



De la molécule à sa synthèse : de la chimie biomimétique à la biocatalyse à l'ère de la génomique

Carine Vergne-Vaxelaire

► To cite this version:

Carine Vergne-Vaxelaire. De la molécule à sa synthèse : de la chimie biomimétique à la biocatalyse à l'ère de la génomique. Chimie thérapeutique. Université Evry-Val d'Essonne; Université Paris Saclay, 2020. <tel-04328810>

HAL Id: tel-04328810

<https://hal.science/tel-04328810v1>

Submitted on 7 Dec 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.



HAL Authorization

Université Evry Val d'Essonne
Université Paris Saclay

Année 2020

Mémoire d'Habilitation à Diriger des Recherches

Discipline : Chimie

Présenté par :

Carine VAXELAIRE

Chercheur CEA
UMR Génomique Métabolique
Laboratoire de Biocatalyse, Biorémédiation et Métabolisme Synthétique
CEA/Institut de Biologie François Jacob/Genoscope

De la molécule à sa synthèse : de la chimie biomimétique à la biocatalyse à l'ère de la génomique.

Présenté devant le jury constitué de :

Rapporteurs :

Bastien NAY (DR, CNRS, HDR)
Frédéric AVENIER (MCU, HDR)
Olivier BERTEAU (DR, INRAE, HDR)

Ecole Polytechnique, Palaiseau
Université Paris-Sud, Orsay
Institut Micalis, Jouy-en-Josas

Examineurs :

Muriel GONDROY (Chercheur, CEA, HDR)
Isabelle ANDRE (DR, CNRS, HDR)
Anne ZAPARUCHA (PR, HDR)

I2BC, Gif-Sur-Yvette
Toulouse Biotechnology Institute, Toulouse
Université Evry Val d'Essonne, Evry

Remerciements

Ali, je commence par toi cette page dédiée aux remerciements, car je crois que c'est ton nom qui restera en premier dans la case "boulot_recherche scientifique" de ma mémoire. Tout simplement car c'est toi qui a guidé mes premiers pas dans ce monde passionnant de la recherche scientifique et qui continue à être pour moi un modèle de gestion d'équipe et un modèle humain. Merci.

Quant à toi Anne, un immense merci de m'avoir permis de continuer des travaux de recherche à tes côtés en intégrant ton équipe ici au Genoscope. On aura toutes les deux parcouru un bon bout de chemin de notre carrière scientifique ensemble. Je suis très fière de grandir scientifiquement à tes côtés. Tu me donnes au quotidien l'opportunité de mûrir mes pensées scientifiques tout en veillant avec empathie à mon équilibre familial, et c'est donc en grande partie grâce à toi que je peux prétendre présenter cette habilitation aujourd'hui.

Un immense merci à toi Véro car tu m'ouvres au quotidien à tous ces domaines qui m'étaient totalement étrangers il y a encore 12 ans. Toutes les discussions que l'on a pu avoir m'ont permis de construire des projets plus pluridisciplinaires. Ton soutien scientifique et humain est sans faille, merci. Marcel, un très grand merci à toi aussi. En plus de l'ouverture thématique que tu m'as également permis d'aborder, nous avons pu partager de nombreuses réflexions sur la vie d'une unité de recherche; ceci sera plus que bénéfique pour la suite de ma carrière.

Des remerciements tous spéciaux à tous mes collègues actuels, notamment à Aurélie, Jean-Louis, Madeleine, Volker, Alain, David, Karine, Katia, Virginie, Adrien, Aline, Peggy, Christine, Eric... qui m'ont aidé à surmonter les défis scientifiques que nous avons relevés ces dernières années; sans oublier bien sûr toutes les autres personnes de l'unité avec qui je peux interagir de près ou de loin.

Ombeline, Chloé, Adam, Franck, Damien et maintenant Laurine, ainsi que les étudiants ingénieur/master/licence encadrés : sans vous tout ceci ne serait pas possible car quasiment aucun résultat ne pourrait être affiché ! Merci beaucoup pour votre investissement expérimental mais aussi stratégique dans ces projets que j'espère vecteurs de maturation scientifique pour vous.

Je n'oublie pas bien sûr mes anciens collègues de l'ICSN et de Servier que j'ai plaisir à recôtoyer et avec qui, pour certains, des collaborations scientifiques naissent encore.

Je tiens chaleureusement à remercier Bastien Nay, Olivier Berteau, Frédéric Avenier, Muriel Gondry, Isabelle André et Anne d'avoir accepté de prendre part au comité d'évaluation de ces travaux et avec qui je l'espère de nouvelles collaborations seront entreprises.

Enfin, un remerciement spécial à mon mari François, mes deux enfants Bastien et Simon, ma famille et mes amis pour leur amour et amitié qui me permettent de m'épanouir au mieux dans un équilibre professionnel et privé qui m'est cher.

Table des matières

Introduction.....	1
<i>Curriculum vitae</i>	3
I. Activités scientifiques de recherche.....	8
I.1. Synthèse organique de molécules biologiquement actives.....	8
I.2. Synthèse biomimétique et extraction de produits naturels (doctorat).....	11
I.2.1. Présentation succincte des métabolites ciblés	11
I.2.2. Principaux résultats obtenus.....	13
I.3. Synthèse biocatalysée de synthons chimiques	17
I.3.1. La biocatalyse et les missions de l'équipe de biocatalyse	17
I.3.2. Présentation succincte des projets supervisés et co-supervisés	18
I.3.3. Participation active au sein de projets de biocatalyse	33
I.3.4. Support de projets internes	36
I.4. Bilan	37
II. Projets scientifiques	39
II.1. Projets de recherche à court terme : projet MODAMDH.....	39
II.1.1. Définition du réservoir d'enzymes à cribler	40
II.1.2. Approches méthodologiques	41
II.1.3. Criblage in silico	42
II.1.4. Validation in-vitro des candidats et analyse in silico des AmDHs validées	42
II.2. Evolution des expertises/thématiques.....	43
II.2.1. Biotransformations.....	44
II.2.2. Projets axés vers la fermeture du cycle de carbone	46
III. Activités annexes.....	48
III.1. Activités d'encadrement	48
III.2. Activités d'enseignement	49
III.3. Responsabilités collectives	50
IV. Conclusion	50
V. Références.....	51
ANNEXE 1 Approche génomique utilisée pour la recherche d'enzymes.....	55
ANNEXE 2 Fonctionnement des GM3.....	57
ANNEXE 3 Publication soumise (Caparco et al).....	59

Introduction

Les organismes vivants ont fasciné et fascineront encore longtemps les chercheurs à la fois pour la complexité de leur fonctionnement, en particulier de leur métabolisme, et leur aptitude à produire une grande variété de métabolites secondaires. Ces derniers ont inspiré nombre de structures de principes actifs pharmaceutiques, de parfums, de produits agroalimentaires, du fait de leurs activités biologiques et/ou de leurs propriétés physico-chimiques. Cette énorme faculté de la nature à générer cette diversité structurale est une source d'inspiration qui rassemble biologistes et chimistes. Que ce soit ceux qui utilisent ces structures naturelles pour synthétiser des molécules plus ou moins analogues et tester leur activité, ceux qui s'intéressent à l'origine de ces produits naturels ou enfin ceux qui valorisent les capacités catalytiques de ces organismes à des fins synthétiques. Par ces trois aspects, on relie alors la biologie à trois grandes spécialités de la chimie de synthèse : la chimie médicinale (plus généralement la chimie de molécules biologiquement actives), la chimie biomimétique et la biocatalyse. Dans le cas de ces deux derniers pendants de la chimie organique, les voies de synthèse des molécules plus ou moins complexes sont pensées rétrosynthétiquement avec la connaissance des capacités du vivant. Pour la chimie biomimétique, ces étapes de synthèses sont réalisées par des réactifs/catalyseurs chimiques pas toujours dans des conditions environnementales biocompatibles. Et dans le cas de la biocatalyse, les transformations sont en partie ou totalement effectuées par les catalyseurs fournis par la nature, les enzymes, sous forme natives ou évoluées par ingénierie protéique.

L'activité biologique des métabolites secondaires est depuis tout temps utilisée par l'homme, d'abord de manière empirique *via* les médecines traditionnelles qui ont rapidement exploité leurs vertus thérapeutiques ou autre par la préparation d'extraits actifs, et depuis le 18^{ème} siècle et la naissance de la chimie moderne, par l'industrie pharmaceutique/agroalimentaire/cosmétique qui s'inspirent ou non de ces mêmes vertus en bénéficiant de la connaissance de leurs structures chimiques et des mécanismes biologiques associés. Depuis l'isolement de la morphine en Allemagne au 19^{ème} siècle, les chimistes n'ont cessé d'extraire et de déterminer la structure de métabolites secondaires, et d'étudier leurs propriétés biologiques pour créer entre autre des médicaments, plus ou moins analogues à ces molécules naturelles (Braña et al. 2006). La chimie biomimétique est quant à elle née à la fin du 19^{ème} siècle, avec les travaux de Collie sur l'origine des polycétides puis ceux de Robinson avec la synthèse de la tropinone à partir de produits initiaux très simples (De Brabander et al. 2011 et références associées). De là est née la preuve que la nature utilise de simples structures (= synthons, = « building blocks ») pour générer toute sa diversité structurale. Les synthèses imitant les réactions naturelles et les procédés enzymatiques, dans un but de construire des voies de synthèses plus efficaces que celles de la chimie conventionnelle, se sont alors développées dès les années 1970 et le concept de chimie biomimétique est apparu. Désormais, ce pendant de la chimie organique donne lieu à d'impressionnants résultats, où les voies d'accès imaginées d'un point de vue global selon les structures des métabolites, contrastent avec les synthèses « étape par étape » de la chimie classique (Nay et al. 2011 et références associées). Les travaux sur la biosynthèse des composés naturels révèlent ainsi souvent le lien entre plusieurs métabolites, comme j'ai pu le découvrir pendant mes travaux de thèse. Quant à la biocatalyse, on peut considérer qu'elle a débuté il y a plus d'un siècle avec l'idée que les composés des cellules vivantes pourraient être utilisés pour des transformations chimiques (ou bien plus tôt si l'on englobe les procédés de fermentations) (Bornscheuer et al. 2012 et références

associées). La synthèse du (*R*)-mandelonitrile à partir du formaldéhyde et de cyanure d'hydrogène en présence d'extrait de plante en 1908 par Rosenthaler est un des premiers exemples. Malheureusement, son développement est resté limité jusqu'à dans les années 1970 du fait de la difficulté à obtenir les enzymes en quantité suffisante pour comprendre les réactions mises en jeu et ainsi permettre le développement d'applications. Ces dernières n'étaient alors obtenables uniquement qu'à partir de l'organisme producteur naturel. La méthodologie de l'ADN recombinant, qui permet d'exprimer une enzyme d'un autre organisme dans un organisme hôte, a permis de déverrouiller la biocatalyse dans les années 1970. Mais le retard pris par rapport à la chimie conventionnelle fait qu'encore aujourd'hui, la biocatalyse n'est pas universellement considérée lors des constructions des voies de synthèses. Néanmoins, les technologies d'ingénierie protéique, d'évolution dirigée dès la fin des années 1990 (qui permettent de modifier à grande échelle les biocatalyseurs) (Cirino et al. 2002), et maintenant le faible coût de synthèse des gènes et le développement d'outils bio-informatiques et d'analyses à haut débit, permettent d'envisager la biocatalyse comme une chimie de synthèse au même niveau que la chimie conventionnelle (Reetz 2013). Nous sommes en train de vivre l'âge d'or de la biocatalyse, pour reprendre les termes de N. Turner et R. Kumar (Turner et al. 2018), qui devient une réelle perspective, particulièrement en chimie pharmaceutique (Devine et al. 2018, Braun et al. 2019). Les connaissances grandissantes issues des séquences génomiques et métagénomiques et des structures tridimensionnelles des enzymes, faciliteront la construction des biocatalyseurs de demain (Bornscheuer 2018).

De plus en plus d'industries pharmaceutiques considèrent les voies biocatalytiques lors de l'élaboration des schémas de synthèses des principes actifs, notamment du fait du contexte international de transition énergétique, où la communauté chimique se doit de fabriquer les composés de manière économe en énergie. De nombreux produits sont d'ailleurs déjà produits par biocatalyse (Bornscheuer et al. 2012) et ceci devrait s'accroître avec les progrès considérables en ingénierie protéique. Mais plus de liens doivent aussi se créer entre la chimie biomimétique et la biocatalyse. Les hypothèses établies par la chimie biomimétique sont une excellente porte d'entrée pour les chimistes de biocatalyse pour construire des voies rétrosynthétiques alternatives (Turner et al. 2013) et rechercher l'existence d'autres activités enzymatiques. L'exemple des Pictet-Spenglerase illustre cette connection (Stockigt et al. 2011).

C'est dans ce triangle chimie pharmaceutique/chimie biomimétique/biocatalyse, directement relié aux organismes vivants et à leur métabolisme primaire et secondaire, que mon parcours professionnel se situe. Après une formation générale de chimiste et des premières expériences en chimie pharmaceutique, j'ai rejoint le vaste champ de la chimie biomimétique et de l'extraction des métabolites secondaires lors de mon expérience doctorale. Après un nouveau passage par la chimie médicinale, j'ai alors pris le parti de participer à l'expansion de la biocatalyse en France pour maintenant pouvoir activement proposer des solutions alternatives de synthèses utilisant les catalyseurs des organismes vivants, plus particulièrement les enzymes issues de procaryotes. L'ensemble de ce parcours professionnel ainsi que les activités d'encadrement associées sont présentés dans ce document, complété d'une description des projets de recherche que je souhaite mener par la suite, axés sur la biocatalyse et ses nouvelles technologies associées qui offrent des possibilités de recherches encore plus diverses.

Curriculum vitae

Carine VERGNE-VAXELAIRE

Laboratory of Biocatalysis, Bioremediation and Synthetic Metabolism
Génomique Métabolique, Genoscope, Institut François Jacob
CEA, CNRS, University Evry, University Paris-Saclay
Evry, FRANCE
+33-686789870
carine.vergne@genoscope.cns.fr
French, 41 years old



Expertises

<i>Scientific</i>	<i>Management/project</i>	<i>Transversal</i>
Organic chemistry Biocatalysis Biochemistry	Project leading Supervision of PhD students, M2, students and technician	Team spirit Meaning of analysis Proposal force

Main Research Interests

- Exploration of the enzymatic resource among the sequenced bacterial biodiversity
- Discovery of novel enzymes catalyzing difficult reactions in conventional synthesis and their applications as biocatalysts in biotechnology and production of key building blocks. Recent work was focused on:
 - Amine dehydrogenases (formation of chiral amines from prochiral ketones and ammonia)
 - Adenylating enzymes for amide bond formation
 - α -Ketoglutarate-dependent dioxygenases and their use in cascade reactions with decarboxylases to access hydroxylated amines
 - Nitrilases to access carboxylic acid from nitriles under mild conditions

In each case, a multidisciplinary approach is undertaken, involving different expertises present in the research unit (genome mining, bioinformatics, molecular biology and synthetic chemistry) or external collaborations (X-ray crystallography, chemical process)

Professional experience and positions

Since 12-2007: CEA researcher, Laboratory of Biocatalysis Bioremediation and Synthetic Metabolism (CEA, Genoscope, Génomique Métabolique, Evry, France)
project leader in biocatalysis
5 PhD and 8 student supervision

skills in synthetic chemistry, analytical chemistry (UHPLC-UV-MS, GC-MS-FID, spectrophotometry, NMR), screening assays and biochemistry, biocatalysis, genome mining, structural analysis of enzyme
co-implementation of the chemistry/biocatalysis laboratory

2006-2007: Industrial post-doctoral position (Chemistry Division of P. Casara, **Institut de Recherche Servier**, Croissy-Sur-Seine, France) under the supervision of Thierry-Le-Diguarher
medicinal chemistry (oncology)

2003-2006: PhD student (**Institut de Chimie des Substances Naturelles**, CNRS, Gif-sur-Yvette, France)
PhD project: *Research of biosynthetic intermediates of pyrrole-2-aminoimidazoles by extraction of marine sponges and biomimetic synthesis (Dir: Ali Al-Mourabit)*
Extraction of marine sponges
Purification of new metabolites (silica columns, HPLC)
Structural elucidation of new compounds (NMR)
Synthetic chemistry
Biomimetic strategies

2003: *master degree traineeship*, (Chemistry Division of J.-L. Pégliion, **Institut de Recherche Servier**, Suresnes, France) under the supervision of Christophe Poitevin
medicinal chemistry (cardiology)

2001-2002: chemist Engineer, Millennium Pharmaceuticals (Cambridge, UK)
medicinal chemistry
combinatorial chemistry

Education

2006: PhD in chemistry, University of Paris-Sud (Orsay, France)
2003: Engineer in chemistry (Chemistry Physics and Electronic Engineer School CPE (Lyon, France)
2003: Master 2 degree in Fine Organic Chemistry (University Lyon I, France)

Supervising activities

5 PhD students (Franck Bordier 2010-2013, Ombeline Mayol 2015-2018, Chloé Lelièvre 2017-2020, Adam Caparco in cosupervision 2018-2019, Laurine Ducrot 2019-2022)
4 students in Master degree

Collaborations

National: Marielle Lemaire and Christine Guerard-Helaine (Clermont Université - Université Blaise Pascal, Institut de Chimie de Clermont-Fd, UMR 6296 CNRS)

International :

- Gideon Grogan (York Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York, UK)
- Andreas S. Bommarius (Georgia Tech School of Chemical and Biomolecular Engineering, Atlanta, US): Chateaubriand fellowship obtained by the PhD student Adam Caparco
- Shizhen Wang (Department of Chemical and Biochemical Engineering, Xiamen University, China): Fellowship "Bourse Jean d'Alembert" from the Paris-Saclay University
- Caroline Paul (Biocatalysis, Department of Biotechnology, TU Delft, The Netherlands)
- Jack Rowbotham, Kylie Vincent (Université d'Oxford, UK)

Collective Responsibilities

Member of the laboratory council of the UMR8030 since 2013

List of publications (22 + 1 patent)

Lelièvre, C., Balandras, M., Petit, J.-L., **Vergne-Vaxelaire, C.***, Zaparucha, A.* ATP regeneration system in chemoenzymatic amide bond formation with thermophilic CoA ligase (2020). ChemCatChem <https://doi.org/10.1002/cctc.201901870>

Domergue, J., Erdmann, D., Fossey-Jouenne, A., Petit, J.-L., Debard, A., de Berardinis, V., **Vergne-Vaxelaire, C.**, Zaparucha, A. XszenFHal, a novel tryptophan 5-halogenase from *Xenorhabdus szentirmaii* (2019). AMB Express 9, 175. <https://doi.org/10.1186/s13568-019-0898-y>

Laurent, V., Hélaine, V., **Vergne-Vaxelaire, C.**, Nauton, L., Traikia, M., Petit, J.-L., Salanoubat, M., de Berardinis, V., Lemaire, M., Guérard-Hélaine, C. (2019) Achiral Hydroxypyruvaldehyde Phosphate as a Platform for Multi-Aldolases Cascade Synthesis of Diuloses and for a Quadruple Acetaldehyde Addition Catalyzed by 2-Deoxyribose-5-Phosphate Aldolases. ACS 9, 10, 9508-9512. <https://doi.org/10.1021/acscatal.9b02668>

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. Nat. Catal. 324, 2, 324–333. <https://doi.org/10.1038/s41929-019-0249-z>

Vergne-Vaxelaire, C., de Berardinis, V., Zaparucha, A. (2019) Method for preparing amines from aldehydes and ketones by biocatalysis. WO2019008110.

Zaparucha, A., de Berardinis, V., **Vaxelaire-Vergne, C.** (2018). Chapter 1. Genome Mining for Enzyme Discovery, in *Modern Biocatalysis: Advances Towards Synthetic Biological Systems*. (RSC, Cambridge, UK, 2018). <https://pubs.rsc.org/en/content/ebook/978-1-78262-726-5>

Vergne-Vaxelaire, C., Mariage, A., Petit, J.L., Fossey-Jouenne, A., Guerard-Helaine, C., Darii, E., Debard, A., Nepert, S., Pellouin, V., Lemaire, M., et al. (2018). Characterization of a thermotolerant ROK-type mannofructokinase from *Streptococcus mitis*: application to the synthesis of phosphorylated sugars. Appl Microbiol Biotechnol. 102, 3, 5569-5583. <https://doi.org/10.1007/s00253-018-9018-1>

Fossey-Jouenne, A., **Vergne-Vaxelaire, C.**, Zaparucha, A. (2018). Enzymatic cascade reactions for the synthesis of chiral amino-alcohols from L-lysine. J. Vis. Exp. 132, e56926. doi:10.3791/56926

Baud, D., Peruch, O., Saaidi, P.-L., Fossey, A., Mariage, A., Petit, J.-L., Salanoubat, M., **Vergne-Vaxelaire, C.**, de Berardinis, V. and Zaparucha, A. (2017). Biocatalytic Approaches towards the Synthesis of Chiral Amino Alcohols from Lysine: Cascade Reactions Combining alpha-Keto Acid Oxygenase Hydroxylation with Pyridoxal Phosphate- Dependent Decarboxylation. Adv. Synth. Catal. 359, 1563-1569. <https://doi.org/10.1002/adsc.201600934>

Bastard, K., Perret, A., Mariage, A., Bessonnet, T., Pinet-Turpault, A., Petit, J.-L., Darii, E., Bazire, P., **Vergne-Vaxelaire, C.**, Brewée, C., et al. (2017). Parallel evolution of non-homologous isofunctional enzymes in methionine biosynthesis. Nature chemical biology 13, 858. <https://doi.org/10.1038/nchembio.2397>

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 *Petrogorgia mobilis*. Catalysis Science & Technology 6, 7421-7428. doi: 10.1039/C6CY01625A

Bordier, F., Stam, M., Darii, E., Tricot, S., Fossey, A., Rohault, J., Debard, A., Mariage, A., Pellouin, V., Petit, J.-L., Perret, A., Vallenet, D., Salanoubat, M., Weissenbach, J., **Vergne-Vaxelaire, C.**, de Berardinis, V., Zapparucha, A. (2014). Large α -aminonitrilase activity screening of nitrilase superfamily members: Access to conversion and enantiospecificity by LC–MS. *Journal of Molecular Catalysis B: Enzymatic* 107, 79-88. <https://doi.org/10.1016/j.molcatb.2014.05.019>

Bastard, K., Smith, A.A.T., **Vergne-Vaxelaire, C.**, Perret, A., Zapparucha, A., De Melo-Minardi, R., Mariage, A., Boutard, M., Debard, A., Lechaplais, C., *et al.* (2014). Revealing the hidden functional diversity of an enzyme family. *Nature chemical biology* 10, 42-49. <https://doi.org/10.1038/nchembio.1387>

Vergne-Vaxelaire, C., Bordier, F., Fossey, A., Besnard-Gonnet, M., Debard, A., Mariage, A., Pellouin, V., Perret, A., Petit, J.-L., Stam, M., *et al.* (2013). Nitrilase Activity Screening on Structurally Diverse Substrates: Providing Biocatalytic Tools for Organic Synthesis. *Adv. Synth. Catal.* 355, 1763-1779. <https://doi.org/10.1002/adsc.201201098>

Perret, A., Lechaplais, C., Tricot, S., Perchat, N., **Vergne, C.**, Pellé, C., Bastard, K., Kreimeyer, A., Vallenet, D., Zapparucha, A., *et al.* (2011). A Novel Acyl-CoA Beta-Transaminase Characterized from a Metagenome. *PLoS one* 6, e22918. <https://doi.org/10.1371/journal.pone.0022918>

Bellinzoni, M., Bastard, K., Perret, A., Zapparucha, A., Perchat, N., **Vergne, C.**, Wagner, T., de Melo-Minardi, R.C., Artiguenave, F., Cohen, G.N., *et al.* (2011). 3-Keto-5-aminohexanoate cleavage enzyme: a common fold for an uncommon Claisen-type condensation. *The Journal of biological chemistry* 286, 27399-27405. doi: 10.1074/jbc.M111.253260

Fonknechten, N., Perret, A., Perchat, N., Tricot, S., Lechaplais, C., Vallenet, D., **Vergne, C.**, Zapparucha, A., Le Paslier, D., Weissenbach, J., *et al.* (2009). A Conserved Gene Cluster Rules Anaerobic Oxidative Degradation of L-Ornithine. *Journal of Bacteriology* 191, 3162-3167. <https://doi.org/10.1128/JB.01777-08>

Ermolenko, L., Zhaoyo, H., Lejeune, C., **Vergne, C.**, Ratinaud, C., Nguyen, T.B., Al-Mourabit, A. (2014). Concise synthesis of Didebromohamacanthin A and Demethylaplysinsopine: Addition of Ethylenediamine and Guanidine Derivatives to the Pyrrole-Amino Acid Diketopiperazines in Oxidative Conditions. *Organic Letters* 16, 3, 872-875. <https://doi.org/10.1021/ol4036177>

Tian, H., Ermolenko, L., Gabant, M., **Vergne, C.**, Moriou, C., Retaillau, P., Al-Mourabit, A. (2011). Pyrrole-Assisted and Easy Oxidation of Cyclic α -Amino Acid-Derived Diketopiperazines under Mild Conditions. *Advanced Synthesis & Catalysis*, 353, 9, 1525-1533. <https://doi.org/10.1002/adsc.201100112>

Ratinaud, C., **Vergne, C.**, Ermolenko, L., Al-Mourabit, A. (2008). Chemiluminescent oxidative rearrangement of diketopiperazines involved in the biomimetic synthesis of marine pyrrole-2-aminoimidazole metabolites. *Luminescence*, 23, 88.

Vergne, C., Appenzeller, J., Ratinaud, C., Martin, M.-T., Debitus, C., Zapparucha, A., Al-Mourabit, A. (2008). Debromodispacamides B and D: Isolation from the Marine Sponge *Agelas mauritiana* and Stereoselective Synthesis Using a Biomimetic Proline Route. *Organic Letters*, 10, 3, 493-496. <https://doi.org/10.1021/ol702866m>

Vergne, C., Boury-Esnault, N., Perez, T., Martin, M.-T., Adeline, M.-T., Tran Huu Dau, E., Al-Mourabit, A. (2006). Verpacamides A–D, a Sequence of C11N5 Diketopiperazines Relating Cyclo(Pro-Pro) to Cyclo(Pro-Arg), from the Marine Sponge *Axinella vaceleti*: Possible Biogenetic Precursors of Pyrrole-2-aminoimidazole Alkaloids. *Organic Letters*, 8, 11, 2421-2424. <https://doi.org/10.1021/ol0608092>

Girard, E., Jaimovich, E., **Vergne, C.**, Al-Mourabit, A., Laurent, D., Molgo, J. (2004). Alcaloïdes bioactifs extraits d'éponges : les xestospongines dans l'étude de la signalisation du Ca^{2+} intracellulaire dépendante de l'inositol 1,4 ;5-triphosphate [Bioactive alkaloids extracted from sponges: the

xestospongins in the study of intracellular Ca^{2+} signalling mediated by inositol 1,4,5-trisphosphate]. *Envenimations, Intoxications* 141-152 (Coll. Rencontres en toxicologie). <https://hal-mnhn.archives-ouvertes.fr/NBCM/hal-00184032v1>

Oral communications (OC) and posters (P)

OC_Amine Biocatalysis 4.0, 2020, Stuttgart (Allemagne)
OC Speed talk_Novel Enzymes, 2018, Darmstadt (Allemagne)
OC_AmineBioCat, 2017, Manchester (UK)
OC_BioTrans, 2017, Budapest (Hungary)
OC_Novel Enzymes, 2016, Groningen (Netherlands)
OC_CBSO, 2014, Carry-Le-Rouet (France)
P_BioTrans, 2013, Manchester (UK)

Prize, distinction

Poster Prize Green Chemistry, 2018, Novel Enzymes (Darmstadt, Allemagne)
Junior Expert CEA (biocatalyse, chimie), 2019

Reviewing

4 papers (ACS catalysis 2019, ChemBioChem 2019, Catalysts 2019, Catalysis Comm. 2015)
External examiner of Benjamin Rowlinson's M.Sc. report (M. Sc. by research, G. Grogan's team)

I. Activités scientifiques de recherche

L'ensemble de mes activités scientifiques de recherche conduites depuis ma formation initiale se répartissent en trois thématiques :

- la synthèse organique de molécules biologiquement actives
- la synthèse biomimétique et l'extraction de produits naturels
- la synthèse biocatalysée de synthons chimiques

Ces trois axes de recherche ont été menés quasi chronologiquement au cours de mes expériences en formation et en vie professionnelle. Ils sont détaillés dans les paragraphes suivants en mettant l'accent sur ce lien « de la molécule à sa synthèse » qui les relie et où l'interface chimie-biologie prend tout son sens. Mon implication dans ces trois thématiques est une richesse pour mes recherches. Comme le mentionne Marc Fontecave, d'après des réflexions de Jean-Marie Lehn, les découvertes qui auront le plus d'impact par leur diversité et leur complexité sont celles qui naîtront aux frontières de ces deux disciplines, *ie* en biologie synthétique/ingénierie métabolique et en chimie bio-inspirée/enzymologie artificielle.

I.1. Synthèse organique de molécules biologiquement actives

Même si mon expérience dans le domaine de la synthèse médicamenteuse reste très limitée, je la mentionne ici car elle a réellement influé sur mon parcours professionnel et elle continue d'impacter mes recherches actuelles. Comme discuté en introduction, la synthèse organique de molécules biologiquement actives est un pilier important qu'il est bénéfique de connaître pour appréhender au mieux les recherches à l'interface chimie-biologie.

Lors de ma formation d'ingénieur chimiste, j'ai tout d'abord souhaité découvrir la synthèse organique à visée thérapeutique en effectuant un stage de 1 an dans l'entreprise Millenium Pharmaceuticals (Cambridge, UK). J'ai été impliquée activement dans trois projets de recherche. Un premier projet appliqué au traitement du diabète et de l'obésité m'a permis de me familiariser avec les réactions chimiques d'amination réductrice, de couplage acide / amine et d'oxydations d'alcool, toutes très courantes en synthèse organique. Lors d'un second projet de recherche plus en amont, j'ai été responsable de la validation expérimentale d'une nouvelle synthèse chimique sur voie solide nécessaire à l'élaboration d'une bibliothèque de composés. Enfin, j'ai été affectée dans une équipe créée rapidement pour la validation et l'optimisation de la synthèse en quantité importante d'une molécule alors très prometteuse dans le domaine de la cancérologie (structurale générale aminopyridine présentée en figure 1). L'efficacité et le travail en étroite liaison entre tous les membres de l'équipe étaient alors les mots d'ordre du projet pour respecter les courtes échéances imposées.

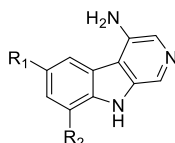


Figure 1. Structure générique de la molécule cible (Millenium Pharmaceuticals)

Lors de cette riche expérience en laboratoire pharmaceutique, j'ai donc développé mes connaissances en chimie organique, ma pratique en laboratoire et utilisé quotidiennement les techniques analytiques utiles au suivi des réactions chimiques (Chromatographie Liquide Haute Performance (HPLC), éventuellement couplée à la spectrométrie de masse (MS), et Résonance Magnétique Nucléaire (RMN) principalement). L'évolution constante des projets industriels m'a forcé à m'adapter rapidement aux nouvelles situations rencontrées. Enfin, cette expérience en pays anglosaxon a largement contribué à ma facilité actuelle à communiquer au sein de la communauté scientifique internationale.

J'ai poursuivi ce volet « synthèse médicinale » lors de deux expériences au sein de centres de recherche de l'Institut de Recherche Servier en région parisienne. La première s'est déroulée en 2003 au centre de Suresnes lors de mon stage de fin d'étude d'ingénieur/DEA (8 mois) dans la direction de Jean-Louis Pégion sous la supervision de Christophe Poitevin, chercheur dans cette direction. Le but de ce travail était de synthétiser des successeurs potentiels au principe actif ivabradine du Procoralan®, médicament récemment découvert par cette division qui a reçu l'Autorisation Européenne de Mise sur le Marché deux ans plus tard en 2005. L'ivabradine est un bradycardisant indiqué dans l'insuffisance cardiaque ou coronaire. Il permet de diminuer la fréquence cardiaque et donc la demande du cœur en oxygène, sans modifier les paramètres hémodynamiques contrairement aux autres médicaments de cette famille alors commercialisés (Vilaine et al. 2003). Les sites d'action uniques de l'ivabradine sont les canaux HCN (« Hyperpolarization-activated Cycloc Nucleotide-gated channels ») situés dans les cellules pacemaker de l'oreillette droite. Ces canaux permettent le passage des courants ioniques lors de la polarisation et la dépolarisation de ces cellules responsables des battements cardiaques. La molécule ivabradine a la capacité d'inhiber sélectivement ces canaux HCN et non les dix autres types de canaux également impliqués lors d'un battement cardiaque, et donc de moduler la fréquence cardiaque sans modifier les autres paramètres cardiovasculaires. Sa structure chimique est un squelette benzazepinone relié par une chaîne alkyle à une partie benzocyclobutane substituée (Figure 2a).

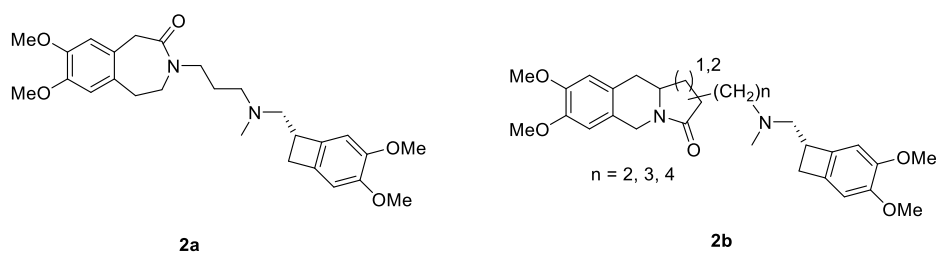


Figure 2. **2a** : Structure de l'ivabradine. **2b** : Structure des analogues cibles (Servier, Suresnes)

Malgré son activité démontrée et sa validation en tant que médicament, il était nécessaire de trouver une seconde génération suite aux observations faites pendant son développement clinique. En particulier, l'ivabradine est totalement métabolisée par l'enzyme 3A4 du cytochrome P450, or cette enzyme métabolise d'autres médicaments qui peuvent parfois être co-administrés au malade avec le Procoralan®. Une compétition peut donc avoir lieu lors de la métabolisation et provoquer une concentration trop élevée en principe actif dans le corps. De plus, il entraîne parfois des problèmes visuels car certains sous-types de canaux HCN se trouvent aussi au niveau de la rétine. Le but était donc de trouver une molécule analogue qui ait toujours la même activité, la même sélectivité mais qui soit métabolisée par une autre enzyme que l'enzyme 3A4 et plus sélective des canaux HCN du cœur (HCN4 contre HCN1-2 de la rétine). Les analogues cibles étaient des variants structuraux du tricycle

benzazepinone, le tricycle lactame « 665 » et lactame « 666 » (Figure 2b). Une recherche bibliographique a permis de choisir deux stratégies de synthèse prioritaires pour ces tricycles ainsi que leur fonctionnalisation pour accéder aux analogues cibles de l'ivabradine. Après plusieurs essais sur les deux voies de synthèses, une a été validée expérimentalement en série racémique (9 et 12 étapes pour « 665 » « 666 » respectivement) et étudiée avec succès en série chirale pour accéder aux deux énantiomères générés. On a ainsi obtenu un précurseur suffisamment versatile pour envisager la synthèse en un nombre limité d'étapes d'analogues. Un des diastéréoisomères a pu être synthétisé avec une grande pureté pour être testé biologiquement (Figure 2b). Le large éventail de réactions chimiques mises en pratique pendant cette expérience m'a permis de continuer à développer mes connaissances théoriques et pratiques en synthèse organique multi-étapes tout en travaillant dans un environnement industriel propice aux découvertes de technologies, notamment d'aide à la purification des produits synthétisés et d'analyses.

Pour la seconde expérience dans ce même Institut, j'ai travaillé au centre de Croissy-sur-Seine dans la direction de Patrick Casara sous la supervision de Thierry Le-Diguarher, en tant que cadre post-doctoral (octobre 2006 à décembre 2007). La thématique du projet de recherche qui m'a été confié, avec 4 autres personnes de l'équipe, concernait la synthèse de nouvelles molécules anticancéreuses agissant sur l'apoptose des cellules. A la différence du projet précédent, ce dernier était bien plus avancé d'un point de vue développement, se situant dans la catégorie « lead optimization ». Une démarche de Relation-Structure-Activité (SAR) était entreprise pour optimiser les propriétés pharmacologiques d'une molécule référence sélectionnée pour cette application (Figure 3). L'ensemble de la synthèse (10 à 15 étapes) devait être réalisée jusqu'à l'obtention des molécules finales, testées biologiquement par l'équipe de pharmacologues responsable du projet. La stratégie de synthèse étant établie et validée, la chimie expérimentale était donc moins variée que celle du projet de DEA, mais nécessitait une grande rigueur du point de vue des rendements et des puretés des produits finaux qui devaient respecter un cahier des charges de qualité pour pouvoir être testés. En plus de cet important travail expérimental de laboratoire soumis à des pressions de résultats, j'étais en relation avec les pharmacologues du projet pour suivre l'évolution des tests de ces molécules. Etant donné le stade de développement avancé, des collaborations avec des équipes d'Oril Industrie et de Servier Orléans étaient aussi entreprises pour appréhender les synthèses sur plus grande échelle, et tester des alternatives enzymatiques *via* l'utilisation de lipases. Mon activité de recherche actuelle en biocatalyse me permet de côtoyer de nouveau certains de ces interlocuteurs industriels de Servier investis dans le domaine de la biocatalyse.

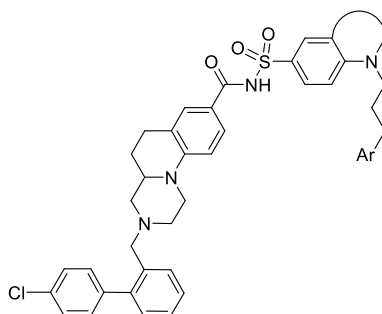


Figure 3. Structure des molécules cibles (Servier, Croissy-Sur-Seine)

Pour des raisons de confidentialité, je n'ai pu garder de notes détaillées de ces expériences en chimie médicinale, exceptée de la deuxième pour mon stage de DEA. Je ne peux donc restituer le détail des synthèses entreprises mais les acquis de laboratoire, de connaissances de chimie de synthèse et du travail sur projet par équipe sont toujours présents. Ils me permettent actuellement d'assurer ou encadrer les travaux de synthèses conventionnelles lorsqu'elles sont requises pour l'équipe ou même l'unité de recherche. Mais surtout ces acquis sont bénéfiques pour l'orientation des projets actuels en biocatalyse de l'unité. Ils m'apportent un regard critique sur les étapes et structures clés en synthèse chimique, point souvent à l'origine de nos projets.

1.2. Synthèse biomimétique et extraction de produits naturels (doctorat)

Mon attrait pour la complexité et diversité des structures chimiques synthétisées par les organismes vivants m'ont incité à découvrir cet autre volet de la chimie organique qu'est la synthèse biomimétique et la recherche de nouveaux métabolites secondaires. Forte de l'obtention d'un financement CNRS pour effectuer mon expérience doctorale dans l'un des centres de recherche clé dans ce domaine, l'Institut de Chimie des Substances Naturelles (ICSN), j'ai pu effectuer des travaux de recherche dans ce domaine, sous la direction de Ali Al-Mourabit (octobre 2003 à octobre 2006).

Lors de ce travail de recherche, je me suis consacrée à l'extraction d'éponges marines de la famille des Agelasidae et Axinellidae et à la synthèse biomimétique d'alcaloïdes pyrrole-2-aminoimidazoles isolés uniquement de ces familles d'éponges.

1.2.1. Présentation succincte des métabolites ciblés

L'intérêt des métabolites secondaires marins est double. Tout d'abord ils possèdent des structures originales souvent sans antécédent dans le monde terrestre et des activités biologiques démontrées (antitumorales, anti-inflammatoire, analgésiques, antivirales...). Leur potentialité comme source d'inspiration de molécules à visée thérapeutique est indéniable. Les structures complexes de ces produits impliquent de mieux connaître leurs biosynthèses pour pouvoir envisager de synthétiser en un nombre limité d'étapes ces molécules naturelles mais aussi des analogues.

Parmi les métabolites secondaires marins d'intérêts et plus particulièrement ceux contenant des atomes d'azote, les pyrrole-2-aminoimidazoles (P2AI) occupent une place remarquable. Cette famille largement étudiée d'un point de vue structures, biosynthèses, activités biologiques, a fait l'objet de nombreuses revues, dont la plus récente de 2017 par Lindel (Lindel 2017). Seules quelques publications sont citées ici, se référer à cette revue pour avoir une vision plus exhaustive.

Le premier représentant de cette vaste famille, la (-)-dibromophakelline, a été isolée en 1969 de l'éponge marine *Phakellia flabellata* (Figure 4a). Depuis, plus de 220 membres ont été isolés, exclusivement d'éponges marines et particulièrement des genres Agelasidae, Axinellidae et Halichondridae, et structurellement identifiés. Leur diversité structurale permet d'émettre des hypothèses de biogénèse. Il est ainsi apparu assez rapidement que les parents structuraux des membres complexes de cette famille sont la clathrodine et ses équivalents bromés hymenidine et oroidine (Figure 4b). Ce qui fut confirmé expérimentalement en 2012 par Molinski et Romo (Stout et

al. 2012). Ces molécules en $C_{11}N_5$ sont constituées d'une partie pyrrole carboxamide liée à un noyau 2-aminoimidazole *via* une chaîne propénamine. En plus de changement d'état d'oxydation/degré de bromation, différentes étapes de cyclisations, dimérisations ou tetramérisations permettent de relier ces précurseurs aux autres métabolites, alors usuellement classifiés en « monomères » (ex : dibromophakelline et dispacamide A, Figure 4a), « dimères » (ex : sceptrine et palau'amine, Figure 4c-d) et « tetramères » d'oroidine. Les hypothèses sur leur biosynthèse, initiées en 2001 par Al-Mourabit et Potier, (Al Mourabit et al. 2001) se poursuivent encore intensément aujourd'hui, principalement *via* des synthèses biomimétiques (Lindel 2017). Ces métabolites représentent de 0.001% à plus de 1% du poids sec d'éponge pour plus de 10% d'entre eux.

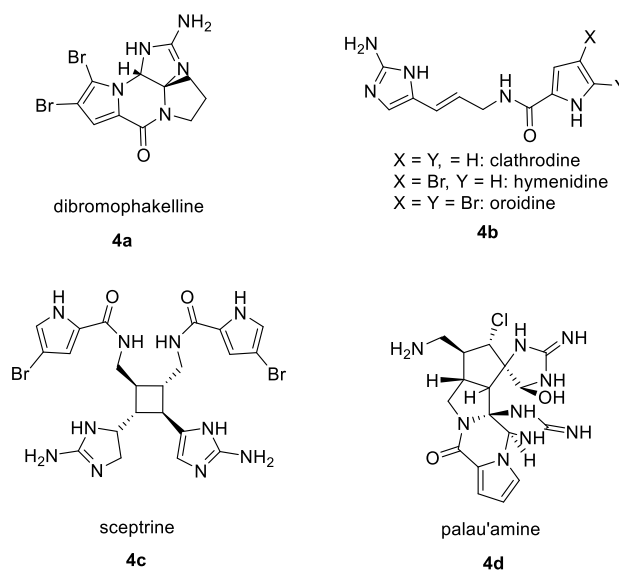


Figure 4. Structures de quelques P2AI

D'un point de vue activités biologiques, plusieurs ont été notées à leur isolement, puis des études ultérieures ont suivi, notamment grâce aux synthèses multi-grammes qui ont permis l'accès aux quantités nécessaires pour les tests. Des activités anti-cancéreuses ont par exemple été montrées pour la sceptrine *via* son activité inhibitrice de la motilité cellulaire des lignées de cellules cancéreuses sans être cytotoxique. La cytotoxicité de l'agelastatine A a été confirmée en 2013 sur un panel de lignées cellulaire humaines, en particulier leucémiques et lymphoïdes (IC_{50} : 20 -190 nM). Des activités antibactériennes des axinellamines, sur à la fois les bactéries Gram-positives et Gram-négatives, ont été validées en 2014, *via* un mode d'action basée sur la déstabilisation des membranes. Quant à l'agéliférine, ses propriétés inhibitrices de biofilms donnent lieu à la synthèse de nombreux analogues pour améliorer les activités.

Au début de mes travaux de thèse sur ce sujet en 2003, environ 100 P2AI avaient été recensés et caractérisés et les premières publications soulignant les liens biogéniques entre ces structures émergeaient juste, notamment par le groupe de Al-Mourabit (Al Mourabit et al. 2001, Travert et al. 2004). Pour étoffer ces réflexions et hypothèses biomimétiques sur cette famille, il était nécessaire de poursuivre l'exploration exhaustive des métabolites présents dans ces familles d'éponges.

Le projet de thèse s'inscrivait ainsi exactement dans cette démarche en se basant sur une stratégie d'étude double :

- la recherche de molécules potentiellement précurseurs de ces molécules complexes par extraction exhaustive d'éponges des familles considérées, puis identification structurale,
- la compréhension de la réactivité de ces molécules par synthèse biomimétique.

Ces deux moyens de recherche nous ont permis d'obtenir des résultats prometteurs et d'émettre des hypothèses de biosynthèse qui n'ont depuis pas été discréditées après l'isolement de plus de 100 autres P2AI et les expériences complémentaires discutées succinctement à la fin de cette partie.

1.2.2. Principaux résultats obtenus

L'extraction des métabolites secondaires d'éponges consiste expérimentalement à agiter vigoureusement l'éponge broyée puis lyophilisée dans un solvant le plus universel possible, afin d'obtenir après concentration du filtrat un extrait brut. Quand l'extraction est non-bioguidée, cet extrait brut est ensuite fractionné lors de purifications sur différents supports jusqu'à obtenir un maximum de composés suffisamment purs pour être analysés pour détermination structurale.

L'extraction de l'éponge *Agelas ceylonica* (269 g), extraite au méthanol et partitionnée par trituration au dichlorométhane, a fourni 5 g de composés apolaires type stérols, identifiés par chromatographie gazeuse (GC) couplée à la spectrométrie de masse (MS), et 68 g de composés polaires dont des sels. Une quantité exceptionnelle de sceptrine a été obtenue de cet extrait polaire (15 g purs à 90%), ainsi que deux autres composés connus, l'acide nakamurique et son ester méthylique. Les autres P2AI minoritaires détectés par HPLC/MS n'ont pas été isolés, leur structure semblant connue. Des études de réactivité menées sur la sceptrine isolée ont permis de confirmer sa transformation en agéliférine sous traitement micro-onde, comme montré par Baran et al auparavant (Figure 5) (Baran et al. 2004), mais aucun autre produit de transformation.

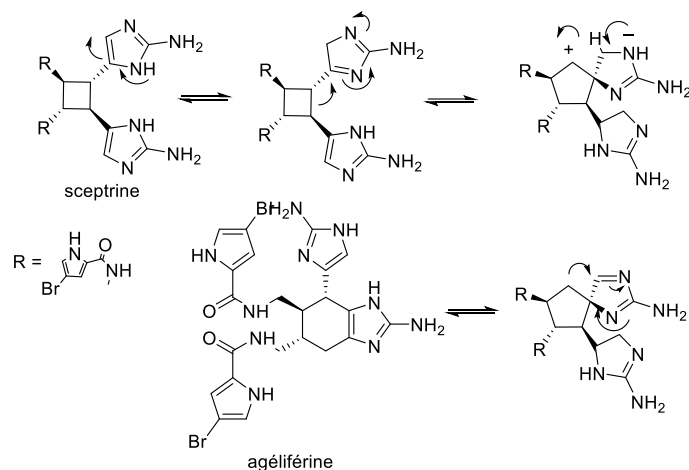


Figure 5. Hypothèse mécanistique proposée par Baran et al pour la transformation observée expérimentalement de sceptrine en agéliférine.

L'extraction de deux lots différents de l'éponge *Axinella vaceleti* (12 et 16 g seulement, l'éponge étant plus rare), a permis d'isoler 7 produits purs (1,2 – 185 mg), en plus des stérols non étudiés. Les produits majoritaires (4 – 185 mg) ont rapidement pu être identifiés car déjà répertoriés (Figure 6a). Quatre autres composés plus minoritaires (1,2 – 11,3 mg) se sont révélés être de nouveaux métabolites suite aux premières analyses en RMN 1D et Spectrométrie de Masse Haute Résolution (HRMS). Leur détermination structurale a nécessité un important travail d'analyses RMN 1D et 2D et une synthèse

chimique pour valider l'hypothèse d'une structure dicétopipérazine proline – arginine. Ils ont été nommés verpacamides A-D (Figure 6b).

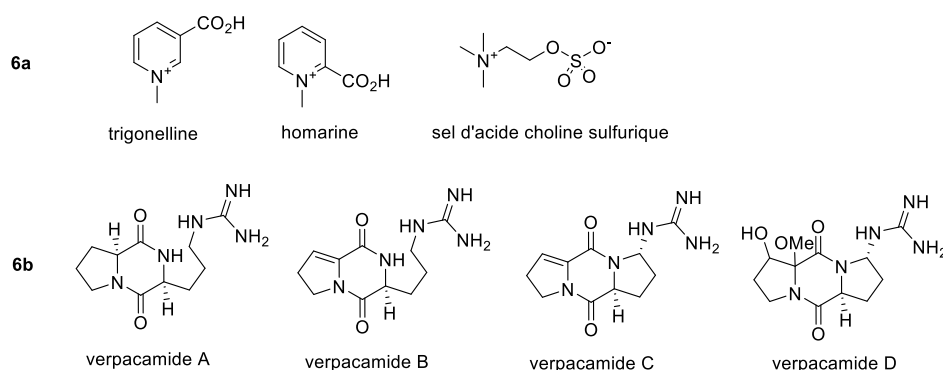


Figure 6. Métabolites isolés de l'éponge *Axinella vaceleti*.

Malgré le continuel questionnement sur l'origine microbienne/spongiaire des métabolites isolés d'éponges, un rôle des métabolites verpacamides A-D dans la formation des P2AI pouvait être proposé, ceux-ci étant aussi en $C_{11}N_5$. Même si plusieurs hypothèses biogéniques étaient posées pour la transformation de l'oroïdine en P2AI plus complexes, très peu de travaux avaient été conduits sur la biosynthèse elle-même du précurseur oroïdine. L'isolement des verpacamides A-D nous a permis de travailler sur ce point. Précédent ces travaux, une hypothèse biogénique basée sur l'oxydation spontanée des systèmes dicétopipérazines pyrrole/proline et leur rôle comme intermédiaires biosynthétiques de composés linéaires comme le dispacamide A, métabolite central de la famille des pyrrole-2-aminoimidazoles, avait été postulée et publiée (Travert et al. 2004). Une étude plus approfondie de cette oxydation m'a permis de mettre en évidence expérimentalement les intermédiaires postulés dioxétanone et la forme hydroperoxyde, et donc de conforter la viabilité de cette hypothèse de voie biosynthétique (Figure 7).

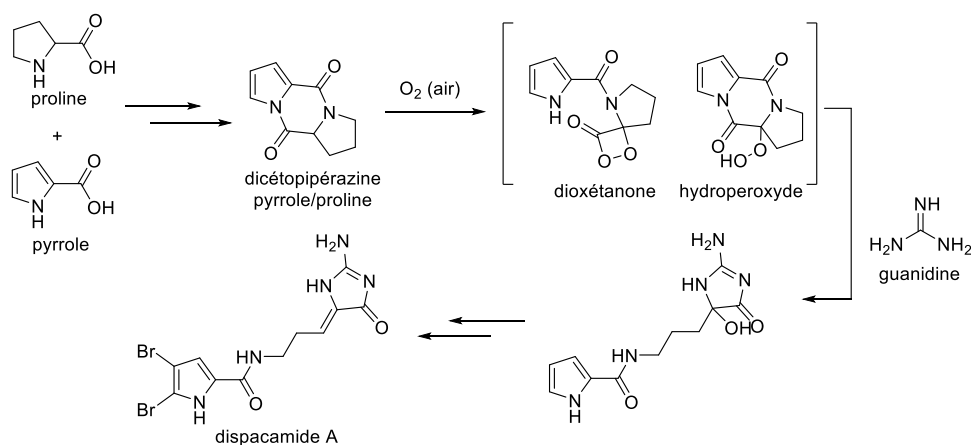


Figure 7. Hypothèse biosynthétique d'un précurseur du dispacamide A

Les précurseurs simples proline et guanidine disponibles dans la nature conduiraient à la dicétopipérazine qui s'oxyde spontanément pour donner ces intermédiaires oxydés. La dioxétanone très réactive est alors facilement attaquée par une entité nucléophile comme la guanidine pour conduire à l'analogue débromé du produit naturel dispacamide A. L'hypothèse biogénétique basée sur

ce réarrangement oxydatif d'unité proline a été confortée par l'isolement ultérieur des métabolites débromodispacamides B et D de l'éponge marine *Agelas mauritiana* collectée dans les eaux des îles Solomon. Même si ces hypothèses sont difficile à valider, il n'en reste pas moins que la chimie entreprise inspirée de celles-ci est forte : à partir d'unité proline (modifiée sous forme d'ester méthylique) et pyrrole, les débromodispacamide B et D ont été obtenus en une et deux étapes avec des rendements corrects (56 et 27% respectivement) en présence de carbonate de guanidine, avec passage probable par la dicétopipérazine pyrrole/proline oxydée (Figure 8).

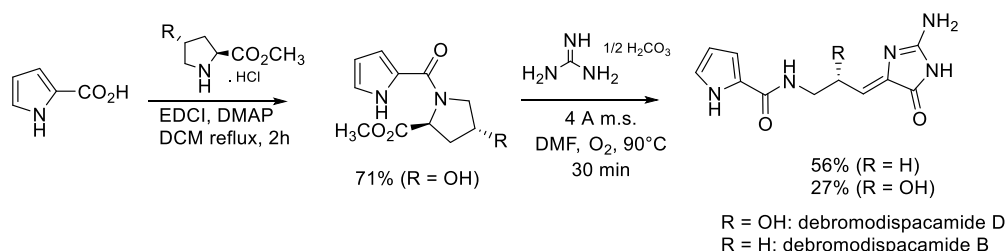


Figure 8. Synthèse biomimétique des débromodispacamide B et D

Cette même stratégie a mené à la publication d'une approche concise de synthèse d'autres dérivés de cette large famille, le didebromohamacanthine A et demethylaplysinsine (Ermolenko et al. 2014).

La meilleure compréhension de ce phénomène d'oxydation spontanée nous a permis de synthétiser rapidement plusieurs analogues en faisant varier les nucléophiles ou l'acide aminé proline. Une petite chimiothèque de produits analogues à des produits naturels possédant des activités biologiques a ainsi été constituée (Figure 9).

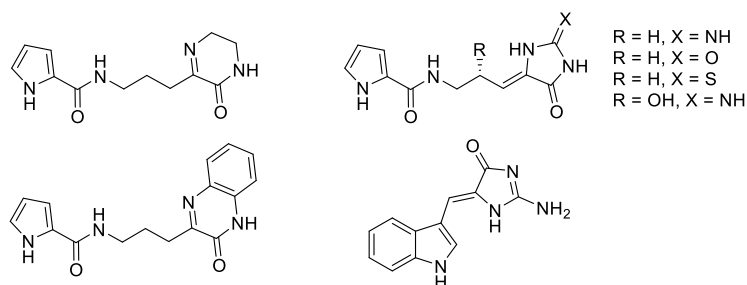


Figure 9. Analogues synthétisés

Une étude détaillée, ultérieure à ces recherches, de la réactivité des dérivés α -aminoacides vis-à-vis du dioxygène en absence d'activateur, a montré la peroxydation aérobie aisée en position α des α -aminoacides acylés par un acide pyrrolo-carboxylique par le dioxygène dans le diméthylformamide. Cette réactivité est généralisable à une grande variété de dérivés pyrrole-aminoacides (Tian et al. 2011).

L'isolement des verpacamides A-C au sein d'éponges de la même famille nous a permis d'émettre l'hypothèse que cette oxydation spontanée existe biosynthétiquement de manière intramoléculaire à partir de composé type proline-proline-guanidine (Figure 10) (Vergne et al. 2006). Cette hypothèse avancée pour relier les unités acides aminés aux P2AI contrastait avec les autres hypothèses reportées dans la littérature qui ne faisaient pas intervenir les dicétopipérazines comme intermédiaires (Kitagawa et al. 1983, Andrade et al. 1999).

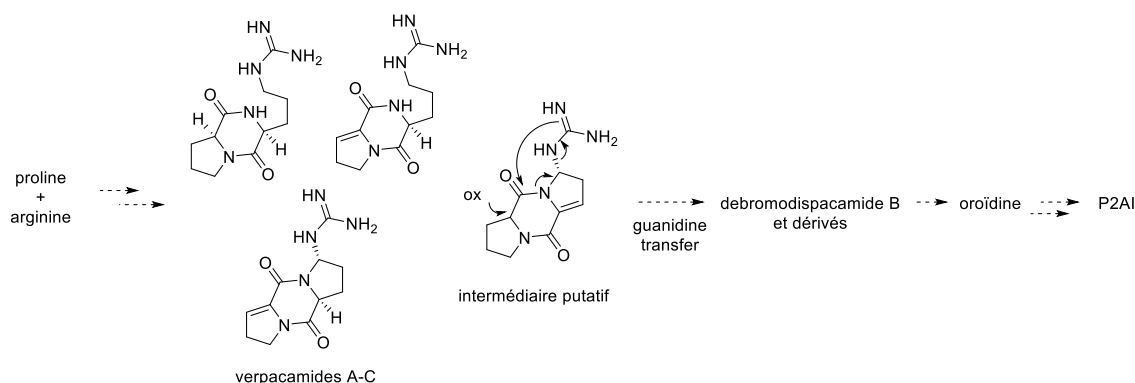


Figure 10. Verpacamides A-C comme intermédiaires possibles des métabolites dispacamide A et analogues

Des résultats ultérieurs obtenus après incubation de l'éponge *Axinella damicornis* dans de l'eau de mer supplémentée par de la proline, arginine, lysine, histidine et ornithine, marquées au carbone 14, montrent, pour la 2-aminoimidazole des P2AI, une bien plus grande incorporation de la lysine que de l'arginine ou de l'ornithine, et aucune incorporation d'histidine au contraire des expériences précédemment réalisées par Kerr (Andrade et al. 1999). Cette expérience a confirmé la biosynthèse du pyrrole à partir de la proline (Genta-Jouve et al. 2011). Ces résultats n'annulent pas l'hypothèse biomimétique avancée, mais ne la conforte pas non plus en faisant intervenir un acteur non prévu, la lysine. Les incorporations mesurées étant très faibles, du même ordre de grandeur que celles observées par Kerr (Andrade et al. 1999), ce type d'expérience n'a pas été poursuivi. D'autres études auraient été nécessaires pour également étudier une éventuelle influence de l'espèce étudiée et l'impact des conditions de culture sur les résultats. Ce point de l'impact de l'espèce utilisée a été souligné par Molinski et al lors de la mise en évidence du squelette oroïdine comme précurseur de P2AI dimériques, où seulement certains des extraits brut d'éponge de la famille donnaient ces résultats (Stout et al. 2012).

Beaucoup d'efforts sont entrepris depuis 2005 pour accomplir la synthèse totale des molécules complexes P2AI, notamment avec l'aide des hypothèses biosynthétiques, qui inspirent notamment les étapes les plus critiques (Nay et al. 2011, Lindel 2017).

Les courtes expériences professionnelles réalisées en synthèse médicinale m'ont permis d'acquérir une expertise en synthèse conventionnelle et en analyses associées qui ont été essentielles pour aborder les recherches doctorales d'extraction d'éponges et de synthèses biomimétiques. Des connaissances en réactivité chimique, propriétés physico-chimiques des composés organiques sont indispensables pour appréhender avec l'angle le plus approprié les réflexions et expérimentations sur la biogénèse des produits naturels. L'exemple des P2AI et des considérations biosynthétiques associées reflète bien l'intérêt des recherches aux interfaces des disciplines pour réaliser les découvertes les plus riches. En plus de chimistes organiciens, ces recherches font et doivent continuer à associer des experts en écologie marine, radiomarquage, bio-informatique etc pour diversifier les expérimentations destinées à conforter les hypothèses de biosynthèse. De plus, la synthèse biomimétique permet souvent de diminuer le nombre d'étapes des synthèses en forçant le chimiste à élaborer des stratégies de synthèses « non conventionnelles ». En ce sens, la synthèse biomimétique peut être considérée comme une « chimie verte », à l'instar de la biocatalyse. Les recherches que j'effectue désormais en biocatalyse et qui sont présentées dans la partie suivante, ne sont donc que peu éloignées conceptuellement de ces précédents travaux.

1.3. Synthèse biocatalysée de synthons chimiques

Les paragraphes suivants relatent les principales démarches entreprises et résultats obtenus dans ce domaine de la biocatalyse que j'explore depuis 2007 au sein de l'équipe de biocatalyse du Laboratoire de Biocatalyse, Biorémédiation et Métabolisme Synthétique (LB2MS) de l'unité mixte de recherche Génomique Métabolique du Genoscope, CEA.

1.3.1. La biocatalyse et les missions de l'équipe de biocatalyse

Dans un contexte international de transition énergétique, la communauté chimique doit fabriquer les composés cibles de manière économique et économe en énergie. À cet égard, les processus catalytiques ont progressivement remplacé bon nombre des réactions stœchiométriques utilisées ; 80% des produits chimiques sont maintenant produits par catalyse. Au cours des dix dernières années, les progrès scientifiques et technologiques ont rendu une autre chimie catalysée viable, réelle alternative pratique et écologique à ces catalyseurs traditionnels : la biocatalyse. Définie comme la catalyse à base d'enzymes isolées ou d'organismes entiers pour la réalisation de transformations chimiques, la biocatalyse fait partie intégrante des biotechnologies dites blanches qui sont en plein essor au vu du besoin mondial de technologies propres.(Turner et al. 2018) Les enzymes sont de puissants biocatalyseurs et améliorent la vitesse d'un grand nombre de réactions biologiques et chimiques. Issus de ressources renouvelables, biodégradables et non-toxiques, les biocatalyseurs offrent souvent des hautes sélectivités apportées par leur structure intrinsèquement chirale, simplifiant ainsi les traitements et améliorant les rendements. Tous ces avantages, combinés à leur action majoritairement à température ambiante, sous pressions atmosphérique et pH proche de la neutralité, expliquent la réussite de la biocatalyse dans plusieurs procédés, dont certains récompensés par des institutions comme l'Agence de Protection Environnementale Américaine.¹ Les enzymes sont considérées comme des alternatives plus écologiques pour la synthèse chimique de produits, particulièrement ceux de hautes valeurs ajoutées.(Tufvesson et al. 2011, Sheldon et al. 2018) Il existe donc une volonté de poursuivre le développement des biotechnologies blanches pour assurer l'avenir économique durable de la société industrialisée moderne.

De nombreux groupes de recherche exploitent le potentiel de biocatalyseurs déjà décrits dans beaucoup de réactions biocatalytiques, en une étape ou en cascade. Même si ces travaux sont essentiels, de nouveaux scaffolds sont sans aucun doute nécessaires pour élargir la diversité des réactions catalysées et leurs applicabilités. Pour que la biocatalyse devienne une alternative encore plus puissante à la synthèse conventionnelle, les biocatalyseurs doivent satisfaire plusieurs critères, dont : un large spectre de substrats (= promiscuité de substrats), une grande stabilité vis-à-vis des conditions de réactions (température, pH) et stockage, des capacités à convertir d'importantes concentrations de substrats, des stéréosélectivités importantes (facteur d'énantiosélectivité $E > 200$) et si possible complémentaire, des turnover élevés... Même si l'ingénierie des protéines est un outil puissant d'évolution qui permet souvent d'améliorer ces critères, les enzymes modifiées résultantes sont quasi identiques d'un point de vue séquence et n'apportent donc pas de diversité protéique, pourtant essentielle pour accéder à tout le potentiel de la biocatalyse. Un thème de recherche essentiel doit être l'identification de plus d'enzymes natives. La nature pouvant être considérée

¹ "Green Chemistry Challenge Winners." from <https://www.epa.gov/greenchemistry/green-chemistry-challenge-winners>.

comme le vétéran de l'ingénierie des protéines, la recherche dans la biodiversité a tout son sens pour trouver des enzymes aux propriétés variées. Ces enzymes natives sont des nouveaux points de départ pour l'ingénierie protéique qui reste généralement nécessaire pour obtenir des biocatalyseurs industriels.

L'unité de recherche Génomique Métabolique, dans laquelle a été créée l'équipe de chimie et biocatalyse en 2006, avait une activité historique d'analyses génomiques de la biodiversité en bénéficiant de l'expertise en séquençage et analyse de génomes du Genoscope. Après avoir décidé en 2004 d'étendre son expertise à l'étude expérimentale de la fonction des gènes pour compléter la connaissance du métabolisme, l'unité a prolongé ce virage en 2005 vers l'exploration des activités biocatalytiques présentes dans les organismes vivants et leur utilisation à des fins biotechnologiques *via* la création du Laboratoire de clonage et Criblage des Activités de Bioconversions (LCAB) et celle en 2006 du Laboratoire de Chimie Organique et Biocatalyse (LCOB) dont l'activité est étroitement liée à celle du LCAB. J'ai ainsi été en charge avec la responsable du laboratoire, Anne Zaparucha, d'appuyer cette thématique par l'apport d'une expertise en chimie et de biocatalyse dans l'unité. En tirant profit des expertises locales en séquençage, bio-informatique et biologie moléculaire, l'accent des recherches est donc mis sur la recherche de nouveaux biocatalyseurs natifs par approche génomique au sein de la biodiversité procaryote par l'équipe CAB et leur utilisation en synthèse par l'équipe COB.² Ces enzymes doivent permettre d'élargir le catalogue d'enzymes disponibles pour le chimiste et de substituer, soit directement soit après amélioration, une étape de chimie conventionnelle par un étape biocatalytique dans un processus chimique existant ou même de dessiner de nouvelles voies de synthèses pour conduire plus efficacement à des produits d'intérêt dans des conditions compatibles avec la préservation de l'environnement.

La partie recherche de nouvelles enzymes dans la biodiversité (Zaparucha et al. 2018) est effectuée par la plateforme de clonage et de criblage des activités enzymatiques de l'équipe CAB selon une approche génomique décrite en annexe 1.

Après avoir participé à la mise en place du laboratoire, j'ai mené des projets de recherches propres à l'équipe ou collaboratifs, et été en support de différents projets avec des équipes internes ou externes. Parmi tous les projets présentés ici, les travaux expérimentaux de bio-informatique dédiés à la sélection des candidats enzymes au sein de la biodiversité, de biologie moléculaire et de production des enzymes, ont toujours été réalisés par les équipes avec lesquelles nous travaillons en interne quotidiennement (équipe CAB, Laboratoire de Génomique et Biochimie du Métabolisme LGBM, équipe de Métabolisme Synthétique du LB2MS) et non par moi-même.

1.3.2. Présentation succincte des projets supervisés et co-supervisés

Les projets décrits dans cette partie sont ceux pour lesquels j'ai activement dirigé ou co-dirigé avec Anne Zaparucha les recherches.

² Le LCOB fait désormais partie du LB2MS et le LCAB du Laboratoire de Génomique et Biochimie du Métabolisme (LGBM). Ils sont donc nommés respectivement équipe COB et CAB dans la suite du document.

a) Recherche et utilisation de nitrilases

Le premier projet de recherche mené par notre équipe a été la recherche d'enzymes catalysant l'hydrolyse de nitriles en acides carboxyliques, les nitrilases. Cette activité cible avait été préalablement à mon arrivée choisie par Anne Zaparucha. Nous avons ensuite dirigé conjointement ce projet.

La transformation de nitriles en acides carboxyliques est souvent opérée en chimie conventionnelle en conditions dures, acides ou basiques, souvent à températures élevées, générant ainsi des produits secondaires non désirés. Elle est cependant courante dans les stratégies de synthèses de composés d'intérêt car de nombreuses méthodes de synthèses existent pour les substrats nitriles et les produits obtenus sont facilement fonctionnalisables et sont donc des intermédiaires clés dans les schémas de synthèse.

L'hydrolyse enzymatique est une alternative intéressante de par son efficacité et sa stéréospécificité associée aux conditions douces utilisées, tolérant ainsi les groupes fonctionnels fragiles. Les nitrilases sont également des enzymes promiscuitaires, ce qui est un avantage indéniable pour leur utilisation sur des substrats structurellement très divers rencontrés dans les schémas de synthèse. Plusieurs nitrilases étaient répertoriées au début du projet, mais disposer de nouvelles séquences permettait de diversifier le panel de séquences mais aussi de capacités biocatalytiques et fournir ainsi des biocatalyseurs plus adaptés selon les applications synthétiques.

Via la plateforme de clonage et criblage de l'équipe CAB (annexe 1), 290 enzymes candidates ont été sélectionnées par approche génomique parmi environ 700 espèces procaryotes et un métagénome issu du bassin de traitement des eaux d'Evry. Après clonage des gènes correspondants et surexpression des enzymes chez *Escherichia coli* (*E. coli*), 125 enzymes ont montré une activité nitrilase sur au moins un des 25 substrats nitriles testés. Ces derniers avaient été choisis pour représenter au mieux la diversité structurale et la réactivité chimique. Cette large étude a ainsi permis de créer un catalogue d'enzymes pour le chimiste indiquant la(es) bonne(s) enzyme(s) candidate(s) à tester en premier lieu selon la nature du substrat. Comme pour toute autre activité, d'autres enzymes natives peuvent ensuite être recherchées en utilisant la séquence de l'enzyme sélectionnée comme hameçon pour trouver une nouvelle enzyme encore plus active sur le(s) substrat(s) cible(s). La purification d'une partie de ces enzymes a permis de valider les activités sur l'ensemble des 25 substrats et donc de proposer une boîte à outil de 31 enzymes très complémentaires, d'un point de vue diversité séquentielle et spectre de substrats, des nitrilases déjà décrites dans la littérature (Figure 11).

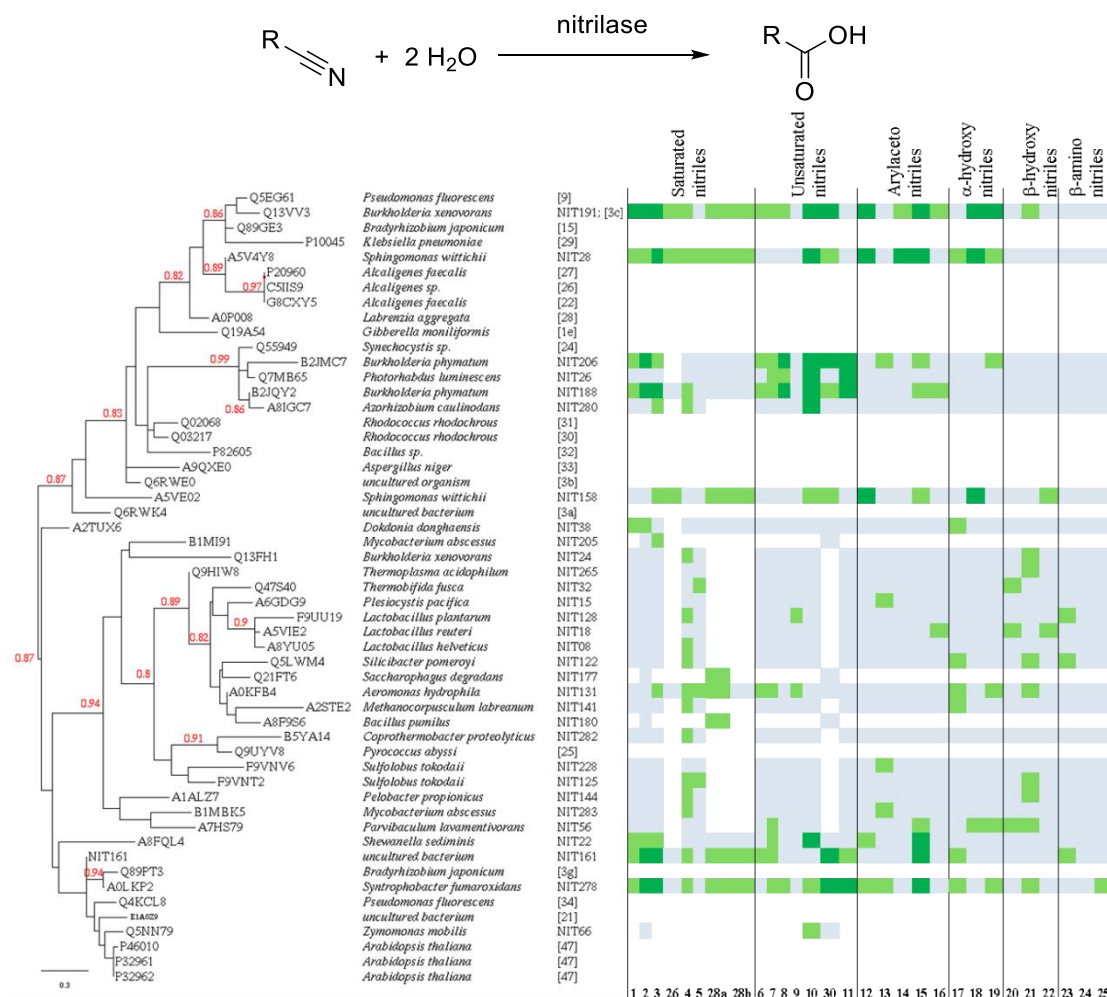


Figure 11. Analyse phylogénétique des 31 nitrilases expérimentalement validées et de 23 nitrilases précédemment décrites dans la littérature.

(les nitrilases identifiées lors de ce travail sont nommées NITX et les références associées aux nitrilases déjà répertoriées sont indiquées entre crochets, les références étant celles de la publication associée (Vergne-Vaxelaire et al. 2013).

- = activité > 1 mM de substrat convertit en 4h
- = activité < 1 mM de substrat convertit en 4h
- = pas d'activité
- = pas testé.

Le potentiel biocatalytique de 3 de ces enzymes a été illustré par la synthèse biocatalysée de 3 acides aux structures variées : l'acide 4-methoxy-4-oxo-3-phenylbutanoïque, l'acide 3-oxocyclopentane carboxylique et l'acide 2-[(cyanométhyl)amino]acétique. L'amélioration des conditions opératoires (tampon, pH, molarité en tampon et substrat, température, cosolvant) a permis de synthétiser ces acides à l'échelle de 30 – 47 mg avec des rendements de 73 - 96 % mais avec des excès énantiomériques modérés (42 – 48 %, respectivement) (Figure 12). L'ensemble de ces résultats a fait l'objet d'une publication dans *Adv. Synth. Catal.* en 2013 (Vergne-Vaxelaire et al. 2013).

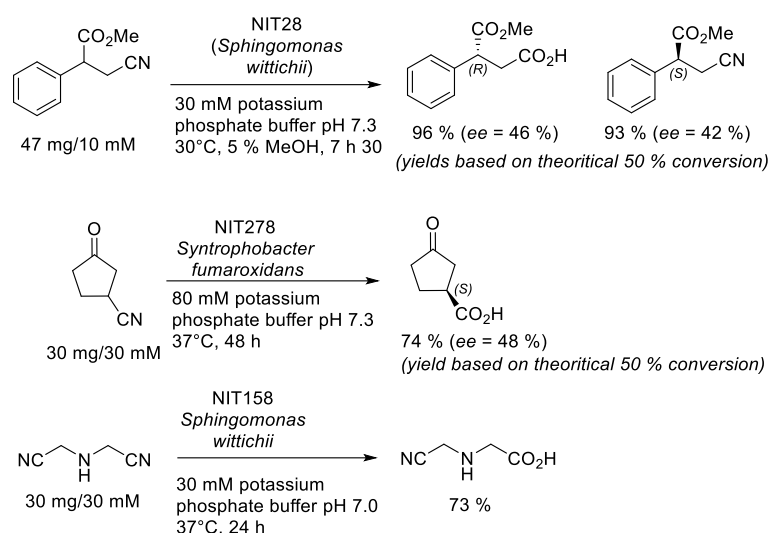


Figure 12. Exemples de synthèses biocatalysées réalisées avec les nitrilases identifiées.

Une seconde collection de nitrilases, focalisée cette fois-ci sur la formation d' α -aminoacides, a ensuite été créée. Ces molécules chirales, naturelles ou non, sont en effet présentes dans de nombreux intermédiaires clés de synthèse et sont donc des cibles pour des voies alternatives biocatalysées. En plus d'une sélection par approche génomique basée sur l'identité de séquence tout comme pour le premier criblage, des candidats ont été sélectionnés selon leur signatures/domaines protéiques en utilisant les bases PFAM (Punta et al. 2012), Gene 3D (Lees et al. 2012) et Prosite (Sigrist et al. 2013). Après production des enzymes correspondantes, un criblage des 425 nitrilases obtenues combinées aux 163 nitrilases du criblage précédent, a été réalisé par LC-MS sur 6 α -aminonitriles après dérivatisation des milieux par un agent chiral. Le développement de cette méthode de criblage sur lysat cellulaire a été mené avec l'aide de Ekaterina Darii et Alain Perret en charge du système LC couplé au spectromètre de masse QTRAP (Hybrid triple quadrupole-linear ion trap mass spectrometer). Grâce à l'optimisation des paramètres de masse, plusieurs α -aminoacides cibles ont pu être suivis simultanément par méthode MRM (Multiple Reaction Monitoring), réduisant ainsi considérablement le temps d'analyse global. La méthode mise en place a ainsi permis de cribler une grande collection en obtenant des informations à la fois de conversions et d'énantiospécificités sur les enzymes testées. Cinq enzymes dont NIT191 et NIT28 ont montré une activité d'hydrolyse, les substrats actifs testés étant uniquement le 2-aminovaléronitrile et le 2-amino-2-phenylacetonitrile (Figure 13). L'ensemble de ces données ont été publiées en 2014 dans *J. Mol. Catal. B: Enz.* (Bordier et al. 2014).

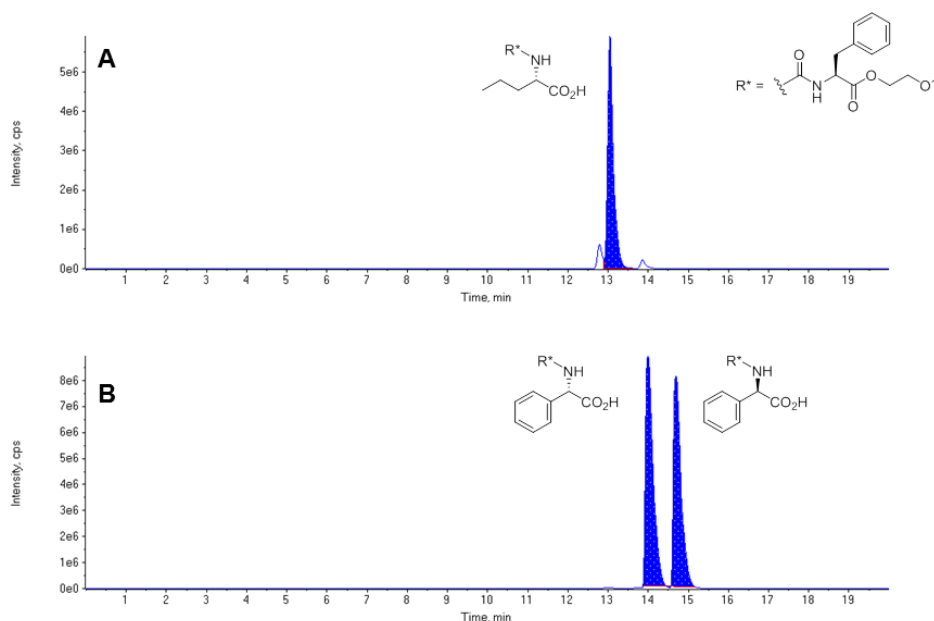


Figure 13. Projet nitrilase : analyse MRM des milieux réactionnels dérivatisés contenant A) NIT191: détection de la (S)-norvaline (*ee* = 94 %) et B) NIT28 : détection de la (R)-phenylglycine (*ee* = 9%)

Ce premier projet mené dans l'équipe a duré de 2009 à 2013 avec la supervision d'un étudiant en thèse et de deux étudiants d'école d'ingénieur (cf III.1). La collection est toujours en cours d'étude dans l'équipe CAB, en particulier pour la recherche de nitrilases thermoactives et thermostables. Je reste donc en support pour la continuité de ce projet. La collection générée est encore à exploiter, des projets futurs de recherche l'utilisant sont à mettre en place. Des projets collaboratifs avec d'autres équipes du CEA sont aussi menés pour utiliser certaines des enzymes découvertes en combinaison avec d'autres enzymes (systèmes en « cascade ») dans des systèmes microfluidiques ; cependant je ne prends pas part à ces recherches.

En plus des résultats scientifiques obtenus, ce premier projet publié de biocatalyse nous a permis de nous approprier les techniques expérimentales propres à la biocatalyse mais aussi d'avoir une meilleure vision de la façon dont ce type de projet doit être mené. Suite aux discussions que nous avons pu avoir lors de congrès scientifiques de biocatalyse où nous avons eu l'opportunité de présenter ces résultats, nous avons pu partager des retours d'expériences qui ont été bénéfiques pour l'ensemble des projets menés par la suite dans l'équipe et l'unité.

b) Amines déshydrogénases pour la synthèse biocatalysée d'amines chirales : recherche, application et évolution

Forte des résultats obtenus sur le projet nitrilase et sur les autres projets pour lesquels j'ai été en support (cf I.3.3), j'ai ensuite décidé de mettre à profit les expertises présentes en interne pour rechercher dans la biodiversité une autre activité enzymatique, les amines déshydrogénases. Ces enzymes catalysent la réaction d'amination réductrice de cétones prochirales en amines chirales en présence d'ammoniaque (Figure 14). Contrairement aux nitrilases, ces enzymes étaient encore orphelines de gènes avant le début de ce projet, la famille restait donc entièrement à découvrir. Les compétences en analyse de génomes du Genoscope, couplées à la plateforme de clonage et criblage

(annexe 1), devaient permettre de combler ce manque de données intéressant un grand nombre de chimistes. La synthèse asymétrique en une étape d'amines chirales à partir de cétones prochirales et d'ammoniaque comme source d'amine avait en effet été listée comme l'une des réactions les plus prometteuses pour l'industrie pharmaceutique (Constable et al. 2007). Les amines chirales sont présentes dans plus de 40% des principes actifs et dans de nombreux composés utilisés en agroalimentaire (Schritt Wieser et al. 2015, Afanasyev et al. 2019). Or la synthèse des amines chirales par voie conventionnelle n'est pas réalisable en une étape, une étape intermédiaire de formation d'imine secondaire ou d'amine substituée étant nécessaire pour apporter la chiralité. La biocatalyse est donc dans ce cas une solution alternative remarquable, l'environnement chiral fourni par l'enzyme donnant accès en une étape à d'excellents excès énantiomériques, dans des conditions douces et sans utilisation de métaux polluants (Höhne et al. 2009, Grogan 2017, Cosgrove et al. 2018).

Au départ du projet, les seuls exemples répertoriés dans la littérature étaient des activités sans gène associé (Itoh et al. 2000, Wang et al. 2013, Wang et al. 2016) ou des enzymes modifiées par ingénierie protéique. Vu l'intérêt porté à cette alternative biocatalytique, de lourds travaux de mutagenèse avaient en effet été entrepris tout d'abord par l'équipe de Andreas Bommarius (Georgia Tech Institute, Atlanta) (Abrahamson et al. 2012, Abrahamson et al. 2013) puis par d'autres équipes lors du projet (Wetzel et al. 2016, Lowe et al. 2018, Pushpanath et al. 2017, Chen 2018) pour modifier le spectre de substrats des acides aminés déshydrogénases (AADHs) qui catalysent la même réaction chimique mais uniquement sur des α/β céto acides. Très récemment, c'est une ϵ -deaminating L-lysine déshydrogénase qui a été modifiée à ces fins par un remarquable travail de l'équipe de F. Mutti (Tseliou et al. 2019). Une autre classe d'oxydoréductases permet l'accès à des amines, les imines réductases (IRED) mais ces dernières catalysent uniquement la réduction de l'imine intermédiaire et non sa formation, même si des homologues nommés « reductive aminases » (RedAms) ont montré la double activité (Aleku et al. 2017, Sharma et al. 2018) comme pour les AmDHs, avec cependant une faible activité sur l'ammoniaque par rapport aux amines substituées (Mangas-Sanchez et al. 2017). Au contraire des transaminases largement décrites pour la synthèse d'amines chirales, les AmDHs ont plusieurs avantages dont uniquement le besoin d'ammoniaque, source d'amine très facilement accessible, et d'un cofacteur nicotinamide pour lequel de nombreux systèmes de régénérations efficaces existent. La synthèse asymétrique d'amines par les AmDHs est donc l'une des voies idéales pour produire les amines chirales (Patil et al. 2018). L'objectif est donc de trouver des enzymes transformant des cétones sans prérequis structural afin de couvrir au mieux les demandes en synthèse.

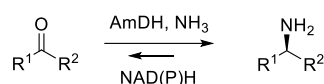


Figure 14. Réaction catalysée par les amines déshydrogénases (AmDHs)

En 2014, j'ai donc mis en place et supervisé ce projet. Cette activité étant orpheline de gène, la stratégie basée sur la définition d'un set de référence (= ensemble des gènes naturels identifiés codant pour les protéines catalysant cette réaction) comme point de départ pour l'approche génomique utilisée usuellement par la plateforme CAB (annexe 1), n'était pas applicable en tant que tel. Or cette étape est déterminante pour la réussite des projets. Un set de référence a donc été constitué selon une vision chimique et non biologique, en utilisant les modes de recherche par sous-structure des bases de données métaboliques. Les enzymes à cofacteur nicotinamide répertoriées comme catalysant une réaction de transformation de cétone en amine non terminale et sans fonctionnalisation sur le

carbone α , afin d'éviter de rapatrier des AADHs, ont été recensées à l'aide des bases KEGG et BRENDA. Les deux enzymes issues de cette recherche, L-erythro-diaminohexanoate déshydrogénase « kdd » (Kreimeyer et al. 2007) et 2,4-diaminopentanoate « 2,4-DAPDH » de *Clostridium sticklandii* (identifiant UniprotKB : E3PY99 (Fonknechten et al. 2009)) n'avaient pas encore été étudiées d'un point de vue potentiel biocatalytique (Figure 15A). Selon la démarche de la plateforme du CAB (annexe 1), une sélection d'enzymes de la biodiversité rapatriées par ces enzymes références a été réalisée et testée sur différents substrats cétones proches structurellement des substrats natifs de kdd et 2,4-DAPDH. Parmi les 9 enzymes identifiées comme actives sur au moins l'un d'entre eux (AmDH1-9), AmDH4 issue de l'organisme thermophile *Petrotoga mobilis* a été plus amplement étudiée. L'activité native de cette enzyme reste une 2,4-DAPDH (l'analyse du contexte génomique et les mesures d'activités vont dans ce sens), mais les cétones acide 4-oxopentanoïque et 5-oxohexanoïque sont aussi substrats, ce qui en fait des AmDHs d'un point de vue biocatalyse. Cette activité a permis d'assigner pour la première fois un gène à cette activité AmDH. Après optimisation des conditions réactionnelles (température, système de recyclage du cofacteur nicotinamide, concentration et nature du tampon, concentration en substrat et enzyme), une synthèse de 465 mg de (4S)-4-aminopentanoate (*ee* > 99.5 %) a permis de montrer l'efficacité de AmDH4 pour une synthèse à échelle semi-préparative (Figure 15B) avec un « Space Time Yield » de 53 g L⁻¹ d⁻¹. Plus de détails concernant ces résultats sont disponibles dans la publication associée (Mayol et al. 2016).

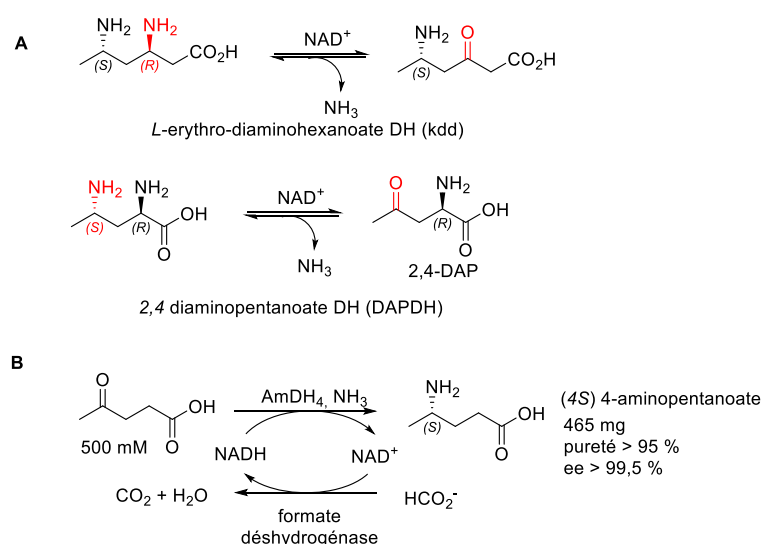


Figure 15 A : enzymes natives choisies comme set de référence ; **B** : Synthèse biocatalysée du (4S)-4-aminopentanoate par l'enzyme AmDH4.

En plus de l'aspect synthèse, une étude biochimique et structurale de cette enzyme a été menée. L'absence de similarité de séquence avec les enzymes répertoriées dans la Protein Data Bank (PDB), pour pouvoir effectuer une modélisation de sa structure par homologie, et son activité prouvée AmDH, ont motivé mon choix d'entreprendre une collaboration avec une équipe de cristallographie intéressée aussi par l'activité AmDH afin de réellement collaborer et non sous-traiter la partie cristallographie. Dès fin 2016, j'ai donc établi des échanges fructueux avec Gideon Grogan en charge du Laboratoire de « Structural and Applied Enzymology » à l'Université de York, UK. La structure de AmDH4 a pu être déterminée ce qui a ouvert le champ d'exploration structurale à ce projet, et une collaboration également en interne pour la modélisation d'enzymes avec Karine Bastard du Laboratoire d'Analyses Bioinformatiques pour la Genomique et le Metabolisme (LABGEM) de l'unité. Le monomère de cette

structure dimérique est composée d'un domaine N-terminal avec un repliement de type Rossman et un domaine C-terminal à feuillet β 6-brins, avec une grande ressemblance avec les dihydropicolinate reductase et les meso-diaminopimelate déshydrogénase malgré leur très faible similarité en séquence. Des différences significatives sont cependant observées, notamment au niveau de la structure des boucles. Un des monomères au sein des deux dimères obtenus montre une configuration « fermée » au-dessus du cofacteur qui semble modifier la topologie du site actif (Figure 16A).

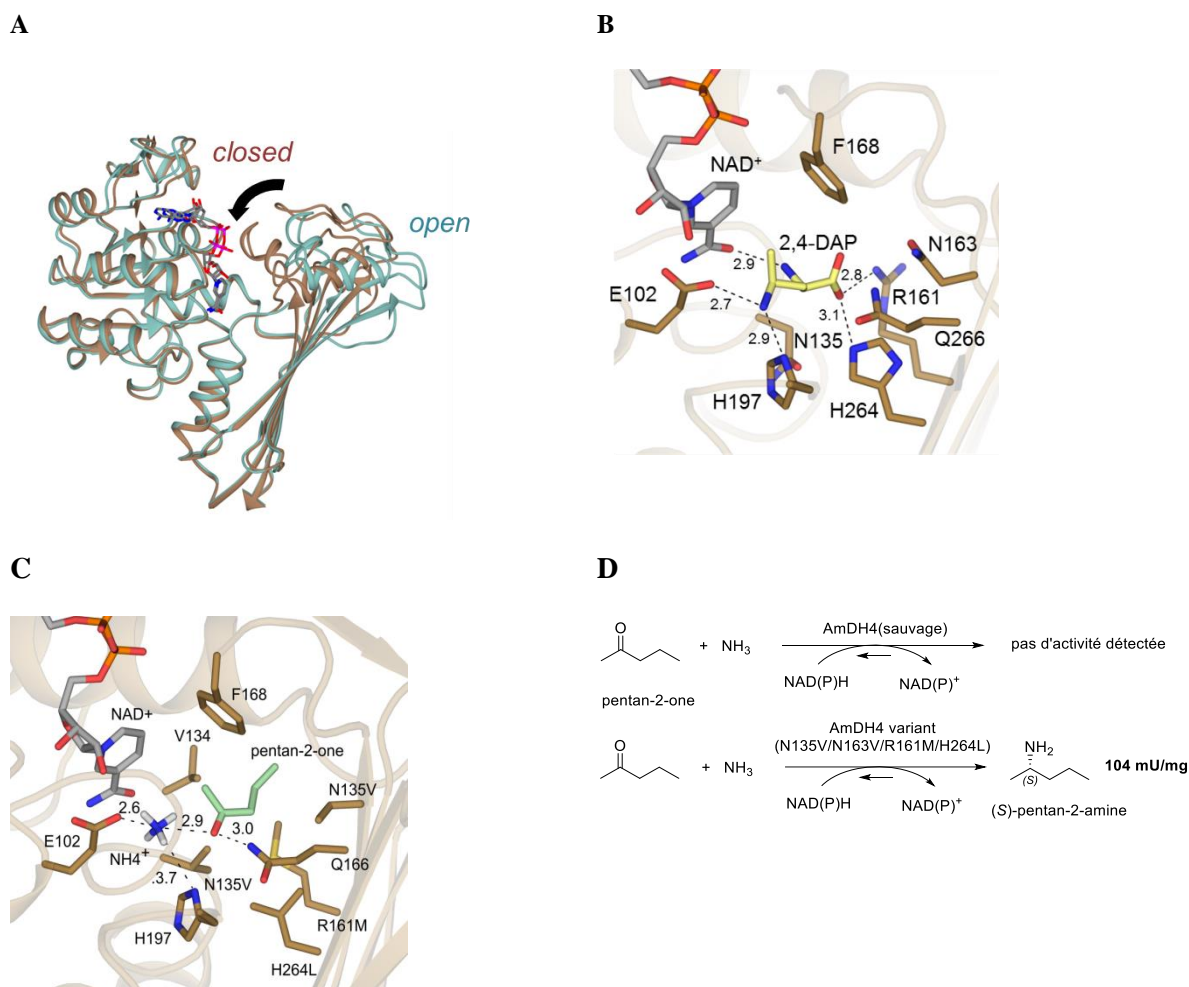


Figure 16. **A** : Structure de AmDH4 en complexe avec NAD^+ : superposition des monomères ouvert (bleu) et fermé (marron) (source : Gideon Grogan); **B** : Site actif de AmDH4 incluant la modélisation avec son substrat 2,4-DAP (source : Karine Bastard); **C** : Modèle du variant AmDH4 (N135V/N163V/R161M/H264L) en complexe avec NAD^+ , NH_4^+ et la pentan-2-one ; **D** : activité du variant AmDH4 sur la pentan-2-one.

L'analyse de la structure du site actif de AmDH4 (Figure 16B) nous a permis d'émettre des hypothèses sur le mécanisme de catalyse et donc les résidus essentiels (discuté plus loin). Ceci nous a ainsi conduit à substituer les résidus polaires Asn135, Arg161, Asn163 et His264, postulés pour interagir avec le groupement carboxylate du substrat natif 2,4-DAP, par des équivalents apolaires afin de modifier la spécificité de cette enzyme pour des substrats sans ce prérequis fonctionnel. La production de 14 mutants (simples ou multiples) en collaboration avec l'équipe CAB a donné lieu à un mutant (N135V/N163V/R161M/H264L) montrant une activité spécifique de 104 mU mg^{-1} avec la pentan-2-one et une absence d'activité sur le substrat natif 2,4-DAP (Figure 16C et D). Cette activité restait faible, en

particulier vis-à-vis de l'activité de l'enzyme native sur son substrat natif (28 U mg^{-1}), mais elle était prometteuse pour le projet.

Les résultats encourageants obtenus lors de cette première partie de projet nous ont ensuite incités à poursuivre l'exploration de la biodiversité pour rechercher des AmDHs avec le minimum de restriction structurale sur le substrat. Les 9 enzymes AmDH1-9 précédemment découvertes n'étant actives que sur des substrats fonctionnalisés par un acide carboxylique en position γ . Une même approche génomique, en utilisant cette fois-ci en plus les séquences protéiques des AmDHs découvertes comme séquences de référence, nous a permis de trouver les AmDHs nommées *Msme*AmDH et *Cfus*AmDH, puis 4 de leurs homologues nommés *Micro*AmDH, *Apau*AmDH, *Mvac*AmDH, et *Myco*AmDH. Ces 6 AmDHs (nat-AmDHs) sont actives sur une grande variété de cétones et aldéhydes aliphatiques, et non actives sur les substrats type 2,4-DAP ou acide 4-oxopentanoïque des enzymes des sets de référence successifs utilisés (Figure 17).

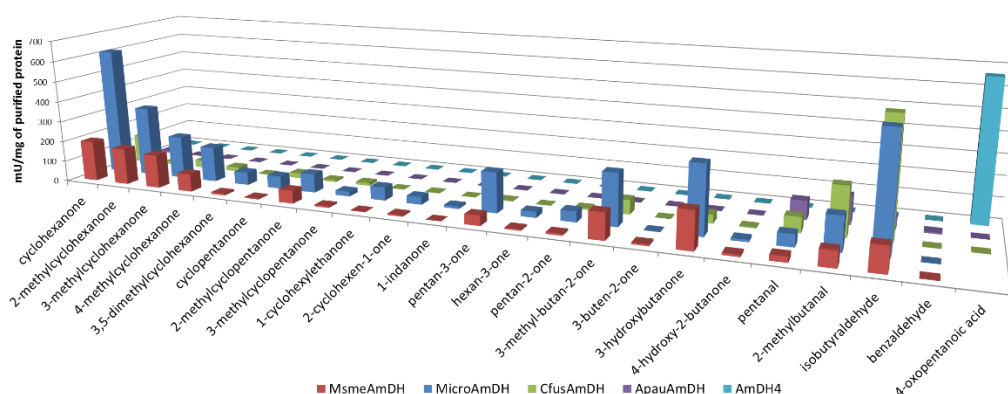
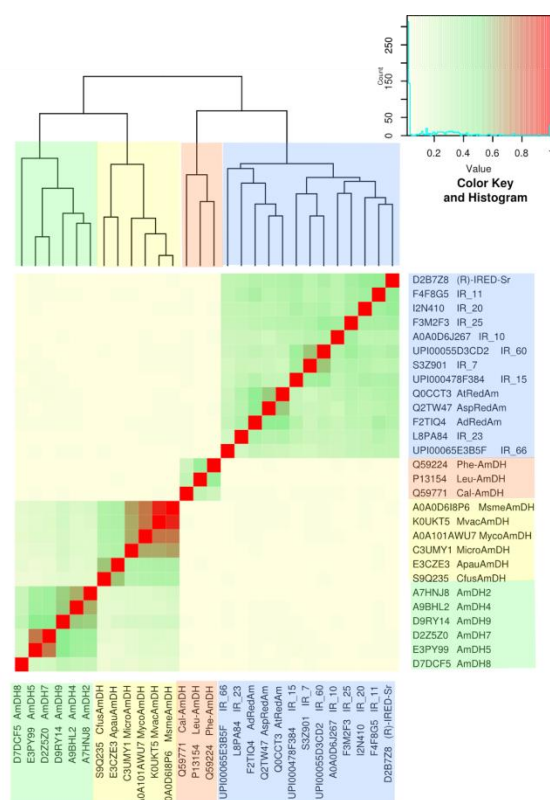


Figure 17. Spectre de substrats de 5 des nat-AmDHs découvertes. Activité spécifique obtenue par test spectrophotométrique dans les conditions suivantes : 2 M $\text{NH}_4\text{HCO}_2/\text{NH}_4\text{OH}$ pH 9.5, NADH ou NADPH (0.2 mM), cétone ou aldéhyde (10 mM), enzyme purifiée (0.01 – 0.5 mg/ml), 20 °C.

Cette démarche itérative de criblage de la biodiversité utilisée pour la première fois au laboratoire a donc montré sa puissance pour accéder à des enzymes catalysant la même réaction d'un point de vue chimique, mais sur des substrats très différents structurellement. A partir de la 2,4-DAPDH comme séquence de référence, les AmDH1-9 identifiées possèdent des similarités de séquences de 30,5 à 75,5%, qui à leur tour ont permis l'identification de *Msme*AmDH homologue à 28% et 25% de AmDH4 et de l'enzyme de référence initiale E3PY99 (=AmDH5), respectivement. Ces nouvelles protéines constituent une intéressante ressource pour les recherches futures car elles n'ont pas de similarité de séquences avec les autres enzymes catalysant les aminations réductrices comme les AADHS mutées en AmDHs et les IREDs et RedAms, découvertes pendant ce projet (Figure 18).

A



B

Uniprot entry	Name	Organism
A0A0D6I8P6	Msme AmDH	<i>Mycobacterium smegmatis</i>
K0UKT5	Mvac AmDH	<i>Mycobacterium vaccae</i>
A0A101AWU7	Myco AmDH	<i>Mycobacterium</i> sp.
C3UMY1	Micro AmDH	<i>Microbacterium</i> sp. MA1
E3CZE3	Apau AmDH	<i>Aminomonas paucivorans</i>
S9Q235	Cfus AmDH	<i>Cystobacter fuscus</i>
A9BHL2	AmDH4	<i>Petratoga mobilis</i>
E3PY99	AmDH5	<i>Clostridium sticklandii</i>
Q2TW47	Asp RedAm	<i>Aspergillus oryzae</i>
Q59224	PheDH	<i>Bacillus badius</i>
P13154	LeuDH	<i>Geobacillus stearothermophilus</i>
D2B7Z8	(R)-IRED-Sr	<i>Streptosporangium roseum</i>
A0A0D6I267	IR_10	<i>Mycobacterium smegmatis</i>
P04036	DapB	<i>Escherichia coli</i>

Figure 18. Similarité de séquences des nat-AmDHs, AmDH1-9 et d'IRED/RedAms décrites comme ayant une activité AmDH. **A** : Heatmap et clustering hiérarchique de ces enzymes. La « heatmap » représente la matrice de similarité de séquence deux à deux. Les valeurs d'identité de séquence, variant de la valeur 0 (aucune identité, en jaune pâle) à la valeur 1 (séquence identique, en rouge) ; **B** : matrice d'identité de séquences de ces enzymes.

Les structures de *MsmeAmDH* et *CfusAmDH* ont été résolues toujours en collaboration avec l'équipe de Gideon Grogan. Leurs structures dimériques sont très similaires à celle de AmDH4, montrant également une forme monomère « fermée » en présence de substrat/produit lors de la cristallisation, ce qui a permis d'appuyer l'hypothèse formulée suite à l'obtention de la structure de AmDH4 (Figure 19A). Cette conformation semble cruciale pour l'effectivité de la catalyse au sein de cette famille, même si des expériences de mutagenèse dirigée sont nécessaires pour affiner cette conclusion. L'analyse des sites actifs, appuyée de quelques expériences de mutagenèse dirigée et de détermination de paramètres cinétiques, a permis de postuler un mécanisme réactionnel pour ces enzymes, impliquant entre autres le résidu catalytique clé glutamate pour l'activation de l'ammoniaque, et des résidus de stabilisation de la cétone (Tyr168 pour *CfusAmDH*, Trp164 pour *MsmeAmDH*, Gln266 pour AmDH4) (Figure 19B).

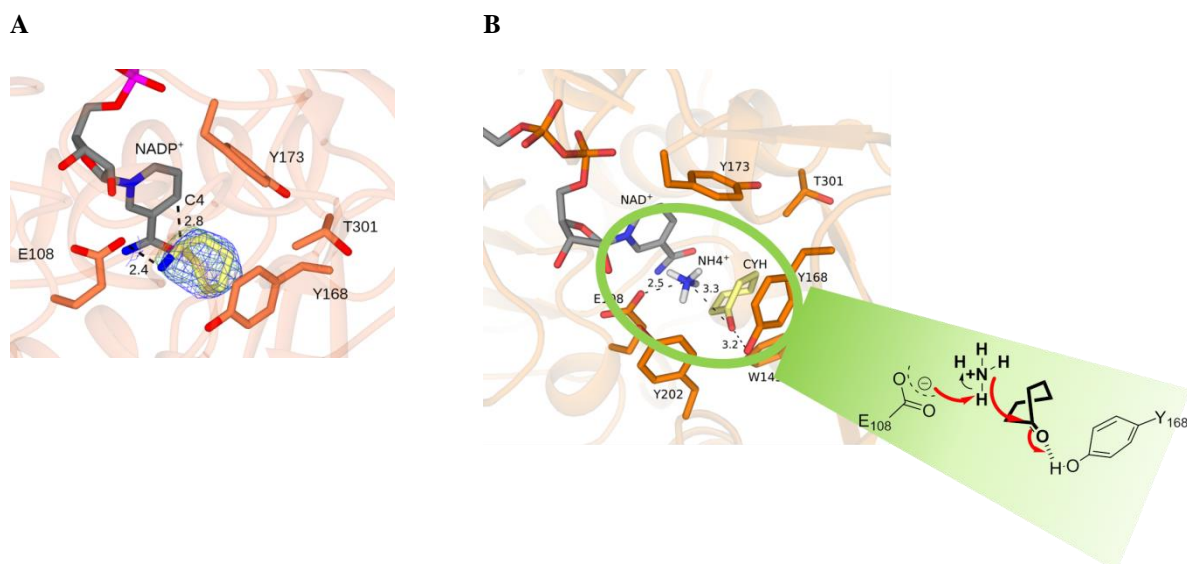


Figure 19. A: Structure du site actif de *CfusAmDH* complexée avec NADP⁺ et la cyclohexylamine; **B:** modèle d'association de NADP⁺, NH₄⁺ et cyclohexanone dans la structure cristallisée *CfusAmDH*.

Tous les membres de la famille AmDH découverte ont été classés selon leur site actif selon la méthode Active Site Modelling and Clustering (ASMC) mise au point par l'unité (de Melo-Minardi et al. 2010, Bastard et al. 2014) ouvrant un vaste champ d'exploration pour la communauté scientifique intéressée par cette activité. L'approche génomique, le criblage des enzymes candidates, l'étude biochimique des enzymes actives, leur utilisation en synthèse biocatalysée illustrée par la synthèse de 3 amines, ainsi que leur résolution structurale sont décrits dans la publication de 2019 dans *Nature Catalysis* (Mayol et al. 2019). L'activité de ces enzymes fait aussi l'objet d'un brevet (Vergne-Vaxelaire et al. 2018).

Les résultats obtenus sur le projet AmDH sont le fruit d'un important travail initié expérimentalement par Sylvain David (stagiaire M2) et remarquablement poursuivi par Ombeline Mayol (étudiante en thèse, 2015-2018) (cf III.1).

c) Travaux actuels et perspectives à court terme sur le projet AmDH :

Ce projet a donc donné lieu à de nombreux résultats qui continuent d'ouvrir des perspectives d'exploration de cette famille encore non caractérisée au début du projet. J'ai entre autre entrepris plusieurs collaborations externes et internes pour étendre les opportunités de découvertes autour de cette activité dans les mois et années à venir.

➤ Evolution dirigée *in vivo* des enzymes

En collaboration avec l'équipe de métabolisme synthétique du LB2MS, des essais d'évolution dirigée *in vivo* des AmDHs sont menés dans le but d'atteindre des meilleures activités catalytiques par une approche alternative à l'évolution dirigée *in vitro* et/ou la recherche de nouvelles enzymes dans la biodiversité. L'unité dispose en effet d'automates de culture continue microbienne nommés GM3 (Marliere et al. 2011). Ces automates permettent l'évolution dirigée à long terme de populations de cellules soumises à des pressions de sélection par application de différents régimes de culture. L'objectif est donc d'utiliser cette technologie pour obtenir des populations fixées de cellules

contenant les mutations adaptatives donnant le phénotype attendu, c'est-à-dire dans notre cas une meilleure activité catalytique envers les substrats ciblés. Le principe du GM3 est fourni en annexe 2. Le crible de sélection choisi pour cette évolution est basé sur la libération d'ammoniaque lors de la réaction réverse d'amination réductrice, à savoir la déamination oxydative. L'azote étant nécessaire à la croissance cellulaire d'*E. coli*, les populations ne fournissant pas assez d'azote *via* la réaction AmDH reverse seront contre-sélectionnées au profit de celles ayant fixées les mutations le permettant. L'objectif est d'obtenir des mutations au niveau des gènes codant pour les AmDHs, mais des mutations engendrant des changements d'expression au niveau des gènes sont aussi à attendre. Bien évidemment, cette approche est basée sur l'hypothèse que l'activité réversible améliorée dans le sens oxydatif, sera également améliorée dans le sens réductif. Une souche modifiée d'*E. coli* a été créée, par l'équipe de métabolisme synthétique, afin d'éviter la sélection de souches possédant des mutations au niveau d'autres gènes codant pour l'expression d'enzymes susceptibles elles aussi de former une source azotée à partir des substrats envisagés. Ainsi, trois gènes ont été délétés sur la souche sauvage : *puuE* et *gapT* codant pour des transaminases et *dadA* codant pour une diaminoacide oxydase. Le gène codant pour l'AmDH ciblée est ensuite introduit dans cette souche.

Les premiers essais menés sur les enzymes *MicroAmDH*/*CfusAmDH*/*MsmeAmDH*/*LAmDH* se sont révélés infructueux, la croissance cellulaire étant initialement trop faible voire inexistante avec les constructions plasmidiques testées et les substrats cyclohexylamine, pentanamine et 4-méthylpentan-2-amine. Afin de tenter cette évolution *in vivo* sur des bases plus solides, une évolution des DAPDHs est désormais planifiée. Les activités natives de déamination sur leur substrats (2*R*,4*S*)-DAP ou (4*S*)-AP sont plus importantes que celles mesurées avec les autres AmDHs et ces substrats plus hydrophiles sont certainement mieux transportés à l'intérieur des cellules. La démarche envisagée repose donc sur l'évolution de l'activité de la DAPDH AmDH5 de *Clostridium sticklandii* (et non de AmDH4 du fait de son activité maximale à plus haute température) sur son substrat natif (2*R*,4*S*)-DAP vers son substrat moins affiné 4AP puis vers des substrats type isobutyrylamine/cyclohexylamine/pentan-2(ou3)-amine etc... Le (2*R*,4*S*)-DAP étant difficilement accessible, l'ornithine aminomutase catalysant sa formation à partir de la D-ornithine dans la voie de fermentation de cette dernière, sera également introduite dans la souche afin de le fournir *in situ* (Figure 20).

Soumis à un régime de sélection de type medium swap (cf annexe 2), les souches contenant les gènes ornithine aminomutase (OA) et AmDH5 seront évoluées jusqu'à obtenir des populations fixées pouvant croître avec le substrat D-ornithine comme seule source d'azote. Cette évolution devrait être assez rapide, le substrat généré *in situ* (2*R*, 4*S*)-DAP étant le substrat natif de AmDH5. L'évolution suivante plus ambitieuse devrait quant à elle permettre d'obtenir des souches capables de croître uniquement avec le (4*S*)-AP comme source d'azote, et donc disposant potentiellement de gènes évolués AmDH5 plus affins pour ce substrat. Même si cette amine n'est pas forcément une amine cible d'un point de vue biocatalyse, les résultats attendus pour cette évolution devraient être très riches à analyser notamment *via* une rationalisation structurale. La structure cristallographique de son homologue AmDH4 étant résolue, l'analyse *in silico* des mutations intragènes attendues sera possible et à relier avec des résultats expérimentaux.

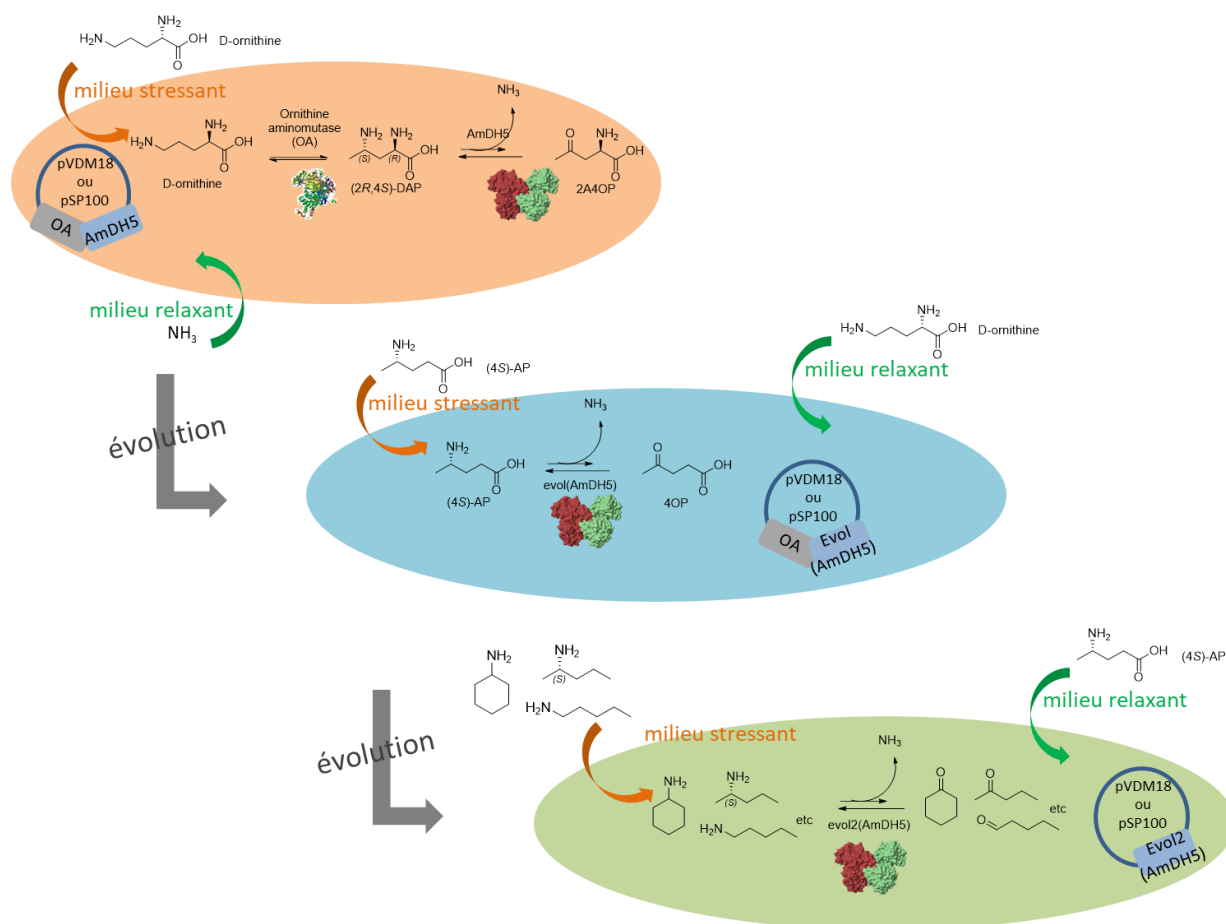


Figure 20. Approche envisagée pour l'évolution dirigée *in vivo* des AmDHs

Enfin, l'évolution suivante a pour objectif de sélectionner des gènes AmDH5 mutés dont les enzymes correspondantes posséderaient des activités proches des nat-AmDHs, c'est-à-dire sur des substrats sans contraintes fonctionnelles type acide carboxylique. Dans ce cas, la contrainte d'entrée des substrats dans la cellule sera à considérer, mais on saura que les souches travaillées sont déjà bien aptes à générer l'azote nécessaire par l'expression de gènes AmDH modifiés (Figure 20). Il est à noter que cette évolution est également envisagée avec une enzyme AmDH nouvellement identifiée (*TtherAmDH*, 1 des 18 issues de l'exploration préliminaire de métagénomes, cf paragraphe « Exploration préliminaire de métagénomes »), car l'enzyme correspondante possède déjà une activité mixte sur le (2R, 4S)-DAP, le (4S)-AP et la cyclohexylamine/pentanamine. Ces travaux débutent avec Laurine Ducrot (doctorante depuis octobre 2019 qui a obtenu un financement ministériel pour la période 2019-2022, cf III.1) après les essais préliminaires infructueux menés par l'équipe de métabolisme synthétique avec Ombeline Mayol (sans intégration de l'OA) et Adam Caparco (évolution directement sur nat-AmDH et LAmDH).

➤ **Identification de gènes codant pour des AmDHs dans des souches marines**

Des approches de génomique comparative ont été menées pour identifier les gènes responsables de cette activité au sein de deux souches marines détectées actives et brevetées par l'équipe de Shizhen Wang (Université de Xiamen, Chine). Contactée après lecture de ses brevets (Wang et al. 2013, Wang

2016) j'ai été à l'initiative d'un projet qui lui a permis l'obtention d'une bourse Jean d'Alembert de l'université Paris Saclay en 2017 et donc de cette collaboration (séjour de 5 mois au sein de notre unité). L'approche utilisée, qui a impliquée plusieurs équipes du Genoscope pour la partie séquençage ainsi que David Vallenet du LABGEM pour l'identification des gènes candidats au sein des génomes « positifs AmDHs », est décrite en Figure 21. Les gènes n'ont pas pu pour l'instant être trouvés mais les séquences de deux génomes utilisés dans ce projet seront soumises à publication dans *Microbiology Resource Announcements*.

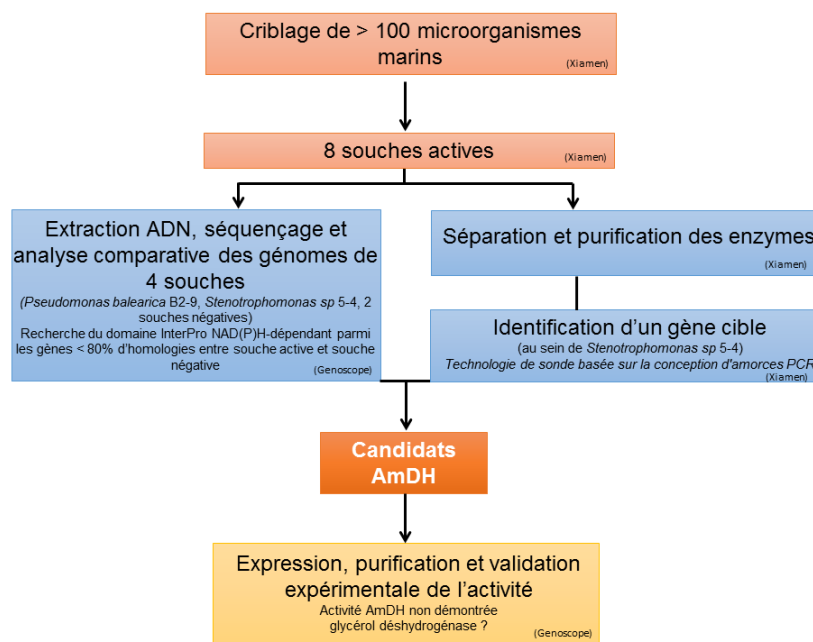


Figure 21. Protocole de recherches d'enzymes AmDHs au sein de souches marines (collaboration avec Shizhen Wang, Université de Xiamen, Chine)

➤ Exploration préliminaire de métagénomes

Des recherches d'autres AmDHs au sein de trois ressources métagénomiques,³ générées en partie au Genoscope, ont pu être effectuées par Adam Caparco, un étudiant en thèse de Andreas Bommarius et de Julie Champion (Georgia Tech Institute, USA) accueilli au sein de notre unité pendant 4 mois grâce à l'obtention d'une bourse Chateaubriand. Cette collaboration a été initiée suite à des discussions lors de congrès internationaux. L'obtention des séquences nat-AmDHs a en effet ouvert la voie à une première exploration de ces données métagénomiques encore très peu exploitées à des fins biocatalytiques. Les ressources explorées de manière très préliminaire sont celles issues des projets Tara Ocean (Alberti et al. 2017) et MetaHit (European Metagenomics of the Human Intestinal Tract)⁴ explorant respectivement le réservoir de gènes des communautés planctoniques des océans et du microbiome intestinal. Les bases de données associées sont Ocean Microbial reference Gene Catalog (OM-RGC,(Sunagawa et al. 2015)) Marine Atlas of Tara Oceans Unigenes (MATOUv1 (Carradec et al. 2018)) et Integrated Gene Catalog (IGC (Li et al. 2014)). Ce sous-projet a été possible grâce à une

³ Une brève description de la métagénomique est développée en début de partie projet

⁴ <http://www.metahit.eu/>

collaboration interne avec le Laboratoire d'Analyses des Génomes Eucaryotes (LAGE) du Genoscope. L'objectif scientifique de ce travail est de montrer la présence de gènes codant pour des AmDHs au sein de ces communautés tout en ayant un premier aperçu de la diversité AmDH (génomique, structurale, activités) accessible *via* ces ressources. Les résultats obtenus sur des protéines obtenues à partir de gènes synthétiques sélectionnés dans ces métagénomes sont encourageants. En effet, 18 AmDHs ont pu être prouvées expérimentalement avec des variabilités au niveau site actif, spectre de substrats et stabilité. Les enzymes testées issues des métagénomes marins sont pour la plupart très éloignées en séquence (< 35 %) des nat-AmDHs précédemment identifiées, ce qui permet d'élargir la gamme en séquence des enzymes caractérisées de cette superfamille. Ces résultats font l'objet d'une publication en collaboration soumise récemment dans *Green Chemistry* (annexe 3). Ce travail préliminaire est à la base d'un des projets de recherche présenté dans la suite de ce mémoire (cf II.1).

➤ **Exploration structurale de la famille**

Par ailleurs, les données structurales obtenues ouvrent la voie à plusieurs champs d'études de cette famille qui vont pouvoir être travaillés par la doctorante Laurine. Des enzymes au spectre de substrats ou stéréosélectivités différentes seront recherchées par analyse *in silico* des membres de la classification ASMC et des nouvelles AmDHs identifiées suite au criblage des (méta)génomes mentionnés ci-dessus, couplée à des considérations d'ingénierie protéique dirigée. Les candidats sélectionnés au cours de ces analyses *in silico* seront testés *in vitro* après expression des gènes correspondants chez *E. coli* en essayant de rationaliser l'ensemble des résultats pour une meilleure compréhension de la catalyse de ces enzymes. La connaissance des sites actifs de AmDH4 NADH-dépendante et de *Cfus/Msme*AmDH NADPH-dépendantes permet aussi d'envisager des études de changement de cofacteur par ingénierie protéique. L'analyse des sites actifs des AmDHs expérimentalement prouvées autorise également des recherches 3D d'homologues dans la biodiversité : cet aspect est développé en partie projet II.1. L'ensemble du volet structural du projet AmDH sera entrepris en lien étroit avec Gideon Grogan et une de ses étudiantes en thèse, Megan Bennett, financée également pour 2019-2022, pour la partie élucidation structurale des nouvelles AmDHs validées expérimentalement et discussions sur les considérations mécanistiques. Gwenaëlle André Leroux (INRAE, Unité MaIAGE) sera également investie dans ce projet, notamment en support pour les aspects modélisations/docking.

➤ **Applications des AmDHs en synthèse**

Enfin, le volet synthétique est développé avec Caroline Paul (Université de Delft, Pays-Bas) sur un sous-projet de cascade enzymatique impliquant les AmDHs découvertes, des Ene-réductases et les mimiques de cofacteurs nicotinamides (Paul et al. 2016). Après des études préliminaires réalisées par un étudiant en M2 (Rudy Bayakimissa), la mise en place de cette cascade est effectuée fin 2019-début 2020 par un stagiaire de Caroline Paul et en interne au Genoscope. L'intérêt principal de cette cascade réside dans le contrôle de la stéréochimie au niveau des centres en α ou β de l'amine créé, peu contrôlé pour l'instant avec les AmDHs natives découvertes (cis/trans ratio = 65/35 à 92/8 selon les enzymes et les substrats) (Mayol et al. 2019). L'utilisation de mimiques de cofacteur synthétiques type 1-benzyl-1,4-dihydronicotinamide (BNAH), au détriment des cofacteurs naturels plus onéreux, n'a pour l'instant pas donné lieu à des résultats prometteurs pour cette application (Figure 22).

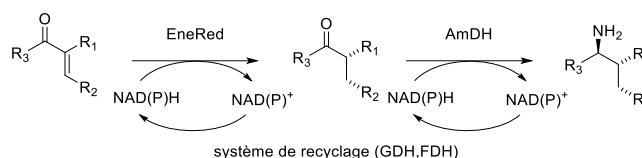


Figure 22. Cascade couplant EneRed et AmDH pour la synthèse stéréocontrôlée d'amines (collaboration Dr. Caroline Paul, Université de Delft, Pays-Bas)

Les premières expériences ont montré que le recyclage du cofacteur NADPH par ces mimiques est trop lent pour permettre la réalisation de cet objectif initial. Les travaux se concentrent donc sur un criblage de différents substrats pour cette cascade avec recyclage du cofacteur par les systèmes communs enzymatiques.

Une première collaboration avec Jack Rowbotham (Université d'Oxford, UK) a également été entamée suite à des discussions avec la responsable du groupe Kylie Vincent lors du congrès Novel Enzymes en 2018. Ces travaux visent à utiliser leur technologie HydRegen (Reeve et al. 2015) pour recycler le cofacteur nicotinamide lors des réactions AmDHs et à développer cette technologie en laboratoire sans équipement critique. Les équivalents réducteurs nécessaires au recyclage proviennent alors de l'oxydation de l'hydrogène par une hydrogénase immobilisée sur particules de carbone. Les premiers résultats indiquent des conversions de l'ordre de 20% sur le substrat test cyclohexanone, mais les conditions opératoires nécessitent pour l'instant l'utilisation d'une boîte à gant pour minimiser la présence d'oxygène inhibant l'activité de l'hydrogénase essentielle à cette technologie (Figure 23). Un court séjour du Dr. Jack Rowbotham en mars 2019 a déclenché l'effectivité de cette collaboration avec comme objectif notre prise en main des manipulations utilisant ce système. Des essais complémentaires sont en cours dans leur laboratoire avant d'envisager la suite du projet.

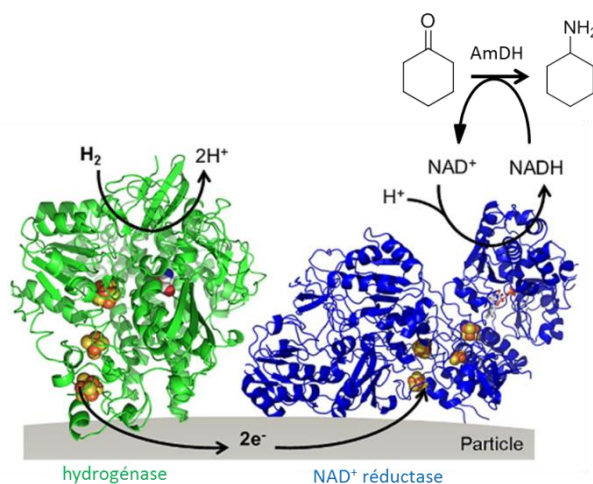


Figure 23. Projet de transformation AmDH utilisant le système HydRegen (collaboration Pr. Kylie Vincent, Université d'Oxford, UK) (en partie extrait de la publication Reeve et al. 2015)

1.3.3. Participation active au sein de projets de biocatalyse

En plus du projet nitrilase co-supervisé et AmDH que je dirige, j'ai contribué principalement à deux autres projets de biocatalyse, l'un interne à l'unité sur l'activité d'oxygénase α -cétooglutarate dépendante et l'autre en collaboration sur la phosphorylation de sucres C6 par les kinases ATP-dépendantes.

a) Découverte et utilisation en réaction cascade de dioxygénases Fe(II)/ α -cétoacides dépendantes.

Toujours dans le but d'exploiter l'expertise en génomique/bio-informatique de l'unité à des fins de biocatalyse, un projet visant à la fonctionnalisation de liaisons C-H inactivées a débuté au laboratoire en 2010. Les enzymes cibles étaient les dioxygénases Fe(II)/ α -cétoacides dépendantes (α KAO), répertoriées pour hydroxyler la chaîne latérale d'acides aminés. En combinant une approche par comparaison de séquences deux à deux, et par partage de domaines protéiques, des nouvelles dioxygénases catalysant l'hydroxylation de la L-lysine (KDO1-6) et de l'ornithine (ODO), ont été découvertes et utilisées pour la synthèse des produits mono et dihydroxylés correspondants (Baud et al. 2014). La haute stéréosélectivité et les différentes régiosélectivités observées ont été appréhendées par des analyses structurales, notamment grâce à l'obtention de la structure de KDO1 (*Catenulispora acidiphila*) et de KDO5 (*Flavobacterium* sp.) qui réalisent une hydroxylation en C3 et en C4 respectivement. Confortée par la modélisation des 523 sites actifs de l'ensemble de cette famille (classification ASMC déjà mentionnée dans le cas des AmDHs), la régiosélectivité semble être expliquée par la présence ou non d'une arginine responsable du positionnement précis du substrat au sein de la poche pour favoriser une hydroxylation en C4. La présence d'une boucle flexible et d'une sous-poche formée après liaison du substrat pourrait être à l'origine de cette catalyse très stéréosélective, même si l'importante flexibilité de cette boucle empêche une analyse poussée des interactions mises en jeu par manque de résolution (Bastard et al. 2018).

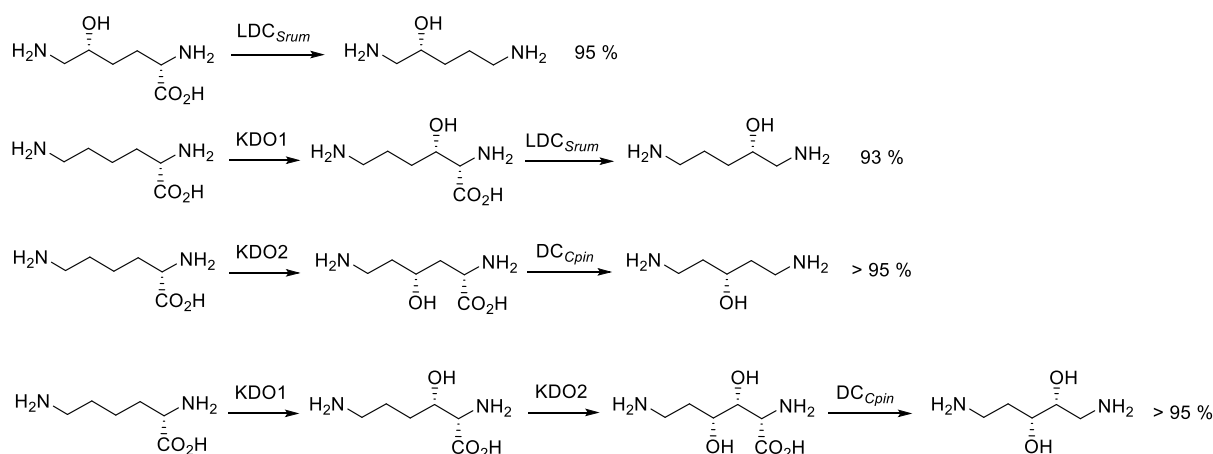


Figure 24. Synthèse des β -et γ - aminoalcools par couplage enzymatique de dioxygénases fer(II)/ α -céto acid-dépendantes et de décarboxylases

Le projet a été poursuivi par la mise en place d'une cascade enzymatique impliquant ces enzymes et des décarboxylases (lysine décarboxylase de *Selenomonas rumirantium* LDC_{Srum}, et la décarboxylase de *Chitinophaga pinensis* DC_{Cpin}) afin de synthétiser des amino-alcools aliphatiques chiraux, molécules très utilisées comme synthons en chimie organique. L'optimisation des conditions opératoires pour coupler en réaction ces deux enzymes, a permis la synthèse de 4 β -et γ - aminoalcools avec d'excellents rendements (Figure 24) (Baud et al. 2017).

J'ai proposé activement tout au long du projet des choix techniques pour pallier des problèmes expérimentaux, et des choix plus stratégiques en réelle discussion biciphale avec Anne Zaparucha en charge de ce projet. J'ai ainsi été en support des étudiants et de la technicienne du laboratoire Aurélie Jouenne travaillant sur ce projet.

b) Caractérisation d'une kinase ATP-dépendante thermotolérante et application pour la synthèse biocatalysée de sucres phosphorylés.

L'autre projet auquel j'ai activement participé est la synthèse biocatalysée de sucres phosphorylés par une kinase ATP-dépendante nouvellement caractérisée. Ce travail de recherche est né du projet Genozyme (ANR Genozymes (2008 – 2012), projet du CAB/LGBM en collaboration avec l'équipe SEESIB de l'Institut de Chimie de Clermont Ferrand et Isthmus. Les nombreux résultats obtenus sur ce projet portant sur la découverte d'aldolases par approche génomique, ont donné lieu depuis à de nombreux projets annexes. Parmi ces derniers, la recherche d'hexokinases pour synthétiser enzymatiquement des stéréoisomères d'hexoses-6-phosphates. Une collection de 90 protéines représentatives de la biodiversité bactérienne a été créée, par la même approche génomique que celle décrite pour le projet nitrilase ou AmDH, en utilisant la fructokinase FrK de *Zymomonas mobilis* comme référence. Parmi les protéines produites, 61 se sont révélées actives sur au moins un des 7 hexoses testés (D-fructose, D-tagatose, D-sorbose, D-psicose, D-glucose, D-mannose et D-galactose) via un suivi spectrophotométrique enzymatique. Une étude de thermotolérance a amené à sélectionner l'enzyme issue de *Streptococcus mitis* (Uniprot id: E1M3M4, ScrK_{mitis}). En plus d'être purifiable par choc thermique et thermoactive (activité maximum à 55°C), cette enzyme possède un très large spectre de substrats allant des hexoses naturels ou non à des C4 et C5 oses. Son efficacité catalytique est très importante avec le D-mannose ($1,5 \cdot 10^6 \text{ s}^{-1} \text{ M}^{-1}$) et le D-fructose ($2,7 \cdot 10^5 \text{ s}^{-1} \text{ M}^{-1}$) mais aussi avec le D-tagatose ($3,5 \cdot 10^2 \text{ s}^{-1} \text{ M}^{-1}$). Des activités spécifiques significatives ont été mesurées avec tous les hexoses naturels ($> 0.1 \text{ U mg}^{-1}$) excepté le D-galactose, ainsi qu'avec des sucres rares comme le D-psicose, le D-allose et le D-gulose ($< 0,1 \text{ U mg}^{-1}$ d'enzyme). L'applicabilité potentielle de cette enzyme a été démontrée *via* la synthèse biocatalysée ATP-dépendante de 8 hexoses phosphorylés, en utilisant la régénération *in situ* de l'ATP par le système phosphoenolpyruvate/pyruvate kinase. La position de phosphorylation a pu être prouvée par RMN (Table 1). Cette étude biochimique et biocatalytique est à notre connaissance la plus vaste répertoriée pour une fructokinase encodée par un gène identifié.

Table 1. Liste des hexoses phosphorylés synthétisés biocatalytiquement avec l'enzyme ScrK_{mitis}.

Produits	Conversion (%) (temps)	Rendement isolé (%)	Forms présentes (%) ^a			
			α-furanose	β-furanose	α-pyranose	β-pyranose
D-fructose-6P	>99 (30 min)	72	21	79	-	-
D-psicose-6P	70 (4 days)	51	35	65	-	-
L-sorbose-5P	>99 (24 h)	67	-	-	> 95	-
D-altrose-6P	90 (10 h)	57	15	11	31	43
D-tagatose-6P	>99 (2 h)	51	25	75	-	-
D-glucose-6P	>99 (30 h)	56	-	-	35	65
D-mannose-6P	>99 (15 h)	77	-	-	66	34
D-talose-6P	95 (7 h)	77	13	10	41	36

(-) : non détecté ; ^a déduit des spectres ¹³C RMN

Mon implication au sein de ce projet a été principalement sur l'ensemble du développement applicatif de ScrK_{mitis}, avec notamment l'optimisation des conditions de réactions, la réalisation des synthèses biocatalysées à l'échelle de 0,15 mmol et la purification des produits phosphorylés par HPLC préparative pour obtenir 20-30 mg de composés purs analysés ensuite par RMN (1D ¹H et ¹³C, mais

aussi 2D ^1H - ^{31}P en collaboration avec Jean-François Gallard de l'ICSN). J'ai réalisé une partie de ce travail expérimental et supervisé le complément (Stessy Nepert, stagiaire de M2, et Aurélie Jouenne). J'ai de plus pris part aux discussions et décisions relatives à l'avancement du projet, ainsi qu'à la rédaction de la publication associée à ces résultats (Vergne-Vaxelaire et al. 2018).

1.3.4. Support de projets internes

Outre les projets de biocatalyse décrits ci-dessus, je suis aussi impliquée dans d'autres projets internes *via* mon expertise en chimie organique. Ce volet intervient en effet dans plusieurs projets menés dans l'unité, que ce soit *via* la synthèse de composés nécessaires aux expériences du projet ou *via* une expertise sur la structure chimique et la réactivité des substrats pour orienter/expliciter les résultats.

J'ai ainsi réalisé la synthèse des standards chimiques *O*-succinyl- et *N*-succinyl-L-serine pour permettre la validation expérimentale d'enzymes acyl-L-homoserine transférases impliquées dans la voie de biosynthèse de la méthionine, tout en étant aussi en support des réflexions liées à leur réactivité (Bastard et al. 2017). Au début de mon travail au sein de l'équipe de biocatalyse, j'ai également synthétisé par biocatalyse le (2*R*,4*S*)-diaminopentanoate et par voie de chimie conventionnelle son diastéréoisomère (2*R*,4*R*)-diaminopentanoate. Ces synthèses ont permis l'étude et la caractérisation de deux des enzymes impliquées dans la voie de dégradation anaérobie de l'ornithine élucidée au sein de l'unité en 2009 (Fonknechten et al. 2009). Mes compétences chimiques de réactivité de molécules m'ont également amené à proposer et à réaliser avec Anne Zaparucha la synthèse d'analogues d'acyl-CoA (*N*-acetylcysteamine thioesters) pour faciliter l'étude d'un projet portant sur une voie alternative de fermentation de la lysine (Perret et al. 2011). J'ai aussi mis d'importants efforts dans la synthèse de standards pour valider des hypothèses structurales de nouveaux métabolites issus du projet d'étude du métabolisme d'*Acinetobacter baylyi* ADP1 par approche métabolomique (Thomas et al. 2019).

Enfin, l'équipe COB a également été impliquée dans un important projet collaboratif de l'unité s'intéressant aux « β -keto acid cleavage enzymes » (BKACE) *via* une stratégie intégrée de découverte d'activités enzymatiques au sein de famille de protéines aux fonctions inconnues (Bastard et al. 2014). La caractérisation préliminaire biochimique et structurale de l'enzyme « Kce » impliquée elle aussi dans la voie précédemment énoncée de fermentation de la lysine (Bellinzoni et al. 2011), nous a en effet amené à penser que la réaction catalysée (la condensation du β -keto-5-amino-hexanoate et de l'acetyl-CoA) pouvait être généralisée à l'ensemble des membres de sa famille Pfam possédant un domaine conservé sans fonction associée au début du projet (DUF849). L'analyse de la réactivité des substrats nécessaires à la réaction nous a permis de proposer une liste de différents β -cétoacides pour analyser puis comprendre la variabilité des activités catalysées par ces enzymes, pas toutes issues d'organismes capables de fermenter la lysine (Figure 25). Les résidus impliqués dans la catalyse et la reconnaissance du substrat ont été identifiés dans les sous-familles définies par la classification ASMC réalisée sur cette famille, en collaboration avec le LABGEM et le LGBM.

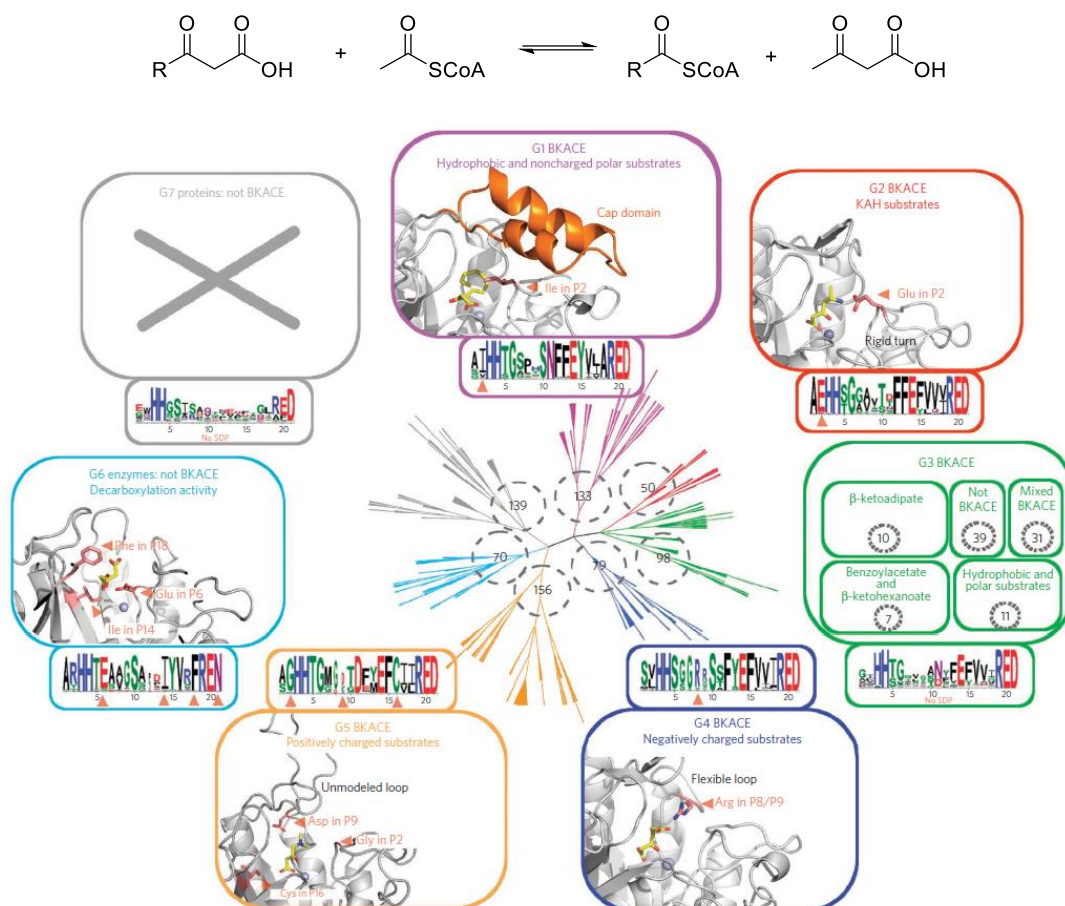


Figure 25. Classification ASMC de la famille BKACE catalysant la condensation de β-cétoacides et d'acétyl-CoA, et type de substrats associés à chaque sous-groupe de la famille.

1.4. Bilan

Lors de ces différentes expériences de recherche, j'ai donc pu acquérir, à partir d'une formation initiale en chimie organique, des compétences dans des domaines connexes où l'expertise initiale en synthèse organique et en méthodes analytiques associées est un atout. De ces compétences acquises en synthèse des produits naturels et biocatalyse sont nées des connaissances dans différentes thématiques de recherche en interaction directe ou indirecte avec ces dernières. En particulier, les travaux de recherche actuels, menés de façon très collaborative et multidisciplinaire au sein de l'unité de recherche, me permettent aujourd'hui d'élaborer, diriger et/ou participer à des projets nés de ces disciplines très éloignées de ma formation d'origine (Figure 26). Certains de ces projets sont développés dans la partie suivante.

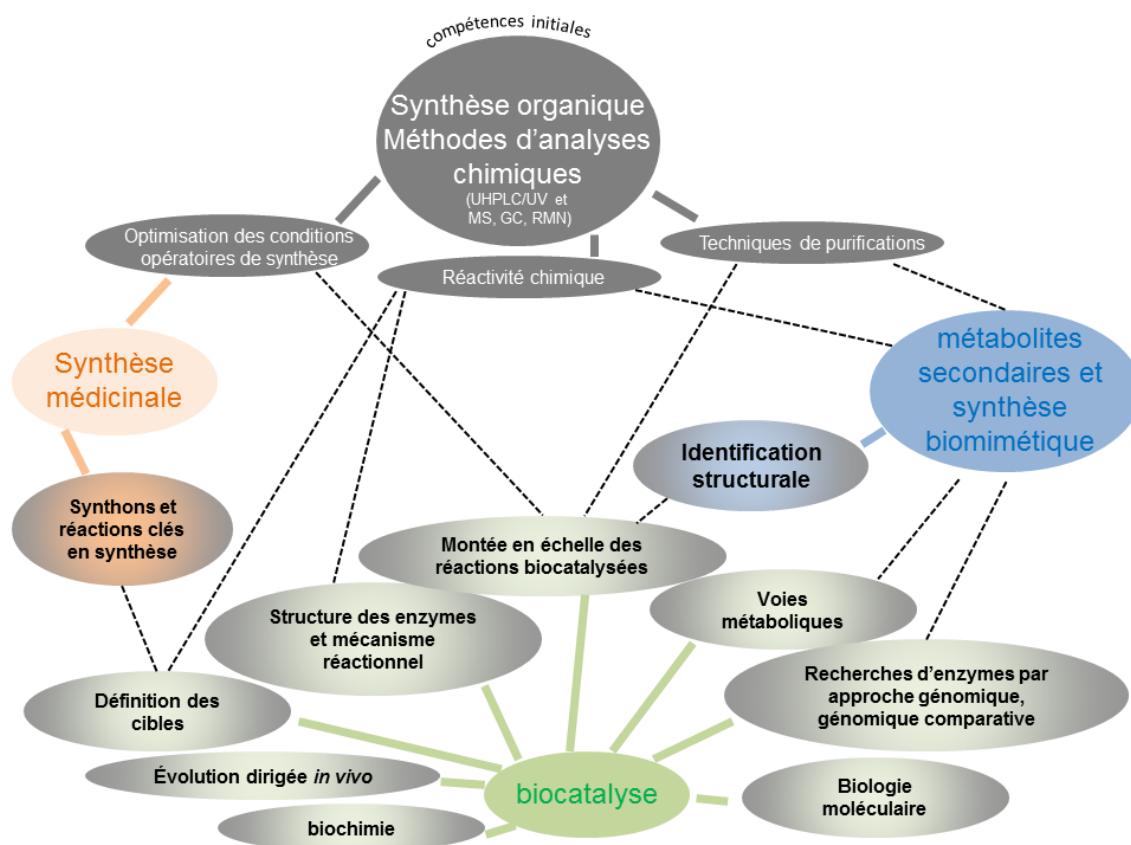


Figure 26. Compétences initiales et connaissances acquises au cours de mes travaux de recherche, et interaction entre celles-ci.

Mon implication dans les trois thématiques de recherche – synthèse organiques de molécules biologiquement actives, synthèse biomimétique et extraction de produits naturels, synthèse biocatalysée de synthons chimiques – est, et restera, une richesse pour mes recherches.

II. Projets scientifiques

La suite de mes recherches vont majoritairement s'orienter sur la niche ouverte par la découverte des AmDHs natives avec un objectif de mettre en place des techniques / développer des expertises nouvelles pour moi et/ou l'unité, en collaboration, pour étendre les moyens d'étude de cette famille et de ses applications en biocatalyse. Les démarches appliquées sur cette famille pourront alors être étendues à d'autres activités enzymatiques. En plus des travaux en cours/perspectives déjà évoquées dans le paragraphe I.3.2.c, des projets à court terme et à moyen/long terme sont envisagés.

II.1. *Projets de recherche à court terme : projet MODAMDH*

Comme explicité plus haut, la recherche de nouvelles AmDHs au sein de la biodiversité se justifie par l'intérêt de disposer d'un panel d'enzymes aux propriétés catalytiques et physico-chimiques variées pour la synthèse d'amines structurellement diverses, fonctionnalité présente dans de nombreux composés d'intérêt. Les nat-AmDHs récemment découvertes au laboratoire sont actives sur des cétones et aldéhydes de taille restreinte et non fonctionnalisées par un groupe aromatique, comme la (methyl)cyclohexanone, la pentan-2-one ou le 2-methylpropanal. Les nouvelles enzymes identifiées au sein des métagénomomes étudiés possèdent des variabilités au niveau de la structure du site actif mais qui ne modifient pas considérablement leur spectre de substrat. De plus, seule la stéréochimie (S) au niveau du centre chiral créé est accessible. Des candidats actifs sur des substrats structurellement plus variés et complémentaires d'un point de vue stéréosélectivité permettraient d'élargir leur applicabilité en biocatalyse (Afanasyev et al. 2019). Leur stabilité et aptitude à transformer des concentrations élevées en substrats sont aussi des cibles.

Le projet (nommé MODAMDH) se focalise donc sur la recherche d'homologues et d'analogues AmDHs au sein de la biodiversité, très divers en terme de séquences protéiques, spectres de substrat, sélectivités, stabilités et structures, en utilisant des approches encore peu répandues. Ce projet se différencie nettement des projets précédemment menés par les méthodes employées.

Pour atteindre les objectifs escomptés, des méthodes de bio-informatiques basées plus particulièrement sur la structure tri-dimensionnelle des enzymes et de motifs spécifiques de leur site actif, ainsi que la recherche d'homologues lointains, seront employées. Pour élargir le catalogue, la biodiversité sera criblée non seulement *via* les bases communes type UniProtKB, mais aussi des ressources métagénomiques publiques et/ou internes, encore très peu étudiées d'un point de vue vivier pour la biocatalyse (Figure 27). L'exploration de ces données métagénomiques sera bien plus approfondie que la recherche préliminaire effectuée lors du projet avec Adam Caparco.

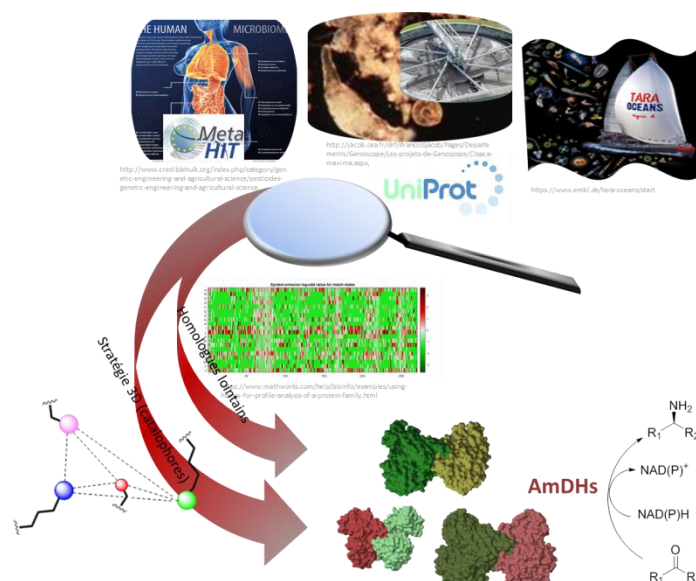


Figure 27. Schéma du projet MODAMDH

II.1.1. Définition du réservoir d'enzymes à cribler

La métagénomique est désormais reconnue comme une technique moléculaire avérée pour fournir des nouvelles enzymes, avec de ce fait un impact considérable sur le succès économique de la biocatalyse (Alma'abadi et al. 2015, Jeffries et al. 2016, Ahmad et al. 2019, Ferrer et al. 2019). Introduite pour la première fois en 1998 pour contourner l'impossibilité de cultiver la plupart des micro-organismes, et donc l'accessibilité de leurs produits, la métagénomique offre une grande opportunité d'étudier l'ADN de toute une communauté et ainsi de prospecter de nouveaux biocatalyseurs. Néanmoins, les nouvelles enzymes découvertes grâce à l'approche métagénomique sont actuellement principalement limitées aux hydrolases (Madhavan et al. 2017), même si d'autres classes d'enzymes comme les transaminases (Baud et al. 2017, Leipold et al. 2019) et déshydrogénases (Jeffries et al. 2016) ont commencé à être étudiées. La recherche d'AmDHs parmi les données métagénomiques peut considérablement augmenter les chances d'accéder à de la nouveauté.

Les protéines à cribler seront donc sélectionnées au sein des bases génomiques (UniProtKB (Apweiler et al. 2004) et la base protéique du NCBI) et métagénomiques (Steinegger et al. 2019) dont les bases déjà brièvement explorées par les travaux d'Adam Caparco. Toutes les séquences géniques codant pour des protéines partageant un domaine de liaison au cofacteur nicotinamide NAD(P) (domaine InterPro IPR036291) seront rapatriées et catégorisées en familles homologues, grâce à des programmes comme MMseq2 (Steinegger et al. 2017) et LINCLUST (Steinegger et al. 2018) développés pour l'analyse massive de données. De plus, l'ensemble des protéines correspondantes seront modélisées. Des modèles 3D seront générés par homologie via des logiciels de modélisation tels que Modeler. Ce type d'approche est envisageable grâce au nombre désormais conséquent de structures protéiques répertoriées. Avec près de 139 000 entrées de structures comprenant plus de 4 000 protéines non caractérisées, la banque de données sur les protéines PDB est maintenant une excellente ressource de structures références pour créer des modèles, en plus d'accéder à de nouvelles enzymes pour une fonction souhaitée (Nadzirin et al. 2012). La détection du site actif dans tous les modèles validés sélectionnés sera effectuée à l'aide des logiciels Fpocket (Le Guilloux et al. 2009).

II.1.2. Approches méthodologiques

a) Définition de profils HMM

Les stratégies basées sur les comparaisons linéaires de séquences ont conduit à la découverte de nouvelles enzymes aux propriétés biocatalytiques variées. Nous pouvons par exemple citer les travaux effectués sur la famille des nitrilases à l'aide des programmes BLASTP qui ont permis la découverte de nouvelles nitrilases, dans des champignons filamenteux (Kaplan et al. 2011) ou dans diverses bactéries (cf les résultats obtenus par notre laboratoire, I.3). Des biocatalyseurs aux spectres de substrats différents sont recherchés dans les protéines possédant des faibles homologies de séquences, tels que la transaminase de *Chromobacterium violaceum* catalysant l'amination stéréosélective d'un large éventail de substrats (Kaulmann et al. 2007), ou nat-AmDHs avec une identité de séquence inférieure à 30 % avec les protéines références. Malgré la pertinence de cette méthodologie, une méthodologie plus puissante est envisagée pour la suite des recherches, principalement en raison de la quantité de données à explorer. Pour cela, des alignements multiples (améliorés manuellement du fait de la connaissance des résidus clés) et des profils HMM (Hidden Markov Models)(Voshol et al. 2017) seront générés à partir de la liste des AmDHs expérimentalement validées, en excluant les domaines de liaison NAD(P).

b) Définition de catalophores

Bien que très substantielles, les stratégies qui traitent des seules séquences primaires conduisent généralement à des enzymes appartenant à des familles déjà connues. L'idée est donc de combiner cette approche avec une approche complémentaire encore très peu répandue pour la biocatalyse : la recherche de familles structurellement apparentées et donc la recherche d'analogues et non d'homologues. À la suite des travaux pionniers du groupe Thornton sur la prévision de la fonction enzymatique basée sur la reconnaissance de modèles de sites actifs 3D (Barker et al. 2003), quelques groupes ont contribué au succès de telles approches. Cependant, ces méthodologies restent rarement appliquées pour la recherche de nouveaux biocatalyseurs à des fins synthétiques. Ces études soulignent que malgré des séquences et / ou des structures en acides aminés complètement différentes, des modes de liaison équivalents des substrats et des ligands d'enzymes peuvent se produire, conduisant à la même réaction catalysée. Dans ces études, la première étape consiste toujours à identifier les résidus clés (ceux qui se lient à des cofacteurs et / ou à des substrats) sur la base de l'analyse de la structure 3D d'au moins un membre de la famille étudiée. La topologie du site actif, à savoir les positions spatiales relatives des résidus essentiels formant la constellation minimale de sites actifs catalytiques (= catalophore), est alors utilisée comme modèle dans les bases de données structurales. Steinkellner et al. ont par exemple identifié par cette approche deux « Old Yellow Enzyme » (OYE) structurellement non apparentés à celles précédemment décrites (Steinkellner et al. 2014). La toute récente spin-off Innophore GmbH propose aux partenaires industriels une base de données exclusive sur la cavité des enzymes 3D *via* la plate-forme Catalophore™ afin d'identifier les enzymes candidates par une recherche ciblée permettant de réduire le délai de mise sur le marché des produits. La méthode bio-informatique repose sur le même concept : exploiter une base de données structurale à l'aide de catalophores.

Dans le cas du projet AmDH, les structures de trois des AmDH - AmDH4, *Msme*AmDH et *Cfus*AmDH - ont été résolues par diffraction aux rayons X grâce à la collaboration avec Gideon Grogan (cf I.3.2.b)

Nous avons ainsi identifié des résidus clés au sein du site actif permettant de postuler un mécanisme. Une famille AmDH a également été définie en analysant la distribution de la poche catalytique 3D parmi la diversité des séquences bactériennes à l'aide de la méthode de classification ASMC. À l'aide de ces données et des données obtenues récemment sur la recherche préliminaire d'AmDHs au sein des microbiomes marins et intestinaux (cf I.3.2.c) les stratégies de recherche basées sur une approche 3D peuvent donc être entreprises. Différents catalophores seront définis, en introduisant plus ou moins de flexibilité (spatiale ou nature des résidus) tout en préservant les résidus catalytiques clés identifiés. La spécificité AmDH de ces constellations sera vérifiée en les comparant à d'autres déshydrogénases identifiées, telles que des alcool déshydrogénases (ADH). Les catalophores seront revus/complétés tout au long du projet suivant les informations déduites de nouvelles structures résolues.

II.1.3. Criblage *in silico*

Le réservoir d'enzymes sélectionné sera criblé pour explorer la diversité suivant les deux approches méthodologiques exposées. Les profils HMM des AmDHs validées seront utilisés comme requête pour rechercher des homologues distants dans l'ensemble des familles de protéines NAD(P)H-dépendantes définies en utilisant l'alignement par paires des profils HMM (logiciel HHsearch (Soding 2005)). Les candidats seront classés selon leur score d'alignements (E_{value}) et la présence du(des) résidu(s) catalytique(s) vérifiée. Parallèlement, les catalophores définis serviront de références pour cribler les sites actifs extraits des modèles générés des enzymes NAD(P)H-dépendantes (logiciel Fit3D (Kaiser et al. 2016)). Les candidats positifs seront analysés manuellement avec Pymol (Schrödinger®) afin de visualiser la structure entière et la géométrie du site actif par rapport aux modèles d'origine à partir desquels les catalophores ont été extraits. La flexibilité et la nature des catalophores pourront être revues selon le nombre de séquences positives. L'approche diffère considérablement de ce qui est couramment fait à des fins de biocatalyse car elle est indépendante de la structure globale, du repli protéique et de la séquence d'acides aminés. Les AmDH identifiées avec cette technique auront potentiellement des propriétés modifiées, dont certaines peuvent correspondre aux critères de la synthèse d'amine ciblée.

II.1.4. Validation *in-vitro* des candidats et analyse *in silico* des AmDHs validées

Les candidats sélectionnés suite au criblage *in silico* seront ensuite testés *in vitro* pour leur activité AmDH sur un panel de substrats, afin de valider le cas échéant par l'analyse des sites actifs *via* les modèles 3D générés. Pour cela, les gènes correspondants seront clonés dans un hôte d'expression (*E. coli*) à l'aide de la plate-forme de clonage du Genoscope. Après la surexpression de ces gènes recombinants, les cellules seront lysées et le lysat cellulaire brut obtenu sera testé pour l'activité AmDH. L'activité en l'absence de donneur d'amine sera vérifiée pour éliminer tout ADH.

Des analyses structurales incluant des études de « docking » de substrats / produits sera effectuée sur les modèles d'AmDHs pour comparer leur structure globale et leur site actif. Des essais de résolutions structurales seront entrepris pour certains, ceci afin de continuellement fournir des informations supplémentaires et ainsi d'avoir un projet fonctionnant en « active learning » (Figure 28), notamment pour la définition des catalophores. Ces analyses seront aussi précieuses pour proposer des mutants répondant aux objectifs. Le positionnement du site actif des différentes enzymes NAD(P)H dépendantes dans la structure globale modélisée/réolue sera aussi étudiée afin d'éventuellement

observer une corrélation avec la gamme de substrats. Il est à noter que des propriétés autres que leur spécificité de substrats seront largement recherchées, dont la stabilité, la résistance à des fortes concentrations en substrats/produits, la stéréosélectivité...

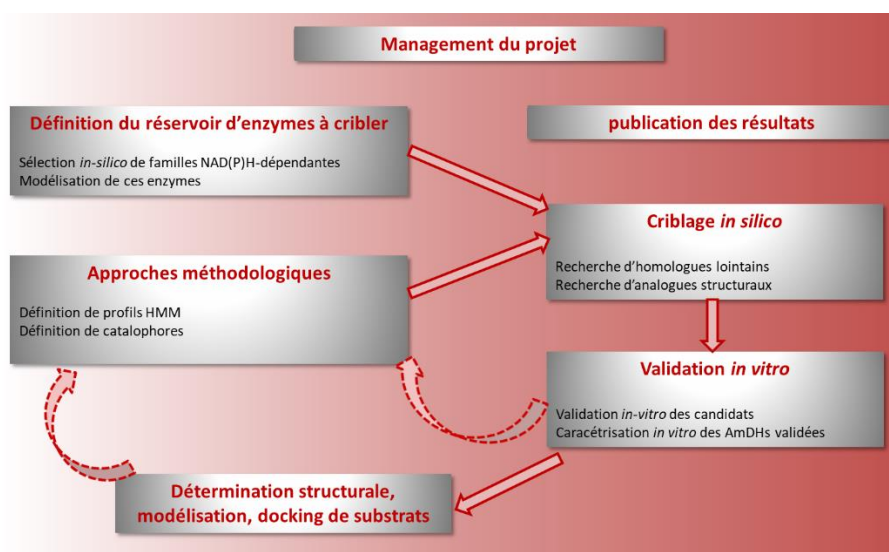


Figure 28. Schéma récapitulatif des étapes envisagées pour le projet MODAMDH

À court terme, ce projet complétera le catalogue d'enzymes pour la synthèse d'amines en fournissant de nouveaux châssis AmDHs et une meilleure compréhension de la superfamille associée. Ceci est essentiel pour élargir les capacités enzymatiques et ainsi fournir des solutions biocatalytiques répondant le mieux aux besoins industriels.

À moyen et plus long termes, les résultats escomptés publiés devraient favoriser le développement des approches de recherche d'enzymes dans la biodiversité par catalophores ou homologues lointains par les groupes de recherche internationaux orientés vers la biocatalyse. Ils seront aussi je l'espère une motivation supplémentaire pour le développement d'outils bio-informatiques encore plus performants et adaptés à la réalisation de tels criblages, en particulier pour l'exploration 3D.

Ce travail nécessite des compétences très variées allant de la bio-informatique aux tests biochimiques et réactions biocatalysées, et sera donc très collaboratif. L'obtention d'une ANR (JCJC) en 2019 permet entre autre le financement du post-doctorant Eddy Elisée pour la période 2020-2022 dont j'axerai les travaux sur le volet bio-informatique, supervisés au quotidien par David Vallenet et Mark Stam du LABGEM. Le volet test *in vitro* sera traité par l'étudiante en thèse Laurine Ducrot qui participera activement aussi à l'analyse structurale des enzymes sélectionnées. Le réseau de collaboration entre l'Université de York, Georgia Tech Institute et notre Unité de Recherche sera renforcé par ce projet.

II.2. Evolution des expertises/thématiques

Depuis la création de l'équipe de biocatalyse en 2007, les thématiques de recherche sont restées centrées sur la recherche, par approche génomique, d'enzymes aux propriétés d'intérêts et leur utilisation en synthèse de synthons.

Les travaux de recherche pour l'instant réalisés dans l'unité en biocatalyse pour la synthèse chimique se sont centrés uniquement, pour le côté laboratoire humide, sur des études *in vitro*. La recherche de nouvelles activités enzymatiques au sein de la biodiversité pour synthétiser biocatalytiquement des composés chimiques d'intérêt explique ce choix initial. Cette exploration de la biodiversité *via* la constitution de collections de gènes clonés pour assigner une activité à l'enzyme correspondante, surexprimée dans un organisme hôte, est en effet réalisée *in vitro* pour assurer la compréhension et la maîtrise de la réaction étudiée. Les résultats obtenus et l'expertise acquise en biocatalyse nous permettent maintenant d'envisager de réaliser des projets de synthèse en biotransformation, c'est-à-dire en utilisant des cellules entières. Je souhaite être motrice dans ce tournant d'activité, en l'appliquant entre autre à des synthèses mettant en jeu plusieurs étapes, des cascades enzymatiques. Après plusieurs années d'exploration qui ont permis d'enrichir grandement la boîte à outils au travers de la découverte de nombreux biocatalyseurs appartenant à de nombreuses familles variées, nous pouvons en effet désormais envisager de coupler plusieurs de ces enzymes découvertes à des fins de synthèses, en plus des enzymes déjà décrites dans la littérature. Les recherches en biotransformations en cellules entières sont de plus en plus étudiées par les groupes de recherche publics de biocatalyse et devraient continuer à prendre de l'ampleur (Rudroff 2019). Tout en ne laissant pas de côté la recherche de nouvelles activités enzymatiques *in vitro* et les études structurales associées, étudier le volet biotransformation est une activité qu'il est selon moi nécessaire de développer dans l'unité. Nous devons prendre ce virage tout en valorisant des expertises qui nous sont plus propres afin de proposer des résultats novateurs. Ceci permettra également une plus grande synergie entre les équipes de biorémédiation et de métabolisme synthétique qui composent désormais depuis janvier 2020 le nouveau laboratoire LB2MS avec notre équipe de biocatalyse. Une vision intégrée de l'exploration de la biodiversité, dont les métagénomiques, de l'expertise en évolution dirigée *in vivo* propre à l'unité, et des besoins en biocatalyse est au cœur des projets envisagés où la multidisciplinarité de l'unité est un élément clé.

II.2.1. Biotransformations

La formulation du biocatalyseur est un point clé pour un procédé. Actuellement, environ 60% des procédés industriels biocatalytiques se font sur cellules entières et non sur enzymes isolées. Même si le coût de la lyse et de la purification des enzymes est en soi un facteur qui peut expliquer ces choix (l'utilisation de cellules entières réduisant le coût du catalyseur par environ 10), le point important est surtout la contribution du biocatalyseur au coût du produit cible, la valeur ajoutée de l'utilisation de la biocatalyse par rapport aux autres méthodes de production. Ceci étant bien évidemment lié au tonnage du produit cible et donc à sa valeur ajoutée. Les productivités requises du système enzymatique ne sont donc pas les mêmes suivant le volume de production de la cible, l'intérêt d'un système *in vivo* vs *in vitro* dépend donc surtout de la cible (Tufvesson et al. 2011).

A l'échelle du laboratoire, même si le but n'est pas, pour l'unité, le développement d'un procédé optimisé pour la production, il est important d'explorer en recherche la viabilité ou non de telle synthèse/cascade enzymatique en biotransformation. D'un point de vue pratique, un des intérêts majeurs est la non nécessité d'implémenter un système de régénération des cofacteurs, déjà présents dans les voies métaboliques toujours actives de la cellule hôte, et d'éviter les étapes de lyses des cellules et d'isolement/purification des protéines.

Pour débiter cette activité de recherche, les activités enzymatiques étudiées seront entre autre les AmDHs afin de valoriser les résultats déjà obtenus et futurs sur cette réaction, la synthèse d'amines restant une cible importante en biocatalyse. Le projet envisagé serait donc d'implémenter *in vivo* des cascades enzymatiques visant une amine cible *via* une étape d'amination réductrice catalysée par une AmDH. L'intérêt des AmDH demeurant dans le seul prérequis en ammoniac. Des travaux ont déjà été publiés sur la mise en place de telles cascades enzymatiques, notamment par le groupe de F. Mutti. Cependant, comme dans la plupart des études de biotransformations, la productivité est limitée par la toxicité des substrats/produits, ici l'amine cible. Ainsi, lors de la biotransformation qui peut durer jusqu'à 48h, une partie des cellules meurt, diminuant ainsi les productivités (Figure 29) (Houwman et al. 2019).

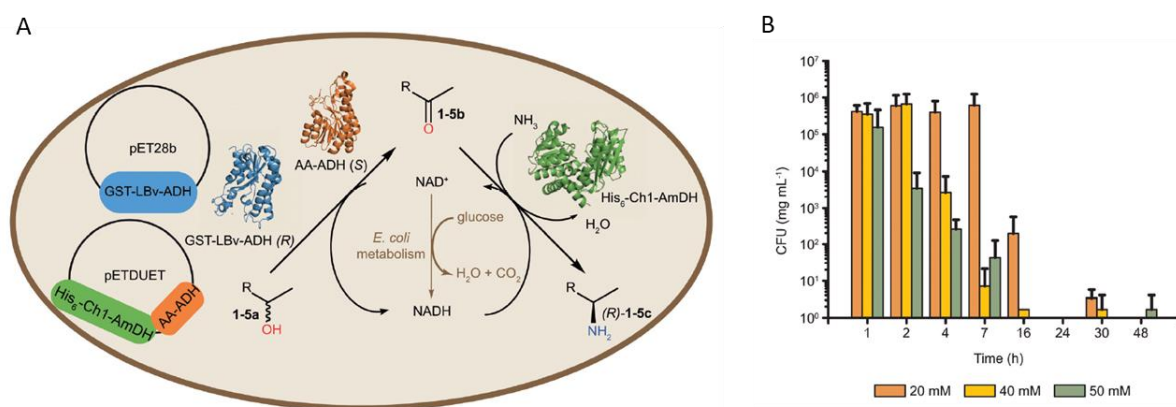


Figure 29. Cascade enzymatique alcools vers amines réalisée en biotransformation, travaux de l'équipe de F. Mutti. **A** Cellule d'*E. Coli* modifiées pour la réalisation de la cascade d'amination d'alcools. **B** Quantification de la survie de ces cellules (CFU) à différentes concentrations en substrat 2-hexanol. Schémas extraits de la publications associée (Houwman et al. 2019).

Une partie du projet sera donc d'étudier l'influence de différentes souches d'*E. coli*, obtenues par évolution *in vivo*, sur la productivité en amine cible et d'obtenir ainsi des souches qui permettent d'accéder à des productivités plus élevées que la souche sauvage.

En utilisant la technologie des GM3 qui permettent d'évoluer des souches bactériennes soumises à des pressions de sélection (cf annexe 2), l'idée serait donc de développer des châssis de production plus adaptés pour la synthèse d'amines, c'est-à-dire moins impactés d'un point de vue survie cellulaire par des concentrations élevées en amines cible/substrat ammoniac ou autre composé impliqué dans la synthèse à implémenter. Contrairement aux travaux décrits en partie I.3.2.c où l'objectif est d'utiliser les GM3 à des fins d'amélioration des biocatalyseurs, il s'agit dans ce cas d'adaptation des souches à des substrats toxiques/inhibiteurs.

Dans un premier temps, cette étude sera menée sur une réaction test comme celle détaillée par Mutti *et al* (Houwman et al. 2019) afin d'évaluer la potentialité de tels travaux. Les cibles sont alors des alkylamines simples (hexan-2-amine dans ce cas), mais les spécificités de substrats des AmDHs de notre équipe nous permettent de cibler la *n*-butylamine, l'isobutylamine ou la cyclohexylamine. Ces molécules à forts volumes de production (la cyclohexylamine est un des plus important produits d'un point de vue commercial avec une production de plus de 1000 mt par an), sont actuellement produites à partir de l'alcool ou cétone/aldéhyde correspondant par amination réductrice à très fortes températures (150 -210 °C) en présence de métaux (Hayes 2001). L'implémentation d'une synthèse

biocatalysée *in vivo* dans des souches améliorées pour l'une de ces amines serait un axe de recherche intéressant, qui pourrait donner lieu à des résultats complémentaires des travaux déjà publiés (Wendisch et al. 2018, Yu et al. 2018, Houwman et al. 2019). Si des résultats notables sont obtenus, des châssis de production adaptés à la synthèse de molécules cibles différentes (amino-cyclocétones, précurseur du tramadol...), *via* une cascade enzymatique, seront alors envisagés.

Lors de ces études, je souhaite également mener des essais de biotransformations en phase de croissance et non en phase de repos. En effet, alors que la majorité des biotransformations sont décrites avec des cellules en repos dans un milieu uniquement dédié à la réaction étudiée, on peut se demander si maintenir une phase de croissance minimum pendant la biotransformation ne permettrait pas de maintenir, en plus d'un taux stabilisé de cellules actives, le stock de cofacteurs. Ceci afin d'éviter leur décroissance au fur et à mesure de la réaction, qui peut être assez longue, en particulier avec les AmDHs (24 – 48h). Ceci serait d'autant plus envisageable avec l'utilisation des souches évoluées pour leur résistances aux conditions opératoires des réactions cibles.

La reproductibilité des transformations opérées sur cellules, dont le métabolisme est encore actif au contraire du travail avec des enzymes isolées, est un point délicat qui nécessite une réelle prise en main. Les réactions secondaires intervenant dans la cellule et les procédures d'extraction du produit à analyser complexifient les expériences, en plus de la maîtrise des cultures. C'est donc une activité de recherche à part entière qui nécessitera du temps et du personnel.

Il est à noter que je participe également à des discussions portant sur la volonté de mettre en place des criblages *in vivo* de différentes enzymes impliquées dans une cascade cible. Transposer des candidats sélectionnés pour les meilleurs rendements obtenus *in vitro* dans des systèmes *in vivo* est en effet généralement générateur de résultats décevants, les meilleurs enzymes *in vivo* étant souvent d'autres candidats non retenus au préalable (Hepworth et al. 2017, Rudroff 2019).

II.2.2. Projets axés vers la fermeture du cycle de carbone

Sous l'impulsion du Haut-Commissaire du CEA, la mise en place de projets interdisciplinaires en rupture axés sur la fermeture du cycle de carbone sont incités et soutenus au sein d'un programme FOCUS. J'ai répondu à cet appel avec l'équipe de métabolisme synthétique du laboratoire en proposant un projet ambitieux tourné vers le développement de souches d'*E. coli* autotrophes pour des applications biotechnologiques. Toujours dans cette même ambition d'inciter des projets communs entre les trois équipes du laboratoire autour de la synthèse de molécules d'intérêts par des cellules entières, ce projet permettrait également de progressivement orienter les thématiques de recherche vers les sujets prioritaires du CEA et de la société.

a) Contexte

En utilisant massivement des ressources carbonées fossiles, notre système énergétique actuel génère au niveau mondial 37,1 milliard de tonnes de CO₂/an dont près de la moitié s'accumule dans l'atmosphère. Dans ce contexte, le développement de technologies de transformation du CO₂ et de ses dérivés en produits synthétiques utiles à partir d'énergie bas-carbone est un axe de recherche à privilégier. Ceci est communément appelé « fermeture du cycle de carbone ».

En plus des solutions technologiques apportées par la chimie et les matériaux, des solutions biologiques sont envisageables. Les organismes photosynthétiques ainsi que certains microorganismes non photosynthétiques, dits chémolitho-autotrophes, convertissent le CO_2 en biomasse en obtenant l'énergie nécessaire pour la réduction du CO_2 , respectivement par captation de photons ou par oxydation de molécules inorganiques, en particulier l'hydrogène. La capacité de fixation de CO_2 de ces microorganismes peut être exploitée dans ce cadre de fermeture du cycle de carbone.

b) Description du projet

Le projet exploratoire conjoint a pour objectif d'adapter *E. coli* à un mode de croissance autotrophe, où le CO_2 , seule source de carbone, est assimilé par une voie synthétique dont l'intermédiaire fondamental est le formiate, et l' H_2 est la source d'énergie. Une telle souche constituerait un châssis biotechnologique pour implémenter des voies hétérologues de production de molécules d'intérêt à partir de CO_2 et d' H_2 (Figure 30). Les connaissances actuelles du métabolisme bactérien et du répertoire des activités biocatalytiques, alliées au large éventail des techniques de modifications des génomes dont celui de *E. coli*, permettent en effet d'implémenter un régime métabolique autotrophe dans un organisme hétérotrophe (Gleizer et al. 2019).

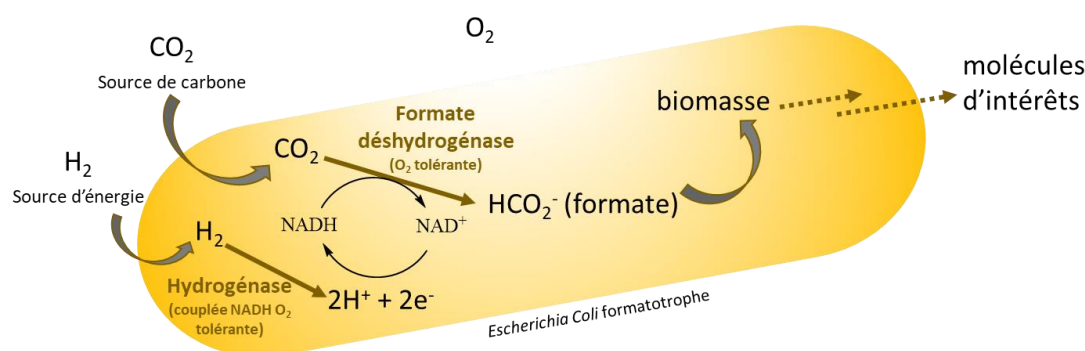


Figure 30. Projet de développement d'une souche autotrophe d'*E. coli* à des fins d'utilisation du CO_2 .

Le projet a d'abord été construit avec l'idée de valoriser l'aptitude de certaines formiate déshydrogénases (FDH) à catalyser la réduction du CO_2 en plus de sa formation par oxydation du formiate. Déjà très utilisées dans le sens oxydatif pour le recyclage des cofacteurs NAD(H) comme nous le faisons dans le projet AmDH par exemple, les FDHs restent à explorer et améliorer pour le sens réductif. Les FDHs forment une famille d'enzymes structurellement diverses, faisant appel à des donneurs/accepteurs d'électrons variés (NAD(H), ferrédoxine). Elles se distinguent notamment par le sens préférentiel de la réaction catalysée. Dans un but d'avoir un système peu contraignant et donc peu sensible à l'oxygène tout en restant dans des efficacités catalytiques notables, nous avons souhaité nous focaliser sur des FDHs NADH/métal dépendantes. Ces dernières, exprimées chez des organismes aérobies, ont montré en présence d'oxygène des activités réductrices *in vitro* intéressantes (Yu et al. 2019) (maximum $k_{\text{cat}} = 10 \text{ s}^{-1}$, en comparaison de $0,3 \text{ s}^{-1}$ pour les FDH NADH-dépendantes insensibles à O_2 et $2,5 \cdot 10^3 \text{ s}^{-1}$ pour les FDH métal-dépendantes NADH- indépendantes O_2 sensibles). Le premier objectif consisterait à évoluer ces activités d'un facteur 10 afin d'ouvrir des champs d'applications aussi bien *in vivo* (assimilation du carbone dans la biomasse) qu'*in vitro* (immobilisation avec couplage électrochimique, insertion dans des cascades enzymatiques...).

Pour ce faire, après vérification de la fonctionnalité de telles enzymes dans *E. coli* et de la réduction du CO₂ en formate par ces FDHs *in vivo*, une évolution utilisant les GM3 est envisagée dans une souche dont la croissance dépend de l'activité réductrice FDH (Doring et al. 2018) dans des conditions de cultures contrôlées (enrichissement en CO₂ de la phase gazeuse, sources d'énergie appropriées). Si des FDHs mutées issues de cette adaptation sont obtenues, ces dernières seront alors caractérisées enzymatiquement et structuralement pour rationaliser l'effet des mutations. Leur incorporation dans une matrice d'hydrogel conductrice est actuellement discutée avec Luca Albertin (CEA, Institut de Recherche Interdisciplinaire de Grenoble (IRIG)).

Afin de fournir l'énergie nécessaire aux cellules pour permettre leur croissance sans apport organique et ainsi se positionner réellement dans un procédé très bas carbone, cette idée initiale a donné naissance à un projet plus complet de développement de souches d'*E. coli* autotrophes où l'énergie est fournie par H₂. L'autre partie importante de ce projet est donc l'implémentation dans cet organisme d'hydrogénases (Hase) NAD couplées montrées tolérantes à l'oxygène (Lenz et al. 2010, Lenz et al. 2018). Pour ce faire, une collaboration avec l'équipe de Christine Cavazza de l'IRIG spécialisée dans les enzymes métal-dépendantes et notamment les hydrogénases a été établie. Leur fonctionnalité dans *E. coli* sera le premier jalon important à valider avant d'envisager de les implémenter conjointement avec les FDHs évoluées dans une souche *E. coli* formatotrophe pour générer ainsi de la biomasse à partir d'uniquement CO₂ et H₂.

Ce projet a fait l'objet d'un dépôt de demande de thèse (directeur de thèse Volker Döring) dans le cadre des programmes FOCUS du CEA.

III. Activités annexes

III.1. Activités d'encadrement

L'ensemble des travaux de recherche réalisés au Genoscope et présentés ici sont les résultats d'expérimentations réalisées en grande partie par des étudiants, hormis quelques projets de support en synthèse de standards pour lesquels j'ai été la seule contributrice. Le laboratoire a en effet eu la volonté et la possibilité financière d'accueillir régulièrement des stagiaires.

Pour l'ensemble des stagiaires accueillis dans l'équipe, j'ai réalisé leur supervision dans le laboratoire, que ce soit au niveau de leur formation aux logiciels et instruments d'analyses (UHPLC, GC, spectrophotométrie) que des techniques opératoires, liées entre autre à la manipulation des enzymes. De plus, j'ai à chaque fois présenté les aspects concrets de sécurité liés à notre laboratoire en complément de leur formation générale réalisée par le CEA. Je poursuis moi-même des travaux en laboratoire, à hauteur environ de 10% de mon activité à l'heure actuelle, afin de continuer à maîtriser les techniques expérimentales dans ce domaine pour procurer les conseils/orientations au plus juste aux étudiants.

Pour les stagiaires impliqués dans les projets présentés ici, c'est-à-dire auxquels j'ai pris part en direction ou en support, je les ai en plus encadrés d'un point de vue analyses des résultats, décisions d'avancement stratégiques du projet, préparation de présentations orales et rédaction de rapports

et/ou publications, toujours en travail biciphale avec Anne Zapparucha. J'ai supervisé avec Anne Zapparucha les travaux de thèse de Franck Bordier et j'ai dirigé pleinement la thèse d'Ombeline Mayol. La collaboration établie avec Andreas Bommarius sur le projet AmDH me permet de codiriger une partie des travaux de thèse de Adam Caparco accueilli 4 mois dans notre laboratoire. J'assure depuis octobre dernier la supervision de la thèse de Laurine Ducrot et je dirigerai les travaux du post-doctorant Eddy Elisée nouvellement recruté dans le cadre du projet ANR JCJC obtenu, en collaboration étroite avec David Vallenet du LABGEM. Je m'investis également dans l'analyse des résultats et les décisions d'orientations du projet de thèse de Chloé Lelièvre dirigé par Anne Zapparucha et dans les discussions relatives aux projets sur lesquels la technicienne du laboratoire Aurélie Jouenne est en support. Les étudiants/thésards encadrés sont listés en table 2.

Table 2. Etudiants encadrés pendant mes activités de recherche au Genoscope

Doctorants			
	Franck Bordier	2010 - 2013	Université Evry Val d'Essonne
	Ombeline Mayol	2015 - 2018	Université Paris Saclay
	Chloé Lelièvre	2017 - 2020	Université Paris Saclay
	Adam Caparco	2018 - 2019	Georgia Tech Institute, USA
	Laurine Ducrot	2019 - 2022	Université Paris Saclay
Stagiaires			
	Yanis Frikha	2009 - 2010	élève ingénieur en année de césure
	Shangwen Zhuang	2010	élève ingénieur en demi-année de césure
	Anthony Flourieusse	2011 - 2012	L3Pro en alternance
	Johanna Rohault	2013	élève ingénieur
	Charlène Rousseau	2015	L3Pro en alternance
	Stessy Nepert	2015	stage ingénieur et M2
	Sylvain David	2015	stage ingénieur
	Rudy Bayakimissa	2019	stage M2 cursus pharmacie

Toutes ces activités d'encadrement n'ont cependant été constructives scientifiquement que grâce à l'implication de nombreux collègues des équipes du CAB, LGBM, LAGE et LABGEM.

III.2. Activités d'enseignement

Je n'ai eu aucune activité d'enseignement au cours de ces années de recherche ou de formation.

III.3. Responsabilités collectives

Pendant les 12 années passées au Genoscope, j'ai volontairement pris part à plusieurs responsabilités collectives. Je suis notamment membre élue depuis 2013 du conseil de laboratoire de l'unité de recherche Génomique Métabolique. De surplus, j'ai mis en place en 2017, avec 2 autres collègues, des séminaires hebdomadaires destinés à l'ensemble du personnel Genoscope. Le format décidé pour ces présentations orales est issu de discussions avec l'ensemble du personnel volontaire afin de répondre à un besoin exprimé en conseil de laboratoire. Nous assurons la régularité de passage des orateurs en maximisant les séminaires de stagiaires de recherche. Le besoin est pour l'instant satisfait, les séminaires sous ce format se poursuivent.

D'un point de vue plus restreint au laboratoire, j'ai mis en place techniquement le laboratoire de chimie à mon arrivée en 2007-2008 sous la direction de Anne, en gérant entre autre l'achat de l'ensemble du matériel *via* des appels d'offre avec le support du service achat et financier du CEA. J'ai depuis géré l'achat de 3 équipements (2 UHPLC et 1 HPLC préparative) et participé à la réflexion de l'achat d'autres équipements (GC-MS, fluorimètre).

Enfin, je me suis investie dans le comité de pilotage de rénovation des locaux en 2017-2018.

IV. Conclusion

L'ensemble des activités de recherche menées depuis ma formation initiale m'ont été très bénéfiques d'un point de vue scientifique mais aussi relationnel. Les projets collaboratifs, l'interdisciplinarité partagée au quotidien et la participation active à la vie du laboratoire et de l'unité sont un tremplin permanent pour améliorer le travail en équipe et le partage des savoirs. Mon expertise acquise ces 10 dernières années en biocatalyse me permettra d'assurer la tenue des objectifs premiers des projets discutés. La vision plus globale des objectifs atteignables selon les démarches projet effectuées me permet d'avoir désormais une supervision plus à moyen/long terme incluant la nature des verrous scientifiques à lever et les perspectives collaboratives alors ouvertes. Ces projets élargiront mes compétences tout en me permettant de m'affirmer de plus en plus dans leur gestion réfléchie au niveau global de l'unité avec les autres participants aux projets.

Au-delà des projets présentés ici, des projets innovants aux frontières des thématiques de chimie bio-inspirée, chimie biomimétique et biocatalyse doivent pouvoir être mis en place. Notamment grâce à l'évolution de mes compétences en structures des sites actifs protéiques, des catalyseurs bio-inspirés pourraient être réfléchis, que ce soit pour la synthèse de synthons/molécules intermédiaires d'actifs, ou pour les réactions d'oxydoréductions liées à la conversion énergétique/thématique de la fermeture du cycle de carbone. Quant aux connections entre biocatalyse et chimie biomimétique, c'est notamment au niveau de l'élaboration des voies rétrosynthétiques pensées « biocatalyse » de molécules cibles (Turner et al. 2013) que les synthèses biomimétiques peuvent avoir une contribution bénéfique. Même si je n'ai pas approfondi ce point, des idées seront à développer dans les prochaines années, en s'appuyant sur les relations que je peux avoir dans ces deux domaines.

V. Références

- Abrahamson, M. J., et al. (2012). "Development of an amine dehydrogenase for synthesis of chiral amines." *Ang Chem Int Ed Engl* 51(16): 3969-3972.
- Abrahamson, M. J., J. W. Wong and A. S. Bommarius (2013). "The Evolution of an Amine Dehydrogenase Biocatalyst for the Asymmetric Production of Chiral Amines." *Adv Synth Catal* 355(9): 1780-1786.
- Afanasyev, O. I., E. Kuchuk, D. L. Usanov and D. Chusov (2019). "Reductive Amination in the Synthesis of Pharmaceuticals." *Chem Rev*
- Ahmad, T., et al. (2019). "Metagenomics in the Search for Industrial Enzymes." *Adv Enz Technol* 419-451.
- Al Mourabit, A. and P. Potier (2001). "Sponge's Molecular Diversity Through the Ambivalent Reactivity of 2-Aminoimidazole: A Universal Chemical Pathway to the Oroidin-Based Pyrrole-Imidazole Alkaloids and Their Palau'amine Congeners." *Eur J Org Chem* 2001(2): 237-243.
- Alberti, A., et al. (2017). "Viral to metazoan marine plankton nucleotide sequences from the Tara Oceans expedition." *Sci Data* 4: 170093.
- Aleku, G. A., et al. (2017). "A reductive aminase from *Aspergillus oryzae*." *Nat Chem* 9: 961-969.
- Alma'abadi, A. D., T. Gojbori and K. Mineta (2015). "Marine Metagenome as A Resource for Novel Enzymes." *Genomics Proteomics Bioinformatics* 13(5): 290-295.
- Andrade, P., R. Willoughby, S. A. Pomponi and R. G. Kerr (1999). "Biosynthetic studies of the alkaloid, stevensine, in a cell culture of the marine sponge *Teichaxinella morchella*." *Tetrahedron Letters* 40(26): 4775-4778
- Apweiler, R., et al. (2004). "UniProt: the Universal protein knowledgebase." *Nucleic Acids Res* 32.
- Aslanidis, C. and P. J. de Jong (1990). "Ligation-independent cloning of PCR products (LIC-PCR)." *Nucleic Acids Res* 18(20): 6069-6074.
- Baran, P. S., D. P. O'Malley and A. L. Zografos (2004). "Sceptrin as a Potential Biosynthetic Precursor to Complex Pyrrole-Imidazole Alkaloids: The Total Synthesis of Ageliferin." *Ang Chem Int Ed* 43(20): 2674-2677.
- Barker, J. A. and J. M. Thornton (2003). "An algorithm for constraint-based structural template matching: application to 3D templates with statistical analysis." *Bioinformatics* 19(13): 1644-1649.
- Bastard, K., et al. (2018). "Structural Studies based on two Lysine Dioxygenases with Distinct Regioselectivity Brings Insights Into Enzyme Specificity within the Clavamine Synthase-Like Family." *Sci Rep* 8(1): 16587.
- Bastard, K., et al. (2017). "Parallel evolution of non-homologous isofunctional enzymes in methionine biosynthesis." *Nat Chem Biol.* 13(8): 858-866
- Bastard, K., et al. (2014). "Revealing the hidden functional diversity of an enzyme family." *Nat Chem Biol* 10(1): 42-49.
- Baud, D., et al. (2017). "A metagenomics approach for new biocatalyst discovery: application to transaminases and the synthesis of allylic amines." *Green Chemistry* 19(4): 1134-1143.
- Baud, D., et al. (2017). "Biocatalytic Approaches towards the Synthesis of Chiral Amino Alcohols from Lysine: Cascade Reactions Combining α -Keto Acid Oxygenase Hydroxylation with Pyridoxal Phosphate- Dependent Decarboxylation." *Adv Synth Catal* 359(9): 1563-1569.
- Baud, D., et al. (2014). "Synthesis of Mono- and Dihydroxylated Amino Acids with New α -Ketoglutarate-Dependent Dioxygenases: Biocatalytic Oxidation of C-H Bonds." *ChemCatChem* 6(10): 3012-3017.
- Bellinzoni, M., et al. (2011). "3-Keto-5-aminoheptanoate cleavage enzyme: a common fold for an uncommon Claisen-type condensation." *J Biol Chem* 286(31): 27399-27405.
- Bordier, F., et al. (2014). "Large α -aminonitrilase activity screening of nitrilase superfamily members: Access to conversion and enantiospecificity by LC-MS." *Journal of Molecular Catalysis B: Enzymatic* 107: 79-88.
- Bornscheuer, U. T. (2018). "The fourth wave of biocatalysis is approaching." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376(2110).
- Bornscheuer, U. T., et al. (2012). "Engineering the third wave of biocatalysis." *Nature* 485(7397): 185-194.
- Braña, M. F. and A. Sánchez-Migallón (2006). "Anticancer drug discovery and pharmaceutical chemistry: a history." *Clinical and Translational Oncology* 8(10): 717-728.
- Braun, M.-G., et al. (2018). "Green Chemistry Articles of Interest to the Pharmaceutical Industry." *Organic Process Research & Development.* 22(12): 1699-1711
- Carradec, Q., et al. (2018). "A global ocean atlas of eukaryotic genes." *Nat Commun* 9(1): 373.
- Chen, H. (2018). "Engineered Imine Reductase and Methods for the Reductive Amination of Ketone and Amine Compounds". Codexis WO2013170050.
- Cirino, P. C. and F. H. Arnold (2002). "Protein engineering of oxygenases for biocatalysis." *Curr Opin Chem Biol* 6(2): 130-135.

- Constable, D. J., et al. (2007). "Key green chemistry research areas—a perspective from pharmaceutical manufacturers." *Green Chemistry* 9(5): 411-420.
- Cosgrove, S. C., et al. (2018). "Imine Reductases, Reductive Aminases, and Amine Oxidases for the Synthesis of Chiral Amines: Discovery, Characterization, and Synthetic Applications." *Methods Enzymol.* 608:131-149
- De Brabander, J. K. and M. Razzak (2011). "Lessons and revelations from biomimetic syntheses." *Nat Chem Biol*: 865-875.
- de Melo-Minardi, R. C., K. Bastard and F. Artiguenave (2010). "Identification of subfamily-specific sites based on active sites modeling and clustering." *Bioinformatics* 26(24): 3075-3082.
- Devine, P. N., et al. (2018). "Extending the application of biocatalysis to meet the challenges of drug development." *Nature Reviews Chemistry* 2(12): 409-421.
- Doring, V., et al. (2018). "Implementation of a Reductive Route of One-Carbon Assimilation in *Escherichia coli* through Directed Evolution." *ACS Synth Biol* 7(9): 2029-2036.
- Ermolenko, L., et al. (2014). "Concise synthesis of Didebromohamacanthin A and Demethylaplysinsopine: Addition of Ethylenediamine and Guanidine Derivatives to the Pyrrole-Amino Acid Diketopiperazines in Oxidative Conditions." *Organic Letters* 16(3): 872-875.
- Ferrer, M., et al. (2019). "Decoding the ocean's microbiological secrets for marine enzyme biodiscovery." *FEMS Microbiol Lett* 366(1).
- Fonknechten, N., et al. (2009). "A Conserved Gene Cluster Rules Anaerobic Oxidative Degradation of L-Ornithine." *J Bacteriol* 191(9): 3162-3167.
- Genta-Jouve, G., et al. (2011). "New insight into marine alkaloid metabolic pathways: revisiting oroidin biosynthesis." *Chembiochem* 12(15): 2298-2301.
- Gleizer et al. 2019. "Conversion of *Escherichia coli* to Generate All Biomass Carbon from CO₂". *Cell* 179(6): 1255-1263
- Grogan, G. (2017). "Synthesis of chiral amines using redox biocatalysis." *Curr Opin Chem Biol* 43: 15-22.
- Hayes, K. S. (2001). "Industrial processes for manufacturing amines." *Applied Catalysis General* 221 187-195.
- Hepworth, L. J., et al. (2017). "Enzyme Cascades in Whole Cells for the Synthesis of Chiral Cyclic Amines." *ACS Catalysis* 7(4): 2920-2925.
- Höhne, M. and U. T. Bornscheuer (2009). "Biocatalytic Routes to Optically Active Amines." *ChemCatChem* 1(1): 42-51.
- Houwman, J. A., T. Knaus, M. Costa and F. G. Mutti (2019). "Efficient synthesis of enantiopure amines from alcohols using resting *E. coli* cells and ammonia." *Green Chemistry* 21(14): 3846-3857.
- Itoh, N., C. Yachi and T. Kudome (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.
- Jeffries, J. W. E., et al. (2016). "Metagenome Mining: A Sequence Directed Strategy for the Retrieval of Enzymes for Biocatalysis." *ChemistrySelect* 1(10): 2217-2220.
- Kaiser, F., A. Eisold, S. Bittrich and D. Labudde (2016). "Fit3D: a web application for highly accurate screening of spatial residue patterns in protein structure data." *Bioinformatics* 32(5): 792-794.
- Kaplan, O., et al. (2011). "Genome mining for the discovery of new nitrilases in filamentous fungi." *Biotechnol Lett* 33(2): 309-312.
- Kaulmann, U., et al. (2007). "Substrate spectrum of ω-transaminase from *Chromobacterium violaceum* DSM30191 and its potential for biocatalysis." *Enzyme and Microbial Technology* 41(5): 628-637.
- Kitagawa, I., et al. (1983). "Marine Natural Products. XII. On the Chemical Constituents of the Okinawan Marine Sponge *Hymeniacidon aldis*." *Chem Pharm Bull* 31(7): 2321-2328
- Kreimeyer, A., et al. (2007). "Identification of the last unknown genes in the fermentation pathway of lysine." *Journal of Biological Chemistry* 282(10): 7191-7197.
- Le Guilloux, V., P. Schmidtke and P. Tuffery (2009). "Fpocket: an open source platform for ligand pocket detection." *BMC Bioinformatics* 10: 168.
- Le, J. M. and K. L. Bren (2019). "Engineered Enzymes and Bioinspired Catalysts for Energy Conversion." *ACS Energy Letters*: 2168-2180.
- Lees, J., et al. (2012). "Gene3D: a domain-based resource for comparative genomics, functional annotation and protein network analysis." *Nucleic Acids Res* 40(Database issue): D465-471.
- Leipold, L., et al. (2019). "The identification and use of robust transaminases from a domestic drain metagenome." *Green Chemistry* 21(1): 75-86.
- Lenz, O., L. Lauterbach and S. Frielingsdorf (2018). Chapter Five - O₂-tolerant [NiFe]-hydrogenases of *Ralstonia eutropha* H16: Physiology, molecular biology, purification, and biochemical analysis. *Methods in Enzymology*. F. Armstrong, Academic Press. **613**: 117-151.
- Lenz, O., et al. (2010). "H₂ conversion in the presence of O₂ as performed by the membrane-bound [NiFe]-hydrogenase of *Ralstonia eutropha*." *Chemphyschem* 11(6): 1107-1119.
- Li, J., et al. (2014). "An integrated catalog of reference genes in the human gut microbiome." *Nat Biotechnol* 32: 834.

- Lindel, T. (2017). "Chemistry and Biology of the Pyrrole-Imidazole Alkaloids." *Alkaloids Chem Biol* 77: 117-219.
- Lowe, J., A. A. Ingram and H. Groger (2018). "Enantioselective synthesis of amines via reductive amination with a dehydrogenase mutant from *Exigobacterium sibiricum*: Substrate scope, co-solvent tolerance and biocatalyst immobilization." *Bioorg Med Chem*. 26(7): 1387-1392
- Madhavan, A., et al. (2017). "Metagenome Analysis: a Powerful Tool for Enzyme Bioprospecting." *Appl Biochem Biotechnol* 183(2): 636-651.
- Mangas-Sanchez, J., et al. (2017). "Imine reductases (IREDs)." *Curr Opin Chem Biol* 37: 19-25.
- Marliere, P., et al. (2011). "Chemical evolution of a bacterium's genome." *Ang. Chem. Int. Ed. Engl* 50(31): 7109-7114.
- Mayol, O., et al. (2019). "A family of native amine dehydrogenases for the asymmetric reductive amination of ketones." *Nature Catalysis* 2: 324-333
- Mayol, O., et al. (2016). "Asymmetric reductive amination by a wild-type amine dehydrogenase from the thermophilic bacteria *Petrotoga mobilis*." *Catalysis Science & Technology* 6(20): 7421-7428.
- Nadzirin, N. and M. Firdaus-Raih (2012). "Proteins of unknown function in the Protein Data Bank (PDB): an inventory of true uncharacterized proteins and computational tools for their analysis." *Int J Mol Sci* 13(10): 12761-12772.
- Nay, B. and E. Poupon (2011). *Biomimetic Organic Synthesis*. Wiley
- Patil, M. D., G. Grogan, A. S. Bommarius and H. Yun (2018). "Oxidoreductase-Catalyzed Synthesis of Chiral Amines." *ACS Catalysis*.
- Paul, C. E. and F. Hollmann (2016). "A survey of synthetic nicotinamide cofactors in enzymatic processes." *Applied Microbiology and Biotechnology* 100(11): 4773-4778.
- Perret, A., et al. (2011). "A Novel Acyl-CoA Beta-Transaminase Characterized from a Metagenome." *PLoS One* 6(8): e22918.
- Punta, M., et al. (2012). "The Pfam protein families database." *Nucleic Acids Res* 40(Database issue): D290-301.
- Pushpanath, A., et al. (2017). "Understanding and Overcoming the Limitations of *Bacillus badius* and *Caldalkalibacillus thermarum* Amine Dehydrogenases for Biocatalytic Reductive Amination." *ACS Catalysis*: 3204-3209.
- Reetz, M. T. (2013). "Biocatalysis in organic chemistry and biotechnology: past, present, and future." *J Am Chem Soc* 135(34): 12480-12496.
- Reeve, H. A., L. Lauterbach, O. Lenz and K. A. Vincent (2015). "Enzyme-Modified Particles for Selective Biocatalytic Hydrogenation by Hydrogen-Driven NADH Recycling." *ChemCatChem* 7(21): 3480-3487.
- Rudroff, F. (2019). "Whole-cell based synthetic enzyme cascades—light and shadow of a promising technology." *Curr Opin Chem Biol* 49: 84-90.
- Schrittwieser, J. H., S. Velikogne and W. Kroutil (2015). "Biocatalytic Imine Reduction and Reductive Amination of Ketones." *Adv Synth Catal* 357(8): 1655-1685.
- Sharma, M., et al. (2018). "A Mechanism for Reductive Amination Catalyzed by Fungal Reductive Aminases (RedAms)." *ACS Catalysis* 8(12): 11534-11541.
- Sheldon, R. A. and J. M. Woodley (2018). "Role of Biocatalysis in Sustainable Chemistry." *Chem Rev* 118(2): 801-838.
- Sigrist, C. J., et al. (2013). "New and continuing developments at PROSITE." *Nucleic Acids Res* 41(Database issue): D344-347.
- Soding, J. (2005). "Protein homology detection by HMM-HMM comparison." *Bioinformatics* 21(7): 951-960.
- Steinegger, M., M. Mirdita and J. Soding (2019). "Protein-level assembly increases protein sequence recovery from metagenomic samples manyfold." *Nat Methods* 16(7): 603-606.
- Steinegger, M. and J. Soding (2017). "MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets." *Nat Biotechnol* 35(11): 1026-1028.
- Steinegger, M. and J. Soding (2018). "Clustering huge protein sequence sets in linear time." *Nat Commun* 9(1): 2542.
- Steinkellner, G., et al. (2014). "Identification of promiscuous ene-reductase activity by mining structural databases using active site constellations." *Nat Commun* 5: 4150.
- Stockigt, J., A. P. Antonchick, F. Wu and H. Waldmann (2011). "The Pictet-Spengler reaction in nature and in organic chemistry." *Ang Chem Int Ed Engl* 50(37): 8538-8564.
- Stout, E. P., et al. (2012). "De novo synthesis of benzozseprin C and nagelamide H from 7-15N-oroidin: implications for pyrrole-aminoimidazole alkaloid biosynthesis." *J Nat Prod* 75(4): 527-530.
- Sunagawa, S., L. P. Coelho and S. Chaffron (2015). "Structure and function of the global ocean microbiome." *Science* 348(6237): 1261359.
- Thomas, M., et al. (2019). "De novo structure determination of 3-((3-aminopropyl)amino)-4-hydroxybenzoic acid, a novel and abundant metabolite in *Acinetobacter baylyi* ADP1." *Metabolomics* 15(3): 45.
- Tian, H., et al. (2011). "Pyrrole-Assisted and Easy Oxidation of Cyclic α -Amino Acid- Derived Diketopiperazines under Mild Conditions." *Adv Synth Catal* 353(9): 1525-1533.

- Travert, N. and A. Al-Mourabit (2004). "A Likely Biogenetic Gateway Linking 2-Aminoimidazolinone Metabolites of Sponges to Proline: Spontaneous Oxidative Conversion of the Pyrrole-Proline-Guanidine Pseudo-peptide to Dispacamide A." *Journal of American Chemical Society* 126: 10252-10253.
- Tseliou, V., et al. (2019). "Generation of amine dehydrogenases with increased catalytic performance and substrate scope from epsilon-deaminating L-Lysine dehydrogenase." *Nat Commun* 10(1): 3717.
- Tufvesson, P., J. Lima-Ramos, M. Nordblad and J. M. Woodley (2011). "Guidelines and Cost Analysis for Catalyst Production in Biocatalytic Processes." *Organic Process Research & Development* 15: 266-274.
- Turner, N. J. and R. Kumar (2018). "Editorial overview: Biocatalysis and biotransformation: The golden age of biocatalysis." *Curr Opin Chem Biol* 43: A1-A3.
- Turner, N. J. and E. O'Reilly (2013). "Biocatalytic retrosynthesis." *Nat Chem Biol* 9(5): 285-288.
- Vergne-Vaxelaire, C., et al. (2013). "Nitrilase Activity Screening on Structurally Diverse Substrates: Providing Biocatalytic Tools for Organic Synthesis." *Adv Synth Catal* 355(9): 1763-1779.
- Vergne-Vaxelaire, C., et al. (2018). "Characterization of a thermotolerant ROK-type mannofructokinase from *Streptococcus mitis*: application to the synthesis of phosphorylated sugars." *Applied Microbiology and Biotechnology* 102(13):5569-5583.
- Vergne-Vaxelaire, C., A. Zaparucha and V. de Berardinis (2018). Method for preparing amines from aldehydes and ketones by biocatalysis WO2019008110.
- Vergne, C., et al. (2006). "Verpacamides A–D, a Sequence of C11N5 Diketopiperazines Relating Cyclo(Pro-Pro) to Cyclo(Pro-Arg), from the Marine Sponge *Axinella vaceleti*: Possible Biogenetic Precursors of Pyrrole-2-aminoimidazole Alkaloids." *Organic Letters* 8(11): 2421-2424.
- Vilaine, J. P., C. Thollon, N. Villeneuve and J. L. Peglion (2003). "Procoralan, a new selective If current inhibitor." *European Heart Journal Supplements* 5: G26—G35.
- Voshol, G. P., E. Vijgenboom and P. J. Punt (2017). "The discovery of novel LPMO families with a new Hidden Markov model." *BMC Research Notes* 10(1).
- Wang, S. (2016). Marine strain and its preparation method of chiral amines amine dehydrogenase catalyzes. CN105567756
- Wang, S. and B. Fang (2013). Method for preparing chiral amine through asymmetric reduction under catalysis of marine strain OTI CN103224963
- Wendisch, V. F., M. Mindt and F. Pérez-García (2018). "Biotechnological production of mono- and diamines using bacteria: recent progress, applications, and perspectives." *Applied Microbiology and Biotechnology* 102(8):3583-3594.
- Wetzel, D., et al. (2016). "Asymmetric Reductive Amination of Ketones Catalyzed by Imine Reductases." *ChemCatChem* 8(12): 2023-2026.
- Yu, H. L., et al. (2018). "Bioamination of alkane with ammonium by an artificially designed multienzyme cascade." *Metab Eng* 47: 184-189.
- Yu, X., et al. (2019). "Synthesis of Formate from CO₂ Gas Catalyzed by an O₂-Tolerant NAD-Dependent Formate Dehydrogenase and Glucose Dehydrogenase." *Biochemistry* 58(14): 1861-1868.
- Zaparucha, A., V. de Berardinis and C. Vaxelaire-Vergne (2018). "Chapter 1. Genome Mining for Enzyme Discovery." 1-27.

ANNEXE 1

Détails de l'approche génomique utilisée pour la recherche de nouvelles enzymes dans la biodiversité procaryote

La stratégie génomique pour la recherche de nouvelles enzymes au sein de la biodiversité a été mise en place par la plateforme de clonage et criblage d'activités enzymatiques. Elle est basée sur la recherche de séquences protéiques homologues à des enzymes références décrites expérimentalement comme possédant l'activité cible (ou la plus proche possible) et/ou à la conservation de signatures/domaines protéiques avec ces mêmes séquences références (Figure 31).

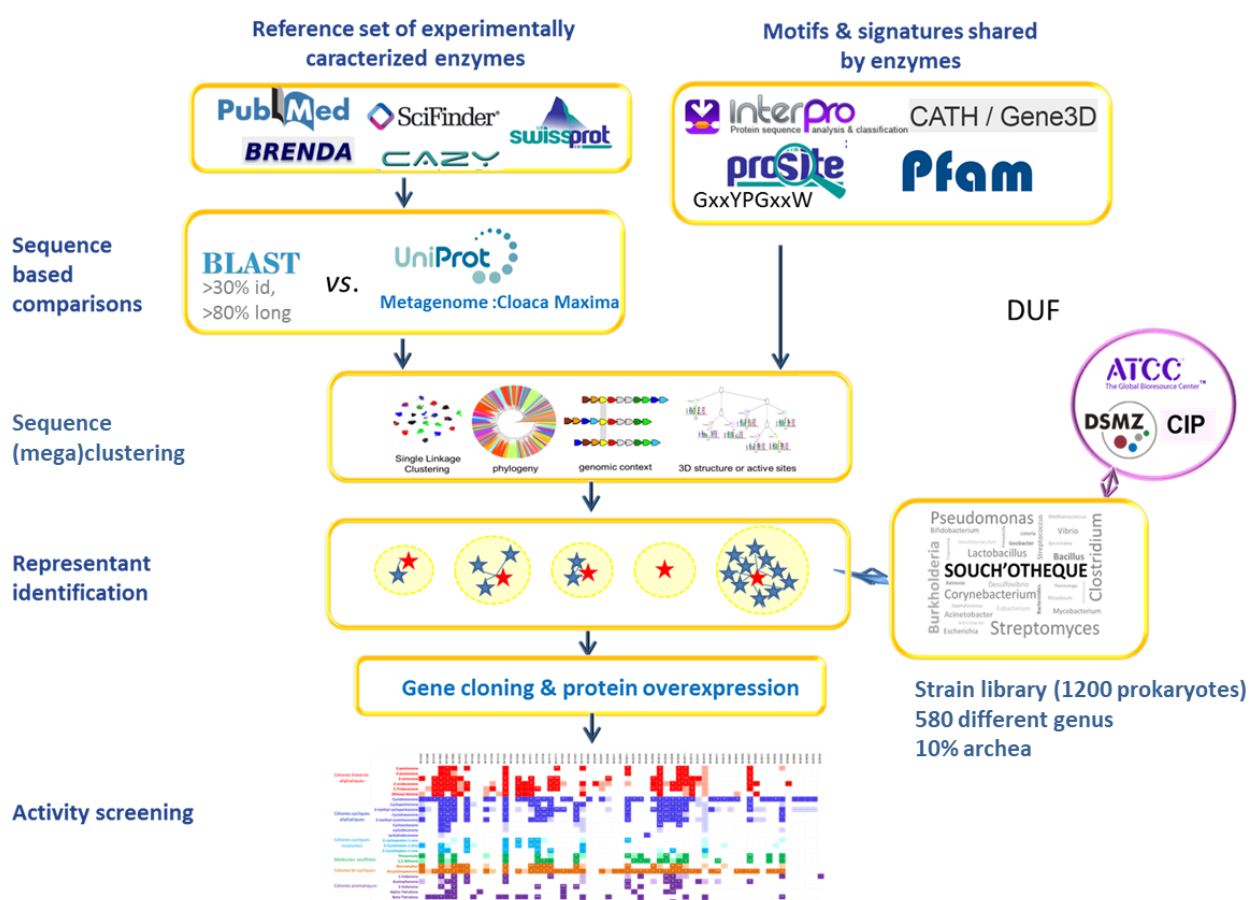


Figure 31. Principe utilisé par l'équipe CAB pour la recherche de nouvelles enzymes au sein de la biodiversité.

La démarche est la suivante :

- *inventaire des enzymes références* : les enzymes déjà décrites comme catalysant la réaction ciblée sont répertoriées par recherche bibliographique d'articles décrivant cette activité avec preuve expérimentale (PubMed, Google Scholar...). Les bases de données enzymatiques (BRENDA, SwissProt principalement) sont également utilisées pour cette recherche. L'exhaustivité de cet inventaire est

important pour décrire au mieux la diversité déjà prouvée et ainsi fournir le plus de diversité au niveau des résultats de candidats dans l'étape suivante.

- *identification des candidats au sein de la biodiversité procaryote* : le set de référence est utilisé pour trouver des homologues au sein des bases UniprotKB et du métagénome Cloaca maxima (métagénome du bassin anaérobie de la station d'épuration d'Evry étudié par le Laboratoire des Métagénomés Procaryotes de l'Unité). Deux types de recherches peuvent être effectuées : soit par homologie de séquences grâce au programme d'alignement BLAST (en général 80% d'identité sur 30% de la longueur), soit par partage de domaines ou signatures prédites (InterPro, Pfam).

- *regroupement des séquences candidates* : cette étape a pour but de minimiser le nombre de protéines qui seront expérimentalement testées pour l'activité cible. Le set minimal généré se doit d'être représentatif de la diversité naturelle de la famille explorée. Pour cela, les séquences rapatriées précédemment sont regroupées par regroupement hiérarchique basé en général sur la séquence protéique (80% d'identité, mais ce paramètre est ajustable selon la grandeur de la famille étudiée). Des regroupements basés sur la distribution 3D en acides aminés des sites actifs des enzymes étudiées, sur la phylogénie ou sur le contexte génomique des gènes correspondants peuvent aussi être effectués.

- *sélection des candidats à tester* : en partant de l'hypothèse que chaque groupe généré à l'étape précédente est isofonctionnel, un représentant de chaque groupe est sélectionné. Cette sélection se fait sur plusieurs critères, avec en priorité la disponibilité de l'ADN du gène correspondant dans la collection d'ADN génomique (« souchothèque »), qui sera utilisé pour cloner le gène dans *E. coli*. En cas d'absence pour l'ensemble d'un groupe, le programme propose des souches commerciales au moyen de leur identifiant (DSM, ATCC). Les gènes sélectionnés sont à chaque fois séquencés pour éviter toute erreur, en particulier lors de l'utilisation de souches dépourvues d'identification précise. Dans ce cas, l'hypothèse est faite que la souche utilisée est suffisamment proche de celle d'où provient le gène sélectionné dans les bases, pour que les amorces choisies sur le gène candidat puissent amplifier le gène.

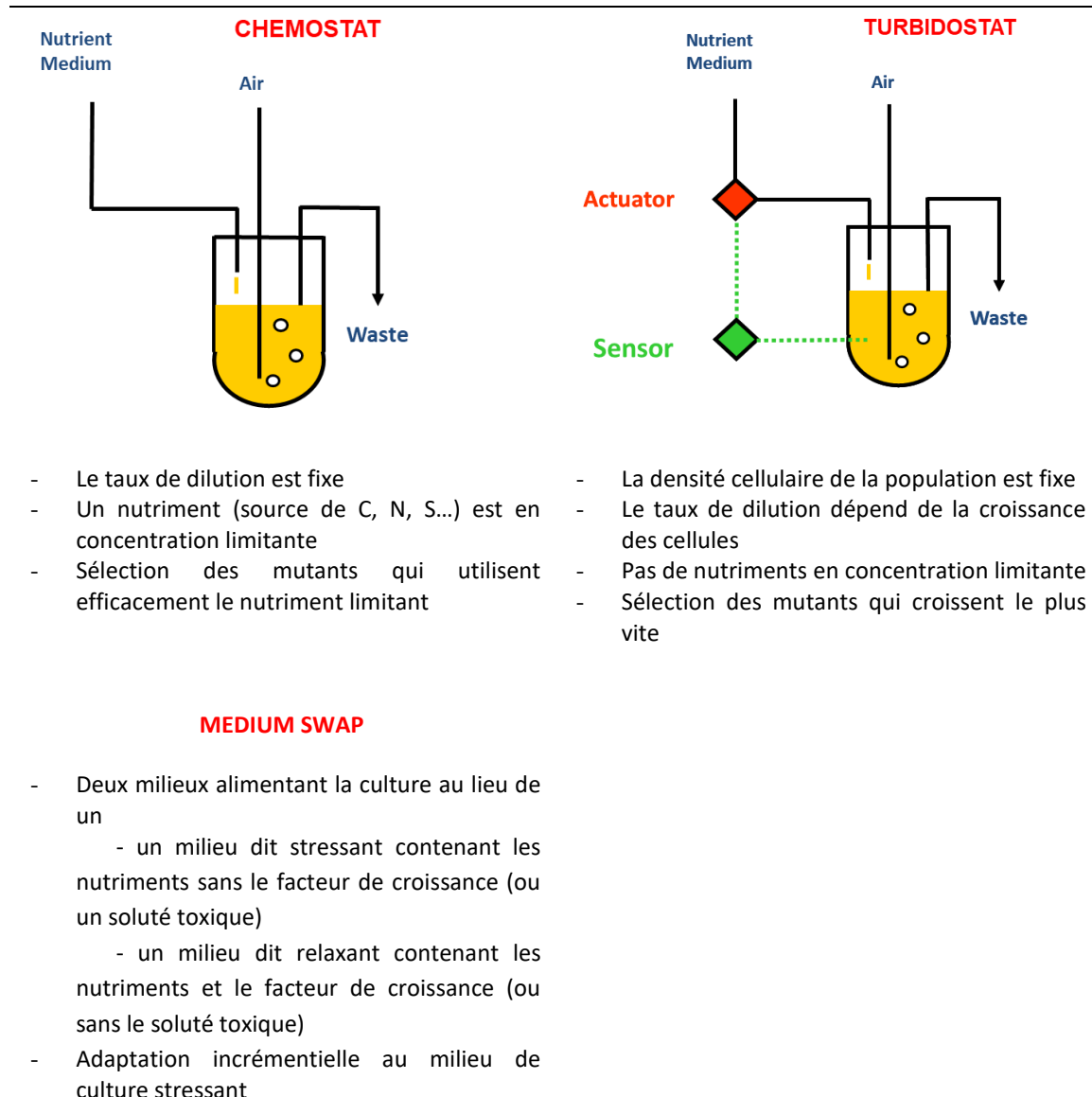
- *clonage en expression des protéines* : après choix automatique des amorces de PCR (et potentiel ajout d'étiquette de 6 histidines en N-terminal de la protéine juste avant la méthionine quand le criblage d'activité nécessite de travailler sur enzymes purifiées), le clonage est réalisé selon une méthode générique compatible pour une utilisation en plaque 96 puits (Ligand Independent Cloning) (Aslanidis et al. 1990), avec un vecteur d'expression de type pET22 modifié. Les protéines sont ensuite produites dans une souche d'*E. coli* optimisée, en format plaque 96 puits, et les extraits cellulaires préparés après sonication des plaques. Une étape de purification des enzymes en plaques 96 puits peut parfois être réalisée, notamment dans le cas de tests de criblages trop bruyants sur extrait cellulaire.

- *criblage des activités enzymatiques* : après mise au point du test sur des enzymes tests et adaptation en plaques 96 puits, le criblage des activités est effectué soit en plaque 96 puits, soit en plaque 384 puits après réarrangement, de préférence par des méthodes spectrophotométriques ou colorimétriques du fait de leur rapidité. Dans le cas où un tel test ne peut pas être utilisé, un criblage par LC-MS (pilote par Ekaterina Darii du LGBM) est effectué (MS : Qtrap 5500 ABSciex).

ANNEXE 2

Fonctionnement des GM3 (Automates permettant la sélection des mutations adaptées en culture continue)

Les différents régimes de culture continue :



Pour éviter la formation de biofilms, les automates de cultures GM3 disposent de deux chambres de culture par expérience. Un transfert de la culture, avec nettoyage de la chambre venant de contenir la culture par du NaOH concentré, est réalisé périodiquement pour permettre la contre-sélection des biofilms (Figure 32).

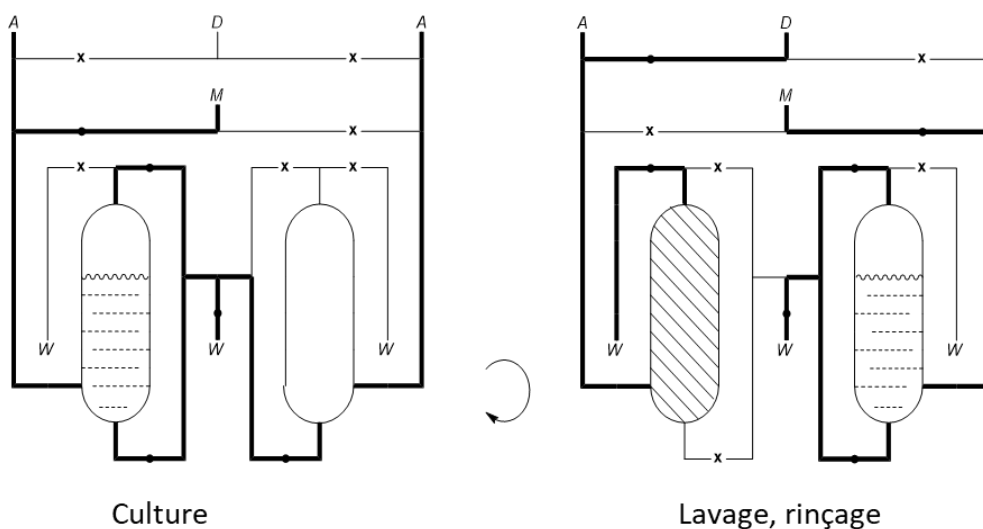


Figure 32. Schéma détaillant les chambres de cultures en phase de culture et lavage/rinçage.

Détails techniques :

- Volume de culture : 16 mL
- Pression 0.3 bar
- Nombre d'électrovannes : 60
- Plage de température : TA – 45°C
- Tubes en silicone
- Possibilité d'atmosphère avec ajout de max 20% de CO₂
- 1 dispositif permettant une culture en anaérobie (100% azote)

Limites :

- Le milieu doit être homogène et sans composés agressifs vis-à-vis du silicone
- Le temps de génération ne peut pas excéder 24h

Publication sur l'exploration préliminaire de métagénomiques pour l'identification d'AmDHs (soumise)

ARTICLE

Received 00th January 20xx,
Accepted 00th January 20xx
DOI: 10.1039/x0xx00000x

Metagenomic Mining for Amine Dehydrogenase Discovery

Adam A. Caparco,^a Eric Pelletier,^b Jean Louis Petit,^b Aurélie Jouenne,^b Bettina R. Bommarius,^a Véronique de Berardinis,^b Anne Zaparucha,^b Julie A. Champion,^a Andreas S. Bommarius,^{*a} Carine Vergne-Vaxelaire^{*b}

Amine dehydrogenases (AmDHs) catalyze the enzymatic reduction of ketones to amines, serving as a suitable biocatalytic route for amine synthesis. A limited number of experimentally validated native AmDHs (nat-AmDHs) have been reported recently, expanding the sequences with this function to complement the engineered set of enzymes. Since researchers can now probe into the vast diversity of enzymes within niche environments by a metagenomics approach, the previously untapped biocatalytic reservoirs of the ocean environment and human microbiome databases were screened for AmDH discovery. Among the hundreds of hits, a subset of 18 enzymes was selected for further characterization and were confirmed to display AmDH activity. Their abundance among the screened environments was also described. Additional analysis on six enzymes confirmed altered cofactor specificities and variation in substrate scopes, catalytic efficiencies, and active site residues compared to the reference nat-AmDHs previously described. Particularly, one of the enzymes from an eukaryotic organism demonstrated unprecedented nat-AmDH specific activity toward isobutyraldehyde and 1,2-cyclohexadione. The protein sequence diversity of validated AmDHs reached by this metagenomics mining strategy highlights the success of such an approach. Metagenomically mined proteins, including eukaryotic ones, stand to increase the reach of biocatalysis towards environmentally benign processes.

Introduction

Over the past ten years, scientific and technological advances have established biocatalysis as a practical and environmentally friendly alternative to some traditional catalysts by fulfilling key criteria of green chemistry.¹ The potential for biocatalysis as a robust alternative to traditional catalysis is still limited by the scope of known enzymatic reactions and protein stabilities. To overcome these limitations, novel biocatalysts must be discovered and characterized in the search for new properties.^{2,3,4} Equipped with a growing number of bioinformatic tools, exploring biodiversity among previously uncharacterized organisms is an attractive approach for enzyme discovery.^{5,6,7,8} Introduced for the first time in 1998, metagenomics is the method of studying the genetic content of an entire community from an environmental sample, circumventing the need to isolate and grow potentially unculturable micro-organisms.⁹ Analyzing the genetic material of a community allows for the identification of potential protein coding regions, which may house novel biocatalysts.⁸ Thus far, enzymes used in biocatalytic synthesis discovered through the

metagenomics approach are mainly limited to hydrolases,^{10,11,12} with some examples of transaminases,^{13,14,11} transketolases,⁸ nitrilases¹¹ and dehydrogenases.⁸

The search for robust biocatalytic solutions continues for many classes of compounds, including dehydrogenases for the synthesis of chiral amines.^{15,16} A direct synthesis of the corresponding primary amines from prochiral ketones and ammonia by reductive amination serves as an ideal route with high atom economy. This can be achieved enzymatically via amine dehydrogenases (AmDHs).^{17,18,19,20} In contrast to lipases or transaminases^{21,22,23} widely reported for the biocatalytic synthesis of chiral amines, NAD(P)H-dependent AmDHs have many advantages such as favorable equilibrium position enabling full conversion, facile coenzyme regeneration, and the use of the abundant and inexpensive ammonia as an amine source. Such advantages align very well with most of the principles of green chemistry.²⁴ Previously restricted to engineered amino acid dehydrogenases,^{17,25,26,27,28,29,30,31,32} the family of AmDHs has grown in recent years due to the discovery of genes for native AmDHs (nat-AmDHs).^{33,34} Similarly, reductive aminases (RedAms) can catalyze this reaction, but with a preference for the synthesis of secondary amines due to low activity with ammonia as an amine donor.^{35,36,37,38}

For AmDH-based biocatalysis to become a realistic alternative for amine production in industrial green biotechnology, several criteria are required. Most notably, a large substrate scope, stability with respect to temperature and pH, tolerance for high substrate concentrations, complementary stereoselectivity

^a School of Chemical & Biomolecular Engineering, Georgia Institute of Technology, Atlanta, GA, USA.

^b Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, 91057 Evry, France

† Footnotes relating to the title and/or authors should appear here.

Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

between enzymes, and high turnover rates can improve the greenness and efficiency of AmDH-catalyzed processes. Searching for amine dehydrogenases among metagenomic data can greatly increase the chances of finding isofunctional enzymes with unique features,^{39,40} potentially resulting in decreased energy consumption and cost savings for the chemical industry.¹¹ Other native AmDHs were sought among three metagenomic databases sourced from the human microbiome and marine samples to increase the diversity of biocatalysts in the newly emerging catalog of AmDHs. The European Metagenomics of the Human Intestinal Tract (MetaHit) project and Human Microbiome Project (HMP) in the United States sought to connect the human microbiome with health indicators and have resulted in an extensive reference catalog of 9.9 million non-redundant microbial genes (Integrated Gene Catalog, IGC).⁴¹ The international Tara Oceans pan-oceanic expeditions have collected samples to quantify planktonic communities by covering the whole spectrum, from virus to larvae, coupled with various biogeochemical measurements to lay the foundations for an exhaustive description of marine ecosystems.⁴² This holistic strategy represents one of the largest efforts in environmental genomics sampling which can now also be valued as a huge reservoir of data for biocatalysis. Combined with other projects of marine environment, these gene sets associated with the sequencing led to the Ocean Microbial Reference Gene Catalog (OM-RGC) which groups more than 40 million prokaryotic genes.⁴³ The second database established from this effort (Marine Atlas of Tara Oceans Unigenes, MATOUv1) comprises 116 million unigenes obtained from metatranscriptomic sequences of eukaryotes ranging in sizes from 0.8 to 2000 μm .⁴⁴ The goal of the presented work was to access unexplored members of the recently identified natural AmDH family,³⁴ and their abundance in the biosphere by an unprecedented metagenomics screen which includes eukaryotic members. This approach aims to rapidly enrich the sequence space and known properties of AmDHs for wider green chemical applications.⁴⁵ After identification and selection of candidates among the three metagenomics databases IGC, OM-RGC and MATOUv1, sixteen potential AmDHs homologs were successfully cloned and overexpressed in *Escherichia coli* (*E. coli*). Their activity, sequence similarity and modelled 3D-structures were compared to the template nat-AmDHs. Selected AmDHs were then further characterized and their suitability for biocatalytic synthesis was investigated.

Results and Discussion

Initial screening of 18 AmDHs identified from marine environment and human microbiome

Environmental sequences that matched experimentally validated nat-AmDHs (*MsmeAmDH*, *MvacAmDH*, *MycoAmDH*, *MicroAmDH*, *ApauAmDH* and *CfusAmDH*)³⁴ via a Hidden Markov Model (HMM) genome mining approach, were retrieved from the OM-RGC (12 proteins), MATOUv1 (9 proteins) and IGC (251

proteins) databases (see Materials and Methods for details). Some representatives were selected for experimental assays. The selection criteria, detailed in ESI Fig. S1-S3, were established with the objectives of i) discovering nat-AmDHs among the metagenomes including eukaryote microorganisms, and of ii) screening distant homologs with higher potential for alternative features. The resulting 18 candidates spanned all three databases, with 11 stemming from prokaryotes in the human microbiome, as well as five prokaryotes and two eukaryotes from marine environment (Table 1). Eight of the candidates have greater than 98% amino acid identity with unreviewed proteins from UniprotKB database, whereas four of them have no reported close homologs (sequence identity < 60%).

In Figure 1, a sequence identity matrix demonstrates the diversity within the selected candidate set and its similarity with the characterized AmDH reference set, as illustrated by the phylogenetic tree (see ESI Fig. S4). The protein sequence identities of the candidates compared to the reference set are typically in the 30-50% range, with a few exceptions in the 20-30% and 60-70% range. Notably, four of the known nat-AmDH are very similar and form a group including only one new candidate (*PortiAmDH*, 64-70% of identity) distinct to the other proteins. Proteins from the OM-RGC database feature very low similarity, whereas the ones from IGC database share identities between 46 and 82%. The eukaryotic genes are grouped together, with sequence similarities < 36% compared to all the other enzymes. It is interesting to note that the amino acid sequence of *TtherAmDH* is quite close to AmDH4, one of the first reported nat-AmDHs, though it was not used in the reference set.³³ The genomic context of the corresponding gene of its identical homolog in UniprotKB database (Uniprot entry M8DJ2) suggests *TtherAmDH* is part of the ornithine synthesis pathway, described to be a diamino-pentanoate dehydrogenase like AmDH4. This suggests 2,4-diamino-pentanoate may also be the native substrate of this enzyme^{46,33} (see ESI Fig. S5). No other reliable information could be extracted from the analysis of the genomic context and/or annotation of the other seven candidates with close homologs in UniprotKB database (UniprotKB entries A0A2G1Y1R1, A8AV93, A7B1V2, D3AE01, A0A120KMV0, R7CEI0 and B0P8I0). Many of them lack a predicted function or are annotated as putative dihydrodipicolinate reductases, as were the reference enzymes.³⁴

After outsourced synthesis, the genes were successfully cloned into a BL21 expression line of *E. coli* and the clones were sequenced (see ESI, Table S1-S2). The corresponding proteins were overexpressed (see Materials and Methods for details). Despite the highly diverse prokaryotic and eukaryotic origins of the proteins, their level of expression in cell-free extract was satisfactory (except for OM-RGCAmDH1 and MATOUAmDH1), enabling investigation of the potential AmDHs in cell-free lysates.

An activity screen toward four known substrates from the reference set, cyclohexanone (1a), isobutyraldehyde (2a), 2-methylcyclohexanone (3a) and 2-pentanone (4a), was

conducted with cell-free extracts and cofactor regeneration system using UHPLC-UV/MS monitoring (Table 2, see Materials and Methods for details). All tested proteins were active with at least one of the substrates in the test set. Apart from MATOUAmDH2, which showed the highest conversion with each of the substrates, higher conversions of **2a** to isobutylamine (**2b**) were found than of the cyclic ketones **1a** and **3a** or the linear ketone **4a**.

The stereoselectivity of the enzymes was evaluated towards **3a** and **4a**, which gave in general lower conversion rates than the less hindered **1a** and **2a**. Like the reference set, all the enzymes preferentially formed (*S*)-amine products **3b** and **4b**. For the enzymes with > 5% conversions towards **3a**, the ratio of each diastereoisomer of **3b** could be determined and assigned based on previous reported results.³⁴ The *trans* products (1*S*,2*S*) and (1*R*,2*R*) were detectable only for IGCAmDH4, *Tther*AmDH, and MATOUAmDH2 at low ratios (0.7 – 5.2%), indicating a high preference for the formation of the *cis* product (1*S*, 2*R*)-**3b** (see ESI Fig. S6 for examples of UHPLC-UV chromatograms).

Comparison of Active Site Residues

In addition to the preliminary activity test, the enzyme structures were modeled by homology using *Cfus*AmDH (PDB: 6IAU) and *Msme*AmDH (PDB: 6IAQ) crystal structures as templates (see Materials and Methods). After superimposing these models to the reference structures, the residues identified in the Active Site Modeling and Clustering (ASMC) previously reported for the AmDH family were compared. The resulting residues at spatial positions P1 to P20 compose the active site pocket, and a mechanistic and structural role have been assigned to several of them.³⁴ P3 is the catalytic residue, while P9 binds the ketone, and P11 is hypothesized to be essential for closing the active site during catalysis. Alignment of the screened AmDHs according to their ASMC structure similarity for P1 – P20 is presented in Figure 2. An important observation is the conservation of the postulated catalytic glutamate at position P3, apart from OM-RGCAmDH2 bearing an aspartate instead of a glutamate at this position. With an arginine at P8 (residue hypothesized to form an electrostatic interaction with the carboxylate of 2,4-DAP and 4-AP),^{33, 34} *Tther*AmDH matches closest with AmDH4, as suggested by the sequence similarity and genomic context analysis (Fig. 1 and ESI Fig. S5). Interestingly, the 3D alignment is quite poor with AmDH4 and better with the other reference, *Cfus*AmDH, despite many differences in nature of the residues (see ESI Fig. S7). Similar to the sequence comparison (Fig. 1), *Porti*AmDH falls into the group of already described enzymes *Mvca*-, *Mvco*- and *Micro*AmDH, bearing a nearly identical active site to *Msme*AmDH. Most of the other active sites are very similar to *Cfus*AmDH, with few differences arising at positions P5, P7 and P15. These positions define the shape of the active site. The eukaryotic proteins MATOUAmDH1 and MATOUAmDH2 displayed the same noticeable differences from the others, particularly at position P5 (L/I instead of Q for the reference enzymes), P6 (A instead of F), P7 (C instead of W) and P14 (K instead of P), expanding the set of known active sites residues

with AmDH activity. Finally, there are two standalone candidates with no clear grouping, OM-RGCAmDH1 and OM-RGCAmDH2. OM-RGCAmDH1 gave low conversion rates toward the tested substrates, corroborated by a lack of consensus in key residues P9, P11 and P16, the latter being probably implicated in position of the catalytic residue (P3). OM-RGCAmDH2 has a positive rather than neutral residue at P15 and a different residue at position P20, hypothesized to interact with the nicotinamide ring of the cofactor. Despite these differences, its cell-free extract gave satisfactory conversions of **1a** and **2a**, which shows the possibility of variation of these residues with maintained activity. This analysis serves to highlight that the reaction may occur with slight differences in the binding residues for substrates/cofactors, reinforcing the interest in looking for more diverse enzymes to span the space of potentially beneficial substitutions.

Selection of Candidates for Further Characterization

Moving forward from these preliminary studies, a subset of enzymes was selected for further characterization. The main objectives were to cover AmDHs from all three selected metagenomes while providing insight into the potential diversity of substrate spectrum and stability. The criteria for picking the subset were ease of expression, utilization of all three databases, storage stability, conversion, stereoselectivity, and active site variability. The seven resulting enzymes (IGCAmDH1, IGCAmDH5, *Sgar*AmDH, *Chat*AmDH, *Tther*AmDH, and *Aca*AmDH from prokaryote organisms, and MATOUAmDH2 from an eukaryote organism) were overexpressed in *E. coli* and purified (see Materials and Methods). For *Tther*AmDH, stability issues prevented further characterization, but some limited data have been generated with the cell-free extract.

Conversion with cofactor regeneration for selected AmDHs

To gain insight into readiness for synthesis, the conversions of isobutyraldehyde (**2a**) into isobutylamine (**2b**) by the purified AmDHs (with the exception of *Tther*AmDH) were assessed at several temperatures with glucose/glucose dehydrogenase as a cofactor regenerating system (Fig. 3). At the same enzyme concentration (0.1 mg mL⁻¹), conversions higher than 70% were obtained with all the tested AmDHs for at least one temperature, excluding for MATOUAmDH2 (23 – 67 %). This is likely due to purified enzyme instability and not to competitive product inhibition since the specific activity was only slightly affected by 10 mM of **1b** (85% of the activity compared to the activity in absence of **1b**; see ESI Fig. S8). For enzymes *Chat*AmDH, IGCAmDH1, IGCAmDH5, and *Aca*AmDH, there is almost no effect with temperature (maximum 10% of loss of conversion between 30 and 50°C), indicating these enzymes are quite stable even at 50°C for 24-hour conversions. This feature can be beneficial in synthesis reactions with low substrate and product solubilities, which can be increased by elevating the temperature in the reaction media. For *Sgar*AmDH and MATOUAmDH2, there is a significant drop in 24-hour conversions with temperature, from 83-85% at 30-40°C to 32% at 50°C, and 59-67% at 30-40°C to 23% at 50°C, respectively. The

high conversions with IGCAmDH1 and *ChatAmDH* prompted the examination of conversions at higher substrate concentrations. Despite lower conversions (52.9% and 28.8% at 50 mM and 200 mM, respectively), this preliminary study with IGCAmDH1 showed a relevant turnover number can be obtained with this enzyme (45 089 at 200 mM **2a** with 0.1 mg mL⁻¹ of purified IGCAmDH1, 30°C, pH 8.2) without any optimization (see Materials and Methods for details).

Stability of purified AmDHs

To gain insight on the structural stability of the enzymes, the melting temperature for each purified AmDH was determined (Table 3). Most of the enzymes seem to have one transition around 50–56°C, consistent with the activity temperature correlations showing maximum activity at 50–60°C (Materials and Methods, ESI Fig. S9). Operating near the equilibrium melting temperatures could explain the decreased conversions at 50°C during the studies described above. Moreover, some form of transition below 35°C were observed for *SaarAmDH*, MATOUAmDH2, and *TtherAmDH* through a noticeable difference in the rate of change of fluorescence ratio (ESI Fig. S10). Particularly for MATOUAmDH2 and *TtherAmDH*, this transition could be indicative of the storage stability and conversion inconsistencies compared to conversion with the freshly prepared crude cell lysate. As expected, due to its thermophilic origin, *TtherAmDH* has another melting temperature calculated at a much higher value than the rest of the set. It should be noted that these enzymes do not show a clear two-state transition in folded states typically seen in a well-developed, stable enzyme. An inconclusive oligomeric status for each of these enzymes, as measured by gel permeation chromatography (GPC), suggests that these enzymes are not formed in their native oligomeric state in *E. coli*. As these genes come from a group of diverse microorganisms, it is not surprising that the expression conditions preferred by *E. coli* may not be ideal. Therefore, both optimization of expression conditions and protein engineering to stabilize a predominant oligomeric state is envisioned for the enzymes of highest biocatalytic potential, such as MATOUAmDH2, with the overall goal of activity retention and improved biocatalyst stability.

Substrate scope of selected AmDHs

A substrate screen was conducted on the selected set of purified enzymes at 50°C with both NADH and NADPH (Table 4). All the selected enzymes have a strong preference for NADH over NADPH, except MATOUAmDH2 which showed a 71:1 NADPH:NADH ratio in activity towards **2a**. The highest specific activities were obtained with aliphatic aldehydes instead of ketones for all these AmDHs. Isobutyraldehyde (**2a**) was the top tested substrate, with specific activities above 3.2 U mg⁻¹ for *SaarAmDH*, and up to 11.1 U mg⁻¹ for MATOUAmDH2. Notably, activity was also detected with pentaldehyde and benzaldehyde, unlike for the reference enzymes.³⁴ The dialdehyde glutaraldehyde was not able to be tested, due to crosslinking of the enzymes.⁴⁷ MATOUAmDH2 was unique in its

specific activity levels for 1,2-cyclohexadione (878.8 mU min⁻¹), methyl isobutylketone (MIBK) (44.4 mU min⁻¹), and norcamphor (36.8 mU min⁻¹). This broader range of substrates for MATOUAmDH2 offers a promising protein engineering template to enhance substrate range and stability. γ -ketones, bulkier ketones such as ethylacetoacetate methyl ester, and aromatic ones such as p-tolyl phenylacetone, had very low or no detectable activity for any of the purified enzymes. The substrate scope of these purified enzymes was similar to the reference set, however some differences are clear. Particularly, their preference for NADH in lieu of NADPH gives more flexibility for cofactor recycling enzymes. More importantly, substrates such as 1,2-cyclohexadione and MIBK, previously not reported to be transformed by nat-AmDHs, are possible using MATOUAmDH2 (see ESI Fig. S11 for details). The similarity of *TtherAmDH* to AmDH4 and the presence of the DAPDH pathway in a genomic context analysis motivated an expanded substrate screen for this enzyme. *TtherAmDH* showed activity toward 2-amino-4-oxopentanoate (0.78 U mg⁻¹ and 1.9 U mg⁻¹ at 30°C and 50°C, respectively), the native substrate of the DAPDH.^{46,33} It also retained activity on the same order of magnitude toward the other substrates **1a** and **2a** (0.4–0.9 U mg⁻¹) not accepted by DAPDHs AmDH4 and AmDH5 (see ESI Table S3 for detailed data). The mixed substrate spectrum and sequence similarity to AmDH4 suggest *TtherAmDH* may represent an enzyme between AmDH4 and the reference nat-AmDHs. Ammonia remains the preferred amine source for all tested enzymes. In comparison, methylamine results in 0.8% (*SaarAmDH*) to 40% (*ChatAmDH*) of the original activity with ammonia at 100 equivalents of amine (see Materials and Methods and ESI Table S4 for details). This behavior is similar to that previously reported for nat-AmDHs.³⁴ In the case of MATOUAmDH2, significant activity was found also with ethylamine (6% and 16% of ammonia and methylamine activity, respectively), which again positions this enzyme as a very promising target for development.

Kinetic parameter determination

Kinetic parameters were determined on **1a** and **2a** for the top performing enzymes (Table 5, see Materials and Methods for details). This set of six purified enzymes has a much higher catalytic efficiency towards **2a** than **1a** (19 to 90 times), similar to *CfusAmDH*. The K_M values obtained for ammonia were similar to those previously reported for AmDHs, with potential inhibition in case of MATOUAmDH2 (see ESI Table S5). Excluding MATOUAmDH2, the tested enzymes displayed higher K_M values and lower catalytic efficiencies towards both substrates **1a** and **2a** than previously reported nat-AmDHs (24.5–109.9 mM and 6.8–41.0 mM compared to 0.8 and 0.6 mM for *CfusAmDH* for **2a** and **1a**, respectively). Unlike *CfusAmDH*, no substrate inhibition occurred at concentrations 150–300 mM which is a benefit for substrate loading on preparative scale. MATOUAmDH2 exhibits an unprecedented catalytic efficiency towards **2a** (16,100 s⁻¹M⁻¹), which again positions this enzyme as a top performer.

Abundance of selected AmDHs

As illustrated in Figures 4a and 4b, the eukaryotic MATOUAmDH2 is expressed in various ocean environments but is mostly located in the Pacific Ocean and off South Africa in the samplings of larger microorganisms (5 – 200 µm) near the surface. The corresponding organism is a diatom (*Lithodesmiales/Lithodesmaceae*). The gene coding for *SaorAmDH* is not widely abundant among the oceans and is mainly present in the Mediterranean Sea in the DCM (deep chlorophyll maximum) layer fraction of 0.2 – 1.6 µm. The gene coding for *TtherAmDH* is not present in Tara sampling but is present in OM-RGC data as a marine environment gene. Concerning genes from human microbiome IGC database, the genes corresponding to IGCAmDH1 and *AcoAmDH* have been detected in many patients (44 and 29 % of patients, respectively) while the genes coding for *ChatAmDH* and especially for IGCAmDH5 are much less prevalent, with some variability by country (Fig. 4c). These data demonstrate that the targeted AmDH activity is widely present in human microbiomes and expressed heterogeneously in marine environments. The lack of genomic context conservation among these new prokaryote AmDHs also suggests various physiological functions. Exploring these metagenomes is unquestionably beneficial for accessing variability in sequence and activity related to their distribution and environmental abundance.

Conclusions

Industrial biotechnology stands to benefit from a more diverse set of enzymes which can perform reductive amination with inherently good atom economy. Further, a more generalized approach for enzyme discovery is necessary to expand the space of available biocatalysts. Using a metagenomics approach, six experimentally validated nat-AmDHs were used to mine three marine and one human intestinal environmental databases, resulting in the identification of hundreds of hits. Several candidates had no close homologs reported in UniprotKB database, demonstrating the implicit benefit of using this type of approach to discover previously unreported enzyme sequences. Apart from *PortiAmDH* and IGCAmDH2, all the representative enzymes tested have less than 62% amino acid sequence identity compared to the previously reported nat-AmDHs. More importantly, biocatalysts for green chemistry should be sought among eukaryotic databases; the very active and more distant homolog MATOUAmDH2 serves as proof. This diversity widens the range of characterized sequences of this superfamily. Multiple sequence alignments and analysis of the active site residues of some selected enzymes showed cluster partitioning, highlighting the variability in active sites for AmDH activity. The set of active site residues which were validated for *Cfus*- and *MsmAmDH* activity are expanded by the newly discovered enzymes in this work. Importantly, the identified enzymes exhibited different cofactor specificities and specific activities from the previously reported nat-AmDHs. MATOUAmDH2 is not only the first eukaryotic AmDH identified

and characterized but it proved also to be the most promising enzyme due to its broad substrate scope and high specific activities. Potential shortcomings with its stability in the absence of cofactor can be overcome with further optimization. *ChatAmDH* and IGCAmDH1 are also highly interesting biocatalysts with high conversions and high turnovers towards isobutyraldehyde. Both protein engineering and reaction engineering are required for further improvements. Based on this newly diversified set of AmDHs identified from different environments, future iterations of metagenomic screens should give rise to enzymes of higher sequence diversity, particularly from the rich IGC database.

Materials and Methods

Materials

All chemicals were purchased from commercial sources and used without additional purification. All the ketones and aldehydes used in this study as well as benzoyl chloride (BzCl), 1-fluoro-2-4-dinitrophenyl-5-L-alanine amide (FDAA), cyclohexylamine (1a), isobutylamine (2a), racemic 2-methylcyclohexylamine (3b), racemic pentan-2-amine (4b), 1,2-cyclohexyldiamine, NADH, NADPH, NAD⁺ and NADP⁺ were bought from Sigma-Aldrich. 2-aminocyclohexanone was bought from Enamine. Glucose dehydrogenase (GDH) GDH105 was generously donated by Codexis®. UHPLC analyses were performed on a UHPLC U3000 RS 1034 bar system (Thermo Fisher Scientific, Waltham, USA) equipped with a DAD3000 diode array detector and a MSQ Plus™ Single Quadrupole Mass Spectrometer with electrospray ESI in positive mode (cone voltage = 75 V, the probe temperature set up at 450°C). The UHPLC column was a Kinetex® F5 (Phenomenex) column (100 × 2.1 mm; 1.7 µm). Spectrophotometric assays were recorded on a Safas UVMC2 (Safas, Monaco) thermostated when specified with a refrigerated/heating circulator Corio CD-200F (Jubabo®, Seelbach, Germany) using microcells high-precision cell quartz with 10-mm light path (Hellma Analytics, Müllheim, Germany).

Selection of candidates among metagenomics databases

Sequences of the six experimentally validated nat-AmDHs (*MsmAmDH*, *MvacAmDH*, *MycoAmDH*, *MicroAmDH*, *ApauAmDH* and *CfusAmDH*)³⁴ were aligned using the MAFFT sequence alignment software (version 7.407, auto mode),⁴⁸ and the resulting alignment was used to generate an HMM model with the hmmbuild tool of the Hmmer package (version 3.1b2).⁴⁹ Three databases of environmental sequences, OM-RGCv1, MATOUv1 and IGC were screened using the hmmsearch tool, and sequences showing significant matches with this nat-AmDHs HMM (corresponding E values are indicated in Table 1 and ESI Fig. S1-S3) were selected (12 from OM-RGCv1, 11 from MATOUv1 and 251 from IGC). A sequence similarity matrix of each set of candidates was built using Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) and 18 representative sequences were selected to screen this

biodiversity with a limited number of candidates to be produced and experimentally tested. Criteria used for this selection are detailed in ESI Fig. S1-S3.

Modeling by homology of AmDH candidates

All 18 AmDH candidates were modeled with the automated protein structure homology-modelling server Swiss-Model⁵⁰ using *MsmeAmDH* and *CfusAmDH* as templates. The accuracy of the models were assessed using the Z-score outputs from the software. Structural analysis including alignments were performed with PyMOL (Schrödinger®).

Enzyme preparation

The 18 selected genes were synthesized by Twist Bioscience (San Francisco, United States) and optimized for expression in *E. coli* (see ESI Table S1). Genes were then amplified from these synthetic fragments by adding to the primers (see ESI, Table S2) specific extensions for cloning into pET22b(+) (Novagen) modified for ligation independent cloning (LIC).⁵¹ The forward primers introduced a hexahistidine tag sequence in the proteins after the initial methionine for purification purposes. The cloned genes were sequenced (see ESI Table S1). In the case of all proteins except for MATOUAmDH1, MATOUAmDH2 and *AcoAmDH*, the plasmids were transformed into BL21-CodonPlus (DE3)-RIPL competent cells (Agilent technologies) in 96 micro-well plates. An autoinduction protocol was used by adding 2 g L⁻¹ of lactose and 0.5 g L⁻¹ of glucose. After inoculation with TB medium, the cultures were incubated at 30°C overnight. After centrifugation, the cell pellets were kept at -80°C for a few hours before lysis (100 mM phosphate buffer pH7.5, 50 mM NaCl, 10% glycerol, pepabloc 1 mM, lysonase) and sonication. In the cases of MATOUAmDH1, MATOUAmDH2 and *AcoAmDH*, the preparation was done from a 50-mL culture. After growing over-night at 37°C, a TB medium was inoculated and grown to an O.D. = 0.8-1. The proteins were induced with 0.5 mM of isopropyl β-D-1-thiogalactopyranoside (IPTG) and the cultures were incubated at 20°C over-night. Total protein concentrations were determined by the Bradford method with bovine serum albumin as the standard.⁵² The samples were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) using the Invitrogen NuPAGE system. All the proteins were visibly expressed (see ESI Fig. S12).

Purification of selected enzymes

Purification of IGCAmDH1, IGCAmDH5, *SaorAmDH*, *ChatAmDH*, *TtherAmDH*, *AcoAmDH*, MATOUAmDH1, and MATOUAmDH2 were conducted from a 500 mL culture by nickel affinity chromatography in tandem with gel filtration (Hi Load 16/600 Superdex 200pg) as described previously.⁵³ The storage buffer was 50 mM Tris-HCl pH 8.0, 50 mM NaCl, 15% glycerol and 1 mM DTT. Protein concentrations were determined by the Bradford method with bovine serum albumin as the standard.⁵² The samples were analyzed by SDS-PAGEs using the Invitrogen NuPAGE system (see ESI Fig. S13). The purified proteins were stored at -80 °C.

UHPLC-UV/MS screening assay

To a reaction mixture (total volume = 100 μL) containing 10 mM of substrate, 0.2 mM NADPH, 0.2 mM NADH, 11 mM glucose, 3 U/mL glucose dehydrogenase (GDH105) in 1 M ammonium formate buffer (pH 8.2) was added 30 μL of cell-free extract with the targeted overexpressed enzyme (0.06 - 0.4 mg total protein). The reaction was shaken at 400 rpm for 48 h at 30°C. All the reaction mixtures were analyzed by UHPLC UV/MS after derivatization with benzoyl chloride (BzCl) for conversion rate determination and after derivatization with Marfey's reagent (FDAA) for stereochemical analysis. Background reactions were performed in the same manner but with mixtures lacking the substrate or with cell-free extract without overexpressed enzyme. Amounts of amines in reaction mixtures were deduced from calibration curves obtained after BzCl or FDAA derivatization of standard commercial amines in the same reaction conditions except in the absence of enzyme and nicotinamide cofactor.

BzCl derivatization: To 20 μL of reaction mixture were added 10 μL of a 1 M Na₂CO₃/NaHCO₃ aqueous solution pH 9.5, 40 μL of H₂O, and 30 μL of a 50 mM BzCl solution in acetonitrile. The mixture was vortexed for 30 s and then quenched by the addition of 20 μL of a 1 M HCl aqueous solution and 30 μL of a 1/1 solution of H₂O/acetonitrile. After centrifugation (10 krpm, 10 min, room temperature) and filtration (0.22 μm), the mixture was analyzed by UHPLC-UV/MS (eluent MeCN/H₂O + 0.1% formic acid with a linear gradient 20/80 during 1 min, then 20/80 to 70/30 in 3 min, then 70/30 during 2 min followed by re-equilibration time; flow 0.5 mL min⁻¹; temperature 25°C; injection volume 3 μL; UV detection at λ = 250 nm; MS detection parameter: electrospray ionization in the positive mode, probe 450°C, 50 V).

FDAA derivatization: To 20 μL of reaction mixture were added 8 μL of a 1 M NaHCO₃ aqueous solution, and 20 μL of a 15 mM FDAA solution in ethanol. The mixture was stirred for 2 h at 65°C and then quenched by the addition of 4 μL of a 2 M HCl aqueous solution and 40 μL of acetonitrile. After centrifugation (10 krpm, 10 min, room temperature) and filtration (0.22 μm), the mixture was analyzed by UHPLC-UV/MS (eluent MeOH/H₂O + 0.1% formic acid with an isocratic elution (70/30) during 10 min; flow 0.3 mL min⁻¹; temperature 25°C; injection volume 3 μL; UV detection at λ = 340 nm; MS detection parameter: electrospray ionization in the positive mode, probe 450°C, 50 V).

Specific activity measurements

Temperature profile: A reaction mixture of ketone/aldehyde (10 mM) in ammonium formate buffer (1M, pH 8.0) was preincubated at the specified temperature in a ThermoMixer® (Eppendorf). After addition of a mixture of NADH/NADPH (0.1 mM each), and transfer in a thermostat spectrophotometric cell, an appropriate amount of purified enzyme (0.005–0.05 mg mL⁻¹) was added to initiate the reaction. The initial slope measured at 340 nm, in a temperature-controlled UV-spectrophotometer set to the specified temperature,

determined the specific activity of the enzyme according to Beer–Lambert's law and the molar absorptivity of β -NADH ($\epsilon = 6,220 \text{ M}^{-1} \text{ cm}^{-1}$) after subtraction of the slope obtained under the same conditions except without substrate and/or without enzyme.

Substrate screen: The same protocol as above was used at 50°C with the specified substrate.

Amine profile: To a mixture of Tris HCl buffer (100 mM, pH 10.0), 0.2 mM NADH (NADPH in the case of MATOUAmDH2), 10 mM of isobutyraldehyde and amine (100 mM) was added an appropriate amount of purified enzyme to initiate the reaction (reaction final volume 100 μ l). The specific activity was determined as described above.

Kinetic parameter determination

Kinetic parameters were determined by spectrophotometric NAD(P)H-monitoring at 340 nm. A reaction mixture of ketone/aldehyde in ammonium formate buffer pH 8.0 was preincubated at 50°C in a ThermoMixer® (Eppendorf). After addition of a mixture of NADH (or NADPH in the case of MATOUAmDH2), the reaction mixture was transferred to a thermostat spectrophotometric cell (optical path of 1 cm) preheated to 50°C and an appropriate amount of purified enzyme was added quickly after stabilization to initiate the reaction. Initial rates of the reaction were measured with various concentrations of substrate and saturated concentrations of the other substrates (or optimal concentration in case of NH_3 for MATOUAmDH2, see ESI Fig. S14). The uncertainties are those generated by the fitting and the data are averages of three experiments. Details regarding the fitting (Sigma Plot software) and plots are provided in ESI Figure S14.

Conversion rates for isobutylamine (2b)

Reactions at 10 mM of substrate: To a reaction mixture (total volume = 100 μ l) containing 10 mM of isobutyraldehyde (2a), 0.2 mM NAD^+ (NADP^+ in case of MATOUAmDH2), 1.2 eq of glucose, 3 U/mL of glucose dehydrogenase in 2 M ammonium formate buffer (pH 8.2) was added 0.1 mg mL^{-1} of purified enzyme. The reaction was shaken at 400 rpm for 24 h at 30°C, 40°C or 50°C. All the reaction mixtures were analyzed by UHPLC-UV/MS after derivatization with BzCl as described. Background reactions were performed in the same manner but with mixtures lacking the substrate. Amounts of amine in reaction mixtures were deduced from calibration curves obtained after BzCl derivatization of the corresponding standard commercial amine in the same reaction conditions except in presence of elution purification buffer instead of enzyme solution, and in absence of nicotinamide cofactor.

Reaction at 50 or 200 mM of substrate: The protocol was the same as above, except that before BzCl derivatization, the reaction mixture was diluted in water to 10 mM maximum theoretical amount of amine. Results obtained with reaction of IGCAmDH1 (dimer) with 50 and 200 mM of isobutyraldehyde (reaction at 40°C) are detailed in Table S6.

Stability studies

Seven of the candidates and three of the reference enzymes were expressed and purified. The solutions were diluted to 0.5 mg mL^{-1} of protein and loaded into capillaries for analysis by nanoDSF (differential scanning fluorimetry). The capillaries were loaded into a Prometheus NT.48 (NanoTemper Technologies GmbH, München, Germany) and a 1°C/min temperature ramp from 20 to 90°C was used at a 30% fluorescence intensity setting. The fluorescence ratio at 350 and 330 nm was tracked, and the inflection points corresponding to melting temperatures were outputted by the instrument's software.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors thank M. Salanoubat for supporting the project, Christine Pelle and Peggy Sirvain for large-scale purification of the described enzymes. The LABGeM (CEA/Genoscope & CNRS UMR8030), the France Génomique and French Bioinformatics Institute national infrastructures (funded as part of Investissement d'Avenir program managed by Agence Nationale pour la Recherche, contracts ANR-10-INBS-09 and ANR-11-INBS-0013) are acknowledged for support within the MicroScope annotation platform. This work was supported by Commissariat à l'énergie atomique et aux énergies alternatives (CEA), the CNRS, the University of Evry Val d'Essonne and the Georgia Institute of Technology. The authors thank the French Embassy at United States of America for a fellowship to Adam Caparco under the STEM Chateaubriand Fellows 2019 program. J.A.C. and A.S.B. acknowledge the support of NSF-CBET grants 1510551 and 1512848, respectively. We thank Codexis for the donation of glucose dehydrogenase for our experiments. We wish to acknowledge the core facilities at the Parker H. Petit Institute for Bioengineering and Bioscience at the Georgia Institute of Technology for the use of their shared equipment, services and expertise.

References

1. R. A. Sheldon and J. M. Woodley, *Chem. Rev.*, 2018, **118**, 801-838.
2. R. A. Sheldon and D. Brady, *Chemical Communications*, 2018, 54 6088-6104.
3. N. J. Turner and R. Kumar, *Current opinion in chemical biology*, 2018, **43**, A1-A3.
4. J. M. Woodley and N. J. Turner, in *Handbook of Green Chemistry, Green Chemical Engineering*, ed. A. A. Lapkin, , 2019, vol. 12.
5. M. Ferrer, M. Martinez-Martinez, R. Bargiela, W. R. Streit, O. V. Golyshina and P. N. Golyshin, *Microb Biotechnol*, 2016, **9**, 22-34.
6. M. A. Distaso, H. Tran, M. Ferrer and P. N. Golyshin, 2016, DOI: 10.1007/978-3-319-31421-1_216-1, 1-25.

7. A. Zapparucha, V. de Berardinis and C. Vaxelaire-Vergne, 2018, DOI: 10.1039/9781788010450-00001, 1-27.
8. J. W. E. Jeffries, N. Dawson, C. Orenco, T. S. Moody, D. J. Quinn, H. C. Hailes and J. M. Ward, *ChemistrySelect*, 2016, 1, 2217-2220.
9. A. Escobar-Zepeda, A. Vera-Ponce de Leon and A. Sanchez-Flores, *Front Genet*, 2015, 6, 348.
10. A. Madhavan, R. Sindhu, B. Parameswaran, R. K. Sukumaran and A. Pandey, *Applied biochemistry and biotechnology*, 2017, 183, 636-651.
11. W. Streit, J. Chow, K.-E. Jaeger, N. Katzke, R. Koch, M. Borchert, M. Björnqvist, M. Roggenbuck, F. Lippold, P. Golyshin, O. Golyshina, H. Tran, M. Ferrer, R. Daniel and J. Hollensteiner, *Impact*, 2018, 2018, 82-84.
12. M. Hajighasemi, A. Tchigvintsev, B. Nocek, R. Flick, A. Popovic, T. Hai, A. N. Khusnutdinova, G. Brown, X. Xu, H. Cui, J. Anstett, T. N. Chernikova, T. Bröls, D. Le Paslier, M. M. Yakimov, A. Joachimiak, O. V. Golyshina, A. Savchenko, P. N. Golyshin, E. A. Edwards and A. F. Yakunin, *Environmental Science & Technology*, 2018, 52, 12388-12401.
13. D. Baud, J. W. E. Jeffries, T. S. Moody, J. M. Ward and H. C. Hailes, *Green Chemistry*, 2017, 19, 1134-1143.
14. L. Leipold, D. Dobrijevic, J. W. E. Jeffries, M. Bawn, T. S. Moody, J. M. Ward and H. C. Hailes, *Green Chemistry*, 2019, 21, 75-86.
15. M. Höhne and U. T. Bornscheuer, *ChemCatChem*, 2009, 1, 42-51.
16. H. Groger, *Appl Microbiol Biotechnol*, 2018, DOI: 10.1007/s00253-018-9452-0.
17. M. J. Abrahamson, E. Vazquez-Figueroa, N. B. Woodall, J. C. Moore and A. S. Bommaris, *Angew Chem Int Ed Engl*, 2012, 51, 3969-3972.
18. G. Grogan, *Current opinion in chemical biology*, 2017, 43, 15-22.
19. S. C. Cosgrove, A. Brzezniak, S. P. France, J. I. Ramsden, J. Mangas-Sanchez, S. L. Montgomery, R. S. Heath and N. J. Turner, *Methods in enzymology*, 2018, DOI: 10.1016/bs.mie.2018.04.022.
20. M. D. Patil, G. Grogan, A. S. Bommaris and H. Yun, *ACS Catalysis*, 2018, DOI: 10.1021/acscatal.8b02924.
21. A. Gomm and E. O'Reilly, *Current opinion in chemical biology*, 2017, 43, 106-112.
22. I. Slabu, J. L. Galman, R. C. Lloyd and N. J. Turner, *ACS Catalysis*, 2017, DOI: 10.1021/acscatal.7b02686, 8263-8284.
23. M. D. Patil, G. Grogan, A. Bommaris and H. Yun, *Catalysts*, 2018, 8, 254.
24. P. T. Anastas and J. C. Warner, *Green chemistry : theory and practice*, Oxford University Press, Oxford [England]; New York, 1998.
25. M. J. Abrahamson, J. W. Wong and A. S. Bommaris, *Advanced Synthesis & Catalysis*, 2013, 355, 1780-1786.
26. S. K. Au, B. R. Bommaris and A. S. Bommaris, *ACS Catalysis*, 2014, 4, 4021-4026.
27. B. R. Bommaris, M. Schurmann and A. S. Bommaris, *Chemical Communications*, 2014, 50, 14953-14955.
28. L. J. Ye, H. H. Toh, Y. Yang, J. P. Adams, R. Snajdrova and Z. Li, *ACS Catalysis*, 2015, 5, 1119-1122.
29. J. Lowe, A. A. Ingram and H. Groger, *Bioorg Med Chem*, 2017, 26, 1387-1392.
30. A. Pushpanath, E. Sirola, A. Bornadel, D. Woodlock and U. Schell, *ACS Catalysis*, 2017, DOI: 10.1021/acscatal.7b00516, 3204-3209.
31. F. Chen, G.-W. Zheng, L. Liu, H. Li, Q. Chen, f. li, C.-X. Li and J.-H. Xu, *ACS Catalysis*, 2018, 8, 2622-2628.
32. V. Tseliou, T. Knaus, M. F. Masman, M. L. Corrado and F. G. Mutti, *Nat Commun*, 2019, 10, 3717.
33. O. Mayol, S. David, E. Darii, A. Debar, A. Mariage, V. Pellouin, J.-L. Petit, M. Salanoubat, V. de Berardinis, A. Zapparucha and C. Vergne-Vaxelaire, *Catalysis Science & Technology*, 2016, 6, 7421-7428.
34. O. Mayol, K. Bastard, L. Beloti, A. Frese, J. P. Turkenburg, J.-L. Petit, A. Mariage, A. Debar, V. Pellouin, A. Perret, V. de Berardinis, A. Zapparucha, G. Grogan and C. Vergne-Vaxelaire, *Nature Catalysis*, 2019, DOI: 10.1038/s41929-019-0249-z.
35. G. A. Aleku, J. Mangas-Sanchez, J. Citoler, S. P. France, S. L. Montgomery, R. S. Heath, M. P. Thompson and N. J. Turner, *ChemCatChem*, 2018, DOI: 10.1002/cctc.201701484.
36. S. P. France, R. M. Howard, J. Steflik, N. J. Weise, J. Mangas-Sanchez, S. L. Montgomery, R. Crook, R. Kumar and N. J. Turner, *ChemCatChem*, 2018, 10, 1-6.
37. M. Sharma, J. Mangas-Sanchez, S. P. France, G. A. Aleku, S. L. Montgomery, J. I. Ramsden, N. J. Turner and G. Grogan, *ACS Catalysis*, 2018, 8, 11534-11541.
38. A. Bornadel, S. Bisagni, A. Pushpanath, S. L. Montgomery, N. J. Turner and B. Dominguez, *Organic Process Research & Development*, 2019, DOI: 10.1021/acs.oprd.9b00123.
39. I. A. Castilla, D. F. Woods, F. J. Reen and F. O'Gara, *Mar Drugs*, 2018, 16.
40. M. Ferrer, C. Mendez-Garcia, R. Bargiela, J. Chow, S. Alonso, A. Garcia-Moyano, G. E. K. Bjerga, I. H. Steen, T. Schwabe, C. Blom, J. Vester, A. Weckbecker, P. Shahgaldian, C. de Carvalho, R. Meskys, G. Zanaroli, F. O. Glockner, A. Fernandez-Guerra, S. Thambisetty, F. de la Calle, O. V. Golyshina, M. M. Yakimov, K. E. Jaeger, A. F. Yakunin, W. R. Streit, O. McMeel, J. B. Calewaert, N. Tonne and P. N. Golyshin, *FEMS Microbiol Lett*, 2019, 366.
41. J. Li, H. Jia, X. Cai, H. Zhong, Q. Feng, S. Sunagawa, M. Arumugam, J. R. Kultima, E. Prifti, T. Nielsen, A. S. Juncker, C. Manichanh, B. Chen, W. Zhang, F. Levenez, J. Wang, X. Xu, L. Xiao, S. Liang, D. Zhang, Z. Zhang, W. Chen, H. Zhao, J. Y. Al-Aama, S. Edris, H. Yang, J. Wang, T. Hansen, H. B. Nielsen, S. Brunak, K. Kristiansen, F. Guarner, O. Pedersen, J. Doré, S. D. Ehrlich, H. I. T. C. Meta, N. Pons, E. Le Chatelier, J.-M. Batto, S. Kennedy, F. Haimet, Y. Winogradski, E. Pelletier, D. LePaslier, F. Artiguenave, T. Bruls, J. Weissenbach, K. Turner, J. Parkhill, M. Antolin, F. Casellas, N. Borruel, E. Varela, A. Torrejon, G. Denariar, M. Derrien, J. E. T. van Hylckama Vlieg, P. Viera, R. Oozeer, J. Knoll, M. Rescigno, C. Brechot, C. M'Rini, A. Mérieux, T. Yamada, S. Tims, E. G. Zoetendal, M. Kleerebezem, W. M. de Vos, A. Cultrone, M. Leclerc, C. Juste, E. Guedon, C. Delorme, S. Layec, G. Khaci, M. van de Guchte, G. Vandemeulebrouck, A. Jamet, R. Dervyn, N. Sanchez, H. Blottière, E. Maguin, P. Renault, J. Tap, D. R. Mende, P. Bork and J. Wang, *Nature biotechnology*, 2014, 32, 834.
42. A. Alberti, J. Poulain, S. Engelen, K. Labadie, S. Romac, I. Ferrera, G. Albini, J. M. Aury, C. Belser, A. Bertrand, C. Cruaud, C. Da Silva, C. Dossat, F. Gavory, S. Gas, J. Guy, M. Haquell, E. Jacoby, O. Jaillon, A. Lemainque, E. Pelletier, G. Samson, M. Wessner, S. G. Acinas, M. Royo-Lluch, F. M.

- Cornejo-Castillo, R. Logares, B. Fernandez-Gomez, C. Bowler, G. Cochrane, C. Amid, P. T. Hoopen, C. De Vargas, N. Grimsley, E. Desgranges, S. Kandels-Lewis, H. Ogata, N. Poulton, M. E. Sieracki, R. Stepanauskas, M. B. Sullivan, J. R. Brum, M. B. Duhaime, B. T. Poulos, B. L. Hurwitz, S. Pesant, E. Karsenti and P. Wincker, *Scientific data*, 2017, **4**, 170093.
43. S. Sunagawa, L. P. Coelho and S. Chaffron, *Science*, 2015, **348**, 1261359.
44. Q. Carradec, E. Pelletier, C. Da Silva, A. Alberti, Y. Seeleuthner, R. Blanc-Mathieu, G. Lima-Mendez, F. Rocha, L. Tirichine, K. Labadie, A. Kirilovsky, A. Bertrand, S. Engelen, M. A. Madoui, R. Meheust, J. Poulain, S. Romac, D. J. Richter, G. Yoshikawa, C. Dimier, S. Kandels-Lewis, M. Picheral, S. Searson, O. Jaillon, J. M. Aury, E. Karsenti, M. B. Sullivan, S. Sunagawa, P. Bork, F. Not, P. Hingamp, J. Raes, L. Guidi, H. Ogata, C. de Vargas, D. Iudicone, C. Bowler and P. Wincker, *Nat Commun*, 2018, **9**, 373.
45. A. Rodríguez Benítez and A. R. H. Narayan, *ACS Central Science*, 2019, **5**, 1747-1749.
46. N. Fonknechten, A. Perret, N. Perchat, S. Tricot, C. Lechaplais, D. Vallenet, C. Vergne, A. Zapparucha, D. Le Paslier, J. Weissenbach and M. Salanoubat, *Journal of bacteriology*, 2009, **191**, 3162-3167.
47. E. A. Hoffman, B. L. Frey, L. M. Smith and D. T. Auble, *Journal of Biological Chemistry*, 2015, **290**, 26404-26411.
48. K. Katoh, J. Rozewicki and K. D. Yamada, *Briefings in bioinformatics*, 2019, **20**, 1160-1166.
49. S. C. Potter, A. Luciani, S. R. Eddy, Y. Park, R. Lopez and R. D. Finn, *Nucleic acids research*, 2018, **46**, W200-W204.
50. A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. de Beer, C. Rempfer, L. Bordoli, R. Lepore and T. Schwede, *Nucleic acids research*, 2018, **46**, W296-W303.
51. K. Bastard, A. A. T. Smith, C. Vergne-Vaxelaire, A. Perret, A. Zapparucha, R. De Melo-Minardi, A. Mariage, M. Boutard, A. Debar, C. Lechaplais, C. Pelle, V. Pellouin, N. Perchat, J.-L. Petit, A. Kreimeyer, C. Medigue, J. Weissenbach, F. Artiguenave, V. De Berardinis, D. Vallenet and M. Salanoubat, *Nature chemical biology*, 2014, **10**, 42-49.
52. M. M. Bradford, *Analytical biochemistry*, 1976, **72**, 248-254.
53. N. Perchat, P.-L. Saaidi, E. Darii, C. Pelle, J.-L. Petit, M. Besnard-Gonnet, V. de Berardinis, M. Dupont, A. Gimbernat, M. Salanoubat, C. Fischer and A. Perret, *Proceedings of the National Academy of Sciences*, 2017, **115**, E4358-E4367.

	Tther AmDH	AmDH4	MATOUAmDH2	MATOUAmDH1	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	Misme AmDH	Mvac AmDH
Tther AmDH	100	54	21	22	29	24	25	25	26	27	26	30	31	29	28	29	27	28	29	29	30	29	29	31	31
AmDH4	54	100	22	24	23	24	24	22	23	22	23	25	25	25	23	27	26	23	24	24	25	24	25	26	25
MATOUAmDH2	21	22	100	53	20	32	35	33	29	33	33	34	31	34	30	35	32	32	33	28	31	32	30	33	32
MATOUAmDH1	22	24	53	100	21	32	34	32	30	32	33	34	33	34	30	36	34	33	33	29	28	29	29	30	30
OM-RGCAmDH2	29	23	20	21	100	26	29	27	29	29	27	27	30	28	27	26	27	27	27	33	34	33	32	33	32
Rgna AmDH	24	24	32	32	26	100	60	49	45	54	52	47	46	46	49	49	50	51	49	35	35	38	35	36	35
Dfai AmDH	25	24	35	34	29	60	100	49	45	59	52	50	49	53	51	51	52	51	51	37	36	38	37	40	39
Cfus AmDH	25	22	33	32	27	49	49	100	41	59	50	38	40	45	47	44	46	47	46	34	38	37	37	37	38
Sgor AmDH	26	23	29	30	29	45	45	41	100	52	57	54	58	53	58	55	56	60	56	37	40	41	37	39	39
Apau AmDH	27	22	33	32	29	54	59	59	52	100	66	50	51	57	61	57	62	60	60	41	38	41	38	39	38
IGCAmDH2	26	23	33	33	27	52	52	50	57	66	100	57	58	59	66	63	65	66	64	37	39	40	36	40	40
IGCAmDH6	30	25	34	34	27	47	50	38	54	50	57	100	65	65	63	66	65	67	65	40	39	40	40	39	38
IGCAmDH4	31	25	31	33	30	46	49	40	58	51	58	65	100	62	59	67	65	68	66	39	39	41	39	40	40
IGCAmDH5	29	25	34	34	28	46	53	45	53	57	59	65	62	100	67	70	69	72	67	38	41	41	40	42	39
IGCAmDH3	28	23	30	30	27	49	51	47	58	61	66	63	59	67	100	65	70	74	68	34	42	40	38	39	37
Crypt AmDH	29	27	35	36	26	49	51	44	55	57	63	66	67	70	65	100	74	78	73	38	37	41	39	42	39
IGCAmDH1	27	26	32	34	27	50	52	46	56	62	65	65	65	69	70	74	100	80	75	36	38	41	38	38	37
Chat AmDH	28	23	32	33	27	51	51	47	60	60	66	67	68	72	74	78	80	100	82	38	41	41	37	39	37
Acol AmDH	29	24	33	33	27	49	51	46	56	60	64	65	66	67	68	73	75	82	100	39	41	41	38	39	37
OM-RGCAmDH1	29	24	28	29	33	35	37	34	37	41	37	40	39	38	34	38	36	38	39	100	41	44	42	43	43
Micro AmDH	30	25	31	28	34	35	36	38	40	38	39	39	39	41	42	37	38	41	41	41	100	64	63	66	65
Porti AmDH	29	24	32	29	33	38	38	37	41	41	40	40	41	41	40	41	41	41	41	44	64	100	70	68	66
Myco AmDH	29	25	30	29	32	35	37	37	37	38	36	40	39	40	38	39	38	37	38	42	63	70	100	75	75
Misme AmDH	31	26	33	30	33	36	40	37	39	39	40	39	40	42	39	42	38	39	39	43	66	68	75	100	90
Mvac AmDH	31	25	32	30	32	35	39	38	39	38	40	38	40	39	37	39	37	37	37	43	65	66	75	90	100

Fig. 1. Percentage of identity matrix of the 18 candidates. The color code refers to the database from which the sequences were selected (purple: OM-RGC, green: MATOUv1, blue: IGC). The enzymes from the reference set (red) together with AmDH4 (black) are also reported for comparisons. This matrix was built using Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>).

a)

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20
<i>AmDH4*</i>	S78	A101	E102	G133	N135	F138	V139	R161	N163	D164	F168	V172	Q176	H197	I198	G199	H264	Q266	G299	T303
<i>Tther AmDH</i>	S80	A103	E104	G135	N137	F140	V141	R163	N165	D166	F170	V174	Q178	A197	I200	G201	H266	Q268	I299	T305
<i>OM-RGCAmDH2</i>	T80	D103	D104	G135	Q137	L140	G141	V162	T164	N165	A169	T173	F177	F195	R196	P197	Y264	V267	V298	R302
<i>OM-RGCAmDH1</i>	T78	S101	E102	G133	Q135	F138	W139	T160	I162	N163	L167	V171	Y175	P195	L196	G197	Q269	A271	T301	T305
<i>MATOUAmDH1</i>	S107	A130	E131	G162	I164	A167	C168	L189	Y191	D192	N196	L200	H204	K225	S226	Y227	Q304	G306	T338	T342
<i>MATOUAmDH2</i>	S87	A110	E111	G142	L144	A147	C148	L169	Y171	N172	Y176	L180	H184	K209	S210	Y211	Q278	G280	T312	T316
<i>MsmAmDH*</i>	S80	S103	E104	G135	Q137	F140	W141	A162	W164	N165	F169	L173	Q177	P197	T198	F199	M266	G268	T298	T302
<i>Mvac AmDH*</i>	S82	S105	E106	G137	Q139	F142	W143	A164	F166	N167	F171	L175	Q179	P198	T199	F200	M267	G269	T299	T303
<i>Myco AmDH*</i>	S80	S103	E104	G135	Q137	F140	W141	A162	W164	N165	F169	L173	Q177	P197	T198	F199	M266	G268	T298	T302
<i>Micro AmDH*</i>	S80	S103	E104	G135	Q137	F140	W141	S162	W164	N165	Y169	L173	Q177	P196	T197	F198	M265	G267	T297	T301
<i>Porti AmDH</i>	S79	S102	E102	G134	Q136	Y139	W140	A161	W163	N164	Y168	L172	Q176	P196	T197	F198	M265	G267	T297	T301
<i>Cfus AmDH*</i>	S84	G107	E108	G139	Q141	F144	W145	T166	Y168	N169	Y173	L177	H181	P200	S201	Y202	C269	G271	T301	T305
<i>Rgna AmDH</i>	S94	C117	E118	G149	P151	Y154	W155	S176	Y178	N179	Y183	L187	H191	P224	S225	Y226	C293	G295	T325	T329
<i>Dfai AmDH</i>	S81	N104	E105	G136	Q138	F141	W142	S163	Y165	N166	Y170	L174	H178	P213	S214	Y215	C282	G284	T314	T318
<i>Apau AmDH*</i>	S80	C103	E104	G135	Q137	F140	W141	T162	Y164	N165	Y169	L173	H177	P197	S198	Y199	C266	G268	T298	T302
<i>IGCAmDH2</i>	S84	C107	E108	G139	Q141	F144	W145	T166	Y168	N169	Y173	L177	H181	P201	A202	Y203	C270	G272	T302	T306
<i>Sgor AmDH</i>	S82	C105	E106	G137	Q139	Y142	W143	V164	Y166	N167	Y171	L175	H179	P199	S200	Y201	C268	G270	T300	T304
<i>IGCAmDH6</i>	S80	C103	E104	G135	E137	F140	W141	V162	Y164	N165	Y169	L173	H177	P197	S198	Y199	C266	G268	T298	T302
<i>IGCAmDH4</i>	S80	C103	E104	G135	Q137	F140	W141	Y162	Y164	N165	F169	L173	Q177	A197	S198	Y199	T266	G268	T298	T302
<i>IGCAmDH3</i>	S81	C104	E105	G136	Q138	F141	W142	V163	Y165	N166	Y170	L174	H178	P198	S199	Y200	C267	G269	T299	T303
<i>IGCAmDH5</i>	S82	C105	E106	G137	Q139	Y142	W143	V164	Y166	N167	Y171	L175	H179	P199	S200	Y201	C268	G270	T300	T304
<i>Crypt AmDH</i>	S82	C105	E106	G137	Q139	Y142	W143	T164	Y166	N167	Y171	L175	H179	P199	S200	Y201	C268	G270	T300	T304
<i>Chat AmDH</i>	S105	C128	E129	G160	Q162	Y165	W166	V187	Y189	N190	Y194	L198	H202	L221	A224	Y225	C292	G294	T324	T328
<i>IGCAmDH1</i>	S83	C106	E107	G138	Q140	F143	W144	V165	Y167	N168	Y172	L176	H180	L199	A202	Y203	C270	G272	T302	T306
<i>Acol AmDH</i>	S83	C106	E107	G138	Q140	F143	W144	V165	Y167	N168	Y172	L176	H180	L199	A202	Y203	C270	G272	T302	T306

b)

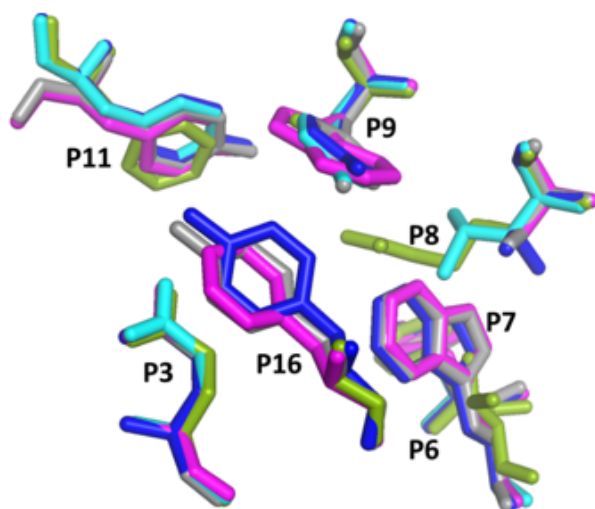


Fig. 2. Comparison of the active site residues P1-P20 of selected AmDHs found among screened metagenomes with those of *CfusAmDH*. All the P1 to P20 residues listed come from an alignment with *CfusAmDH*. a) The AmDH candidates and references (indicated by a star) are grouped in sub-clusters according to their residues P1 – P20 defining the active site pocket. Position P3 is hypothesized to be the catalytic residue. The residues are color coded by polarity and charge (blue: polar residues, yellow: hydrophobic residues, orange: aromatic residues, red: negatively charged residues, and teal: positively charged residues (charges at physiological pH)). b) overlay of the active sites of representatives of 3D sub-clusters: *TtherAmDH* (green), *MATOUAmDH2* (cyan), *PortiAmDH* (magenta), *ChatAmDH* (blue) and reference *CfusAmDH* (grey).

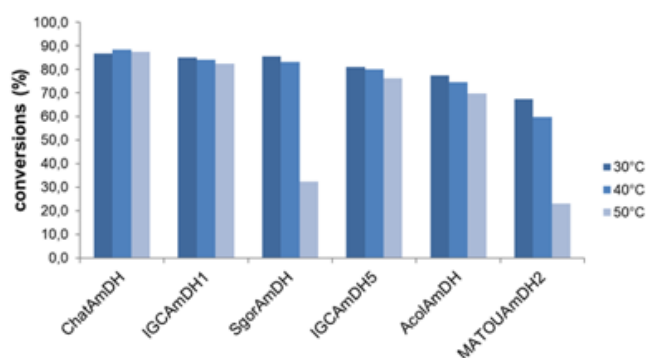


Fig. 3. Conversion of 10 mM of isobutyraldehyde (**2a**) into isobutrylamine (**2b**) at 30, 40 and 50°C with GDH-mediated cofactor recycling system for selected AmDHs. Reactions conditions: : 0.2 mM NAD⁺ (NADP⁺ for MATOUAmDH2), 10 mM substrate, 1 M ammonium formate buffer pH 8.2, GDH/glucose system for cofactor regeneration, 0.1 mg mL⁻¹ AmDH, 30°C, 24h. For more details, see Materials and Methods.

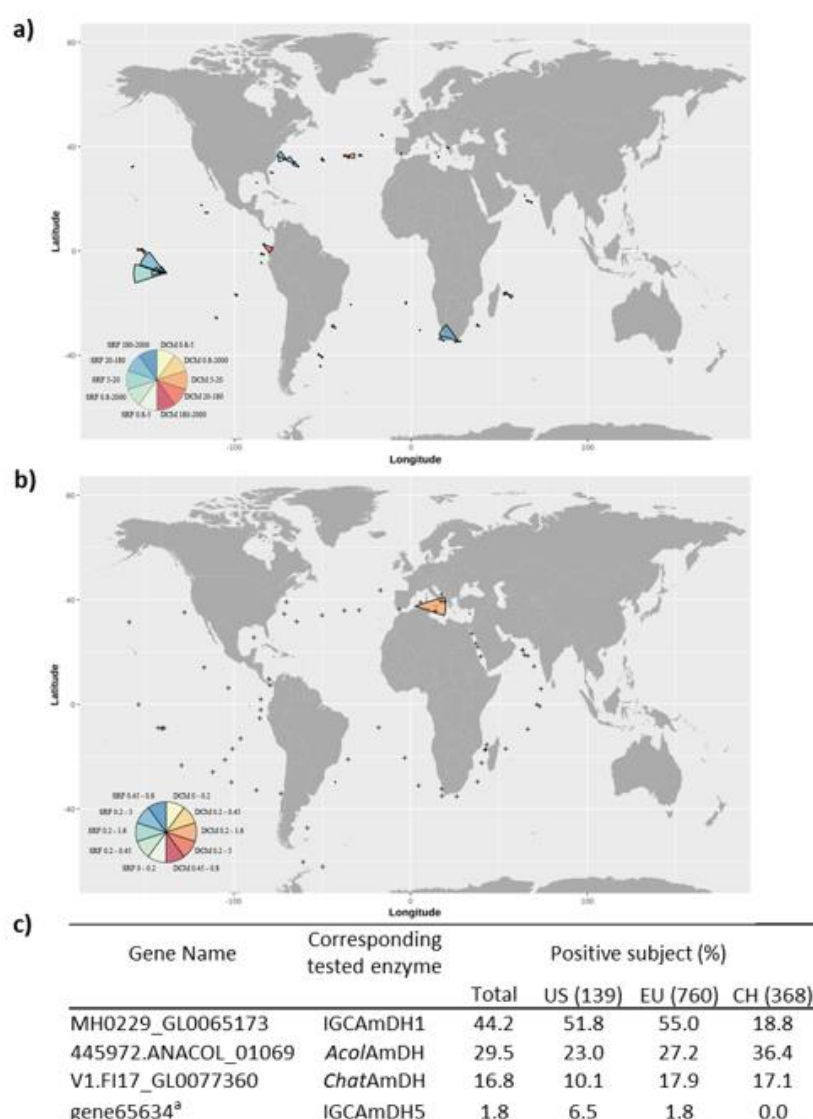


Fig. 4. Abundance of selected genes among the database sources. a) expression map of the gene coding for MATOUAmDH2 in Tara Oceans data (MATOU-v1 database), b) abundance map of the gene coding for SgorAmDH among Tara Oceans data (OM-RGC database). Black crosses represent sampling locations, c) percentage of patients where the genes coding for the corresponding AmDHs have been detected and distribution by cohort tested (IGC database). US (X) = X patients in United States of America, EU = European Union, CH = Switzerland; ^a full identifier = 160704339-stool1_revised_scaffold32525_1_gene65634.

Table 1. List of selected candidates from the screened metagenomics databases.

Resource	Environmental Identified Candidates	Name	First Homologs in UniprotKB (protein sequence identity, %)	Organism corresponding to the first homolog
OM-RGCv1 12 candidates (E value < 10 ⁻⁵⁰)	OM-RGC.v1.005913914	<i>PortiAmDH</i>	A0A2G1Y1R1 ^d (98)	<i>Porticoccus</i> sp.
	OM-RGC.v1.005006757	<i>SaorAmDH</i>	A8AV93 ^d (100)	<i>Streptococcus gordonii</i>
	OM-RGC.v1.004954749 ^d	OM-RGCAmDH1	L0AKZ3 (49)	<i>Natronobacterium gregarii</i>
	OM-RGC.v1.005531367 ^d	OM-RGCAmDH2	L0AKZ3 (35)	<i>Natronobacterium gregarii</i>
	OM-RGC.v1.004932010	<i>TtherAmDH</i>	M8DJE2 ^d (100)	<i>Thermoanaerobacter thermohydrosulfuricus</i>
MATOUv1 9 candidates (E value < 1e ⁻⁴⁰)	MATOU-v1_116481226_1 ^{a,d}	MATOUAmDH1	A0A143WV43 (35)	<i>Clostridiales</i> bacterium
	MATOU-v1_94466568_6 ^{a,d}	MATOUAmDH2	A0A437S6J8 (33)	<i>Anaerospira</i> sp.
IGC (MetaHIT + HMP) 251 candidates (E value < 10 ⁻¹⁰⁰)	MH0014_GL0085914	<i>RanaAmDH</i>	A7B1V2 ^d (99)	<i>Ruminococcus gnavus</i>
	V1.F117_GL0077360	<i>ChatAmDH</i>	D3AE01 ^d (100)	<i>Hungateella bathewayi</i>
	O2.CD3-O-PT_GL0003322	<i>DfaiAmDH</i>	A0A120KMV0 ^d (99)	<i>Desulfovibrio fairfieldensis</i>
	O2.UC48-O_GL0186261	<i>CryptAmDH</i>	R7CEI0 ^d (99)	<i>Cryptobacterium</i> sp.
	MH00229_GL0065173 ^d	IGCAmDH1	A0A084JH42 (80)	<i>Clostridium celerectescens</i>
	445972.ANACOL_01069	<i>AcolAmDH</i>	B0P8I0 ^d (100)	<i>Anaerotruncus colihominis</i>
	MH00233_GL0007304 ^d	IGCAmDH2	A0A0E2H951 (70)	<i>Clostridium clostridioforme</i>
	gene122828 ^{b,d}	IGCAmDH3	C0D1J1 (76)	<i>Clostridium asparagiforme</i>
	MH0224_GL0149939	IGCAmDH4	R6TLQ8 (88)	<i>Staphylococcus</i> sp.
	gene65634 ^{c,d}	IGCAmDH5	A0A136Q7Y1 (75)	<i>Christensenella minuta</i>
	MH0357_GL0138777 ^d	IGCAmDH6	A0A073J7A1 (76)	<i>Synergistes jonesii</i>

For each resource, the number of candidate below the fixed E value (parameter that describes the number of hits one can "expect" to see by chance when searching a database of a particular size) are indicated (see ESI Fig. S1-S3 for details). ^a sequence has been designed (see ESI Fig. S1 for details) ^b full identifier = 763860675-stool1_revised_scaffold58455_1_gene122828; ^c full identifier = 160704339-stool1_revised_scaffold32525_1_gene65634. ^d proteins used for cloning: in case of high protein sequence identity (≥ 98%) with a homologous protein present in UniprotKB database, the latter nucleic acid sequence was chosen instead of the metagenomic sequence from metagenomics data.

Table 2. Conversions and enantioselectivities/diastereoselectivities for the candidate AmDHs

	cyclohexylamine (1b)		2-methylcyclohexylamine (3b)				2-pentanamine (4b)		
	conv. (%)	conv. (%)	conv. (%)	trans. (%)	cis. (%)		conv. (%)	%	
				(1S,2S)	(1R,2R)	(1S,2R)	(1R,2S)	(S)	(R)
OM-RGCAmDH1	-	2.9							
OM-RGCAmDH2	3.1	13.7	7.2	-	-	96.8	3.2	>99.9	-
IGCAmDH1	4.5	13.7	13.2	-	-	97.2	2.8	>99.9	-
IGCAmDH2	2.7	15.1	0.4	na	na	na	na	-	na
IGCAmDH3	2.1	14.5	2.2	na	na	na	na	-	na
IGCAmDH4	3.2	11.9	8.9	-	0.8	91.7	7.4	>99.9	-
IGCAmDH5	2.5	13.2	0.6	na	na	na	na	-	na
IGCAmDH6	0.8	7.1	5.9	-	-	74.3	25.7	>99.9	-
<i>PortiAmDH</i>	3.8	11.8	0.4	na	na	na	na	-	na
<i>SaorAmDH</i>	1.3	8.7	1.9	na	na	na	na	-	na
<i>TtherAmDH</i>	17.1	14.2	23.9	5.2	0.7	90.3	3.8	>99.9	-
<i>RanaAmDH</i>	1.5	10.5	0.4	na	na	na	na	-	na
<i>ChatAmDH</i>	11.8	12.3	-	na	na	na	na	0.5	>99.9
<i>DfaiAmDH</i>	7.5	14.6	6.8	-	-	94.7	5.3	1.5	>99.9
<i>CryptAmDH</i>	0.8	9.9	0.4	na	na	na	na	-	na
<i>AcolAmDH</i>	5.8	13.9	-	na	na	na	na	-	na
MATOUAmDH1	0.1	0.1							
MATOUAmDH2	99.1	28.2	77.8	3.0	3.2	72.1	21.8	26.8	98.8

(-) = not detected; na: not applicable/not determined; empty cells: not done; conv. = conversions. Reactions conditions: 0.2 mM NADH/NADPH, 10 mM substrate, 1M ammonium formate buffer pH 8.2, GDH/glucose system for cofactor regeneration, cell-free extract, 30°C, 48 h. Conversions for **1b** and **2b** were determined after derivatization with benzoyl chloride; conversions and enantioselectivities/diastereoselectivities for **3b** and **4b** were determined after derivatization with Marfey's reagent (see Materials and Methods for details).

Table 3. Melting temperatures for selected AmDHs using DSF.

Sample ID	T1	T2	T3
<i>Micro</i> AmDH		54.2	
<i>Cfus</i> AmDH		52.4	
<i>Msmc</i> AmDH	25.4	53.6	
<i>Aco</i> AmDH		54.1	
<i>Trher</i> AmDH	27.9	54.0	68.1
<i>Saor</i> AmDH	31.5	53.8	
<i>Chat</i> AmDH		52.2	
MATOUAmDH2	33.5	53.1	
IGCAmDH1		55.7	
IGCAmDH5		53.8	

Temperatures T1, T2 and T3 are indicated in °C. See Materials and Methods for protocol details.

Table 4. Ketones/aldehydes substrate spectrum of AmDHs.

Substrate	Specific Activity (mU mg ⁻¹)														
	<i>Chat</i> AmDH		IGCAmDH1		<i>Saor</i> AmDH		IGCAmDH5		MATOUAmDH1		MATOUAmDH2		<i>Aco</i> AmDH		
	NADH	NADPH	NADH	NADPH	NADH	NADPH	NADH	NADPH	NADH	NADPH	NADH	NADPH	NADH	NADPH	
pyruvaldehyde	-	-	17.1	-	-	-	33.5	-	-	-	0.8	34.7	-	-	
isobutyraldehyde	578.1	152.7	3663.9	26.4	3295.8	201.0	1816.7	160.8	122.5	99.5	154.7	11070.6	286.5	30.1	
pentanaldehyde	395.5	6.8	112.5	5.8	142.3	6.4	76.4	6.4	21.1	34.5	3.7	609.0	20.5	11.7	
glutaraldehyde	-	-	-	-	-	-	-	-	-	-	7.8	-	-	-	
benzaldehyde	65.9	7.2	23.9	< 3.0	12.5	-	29.7	15.3	24.9	19.1	< 3.0	36.3	< 3.0	-	
2-pentanone	9.6	< 3.0	< 3.0	< 3.0	7.2	< 3.0	-	< 3.0	26.8	49.8	< 3.0	50.6	4.9	-	
MIBK	7.2	3.2	4.2	< 3.0	4.0	4.0	4.8	< 3.0	17.2	90.0	21.8	44.4	6.3	-	
3-hexanone	18.5	2.4	6.6	< 3.0	10.5	3.2	8.8	< 3.0	13.4	28.7	< 3.0	17.9	10.5	12.2	
cyclohexanone	192.9	12.9	32.8	3.2	23.3	< 3.0	35.4	5.6	42.1	49.8	33.9	315.4	< 3.0	-	
2-methyl cyclohexanone	11.3	4.0	4.8	< 3.0	4.0	6.4	5.6	-	26.8	80.4	20.0	128.1	< 3.0	-	
acetophenone	5.6	3.2	< 3.0	< 3.0	4.8	3.2	6.4	< 3.0	13.4	15.3	< 3.0	3.2	16.4	15.4	
p-tolyl phenylacetone	5.2	3.2	3.7	-	6.4	3.2	< 3.0	< 3.0	-	-	< 3.0	3.7	6.2	-	
4-oxopentanoate	15.3	8.0	10.6	7.7	12.1	23.3	16.9	16.9	36.4	59.3	-	< 3.0	-	-	
2,4 pentadione	-	-	-	-	-	-	-	-	-	-	< 3.0	< 3.0	< 3.0	-	
1,2-cyclohexadione	-	-	17.6	-	-	-	< 3.0	-	-	-	19.2	878.8	4.6	-	
peracmahor	7.2	3.2	5.8	-	3.2	4.8	7.2	7.2	7.7	15.3	17.1	36.8	16.6	13.2	
methyl 2-ethylacetoacetate	-	-	< 3.0	4.2	< 3.0	< 3.0	3.2	3.2	-	5.7	5.5	8.6	5.7	-	
ethyl 2-methylacetoacetate	< 3.0	-	4.8	< 3.0	< 3.0	4.0	4.8	-	7.7	28.7	< 3.0	7.7	< 3.0	-	
benzyl acetoacetate	-	< 3.0	-	< 3.0	< 3.0	< 3.0	-	-	-	-	< 3.0	5.9	4.8	-	

-, no detected activity. A specific activity of 1 U mg⁻¹ corresponds to 1 µmol of substrate converted per minute and per mg of enzyme. An average value of three independent experiments is marked. Reactions conditions: 0.1 mM NADH, 0.1 mM NADPH, 10 mM substrate, 1M ammonium formate buffer pH 8.0, appropriate amount of purified AmDH, 50°C. For more details, see Materials and Methods.

Table 5. Kinetic parameters for the selected enzymes.

	isobutyraldehyde (2a)			cyclohexanone (1a)		
	k_{cat}	K_M	k_{cat}/K_M	k_{cat}	K_M	k_{cat}/K_M
IGCAmDH1	7.90 ± 0.51 ^a	109.9 ± 18.1	71.9	0.07 ± 0.00	23.1 ± 3.5	2.8
IGCAmDH5	3.44 ± 0.15	55.7 ± 6.7	61.8	0.05 ± 0.00	41.0 ± 6.9	1.2
<i>Chat</i> AmDH	5.28 ± 0.31	26.6 ± 5.6	198.7	0.04 ± 0.00	19.4 ± 2.1	2.2
<i>Saor</i> AmDH	7.11 ± 0.32	38.1 ± 5.2	186.7	0.04 ± 0.00	6.8 ± 0.9	5.7
<i>Aco</i> AmDH	1.87 ± 0.09	24.5 ± 4.1	76.2	0.01 ± 0.00	9.6 ± 1.6	1.4
MATOUAmDH2 ^a	12.94 ± 0.59 ^b	0.80 ± 0.16	16,100.4	0.52 ± 0.02 ^b	0.6 ± 0.1	837.5

^a NADPH is used for each experiment instead of NADH; ^b apparent k_{cat} ; ^c overestimated values induced by the fitting. k_{cat} is expressed in s⁻¹, K_M in mM and k_{cat}/K_M in s⁻¹ M⁻¹. Temperature = 50°C. The uncertainties are those generated by the fitting of data averaged over three experiments. For more details, including kinetic parameters for NH₃, see Materials and Methods and ESI Table S5 and Fig. S14