of DNA homology. *Nature Biotechnology, 17*(12), 1205-1209. doi:10.1038/70754

Osuna, S. (2021). The challenge of predicting distal active site mutations in computational enzyme design. *WIREs Computational Molecular Science, 11*(3), e1502. doi:10.1002/wcms.1502

Ovchinnikov, S., Park, H., Kim, D. E., DiMaio, F., & Baker, D. (2018). Protein structure prediction using Rosetta in CASP12. *Proteins: Structure, Function, and Bioinformatics, 86*(S1), 113-121. doi:10.1002/prot.25390

Owen, J. G., Reddy, B. V., Ternei, M. A., Charlop-Powers, Z., Calle, P. Y., Kim, J. H., & Brady, S. F. (2013). Mapping gene clusters within arrayed metagenomic libraries to expand the structural diversity of biomedically relevant natural products. *Proceedings of the National Academy of Sciences USA, 110*(29), 11797-11802. doi:10.1073/pnas.1222159110

Packer, M. S., & Liu, D. R. (2015). Methods for the directed evolution of proteins. *Nature Reviews Genetics, 16*(7), 379-394. doi:10.1038/nrg3927

Patil, M. D., Grogan, G., Bommarius, A. S., & Yun, H. (2018). Oxidoreductase-Catalyzed Synthesis of Chiral Amines. *ACS Catalysis*, *8*(12), 10985-11015. doi:10.1021/acscatal.8b02924

Pavelka, A., Chovancova, E., & Damborsky, J. (2009). HotSpot Wizard: a web server for identification of hot spots in protein engineering. *Nucleic Acids Research, 37*(suppl_2), W376-W383. doi:10.1093/nar/gkp410

Perchat, N., Saaidi, P.-L., Darii, E., Pellé, C., Petit, J.-L., Besnard-Gonnet, M., De Berardinis, V., Dupont, M., Gimbernat, A., Salanoubat, M., Fischer, C., Perret, A. (2018). Elucidation of the trigonelline degradation pathway reveals previously undescribed enzymes and metabolites. *Proceedings of the National Academy of Sciences, 115*(19), E4358-E4367. doi:10.1073/pnas.1722368115

Poewe, W. H., Rascol, O., Quinn, N., Tolosa, E., Oertel, W. H., Martignoni, E.,Rupp, M., Boroojerdi, B. (2007). Efficacy of pramipexole and transdermal rotigotine in advanced Parkinson's disease: a double-blind, double-dummy, randomised controlled trial. *The Lancet Neurology, 6*(6), 513-520. doi:10.1016/S1474-4422(07)70108-4

Posner, B. A., Li, L., Bethell, R., Tsuji, T., & Benkovic, S. J. (1996). Engineering Specificity for Folate into Dihydrofolate Reductase from Escherichia coli. *Biochemistry, 35*(5), 1653-1663. doi:10.1021/bi9518095

Prather, K. L. J. (2020). Accelerating and expanding nature to address its greatest challenges. *Nature Catalysis, 3*(3), 181-183. doi:10.1038/s41929-020-0422-4

Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., van der Spoel, D., Hess, B., Lindahl, E. (2013). GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics, 29*(7), 845-854. doi:10.1093/bioinformatics/btt055

Pushpanath, A., Siirola, E., Bornadel, A., Woodlock, D., & Schell, U. (2017). Understanding and Overcoming the Limitations of Bacillus badius and *Caldalkalibacillus thermarum* Amine Dehydrogenases for Biocatalytic Reductive Amination. *ACS Catalysis*, *7*(5), 3204-3209. doi:10.1021/acscatal.7b00516

Pyser, J. B., Chakrabarty, S., Romero, E. O., & Narayan, A. R. H. (2021). State-of-the-Art Biocatalysis. *ACS Central Science, 7*(7), 1105-1116. doi:10.1021/acscentsci.1c00273

Rajakumara, E., Abhishek, S., Nitin, K., Saniya, D., Bajaj, P., Schwaneberg, U., & Davari, M. D. (2022). Structure and Cooperativity in Substrate–Enzyme Interactions: Perspectives on Enzyme Engineering and Inhibitor Design. *ACS Chemical Biology, 17*(2), 266-280. doi:10.1021/acschembio.1c00500

Ravikumar, A., Arrieta, A., & Liu, C. C. (2014). An orthogonal DNA replication system in yeast. *Nature Chemical Biology, 10*(3), 175-177. doi:10.1038/nchembio.1439

Ravikumar, Y., Nadarajan, S. P., Hyeon Yoo, T., Lee, C.-s., & Yun, H. (2015). Unnatural amino acid mutagenesis-based enzyme engineering. *Trends in Biotechnology, 33*(8), 462-470. doi:10.1016/j.tibtech.2015.05.002

RCSB PDB website, all released structures can be found under https://www.rcsb.org/stats/all-released-structures

Reetz, M. T., Bocola, M., Carballeira, J. D., Zha, D., & Vogel, A. (2005). Expanding the Range of Substrate Acceptance of Enzymes: Combinatorial Active-Site Saturation Test. *Angewandte Chemie International Edition, 44*(27), 4192-4196. doi:10.1002/anie.200500767

Reetz, M. T., Carballeira, J. D., & Vogel, A. (2006). Iterative Saturation Mutagenesis on the Basis of B Factors as a Strategy for Increasing Protein Thermostability. *Angewandte Chemie International Edition, 45*(46), 7745-7751. doi:10.1002/anie.200602795

Riziotis, I. G., Ribeiro, A. J. M., Borkakoti, N., & Thornton, J. M. (2022). Conformational Variation in Enzyme Catalysis: A Structural Study on Catalytic Residues. *Journal of Molecular Biology, 434*(7), 167517. doi:10.1016/j.jmb.2022.167517

Robinson, S. L., Piel, J., & Sunagawa, S. (2021). A roadmap for metagenomic enzyme discovery. *Natural Product Reports, 38*(11), 1994-2023. doi:10.1039/D1NP00006C

Roddan, R., Ward, J. M., Keep, N. H., & Hailes, H. C. (2020). Pictet–Spenglerases in alkaloid biosynthesis: Future applications in biocatalysis. *Current Opinion in Chemical Biology, 55*, 69-76. doi:10.1016/j.cbpa.2019.12.003

Roiban, G.-D., Kern, M., Liu, Z., Hyslop, J., Tey, P. L., Levine, M. S., . . . Brown, M. J. B. (2017). Efficient Biocatalytic Reductive Aminations by Extending the Imine Reductase Toolbox. *ChemCatChem, 9*(24), 4475-4479. doi:10.1002/cctc.201701379

Rokas, A. (2011). Phylogenetic Analysis of Protein Sequence Data Using the Randomized Axelerated Maximum Likelihood (RAXML) Program. *Current Protocols in Molecular Biology, 96*(1), 19.11.11-19.11.14. doi:10.1002/0471142727.mb1911s96

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., Baker, D. (2008). Kemp elimination catalysts by computational enzyme design. *Nature, 453*(7192), 190-195. doi:10.1038/nature06879

Roughley, S. D., & Jordan, A. M. (2011). The Medicinal Chemist's Toolbox: An Analysis of Reactions Used in the Pursuit of Drug Candidates. *Journal of Medicinal Chemistry, 54*(10), 3451-3479. doi:10.1021/jm200187y

Sablin, S. O., Yankovskaya, V., Bernard, S., Cronin, C. N., & Singer, T. P. (1998). Isolation and characterization of an evolutionary precursor of human monoamine oxidases A and B. *Eur J Biochem, 253*(1), 270-279. doi:10.1046/j.1432-1327.1998.2530270.x

Sali, A., & Blundell, T. L. (1993). Comparative protein modeling by satisfaction of spatial restraints. *Journal of Molecular Biology, 234*(3), 779-815. doi:10.1006/jmbi.1993.1626

Salwan, R., Sharma, V., Surajit, S. (2021). Genome Mining, Phylogenetic and Structural Analysis of Bacterial Nitrilases for the Biodegradation of Nitrile Compounds, PREPRINT (Version 1)

Sangster, J. J., Marshall, J. R., Turner, N. J., & Mangas-Sanchez, J. (2022). New trends and future opportunities in the enzymatic formation of C-C, C-N, and C-O bonds. *Chembiochem, 23*(6), e202100464. doi:10.1002/cbic.202100464

Savile, C. K., Janey, J. M., Mundorff, E. C., Moore, J. C., Tam, S., Jarvis, W. R.,, Colbeck, J. C., Krebber, A., Fleitz, F. J., Brands, J., Devine, P. N., Huisman, G. W., Hughes, G. J. (2010). Biocatalytic Asymmetric Synthesis of Chiral Amines from Ketones Applied to Sitagliptin Manufacture. *Science, 329*(5989), 305-309. doi:10.1126/science.1188934

Sawitzke, J. A., Costantino, N., Li, X.-t., Thomason, L. C., Bubunenko, M., Court, C., & Court, D. L. (2011). Probing Cellular Processes with Oligo-Mediated Recombination and Using the Knowledge Gained to Optimize Recombineering. *Journal of Molecular Biology, 407*(1), 45-59. doi:10.1016/j.jmb.2011.01.030

Scheller, P. N., Fademrecht, S., Hofelzer, S., Pleiss, J., Leipold, F., Turner, N. J., Nestl, B. M., Hauer, B. (2014). Enzyme Toolbox: Novel Enantiocomplementary Imine Reductases. *Chembiochem, 15*(15), 2201-2204. doi:10.1002/cbic.201402213

Scheller, P. N., Lenz, M., Hammer, S. C., Hauer, B., & Nestl, B. M. (2015). Imine Reductase-Catalyzed Intermolecular Reductive Amination of Aldehydes and Ketones. *ChemCatChem, 7*(20), 3239-3242. doi:10.1002/cctc.201500764

Schilling, B., & Lerch, K. (1995) (A). Amine oxidases from *Aspergillus niger* : identification of a novel flavin-dependent enzyme. *Biochimica et Biophysica Acta (BBA) - General Subjects, 1243*(3), 529-537. doi:10.1016/0304-4165(94)00183-X

Schilling, B., & Lerch, K. (1995) (B). Cloning, sequencing and heterologous expression of the monoamine oxidase gene from *Aspergillus niger*. *Molecular and General Genetics, 247*(4), 430-438. doi:10.1007/bf00293144

Schirò, A., Carlon, A., Parigi, G., Murshudov, G., Calderone, V., Ravera, E., & Luchinat, C. (2020). On the complementarity of X-ray and NMR data. *Journal of Structural Biology: X, 4*, 100019. doi:10.1016/j.yjsbx.2020.100019

Schmidtke, P., Le Guilloux, V., Maupetit, J., & Tufféry, P. (2010). fpocket: online tools for protein ensemble pocket detection and tracking. *Nucleic Acids Res, 38*, W582-589. doi:10.1093/nar/gkq383

Schober, M., MacDermaid, C., Ollis, A. A., Chang, S., Khan, D., Hosford, J., Latham, J., Ihnken, L. A., Brown, M. J. B., Fuerst, D., Sanganee, M. J., Roiban, G.-D. (2019). Chiral synthesis of LSD1 inhibitor GSK2879552 enabled by directed evolution of an imine reductase. *Nature Catalysis, 2*(10), 909-915. doi:10.1038/s41929-019-0341-4

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research, 13*(11), 2498-2504. doi:10.1101/gr.1239303

Sharma, M., Mangas-Sanchez, J., France, S. P., Aleku, G. A., Montgomery, S. L., Ramsden, J. I., Turner, N. J., Grogan, G. (2018). A Mechanism for Reductive Amination Catalyzed by Fungal Reductive Aminases. *ACS Catalysis, 8*(12), 11534-11541. doi:10.1021/acscatal.8b03491

Sharma, M., Mangas-Sanchez, J., & Grogan, G. (2017). NAD(P)H-Dependent Dehydrogenases for the Asymmetric Reductive Amination of Ketones: Structure, Mechanism, Evolution and Application. *Advanced Synthesis & Catalysis, 359*, 2011-2025. doi:10.1002/adsc.201700356

Siddiq, M. A., Hochberg, G. K. A., & Thornton, J. W. (2017). Evolution of protein specificity: insights from ancestral protein reconstruction. *Current Opinion in Structural Biology, 47*, 113-122. doi:10.1016/j.sbi.2017.07.003

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology, 7*(1), 539. doi:10.1038/msb.2011.75

Smits, S. H., Meyer, T., Mueller, A., van Os, N., Stoldt, M., Willbold, D., Schmitt, L., Grieshaber, M. K. (2010). Insights into the mechanism of ligand binding to octopine dehydrogenase from Pecten maximus by NMR and crystallography. *PLoS One, 5*(8), e12312. doi:10.1371/journal.pone.0012312

Smits, S. H., Mueller, A., Schmitt, L., & Grieshaber, M. K. (2008). A structural basis for substrate selectivity and stereoselectivity in octopine dehydrogenase from Pecten maximus. *Journal of Molecular Biology, 381*(1), 200-211. doi:10.1016/j.jmb.2008.06.003

Söding, J. (2005). Protein homology detection by HMM-HMM comparison. *Bioinformatics, 21*(7), 951-960. doi:10.1093/bioinformatics/bti125

Souterre, T. (2017). Optimization of enzymes via directed evolution *in vivo*. UMR 8030 Genomics Meabolics, Evry.

Springer Handbook of Enzymes, (2005). *d-Octopine dehydrogenase*, in Springer Handbook of Enzymes, 23 , 108–121. Schomburg D., Schomburg I.

St-Jacques, A. D., Gagnon, O., & Chica, R. A. (2018). Chapter 4 Computational Enzyme Design: Successes, Challenges, and Future Directions *Modern Biocatalysis: Advances Towards Synthetic Biological Systems*, 88-116. The Royal Society of Chemistry.

Steinkellner, G., Gruber, C. C., Pavkov-Keller, T., Binter, A., Steiner, K., Winkler, C., Lyskowski, A., Schwamberger, O., Oberer, M., Scwab, H., Faber, K., Macheroux, P., Gruber, K. (2014). Identification of promiscuous ene-reductase activity by mining structural databases using active site constellations. *Nature Communications, 5*. doi:10.1038/ncomms5150

Stemmer, W. P. (1994). DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution. *Proceedinds of the National Academy of Sciences U S A, 91*(22), 10747-10751. doi:10.1073/pnas.91.22.10747

Stemmer, W. P. C. (1994). Rapid evolution of a protein in vitro by DNA shuffling. *Nature, 370*(6488), 389-391. doi:10.1038/370389a0

Stemmer, W. P. C., Crameri, A., Ha, K. D., Brennan, T. M., & Heyneker, H. L. (1995). Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene, 164*(1), 49-53. doi:10.1016/0378-1119(95)00511-4

Stourac, J., Vavra, O., Kokkonen, P., Filipovic, J., Pinto, G., Brezovsky, J., Damborsky, J., Bednar, D. (2019). Caver Web 1.0: Identification of tunnels and channels in proteins and analysis of ligand transport. *Nucleic Acids Research*, *47*, W414-W422. doi:10.1093/nar/gkz378

Studer, G., Rempfer, C., Waterhouse, A. M., Gumienny, R., Haas, J., & Schwede, T. (2019). QMEANDisCo—distance constraints applied on model quality estimation. *Bioinformatics, 36*(6), 1765-1771. doi:10.1093/bioinformatics/btz828

Suhre, K., Sanejouand, Y. K., (2004) (A). ElNémo: a ‚ormal mode web-server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Research*, *32*, W610-W614. doi: 10.1093/nar/gkh368

Suhre, K., Sanejouand, Y. K., (2004) (B). On the potential of normal mode analysis for solving difficult molecular replacement problems. *Acta Crystallographica, D*, *60*, 796-799. doi: 10.1107/S0907444904001982

Takashima, H. (2006). High-Resolution Protein Structure Determination by NMR. In G. A. Webb (Ed.), *Annual Reports on NMR Spectroscopy*, 59, 235-273. Academic Press.

Thorpe, T. W., Marshall, J. R., Harawa, V., Ruscoe, R. E., Cuetos, A., Finnigan, J. D., Angelastro, A., Heath, R. S., Parmeggani, F., Charnock, S. J., Howard, R. M., Kumar, R., Daniels, D. S. B., Grogan, G., Turner, N. J. (2022). Multifunctional biocatalyst for conjugate reduction and reductive amination. *Nature, 604*(7904), 86-91. doi:10.1038/s41586-022-04458-x

Tong, F., Qin, Z., Wang, H., Jiang, Y., Li, J., Ming, H., Qu, G., Xiao, Y., Sun, Z. (2021). Biosynthesis of Chiral Amino Alcohols via an Engineered Amine Dehydrogenase in *E. coli*. *Frontiers in Bioengineering and Biotechnolohy, 9*, 778584. doi:10.3389/fbioe.2021.778584

Tseliou, V., Knaus, T., Masman, M. F., Corrado, M. L., & Mutti, F. G. (2019) (A). Generation of amine dehydrogenases with increased catalytic performance and substrate scope from epsilon-deaminating *L*-Lysine dehydrogenase. *Nature Communications, 10*(1), 3717. doi:10.1038/s41467-019-11509-x

Tseliou, V., Masman, M. F., Böhmer, W., Knaus, T., & Mutti, F. G. (2019) (B). Mechanistic Insight into the Catalytic Promiscuity of Amine Dehydrogenases: Asymmetric Synthesis of Secondary and Primary Amines. *Chembiochem, 20*(6), 800-812. doi:10.1002/cbic.201800626

Udit, A. K., Silberg, J. J., & Sieber, V. (2003). Sequence homology-independent protein recombination (SHIPREC). *Methods in Molecular Biology, 231*, 153-163. doi:10.1385/1-59259-395-x:153

United Nations website, "Take Action for Sustainable Development Goals" can be foudn under https://www.un.org/sustainabledevelopment/sustainable-development-goals/

Vagin, A., & Teplyakov, A. (1997). MOLREP: an Automated Program for Molecular Replacement. *Journal of Applied Crystallography, 30*(6), 1022-1025. doi:10.1107/S0021889897006766

van Beek, H. L., Gonzalo, G. d., & Fraaije, M. W. (2012). Blending Baeyer–Villiger monooxygenases: using a robust BVMO as a scaffold for creating chimeric enzymes with novel catalytic properties. *Chemical Communications, 48*(27), 3288-3290. doi:10.1039/C2CC17656D

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. C. (2005). GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry, 26*(16), 1701-1718. doi:10.1002/jcc.20291

Vanhooke, J. L., Thoden, J. B., Brunhuber, N. M. W., Blanchard, J. S., & Holden, H. M. (1999). Phenylalanine Dehydrogenase from Rhodococcus sp. M4: High-Resolution X-ray Analyses of Inhibitory

Ternary Complexes Reveal Key Features in the Oxidative Deamination Mechanism. *Biochemistry, 38*(8), 2326-2339. doi:10.1021/bi982244q

Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., . . . Velankar, S. (2021). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research, 50*(D1), D439-D444. doi:10.1093/nar/gkab1061

Vasiliki Paraskevopoulo, F. H. F. (2018). Polyionic Tags as Enhancers of Protein Solubility in Recombinant Protein Expressio. *Microorganisms, 6*(2), 47. doi:10.3390/microorganisms6020047

Vavra, O., Filipovic, J., Plhak, J., Bednar, D., Marques, S. M., Brezovsky, J., Stourac, J., Matyska, L., Damborsky, J. (2019). CaverDock: a molecular docking-based tool to analyse ligand transport through protein tunnels and channels. *Bioinformatics, 35*(23), 4986-4993. doi:10.1093/bioinformatics/btz386

Verberkmoes, N. C., Russell, A. L., Shah, M., Godzik, A., Rosenquist, M., Halfvarson, J., Lefsrud, M. G., Apajalahti, J., Tysk, C., Hettich, R. L., Jansson, J. K. (2009). Shotgun metaproteomics of the human distal gut microbiota. *The ISME Journal, 3*(2), 179-189. doi:10.1038/ismej.2008.108

Vergne-Vaxelaire, C., Bordier, F., Fossey, A., Besnard-Gonnet, M., Debard, A., Mariage, A., Pellouin, V., Perret, A., Petit, J.-L., Stam, M., Salanoubat, M., Weissenbach, J., De Berardinis, V., Zaparucha, A. (2013). Nitrilase Activity Screening on Structurally Diverse Substrates: Providing Biocatalytic Tools for Organic Synthesis. *Advanced Synthesis & Catalysis, 355*(9), 1763-1779. doi:10.1002/adsc.201201098

Vidal, L. S., Kelly, C. L., Mordaka, P. M., & Heap, J. T. (2017). Review of NAD(P)H-dependent oxidoreductases: Properties, engineering and application. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 1866*, 327-347. doi:10.1016/j.bbapap.2017.11.005

Voigt, C. A., Martinez, C., Wang, Z. G., Mayo, S. L., & Arnold, F. H. (2002). Protein building blocks preserved by recombination. *Nature Structural Biology, 9*(7), 553-558. doi:10.1038/nsb805

Wallace, A. C., Borkakoti, N., & Thornton, J. M. (1997). TESS: a geometric hashing algorithm for deriving 3D coordinate templates for searching structural databases. Application to enzyme active sites. *Protein Sci, 6*(11), 2308-2323. doi:10.1002/pro.5560061104

Wang, D.-H., Chen, Q., Yin, S.-N., Ding, X.-W., Zheng, Y.-C., Zhang, Z., Zhang, Y.-H., Chen, F._F., Xu, J.-H., Zheng, G.-W. (2021). Asymmetric Reductive Amination of Structurally Diverse Ketones with Ammonia Using a Spectrum-Extended Amine Dehydrogenase. *ACS Catalysis, 11*(22), 14274-14283. doi:10.1021/acscatal.1c04324

Wang, D. D., Ou-Yang, L., Xie, H., Zhu, M., & Yan, H. (2020) (A). Predicting the impacts of mutations on protein-ligand binding affinity based on molecular dynamics simulations and machine learning methods. *Computational and Structural Biotechnology Journal, 18*, 439-454. doi:10.1016/j.csbj.2020.02.007

Wang, H., Qu, G., Li, J.-K., Ma, J.-A., Guo, J., Miao, Y., & Sun, Z. (2020) (B). Data mining of amine dehydrogenases for the synthesis of enantiopure amino alcohols. *Catalysis Science & Technology, 10*(17), 5945-5952. doi:10.1039/D0CY01373K

Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R., & Church, G. M. (2009). Programming cells by multiplex genome engineering and accelerated evolution. *Nature, 460*(7257), 894-898. doi:10.1038/nature08187

Wang, L., & Schultz, P. G. (2001). A general approach for the generation of orthogonal tRNAs. *Chemistry & Biology, 8*(9), 883-890. doi:10.1016/S1074-5521(01)00063-1

Wang, S., Fang, B. (2013). Method for preparing chiral amine through asymmetric reduction under catalysis of marine strain. Chinese Patent No.:103224963

Wang, S. (2016). Marine strain and its preparation method of chiral amines using amine dehydrogenases. Chinese Patent No.:105567756

Wang, Y., Hou, Y., Wang, Y., Zheng, L., Xu, X., Pan, K., Li, R., Wang, Q. (2018). A Novel Cold-Adapted Leucine Dehydrogenase from Antarctic Sea-Ice Bacterium Pseudoalteromonas sp. ANT178. *Marine Drugs, 16*(10), 359. doi:10.3390/md16100359

Warnecke, F., & Hess, M. (2009). A perspective: Metatranscriptomics as a tool for the discovery of novel biocatalysts. *Journal of Biotechnology, 142*(1), 91-95. doi:10.1016/j.jbiotec.2009.03.022

Watanabe, S., Sueda, R., Fukumori, F., Watanabe, Y. (2015). Characterization of Flavin-Containing Opine Dehydrogenase from Bacteria. PLoS One, *10*(9), e0138434. doi:10.1371/journal.pone.0138434

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T. (2018). SWISS-MODEL: homology modeling of protein structures and complexes. *Nucleic Acids Research, 46*(W1), W296-W303. doi:10.1093/nar/gky427

Webb, B., & Sali, A. (2016). Comparative Protein Structure Modeling Using MODELLER. *Current Protocols in Bioinformatics, 54*, 5.6.1-5.6.37. doi:10.1002/cpbi.3

Wei Jiang, Y. W. (2020). Improving Catalytic Efficiency and Changing Substrate Spectrum for Asymmetric Biocatalytic Reductive Amination. *Journal of Micobiology and Biotechnology, 30*(1), 146-154. doi:10.4014/jmb.1907.07015

Wetzl, D., Berrera, M., Sandon, N., Fishlock, D., Ebeling, M., Müller, M.,, Hanlon, S., Wirz, B., Iding, H. (2015). Expanding the Imine Reductase Toolbox by Exploring the Bacterial Protein-Sequence Space. *Chembiochem, 16*(12), 1749-1756. doi:10.1002/cbic.201500218

Wheeler, L. C., Lim, S. A., Marqusee, S., & Harms, M. J. (2016). The thermostability and specificity of ancient proteins. *Current Opinion in Structural Biology, 38*, 37-43. doi:10.1016/j.sbi.2016.05.015

Wijma, H. J., Floor, R. J., Jekel, P. A., Baker, D., Marrink, S. J., & Janssen, D. B. (2014). Computationally designed libraries for rapid enzyme stabilization. *Protein Engineering, Design and Selection, 27*(2), 49-58. doi:10.1093/protein/gzt061

William Finnigan, J. C., Sebastian Cronin Cosgrove, Nicholas J. Turner. (2020). Rapid model-based optimisation of a two-enzyme system for continuous reductive amination in flow. *Organic Process Research & Development*, *24*(10), 1969-1977. doi:10.1021/acs.oprd.0c00075

Wiltschi, B., Cernava, T., Dennig, A., Galindo Casas, M., Geier, M., Gruber, S., . . . Wriessnegger, T. (2020). Enzymes revolutionize the bioproduction of value-added compounds: From enzyme discovery to special applications. *Biotechnology Advances, 40*, 107520. doi:10.1016/j.biotechadv.2020.107520

Winter, G. (2010). xia2: an expert system for macromolecular crystallography data reduction. *Journal of Applied Crystallography, 43*(1), 186-190. doi:10.1107/S0021889809045701

Winzer, T., Kern, M., King, A. J., Larson, T. R., Teodor, R. I., Donninger, S. L., . . . Graham, I. A. (2015). Morphinan biosynthesis in opium poppy requires a P450-oxidoreductase fusion protein. *Science, 349*(6245), 309-312. doi:10.1126/science.aab1852

Wong, T. S., Tee, K. L., Hauer, B., & Schwaneberg, U. (2004). Sequence saturation mutagenesis (SeSaM): a novel method for directed evolution. *Nucleic Acids Research, 32*(3), e26. doi:10.1093/nar/gnh028

Wu, S., Snajdrova, R., Moore, J. C., Baldenius, K., & Bornscheuer, U. T. (2021). Biocatalysis: Enzymatic Synthesis for Industrial Applications. *Angewandte Chemie International Edition, 60*(1), 88-119. doi:10.1002/anie.202006648

Yang, J., & Zhang, Y. (2015). Protein Structure and Function Prediction Using I-TASSER. *Current Protocols in Bioinformatics, 52*, 5.8.1-5.8.15. doi:10.1002/0471250953.bi0508s52

Yang, Z.-Y., Hao, Y.-C., Hu, S.-Q., Zong, M.-H., Chen, Q., & Li, N. (2021). Direct Reductive Amination of Biobased Furans to N-Substituted Furfurylamines by Engineered Reductive Aminase. *Advanced Synthesis & Catalysis, 363*(4), 1033-1037. doi:10.1002/adsc.202001495

Yasukawa, K., Motojima, F., Ono, A., & Asano, Y. (2018). Expansion of the Substrate Specificity of Porcine Kidney D-Amino Acid Oxidase for S-Stereoselective Oxidation of 4-Cl-Benzhydrylamine. *ChemCatChem, 10*(16), 3500-3505. doi10.1002/cctc.201800614

Yasukawa, K., Nakano, S., & Asano, Y. (2014). Tailoring D-Amino Acid Oxidase from the Pig Kidney to R-Stereoselective Amine Oxidase and its Use in the Deracemization of α-Methylbenzylamine. *Angewandte Chemie International Edition, 53*(17), 4428-4431. doi:10.1002/anie.201308812

Ye, L. J., Toh, H. H., Yang, Y., Adams, J. P., Snajdrova, R., & Li, Z. (2015). Engineering of Amine Dehydrogenase for Asymmetric Reductive Amination of Ketone by EvolvingRhodococcusPhenylalanine Dehydrogenase. *ACS Catalysis, 5*(2), 1119-1122. doi:10.1021/cs501906r

Zanghellini, A. (2014). de novo computational enzyme design. *Current Opinion in Biotechnol, 29*, 132-138. doi:10.1016/j.copbio.2014.03.002

Zaparucha, A., de Berardinis, V., & Vaxelaire-Vergne, C. (2018). Chapter 1 Genome Mining for Enzyme Discovery *Modern Biocatalysis: Advances Towards Synthetic Biological Systems*, 1-27. The Royal Society of Chemistry. doi:10.1039/9781788010450-00001

Zawodny, W., & Montgomery, S. L. (2022). Evolving New Chemistry: Biocatalysis for the Synthesis of Amine-Containing Pharmaceuticals. *Catalysts, 12*(6), 595. doi :10.3390/catal12060595

Zhang, J., Liao, D., Chen, R., Zhu, F., Ma, Y., Gao, L., Qu, G., Cui, C., Sun, Z., Lei, X., Gao, s. (2022). Tuning an Imine Reductase for the Asymmetric Synthesis of Azacycloalkylamines by Concise Structure-Guided Engineering. *Angewandte Chemie International Edition, 61*(24), e202201908. doi:10.1002/anie.202201908

Zhang, Q., Lu, X.-F., Zhang, Y., Tang, X.-L., Zheng, R.-C., & Zheng, Y.-G. (2020). Development of a robust nitrilase by fragment swapping and semi-rational design for efficient biosynthesis of pregabalin precursor. *Biotechnol Bioeng, 117*(2), 318-329. doi:10.1002/bit.27203

Zhao, H., Giver, L., Shao, Z., Affholter, J. A., & Arnold, F. H. (1998). Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nature Biotechnology, 16*(3), 258-261. doi:10.1038/nbt0398-258

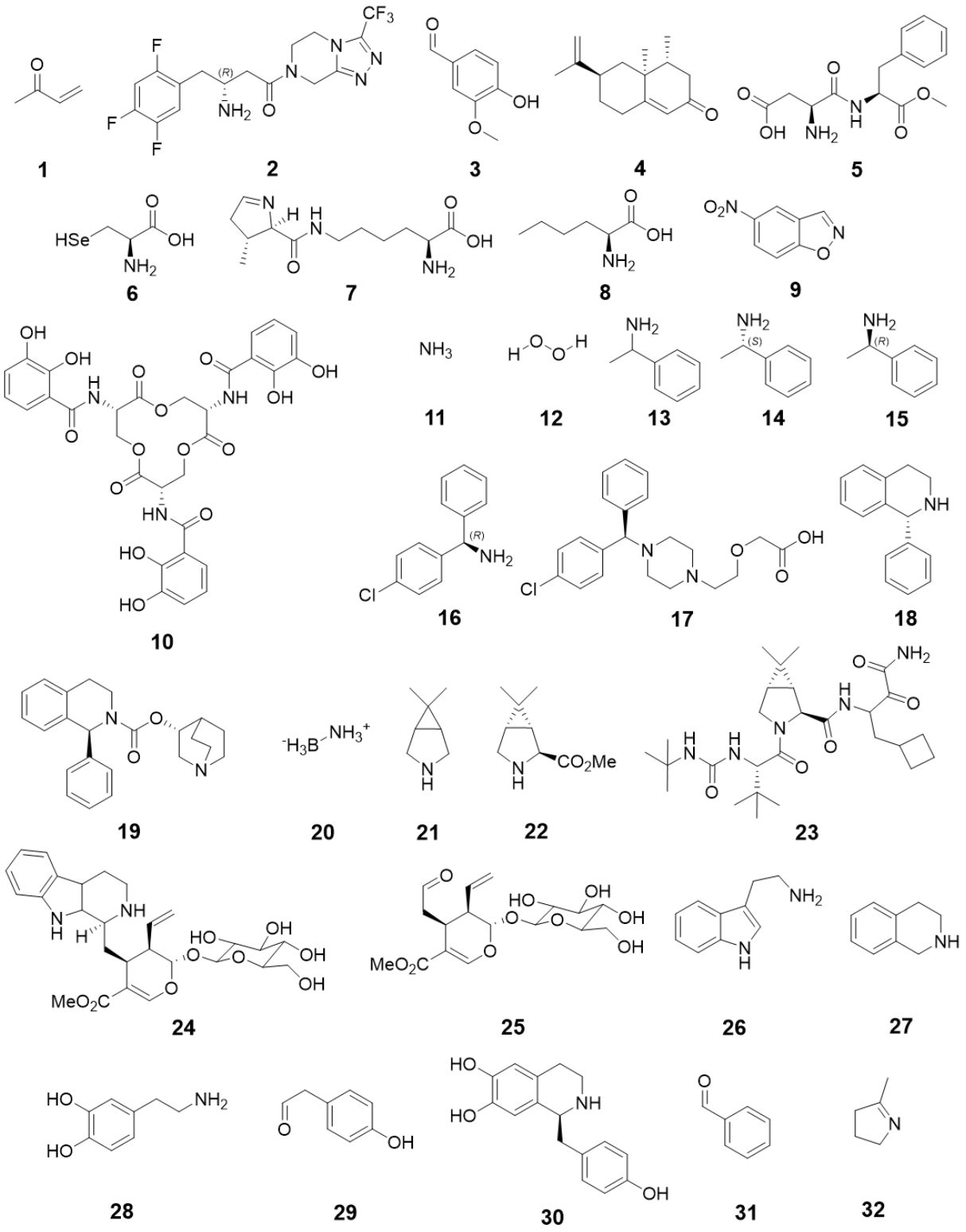
Zhao, S., Kumar, R., Sakai, A., Vetting, M. W., Wood, B. M., Brown, S., . . . Jacobson, M. P. (2013). Discovery of new enzymes and metabolic pathways by using structure and genome context. *Nature, 502*(7473), 698-702. doi:10.1038/nature12576

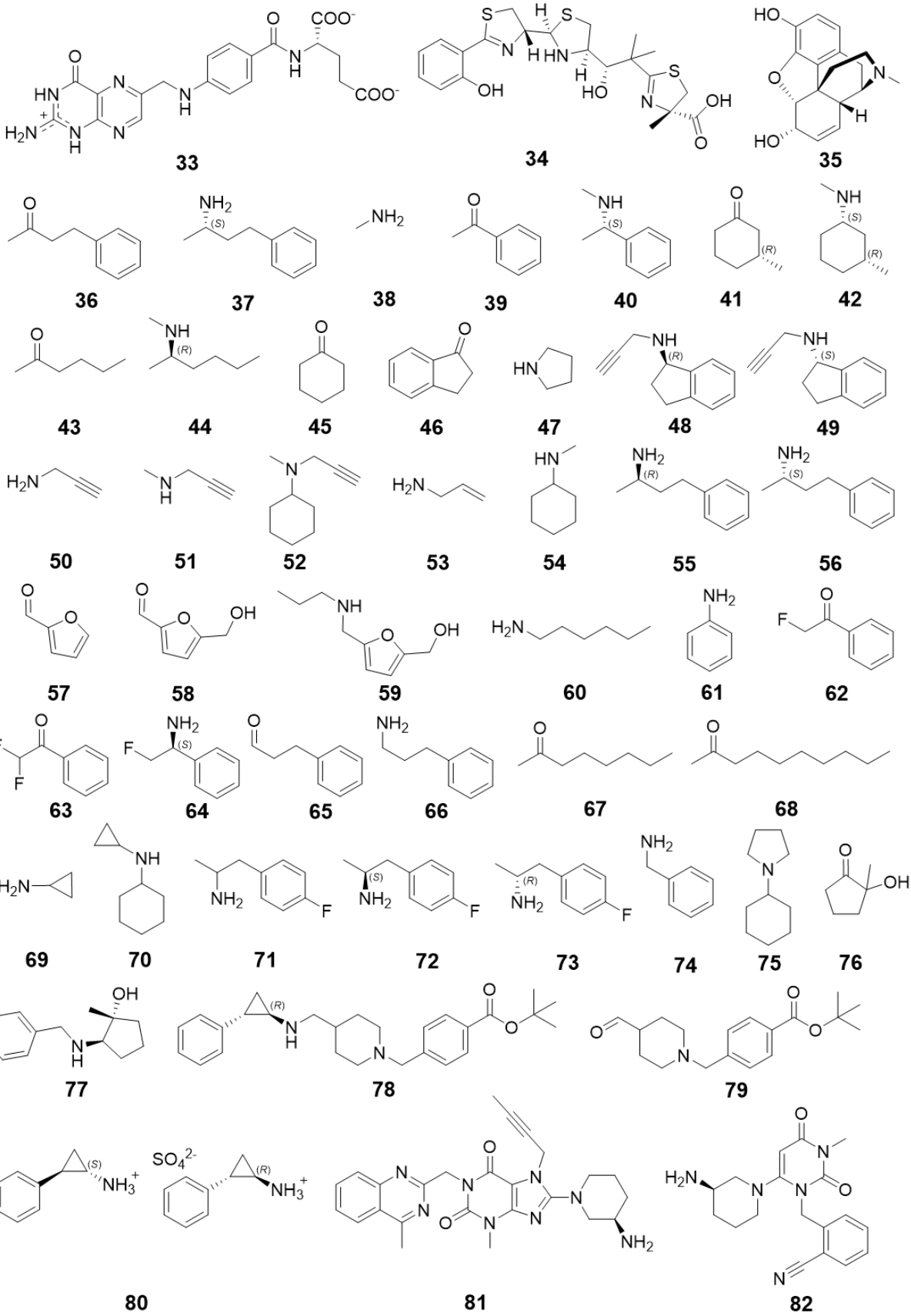
Zhou, F., Xu, Y., Mu, X., & Nie, Y. (2022). A Sustainable Approach for Synthesizing (R)-4-Aminopentanoic Acid From Levulinic Acid Catalyzed by Structure-Guided Tailored Glutamate Dehydrogenase. *Frontiers in Bioengineering and Biotechnology, 9*. doi:10.3389/fbioe.2021.770302

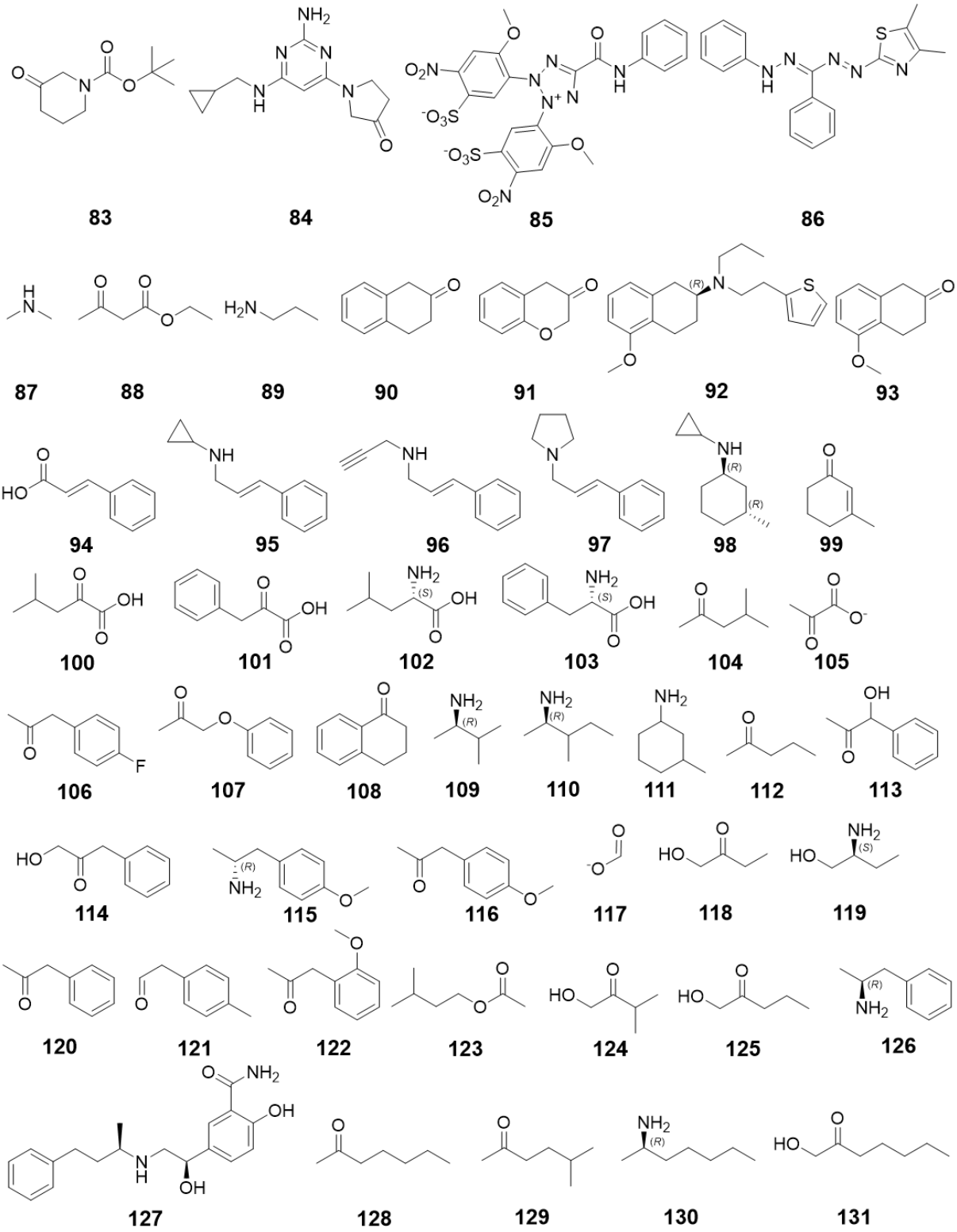
Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F., Söding, J., Lupas, A. N., Alva, V. (2018). A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *Journal of Molecular Biology, 430*(15), 2237-2243. doi:10.1016/j.jmb.2017.12.007

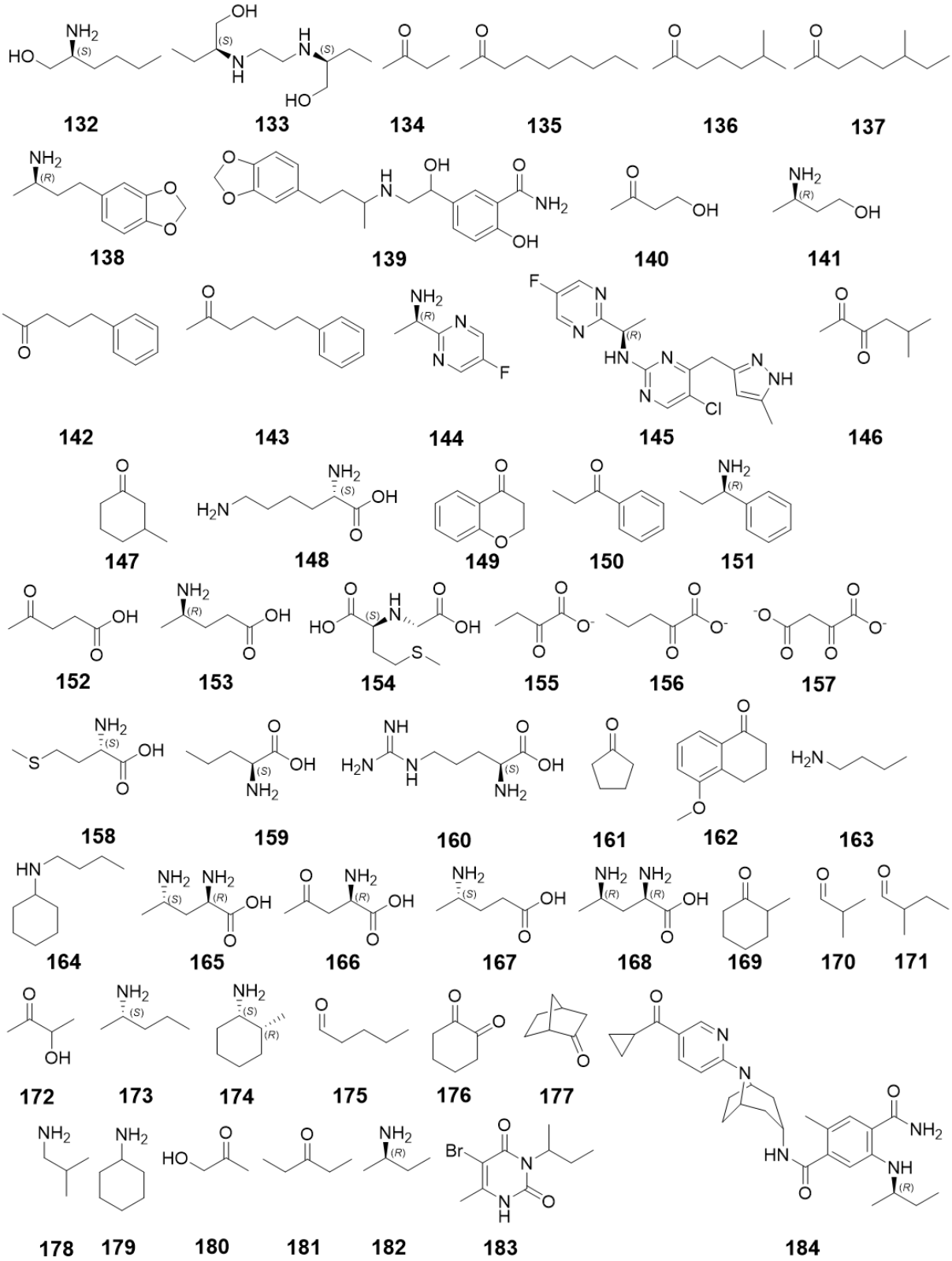
APPENDIX

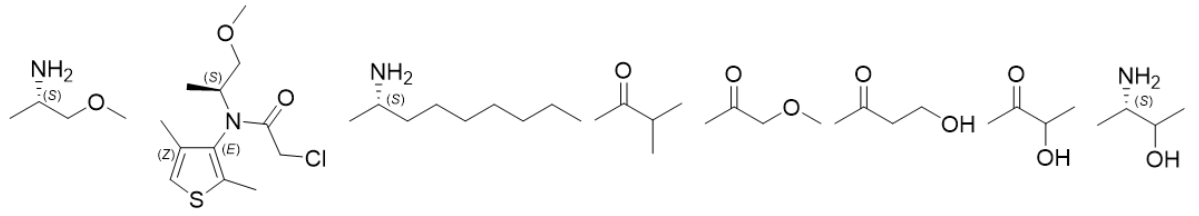
Appendix 1. List of molecules mentioned in the manuscript. The numbering corresponds to the order of appearance in the text. For sake of clarity, the cofactors were drawn at the end, the amino acids were drawn only if they appeared as substrate or product of a reaction and the molecules mentioned only in Experimental section or in Appendix were omitted. The molecule numbering was not used in the text when an abbreviation was given to a molecule (DMSO, IPTG, etc.).

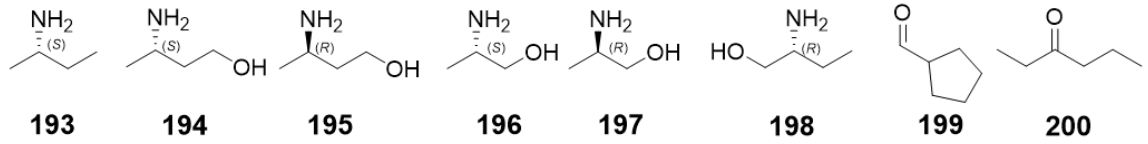


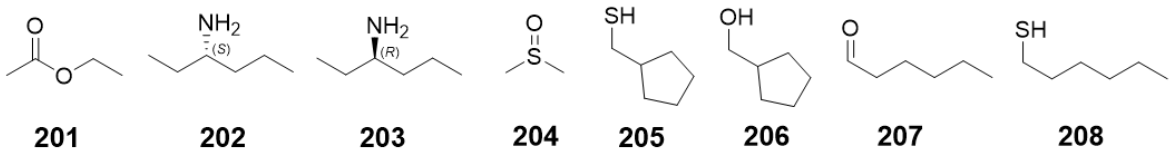

33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82

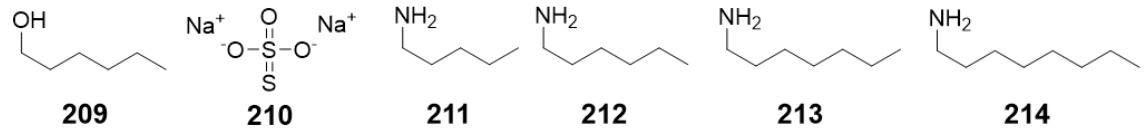

83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131

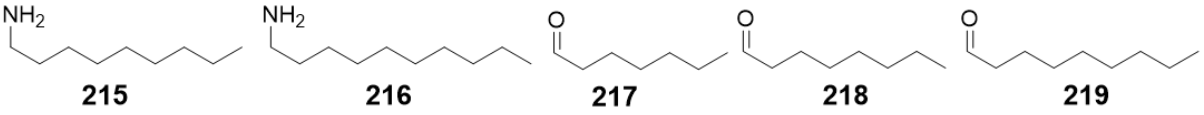

132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184

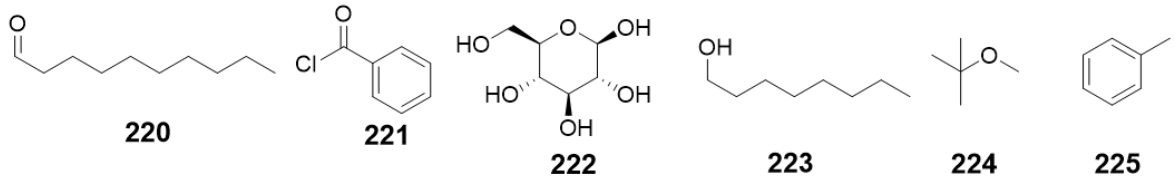

185
186
187
188
189
190
191
192

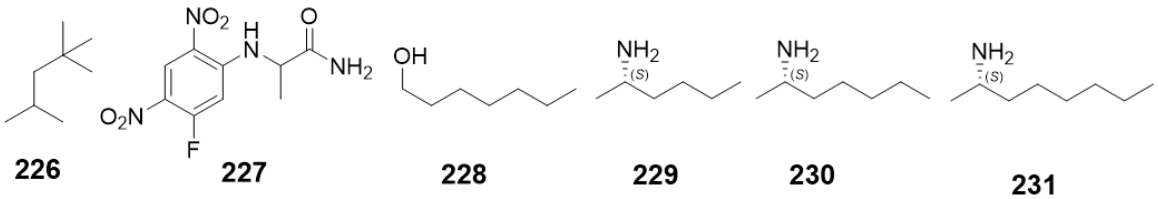

193
194
195
196
197
198
199
200

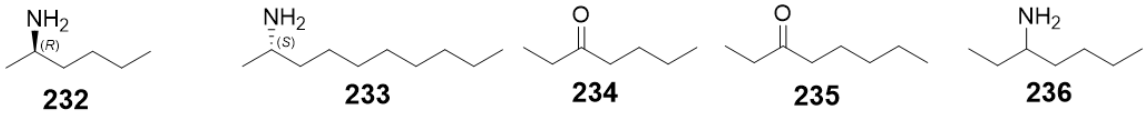

201
202
203
204
205
206
207
208

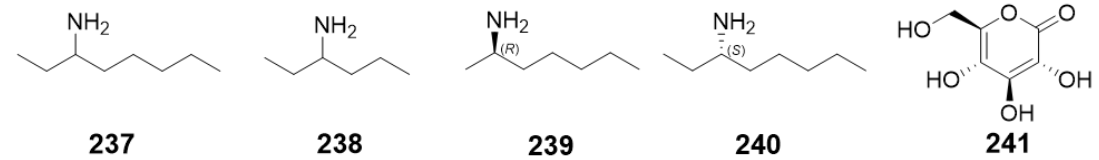

209
210
211
212
213
214

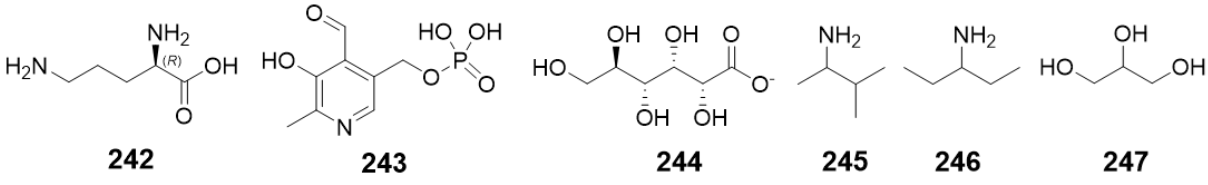

215
216
217
218
219

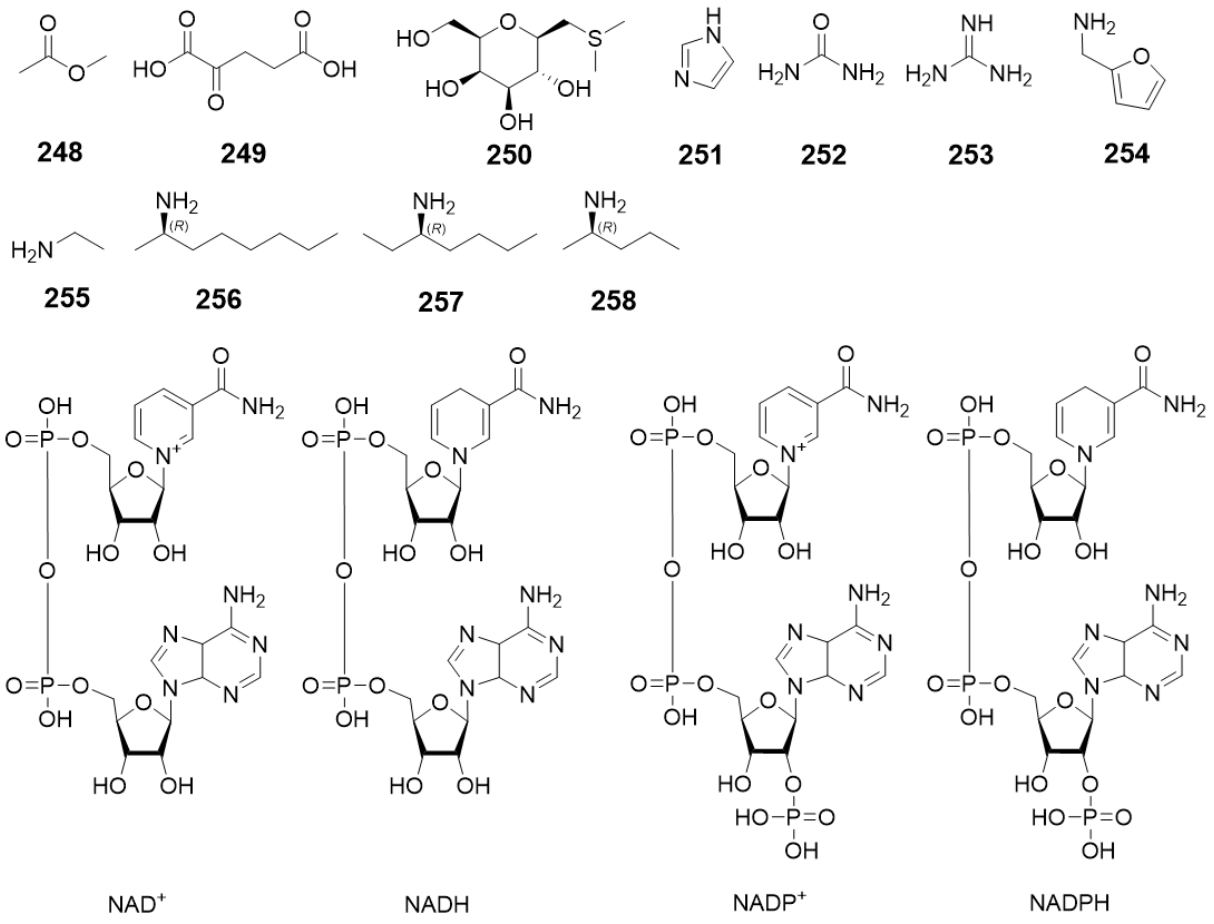

220
221
222
223
224
225


226
227
228
229
230
231


232
233
234
235
236


237
238
239
240
241


242
243
244
245
246
247

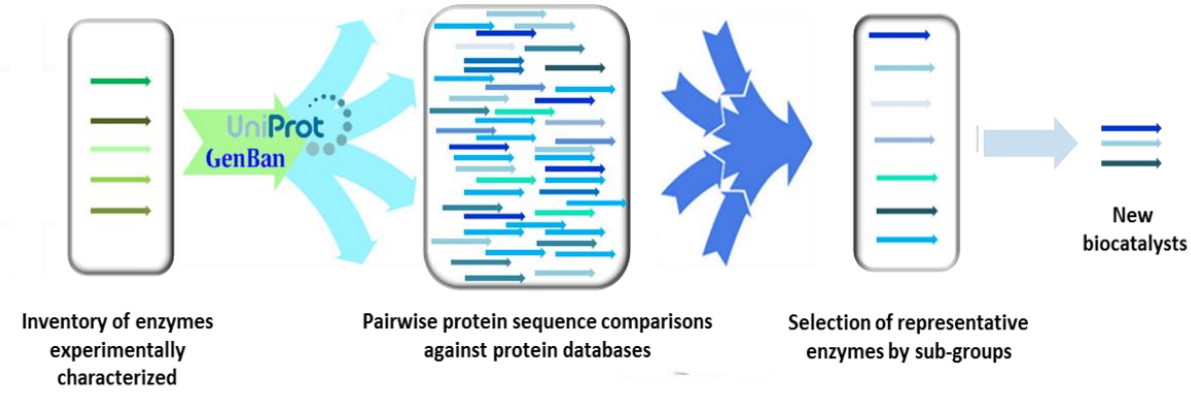


Appendix 2. (Chapter I, I.1.3.1) Algorithms for protein fold prediction. Extracted from Raphaël Guérois's presentation at the Genoscope on 06/09 "New frontiers in structural modeling of interactomes with the advent of AlphaFold2".

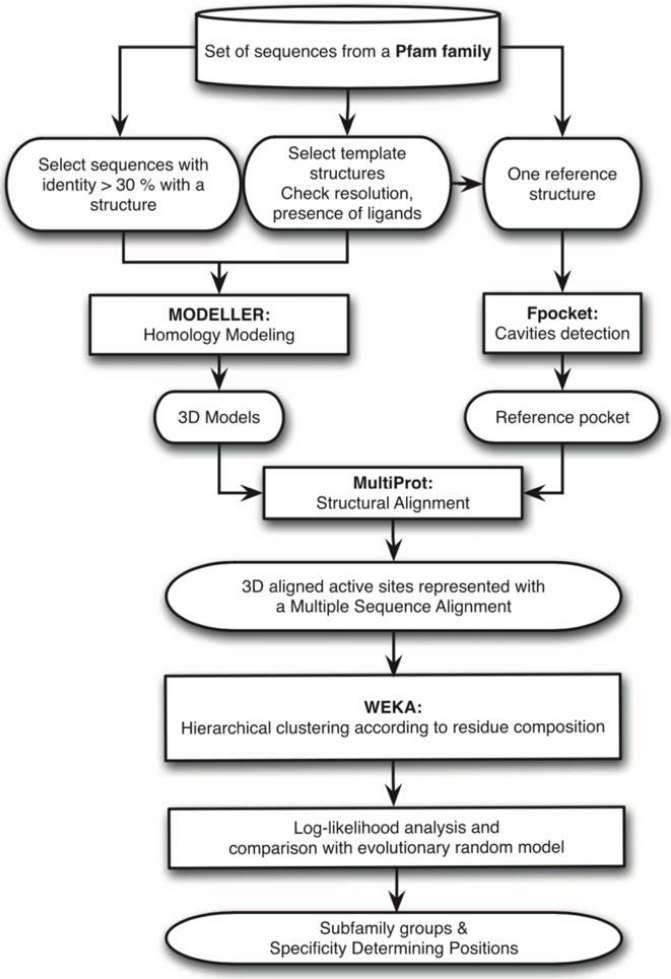
Method	Model_Input	Multimers	Webserver_Colab	Code
AlphaFold2	MSA-based	yes	yes	https://github.com/deepmind/alphafold
ColabFold	MSA-based	yes	yes	https://github.com/sokrypton/ColabFold
FastFold	MSA-based	no*	no*	https://github.com/hpcaitech/FastFold
HelixFold	MSA-based	no*	no*	https://github.com/PaddlePaddle/PaddleHelix/tree/
MEGA-Fold	MSA-based	no*	no*	https://gitee.com/mindspore/mindscience/tree/mas
OpenFold	MSA-based	pending*	yes	https://github.com/aqlaboratory/openfold
RoseTTAFold	MSA-based	no*	yes*	https://github.com/RosettaCommons/RoseTTAFold
Uni-Fold	MSA-based	yes	yes	https://github.com/dptech-corp/Uni-Fold
Uni-Fold-jax	MSA-based	yes	yes	https://github.com/dptech-corp/Uni-Fold-jax
ESM-Fold	pLM-based	no*	no*	N/A
EMBER3D	pLM-based	no*	no*	https://github.com/kWeissenow/EMBER3D
HelixFold-singl	pLM-based	no*	yes	https://github.com/PaddlePaddle/PaddleHelix/tree/
IgFold	pLM-based	no*	yes(mono)	https://github.com/Graylab/IgFold
OmegaFold	pLM-based	pending*	yes	https://github.com/HeliXonProtein/OmegaFold

https://github.com/biolists

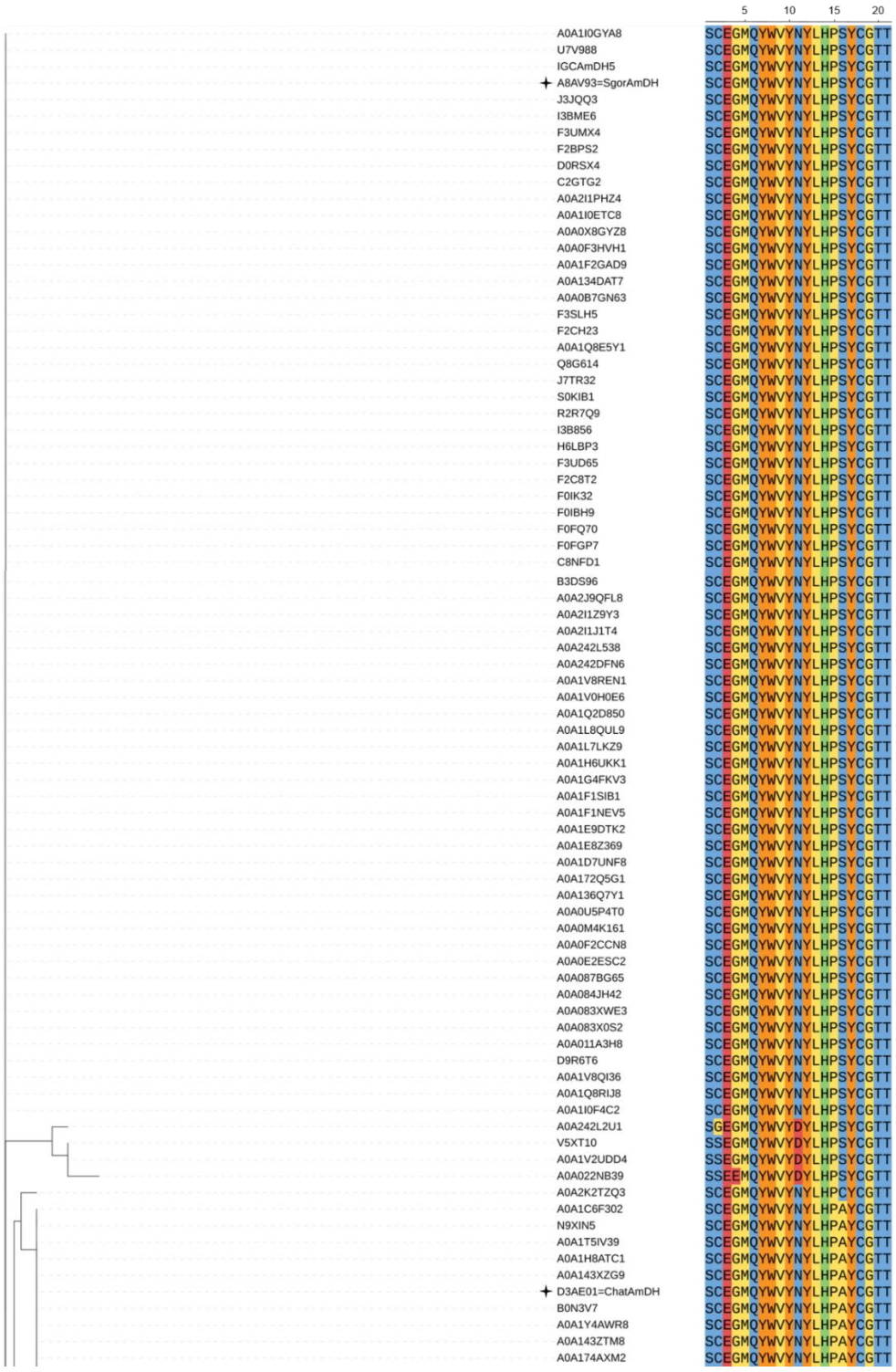
Appendix 3. (Chapter I, II.2.6.1) Genome mining approach of the Cloning and Screening Platform (Genoscope).

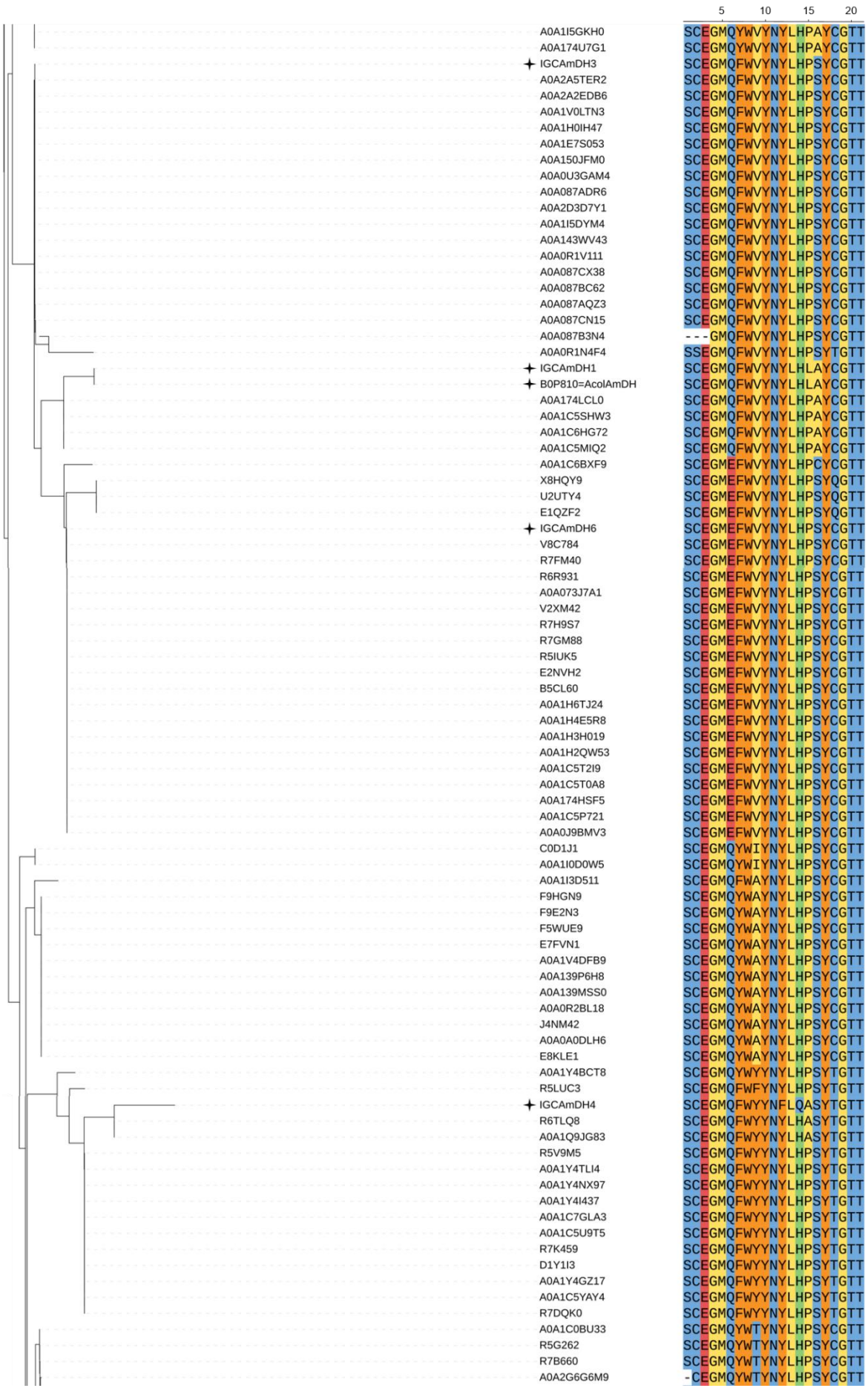


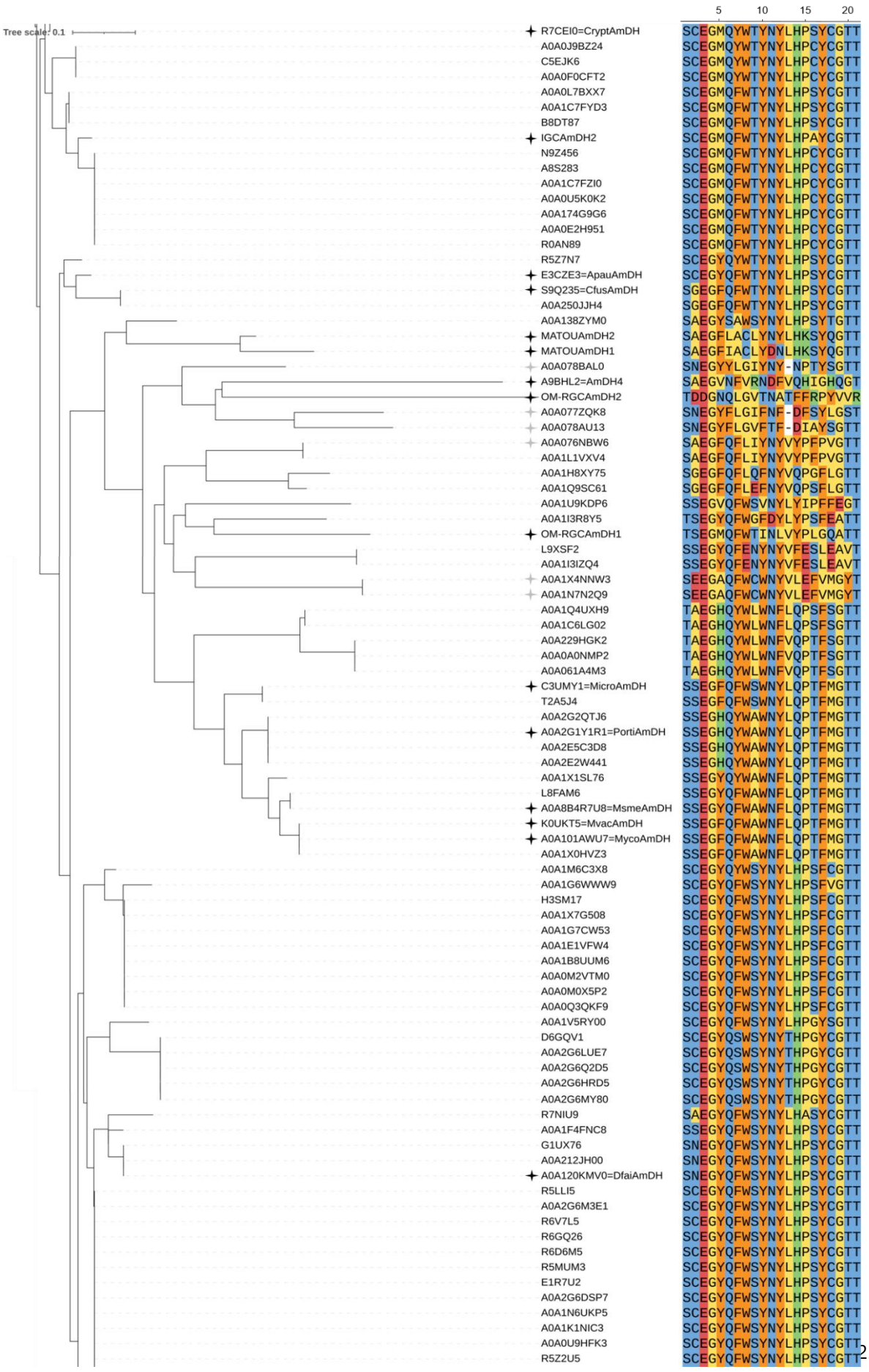
Appendix 4. (Chapter I, II.2.6.2) ASMC pipeline.

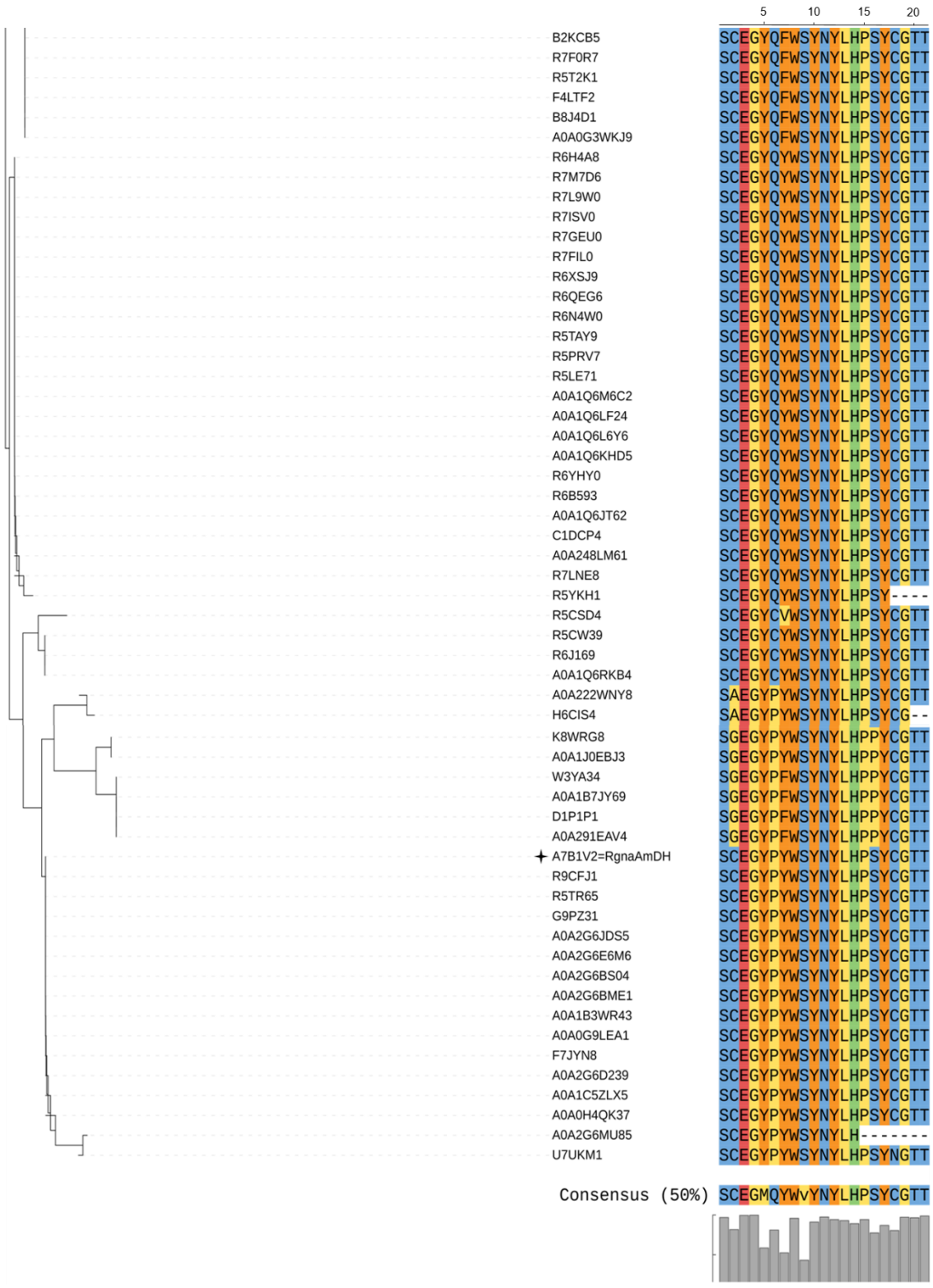


Appendix 5. (Chapter II, II) Comparison of the P1–P21 positions of *Msme*AmDH (G3) and *Cfus*AmDH (G4) active site homologs from ASMC analysis (Mayol *et al.*,2019). All the P1 to P21 residues listed come from an alignment with *Cfus*AmDH (PDB: 6IAU). The tree was generated using the webservice iTOL after alignment of the P1-P21 sequences with Clustal Omega. The color code used refers to the polarity and charge of the corresponding residue [blue: polar residues, yellow: hydrophobic residues, orange: aromatic residues, red: negatively charged residues, and green: positively charged residues (charges at physiological pH)]. The characterized nat-AmDHs from Mayol *et al.* (2016 and 2019) and Caparco *et. al.* (2020) and the nat-AmDHs reported in the main text are annotated with black and grey stars, respectively.

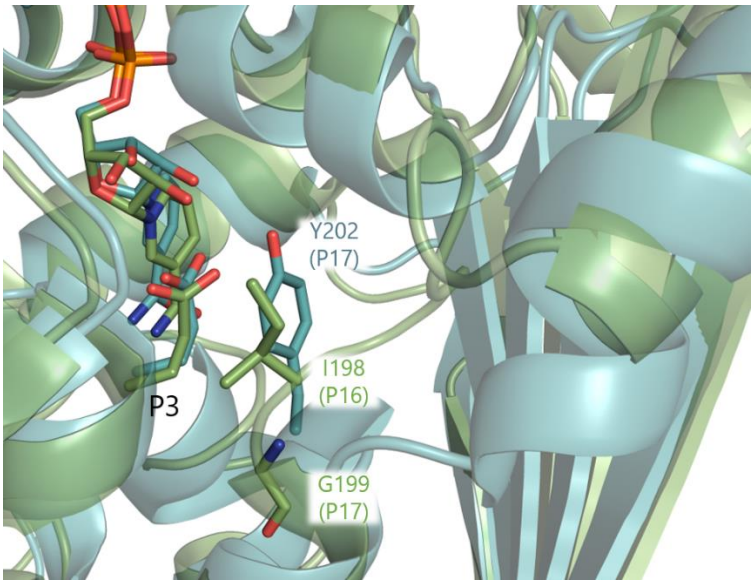



5
10
15
20
A0A1I5GKH0 SCEGMQYWVYNYLHPAYCGTT
A0A174U7G1 SCEGMQYWVYNYLHPAYCGTT
IGCAmDH3 SCEGMQFWVYNYLHPSYCGTT
A0A2A5TER2 SCEGMQFWVYNYLHPSYCGTT
A0A2A2EDB6 SCEGMQFWVYNYLHPSYCGTT
A0A1V0LTN3 SCEGMQFWVYNYLHPSYCGTT
A0A1H0IH47 SCEGMQFWVYNYLHPSYCGTT
A0A1E7S053 SCEGMQFWVYNYLHPSYCGTT
A0A150JFM0 SCEGMQFWVYNYLHPSYCGTT
A0A0U3GAM4 SCEGMQFWVYNYLHPSYCGTT
A0A087ADR6 SCEGMQFWVYNYLHPSYCGTT
A0A2D3D7Y1 SCEGMQFWVYNYLHPSYCGTT
A0A1I5DYM4 SCEGMQFWVYNYLHPSYCGTT
A0A143WV43 SCEGMQFWVYNYLHPSYCGTT
A0A0R1V111 SCEGMQFWVYNYLHPSYCGTT
A0A087CX38 SCEGMQFWVYNYLHPSYCGTT
A0A087BC62 SCEGMQFWVYNYLHPSYCGTT
A0A087AQZ3 SCEGMQFWVYNYLHPSYCGTT
A0A087CN15 SCEGMQFWVYNYLHPSYCGTT
A0A087B3N4 ---GMQFWVYNYLHPSYCGTT
A0A0R1N4F4 SSEGMQFWVYNYLHPSYTGTT
IGCAmDH1 SCEGMQFWVYNYLHLAYCGTT
B0P810=AcolAmDH SCEGMQFWVYNYLHLAYCGTT
A0A174LCL0 SCEGMQFWVYNYLHPAYCGTT
A0A1C5SHW3 SCEGMQFWVYNYLHPAYCGTT
A0A1C6HG72 SCEGMQFWVYNYLHPAYCGTT
A0A1C5MIQ2 SCEGMQFWVYNYLHPAYCGTT
A0A1C6BXF9 SCEGMEFWVYNYLHPCYCGTT
X8HQY9 SCEGMEFWVYNYLHPSYQGTT
U2UTY4 SCEGMEFWVYNYLHPSYQGTT
E1QZF2 SCEGMEFWVYNYLHPSYQGTT
IGCAmDH6 SCEGMEFWVYNYLHPSYCGTT
V8C784 SCEGMEFWVYNYLHPSYCGTT
R7FM40 SCEGMEFWVYNYLHPSYCGTT
R6R931 SCEGMEFWVYNYLHPSYCGTT
A0A073J7A1 SCEGMEFWVYNYLHPSYCGTT
V2XM42 SCEGMEFWVYNYLHPSYCGTT
R7H9S7 SCEGMEFWVYNYLHPSYCGTT
R7GM88 SCEGMEFWVYNYLHPSYCGTT
R5IUK5 SCEGMEFWVYNYLHPSYCGTT
E2NVH2 SCEGMEFWVYNYLHPSYCGTT
B5CL60 SCEGMEFWVYNYLHPSYCGTT
A0A1H6TJ24 SCEGMEFWVYNYLHPSYCGTT
A0A1H4E5R8 SCEGMEFWVYNYLHPSYCGTT
A0A1H3H019 SCEGMEFWVYNYLHPSYCGTT
A0A1H2QW53 SCEGMEFWVYNYLHPSYCGTT
A0A1C5T2I9 SCEGMEFWVYNYLHPSYCGTT
A0A1C5T0A8 SCEGMEFWVYNYLHPSYCGTT
A0A174HSF5 SCEGMEFWVYNYLHPSYCGTT
A0A1C5P721 SCEGMEFWVYNYLHPSYCGTT
A0A0J9BMV3 SCEGMEFWVYNYLHPSYCGTT
C0D1J1 SCEGMQYWIYNYLHPSYCGTT
A0A1I0D0W5 SCEGMQYWIYNYLHPSYCGTT
A0A1I3D511 SCEGMQFWAYNYLHPSYCGTT
F9HGN9 SCEGMQYWAYNYLHPSYCGTT
F9E2N3 SCEGMQYWAYNYLHPSYCGTT
F5WUE9 SCEGMQYWAYNYLHPSYCGTT
E7FVN1 SCEGMQYWAYNYLHPSYCGTT
A0A1V4DFB9 SCEGMQYWAYNYLHPSYCGTT
A0A139P6H8 SCEGMQYWAYNYLHPSYCGTT
A0A139MSS0 SCEGMQYWAYNYLHPSYCGTT
A0A0R2BL18 SCEGMQYWAYNYLHPSYCGTT
J4NM42 SCEGMQYWAYNYLHPSYCGTT
A0A0A0DLH6 SCEGMQYWAYNYLHPSYCGTT
E8KLE1 SCEGMQYWAYNYLHPSYCGTT
A0A1Y4BCT8 SCEGMQYWYYNYLHPSYTGTT
R5LUC3 SCEGMQFWFYNYLHPSYTGTT
IGCAmDH4 SCEGMQFWYYNFLQASYTGTT
R6TLQ8 SCEGMQFWYYNYLHASYTGTT
A0A1Q9JG83 SCEGMQFWYYNYLHASYTGTT
R5V9M5 SCEGMQFWYYNYLHPSYTGTT
A0A1Y4TLI4 SCEGMQFWYYNYLHPSYTGTT
A0A1Y4NX97 SCEGMQFWYYNYLHPSYTGTT
A0A1Y4I437 SCEGMQFWYYNYLHPSYTGTT
A0A1C7GLA3 SCEGMQFWYYNYLHPSYTGTT
A0A1C5U9T5 SCEGMQFWYYNYLHPSYTGTT
R7K459 SCEGMQFWYYNYLHPSYTGTT
D1Y1I3 SCEGMQFWYYNYLHPSYTGTT
A0A1Y4GZ17 SCEGMQFWYYNYLHPSYTGTT
A0A1C5YAY4 SCEGMQFWYYNYLHPSYTGTT
R7DQK0 SCEGMQFWYYNYLHPSYTGTT
A0A1C0BU33 SCEGMQYWTYNYLHPSYCGTT
R5G262 SCEGMQYWTYNYLHPSYCGTT
R7B660 SCEGMQYWTYNYLHPSYCGTT
A0A2G6G6M9 -CEGMQYWTYNYLHPSYCGTT

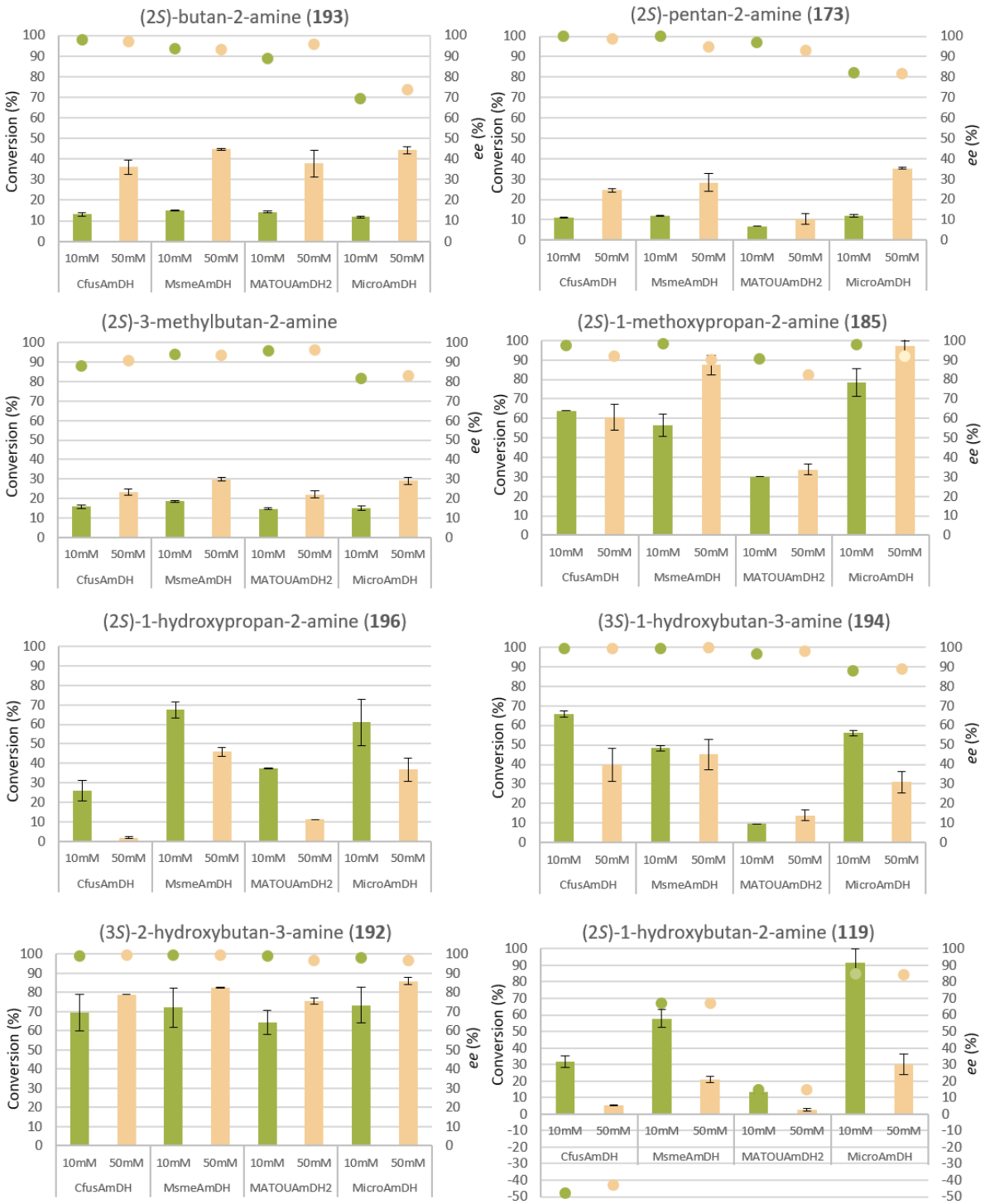

Tree scale: 0.1
5
10
15
20
R7CEI0=CryptAmDH
A0A0J9BZ24
C5EJK6
A0A0F0CFT2
A0A0L7BXX7
A0A1C7FYD3
B8DT87
IGCAmDH2
N9Z456
A8S283
A0A1C7FZI0
A0A0U5K0K2
A0A174G9G6
A0A0E2H951
R0AN89
R5Z7N7
E3CZE3=ApauAmDH
S9Q235=CfusAmDH
A0A250JJH4
A0A138ZYM0
MATOUAmDH2
MATOUAmDH1
A0A078BAL0
A9BHL2=AmDH4
OM-RGCAmDH2
A0A077ZQK8
A0A078AU13
A0A076NBW6
A0A1L1VXV4
A0A1H8XY75
A0A1Q9SC61
A0A1U9KDP6
A0A1I3R8Y5
OM-RGCAmDH1
L9XSF2
A0A1I3IZQ4
A0A1X4NNW3
A0A1N7N2Q9
A0A1Q4UXH9
A0A1C6LG02
A0A229HGK2
A0A0A0NMP2
A0A061A4M3
C3UMY1=MicroAmDH
T2A5J4
A0A2G2QTJ6
A0A2G1Y1R1=PortiAmDH
A0A2E5C3D8
A0A2E2W441
A0A1X1SL76
L8FAM6
A0A8B4R7U8=MsmeAmDH
K0UKT5=MvacAmDH
A0A101AWU7=MycoAmDH
A0A1X0HVZ3
A0A1M6C3X8
A0A1G6WWW9
H3SM17
A0A1X7G508
A0A1G7CW53
A0A1E1VFW4
A0A1B8UUM6
A0A0M2VTM0
A0A0M0X5P2
A0A0Q3QKF9
A0A1V5RY00
D6GQV1
A0A2G6LUE7
A0A2G6Q2D5
A0A2G6HRD5
A0A2G6MY80
R7NIU9
A0A1F4FNC8
G1UX76
A0A212JH00
A0A120KMV0=DfaiAmDH
R5LLI5
A0A2G6M3E1
R6V7L5
R6GQ26
R6D6M5
R5MUM3
E1R7U2
A0A2G6DSP7
A0A1N6UKP5
A0A1K1NIC3
A0A0U9HFK3
R5Z2U5


5
10
15
20
B2KCB5 SCEGYQFWSYNYLHPSYCGTT
R7F0R7 SCEGYQFWSYNYLHPSYCGTT
R5T2K1 SCEGYQFWSYNYLHPSYCGTT
F4LTF2 SCEGYQFWSYNYLHPSYCGTT
B8J4D1 SCEGYQFWSYNYLHPSYCGTT
A0A0G3WKJ9 SCEGYQFWSYNYLHPSYCGTT
R6H4A8 SCEGYQYWSYNYLHPSYCGTT
R7M7D6 SCEGYQYWSYNYLHPSYCGTT
R7L9W0 SCEGYQYWSYNYLHPSYCGTT
R7ISV0 SCEGYQYWSYNYLHPSYCGTT
R7GEU0 SCEGYQYWSYNYLHPSYCGTT
R7FIL0 SCEGYQYWSYNYLHPSYCGTT
R6XSJ9 SCEGYQYWSYNYLHPSYCGTT
R6QEG6 SCEGYQYWSYNYLHPSYCGTT
R6N4W0 SCEGYQYWSYNYLHPSYCGTT
R5TAY9 SCEGYQYWSYNYLHPSYCGTT
R5PRV7 SCEGYQYWSYNYLHPSYCGTT
R5LE71 SCEGYQYWSYNYLHPSYCGTT
A0A1Q6M6C2 SCEGYQYWSYNYLHPSYCGTT
A0A1Q6LF24 SCEGYQYWSYNYLHPSYCGTT
A0A1Q6L6Y6 SCEGYQYWSYNYLHPSYCGTT
A0A1Q6KHD5 SCEGYQYWSYNYLHPSYCGTT
R6YHY0 SCEGYQYWSYNYLHPSYCGTT
R6B593 SCEGYQYWSYNYLHPSYCGTT
A0A1Q6JT62 SCEGYQYWSYNYLHPSYCGTT
C1DCP4 SCEGYQYWSYNYLHPSYCGTT
A0A248LM61 SCEGYQYWSYNYLHPSYCGTT
R7LNE8 SCEGYQYWSYNYLHPSYCGTT
R5YKH1 SCEGYQYWSYNYLHPSY----
R5CSD4 SCEGYCVWSYNYLHPSYCGTT
R5CW39 SCEGYCYWSYNYLHPSYCGTT
R6J169 SCEGYCYWSYNYLHPSYCGTT
A0A1Q6RKB4 SCEGYCYWSYNYLHPSYCGTT
A0A222WNY8 SAEGYPYWSYNYLHPSYCGTT
H6CIS4 SAEGYPYWSYNYLHPSYCG--
K8WRG8 SGEGYPYWSYNYLHPPYCGTT
A0A1J0EBJ3 SGEGYPYWSYNYLHPPYCGTT
W3YA34 SGEGYPFWSYNYLHPPYCGTT
A0A1B7JY69 SGEGYPFWSYNYLHPPYCGTT
D1P1P1 SGEGYPFWSYNYLHPPYCGTT
A0A291EAV4 SGEGYPFWSYNYLHPPYCGTT
A7B1V2=RgnaAmDH SCEGYPYWSYNYLHPSYCGTT
R9CFJ1 SCEGYPYWSYNYLHPSYCGTT
R5TR65 SCEGYPYWSYNYLHPSYCGTT
G9PZ31 SCEGYPYWSYNYLHPSYCGTT
A0A2G6JDS5 SCEGYPYWSYNYLHPSYCGTT
A0A2G6E6M6 SCEGYPYWSYNYLHPSYCGTT
A0A2G6BS04 SCEGYPYWSYNYLHPSYCGTT
A0A2G6BME1 SCEGYPYWSYNYLHPSYCGTT
A0A1B3WR43 SCEGYPYWSYNYLHPSYCGTT
A0A0G9LEA1 SCEGYPYWSYNYLHPSYCGTT
F7JYN8 SCEGYPYWSYNYLHPSYCGTT
A0A2G6D239 SCEGYPYWSYNYLHPSYCGTT
A0A1C5ZLX5 SCEGYPYWSYNYLHPSYCGTT
A0A0H4QK37 SCEGYPYWSYNYLHPSYCGTT
A0A2G6MU85 SCEGYPYWSYNYLH-------
U7UKM1 SCEGYPYWSYNYLHPSYNGTT
Consensus (50%) SCEGMQYWvYNYLHPSYCGTT

Appendix 6. (ChapterII, II.3) Structural alignment of *Cfus*AmDH and AmDH4 of comparison of P17 positions and surroundings. *Cfus*AmDH (PDB: 6IAU, chain B) and AmDH4 (PDB: 6G1M, chain C) are colored in blue and green, respectively. P3 for both enzymes, Y202 (P17) of *Cfus*AmDH and I198 (P16) and G199 (P17) of AmDH4 are shown in stick format. RMSD (AmDH4-*Cfus*AmDH) = 1.897 Å.

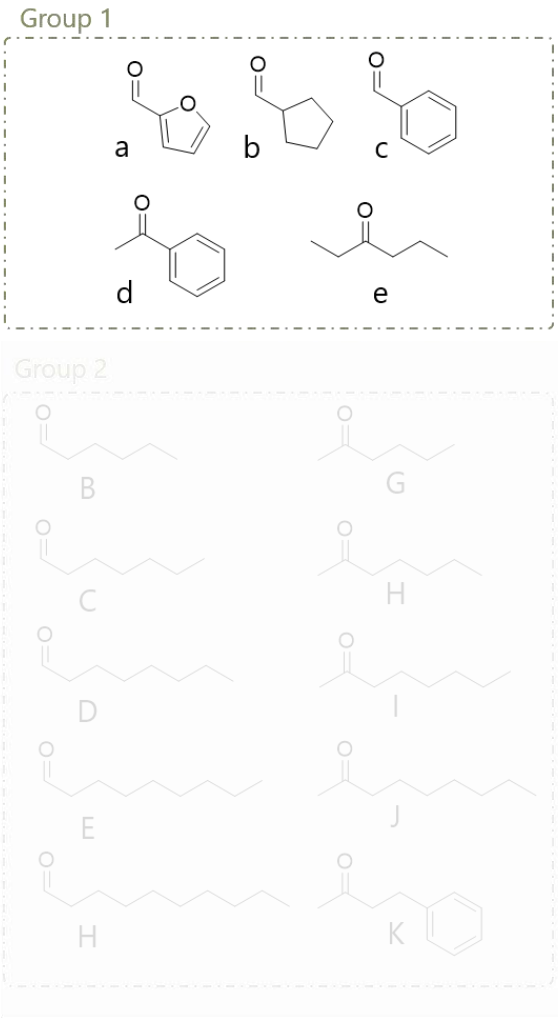


Appendix 7. (Chapter II, IV.1.1) Conversions and *ee* results obtained with *Cfus*AmDH, *Msme*AmDH, MATOUAmDH2 and *Micro*AmDH towards short chiral amines. Reaction conditions: 10 mM or 50 mM substrate, 2 M NH_4HCO_2 buffer pH 9.0, 0.2 mM $NADP^+$, 1.1 eq. glucose (**222**), 3 U mL^{-1} GDH-105, 0.5 mg mL^{-1} purified nat-AmDH, 24 h, 30°C. Bars refer to the conversion rate and the dots to the corresponding *ee* (for the (*S*)-enantiomer if >0 and for the (*R*)-enantiomer if <0). Error bars represent standard deviation of two or three independent experiments. Due to coelution with FDAA-derivatized co-product, the *ee* of (2*S*)-1-hydroxy-propan-2-amine (**196**) could not be calculated, but the (*S*)-enantiomer was the major peak.



Appendix 8. (Chapter II, IV.2) Enzymes and substrates annotations for GC-FID and UHPLC-UV chromatograms from the screening of nat-AmDHs towards a range of short aliphatic and aromatic aldehydes and ketones. The group of substrates named Group 2 and the mutated enzymes that are indinstinct concern the rest of the screening that is discussed in Chapter III, I.4. The chromatograms given in Appendix 9 to Appendix 15 correspond to the whole screening but the enzymes numbering and substrates annotation enable to only focus on the results discussed here.

A

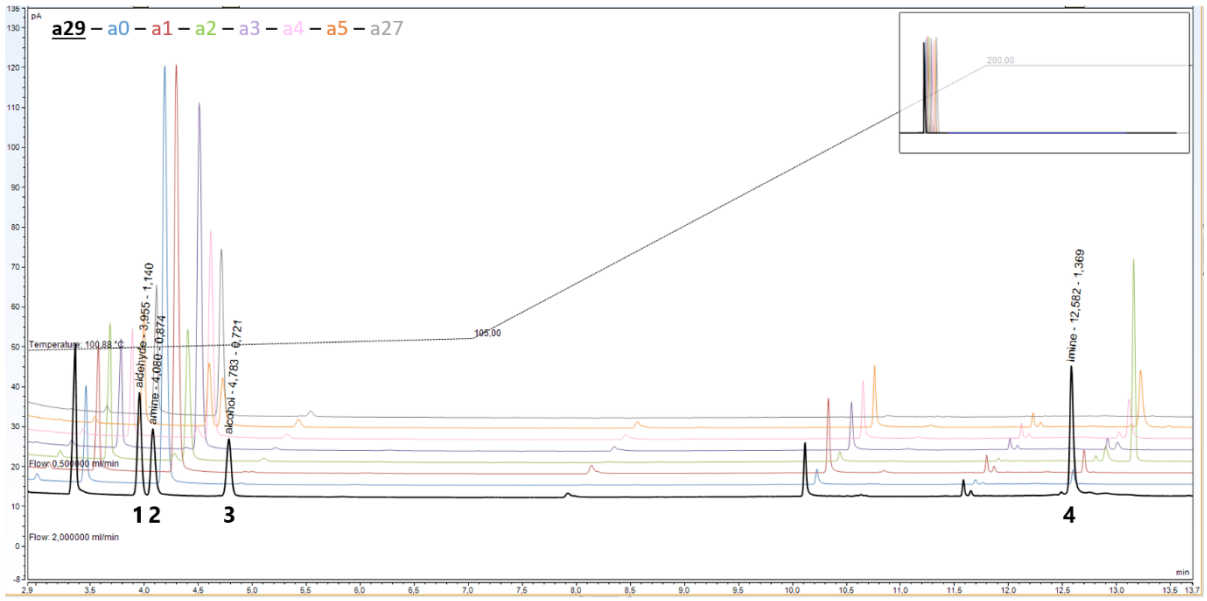


B

0	***Cfus* AmDH**
1	*Cfus* AmDH-F140A
2	*Cfus* AmDH-W145A
3	***Apau* AmDH**
4	*Apau* AmDH-Y136A
5	*Apau* AmDH-W141A
6	***Chat* AmDH**
7	*Chat* AmDH-M161A
8	*Chat* AmDH-W166A
9	**IGCAmDH1**
10	IGCAmDH1-M139A
11	IGCAmDH1-W144A
12	**MATOUAmDH2**
13	MATOUAmDH2-F143A
14	MATOUAmDH2-C148A
15	***Micro* AmDH**
16	*Micro* AmDH-F136A
17	*Micro* AmDH-W141A
18	***Msme* AmDH**
19	*Msme* AmDH-Y136A
20	*Msme* AmDH-W141A
21	***Porti* AmDH**
22	*Porti* AmDH-H135A
23	*Porti* AmDH-W140A
24	***Rgna* AmDH**
25	*Rgna* AmDH-Y150A
26	*Rgna* AmDH-W155A
27	without enzyme
28-31	commercial product

Appendix 9. (Chapter II, IV.2.1.1 and Chapter III, I.4.4) GC-FID chromatograms for reaction of furfural (**57**) with various nat-AmDHs and mutants. Chromatograms of a commercial standard of furfurylamine (**254**) (2 mM), **57** (7.2 mM), and corresponding alcohol (0.8 mM), blank reaction mixture without enzyme and reaction mixture with (A) *Cfus*AmDH, *Apau*AmDH and their mutants (B) *Chat*AmDH, IGCAmDH1 and their mutants (C) MATOUAmDH2, *Micro*AmDH and their mutants (D) *Msme*AmDH, *Porti*AmDH and their mutants and (E) *Rgna*AmDH and its mutants after extraction in EtOAc. The imine formed in the extraction layer between the amine and the remaining aldehyde could be detected at 12.6 min. The legend above matches the color of the chromatogram with the enzyme/substrate couple according to Appendix 8. The legend below indicates: compound; retention time (min); area.

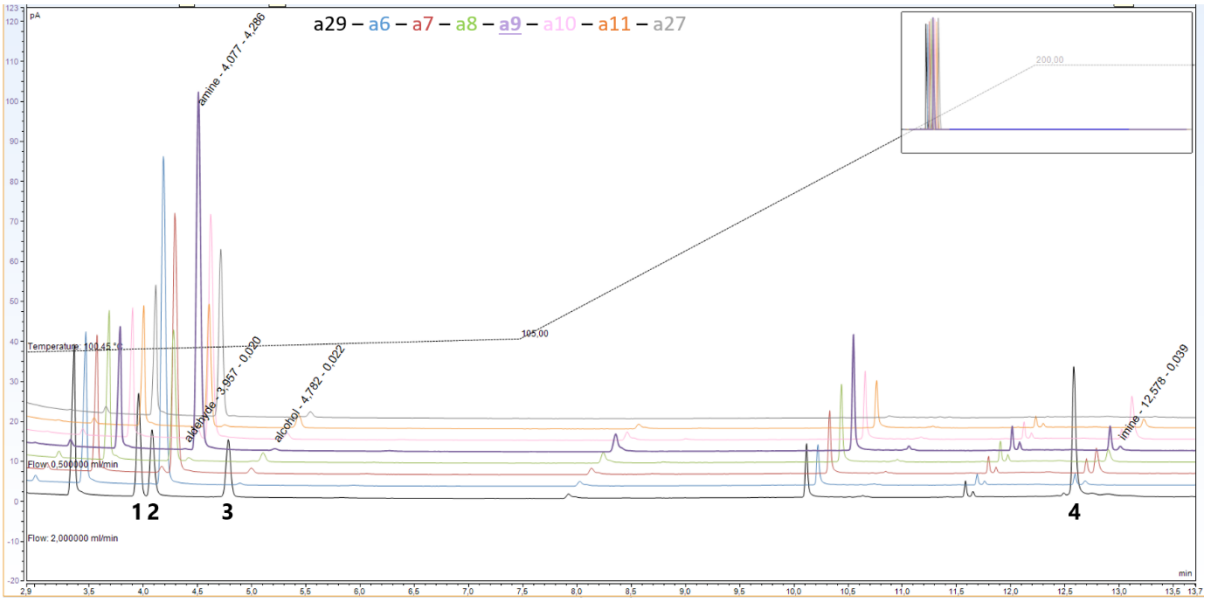
A



1: furfural; 3.955; 1.140 **2**: furfurylamine; 4.080; 0.874

3: furfurylalcohol; 4.783; 0.721 **4**: *N*-furfuryl-1-furfurylimine; 12.582; 1.369

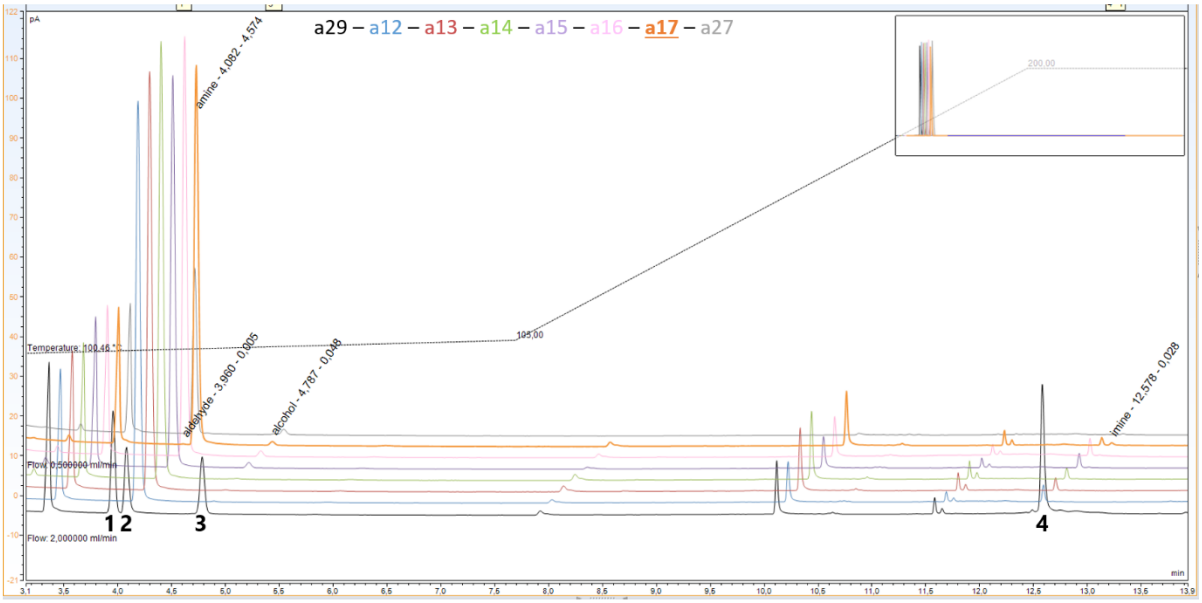
B



1: furfural; 3.957; 0.020 **2**: furfurylamine; 4.077; 4.286

3: furfurylalcohol; 4.782; 0.022 **4**: *N*-furfuryl-1-furfurylimine; 12.578; 0.039

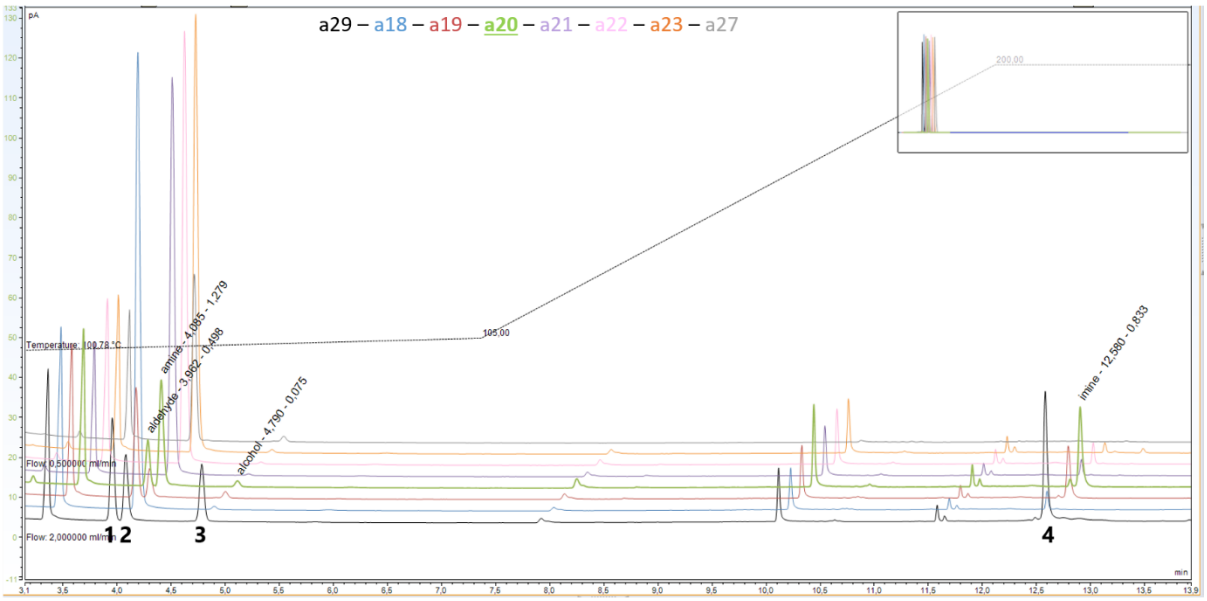
C



1: furfural; 3.960; 0.005 **2**: furfurylamine; 4.082; 4.574

3: furfurylalcohol; 4.787; 0.048 **4**: *N*-furfuryl-1-furfurylimine; 12.578; 0.028

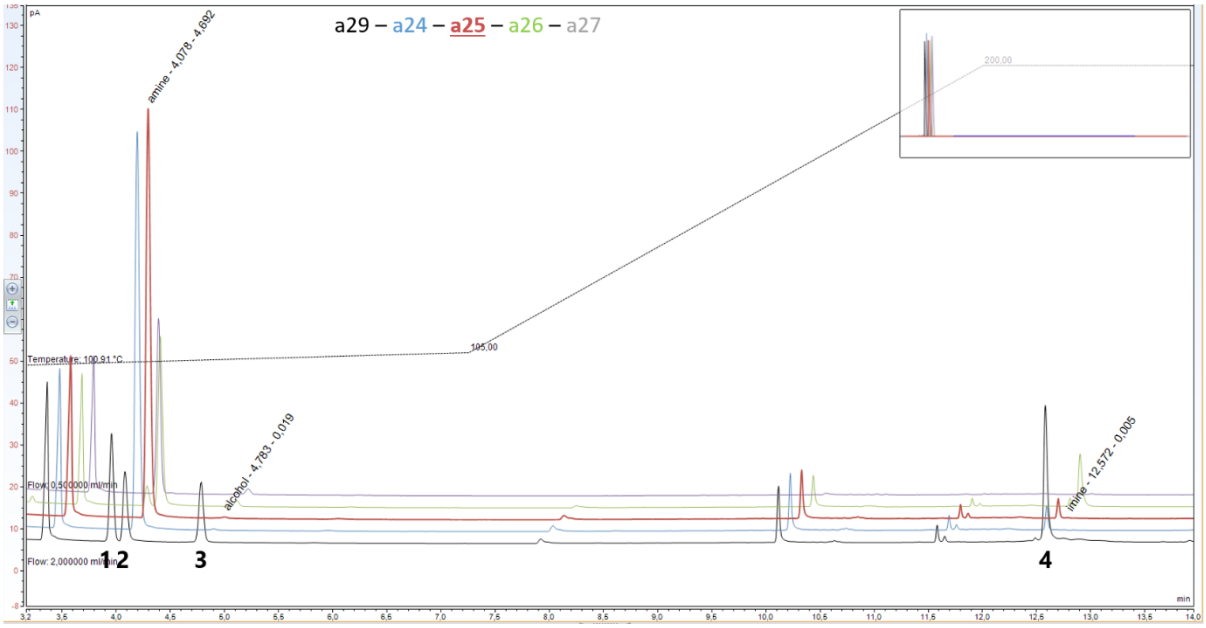
D



1: furfural; 3.962; 0.498 **2**: furfurylamine; 4.085; 1.279

3: furfurylalcohol; 4.790; 0.075 **4**: *N*-furfuryl-1-furfurylimine; 12.580; 0.833

E



1: /

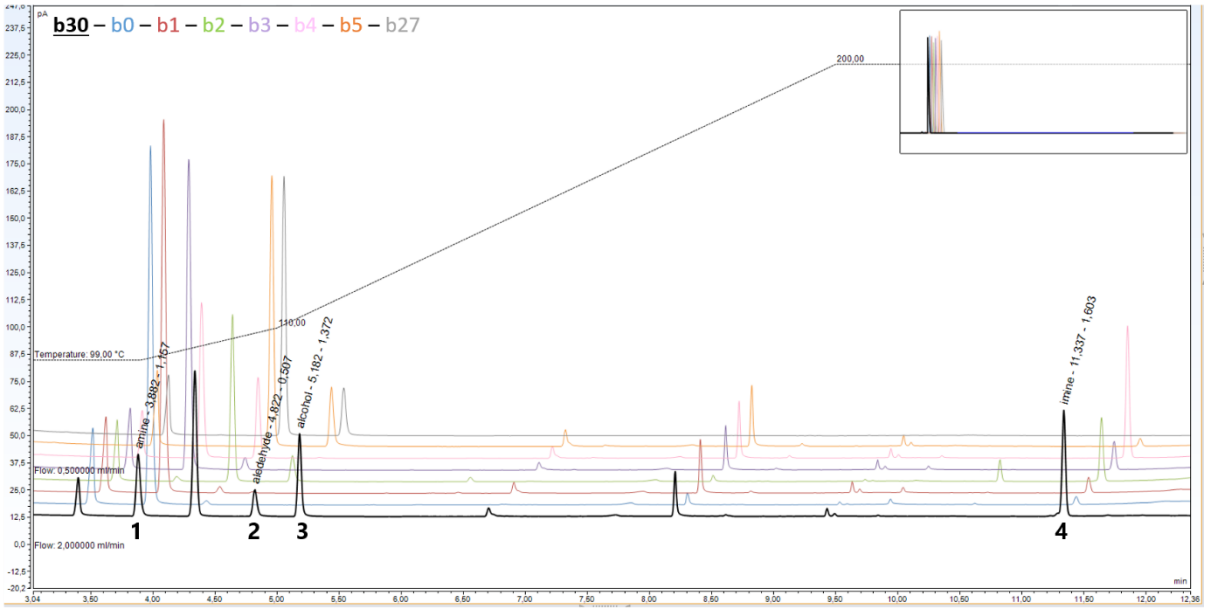
2: furfurylamine; 4.078; 4.692

3: furfurylalcohol; 4.783; 0.019

4: *N*-furfuryl-1-furfurylimine; 12.572; 0.005

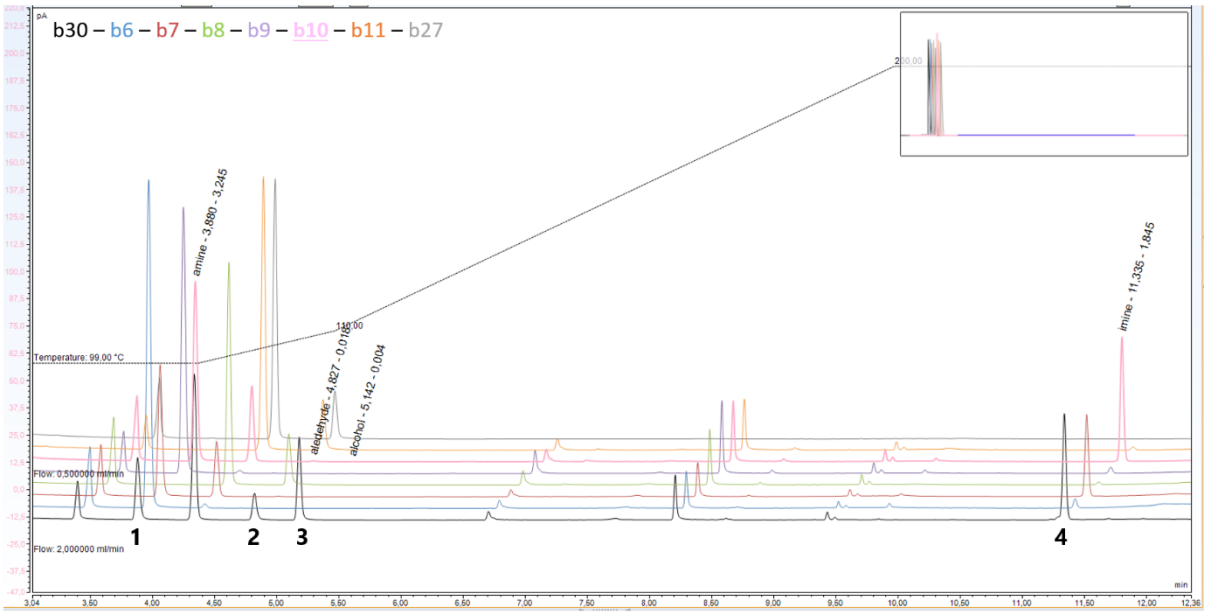
Appendix 10. (Chapter II, IV.2.1.1 and Chapter III, I.4.4) GC-FID chromatograms for reaction of cyclopentancarbaldehyde (**199**) with various nat-AmDHs and mutants. Chromatograms of a commercial standard of amine (3 mM), **199** (6.4 mM), and corresponding alcohol (0.6 mM), blank reaction mixture without enzyme and reaction mixture with (A) *Cfus*AmDH, *Apau*AmDH and their mutants (B) *Chat*AmDH, IGCAmDH1 and their mutants (C) MATOUAmDH2, *Micro*AmDH and their mutants (D) *Msme*AmDH, *Porti*AmDH and their mutants and (E) *Rgna*AmDH and its mutants after extraction in EtOAc. The imine formed in the extraction layer between the amine and the remaining aldehyde could be detected at 11.3 min. The peak at 4.3 min is yet unidentified but does not impact the linearity of the amine or alcohol response in the calibration curves. It is currently under further analyses. The legend above matches the color of the chromatogram with the enzyme/substrate couple according to Appendix 8. The legend below indicates: compound; retention time (min); area.

A



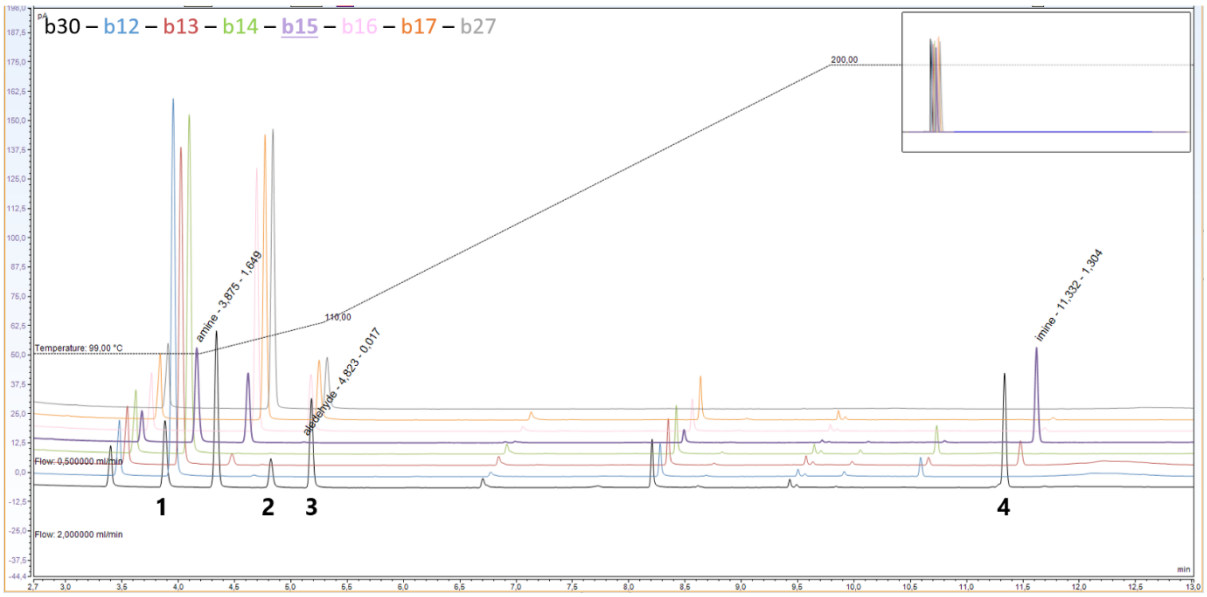
1: cyclopentanmethylamine; 3.882; 1.157 **2**: cyclopentancarbaldehyde; 4.822; 0.507
3: cyclopentylmethanol; 5.182; 1.372 **4**: *N*-cyclopentyl-1-cyclopentylimine; 11.337; 1.603

B



1: cyclopentanmethylamine; 3.880; 3.245 **2**: cyclopentancarbaldehyde; 4.827; 0.018
3: cyclopentylmethanol; 5.142; 0.004 **4**: *N*-cyclopentyl-1-cyclopentylimine; 11.335; 1.845

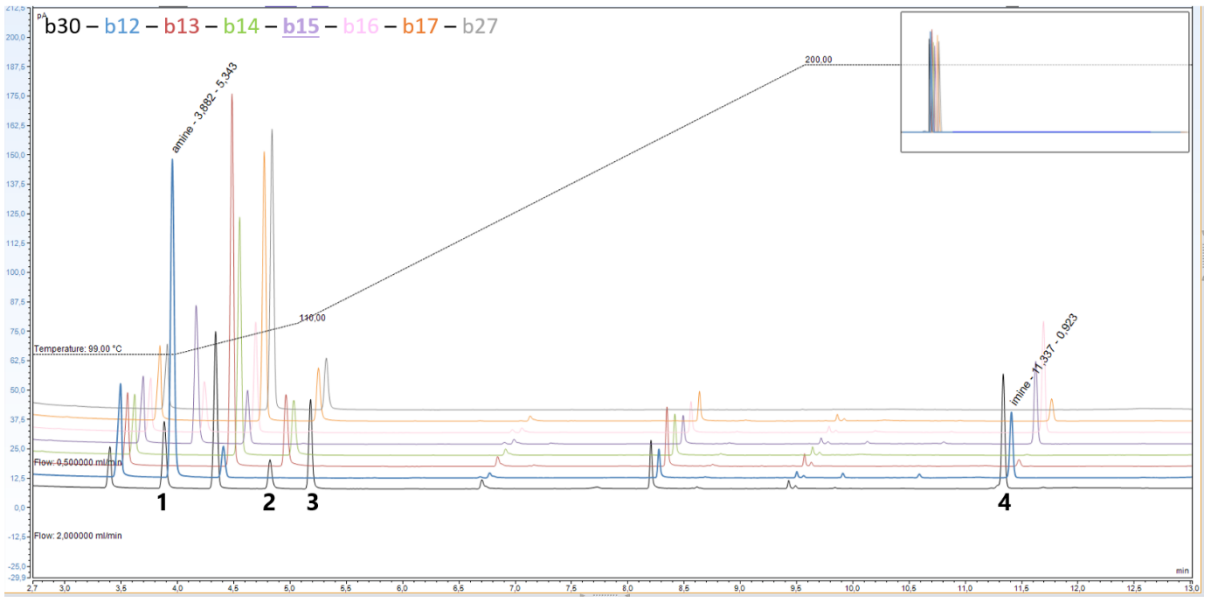
C



1: cyclopentanmethylamine; 3.875; 1.649 **2**: cyclopentancarbaldehyde; 4.823; 0.017

3: / **4**: *N*-cyclopentyl-1-cyclopentylimine; 11.332; 1.304

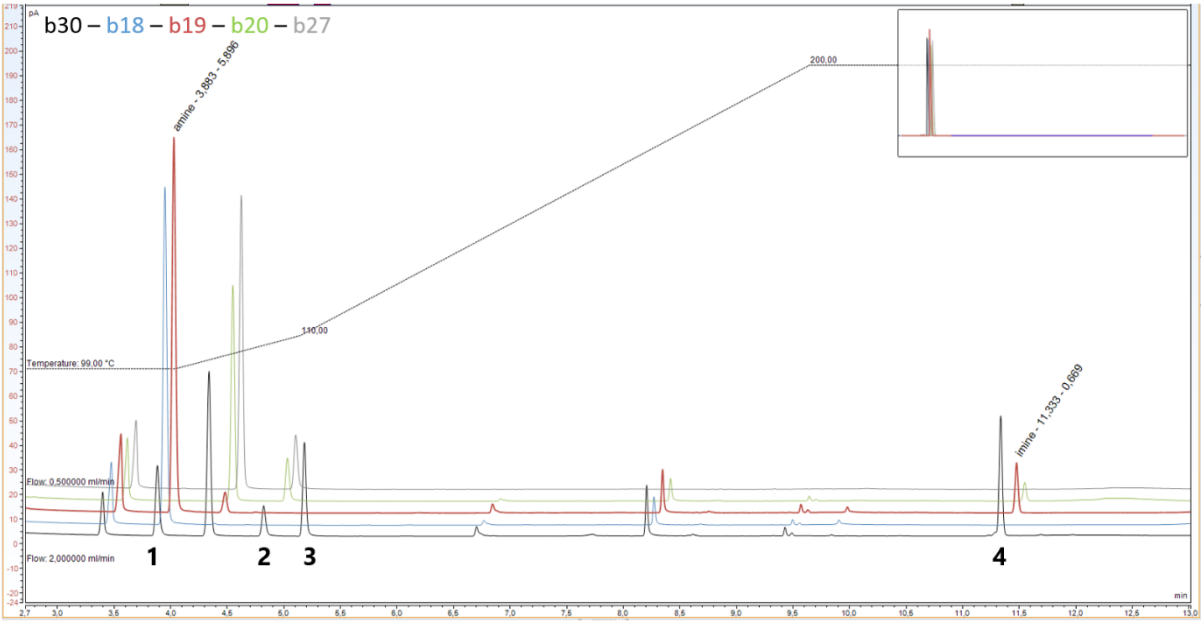
D



1: cyclopentanmethylamine; 3.875; 5.343 **2**: /

3: / **4**: *N*-cyclopentyl-1-cyclopentylimine; 11.337; 0.923

E

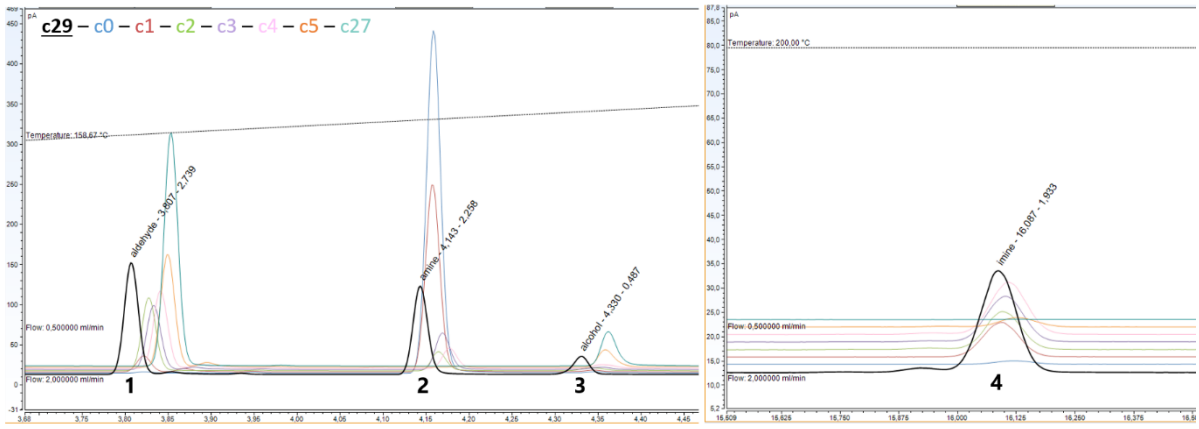


1: cyclopentanmethylamine; 3.883; 5.896 **2**: /

3: / **4**: *N*-cyclopentyl-1-cyclopentylimine; 11.333; 0.669

Appendix 11. (Chapter II, IV.2.1.1 and Chapter III, I.4.4) GC-FID chromatograms for reaction of benzaldehyde (**31**) with various nat-AmDHs and mutants. Chromatograms of a commercial standard of benzylamine **74** (5 mM), **31** (4.6 mM), and corresponding alcohol (0.4 mM), blank reaction mixture without enzyme and reaction mixture with (A) *Cfus*AmDH, *Apau*AmDH and their mutants (B) *Chat*AmDH, IGCAmDH1 and their mutants (C) MATOUAmDH2 and mutants, *Micro*AmDH and *Msme*AmDH (D) *Porti*AmDH, *Rgna*AmDH and their mutants. The imine formed in the extraction layer between the amine and the remaining aldehyde could be detected at 16.1 min. The legend above matches the color of the chromatogram with the enzyme/substrate couple according to Appendix 8. The legend below indicates: compound; retention time (min); area.

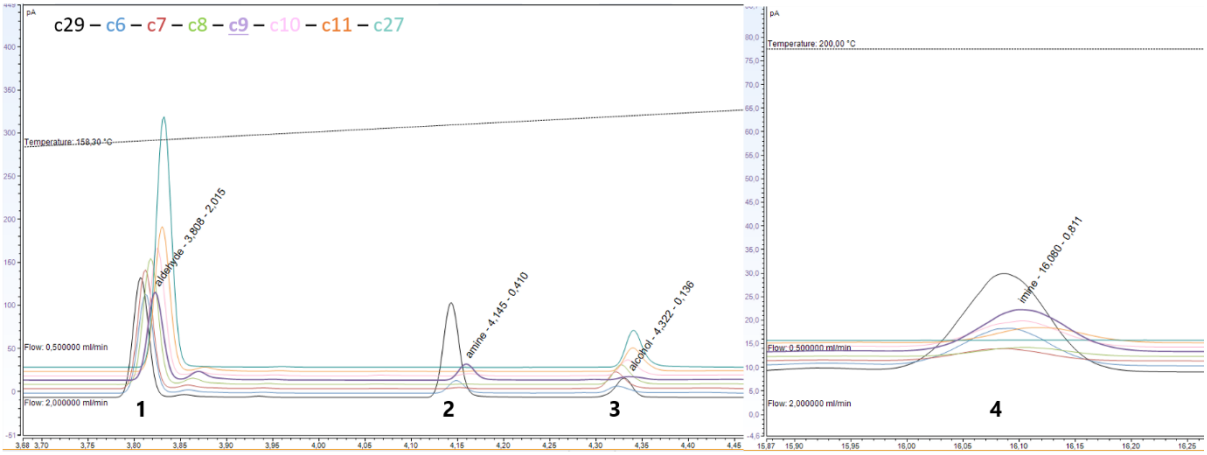
A



1: benzaldehyde; 3.807; 2.739 **2**: benzylamine; 4.143; 2.258

3: benzylalcohol; 4.330; 0.487 **4**: *N*-benzyl-1-benzylimine; 16.087; 1.933

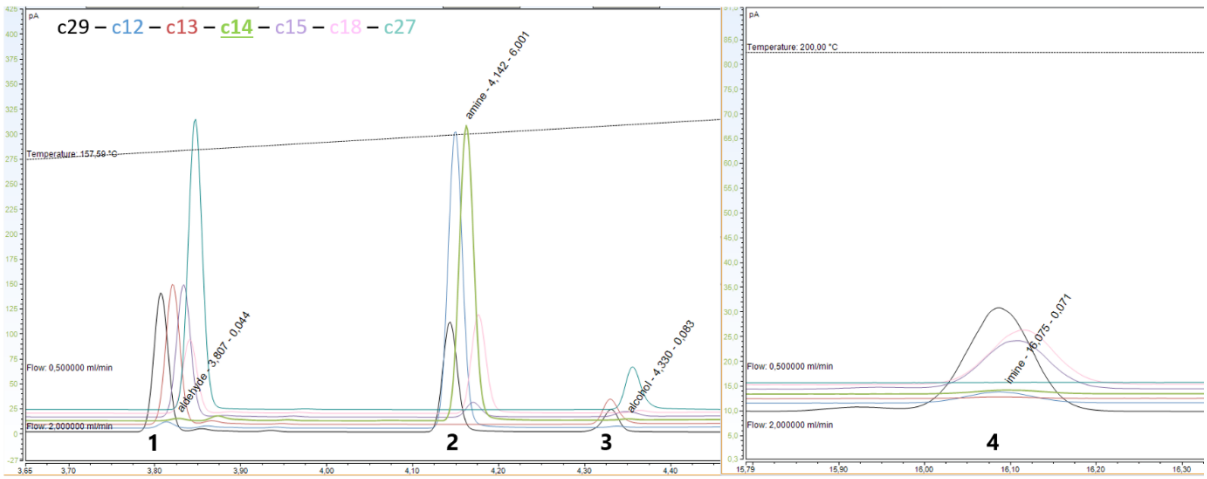
B



1: benzaldehyde; 3.808; 2.015 **2**: benzylamine; 4.145; 0.410

3: benzylalcohol; 4.322; 0.136 **4**: *N*-benzyl-1-benzylimine; 16.080; 0.811

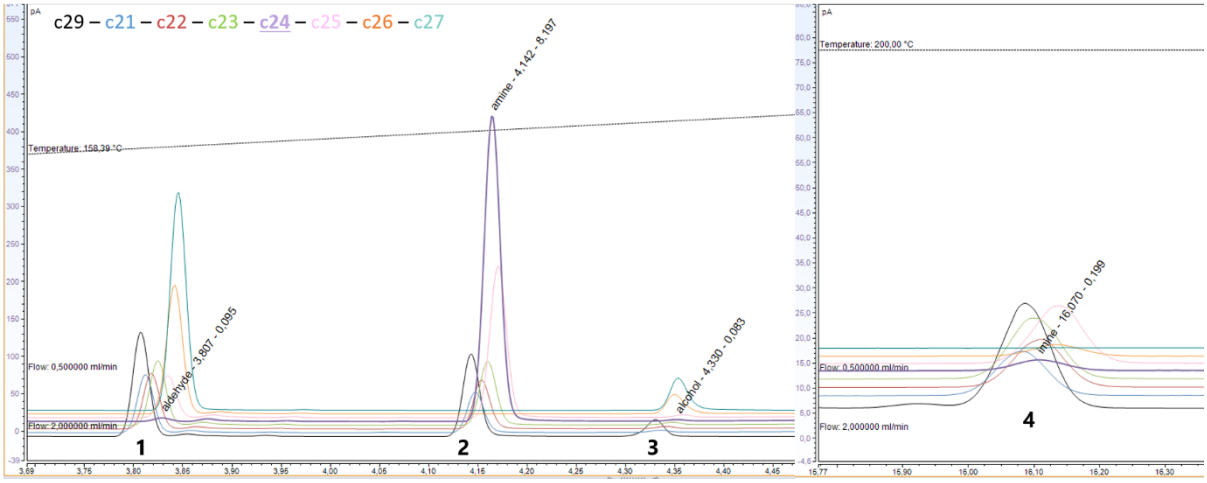
C



1: benzaldehyde; 3.807; 0.044 **2**: benzylamine; 4.142; 6.001

3: benzylalcohol; 4.330; 0.083 **4**: *N*-benzyl-1-benzylimine; 16.075; 0.071

D

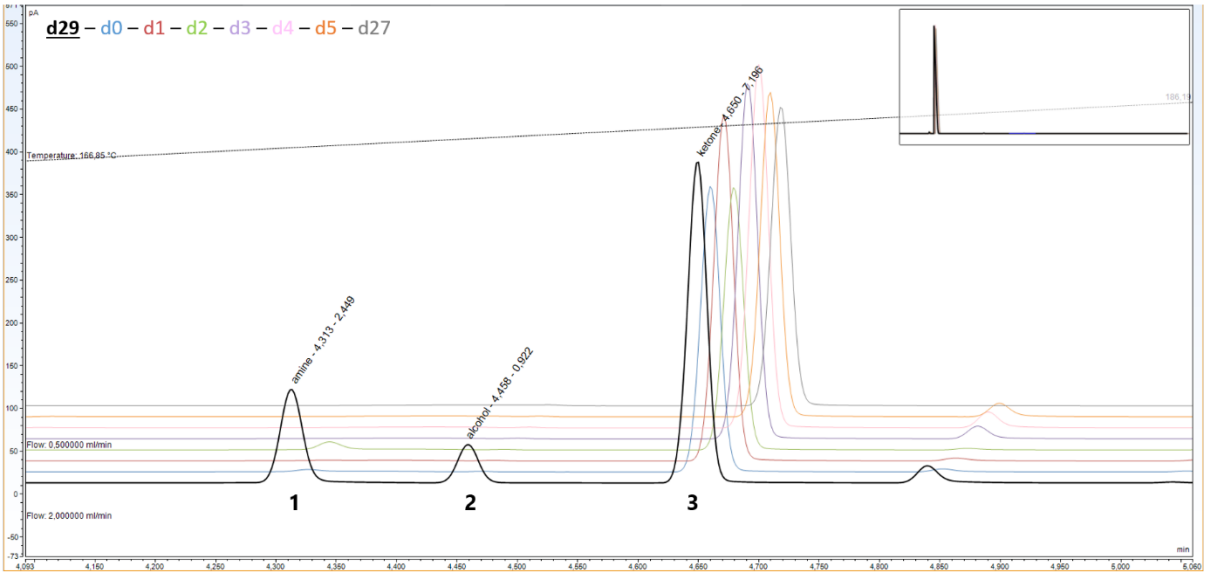


1: benzaldehyde; 3.807; 0.095 **2**: benzylamine; 4.142; 8.197

3: benzylalcohol; 4.330; 0.083 **4**: *N*-benzyl-1-benzylimine; 16.070; 0.199

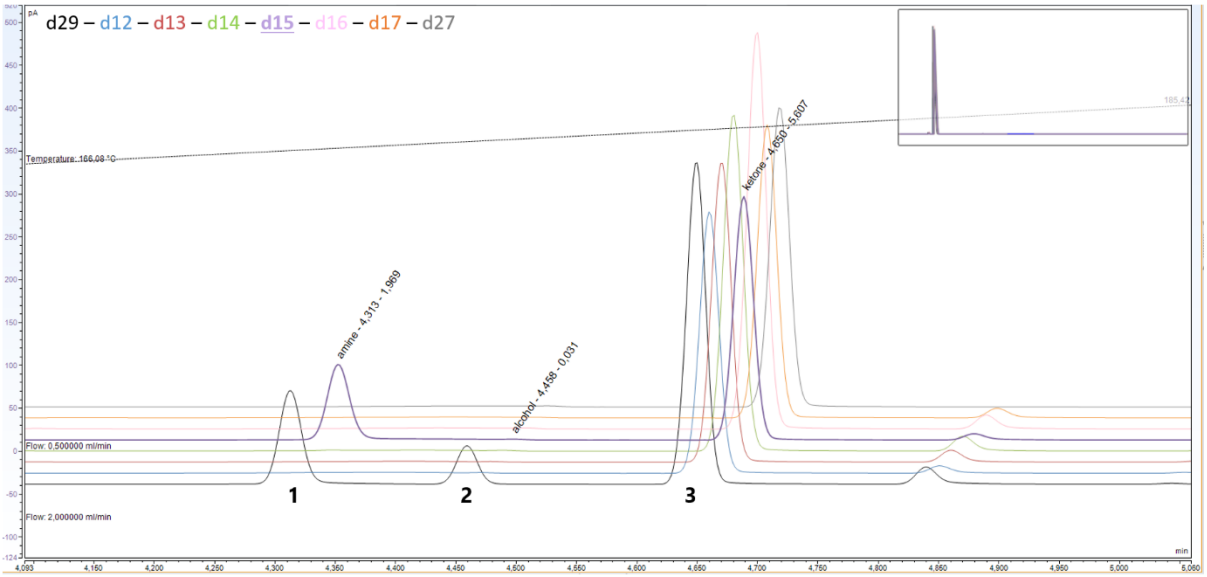
Appendix 12. (Chapter II, IV.2.1.1 and Chapter III, I.4.4) GC-FID chromatograms for reaction of acetophenone (**39**) with various nat-AmDHs and mutants. Chromatograms of a commercial standard of 1-phenylethylamine (**14**) (2 mM), **39** (7.2 mM), and corresponding alcohol (0.8 mM), blank reaction mixture without enzyme and reaction mixture with (A) *Cfus*AmDH, *Apau*AmDH and their mutants (B) MATOUAmDH2, *Micro*AmDH and their mutants (C) MsmeAmDH, *Porti*AmDH and their mutants. The legend above matches the color of the chromatogram with the enzyme/substrate couple according to Appendix 8. The legend below indicates: compound; retention time (min); area.

A



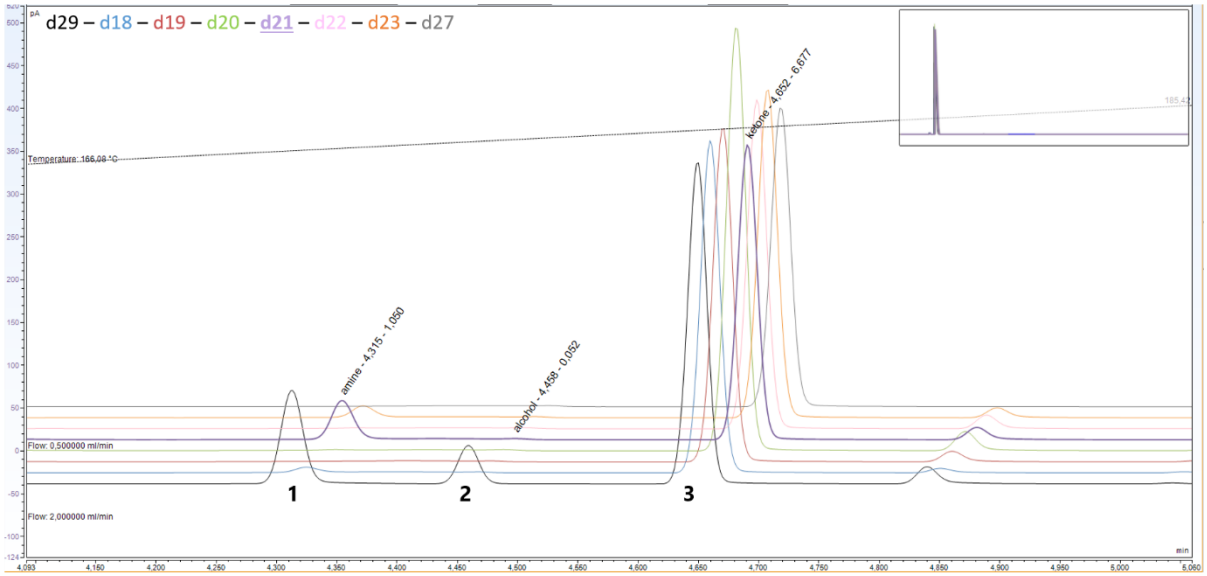
1: 1-phenylethylamine; 4.313; 2.449 **2**: 1-phenylethanol; 4.458; 0.922 **3**: acetophenone; 4.650; 7.196

B



1: 1-phenylethylamine; 4.313; 1.969 **2**: 1-phenylethanol; 4.458; 0.031 **3**: acetophenone; 4.650; 5,607

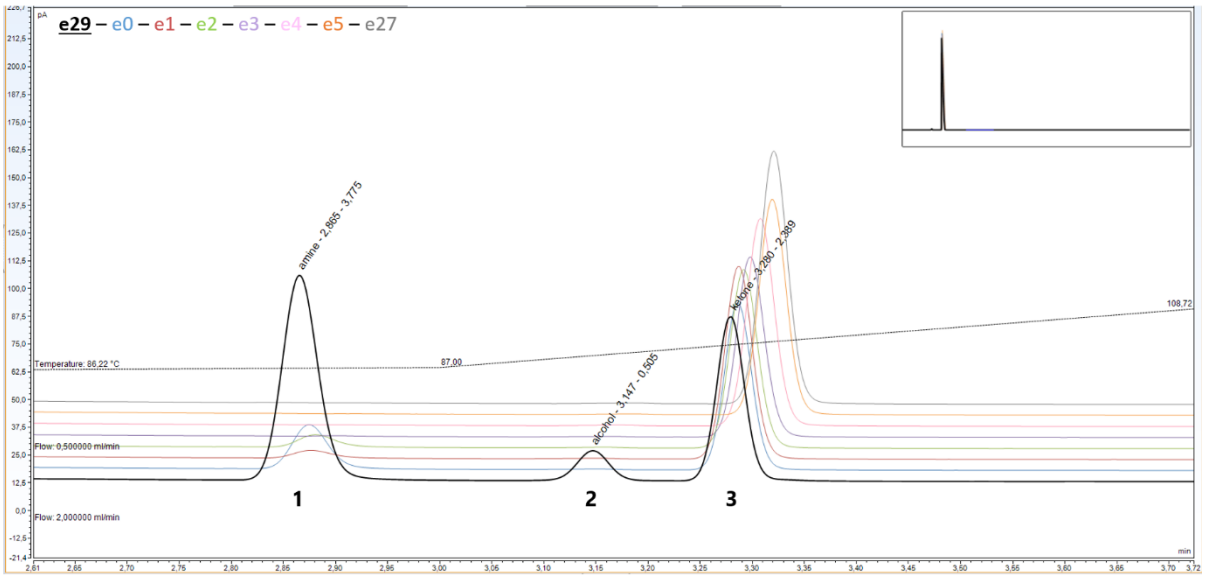
C



1: 1-phenylethylamine; 4.315; 1.050 **2**: 1-phenylethanol; 4.458; 0.052 **3**: acetophenone; 4.652; 6.677

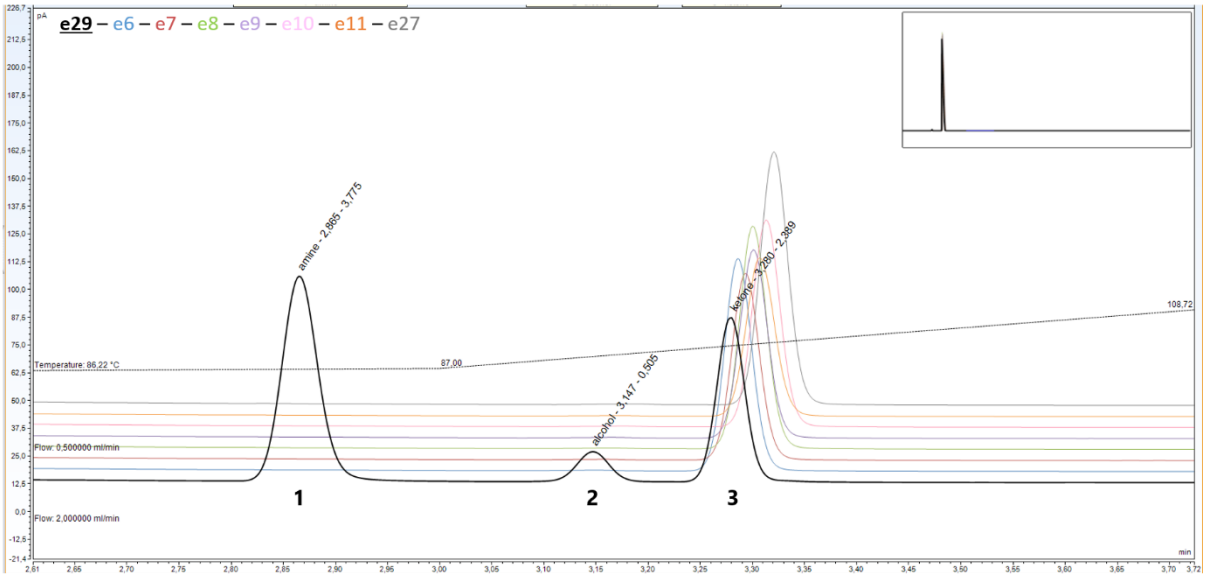
Appendix 13. (Chapter II, IV.2.1.1 and Chapter III, I.4.4) GC-FID chromatograms for reaction of hexan-3-one (**200**) with various nat-AmDHs and mutants. Chromatograms of a commercial standard of hexan-3-amine (**238**) (3 mM), **200** (6.4 mM), and corresponding alcohol (0.6 mM), blank reaction mixture without enzyme and reaction mixture with (A) *Cfus*AmDH, *Apau*AmDH and their mutants (B) *Chat*AmDH, IGCAmDH1 and their mutants (C) MATOUAmDH2, *Micro*AmDH and their mutants (D) *Msme*AmDH, *Porti*AmDH and their mutants and (E) *Rgna*AmDH and its mutants after extraction in EtOAc. The legend above matches the color of the chromatogram with the enzyme/substrate couple according to Appendix 8. The legend below indicates: compound; retention time (min); area.

A



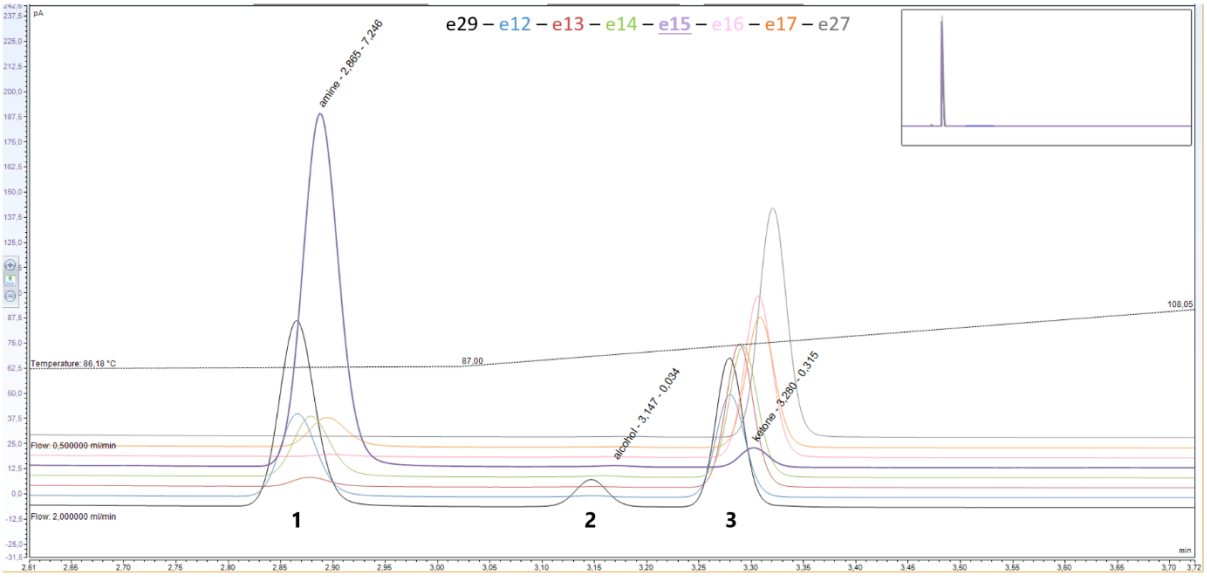
1: hexan-3-amine; 2.865; 3.775 **2**: hexan-3-ol; 3.147; 0.505 **3**: hexan-3-one; 3.280; 2.389

B



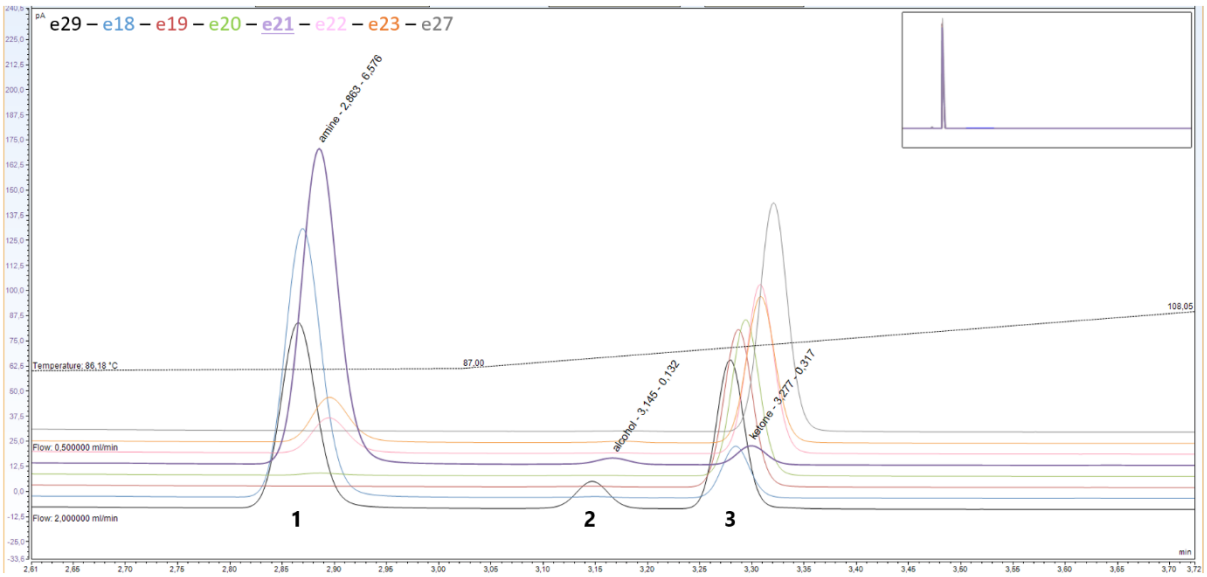
1: hexan-3-amine; 2.865; 3.775 **2**: hexan-3-ol; 3.147; 0.505 **3**: hexan-3-one; 3.280; 2.389

C



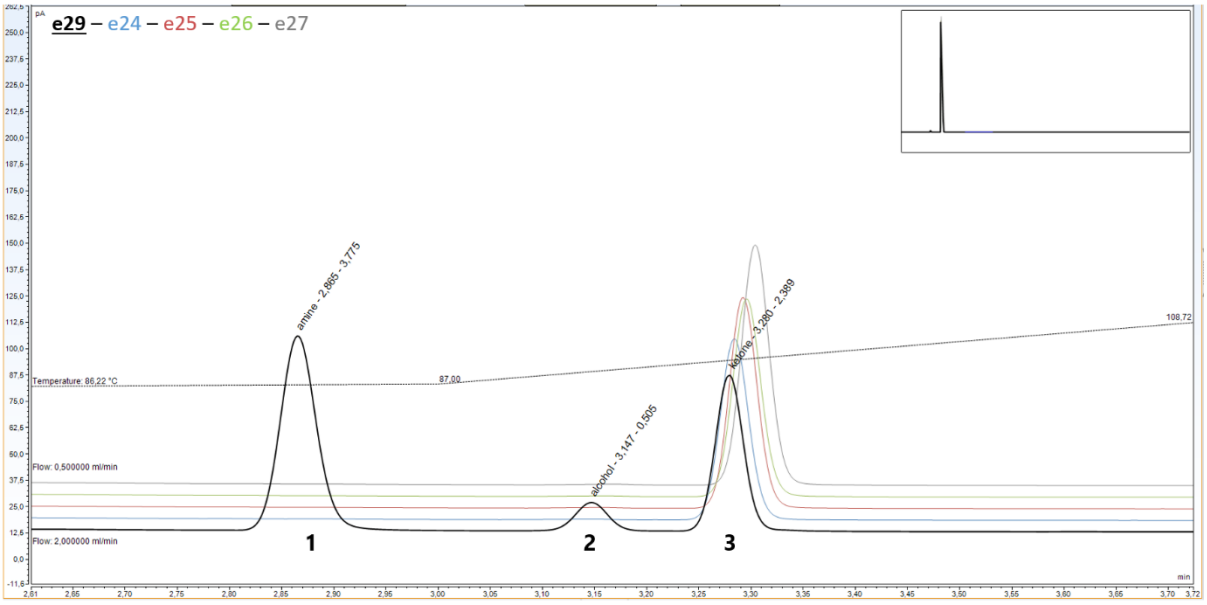
1: hexan-3-amine; 2.865; 7.246 **2**: hexan-3-ol; 3.147; 0.034 **3**: hexan-3-one; 3.280; 0.315

D



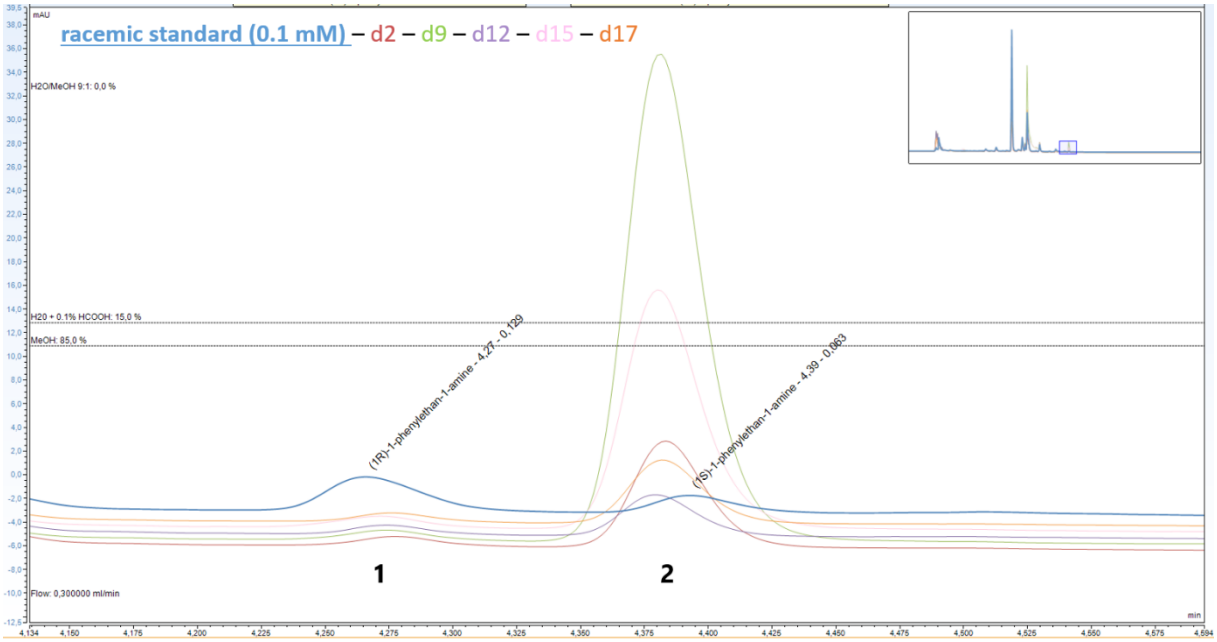
1: hexan-3-amine; 2.863; 6.576 **2**: hexan-3-ol; 3.145; 0.132 **3**: hexan-3-one; 3.277; 0.317

E



1: hexan-3-amine; 2.865; 3.775 **2**: hexan-3-ol; 3.147; 0.505 **3**: hexan-3-one; 3.280; 2.389

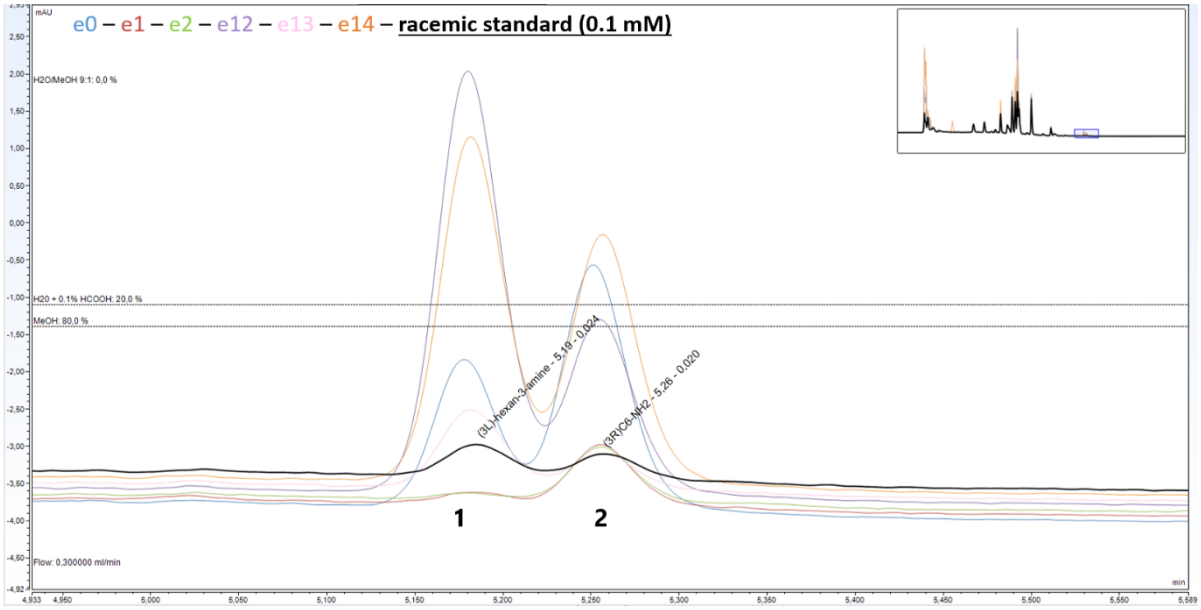
Appendix 14. (Chapter II, IV.2.1.1 and Chapter III, I.4.4) UHPLC-UV chromatograms for reaction of acetophenone (**39**) with various nat-AmDHs and mutants. UHPLC-UV chromatograms (340 nm) of a commercial standard of racemic amine (0.1 mM) and reaction mixture with *Cfus*AmDH-W145A, *Micro*AmDH, *Msme*AmDH, *Porti*AmDH and *Porti*AmDH-W140A after derivatization with FDAA. The legend above matches the color of the chromatogram with the enzyme/substrate couple according to Appendix 8. The legend below indicates: compound; retention time (min); area.



1: (1*R*)-1-phenylethan-1-amine; 4.27; 0.129 **2**: (1*S*)-1-phenylethan-1-amine; 4.39; 0.063

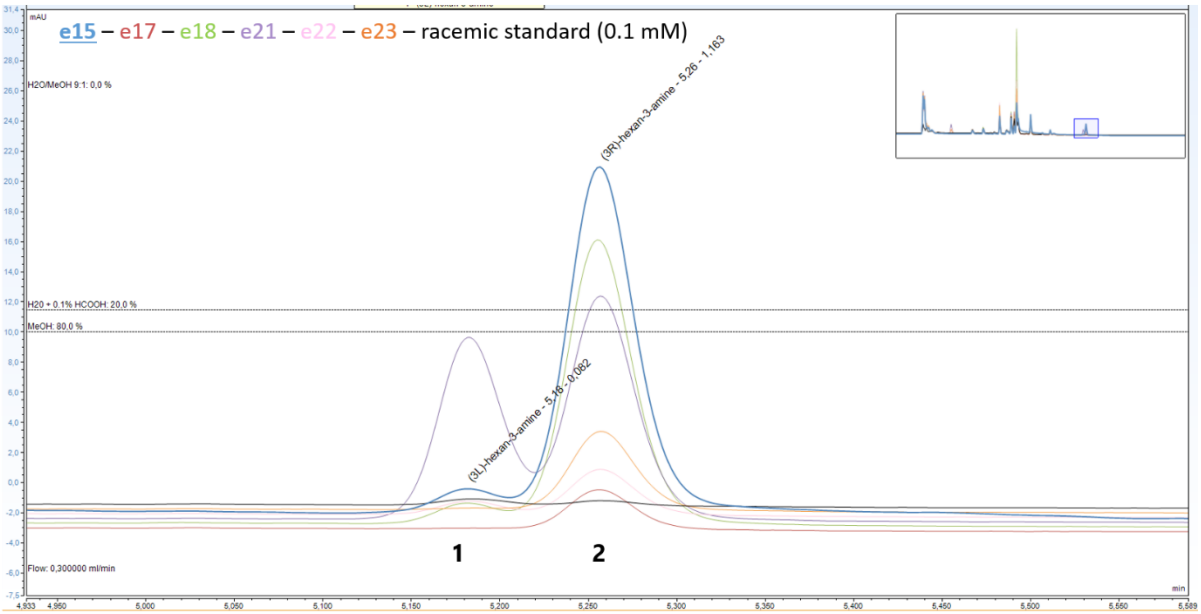
Appendix 15. (Chapter II IV.2.1.1 and Chapter III, I.4.4) UHPLC-UV chromatograms for reaction of hexan-3-one (**200**) with various nat-AmDHs. UHPLC-UV chromatograms (340 nm) of standard expected racemic amine (**238**) (0.1 mM) and reaction mixture with (A) *Cfus*AmDH, MATOUAmDH2 and mutants and (B) *Msme*AmDH, *Micro*AmDH, *Micro*AmDH-W141A and *Porti*AmDH and mutants after derivatization with FDAA. The legend above matches the color of the chromatogram with the enzyme/substrate couple according to Appendix 8. The legend below indicates: compound; retention time (min); area.

A



1: hexan-3-amine (left peak); 5.19; 0.024 **2**: hexan-3-amine (right peak); 5.26; 0.024

B

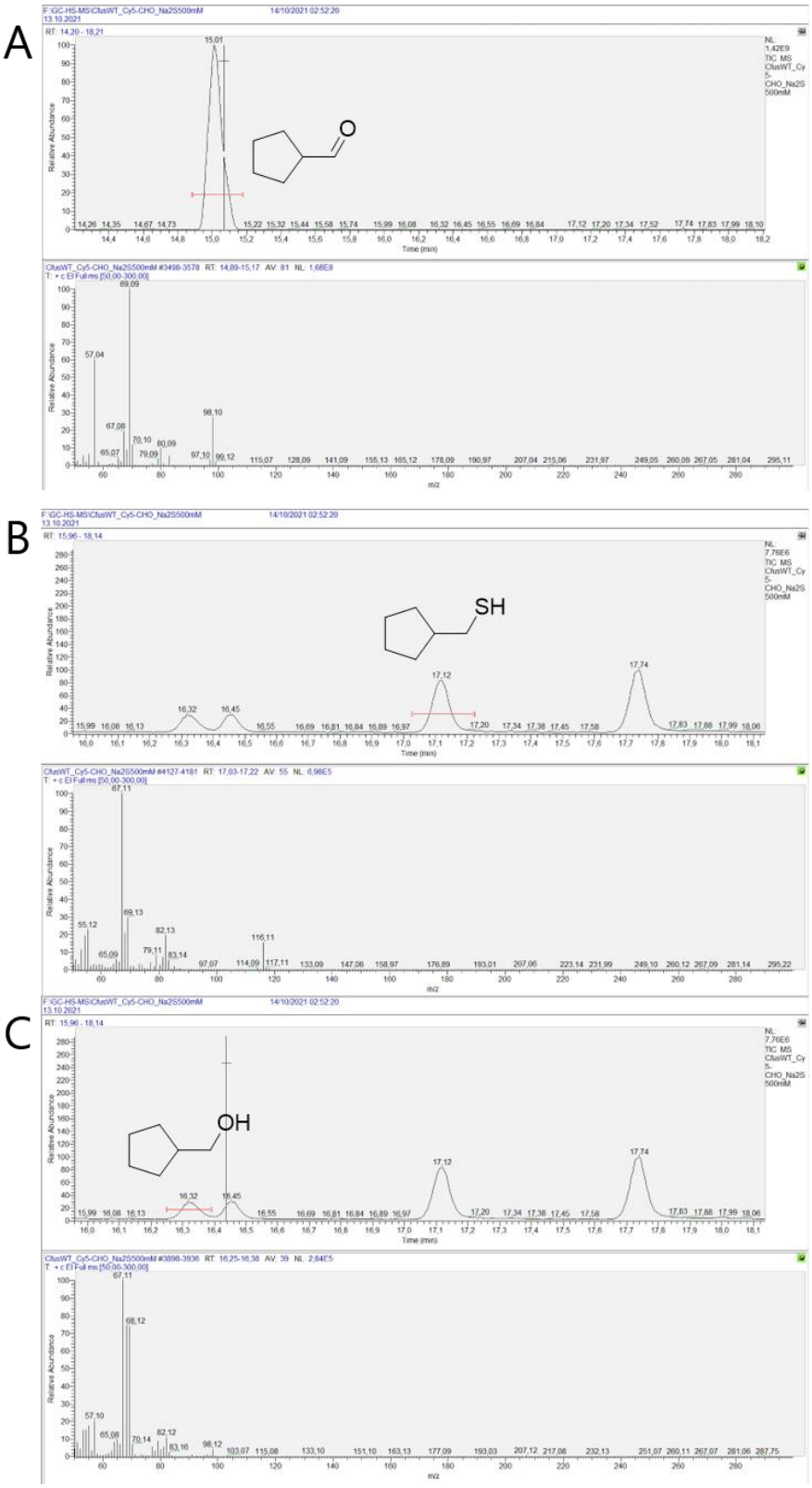


1: hexan-3-amine (left peak); 5.18; 0.082 **2**: hexan-3-amine (right peak); 5.26; 1.163

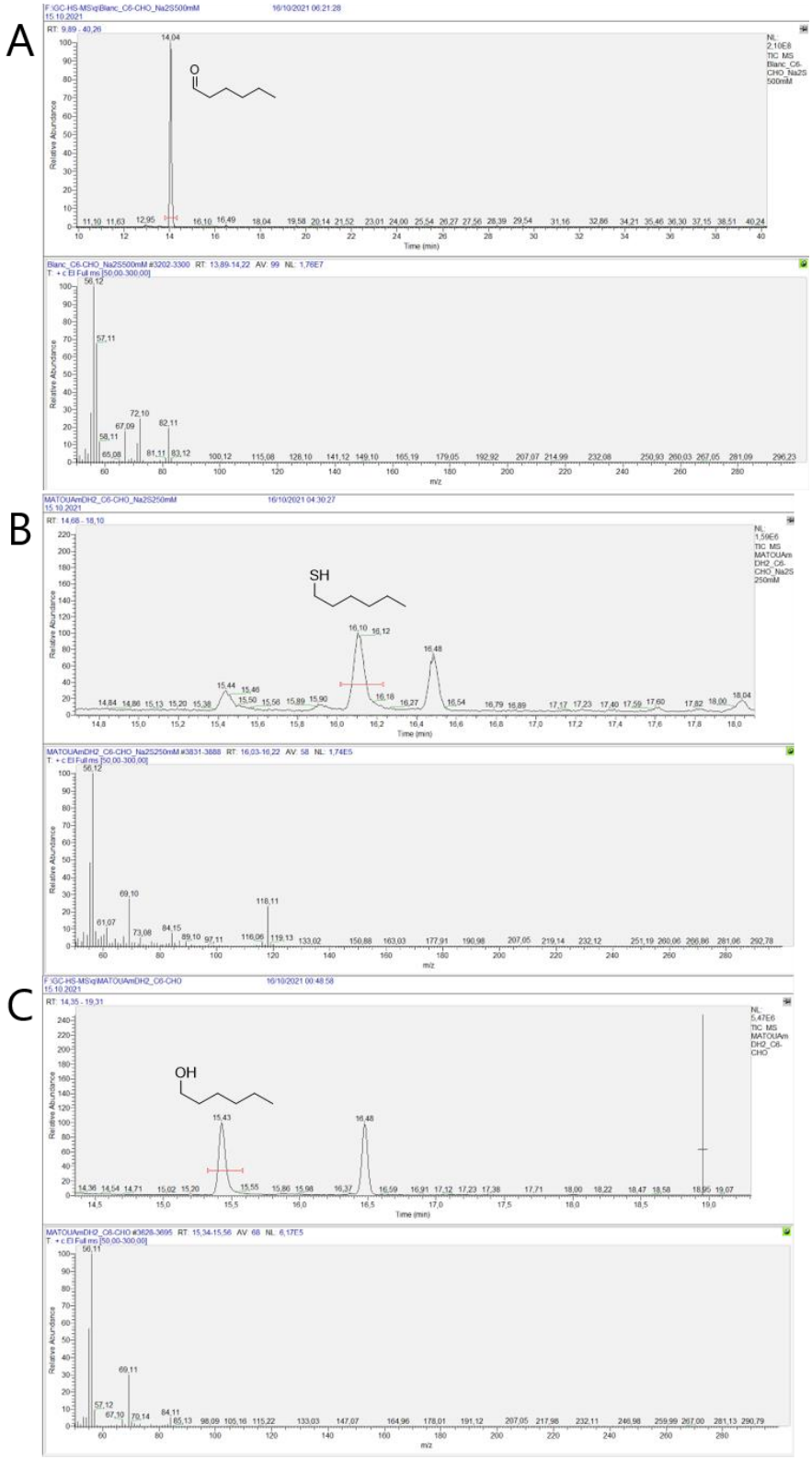
Appendix 16. (Chapter II, IV.2.1.2 and Chapter III, I.4.4) Results in alcohol formation for the screening of nat-AmDHs WT (described in Chapter II, IV.2.1.2) and mutants (described in Chapter III, I.4.4) towards furfural (**57**), cyclopentancarbaldehyde (**199**), benzaldehyde (**31**), acetophenone (**39**) and hexan-3-one (**200**). The cells highlighted in shades of green give the analytical conversion (%) in the corresponding alcohol. *n.t.*: not tested; *n.d.*: not detected.

	Chat AmDH			IGCAmDH1			*Rgna* AmDH			*Apau* AmDH			MATOUAmDH2			*Cfus* AmDH			*Msme* AmDH			*Porti* AmDH			*Micro* AmDH		
	WT	M161A	W166A	WT	M139A	W144A	WT	Y150A	W155A	WT	Y136A	W141A	WT	F143A	C148A	WT	F140A	W145A	WT	Y136A	W141A	WT	H135A	W140A	WT	F136A	W141A
57	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*
199	<1	2.8	1.5	<1	4.9	1.0	*n.d.*	*n.d.*	5.8	<1	4.5	6.6	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	1.7	*n.d.*	6.1	9.3	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	<1
31	1.2	2.4	2.9	<1	2.4	2.9	<1	<1	3.3	<1	<1	2.7	<1	3.1	<1	<1	<1	<1	<1	*n.t.*	*n.t.*	<1	<1	<1	1.1	*n.t.*	*n.t.*
39	*n.t.*	*n.t.*	*n.t.*	*n.t.*	*n.t.*	*n.t.*	*n.t.*	*n.t.*	*n.t.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*
200	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	1.2	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*

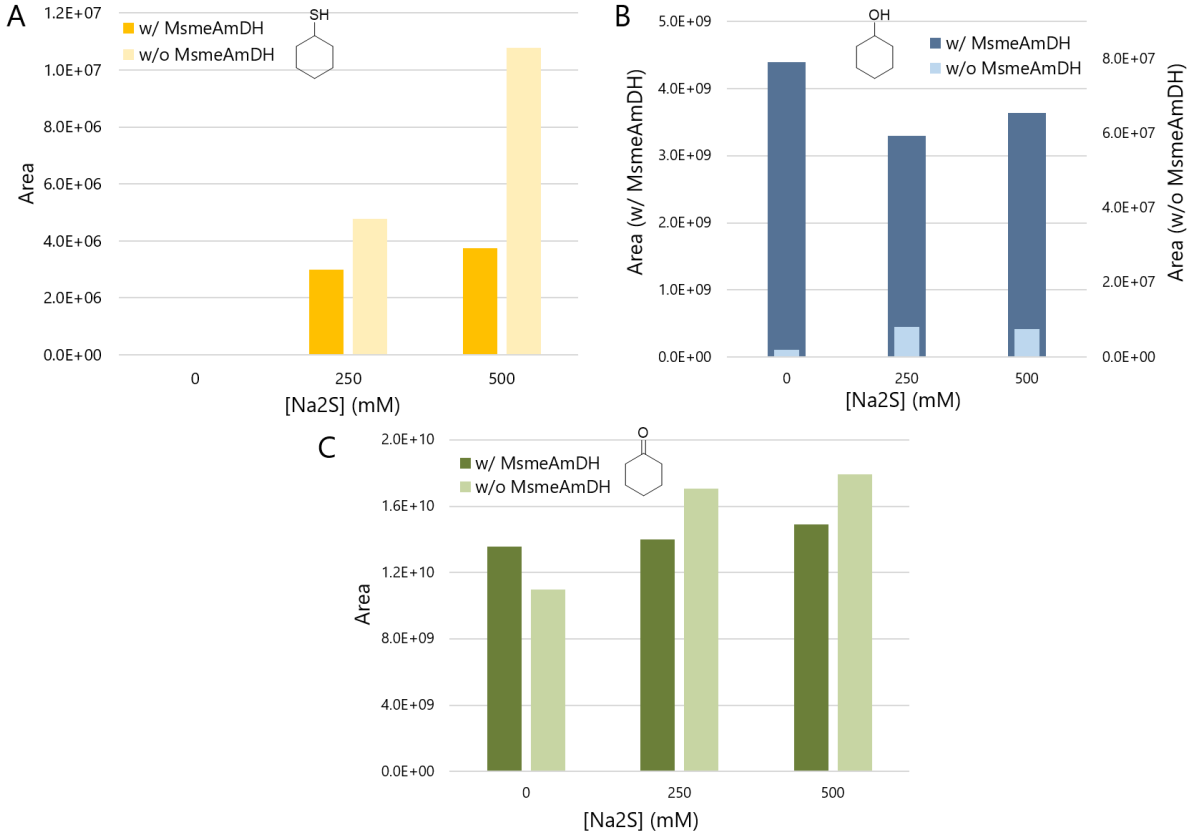
Appendix 17. (Chapter II, V) GC-HS-MS chromatograms of (A) cyclopentancarbaldehyde (**199**) (15.01 min), (B) cyclopentylmethanethiol (**205**) (17.12 min), (C) cyclopentylmethanol (**206**) (16.32 min) of reaction with *Cfus*AmDH in presence of 500 mM Na_2S.



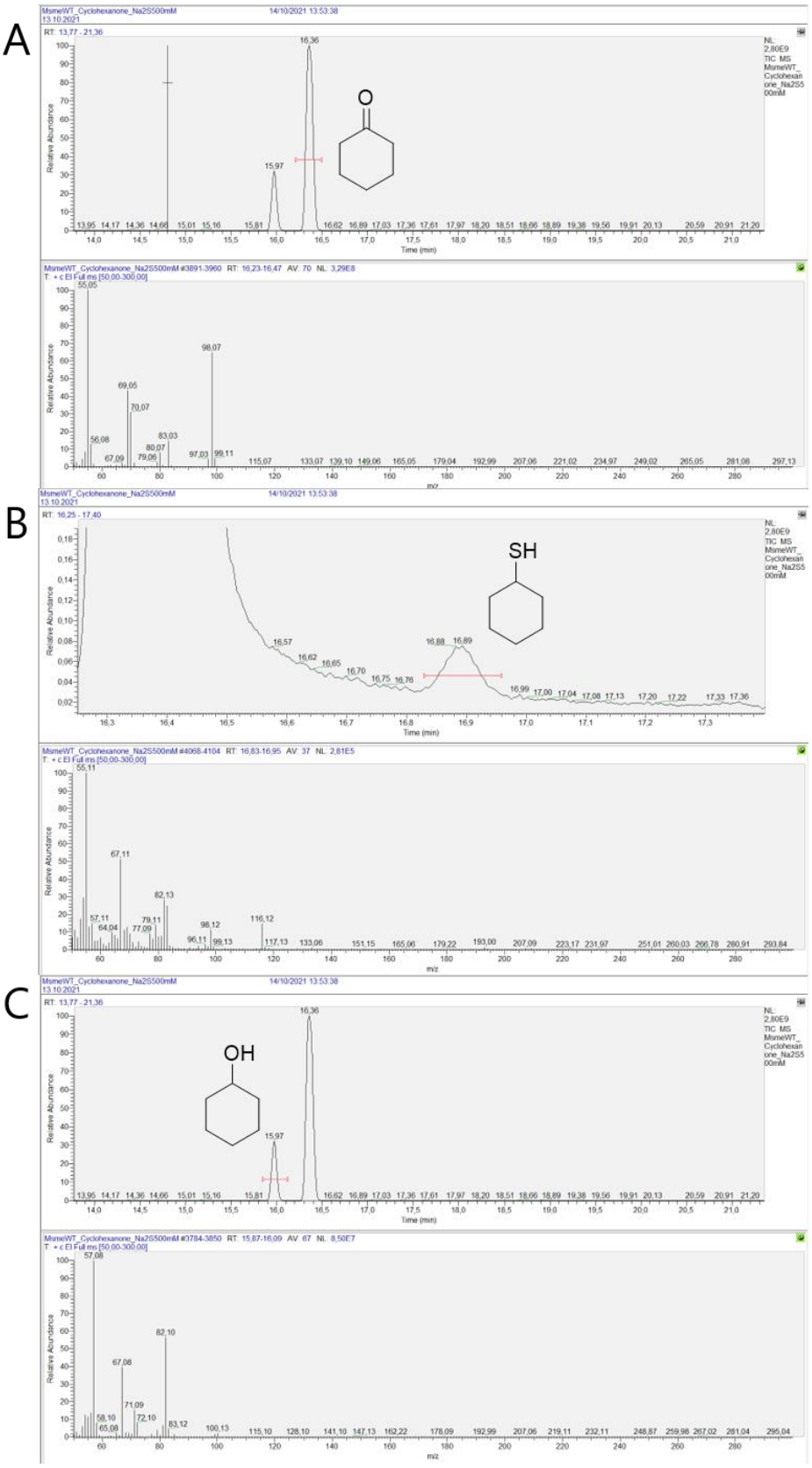
Appendix 18. (Chapter II, V) GC-HS-MS chromatograms of (A) hexanal (**207**) (14.04 min), (B) hexane-1-thiol (**208**) (16.10 min), (C) hexane-1-ol (**209**) (15.43 min) of reaction with MATOUAmDH2 in presence of 250 mM Na_2S.



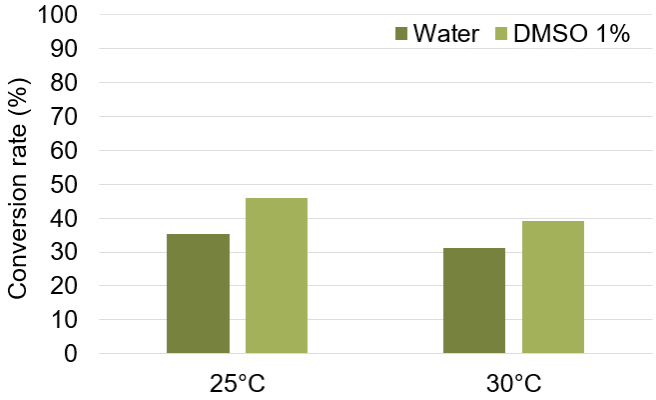
Appendix 19. (Chapter II, V) Thiol and alcohol formation from cyclohexanone (**45**) in presence of Na_2S and *Msme*AmDH. Area resulted from the transformation of **45** in presence or absence of *Msme*AmDH and 0, 250 and 500 mM Na_2S. (A) cyclohexanethiol formed, (B) cyclohexanol formed and (C) remaining aldehyde. The reaction conditions were as follows: final volume 100 µL, 10 mM **45**, 0-500 mM Na_2S, 100 mM potassium phosphate buffer pH 9, 5 mM NADPH, 5 mM NADH, 0.5 mg mL^{-1} purified *Msme*AmDH, 30°C, 24 h.



Appendix 20. (Chapter II, V) GC-HS-MS chromatograms of (A) **45** (16.36 min), (B) cyclohexanethiol (16.89 min), (C) cyclohexanol (15.97 min) of reaction with *Msme*AmDH in presence of 500 mM Na_2S.

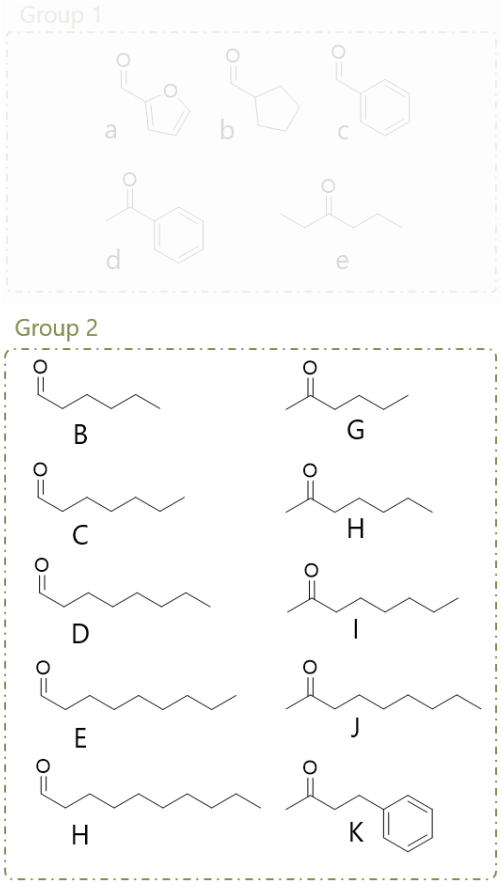


Appendix 21. (Chapter III, I.3) Effect of DMSO and temperature on conversion rates towards heptanal (**217**). Conversion rate obtained with *Cfus*AmDH-W145A towards **217** at 25°C and 30°C with or without 1% DMSO for substrate solubilization.



Appendix 22. (Chapter II, IV.2) Enzymes and substrates annotations for GC-FID and UHPLC-UV chromatograms from the screening of nat-AmDHs and mutants towards a range of carbonyl-containing compouds. The group of substrates named Group 1 that is indinstinct concern the rest of the screening discussed in Chapter II, IV.2. The chromatograms given in Appendix 9 to Appendix 15 correspond to the whole screening but the enzymes numbering and substrates annotation enable to focus on the results discussed here.

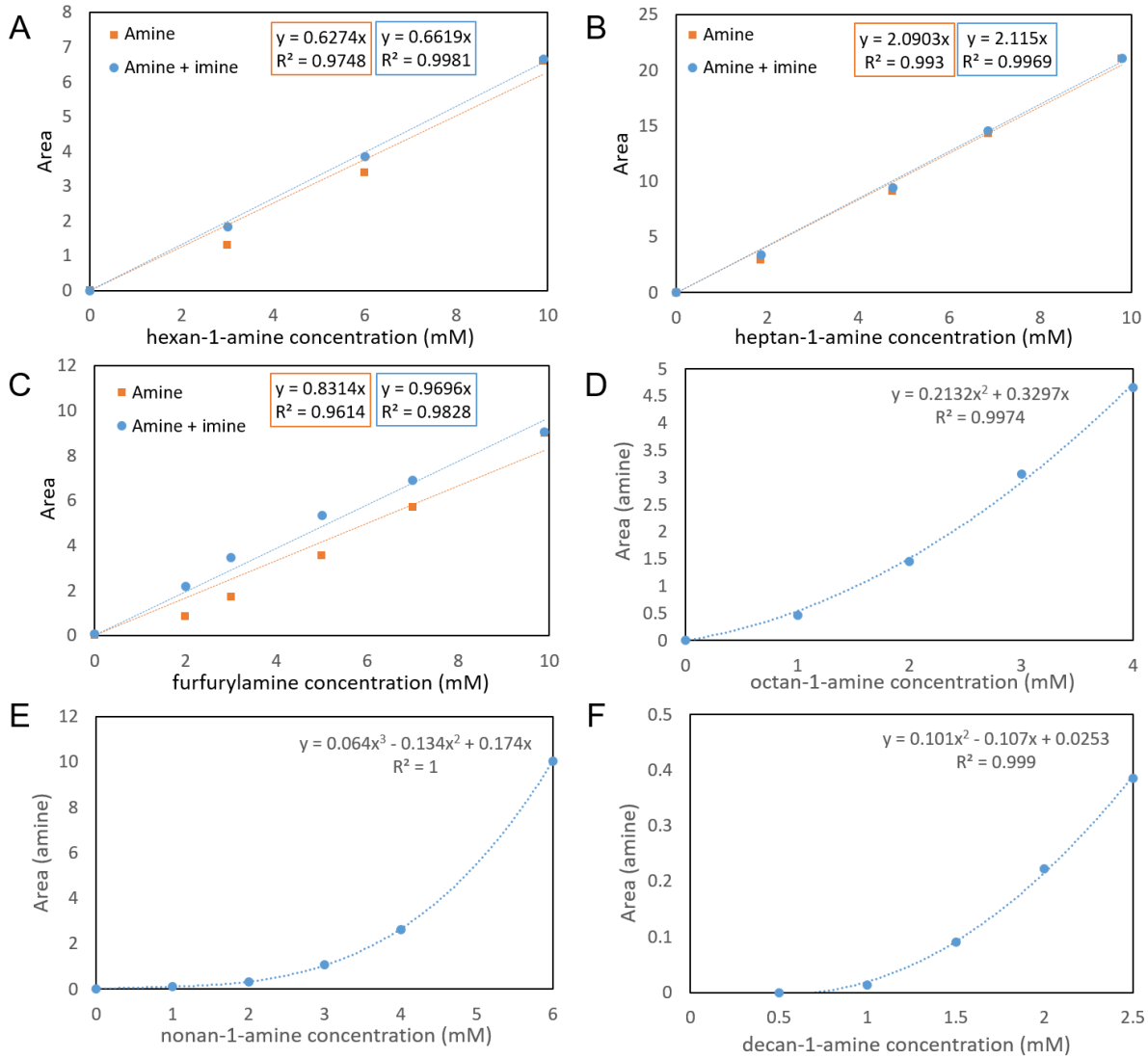
A



B

0	***Cfus* AmDH**
1	*Cfus* AmDH-F140A
2	*Cfus* AmDH-W145A
3	***Apau* AmDH**
4	*Apau* AmDH-Y136A
5	*Apau* AmDH-W141A
6	***Chat* AmDH**
7	*Chat* AmDH-M161A
8	*Chat* AmDH-W166A
9	**IGCAmDH1**
10	IGCAmDH1-M139A
11	IGCAmDH1-W144A
12	**MATOUAmDH2**
13	MATOUAmDH2-F143A
14	MATOUAmDH2-C148A
15	***Micro* AmDH**
16	*Micro* AmDH-F136A
17	*Micro* AmDH-W141A
18	***Msme* AmDH**
19	*Msme* AmDH-Y136A
20	*Msme* AmDH-W141A
21	***Porti* AmDH**
22	*Porti* AmDH-H135A
23	*Porti* AmDH-W140A
24	***Rgna* AmDH**
25	*Rgna* AmDH-Y150A
26	*Rgna* AmDH-W155A
27	without enzyme
28-31	commercial product

Appendix 23. (Chapter III, I.4.2) Calibration curves from the screening with (A) hexanal (**207**), (B) heptanal (**217**), (C) furfural (**57**), (D) octanal (**218**), (E) nonanal (**219**) and (F) decanal (**220**). For the first three substrates (A), (B) and (C), two curves are given one considering solely the amine peak (orange) and the other one with addition of the area of the peaks of amine and imine (blue). For the last three compounds (D), (E) and (F), the imine peak was not detected and the curves obtained considering solely the amine peak were defined following polynomial models.



Appendix 24. (Chapter III, I.4.3) Results in amine formation for the screening of nat-AmDHs WT and mutants towards hexanal (**207**), heptanal (**217**), octanal (**218**), nonanal (**219**), decanal (**220**), hexan-2-one (**43**), heptan-2-one (**128**), octan-2-one (**67**), nonan-2-one (**135**) and 4-phenylbutan-2-one (**36**). The cells highlighted in shades of green gives the analytical conversion (%) obtained for the synthesis of the corresponding amines and the cells colored in shades of yellow gives the corresponding *ee* in the case of the formation of chiral amine (%). If not indicated, the *ee* is given for the (*S*)-amine. conv.: analytical conversion ; *n.d.*: not detected; *n.t.*: not tested

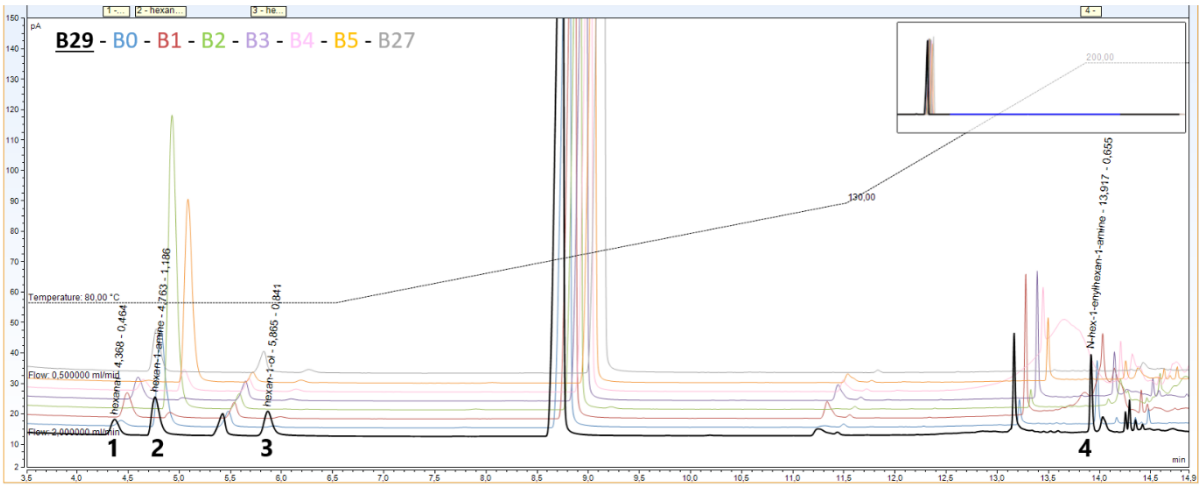
		Chat AmDH			IGCAmDH1			*Rgna* AmDH			MATOUAmDH2			*Cfus* AmDH			*Apau* AmDH			*Porti* AmDH			*Micro* AmDH			*Msme* AmDH		
		WT	M161A	W166A	WT	M139A	W144A	WT	Y150A	W155A	WT	F143A	C148A	WT	F140A	W145A	WT	Y136A	W141A	WT	H135A	W140A	WT	F136A	W141A	WT	Y136A	W141A
207	conv. (%)	4.6	<1	31.6	2.4	1.7	29.7	12.5	95.2	85.6	72.7	6.9	79.2	39.9	9.1	99.5	5.3	11.3	59.6	60.8	15.1	60.0	65.1	<1	2.0	54.9	<1	3.0
217	conv. (%)	<1	<1	34.6	<1	<1	26.9	<1	35.6	36.6	26.9	21.4	32.5	2.3	15.3	40.4	<1	23.2	34.4	7.9	3.3	23.8	18.1	<1	<1	6.0	<1	2.0
218	conv. (%)	*n.d.*	*n.d.*	31.9	*n.d.*	*n.d.*	32.3	*n.d.*	40.0	39.1	13.8	37.3	21.5	*n.d.*	33.2	38.8	*n.d.*	31.6	34.9	*n.d.*	3.3	25.0	3.1	*n.d.*	5.2	*n.d.*	*n.d.*	6.5
219	conv. (%)	*n.d.*	<1	26.8	*n.d.*	*n.d.*	26.2	*n.d.*	44.6	33.1	20.5	35.8	21.5	*n.d.*	40.2	35.3	*n.d.*	34.2	28.2	*n.d.*	1.8	22.0	*n.d.*	*n.d.*	6.7	*n.d.*	*n.d.*	<1
220	conv. (%)	*n.d.*	*n.d.*	1.1	*n.d.*	<1	*n.d.*	*n.d.*	11.5	*n.d.*	*n.d.*	8.9	*n.d.*	*n.d.*	8.8	4.1	<1	7.8	<1	*n.d.*	<1	<1	*n.d.*	*n.d.*	1.1	*n.d.*	*n.d.*	<1
43	conv. (%)	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	<1	<1	<1	2.0	2.9	3.0	3.1	10.6	39.7	*n.d.*	<1	4.0	26.5	2.0	18.2	22.6	<1	1.2	8.4	*n.d.*	<1
	ee (%)										77.5	89.1	91.8	94.7	99.5	99.3			99.2	79.4	90.2	99.5	94.6		>99.9	98.8		
128	conv. (%)	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	6.2	<1	<1	11.0	53.7	<1	<1	18.5	4.2	2.6	26.6	1.5	<1	<1	<1	<1	4.6
	ee (%)											95.2		57.9	99.3	99.7			99.3	33.0	96.7	99.6						>99.9
67	conv. (%)	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	4.6	3.8	*n.d.*	9.8	40.3	*n.d.*	*n.d.*	9.7	3.4	2.4	19.9	*n.d.*	*n.d.*	3.2	*n.d.*	*n.d.*	6.7
	ee (%)											95.1	>99.9		98.5	99.5			99.2	49.3 (*R*)	92.8	99.0			>99.9			>99.9
135	conv. (%)	<1	1.0	1.3	2.8	1.3	1.7	<1	3.5	<1	<1	2.5	1.3	<1	11.4	16.0	<1	2.9	4.6	<1	1.8	13.0	<1	<1	2.5	<1	<1	3.2
	ee (%)											>99.9	>99.9		>99.9	>99.9		>99.9	>99.9		>99.9	>99.9			>99.9			>99.9
36	conv. (%)	*n.t.*	*n.t.*	*n.t.*	*n.t.*	*n.t.*	*n.t.*	*n.t.*	*n.t.*	*n.t.*	<1	1.2	<1	<1	17.1	52.5	<1	5.8	22.2	2.4	<1	23.5	<1	<1	28.1	<1	<1	34.1
	ee (%)											95.2			>99.9	99.6		>99.9	99.2	49.6 (*R*)	84.7	99.6			95.5			99.9

Appendix 25. (Chapter III, I.4.3) Results in alcohol formation for the screening of nat-AmDHs WT and mutants towards hexanal (**207**), heptanal (**217**), octanal (**218**), nonanal (**219**), decanal (**220**), hexan-2-one (**43**), heptan-2-one (**128**), octan-2-one (**67**), nonan-2-one (**135**) and 4-phenylbutan-2-one (**36**). The cells highlighted in shades of green give the conversion rate (%) in corresponding alcohol. *n.d.*: not detected; *n.t.*: not tested

	Chat AmDH			IGCAmDH1			*Rgna* AmDH			MATOUAmDH2			*Cfus* AmDH			*Apau* AmDH			*Porti* AmDH			*Micro* AmDH			*Msme* AmDH		
	WT	M161A	W166A	WT	M139A	W144A	WT	Y150A	W155A	WT	F143A	C148A	WT	F140A	W145A	WT	Y136A	W141A	WT	H135A	W140A	WT	F136A	W141A	WT	Y136A	W141A
207	*n.d.*	<1	<1	*n.d.*	<1	1.8	<1	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	<1	*n.d.*	*n.d.*	*n.d.*	<1	*n.d.*	*n.d.*	*n.d.*
217	*n.d.*	<1	*n.d.*	*n.d.*	<1	<1	<1	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	<1	<1	*n.d.*	*n.d.*	<1	<1	*n.d.*	*n.d.*	*n.d.*
218	<1	2.2	<1	<1	2.1	4.0	3.2	<1	<1	<1	*n.d.*	1.2	<1	*n.d.*	*n.d.*	<1	*n.d.*	<1	3.4	2.6	2.2	<1	2.3	2.8	<1	1.1	1.4
219	<1	<1	<1	<1	<1	1.3	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1
220	*n.d.*	<1	<1	*n.d.*	<1	1.1	*n.d.*	1.7	*n.d.*	*n.d.*	<1	*n.d.*	*n.d.*	1.1	*n.d.*	*n.d.*	<1	<1	*n.d.*	<1	<1	*n.d.*	<1	<1	*n.d.*	*n.d.*	*n.d.*
43	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*
128	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	<1	*n.d.*	*n.d.*	*n.d.*	<1	*n.d.*	<1	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*
67	*n.d.*	*n.d.*	<1	*n.d.*	<1	*n.d.*	*n.d.*	*n.d.*	<1	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	<1	*n.d.*	*n.d.*	<1	<1	*n.d.*	<1	*n.d.*	*n.d.*	*n.d.*	<1	*n.d.*	*n.d.*
135	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*
36	*n.t.*	*n.t.*	*n.t.*	*n.t.*	*n.t.*	*n.t.*	*n.t.*	*n.t.*	*n.t.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*	*n.d.*

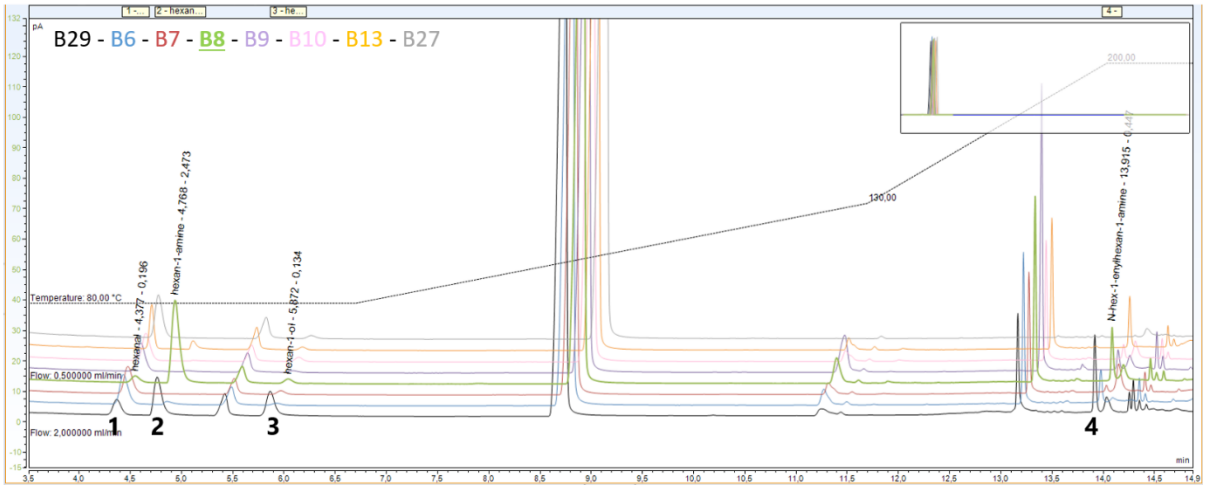
Appendix 26. (Chapter III, I.4.3) GC-FID chromatograms for reaction of hexanal (**207**) with various nat-AmDHs and mutants. Chromatograms of a commercial standard of hexan-1-amine (**212**) (3 mM), **207** (6.4 mM), and hexan-1-ol (**209**) (0.6 mM), blank reaction mixture without enzyme and reaction mixture with (A) *Cfus*AmDH, *Apau*AmDH and their mutants (B) *Chat*AmDH, IGCAmDH1 and their mutants (C) MATOUAmDH2, *Micro*AmDH and their mutants (D) *Msme*AmDH, *Porti*AmDH and their mutants and (E) *Rgna*AmDH and its mutants after extraction in EtOAc. The imine formed in the extraction layer between the amine and the remaining aldehyde could be detected at 13.9 min. The legend above matches the color of the chromatogram with the enzyme/substrate couple according to Appendix 22. The legend below indicates: compound; retention time (min); area.

A



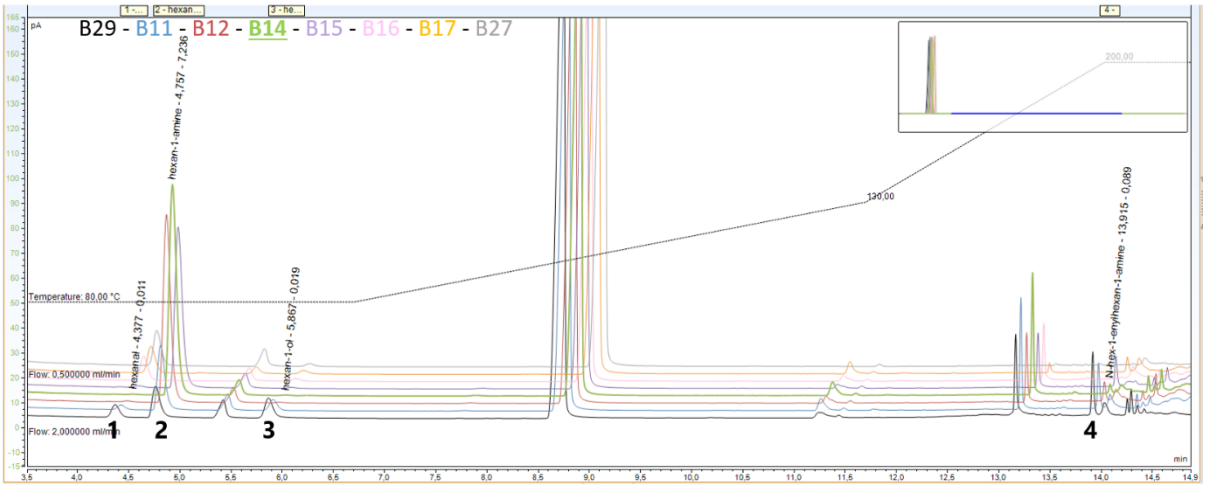
1: hexanal; 4.368; 0.464
2: hexan-1-amine; 4.763; 1.186
3: hexan-1-ol; 5.865; 0.841
4: *N*-hexyl-1-hexanimine; 13.917; 0.655

B



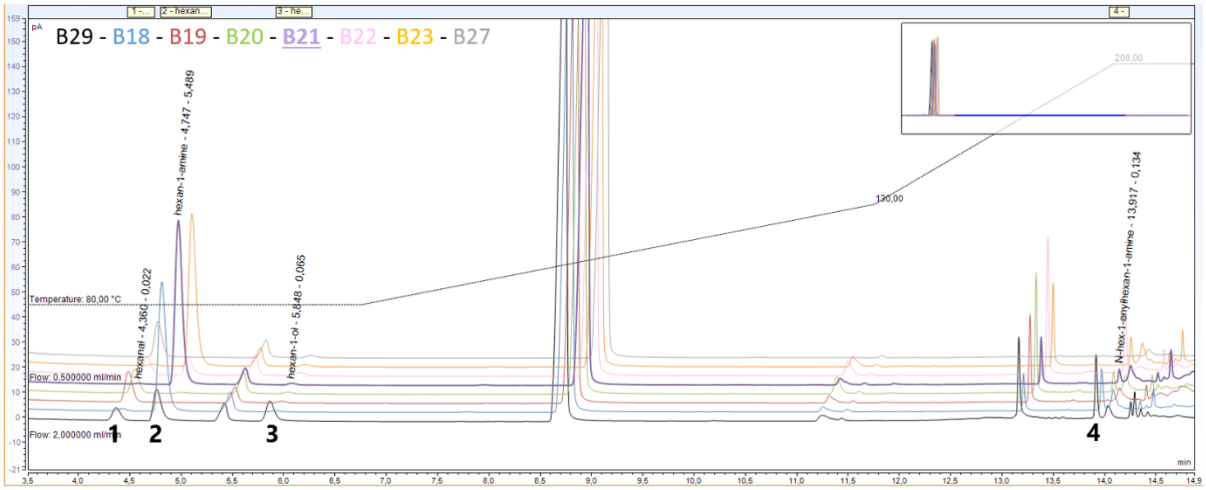
1: hexanal; 4.377; 0.196
2: hexan-1-amine; 4.768; 2.473
3: hexan-1-ol; 5.872; 0.134
4: *N*-hexyl-1-hexanimine; 13.915; 0.447

C



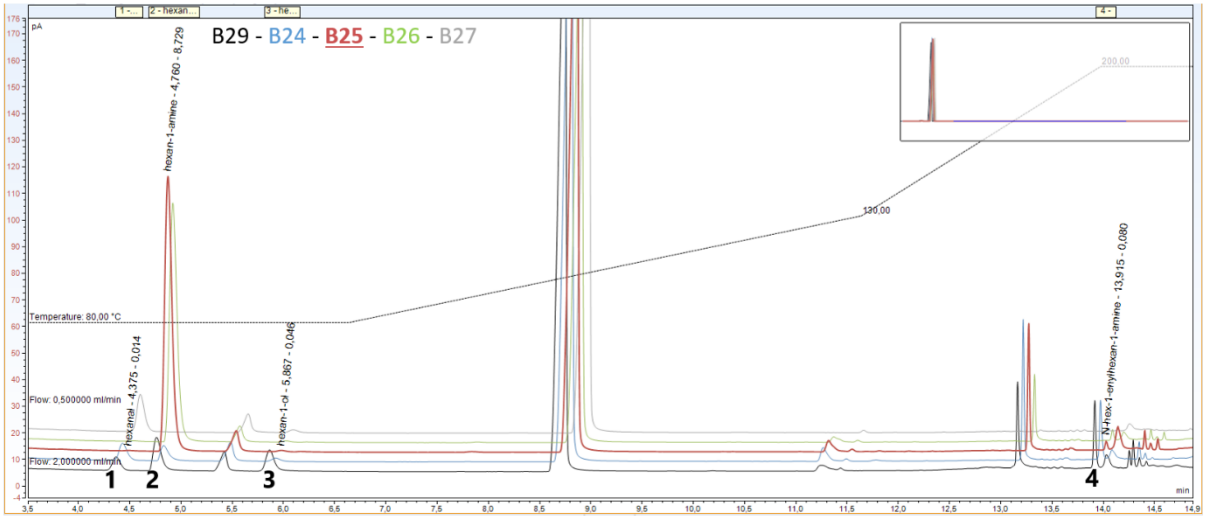
1: hexanal; 4.377; 0.011
2: hexan-1-amine; 4.757; 7.236
3: hexan-1-ol; 5.867; 0.019
4: *N*-hexyl-1-hexanimine; 13.915; 0.089

D



1: hexanal; 4.360; 0.022
2: hexan-1-amine; 4.747; 5.489
3: hexan-1-ol; 5.848; 0.065
4: *N*-hexyl-1-hexanimine; 13.917; 0.134

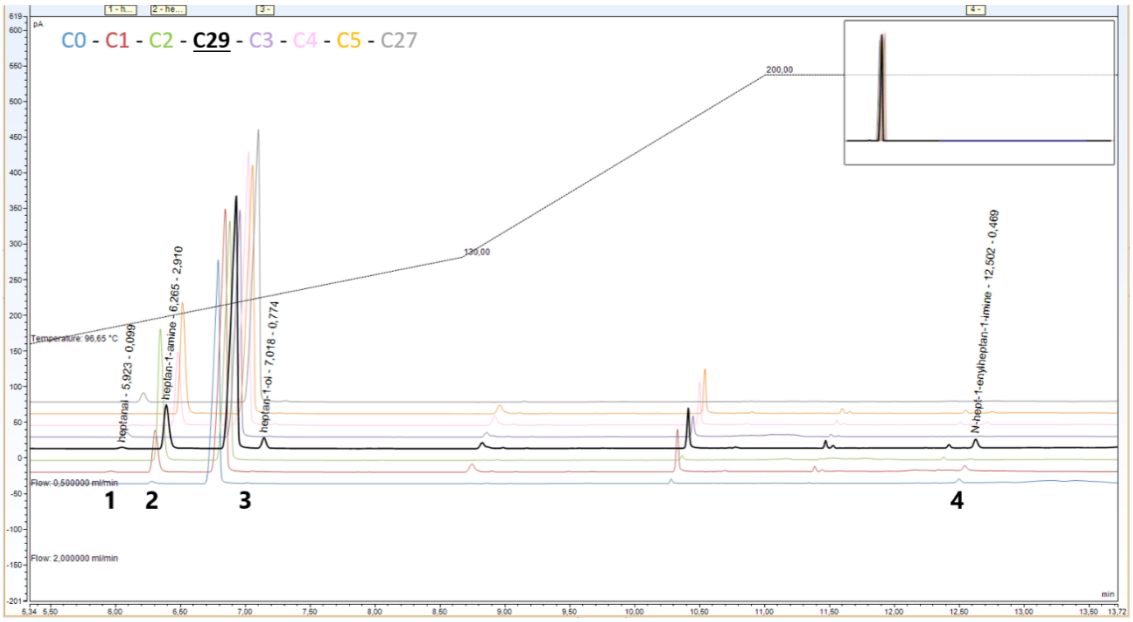
E



1: hexanal; 4.375; 0.014
2: hexan-1-amine; 4.760; 8.729
3: hexan-1-ol; 5.867; 0.046
4: *N*-hexyl-1-hexanimine; 13.915; 0.080

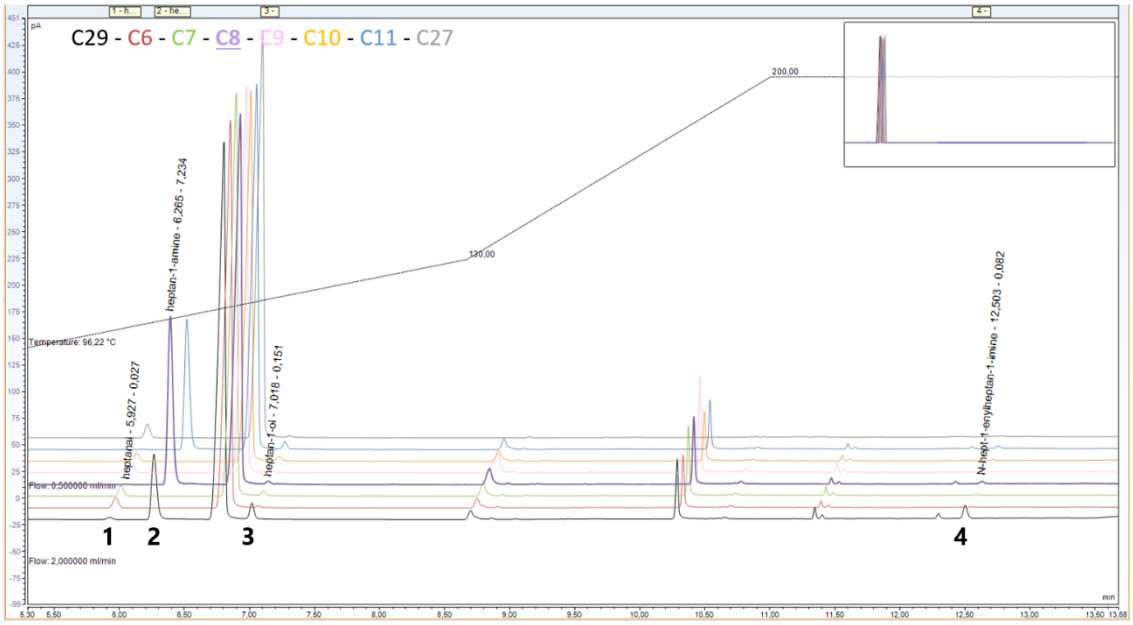
Appendix 27. (Chapter III, I.4.3) GC-FID chromatograms for reaction of heptanal (**217**) with various nat-AmDHs and mutants. Chromatograms of a commercial standard of heptan-1-amine (**213**) (3 mM), **217** (6.4 mM), corresponding alcohol (0.6 mM), blank reaction mixture without enzyme and reaction mixture with (A) *Cfus*AmDH, *Apau*AmDH and their mutants (B) *Chat*AmDH, IGCAmDH1 and their mutants (C) MATOUAmDH2, *Micro*AmDH and their mutants (D) *Msme*AmDH, *Porti*AmDH and their mutants and (E) *Rgna*AmDH and its mutants after extraction in EtOAc. The imine formed in the extraction layer between the amine and the remaining aldehyde could be detected at 12.5 min. The legend above matches the color of the chromatogram with the enzyme/substrate couple according to Appendix 22. The legend below indicates: compound; retention time (min); area.

A



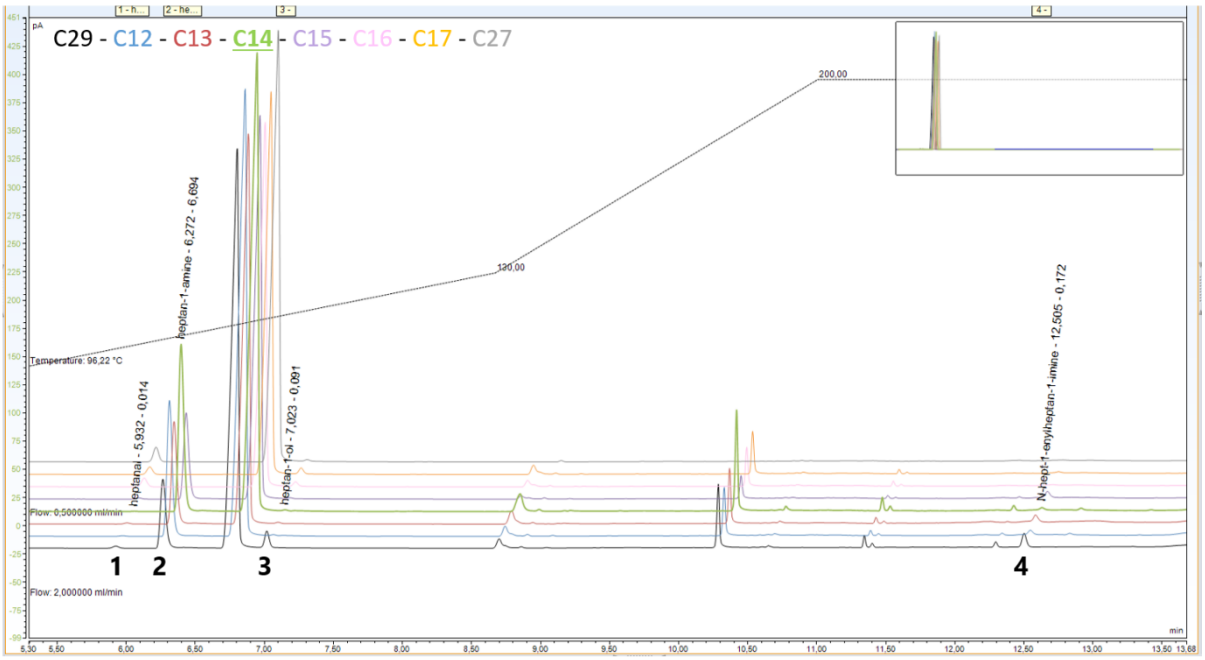
1: heptanal; 5.923; 0.099
2: heptan-1-amine; 6.265; 2.910
3: heptan-1-ol; 7.018; 0.774
4: *N*-heptyl-1-heptanimine; 12.502; 0.469

B



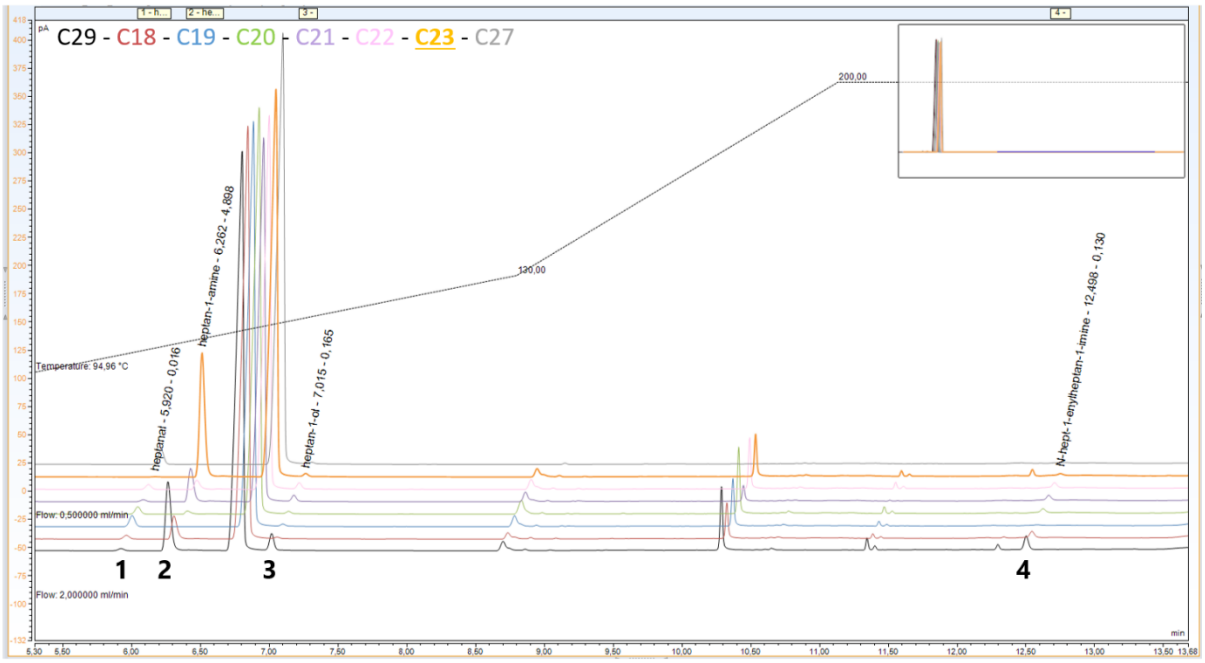
1: heptanal; 5.927; 0.027
2: heptan-1-amine; 6.265; 7.234
3: heptan-1-ol; 7.018; 0.151
4: *N*-heptyl-1-heptanimine; 12.503; 0.082

C



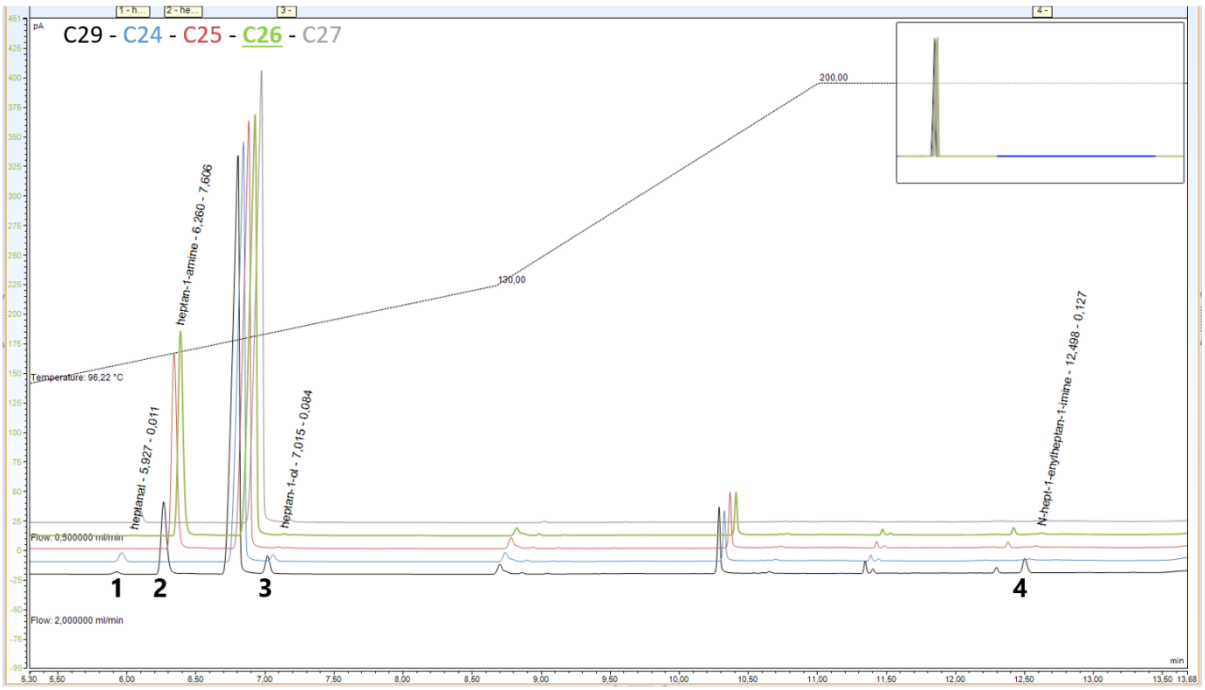
1: heptanal; 5.932; 0.014
2: heptan-1-amine; 6.272; 6.694
3: heptan-1-ol; 7.023; 0.091
4: *N*-heptyl-1-heptanimine; 12.505; 0.172

D



1: heptanal; 5.920; 0.016
2: heptan-1-amine; 6.262; 4.898
3: heptan-1-ol; 7.015; 0.165
4: *N*-heptyl-1-heptanimine; 12.498; 0.130

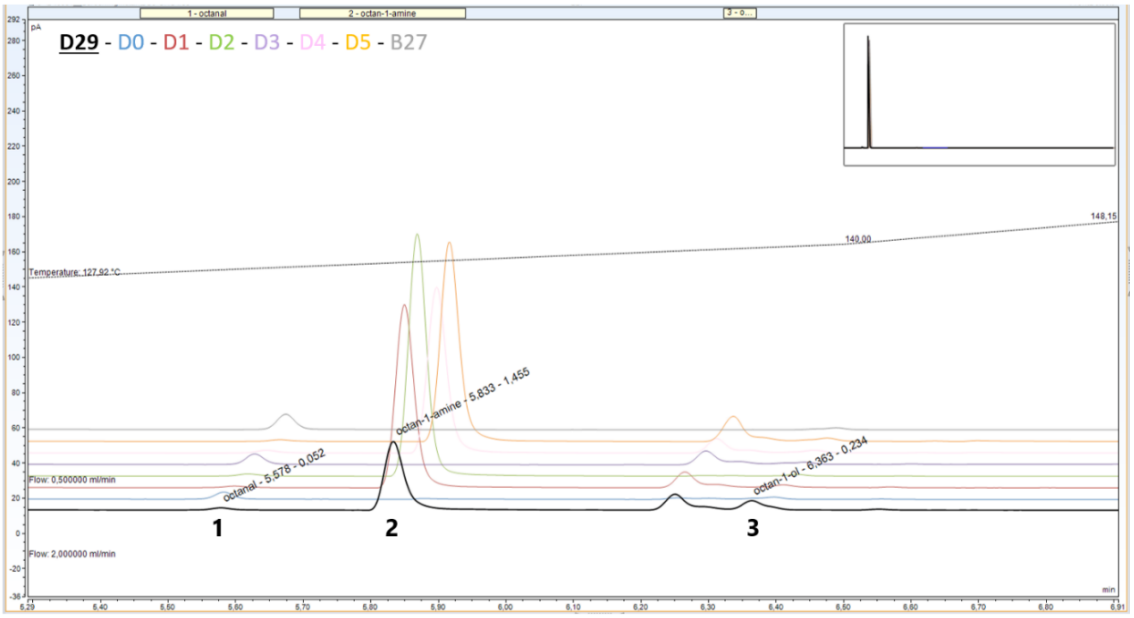
E



1: heptanal; 5.927; 0.011
2: heptan-1-amine; 6.260; 7.606
3: heptan-1-ol; 7.015; 0.084
4: *N*-heptyl-1-heptanimine; 12.498; 0.127

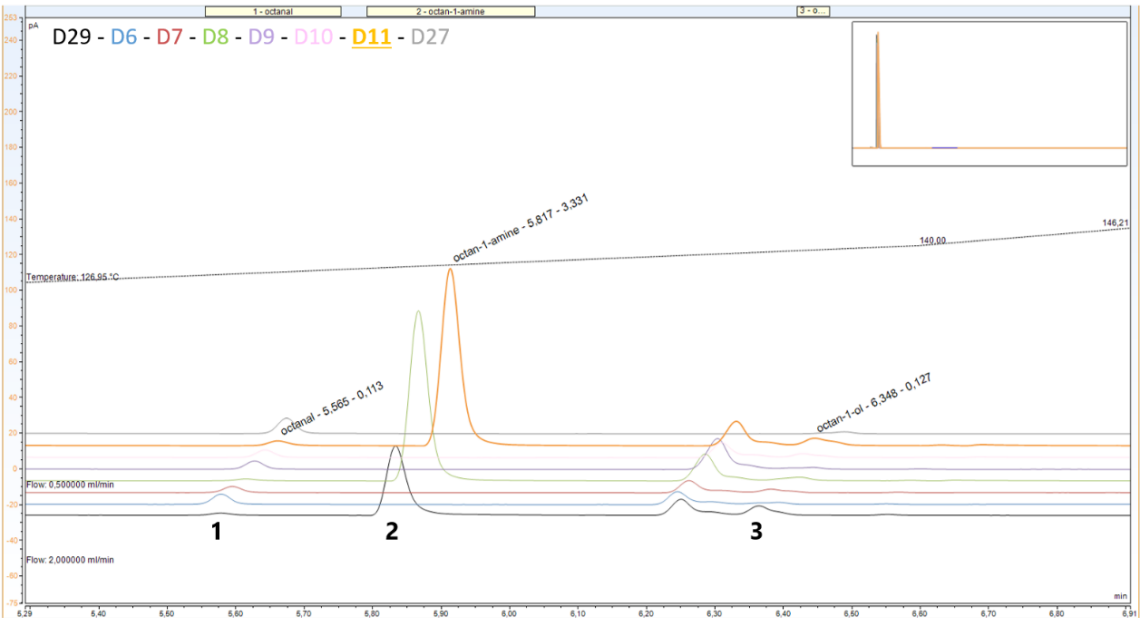
Appendix 28. (Chapter III, I.4.3) GC-FID chromatograms for reaction of octanal (**218**) with various nat-AmDHs and mutants. Chromatograms of a commercial standard of octan-1-amine (**214)** (3 mM), **218** (6.4 mM), corresponding alcohol (0.6 mM), blank reaction mixture without enzyme and reaction mixture with (A) *Cfus*AmDH, *Apau*AmDH and their mutants (B) *Chat*AmDH, IGCAmDH1 and their mutants (C) MATOUAmDH2, *Micro*AmDH and their mutants (D) *Msme*AmDH, *Porti*AmDH and their mutants and (E) *Rgna*AmDH and its mutants after extraction in EtOAc. The legend above matches the color of the chromatogram with the enzyme/substrate couple according to Appendix 22. The legend below indicates: compound; retention time (min); area.

A



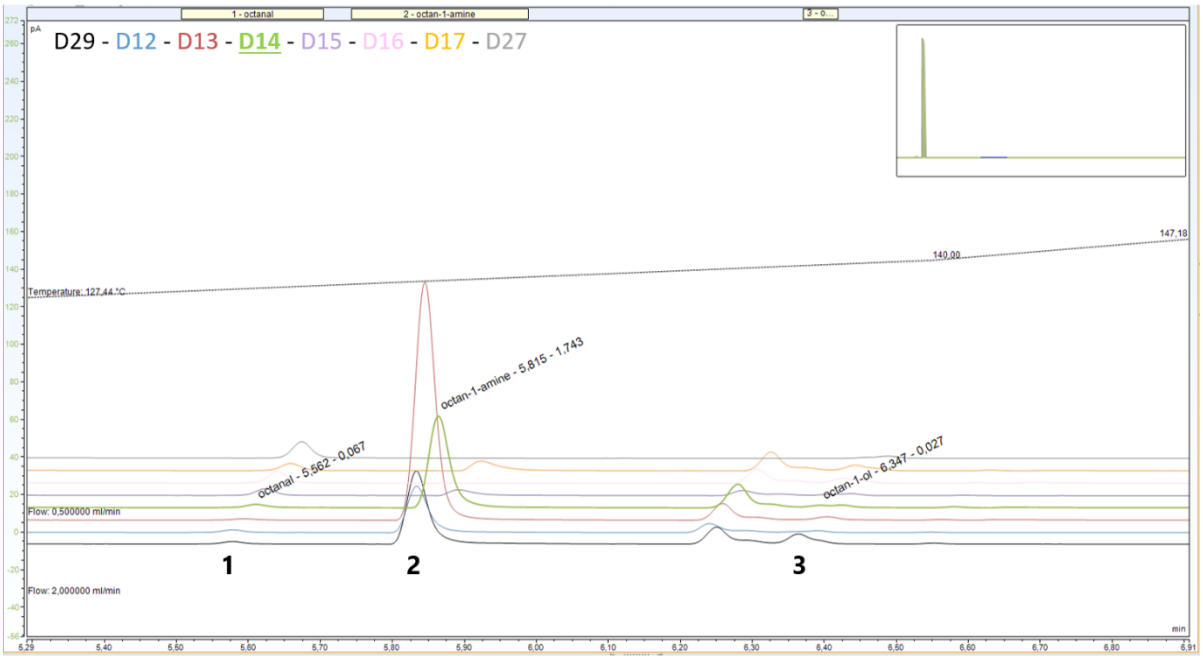
1: octanal; 5.578; 0.052 **2**: octan-1-amine; 5.833; 1.455 **3**: octan-1-ol; 6.363; 0.234

B



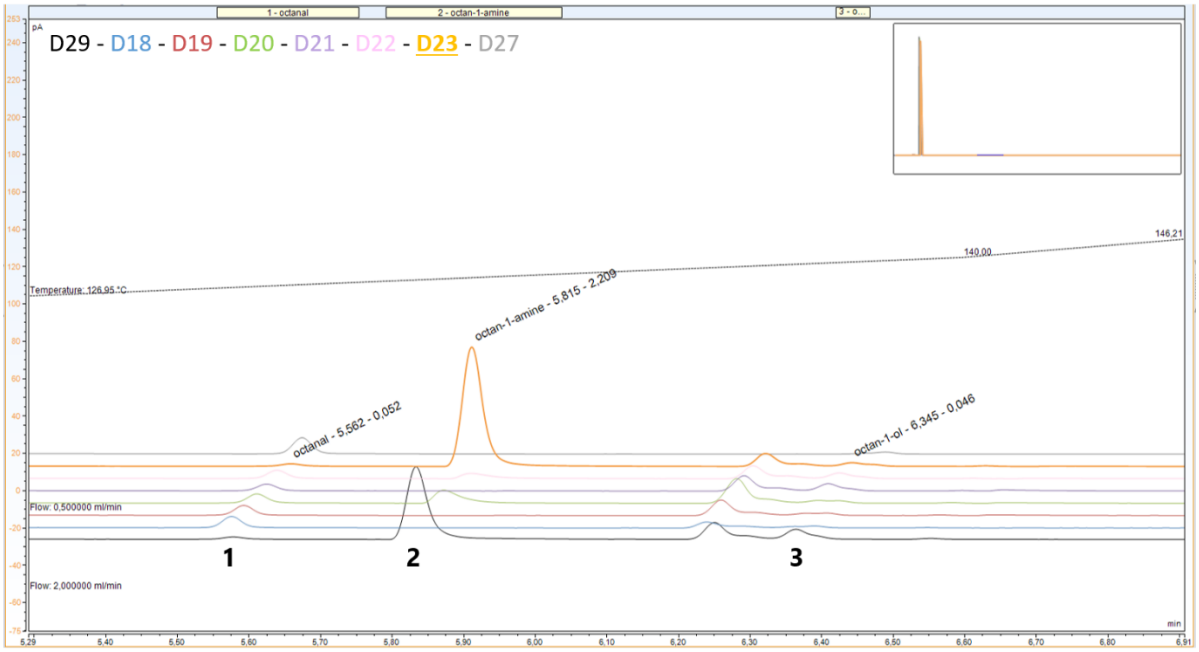
1: octanal; 5.565; 0.113 **2**: octan-1-amine; 5.817; 3.331 **3**: octan-1-ol; 6.348; 0.127

C



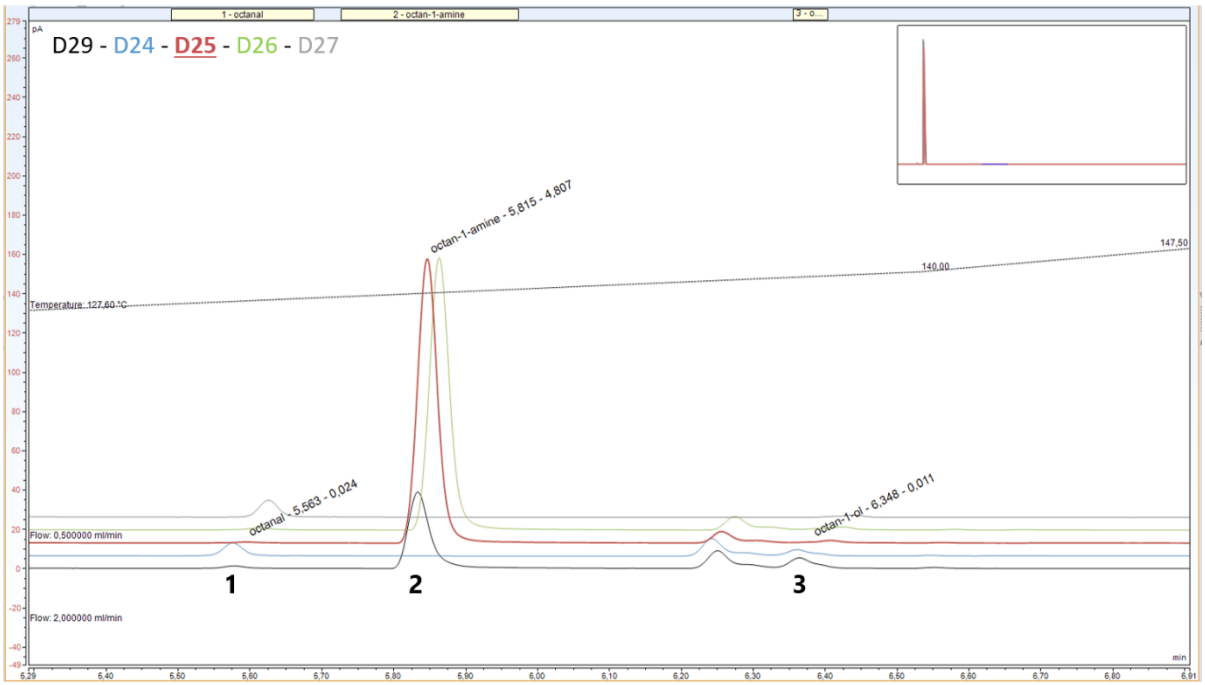
1: octanal; 5.562; 0.067 **2**: octan-1-amine; 5.815; 1.743 **3**: octan-1-ol; 6.347; 0.027

D



1: octanal; 5.562; 0.052 **2**: octan-1-amine; 5.815; 2.209 **3**: octan-1-ol; 6.345; 0.046

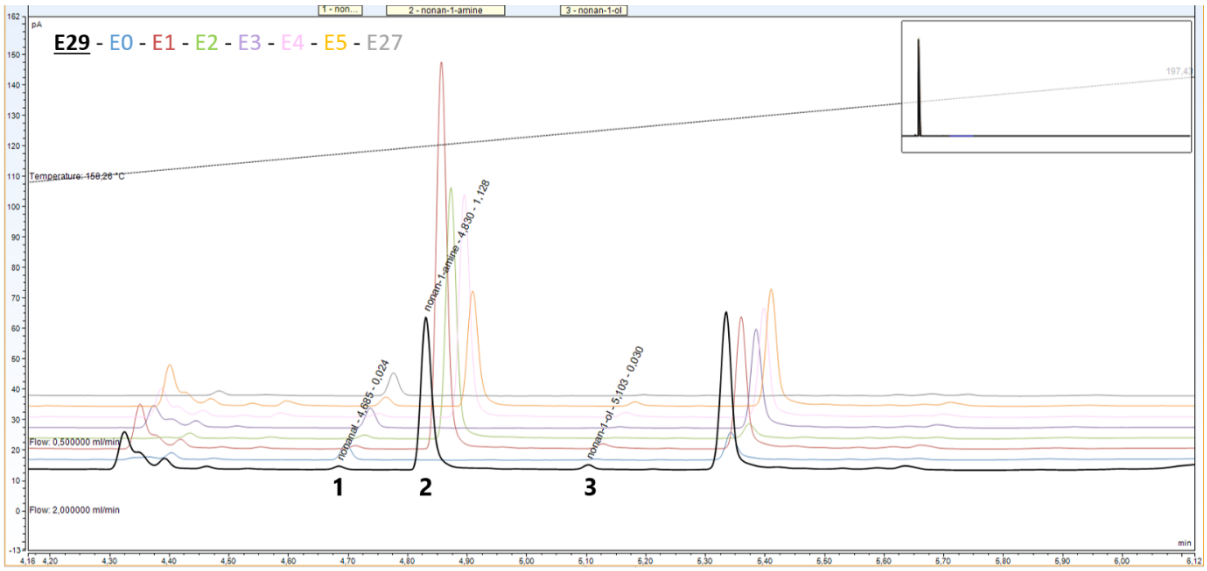
E



1: octanal; 5.563; 0.024 **2**: octan-1-amine; 5.815; 4.807 **3**: octan-1-ol; 6.348; 0.011

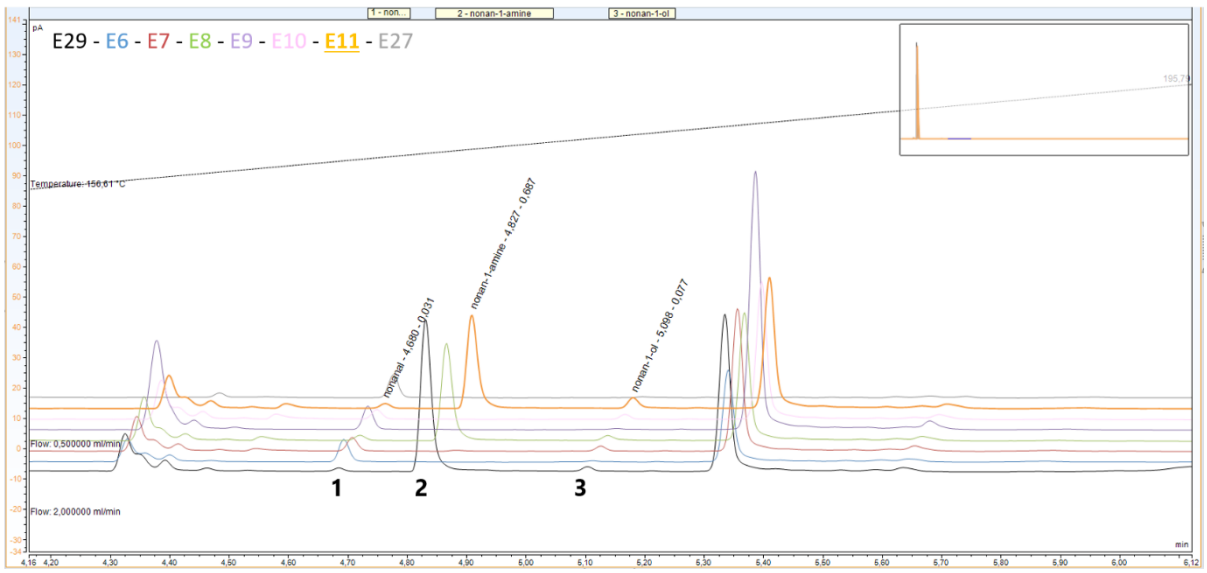
Appendix 29. (Chapter III, I.4.3) GC-FID chromatograms for reaction of nonanal (**219**) with various nat-AmDHs and mutants. Chromatograms of a commercial standard of nonan-1-amine (**215**) (3 mM), **219** (6.3 mM), corresponding alcohol (0.7 mM), blank reaction mixture without enzyme and reaction mixture with (A) *Cfus*AmDH, *Apau*AmDH and their mutants (B) *Chat*AmDH, IGCAmDH1 and their mutants (C) MATOUAmDH2, *Micro*AmDH and their mutants (D) *Msme*AmDH, *Porti*AmDH and their mutants and (E) *Rgna*AmDH and its mutants after extraction in EtOAc. The legend above matches the color of the chromatogram with the enzyme/substrate couple according to Appendix 22. The legend below indicates: compound; retention time (min); area.

A



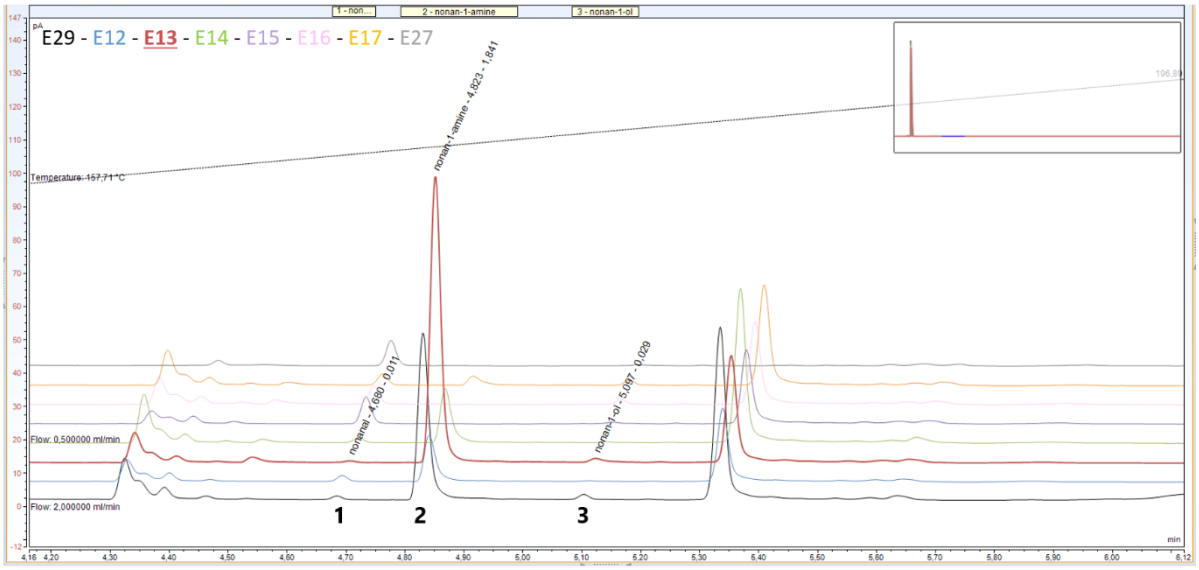
1: nonanal; 4.685; 0.024 **2**: nonan-1-amine; 4.830; 1.128 **3**: nonan-1-ol; 5.103; 0.030

B



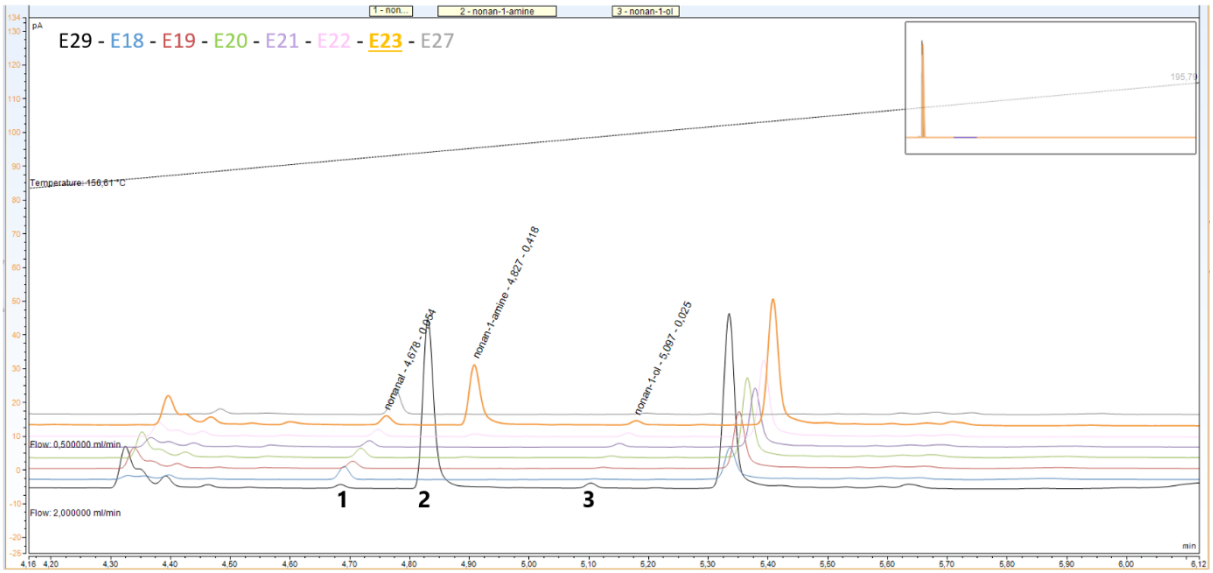
1: nonanal; 4.680; 0.031 **2**: nonan-1-amine; 4.827; 0.687 **3**: nonan-1-ol; 5.098; 0.077

C



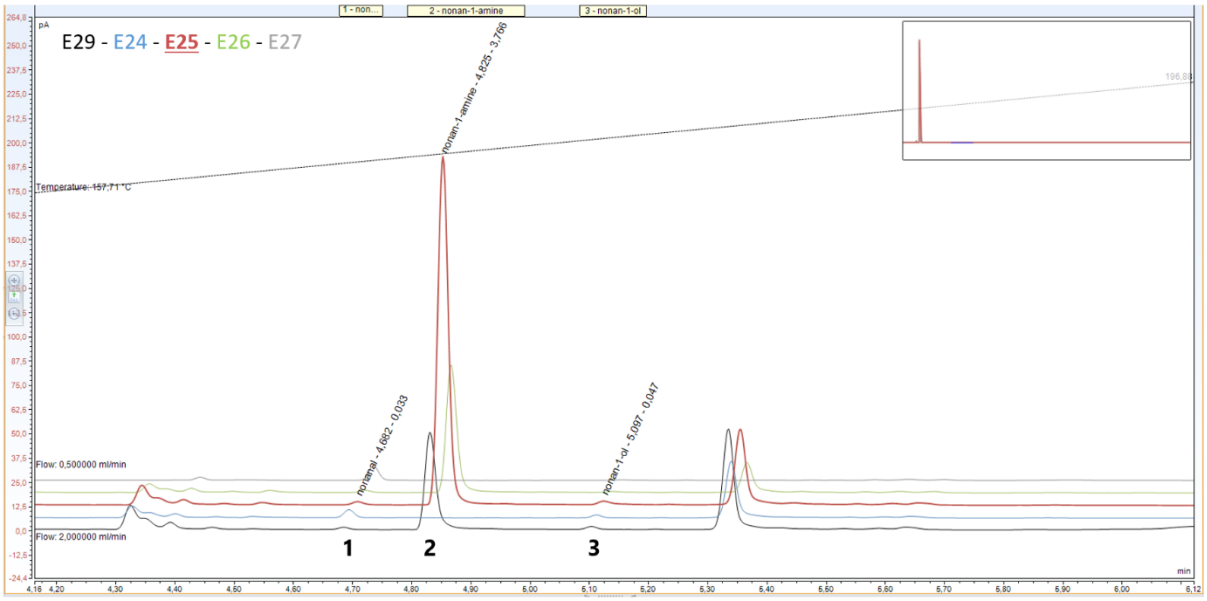
1: nonanal; 4.680; 0.011 **2**: nonan-1-amine; 4.823; 1.841 **3**: nonan-1-ol; 5.097; 0.029

D



1: nonanal; 4.678; 0.054 **2**: nonan-1-amine; 4.827; 0.418 **3**: nonan-1-ol; 5.097; 0.025

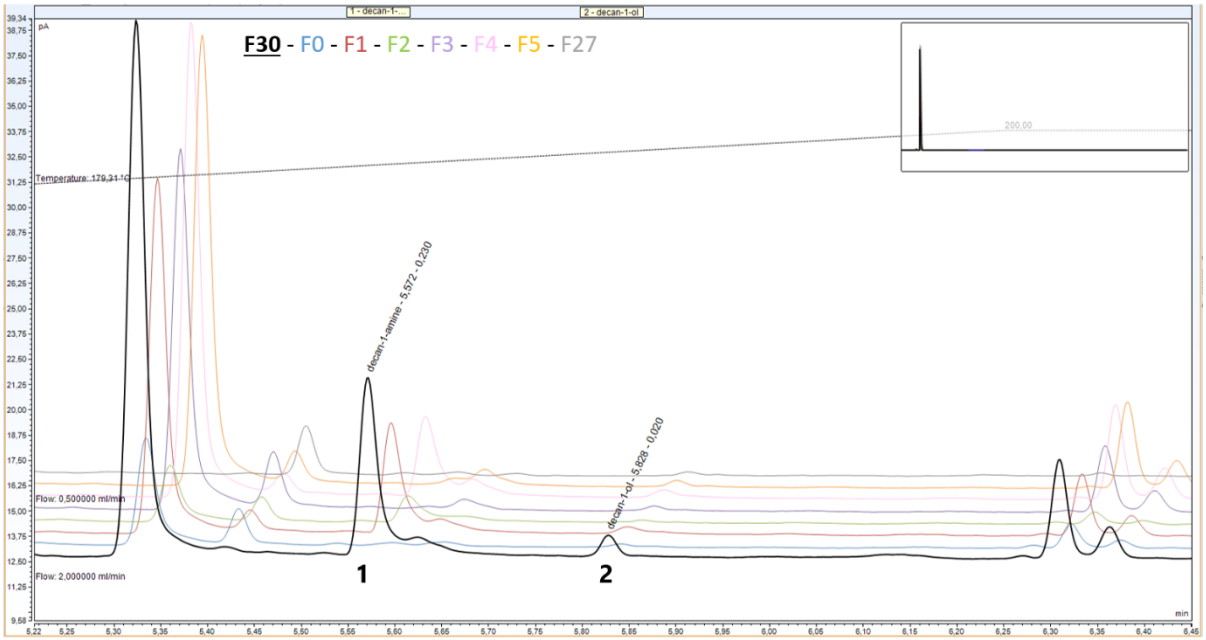
E



1: nonanal; 4.682; 0.033 **2**: nonan-1-amine; 4.825; 3.766 **3**: nonan-1-ol; 5.097; 0.047

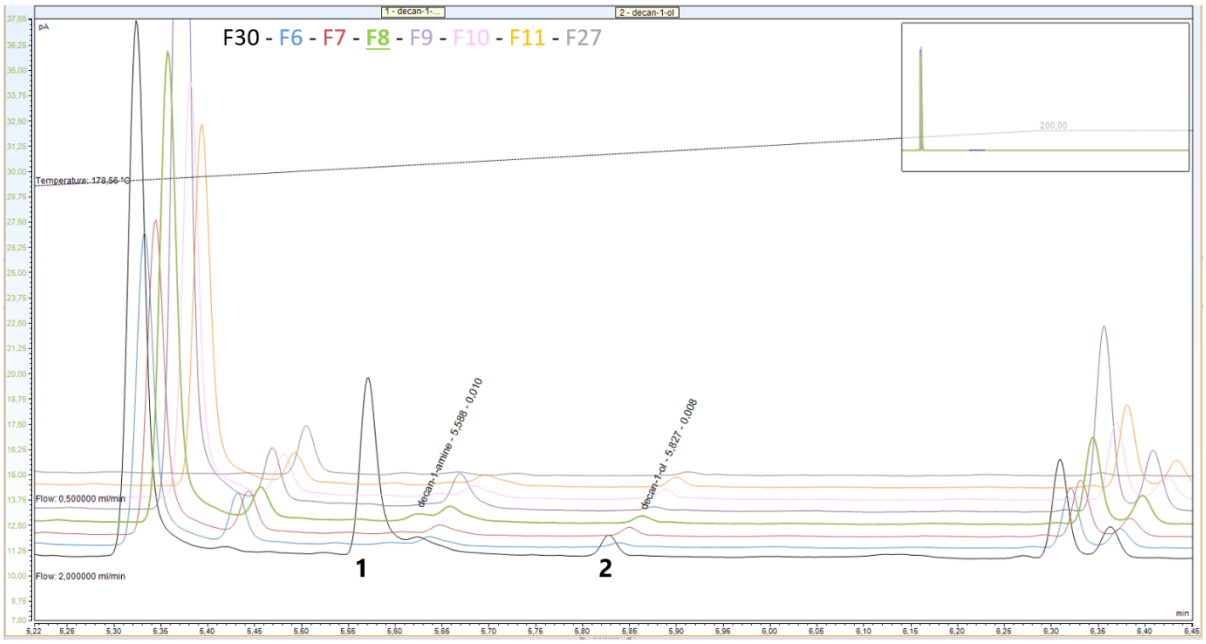
Appendix 30. (Chapter III, I.4.3) GC-FID chromatograms for reaction of decanal (**220**) with various nat-AmDHs and mutants. Chromatograms of a commercial standard of decan-1-amine (**216**) (2 mM), corresponding alcohol (0.6 mM), blank reaction mixture without enzyme and reaction mixture with (A) *Cfus*AmDH, *Apau*AmDH and their mutants (B) *Chat*AmDH, IGCAmDH1 and their mutants (C) MATOUAmDH2, *Micro*AmDH and their mutants (D) *Msme*AmDH, *Porti*AmDH and their mutants and (E) *Rgna*AmDH and its mutants after extraction in EtOAc. The legend above matches the color of the chromatogram with the enzyme/substrate couple according to Appendix 22. The legend below indicates: compound; retention time (min); area.

A



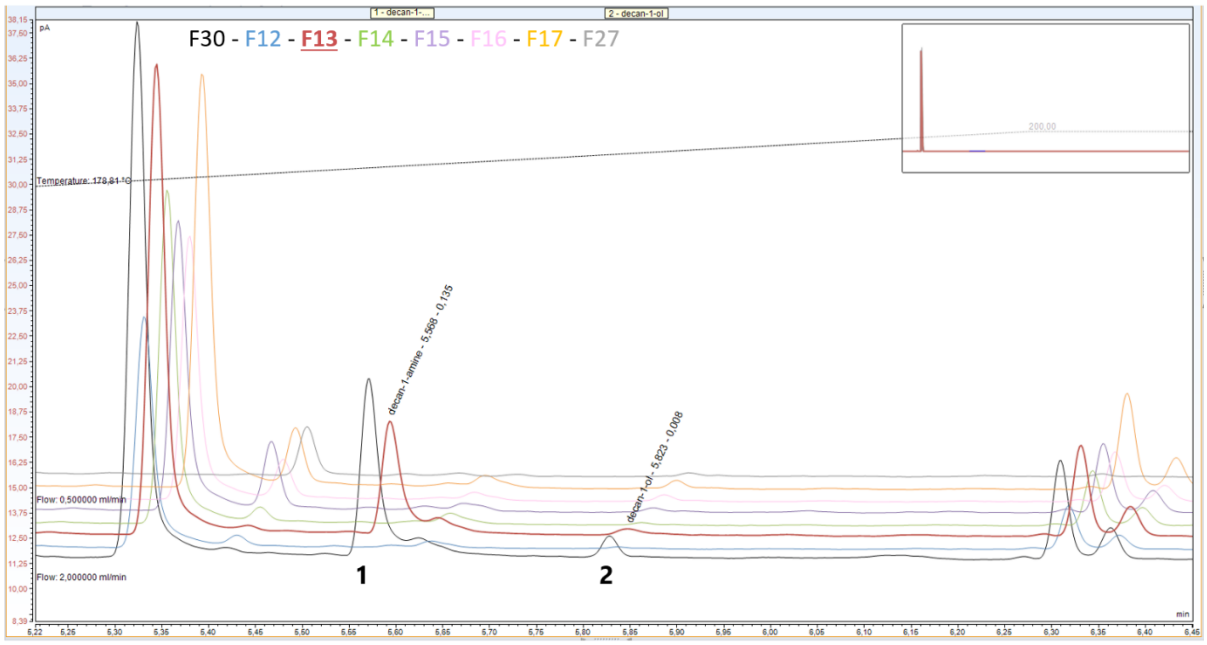
1: decan-1-amine; 5.572; 0.230 **2**: decan-1-ol; 5.828; 0.020

B



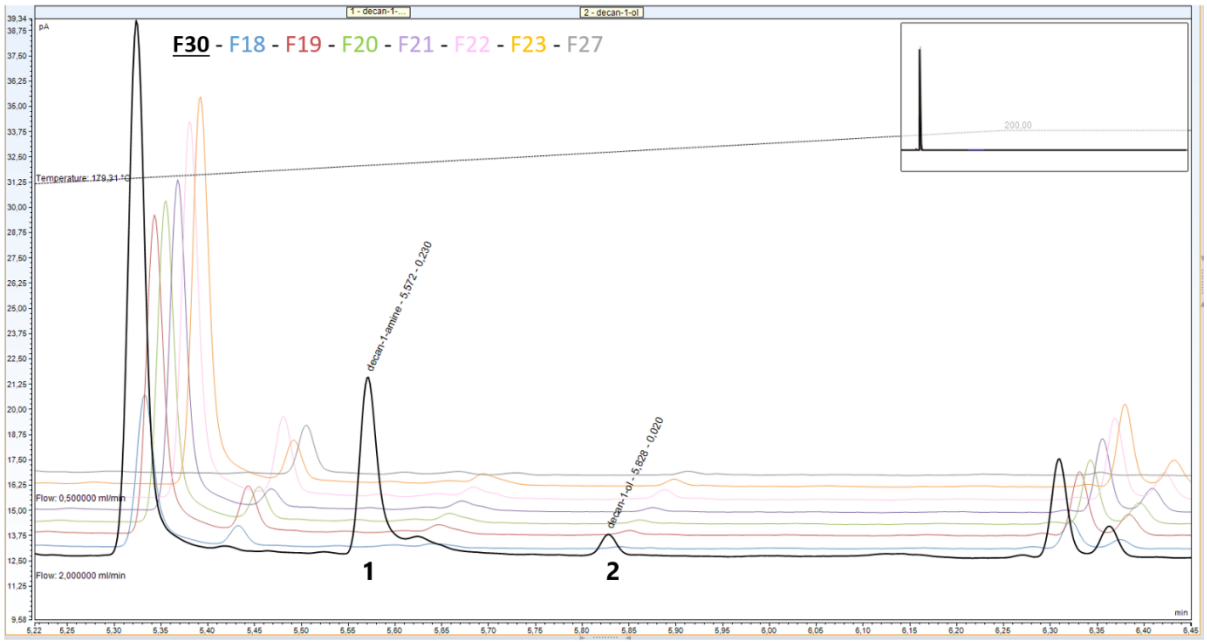
1: decan-1-amine; 5.588; 0.010 **2**: decan-1-ol; 5.827; 0.008

C



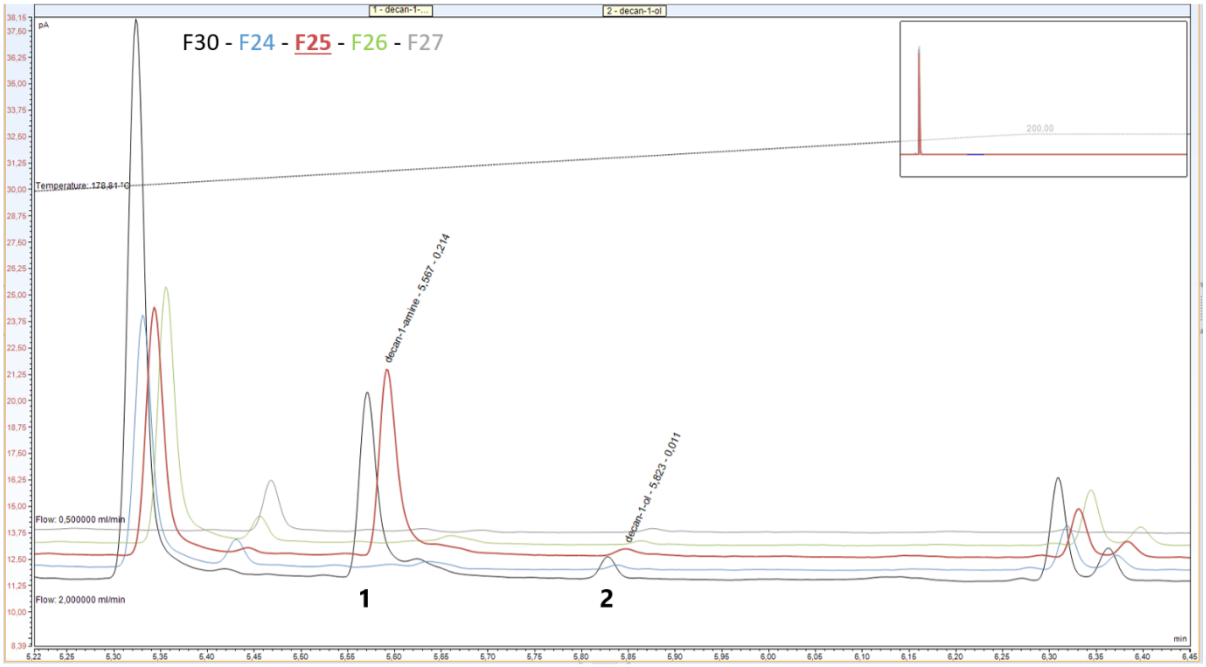
1: decan-1-amine; 5.568; 0.135 **2**: decan-1-ol; 5.823; 0.008

D



1: decan-1-amine; 5.572; 0.230 **2**: decan-1-ol; 5.828; 0.020

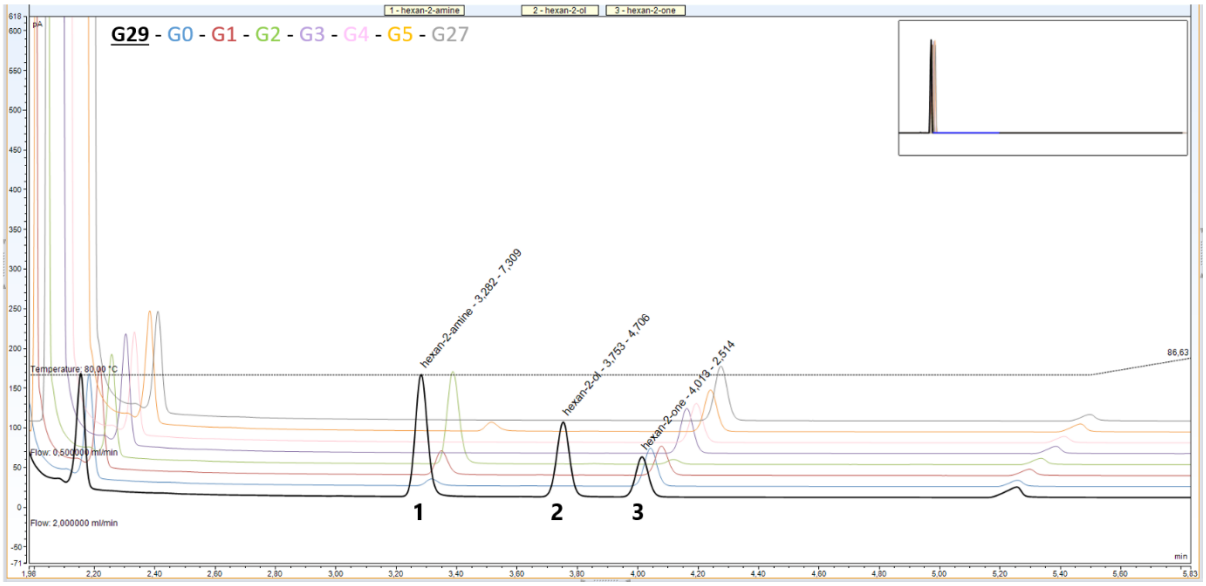
E



1: decan-1-amine; 5.567; 0.214 **2**: decan-1-ol; 5.823; 0.011

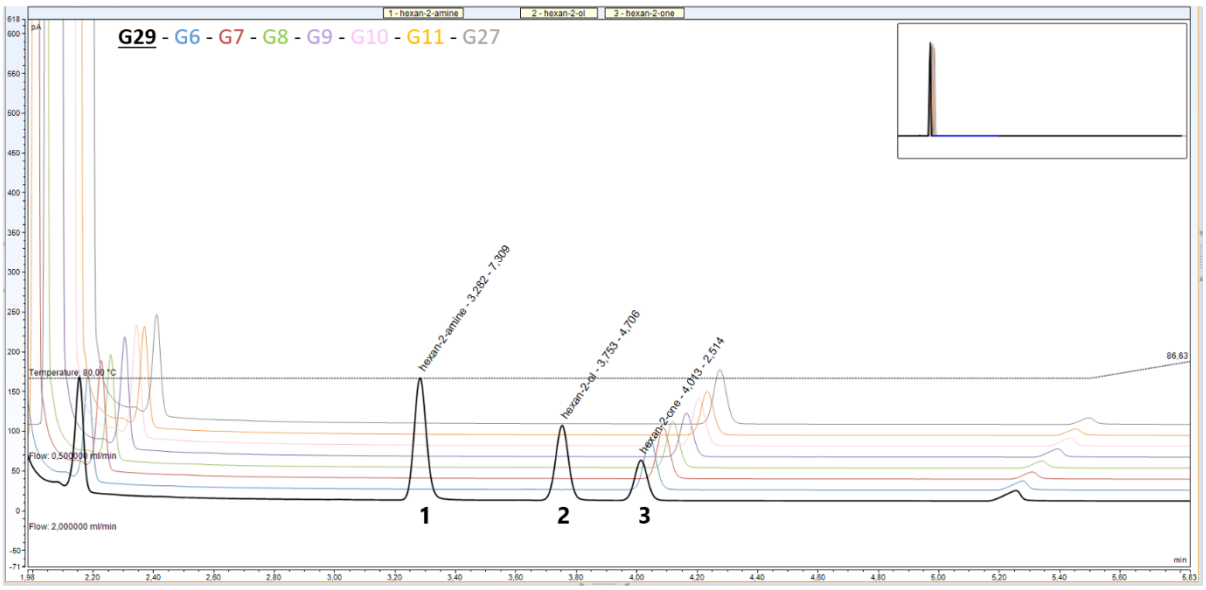
Appendix 31. (Chapter III, I.4.3) GC-FID chromatograms for reaction of hexan-2-one **(43** with various nat-AmDHs and mutants. Chromatograms of a commercial standard of amine (**229**) (3 mM), **43** (6.4 mM), the corresponding alcohol (0.6 mM), blank reaction mixture without enzyme and reaction mixture with (A) *Cfus*AmDH, *Apau*AmDH and their mutants (B) *Chat*AmDH, IGCAmDH1 and their mutants (C) MATOUAmDH2, *Micro*AmDH and their mutants (D) *Msme*AmDH, *Porti*AmDH and their mutants and (E) *Rgna*AmDH and its mutants after extraction in EtOAc. The legend above matches the color of the chromatogram with the enzyme/substrate couple according to Appendix 22. The legend below indicates: compound; retention time (min); area.

A



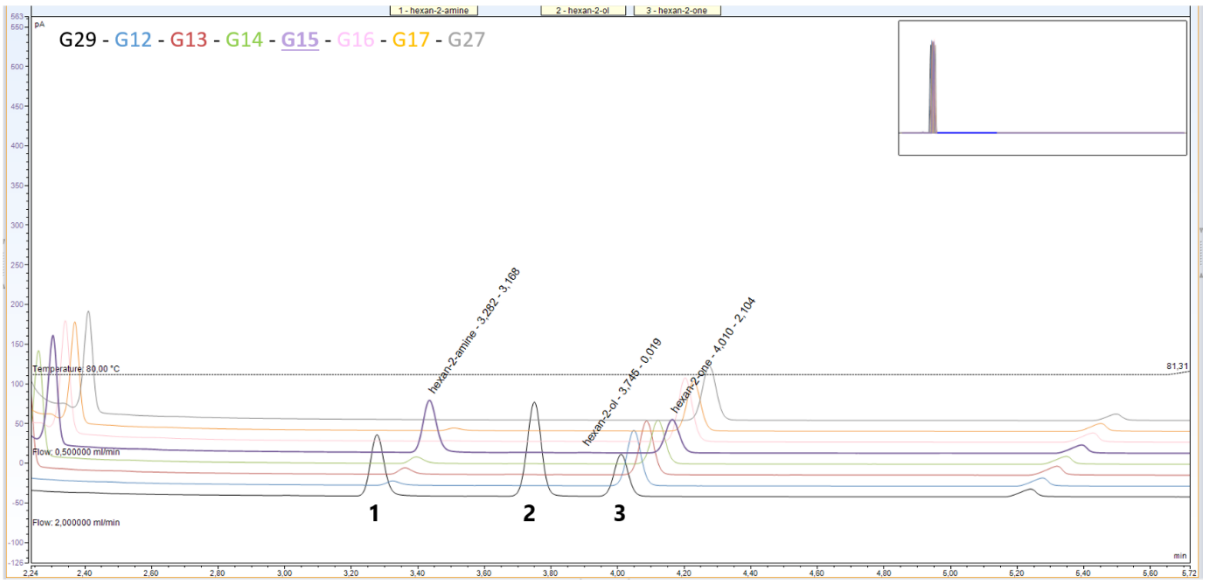
1: hexan-2-amine; 3.282; 7.309 **2**: hexan-2-ol; 3.753; 4.706 **3**: hexan-2-one; 4.013; 2.514

B



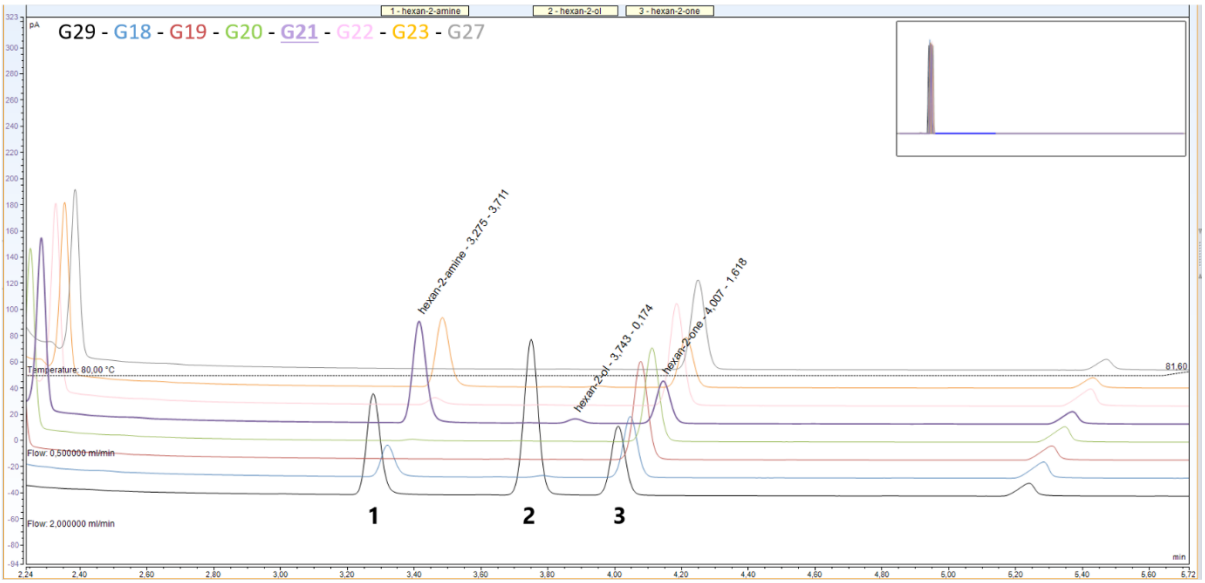
1: hexan-2-amine; 3.282; 7.309 **2**: hexan-2-ol; 3.753; 4.706 **3**: hexan-2-one; 4.013; 2.514

C



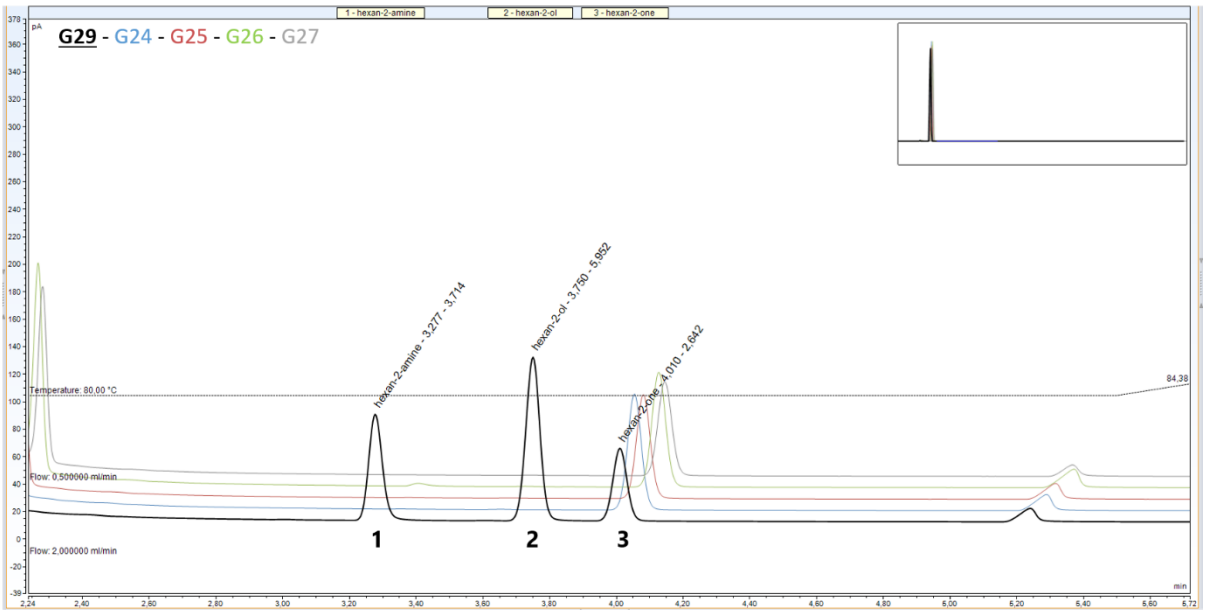
1: hexan-2-amine; 3.282; 3.168 **2**: hexan-2-ol; 3.745; 0.019 **3**: hexan-2-one; 4.010; 2.104

D



1: hexan-2-amine; 3.275; 3.711 **2**: hexan-2-ol; 3.743; 0.174 **3**: hexan-2-one; 4.007; 1.618

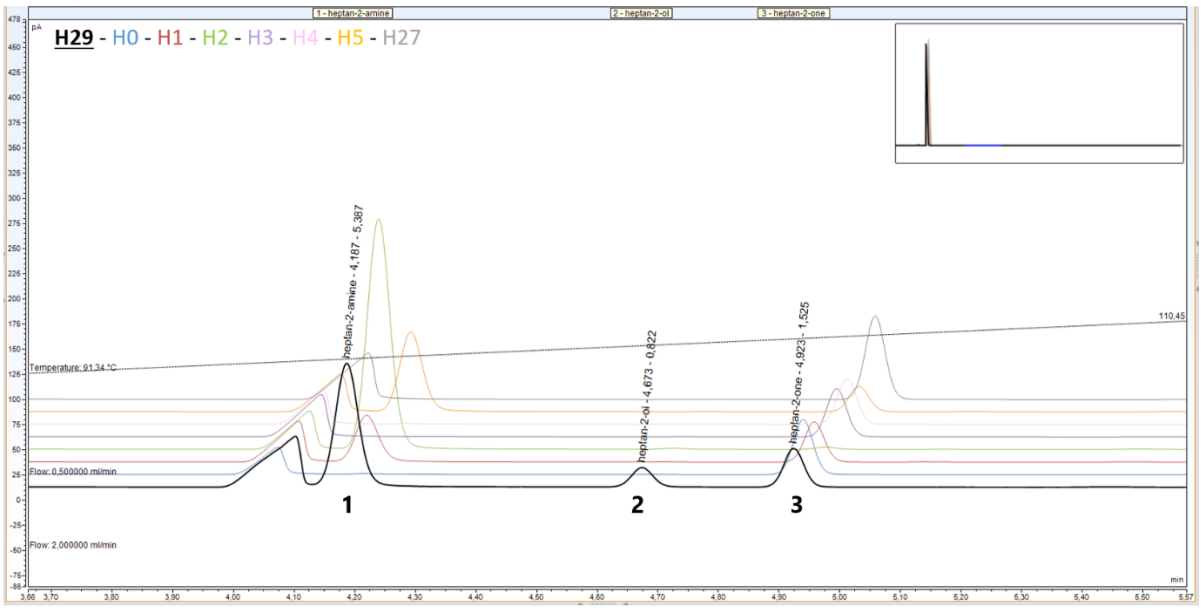
E



1: hexan-2-amine; 3.277; 3.714 **2**: hexan-2-ol; 3.750; 5.952 **3**: hexan-2-one; 4.010; 2.642

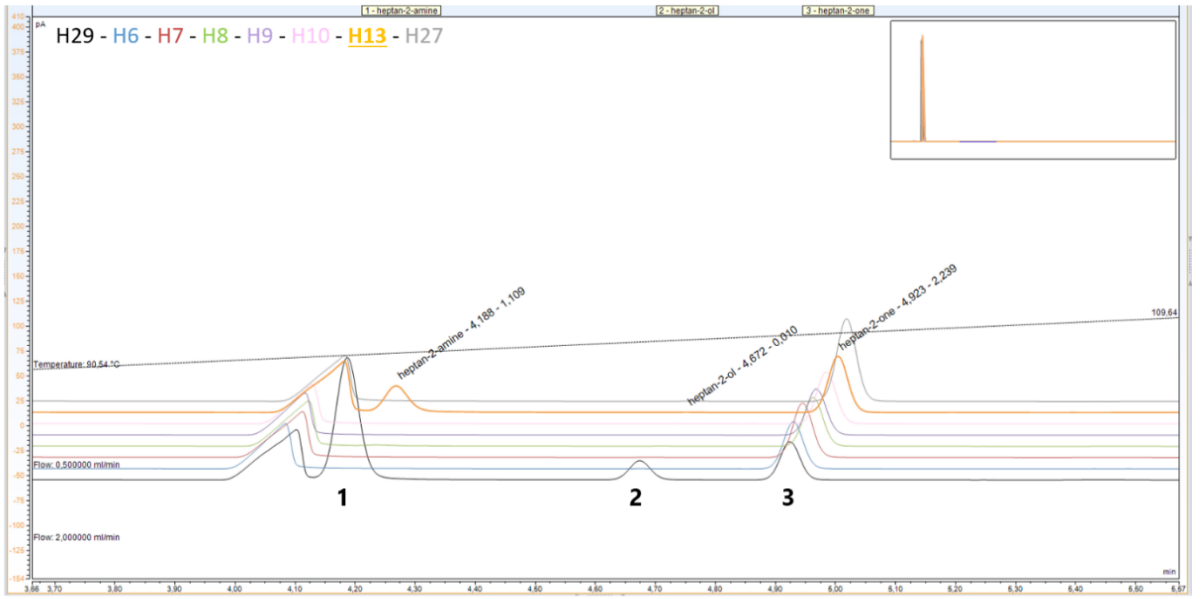
Appendix 32. (Chapter III, I.4.3) GC-FID chromatograms for reaction of heptan-2-one (**128**) with various nat-AmDHs and mutants. Chromatograms of a commercial standard of heptan-2-amine **230** (3 mM), **128** (6.4 mM), the corresponding alcohol (0.6 mM), blank reaction mixture without enzyme and reaction mixture with (A) *Cfus*AmDH, *Apau*AmDH and their mutants (B) *Chat*AmDH, IGCAmDH1 and their mutants (C) MATOUAmDH2, *Micro*AmDH and their mutants (D) *Msme*AmDH, *Porti*AmDH and their mutants and (E) *Rgna*AmDH and its mutants after extraction in EtOAc. The legend above matches the color of the chromatogram with the enzyme/substrate couple according to Appendix 22. The legend below indicates: compound; retention time (min); area.

A



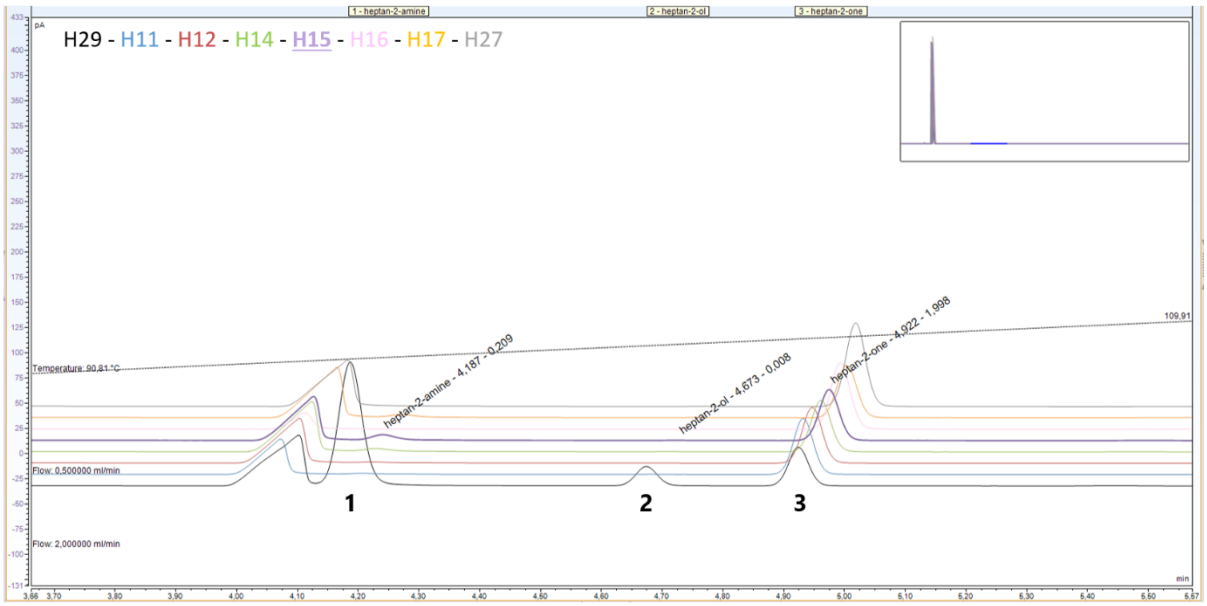
1: heptan-2-amine; 4.187; 5.387 **2**: heptan-2-ol; 4.673; 0.822 **3**: heptan-2-one; 4.923; 1.525

B



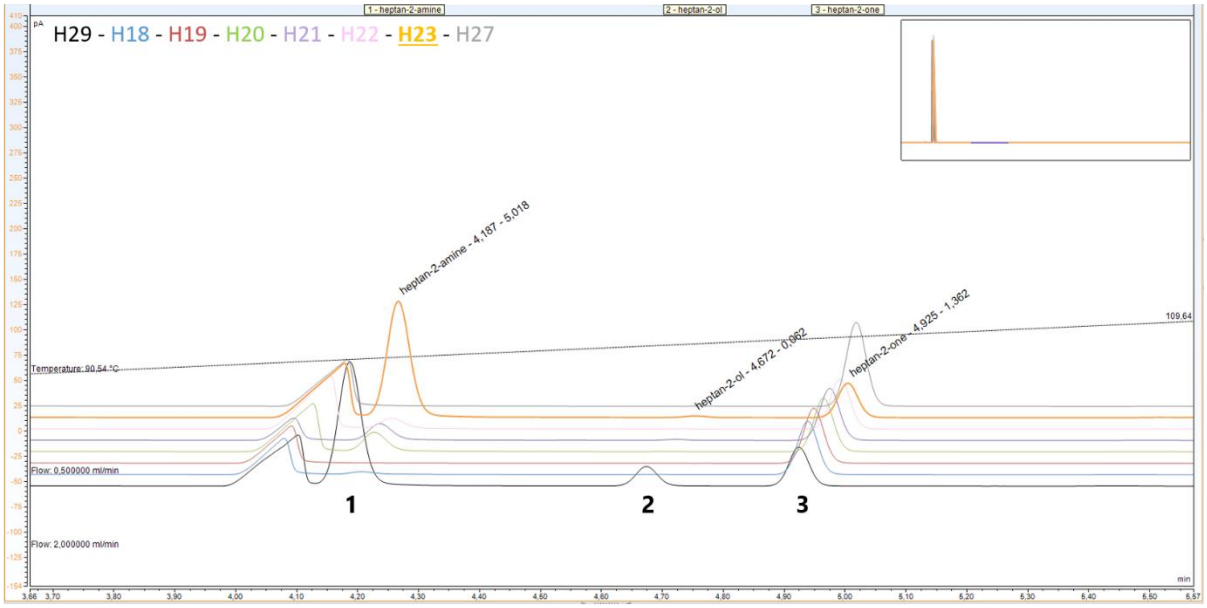
1: heptan-2-amine; 4.188; 1.109 **2**: heptan-2-ol; 4.672; 0.010 **3**: heptan-2-one; 4.923; 2.239

C



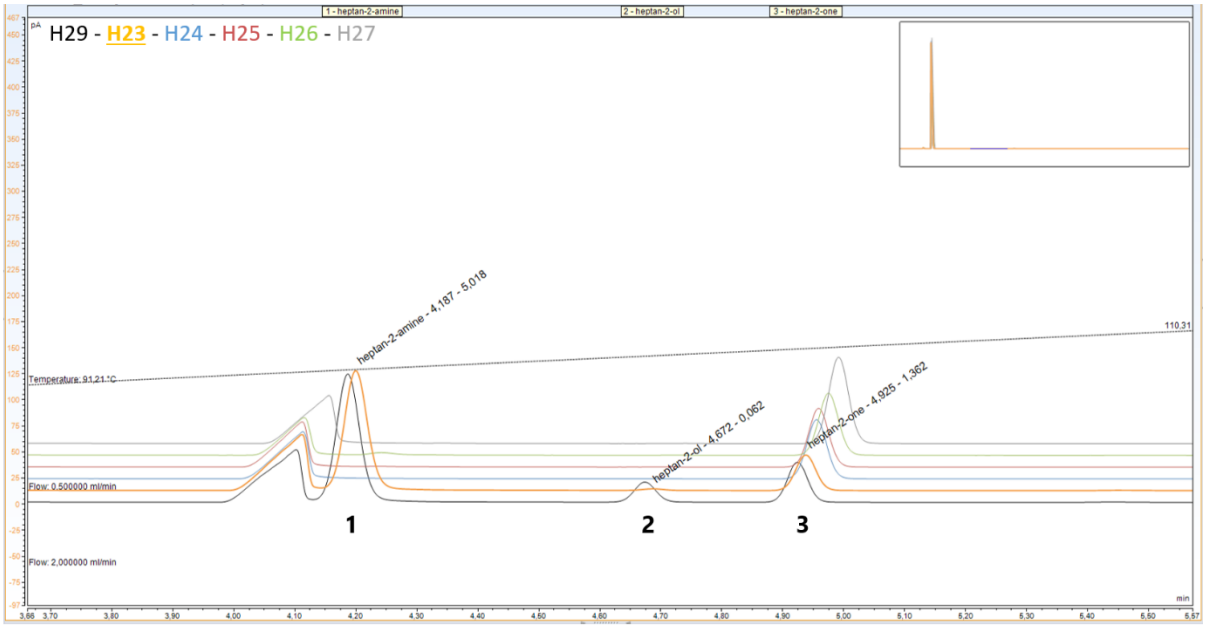
1: heptan-2-amine; 4.187; 0.209 **2**: heptan-2-ol; 4.673; 0.008 **3**: heptan-2-one; 4.922; 1.998

D



1: heptan-2-amine; 4.187; 5.018 **2**: heptan-2-ol; 4.672; 0.062 **3**: heptan-2-one; 4.925; 1.362

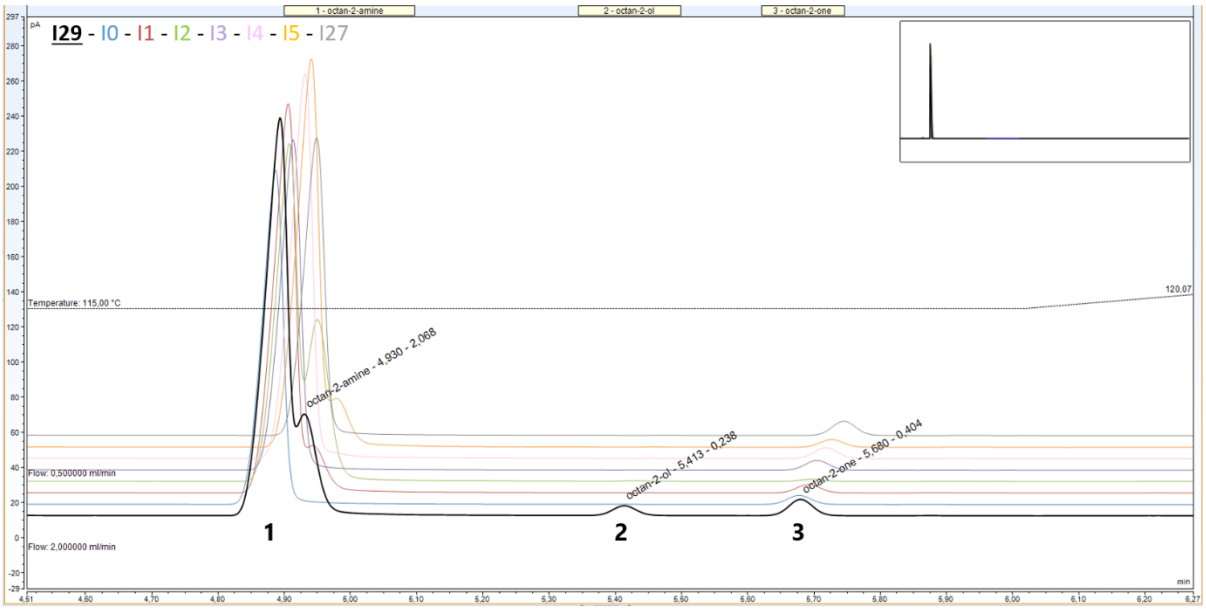
E



1: heptan-2-amine; 4.187; 5.018 **2**: heptan-2-ol; 4.672; 0.062 **3**: heptan-2-one; 4.925; 1.362

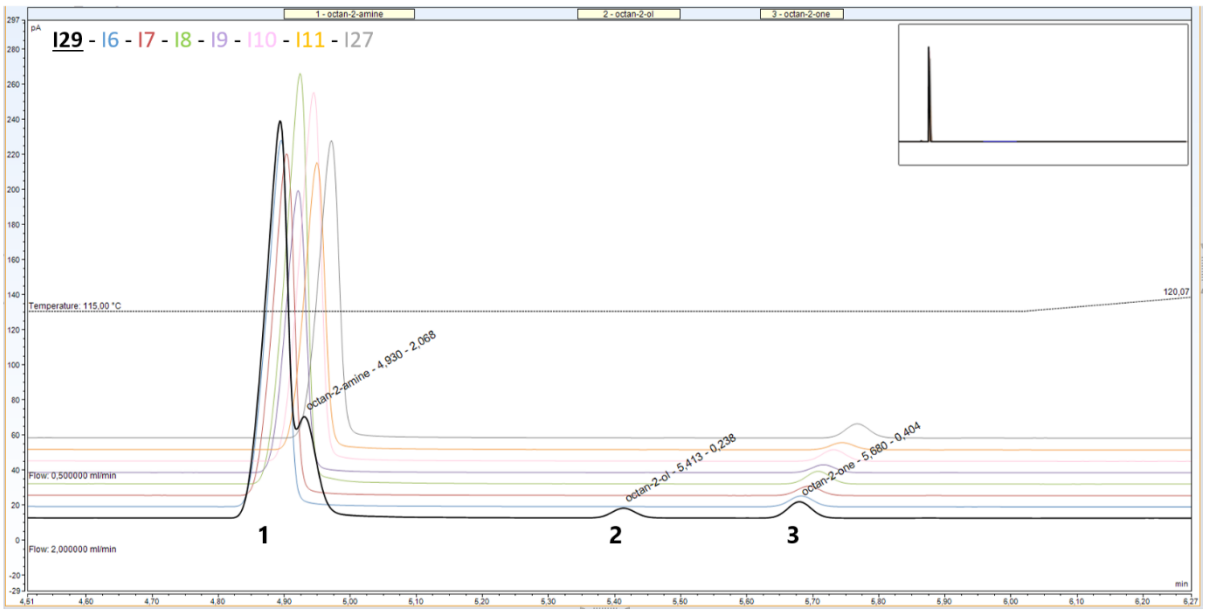
Appendix 33. (Chapter III, I.4.3) GC-FID chromatograms for reaction of octan-2-one (**67**) with various nat-AmDHs and mutants. Chromatograms of a commercial standard of octan-2-amine **231** (3mM), **67** (6.4 mM), the corresponding alcohol (0.6 mM), blank reaction mixture without enzyme and reaction mixture with (A) *Cfus*AmDH, *Apau*AmDH and their mutants (B) *Chat*AmDH, IGCAmDH1 and their mutants (C) MATOUAmDH2, *Micro*AmDH and their mutants (D) *Msme*AmDH, *Porti*AmDH and their mutants and (E) *Rgna*AmDH and its mutants after extraction in EtOAc. The legend above matches the color of the chromatogram with the enzyme/substrate couple according to Appendix 22. The legend below indicates: compound; retention time (min); area. Due to co-elution, the area was estimated by measuring the right half of the peak corresponding to **231** from the edge to the end and multiply it by two.

A



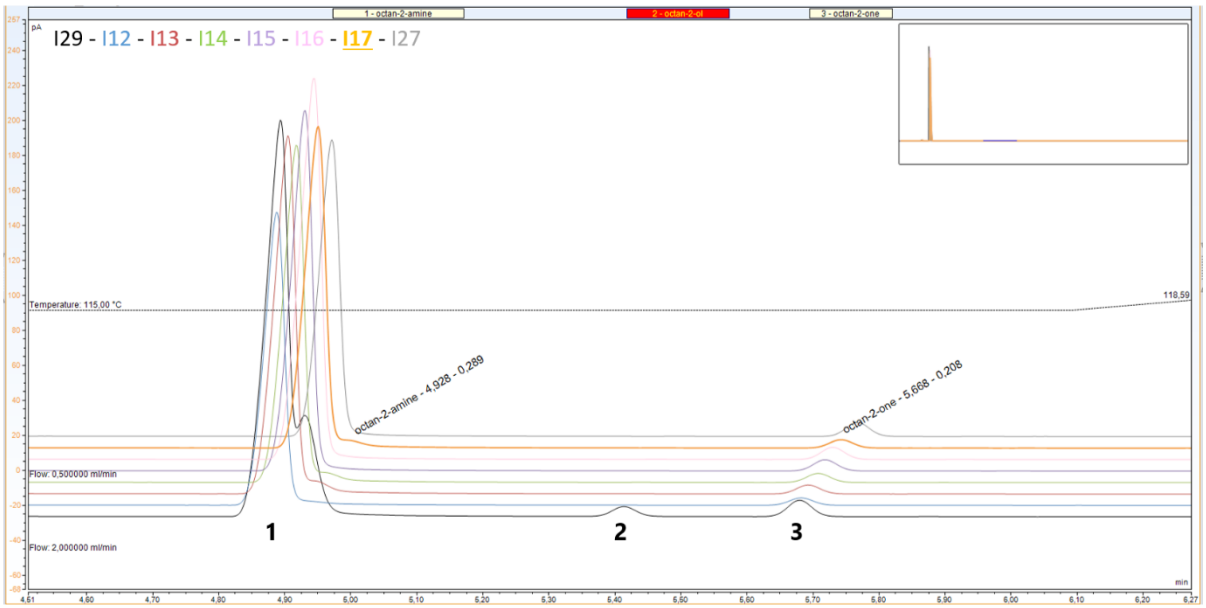
1: octan-2-amine; 4.930; 2.068 **2**: octan-2-ol; 5.413; 0.238 **3**: octan-2-one; 5.680; 0.404

B



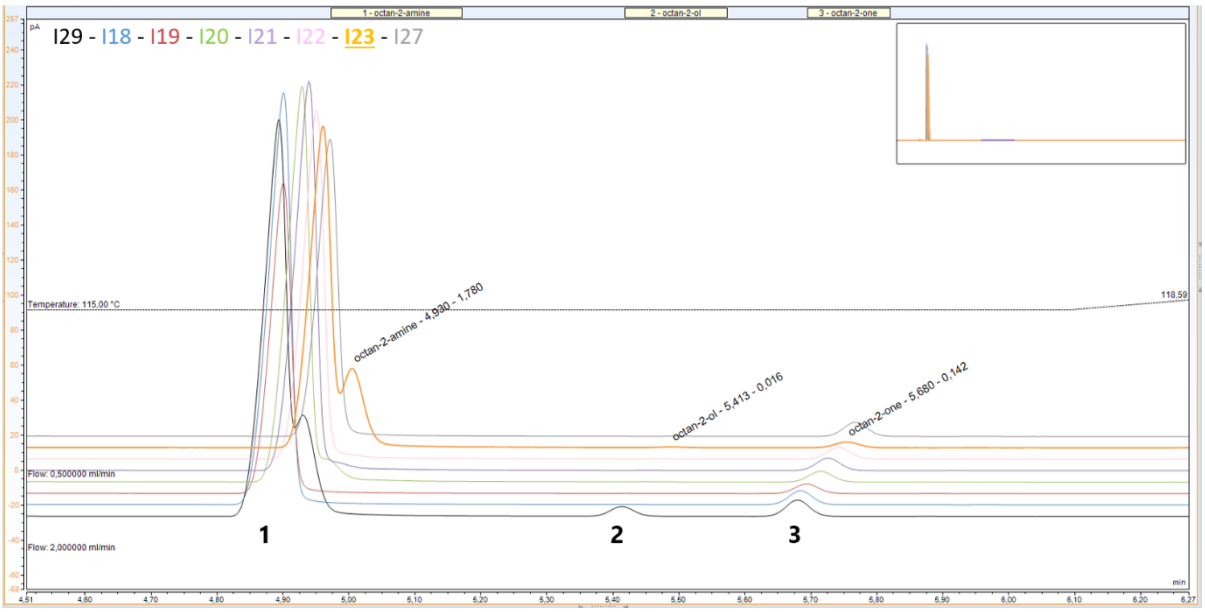
1: octan-2-amine; 4.930; 2.068 **2**: octan-2-ol; 5.413; 0.238 **3**: octan-2-one; 5.680; 0.404

C



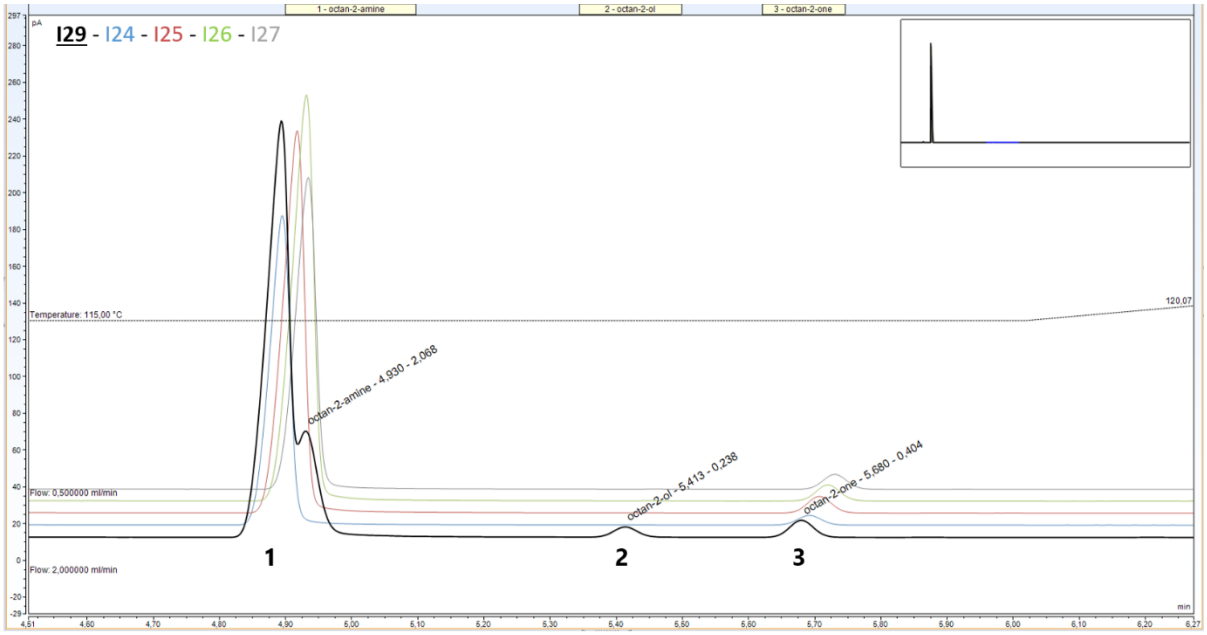
1: octan-2-amine; 4.928; 0.289 **2**: / **3**: octan-2-one; 5.668; 0.208

D



1: octan-2-amine; 4.930; 1.780 **2**: octan-2-ol; 5.413; 0.016 **3**: octan-2-one; 5.680; 0.142

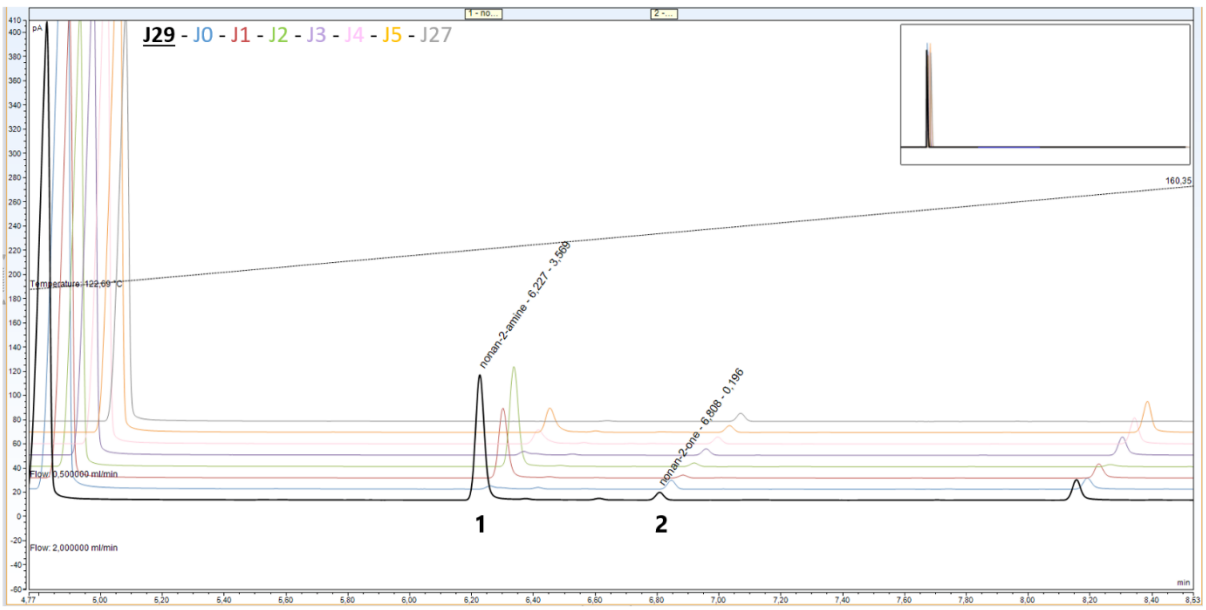
E



1: octan-2-amine; 4.930; 2.068 **2**: octan-2-ol; 5.413; 0.238 **3**: octan-2-one; 5.680; 0.404

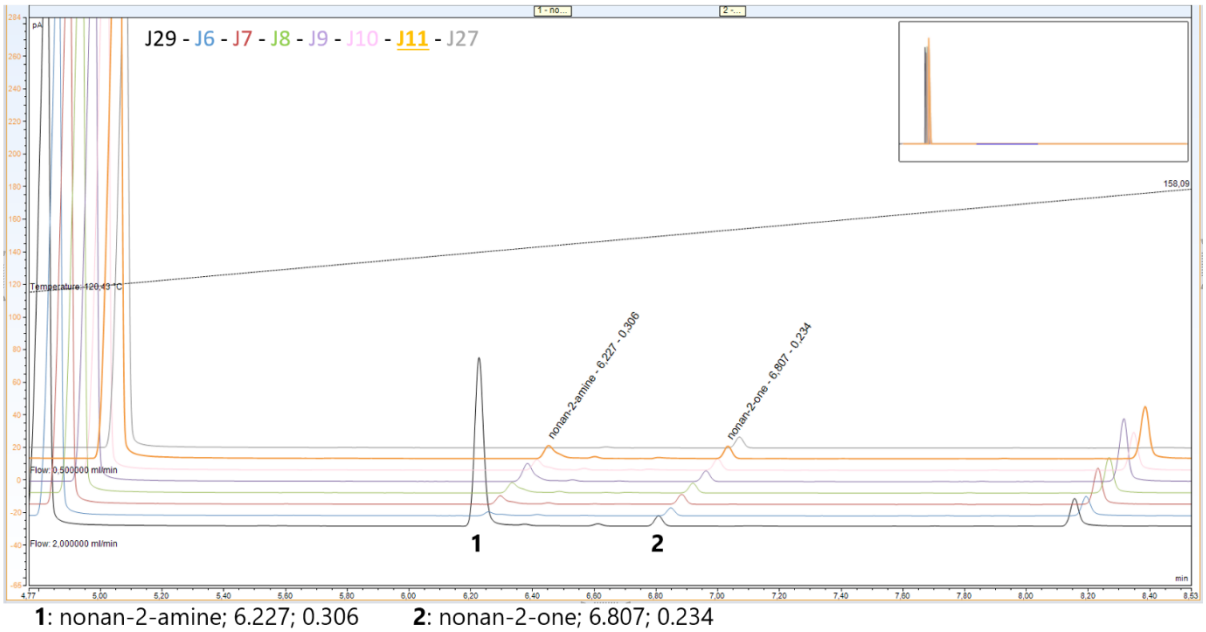
Appendix 34. (Chapter III, I.4.3) GC-FID chromatograms for reaction of nonan-2-one (**135**) with various nat-AmDHs and mutants. Chromatograms of a commercial standard of nonan-2-amine **187** (3 mM), **135** (6.4 mM), the corresponding alcohol (0.6 mM), blank reaction mixture without enzyme and reaction mixture with (A) *Cfus*AmDH, *Apau*AmDH and their mutants (B) *Chat*AmDH, IGCAmDH1 and their mutants (C) MATOUAmDH2, *Micro*AmDH and their mutants (D) *Msme*AmDH, *Porti*AmDH and their mutants and (E) *Rgna*AmDH and its mutants after extraction in EtOAc. The legend above matches the color of the chromatogram with the enzyme/substrate couple according to Appendix 22. The legend below indicates: compound; retention time (min); area.

A



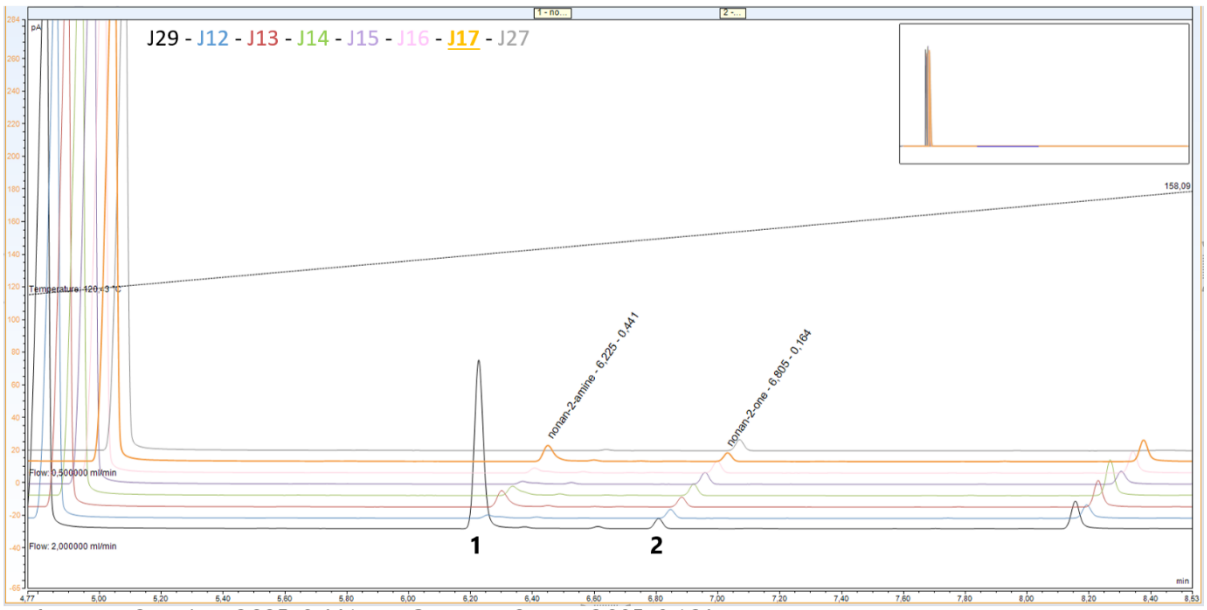
1: nonan-2-amine; 6.227; 3.569 **2**: nonan-2-one; 6.808; 0.196

B



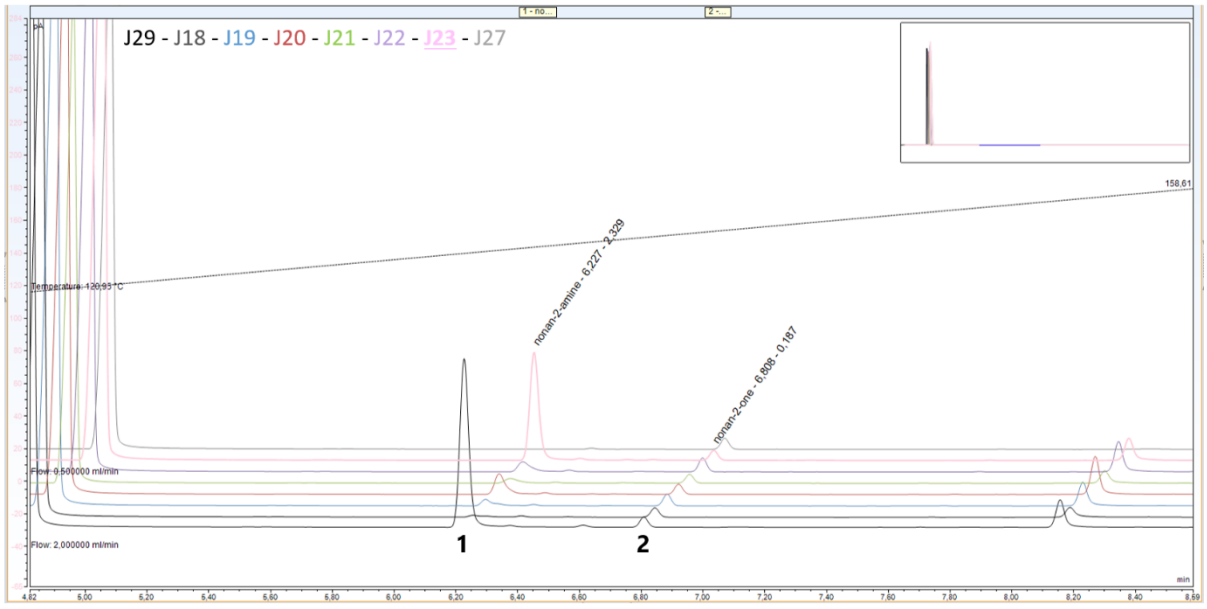
1: nonan-2-amine; 6.227; 0.306 **2**: nonan-2-one; 6.807; 0.234

C



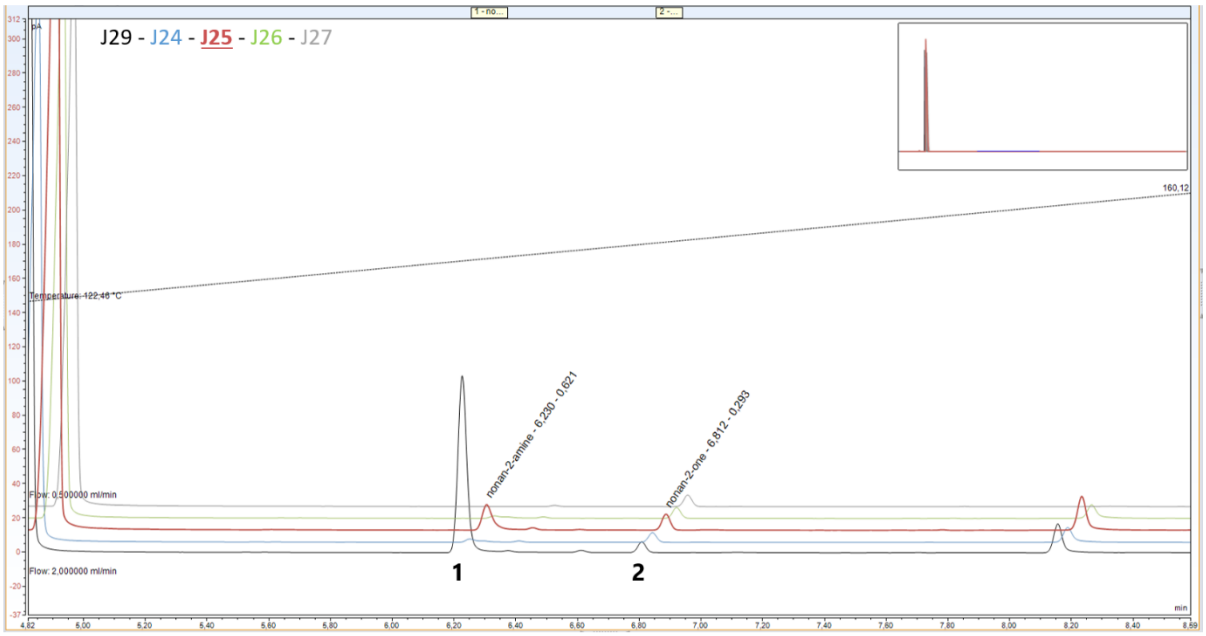
1: nonan-2-amine; 6.225; 0.441 **2**: nonan-2-one; 6.805; 0.164

D



1: nonan-2-amine; 6.227; 2.329 **2**: nonan-2-one; 6.808; 0.187

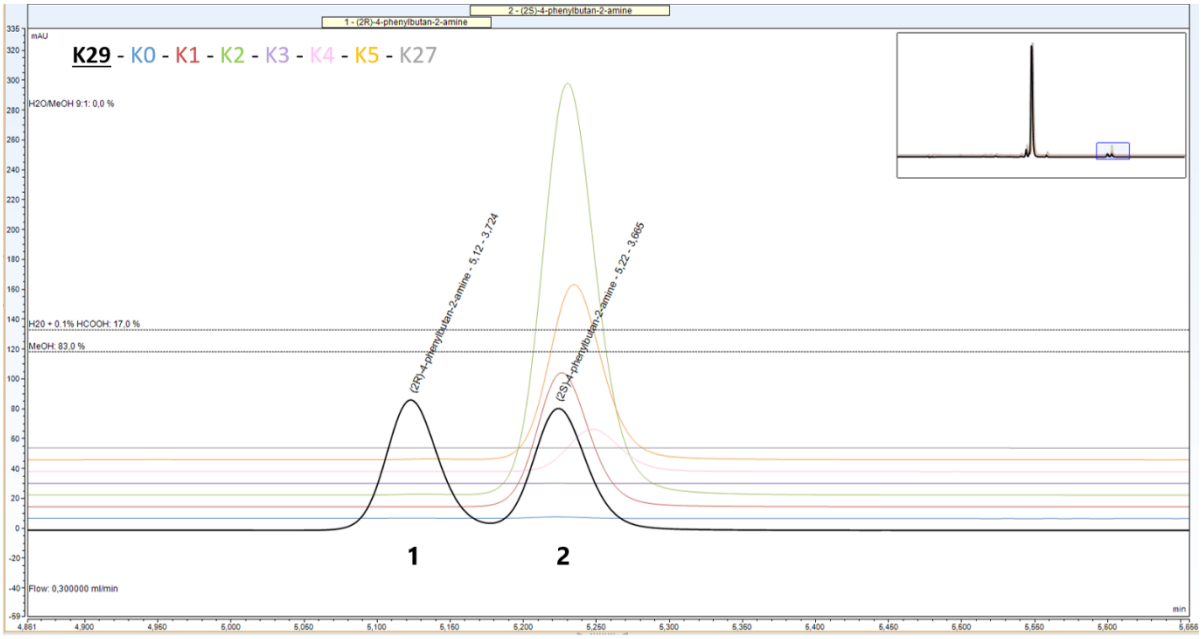
E



1: nonan-2-amine; 6.230; 0.621 **2**: nonan-2-one; 6.812; 0.293

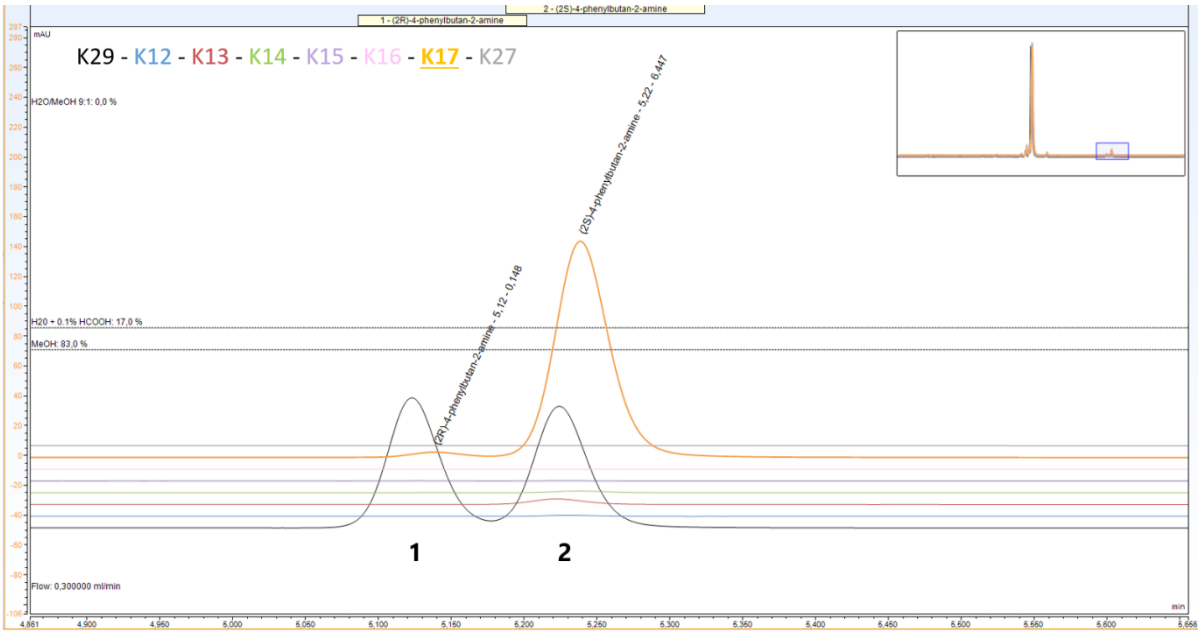
Appendix 35. (Chapter III, I.4.3) UHPLC-UV chromatograms for reaction of 4-phenylbutan-2-one (**39**) with various nat-AmDHs and mutants for conversion and *ee* estimation. UHPLC-UV chromatograms (340 nm) of a commercial standard of racemic amine (3 mM), blank reaction mixture without enzyme and reaction mixture with (A) *Cfus*AmDH, *Apau*AmDH and their mutants (B) MATOUAmDH2, *Micro*AmDH and their mutants (C) *Msme*AmDH, *Porti*AmDH and their mutants after derivatization with FDAA. The legend above matches the color of the chromatogram with the enzyme/substrate couple according to Appendix 22. The legend below indicates: compound; retention time (min); area.

A



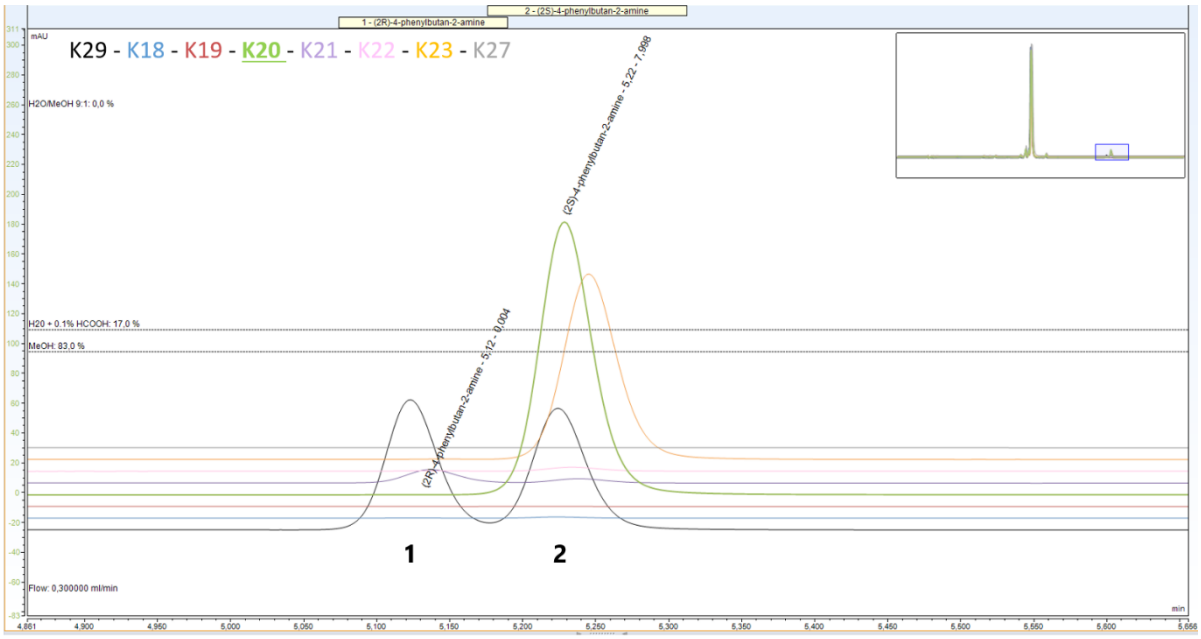
1: (2*R*)-4-phenylbutan-2-amine; 5.12; 3.724 **2**: (2*S*)-4-phenylbutan-2-amine; 5.22; 3.665

B



1: (2*R*)-4-phenylbutan-2-amine; 5.12; 0.148 **2**: (2*S*)-4-phenylbutan-2-amine; 5.22; 6.447

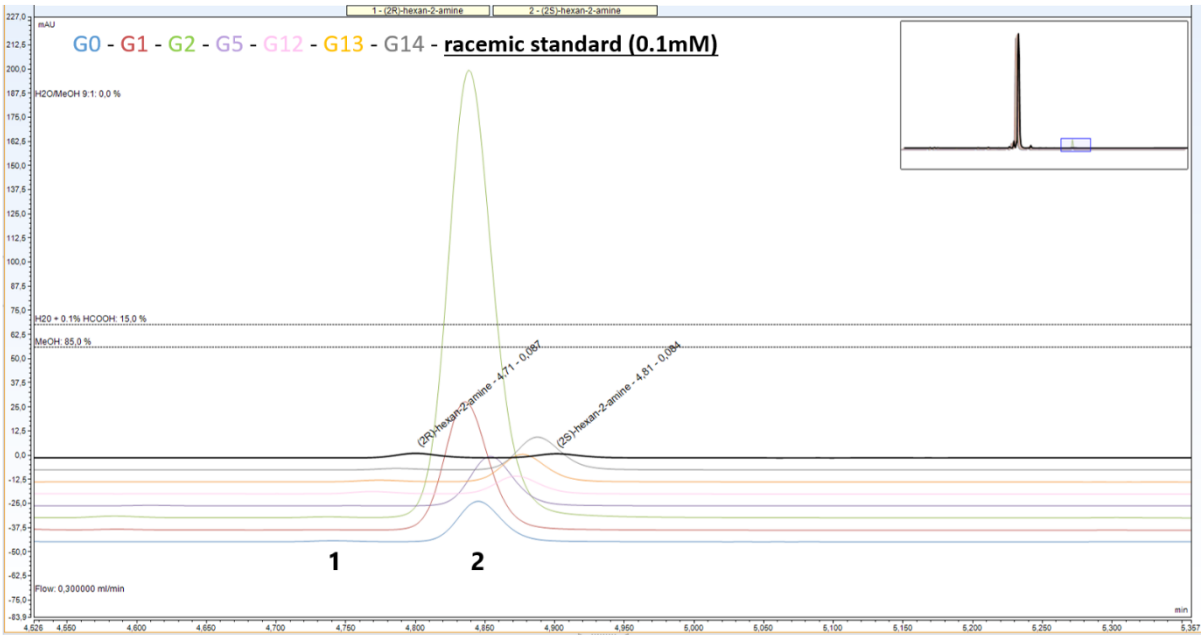
C



1: (2*R*)-4-phenylbutan-2-amine; 5.12; 0.004 **2**: (2*S*)-4-phenylbutan-2-amine; 5.22; 7.998

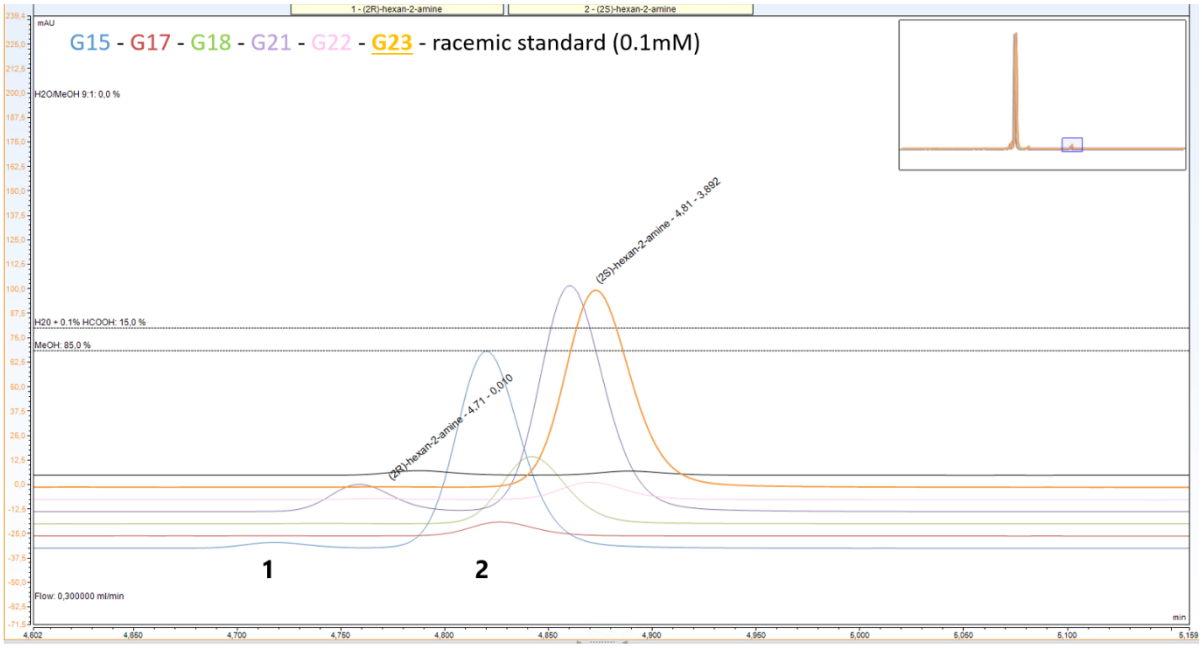
Appendix 36. (Chapter III, I.4.3) UHPLC-UV chromatograms for reaction of hexan-2-one (**43**) with various nat-AmDHs and mutants for *ee* estimation. UHPLC-UV chromatograms (340 nm) of a commercial standard of racemic amine (0.1 mM) and reaction mixture with (A) *Cfus*AmDH and mutants, *Apau*AmDH-W141A and MATOUAmDH2 and mutants, (B) *Micro*AmDH, *Micro*AmDH-W141A, *Msme*AmDH, *Porti*AmDH and mutants after derivatization with FDAA. The legend above matches the color of the chromatogram with the enzyme/substrate couple according to Appendix 22. The legend below indicates: compound; retention time (min); area.

A



1: (2*R*)-hexan-2-amine; 4.71; 0.087 **2**: (2*S*)-hexan-2-amine; 4.81; 0.84

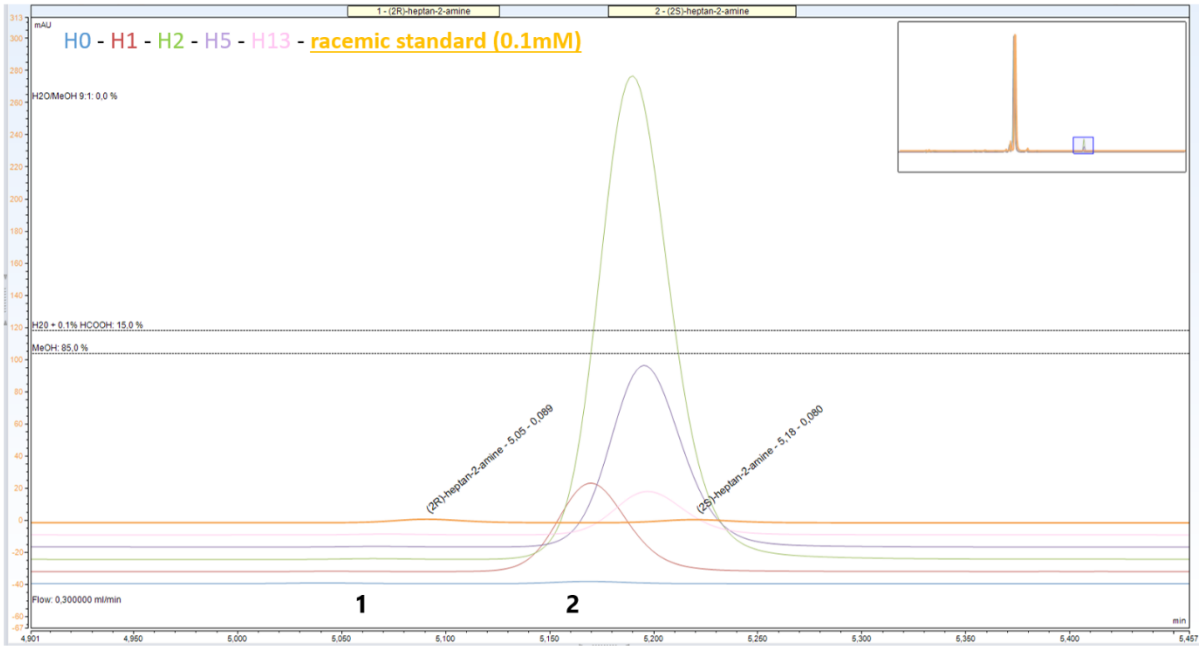
B



1: (2*R*)-hexan-2-amine; 4.71; 0.010 **2**: (2*S*)-hexan-2-amine; 4.81; 3.892

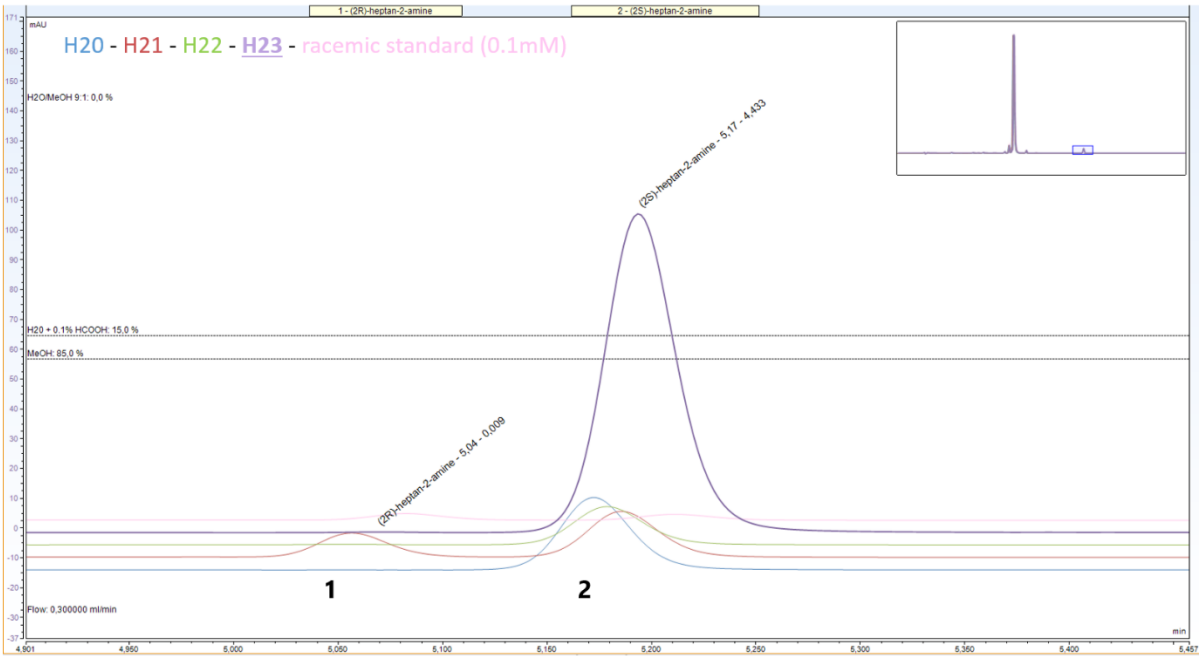
Appendix 37. (Chapter III, I.4.3) UHPLC-UV chromatograms for reaction of heptan-2-one (**128**) with various nat-AmDHs and mutants for *ee* estimation. UHPLC-UV chromatograms (340 nm) of a commercial standard of racemic amine (0.1 mM) and reaction mixture with (A) *Cfus*AmDH and mutants, *Apau*AmDH-W141A and MATOUAmDH2-F143A, (B) *Msme*AmDH-W141A, *Porti*AmDH and mutants after derivatization with FDAA. The legend above matches the color of the chromatogram with the enzyme/substrate couple according to Appendix 22. The legend below indicates: compound; retention time (min); area.

A



1: (2*R*)-heptan-2-amine; 5.05; 0.089 **2**: (2*S*)-heptan-2-amine; 5.18; 0.080

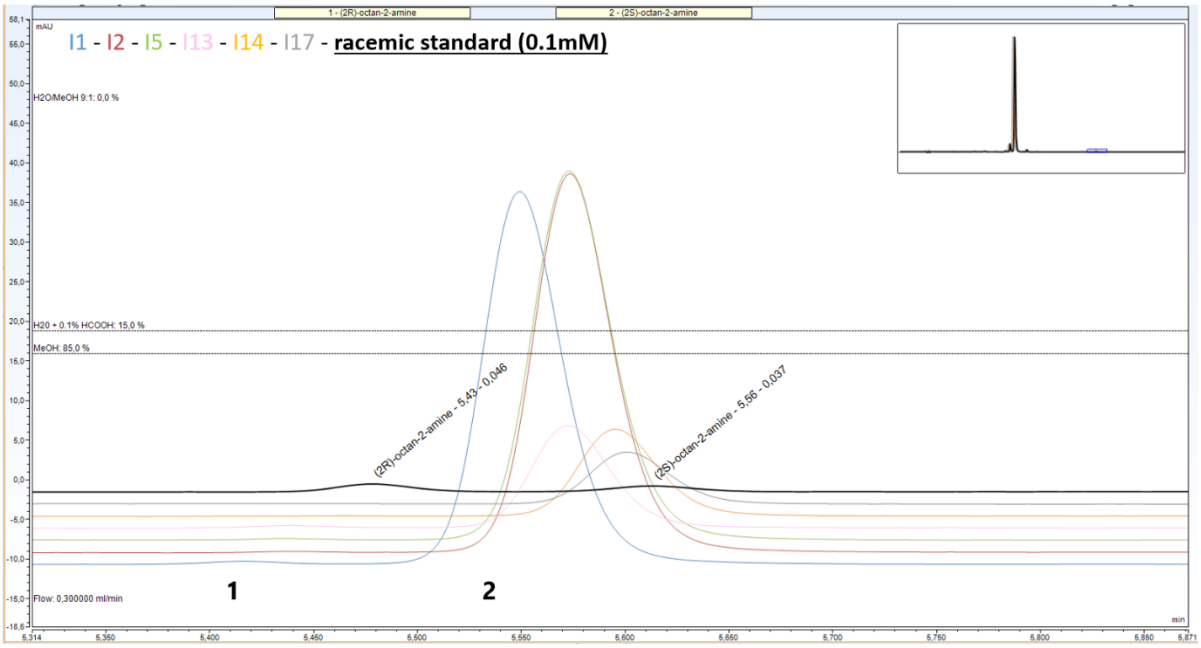
B



1: (2*R*)-heptan-2-amine; 5.04; 0.009 **2**: (2*S*)-heptan-2-amine; 5.17; 4.433

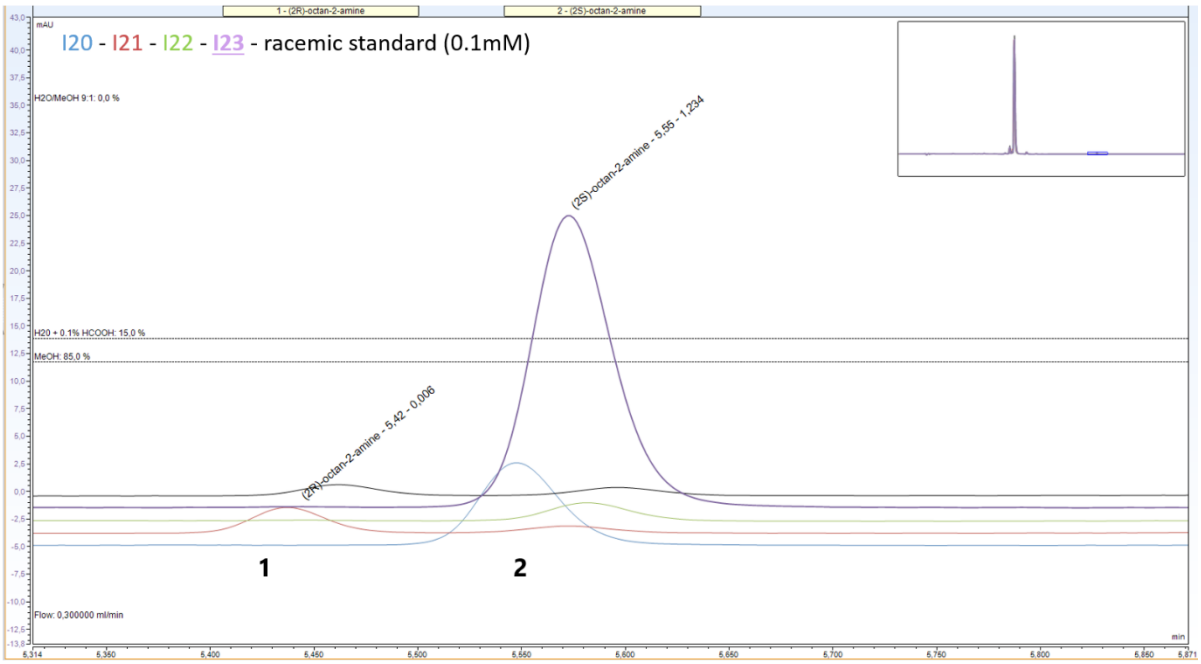
Appendix 38. (Chapter III, I.4.3) UHPLC-UV chromatograms for reaction of octan-2-one (**67**) with various nat-AmDHs and mutants for *ee* estimation. UHPLC-UV chromatograms (340 nm) of a commercial standard of racemic amine (0.1 mM) and reaction mixture with (A) *Cfus*AmDH-F140A, *Cfus*AmDH-W145A, *Apau*AmDH-W141A, MATOUAmDH2-F143A, MATOUAmDH2-C148A and *Micro*AmDH-W141A and (B) *Msme*AmDH-W141A, *Porti*AmDH and mutants after derivatization with FDAA. The legend above matches the color of the chromatogram with the enzyme/substrate couple according to Appendix 22. The legend below indicates: compound; retention time (min); area.

A



1: (2*R*)-octan-2-amine; 5.43; 0.046 **2**: (2*S*)-octan-2-amine; 5.56; .0,037

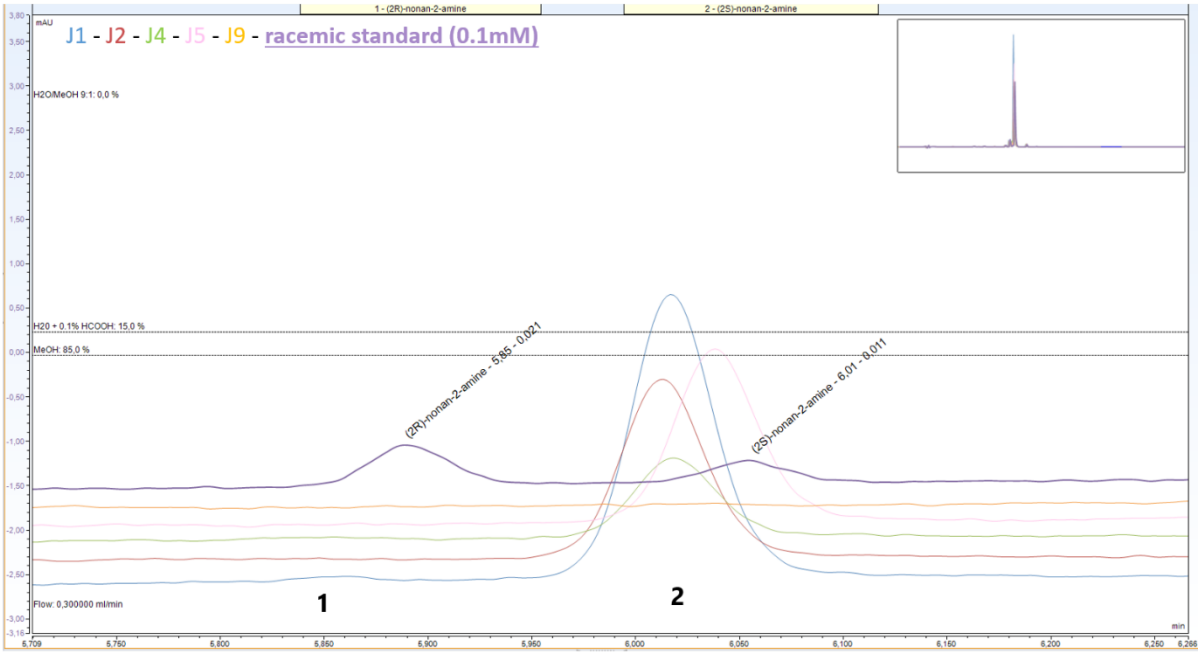
B



1: (2*R*)-octan-2-amine; 5.42; 0.006 **2**: (2*S*)-octan-2-amine; 5.55; 1.234

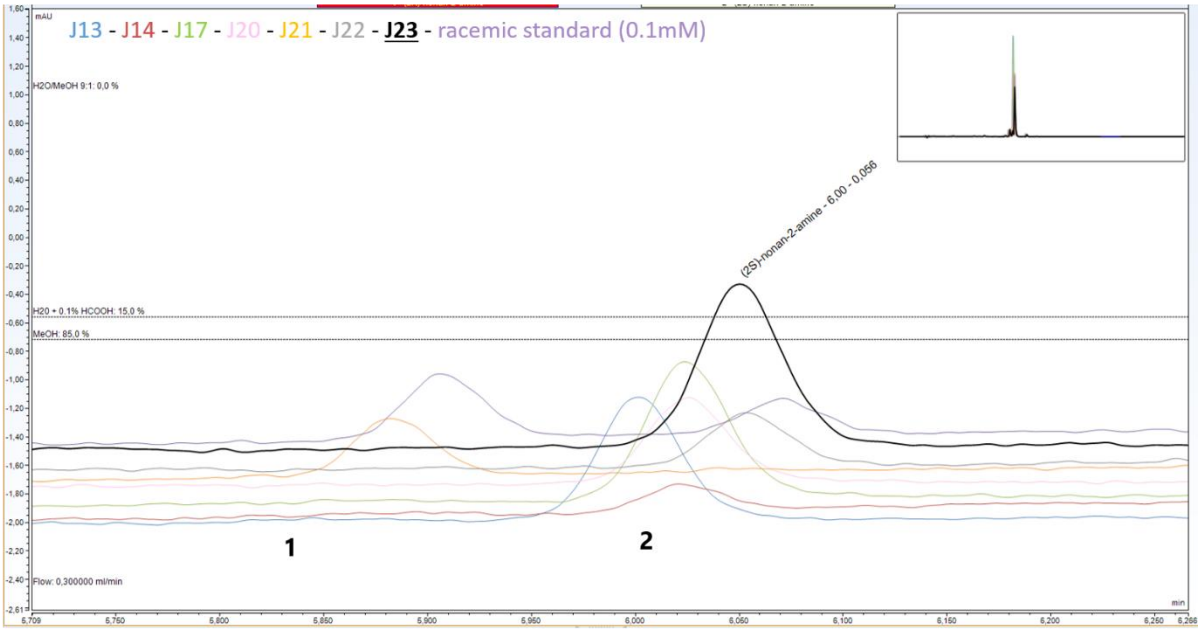
Appendix 39. (Chapter III, I.4.3) UHPLC-UV chromatograms for reaction of nonan-2-one (**135**) with various nat-AmDHs and mutants for *ee* estimation. UHPLC-UV chromatograms (340 nm) of a commercial standard of racemic amine (0.1 mM) and reaction mixture with (A) *Cfus*AmDH-F140A, *Cfus*AmDH-W145A, *Apau*AmDH-W141A, MATOUAmDH2-F143A, MATOUAmDH2-C148A and *Micro*AmDH-W141A and (B) *Msme*AmDH-W141A, *Porti*AmDH and mutants after derivatization with FDAA. The legend above matches the color of the chromatogram with the enzyme/substrate couple according to Appendix 22. The legend below indicates: compound; retention time (min); area.

A



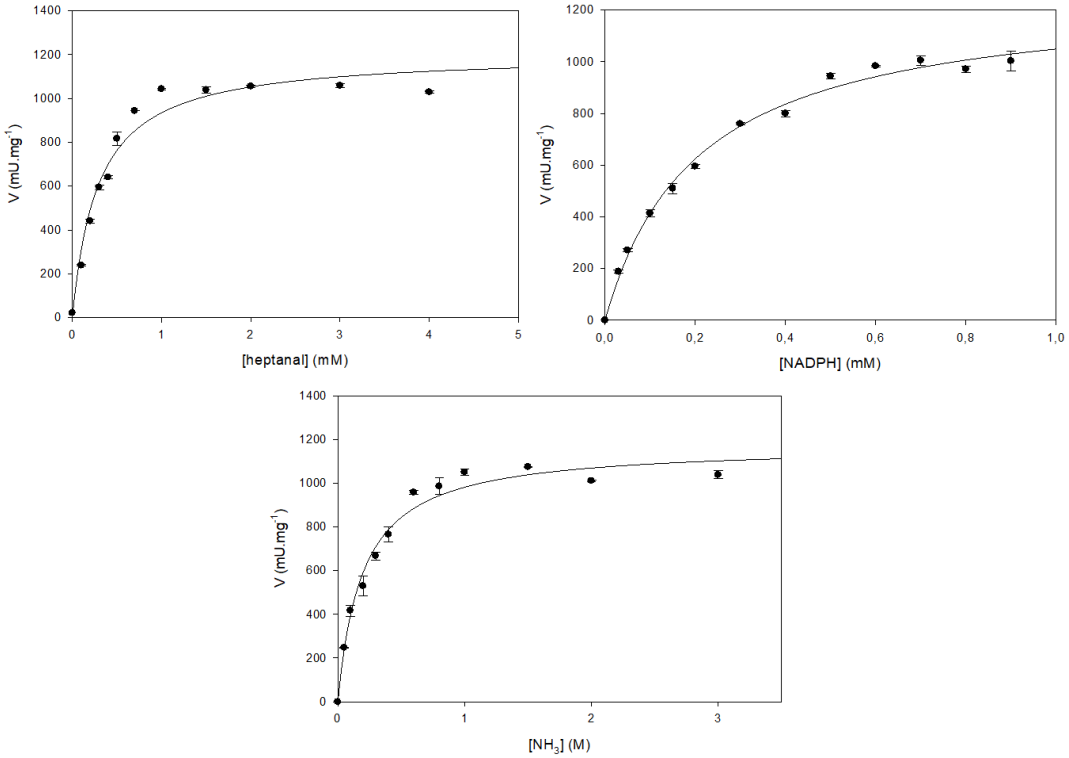
1: (2*R*)-nonan-2-amine; 5.85; 0.021 **2**: (2*S*)-nonan-2-amine; 6.01; 0.011

B

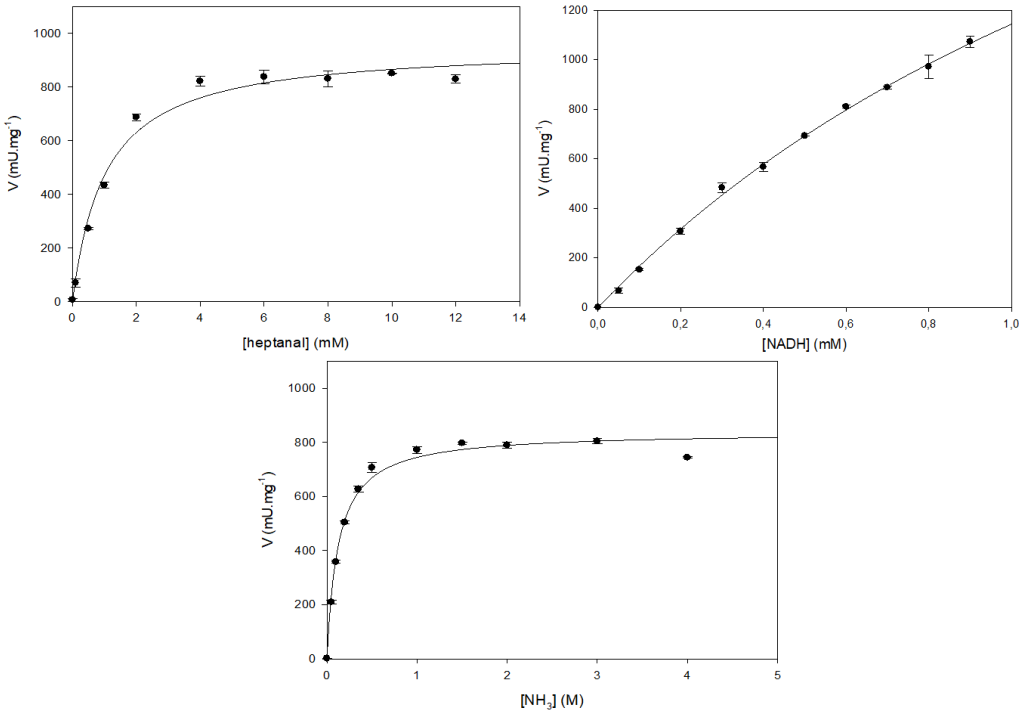


1: / **2**: (2*S*)-nonan-2-amine; 6.00; 0.056

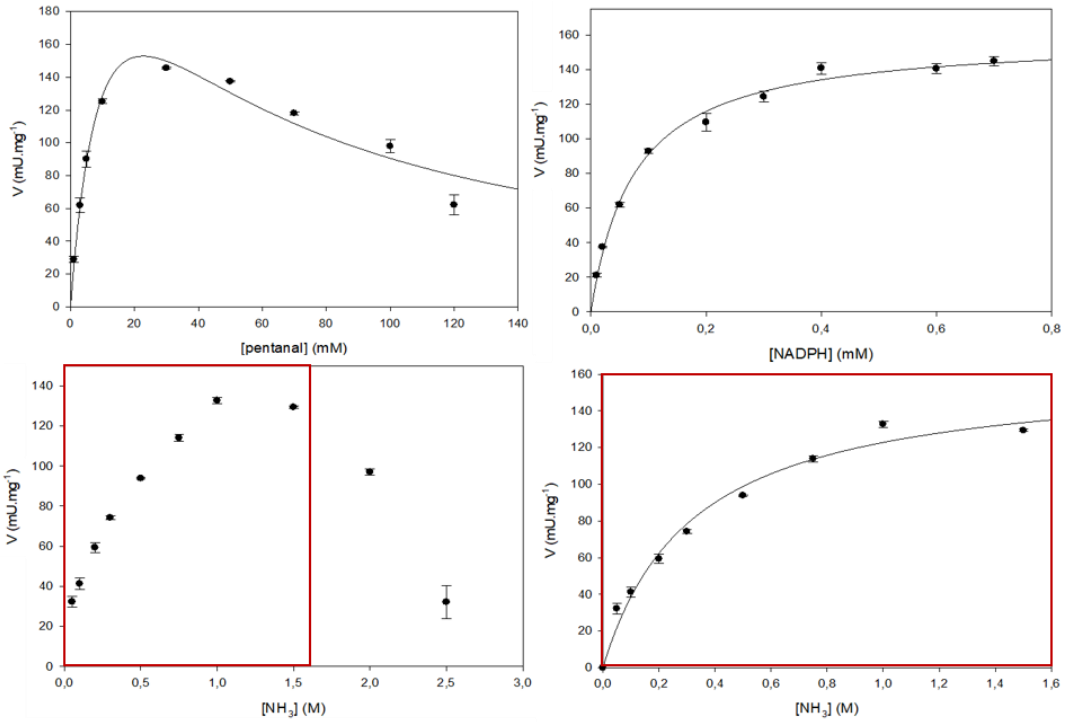
Appendix 40. (Chapter III, I.4.6.1) Michaelis-Menten plots for the determination of *Cfus*AmDH-W145A kinetic parameters in the system **217**/NADPH/NH_3. Error bars represent the standard deviation of two to three independent experiments. Saturating conditions: **217** 2 mM, NADPH 0.6 mM and ammonia (**11**) 2 M. The uncertainties are those generated out of two to three experiments.



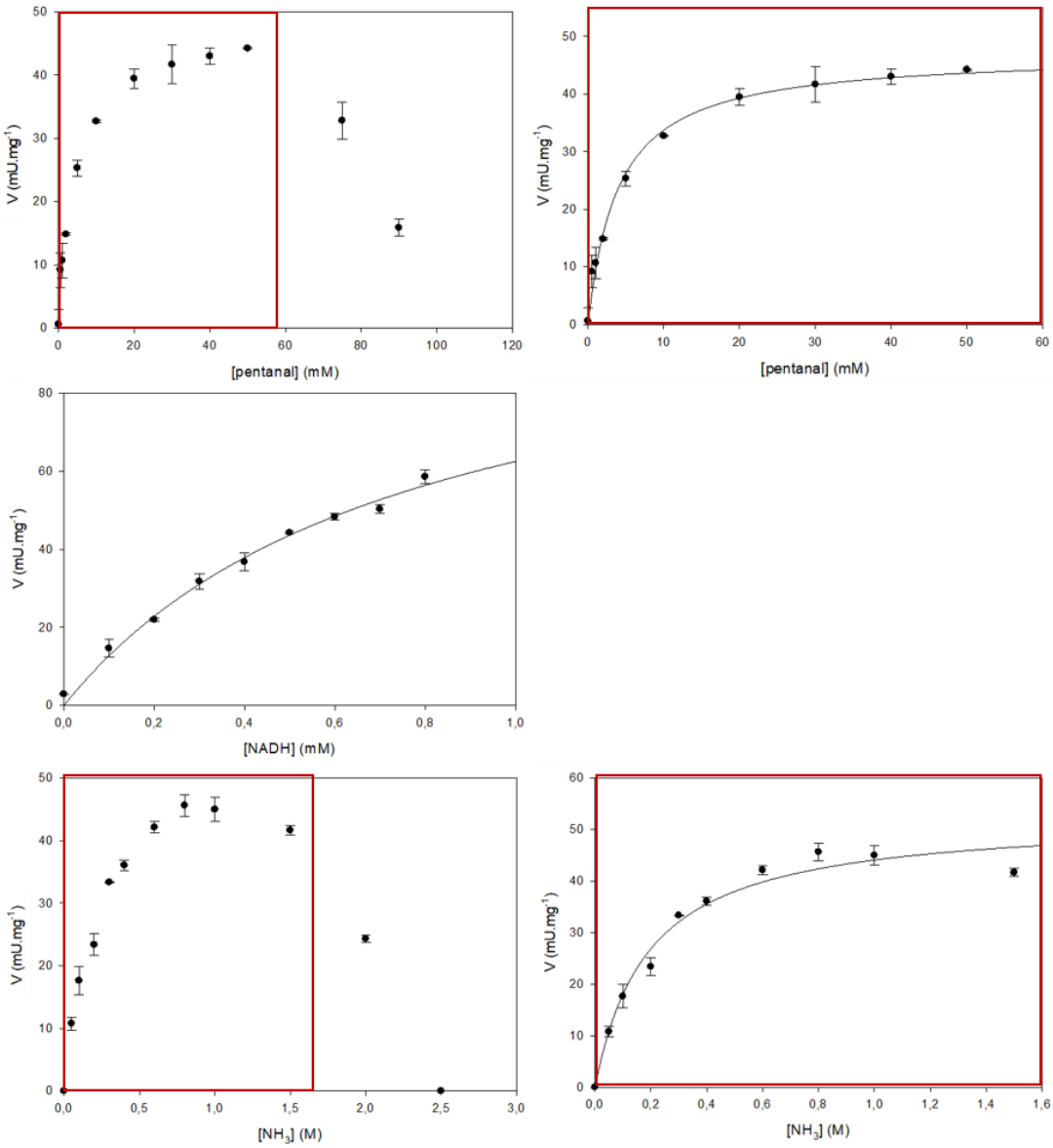
Appendix 41. (Chapter III, I.4.6.1) Michaelis-Menten plots for the determination of *Cfus*AmDH-W145A kinetic parameters in the system **217**/NADH/NH_3. Error bars represent the standard deviation of two to three independent experiments. Saturating conditions: **217** 10 mM, NADH 0.6 mM and ammonia (**11**) 2 M. The uncertainties are those generated out of two to three experiments.



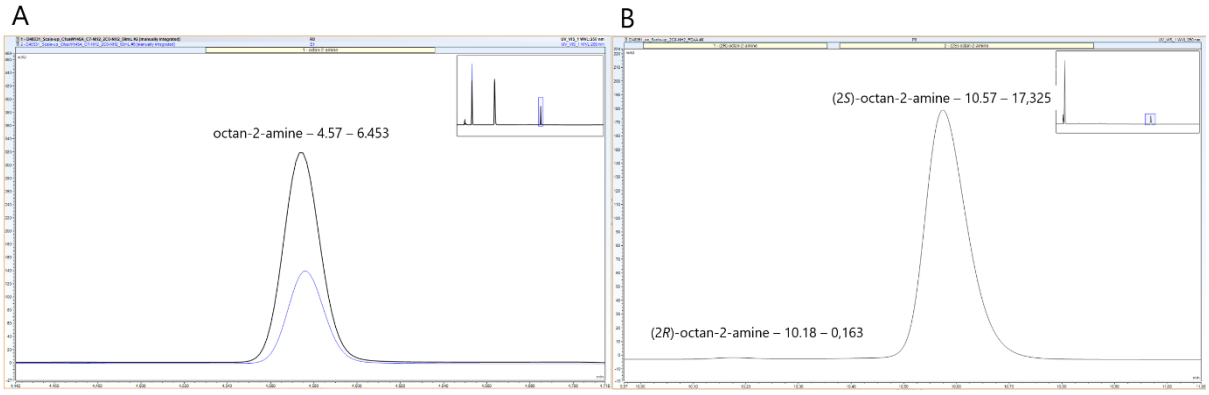
Appendix 42. (Chapter III, I.4.6.1) Michaelis-Menten plots for the determination of *Cfus*AmDH-W145A kinetic parameters in the system **175**/NADPH/NH_3. Error bars represent the standard deviation of two to three independent experiments. For NH_3 kinetic parameters, the curve could not fit an equation of inhibition by excess of substrate, therefore only the first plots have been used, indicated in a red frame, to find a correct fitting and determine an apparent K_M without taking the inhibition into account. Saturating conditions: **175** 30 mM, NADPH 0.6 mM and ammonia (**11**) 1 M. The uncertainties are those generated out of two to three experiments.



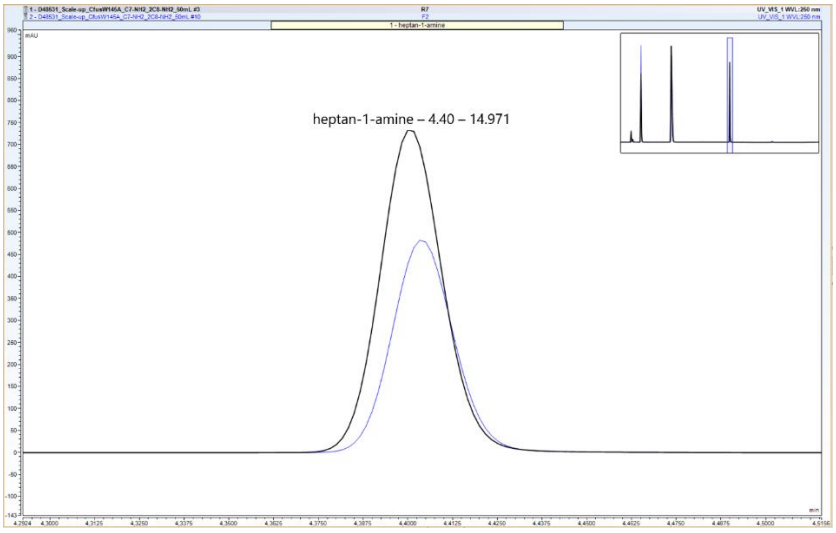
Appendix 43. (Chapter III, I.4.6.1) Michaelis-Menten plots for the determination of *Cfus*AmDH-W145A kinetic parameters in the system **175**/NADH/NH_3. Error bars represent the standard deviation of two to three independent experiments. For NH_3 and **175** kinetic parameters, the curves could not fit an equation of inhibition by excess of substrate, therefore only the first plots, indicated in red frames, have been used to find a correct fitting and determine an apparent K_M without taking the inhibition into account. Saturating conditions: **175** 30 mM, NADH 0.6 mM and ammonia (**11**) 1 M. The uncertainties are those generated out of two to three experiments.



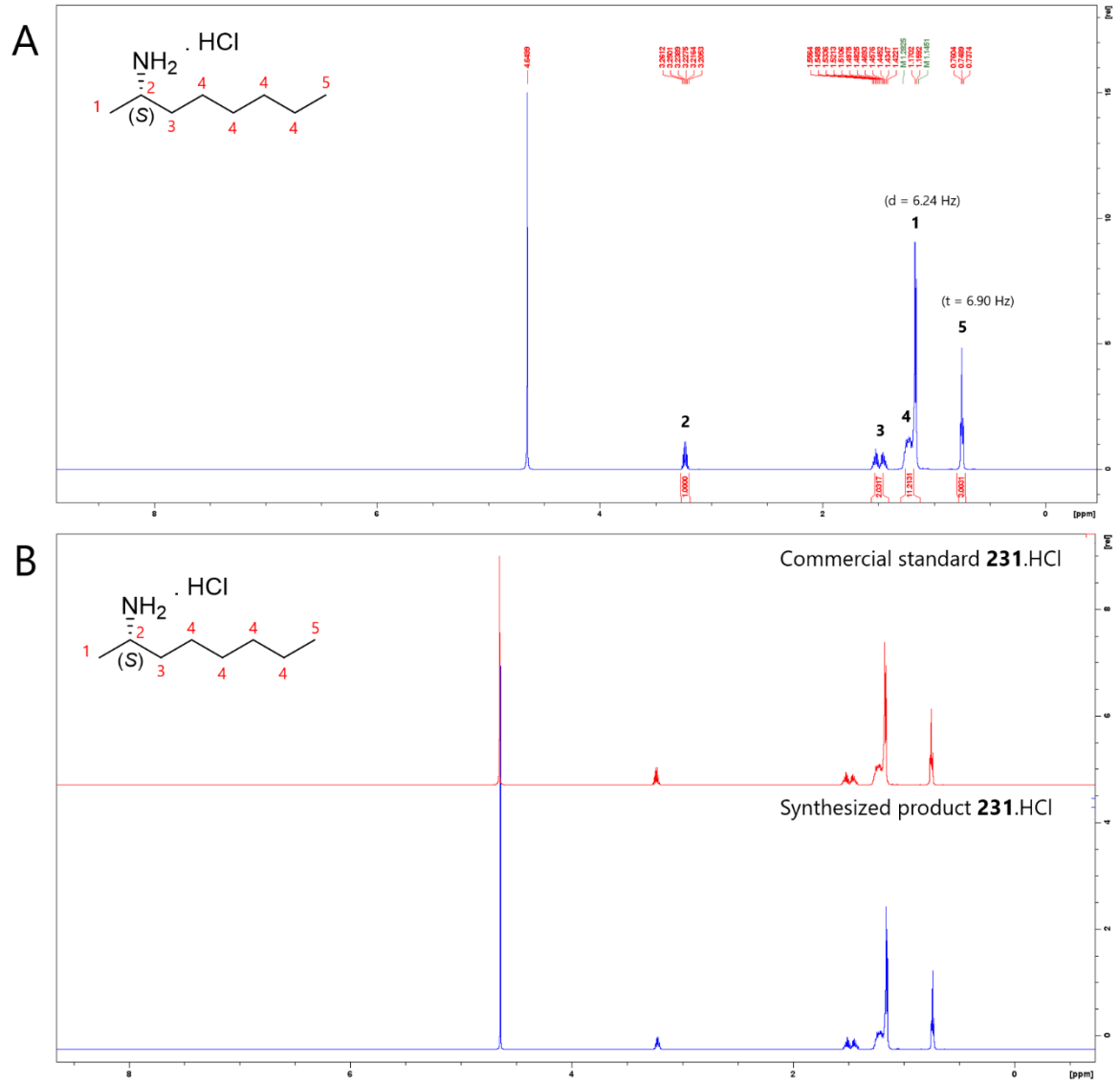
Appendix 44. (Chapter III, I.4.6.4) UHPLC-UV chromatograms chromatograms of (2*S*)-octan-2-amine (**231**) recovered from the semi-preparative scale reaction with *Cfus*AmDH-W145A. (A) UHPLC-UV chromatograms of standard expected amine **231** (5 mM) in blue and reaction mixture in black after derivatization with BzCl. (B) UHPLC-UV chromatograms (340 nm) of reaction mixture after derivatization with FDAA. The labelling corresponds to the retention time (min) and peak area.



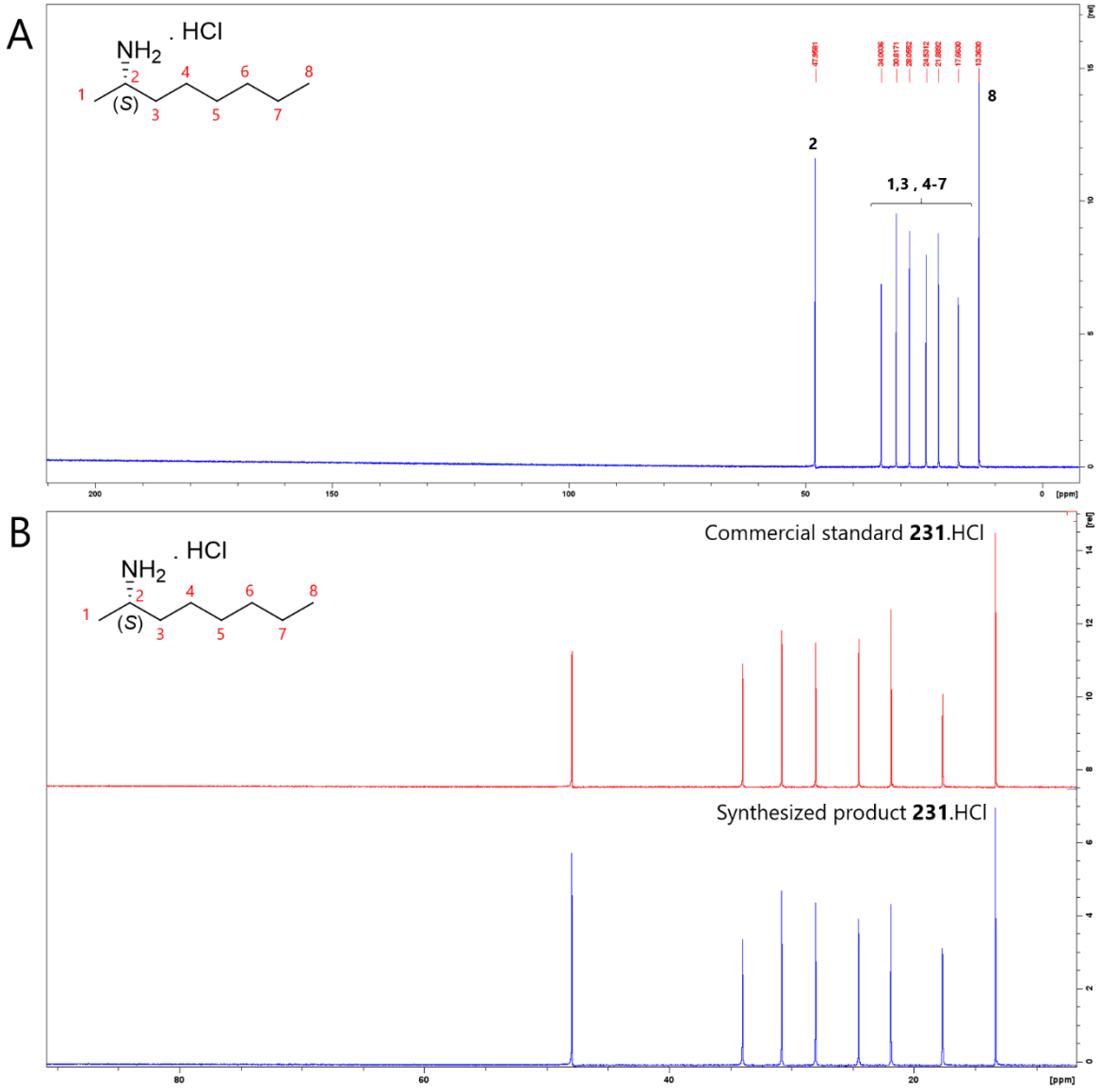
Appendix 45. (Chapter III, I.4.6.4) UHPLC-UV chromatograms of heptan-1-amine (**213**) recovered from the semi-preparative scale reaction with *Cfus*AmDH-W145A. UHPLC-UV chromatograms of standard expected amine (10 mM) in blue and reaction mixture in black after derivatization with BzCl. The labelling corresponds to the retention time (min) and peak area.



Appendix 46. (Chapter III, I.4.6.4) ^{1}H NMR spectrum of **231** recovered from the preparative scale reaction with *Cfus*AmDH-W145A. (A) ^{1}H NMR spectrum of synthesized **231** as monohydrochloride salt in D_2O. (B) Comparison of ^{1}H NMR spectra of commercial **231** as monohydrochloride salt and synthesized **231** as monohydrochloride salt in D_2O. ^{1}H NMR (600MHz) δ 3.22 (m, 1H), 1.48 (m, 2H), 1.27-1.14 (m, 8H; d, J = 6.24 Hz, 3H), 0.74 (t, J = 6.90 Hz, 3H).



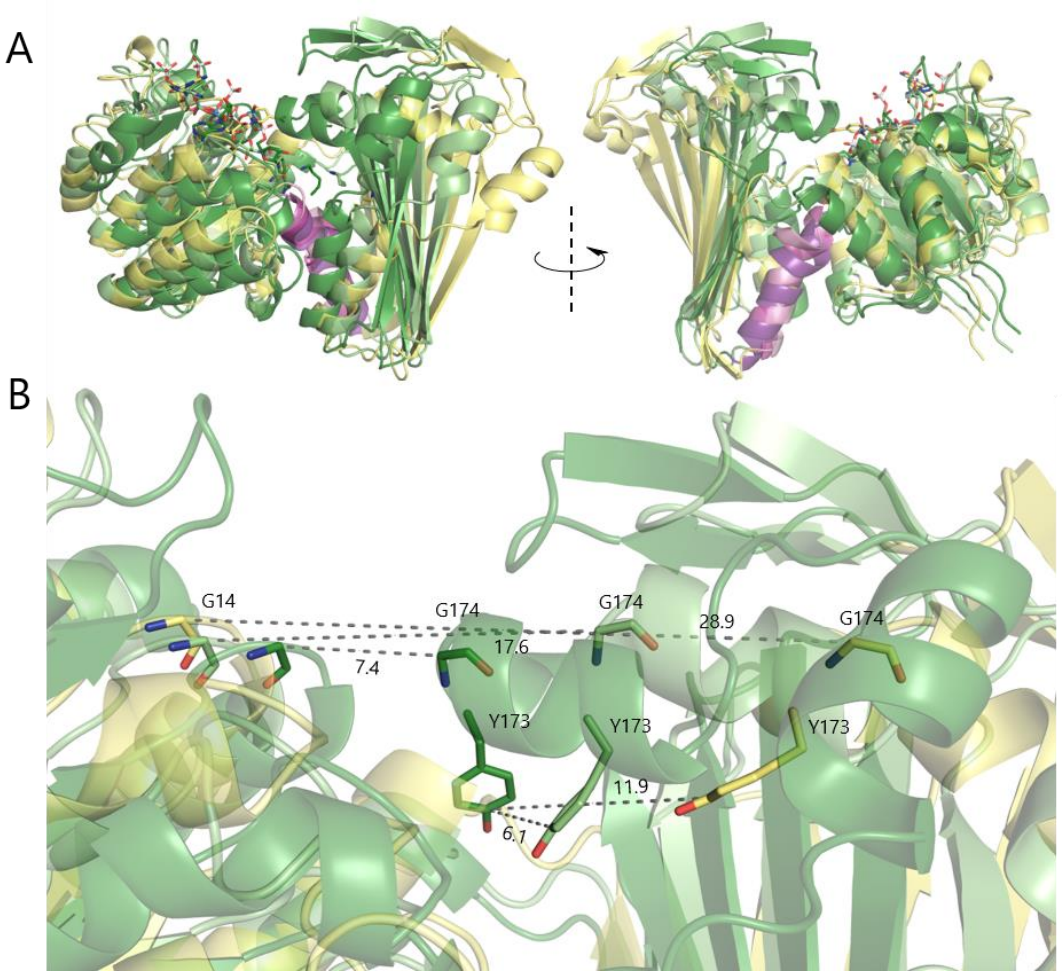
Appendix 47. (Chapter III, I.4.6.4) ^{13}C NMR spectrum of **231** recovered from the preparative scale reaction with *Cfus*AmDH-W145A. (A) ^{13}C NMR spectrum of synthesized **231** as monohydrochloride salt in D_2O. (B) Comparison of ^{13}C NMR spectra of commercial **231** as monohydrochloride salt and synthesized **231** as monohydrochloride salt in D_2O. ^{13}C NMR (150 MHz, D_2O) δ 47.9, 34.0, 30.8, 28.0, 24.5, 21.9, 17.7, 13.4.



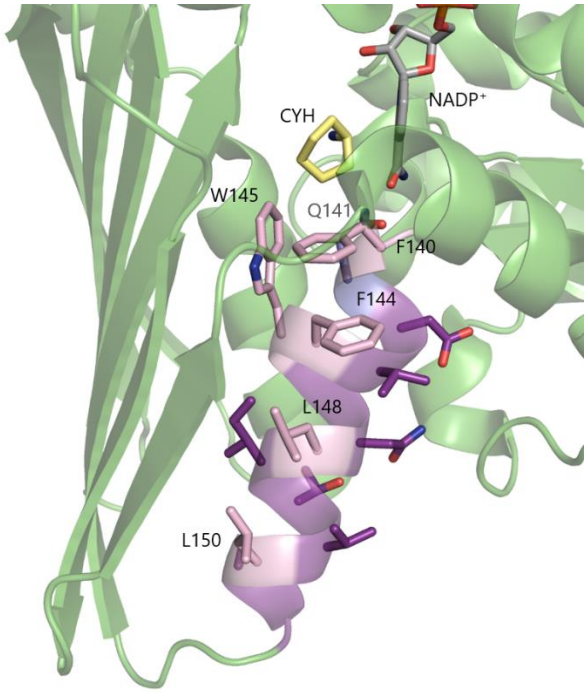
Appendix 48. (Chapter III, **Erreur ! Source du renvoi introuvable.**) Data Collection and Refinement Statistics for *Cfus*AmDH-W145A in complex with NAD^+ and $NADP^+$ with pentan-1-amine (**211**). Numbers in brackets refer to data for highest resolution shells.

	*Cfus*AmDH-W145A with NAD^+	*Cfus*AmDH-W145A with $NADP^+$ and **211**
Beamline	Diamond I03	Diamond I03
Wavelength (Å)	0.976261	0.976284
Resolution (Å)	48.05-1.64 (1.67-1.64)	47.36-1.50 (1.53-1.50)
Space Group	$P2_12_12_1$	$C222_1$
Unit cell (Å)	a = 51.02; b = 86.23; c = 142.94	a = 56.69; b = 84.31; c = 284.13.
No. of molecules in asu	2	2
Unique reflections	78152 (3786)	109306 (5350)
Completeness (%)	100.0 (100.0)	100.0 (100.0)
R_{merge} (%)	0.07 (1.30)	0.08 (0.96)
$R_{p.i.m.}$	0.03 (0.54)	0.04 (0.40)
Multiplicity	11.8 (12.8)	12.2 (13.0)
$<I/\sigma(I)>$	19.1 (2.0)	17.1 (3.1)
Overall *B* factor from Wilson plot ($Å^2$)	22	17
$CC_{1/2}$	1.00 (0.73)	1.00 (0.90)
R_{cryst}/ R_{free} (%)	17.8/21.4	16.4/18.8
r.m.s.d 1-2 bonds (Å)	0.011	0.012
r.m.s.d 1-3 angles (°)	1.71	1.74
Avge main chain B ($Å^2$)	27	20
Avge side chain B ($Å^2$)	30	23
Avge water B ($Å^2$)	38	31
Avge cofactor B ($Å^2$)	42	15
Avge Ligand B ($Å^2$)	-	30

Appendix 49. (Chapter III, I.4.7.2.1) Open-to-closed event observed in *Cfus*AmDH-W145A MD simulations. (A) 3D representation at 180° of *Cfus*AmDH monomer in complex with $NADP^+$ and **211** (M_NADP_PEN_1). Superimposition of three discrete conformations along the trajectory: the closed one in dark green at 0 ns, the medium "relaxed" conformation in light green at 19.2 ns and the fully open monomer in yellow at 21.5 ns. The spine helix F140-G154 discussed in this work is highlighted in purple, dark pink and light pink, respectively. (B) Close-view on opening loops of M_NADP_PEN_1. The translation motion is highlighted by residues G14, Y173, and G174 shown as sticks. Distance G14-G174 and Y173 with itself in the three frames are given in Å.

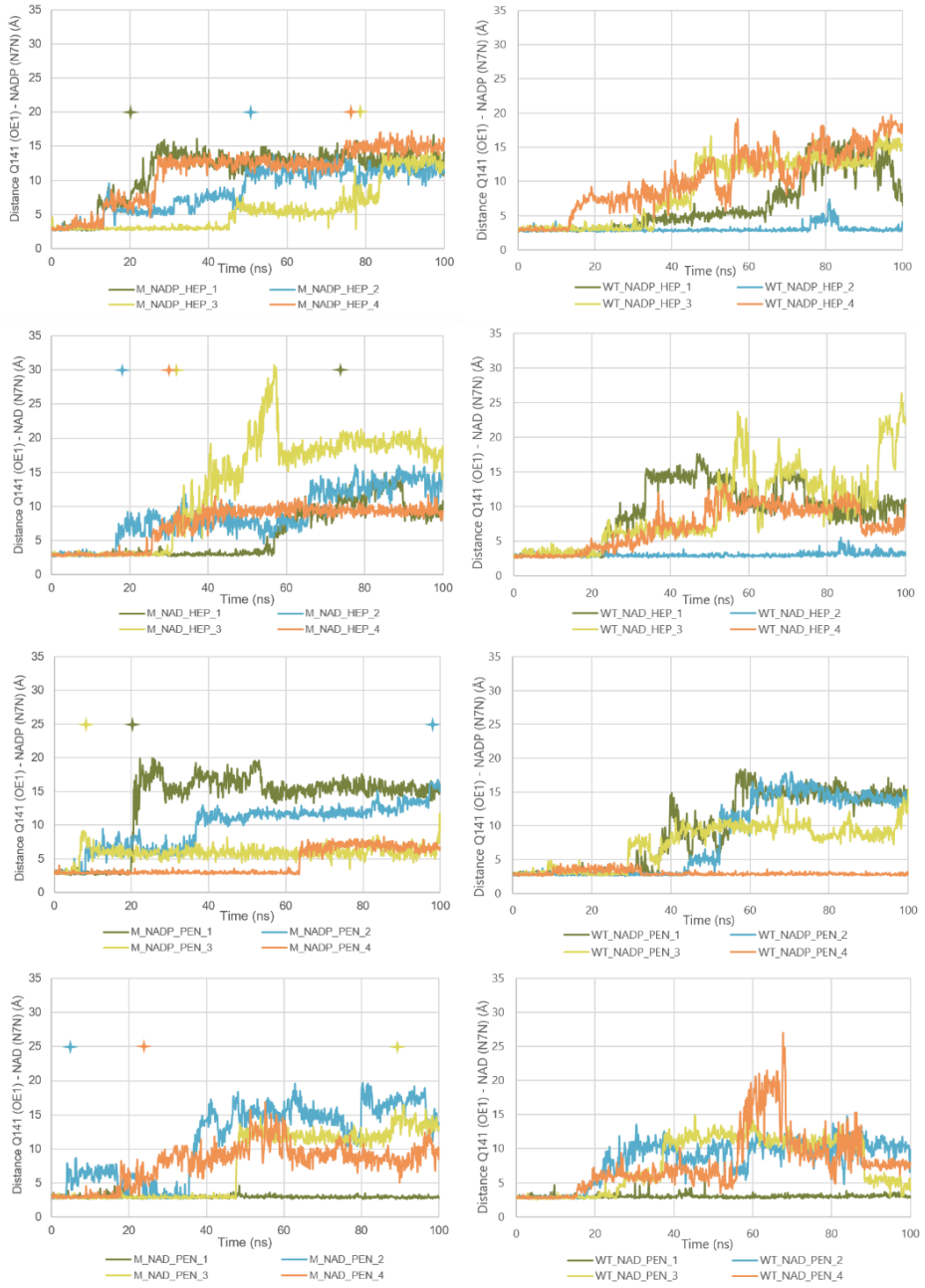


Appendix 50. (Chapter III, I.4.7.2.3) 3D representation of F140-G154 spine helix in *Cfus*AmDH (PDB: 6IAU). The residues of the spine helix are highlighted in purple and the one discussed in this part (F140, F144, W145, L148 and L150) in pink. $NADP^+$ and cylochexanone (**45**) ligands are represented in grey and yellow, respectively.



Appendix 51. (Chapter III, I.4.7.2.3) Distance between Q141 and cofactor along the time of the simulation. (A) Distance between Q141 (OE1) and NAD(P)$^+$ (N7N) over the course of the four independent simulations in green, blue, yellow and orange. The stars indicate the conformational switch of Q141 in M systems as shown in Figure 54. Distances are given in Å and time in ns. (B) Estimated time between the loss of the interaction between Q141 and the NAD(P)$^+$ and the conformational switch of Q141. The time between the hydrogen bond breakage and the conformational switch of Q141 can vary from an almost immediate event, such as 0.1 ns in M_NAD_PEN, to longer time frames such as 89.3 ns in the simulation with M_NADP_PEN_4, which is mainly due to steric hindrance brought about by HEP or PEN ligand. * Arbitrarily set at 100 ns but the conformational switch never took place in this simulation. *n.d.* not determined.

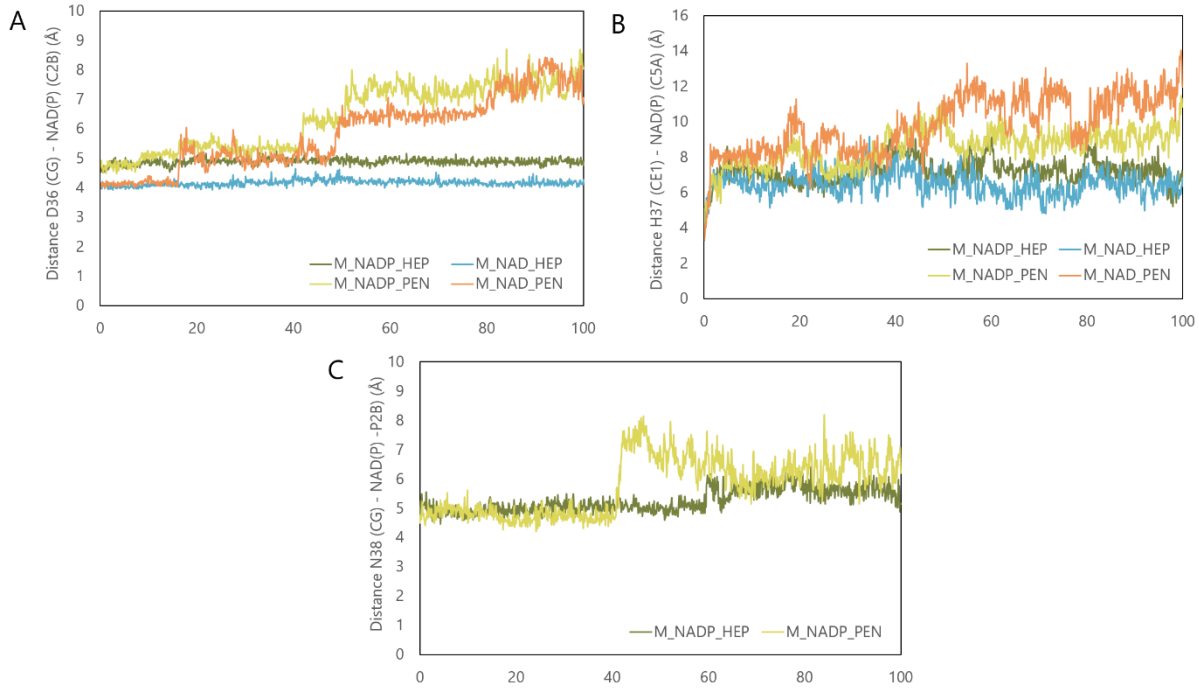
A



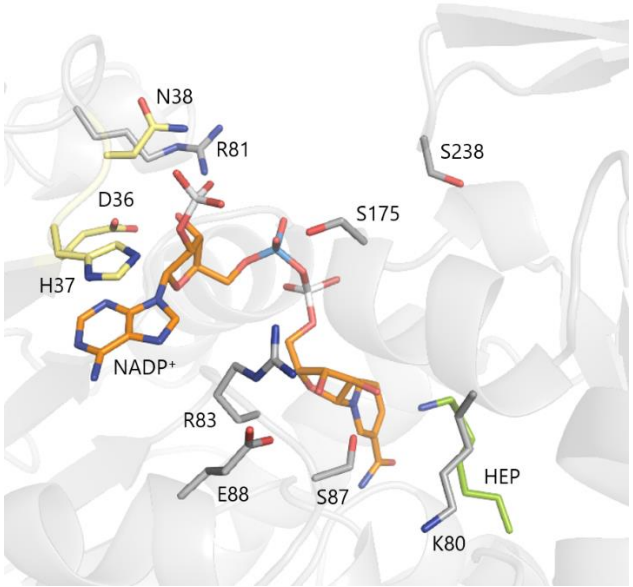
B

Amine	Cofactor	Simulation 1	2	3	4	Average
HEP	NADP	8.4	36.9	32.3	62.5	35.0
HEP	NAD	16.4	2.2	0.3	4.5	5.9
PEN	NADP	0.8	89.3	64.5	100*	63.7
PEN	NAD	*n.d.*	0.1	41.4	5.3	15.6

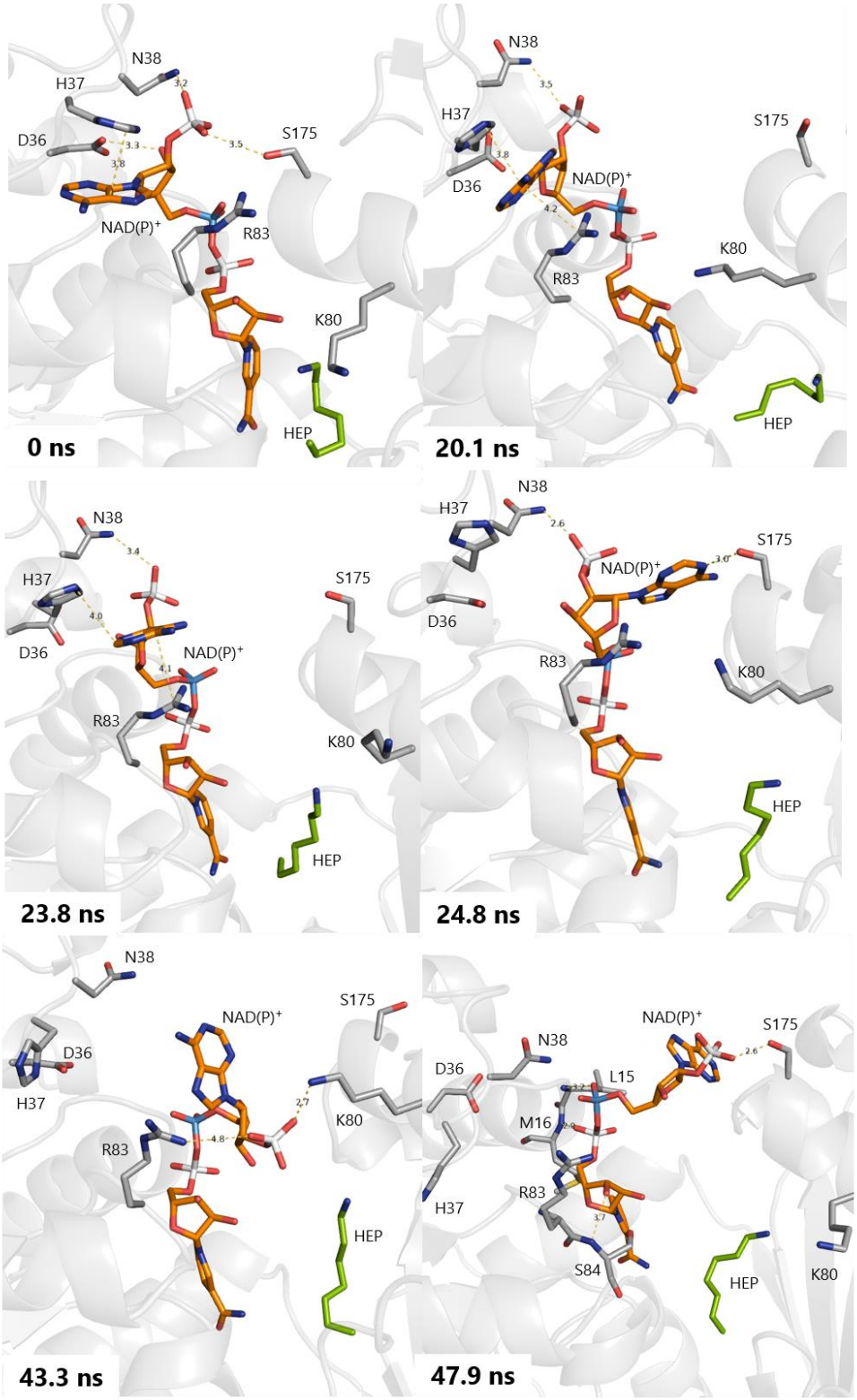
Appendix 52. (Chapter III, I.4.7.2.4) Distance between the adenosine moiety of the cofactor and its binding site. Distance between (A) H37 (CE1) and NA(D)P (Adenine-C5A) (B) D36 (CG) and NA(D)P$^+$ (ribose-C2B) and (C) N38 (CG) and NADP$^+$ (PO3-P2B). Distances are given in Å and time in ns.



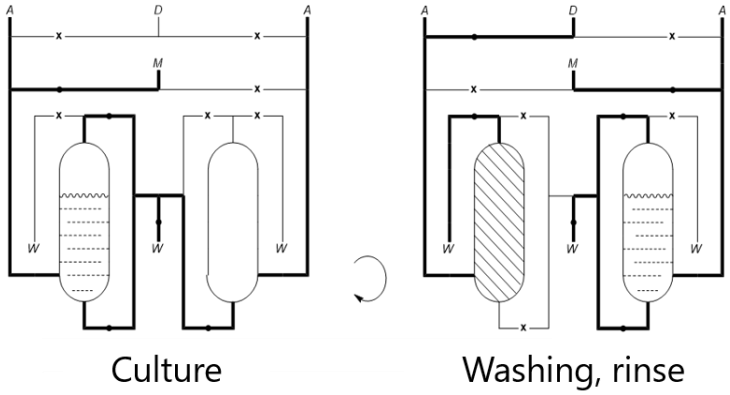
Appendix 53. (Chapter III, I.4.7.2.4) 3D representation of the adenosine binding site of NADP and residues implied in its mobility. The three main residues at t=0 ns are shown in yellow while the other residue that can interact with one part of the adenosine during the simulation are shown in grey.



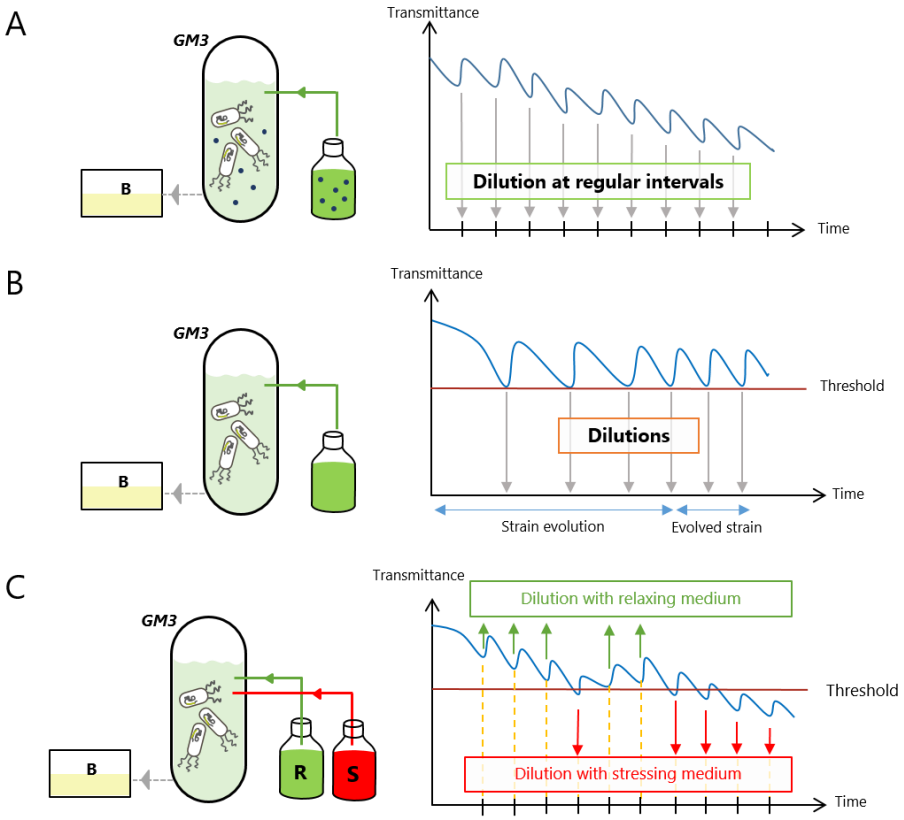
Appendix 54. (Chapter III, I.4.7.2.4) Close-view of NADP adenosine interactions with protein residues along the simulation (WT_NADP_HEP_1). NADP and HEP ligands are shown in orange and green, respectively. Distances are given in Å.



Appendix 55. (Chapter III, II.1) GM3 culture vessels during the culture phase and the whashing/rinse phase.



Appendix 56. (Chapter III, II.1) Principle of cultivation in (A) chemostat regime, (B) turbidostat regime and (C) medium swap regime. B = bin; R = relaxing medium; S = stressing medium. Adapted from Mayol, 2019.

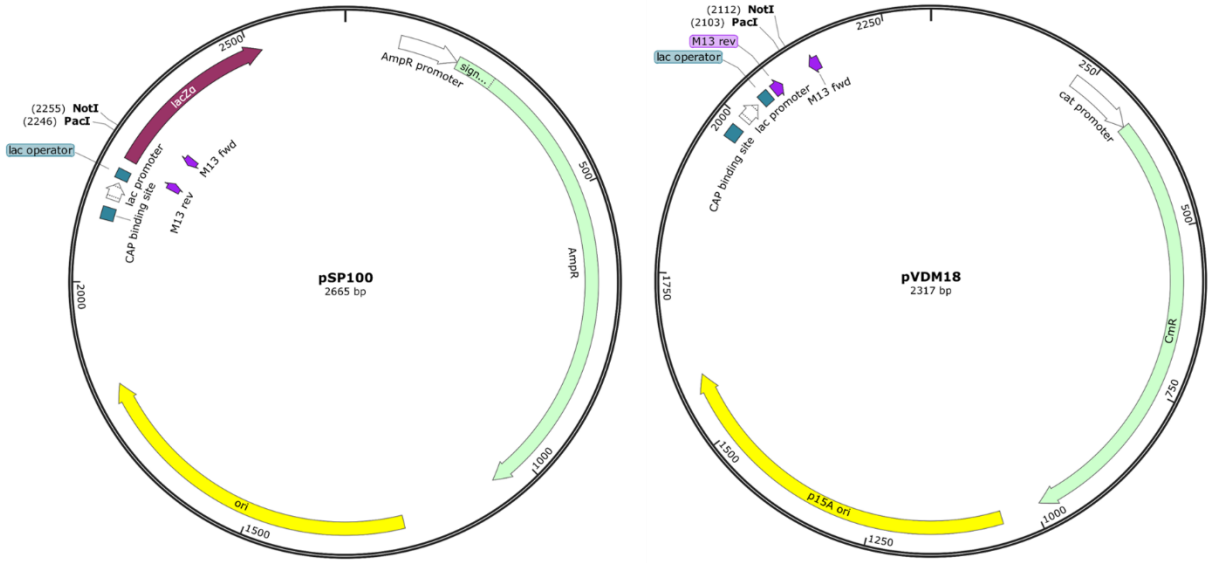


Appendix 57. (Chapter III, II.2 and II.3) *E. coli* strains and constructions used for the *in vivo* evolution of nat-AmDH/RedAm activity.

Strain	Genotype	Insertion
G4480	E. coli K12 C> Δ puuE ::FRT+ Δ gabT:: FRT+ ΔdadA :: FRT+	Ø
G4265	E. coli K12 C> Δ puuE ::FRT+ Δ gabT::kan+ FRT+ ΔdadA :: cat+ FRT+ P> pGEN1121 bla+ AmDH5 (*Clostridium sticklandii*)+	AmDH5 - pSP100
G5471	E. coli K12 C> DpuuE ::FRT+ Δ gabT::FRT+ ΔdadA :: FRT+ P> pGEN1121 bla+ AmDH5 (*Clostridium sticklandii*)+	AmDH5 - pSP100
G4488	E. coli K12 C> Δ puuE ::FRT+ Δ gabT:: FRT+ ΔdadA :: FRT+ ΔkdgK :: AmDH5 (*Clostridium sticklandii*)+ kan+	AmDH5 - chromosome
G4489	E. coli K12 C> Δ puuE ::FRT+ Δ gabT:: FRT+ DdadA :: FRT+ ΔkdgK :: AmDH5 (*Clostridium sticklandii*)+ kan+	AmDH5 - chromosome
G5376	E. coli K12 C> Δ puuE ::FRT+ Δ gabT::kan+ FRT+ ΔdadA :: FRT+ P> pGEN1285 cat+ AmDH5 (*Clostridium sticklandii*)+	AmDH5 - pVDM18
G5377	E. coli K12 C> Δ puuE ::FRT+ ΔgabT:: FRT+ ΔdadA :: FRT+ P>pGEN1286 cat+ OA (*Clostridium sticklandii*)+	OA - pVDM18
G5378	E. coli K12 C> Δ puuE ::FRT+ ΔgabT:: FRT+ ΔdadA :: FRT+ P>pGEN1287 cat+ *Tther*AmDH (*Thermohydrosulfuricus* WC1)+	*Tther*AmDH - pVDM18
G5379	E. coli K12 C> Δ puuE ::FRT+ ΔgabT:: FRT+ ΔdadA :: FRT+ P>pGEN1121 bla+ AmDH5 (*Clostridium sticklandii*)+ pGEN1286 cat+ OA (Clostridium sticklandii)+	AmDH5 - pSP100 OA -pVDM18
G5393	E. coli K12 C> Δ puuE ::FRT+ ΔgabT:: FRT+ ΔdadA :: FRT+ P>pGEN1293 bla+ *Tther*AmDH (*Thermohydrosulfuricus* WC1)+	TtherAmDH - pSP100
G5392	E. coli K12 C> Δ puuE ::FRT+ ΔgabT:: FRT+ ΔdadA :: FRT+ P>pGEN1292 bla+ OA (*Clostridium sticklandii*)+	OA - pSP100
G5394	E. coli K12 C> Δ puuE ::FRT+ ΔgabT:: FRT+ ΔdadA :: FRT+ P>pGEN1285 cat+ AmDH5 (*Clostridium sticklandii*)+ pGEN1292 bla+ OA (Clostridium sticklandii)+	DH2B8 - pVDM18 OA - pSP100
G5422	E.coli K12 C> ΔkdgK :: AmDH5 (*Clostridium sticklandii*)+ kan+ P>pORTMAGE-2 bla+ mutL+	AmDH5 - chromosome
G5447	E.coli K12 C> ΔkdgK :: AmDH5* (*Clostridium sticklandii*)+ kan+	AmDH5 - chromosome
G5488	E.coli K12 C>b2033 P>pGEN1302 bla+ OA(*Clostridium sticklandii*)+ AmDH5(*Clostridium sticklandii*)+	AmDH5+OA - pSP100
G5469	E. coli K12 C> DpuuE ::FRT+ Δ gabT:: FRT+ ΔdadA :: FRT+ ΔkdgK :: AmDH5* (*Clostridium sticklandii*)+ kan+	AmDH5 - chromosome with mutation in kdgK promoter

G5492	E. coli K12 C>DpuuE ::FRT+ DgabT:: FRT+ DdadA :: FRT+ P>pGEN1302 bla+ OA(Clostridium sticklandii)+ DH2B8(Clostridium sticklandii)+	AmDH5 and OA – pSP100
G5490	E. coli K12 C>DpuuE ::FRT+ DgabT:: FRT+ DdadA :: FRT+ DkdgK :: DH2B8* (Clostridium sticklandii)+ kan+ P>pGEN1292 bla+ OA(Clostridium sticklandii)+	AmDH5 – chromosome + mutation in kdgK promotor OA – pSP100
G5618	E. coli K12 C> ΔpuuE ::FRT+ ΔgabT:: FRT+ ΔdadA :: FRT+ P>pGEN1124 bla+ AmDH4 (*Petrotoga mobilis*)+	AmDH4 - pSP100
G5592	E. coli K12 C> ΔpuuE ::FRT+ ΔgabT: FRT+ ΔdadA :: FRT+ P> pGEN1308 bla+ AmDH4* (*Petrotoga mobilis*)+	AmDH4-I80T/P224S/E296G - pSP100
G5313	E. coli K12 C> Δzwf ΔmaeB ΔudhA ΔpntAB Δicd::FRT+ ΔmaeA::FRT+ Δlpd::FRT+	Ø
G5641	E. coli K12 C> Δzwf ΔmaeB ΔudhA ΔpntAB Δicd::FRT+ ΔmaeA::FRT+ Δlpd::FRT+ P>pGEN1237 *Micro*AmDH+ (*Microbacterium* sp. MA1) cat+	*Micro*AmDH - pVDM18
G5647	E.coli K12 C> Δzwf ΔmaeB ΔudhA ΔpntAB Δicd::FRT+ ΔmaeA::FRT+ Δlpd::FRT+ P>pGEN1330 cat+ AspRedAm+ (*Aspergillus oryzae*)	*Asp*RedAm - pVDM18
G5648	E.coli K12 C> Δzwf ΔmaeB ΔudhA ΔpntAB Δicd::FRT+ ΔmaeA::FRT+ Δlpd::FRT+ P>pGEN1332 bla+ AspRedAm+ (*Aspergillus oryzae*)	*Asp*RedAm - pSP100
G5780	E. coli K12 C> Δzwf ΔmaeB ΔudhA ΔpntAB Δicd::FRT+ ΔmaeA::FRT+ Δlpd::FRT+ ΔkdgK:: *Asp*RedAm (*Aspergillus oryzae*)+ aad+	*Asp*RedAm - chromosome
G5782	E.coli K12 C>DkdgK::aspRedAm (Aspergillus oryzae)+ aad+ P> pORTMAGE-2 bla+ mutL E32K+	*Asp*RedAm - chromosome

Appendix 58. (Chapter III, II) Maps of pSP100 and pVDM18 vectors.



Appendix 59. (Chapter III, II.2 and II.3) Composition of the culture medium used in cells growth assays in liquid medium, bioscreen assays and evolution in the GM3.

MS	
Citric acid.H_2O	4 mM
MgSO4.7H_2O	1 mM
NH_4Cl	20 mM
K_2HPO_4	50 mM
D-glucose	0.2% (m/v)
NTAmix (1000X)	0.1% (v/v)

MSN-	
Citric acid.H_2O	4 mM
MgSO4.7H_2O	1 mM
K_2HPO_4	50 mM
D-glucose	0.2% (m/v)
NTAmix (1000X)	0.1% (v/v)

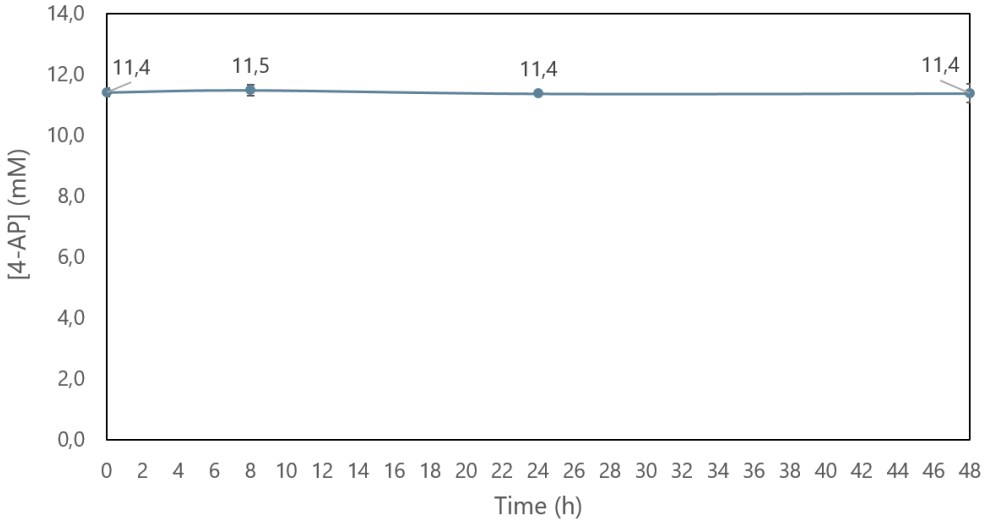
MA	
Na_2HPO_4.12H_2O	31 mM
MgSO4.7H_2O	1 mM
NH_4Cl	18 mM
K_2HPO_4	25 mM
Nitrilotriacetic acid trisodium	40 µM
NTAmix (1000X)	0.1% (v/v)

MAN-	
Na_2HPO_4.12H_2O	31 mM
MgSO4.7H_2O	1 mM
K_2HPO_4	25 mM
Nitrilotriacetic ac	40 µM
NTAmix (1000X)	0.1% (v/v)

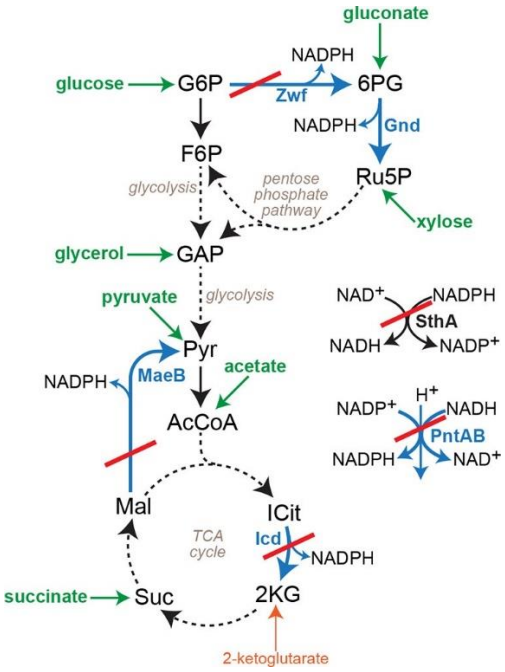
Appendix 60. (Chapter III, II.2.3.1.1) Growth assays using G4480 strain in presence of NH_4Cl and/or 4-AP. NH_4Cl eq. 2% corresponds to the addition of 4-AP 10mM if contaminated with eq. 2% v/v of NH_4Cl.

Strain	Time (h)	Nitrogen source			
		NH_4Cl (5 mM)	NH_4Cl eq. 2% (0.57 mM)	4-AP (10 mM) NH_4Cl eq. 2%	4-AP (10 mM)
G4480	24	1.20	0.17	0.36	0.19
	48	1.14	0.19	0.36	0.25

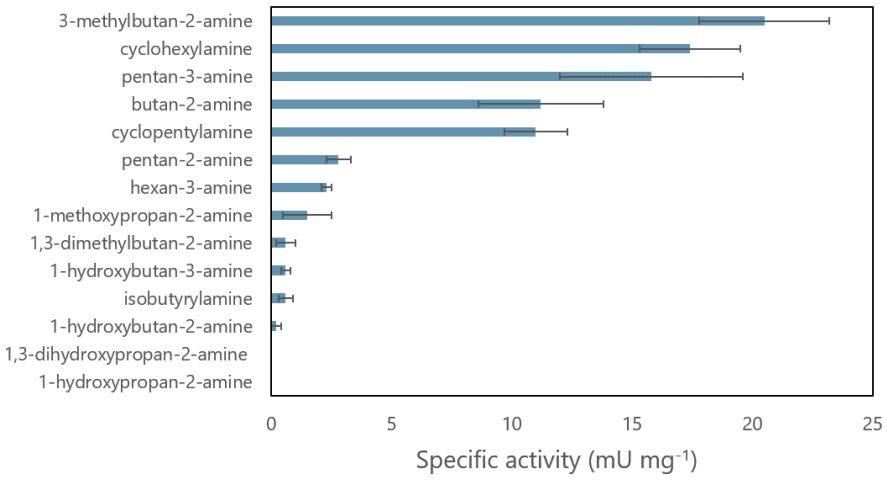
Appendix 61. (Chapter III, II.2.3.1.1) Monitoring of 4-AP contained in the culture medium over time in presence of G4480. A calibration curve containing various concentrations of 4-AP enabled the estimation of 4-AP concentration. Error bars represent the standard deviation of two independent experiments.



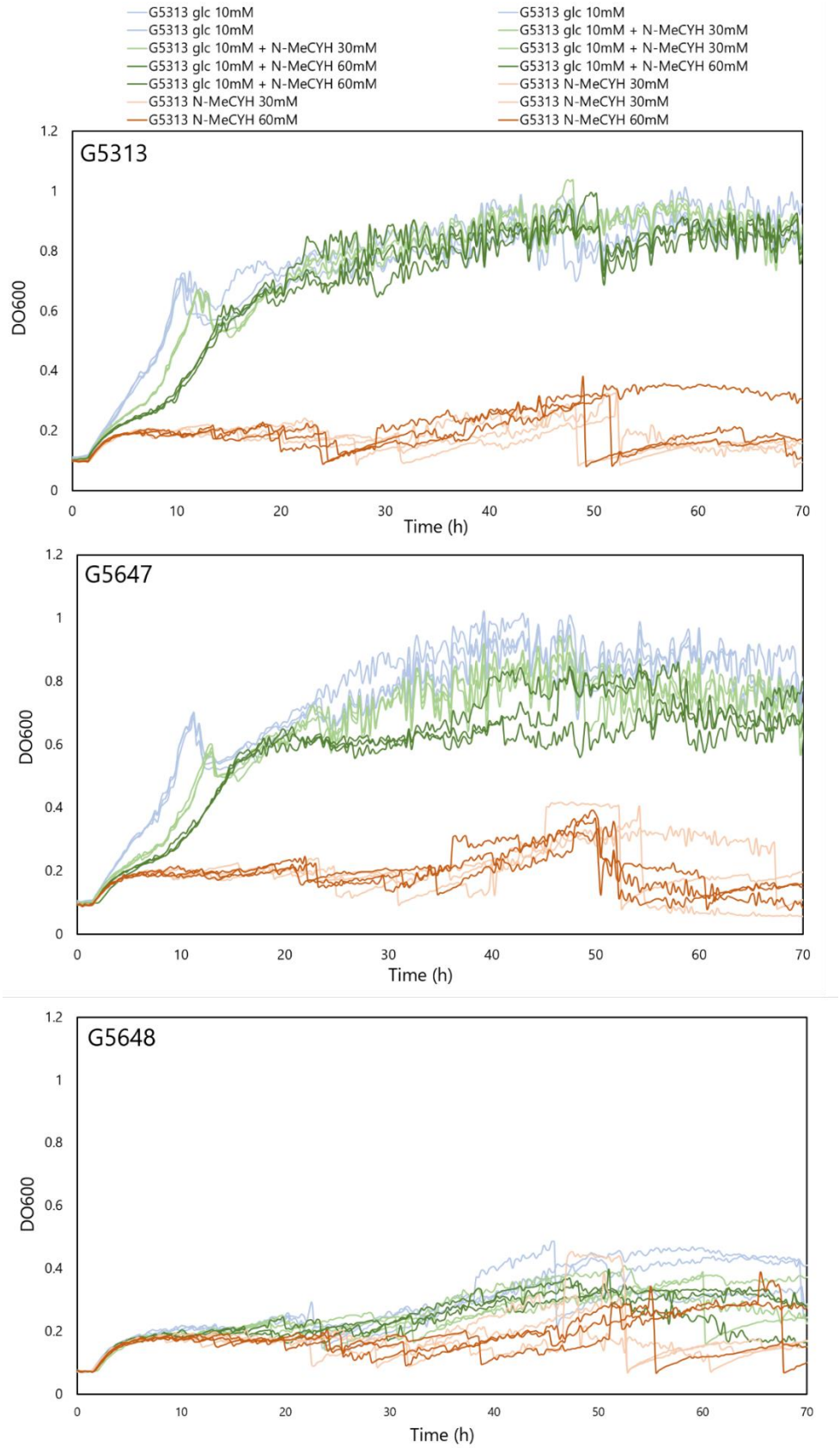
Appendix 62. (Chapter III, II.3) Schematic representation of the NADPH auxotrophic *E. coli*. Enzymes producing NADPH are shown in blue arrows and delete ones are indicated using a red cross line. Carbone sources are shown in green. Extracted from Lindner *et al.* (2018).



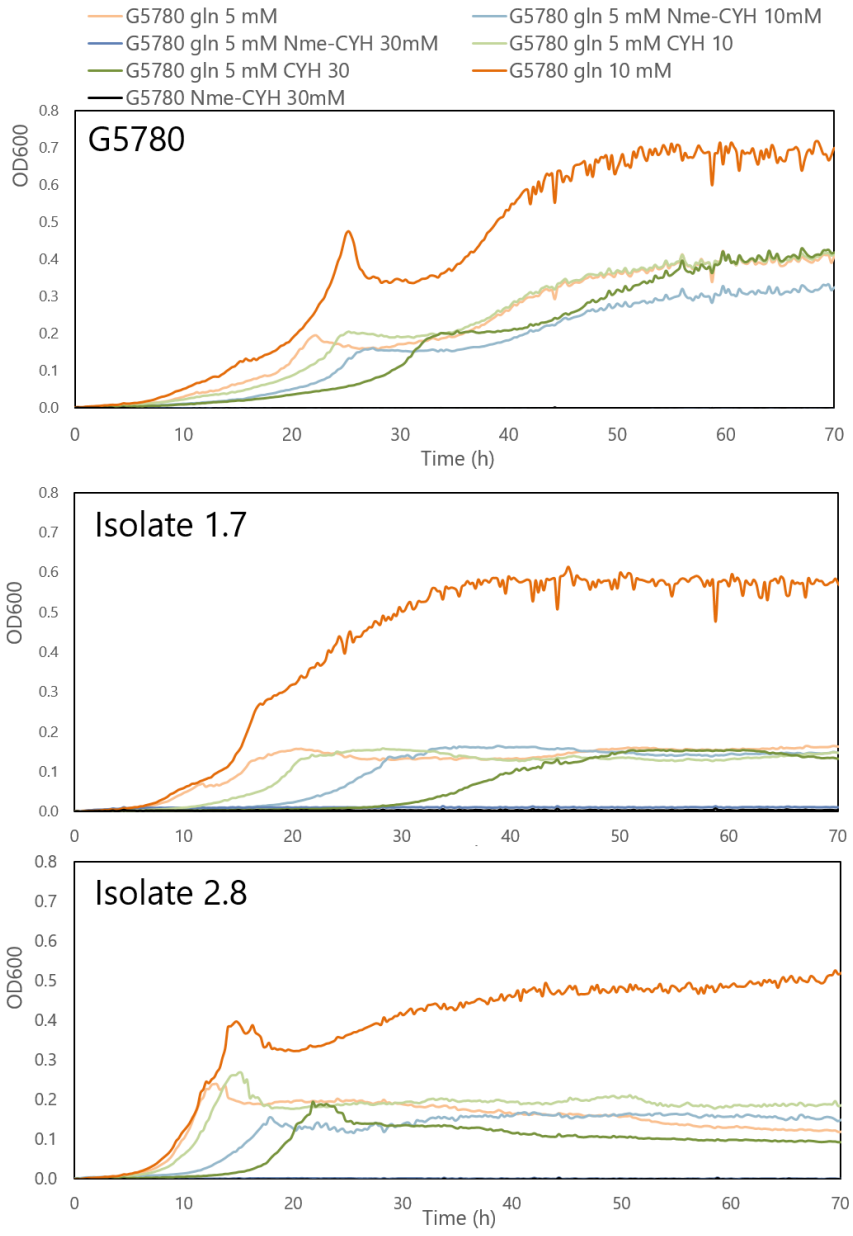
Appendix 63. (Chapter III, II.3.1) Specific activity (mU mg^{-1}) of *Micro*AmDH towards a panel of short amines in the oxidative deamination direction. Reaction conditions were as follows: Final volume 100 µL, 10 mM amine substrate, 0.4 mM $NADP^+$, Tris HCl buffer (50 mM, pH 9), 0.2 - 0.4 mg mL^{-1} purified *Micro*AmDH, 37°C. The uncertainties are those generated out of two to three experiments (Experimental section, II.3.1).



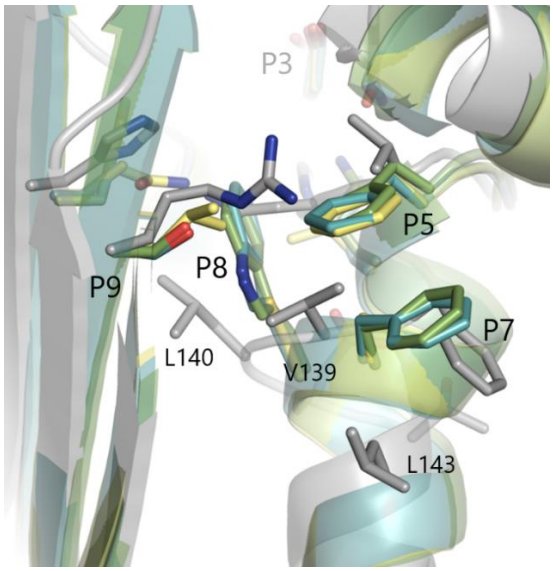
Appendix 64. (Chapter III, II.3.2) Growth curves of G5313, G5647 and G5648 on minimum medium containing gluconate (**244**) and/or *N*-methylcyclohexylamine (N-MeCYH, **54**) as the NADPH source. Growth conditions: MA medium, glycerol (**247**) 10 mM, acetate (**248**) 10 mM and α-ketoglutarate (**249**) 5 mM, ± **244** 10 mM and ± **54** 30-60 mM, 30°C, 180 rpm. For clarity, the legend is only given for G5313 but remains the same for the two other strains.



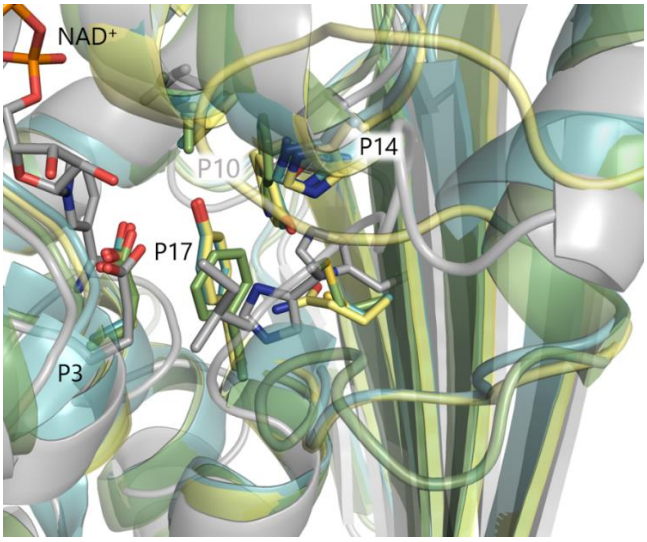
Appendix 65. (Chapter III, II.3.2) Growth curves of G5780 and two isolates sampled from the first evolution of G5780 in the GM3 on minimum medium containing **244** and/or **54** (N-meCYH) and/or **45** (CYH). Growth conditions: MA medium, glycerol (**247**) 10 mM, acetate (**248**) 10 mM and α-ketoglutarate (**249**) 5 mM, ± **244** 5-10 mM and ± **54** 10-30 mM, **45** 10-30 mM, 30°C, 180 rpm. For clarity, the legend is only given for G5780 but remains the same for the two other strains.



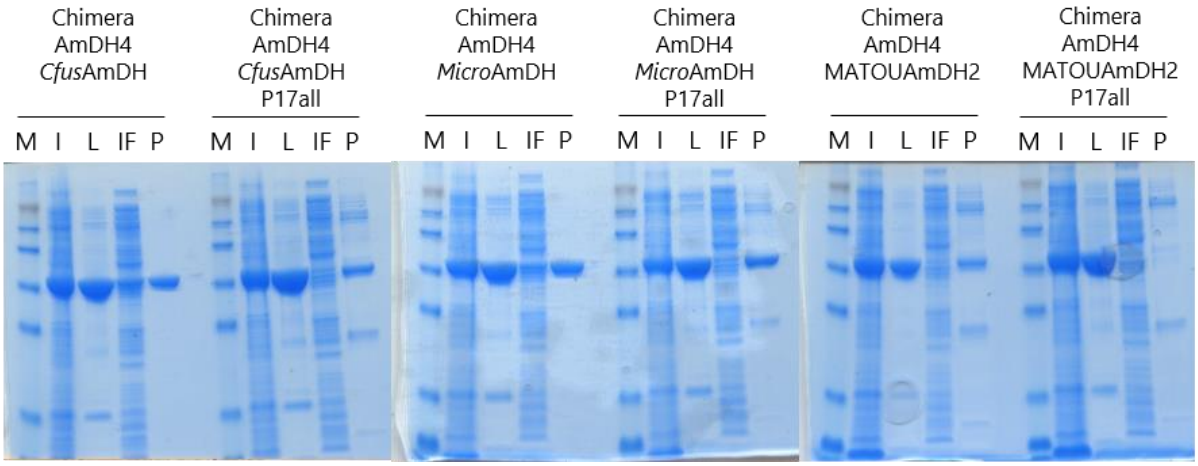
Appendix 66. (Chapter III, III) Structural alignment of *Cfus*AmDH, *Micro*AmDH, MATOUAmDH2 (template 2) on AmDH4 (template 1) focusing on the top of the spine helix V134-I150. AmDH4, *Cfus*AmDH, *Micro*AmDH and MATOUAmDH2 are colored in grey, blue, green and yellow, respectively. Some P1-P21 are represented in stick format. RMSD (*Cfus*AmDH-AmDH4) = 1.936 Å; RMSD (*Micro*AmDH-AmDH4) = 1.442 Å; RMSD (MATOUAmDH2-AmDH4) = 1.906 Å;



Appendix 67. (Chapter III, III) Structural alignment of *Cfus*AmDH, *Micro*AmDH, MATOUAmDH2 (template 2) on AmDH4 (template 1) focusing on the linker segment G191-G199. AmDH4, *Cfus*AmDH, *Micro*AmDH and MATOUAmDH2 are colored in grey, blue, green and yellow, respectively. Some P1-P21 are represented in stick format. RMSD (*Cfus*AmDH-AmDH4) = 1.936 Å; RMSD (*Micro*AmDH-AmDH4) = 1.442 Å; RMSD (MATOUAmDH2-AmDH4) = 1.906 Å;



Appendix 68. (Chapter III, III) SDS-PAGE gels of the chimeras designed between AmDH4 and *Cfus*AmDH, *Micro*AmDH and MATOUAmDH2. M = molecular marker SeeBlue® Plus2 pre-strained standard (5 µg deposited); L = cell-free extracts (15 µg deposited); IF = insoluble fraction of the lysis step (small fraction diluted in water); P = purified enzymes (5 µg deposited).



Appendix 69. (Chapter IV, I) Sequences of the nat-AmDHs selected within ASMC1 for accomodation of larger carbonyl substrates.

Uniprot ID	**Sequence**
A0A076NBW6	MTEPKIRAIVWGCGHMGQIIIKYLKDKGVELVGAIDRNVSRVGKDAGEISALKRSLGTFIHHPDEAKQV FHKANANVCIISTRSLMIDIYDSLETSARHGVNAITLAEEAFYPWNTSPDLTQKLDKLAKEHNCTLTGSG FQDVFLGNLISILAAGTHRIDRIQAIIQYNIDGYGSSVAAQYGVGLSVEAFKQQISGYTEPFPSWYMNE WLCAHLGLAIRTQKQEFWPITNIKTIHSTSLKTDISVGSVTGMKAVVTTETEKGVIIENQVLGKIYTPDEID TYECTLVGEPTTTLTIHRPATAQMTCASLVNRLQHLIDAPPGFVTTDKMQPPNYWM
A0A1H8XY75	MKPIRAALVGVGAMNRIVAGLLPEKGVEIVGAVGRSKAGQDLGEVLELGHRLGISVTTDAGAMYASAR PDIALVATSSFMTGQYAVLADCARHGVNAITIGEELLHPWHTAPERTRELDALARRHGVTLAGGGFQD FFLVHQLAGMLGTMNRLDRLSGQQTFNVDDYGAEVARDQQVGAGVADFARWRQSDERPPGFGLPI LAAIAAASDLTVTGTGFEIRPEVAERATPCRSLGITVGPGDLIGFTTVDTVTTEEGPVLVMELSGRIYQPGE TDRFAWTASGDPDVRVVNPGVDTLRTTCTQMVNRIPDVINAAPGFVGPAGLPPLRYRARPFAEYLLER PRLAVN
A0A1I3IZQ4	MSKIRTIVYGVGATGQNVTRHLAERGVDIVGAIGRVENVGADLGTVAELDRELNVEINDDADDVLENT EADVAIVSIASTLEEMYPHLKRCLEAGVNVITTSEETLYPWYTSPDLAGRLDRIAIENDVTFTGGGYQDIF EVNLPVMLTGASRKLDGVDGTNRYNIDDYGPAVVEYFFGGKDRDDVEEKLERGGAPESLFRTTLEAQI AELGLSEADASQKATPIVADRAVESEALGRTIQEGEVLGLEIDVTIETHEGITFTGTEIAKVYDTEEEDRNE WVIEGTPEMHVVDQPTPVEVGTSTQIVNRLPDIINYESGFVSVTELPRPKYRVGPLETYLD
A0A078AU13	MGEKIALWGFGVMNKIMLKYLIENNYQIATVIGGRDHGKDAGEAAGLEYLGVKVIRPEEADDVLRETR PVLCILSTISGLSLIMDQLVILARNRVNVLTINEDAFYSWSIQPQLTQEIDSMFRQNLVTFTGTGYFDLLG CYFGSLLVGISHKVDKISMKVQFTLDDFGVGDEDGVGLSIEEFEKKFAESKGEIAYIWPPVEALAFRLGW KILAISQKHEPTLAPVDIQSKQYGVIEQGRATGFKAIVTCKALRQDNQVAEIEAQSIGCIYHGDMEDYCE WETIGVPSLTLKMPKPDTNAITCASTVNRIKDCIKARPGYVTVVEMGPIA
A0A078BAL0	MGEKIALWGFGLMNKVMLKYLIESDYQIATVIDNHNHGKDAGEIAGMEQLGVNIISREEADDVLQETR PIACIVSTKSKLSQIMDSLKILAKNRVNVLTINEEAFYSWNIQPQLTYEIDTMFRQNLVTFTGTGYYDLLG CYFGSTLVGISHKVDKLIVRIQYNFDDYGVGDNGIGLTMEEFDEKFVRTEREPTYFWPPVEALAIRLGWK VLSISQKHEPTLAPIDIQSKQYGFIPQGRATGFKAIVYCQLLRQDGKVAEIEAQSIGCIYHGEMEDYCEW ETVGVPSLTLKMPKPDTTAITCASTVNRIKDCIKAKPGVVTVVEMGPIS

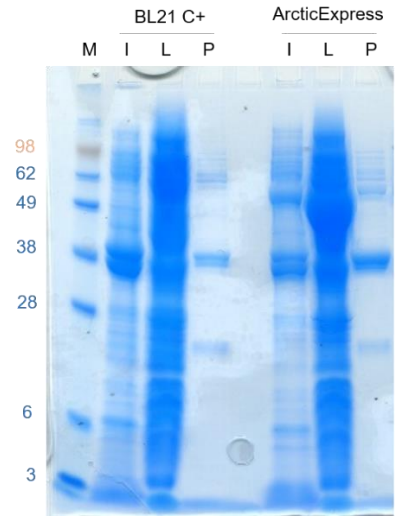
A0A1Q9SC61	MTTIRAVLVGVGAMNAIVAHLLAEKGVEVVGAVGRSPEKVGRDAGEVMGLPHPLGVEVAADAGRLY DSVRPDIALVATSSFMADQYDVLATCARHGVNAITIGEEMLFPWHTAPERTGELDGLAREHGVTLAGG GFQDFFLVNQVAGLLGAVNRFDRLVGRESFNVDDYGAEVARDQRIGTTPDDFAEWQTSSTRPPSFAL PILAAIASSAGLTITASTVELRPDIATVPIESRSLATVVEPGLLIGFTTVDTVTTGEGPILQMELSGHIYRPGE TDRCEWAVQGDPDVRVVNPDVDTLRTTCTQMVNRVADVINAAPGFVTPADLPPLRYRLRPLNEYVL ART

Appendix 70. (Chapter III, III) Primers used for the cloning and production of the nat-AmDHs selected within ASMC1 for accomodation of larger carbonyl substrates. These represent the different His-tag locations in the sequence and soluble tag tested. The segment for pLIC cloning, the start or stop codon and the tags are respectively highlighted in yellow, red and green.

Name	Tag_position	Primer 5'-3'
A0A076NBW6_His6_F	His-tag_N-ter	AAAGAAGGAGATAGGATCATGCATCATCACCATCACCATACGGAACCGAAAATCCGCGCT
A0A076NBW6_F	no His-tag	AAAGAAGGAGATAGGATCATGACGGAACCGAAAATCCGCGCT
A0A076NBW6_R	no His-tag	GTGTAATGGATAGTGATCTTACATCCAGTAATTGGGAGGTTGCAT
A0A076NBW6_His6_R	His-tag_C-ter	GTGTAATGGATAGTGATCTTAATGGTGATGGTGATGATGCATCCAGTAATTGGGAGGTTGCAT
A0A076NBW6_Lys5_F	Lys-tag_N-ter	AAAGAAGGAGATAGGATCATGAAAAAGAAAAAGAAAACGGAACCGAAAATCCGCGCT
A0A076NBW6_Arg5_F	Arg-tag_N-ter	AAAGAAGGAGATAGGATCATGCGTCGCCGTCGCCGTACGGAACCGAAAATCCGCGCT
A0A076NBW6_Asp5_F	Asp-tag_N-ter	AAAGAAGGAGATAGGATCATGGATGATGACGATGACACGGAACCGAAAATCCGCGCT
A0A1H8XY75_His6_F	His-tag_N-ter	AAAGAAGGAGATAGGATCATGAAACCGATTCGCGCTGCACTG
A0A1H8XY75_F	no His-tag	AAAGAAGGAGATAGGATCATGCATCATCACCATCACCATAAACCGATTCGCGCTGCACTG
A0A1H8XY75_R	no His-tag	GTGTAATGGATAGTGATCTTAATTCACGGCCAAACGTGGAC
A0A1H8XY75_His6_R	His-tag_C-ter	GTGTAATGGATAGTGATCTTAATGGTGATGGTGATGATGATTCACGGCCAAACGTGGAC
A0A1H8XY75_Lys5_F	Lys-tag_N-ter	AAAGAAGGAGATAGGATCATGAAAAAGAAAAAGAAAAAACCGATTCGCGCTGCACTG
A0A1H8XY75_Arg5_F	Arg-tag_N-ter	AAAGAAGGAGATAGGATCATGCGTCGCCGTCGCCGTAAACCGATTCGCGCTGCACTG
A0A1H8XY75_Asp5_F	Asp-tag_N-ter	AAAGAAGGAGATAGGATCATGGATGATGACGATGACAAACCGATTCGCGCTGCACTG

Appendix 71. (Chapter III, I) SDS-PAGE gels of A0A1H8XY75 overexpressed in BL21 C+ or ArcticExpress. M = molecular marker SeeBlue® Plus2 pre-strained standard (5 µg deposited), L = cell-free extracts (15 µg deposited)

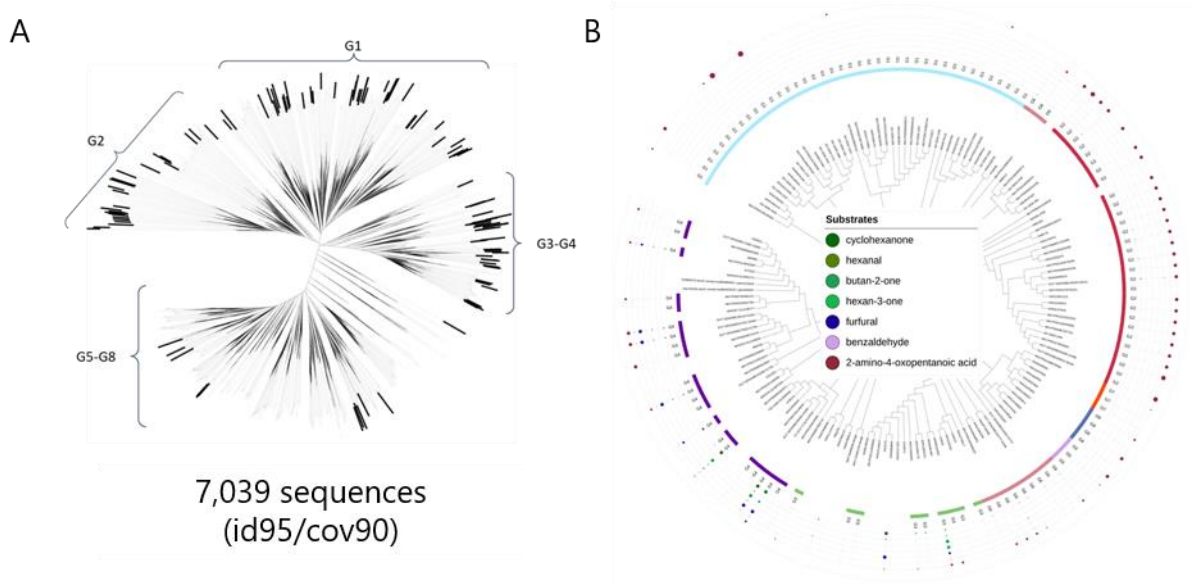
and P = purified enzymes (5 µg deposited).



Appendix 72. (Chapter IV, II) (Meta)genomic databases used for nat-AmDH family expansion

Resources	Number of sequences	Volumes
UniProtKB SwissProt	562,253	266.2 MiB (release 2020_02)
UniProtKB TrEMBL	184,998,855	81.6 GiB (release 2020_03)
GEM (Genome from Earth's Microbiomes)	111,428,992	25 GiB (download 2021_02)
OM-RGC (Ocean Microbial Reference Gene Catalog)	46,828,091	15.4 GiB (download 2020_05)
MetDB (Marine Eukaryotes Transcriptomes)	16,231,949	7.4 GiB (download 2020_07)
SMAGs (Tara Ocean Eukaryote Metagenome Assembled Genomes)	10,207,435	3.6 GiB (download 2020_09)
IGC (Integrated Gene Catalog of Human Gut)	9,878,647	8.9 GiB (download 2020_07)
UHGP (Unified Human Gastrointestinal Protein)	170,602,708	67.3 GiB (release 2020_01)
Mgnify (EMBL-EBI)	1,106,951,200	133.6 GiB (release 2019_05)
MATOUv2 (Tara Oceans Eukaryote Gene Catalog)	950,799,894	174.1 GiB (download 2020_05)
	Total 2,608,490,024	

Appendix 73. (Chapter III, II.3) Preliminary results of the screening carried out to describe the extended nat-AmDH family. (A) Phylogenetic tree of the extended set with id95/cov90 threshold (B) Activity results labelled as circles of proportional size according to the activity measured towards cyclohexanone (**45**), hexanal (**207**), butan-2-one (**134**), hexan-3-one (**200**), furfural (**57**), benzaldehyde (**31**) and (2*R*)-2-amine-4-oxopentanoic acid (**166**).



Appendix 74. (Chapter IV, II.5.5) Example of the design of a N-ter truncated enzyme (MGYP000882735950, in a black frame) selected within the ASMC2. The figure represent a MSA between the target enzyme and homologs with complete sequences use as templates. The sequence added in this case was deduced from a consensus approach of the templates. TVVKIARPATVELTCATIVNRIPQLLEAPAGYYTTEKMPPAEYRTYPLHFYVDF was added after the GEPN terminal.

```
|A0A1M6C3X8|A0A1M6C IVSECIGKVYGEGEIDRNDWKILGEPNTEVKIARPSTVELTCATIVNRIPQLIEAPAGFYTTEKMPPARYRTYPLHYYVR----
MGYP000854504544    IVSECIGKVYAEGEYDRNDWSIKGEPDTVLKIARPSTVELTCATIVNRIPQLIESPPGYYTTEKMPPAKYRTYPLHFYVK----
MGYP000882735950    LVSECIGKVYAEGEFDRNDWTITGEPN------------------------------------------------------------
MGYP000883079973    VVSECIGKVYAEGEVDRNDWKIIGEPNTVVNIARPATVELTCATIVNRIPQLIEAPPGYYTTEKMPPAQYRTYPLHFYVK----
MGYP001021160058    IVSECIGKVYAEGEFDRNDWKILGEPDTVVNIARPATVELTCATIVNRIPQLIDAPPGYYTTEKMPPAQYRTFPLHFYVK----
MGYP000969854580    IVSECIGKVYAEGEFDRNDWSIIGEPNTLVKIARPATVELTCATVVNRIPQLLQAPAGFYTTEKMPPAEYRTYPLHFYVEW---
MGYP000901104217    IVSECIGKVYAEGEFDRNDWSIIGEPNTVVKIARPATVELTCATIVNRIPQLLEAPAGFYTTEKMAPAQYRTYPLHFYVG----
MGYP000852032497    IVSECIGKVYAEGEFDRNDWSIIGEPNTVVKIARPATVELTCATIVNRIPQLLEAPAGYYTTEKMPPAEYRTYPLHFYVD----
MGYP001014780063    IVSECIGKVYAEGEFDRNDWSIMGEPNTVVKIARPATVELTCATIVNRIPQLLEAPAGYCTTEKMPPAEYRTYPLHFYVDF---
MGYP000902496227    IVSECIGKVYAEGEFDRNDWSIMGEPNTVVKIARPATVELTCATIVNRIPQLLEAPAGYCTTEKMPPAEYRTYPLHFYVDF---
MGYP000855550318    LVTQCIGKVYPEGEVDTNLWTIYGEPNTTVKIEQPATVELTCATIVNRIPQLLMAPAGFYTTDKWDPAKYLTYPMNLYVEEIY-
tr|F4LTF2|F4LTF2_TE LETQCIGKVYPEGEVDTNLWTIYGEPNTTLKIEQPATVELTCATIVNRIPQLIMAPPGFYTTEKMPPAKYLAYPMNLYIDDCCW
```

La forte présence d'amines dans les composés chimiques issus des industries pharmaceutiques, agrochimiques et plus généralement des industries de chimie fine et de gros tonnage, nécessite le développement de nouveaux procédés durables de synthèse d'amines prenant en compte les défis économiques et environnementaux actuels. La biocatalyse, qui consiste en l'utilisation d'enzymes dans des procédés de synthèse, représente une alternative plus verte aux procédés chimiques conventionnels. Elle permet de travailler de manière catalytique dans des conditions douces (température, pression, en milieu aqueux, etc.) et généralement au sein de schémas réactionnels plus courts grâce à l'exceptionnelle chimio et stereosélectivité des enzymes. D'abord centrées autour des lipases et des transaminases, les solutions biocatalytiques pour la synthèse d'amines se sont étendues au milieu des années 2010 avec la découverte d'enzymes catalysant l'amination réductrice de composés carbonylés. Ces solutions comprennent plusieurs familles, les opine déshydrogénases, les imine réductases, les réductive aminases (RedAm) et les amine déshydrogénases (AmDHs). Ces dernières étaient d'abord restreintes aux acide aminés déshydrogénases ingénierées, jusqu'à la découverte, en 2019, de gènes codant pour des enzymes natives (nat-AmDHs) catalysant la formation de l'imine en présence d'ammoniac et sa réduction par un cofacteur nicotinamide. Dans le cas de réactions sur des cétones prochirales, elles donnent lieu très préférentiellement à la formation de l'énantiomère *S*, stéréochimie rarement obtenue par les autres familles d'enzymes catalysant la même réaction. La famille des nat-AmDHs s'est étendue grâce à plusieurs cycles d'exploration de génomes et metagénomes, restant cependant limitée par une faible diversité de substrats. Les recherches présentées cherchaient à apporter cette diversité manquante par des approches basées sur des analyses *in silico* de structures cristallographiques, obtenues grâce à un travail collaboratif extérieur, ou de modèles générés par homologie. Dans un premier temps, une analyse structurale poussée de la famille couplée à un travail d'ingénierie protéique a révélé l'importance de certains résidus du site actif. Même si le mécanisme réactionnel des nat-AmDHs n'est encore qu'hypothétique, cette étude a permis de mieux comprendre le rôle de ces résidus dans la réaction et a aussi aidé à définir plusieurs configurations de sites actifs minimaux dans le cadre d'une recherche d'enzymes natives analogues. En collaboration avec l'équipe du Pr. Bommarius, une étude portant sur la stabilité d'un sous-ensemble de nat-AmDHs a été entreprise pour corréler ces données de stabilité avec l'état oligomérique des enzymes étudiées. De notre côté, il s'agissait de suivre l'activité spécifique des nat-AmDHs après différentes conditions de stockage. Comme attendu, la température s'est révélé être un paramètre clé influençant la stabilité des enzymes et deux nat-AmDHs se sont démarquées, *Acol*AmDH et IGCAmDH5, en conservant plus de 80% d'activité spécifique après six jours à 37°C. La connaissance du spectre de substrats des nat-AmDHs, indiqué comme une limite de la famille, a été

complétée par rapport aux données déjà publiées. Parmi les enzymes testées, certaines ont montré une activité notable pour la synthèse d'amines plateformes pour la synthèse de composés pharmaceutiques ou agrochimiques [furfurylamine, (2*S*)-methoxypropan-2-amine, (2*S*)-butan-2-amine, (3*S*)-2-hydroxybutan-3-amine, etc.]. Une étude très préliminaire a également été débutée afin d'estimer le potentiel de certaines nat-AmDHs à catalyser la réaction de sulfidation réductrice de fonctions carbonyles en thiols correspondants en présence d'un donneur du soufre (Na_2S), remplaçant hypothétique du donneur d'amine NH_3. L'enzyme MATOUAmDH2 s'est révélée une candidate potentielle pour la formation d'hexan-1-thiol, mais ces résultats restent à confirmer, notamment du fait d'un fort taux de formation spontanée des thiols dans les conditions utilisées. Plusieurs approches ont été utilisées pour étendre le spectre de substrats des nat-AmDHs restreints aux petits composés carbonylés (<C6). La première s'est basée sur l'identification *in silico* de deux résidus encombrants fortement conservés au sein de la famille et ainsi considérés comme des cibles de choix pour l'agrandissement du site actif. Ces deux résidus ont été mutés dans dix nat-AmDHs possédant une structure similaire mais des sites actifs légèrement différents. La majorité des mutants se sont montrés actifs sur des substrats encore non atteints par les enzymes natives, tels que les aldéhydes hexanal (jusqu'à 99.5% de conversion non optimisée) à nonanal (44.5%), les cétones hexan-2-one (39.5%) à nonan-2-one (16%) et la 4-phenylbutan-2-one (52.5%), tout en conservant leur forte (*S*)-enantiosélectivité (> 90% *ee*). De bonnes conversions ont été obtenues par certaines enzymes, natives ou mutées, sur l'hexan-3-one (51.2%) et dans une moindre mesure sur les homologues heptan-3-one et octan-3-one, un type de substrats encore assez peu décrit, même dans les autres familles d'enzymes permettant l'amination réductrice. Le meilleur mutant *Cfus*AmDH-W145A a été étudié plus en détails et utilisé pour la synthèse d'environ 100 mg de (2*S*)-octan-2-amine et heptan-1-amine. La cristallisation et la résolution structurale de *Cfus*AmDH-W145A, réalisée au sein de l'équipe du Pr. Grogan, couplée à des études de dynamiques moléculaires ont permis d'émettre des hypothèses sur sa spécificité de cofacteur et de substrat. Une autre approche, cette fois d'évolution *in vivo*, utilisant les automates de culture en continu disponibles au Genoscope, a été tentée pour l'amélioration des activités de certaines nat-AmDHs et d'*Asp*RedAm. Les deux cribles de sélection se basaient sur l'activité de déamination oxidative permettant la libération de NH_3 ou de NAPDH nécessaire à la croissance des souches *E. coli* utilisées. Cependant l'activité de ces enzymes dans ce sens réactionnel s'est révélé hypothétiquement trop faible pour engendrer une croissance cellulaire suffisante pour permettre une évolution. Une dernière approche a été de créer des chimères entre deux nat-AmDHs afin de combiner leurs caractéristiques les plus utiles dans un but de synthèse. Ce travail préliminaire a été testé en couplant d'un côté AmDH4, une enzyme thermostable et thermoactive mais non active sur des cétones sans prérequis fonctionnel, et de l'autre *Cfus*AmDH, *Micro*AmDH ou MATOUAmDH2, des enzymes actives sur des cétones et

aldéhydes simples mais instables dès 30-40°C. Les premiers designs ont été infructueux mais le travail pourra être poursuivi de manière collaborative. Enfin, dans le cadre du projet en collaboration interne MODAMDH cherchant à trouver plus de diversité de séquences AmDHs, la famille a été drastiquement étendue de 2 011 à 17 039 membres. En se basant sur les analyses structurales et résultats expérimentaux obtenus jusqu'alors, une sélection *in silico* de 92 enzymes a été faite afin de trouver de nouvelles nat-AmDHs possédant des caractéristiques structurales particulières pouvant influer sur leur spectre de substrats amines ou carbonyles ainsi que sur leur stéréosélectivité. Grâce à ce travail, nous avons une vue et une compréhension quasi complète de la diversité structurale au sein des nat-AmDHs ce qui aidera à leur amélioration et à leur application en synthèse.

université PARIS-SACLAY

ÉCOLE DOCTORALE
Sciences Chimiques : Molécules, Matériaux, Instrumentation et Biosystèmes (2MIB)

Titre : Identification et modification d'amine déshydrogénases par approches génomique et structurale pour la synthèse biocatalysée d'amines

Mots clés : biocatalyse, amination réductrice, analyse structurale d'enzymes, amine déshydrogénases

Résumé : Pour faire face à la forte demande en amines des industries pharmaceutiques, agrochimiques et plus généralement des industries de chimie fine et de gros tonnages, de nouveaux procédés durables de synthèse d'amines doivent être développés en prenant en compte les challenges économiques et environnementaux actuels. La biocatalyse, qui consiste en l'utilisation d'enzymes pour la synthèse, représente une alternative plus verte aux procédés chimiques conventionnels. Elle permet de travailler de manière catalytique dans des conditions douces et généralement dans des temps réactionnels plus court grâce à la chimio et stereosélectivité des enzymes. D'abord centrées autour des lipases et des transaminases, les solutions biocatalytiques pour la synthèse d'amines se sont étendues dans les années 2010 avec la découverte d'enzymes catalysant l'amination réductrice de composés carbonylés. Ces solutions comprennent plusieurs familles telles que les opine déshydrogénases, les imine réductases, les réductive aminases (RedAms) et les amine déshydrogénases (AmDHs). Ces dernières étaient restreintes aux acide aminé déshydrogénases ingénierées jusqu'à la découverte d'enzymes natives (nat-AmDHs) catalysant la formation de l'imine en présence d'ammoniac et sa réduction par un cofacteur NAD(P)H. Elles favorisent la formation de l'amine (*S*), rarement décrit dans d'autres familles d'enzymes catalysant la même réaction. La famille des nat-AmDHs s'est étendue grâce à plusieurs cycles d'exploration de (méta)génomes mais manquait encore de diversité de séquences. Les recherches présentées cherchent à apporter cette diversité manquante par des approches basées sur des analyses *in silico* de modèles par homologie ou structures crystallographiques, obtenues grâce à un travail collaboratif extérieur. Une analyse structurale poussée de la famille couplée à un travail d'ingénierie protéique a révélé l'importance de certains résidus du site actif. La caractérisation de quelques nat-AmDHs a montré leur activité sur des substrats encore peu étudiés tels que des petits aldéhydes aromatiques ou cétones hydroxylées. Plusieurs approches ont été utilisées pour étendre le spectre de substrats des nat-AmDHs restreints aux petits composés carbonylés (<C6). La première a permis l'identification *in silico* de deux résidus encombrants et conservés au sein de la famille, considérés comme des cibles pour l'agrandissement du site actif. Ces deux résidus ont été mutés dans dix nat-AmDHs possédant une structure similaire mais des sites actifs différents. La majorité des mutants se sont révélés actifs sur des substrats encore non atteints tels que les aldéhydes de hexanal (jusqu'à 99.5% de conversion non optimisée) à nonanal (44.5%), les cétones de hexan-2-one (39.5%) à nonan-2-one (16%) et la 4-phenylbutan-2-one (52.5%), en conservant leur forte (*S*)-enantiosélectivité (> 90% *ee*). Le meilleur mutant *Cfus*AmDH-W145A a été utilisé pour la synthèse d'environ 100mg de (2*S*)-octan-2-amine et heptan-1-amine. Une autre approche d'évolution *in vivo* en continu, utilisant les automates de culture du Genoscope, a été tentée pour l'amélioration des activités de certaines nat-AmDHs et d'*Asp*RedAm. Les deux sélections se basaient sur l'activité de déamination oxidative de ces enzymes qui s'est révélé hypothétiquement trop faible pour permettre une croissance cellulaire suffisante pour permettre une évolution. Dans le cadre du projet MODAMDH cherchant à trouver plus de diversité de séquences AmDH, la famille a été drastiquement étendue de 2011 à 17039 membres. En se basant sur les analyses structurales et résultats expérimentaux obtenus, une sélection de 92 enzymes possédant des caractéristiques particulières a été faite. Grâce à ce travail, nous avons une vue et une compréhension quasi complètes de la diversité des nat-AmDHs ce qui aidera à leur amélioration et à leur application en synthèse.

université PARIS-SACLAY

ÉCOLE DOCTORALE
Sciences Chimiques : Molécules, Matériaux, Instrumentation et Biosystèmes (2MIB)

Title : Identification and modification of amine dehydrogenases by genomic and structural approaches for the biocatalytic amine synthesis

Keywords : biocatalysis, reductive amination, enzyme structural analysis, amine dehydrogenases

Abstract: Facing the great demand of amine intermediates in pharmaceutical, agrochemical as well as in fine and bulk chemical industries, new sustainable processes for amine synthesis need to be developed considering the current economic and environmental challenges. Biocatalysis, *i.e.* the use of enzymes in synthesis, represents a greener alternative to traditional chemical processes. It enables to work catalytically in mild conditions and generally with shorter reaction times thanks to the high chemo- and stereoselectivity of enzymes. First mainly limited to lipases and transaminases, the biocatalytic alternatives for amine synthesis have expanded since the 2010s with the discovery of enzymes catalyzing the reductive amination of carbonyl-containing compounds. These solutions encompass several enzyme families such as opine dehydrogenases, imine reductases, reductive aminases (RedAms) and amine dehydrogenases (AmDHs). The latter, previously restricted to engineered amino acid dehydrogenases, has grown in recent years thanks to the identification of genes coding for native AmDHs (nat-AmDHs) that catalyze the imine formation from a carbonyl compound and ammonia and its subsequent reduction by NAD(P)H cofactor. These enzymes harbor a high (*S*)-enantioselectivity rarely obtained from other enzyme families catalyzing the same reaction. The nat-AmDH family has been extended over the years through several iterations of (meta)genome mining but still lacked in active site diversity. The studies described in this research work aimed at filling this missing diversity using structure-guided approaches based on homology models or 3D structures obtained from an external collaborative work. First, a deep structural analysis of the family members coupled with targeted protein engineering highlighted the importance of some residues in the active site. The family was furthered characterized for reductive amination of still undescribed key substrates such as short (hydroxy)ketones or aromatic aldehydes. Some approaches were carried out to expand the substrate scope of the native AmDHs, restricted to short carbonyl substrates. One strategy relied on the *in silico* identification of two bulky residues highly conserved within the family which were considered as hot spots for the active site expansion. These two residues were mutated in ten nat-AmDHs bearing similar scaffolds but slightly different active sites. The majority of the mutants were active towards still unreached substrates such as aldehydes from hexanal (non-optimized conversions up to 99.5%) to nonanal (44.5%), ketones from hexan-2-one (39.5%) to nonan-2-one (16%) and 4-phenylbutan-2-one (52.5%) while still retaining their high (*S*)-enantioselectivity (>90% *ee*). The best hit *Cfus*AmDH-W145A was furthered characterized and used in a 100-mg scale synthesis of (2*S*)-octan-2-amine and heptan-1-amine. Secondly, an approach of continuous *in vivo* evolution using automated devices available at the Genoscope was attempted to enhance the activity of some nat-AmDHs and *Asp*RedAm. The selection screens were based on the activity of these enzymes in the oxidative deamination way that was hypothetically too low to enable sufficient initial cells growth thus preventing good starting point for evolution. Eventually, within the scope of the MODAMDH project to find AmDHs with high sequence diversity among 2.6 billion sequences, the family was drastically extended from 2,011 members to 17,039 members. Thus, taking advantage of all the structural and experimental data collected on the characterized nat-AmDHs, a promising structure-guided selection of 92 new enzymes harboring specific features has been done. Thanks to this work, we have a nearly complete structural picture and understanding of the nat-AmDHs diversity that will help further improvements and applications in synthetic chemistry.