Towards the construction of an autotrophic Escherichia coli strain

Marion Schulz

THÈSE DE DOCTORAT

NNT : 2023UPASL140

Towards the construction of an autotrophic *Escherichia coli* strain

Vers le développement d'une souche d'Escherichia coli autotrophe

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 577 : structure et dynamique des systèmes vivants (SDSV)
Spécialité de doctorat : biotechnologies
Graduate School : Sciences de la vie et santé. Référent : Université d'Évry Val d'Essonne

Thèse préparée dans l'unité de recherche **Génomique Métabolique** (Université Paris-Saclay, Univ Évry, CNRS, CEA), sous la direction de **Volker DÖRING**, directeur de recherche et le co-encadrement de **Christine CAVAZZA**, directrice de recherche

Thèse soutenue à Paris-Saclay, le 14 décembre 2023, par

Marion SCHULZ

Composition du Jury

Membres du jury avec voix délibérative

Stéphanie BURY-MONÉ Professeure, I2BC, Université Paris-Saclay	Présidente
Axel MAGALON Directeur de recherche, CNRS, Université Aix-Marseille	Rapporteur
Stéphane GUILLOUET Professeur, TBI, INSA Toulouse	Rapporteur
Olga SOUTOURINA Professeure, I2BC, Université Paris-Saclay	Examinatrice

ACKNOWLEDGMENTS

En me lançant dans cette aventure scientifique, j'étais loin de penser que ce serait une telle aventure humaine.

Tout d'abord, je tiens à remercier Volker Döring, mon Directeur de thèse, et Madeleine Bouzon-Bloch, Cheffe de l'équipe Métabolisme Synthétique, de m'avoir offert l'opportunité d'effectuer ma thèse au sein de leur laboratoire. Je les remercie également pour leur implication tout au long du projet, jusqu'à l'écriture du manuscrit et de l'article, malgré les nombreuses difficultés. Je leur suis reconnaissante notamment de m'avoir laissé la liberté d'explorer certains sujets qui me motivaient particulièrement. Toute ma gratitude à Carine Vergne-Vaxelaire pour sa bienveillance, ses encouragements et son écoute tout au long de ma thèse; ainsi qu'à Anne Zaparucha pour son soutien et la qualité de nos échanges toujours très sympathiques et constructifs. Je remercie également Véronique de Berardinis et Patrick Wincker d'avoir œuvré au bon déroulement de ma thèse.

Je voudrais également adresser mes sincères remerciements aux membres du jury qui ont accepté de participer à l'évaluation de ces travaux : Olga Soutourina, Stéphanie Bury-Moné, Stéphane Guillouet et Axel Magalon.

Je remercie aussi les deux membres de mon comité de thèse, Eliane Meilhoc et François Wasels, d'avoir accepté de m'accompagner durant ce projet ainsi que pour leurs précieux conseils et leur bienveillance.

Merci chaleureusement à Erika Brunet qui m'a accompagnée dans le cadre d'un programme de mentorat de l'association Femmes & Sciences durant la seconde année de ma thèse, merci d'avoir été à l'écoute, de m'avoir conseillée et aidée à prendre du recul.

Ensuite, je voudrais remercier l'équipe co-encadrante de mon projet de thèse, au CEA Grenoble, Christine Cavazza ainsi que Julien Perard qui m'ont encadrée lors de mes deux séjours à Grenoble. Je remercie par la même occasion tous les membres de ce laboratoire qui m'ont accueillie chaleureusement : Patrice, Roger, Thomas, Greg et Laurence. Un merci spécial à Jade, qui a été une très chouette rencontre et à qui je souhaite une « belle » fin de thèse, de la réussite dans sa carrière de photographe et de future prof et beaucoup de joies pour la suite.

Revenons au Genoscope, où j'ai passé la majeure partie de ma thèse et où j'ai rencontré beaucoup de personnes toujours prêtes à aider, ce qui a été très précieux pour moi durant ces trois années. Pour ne pas les citer : Christophe, Sébastien, Agnès, Nadia, Aline, Camille, Nathalie et tout le 1er étage... Un grand merci spécial à Peggy S., ma partenaire d'open space, qui avait toujours un mot gentil et qui s'est beaucoup impliquée dans toutes les purifications de protéines et tests ainsi que le séquençage chaque

semaine. Merci évidemment à Alain qui s'est également beaucoup investi pour la production de mes enzymes. Je remercie également tout particulièrement Jean-Louis, qui m'a sauvée avec ses PCR touchdown, sans lesquelles je serais encore en train d'essayer d'amplifier mon opéron. Merci aussi à Adrien d'avoir fait toutes ces plaques de séquences et merci à Emilie pour les qPCR. Un grand merci à David R. et Stéphanie pour l'analyse des données de séquençage génomiques. Merci également à Ioana pour sa pédagogie et de nous avoir aidés avec la mutagenèse. Je tiens à remercier aussi Valérie Pezo, j'ai beaucoup apprécié nos petits échanges réguliers scientifiques et moins scientifiques. Merci aussi à Catherine C. pour son aide et son efficacité dans la gestion des aspects plus logistiques.

Un très grand merci aux membres de mon laboratoire, qui se sont tous beaucoup impliqués durant ces trois années et qui ont été d'un grand soutien. Merci à Valérie, merci de m'avoir autant transmis et appris, on partait de si loin, également pour toutes ces discussions plus personnelles et très enrichissantes. Je lui souhaite tout le bonheur du monde et que la vie soit aussi douce que possible. Merci à Anne B. également, qui est partie dans un autre labo à la fin de ma deuxième année mais qui s'est beaucoup investie aussi, je lui souhaite beaucoup de joie pour ce nouvel épisode de sa carrière et de sa vie. Merci à Mélodie qui nous a rejoint pour ma dernière année de thèse et qui s'est également énormément impliquée dans ces travaux, je lui souhaite une belle poursuite de carrière et qu'elle réalise ses projets personnels également. Merci à Ivan pour sa gentillesse et pour s'être toujours préoccupé de savoir comment j'allais, de m'aider avec toutes ces histoires de manips innovantes avec du CO_2 que je voulais faire et de trouver des solutions. Merci à Laurent également de s'être impliqué dans ces manips en atmosphère modifiée qui n'étaient pas si simples à mettre en place.

Je remercie également tous les doctorants du Genoscope avec qui j'ai pu échanger et qui m'ont apporté du soutien durant ces trois années : Laurine, Déborah, Océane, Brieuc, Humbeline, Tom, et tous les autres. Evidemment un merci très spécial pour ma co-bureau qui est devenue une amie, Chloé. Nous n'avons partagé au labo que la 1ère année pour moi et la dernière pour elle, mais quelle rencontre ! On peut le dire, ça a été court mais intense. Nous nous sommes mutuellement accompagnées pendant nos fins de thèse respective et pourvu que ça dure !

Un merci également spécial à Solange qui a rejoint le Genoscope récemment mais qui m'a beaucoup apporté en peu de temps. J'ai apprécié nos conversations scientifiques, ton aide précieuse pour développer des outils de biologie moléculaire et ton partage de connaissances. D'un point de vue moins scientifique, merci pour ta bonne humeur et nos petites sessions gossips ! Je te souhaite tout le bonheur du monde à toi et à ta famille à venir.

C'est grâce à toutes ces personnes que j'ai rencontrées que, jour après jour, mes manipulations ont pu progresser, beau travail d'équipe ! J'ai tellement appris professionnellement et humainement. Ces 3

années resteront une belle parenthèse dans ma vie, avec le sentiment d'avoir parcouru un bon petit bout de chemin dans le monde de la recherche. Cela restera indéniablement une source inépuisable de souvenirs et un socle sur lequel je pourrai toujours m'appuyer pour la suite de ma carrière.

Je remercie du fond du cœur Oriane, qui est la parfaite transition entre le Genoscope et le monde extérieur. Amie de longue date, nous avons également été collègues pendant 2 ans au sein du même laboratoire et ta présence au quotidien a été très précieuse. Merci pour les gâteaux, les rires, d'avoir été un super bureau des plaintes et un soutien.

Je remercie évidemment tous mes amis avec qui j'ai passé de très bons moments tout au long de ces trois dernières années : Pauline R., Pauline B., Axelle, Clément, Gwendal, Thibault, Jérémie, et tous les autres. Ces moments ont été importants pour moi et m'ont permis penser à autre chose le temps d'un week-end, d'une soirée ou d'une balade.

Un grand merci maintenant à ma famille qui a aussi été d'un très grand soutien durant toute cette période : merci à tous, tantes, oncles, cousins et cousines, de m'avoir régulièrement demandé comment ça se passait, je me sentais soutenue. Evidemment, une pensée très spéciale pour mon cousin Germain, passionné de sciences, tu es et seras pour toujours dans mon cœur. Un immense merci bien sûr à mon grand frère Clément d'avoir été là avec Floflo et de nous avoir permis de partager de chaleureux moments en famille autour de bons repas. Merci aussi à eux deux de m'avoir offert la grande joie d'être tata. Merci à ma super-maman, ma supportrice de la première heure, celle qui m'a appris « quand on veut, on peut », elle est un exemple de courage et de détermination.

Enfin, merci à toi Nadir, merci d'avoir été si présent, de m'avoir soutenue au quotidien, d'avoir été si compréhensif, de me faire rire tous les jours et de t'impliquer autant. Ces mots que tu me disais tous les matins en partant, « allez, t'es la meilleure et tu vas y arriver » ont été de supers encouragements qui ont sans aucun doute joué un rôle déterminant à l'aboutissement de ce projet.

TABLE OF CONTENTS

ABBREVIATIONS

Abbreviation	Full name
ALE	Adaptive laboratory evolution
Amp	Ampicillin
aTc	Anhydrotetracycline
ATP	Adenosine triphosphate
bp	Base pair
C. necator	Cupriavidus necator
CBB	Calvin-Benson-Bassham
CEA	Commissariat à l'Energie Atomique et aux Energies Alternatives
$CHCl_3$	Chloroform
Cm	Chloramphenicol
CO	Carbon monoxide
CO_2	Carbon dioxide
CoA	Coenzyme A
CODH	carbon monodioxide dehydrogenase
CRISPR	Clustered regularly interspaced short palindromic repeats
CRISPRi	Clustered regularly interspaced short palindromic repeats interference
DAPG	Di-acetyl-phloroglucinol
DC/HB	Dicarboxylate/4-hydroxybutyrate
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotide triphosphates
dsDNA	Double-stranded deoxyribonucleic acid
DTT	Dithiothreitol
E. coli	Escherichia coli
FDH	Formate Dehydrogenase
Fe	Fer
FHL	Formate hydrogenlyase
FMN	Flavin mononucleotide
FRT	Flippase recognition target
G3P	Glyceraldehyde 3-phosphate
GCS	Glycine cleavage system
GM3	Genemat 3rd generation– directed evolution automaton
H^+	Proton

H_2	Dihydrogen
$HCO3^-$	Bicarbonate
HP/HB	3-hydroxypropionate/4-hydroxybutyrate
i.e.	That is
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Kan	Kanamycin
kbp	Kilo base pair
k_{cat}	Turnover number
kDa	kilodalton
K_m	Michaelis constant
LB	Lysogeny Broth
lpd	Dihydrolipoyl dehydrogenase
M, mM	molaire (mol.L^{-1}), mili-molaire
MAGE	Multiplex automated genome engineering
MAGE	Multiplex Automated Genome Engineering
mg	milligramme
MGD	Molybdopterin guanine dinucleotide
Mo	Molybdenum
MOG	Malonyl-CoA-oxaloacetate-glyoxylate
mol, pmol, fmol	Mole, picomol, femtomole
MOPS	Morpholinopropane sulfonate
MS	Murashige and Skoog
MV	Methylviologen
Na^+	Sodium
NAD	Nicotinamide adenine dinucleotide (oxidized form)
NADH	Nicotinamide adenine dinucleotide (reduced form)
NADP	Nicotinamide adenine nucleotide phosphate (oxidized form)
NADPH	Nicotinamide adenine nucleotide phosphate (reduced form)
ND	Not determined
nm	nanometre
NTA	Nitrilotriacetic acid
O_2	Dioxygen
OD	Optical Density
ORF	Open reading frame
Ori	Origin of replication

PCR	Polymerase chain reaction
PEP	Phosphoenolpyruvate
PFL	Pyruvate formate-lyase
ppm	Parts per million
PTDH	Phosphite dehydrogenase
R. capsulatus	Rhodobacter capsulatus
RBS	Ribosome binding site
RGP	Reductive glycine pathway
RNA	Ribonucleic acid
Rpm	Rotation per minute
rTCA	Reductive tricarboxylic acid
RT-qPCR	Quantitative reverse transcription polymerase chain reaction
RuBisCO	Ribulose-1,5-bisphosphate carboxylase/oxygenase
RuBP	Ribulose-bisphosphate
rWGS	Reverse water gas shift
S	Sulfur
SDM	site-directed mutagenesis
SDS	sodium dodecyl sulfate
SOC	Super Optimal Broth
Spec	Spectinomycin
ssDNA	Single-stranded deoxyribonucleic acid
TB	Terrific Broth
TCA	Tricarboxylic acid
TE	Tris-EDTA
Tris	tris(hydroxymethyl)aminomethane
UNFCCC	United Nations Framework Convention on Climate Change
UV	Ultra Violet
V_{max}	Maximum velocity
W	Tungsten
WT	Wild type

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

I. Introduction

1. Context of the doctoral project

The PhD project is part of the "FOCUS FCC" program funded by the Commissariat à l'Énergie Atomique et aux Énergies Alternatives (CEA). The FCC stands for "Fermeture du Cycle du Carbone" or Carbon Cycle Closure. Within this program, involving 25 PhD projects, 5 axes were developed and one of which is "New approaches for 3G biofuels". The project presented in this manuscript is part of this axis.

In order to respond to this call for proposals, two CEA laboratories partnered for this project: Biocatalysis, Bioremediation and Synthetic Metabolism Laboratory (L2BMS, Genoscope, CEA Evry) represented by Dr. Volker Döring, and Chemistry and Biology of Metals Laboratory (LCBM, CEA Grenoble) represented by Dr. Christine Cavazza. The ultimate objective of the project, as depicted in Figure 1, was to engineer an autotrophic *Escherichia coli* strain able to use carbon dioxide (CO_2) as sole carbon source and dihydrogen (H_2) as energy source. For this purpose, an O_2-tolerant formate dehydrogenase (FDH) capable of CO_2 reduction using NADH as electron donor was to be introduced in a formatotrophic *E. coli* strain previously developed in the laboratory of Évry, thus enabling CO_2 to become a carbon source upon reduction. The CO_2-reduction reaction consumes NADH, entailing the need of a reduced cofactor generation system operating in the cell. In the original scheme, an O_2-tolerant hydrogenase capable of H_2 cleavage upon NAD^+ reduction was envisioned to regenerate NADH.

The L2BMS possesses the expertise for metabolic engineering, in particular in *E. coli,* having conducted several projects dealing with the implementation of C1 assimilatory pathways in this biotechnological platform strain. In addition, the availability of GM3 devices at Genoscope, enabling automated continuous culture of microorganisms, is a major asset for such a metabolic engineering project. On the other hand, the LCBM at Grenoble has expertise in the manipulation of metal-dependent enzymes such as FDH and hydrogenases that will be heterologously expressed in *E. coli* and is equipped to handle experiments involving H_2.

In parallel to the NADH generation system involving hydrogenase which are complex metalloenzymes encoded by large operons, a less complex alternative was also tested relying on the NAD^+-dependent oxidation of phosphite to phosphate catalyzed by a phosphite dehydrogenase. A global view of the project is shown schematically in Figure 1.

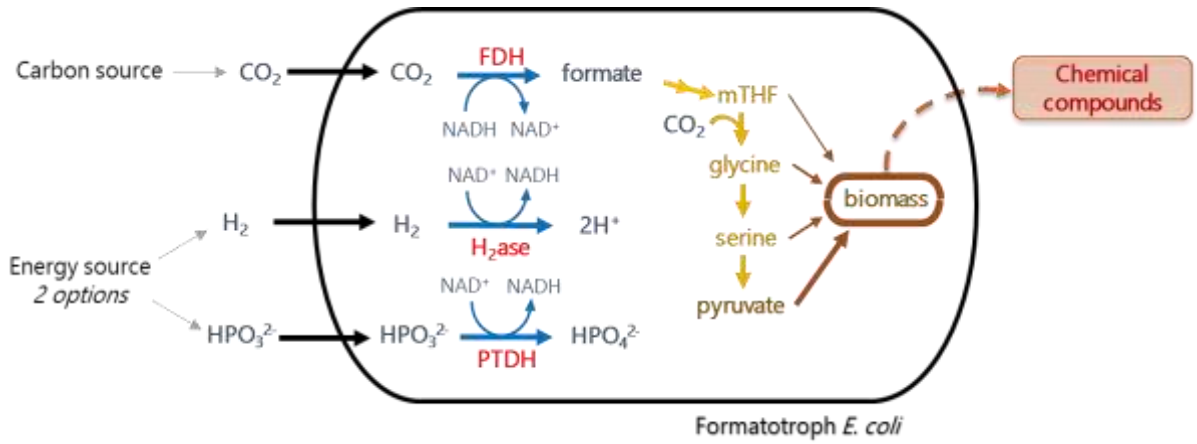


Figure 1. Global schematic representation of the PhD project.

A number of experimental challenges were faced during this highly exploratory research project. A choice was made to focus on the expression of a complex FDH in *E. coli* with the formate oxidation as selectable reaction sense. The results were conclusive and were further explored in parallel with the initial roadmap.

Throughout the experiments carried out in this project, I benefited from the contribution of several technicians and researchers. Valérie Delmas, Mélodie Cadillon and Anne Berger contributed to some strain constructions and metabolic tests. Ivan Dubois and Laurent Gaillon carried out the inoculation and the monitoring of the GM3 evolutions. Peggy Sirvain and Alain Perret assisted with the protein purification and *in vitro* testing experiments. Julien Perard helped with the hydrogenase investigations. Finally, David Roche and Stéphanie Fouteau performed whole genome sequencing data analysis.

2. CO_2 capture and assimilation

1. The necessity of CO_2 capture

Carbon dioxide, CO_2, is one of the three main greenhouse gases along with methane and nitrous oxide. Its rapid increase in the atmosphere is clearly documented in the datasets available. Human activities and the combustion of fossil fuels are responsible for the largest part of CO_2 emissions. Before the industrial revolution, the CO_2 concentration in the atmosphere was 30 % lower than today, reaching 420 ppm (2 degrees institute). Natural CO_2 assimilation pathways operating in nature are not able to absorb an ever-increasing quantity of CO_2. Thus, while the reduction of CO_2 emission is the first axis of action, efficient CO_2 capture and storage and CO_2 conversion are also important avenues being explored by the

United Nations Framework Convention on Climate Change (UNFCCC). To tackle this global issue, the Paris Agreement was signed by 196 Parties during COP21 in 2015 (UNFCCC website).

To achieve the objectives of this international treaty, and to move away from fossil carbon sources, the development of efficient CO_2 recycling processes into valuable chemicals or fuels is an axis of action to decrease the CO_2 emissions and the environmental impact of the production of such compounds. Currently, CO_2 conversion technologies are available and will be described in the following section. Among these technologies, CO_2 reduction by living organisms is an option being increasingly explored and this project is in line with this research axis.

2. CO_2 conversion technologies

Carbon capture, utilization and storage, abbreviated as CCUS, covers several concepts. The carbon capture and storage focuses on the recovery and transportation of the carbon; while the utilization refers to the processes and reactions of CO_2 conversion. The latter concept is the one that will be developed in this paragraph.

The principal technologies that can perform CO_2 conversion are various and can be classified as follows: photochemical, biochemical, electrochemical and thermochemical (Parvez et al. (2020); Yaashikaa et al. (2019); Yang et al. (2022); Serafini et al. (2023)). However, CO_2 is the most oxidized carbon form and the most stable under atmospheric conditions, thus requiring high amount of energy for its conversion.

1. Thermochemical conversion

The thermochemical conversion of CO_2 is largely dominated by the hydrogenation of CO_2, reaction in which H_2 reacts with CO_2 to form methane (CH_4), carbon monoxide (CO) and methanol (CH_3OH). The reverse water gas shift (rWGS) reaction (Equation 1) is an extensively studied thermochemical conversion of CO_2 and can be catalyzed by different catalysts (reviewed by Daza and Kuhn (2016)). In this reaction, the CO_2 source can be air, despite the low concentration of CO_2 (400 ppm), or gases containing much higher CO_2 concentrations. This reaction takes place under specific temperature and pressure conditions of 550−750 °C and 0−5 bar respectively (Roy, Cherevotan, and Peter 2018). The catalyst involved in this reaction can be a metal (copper, platinum, rhodium), copper-based, noble metals or oxides (zinc oxide, indium III oxide and many others), for example (Daza and Kuhn 2016).

Equation 1. The reverse water gas shift reaction. This reaction takes place in the presence of a catalyst.

$$CO_2 + H_2 \leftrightarrow H_2O + CO$$

The CO formed by this reaction can be further used as a precursor for other hydrocarbons through the Fischer-Tropsch process. Thus, the rWGS offers interesting perspectives through the wide range of possibilities that CO provides (Daza and Kuhn 2016).

Kothandaraman et al. (2016) reported efficient direct CO_2 conversion to methanol (CH_3OH) *via* a metal-based thermochemical reaction using a Ru-based catalyst and polyamine in an ethereal solvent. Formic acid generation using a very close mechanism for H_2 storage purposes was also reported by the catalytic hydrogenation of CO_2 (Filonenko et al. 2014).

Despite the high productivity of this type of CO_2 conversion, the major drawbacks are the conditions required for the reaction, *i.e.* high temperature and pressure needed for the reaction to occur, and the degradation of the catalyst during the process (Roy, Cherevotan, and Peter 2018).

2. Photochemical conversion

The photochemical conversion consists of CO_2 reduction using light energy. This reaction takes place in cells subjected to light-containing semiconductors such as TiO_2, ZnO or SiC among many others suspended in water where CO_2 is bubbled (Inoue et al. 1979). These reactions, depending on the photocatalyst, can lead to various chemicals such as formic acid (HCOOH), formaldehyde (CH_2O), methane (CH_4), carbon monoxide (CO) or hydrocarbons (Inoue et al. (1979); Wang, Chen, and Tang (2022)).

At present, this type of CO_2 conversion suffers from low productivity, preventing it from becoming an industrially valuable solution. This is partly due to the poor understanding of the mechanism of this reaction along with inherent reasons such as the high CO_2 stability, unfavorable thermodynamics of CO_2 reduction by water and low selectivity (Wang, Chen, and Tang 2022).

3. Electrochemical conversion

CO_2 reduction can be performed *via* an electrochemical process enabling the production of more reduced carbon products such as CO, CH_4, HCOOH, CH_3OH, and even C_2 and C_3 compounds (Serafini et al. 2023). This solution is both studied as a sustainable way to produce value-added chemicals and also an approach for the storage of excess electricity produced by renewable sources as they are intermittent by nature (Birdja et al. 2019).

1. Inorganic catalysis

CO_2 electrolysis has been mostly studied in aqueous media using metal or metal-derived electrocatalysts, the products of the reaction mostly depend on this choice. Formate has been efficiently

produced using lead (Pb) or palladium (Pd) nanoparticles dispersed on carbon particles, with high selectivity of 97.4 and 99 % (Hori et al. (1994); Min and Kanan (2015)).

This type of conversion does not require high temperature or pressure and researchers are currently working on processes requiring low overpotential (less than 1 V) for CO_2 reduction which can be achieved by using certain types of catalysts such as Pd on carbon (Pd/C) electrodes (Min and Kanan 2015).

2. *Enzyme catalysis*

The enzyme-dependent conversion of CO_2 is a promising method due to the high specificity of enzymatic reactions performed under mild conditions.

Two main families of enzymes were described as having CO_2 reduction activity, thus being of potential interest for biotechnological purposes:

- Formate dehydrogenases (FDH): reduction of CO_2 to formate
- Carbon monoxide dehydrogenases (CODH): reduction of CO_2 to CO

CODH are enzymes able to catalyze the reversible conversion of CO_2 to CO under anaerobic conditions and are found in organisms using the Wood-Ljungdahl pathway (described in I. 2. 3. 3, Shi et al. (2015)). CO can then be converted to methanol, which is used as a liquid fuel, or other hydrocarbons (as mentioned in the Thermochemical conversion paragraph, I. 2. 2. 1). The organisms using this pathway are strict anaerobes and experiments using CODH for CO_2 reduction are always performed in anoxic conditions. The first CODH reported to catalyze this reaction efficiently *in vitro* is the one from *Moorella thermoacetica*, the conversion was mediated with an electrode and methylviologen (MV) as electron mediator (Shin et al. 2003). Light-dependent CO_2 conversion to CO with CODH, an interesting approach to exploit light as the energy source, was also broadly studied by Armstrong's group (Woolerton et al. (2010); Chaudhary et al. (2012); Bachmeier et al. (2013)). However, the necessity of anaerobic conditions for these conversions represents an important drawback for industrial-scale perspectives.

FDHs form a class of enzymes of foremost interest for this research project. They catalyze the reversible conversion of formate to CO_2, described in detail in section I. 3. The formate obtained by this CO_2 reduction can be further used as an energy source or as a chemical synthon. FDH immobilization on various supports has been extensively studied for *in vitro* reduction of CO_2 to formate and was reviewed in detail by Amao (2018). This reaction mostly uses NADH as the electron carrier, but this cofactor has the disadvantages of being expensive and unstable, thus the direct electron transfer using an electrode or a synthetic cofactor such as MV was explored to circumvent these limitations. The efficient CO_2 reduction to formate reported for diverse FDHs from sources such as *Desulfovibrio vulgaris*

Hildenborough, *Syntrophobacter fumaroxidans* or *E. coli* was demonstrated using immobilized proteins (Alvarez-Malmagro et al. (2021); Reda et al. (2008); Bassegoda et al. (2014)). This method allows to bypass the need for a cofactor using direct electron transfer from the electrode to the immobilized protein. Still, these reductions were performed in strictly anaerobic conditions, demonstrating the difficulty of working with these enzymatic complexes under aerobic conditions.

Yu et al. (2019) designed an efficient and aerobic biocatalytic process by coupling the FDH from *Cupriavidus necator* for CO_2 reduction to formate (further described in the section I. 3. 2) and a glucose dehydrogenase (GDH) from *Pseudomonas sp.* for NADH regeneration. Thus, if the formate synthesized by the FDH is removed from the buffer to favor the desired reaction at the maximum rate, this system can become industrially interesting.

Formate, as being the product of CO_2 reduction by FDHs and thus important for this research project, the following paragraph will provide more details of the various applications of formate.

4. Formate: a renewable resource efficiently obtained from CO_2 reduction

Formic acid (HCOOH) is a promising compound for connecting physiochemical and biological processes, offering a sustainable solution to produce value-added chemicals from CO_2. Formic acid can be generated through various methods, such as electrochemical reduction of CO_2, photo-reduction of CO_2, hydrogenation of CO_2, selective oxidation of biomass, partial oxidation of natural gas and hydration of syngas (Yishai et al. 2016). Of special interest is the efficient electrochemical production of formate (> 40 % energetic efficiency, Jouny, Luc, and Jiao (2018)) from CO_2 *via* a two-electron reduction. This compound has advantageous properties, being non-toxic, safe, and highly soluble in water, unlike carbon monoxide (CO), which can also be obtained from direct CO_2 conversion, as described in the previous sections (I. 2. 2).

Hydrogen is seen as a potential "clean" energy carrier as its combustion only produces water. In a hydrogen-based economy, the demand for H_2 storage is high considering the difficulty of its storage and its relatively low energy density (2.5 $W.h.L^{-1}$) compared to gasoline (8.07 $W.h.L^{-1}$). It is in this context that formic acid was largely discussed as a potential hydrogen-storage solution in many reviews, showing the interest around this compound (Joó (2008); Mellmann et al. (2016); Singh, Singh, and Kumar (2016)).

Furthermore, formate is one of the simplest organic compounds that can be used as a growth substrate by living cells, such as methylotrophic organisms (Chistoserdova 2011). Methylotrophic organisms, like *Methylorubrum extorquens*, are natural formatotrophs (Crowther, Kosály, and Lidstrom 2008) that have been studied for their potential as biotechnological chassis strains. However, the level of knowledge

about their metabolism, its regulation and the availability of appropriate metabolic engineering tools is low (Singh et al. 2022). Hence, the implementation of synthetic formatotrophy in biotechnological platform organisms such as *Escherichia coli*, has been undertaken (see paragraph *E. coli* as a chassis for synthetic autotrophy I. 2. 5)

Despite the promising and significant progress achieved for the development of sustainable and cost-effective production of formate, some challenges still need to be tackled. For example, the sustainable and efficient electrochemical reduction of CO_2 to formate relies mainly on electrodes whose cost and durability make the production process not economically competitive (Martín, Larrazábal, and Pérez-Ramírez 2015). Therefore, there is interest for microbial strains that can directly use CO_2 as a carbon source without the need for a previous *ex vivo* CO_2 reduction step.

While the above-mentioned technologies are very promising, CO_2 capture is already performed by living organisms and research projects are also explored to take advantage of these pathways. The comprehension of these natural processes, which are described in the following section, can also offer opportunities for biotechnological applications.

3. Natural CO_2 assimilation pathways

"Autotrophy is the ability of an organism to synthesize all cell carbon from CO_2 alone" (Alber 2019).

In the biosphere, carbon fixation, especially CO_2 fixation constitutes the primary biochemical process for the production of organic compounds through the conversion of inorganic carbon into biomass and supplies organic carbon sources essential for all living organisms (Bar-Even, Noor, and Milo 2011).

In nature, the reductive pentose phosphate (Calvin-Bassham-Benson or CBB) cycle is the most prevalent carbon assimilation pathway. In addition to this pathway, five alternative routes for carbon fixation have been identified: the reductive tricarboxylic acid (rTCA) pathway, the reductive acetyl-CoA pathway (Wood-Ljungdahl pathway), the dicarboxylate/4-hydroxybutyrate cycle, the 3-hydroxypropionate/4-hydroxybutyrate pathway and the 3-hydroxypropionate bicycle (Bar-Even, Noor, and Milo 2011). The last two cycles exclusively assimilate bicarbonate, HCO_3^-, molecules. Along with these six carbon-fixation routes, three additional natural alternatives have recently been described: the reductive hexulose-phosphate pathway, the natural reductive glycine cycle, and the reverse oxidative TCA cycle (Santos Correa et al. 2023).

1. The reductive pentose phosphate cycle

The reductive pentose phosphate cycle is the first carbon fixation cycle to have ever been described (Calvin and Benson 1948), and stands as the primary carbon fixation pathway accounting for approximately 90 % of all CO_2 fixation (Gong et al. 2018). This cycle is present in plants, algae, chemolithoautotrophic bacteria, photoautotrophic microorganisms such as cyanobacteria and aerobic proteobacteria (Berg 2011).

Through this cycle, the CO_2 is fixed by the ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO), which is the most abundant enzyme on earth (Cleland et al. 1998). Following a cascade of reactions (Figure 2), three molecules of CO_2 are converted into one molecule of glyceraldehyde 3-phosphate (G3P). Subsequently, two molecules of G3P are condensed to form glucose. The energy needed for this cycle is derived from NAD(P)H and ATP generated by the light-dependent reactions of photosynthesis, utilizing photon energy.

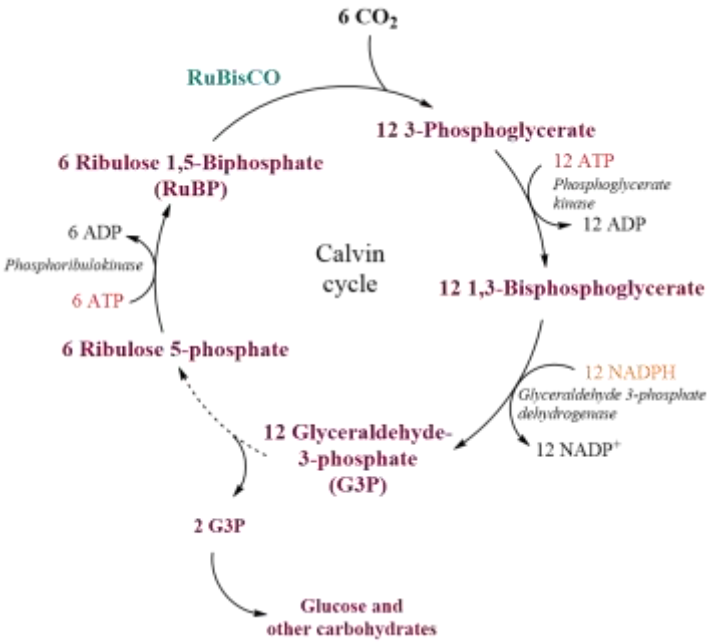


Figure 2. Schematic representation of the Calvin cycle.

2. The reductive tricarboxylic acid pathway

The rTCA pathway was first described in a green sulfur bacterium, *Chlorobium limicola* by Evans, Buchanan, and Arnon (1966). It implicates O_2-sensitive enzymes containing iron-sulfur clusters and involving free radical intermediates, exclusively found in anaerobic bacteria or microaerophiles (Thauer 2007). As represented in Figure 3, this cycle operates in the opposite direction of the oxidative TCA cycle (Krebs cycle) and two molecules of CO_2 are used to form its main product, acetyl-CoA. Acetyl-CoA is then converted to other carbon intermediates: pyruvate, phosphoenolpyruvate, oxaloacetate and 2-

oxoglutarate, using ferredoxin and NAD(P)H as electron donors (Berg 2011). Only two molecules of ATP are required to form one molecule of pyruvate.

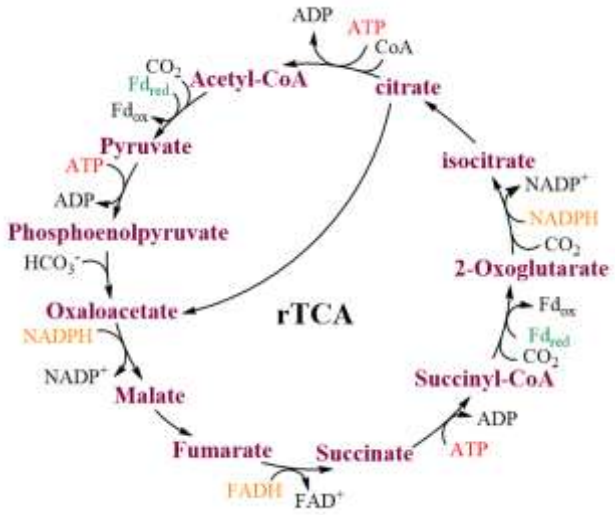


Figure 3. Schematic representation of the reductive tricarboxylic acid cycle.

3. The reductive acetyl-CoA pathway

The reductive acetyl-CoA pathway, also known as the Wood-Ljungdahl pathway was described by Ljungdahl in 1986. It is a linear route that reduces two molecules of CO_2 to one molecule of acetyl-CoA in anaerobic microorganisms. As represented in Figure 4, this pathway possesses two branches. The methyl branch where one molecule of CO_2 goes through a catalyzed six-electron reduction process starting with FDH to produce formate which is further reduced to a methyl group. The carbonyl branch catalyzes the reduction of CO_2 to CO, which is then condensed with the methyl group from the methyl branch and CoA to produce acetyl-CoA (Ljungdhal 1986).

This pathway is primarily found in acetogenic and methanogenic bacteria, that live very close to the thermodynamic limit, and this pathway plays not only the role of CO_2 fixation but also energy conservation (Ragsdale and Pierce (2008); Thauer et al. (2008)). During autotrophic growth, acetogens produce acetate as the end product and methanogens generate methane from acetate, H_2 and CO_2 using unique coenzymes such as methanopterin and methanofuran (Ljungdhal 1986).

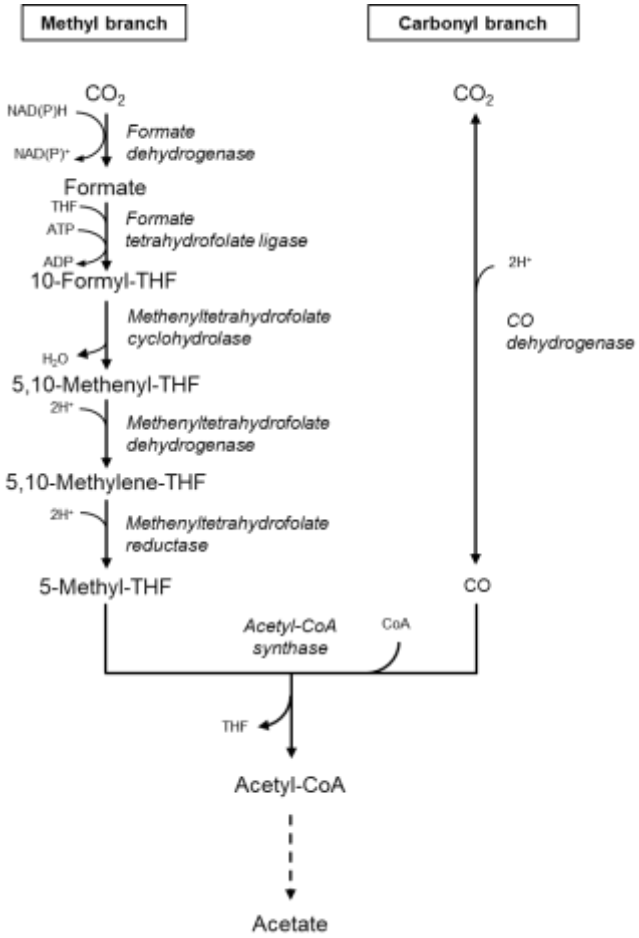


Figure 4. Reductive acetyl-CoA pathway from acetogens. THF = tetrahydrofolate

4. The 3-hydroxypropionate bicycle

The 3-hydroxypropinate bicycle was initially described in the thermophilic photosynthetic bacterium *Chloroflexus aurantiacus* (Holo 1989). This cycle does not involve oxygen-sensitive enzymes and is also found in related *Chloroflexi* bacteria (Bar-Even, Noor, and Milo 2011). As represented in Figure 5, the first cycle fixes two molecules of bicarbonate to form one molecule of glyoxylate. The second cycle starts with the assimilation of glyoxylate and catalyzes its five-step conversion to pyruvate, which will enter the central metabolism, and acetyl-CoA which will serve as a precursor for the next cycle (Fuchs 2011). In total, one molecule of pyruvate is formed from three molecules of bicarbonate fixed by the propionyl-CoA and the acetyl-CoA carboxylases.

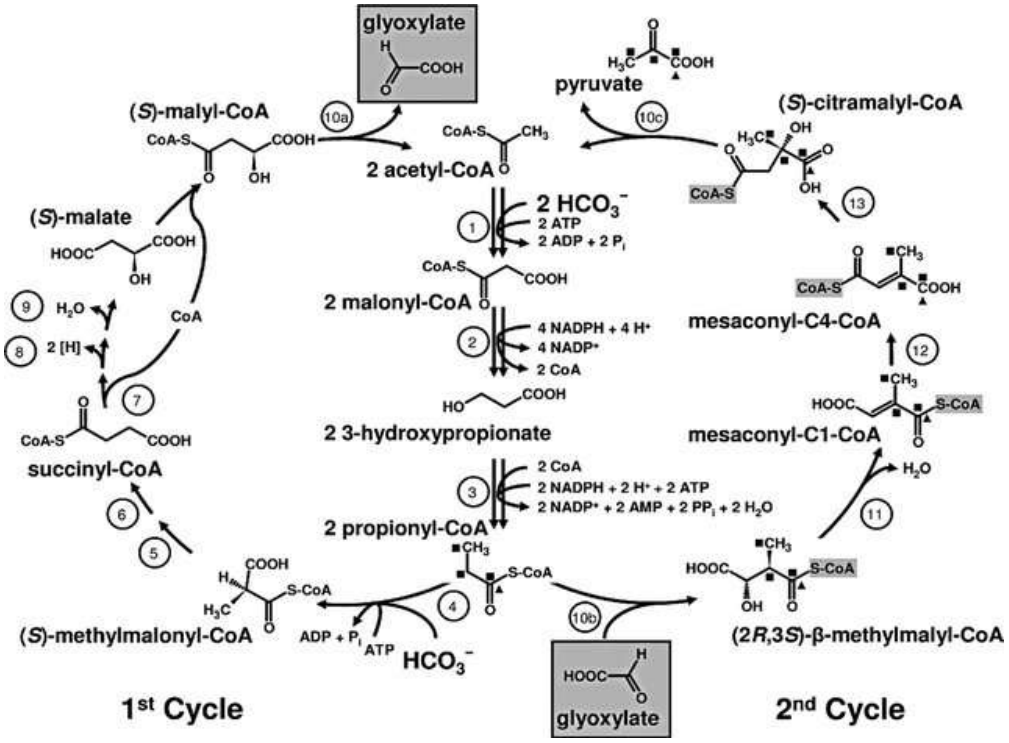


Figure 5. Schematic representation of the 3-Hydroxypropionate bicycle.[1] Acetyl-CoA carboxylase, [2] malonyl-CoA reductase, [3] propionyl-CoA synthase, [4] propionyl-CoA carboxylase, [5] methylmalonyl-CoA epimerase, [6] methylmalonyl-CoA mutase, [7] succinyl-CoA:(S)-malate-CoA transferase, [8] succinate dehydrogenase, [9] fumarate hydratase, [10 a, b, c] (S)-malyl-CoA/-methylmalyl-CoA/(S)-citramalyl-CoA (MMC) lyase, [11] mesaconyl-C1-CoA hydratase (-methylmalyl-CoA dehydratase), [12] mesaconyl-CoA C1-C4 CoA transferase, [13] mesaconyl-C4-CoA hydratase. From Zarzycki et al. (2009).

5. The 4-hydroxybutyrate cycles

The dicarboxylate/4-hydroxybutyrate (DC/HB) cycle and the 3-hydroxypropionate/4-hydroxybutyrate (HP/HB) cycle found in archaea share common reactions and are producing the same final product, acetyl-CoA (Berg, Ramos-Vera, et al. 2010). As represented in Figure 6, the two cycles follow the same cascade of enzymatic reactions from succinyl-CoA to acetyl-CoA. However, the carbon-fixing enzyme differs between the cycles. In the DC/HB cycle, the pyruvate synthase and the phosphoenolpyruvate carboxylase are fixing CO_2 and HCO_3^- respectively. By contrast, in the HP/HB cycle, acetyl-CoA/propionyl-CoA carboxylases are fixing two molecules of HCO_3^-. Another important difference between the two cycles is their sensitivity to oxygen. The enzymes from HP/HB can tolerate oxygen while the pyruvate synthase involving ferredoxin from the DC/HB cycle cannot, which limits this cycle to anaerobic conditions (Berg 2011).

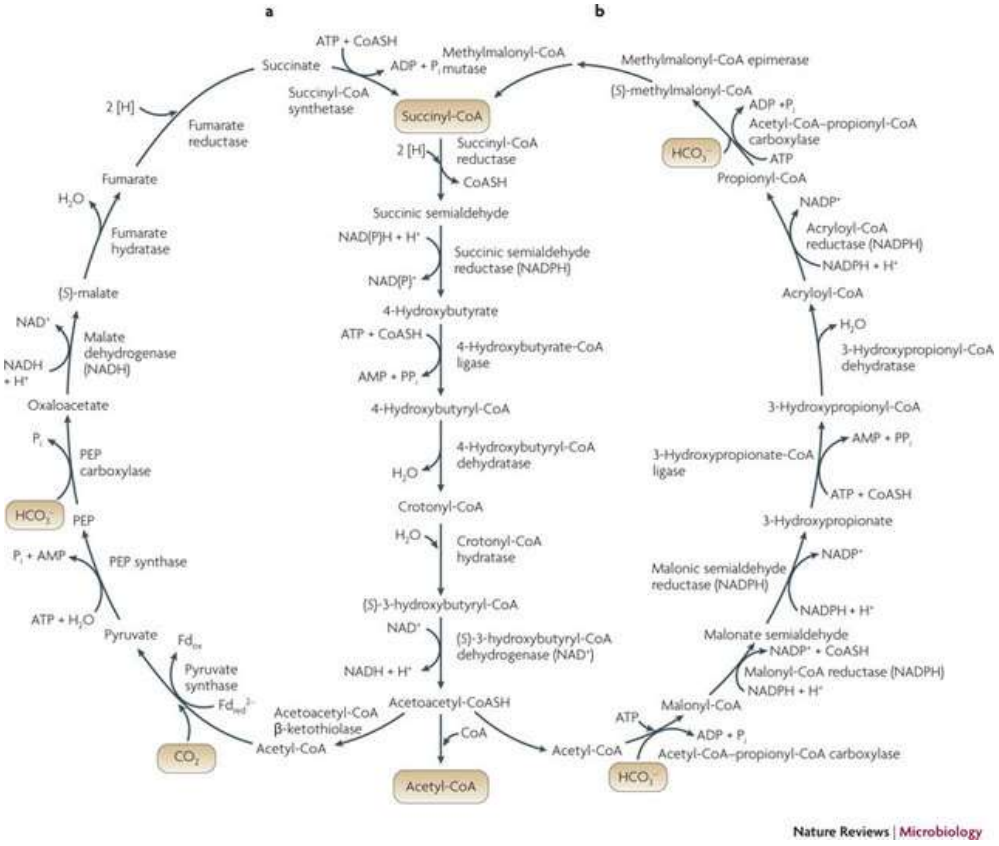


Figure 6. The dicarboxylate/4-hydroxybutyrate cycle (a) and the 3-hydroxypropionate/4-hydroxybutyrate (b). From Berg, Kockelkorn, et al. (2010).

This overview of the main natural CO_2 assimilation pathways selected during biological evolution gives insights and points to biotechnological opportunities. One possibility would be to use these organisms for autotrophic production of valuable compounds. Alternatively, synthetic pathways could be designed based on these reactions. The scientific community exploited the gained knowledge to pursue both approaches, the design of synthetic assimilation pathways and the transfer of existing ones into better described and easier to manipulate chassis strains.

6. Metabolic engineering of autotrophic strains for bioproduction

A great effort has been put into engineering natural autotrophic microbes, mostly cyanobacteria and *C. necator* that are using the Calvin cycle to produce valuable products (Angermayr, Gorchs Rovira, and Hellingwerf 2015) through implemented heterologous biosynthetic routes. Some examples are described below.

Cupriavidus necator, a gram-negative bacterium and facultative chemolithotroph, capable of growing on H_2 and CO_2 in the absence of other organic compounds, is an interesting chassis strain for sustainable bioproduction. This organism uses the reductive pentose phosphate pathway (Calvin cycle, described above, I. 2. 3. 1) to fix CO_2 and a soluble hydrogenase (see details of this hydrogenase in section I. 5. 3)

to produce NADH from H_2. Additionally, O_2 or possibly NO_3^- plays the role of final electron acceptors (Pohlmann et al. 2006). Under atmospheric CO_2 concentrations, *C. necator* grows autotrophically with a doubling time of approximately 20 h. The metabolic engineering of this strain, under autotrophic growth mode, enabled the production of a wide range of chemicals of industrial interest, such as methyl ketones, polyhydroxyalkanoates, lipo-chitooligosaccharides, alka(e)nes (Müller et al. (2013); Nangle et al. (2020); Crépin, Lombard, and Guillouet (2016); Panich, Fong, and Singer (2021)).

Cyanobacteria have also been considered as potential autotrophic production strains from CO_2 and light. The development of genetic tools for the engineering of their metabolism enabled strain constructions for the autotrophic production of a large panel of chemicals (extensively reviewed by Oliver and Atsumi (2014) and Angermayr, Gorchs Rovira, and Hellingwerf (2015)). This includes sucrose, butanol, fatty acids, among others (Ducat et al. (2012); Lan and Liao (2012); Liu, Sheng, and Curtiss III (2011); Gao et al. (2012)).

Furthermore, other metabolic engineering processes for autotrophic bioproduction were developed in acetogenic strains and the successful production of diverse compounds such as acetate, ethanol, butyrate, lactate and formate was demonstrated (reviewed by Liew et al. (2016)). LanzaTech developed commercially viable processes of bioproductions using acetogens.

In general, the production yields obtained for commodities produced by the above mentioned strains are still low, which is due to multiple factors, like the low energy efficiency of the Calvin cycle, sub-optimal carbon partitioning and low cell-densities in photosynthetic reactors (Ducat, Way, and Silver (2011); Cotton et al. (2015); Gomaa, Al-Haj, and Abed (2016)). Research to better understand the regulation of the central metabolism is needed to optimize carbon flow in these autotrophic production chassis.

Alternatively, natural CO_2 fixing pathways can be implemented in well-studied heterotrophic platform organisms to entail these strains with enlarged trophic modes.

4. Engineering CO_2 fixing pathways in heterotrophs

1. Implementation of natural CO_2-reducing pathways into heterotrophic chassis strains

Being the most widespread carbon-fixation route, the Calvin cycle, despite its low energy efficiency, is one of the most studied option for implementation into heterotrophic host strains to achieve autotrophic growth.

The CBB cycle has been successfully implemented in *E. coli* by the heterologous expression of two enzymes, the Rubisco that fixes CO_2 and the phosphoribulokinase *(prk)* for ribulose-bisphosphate (RuBP)

regeneration. These are the only two enzymes necessary for the cycle functionality not already present in *E. coli* (Zhuang and Li (2013); Gong et al. (2015); Antonovsky et al. (2016); Gleizer et al. (2019). This carbon-fixation route was also successfully implemented in the yeast *Saccharomyces cerevisiae* (Guadalupe-Medina et al. (2013); Li et al. (2017)).

In an attempt to engineer a heterotrophic chassis for direct CO_2 assimilation into biomass, Antonovsky et al. (2016) expressed the Calvin cycle in *E. coli* and were able to demonstrate the synthesis of all sugar-derived metabolic building blocks from CO_2. As already stated above, only two enzymes are theoretically necessary for the execution of the Calvin cycle in *E. coli* cells: the Rubisco and the phosphoribulokinase. In addition, a carbonic anhydrase that catalyzes the interconversion of CO_2 to HCO_3^- was overexpressed. All the other reactions necessary for the rest of the sugar synthesis using this pathway are already present in *E. coli*. The deletion of the phosphoglycerate mutase genes (*gpmA* and *gpmM*) allowed to decouple the glycolytic and gluconeogenic fluxes, hence creating a system where CO_2 can be the source of sugar-carbons through the Calvin cycle, the energy requirement and additional metabolites were derived from the externally provided pyruvate. This strain construction resulted in a context where CO_2 was supplying one-third of the total carbon needed for growth. As the strain was not able to grow initially in these conditions, researchers performed an adaptation using a chemostat with a limiting xylose supply which fed the Calvin cycle but made it dependent on the use of Rubisco. After around 150 generations, they obtained cells capable of growing on a minimal medium supplemented with pyruvate and CO_2 (25 % in air). The doubling time of the resulting strains was around 6 h and the maximum OD_{600} of 0.5-0.6. The resulting growth mode was named "hemiautotrophic" by the authors.

In a subsequent work, the same research group reported an improvement in the strain described above (Gleizer et al. 2019). In this engineered strain, formate served as the electron donor for CO_2 fixation, which was achieved through the heterologous expression of the NAD^+-coupled formate dehydrogenase (FDH) from *Pseudomonas sp.* 101. For the carbon source, researchers heterologously expressed the same enzymes for the implementation of synthetic Calvin cycle, *i.e.* Rubisco and phosphoribulokinase, and carbonic anhydrase. In this genetic context, all the carbon necessary for biomass was supplied *via* the Calvin cycle. Initially, the constructed strain did not exhibit immediate growth on the minimal medium supplemented with formate (30 mM) and CO_2 (10 % in air). After continuous cultivation in a chemostat under selective conditions and a gradual reduction of sugar in the feeding medium, evolved clones emerged with the ability to grow autotrophically using CO_2 as the sole carbon source and formate as the energy source. This study demonstrated that a combination of rational metabolic engineering and evolution experiments under selective conditions can successfully rewire the central metabolism of the versatile strain, *E. coli*. However, while this work is a significant achievement, the strain's doubling time of 18 h and its maximum OD_{600} of 0.2 indicate that this synthetic autotrophic strain still has a long way

to go before it can be considered as a viable biotechnological chassis. In addition, the use of formate oxidation as energy module results in no net CO_2 consumption as CO_2 is formed for energy supply. This underlines the importance of the development of other energy modules such as the ones described in the *Energy generation systems* paragraph (I. 5) to circumvent this issue. The resulting strains would be able to grow without any organic carbon substrate. Still, the performance of these strains would remain limited due to the low activity of the Rubisco.

The Calvin cycle is not the best choice for being transferred to non-autotrophic platform strains due to its low energy efficiency, the low catalytic activity exhibited by RuBisCo ($k_{cat} \approx 1\text{-}10\ s^{-1}$, Sage (2002)) and the enzyme's intrinsic affinity to oxygen leading to the futile cycle of photorespiration (Sage 2002)).

Mattozzi et al. (2013) successfully expressed the 3-hydroxpropionate bicycle, which is not affected by O_2, in *E. coli* and obtained sufficient activity to complement the introduced auxotrophy to diaminopimelic acid and to participate in cell growth. To our knowledge, no further improvement of this strain towards the implementation of an autotrophic growth mode was reported. In another genetic context, the extremophile archaeon *Pyrococcus furiosus* was engineered to integrate a part of the HP/HB cycle to produce 3-hydroxypropionic acid, an industrial chemical building block, with H_2 and CO_2 as substrates (Keller et al. 2013).

Unlike the Calvin cycle, the other natural CO_2 assimilation pathways described in the precedent section were not extensively studied for heterologous expression in a heterotroph host, probably due to their complexity.

2. Synthetic pathways design

As exposed in the previous sections, CO_2-fixing enzymes like the RuBisCO can have low activity; the attempts to improve their catalytic properties led to limited improvements (Genkov et al. (2010); Ishikawa et al. (2011); Cai et al. (2014)). Others, like most of the metal-dependent FDHs are sensitive to oxygen, making them not suitable for large-scale biotechnological use.

Bar-Even et al. (2010) performed a survey of carbon fixing enzymes and identified the phosphoenolpyruvate (PEP) carboxylase that catalyzes the addition of HCO_3^- to one PEP molecule resulting in oxaloacetate and the pyruvate kinase as the most kinetically efficient catalysts. A computational analysis of potential pathways employing these enzymes resulted in the design of the malonyl-CoA-oxaloacetate-glyoxylate (MOG) cycles family sharing the same pathway structure. The MOG pathways have the highest calculated specific activities (two to three times faster than the CBB cycle). This study was performed within the context of photosynthetic organisms, yet the researchers underlined the practical challenges that the implementation of a MOG pathway in cell's metabolism

represents, despite the theoretical demonstration pointing to improved efficiency. Consequently, to our knowledge, there is no report of a successful implementation of such synthetic pathways in photoautotrophs. This suggests that significant difficulties likely emerge when metabolic routes designed on *in silico* data are transferred into living organisms.

The difficulties encountered upon the implementation of efficient and direct CO_2 assimilation in heterotroph chassis, led to the exploration of hybrid *in vitro/in vivo* pathways with a combination of electrochemical and biological processes. As described in the Electrochemical conversion paragraph (I. 2. 2. 3), direct reduction of CO_2 can be efficiently performed to produce carbon monoxide (CO) or formate (HCOOH). Due to the toxicity and high flammability of CO, formate with its high solubility and low toxicity has been favored (Yishai et al. 2016). The first proof-of-concept study, from Li et al. (2012), integrates the electrochemical reduction of CO_2 to formate with the re-oxidation of formate by an engineered *Cupriavidus necator* (formerly known as *Ralstonia eutropha*) yielding the NADH necessary for producing the biofuels isobutanol and 3-methyl-1-butanol. Another setup, inspired by the previous one, took advantage of the ability of *E. coli* to grow under anaerobic conditions, thus reducing the exposition of the cells to active oxygen species that are produced in the cathodic chamber during CO_2 reduction known to inhibit bacterial growth. The electrochemically produced formate was reused by an engineered formate-dependent *E. coli* strain for growth (Tashiro et al. 2018). Such hybrid systems for CO_2 fixation can be an alternative to circumvent the low carbon fixation rate of phototrophs.

Very recently, a study in *Pseudomonas putida* demonstrated "partial" autotrophic growth by the implementation of the reductive glycine pathway (see detailed description of the pathway in section I. 2. 5. 2. 1) and the activity of an endogenous NAD^+-dependent FDH in a serine auxotrophic strain (Bruinsma et al. 2023). This work is discussed in more detail in the Results and discussion chapter (III. 2) as the strategy employed is very similar to the one explored during the PhD work.

While these above-mentioned pathways bring interesting alternatives, direct CO_2 reduction *in vivo* remains an open challenge due to the very few options available. This project takes up this research challenge.

5. *E. coli* as a chassis for synthetic autotrophy

As described in the previous section, many attempts were performed for CO_2 assimilation by microorganisms and in particular, *E. coli*. Undeniably, this strain is seen as a promising candidate for such modifications of its central metabolism because of its flexibility and biotechnological potential (see section I. 4). However, considering the high level of difficulty, the very few options available for direct assimilation of CO_2 and the limitations of the previously explored pathways to support autotrophic

growth in a heterotrophic biotechnological strain, alternative growth modes have been developed, including synthetic formatotrophy and methylotrophy. The latter growth mode will be briefly overviewed below, whereas synthetic formatotrophy, an important aspect of this PhD work, will be explored in more detail.

1. Methylotrophy

Methylotrophy refers to the assimilation of methane (CH_4) or methanol (CH_3OH) as carbon and energy sources. CH_4 is the main component of natural gas representing 80–95 % (v/v), thus is easily available (Fei et al. 2014). Methanol can be efficiently produced from H_2 and CO_2 by a two-step process of first H_2 production through water electrolysis, which further reacts with CO_2 to produce CH_3OH (Szima and Cormos 2018). Direct assimilation of methane or methanol by anaerobic organisms exists naturally but the theoretical yield, the compound spectrum and the genetic toolkits available for these types of strains are rather limited (Haynes and Gonzalez 2014). However, the exploration of these organisms expanded the knowledge of their metabolism and the biotechnological possibilities for heterologous expression of key enzymes in well-described hosts (Schrader et al. 2009). In addition, methanol, being more reduced than glucose, is a favorable substrate for improved production yield (Whitaker et al. 2015). Thus, the implementation of this growth mode was explored in non-methylotrophic organisms, including *E. coli*. Several studies were published on this question and a few are described in this section.

Synthetic methylotrophy was established in an *E. coli* strain engineered by Kim et al. (2020). This strain was initially modified to implement formatotrophy using the reductive glycine pathway (a more detailed description of this metabolic engineering part is available in the following section, I. 2. 5. 2. 1). An additional enzyme was expressed to implement methylotrophy, the methanol dehydrogenase from *Bacillus methanolicus* that converts methanol to formaldehyde. Formaldehyde is oxidized to formate by the endogenous glutathione system from *E. coli* and formate is assimilated through the RGP pathway. However, the growth rate obtained for this strain was very low, 54 h and a maximum OD_{600} of 0.2.

A report of Chen et al. (2020) described the implementation of synthetic methylotrophy in an engineered *E. coli* strain capable of growing on methanol as sole carbon source following a different pathway from the previously cited study. The ribulose monophosphate pathway, a typical pathway operating in methylotrophic strains, was implemented in *E. coli*. The genome of this strain was highly engineered with the expression of eight heterologous genes integrated into the chromosome for better genetic stability and multiple gene deletions to ensure methanol dependence of growth. To maximize the flexibility of their strain, the authors co-expressed isofunctional enzymes. They used a stepwise protocol by initially implementing methanol dependence that necessitated an additional carbon source, xylose, for the regeneration of the essential co-substrate Ru5P. Then, after a short-term laboratory evolution of 80

generations, an isolate capable of reaching a final OD_{600} of 1.0 was obtained. The subsequent implementation of synthetic methylotrophy being unsuccessful, a metabolic modeling was performed to identify hurdles in their engineered strain, which gave insights for further fine-tuning of gene expression. Eventually, after highly progressive adaptation steps, a culture that could grow on methanol only was obtained. Very interestingly, whole genome sequencing of isolates revealed a mixed population of two strains, one being methylotrophy, the other one not. The methylotrophic isolate exhibited a favorable growth rate of 8.5 h, close to natural methylotrophic strains such as *Methylorubrum extorquens* AM1, and a maximum OD_{600} of 2.0.

2. Formatotrophy: a promising option

In the Formate: a renewable resource efficiently obtained from CO_2 reduction section (I. 2. 4), the advantages of the use of formate as a reduced compound directly derived from CO_2 were described. In this paragraph, an overview of the work undertaken to implement formatotrophic growth in the heterotrophic chassis strain *E. coli* is provided. As mentioned earlier, some methylotrophic organisms such as *M. extorquens* AM1, can grow on formate as their sole carbon and energy source. However, these organisms are challenging to manipulate which incented the implementation of formatotrophy in organisms such as *E. coli* or *S. cerevisiae* (Bar-Even et al. 2013), as most bioproduction pathways are designed for these model organisms.

1. Focus on the reductive glycine pathway

The work of Arren Bar-Even has been particularly influential in providing insights into relevant pathways and approaches for formate assimilation in microorganisms. With a view to building an electricity-dependent system for formatotrophic microorganisms, Bar-Even et al. (2013) realized a computational analysis to determine the most favorable pathways for such growth mode and found that a promising option was the reductive glycine pathway (RGP, Figure 7). The RGP represents the most efficient route, in terms of resource consumption and biomass production, to assimilate formate in aerobic conditions (Bar-Even et al. 2013). In this pathway, the glycine cleavage system (GCS) needs to operate in the reverse direction for glycine biosynthesis, which represented the main challenge when this pathway was first described. Most importantly, this linear route can operate aerobically since no O_2-sensitive enzymes are involved. For the synthesis of one molecule of G3P (C3), the Calvin cycle will require nine ATP and six NAD(P)H molecules while the RGP pathway only uses two ATP and three NAD(P)H for the synthesis of one pyruvate (C3), making this cycle a highly efficient option.

The first step of this pathway is formate activation through the ligation to one molecule of tetrahydrofolate (THF) giving 10-formyl-THF, the only reaction not catalyzed by an *E. coli* enzyme. 10-

formyl-THF is then further reduced to 5,10-methylene-THF and condensed with CO_2 and NH_3 to form glycine through the GCS. Then, the glycine molecule is condensed with another 5,10-methylene-THF to generate serine and then pyruvate, a central biomass precursor of carbon metabolism.

More recently, two organisms have been described to use this pathway naturally. The first organism that was proposed with this metabolism is the uncultivated bacterium *Candidatus* Phosphitivorax anaerolimi strain Phox-21. This organism would couple phosphite oxidation and CO_2 reduction to formate, which would be assimilated through the RGP (Figueroa et al. 2018). Yet this study was mainly based on metagenomic data analysis and complex cultures but no experimental demonstration using a pure culture of this microorganism was provided. The other microorganism is *Desulfovibrio desulfuricans*, which is capable of assimilating CO_2 as its carbon source via the RGP with hydrogen and sulfate as energy sources (Sánchez-Andrea et al. 2020).

To implement synthetic formatotrophy in *E. coli*, two strategies were followed, relying either on directed evolution (Döring et al. (2018); Delmas et al. (2022), Genoscope laboratory) or on rational design (Kim et al. (2020), Bar-Even laboratory). Both strategies are based on a modular pathway implementation using selection strains auxotrophic for pathway intermediates, with formate supplying an ever greater part of biomass carbon. While the Bar-Even laboratory either replaced the key pathway enzymes with heterologous homologs more active in the reductive sense or modulated the expression of native genes, the Genoscope team used continuous culture evolution to select for mutated genes coding for variants of native enzymes favoring catalysis in the reductive sense. A restricted number of mutations were selected, affecting either the expression of the heterologous gene *ftfL* coding for the formate THF ligase, or modulating the activity of the methenyl-THF cyclohydrolase/dehydrogenase (gene *folD*) or the lipoamide dehydrogenase (gene *lpd*). The selection strains used for RGP implementation, characterized by their respective formate requirement for growth, offered the possibility to select for CO_2 reduction catalyzed by the FDH studied in this work. Their exact genotype will be detailed in the results section (III. 2) of this manuscript (Döring et al. (2018); Delmas et al. (2022)).

The RGP was successfully implemented in *Cupriavidus necator* and *Saccharomyces cerevisiae* as well (Claassens et al. (2020); Gonzalez de la Cruz et al. (2019)).

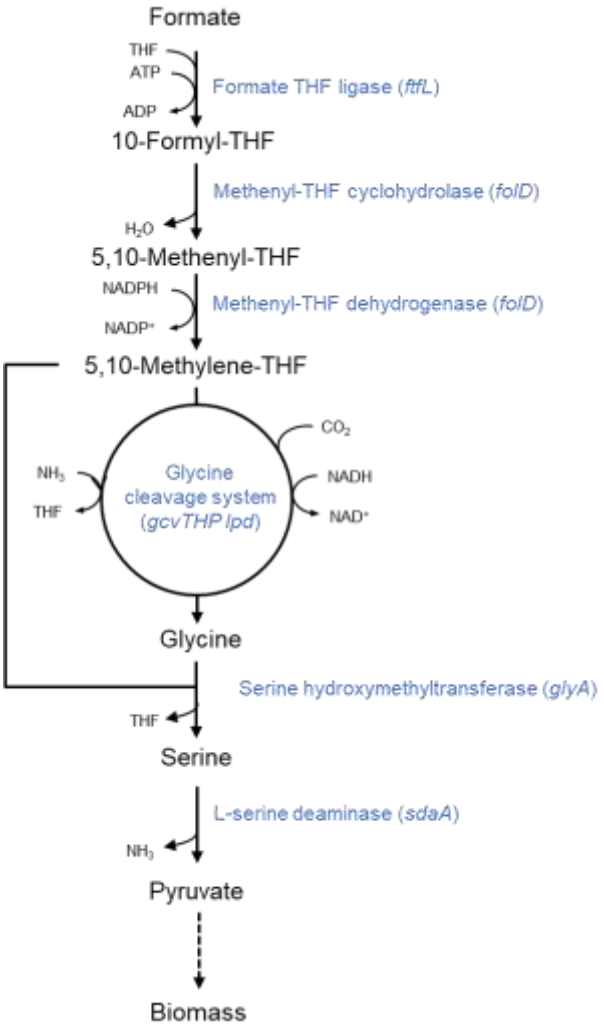


Figure 7. Scheme of the reductive glycine pathway (RGP).

2. *Metabolic engineering work around formatotrophy*

Along with the RGP, other innovative C1 assimilation routes involving formic acid as sole or in combination with other carbon sources were explored in *E. coli*, the major works in this field are reviewed in this paragraph (summarized in Table 1).

Zelcbuch et al. (2016) described an innovative construction by using the pyruvate formate-lyase (PFL) from *E. coli* to create a strain able to grow on acetate and formate. PFL is reported as an oxygen-sensitive enzyme, functioning exclusively under anaerobic conditions that catalyzes the reversible reaction described by Equation 2 (Knappe et al. 1974). During anaerobic growth, the ATP availability is lower compared to aerobic conditions, thus limiting growth. By catalyzing the transformation of pyruvate to formate and acetyl-coA (without the need for NAD^+), the PFL complex becomes crucial. The acetyl-CoA produced by the PFL activity is then further metabolized to acetate and ethanol which generates one

additional ATP per glucose molecule, resulting in a total of three ATP molecules (Figure 9), giving an important advantage for growth compared to a bacteria that does not possess this enzyme (Hasona et al. 2004).

Equation 2. Pyruvate formate-lyase reaction.

$$\text{Formate} + \text{Acetyl-CoA} \leftrightarrow \text{Pyruvate} + \text{CoA}$$

The goal of the work around this complex was to explore the reversibility of this reaction *in vivo* as it was demonstrated *in vitro* thus being a potential route for formate assimilation. In their genetic context, *aceA* encoding isocitrate lyase, necessary for the assimilation of acetate through the glyoxylate shunt, is disrupted (Figure 9) as well as *pflB*, the gene coding for PFL. In this Δ*pflB* Δ*aceA* double mutant, a plasmid for overexpression of the *plfB* gene along with the gene coding for the activation enzyme *pflA* was transformed. In this strain, the presence of the plasmid restored growth under anaerobic conditions on acetate (15 mM), formate (15 mM) and nitrate as electron acceptor. Thus, the *in vivo* reversibility of the PFL complex that catalyzes the condensation of acetyl-CoA and formate was demonstrated and showed a pathway for formate assimilation in *E. coli* (Zelcbuch et al. 2016). Although the growth was successfully restored, it remained very low in terms of maximum OD_{600} (0.18 after 50 hours of culture) and required anaerobic conditions.

In the work carried out by Kim et al. (2020), the implementation of the RGP was performed in *E. coli* allowing direct assimilation of formate and CO_2. For this purpose, they performed the expression of four heterologous enzymes from *M. extorquens* for the first three reactions of the pathway as well as the L-serine deaminase (Figure 7). In addition, the FDH from *Pseudomonas sp.* for energy and reducing power regeneration was used and gene disruptions to redirect the carbon fluxes were performed. The implementation of the entire pathway was performed sequentially with validation at each step of the different modules implemented. The formate dependency of the strain was gradually increased until it reached 100 % for biomass and energy production. The final constructed strain had a doubling time between 65 to 80 hours. After a short-term evolution in fed-batch mode in tubes, the doubling time dropped under 10 h and an OD_{600} maximum of 1 with 100 mM formate and 10 % CO_2 was obtained.

Another *E. coli* engineered strain was shown to grow aerobically on formate and CO_2 exclusively by implementing the RGP with a similar approach to the previous one, the first three reactions of the pathway being performed by heterologous genes. Along with this, two FDH enzymes, one from *Candida boidinii* and one from *Arabidopsis thaliana* are employed for NADH and NADPH regeneration respectively (Bang et al. 2020). In this engineered system, formate served as the energy source *via* FDHs, and CO_2 served as the carbon source *via* the reductive glycine pathway (RGP). The researchers stated that two separate experiments were conducted, resulting in significantly different generation times of

65.9 hours and 250.6 hours. Additionally, the variability in the maximum OD_{600} reached for their modified strain ranged from 7 to 11. In the experiments presented, the variability observed was very important, which requires further explanation.

These widely scattered values between published works may be influenced, at least in part, by the cultivation conditions, making direct comparisons challenging.

Table 1. Overview of engineered *E. coli* strains for growth on formate. Adapted and completed from Bang et al. (2021).

Strain	Substrate	Doubling time (h)	Formate and/or CO_2 assimilation pathways	Description	Reference
E. coli	Formate and acetate	4.7 ± 0.1	Pyruvate formate-lyase (PFL)	• Constructed by rational design • Demonstration of the reversibility of PFL activity *in vivo*	Zelcbuch et al. (2016)
E. coli	Formate and CO_2	18 ± 4	CBB cycle and FDH (*Pseudomonas sp. 101*)	• Constructed by rational design and adaptive laboratory evolution • CO_2 is the carbon source and formate the energy source	Gleizer et al. (2019)
E. coli	Formate and CO_2	157.6	RGP and FDHs (*C. boidinii* and *A. thaliana*)	• Constructed by rational design • CO_2 is the carbon source and formate the energy source • Experiments were performed in bioreactor	Bang et al. (2020)
E. coli	Formate and CO_2	9	RGP and FDH (*Pseudomonas sp.* 101)	• Constructed by rational design and adaptive laboratory evolution • Shortest doubling time obtained	Kim et al. (2020)
E. coli	Formate and CO_2	6	RGP and FDH (*Pseudomonas sp.* 101)	• Further adaptation of the strain developed in their 2020 study • Growth was also performed in bioreactor • Highest maximal cell growth on formate and CO_2	Kim et al. (2023)

The modified *E. coli* strains described in this section achieved decent growth rates, although not yet ideal for industrial use. Still, these successes demonstrated *E. coli*'s adaptability and flexibility in using different carbon sources.

3. Formate dehydrogenases

Formate dehydrogenases (FDHs) are biocatalysts capable of the reversible two-electron oxidation of formate to CO_2, as described in Equation 3 (Ferry 1990).

Equation 3. Formate dehydrogenase reaction.

$$HCOOH \leftrightarrow CO_2 + 2H^+ + 2e^-$$

While the reaction can be reversible, most of the known FDHs are only catalyzing formate oxidation under physiological conditions, which is why these enzymes are referred to as formate dehydrogenases and not CO_2-reductases.

It has been unclear for a long time whether FDHs reduce CO_2 or bicarbonate (HCO_3^-) but it has recently been confirmed that these enzymes favor CO_2 as the substrate for the reverse reaction, at least for the W/Mo-dependent FDHs (Meneghello et al. 2021). The substrate study was also performed for a metal-independent FDH, *Candida boidinii* and was determined to be CO_2 as well (Sato and Amao 2020).

1. Function in microorganisms

Formate is a metabolite involved in many metabolic pathways of aerobic and anaerobic microorganisms. To produce or consume this compound, these organisms use FDHs.

As an example, the model-organism facultative methylotroph *Methylorubrum extorquens* AM1 possesses three FDHs involving molybdenum (Mo) or tungsten (W) which brings to the microorganism flexibility regarding environmental conditions. These enzymes are NAD^+-dependent and allow the bacteria to grow on formate by using it as an NADH-generating system in the absence of methanol in the medium (Chistoserdova et al. 2004).

In anaerobic microorganisms such as the acetogen *Moorella thermoacetica* (formerly *Clostridium thermoaceticum)* using the Wood-Ljungdahl pathway (described in I. 2. 3.), FDHs catalyze the reduction of CO_2 to formate in the methyl branch, which is further processed for the synthesis of acetate (Ljungdahl and Andreesen 1975).

E. coli, a facultative anaerobe, produces large amounts of formate during anaerobic fermentative growth on glucose *via* the pyruvate formate-lyase (PFL) complex. In this organism, three FDHs are reported: FDH-F, FDH-N and FDH-O (Sawers 1994). Within the formate hydrogenlyase (FHL) complex, FDH-F realizes the oxidation of formate while producing electrons (Equation 3) further used in dihydrogen production (McDowall et al. 2014). FDH-N is involved in the formate-nitrate respiratory chain operational under anaerobic growth conditions and catalyzes the oxidation of formate, the reducing equivalents

obtained are transferred to nitrate (Garland, Downie, and Haddock 1975). The role of the membrane-bound FDH-O seems to be the rapid adaptation of *E. coli* metabolism upon switching from aerobiosis to anaerobiosis (Benoit, Abaibou, and Mandrand-Berthelot 1998).

To summarize, formate dehydrogenases are mainly involved in the usage of formate as an energy source either by transferring the electrons to another enzyme (mainly in anaerobic organisms) or by direct regeneration of reducing power, such as NADH (mainly in aerobic organisms), at the exception of the FDH involved in the Wood-Ljungdahl pathway.

2. A large family

Formate dehydrogenases form a heterologous group of enzymes diverse in terms of subunit composition, structure and mode of action. Hence, the enzymes can be classified according to multiple criteria, one being their dependence on an NAD^+ cofactor for the exchange of electrons. The main features of the two categories thus defined, *i.e.* NAD^+-dependent and independent enzymes will be discussed in the following sections with a focus on NAD^+-dependent FDH. In addition, a more detailed description of the metal and NAD^+-dependent multi-subunits FDH from *C. necator* is provided, as it was the main enzyme tested for its suitability as a catalyst for autotrophic growth in aerobic conditions.

While we focus on the NAD^+ co-substrate of the enzyme as it is of particular interest for our study, other co-substrates are also used by FDHs, namely cytochromes, quinone and NAD^+-ferredoxin (Amao 2018).

Besides the cofactor, other classifications of the FDH family based on the metal content of the enzymes have also been proposed (Maia, Moura, and Moura 2015) but are less relevant to this study.

1. NAD^+-independent formate dehydrogenases

As stated above, FDHs are in general active as formate oxidizing catalysts. However, activity in the reductive sense has been reported for a number of these enzymes. The class of NAD^+-independent FDHs harbors examples exhibiting the highest CO_2-reductive activities. These enzymes contain a molybdopterin guanine dinucleotide (MGD) cofactor where molybdenum, or alternatively tungsten, constitutes the active center, and involve diverse forms of Fe-S clusters. In general, these redox cofactors entail high sensitivity to oxygen and consequently, the enzymes are predominantly found in anaerobic microorganisms or are only induced under anoxic growth conditions in facultative aerobes (Jormakka, Byrne, and Iwata 2003).

As an example, the FDH from the acetogenic bacterium *Acetobacterium woodii* (AwFDH) where it is part of the hydrogen-dependent carbon dioxide reductase (HDCR) complex was reported to reduce CO_2 highly efficiently using H_2 or methylviologen (MV) as electron acceptor, with k_{cat} of 28 and 372 s^{-1}

respectively (Schuchmann and Müller (2013); Nielsen, Lange, and Meyer (2019)). More recently, another FDH from the HDCR complex of *Thermoanaerobacter kuvui* was reported with a k_{cat} for CO_2 reduction of 2654 s^{-1} (Schwarz, Schuchmann, and Müller 2018).

Other FDHs have been reported with a CO_2-reducting activity such as the one from *Syntrophobacter fumaroxidans* tested with MV and on an electrode with a k_{cat} of 282 and 112 s^{-1} respectively (de Bok et al. (2003); Reda et al. (2008)). The FDH activity from *Desulfovibrio desulfuricans* was determined to be 47 s^{-1} with MV as electron donor (Maia et al. 2016). Finally, the FDH-H from *E. coli* was tested on an electrode and a k_{cat} of 112 s^{-1} was determined for CO_2 reduction (Bassegoda et al. 2014).

As described in this section, NAD^+-independent FDHs with a demonstrated CO_2 reduction activity are sensitive to oxygen; the reported activities were all performed under anaerobic conditions. In addition, as described for *Acetobacterium woodii,* the kinetic parameters can vary greatly when changing the electron donor, making the comparison between all the reported activities quite challenging and not necessarily reflecting their activity with their "natural" electron donor. Consequently, these enzymes were not suitable for the project of engineering an autotroph *E. coli* strain with a direct reduction of CO_2 to formate in aerobic conditions.

2. NAD^+-dependent formate dehydrogenases

NAD^+-dependent formate dehydrogenases (EC 1.17.1.9) are mainly found in methylotrophic organisms and play a vital role in the catabolism of C1 compounds. They mainly act as regulators of the reducing equivalent pool in the cells and are expressed when the energy availability is low (Popov and Lamzin 1994).

Within this class, two other sub-categories can be defined regarding the numbers of subunits: either the enzyme is composed of one subunit encoded by one gene and the resulting protein is usually metal-free or the enzyme holds multiple subunits with a complex quaternary structure involving metals. These two subclasses will be further referred to as "simple" and "complex" FDHs. The reaction catalyzed by these enzymes is described in Equation 4.

Equation 4. NAD^+-dependent formate dehydrogenase reaction.

$$HCOOH + NAD^+ \leftrightarrow CO_2 + NADH/H^+$$

Table 2. Comparison of several NAD^+-dependent FDHs described to reduce CO_2 in the literature. This list is not exhaustive. K_m for CO_2 reduction where calculated with CO_2 or alternatively bicarbonate as substrate. K_m for formate oxidation were determined for formate as substrate. ND: not determined. *: activity test performed with bicarbonate instead of CO_2.

Microorganism	Classification	CO_2 reduction			Formate oxidation			Reference
		k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (mM^{-1} s^{-1})	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m (mM^{-1} s^{-1})	
Thiobacillus sp. KNK65MA	Non-metal/simple	0.32*	9.23*	0.034*	1.77	16.24	0.11	Choe et al. (2014)
Chaetomium thermophilum	Non-metal/simple	0.032	0.2	0.16	2.04	3.3	0.62	Çakar et al. (2020) Aslan et al. (2016)
Candida boidinii	Non-metal/simple	0.02	31.28	0.0004	1.08	8.55	0.13	Choe et al. (2014)
Myceliophthora thermophile	Non-metal/simple	0.1	0.43	0.23	0.32	7.2	0.04	Altaş et al. (2017)
Clostridium ljungdahlii	Mo-containing/simple	0.73*	7.27*	0.1*	14.77	1.4	10.55	Çakar et al. (2018)
Rhodobacter capsulatus	Mo-containing/complex	1.48*	ND	ND	36.48	0.28	130.3	Hartmann and Leimkühler (2013)
Cupriavidus necator	Mo-containing/complex	4.8	ND	ND	99	0.26	380.8	Yu et al. (2019)

1. *Simple FDHs*

Simple monomeric non-metal/NAD^+-dependent FDHs have been widely used for their high NADH-regeneration capabilities through the oxidation of formate (Gleizer et al. (2019); Bang et al. (2020); Wenk et al. (2020)). In addition, for some of these enzymes, a CO_2 reduction activity was reported, yet the reported activities were low (Table 2).

Among these simple metal-independent FDHs, the FDH from *Thiobacillus sp.* KNK65MA is the most promising enzyme considering the k_{cat} of 0.32 s^{-1} for CO_2 reduction. Kinetic and structural studies were conducted showing the enzyme to be a homodimer encoded by a 1.2 kbp gene (Nanba, Takaoka, and Hasegawa 2003). However, the determined K_m remains high (9.23 mM) showing a low affinity of the enzyme for CO_2 as substrate (Choe et al. 2014). Still, this enzyme was considered an interesting option for this study and was used in a number of experiments that are detailed in the Results and discussion chapter (III. 1). The structural analysis gave important insights into the mechanism of this protein, in particular regarding the reduction activity for which residues such as Lys287 play a crucial role in the capture of formate. Moreover, the activation energy was revealed to be about 10 $kcal.mol^{-1}$ lower than the one determined for the other well-studied simple FDH, such as the CbFDH from *Candida boidinii*

(Choe et al. 2015). In this study, the *Thiobacillus* enzyme was used as a benchmark reference for CO_2/formate oxidation-reduction.

Another possibility is the FDH from *Clostridium ljungdahlii* since it is encoded by a single gene and harbors the best activity towards CO_2 reduction between the simple FDHs. Interestingly, to date, it is the only simple metal-FDH reported with such activity. The kinetic parameters described in Table 2 are the ones reported by Çakar et al. (2018) where the pH optimum for CO_2 reduction (k_{cat} = 0.73 s^{-1} and $K_{m,bicarbonate}$ = 7.27 mM) was determined to be around 8.5-9 at 25 °C. In addition, the best metal cofactors for this enzyme was tested between Mo and W and was reported to be Mo. Surprisingly, another study on the same enzyme was published very recently and stated different conclusions. Moon et al. (2022) determined kinetic parameters for ClFDH towards CO_2 reduction, with a k_{cat} of 5.66 s^{-1} and a K_m of 66.18 mM. In addition, the pH optimum for this reaction was described to be 7 at 30 °C and the enzyme was tested as a W-dependent FDH as was reported in another study (Kuk et al. 2019). While the results from the first publication are mentioned in the more recent paper, no discussion about what could explain such differences is made. This shows that very little is known about this specific enzyme and that further study should be performed to determine its potential for the biotechnological field.

Simple FDHs are harboring relatively low CO_2 reduction rates but could represent an option for direct CO_2 reduction, especially if it was combined with a mutagenesis approach to increase their capabilities. To date, there is no reported demonstration of their *in vivo* activity for CO_2 reduction. On the other hand, during the past decade, NAD^+-dependent FDHs with multi-subunits were reported with a higher activity towards CO_2 reduction.

2. *Complex FDHs*

Recently, some complex NAD^+-and metal-dependent FDHs gained interest for both structure/function studies and as potential CO_2 assimilating catalysts for biotechnological applications. This interest is motivated by the fact that these enzymes, despite the presence of the MGD cofactor, Fe-S clusters and metals (in general Mo or W) were described as oxygen tolerant. The best studied examples are the RcFDH from *Rhodobacter capsulatus* (Hartmann and Leimkühler 2013) and the CnFDH from *Cupriavidus necator* (Yu et al. 2017). A significant activity towards CO_2 reduction in the presence of O_2 was demonstrated for these two enzymes *in vitro*.

These two enzyme complexes offer the possibility to demonstrate the activity of *in vivo* CO_2 reduction to formate in a model organism such as *E. coli* and eventually obtain autotrophic growth thanks to this heterologous enzyme. However, the high complexity of their quaternary structure represents a challenge for their efficient heterologous expression in *E. coli*.

The first FDH of this type to have been described is the one from *Rhodobacter capsulatus*, a facultative anaerobic and photosynthetic nonsulfur purple bacterium (Weaver, Wall, and Gest (1975); Hartmann and Leimkühler (2013)). RcFDH was heterologously expressed in *E. coli*, purified, characterized and was initially described as a $(\alpha\beta Y)_2$ heterotrimer encoded by a five genes operon, *fdsGBACD* with *fdsC* and *fdsD* encoding maturases. The α-subunit FdsA is the largest subunit (105 kDa) and contains the bis-MGD cofactor, four $[Fe_4S_4]$ clusters and one $[Fe_2S_2]$ cluster. FdsB, the β-subunit of 52 kDa, harbors one additional $[Fe_4S_4]$ cluster and the FMN cofactor. Finally, FdsG (15 kDa), the Y-subunit binds one $[Fe_2S_2]$ cluster. The kinetics parameters of this enzyme are detailed in Table 2 and these parameters combined with the O_2-tolerance of this enzyme make it a target for biotechnological applications. Later, the same group of researchers published the structure of this enzyme from a culture of *R. capsulatus* and the cryo-EM structures revealed a dimer of heterotetramers with FdsD (7 kDa) subunit bound to FdsA as represented in Figure 8 (Radon et al. 2020).

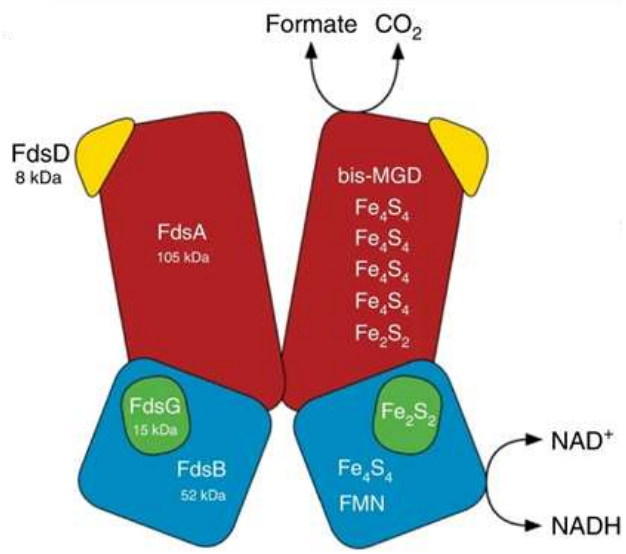


Figure 8. Schematic representation of RcFDH subunits. Cofactors are indicated. FdsA, red; FdsB, blue; FdsG, green; FdsD, yellow. From Radon et al. (2020).

The second FDH of this subclass to be described was from *Cupriavidus necator*, formerly known as *Ralstonia eutropha* (Yu et al. 2017). Interestingly, when the FDH from this strain was initially characterized by Friedebold and Bowien (1993), no CO_2 reduction activity was detected. This strain is an aerobic facultative chemoautotrophic bacterium capable of both heterotrophic (on various organic substrates) and lithoautotrophic growth with H_2 and CO_2 as sole carbon and energy source (Friedebold and Bowien

1993). This enzyme, CnFDH, harbors very similar genetic and structural features compared to RcFDH with *fdsGBACD* constituting the five genes operon (Table 3).

Table 3. Protein alignment between CnFDH and RcFDH. The alignment was performed using SnapGene software.

	FdsG	FdsB	FdsA	FdsC	FdsD
% identity	50.7	61.3	69.13	47.5	45.0
% similarity	66.9	74.4	80.5	64.6	53.3

The size of the β and Y subunits vary slightly compared to RcFDH being respectively 55 and 19 kDa, in contrast the α-subunit has the same size (Yu et al. 2017). There is no structure available for this protein and, to date, the involvement of FdsD in the active enzyme is not confirmed. The kinetic parameters are detailed in Table 2, CnFDH exhibits better features in terms of kinetic parameters. However, for RcFDH the CO_2-reduction activity was measured using bicarbonate as substrate while the CO_2 reduction activity for CnFDH was determined with a CO_2-saturated solution. It was demonstrated that the preferred substrate for FDH is CO_2, not bicarbonate (Yu et al. (2017); Ruschig et al. (1976); Meneghello et al. (2021)). Importantly, the kinetic parameters discussed here were obtained either from proteins heterologously expressed in *E. coli* and purified under aerobic conditions or from proteins homologously expressed and purified from their natural host strain. For RcFDH, the homologous expression in *Rhodobacter capsulatus* yielded an enzyme for which the k_{cat} for formate oxidation was determined as 60 ± 2 s^{-1} (Radon et al. 2020), a 40 % increase compared to heterologous expression (Table 2). CnFDH was expressed and purified in strain *Cupriavidus necator* H16 and the k_{cat} obtained for formate oxidation was 201 s^{-1} (Niks et al. 2016), representing a 50 % activity loss as compared to the heterologously expressed version, which is consistent with the results obtained for RcFDH. While these enzymes are O_2-tolerant and an efficient expression was obtained from heterologous expression in *E. coli*, they still show some sensitivity to oxygen, notably in the absence of a stabilizing inhibitor such as azide, which can distort the results found during enzymatic assays (Hakopian, Niks, and Hille 2022).

Because of the better results reported for the FDH from *C. necator*, this enzyme was favored as part of this research project.

4. Escherichia coli

1. General features

Escherichia coli is a Gram-negative eubacterium belonging to the Enterobacteria family and is part of the intestinal microbiome of humans and other mammals. These commensal *E. coli* strains are most of the time harmless but the acquisition of a few genetic elements can transform the cells into pathogens (Kaper, Nataro, and Mobley 2004). *E. coli* can grow on various carbon sources and has the ability to grow aerobically and anaerobically, making it a facultative anaerobe (Palsson 2015).

Escherichia coli strain MG1655 is a derivative of the wild-type strain K-12, and was one of the first bacteria to have its whole genome sequenced (Blattner et al. 1997). This strain is one of the best-characterized organisms and a model for many research projects.

E. coli is a heterotrophic organism typically growing on carbohydrates as carbon and energy source (Figure 9 shows the metabolic routes for glucose assimilation). Glucose is phosphorylated to glucose-6-phosphate and yields two molecules of pyruvate along with two ATP and two NADH molecules *via* the Embden-Meyerhoff-Parnass (EMP) glycolysis pathway. In aerobic growth mode, pyruvate is converted to acetyl-CoA by the pyruvate dehydrogenase and acetyl-CoA enters the TCA cycle where it is oxidized to CO_2. Under anaerobic conditions, the Krebs cycle is downregulated, pyruvate formate-lyase replaces pyruvate dehydrogenase and produces formate and acetyl-CoA which is converted to acetate (Förster and Gescher 2014).

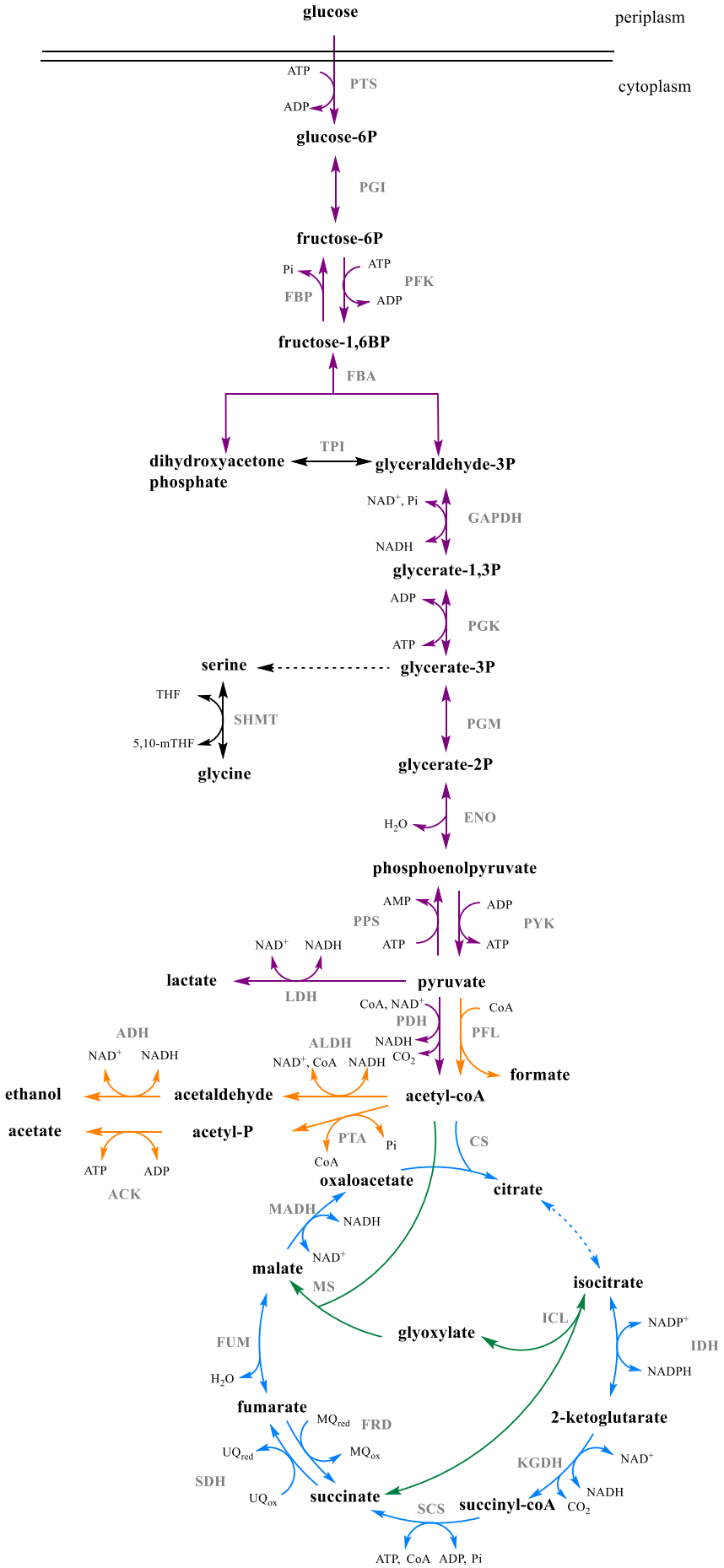

glucose
periplasm
cytoplasm
ATP
PTS
ADP
glucose-6P
PGI
fructose-6P
Pi
FBP
ATP
PFK
ADP
fructose-1,6BP
FBA
TPI
dihydroxyacetone phosphate
glyceraldehyde-3P
NAD+, Pi
GAPDH
NADH
glycerate-1,3P
ADP
PGK
ATP
serine
glycerate-3P
THF
SHMT
5,10-mTHF
glycine
PGM
glycerate-2P
H2O
ENO
phosphoenolpyruvate
AMP
PPS
ATP
ADP
PYK
ATP
NAD+
NADH
lactate
LDH
pyruvate
CoA, NAD+
PDH
NADH
CO2
CoA
PFL
formate
ADH
NAD+
NADH
ALDH
NAD+, CoA
NADH
ethanol
acetaldehyde
acetyl-coA
acetate
acetyl-P
PTA
CoA
Pi
CS
ATP
ADP
ACK
oxaloacetate
citrate
MADH
NADH
NAD+
malate
MS
isocitrate
ICL
glyoxylate
NADP+
IDH
NADPH
FUM
H2O
fumarate
MQred
FRD
MQox
UQred
SDH
UQox
succinate
2-ketoglutarate
KGDH
NAD+
NADH
CO2
SCS
succinyl-coA
ATP, CoA
ADP, Pi

Figure 9. Central metabolism of *E. coli*. The EMP glycolysis (purple arrows), the Krebs cycle (blue arrows) and glyoxylate shunt (green arrows) are represented. The orange arrows represent reactions that are occurring in anaerobic growth conditions. Dashed arrows: at least two reactions are occurring between the two molecules. ACK, acetate kinase; ADH, alcohol dehydrogenase; ALDH, aldehyde dehydrogenase; CS, citrate synthase; ENO, enolase; FBA, fructose-bisphosphate aldolase; FBP, Fructose-1,6-bisphosphatase; FRD, fumarate reductase; FUM, fumerase; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; IDH, isocitrate dehydrogenase; KGDH, 2-ketoglutarate dehydrogenase complex; LDH, lactate dehydrogenase; MADH, malate dehydrogenase; MS, malate synthase; PDH, pyruvate dehydrogenase; PFK, 6-phosphofructokinase; PFL, pyruvate formate-lyase; PGI, phosphoglucose isomerase; PGK, phosphoglycerate kinase; PGM, phosphoglycerate mutase; PPS, phosphoenolpyruvate synthetase; PTA, phosphate acetyltransferase; PTS, phosphotransferase system; PYK, pyruvate kinase; SCS, succinyl-CoA synthetase; SDH, succinate dehydrogenase; SHMT, serine hydroxymethyltransferase; TPI, triose-phosphate isomerase.

2. Versatile biotechnological strain

Escherichia coli stands as one of the most studied organisms, and the intimate knowledge of its metabolism and genotype serves as valuable assets in the field of synthetic biology. Furthermore, the large set of tools and protocols for genome editing and precise modulation of homologous and heterologous gene expression constitute one of the most extensive toolkits for molecular genetics available (Richter and Gescher 2012). In the current context where bioproduction of chemicals is increasingly preferred over fossil fuel-based methods, these attributes position *E. coli* as a prime candidate for being a universal biotechnological chassis.

As described in the *E. coli as a chassis for synthetic autotrophy* section (I. 2. 5), numerous instances of metabolic rewiring have been carried out in *E. coli*, demonstrating its metabolic flexibility. Beyond the modifications made to the central metabolism of the bacterium, *E. coli* has also been widely employed as a chassis for various bioproductions.

E. coli is widely used to produce various chemical commodities through metabolic engineering and the introduction of synthetic pathways. An illustrative example is ethanol, a biofuel of significance, which has been successfully produced under anaerobic conditions by heterologously expressing pyruvate decarboxylase and alcohol dehydrogenase genes from *Zymomonas mobilis* (reviewed by Förster and Gescher (2014)). Conversely, ethanol was efficiently produced in *E. coli* without the introduction of heterologous genes. *E. coli* achieved efficient ethanol production from glucose or xylose fermentation thanks to a mutation in the pyruvate dehydrogenase gene, which is normally active during aerobic growth (Kim, Ingram, and Shanmugam 2007). Similarly, acetate and pyruvate have been efficiently synthesized during glucose fermentation by modifying *E. coli*'s native pathways. This yielded 86% and 77.9% of the theoretical yield, respectively, without requiring the expression of any heterologous genes (Causey et al. (2003); Causey et al. (2004)).

Lactic acid is a compound of great interest as it can be used as a precursor to produce polylactic acid-based plastic materials. This molecule is naturally synthesized by various microorganisms including yeasts, *Bacillus subtilis*, *Lactobacillus* strains, and *Escherichia coli* (Tengku et al. 2009) but *E. coli* emerges as an attractive choice for lactic acid bioproduction, particularly due to the challenges associated with the complex media required for growth of other organisms such as *Lactobacillus* strains (van Niel and Hahn-Hägerdal 1999). The production of pure D- and L-lactate isomers was successfully performed in various studies. D-lactate, naturally generated through fermentation in *E. coli*, had its production yield optimized to approximately 90% of the theoretical yield in mutant strains engineered to minimize competing pathways (Chang et al. 1999). Moreover, the gene encoding the L-lactate dehydrogenase

from *Streptococcus bovis* was introduced into an engineered *E. coli* strain resulting in the production of L-lactate at a yield of 93% of the theoretical maximum (Dien, Nichols, and Bothast 2001).

More recently, a chassis strain, previously described earlier (*Metabolic engineering work around formatotrophy* section, I. 2. 5. 2. 2; Kim et al. (2020)), which assimilates formate and CO_2 through the reductive glycine pathway, was further modified for the production of D-lactate. The overexpression of the *ldhA* gene from *E. coli* was performed; however, the yield achieved was 10 % of the theoretical maximum, indicating potential for further improvement (Kim et al. 2023). Still, this demonstrates the feasibility of the production of value-added chemicals using C1-assimilating *E. coli* strains.

In addition to the chemical commodities, *E. coli* is also used as a host for the expression of heterologous proteins, including biopharmaceuticals. The major example is insulin which was the first human drug produced using recombinant DNA in *E. coli* in 1982 (Chance and Frank 1993). While the production of recombinant proteins in *E. coli* faces challenges such as codon usage and post-translational modifications, strategies to overcome these limitations are explored and research to expand the scope of bioproduction of molecules in *E. coli* is constantly ongoing (Baeshen et al. (2015); Choi, Keum, and Lee (2006)).

These examples illustrate the wide range of compounds that can be efficiently produced using *E. coli* as a cell factory. To make these bioproductions even more sustainable, the use of CO_2 as a growth substrate for *E. coli* would represent a remarkable achievement.

As described in the previous sections, the selected FDH utilized for CO_2 reduction to formate relies on NADH as electron donating cofactor. Consequently, an energy generation system in the cell is necessary to support cell growth.

5. Energy generation systems

For the construction of the biotechnological chassis described in the first part of this introduction (I. 1), there is a need for an energy generation system. Initially, the system proposed was the use of a hydrogenase, this would make H_2 the source of energy of our strain. However, considering the complexity of such an enzymatic system, which will be described in the next part, an alternative energy generation system was considered using a phosphite dehydrogenase.

Whether phosphite or hydrogen, the common feature is that these energy sources are non-carbon sources, a highly important point when considering environmental aspects.

Indeed, the systems used in most of the propositions for this type of biotechnological strains are carbon-based (pyruvate or formate for example), like it was already performed in most of the propositions using formate as the energy source through FDH activity, thus rejecting CO_2.

1. Hydrogenase

Hydrogenases are multimeric metalloenzymes capable of the reversible splitting of molecular hydrogen (H_2) into protons and electrons.

Most of the known hydrogenases are iron-sulfur enzymes and can be classified under three different categories depending on the metal atom(s) at their active site: the [NiFe]-, the [FeFe]-, and the [Fe]-hydrogenases (Vignais and Billoud 2007). The hydrogenases of *Escherichia coli* and from *Cupriavidus necator* that will be further described in this part are [NiFe]-hydrogenases. In addition to their metal active sites, these highly complex enzymes contain [Fe-S] clusters (Burgdorf, Lenz, et al. (2005); Forzi and Sawers (2007)).

[NiFe]-hydrogenases are composed of two subunits, one large and one small. The [Ni-Fe] active site is located in the large subunit of approximately 60 kDa while the small subunit varies greatly in size and carries at least one [4Fe-4S] and other [Fe-S] clusters (Burgdorf, Lenz, et al. 2005).

An O_2-tolerant NAD^+-dependent hydrogenase from *C. necator* has been identified in the literature (Friedrich, Fritsch, and Lenz 2011), these two characteristics are crucial for the project since the goal is to build an aerobic biotechnological chassis and the FDH reducing CO_2 requires a NADH generating system.

1. H_2 as a sustainable energy source

Since two decades, hydrogen is seen as a promising and renewable energy source that could replace fossil fuels in a hydrogen economy (Sherif, Barbir, and Veziroglu 2005). As it is not directly available in nature in important amounts, hydrogen must be produced from a primary energy source, as illustrated in Figure 10. In a hydrogen economy scenario, hydrogen is efficiently produced from a renewable energy source through water electrolysis or biomass processes, powered by solar or wind energy (Marbán and Valdés-Solís (2007); Sarmah et al. (2023)). The potential and high expectations for H_2 as a global energy source lie in the fact that it is a zero-emission carrier with water as the only byproduct, if produced from a renewable primary energy source.

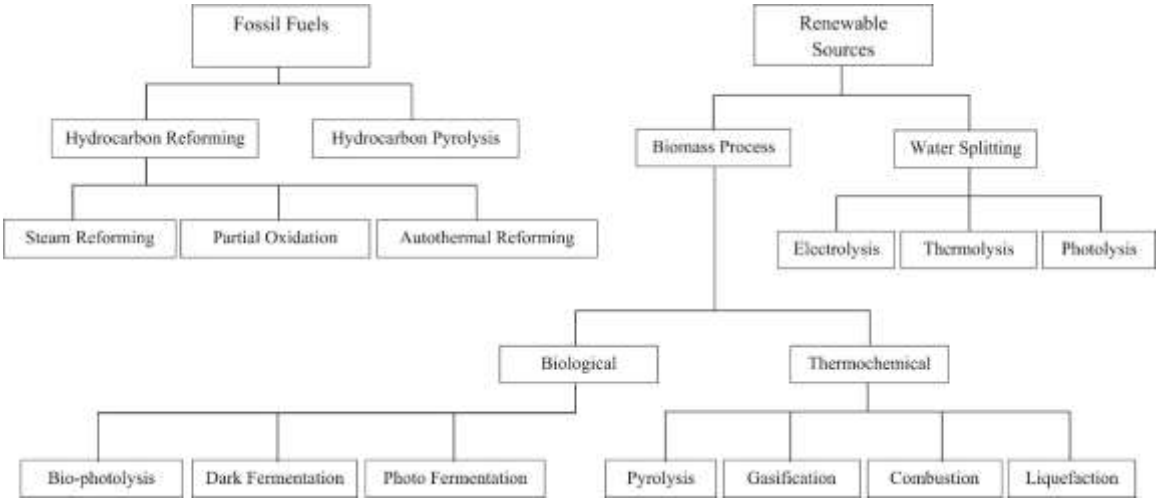


Figure 10. Hydrogen production methods. From Nikolaidis and Poullikkas (2017).

A major prerequisite for H_2 to become the energy source of the future is the need for low-cost production processes in addition to its renewability. For now, H_2 is mainly produced by energy-expensive processes that require high temperatures (above 850 °C), the most important process being steam-methane reforming (Ishaq, Dincer, and Crawford 2022).

In parallel to these mature production processes, the development of technologies based on renewable resources is increasing. Among these technologies, the use of biomass is extensively studied for bio-hydrogen production. It can be realized by bio-photolysis of water by algae which can split water into hydrogen ion and O_2, the dihydrogen gas is then produced by a hydrogenase. Another means is dark-fermentation of organic matter by anaerobic organisms such as Clostridia (Kapdan and Kargi 2006).

It is in this context of growing interest and innovative technologies that hydrogen was seen as a potential energy source for the construction of our biotechnological chassis strain. Still, there are challenges to be met such as its low solubility and its explosivity.

2. Hydrogenases from *Escherichia coli*

To implement a system of hydrogen oxidation to regenerate NADH cofactor in our biotechnological strain, it is important to study the hydrogen metabolism from *Escherichia coli*.

Escherichia coli possesses four NAD^+-independent hydrogenase enzymes, named Hyd-1 to Hyd-4 in order of discovery, catalyzing the reversible reaction of hydrogen oxidation. These four hydrogenases have a complex structure and a metal-based active site which is probably one of the reasons for their instability in the presence of molecular oxygen (Forzi and Sawers 2007). They are all encoded by a

different operon: hydrogenase 1 is encoded by 6 genes forming the *hya* operon, hydrogenase 2 is encoded by 8 genes grouped in the *hyb* operon, hydrogenase 3 is encoded by the 9 genes of the *hyc* operon and hydrogenase 4 is encoded by 11 genes forming the *hyf* operon (Maeda et al. 2018).

Hydrogenase 1 and hydrogenase 2 are expressed in the periplasm, are membrane-bound and responsible for hydrogen uptake under glucose metabolism (Menon et al. (1991); Menon et al. (1994)). Hyd-1 is a 200 kDa multimeric [NiFe]-enzyme that is mainly expressed under anaerobic growth conditions and catalyzes hydrogen oxidation. Hyd-2 is a heterodimer, expressed under anaerobic conditions as well. Hydrogenase 3 and hydrogenase 4 form two FHL complexes respectively FHL-1 and FHL-2. The FHL-1 comprises Hyd-3 and FdhH and was demonstrated to be a reversible enzyme capable of both hydrogen uptake and hydrogen synthesis, unlike Hyd-1 and 2. It is the main responsible for *E. coli* hydrogen production capabilities (Maeda, Sanchez-Torres, and Wood 2007). Finally, FHL-2 is composed of Hyd-4 and FdhH, and was reported with hydrogen production capabilities as well (Mirzoyan et al. 2017).

3. NAD^+-dependent hydrogenase from *Cupriavidus necator*

Cupriavidus necator H16 (formerly *Ralstonia eutropha* and *Alcaligenes eutrophus*) is a Gram-negative lithoautotrophic bacterium from the *Pseudomonata* (or *Proteobacteria*) phylum, is found in soil and freshwater and is able to survive transient anoxia. This versatile organism is a hydrogen-oxidative ("Knallgas") bacterium able to use H_2 as its source of energy in the absence of organic substrates (Pohlmann et al. 2006).

Cupriavidus necator H16 possesses at least three hydrogenases that are encoded on the megaplasmid pHG1. The membrane-bound hydrogenase (MBH) is encoded by the genes *hoxG*, *hoxK* and *hoxZ*. The soluble hydrogenase (SH), which is located in the cytoplasm, is the one with the NAD^+-reduction activity and is encoded by the genes *hoxH*, *hoxY*, *hoxU*, *hoxF* as well as two accessory genes *hoxW* and *hoxI*. The Regulatory Hydrogenase (RH) acts as an H_2 sensor for the expression of MBH and SH and is encoded by the genes *hoxC* and *hoxB* in link with *hoxJ* (histidine protein kinase). A fourth hydrogenase was identified following sequence analysis of the pHG1, although its precise function remains unclear and there is limited documentation available (Schwartz et al. 2003). The three main enzymatic complexes are represented in Figure 11 below.

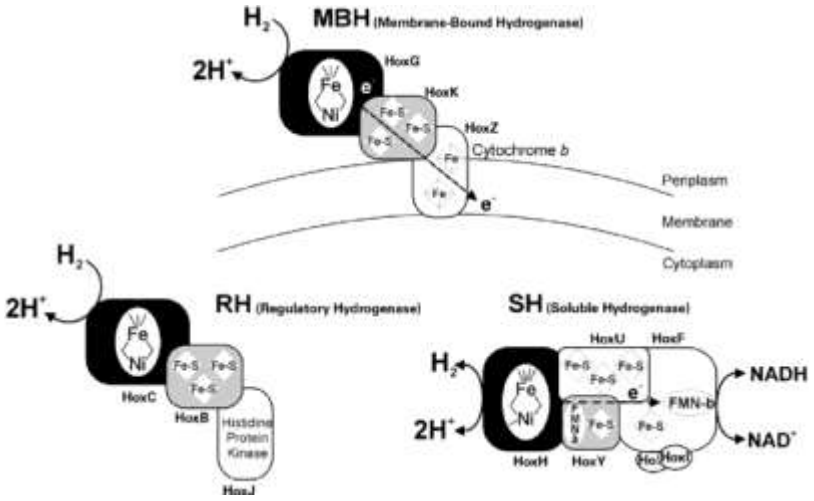


Figure 11. The three [NiFe]-hydrogenases of *Cupriavidus necator* H16. From Burgdorf, Lenz, et al. (2005).

For this study, the focus was placed on the SH of *C. necator* since this enzyme was described as an O_2-tolerant NAD^+-dependent H_2-oxidizing enzyme (Burgdorf, Lenz, et al. 2005). In addition, this enzyme was successfully expressed in *E. coli*, which makes it a promising candidate, as elaborated in the next section (I. 5. 1. 4).

The soluble hydrogenase from *Cupriavidus necator* (CnHyd) is formed of two main modules, the heterodimeric hydrogenase (*hoxH* and *hoxY*) and the FMN-containing NADH:oxidoreductase (diaphorase) module (*hoxF* and *hoxU*). Additionally, *hoxI* encodes a homodimer linked to the diaphorase subunit, which provides a binding site for NADPH (Burgdorf, van der Linden, et al. 2005). As a [NiFe]-hydrogenase, CnHyd is characterized by the need for an extensive maturation machinery involving numerous auxiliary genes, *hypA2B2F2C1D1E1*, *hypX and hoxW* (Schiffels et al. (2013); Lamont and Sargent (2017)).

4. CnHyd expression in *Escherichia coli*

The soluble hydrogenase from *Cupriavidus necator* (CnHyd) has been successfully expressed in *E. coli* strains, under various formats and several successful attempts are outlined in this paragraph.

The first study, conducted by Schiffels et al. (2013), focused on assembling two plasmids carrying every gene necessary for a functional CnHyd, as mentioned in the previous section (I. 5. 1. 3). The cloning was performed using native genes amplified from *C. necator* megaplasmid pHG1. One plasmid was dedicated to the structural genes, while the other one carried the auxiliary proteins. Since two copies of

the auxiliary genes *hypABCDEF* exist in *C. necator*, two separate constructs for the latter plasmid were created. Each gene was placed under the control of individual promoters and terminators. This study aimed to demonstrate the efficiency of a gene assembly technique for enzyme production. Hence, the authors exclusively performed *in vitro* studies of the resulting protein, which proved to be functional for hydrogen oxidation.

Another study involving the same enzymatic complex but focused on *E. coli* metabolic engineering was conducted by Ghosh, Bisaillon, and Hallenbeck (2013). In this study, the authors used native genes from *C. necator* and cloned the entire operon, including the auxiliary genes, into a single plasmid. They successfully demonstrated the *in vivo* activity of CnHyd for H_2 production in anaerobic conditions, using an engineered strain unable of NAD^+ regeneration.

The last work that is cited is the one from Lamont and Sargent (2017) in which a different approach using synthetic genes was adopted. The cloning of CnHyd was carried out on two different plasmids, one containing the structural genes and another one with the auxiliary genes, in an organization similar to the study from Schiffels mentioned above. The genes on each plasmid were organized in operons, with a promoter upstream of the genes and RBS placed between each gene. To demonstrate the successful expression, the authors measured hydrogen production by the complex in *E. coli* strains lacking endogenous hydrogenases. The genetic constructs from this study were graciously provided for our project by Frank Sargent.

As demonstrated in this section, hydrogenases are highly complex, multimeric enzymes that are challenging to express heterologously. Consequently, in parallel to this initial strategy, we also explored the option of phosphite as the energy source of our system.

2. Phosphite dehydrogenase

1. Phosphite as a potential source of energy for biotechnological application

Phosphite, HPO_3^{2-}, is seen as a highly promising energy source for microorganisms thanks to its very low redox potential (-650 mV). Additionally, it has favorable features, such as high solubility and low microbial toxicity. Unlike hydrogen, phosphite has a low flammability, which makes it safer to use.

Currently, only two organisms are known to use phosphite as their sole energy source. The first to have been described is *Desulfotignum phosphitoxidans* FiPS-3, isolated from a marine sediment. This microorganism is capable of coupling phosphite oxidation to CO_2 or sulfate reduction (Schink et al. 2002). The second organism that was described with a similar metabolism is the uncultured bacterium

Candidatus Phosphitivorax. Genomic analysis has provided data suggesting that it also couples CO_2-reduction and phosphite oxidation (Figueroa et al. 2018). Both of these organisms are expected to use a *ptx-ptd* gene cluster for phosphite oxidation, although there is limited knowledge about these organisms and the mechanism of this gene cluster.

While these phosphite oxidizing systems are poorly described, a promising option from a unique NAD^+-dependent D-hydroxy acid dehydrogenase found in *Pseudomonas stutzeri* is described. This enzyme, a phosphite dehydrogenase, can oxidize phosphite using NAD^+ as its cofactor, leading to the formation of phosphate and NADH, was described in several studies (Costas, White, and Metcalf (2001); White and Metcalf (2004b); Figueroa and Coates (2017)).

2. Phosphite dehydrogenase from *Pseudomonas stutzeri*

Phosphite dehydrogenase (PTDH) was seen as an alternative for the complex hydrogenase system initially selected. This enzyme produces phosphate and NADH from NAD^+ and inorganic phosphite. Indeed, contrary to the described system above, PTDH consists of a single gene, *ptxD*, encoding the active enzyme phosphite dehydrogenase and three other genes *ptxA*, *ptxB* and *ptxC* which encode a phosphite transporter. These four genes are organized in an operon along with a regulator, *ptxE* (White and Metcalf 2007).

The only enzyme of this kind to be precisely described is the one from *Pseudomonas stutzeri* WM88 (Costas, White, and Metcalf (2001); Figueroa and Coates (2017)) and was chosen for this study. This enzyme, encoded by *ptxD* gene, was characterized as a homodimer by heterologous over-expression in *E. coli* (Costas, White, and Metcalf 2001). This study showed a high affinity of the enzyme for its phosphite (K_m = 53.1 ± 6.7 µM) and NAD^+ (K_m = 54.6 ± 6.7 µM) substrates with a turnover number (k_{cat}) of 7.3 s^{-1}, indicating efficient catalytic activity. *In vitro*, the enzyme operates optimally at 35 °C and tolerates a wide range of pH (pH 5-9), with an optimum between 7.25 to 7.75. Importantly, the researchers did not observe the reverse reaction during their study. This observation is expected considering the thermodynamic metrics of this reaction, where the reduction potential of the phosphate/phosphite couple is -650 mV and the one from the NAD^+/NADH couple is -320 mV. Thus, the reduction of NAD^+ by phosphite oxidation is the most favorable reaction (Costas, White, and Metcalf 2001).

Pseudomonas stutzeri has a respiratory metabolism with oxygen as the final electron acceptor and can grow in a wide range of temperatures with an optimum of 35 °C (Lalucat et al. 2006), very close to *E. coli* characteristics. In *P. stutzeri* metabolism, the unique role of *ptxD* is to provide phosphate for the cells when no other P sources are available (White and Metcalf 2004a). However, despite the NADH generation in the phosphite oxidation process, this strain was not described as capable of using phosphite as its energy source.

Two studies have been conducted for heterologous expression of *ptxD* in *E. coli* and *P. putida* for biocontainment purposes in which the enzyme was used for its phosphate generation capabilities. In *E. coli*, a strain dependent on phosphite/hypophosphite for growth was engineered making it unable to grow in the environment due to the low concentrations of such compounds. This was carried out by disrupting inorganic phosphate transporters (*pit, pst, phn*) and Pi-ester transporters (*uhp, glpT*) from *E. coli* chromosome. The expression of the exogenous hypophosphite/phosphite specific transporter, HtxBCDE, along with HtxA for conversion of hypophosphite to phosphite and PtxD for phosphite oxidation to phosphate was performed. This study proves the efficient activity of PtxD *in vivo* in *E. coli* to sustain the phosphate needs of the engineered *E. coli* strain (Hirota et al. 2017).

A similar strategy was followed in *Pseudomonas Putida*, where all phosphate and phosphate esters transporters were deleted and HtxBCDE for phosphite import in the cells and PtxD for its conversion to phosphate were expressed (Asin-Garcia et al. 2022). Both studies validated the *in vivo* activity of PtxD in *E. coli* and *P. putida* to support growth by efficiently converting phosphite to phosphate.

While PTDH has been proposed for its potential application as a cofactor generation system to sustain growth (Claassens et al. 2018), an actual experimental demonstration of this application has not yet been published and was part of the objectives of this thesis project.

6. Mutagenesis

1. General features

Mutagenesis describes the process of modification of the genetic material, in general DNA, of an organism. Mutagenesis occurs spontaneously in nature and is the fundamental process underlying evolution. The modifications involve base pair exchange, insertions or deletions of small fragments, and deletions or rearrangements of larger DNA fragments. Mutations are introduced *via* errors during replication or upon exposition to a mutating chemical or to ionizing radiation. Most living cells possess DNA reparation systems that ensure a low mutation rate, which is crucial for genetic integrity. For a wild-type *E. coli*, the mutation rate was determined to be $\sim 1 \times 10^{-3}$ per genome per generation (Lee et al. 2012). Defective reparation or proofreading systems can lead to mutator phenotypes exhibiting up to 1000 times higher mutation rates (Taddei et al. (1997); (Kang et al. 2019)), often resulting in unstable genetic contexts and growth fluctuations.

It could be anticipated that the oxygen-tolerant formate dehydrogenases, principal subject of this Ph.D. work, do not have an intrinsic activity sufficient to sustain autotrophic growth (see section above I. 3. 2. 2). In addition, due to a lack of structure/function knowledge, targeted mutagenesis of candidate residues prone to enhance enzyme activity cannot be envisioned. Evolution in automated continuous culture of strains enabling selection for FDH-catalyzed formate/CO_2 redox conversion is a promising option for obtaining more active FDH variants. However, constraints imposed by the genetic code and low mutation rates lower the probability to obtain enhanced mutants during reasonable time spans when recurring to naturally occurring mutations. The process of evolution can be accelerated in the laboratory using mutagenesis protocols that target the gene in question (Figure 12). These methods can be divided into *in vitro* and *in vivo* techniques, *in vitro* techniques comprising at least one *ex vivo* step, as exemplified by the creation of a gene library through error-prone PCR. These methods in general suffer from limited mutant diversity since the library needs to be introduced into the selection strain, step notoriously inefficient. *In vivo* approaches have the advantage of unity of time and space, the mutant gene variants, created in the selection strain, can be screened by switching to selective growth conditions.

In recent years, innovative *in vivo* mutagenesis methods have been developed, allowing accumulation of mutations over a whole gene with a broad spectrum of mutations and low off-target mutation rates. In the following paragraphs, a brief overview of *in vitro* and promising *in vivo* methods is provided, with an emphasis on the Evolution.T7 method, which was initiated during this Ph.D. project.

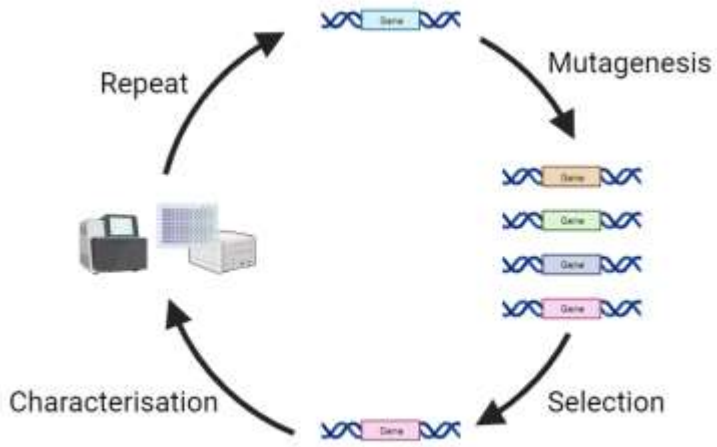


Figure 12. Directed evolution workflow.

2. Methods of *in vitro* mutagenesis

1. Site-directed mutagenesis

The site-directed mutagenesis (SDM) is a method used to introduce specific nucleotide modifications in a double-stranded plasmid DNA and can be used to insert, delete or substitute a nucleotide. This method can be useful to introduce an amino acid modification in the protein encoded by the targeted gene, disrupt a gene by the introduction of a stop codon or delete a base to shift the reading frame, among others. SDM is usually performed by using two primers in forward and reverse directions centered on the mutation to introduce (Bachman 2013).

2. Overlap extension PCR mutagenesis

The overlap extension PCR mutagenesis is a tool used for insertion, deletion or substitution of larger DNA fragments compared to the SDM method described above. This method is based on the production of PCR products using usually two pairs of primers creating overlaps to allow their recombination. Then, the two fragments are annealed and another PCR is performed to obtain one full-length PCR product containing the desired mutation that can be further cloned in a plasmid vector (Lee et al. 2010). Although this method can be useful in some cases, it has limits due to inherent constraints in the primer design and especially the length of the insertion achievable. Moreover, it is principally applicable when the mutation to introduce is precisely defined at a specific position.

3. Error-prone PCR

The error-prone PCR is utilized to introduce random mutations in a DNA template using a DNA polymerase with low fidelity and thus create a library of variants. This technique is very useful for screening a large number of variants of a protein with very little knowledge of its structure (Matteo, Alex, and Narendra 2018). Error-prone PCR is able to drastically reduce the fidelity of the replication from 10^{-10} to 10^{-4} (Wang, Xue, et al. 2021), thus efficiently introducing a high number of mutations. The low-fidelity DNA polymerase error rate can be further modulated by modifying the buffer composition such as the addition of Mn^{2+}, increasing the Mg^{2+} concentration, and using an unequal dNTP mix (Wilson and Keefe 2000). The mix of variants produced through this protocol is then transformed into a cloning strain and the library is screened to look for variants with enhanced activity. However, this method is subjected to bias in terms of nucleotide introduced. Depending on the design of the primers, the target can be the gene, which is then sub-cloned into a vector before transformation or the target can be the entire plasmid but it will inevitably introduce off-target mutations in the antibiotic resistance gene used for the selection or the origin of replication for example. To summarize, the error-prone PCR approach by itself

is tedious and requires a transformation strain with high efficiency to maintain as much diversity as possible, which is not always the case for the strain used for selection.

4. Chemical mutagenesis

In vitro mutagenesis can be performed using specific chemicals that will alter the double-stranded DNA. To this end, the major compounds are sodium bisulfite, formic acid, and hydrazine. Sodium bisulfite will deaminate unmethylated cytosine bases changing it to uracil; this altered base will be paired with adenosine during DNA replication instead of guanosine (Darst et al. 2010). The last two mentioned compounds will create abasic sites by removing the base from the DNA strand, that will be paired with any of the four nucleotides during DNA replication (Arkin 2001).

In vitro mutagenesis is the most suitable when the desired modification is targeted to a specific location. However, when used for the creation of libraries of mutations, inescapable bias of the different steps of mutation generation, sub-cloning in an expression vector and transformation in a recipient strain for selection lead to a loss of variability, which is the major limitation of these techniques.

3. Methods of *in vivo* mutagenesis

In vivo methods have been largely developed in the last decades in order to circumvent the above-mentioned drawbacks of the *in vitro* approaches. Moreover, the available methods of *in vivo* mutagenesis, especially the ones involving the exposition of the cells to UV of chemicals, were highly problematic because of the off-target mutations that can affect essential genes and be deleterious for the cells.

CRISPR-Cas9 assisted methods were proposed and will be rapidly discussed in the next subsection. In parallel, a wide range of methods using the RNA polymerase of bacteriophage T7 (T7 RNAP) were developed, including the one explored for this research project, the Evolution.T7 method.

1. Recombination-based methods

1. CRISPR-Cas9 assisted methods

The CRISPR-derived systems allow for efficient DNA editing, such as large fragment deletions, insertions or replacement, as well as multiplex editing. Technologies were developed for mammalian cells, yeast and bacteria, including *E. coli* (Wang, Xue, et al. 2021).

As an illustrative example, the system named EvolvR uses engineered DNA polymerase I associated with a CRISPR tool, coupled with a nickase Cas9 (nCas9). The nCas9 enzyme creates single-strand nicks in the chromosome at a target locus guided by a gRNA (Halperin et al. 2018). The DNA polymerase involved in this system is error-prone, and when this enzyme is recruited to repair the nick created by the nCas9, mutations are incorporated at this locus. Using this system, substitutions of all four nucleotides were detected, although not in equivalent proportions. The window of mutations was optimized and reached around 60 bp from the nick for most of the mutations observed (Halperin et al. 2018). While this technology can be useful for studying a specific locus within a gene, it requires a large number of gRNA to cover an entire gene.

Another powerful system, known as CRISPR/Cas9-mediated genomic error-prone editing (CREPE), works in combination with a library of error-prone PCR products corresponding to the target genome region flanked with unmutated recombination arms (Choudhury et al. 2020). A plasmid, pCREPE, containing the Cas9 enzyme and the λ-red recombination system is employed and significantly improves the efficiency of the recombination of the repair templates, *i.e.* the library of double-strand DNA fragments generated by error-prone PCR. This is particularly important due to the naturally low efficiency of homologous recombination in *E. coli*. Another plasmid harboring a pre-validated gRNA for the targeted gene is also used in this system. However, this system was used to introduce mutations in sequences of 300 to 400 bp long (Choudhury et al. 2020). For longer sequences, the recombination efficiency, which is a key success factor, decreases significantly, thus leading to a loss of diversity from the library. As a side note, in this method, diversity is created *in vitro* through PCR, but the actual genome editing takes place in the cells during the homologous recombination step, which allows it to be classified as an *in vivo* method.

However, Cas9 systems show limitations when it comes to targeting large loci, thus requiring the use of several gRNAs to target an entire gene and even more to cover entire operons.

2. *MAGE*

A method commonly referred to as MAGE which stands for multiplex automated genome engineering was developed to incorporate simultaneously a large number of chromosomal modifications based on the homologous recombination of single-stranded DNA (ssDNA) assisted by a λ-red system, in *E. coli* (Wang et al. (2009); Wang and Church (2011)). This tool involves oligonucleotides of 90 bp harboring the desired mutation or a library of mutations using degenerate oligos, which can be insertions, deletions or mismatches. To prevent these oligos from being degraded by exonucleases when the transformation is performed, phosphorothioated bases on both extremities of the ssDNA are used. However, the larger the modification, the lower the replacement efficiency tends to be. Thus, Wang and his collaborators set

up a protocol of recombination cycles in order to increase the cells harboring the desired chromosomal modification. Still, the authors demonstrated that for mismatches and insertions, the efficiency decreases with the homology loss between the target sequence and the oligo (Wang et al. 2009). Furthermore, the strain used to perform these mutations lacks the mismatch repair system encoded by the gene *mutS* to avoid the reparation of the mutation back to the wild-type sequence, especially when incorporating point mutations through mismatches. Nevertheless, such a knock-out of the *mutS* gene increases the mutation rate from 10^{-10} to 10^{-8}, leading inevitably to off-target mutations (Wang and Church 2011).

Later, this method was further developed by Nyerges et al. (2016) by the construction of a plasmid named pORTMAGE which is a system that can be introduced in any cells to incorporate mutations without the requirement of prior modification (*mutS* deletion). pORTMAGE is a plasmid expressing the λ-red recombination system as well as a dominant-negative allele of *mutL*, placed under a temperature-sensitive promoter to prevent the expression of the mismatch repair complex while the oligonucleotide harboring the mutation is incorporated. Such a transitory system decreases the number of off-target mutations compared to the permanent deletion of the mismatch repair machinery (Nyerges et al. 2016).

While this technology is very useful to incorporate point mutations easily in the chromosome, it is still limiting in terms of covered windows as the oligo is 90 bp long and a minimum of 15 bp long homology sequence should be left on each side of the desired mutation due to the degradation of extremities during integration (Wang and Church 2011). In addition, even if a library of oligonucleotides can be used to incorporate many mutations at a desired position in the genome, this method is rather expensive due to the phosphorothioated modification at the extremities of the oligonucleotides coupled to the long size (90 bp) necessary. As a result, this method is better suited for the incorporation of pre-determined point mutations than large-scale mutagenesis efforts.

2. T7 RNA polymerase methods

In the last few years, several research papers developed methods using the T7 RNA polymerase (T7RNAP) from the T7 bacteriophage fused with a cytidine deaminase.

The first system proposed by Moore, Papa, and Shoulders (2018) was called MutaT7 and consisted of the insertion of mutations using a cytidine deaminase (rApoBEC1) on plasmid DNA sequences downstream of the P_{T7} promoter making this system specific to a target gene or region. A cytidine deaminase converts cytidine to uridine in the DNA sequence that will pair with adenosine, ultimately leading to C → T transition. The deletion of the *ung* gene in the *E. coli* host strain, encoding an uracil N-DNA glycosylase, is beneficial to the mutagenesis efficiency by preventing the repair of the uridine bases introduced by the deaminase (Ramiro et al. 2003). This method was successful in obtaining mutants resistant to kanamycin and specific to the presence of the T7 promoter upstream of the resistance gene,

which is an important feature to avoid off-target mutations. However, the mutation rate remained low with 0.34 mutations day^{-1} kb^{-1}. Due to the high processivity of the T7 RNAP, authors pointed out the high importance of the insertion of multiple terminators at the end of the targeted locus, and the number determined for efficiently stopping the MutaT7 system was four terminators in a row (Moore, Papa, and Shoulders 2018).

From this MutaT7 tool developed, another research team developed an enhanced version, eMutaT7. To achieve higher efficiency, another cytosine deaminase, PmCDA1 was used. Following their protocol, they were able to report an increased mutation rate of ~ 4 mutations day^{-1} kb^{-1} (Park and Kim 2021).

Another approach based on the same principle, designated as T7-DIVA, involves T7RNAP with base deaminases, both cytosine and adenosine, but using gRNA/dCas9 complex to block the elongation from the T7RNAP in place of the four terminators employed by the previous system (Álvarez et al. 2020). The authors placed the P_{T7} promoter at the 3' end of the target gene, in *E. coli* chromosome, to preserve the endogenous expression of this gene from the host promoter and validated their system by evolving an antibiotic resistance gene *TEM-1*. By its design where the gRNA for the dCas9 that will block the base deaminase activity can be chosen at the extremity of a gene or in the middle of the sequence, the approach brings flexibility. When compared to the previously described method of MutaT7 which needed four terminators to stop the mutagenesis systems and that can obviously not be introduced in the middle of a gene without disrupting the sequence, this innovative approach can be interesting in specific contexts. However, for our objectives of mutations across an entire gene, the use of this method does not bring major advantages compared to the above-mentioned method with terminators. As the objective is the use of this method for targeting heterologous genes, the need for prior gene insertion in the chromosome for each mutagenesis experiment is a clear disadvantage. Another interesting aspect of the Álvarez et al. (2020) study is the use of a diversity of base deaminases: AID, rAPOBEC1, pmCDA1 and TadA* that proves to be efficient when fused with the T7RNAP.

An additional interesting system, TRIDENT, which stands for TaRgeted In vivo Diversification ENabled by T7 RNAP, was developed in *S. cerevisiae* and involved a cytosine deaminase, as the other systems described as well as an adenosine deaminase and DNA-repair factors to enhance mutational diversity, making the originality of the system (Cravens et al. 2021). TRIDENT showed higher specificity with lower off-target mutations and was efficient in introducing transitions C↔T, G↔A and more singularly, T→G transversion.

Therefore, the method called Evolution.T7 described in the following paragraph that offers the possibility of mutations across the entire gene looked better suited for our problematics.

3. Evolution.T7

Evolution.T7 is a method developed by the iGEM 2021 team from Evry and led by Dr. Ioana Popescu. While the method they developed is not published, yet, the website they created for the description of their project is used as a reference (iGEM 2021 Evry Paris-Saclay).

This approach is a combination of the fused T7RNAP with a base deaminase (cytosine and adenosine, Figure 13) and an innovative design that will produce mutations on both strands of the plasmid DNA using two distinct promoters flanking the region to evolve. Indeed, the systems described in the previous section produce mutation with a strong bias towards the non-template strand, probably due to the lack of accessibility of the template strand during transcription (Álvarez et al. 2020). Thus, Evolution.T7 aims at increasing the mutation rate and reducing bias compared to other systems that are performed on only one strand of the DNA. These two promoters, P_{T7} and P_{T7CGG}, are recognized by the native T7RNAP and a modified T7RNAP, T7RNAP-CGG-R12-KIRV, specific to this mutated promoter, respectively. Moore, Papa, and Shoulders (2018) tested a similar approach by flanking the target gene with P_{T7} promoters on both extremities but this resulted in a strong bias towards C to T compared to G to A transitions, with means of 2.7 mutations and 1.6 mutations respectively per clone after 12 days of continuous culture.

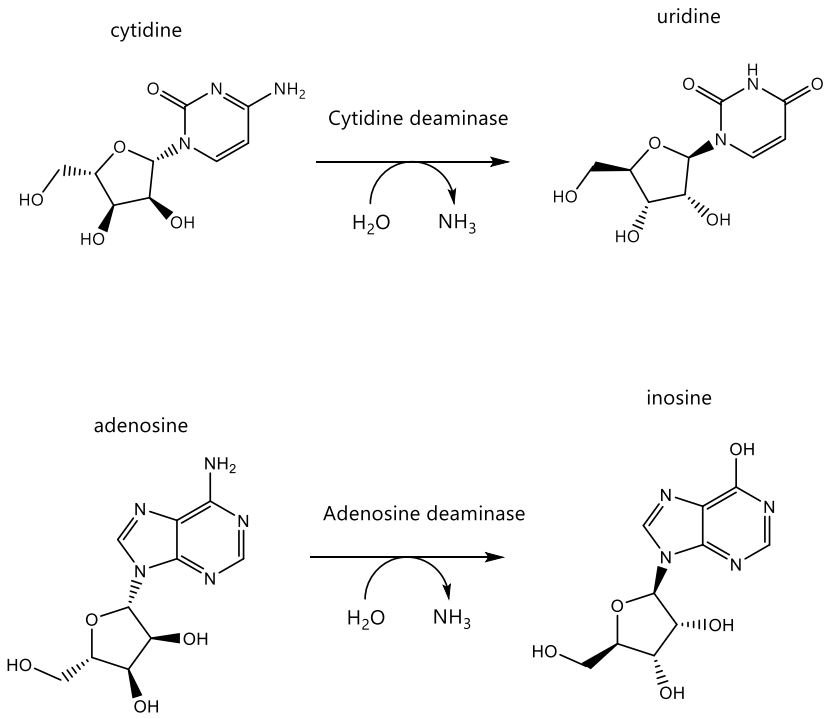


Figure 13. Cytidine and adenosine deaminases as a source of mutations. Uridine produced by cytosine deaminases will pair with adenosine, leading to C → T transition. Inosine produced by adenosine deaminases will pair with cytidine, leading to A → G transition.

In the Evolution.T7 tool design (Figure 14), the two T7RNAP are inducible using two distinct inducers (tetracycline and L-arabinose). This allows for the activation of the transcription and the mutagenesis sequentially to avoid any collision between the two polymerases which is hypothesized to be partly responsible for an uneven distribution of mutations along the gene (iGEM 2021 Evry Paris-Saclay). As demonstrated by Moore, Papa, and Shoulders (2018), the number of terminators to efficiently block the highly processive T7RNAP is four, a finding that was used for Evolution.T7 as well. In addition, an *E. coli* strain Δ*ung* Δ*nfi* was used for these experiments, *ung* was described in the previous section (I. 6. 3. 2) and *nfi* encodes the endonucleases V that eliminates inosine bases (Vik et al. 2013), bases created by adenosine deaminases (Figure 13).

Figure 14. Evolution T7 design. BD = base deaminase. P_{T7*} designates P_{T7CGG} and T7RNAP* is T7RNAP-CGG-R12-KIRV. Adapted from iGEM 2021 Evry Paris-Saclay website.

To enlarge the T7RNAP/deaminase-based toolbox, the iGEM team used a broader range of base deaminases compared to previously described systems, bringing it to eight enzymes: AID, pmCDA1, rAPOBEC1, TadA*, ABE8.20m, evoAPOBEC1-BE4max, evoCDA1-BE4max. With this variety of base deaminases and the two T7RNAP variants, they explored a total of 64 combinations to identify the most efficient ones.

The iGEM team used a similar proof of concept approach as other mutagenesis development studies to validate their system. They employed an antibiotic resistance gene specific to ampicillin and part of the β-lactamases enzyme family, as their target for mutagenesis. The goal was to select mutants of this enzyme capable of conferring antibiotic resistance to aztreonam, another β-lactam, antibiotic for which the gene to evolve has less affinity compared to ampicillin. After inducing their *in vivo* mutagenesis system for 24 hours and applying a selection protocol, they successfully selected *E. coli* mutants resistant to aztreonam. Sequencing confirmed the presence of mutations in the targeted gene, resulting in aztreonam resistance from the mutated AmpR gene (iGEM 2021 Evry Paris-Saclay).

Using both wet lab experiments and mathematical modeling, they selected the five most efficient pairs of T7RNAP/deaminases. Thus, these specific combinations were used in the protocol applied for this PhD project, which is described in detail in the Results chapter (III. 5. 1).

Importantly, the techniques described in this section proved to be performant in the evolution of antibiotic resistance genes. The screening process is relatively straightforward with this type of gene, which is significantly different compared to the objective of our project, which is the evolution of a key enzyme of carbon assimilation. The challenge of directed evolution using mutagenesis lies in the difficulty of finding an efficient *in vivo* screening test and the lack of published examples demonstrating the successful application of these methods indicates the challenge it may be.

7. Objectives of the presented work

As presented in this introduction, the PhD project was conceived to match with the specific context of carbon cycle closure. In the frame of this context, the research work aimed to demonstrate the direct reduction of CO_2 in the versatile platform strain *E. coli* by employing enzymatic systems that are recognized for their biotechnological potential but for which experimental demonstration of *in vivo* activity is lacking. The questions that were explored were diverse:

- Can a complex metal-dependent FDH, such as the one from *C. necator* (CnFDH), be stably active in *E. coli* cells depending on its formate-oxidation activity under aerobic conditions?
- Is the CnFDH able to reduce CO_2 to formate *in vivo*, as it was demonstrated *in vitro*, when expressed in a metabolically engineered formate-dependent *E. coli* strains?
- Can a non-carbon energy-generating system, using hydrogen or phosphite as electron donors, be efficiently implemented in *E. coli* cells?
- Can an *in vivo* mutagenesis method be used for the generation and selection of favorable enzyme variants directly in metabolic screening strains?

II. Material and methods

1. Strains and plasmids

1. Strains

All the strains used for this project are described in Table 4.

Table 4. *E. coli* strains used for this study.

Strain	Genotype	Reference
MG1655	*λ- rph-1*	CGSC Collection, Yale
BW19610	*(lac)X74 ΔuidA::pir-116 recA1 ΔphoA532 Δ(phnC?DEFGHIJKLMNOIP)33-30*	Metcalf, Jiang, and Wanner (1994)
DH5α	*supE44 ΔlacU169 (Φ80lacZ ΔM15) ΔardF hsdR17 recA1 endA1 gyrA96 thi-1 relA1*	Laboratory collection
DH5αPir	*supE44 ΔlacU169 (Φ80lacZ ΔM15) ΔardF hsdR17 recA1 endA1 gyrA96 thi-1 relA1 uidA::pir+*	Platt et al. (2000)
β2033	*pro thi rpsL hsds ΔlacZ* *F'(ΔlacZ M15 laclq traD36 proA⁺ proB⁺)*	Laboratory collection
ST18	pro thi hsdR+ Tp^r Sm^r; chromosome::RP4-2 Tc::Mu-Kan::Tn7/λpir Δ*hemA*	Max Planck Institute
+2440	pKD46 *ampR*	Laboratory collection
JW0112	Δ*lpd*::*kanR*	Baba et al. (2006)
JW0374	Δ*phoA*::*kanR*	Baba et al. (2006)
JW0855	Δ*poxB*::*kanR*	Baba et al. (2006)
JW0889	ΔycaP::*kanR*	Baba et al. (2006)
JW0955	Δ*hyaB*::*kanR*	Baba et al. (2006)
JW1228	Δ*adhE*::*kanR*	Baba et al. (2006)
JW2278	Δ*nuoG*::*kanR*	Baba et al. (2006)
JW2472	Δ*hyfG*::*kanR*	Baba et al. (2006)
JW2515	Δ*iscR*:: *kanR*	Baba et al. (2006)
JW2564	Δ*ung*:: *kanR*	Baba et al. (2006)
JW2691	Δ*hycE*:: *kanR*	Baba et al. (2006)
JW2962	Δ*hybC*::*kanR*	Baba et al. (2006)
JW3865	Δ*fdoG*::*kanR*	Baba et al. (2006)
JW4060	Δ*phnI*::*kanR*	Baba et al. (2006)
JW5547	Δ*nfi*:: *kanR*	Baba et al. (2006)
E491	Δ*flu* Δ*pyrF* Δ*ung* Δ*nfi*	Gift from Ioana Popescu

G3806	Evolved strain Δ*glyA::aad* Δ(*tdh-kbl*) Δ*ltaE* Δ*gcvTHP::erm* Δ*gcvR* Δ*kdgK*::*ftfL*(*C.kluyveri*)+ kan+	(Döring et al. 2018) Laboratory collection
G4248	Derived from G3806 Δ*gcv::erm* Δ(*tdh-kbl*) Δ*ltaE* Δ*gcvR* Δ*kdgK* :: *ftfL* (*C. kluyveri*)+ Δ*serA*	Laboratory collection
G4463	Derived from G3806 - Evolved strain Δ*glyA*::aad Δ(*tdh-kbl*) Δ*ltaE* Δ*gcvR* Δ*kdgK*::ftfL (*C. kluyveri*)+ Δ*aceBAK*	(Delmas et al. 2022) Laboratory collection
G5416	Δ*lpd*::FRT	Laboratory collection
G5620	Δ*hyaB* Δ*hybC* Δ*hycE* Δ*hyfG*	This study
G5626	Δ*phoA* ΔphnI	This study
G5654	Δ*phoA* ΔphnI; pGEN1325 *ptxD*	This study
G5656	Δ*phoA* ΔphnI; pGEN1337 *ptxA ptxB ptxC*	This study
G5660	Δ*phoA* ΔphnI; pGEN1325 *ptxD*; pGEN1337 *ptxA ptxB ptxC*	This study
G5658	Δ*lpd* Δ*phoA* Δ*phnI*	This study
G5663	Δ*lpd*; pGEN1340 amp+ *fdsA fdsB fdsG fdsC fdsD*	This study
G5667	Δ*lpd* Δ*phoA* ΔphnI; pGEN1325::*ptxD* (*P. stutzeri*)	This study
G5684	Δ*phoA* Δ*phnGHIJK*	This study
G5726	Δ*lpd*; pGEN1320 cm+ *fdsC fdsD* (codon optimized); pGEN1323 amp^{+} *fdsA fdsB fdsG* (codon optimized)	This study
G5730	Δ*lpd*; pGEN1348 cm+ *fdsC fdsD* ; pGEN1338 amp+ *fdsA fdsB fdsG*	This study
G5743	Δ*glyA*::aad Δ(*tdh-kbl*)+ Δ*ltaE* Δ*gcvR* Δ*kdgK*::ftfL (*C. kluyveri*)+ Δ*aceBAK* ; pGEN1340 CnFDH	This study
G5785	Δ*gcv*::erm Δ(*tdh-kbl*) Δ*ltaE* Δ*gcvR* Δ*kdgK* :: *ftfL* (*C. kluyveri*)+ Δ*serA* ; pGEN1340 CnFDH	This study
G5791	Δ*phoA* ΔphnI Δlpd *kdgK* :: *ptxD*	This study
G5823	Evolved G5663 strain (UOF1) Δ*lpd*; pGEN1340 bla+ *fdsA fdsB fdsG fdsC fdsD* (*C. necator*)	This study
G5824	Evolved G5663 strain (UOF1) Δ*lpd*; pGEN1340 bla+ *fdsA fdsB fdsG fdsC fdsD* (*C. necator*)	This study
G5825	Evolved G5663 strain (UOF1) Δ*lpd*; pGEN1340 bla+ *fdsA fdsB fdsG fdsC fdsD* (*C. necator*)	This study
G5848	Evolved G5663 strain (UOF2) Δ*lpd*; pGEN1340 bla+ *fdsA fdsB fdsG fdsC fdsD* (*C. necator*)	This study
G5849	Evolved G5663 strain (UOF2) Δ*lpd*; pGEN1340 bla^{+} *fdsA fdsB fdsG fdsC fdsD* (*C. necator*)	This study
G5850	Evolved G5663 strain (UOF2)	This study

	Δ*lpd*; pGEN1340 amp$^+$ *fdsGBACD* (*C. necator*)	
G5876	Evolved strain G5823 cured Δ*lpd*	This study
G5956	Δ*hyaB* Δ*hybC* Δ*hycE* Δ*hyfG* Δ*iscR*	This study
G5968	Δ*hyaB* Δ*hybC* Δ*hycE* Δ*hyfG* Δ*lpd*	This study
G5984	Δ*hyaB* Δ*hybC* Δ*hycE* Δ*hyfG* Δ*iscR* Δ*adhE*	This study
G6012	IS10::CnFDH kan+ (Δ62 bp in *fdsG*)	This study
G6085	Δ*lpd* IS10::CnFDH (Δ62 bp in *fdsG*)	This study
G6106	Δ*hyaB* Δ*hybC* Δ*hycE* Δ*hyfG* Δ*lpd* Δ*iscR*	This study
G6114	Δ*lpd*; pZE-CnFDH (pGEN1393) spec$^+$	This study
G6165	Δ*flu* Δ*pyrF* Δ*ung* Δ*nfi;* pGEN1410	This study
G6178	Δflu ΔpyrF Δung Δnfi; p112; p119; pGEN1410	This study
G6217	Δ*lpd*; pZE-TsFDH (pGEN1395) spec$^+$	This study
G6219	Δflu ΔpyrF Δung Δnfi; p115; p118; pGEN1410	This study
G6221	Δflu ΔpyrF Δung Δnfi; p113; p103; pGEN1410	This study
G6223	Δflu ΔpyrF Δung Δnfi; p111; p119; pGEN1410	This study
G6234	Evolved strain – G5823 (mutated *focA* reverted) Δ*lpd*; Δ*ycaP*::kan+; pGEN1340	This study
G6235	Evolved strain – G5848 (mutated *focA* reverted) Δ*lpd*; Δ*ycaP*::kan+; pGEN1340	This study
G6252	Δ*lpd* Δ*poxB*	This study
G6258	Evolved strain (OCF 4) Δ*lpd*; IS10::CnFDH (Δ63 bp in *fdsG*)	This study
G6266	Evolved strain (OCF 3) Δ*lpd*; IS10::CnFDH (Δ63 bp in *fdsG*)	This study
G6272	Evolved strain G5823 cured Δ*lpd*; pZE-TsFDH (pGEN1395) spec$^+$	This study
G6280	Evolved strain G5823 cured Δ*lpd*; pZE-CnFDH (pGEN1393) spec$^+$	This study
G6282	Δ*lpd* Δ*poxB*; pTRC-CnFDH (pGEN1340) amp$^+$	This study
G6284	Δ*lpd* Δ*poxB*; pZE-TsFDH (pGEN1395) spec$^+$	This study
G6304	Δ*hyaB* Δ*hybC* Δ*hycE* Δ*hyfG* Δ*lpd* Δ*iscR* Δ*adhE*	This study
G6384	Δ*glyA*::aad Δ(*tdh-kbl*)+ Δ*ltaE* Δ*gcvR* Δ*kdgK*::ftfL (*C. kluyveri*)+ Δ*aceBAK* Δ*nuoG*	This study
G6385	Δ*gcv*::erm Δ(*tdh-kbl*)+ Δ*ltaE* Δ*gcvR* Δ*kdgK* :: *ftfL* (*C. kluyveri*)+ Δ*serA* Δ*nuoG*	This study

G6403	Evolved strain G5823 cured Δ*lpd*; pGEN1425 CnFDH (Δ63 bp in *fdsG*)	This study
G6405	Evolved strain G5823 cured Δ*lpd*; pGEN1426 CnFDH (Δ62 bp in *fdsG*)	This study
G6408	IS10::CnFDH kan$^+$	This study
G6435	Δ*lpd*; IS10::CnFDH	This study
G6440	Δ*gcv*::erm Δ(*tdh-kbl*)+ Δ*ltaE* Δ*gcvR* Δ*kdgK* :: *ftfL* (*C. kluyveri*)+ FRT+ Δ*serA* Δ*fdoG*	This study
G6441	Δ*gcv*::erm Δ(*tdh-kbl*) Δ*ltaE* Δ*gcvR* Δ*kdgK* :: *ftfL* (*C. kluyveri*)+ Δ*serA* Δ*fdoG* Δ*nuoG*	This study
G6446	Δ*glyA*::aad Δ(*tdh-kbl*) Δ*ltaE* Δ*gcvR* Δ*kdgK*::ftfL (*C. kluyveri*)+ Δ*aceBAK* Δ*fdoG*	This study
G6447	Δ*glyA*::aad Δ(*tdh-kbl*) Δ*ltaE* Δ*gcvR* Δ*kdgK*::ftfL (*C. kluyveri*)+ Δ*aceBAK* Δ*fdoG* Δ*nuoG*	This study
G6475	Δ*gcv*::erm Δ(*tdh-kbl*) Δ*ltaE* Δ*gcvR* Δ*kdgK* :: *ftfL* (*C. kluyveri*)+ Δ*serA* Δ*fdoG* IS10::CnFDH kan$^+$	This study
G6476	Δ*lpd* Δ*nfi* Δ*ung*	This study
G6488	Δ*gcv*::erm Δ(*tdh-kbl*)+ Δ*ltaE* Δ*gcvR* Δ*kdgK* :: *ftfL* (*C. kluyveri*)+ Δ*serA* Δ*fdoG* Δ*nuoG* IS10::CnFDH kan$^+$	This study
G6493	Δ*gcv*::erm Δ(*tdh-kbl*) Δ*ltaE* Δ*gcvR* Δ*kdgK* :: *ftfL* (*C. kluyveri*)+ FRT+ Δ*serA* Δ*fdoG*; pGEN1340::CnFDH	This study
G6495	Δ*gcv*::erm Δ(*tdh-kbl*) Δ*ltaE* Δ*gcvR* Δ*kdgK* :: *ftfL* (*C. kluyveri*)+ FRT+ Δ*serA*+ Δ*fdoG* Δ*nuoG*; pGEN1340::CnFDH	This study
G6497	Δ*glyA*::aad Δ(*tdh-kbl*) Δ*ltaE* Δ*gcvR* Δ*kdgK*::ftfL (*C. kluyveri*)+ Δ*aceBAK* Δ*fdoG* ; pGEN1340::CnFDH	This study
G6499	Δ*glyA*::aad Δ(*tdh-kbl*) Δ*ltaE* Δ*gcvR* Δ*kdgK*::ftfL (*C. kluyveri*)+ Δ*aceBAK* Δ*fdoG* Δ*nuoG* ; pGEN1340::CnFDH	This study
G6511	Δ*lpd* Δ*nfi* Δ*ung*; p114; p119; pGEN1410	This study
G6512	Δ*glyA*::aad Δ(*tdh-kbl*) Δ*ltaE* Δ*gcvR* Δ*kdgK*::ftfL (*C. kluyveri*)+ Δ*aceBAK* Δ*fdoG* Δ*nuoG* IS10::CnFDH kan$^+$	This study
G6539	Evolved G6435 strain (OCF5) Δ*lpd* IS10::CnFDH	This study
G6540	Evolved G6435 strain (OCF5) Δ*lpd* IS10::CnFDH	This study
G6541	Evolved G6435 strain (OCF6) Δ*lpd* IS10::CnFDH	This study
G6542	Evolved G6435 strain (OCF6) Δ*lpd* IS10::CnFDH	This study

2. Plasmids

All the plasmids used for this project are described in Table 5.

Table 5. Plasmids used for this study.

Plasmid	Description	Reference
pSW23	$oriV_{R6K\gamma}$; Cm^+	Demarre et al. (2005)
pTRC99	ColE1ori; Amp^+	Amann, Ochs, and Abel (1988)
pAV10	pSC101 ori; P_{PhlF}-Cas9-gRNA; $Spec^+$	Gift from David Bikard (Pasteur Institute)
pUC18	pBR322 ori; Amp^+	Norrander, Kempe, and Messing (1983)
pSU18	p15Aori; Cm^+	Männistö et al. (2003)
pSEVA221	oriV; kan^+	Gift from Ioana Popescu
pSEVA471	pSC101ori; $Spec^+$	Gift from Ioana Popescu
pSEVA721	oriV; Tmp^+	Gift from Ioana Popescu
pCP20	FLP recombinase; Amp^+	Cherepanov and Wackernagel (1995)
pKD46	λ-red recombinase (γ, β, exo genes); Amp^+	Datsenko and Wanner (2000)
pDM4_IS10	$ori_{R6K\gamma}$; Plasmid for chromosomal integration at IS10 locus (*ula* operon) in *E. coli*; Cm^+; Kan^+	Gift from Wenk et al. (2018)
pFD152	d*cas9*; $spec^+$	Depardieu and Bikard (2020)
pEVL410	pSW23-500 bp *kdgK* regions upstream and downstream	Laboratory collection
pTHfds05	pTRC99-His-*fdsGBACD* (*R. capsulatus*); Amp^+	Gift from Hartmann and Leimkühler (2013)
pQE80-SH	*hoxF*, *hoxU*, *hoxY*, *hoxH*, *hoxW*, *hoxI* (*C. necator*, codon optimized); Amp^+	Gift from Lamont and Sargent (2017)
pSUhypA2-X	pSU-*hypA2*, *hypB2*, *hypF2*, *hypC1*, *hypD1*, *hypE1*, *hypX* (*C. necator*, codon optimized); Kan^+	Gift from Lamont and Sargent (2017)
pFDH	pZE21; pBR322 ori; Spec+; *fdh Pseudomonas sp101*	Addgene #131706 (Gleizer et al. 2019)
p103	pSEVA221-rAPOBEC1-T7RNAP; kan^+	Gift from Ioana Popescu
p111	pSEVA471-pmCDA1-T7RNAP-CGG-R12-KIRV; $spec^+$	Gift from Ioana Popescu
p112	pSEVA471-rAPOBEC1-T7RNAP-CGG-R12-KIRV; $spec^+$	Gift from Ioana Popescu
p113	pSEVA471-TadA*-T7RNAP-CGG-R12-KIRV; $spec^+$	Gift from Ioana Popescu
p114	pSEVA471-evoAPOBEC1-BE4max-T7RNAP-CGG-R12-KIRV; $spec^+$	Gift from Ioana Popescu

p115	pSEVA471-evoCDA1-T7RNAP-CGG-R12-KIRV; $spec^+$	Gift from Ioana Popescu
p118	pSEVA221-evoCDA1-T7RNAP; kan^+	Gift from Ioana Popescu
p119	pSEVA221::ABE8.20- m-T7RNAP; kan^+	Gift from Ioana Popescu
pGEN1320	pSU18::*fdsC fdsD* (*C. necator,* codon optimized); Cm^+	This study
pGEN1323	pUC18::*fdsA fdsB fdsG* (*C. necator,* codon optimized); Amp^+	This study
pGEN1325	pUC18::*ptxD (P. stutzeri,* codon optimized); Amp^+	This study
pGEN1337	pSU18::*ptxA ptxB ptxC* (*P. stutzeri*, codon optimized); Cm^+	This study
pGEN1338	pUC18::*fdsA fdsB fdsG* (*C. necator*); Amp^+	This study
pGEN1340	pTRC99::*fdsGBACD* (*C. necator*); Amp^+	This study
pGEN1348	pSU18::fdsC fdsD (*C. necator*); Cm^+	This study
pGEN1353	pTRC99::His-*fdsGBACD* (*C. necator*); Amp^+	This study
pGEN1378	pDM4_IS10::*fdsGBACD* (*C. necator*); Cm^+; kan^+	This study
pGEN1393	pZE21::*fdsGBACD* (*C. necator*); $Spec^+$	This study
pGEN1395	pZE21::*fdh* (*Thiobacillus sp.*); $Spec^+$	This study
pGEN1410	pSEVA721::J23110_4ter_P-moderate_RBS-B_TsFDHopt; Tmp^+	This study
pGEN1416	pFD152::gRNA1; $Spec^+$	This study
pGEN1418	pFD152::gRNA2; $Spec^+$	This study
pGEN1420	pFD152::gRNA3; $Spec^+$	This study
pGEN1422	pFD152::gRNA4; $Spec^+$	This study
pGEN1425	pTRC::*fdsGBACD* (Δ63 bp in *fdsG*); Amp^+	This study
pGEN1426	pTRC::*fdsGBACD* (Δ62 bp in *fdsG*); Amp^+	This study
pGEN1437	pSW23::1000 bp *IS10* regions upstream and downstream TsFDH	This study
pGEN1444	pFD152::gRNA5; $Spec^+$	This study
pGEN1445	pFD152::gRNA6; $Spec^+$	This study
pGEN1446	pFD152::gRNA7; $Spec^+$	This study
pGEN1447	pFD152::gRNA8; $Spec^+$	This study

2. Culture media

The *E. coli* K-12 strain was cultivated in both rich and mineral media as describes in the subsequent sections. In the case of each medium (excluding SOC and TB), a 2X stock solution was prepared and solidified using 15 g/L agar for solid cultures on Petri dishes, if needed. Chemicals were purchased from Sigma-Aldrich.

1. Rich media

Various nutrient-rich media were used for the experimental protocols, and their specific compositions can be found in Table 6. These media include lysogeny broth (LB), super optimal broth with catabolite repression (SOC), and terrific broth (TB). The utilization of each medium is described in relation to the respective experiment.

Table 6. SOC, LB and TB media composition.

Compound	SOC	LB	TB
Tryptone	20 g	10 g	12 g
Yeast extract	5 g	5 g	24 g
NaCl	0.5 g	10 g	-
KCl	0.186 g	-	-
$MgCl_2$	0.952 g	-	-
$MgSO_4$	1.204 g	-	-
K_2HPO_4	-	-	12.5 g
KH_2PO_4	-	-	2.3 g
Glycerol	-	-	4 mL
Glucose	3.6 g	-	-
ddH_2O	Up to 1 L	Up to 1 L	Up to 1 L

M9ZB medium (Table 7) was exclusively used for tests involving hydrogenase from *C. necator*.

Table 7. M9ZB medium preparation.

Compound	Description	Amount
ZB	10 g NZ-Amine A + 5 g NaCl up to 860 mL sterile water	860 mL
10X M9 salts	1 g NH_4Cl + 3 g KH_2PO_4 + 6 g Na_2HPO_4 up to 100 mL sterile water	100 mL
10 % glucose		40 mL
$MgSO_4$ 1 M		1 mL

2. Mineral media

Two types of mineral medium were used for *E. coli* K-12 cultures: Murashige and Skoog medium (MS medium) and morpholinopropane sulfonate medium (MOPS). These media were used for growth tests and supplemented with various carbon and energy sources.

For the preparation of MS minimal medium in liquid culture, the formulation outlined in Table 8 was followed.

Table 8. Mineral medium Murashige and Skoog composition.

Compound	Amount
Citric acid ($C_6H_8O_7$, H_2O)	0.84 g
NH_4Cl	1.06 g
K_2HPO_4, 3 H_2O or K_2HPO_4 anhydrous	11.41 g 8.75 g
NTA mix (see composition Table 11)	1 ml
ddH_2O	Up to 1 L

For the functional tests of phosphite dehydrogenase, a specific minimal medium lacking phosphate was employed, MOPS. For its preparation, a 10X MOPS solution (as specified in Table 9) was sterilized through filtration and stored at -20 °C. Subsequently, the MOPS minimal medium was prepared as detailed in Table 10.

Table 9. MOPS 10X solution.

Compound	Amount
MOPS (potassium morpholinopropane sulfonate)	83.72 g
Tricine ($C_6H_{13}NO_5$)	7.17 g
$FeSO_4$-7H_2O	0.028 g
NH_4Cl	5.08 g
K_2SO_4	0.48 g
$MgCl_2$	1.067 g
NaCl	29.22 g
NTA mix (see composition Table 11)	10 ml
ddH_2O	Up to 1 L

Table 10. MOPS minimal medium.

Compound	Amount
10X MOPS solution	100 mL
0.132 M K_2HPO_4	10 mL
ddH_2O	890 mL

Finally, a medium was especially designed for the metabolic screening the CnHyd activity and its composition is detailed in the Table 12 below.

Table 11. Nitrilotriacetic acid (NTA) mix composition.

Compound	Amount
Nitrilotriacetic acid ($C_6H_9NO_6$)	1,91 g
$CaCl_2$, 2 H_2O	0,45 g
$FeCl_3$, 6 H_2O	0,81 g
$MnCl_2$, 4 H_2O	0,20 g
$ZnCl_2$	40 mg
H_3BO_3	20 mg
$CrCl_3$, 6 H_2O	80 mg
$CoCl_2$, 6 H_2O	70 mg
$CuCl_2$, 2 H_2O	50 mg
Ni_2Cl, 6 H_2O	70 mg
Na_2MoO_4, 2 H_2O	70 mg
Na_2SeO_3, 5 H_2O	80 mg
ddH_2O	Up to 1 L

Table 12. Medium composition for the metabolic test of hydrogen as a source of NADH.

Compound	Finale concentration
M9 salt	1 X
Acetate	20 mM
Pyruvate	20 mM
$MgCl_2$	2 mM
$CaCl_2$	0.1 mM
Thiamine	0.10 %
$FeCl_2$	0.10 %
Fe^{2+}	100 µM
$NiCl_2$	100 µM
IPTG	500 µM

3. Stock solutions

Rich and mineral medium described above were supplemented with various compounds that are described in the Table 13.

Table 13. Stock solutions. The final concentrations are indicated as references but may vary considerably depending on the experiments.

Component	Stock solution	Solvent	Final concentration
Glucose	20 %	ddH_2O	0.2 %
Acetate	2 M	ddH_2O	20 mM
Pyruvate	2 M	ddH_2O	20 mM
Formate	5 M	ddH_2O	60 mM
Glycerol	20 %	ddH_2O	0.2 %
Phosphite	2 M	ddH_2O	10-80 mM
IPTG	1 M	ddH_2O	100 µM
Glycine	2 M	ddH_2O	2 mM
Serine	2 M	ddH_2O	2 mM
aTc	1 $mg.mL^{-1}$	Ethanol	0.5 $µg.mL^{-1}$
DAPG	250 mM	Ethanol	25 µM
Citrate	1 M	ddH_2O	0.1 M
L-arabinose	1 M	ddH_2O	10 mM

4. Antibiotics

Antibiotic solutions used for this study are detailed in Table 14. The final concentrations are indicated as references but may vary depending on the experiments.

Table 14. Antibiotic stock solutions and usage concentrations. Carbenicillin is used instead of Ampicillin due do its superior stability.

Antibiotic	Stock solution	Solvent	Final concentration
Carbenicillin	100 $mg.ml^{-1}$	ddH_2O	100 $µg.ml^{-1}$
Kanamycin	30 $mg.ml^{-1}$	ddH_2O	30 $µg.ml^{-1}$
Chloramphenicol	25 $mg.ml^{-1}$	EtOH	25 $µg.ml^{-1}$
Spectinomycin	100 $mg.ml^{-1}$	ddH_2O	100 $µg.ml^{-1}$
Trimethoprim	50 $mg.ml^{-1}$	DMF	10 $µg.ml^{-1}$

3. Culture conditions

1. Solid cultures

Solid cultures were used for colony isolation and metabolic tests. These cultures were performed using 2X culture media (LB or MS) described in section II. 2. with 15 g/L agar for Petri dishes preparation. Plates were supplemented with antibiotic and other supplements at the concentration described in Table 13 and Table 14, unless stated differently. Plates were incubated at 30 or 37 °C depending on the strain and the experiment.

For experiments that required CO_2-enriched atmosphere conditions, a specific incubator with a CO_2-injection system was used. This allowed to control CO_2-concentration in the incubator atmosphere in a range from 5 to 20 % of CO_2 in 95 to 80 % air.

2. Liquid cultures

1. Cultures under aerobic conditions

Liquid cultures were conducted either in sterile tubes or Erlenmeyer flasks, depending on the specific requirements and volume needed for each experiment. Erlenmeyer flasks were equipped with porous pads to ensure proper air exchange while maintaining sterile conditions.

For liquid culture in rich media, a single colony was harvested to inoculate the culture. For minimal media cultures, biomass from an adaptation Petri dish was utilized to inoculate the liquid medium. Subsequently, the cultures were placed in a thermostated incubator under agitation, at 30 or 37 °C, based on the specific experimental requirements.

2. Culture under CO_2-enriched atmosphere

1. *Minitron*

For liquid culture CO_2 reduction experiments, a specific device to set CO_2-enriched atmospheric conditions was used, the Minitron (INFORS HT). This equipment enabled to control and maintain CO_2 concentrations within the incubator, with a range spanning from 0 to 20 % in air.

Cultures were performed in test tubes of 15 mL with 3 mL of culture of 50 mL with 5 mL of culture with a porous pad in place of the cap to favor air-exchange and are placed in the Minitron incubator. The CO_2 content was adjusted depending on the experiment. The humidity-control device available in the

incubator was used to prevent evaporation as the culture were only sealed with porous pads to promote gas exchanges between the culture and the atmosphere. In addition, the temperature inside the Minitron was also set and controlled to 30 °C.

2. *Flask with CO_2 injection*

Considering the inherent limitation of the Minitron for the CO_2 range available, additional tests were conducted using hermetic flasks for CO_2 concentrations above the 20 % threshold, up to 50 %. 100 mL volume flasks were inoculated with 10 mL liquid media and a pre-culture at an initial OD_{600} of 0.05. A specific volume of CO_2 was injected in the flask using a syringe to implement the desired atmosphere or the flask was flushed with a mix of 50 % CO_2 in air. Flasks were then incubated in a shaking incubator at 30 °C.

3. Anaerobic culture

For anaerobic *E. coli* cultures, pre-cultures in a semi-anaerobic regime were performed in 2 mL Eppendorf tube filled with 2 mL of the desired liquid medium inoculated with a single colony without agitation. In parallel, Hungate tubes were prepared with 3 mL of the culture medium and sealed hermetically before establishing a vacuum. The next day, prepared Hungate tubes were inoculated with 100 µL of semi-anaerobic overnight pre-culture. Tubes were flushed with N_2 to ensure anaerobic growth condition. Anaerobic cultures were incubated with or without agitation at the set temperature (30 or 37 °C) depending on the strains and the test.

3. GM3

Continuous culture evolution experiments were conducted using the GM3 self-cleaning automated device (Mutzel and Marliere 2004). This device facilitates prolonged cultivation by diluting the culture with a predetermined volume of fresh medium while simultaneously removing an equivalent volume of the culture. The operations involving the injection of fresh medium, cleaning of the cultivation chambers, and the transfer of cultures between chambers are controlled by an automate which operates electro valves, according to a programmed culture regime. For this study, turbidostat or medium swap regimes were conducted, the first employing one, the second two culture media.

Continuous culture of cell suspensions often faces challenges due to the development of biofilms on the inner surfaces of culture vessels. Under dilution-based growth conditions, mutants with an affinity for adhering to vessel surfaces emerge and become selected. These variants create biofilms, allowing them to escape the selective pressure exerted by dilution cycles, thus preventing the acquisition of the desired improved phenotype.

To overcome this difficulty, the GM3 self-cleaning device consists of two twin growth vessels whish are connected by silicon tubings. They are alternatively utilized for the cultivation of the evolution population, allowing the second vessel to be cleaned by rinsing with a cleaning solution (Figure 15). Figure 15 provides a simplified representation of the fluidic movements within the GM3 device, moving in a clockwise direction starting from the top left image. The vessel located on the left side contains the actively growing culture. Following a predefined time interval, the entire culture is transferred into the second vessel. Meanwhile, the first vessel undergoes cleaning through a 5 M NaOH treatment, effectively eliminating any biofilm formation. As the culture thrives within the second vessel, the cleaning agent from the first vessel is purged and the vessel is subjected to thorough rinsing. Subsequently, the culture is reintroduced into the first vessel. The second vessel then goes through a sterilization process involving 5 M NaOH treatment and rinsing, after which the entire cycle can restart. Usually, within a 24-hour timeframe, two to four cleaning cycles are programmed.

The cell suspension (approximately 16 ml) cultivated in the vessel is consistently aerated using compressed air or other gas mixtures, depending on the culture needs. This continuous aeration prevents cell sedimentation. The real-time acquisition of the culture's turbidity allows for the establishment of thresholds necessary to implement various continuous culture strategies. Indeed, GM3 was used under two types of culture regimes: turbidostat and medium swap.

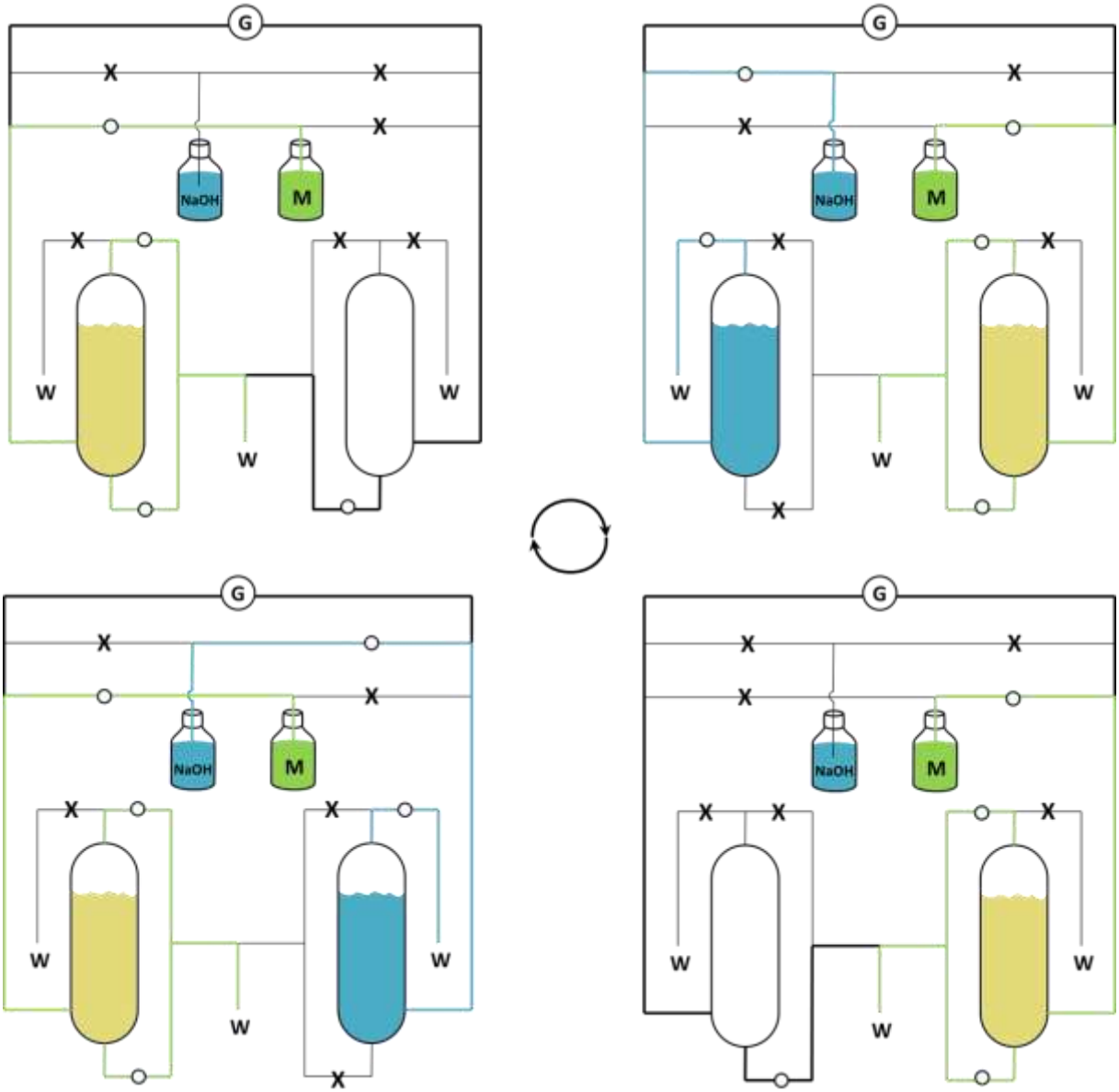


Figure 15. GM3 automated device with twin chambers. Pressurized gas supply (G); medium source (M in green); sterilizing agent (NaOH 5 M in blue); waste (W); Bacterial cultures is shown in yellow. Valves are indicated by X and O for the closed and open states, respectively. Flows in the tubing are indicated by bold lines. From Souterre (2017).

1. Turbidostat

The turbidostat regime selects for cells with an improved generation time by diluting the cells in a permissive medium (Figure 16). A pre-culture is prepared in the same medium to inoculate the device. Turbidity is measured automatically every 10 minutes and cells are diluted if the OD_{600} value is above a set threshold (approx. equivalent to an OD_{600} of 0.4). These dilutions ensure that the cells are always growing at their maximum growth rate.

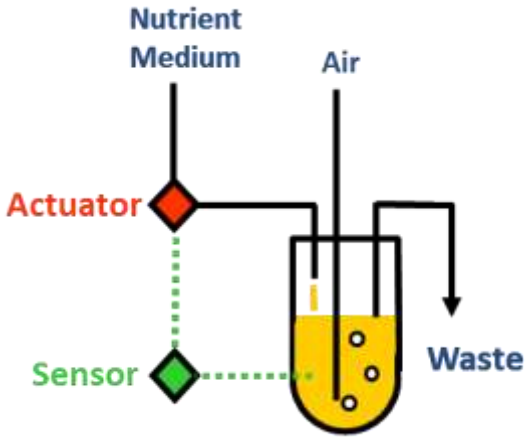


Figure 16. Turbidostat regime.

2. Medium swap

The medium swap regime was used to gradually adapt cells to a stressing medium in which the cells are not initially growing (Figure 17). Cells are diluted with fresh medium under regular intervals either with permissive or with stressing medium. The GM3 automaton is automatically making the choice of the dilution medium depending on the measured turbidity. If the measure is above the set threshold (approx. equivalent to an OD_{600} of 0.4), cells are diluted with the stressing medium, otherwise with the permissive medium.

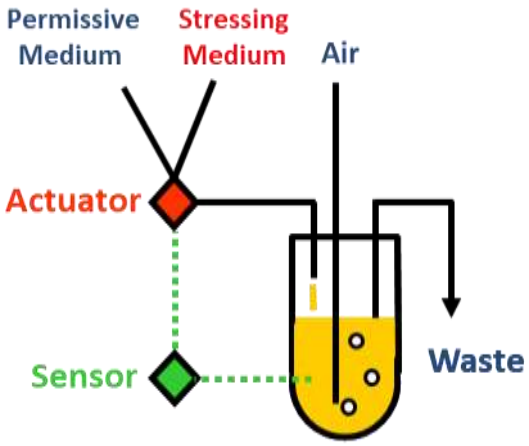


Figure 17. Medium SWAP regime.

4. Microbiologic protocols

1. Transformation by electroporation

1. Electrocompetent cells preparation

To insert plasmid or linear DNA fragments into *E. coli*, electroporation was mainly employed. A 2 mL overnight pre-culture was used to inoculate 10 mL of LB medium with supplements if necessary at an initial optical density (OD_{600}) of 0.05. The culture was incubated until an approx. OD_{600} of 0.6. The cells were then centrifuged 3 times in a 4 °C refrigerated centrifuge for 10 minutes at 4000 rpm. The volume of ice-cold water to resuspend the pellets obtained after each cycle was decreasing: after the first cycle, the pellet was resuspended in the same volume, *i.e.* 10 mL; for the second wash, the initial volume was divided by 2, *i.e.* 5 mL of ice-cold water; the third wash was made with 1 mL of ice cold water. Finally, the cells were resuspended in 100 µL of a sterile ice-cold 10 % glycerol solution and aliquots of 50 µL were stored at – 80 °C for further usage.

2. Transformation

To transform a plasmid, 50 µL electrocompetent cells were placed in a 1 mm wide electroporation cuvette along with 1 µL of a miniprep. The mix was pulsed at a voltage of 1.8 kV using an electroporator device. The cells were immediately resuspended in 1 mL of SOC medium, with supplements if necessary, for 20 minutes if the plasmid is exposing a carbenicillin resistance or 45 minutes for any other antibiotic resistance. The regeneration times were adjusted if the cells to be transformed were slow-growth strains. Finally, transformants were selected on LB plates supplemented with the appropriate antibiotics and complements at 37 °C or 30 °C depending on the strain and the plasmid.

To transform a ligation or PCR product, the amount used was usually 3 µL. PCR product was purified prior to transformation and the ligation reaction dialyzed on a 0.22 µm membrane. The rest of the protocol remained unchanged.

2. Transduction

Transduction was used to move DNA fragments from one *E. coli* strain to another was conducted following P1 phage based protocol using a various antibiotic resistance cassettes.

For phage preparation, an overnight 3 mL culture of the donor strain was made in LB medium supplemented with antibiotic and supplements if necessary. This culture was supplemented with 2.5 mM of $CaCl_2$ and incubated again at 37 °C for 30 minutes. A P1 *E. coli* MG1655 phage stock was diluted,

typically 10^{-4} and 10^{-5} in LB medium. After 30 minutes of incubation, 300 µL of P1 dilution was mixed with 300 µL of the donor strain culture incubated with $CaCl_2$. This mix was incubated at 37 °C without shaking for 20 minutes. The 600 µL mix was then transferred in 12 mL of soft agar maintained at 55 °C and immediately distributed on 3 LB agar plates. The plates were then incubated overnight at 37 °C.

The next day, phages were harvested on plates with confluent lysis plaques and transferred in a 50 mL centrifuge tube with 800 µL of chloroform ($CHCl_3$) for each plate. The tube was vortexed for 15 seconds to obtain a homogenous mix and centrifuged for 10 mL at 10,000 rpm. The supernatant containing the phages was collected in a glass tube, 100 µL of $CHCl_3$ was added for a better conservation and stored at 4 °C for further use.

To perform P1 transduction, an overnight 5 mL culture of the recipient strain was conducted in LB medium. This culture was supplemented with 2.5 mM of $CaCl_2$ and incubated again at 37 °C for 30 minutes. 1 mL of recipient strain culture was then distributed in 3 Eppendorf tubes: one with no phage, one with 5 µL of phages and one with 50 µL of phages. To allow the phages to adsorb to the cells, the 3 tubes were incubated for 15 min at 37 °C without shaking. The tubes were then vortexed and centrifuged at 8,000 rpm for 1 minute. The supernatant was discarded and the cell pellet was resuspended in LB supplemented with 0.1 $mol.L^{-1}$ of citrate and other complements if necessary. The resuspended cells were then incubated for 45 minutes at 37 °C with agitation. Finally, the cells were centrifuged again and the pellets, resuspended in the remaining medium, were plated on selective LB Petri dishes overnight at 37 °C. The following day, clones are re-streak on fresh selective plates and controlled by colony PCR.

3. Recombination using pKD46 plasmid

Recombination using λ-red recombinase from the λ red bacteriophage was used mainly to replace a gene by an antibiotic resistance cassette using a method developed by Datsenko and Wanner (2000) (Figure 18).

This λ-red recombinase is encoded by three genes, exo, beta and gam. Gene *exo* which codes for an exonuclease that will digest the linear DNA fragment from the 5′ extremity, creating a single strand DNA (ssDNA) which will recombine to the targeted region with more efficiency. Beta protects the ssDNA and promotes its recombination with a complementary strand. Gam inhibits RecBCD, the endogenous nuclease from digesting the linear DNA fragment introduced (Poteete 2001).

An antibiotic resistance gene was amplified by PCR containing extremities that had 50 bp homology to the region or gene to be replaced, the correct PCR amplification was controlled by gel analysis and the fragment purified as described in the PCR amplification and purification paragraph (II. 5. 1).

A 10 mL overnight culture of the *E. coli* MG1655 strain +2440 (Table 4) containing the plasmid pKD46 expression the λ-red recombinase was left on the bench without agitation. The next day, when the OD_{600} reached 0.3 – 0.4, the λ-red recombinase was induced using 10 mM of L-arabinose for 45 minutes at 30 °C. Electrocompetent cells were prepared with this induced culture (refer to Electrocompetent cells preparation section, II. 4. 1. 1).

Subsequently, 3 µL of the purified PCR product was transformed in 50 µL of electrocompetent cells, regenerated for 1 h at 37 °C in SOC medium and recombinants were selected on LB supplemented with the appropriate antibiotic.

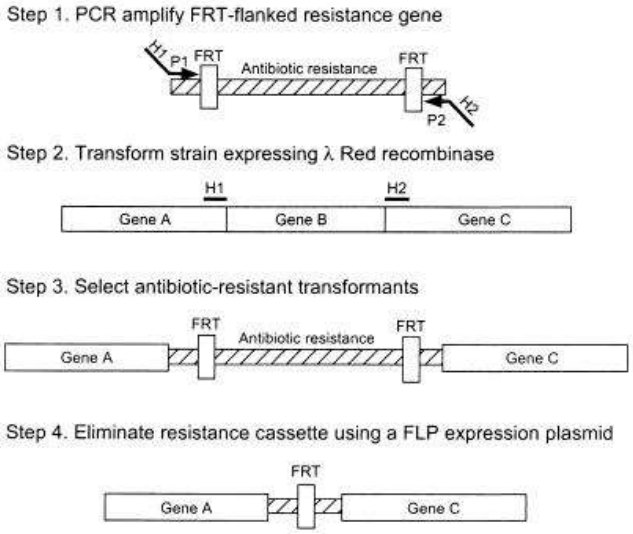


Figure 18. Gene disruption strategy by homologous recombination. H1 and H2 refer to the homology extensions or regions. P1 and P2 refer to priming sites. From Datsenko and Wanner (2000).

4. Resistance cassette excision using pCP20 plasmid

To modify the genome of *E. coli* strains used in this project, antibiotic resistance genes were inserted to select for these modifications. To further engineer or transform these modified strains with expression plasmids, the resistance genes introduced were eliminated by the Flp recombinase which recognizes the FRT regions present on either side of the cassette.

The Flp recombinase is expressed from plasmid pCP20 under a constitutive promoter and has a temperature-sensitive replication origin (Cherepanov and Wackernagel 1995). Electrocompetent cells derived from the target strain were prepared and transformed with the pCP20 plasmid, as described in the preceding protocol. Transformed colonies were selected on LB plates supplemented with carbenicillin at 30 °C, overnight. Four isolated colonies were chosen for re-streaking under similar conditions at 30 °C. Then, these four clones were streaked again on antibiotic-free agar plates at 42 °C, which facilitated the plasmid loss, as the replication efficiency of pCP20 is compromised at this

temperature. Subsequently, these four clones were successively streaked on LB Petri dishes, containing the relevant antibiotic corresponding to the targeted resistance cassette removal, LB plates supplemented with carbenicillin, which is the pCP20 plasmid resistance, and LB plates devoid of any antibiotic, at 37 °C. The selection of clones that displayed the anticipated growth phenotype was further confirmed through PCR analysis.

It is important to note that this method is not scarless, a FRT site (34 bp) is left in the genome (as illustrated by Figure 18).

5. Bioscreen

For growth experiments, a Microbiology Reader Bioscreen C apparatus (Thermo Fisher Scientific) was used. It consists of a thermostatic incubator and a culture growth monitoring device (OD reader). Log-phase pre-cultures in minimal medium with appropriate supplements were washed once using minimal medium devoid of any additional compound. Then, cells were diluted in growth medium to an initial OD_{600} of 0.05. 200 µL aliquot of each condition were distributed into honeycomb 100-well plates. Each condition was performed in triplicate. The plates were incubated in the device at 30 or 37 °C under continuous agitation. Bacterial growth was followed by optical density measurements at 600 nm every 15 minutes for variable periods depending on the growth pace of the tested strains.

6. Chromosomal insertion methods

1. pDM4

This method is adapted from Wenk et al. (2018) who kindly provided the plasmid for chromosomic insertion.

CnFDH was cloned into a plasmid, pDM4_IS10 harboring homologous regions of 600 bp upstream and downstream the targeted region IS10 (*ula* operon in *E. coli*), *sacB* gene as well as a chloramphenicol resistance gene, a kanamycin resistance, a R6K replication origin that can only be replicated by Pir strains. CnFDH operon was cloned between the two homologous regions using the

Gibson cloning technique described in II. 5. 6 section. The plasmid obtained is represented in Figure 19.

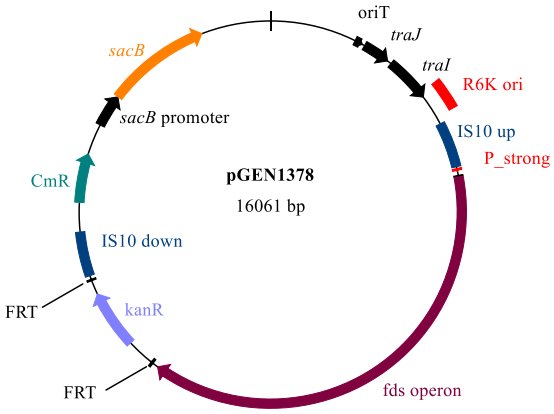


Figure 19. Map of pDM4_IS10-CnFDH (pGEN1378).

The plasmid was integrated in the chromosome of an *E. coli* MG1655 by conjugation following the protocol described in Appendix 4: Conjugation protocol adapted from Wenk et al. (2018). The integration was confirmed by colony PCR (II. 5. 1).

The selection of the second event of homologous recombination to get rid of the plasmid backbone was made using a selection on kanamycin and sucrose (Wenk et al. 2018). The positive transconjugants are grown in 5 mL of LB medium for 5 hours in a 37 °C shaking incubator. Then, serial dilutions from 10^0 to 10^{-4} are plated on LB Petri dishes supplemented with 10 % sucrose and kanamycin and grown for 24 hours at 37 °C. *sacB* gene encodes for Levansucrase that will convert sucrose into toxic levan counter-selecting the cells that still contain the plasmid backbone harboring *sacB*.

The colonies obtained are then streaked on LB plates with kanamycin and LB plates with kanamycin and chloramphenicol overnight at 37 °C. Colony PCR is performed on colonies that only grew on LB kanamycin plates.

Finally, the kanamycin resistance cassette is removed using the FLP recombinase as described in II. 4. 4.

2. pEVL410

The gene *ptxD* was integrated in the chromosome of an *E. coli* strain using a method based on homologous recombination assisted by λ-red recombinase. For this purpose, *ptxD* was cloned into the pEVL410 vector which is a pSW23 backbone containing 500 bp homologous regions (upstream and downstream) specific to the *kdgK* locus in the *E. coli* chromosome, *kdgK* being a non-essential gene. The cloning was performed using restriction-ligation technique in a BW19610 strain, which is capable of replicate the R6K origin of pSW23 plasmid. The integrity of the cloning was verified through sequencing, and the resulting plasmid was subjected to digestion with KpnI and SacI restriction enzymes to generate

a linear double-stranded fragment. This fragment of 2.8 kbp included the two 500 bp homologous regions, *ptxD*, and a spectinomycin resistance gene for selection (Figure 20).

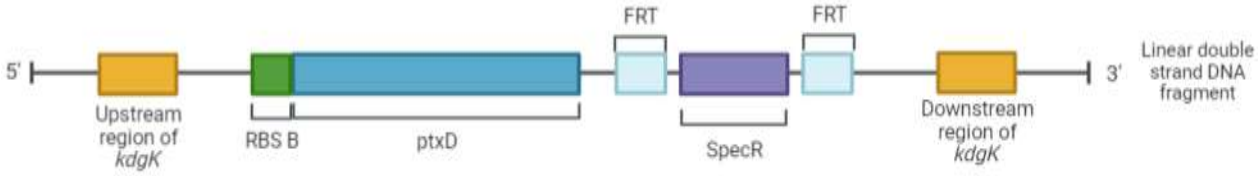


Figure 20. Linear double strand DNA fragment containing *ptxD* gene from *P. stutzeri* used for homologous recombination at *kdgK* locus in *E. coli* chromosome. Designed on Biorender.

The linear fragment was then transformed into the +2440 strain, which carries the pKD46 plasmid expressing the λ-red recombinase protein. Recombinant clones were selected on spectinomycin plates. Further validation was performed through colony PCR, and a clone with the correct PCR profile was chosen to produce P1 phages for transduction into the recipient strain. Finally, the spectinomycin resistance gene was excised using the pCP20 plasmid containing the FLP-recombinase.

3. pSW23

Another chromosomal insertion method was used for insertion at the IS10 locus using homologous recombination. The gene to insert in the chromosome was cloned in a pSW23 vector backbone (R6K replication origin) with 1000 bp upstream and downstream of the IS10 flanking this gene. This plasmid was transformed in an *E. coli* MG1655 WT strain, unable to replicate such plasmid with a R6K origin, thus enabling the selection of the clones that incorporated the plasmid in their genome using LB supplemented with chloramphenicol. The clones are further verified by PCR.

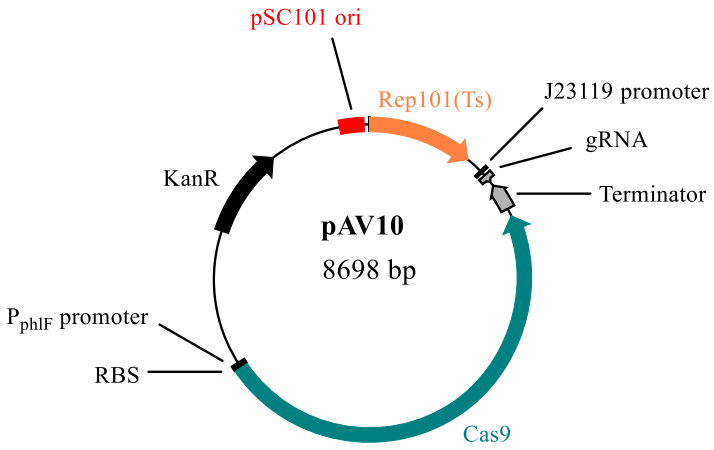


Figure 21. Map of pAV10 plasmid. P_{phlF} is a DAPG-inducible promoter for Cas9 expression. This plasmid possesses a thermo-sensitive origin of replication (Rep101Ts). The gRNA targets specifically the CmR resistance gene. Gift from David Bikard.

The selection of the second event of homologous recombination was performed using CRISPR-Cas9 assisted counter-selection. For this purpose, the plasmid pAV10 (Figure 21), a gift from David Bikard (Pasteur Institute), harboring the dCas9 protein under the control of a P_{PhlF} promoter inducible using Di-acetyl-phloroglucinol (DAPG), along with a RNA guide targeting the chloramphenicol resistance guide. The surviving clones are the ones that has lost the pSW23 backbone.

As this method was elaborated during the PhD project, it will be developed in more details in the Results and discussion chapter.

7. Gene silencing

Gene silencing was performed using a method from Depardieu and Bikard (2020). Specific gRNAs for the target sequence were designed using the tool developed by Calvo-Villamañán et al. (2020). Then, the selected gRNAs were cloned into a plasmid carrying dCas9 as well as a Golden Gate cloning platform (BsaI), represented in Figure 22. The cloning procedure in available in II. 5. 8.

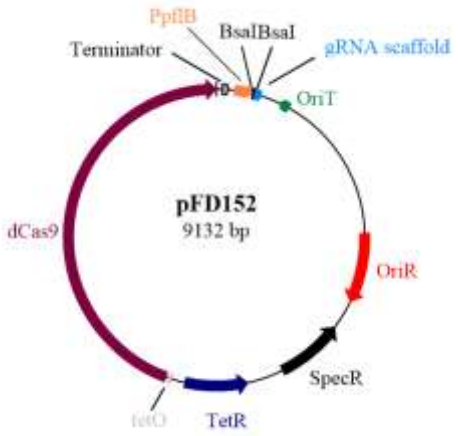


Figure 22. pFD152 plasmid map for CRISPRi from Depardieu and Bikard (2020).

The strain of interest was transformed by electroporation with the four plasmids constructed with different gRNA as well as the parent plasmid with a non-targeting gRNA (pFD152). Two clones of each transformation were picked and adapted to minimal medium corresponding to the permissive condition supplemented with spectinomycin (plasmid pFD152 resistance).

An overnight 2 mL culture was performed for each clone. The next day, dilutions in minimal medium were made from 10^{0} to 10^{-8}. 5 µL of these dilutions were then plated on four square Petri dishes with permissive and test conditions with and without anhydrotetracycline (aTc) induction (0.5 µg.ml^{-1}).

Plates were then incubated at 30 °C from several days to one week, depending on growth rate.

5. Molecular biology protocols

1. PCR amplification and purification

Polymerase Chain Reaction (PCR) primers were purchased from Sigma. Q5 High-Fidelity DNA polymerase (NEB) was used for the amplification of fragments for cloning purposes (Table 15). Two types of cycle (Table 16 and Table 17) were performed for amplification depending on the size and the GC content of the PCR product expected. Annealing temperature were calculated by NEB Tm calculator.

Table 15. PCR mix using Q5 DNA polymerase for cloning purposes.

Component	Volume per 50 µL reaction
Q5 reaction buffer (5X)	10 µL
Q5 High GC Enhancer (5X)	10 µL - optional
dNTP (2,5 mM)	4 µL (or 8 µL for PCR > 3 kbp)
DNA Template	Variable (depends on the template concentration)
5 µM forward primer	2 µL
5 µM reverse primer	2 µL
Q5 DNA polymerase	0.5 µL
ddH_20	Up to 50 µL

Table 16. Regular PCR cycle.

Step	Temperature	Time	Number of cycle
Initial denaturation	98 °C	5 min	1
Denaturation	98 °C	45 s	25 to 30
Annealing	72 to 55 °C	45 s	
Extension	72 °C	30 s per kb	
Final extension	72 °C	5 min	1

Table 17. Touchdown PCR cycle for long and high GC content fragments.

Step	Temperature	Time	Number of cycle
Initial denaturation	98 °C	5 min	1
Denaturation	98 °C	45 s	10
Annealing	72-60 °C (-0,5 °C/cycle)	45 s	
Extension	72 °C	30 s per kb	
Denaturation	98 °C	45 s	25
Annealing	Last annealing temperature reached by the previous cycle	45 s	
Extension	72 °C	30 s per kb	
Final extension	72 °C	5 min	1

All PCR fragments were purified using QIAquick PCR Purification Kit (QIAGEN) following the protocol from the supplier and DNA was eluted in DNAase-free water.

If the PCR was made using a template that had the same antibiotic resistance than the recipient vector used for the cloning, a DpnI digestion was performed. 1 µL of DpnI (NEB) was added directly to the PCR and incubated at 37 °C overnight. Then, heat-inactivation of the enzyme was made at 80 °C for 20 min and the PCR product was purified using the QIAquick PCR purification kit.

Colony PCR were performed using DreamTaq DNA polymerase (Table 18) following the cycle described in Table 19.

Table 18. PCR mix using DreamTaq DNA polymerase for colony PCR.

Component	Volume per 25 µL reaction
DreamTaq green buffer (5X)	10 µL
dNTP 2,5 mM	2 µL
DNA Template	1 colony
5 µM forward primer	2 µL
5 µM reverse primer	2 µL
DreamTaq DNA polymerase	0.5 µL
ddH_20	Up to 25 µL

Table 19. Colony PCR cycle.

Step	Temperature	Time	Number of cycle
Initial denaturation	95 °C	5 min	1
Denaturation	95 °C	30 s	25 to 30
Annealing	58 °C	30 s	
Extension	72 °C	1 min per kb	
Final extension	72 °C	5 min	1

If the PCR product from the colony PCR was to be sequenced, purification was made using a MultiScreen-PCR 96-weels filter plate (Milipore) and DNA was eluted in water.

2. Plasmid extraction

All the plasmid extractions were conducted using QIAGEN miniprep kit with a culture made in LB medium complemented with the appropriate antibiotic and supplements. For the purification of medium and high copy plasmids, the volume of culture used was 3 mL, for low copy plasmid, 10 mL culture were used. Plasmids were further eluted in EB or water depending on usage.

The quality of the plasmid extraction was controlled using NanoDrop spectrophotometer (DNA quantification section, II. 5. 4).

3. Genomic DNA extraction & sequencing

Genomic DNA was extracted using the Sigma GenElute Bacterial Genomic DNA kit with a 5 mL overnight culture in rich medium. The DNA extraction quality was verified by NanoDrop and Qubit (described in DNA quantification section, II. 5. 4). Once the DNA preparation is validated, the DNA was sequenced either by an Illumina MiSeq instrument performed at the Genoscope or by an Oxford Nanopore Technologies device performed by Plasmidsaurus.

Subsequently, reads are compared to the reference genome for *Escherichia coli* K-12 MG1655 (GenBank: U00096.3) using the Microscope Platform (Vallenet et al. 2017) to identify mutations.

4. DNA quantification

Genomic DNA, PCR products and minipreps were quantified using Qubit or NanoDrop depending on the needs.

DNA purity verification and quantification were performed using a NanoDrop 2000 spectrophotometer. The device realizes three absorbance measurements at different wavelength: 230 nm (absorbance of contaminants such as carbohydrates, phenol, EDTA), 260 nm (nucleic acid absorbance) and 280 nm (protein absorbance). The device give two ratios: $A_{260/280}$ which should be around 1.8 (indicator of the presence of proteins) and $A_{260/230}$ which should be around 2 (indicator of the presence of solvents). The concentration was also calculated by the NanoDrop.

If necessary for further usage of the DNA sample, a more precise quantification using a Qubit fluorometer (Invitrogen) was performed following the supplier's guidelines.

5. CPEC cloning

Circular Polymerase Extension Cloning (CPEC) is a cloning technique based on overlapping DNA fragments (Quan and Tian 2011).

Primers possessing homologous extensions corresponding to the vector sequences were designed to amplify the fragment(s) intended for cloning. Simultaneously, the vector was amplified utilizing primers generating linear double-stranded DNA (dsDNA) fragments. Then, the resulting products were verified by gel analysis, purified and quantificatied using the Qubit quantification technique.

Cloning reactions were made using an equimolar vector:insert ratio using 150 ng of vector (Table 20).

Table 20. CPEC cloning reaction mix.

Component	Volume per 15 µL reaction
Q5 reaction buffer (5X)	3 µL
dNTP 2,5 mM	2 µL
Vector	Variable (depends on the PCR)
Insert	Variable (depends on the PCR)
Q5 DNA polymerase	0.5 µL
ddH_20	Up to 15 µL

The mix described above was placed in a thermocycler to follow the cycle describe in Table 21. 3 µL of the cloning reaction was dialyzed on a 0.22 µm membrane and transformed in β2033 electrocompetent cells, unless stated otherwise. Colonies were selected on LB agar plates supplemented with the

appropriate antibiotic. Clones were isolated and first controlled by colony PCR. Clones with the expected PCR profile were cultured, plasmid extraction was performed and finally validated by Sanger sequencing performed at the Genoscope or by sending it to Plasmidsaurus for whole plasmid sequencing.

Table 21. CPEC cloning cycle.

Step	Temperature	Time	Number of cycle
Initial denaturation	98 °C	5 min	1
Denaturation	98 °C	30 s	25
Slow ramp annealing	70 °C to 55 °C (-0,1 °C.s^{-1})	3 min	
Annealing	55 °C	30 s	
Extension	72 °C	30 s per kb	
Final extension	72 °C	5 min	1

6. Gibson cloning

Gibson cloning was performed using NEBuilder HiFi DNA Assembly Cloning Kit (NEB). Primers were designed using SnapGene Gibson Assembly Tool with 25 bp homology between vector and insert. PCRs were performed, purified and quantified as described above (sections II. 5. 1. and II. 5. 4.). For a two and three fragments assembly, a vector:insert ratio of 1:2 was used with a total amount for all fragments of 0.2 pmols in a 20 µL reaction. For assemblies with more fragments, a vector:insert ratio of 1:1 was used with a total amount for all fragments of 0.5 pmols in a 20 µL reaction.

All Gibson cloning reactions were incubated for 60 min at 50 °C in a thermocycler. 3 µL of the cloning reaction was dialyzed on a 0.22 µm membrane and transformed in β2033 electrocompetent cells, unless stated otherwise. Colonies were selected on LB agar plates supplemented with the appropriate antibiotic. Clones were isolated and first controlled by colony PCR. The ones with the expected PCR profile were cultured, plasmid extraction was performed and finally validated by Sanger sequencing and/or sent to Plasmidsaurus for whole plasmid sequencing.

7. Golden Gate Cloning

Golden Gate cloning was performed using type IIS restriction enzymes and T4 DNA ligase (NEB). PCRs were performed, purified and quantified as described in previous sections (II. 5. 1 and 4.). Cloning reactions were made using 200 fmol of each cloning fragment (quantity was calculated with NEBioCalculator). Cloning reaction mixes were made as stated in Table 22, then placed in a thermocycler following the cycle described in Table 23.

Table 22. Golden Gate cloning mix.

Component	Volume per 20 µL reaction
T4 DNA ligase buffer (10X)	3 µL
Vector	Variable (depends on the PCR)
Insert(s)	Variable (depends on the PCR)
T4 DNA ligase	1 µL
BsaI or Esp3I	1 µL
ddH_20	Up to 20 µL

Table 23. Golden Gate reaction cycle.

Step	Temperature	Time	Number of cycle
Cutting	37 °C	5 min	50
Annealing	16 °C	5min	
Enzyme denaturation	80 °C	5 min	1

Subsequently, 3 µL of the cloning reaction was dialyzed on a 0.22 µm membrane and transformed in β2033 electrocompetent cells, unless stated otherwise. Colonies were selected on LB agar plates supplemented with the appropriate antibiotic. Clones were isolated and initially controlled by colony PCR. The ones with the expected PCR profile were cultured, plasmid extraction was performed and finally validated by Sanger sequencing and/or sent to Plasmidsaurus for whole plasmid sequencing.

8. CRISPRi

CRISPRi method was used to study the impact of CnFDH operon silencing on the phenotype of a test strain. This method is based on the use of a dCas9 guided with a gRNA specific to a target region that will prevent the transcription initiation or elongation (Depardieu and Bikard 2020). Specific gRNA were designed using the CRISPR Pasteur guide RNA design tool (CRISPR@Pasteur). The four gRNA with the highest scores were cloned in pFD152 (Table 5 and Figure 22) harboring dCas9 and a cloning region for gRNAs. The cloning procedure was made using the protocol described in the reference article (Depardieu and Bikard 2020) and detailed below.

Oligonucleotides corresponding to gRNA were first phosphorylated by incubating the mix described in Table 24 for 30 min at 37 °C.

Table 24. Phosphorylation reaction mix

Component	Volume per 50 µL reaction
Forward 10 µM primer	15 µL
Reverse 10 µM primer	15 µL
T4 ligase buffer 10 X	5 µL
T4 PNK	1 µL
ddH_20	Up to 50 µL

Oligonucleotides were then denatured by adding 2 µL of NaCl 1 mM to the phosphorylated primer mix and incubated for 5 minutes at 95 °C in a dry heating block. Primers are then annealed by removing the block from the heater and leaving in on the bench until the block achieve ambient temperature for a slow decrease of the temperature that will promote proper annealing.

Annealed oligonucleotides are then cloned using Golden Gate as described in II. 5. 7. with BsaI enzyme (Table 25).

Table 25. Golden Gate reaction mix for gRNA cloning.

Component	Volume per 10 µL
Plasmid DNA (pFD152)	2 µL
Annealed oligonucleotides diluted 1 : 10 in water	2 µL
T4 DNA ligase buffer 10X	1 µL
BsaI-HF	1 µL
T4 DNA ligase	1 µL
ddH_2O	Up to 10 µL

Subsequently, this mix is incubated in a thermocycler with the conditions described in Table 26.

Table 26. Cycle for Golden Gate cloning of gRNA in pFD152.

Step	Temperature	Time	Number of cycle
Cutting	37 °C	3 min	25
Annealing	16 °C	4min	
Enzyme denaturation	80 °C	5 min	1

The cloning reactions are dialyzed on a 0.22 µm filter membrane for 20 min at room temperature and 2 µL is transformed in β2033 electrocompetent cells using the protocol described above (II. 4. 1. 2).

Four clones obtained after transformation are re-streaked for each gRNA, plasmids are extracted as described above (II. 5. 2) and validated by Sanger sequencing at the gRNA cloning locus. Finally, the resulting plasmids are extracted and transformed in the targeted strain.

9. RT-qPCR

Quantitative reverse transcription polymerase chain reaction (RT-qPCR) was performed to determine the level of expression for genes of interest in specific conditions. Pre-cultures were prepared to inoculate test media at an OD_{600} around 0.2-0.3 in 5 mL of the test medium. The cultures were incubated at 30 or 37 °C with continuous shaking until the final OD_{600} reached 0.5-0.7.

Two volumes of RNA protect Bacteria Reagent was added to one volume of bacterial culture. The mix is vortexed, incubated for 5 minutes at room temperature and centrifuged for 10 minutes at 7000 rpm and 4 °C. The supernatants are discarded.

RNA extraction was performed using RNAeasy kit (QIAGEN) following the supplier protocol guidelines and eluted in 50 µL of RNase free water. Extracted RNA are then treated with DNaseI (NEB) and the absence of DNA was verified by PCR. RNA extractions are stored at -20 °C for further use.

Reverse transcription is performed using High capacity cDNA Reverse transcription kit (Applied Biosystem). cDNA obtained are then quantified using Qubit single strand DNA quantification kit (ThermoFisher). The KAPA SYBR FAST kit (Roche) and Mx3005 P real-time PCR system (Agilent Technologies) were used for qPCR.

The primer pairs utilized for qPCR were designed to amplify a region spanning approximately 200 bp. The chosen housekeeping gene is *panB* and was amplified using primers 4463 and 4464 (sequences available in Appendix 1: Genetic sequences of diverse building blocks). Three pairs of primers on the

fdsGBACD operon were designed, 6419 and 6420 for *fdsG*, 6422 and 6423 for *fdsA*, 6427 and 6428 for *fdsD* (sequences available in Appendix 1: Genetic sequences of diverse building blocks). Amplifications were conducted in duplicate using cDNA corresponding to 20 ng of total RNA, with the cDNA being subjected to four successive three-fold dilutions. The amplification cycling procedure consisted of an initial denaturation at 95 °C for 3 minutes, followed by 40 cycles of denaturation at 95 °C for 30 seconds, annealing at 60 °C for 30 seconds, and extension at 72 °C for 30 seconds. The acquired data were analyzed with the MxPro QPCR software from Agilent. The abundance of mRNAs transcripts were compared between the cycle threshold of the internal control (*panB*) and the one from *fdsG*, *fdsA* and *fdsD* genes.

6. Biochemistry

1. Production and purification of proteins

E. coli DH5α cells were transformed with expression plasmids of CnFDH and RcFDH (pGEN1353 and pTHfds05, Table 5). One colony was picked to start a pre-culture in LB medium supplemented with carbenicillin and sodium molybdate at 1 mM. The next day, the overnight pre-culture was used to inoculated 500 mL culture in Terrific Broth culture supplemented with carbenicillin, sodium molybdate (1 mM) and riboflavin (0.2 g/L), at an initial OD600 of 0.1. Riboflavin was added to the culture as it is FMN precursor which is a cofactor essential for FDH functionality. Gene expression was induced with 50 µM isopropyl β-D-1-thiogalactopyranoside (IPTG) and the culture was incubated for 24 h at 28 °C under agitation as described by Hartmann and Leimkühler (2013). The culture is then centrifuged at 4 °C, 12000 rpm for 30 min, supernatant is discarded and the pellet is frozen at – 80 °C.

The cell pellet is resuspended in 32 mL lysis buffer (50 mM phosphate, 300 mM NaCl, 30 mM imidazole, 10 mM KNO_3, pH is adjusted to 7.5) supplemented with 1 mM Dithiothreitol (DTT), 1 mM Pefabloc SC (Roche Applied Sciences), 26 µL of LysonaseTM bioprocessing reagent and 3,6 mL of BugBuster Protein Extraction Reagent 10X (Novagen). Lysis mix is incubated for 30 min under agitation at ambient temperature then centrifuged at 4 °C, 12000 rpm for 30 min. Total protein content of the lysate is determined using the Bradford protein assay.

The lysate is then filtered, diluted to an approximate concentration of 3 mg/ml before being charged on to Nickel affinity column. This column is coupled to a gel permeation chromatography to obtain fractions containing the proteins retained by the affinity column (HiLoad 16/600 Superdex 200-pg) by size. Proteins were eluted in a buffer containing 50 mM Tris pH 7.0, 10 mM KNO_3 and 1 mM DTT. Fractions containing purified proteins are analyzed using SDS-PAGE gel.

Protein content of each fraction obtained is then quantified using the Bradford protein assay (Bradford 1976). Samples were diluted if necessary to enter in the standard curve that was established with a stock solution of BSA at 1 mg/ml diluted in a range from 5 to 20 µg/ml. Absorbance is measured using SAFAS UV mc2 spectrophotometer at 595 nm in 1 cm optical path cuvettes.

2. In vitro activity test

FDH activity was assessed using a SAFAS UV mc2 double beam spectrophotometer at a temperature of 25 °C. The enzymatic reactions were monitored by measuring the change in NADH absorption at 340 nm. Typical activity assays were conducted in Tris buffer pH 9 with 1 mM NAD^+ and 3 mM formate as substrates. The molar extinction coefficient of NADH, which is used to calculate kinetic parameters is 6220 $M^{-1} \cdot cm^{-1}$. All activities were calculated were calculated with respect to one trimer (*fdsA, fdsB* and *fdsG*) with a molecular mass of 172 kDa for RcFDH and 178 kDa for CnFDH.

The activity assays were initiated by the addition of formate to the reaction mix and were monitored for 30 seconds. Kinetic constants were determined by nonlinear analysis of initial rates from duplicate experiments using SigmaPlot 9.0 (Systat Software, Inc.).

3. Measure of H_2 in Hungate tubes using gas chromatography

The gaseous composition of the culture atmosphere was determined using gas chromatography (GC) analysis with direct injection of 50 µL of the gaseous atmosphere from the culture. The detailed analytic method used for GC is available in Appendix 10: GC method for H_2 detection. The measures are performed using the Clarus 580 GC model from Perkin Elmer with a GC column Porapak Q, 80/100, SS, 6' x 1/8'' x 6''. The carrier gas is N_2.

III.Results and discussion

1. Expression of a functional complex formate dehydrogenase in *E. coli*

In this chapter, the experiments described were conducted to demonstrate the *in vivo* functionality of the O_2-tolerant NAD^+-dependent formate dehydrogenase from *Cupriavidus necator* (CnFDH) in *E. coli.* For this, the enzyme was cloned and expressed from different formats in an *E. coli* selection strain depending on FDH-catalyzed formate oxidation to CO_2 to provide NADH for cell growth, the oxidizing reaction being favored over the reaction in the reductive sense.

1. Expression of the CnFDH using plasmids

1. Plasmids & strains constructions

1. Plasmids

The construction of the main plasmid of this study harboring the operon encoding the formate dehydrogenase from *Cupriavidus necator* (CnFDH) posed significant challenges due to the substantial size of this operon. As mentioned in the Introduction (*Formate dehydrogenases* section, I. 3), the operon is composed of five genes, has a size of approximately 6100 base pairs (bp) and a high GC content of 67 % (see the complete sequence in Appendix 2: Formate dehydrogenase from *C. necator* sequence). Numerous attempts were undertaken to amplify the operon by PCR from the genomic DNA of DSM13513, the best results were obtained using a touchdown PCR protocol (*PCR amplification and purification* section, II. 5. 1). Regarding the vector backbone, pTRC99a was finally chosen for the inducibility of expression after a number of unsuccessful trials involving plasmids with constitutive expression, indicating toxicity. Multiple experiments were conducted using restriction-ligation cloning, yet, no successful results were achieved. Finally, the successful cloning of the operon was accomplished through Gibson cloning and was validated using Sanger sequencing, resulting in the plasmid named pGEN1340 represented in Figure 23.

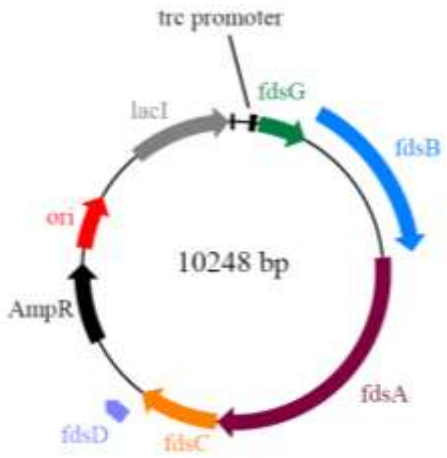


Figure 23. pTRC-CnFDH plasmid (pGEN1340).

To assess the functionality of the CnFDH operon, additional constructions were performed using the same native genes, but with ribosome binding sites (RBS) incorporated at the 5'-extremity of each gene. The three genes constituting the active enzyme, namely *fdsA*, *fdsB*, and *fdsG*, were cloned into a high-copy pUC plasmid via Gibson cloning methodology (Figure 24, A), resulting in the plasmid named pGEN1338. Following the same design strategy, the chaperone proteins encoded by the *fdsC* and *fdsD* genes were cloned into a pSU plasmid (Figure 24, B) using CPEC protocol, resulting in the plasmid named pGEN1348.

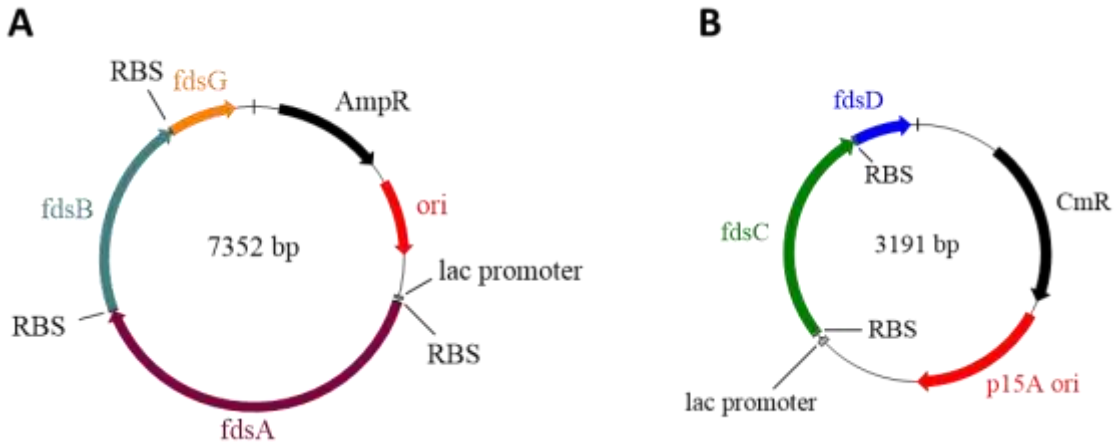


Figure 24. Formate dehydrogenase operon from *Cupriavidus necator* cloned into two plasmids. (A) pUC-*fdsABG* plasmid (pGEN1338). (B) pSU-*fdsCD* plasmid (pGEN1348).

CnFDH expression using synthetic genes was also tested, the five genes of the operon were individually codon optimized using the *E. coli* codon usage by Twist Biosciences. Subsequently, these genes were cloned in the same way as were pGEN1338 and pGEN1348, using CPEC cloning resulting in the plasmids

pGEN1323 (pUC-*fdsABG*) and pGEN1320 (pSU-*fdsCD*) which was constructed by restriction-ligation cloning.

Additional constructions in a constitutive plasmid backbone, pZE21, were also used for the cloning of CnFDH and the FDH from *Thiobacillus* sp. (TsFDH). CnFDH was cloned in this vector using Gibson method, resulting in the plasmid pGEN1393. TsFDH was cloned by CPEC, resulting in the plasmid pGEN1395. Both constructions are represented in Figure 25 below.

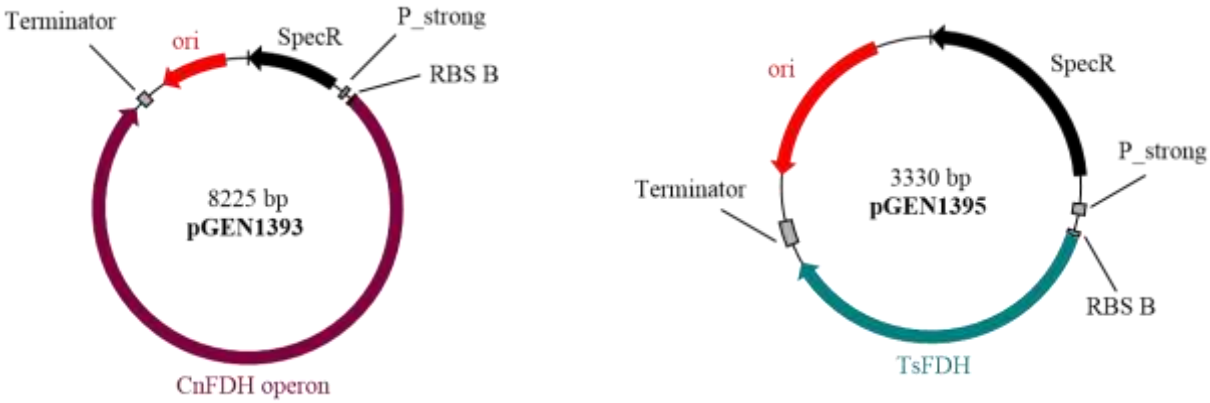


Figure 25. Plasmids maps of pGEN1393 (pZE21-CnFDH) and pGEN1395 (pZE21-TsFDH).

2. *Strains*

An "energy-auxotroph" strain was described by Wenk et al. (2020), which can be used as an *in vivo* platform for testing NADH-generating enzymes. To do so, the authors deleted the *lpd* gene encoding the lipoamide dehydrogenase. The lipoamide dehydrogenase is part of three enzyme complexes in *E. coli*: pyruvate dehydrogenase, 2-oxoglutarate dehydrogenase complex, and the glycine cleavage system (Guest and Creaghan (1972); Pettit and Reed (1967); Steiert, Stauffer, and Stauffer (1990)). By deleting this gene, the strain is auxotrophic for reducing power requiring an energy source and acetate as fatty acid precursor. Acetate can be activated to acetyl-CoA and further assimilated *via* the glyoxylate shunt, thus serving as a carbon source. By contrast, no energy is produced in the process in a Δ*lpd* context. When carbon sources entering the upper metabolism (such as glucose or glycerol) are provided to the cells, they cannot synthesize acetyl-CoA due to the pyruvate dehydrogenase disruption (Figure 26). As demonstrated in this reference paper, the resulting strain can be used to test for energy generation enzymes such as FDHs, hydrogenases or phosphite dehydrogenases, which are tested in this research project. The NADH that is generated by the heterologous system can subsequently be used to regenerate ATP through oxidative phosphorylation.

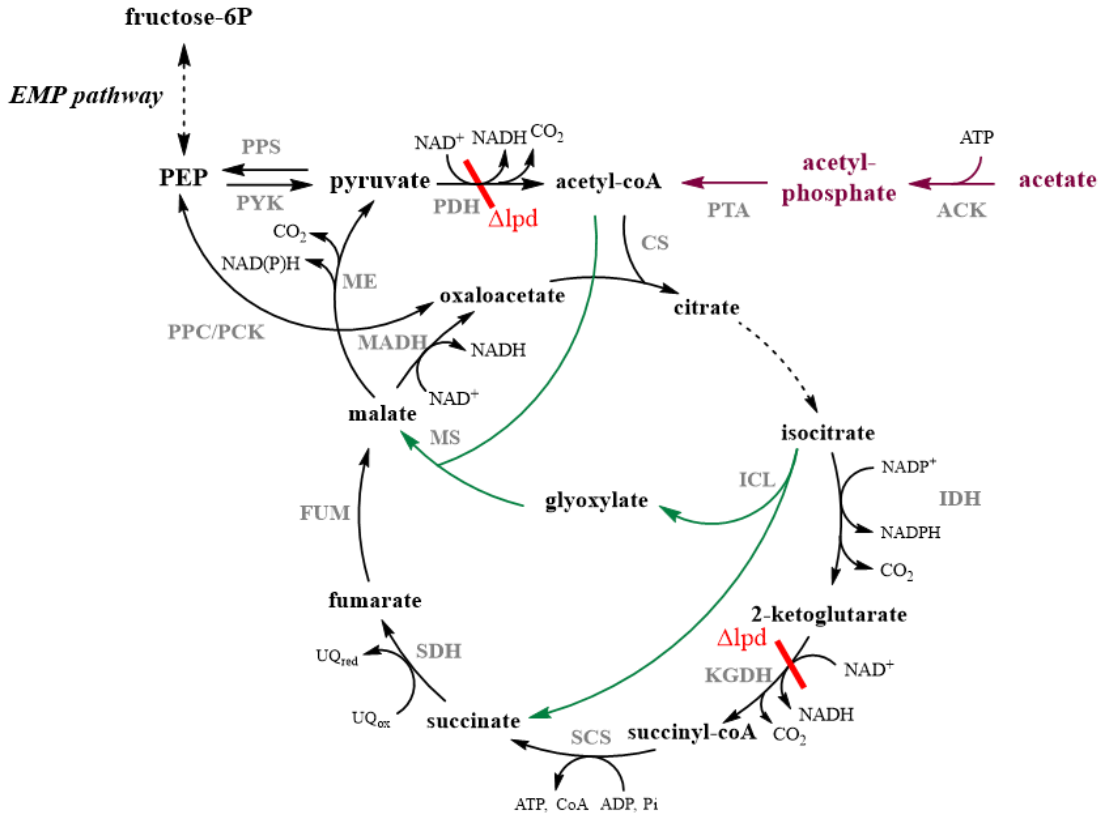


Figure 26. *E. coli* Δ*lpd* strain metabolism. The red bars represent the *lpd* gene disruption. ACK, acetate kinase; CS, citrate synthase; FUM, fumarase; IDH, isocitrate dehydrogenase; KGDH, 2-ketoglutarate dehydrogenase complex; MADH, malate dehydrogenase; ME, malic enzymes; MS, malate synthase; PDH, pyruvate dehydrogenase; PCK, PEP carboxykinase; PPC, PEP carboxylase; PPS, PEP synthetase; PTA, phosphate acetyltransferase; PYK, pyruvate kinase; SCS, succinyl-CoA synthetase; SDH, succinate dehydrogenase. Redesigned from Wenk et al. (2020).

To generate this strain, the *lpd* gene was deleted through P1 transduction. Our laboratory possesses the Keio collection (Baba et al. 2006), which consists of *E. coli* K-12 strains where each non-essential gene under rich medium condition has been replaced with a kanamycin resistance cassette. P1 phages of the JW0112 Δ*lpd*::*kan* strain from the collection were prepared and used to transduce the deleted *lpd* allele into *E. coli* MG1655 WT. Clones were selected on LB plates supplemented with kanamycin and their genotype was confirmed through colony PCR, the kanamycin resistance cassette was excised using the FLP-recombinase protocol, resulting in the strain G5416.

Subsequently, the plasmid pGEN1340 (pTRC-CnFDH) was transformed in this metabolic screening strain, resulting in strain G5663 Δ*lpd* pTRC-CnFDH. In addition, the plasmids pGEN1338 and pGEN1348 with the native genes and the plasmids pGEN1320 and pGEN1323 with codon-optimized genes were also inserted in strain G5416 to obtain strains G5730 and G5726, respectively.

2. FDH dependent growth on formate

1. *Native operon expression*

The *in vivo* activity of plasmid-borne CnFDH was tested in the Δ*lpd* context with formate as the sole energy source. Various growth media compositions were tested, the best growth was observed in MS medium supplemented with acetate (20 mM), pyruvate (20 mM), formate (60 mM), and IPTG (100 µM). A long lag phase of 3 to 4 days was observed when the medium was changed from a permissive minimal medium containing glucose to a test medium lacking glucose. To facilitate the growth of the G5663 strain and to obtain more representative growth curves, the G5663 strain was pre-cultured in the test medium before inoculating Bioscreen plates for acquiring growth curves (Figure 27).

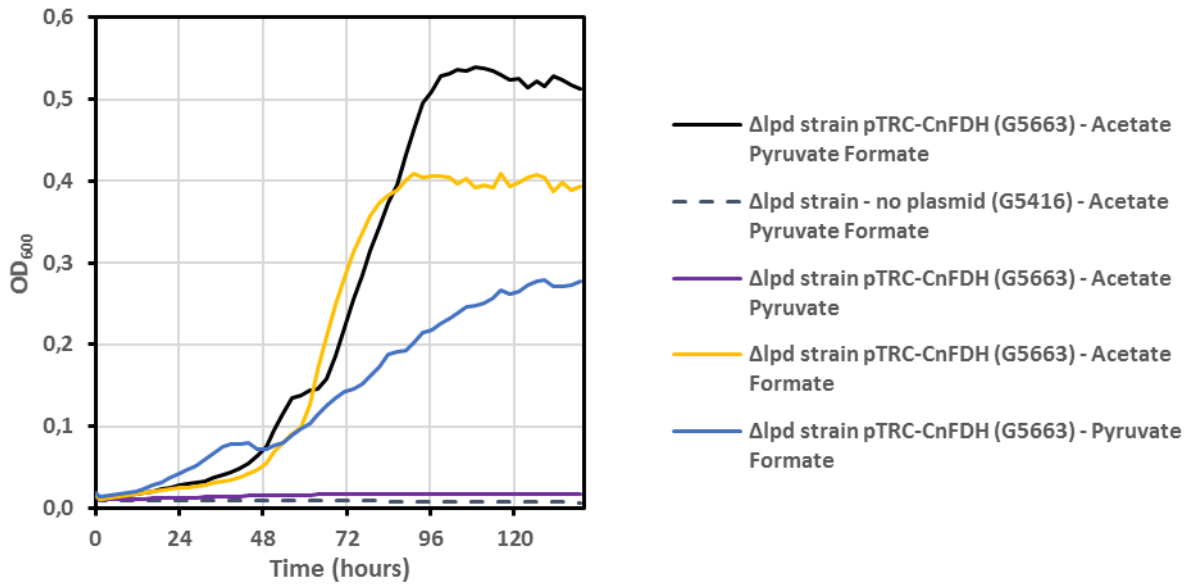


Figure 27. Growth of *E. coli* Δ*lpd* strains G5663 (Δ*lpd* pTRC-CnFDH) and G5416 (no plasmid). Cells were grown in MS mineral medium supplemented with the indicated compounds at 60 mM formate, 20 mM pyruvate and 20 mM acetate. IPTG was added to all cultures at 100 µM. Growth temperature was adjusted to 30°C.

In the presence of acetate (20 mM) and formate (60 mM) solely (yellow curve in Figure 27), the maximum growth yield (OD_{600} of 0.4) is consistent with what was described in the reference publication (Wenk et al. 2020). It was noted that increasing the acetate supplementation in the medium did not improve growth suggesting that it is not the limiting factor for growth in this context.

To optimize growth from the G5663 strain, pyruvate was provided as an additional carbon source but without impeding the metabolic test since pyruvate was demonstrated to not be sufficient to support

growth by itself in the Δ*lpd* genetic context (Wenk et al. (2020); this study). To better understand these results, the contribution of pyruvate to growth was further investigated.

1) *poxB* deletion

The hypothesis was formulated that *poxB* gene may be involved in the observed phenotype, referring to the study of Abdel-Hamid, Attwood, and Guest (2001). The *poxB* gene encodes a pyruvate oxidase, which catalyzes the oxidative decarboxylation of pyruvate to form acetate and CO_2. The authors constructed an *E. coli* strain lacking the pyruvate dehydrogenase complex (PDHC) by deleting *aceE* gene. They demonstrated that in this strain, growth could be restored either by acetate supplementation or by achieving a high expression rate of *poxB*, suggesting that elevated expression of this enzyme could compensate for the inactive PDHC by providing sufficient acetate. In the case of our G5416 Δ*lpd* strain, the PDHC is inactive due to the deletion of the *lpd* gene. Therefore, the hypothesis was formulated that the observed growth in our strain was supported by PoxB, which increased the availability of acetate within the cells.

To construct the strain Δ*lpd* Δ*poxB*, P1 phages were prepared from strain JW0855 (Δ*poxB*::*kan*) of the Keio collection and used for transduction of strain G5416 resulting in strain Δ*lpd* Δ*poxB::kan+.* Finally, the kanamycin resistance cassette was excised using pCP20 recombination, resulting in the strain G6252 (Δ*lpd* Δ*poxB).* The plasmids harboring FDHs from *Thiobacillus sp.* and *C. necator*, pGEN1395 (pZE21-TsFDH) and pGEN1340 (pTRC-CnFDH) respectively, were subsequently transformed in G6252 strain by electroporation yielding the strains G6282 (Δ*lpd* Δ*poxB* pGEN1340) and G6284 (Δ*lpd* Δ*poxB* pGEN1395).

These strains were cultured with 20 mM pyruvate and 60 mM formate with or without 20 mM acetate to test if, without *poxB*, the conversion of pyruvate to acetate was still observed (Figure 28). It can be deduced from these growth curves that the impact of the *poxB* deletion is different depending on the FDH used. The construction harboring TsFDH is not affected by the *poxB* deletion since growth in the absence of acetate is sustained and identical between the two strains, meaning that another system than *poxB* must convert pyruvate to acetate. For the strains carrying the CnFDH, both conditions with and without acetate in the culture medium showed impaired growth when *poxB* was deleted. This suggests that the additional carbon source for this genetic context is necessary to support efficient growth. In one case or another, although altered, growth is observed in the absence of *poxB* and a medium lacking acetate, indicating that pyruvate serves as a source of acetate through an alternative pathway, acetate being essential for growth in the genetic context.

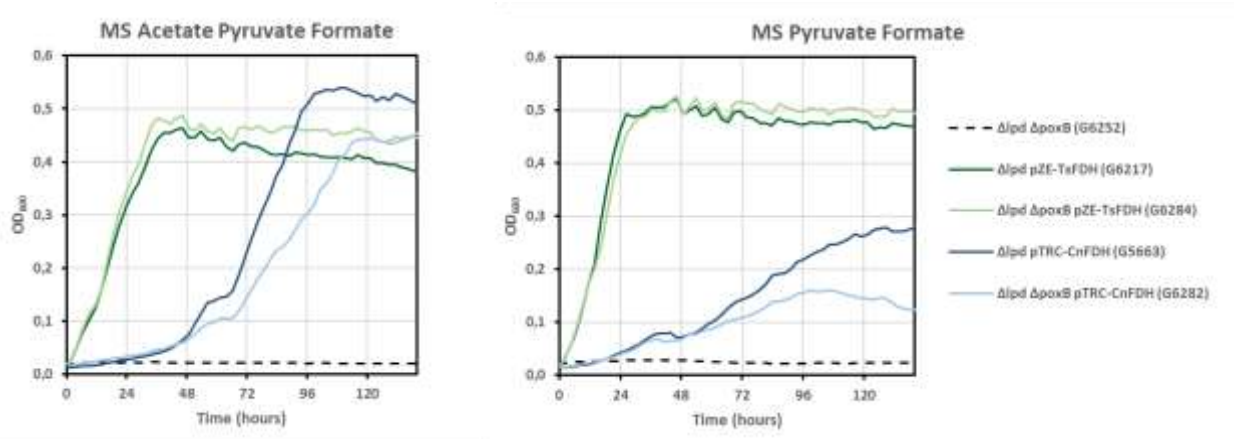


Figure 28. Growth of *poxB* deletion strains expressing CnFDH and TsFDH. Depending on the conditions, MS mineral medium was supplemented with the indicated compounds at 60 mM formate, 20 mM pyruvate and 20 mM acetate. Strains carrying pTRC-CnFDH were supplemented with 100 µM IPTG for gene expression.

2) IPTG and molybdenum supplementation

Following the initial tests, we further investigated the G5663 (Δ*lpd* pTRC-CnFDH) strain in the test medium supplemented with molybdenum, as this metal is involved in FDH activity (see Complex FDHs section, I. 3 2. 2. 2). Additionally, we examined the impact of IPTG supplementation on the test medium as we noticed growth without it during preliminary tests suggesting leakiness from the P_{trc} promoter (see sequence in Appendix 1: Genetic sequences of diverse building blocks).

Figure 29 clearly indicates that molybdenum supplementation to the liquid medium has no significant impact on growth. As molybdenum is already present in the NTA mix added to the MS minimal medium (as detailed in the Material and methods chapter II. 2. 2), it appears to be sufficient for efficient FDH activity in this particular context.

Furthermore, the addition of IPTG to the liquid medium, which induces FDH expression, results in a shortened lag phase. However, it is noteworthy that the strain was still able to grow without induction, supporting the hypothesis that the P_{trc} promoter is leaky causing some level of expression even in the absence of the inducer, which was already reported in the literature (Qiu, Swartz, and Georgiou (1998); Tegel, Ottosson, and Hober (2011)).

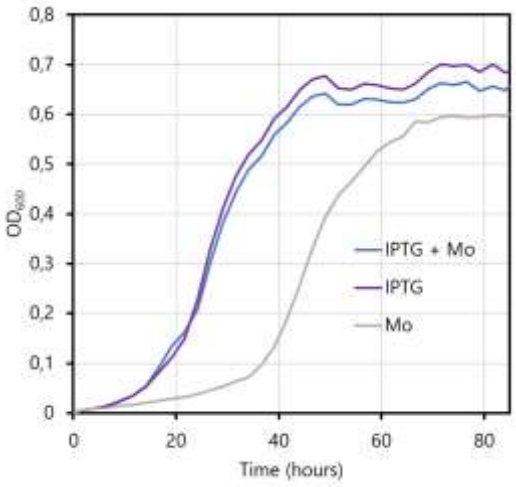


Figure 29. Growth impact of Mo and IPTG on strain G5663 (Δ*lpd* pTRC-CnFDH). Depending on the conditions, MS mineral medium was supplemented with the indicated compounds at 60 mM formate, 20 mM pyruvate, 20 mM acetate, 100 µM Mo and 100 µM IPTG.

3. Growth of strains harboring the CnFDH operon on two plasmids

The Δ*lpd* strain G5730 carrying the plasmids pGEN1338 (pUC-*fdsABG*) and pGEN1348 (pSU-*fdsCD*) was also tested. Growth of this strain was monitored under the same conditions as described above, MS minimal medium supplemented with acetate (20 mM), pyruvate (20 mM), and formate (60 mM). Since the genes were placed under constitutive promoters, no inducer was added to the cultures for this strain, the growth curves obtained through Bioscreen plate reader experiments are depicted in Figure 30.

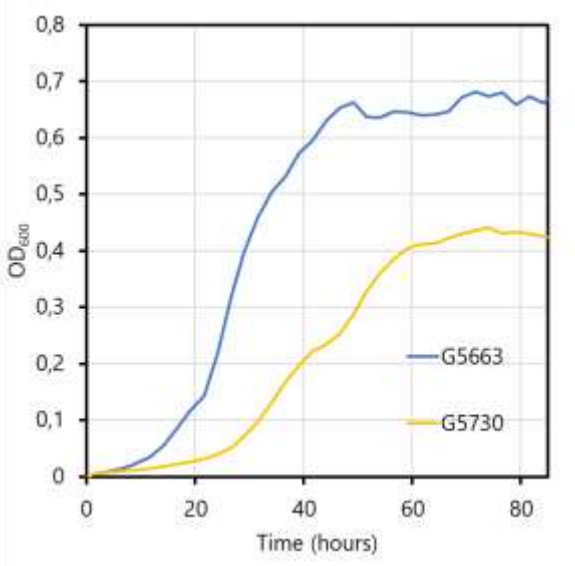


Figure 30. Comparison of the growth of strains G5730 (Δ*lpd* pGEN1338 pGEN1348) and G5663 (Δ*lpd* pGEN1340). The strains were grown in MS minimal medium supplemented with acetate (20 mM), pyruvate (20 mM) and formate (60 mM). For G5663 strain, growth medium was supplemented with 100 µM IPTG. G5730 and G5663 are represented by yellow and blue lines respectively.

Figure 30 reveals that G5730 strain exhibits a longer lag phase and a maximum OD_{600} more than 30% lower compared to the strain harboring the native operon under IPTG-inducible promoter (G5663). To explain these results, a few hypothesis can be proposed. Firstly, the two compared plasmid contexts (pTRC *vs* pUC and pSU) are different in terms of copy number and constitutive/inducible expression, which will inevitably create differences in expression levels, thus leading to phenotype variations. Secondly, the expression system with two plasmids leads to different expression levels between the constitutive genes *fdsABG*, cloned in a pUC plasmid, and the genes coding for the maturases *fdsCD*, cloned in a pSU plasmid with lower expression level. If FdsC and FdsD are not sufficiently expressed to ensure an efficient maturation of the protein complex, it could alter the functionality of the FDH.

Furthermore, only one restructuration of the CnFDH operon was chosen and might not be the most favorable one. For fine-tuning and optimization of the operon expression, other constructions could be designed and tested in the future using tailored cloning tools like a MoClo kit, for example the one developed by Paul Freemont's laboratory, EcoFlex MoClo kit designed especially for *E. coli* (Moore et al. 2016). This enables to test various gene orders, promoters, and RBS to identify the best operon conformation for a specific genetic context (Weber et al. 2011). To date, there are no published results regarding the remodeling of this CnFDH operon improving its expression or activity.

Ultimately, strain G5726 (Δ*lpd* pUC-*fdsABG* pSU-*fdsCD*) carrying plasmids containing the codon-optimized synthetic genes was tested in the same conditions. Yet, no growth was obtained with this strain after several attempts. One possibility is that codon optimization of the synthetic genes, performed by Twist Biosciences using the *E. coli* codon usage table, might have impacted the 3D structure of the protein due to a faster translation process. This issue was raised after a discussion about codon optimization for complex metal-dependent enzymes with Prof. Silke Leimkühler, who is a specialist of complex FDHs. The modification of gene order from the natural operon in this synthetic gene construction could also have led to alterations in gene expression and negatively impacted protein function.

Since the results obtained with the native operon cloned in the pTRC99a vector were conclusive, the decision was made to proceed with this construction (G5663 Δ*lpd* pTRC-CnFDH). Nevertheless, operon optimization remains an interesting option to be explored in future research.

4. Continuous culture in GM3 devices

The growth tests conducted showed that the native CnFDH operon was sufficiently active in catalyzing formate oxidation to support growth of the selection strain G5663. However, the generation time remained high, around 20 hours when cultured in 15 mL tubes. By contrast, the wild-type *E. coli* strain exhibits a generation time of around 65 minutes in a minimal medium supplemented with glucose.

To address this issue and select strains with faster growth rates, we conducted an evolution of strain G5663 (Δ*lpd* pTRC-CnFDH) in turbidostat regime in the GM3 device under the project name UOF. A pre-culture grown in the test medium was used for GM3 inoculation, consisting of MS minimal medium supplemented with acetate (20 mM), pyruvate (20 mM), formate (60 mM), and IPTG (100 µM). The evolution was then carried out in this medium.

1. *Isolates obtained from GM3 evolution*

Two turbidostat cultures, referred to as UOF1 and UOF2, were conducted in parallel in a GM3 device. This regime, functioning with a growth medium where all components are in excess, is best suited to select for fastest growing cells of a population. Figure 31 illustrates the evolution of generation times for both cultures, which showed similar evolution profiles. Both evolutions were characterized by a short adaptation phase during which the generation time dropped from about 4h40 to 3h40. A period of constant acceleration followed, until a generation time around 2h15 was reached after 30 days for both evolutions, corresponding to approximately 250 generations.

After 30 days of cultivation for UOF1 and 44 days for UOF2 (this slight difference is due to technical issues encountered with the samples), samples were taken from the cultures. These samples were then plated on Petri dishes with the same culture medium used for the evolution process. Three isolated colonies from each culture were streaked again and subsequently cultured for storage at -80 °C. The purpose of plating the samples was to obtain clones that could be further characterized through growth curve acquisition and genomic sequencing to understand the phenotypic differences observed between the parental strain and these evolved isolates.

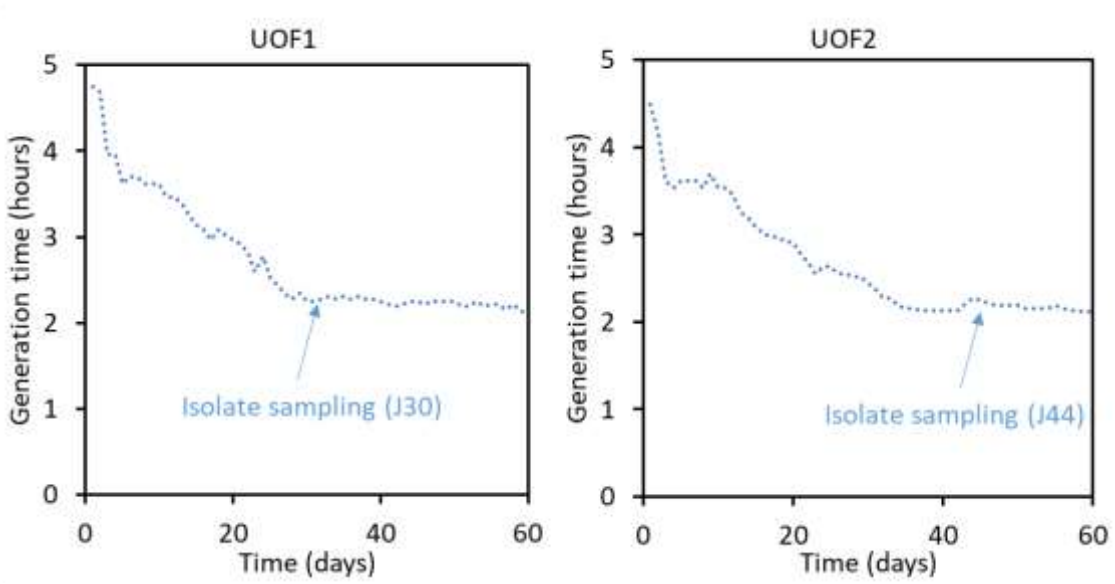


Figure 31. Evolution of the generation time of the two UOF evolutions from the G5663 (Δ*lpd* pTRC-CnFDH) strain. The evolutions were conducted in parallel in turbidostat regime using MS minimal medium supplemented with acetate (20 mM), pyruvate (20 mM), formate (60 mM) and IPTG (100 µM).

These isolated clones were also plated on Petri dishes containing permissive medium (MS medium supplemented with glucose (0.2 %), acetate (20 mM), and carbenicillin (50 µg/ml)) to control that the plasmid pGEN1340 was conserved during the evolution. All the tested clones still were carbenicillin resistant indicating the persisting presence of the pGEN1340 plasmid. This suggests that the strains remained dependent on CnFDH for NADH regeneration, despite being a burden.

1) Growth curves

In the Bioscreen analysis, three isolates from each evolution chamber, UOF1 (designated as G5823 to G5825) and UOF2 (designated as G5848 to G5850) were tested. Growth curves were generated to compare the growth in selective conditions of the evolved isolates with the parental strain G5663.

The results displayed in Figure 32 indicate a clear growth improvement through a shortened lag phase and a higher growth rate for all the tested isolates originating from both evolutions UOF1 and UOF2. This improvement suggests that there was an overall enhancement in growth characteristics compared to the parent strain G5663. In addition, there were no significant differences in the growth profiles among the different isolated strains, indicating a homogeneity of the population within the cultures.

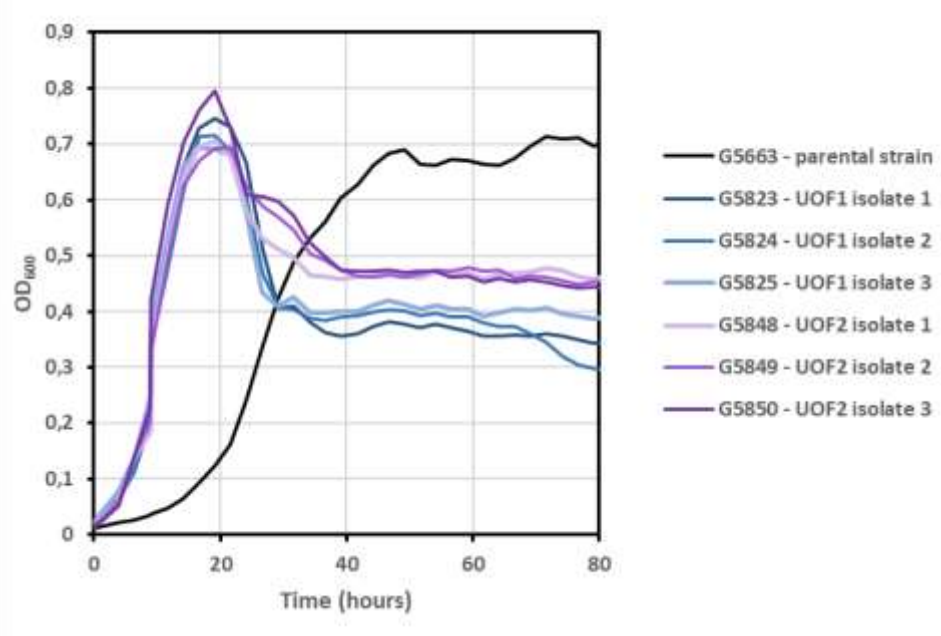


Figure 32. Growth of isolates obtained from UOF evolution in GM3. UOF1 isolates are represented with purple shades and UOF2 with blue shades. The growth medium consisted of MS minimal medium supplemented with acetate (20 mM), pyruvate (20 mM), formate (60 mM) and IPTG (100 µM). These data were obtained using a Bioscreen plate reader.

2) Genome sequencing

To better understand the evolution process of UOF1 and UOF2 cultures, we performed the genome sequencing of the six above-mentioned isolates using Illumina technology. We obtained sequencing data for both the chromosome and the plasmid pGEN1340 carrying the heterologous CnFDH operon. The sequencing reads were then mapped onto the genome of the ancestor strain to identify any mutations that may have occurred during the evolution process.

The results of the whole genome sequencing and the detected mutations are provided in detail in Appendix 5: UOF mutation analysis. The analysis revealed a total of 11 point mutations, with 8 of them being non-synonymous mutations. Additionally, 1 point mutation was found to affect an intergenic region, which could have implications for gene regulation and expression.

Notably, no mutations were observed in the pTRC-CnFDH plasmid (pGEN1340), indicating that the measured growth improvements were not directly linked to changes in the heterologous FDH operon affecting coding regions or expression regulating elements. The finding that no mutations were identified in the heterologous operon is not unexpected, it was already underlined in other evolution experiments with no mutations found in the foreign genes introduced in the host cells to establish a selectable phenotype (Antonovsky et al. 2016).

The identification of mutations in the *focA* gene, specifically the F7C and V97I, is of particular interest as these were found in isolates from independent evolutions conducted in parallel (UOF1 and UOF2). Interestingly, the F7C mutation was detected in two out of three isolates from UOF1 and one isolate from UOF2.

FocA is a pH-dependent bidirectional formate transporter that regulates intracellular formate levels, especially during fermentation (Kammel, Pinske, and Sawers 2022). This mutation detected in UOF evolutions, could influence the availability of formate as a substrate for formate oxidation by the FDH from *C. necator*. To evaluate the impact of these mutations in the isolates, further testing is detailed in the following section.

We decided to focus on the investigation of the FocA mutations because of the direct link that can be made with the observed phenotype. Interpreting the impact of mutations in intergenic regions and less-characterized genes is more challenging and necessitates extensive experiments, including the reversion of each mutation independently or in combination, to understand their impact on the reduction of the generation time observed during evolution.

2. *Tests with UOF isolates*

1) FocA WT restitution

To investigate the impact of the mutations in *focA*, we carried out a reverse mutagenesis by replacing the mutated allele of evolved UOF isolates with the WT allele through co-transduction. For this purpose, we generated P1 phages from the Keio collection strain carrying the *ycaP*::kan mutation, which is a non-essential gene located approximately 2000 bp downstream of *focA*. Next, the UOF isolates G5823 (*focA* F7C, UOF1) and G5848 (*focA* V97I, UOF2) were transduced with P1 *ycaP*::kan phages, and the transductants were selected on LB plates supplemented with kanamycin and carbenicillin to maintain the pTRC-CnFDH plasmid. The *focA* locus of the resulting clones was then sequenced to verify the restitution of the WT allele.

Subsequently, growth of isolates G5823 and G5848 and the corresponding strains harboring WT *focA* gene in the medium used for GM3 evolution was tested, the growth curves are available in Appendix 6: *focA* reversion in UOF isolates G5823 and G5848. The doubling times were determined graphically and the results are summarized in Table 27.

Table 27. Doubling times of evolved UOF isolates versus derivatives reverted to *focA* wild type. Cells were grown at 30°C in mineral MS medium supplemented with acetate 20 mM pyruvate 20 mM formate 60 mM in a Bioscreen plate reader in triplicate. Doubling times in exponential growth phase were graphically estimated.

lineage	strain	*focA*	doubling time
UOF1	G5823	F7C	5h30
	G6234	WT	17h30
UOF2	G5848	V97I	5h40
	G6235	WT	23h20

It is important to note that the calculated doubling times obtained from UOF isolates (approximately 5h30, Table 27) differ from those recorded by the GM3 device (approximately 2h, Figure 31). This variation can be attributed to differences in the culture conditions between a Bioscreen microtiter plate and the GM3 culture chamber. Various parameters are different between these two cultivation devices, including volume, aeration, and agitation of the cultures. Besides, in Bioscreen culture conditions, the calculated doubling time of the parental strain G5663 was around 20 h (Figure 27). As observed from the results in Table 27 above, for both strains, the restitution of the wild-type allele made the strains achieve nearly identical generation time to that of the parental strain. Consequently, it appears that *focA* mutations are primarily responsible for the phenotypic improvements observed during evolution.

This reverse mutagenesis experiment helped to understand the impact of the *focA* mutations detected on this specific phenotype. However, to better determine the role of these mutations on this bi-directional transporter capable of both import and export of formate, additional investigations would be necessary. The activity of this transporter is highly linked to fermentation since formate is the product of the pyruvate formate-lyase PflB complex, only active under anaerobic conditions. Especially at low pH, formate is then oxidized to CO_2 and H_2 by the FHL complex. The role of FocA is to regulate the formate concentration in cell compartments, cytoplasm and periplasm, for pH control. Under fermentation conditions, formate is first excreted outside the cytoplasm, either in the periplasm or in the medium and when the culture reaches late-exponential phase, formate is transported back into the cells (Kammel, Pinske, and Sawers 2022). To investigate the influence of these mutations on formate metabolism, particularly in terms of import, export or both mechanisms, an experiment could be carried out. An *E. coli* WT strain and a strain only carrying the *focA* mutations would be cultured in anaerobic conditions and samples would be collected at different time points during these cultures. Formate, produced by PflB, would be quantified in the cells and in the medium to evaluate the impact of the *focA*

mutations on formate fluxes. The interpretation of the results might be challenging as the activity of the two other complexes FHL and PflB are co-regulated with FocA and control formate homeostasis (Kammel, Pinske, and Sawers 2022). In parallel, transcriptional analysis of the three enzymes, using RT-qPCR for instance, in each condition might be necessary as well to assess gene expression levels and formate import/export.

2) Construction of an evolved background cured from the plasmid

To test alternative constructions, an isolate from UOF evolution, G5823, was cured from its plasmid. To do so, a growing culture of the strain was diluted successively five times in permissive mineral medium without antibiotic, *i.e.* MS medium supplemented with glucose (0.2 %) and acetate (20 mM). The loss of the plasmid was verified on plates and a clone sensitive to carbenicillin was selected and named G5876.

To validate this construction, we electroporated back the initial plasmid with which the evolution in GM3 was carried out, pGEN1340 (pTRC-CnFDH), creating the strain G6504. Then, a Bioscreen analysis was conducted with the initial isolate G5823 as a positive control (Figure 33). This test demonstrates that strains G5823 (UOF isolate) and G6504 (cured and transformed UOF isolate) exhibit similar growth patterns, thus validating the evolved and cured background of the strain G5876.

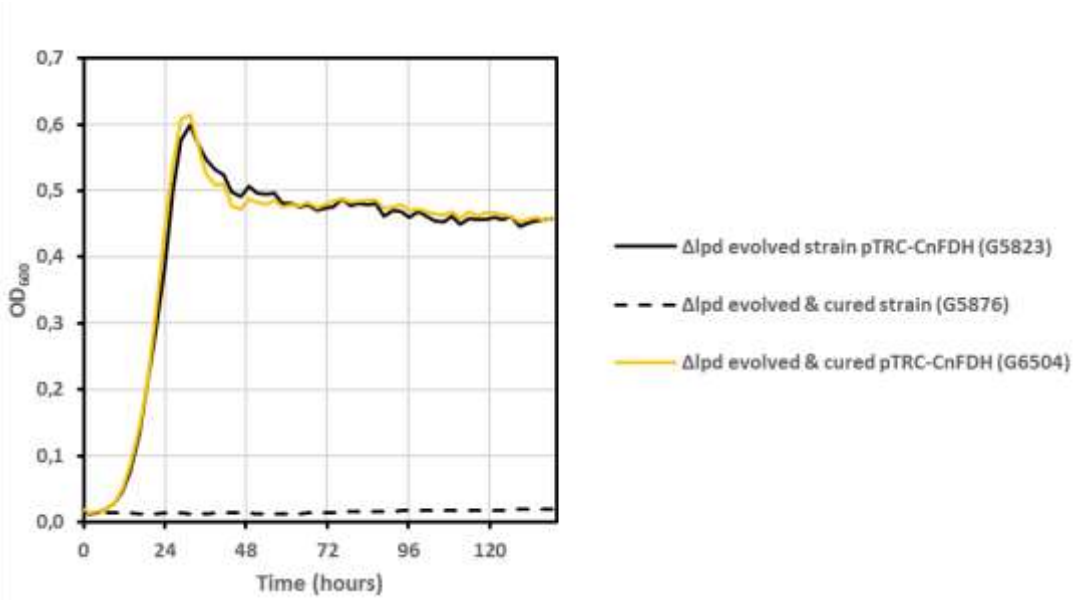


Figure 33. Dependence of evolved isolate G5823 (turbidostat culture UOF1) on the presence of plasmid-borne CnFDH for growth on formate as sole energy source. Strains G5823, G5876 (derivative of G5823 cured from the plasmid pGEN1340), and strain G6504 (derivative of G5876 retransformed with plasmid pGEN1340) were grown at 30°C on mineral MS medium supplemented with formate (20 mM), acetate (60 mM), pyruvate (20 mM) and IPTG (100 µM) and growth recorded with a Bioscreen plate-reader.

3) Influence of the evolved background on FDH activity

To investigate whether the genotypic context obtained during the evolution was specific to CnFDH or had broader applicability, we used the FDH gene from *Thiobacillus sp.* KNK65MA (TsFDH, Nanba, Takaoka, and Hasegawa (2003)) cloned into a high-copy plasmid (pZE21 backbone) and CnFDH cloned into this plasmid as well in order to obtain the same plasmidic context for comparison. Both constructions, pGEN1393 (pZE-CnFDH) and pGEN1395 (TsFDH), were described above (III. 1. 1. 1) and are represented in Figure 25.

In this study, TsFDH was chosen as a reference representing simple NAD^+-dependent enzymes. It was shown to have relatively high CO_2 reduction activity as compared to other simple FDH (see comparison Table 2 from the Introduction), although its turnover number (k_{cat} = 0.318 s^{-1}) is much lower compared to CnFDH (k_{cat} = 11 s^{-1}).

To test these constructions for their formate-oxidation properties, plasmids pGEN1393 (pZE-CnFDH) and pGEN1395 (pZE-TsFDH) were transformed in the metabolic screening strain G5416 Δ*lpd* unevolved as well as an UOF evolved background cured from the pGEN1340 plasmid, the strain G5876 (III. 1. 1. 4. 2. 2)). The strains with the unevolved background were named G6114 (Δ*lpd* pZE-CnFDH) and G6217 (Δ*lpd* pZE-TsFDH) and with the evolved background, G6280 (Δ*lpd* evolved pZE-CnFDH) and G6272 (Δ*lpd* evolved pZE-TsFDH). Subsequently, the resulting strains were tested in the MS minimal medium supplemented with acetate (20 mM), pyruvate (20 mM) and formate (60 mM), without inducers as the expression from this plasmid is constitutive (Figure 34).

From Figure 34, it is observed that independently of the FDH expressed, the strains with the evolved background (solid lines) show faster growth compared to the parental strains (dashed lines), indicating the adaptability of the evolved background.

For TsFDH expression, both the parental (G6217) and the evolved (G6272) background are showing better growth compared to the corresponding strains with CnFDH, with approximately 50 % longer generation times obtained with the strains carrying the CnFDH. This result suggests that besides less favorable kinetic parameters determined *in vitro* for TsFDH (see Table 2 from the introduction) compared to CnFDH for formate oxidation, it does not reflect in this experiment *in vivo* in *E. coli* (this point is further discussed in the III. 6. 1).

Taking into consideration all the experiments performed with CnFDH, including cloning and cell culture with long lag phases, there are clear indications that expression of this enzyme perturbs growth of the *E. coli* host cells, possibly due to incorrect folding of this highly complex catalyst. However, we can observe that UOF evolution substantially reduced the lag phase, and the difference in growth between the TsFDH and CnFDH expressing cells is much reduced in this genetic context (Figure 34). Hence, it is

most likely that the evolution process both optimized the carbon fluxes and improved the expression of CnFDH. For the first observation, the direct link between genotype and phenotype was demonstrated through the *focA* reversion experiment. Yet, the second observation could not directly be linked to a mutation detected with the sequencing data analysis.

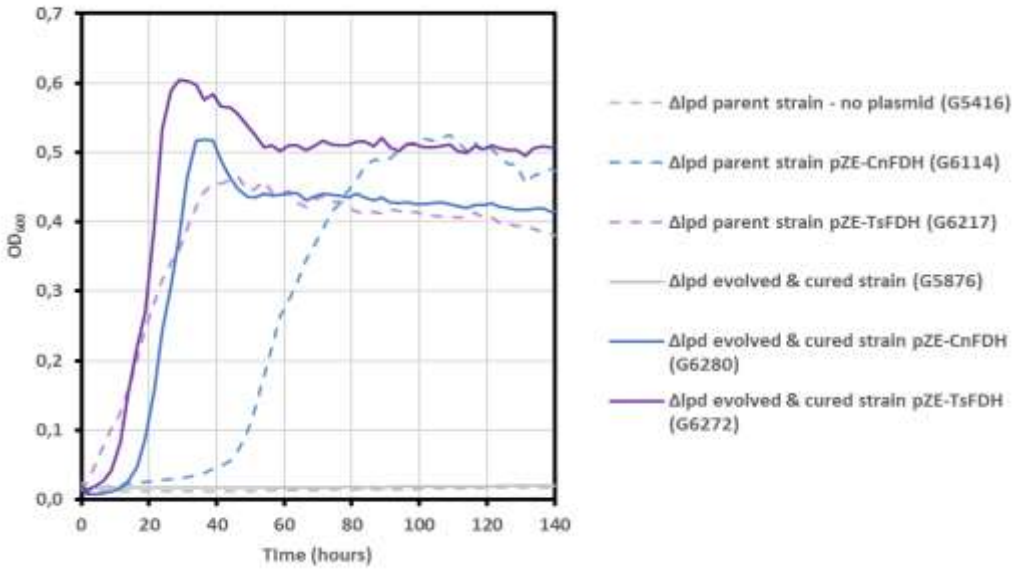


Figure 34. Growth of the Δ*lpd* evolved and cured background (G5786) with CnFDH and TsFDH in pZE plasmid compared to the parent strain with the same plasmids. Cultures were conducted in MS minimal medium supplemented with acetate (20 mM), pyruvate (20 mM) and formate (60 mM).

2. Expression of the CnFDH from the chromosome

Plasmid-borne expression of genes, especially heterologous ones, in selection contexts can cause difficulties due to genetic burden and diminished strain stability. Plasmids can be lost upon non-selective passages, can show heterogeneous replication patterns or undergo « illegal » recombination with the chromosome.

To enhance the stability of our engineered genetic system, especially with the perspective of long-term directed evolution experiments, we chose to integrate the CnFDH operon into the chromosome of strain G5876. This strain, cured from the pTRC-CnFDH plasmid, possesses the evolved UOF background favorable for FDH activity (III. 1. 1. 4. 2. 2)).

The integration process is detailed in the Material and methods (II. 4. 6) section. In brief, we cloned the CnFDH operon into a plasmid (pDM4_IS10) that facilitates integration at a specific locus known as IS10.

This locus has been previously identified as an efficient integration site and corresponds to the non-essential *ula* operon (Bassalo et al. 2016). In addition, these "safe spots" were already used and reported to allow successful expression of heterologous enzymes in *E. coli* genome (Kim et al. 2020). The resulting genetic context is illustrated in Figure 35.

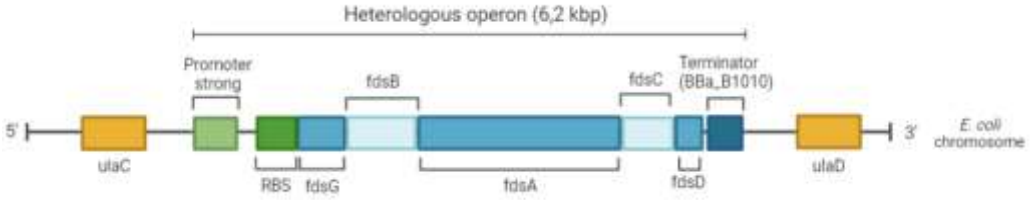


Figure 35. CnFDH operon integrated in *E. coli* chromosome.Designed on Biorender.

1. Plasmid construction for genome integration of CnFDH

The process of cloning the CnFDH operon into the plasmid proved to be challenging, primarily due to the substantial size of the resulting construction, which reached 16 kbp (10 kbp for the plasmid backbone and 6 kbp for the operon) in addition to the high GC content of the CnFDH already mentioned above. Traditional restriction-based methods, including Golden Gate cloning, were not possible for this particular construct. Consequently, we decided to use Gibson cloning which necessitates an amplification of the entire backbone, but we encountered difficulties in obtaining a successful PCR product for the 10 kbp fragment necessary. As a result, we divided the backbone into two parts. Subsequently, the Gibson cloning method was completed, with three segments: two for the vector and one for the CnFDH operon. After these efforts, we successfully obtained one clone, which was validated through whole plasmid sequencing (Nanopore sequencing performed by Plasmidsaurus) to confirm the integrity of the construct. The resulting construction was named pGEN1378 (pDM4_IS10::CnFDH) and is represented in Figure 19.

2. Strain construction

The construction procedure is elaborated in detail in the Material and methods section (II. 4. 6). In brief, the construction involved a conjugation process between two strains, where the recipient strain (*E. coli* MG1655 WT) lacked the ability to replicate a plasmid with a Pir origin as the pDM4 plasmid. This setup allowed for the selection of clones with the plasmid integrated into the chromosome upon recombination. Subsequently, the excision of the plasmid backbone was positively selected on a medium supplemented with sucrose, which is toxic in *E. coli* expressing sacB gene.

To transfer the integrated CnFDH operon from *E. coli* MG1655 IS10::CnFDH kan+ (G6012) to the recipient strain with the UOF evolved background (G5876), P1 phages were prepared. These phages were then used to transduce the G5876 strain, and clones were selected on LB plates supplemented with kanamycin. Subsequently, the kanamycin resistance gene was excised using the FLP-mediated system (II. 4), resulting in the strain G6085 (Δ*pd* evolved IS10::CnFDH). To confirm the successful integration of the entire operon, verification was conducted through colony PCR.

3. GM3 evolution

1. *First attempt of GM3 evolution*

The strain G6085 was tested for growth on formate as energy source using MS medium supplemented with acetate (20 mM), pyruvate (20 mM) and formate (60 mM). Unfortunately, no growth was obtained in these conditions. The hypothesis was made that despite the strong promoter and RBS that were employed, the expression from the single copy of CnFDH in the chromosome was not sufficient to support growth.

Consequently, it was decided to evolve the strain into a GM3 evolution under a medium swap regime using the conditions detailed in Table 28, under the project name of OCF. This regime, alternating permissive and stressing dilution pulses according to the density of the population with respect to a set threshold, can fix mutations favoring the shift from permissive to non-permissive growth conditions in an incremental way. The glucose concentration of the permissive medium was lowered to 0.1 % (usually used at 0.2 %) to decrease the adaptation step between the permissive and the stressing media. First, the strain is cultured under a turbidostat regime in the permissive medium to ensure adaptation of the culture to growth in the GM3 and to determine the generation time to impose for the medium swap regime. Subsequently, the medium swap is started.

Table 28. Description of the medium used for GM3 evolution (OCF cultures) of the strain G6085 Δ*pd* evolved IS10::CnFDH

MS medium	Glucose	Acetate	Pyruvate	Formate
Permissive medium	0.1 %	20 mM	20 mM	60 mM
Stressing medium	-	20 mM	20 mM	60 mM

Figure 36 depicts the evolution of the generation time and the percentage of stressing medium in the two separate cultures, OCF3 and OCF4, of strain G6085 during the GM3 evolution. Notably, both OCF3 and OCF4 cultures reached complete adaptation to the stressing medium, corresponding to 100 % of

stressing medium dilution pulses, after 65 days and 63 days, respectively. The cultures were continued in a turbidostat regime using a medium with the same composition as the stressing medium. The turbidostat enabled both cultures to lower their generation time to about 2 hours (GM3 culturing conditions).

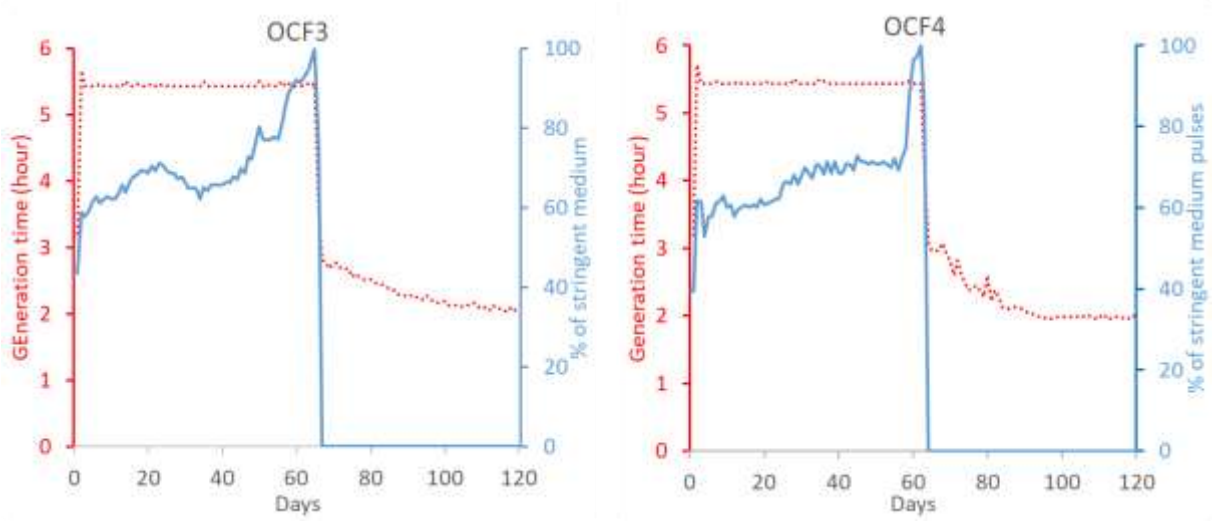


Figure 36. Evolutionary kinetics of OCF3 and OCF4 cultures of strain G6085 (Δ*lpd* evolved IS10::CnFDH) conducted in a GM3 device. The red line represents the evolution of the generation time while the blue line represents the percentage of stressing medium relative to the permissive medium fed to the cultures.

After completing the adaptation process, samples were taken from both chambers, and individual clones were selected on plates for sequencing. Initially, we focused on sequencing the IS10 locus containing the CnFDH operon. To our surprise, we discovered a deletion of 63 bp in the first gene of the operon, *fdsG* (see Figure 37), from the base pair 33 to 94 and one deletion at base pair 19 of the 528 bp of the gene. This deletion was present in all the sequenced isolates. To better understand our findings, we examined the parent strain (G6085) and identified a deletion of 62 bp in the exact same region, from position 33 to 94, which was unfortunately not detected before the evolution. This deletion in the parental strain is one base pair less than the mutation observed in the evolved clones. It appears that during the evolutionary process, a specific mutation (deletion of cytosine 19 of *fdsG*) was selected that restored the reading frame of the gene, 63 being a multiple of 3, apparently restoring the activity of the heterologous CnFDH.

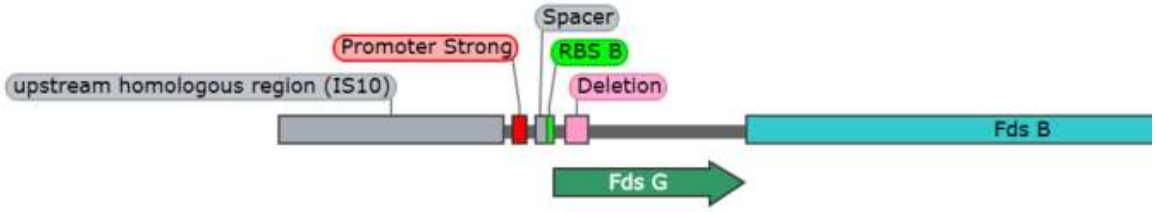


Figure 37. Representation of the deletion found in *fdsG* gene from the CnFDH operon integrated in the chromosome of the strain G6085. The deletion is represented in pink.

To verify whether this truncated CnFDH was indeed functional, we subcloned the unevolved version (62 bp deletion in *fdsG*) and the evolved version harboring the additional 1 bp deletion in the pTRC vector. The resulting constructions, obtained with the Gibson cloning technique, were named pGEN1425 (pTRC-CnFDH Δ63 bp, after evolution) and pGEN1426 (pTRC-CnFDH Δ62 bp, before evolution). These two plasmids were transformed in the UOF evolved and cured background (strain G5876) and strains G6403 (Δ*lpd* evolved pTRC-CnFDH Δ63 bp) and G6405 (Δ*lpd* evolved pTRC-CnFDH Δ62 bp) were obtained. The strains, along with G5823 as a positive control (UOF1 isolate containing pTRC-CnFDH), were tested on plates for growth on formate as energy source in MS minimal medium supplemented with acetate (20 mM), pyruvate (20 mM), formate (60 mM) and IPTG (100 µM). Growth on plates was observed for strain G6403 but not for strain G6405. Compared to the control strain G5823, both the number and the size of the colonies obtained were reduced with the strain G6403. Eventually, the GM3 evolution has enabled the selection of a functional CnFDH despite the initial unintended deletion in *fdsG*, although the selected mutant with Δ63 bp seems less active than the native operon. Still, it is an interesting result to observe that this type of medium swap evolution was able to select such a mutation even in a heterologous gene, which in the past has been found to be difficult targets for GM3 evolution with respect to native genes. It also attests the tightness of the selection and the pressure exerted on the cells to access formate as energy source.

2. *Second strain construction and evolution*

1) Construction and testing

Given that the objectives of the project required a fully functional *C. necator* FDH, we re-attempted the chromosomal integration of the CnFDH operon to obtain a strain with the complete operon in the chromosome. To do so, the integration step in wild-type *E. coli* MG1655 was performed again without changes in the protocol and a clone with the complete sequence of CnFDH was selected. The selected clone, G6408 (MG1655 IS10::CnFDH kan+), was used to produce P1 phages and transduce the same

evolved background (G5876) which resulted in the construction of the strain G6435 (Δ*lpd* evolved IS10::CnFDH). Interestingly, the 62 bp deletion that was observed in the selected clone from the previous experiment was found again in many sequenced clones from this new experiment. A few hypotheses can be made as to why this event tends to happen. Firstly, the CnFDH is still toxic and represents a burden for the strain, even more if considering the strong constitutive promoter placed upstream of this operon and the growth impact that the lipoamide dehydrogenase (Lpd) deletion has for the strain. Secondly, this specific part of the *fdsG* gene has a high GC content of 74 % probably making the replication process challenging for DNA polymerases. Nevertheless, we were finally able to obtain a correct chromosomal integration of the *fdsGBACD* operon.

We then conducted growth tests with strain G6435 in MS acetate/formate/pyruvate medium. The growth curve obtained for this strain is presented in Figure 38. The graphically calculated generation times were 5h20 and 7h50 for strains G5823 and G6435, respectively.

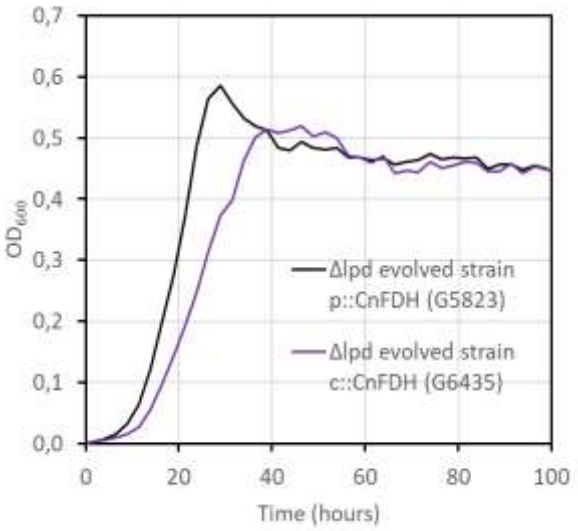


Figure 38. Comparison of growth of the strains G5823 and G6435 in test medium. The test medium consisted of MS medium supplemented with acetate (20 mM), pyruvate (20 mM), formate (60 mM) and IPTG (100 µM) for G5823.

Additionally, we conducted a comparative analysis between evolved isolates from UOF evolution, G5823, expressing CnFDH from a pTRC plasmid, and the strain containing CnFDH into the chromosome, G6435 (Figure 39). The objective of this comparison was to assess growth yields under varying formate concentrations in the growth medium.

The results from this experiment suggest that the expression from the multi-copy inducible plasmid pTRC (in strain G5823) is more sensitive to the formate concentration available in the medium. At the

lowest formate concentrations tested, specifically 20 and 40 mM, the G6435 strain, expressing CnFDH from the chromosome, demonstrated a more efficient growth. This indicates that the expression obtained from the chromosomal integration is highly efficient, potentially counter-balancing the burden associated with the multi-copy plasmid, especially in low-formate conditions. For the higher concentrations tested, the growth curves are consistent with our previous experiment. Overall growth rate was higher for G5823 cells expressing FDH from the plasmid. The final OD_{600} of both cultures increased with increased formate concentration, the maximum OD being reached at 80 mM for both strains.

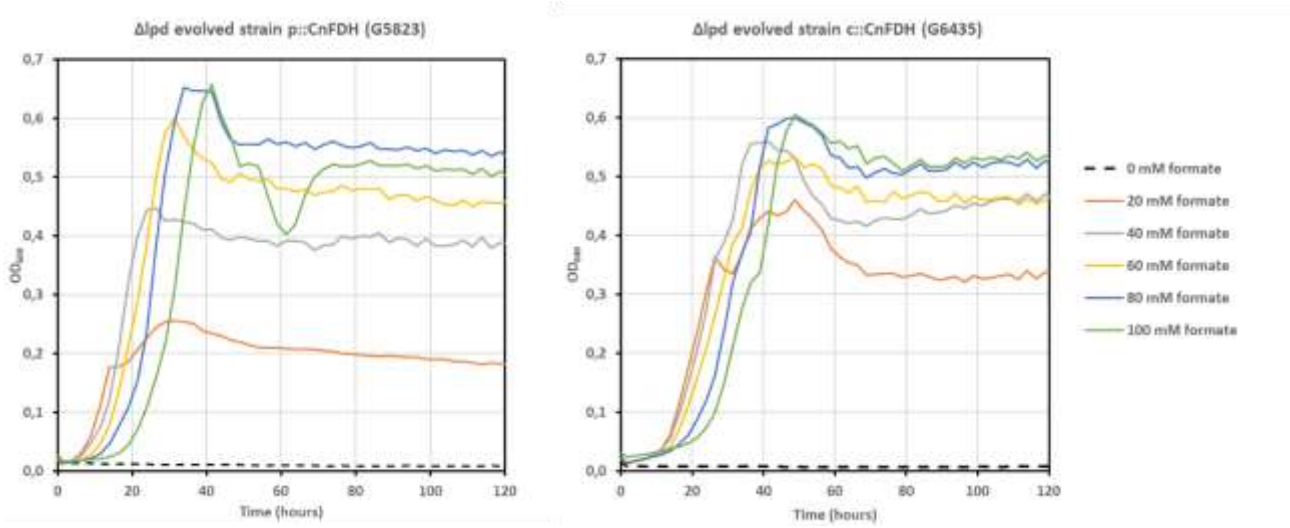


Figure 39. Growth yield of *C. necator* FDH expressing Δ*lpd* strains depends on formate concentration in the medium. G5823 (A) and G6435 (B) cells were grown in medium MS acetate 20 mM pyruvate 20 mM supplemented with formate at concentrations ranging from 0 to 100 mM. Experiments were conducted with a Bioscreen plate reader as described in the material and method section.

Finally, an RT-qPCR comparison of the transcription level of the *fds*-operon between the strains G5663 (pTRC-CnFDH), its evolved derivative G5823 and strain G6435 containing the operon in the chromosome was performed to verify that the growth differences on formate as energy source observed between strains were linked to differences in *fds* gene expression. The results are presented in Figure 40.

Without surprise, expression of the three genes *fdsG*, *fdsA* and *fdsD* in the G5663 strain (non-evolved Δ*lpd* pTRC-CnFDH) was lower in permissive than in selective medium. For all context tested, the expression level decreased for genes more distant from the promoter indicating distance-dependency of the processivity of the RNA polymerase (Wang, Yue, et al. 2021). The expression levels between strains G5663 and G5823 are similar for genes *fdsG* and *fdsA*, with a difference observed for *fdsD*. The mutational analysis did not reveal any mutations in the plasmid pGEN1340, the expression level is then

expected to remain similar, and the difference observed for *fdsD* is not easily explicable. When comparing the two expression formats, plasmid and chromosome, a weaker expression is found for the G6435 strain expressing CnFDH from its chromosome. This is consistent with the lower growth rate observed in the comparative growth experiment described above (see Figure 38). We note the relatively high expression of the operon in the non-induced G5663 strain in permissive medium. This demonstrates the leakiness of the P_{trc} promoter, which was suggested in *IPTG and molybdenum supplementation* section (III. 1. 1. 2. 1. 2)), and is here confirmed with this RT-qPCR experiment.

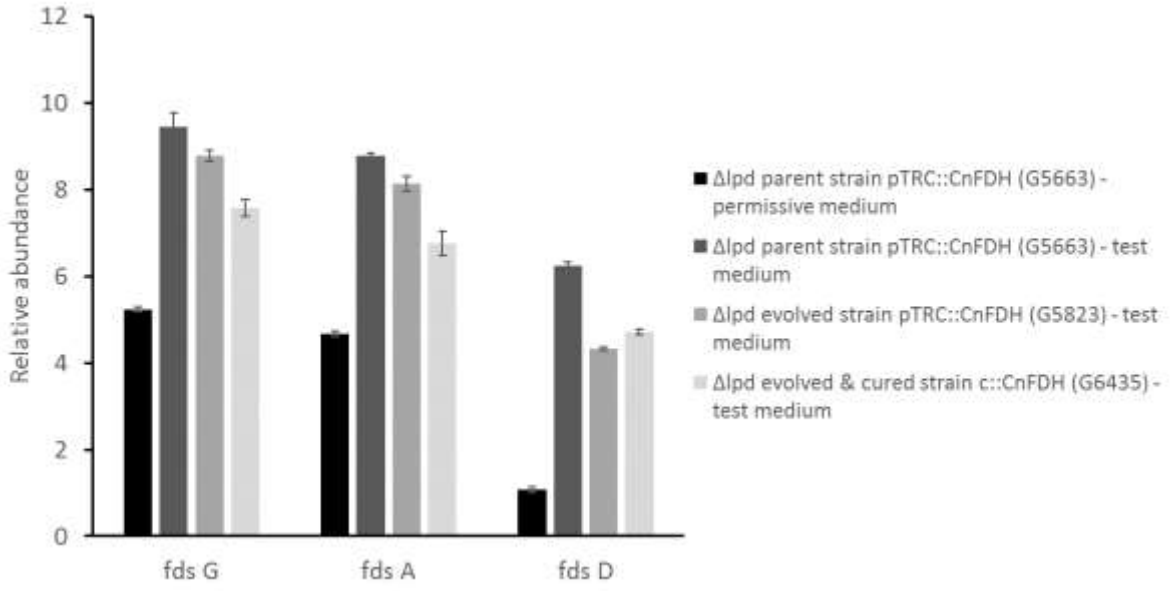


Figure 40. RT-qPCR of *fdsG, fdsA* and *fdsD* gene transcripts from the CnFDH operon, expressed in the indicated strains during growth in the indicated conditions. The permissive medium used was MS minimal medium supplemented with glucose (0.2 %), acetate (20 mM) and carbenicillin (50 µg/ml). The test medium was MS minimal medium supplemented with acetate (20 mM), pyruvate (20 mM), formate (60 mM) and IPTG (100 µM) for G5663 strain. This experiment was performed as described in the Material and methods section, with *panB* used as the housekeeping reference gene.

2) Global comparison

Many genetic constructions were described throughout this chapter; Table 29 summarizes the main results. For some experiments, we encountered difficulties to obtain reproducible results especially when testing the unevolved Δ*lpd* background possibly due to the toxicity and the burden that the heterologous system imposed to the cells, the Δ*lpd* strain being already impaired as compared to WT. However, clear trends emerged and the growth ameliorations obtained through GM3 evolution were compelling.

Table 29. Summary of the growth results obtained using Δ*lpd* parent and evolved strains with different FDHs and expression formats.

Strain	Expression format	Expression type	Generation time	Maximum OD_{600}
G5663 **Δ*lpd* - parent**	pTRC-CnFDH (pGEN1340) native genes	Inducible	10h55-19h38	0.55-0.7
G5726 **Δ*lpd* - parent**	pUC-*fdsABG* (pGEN1323) + pSU-*fdsCD* (pGEN1320) synthetic genes	Constitutive	No growth	NA
G5730 **Δ*lpd* - parent**	pUC-*fdsABG* (pGEN1338) + pSU-*fdsCD* (pGEN1348) native genes	Constitutive	13h32	0.45
G5823 **Δ*lpd* - evolved**	pTRC-CnFDH (pGEN1340) native genes	Inducible	5h12	0.6-0.8
G6114 **Δ*lpd* - parent**	pZE-CnFDH (pGEN1393) native genes	Constitutive	18h54	0.5
G6217 **Δ*lpd* - parent**	pZE-TsFDH (pGEN1395) synthetic gene	Constitutive	12h09	0.45
G6272 **Δ*lpd* - evolved**	pZE-TsFDH (pGEN1395) synthetic gene	Constitutive	4h36	0.6
G6280 **Δ*lpd* - evolved**	pZE-CnFDH (pGEN1393) native genes	Constitutive	6h48	0.5
G6435 **Δ*lpd* - evolved**	IS10::CnFDH native genes	Constitutive	7h48	0.5

3) GM3 evolution of the strain expressing CnFDH from the chromosome

We then conducted a GM3 evolution in a turbidostat regime to obtain faster growing cell populations. The strain G6435 was cultured in the same test medium as described above. These new evolutions were named OCF5 and OCF6 and are presented in Figure 41.

Samples were taken after 56 and 60 days (around 500 generations) from OCF5 and OCF6 cultures respectively, for functional and mutational analysis. Four isolates, two from each evolution chamber, OCF5 (G6539 and G6540) and OCF6 (G6541 and G6542) were tested.

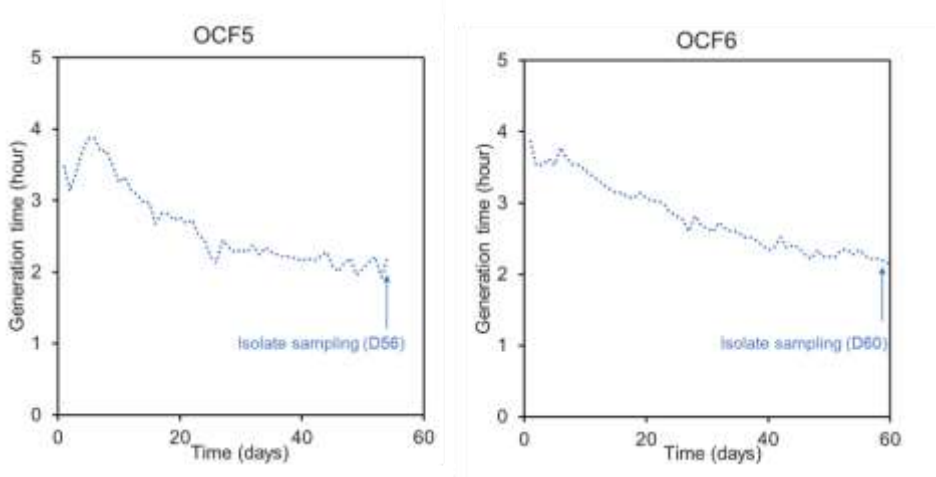


Figure 41. Evolution kinetics of cultures OCF5 and OCF6 with G6435 cell populations. Cells were grown under a turbidostat regime in MS medium supplemented with acetate (20 mM), pyruvate (20 mM) and formate (60 mM).

In the Bioscreen analysis, the four isolates were included and growth curves were generated to compare the growth in selective conditions of the evolved isolates with the parental strain G6435 (Figure 42). As depicted in the Figure 42 below, the growth of both isolates from OCF5 and OCF6 showed improved growth rates and shorter lag phases compared to the parent strain G6435. The generation times were calculated graphically and are homogenous between all the isolates, ranging from 3h40 to 4h20, in line with the data provided by the GM3 with similar growth rates between both cultures as well. However, we observe a slightly shorter lag phase for OCF6 isolates.

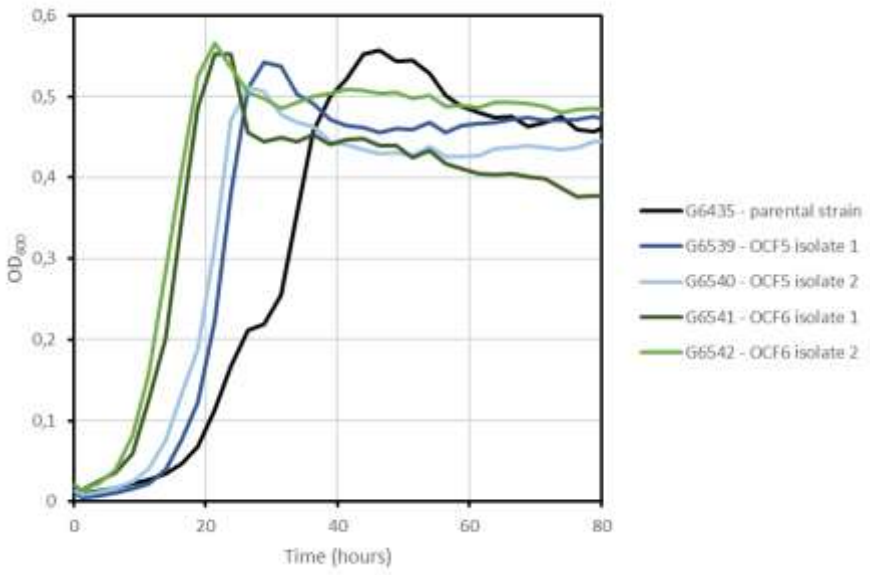


Figure 42. Growth of isolates obtained from OCF5 and OCF6 evolutions in GM3. OCF 5 isolates are represented in blue shades and OCF6 with green shades. Growth medium: MS minimal medium supplemented with acetate (20 mM), pyruvate (20 mM) and formate (60 mM). These data were obtained using a Bioscreen plate reader.

Genomic DNA was extracted and sent for Nanopore sequencing to Plasmidsaurus. Mutational analysis was performed using the computational pipeline breseq at the Genoscope. The results of the whole genome sequencing and the identified mutations are provided in detail in Appendix 7: OCF5 and OCF6 mutation analysis. The analysis of the data revealed a different evolution path between the isolates of OCF5 and OCF6 populations.

OCF5 isolates G6539 and G6540 differed from the ancestral strain G5876 (cured derivative of UOF1 isolate G5823) by one non-synonymous point mutation fixed in the gene *amiB* (N445S). However, this mutation was detected in the two isolates of the OCF6 lineage as well, suggesting that it was already present in the cell population used to inoculate the two parallel GM3 devices. The fact that no mutations were fixed in the genome of the evolved isolates over 500 generations in turbidostat is surprising in view of the clear decrease of the generation time (almost 2 hours) of the OCF5 population between the start of the culture and the sampling at Day 50 from which the two sequenced isolates originate. A metagenomic analysis of the population sampled at Day 50 could be considered to evaluate the diversity and abundance of acquired mutations in the evolved population.

By contrast, 4 and 2 point mutations were detected in OCF6 isolates G6541 and G6542, respectively. None of these mutations were common between the two isolates, but, interestingly, impacted the same gene or function. Short deletions were identified in the gene *sspA* encoding stringent starvation protein A. They lead to localized modifications of the protein through the loss of one (tyrosine 21) or seven contiguous amino acids (VDRELTL63-69). Non-sense mutations were detected in the genes *nuoC* (W11* in isolate G6541) and *nuoI* (E165* in isolate G6542) that generate shortened inactive proteins. The two *nuo* genes code for subunits of the NADH:ubiquinone oxidoreductase, whose activity depends on the integrity of all its components (Erhardt et al. 2012). Thus, it seems that the evolutionary process selected for the inactivation of the NADH dehydrogenase complex I. It can be hypothesized that a non-functional NADH dehydrogenase complex I modifies the intracellular NAD^+/NADH ratio and benefits to the NADH requirement of the energy auxotrophic cells.

4. CRISPRi

1. *gRNAs in fdsG and promoter region*

To validate the essentiality of the chromosomal insertion of CnFDH in the strain G6435 (Δ*lpd* evolved IS10::CnFDH) for growth on formate as energy source, a CRISPR interference protocol for gene silencing (CRISPRi) was applied. As elaborated in the Material and methods chapter (II. 5. 8), four guides specific to the heterologous insertion, especially the promoter region and *fdsG* as it is the first gene of the operon, were designed using an online Tool from Pasteur Institute (CRISPR@Pasteur). As it was reported

in the reference publication, the silencing of a gene within an operon will block the expression of all the downstream genes (Depardieu and Bikard 2020). The three guides with the highest prediction score for site-specific recognition were chosen, and one additional guide, gRNA1 showing a lower score but located in the promoter region was also used (Table 30 and Figure 43).

Table 30. gRNAs designed for CnFDH operon silencing. gRNAs and prediction scores were obtained from the CRISPR @Pasteur tool.

Guide	Sequence	Prediction score
gRNA1	TTTCTATTAACTAGTGAATT	0.32
gRNA2	GACCTGTACCACGTGGCGCC	1.38
gRNA3	GGTGATCACGCCATGCACCT	1.34
gRNA4	CGCATGCTCGGCCAGCGCTT	1.16

Figure 43. Representation of the positions targeted by the four gRNAs selected for CRISPRi gene silencing.

Following the successful cloning of all four gRNAs in the pFD152 plasmid carrying the gene encoding dCas9 under a promoter inducible with anhydrotetracycline (aTc) and a cloning site for gRNA, four plasmids were obtained, namely pGEN1416 (pFD152-gRNA1), pGEN1418 (pFD152-gRNA2), pGEN1420 (pFD152-gRNA3) and pGEN1422 (pFD152-gRNA4). Each plasmid was transformed in the G6435 strain along with the pFD152 plasmid without any targeting guide as a control. Two isolated colonies were selected from each transformation and adapted to permissive minimal media composed of MS medium supplemented with glucose (0.2 %), acetate (20 mM) and spectinomycin (50 µg/ml) to ensure the strains retained the pFD152 derivative plasmids. These adaptations were used to perform liquid cultures in the same conditions. Subsequently, dilutions (10^{0} to 10^{-8}) of these overnight cultures were prepared in MS minimal medium.Then, 5 µL of each dilution were patched onto four different plates, media compositions are detailed in Table 31. Finally, the plates were incubated at 30 °C.

Table 31. Medium composition for CnFDH gene silencing experiments conducted in strain G6435 (Δ*lpd* evolved IS10::CnFDH). aTc = anhydrotetracycline

	Glucose	Acetate	Pyruvate	Formate	aTc	Spec
Permissive w/o induction	0.1 %	20 mM	20 mM	60 mM	-	2000e
Permissive with induction	0.1 %	20 mM	20 mM	60 mM	0.5 µg/ml	2000e
Test w/o induction	-	20 mM	20 mM	60 mM	-	2000e
Test with induction	-	20 mM	20 mM	60 mM	0.5 µg/ml	2000e

After incubation, pictures of the plates are taken (Figure 44). From these pictures, a few observations can be made starting with a clear toxicity of dCas9 upon induction with aTc. In the permissive plates, where this strain should grow actively, there is significantly reduced growth on the plate containing the inducer. Harboring the *lpd* deletion impairing growth, strain G6435 shows a clear sensitivity to the dCas9 enzyme. The toxicity of this modified Cas9 enzyme was already reported in the literature (Cui et al. (2018); Depardieu and Bikard (2020)). It was later demonstrated to result from off-target binding of gRNA/dCas9 complex to the promoters of essential genes that share no more than four nucleotides of identity in the PAM-proximal sequence (Rostain et al. 2023). Consequently, there is almost no growth observed in test conditions even with the strain only carrying the native pFD152 plasmid with no specific gRNA. Besides, this aTc-inducible promoter seems to be leaky as an effect on growth depending on the gRNA can be observed even without induction.

The second observation is that the two clones of G6435 strain carrying the plasmid pFD152 with a non-targeting gRNA showed growth in the selective medium without aTc-induction which is a validation for this experiment. Additionally, the gRNA1 that is located between the promoter and the RBS (the last two lines on each picture of Figure 44) shows a very weak silencing effect. There could be two explanations for this phenomenon. Either this specific gRNA is not strong enough to ensure proper interaction between the gRNA/dCas9 complex with the DNA, which is probable considering the low prediction score (0.32, Table 30), or the transcription starts efficiently enough after the promoter to ensure a basal transcription level that can support growth.

Lastly, the plasmids harboring gRNAs 2 to 4 targeting locus inside the *fdsG* gene show a strong silencing effect in this context since very low growth is observed in the test medium condition with these guides. Therefore, this CRISPRi experiment shows the essentiality of the expression of the CnFDH operon for the phenotype observed when the strain G6435 is cultivated with acetate and pyruvate as carbon sources and formate as the unique energy source.

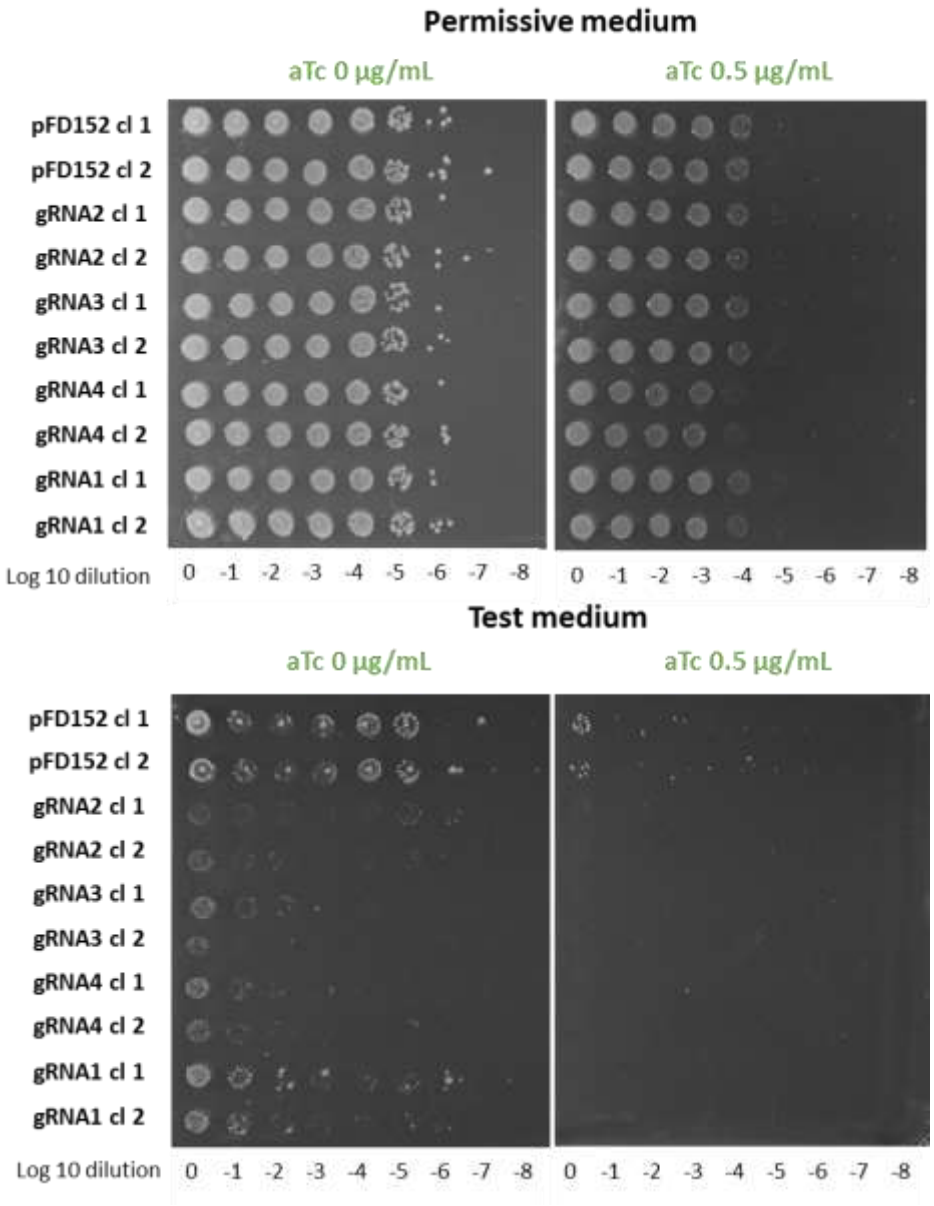


Figure 44. CRISPRi silencing experiment on G6435 strain.

2. gRNA targeting fdsC and fdsD genes from CnFDH operon

Another CRISPRi experiment was performed with gRNA targeting specifically *fdsC* and *fdsD*, the last two genes of the operon that are encoding chaperones, to determine their essentiality for the observed phenotype. Two guides were designed for each gene (Table 32 and Figure 45), named gRNA from 5 to 8, and cloned in plasmid pFD152, resulting in the plasmids pGEN1444 to pGEN1447.

Table 32. gRNA designed for chaperone proteins of the CnFDH for operon silencing. gRNAs and prediction scores were obtained from the CRISPR @Pasteur tool.

gRNA	Targeted gene	Sequence	Prediction score
gRNA5	*fdsC*	CACCAGCGATACATGGCCAT	1.18
gRNA6	*fdsC*	GGCATAGACCACGTGCGCGC	1.03
gRNA7	*fdsD*	GAGACAGCCTCTTCCCGATC	0.87
gRNA8	*fdsD*	CAACAGCCCGTTCCCTGCTT	0.9

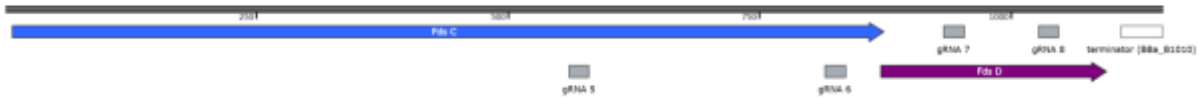


Figure 45. Representation of gRNA targeting *fdsC* and *fdsD* genes from CnFDH operon.

These constructions were transformed in the metabolic screening strain, G6435 and cell cultures were started and incubated on plates following the same procedure as described above (see pictures of the plates in Appendix 8: CRISPRi experiment with guides targeting *fdsC* and *fdsD* genes from CnFDH). However, no noticeable growth effect of these guides was observed. The prediction scores for these guides, while high, were lower than for the efficient guides (gRNA 2 to 4) from the precedent experiment, possibly their efficiency is not high enough for successful silencing. Alternatively, the chaperones might not be essential for the correct folding of CnFDH in *E. coli.*

FdsC and FdsD were described as essential for the maturation and the insertion of the bis-MGD cofactor in the CnFDH complex (Walker et al. 2019). In addition, a study on the FDH from *Rhodobacter capsulatus*, which is highly similar to CnFDH as discussed in the introduction (I. 3), investigated the essentiality of these two subunits, as well. It was demonstrated that heterologous expression of *fdsG, fdsB, fdsA* and *fdsD* (lacking *fdsC*) from RcFDH in a WT *E. coli* strain, gave a functional protein, while if the same complex was expressed in an *E. coli* Δ*fdhD* strain, the complex was inactive. This demonstrated that FdhD from *E. coli* can substitute for the activity of the protein FdsC from *R. capsulatus* (Hartmann and Leimkühler 2013). FdsC from CnFDH shares high similarity with FdhD from *E. coli*, with 43 % identity and 60 % similarity (Oh and Bowien 1998). This could explain the absence of effect observed when silencing *fdsC* gene from the CnFDH operon, which should be further confirmed by the construction a G6435 *fdhD* strain.

However, no homology was found for *fdsD* and this protein is described to be essential for the complex activity which seems inconsistent with our observations. The *fdsD* gene is located downstream of *fdsC*

and its translation should be blocked by the gRNA/dCas9 complex when *fdsC* is silenced, which is not observed in this experiment. Possibly the translation resumes behind the gRNA/dCas9 complex or the silencing is inefficient. To determine the essentiality of these genes in this genetic context, deletion of each gene, both individually and in combination, should be performed to assess the origin of the maintenance of growth on formate.

Moreover, using the plasmid constructions containing *fdsA*, *fdsB* and *fdsG* (pGEN1338) and the chaperones *fdsC* and *fdsD* (pGEN1348), another test was performed. For this, either both plasmids were transformed in the G5416 (Δ*/pd*) selection strain or only plasmid pGEN1338, the resulting strains were then tested for growth in the same test medium. Growth was only obtained when both plasmids were co-expressed in the cell. This additional result is another demonstration of the essentiality of the accessory proteins for efficient CnFDH activity. This also suggests that the observation made with the CRISPRi system is probably due to the lack of efficient silencing.

3. Discussion and perspectives

We were able to demonstrate the activity of the complex metal-dependent formate dehydrogenase from *Cupriavidus necator* towards formate oxidation, *in vivo*, in *E. coli,* in the presence of O_2. The genetic selection strain (Δ*/pd*) presented difficulties to grow in the metabolic test conditions when formate was the source of NADH through the activity of CnFDH expressed on a plasmid, especially characterized by a long lag phase. Directed evolution in GM3 allowed the selection of mutant cells with a lower generation time and drastically reduced lag phase. The isolates from GM3 evolution presented a generation time 3-fold lower than the parental strain. The genetic background of the evolved strains is more favorable for FDH-catalyzed formate oxidation, as demonstrated for CnFDH, but also for TsFDH, and thus is more convenient for the testing of various NAD^+-dependent FDHs, making it versatile. Using this evolved background and overcoming several difficulties, we were able to achieve successful integration of the CnFDH operon in the chromosome of *E. coli* and to obtain efficient expression to restore growth in test conditions. An additional evolution in GM3 proved the stability of the CnFDH insertion and the stable FDH expression throughout more than 400 generations. These results proved the feasibility of using such complexes within *E. coli* when the metabolic test is properly sealed.

Mutations in gene *focA* were identified as being the main cause of the gain in growth rate observed throughout evolution. However, the evolution curves obtained for the UOF cultures with CnFDH expressed on a plasmid show a progressive evolution. Thus, the sequencing of isolates at different time points of this evolution could allow determining the moment of emergence of these mutations in the evolution process and could help to understand the potential role of the other mutations detected.

Although the activity of CnFDH was high enough to support growth in this test, the growth rate obtained with the strains expressing it compared to TsFDH is not reflecting *in vitro* data. The kinetic parameters for these FDHs were described in the introduction (Table 2) and were demonstrated to be less favorable for TsFDH (both K_m and k_{cat}) compared to CnFDH. From the literature, the pH optimum for both FDHs was determined for formate oxidation. For CnFDH, it was determined to be around 7.5 with a rapid decrease in the activity as the pH changes (Niks et al. 2016). For TsFDH, the pH effect in the range of 5.5 to 7.0 was tested and no significant effect of this parameter was observed (Choe et al. 2014). The pH of the test medium used for this study is around 7.2, hence optimal for formate oxidation activities of both enzymes. These results demonstrate the difficulty of the heterologous expression of such complex proteins in *E. coli* despite their theoretical potential for their use in biotechnology.

A more extensive work on the expression format, *i.e.* plasmid or chromosomic, native or optimized genes, operon re-arrangement, would be interesting to pursue in order to test if a better expression could be obtained from the CnFDH to optimize growth.

From these experiments, a publication was deposited in bioRxiv, the article is available in the Appendix 9: Publication submitted in BioRxiv.

While this demonstration of the activity of the CnFDH in *E. coli* towards formate oxidation was a promising result, the goal was to investigate the reversibility of this reaction in the *E. coli* context by studying the CO_2 reduction mediated by the CnFDH, which was only demonstrated *in vitro* to date.

2. Formate dehydrogenases for *in vivo* CO_2 reduction

This section describes the work conducted on CO_2 reduction using the FDH from *C. necator*. Two strains previously developed by the laboratory were used as a base for these experiments. Both strains depend on formate for growth and are using the reductive glycine pathway (RGP) for formate assimilation but differ in the amount of formate incorporated into the total carbon for biomass. The metabolism of these two strains is outlined in the following paragraph. Then, the results of the experiments carried out are described and discussed.

1. Metabolism of the selection strains

The two strains used are: G4248, which will be referred to as the serine auxotroph strain, and G4463, referred to as the glycine auxotroph strain. Their detailed genotype is shown in Table 33 below.

Table 33. Genotype of formate-dependent strains G4248 and G4463. The composition of (A) permissive and (B) non-permissive (selective) growth medium is indicated for each strain.

Strain	Genotype	Growth medium
G4248	Δ*gcv* Δ*(tdh-kbl)* Δ*ltaE* Δ*gcvR* Δ*kdgK*::*ftfL (C. kluyveri)* Δ*serA*	(A) MS glucose serine (B) MS glucose glycine formate
G4463	Δ*glyA*::aad Δ*(tdh-kbl)* Δ*ltaE* Δ*gcvR* Δ*kdgK*::*ftfL* (*C. kluyveri*) Δ*aceBAK*	(A) MS glucose glycine (B) MS glucose formate CO_2

To implement such carbon flux rewiring and dependency on formate assimilation, genetic engineering, heterologous gene insertions and deletions were performed. The first key modification is the insertion of the *ftfL* gene from *Clostridium kluyveri,* encoding a formate THF-ligase. It was introduced at the *kdgK* locus (non-essential gene) in *E. coli* chromosome, it catalyzes the first reaction of formate assimilation, the activation of formate through ligation to THF (Figure 7). Secondly, specific gene deletions were performed to guide the carbon fluxes enabling metabolic selection. The two deletions of the gene *ltaE* and the operon *tdh-kbl* were introduced to prevent the synthesis of glycine from threonine. *gcvR* is a transcriptional regulator that is described as negatively regulate the glycine cleavage system (GCS) (Ghrist, Heil, and Stauffer 2001) which could impede the efficient activity of this complex.

Both strains G4248 and G4463 are derived from the glycine and C1 auxotroph isolate G3806 (Δ*glyA* Δ*gcvTHP* Δ(*tdh-kbl*) Δ*ltaE* Δ*kdgK::ftfL* (*C. kluyveri*) obtained by GM3 evolution. For this strain, formate is necessary for the synthesis of compounds incorporating THF-carried C1 moieties in their structure (C1-compounds, Döring et al. (2018)). The formate-dependent phenotype was established by directed

evolution in GM3, and mutations affecting the *folD* gene (R191S) and heterologous *ftfL* region (Δ 3 bp upstream the gene insertion) were detected (Döring et al. (2018); Delmas et al. (2022)). For the construction of the serine auxotrophic strain G4248, the WT allele of *glyA* was restored and *serA* was deleted. The glycine auxotrophic strain G4463 was obtained from a G3806 descendent restored in the *gcv*THP genes and deleted in *aceBAK* genes (strain G4369, Döring et al. (2018)), which was subjected to another continuous culture experiment where additional mutations were selected, notably a mutation in the *lpd* gene (R273H), which encodes the lipoamide dehydrogenase involved in the GCS (Delmas et al. 2022).

Concerning strain G4248, the absence of genes *gcvTHP* coding for the glycine cleavage system prevents glycine cleavage to methylene-THF, NH_3 and CO_2. The *serA* deletion prevents serine synthesis from 3-phosphoglycerate. Hence, this strain is serine and C1 auxotroph as illustrated by Figure 46 and strictly depends on the activation of formate into formyl-THF and subsequent reduction into methylene-THF. In this genetic context, formate accounts for 3.8 % of total carbon biomass.

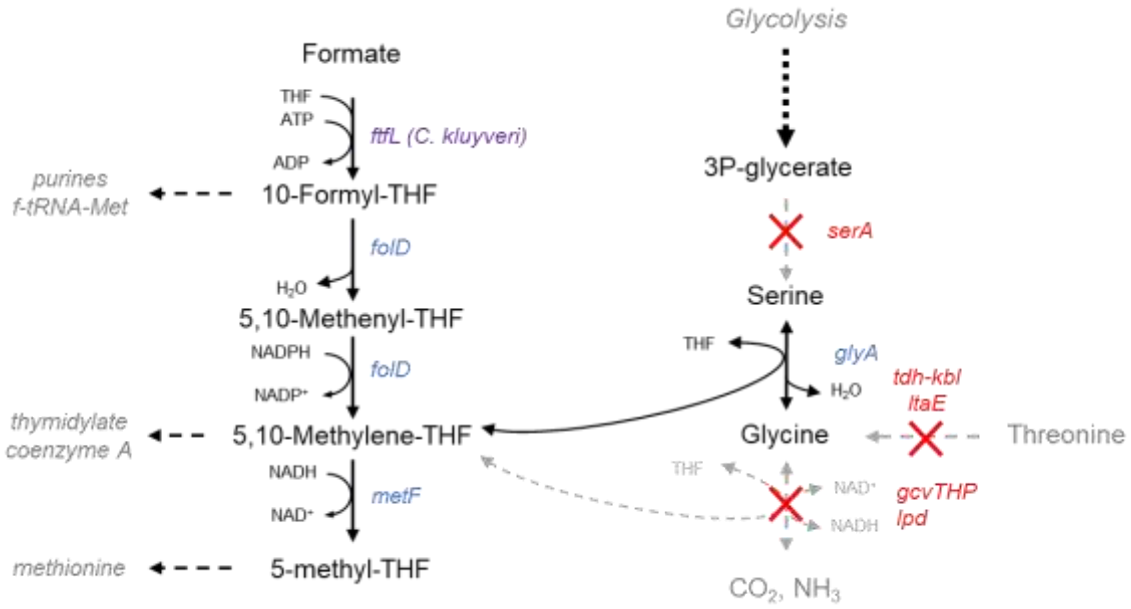


Figure 46. Serine, glycine and C1 metabolism of the serine auxotrophic strain G4248. Adapted from Delmas et al. (2022).

As for the G4463 metabolic screening strain, the absence of gene *glyA* impedes the production of glycine from serine and the deletion of *aceBAK* blocks the glyoxylate shunt preventing a possible glycine synthesis from glyoxylate by a mutated transaminase. Hence, this strain is a glycine auxotroph, its

metabolism is represented in Figure 47. This strain presents a higher selection stringency compared to G4248 with formate accounting for 7.8 % of total carbon biomass (Döring et al. 2018).

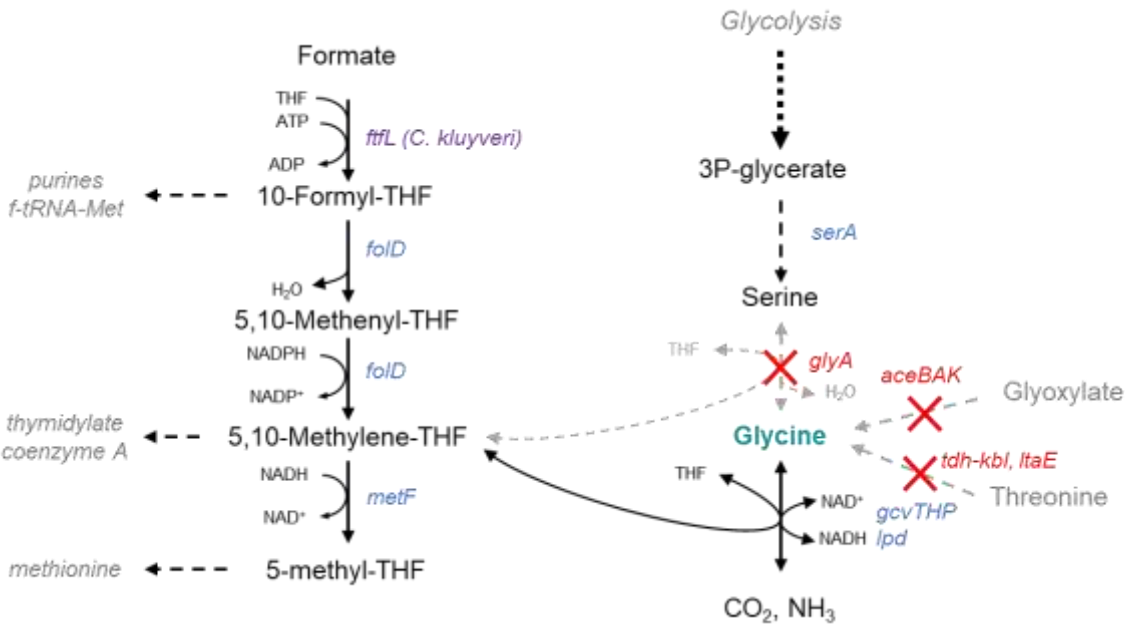


Figure 47. Serine, glycine and C1 metabolism from the glycine auxotrophic strain G4463. Adapted from Delmas et al. (2022).

For the two strains presented, formate constitutes a fraction of the total carbon assimilated for biomass production. Our hypothesis was that these genetic contexts enable the selection of CO_2 reduction activity of CnFDH effectively transforming CO_2 into formate, which is then integrated into the biomass. To favor this non-favorable reaction of CnFDH, it is necessary to use elevated CO_2 concentrations in the aeration gas of the culture. Additionally, maintaining an elevated pool of NADH/NAD^+ is crucial to support NADH-dependent CO_2 reduction.

2. Experimental setups

To assess the activity of CnFDH for CO_2 reduction in the above-mentioned strains, diverse culture methods were employed and the reasons for these choices are detailed below. Specific experiment descriptions are available in the Material and methods chapter (II. 3. 2. 2).

Initially, plate assays were performed but were not conclusive because of sporadic growth observed during incubation, even on negative controls. These inconsistencies were attributed to the incubation time of the plates and the possible presence of trace compounds in the agar that could lead to "pop

out" colonies. Consequently, subsequent tests were exclusively conducted in liquid culture conditions in tubes under a CO_2-enriched atmosphere and agitation using an Infors Minitron device. Humidity inside the incubator was controlled *via* a specific device, which allowed culturing for extended time periods. Additionally, another experimental setup utilized 100 mL hermetically sealed flasks containing 10 mL of growth medium. The inoculated flasks were flushed with a mix of 50 % CO_2 in air unless stated differently. This method exposed the cells to higher CO_2 concentrations to promote the CO_2-reduction activity of the CnFDH. The flasks were then placed under agitation at 30 °C.

In addition to these tests in tubes and hermetic flasks, GM3 evolution experiments were conducted with cultures aerated with 20 % CO_2-enriched air.

3. Preliminary tests

As a first attempt, the previously validated pTRC-CnFDH plasmid construction (pGEN1340) for formate oxidation was transformed in the two described genetic contexts. This resulted in the strains G5785 (G4248 pTRC-CnFDH) and G5743 (G4463 pTRC-CnFDH).

To perform preliminary tests, the strains were adapted on plates on a minimal permissive medium described in Table 33 above. An overnight liquid culture in this medium was prepared to inoculate test cultures, *i.e.* MS mineral medium supplemented with glucose and glycine for the serine auxotrophic strain (G4248) and glucose only for the glycine auxotrophic strain (G4463), along with IPTG (100 µM) for CnFDH induction. For tests with the latter strain, ammonia concentration was augmented to 80 mM to favor glycine synthesis by the GCS reaction (Döring et al. 2018). The test tubes were incubated in the Minitron with 20 % CO_2-enriched air under agitation at 30 °C. As such, no growth was detected after 2 weeks of cultivation repeated three times.

Considering these preliminary tests we tried to lower the formate consumption by interrupting unnecessary pathways to maintain the formate pool that was potentially produced by the CnFDH thus keeping it available for assimilation.

4. Enhancement of growth test sensitivity by strain engineering

1. *fdoG* deletion

As mentioned in the introduction (I. 3), FDH-O is the only FDH from *E. coli* reported to be active under aerobic conditions out of the three that the strain possesses. It is membrane-bound and composed of three subunits encoded by the operon *fdoGHI*. This enzyme is non-essential in the genetic context and oxidizes formate to CO_2, which could increase the formate needs of the test strains. To inactivate this

complex, the α-subunit, which is the largest one of this complex (113 kDa) encoded by *fdoG* gene (Benoit, Abaibou, and Mandrand-Berthelot 1998), was chosen as the target for gene deletion to lower the formate requirements.

In the two metabolic screening strains, *fdoG* was deleted via P1 transduction. P1 phages were prepared using the Keio strain JW3865 carrying the deletion Δ*fdoG*::*kan* and used to transduce G4248 and G4463 strains. Then, the resistance cassette was excised by the FLP-recombinase system and the resulting strains were named G6440 (G4248 Δ*fdoG*) and G6446 (G4463 Δ*fdoG).*

2. *nuoG* deletion

During the course of this PhD project, an interesting study demonstrated the *in vivo* CO_2 to formate reduction by a homologous FDH in a formate-dependent *Pseudomonas putida* strain (Bruinsma et al. 2023). In this work, the authors demonstrated the complementation of a serine auxotrophy in an engineered *P. putida* strain *via* the overexpression of the native PpFDH for CO_2 reduction to formate and three heterologous genes *(fhs*, *fchA* and *folD*) from *C. ljungdhalii* to support C1 biosynthesis catalyzing the reactions from formate to 5,10-methylene-THF. Their work combined rational design along with directed evolution. Genome sequencing performed after short-term evolution steps revealed a mutation in gene *nuoG*, introducing a stop codon in the coding sequence. *nuoG* encodes the subunit G of the NADH:quinone oxidoreductase (complex I) from the respiratory chain that is regenerating NAD^+ within the cells. This interruption in the sequence most likely impacts the activity of the complex, hence increasing the NADH pool and favoring FDH's CO_2 reduction activity. Consequently, it was decided to implement this deletion in *E. coli* to study its effect on the metabolic selection strains. In *E. coli*, *nuoG* is essential for complex I activity but not for cell survival and plays a role for the stability and the assembly of the complex (Pohl, Bauer, et al. 2007).

In the two metabolic selection strains, *nuoG* was deleted via P1 transduction. P1 phages were prepared using the Keio strain JW2278 (Δ*nuoG*::*kan*) and used to transduce G4248 and G4463 strains. Then, the resistance cassette was excised using the FLP-recombinase system and the resulting strains are G6385 (G4248 Δ*nuoG*) and G6384 (G4463 Δ*nuoG)*.

Finally, a combination of the two above-mentioned deletions was performed using the same method, creating the strain G6441 (G4248 Δ*fdoG* Δ*nuoG*) and G6447 (G4463 Δ*fdoG* Δ*nuoG*).

3. Test of formate requirements

To quantify the effect of those deletions, growth rate and yield of the engineered strains described in the previous paragraphs were evaluated in a range of formate concentrations. Strains G4248 and G4463

and their derivatives were pre-cultured overnight in permissive medium with excess formate to ensure efficient growth. Then, the cells were washed twice with MS mineral medium to discard formate and prevent carry-over. Finally, the strains were grown with different concentrations of formate at 30 °C in tubes with 3 mL of liquid medium.

1. G4248 (serine and C1 auxotroph) strain

The OD_{600} values measured at 24 and 48h for strain G4248 and its derivatives grown in a formate concentration range from 0 to 10 mM are represented in Figure 48 below.

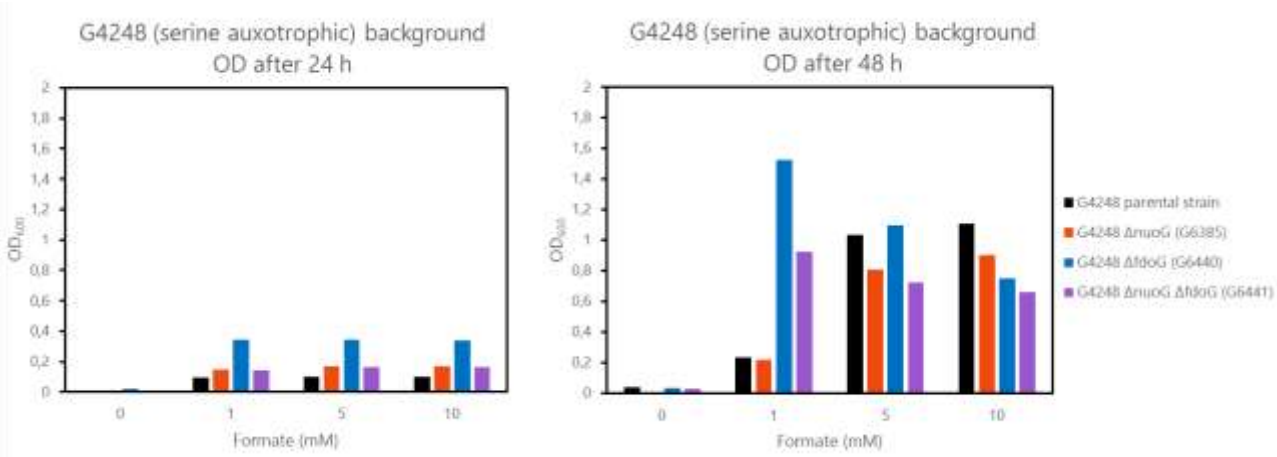


Figure 48. G4248 and derivative strains grown in a formate concentration range under selective conditions. Strains were grown in MS supplemented with glucose (0.2 %), glycine (2 mM) and formate at 0, 1 mM, 5 mM and 10 mM. Cultures were incubated at 30 °C under a 20 % CO_2-enriched atmosphere.

Figure 48 shows that after 24 h of incubation, strain G4248 and its Δ*nuoG*, Δ*fdoG* and Δ*nuoG* Δ*fdoG* derivatives grew to similar cell density within the tested formate range (1 to 10 mM). Obviously, during early growth phase formate at 1 mM is not growth limiting, demonstrating the low requirement for formate of the C1/serine selection. Strain G6440 (represented in blue in the figure), only deleted in *fdoG,* exhibited higher growth at 24 h compared to the other strains, the formate entering the cells is probably better directed to support growth rather than being oxidize by the endogenous FDH-O. This shows that formate dehydrogenase O is indeed a formate sink in the cells. After 48 h of incubation, G6440 and G6441 strains (represented in blue and purple in Figure 48), both lacking *fdoG*, showed more efficient growth at 1 mM formate in the test medium, demonstrating a strong advantage of these genetic constructions at low formate concentration, as compared to the parental strain. However, for 5 and 10 mM formate, the gene deletion do not entail a significant advantage and represented rather a burden

at 10 mM formate. This weaker growth under 10 mM of formate could indicate toxicity that could be due to an accumulation of formate within the cells.

The lower growth observed for the Δ*nuoG* strains was expected as it was reported that *nuo* gene inactivation impacted growth rate due to a non-functional NADH:quinone oxidoreductase (complex I). This is probably due to a NAD^+/NADH ratio disturbance created in the cells (Erhardt et al. 2012).

Given the observed growth advantage at low formate, it appears that strains G6440 (Δ*fdoG*) and G6441 (Δ*fdoG* Δ*nuoG*) are better suited as selection strain than the parental genotype for the detection of presumably weak reductive CnFDH activity.

2. *G4463 (glycine auxotroph) strain*

The growth obtained for G4463 strain and its derivatives for a formate concentration range from 0 to 10 mM are represented in Figure 49 below.

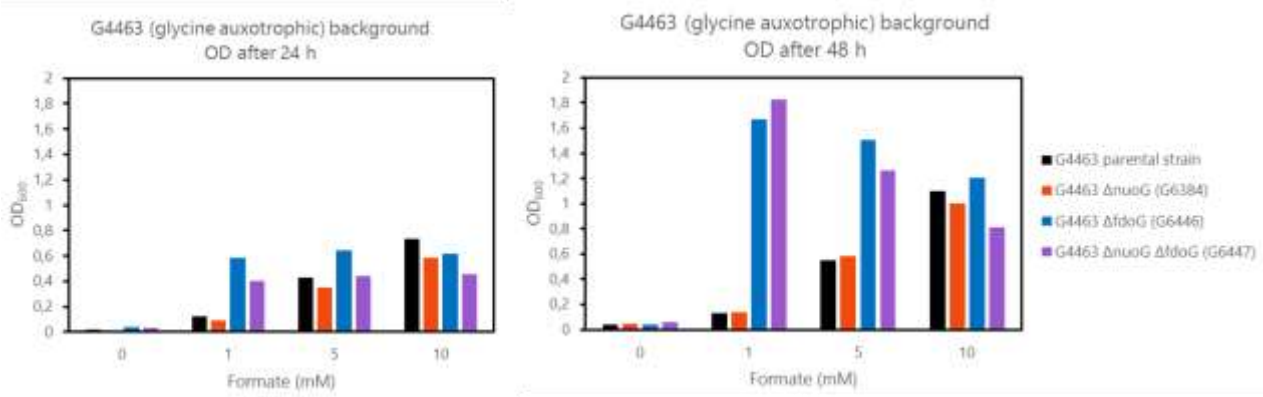


Figure 49. G4463 and derivative strains grown in a formate concentration range under selective conditions. Strains were grown in MS supplemented with glucose (0.2 %), and formate at 0, 1, 5 and 10 mM. Cultures were incubated at 30 °C under a 20 % CO_2-enriched atmosphere. NH_4Cl (80 mM) is added to the medium to promote the activity of the GCS complex (*gcvTHP lpd*) as NH_3 is a substrate for its glycine production activity.

At low formate concentrations of 1 mM and 5 mM in the growth medium, strains lacking *fdoG*, namely G6446 and G6447 (in blue and purple in Figure 49), exhibited significantly higher growth yields, especially after 48 h of incubation, in contrast to the two other strains from the test. At 1 mM of formate, G4463 and G6384 showed nearly no growth, while the two other strains reached the highest OD_{600} values measured. Additionally, as observed in the previous experiment with the serine auxotroph strain, formate toxicity is observed for the strains lacking *fdoG*.

For this metabolic selection, the deletion of *nuoG* did not impair growth unlike the serine auxotroph strain. At low formate concentrations, after 48 h of incubation, the Δ*fdoG* Δ*nuoG* double mutant showed similar growth to that of the Δ*fdoG* strain. This might be due to the presumably higher NADH/NAD^+ ratio that would favor the activity of the GCS complex acting in the glycine synthesis sense where NADH is a substrate.

In summary, the deletion of *fdoG* in this genetic context is dramatically lowering the strain's formate requirements and the additional *nuoG* does not impair growth yields.

3. *CnFDH expression in these genetic contexts*

1) Plasmid expression

The plasmid pTRC-CnFDH (pGEN1340) described in the *Expression of a functional complex formate* dehydrogenase in *E. coli* section (III. 1. 1. 1. 1) was used again for CnFDH expression in the formate-dependent selection strains. The plasmid was transformed in the engineered strains described in the previous section and the resulting strains are detailed in Table 34.

Table 34. Formate-dependent strains constructed for the test of the CnFDH CO_2 reduction activity expressed from a plasmid.

Strain	Genotype
G6493	G4248 (serine auxotroph) Δ*fdoG*; pTRC-CnFDH (pGEN1340)
G6495	G4248 (serine auxotroph) Δ*fdoG* Δ*nuoG*; pTRC-CnFDH (pGEN1340)
G6497	G4463 (glycine auxotroph) Δ*fdoG*; pTRC-CnFDH (pGEN1340)
G6499	G4463 (glycine auxotroph) Δ*fdoG* Δ*nuoG*; pTRC-CnFDH (pGEN1340)

The strains were pre-cultured overnight in the permissive minimal medium described in Table 33, then the cells were washed twice with MS medium and used to inoculate 100 mL hermetic flasks containing 10 mL of test medium at an initial OD_{600} of 0.05. For the strains with G4248 background (serine auxotroph), the test medium consisted of MS minimal medium supplemented with glucose (0.2 %), glycine (2 mM), IPTG (100 µM) and 50 % CO_2 in the headspace. In this context, if the CnFDH is efficiently active for CO_2 reduction, formate produced by the heterologous enzyme should be incorporated in the biomass and become the source of serine and C1. As for the G4463 background (glycine auxotroph), the test conditions were MS minimal medium supplemented with glucose (0.2 %), IPTG (100 µM) and 50 % CO_2 in the headspace where formate produced by the CnFDH should complement the glycine auxotrophy. The flasks were incubated at 30°C under agitation.

Unfortunately, after 8 weeks of culture, none of the test cultures showed any growth. Interestingly, the positive controls performed in parallel in permissive conditions for each test strain in both regular atmosphere and CO_2-enriched atmosphere showed a longer lag phase of one day for the strains cultured with CO_2. *E. coli* culture in this type of atmospheric conditions is not described but it can be hypothesized that the lower O_2 content that this CO_2-enrichment creates is disturbing the global metabolism of the cells.

Considering this absence of positive results, it was decided to integrate the operon in the chromosome for higher stability of the heterologous FDH that is better suited for long-term continuous culture evolution in GM3 devices.

2) Genome integration and preliminary tests

The genomic integration of CnFDH was performed following the same procedure as described in the *Expression of a functional complex formate* dehydrogenase in *E. coli* section (III. 1. 2) using the same P1 phages preparation that was validated from the strain G6408 (MG1655 IS10::CnFDH kan+). The formatotrophic strains were transduced with these phages which resulted in the strains described in Table 35. For the serine auxotroph, the role of the *nuoG* deletion was unclear, thus the following tests in tubes were performed with the G4248 derivative only harboring the *fdoG* deletion. However, the double mutant with the IS10::CnFDH integration was constructed for GM3 evolution. For the glycine-auxotroph background, it was decided to focus on the strain with both *fdoG* and *nuoG* deletions as it showed more favorable growth characteristics under low formate concentrations.

Table 35. Formate-dependent strains constructed for the test of the CnFDH CO_2 reduction activity expressed from the chromosome.

Strain	Genotype
G6475	G4248 (serine auxotroph) Δ*fdoG* IS10::CnFDH kan+
G6488	G4248 (serine auxotroph) Δ*fdoG* Δ*nuoG* IS10::CnFDH kan+
G6512	G4463 (glycine auxotroph) Δ*fdoG* Δ*nuoG* IS10::CnFDH kan+

The strains described in Table 35 above were initially tested in test tubes in the Minitron with 20 % CO_2-enriched air and the corresponding growth medium for each strain but no growth without formate was detected, which is consistent with the precedent experiment. We additionally performed the same tests in flask under 50 % CO_2 but no growth was observed either. However, a formate range with concentrations below 1 mM was tested to see if using a low amount of formate could highlight a reductive activity from the CnFDH, albeit very low. In addition, this experiment would give a more precise idea of the formate concentration necessary for growth in these genetic contexts.

As depicted in Figure 50, for G4248 and derivatives strains, a clear advantage is seen for Δ*fdoG* strains using these low formate concentrations, which is consistent with the previous experiment (Figure 48). However, the advantage of the presence of the CnFDH is only slightly visible with the OD_{600} measurements performed after 24 h. After 48 h of incubation, the strain G6440 (represented in red), without CnFDH, showed better growth. The formate needs of these strains are very low, depending on the final OD_{600} that is targeted but efficient growth is reached with as low as 0.25 mM of formate in the medium.

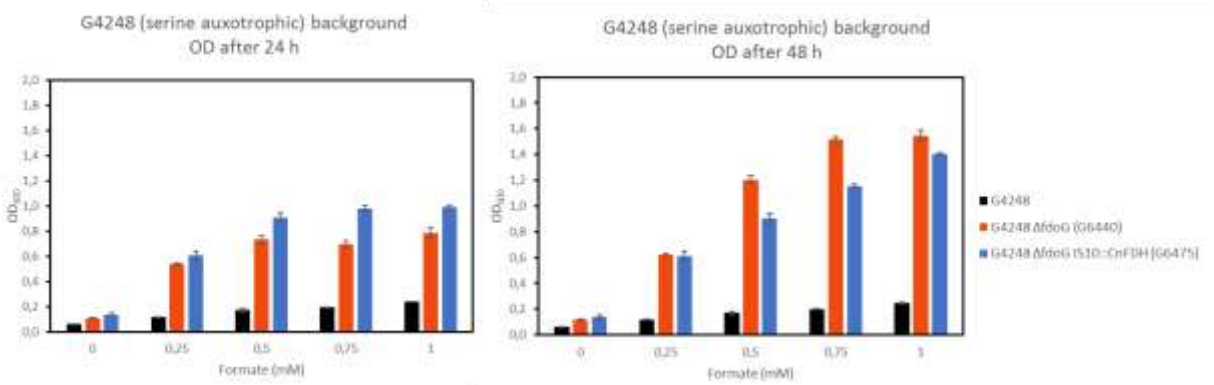


Figure 50. G4248 derivatives strain cultures with formate concentration below 1 mM. The strains were grown in MS mineral medium supplemented with glucose (0.2 %), glycine (2 mM) and formate in the range of 0 to 1 mM. The cultures were performed at 30 °C in a 20 % CO_2-enriched atmosphere in Minitron device. Each condition was performed in triplicates, error bars are representing the variation observed between this triplicates.

To assess the growth capacities of the glycine auxotroph selection strain (G4463) derivatives lacking genes *fdoG* and *nuoG*, the same formate concentration range between 0 to 1 mM was monitored (Figure 51). For this strain, after 24 h of growth, no difference in OD_{600} measurement is detected. However, the heterologous CnFDH seems to represent a burden for the cells that is noticeable after 48 h of growth, especially with formate concentrations of 0.5 and 1 mM. In conclusion, from this experiment, no positive effect of the CnFDH on growth can be detected.

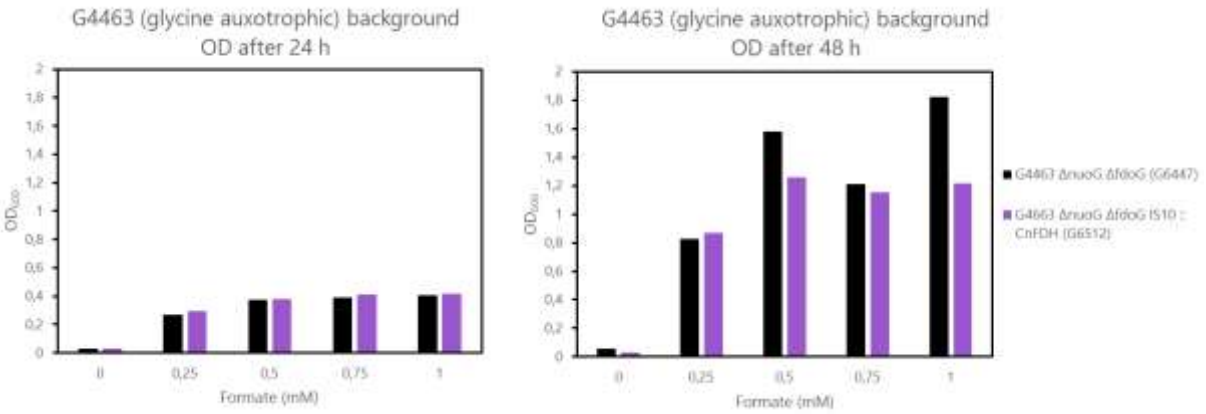


Figure 51. G4463 derivative strain cultures with formate concentration below 1 mM. The strains were grown in MS mineral medium supplemented with glucose (0.2 %) and formate in the range of 0 to 1 mM. The cultures were performed at 30 °C in a 20 % CO_2-enriched atmosphere in Minitron device.

Following these experiments that were rather not conclusive on the CnFDH activity for formate production, a GM3 evolution was started.

3) GM3 evolution

For the GM3 experiment, the three strains described in Table 35 above were cultured in the permissive medium (1 mM formate for every strain which proved sufficient for efficient growth) and were used to inoculate GM3 chambers with a medium swap regime under the conditions described in the table below (Table 36).

Initially, the GM3 culture chambers are inoculated and the cultures are stabilized under turbidostat in the permissive medium. This ensures a proper fixation of the strains in the device and is used to determine the generation time to set for dilution rate under a medium swap regime. Once the cultures are stabilized, the medium swap experiment is started.

Table 36. GM3 culture conditions of the serine and glycine auxotroph strains (FRS cultures). For FRS 1 and 2, only one evolution was performed for each genotype (G4248 Δ*fdoG* ±Δ*nuoG*). For FRS 3 and 4, the same strain is under two distinct evolutions in parallel.

Culture name	Strain	Permissive medium	Stressing medium
FRS 1	G6475	Glucose (0.2 %), glycine (2mM), formate (1 mM), CO_2 (20 %)	Glucose (0.2 %), glycine (2mM), CO_2 (20 %)
FRS 2	G6488	Glucose (0.2 %), glycine (2mM), formate (1 mM), CO_2 (20 %)	Glucose (0.2 %), glycine (2mM), CO_2 (20 %)
FRS 3&4	G6512	Glucose (0.2 %), formate (1mM), CO_2 (20 %)	Glucose (0.2 %), CO_2 (20 %)

The evolution graphs are shown in Figure 52 below. While the ratio of stressing to permissive dilution pulses did augment for the four cultures, no complete adaptation to the stressing medium was obtained after 100 days for FRS 1 and 2 cultures and 50 days for FRS 3 and 4 cultures. The differences in generation time were due to technical difficulties (loss of temperature control) encountered for cultures FRS 2 and FRS 3. The number of GM3 devices enabling culturing with CO_2/air mixes being limited at 2, no transfers to other devices were possible. To date, these evolution experiments are still running, the stressing/permissive ratio staying stable for the moment.

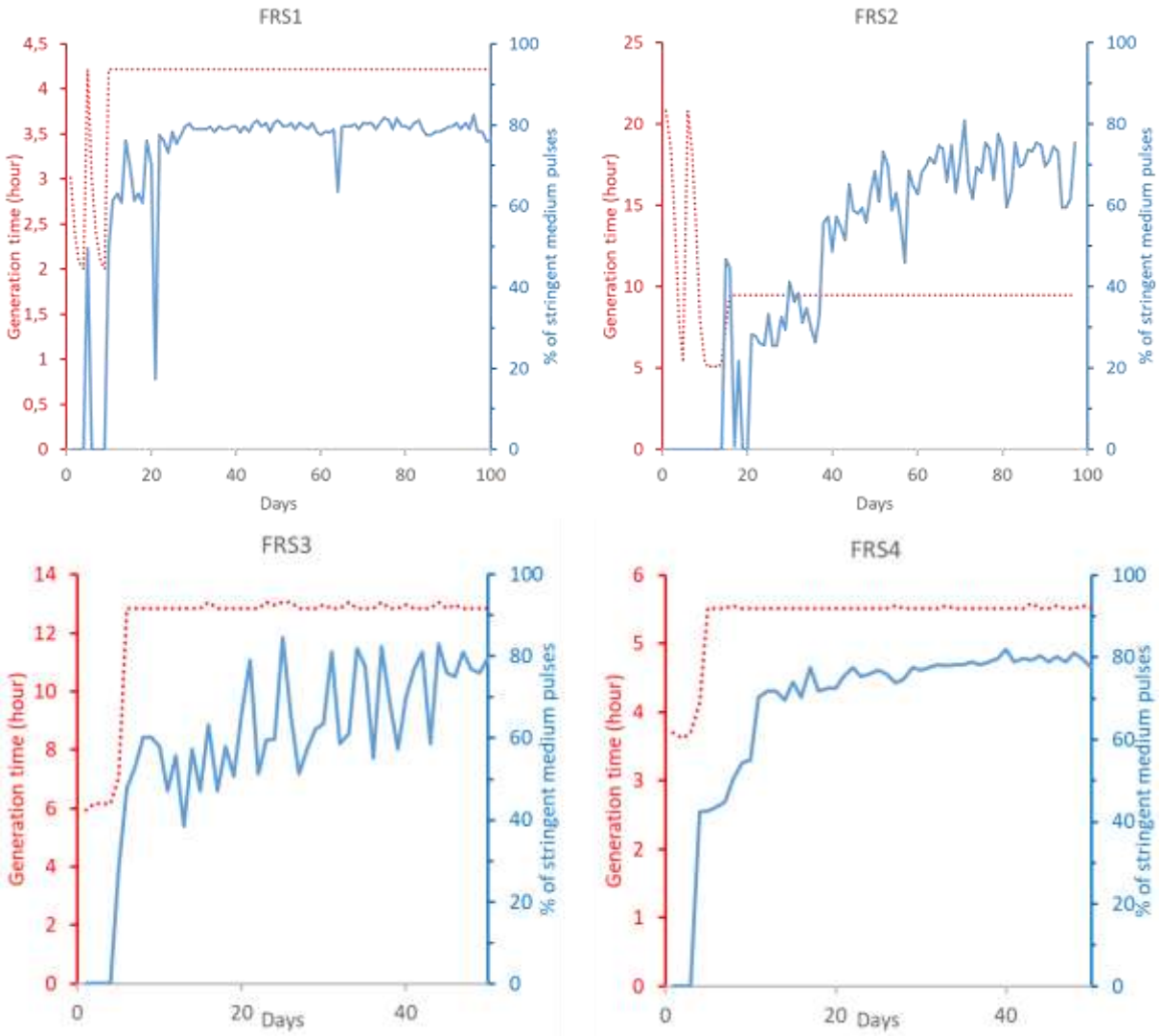


Figure 52. GM3 evolution of the FRS cultures. FRS1 and 2 correspond to the evolution of strains G6475 and G6488 respectively. FRS 3 and 4 are two evolutions conducted in parallel for the strain G6512. The red line represents the evolution of the generation time while the blue line represents the progression of stressing medium dilution pulses relative to permissive medium dilution pulses.

5. Discussion and perspectives

In this chapter, formate-dependent selection strains to test and evolve FDH reductive activity are described. The formate requirements, in other words the stringency of the selections, were scalable; the serine auxotrophic selection requiring 3 %, the glycine auxotrophic selection around 8 % of biomass carbon to origin from formate assimilation. Deletion of gene *fdoG* coding for formate dehydrogenase O reduced the formate need of the strains.

However, despite the efforts deployed for the demonstration of the CnFDH activity for *in vivo* CO_2 reduction, we were not able to achieve this objective. One of the difficulties is related to the thermodynamics of the reaction, considering the CO_2/formate redox potential of -430 mV and the NAD^+/NADH couple's potential of -320 mV. While these redox potentials are not extremely different, making the reaction thermodynamically possible, it remains unfavorable. To favor this reaction, two key parameters must be explored: increasing the CO_2 concentration, which was explored by pushing the CO_2-content up to 50 %, and enhancing NADH availability within the cells. The NAD^+/NADH ratio is tightly controlled within the cells, with a ratio of 10.6 (2.56 and 0.24 µmol per ml of culture for NAD^+ and NADH respectively) in *E. coli* grown on glucose under aerobic conditions (Leonardo, Dailly, and Clark 1996). The deletion of gene *nuoG* coding for an essential subunit of NADH dehydrogenase I described to affect the NADH oxidation by the respiratory chain favored formate reduction in some circumstances but was not sufficient to establish the desired growth phenotype. A possibility to augment the NADH-concentration in the cells would be to express the NAD^+ transporter NTT4, which is described to transport both NAD^+ and NADH in counter-exchange with ATP/ADP (Haferkamp et al. (2004); Zhou et al. (2011)), and supplement the growth medium with NADH. However, NADH being unstable and expensive, this strategy would be suitable for small-scale tests rather than for larger evolution setups.

We were able to cultivate selection strains for 100 days (serine and C1) and 50 days (glycine selection) in a medium swap continuous culture regime without the appearance of cells producing formate in an FDH-independent way. Such cells would have rapidly overgrown the cultures. A candidate gene introducing leakiness in the selection is *pflB* coding for pyruvate formate-lyase. Described for being inactivated by O_2 due to a glycyl radical (G737) involved in catalysis, it was found that a glycyl radical factor (GRC), *grcA* gene, was potentially able to rescue certain G737 mutants, which in turn would become candidates for activity in an O_2 environment (Knappe et al. (1974); Wagner et al. (1992); Kammel and Sawers (2022)). This is additionally favored by the large fraction of CO_2 in the gas mixture aerating the cultures, necessarily decreasing the O_2 proportion of the atmosphere. Besides PflB, gene *tdcE* coding for an enzyme with 82 % identity with PflB has been reported as having pyruvate formate-lyase and 2-ketobutyrate formate-lyase activity (Heßlinger, Fairhurst, and Sawers 1998). The deletion of these genes in the selection strains to enable long-term GM3 evolution is currently undertaken.

Even though the activity of the CnFDH for formate production is not sufficient to support growth of the test strains, this does not mean the enzyme is completely inactive. To assess this question, experiments using ^{13}C-CO_2-enriched air with cultures in limiting formate concentrations would give valuable information on the eventual CnFDH activity. The analysis of the ^{13}C content in specific metabolites using LC-MS could potentially reveal any CO_2-reducing activity from the CnFDH.

A number of measures can be taken to enhance the mutation rate of the cells during evolution. Lowering the available carbon source concentration is an option, or the deletion of genes coding for error corrections thus creating a mutator phenotype (Taddei et al. (1997); Kang et al. (2019)). These mutator strains exhibit up to 1000x higher mutation rates. While these approaches affect the genome in a random way, *in vivo* methods of targeted mutation have been developed enabling the accumulation of mutations throughout a gene with low off-target changes. In the frame of this study, first steps to establish such protocols in the laboratory with a metal-independent FDH as target were undertaken (III. 5. 1).

3. Energy generation systems

As elaborated in the introduction (I. 5), the autotrophic strain that was targeted requires an energy-generation system. Formate has been used as energy source through the formate oxidation activity of FDHs expressed in *E. coli* engineered strains. However, it is evident that in our system, this energy source cannot be chosen. Instead, we focused on two potential, non-carbon energy sources for our system: hydrogen and phosphite.

1. Hydrogenase

The study of hydrogen as electron donor was explored through the expression of an NAD^+-dependent hydrogenase from *C. necator*. This enzyme, described to be oxygen tolerant, is encoded by a 14-genes operon and involves metals, essential for its activity. It catalyzes the reversible cleavage of H_2 to two protons and two electrons with NAD^+ as cofactor. We obtained the plasmid constructions containing the operon from the authors of a previous study and introduced these in our energy-auxotroph *E. coli* strain described above harboring additional genomic modifications. The functionality of the complex in our engineered strain was initially verified for hydrogen production. Subsequently, we performed *in vivo* tests with H_2 as the energy source.

1. Strains and plasmids

To create an expression platform for *C. necator* NAD^+-dependent hydrogenase (CnHyd), the deletion of the four endogenous hydrogenases from *E. coli* was a prerequisite to be able to verify the functionality of the complex by GC analysis in the preliminary tests. As these four hydrogenases are encoded by large operons as described in the Introduction (I. 5. 1), our strategy was to knockout one essential gene of each operon. For hydrogenase 1, the disrupted gene was *hyaB*, for hydrogenase 2, *hybC*, for hydrogenase 3, *hycE* and for hydrogenase 4, *hyfG*. These four deletions were carried out sequentially by P1 transduction using strains from the Keio collection, respectively JW0955, JW2962, JW2691 and JW2472 (Material and methods chapter, II. 4. 2). Between each transduction, the resistance cassette was erased using the FLP-recombinase system.

Following the experimental scheme of Lamont and Sargent (2017), we also deleted gene *iscR* coding for a transcriptional regulator, which was reported to be beneficial for heterologous production of metalloenzymes in *E. coli* and gene *adhE* coding for an acetaldehyde-CoA dehydrogenase, a known NADH sink in O_2-depleted environments, to increase NADH availability, necessary for H_2 production, in the cells when cultured in anaerobic conditions. In addition, gene *lpd* was deleted to construct the

metabolic screening strain for hydrogenase activity tests with H_2 as NADH source in aerobic conditions. All the strains were successfully constructed and summarized in Table 37 below.

Table 37. Strains constructed and used for *in vivo* tests of hydrogen production and consumption by the CnHyd system.

Strain	Genotype
G5620	Δ*hyaB* Δ*hybC* Δ*hycE* Δ*hyfG*
G5956	Δ*hyaB* Δ*hybC* Δ*hycE* Δ*hyfG* Δ*iscR*
G5984	Δ*hyaB* Δ*hybC* Δ*hycE* Δ*hyfG* Δ*iscR* Δ*adhE*
G6304	Δ*hyaB* Δ*hybC* Δ*hycE* Δ*hyfG* Δ*lpd* Δ*iscR* Δ*adhE*

For the cloning of the hydrogenase from *Cupriavidus necator*, CnHyd, it was initially planned that the cloning would be performed in our laboratory using native DNA from *C. necator* and Gibson cloning following the same operon organization described by Lamont and Sargent (2017) but using native DNA instead of ordering synthetic genes. However, considering the complexity of the cloning process, it was decided to contact the research team who already made this cloning process using *E. coli* optimized synthetic genes and demonstrated activity for these constructions. Frank Sargent kindly passed the plasmids harboring CnHyd operon to our laboratory which we used for the project.

Hydrogenase subunits of interest are separated into two plasmids. The first one is an IPTG-inducible pUC plasmid harboring all the subunits from the soluble hydrogenase, namely *hoxF*, *hoxU*, *hoxY*, *hoxH*, *hoxW* and *hoxI* (pQE80-SH, Figure 53 A). In addition to these constitutive subunits, a set of auxiliary genes are required (Schiffels et al. 2013) and were cloned in a pSU backbone: *hypA2*, *hypB2*, *hypF2*, *hypC1*, *hypD1*, *hypE1* and *hypX* (pSUhypA2-X, Figure 53 B).

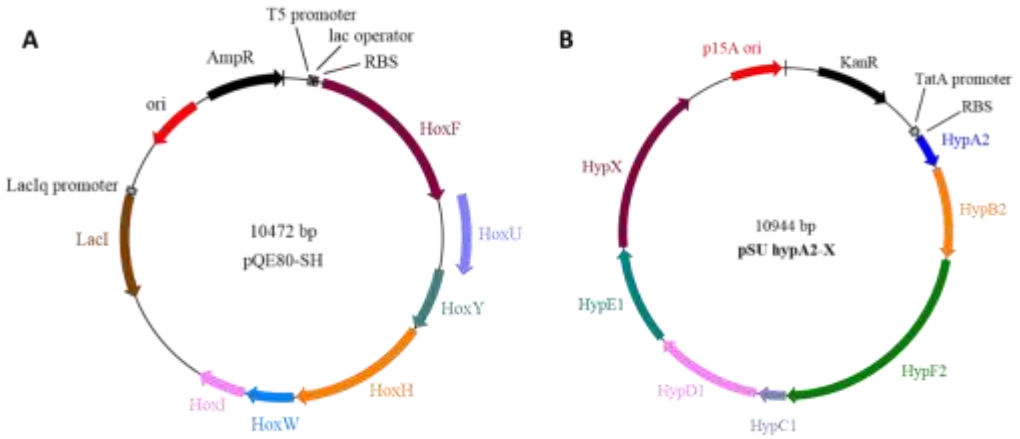


Figure 53. Plasmid maps for *Cupriavidus necator* hydrogenase (CnHyd) operon.(A) pQE80-SH plasmid map with *hoxF*, *hoxU*, *hoxY*, *hoxH*, *hoxW* and *hoxI* genes. (B) pSUhypA2-X plasmid map with *hypA2*, *hypB2*, *hypF2*, *hypC1*, *hypD1*, *hypE1* and *hypX* genes. Plasmids were a gift from Frank Sargent (Lamont and Sargent 2017).

The plasmids were inserted by transformation in the constructed strains described above and clones were selected on LB plates supplemented with carbenicillin (50 µg/ml) and kanamycin (30 µg/ml).

2. Test setup

After inoculating anaerobic cultures in Hungate tubes following the protocol described in Material and methods chapter (II. 3. 2. 3), OD_{600} was monitored with direct measure through the glass of the tubes. Once the cultures reached the log phase (OD_{600} ~ 0.5), the hydrogenase expression was induced by adding IPTG (1 mM). In addition, Nickel(II) sulfate hexahydrate (1 mM) and Ammonium Iron(II) sulfate hexahydrate (1 mM) salts were added to provide metals essential for the catalytic activity, plus L-cysteine (500 µM) as a source of sulfur. Subsequently, the cultures were incubated with agitation at 30 °C, unless stated otherwise.

3. H_2 production

To assess the functionality of CnHyd in the *E. coli* Δ*hyd* strains and validate the experimental setup, a first test of H_2 production was performed under anaerobic conditions in Hungate tubes. To do so, different strains harboring hydrogenase deletions along with *iscR*, *adhE* and/or *lpd* genes were tested: G5620, G5956, G5984, G6304 and MG1655 WT as a control. Each of these strains was transformed with the two plasmids of interest, pQE80-SH and pSUhypA2-X. All the strains were cultured in M9ZB medium supplemented with the appropriate antibiotics, the detailed composition of this medium is listed in Table 7 in the Material and methods chapter (II. 2. 1). The test was conducted following the procedure described above.

The composition of the headspace of the Hungate tubes was analyzed through gas chromatography (GC), as described in Material and methods chapter (II. 6. 3), 2 hours and 16 hours after hydrogenases induction. The results of these analyses are presented in Table 38. As expected, based on the literature (see introduction, I. 5. 1), the MG1655 WT strain is able to produce H_2 through its endogenous hydrogenases. On the contrary, the strains lacking these hydrogenases, G5620, G5956, G5984 and G6304, are not able to produce H_2, thus validating the metabolic screening strains. In addition, it is noteworthy that the amount of H_2 produced by MG1655 WT is much higher (after 2 hours of induction, close to 200-fold) than the hydrogen produced by the heterologous CnHyd in G5984 strain. This observation is coherent with the results published by Frank Sargent stating that they have a 50-fold difference in their value after 48 to 72 hours of incubation.

Table 38. Results obtained for the H_2 production experiment. The plasmids are pQE80-SH and pSUhypA2-X containing constitutive genes from CnHyd along with accessory proteins. For strains harboring plasmids, tubes were made in triplicates, the results showing the average values. *: the culture for the control strain G6304 turned black, probably due to iron sulfide, the OD_{600} measurement was not possible. The GC method used for these measurements is available in Appendix 10: GC method for H_2 detection.

Strain	Genotype	Plasmids	2 h after induction		16 h after induction	
			OD_{600}	H_2 peak area ($\mu V.s^{-1}$)	OD_{600}	H_2 peak area ($\mu V.s^{-1}$)
MG1655	WT	No	2.36	146655	> 2.5	433045
G5620	Δ*hyaB* Δ*hybC* Δ*hycE* Δ*hyfG*	No	2.12	0	2.253	0
G5956	Δ*hyaB* Δ*hybC* Δ*hycE* Δ*hyfG* Δ*iscR*	No	2.08	0	2.22	0
G5984	Δ*hyaB* Δ*hybC* Δ*hycE* Δ*hyfG* Δ*iscR* Δ*adhE*	No	1.20	0	1.338	0
G6304	Δ*hyaB* Δ*hybC* Δ*hycE* Δ*hyfG* Δ*lpd* Δ*iscR* Δ*adhE*	No	0.76	0	*	0
MG1655	WT	Yes	1.83	27784	1.83	180819
G5620	Δ*hyaB* Δ*hybC* Δ*hycE* Δ*hyfG*	Yes	1.84	0	2.18	0
G5956	Δ*hyaB* Δ*hybC* Δ*hycE* Δ*hyfG* Δ*iscR*	Yes	1.74	0	2.17	0
G5984	Δ*hyaB* Δ*hybC* Δ*hycE* Δ*hyfG* Δ*iscR* Δ*adhE*	Yes	1.00	324	1.68	3701
G6304	Δ*hyaB* Δ*hybC* Δ*hycE* Δ*hyfG* Δ*lpd* Δ*iscR* Δ*adhE*	Yes	0.93	654	1.91	2735

Figure 54 below represents the data for the hydrogen produced by *E. coli* strains lacking endogenous hydrogenases. As depicted on the graphs, only the strains lacking *adhE* (G5984 and G6304) are capable of detectable H_2 production from CnHyd. As previously described above (III. 3. 1. 1), this deletion increases NADH availability by preventing the reduction of acetyl-CoA to acetaldehyde and then to ethanol, oxidizing two molecules of NADH in the process. The higher availability of NADH in the cells appears to be necessary for H_2 production, which is consistent with the published data. After 16 hours of induction, the hydrogen yield is about 160 nmol $H_2.ml^{-1}.OD_{600}^{-1}$ for the G5984 strain. In the reference publication by Lamont and Sargent (2017), they reported approximately 240 nmol $H_2.ml^{-1}.OD_{600}^{-1}$ after 72 hours of incubation, with the corresponding strain. Thus, the value we obtained is within the same order of magnitude, indicating consistency in our results. For MG1655 wild-type strain, the hydrogen yield was about 4.7 µmol $H_2.ml^{-1}.OD_{600}^{-1}$ after 2 hours of incubation (the calculation after 16 hours was not possible due to the high OD_{600}, as indicated in Table 38). The same reference publication indicated a yield of approximately 12 µmol $H_2.ml^{-1}.OD_{600}^{-1}$ after 48 hours. The difference observed can be attributed to the drastically inferior incubation time from our experiments, but it still remains in the same order of magnitude. These measures validate the experimental setup and the efficient handling of the CnHyd plasmids.

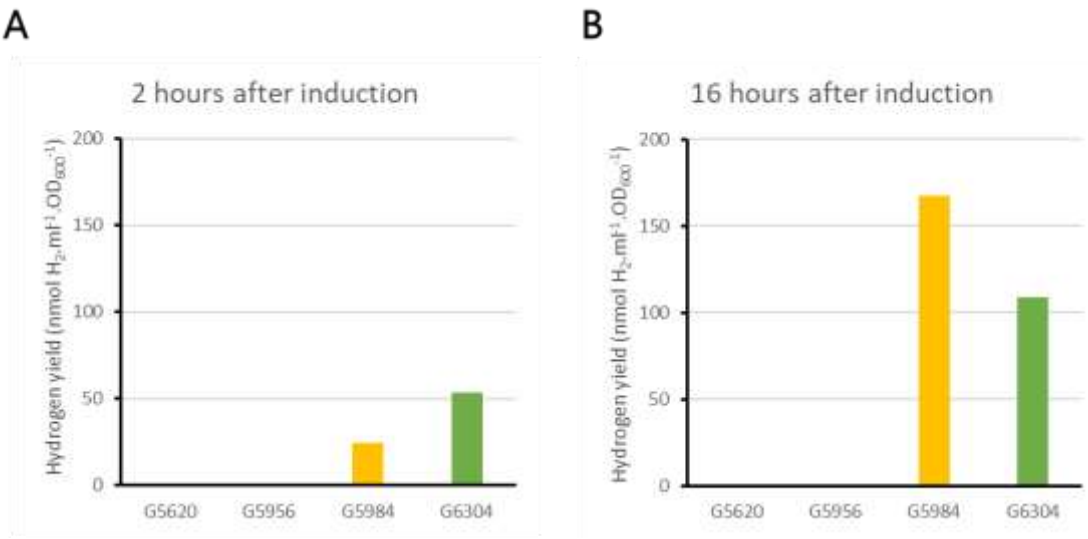


Figure 54. H_2 production of the strains lacking endogenous hydrogenases and harboring CnHyd on plasmids. The hydrogen yield was calculated from the GC and the OD_{600} measurements (Table 38). The peak area corresponding to the amount of H_2 produced was normalized with OD_{600} determined for the cultures. (A) Hydrogen yield 2 hours after induction. (B) Hydrogen yield 16 hours after induction.

In conclusion, this experiment produced consistent results compared to the literature on H_2 production using the synthetic expression system consisting of two plasmids, pQE80-SH and pSUhypA2-X, for CnHyd expression. Additionally, it demonstrates the functionality of the system in the H_2 production sense even in the impaired genetic context of the Δ*lpd* strain G6304 which lacks endogenous hydrogenases as well as the *adhE*, *iscR* and *lpd* genes. This result was the basis for the subsequent steps of this project.

4. NADH generation using hydrogenase

The functionality of the synthetic CnHyd system was demonstrated for H_2 production, the tests for its functionality for NADH generation are presented in this section.

For these tests, pre-cultures of the strain G6304 (Δ*hyaB* Δ*hybC* Δ*hycE* Δ*hyfG* Δ*lpd* Δ*iscR* Δ*adhE*) with and without the two plasmids pQE80-SH and pSUhypA2-X are performed in LB medium supplemented with ampicillin (50 µg/ml) and kanamycin (25 µg/ml) in aerobic condition. Subsequently, the cells are centrifuged and resuspended in M9ZB medium, growth is obtained and cells are once again resuspended in M9ZB supplemented with ampicillin (50 µg/ml) and kanamycin (25 µg/ml) in aerobic condition. This step is performed twice to ensure the complete removal of any remaining traces of LB medium from the cells. Finally, the cells are centrifuged and resuspended in the test medium described in Table 12. In parallel, cultures of the strain MG1655 WT with and without the plasmids are performed as controls following the same procedure.

The test cultures are performed in Hungate tubes with 5 mL of culture medium in aerobic conditions. The tubes are hermetically sealed in air and injection of 5 mL of nearly pure H_2 (98 % H_2, 2 % O_2) is realized using a syringe every 24 h. Unfortunately, after 72 h of cultures, no growth in these conditions was observed in the metabolic screening strain G6304 with the two plasmids of interest. As expected, the MG1655 wild-type strain with or without plasmids grew stably under these conditions.

5. Discussion and perspectives

The functionality of the CnHyd, encoded by optimized genes organized in a synthetic operon, was successfully demonstrated under anaerobic conditions in *E. coli* strains lacking endogenous hydrogenases. In addition, the deletion of *adhE* gene was shown to be essential for such activity in these conditions. However, the reaction direction targeted for the metabolic test being the opposite one, *i.e.* the oxidation of H_2 to produce NADH, this deletion is unnecessary for this specific context. In aerobic conditions, this gene is repressed (Leonardo, Dailly, and Clark 1996) and its presence would not affect the phenotype. Nevertheless, for the initial validation of the Δ*lpd* screening strain and the demonstration of the heterologous operon efficient expression, this deletion was necessary.

The CnHyd is physiologically working in the H_2 oxidative direction in *Cupriavidus necator* as H_2 is used as electron donor under autotrophic growth mode. However, its heterologous activity in *E. coli* in this direction has never been demonstrated.

Due to a lack of time and the complexity of the protocol, we were not able to perform a large number of assays for H_2 oxidation. Those performed failed to reveal this activity in the test setup. A few hypotheses can be formulated to explain these findings. First, there is an important difference between the medium used for the H_2 production assays, which is a rich medium, and the minimal medium for the metabolic test only containing acetate and pyruvate as carbon sources compounds. Growth in this strictly minimal medium might not be compatible with the functioning of a complex enzymatic system like CnHyd, requiring the availability of amino acids and diverse growth factors for efficient expression and protein folding. The addition of Ni and Fe supplements to the medium, required for the maturation of the complex, was not sufficient for its activity. Adjustments of this medium without compromising the metabolic test are probably required. Lamont and Sargent (2017) added casaminoacids (0.8 %) as a supplement in the culture medium. This supplement did not allow growth of the metabolic screening strain without hydrogenases, its addition will be tested for the metabolic test. Additionally, the *hoxN1* gene from *C. necator* encoding a high-affinity nickel transporter should be added to the set of auxiliary genes as it was reported to improve the maturation efficiency of CnHyd expressed in *E. coli* under aerobic conditions (Schiffels et al. 2013).

As mentioned in the *Expression of a functional complex formate* dehydrogenase in *E. coli* section (III. 1), the use of synthetic genes for the expression of complex proteins such as this hydrogenase can lead to lower activity levels, probably due to inefficient maturation (Dr. Silke Leimkühler). This hypothesis is supported by the study published by Ghosh, Bisaillon, and Hallenbeck (2013) where the heterologous expression of the native CnHyd on plasmids was able to restore growth in Δ*adhE* strain under fermentative conditions, while the synthetic, codon optimized operon was not (Lamont and Sargent 2017). Their metabolic test consisted on taking advantage of the NAD^+-generation coupled to H_2 production by CnHyd in a strain lacking *adhE*, *ldhA* and *mdh* genes, the main NAD^+-generating systems in these anaerobic growth conditions. The fact that this activity could not be achieved with synthetic genes, as stated by F. Sargent, emphasizes the potential lower efficiency of their expression. Consequently, it would be interesting to perform the cloning with native genes, following the same construction patterns as described in the literature to test if better results are obtained.

Overall, many options are still to be explored to be able to state on the feasibility of NAD^+-dependent hydrogenases as energy regeneration systems. However, even if such activity was demonstrated, the explosivity of H_2 gas will remain a challenge for its manipulation, as well as its relative insolubility.

The redox potential for hydrogen at pH 7 is -410 mV while the one of NADH is -320 mV making H_2 oxidation/NAD^+-reduction a favorable redox couple under standard conditions, even so the difference between these two potentials is not very high. The NADH-generating enzyme, phosphite dehydrogenase, described in the following paragraph is hypothesized to be more favorable, partly because of the redox potential of phosphite of -650 mV, which is highly negative, thus more favorable for NAD^+ reduction.

2. Phosphite dehydrogenase

The study of the phosphite dehydrogenase activity was initiated by Gaëlle Girard in the frame of an M2 internship under my supervision in the laboratory.

As an alternative system to the highly complex hydrogenase described above, the option of phosphite as energy source for our autotrophic strain was explored. In this context, the NAD^+-dependent phosphite dehydrogenase from *Pseudomonas stutzeri*, catalyzing the reaction described below (Equation 5), was studied.

Equation 5. Phosphite oxidation catalyzed by PTDH from P. stutzeri.

$$HPO_3^{2-} + NAD^+ + H_2O \rightarrow PO_4^{3-} + NADH + 2H^+$$

This reaction is highly favorable considering the low redox potential of the phosphate/phosphite couple of -650 mV.

1. Plasmid & strain construction

1. *Strains*

To evaluate the activity of the phosphite dehydrogenase (PTDH) from *P. stutzeri* in oxidizing phosphite to phosphate while generating NADH, the Δ*lpd* background already used in other contexts for NADH production, served again as selection scheme. However, to ensure that native *E. coli* genes described for having phosphite oxidation activity are not interfering with the selection, two additional deletions are necessary.

The first system is the bacterial alkaline phosphatase (BAP), encoded by the gene *phoA.* This enzyme allows *E. coli* to use phosphate esters as alternative sources of phosphorus (Yang and Metcalf 2004). Its deletion eliminates the potential for phosphite oxidation through the BAP-mediated pathway. The second system is the C-P lyase enzyme encoded by the *phn* operon. The *phn* operon consists of 14 genes spanning over 10 kbp (Metcalf and Wanner 1991). For testing heterologous phosphite dehydrogenase (PTDH) activity, one gene from the *phn* operon, *phnI*, which is essential for the phosphite oxidation activity of the complex (Metcalf and Wanner 1993), was specifically deleted. It has been demonstrated that an *E. coli* strain lacking these two pathways (BAP and C-P lyase) is unable to grow on phosphite in a medium where phosphate is absent (Yang and Metcalf 2004).

Therefore, in addition to the *lpd* gene deletion, the metabolic screening strain must also carry deletions of the *phoA* gene and the *phn* operon to create a background where native phosphite oxidation

pathways are disrupted. These deletions ensure that the phosphite supplied to the cells will serve for phosphite oxidation through the heterologous expression of the cytoplasmic PTDH from *P. stutzeri*.

Both the Phn and BAP enzymes are periplasmic, supporting the hypothesis that phosphite is not transported to the cytoplasm in *E. coli*. Hence, it is supposed that the presence of the specific membrane phosphite transporter encoded by the three genes *ptxABC* in *P. stutzeri* genome would be essential for the cytoplasmic activity of PTDH, enabling the import of phosphite into the cytoplasm for further enzymatic oxidation.

To construct the *E. coli* strain lacking these three enzymatic systems, the genes *phoA*, *phnI* and *lpd* were sequentially deleted using transduction. P1 phages were prepared with the corresponding Keio strains: JW0374 for Δ*phoA*::*kan* allele, JW4060 for Δ*phnI*::*kan* and JW0112 for Δ*lpd*::*kan*. Each transduction was followed by a resistance cassette excision using the FLP-recombinase system. The G5626 Δ*phoA* Δ*phnI* and G5658 Δ*lpd* Δ*phoA* Δ*phnI* strains were successfully obtained.

2. *Plasmids*

The four genes of interest, namely *ptxD* (phosphite dehydrogenase) and the three genes coding for the subunit of *ptxABC* (phosphite transporter), were codon optimized and synthesized by Twist Biosciences.

Subsequently, *ptxD* was cloned into a pUC18 plasmid, creating the plasmid pGEN1325, while the transporter genes (*ptxA*, *ptxB*, and *ptxC*) were cloned into a pSU18 plasmid, designated as pGEN1337. The genes were cloned using restriction-ligation cloning and the obtained plasmids are represented in Figure 55.

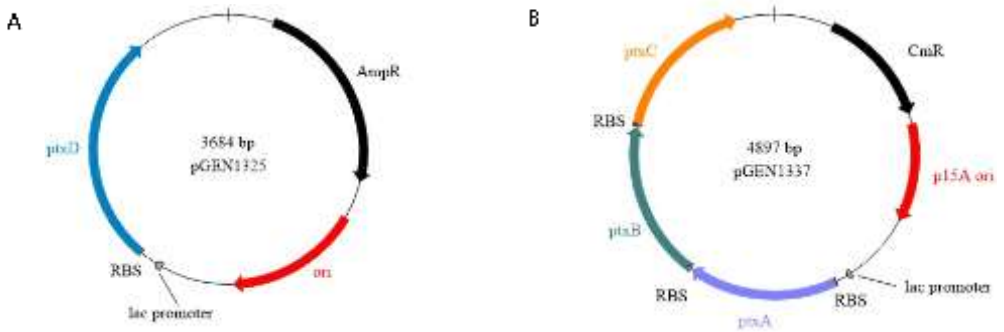


Figure 55. Plasmid for the expression of the phosphite dehydrogenase from *P. stutzeri*. (A) pUC18-*ptxD* plasmid (pGEN1325). (B) pSU18-*ptxABC* plasmid (pGEN1337).

2. Test of the PTDH activity towards phosphite as phosphate source

1. *Limiting phosphate concentration*

The metabolic test was carried out using MOPS minimal medium, which lacks phosphate (II. 2. 2). This medium was chosen specifically to assess the activity of PTDH when phosphite is used as the sole phosphate source for the cells.

Initially, a test to determine the limiting phosphate concentration to support growth in *E. coli* was conducted using a phosphate concentration range. An overnight pre-culture of the MG1655 *E. coli* WT strain was prepared in MOPS medium supplemented with glucose (0.2 %) and phosphate (10 mM). To remove any residual phosphate that could potentially interfere with the subsequent test, the cells were washed twice using MOPS medium only. Based on the preliminary test conducted, it has been determined that the minimal phosphate concentration necessary for cell growth is between 0 to 0.5 mM. Thus, growth of MG1655 strain on a phosphate concentration range between these two values was performed. The results are summarized in the Table 39 below.

Table 39. Limiting phosphate concentration for *E. coli* MG1655 WT. Cells were grown in MOPS minimal medium supplemented with glucose (0.2 %) and phosphate from 0 to 0.5 mM.

Phosphate (mM)		0	0.1	0.2	0.3	0.4	0.5
OD_{600}	24 h	0.014	0.84	1.50	1.78	1.99	1.96
	48 h	0.014	0.80	1.24	1.60	1.74	1.70
	72 h	0.015	0.76	1.21	1.54	1.70	1.72

These results clearly indicate that growth is significantly limited when the phosphate concentration falls below 0.2 mM. This experiment provides valuable data for the following experiments, especially for the test of phosphite as a phosphate source by establishing the minimal phosphate concentration required for cell growth.

2. *Phenotype verification*

To confirm the expected phenotype of the strain G5626 (Δ*phoA* Δ*phnI*), *i.e.* no growth in the absence of phosphate and the presence of phosphite in the medium, a growth test was conducted. Both the MG1655 WT and G5626 strains were pre-cultured overnight in permissive conditions with MOPS medium supplemented with glucose (0.2 %) and phosphate (25 mM). Then, the cells were washed twice

with MOPS medium and tested in MOPS medium supplemented with 0.2 % glucose and a range of phosphite concentrations from 0 to 100 mM. Testing phosphite concentrations up to 100 mM appeared to be necessary on the hypothesis that the amount of phosphite required to support growth as a phosphate and energy source would be significant. Thus, it was necessary to verify the phenotype of the constructed strain under these high phosphite concentrations. The results are presented in Figure 56.

As shown in Figure 56, when comparing the growth rate and yield between the WT strain and strain G5626 under permissive conditions with 25 mM of phosphate (brown curve), no deleterious effect of the two gene deletions on the growth phenotype of G5626 is detected. Additionally, at low phosphite concentrations of 1, 5 and 10 mM, very little to no growth was observed for strain G5626, which is consistent with data reported in the literature (Yang and Metcalf 2004). In contrast, at higher phosphite concentrations from 25 mM and above, similar growth patterns were observed between the two strains. As a side note, the slight growth observed in the condition with no phosphite may be attributed to carry-over effects from the pre-culture used for the metabolic test. The deletion of the two systems, BAP and C-P lyase, appears to have a notable impact on the G5626 strain's ability to grow at low phosphite concentrations, but does not impede growth at high phosphite concentrations.

Furthermore, the experiment points out a moderate toxic effect of phosphite concentrations of 50 and 100 mM with a longer lag phase compared to lower concentrations for both strains but the final OD_{600} reached by these cultures is very similar. This observation is crucial considering the probable high concentrations that will be required to support growth in phosphite as energy source tests.

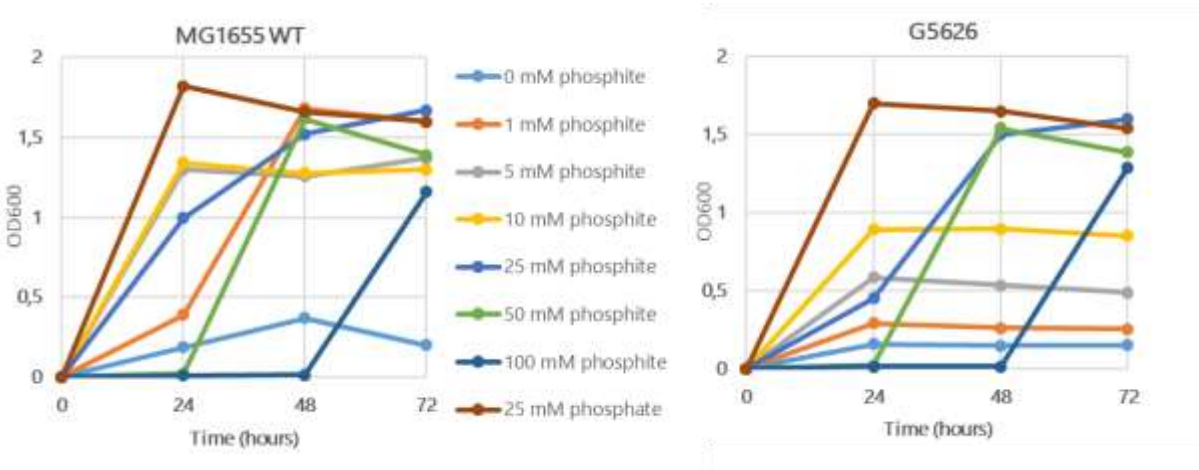


Figure 56. Growth of *E. coli* MG1655 WT and G5626 (Δ*phoA* Δ*phnI*) on a range of phosphite concentrations. The strains were grown in MOPS mineral medium supplemented with 0.2 % glucose and a phosphite range from 0 to 100 mM. A positive control with 25 mM phosphate was performed (brown curve).

As there were no other phosphite oxidizing systems reported for *E. coli* that could explain the growth observed with high phosphite concentrations, it was hypothesized that the sole deletion of the *phnI* gene might not be sufficient to completely prevent the activity of the C-P lyase. To further investigate this, the *phnGHIJK* genes were deleted using Wanner recombination (II. 4. 3), resulting in the strain G5684 (Δ*phoA* Δ*phnGHIJK*). However, it was observed that these additional deletions did not alter the growth profile observed previously. As a result, the original strain G5626 (Δ*phoA* Δ*phnI*) was used for further tests and analysis (see Appendix 11: Test of strain G5684 in phosphite range).

These results suggest that phosphite might enter the cells using Pi transporters such as PstSCAB, PitA/B, PhnCDE, as reported by Hirota et al. (2010). Subsequently, phosphite could be further oxidized to phosphate by an unidentified system, periplasmic of cytoplasmic, which appears to be less efficient compared to the two phosphite oxidation systems already reported (BAP and C-P lyase).

It is important to note that the available literature does not mention experiments conducted at such high phosphite concentrations. The absence of growth observed in *E. coli* with *phn* and *phoA* deletions was only demonstrated at 0.5 mM phosphite (Yang and Metcalf 2004).

3. *PTDH activity test*

To investigate the ability of the strain G5626 (Δ*phoA* Δ*phnI)* to utilize phosphite as a phosphate source, several descending strains were constructed and tested:

- G5626 (Δ*phoA* Δ*phnI*): base strain lacking the *phoA* and phnI genes.
- G5654 (Δ*phoA* Δ*phnI* p::*ptxD*): strain carrying the pUC18-*ptxD* plasmid (pGEN1325).
- G5656 (Δ*phoA* Δ*phnI* p::*ptxABC*): strain carrying the pSU18-*ptxABC* transporter plasmid (pGEN1337).
- G5660 (Δ*phoA* Δ*phnI* p::*ptxD*, *ptxABC*): strain carrying both the pUC18-*ptxD* plasmid and the pSU18-*ptxABC* transporter plasmid.

These strains were used to conduct growth experiments in MOPS medium supplemented with 0.2 % glucose and a range of phosphite concentrations: 0 mM, 0.1 mM, 0.5 mM, 1 mM, and 5 mM phosphite. The results are depicted in Figure 57.

Figure 57 demonstrates the activity of *ptxD* in phosphite oxidation, as evidenced by the differences in growth, between strains lacking this gene (G5626 and G5656) and strains expressing it (G5654 and G5660), the latter showing faster growth and lower lag phase. Additionally, growth was observed at 0.1 mM and 0.5 mM phosphite concentrations when *ptxD* was expressed, the co-expression of the transporter only having a limited impact on growth. Concentrations of 1 and 5 mM supported growth also in the absence of PtxD albeit lower than in its presence. The transporter does not have a substantial

impact on growth, when working with high concentrations of phosphite, which will be necessary for testing phosphite as an energy source, its utilization will not be necessary. Phosphite enters the cells without this specific transporter.

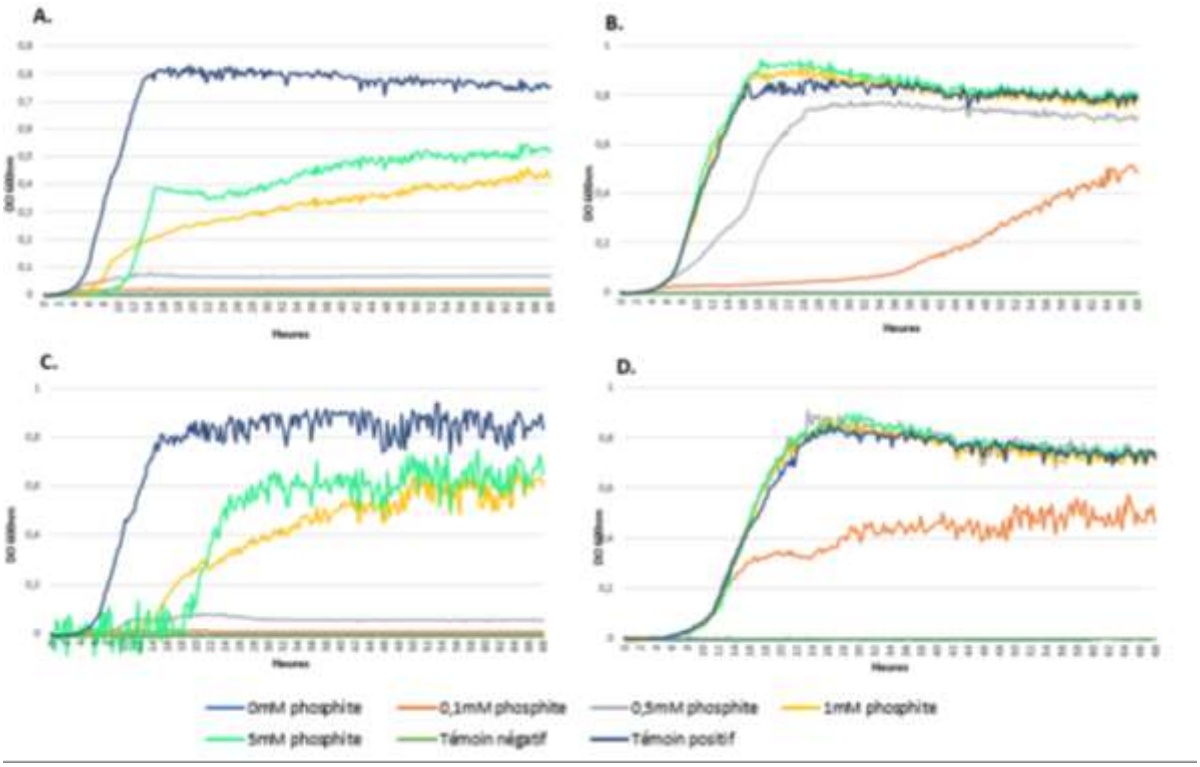


Figure 57. Growth of strains expressing gene *ptxD* with phosphite as phosphate source. (A) G5626 (Δ*phoA* Δ*phnI*). (B) G5654 (Δ*phoA* Δ*phnI* p::*ptxD*). (C) G5656 (Δ*phoA* Δ*phnI* p::*ptxABC*). (D) G5660 (Δ*phoA* Δ*phnI* p::*ptxD*, *ptxABC*). In this experiment, the negative control corresponds to the medium test without inoculum and the positive control corresponds to strains grown in MOPS minimal medium supplemented with glucose (0.2 %) and phosphate (25 mM). These growth curves were took from Gaëlle Girard internship report without access to the original data to harmonize the lay out.

In the context of phosphite metabolism, extensive studies have been conducted in *E. coli* and *P. putida* for biocontainment purposes. In these studies, researchers engineered the phosphite-dependent strains and deleted every known phosphate transporter to prevent the modified strains from using phosphate from the environment (Hirota et al. (2017); Asin-Garcia et al. (2022)). In our experiment, as the goal was not to eliminate all possibilities of phosphate entry into the cells, these additional deletions were not performed. Thus, it can be hypothesized that the native *E. coli* phosphate (Pi) transporters have the ability to transport phosphite (Pt) due to their structural similarity. Indeed, it has been demonstrated that an *E. coli* strain lacking the four inorganic Pi transporters and the three Pi-ester transporters was unable to grow if phosphate was provided in the culture medium (Hirota et al. 2017). This could explain why the

expression of *ptxABC* (G5660) did not exhibit a significant advantage compared to the strain without the corresponding plasmid (G5654).

3. Test of the PTDH activity towards phosphite as energy source

1. *Activity test from plasmid expression*

To evaluate the possibility of using phosphite as phosphate and energy sources using PTDH from *P. stutzeri*, the plasmid carrying *ptxD* (pGEN1325) was transformed into the strain G5658 resulting in the strain G5667 (Δ*phoA* Δ*phnI* Δ*lpd* p::pUC18-*ptxD*). In this experiment, we focused on the expression of *ptxD* alone, as the previous experiments demonstrated that the *ptxABC* transporter was not significantly advantageous at high phosphite concentrations. In this selection strain, acetate and pyruvate are the carbon sources and phosphite the NADH source.

It appeared that the growth of the strain G5667 in the permissive medium, MOPS supplemented with glucose (0.2 %), acetate (20 mM), phosphate (10 mM), and carbenicillin (50 µg/ml), was slow in liquid conditions and only tiny colonies were obtained after 5-7 days on adaptive plates. This is likely due to the Δ*lpd*-derived strains that show impaired growth (see III. 1). It also indicates that MOPS minimal medium may not be suitable to support growth from this strain. Consequently, another approach was performed in parallel. The same adaptation process was carried out on MS minimal medium supplemented with glucose (0.2%), acetate (20 mM), and carbenicillin (50 µg/ml). In this context, the growth of the strain slightly improved.

It is important to note this specific context of testing phosphite as an energy source, the use of MOPS minimal medium is not mandatory. Indeed, since the cells will need to use phosphite to produce NADH and support growth, the presence of phosphate in MS minimal medium does not interfere with the metabolic test. However, the presence of phosphate in MS medium might not favor the PtxD activity.

A range of phosphite concentrations from 25 mM to 100 mM was tested in both MOPS or MS media supplemented with acetate (20 mM) and pyruvate (20 mM). However, no growth was demonstrated. This could be linked to the impaired metabolism of the Δ*lpd* strain, which might struggle to handle this enzyme. However, it is rather unlikely the main reason as previous experiments with much more complex enzymatic systems in the same genetic background showed positive results. The other hypothesis is the lack of efficiency from the heterologous PTDH to support growth in these conditions. The rate of NADH generation might not be high enough to support the cell's needs. Additionally, we observed that an undescribed system was able to oxidize phosphite to phosphate, probably creating competition with PtxD.

2. *Activity test from chromosomic expression*

1) Strain construction with PtxD inserted in the chromosome

An additional attempt was made by integrating *ptxD* into the chromosome of the G5658 (Δ*phoA* Δ*phnI* Δ*lpd*) strain through homologous recombination. The purpose of this chromosomal insertion was to establish a more stable system and enable the evolution of this strain in a medium swap continuous culture experiment in a GM3 for the selection of cells growing with phosphite as sole energy source.

The genome modification was performed following the procedure described in the Material and methods chapter (II. 4. 6. 2). The recombination of the double-stranded linear fragment containing *ptxD* was performed in G5658 strain and the SpecR cassette was excised, leading to the construction of the G5791 strain (Δ*phoA* Δ*phnI* Δ*lpd kdgK*::*ptxD*), where *ptxD* was successfully integrated into the *kdgK* locus (Figure 58). This gene, coding for the non-essential enzyme 2-dehydro-3-deoxugluconokinase, has been used as a chromosomal integration site in previous studies of the laboratory (Bouzon et al. 2017).

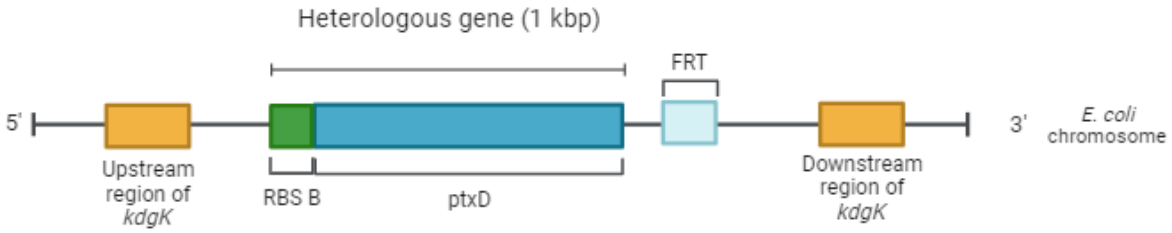


Figure 58. *ptxD* gene from *P. stutzeri* integrated in the *E. coli* chromosome at the *kdgK* locus. Designed on Biorender.

2) Test of the strain expressing PtxD from the chromosome

The growth on phosphite as phosphate source of strain G5791 was monitored to validate the activity of *ptxD* from a chromosomal expression. The results are represented in Figure 59.

At phosphite concentrations of 5 mM and 10 mM, the presence of the *ptxD* gene reduced the lag phase, indicating an improvement in the time taken for the strain to initiate growth. However, there was no significant difference observed in biomass production (final OD_{600}) between the strain with *ptxD* and the strain without it.

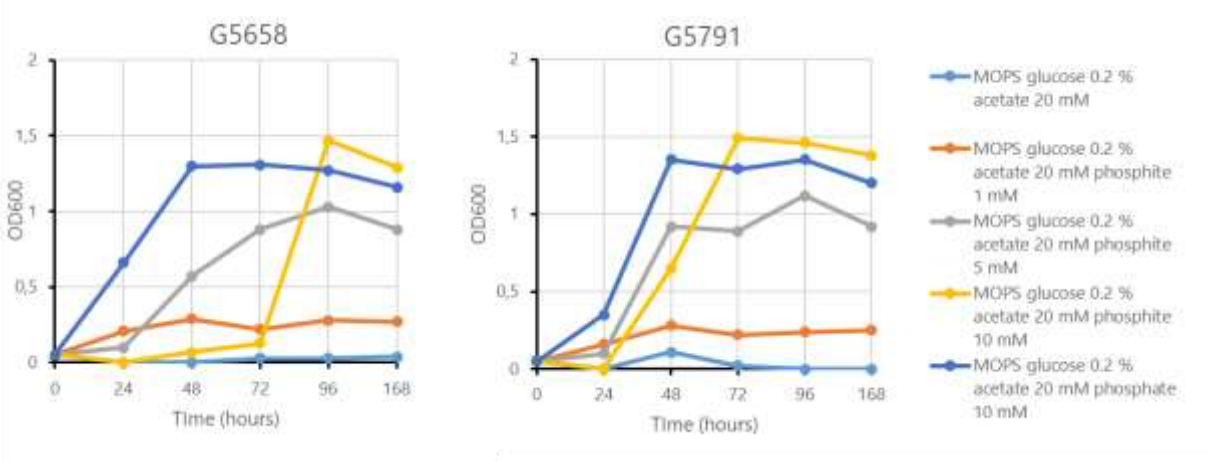


Figure 59. Growth of strains G5658 (Δ*phoA* Δ*phnI* Δ*lpd*) and G5791 (Δ*phoA* Δ*phnI* Δ*lpd kdgK::ptxD*) within a range of phosphite concentrations from 1 to 10 mM. Strains were grown in MOPS mineral medium supplemented with glucose 0.2 % acetate 20 mM medium supplemented with the indicated concentration of phosphite ranging from 0 to 10 mM.

In the case of 1 mM phosphite in the medium, both strains exhibited very poor growth. Comparing this growth profile to the one obtained with *ptxD* expressed from a plasmid, the growth obtained with genomic integration is much lower, indicating a lower expression from the chromosome. This is probably due to the fact that the expression from a single copy in the chromosome is much lower compared to the one obtained from the high-copy plasmid, pUC18. The same strong RBS (RBS A, see Appendix 1: Genetic sequences of diverse building blocks for sequence) was used for both plasmidic and chromosomic expression. However, the promoter differs with chromosomic expression as *ptxD* is placed under the endogenous *kdgK* promoter while the expression from the plasmid was under the control of P_{lac}. These results highlight the impact of gene expression levels between the two constructed strains.

We judged from the growth curves that strain G5791 (Δ*phoA* Δ*phnI* Δ*lpd kdgK*::*ptxD*) utilized phosphite as a phosphate source upon PtxD activity. The test of phosphite as an energy source, conducted under the same conditions described above, yielded the negative result expected.

3) GM3 evolution

Subsequently, we decided to employ a directed evolution approach using the GM3 device. To facilitate this process, we conducted a medium swap evolution as described in the Material and methods chapter (II. 3. 3). The permissive medium consisted of MOPS minimal medium supplemented with glycerol (0.2%), acetate (20 mM), pyruvate (20 mM), and phosphite (50 mM). The stressing medium was identical to the permissive medium except for the absence of glycerol. Glycerol was chosen as the energy and carbon source for the permissive medium due to its higher level of reduction compared to glucose. It was

expected that this choice would minimize the metabolic gap and facilitate the transition of enzymes from gluconeogenesis metabolism between the permissive and stressing conditions.

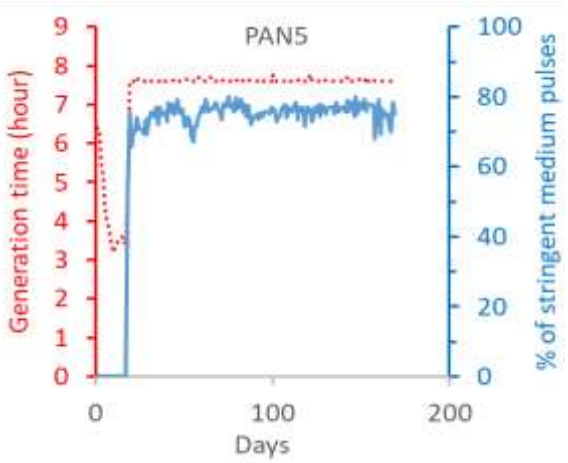


Figure 60. GM3 evolution of the strain G5791 in medium SWAP regime for growth on phosphite as sole energy source. Permissive medium: MOPS medium supplemented with glycerol (0.2 %), acetate (20 mM), pyruvate (20 mM), and phosphite (50 mM), stressing medium: permissive medium without glycerol. Cultures were grown at 30°C.

The directed evolution process was carried out for approximately 5 months, representing approximately 580 generations. As depicted in Figure 60, growth in 100 % of stressing medium was not obtained, during the evolution. Despite these efforts, we were unable to identify mutants capable of growing on phosphite as the sole energy source.

4. Discussion and perspectives

The project aimed to investigate the potential of phosphite as a NADH generation system, following promising hypotheses from the existing literature (Claassens et al. 2018). We were able to demonstrate the phosphite oxidizing activity of the phosphite dehydrogenase from *Pseudomonas stutzeri in vivo* to produce phosphate. This demonstration has already been reported for *E. coli* and *P. putida* using PtxD as well (Hirota et al. (2017); Asin-Garcia et al. (2022)). However, the enzyme did not support growth with phosphite as sole energy source. This result can either be due to a lack of activity from the PTDH or due to the presence of an unidentified enzyme or pathway in *E. coli* that operates independently of NAD^+ and is capable of oxidizing phosphite to phosphate.

A lack of activity is unlikely to be responsible for the negative result by itself because of the kinetic parameters that were calculated, demonstrating a k_{cat} of 7.3 s^{-1} and a very low K_m of 53.1 µM for

phosphite (Costas, White, and Metcalf 2001). However, these *in vitro* data cannot automatically be extrapolated to the *in vivo* situation, low expression or imperfect folding of the peptide chains could occur. As for the presence of unidentified pathways of phosphite conversion, we demonstrated that growth in phosphate-free medium is restored in strains lacking the described phosphite oxidizing enzymes (BAP and C-P lyase) with high phosphite concentrations. An unknown process operating either in the periplasm, preventing phosphite access to the cytoplasm where the heterologous PTDH is located, or within the cytoplasm itself, potentially with a higher efficiency or affinity for phosphite compared to PtxD, possibly using an electron carrier different from NAD^+, could be involved. However, this latter hypothesis is questionable since if such system exists, the large phosphite excess provided to the cells should be sufficient to also be used as a substrate by the PtxD enzyme.

To gain a deeper understanding of this potential NADH regeneration system, further investigations are necessary. It would be essential to perform a quantification of phosphite consumption in the test medium. This quantification can be performed enzymatically by the same PTDH as the one described here, using a cascade of reactions starting with the oxidation of phosphite to phosphate and ending with the reduction of resazurin to resofurin which is highly fluorescent, thus is easy to measure (Berkowitz et al. (2011); Bailey and Greene (2023)). This would give insights into whether phosphite is effectively oxidized and to which extent. In addition, constructing and testing a library of knock-out mutants could help identify the precise source of phosphite oxidation to be deleted to prevent interference with PTDH. Finally, efforts should be made to determine the specific location within the cell where phosphite oxidation occurs, whether in the periplasm or cytoplasm. Collectively, these investigations would contribute to clarify if phosphite can effectively be used as an NADH source.

As elaborated in the introduction, the phosphite oxidizing system that was tested was from *P. stutzeri*, an organism that only exploits phosphite as a phosphate source. It was hypothesized that the reason for this is the consumption of two ATP molecules for the transport of one phosphite molecule by *ptxABC*. Thus, the use of such system for the cells would be energy-neutral at best, and energy-consuming most probably (Figueroa and Coates 2017). However, *Desulfotignum phosphitoxidans* FiPS-3 is an organism capable of consuming phosphite use as energy source. In this organism, the *ptx-ptd* gene cluster is most probably responsible for phosphite oxidation, but is not well-described. PtxD is present as well, like in *P. stutzeri* but the transport is realized by PtdC, an energy-neutral antiporter, and an energy conservation mechanism, mediated by *ptdFGHI* is most likely necessary for the use of phosphite as an energy source (Figueroa and Coates 2017). The exploration of these systems, although poorly described, is probably necessary for a better understanding of phosphite metabolism.

4. Formate dehydrogenase purification for *in vitro* activity testing

1. BECOOL-3D project presentation

This project, named BECOOL-3D which stands for 3D-Bioelectrodes for the electrocatalytic conversion of CO_2 into a liquid energy vector aims to build oxygen-tolerant 3D bioelectrodes for the selective conversion of CO_2 to formic acid (Figure 61). To this end, metal-dependent oxygen-tolerant FDHs will be immobilized into a redox hydrogel film that can self-assemble on the surface of the electrode. These hydrogels will serve as a protective and solvated environment for the enzyme while optimizing the electrical connectivity of the biocatalyst with the electrode.

This project, funded by the CNRS, was conducted in collaboration between the laboratory of bioremediation, biocatalyse and synthetic metabolism (L2BMS) represented by Dr. Volker Döring from Genoscope and the Chemistry and Biology of Metals Laboratory (LCBM) from CEA Grenoble represented by Dr. Luca Albertin.

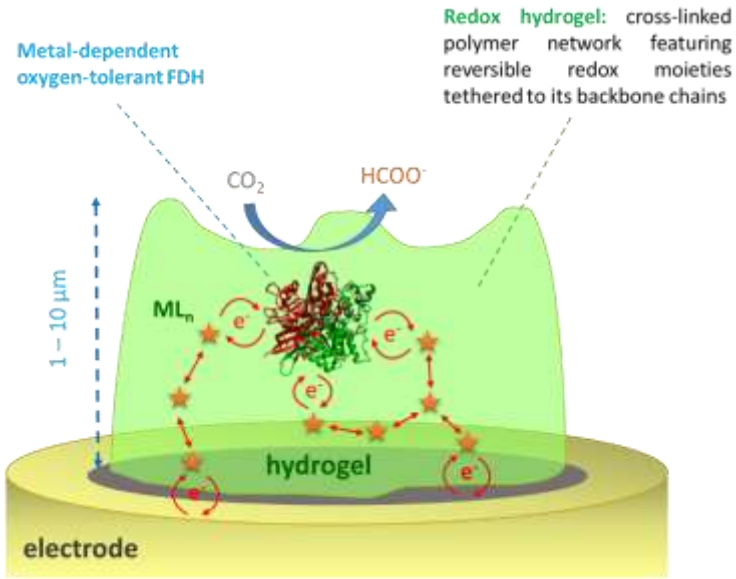


Figure 61. Schematic representation of the experimental setup for the BECOOL-3D project. This illustration is from Luca Albertin (LCBM, CEA Grenoble).

In the context of this project, two metal/NAD^+-dependent FDHs were used: *Rhodobacter capsulatus* and *Cupriavidus necator*. The responsibility of the Genoscope in this project was to provide batches of functional formate dehydrogenases to LCBM. LCBM then utilized these enzymes to immobilize them on the hydrogel and perform CO_2 reduction experiments.

2. Plasmid construction

For the expression and purification of the FDH from *Cupriavidus necator*, the previously constructed plasmid pTRC-CnFDH (pGEN1340) was used as a base to introduce a His-Tag at the N-terminal end of the enzyme. This cloning process was successfully conducted through Gibson cloning, resulting in the generation of the plasmid pGEN1353 (pTRC-HisTagCnHyd), which was sequenced by Sanger sequencing and validated.

Furthermore, a plasmid containing the FDH from *Rhodobacter capsulatus* was kindly provided to us by Dr. Silke Leimkühler (University of Potsdam, Germany). This plasmid shared the same construction items as the one described above, *i.e.* a pTRC backbone with a His-Tag fused at the N-terminal extremity of the RcFDH operon under the control of an IPTG-inducible promoter. This construction, named pTHfds05, is described and published by Hartmann and Leimkühler (2013).

3. Expression and purification

The two plasmids harboring the FDHs from *C. necator* and *R. capsulatus* (pGEN1353 and pTHfds05) were transformed in DH5α electrocompetent cells by electroporation. The protocol for protein purification is described in the Material and methods chapter (II. 6. 1).

1. FDH from *Rhodobacter capsulatus*

For the production and purification of RcFDH, cultures of 500 mL were performed. A fully automated two-step method was set up for each protein in which a Ni^{2+} affinity column was used in the first purification step. The eluted peak was redirected on a size exclusion column (method described in Perchat et al. (2018)). The size exclusion chromatography graph obtained is shown in Figure 62.

As shown in Figure 62, four peaks were obtained after the gel filtration column. A sample from each fraction is then analyzed by SDS-PAGE. Peaks 2 and 3 have the same protein profile on the gel and harbored three bands corresponding to the size of the three subunits FdsA (105 kDa), FdsB (52 kDa) and FdsG (15 kDa). The fact that two distinct peaks with the same SDS-PAGE gel profile were obtained suggests potential differences in their quaternary structure or the partial unfolding of one of the protein species in one of the peak fractions. Peak 4 fraction analysis on SDS-PAGE only reveals two major bands corresponding to FdsB and FdsG. This elution profile is similar to the one published by Hartmann and Leimkühler (2013) (Figure 63 below). In both cases, four peaks were obtained with two smaller peaks on both sides of the two principal ones. However, comparing the SDS-PAGE gel of the two major peaks from their study with what we obtained, revealed a difference between the profiles. In their study, the two main peaks corresponded to the two structures FdsABG and FdsBG, making it different from what

we observed. Their culture conditions are highly similar to the ones we performed, making it challenging to find an obvious reason for the gel variation we observed.

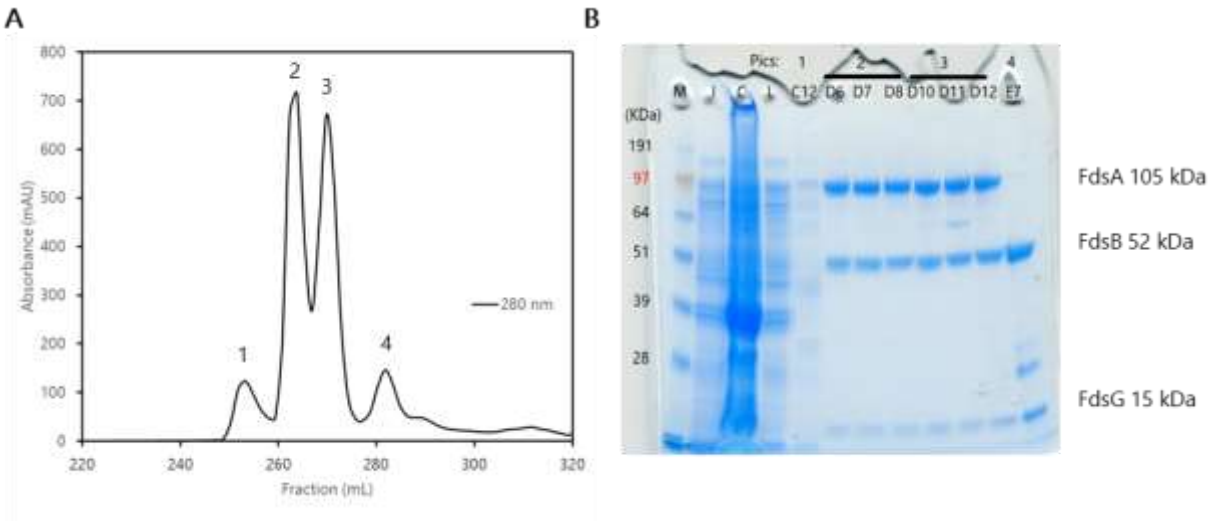


Figure 62. Size exclusion chromatography of RcFDH. (A) The black curve is the absorbance at 280 nm after column elution. (B) SDS-PAGE gel analysis. M: ladder of protein of known molecular weights. I: induction. C: pellet. L: cell lysate. Wells "Pic 1" to "Pic 4": proteins from collected fractions corresponding to the elution graph (A).

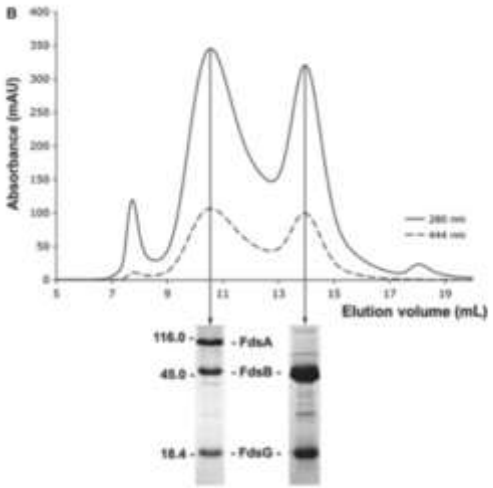


Figure 63. Size exclusion chromatography and purification of pTRC-HisTagRcHyd (pTHfds05) with SDS-PAGE gel analysis obtained by Hartmann and Leimkühler (2013).

The elution fractions obtained from peaks 2 and 3 were pooled because of the similarity of the profiles after SDS-PAGE analysis, and the total protein content was determined using the Bradford method. From a 1 L cell culture, a total of 8.5 mg of protein exhibiting the appropriate SDS-PAGE profile was obtained. The reference publication from Hartmann and Leimkühler (2013) describing the plasmid construction

used exhibited a yield of 6 mg per liter of *E. coli* cells, which remains consistent with the amount we obtained. The peak 4 corresponds to the diaphorase subunit (NADH:O_2 oxidoreductase), FdsBG, was quantified and only 0.7 mg of this conformation was obtained from 1 L of *E. coli* cell culture.

2. FDH from *Cupriavidus necator*

For the production and purification of CnFDH, cultures of 500 mL were performed. The same purification protocol was applied. The size exclusion chromatography graph obtained along with the corresponding SDS-PAGE gel is shown in Figure 64.

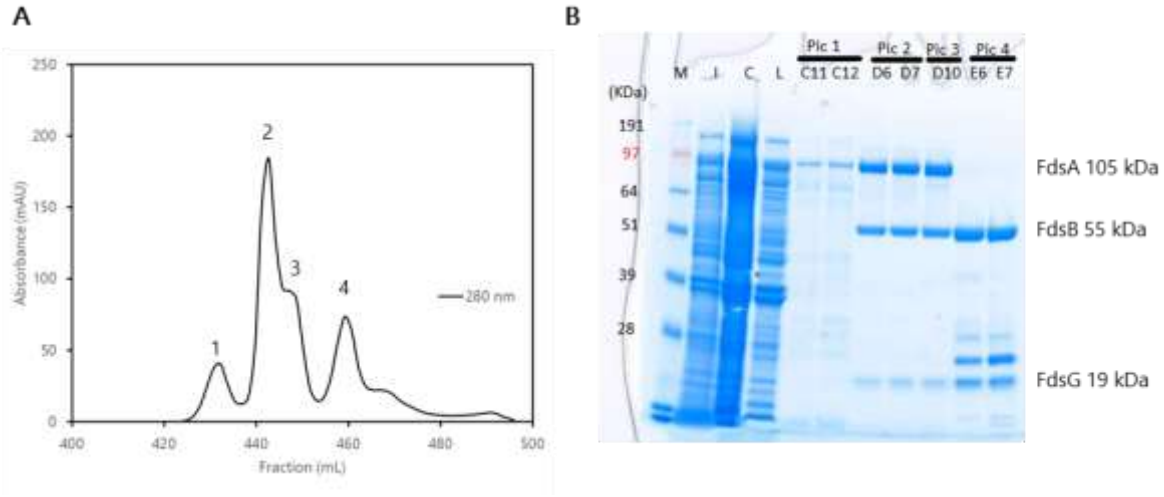


Figure 64. Size exclusion chromatography of CnFDH. (A) The black curve is the absorbance at 280 nm after column elution. (B) SDS-PAGE gel analysis. M: ladder of protein of known molecular weights. I: induction. C: pellet. L: cell lysate. Wells "Pic 1" to "Pic 4": proteins from collected fractions corresponding to the elution graph (A).

As shown in Figure 64, four peaks were obtained after gel filtration column with a slightly different elution profile compared to the one obtained for RcFDH. A sample from each fraction is then analyzed on an SDS-PAGE gel. Peaks 2 and 3 have the same profile on SDS-PAGE, which corresponds to the one expected from this purification, showing the three subunits of the FDH complex, FdsA, FdsB and FdsG. Similar to our observation with RcFDH production, we noted the production of a dimeric structure composed of FdsB and FdsG subunits (Peak 4) as was already seen for CnFDH Yu et al. (2019).

In general, the elution profile obtained during the purification of the CnFDH complex is consistent with what has been reported for heterologous expression of CnHyd in DH5α *E. coli* cells in the literature, as shown in Figure 65 below (Yu et al. 2019).

Finally, the elution fractions obtained from peaks 2 and 3 were pooled together, and the total protein content was determined using the Bradford method. From a 1 L cell culture, a total of 2.2 mg of protein

exhibiting the appropriate SDS-PAGE profile was obtained. Notably, the publication by Yu et al. (2019) was followed as a reference for the protein purification protocol but did not provide specific information regarding the total amount of protein obtained in their study. Additionally, this purification process resulted in an almost 4-fold lower yield compared to the purification of the FDH from *R. capsulatus*. The lack of available data in the literature makes it challenging to determine whether the plasmid construction is responsible for this difference or if the protein inherently yields lower amount during purification.

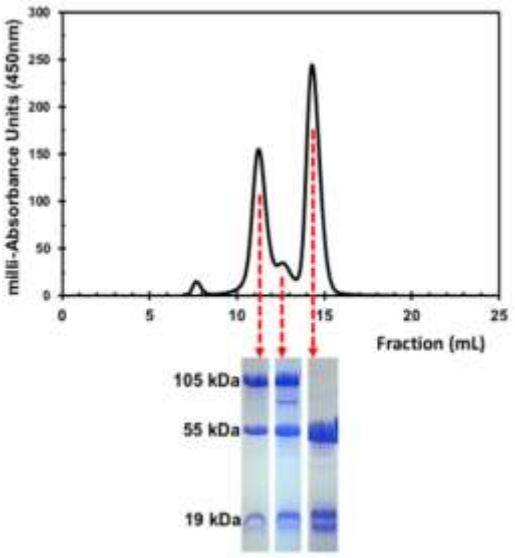


Figure 65. Size exclusion chromatography purification and SDS-PAGE gel analysis of pTRC-HisTagCnHyd obtained by Yu et al. (2019).

4. *In vitro* activity test

The *in vitro* activity test for formate oxidation was conducted by following the change in absorption at 340 nm from NADH formation. These tests were made to verify the quality of the enzymes produced and the consistency with the kinetic parameters available in the literature before employing the batches by Luca Albertin's laboratory for electrode experiments.

1. RcFDH

Typical enzyme assays were performed following the protocol described in the Material and methods section, the results are detailed in the Table 40.

Table 40. RcFDH reaction rates with formate as substrate and comparison with literature values. The activity was determined using 1 mM NAD^+, 3 mM formate and 0.1 mM FMN, if mentioned, in 50 mM Tris buffer pH 9.0 in 120 µL total reaction volume. Each enzyme assay was performed five times at 25 °C

with 5 µL of purified enzyme for assays without FMN and with 2 µL of purified enzyme if FMN was added to the reaction mix due to a higher activity detected.

Condition	Specific activity (s^{-1}) from this study	k_{cat} (s^{-1}) from Hartmann and Leimkühler (2013)
RcFDH activity (without FMN)	10.9 ± 4.0	36.5 ± 0.6
RcFDH activity (with FMN)	36.4 ± 6.4	-

The main objective of these *in vitro* tests was to assess the quality of the batch produced, thus we did not perform a complete kinetic study for k_{cat} and K_m determination. For RcFDH, $K_{m\ Formate}$ and $K_{m\ NAD+}$ were determined to be 281 ± 10 µM and 173 ± 19 µM respectively (Hartmann and Leimkühler 2013). Based on the literature and the highly similar profiles obtained from the purification step, it was hypothesized that the specific activity calculated for our RcFDH purification could be compared to the k_{cat} published using saturating substrate concentrations.

The addition of FMN to the reaction mixture significantly enhanced the enzyme activity, resulting in a three-fold improvement. This observation suggests that a portion of the RcFDH produced may have lacked the FMN cofactor bound to FdsB, making them inactive. The k_{cat} value reported by Hartmann and Leimkühler (2013) was 36.5 ± 0.6 s^{-1}, and this measurement was conducted without the addition of FMN in the reaction mix and was not corrected by the FMN saturation measured and determined to be 68 %.

A dosage by mass spectrometry of the FMN content on a precedent batch of RcFDH purification that was showing a lower specific activity revealed an FMN content of 10 % (see Appendix 12: FMN and Mo content from FDH production). This suggests that the FMN saturation of the enzyme was not achieved, thus corroborating the better results obtained with the addition of this cofactor to the reaction mix. A more extensive study of the culture conditions and protein purification steps should be performed to improve the proportion of FDH containing the FMN cofactor.

Reduction spectra of the enzyme using formate and dithionite were conducted to determine graphically the molybdenum saturation of RcFDH (see Appendix 12: FMN and Mo content from FDH production). In our analysis, we found that the protein was only 24% saturated with Mo, which is significantly lower than the 41% reported in the reference publication. These data indicate that the low FMN content of the purified protein is not the only factor limiting activity, and may explain, at least in part, why the addition of FMN to a 10% FMN-saturated enzyme did not result in a 10-fold increase in activity. The measured Mo and FMN saturation levels of the RcFDH we produced are lower than the one reported in the literature, which helps to explain the lower specific activity measured during *in vitro* tests.

The discrepancies observed between the reported k_{cat} for this enzyme and our results, 36.5 s^{-1} and 10.9 s^{-1} (activity without FMN) respectively, could be attributed to variations in the elution profiles obtained from the size exclusion chromatography. Despite these differences, the k_{cat} value we obtained has the same order of magnitude as the reported value, indicating that the enzyme quality is validated. Consequently, this batch was sent over to the LCBM team (CEA Grenoble).

2. CnFDH

Typical enzyme assays were performed at 25 °C using 1 µL of purified enzyme along with NAD^+ (1 mM), formate (3 mM) and Tris buffer pH 9.0, in 120 µL total reaction volume. As FMN is reported as a cofactor of this enzyme, and showed an improvement of the activity for RcFDH, we also performed activity assays with the addition of FMN in the reaction mix at 0.1 mM. Table 41 below shows the results.

Table 41. CnFDH reaction rates with formate as substrate and comparison with literature values. The activity was determined using 1 mM NAD^+, 3 mM formate and 0.1 mM FMN, if mentioned, in 50 mM Tris buffer pH 9.0 in 120 µL total reaction volume. Each enzymatic assay was performed two times at 25 °C with 1 µL of purified enzyme.

Condition	Specific activity (s^{-1}) from this study	k_{cat} (s^{-1}) from Yu et al. (2019)
CnFDH activity (without FMN)	94.8 ± 7.3	99
CnFDH activity (with FMN)	87.3 ± 2.3	-

The k_{cat} value reported by Yu et al. (2019) for heterologous expression of CnFDH in *E. coli* is 99 s^{-1}. Thus, the results we obtained are consistent with the results reported in the literature. Unlike RcFDH, the addition of FMN did not increase the activity of the enzyme, which suggests that each protein contains this essential cofactor. The k_{cat} value reported by Niks et al. (2016) from a purification of the native enzyme isolated from *C. necator* was 201 s^{-1}. The heterologous expression of the CnFDH shows an activity of 50 % compared to the native enzyme.

As the results for *in vitro* activity of CnFDH were consistent with the literature, this batch was sent to the LCBM team (CEA Grenoble) for further usage.

5. Discussion and perspectives of the BECOOL project

As elaborated in this section, we were able to produce efficiently active CnFDH and RcFDH with similar activity rates compared to the values reported in the literature.

From the LCBM team perspective, the first attempts to detect CO_2 reduction (either using bubbled CO_2 in the electrolyte or using HCO_3^-) from RcFDH immobilized in the hydrogel did not demonstrate such activity. This may be attributed to a lack of electrically connected enzyme to the electrode. Further tests are planned, including the use of redox mediators with a more negative reduction potential for CO_2 reduction. Additionally, another approach being considered involves the use of a truncated FDH version. It has already been reported that a truncated form of the FdsA subunit (approximately 80 kDa compared to the WT version of 105 kDa) retained some activity, ~ 12 % compared to WT FDH complex, for formate oxidation when the bis-MGD cofactor was present (Walker et al. 2019). This truncated version was not described for CO_2 reduction but it could be tested on the hydrogel to assess its ability to be immobilized in the hydrogel efficiently and its potential for CO_2 reduction.

If the project were to be continued, it would be interesting to produce more batches of both FDH to try to improve the yield and supply the LCBM with highly efficient FDH proteins. In addition, a systematic analysis of the FMN and the Mo content of the productions should be performed to better understand the influence of the culture conditions on those two parameters. Indeed, FMN and Mo are essential cofactors for FDH activity.

5. Method development & improvement

During this PhD work, an important part of the research consisted of method development and improvement due to the complexity of the manipulated heterologous enzymes and strains. Part of the work that was carried out on these aspects is detailed in this section.

The first part will describe the effort put into the implementation of an *in vivo* mutagenesis method, known as Evolution.T7. This method was developed to introduce mutations into a specific targeted region for the discovery of improved mutants towards an activity of interest. The second part will regroup a wide range of tools that were introduced in the laboratory throughout the doctoral work. These tools were necessary for overcoming technical issues and achieving research objectives.

1. In vivo mutagenesis

As discussed in the introduction (I. 6), an *in vivo* mutagenesis method, Evolution.T7, was explored during this PhD project. Dr. Ioana Popescu who developed the method in the frame of the iGEM competition 2021, provided us with the genetic material necessary (plasmids). This project was partly carried out by an M2 intern, Afra Gaaden, under my supervision.

1. Genetic construction and preliminary tests

1. *Plasmid construction with TsFDH*

As a target for the implementation of this method in the laboratory, we chose a "simple" FDH encoded by a single gene of 1.2 kbp, the FDH from *Thiobacillus sp*. Implementing this type of methodology by targeting a complex and multimeric enzyme like CnFDH, the main object of the PhD work, would have posed significant challenges outside of the method itself.

The methodology relies on the mutating activity of two T7RNA polymerase-nucleobase deaminase fusion complexes. Each polymerase recognizes a different promoter, which when placed at the 5' and 3' ends of the gene to be mutated, enables transcription and thus mutagenesis of both DNA strands. The two mutator-fusions are expressed from two compatible plasmids upon induction (gift from I. Popescu).

The plasmid containing the target gene of the *in vivo* mutagenesis is the third component of the system, it is represented in Figure 66.

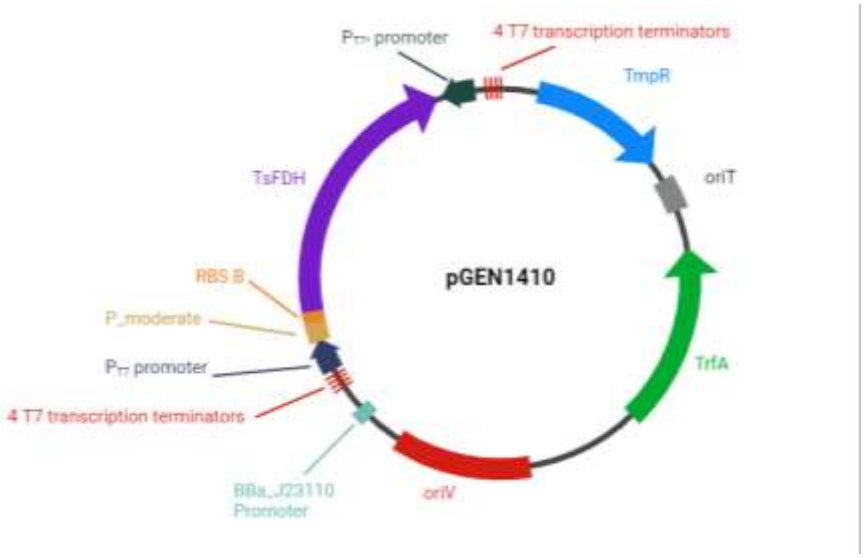


Figure 66. Map of the plasmid for *in vivo* mutagenesis (pGEN1410). The entire plasmid is 5268 bp long. The four terminators are placed in the same order, as follows, on each side of the region to evolve: BBa_B0015, SBa_000587, T7 and SBa_000451 (DNA sequences are available in Appendix 13: *In vivo* mutagenesis Evolution.T7 method). Designed on Biorender.

The plasmid used for this method has a pSEVA721 backbone (Standard European Vector Architecture), featuring an RK2 replication system consisting of the replication origin oriV and the *trfA* gene. The TrfA protein binds to a specific sequence of the replication origin to initiate replication (Toukdarian, Helinski, and Perri 1996). Hence, the plasmid maintains a very low copy number, typically from 1 to 3 copies in the cells (iGEM 2021 Evry Paris-Saclay). This low copy number is an important feature of the plasmid as it helps to limit the diversity of mutated plasmid within a single cell, but it also makes it more challenging to manipulate. Golden Gate cloning was applied to insert the TsFDH coding region into this plasmid along with a promoter and an RBS sequence. The primary difficulty of this cloning was provoked by the presence of four T7 transcription terminators in a row at each side of the gene, forming blocs of repetitive sequences. The detailed sequences of the plasmid elements are available in Appendix 13: *In vivo* mutagenesis Evolution.T7 method.

This plasmid was tested in a G5416 Δ*lpd* strain to verify the functionality of the TsFDH in this plasmid context. The test showed that the pGEN1410 construct successfully complemented the energy-auxotrophy of the metabolic screen strain using formate as energy source.

2. *Test of the system efficiency by sequencing*

The implementation of this new method in the lab required preliminary experiments of protocol validation. As such, the first focus was the functional testing of the *in vivo* mutagenesis on our targeted gene (TsFDH) using the above-described plasmid construction (pGEN1410).

1) Strain & plasmids used

For those preliminary tests, the recipient strain used was supplied by Ioana Popescu, strain E491 *E. coli* MG1655 Δ*flu* Δ*pyrF* Δ*ung* Δ*nfi*. As it was described in the introduction (I. 6), using a strain lacking two DNA reparation systems (*ung* and *nfi* genes) improves greatly the efficiency of deaminase-mediated mutation methods, such as Evolution.T7. Briefly, the *ung* gene encodes a uracil DNA N-glycosylase, responsible for uracil repair in DNA. The *ung* deletion was proved to be essential for a similar system, with the ung^+ strain harboring less than 1 % of the mutations detected in Δ*ung* strains (Álvarez et al. 2020). The other reparation enzyme, encoded by *nfi*, is the endonuclease V of *E. coli*, which plays a role in eliminating inosines. Both uracil and inosine are nucleobases generated from the activity of cytosine and adenine deaminases, respectively, which are the mutator domains in the Evolution.T7 method. The two other deletions, *flu* and *pyrF* genes were not of importance for this experiment.

The iGEM team conducted an extensive survey in which they tested 64 combinations of different T7RNAP/deaminases complexes, using deaminases that were previously described in the literature. From the results obtained in the iGEM project (iGEM 2021 Evry Paris-Saclay), five pairs of T7RNAP/deaminases demonstrated higher efficiencies and were consequently selected for our experiments. Detailed information regarding the composition of these five pairs is provided in Appendix 13: *In vivo* mutagenesis Evolution.T7 method.

These five selected pairs of T7RNAP/deaminase complexes were transformed in E491 (Δ*flu* Δ*pyrF* Δ*ung* Δ*nfi)* electrocompetent cells along with the plasmid harboring the TsFDH (pGEN1410). The clones that incorporated the three plasmids are selected on LB plates supplemented with the three antibiotics for which resistance cassettes reside on the plasmids: kanamycin (12.5 µg/ml), spectinomycin (12.5 µg/ml) and trimethoprim (5 µg/ml). The strong selection with three antibiotics led to a reduction of the concentration usually used for these. The five resulting strains were named G6165, G6178, G6219, G6221 and G6223 each carrying a pair of mutators (see details of these pairs in Appendix 13: *In vivo* mutagenesis Evolution.T7 method).

2) Mutagenesis protocol

Mutagenesis initiates upon induction of the expression of the two T7RNAP/deaminase mutator complexes; wild-type T7RNAP is placed under an anhydrotetracycline (aTc) inducible promoter, and T7RNAP* (CGG-R12-KIRV variant) is placed under an L-arabinose inducible promoter. The flow sheet detailing the experimental steps is depicted in Figure 67.

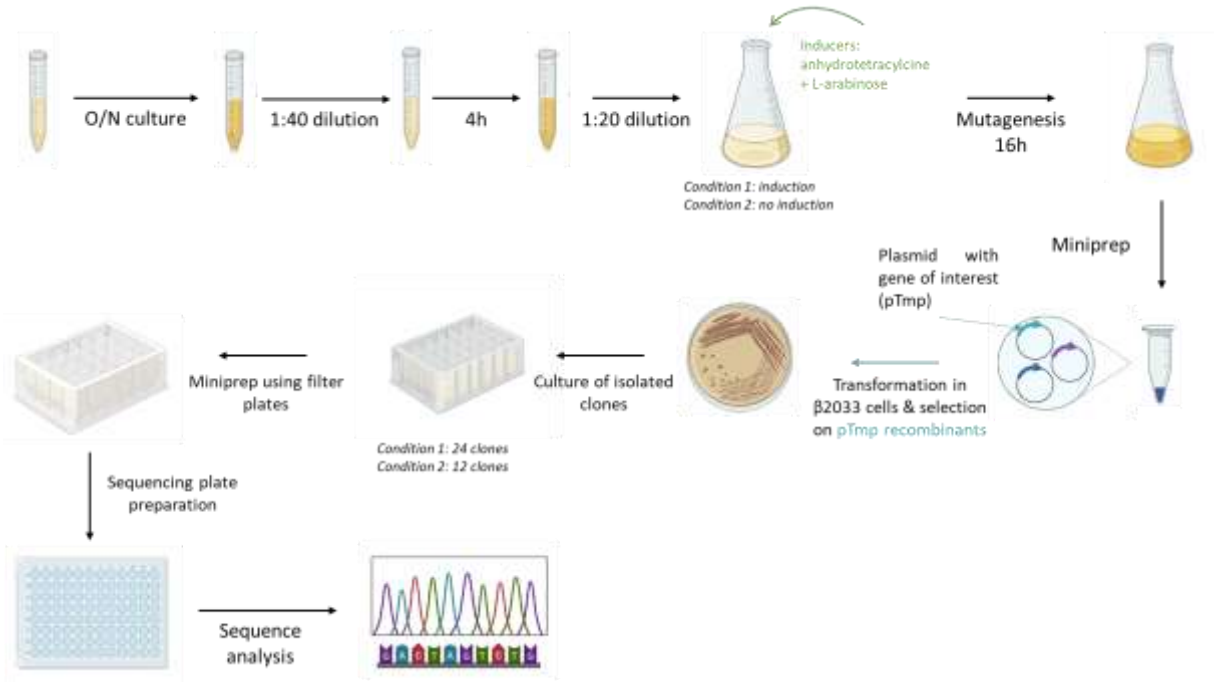


Figure 67. Experimental flow sheet of the functional testing of the Evolution.T7 system. The L-arabinose inducer was used at 1.5 mM and aTc was used at 200 ng/ml final concentrations.

In brief, strains carrying the three plasmids necessary are cultured and diluted to obtain early log phase growth. Then, mutagenesis is induced with L-arabinose and aTc added to the culture at the same time, the cultures are incubated for 16 hours and plasmids are purified from 5 mL of each culture (miniprep using QIAGEN kit). In parallel, a culture with no inducers is performed as a negative control. The plasmid minipreps obtained from these cultures are a mix of the three plasmids that the strains contained. To select only the plasmids that were targeted for mutagenesis containing the TsFDH gene, a fraction of these minipreps are transformed in β2033 electrocompetent cells and clones are selected on LB plates supplemented with trimethoprim (10 µg/ml) at 30 °C. Colonies are piqued to inoculate cultures in 96-well plates and minipreps using these cultures are prepared with filter plates. Subsequently, the region to evolve within the plasmid pGEN1410, *i.e.* the TsFDH coding region, is sequenced using Sanger

sequencing. Finally, the sequences obtained are mapped onto the reference sequence and analyzed to determine the number of mutations in the target gene for each isolated clone.

For each strain, 24 colonies are selected for the condition with induction and 12 colonies are piqued for the negative control condition. The negative control experiment will provide important information about the mutations that can effectively be attributed to the mutagenesis system and not to spontaneous mutation appearance, if the assumption is made that the promoters used are strict on/off systems.

3) Results from the preliminary experiments

The average number of mutations per gene was determined by analysis of the Sanger sequencing data, and this analysis was conducted using JalView and Snapgene softwares. Due to certain experimental constraints, it was not possible to read the sequences of all 24 clones for each mutagenesis experiment. Most problematic, was the plasmid extractions in 96-well plates, since the volume of the culture was restricted to 1 mL, which is insufficient for preparing minipreps of low-copy plasmids with high yield. Nevertheless, clones could be successfully analyzed, the results are presented below (Table 42).

Table 42. Number and type of mutations per strain and per gene in induced conditions. When the number of readable clones has digits, it means that some clones were only partially readable.

Strain	Readable clones	Detected mutations			Average number of mutations per gene (1.2 kbp)
		T ↔ C	A ↔ G	Total	
Strain 1 (G6165)	21/24	42	21	63	3
Strain 2 (G6178)	15.5/24	10	22	32	2.1
Strain 3 (G6219)	*4.2/24*	*33*	*21*	*54*	*12.9*
Strain 4 (G6221)	22.5/24	7	17	27	1.2
Strain 5 (G6223)	14/24	6	9	15	1.1

Several important observations can be made from Table 42. Firstly, the data obtained for strain 3 could not be exploited due to the very limited number of readable clones, in addition to the low quality of the sequences obtained that artificially increased the number of mutations detected. However, for the other four constructs, the number of readable clones was sufficient for significant comparisons between strains. Notably, the first two pairs of T7RNAP/deaminase complexes provided the highest number of mutations. An average of 3 and 2.1 mutations per gene was considered the most favorable results since it represented a significant impact on the protein without introducing an excessively high number of

mutations that might cause extensive protein damage. In addition, all the mutations that were detected are transitions mutations (T↔C or A↔G) which is consistent with the activity of the cytosine and adenosine deaminases (see I. 6. 3. 2 from the introduction chapter).

The results obtained for the non-induced condition exhibited poor quality, resulting in a limited number of readable clones out of the 12 sequenced clones. However, among the clones that could be analyzed, the average number of mutations detected was less than one mutation per gene (1.2 kbp). Consequently, it was deduced that the mutations observed in the induced condition were primarily due to the activity of the mutagenic system.

Finally, an analysis of the distribution of the mutations over the entire length of the TsFDH gene was conducted. The results obtained from six clones obtained with "strain 1" (G6165) are available in Figure 68 for illustration. It was observed that no specific mutation hotspots were detected, and mutations were distributed along the entire length of the targeted DNA sequence. However, a slight accumulation of mutations in the 5' and 3' extremities of the gene was noted. This observation could be partly attributed to transcriptional abortion, a phenomenon reported for all RNA polymerases, resulting in very short RNA transcripts from to 2 to 8 nucleotides. The transcription complex tends to be more stable from the 9th nucleotide on (Kochetkov, Rusakova, and Tunitskaya (1998); Severinov (2001); Dousis et al. (2023)). Our observation shows accumulation within larger DNA length than these very short abortive RNA molecules. Another possible explanation is the collision between the two complexes moving in both directions along the DNA sequence, as both complexes were induced at the same time.

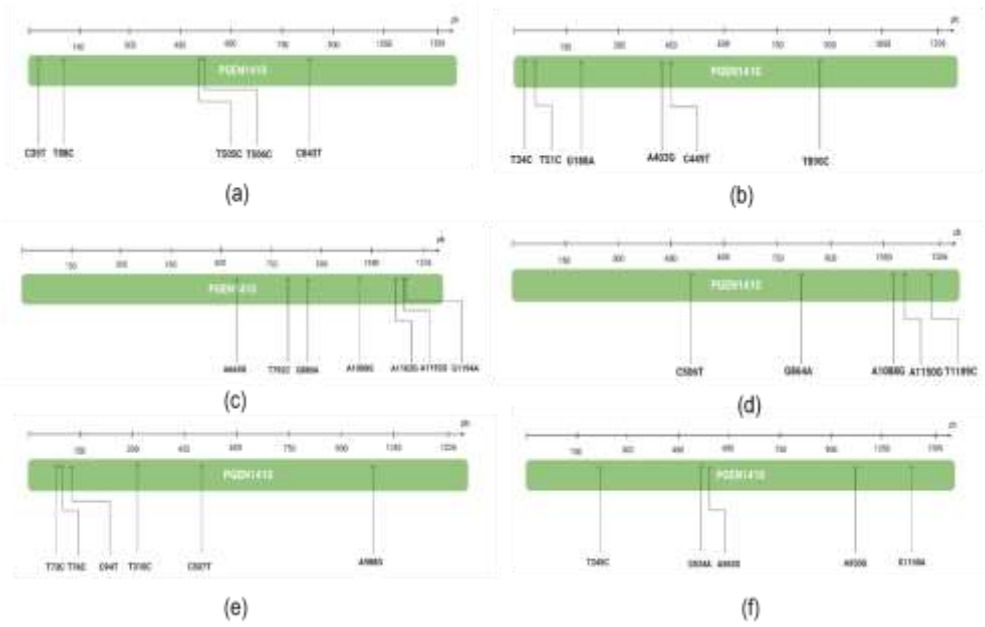


Figure 68. *TsFDH* gene mutations distribution of six sequenced clones obtained after mutagenesis induction with strain 1 (G6165, Table 42). (a) clone 12, (b) clone 14, (c) clone17, (d) clone 19, (e) clone 23, (f) clone 24.

In conclusion, considering the results obtained in this experiment, it was decided to focus on the first pair of the five T7RNAP/deaminase complexes tested. This complex pair consists of evoAPOBEC1-BE4max deaminase fused with T7RNAP* and ABE8.20-m deaminase fused with wild-type T7RNAP.

3. *Strain construction for metabolic screening*

As the targeted gene codes for a formate dehydrogenase (TsFDH), the easiest way to select improved variants of this enzyme generated by *in vivo* mutagenesis is by screening enhanced NADH-generation activity through formate oxidation. To combine mutagenesis and selection, the strain must harbor the plasmids of the mutagenesis system, have the mutator phenotype (Δ*ung*, Δ*nfi*) and carry the *lpd* deletion. To construct this strain, strain G5416 (MG1655 Δ*lpd*) was reused for further genetic modifications. A P1 phage preparation was obtained from strain JW5547 (Δ*nfi::kan*) of the Keio collection and used to transduce the G5416 strain, resulting in the strain G6420 Δ*lpd* Δ*nfi::kan*. The resistance cassette was excised using the FLP-recombinase system. Subsequently, a P1 phage preparation was made using the strain JW2564 (Δ*ung::kan*) from the Keio collection. The double mutant Δ*lpd* Δ*nfi* strain was transduced with this latter P1 phages and the kanamycin resistance cassette was removed resulting in the final strain G6476 Δ*lpd* Δ*nfi* Δ*ung.*

2. Development of a selection protocol

The next step of the implementation of the Evolution.T7 method was the development of a selection protocol targeting the enzymatic activity involved in metabolism. This was a challenging task given the fact that these *in vivo* methods were mainly validated for the acquisition of new antibiotic resistance profiles. In our case, descendants of the metabolic screening strain G6476 (Δ*lpd* Δ*nfi* Δ*ung*) were used for the development of an *in vivo* mutagenesis/selection protocol. The two plasmids containing the selected T7RNAP/deaminase pair 1 (evoAPOBEC1-BE4max-T7RNAP* and ABE8.20-m-T7RNAP) were transformed in this strain, along with the pGEN1410 plasmid carrying the gene to be evolved, TsFDH, resulting in the strain G6511. The detailed steps of the protocol are represented in Figure 69.

The objective of this test was to evaluate the impact of *in vivo* mutagenesis applied to the TsFDH gene on the activity of this enzyme as judged by a loss of function on formate, revealed by plating on selective medium and cell counting. Through parallel plating on permissive medium, possible growth effects of off-target mutations can be subtracted. To realize this test, a pre-culture is performed and diluted to reach early log-phase before induction of the mutagenesis system. Two parallel cultures were made, one with the inducers (aTc and L-arabinose) and the other without. Three induction times, 16, 24 and 40 hours, were tested to analyze the influence of this parameter on the mutagenesis. These cultures were then adapted to a minimal medium under permissive conditions, *i.e.* cell growth did not depend on the

activity of the TsFDH gene. Finally, dilutions of these cultures were plated on two types of plates: one plate with permissive medium and the other one with test medium (medium composition detailed in Figure 69), where only the cells harboring a functional TsFDH gene should survive.

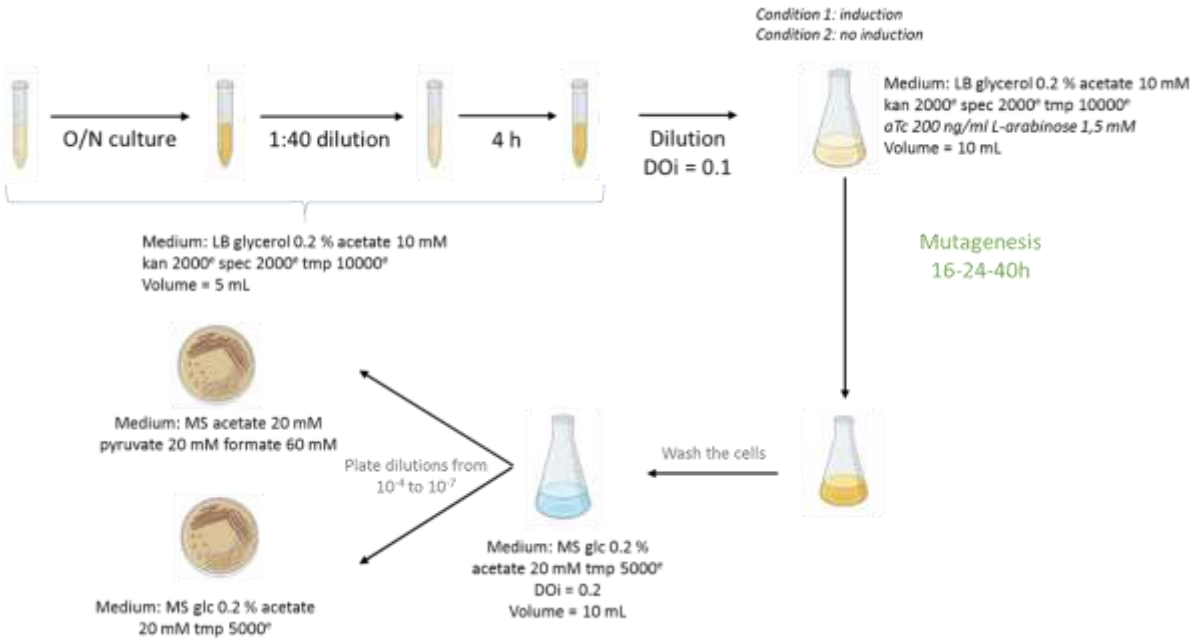


Figure 69. Experimental flow sheet of the mutagenesis targeting TsFDH gene in metabolic screening conditions. The TsFDH is carried by pGEN1410 plasmid and the mutagenesis is realized with pair 1 of T7RNAP/deaminase complexes (evoAPOBEC1-BE4max-T7RNAP* and ABE8.20-m-T7RNAP) using G6511 strain.

Following this protocol the plates were incubated at 30 °C. After 3 days, the colonies on permissive condition plates had grown enough for counting, for test conditions, the counting was performed after 6 days of incubation. The results are summarized in Table 43 below.

Table 43. Results of the mutagenesis experiment on TsFDH followed by metabolic selection. Permissive medium consisted of MS mineral medium supplemented with glucose (0.2 %), acetate (20 mM) and trimethoprim (10 µg/ml). Test medium consisted of MS mineral medium supplemented with acetate (20 mM), pyruvate (20 mM) and formate (60 mM). The star signs mean that the plates were contaminated.

Induction time		**16 h**		**24 h**		**40 h**	
Type of culture		**Induced**	**Non-induced**	**Induced**	**Non-induced**	**Induced**	**Non-induced**
OD_{600} of the culture in rich medium after mutagenesis incubation		0.864	2.032	1.104	2.45	1.43	2.67
OD_{600} after 24 hours of incubation in minimal medium		0.995	1.075	1.190	1.345	0.702	1.036
Counting of permissive plates (3 days of incubation)	**10^{-5}**	95	580	85	532	52	408
	10^{-6}	20	85	12	47	4	53
Counting of test plates (6 days of incubation)	**10^{-5}**	5	420	2	360	*	456
	10^{-6}	1	52	0	30	*	16

From Table 43, the first observation that can be made is that independently of the induction time or the type of plate, there are constantly fewer clones with the induced cultures. While this can be partly attributed to the lower OD_{600} reached by the induced culture, other factors must be implicated. For instance, in the 16 h-induction time condition, very close final OD_{600} were reached upon adaptation in minimal medium for both induced and non-induced cultures, 0.995 and 1.075 respectively, but a factor of approximately 5 is found in favor of the colony count on permissive plates for non-induced cultures. This could be due to the off-target activity of the mutagenesis system or the burden of the system that impedes a higher survival ratio when cells are plated.

The other observation that can be made is that the longer the induction is, the lower the number of survivors is as well. This is probably because of the cell death that started in the culture in rich medium and when washed for incubation in mineral medium, part of the cells were probably already dead. For this reason, the condition of 40 hours of induction can be considered as inappropriate. In addition, the colonies obtained especially for induced conditions, either on permissive or test plates showed high heterogeneity in colony shape, pointing to the toxicity of the mutagenesis system (see Appendix 14: Results of the Evolution.T7 method targeting TsFDH).

Most importantly, while the number of clones counted on permissive plates under induced and non-induced conditions consistently exhibits lower survival rates for the induced condition, this difference is much larger when comparing the test plate's condition. For instance, in the 16 h-induction time condition (counting with 10^{-5} dilution), 5 clones were obtained for the induced condition while 420 were counted with the non-induced culture when plated on test medium, corresponding to a multiplication factor of 84. For permissive plates, the factor was only 5, indicating a clear influence of the mutagenesis system on the activity of the targeted gene coding for TsFDH required for growth in the metabolic screening test. These results are a first demonstration of the functionality of the mutagenesis system in Δ*pd*-derivative selection strain.

3. Discussion and perspectives

The results presented in this section are preliminary and a lot is still to explore.

We were able to demonstrate the activity of the mutagenesis system on the plasmid pGEN1410 carrying the targeted gene to evolve, thus validating this construction and the overall experimental setup. On this aspect, it would be highly interesting to conduct more extensive parameters analysis such as induction time and perform sequential induction of the two mutagenic complexes. In addition, the use of high-throughput sequencing would both facilitate the interpretation of the results and give more accurate results, especially in terms of mutation rates. This approach was used by the iGEM team for the development of Evolution.T7.

Regarding the metabolic test, a detailed analysis of the clones growing on selection plates is required to confirm the presence of mutated TsFDH still capable of formate oxidation. If so, the characterization of these mutants compared to the WT allele would be interesting to determine if a more favorable mutant was selected. This characterization could be performed *in vivo* upon growth tests and subsequent sub-cloning of candidate alleles in production plasmids such as pET vectors for purification and *in vitro* activity characterization (Rosano and Ceccarelli (2014); Shilling et al. (2020)).

Furthermore, a protocol for the selection of improved variants should be developed. The first step is diversity enrichment, which can be carried out as described in this results section or further improved accordingly to the few above-mentioned parameters, *i.e.* time of induction, pairs of mutators, sequential induction. The selection step could be performed by successive dilutions of the genetically-enriched culture in test medium for the emergence and the fixation of favorable mutants.

Our results suggest that the inducible promoters for the expression of the mutator complexes are strongly regulated as very few mutations were detected in non-induced cultures. However, this type of *in vivo* mutagenesis will always involve risks of off-target mutations, outside the targeted region in the

chromosome for example, that could impair the metabolic selection or the metabolism in general. This highly important aspect should be explored to ensure a tight-regulated system. This could be done by genomic sequencing but would rather be expensive and require extensive data analysis. Alternatively, as was performed by Álvarez et al. (2020), a reporter gene could be used, such as *rpoB* to assess this question. *rpoB* is a gene from *E. coli* that is known to be mutated in rifampicin-resistant colonies. By plating induced and non-induced cultures on rifampicin, plates would give an off-target rate of mutant frequency by comparing spontaneous mutants resulting from the intrinsic mutation rate of *E. coli* compared to a culture where the mutagenesis system was induced.

For long-term perspective, the mutagenesis system could be integrated in the genome to stabilize it and to avoid the burden of having three plasmids co-existing within the cell. In addition, this stabilization would allow the use of the system coupled with a GM3-mediated selection alternating automated mutagenesis and selection phases, for example. In such context, the off-target mutation frequency should be carefully monitored as these cycles would necessarily increase these undesired mutations and are combined with strains lacking DNA damage repairing systems. The most active deaminases have been reported to increase by 20 to 30-fold the number of mutations outside the target region compared to the "natural" mutation rate in *E. coli* cells (Álvarez et al. 2020).

2. Implementation of diverse techniques in the laboratory

1. Cloning techniques

The main enzyme used in this study, the formate dehydrogenase from *Cupriavidus necator*, is encoded by five genes in an operon spanning 6.1 kbp and possesses a high GC content of 67 %. This represented a challenge at every step of the project. Due to this size, every plasmid constructed with the operon represented a substantial genetic construct. For example, the plasmid used for genomic integration, pDM4_IS10, is 10 kbp long, resulting in a construction of more than 16 kbp when the CnFDH was cloned inside. In addition, the difficulties for cloning also pointed out the toxicity of such enzymatic complexes when overexpressed in *E. coli*.

Consequently, with many failed trials applying regular techniques of restriction-ligation cloning, the necessity of using higher efficiency cloning techniques imposed itself. In addition, restriction-ligation cloning offers very limited possibilities of plasmid rearrangements, unless a standardized library of vectors is available. Finally, the amplification of the native operon itself represented a challenge and the use of a touchdown PCR was necessary in some cases.

Among all the techniques tested, the use of Gibson cloning showed the best performance and versatility. However, it required the amplification of the entire backbone to insert the operon at a specific locus in the vector. Indeed, as most of the vectors available were not standardized, no restriction enzyme could be used to prevent the entire amplification of the vector or to clone using Golden Gate cloning. In the future, the implementation of a set of standardized vectors with different levels of expression, copy numbers, types of expression (inducible/constitutive), Golden Gate cloning sites or well-designed restriction enzymes for Gibson cloning that would not necessitate the amplification of the entire vector, could be beneficial.

The success of ambitious projects is significantly dependent on the technical ease of manipulating strains and plasmids, allowing to test various parameters to achieve fine-tuning expression adapted to the objectives, which was demonstrated, for example, by the previously cited work by Chen et al. (2020) for the implementation of synthetic methylotrophy.

This PhD project was the opportunity to explore a wide range of techniques for PCR, cloning design and techniques that were not previously commonly used in the laboratory. This also pointed out the limitations of some systems, especially for genome editing in general and chromosomic integration in particular, which was the opportunity to develop other tools. A tool for genome integration was implemented and is described in the following paragraph.

2. Chromosomal integration using CRISPR-Cas9 selection

1. *Design and proof of concept*

Genome integration of CnFDH represented one of the most important technical challenges in this PhD work. As previously explained, the total plasmid size was approximately 16 kbp when the pDM4_IS10 construct was used, and each step from cloning to integration into the chromosome came with difficulties. Nevertheless, it was successfully accomplished but it became evident that an alternative method was needed. Consequently, we implemented a method based on the one developed by Vigouroux et al. (2018), in David Bikard's laboratory at Pasteur Institute. The method relies on homology recombination and CRISPR/Cas9-mediated counter-selection to select the cells carrying the heterologous gene. It was briefly explained in the material and methods chapter (II. 4. 6. 2) but will be described in detail in the paragraphs below.

The first step of this method development was the construction of a plasmid with homology arms (1 kbp upstream and downstream) for the insertion locus selected, IS10 (ula operon), in a pSW23 backbone, which has a R6K origin that can only be replicated by Pir strains. This IS10 locus was selected since it proved to be efficient from the previous experiments described in this study. The plasmid needed to

include a strong promoter and RBS for efficient expression of the heterologous gene, which would be in a single copy once inserted into the chromosome. The promoter and RBS from the pDM4_IS10 showed efficient expression previously, and thus were chosen for this method. As a proof of concept, a simple gene, FDH from *Thiobacillus sp.*, was initially selected to ensure the efficiency of this method. This construction resulted in a five-part cloning performed by Gibson cloning, using an ST18 strain. The cloning was successful and the resulting plasmid pre-validated by Sanger sequencing was then fully sequenced to confirm the integrity of every part of this construction, resulting in the plasmid pGEN1437 (represented Figure 70).

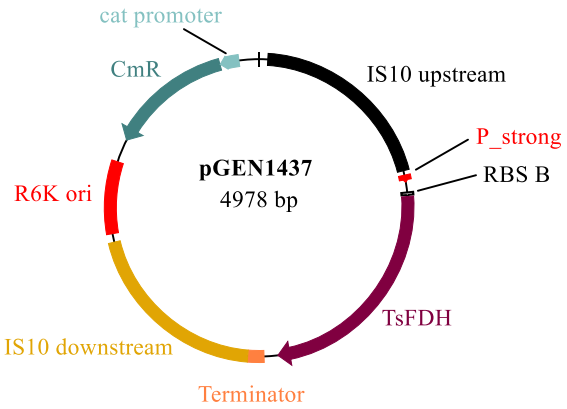


Figure 70. Map of pSW23_IS10::TsFDH (pGEN1437) plasmid for chromosomic integration.

Following the validation of this construction, the plasmid pGEN1437 was transformed into WT MG1655 electrocompetent cells, and clones were selected on LB plates supplemented with chloramphenicol. With this selection, only the clones that had integrated the entire plasmid into their chromosome would be able to survive since this type of plasmid cannot replicate in *E. coli* MG1655 cells. Although homologous recombination in *E. coli* is an inefficient mechanism, the 1 kbp-long homology sequences and highly competent cells were used to increase the probability of obtaining clones. Only a small number of clones, from six to ten, were obtained, a control by colony PCR confirmed the integration of the plasmid. This PCR also allowed us to determine the orientation of the vector insertion (see Figure 71). It was observed that there was no bias in orientation, the two possibilities were equally represented. Then, following the protocol described by Vigouroux et al. (2018), one clone from each orientation was selected to proceed to the next steps.

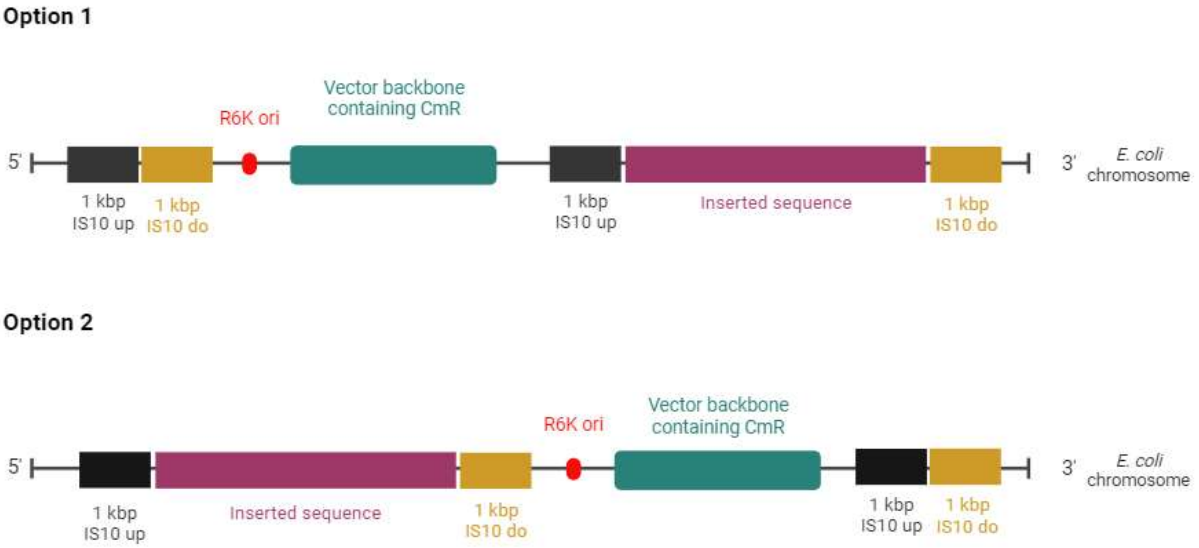


Figure 71. The two orientations possible for plasmid pGEN1437 insertion in *E. coli* chromosome.

To select for the second recombination event, *i.e.* the excision of the plasmid backbone, electrocompetent cells were prepared using the two selected clones. These cells were then transformed with the plasmid pAV10 (Figure 21), which harbors a DAPG-inducible promoter for a Cas9 gene and a gRNA targeting the chloramphenicol resistance gene of pGEN1437. The transformed cells were plated on LB dishes supplemented with DAPG (25 µg/ml) and spectinomycin (antibiotic resistance from pAV10 plasmid) and incubated at 30 °C as pAV10 replication is thermo-sensitive. This transformation showed high efficiency with approx. 100 clones obtained on selection plates.

If the second event of recombination did not occur, the cells are killed due to the double-strand break produced by the Cas9/gRNA targeting the CmR. Clones obtained were then re-streaked on plates with the same composition before being patched onto LB plates with and without chloramphenicol at 42 °C. This step was performed to both verify their sensitivity to chloramphenicol, ensure they were not escapers from the Cas9, and lose the pAV10 plasmid. Finally, the clones are PCR-verified with primers external to the insertion to prove the excision of the backbone by controlling their profile. This PCR step also ensures the presence of the inserted gene, as there is a possibility that cells could potentially excise the entire plasmid. Importantly, for all trials that were performed, no clones that retained the backbone vector (CmR) were detected after the second patch on selecting plates with the inducer, indicating the high efficiency of the gRNA. Additionally, no clones that lost the heterologous gene were detected either.

In summary, we have successfully implemented an alternative method for genomic insertion and obtained edited strains. The limited number of assays conducted does not allow for precise statistical analysis but the insertion tests conducted resulted in the selection of edited clones exclusively.

2. *Further experiments with this construction*

For the transfer of the heterologous gene insertion, two possibilities are available. Either the first recombination event is performed in MG1655 WT cells and phages are prepared using this edited strain to transduce the target strain and the second event of recombination is selected in this latter strain. This could allow the editing of strains that are difficult to handle as the first event of recombination is the one with the lowest efficiency. This possibility was confirmed by using a clone from the precedent experiment after the first event of recombination to prepare phages and to transduce the G5416 (Δ*pd*) strain and the evolved G5876 strain. The transduction and the subsequent transformation of the pAV10 plasmid for the second recombination event selection showed high efficiency. This experiment revealed 100 % of correctly edited clones, *i.e.* with the TsFDH gene integrated at the IS10 locus and the plasmid backbone excised from the chromosome. However, this adds a time-consuming step to the protocol.

To further test the method and to assess its potential, the insertion of the TsFDH using the same plasmid, pGEN1437, was tested directly in an engineered strain, G6447 (described in III. 2). This editing also showed successful results.

Finally, the cloning and the integration of the CnFDH operon will be conducted as the next step for validation of this alternative method.

3. *Discussion and perspectives*

In this section, a method for genomic insertion was adapted from a publication (Vigouroux et al. 2018), and the initial results are promising. However, further testing of this method is required to validate its efficiency. Still, handling this method proved to be much more convenient compared to the one initially used with pDM4 plasmid. Furthermore, there is room for improvement and several possible adjustments are proposed here.

To facilitate the initial cloning into the pSW23 vector, a construction with Golden Gate cloning sites, along with a colorful selectable marker such as mCherry should be performed. Although the number of clones obtained after the first event of recombination could be increased, perhaps by using a λ-red assisted system, the obtained results were sufficient for our needs.

One important consideration that should be further assessed is the direct insertion into engineered strains rather than WT *E. coli* cells. Since the laboratory mainly works with genetically modified cells, having a system that can be directly used in these cells would represent a major asset. If the recovery efficiency of these engineered cells is not sufficiently high and direct integration is not an option, an alternative approach can be to perform the first recombination event in WT cells, preparing P1 phages with the resulting recombined cells and using these phages to transduce the engineered strain. Indeed,

transduction showed high efficiency, and the second homologous recombination event proved to be more efficient. This protocol was successfully tested for highly impaired Δ*lpd* strains (G5416 and G5876), and positive results were obtained.

Finally, to expand the possibilities of chromosomic integration locus, homology arms in the plasmid pGEN1437 can be modified to target other regions from the chromosome such as IS7 (*nupG* gene) or SS9 (intergenic region) that were already described with a high integration efficiency (Bassalo et al. 2016).

6. Global discussion

After exposing the experiments carried out during this PhD project, a more global discussion is proposed in this paragraph to put the results we obtained in perspective.

1. Focus on formate dehydrogenase activity

The project aimed at studying the potential of the multimeric O_2-tolerant formate dehydrogenase described with CO_2-reducing capabilities *in vitro* from *C. necator*.

Typically, simple NAD^+-dependent FDHs are used as energy generating systems through their formate oxidative capabilities, and their heterologous expression in *E. coli* was proved to be efficient. The study from Wenk et al. (2020), describing the Δ*lpd* energy-auxotroph strain involved the FDHs from *Pseudomonas sp*. 101 (PsFDH) and *Candida boidinii* (CbFDH) for NADH generation and is used as a reference for comparison with our results. PsFDH and CbFDH were both found to be able to support growth with generation times of 5.2 h and 15 h respectively. Their reported k_{cat} values for formate oxidation are 10 s^{-1} (Tishkov et al. 1993) and 1.08 s^{-1} (Choe et al. 2014). In their study, the comparative growth profiles obtained with formate as energy source were consistent with these *in vitro* data, with the strain expressing the PsFDH showing better growth performances. For our study, we tested CnFDH and TsFDH expressed from the same plasmid backbone as the reference study and in the same genetic background, resulting in generation times of 19 h and 12 h respectively (see Table 29). For these FDHs, the corresponding k_{cat} values reported for formate oxidation are 99 s^{-1} and 1.77 s^{-1} (see Table 2 from the introduction chapter). While the results obtained for TsFDH are in line with the cited study, the growth obtained for the CnFDH was low. This suggests that *in vitro* data provide indications but may not fully reflect what can be achieved *in vivo*.

Notably, two interesting publications exploring these aspects were published by van Eunen and collaborators (van Eunen et al. (2012); van Eunen and Bakker (2014)). These studies highlight the discrepancies in enzyme kinetics and metabolite accumulation that can result from the direct integration of *in vitro* kinetic parameters of an enzyme into mathematical models intended to represent yeast metabolism, when compared with experimental data. These inconsistencies were resolved by including allosteric regulation and new *in vitro* activity measurements conducted in conditions closer to the intracellular environment, called "*in vivo*-like". Our objective was not to build a mathematical model, yet, but the study of the latter point with "*in vivo*-like" activity measurement could give valuable insights for the continuation of our project.

2. Heterologous expression challenges

Another aspect that could be discussed is the metabolic load (also called metabolic burden) that is inherent to overexpression, whether the gene is native or heterologous. This metabolic load can divert nutrient fluxes in the cells that then become unavailable to support cellular metabolism (Cardinale and Arkin 2012). The overexpression of such a big enzymatic complex as the CnFDH, especially when expressed from a multi-copy (~15-20) plasmid such as the pTRC99a used for this study, can create a substantial metabolic load that can lower growth, which may partly explain the results obtained with the CnFDH expressed in the energy auxotroph strain. This points to the advantage of genomic integration, with only one copy of the heterologous gene, necessarily decreasing the metabolic load. It was observed in our study that competitive growth could be achieved with this single-copy format using a strong promoter and RBS.

As we observed during our experiments, *E. coli* seems to be struggling to handle the CnFDH complex but decent growth was obtained with TsFDH for formate oxidation despite its very poor kinetic properties (k_{cat} of 1.77 s^{-1}, k_{cat}/K_m, 0.11 $mM^{-1}.s^{-1}$) compared to k_{cat} of 99 s^{-1} and k_{cat}/K_m of 380.8 $mM^{-1}.s^{-1}$ for CnFDH. This demonstrates the highly efficient expression of this simple FDH in *E. coli* as heterologous host. As elaborated in the introduction, the TsFDH has the best reported CO_2 reducing capabilities among the simple FDHs. Additionally, it was demonstrated that the CO_2 reducing efficiency from simple FDH could be improved. The simple FDH from *Chaetomimum thermophilum* was engineered by constructing a library of variants for five residues in the active site. The screening of this library allowed the discovery of a variant with a catalytic efficiency for CO_2-reduction 5.4-fold higher than the WT version of the gene (Çakar et al. 2020). This suggests that there is room for improvement in these simple FDHs that justify the use of mutagenesis tools for the discovery of variants. However, the final catalytic efficiency still does not match the catalytic efficiency measured for wild-type CnFDH, it remains questionable if simple FDHs can be obtained to meet the requirements of autotrophic growth.

3. CO_2 consumption at the cell scale

A highly challenging point to assess is the quantification of net CO_2 assimilation in synthetic autotrophic or C1-assimilating strains. This aspect is crucial to determine the efficiency of a bioproduction strain to effectively participate in the carbon capture effort. As an illustrative example, the autotrophic *E. coli* strain already described in the introduction (I. 2. 4. 1) in which the Calvin cycle was integrated in combination with the expression of a formate dehydrogenase as the energy-generation system can be discussed (Gleizer et al. 2019). It is obvious and recognized by the authors that upon the employment of the Calvin cycle, notoriously exhibiting a poor energy efficiency entailing high energy demands for CO_2 fixation, and the use of formate oxidation for energy supply, the CO_2 release exceeds CO_2

consumption in this system. This could be acceptable if the formate used as the energy source was produced from CO_2 through reduction using a sustainable energy source, as mentioned in the introduction (I. 2. 2) or if another energy-generating system was employed. While exact quantification of CO_2 fluxes requires more extensive metabolic modeling, this basic approach indicates that the construction of an autotrophic strain with a net CO_2 assimilation is still an open challenge.

CO_2 fixation through carboxylation is much more widespread in nature due to more favorable thermodynamics of the reaction and participate in the overall assimilation of CO_2 by engineered strains. However, CO_2 reduction pathways are less resource-consuming thus improving the biomass yield that can be expected (Cotton, Edlich-Muth, and Bar-Even 2018), making direct CO_2 reduction an objective to pursue. In this regard, the reductive glycine pathway (I. 2. 5. 2. 1), similar to the highly efficient rTCA anaerobic pathway, is the only pathway that can compete in terms of energy conservation efficiency. Both metabolic routes start with CO_2 reduction to formate, which is subsequently reduced by involving tetrahydrofolate as a coenzyme. However, CO_2 reducing enzymes involved in the rTCA suffer from high O_2-sensitivity and/or low CO_2 reduction rates preventing them from being incorporated in aerobic biotechnological chassis without previous engineering.

IV. Conclusion and perspectives

Utilizing CO_2 or its reduced derivatives, notably formate and methanol, as feedstock for biomass production is an important objective of the bioindustry. Progress has been made in the implementation of synthetic pathways of reduced C1 compound assimilation in *E. coli*, the most versatile and easy-to-handle platform organism of bioproduction. Constructing *E. coli* strains for direct CO_2 assimilation was also undertaken by expressing the enzymes of the Calvin cycle in the cells. This approach, while being successful in incorporating CO_2 carbon in the metabolites of the cell, suffers from the low processivity and energy efficiency of the cycle.

In this PhD project, we pursued a strategy for the construction of an autotrophic *E. coli* strain which was based on the reduction of CO_2 by formate dehydrogenases. Formate-dependent strains previously obtained through the implementation of the reductive glycine pathway were available in the laboratory, representing scalable *in vivo* selection schemes for formate production.

Numerous technical difficulties were encountered throughout the realization of this project. Most were caused by the choice of the FDH we made. The complex metal containing CnFDH was a rational choice due to its tolerance for O_2 and its NAD^+-dependence which, together with a reported activity for CO_2 reduction 10-fold higher than measured for the best monomeric metal-free enzymes, made this FDH ideal for the development of an aerobic *E. coli* autotroph.

However, the *in vivo* demonstration of the CO_2-reducing activity from the CnFDH, activity that was exclusively reported *in vitro*, proved to be much more challenging than initially expected. Unfortunately, we were not able to observe such activity in the hemi-formatotrophic strains used for this project. We expected that, even in the absence of initial growth, a medium swap regime would, through fixation of adaptive mutations, enable carbon and energy fluxes to adapt serine auxotroph strains to some growth on formate and CO_2, given the low stringency (about 3 % of cell carbon from formate) of this selection. In the past, medium swap regimes were successfully employed to fix adaptive mutations in evolving populations not showing any initial growth under selective conditions (Bouzon et al. 2021). The GM3 evolution, now growing for up to 600 generations, revealed the selection to be stable with no rapid escapers selected *via* some endogenous activities from *E. coli* that produce formate as a co-product (genes *ribA* and *ribB* involved in purine biosynthesis for instance). The requirements to obtain formate complementation are much higher than expected. However, while these O_2-tolerant CO_2-reducing enzymes attract a lot of attention, to date, there is no report of their efficient *in vivo* activity in the reductive sense.

In contrast, we were able to demonstrate the *in vivo* activity of the *C. necator* formate dehydrogenase in an energy auxotrophic selection strain for formate oxidation coupled with NADH generation. This specific finding is the major achievement of this work. Additionally, GM3 continuous culture evolution

allowed the selection of strain backgrounds favoring the CnFDH oxidative activity, which proved to be versatile. When cured from the initial plasmid, this evolved background could be used for the test of the NADH-regenerating activity of other FDHs much easier than the initial unevolved strain.

The energy-generating systems tested also came with their share of difficulties. For the hydrogen-based system, this could be expected considering the size of the enzymatic complex employed, the soluble hydrogenase from *C. necator*. Further experiments surely need to be pursued to determine whether hydrogen is an option as energy source for CO_2 fixing autotrophic production processes. As for the phosphite-based system, this proved to be much more challenging than expected as well. The phosphite metabolism of *E. coli* still needs to be explored. Our experiments revealed the presence of an unknown system capable of phosphite oxidation to phosphate. In addition, the phosphite dehydrogenase from *P. stutzeri* may lack activity to support the NADH needs of the cells.

Furthermore, an *in vivo* mutagenesis system was implemented in *E. coli* that proved to be active. Further testing and the development of protocols for iterative, automated mutation/selection cycles should provide a powerful method to select for enzyme activities necessitating complex mutation patterns. Finally, the overall set of molecular biology methods explored during this PhD to address the various research questions proved to be diverse, and many of these methods were successfully employed and are now available for further use and adjustments.

To address the challenge of the construction of an autotrophic *E. coli* strain through an efficient CO_2-reduction activity from the CnFDH, an *in silico* modeling of the global metabolism would be highly interesting. Such modeling could identify potential metabolic bottlenecks, helping for further engineering of our metabolic screening strain. Conveniently, a large set of tools for metabolic modeling are available and could be used for this purpose (Lerman et al. (2012); Sridhara et al. (2014); Jung et al. (2009); Tafur Rangel et al. (2021)). This approach proved to be efficient for bioproduction purposes and the implementation of synthetic methylotrophy (Park et al. (2007); Chen et al. (2020)).

Another potential approach that could be considered to achieve efficient bioproduction from CO_2 is co-culturing, to still be able to take advantage of *E. coli* as a convenient bioproduction strain, a technique that showed efficiency in multiple contexts. *C. necator* naturally uses the Calvin cycle to grow autotrophically, but this pathway is inherently inefficient. One strategy could be to rewire *C. necator* metabolism to obtain an FDH-dependent autotrophy in this strain. The Calvin cycle was already successfully replaced by the reductive glycine pathway in this strain (Claassens et al. 2020), this would potentially allow the characterization of the CO_2-reducing activity of the FDH thanks to the advantages of homologous expression (Yu et al. 2019). The study from Bruinsma et al. (2023) proved that a CO_2-reducing NAD^+-dependent FDH could be efficient when expressed in its native host. Since there is no

described FDH of this kind in *E. coli*, this does not constitute an option for this organism. Still, to take advantage of the important capacities of *E. coli* as a bioproduction platform, a heterologous pathway could be expressed in a formatotrophic *E. coli* strain, dependent on the formate production from the engineered autotrophic *C. necator* strain. Co-cultures of *E. coli-C. necator* strains have been shown to be possible (Nangle et al. 2020), but this approach raises questions about the co-culture stability, which should be addressed by creating an inter-dependence between the two strains.

From a more personal perspective, this highly ambitious and multi-axis PhD project offered me the possibility to obtain both technical and theoretical knowledge in the field of metabolic engineering. Before starting my PhD, this field was quasi-unknown to me and the step was high. While navigating into this highly technical research area was definitely demanding, it was a rewarding experience. With the numerous challenges I faced along the way, the knowledge I gained is substantial and it expanded my "problem-solving" vision.

Résumé de la thèse en français

Introduction

Ce projet de doctorat, financé par le CEA dans le cadre du programme "FOCUS FCC", vise à développer une souche d'*Escherichia coli* capable d'utiliser le dioxyde de carbone (CO_2) comme source de carbone et le dihydrogène (H_2) comme source d'énergie. Les laboratoires impliqués, L2BMS et LCBM, apportent respectivement leur expertise en ingénierie métabolique et en manipulation des metalloenzymes. L'objectif est d'introduire une formiate déshydrogénase (FDH) tolérante à l'oxygène dans une souche d'*E. coli*, permettant la réduction du CO_2. La réduction du CO_2 consommant du NADH, deux approches de génération de NADH sont explorées : l'utilisation d'une hydrogénase ou d'une phosphite déshydrogénase. Le schéma du projet est présenté dans la Figure 1.

Ce projet de recherche exploratoire a rencontré plusieurs défis expérimentaux. La décision a été prise de se concentrer principalement sur l'expression d'une FDH complexe dans *E. coli*, avec l'oxydation du formiate couplé à la réduction du NAD^+ comme sens de réaction sélectionnable. Les résultats étaient concluants et ont été explorés de manière approfondie en parallèle avec la feuille de route initiale.

La quantité de CO_2, l'un des principaux gaz à effet de serre, a considérablement augmenté dans l'atmosphère en raison des activités humaines et de la combustion d'énergies fossiles. Avant la révolution industrielle, la concentration était 30 % plus basse qu'actuellement. Comme souligné par la signature des Accords de Paris en 2015, la réduction des émissions de CO_2 est cruciale. En parallèle, la capture, le stockage du CO_2 et sa conversion en produits chimiques d'intérêt ou carburants sont des axes importants de la lutte contre le changement climatique. Actuellement, des technologies de conversion du CO_2 sont disponibles et la fixation ou la réduction de ce dernier par des organismes vivants est une option de plus en plus explorée.

La conversion du CO_2 peut se faire *via* des procédés thermochimique, photochimique ou électrochimique, notamment à l'aide d'enzymes. Les enzymes réalisant cette réduction sont des monoxydes de carbone déshydrogénase ou des formate déshydrogénases. Cette deuxième catégorie d'enzymes, pouvant catalyser la réduction du CO_2 en formiate (HCOOH) est particulièrement intéressante, le formiate étant non-toxique, sûr et très soluble dans l'eau. Par ailleurs, il est la plus simple molécule organique pouvant être naturellement utilisé comme substrat de croissance par certains organismes, comme les méthylotrophes (Chistoserdova 2011). Les organismes naturellement formatotrophes étant généralement plus difficile à manipuler, l'implémentation de la formatotrophie chez des organismes modèles tels qu'*Escherichia coli* a été explorée ces dernières années (Kim et al.

(2020) ; Delmas et al. (2022)). Cependant, la production efficace et durable du formate dépend souvent de l'utilisation d'électrodes dont le coût et la durabilité ne sont pas économiquement compétitifs (Martín, Larrazábal, and Pérez-Ramírez 2015). Par conséquent, il y a un intérêt important à l'assimilation directe du CO_2 par des souches microbiennes.

Cette assimilation directe du CO_2 est déjà opérée naturellement par des organismes autotrophes et peut être faite par différentes voies métaboliques. Dans la nature, la voie de réduction des pentoses phosphates (ou cycle de Calvin) est la voie principale d'assimilation du CO_2. En plus de cette voie, cinq autres ont été identifiées : le cycle de réduction de l'acide tricarboxylique (rTCA), la voie réductrice de l'acétyl-CoA (voie des Wood-Ljungdahl), le cycle dicarboxylate/4-hydroxybutyrate, la voie 3-hydroxypropionate/4-hydroxybutyrate et le cycle 3-hydroxypropionate (Bar-Even, Noor, and Milo 2011). En plus de ces six voies de fixation, trois alternatives ont été décrites récemment : la voie réductrice de l'hexulose-phosphate, la voie naturelle du RGP (Reductive Glycine Pathway), et le cycle inverse de réduction de l'acide tricarboxylique (Santos Correa et al. 2023).

Dans un objectif de développement d'une souche châssis pour la bioproduction en condition aérobie, la voie la plus facilement utilisable est celle du cycle de Calvin (Figure 2) compte tenu de l'absence d'enzymes sensibles à l'oxygène notamment. En revanche, compte tenu de sa faible efficacité énergétique et de la moindre facilité à modifier les organismes qui le portent, l'exploration de voies synthétiques d'assimilation du CO_2 ont été envisagées.

Le cycle de Calvin a été exprimé avec succès de manière hétérologue chez *E. coli*, faisant de cet organisme naturellement hétérotrophe, un autotrophe (Gleizer et al. 2019). Cependant, le cycle de Calvin ainsi implémenté, nécessitait une source d'énergie et le formiate a été utilisé *via* l'expression d'une FDH couplant l'oxydation du formiate à une réduction du NAD^+. Par conséquent, le bilan de cette souche pour la capture du CO_2 n'est pas positif et la faible efficacité du cycle de Calvin en fait une option peu favorable à son utilisation pour la bioproduction.

La difficulté de l'implémentation de l'autotrophie chez des organismes hétérotrophes et le faible nombre d'options disponibles, a poussé les chercheurs à développer des modes de croissances alternatifs comme la méthylotrophie ou la formatotrophie synthétique.

Escherichia coli, eubactérie à Gram négatif et anaérobe facultatif, est choisie comme souche châssis en raison du niveau de connaissance disponible sur son génome et son édition, son métabolisme et les voies de bioproduction de composés d'intérêt développées pour cet organisme. *E. coli* utilise habituellement des sucres pour produire de la biomasse qui sont assimilés *via* la voie d'Embden-Meyerhoff-Parnass (EMP) jusqu'au pyruvate qui est ensuite converti en acétyl-CoA puis entre dans le cycle de Krebs en condition aérobie (Figure 9).

La flexibilité du métabolisme central d'*E. coli* a été montrée à de nombreuses reprises, notamment lors de l'implémentation de l'autotrophie, de la formatotrophie ou de la méthylotrophie. En outre, elle a été utilisée pour produire une grande palette de produits chimiques d'intérêts comme l'éthanol, le pyruvate, l'acétate ou l'acide lactique ainsi que des protéines. Dans le but de rendre ces bioproductions davantage durables, la construction d'une souche d'*E. coli* capable de réaliser ces bioproductions en utilisant le CO_2 pour produire de la biomasse présente donc un intérêt majeur.

La formatotrophie est le mode de croissance favorisé en raison des caractéristiques avantageuses du formiate. Une étude systématique des différentes possibilités d'implémentation de la formatotrophie chez *E. coli* a été menée et a démontré que le RGP était la voie la plus favorable, compte-tenu de critères tels que le rendement théorique ou la faisabilité thermodynamique (Bar-Even et al. (2013) ; Figure 7). Cette voie a été implémentée avec succès, de façon modulaire chez *E. coli* à plusieurs reprises, en utilisant différentes stratégies (Kim et al. (2020) ; Bang et al. (2020) ; Delmas et al. (2022)).

Ces souches formatotrophes peuvent être utilisées comme châssis pour l'implémentation de l'autotrophie en y exprimant une FDH capable de réaliser la réduction du CO_2 en formiate qui sera ensuite utilisé comme source de carbone pour la biomasse. Les FDHs sont une large famille d'enzymes catalysant la réaction réversible de l'oxydation du formiate en CO_2. Cette réaction est réversible mais la plupart des FDHs ne catalysent que l'oxydation du formiate en conditions physiologiques.

Chez les microorganismes, les FDHs jouent des rôles variés. Par exemple, *Methylorubrum extorquens* AM1, organisme méthylotrophe facultatif, possède trois FDHs NAD^+-dépendantes qui sont utilisées en cas d'absence de méthanol pour que le formiate puisse être utilisé comme source d'énergie (Chistoserdova et al. 2004). Chez les organismes anaérobes utilisant la voie de Wood-Ljungdahl, comme l'organisme acétogène *Moorella thermoacetica*, les FDHs catalysent la réduction du CO_2 en formiate qui est ensuite assimilé pour la synthèse d'acétate (Ljungdahl and Andreesen 1975). En résumé, les FDHs sont principalement impliquées dans l'utilisation du formiate comme source d'énergie, soit en transférant les électrons à une autre enzyme (principalement chez les organismes anaérobies), soit en régénérant directement le pouvoir réducteur, tel que le NADH (principalement chez les organismes aérobies), à l'exception des FDHs impliquées dans la voie de Wood-Ljungdahl.

Les FDHs sont des enzymes variées en composition, structure et mode d'action, et peuvent être classées selon leur dépendance au cofacteur NAD^+ pour l'échange d'électrons. Pour cette étude, les enzymes dépendantes du NAD^+ présentent un intérêt particulier compte-tenu de leurs caractéristiques. Bien que l'étude se concentre sur le co-substrat NAD^+, d'autres co-substrats comme les cytochromes, la quinone et le NAD^+-ferrédoxine sont utilisés par les FDHs (Amao 2018). D'autres classifications basées sur la

teneur en métal des enzymes ont été proposées mais sont moins pertinentes ici (Maia, Moura, and Moura 2015).

Les FDH NAD^+-dépendantes (EC 1.17.1.9) jouent un rôle crucial pour le catabolisme des composés C1 en jouant sur le pool d'équivalents réducteurs dans les cellules et sont exprimées lorsqu'il y a peu d'énergie disponible (Popov and Lamzin 1994). Au sein de cette catégorie, certaines FDHs ont été rapportées comme ayant une activité de réduction du CO_2 en présence d'O_2. Ces FDH peuvent être classés en deux catégories : les FDHs monomériques simples et les FDHs multimériques complexes codées par des opérons. Les paramètres cinétiques décrits dans la littérature pour certaines FDHs simples et complexes sont résumées dans la Table 2. L'analyse de ces données montre que quelle que soit la réaction considérée, l'oxydation du formate ou la réduction du CO_2, les FDHs complexes présentent des caractéristiques cinétiques plus favorables et parmi elles, celle de *Cupriavidus necator* apparait comme la meilleure candidate.

La FDH de *C. necator* (CnFDH) est codée par un opéron de cinq gènes dont la structure n'a pas encore été déterminée à ce jour mais présente une forte similarité de séquence protéique (68 %) avec celle de *Rhodobacter capsulatus* dont la structure est connue (Radon et al. 2020). L'hypothèse peut être faite que les deux structures sont proches et que la CnFDH est donc un dimère d'hétérotétramères. CnFDH fait intervenir 7 centres Fe-S et un cofacteur molybdoptérine guanine dinucléotide (MGD). Cette FDH candidate pour réaliser la réduction du CO_2 en formiate *in vivo* étant NAD^+-dépendante, un système de régénération de ce cofacteur dans les cellules est donc indispensable. Pour cela, deux sources ont été considérées, l'hydrogène et le phosphite.

L'hydrogène est considéré comme une source prometteuse et renouvelable d'énergie qui pourrait remplacer les ressources fossiles dans le cadre d'une économie basée sur l'hydrogène, son utilisation ne produisant que de l'eau comme sous-produit (Sherif, Barbir, and Veziroglu 2005). Pour que l'hydrogène puisse être source d'énergie des cellules, l'emploi d'hydrogénases, metalloenzymes multimériques réalisant la réaction réversible de l'oxydation du dihydrogène en protons et électrons, est indispensable. La plupart d'entre elles font intervenir des atomes de métal (Ni ou Fe) dans leur centre actif (Vignais and Billoud 2007), ainsi que des centres Fe-S (Burgdorf, Lenz, et al. (2005) ; Forzi and Sawers (2007)). L'hydrogénase à Ni-Fe NAD^+-dépendante et O_2-tolérante de *C. necator* (CnHyd) a été identifiée dans la littérature comme une candidate pour notre projet car elle possède les caractéristiques essentielles (Friedrich, Fritsch, and Lenz 2011). Cette enzyme cytoplasmique comprend deux modules principaux : l'hydrogénase hétérodimérique et le module NADH:oxydoréductase (diaphorase) contenant le FMN. En tant qu'hydrogénase à Ni-Fe, CnHyd nécessite une machinerie de maturation complexe impliquant plusieurs gènes auxiliaires (Schiffels et al. (2013) ; Lamont and Sargent (2017)). Par ailleurs, cette enzyme a déjà été exprimée avec succès de manière hétérologue chez *E. coli via* différentes stratégies de clonage

impliquant les gènes natifs (Schiffels et al. (2013) ; Ghosh, Bisaillon, and Hallenbeck (2013)) ou des gènes synthétiques optimisés pour leur expression chez *E. coli* (Lamont and Sargent 2017).

Compte-tenu de la complexité des hydrogénases ainsi que des difficultés liées à l'utilisation d'H_2 en présence d'O_2, une alternative, le phosphite a également été exploré. Le phosphite, HPO_3^{2-}, est considéré comme une source d'énergie très prometteuse pour les micro-organismes en raison de son potentiel redox très bas (-650 mV). De plus, il présente des caractéristiques favorables telles qu'une solubilité élevée et une faible toxicité pour les micro-organismes. Contrairement à l'hydrogène, le phosphite a une faible inflammabilité, ce qui le rend plus sûr à utiliser.

Aujourd'hui, seulement deux micro-organismes sont décrits comme capables d'utiliser le phosphite comme source d'énergie et les systèmes enzymatiques utilisés par ces organismes sont peu décrits (Schink et al. (2002) ; Figueroa et al. (2018)). En revanche, la phosphite déshydrogénase NAD^+-dépendante et monomérique de *Pseudomonas stutzeri* est plus étudiée (Costas, White, and Metcalf (2001) ; White and Metcalf (2004b) ; Figueroa and Coates (2017)). Cette enzyme est codée par un unique gène, *ptxD*, et est exprimée sous forme d'opéron chez son hôte avec un transporteur de type ABC codé par les gènes *ptxABC* (White and Metcalf 2007). Son surexpression et sa production chez *E. coli* ont permis de déterminer ses caractéristiques cinétiques favorables (K_m = 53.1 ± 6.7 µM pour le phosphite et une constante catalytique (k_{cat}) de 7.3 s^{-1}). Par ailleurs, cette enzyme a été utilisée avec succès pour complémenter les besoins en phosphate chez *E. coli* ou *P. putida* (Hirota et al. (2017) ; Asin-Garcia et al. (2022)).

Comme dernier axe de ce projet doctoral, une méthode de mutagenèse *in vivo* a également été exploré. En effet, il peut être anticipé que les FDH O_2-tolérantes considérées pour cette thèse ne possèdent pas une activité intrinsèque suffisante à l'obtention de l'autotrophie. Par ailleurs, l'emploi de méthode d'évolution dirigée en culture continue peut être un outil puissant mais le faible taux de mutation d'*E. coli*, crucial pour l'intégrité génétique, peut ralentir l'émergence de mutants nécessaire à l'obtention des phénotypes recherchés. Par conséquent, un système de mutagenèse *in vivo*, Evolution.T7 (iGEM 2021 Evry Paris-Saclay), a été exploré. Il existe des méthodes de mutagenèse *in vitro* et *in vivo* qui peuvent être utiles selon les objectifs recherchés et l'état des connaissances de l'enzyme ou des enzymes ciblées par le système de mutagenèse. Les connaissances, notamment sur la relation structure-fonction des FDHs complexes, n'étant pas complètes, une méthode permettant d'introduire des mutations sur l'intégralité des gènes parait plus appropriée que l'emploi de méthodes ciblées.

Ces dernières années, des méthodes employant la très processive ARN polymérase T7 (T7RNAP) provenant du phage T7, fusionnée à des bases déaminases se sont développées (Moore, Papa, and Shoulders (2018) ; Park and Kim (2021) ; Álvarez et al. (2020) ; Cravens et al. (2021)). Le système exploré

durant cette thèse, Evolution.T7, est basé sur ces précédents travaux et a été développé par Ioana Popescu lors du concours iGEM 2021 (Université Paris-Saclay, équipe d'Evry). Ce système innovant se base sur l'utilisation de deux T7RNAP fusionnées à des bases déaminases reconnaissant deux promoteurs différents se situant de part et d'autre de la région à évoluer, permettant l'introduction de mutations sur les deux brins de l'ADN (Figure 14). Ce système a été testé et une preuve de concept a été établie sur un gène codant pour une β-lactamase conférant une résistance à l'ampicilline et a permis de sélectionner des mutants capables de résister à l'aztréonam, un autre antibiotique faisant partie de la famille des β-lactames. Il est essentiel de noter que la plupart des techniques de mutagenèse décrites ont montré leur efficacité dans l'évolution des gènes de résistance aux antibiotiques, mais leur application dans l'évolution d'une enzyme clé de l'assimilation du carbone représente un défi. La difficulté réside dans la recherche d'un test de criblage *in vivo* efficace, et le manque d'exemples de réussite sur ce type de gène comme cible souligne la complexité de cette démarche.

Le projet doctoral avait pour objectif de démontrer la réduction directe du CO_2 dans le châssis biotechnologique *E. coli* en utilisant des systèmes enzymatiques prometteurs, mais non validés *in vivo*. Les questions explorées étaient diverses et comprenaient la stabilité de la FDH de *C. necator* dans des conditions aérobies chez *E. coli*, sa capacité à réduire le CO_2 *in vivo* dans des souches modifiées d'*E. coli*, l'implémentation efficace d'un système non carboné de génération de l'énergie dans les cellules, et l'utilisation d'une méthode de mutagenèse *in vivo* pour générer des variants enzymatiques favorables.

Résultats

Dans un premier temps, les expériences menées ont eu pour objectif de démontrer la fonctionnalité *in vivo* de la formate déshydrogénase tolérante à l'oxygène et dépendante du NAD^+ de *Cupriavidus necator* (CnFDH) dans *E. coli*. À cette fin, l'enzyme a été clonée et exprimée dans une souche de sélection d'*E. coli* dépendante de l'oxydation du formiate en CO_2 par la FDH pour fournir du NADH, indispensable à la croissance cellulaire, la réaction d'oxydation étant favorisée par rapport à la réduction.

Ce crible de sélection est dit « énergie-auxotrophe » et consiste en une souche d'*E. coli* dans laquelle le gène codant pour la lipoamide déshydrogénase, *lpd*, a été interrompu (Wenk et al. (2020) ; Figure 26). Le gène *lpd* est impliqué dans trois complexes enzymatique dans le métabolisme d'*E. coli* que sont : la pyruvate déshydrogénase, la 2-oxoglutarate déshydrogénase et le système de clivage de la glycine. Cette délétion rend la souche auxotrophe au pouvoir réducteur et nécessite l'apport d'acétate en tant que précurseur des acides gras, qui est assimilé *via* le shunt glyoxylique, ainsi que d'une source d'énergie. Ainsi, ce crible peut servir de test métabolique pour différentes enzymes de génération d'énergie que sont les FDHs, les hydrogénases ou les phosphite déshydrogénases, qui sont testées dans ce projet de

recherche. Le NADH généré par l'expression de ces systèmes hétérologues peut ensuite être utilisé pour produire de l'ATP *via* la phosphorylation oxydative.

Grâce aux expériences menées, nous avons pu démontrer l'activité de la formate déshydrogénase NAD^+/Mo-dépendante de *Cupriavidus necator* pour l'oxydation du formiate, *in vivo*, dans *E. coli*, en présence d'O_2. Le crible de sélection (Δ*ldp*) présentait des difficultés de croissance dans les conditions du test métabolique lorsque le formiate était la source de NADH par l'activité de CnFDH exprimée sur un plasmide, notamment caractérisée par une longue phase de latence et un long temps de génération. Une évolution dirigée en culture continue (automate de culture GM3, Mutzel and Marliere (2004)) a permis la sélection de variants présentant un temps de génération trois fois plus court que la souche parentale et une phase de latence considérablement réduite. Le contexte génétique des isolats issus de l'évolution en culture continue s'est montré plus favorable à l'oxydation du formiate catalysée par des FDHs, comme démontré pour CnFDH, mais aussi pour la FDH simple de *Thiobacillus sp*. Il est donc plus adapté pour le test de diverses FDH dépendantes du NAD^+, le rendant polyvalent.

Un séquençage génomique complet de plusieurs isolats issus des évolutions en culture continue ont permis de mettre en évidence des mutations dans le gène *focA*, codant pour un transporteur de formate bidirectionnel. L'étude de ces mutations par mutagenèse reverse a permis de montrer qu'elles étaient les principales causes de l'augmentation du taux de croissance observé lors de l'évolution.

En utilisant ce contexte génétique évolué et surmontant de nombreuses difficultés techniques liées à la complexité du système manipulé, nous avons réussi à intégrer avec succès l'opéron CnFDH dans le chromosome d'*E. coli* (locus IS10) et à obtenir une expression efficace pour restaurer la croissance dans des conditions de test. Une évolution supplémentaire en culture continue (GM3) a confirmé la stabilité de l'insertion et de l'expression de la CnFDH pendant plus de 400 générations. Ces résultats ont prouvé la faisabilité de l'utilisation de tels complexes dans *E. coli* lorsque le test métabolique est correctement verrouillé.

Bien que cette démonstration de l'activité de la CnFDH dans *E. coli* pour l'oxydation du formiate soit un résultat prometteur, l'objectif était d'explorer la réversibilité de cette réaction chez *E. coli* en étudiant la réduction du CO_2 par la CnFDH, qui n'a été démontrée jusqu'à présent qu'*in vitro*.

La suite des travaux s'est donc concentrée sur la réduction du CO_2 en utilisant la FDH de *C. necator*. Deux souches précédemment développées par le laboratoire ont été utilisées comme base pour ces expériences, une souche serine auxotrophe et une autre glycine auxotrophe. Les deux souches dépendent du formiate pour leur croissance et utilisent le RGP pour l'assimilation du formiate, mais diffèrent par la quantité de formiate incorporée dans le carbone total nécessaire pour la production de biomasse (Figure 46 et Figure 47). Les besoins en formiate, autrement dit la stringence de la sélection,

étaient ajustables ; la souche auxotrophe à la sérine représentait une stringence de 3 % (quantité de formiate incorporé par rapport au carbone total) et la souche auxotrophe à la glycine environ 8 %.

Les premières expériences réalisées en exprimant la CnFDH n'ayant pas permis de mettre en évidence l'activité hétérologue de réduction du CO_2, les souches de cribles ont été davantage modifiées, notamment *via* deux délétions supplémentaires. Tout d'abord, le gène *fdoG* codant pour la FDH-O de *E. coli* active en condition aérobie (Benoit, Abaibou, and Mandrand-Berthelot 1998) a été interrompu dans le but de réduire les besoins en formate des cribles. Ensuite, le gène *nuoG*, codant pour une sous-unité essentielle du complexe I de la chaine respiratoire, qui oxyde le NADH (Pohl, Uhlmann, et al. 2007) a été interrompu, dans le but d'augmenter le pool de NADH dans les cellules et favoriser la réaction recherchée qui utilise ce cofacteur comme substrat. La suppression du gène *fdoG* codant pour la formiate déshydrogénase O a réduit drastiquement les besoins en formiate des souches.

Cependant, malgré les efforts déployés pour démontrer l'activité de la CnFDH pour la réduction *in vivo* du CO_2, nous n'avons pas pu atteindre cet objectif. L'une des difficultés est liée à la thermodynamique de la réaction, compte tenu du potentiel redox CO_2/formiate de -430 mV et du potentiel du couple NAD^+/NADH de -320 mV. Bien que ces potentiels redox ne soient pas extrêmement différents, rendant la réaction thermodynamiquement possible, elle reste défavorable. Pour favoriser cette réaction, deux paramètres clés doivent être explorés : augmenter la concentration en CO_2, ce qui a été exploré en augmentant la teneur en CO_2 jusqu'à 50 % dans l'air, et améliorer la disponibilité du NADH à l'intérieur des cellules. Le pool NAD^+/NADH est étroitement régulé dans les cellules, avec un rapport de 10,6 (2,56 et 0,24 µmol par ml de culture pour NAD^+ et NADH respectivement) chez *E. coli* cultivé en glucose dans des conditions aérobies (Leonardo, Dailly, and Clark 1996). La suppression du gène *nuoG* a favorisé la réduction du formiate dans certaines circonstances, mais n'a pas été suffisante pour établir le phénotype de croissance souhaité. Une option à explorer pour augmenter la concentration en NADH dans les cellules serait d'exprimer le transporteur de NAD^+ NTT4, qui est décrit pour transporter à la fois le NAD^+ et le NADH (Haferkamp et al. (2004) ; Zhou et al. (2011)), et de supplémenter le milieu de croissance avec du NADH. Cependant, le NADH étant instable et coûteux, cette stratégie serait plus adaptée pour des tests à petite échelle que pour des installations d'évolution plus importantes.

Par ailleurs, nous avons pu cultiver ces souches de sélection pendant un grand nombre de génération (400 et 800 respectivement pour les deux cribles testés serine et glycine auxotrophe) dans un régime de culture continue (GM3) sans l'apparition de cellules produisant du formiate de manière indépendante de la FDH. De telles cellules auraient rapidement pris le dessus dans les cultures.

Comme décrit en introduction, la construction de la souche autotrophique qui était ciblée nécessite un système de génération de l'énergie. Dans ce cadre, deux systèmes de régénération d'énergie ont été explorés *via* l'utilisation de l'hydrogène ou du phosphite.

L'étude de l'hydrogène en tant que donneur d'électrons a été explorée par l'expression d'une hydrogénase dépendante du NAD^+ provenant de *C. necator* (CnHyd). Cette enzyme, décrite comme tolérante à l'oxygène, est codée par un opéron de 14 gènes et implique des métaux, essentiels pour son activité. Elle catalyse la clivage réversible de l'H_2 en deux protons et deux électrons avec le NAD^+ comme cofacteur. Nous avons obtenu des constructions plasmidiques portant les gènes synthétiques codant pour l'opéron CnHyd des auteurs d'une étude précédente (Lamont and Sargent 2017) et les avons introduites dans la souche d'*E. coli* auxotrophe en énergie (Δ*lpd*) décrite ci-dessus, comportant des modifications génomiques supplémentaires que sont les délétions des gènes codants pour les quatre hydrogénases endogènes d'*E. coli*. La fonctionnalité du complexe dans notre souche génétiquement modifiée a été initialement vérifiée pour la production d'hydrogène. Ensuite, nous avons réalisé des tests afin de tester l'utilisation du dihydrogène comme source d'énergie.

La fonctionnalité de la CnHyd, encodée par des gènes optimisés organisés dans un opéron synthétique, a été démontrée avec succès pour la production d'H_2 en des conditions anaérobies dans des souches d'*E. coli* dépourvues des hydrogénases endogènes. La CnHyd fonctionne physiologiquement dans la direction oxydative de clivage de l'H_2 dans *Cupriavidus necator* car le dihydrogène est utilisé comme donneur d'électrons en mode de croissance autotrophe. Cependant, son activité hétérologue dans *E. coli* dans cette direction n'a pas pu être démontrée dans notre étude.

En raison d'un manque de temps et de la complexité du protocole, nous n'avons pas pu réaliser un grand nombre de tests pour l'oxydation de H_2. Ceux qui ont été réalisés n'ont pas permis de révéler cette activité dans le dispositif de test. Quelques hypothèses peuvent être formulées pour expliquer ces résultats. Tout d'abord, il existe une différence importante entre le milieu utilisé pour les essais de production de H_2, qui est un milieu riche, et le milieu minimal pour le test métabolique ne contenant que de l'acétate et du pyruvate comme sources de carbone. La croissance dans ce milieu strictement minimal pourrait ne pas être compatible avec le fonctionnement d'un système enzymatique complexe comme la CnHyd, nécessitant la disponibilité d'acides aminés et de divers facteurs de croissance pour une expression efficace et le repliement des protéines. L'ajout des métaux essentiels Ni et Fe au milieu, nécessaire pour la maturation du complexe, n'était pas suffisant pour démontrer son activité. Des ajustements de ce milieu sans compromettre le test métabolique sont probablement nécessaires. De plus, le gène *hoxN1* de *C. necator* codant pour un transporteur de nickel à haute affinité devrait être ajouté à l'ensemble des gènes auxiliaires, car il a été rapporté qu'il améliore l'efficacité de maturation de la CnHyd exprimée dans *E. coli* dans des conditions aérobies (Schiffels et al. 2013). Par ailleurs, l'emploi

de gènes synthétiques pour l'expression de metalloenzymes complexes peut conduire à une efficacité d'expression plus faible comme démontré dans la littérature (Ghosh, Bisaillon, and Hallenbeck (2013) ; Lamont and Sargent (2017)). Par conséquent, il serait intéressant de réaliser le clonage avec les gènes natifs, suivant les mêmes schémas de construction décrits dans la littérature pour tester si de meilleurs résultats sont obtenus. Dans l'ensemble, de nombreuses options restent à explorer pour pouvoir se prononcer sur la faisabilité de l'emploi d'hydrogénases dépendantes du NAD^+ en tant que systèmes de régénération d'énergie. Cependant, même si une telle activité était démontrée, l'explosivité du gaz H_2 restera un défi pour sa manipulation, ainsi que son insolubilité relative.

Un autre système de génération du NADH, la phosphite déshydrogénase de *Pseudomonas stutzeri* (PtxD), a été étudié et est supposé être plus favorable, en partie en raison du potentiel redox très bas du couple phosphate/phosphite (-650 mV), donc plus favorable à la réduction de NAD^+. Ce système avait également été mentionné dans la littérature comme étant prometteur (Claassens et al. 2018).

Lors de nos tests, nous avons pu démontrer l'activité oxydante du phosphite de la phosphite déshydrogénase de *Pseudomonas stutzeri in vivo* pour produire du phosphate nécessaire à la croissance des cellules. Cette démonstration avait déjà été rapportée pour *E. coli* et *P. putida* en utilisant cette même PTDH (Hirota et al. (2017) ; Asin-Garcia et al. (2022)). Cependant, ce système n'a pas permis de complémenter l'auxotrophie du crible Δ*lpd* avec le phosphite comme seule source d'énergie. Ce résultat peut être dû soit à un manque d'activité de la PTDH, soit à la présence d'une enzyme ou d'une voie non identifiée dans *E. coli* indépendante du NAD^+ et capable d'oxyder le phosphite en phosphate qui a été mise en évidence lors des expériences menées.

L'hypothèse du manque d'activité de PtxD est peu probable compte tenu des paramètres cinétiques très favorables qui ont été calculés pour cette enzyme (Costas, White, and Metcalf 2001). Cependant, ces données *in vitro* ne peuvent pas automatiquement être extrapolées à la situation *in vivo* ; une faible expression ou un repliement imparfait des chaînes peptidiques pourraient se produire. En ce qui concerne la présence de voies non identifiées de conversion du phosphite en phosphate, nous avons démontré que la croissance en milieu exempt de phosphate est restaurée dans des souches dépourvues des enzymes endogènes de *E. coli* décrites comme oxydant de phosphite (BAP, gène *phoA* et C-P lyase, opéron *phn*) avec des concentrations élevées de phosphite. Un processus inconnu opérant soit dans le périplasme, empêchant l'accès du phosphite au cytoplasme où se trouve la PTDH hétérologue, soit dans le cytoplasme lui-même, potentiellement avec une efficacité ou une affinité plus élevée pour le phosphite par rapport à PtxD, pourrait être impliqué. Cependant, cette dernière hypothèse est discutable car si un tel système existe, l'excès important de phosphite fourni aux cellules devrait également être suffisant pour être utilisé comme substrat par l'enzyme PtxD.

Pour approfondir la compréhension de ce système potentiel de régénération de NADH, des investigations supplémentaires sont nécessaires. De plus, la construction et le test d'une bibliothèque de mutants « knock-out » pourraient aider à identifier la source précise de l'oxydation du phosphite à supprimer pour éviter les interférences avec PtxD. Enfin, l'exploration de la biodiversité afin de trouver d'autres phosphite déshydrogénases à tester serait également intéressante, avec notamment celle de l'organisme *Desulfotignum phosphitoxidans* qui utilise effectivement le phosphite comme source d'énergie (Figueroa and Coates 2017).

Au cours de ce travail de doctorat, une partie importante des travaux a consisté à développer et à améliorer des méthodes en raison de la complexité des enzymes hétérologues et des souches manipulées.

Une partie de ce travail méthodologique a consisté en la mise en place d'une méthode de mutagenèse *in vivo*, connue sous le nom d'Evolution.T7 (iGEM 2021 Evry Paris-Saclay). Cette méthode a été développée pour introduire des mutations dans une région ciblée afin de sélectionner des mutants améliorés pour une activité d'intérêt. Les résultats obtenus sur ce projet sont préliminaires et restent encore à approfondir. Toutefois, nous avons pu démontrer l'activité du système de mutagenèse sur le gène choisi comme cible, la FDH de *Thiobacillus sp.*, clonée dans un plasmide portant les caractéristiques nécessaires au système de mutagenèse (Figure 66), validant ainsi cette construction et l'ensemble du dispositif expérimental. À cet égard, il serait très intéressant de réaliser une analyse plus approfondie des paramètres expérimentaux tels que le temps d'induction et d'effectuer une induction séquentielle des deux complexes mutateurs. De plus, l'utilisation du séquençage à haut débit faciliterait l'interprétation des résultats et fournirait des résultats plus précis, en particulier en termes de taux de mutation. Cette approche a été utilisée par l'équipe iGEM pour le développement d'Evolution.T7.

Après la validation fonctionnelle du système, nous avons débuté un test métabolique, formate source d'énergie *via* l'activité de la TsFDH. Les résultats obtenus sont préliminaires et nécessitent une analyse plus détaillée des clones obtenus sur les boites de test métabolique après induction de la mutagenèse. Par ailleurs, ce type de mutagenèse *in vivo* impliquera toujours des risques de mutations « off-target », dans le chromosome notamment, qui pourraient compromettre la sélection métabolique ou le métabolisme en général. Cet aspect très important devrait être exploré pour garantir un système étroitement régulé.

La deuxième partie méthodologique a consisté en l'optimisation d'outils de clonage et de manipulation des génomes. Ces outils étaient nécessaires pour surmonter les problèmes techniques et atteindre les objectifs du projet. Ce projet de thèse a été l'occasion d'explorer un large éventail de techniques de PCR et de clonage notamment qui n'étaient pas couramment utilisées dans le laboratoire. Cela a également

mis en évidence les limitations de certains systèmes, pour l'édition du génome en général et l'intégration chromosomique en particulier. Dans ce contexte, une méthode d'insertion génomique a été adaptée à partir d'une publication (Vigouroux et al. 2018), et les premiers résultats sont prometteurs. Cependant, des tests supplémentaires de cette méthode sont nécessaires pour valider son efficacité. Néanmoins, la manipulation de cette méthode s'est avérée beaucoup plus pratique par rapport à celle initialement utilisée avec le plasmide pDM4 (Wenk et al. 2018).

Conclusion et perspectives

Utiliser le CO_2 ou ses dérivés réduits, notamment le formiate et le méthanol, comme matière première pour la production de biomasse, constitue un objectif important de la bio-industrie. Des progrès ont été réalisés dans la mise en œuvre de voies synthétiques d'assimilation de ces composés C1 réduits dans *E. coli*, l'organisme de production biologique le plus polyvalent et facile à manipuler.

Dans ce projet de thèse, la stratégie suivie est celle de la réduction du CO_2 par une FDH pour la construction d'une souche autotrophe d'*E. coli*. Des souches dépendantes du formiate, obtenues précédemment par la mise en œuvre de la voie RGP et disponibles au laboratoire, ont été utilisés comme cribles de sélection *in vivo* pour la production de formiate *via* la CnFDH.

De nombreuses difficultés techniques ont été rencontrées tout au long de la réalisation de ce projet, principalement causées par le choix de la FDH que nous avions fait. La CnFDH était un choix rationnel en raison de sa tolérance à l'O_2 et de sa dépendance au NAD^+, ce qui, associé à une activité rapportée pour la réduction du CO_2 10 fois supérieure à celle mesurée pour les meilleures enzymes monomériques, en faisait une FDH idéale pour le développement d'une souche autotrophe aérobie d'*E. coli*. Cependant, la démonstration *in vivo* de l'activité de réduction du CO_2 de la CnFDH, activité exclusivement démontrée *in vitro*, s'est avérée beaucoup plus difficile qu'initialement prévu.

Malheureusement, nous n'avons pas pu observer une telle activité dans les souches hémi-formatotrophes utilisées pour ce projet. Bien que ces enzymes réductrices du CO_2 tolérantes à l'O_2 attirent beaucoup d'attention, à ce jour, aucune démonstration expérimentale de leur activité *in vivo* dans le sens réducteur n'a été rapportée. En revanche, nous avons pu démontrer l'activité *in vivo* de la FDH de *C. necator* dans un crible dépendant de la génération du NADH couplée à l'oxydation du formiate. Cette démonstration est la réalisation majeure de ce travail.

Les systèmes de génération d'énergie testés ont également présenté leur lot de difficultés. Pour le système à base d'hydrogène, cela pouvait être attendu compte tenu de la taille du complexe enzymatique utilisé, l'hydrogénase soluble de *C. necator*. D'autres expériences doivent être menées afin

de déterminer si l'hydrogène est une option en tant que source d'énergie pour les processus de fixation autotrophes du CO_2. En ce qui concerne le système à base de phosphite, cela s'est avéré beaucoup plus difficile que prévu également. Le métabolisme du phosphite chez *E. coli* doit encore être exploré. Nos expériences ont révélé la présence d'un système inconnu capable d'oxyder le phosphite en phosphate.

Enfin, un système de mutagenèse *in vivo* a été mis en œuvre dans *E. coli* et s'est avéré fonctionnel. Des tests supplémentaires et le développement de protocoles pour des cycles itératifs de mutation/sélection automatisés devraient fournir une méthode robuste pour sélectionner des activités enzymatiques nécessitant des schémas de mutation complexes. Enfin, l'ensemble des méthodes de biologie moléculaire explorées au cours de cette thèse pour répondre aux diverses questions de recherche s'est révélé diversifié, et bon nombre de ces méthodes ont été mises en place avec succès et sont maintenant disponibles pour leur utilisation et amélioration ultérieures.

Pour relever le défi de la construction d'une souche autotrophe d'*E. coli* grâce à une activité efficace de réduction du CO_2 de la CnFDH, une modélisation *in silico* du métabolisme global serait très intéressante afin d'identifier les éventuels points critiques au sein des flux métaboliques pour la construction d'une telle souche.

Une autre approche potentielle qui pourrait être envisagée pour atteindre une bioproduction efficace à partir du CO_2 est la co-culture, afin de tirer parti d'*E. coli* en tant que souche de bioproduction versatile, une technique qui a montré son efficacité dans de multiples contextes. *C. necator* utilise naturellement le cycle de Calvin pour croître de manière autotrophe, mais cette voie est intrinsèquement inefficace. Une stratégie pourrait consister à rediriger le métabolisme de *C. necator* pour obtenir une autotrophie dépendante de l'activité réductrice de la FDH. Le cycle de Calvin a déjà été remplacé avec succès par la voie RGP dans cette souche (Claassens et al. 2020), ce qui permettrait potentiellement la caractérisation de l'activité de réduction du CO_2 de la FDH grâce aux avantages de l'expression homologue. La co-culture de souches *E. coli-C. necator* a été montrée comme faisable (Nangle et al. 2020), mais cette approche soulève des questions sur la stabilité de la co-culture, qui devraient être résolues en créant une interdépendance entre les deux souches.

REFERENCES

Abdel-Hamid, Ahmed M, Margaret M Attwood, and John R Guest. 2001. "Pyruvate oxidase contributes to the aerobic growth efficiency of Escherichia coli." *Microbiology* 147 (6):1483-1498. doi: https://doi.org/10.1099/00221287-147-6-1483.

Alber, Birgit E. 2019. "Autotrophic CO2 Metabolism." In *Encyclopedia of Microbiology (Fourth Edition)*, edited by Thomas M. Schmidt, 293-306. Oxford: Academic Press.

Altaş, Nilay, Aşkın Sevinç Aslan, Ersin Karataş, Evangelia Chronopoulou, Nikolaos E. Labrou, and Barış Binay. 2017. "Heterologous production of extreme alkaline thermostable NAD+-dependent formate dehydrogenase with wide-range pH activity from Myceliophthora thermophila." *Process Biochemistry* 61:110-118. doi: https://doi.org/10.1016/j.procbio.2017.06.017.

Alvarez-Malmagro, Julia, Ana R. Oliveira, Cristina Gutiérrez-Sánchez, Beatriz Villajos, Inês A. C. Pereira, Marisela Vélez, Marcos Pita, and Antonio L. De Lacey. 2021. "Bioelectrocatalytic Activity of W-Formate Dehydrogenase Covalently Immobilized on Functionalized Gold and Graphite Electrodes." *ACS Applied Materials & Interfaces* 13 (10):11891-11900. doi: https://doi.org/10.1021/acsami.0c21932.

Álvarez, Beatriz, Mario Mencía, Víctor de Lorenzo, and Luis Ángel Fernández. 2020. "In vivo diversification of target genomic sites using processive base deaminase fusions blocked by dCas9." *Nature Communications* 11 (1):6436. doi: https://doi.org/10.1038/s41467-020-20230-z.

Amann, Egon, Birgit Ochs, and Karl-Josef Abel. 1988. "Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in Escherichia coli." *Gene* 69 (2):301-315. doi: https://doi.org/10.1016/0378-1119(88)90440-4.

Amao, Yutaka. 2018. "Formate dehydrogenase for CO2 utilization and its application." *Journal of CO2 Utilization* 26:623-641. doi: https://doi.org/10.1016/j.jcou.2018.06.022.

Angermayr, S. Andreas, Aleix Gorchs Rovira, and Klaas J. Hellingwerf. 2015. "Metabolic engineering of cyanobacteria for the synthesis of commodity products." *Trends in Biotechnology* 33 (6):352-361. doi: https://doi.org/10.1016/j.tibtech.2015.03.009.

Antonovsky, Niv, Shmuel Gleizer, Elad Noor, Yehudit Zohar, Elad Herz, Uri Barenholz, Lior Zelcbuch, Shira Amram, Aryeh Wides, Naama Tepper, Dan Davidi, Yinon Bar-On, Tasneem Bareia, David G Wernick, Ido Shani, Sergey Malitsky, Ghil Jona, Arren Bar-Even, and Ron Milo. 2016. "Sugar Synthesis from CO2 in Escherichia coli." *Cell* 166 (1):115-125. doi: https://doi.org/10.1016/j.cell.2016.05.064.

Arkin, M. 2001. "In Vitro Mutagenesis." In *Brenner's Encyclopedia of Genetics (Second Edition)*, edited by Stanley Maloy and Kelly Hughes, 46-50. San Diego: Academic Press.

Asin-Garcia, Enrique, Christos Batianis, Yunsong Li, James D. Fawcett, Ivar de Jong, and Vitor A. P. Martins dos Santos. 2022. "Phosphite synthetic auxotrophy as an effective biocontainment strategy for the industrial chassis Pseudomonas putida." *Microbial Cell Factories* 21 (1):156. doi: https://doi.org/10.1186/s12934-022-01883-5.

Aslan, Aşkın Sevinç, Jarkko Valjakka, Jouni Ruupunen, Deniz Yildirim, Nicholas J. Turner, Ossi Turunen, and Barış Binay. 2016. "Chaetomium thermophilum formate dehydrogenase has high activity in

the reduction of hydrogen carbonate (HCO3−) to formate." *Protein Engineering, Design and Selection* 30 (1):47-55. doi: https://doi.org/10.1093/protein/gzw062.

Baba, T., T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A. Datsenko, M. Tomita, B. L. Wanner, and H. Mori. 2006. "Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection." *Mol Syst Biol* 2:2006.0008. doi: https://doi.org/10.1038/msb4100050.

Bachman, Julia. 2013. "Chapter Nineteen - Site-Directed Mutagenesis." In *Methods in Enzymology*, edited by Jon Lorsch, 241-248. Academic Press.

Bachmeier, Andreas, Vincent C. C. Wang, Thomas W. Woolerton, Sophie Bell, Juan C. Fontecilla-Camps, Mehmet Can, Stephen W. Ragsdale, Yatendra S. Chaudhary, and Fraser A. Armstrong. 2013. "How Light-Harvesting Semiconductors Can Alter the Bias of Reversible Electrocatalysts in Favor of H2 Production and CO2 Reduction." *Journal of the American Chemical Society* 135 (40):15026-15032. doi: https://doi.org/10.1021/ja4042675.

Baeshen, Mohammed N., Ahmed M. Al-Hejin, Roop S. Bora, Mohamed M. M. Ahmed, Hassan A. I. Ramadan, Kulvinder S. Saini, Nabih A. Baeshen, and Elrashdy M. Redwan. 2015. "Production of Biopharmaceuticals in E. coli: Current Scenario and Future Perspectives." *J Microbiol Biotechnol* 25 (7):953-62. doi: https://doi.org/10.4014/jmb.1412.12079.

Bailey, Clara A., and Brandon L. Greene. 2023. "A fluorometric assay for high-throughput phosphite quantitation in biological and environmental matrices." *Analyst* 148 (15):3650-3658. doi: https://doi-org.insb.bib.cnrs.fr/10.1039/D3AN00575E.

Bang, Junho, Jung Ho Ahn, Jong An Lee, Chang Hwang, Gi Kim, Jinwon Lee, and Sang Yup Lee. 2021. "Synthetic Formatotrophs for One-Carbon Biorefinery." *Advanced Science* 8. doi: https://doi.org/10.1002/advs.202100199.

Bang, Junho, Chang Hun Hwang, Jung Ho Ahn, Jong An Lee, and Sang Yup Lee. 2020. "Escherichia coli is engineered to grow on CO2 and formic acid." *Nature Microbiology* 5 (12):1459-1463. doi: https://doi.org/10.1038/s41564-020-00793-9.

Bar-Even, A., E. Noor, A. Flamholz, and R. Milo. 2013. "Design and analysis of metabolic pathways supporting formatotrophic growth for electricity-dependent cultivation of microbes." *Biochim Biophys Acta* 1827 (8-9):1039-47. doi: https://doi.org/10.1016/j.bbabio.2012.10.013.

Bar-Even, Arren, Elad Noor, Nathan E. Lewis, and Ron Milo. 2010. "Design and analysis of synthetic carbon fixation pathways." *Proceedings of the National Academy of Sciences* 107 (19):8889-8894. doi: https://doi.org/10.1073/pnas.0907176107.

Bar-Even, Arren, Elad Noor, and Ron Milo. 2011. "A survey of carbon fixation pathways through a quantitative lens." *Journal of Experimental Botany* 63 (6):2325-2342. doi: https://doi.org/10.1093/jxb/err417.

Bassalo, Marcelo C., Andrew D. Garst, Andrea L. Halweg-Edwards, William C. Grau, Dylan W. Domaille, Vivek K. Mutalik, Adam P. Arkin, and Ryan T. Gill. 2016. "Rapid and Efficient One-Step Metabolic Pathway Integration in E. coli." *ACS Synthetic Biology* 5 (7):561-568. doi: https://doi.org/10.1021/acssynbio.5b00187.

Bassegoda, Arnau, Christopher Madden, David W. Wakerley, Erwin Reisner, and Judy Hirst. 2014. "Reversible Interconversion of CO2 and Formate by a Molybdenum-Containing Formate

Dehydrogenase." *Journal of the American Chemical Society* 136 (44):15473-15476. doi: https://doi.org/10.1021/ja508647u.

Benoit, S., H. Abaibou, and M. A. Mandrand-Berthelot. 1998. "Topological analysis of the aerobic membrane-bound formate dehydrogenase of Escherichia coli." *J Bacteriol* 180 (24):6625-34. doi: https://doi.org/10.1128/jb.180.24.6625-6634.1998.

Berg, Ivan A. 2011. "Ecological aspects of the distribution of different autotrophic CO2 fixation pathways." *Appl Environ Microbiol* 77 (6):1925-36. doi: https://doi.org/10.1128/AEM.02473-10.

Berg, Ivan A., W. Hugo Ramos-Vera, Anna Petri, Harald Huber, and Georg Fuchs. 2010. "Study of the distribution of autotrophic CO2 fixation cycles in Crenarchaeota." *Microbiology* 156 (1):256-269. doi: https://doi.org/10.1099/mic.0.034298-0.

Berg, Ivan, Daniel Kockelkorn, W. Ramos-Vera, Rafael Say, Jan Zarzycki, Michael Hügler, Birgit Alber, and Georg Fuchs. 2010. "Autotrophic Carbon Fixation in Archaea." *Nature reviews. Microbiology* 8:447-60. doi: https://doi.org/10.1038/nrmicro2365.

Berkowitz, Oliver, Ricarda Jost, Stuart J. Pearse, Hans Lambers, Patrick M. Finnegan, Giles E. St J. Hardy, and Philip A. O'Brien. 2011. "An enzymatic fluorescent assay for the quantification of phosphite in a microtiter plate format." *Analytical Biochemistry* 412 (1):74-78. doi: https://doi.org/10.1016/j.ab.2011.01.014.

Birdja, Yuvraj Y., Elena Pérez-Gallent, Marta C. Figueiredo, Adrien J. Göttle, Federico Calle-Vallejo, and Marc T. M. Koper. 2019. "Advances and challenges in understanding the electrocatalytic conversion of carbon dioxide to fuels." *Nature Energy* 4 (9):732-745. doi: https://doi.org/10.1038/s41560-019-0450-y.

Blattner, Frederick R., Guy 3rd Plunkett, C raig A. Bloch, Nicole T. Perna, Valerie Burland, Monica Riley, Julio Collado-Vides, Jeremy D. Glasner, Christopher K. Rode, George F. Mayhew, Jason Gregor, Nelson W. Davis, Heather A. Kirkpatrick, Michael A. Goeden, Debra J. Rose, Bob Mau, and Ying Shao. 1997. "The complete genome sequence of Escherichia coli K-12." *Science* 277 (5331):1453-62. doi: https://doi.org/10.1126/science.277.5331.1453.

Bouzon, Madeleine, Volker Döring, Ivan Dubois, Anne Berger, Gabriele M. M. Stoffel, Liliana Calzadiaz Ramirez, Sophia N. Meyer, Marion Fouré, David Roche, Alain Perret, Tobias J. Erb, Arren Bar-Even, and Steffen N. Lindner. 2021. "Change in Cofactor Specificity of Oxidoreductases by Adaptive Evolution of an Escherichia coli NADPH-Auxotrophic Strain." *mBio* 12 (4):10.1128/mbio.00329-21. doi: https://doi.org/10.1128/mbio.00329-21.

Bouzon, Madeleine, Alain Perret, Olivier Loreau, Valérie Delmas, Nadia Perchat, Jean Weissenbach, Frédéric Taran, and Philippe Marlière. 2017. "A Synthetic Alternative to Canonical One-Carbon Metabolism." *ACS Synthetic Biology* 6 (8):1520-1533. doi: https://doi.org/10.1021/acssynbio.7b00029.

Bradford, Marion M. 1976. "A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding." *Anal Biochem* 72:248-54. doi: https://doi.org/10.1006/abio.1976.9999.

Bruinsma, Lyon, Sebastian Wenk, Nico J. Claassens, and Vitor A. P. Martins dos Santos. 2023. "Paving the way for synthetic C1 - Metabolism in Pseudomonas putida through the reductive glycine pathway." *Metabolic Engineering* 76:215-224. doi: https://doi.org/10.1016/j.ymben.2023.02.004.

Burgdorf, T., E. van der Linden, M. Bernhard, Q. Y. Yin, J. W. Back, A. F. Hartog, A. O. Muijsers, C. G. de Koster, S. P. Albracht, and B. Friedrich. 2005. "The soluble NAD+-Reducing [NiFe]-hydrogenase from Ralstonia eutropha H16 consists of six subunits and can be specifically activated by NADPH." *J Bacteriol* 187 (9):3122-32. doi: https://doi.org/10.1128/jb.187.9.3122-3132.2005.

Burgdorf, Tanja, Olivier Lenz, Thorsten Buhrke, Eddy van der Linden, Anne K. Jones, Simon P. Albracht, and Bärbel Friedrich. 2005. "[NiFe]-hydrogenases of Ralstonia eutropha H16: modular enzymes for oxygen-tolerant biological hydrogen oxidation." *J Mol Microbiol Biotechnol* 10 (2-4):181-96. doi: https://doi.org/10.1159/000091564.

Cai, Zhen, Guoxia Liu, Junli Zhang, and Yin Li. 2014. "Development of an activity-directed selection system enabled significant improvement of the carboxylation efficiency of Rubisco." *Protein & Cell* 5 (7):552-562. doi: https://doi.org/10.1007/s13238-014-0072-x.

Çakar, Mehmet M., Jouni Ruupunen, Juan Mangas-Sanchez, William R. Birmingham, Deniz Yildirim, Ossi Turunen, Nicholas J. Turner, Jarkko Valjakka, and Barıs Binay. 2020. "Engineered formate dehydrogenase from Chaetomium thermophilum, a promising enzymatic solution for biotechnical CO2 fixation." *Biotechnol Lett* 42 (11):2251-2262. doi: https://doi.org/10.1007/s10529-020-02937-7.

Çakar, Mervan M., Juan Mangas-Sanchez, William R. Birmingham, Nicholas J. Turner, and Barış Binay. 2018. "Discovery of a new metal and NAD(+)-dependent formate dehydrogenase from Clostridium ljungdahlii." *Prep Biochem Biotechnol* 48 (4):327-334. doi: https://doi.org/10.1080/10826068.2018.1446150.

Calvin, M., and A. A. Benson. 1948. "The Path of Carbon in Photosynthesis." *Science* 107 (2784):476-480. doi: https://doi.org/10.1126/science.107.2784.476.

Calvo-Villamañán, Alicia, Jérome Wong Ng, Rémi Planel, Hervé Ménager, Arthur Chen, Lun Cui, and David Bikard. 2020. "On-target activity predictions enable improved CRISPR–dCas9 screens in bacteria." *Nucleic Acids Research* 48 (11):e64-e64. doi: https://doi.org/10.1093/nar/gkaa294.

Cardinale, S., and A. P. Arkin. 2012. "Contextualizing context for synthetic biology--identifying causes of failure of synthetic biological systems." *Biotechnol J* 7 (7):856-66. doi: https://doi.org/10.1002%2Fbiot.201200085.

Causey, T. B., K. T. Shanmugam, L. P. Yomano, and L. O. Ingram. 2004. "Engineering Escherichia coli for efficient conversion of glucose to pyruvate." *Proc Natl Acad Sci U S A* 101 (8):2235-40. doi: https://doi.org/10.1073/pnas.0308171100.

Causey, T. B., S. Zhou, K. T. Shanmugam, and L. O. Ingram. 2003. "Engineering the metabolism of Escherichia coli W3110 for the conversion of sugar to redox-neutral and oxidized products: Homoacetate production." *Proceedings of the National Academy of Sciences* 100 (3):825-832. doi: https://doi.org/10.1073/pnas.0337684100.

Chance, Ronald E, and Bruce H Frank. 1993. "Research, Development, Production, and Safety of Biosynthetic Human Insulin." *Diabetes Care* 16 (Supplement_3):133-142. doi: https://doi.org/10.2337/diacare.16.3.133.

Chang, D. E., H. C. Jung, J. S. Rhee, and J. G. Pan. 1999. "Homofermentative production of D- or L-lactate in metabolically engineered Escherichia coli RR1." *Appl Environ Microbiol* 65 (4):1384-9. doi: https://doi.org/10.1128/aem.65.4.1384-1389.1999.

Chaudhary, Yatendra S., Thomas W. Woolerton, Christopher S. Allen, Jamie H. Warner, Elizabeth Pierce, Stephen W. Ragsdale, and Fraser A. Armstrong. 2012. "Visible light-driven CO2 reduction by enzyme coupled CdS nanocrystals." *Chemical Communications* 48 (1):58-60. doi: https://doi.org/10.1039%2Fc1cc16107e.

Chen, Frederic Y. H., Hsin-Wei Jung, Chao-Yin Tsuei, and James C. Liao. 2020. "Converting Escherichia coli to a Synthetic Methylotroph Growing Solely on Methanol." *Cell* 182 (4):933-946.e14. doi: https://doi.org/10.1016/j.cell.2020.07.010.

Cherepanov, Peter P., and Wilfried Wackernagel. 1995. "Gene disruption in Escherichia coli: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant." *Gene* 158 (1):9-14. doi: https://doi.org/10.1016/0378-1119(95)00193-A.

Chistoserdova, L. 2011. "Modularity of methylotrophy, revisited." *Environ Microbiol* 13 (10):2603-22. doi: https://doi.org/10.1111/j.1462-2920.2011.02464.x.

Chistoserdova, L., M. Laukel, J. C. Portais, J. A. Vorholt, and M. E. Lidstrom. 2004. "Multiple formate dehydrogenase enzymes in the facultative methylotroph Methylobacterium extorquens AM1 are dispensable for growth on methanol." *J Bacteriol* 186 (1):22-8. doi: https://doi.org/10.1128/jb.186.1.22-28.2004.

Choe, Hyunjun, Jung Min Ha, Jeong Chan Joo, Hyunook Kim, Hye-Jin Yoon, Seonghoon Kim, Sang Hyeon Son, Robert M. Gengan, Seung Taeg Jeon, Rakwoo Chang, Kwang Deog Jung, Yong Hwan Kim, and Hyung Ho Lee. 2015. "Structural insights into the efficient CO2-reducing activity of an NAD-dependent formate dehydrogenase from Thiobacillus sp. KNK65MA." *Acta Crystallogr D Biol Crystallogr* 71 (Pt 2):313-23. doi: https://doi.org/10.1107/s1399004714025474.

Choe, Hyunjun, Jeong Chan Joo, Dae Haeng Cho, Min Hoo Kim, Sang Hyun Lee, Kwang Deog Jung, and Yong Hwan Kim. 2014. "Efficient CO2-reducing activity of NAD-dependent formate dehydrogenase from Thiobacillus sp. KNK65MA for formate production from CO2 gas." *PLoS One* 9 (7):e103111. doi: https://doi.org/10.1371/journal.pone.0103111.

Choi, Jong Hyun, Ki Chang Keum, and Sang Yup Lee. 2006. "Production of recombinant proteins by high cell density culture of Escherichia coli." *Chemical Engineering Science* 61 (3):876-885. doi: https://doi.org/10.1016/j.ces.2005.03.031.

Choudhury, Alaksh, Jacob A Fenster, Reilly G Fankhauser, Joel L Kaar, Olivier Tenaillon, and Ryan T Gill. 2020. "CRISPR/Cas9 recombineering-mediated deep mutational scanning of essential genes in Escherichia coli." *Molecular Systems Biology* 16 (3):e9265. doi: https://doi.org/10.15252/msb.20199265.

Claassens, Nico J., Guillermo Bordanaba-Florit, Charles A. R. Cotton, Alberto De Maria, Max Finger-Bou, Lukas Friedeheim, Natalia Giner-Laguarda, Martí Munar-Palmer, William Newell, Giovanni Scarinci, Jari Verbunt, Stijn T. de Vries, Suzan Yilmaz, and Arren Bar-Even. 2020. "Replacing the Calvin cycle with the reductive glycine pathway in Cupriavidus necator." *Metabolic Engineering* 62:30-41. doi: https://doi.org/10.1016/j.ymben.2020.08.004.

Claassens, Nico Joannes, Irene Sánchez-Andrea, Diana Zita Sousa, and Arren Bar-Even. 2018. "Towards sustainable feedstocks: A guide to electron donors for microbial carbon fixation." *Curr Opin Biotechnol* 50:195-205. doi: https://doi.org/10.1016/j.copbio.2018.01.019.

Cleland, W. Wallace, T. John Andrews, Steven Gutteridge, Fred C. Hartman, and George H. Lorimer. 1998. "Mechanism of Rubisco: The Carbamate as General Base." *Chemical Reviews* 98 (2):549-562. doi: https://doi.org/10.1021/cr970010r.

Costas, A. M., A. K. White, and W. W. Metcalf. 2001. "Purification and characterization of a novel phosphorus-oxidizing enzyme from Pseudomonas stutzeri WM88." *J Biol Chem* 276 (20):17429-36. doi: https://doi.org/10.1074/jbc.m011764200.

Cotton, Charles A. R., Jeffrey S. Douglass, Sven De Causmaecker, Katharina Brinkert, Tanai Cardona, Andrea Fantuzzi, A. William Rutherford, and James W. Murray. 2015. "Photosynthetic Constraints on Fuel from Microbes." *Frontiers in Bioengineering and Biotechnology* 3. doi: https://doi.org/10.3389/fbioe.2015.00036.

Cotton, Charles A. R., Christian Edlich-Muth, and Arren Bar-Even. 2018. "Reinforcing carbon fixation: CO2 reduction replacing and supporting carboxylation." *Current Opinion in Biotechnology* 49:49-56. doi: https://doi.org/10.1016/j.copbio.2017.07.014.

Cravens, Aaron, Osman K. Jamil, Deze Kong, Jonathan T. Sockolosky, and Christina D. Smolke. 2021. "Polymerase-guided base editing enables in vivo mutagenesis and rapid protein engineering." *Nature Communications* 12 (1):1579. doi: https://doi.org/10.1038/s41467-021-21876-z.

Crépin, Lucie, Eric Lombard, and Stéphane E. Guillouet. 2016. "Metabolic engineering of Cupriavidus necator for heterotrophic and autotrophic alka(e)ne production." *Metabolic Engineering* 37:92-101. doi: https://doi.org/10.1016/j.ymben.2016.05.002.

Crowther, G. J., G. Kosály, and M. E. Lidstrom. 2008. "Formate as the main branch point for methylotrophic metabolism in Methylobacterium extorquens AM1." *J Bacteriol* 190 (14):5057-62. doi: https://doi.org/10.1128/jb.00228-08.

Cui, Lun, Antoine Vigouroux, François Rousset, Hugo Varet, Varun Khanna, and David Bikard. 2018. "A CRISPRi screen in E. coli reveals sequence-specific toxicity of dCas9." *Nature Communications* 9 (1):1912. doi: https://doi.org/10.1038/s41467-018-04209-5.

Darst, Russell P., Carolina E. Pardo, Lingbao Ai, Kevin D. Brown, and Michael P. Kladde. 2010. "Bisulfite sequencing of DNA." *Curr Protoc Mol Biol* Chapter 7:Unit 7.9.1-17. doi: https://doi.org/10.1002%2F0471142727.mb0709s91.

Datsenko, Kirill A., and Barry L. Wanner. 2000. "One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products." *Proc Natl Acad Sci U S A* 97 (12):6640-5. doi: https://doi.org/10.1073/pnas.120163297.

Daza, Yolanda A., and John N. Kuhn. 2016. "CO2 conversion by reverse water gas shift catalysis: comparison of catalysts, mechanisms and their consequences for CO2 conversion to liquid fuels." *RSC Advances* 6 (55):49675-49691. doi: https://doi.org/10.1039/C6RA05414E.

de Bok, Frank A., Peter-Leon Hagedoorn, Pedro J. Silva, Wilfred R. Hagen, Emile Schiltz, Kathrin Fritsche, and Alfons J. Stams. 2003. "Two W-containing formate dehydrogenases (CO2-reductases) involved in syntrophic propionate oxidation by Syntrophobacter fumaroxidans." *Eur J Biochem* 270 (11):2476-85. doi: https://doi.org/10.1046/j.1432-1033.2003.03619.x.

Delmas, Valérie A., Nadia Perchat, Oriane Monet, Marion Fouré, Ekatarina Darii, David Roche, Ivan Dubois, Emilie Pateau, Alain Perret, Volker Döring, and Madeleine Bouzon. 2022. "Genetic and

biocatalytic basis of formate dependent growth of Escherichia coli strains evolved in continuous culture." *Metabolic Engineering* 72:200-214. doi: https://doi.org/10.1016/j.ymben.2022.03.010.

Demarre, Gaëlle, Anne-Marie Guérout, Chiho Matsumoto-Mashimo, Dean A. Rowe-Magnus, Philippe Marlière, and Didier Mazel. 2005. "A new family of mobilizable suicide plasmids based on broad host range R388 plasmid (IncW) and RP4 plasmid (IncPα) conjugative machineries and their cognate Escherichia coli host strains." *Research in Microbiology* 156 (2):245-255. doi: https://doi.org/10.1016/j.resmic.2004.09.007.

Depardieu, Florence, and David Bikard. 2020. "Gene silencing with CRISPRi in bacteria and optimization of dCas9 expression levels." *Methods* 172:61-75. doi: https://doi.org/10.1016/j.ymeth.2019.07.024.

Dien, B. S., N. N. Nichols, and R. J. Bothast. 2001. "Recombinant Escherichia coli engineered for production of L-lactic acid from hexose and pentose sugars." *J Ind Microbiol Biotechnol* 27 (4):259-64. doi: https://doi.org/10.1038/sj.jim.7000195.

Döring, Volker, Ekaterina Darii, Oren Yishai, Arren Bar-Even, and Madeleine Bouzon. 2018. "Implementation of a Reductive Route of One-Carbon Assimilation in Escherichia coli through Directed Evolution." *ACS Synthetic Biology* 7 (9):2029-2036. doi: https://doi.org/10.1021/acssynbio.8b00167.

Dousis, Athanasios, Kanchana Ravichandran, Elissa M. Hobert, Melissa J. Moore, and Amy E. Rabideau. 2023. "An engineered T7 RNA polymerase that produces mRNA free of immunostimulatory byproducts." *Nature Biotechnology* 41 (4):560-568. doi: https://doi.org/10.1038/s41587-022-01525-6.

Ducat, Daniel C., J. Abraham Avelar-Rivas, Jeffrey C. Way, and Pamela A. Silver. 2012. "Rerouting Carbon Flux To Enhance Photosynthetic Productivity." *Applied and Environmental Microbiology* 78 (8):2660-2668. doi: https://doi.org/10.1128/AEM.07901-11.

Ducat, Daniel C., Jeffrey C. Way, and Pamela A. Silver. 2011. "Engineering cyanobacteria to generate high-value products." *Trends in Biotechnology* 29 (2):95-103. doi: https://doi-org.insb.bib.cnrs.fr/10.1016/j.tibtech.2010.12.003.

Erhardt, H., S. Steimle, V. Muders, T. Pohl, J. Walter, and T. Friedrich. 2012. "Disruption of individual nuo-genes leads to the formation of partially assembled NADH:ubiquinone oxidoreductase (complex I) in Escherichia coli." *Biochim Biophys Acta* 1817 (6):863-71. doi: https://doi.org/10.1016/j.bbabio.2011.10.008.

Evans, M C, B B Buchanan, and D I Arnon. 1966. "A new ferredoxin-dependent carbon reduction cycle in a photosynthetic bacterium." *Proceedings of the National Academy of Sciences* 55 (4):928-934. doi: https://doi.org/10.1073%2Fpnas.55.4.928.

Fei, Qiang, Michael T. Guarnieri, Ling Tao, Lieve M. L. Laurens, Nancy Dowe, and Philip T. Pienkos. 2014. "Bioconversion of natural gas to liquid fuel: Opportunities and challenges." *Biotechnology Advances* 32 (3):596-614. doi: https://doi.org/10.1016/j.biotechadv.2014.03.011.

Ferry, James G. 1990. "Formate dehydrogenase." *FEMS Microbiology Reviews* 7 (3-4):377-382. doi: https://doi.org/10.1111/j.1574-6968.1990.tb04940.x.

Figueroa, Israel A., Tyler P. Barnum, Pranav Y. Somasekhar, Charlotte I. Carlström, Anna L. Engelbrektson, and John D. Coates. 2018. "Metagenomics-guided analysis of microbial chemolithoautotrophic

phosphite oxidation yields evidence of a seventh natural CO2 fixation pathway." *Proceedings of the National Academy of Sciences* 115 (1):E92-E101. doi: https://doi.org/10.1073/pnas.1715549114.

Figueroa, Israel A., and John D. Coates. 2017. "Microbial Phosphite Oxidation and Its Potential Role in the Global Phosphorus and Carbon Cycles." *Adv Appl Microbiol* 98:93-117. doi: https://doi.org/10.1016/bs.aambs.2016.09.004.

Filonenko, Georgy A., Robbert van Putten, Erik N. Schulpen, Emiel J. M. Hensen, and Evgeny A. Pidko. 2014. "Highly Efficient Reversible Hydrogenation of Carbon Dioxide to Formates Using a Ruthenium PNP-Pincer Catalyst." *ChemCatChem* 6 (6):1526-1530. doi: https://doi.org/10.1002/cctc.201402119.

Förster, Andreas H., and Johannes Gescher. 2014. "Metabolic Engineering of Escherichia coli for Production of Mixed-Acid Fermentation End Products." *Frontiers in Bioengineering and Biotechnology* 2. doi: https://doi.org/10.3389/fbioe.2014.00016.

Forzi, L., and R. G. Sawers. 2007. "Maturation of [NiFe]-hydrogenases in Escherichia coli." *Biometals* 20 (3-4):565-78. doi: https://doi.org/10.1007/s10534-006-9048-5.

Friedebold, Jörg, and Botho Bowien. 1993. "Physiological and biochemical characterization of the soluble formate dehydrogenase, a molybdoenzyme from Alcaligenes eutrophus." *J Bacteriol* 175 (15):4719-28. doi: https://doi.org/10.1128%2Fjb.175.15.4719-4728.1993.

Friedrich, Bärbel, Johannes Fritsch, and Oliver Lenz. 2011. "Oxygen-tolerant hydrogenases in hydrogen-based technologies." *Current Opinion in Biotechnology* 22 (3):358-364. doi: https://doi.org/10.1016/j.copbio.2011.01.006.

Fuchs, GEORG. 2011. "Alternative pathways of carbon dioxide fixation: insights into the early evolution of life?" *Annu Rev Microbiol* 65:631-58. doi: https://doi.org/10.1146/annurev-micro-090110-102801.

Gao, Zhengxu, Hui Zhao, Zhimin Li, Xiaoming Tan, and Xuefeng Lu. 2012. "Photosynthetic production of ethanol from carbon dioxide in genetically engineered cyanobacteria." *Energy & Environmental Science* 5 (12):9857-9865. doi: https://doi.org/10.1039/C2EE22675H.

Garland, Peter B, J. Allan Downie, and Bruce A. Haddock. 1975. "Proton translocation and the respiratory nitrate reductase of Escherichia coli." *Biochemical Journal* 152 (3):547-559. doi: https://doi.org/10.1042/bj1520547.

Genkov, Todor, Moritz Meyer, Howard Griffiths, and Robert J. Spreitzer. 2010. "Functional hybrid rubisco enzymes with plant small subunits and algal large subunits: engineered rbcS cDNA for expression in chlamydomonas." *Journal of Biological Chemistry* 285 (26):19833-19841. doi: https://doi.org/10.1074/jbc.M110.124230.

Ghosh, Dipankar, Ariane Bisaillon, and Patrick C. Hallenbeck. 2013. "Increasing the metabolic capacity of Escherichia coli for hydrogen production through heterologous expression of the Ralstonia eutropha SH operon." *Biotechnology for Biofuels* 6 (1):122. doi: https://doi.org/10.1186/1754-6834-6-122.

Ghrist, Angela C., Gary Heil, and Geroge V. Stauffer. 2001. "GcvR interacts with GcvA to inhibit activation of the Escherichia coli glycine cleavage operon." *Microbiology (Reading)* 147 (Pt 8):2215-2221. doi: https://doi.org/10.1099/00221287-147-8-2215.

Gleizer, Shmuel, Roee Ben-Nissan, Yinon M. Bar-On, Niv Antonovsky, Elad Noor, Yehudit Zohar, Ghil Jona, Eyal Krieger, Melina Shamshoum, Arren Bar-Even, and Ron Milo. 2019. "Conversion of Escherichia coli to Generate All Biomass Carbon from CO2." *Cell* 179 (6):1255-1263.e12. doi: https://doi.org/10.1016/j.cell.2019.11.009.

Gomaa, M. A., L. Al-Haj, and R. M. Abed. 2016. "Metabolic engineering of Cyanobacteria and microalgae for enhanced production of biofuels and high-value products." *J Appl Microbiol* 121 (4):919-31. doi: https://doi.org/10.1111/jam.13232.

Gong, Fuyu, Guoxia Liu, Xiaoyun Zhai, Jie Zhou, Zhen Cai, and Yin Li. 2015. "Quantitative analysis of an engineered CO2-fixing Escherichia coli reveals great potential of heterotrophic CO2 fixation." *Biotechnology for Biofuels* 8 (1):86. doi: https://doi.org/10.1186/s13068-015-0268-1.

Gong, Fuyu, Huawei Zhu, Yanping Zhang, and Yin Li. 2018. "Biological carbon fixation: From natural to synthetic." *Journal of CO2 Utilization* 28:221-227. doi: https://doi.org/10.1016/j.jcou.2018.09.014.

Gonzalez de la Cruz, Jorge, Fabian Machens, Katrin Messerschmidt, and Arren Bar-Even. 2019. "Core Catalysis of the Reductive Glycine Pathway Demonstrated in Yeast." *ACS Synthetic Biology* 8 (5):911-917. doi: https://doi.org/10.1021/acssynbio.8b00464.

Guadalupe-Medina, Víctor, H. Wouter Wisselink, Marijke A. H. Luttik, Erik de Hulster, Jean-Marc Daran, Jack T. Pronk, and Antonius J. A. van Maris. 2013. "Carbon dioxide fixation by Calvin-Cycle enzymes improves ethanol yield in yeast." *Biotechnology for Biofuels* 6 (1):125. doi: https://doi.org/10.1186/1754-6834-6-125.

Guest, J. R., and I. T. Creaghan. 1972. "Lipoamide dehydrogenase mutants of Escherichia coli K 12." *Biochem J* 130 (1):8p. doi: https://doi.org/10.1042%2Fbj1300008p.

Haferkamp, Ilka, Stephan Schmitz-Esser, Nicole Linka, Claude Urbany, Astrid Collingro, Michael Wagner, Matthias Horn, and H. Ekkehard Neuhaus. 2004. "A candidate NAD+ transporter in an intracellular bacterial symbiont related to Chlamydiae." *Nature* 432 (7017):622-625. doi: https://doi.org/10.1038/nature03131.

Hakopian, Sheron, Dimitri Niks, and Russ Hille. 2022. "The air-inactivation of formate dehydrogenase FdsDABG from Cupriavidus necator." *J Inorg Biochem* 231:111788. doi: https://doi.org/10.1016/j.jinorgbio.2022.111788.

Halperin, Shakked O., Connor J. Tou, Eric B. Wong, Cyrus Modavi, David V. Schaffer, and John E. Dueber. 2018. "CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window." *Nature* 560 (7717):248-252. doi: https://doi.org/10.1038/s41586-018-0384-8.

Hartmann, Tobias, and Silke Leimkühler. 2013. "The oxygen-tolerant and NAD+-dependent formate dehydrogenase from Rhodobacter capsulatus is able to catalyze the reduction of CO2 to formate." *Febs j* 280 (23):6083-96. doi: https://doi.org/10.1111/febs.12528.

Hasona, Adnan, Youngnyun Kim, F. G. Healy, L. O. Ingram, and K. T. Shanmugam. 2004. "Pyruvate formate lyase and acetate kinase are essential for anaerobic growth of Escherichia coli on xylose." *J Bacteriol* 186 (22):7593-600. doi: https://doi.org/10.1128/jb.186.22.7593-7600.2004.

Haynes, Chad A., and Ramon Gonzalez. 2014. "Rethinking biological activation of methane and conversion to liquid fuels." *Nature Chemical Biology* 10 (5):331-339. doi: https://doi.org/10.1038/nchembio.1509.

Heßlinger, Christian, Shirley A. Fairhurst, and Gary Sawers. 1998. "Novel keto acid formate-lyase and propionate kinase enzymes are components of an anaerobic pathway in Escherichia coli that degrades L-threonine to propionate." *Molecular Microbiology* 27 (2):477-492. doi: https://doi.org/10.1046/j.1365-2958.1998.00696.x.

Hirota, R., A. Kuroda, J. Kato, and H. Ohtake. 2010. "Bacterial phosphate metabolism and its application to phosphorus recovery and industrial bioprocesses." *J Biosci Bioeng* 109 (5):423-32. doi: https://doi.org/10.1016/j.jbiosc.2009.10.018.

Hirota, Ryuichi, Kenji Abe, Zen-ichiro Katsuura, Reiji Noguchi, Shigeaki Moribe, Kei Motomura, Takenori Ishida, Maxym Alexandrov, Hisakage Funabashi, Takeshi Ikeda, and Akio Kuroda. 2017. "A Novel Biocontainment Strategy Makes Bacterial Growth and Survival Dependent on Phosphite." *Sci Rep* 7:44748. doi: https://doi.org/10.1038/srep44748.

Holo, Helge. 1989. "Chloroflexus aurantiacus secretes 3-hydroxypropionate, a possible intermediate in the assimilation of CO2 and acetate." *Archives of Microbiology* 151 (3):252-256. doi: https://doi.org/10.1007/BF00413138.

Hori, Yoshio, Hidetoshi Wakebe, Toshio Tsukamoto, and Osamu Koga. 1994. "Electrocatalytic process of CO selectivity in electrochemical reduction of CO2 at metal electrodes in aqueous media." *Electrochimica Acta* 39 (11):1833-1839. doi: https://doi.org/10.1016/0013-4686(94)85172-7.

Inoue, Tooru, Akira Fujishima, Satoshi Konishi, and Kenichi Honda. 1979. "Photoelectrocatalytic reduction of carbon dioxide in aqueous suspensions of semiconductor powders." *Nature* 277 (5698):637-638. doi: https://doi.org/10.1038/277637a0.

Ishaq, Haris, Ibrahim Dincer, and Curran Crawford. 2022. "A review on hydrogen production and utilization: Challenges and opportunities." *International Journal of Hydrogen Energy* 47 (62):26238-26264. doi: https://doi.org/10.1016/j.ijhydene.2021.11.149.

Ishikawa, Chie, Tomoko Hatanaka, Shuji Misoo, Chikahiro Miyake, and Hiroshi Fukayama. 2011. "Functional Incorporation of Sorghum Small Subunit Increases the Catalytic Turnover Rate of Rubisco in Transgenic Rice " *Plant Physiology* 156 (3):1603-1611. doi: https://doi.org/10.1104/pp.111.177030.

Joó, Ferenc. 2008. "Breakthroughs in Hydrogen Storage—Formic Acid as a Sustainable Storage Material for Hydrogen." *ChemSusChem* 1 (10):805-808. doi: https://doi.org/10.1002/cssc.200800133.

Jormakka, Mika, Bernadette Byrne, and So Iwata. 2003. "Formate dehydrogenase - a versatile enzyme in changing environments." *Curr Opin Struct Biol* 13 (4):418-23. doi: https://doi.org/10.1016/s0959-440x(03)00098-8.

Jouny, Matthew, Wesley Luc, and Feng Jiao. 2018. "General Techno-Economic Analysis of CO2 Electrolysis Systems." *Industrial & Engineering Chemistry Research* 57 (6):2165-2177. doi: https://doi.org/10.1021/acs.iecr.7b03514.

Jung, Tae-Sung, Hock Chuan Yeo, Satty G. Reddy, Wan-Sup Cho, and Dong-Yup Lee. 2009. "WEbcoli: an interactive and asynchronous web application for in silico design and analysis of genome-scale E.coli model." *Bioinformatics* 25 (21):2850-2852. doi: https://doi.org/10.1093/bioinformatics/btp496.

Kammel, M., and R. G. Sawers. 2022. "The Autonomous Glycyl Radical Protein GrcA Restores Activity to Inactive Full-Length Pyruvate Formate-Lyase In Vivo." *J Bacteriol* 204 (5):e0007022. doi: https://doi.org/10.1128%2Fjb.00070-22.

Kammel, Michelle, Constanze Pinske, and R. Gary Sawers. 2022. "FocA and its central role in fine-tuning pH homeostasis of enterobacterial formate metabolism." *Microbiology* 168 (10). doi: https://doi.org/10.1099/mic.0.001253.

Kang, Minjeong, Kangsan Kim, Donghui Choe, Suhyung Cho, Sun Chang Kim, Bernhard Palsson, and Byung-Kwan Cho. 2019. "Inactivation of a Mismatch-Repair System Diversifies Genotypic Landscape of Escherichia coli During Adaptive Laboratory Evolution." *Frontiers in Microbiology* 10. doi: https://doi.org/10.3389/fmicb.2019.01845.

Kapdan, Ilgi Karapinar, and Fikret Kargi. 2006. "Bio-hydrogen production from waste materials." *Enzyme and Microbial Technology* 38 (5):569-582. doi: https://doi.org/10.1016/j.enzmictec.2005.09.015.

Kaper, James B., James P. Nataro, and Harry L. T. Mobley. 2004. "Pathogenic Escherichia coli." *Nature Reviews Microbiology* 2 (2):123-140. doi: https://doi.org/10.1038/nrmicro818.

Keller, Matthew W., Gerrit J. Schut, Gina L. Lipscomb, Angeli L. Menon, Ifeyinwa J. Iwuchukwu, Therese T. Leuko, Michael P. Thorgersen, William J. Nixon, Aaron S. Hawkins, Robert M. Kelly, and Michael W. W. Adams. 2013. "Exploiting microbial hyperthermophilicity to produce an industrial chemical, using hydrogen and carbon dioxide." *Proceedings of the National Academy of Sciences* 110 (15):5840-5845. doi: https://doi.org/10.1073/pnas.1222607110.

Kim, Seohyoung, Néstor Giraldo, Vittorio Rainaldi, Fabian Machens, Florent Collas, Armin Kubis, Frank Kensy, Arren Bar-Even, and Steffen N. Lindner. 2023. "Optimizing E. coli as a formatotrophic platform for bioproduction via the reductive glycine pathway." *Front Bioeng Biotechnol* 11:1091899. doi: https://doi.org/10.3389/fbioe.2023.1091899.

Kim, Seohyoung, Steffen N. Lindner, Selçuk Aslan, Oren Yishai, Sebastian Wenk, Karin Schann, and Arren Bar-Even. 2020. "Growth of E. coli on formate and methanol via the reductive glycine pathway." *Nature Chemical Biology* 16 (5):538-545. doi: https://doi.org/10.1038/s41589-020-0473-5.

Kim, Youngnyun, L. O. Ingram, and K. T. Shanmugam. 2007. "Construction of an Escherichia coli K-12 mutant for homoethanologenic fermentation of glucose or xylose without foreign genes." *Appl Environ Microbiol* 73 (6):1766-71. doi: https://doi.org/10.1128%2FAEM.02456-06.

Knappe, Joachim, Hans P. Blaschkowski, Peter Gröbner, and Thomas Schmttt. 1974. "Pyruvate Formate-Lyase of Escherichia coli: the Acetyl-Enzyme Intermediate." *European Journal of Biochemistry* 50 (1):253-263. doi: https://doi.org/10.1111/j.1432-1033.1974.tb03894.x.

Kochetkov, S. N., E. E. Rusakova, and V. L. Tunitskaya. 1998. "Recent studies of T7 RNA polymerase mechanism." *FEBS Letters* 440 (3):264-267. doi: https://doi.org/10.1016/S0014-5793(98)01484-7.

Kothandaraman, Jotheeswari, Alain Goeppert, Miklos Czaun, George A. Olah, and G. K. Surya Prakash. 2016. "Conversion of CO2 from Air into Methanol Using a Polyamine and a Homogeneous Ruthenium Catalyst." *Journal of the American Chemical Society* 138 (3):778-781. doi: https://doi.org/10.1021/jacs.5b12354.

Kuk, Su Keun, Krishnasamy Gopinath, Raushan K. Singh, Tae-Doo Kim, Youngjun Lee, Woo Seok Choi, Jung-Kul Lee, and Chan Beum Park. 2019. "NADH-Free Electroenzymatic Reduction of CO2 by

Conductive Hydrogel-Conjugated Formate Dehydrogenase." *ACS Catalysis* 9 (6):5584-5589. doi: https://doi.org/10.1021/acscatal.9b00127.

Lalucat, Jorge, Antoni Bennasar, Rafael Bosch, Elena García-Valdés, and Norberto J. Palleroni. 2006. "Biology of Pseudomonas stutzeri." *Microbiol Mol Biol Rev* 70 (2):510-47. doi: https://doi.org/10.1128%2FMMBR.00047-05.

Lamont, Ciaran M., and Frank Sargent. 2017. "Design and characterisation of synthetic operons for biohydrogen technology." *Arch Microbiol* 199 (3):495-503. doi: https://doi.org/10.1007%2Fs00203-016-1322-5.

Lan, Ethan I., and James C. Liao. 2012. "ATP drives direct photosynthetic production of 1-butanol in cyanobacteria." *Proceedings of the National Academy of Sciences* 109 (16):6018-6023. doi: https://doi.org/10.1073/pnas.1200074109.

Lee, Heewook, Ellen Popodi, Haixu Tang, and Patricia L. Foster. 2012. "Rate and molecular spectrum of spontaneous mutations in the bacterium Escherichia coli as determined by whole-genome sequencing." *Proceedings of the National Academy of Sciences* 109 (41):E2774-E2783. doi: https://doi.org/10.1073/pnas.1210309109.

Lee, Jehan, Myeong-Kyun Shin, Dong-Kyun Ryu, Seahee Kim, and Wang-Shick Ryu. 2010. "Insertion and deletion mutagenesis by overlap extension PCR." *Methods Mol Biol* 634:137-46. doi: https://doi.org/10.1007/978-1-60761-652-8_10.

Leonardo, Michael R., Yves Dailly, and David P. Clark. 1996. "Role of NAD in regulating the adhE gene of Escherichia coli." *Journal of Bacteriology* 178 (20):6013-6018. doi: https://doi.org/10.1128/jb.178.20.6013-6018.1996.

Lerman, Joshua A., Daniel R. Hyduke, Haythem Latif, Vasiliy A. Portnoy, Nathan E. Lewis, Jeffrey D. Orth, Alexandra C. Schrimpe-Rutledge, Richard D. Smith, Joshua N. Adkins, Karsten Zengler, and Bernhard O. Palsson. 2012. "In silico method for modelling metabolism and gene product expression at genome scale." *Nature Communications* 3 (1):929. doi: https://doi.org/10.1038/ncomms1928.

Li, Han, Paul H. Opgenorth, David G. Wernick, Steve Rogers, Tung-Yun Wu, Wendy Higashide, Peter Malati, Yi-Xin Huo, Kwang Myung Cho, and James C. Liao. 2012. "Integrated Electromicrobial Conversion of CO2 to Higher Alcohols." *Science* 335 (6076):1596-1596. doi: https://doi.org/10.1126/science.1217643.

Li, Yun-Jie, Miao-Miao Wang, Ya-Wei Chen, Meng Wang, Li-Hai Fan, and Tian-Wei Tan. 2017. "Engineered yeast with a CO2-fixation pathway to improve the bio-ethanol production from xylose-mixed sugars." *Scientific Reports* 7 (1):43875. doi: https://doi.org/10.1038/srep43875.

Liew, FungMin, Michael E. Martin, Ryan C. Tappel, Björn D. Heijstra, Christophe Mihalcea, and Michael Köpke. 2016. "Gas Fermentation—A Flexible Platform for Commercial Scale Production of Low-Carbon-Fuels and Chemicals from Waste and Renewable Feedstocks." *Frontiers in Microbiology* 7. doi: https://doi.org/10.3389/fmicb.2016.00694.

Liu, Xinyao, Jie Sheng, and Roy Curtiss III. 2011. "Fatty acid production in genetically modified cyanobacteria." *Proceedings of the National Academy of Sciences* 108 (17):6899-6904. doi: https://doi.org/10.1073/pnas.1103014108.

Ljungdahl, Lars G., and Jan R. Andreesen. 1975. "Tungsten, a component of active formate dehydrogenase from Clostridium thermoaceticum." *FEBS Letters* 54 (2):279-282. doi: https://doi.org/10.1016/0014-5793(75)80092-5.

Ljungdhal, Lars G. 1986. "The Autotrophic Pathway of Acetate Synthesis in Acetogenic Bacteria." *Annual Review of Microbiology* 40 (1):415-450. doi: https://doi.org/10.1146/annurev.mi.40.100186.002215.

Maeda, Toshinari, Viviana Sanchez-Torres, and Thomas K. Wood. 2007. "Escherichia coli hydrogenase 3 is a reversible enzyme possessing hydrogen uptake and synthesis activities." *Applied Microbiology and Biotechnology* 76 (5):1035-1042. doi: https://doi.org/10.1007/s00253-007-1086-6.

Maeda, Toshinari, Kien Trung Tran, Ryota Yamasaki, and Thomas K. Wood. 2018. "Current state and perspectives in hydrogen production by Escherichia coli: roles of hydrogenases in glucose or glycerol metabolism." *Appl Microbiol Biotechnol* 102 (5):2041-2050. doi: https://doi.org/10.1007/s00253-018-8752-8.

Maia, Luisa B., Luis Fonseca, Isabel Moura, and José J. G. Moura. 2016. "Reduction of Carbon Dioxide by a Molybdenum-Containing Formate Dehydrogenase: A Kinetic and Mechanistic Study." *Journal of the American Chemical Society* 138 (28):8834-8846. doi: https://doi.org/10.1021/jacs.6b03941.

Maia, Luisa B., José J. G. Moura, and Isabel Moura. 2015. "Molybdenum and tungsten-dependent formate dehydrogenases." *J Biol Inorg Chem* 20 (2):287-309. doi: https://doi.org/10.1007/s00775-014-1218-2.

Männistö, Riina H., A. Marika Grahn, Dennis H. Bamford, and Jaana K. Bamford. 2003. "Transcription of bacteriophage PM2 involves phage-encoded regulators of heterologous origin." *J Bacteriol* 185 (11):3278-87. doi: https://doi.org/10.1128/jb.185.11.3278-3287.2003.

Marbán, Gregorio, and Teresa Valdés-Solís. 2007. "Towards the hydrogen economy?" *International Journal of Hydrogen Energy* 32 (12):1625-1637. doi: https://doi.org/10.1016/j.ijhydene.2006.12.017.

Martín, Antonio J., Gastón O. Larrazábal, and Javier Pérez-Ramírez. 2015. "Towards sustainable fuels and chemicals through the electrochemical reduction of CO2: lessons from water electrolysis." *Green Chemistry* 17 (12):5114-5130. doi: https://doi.org/10.1039/C5GC01893E.

Matteo, Forloni, Y. Liu Alex, and Wajapeyee Narendra. 2018. "Random Mutagenesis Using Error-Prone DNA Polymerases." *Cold Spring Harbor Protocols* 2018. doi: https://doi.org/10.1101/pdb.prot097741.

Mattozzi, Matthew d, Marika Ziesack, Mathias J. Voges, Pamela A. Silver, and Jeffrey C. Way. 2013. "Expression of the sub-pathways of the Chloroflexus aurantiacus 3-hydroxypropionate carbon fixation bicycle in E. coli: Toward horizontal transfer of autotrophic growth." *Metabolic Engineering* 16:130-139. doi: https://doi.org/10.1016/j.ymben.2013.01.005.

McDowall, Jennifer S., Bonnie J. Murphy, Michael Haumann, Tracy Palmer, Fraser A. Armstrong, and Frank Sargent. 2014. "Bacterial formate hydrogenlyase complex." *Proceedings of the National Academy of Sciences* 111 (38):E3948-E3956. doi: https://doi.org/10.1073/pnas.1407927111.

Mellmann, Dörthe, Peter Sponholz, Henrik Junge, and Matthias Beller. 2016. "Formic acid as a hydrogen storage material – development of homogeneous catalysts for selective hydrogen release." *Chemical Society Reviews* 45 (14):3954-3988. doi: https://doi.org/10.1039/C5CS00618J.

Meneghello, Marta, Ana Rita Oliveira, Aurore Jacq-Bailly, Inês A. C. Pereira, Christophe Léger, and Vincent Fourmond. 2021. "Formate Dehydrogenases Reduce CO2 Rather than HCO3−: An Electrochemical Demonstration." *Angewandte Chemie International Edition* 60 (18):9964-9967. doi: https://doi.org/10.1002/anie.202101167.

Menon, Nanda K., Corinne Y. Chatelus, Marie Dervartanian, Julia C. Wendt, K. T. Shanmugam, Harry D. Peck, Jr., and Allan E. Przybyla. 1994. "Cloning, sequencing, and mutational analysis of the hyb operon encoding Escherichia coli hydrogenase 2." *J Bacteriol* 176 (14):4416-23. doi: https://doi.org/10.1128/jb.176.14.4416-4423.1994.

Menon, Nanda K., J. Robbins, Julia C. Wendt, K. T. Shanmugam, and Allan E. Przybyla. 1991. "Mutational analysis and characterization of the Escherichia coli hya operon, which encodes [NiFe] hydrogenase 1." *J Bacteriol* 173 (15):4851-61. doi: https://doi.org/10.1128/jb.173.15.4851-4861.1991.

Metcalf, William W., Weihong Jiang, and Barry L. Wanner. 1994. "Use of the rep technique for allele replacement to construct new Escherichia coli hosts for maintenance of R6K gamma origin plasmids at different copy numbers." *Gene* 138 (1-2):1-7. doi: https://doi.org/10.1016/0378-1119(94)90776-5.

Metcalf, William W., and Barry L. Wanner. 1991. "Involvement of the Escherichia coli phn (psiD) gene cluster in assimilation of phosphorus in the form of phosphonates, phosphite, Pi esters, and Pi." *J Bacteriol* 173 (2):587-600. doi: https://doi.org/10.1128/jb.173.2.587-600.1991.

Metcalf, William W., and Barry L. Wanner. 1993. "Mutational analysis of an Escherichia coli fourteen-gene operon for phosphonate degradation, using TnphoA' elements." *J Bacteriol* 175 (11):3430-42. doi: https://doi.org/10.1128/jb.175.11.3430-3442.1993.

Min, Xiaoquan, and Matthew W. Kanan. 2015. "Pd-Catalyzed Electrohydrogenation of Carbon Dioxide to Formate: High Mass Activity at Low Overpotential and Identification of the Deactivation Pathway." *Journal of the American Chemical Society* 137 (14):4701-4708. doi: https://doi.org/10.1021/ja511890h.

Mirzoyan, Satenik, Pablo Maria Romero-Pareja, Maria Dolores Coello, Armen Trchounian, and Karen Trchounian. 2017. "Evidence for hydrogenase-4 catalyzed biohydrogen production in Escherichia coli." *International Journal of Hydrogen Energy* 42 (34):21697-21703. doi: https://doi.org/10.1016/j.ijhydene.2017.07.126.

Moon, Myounghoon, Gwon Woo Park, Joon-Pyo Lee, Jin-Suk Lee, and Kyoungseon Min. 2022. "Recombinant expression and characterization of formate dehydrogenase from Clostridium ljungdahlii (ClFDH) as CO2 reductase for converting CO2 to formate." *Journal of CO2 Utilization* 57:101876. doi: https://doi.org/10.1016/j.jcou.2021.101876.

Moore, Christopher L., Louis J. Papa, 3rd, and Matthew D. Shoulders. 2018. "A Processive Protein Chimera Introduces Mutations across Defined DNA Regions In Vivo." *J Am Chem Soc* 140 (37):11560-11564. doi: https://doi.org/10.1021/jacs.8b04001.

Moore, Simon J., Hung-En Lai, Richard J. R. Kelwick, Soo Mei Chee, David J. Bell, Karen Marie Polizzi, and Paul S. Freemont. 2016. "EcoFlex: A Multifunctional MoClo Kit for E. coli Synthetic Biology." *ACS Synthetic Biology* 5 (10):1059-1069. doi: https://doi.org/10.1021/acssynbio.6b00031.

Müller, Jana, Daniel MacEachran, Helcio Burd, Noppadon Sathitsuksanoh, Changhao Bi, Yi-Chun Yeh, Taek Soon Lee, Nathan J. Hillson, Swapnil R. Chhabra, Steven W. Singer, and Harry R. Beller. 2013. "Engineering of Ralstonia eutropha H16 for Autotrophic and Heterotrophic Production of Methyl Ketones." *Applied and Environmental Microbiology* 79 (14):4433-4439. doi: https://doi.org/10.1128/AEM.00973-13.

Mutzel, Rupert, and Philippe Marliere. 2004. Method and device for selecting accelerated proliferation of living cells in suspension. Google Patents.

Nanba, H., Y. Takaoka, and J. Hasegawa. 2003. "Purification and characterization of an alpha-haloketone-resistant formate dehydrogenase from Thiobacillus sp. strain KNK65MA, and cloning of the gene." *Biosci Biotechnol Biochem* 67 (10):2145-53. doi: https://doi.org/10.1271/bbb.67.2145.

Nangle, Shannon N., Marika Ziesack, Sarabeth Buckley, Disha Trivedi, Daniel M. Loh, Daniel G. Nocera, and Pamela A. Silver. 2020. "Valorization of CO2 through lithoautotrophic production of sustainable chemicals in Cupriavidus necator." *Metabolic Engineering* 62:207-220. doi: https://doi.org/10.1016/j.ymben.2020.09.002.

Nielsen, Christian Førgaard, Lene Lange, and Anne S. Meyer. 2019. "Classification and enzyme kinetics of formate dehydrogenases for biomanufacturing via CO2 utilization." *Biotechnology Advances* 37 (7):107408. doi: https://doi.org/10.1016/j.biotechadv.2019.06.007.

Nikolaidis, Pavlos, and Andreas Poullikkas. 2017. "A comparative overview of hydrogen production processes." *Renewable & Sustainable Energy Reviews* 67:597-611. doi: https://doi.org/10.1016/j.rser.2016.09.044.

Niks, D., J. Duvvuru, M. Escalona, and R. Hille. 2016. "Spectroscopic and Kinetic Properties of the Molybdenum-containing, NAD+-dependent Formate Dehydrogenase from Ralstonia eutropha." *J Biol Chem* 291 (3):1162-74. doi: https://doi.org/10.1074/jbc.m115.688457.

Norrander, Jan, Tomas Kempe, and Joachim Messing. 1983. "Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis." *Gene* 26 (1):101-6. doi: https://doi.org/10.1016/0378-1119(83)90040-9.

Nyerges, Ákos, Bálint Csörgő, István Nagy, Balázs Bálint, Péter Bihari, Viktória Lázár, Gábor Apjok, Kinga Umenhoffer, Balázs Bogos, György Pósfai, and Csaba Pál. 2016. "A highly precise and portable genome engineering method allows comparison of mutational effects across bacterial species." *Proceedings of the National Academy of Sciences* 113 (9):2502-2507. doi: https://doi.org/10.1073/pnas.1520040113.

Oh, Jeong-Il, and Botho Bowien. 1998. "Structural Analysis of the fds Operon Encoding the NAD+-linked Formate Dehydrogenase of Ralstonia eutropha *." *Journal of Biological Chemistry* 273 (41):26349-26360. doi: https://doi.org/10.1074/jbc.273.41.26349.

Oliver, John W. K., and Shota Atsumi. 2014. "Metabolic design for cyanobacterial chemical synthesis." *Photosynthesis Research* 120 (3):249-261. doi: https://doi.org/10.1007/s11120-014-9997-4.

Palsson, Bernhard Ø. 2015. "Metabolism in Escherichia coli." In *Systems Biology: Constraint-based Reconstruction and Analysis*, edited by Bernhard Ø Palsson, 50-74. Cambridge: Cambridge University Press.

Panich, Justin, Bonnie Fong, and Steven W. Singer. 2021. "Metabolic Engineering of Cupriavidus necator H16 for Sustainable Biofuels from CO2." *Trends in Biotechnology* 39 (4):412-424. doi: https://doi.org/10.1016/j.tibtech.2021.01.001.

Park, Hyojin, and Seokhee Kim. 2021. "Gene-specific mutagenesis enables rapid continuous evolution of enzymes in vivo." *Nucleic Acids Res* 49 (6):e32. doi: https://doi.org/10.1093/nar/gkaa1231.

Park, Jin Hwan, Kwang Ho Lee, Tae Yong Kim, and Sang Yup Lee. 2007. "Metabolic engineering of Escherichia coli for the production of L-valine based on transcriptome analysis and in silico gene knockout simulation." *Proceedings of the National Academy of Sciences* 104 (19):7797-7802. doi: https://doi.org/10.1073/pnas.0702609104.

Parvez, Ashak Mahmud, Muhammad T. Afzal, Thayne George Victor Hebb, and Max Schmid. 2020. "Utilization of CO2 in thermochemical conversion of biomass for enhanced product properties: A review." *Journal of CO2 Utilization* 40:101217. doi: https://doi.org/10.1016/j.jcou.2020.101217.

Perchat, Nadia, Pierre-Loïc Saaidi, Ekaterina Darii, Christine Pellé, Jean-Louis Petit, Marielle Besnard-Gonnet, Véronique de Berardinis, Maeva Dupont, Alexandra Gimbernat, Marcel Salanoubat, Cécile Fischer, and Alain Perret. 2018. "Elucidation of the trigonelline degradation pathway reveals previously undescribed enzymes and metabolites." *Proceedings of the National Academy of Sciences* 115 (19):E4358-E4367. doi: https://doi.org/10.1073/pnas.1722368115.

Pettit, F. H., and L. J. Reed. 1967. "Alpha-keto acid dehydrogenase complexes. 8. Comparison of dihydrolipoyl dehydrogenases from pyruvate and alpha-ketoglutarate dehydrogenase complexes of Escherichia coli." *Proc Natl Acad Sci U S A* 58 (3):1126-30. doi: https://doi.org/10.1073/pnas.58.3.1126.

Platt, R., C. Drescher, S. K. Park, and G. J. Phillips. 2000. "Genetic system for reversible integration of DNA constructs and lacZ gene fusions into the Escherichia coli chromosome." *Plasmid* 43 (1):12-23. doi: https://doi.org/10.1006/plas.1999.1433.

Pohl, Thomas, Theresa Bauer, Katerina Dörner, Stefan Stolpe, Philipp Sell, Georg Zocher, and Thorsten Friedrich. 2007. "Iron−Sulfur Cluster N7 of the NADH:Ubiquinone Oxidoreductase (Complex I) Is Essential for Stability but Not Involved in Electron Transfer." *Biochemistry* 46 (22):6588-6596. doi: https://doi.org/10.1021/bi700371c.

Pohl, Thomas, Mareike Uhlmann, Miriam Kaufenstein, and Thorsten Friedrich. 2007. "Lambda Red-Mediated Mutagenesis and Efficient Large Scale Affinity Purification of the Escherichia coli NADH:Ubiquinone Oxidoreductase (Complex I)." *Biochemistry* 46 (37):10694-10702. doi: https://doi.org/10.1021/bi701057t.

Pohlmann, Anne, Wolfgang Florian Fricke, Frank Reinecke, Bernhard Kusian, Heiko Liesegang, Rainer Cramm, Thomas Eitinger, Christian Ewering, Markus Pötter, Edward Schwartz, Axel Strittmatter, Ingo Voß, Gerhard Gottschalk, Alexander Steinbüchel, Bärbel Friedrich, and Botho Bowien. 2006. "Genome sequence of the bioplastic-producing "Knallgas" bacterium Ralstonia eutropha H16." *Nature Biotechnology* 24 (10):1257-1262. doi: https://doi.org/10.1038/nbt1244.

Popov, Vladimir O., and Victor S. Lamzin. 1994. "NAD(+)-dependent formate dehydrogenase." *Biochem J* 301 (Pt 3) (Pt 3):625-43. doi: https://doi.org/10.1042/bj3010625.

Poteete, Anthony R. 2001. "What makes the bacteriophage λ Red system useful for genetic engineering: molecular mechanism and biological function." *FEMS Microbiology Letters* 201 (1):9-14. doi: https://doi.org/10.1111/j.1574-6968.2001.tb10725.x.

Qiu, J., J. R. Swartz, and G. Georgiou. 1998. "Expression of active human tissue-type plasminogen activator in Escherichia coli." *Appl Environ Microbiol* 64 (12):4891-6. doi: https://doi.org/10.1128%2Faem.64.12.4891-4896.1998.

Quan, Jiayuan, and Jingdong Tian. 2011. "Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries." *Nature Protocols* 6 (2):242-251. doi: https://doi.org/10.1038/nprot.2010.181.

Radon, Christin, Gerd Mittelstadt, Benjamin R. Duffus, Jörg Burger, Tobias Hartmann, Thorsten Mielke, Christian Teutloff, Silke Leimkuhler, and Petra Wendler. 2020. "Cryo-EM structures reveal intricate Fe-S cluster arrangement and charging in Rhodobacter capsulatus formate dehydrogenase." *Nat Commun* 11 (1):1912. doi: https://doi.org/10.1038/s41467-020-15614-0.

Ragsdale, Stephen W., and Elizabeth Pierce. 2008. "Acetogenesis and the Wood–Ljungdahl pathway of CO2 fixation." *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1784 (12):1873-1898. doi: https://doi.org/10.1016/j.bbapap.2008.08.012.

Ramiro, Almudena R., Pete Stavropoulos, Mila Jankovic, and Michel C. Nussenzweig. 2003. "Transcription enhances AID-mediated cytidine deamination by exposing single-stranded DNA on the nontemplate strand." *Nature Immunology* 4 (5):452-456. doi: https://doi.org/10.1038/ni920.

Reda, T., C. M. Plugge, N. J. Abram, and J. Hirst. 2008. "Reversible interconversion of carbon dioxide and formate by an electroactive enzyme." *Proc Natl Acad Sci U S A* 105 (31):10654-8. doi: https://doi.org/10.1073/pnas.0801290105.

Richter, K., and J. Gescher. 2012. "The molecular toolbox for chromosomal heterologous multiprotein expression in Escherichia coli." *Biochem Soc Trans* 40 (6):1222-6. doi: https://doi.org/10.1042/bst20120143.

Rosano, Germán L., and Eduardo A. Ceccarelli. 2014. "Recombinant protein expression in Escherichia coli: advances and challenges." *Frontiers in Microbiology* 5. doi: https://doi.org/10.3389/fmicb.2014.00172.

Rostain, William, Theophile Grebert, Danylo Vyhovskyi, Paula Thiel Pizarro, Gatwa Tshinsele-Van Bellingen, Lun Cui, and David Bikard. 2023. "Cas9 off-target binding to the promoter of bacterial genes leads to silencing and toxicity." *Nucleic Acids Research* 51 (7):3485-3496. doi: https://doi.org/10.1093/nar/gkad170.

Roy, Soumyabrata, Arjun Cherevotan, and Sebastian C. Peter. 2018. "Thermochemical CO2 Hydrogenation to Single Carbon Products: Scientific and Technological Challenges." *ACS Energy Letters* 3 (8):1938-1966. doi: https://doi.org/10.1021/acsenergylett.8b00740.

Ruschig, U., U. Müller, P. Willnow, and T. Höpner. 1976. "CO2 reduction to formate by NADH catalysed by formate dehydrogenase from Pseudomonas oxalaticus." *Eur J Biochem* 70 (2):325-30. doi: https://doi.org/10.1111/j.1432-1033.1976.tb11021.x.

Sage, Rowan F. 2002. "Variation in the kcat of Rubisco in C3 and C4 plants and some implications for photosynthetic performance at high and low temperature." *Journal of Experimental Botany* 53 (369):609-620. doi: https://doi.org/10.1093/jexbot/53.369.609.

Sánchez-Andrea, Irene, Iame Alves Guedes, Bastian Hornung, Sjef Boeren, Christopher E. Lawson, Diana Z. Sousa, Arren Bar-Even, Nico J. Claassens, and Alfons J. M. Stams. 2020. "The reductive glycine pathway allows autotrophic growth of Desulfovibrio desulfuricans." *Nature Communications* 11 (1):5090. doi: https://doi.org/10.1038/s41467-020-18906-7.

Santos Correa, Sulamita, Junia Schultz, Kyle J. Lauersen, and Alexandre Soares Rosado. 2023. "Natural carbon fixation and advances in synthetic engineering for redesigning and creating new fixation pathways." *Journal of Advanced Research* 47:75-92. doi: https://doi.org/10.1016/j.jare.2022.07.011.

Sarmah, M. K., T. P. Singh, P. Kalita, and A. Dewan. 2023. "Sustainable hydrogen generation and storage - a review." *RSC Adv* 13 (36):25253-25275. doi: https://doi.org/10.1039/D3RA04148D.

Sato, Ryohei, and Yutaka Amao. 2020. "Can formate dehydrogenase from Candida boidinii catalytically reduce carbon dioxide, bicarbonate, or carbonate to formate?" *New Journal of Chemistry* 44 (28):11922-11926. doi: https://doi.org/10.1039/D0NJ01183E.

Sawers, G. 1994. "The hydrogenases and formate dehydrogenases of Escherichia coli." *Antonie Van Leeuwenhoek* 66 (1-3):57-88. doi: https://doi.org/10.1007/bf00871633.

Schiffels, Johannes, Olaf Pinkenburg, Maximilian Schelden, El-Hussiny A. A. Aboulnaga, Marcus E. M. Baumann, and Thorsten Selmer. 2013. "An Innovative Cloning Platform Enables Large-Scale Production and Maturation of an Oxygen-Tolerant [NiFe]-Hydrogenase from Cupriavidus necator in Escherichia coli." *PLOS ONE* 8 (7):e68812. doi: https://doi.org/10.1371/journal.pone.0068812.

Schink, Bernhard, Volker Thiemann, Heike Laue, and Michael W. Friedrich. 2002. "Desulfotignum phosphitoxidans sp. nov., a new marine sulfate reducer that oxidizes phosphite to phosphate." *Arch Microbiol* 177 (5):381-91. doi: https://link.springer.com/article/10.1007/s00203-002-0402-x#citeas.

Schrader, Jens, Martin Schilling, Dirk Holtmann, Dieter Sell, Murillo Villela Filho, Achim Marx, and Julia A. Vorholt. 2009. "Methanol-based industrial biotechnology: current status and future perspectives of methylotrophic bacteria." *Trends in Biotechnology* 27 (2):107-115. doi: https://doi.org/10.1016/j.tibtech.2008.10.009.

Schuchmann, K., and V. Müller. 2013. "Direct and Reversible Hydrogenation of CO2 to Formate by a Bacterial Carbon Dioxide Reductase." *Science* 342 (6164):1382-1385. doi: https://doi.org/10.1126/science.1244758.

Schulz, Marion, Anne Berger, Ivan Dubois, Valérie Delmas, Mélodie Cadillon, Madeleine Bouzon, and Volker Döring. 2023. "Functional expression of a Mo-dependent formate dehydrogenase in Escherichia coli under aerobic conditions." *bioRxiv*:2023.10.27.564357. doi: 10.1101/2023.10.27.564357.

Schwartz, Edward, Anke Henne, Rainer Cramm, Thomas Eitinger, Bärbel Friedrich, and Gerhard Gottschalk. 2003. "Complete Nucleotide Sequence of pHG1: A Ralstonia eutropha H16 Megaplasmid Encoding Key Enzymes of H2-based Lithoautotrophy and Anaerobiosis." *Journal of Molecular Biology* 332 (2):369-383. doi: https://doi.org/10.1016/S0022-2836(03)00894-5.

Schwarz, Fabian M., Kai Schuchmann, and Volker Müller. 2018. "Hydrogenation of CO2 at ambient pressure catalyzed by a highly active thermostable biocatalyst." *Biotechnology for Biofuels* 11 (1):237. doi: https://doi.org/10.1186/s13068-018-1236-3.

Serafini, Martina, Federica Mariani, Francesco Basile, Erika Scavetta, and Domenica Tonelli. 2023. "From Traditional to New Benchmark Catalysts for CO(2) Electroreduction." *Nanomaterials (Basel)* 13 (11). doi: https://doi.org/10.3390/nano13111723.

Severinov, Konstantin. 2001. "T7 RNA polymerase transcription complex: What you see is not what you get." *Proceedings of the National Academy of Sciences* 98 (1):5-7. doi: https://doi.org/10.1073%2Fpnas.021535298.

Sherif, S. A., Frano Barbir, and T. N. Veziroglu. 2005. "Towards a Hydrogen Economy." *The Electricity Journal* 18 (6):62-76. doi: https://doi.org/10.1016/j.tej.2005.06.003.

Shi, Jiafu, Yanjun Jiang, Zhongyi Jiang, Xueyan Wang, Xiaoli Wang, Shaohua Zhang, Pingping Han, and Chen Yang. 2015. "Enzymatic conversion of carbon dioxide." *Chemical Society Reviews* 44 (17):5981-6000. doi: https://doi.org/10.1021%2Facs.chemrev.2c00581.

Shilling, Patrick J., Kiavash Mirzadeh, Alister J. Cumming, Magnus Widesheim, Zoe Köck, and Daniel O. Daley. 2020. "Improved designs for pET expression plasmids increase protein production yield in Escherichia coli." *Communications Biology* 3 (1):214. doi: https://doi.org/10.1038/s42003-020-0939-8.

Shin, Woonsup, Sang Hee Lee, Jun Won Shin, Sang Phil Lee, and Yousung Kim. 2003. "Highly Selective Electrocatalytic Conversion of CO2 to CO at −0.57 V (NHE) by Carbon Monoxide Dehydrogenase from Moorella thermoacetica." *Journal of the American Chemical Society* 125 (48):14688-14689. doi: https://doi.org/10.1021/ja037370i.

Singh, Ashish Kumar, Suryabhan Singh, and Abhinav Kumar. 2016. "Hydrogen energy future with formic acid: a renewable chemical hydrogen storage system." *Catalysis Science & Technology* 6 (1):12-40. doi: https://doi.org/10.1039/C5CY01276G.

Singh, Hawaibam Birla, Min-Kyoung Kang, Moonhyuk Kwon, and Seon-Won Kim. 2022. "Developing methylotrophic microbial platforms for a methanol-based bioindustry." *Front Bioeng Biotechnol* 10:1050740. doi: https://doi.org/10.3389%2Ffbioe.2022.1050740.

Souterre, Tiffany. 2017. "Optimization of enzymes via directed evolution in vivo." http://www.theses.fr/2017SACLE016.

Sridhara, Viswanadham, Austin G. Meyer, Piyush Rai, Jeffrey E. Barrick, Pradeep Ravikumar, Daniel Segrè, and Claus O. Wilke. 2014. "Predicting Growth Conditions from Internal Metabolic Fluxes in an In-Silico Model of E. coli." *PLOS ONE* 9 (12):e114608. doi: https://doi.org/10.1371/journal.pone.0114608.

Steiert, Paula S., Lorraine T. Stauffer, and George V. Stauffer. 1990. "The lpd gene product functions as the L protein in the Escherichia coli glycine cleavage enzyme system." *J Bacteriol* 172 (10):6142-4. doi: https://doi.org/10.1128/jb.172.10.6142-6144.1990.

Szima, Szabolcs, and Calin-Cristian Cormos. 2018. "Improving methanol synthesis from carbon-free H2 and captured CO2: A techno-economic and environmental evaluation." *Journal of CO2 Utilization* 24:555-563. doi: https://doi.org/10.1016/j.jcou.2018.02.007.

Taddei, F., M. Radman, J. Maynard-Smith, B. Toupance, P. H. Gouyon, and B. Godelle. 1997. "Role of mutator alleles in adaptive evolution." *Nature* 387 (6634):700-702. doi: https://doi.org/10.1038/42696.

Tafur Rangel, Albert Enrique, Wendy Ríos, Daisy Mejía, Carmen Ojeda, Ross Carlson, Jorge Mario Gómez Ramírez, and Andrés Fernando González Barrios. 2021. "In silico Design for Systems-Based Metabolic Engineering for the Bioconversion of Valuable Compounds From Industrial By-Products." *Frontiers in Genetics* 12. doi: https://doi.org/10.3389/fgene.2021.633073.

Tashiro, Yohei, Shinichi Hirano, Morgan M. Matson, Shota Atsumi, and Akihiko Kondo. 2018. "Electrical-biological hybrid system for CO2 reduction." *Metabolic Engineering* 47:211-218. doi: https://doi.org/10.1016/j.ymben.2018.03.015.

Tegel, H., J. Ottosson, and S. Hober. 2011. "Enhancing the protein production levels in Escherichia coli with a strong promoter." *Febs j* 278 (5):729-39. doi: https://doi.org/10.1111/j.1742-4658.2010.07991.x.

Tengku, Elida, Zainal Mulok, Mei-Ling Chong, Yoshihito Shirai, Raha Rahim, and Mohd Hassan. 2009. "Engineering of E. coli for increased production of L- lactic acid." *AFRICAN JOURNAL OF BIOTECHNOLOGY* 8. doi: https://doi.org/10.5897/AJB09.614.

Thauer, Rudolf K. 2007. "A Fifth Pathway of Carbon Fixation." *Science* 318 (5857):1732-1733. doi: https://doi.org/10.1126/science.1152209.

Thauer, Rudolf K., Anne-Kristin Kaster, Henning Seedorf, Wolfgang Buckel, and Reiner Hedderich. 2008. "Methanogenic archaea: ecologically relevant differences in energy conservation." *Nature Reviews Microbiology* 6 (8):579-591. doi: https://doi.org/10.1038/nrmicro1931.

Tishkov, V. I., A. G. Galkin, G. N. Marchenko, O. A. Egorova, D. V. Sheluho, L. B. Kulakova, L. A. Dementieva, and A. M. Egorov. 1993. "Catalytic Properties and Stability of a Pseudomonas sp.101 Formate Dehydrogenase Mutants Containing Cys-255-Ser and Cys-255-Met Replacements." *Biochemical and Biophysical Research Communications* 192 (2):976-981. doi: https://doi.org/10.1006/bbrc.1993.1511.

Toukdarian, Aresa E., Donald R. Helinski, and Silvia Perri. 1996. "The Plasmid RK2 Initiation Protein Binds to the Origin of Replication as a Monomer (∗)." *Journal of Biological Chemistry* 271 (12):7072-7078. doi: https://doi.org/10.1074/jbc.271.12.7072.

Vallenet, David, Alexandra Calteau, Stéphane Cruveiller, Mathieu Gachet, Aurélie Lajus, Adrien Josso, Jonathan Mercier, Alexandre Renaux, Johan Rollin, Zoé Rouy, David Roche, Claude Scarpelli, and Claudine Médigue. 2017. "MicroScope in 2017: an expanding and evolving integrated resource for community expertise of microbial genomes." *Nucleic Acids Res* 45 (D1):D517-d528. doi: https://doi.org/10.1093/nar/gkw1101.

van Eunen, Karen, and Barbara M. Bakker. 2014. "The importance and challenges of in vivo-like enzyme kinetics." *Perspectives in Science* 1 (1):126-130. doi: https://doi.org/10.1016/j.pisc.2014.02.011.

van Eunen, Karen, José A. L. Kiewiet, Hans V. Westerhoff, and Barbara M. Bakker. 2012. "Testing Biochemistry Revisited: How In Vivo Metabolism Can Be Understood from In Vitro Enzyme Kinetics." *PLOS Computational Biology* 8 (4):e1002483. doi: https://doi.org/10.1371/journal.pcbi.1002483.

van Niel, E. W. J., and B. Hahn-Hägerdal. 1999. "Nutrient requirements of lactococci in defined growth media." *Applied Microbiology and Biotechnology* 52 (5):617-627. doi: https://doi.org/10.1007/s002530051569.

Vignais, Paulette M., and Bernard Billoud. 2007. "Occurrence, Classification, and Biological Function of Hydrogenases: An Overview." *Chemical Reviews* 107 (10):4206-4272. doi: https://doi.org/10.1021/cr050196r.

Vigouroux, Antoine, Enno Oldewurtel, Lun Cui, David Bikard, and Sven van Teeffelen. 2018. "Tuning dCas9's ability to block transcription enables robust, noiseless knockdown of bacterial genes." *Molecular Systems Biology* 14 (3):e7899. doi: https://doi.org/10.15252/msb.20177899.

Vik, Erik Sebastian, Meh Sameen Nawaz, Pernille Strøm Andersen, Cathrine Fladeby, Magnar Bjørås, Bjørn Dalhus, and Ingrun Alseth. 2013. "Endonuclease V cleaves at inosines in RNA." *Nat Commun* 4:2271. doi: https://doi.org/10.1093/nar/gkaa115.

Wagner, A F, M Frey, F A Neugebauer, W Schäfer, and J Knappe. 1992. "The free radical in pyruvate formate-lyase is located on glycine-734." *Proceedings of the National Academy of Sciences* 89 (3):996-1000. doi: https://doi.org/10.1073/pnas.89.3.996.

Walker, Lindsey M., Bin Li, Dimitri Niks, Russ Hille, and Sean J. Elliott. 2019. "Deconvolution of reduction potentials of formate dehydrogenase from Cupriavidus necator." *J Biol Inorg Chem* 24 (6):889-898. doi: https://doi.org/10.1007/s00775-019-01701-1.

Wang, Harris H., and George M. Church. 2011. "Multiplexed Genome Engineering and Genotyping Methods: Applications for Synthetic Biology and Metabolic Engineering." In *Methods in Enzymology*, edited by Christopher Voigt, 409-426. Academic Press.

Wang, Harris H., Farren J. Isaacs, Peter A. Carr, Zachary Z. Sun, George Xu, Craig R. Forest, and George M. Church. 2009. "Programming cells by multiplex genome engineering and accelerated evolution." *Nature* 460 (7257):894-898. doi: https://doi.org/10.1038/nature08187.

Wang, Yajie, Pu Xue, Mingfeng Cao, Tianhao Yu, Stephan T. Lane, and Huimin Zhao. 2021. "Directed Evolution: Methodologies and Applications." *Chemical Reviews* 121 (20):12384-12444. doi: https://doi.org/10.1021/acs.chemrev.1c00260.

Wang, Ye, Xin-jing Yue, Shu-fei Yuan, Yu Hong, Wei-feng Hu, and Yue-zhong Li. 2021. "Internal Promoters and Their Effects on the Transcription of Operon Genes for Epothilone Production in Myxococcus xanthus." *Frontiers in Bioengineering and Biotechnology* 9. doi: https://doi.org/10.3389/fbioe.2021.758561.

Wang, Yiou, Enqi Chen, and Junwang Tang. 2022. "Insight on Reaction Pathways of Photocatalytic CO2 Conversion." *ACS Catalysis* 12 (12):7300-7316. doi: https://doi.org/10.1021/acscatal.2c01012.

Weaver, Paul F., Judy D. Wall, and Howard Gest. 1975. "Characterization of Rhodopseudomonas capsulata." *Archives of Microbiology* 105 (1):207-216. doi: 10.1007/BF00447139.

Weber, Ernst, Carola Engler, Ramona Gruetzner, Stefan Werner, and Sylvestre Marillonnet. 2011. "A Modular Cloning System for Standardized Assembly of Multigene Constructs." *PLOS ONE* 6 (2):e16765. doi: https://doi.org/10.1371/journal.pone.0016765.

Wenk, Sebastian, Karin Schann, Hai He, Vittorio Rainaldi, Seohyoung Kim, Steffen N. Lindner, and Arren Bar-Even. 2020. "An "energy-auxotroph" Escherichia coli provides an in vivo platform for assessing NADH regeneration systems." *Biotechnology and Bioengineering* 117 (11):3422-3434. doi: https://doi.org/10.1002/bit.27490.

Wenk, Sebastian, Oren Yishai, Steffen N. Lindner, and Arren Bar-Even. 2018. "An Engineering Approach for Rewiring Microbial Metabolism." *Methods Enzymol* 608:329-367. doi: https://doi.org/10.1016/bs.mie.2018.04.026.

Whitaker, William B., Nicholas R. Sandoval, Robert K. Bennett, Alan G. Fast, and Eleftherios T. Papoutsakis. 2015. "Synthetic methylotrophy: engineering the production of biofuels and chemicals based on the biology of aerobic methanol utilization." *Current Opinion in Biotechnology* 33:165-175. doi: https://doi.org/10.1016/j.copbio.2015.01.007.

White, Andrea K. , and William W. Metcalf. 2007. "Microbial Metabolism of Reduced Phosphorus Compounds." *Annual Review of Microbiology* 61 (1):379-400. doi: https://doi.org/10.1146/annurev.micro.61.080706.093357.

White, Andrea K., and William W. Metcalf. 2004a. "The htx and ptx operons of Pseudomonas stutzeri WM88 are new members of the pho regulon." *J Bacteriol* 186 (17):5876-82. doi: https://doi.org/10.1128/jb.186.17.5876-5882.2004.

White, Andrea K., and William W. Metcalf. 2004b. "Two C-P lyase operons in Pseudomonas stutzeri and their roles in the oxidation of phosphonates, phosphite, and hypophosphite." *J Bacteriol* 186 (14):4730-9. doi: https://doi.org/10.1128/jb.186.14.4730-4739.2004.

Wilson, David S., and Anthony D. Keefe. 2000. "Random Mutagenesis by PCR." *Current Protocols in Molecular Biology* 51 (1):8.3.1-8.3.9. doi: https://doi.org/10.1002/0471142727.mb0803s51.

Woolerton, Thomas W., Sally Sheard, Erwin Reisner, Elizabeth Pierce, Stephen W. Ragsdale, and Fraser A. Armstrong. 2010. "Efficient and Clean Photoreduction of CO2 to CO by Enzyme-Modified TiO2 Nanoparticles Using Visible Light." *Journal of the American Chemical Society* 132 (7):2132-2133. doi: https://doi.org/10.1021/ja910091z.

Yaashikaa, P. R., P. Senthil Kumar, Sunita J. Varjani, and A. Saravanan. 2019. "A review on photochemical, biochemical and electrochemical transformation of CO2 into value-added products." *Journal of CO2 Utilization* 33:131-147. doi: https://doi.org/10.1016/j.jcou.2019.05.017.

Yang, Dongtai, Sheng Li, Song He, and Yawen Zheng. 2022. "Can conversion of CO2 into fuels via electrochemical or thermochemical reduction be energy efficient and reduce emissions?" *Energy Conversion and Management* 273:116425. doi: https://doi.org/10.1016/j.enconman.2022.116425.

Yang, Kechao, and William W. Metcalf. 2004. "A new activity for an old enzyme: Escherichia coli bacterial alkaline phosphatase is a phosphite-dependent hydrogenase." *Proc Natl Acad Sci U S A* 101 (21):7919-24. doi: https://doi.org/10.1073/pnas.0400664101.

Yishai, Oren, Steffen N. Lindner, Jorge Gonzalez de la Cruz, Hezi Tenenboim, and Arren Bar-Even. 2016. "The formate bio-economy." *Curr Opin Chem Biol* 35:1-9. doi: https://doi.org/10.1016/j.cbpa.2016.07.005.

Yu, Xuejun, Dimitri Niks, Xin Ge, Haizhou Liu, Russ Hille, and Ashok Mulchandani. 2019. "Synthesis of Formate from CO2 Gas Catalyzed by an O2-Tolerant NAD-Dependent Formate Dehydrogenase and Glucose Dehydrogenase." *Biochemistry* 58 (14):1861-1868. doi: https://doi.org/10.1021/acs.biochem.8b01301.

Yu, Xuejun, Dimitri Niks, Ashok Mulchandani, and Russ Hille. 2017. "Efficient reduction of CO2 by the molybdenum-containing formate dehydrogenase from Cupriavidus necator (Ralstonia eutropha)." *J Biol Chem* 292 (41):16872-16879. doi: https://doi.org/10.1074/jbc.m117.785576.

Zarzycki, Jan, Volker Brecht, Michael Müller, and Georg Fuchs. 2009. "From the Cover: Identifying the missing steps of the autotrophic 3-hydroxypropionate CO2 fixation cycle in Chloroflexus aurantiacus." *Proceedings of the National Academy of Sciences of the United States of America* 106:21317-22. doi: https://doi.org/10.1073/pnas.0908356106.

Zelcbuch, Lior, Steffen N. Lindner, Yonatan Zegman, Ilya Vainberg Slutskin, Niv Antonovsky, Shmuel Gleizer, Ron Milo, and Arren Bar-Even. 2016. "Pyruvate Formate-Lyase Enables Efficient Growth of Escherichia coli on Acetate and Formate." *Biochemistry* 55 (17):2423-2426. doi: https://doi.org/10.1021/acs.biochem.6b00184.

Zhou, Y., L. Wang, F. Yang, X. Lin, S. Zhang, and Z. K. Zhao. 2011. "Determining the extremes of the cellular NAD(H) level by using an Escherichia coli NAD(+)-auxotrophic mutant." *Appl Environ Microbiol* 77 (17):6133-40. doi: https://doi.org/10.1128/AEM.00630-11.

Zhuang, Zong-Yu, and Si-Yu Li. 2013. "Rubisco-based engineered Escherichia coli for in situ carbon dioxide recycling." *Bioresource Technology* 150:79-88. doi: https://doi.org/10.1016/j.biortech.2013.09.116.

APPENDICES

Appendix 1: Genetic sequences of diverse building blocks

Building block	Sequence
P_strong	aatacttgacatatcactgtgattcacatataatatgcg
P_moderate	acctattgacaattaaaggctaaaatgctataattccac
P_{trc} promoter	ttgacaattaatcatccggctcgtataatg
RBS A	aggaggtttgga
RBS B	aacaaaatgaggaggtactgag
Primer 4463 (*panB*) – RTqPCR	ttagaagctgctggggcaca
Primer 4464 (*panB*) – RTqPCR	ccgtttcggcgaggaaatt
Primer 6419 (*fdsG*) – RTqPCR	atcctgcatgagatccaggacac
Primer 6420 (*fdsG*) – RTqPCR	aagtggtggtagaaggtgatcacg
Primer 6422 (*fdsA*) – RTqPCR	aacggcaattgcgaactgcag
Primer 6423 (*fdsA*) – RTqPCR	attcgtccttcttcatctgcgtgtg
Primer 6427 (*fdsB*) – RTqPCR	acaacctcatcaccatggccaac
Primer 6428 (*fdsB*) – RTqPCR	atatccaacagcccgttccctg

Appendix 2: Formate dehydrogenase from *C. necator* sequence

Sequence of the native formate dehydrogenase from *Cupriavidus necator* N-1 strain: the start codon are highlighted in blue, the stop codon are in red, if overlap, in purple. The five genes are placed in this order: *fdsGBACD*.

ATGCCAGAAATTGCCCCCCACGCAGCGGCATCCGCCGATGCCACGCGCATCGCCGCCATCGTGGCGGCACGCCAGGAAATGCCGGGCGC
CTTGCTGCCGATCCTGCATGAGATCCAGGACACACAGGGCTATATTCCCGACGCCGCCGTGCCCGTGATTGCCCGCGCGCTGAACCTGTCG
CGCGCCGAGGTGCATGGCGTGATCACCTTCTACCACCACTTCCGCCAGCAGCCGGCCGGGCGCCACGTGGTACAGGTCTGCCGCGCCGA
AGCCTGCCAGTCGGTCGGTGCCGAAGCGCTGGCCGAGCATGCGCAGCGCGCACTTGGCTGTGGCTTCCATGAAACCAGCGCGGACGGGC
AGGTGACGCTGGAGCCGGTTTATTGCCTGGGCCAGTGCGCCTGCGGCCCGGCGGTGATGGTCGGCGAGCAACTGCACGGCTATGTCGAC
GCGAAGCGCTTCGACGCGCTGGTGCGCTCGCTGCGCGAGTCTTCGAAAGAAACCCCGGAAGCGGCGGAGGCACAGGCATGATCACGGTC
ACCACCATCTTCGTGCCGCGCGACTCCACCGCGCTGGCGCTGGGCGCCGACGATGTGGCGCGCGCCATTGCGCGCGAAGCCGCCGCGCG
CAACCAGCACGTGCGCATTGTCCGCAACGGCTCGCGCGGCATGTTCTGGCTGGAGCCGCTGGTCGAGGTGCAGACCGGAGCCGGCCGCG
TGGCCTATGGCCCGGTCAGCGCCGCAGACGTGCCGGAATTGTTCGACGCTGGCCTGCTGCAAGGCGGCGAGCACGCGCTGTCGCAGGGC
GTGACCGAAGAGATCCCCTTCCTGAAGCAGCAGGAGCGCCTGACCTTCGCCCGCGTCGGCATCACCGATCCGCTGTCGCTGGACGACTAC
CGCGCGCATGAGGGCTTTGCCGGCCTGGAGCGCGCGCTGGCGATGCAGCCCGCCGAGATCGTGCAGGAGGTCGCCGACTCCGGCCTGCG
CGGCCGCGGCGGCGCGGCATTCCCGACCGGCATCAAGTGGAAGACCGTGCTGGGCGCGCAGTCCGCGGTCAAGTACATCGTCTGCAATG
CCGACGAAGGCGACTCGGGCACGTTCTCTGACCGCATGGTGATGGAAGACGACCCGTTCATGCTGATCGAAGGCATGACCATTGCCGCCC
TGGCGGTGGGTGCTGAGCAGGGCTACATCTACTGCCGTTCCGAATACCCGCACGCGATTGCCGTGCTGGAAAGCGCGATCGGCATCGCCA
ACGCCGCCGGCTGGCTCGGCGACGACATCCGCGGCAGCGGCAAGCGCTTCCACCTGGAAGTGCGCAAGGGCGCCGGCGCCTATGTCTGC
GGCGAGGAAACCGCGCTGCTGGAAAGCCTGGAAGGACGGCGCGGCGTGGTGCGCGCCAAGCCGCCGCTGCCGGCGCTGCAGGGGCTGT
TCGGCAAGCCCACGGTGATCAACAACGTGATCTCGCTGGCCACCGTGCCGGTGATCCTGGCGCGCGGCGCGCAGTACTACCGCGACTACG
GCATGGGCCGTTCGCGCGGCACGCTGCCGTTCCAGCTGGCCGGAAACATCAAGCGGGGCGGACTGGTGGAGAAAGCGTTCGGCGTCACG
CTGCGCGAGCTGCTGGTCGACTACGGCGGCGGCACGCGCAGCGGCCGCGCCATCCGCGCGGTGCAGGTGGGCGGGCCGCTGGGCGCCT
ACCTGCCCGAGTCGCGCTTCGACGTGCCGCTGGACTATGAAGCCTATGCCGCGTTCGGCGGCGTGGTCGGCCACGGCGGCATCGTGGTGT
TCGATGAAACCGTCGACATGGCAAAGCAGGCCCGCTACGCGATGGAGTTCTGTGCGATCGAATCGTGCGGCAAGTGCACCCCGTGCCGG
ATCGGCTCGACCCGCGGGGTCGAAGTGATGGACCGCATCATTGCCGGCGAGCAGCCGGTCAAGCATGTCGCACTGGTGCGCGACCTGTG
CGACACCATGCTTAACGGCTCGCTGTGCGCGATGGGCGGCATGACACCGTACCCGGTGCTGTCCGCGCTGAATGAATTCCCCGAGGACTT
CGGCCTCGCCTCCAACCCAGCCAAGGCCGCCTGAGCCAGGTCCAGCAGAGACACGGGAGACAAACCGCCATGAACGCCCGCAACGAGAT
CGATTTCGGCACGCCTGCCAGCCCATCCACTGAACTGGTCACCCTGGAGGTCGATGGCGTCAGCGTCACCGTGCCCGCCGGCACCTCAGT
GATGCGCGCCGCGATGGAAGCGCAGATCGCCGTCCCCAAGCTGTGCGCCACCGACAGCCTCGAAGCCTTCGGCTCGTGCCGGCTGTGCCT
GGTCGAGATCGAAGGGCGCCGCGGCTATCCGGCATCGTGCACCACGCCGGTCGAAGCCGGCATGAAGGTCAAGACCCAGAGCGACAAG
CTGGCCGACCTGCGTCGCGGCGTGATGGAGCTGTATATCTCCGACCACCCGCTCGATTGCCTGACCTGCCCGACCAACGGCAATTGCGAA
CTGCAGGACATGGCCGGCGTGGTCGGCCTGCGTGAAGTGCGCTACAACGACGGCGGCCCGGAAGCTGCGCCGATCGCGACGCACACGC
AGATGAAGAAGGACGAATCCAATCCTTACTTCACCTATGACCCCTCCAAGTGCATCGTCTGCAACCGCTGCGTGCGTGCCTGCGAGGAAAC
GCAGGGTACCTTCGCCCTGACCATCAGCGGCCGCGGCTTCGATTCCCGCGTCTCGCCCGGCACCAGCCAGTCGTTCATGGAATCGGACTG
CGTCTCGTGCGGCGCCTGCGTGCAGGCGTGCCCGACCGCGACGCTGACCGAGACCTCGGTGATCAAGTTCGGCCAGCCCTCGCACAGCA
CCGTGACTACCTGTGCCTATTGCGGCGTGGGCTGTTCGTTCAAGGCCGAGATGAAGGGCAATGAAGTGGTGCGCATGGTGCCGTACAAGG
ACGGCAAGGCCAATGAAGGTCACGCCTGCGTCAAGGGCCGCTTTGCCTGGGGCTACGCCACGCACAAGGACCGCATTCTCAAGCCGATG
ATCCGCGCCAAGATCACCGATCCGTGGCGCGAGGTGTCGTGGGAAGAGGCGATCGACTACGCCGCGTCGCAGTTCAAACGTATCCAGGC
CGAGCACGGCAAGGATTCGATCGGCGGCATCGTGTCGTCGCGCTGCACCAATGAAGAGGGCTATCTGGTCCAGAAACTGGTGCGCGCCG
CCTTCGGCAACAACAACGTCGACACCTGCGCGCGCGTGTGCCATTCGCCGACCGGCTATGGCCTGAAGCAGACGCTGGGTGAATCGGCC

GGCACGCAGACCTTCAAGTCGGTCGAGAAGGCCGACGTGATCATGGTGATCGGCGCCAACCCGACCGACGGACACCCGGTCTTTGCCTC
GCGCATGAAGAAGCGCCTGCGCGCCGGTGCGAAGCTGATTGTGGTCGATCCGCGCCGCATCGACCTGGTCGACTCGCCGCATATCCGTGC
CGACTATCACCTGCAGCTGCGCCCGGGCACCAACGTGGCACTGGTGACCTCGCTGGCCCACGTGATCGTCACCGAAGGCCTGCTCAACGA
AGCCTTCATCGCTGAGCGCTGCGAGGACCGCGCCTTCCAGCAATGGCGCGATTTCGTCTCGTTGGCGGAGAACTCGCCCGAGGCCATGGA
AAGCGTCACCGGCATTCCGGCGGAACAGCTGCGCGGCGCCGCGCGCCTGTACGCTACCGGCGGCAATGCAGCGATCTACTACGGCCTGG
GCGTGACCGAGCATGCGCAAGGCTCCACCACCGTGATGGGCATTGCCAACTTGGCCATGGCCACCGGCAATATCGGTCGCGAAGGCGTG
GGCGTGAACCCGCTGCGCGGGCAGAACAACGTGCAGGGCTCGTGCGATATCGGCTCGTTCCCGCATGAGCTGCCGGGCTACCGCCACGT
GTCGGACTCGACCACGCGTGGCCTGTTCGAAGCTGCGTGGAACGTCGAGATCAGCCCCGAGCCGGGCCTGCGCATTCCCAATATGTTCGA
AGCCGCGCTGGCTGGCAGCTTCAAGGGCCTGTACTGCCAGGGCGAGGACATCGTCCAGTCCGACCCGAATACGCAGCACGTGTCCGAGG
CGCTGTCGTCGATGGAATGCATTGTGGTGCAGGACATCTTCCTCAACGAGACCGCCAAGTACGCGCACGTGTTCCTGCCGGGCTCGTCCTT
CCTGGAAAAGGACGGCACCTTCACCAACGCCGAGCGCCGCATCTCGCGCGTGCGCAAGGTGATGCCGCCCAAGGCGCGCTATGCCGACT
GGGAAGCGACCATCCTGCTGGCCAATGCGCTCGGCTATCCGATGGAGTACAAGCATCCGTCGGAGATCATGGACGAGATCGCGCGCCTG
ACGCCGACCTTCGCGGGCGTCAGCTACAAGCGCCTGGACCAGCTGGGCAGCATCCAGTGGCCGTGCAACGCCGACGCGCCGGAAGGCAC
GCCGACCATGCATATCGACGCCTTCGTGCGTGGCAAGGGCAAATTCATCATCACCAAGTACGTGCCCACCACCGAGAAGATCACGCGCGC
CTTCCCGCTGATCCTGACCACTGGCCGCATCCTGTCGCAATACAACGTCGGCGCGCAGACGCGCCGCACCGACAACGTCTACTGGCATGC
CGAGGACCGGCTCGAGATCCATCCGCACGATGCCGAGGAGCGCGGCATCAAGGACGGCGACTGGGTCGGGGTGCAGAGCCGGGCAGG
GGACACCGTGCTGCGCGCGATTGTCAGCGAGCGCATGCAGCCGGGTGTGGTCTACACCACCTTCCACTTCCCGGAATCCGGCGCCAATGT
GATCACCACCGACAATTCCGACTGGGCCACCAACTGCCCGGAATACAAGGTGACCGCGGTGCAGGTGCTGCCGGTGGCGCAGCCTTCGG
CGTGGCAGCGGGAGTACCAGGAGTTCAACACCCAGCAGCTGCAACTGCTGGAAGCAGCGAGCGCCGATCCGGCGCAGGCCGCAGGGTG
AGCGGAGGGTTACGCCATGATGCGCTGCATGCAGTCACCGGAAATTGATCCGGACGCGGCGGAAGACGCCGTGCCCGCCACCCACAGCA
CCTTCGCCGTCAGCCGCTGGCGCCGCGGCGAGCTGATGCTGAGCCCCGATGAAGTGGCCGAGGAAGTGCCGGTCGCGCTGGTGTACAAC
GGCATCTCGCACGCGGTGATGCTGGCGACGCCGGCCGACCTGGAGGACTTTGCGCTCGGGTTCAGCCTGAGCGAGGGCATCGTTACCCGT
GCCAGCGACGTCTATGACATCGAGATCGATATGCGCGAGCACGGCATCGCCGTGCAGCTGGAGATCGCCTCCGAAGCCTTCATGCGGCTC
AAGGATCGCCGCCGCTCGCTGGCCGGGCGCACCGGCTGCGGGCTGTGCGGCACTGAATCGCTGGAGCAGGTGATGCGCCTGCCGGCACC
TGTGCGCAGCGAAGCCAGCTTCCATACCGACGTGATTCAGGCCGCCTTCGTGCAACTGCAACTGCGACAGGAGCTGCAGCAACACACGGG
TGCGACGCACGCTGCCGCATGGCTGCGTGCCGATGGCCATGTATCGCTGGTGCGCGAAGACGTGGGCCGCCACAACGCGCTGGACAAGC
TGGCGGGCGCGCTTGCCCGCAGCGGCGAGGACATCTCCAGCGGCGCGGTGCTGGTGACCAGTCGCGCGAGCTATGAAATGGTGCTGAAG
ACCGCCGCCATCGGCGCCGGCGTGCTGGCCGCAGTGTCCGCACCGACGGCGCTGGCCGTGCGCCTCGCCGAGCAGGCCAACATCACCCT
GGCCGGCTTCGTGCGCGCCGGCGCGCACGTGGTCTATGCCCATCCCCAACGTTTGCAGCACGAAGCGAGTCTGGCATGAAGATCGACAAC
CTCATCACCATGGCCAACCAGATCGGCAGCTTCTTCGAGGCCATGCCGGATCGGGAAGAGGCTGTCTCGGATATTGCGGGGCATATCAAG
CGGTTCTGGGAGCCGCGGATGCGCAAGGCGCTGCTGGGGCATGTGGATGCCGAAGCAGGGAACGGGCTGTTGGATATCGTGCGCGAGG
CGCTGGGGCGGCATCGGGCGATGCTGGAGTAA

Appendix 3: Formate dehydrogenase from *Thiobacillus sp.* sequence

Sequence of the codon-optimized formate dehydrogenase from *Thiobacillus sp.*: the start codon is highlighted in blue, the stop codon is in red.

ATGGCAAAGATCTTATGTGTGTTATACGACGACCCAGTGGATGGTTACCCCAAAACATACGCCCGTGATGACCTTCCTAAAATTGATCATTA
CCCAGGTGGACAAACTTTGCCAACTCCCAAAGCAATTGATTTTACACCGGGTCAATTACTGGGGTCAGTAAGCGGTGAATTGGGTCTCCGA
AAGTACCTGGAAGCAAATGGGCATACGTTTGTAGTAACATCAGATAAAGATGGTCCTGACTCTGTATTTGAAAAGGAACTGGTAGATGCAG
ATGTAGTTATTTCACAACCATTTTGGCCTGCTTACCTCACGCCAGAAAGAATTGCTAAAGCCAAGAATCTTAAATTAGCCTTAACGGCGGGA
ATTGGGTCGGACCATGTGGACCTGCAATCTGCTATAGATAGAGGTATTACGGTTGCAGAGGTGACTTATTGTAATTCTATATCAGTTGCGGA
ACACGTAGTCATGATGATTTTGGGTCTCGTTCGGAATTATATACCTTCCCACGATTGGGCTCGTAAAGGTGGTTGGAATATTGCAGATTGTG
TTGAACATTCATATGATTTAGAAGGTATGACAGTGGGTAGTGTAGCAGCTGGTCGAATTGGGCTGGCAGTTCTCCGTCGTTTAGCCCCGTTT
GATGTTAAACTTCATTATACGGATCGGCATCGTCTTCCCGAGGCTGTGGAGAAAGAACTTGGTTTAGTGTGGCATGATACGAGAGAAGATA
TGTATCCGCACTGTGATGTTGTAACCCTGAATGTTCCTCTTCACCCAGAAACAGAGCATATGATTAACGATGAAACACTTAAATTATTTAAAC
GTGGGGCATACATTGTTAATACAGCGCGAGGTAAACTGGCTGATCGTGATGCTATTGTTCGGGCCATAGAAAGTGGCCAACTGGCGGGTT
ACGCCGGTGATGTTTGGTTTCCCCAACCAGCTCCTAAAGACCATCCTTGGCGGACGATGAAATGGGAAGGTATGACACCCCATATCTCTGG
TACTAGTTTATCTGCGCAAGCCAGATACGCAGCTGGAACCAGAGAAATTCTGGAGTGTTTCTTTGAGGGGCGTCCAATTCGAGATGAATAT
CTTATTGTTCAAGGTGGTGCCTTAGCTGGAACAGGTGCTCACTCATATAGTAAAGGAAACGCTACTGGCGGGTCTGAGGAAGCGGCAAAA
TTTAAGAAAGCGGGTTAA

Appendix 4: Conjugation protocol adapted from Wenk et al. (2018)

1. Transform *E. coli* ST18 cells with the pDM4 genome integration vector by electroporation and select for clones on LB plates supplemented with 50 mg/ml 5-aminolevulinic acid (ALA) and chloramphenicol and/or kanamycin.

2. Inoculate the ST18 donor strain and the recipient strain (MG1655 unless stated differently) in 5 mL of appropriate medium (e.g. LB ALA) in the morning and incubate at 37 °C in a shaking incubator until OD_{600} of 0.5.

3. Mix donor and recipient strain in a 1:1 OD ratio by combining 500 µL of each culture in a microfuge tube. If the OD differs, harvest higher volume of the strain with the lower OD and resuspend in 500 µL. Mix donor and recipient strain by inverting the tube.

4. Centrifuged the mixture for 3 min at 4000 rcf in a benchtop centrifuge.

5. Discard the supernatant and resuspend the cell pellet in the remaining 100 µL of the supernatant.

6. Pipette the entire cell suspension as one drop in the middle of an LB-ALA plate and incubate for 4–5 h upside up at room temperature under sterile conditions to allow for mating. Keep the lid of the plate open for the first 30 min to allow the spot to dry.

7. After 4–5 h scoop the cells from the mating spot using an inoculation loop and resuspend 100 µL LB medium. Vortex to stop the mating.

8. Plate 2 µL (20 µL of a 1:10 dilution), 20 µL and the rest of the cell suspension on plates optimized for the recipient strain (e.g., LB) containing just chloramphenicol and NO ALA and incubate the plates at 37 °C overnight in a plate incubator.

9. Pick single colonies and analyze plasmid integration by PCR.

10. Although *E. coli* ST18 cannot grow without the addition of ALA, plasmid integration strains should be transferred to a new plate at least two times by picking single colonies.

Appendix 5: UOF mutation analysis

UOF1

chromosomal position*	mutation type	mutation event	*mutated gene*	intragenic position	amino acid change	intergenic mutation (distance to the flanking genes)	mutations identified in the indicated strain (•)		
							G5823	G5824	G5825
953670	A/C	SNP	*focA*	20	F7C		•	•	•
1029182	C/A	SNP				yccW/yccX (77/105)	•	•	•
1798380	C/T	SNP	*infC*	283	E95K		•	•	•

UOF2

chromosomal position*	mutation type	mutation event	*mutated gene*	intragenic position	amino acid change	intergenic mutation (distance to the flanking genes)	mutations identified in the indicated strain (•)		
							G5848	G5849	G5850
88110	A/G	SNP	*cra*	83	Y28C				•
953401	C/T	SNP	*focA*	289	V97I		•	•	
953670	A/C	SNP	*focA*	20	F7C				•
1785132	G/A	SNP	*pps*	5	S2F		•	•	
2533860	G/T	SNP	*crr*	5	G2V		•	•	
4159247	T/G	SNP	*fabR*	158	L53W		•	•	

Appendix 6: *focA* reversion in UOF isolates G5823 and G5848

Figure: growth curves of strains G5823 (Δ*lpd* pTRC-CnFDH UOF1 evolved strain with *focA* F7C) and G6234 (G5823 strain with WT *focA*) for analysis of F7C mutation reversion. The strains were grown in MS minimal medium supplement with acetate (20 mM), pyruvate (20 mM), formate (60 mM) and IPTG (100 µM).

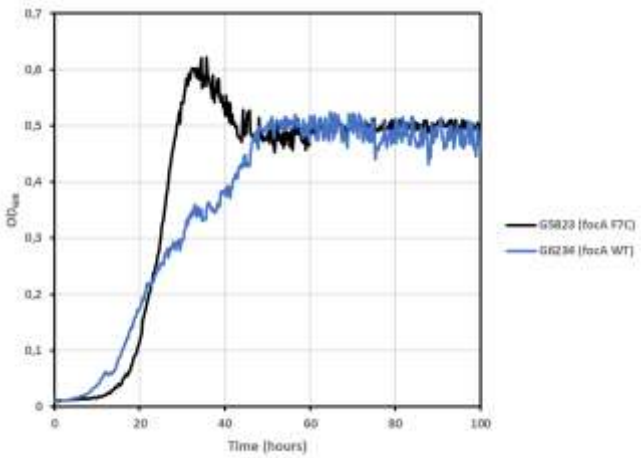
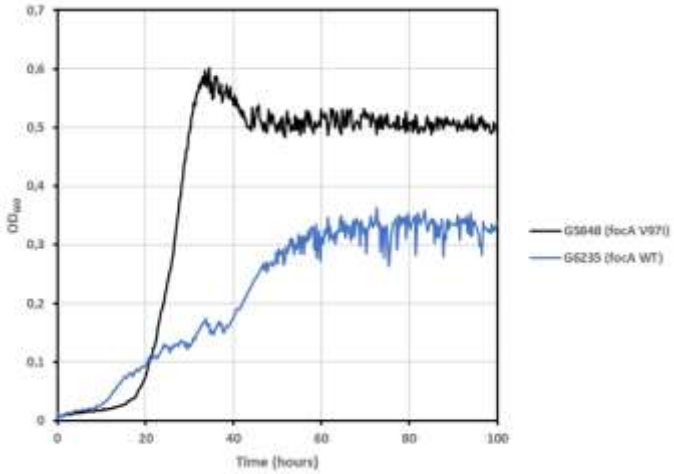


Figure: growth curves of strains G5848 (Δ*lpd* pTRC-CnFDH UOF1 evolved strain with *focA* V97I) and G6235 (G5848 strain with WT *focA*) for analysis of F7C mutation reversion. The strains were grown in MS minimal medium supplement with acetate (20 mM), pyruvate (20 mM), formate (60 mM) and IPTG (100 µM).



Appendix 7: OCF5 and OCF6 mutation analysis

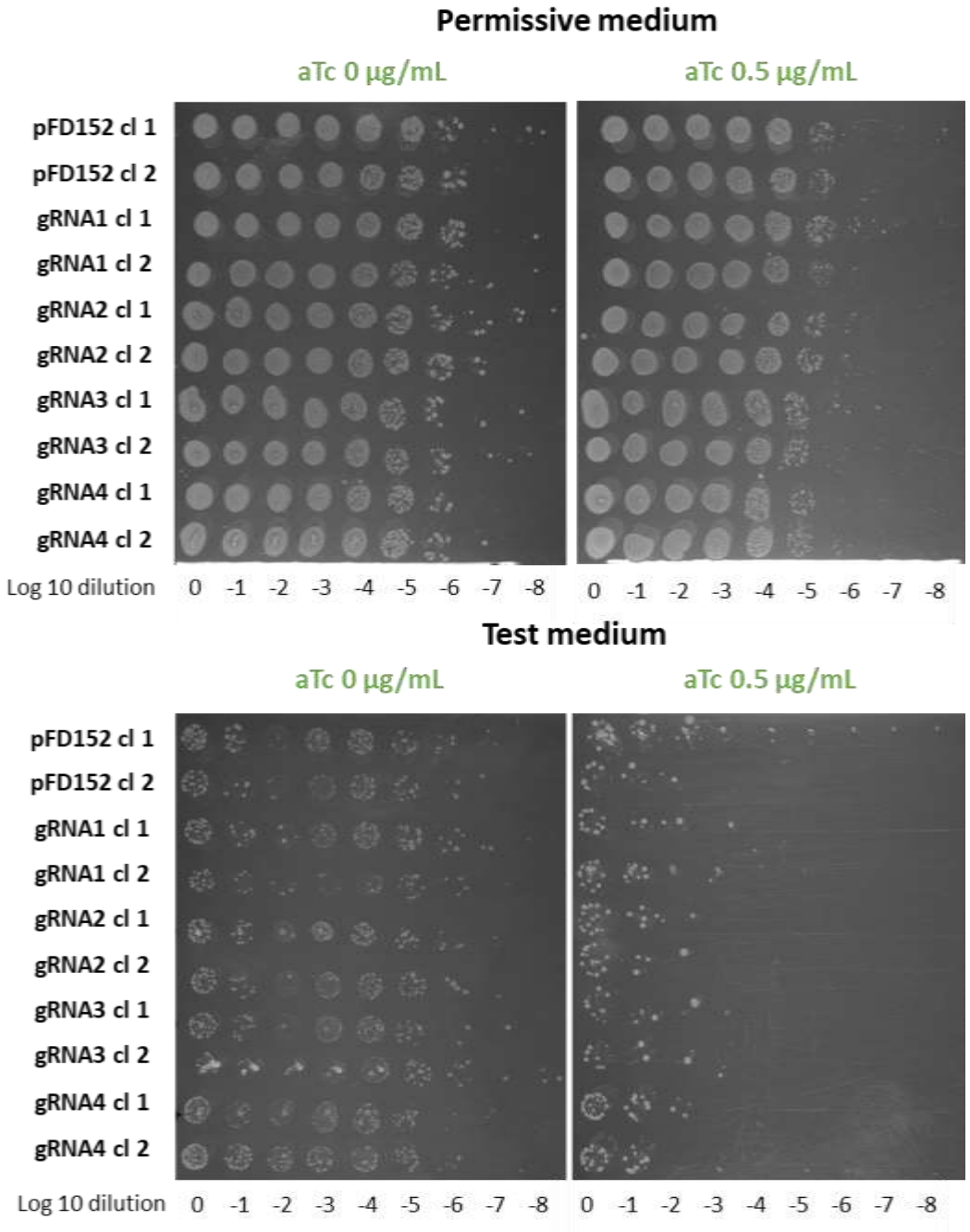
OCF5

chromosomal position*	mutation type	mutation event	*mutated gene*	intragenic position	amino acid change	intergenic mutation (distance to the flanking genes)	mutations identified in the indicated strain (•)	
							G6539	G6540
953670	A/C	SNP	*focA*	20	F7C		•	•
1029182	C/A	SNP				*yccW/yccX* (-77/-105)	•	•
1798380	C/T	SNP	*infC*	283	E95K		•	•
4395421	A/G	SNP	*amiB*	1334	N445S		•	•

OCF6

chromosomal position*	mutation type	mutation event	*mutated gene*	intragenic position	amino acid change	intergenic mutation (distance to the flanking genes)	mutations identified in the indicated strain (•)	
							G6541	G6542
953670	A/C	SNP	*focA*	20	F7C		•	•
1029182	C/A	SNP				*yccW/yccX* (-77/-105)	•	•
1798380	C/T	SNP	*infC*	283	E95K		•	•
2304758	A/-	DEL 1bp	*mqo*	19	frameshift		•	
2393980	G/T	SNP	*nuoI*	493	E165*			•
2401836	G/A	SNP	*nuoC*	32	W11*		•	
3375243	DEL	D21bp	*sspA*		$DVDRELTL_{83-89}$		•	
3375381	DEL	D3 bp	*sspA*		DY_{211}			•
399149	T/G	SNP				*mreB/csrD* (-40/+265)	•	
4395421	A/G	SNP	*amiB*	1334	N445S		•	•

Appendix 8: CRISPRi experiment with guides targeting *fdsC* and *fdsD* genes from CnFDH



Appendix 9: Publication submitted in BioRxiv

doi: https://doi.org/10.1101/2023.10.27.564357 (Schulz et al. 2023)

Functional expression of a Mo-dependent formate dehydrogenase in *Escherichia coli* under aerobic conditions

Marion Schulz, Anne Berger, Ivan Dubois, Valérie Delmas, Mélodie Cadillon, Madeleine Bouzon, Volker Döring*

Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, 91057, Evry-Courcouronnes, France. *email : vdoring@genoscope.cns.fr

Abstract

Background: Oxygen tolerant complex metal-dependent formate dehydrogenases hold potential for biotechnological applications.

Principle Findings: In this work, we report the functional expression of the complex, molybdenum-dependent soluble formate dehydrogenase encoded by the *fdsGBACD* operon from *Cupriavidus necator* (CnFDH) in *Escherichia coli.* Expression of the operon from plasmids or from a copy integrated in the chromosome enabled growth of an energy-auxotrophic selection strain on formate as sole energy source under aerobic conditions. Growth could be accelerated in turbidostat, leading to a drop of the generation time of 1 hour. While no mutation was found in the operon of evolved isolates, genome sequencing revealed non-synonymous point mutations in the gene *focA* coding for a bidirectional formate transporter carried in all isolates sequenced. Reverting the mutations led to a drop in the growth rate demonstrating the *focA* mutation as principle target of continuous culture adaptation.

Significance: A member of the oxygen-tolerant subclass of complex FDH showed stable formate oxidation activity when expressed in the heterologous host *E. coli*, a model organism of biotechnology. The integration of the operon in the chromosome offers the possibility of structure/function studies and activity enhancements through *in vivo* mutagenesis, which can also be applied to CO_2 reduction in appropriate selection hosts.

Introduction

Formate dehydrogenases (FDH), a diverse family of enzymes, catalyze the reversible conversion of formate to CO_2, using NAD(P)/NAD(P)H or other compounds as redox co-substrate [1]. Members of this family can be divided into metal-dependent and metal-independent FDHs. The latter are monomeric proteins that do not contain redox-active centers, they are oxygen-insensitive and depend on NADH as redox cofactor [2]. Due to their simple structures and complete O_2 tolerance, metal-independent enzymes are the ones mostly used in biotechnological applications, notably for the regeneration of NADH, but also in the reductive sense in electrochemical [3] and photoelectrochemical processes [4]. By contrast, the metal-dependent enzymes contain either a molybdenum or a tungsten atom as part of a pyranopterin guanosine dinucleotide (PGD) cofactor, at least one Fe/S-center and have a complex quaternary structure [5]. Although most of these enzymes are oxygen sensitive, membrane bound and require electron donors/acceptors other than NAD(H), a few O_2-tolerant, NAD(H)-dependent and soluble enzymes have been found among this class, notably from the metabolically versatile bacteria *Cupriavidus necator* (CnFDH) [6, 7] and *Rhodobacter capsulatus* (RcFDH). The cryo-EM structure of this latter enzyme was solved, providing first insights into their mechanism of catalysis [8].

Most FDHs preferentially catalyze the exergonic oxidation of formate to CO_2. However, under appropriate thermodynamic conditions, they can reduce CO_2 to formate [9], thus having the potential to become valuable catalysts in the circular carbon economy: the greenhouse gas CO_2 is converted to value-added formate, that can be used as hydrogen storage material, as fuel in "Direct Formic Acid Fuel Cells" [10], as a versatile C_1 synthon for chemical synthesis and as sustainable feedstock for the bioindustry [11, 12]. The different FDH enzyme classes vary in this capacity [13], with the O_2-sensitive metal-dependent dehydrogenases from anaerobic bacteria like *Acetobacter woodii* [14] being the most active (k_{cat} > 500 sec^{-1}), and the metal-independent dehydrogenases being the least active catalysts (k_{cat}< 1 sec^{-1}). While examples exist in the literature in which the activity of a metal-independent FDH was enhanced up to 3-fold by site directed mutagenesis [15], it can be assumed that the subclass of O_2-tolerant, NAD- and metal-dependent enzymes have a higher potential to become the enzyme workhorses of CO_2 reduction under aerobic conditions. The cytoplasmic FDH purified from *Cupriavidus necator* was shown to catalyze this reaction with a k_{cat}= 11 sec^{-1} under anaerobic conditions [6]. However, enzyme purification was conducted under fully aerobic conditions demonstrating oxygen tolerance despite the presence of a molybdenum-containing CO_2-formate redox active site and four [4Fe-4S] centers and one [2Fe-2S] center in the α-subunit (FdsA, 105 kDa), a FMN cofactor for NAD/NADH electron transfer and a [4Fe-4S] center in the β-subunit (FdsB, 55 kDa) and a [2Fe-2S] in the γ-subunit (FdsG, 19 kDa).

While *ex vivo* structural and activity studies with complex FDHs have been conducted in recent years, studies of their activity in a cellular context are scarce. Recently, a Mo-dependent enzyme homologously expressed in *Pseudomonas putida* was shown to be active in a selective context [16]. In this report, we describe the cloning and heterologous plasmid-borne expression of the *fdsGBACD* operon coding for CnFDH in an *E. coli* MG1655 derived energy auxotrophic selection strain. We obtained expression-dependent aerobic growth on formate as

sole source of energy and isolated faster growing strain descendants upon evolution in continuous culture harboring genetic background mutations. Sustained formate dependent growth was also obtained when the operon was inserted into the chromosome of an evolved isolate cured from the plasmid. Expression from one copy upon genomic integration stabilizes the construct and will enable long-term strain adaptation and evolution in chosen genetic backgrounds to ameliorate enzyme activity and tolerance to O_2.

Results

Rescue of an energy auxotrophic E. coli strain through formate oxidation by FDH from C. necator.

The soluble NAD- and Mo-dependent native formate dehydrogenase from *C. necator* is coded by the *fdsGBACD* operon, with the genes *fdsGBA* specifying the three enzyme subunits and the genes *fdsCD* specifying two chaperones shown to be essential for enzyme activity [17, 18]. FdsD was recently shown to be part of the FdsGBAD heterotetrametric functional unit of the closely related FDH from *Rhodobacter capsulatus* [8]. The operon was amplified by PCR from chromosomal *C. necator* DNA and cloned into plasmid pTRC99a. We chose this vector for its strong inducible tac promoter assuring high operon expression. The resulting plasmid pTRC-CnFDH (pGEN1340) was introduced into an *E. coli* MG1655 strain deleted for the gene *lpd* coding for lipoamide dehydrogenase (strain G5416) yielding strain G5663 (for strain and plasmid description, refer to Table 1). This enzyme, a component of the pyruvate dehydrogenase and the 2-oxoglutarate dehydrogenase complexes, catalyzes electron transfer from pyruvate and 2-oxoglutarate to NAD^+, respectively. *E. coli* strains lacking lipoamide dehydrogenase activity require, when fed with acetate as sole carbon source, an energy source in addition for growth. In this context, NAD-dependent formate oxidation to CO_2 can provide the necessary energy, as was shown for the monomeric NAD-dependent Fdh of *Pseudomonas* sp.101 [19]. When the expression of the plasmid-borne *C. necator fdsGBACD operon* was induced by IPTG addition in the culture of strain G5663 [Δ*lpd* pTRC-CnFDH], growth was obtained in mineral medium in the presence of formate (60 mM), acetate (20 mM) and pyruvate (20 mM) (Fig. 1). In contrast, no growth was observed when formate was omitted in the culture medium or in the case the energy auxotroph did not harbor the plasmid pTRC-CnFDH (Fig. 1), demonstrating that the *C. necator* NAD-dependent cytoplasmic formate dehydrogenase is functional when expressed in the *E. coli* host. As previously reported [20], we observed low growth yield on formate (60 mM) and acetate (20 mM) as sole carbon source, which was enhanced through pyruvate addition. Pyruvate when replacing acetate supported sustainable growth on formate as energy source. We speculated that the production of acetate through the action of pyruvate oxidase,

catalyzing the oxidative decarboxylation of pyruvate to acetate [21, 22] was responsible for this supporting effect. However, deletion of the gene *poxB* specifying the enzyme did only slightly affect growth, pyruvate still being a growth-enhancing factor (not shown). Acetate might be produced from pyruvate by an activity other than pyruvate oxidase, pyruvate formate lyase (Pfl) being a candidate, even so this enzyme is described inactive in the presence of oxygen. In addition, pyruvate might function as supplementary carbon source through gluconeogenesis and as precursor of several amino acids, while it cannot function as electron donor for growth.

Acceleration of formate dependent growth in continuous culture.

To accelerate formate-dependent growth, a cell population of NADH-requiring strain G5663 growing in selective medium (formate/acetate/pyruvate, see materials and methods) in the presence of IPTG was subjected to a turbidostat in GM3 continuous culture automatons (see materials and methods). Two independent cultures (UOF1 and UOF2) were launched in parallel. Both cultures were characterized by a short adaptation phase during which the initial generation time dropped rapidly from around 4h40 to stabilize at 3h40. Growth of both cell populations continuously accelerated (Fig. 2) until reaching a plateau with a generation time of 2h15 for both cultures, representing a diminution of about 1h25 as counted from the first stabilized plateau. Three isolates were obtained from each culture and formate dependence of growth verified. Isolate G5823 from UOF1 culture was cured from plasmid pTRC-CnFDH upon serial culture in selective medium supplemented with glucose. Plasmid loss was verified by sensitivity to ampicilline and the absence of PCR amplification of the *fdsGBACD* operon. The cured cells (strain G5876) lost their capacity to use formate as an energy source. Introducing plasmid pTRC-CnFDH in the cured cells restored growth on selective medium, showing that the dependency on FDH-catalyzed formate oxidation for energy supply was maintained during strain evolution (Fig. 3).

To identify adaptive mutations entailing improved growth under selective conditions, we proceeded to whole genome Illumina sequencing of all six isolates and mapped the sequence reads onto the genome (chromosome and plasmid pTRC-CnFDH) of the ancestor strain (see materials and methods). The plasmid pTRC-CnFDH from all six isolates remained unmutated. Sequencing of the genomes of the isolates identified a total of nine point mutations, with eight genes harboring a non-synonymous mutation in their coding region and one mutation affecting an intergenic region (Supplementary Table 1). The only gene found to be affected in all isolates, albeit not carrying the same mutation, was *focA* coding for a bidirectional formate transporter, differing in the changed codon between isolates. Interestingly, one mutation (*focA* F7C) was fixed in isolates obtained from both evolved populations from UOF1 and UOF2 independent cultures. The pH-dependent channel FocA plays an important role

in the regulation of intracellular formate concentration, notably during mixed-acid fermentation [23, 24]. We tested the impact of *focA* mutations F7C (isolate G5823) and V97I (isolate G5848) on formate-dependent growth by exchanging the mutated with the wild type *focA* allele in the two isolates. A strong augmentation of the doubling time (3,2x for F7C, 4,3x for V97I) was observed for both derived strains harboring wt *focA* when grown in the formate/acetate/pyruvate test medium (Table 2). FocA regulates formate concentration in the cytoplasm by favoring formate influx or efflux depending on medium pH and the growth phase of cultures in anaerobic environments. In the UOF turbidostats, the population is constantly growing in the logarithmic phase (OD=0,4) at a pH favoring FocA activity in the formate efflux sense (culture medium pH=7.2). Possibly, the mutant variants which arose during evolution impact the fine-tuned regulation of the channel favoring influx or impeding efflux of formate as response to the pressure imposed by the selection.

To test whether the growth rate enhancement effect of the UOF-background was somehow related to the activity of the *C. necator* FDH, we transformed the cured strain G5876 with a pZE21 plasmid (Expressys) for constitutive expression containing the gene for formate dehydrogenase of *Thiobacillus* sp. KNK65MA [25] and compared its growth on formate as sole energy source with the unevolved strain G6272, also harboring the pZE-Ts*fdh* plasmid. In contrast to the *C. necator* enzyme, the enzyme of *Thiobacillus* is monomeric not involving metal centers. Fig. 4 shows that the evolved background also had an enhancing effect on growth under selective conditions when this enzyme was expressed, indicating an adaptation on carbon flux during evolution rather than a regulation of FDH activity related to the structure.

Chromosomal integration of the complex FDH.

To create a platform of stable *C. necator* FDH expression in *E. coli* enabling *in vivo* structure/function studies and the evolution of activity in continuous culture, we integrated the *fdsGBACD* operon in the IS10 site of *E. coli* strain G5876 behind a strong promoter and an RBS following a described protocol [26]. We used the evolved and cured energy auxotrophic strain G5876 originating from culture UOF1 for chromosomal integration to favor our chances to obtain growth on formate. Formate dependent growth was observed for the resulting strain G6435 and the essential implication of the integrated FDH operon demonstrated by CRISPR interference [27]. This method is based on the concomitant expression of a catalytically inactive dCas9 protein and a guide RNA targeting the gene or the operon to be silenced. The inactive dCas9 protein binds – guided by the gRNA - to the promoter or a gene locus near the N-terminus thus interfering with initiation or elongation of DNA transcription by the RNA polymerase. Fig. 5 shows the results for the *C. necator fdsGBACD* operon silenced with three different gRNAs specific for the operon (see materials and methods). Overnight culture samples were serially diluted and dotted on

permissive or test plates containing or not the dCas9 expression inducer anhydrotetracycline. Slight growth inhibition was noticed on permissive medium in the presence of the inducer, suggesting residual DNA cleavage activity of dCas9 independent of the presence of a gRNA. On test medium without inducer of dCas9 expression, growth of control cells was observed. Expression of a specific gRNA strongly impeded cell growth under these conditions, demonstrating loss of growth on formate through silencing of the formate dehydrogenase. In the presence of the inducer, virtually no growth was observed, reflecting the deleterious effect of residual dCas9 activity on the Δ*lpd* strains due to their attenuated growth on the test medium.

Strain G6435 (Table 1) grew with a generation time of 3h30 in the formate/acetate/pyruvate test medium. As expected, induced FDH expression from the multicopy-plasmid pTrc99a supported faster growth on formate (Tgen=1h30) than expression from the single chromosomal copy integrated in strain G6435 (Fig. 6). The linear correlation between formate concentration and growth was demonstrated for the UOF1 isolate G5823 (Fig. 7A) and from its derived strain G6435 (Fig. 7B) for a formate range of 5 to 100 mM.

We conducted quantitative PCR to compare the expression levels of the enzyme in the two expression formats (normalized for *fdsA*, Table 3). Without surprise, expression of the three genes in the non-evolved plasmid-bearing G5663 context was higher in permissive than in selective medium. For all contexts tested, expression diminished for genes more distant from the promoter, reflecting lower processivity of the RNA polymerase towards the 3' end of the polycistronic reading frame [28]. The expression level between strain G5663 and its evolved descendant G5823 were comparable for the genes *fdsG* and *fdsA*, but differed for *fdsD*. Given that the mutational analysis of pGEN1340 from strain G5823 did not reveal mutations, the difference observed for *fdsD* expression is not easily explainable. Finally, when comparing the two expression formats, plasmid and chromosome, a lower expression was seen for the G6435 strain which is consistent with the lower growth rate observed when this strain was compared with strain G5823 (Table 3).

Strain G6435 was subjected to turbidostat evolution in a GM3 and a decrease in generation time of about 1h within 55 days (around 500 generations) was observed for two parallel cultures which reached Tgen=2h32 and 2h14 for cultures OCF5 and OCF6 respectively (Fig. 8).

Conclusion

Complex metal- and NAD-dependent formate dehydrogenases have been identified and studied in recent years. Purification and *in vitro* activity tests under oxic conditions demonstrated oxygen resistance of these enzymes, which was enhanced by stabilizing factors like nitrogen azide [6].

In this study we addressed the question whether such enzymes could stably function *in vivo* under aerobic conditions. An energy auxotrophic *E. coli* strain was used as test system to validate FDH activity for formate oxidation to generate NADH necessary for cell growth [20]. The soluble formate dehydrogenase from *C. necator* expressed from plasmids or from the chromosome supported stable formate-dependent growth in the presence of O_2. The plasmid-bearing and the chromosomal insertion strains were evolved in continuous culture for faster growth for up to 400 generations, without loss of the selective formate/FDH dependency of the populations. Genomic sequencing revealed adaptive mutations in the genetic background of evolved isolates, while the sequence of the heterologous FDH operon was found unchanged.

The FDH chromosomal insertion construct is of special interest as it provides a stable expression platform not only for continuous culture evolution, but also for *in vivo* site directed or targeted random mutagenesis. In recent years, methods were developed to enable mutagenesis and selection in the same cellular background. Key residues directly involved in the catalytic activity could be identified giving insights into structure/function relationships of these complex enzymes by *in vivo* activity screens avoiding protein overexpression and purification.

As a further perspective, FDH activities could be enhanced for the reductive reaction, using recently constructed formate dependent *E. coli* strains as selection chassis [29, 30]. Efficient enzymatic CO_2 reduction to formate can be envisioned as an entry point for CO_2 assimilation for biomass production engineered in initially heterotrophic model strains like *E. coli*. The metal- and NAD-dependent formate dehydrogenases, characterized by their oxygen tolerance and a CO_2 reduction activity up to 20x higher as compared to non-metal FDHs, are promising candidates for the implementation of synthetic autotrophic growth modes.

Materials and Methods

FDH plasmid constructions

The *fdsGBACD* operon (gene IDs 10917038-10917042) coding for the soluble Mo-dependent formate dehydrogenase of *Cupriavidus necator* DSM 13513 (CnFDH) was PCR amplified from genomic DNA using oligonucleotide primers 6125 GAGGTTAATTAAATGCCAGAAATTGCCCCCCAC (fwd) and 6126 ACAGCCAAGCTTTTACTCCAGCATCGCCCGATG (rev). The PCR fragment was gel purified and inserted into plasmids pTRC99a and pKI_IS10 (gift of S. Wenk) using HiFi DNA Assembly Cloning Kit (NEB. pZE21) for Gibson cloning. To obtain integrative plasmid pKI_IS10_CnFDH cloning was conducted in *E. coli* DH5α λ pir cells. A version optimized for *E. coli* codon usage of the gene coding for formate dehydrogenase from *Thiobacillus sp*. (AB106890) was synthetized by Twist Biosciences, California and cloned into plasmid pZE21 (Expressys, kan^r, colE1 origin, tet promoter) using the CPEC protocol [31].

Strain constructions

The strains used or constructed in this study were all derivatives of the wild type *E. coli* K12 strain MG1655. Their relevant genotypes and filiations are listed in Table 1. The desired genetic contexts were obtained by phage P1-mediated transductions of gene knockouts substituted by antibiotic resistance cassettes according to the method of [32]. Genes of interest were mobilized in the desired recipient cells by co-transduction with closely linked kanamycin markers originating from the Keio *E. coli* knockout collection [33]. Resistance cassettes were removed by flippase reaction after transformation with the plasmid pCP20 . The *fdsGBACD* operon flanked in 3′ by a strong promotor and an RBS was inserted in the chromosomal IS10 site [26] of MG1655 by recombination with plasmid pKI_IS10_CnFDH. The plasmid was transformed into the *E. coli pir*$^+$ donor strain ST18 [34] and transferred into recipient strain MG1655 by conjugation. Plasmid integration and subsequent removal of the plasmid backbone were selected as described [26]. The *fdsGBACD* operon was PCR amplified and correct integration verified by sequencing.

Continuous culture

Evolution experiments in continuous culture were carried out using GM3 fluidic self-cleaning cultivation devices. This device automatically dilutes growing cell suspensions with nutrient medium by keeping the culture volume constant. A continuous gas flow of controlled composition through the culture vessel ensures constant aeration and counteracts cell sedimentation. Twin culture vessels connected with silicon tubing enable the periodical transfer of the evolving culture between vessels and their cleaning upon rinsing with a 5N NaOH solution to remove biofilms [35].

To evolve G5663 cells to faster growth on formate as energy source, a turbidostat regime was programmed. This cultivation regime enables the selection of optimized growth in permissive conditions. Every 10 min, the optical density of the culture is automatically measured and compared to a fixed threshold (OD_{600} value of 0.4). When the measured OD_{600} exceeds the threshold, a pulse of fresh nutrient medium is injected into the culture and the same volume of used culture discarded. The dilutions ensure that the biomass in the vessel remains constant and that the bacteria grow at their maximal growth rate. A preculture of G5663 cells was grown in MS formate (60 mM) acetate (20 mM) pyruvate (20 mM) medium supplemented with IPTG (100 µM) at 30°C to an OD_{600nm} of 0,8 and used to inoculate two independent culture vessels (UOF1 and UOF2) with the same medium composition. Samples of the growing cultures were taken once a week and kept at -80°C. Growth was stopped after 230 (UOF1)and 380 (UOF2) generations and culture samples plated on semisolid MS formate acetate pyruvate medium to obtain isolates from colonies for further analysis. Cultures OCF5 and OCF6 were inoculated in a GM3 device with a preculture of strain G6435 grown in MS formate (60 mM) acetate (20 mM) pyruvate (20 mM) medium at 30°C to an OD_{600nm} of 0,8 and used to inoculate two independent culture vessels. Samples of the growing cultures were taken once a week and kept at -80°C. Growth was stopped after 500 generations and culture samples plated on semisolid MS formate acetate pyruvate medium to obtain isolates from colonies for further analysis.

Bacterial growth assays

A Microbiology Reader Bioscreen C apparatus (Thermo Fisher Scientific) was used for growth curve recordings. It consists of a thermostatic incubator and a culture growth monitoring device (OD reader). Overnight bacterial cultures were washed once in MS medium and diluted 100-fold in the respective growth medium; 200 µl aliquots of the cell suspensions were distributed into honeycomb 100-wells plates. Each experiment was performed in triplicate. The plates were incubated at 30° or 37°C under continuous agitation. Bacterial growth was followed by recording optical densities at 600 nm every 15 minutes during the indicated time.

Whole genome sequencing and mutation analysis

Pair-end libraries (2x150 bp) were prepared from 1 µg of genomic DNA of the evolved isolates and sequenced using an MiSeq sequencer (Illumina).

High-throughput sequencing data were analyzed using the PALOMA bioinformatic pipeline implemented in the MicroScope platform [36] (https://mage.genoscope.cns.fr /microscope/home/). In a

first step, reads were mapped onto the *E. coli* MG1655 reference (NC_000913.3) using the SSAHA2 package (v.2.5.1). Only unique matches having an alignment score equal to at least half of their length were retained as seeds for full Smith-Waterman realignment [37] with a region extended on both sides by five nucleotides of the reference genome. All computed alignments then were screened for discrepancies between read and reference sequences and a score based on coverage, allele frequency, quality of bases, and strand bias was computed for each detected event to assess its relevance. The mutations (single nucleotide variations and short insertions or deletions) with a score superior to 0.8 with at least five supporting reads were retained.

Gene silencing

Gene silencing was performed using CRISPRi method [27]. Specific gRNAs were cloned into plasmid pFD152 (gift from Solange Miele) harboring the inducible gene coding for dCas9 and a gRNA cloning sites. Three gRNAs (gRNA1: GGCGCCACGTGGTACAGGTC, gRNA2: AGGTGCATGGCGTGATCACC, gRNA3: AAGCGCTGGCCGAGCATGCG) were cloned into plasmid pFD152 using Golden Gate technique (Bsa I) and tested for silencing of the *fdsGBACD* operon. Plasmids expressing a specific gRNA were transformed into the strain G6435 and serial dilutions of overnight cultures were dotted on large Petri dishes in permissive and test conditions with and without induction of dCas9 by 0.5 µg/ml anhydrotetracycline and incubated for 2 to 7 days at 30°C.

Expression analysis by reverse transcriptase quantitative PCR (RT-qPCR)

The mRNA levels of *fdsA, fdsG* and *fdsD* genes were determined by RT-qPCR with *panB* as internal standard for expression normalization. Cultures were perfomed in MS mineral medium supplemented with either glucose 0.1 %, acetate 20 mM, pyruvate 20 mM, formate 60 mM, IPTG 0.1 mM and carbenicillin 100 mg/L if necessary or acetate 20 mM, pyruvate 20 mM, formate 60 mM and IPTG 0.1 mM if necessary. Cells were harvested in exponential phase (OD_{600} 0.5-0.6). Total RNA was extracted using the RNeasy Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. In summary, 2 volumes of RNAprotect Bacteria Reagent (Qiagen, Hilden, Germany) are added to one volume of bacterial culture. After pelleting the cells, RNA extraction was performed and RNA treated with DNase I (NEB). The quality of the extraction was verified by agarose gel electrophoresis. cDNA were generated through reverse transcription using High capacity cDNA Reverse transcription kit (Applied Biosystems) and their concentrations were determined using Qubit™ ssDNA Assay Kit (Invitrogen). Quantitative real-time PCR experiments were performed with three biological replicates, each analyzed

in three technical replicates using the KAPA SYBR FAST kit (Roche). Primer pairs used for amplification of genes *fdsA, fdsG* and *fdsD* and *panB* are listed in the table below:

panB	4463	TTAGAAGCTGCTGGGGCACA
	4464	CCGTTTCGGCGAGGAAATT
fdsG	6419	ATCCTGCATGAGATCCAGGACAC
	6420	AAGTGGTGGTAGAAGGTGATCACG
fdsA	6422	AACGGCAATTGCGAACTGCAG
	6423	ATTCGTCCTTCTTCATCTGCGTGTG
fdsD	6427	ACAACCTCATCACCATGGCCAAC
	6428	ATATCCAACAGCCCGTTCCCTG

Funding

This work was supported by Commissariat à l'Énergie Atomique et aux Énergies Alternatives (CEA), Fundamental Research Division (DRF), the CNRS, and the University of Évry Val d'Essonne. M. Schulz was recipient of PhD scholarship from CEA.

References

1. Ferry JG. Formate dehydrogenase. FEMS microbiology reviews. 1990;7(3-4):377-82. Epub 1990/12/01. doi: 10.1111/j.1574-6968.1990.tb04940.x. PubMed PMID: 2094290.

2. Popov VO, Lamzin VS. NAD(+)-dependent formate dehydrogenase. The Biochemical journal. 1994;301 (Pt 3)(Pt 3):625-43. Epub 1994/08/01. doi: 10.1042/bj3010625. PubMed PMID: 8053888; PubMed Central PMCID: PMCPMC1137035.

3. Zhang L, Liu J, Ong J, Li SF. Specific and sustainable bioelectro-reduction of carbon dioxide to formate on a novel enzymatic cathode. Chemosphere. 2016;162:228-34. Epub 2016/08/09. doi: 10.1016/j.chemosphere.2016.07.102. PubMed PMID: 27501309.

4. Son EJ, Ko JW, Kuk SK, Choe H, Lee S, Kim JH, et al. Sunlight-assisted, biocatalytic formate synthesis from CO2 and water using silicon-based photoelectrochemical cells. Chemical communications (Cambridge, England). 2016;52(62):9723-6. Epub 2016/07/15. doi: 10.1039/c6cc04661d. PubMed PMID: 27411734.

5. Maia LB, Moura I, Moura JJG. Molybdenum and tungsten-containing formate dehydrogenases: Aiming to inspire a catalyst for carbon dioxide utilization. Inorganica Chimica Acta. 2017;455:350-63. doi: https://doi.org/10.1016/j.ica.2016.07.010.

6. Yu X, Niks D, Mulchandani A, Hille R. Efficient reduction of CO(2) by the molybdenum-containing formate dehydrogenase from Cupriavidus necator (Ralstonia eutropha). The Journal of biological chemistry. 2017;292(41):16872-9. Epub 2017/08/09. doi: 10.1074/jbc.M117.785576. PubMed PMID: 28784661; PubMed Central PMCID: PMCPMC5641872.

7. Yu X, Niks D, Ge X, Liu H, Hille R, Mulchandani A. Synthesis of Formate from CO(2) Gas Catalyzed by an O(2)-Tolerant NAD-Dependent Formate Dehydrogenase and Glucose Dehydrogenase.

Biochemistry. 2019;58(14):1861-8. Epub 2019/03/07. doi: 10.1021/acs.biochem.8b01301. PubMed PMID: 30839197.

8. Radon C, Mittelstädt G, Duffus BR, Bürger J, Hartmann T, Mielke T, et al. Cryo-EM structures reveal intricate Fe-S cluster arrangement and charging in Rhodobacter capsulatus formate dehydrogenase. Nature communications. 2020;11(1):1912. Epub 2020/04/22. doi: 10.1038/s41467-020-15614-0. PubMed PMID: 32313256; PubMed Central PMCID: PMCPMC7171172.

9. Amao Y. Formate dehydrogenase for CO2 utilization and its application. Journal of CO2 Utilization. 2018;26:623-41. doi: https://doi.org/10.1016/j.jcou.2018.06.022.

10. Ma Z, Legrand U, Pahija E, Tavares JR, Boffito DC. From CO2 to Formic Acid Fuel Cells. Industrial & Engineering Chemistry Research. 2021;60(2):803-15. doi: 10.1021/acs.iecr.0c04711.

11. Yishai O, Lindner SN, Gonzalez de la Cruz J, Tenenboim H, Bar-Even A. The formate bio-economy. Current opinion in chemical biology. 2016;35:1-9. Epub 2016/07/28. doi: 10.1016/j.cbpa.2016.07.005. PubMed PMID: 27459678.

12. Yishai O, Goldbach L, Tenenboim H, Lindner SN, Bar-Even A. Engineered Assimilation of Exogenous and Endogenous Formate in Escherichia coli. ACS synthetic biology. 2017;6(9):1722-31. Epub 2017/05/31. doi: 10.1021/acssynbio.7b00086. PubMed PMID: 28558223.

13. Nielsen CF, Lange L, Meyer AS. Classification and enzyme kinetics of formate dehydrogenases for biomanufacturing via CO(2) utilization. Biotechnology advances. 2019;37(7):107408. Epub 2019/06/15. doi: 10.1016/j.biotechadv.2019.06.007. PubMed PMID: 31200015.

14. Schuchmann K, Müller V. Direct and reversible hydrogenation of CO2 to formate by a bacterial carbon dioxide reductase. Science (New York, NY). 2013;342(6164):1382-5. Epub 2013/12/18. doi: 10.1126/science.1244758. PubMed PMID: 24337298.

15. Çakar MM, Ruupunen J, Mangas-Sanchez J, Birmingham WR, Yildirim D, Turunen O, et al. Engineered formate dehydrogenase from Chaetomium thermophilum, a promising enzymatic solution for biotechnical CO(2) fixation. Biotechnology letters. 2020;42(11):2251-62. Epub 2020/06/20. doi: 10.1007/s10529-020-02937-7. PubMed PMID: 32557118.

16. Bruinsma L, Wenk S, Claassens NJ, Martins Dos Santos VAP. Paving the way for synthetic C1 - Metabolism in Pseudomonas putida through the reductive glycine pathway. Metabolic engineering. 2023;76:215-24. Epub 2023/02/23. doi: 10.1016/j.ymben.2023.02.004. PubMed PMID: 36804222.

17. Oh JI, Bowien B. Structural analysis of the fds operon encoding the NAD+-linked formate dehydrogenase of Ralstonia eutropha. The Journal of biological chemistry. 1998;273(41):26349-60. Epub 1998/10/03. doi: 10.1074/jbc.273.41.26349. PubMed PMID: 9756865.

18. Hille R, Hall J, Basu P. The mononuclear molybdenum enzymes. Chemical reviews. 2014;114(7):3963-4038. Epub 2014/01/29. doi: 10.1021/cr400443z. PubMed PMID: 24467397; PubMed Central PMCID: PMCPMC4080432.

19. Wenk S, Schann K, He H, Rainaldi V, Kim S, Lindner SN, et al. An "energy-auxotroph" Escherichia coli provides an in vivo platform for assessing NADH regeneration systems. Biotechnology and Bioengineering. 2020;117(11):3422-34. doi: https://doi.org/10.1002/bit.27490.

20. Wenk S, Schann K, He H, Rainaldi V, Kim S, Lindner SN, et al. An "energy-auxotroph" Escherichia coli provides an in vivo platform for assessing NADH regeneration systems. Biotechnology and bioengineering. 2020;117(11):3422-34. Epub 2020/07/14. doi: 10.1002/bit.27490. PubMed PMID: 32658302.

21. Gennis RBH, L. P. Pyruvate oxidase. In The Enzymes and Biological Membranes. 1976;2:493-504.

22. Abdel-Hamid AM, Attwood MM, Guest JR. Pyruvate oxidase contributes to the aerobic growth efficiency of Escherichia coli. Microbiology (Reading). 2001;147(Pt 6):1483-98. doi: 10.1099/00221287-147-6-1483. PubMed PMID: 11390679.

23. Kammel M, Pinske C, Sawers RG. FocA and its central role in fine-tuning pH homeostasis of enterobacterial formate metabolism. Microbiology (Reading). 2022;168(10). Epub 2022/10/06. doi: 10.1099/mic.0.001253. PubMed PMID: 36197793.

24. Peters K, Sargent F. Formate hydrogenlyase, formic acid translocation and hydrogen production: dynamic membrane biology during fermentation. Biochimica et biophysica acta Bioenergetics. 2023;1864(1):148919. Epub 2022/09/25. doi: 10.1016/j.bbabio.2022.148919. PubMed PMID: 36152681.

25. Nanba H, Takaoka Y, Hasegawa J. Purification and characterization of an alpha-haloketone-resistant formate dehydrogenase from Thiobacillus sp. strain KNK65MA, and cloning of the gene. Bioscience, biotechnology, and biochemistry. 2003;67(10):2145-53. Epub 2003/10/31. doi: 10.1271/bbb.67.2145. PubMed PMID: 14586102.

26. Wenk S, Yishai O, Lindner SN, Bar-Even A. An Engineering Approach for Rewiring Microbial Metabolism. Methods in enzymology. 2018;608:329-67. Epub 2018/09/04. doi: 10.1016/bs.mie.2018.04.026. PubMed PMID: 30173769.

27. Depardieu F, Bikard D. Gene silencing with CRISPRi in bacteria and optimization of dCas9 expression levels. Methods. 2020;172:61-75. doi: https://doi.org/10.1016/j.ymeth.2019.07.024.

28. Wang Y, Yue XJ, Yuan SF, Hong Y, Hu WF, Li YZ. Internal Promoters and Their Effects on the Transcription of Operon Genes for Epothilone Production in Myxococcus xanthus. Frontiers in bioengineering and biotechnology. 2021;9:758561. Epub 2021/11/16. doi: 10.3389/fbioe.2021.758561. PubMed PMID: 34778232; PubMed Central PMCID: PMCPMC8579030.

29. Kim S, Lindner SN, Aslan S, Yishai O, Wenk S, Schann K, et al. Growth of E. coli on formate and methanol via the reductive glycine pathway. Nature chemical biology. 2020;16(5):538-45. Epub 2020/02/12. doi: 10.1038/s41589-020-0473-5. PubMed PMID: 32042198.

30. Delmas VA, Perchat N, Monet O, Fouré M, Darii E, Roche D, et al. Genetic and biocatalytic basis of formate dependent growth of Escherichia coli strains evolved in continuous culture. Metabolic engineering. 2022;72:200-14. Epub 2022/03/29. doi: 10.1016/j.ymben.2022.03.010. PubMed PMID: 35341982.

31. Quan J, Tian J. Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. Nature protocols. 2011;6(2):242-51. Epub 2011/02/05. doi: 10.1038/nprot.2010.181. PubMed PMID: 21293463.

32. Miller JH. Experiments in Molecular Genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. 1972.

33. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, et al. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Molecular systems biology. 2006;2:2006.0008. Epub 2006/06/02. doi: 10.1038/msb4100050. PubMed PMID: 16738554; PubMed Central PMCID: PMCPMC1681482.

34. Jackson SA, Fellows BJ, Fineran PC. Complete Genome Sequences of the Escherichia coli Donor Strains ST18 and MFDpir. Microbiology resource announcements. 2020;9(45). Epub 2020/11/07. doi: 10.1128/mra.01014-20. PubMed PMID: 33154010; PubMed Central PMCID: PMCPMC7645665.

35. Mutzel R, Marliere, P. Method and device for selecting accelerated proliferation of living cells in suspension. Patent WO2000034433 A1. 2000.

36. Vallenet D, Calteau A, Dubois M, Amours P, Bazin A, Beuvin M, et al. MicroScope: an integrated platform for the annotation and exploration of microbial gene functions through genomic, pangenomic

and metabolic comparative analysis. Nucleic acids research. 2020;48(D1):D579-d89. Epub 2019/10/28. doi: 10.1093/nar/gkz926. PubMed PMID: 31647104; PubMed Central PMCID: PMCPMC7145621.

37. Smith TF, Waterman MS. Identification of common molecular subsequences. Journal of molecular biology. 1981;147(1):195-7. Epub 1981/03/25. doi: 10.1016/0022-2836(81)90087-5. PubMed PMID: 7265238.

Tables

Table 1. Strains and plasmids used in this study.

Strain	Genotype	Origin/modification
MG1655	F-, LAM-, *rph-1*	CGSC Collection, Yale
JW0889	Δ*ycaP ::kanR*	Keio Collection,
G5416	Δ*lpd*	Laboratory collection
G5663	Δ*lpd* pGEN1340	Transformation of G5416
G5823	Δ*lpd focA* F7C *infC* E95K pGEN1340	Evolvant of G5663 selected in GM3 under turbidostat regime – UOF1 lineage
G5824	Δ*lpd focA* F7C *infC* E95K pGEN1340	Evolvant of G5663 selected in GM3 under turbidostat regime – UOF1 lineage
G5825	Δ*lpd focA* F7C *infC* E95K pGEN1340	Evolvant of G5663 selected in GM3 under turbidostat regime – UOF1 lineage
G5848	Δ*lpd focA* V97I *pps* S2F *crr* G2V *fabR* L53W pGEN1340	Evolvant of G5663 selected in GM3 under turbidostat regime – UOF2 lineage
G5849	Δ*lpd focA* V97I *pps* S2F *crr* G2V *fabR* L53W pGEN1340	Evolvant of G5663 selected in GM3 under turbidostat regime – UOF2 lineage
G5850	Δ*lpd focA* F7C *cra* Y28C pGEN1340	Evolvant of G5663 selected in GM3 under turbidostat regime – UOF2 lineage
G5876	Δ*lpd focA* F7C *infC* E95K	Cured G5823
G6234	Δ*lpd infC* E95K Δ*ycaP ::kanR* pGEN1340	Transduction G5823 x P1 (Keio JW0889)
G6235	Δ*lpd pps* S2F *crr* G2V *fabR* L53W Δ*ycaP ::kanR* pGEN1340	Transduction G5848 x P1 (Keio JW0889)

Strain	Genotype	Origin/modification
G6114	Δ*lpd* pGEN1393	Transformation of G5416
G6217	Δ*lpd* pGEN1395	Transformation of G5416
G6272	Δ*lpd.focA* F7C *infC* E95K pGEN1395	Transformation of G5876
G6280	Δ*lpd focA* F7C *infC* E95K pGEN1393	Transformation of G5876
G6408	IS10::*fdsGBACD C.n.::kan*	Insertion at IS10 by recombination plasmid pGEN1378 in MG1655
G6435	Δ*lpd focA* F7C *infC* E95K IS10::*fdsGBACD C.n.::kan*	Transduction G5876 x P1(G6408)
G6504	Δ*lpd focA* F7C *infC* E95K pGEN1340	Transformation of G5876

Plasmid	Description	Origin
pTrc99a	IPTG-inducible expression vector, pBR322 origin, *bla*$^{+}$ *lacI*$^{+}$	CGSC Collection, Yale
pFDH	constitutive expression pZE21 vector, pBR322 origin, streptomycine^{R}, *fdh Pseudomonas sp101*	Addgene (#131706)
pFD152	dCas9 (aTc-inducible), spectinomycine^{R}, gRNA cloning site	Depardieu & Bikard (2019)
pKI_IS10	R6K origin, conjugative, chloramphenicol^{R} kanamycine^{R},	Wenk *et al.* (2018)
pGEN1340	pTrc99a::*fdsGBACD Cupriavidus necator* native operon	This study
pGEN1378	pKI_IS10::*fdsGBACD Cupriavidus necator* native operon	This study
pGEN1393	pZE21::*fdsGBACD Cupriavidus necator* native operon	This study
pGEN1395	pZE21::*fdhThiobacillus sp. E. coli* adapted synthetic gene	This study

Table 2. Doubling times of evolved UOF isolates versus derivatives reverted to *focA* wild type.

lineage	strain	*focA*	doubling time[a]
UOF1	G5823	F7C	5h30
	G6234	wt	17h30
UOF2	G5848	V97I	5h40
	G6235	wt	23h20

[a] Cells were grown at 30°C in mineral MS medium supplemented with acetate 20 mM pyruvate 20 mM formate 60 mM in a Bioscreen C plate reader in triplicate. Doubling times in exponential growth phase were graphically estimated.

Table 3. Comparison of the transcriptional levels of the genes *fdsG, fdsA* and *fdsD* by quantitative PCR.

Comparison		**ΔΔCt (FC)[a]**		
Inter-conditions	**strain**	***fdsG***	***fdsA***	***fdsD***
selective medium + IPTG / permissive medium + IPTG	**G5663**	-0.14 (1.1)	1.98 (0.25)	3.29 (0.1)
Inter-strains	**strains**	***fdsG***	***fdsA***	***fdsD***
selective medium + IPTG	**G5823 - G5663**	0.68 (0.62)	0.65 (0.64)	1.92 (0.26)
selective medium + IPTG	**G6435 - G5823**	1.2 (0.43)	1.37 (0.39)	-1.12 (2.2)

[a] The relative levels of expression of *fds* genes were evaluated by the calculation of ΔCt (Ct Cycle threshold) using *E. coli panB* gene as internal control. Inter-strain or inter-condition differences of the level of expression of each *fds* gene were evaluated by the calculation of ΔΔCt. FC: fold change.

Figures

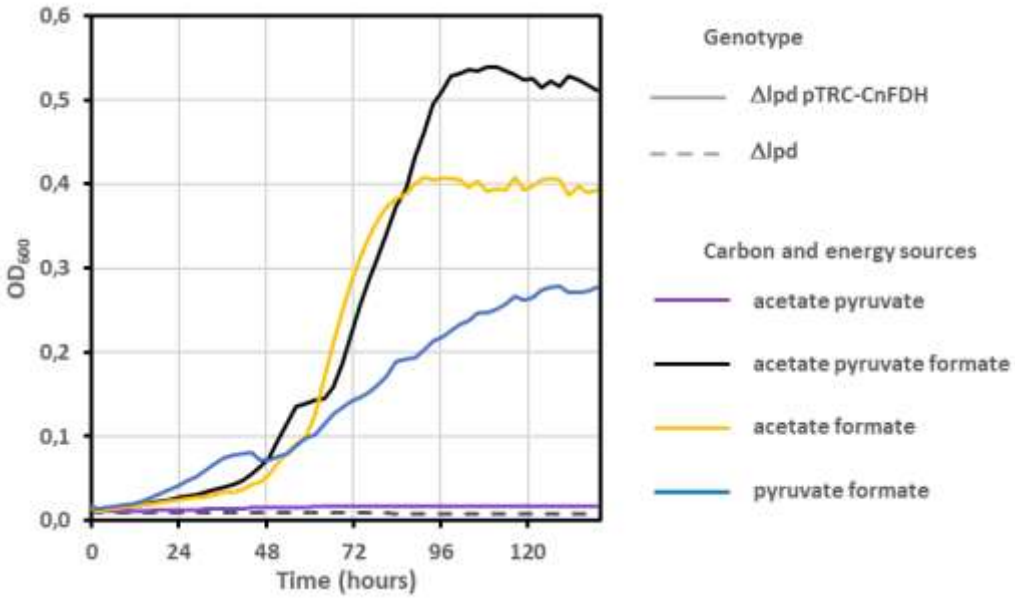


Fig 1. Expression of *C. necator* NAD-dependent formate dehydrogenase allows *E. coli* NADH auxotroph strain Δ*lpd* to use formate as energy source. Strains G5416 (Δ*lpd*) (broken line) and G5663 (Δ*lpd* pTRC-CnFDH) (plain line) were grown at 30°C on mineral MS medium supplemented with the indicated compounds. Concentrations of formate, acetate and pyruvate were 60, 20 and 20 mM, respectively. Growth was recorded with a Bioscreen C plate reader in triplicates.

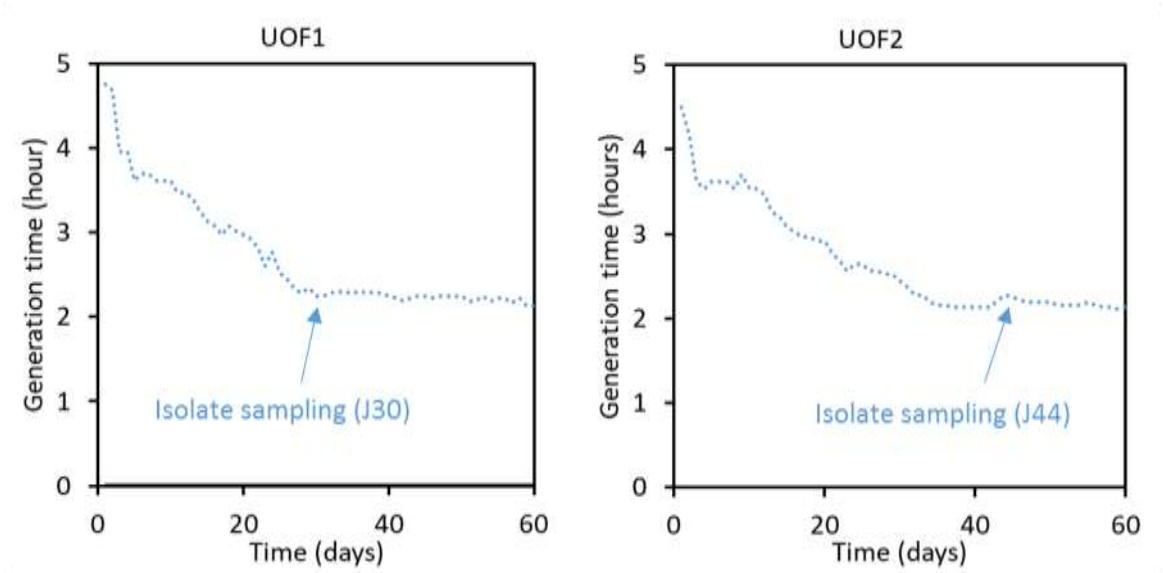


Fig 2. Growth acceleration of NADH auxotroph G5663 bacteria (Δ*lpd* pTRC-CnFDH) in turbidostat. Cells were grown in two independent cultures (UOF1 and UOF2) at 30°C in mineral MS medium supplemented with acetate (20 mM), pyruvate (20 mM) and formate (60 mM) in a GM3 device for 60 days. The time points of isolate samplings are indicated, corresponding to 230 generations in turbidostat for UOF1 and 380 generations in turbidostat for UOF2.

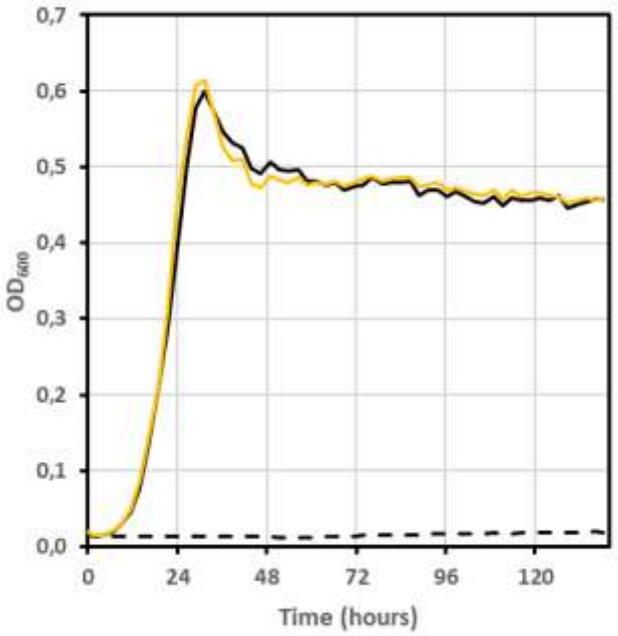


Fig 3. Dependence of evolved isolate G5823 (turbidostat culture UOF1) on the presence of plasmid-borne *C. necator* FDH for growth on formate as energy source. Strains G5823 (black line), G5876 (derivative of G5823 cured from the plasmid pTRC-CnFDH) (black broken line), and strain G6504 (derivative of G5876 transformed with plasmid pTRC-CnFDH) (yellow line) were grown at 30°C on mineral MS medium supplemented with formate (60 mM), acetate (20 mM) and pyruvate (20 mM). Growth was recorded with a Bioscreen C plate reader in triplicates.

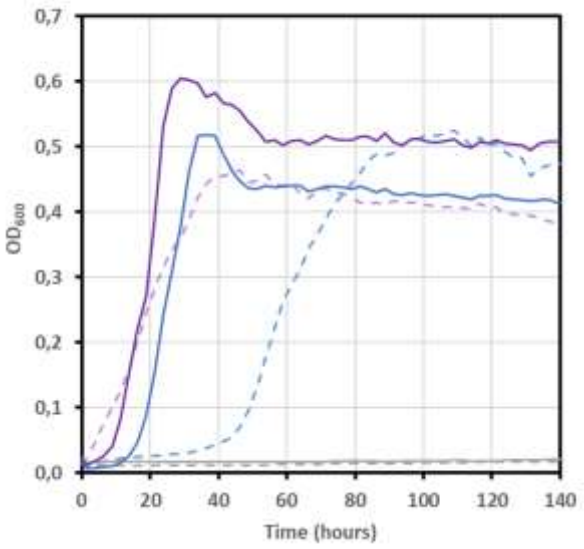


Fig 4. Influence of the genetic background on FDH-dependent growth. Formate-complemented growth of Δ*lpd* strains expressing *C. necator* or *Thiobacillus sp.* formate dehydrogenase was compared in unevolved genetic background of strain G5416 (broken line) and evolved genetic background of strain G5876 (cured derivative of UOF1 isolate G5823) (plain line). Growth of strains G6114 and G6217, derivatives of strain G5416 harboring plasmid pZE::CnFDH (blue broken line) or pZE::TsFDH (purple broken line), respectively, was compared with strains G6280 and G6272, derivatives of strain G5876 likewise harboring plasmids pZE::CnFDH (blue line) or pZE::TsFDH (purple line), respectively. Lack of growth of plasmid-free strains G5416 and G5876 (grey lines) demonstrate the dependence on heterologous FDH activity for cell proliferation under selective conditions. Bacteria were grown on mineral MS medium supplemented with formate (60 mM) acetate (20 mM) and pyruvate (20 mM) at 30°C in a Bioscreen C plate reader in triplicates.

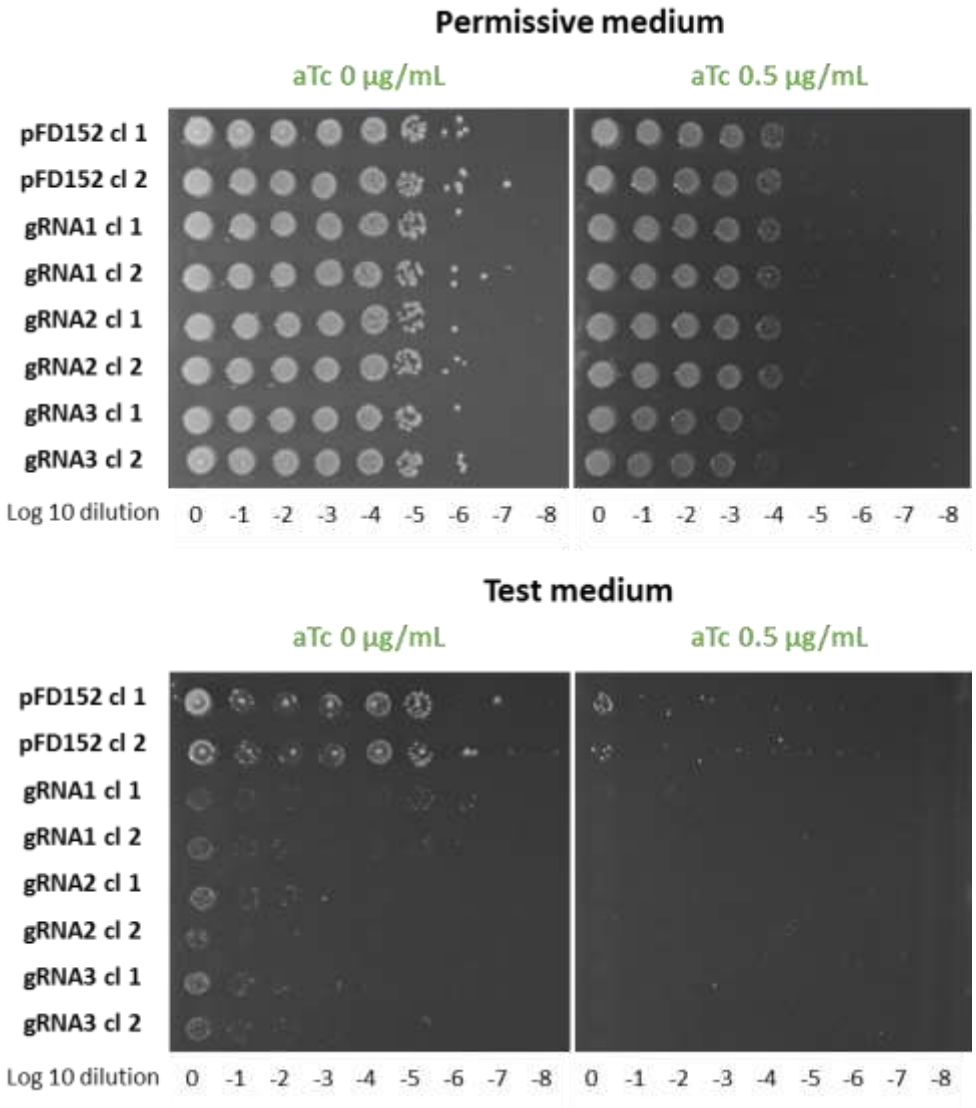


Fig 5. dCas9 silencing of *C. necator fdsGBACD* operon inserted in the chromosome of *E. coli* Δ*lpd* strain. Cells of strain G6435 (Δ*lpd* IS10::*fdsGBACD C.n.)* were grown overnight in permissive medium (MS glucose 0.2% acetate 20 mM), then serially diluted and dotted on semi-solid permissive or test medium (MS formate 60mM acetate 20mM pyruvate 20mM) containing or not the dCas9 inducer anhydrotetracycline (aTc) as indicated. Plates were incubated at 30°C for a maximum of seven days Results are shown for G6435 cells harboring the empty plasmid pFD152 as control, and G6435 cells expressing one of three different gRNAs cloned in pFD152.

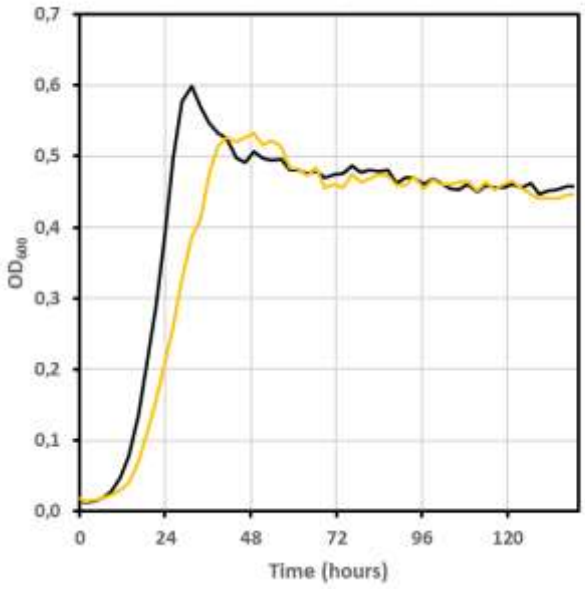


Fig 6. Impact of FDH expression context on growth. Growth of the UOF1 isolate G5823, which harbors plasmid pTRC-CnFDH (black line) was compared with growth of G5823 descendant strain G6435 containing of *C. necator* FDH operon *fdsGBACD* on the chromosome (yellow line). Bacteria were grown on mineral MS medium supplemented with formate (60 mM), acetate (20 mM) and pyruvate (20 mM) at 30°C in a Bioscreen C plate reader in triplicates.

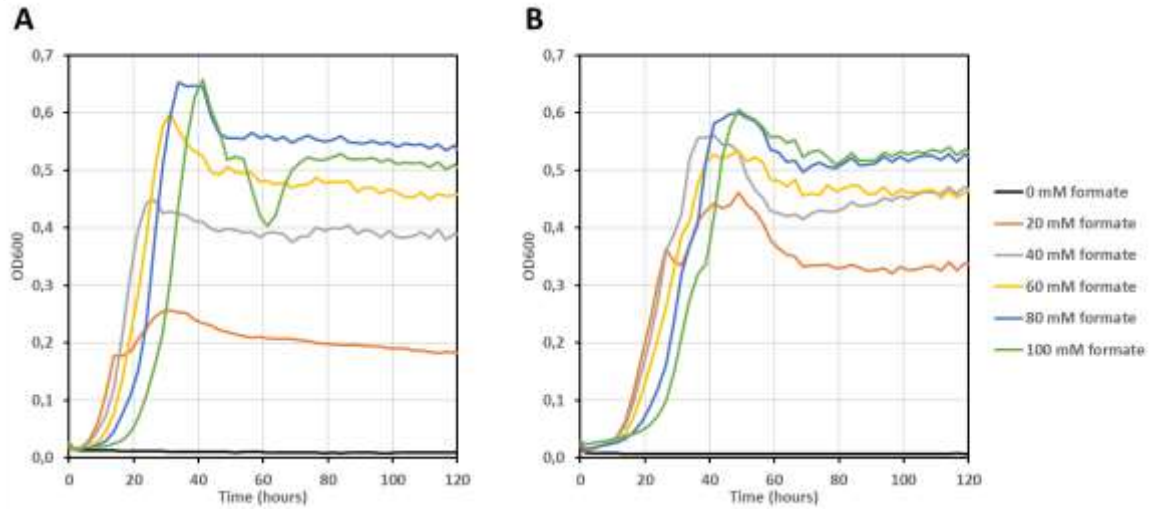


Fig 7. Growth of *C. necator* FDH expressing Δ*lpd* strains depends on formate concentration in the medium. Bacteria from UOF1 isolate G5823 (**A**) and from its derivative strain G6435 (**B**) harboring *C. necator fdsGBACD* operon on the chromosome were grown in mineral MS medium supplemented with acetate 20 mM, pyruvate 20 mM and formate at concentrations ranging from 0 to 100 mM. Overall growth rate was higher for G5823 cells expressing CnFDH from plasmid. The final OD_{600} of both cultures increased with increased formate concentration, the maximum OD being reached at 80 mM for both strains. Experiments were conducted with Bioscreen C plate reader as described in the Material and method section.

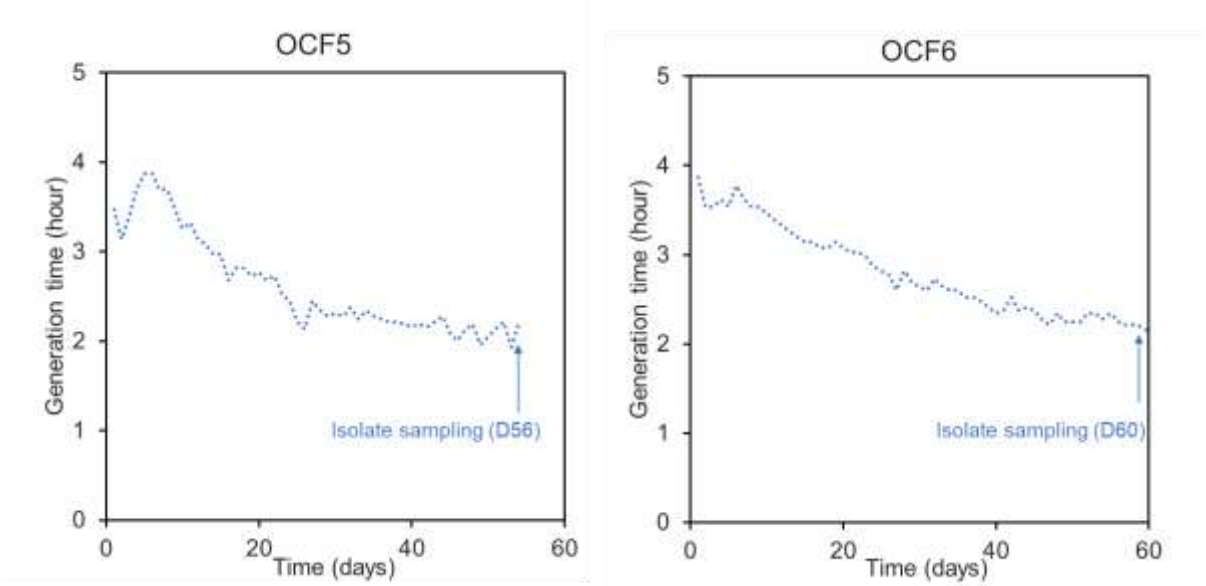


Fig 8. Growth acceleration of NADH auxotroph G6435 bacteria (Δ*lpd* IS10::*fdsGBACD C.n.*) in turbidostat. Cells were grown in two independent cultures (OCF5 and OCF6) at 30°C in mineral MS medium supplemented with acetate (20 mM), pyruvate (20 mM) and formate (60 mM) in a GM3 device for 60 days. The time points of isolate samplings are indicated, corresponding to about 500 generations in turbidostat for both cell populations.

Supplementary Table 1. Mutations fixed in the genome of the evolved strains

UOF1	chromosomal position*	mutation type	mutation event	*mutated gene*	intragenic position	amino acid change	intergenic mutation (distance to the flanking genes)	mutations identified in the indicated strain (•)		
								G5823	**G5824**	**G5825**
	953670	A/C	SNP	*focA*	20	F7C		•	•	•
	1029182	C/A	SNP				yccW/yccX (-77/-105)	•	•	•
	1798380	C/T	SNP	*infC*	283	E95K		•	•	•

UOF2	chromosomal position*	mutation type	mutation event	*mutated gene*	intragenic position	amino acid change	intergenic mutation (distance to the flanking genes)	mutations identified in the indicated strain (•)		
								G5848	**G5849**	**G5850**
	88110	A/G	SNP	*cra*	83	Y28C				•
	953401	C/T	SNP	*focA*	289	V97I		•	•	
	953670	A/C	SNP	*focA*	20	F7C				•
	1785132	G/A	SNP	*pps*	5	S2F		•	•	
	2533860	G/T	SNP	*crr*	5	G2V		•	•	
	4159247	T/G	SNP	*fabR*	158	L53W		•	•	

Appendix 10: GC method for H_2 detection

Used method: direct injection, CO_2 10 minutes.

- Carrier gas flow rate: 12 psi (20 mL/min), auxiliary gas flow rate: 0 mL/min.
- Temperature settings:

Oven temperature: 40 °C.

Detector temperature: 100 °C.

- Flame Ionization Detector (FID) settings:

FID temperature: 350 °C.

Hydrogen flow rate: 45 mL/min.

Air flow rate: 450 mL/min.

- Injection duration: each injection lasted for 10 minutes.

FID Analysis (Flame Ionization Detector): enables the quantification of organic gases (CH_4, CO_2, CO) present in the injected gas.

TCD Analysis (Thermal Conductivity Detector): enables the quantification of inorganic gases (H_2, O_2) present in the sample. Other organic gases can also be detected, albeit with lower sensitivity (e.g., CH_4).

GC injections: 50 µL of the gaseous atmosphere were sampled and injected into the GC using a gastight glass Hamilton syringe with an integrated needle. The amount of gas injected into the GC depends on the volume injected as well as the pressure and temperature conditions in the ambient atmosphere (ideal gas law). These latter parameters were measured using a barometric altimeter.

Precautions: before and between each sampling, the syringe needle and each pierced septum were disinfected with ethanol to prevent tube contamination. Subsequently, the syringe was purged with sterile nitrogen-filled tubing to avoid contaminating the culture tube with air or gas from previous samples.

Appendix 11: Test of strain G5684 in phosphite range

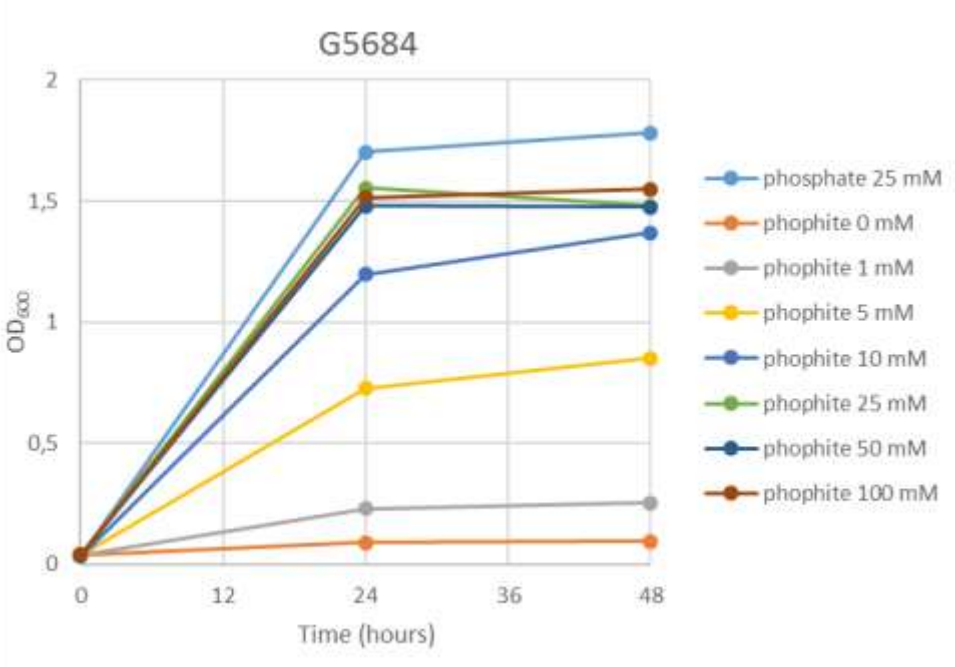


Figure: Growth of E. coli strain G5684 (Δ*phoA* Δ*phnGHIJK*) on a range of phosphite concentrations. The strains were grown in MS mineral medium supplemented with 0.2 % glucose and a phosphite range from 0 to 100 mM. A positive control with 25 mM phosphate was performed as well.

Appendix 12: FMN and Mo content from FDH production

The FMN and Mo content quantification were performed on one sample of RcFDH from a test batch.

The FMN analysis was performed by mass spectrometry, using a method described in a paper published by a laboratory at the Genoscope, LGBM (Perchat et al. 2018). This quantification was performed in collaboration with Peggy Sirvain, Ekaterina Darii and Alain Perret.

The calibration line used is available on the Figure below. This calibration line was used to determine the quantity of FMN in our samples by measuring it in two replicates.

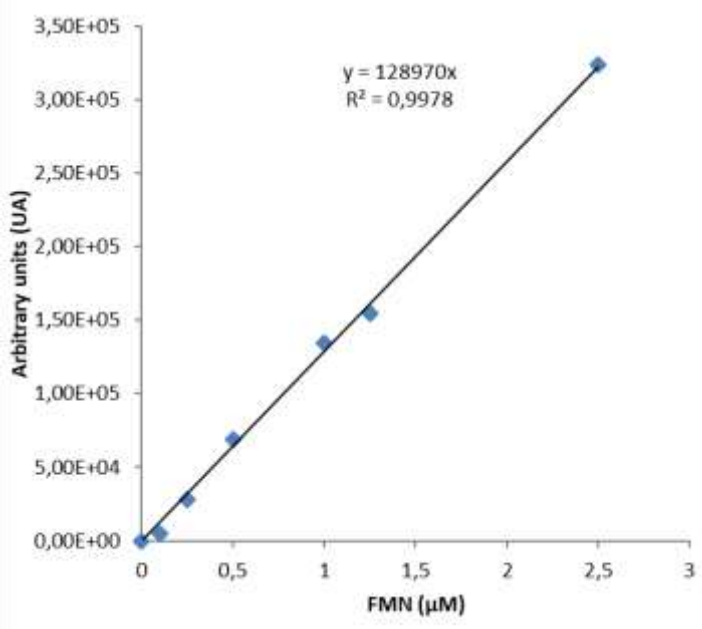
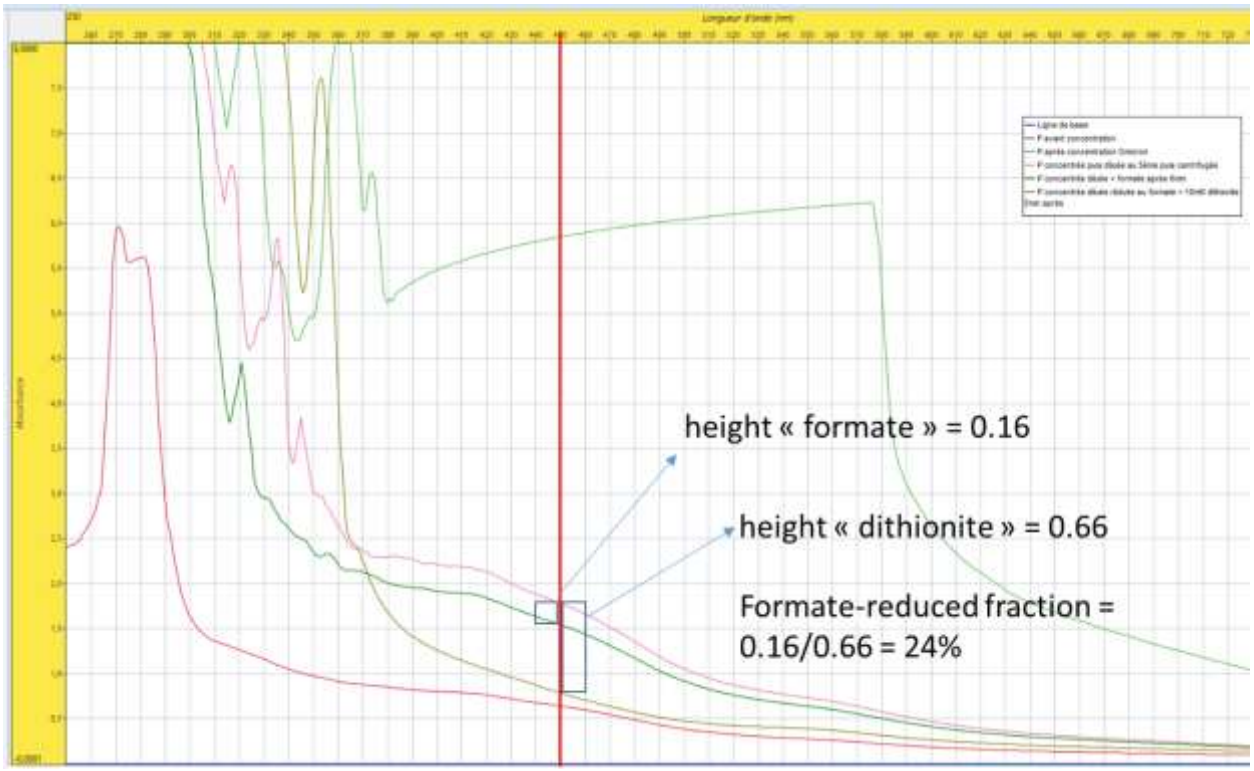


Figure: calibration line for FMN quantification.

To quantify the Mo-content of the RcFDH produced, reduction spectra were conducted (see Figure below). Initially, a UV-visible absorption spectrum of the oxidized protein was obtained (pink curve). Then, the protein sample was reduced using 10 mM formate, and after 5 minutes, another spectrum was recorded (dark green curve). Finally, 10 mM dithionite was added to the sample, allowed to stand for 5 minutes, and a final spectrum of the fully-reduced protein was acquired (greenish-brown curve). The Mo-content was graphically determined by the fraction of protein reduced by formate compared to a protein fully-reduced with dithionite.


height « formate » = 0.16
height « dithionite » = 0.66
Formate-reduced fraction =
0.16/0.66 = 24%

Appendix 13: *In vivo* mutagenesis Evolution.T7 method

This appendix will describes details of the different parts used with the Evolution.T7 method.

For the transcription/mutagenesis lead by the T7 RNA polymerase, the wild-type version as well as the mutated version, as explained above, for terminators were necessary to efficiently stop the translation process and placed behind the region to evolve. In the following table, the sequence of this four terminators is detailed. In addition, the specific promoters used for the construction of the plasmid are detailed in the same table.

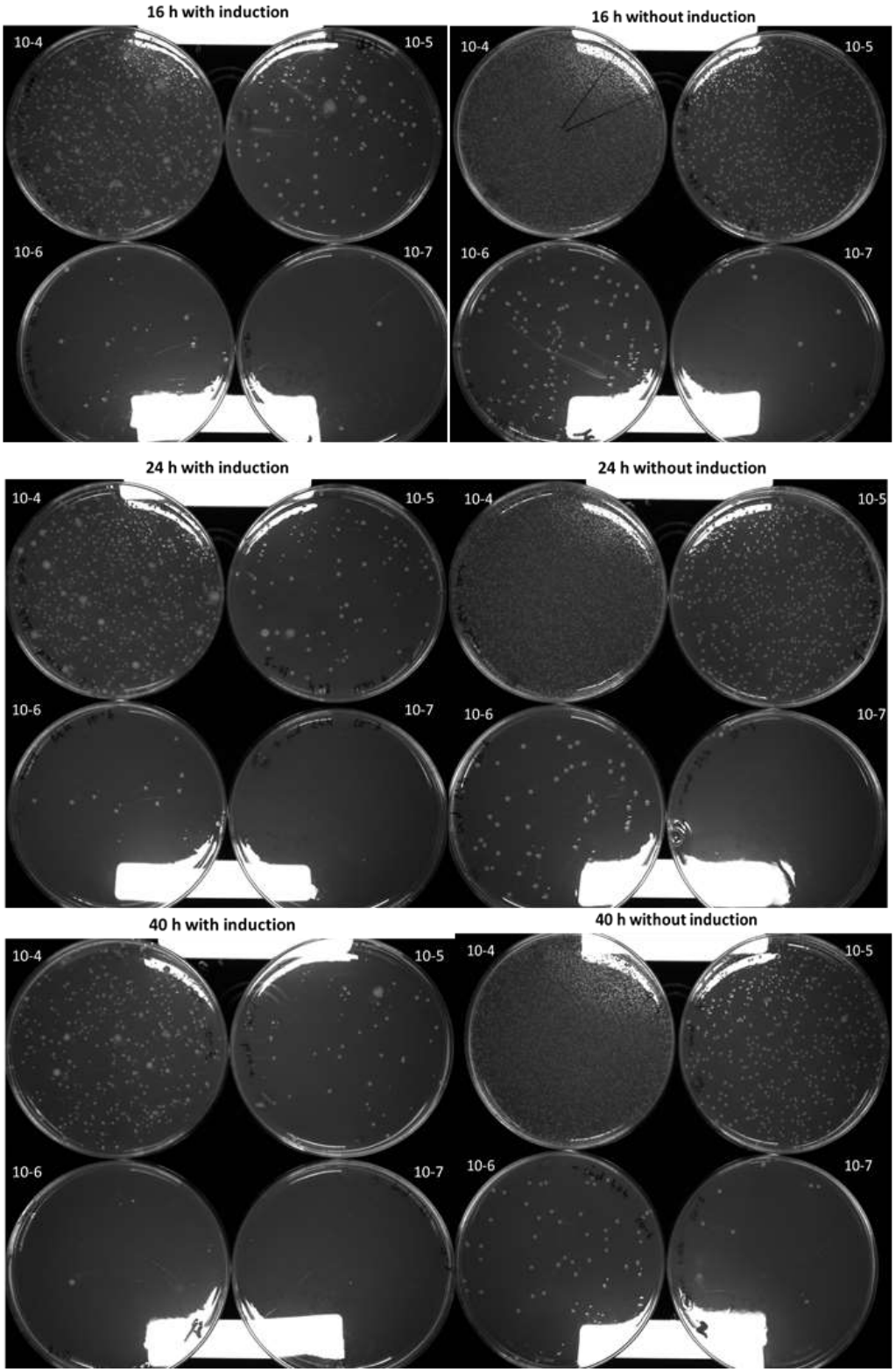
Building block	Sequence
BBa_B0015 Terminator	ccaggcatcaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttatctgttgtttgtc ggtgaacgctctctactagagtcacactggctcaccttcgggtgggcctttctgcgtttata
SBa_000587 Terminator	tactcgaacccctagcccgctcttatcgggcggctaggggttttttgt
T7 terminator	tagcataaccccttggggcctctaaacgggtcttgaggggttttttgt
SBa_000451 terminator	tacatatcgggggggtaggggttttttgt
BBa_J23110 Promoter	tttacggctagctcagtcctaggtacaatgctagc
P_{T7} promoter	taatacgactcactatagg
P_{T7*} promoter	tatagtgaccggtatta
P_moderate	acctattgacaattaaaggctaaaatgctataattccac

For the functional testing experiment, five couples of T7RNAP/deaminases were used and are detailed in the following table. In this table, the wild-type T7RNAP is designated as T7RNAP and the one recognizing the mutated T7 promoter, in named T7RNAP-CGG-R12-KIRV. For wild-type T7RNAP plasmid, the backbone used is pSEVA221, for the T7RNAP*, the backbone is a pSEVA471.

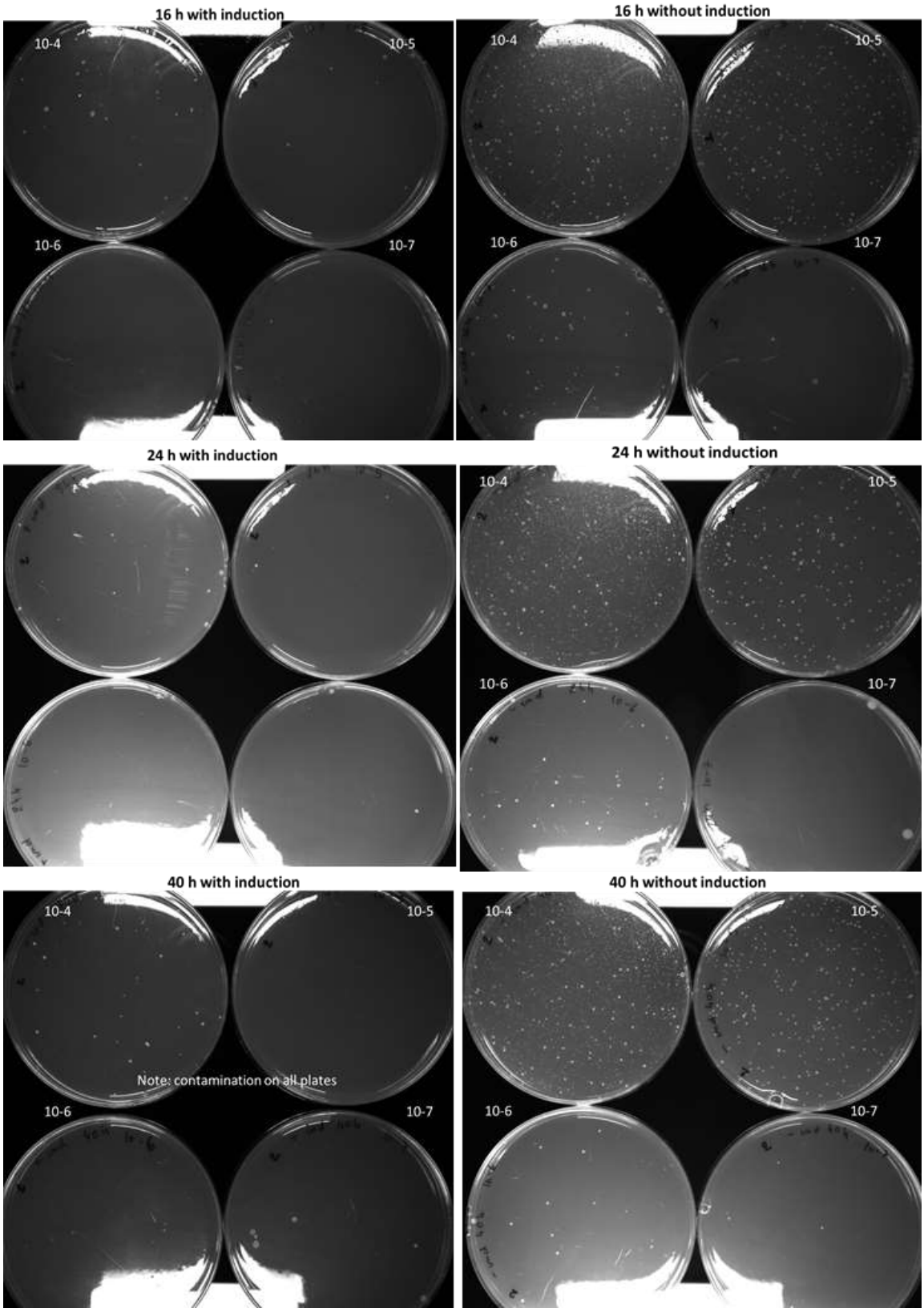
Name	T7RNAP/deaminase combination
Couple 1	evoAPOBEC1-BE4max-T7RNAP-CGG-R12-KIRV (p114) + ABE8.20- m-T7RNAP (p119)
Couple 2	rAPOBEC1-T7RNAP-CGG-R12-KIRV (p112) + ABE8.20-m-T7RNAP (p119)
Couple 3	evoCDA1-T7RNAP-CGG-R12-KIRV (p115) + evoCDA1-T7RNAP (p118)
Couple 4	TadA*-T7RNAP-CGG-R12-KIRV (p113) + rAPOBEC1-T7RNAP (p103)
Couple 5	pmCDA1-T7RNAP-CGG-R12-KIRV (p111)+ ABE8.20-m-T7RNAP (p119)

Appendix 14: Results of the Evolution.T7 method targeting TsFDH

The pictures of the plates with permissive medium are below:



The pictures of the plates with test medium are below:



ÉCOLE DOCTORALE
Structure et dynamique des systèmes vivants (SDSV)

université PARIS-SACLAY

Titre : Vers le développement d'une souche d'*Escherichia coli* autotrophe

Mots clés : Escherichia coli ; Autotrophie synthétique ; Formate deshydrogénase ; ingénierie métabolique

Résumé : Les micro-organismes autotrophes convertissent le CO_2 en biomasse en utilisant l'énergie lumineuse ou en oxydant des molécules inorganiques. En y exprimant des voies de biosynthèse pour la production de produits chimiques ou de carburants, elles offrent la possibilité de valoriser industriellement le CO_2. Cependant, ces organismes manquent souvent de flexibilité génétique et sont limités en termes de gamme de produits. Les connaissances actuelles sur le métabolisme bactérien et les diverses techniques de modification du génome rendent la mise en place d'un régime métabolique autotrophe dans un contexte hétérotrophe réalisable. *Escherichia coli*, un hétérotrophe anaérobie facultatif, est un organisme polyvalent doté d'un réseau métabolique flexible et de qualités prouvées en tant que plateforme biotechnologique, en faisant ainsi une cible privilégiée pour ce type d'adaptation. Le projet s'appuie sur l'utilisation de souches d'*E. coli* qui dépendent de l'apport de formate pour la biosynthèse de métabolites essentiels, pour manifester l'activité réductrice de formate déshydrogénases. Les formate déshydrogénases (FDH) NAD-dépendantes capables de réduire efficacement le CO_2 en formate ont une structure complexe et fonctionnent avec des centres métalliques, le plus souvent inactivés en présence d'O_2. Récemment, la réduction significative du CO_2 en formate lors de l'oxydation du NADH en présence d'O_2 a été montrée in vitro pour la FDH soluble et Mo-dépendante de *Cupriavidus necator* (CnFDH). Ce projet de recherche avait plusieurs objectifs. Tout d'abord, il visait à démontrer la fonctionnalité de la CnFDH in vivo au sein d'*E. coli* en utilisant diverses souches génétiquement modifiées. Deuxièmement, il avait pour but d'explorer des sources d'énergie alternatives et non carbonées, l'hydrogène et le phosphite, pour soutenir la croissance d'*E. coli*. Pour atteindre ces objectifs, une combinaison de design rationnel des souches et de culture continue automatisée dans des conditions sélectives (technologie GM3) a été utilisée pour exploiter le potentiel des complexes enzymatiques testés. De plus, l'adaptation d'une méthode de mutagenèse in vivo pour générer et sélectionner des mutants favorables d'enzymes régénérant le NADH a été explorée.

Title : Towards the construction of an autotrophic *Escherichia coli* strain

Keywords: *Escherichia coli* ; Synthetic autotrophy ; Formate dehydrogenase ; Metabolic engineering

Abstract: Autotrophic microorganisms convert CO_2 into biomass using energy from light or from the oxidation of inorganic molecules. When implementing biosynthetic pathways for the production of commodity chemicals or fuels, they offer the possibility to industrially valorize CO_2. However, these organisms often lack genetic malleability and are limited in product range. Current knowledge of bacterial metabolism and various genome modification techniques make the implementation of an autotrophic metabolic regime in a heterotrophic background feasible. *Escherichia coli*, a facultative anaerobic heterotroph, is a versatile organism with a flexible metabolic network and proven qualities as a biotechnological platform organism, thus being a preferred target for such adaptation. The project is based on the usage of *E. coli* strains that rely on the supply of formate for the biosynthesis of essential metabolites for bringing to light the reductive activity of Formate dehydrogenases. NAD-dependent formate dehydrogenases (FDH) capable of the efficient conversion of CO_2 to formate are structurally complex and operate with metal centers often entailing O_2 toxicity. Recently, efficient *in vitro* reduction of CO_2 to formate upon oxidation of NADH in the presence of O_2 was reported for the Mo-dependent complex and soluble FDH from *Cupriavidus necator* (CnFDH). This research project had several objectives. First, it aimed to demonstrate the functionality of CnFDH *in vivo* within *E. coli* using various genetically engineered strains. Second, it explored alternative non-carbon energy sources, namely hydrogen and phosphite, to support *E. coli* growth. To achieve these goals, a combination of rational strain design and automated continuous culture in selective conditions (GM3 technology) was used to harness the potential of the enzymatic complexes tested. Additionally, the adaptation of an *in vivo* mutagenesis method for generating and selecting favorable mutants of NADH-regenerating enzymes was explored.