# Characterization of the metabolism of Eubacterium limosum by a Quantitative Systems Biology Approach

Guillaume Pregnon


# THÈSE

## En vue de l'obtention du
## DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

**Délivré par l'Institut National des Sciences Appliquées de Toulouse**

---

**Présentée et soutenue par**

**GUILLAUME PREGNON**

Le 14 septembre 2022

## Caractérisation du métabolisme de Eubacterium limosum par une approche quantitative de biologie des systèmes

---

Ecole doctorale : **SEVAB - Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingenieries**

Spécialité : **Ingénieries microbienne et enzymatique**

Unité de recherche :
**TBI - Toulouse Biotechnology Institute, Bio & Chemical Engineering**

Thèse dirigée par
**Philippe SOUCAILLE**

Jury

**M. Kaspar VALGEPEA,** Rapporteur
**M. Benjamin WOOLSTON,** Rapporteur
**M. Peter DURRE,** Rapporteur
**M. Philippe SOUCAILLE,** Directeur de thèse
**Mme ISABELLE MEYNIAL-SALLES,** Présidente

**Résumé**

*Eubacterium limosum* est une bactérie anaérobie stricte, faisant partie du groupe des acétogènes. Son intérêt réside dans sa capacité à transformer des sources de carbone à un atome de carbone (C1) telle que le méthanol en butyrate, composé en C4 présentant un intérêt industriel. Néanmoins, les rendements et productivité doivent être améliorés pour se rapprocher des performances de l'industrie chimique. L'ingénierie métabolique des souches permettrait d'atteindre cet objectif mais pour cela il est nécessaire d'améliorer nos connaissances du métabolisme central.

L'objectif principal de ce projet de thèse était de caractériser le métabolisme central de *E. limosum* B2 cultivé en chémostat sur méthanol ou sur glucose comme source de carbone, par une approche de biologie des systèmes.

Suivant un processus d'adaptation évolutive en laboratoire, la souche sauvage a été adaptée à un milieu méthanol défini sans extrait de levure. Le taux de croissance maximal chez le clone isolé a été amélioré d'un facteur 2,72. De plus, nous sommes parvenus à substituer la cystéine du milieu défini, décrite comme indispensable, par du thiosulfate de sodium. La souche sauvage s'est montrée incapable de croître dans ces conditions, suggérant l'apparition de mutations favorable chez le mutant. Le génome de la souche sauvage a d'abord été séquencé *de novo*, suivi des génomes de la population sur plusieurs générations et de trois clones isolés pour étudier la dynamique d'évolution génétique. Aucune mutation pouvant expliquer ce changement de phénotype n'a a été détectée dans le métabolisme central. Une étude protéomique a été réalisée sur la souche sauvage et adaptée afin d'éclaircir les mécanismes d'adaptations. Parmi les protéines surproduites chez la souche adaptée, deux dans la WLP, deux dans la gluconéogenèse et un complexe lié à l'assimilation du soufre ont été identifiés. De plus, l'exploration du méthylome a révélé un évènement de recombinaison homologue au niveau du système de restriction-modification de type I entre les deux souches. Cette modification pourrait jouer un rôle clé dans le mécanisme de régulation sur milieu minéral méthanol.

La souche adaptée a été cultivée en chémostat pour la caractérisation de son métabolisme central sur méthanol ou sur glucose comme seule source de carbone. Les paramètres physiologiques ont été déterminés et les flux d'entrée et de production ont été calculés. Ces valeurs ont été intégrées dans un modèle *in silico* à l'échelle du génome pour estimer les flux spécifiques de chaque enzyme du métabolisme primaire. De ces estimations, des modèles de conversation de l'énergie ont été élaborés pour les deux conditions étudiées. Le nombre absolu de copies de protéines par cellule a ensuite été déterminé puis ces données ont permis d'estimer la constante catalytique *in vivo* de chaque enzyme du métabolisme central. Ce paramètre nous a été utile pour comparer les activités des enzymes et identifier les plus limitantes.

Pour la partie transcriptomique, le développement d'une méthode d'extraction totale des ARNm a été un défi majeur. Nous avons élaboré une stratégie de quantification absolue des ARNm par cellule renforcée par l'ajout de deux mélanges de standards externes lors du procédé d'extraction des ARNm. Le premier a été ajouté après l'étape de lyse cellulaire et le second avant la ribodéplétion. Les résultats de séquençage par RNA-Seq ont été satisfaisants, avec une bonne corrélation entre le nombre de « reads » obtenus pour chaque standard et leur concentration théorique, indiquant une perte minimale en ARNm durant le processus d'extraction. La détermination de la quantité absolue de chaque transcrits est désormais possible.

En combinant les données de protéomique et de transcriptomique, obtenues à l'échelle absolue, la vitesse spécifique de traduction *in vivo* de chaque protéine a été déterminée. Nos travaux ont permis de mieux comprendre le métabolisme de *E. limosum* sur substrat méthanol et glucose. Nos données ont permis d'améliorer le modèle à l'échelle du génome de *E. limosum* et permettra de futures approches rationnelles d'ingénierie génétique pour améliorer les rendements de production de composés d'intérêts.

## Abstract

*Eubacterium limosum* is a strict anaerobic bacterium, belonging to the group of acetogens. Its interest lies in its ability to transform one-carbon (C1) feedstock into butyrate, a C4 molecule, through the oldest metabolic pathway on earth, called the Wood-Ljungdahl Pathway (WLP).

Nevertheless, yields and productivity must be improved to reach performances close to chemical processes. For this purpose, metabolic engineering strategies are needed but improving our knowledge of primary metabolism is of paramount importance to reach this goal.

The work performed in this thesis project aimed to characterize, by a systems biology approach, the central metabolism of *E. limosum* B2 growing in chemostat cultures on methanol or glucose as a carbon source.

Following an adaptive laboratory evolution process, the wild strain was cultivated on a synthetic methanol medium without yeast extract. The maximum growth rate for the evolved clone was improved by 2.72. In addition, we succeeded in substituting the cysteine of the synthetic medium, described as essential, with sodium thiosulfate. The wild-type strain was unable to grow under these conditions, suggesting positive mutations in the evolved clone. The complete genome of the wild strain was first sequenced, *de novo*, followed by the genomes of the population on several generations and the isolated clone to identify and study the dynamics of genetic evolution. No mutation that could explain this change in phenotype was detected in the genes encoding enzymes of the primary metabolism. A proteomic study was carried out on the wild type and the adapted strain to clarify the adaptation mechanisms. Among overexpressed proteins measured in the adapted strain, four proteins involved in central metabolism, including two in the WLP and two in gluconeogenesis, as well as a protein complex linked to sulfur assimilation were identified. In addition, exploration of the methylome revealed a homologous recombination event on the type I restriction-modification system between the two strains. This event by changing the methylome of the evolved clone could play a key role in the regulatory mechanism on synthetic methanol medium.

The adapted strain was grown in chemostat cultures for the characterization of its central metabolism on methanol or glucose as a carbon source. The physiological parameters were determined as well as all the input and output fluxes. These values were integrated into an *in silico* genome-scale model to estimate the specific fluxes of each enzyme of primary metabolism. From these estimations, energy conversation models were developed for the two conditions studied. The absolute number of protein copies per cell was then determined and these data allowed the estimation of the *in vivo* turnover rates

for each enzyme. This parameter was useful to compare enzyme activities and identify the most limiting ones.

For the transcriptomic part, the development of a total mRNA extraction method was a major challenge. We developed a strategy for the absolute quantification of mRNA per cell by adding two mixes of external standards during the extraction process. The first was added after the cell lysis step and the second before ribodepletion. RNA-Seq sequencing showed satisfactory results with a good correlation between the number of reads obtained for each standard and their theoretical concentration, indicating a minimal loss of mRNA during the extraction process. The determination of the absolute mRNA content for each gene is now possible.

By combining proteomics and transcriptomics data obtained at the absolute scale, each protein's specific in vivo translation rate was determined. Our work has provided a better understanding of the metabolism of *E. limosum* on methanol and glucose substrate. Our data have improved the genome-wide model of *E. limosum* and will enable future rational genetic engineering approaches to improve yields of compounds of interest.

# Remerciements

Je souhaite tout d'abord remercier Isabelle Meynial-Salles pour son accueil dans son équipe PEEP (Engineering and Evolution Pathways in Prokaryotes) au sein du Toulouse Biotechnology Institute.

Merci infiniment à Philippe Soucaille pour m'avoir donné la chance de travailler sur un sujet de thèse aussi intéressant que varié, pour son soutien et ses précieux conseils qu'il m'a apporté tout au long de ces travaux. Surtout, l'intégration de ce projet dans le programme européen BIOMETCHEM a été une occasion unique pour moi de partager mes avancées avec des scientifiques de haut niveau, Nigel Minton de l'université de Nottingham, Peter Dürre de l'université d'Ulm et Volker Muller de l'université Johann Wolfgang Goethe. Ce fût tellement enrichissant, merci encore.

Je tiens à remercier les membres du jury de ma thèse, Isabelle Meynial-Salles, Peter Dürre, Benjamin Woolston et Kaspar Valgepea pour avoir accepté d'évaluer mon travail. Leurs recommandations m'ont été très utiles pour améliorer mon manuscrit.

Mes remerciements vont également au Professeur Byung-Kwan Cho qui m'a accueilli au sein de son équipe dans le Korean Advanced Institute of Science and Technology (KAIST) en Corée du Sud à deux reprises. Je pense surtout à Yoseb Song, Sangrak Jin et Jiyun Bae qui ont tant donné pour que mon séjour se passe le mieux possible, je n'oublierai pas ces moments et tous ces délicieux restaurants coréens.

Merci beaucoup à David Boocock pour la quantité de travail qu'il a accompli sur la génération des données de protéomique et pour avoir pris le temps d'expliquer en détail les étapes analytiques. Aussi je tiens à remercier Nick Owens qui, au moment où je rédige ces lignes, nous aide dans la normalisation des données de transcriptomique. Les briques d'une superbe publication scientifique se mettent en place, il me tarde de voir le résultat final…

Merci à la plateforme de séquençage Gentyane à Clermont-Ferrand et à la plateforme de génomique Get Biopuces à TBI pour leurs prestations de grande qualité.

Je tiens à remercier Nigel Minton, Elizabeth Redfern, et James Millard pour leurs travaux sur la détermination de l'essentialité des gènes chez *E. limosum* qui nous ont été très utiles et pour avancer dans la compréhension du métabolisme. Un grand merci à Nicole Pearcy pour ses travaux sur

l’intégration des données physiologiques et protéomiques dans le modèle à l’échelle du génome. Cela nous a fait grandement avancer dans la construction de nos modèles de conservation de l’énergie et nous a permis d’émettre d’intéressantes théories sur le fonctionnement de certains éléments clés du métabolisme central du méthanol.

Mes pensées vont bien évidement à toute l’équipe PEEP, merci pour tous ces moments. Merci à Antoine et à Philippe Copillet pour ces discussions de bon matin qui m’auront bien fait marrer. Merci à Caro pour tes gâteaux qui n’ont cessés de monter en gamme ! Merci à Eglantine, à Paul pour tous les bons moments passés ensemble, les barbeuc party !

Merci à Céline pour m’avoir initié à l’art du chémostat, je n’ai pas oublié ! Ce fut le début d’un long périple...Que dire de l’inénarrable Quentin et de toutes ses histoires plus rocambolesques les unes que les autres. Je pense à ces innombrables pauses midi-vidéo avec Quentin, Paul, Eglantine et Julian. Aux post-docs Tom le fan d’escalade et Fernando qui s’est aussi lancé dans l’aventure *limosum* ! Merci à Gaoxin, toujours là pour aider, à Emilie ainsi que toute l’équipe pour ses conseils lors de la préparation de ma défense. Bon courage également au jeune Guigui qui se lance dans sa thèse, tous mes vœux de réussite !

Évidemment, bien entendu, un immense merci à Alexia pour tout ce qu’elle m’apporte, notamment l’oxygène qui se fait rare dans le monde anoxique des acétogènes, mais surtout son sourire et sa joie de vivre qui me rendent heureux.

Un immense merci à ma famille qui m’aura soutenu jusqu’au bout et qui m’aura toujours encouragé à aller plus loin. Sans oublier Pepito (et son poto Rico), à mes côtés depuis mes débuts à Toulouse, qui m’aura soutenu durant toutes mes phases de repos !

L’aventure continue !!!

# List of contents

# List of figures

# List of tables

# Abbreviations

**ACK:** Acetate kinase

***A. woodii*:** *Acetobacterium woodii*

**ACS:** Acetyl Coenzyme A Synthase

**ALE:** Adaptive laboratory evolution

**ATP:** Adenosine-5’ triphosphate

**ATPase:** Adenosine-5’ triphosphate synthase

**BCD:** Butyryl-CoAdehydrogenase

**Bp:** base pair

***C. autoethanogenum*:** *Clostridium autoethanogenum*

***C. ljungdahlii**: Clostridium ljungdahlii*

**CDM:** Cell dry mass

**CoA:** Coenzyme A

**CODH:** Carbon monoxide dehydrogenase

**CoP:** Corrinoid protein

**CRT:** Crotonase

**DNA:** Desoxyribonucleic Acid

**EDTA:** Ethylenediaminetetraacetic acid

***E. limosum***: *Eubacterium limosum*

***E. callanderi:*** *Eubacterium callanderi*

**ERCC:** External RNA Controls Consortium

**EtfA:** Electron transferring factor A

**EtfB:** Electron transferring factor B

**Fd:** ferrdoxin

**FDH:** Formate dehydrogenase

**FMN:** Flavin mononucleotide

**FTS:** Formyltetrahydrofolate synthetase

**GEM:** Genome-scale metabolic model

**$H_2$:** Dihydrogen

**Hbd:** 3-hydroxybutyryl-CoA dehydrogenase

**Kapp :** *in vivo* catalytic rates

**MeOH:** Methanol

**mM**: Millimolar

**mM-C**: Millimolar of carbon

**MTI:** Methyltransferase I

**MTII:** Methyltransferase II

**MTC:** Methylenetetrahydrofolate cyclohydrolase

**MTD:** Methyletetrahydrofolate dehydrogenase

**MTHFR:** Methylenetetrahydrofolate reductase

**$NAD^+$/NADH:** nicotinamide adenine dinucleotide (oxidized/reduced form)

**NGS:** Next-Generation Sequencing

**Nt**: Nucleotides (nt)

**PBS:** Phosphate Buffered Saline

**PCR:** polymerase chain reaction

**PFOR:** Pyruvate ferredoxin oxidoreductase

**PTA:** Phosphotransacetylase

**RNA:** Ribonucleic Acid

**rRNA:** Ribosomal RNA

**mRNA:** messenger RNA

**Rnf:** *Rhodobacter* nitrogen fixation

**SDS:** Sodium Dodecyl Sulfate

**SPEED:** Sample Preparation by Easy Extraction and Digestion

***S. ovata***: *Sporomusa ovata*

**SWATH-MS:** Sequential Window Acquisition of All Theoretical Mass Spectra

**MRM:** Multiple Reaction Monitoring

**TFA:** Trifluoroacetic acid

**THF:** Tetrahydrofolate

**THL:** Thiolase

**TRIS:** Tris(hydroxymethyl)aminomethane

**WT:** Wild Type

# Chapter 1:  Background

## 1.1. Developing bioeconomy, a major challenge

The extensive use of energetic fossil fuels for transportation, industry, and building heating led to a tremendous enhancement of greenhouse gas (GHG) emissions in the atmosphere these last decades, passing from 200-280 ppm of $CO_2$ during more than 400 000 years to 420 ppm in 2022 in less than 50 years[1,2]. The last report of the Intergovernmental Panel on Climate Change (IPCC) published in April 2022 mentioned unavoidable global warming of 1.5°C for the next two decades[3]. Faced with this alarming observation, a quick change in our ways to produce and consume energy are necessary. Also, the economy based on a take-make-waste behavior with a global carbon recycling rate of only 15% represents a huge loss[4]. For example, the value of wastes after landfilling or burning processes is decreasing by 95%, with a range value estimated from 80 to 120 billion US dollars of loss[5].

The situation pushed the scientific community to work on two fronts to find efficient solutions to i) improve the energy capture technology from renewable resources such as the sun with solar panels or the wind energy with wind turbines and ii) capture carbon emissions from unsustainable energy sources, such as carbon dioxide ($CO_2$) and carbon monoxide (CO), mainly responsible of the global warming, to create a virtuous circular carbon economy[6].

The conversion of synthesis gases (syngas), which includes CO, $CO_2$, and $H_2$, into value-added chemicals is possible using a chemical catalytic route like the Fischer-Tropsch process. This process consists of warming CO and $H_2$ at 150-300°C under 30 bar to convert gases into liquid hydrocarbons, but it requires high amounts of energy[7]. Another option is the utilization of biocatalysts, driven by microorganisms. Many studies are focused on the utilization of bacteria to efficiently fix $CO_2$ and CO with the production of molecules of valuable interest. Much attention was particularly paid to a group of prokaryotes, called acetogens, due to their unique ability to fix carbon from C1 feedstock, including synthesis gases (syngas) or liquid $C_1$ feedstock such as methanol. Using such organisms which can grow on carbon sources considered as pollutants represents a strong asset as these carbon sources are available in large quantities and are systemically produced during combustion of any organic components such as biomass from municipal wastes, gases produced by steel mills, or during coal production. Acetogens would then have the double benefit to recycle the carbon released from industrial infrastructures and producing valuable chemicals usually produced by synthetic chemistry which is efficient by highly polluting.

## 1.2. Acetogens

### 1.2.1. History of acetogens

The acetogens constitute a phylogenetic clade from the family of the *Bacteriaceae*. This group is represented by 23 different genera and more than 100 species have been identified so far. *Acetobacterium* and *Clostridium* harbor the most known acetogenic species[8]. This group has the particularity to use two molecules $CO_2$ as a carbon substrate and $H_2$ or CO as an electron donor to produce acetate, a $C_2$ compound, by using the reductive acetyl-CoA pathway, also called the Wood-Ljungdahl pathway (WLP). A large variety of substrates can be used by acetogens such as alcohols, hexoses, acids, and syngas. Depending on the specie and the substrate considered, acetate is not the only molecule produced by acetogens. The utilization of syngas as the substrate has been particularly studied due to the ability of acetogens to convert toxic gases into valuable products such as ethanol, 2,3-butanediol, butyric acid, and butanol[9,10]. Such versatility of substrate consumption allows the bacteria to grow heterotrophically and autotrophically.

The metabolic acetogenesis process, *i.e.* the reduction of two molecules of $CO_2$ with hydrogen in one molecule of acetate, has been first discovered in 1932 in a microbial population from sewage by F. Fisher and associates[11].

A few years after, in 1936, the microbiologist Klaas T. Wieringa isolated for the first time an acetogen, *Clostridium aceticum*, from the soil. The production of acetate from $CO_2$ and $H_2$ was defined by the following stoichiometry[12] :

$$2CO_2 + 4H_2 \rightarrow CH_3COOH + 2\,H_2O$$

It was the first carbon fixation metabolism discovered. Although the potential of acetogens was noticed, the strain was lost, probably due to World War II.

Then, the second acetogen to be isolated from horse manure was *C. thermoaceticum* in 1942 by Francis E. Fontaine and became the only strain studied for decades[13]. Reclassified as *Moorella thermoacetica*, the specie was observed to convert one molecule of glucose to almost 3 molecules of acetate following this reaction:

$$C_6H_{12}O_6 \rightarrow 3CH_3COOH$$

With *M. thermoacetica* as a model of acetogens, the link between heterotrophic and autotrophic acetogenesis processes has been highlighted[14]. The reduction of glucose to acetate results from two distinctive reactions with an oxidative and a reductive part:

Oxidation reaction: $C_6H_{12}O_6 + 2H_2O \rightarrow 2CH_3COOH + 2CO_2 + 8H$

Reduction reaction: $2CO_2 + 8H \rightarrow CH_2COOH + 2H_2O$

Net reaction: $C_6H_{12}O_6 \rightarrow 3CH_3COOH$

Fontaine emitted the hypothesis that $CO_2$ was further absorbed to allow the production of almost 3 moles of acetate per mole of glucose. The individual enzymatic steps underlying the reduction reaction of $CO_2$ into acetate were deciphered by Harland. G. Wood and Lars. G. Ljungdahl during the 80's[15,16]. The metabolic pathway was named the reductive acetyl-CoA pathway or the Wood-Ljungdahl Pathway (WLP).

Some acetogens such as *Acetobacterium woodii*, *Clostridium ljungdahlii*, or *M. thermoacetica* are extensively studied and contribute to bringing light to the metabolism of acetogens[17].

The elucidation of WLP marked a great step forward in the understanding of acetogens and their potential applications. The metabolic pathway will be detailed in the next part of the introduction.

### 1.2.2. Biotechnological applications

Acetogens represent good candidates to produce bulk chemicals sustainably. Therefore, the economic interest in acetogens rose this last decade[18,19]. Taking profit from their natural metabolic route of production or after genetic engineering, acetogens can produce acetate, ethanol, isopropanol, 3-hydroxybutyrate (3-HB), butyrate, butanol, and 2,3-BDO (Tab.1.1). This interesting versatility is a strong argument for the utilization of acetogens as value-added chemical producers to substitute petrochemistry.

**Table 1.1:** Non-exhaustive list showing the versatility of value-added chemicals produced by acetogens, whether native or non-native strains.

| Organism | Substrate | Product | Strain | Reference |
|---|---|---|---|---|
| ***A. woodii*** | Methanol | Acetate | Native | [20] |
| ***Eubacterium limosum*** | Methanol, syngas | Acetate, Butyrate | Native | [21] |
| | Methanol and formate | Butanol, Hexanoate | Native | [22] |
| ***Clostridium aceticum*** | CO | Ethanol | Native | [23] |
| ***C. ljungdahlii*** | Syngas | Isopropanol | Non-native | [24] |
| | Methanol and formate | Butanol, Hexanoate | Native | [22] |
| ***Clostridium autoethanogenum*** | Syngas | 2,3 BDO, acetate, ethanol | Native | [25] |

However, the replacement of petrochemistry with biotechnology which would be used in smalls industrials plants was considered for a long time as not competitive enough due to the high price of feedstock which represented more than 50% of production costs[26]. High prices were notably due to transportation costs. Nevertheless, thanks to the incredible progression of genetic engineering combined with the improvement of fermentation infrastructures to boost the production yields of microorganisms, the possibility of seriously using bacteria as an alternative to chemistry emerged. Besides, Lanzatech was the first company to industrialize a process using *C. autoethanogenum*, an acetogen, as a biological tool to produce ethanol from syngas emitted by a steel mill. Functional since 2018, the facility based in China allowed the complete recycling of syngas from the steel mill and the production of 60 000 tons of ethanol avoiding the emission of over 100 000 tons of $CO_2$[27].
Among the other applications considered for the acetogens, the production of biohydrogen from formate was mentioned with the discovery of a hydrogen-dependent $CO_2$ reductase (HDCR) enzyme complex, able to produce formate by the reduction of $CO_2$ and $H_2$ in *A. woodii*[28,29].

The acetogens showed a strong interest in the production of molecules of valuable interest, notably due to their specific metabolism. However, the exploitation of syngas as substrate by acetogens is difficult to set up due to the low water solubility of gases, limiting the mass transfer and therefore the

biomass concentration. Furthermore, the transportation costs of syngas are problematic and require under-pressure confinement to be profitable. In contrast, using methanol as substrate, a $C_1$ compound entirely miscible in the water and carrying a higher energy value than $C_1$ in the gaseous phase is promising. Among acetogens able to consume methanol, *E. limosum* takes a special part as it can produce complex molecules in addition to acetate.

The next part will be focused on a description of *E. limosum* B2, the acetogen specie studied in this thesis project and the specific metabolism of acetogens will be detailed.

## 1.3. The methylotrophic acetogen: *E. limosum*

*E. limosum* was first isolated in 1980 from sewage by Genthner et al., and was described as a non-pathogenic, rod-shaped, gram-positive, and strictly anaerobic bacterium[30]. This microorganism lives in the intestinal flora of numerous mammalians, including humans where *E. limosum* was detected in the intestinal tract and feces[31]. Versatile, this specie can grow from a large variety of substrates including $C_1$ feedstock with syngas, methanol, and formate but also hexoses, acids, alcohols, and methoxylated aromatic compounds[30–33]. *E. limosum* is therefore considered a methylotroph as it can consume methylated substrates like methanol or L-carnitine[34]. The main products are acids and are mainly constituted of acetate and butyrate. Other products such as pentanoic, hexanoic, and also butanol were observed to a lesser extent[22,35]. Furthermore, *E. limosum* was shown to biotransform isoflavonoids such as biochanin A, formononetic, and glycitein to the estrogenic metabolites genistein, daidzein, and 6,7,4'-trihydroxyisoflavone[33]. These molecules showed biological activities including anticancer, anti-inflammatory, and cardioprotective effects[36–38]. Still in this domain, evidence showed that *E. limosum* plays a role in human health, with positive effects on colitis, anti-inflammatory, and aging[39–41].

Above all, *E. limosum* receives particular attention for its ability to produce butyrate, a $C_4$ compound, from $C_1$ compounds such as methanol and syngas. This species is one of the few acetogens known to naturally produce butyrate from methanol among *C. carboxidovorans*, *C. drakei,* and *Oxobacter pfennigii*[42,43]. Moreover, *E. limosum* is one of the only known species able to naturally produce a $C_4$ compound, butyrate, from syngas[44].

Our attention will be focused on the $C_1$ feedstock methanol only, as the metabolism of *E. limosum* on this substrate harbors interesting properties.

## 1.4. Interest in methanol as $C_1$ feedstock

Methanol ($CH_3OH$) is one of the building blocks of the chemical industry and represents an interesting $C_1$ feedstock for several reasons. Widely available in the world, methanol constitutes a promising feedstock for future applications involving methylotrophs. Almost all methanol is produced from methane or syngas but it can also be produced from biomass by gasification[45]. Initially, methanol was made from wood as a co-product of charcoal production and was commonly named wood alcohol. Methanol can be manufactured from natural gas, refinery off-gas, coal, petroleum, and syngas. The process consists of the hydrogenation of $CO_2$ and CO by $H_2$ with the following reaction[45]:

$$2H_2 + CO \rightarrow CH_3OH$$

$$3H_2 + CO_2 \rightarrow CH_3OH + H_2O$$

This production process is relatively low cost with an estimation of around 500 €/ton[46]. As was previously mentioned, syngas can be easily obtained by the combustion of organic municipal waste as well as residues from the food and materials sectors. Furthermore, old energy resources, such as oil, fuel, or coal are limited and under the monopoly of oil-importing countries. The synthesis process to generate methanol could be adapted for any country.

Several full-scale installations are currently under construction using syngas with a production capacity of 1 million tons of methanol per year[47].

Furthermore, methanol constitutes an excellent fuel (octane number of 100) for internal combustion engines and is safer than gasoline as it's less inflammable. The price of methanol is similar to glucose and shows higher energy density with a degree of reduction per mole of carbon 50% greater than[48]. In 2006, the worldwide production capacity has been estimated to be 46 million tons, and 35 million tons are consumed each year[49]. In the bioindustry, single-cell protein production used as feedstock for animals by methylotrophic bacteria is considered. For example, high-cell density fermentation of *Methylobacterium extorquens* is possible in large-scale volumes and mineral medium, avoiding contamination while reducing prices compared to sugar-based fermentation with yeasts[50]. Easily produced in large quantities, methanol represents a C1 feedstock of interest to produce industrial valuable molecules. In contrast to syngas, using methanol shows an interesting advantage to avoid the gas-to-liquid mass transfer issue as methanol is entirely miscible in water. Furthermore, methanol carries a higher energy content than syngas and is easily and safely portable. Whether for methanogens or acetogens, the incorporation of methanol in the metabolism is allowed by a specific enzymatic complex called methyltransferase. The mechanism will be detailed in the metabolism part

of the introduction. For some acetogens able to consume methanol in acetogenesis only, such as *A. woodii* or *Sporomusa ovata*, the substrate is metabolized following this equation[50,51] :

$$4\ CH_3OH + 2\ CO_2 \rightleftharpoons 3\ CH_3COOH + 2H_2O$$

Acetate is mainly produced from the degradation of methanol. For *A. woodii*, the net gain of ATP per acetate is 0.83[20].

A part of the interest in *E. limosum* relies on its butyrogenesis ability during methylotrophy. The carbon chain elongation from $C_1$ bulk chemical is particularly interesting for the synthesis of multi-carbon molecules as the addition of one carbon to the chain length of the product enhances its value by a factor of 1.5–3, depending on the molecule[53]. Using a microorganism to build such molecules, from $C_1$ to $C_4$ compounds by biosynthesis is therefore attractive since each step chemical synthesis process is costly. Butyrate shows also several interesting properties that will be detailed in the next part.

## 1.5. The interest of butyrate in methylotrophic growth

The ability of *E. limosum* B2 to synthesize butyrate from methanol has been characterized in 1985 and recently for *E. callanderi*, previously *E. limosum KIST612* strain, a close genetic neighbor [54]. Butyrate or butyric acid is a short-chain fatty acid four-carbon molecule with wide applications in chemical, pharmaceutical industries, and animal feed[55].

Also, butyric acid is known for its anticancer properties by inducing biochemical and morphological differentiation in a variety of cells leading to the suppression of cancer cells[56]. In industry, butyric acid is mainly used for plastic production. The addition of the butyryl group to cellulose acetate allows better solubility in organic solvents, better flexibility, and enhances light and cold resistance[57]. Moreover, butyric acid can be used to produce esters such as ethyl butyrate and butyl butyrate. These products are involved in perfume, solvent, and fuel production[58].

Currently, the molecule is mainly chemically synthesized by the oxidation of butyraldehyde obtained from the oxosynthesis of propylene[59]. The global production was estimated at 80 000 metric tons per year for a mean price estimated at 1,8$/kg [58]. However, butyrate is also synthesized at an industrial scale by Metabolic Explorer, using recombinant strains of *C. acetobutylicum*[60].

Relatively low cost, butyric acid is especially of great interest for its metabolic pathway to further produce close $C_4$ valuable products such as butanol in a two-step enzymatic process for example. The WLP is used to reduce $C_1$ compounds in acetyl-CoA, which is used for butyrate synthesis. The

biosynthesis steps will be detailed in the next part. The biosynthetic pathway for butyric acid represents an interesting alternative for biofuel production. Regarding alternative energies, biobutanol constitutes a promising biofuel for replacing gasoil due to its thermodynamic properties. The combustion of butanol breaks the three C-C bonds of the molecule, releasing a high energy amount compared to the combustion of two molecules of ethanol with two C-C bonds[55].

Some native acetogens, like *E. limosum*[61], *Butyribacterium methylotrophicum*[62], or *C. carboxidovorans*[63] were reported to produce butanol on CO but in low amounts. Production of butanol at low concentrations was also shown for the *Eubacterium* genus after the metabolic engineering of *E. callanderi* KIST612[64] with glucose as substrate. Although progress was made on synthetic methylotrophs for methanol assimilation and product yields, native methylotrophs exhibit higher growth and product formation performances. For example, *E. limosum* showed a yield of 0.35 g cell/ g methanol against 0.198 g cell/ g methanol for the best metabolic-engineered strain of *E. coli*. Furthermore, the product yield of acetate and butyrate was demonstrated to be around 20 and 2 times higher respectively compared to *E. coli* in methanol condition[65]. Such results contributed to the extensive study of natural methylotrophs to improve metabolic engineering tools for native and synthetic microorganisms.

In this thesis project, the metabolism of *E. limosum* B2 strain was characterized by methanol and glucose mineral media. Numerous studies were performed on this strain at INSA Toulouse in the 80s. The wild-type (WT) strain of *E. limosum* B2 was isolated by Eric Samin in 1983. The physiologic characteristics of the strain on methanol-defined media were extensively studied in our laboratory. The present work proposes an in-depth approach to the characterization of the metabolism of this strain using recent omics technologies. The second part of the introduction will be focused on the metabolism of the acetogen *E. limosum* B2 in heterotrophic and methylotrophic conditions.

## 1.6. Primary metabolism of acetogens

### 1.6.1. Hexose fermentation in acetogens

The hexose fermentation for acetogens follows the same process as other prokaryotes. The glucose is metabolized through a succession of 9 metabolic reactions called glycolysis or Embden-Meyerhof-Parnas (EMP) pathway, giving two molecules of pyruvate (Fig.1.1).

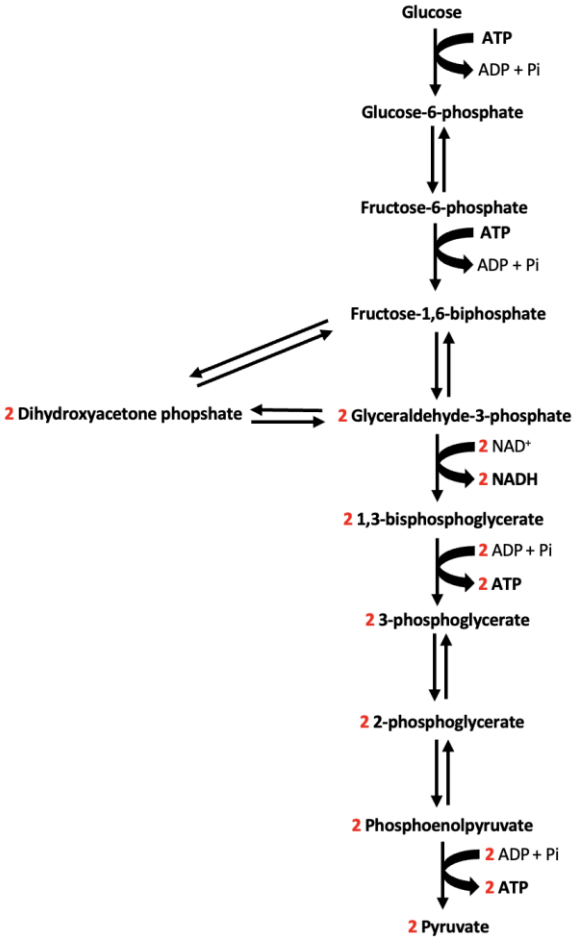



**Figure 1.1 :** Embden-Meyerhof-Parnas pathway (EMP) or glycolysis. Glucose is converted to pyruvate by a succession of nine reactions.

The oxidation process will generate two molecules of Adenosine TriPhosphate (ATP), the central compound to fuel metabolic reactions, and two reduced cofactors Nicotinamide Adenine Dinucleotide (NADH). The two molecules of pyruvate produced are then oxidized to acetyl-coenzyme A (acetyl-CoA) by the Pyruvate: Ferredoxin(Fd) oxidoreductase (PFOR) (Fig. 1.2). To achieve this reaction, ferredoxin is used as a cofactor. The ferredoxin is a low molecular weight protein containing an iron-sulfur cluster with a lower standard redox potential ($E_0'$ $Fd_{ox}/Fd^{2-}$ = -450mV) compared to the $NAD^+$/NAD couple ($E_0'$= -320mV). During the oxidation of pyruvate to acetyl-CoA, ferredoxin is usually used as an electron acceptor and carries two electrons ($Fd^{2-}$)[66]. Acetyl-CoA is then converted to acetate in a couple of reactions involving a phosphotransacetylase and an acetate kinase, generating one molecule of ATP per mole of acetyl-phosphate transformed.

The glucose oxidation is coupled with the production of 4 moles of ATP by substrate-level phosphorylation (SLP) following this net yield reaction:

$$C_6H_{12}O_6 + 2NAD^+ + 2\ Fd_{ox} + 4ADP + 4P_i \rightarrow 2CH_3COOH + 2CO_2 + 4ATP + 2H_2O + 2\ NADH + 2\ Fd^{2-}$$

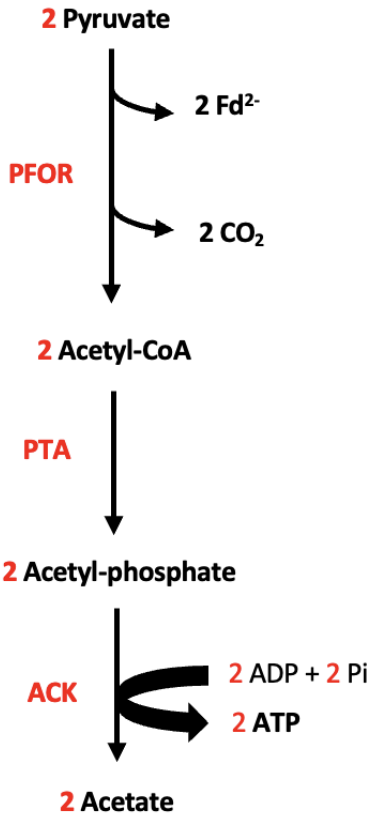



**Figure 1.2 :** Reduction of pyruvate to acetate during heterotrophic acidogenesis. PFOR, pyruvate ferredoxin oxidoreductase; PTA, phosphotransacetylase; ACK, acetate kinase; $Fd^{2-}$, reduced ferredoxin.

In summary, one molecule of glucose is oxidized in two molecules of pyruvate during glycolysis releasing energy and two moles of NADH. The pyruvate is then oxidized in acetyl-CoA, the central metabolic intermediary, by the PFOR, releasing two moles of $CO_2$ and two moles of $Fd^{2-}$. Acetyl-CoA is then converted into acetate via a two-step enzymatic reaction. Acetyl-CoA is also involved in numerous biomolecule production including fatty acids, amino acids, biomass, etc. In the case of acetogens, the two molecules of $CO_2$ produced during the oxidation of pyruvate to acetyl-CoA will

be recruited in the WLP to produce additional acetate. Of first importance, the WLP which is involved in carbon fixation will be extensively detailed in the next paragraphs.

### 1.6.2. The Wood-Ljungdahl pathway

As an acetogen, *E. limosum* can reduce two molecules of $CO_2$ to acetate via the reductive coenzyme A pathway, also known as the Wood-Ljungdahl Pathway. The WLP is the only known pathway allowing the $CO_2$ fixation into acetate coupled to a net synthesis of ATP produced by a complex system anchored in the cell membrane[67].

In the WLP, electrons are essential to driving the reductive process; they may derive from molecular hydrogen or carbon monoxide in autotrophic conditions. In heterotrophic growth, electrons are derived from organic acids, hexose, pentoses, alcohol, or methyl donor such as methanol[68].

The electron carriers, such as $Fd^{2-}$ or NADH can derive from the oxidation of hexose in heterotrophic acidogenesis or from $H_2$ oxidation during autotrophic acidogenesis.

The WLP is divided into two branches, the methyl branch and the carbonyl branch with one molecule of $CO_2$ at the onset of each branch (Fig. 1.3). The presented scheme is valid for *E. limosum*, as the conformation of enzymes such as FDH-Hydrogenase complex or the nature of electron carrier varies depending on the acetogen.

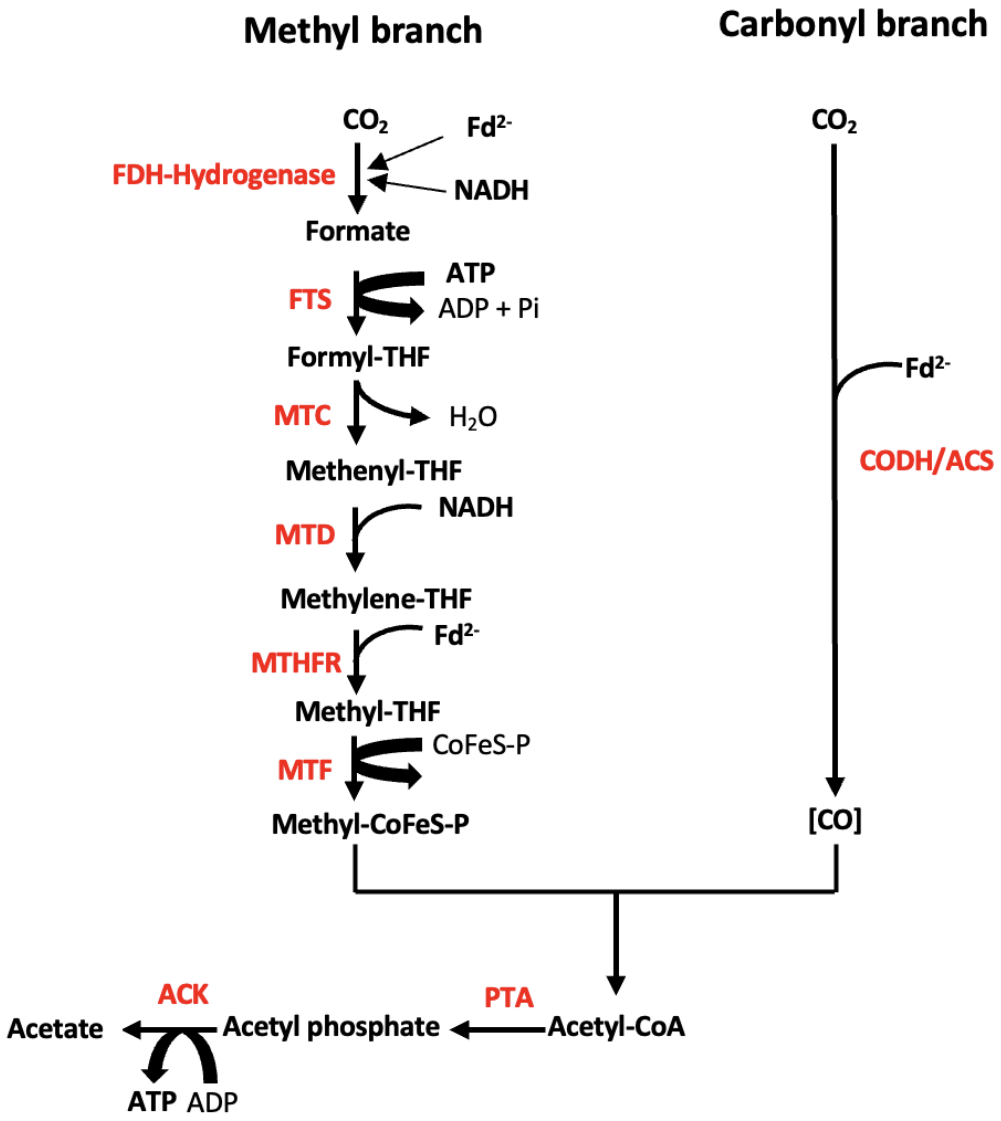



**Figure 1.3 :** Reductive acetyl-CoA pathway or Wood-Ljungdahl pathway (WLP) during autotrophy on $CO_2$. The scheme is under the assumption for *E. limosum* as the nature of electron carriers and enzymes (notably the FDH-hydrogenase complex, or the nature of electron carrier for MTHFR enzyme) can vary depending on the specie or still needs experimental validation. Two molecules of $CO_2$ are required to produce one molecule of acetate. FDH-Hydrogenase, formate dehydrogenase-hydrogenase complex; FTS, formyl-THF synthetase; MTC, methyl-THF cyclohydrolase; MTD, methyl-THF dehydrogenase; MTR, methylene-THF reductase; MTF, methyltransferase; CODH/ACS, carbon monoxide dehydrogenase/acetyl-CoA synthase; PTA, phosphotransacetylase; ACK, acetate kinase; $Fd^{2-}$, reduced ferredoxin.

#### *1.6.2.1. Methyl branch of WLP*

In the methyl branch, one molecule of CO2 is reduced to a methyl group bound to a corrinoid iron-sulfur protein (CoFeS), giving $CH_3$-CoFeSP by a succession of six metabolic reactions. Although the enzymes involved and the metabolites produced at each step are known, information concerning the redox carrier involved, for example, remains uncertain. In this part, the state art of each metabolic reaction will be described.

First, one molecule of $CO_2$ is reduced into formate by the action of the formate dehydrogenase (FDH)[69]. An electron carrier is required to allow the reaction, and its nature can be different depending on the specie. The use of NADH as an electron donor was demonstrated for *C. carboxidovorans*. Interestingly, a hydrogen-dependent carbon dioxide reductase (HDCR) for the direct interconversion of $CO_2$ into formate using $H_2$ was discovered for *A. woodii*[28]. Such a system could represent a natural and less energy-intensive way to store hydrogen in the form of formic acid (HCOOH) by the hydrogenation of $CO_2$. Recently, the isolation of FDH on *E. callanderi* KIST 612 showed a hydrogenase activity, suggesting functional connectivity with a hydrogenase[70]. Furthermore, the catalytic activity of $CO_2$ reduction with $H_2$ as the electron donor increased by 3-folds, and the $H_2$ production was increased by 2 with the addition of Fd and $NAD^+$, showing evidence that FDH and hydrogenase form a complex and work together. A theoretical structural model was proposed for the FDH-Bifurcating hydrogenase complex where the catalytic reaction reduction of $CO_2$ into formate by the FDH is enhanced by the hydrogenase that can bifurcate electron through iron cluster presents in the subunits (Fig. 1.4)[70]. The mechanism of electron bifurcation will be detailed in the next part.

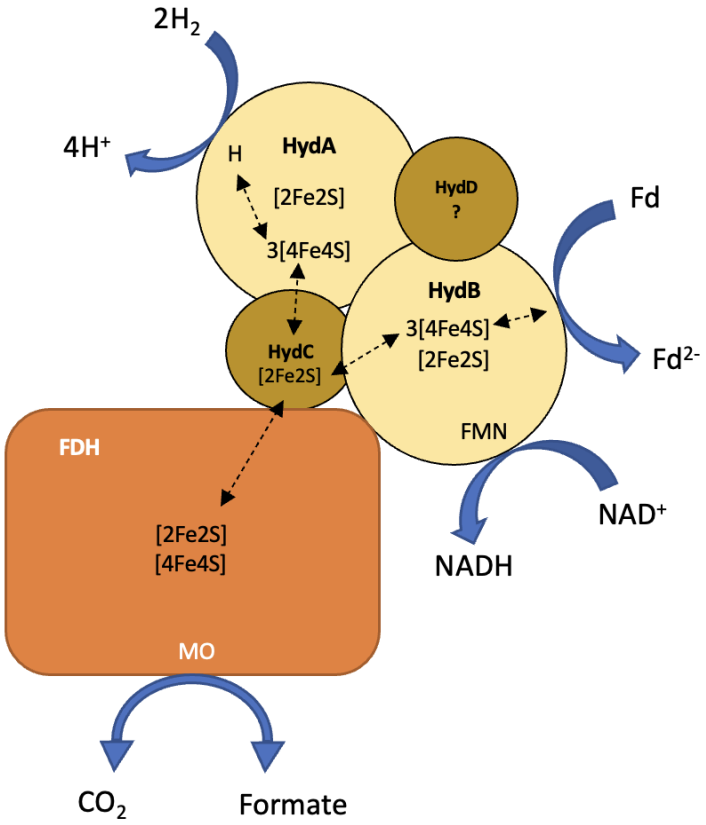



**Figure 1.4 :** Proposed structure of the formate dehydrogenase-hydrogenase complex (Fdh/HydABCD) of *E. callanderi* KIST612 that allows the catalytic reduction of $CO_2$ into formate and reversely with NADH and $Fd^2$ as electron carriers. The presence of HydD, a constitutive part of hydrogenase, was not detected in the active complex[70]. $Fd^{2-}$ reduced ferredoxin; Fd, oxidized ferredoxin; MO, molybdopterin cofactor; FMN, flavin mononucleotide.

The formyl group is bound to the $C_1$ carrier tetrahydrofolate (THF), giving formyl-THF by the action of the formyl-THF synthetase (FTS). This reaction requires the consumption of one molecule of ATP [16].

The formyl-THF is then converted into methenyl-THF by a dehydration reaction, *i.e.* a loss of water molecule ($H_2O$) catalyzed by the formyl-THF cyclohydrolase (MTC). The methenyl-THF is then reduced to methylene-THF by the methylene-THF dehydrogenase (MTD). This reaction necessitates an electron, given by the electron donor NADH. Next, the methylene-THF is reduced to methyl-THF via the methylene-THF reductase (MTHFR). The enzyme catalyzing this last reaction can differ among species. It exists 4 different types of MTHFR showing different organizations (Fig. 1.5).

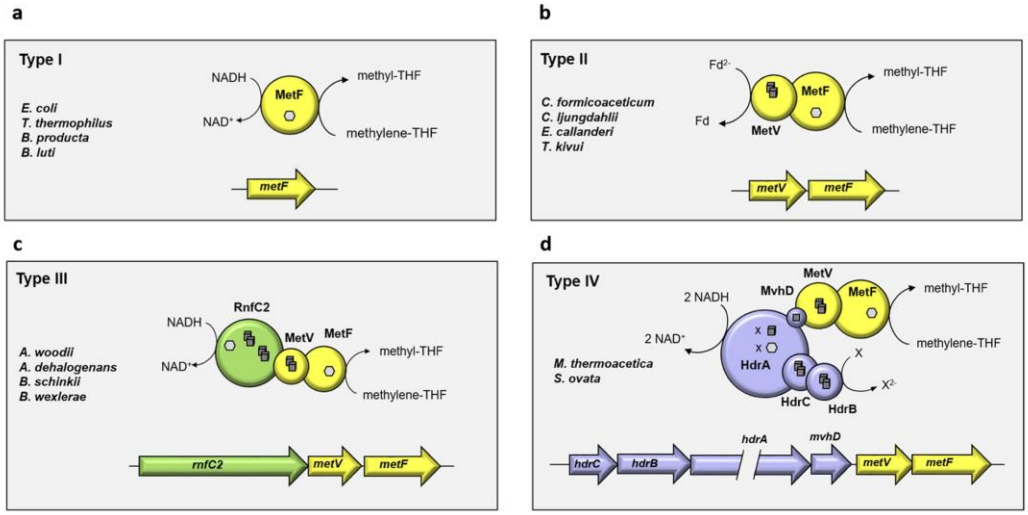



**Figure 1.5 :** The four different types of MTHFR[71]. For the MTHFR of type I, only one MetF subunit catalyzes the reduction of methylene-THF with NADH as the electron donor. The MTHFR type II forms a MetFV complex. The reduction of methylene-THF is catalyzed with $Fd^{2-}$ as the electron donor. Type III forms a three subunits complex, RnfC2MetFV, using NADH as the electron donor. The type IV MFTHR forms a complex between MetFV subunits and the heterodisulfide (Hdr) like subunits HdrCBA, linked by a methyl viologen subunit MvhD. Flavins are represented by hexagons, cubes are for [4Fe4S] clusters, and squares are the [2Fe2S] clusters.

Concerning *E. limosum*, the MTHFR was demonstrated to belong to type II after purification and enzymatic assays for *E. callanderi* KIST612, a very close genomic neighbor specie[70]. The ferredoxin was identified as the electron donor rather than NADH to catalyze the reduction of methylene-THF in methyl-THF. Theoretic energy conservation models were modelized in acetogenesis and butyrogenesis for methanol and CO conditions where MTHFR was assumed to bifurcate electrons or not. If the models were possible in both conditions previously mentioned, the model did not fit with the $CO_2 + H_2$ condition, showing a negative net gain of ATP during acetogenesis. Therefore, the use of ferredoxin as an electron donor is uncertain despite experimental evidence obtained *in vitro*, so the consideration of a type I MTHFR NAD dependent was privileged. Also, an MTHFR of type III for *E. callanderi* KIST612 was excluded since no *rnfC2* gene was found next to the *metFV* genes.
Nevertheless, a theory recently emerged to adapt the results obtain *in vitro* to the theoretic models. The RnfC2 from the MTHFR type III complex harbors a MetV domain N-terminal to the RnfC domain and was supposed to bind the MTHFR complex to the Rnf complex by docking the RnfC2 subunit to the Rnf complex[67]. For the MTHFR type II complex without the RnfC2 subunit, an electrostatic interaction would be possible between the RnfC subunit from the Rnf complex and the

MetV subunit. Therefore, the NADH would be used as a reductant in $CO_2$ + $H_2$ condition while $Fd^{2-}$ would be used as a reductant in other conditions for *E. limosum*. This hypothesis needs to be checked but would allow viable growth during methylotrophy or heterotrophy.

The final enzymatic step of the methyl branch consists of the transfer of the methyl group carried by the methyl-THF to a corrinoid iron-sulfur protein (CoFeSP) by a methyltransferase subunit of the CO dehydrogenase/acetyl-CoA synthase (CODH/ACS)[16]. The methyl group is transferred to the cobalt site in the cobalamin cofactor bound to the CoFeSP, forming an organometallic methyl-Co(III) intermediate (methyl-CoFeSP). The intermediary metabolite is the final molecule synthesized in the methyl branch before its utilization to synthesize acetyl-CoA.

#### *1.6.2.2. The carbonyl branch of WLP*

In this branch, the second molecule of $CO_2$ is reduced to CO by the CODH/ACS complex. Showing a low redox potential ($E_0'$ $CO_2$/CO = -520mV), the reduction reaction of $CO_2$ to CO can only be achieved with the reduced form of the ferredoxin ($Fd^{2-}$) as electron donor [17]. The metabolic oxidation of CO by CODH follows this reaction:

$$CO + H_2O \leftrightarrow CO_2 + 2\ H^+ + 2e^-$$

In acetogens, the enzymatic complex CODH/ACS is the key enzyme for the CO/$CO_2$ interconversion and the synthesis of Acetyl-CoA [72]. The CODH/ACS forms an $(\alpha\beta)_2$ complex with CODH and ACS constituted by two subunits respectively. The complex contains a total of seven metallic clusters with two [4Fe-4S] Ni bridged A-cluster only present in the two ACS subunits and the two CODH subunits contain two [4Fe-4S] B-clusters, two [Ni-3Fe-4S] C-cluster (C-cluster) and one [4Fe-4S] D-cluster (fig. 1.6)[73].

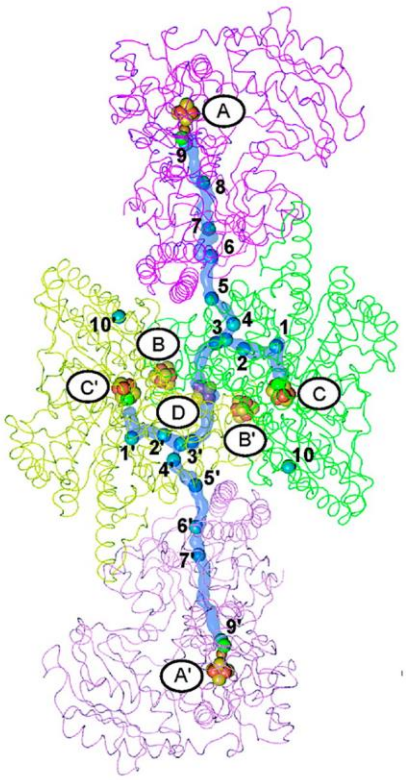



**Figure 1.6 :** Crystal structure of the CODH/ACS of *Moorella thermoacetica* showing the four metallic clusters A, B, C, and D[73]. The CODH shows homodimeric forms (green parts) as well as ACS (purple parts). The blue tunnel represents the hydrophobic CO tunnel through the enzymatic complex. Blue balls represent Xenon atoms (Xe) used to determine the structure of the tunnel.

The C-cluster is the active site of CODH and catalyzes the reversible oxidation of CO to $CO_2$. The A-cluster is the active site of ACS and catalyzes the synthesis of acetyl-CoA from CO, CoA, and the methyl group from the methyl branch of WLP.

The coenzyme B12 (5'-deoxyadenosylcobalamin) is used to transfer the methyl group from methyl-CoFeSP, the last metabolic intermediary of the methyl branch of WLP, to CODH [74]. The methyltransferase catalyzes the reaction while the corrinoid protein accepts the methyl group. The coenzyme A (CoA) binds to the CODH/ACS complex and then, the enzyme catalyzes the synthesis of acetyl-CoA with the addition of the carbonyl and methyl groups from both branches of the WLP. Acetyl-CoA represents the final product of the WLP and is further used to synthesize a large variety of products including acetate and butyrate, the two main products produced by *E. limosum* in methanol condition[21]. The biosynthesis of these two products will be described in the next paragraphs.

### 1.6.3. Acetate production from acetyl-CoA

Acetate biosynthesis results from the conversion of acetyl-CoA by two enzymatic steps. First, acetyl-CoA is phosphorylated by a phosphotransacetylase, giving acetyl-phosphate. The acetyl-phosphate is then converted to acetate by the enzymatic action of an acetate kinase (AK) allowing the synthesis of one molecule of ATP by substrate-level phosphorylation (SLP) per molecule of acetate generated. For one molecule of acetate produced from 2 molecules of $CO_2$, one molecule of ATP is consumed for the synthesis of formyl-THF by the formyl-THF synthetase while one molecule of ATP along the synthesis of acetate from acetyl-phosphate. In terms of energy, the net ATP yield is zero in the WLP. Acetogens do not use this pathway for energy production. To sustain growth, acetogens used an additional mechanism to produce energy, using a chemiosmotic ion gradient through intra/extracellular medium to fuel membrane-bound complexes. The function of these complexes will be detailed later in this chapter.

As it is mentioned earlier, *E. limosum* is one of the few acetogens able to produce butyrate from $C_1$ feedstock such as methanol. The biosynthesis of butyrate starts from two molecules of acetyl-CoA.

### 1.6.4. Butyrate production from acetyl-CoA

The synthesis of butyric acid by *E. limosum* depends on the carbon source, showing different production yields. Butyric acid was shown to be produced from methanol, carbon monoxide, carbon dioxide + dihydrogen, lactate, glucose, and dihydroxyacetone [75].

Two molecules of acetyl-CoA are required to produce one molecule of butyrate (Fig. 1.7)

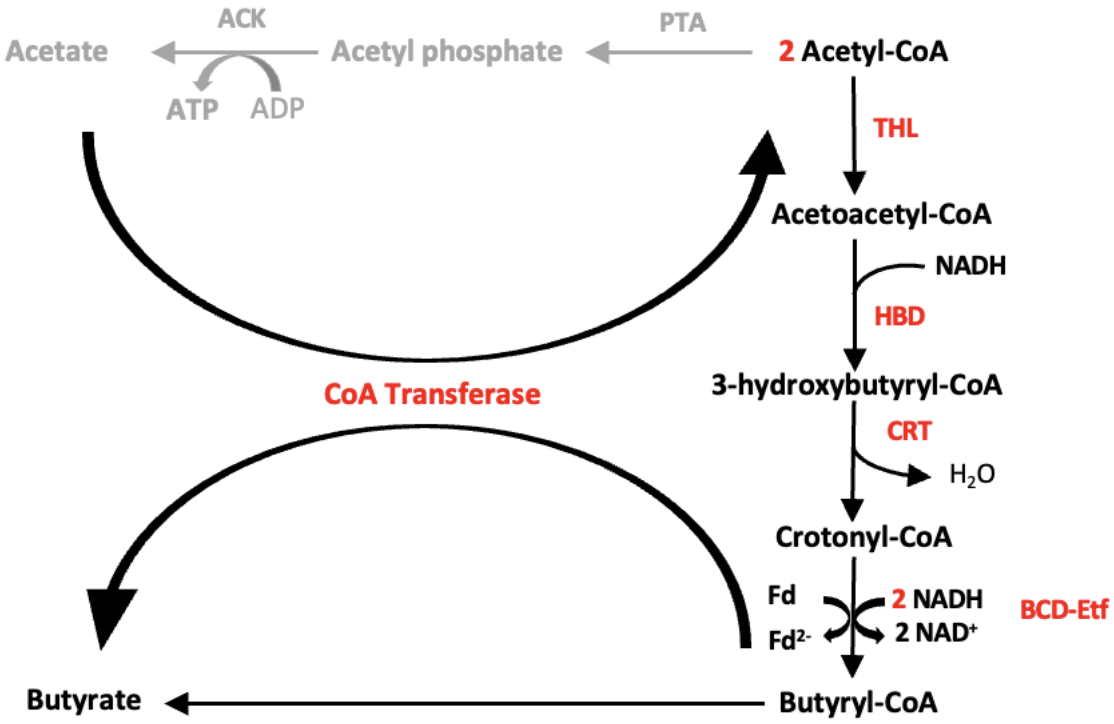



**Figure 1.7 :** Butyrate biosynthesis pathway from acetyl-CoA. Two molecules of acetyl-CoA are required to produce one molecule of butyrate. Since BUK enzyme has not been detected in *E. limosum* genome, production of butyrate by a CoA-transferase is assumed. THL, thiolase; HBD, 3-hydroxybutyryl CoA dehydrogenase; CRT, crotonase; BCD, butyryl CoA dehydrogenase; PTF, phosphotransbutyrylase; BUK, butyrate kinase. Metabolic pathway in grey color is not include in the global stoichiometry of butyrate production.

First, the two molecules of acetyl-CoA are converted to acetoacetyl-CoA by the acetyl-CoA acetyltransferase, also called thiolase (THL). The acetoacetyl-CoA is then hydroxylated into 3-hydroxybutyryl-CoA by the 3-hydroxybutyryl CoA dehydrogenase (HBD) using NADH as a cofactor. Then, the crotronase (CRT) dehydrates the 3-hydroxybutyl-CoA into crotonyl-CoA, releasing one molecule of water. The production of butyryl-CoA from crotonyl-CoA is catalyzed by the enzymatic complex BCD-Etf (Butyryl CoA dehydrogenase-Electron transfer flavoprotein) using 2 molecules of NADH as the electron donor to reduce simultaneously the crotonyl-CoA and one molecule of Fd via a particular mechanism called electron bifurcation. This mechanism, first discovered in *C. kluyveri* was also experimentally confirmed in *E. limosum* KIST612[76,77].

From the butyryl-CoA, butyrate or butanol can be naturally synthesized, depending on the acetogen. The synthesis of butyrate is realized by two reactions. First, the phosphorylation of butyryl-CoA by the phosphotransbutyrylase (PTB) produces butyryl-phosphate. Secondly, the butyryl-phosphate is dephosphorylated by a butyrate kinase (BK), producing butyrate and releasing a molecule of ATP.

The production of butyrate allows the regeneration of reducing equivalent by consumption of electrons and the production of energy by the formation of ATP. The enzymes phosphotransbutyrylase (PTB) and butyrate kinase (BUK) play a key role in the butyric acid biosynthesis in saccharolytic Clostridia like *Clostridium acetobutylicum, Clostridium perbutylacetonicum, Clostridium beijerinckii* but not in *Clostridium tyrobutyricum*[78]. For the acetogens, a gene coding for the butyrate kinase has been identified in *Eubacterium ruminantium.* Among 38 butyrate-producing bacteria species extracted from the human colon, four strains were found to possess detectable butyrate kinase activity, including *E. ruminantium*[79]. Although *E. limosum* was not included in the last study, the gene coding for BUK in *E. limosum* SA11 was not detected, suggesting another mechanism to produce butyrate[32]. The other possibility for the cells to produce butyrate is via the action of a butyryl-coenzyme A(CoA): acetate CoA-transferase instead of the combination of PTB and BUK (Figure 1.7).

In *E. limosum*, the mechanism for butyric acid production remains unclear. In the strain *E. limosum* ATCC 8486, the genes coding for the butyrate kinase has not been detected. To allow the biosynthesis of butyric acid, the cell probably uses a butyryl-CoA/acetate CoA transferase where the gene was identified in *E. limosum* SA11[32]. This enzyme catalyzes the conversion of butyryl-CoA to butyrate by transferring CoA residue to acetate. This transfer results in a new molecule of acetyl-CoA, which is reused to produce more acetate or butyrate, depending on the carbon source. How the cell produces butyrate is a question of great importance since the enzymatic dephosphorylation of butyryl phosphate leads to the production of one molecule of ATP. If the cell uses a CoA transferase, no ATP is directly produced during butyrate biosynthesis. On the other hand, one ATP is produced during acetate biosynthesis that is back converted to acetyl-CoA, giving the two pathways energetically equivalent. However, this theoretic enzymatic mechanism led by CoA transferase needs to be clarified by the creation of mutants and by enzymatic activity analysis. The action of the CoA-transferase for butyrate production will be assumed in the thesis project.

Next, the central metabolism of glucose consumption will be detailed as this substrate was chosen as the reference condition against methanol and lactate conditions.

## 1.7. Complete acetogenesis from glucose using heterotrophic oxidation and WLP

The homoacetogenesis in acetogens is divided into two distinctive parts. First is the oxidation of glucose in two molecules of acetate by glycolysis, PFOR, PTA, and ACK. During this process, two molecules of $CO_2$ are produced and reassimilated via the WLP to produce one additional molecule of acetate and to regenerate reducing equivalents via the oxidation of NADH into $NAD^+$ and $Fd^{2-}$ to Fd [80] (Fig. 1.8).

Theoretically, the simultaneous utilization of EMP and WLP pathways allows the production of three molecules of acetate for one molecule of glucose giving the following reaction:

$$C_6H_{12}O_6 + 4ADP + 4P_i \rightarrow 3CH_3COOH + 4ATP + 4H_2O$$

The heterotrophic acetogenesis produces four molecules of ATP for one molecule of glucose and the net gain of ATP occurs only during the oxidation of glucose into two acetate molecules since the net gain of ATP is null with the WLP.

According to this model, the oxidation of glucose does not require external $CO_2$. Indeed, two molecules of $CO_2$ are produced from the decarboxylation of pyruvate to acetyl-CoA. However, glucose is often not oxidized optimally, or not at all, without supplemental $CO_2$. This phenomenon is mainly due to the recycling of electron carriers. An additional amount of exogenous $CO_2$ has been demonstrated to be mandatory during glucose acetogenesis [81]. Usually, additional carbonates such as $NaHCO_3$ or $KHCO_3$ are added to the medium to allow an optimal metabolization of the substrate. Carbonates are converted to $CO_2$ by the carbonic anhydrase with the following reaction:

$$HCO_3^- + H^+ \leftrightarrow CO_2 + H_2O$$

This zinc-containing enzyme is found in various organisms. Most bacteria and archaea that contain carbonic anhydrase can grow autotrophically and higher expression of the enzyme has been detected during $CO_2$ limitation [82].

During autotrophic growth with syngas as the substrate or with low energy substrate such as methanol, the WLP does not gain net ATP by SLP. Growth relies on the additional formation of ATP

via the transmembrane chemiosmotic mechanism. Thereby, WLP is coupled to this additional mechanism to produce energy, both are closely linked to ensure energy conservation.

However, butyrate production can constitute a way to reduce the excess of reducing equivalent NADH[83].

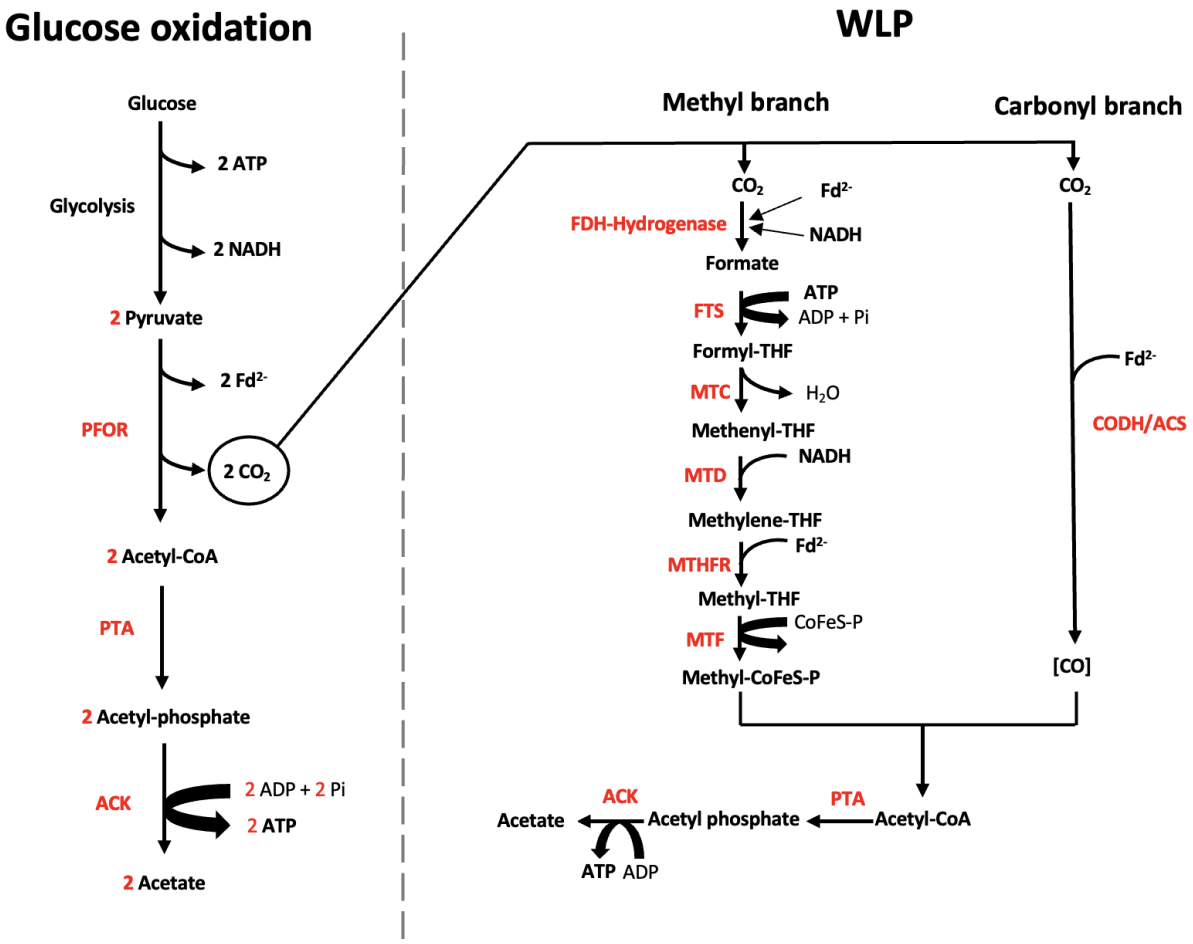



**Figure 1.8 :** Heterotrophic acidogenesis of glucose. Glucose is first oxidized to two molecules of acetate and two molecules of $CO_2$ by glycolysis, pyruvate ferredoxin oxidoreductase, phosphotransacetylase, and acetate kinase. The $CO_2$ produced is then fixed into acetate by the WLP. Also, WLP allows the regeneration of reducing equivalent by the oxidation of NADH to $NAD^+$.

During autotrophic growth with low-energy substrates such as syngas or methanol, the WLP does not generate a net gain of ATP by SLP. Furthermore, a close look at the model shows a deficit of reduced electron carriers, $Fd^{2-}$ and NADH, during the autotrophic condition. To ensure an electron balance and bring a supply of ATP that is mandatory to sustain growth, another system is required, a

transmembrane chemiosmotic mechanism. Thereby, WLP is coupled to this additional mechanism to produce energy, and both are closely linked to ensure energy conservation. The mechanisms underlying this trans-membranous system will be highlighted in the next part. Moreover, the theoretic energy model proposed for *Eubacterium sp* during methylotrophy will be detailed, including the incorporation of methanol in the WLP and the butyrate/acetate synthesis along the chemiosmotic system.

## 1.8. Energy conservation

Energy conservation is the mechanism used by the cells to convert external energy to ATP, the universal energy carrier. To date, two mechanisms of energy conservation are known: substrate level phosphorylation (SLP) and chemiosmotic ion gradient-driven phosphorylation.

The principle of SLP consists in coupling a metabolic reaction directly to the phosphorylation of ATP. In heterotrophic acidogenesis during the conversion of one mole of glucose to two moles of pyruvate, two moles of ATP are produced by SLP in the glycolytic pathway. Thermodynamically, the oxidation of glucose to pyruvate provides more free energy than the amount of energy needed for the phosphorylation of one molecule of ADP. In the case of acetogens, the autotrophic metabolism using WLP does not produce enough energy produced to sustain the growth of the cell. To sustain growth, another mode of energy conservation is required, the chemiosmotic ion gradient-driven phosphorylation. The principle is to take advantage of the energy liberated by an exergonic reaction to drive the translocation of cations across the cellular membrane resulting in a chemiosmotic ion gradient, also called protonmotive force[17]. The nature of cations differs among species, either protons $H^+$ or sodium ions $Na^+$. To date, it exists two different enzymatic complexes, the $NAD^+$:ferredoxin oxidoreductase (Rnf complex) and the Fd: $H^+$ oxidoreductase (Ech) complex[84,85]. Regardless of the ion nature, the chemiosmotic ion gradient from intracellular to extracellular medium is generated by a membrane-bound complex. Then, the importation of ions into the intracellular medium through the ATP synthase complex will generate ATP by phosphorylation of ADP to ATP[86]. Chemiosmosis is an efficient mode of energy conservation. It allows acetogens to live at the thermodynamic limits of life with low free energy substrates such as syngas or methanol. In this part, the mechanisms of energy conservation used in acetogens and particularly by *E. limosum* will be detailed.

### 1.8.1. Generation of ion gradient by membrane-bound complexes

Among the two different protonmotive force-generating membranous complex, the Rnf complex can drive either $Na^+$ or $H^+$ gradient and the Ech complex has been shown to drive only $H^+$ gradient[87]. Evidence of genes coding for the Rnf complex (ELIM_c3879-ELIM_c3884) has been shown in *E. limosum* ATCC 8486 with the release of the complete genome sequence[88]. Sequences appear similar to the electron transfer system of *A. woodii* where the Rnf complex was characterized[89]. Although, the $Na^+$ dependent activity of Rnf complex has been demonstrated by enzymatic activity assays in *E. limosum* KIST 612, in *E. limosum* ATCC 8486 it was only suggested from a bioinformatic analysis based on sequence homology[77,88]. The Rnf complex is a group of respiratory enzymes that catalyzes the oxidation of reduced ferredoxin for the reduction of $NAD^+$. The negative free energy released generates a transmembrane ion gradient used to produce energy in the form of ATP via the membrane-bound ATP synthase. For *C. ljungdahlii*, the Rnf complex was found to use protons to create a proton motive force[90].

#### *1.8.1.1. Structure of the Rnf complex*

Initially discovered in 1993 in *Rhodobacter capsulatus*, the *Rhodobacter* nitrogen fixation (Rnf) complex was associated with the nitrogen fixation ability of the strain[91]. A total of six subunits were identified, ABCDE and G, in gene operon organization *rnf*CDGEAB.
The complex contains iron-sulfur clusters and a flavin binding site, allowing an electron flux through the complex (Fig. 1.9).

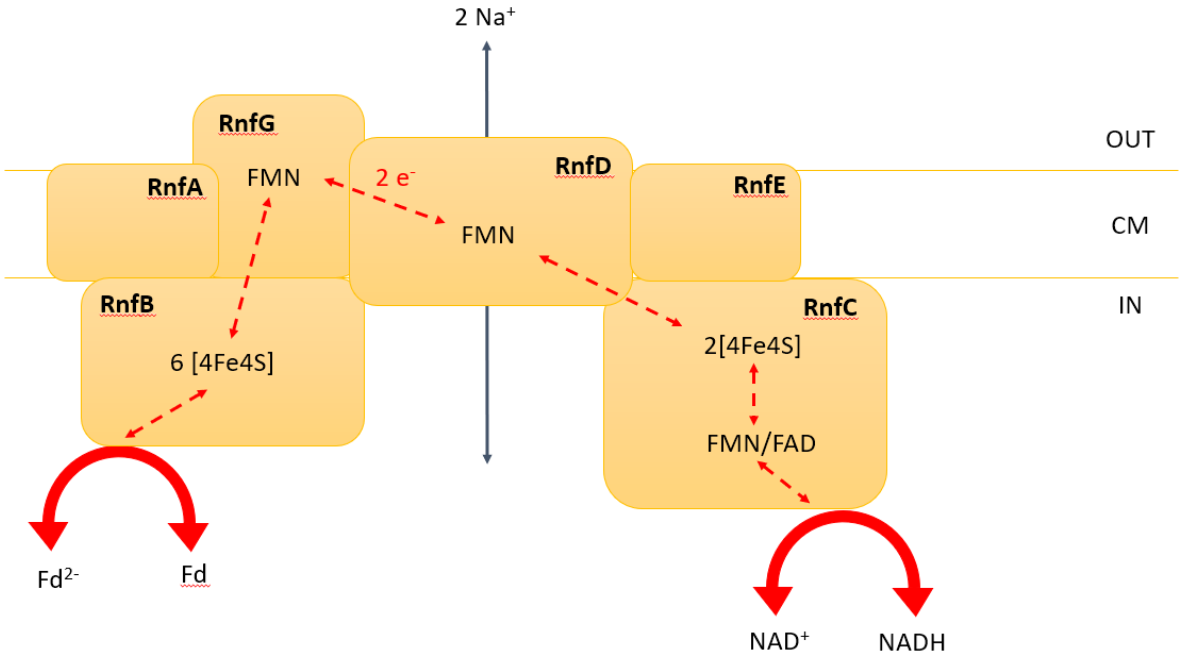



**Figure 1.9 :** Model of the Rnf complex of *A.woodii* and its coupling to the membrane potential[86].

The RnfC was identified as the $NAD^+$/NADH binding site with a molecular mass of 48.7 kDa in *A. woodii*[92]. The RnfC subunit contains two [4Fe-4S] clusters and one flavin binding site where the flavin cofactor was identified. The flavin mononucleotide (FMN) is essential for the oxidation of NADH, as will play the role of electron acceptor. Also, the flavin adenosine dinucleotide (FAD) plays a similar role to FMN. A doubt remained on the nature of the electron acceptor in the RnfC subunit. The FMN was recently strongly assumed in RnfC of *A.woodii*, as the FAD showed only 10% of the activity calculated in vitro for the reduction of $NAD^+$ compared with FMN as electron receptor[92]. The RnfD (35 kDa) has an FMN binding site, and the presence of the cofactor has been experimentally confirmed in *Vibrio cholerae*[93]. RnfG (22.8 kDa) is anchored to the membrane, and the presence of a conserved domain for covalently attached flavin mononucleotide has been proved[93,94]. RnfE (21.6 kDa) and RnfA are transmembrane subunits and did not show evidence of a cofactor binding site, suggesting that the electron course does not pass through these subunits. Finally, the RnfB is suggested to have six [4Fe-4S] clusters and was identified as the ferredoxin binding subunit. The genes coding for the Rnf ion-translocation complex has been identified in the genome of *E. limosum* ATCC 8486 and KIST612[77,88].

#### *1.8.1.2. Electron flow in the Rnf complex of E. limosum*

The Rnf complex catalyzes the reduction of $NAD^+$ with the oxidation of $Fd^{2-}$, exporting $Na^+$ to the extracellular medium. The soluble subunit RnfB is suggested to be the entry point for electrons derived from $Fd^{2-}$ oxidation while the RnfC is supposed to be the exit point of the electron with the reduction of $NAD^+$. However, the reverse mechanism is possible with the oxidation of NADH in the RnfC subunit along with the Fd reduction in the subunit RnfB. In this configuration, the electron flow is inverted, raising the level of intracellular NADH and importing ions into the intracellular compartment[86]. The Rnf complex plays a mandatory role in acetogens, a *rnf* mutant strain of *A. woodii* did not show growth on ethanol, lactate, and $CO_2$ + $H_2$ while it was able to grow on fructose[89]. Experimentations showed that reverse electron transport is essential for heterotrophic growth on substrates like methanol, ethanol, and acetate and during autotrophic growth on syngas[95]. The reverse electron flow is required when reduced $Fd^{2-}$ is insufficient to assume the reduction of $CO_2$ through the WLP. Nevertheless, the inversion of the chemiosmotic ion gradient has an energy cost and ATP hydrolysis by the ATP synthase is required to export the Na+ and create the proton motive force of ions. The mechanism of ATP synthase will be explained in the next paragraph.

### 1.8.2. Ech complex

In some acetogens, such as the acetogen model *M. thermoacetica* or *C. ljungdahlii*, protons $H^+$ are used as chemiosmotic coupling ions instead of $Na^+$ for *E. limosum* or *A. woodii*. Genes coding for an Ech-type ion translocating protein complex has not been detected in the genome of *E. limosum* ATCC 8486[61]. However, while *C. ljungdahlii* uses the Rnf complex for energy conservation with $H^+$ protons, and *M. thermoacetica* uses an energy-conserving hydrogenase (Ech) to drive the $H^+$ gradient. Ech is commonly found in anaerobic or facultative anaerobic bacteria and archaea. These hydrogenases couple $H^+$ export the exergonic electron transfer from reduced ferredoxin to $H^+$ producing $H_2$[96]. Then, hydrogen is used as a source of reducing equivalents. To date, Ech is known to drive only $H^+$ gradient proton *in vivo*, except for an Ech complex anchored in inverted membrane vesicles which were able to drive a $Na^+$ gradient in *Thermoanaerobacter kivui*[97]. All Ech hydrogenases share a core module of bounded membrane subunits, two [NiFe] hydrogenase subunits, and two additional hydrophilic subunits[17]. Also, the Ech complex has been well studied in the methanogenic model, *Methanosarcina barkeri*[98]. If the structure of the Ech complex can slightly differ between methanogenic and acetogenic microorganisms, the mechanism remains similar.

In addition to the large and small subunits, the Ech complex contains at least two additional hydrophilic subunits and two integral membrane subunits (Fig. 1.10). Ech complex is composed of six subunits, corresponding to the genes of the *echABCDEF* operon[98]. The subunit EchF is composed of two [4Fe-4S] clusters and one [4Fe-4S] cluster in EchC. The EchF subunit contains the characteristic binding motif for the [NiFe] site. Dihydrogen $H_2$ is activated at the [NiFe] center and electrons are transferred by the [4Fe-4S] cluster EchC subunit to the [4Fe-4S] cluster of the EchF subunit where ferredoxin is reduced. The [4Fe-4S] clusters mediate the electron transfer from reduced ferredoxin to the [NiFe] center. Electrons are then used to reduce a part of $H^+$ whereas others $H^+$ are guided to a transmembrane $H^+$ channel, creating a chemiosmotic gradient.

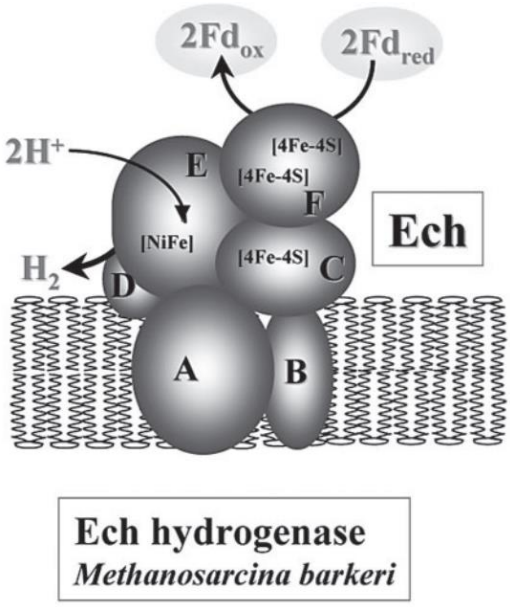



**Figure 1.10 :** Schematic representation of energy-converting hydrogenase in the methanogenic archaea *Methanosacina barkeri* [96].

With Ech of Rnf complex as ion-translocation proteins, the chemiosmotic gradient across the membrane is only permitted by the concomitant action of the ATP synthase along the production or the consumption of ATP.

### 1.8.3. ATP synthase

The role of ATP synthase is central in energy conservation. This membrane-bound complex is found in any living organism, from archaea to eukaryotes. It exists three families of ATP synthase, the $F_1F_0$ type present in bacteria, the $A_1A_0$ type in archaea, and the $V_1V_0$ type present in vacuolar membranes of eukaryotic cells[99]. The main function of these membrane-bound complexes is to import ions to produce ATP. The $F_1F_0$ synthase from *Escherichia coli* is the best-characterized ATP synthase. The hydrophilic anchored-membrane $F_1$, $A_1$, or $V_1$ part is in the cytoplasm. This domain catalyzes the synthesis of ATP from ADP and $P_i$ with an oxidative phosphorylation reaction. The structure of ATP synthase is an ingenious assembly of two motor domains linked by a central and a peripherical stalk (Fig. 1.11). The hydrophobic $F_0$, $A_0$, or $V_0$ domains are incorporated in the cytoplasmic membrane and are constituted by multiple c-ring subunits and a stator α subunit. The membrane-embedded domain conducts $H^+$ or $Na^+$ ions from the extracellular to the intracellular medium. This ion flux induces the rotation of the *c* ring against the stator. The energy produced by the rotation of the *c* ring is transferred via the central stalk to the cytoplasmic domain to drive ATP synthesis.

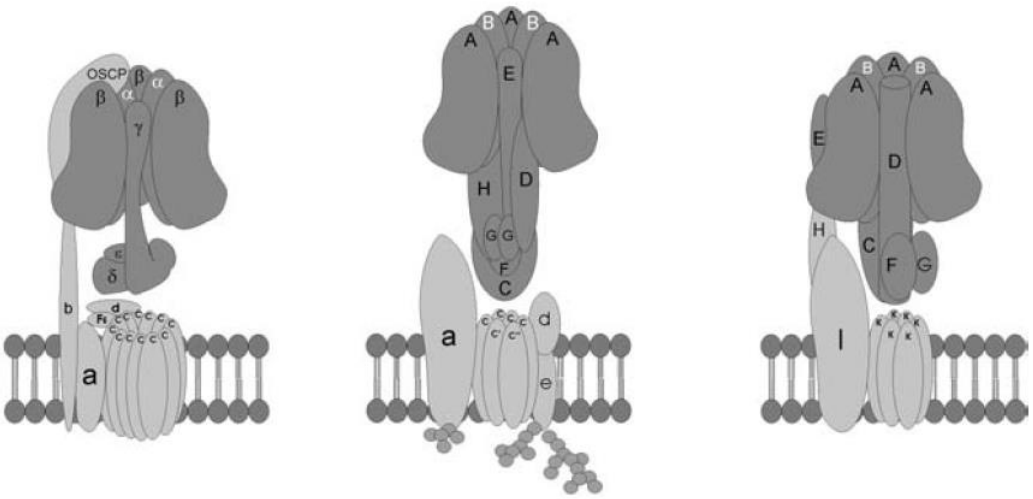



**Figure 1.11 :** Structure of $F_1F_0$ ATP synthase (left), $V_1V_0$ ATP synthase (middle), and $A_1A_0$ ATP synthase (right)[100].

As it was demonstrated for the Rnf complex, the ATPase of *E. callanderi* KIST612 showed an ATPase $Na^+$ dependent[77]. Surprisingly, the ATP synthase identified in *E. callanderi* KIST612 was an $A_1A_0$ with a V-type *c* subunit[77]. This type of ATP synthase was only known to be found in

hyperthermophilic archaea, making E. *callanderi* KIST612, the first mesophilic microorganism harboring an $A_1A_0$ with V-type *c* ATPase.

The ATP synthase allows the production of ATP with the chemiosmotic cation gradient concomitantly generated by the Rnf complex. Of course, the reverse use of ATPase for the exportation of cations from intracellular to extracellular medium is possible during growth on a non-glucosidic substrate to maintain the balance of reduced/oxidized forms between electron carriers. However, the ion exportation by the ATPase requires a supply of energy with the hydrolysis of ATP. The stoichiometry of $Na^+$ translocated per mole of ATP hydrolyzed or synthesized is unknown precisely for *Eubacterium sp* but is assumed to be four[70]. The redox balance between Fd and NADH is not possible with the WLP and the two membrane-anchored complexes Rnf and ATPase only. An additional mechanism is required, hydrogen-dependent, to reduce or oxidize whether $Fd/Fd^{2-}$ or $NADH/NAD^+$. The reduction of a molecule is sometimes thermodynamically unfavorable, such as ferredoxin due to its very low redox potential. Through evolution, a mechanism called electron bifurcation emerged to overcome this energetic barrier, and hydrogenase shows this ability to bifurcate electrons. The mechanism of electron bifurcation will be detailed in the next paragraph.

### 1.8.4. Electron bifurcation

Electrons are essential to driving all the metabolic reactions in the cell. The substrate can be the direct source of electrons, like hexose for instance. For acetogenic bacteria growing on syngas or other bacteria living in an anoxic environment, hydrogen is usually the electron source[101]. When the substrate is oxidized, electrons are transferred to reducing equivalents. It can be either Fd or $NAD^+$ in the case of *E. limosum*. In autotrophic conditions, with $CO_2 + H_2$, the only available electron donor is $H_2$ with a redox potential of -414 mV. The reduction of ferredoxin by $H_2$ as an electron donor is endergonic where the cell is facing a thermodynamic energetic barrier[17]. In the WLP, the reduction of $CO_2$ to CO in the carbonyl branch constitutes the largest energetic barrier, showing a redox potential of $E_0$'=-520mV[102]. Only reduced ferredoxin can provide electrons for the reduction process due to its low redox potential ($E_0$'= -500mV). In autotrophic or heterotrophic conditions with non-glucosidic substrates, the problem is then to reduce enough ferredoxin to maintain the reducing equivalent balance with NADH. However, reducing ferredoxin is highly endergonic. To overcome this energetic barrier and allow the reduction of oxidized ferredoxin to reduced ferredoxin, enzymes have evolved to form a multimeric complex and capture the free energy change of an exergonic

electron transfer reaction to drive an endergonic electron transfer reaction[17]. This mechanism is called electron bifurcation (Fig. 1.12).

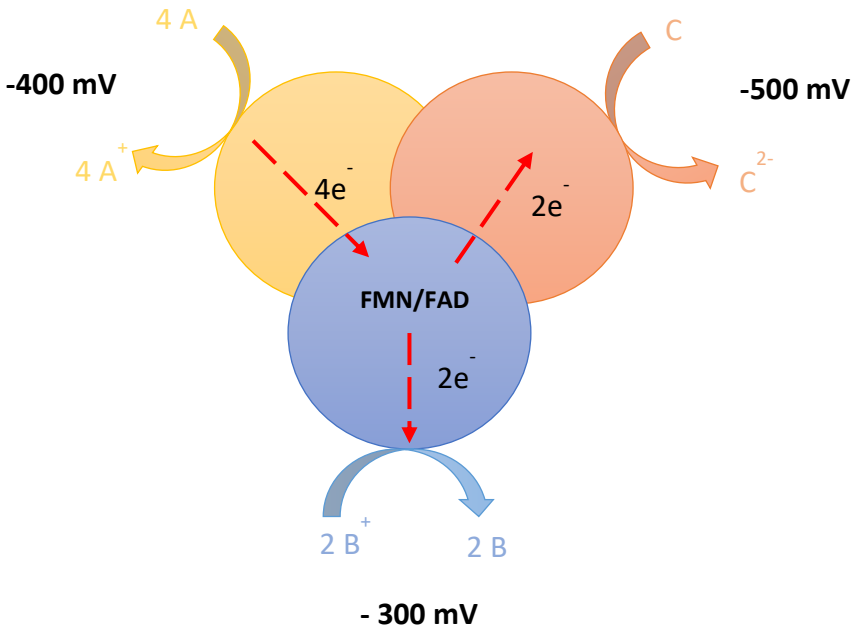



**Figure 1.12** : Electron bifurcation mechanism scheme. A is the electron donor and B,C the electron acceptors. The redox potential of C (-500mV) is lower than B (-300mV), and the redox potential of A (-400mV) is between B and C. The difference of redox potential allows the unfavorable thermodynamic reduction of C with A as electron donor. The electron transfer from A to a positive redox potential B (exergonic reaction) allows the electron transfer from A to more negative potential redox C (endergonic reaction). FMN, flavin mononucleotide; FAD, flavin adenine dinucleotide.

With this mechanism, the reduction of Fd by a higher redox potential electron donor such as $H_2$ is possible without energy loss.

#### 1.8.4.1. Discovery of the first electron bifurcation enzymatic complex

The electron bifurcation mechanism has been highlighted in 1975 by P. Mitchell with the identification of a quinone that drives an electron bifurcation in the complex III of the mitochondria [103]. A similar process is called flavin-based electron bifurcation (FBEB) in anaerobic metabolism and especially in acetogenic bacteria. The FBEB mechanism has been first discovered in *C. kluyveri* within an enzymatic complex containing an electron-transferring flavoprotein and a butyryl-CoA dehydrogenase (Etf/Bcd) [76]. The reduction of crotonyl-CoA to butyryl-CoA occurs naturally in the microorganisms able to produce butyrate or butanol, such as *C. acetobutyricum, C. kluyveri,* or *E. limosum*. This complex couples the exergonic reduction of crotonyl-CoA to butyryl-CoA with the endergonic reduction of ferredoxin with NADH. In this reaction, the electron acceptor with positive redox potential is the couple crotonyl-CoA/butyryl-CoA (E°'= -10 mV) and the electron acceptor with negative redox potential is ferredoxin with -405 mV. NADH is the electron donor with an intermediary redox potential (E°' = -320mV). The Bcd/Etf complex allows the reduction of ferredoxin and follows the reaction [104]:

$$\text{Crotonyl-CoA} + 2\ \text{NADH} + 2\ \text{Fd}_{ox} \rightarrow \text{Butyryl-CoA} + 2\ \text{NAD}^+ + 2\ \text{Fd}_{red}$$

The Etf is divided into two subunits (α and β), both contain one FAD (α-FAD and β-FAD). During the process, NADH reduces the β-FAD which bifurcates one electron to Bcd via α-FAD and the other electron to ferredoxin [104]. The genes coding for the complex Etf/Bcd has been detected in *E. callanderi* KIST 612[77]. Furthermore, the butyryl-CoA dehydrogenase has been purified and oxidation of NADH with a reduction of ferredoxin has been observed in presence of crotonyl-CoA. Also, the enzymatic activity was enhanced by the addition of FAD as a cofactor, confirming that the Etf/Bcd complex shared similar activity to the complex identified in *C. kluyveri*[77].

Soon after the discovery of the electron bifurcation mechanism, the reverse mechanism was discovered in a hydrogenase complex in the anaerobic bacterium *Thermotoga maritima*[105]. This species uses a trimeric [FeFe] hydrogenase to reduce protons $H^+$ into $H_2$ by the oxidation of NADH and $Fd^{2-}$ with the following reaction:

$$\text{NADH} + \text{Fd}^{2-} + 3\ \text{H}^+ \rightleftharpoons \text{NAD}^+ + \text{Fd} + 2\ \text{H}_2$$

This reverse mechanism allows the reduction of one specie, $H^+$ protons for instance, by two different electron donors, NADH, and $Fd^{2-}$. This convergence of electrons from two donors to one acceptor is

called electron confurcation. This mechanism is also known in aerobic chemolithotroph organisms growing on a substrate with higher redox potential than NAD+/NADH. In this case, the cell uses the electrochemical $H^+$ gradient ensured by the membrane-bound electron transport to reduce $NAD^+$ [106].
The FBEB mechanism is essential to fuel the Rnf complex to maintain the chemiosmotic $Na^+/H^+$ gradient and ATP synthesis through the ATP synthase[107]. The Bcd/Etf complex is not the only enzymatic complex able to bifurcate electrons and some enzymes can combine bifurcation and confurcation modes.
Studies have demonstrated the electron bifurcation/confurcation mode of the FBEB hydrogenase in *A. woodii*, with confurcation during growth on lactate and bifurcation during growth on $CO_2$ + $H_2$ [108,109]. In autotrophic conditions, with $CO_2$ + $H_2$, the cell oxidizes dihydrogen with an FBEB hydrogenase to reduce $NAD^+$ and Fd. The excess of $Fd^{2-}$ is oxidized by the Rnf complex which reduces $NAD^+$. The excess NADH is further used through the methyl branch of the WLP during the reduction of methenyl-THF and methylene-THF via the MTHFR. In lactate condition for *A. woodii*, the endergonic oxidation of the substrate is driven by the oxidation of reduced ferredoxin catalyzed by the electron-bifurcating Ldh/EtfAB complex. The reduced ferredoxin is supplied by the oxidation of pyruvate and by the Rnf complex while NADH is supplied by the oxidation of lactate to pyruvate [108]. However, to provide enough $H_2$ that is essential for the reduction of $CO_2$ to formate (instead for *E. limosum*), a hydrogenase is used to produce hydrogen by electron confurcation[108].
This ability to switch between bifurcation and confurcation to balance $NAD^+$/NADH and $Fd/Fd^{2-}$ levels depending on the substrate, shows the great metabolic adaptability of acetogens.

#### *1.8.4.2. FBEB Hydrogenases*

Hydrogenases are widespread in nature and can be found in various organisms. These enzymes catalyze the reversible oxidation of $H_2$ into two protons and two electrons.

$$H_2 \rightleftharpoons 2\ H^+ + 2e^-$$

This reversible reaction occurs in the organometallic active site of hydrogenases. The nature of metals depends on the class of hydrogenase considered. Based on their phylogeny, hydrogenases can be divided into three major classes based on their metal cluster content. The [FeFe] hydrogenases are found in bacteria and some eukaryotes. The second class is the [NiFe] hydrogenases which are found in bacteria and archaea. The last class is the [Fe] hydrogenases, only present in archaea [110–112]. Despite the differences among these classes, the metal clusters share similarities. The metal ions are ligated by inorganic CO and $CN^-$ ligands and are bridged by sulfur atoms. For the [NiFe] and [FeFe] classes,

$H_2$ molecules reach the organometallic active site through a hydrophobic gas channel and are transferred to a [4Fe4S] cluster adjacent to the metal center [113].

All [NiFe] hydrogenases are divided into two subunits, usually called large and small hydrogenase subunits. The large subunit contains the [NiFe] active site and the small subunit contains a conserved amino-acid pattern, which binds to the [4Fe-4S] cluster, called the proximal cluster [114]. Energy-converting [NiFe] hydrogenases identified in 2004 constitute a distinct phylogenetic group of the large family of hydrogenases [115].

The [FeFe] hydrogenases play an important role in energy conservation and carbon metabolism. These hydrogenases are involved in the electron transfer to soluble electron carriers, in the generation of the chemiosmotic ion gradient, and the energetic coupling mechanism of flavin-based electron bifurcation (FBEB).

In autotrophic acidogenesis, with $CO_2 + H_2$, electrons are derived from the oxidation of hydrogen. In *A. woodii*, these electrons are derived from a soluble tetrameric [FeFe] hydrogenase, HydABCD, which contains iron-sulfur centers and flavin. This enzymatic complex has been detected in *E. callanderi* KIST612 and appears to be similar with over 95% genes homology, to the electron bifurcating hydrogenase of *A.woodii* [77].

### 1.8.5. Structure of FBEB hydrogenase

The FBEB hydrogenase enzymatic complex is divided into four subunits: HydABCD in *A. woodii* and *E. callanderi*. The subunit composition is quite similar between acetogens for the subunits HydABCD (Fig. 1.13). In *A. woodii* or *E. callanderi* KIST612, Hyd A (~63kDa), HydB (~65kDa), and HydC (~17kDa) share equivalent size[116]. The presence of HydD (13.3 kDa) in the hydrogenase complex is still under the assumption.

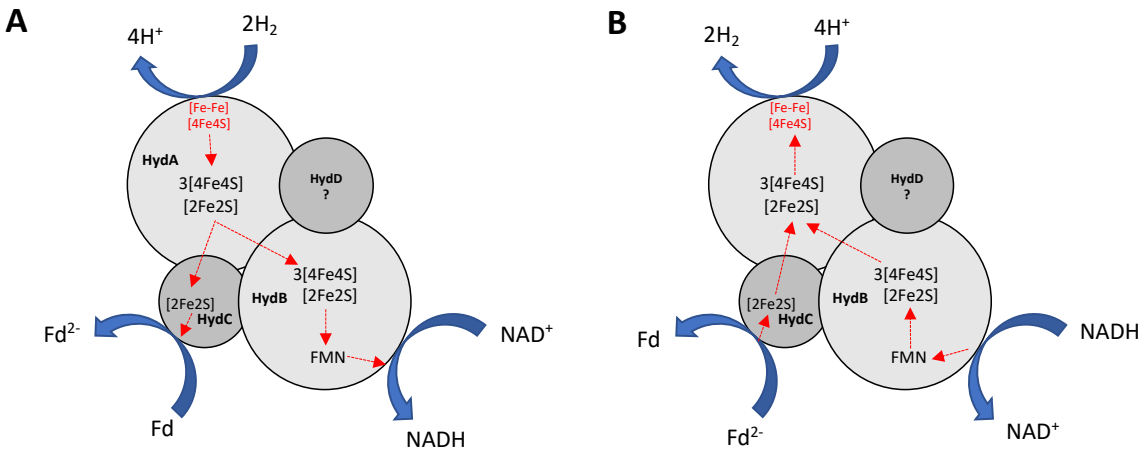



**Figure 1.13**: Structure of FBEB hydrogenase of *A. woodii* [107]. The structure is presumed to be similar in *E. callanderi* KIST612 and *E. limosum* B2. A. Electron bifurcation with $H_2$ as an electron donor. B. Electron confurcation with NADH and $Fd^{2-}$ as an electron donor.

The general structure of the core domain consists of the H-cluster located in HydA which is the active site of the hydrogenase where $H_2$ oxidation occurs (in red in fig. 1.13 A and B). HydB is predicted to contain a flavin binding site, one [2Fe2S], and three [4Fe4S] clusters. This subunit is assumed and the reduction of $NAD^+$ and Fd. HydC contains one [2Fe2S] cluster. HydD has no predicted cofactor and is unlikely to be involved in electron transport.

The FBEB hydrogenase is mandatory to preserve the intracellular homeostasis between ferredoxin and NAD. Combined, with glycolysis for heterotrophy on glucose, WLP, the FBEB hydrogenase, and the membranous complexes Rnf + ATPase, the strategy is perfectly optimized to save energy and sustain growth on the less energetic possible such as syngas. This is the possibility for these components to work as usual or reversely which gives such versatility to acetogens. The complete methanol metabolism will be now detailed from its incorporation into the WLP to the production of acetate and butyrate along with the energy conservation mechanisms involved.

## 1.9. Methanol metabolism

### 1.9.1. Methyltransferase

Acetogenic, methanogenic and other microorganisms using the WLP to convert methanol into biomass and products employs methyltransferases to transfer the methyl group in the WLP. The methanol is transformed to methyl-THF, which is incorporated into the WLP by a transfer of its methyl group ($CH_3$) to a THF molecule, giving methyl-THF a component of the carbonyl branch of the WLP. This transfer is permitted by an enzymatic complex called a methyltransferase system (MT). These are a class of enzymes that catalyze the transfer of a methyl group ($CH_3$) from a donor to an acceptor. MT is substrate specific and allows the assimilation of methyl groups from a large variety of molecules. The MT is composed of three subunits (fig. 1.14). The methyltransferase I (MTI), also known as MtaB, which is substrate specific, cleaves the C-O bound of the methyl group. Then, the methyl group is transferred to a central cobalt atom of a corrinoid protein (CoP). Finally, the methyl group is transferred to a TFH, forming methyl-THF.

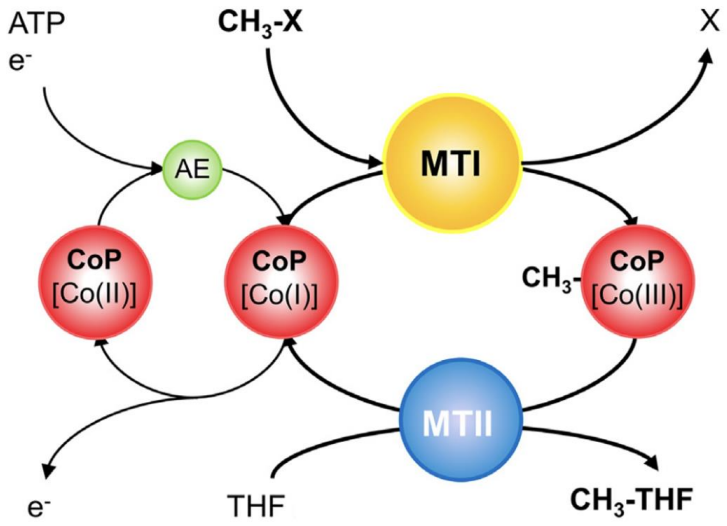



**Figure 1.14 :** Mechanism of methyltransferase system in acetogens. The substrate $CH_3$-X is demethylated by MTI and the methyl group is transferred to the super-reduced form of cob(I)amide, giving $CH_3$-CoP (Co(III)). The methyl group from $CH_3$-CoP is then transferred to THF, giving methyl-THF. The super-reduced state of cob(I)amide is immediately autoxidized in cob(II)amide. An ATP-dependent enzyme (AE) is required to restore the state cob(I)amide[117].

It exists a large variety of methyltransferases. In *A. woodii* for instance, around 30 different methyltransferases have been identified[67]. Recently, a total of 32 different substrate specificity subunits of methyltransferases were identified in the *E. callanderi* KIST612 genome, indicating the importance of the strain to capture a large variety of methyl groups as a carbon source for energy and growth sustainability[70]. However, the nature of the molecules carrying a methyl group targeted by the MT of *E. callanderi* remains unknown and most MT identified were annotated as trimethylamine methyltransferase. For a sake of simplicity, the MT will be represented below (Fig. 1.15)

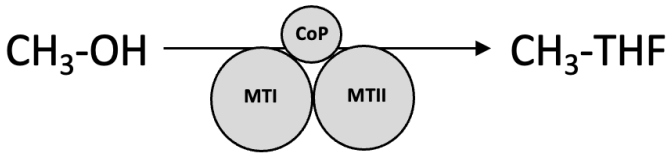



**Figure 1.15 :** Schematic representation of methanol transformation in methyl-THF by the MT complex. This representation will be used in this manuscript to represent the integration of methanol in the WLP.

Methyltransferases are organized in an operon and share a similar genetic organization among acetogens. The operon structure of methanol-specific MT of *A. woodii* is very close to the operon coding for methyltransferase of *E. limosum* KIST612 (ancient name of *E. callanderi* KIST612) (Fig. 1.16)[20].

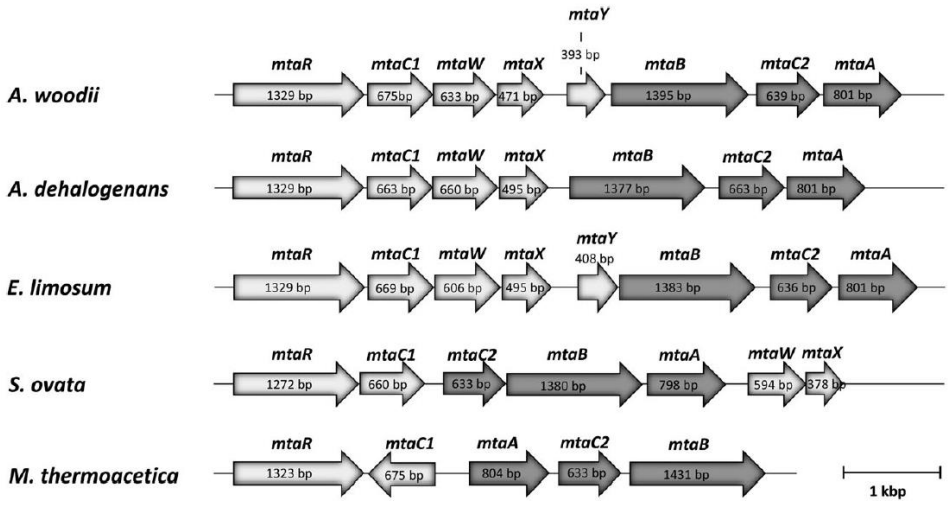



**Figure 1.16 :** Potential methanol-specific methyltransferase operons within acetogenic bacteria[20].

With the assimilation mechanism of methanol, all elements are in place to build an energy conservation model, including carbon and electron fluxes. The next paragraph will focus on the last models of energy conservation on methanol growth proposed for *E. callanderi* KIST 612, the closest genetic neighbor of *E. limosum*.

## 1.10. Energy conservation models of *E. callanderi* KIST612

The last study focused on the methanol metabolism of *E. callanderi* KIST 612 was published in 2021 by the group of Volker Muller at the Johann Wolfgang Goethe University- Frankfurt[70].
Interestingly, the energy conservation model suggests oxidation of a part of the methyl group carried by methyl-THF into $CO_2$ by a reverse use of the methyl branch of WLP to produce hydrogen and reduced NADH.
Such model suggestion on methanol assimilation by an acetogen was proposed for the first time for *A. woodii* after i) comparing the transcriptomic profile on methanol and fructose where the enzymes involved in the WLP such as FTS, MTC, MTD, and MTHFR were not significantly differentially expressed between both conditions nor showed difference enzymatic activity and ii) demonstrating the oxidation of methyl-THF in methylene-THF by the MTHFR and measuring a hydrogen production by the HDCR complex alongside the acetate production[20]. It was then proposed that the oxidation of one $CH_3$ into $CO_2$ yields six reducing equivalents (6 $e^-$) which would be further used to reduce 3 molecules of $CO_2$ into CO through the carbonyl branch. The NADH produced can then be used to produce $Fd^{2-}$ through the Rnf complex and hydrogenase can be solicited to reduce $NAD^+$ and Fd from $H_2$ to balance the intracellular levels of reducing agents. In terms of energy production, in addition to the ATP production by SLP during acetate production, one molecule of ATP is synthesized with the oxidation of formyl-THF to formate via the FTS while ATP is consumed by the ATP synthase to export $Na^+$ out of the cell to fuel the Rnf complex for the reduction of Fd. The reverse use of the methyl branch represents the only energetic solution to produce enough reducing equivalent and sustain growth.

Let us return to *E. callanderi* KIST612. Biochemistry approaches with enzymatic activity measurements bring evidence of similar use of the WLP for methanol consumption to *A. woodii*. The Fdh-Hydrogenase complex showed higher activity during formate oxidation compared to $CO_2$ hydrogenation in formate. Also, the MTD catalyzed the oxidation of methylene-THF into methenyl-THF. However, one uncertainty remains on the function of MTHFR regarding the nature of the electron carrier NADH or Fd as was detailed in a previous paragraph (§1.6.2.1.). Therefore, the

energetic models were based on two assumptions, whether the MTHFR can bifurcate electrons, in this case, $Fd^{2-}$ would be the electron donor or MTHFR does not bifurcate electrons and $NAD^+$ is the electron acceptor although the reaction would be highly endergonic with methyl-THF as electron donor ($\Delta G_0{}' = +23$ kJ/mol)[70].

These present energy conservation models from *E. callanderi* KIST612 show the last knowledge on acetogenesis and butyrogenesis on methanol, assuming electron bifurcation of the MTHFR or not for acetogenesis (Fig. 1.17 A & B) and butyrogenesis (Fig. 1.18 A & B) conditions. During acetogenesis, the ATP yield per mole of acetate is 0.91 without electron bifurcation for MTHFR while the value decreased to 0.75 with electron bifurcation. During butyrogenesis, the ATP yield reaches 2 ATP/butyrate without electron bifurcation while the value decreased to 1.5 ATP/butyrate with electron bifurcation. However, no energy conservation model including stoichiometry of acetate and butyrate measured from the experimental condition during methylotrophic growth on methanol was realized in *Eubacterium* species to date. The proposed energy conservation models will be useful to calculate energetic yield in experimental conditions. The work performed on this point for the strain *E. limosum* B2 will take profit from these theoretic energy conservation models to properly characterize the metabolism of the strain.

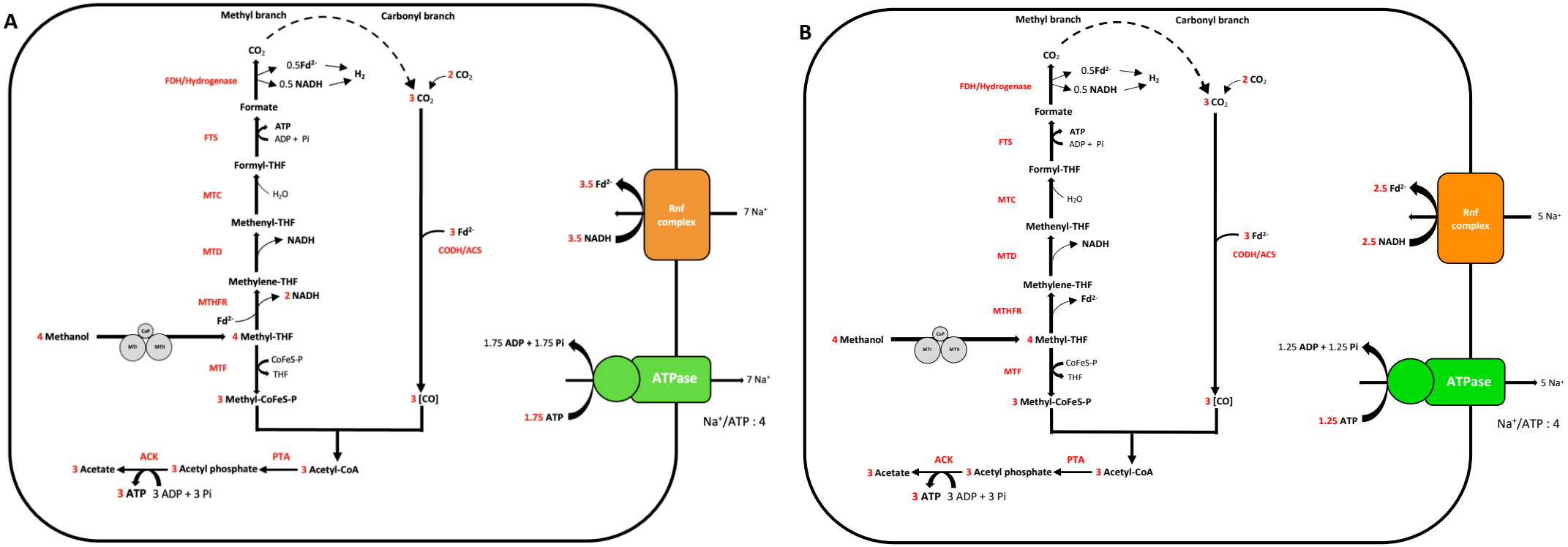



**Figure 1.17** : Energy conservation model of *E. callanderi* KIST612 during acetogenesis on methanol. A) Electron bifurcation is assumed for the MTHFR. B) No electron bifurcation for MTHFR. The list of enzymes is available in the legend of the figure 1.3.

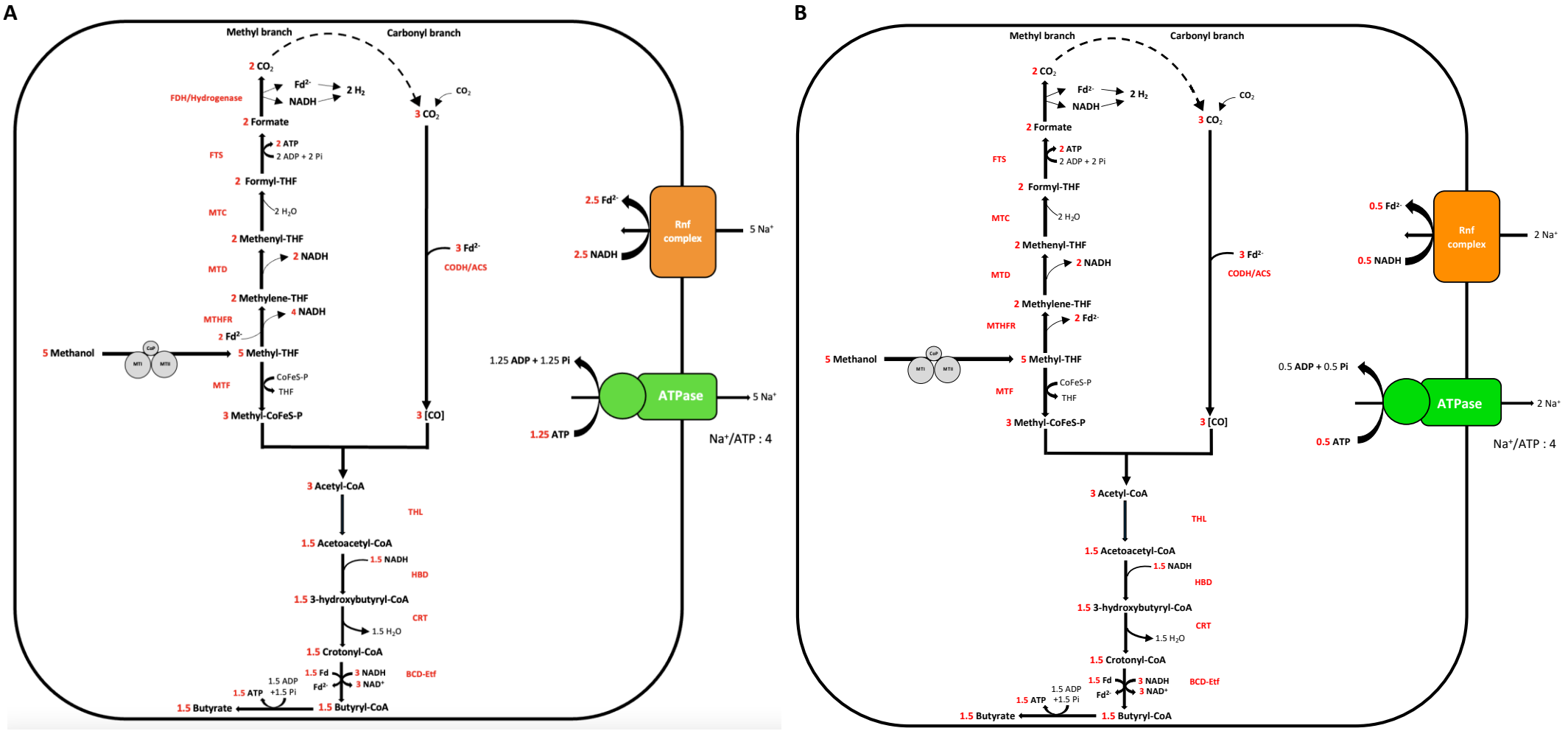



**Figure 1.18** : Energy conservation model of *E. callanderi* KIST612 during butyrogenesis on methanol. A) Electron bifurcation is assumed for the MTHFR. B) No electron bifurcation for MTHFR. The list of enzymes is available in the legend of the figure 1.3.

These models rely on a theoretical understanding of the methanol metabolism of *E. limosum*. However, a physiologic characterization of the strain *E. limosum* B2 on methanol was performed the at INSA laboratory mainly during the 80s. Of first importance for our commitment to deeply characterized this strain on methanol and other substrate conditions by a modern approach, the results obtained decades ago will be resumed in the next part.

## 1.11. Physiology of *E. limosum* growing on methanol

The physiologic characterization of a bacterium consists in measuring its growth parameters, such as optical density, growth rate, and product concentrations. The product yield can be determined from the input concentration value of the substrate added. The physiologic characterization of *E. limosum* has been well documented decades ago in heterotrophic conditions with glucose and methanol as carbon sources. Physiologic characterizations were performed in batch and continuous cultures.

In continuous culture, the culture is constantly fed by a fresh medium while the excess volume is constantly discarded to maintain the culture volume constant. Continuous cultures are performed in chemostats where medium flux rate, pH, and temperature are monitored. The advantage is to be able to work at the same specific growth rate for carbon sources. While the carbon source is completely consumed and becomes limiting, the culture reaches a steady state under normalized conditions to accurately compare the physiology of a strain for several different carbon sources. In chemostat cultures, the physiologic state of a strain is comparable among different conditions as long as the specific growth rate is constant.

### 1.11.1. Physiological characterization of *E. limosum* metabolism growing in batch culture on methanol

With methanol as a carbon source and $CO_2$ as a cosubstrate (under the forms of carbonates), stoichiometry depends on the concentration of acetate and reducing equivalents[21]. Excess acetate addition along with a low concentration of dissolved $CO_2$ led the strain of *E. limosum B2* to perform an homobutyric fermentation following this reaction:

$$10\ CH_3OH + 2\ CO_2 \rightarrow 3\ CH_3CH_2CH_2COOH$$

Consumption of carbon dioxide is mandatory to metabolize methanol. Inversely, an increase of dissolved $CO_2$ in the medium stimulates the production of acetic acid while decreasing the synthesis of butyric acid. The rate of dissolved $CO_2$ in the medium is mediated by the concentration of carbonates added to the medium. In methanol condition, increasing carbonates concentration to 200mM results in a coproduction of acetic acid and butyric acid, with a maximum acetate and butyrate production of 0,84 and 0.2 gram per gram of methanol respectively, giving the following reaction [21]:

$$10\ CH_3OH + 2\ CO_2 \rightarrow 4{,}5\ CH_3COOH + 0{,}75\ CH_3CH_2CH_2COOH$$

### 1.11.2. Physiological characterization of *E. limosum* metabolism growing on methanol in continuous culture

The physiological characterization of *E. limosum* B2 was performed in a chemostat culture using methanol or glucose as a carbon source[75,118]. These experimentations highlighted the capacity of the strain for methanol assimilation. Furthermore, acetate was shown to improve the growth capacity of the strain during methylotrophy. The maximum growth rate sustained was determined to be $0.06h^{-1}$, while a washout was observed above this value[118]. The products concentration between acetate and butyrate reached maximum values at the lowest growth rate tested: $0.0125h^{-1}$, with 60 mM-C of acetate and 68 mM-C of butyrate for 100mM-C of methanol added initially (the rest of the carbon necessary to balance the carbon comes from $CO_2$ consumed). The carbon distribution of methanol into products was almost equimolarly distributed with a slightly higher value in butyrate.

For the glucose condition, the maximum growth rate reached under chemostat conditions was $0.3h^{-1}$ [75]. The fermentation was almost homoacetogenic at dilution lower than $0.1h^{-1}$ with around 30 mM of acetate for 100 mM-C of glucose. Production of butyric acid was noticed at a dilution rate higher than $0.1h^{-1}$ with a maximum value of 15mM-C. This experiment showed the ability of the strain to produce butyrate during heterotrophy on glucose.

Determining the physiological parameters of a strain constitutes the first approach to accurately characterizing the metabolism. The systems biology approaches go further with the characterization of each metabolic intermediary. This branch of science is based on the modern omics approach for the identification and quantification of the total transcriptome, metabolome, and proteome of the cell.

## 1.12. Systems biology

The development of high throughput sequencing technologies allowed the building of metabolic models called genome-scale models (GSM) and the determination of metabolic fluxes with a precision that has never been achieved before. Such attention paid to the metabolic network allowed the identification of bottlenecks hindering the exploitation of fermentation to produce valuable compounds. Combined, omics technologies aimed to better understand how the metabolism is regulated as well as improve internal fluxes determination.

### 1.12.1. Genome-scale model

To optimize the production of targeted molecules by microorganisms and unlock the full potential of a microorganism, an understanding of the associated metabolic pathways is crucial. The rising of "omics" technologies brought a large amount of data on molecular components such as metabolites, transcripts, and proteins. Well analyzed, all this information underlies the mechanisms of a given phenotype or a cellular function. The main challenge is to construct a reliable network to interconnect the data obtained from the different omics analyses to the metabolic network used as a chassis. Based on years of research on biochemical characterization of metabolic reactions including enzymatic catalytic activity, genome sequencing, transcriptomics, and proteomics, genome scales models of metabolism were constructed to obtain a mathematical representation of metabolic reactions occurring in the cell with the determination of flux reactions [119,120]. The GSM aims to predict which modification would be necessary, from the genome or the medium condition, to obtain the desired phenotype. For example, simulations performed on the GSM of *C. autoethanogenum* in CO conditions, suggested an oversupply of CO to enhance the production yield of ethanol[121]. This oversupply would lead to an excess of reducing equivalents that would reorientate the metabolic flux to ethanol production. Furthermore, the model predicted that an excess of hydrogen would lead to a higher production of 2,3-butanediol (2,3-BD). This approach opens new testable experimentations to improve yields.

A metabolic analysis at the genome-scale during autotrophy was performed on *E. limosum* ATCC8486 using RNA-Seq to explore the transcriptome and ribosome profiling (RiboSeq) to decipher the translatome[122]. This last approach consists in sequencing the ribosome-protected mRNA fragment during the translation step. This method allows the determination of the translation efficiency of each mRNA and shows better accuracy compared to proteomic studies where total proteome extraction remains problematic, especially regarding membrane-bound

proteins[123]. However, RiboSeq does not consider post-translational regulation. This technique represents a strong additional feature to cure GSM. Furthermore, the study on *E. limosum* ATCC8486 shed light on the importance of the 5' Untranslated Transcribed Region (UTR) in translation efficiency and suggests that this conformation plays a role in translation regulation of mRNA under certain circumstances such as autotrophy where less energy is available. Indeed, during the translation process, the 5'UTR region of mRNA is broken by the ribosome to translate the transcript, but it requires energy in the form of guanosine triphosphate (GTP) a close neighbor of ATP. Although the transcript abundance was higher in autotrophic conditions, the translation efficiency on the carbonyl branch of WLP, energy conservation system including Rnf complex and ATPase was shown to decrease for the benefit of cellular respiration and electron bifurcation genes in autotrophic ($CO_2/H_2$) condition compared to heterotrophic (glucose) condition. This approach on the genome-scale level showed hidden regulation mechanisms. Another study performed on *E. callanderi* KIST 612 showed that proteomic fold change for most WLP enzymes and energy conservation systems showed high correspondence between methanol/$CO_2$ and $CO_2/H_2$ although the growth rate and $H_2$ consumption rate was different, with 4 and 2.7 folds higher respectively in methanol/$CO_2$ condition. The phenotypic differences observed between methanol and $CO_2/H_2$ observed was not explained by a proteome difference of central carbon mechanism or energy conservation underlying the importance to explore the metabolism on a wider aspect, including transcriptome, proteome, fluxome, etc[124].

Although studies performed on a genome-scale bring answers on the metabolic function of the cells, transcriptome and proteome are often compared relatively to other conditions. Some studies aimed to perform absolute quantification of the total transcriptome and proteome to reach a superior level of precision to improve existing GSM.

## 1.13. Absolute quantification in systems biology

If the integration of omics data in the models pushed the limits of our knowledge, few studies were focused on the absolute quantification of the proteome and the transcriptome. Recently, the first quantitative description of the proteome of an acetogen, *C. authoethanogenum* was published[125]. This approach allowed a better comprehension of the proteome allocation, with the identification of the energetic expenditure managed by the cell in a specific condition. Another interest is to identify undefined proteins of high abundance. Most of all, the determination of *in vivo* catalytic rates $k_{app}$ of enzymes is now possible by knowing the absolute number of proteins per cell and the internal fluxes.

The absolute quantification of the total number of proteins per cell was possible by the addition of external standards of known size and concentrations called stable-isotope-labeled (SIL) spike-ins based on endogenous proteins called anchor proteins. However, this approach is valid only at an equivalent growth rate, therefore chemostat culture is paramount to correctly build the experiment. This method allows the establishment of a linear correlation between the anchor protein concentrations and the mass spectrometry signal. The absolute proteome quantification of *C. autoethanogenum* was performed for three different conditions, CO, $CO/H_2$, and $CO_2/CO/H_2$. The linear correlation obtained from the use of SIL proteins based on the sequence of 16 anchor proteins allowed the absolute proteome quantification of around 1043 label-free proteins (fig. 1.19).

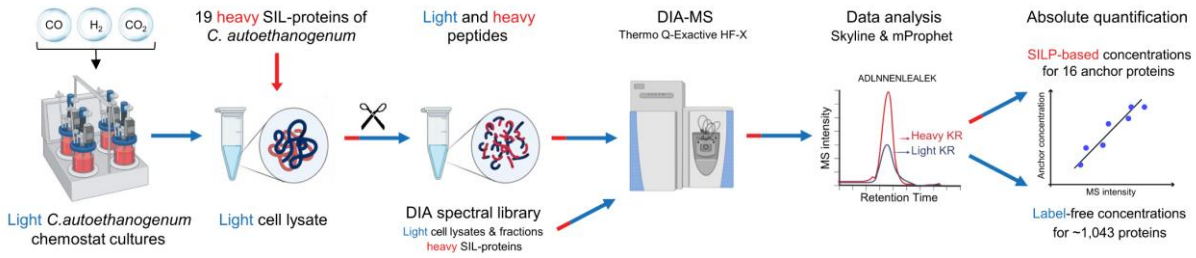



**Figure 1.19** : Schematic overview of absolute proteome quantification using heavy lysine and arginine amino acids for SIL-proteins to determine the absolute concentration of the total protein proteome. Heavy means isotopically marked protein containing arginine and lysine standards while light means nonmarked arginine and lysine amino acids. The heavy-to-light mass spectrometry intensity ratio was calculated to determine the concentration of total protein extracted. KR, lysine-arginine.

## 1.14. Addition of external standards for absolute quantification of mRNA

The addition of external standards during RNA-Seq routinely used is called External RNA Controls Consortium (ERCC). In 2003, industrials groups share with the NIST (National Institute of Standards and Technology) their problems regarding the results reproducibility of the gene expression data from RNA microarray. Soon after, the ERCC was created, an ad-hoc group of academicians, and public and private organizations were gathered to develop a common spike-in RNA molecule controls to overcome the lack of reproducibility. The ERCC spike-in mix is constituted of 92 polyadenylated RNA molecules that differ by their sizes (from 273 to 2022 nucleotides) and their concentrations (from 0.0143 to 30000 attomole/µL). Using transcripts inspired by eukaryotic mRNA allows full compatibility for prokaryotes experiments. The concentration of the ERCC mix span an approximately $10^6$ fold range concentration that is suitable for NGS (Next-Generation Sequencing) technologies.

The wide range concentration of the mix allows determining the limit of detection of size and concentrations of transcripts. A too-low concentration of mRNA can be undetectable by RNA-Seq as the transcript must be sequenced in multiple reads to be detected. If a transcript is present at too low a concentration, the amplification process before sequencing will probably fail. Furthermore, risks of RNA degradation during the extraction process remain important. For this reason, an effective extraction method that allows total transcriptome extraction combined with efficient protection of transcripts is required. Therefore, adding external standards such as ERCC spike-in mix allows the loss assessment of RNA during the extraction process.

Another commercial RNA spike-in kit exists, the Ambion Array controls which contains 8 RNA spikes-in with different sizes and concentrations. However, the spike-in sequences are copied from *the E. coli* genome limiting its general use with bacteria.

To date, only a few studies aimed to quantify the absolute number of transcripts per cell. A study published in 2019 by Gorochowski and his team used a homemade external RNA spike-in mix constituted of 6 RNA molecules, added just after the total extraction step to perform absolute mRNA quantification following the methodology below (Fig. 1.20) [126]. With known concentration, the mapped reads obtained after sequencing are converted into absolute molecule counts and normalized to the total number of cells used for RNA extraction to determine the absolute transcript number per cell [127].

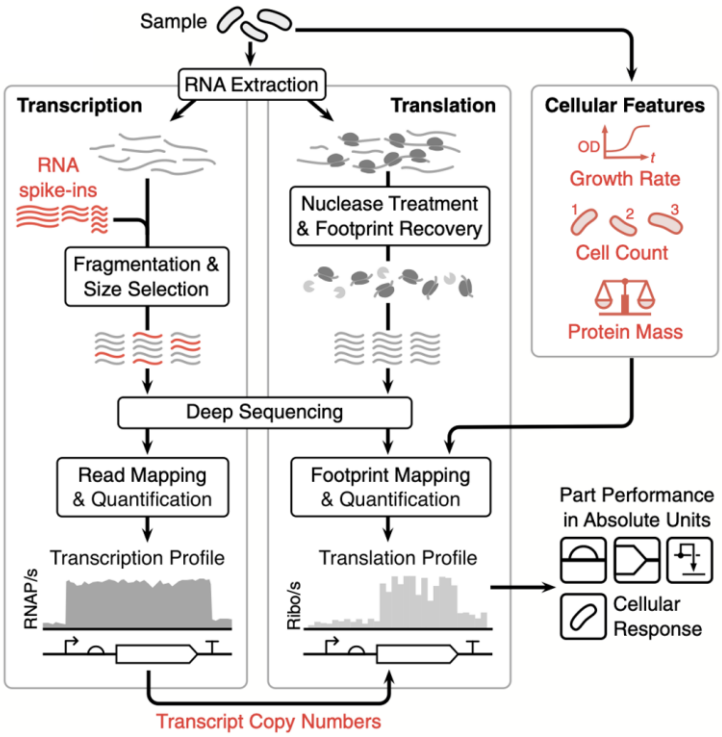



**Figure 1. 20** : Schematic overview of the strategy used by Goroshowski et al. to determine the absolute copy number of transcripts per cell[128]. Elements highlighted in red are required for absolute quantification. The protocol integrated RiboSeq to determine the translation efficiency.

## 1.15. Context and objectives of the thesis

The bibliographic introduction showed the current knowledge on acetogen metabolism with particular attention to *E. limosum* specie and its close genetic neighbor, *E. callanderi* KIST612. The ability of *E. limosum* to produce C4 molecules from $C_1$ compounds such as methanol is of great interest and makes *E. limosum* a promising candidate to produce $C_4$ compounds of industrial interest. Although almost fully characterized by several metabolic studies which bring light to the $CO_2$ fixation pathway, the WLP, remains uncharacterized regarding the nature of reducing equivalents involved in certain enzymatic reactions. The physiological characterization of *E. limosum* B2 was performed on a methanol semi-synthetic medium, mainly during the 80s. At the genome scale, the metabolism of *E. limosum* ATCC8486 was characterized by autotrophic conditions ($CO_2$ + $H_2$, CO). To date, no characterization of the central metabolism of methanol-grown cells was performed using a systems biology approach.

This present thesis is part of the European project BIOMETCHEM, funded by EraCobiotech. The project aimed to unlock the potential of *E. limosum* to produce value-added chemical products from methanol, a $C_1$ feedstock, with butanol and acetone as the main targets by the incorporation of exogenous pathways. Globally, the project connects resources and expertise from three European countries: A team from the Synthetic biology research center in the UK directed by Nigel Minton; a team from the universities of Ulm and Frankfurt in Germany directed by Peter Dürre and Volker Muller respectively, and a team from Toulouse Biotechnology Institute in France directed by Philippe Soucaille. Efforts commonly converged to i) characterize the methanol metabolism of *E. limosum*, ii) develop efficient systems for genetic modification, and iii) genetically design the strain for value-added chemicals production. In parallel to this project, the Korean team directed by Pr. Byang-Kwan Cho worked on the metabolic characterization of *E. limosum* ATCC8486 in autotrophic conditions and on the elaboration of a genome-scale model in collaboration with Nigel Minton and Philippe Soucaille teams.

The objective of the thesis is focused on the first point listed above. It consists of the characterization of the metabolism of *E. limosum* B2 on glucose or methanol as a carbon source using a systems biology approach with absolute quantification of both mRNAs and proteins. Such an approach was initiated in our team with the metabolic characterization of *C. acetobutylicum* by use of omics techniques. The thesis aimed to follow the footsteps of this previous work and go further in the analysis using the potential of RNA-seq and quantitative mass spectrometry to accurately characterized the *E. limosum* B2 metabolism in the context of the BIOMETCHEM project.

To achieve our goal, the E. *limosum* B2 WT strain will be adapted in methanol define medium by a process called adaptive laboratory evolution (ALE). Performed in numerous studies including a large variety of species, ALE was already performed on *E. limosum* ATCC 8486 on carbon monoxide semi-define medium[130]. This adaptation step, accompanied by genetic mutations through the evolutive process, is essential since culture in a rich medium makes impossible the accurate metabolism characterization on a specific substrate. The product profile, the maximal biomass reached and the growth rate for each subculture will be determined to depict the evolution of the growth performance of the strain in a methanol synthetic medium.
In parallel, the genome of the WT strain will be sequenced *de novo* as no complete genomic sequence of the *E. limosum* B2 strain was available. In addition, the complete genome of the $10^{th}$, $25^{th}$, $50^{th}$, and $75^{th}$ generations and three isolated clones from the final evolved population will be

sequenced to follow the genetic variability through the ALE. Furthermore, the growth improvement will be analyzed at the metabolic scale by the comparison of the total proteome (mass spectrometry) between the WT strain and the fastest-growing isolated clone from the final evolved population on a methanol mineral medium. The obtention of a robust strain able to grow on a methanol mineral medium will be valorized by a first publication supported by the physiologic, genomic, and proteomic data obtained.

The second part, the core project of the thesis, will be focused on the characterization of the metabolism of *E. limosum* B2 by a systems biology approach with absolute quantification of mRNAs and proteins. The first step will be the cultivation of the adapted strain previously obtained in substrate-limited continuous culture with similar carbon source concentration and dilution rate (specific growth rate) between the two substrate conditions to standardize the experiment. An enhancement of chemostat equipment for continuous culture will be realized to sustain growth in glucose or methanol mineral medium. The physiologic parameters such as optical density, product concentration, cell density, and dry cell weight will be determined from chemostat cultures at a steady state. From these first results, the elaboration of carbon stoichiometric equations will allow us to estimate the *in vivo* fluxes of enzymes involved in the WLP. These fluxes data will be essential to elaborate a theoretical energy conservation model, based on the last models designed on *E. callanderi* KIST 612 in methanol conditions where no physiologic experimental data were integrated. Furthermore, these data will be integrated into the genome-scale model of *E. limosum* co-developed between Toulouse, Nottingham, and KAIST laboratories to enhance the model.

To perform absolute quantification of mRNA and protein molecule numbers per cell, efficient quantitative extraction methods are required. A part of the thesis project will be focused on the optimization of these methods to ensure the best extraction yield for this strain. Many efforts will be applied to the isolation of total transcripts as it requires i) a particular extraction protocol ii) the addition of external standards at key steps of the extraction, iii) efficient DNase and ribodepletion treatment without losing transcripts of interest. Reaching this level of precision will allow us to precisely determine the absolute protein synthesis rate ($k_x$) by combining transcriptomic and proteomic results and the *in vivo* catalytic rate $k_{app}$ of a lot of enzymes involved in the central metabolism.

Finally, an accurate metabolic characterization at the genome-scale will be possible by combining the complete genomic sequence of *E. limosum* B2, the fluxome of associated proteins to the WLP,

and the absolute number of proteins and transcripts per cell. Through this project, we aimed to push forward the metabolic characterization using a systems biology approach with absolute quantification of mRNAs and proteins and to further extend this approach to other conditions such as lactate and syngas as carbon sources.

# Chapter 2: Genome sequence of *Eubacterium limosum* B2 and evolution for growth on a mineral medium with methanol and $CO_2$ as sole carbon sources.

by Guillaume Pregnon[1], Nigel Minton[2] and Philippe Soucaille[1,2*]

[1] INSA, UPS, INP, Toulouse Biotechnology Institute (TBI), Université de Toulouse, 31400 Toulouse, France

[2]BBSRC/EPSRC Synthetic Biology Research Centre (SBRC), School of Life Sciences, University Park, The University of Nottingham, Nottingham NG7 2RD, UK

[*]Author to whom correspondence should be addressed.

*Microorganisms* **2022**, *10*(9), 1790; **https://doi.org/10.3390/microorganisms10091790**

**Received: 20 July 2022 / Revised: 23 August 2022 / Accepted: 1 September 2022 /Published: 5 September 2022**

## 2.1. Abstract

*Eubacterium limosum* is an acetogen that can produce butyrate along with acetate as the main fermentation end-product from methanol, a promising C1 feedstock. Although physiological characterization of *E. limosum* B2 during methylotrophy was previously performed, the strain was cultured in a semi-defined medium, limiting the scope for further metabolic insights. Here, we sequenced the complete genome of the native strain and performed adaptive laboratory evolution to sustain growth on methanol mineral medium. The evolved population significantly improved its maximal growth rate by 3.45-fold. Furthermore, three clones from the evolved population were isolated on methanol mineral medium without cysteine by the addition of sodium thiosulfate. To identify mutations related to growth improvement, the whole genomes of wild-type *E. limosum* B2, the 10th, 25th, 50th, and 75th generations, and the three clones were sequenced. We explored the total proteomes of the native and the best evolved clone (n°2) and noticed significant differences in proteins involved in gluconeogenesis, anaplerotic reactions, and sulfate metabolism. Furthermore, homologous recombination was found in subunit S of the type I restriction-modification system between both strains, changing the structure of the subunit, its sequence recognition and the methylome of the evolved clone. Taken together, the genomic, proteomic and methylomic data suggest a possible epigenetic mechanism of metabolic regulation.

**Keywords:** *Eubacterium limosum;* acetogen; adaptive laboratory evolution; genome sequence*;* methanol; butyric acid; mineral medium; proteomics

## 2.2. Introduction

Strictly anaerobic acetogenic bacteria are known to fix one carbon (C1) substrates, such as synthesis gas (syngas) $CO_2$ + $H_2$ or CO, using the methyl and carbonyl branches of the Wood–Ljungdahl pathway (WLP). Syngas metabolization consists of the reduction of two molecules of $CO_2$ with $H_2$ as an electron donor or by incorporation of CO in the carbonyl branch by the carbon monoxide dehydrogenase/acetyl-CoA enzyme and results in the production of one molecule of acetyl-CoA[15]. This central intermediate is mainly used to produce ATP and acetate. However, depending on the acetogen, various molecules can be produced from acetyl-CoA, such as ethanol, or more complex products, including lactate, butyrate, butanol or 2,3-butanediol[10,19]. Conversion of C1 compounds to four-carbon products, such as butyrate, by microorganisms has received much attention due to its potential application in numerous fields, including the energy, cosmetics, and pharmaceutical industries[129]. Moreover, butyryl-CoA, an intermediate in the butyrate pathway, can be reduced to n-butanol, an attractive molecule that can be used both as a platform chemical and a biofuel[63]. Driven by growing ecological interest, syngas exploitation by bacteria is booming with the development of the first factories of industrial scale[130]. However, gas substrate utilization faces several technical challenges, such as storage and gas–liquid mass transfer, decreasing bacterial productivity[131132,133]. Methanol constitutes an attractive C1 carbon source due to its wide availability throughout the world, easy production process, storability, and liquid state to avoid mass transfer issues. The acetogen *Eubacterium limosum* can metabolize methanol and remains one of the few bacterial species able to convert it to butyrate with the coproduction of acetate[21]. Although numerous physiological studies of *E. limosum* B2 have been performed in the past[35,75,134–136], its genome sequence is still undetermined, and the first biochemical characterization of anaerobic methanol metabolism, including the energetic aspect, was performed in the model methylotrophic bacterium *Acetobacterium woodii*[20], which does not produce C4 compounds. Two molecules of methanol are converted to two molecules of methyl-THF by a methyltransferase system. One molecule is oxidized into $CO_2$ through the reverse use of the methyl branch to produce the reducing equivalents needed to then reduce $CO_2$ to CO, while the second molecule is used with CO by the ACS/CODH complex to produce acetyl-CoA. This method of methanol assimilation through the WLP was also demonstrated for *Eubacterium callanderi* KIST 612, formerly named *E. limosum* KIST612[70]. Moreover, the first evidence of butanol production by *E. limosum* ATCC 8486 from methanol and formate used as cosubstrates was recently obtained, reinforcing the interest in this bacterial species[22]. Methanol metabolism by *E. limosum* B2 was characterized in semi-defined medium containing yeast extract (YE) and cysteine[21]. To the best of

our knowledge, no study has been performed on any *E. limosum* strain to develop a mineral medium for growth on methanol as a carbon and energy source. A strain that can grow on a mineral medium will have several advantages for in-depth metabolic characterization using flux balance analysis and genome-scale models[137]. The semi-defined medium previously used for *E. limosum* B2 contained YE but also cysteine, as the strain was shown to be an auxotroph for cysteine[21,30]. Synthetic medium free from YE is also of economic interest, as YE is an expansive supplement prohibiting favourable economic scale-up [138]. Growth improvement can be easily achieved by adaptive laboratory evolution (ALE), which consists of culturing a microorganism for several generations in restrictive medium to naturally improve its growth parameters[139]. Adaptive laboratory evolution has already been performed on *E. limosum* ATCC 8486 using CO as a carbon and energy source, improving its growth rate 1.4-fold[140]. Here, we aimed to deepen our knowledge of the *E. limosum* B2 strain at the genome level and to develop a mineral medium allowing good growth on methanol of an evolved strain for further performing systems biology analysis in chemostat culture. For this purpose, we (i) performed de novo whole-genome sequencing by single-molecule real-time technology (SMRT), (ii) developed a defined medium allowing the strain to grow with methanol as a carbon and energy source without the addition of YE but with cysteine, (iii) improved the growth performance of the strain on this medium by ALE, (iv) developed a mineral medium allowing the adapted strain to grow with methanol as a carbon and energy source without the addition of cysteine, (v) studied genomic modifications that occurred during the evolution process and (vi) performed a comparative study of the total proteome between an isolated clone from the evolved population and the wild-type strain.

## 2.3. Materials and Methods

### 2.3.1. Bacterial Strain and Growth Conditions

The *E. limosum* B2 strain used in this study has been studied in our laboratory since 1985 and is stored at −80 °C. Culture growth was performed in serum bottles (in 25 mL of culture medium) under a strict anaerobic nitrogen ($N_2$) atmosphere and incubated at 37 °C at pH 7.4. The wild-type strain was cultured in rich medium for anaerobic bacteria (M187), which comprised the following components: tryptone 30 g/L, yeast extract 20 g/L, glucose 5 g/L, and cysteine hydrochloride 0.5 g/L. The adapted strain in the methanol substrate was cultured in synthetic medium composed of the following components: NaCl 0.6 g/L, $NH_4Cl$ 1 g/L, $MgSO_4$-$7H_2O$ 0.12 g/L, $CaCl_2$ 0.08 g/L, $KH_2PO_4$ 1 g/L, vitamin solution 100X (biotin 20 mg/L, pantothenic acid 50 mg/L, lipoic acid 50 mg/L), sodium thiosulfate pentahydrate 4 mM, potassium carbonate 1 g/L, titanium citrate 2 mM as a reducing agent, and trace element solution 100X ($MnCl_2$ 0.1 g/L, $CoCl_2$-$2H_2O$ 0.1 g/L, $ZnCl_2$ 0.1 g/L, $CuCl_2$ 0.02 g/L, $H_3BO_3$ 0.01 g/L, Na molybdate 0.01 g/L, $Na_2SeO_3$ 0.017 g/L, $NiSO_4$-$6H_2O$ 0.026 g/L, $FeSO_4$,$6H_2O$ 1 g/L and nitrilotriacetic acid 12.8 g/L). Resazurin was added at $1.10^{-4}$% (*wt*/*wt*) as a redox potential indicator. Glucose (6 g/L) and methanol (6.4 g/L) were the two substrates tested, corresponding to 200 mM carbon.

Isolated clones from the methanol-adapted population were cultured in solid synthetic medium with glucose as the substrate but without cysteine. The product concentrations for solid synthetic medium elaboration remained the same as those for liquid synthetic medium supplemented with 15 g/L of agarose. After solidification, plates were placed in an anaerobic chamber overnight before streaking the population to ensure balance of the anaerobic atmosphere on the solid medium (90% $N_2$, 5% $CO_2$ and 5% $H_2$).

### 2.3.2. Adaptive Laboratory Evolution

The ALE was performed from the first generation of a yeast extract-free adapted *E. limosum* B2 strain previously obtained in the laboratory in synthetic medium with 200 mM methanol as the substrate (unpublished data). A 5% volume of culture was transferred to fresh medium every three generations. The population was cultured over 100 generations.

### 2.3.3. Clone Isolation on Methanol Mineral Medium without Cysteine

Single clones from the adapted population (over 100 generations) were isolated in solid glucose mineral medium without cysteine in an anaerobic chamber at 37 °C. Cysteine was replaced by 4 mM sodium thiosulfate in the synthetic medium. Three clones were cultivated in liquid methanol mineral medium without cysteine to follow and compare the growth profile from the 10th, 25th, 50th and 75th generations adapted in MMM with cysteine. Cysteine was replaced by 4 mM sodium thiosulfate.

### 2.3.4. Analytical Methods

Culture density was measured by spectrophotometry at 620 nm (OD620) against a water blank. Growth rates were calculated from these measurements. Substrates and products were measured using a high-performance liquid chromatography instrument (Agilent 1200 series, Massy, France) equipped with a refractive index detector and Aminex HPX 87 H 300 × 7.8 mm. The acid $H_2SO_4$ (0.5 mM) was used as the mobile phase and elution was performed at 48 °C. The biomass formula $C_4H_7O_2N_{0.6}$ was used to convert biomass values into molar cell carbon concentration (mC)[141]. Acetate and butyrate production yields were calculated in millimolar carbon from the product per 100 mM carbon from methanol (mC/100 mC MetOH).

### 2.3.5. Genomic DNA Extraction

For de novo sequencing, the wild-type strain of *E. limosum* B2 was cultured in rich medium for anaerobic bacteria (M187). Genomic DNA of harvested cells in mid-exponential phase from 1 mL culture was carried out with the Wizard® HMW DNA extraction kit (Promega, Madison, WI, USA) according to the manufacturer's instructions with the following modifications. The lysis step with lysozyme treatment was extended to 2 h to optimize the DNA extraction yield and genomic DNA was resuspended in 10 mM Tris-HCl (pH 8.5) solution. The DNA quantification was performed spectrophotometrically by a Nanodrop 2000 spectrophotometer (Thermo Fisher

Scientific, Wilmington, DE, USA) and fluorometrically using a Qubit™ instrument (Thermo Fisher Scientific, Wilmington, DE, USA). Quality assessment was performed by (i) gel electrophoresis, (ii) the A260/A280 and A260/230 ratios (>1.9) and (iii) DNA concentration comparison by Qubit and Nanodrop. The difference between the results from both instruments had to be less than 2 for validation.

For genomic comparison of the evolved population through generations, whole-genomic DNA of the wild-type strain, the 10th, 25th, 50th, and 75th generations and three isolated clones was extracted using the same kit as that used for the WT strain for de novo sequencing, measured by Nanodrop and stored at −20 °C until sequencing.

#### 2.3.6. Whole Genome Sequencing

Single-molecule real-time long read sequencing of the *E. limosum* B2 wild-type strain was performed at the Gentyane Sequencing Platform (Clermont-Ferrand, France, https://gentyane.clermont.inrae.fr, accessed on 25 November 2020) with a PacBio Sequel II Sequencer (Pacific Biosciences, Menlo Park, CA, USA). The SMRTBell library was prepared using a SMRTbell Express 2 Template prep kit (Pacific Biosciences, Menlo Park, CA, USA), following the "procedure and checklist -preparing Multiplexed Microbial Libraries using SMRTbell Express Template prep kit 2.0" protocol. Genomic DNA (1 µg) of each strain was sheared using g-tubes (Covaris, England), generating DNA fragments of approximately 10 kb. A Fragment Analyzer (Agilent Technologies, Santa Clara, California, USA) assay was used to assess the fragment size distribution. Sheared genomic DNA was carried into the enzymatic reactions to remove the single-strand overhangs and to repair any damage that may be present on the DNA backbone. An A-tailing reaction followed by overhang adapter ligation was conducted to generate SMRTbell templates. The sample was then purified with 0.45X AMPure PB beads to obtain the final libraries of approximately 10 kb. The SMRTBell libraries were inspected for quality and quantified on a Fragment Analyzer (Agilent Technologies, Santa Clara, CA, USA) and with a Qubit fluorimeter with the Qubit dsDNA HS reagent Assay kit (Life Technologies, Eugene, OR, USA). A ready-to-sequence SMRTBell Polymerase Complex was created using a Binding Kit 2.0 (Pacific Biosciences, Menlo Park, CA, USA) and the primer V4, and the diffusion loading protocol was used according to the manufacturer's instructions. The PacBio Sequel instrument was programmed to load a 100 pM library and sequenced in CLR mode on a PacBio SMRTcell 8M, with a Sequencing Plate 2.0 (Pacific Biosciences, Menlo Park, CA, USA) and 2 h of pre-extension time and acquiring one movie of 15 h per SMRTcell.

Whole-genomic DNA of the wild-type strain, the 10th, 25th, 50th, and 75th generations and the three isolated clones were sequenced with the Ion S5 system on the Get-Biopuces sequencing platform (Toulouse, France, https://get.genotoul.fr, accessed on 20 March 2021).

#### 2.3.7. Functional Annotation and Genome Analysis

Gene prediction and functional annotation of the wild-type strain of *E. limosum* B2 were performed using the NCBI Prokaryotic Genome Annotation Pipeline PGAP [24]. The genome sequence can be found under the accession number CP097376.1. Whole-genome sequence alignment of *E. limosum* B2 against *E. limosum* ATCC 8486 was performed using ProgressiveMauve (v2.4.0)[142].

#### 2.3.8. Methylome and Type I Restriction-Modification System Analysis

For analysis of the whole DNA methylation level, methylation of adenine on the 6$^{th}$ carbon (m6a) and methylation of cytosine on the 4th carbon (m4c) were determined using the ipdSummary and Motifmaker tools included in the SMRT Link (v10.1). The protein sequence alignment for the analysis of the type I RM system was performed using ClustalW in Ugene (v40.1). Conserved domain analysis was performed using the NCBI conserved domain search tool.

#### 2.3.9. Protein Isolation and Total Proteome Analysis

Wild-type *E. limosum* B2 and clone 2 were cultured in methanol mineral medium containing 0.5 g/L yeast extract in duplicate. Cells were harvested at mid-exponential growth phase, and sample volumes were adjusted to normalize the cell number to $10^9$ for each condition. The samples were then centrifuged at 16,000× *g* for 30 s, and the supernatants were discarded. Cell pellets were washed twice with 1 mL of phosphate-buffered saline solution. After discarding the supernatant, the cell pellets were frozen in liquid nitrogen and stored at −20 °C. Total proteins were isolated using the SPEED extraction method[143]. Each pellet was resuspended in 150 µL of trifluoroacetic acid (TFA, Supelco MS Grade, Darmstadt, Germany), warmed at 70 °C for 3 min and centrifuged at 16,000× *g* for 1 min. A total of 130 µL of supernatant was carefully transferred to a fresh 1.5 mL Eppendorf tube. Fourteen microliters of supernatant were taken for each sample and diluted in 56 µL 75% TFA to measure the protein concentration by the Bradford assay. Each sample was then neutralized with 1.5 mL of TRIS 2 M, prepared in LC–MS grade water (Supelco), transferred to glass vials and freeze dried. Samples were sent to the John van Geest Cancer Research Center

at Nottingham, resolubilized, purified and injected into a TripleTOF 6600 mass spectrometer (Sciex). Two injections per sample were performed, giving a technical duplicate for each sample. Sequential window acquisition of all theoretical spectra (SWATH) and additional data treatment using DIA-NN (https://github.com/vdemichev/DiaNN, accessed on 1 April 2022) were performed to analyse the two groups of samples[144].

## 2.4. Results

### 2.4.1. *De novo* Sequencing, Annotation and Analysis of the E. limosum B2 Genome

Although the genomes of several *E. limosum* strains have been sequenced, the genome sequence of *E. limosum* B2, the most well-studied strain from a physiological point of view, remains unknown[32,88,145]. De novo sequencing was performed on the WT strain using single molecule real-time technology (PacBio). A total of 76,620 long reads were used to produce two linear contigs of 4.06 Mb and 358 kb for a total genome size of 4.424 Mb. The main issue in obtaining one circular contig was the presence of a prophage, present in two copies in the genome and localized at one end of each of the contigs. A complete circular assembly of 4,421,327 base pairs was created using long-range PCR associated with Sanger sequencing. Based on large reads, PacBio sequencing technology can be prone to errors[146]. To correct possible errors from the sequencer, the circular genome sequence was compared to the whole genome sequence of the *E. limosum* B2 WT strain obtained by the IonTorrent S5 sequencer. One nucleotide insertion was observed in the PacBio sequence, localized in an *ompD* homologous gene, which was not confirmed by Ion S5 sequencing. The genome sequence of *E. limosum* B2 has a G+C content of 47.2% and contains 4192 genes, 51 tRNA genes, 16 rRNA genes including 5 23S, 16S and 5S rRNA operons, a supplementary 5S rRNA gene, 1 tmRNA and two clusters of CRISPR genes, similar to the *E. limosum* SA11 and KIST612 strains[32,145]. The genome of *E. limosum* B2 showed 99.837% sequence identity with the *E. limosum* ATCC 8486 strain.

### 2.4.2. Genomic Comparison of *E. limosum* B2 vs. *E. limosum* ATCC 8486

Differences between genes and intergenic regions were identified (Tab. 2.1).

**Table 2.1** : Genomic differences between *Eubacterium limosum* B2 and ATCC 8486 strains.

| Locus Tag ATCC8486 | Locus Tag B2 | Position B2 | Position ATCC8486 | Gene | Nucleotide Variation ATCC8486/B2 | Type | AA Change | Description |
|---|---|---|---|---|---|---|---|---|
| B2M23_RS00035 | M5595_17445 | 3667592 | 4431 | - | G1265T | SNP | $Ala^{421}Glu$ | Phage terminase |
| B2M23_RS00620 | M5595_16850 | 3547040 | 124983 | - | A514G | SNP | $Tyr^{171}His$ | Acyl ACP thioesterase |
| B2M23_RS00770 | M5595_16695 | 3512058 | 159965 | - | C1065T | SNP | $Val^{355}Val$ | Transcriptional activator |
| Intergenic region | Intergenic region | 2725372 | 946652 | - | G/A | SNP | - | - |
| Intergenic region | Intergenic region | 3264417 | 407607 | - | A/- | DEL | - | - |
| B2M23_RS06095 | M5595_11305 | 2338369 | 1333655 | *cheY* | A187G | SNP | $Arg^{63}Gly$ | Response regulator |
| B2M23_RS08865 | M5595_08530 | 1755671 | 1916353 | *aspS* | G1004A | SNP | $Gly^{335}Glu$ | Aspartate–tRNA ligase |
| B2M23_RS08970 | M5595_08425 | 1735036 | 1936988 | - | G1007T | SNP | $Thr^{336}Asp$ | Bacitracin export permease |
| Intergenic region | Intergenic region | 1442364 | 2231325 | - | A/C | SNP | - | Lysin riboswitch |
| B2M23_RS11215 | M5595_20825 | 4406797 | 2387123 | - | -/G | INS | - | Hypothetical protein |
| Intergenic region | M5595_20830 | 4406887 | 2387213 | - | -/A | INS | - | Hypothetical protein |
| Intergenic region | M5595_20830 | 4406916 | 2387242 | - | -/T | INS | - | Hypothetical protein |
| B2M23_RS13215 | M5595_19385 | 4089680 | 2815304 | - | G234T | SNP | $Met^{78}Ile$ | Mobilization protein |
| Intergenic region | M5595_00360 | 73507 | 3183052 | - | -/C | INS | - | Restriction endonuclease subunit S |
| Intergenic region | M5595_00360 | 73556 | 3183101 | - | -/TG | INS | - | Restriction endonuclease subunit S |
| Intergenic region | M5595_00360 | 73990 | 3183516 | - | -/G | INS | - | Restriction endonuclease subunit S |
| Intergenic region | Intergenic region | 367145 | 3476696 | - | T/- | DEL | - | - |
| B2M23_RS17820 | M5595_03295 | 694242 | 3803795 | *ktrA* | A/- | DEL | | TrkA family potassium uptake protein |
| B2M23_RS18850 | M5595_04320 | 887882 | 3997435 | *pucK* | C/T | SNP | $Ile^{116}Thr$ | Uric acid permease PucK/Xanthine permease |
| B2M23_RS18850 | M5595_04320 | 887885 | 3997438 | *pucK* | A/G | SNP | $Gly^{117}Asp$ | Uric acid permease PucK/Xanthine permease |
| B2M23_RS19910 | M5595_05415 | 1121402 | 4230955 | *pbpF* | T/C | SNP | $Ser^{459}Pro$ | Transglycosylase domain-containing protein |

In addition to large genomic rearrangement, as we describe later, a total of 21 small genomic differences, including 12 single nucleotide polymorphisms (SNPs), six nucleotide insertions and three single nucleotide deletions, were observed for the *E. limosum* B2 genome compared to the *E. limosum* ATCC 8486 chromosome. Differences occurred mainly in genes coding for regulators, such as a potential response regulator (M5595_11305) introducing an $Arg^{63}Gly$ substitution in the protein or a transcriptional activator (M5595_16695) introducing a silent mutation. A total of four genetic differences were detected in genes coding for a prophage operon, one in a terminase encoding gene (M5595_17445) leading to an $Ala^{421}Glu$ substitution in the protein, two in intergenic regions between genes of unknown function and one in a gene coding for a hypothetical protein. Furthermore, three insertions were noticed in a gene of *E. limosum* B2 encoding a potential S subunit (M5595_00360) of a type I restriction endonuclease. Other genomic differences were localized in genes encoding transporters, such as a uric acid permease-encoding gene (M5595_04320) or a bacitracin export permease-encoding gene (M5595_08425). Additionally, deletion of an insertion sequence (IS) was detected in the genome of *E. limosum* B2 compared to *E. limosum* ATCC 8486. This IS was present in six and five copies in *E. limosum* ATCC 8486 and B2 strains, respectively. The missing IS in *E. limosum* B2 was located upstream of a sodium/phosphate transporter-encoding gene, reducing the size of the gene by 42 nucleotides in the *E. limosum* ATCC 8486 gene according to the annotation. Whole genome alignment using ProgressiveMauve also revealed genome reorganization of five locally colinear blocks (LCBs) with reverse complement orientation (Fig. 2.1).

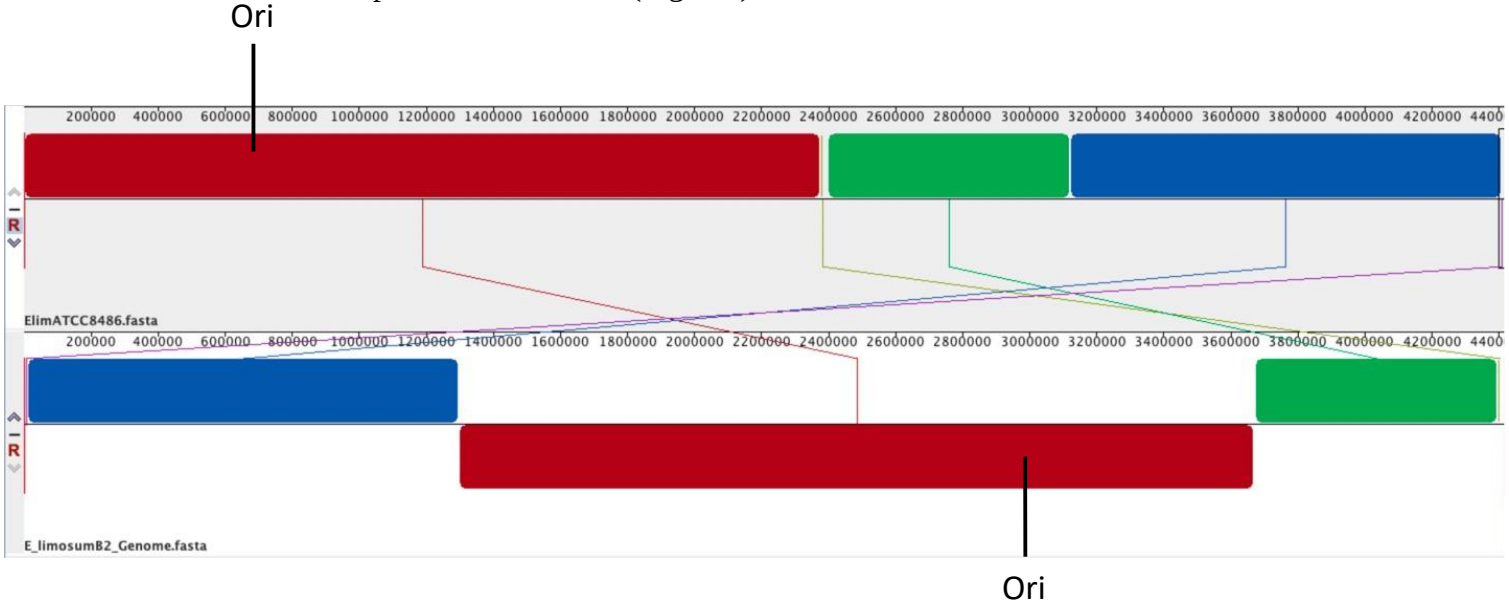



**Figure 2.1** : Whole genome alignment of *E. limosum* B2 (bottom) to *E. limosum* ATCC 8486 (top) using ProgressiveMauve software. Ori, origin of replication at genomic position 689,833 for *E. limosum* ATCC8486 and 2,982,134 for *E. limosum* B2.

### 2.4.3. Whole DNA Methylation Level

The DNA methylation of the chromosome is an epigenetic modification that plays a crucial role in the response to environmental stresses, genome structure rearrangements and gene regulation [147]. To explore the methylome of *E. limosum* B2, a kinetic data analysis was performed on the SMRT sequence obtained on the PacBio system to identify the locations of m6A and m4C methylated nucleotides. The five restriction-modification (RM) systems identified in *E. limosum* ATCC 8486, including one type I RM and four type II RMs, were also present in the *E. limosum* B2 strain. Methylome data from *E. limosum* B2 showed only three different motifs of the m6A type for a total of 3974 methylated motifs out of 8444 probable motifs. The *E. limosum* B2 strain had a modification frequency of 0.96 modifications/kb (Tab. 2.2).

**Table 2.2** : Summary of methylome data from the *E. limosum* B2 and the evolved clone 2 strains.

| Strain | Motif | Predicted RM Type | Modified Position | Methylation Type | Fraction (%) | Number of Methylated Motifs Detected | Number of Total Motifs Detected |
|---|---|---|---|---|---|---|---|
| *E. limosum* B2 | GCGRAG | II | 5 | m6A | 67.6 | 2058 | 3042 |
| | CAAAAAR | II | 6 | m6A | 37.3 | 1623 | 4350 |
| | CNNTAYNNNNNTCC | I | 5 | m6A | 27.9 | 293 | 1052 |
| *E. limosum* B2 clone 2 | GCGRAG | II | 5 | m6A | 81 | 2464 | 3042 |
| | CAAAAAR | II | 6 | m6A | 83.7 | 3642 | 4350 |
| | CNNTAYNNNNGTG | I | 5 | m6A | 79.8 | 576 | 1052 |

The GCGRA$^{m6}$G motif was most prevalent, with 2058 occurrences among a total of 3042 sites. A GCGRA $^{m6}$G-like motif, 'GCGCA $^{m6}$G', was identified in *Rhuminococcus flavefaciens* FD-1 according to the REbase databank. This motif site is recognized by a type II RM (RflFIII), sharing 66% similarity to Class I SAM-dependent methyltransferase from *Eubacterium* species but only 21% similarity with N-6 DNA methylase from *E. limosum* B2.

Compared to the methylome of a close metabolic neighbour, *A. woodii*, almost all the detected motifs carried a methyl group, while the motif fraction that contained methylated nucleotides was lower (67.6%) for *E. limosum* B2[148]. Furthermore, the number of methylated nucleotides expressed per kb of DNA was calculated to be 0.96 modifications/kb for *A. woodii*. The CAAAAA $^{m6}$R methylated motif present in the *E. limosum* B2 genome was also detected in *A. woodii*. This adenine in position 6 in the motif is methylated by a type II methyltransferase, annotated N-6 DNA methylase (Awo_c14460), sharing 48.92% similarity to the N-6 DNA methylase of *E. limosum* B2 (M5595_07210). A total of 27.9% of CNNTAYNNNNNTCC carry a modification of their

adenine at positions 5. Although the associated enzyme remains unknown, these nonpalindromic bipartite motifs are likely recognized by a Type I RM system and most likely encoded by the M5595_00345 to M5595_00360 genes.

### 2.4.4. Strain Adaptation on Methanol-Defined Medium without Yeast Extract

The first step for growth optimization on methanol substrate was to adapt the strain on a YE-free defined medium. The WT strain, initially cultivated in M187-rich medium, was inoculated in methanol-defined medium without yeast extract to obtain a strain that could grow in YE-depleted medium. Two subcultures in freshly defined medium were performed to ensure YE from the original rich medium culture was at least diluted 400- fold and played a minor role in the YE-depleted medium. The cell density of the WT strain in methanol-defined medium was compared to that in methanol-rich medium as a control (Fig. 2.2).

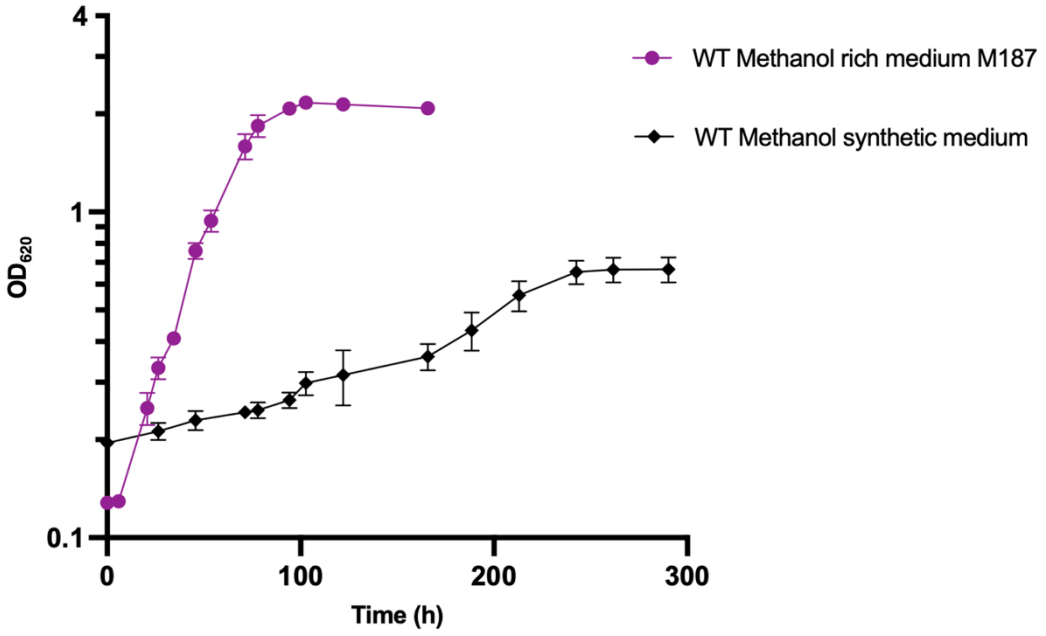



**Figure 2.2** Growth profile of *E. limosum* B2 on methanol-rich and synthetic medium without YE. Cultures were performed in triplicate and errors bars indicates standard deviation.

After a 3-day lag phase, a population started to grow, showing a maximum $OD_{620}$ of 0.670 ± 0.05 and a maximum specific growth rate (µmax) of 0.017 ± 0.002 $h^{-1}$. The WT strain cultivated on M187 rich medium showed a maximum $OD_{620}$ of 2.165 ± 0.05 with a µmax of 0.049 ± 0.003 $h^{-1}$. The growth rate in methanol-defined medium decreased by 65% compared to that in the rich medium, but the WT strain showed its ability to overcome YE depletion. Obtaining a strain that

can grow without YE represents a strong asset to precisely determine metabolic fluxes from different carbon sources using a genome-scale model. However, the specific growth rate of 0.017 $h^{-1}$ remained low compared to the maximum growth rate of 0.05 $h^{-1}$ measured for the *E. limosum* B2 strain in semi-defined medium supplemented with YE 0.5 $g \cdot L^{-1}$ YE[35]. The next step of the study was to improve the physiological characteristics of the strain in methanol-defined medium by ALE.

### 2.4.5. Adaptative Laboratory Evolution of *E. limosum* B2 in a Methanol-Defined Medium

The ALE on *E. limosum* ATCC 8486 was reported on carbon monoxide in semi-defined medium supplemented with 2 g. $L^{-1}$ YE. The results showed a notable improvement in the physiological characteristics of the *E. limosum* ATCC 8486 strain, with a maximal optical density (620 nm) and a growth rate that increased 2.14-fold and 1.44-fold[149] , respectively. To the best of our knowledge, no ALE has been performed using methanol as a substrate with *Eubacterium* species. Few studies have focused on ALE to improve growth on methanol. Such an experiment was performed on the natural methylotroph *Saccharomyces cerevisiae,* improving the final biomass by 44% after 230 generations[150]. Additionally, ALE on methanol was performed for synthetic methylotrophs such as *Escherichia coli* or *Corynebacterium glutamicum* to improve their tolerance[151,152] to methanol. The ALE was performed on *E. limosum* B2 to obtain a strain with the best growth performance on methanol-defined medium. In total, the strain was cultured for more than 100 generations with a subculture performed every three generations. The ALE showed a clear improvement in the maximal growth rate (Fig. 2.3A).

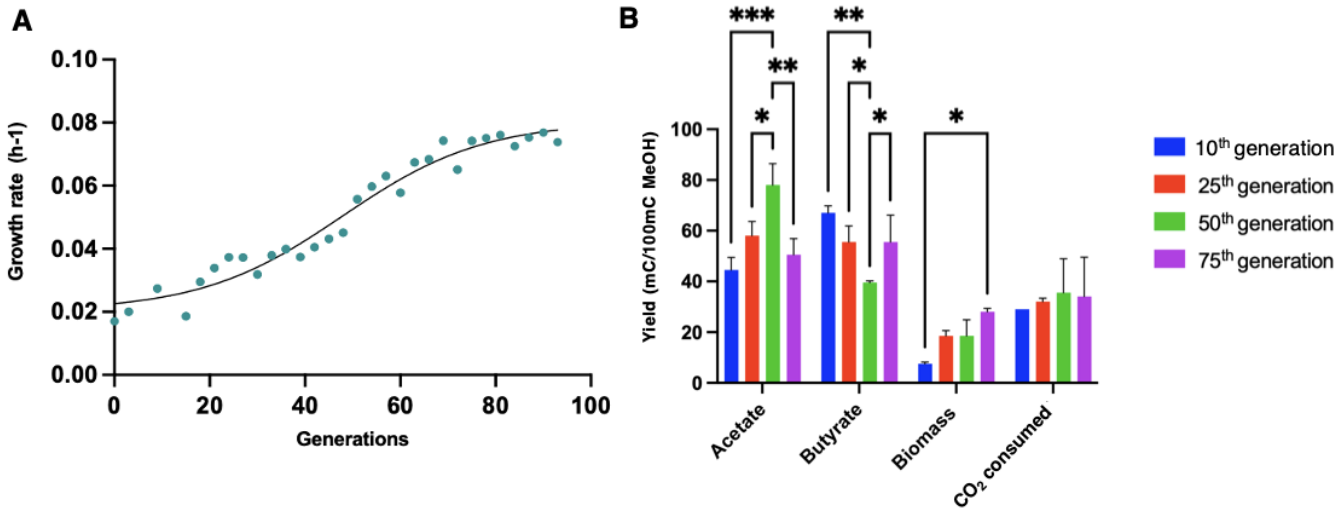



**Figure 2.3 :** Evolution of physiologic characteristics during ALE on methanol methylotrophic growth of *E. limosum* B2. (**A**) Evolution of the maximum growth rate for each generation during ALE. (**B**) Carbon yields obtained from consumed methanol to acetate, butyrate, biomass, and $CO_2$. Yields are expressed in millimolar of carbon (mC) per 100 mC of methanol consumed. Cultures were performed in triplicate and error bars indicates standard deviation. Asterisk indicates that the results were significantly different from the control according to the one-sided Student's *t*-test: *: *p* value ≤ 0.05, **: *p* value ≤ 0.01, ***: *p* value ≤ 0.001.

The lag phase was gradually reduced as the strain was subcultured, from approximately 70 h for the first generation to less than 5 h for the evolved population. The growth rate progressively increased over generations, with a µmax at 0.022 ± 0.001 $h^{-1}$ for the 10th generation, 0.047 ± 0.0084 $h^{-1}$ for the 25th generation, and 0.057 ± 0.0026 $h^{-1}$ for the 50th generation, reaching a maximal value of 0.076 ± 0.0019 $h^{-1}$ for the 75th generation. Beyond this last generation, no significant physiological growth improvement was noticed, suggesting that the population reached its maximal potential. In comparison, a maximum growth rate of 0.07 $h^{-1}$ was observed for *E. limosum* B2 in semi-defined methanol (200 mM) medium supplemented with sodium acetate at 12.7 mM[21].

To further investigate the physiologic parameters of the strain on methanol-defined medium, product measurements were systematically performed before each subculture, and values are presented for the 10th, 25th, 50th and 75th generations (Fig. 2.3B). Significant changes in carbon distribution were observed for acetate production yield between generations, with a significant yield increase of approximately 80% observed between the 10th and 50th generations, from 45 ± 4 to 77 ± 5 mC acetate/100 mC MetOH. The mean acetate yield decreased by approximately 20% between the 50th and 75th generations, reaching 50.5 ± 4.5. In contrast, the butyrate production yield showed a significant decrease of approximately 40% between the 10th and 50th generations from 67 ± 2 to 39.5 ± 0.5 mC butyrate/100 mC MetOH. For the 75th generation, the mean butyrate

yield showed an increase to 55.5 ± 7.5 mC butyrate/100 mC MetOH, approximately 40% that for to the 50th generation. The carbon flux toward acetate tended to increase from the 10th to 50th generation, while the reverse phenomenon was observed for butyrate production. Carbon from methanol tended to have an equimolar distribution between acetate and butyrate for the 75th generation. Globally, the ALE did not drastically affect the carbon balance between the two products. Carbon incorporation into biomass rapidly increased from the 10th generation (7.5 ± 0.5) to reach 28 ± 1 mC biomass/100 mC MetOH for the 75th generation.

According to the statistical analysis method applied, no significant change was noticed over generations for the $CO_2$ consumed, which is a necessary cosubstrate for methylotrophic growth of *E. limosum* B2 [10]. In general, more carbon from the substrate was dedicated to biomass production through ALE, while the $CO_2$ consumption rate remained constant. The evolved population on synthetic methanol medium tended to have an equimolar carbon distribution from methanol and $CO_2$ to produce acetate and butyrate.

### 2.4.6. Isolation of Individual Clones Growing on Methanol Mineral Medium

The next step of the study was to isolate a strain growing on a methanol-defined medium in the absence of cysteine, i.e.**,** a methanol mineral medium (MMM). To obtain a strain able to grow on a defined medium without cysteine and with methanol as substrate, cells from the advanced evolved population (75th generation) were first spread on a solid glucose mineral medium (SGMM) as no colonies were obtained when spread on a solid MMM (SMMM) supplemented or not with cysteine. This phenomenon was probably due to $CO_2$ accessibility issues (although cultures were performed in an anaerobic chamber with 5% $CO_2$) along with possible methanol evaporation. After growth on SGMM, three colonies were selected and cultured in MMM medium at 37 °C. However, after three subcultures, clone growth declined progressively, presumably from cysteine deficiency. Although *E. limosum* B2 was not documented as auxotrophic for cysteine, alternative sources of sulphur were tested to overcome this presumed auxotrophy[153]. Sodium dithionite, sodium metabisulfite and sodium thiosulfate were added separately to the MMM at 1, 2 and 4 mM. Clones grew on all sulphur sources, with a better maximal biomass observed at 4 mM sodium dithionite. Although sodium dithionite slightly improved growth performance compared to sodium thiosulfate at 4 mM, this last sulphur source was chosen for two reasons: (i) it allows similar growth performance compared to a methanol-defined medium with cysteine and, (ii) sodium dithionite is also a reducing agent. Therefore, we chose to dissociate the source of

sulphur from the source of the reducing agent (titanium citrate at 2 mM) to obtain more reproducible fermentation data.

### 2.4.7. Growth Profile of Adapted Clones in Liquid MMM

Growth profiles of the three isolated clones were determined in liquid MMM supplemented with sodium thiosulfate. Optical density monitoring of the three clones showed similar growth profiles, with a maximum OD of 3.69 ± 0.43 for clone 1, 4.03 ± 0.08 for clone 2 and 3.44 ± 0.37 for clone 3. The maximum growth rate was 0.054 ± 0.008 $h^{-1}$ for clone 1, 0.060 ± 0.003 $h^{-1}$ for clone 2 and 0.047 ± 0.004 $h^{-1}$ for clone 3 (Fig. 2.4A). Globally, clone 2 showed the best growth performance compared to the two others.

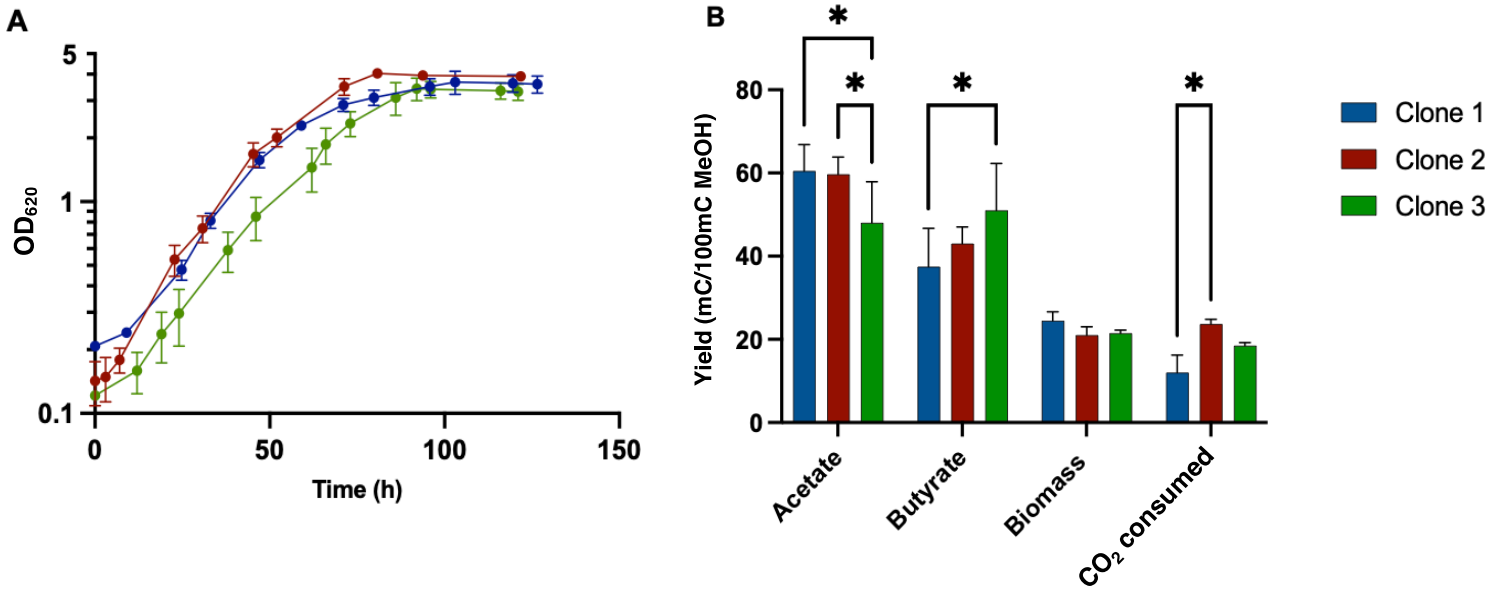



**Figure 2.4 :** Physiological study of isolated clones from evolved populations growing on YE and cysteine-free synthetic medium with 200 mM methanol. (**A**) Growth profile monitoring by measuring the optical density (620 nm). (**B**) Carbon balance from methanol to products, biomass and $CO_2$ consumed. Cultures were performed in triplicate and error bars indicates standard deviation. Asterisk indicates that the results were significantly different from the control according to the one-sided Student's *t*-test: *: *p* value ≤ 0.05.

Analysis of the carbon balance for the three clones showed that the acetate yield of clone 3, with a value of 48 ± 6, was significantly lower than those of clones 1 and 2, with values of 60.5 ± 4.5 and 59.7 ± 4.16 mC acetate/100 mC MetOH, respectively (Figure 4B). The butyrate yield of clone 1, with a value of 37.5 ± 6.5, was significantly lower than those of Clones 2 and 3, with values of 43 ± 4 and 51 ± 8 mC butyrate/100 mC MetOH, respectively. No significant difference was

observed among the clones for biomass yield. However, $CO_2$ consumption was significantly higher for clone 2, with a value of 23.27 ± 1.15 mC $CO_2$ consumed/100 mC MetOH, compared to clone 1 and clone 3, with values of 12 ± 3 mC and 18.5 ± 0.5 $CO_2$ consumed/100 mC MetOH, respectively.

Culture of the wild-type strain showed no significant growth on MMM with sodium thiosulfate addition, indicating that genetic modification/regulation could be responsible for the ability of the strain to grow on MMM. Whole-genome sequencing of the 10th, 25th, 50th and 75th generations from the evolved population and the 3 isolated clones was performed to potentially identify mutated genes responsible for (i) the improved growth on MMM plus cysteine during ALE and (ii) the potential ability of the isolated clones to grow on MMM supplemented with thiosulfate.

### 2.4.8. Mutation Profiles of the ALE Lineage by whole Genome Resequencing

To better understand the underlying mechanisms of growth improvement on MMM supplemented with cysteine, the genome sequences of the evolved populations after the 10th, 25th, 50th, and 75th generations were determined. Furthermore, the genomes of the three selected clones growing on liquid MMM supplemented with thiosulfate were also sequenced (Tab. 2.3).

**Table 2.3.** List of mutations observed for the 10th, 25th, 50th, and 75th generations and Clones 1, 2 and 3 during ALE against the WT strain of *E. limosum* B2. Blue indicates a high rate of mutation among the considered population, while red indicates a low rate of mutation. The numbers represent the percentage of the population carrying genetic changes compared to the WT strain.

| Locus Tag *E. limosum* B2 | Position | WT | Variant | Type | 10th | 25th | 50th | 75th | C1 | C2 | C3 | Gene | Gene in 5′ of Intergenic Region | Locus tag 5′ Gene Intergenic Region | Gene in 3′ of Intergenic Region | Locus tag 3′ Gene Intergenic Region |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| M5595_20825 | 4406858 | - | T | INS | 100 | 100 | 100 | 100 | 100 | 100 | 100 | Hypothetical protein | - | - | - | - |
| M5595_20830 | 4406927 | - | T | INS | 100 | 100 | 100 | 100 | 100 | 100 | | Hypothetical protein | - | - | - | - |
| M5595_00020 | 4683 | G | A | SNP | 100 | | 100 | 100 | 100 | | 100 | Hypothetical protein | - | - | - | - |
| M5595_04930 | 1017540 | C | A | SNP | 100 | | | | | 100 | 100 | Hypothetical protein | - | - | - | - |
| M5595_17525 | 3682586 | A | C | SNP | 36.2 | | 22 | | | | | Hypothetical protein | - | - | - | - |
| - | 1193634 | A | - | DEL | 58.9 | | | | | | | Intergenic region | Ketol-acid reductoisomerase (3′ -> 5′) | | Alpha-amylase family glycosyl hydrolase (3′ -> 5′) | |
| M5595_16535 | 3470810 | G | A | SNP | 100 | | | | | | | Sugar binding transcriptionnal regulator | - | - | - | - |
| M5595_12615 | 2631029 | T | C | SNP | 100 | | | | | | | HAD-IA family hydrolase | - | - | - | - |
| M5595_17525 | 3682598 | C | T | SNP | 31.8 | | | | | | | Hypothetical protein | - | - | - | - |
| - | 4262268 | C | - | DEL | 100 | | | | | | | Intergenic region | Threonine synthase (3′ -> 5′) | M5595_20065 | Homoserine dehydrogenase (5′ -> 3′) | M5595_20070 |
| M5595_06115 | 1274807 | T | C | SNP | 100 | | | | | | | Butyryl-CoA:acetate CoA-transferase | - | - | - | - |
| M5595_11550 | 2401587 | C | T | SNP | | 75.5 | | | | | | Prepilin peptidase | - | - | - | - |
| M5595_08805 | 1810120 | T | - | DEL | | | 95.9 | | | | | Hypothetical protein | - | - | - | - |
| - | 2743930 | A | - | DEL | | | 43.4 | | | | | Intergenic region | Helix-turn-helix transcriptional regulator (5′ -> 3′) | M5595_13160 | Lysozyme family protein (5′ -> 3′) | M5595_13165 |
| M5595_15025 | 3141664 | AT | - | DEL | | | 37 | | | | | Arsenate reductase family protein | - | - | - | - |
| M5595_10115 | 2083561 | A | - | DEL | | | 62.1 | | | | | Glycosyltransferase family 2 protein | - | - | - | - |
| - | 3378354 | T | C | SNP | | | | 70.2 | | | | Intergenic region | Rrf2 family transcriptional regulator (3′ -> 5′) | M5595_16095 | FtsX-like permease family protein (5′ -> 3′) | M5595_16100 |
| - | 2164707 | T | G | SNP | | | | 70.2 | | | | Intergenic region (riboswitch) | Leucine-rich repeat domain-containing protein (5′ -> 3′) | M5595_10385 | Energy-coupling factor ABC transporter permease (3′ -> 5′) | M5595_10390 |
| M5595_09145 | 1871776 | AACTG | - | DEL | | | | 25 | | | | Cysteine synthase A | - | - | - | - |
| - | 4051265 | T | - | DEL | | | | 88.2 | | | | Intergenic region | HPr family phosphocarrier protein (5′ -> 3′) | M5595_19190 | Aminodeoxychorismate/anthranilate synthase component II (5′ -> 3′) | M5595_19195 |
| M5595_03625 | 755658 | C | A | SNP | | | | 35.2 | | | | Peptidoglycan DD-metalloendopeptidase | - | - | - | - |
| - | 825607 | T | - | DEL | | | | 46.7 | | | | Intergenic region | PucR family transcriptional regulator (5′ -> 3′) | M5595_03980 | DUF917 domain-containing protein (5′ -> 3′) | M5595_03985 |
| M5595_00050 | 9361 | T | - | DEL | | | | 93.3 | | | | Hypothetical protein | - | - | - | - |
| - | 370141 | A | - | DEL | | | | 50 | 100 | 100 | | Intergenic region | ParB/RepB/Spo0J family partition protein (5′ -> 3′) | M5595_01730 | Hypothetical protein (5′ -> 3′) | M5595_01735 |
| M5595_05545 | 1145373 | T | C | SNP | | | | | | 100 | | Nucleotide sugar dehydrogenase | - | - | - | - |
| M5595_10115 | 2082950 | A | - | DEL | | | | | | | 100 | Glycosyltransferase family 2 protein | - | - | - | - |
| - | 3881186 | T | C | SNP | | | | 25 | | | | Intergenic region | tRNA-Phe (5′ -> 3′) | M5595_18430 | Cobalamin-dependent protein (5′ -> 3′) | M5595_18435 |
| M5595_00360 | 73656 | - | T | INS | | | | | | | 100 | Restriction endonuclease subunit S | - | - | - | - |

A total of 28 genetic modifications, including two conserved mutations, were identified by comparison with the wild-type strain sequence. Surprisingly, no mutation occurred in genes directly involved in methanol metabolism or the WLP. Two conserved one-base insertions spaced by only 69 bp were found in all generations and clones (except for one insertion that was not observed in clone 3) at the end and the beginning of two adjacent genes (coding for hypothetical proteins) and part of a large prophage operon. Such conserved mutations over generations suggest a response by genomic adaptation of the strain to the restrictive growth conditions on methanol-defined medium without YE. However, understanding the role of insertions in genes coding for hypothetical proteins remains difficult. An ALE experiment performed to improve the growth of the *Sporomusa ovata* acetogen on a liquid methanol mineral medium also showed similar insertions in genes belonging to a prophage operon[52]. In addition to the mutations in genes from a prophage operon, the structure of type I RM encoding genes of the three adapted clones was different compared to the WT strain (Fig. 2.5).

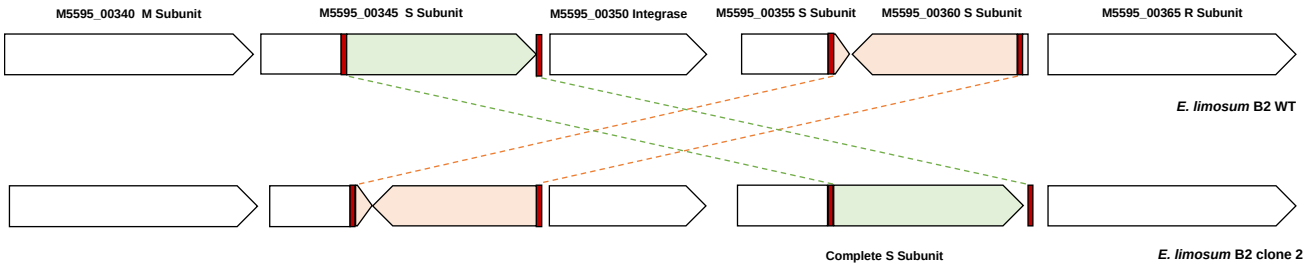



**Figure 2.5 :** Restriction-modification system type I organization in *E. limosum* B2 WT and clone 2. Repeated sequences are highlighted in red while homologous recombination is represented by interchange of green (714 bp) and orange (642 bp) genomic parts. The organization of this region in clones 1 and 3 was the same as clone 2.

The system is composed of one R subunit, one M subunit, one integrase and three S subunits. Among the S subunits annotated, two subunits seem truncated (M5595_00355 and M5595_00360), while one was complete (M5595_00345). We found four repeated sequences, including two repeated sequences in the complete S subunit and two other sequences in truncated S subunits in the WT and adapted strains. Interestingly, we observed a homologous recombination event without mutation in these S subunits between the adapted and WT strains, giving a complete S subunit in the truncated S subunit region and vice versa. The apparent complete S subunit encoding the gene sequence was 1149 bp for the WT strain versus 1236 bp for the clones 1, 2 and 3, indicating a different protein sequence. After the alignment of protein sequences of both S subunits of WT and adapted strains, the structure appeared to be different at the C-terminal side of the protein (Fig. 2.6). Furthermore, the search for conserved domains

indicated two target recognition domains (TRDs), the first located in the 101-192 residue interval, which corresponds to the modified region, and the second in the 236-396 residue interval. The sequence analysis of the S subunit from the adapted strains strongly suggests the recognition of a different DNA sequence. This was confirmed by the determination of the methylome of the clone 2 (Tab. 2.2) which non-only presents a different recognition sequence for the type I RM (CNNTAYNNNNGTG instead of CNNTAYNNNNNTCC) but also a much higher proportion of CAAAAA $^{m6}$R methylated motifs (83.7 versus 37.3%).

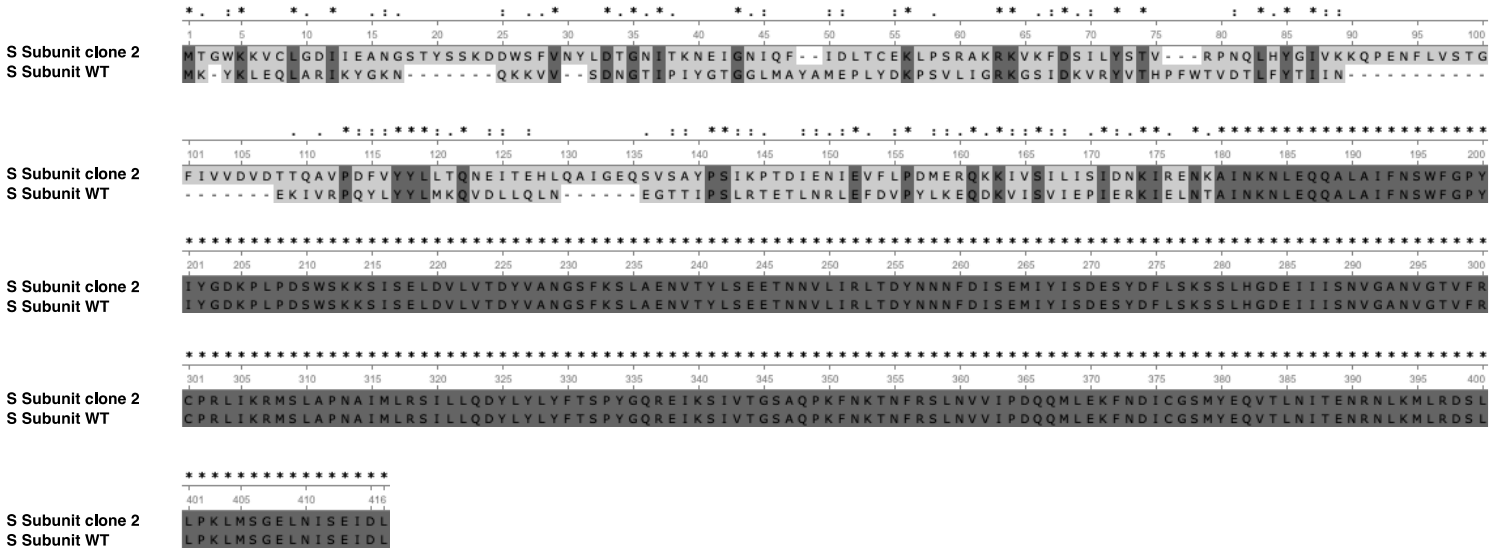



**Figure 2.6.** Protein sequence alignment of the apparent complete S subunit of RM type I system for WT and clone 2 strains. * indicates conserved aminoacids. Clones 1, 2 and 3 produce the same S subunit of RM type I system.

Eleven mutations were detected in the 10th generation compared to the WT strain, with eight mutations at 100% frequency and three other mutations below 58.9% frequency. The only SNP detected in a gene coding for a metabolic enzyme was in the gene encoding the butyryl-CoA:acetate CoA transferase, but this mutation was not detected in the next generations, indicating that it was not needed to improve growth on methanol. Globally, the mutations found in early generations declined for more advanced evolved populations. Furthermore, the three isolated clones showed only a few mutations compared to the WT strain and shared no mutation in common except the ones in the prophage operon and the reorganization of the S subunit encoding genes of the type I RM.

Whole genome sequencing for the 10th, 25th, 50th, and 75th generations and isolated clones during the ALE showed no evidence of genetic modification in genes involved in methanol oxidation or in genes coding for the WLP and enzymes for energy conservation. However, the

WT strain could grow in liquid methanol mineral medium without YE, while the evolved clone 2 could grow if sodium thiosulfate was added. To characterize the effect of the modified methylome on gene expression and also better understand the adaptation mechanisms on synthetic medium, the complete proteome profile was performed for WT and clone 2 strains.

### 2.4.9. Proteomic Analysis of the WT and Evolved Clone 2

#### *2.4.9.1. Global Analysis of the Proteomes*

The total proteome of *E. limosum* B2 was compared between the WT strain and clone 2 on MMM supplemented with 0.5 g/L YE to sustain the growth of the WT strain, as it failed to grow on MMM supplemented with sulphur as the only mineral source. The optimal cell density remained similar for both strains, with an $OD_{620}$ of 4.0. However, the specific growth rate of clone 2 reached 0.084 ± 0.005 $h^{-1,}$ compared to 0.035 ± 0.001 $h^{-1}$ for the WT strain, improving the maximal growth rate 2.4-fold for the adapted strain compared to the WT strain (Fig. 2.7). Although the MMM was supplemented with YE, the clone 2 strain showed a significant improvement in growth compared to the native strain.

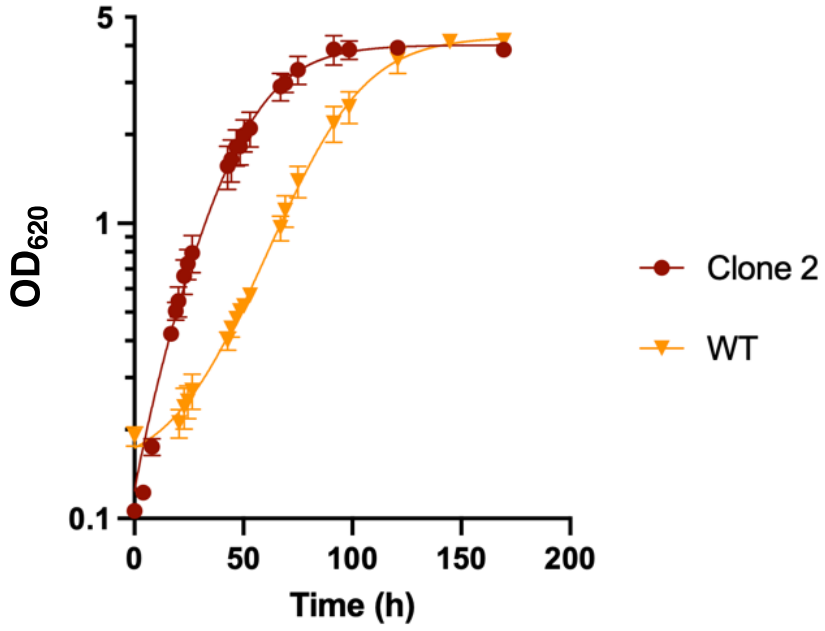
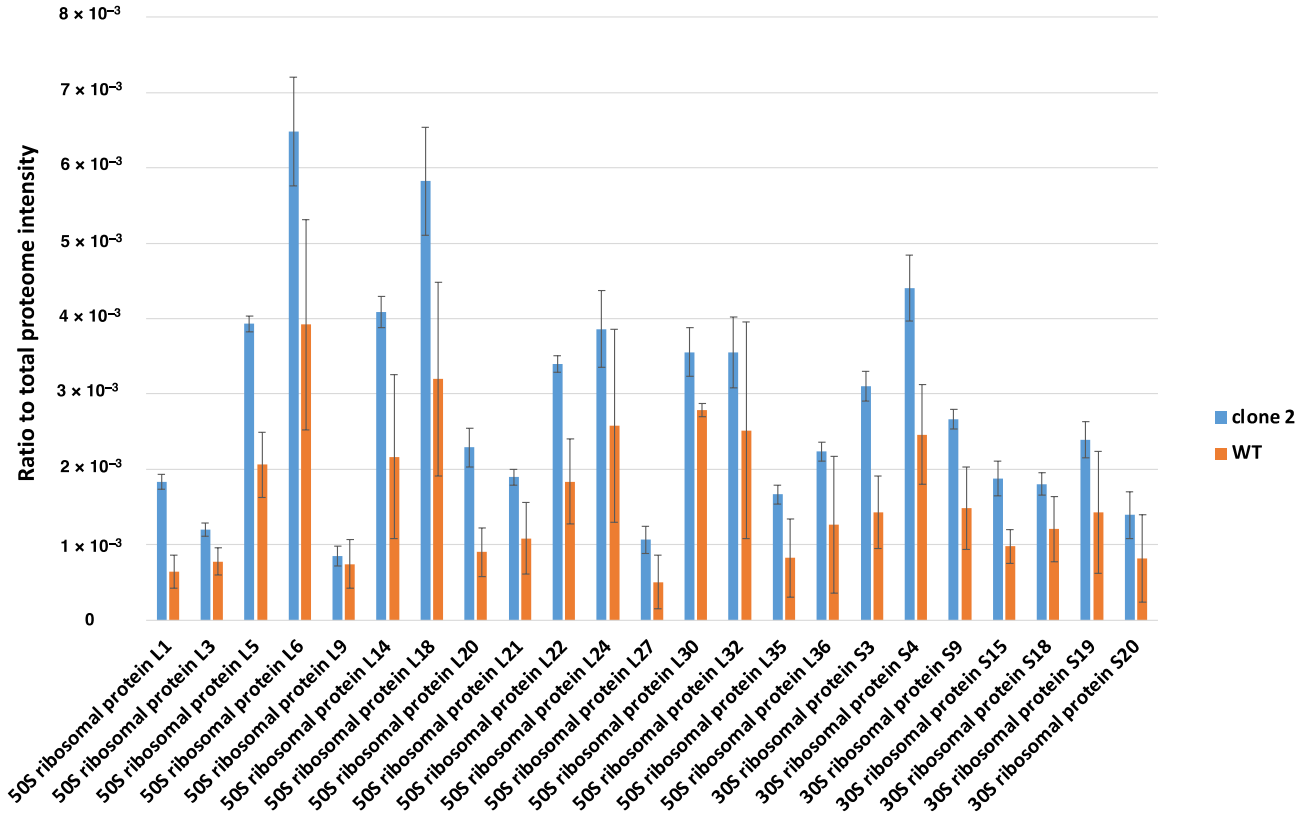



**Figure 2.7 :** Growth profile of the WT and clone 2 strains in methanol synthetic medium supplemented with 0.5 $g.L^{-1}$ YE. Cultures were performed in triplicate and error bars indicates standard deviation.

To evaluate the efficiency of total proteome extraction, the abundance of ribosomal proteins was analysed as a representative of cytosolic proteins, while the extraction of membrane-bound proteins was assessed by examining the proteins of the ATP synthase complex. A total of 46 ribosomal proteins out of 51 were identified for both conditions, indicating efficient extraction using the SPEED method with an average coefficient of variation of 11.5% for the clone 2 strain and 35% for WT (Fig. 2.8A). The variability among cytosolic proteins between replicates for the native strain was superior to that for the adapted strain. This difference might be explained by a different biomass composition between strains. After cell lysis in trifluoroacetic acid (TFA), a larger amount of precipitated compound was observed for the native strain compared to the adapted strain. Based on the ratio of the total ribosomal intensity value to the total proteome intensity value for both conditions, the relative number of ribosomes per cell was 2.57-fold higher in the clone 2 strain than in the native strain. This value was closely related to the specific growth rate difference of both strains, as the clone 2 strain showed a specific growth rate 2.4-fold higher than that of the WT strain.

**A**


Ratio to total proteome intensity
8 × 10⁻³
7 × 10⁻³
6 × 10⁻³
5 × 10⁻³
4 × 10⁻³
3 × 10⁻³
2 × 10⁻³
1 × 10⁻³
0
clone 2
WT
50S ribosomal protein L1
50S ribosomal protein L3
50S ribosomal protein L5
50S ribosomal protein L6
50S ribosomal protein L9
50S ribosomal protein L14
50S ribosomal protein L18
50S ribosomal protein L20
50S ribosomal protein L21
50S ribosomal protein L22
50S ribosomal protein L24
50S ribosomal protein L27
50S ribosomal protein L30
50S ribosomal protein L32
50S ribosomal protein L35
50S ribosomal protein L36
30S ribosomal protein S3
30S ribosomal protein S4
30S ribosomal protein S9
30S ribosomal protein S15
30S ribosomal protein S18
30S ribosomal protein S19
30S ribosomal protein S20

**B**

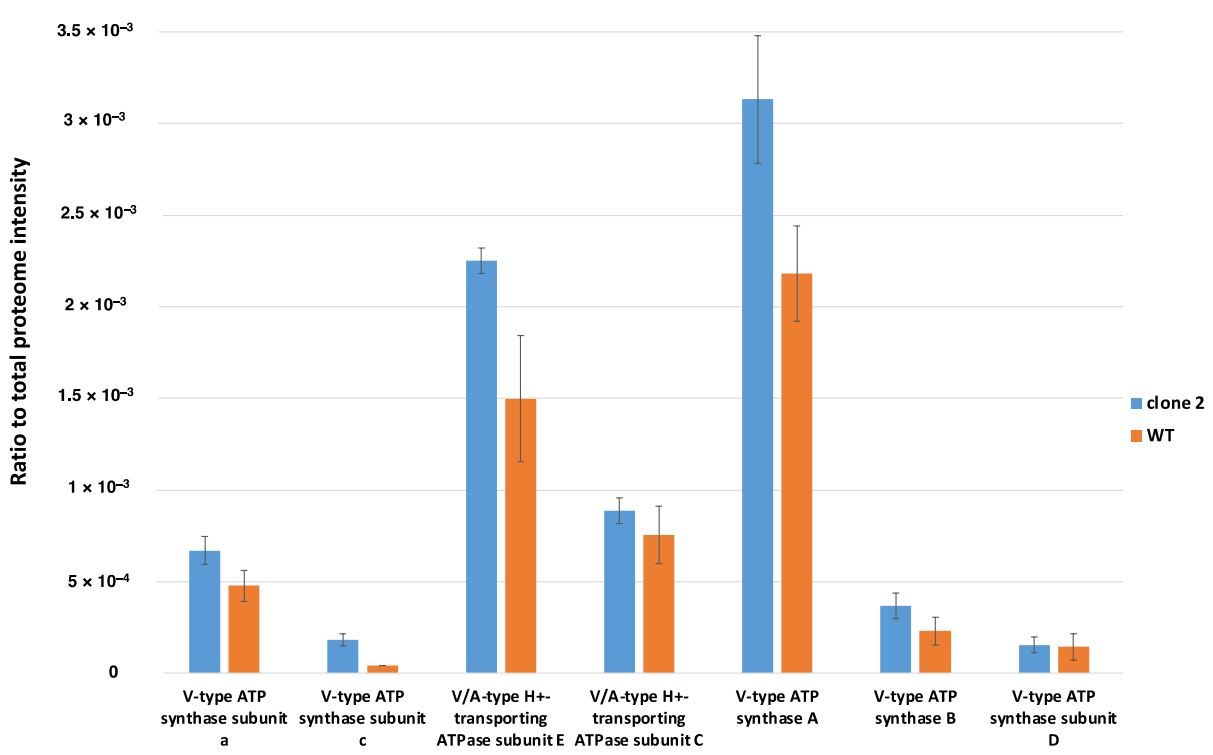

Ratio to total proteome intensity
3.5 × 10⁻³
3 × 10⁻³
2.5 × 10⁻³
2 × 10⁻³
1.5 × 10⁻³
1 × 10⁻³
5 × 10⁻⁴
0
clone 2
WT
V-type ATP synthase subunit a
V-type ATP synthase subunit c
V/A-type H+-transporting ATPase subunit E
V/A-type H+-transporting ATPase subunit C
V-type ATP synthase A
V-type ATP synthase B
V-type ATP synthase subunit D

**Figure 2.8 :** Assessment of the total proteome extraction. (**A**) Assessment of the total cytosolic protein extraction by comparison of all ribosomal proteins extracted. (**B**) Assessment of total proteome extraction on membrane-bound proteins, the example of ATPase complex. Data expressed in proportion to total proteome intensity. Blue bars correspond to the clone 2 strain, and orange bars correspond to the WT strain.

Regarding the efficiency of the total protein extraction method for membrane-bound proteins, seven out of nine constitutive proteins of the ATP synthase complex were identified. The ATP synthase in *E. callanderi* KIST 612 was identified to be a unique combination of a $Na^+$ $A_1A_0$ ATP synthase with a V-type *c* subunit organized in an operon of nine genes[154]. A similar operon was found in the genome of *E. limosum* B2, and seven out of nine constitutive proteins of the ATPase operon were identified by mass spectrometry (Fig. 2.8B). The ATPase complex is organized as an $A_1A_0$ ATP synthase with a catalytic head constituted by A and B subunits, a central stalk comprising the cytoplasmic subunits C, D and F and a rotor located in the membrane including eight to ten *c* and *a* subunits that are stoichiometrically equivalent. In addition, A-type ATP synthases contain subunits H and E[155]. The precise stoichiometry for the ATP synthase of *E. limosum* B2 is currently unknown and requires further investigation using imaging technology to determine the structure of the complex[156]. For example, the stoichiometry of *Pyrococcus furiosus* was established with the following formula: $A_3B_3CDE_2FH_2ac_{10}$[157]. Based on the previous stochiometric formula, the intensity value measured for the different subunits of the ATP synthase complex did not match. A clear difference in the proportion of total proteome intensity was observed between the *a* and *c* subunits, the C and D subunits, and the A and B subunits. The extraction method applied to the samples did not allow total membrane-bound protein extraction but allowed the identification of a sufficient proportion of membranous proteins, as numerous transporters were identified through this extraction method coupled to mass spectrometry.

By combining replicates of both strain cultures, a total of 1256 different proteins were detected, with 1197 proteins identified for the clone 2 strain and 1015 proteins for the WT. A difference of 182 supplementary proteins was observed for the clone 2 strain, suggesting that the mobilization of higher protein diversity led to better growth performance on methanol synthetic medium. The complete list of differentially produced proteins is available in Table S1.

The distribution of differentially produced proteins between the clone 2 strain and native strain (One-sided Student's *t*-test, $p$ value < 0.05) showed over- and underproduced proteins with

more than or less than 1 $Log_2$ fold change. The global proteome showed substantial differences between both conditions, with a total of 91 significantly differentially produced proteins with a fold change (FC) value > 2 or < -2 and a *p* value < 0.05. A total of 52 proteins were overproduced, while 39 proteins were underproduced in the clone 2 strain compared to the WT strain (Fig. 2.9). Proteins from various metabolic functions and pathways were identified as differentially expressed between the conditions, suggesting metabolic flux reorganization between the two strains.

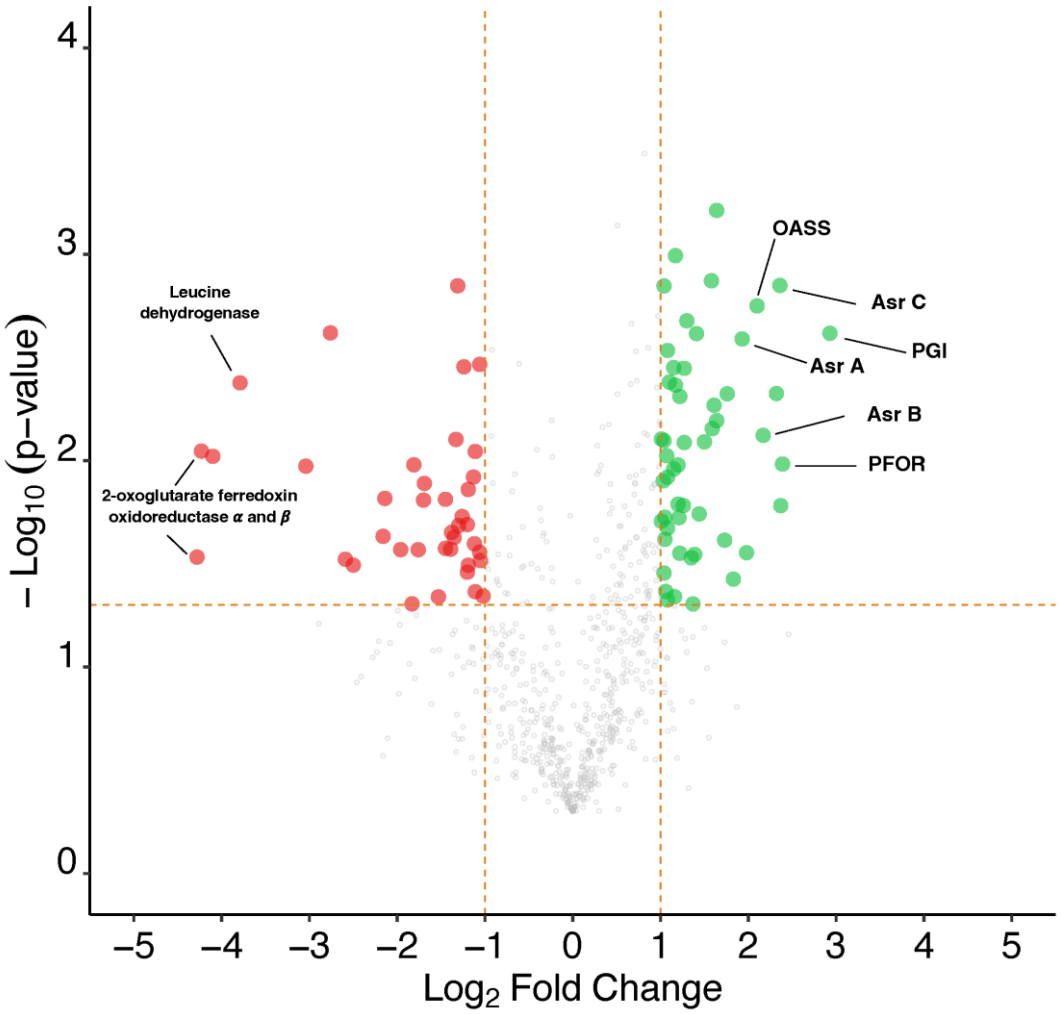



**Figure 2.9.** Volcano plot representing the results of total proteome analysis between the clone 2 strain against the WT strain. Green dots show significantly overproduced proteins, while red dots show significantly underproduced proteins. Horizontal dotted line represents the applied significance threshold of a one-sided student's *t*-test *p* value of 0.05. Vertical dotted lines represent the applied threshold of a $log_2$ fold change of >1 and < -1, corresponding to an absolute fold change > 2 or < −2. Asr, anaerobic sulphite reductase; OASS, O-acetyl-homoserine sulphhydrylase; PFOR, pyruvate ferredoxin oxidoreductase; PGI, glucose-6-phosphate isomerase.

*2.4.9.2. Analysis of the Methanol Oxidation and Wood–Ljungdahl Pathways*

Methanol oxidation has been well characterized in acetogens with evidence of reverse use of the methyl branch of WLP in *E. callanderi* KIST612, a closely related species, to sustain reducing equivalent equilibrium[70]. Additionally, energy production and reducing equivalent balance were reported to be ensured by the $Na^+$-dependent Rnf complex associated with the ATP synthase complex, creating a proton motive force by importing $Na^+$ ions to oxidize NADH and reduce Fd [20,70]. Before entering the WLP, methanol is oxidized and coupled to a tetrahydrofolate (THF) molecule by a methanol-specific methyltransferase complex composed of three proteins containing two catalytic domains[158]. Methanol corrinoid methyltransferase (MTI) transfers the methyl group of methanol to the central cobalt atom of corrinoid protein (CoP), followed by the transfer of the methyl group to a THF molecule by methyltetrahydrofolate cobalamin methyltransferase (MTII). No mutation occurred in the central metabolism coupled to the energy conservation system, and as expected, the proteomic analysis revealed a high abundance of proteins involved in methanol oxidation and the WLP for both conditions, but although the level of all these proteins was higher in the evolved clone, only two were above the 2-fold difference that we set as a threshold (Fig. 2.10). Carbon monoxide dehydrogenase (CooS) was significantly more abundant in the adapted strain, showing an FC value of 2.22, as well as the thiolase, FC value of 2.08, involved in butyrogenesis. Globally, proteins of the CODH/ACS complex in the carbonyl branch, which are involved in $CO_2$ fixation into acetyl-CoA, and proteins of the butyrogenesis pathway were slightly overproduced compared to proteins involved in the methyl branch and acetogenesis.

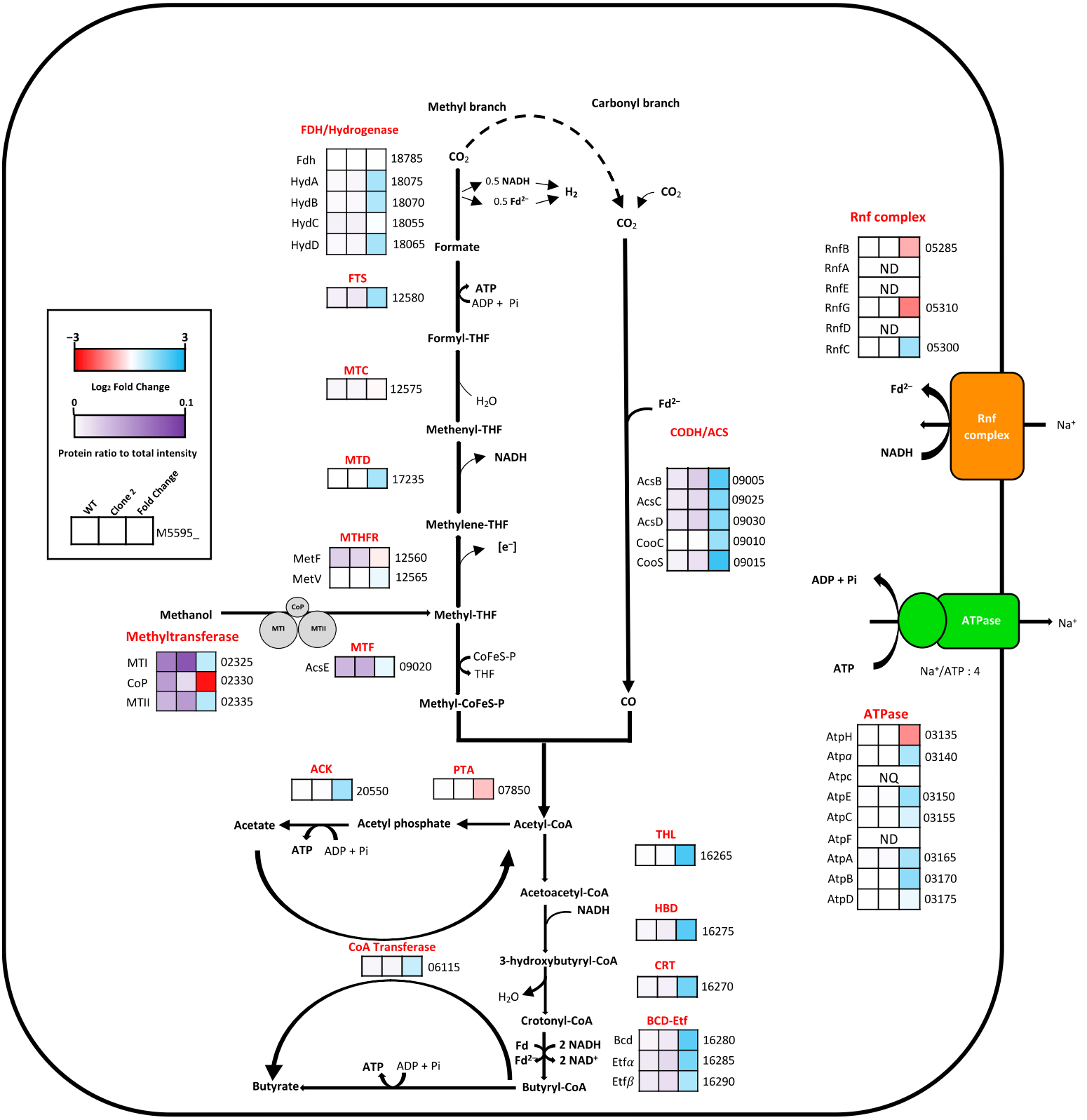



**Figure 2.10 :** Comparative enzyme abundance in the central carbon metabolism between the WT and the adapted strain clone 2 in methanol synthetic medium. The first and second boxes

correspond to the relative protein abundance expressed as a proportion of total intensity value of the WT and clone 2 strains, respectively. The third box corresponds to the $Log_2$ fold change of protein abundance of the clone 2 strain relative to the WT strain. The number indicated at the right of each box corresponds to the *E. limosum* B2 locus tag, starting with M5595; [$e^-$], undefined reducing equivalent; NQ, not quantified; ND, not detected. List of proteins is available in Table S1.

#### *2.4.9.3. Analysis of Gluconeogenesis and Anaplerotic Reactions*

When glucose is not available during autotrophic or methylotrophic conditions, acetyl-CoA, $CO_2$, ATP and reducing power are used in gluconeogenic reactions to produce all the intermediates needed for nucleic acid, amino acid, cell wall and essential cofactor biosynthesis[159]. The first step is catalysed by the pyruvate-ferredoxin oxidoreductase (PFOR) A0A317RSE4 to produce pyruvate [160] from acetyl-CoA, reduced ferredoxin and $CO_2$ (Fig. 2.11).

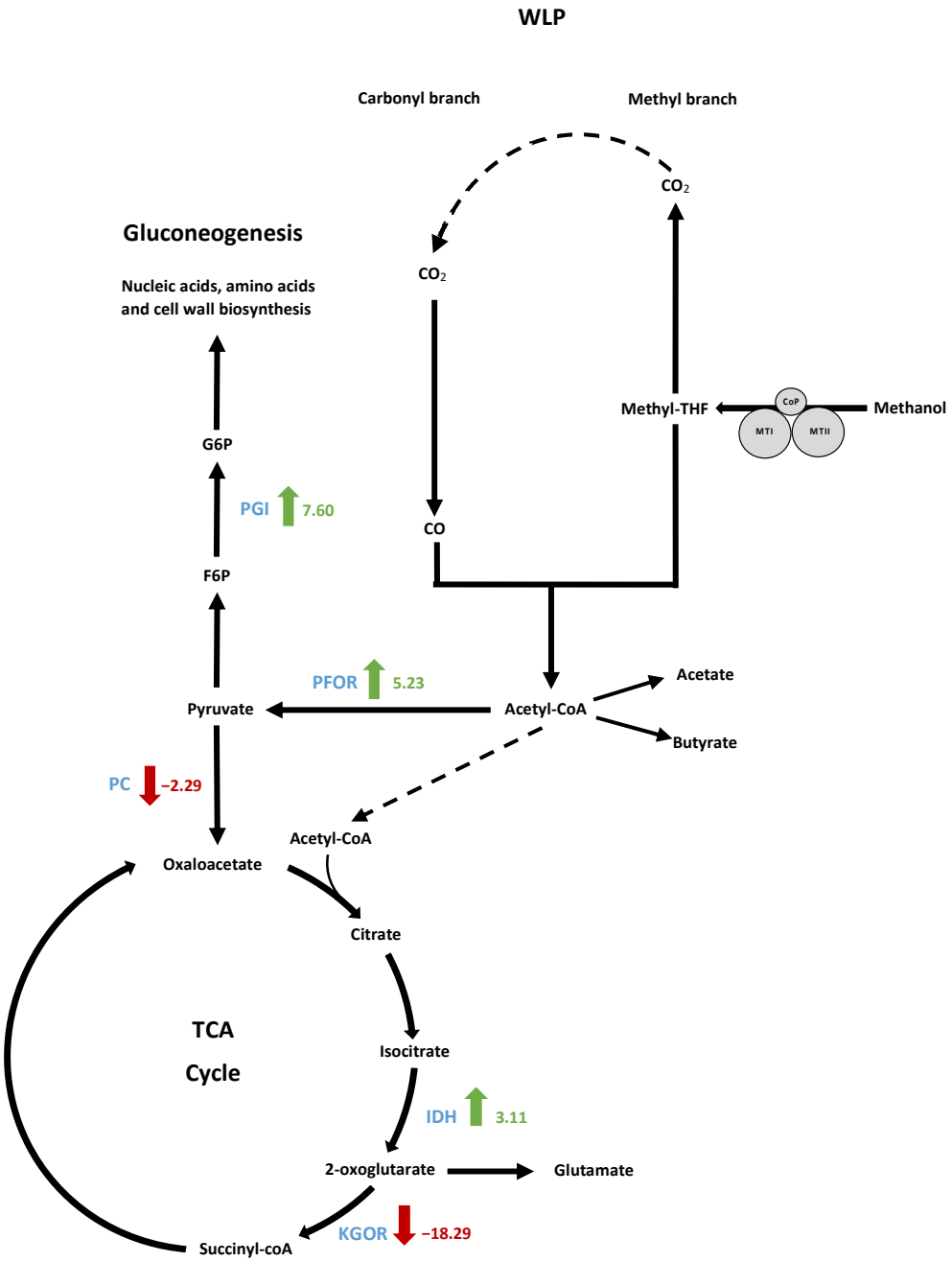



**Figure 2.11 :** Schematic pathway showing significatively differentially produced proteins in clone 2 strain compared to WT strain in gluconeogenesis and tricarboxylic acid cycle alongside WLP. Pathways were simplified to show only proteins of interest. Green arrows indicate protein overproduction and red arrows indicate underproduction in clone 2 strain compared to WT strain. The scores indicate the fold change values. IDH, isocitrate dehydrogenase; KGOR, 2-oxoglutarate ferredoxin oxidoreductase; PC, pyruvate carboxylase; PFOR, pyruvate ferredoxin oxidoreductase; PGI, glucose-6-phosphate isomerase.

Interestingly, substantial overproduction of PFOR was observed, showing an FC value of 5.23, and this was among the most abundant differentially produced proteins detected. Another gluconeogenic enzyme, glucose-6-phosphate isomerase (PGI), was the most overproduced protein, with an FC value of 7.60. Glucose 6-phosphate was demonstrated to represent a key metabolite for peptidoglycan as well as cell wall polysaccharide synthesis. Such upregulation of both the PFOR- and PGI-encoding genes must be a solution developed by the evolved strain to improve its growth capacity in synthetic medium. The carbon recovered in biomass for the clone 2 strain represented approximately 20% of the total carbon metabolized during methylotrophic fermentation against approximately 6.5% for the 10th generation population (Fig. 2.4 and 2.5).

The pyruvate carboxylase (PC), which catalyses the carboxylation of pyruvate into oxaloacetate, the first metabolic intermediary of the TCA cycle, was underproduced (FC value −2.29). Furthermore, the 2-oxoglutarate ferredoxin oxidoreductase subunits alpha and beta (KGOR) were the most highly underproduced proteins (FC values−17.11 and −19.47, respectively). As isocitrate dehydrogenase was upregulated (FC 3.11), this might allow a high flux of alpha-ketoglutarate production that could further be used for glutamate production.

#### *2.4.9.4. Sulphate Metabolism for the Biosynthesis of Sulphur-Containing Amino Acids*

The clone 2 strain can grow on methanol mineral medium with the addition of sodium thiosulfate instead of cysteine, which is required for the WT strain. Interestingly, anaerobic sulphite reductase subunits A, B and C (A0A317RV53, A0A317RVB5 and A0A317RTV7) were among the most statistically overproduced proteins in the adapted clone 2, showing an average FC value of 4.47. In mycobacteria, anaerobic sulphite reductase has previously been demonstrated to be essential for growth on sulphate or sulphite as the sole source of sulphur[161]. Anaerobic sulphite reductase catalyses sulphide ($S^{2-}$) production from sulphite ($SO_3^{2-}$). Cysteine is then synthesized from sulphide and serine by a two-step enzymatic pathway[162]. First, serine is acetylated by a serine transacetylase (CysE) to form O-acetylserine, and second, cysteine is produced from O-acetylserine and sulphide via a reaction catalysed by a cysteine synthase (CysK). Furthermore, CysK (A0A317RWJ0) was overexpressed in the clone 2 strain with an FC value of 1.87. There remains some uncertainty regarding the sulphite supply by thiosulfate. Sulphite can be produced either by the reduction of sulphate ($SO_4^{2-}$) or by the reduction of thiosulfate by a thiosulfate reductase[163]. However, no thiosulfate reductase-encoding gene was

found in the genome of *E. limosum* B2 or among the detected proteins, and $SO_4^{2-}$ from iron sulphate did not sustain the growth of the strain in synthetic medium. Where does the sulphite come from if sulphate is not the sulphur source? Evidence of cysteine production with thiosulfate was reported in *E. coli*[164]: L-serine can be acetylated to O-acetylserine by a serine acetyltransferase, and the product can be further converted with the addition of thiosulfate to S-sulphocysteine by an O-acetyl-homoserine sulphhydrylase OASS (A0A317RPB8). Then, L-cysteine can be synthesized from S-sulphocysteine by thioredoxin/glutaredoxin, liberating sulphite, which would be recycled for further cysteine production. The OASS was highly overproduced in the adapted strain (FC 4.30), suggesting the utilization of this pathway to produce cysteine in the adapted clone 2 (Fig. 2.12).

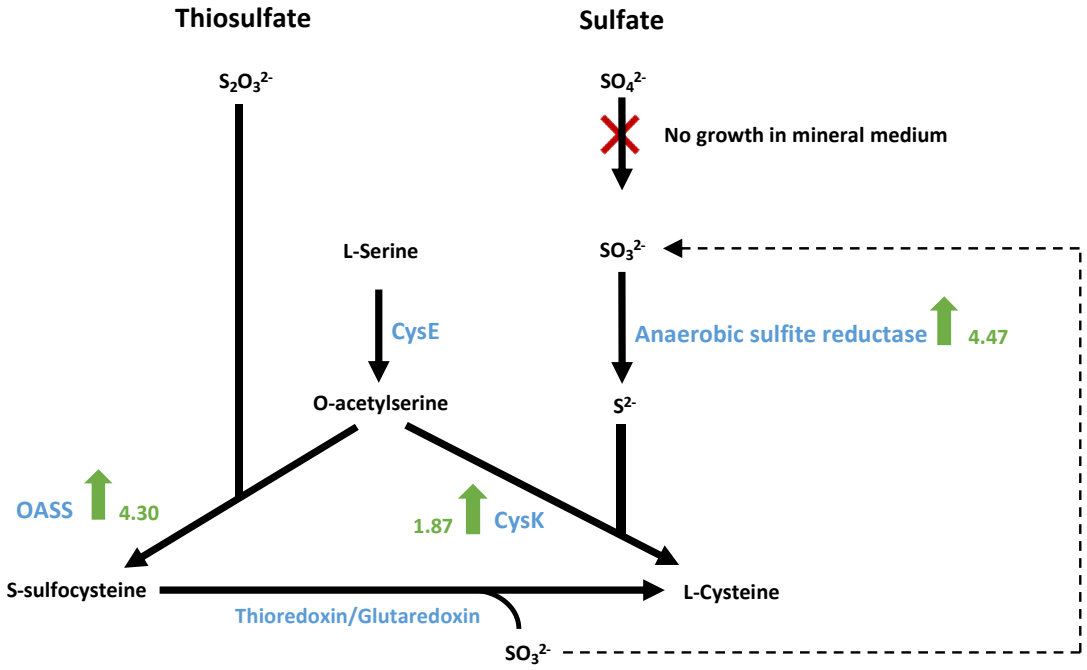



**Figure 2.12 :** Theoretic pathway for sulphur assimilation and cysteine biosynthesis in *E. limosum* B2 in mineral medium. Green arrows indicate protein overproduction in clone 2 strain compared to WT. The scores indicate the fold change values. CysE, serine transacetylase; CysK, cysteine synthase; OASS, O-acetyl-homoserine.

Interestingly, 5-methyltetrahydrofolate-homocysteine methyltransferase, also known as methionine synthase, was the second strongest underproduced protein (FC −3.56) among amino acid biosynthesis-related proteins. This enzyme was shown to have methionine synthase activity[165]. Methionine synthase is involved in the last step of methionine biosynthesis, using homocysteine and methyl-cobalamin as cofactors. Decreased production of methionine synthase could suggest the preservation of homocysteine to produce cysteine via the

transsulfuration pathway[166]. In addition to cysteine production through L-serine transformation and thiosulfate utilization, this pathway was probably used by the clone 2 strain to produce cysteine. This function likely emerged during the evolution of the strain, as cystathionine gamma-synthase (A0A317RYT5) and cystathionine beta-lyase (A0A317RVG5), two key enzymes of the transsulfuration pathway, were only detected in the clone 2 strain.

#### *2.4.9.5. Other Differentially Produced Proteins*

Because all ALE processes were performed without YE, the evolved strain strongly regulated its metabolism to produce the required amino acids. While proteins involved in amino acid biosynthesis were found to be slightly overproduced, others were strongly underproduced, such as leucine dehydrogenase (A0A1H0LY56), with an FC value of −13.81.

Surprisingly, several enzymes involved in propanediol utilization (*pdu* operon) were strongly underproduced in clone 2, indicating an important function in the native strain. The enzyme BMC domain-containing protein (A0A0U3FC86) which is involved in the bacterial microcompartment for 1,2 propanediol (1,2 PDO) or glycerol metabolism exhibits significant underproduction (FC −6.76). The enclosed proteinaceous structure metabolizes 1,2-propanediol to propanol and propionate or to produce 3-hydroxypropionate from glycerol. Encapsulation is necessary to limit cytosolic exposure to propionaldehyde, a toxic metabolic intermediary[167]. The presence of cytosolic microcompartments was described for *E. maltosivorans* with 1,2 PDO or betaine as carbon sources. Furthermore, BMC structures were also reported on *A. woodii*, a close relation, with 1,2 PDO and 2,3 butanediol as carbon sources, but no BMC was detected with methanol as a carbon source. The presence of proteins from the *pdu* operon at high intensity for the native strain remains enigmatic, as no propanol or 3-hydroxypropionate was detected by HPLC.

## 2.5. Discussion

In this study, the genome of the most studied (from a physiological point of view) strain of *E. limosum* was sequenced. The *E. limosum* B2 genome shows 99.837% sequence identity with that of the *E. limosum* ATCC 8486 strain. Although the two genomes are very similar, they exhibit chromosomal rearrangement in prophage operons, insertion sequences and 5 colinear blocks (LCBs) with a reverse complement orientation. Furthermore, several differences were found in genes coding for regulators and an S subunit-encoding gene of a type I RM system.

*Eubacterium limosum* B2 required unknown components in the complex media for growth on methanol[21]. To overcome this problem, we applied an adaptive evolution approach to develop an optimized strain of *E. limosum* B2 capable of growth on a methanol-defined medium without yeast extract. By this approach, the growth rate was enhanced 3.45-fold in 75 generations. From the evolved population, three clones were isolated and grown on methanol mineral medium without cysteine (which was replaced by sodium thiosulfate). Among the three isolated clones, the clone 2 strain showed the best growth characteristics, showing a maximal growth rate 2.72-fold higher than that of the 10th generation population cultured in methanol-defined medium supplemented with cysteine. Genome sequencing of the WT, 10th, 25th, 50th, and 75th generations and three evolved clones showed no specific mutation in central carbon metabolism or conserved mutation over generations except in a large phage operon and in the S subunit encoding genes of a type I RM system. Proteomics data, on the other hand, showed important differences in the proteomic compositions of the WT and clone 2. First, overproduction was observed for the enzymes of both (i) the carbonyl branch of the WLP and (ii) the butyrogenesis pathway. This should lead to an improvement in $CO_2$ fixation ability by both (i) an enhancement in reduced ferredoxin production by the electron bifurcating complex Bcd-Etf in the butyrogenesis pathway and (ii) by an increase in the level if CODH that also requires reduced $Fd^{2-}$ as a source of electrons.

Better growth of clone 2 could also be the result of higher production of enzymes involved in the synthesis of key building blocks of the cells, such as the PFOR involved in pyruvate production from acetyl-CoA, PGI responsible for glucose-6-P production from fructose-6-P or IDH involved in a-keto-glutarate production needed for glutamate production. The ability of clone 2 to grow on a methanol mineral medium in the absence of cysteine could be the result of a combination of higher production of (i) the anaerobic sulphite reductase and (ii) the

enzymes of the transsulphuration pathway allowing cysteine biosynthesis from thiosulfate. The considerable modification of the proteomic profile of the evolved clone 2 was associated to changes of its methylome profile suggesting that it is the result of an epigenetic phenomenon. The rearranged S subunit of the type I RM in the evolved clones, recognize a different motif (CNNTAYNNNNGTG instead of CNNTAYNNNNNTCC) and a higher proportion of these motifs are methylated. Similarly, a much higher proportion of type II motifs (CAAAAA $^{m6}$R) are also methylated (83.7 versus 37.3% in wild type). These modifications in the methylome profile are probably responsible in the large change in gene expression. A similar phenomenon was previously demonstrated in *Bacillus pumilus* when the type I RM-encoding genes were deleted [168]. This study highlighted the metabolic/genetic adaptation of the *E. limosum B2* strain growing on methanol-defined medium and showed the importance of improving our knowledge of cell adaptation mechanisms in a constrained environment. The evolved clone 2 that grows on a mineral medium should allow us to develop and use a genome-scale model of *E. limosum* B2 for both flux analysis and rational metabolic engineering of *E. limosum* B2, as previously performed with *Clostridium acetobutylicum*[169]. Furthermore, to better characterize the metabolism of this strain, it can now be used in chemostat cultures in mineral medium at a constant specific growth rate using different carbon sources to perform a complete systems biology analysis.

# Chapter 3: Characterization of the glucose and methanol metabolism of *E. limosum* B2 by a systems biology approach at the absolute scale

## 3.1. Introduction

The intensive use of fossil fuels led to a massive production amount of greenhouse gases (GHG). Several pieces of evidence show the impact of GHG on climate change and a rise of more than 1.5°C of the atmosphere would have dramatic consequences on biodiversity and cultivable resources [3,170]. Faced with this situation, much attention was focused on finding efficient solutions to develop a virtuous circular bioeconomy. Devices to collect sources of renewable energies such as solar panels or wind turbines are extensively deployed to initiate the energetic transition. To reach net zero carbon emissions, the development of efficient technologies for carbon dioxide capture is paramount.

Some microorganisms showed the ability to fix one carbon element (CO, $CO_2$, or methanol) as the sole source of carbon and energy into multi-carbon branch chemicals, making them an important asset for bioeconomy development. Acetogens belong to the most promising group of natural biocatalysts as they can use the Wood-Ljungdahl Pathway (WLP) known as the most energy-efficient pathway to reduce $CO_2$ into acetyl-CoA, the central metabolic intermediary. Although the acetogen *Clostridium autoethanogenum* is already used as a cell factory for gas fermentation, several technical constraints limit the full potential of such microorganisms notably the gas-to-liquid mass transfer, storage, and transport concerns. Methanol constitutes an interesting alternative to C1 gases and is widely produced throughout the world from biomass, by hydrogenation of $CO_2$ and CO, and from the exhaust of industrial processes[133,171,172]. *Eubacterium limosum* is a promising acetogen able to produce acetate and butyrate from synthesis gases (a mix of CO, $CO_2$, and $H_2$) and methanol. The physiologic characteristics of *E. limosum* B2 on methanol were extensively studied[21,75]. At the genome scale, the autotrophic (CO and $CO_2$ + $H_2$) metabolism of *E. limosum* ATCC8486 was deciphered, revealing transcriptional regulations of the WLP along the energy conservation system and effects of 5'UTR on translation efficiencies[122]. Given the potential of *E. limosum* for the biosynthesis of complex C4 molecules from the C1 methanol feedstock, a deep understanding of its metabolism by a characterization at the genome scale is essential to further improve the actual genome-scale model and perform metabolic engineering.

However, despite the wide usage of last omics technologies, only a few quantitative descriptions were performed to date. Recently, absolute quantification of the total proteome of C. *autoethanogenum* was described on synthesis gases. The results highlighted the total proteome allocation relative to the physiological growth condition and allowed the identification of the biological relevance of isoenzymes by comparison of their abundances[125]. Furthermore, the apparent *in vivo* catalytic flux rate ($k_{app}$) of each enzyme can be calculated once the metabolic flux rates of enzymes and absolute protein number per cell at steady state are known. This parameter constitutes an interesting tool to

identify less-efficient enzymes which can be targeted by metabolic or protein engineering for value-added production enhancement purposes[173].

Previously, we developed a mutant strain of *E. limosum* B2 able to grow on a methanol mineral medium by adaptive laboratory evolution [174]. We used this strain and performed chemostat cultures in methanol and glucose mineral medium, at equal carbon concentration and dilution rate ($\mu = 0.025 h^{-1}$) for both conditions. Based on product concentrations measured at steady state, the carbon balances were calculated, allowing the elaboration of energy conservation models. The determination of the absolute number of proteins per cell for > 1400 proteins using a quantitative protein extraction method allowed us to precisely determine the metabolic flux rates of enzymes in the WLP and the absolute number of proteins per cell. The $k_{app}$ of each enzyme in the WLP was estimated in methanol and glucose conditions. Furthermore, the absolute number of mRNA molecules per cell was calculated using external standards (spike-in mixes) allowing the identification of the strongest promoter for both conditions. Finally, absolute quantification of transcripts and proteins number per cell opened access to the *in vivo* synthesis rate ($k_x$) for each protein allowing the determination of the strongest ribosome binding sites (RBS). Interestingly, our approach which combined flux simulations into the genome-scale model of *E. limosum* B2 with transcriptomic and proteomic data raised questions on the mechanism of the methyltransferase-THF reductase complex in the methyl branch of the WLP, and on the Rnf complex activity during heterotrophy on glucose. Metabolism characterization at the absolute scale allowed the identification of potential bottlenecks for enzymes in the WLP and product synthesis. Furthermore, our results strongly suggest the use of the butyryl-CoA:acetate-Coa transferase for butyrate biosynthesis.

To our knowledge, such an approach provides a unique level of detail on the methylotrophic and heterotrophic metabolism of *E. limosum*.

## 3.2. Materials and methods

### 3.2.1. Bacterial strain and growth condition

The adapted strain selected from a previous study (publication submitted) and growing on a methanol mineral medium was cultured in a chemostat in the following mineral medium: NaCl 0.6 g/L, $NH_4Cl$ 1 g/L, $MgSO_4$-$7H_2O$ 0.12 g/L, $CaCl_2$ 0.08 g/L, $KH_2PO_4$ 1 g/L, vitamin solution 100X (biotin 20 mg/L, pantothenic acid 50 mg/L, lipoic acid 50 mg/L), sodium thiosulfate pentahydrate 4mM, potassium carbonates 1 g/L, titanium citrate 2mM as reducing agent, and trace element solution 100X ($MnCl_2$ 0.1 g/L, $CoCl_2$-$2H_2O$ 0.1 g/L, $ZnCl_2$ 0.1 g/L, $CuCl_2$ 0.02 g/L, $H_3BO_3$ 0.01 g/L, Na molybdate 0.01

g/L, $Na_2SeO_3$ 0.017 g/L, $NiSO_4$-$6H_2O$ 0.026 g/L, $FeSO_4$,$6H_2$ 1 g/L and nitrilotriacetic acid 12.8 g/L. Resazurin was added at $1.10^{-4}$ % (wt/wt) as redox potential indicator. Glucose 6 g/L and methanol 6.4 g/L were the two carbon sources tested, corresponding to 200mM of carbon. Details on chemostat culturing are described in chapter 3.

### 3.2.2. Analytical methods

Culture density was measured by spectrophotometry at 620nm ($OD_{620}$) against a water blank. Substrates and products were measured using high-performance liquid chromatography (Agilent 1200 series, Massy, France) equipped with a refractive index detector and Aminex column HPX 87 H 300 x 7.8 mm. $H_2SO_4$ 0.5mM was used as the mobile phase and elution was performed at 48°C.
The biomass formula $C_4H_7O_2N_{0.6}$ was used to convert biomass values into molar cell carbon concentration (mC)[141]. Acetate and butyrate production yield were calculated in carbon millimolar of product for 100 mM of carbon from methanol (mC/100mC MetOH). Consumption or production of $CO_2$ has stoichiometrically deduced after products and biomass yield determination.

### 3.2.3. Determination of bacterial dry cell weight and cell counting of chemostat cultures

To reach our goal of absolute quantification of the total copy number of proteins and transcripts per cell, it was essential to determine the cell count for each substrate studied in chemostat culture at the steady state. The cell counting was realized with Toma's hemocytometer which consists of a microscope slide with a squared indentation that creates a delimited area where the precise volume is known. The cultures were diluted, and a drop of culture was deposited in the counting chamber. A special glass coverslip was properly positioned on the surface of the counting chamber. The counting surface was divided into large squares, each of them composed of 16 little squares. The volume of a little square was 0.25nL and 4nL for a large square. The cells were observed and counted in at least 3 large squares using a microscope at 400X magnification. The cell count (cell number.$mL^{-1}$.$OD_{620}$) was calculated using the following formula:

$$\text{Cell count} = \frac{Average\ number\ of\ cell\ counted\ per\ large\ square\ \times Dilution\ factor}{\mathrm{OD}^{620} measured} \times\ 2{,}5.10^5$$

**Table 3.1 :** Cell count values for each culture condition performed in steady state chemostat.

| Chemostat culture condition | Cell count (Cells number.mL$^{-1}$.OD$^{620}$) |
|---|---|
| Methanol | 2,55.10$^{8}$ |
| Glucose | 2,58.10$^{8}$ |

The average cell count was measured for methanol and glucose conditions and showed an average value of $2.55 \times 10^{8}$ cells.mL$^{-1}$.OD$^{620}$ (Table 3.1).

The bacterial dry cell weight (DCW) was determined for methanol and glucose conditions at the steady state. The determination of this parameter was essential for biomass value incorporation in the stoichiometric equations. The experiments were realized in triplicate, in 1.5 mL Eppendorf tube, for each condition. The tubes were weighed on a precision scale before the addition of cell cultures. In parallel, three controls consisting of empty tubes were weighed on a precision scale before and after treatment in the oven to take into count the natural loss of weight due to evaporation during the process.

After measuring the optical density of the chemostat cultures, 1ml of culture were added in Eppendorf tubes and were centrifuged at 16 000g for 1 minute and supernatants were discarded. The operation was repeated two other times for a total of 3mL of culture added in each tube. The cell pellets were washed by resuspension in 1mL of mQ water and were centrifuged at 16 000g. The wash step was repeated, then the cell pellets and controls were incubated in the vacuum oven (200mm Hg) heated at 60°C. After 48h, the controls tubes and the samples were rapidly weighed on a precision scale. The cell dry mass was calculated following the formula:

$$\text{DCW} = \frac{Dry\ cell\ weight\ weighed + Tube\ mass\ loss\ (control)}{3 \times OD^{620} measured} \times 10^{3}$$

**Table 3.2 :** Dry cell weight values for each culture condition performed in steady state chemostat.

| Chemostat culture condition | Dry cell weight (g.L$^{-1}$.OD$^{620}$) |
|---|---|
| Methanol | 0.272 |
| Glucose | 0.269 |

The values measured for methanol and glucose conditions were similar (Table 3.2). The dry cell weight of *E. limosum* B2 was estimated by Pascal Loubière (thesis, 1992) at 0.408 $g.L^{-1}.OD^{660}$. This value was determined by membrane filtration at a different wavelength, 660nm. Furthermore, the exact growth conditions where the experiment was performed were not detailed. Depending on the medium and growth phase, whether the cells are in the exponential or stationary growth phase, the dry cell weight of *E. coli* showed important variations[175]. The calculation of biomass in the stochiometric equations for the different chemostat conditions was based on these experimental DCW values. The DCW values were then used to calculate the carbon balance and substrate consumption and acids production flux rates.

### 3.2.4. Genome-scale metabolic network of *E. limosum* and flux simulations

The physiologic profiles determined for both conditions were integrated into the genome-scale metabolic network (GEM) of *E. limosum* (data unpublished) to accurately estimate the metabolic flux rates of enzymes involved in the WLP and to predict the effects of genes expression on the metabolism for a defined phenotype. The construction of the GEM of *E. limosum* was constructed from genes annotated by the NCBI annotation pipeline for *E. limosum* ATCC8486, a highly close genetic neighbor of *E. limosum* B2. Genes were first mapped to biochemicals reactions enzymes available for other highly curated GEM of *Clostridium ljungdahlii* (*i*HN647), *Bacillus subtilis* (*i*YO844), and *E. coli* (*i*ML1515). For the remaining genes unidentified during the previous step, biochemical reactions available in KEGG, BiGG and MetaCyc were prospected to complete the model. To include biomass in the model, the cell composition was first determined by measuring DNA, RNA, and proteins at the exponential phase. The elements constituting the biomass for *C. ljungdahlii* (*i*HN647), were used to define the biomass of *E. limosum*. The final model, named *i*SP652, consists of 937 reactions, 855 metabolites, and 652 genes covering 33.6% of the function predicted by protein-coding genes. Flux simulations were performed by Nicole Percy (University of Nottingham, UK).

### 3.2.5. Flux rate calculation

The substrate and products flux rates from the chemostat cultures were calculated considering the flow rate determined by the formula:

$$d = \mu \times V$$

With d the dilution rate ($mL.h^{-1}$), μ the dilution rate ($h^{-1}$) and V the volume of the chemostat vessel (mL) and the concentrations measured by HPLC. The calculation of flux rate for each substrate and product was determined with the following formula:

$$\text{Flux rate} = \frac{d \times C}{DCW \times OD^{620} measured \times V}$$

With the flux rate (mmol $C.g^{-1}.h^{-1}$), d the dilution rate ($L.h^{-1}$), C the substrate or product concentration in the feed medium (mmol $C.L^{-1}$), the DCW ($g.L^{-1}.OD^{620}$), OD the optical density and V the volume of the chemostat vessel (L). These data were then integrated into the *E. limosum* GEM (*i*SP652) to precisely determine the specific metabolic flux of each enzyme of the WLP.

### 3.2.6. Total protein extraction

The total protein extraction was achieved following the sample preparation by easy extraction and digestion (SPEED) method [143]. Based on trifluoroacatic (TFA) cells lysis, this approach was described as quantitative and allowed the identification of a large variety of proteins from *E. coli* and *Bacillus subtilis*, including membranous proteins[143]. The extraction protocol used in this study was like the protocol followed in the first part of the thesis project.

For proteomic analysis, 1mL of cells were harvested from methanol and glucose chemostat cultures. Samples were immediately pelleted by centrifugation at 16 000g for 30 seconds and supernatants were discarded. Cell pellets were washed by resuspension in 1 mL of phosphate-buffered saline solution and centrifuged at 16 000g for 30 seconds. The supernatant was discarded, and the step was repeated twice. Cell pellets were quickly frozen in liquid nitrogen and stored at -80°C until analysis. Each pellet was resuspended in 150µL of trifluoroacetic acid (TFA, Supelco MS Grade) and warmed at 70°C for 3 minutes and centrifuged at 16000g for 1 minute. 130µL of supernatant was carefully transferred to a fresh 1,5mL Eppendorf tube. 10µL of supernatant was taken for each sample and diluted in TFA 75% to measure protein concentration by Bradford assay against BSA standards. Each sample was then neutralized by 1.5 mL of TRIS 2M, prepared in water LC-MS grade (Supelco) and transferred in glass vials, and freeze-dried. Four biological replicates from both conditions were sent to the John van Geest Cancer Research Center at Nottingham. After resolubilization and purification, samples were injected into a mass spectrometer TripleTOF 6600 instrument (Sciex). Raw data were

treated using the DIA-NN software[176]. Samples were prepared and analyzed by David Boocock (John van Geest Cancer Research Center, Nottingham, UK)

### 3.2.7. Total mRNA isolation, external standards and RNA-Seq

Total RNA extraction was performed for both conditions following the RNAsnap protocol[177]. The detailed protocol including the extraction process and spike-in mixes preparation and concentrations is available in the appendix. Two sets of external standards were added to the total RNA during the extraction process. The first mix was added after cell lysis and consists of 6 different RNA molecules of known sizes and concentrations. The spikes were synthesized by SynbioTechnologies (USA, New Jersey) and diluted in pure formamide to match the target concentrations of the reference study on *E. coli*[178]. A similar quantity of spikes was added to the total RNA for both conditions. Total RNAs were then treated by TURBO Dnase (Thermofisher, USA) and on-column purified using the RNA Clean and concentrator's kit from Zymo (California, USA). A second spike-in, ERCC mix (Thermofisher) was added to total RNA before ribodepletion. ERCC mix concentration was determined based on the reference study for absolute quantification of transcripts molecules per cell in *Xenopus*[179]. The ribodepletion was performed on 500ng of total RNA using a custom-made ribodepletion kit, specifically targeting ribosomal RNAs of *E. limosum* B2, from SiTools (Germany). The RNA concentration measurement was fluorometrically performed using a QubitTM instrument (Thermo Fisher Scientific, Wilmington, DE, USA).

An amount of 10ng of isolated transcripts was used for library preparation using the Ion Total RNA-Seq v2 kit and the 100bp single-end sequencing was performed with the Ion Torrent S5 system equipped with an Ion 540 chip. RNA-seq raw data were normalized and mapped using Torrent_suite 5.12.1.

Ribodepletion, library preparation, and sequencing were performed by the Get-Biopuces platform at TBI (http://get-biopuces.insa-toulouse.fr).

Quality controls were performed by gel electrophoresis on agarose gel 2% and by BioAnalyzer 2100 (Agilent) using a Nano 6000 Nano chip.

### 3.2.8. Determination of the absolute copy number of proteins per cell

One of the main goals of this thesis project was to quantify the total copy number of each protein at the scale of one cell. To perform this, several conditions were required with i) the utilization of a

quantitative method for protein extraction, ii) the dosage of the total protein extracted and iii) knowing the number of cells from which protein extraction was performed. All protein extractions were performed from 1mL of chemostat culture for both conditions. The $OD_{620}$ was systematically measured just before sampling and the cell density value relative to the culture condition allowed the determination of cell number in 1mL of culture.

After analysis on mass spectrometry, the intensity of each protein was determined by software using artificial intelligence called DIA-NN[176]. By the combination of these intensities values for each protein obtained and the information previously cited, the absolute copy number of proteins per cell was calculated following the formula:

$$\text{Absolute copy number of proteins per cell} = A \times \frac{Mp}{Cn} \times \frac{R}{M}$$

With A the Avogadro's number (number of molecules.mol$^{-1}$), Mp the protein mass measured by Bradford assay (g), Cn the cell number, R the ratio of the intensity of an identified protein on the total intensity, and M the protein molecular weight (g.mol$^{-1}$). This formula allowed the determination of the total number of proteins per cell for all the proteins identified.

### 3.2.9. Determination of the absolute copy number of transcripts per cell

The absolute copy number of transcripts per cell was determined for both culture conditions at steady-state, glucose, and methanol. Read counts obtained from RNA-Seq were normalized against the first spike-in mix added by Nick Owens (Institut Pasteur, France).

The regression model was built with the equation following:

$$\log(t) = \beta_I + \beta_C \log c + \beta_L \log(L) + \beta_G I_G$$

With *t* the total spike in fmol, c the normalized read count, *L for* gene length and $I_{Mut}$ as indicator variable that sample is glucose.

The absolute copy number of transcripts per cell was then determined with the following formula:

$$Absolute\ copy\ number\ of\ transcripts\ per\ cell = \frac{t \times A}{Cell\ count}$$

### 3.2.10. Determination of the *in vivo* catalytic rate ($k_{app}$) and the *in vivo* protein synthesis rate ($k_x$).

Before determine the *in vivo* catalytic rate ($k_{app}$), the molar quantity of the absolute copy number of protein in the fermenter (Nprot in mol) was calculated followed by the protein concentration (mmol.gDCW$^{-1}$) was determined by these formulas :

$$\text{Nprot} = \frac{\text{Protein copy number} \times \text{Cell count} \times OD^{620}measured \times \text{V}}{Na} \quad (1)$$

$$Protein\ concentation = \frac{N_{prot}}{V \times DCW \times\ OD^{620}measured} \quad (2)$$

With $OD^{620}$, the optical density measured at 620nm and V the volume chemostat vessel (mL),
The *in vivo* catalytic rate, $k_{app}$ ($s^{-1}$) was then determined with the formula:

$$k_{app} = \frac{Specific\ metabolic\ flux\ rate}{Protein\ concentration\ \times\ 3600}$$

With the specific metabolic flux rate (mmol.gDCW.h$^{-1}$) obtained by the incorporation of the stoichiometry, calculated from products concentration, in the theoretic energy conservation model and adjusted via the genome scale model of *E. limosum*; and the protein concentration (mmol.gDCW) multiplied by 3600 to convert hour into seconds. The formula was applied to each protein and membranous complex identified in the central carbon metabolism in methanol and glucose condition at steady state.

The $k_x$ values were determined for each protein identified in both culture conditions by the following formula:

$$k_x = \frac{Protein\ copy\ number \times D}{Transcript\ copy\ number \times 3600}$$

### 3.2.11. Identification of promoter and RBS strengths

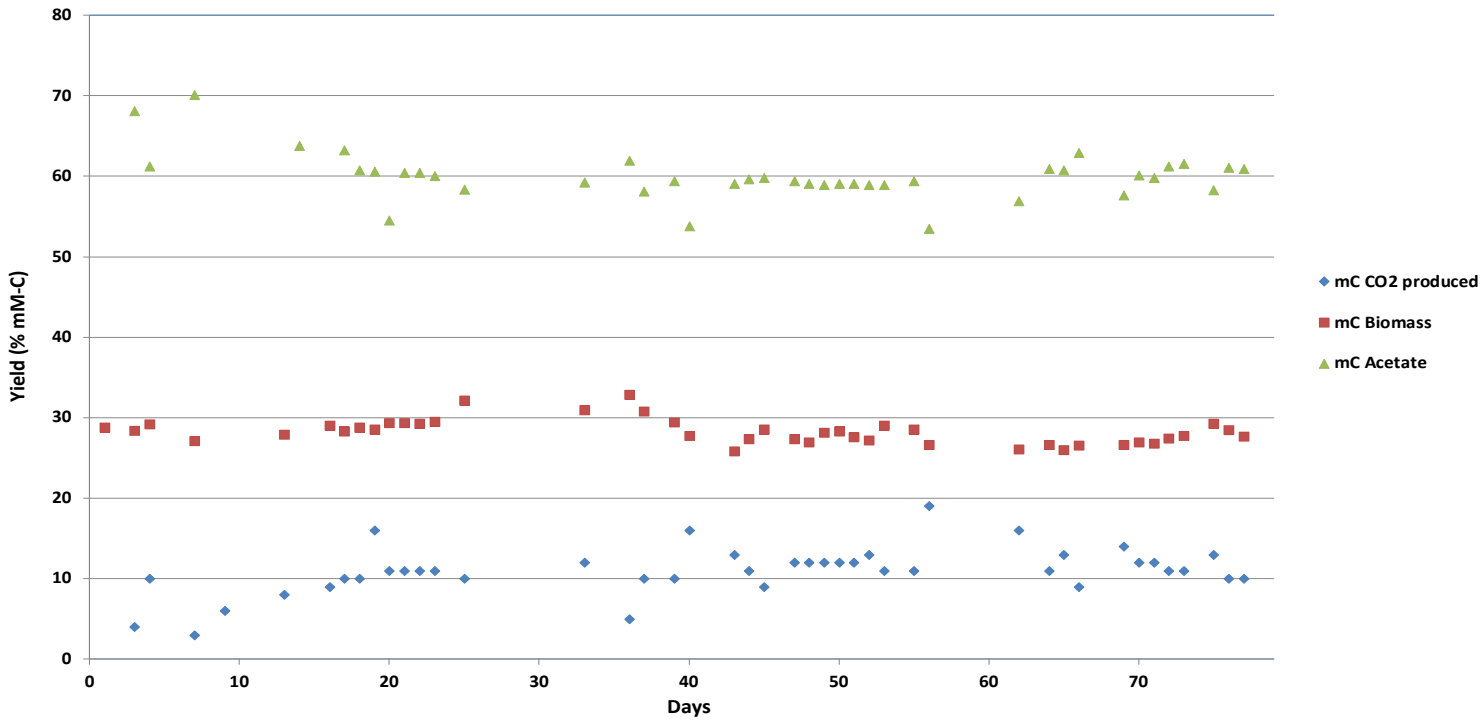
The promoter strength was determined from the absolute copy number of transcripts per cell. The more the transcript number was high, the more the promoter was considered as strong. The transcripts coded by genes belonging to an operon organization were excluded if it was not the first gene. The RBS strengths were assessed by the $k_x$ results as it's representative of the protein synthesis rate.

## 3.3. Results

### 3.3.1. Carbon balance and flux rate determination

The carbon balance and flux rates of substrate consumption and chemicals production of acetate and butyrate were calculated from methanol and glucose chemostat cultures at a steady state (Fig 3.1 and 3.2).

**A**



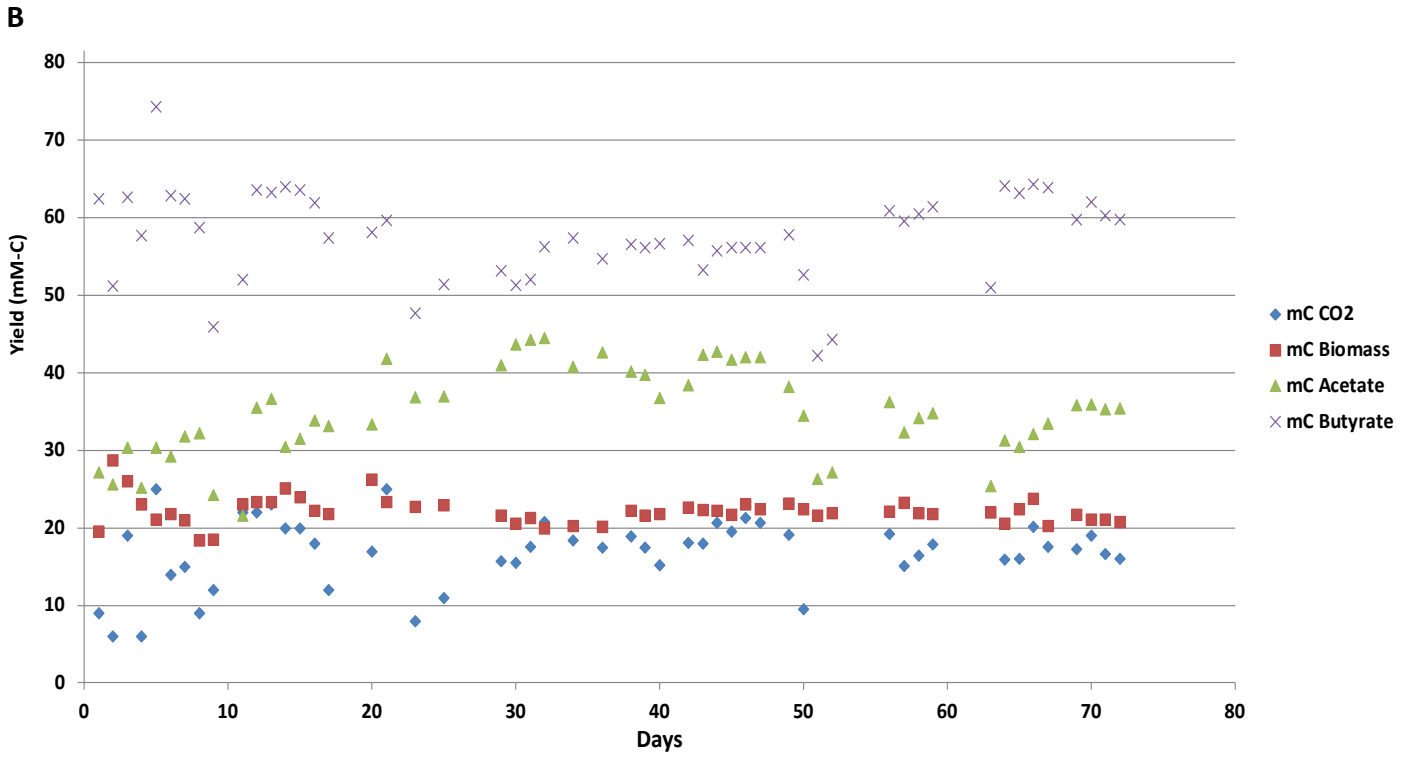



**Figure 3.1** : Carbon yields calculated for *E. limosum* B2 clone 2 in continuous culture on mineral medium. A) Methanol condition, B) glucose condition. Carbon yields for products, biomass, and $CO_2$ consumption/production were determined in percentage of substrate carbon

The substrate-limited condition at 200 mM of carbon to reach the steady state allowed an accurate physiologic and metabolic comparison for methanol and glucose conditions. The biomass concentration was estimated at 1.06 and 1.35 $gDCW.L^{-1}$ for methanol and glucose conditions

respectively while the average $OD^{620}$ was measured at 3.95 for methanol and 5.02 for glucose conditions. Stoichiometries and fluxes were determined for both conditions (Tab. 3.3)

**Table 3.3 :** Physiologic characterization by carbon distribution analysis from chemostat cultures on methanol and glucose substrate at steady state condition. Growth rate was set at $0{,}025h^{-1}$ for both conditions. A) Carbon balance and stoichiometry data, B) Flux rate values including substrate consumption, products biosynthesis rate, and $CO_2$ consumption or production rate depending on the substrate.

**A**

| | **Methanol** | **Glucose** |
|---|---|---|
| **Carbon balance** | 00 mC MetOH + 17 mC $CO_2$ = 36 mC Acetate 59 mC Butyrate + 22 mC $C_4H_7O_2N_{0.6}$ | 100 mC Glucose = 59 mC Acetate + 1 mC Butyrate + 28 mC $C_4H_7O_2N_{0.6}$ + 12mC $CO_2$ |
| **Stoichiometry** | 100 MetOH + 17 $CO_2$ = 18 Acetate + 14,75 Butyrate + 5.5 $C_4H_7O_2N_{0.6}$ | 100 Glucose = 177 Acetate + 1.5 Butyrate + 42 $C_4H_7O_2N_{0.6}$ + 72 $CO_2$ |

**B**

| **Flux (mmol/gDCW/h)** | **Methanol** | **Glucose** |
|---|---|---|
| **Substrate** | 4.72 | 0.61 |
| **Acetate** | 0.85 | 1.08 |
| **Butyrate** | 0.70 | 0.039 |
| **$CO_2$** | 0.84 Consumption | 0.44 Production |

During methylotrophy, the carbon yield for acetate production was around 40% lower compared to the glucose condition. A major part of the carbon substrate was dedicated to butyrate production while almost no butyrate was produced in glucose condition. Unsurprisingly, the glucose fermentation was homoacetic considering that only 1 for 100 mC of glucose was directed to butyrate production. The

carbon consumed from the substrate was around 21,5% higher for biomass production compared to methanol condition. During methylotrophy, the consumption of $CO_2$ is required as a co-substrate along methanol[21]. According to the stochiometric equation, 17 mC of $CO_2$ was co-consumed for 100 mC of methanol, *i.e.* 34 mC of $CO_2$ is required to consume all the 200 mC of substrate present in the feed medium.

The physiological characterization of the WT strain of *E. limousm* B2 was already performed in methanol and glucose conditions. The physiological profiles of both strains were comparable.

A similar carbon balance in methanol semi-synthetic medium was previously determined for *E. limosum* B2 WT at a growth rate 0,025$h^{-1}$ and showed closed yields: 100 mC MetOH + 23 mC CO2 → 48 mC Acetate + 68 mC Butyrate + 7 mC $C_4H_7O_2N_{0,6}$[118]. Product yields were slightly higher while the biomass yield was lower than our experimental results. Using an evolved strain for methanol mineral medium improved its biomass production yield at the expense of product formation yields. Regarding glucose condition, a previous study by Loubière *et al*, on similar steady state conditions showed the following equation: 100 mC Glucose → 55 mC Acetate + 5 mC Butyrate + 40 mC $C_4H_7O_2N_{0,6}$[75]. Although not quantified by instrument, gas was produced in our experiment at low specific growth while no $CO_2$ or $H_2$ was produced in their study at a similar specific growth rate but started to be produced at a minimal growth rate of 0.15$h^{-1}$.

In the study performed by Loubière *et al*, the steady state culture was performed on a semi-synthetic medium, which contained 16.7 mM of glucose, 60 mM of carbonates, 0.5 g.$L^{-1}$ of yeast extract, and 50 mM of acetate. In our experimental conditions, glucose was twofold more concentrated, acetate was not added and the carbonate content was 65% higher. During heterotrophy on glucose, an important amount of $CO_2$ is produced during the decarboxylation of pyruvate into acetyl-CoA by the pyruvate ferredoxin oxidoreductase (PFOR), and a part of this $CO_2$ is fixed via the WLP.

The substrate flux rate consumption of methanol was around 7.7 folds higher compared to glucose flux rate consumption. However, in terms of carbon substrate uptake rate, the value was 27% superior in methanol condition with 4.72 mmol C.gDCW$^{-1}$.$h^{-1}$ against 3.66 mmol C.$g^{-1}$.$h^{-1}$ in glucose condition. Since both conditions were set at the same dilution rate of 0.025$h^{-1}$, such difference indicates that the adapted strain improved its growth by enhancing its methanol assimilation capacity. Concerning the total specific production rate ($q_p$), 4.50 mmol C.$g^{-1}$.$h^{-1}$ was produced for methanol condition against 2.68 mmol C.$g^{-1}$.$h^{-1}$ including $CO_2$ production in glucose condition. The $q_p$ is therefore 1,68 folds higher during methylotrophy, underlying the interest in using methanol as substrate in terms of carbon transformation into valuable products.

### 3.3.2. Methanol metabolism and energy conservation

The physiological characterization previously determined through the carbon balance allowed the elaboration of theoretical models for energy conservation. These models aimed to highlight the carbon and electron fluxes through the central carbon metabolism along with the energy conservation system composed of the Rnf complex and the ATP synthase. Few theoretical energy conservation models were proposed for *Eubacterium* species, current models were elaborated from *E. callanderi* KIST612[70,77]. Therefore, we used these established theoretical energy conservation models to build the energetic profile of *E. limosum* B2 based on our experimental data. The newly discovered formate dehydrogenase/bifurcating dehydrogenase complex involving the reduction of redox carriers ferredoxin (Fd) and hydrogen ($H_2$) dehydrogenase in *E. callanderi* KIST612 (previously *Eubacterium limosum* KIST612) was integrated into the theoretical models as well as $Na^+$ dependent Rnf complex driving ion proton motive force through the membrane to fuel the membrane-bound ATPsynthase[77]. The stoichiometry of $Na^+$ translocated per ATP was assumed to be 4, based on the study performed on *E. callanderi* KIST612. Initially, the methylene-tetrahydrofolate reductase (MTHFR) was suggested to bifurcate electrons. However, *in vitro* enzymatic assays of the MTHFR complex from *E. callanderi* KIST612 showed that the reduction of methylene-THF to methyl-THF was only possible with ferredoxin as the electron donor [70]. No electron bifurcation or confurcation was possible nor using NADH as cofactor. However, *in vivo*, such activity would lead to a loss of ATP yield of 21% per mole of acetate during methylotrophy and appeared to be thermodynamically impossible during autotrophy with $CO_2$ + $H_2$ as substrate. For a sake of simplicity, the MTHFR was considered ferredoxin dependent in the present energetic model, and simulation on the genome-scale model of *E. limosum* was performed considering the MTFHR as Fd-dependent.

Based on products concentrations measured during methylotrophy, the energy conservation model and a global stoichiometric equation including carbon distribution and energy produced was determined using the GEM of *E.limosum* (*i*SP652) :

$$100\ CH_2OH + 17\ CO_2 + 53{,}25\ ADP + 53{,}25\ Pi + 26\ H_2O = 18\ CH_3COOH + 14{,}75\ CH_3CH_2CH_2COOH + 53{,}25\ ATP + 34{,}5H_2 + 5{,}5\ C_4H_7O_2N_{0,6}$$

Growth on methanol gave an ATP yield of 2,95 mol ATP/ mol acetate and 3,6 mol ATP/ mol butyrate. Compared to the substrate consumed, the ATP yield is 0.53 mol ATP produced/ mol of methanol consumed. Such value was lower than the ATP yield calculated on methanol condition for the acetogen *Acetobacterium woodii* with 0,625 mol ATP/mol of methanol[20].

As it was proposed for *A. woodii* and *E. callanderi* KIST612, the reverse use of methyl branch from WLP was also utilized by *E. limosum* B2 to produce the major part of ATP by producing $CO_2$ during the oxidation of formyl-THF into formate by the formyl-THF synthetase.

Surprisingly, the Rnf and the ATPase complexes showed low fluxes, suggesting a minor role. ATP production would mainly come from substrate-level phosphorylation with acetate production.
If the Rnf complex is not used during methylotrophy, the reduced ferredoxin produced during the oxidation methyl-THF to methylene-THF could be transferred to the CODH/ACS to ensure the reduction of $CO_2$ to CO in the carbonyl branch (Fig. 3.2).

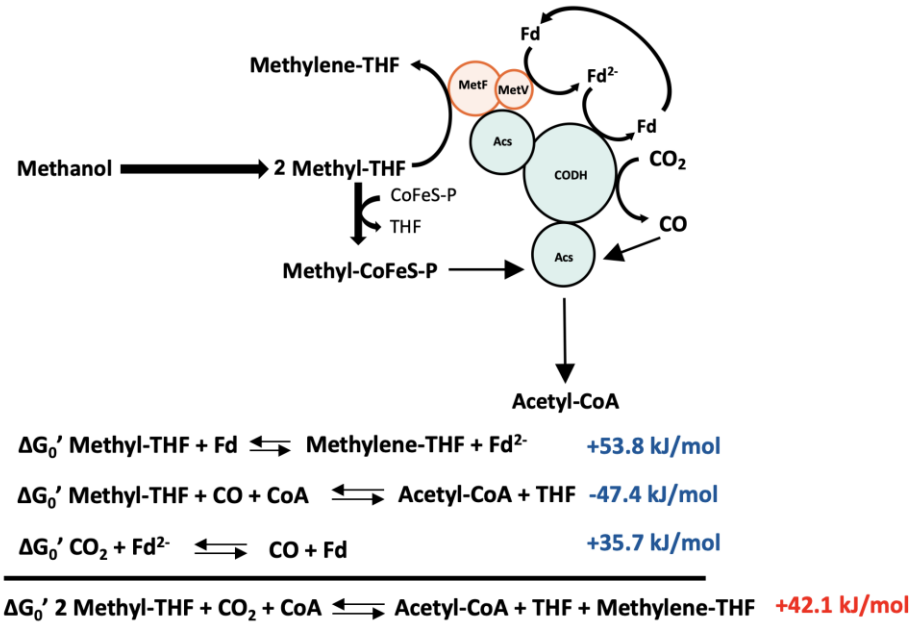



**Figure 3.2**: Hypothetic mechanism of the MTHFR-CODH/ACS complex for the simultaneous oxidation of methyl-THF and acetyl-CoA production using Fd as the cofactor in methylotrophic condition. Thermodynamics data calculated by eQuilibrator 3.0[180].

Such a complex would be thermodynamically unfavorable with a $\Delta G^0$ value of +42.1 kJ/mol. However, the NAD+ dependent oxidation of methyl-THF to methylene-THF is also unfavorable ($\Delta G^0$ value of +40.5 kJ/mol) but was demonstrated to be possible, making this model a possible explanation. The formation of this complex remains to be demonstrated.

In parallel, an important quantity of NADH was produced during the process, further oxidized by the Rnf complex to produce $Fd^{2-}$ along with the import of $Na^+$ into the intracellular medium. The redox balance appeared to be sustained by the action of the hydrogenase which captures the excess of $H_2$

produced from the oxidation of formate into $CO_2$ by the FDH-Hydrogenase complex, allowing the reduction of lacking $Fd^{2-}$ required to supply the demand of the redox carrier in the carbonyl branch. Such hydrogen cycling highlighted in *A. woodii* and suggested in *E. callanderi* could be a convenient way to maintain the redox balance[181]. In this model, the role of hydrogenase to reduce Fd and $NAD^+$ using $H_2$ as the electron donor by electron bifurcation is minor, leaving a residual part of 34,5 $H_2$ for 100 MetOH consumed. However, no gas measurement was performed during the chemostat culture of the adapted strain in methanol condition and the whole device was maintained at atmospheric pressure by flushing a $N_2/CO_2$ gas mixture at the surface of the fermenter to maintain anoxic conditions. In batch culture, no gas was emitted during growth. If residual cytoplasmic $H_2$ was entirely oxidated by hydrogenase to produce $Fd^{2-}$, it would induce an excess of NADH and disturb the redox balance. Further biochemical analyses are needed to confirm the model with the concomitant action of the FDH-hydrogenase complex and hydrogenase alone.

In the proposed model, a major part of $Fd^{2-}$ would be produced by the reverse use of ATP synthase. The membrane-bound ATP synthase would export $Na^+$ ions out of the membrane by the consumption of ATP to create ion proton motive force. $Na^+$ import by the Rnf complex allows the reduction of Fd using NADH as the electron donor. Therefore, the role of the ATPase and Rnf complexes seems to balance the intracellular redox potential at the expense of ATP.

### 3.3.3. Glucose metabolism and energy conservation

In parallel to the methanol condition, the theoretic model for energy conservation in glucose condition was elaborated with the data obtained from chemostat culture at a steady state. The present model was elaborated assuming electron confurcation during the reduction of $CO_2$ into formate by the enzymatic complex FDH/hydrogenase. Global stoichiometry showed the following equation:

$$100\ C_6H_{12}O_6 + 331{,}25\ ADP + 331{,}25\ Pi + 37{,}5\ H_2O = 177\ CH_3COOH + 1{,}5\ CH_3CH_2CH_2COOH + 331{,}25\ ATP + 144\ H_2 + 42\ C_4H_7O_2N_{0,6} + 72\ CO_2$$

Surprisingly, no solution was found by the GEM considering the MTHFR as Fd dependent. Interestingly, the model was possible with the use of NADH as the electron donor to oxidize the methylene-THF into methyl-THF. Such an approach would suggest another biochemical mechanism for the MTHFR.

In *A. woodii*, the MTHFR forms a complex with the cytoplasmic RnfC2 subunit, allowing the use of NADH as an electron acceptor. In *E. limosum,* no rnfC2 gene was identified, suggesting a complex

formation with another protein that contains an NADH binding domain. Such a hypothesis was emitted with the possibility of a complex formation between the MTHFR and RnfC subunit of the Rnf complex for *E. callanderi* during methylotrophy[71]. In our case, the MTHFR could form a complex with a cytoplasmic RnfC subunit using NADH as a cofactor (Fig. 3.3) Proteomic and transcriptomic data should bring answers with notable differences between methylotrophic and heterotrophic condition if the Rnf complex is not used but only RnfC subunit in glucose condition.

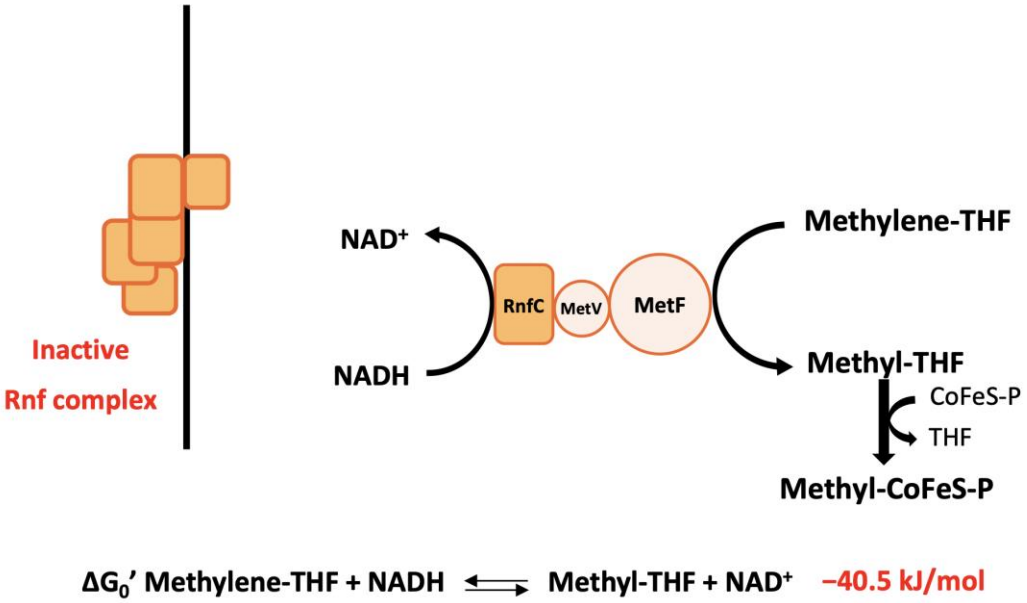



**Figure 3.3**: Hypothetic mechanism of the MTHFR functioning during heterotrophy on glucose.

Growth on glucose showed a net ATP yield of 2.97 mol of ATP/ mol of glucose, *i.e.* 0.5 ATP/ mol of carbon glucose, showing the same energetic value for methanol condition. At the same growth rate, the ATP yield per mol of carbon consumed was similar between both conditions. The adapted strain of *E. limosum* B2 showed a great ability to make an energy profit from methanol consumption. As expected in heterotrophic condition, a major part of ATP was produced during glycolysis. Regarding the redox carriers, the main source of NADH was produced during glycolysis as well as $Fd^{2-}$ which was mainly produced by the reduction of pyruvate into acetyl-CoA through the PFOR. A part of the $CO_2$ produced during this reaction was recycled in the WLP. Contrary to the methylotrophic condition, the methyl branch from the WLP was used in the usual sense, starting with the reduction of $CO_2$ into formate by the FDH along with the consumption of $Fd^{2-}$ and NADH. Also, the ATP synthase was used to produce ATP from the proton motive force driven by the concomitant action of Rnf complex and ATP synthase. As a consequence of the larger amount of redox carriers

with 144 $Fd^{2-}$ and NADH produced for 100 glucose consumed during glucose oxidation, the hydrogenase was much more solicited compared to methanol condition, producing $H_2$ to reduce the excess of reduced redox carriers.

As it was described for acetogens during heterotrophic growth on glucose, a major part of acetate produced came from the successive oxidation reactions started from glucose including the glycolysis to produce two molecules of pyruvate, the action of PFOR to oxidize pyruvate into acetyl-CoA and the double step enzymatic reaction catalyzed by the successive action of PTA and ACK to obtain acetate. Globally, the theoretic energy conservation model showed that WLP played a minor role in heterotrophic condition as well as energy conservation systems where few ATP was produced through the couple Rnf complex/ATP synthase. The energy was mainly produced by substrate-level phosphorylation.

For the first time, an energy conservation model was proposed for *E. limosum* B2 in methanol and glucose synthetic medium at a steady state. These specific culture conditions allowed the absolute quantification of mRNA and protein molecules per cell. First, the proteomic part will be analyzed.

### 3.3.4. Proteomic analysis

#### *3.3.4.1. Assessment of the quantitative method measurement*

Before analysis, we evaluated the efficiency of the extraction method and the accuracy of proteomic data generated by mass spectrometry by examining the correlation between the absolute copy number of proteins per cell and the expected stoichiometries for equimolar and non-equimolar proteins complexes (Fig. 3.4).

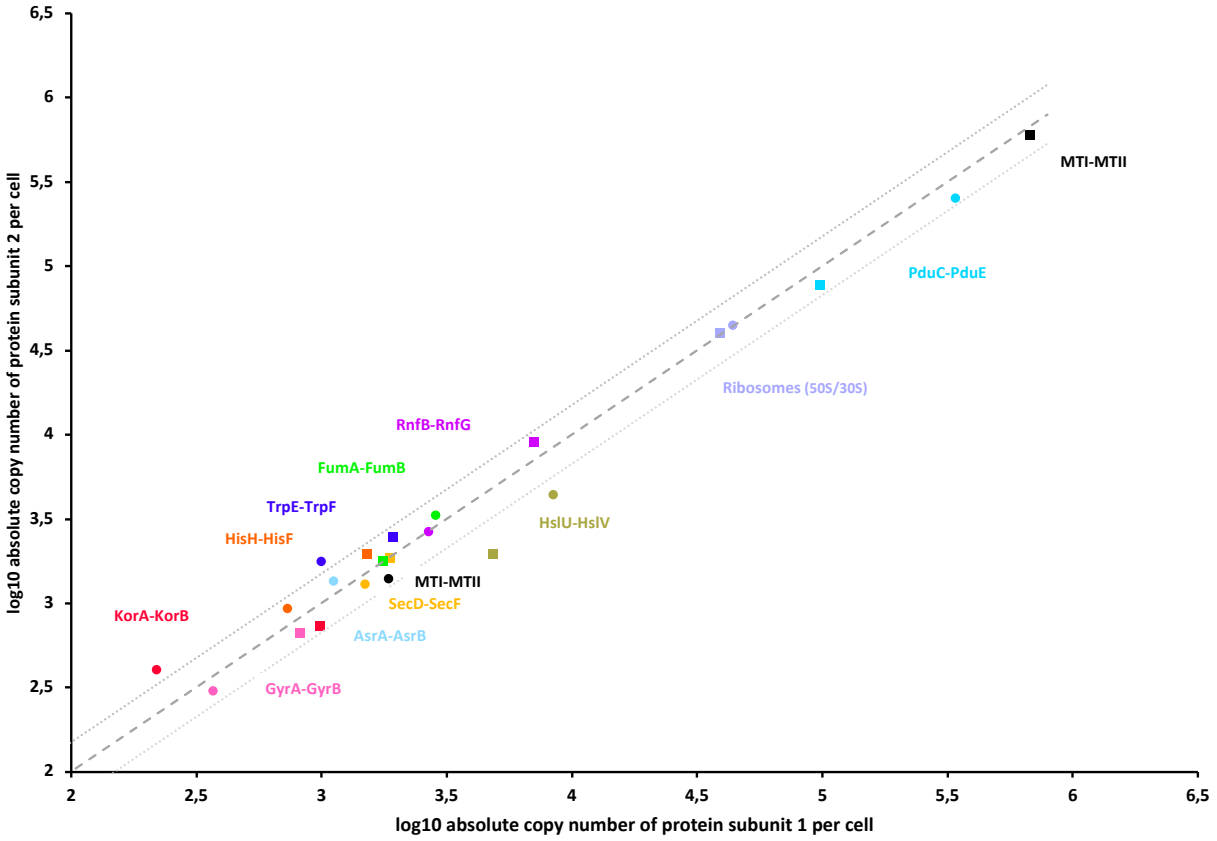



**Figure 3.4** : Correlation between expected equimolar protein complexes and the absolute number of protein subunits copy number per cell. Data were expressed in a log10 scale. Data points sharing the same colors represent data obtained from glucose and methanol conditions. The methanol condition is represented by squared points while the glucose condition is by round points. Grey dotted lines correspond to the 1.5-fold mean fold error zone. A total of 48 different ribosomal proteins were identified and the average number of 50S ribosomal protein number per cell was compared against 30S ribosomal proteins.

The data showed a strong correlation between the expected stoichiometry and the absolute copy number of subunits per cell determined for both conditions. A total of 11 out of 12 compared proteins complexes were found in the 1.5 mean fold error zone (MFE) for at least one of the two conditions tested, which represents the incertitude range calculated and cross-validated by using protein spike-in for an absolute proteins quantification study on C. *autoethanogenum*[125].

Furthermore, the expected stoichiometry and the absolute number of subunits per cell for 50S and 30S ribosomal proteins were strongly correlated for both conditions. A total of 45 out of 51 subunits of the 50S and 30S ribosomal proteins were quantifiable (Fig 3.5).

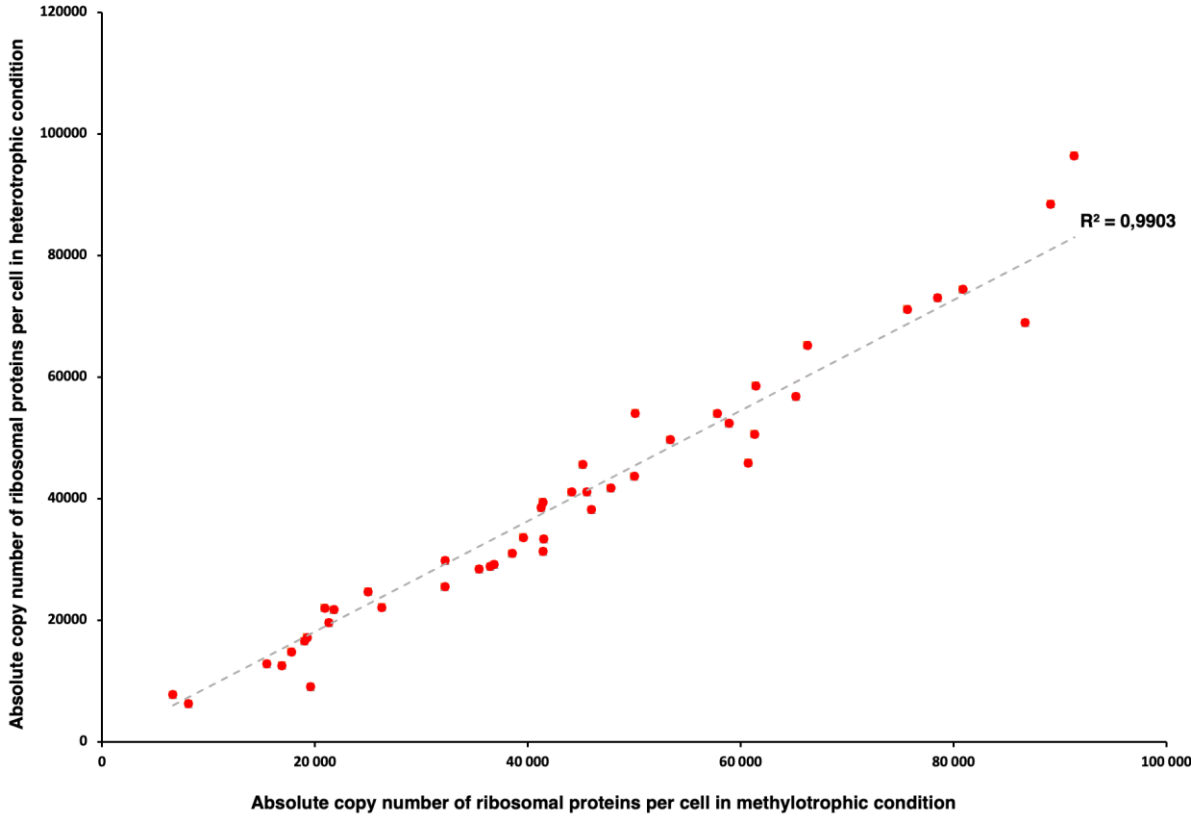



**Figure 3.5** : Linear correlation of the absolute copy number of ribosomal proteins per cell between methylotrophic and heterotrophic conditions.

We found a linear correlation between the absolute copy numbers of ribosomal protein per cell in glucose and methanol conditions. At a similar growth rate, the concentration of tRNA and ribosomes were shown to be similar and to be independent of the carbon source[182]. This physiological property was verified in our experiments, further validating the total proteome extraction process.

We also assessed the protein extraction efficiency on membrane-bound proteins by examining the expected stoichiometry and the absolute number of Rnf complex and ATP synthase, as the method was shown to be efficient on membrane-bound proteins although extraction is known to be difficult for the difficult-to-lyse gram-positive bacteria such as *E. limosum* B2 [143].

Membrane-bound proteins are particularly difficult to extract, usually requiring a combination of physical disruption methods of the cell membrane such as ultra-sonication or grinding with the addition of detergents to improve disruption of lipidic membranes. Furthermore, such methods demand several empirical steps for optimization and often lead to an important variability regarding extraction yield between replicates[183].

Although uncharacterized for *E. limosum* B2, the ATP synthase of its close genetic neighbor, *E. callanderi* KIST 612 was identified to be a unique combination of a $Na^+$ $A_1A_0$ ATP synthase with a V-type *c* subunit[184]. The membrane-bound complex is organized in a 9 genes operon. This structural particularity is probably shared by *E. limosum* B2 as a similar operon was found in its genome, showing highly similar sequences to the ATPase operon of *E. callanderi* KIST 612 (Fig. 3.6).

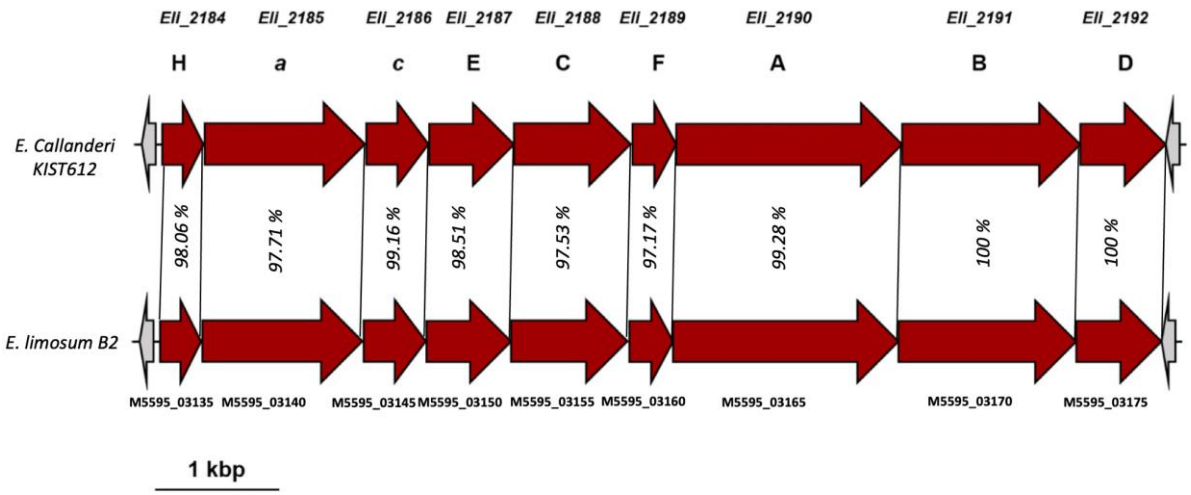



**Figure 3.6 :** Genetic comparison of the ATPase operon between *E. callanderi* KIST 612 and *E. limosum* B2.

A total of 8 out of 9 proteins coded from the ATPase operon were identified by mass spectrometry (Fig. 3.7 A). The subunit F was not detected for both physiological conditions tested. The absolute number of subunits was higher for methanol condition compared to glucose condition suggesting a higher number of ATPase complex per cell in methylotrophic condition.

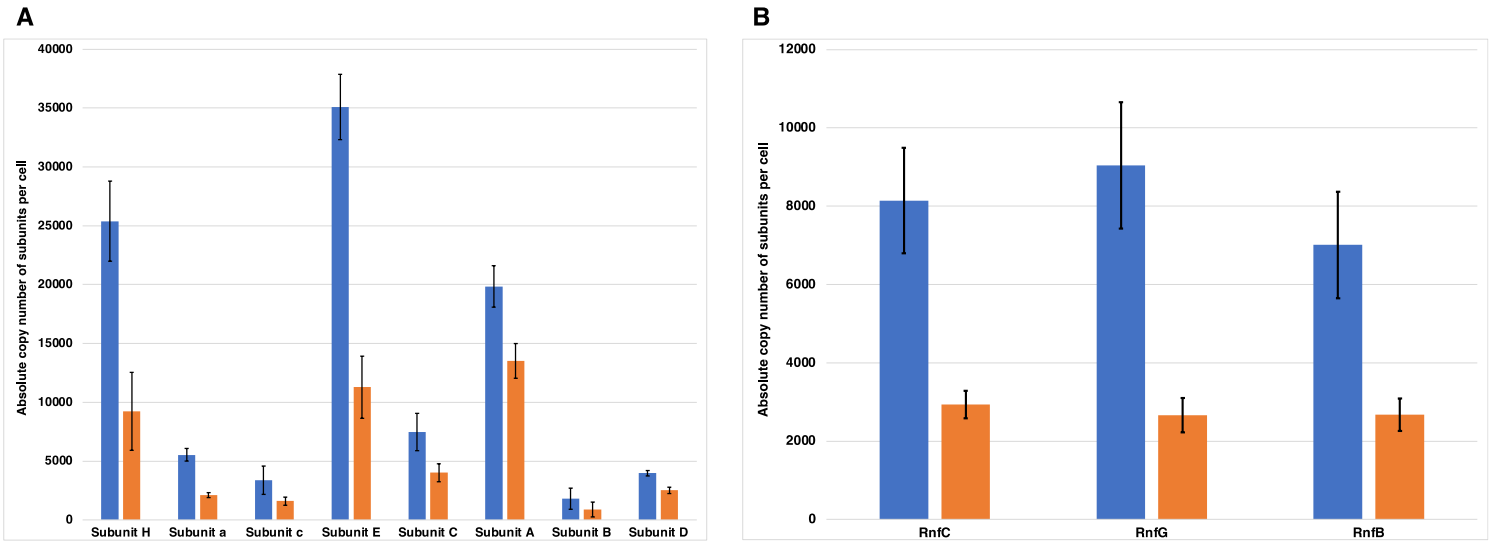



**Figure 3.7** : Absolute copy number of membrane bound complexes subunits involved in the energy conservation system for glucose and methanol conditions. A) Subunits from ATPase complex. B) Subunits from the Rnf complex. Blue bins represent methanol condition and orange bins the glucose condition .

The ATP synthase complex is organized as an $A_1A_0$ type with the cytoplasmic subunits A, B, F, D, E, and H. The addition of 8 to 10 subunits *c* forms the rotor which is anchored to the cell membrane while the subunit *a* play the role of the stator. The exact stoichiometry for ATPase in *E. limosum* is currently unknown as it requires specific experiments including precise imaging technology to determine the structure of the complex[156]. For example, the stoichiometry of *Pyrococcus furiosus*, an $A_1A_0$ type ATP synthase, was established with the following formula: $A_3B_3CDE_2FH_2ac_{10}$[157]. Although the structure of ATPase remains unknown for *E. limosum*, a similar organization to *P. furiosus* is assumed since subunit *c* is the only element that is prone to variation.

For the subunits sharing the same stoichiometry, the absolute number of subunits E and H are relatively close in glucose condition with around 10000 subunits. However, the stoichiometry was not respected for the membrane-bound subunits D, *a*, and *c* which showed significative differences according to their theoretical stoichiometry. Furthermore, cytoplasmic subunits A and B showed significant differences in the absolute number of subunits detected for both conditions. The expected stoichiometry was not reached considering the total number detected compared to other cytoplasmic subunits, especially for subunit B in both conditions. Although the expected stoichiometry was not reached for all constitutive ATPase subunits, the distribution between subunits was homogenous between both conditions suggesting a good reproducibility of the extraction method.

Among the 6 constituting subunits that form the Rnf complex, three subunits were detected by mass spectrometry for both physiological conditions, the cytoplasmic Rnf subunit C and the membrane-bound subunits B and G (Fig. 3.4 B). Theoretically equimolar, the three subunits showed a similar absolute copy number of subunits per cell for both conditions with a higher number of molecules in the methanol condition. Finding more Rnf complex copies per cell in methanol is not surprising since the $Na^+$ ion flux going through is higher during methylotrophy. Similarly to the ATPase complex, the distribution between both conditions was homogenous, reinforcing the method's robustness for further analysis.

Based on the absolute copy number per cell determined for the ATPase and Rnf complexes, the total extraction of membrane-anchored proteins seem incomplete by TFA treatment. Furthermore, a part of proteins might be lost during the neutralization process in TRIS 2M, as protein precipitation occurs during this step[143]. In the reference paper, the authors successfully recovered 95% of the total proteome according to the PeptideAtlas list for the gram-positive bacterium *B. subtilis*. Although efficient in cytoplasmic protein extraction, membrane proteins are still challenging to extract. However, the absolute values quantified showed great repeatability and reproducibility as was shown for the cytoplasmic proteins, which comforts us in our approach to proteome exploration.

#### *3.3.4.2. The total copy number of proteins brings information on enzyme usage of central carbon metabolism*

Proteomic analysis by mass spectrometry allowed the identification and the quantification of 1486 proteins for a total number of $14.1 \pm 2.53 \times 10^6$ proteins per cell in glucose and $14.0 \pm 1.72 \times 10^6$ proteins per cell in methanol condition. The total copy number of proteins per cell was not significatively different between both conditions. This is not surprising since total proteome was demonstrated in *E. coli* to be independent of the physiological conditions and correlated to the specific growth rate [185]. By comparison, the proteome of *Clostridium acetobutylicum* was characterized in steady state condition at a growth rate of $0.05h^{-1}$ for a total number of $6.26 \pm 0.18 \times 10^6$ proteins per cell[186]. In our experiment, the growth rate was set at $0.025h^{-1}$, and the total number of quantifiable proteins was twofold higher in our study, around 1400 for *E. limosum* B2 against more than 700 for *C. acetobutylicum*. However, the variability measured across biological replicates was also higher compared to the study on *C. acetobutylicum*.

The global range of total copy number of proteins per cell was comprised between 38 ± 14 to 7.48 ± 0.66 x $10^5$ for glucose condition and from 60 ± 44 to 6.75 ± 0.25 x $10^5$ copy number for methanol condition.

Globally, the enzymes involved in the WLP coupled to Rnf and ATP synthase were significatively more abundant in methanol condition. Interestingly, all countable proteins in methanol condition were also present in the glucose condition while more protein diversity was observed during heterotrophy. Globally, the average number of proteins per cell in the WLP was 3.80 higher in methanol condition. As expected, the average number of proteins per cell involved in butyrogenesis was 3.9 higher in the methanol condition while the value was 2.4 higher in the glucose condition for enzymes involved in acetate production.

A massive metabolic effort was made during methylotrophy for the methyltransferase complex biosynthesis (M5595_02325; M5595_02330; M5595_02335) with a total amount of 1.58 ± 0.08 x $10^6$ proteins copy number per cell which represents around 11% of total copy number of proteins per cell in methanol. Such value testifies to the energy investment required to assimilate this substrate.

Globally, around 4.65 x $10^6$ proteins were mobilized for the methanol assimilation, the WLP, and acetate/butyrate production which represents around 33% of the total protein copy number per cell in methanol condition. Comparatively, proteins mobilized in WLP and product formation represent only 6.5% of the total copy number of proteins per cell during heterotrophy on glucose.

Among the highest number of proteins per cell, proteins involved in the propanediol utilization were massively present. The highest number of proteins per cell in glucose condition was the microcompartment protein PduB (M5595_11295) with 7.48 ± 0.66 x $10^5$ protein units per cell (Fig. 3.8).

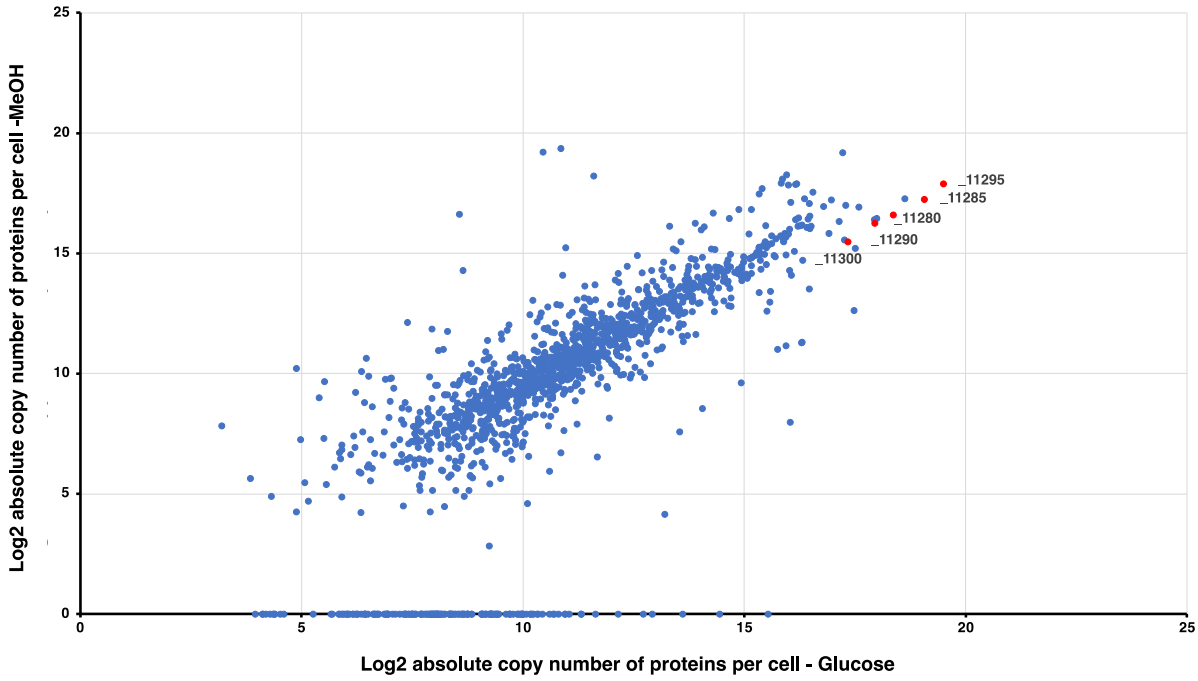



**Figure 3.8**: Absolute copy number of total proteins in methylotrophy vs heterotrophy. The figure highlights the most abundant proteins of the *pdu* operon in red. _XXXXX represents the locus tag.

The *pduB* gene belongs to the large *pdu* operon, first characterized for its role in propanediol assimilation [187]. The bacterial microcompartment (BMC) is an organelle constituted of a shell containing enzymes involved in the metabolization of a large variety of substrates. Toxic metabolites intermediary can then be managed by the cell using such a protective structure[188] Furthermore, BMC was shown to play a catalytic role by concentrating molecules and proteins in a confined zone. Overexpression of such genes was reported for *A. woodii* during growth on 1,2 – propanediol, 2,3 – butanediol, ethylene glycol, and ethanol[189]. Also, BMCs were identified for fucose and rhamnose metabolism in *Clostridium phytofermentens* but gene expression was turned-off during glucose fermentation[190]. BMC were also reported to play a role in cyanobacteria for $CO_2$ fixation by carbon concentrating mechanism within their lumen and fixation with encapsulated carbonic anhydrase and RuBisCO[191]. Although acetate constitutes the unique biochemical product synthesized during steady-state culture in glucose condition, several proteins involved in the utilization of 1,2- propanediol were present in large copy numbers per cell with propanediol dehydratase large and small subunits and propionaldehyde dehydrogenase. These proteins can be associated with other BMC proteins also present in large numbers. *Pdu* operon was studied in *E. maltosovirans*, showing evidences of BMCs by electronic microscopy under 1,2 propanediol + glucose and maltose conditions but no BMC was detected with glucose as only substrate [192]. Interestingly, the *pduA* gene (M5595_11300) in *E.*

*limosum* B2 is the first gene of the operon while the gene is at the last position in *E. maltosivorans* genome[192]. According to this study, the methyltransferase activity consisting of the transfer of a methyl group into the WLP for metabolization could be realized in BMCs. In the case of methylotrophy, BMC might be used to concentrate the cobamide and its different oxidized states to ensure the transfer of the methyl group into the WLP. Although *E. limosum* B2 was grown in a mineral medium compared to *E. maltosivorans* where 1 g/L of yeast extract was added, the production of BMCs during heterotrophy remains unknown. However, cells were grown at a low growth rate, $0{,}025h^{-1}$, which can induce a stress state caused by low nutrient abundance. Also, the reverse organization of the *pdu* operon between both species can play a role.

Further investigations are required to highlight the role of these proteins which are not known to play important role in glucose metabolism among acetogens but are intriguingly massively present in the adapted strain of *E. limosum* B2.

### 3.3.5. Transcriptomic analysis

#### *3.3.5.1. Determination of the absolute copy number of transcripts per cell and the absolute synthesis rate kx*

More than 105 million reads were generated during RNA-Seq with an average length of 99.5 bp (Tab. 3.4)

**Table 3.4 :** Summary of RNA-Seq data.

| Sample | **Total reads** | **Mean length (bp)** |
|---|---|---|
| **M1** | 18 083 661 | 107 |
| **M2** | 16 029 816 | 102 |
| **M3** | 20 721 899 | 100 |
| **G1** | 16 273 038 | 98 |
| **G2** | 16 398 826 | 95 |
| **G3** | 17 500 373 | 95 |

The distribution of spike-in mapped reads was examined to assess the quality of the RNA extraction and to define the minimal number of mapped reads below the range of linearity. The number of mapped reads to the first spike-in mix was normalized to 1kb and plotted against their relative concentrations (Fig. 3.9). The reads count of biological replicates for both conditions showed a similar

profile with an overestimation of counted reads for the lowest concentrated spike-in SpkH (Fig. 3.9 A & B). Therefore, reads counts below 10 were defined as the minimum threshold below the RNA expression is considered technical noise and impossible to accurately quantify. Surprisingly, the SpkG spike-in showed lower reads count than expected for both conditions and each biological replicate. Since this spike-in was of average size and concentration, such a result suggests an issue of concentration of the spike-in in the mix. This indicates that either a mistake was made during the spike-in dilution despite double checking during the spike-in mix preparation or that the concentration indicated by the manufacturer was wrong. Globally, the read count showed satisfying repeatability across the replicates. The read counts between both conditions were compared after the elimination of the two spike-in SpkG and SpkH as they were out of the linearity range (Fig. 3.9 C). The counted reads for both conditions showed a high degree of correlation ($R^2 = 0.999$), indicating strong stability of the spike-in across the transcript enrichment process. This spike-in mix was added right after cell lysis, indicating that DNAse treatment, on-column purification and concentration, ribodepletion, and library preparation did not significatively alter the spikes.

The read counts distribution was examined for the ERCC spike-in mix added to the total RNA before ribodepletion (Fig. 3.10). The biological replicates of the methanol condition showed a correlation coefficient $R^2$ from 0.93 to 0.95 and from 0.90 to 0.94 for the glucose condition. ERCC showed higher dispersion compared to the first spike-in mix. This variability can be imputed to the low volume added to the total RNA before ribodepletion (2µL).

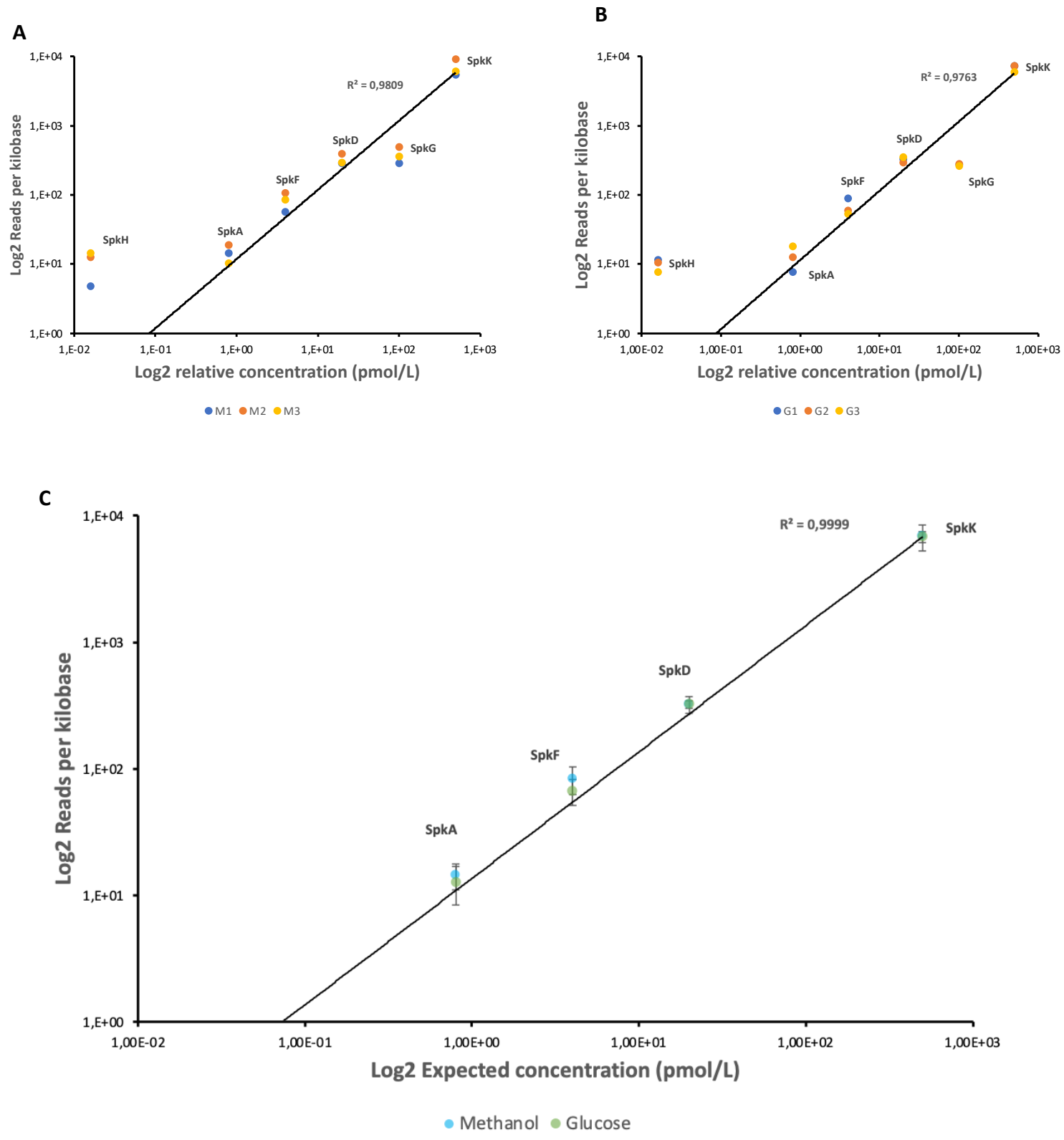



**Figure 3.9:** Reads distribution of the first spike-in mix added in methanol and glucose conditions. A) Reads distribution from the triplicate samples in methanol condition. B) Reads distribution from the triplicate samples in glucose condition. C) Comparison of reads distribution between glucose and methanol condition after correction of spike-in out of the linearity range. Read counts for each spike-in were normalized to 1kb. Relative concentration is expressed in picomole per liter.

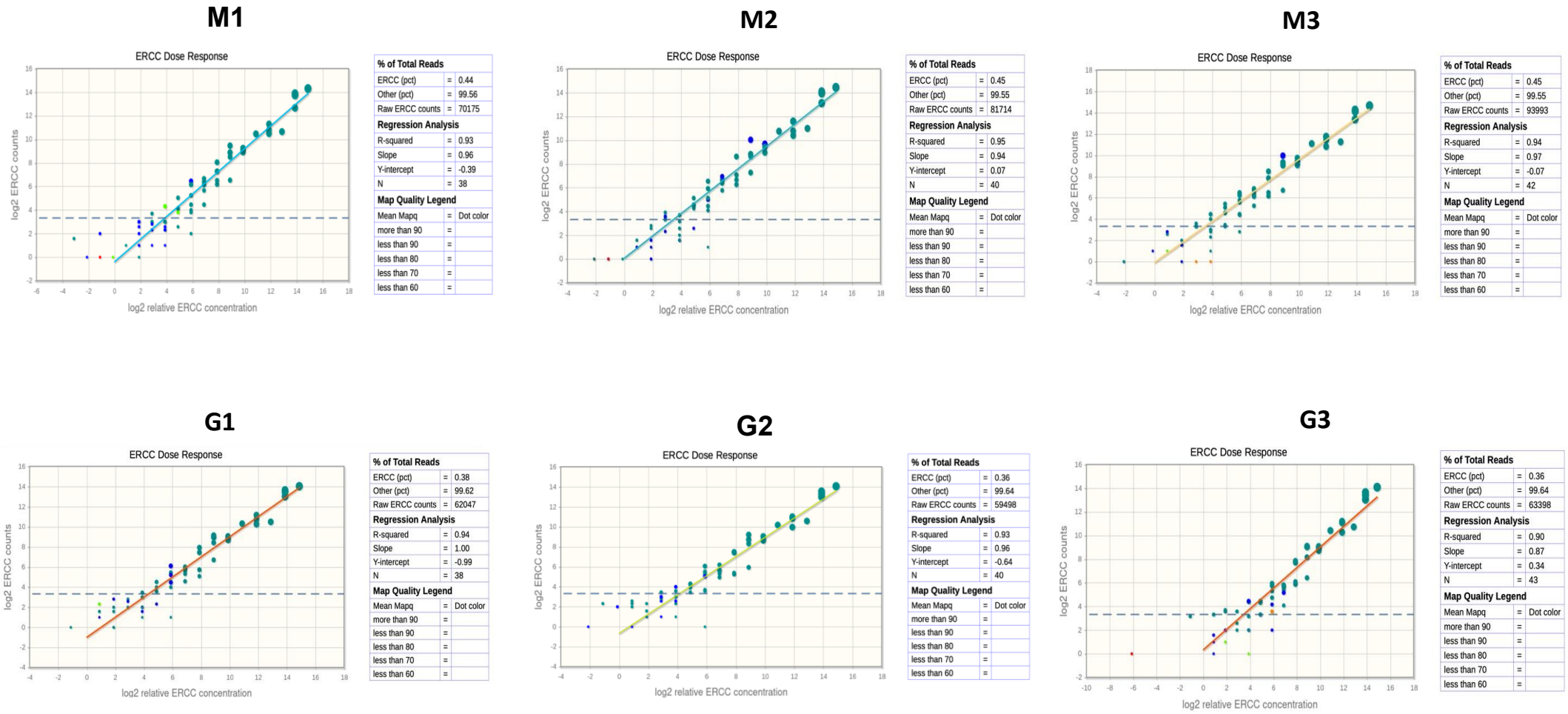



**Figure 3.10 :** Reads count distribution of ERCC spike-in mix and their relative concentrations. M1-M2-M3 correspond to the biological replicates in methanol co while G1-G2-G3 correspond to the biological replicates in glucose condition. The dotted line corresponds to tbe minimal threshold of ERCC read counts below technical noise is considered.

Read counts were then normalized with the first spike-in mix added to the samples just after the cell lysis step. A total of 9621 ± 561 and 10519 ± 1021 transcripts number per cell were found in heterotrophic and methylotrophic conditions respectively. Following the same physiologic dynamic as the total proteome, total RNA concentration per cell remains constant between both conditions set at the same dilution rate[193].

The absolute copy number of transcripts per cell was not significatively different between conditions. To assess the robustness of our approach for absolute quantification, we compared the copy number of ribosomal transcripts for both conditions (Fig 3.11 A). We found a linear correlation ($R^2$ 0.975) between both conditions confirming that absolute quantification at transcriptomic and proteomic levels follows the cell physiology at steady state. The maximum copy number of transcripts per cell was 140.7 ± 6.6 and 157 ± 7.6 for methanol and glucose conditions respectively. Globally, the methanol metabolism including the use of the WLP, product formation, and energy conservation systems (Rnf complex, ATPase, and hydrogenase) represents around 6.5 % of the total cell transcriptome cell in methylotrophy, suggesting an important translational regulation to sustain cellular functions for methanol condition.

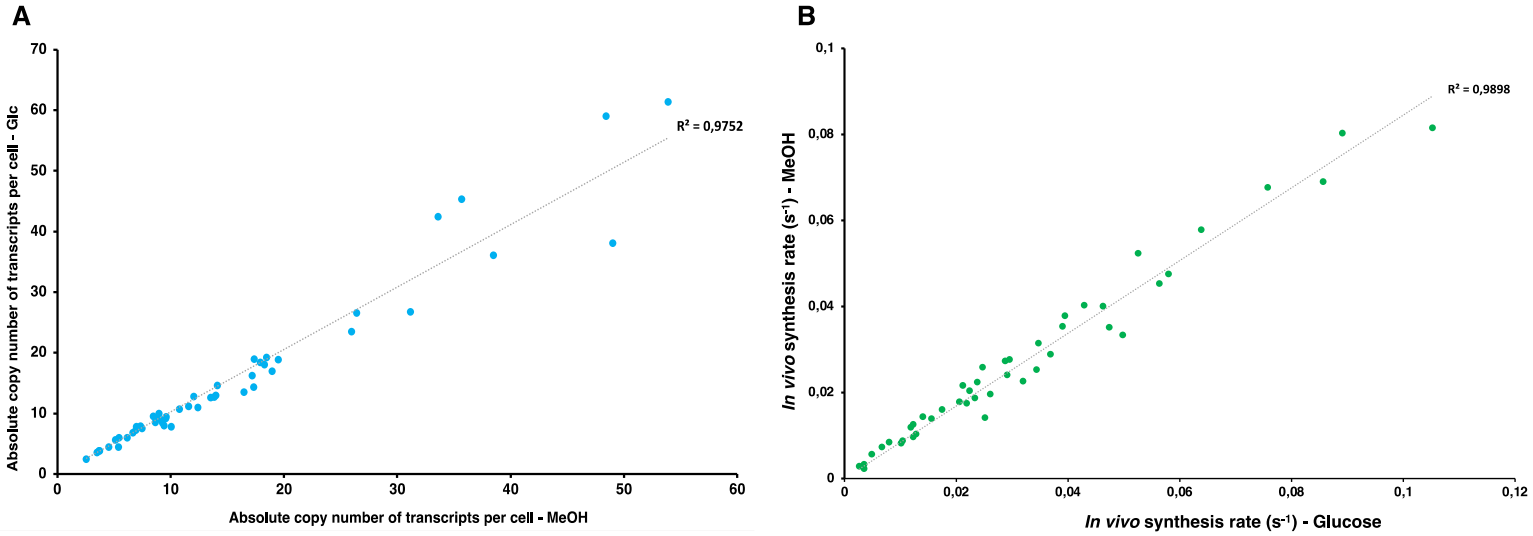



**Figure 3.11** : Assessment of transcriptomic absolute quantification results on transcripts coding for ribosomal proteins. A) Absolute copy number of transcripts for methanol and glucose conditions. B) *In vivo* synthesis rates $k_x$ ($s^{-1}$) for ribosomal proteins in glucose and methanol conditions.

Combined with the proteomic results, the determination of the *in vivo* synthesis rate $k_x$ for each protein is possible. The relevance of the values obtained was assessed by comparing the $k_x$ values for ribosomal proteins between both conditions (Fig. 3.11 B). A strong linear correlation was observed

between both conditions, $R^2$ = 0.99 confirming that our proteomic and transcriptomic data were relevant to properly quantify transcriptional and translational mechanisms.

### *3.3.5.2. Promoters and RBS strengths*

The absolute quantification of copy number of transcripts and proteins per cell allows the identification of the strongest promoters by identifying the highest copy number of transcripts per cell.

Promoters were then classified into three categories based on the log2 fold change of the ratio of mRNA copy per cell in methylotrophic on heterotrophic conditions. Promoters with log2 fold change values higher than one were considered methylotrophic, values below -1 were heterotrophic and between -1 and 1 constitutive. The strength in each category was then assessed by examining the absolute copy number of transcripts per cell for monocistronic genes and potential first genes of operons (Tab. 3.5 A, B and C).

**Table 3.5** : List of the 20th strongest promoters for each categories are presented below. A) Methylotrophic , B) Heterotrophic and C) constitutive promoters. Proteins IDs in red were no detected by mass spectrometry.

**A**

| Protein ID | Locus Tag | Gene | mRNA/cell MeOH | mRNA/cell Glc | log2FC |
|---|---|---|---|---|---|
| A0A1I5RRZ4 | M5595_20750 | hypothetical protein | **38,44** | **14,65** | **1,39** |
| A0A317RMC0 | M5595_12550 | glycine cleavage system protein GcvH | **33,92** | **14,86** | **1,19** |
| A0A317RC31 | M5595_20755 | helix-turn-helix domain-containing protein | **32,57** | **12,82** | **1,35** |
| A0A1I5RRZ4 | M5595_17295 | helix-turn-helix transcriptional regulator | **32,41** | **12,25** | **1,40** |
| A0A0U3FEV5 | M5595_16945 | preprotein translocase subunit SecG | **25,31** | **12,04** | **1,07** |
| A0A317RBF0 | M5595_20740 | helix-turn-helix domain-containing protein | **24,32** | **8,97** | **1,44** |
| A0A317RXG4 | M5595_09030 | acetyl-CoA decarbonylase/synthase complex subunit delta | **23,73** | **7,25** | **1,71** |
| A0A317RUQ7 | M5595_09395 | 4Fe-4S binding protein | **23,18** | **7,23** | **1,68** |
| A0A317RZM3 | M5595_17105 | hypothetical protein | **22,21** | **8,51** | **1,38** |
| A0A317RYC8 | M5595_08250 | hypothetical protein | **20,06** | **1,75** | **3,52** |
| A0A317RFD1 | M5595_06835 | RecX family transcriptional regulator | **18,96** | **7,77** | **1,29** |
| A0A317RH23 | M5595_06360 | DUF111 family protein | **18,39** | **6,48** | **1,51** |
| A0A317RJV4 | M5595_12575 | cyclodeaminase/cyclohydrolase family protein | **17,39** | **6,70** | **1,38** |
| A0A317RH07 | M5595_15380 | cation diffusion facilitator family transporter | **16,70** | **3,50** | **2,25** |
| A0A1H0R952 | M5595_16155 | CarD family transcriptional regulator | **15,41** | **7,50** | **1,04** |
| A0A317RDF4 | M5595_20775 | hypothetical protein | **14,26** | **4,94** | **1,53** |
| A0A317RM40 | M5595_11565 | prepilin-type N-terminal cleavage/methylation domain-containing protein | **12,99** | **4,38** | **1,57** |
| A0A317RJ24 | M5595_12585 | methyltetrahydrofolate cobalamin methyltransferase | **11,20** | **5,42** | **1,05** |
| A0A317S1D5 | M5595_17305 | hypothetical protein | **9,61** | **3,67** | **1,39** |
| A0A6N2YTA1 | M5595_12560 | methylenetetrahydrofolate reductase | **9,60** | **3,42** | **1,49** |

**B**

| Protein ID | Locus Tag | Gene | mRNA/cell MeOH | mRNA/cell Glc | log2FC |
|---|---|---|---|---|---|
| A0A317REL6 | M5595_19190 | HPr family phosphocarrier protein | **15,01** | **37,02** | **-1,30** |
| A0A0U3FC86 | M5595_11300 | propanediol utilization microcompartment protein PduA | **15,93** | **36,26** | **-1,19** |
| A0A317RJC8 | M5595_09920 | HU family DNA-binding protein | **15,46** | **34,79** | **-1,17** |
| A0A317RU86 | M5595_07045 | PRD domain-containing protein | **1,10** | **15,98** | **-3,87** |
| A0A317RK26 | M5595_02635 | aquaporin family protein | **4,10** | **12,17** | **-1,57** |
| A0A317RGV4 | M5595_10630 | CHC2 zinc finger domain-containing protein | **5,70** | **11,40** | **-1,00** |
| A0A0U3FI06 | M5595_09210 | ribosome-associated translation inhibitor RaiA | **5,23** | **10,62** | **-1,02** |
| A0A317RU34 | M5595_05190 | OsmC family protein | **4,51** | **9,58** | **-1,09** |
| A0A317RM40 | M5595_02290 | hypothetical protein | **4,60** | **9,48** | **-1,04** |
| A0A317RNZ7 | M5595_14120 | sulfurtransferase TusA family protein | **4,12** | **8,67** | **-1,07** |
| A0A317RPA7 | M5595_13630 | helix-turn-helix domain-containing protein | **2,59** | **8,19** | **-1,66** |
| A0A317RJW3 | M5595_03030 | TIGR03905 family TSCPD domain-containing protein | **3,16** | **7,83** | **-1,31** |
| A0A317RHN3 | M5595_10420 | GntR family transcriptional regulator | **1,81** | **6,92** | **-1,93** |
| A0A317RH81 | M5595_10480 | transketolase | **0,58** | **6,62** | **-3,51** |
| A0A317RZZ8 | M5595_17175 | CBS domain-containing protein | **2,11** | **6,28** | **-1,57** |
| A0A317RM16 | M5595_11240 | cob(I)yrinic acid a,c-diamide adenosyltransferase | **1,98** | **5,17** | **-1,39** |
| A0A1H0Q7L3 | M5595_07675 | M48 family metallopeptidase | **1,36** | **3,17** | **-1,22** |
| A0A317RF72 | M5595_06665 | LysR family transcriptional regulator | **0,72** | **2,35** | **-1,71** |
| A0A317RMI6 | M5595_12040 | hypothetical protein | **1,00** | **2,00** | **-1,00** |
| A0A317RIT8 | M5595_06280 | hypothetical protein | **0,57** | **1,42** | **-1,32** |

**C**

| Protein ID | Locus Tag | Gene | Average mRNA/cell | log2FC |
|---|---|---|---|---|
| A0A317RVM5 | M5595_09510 | hypothetical protein | **143,58** | **-0,23** |
| A0A317RWM7 | M5595_07810 | hypothetical protein | **128,59** | **0,27** |
| A0A317RPU9 | M5595_14440 | 4Fe-4S binding protein | **128,42** | **0,05** |
| A0A0U3FSH4 | M5595_11095 | cold-shock protein | **88,51** | **-0,32** |
| A0A317RTE3 | M5595_03180 | zinc-ribbon domain-containing protein | **73,68** | **0,12** |
| A0A317RHH2 | M5595_15255 | cyclic lactone autoinducer peptide | **65,04** | **0,39** |
| A0A0U3ECV5 | M5595_09225 | six-cysteine ranthipeptide SCIFF | **54,39** | **0,05** |
| A0A317RWE7 | M5595_07360 | 50S ribosomal protein L28 | **53,77** | **0,29** |
| A0A0U3GC77 | M5595_18405 | 30S ribosomal protein S21 | **43,56** | **-0,36** |
| A0A317RF75 | M5595_06695 | cold-shock protein | **42,30** | **0,10** |
| A0A317RF83 | M5595_06715 | hypothetical protein | **40,94** | **0,07** |
| A0A317RU79 | M5595_08490 | 30S ribosomal protein S20 | **40,51** | **0,35** |
| A0A317RQZ6 | M5595_13635 | SEC-C metal-binding domain-containing protein | **38,89** | **-0,90** |
| A0A317RIG4 | M5595_15400 | rubredoxin | **37,32** | **0,10** |
| A0A317RD75 | M5595_19170 | peptidoglycan DD-metalloendopeptidase family protein | **36,76** | **0,46** |
| A0A317RS31 | M5595_03660 | hypothetical protein | **35,70** | **0,43** |
| A0A317RLS1 | M5595_02560 | CsbD family protein | **33,98** | **0,14** |
| A0A0U3FKE9 | M5595_14510 | rubredoxin | **30,90** | **0,73** |
| A0A0U3G785 | M5595_07180 | transcriptional repressor | **30,88** | **0,02** |
| A0A0U3ECH4 | M5595_08475 | 50S ribosomal protein L31 | **28,98** | **-0,22** |

In methylotrophic condition, the strongest promoter, with around 38.5 mRNA copy numbers per cell, was identified for a gene coding for a hypothetic protein presumably located in a large phage gene cluster (Tab. 3.5 A). Interestingly, genes coding for WLP enzymes appeared to be specifically strong during methylotrophy such as *acsD* (M5595_09030). The presence of a promoter sequence was validated by comparison with data available on *E. limosum* ATCC8486[88]. The glycine cleavage system H protein coding gene *gcvH* was shown to be presumably under the control of a strong promoter during methylotrophy.

During heterotrophic conditions, the *pduA* gene (M5595_11300), coding for BMCs showed higher copy numbers of mRNA/cell compared to the other condition (Tab. 3.5 B) This observation matches with the higher concentration of protein per cell detected by mass spectrometry, highlighting the important role of the operon in this condition. Globally, the genes with an absolute copy number of transcripts per cell with a log2 fold change < -1 were significantly less important in glucose condition compared to methanol condition with 58 genes and 179 genes respectively. However, a clear identification of promoters significantly stronger for each gene condition remains to do.

Regarding the strongest constitutive promoters, most of the genes were involved in the cellular maintenance, such as genes coding for ribosomal proteins (Tab. 3.5 C). Such results were not surprising since the growth rate was equally maintained for both conditions. Interestingly, the number of mRNA/cell was higher compared to the specific methylotrophic and heterotrophic strong promoters.

The RBS strength was assessed by examining the average *in vivo* protein synthesis rate $k_x$ values for both conditions (Tab. 3.6). The $k_x$ values associated with genes showing less than 0.5 mRNA/cell were filtered. In accordance with the absolute quantification of proteins, the $k_x$ for PduA (M5595_11295) was the highest value measured with 0,257 $s^{-1}$. Several genes involved in the WLP and product biosynthesis showed a high protein synthesis rate.

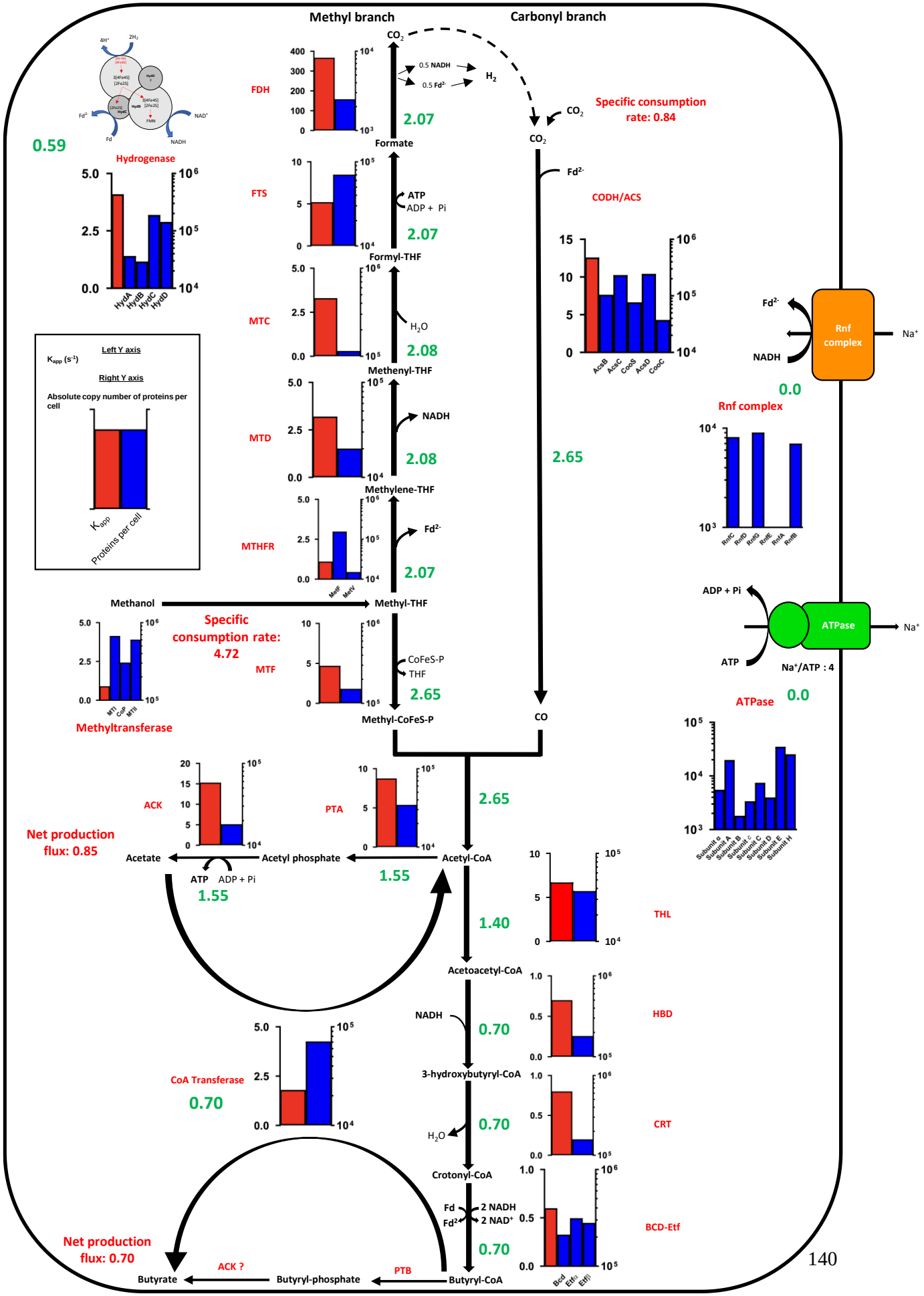
**Table 3.6** : List of the 40th highest *in vivo* protein synthesis rates values ($k_x$)

| Protein ID | Locus Tag | Gene | $k_x$ ($s^{-1}$) |
|---|---|---|---|
| A0A0U3E083 | M5595_11295 | propanediol utilization microcompartment protein PduB | **0,257** |
| A0A317RIV3 | M5595_02325 | methanol--corrinoid methyltransferase | **0,238** |
| A0A317RM29 | M5595_11290 | propanediol/glycerol family dehydratase large subunit | **0,213** |
| A0A0U3G516 | M5595_11285 | propanediol/glycerol family dehydratase medium subunit | **0,164** |
| A0A6N2YTA1 | M5595_12560 | methylenetetrahydrofolate reductase | **0,160** |
| A0A1V0LRE2 | M5595_11670 | elongation factor Tu | **0,160** |
| A0A0U3EME7 | M5595_09020 | carbon monoxide dehydrogenase/acetyl-CoA synthase methytransferase subunit | **0,152** |
| A0A317RKW5 | M5595_11645 | fumarylacetoacetate hydrolase family protein | **0,147** |
| A0A0U3GD27 | M5595_16290 | electron transfer flavoprotein subunit alpha/FixB family protein | **0,147** |
| A0A317RZF7 | M5595_18055 | NADH-quinone oxidoreductase subunit NuoE | **0,129** |
| A0A0U2UJ95 | M5595_09025 | acetyl-CoA decarbonylase/synthase complex subunit gamma | **0,121** |
| A0A317RHC5 | M5595_15510 | dihydroxyacetone kinase subunit DhaK | **0,121** |
| A0A317RVX8 | M5595_04555 | ribosome recycling factor | **0,117** |
| A0A317RV63 | M5595_03670 | 6,7-dimethyl-8-ribityllumazine synthase | **0,114** |
| A0A317RZI7 | M5595_16280 | acyl-CoA dehydrogenase | **0,113** |
| A0A317RZK6 | M5595_16180 | FAD-dependent oxidoreductase | **0,110** |
| A0A317RWP4 | M5595_18070 | 4Fe-4S binding protein | **0,105** |
| A0A317RL08 | M5595_11395 | transketolase | **0,105** |
| A0A317S042 | M5595_16925 | type I glyceraldehyde-3-phosphate dehydrogenase | **0,101** |
| A0A317RSE4 | M5595_00290 | pyruvate:ferredoxin (flavodoxin) oxidoreductase | **0,100** |
| A0A317RT48 | M5595_04835 | SGNH/GDSL hydrolase family protein | **0,098** |
| A0A317RQG9 | M5595_01705 | Rib/alpha-like domain-containing protein | **0,097** |
| A0A317RU34 | M5595_05190 | OsmC family protein | **0,096** |
| A0A317RXN6 | M5595_16285 | electron transfer flavoprotein subunit beta/FixA family protein | **0,096** |
| A0A317RVD8 | M5595_09005 | acetyl-CoA decarbonylase/synthase complex subunit alpha/beta | **0,095** |
| A0A0U3E602 | M5595_11040 | 50S ribosomal protein L16 | **0,093** |
| A0A317RXJ4 | M5595_18075 | NADH-dependent [FeFe] hydrogenase, group A6 | **0,091** |
| A0A317RJL5 | M5595_06115 | butyryl-CoA:acetate CoA-transferase | **0,091** |
| A0A317RHW9 | M5595_19805 | ABC transporter substrate-binding protein | **0,090** |
| A0A317RHJ8 | M5595_10090 | transporter substrate-binding domain-containing protein | **0,088** |
| A0A317RT12 | M5595_20550 | acetate kinase | **0,087** |
| A0A317REQ5 | M5595_20015 | FprA family A-type flavoprotein | **0,086** |
| A0A317RLT3 | M5595_11000 | 50S ribosomal protein L6 | **0,085** |
| A0A317RWF3 | M5595_07070 | 2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-acetyltransferase | **0,084** |
| A0A317RM06 | M5595_10240 | glycogen/starch/alpha-glucan phosphorylase | **0,083** |
| A0A317RIN8 | M5595_20200 | pyruvate kinase | **0,082** |
| A0A317RW29 | M5595_08730 | molecular chaperone DnaK | **0,081** |
| A0A317RZ63 | M5595_16275 | 3-hydroxyacyl-CoA dehydrogenase NAD-binding domain-containing protein | **0,081** |
| A0A317RRY6 | M5595_01135 | pyruvate, phosphate dikinase | **0,081** |
| A0A317RN71 | M5595_14135 | HAD-IIA family hydrolase | **0,080** |

### 3.3.6. Fluxes and apparent *in vivo* catalytic rates of enzymes from primary methanol metabolism

The determination of the absolute copy number of proteins per cell combined with the specific metabolic flux rate obtained from the profile product and the GEM of *E. limosum* (*i*SP652) allowed the estimation of the apparent *in vivo* catalytic rate (Kapp) for enzymes of the central carbon metabolism[173]. The steady-state cultures for both conditions were paramount to reaching the absolute quantification level and precisely determining fluxes and catalytic rates. Products profile measured for both conditions were incorporated in the GEM of *E. limosum i*SP652 (data unpublished) to determine the metabolic flux rate of each enzyme in the WLP. The apparent *in vivo* catalytic rates ($k_{app}$) and synthesis rate ($k_x$) of each enzyme of the WLP, energy conservation complexes, and acid production proteins were determined for both conditions (Fig. 3.12 A & B).

A


Methyl branch
Carbonyl branch
Hydrogenase
0.59
FDH
2.07
Formate
0.5 NADH
0.5 Fd2-
H2
CO2
Specific consumption rate: 0.84
FTS
ATP
ADP + Pi
2.07
Formyl-THF
CODH/ACS
MTC
H2O
2.08
Methenyl-THF
Fd2-
Rnf complex
Na+
NADH
0.0
Rnf complex
MTD
NADH
2.08
Methylene-THF
2.65
MTHFR
Fd2-
2.07
Methanol
Methyl-THF
ADP + Pi
ATPase
Na+
ATP
Na+/ATP : 4
ATPase
0.0
Specific consumption rate: 4.72
MTF
CoFeS-P
THF
2.65
Methyltransferase
Methyl-CoFeS-P
CO
ACK
PTA
2.65
Net production flux: 0.85
Acetate
Acetyl phosphate
Acetyl-CoA
ATP
ADP + Pi
1.55
1.55
THL
1.40
Acetoacetyl-CoA
HBD
NADH
0.70
3-hydroxybutyryl-CoA
CoA Transferase
0.70
CRT
H2O
0.70
Crotonyl-CoA
BCD-Etf
Fd
Fd2+
2 NADH
2 NAD+
0.70
Net production flux: 0.70
ACK ?
PTB
Butyrate
Butyryl-phosphate
Butyryl-CoA
Left Y axis
Kapp (s-1)
Right Y axis
Absolute copy number of proteins per cell
Kapp
Proteins per cell

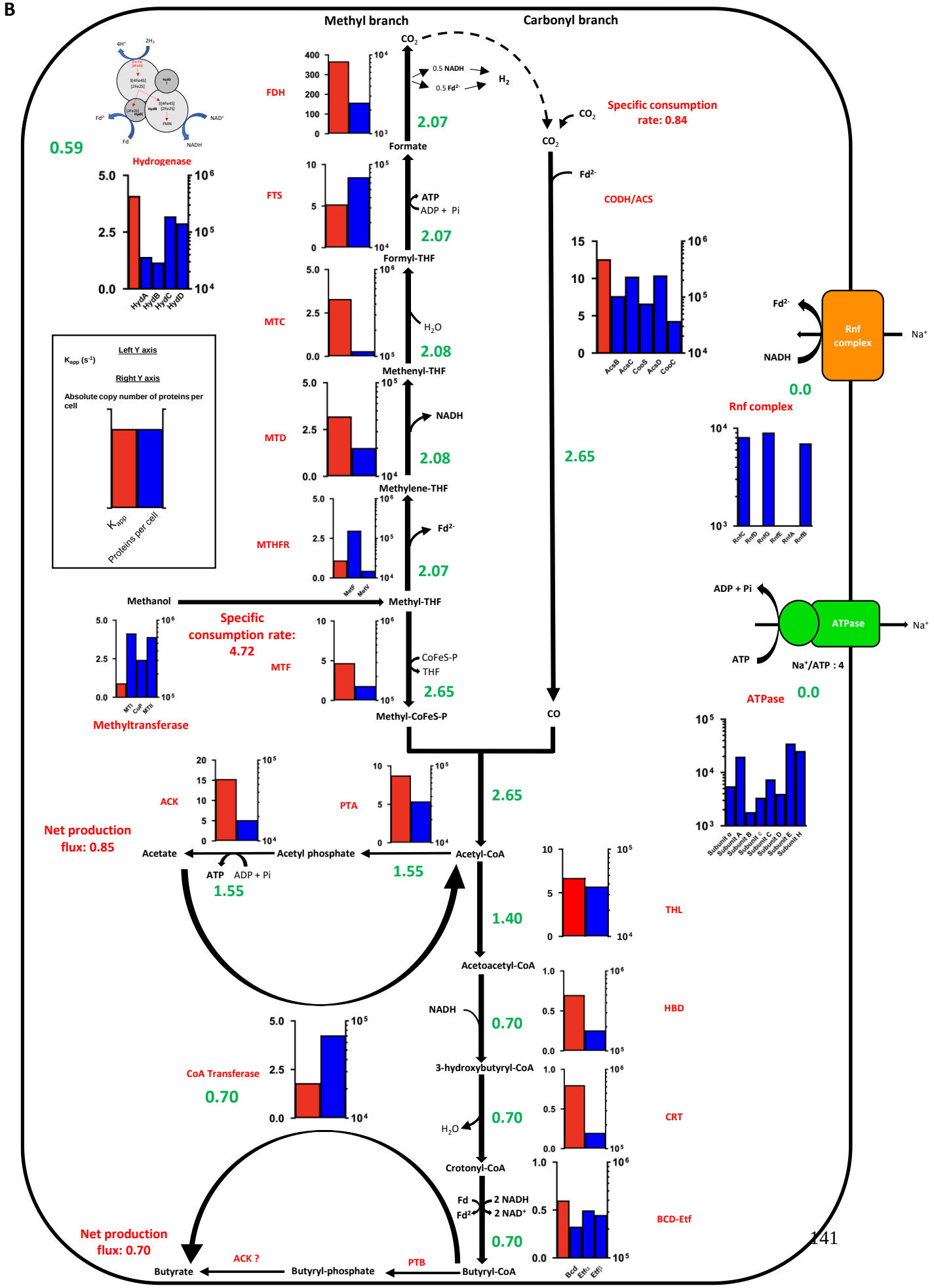

B
Methyl branch
Carbonyl branch
0.59
Hydrogenase
HydA
HydB
HydC
HydD
FDH
FTS
MTC
MTD
MTHFR
MTF
Formate
Formyl-THF
Methenyl-THF
Methylene-THF
Methyl-THF
Methyl-CoFeS-P
CO
Methanol
Methyltransferase
Specific consumption rate: 4.72
Specific consumption rate: 0.84
CODH/ACS
AcsB
AcsC
CooS
AcsD
CooC
2.07
2.07
2.08
2.08
2.07
2.65
2.65
Rnf complex
RnfC
RnfD
RnfG
RnfE
RnfA
RnfB
0.0
ATPase
Na+/ATP : 4
0.0
ACK
PTA
Acetate
Acetyl phosphate
Acetyl-CoA
ATP
ADP + Pi
1.55
1.55
Net production flux: 0.85
1.40
THL
Acetoacetyl-CoA
NADH
0.70
HBD
3-hydroxybutyryl-CoA
H2O
0.70
CRT
Crotonyl-CoA
Fd
Fd2+
2 NADH
2 NAD+
0.70
BCD-Etf
CoA Transferase
0.70
Net production flux: 0.70
ACK ?
PTB
Butyrate
Butyryl-phosphate
Butyryl-CoA

**Figure 3.12 :** Representation of the methanol central carbon metabolism with the fluxes, *in vivo* catalytic rates and the absolute copy number of proteins per cell for each enzymes. A) Methylotrophy on methanol B) Heterotrophy on glucose. The apparent *in vivo* catalytic rate $k_{app}$ ($s^{-1}$) and the absolute copy number of proteins are shown in the histograms for each enzyme of the central carbon metabolism. The $k_{app}$ values (in red) are indicated in the left Y axis, while the absolute copy number of proteins (in blue) are indicated on the right Y axis. Fluxes ($mmol.gDCW^{-1}.h^{-1}$) values are indicated in green beside each enzymes. Net production fluxes for acetate, butyrate and $CO_2$ (heterotrophy) and specific consumption rates ($mmol.gDCW^{-1}.h^{-1}$) for glucose, methanol and $CO_2$ (methylotrophy) are indicated in red beside each corresponding molecule.. Data were collected from four biological replicates from chemostat cultures at steady state. The $K_{app}$ were calculated assuming MTHFR Fd and NADH dependent for methanol and glucose conditions respectively. Omics and flux data are available in appendix. FDH, formate dehydrogenase; THF, tetrahydrofolate; FTS, formyl-THF synthetase; MTC, methenyl-THF cyclohydrolase; MTD, methyl-THF dehydrogenase; MTR, methylene-THF reductase; MTF, methyltransferase; CODH/ACS, carbon monoxide dehydrogenase/acetyl-CoA synthase; PTA, phosphotransacetylase; ACK, acetate kinase; THL, thiolase; HBD, 3-hydroxybutyryl-CoA dehydrogenase; CRT, crotonase; BCD-Etf, butyryl-CoA dehydrogenase Electron transfer flavoprotein complex; PFOR, pyruvate ferredoxin oxidoreductase; MTI, methyltransferase I; CoP, corrinoid protein; MTII, methyltransferase II; $Fd^{2-}$, reduced ferredoxin; [$e^-$], reducing equivalent. The ratio of $Na^+$ translocated per mole of ATP is assumed to be 4.

Globally, the fluxes, protein copy numbers and $k_{app}$ values in the WLP were higher in methanol condition. Under high metabolic flux, a correlation between enzyme abundance and catalytic activity was observed for *C. autoethanogenum, E. coli,* and other microorganisms[125,194,195].

The formate dehydrogenase (Fdh) showed the highest $k_{app}$ at 367 $s^{-1}$ for the methanol condition while the value was 112 $s^{-1}$ in the glucose condition. These high values are due to the very low amount of protein copy number detected per cell for both conditions. Although $k_{app}$ value for the formate dehydrogenase determined for *C. autoethanogenum* in autotrophic condition was considered as high compared to the other enzymes of the WLP with 30$s^{-1}$ the value found for *E. limosum* B2 remains radically higher. The highest value measured for the Fdh *in vitro* was 362$s^{-1}$ with methyl viologen as electron donor[196]. In the WLP, the Fdh was the only one where the synthesis rate was higher during heterotrophy than methylotrophy. Globally, protein concentrations were lower for other enzymes of the WLP in glucose compared to the methanol condition.

Interestingly, the $k_{app}$ values for the methenyl-THF cyclohydrolase (Mtc) and the 5,10-methylenetetrahydrofolate dehydrogenase (Mtd) were the lowest of the WLP with 3.3 $s^{-1}$ and 3.2 respectively in methanol condition. Furthermore, the transcript and protein copy numbers were higher than the other proteins involved in the methyl branch, suggesting a limitation by the catalytic capacity of the enzyme.

The $k_{app}$ of the formyl-THF synthetase (Fts) was estimated at 5.2 and 1.1 $s^{-1}$ for methanol and glucose conditions respectively. The Fts showed the lowest catalytic among WLP enzymes in glucose conditions. An apparent less-efficiency of Fts was also noticed for *C. autoethanogenum* under autotrophic condition, where the $k_{app}$ value measured was the lowest value of the WLP although the enzyme was among the most abundant protein[125]. In our experiments, the Mtc and Mtd showed the lowest $k_{app}$ in the WLP, suggesting a potential bottleneck in the methyl branch.

Regarding the methanol assimilation into the WLP, the average catalytic activity of the methanol-specific methyltransferase was measured at 0.9 $s^{-1}$, highlighting the relatively low efficiency of the complex to incorporate methanol into the central carbon metabolism. Also, the protein synthesis rate of the MTI protein was the highest value measured in the central carbon metabolism with 0,238$s^{-1}$, underlying the energetic cost necessary to integrate the methyl group into the WLP.

As expected, $k_{app}$ values were different between both conditions. The very low butyrate flux under the glucose condition was associated with lower $k_{app}$ values than during methylotrophic growth indicating that this pathway was substrate limited under heterotrophic conditions. However, the catalytic rates of these enzymes under methylotrophic conditions were low with for example 0.7 $s^{-1}$ for the 3-hydroxyacyl-CoA (Hbd) dehydrogenase and 0.6 $s^{-1}$ for the butyryl-Coa dehydrogenase-Etf complex (Bcd-Etf). Interestingly, the butyryl-CoA:acetate-Coa transferase showed a low $k_{app}$ value of 1,75 $s^{-1}$. Since no gene coding for butyrate kinase was found in the genome of *E. limosum* B2, the butyryl-CoA:acetate-Coa transferase most certainly plays a key role in butyrate biosynthesis. The CoA group from butyryl-CoA would be transferred to acetate, giving butyrate and acetyl-CoA.

On the contrary, under methylotrophic conditions the $k_{app}$ of the acetate kinase (Ack) and the phosphotransacetylase were much higher than enzymes involved in butyrogenesis with values of 15.3 and 9 $s^{-1}$. However, under heterotrophic conditions both values were much lower indicating that under this condition the enzyme was also substrate limited.

### 3.3.7. Absolute quantification brings light on the energy conservation mechanism and raises new questions

To try to decipher this uncertain parameter of the WLP, we simulated metabolic and electron fluxes in the GEM of *E. limosum* by considering MTHFR was either NADH or ferredoxin-dependent. Surprisingly, redox balance would not need active Rnf complex in methanol condition considering MTHFR as ferredoxin dependent. ATP would be essentially produced by substrate-level phosphorylation if MTHFR was NADH dependent, and Rnf complex would be used to oxidize the excess of NADH by importing $Na^+$ with a metabolic flux rate of 0.96 $mmol.gDCW^{-1}.h^{-1}$. According to the GEM, this option wouldn't be suitable to sustain growth on methanol.

Regarding the energy conservation system, the catalytic activity was estimated from the cytoplasmic subunits of Rnf and ATPase complexes although the confidence is relatively low due to the apparent difficulty in accurately quantifying membrane-bound proteins. The absolute number of RnfC subunits was around 2.8-fold higher in methanol condition with around 8150 and 2950 copy number of proteins per cell in methanol and glucose respectively.

Transcriptomic and proteomic data showed the biosynthesis of the Rnf complex for both tested conditions. These results raise the question of the very low Rnf fluxes suggested by GEM simulations. The RnfC, RnfB and RnfG subunits were present in equal amounts in glucose condition. Also, the Rnf subunits detected were around three times higher in methanol suggesting the use of the complex, questioning our assumption previously emitted.

## 3.4. Conclusion and perspectives

The absolute quantification of the total copy number of transcripts and proteins per cell was determined for the first time for an acetogen in methanol condition. Our approach was possible by i) standardization of growth rate by substrate-limited chemostat cultures at a defined growth rate, ii) the determination of physiological parameters of the cultures (products concentration, DCW, and cell density), iii) the use of a quantitative protein and RNA extraction methods with external standards spike-in mixes for transcriptomic and iv) an efficient analysis method by mass-spectrometry for proteomic and RNA-Seq for transcriptomic. Calculations of specific metabolic flux rates were determined by the GEM of *E. limosum* (*i*SP652). Finally, the apparent *in vivo* catalytic rates and *in vivo* synthesis rate values for each enzyme of the WLP was determined with the absolute copy number of each enzyme and transcripts per cell. Chemostat cultures at steady state showed an homoacetogenic fermentation during heterotrophy, while a mix of acetate and butyrate was produced during methylotrophy with a larger amount of carbon flux going to the butyrate.

Examination of the estimated copy number of proteins and transcripts per cell for ribosomal subunits showed a strong correlation between conditions, validating our extraction methods and our omics data set for further exploration at the absolute scale. The identifications of the strongest promoters and RBS will be useful to tune gene expression levels in the future.

We identified the methylene-THF cyclohydrolase as a potential flux rate-limiting enzyme of the WLP for methanol and glucose conditions. The MetV appeared to be translationally limited in the WLP as $k_x$ values were the lowest measured. Furthermore, the methanol-specific methyltransferase system appeared to be massively synthesized during methylotrophy involving an important energetic expenditure at the transcriptional and translational levels. A potential modification of the system by genetic engineering might enhance methanol assimilation.

Interestingly, the formate dehydrogenase (Fdh) showed a high catalytic rate capacity with a $k_{app}$ = 367 $s^{-1}$ in methanol condition. Formate is one of the promising $C_1$ feedstock to produce chemicals in a sustainable way [133]. The Fdh kinetic potential for formate assimilation reinforces the interest in such a carbon source, although only acetate was reported to be produced by *E.callanderi* KIST612[54] during formate fermentation. Regarding butyrate biosynthesis, three enzymes including crotonase, the butyryl-CoA dehydrogenase, and the 3-hydroxbutyryl-CoA dehydrogenase showed low catalytic efficiencies.

Importantly, omics data revealed a strong activity of the butyryl-CoA:acetate-Coa transferase in methanol condition suggesting the use of this enzyme to produce butyrate, as no gene coding for the butyrate kinase was found in the genome.

First fluxes simulations by the *E. limosum* GEM (*i*SP652) suggested a very low Rnf complex flux in methanol condition with an MTHFR ferredoxin dependent for the oxidation of methyl-THF to methylene-THF by the reverse use of methyl branch. A pool of reduced ferredoxin would be generated from the oxidation of methyl-THF to methylene-THF by the THFR and consumed during the production of acetyl-CoA by the acetyl-CoA synthase/carbon monoxide dehydrogenase complex in the carbonyl branch. Therefore, we hypothesized the formation of a complex between these enzymes.
In glucose condition, GEM simulations indicated a viable growth only with $NAD^+$ as the electron acceptor during the reduction of methylene-THF to methyl-THF by the MTHFR and inactivation of the Rnf complex. Since the MTHFR was described as ferredoxin-dependent only, $NAD^+$ reduction might be conferred by a missing protein. Our first hypothesis suggested an MTHFR-RnfC complex to ensure the reaction.

The data obtained highlight the methanol and glucose metabolisms of *E. limosum* B2. Our absolute quantification approach will serve as a model for future metabolism characterizations in various physiological conditions such as synthesis gases or substrates with different sizes of carbon chains. The data obtained will help for future rational metabolic engineering and improve the *E.limosum* GEM accuracy to optimize substrate assimilation and product yields.

# Conclusion and perspectives

Advances in technological tools reveal the increasing complexity of the metabolism. The goal of systems biology is first to expand our knowledge of the primary metabolism but also to help us to build *in silico* models that describe all physicochemical phenomena which occur in the system under different physiological conditions. At the cell level, genomic, transcriptomic, proteomic, and fluxomic data are required to accurately design models that fit reality as close as possible. Building *in silico* models of a microorganism's metabolism aimed to predict its phenotype under defined conditions and thus allow rational metabolic engineering to optimize the desired phenotype such as improving a product yield or maximizing a growth rate. The main asset of using the genome metabolic scale model is to direct research toward an objective faster and cheaper than previous approaches. Face climate change, finding alternatives to chemical processes by the development of new ways of production with sustainable resources, and pushing us forward into bioeconomy are more necessary than ever. The use of C1 feedstocks such as methanol to produce value-added products fits perfectly with the bioeconomy plan, and *E. limosum* is a promising candidate to transform methanol into C4 compounds such as butanol.

The use of metabolically engineered microorganisms in cell factories showed promising results at an industrial scale. Therefore, the use of GEMs to rationally improve production yield is crucial. Significant advances in biological models were made to date and are already used to improve production yield [121]. Typical GEM in acetogens includes 600-800 genes and 750-1000 reactions[197]. To further improve these models 1) elucidating the regulation, both at the transcriptional and the translational level, of genes encoding "metabolic enzymes" and 2) estimating the turnover constant of each of these enzymes are of paramount importance.

This thesis project is a part of the European BIOMETCHEM project which aims to use *E. limosum* as a microbial chassis to produce chemicals from methanol. The work performed in this thesis project aimed to characterize, by a systems biology approach, the central metabolism of *E. limosum* B2 growing in chemostat cultures on methanol or glucose as a carbon source. However, to perfectly characterize the metabolic reactions of the central carbon metabolism, the genomic sequence of the strain must be known, and the strain must be cultivated in a defined medium.

First, we sequenced, *de novo*, the complete genome of the WT strain of *E. limosum* B2 using PacBio technology. The sequence was annotated using the NCBI annotation pipeline, the complete sequence is now accessible under the accession number NZ_CP097376.1. Furthermore, the methylome profile of the WT strain was also deposited on the platform.

Then, we evolved a strain to grow on a methanol define medium by an ALE approach. The population was cultivated over 100 generations on a methanol define medium without yeast extract. We isolated three clones from the evolved population ($100^{th}$ generation) and we selected the clone which showed the

best growth performances. In addition, we managed to grow the selected clone on a methanol mineral medium without cysteine, with the addition of sodium thiosulfate. Such a phenotype was ideal to perfectly characterize the metabolism in a mineral medium with glucose or methanol as a carbon source. Batch fermentations showed that the wild-type strain can't grow on a cysteine-depleted mineral medium, suggesting key mutations had occurred in the adapted strains. We also observed significant improvements in the maximal specific growth rate and product yields. A 2.72-fold increase in the maximal growth rate was observed between the native population and the adapted strain in the methanol mineral medium cysteine free. Furthermore, the adapted strain was able to produce 43 ± 4 mC butyrate /100 mC MetOH in a methanol mineral medium, confirming the interest of the strain for $C_4$ compound production from $C_1$ methanol feedstock. To understand by which adaptive mechanism the strain improved its growth ability, we sequenced the genomes of the $10^{th}$, $25^{th}$, $50^{th}$, and $75^{th}$ generations and 3 isolated clones from the evolved population. Surprisingly, no specific mutation in central carbon metabolism nor conserved mutation over generations was noticed except in a large phage operon. Furthermore, the structure of type I RM encoding genes of the adapted strain was different compared to the WT strain. The system is composed of one R subunit, one M subunit, one integrase, and three S subunits. Among the S subunits annotated, two subunits seem truncated while one is complete. We found four repeated sequences including two repeated sequences in the complete S subunit and two other sequences in truncated S subunits in WT and adapted strain. Interestingly, we observed a homologous recombination event without mutation in these S subunits between the adapted and WT strains, giving a complete S subunit at the truncated S subunits zone and inversely. This recombination was not detected in the genome sequencing. After the alignment of protein sequences of both S subunits of WT and the adapted strain, the structure appeared different at the C terminus side of the protein. This modification on type I RM suggests the recognition of a new sequence by the complete S subunit of the adapted strain compared to the WT. After growing the adapted and the WT strain in a methanol synthetic medium with the addition of 0.5 $g.L^{-1}$ of yeast extract, the adapted strain showed a maximal growth rate 2.4-fold higher than the WT strain indicating a clear metabolic adaptation. We explored the total proteome of the WT and the adapted strain to find an explanation for this growth improvement. A significant total of 52 over-produced and 39 down-produced proteins showing a fold change < 2 or > 2 were observed in the adapted strain compared to the WT strain. Two enzymes involved in the primary metabolism, the carbon monoxide dehydrogenase (CooS) and the thiolase with Fc values of 2.22 and 2.08 respectively, and two proteins involved in gluconeogenesis, the pyruvate ferredoxin oxidoreductase and the glucose-6P isomerase with Fc values of 5.23 and 7.6 respectively were shown to be more produced in the adapted clone and potentially explaining the better growth on methanol. Interestingly, the anaerobic sulfite

reductase subunits A, B, and C were also more produced (Fc values, 3.81, 4.49, 5.12 respectively) in the adapted clone. This complex is known to play a key role in sulfate assimilation for cysteine biosynthesis. Mutations in the type I RM system were found to affect a large number of genes involved in the central metabolism of *B. subtilis*[168]. One possible explanation of the improved phenotype of the adapted clone might be that the methylated sequence of the mutated S subunit encoding gene of the type I RM is different.

After the obtention of an efficiently adapted clone growing on a methanol mineral medium, we focused our work on the characterization of its metabolism in methanol and glucose conditions. To standardize the physiological conditions, the cultures were performed in carbon limited chemostat with a specific growth rate set at 0.025 $h^{-1}$. Efforts were made to sustain growth, notably by equipment optimization to perfectly control the purity and the gas mixture composition. The physiological characteristics of the strain were determined by optical density measurement, cell density, dry cell weight, and product concentrations. The carbon balances and flux productions were determined for both conditions, showing an homoacetogenic production in the glucose condition and a mix of acetate and butyrate production for the methanol condition.

Fluxes of acids productions and substrate consumption were integrated into the GEM of *E. limosum* (*i*SP652) to predict specific fluxes for each metabolic reaction of the central carbon metabolism for methanol assimilation. Surprisingly, the Rnf complex was inactive for both conditions with an MTHFR ferredoxin dependent during methylotrophy and NADH dependent during heterotrophy. This metabolic specificity goes against theoretical energy conservation models established in methanol condition [70]. Therefore, hypothetic models were proposed suggesting the formation of a CODH/ACS-MTHFR complex for methanol condition and an RnfC-MTHFR complex in glucose condition.

The absolute quantification of transcript and protein copy number per cell was then investigated. Total proteome was extracted using the SPEED method with treatment by hot trifluoroacetic acid for cell lysis and protein solubilization. The protein concentration was determined using the Bradford assay, and samples were analyzed by mass spectrometry using DIA-NN software that exploits neural networks to accurately identify and quantify proteins.

To fully characterize the metabolism of *E. limosum* B2 on an absolute scale and bring light to underlying regulations which govern the cellular machinery, the absolute quantification of transcripts per cell by RNA-Seq was a crucial step. Many efforts were focused on this part to i) use an adapted method to ensure total RNA extraction, ii) preserve the integrity of RNA through the entire isolation process, iii)

use an efficient ribodepletion method. We found that the RNAsnap method was the most efficient to fully extract RNA while preserving them with a formamide-containing solution. We optimized the protocol to efficiently remove residual DNA in presence of a low amount of formamide allowing the recovery of enough RNA amounts to perform ribodepletion after column purification. The removal of ribosomal RNA is a critical step as it may affect transcriptomic profiles. The ribodepletion step was performed with a custom-made RiboPool kit adapted to *E. limosum* B2 transcriptome. To assess and consider the potential loss of transcripts through extraction steps and normalize the conditions between biological replicates, we added two sets of external spike-in standards. Spike-in mixes, constituted by RNA of different sizes and spanning a $10^6$ folds range concentration were added at two key steps of extraction. The first spike-in mix constituted of 6 RNA molecules synthesized *in vitro* and inspired by bacteriophage sequences, was added after cell lysis. The second mix, ERCC standards composed of 92 transcripts, was added to the purified RNA before ribodepletion. The total transcriptome was sequenced by RNA-Seq and the first analysis based on read counts showed a great correlation between the number of counted reads and the expected concentration for the first and second spike-in mixes.

To assess the efficiency of the extraction methods, we analyzed the correlation between the expected equimolar protein complexes and the absolute number of protein subunits copy number per cell. The results were positively correlated for 14 protein complexes analyzed. Furthermore, the number of total ribosomal proteins per cell, which are known to be growth rate dependent, was similar between both physiological conditions. The quality of transcriptomic results was assessed by comparing transcript copy numbers for ribosomal protein for both conditions. A strong correlation was found between absolute transcript copy numbers of ribosomal protein for both conditions. Also, a strong correlation between the absolute copy number of proteins and transcripts for ribosomal subunits was found for each condition. These results comforted us with the extraction and standardization method used in our experiment. Then, we investigated the extraction efficiency of membrane-bound proteins from Rnf and ATPase complexes. However, the absolute number of subunits per cell did not match the expected stoichiometry, suggesting lower reliability on membranous proteins.

A total of $14.1 \pm 2.53 \times 10^6$ copy number per cell in glucose condition and $14.0 \pm 1.72 \times 10^6$ copy number per cell in methanol condition was estimated. For the transcriptome, a total of $9621 \pm 561$ and $10519 \pm 1021$ transcripts number per cell were found in heterotrophic and methylotrophic conditions respectively.

Absolute quantification of protein and transcripts copy number per cell along flux estimation by simulation, open access to the *in vivo* catalytic rate ($k_{app}$) and the *in vivo* protein synthesis ($k_x$) for the main enzymes of the central carbon metabolism. Our results showed the highest catalytic rate for the formate dehydrogenase for methanol condition while other enzymes of the WLP showed much lower values, specifically the methenyl-THF cyclohydrolase. In the glucose condition, the pyruvate ferredoxin oxidoreductase showed the highest catalytic rate, while enzymes of the WLP were weakly active compared to the methanol condition, which is consistent with the absolute number of proteins estimated for both conditions. Interestingly, the butyryl-COA:acetate-Coa transferase was highly transcripted and translated in methanol condition. As no gene coding for butyrate kinase was found in the genome, the butyryl-COA:acetate-Coa transferase is most likely involved in butyrate biosynthesis by transferring CoA from butyryl-CoA to acetate, giving butyrate and acetyl-CoA.

Analysis of the protein synthesis rates showed a high translation rate for the Acetyl-Coa synthase, the MetF subunit from MTHFR, and the Bcd-Etf complex. Methylotrophic condition revealed strong RBS as one of the highest $k_x$ values was found for the methanol-specific methyltransferase MTI and calculable only in methanol conditions. Also, several proteins involved in the *pdu* operon including PduA and PduB which code for bacterial microcompartments shell (BMC) were translated at a high rate.. Furthermore, these proteins were among the most concentrated molecules per cell for both conditions, suggesting an important role of BMCs. The massive abundance of these proteins remained unknown and may concern culture conditions in mineral medium combined to the low dilution rate at $0{,}025h^{-1}$ used for steady-state chemostats.

## Perspectives

The work achieved allowed the development of a new systems biology approach for primary metabolism characterization based on absolute quantification of both protein and mRNA. Concerning the first part of the thesis, the methylome of WT and the adapted strain of *E. limosum* could be studied on methanol and glucose condition. The use of Nanopore technology should allow the identification of m5C and m4A methylated sites in addition to m6A previously determined for the WT with PacBio technology. Such work would bring information on the impact of methylome on adaptation mechanisms as few studies were performed on bacteria.

Regulation mechanisms respond to a complex dynamic and govern the cell phenotypes. Therefore, their understanding is crucial for GEM-built and rational metabolic engineering purposes. For example, a study already explored regulation mechanisms such as ribosome profiling and 5'UTR identification to

study the post-transcriptional regulations on *E. limosum* ATCC 8486 [122]. Furthermore, studies underlying the importance of metabolite concentration and protein acetylation on global metabolism[198].

The energy conservation mechanism of the strain will be further explored, notably concerning the MTHFR mechanism in methanol and glucose conditions. Currently, knockout mutants *rnfC* and *rnfEAB* are under construction to verify if the Rnf complex might be inactivated while maintaining growth during methylotrophy on methanol and heterotrophy on glucose. The answer should bring strong evidence to support or revise our models.

Since a very copy number of proteins for RnfC was quantified in glucose condition, its role with the MTHFR might be compromised. If RnfC is not involved with the MTHFR in heterotrophic conditions, future investigations might focus on the identification of proteins carrying NADH binding domain which might potentially interact with the MTHFR. Interestingly, the PduS protein, showing similarities at the N-terminal part to the RnfC subunit with an NADH binding domain, was four times more concentrated per cell in heterotrophic conditions. Furthermore, other proteins such as rubredoxin, thioredoxin, or glycine cleavage system protein H were proposed as potential candidates[71]. A work on gene essentiality was performed on *E. limosum* on methanol and glucose conditions by Elizabeth Redfern (unpublished data, University of Nottingham), the results should also bring answers to our questions. Also, a new flux simulation will be performed on the GEM considering the MTHFR-Rnf complex.

Regarding the potential of *E. limosum* B2 to produce C4 compounds from C1 molecules, omics evidence of butyryl-Coa:acetate-Coa transferase activity during methylotrophy strongly suggests that butyrate is produced through the transfer of CoA from butyryl-Coa to acetate. Therefore, deletion of the phosphotransacetylase might lead to homobutyric production using a mix of methanol, $CO_2$, and acetate as carbon sources.

The methodology developed through the study of the primary metabolism of methanol or glucose-grown cells could be extended to other substrates with various carbon content. Continuous culture on lactate at a similar growth rate than previous conditions was already performed. Furthermore, a similar approach will be performed in autotrophic conditions. Combining proteomic and transcriptomic absolute profiles with additional omics techniques such as ribosome profiling, 5'UTR identification, methylome characterization, and so on would bring tremendous advances in GEM building. For the case of *E. limosum* using methanol as C1 feedstock, understanding its metabolism will be very useful for metabolic

engineering and further improve the acetone and butanol yields in the work initiated by Flaiz et al, within the BIOMETCHEM project [64].

# Bibliography


1. Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica | Nature. https://www.nature.com/articles/20859.
2. Earth's CO2 Home Page. https://www.co2.earth/.
3. Climate change: a threat to human wellbeing and health of the planet. Taking action now can secure our future — IPCC. https://www.ipcc.ch/2022/02/28/pr-wgii-ar6/.
4. Projects and publications relating to plastics. https://ellenmacarthurfoundation.org/topics/plastics/projects.
5. Golwala, H., Zhang, X., Iskander, S. M. & Smith, A. L. Solid waste: An overlooked source of microplastics to the environment. *Sci. Total Environ.* **769**, 144581 (2021).
6. Lee, R. P., Keller, F. & Meyer, B. A concept to support the transformation from a linear to circular carbon economy: net zero emissions, resource efficiency and conservation through a coupling of the energy, chemical and waste management sectors. *Clean Energy* **1**, 102–113 (2017).
7. Steynberg, A. P. Chapter 1 - Introduction to Fischer-Tropsch Technology. in *Studies in Surface Science and Catalysis* (eds. Steynberg, A. & Dry, M.) vol. 152 1–63 (2004).
8. Drake, H. L., Küsel, K. & Matthies, C. Acetogenic Prokaryotes. in *The Prokaryotes: Prokaryotic Physiology and Biochemistry* (eds. Rosenberg, E., DeLong, E. F., Lory, S., Stackebrandt, E. & Thompson, F.) 3–60 (2013).
9. Daniell, J., Köpke, M. & Simpson, S. D. Commercial Biomass Syngas Fermentation. *Energies* **5**, 5372–5417 (2012).

10. Henstra, A. M., Sipma, J., Rinzema, A. & Stams, A. J. M. Microbiology of synthesis gas fermentation for biofuel production. *Curr. Opin. Biotechnol.* **18**, 200–206 (2007).
11. F, F. Biologische Gasreaktionen. II. Uber die Bildung von Essigsaure bei der Biologischen Umsetzung von Kohlenoxyd und Kohlensa mit Wasserstoff zn Methan. *Biochem Z* **245**, 2–12 (1932).
12. Wieringa, K. T. Over het verdwijnen van waterstof en koolzuur onder anaerobe voorwaarden. *Antonie van Leeuwenhoek, 3(1)* 263–273 (1936).
13. Fontaine, F. E., Peterson, W. H., McCoy, E., Johnson, M. J. & Ritter, G. J. A New Type of Glucose Fermentation by *Clostridium thermoaceticum*. *J. Bacteriol.* **43**, 701–715 (1942).
14. Barker, H. A. On the Rôle of Carbon Dioxide in the Metabolism of *Clostridium Thermoaceticum*. *Proc. Natl. Acad. Sci. U. S. A.* **30**, 88–90 (1944).
15. Wood, H. G., Ragsdale, S. W. & Pezacka, E. The acetyl-CoA pathway of autotrophic growth. *FEMS Microbiol. Rev.* **2**, 345–362 (1986).
16. Ljungdhal, L. G. The Autotrophic Pathway of Acetate Synthesis in Acetogenic Bacteria. *Annu. Rev. Microbiol.* **40**, 415–450 (1986).
17. Schuchmann, K. & Müller, V. Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria. *Nat. Rev. Microbiol.* **12**, 809–821 (2014).
18. Bengelsdorf, F. R. *et al.* Industrial Acetogenic Biocatalysts: A Comparative Metabolic and Genomic Analysis. *Front. Microbiol.* **7**, (2016).
19. Bengelsdorf, F. R. *et al.* Chapter Four - Bacterial Anaerobic Synthesis Gas (Syngas) and CO2+H2 Fermentation. in *Advances in Applied Microbiology* (eds. Sariaslani, S. & Gadd, G. M.) vol. 103 143–221 (Academic Press, 2018).
20. Kremp, F., Poehlein, A., Daniel, R. & Müller, V. Methanol metabolism in the acetogenic bacterium *Acetobacterium woodii*. *Environ. Microbiol.* **20**, 4369–4384 (2018).

21. Pacaud, S., Loubiere, P. & Goma, G. Methanol metabolism by *Eubacterium limosum* B2: Effects of pH and carbon dioxide on growth and organic acid production. *Curr. Microbiol.* **12**, 245–250 (1985).

22. Wood, J. C., Marcellin, E., Plan, M. R. & Virdis, B. High methanol-to-formate ratios induce butanol production in *Eubacterium limosum*. *Microb. Biotechnol.* (2021) doi:10.1111/1751-7915.13963.

23. Arslan, K., Veiga, M. C. & Kennes, C. Autotrophic (C1-gas) versus heterotrophic (fructose) accumulation of acetic acid and ethanol in *Clostridium aceticum*. *Bioresour. Technol.* **337**, 125485 (2021).

24. Jia, D. *et al.* Metabolic Engineering of Gas-Fermenting *Clostridium ljungdahli*i for Efficient Co-production of Isopropanol, 3-Hydroxybutyrate, and Ethanol. *ACS Synth. Biol.* **10**, 2628–2638 (2021).

25. Brown, S. D. *et al.* Comparison of single-molecule sequencing and hybrid approaches for finishing the genome of *Clostridium autoethanogenum* and analysis of CRISPR systems in industrial relevant *Clostridia*. *Biotechnol. Biofuels* **7**, 40 (2014).

26. Charubin, K. & Papoutsakis, E. T. Direct cell-to-cell exchange of matter in a synthetic *Clostridium* syntrophy enables CO2 fixation, superior metabolite yields, and an expanded metabolic space. *Metab. Eng.* **52**, 9–19 (2019).

27. Beijing Shougang LanzaTech New Energy Technology CCU plant achieves RSB Global Advanced Products Standard. *Bioenergy International* https://bioenergyinternational.com/beijing-shougang-lanzatech-new-energy-technology-ccu-plant-achieves-rsb-global-advanced-products-standard/ (2021).

28. Schuchmann, K. & Müller, V. Direct and reversible hydrogenation of CO2 to formate by a bacterial carbon dioxide reductase. *Science* **342**, 1382–1385 (2013).

29. Müller, V. New Horizons in Acetogenic Conversion of One-Carbon Substrates and Biological Hydrogen Storage. *Trends Biotechnol.* **37**, 1344–1354 (2019).

30. Genthner, B. R., Davis, C. L. & Bryant, M. P. Features of rumen and sewage sludge strains of *Eubacterium limosum*, a methanol- and H2-CO2-utilizing species. *Appl. Environ. Microbiol.* **42**, 12–19 (1981).

31. Sharak Genthner, B. R. & Bryant, M. P. Additional characteristics of one-carbon-compound utilization by *Eubacterium limosum* and *Acetobacterium woodii*. *Appl. Environ. Microbiol.* **53**, 471–476 (1987).

32. Kelly, W. J. *et al.* The complete genome sequence of *Eubacterium limosum* SA11, a metabolically versatile rumen acetogen. *Stand. Genomic Sci.* **11**, (2016).

33. Hur, H.-G. & Rafii, F. Biotransformation of the isoflavonoids biochanin A, formononetin, and glycitein by *Eubacterium limosum*. *FEMS Microbiol. Lett.* **192**, 21–25 (2000).

34. Kountz, D. J., Behrman, E. J., Zhang, L. & Krzycki, J. A. MtcB, a member of the MttB superfamily from the human gut acetogen *Eubacterium limosum*, is a cobalamin-dependent carnitine demethylase. *J. Biol. Chem.* **295**, 11971–11981 (2020).

35. Pacaud, S., Loubiere, P., Goma, G. & Lindley, N. D. Effects of various organic acid supplements on growth rates of *Eubacterium limosum* B2 on methanol. *Appl. Microbiol. Biotechnol.* **24**, 75–78 (1986).

36. Mäkelä, S. *et al.* Inhibition of 17beta-hydroxysteroid oxidoreductase by flavonoids in breast and prostate cancer cells. *Proc. Soc. Exp. Biol. Med. Soc. Exp. Biol. Med. N. Y. N* **217**, 310–316 (1998).

37. Arora, A., Nair, M. G. & Strasburg, G. M. Antioxidant activities of isoflavones and their biological metabolites in a liposomal system. *Arch. Biochem. Biophys.* **356**, 133–141 (1998).

38. Middleton, E. & Kandaswami, C. Effects of flavonoids on immune and inflammatory cell functions. *Biochem. Pharmacol.* **43**, 1167–1179 (1992).

39. Kanauchi, O. *et al. Eubacterium limosum* ameliorates experimental colitis and metabolite of microbe attenuates colonic inflammatory action with increase of mucosal integrity. *World J. Gastroenterol.* **12**, 1071–1077 (2006).

40. Biagi, E. *et al.* Through Ageing, and Beyond: Gut Microbiota and Inflammatory Status in Seniors and Centenarians. *PLOS ONE* **5**, e10667 (2010).

41. Peled, J. U. *et al.* Intestinal Microbiota and Relapse After Hematopoietic-Cell Transplantation. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **35**, 1650–1659 (2017).

42. Krumholz, L. R. & Bryant, M. P. Y. *Clostridium pfennigii* sp. nov. Uses Methoxyl Groups of Monobenzenoids and Produces Butyrate. *Int. J. Syst. Evol. Microbiol.* **35**, 454–456 (1985).

43. Liou, J. S.-C., Balkwill, D. L., Drake, G. R. & Tanner, R. S. *Clostridium carboxidivorans* sp. nov., a solvent-producing clostridium isolated from an agricultural settling lagoon, and reclassification of the acetogen *Clostridium scatologenes* strain SL1 as *Clostridium drakei* sp. nov. *Int. J. Syst. Evol. Microbiol.* **55**, 2085–2091 (2005).

44. Park, S. *et al.* Acetate-assisted increase of butyrate production by *Eubacterium limosum* KIST612 during carbon monoxide fermentation. *Bioresour. Technol.* **245**, 560–566 (2017).

45. Demirbas, A. Biomethanol Production from Organic Waste Materials. *Energy Sources Part Recovery Util. Environ. Eff.* **30**, 565–572 (2008).

46. Hank, C. *et al.* Economics & carbon dioxide avoidance cost of methanol production based on renewable hydrogen and recycled carbon dioxide – power-to-methanol. *Sustain. Energy Fuels* **2**, 1244–1261 (2018).

47. Bergins, C. *et al.* A Technology Review and Cost Analysis of the Production of Low Carbon Methanol and Following Methanol to Gasoline Process. in *Zukünftige Kraftstoffe: Energiewende des Transports als ein weltweites Klimaziel* (ed. Maus, W.) 433–463 (Springer, 2019).

48. Woolston, B. M., King, J. R., Reiter, M., Van Hove, B. & Stephanopoulos, G. Improving formaldehyde consumption drives methanol assimilation in engineered *E. coli*. *Nat. Commun.* **9**, 2387 (2018).

49. Schrader, J. *et al.* Methanol-based industrial biotechnology: current status and future perspectives of methylotrophic bacteria. *Trends Biotechnol.* **27**, 107–115 (2009).

50. Pfeifenschneider, J., Brautaset, T. & Wendisch, V. F. Methanol as carbon substrate in the bio-economy: Metabolic engineering of aerobic methylotrophic bacteria for production of value-added chemicals. *Biofuels Bioprod. Biorefining* **11**, 719–731 (2017).

51. Bache, R. & Pfennig, N. Selective isolation of *Acetobacterium woodii* on methoxylated aromatic acids and determination of growth yields. *Arch. Microbiol.* **130**, 255–261 (1981).

52. Tremblay, P.-L., Höglund, D., Koza, A., Bonde, I. & Zhang, T. Adaptation of the autotrophic acetogen *Sporomusa ovata* to methanol accelerates the conversion of CO2 to organic products. *Sci. Rep.* **5**, 16168 (2015).

53. Kim, H., Jeon, B. S. & Sang, B.-I. An Efficient New Process for the Selective Production of Odd-Chain Carboxylic Acids by Simple Carbon Elongation Using *Megasphaera hexanoica*. *Sci. Rep.* **9**, 11999 (2019).

54. Litty, D. & Müller, V. Butyrate production in the acetogen *Eubacterium limosum* is dependent on the carbon and energy source. *Microb. Biotechnol.* **14(6)**, 2686-2692 (2021).

55. Dwidar, M., Park, J.-Y., Mitchell, R. J. & Sang, B.-I. The Future of Butyric Acid in Industry. *The Scientific World Journal* vol. 2012 e471417 https://www.hindawi.com/journals/tswj/2012/471417/ (2012).

56. Chen, Z.-X. & Breitman, T. R. Tributyrin: A Prodrug of Butyric Acid for Potential Clinical Application in Differentiation Therapy. *Cancer Res.* **54**, 3494–3499 (1994).

57. Cao, Y., Li, H. & Zhang, J. Homogeneous Synthesis and Characterization of Cellulose Acetate Butyrate (CAB) in 1-Allyl-3-Methylimidazolium Chloride (AmimCl) Ionic Liquid. *Ind. Eng. Chem. Res.* **50**, 7808–7814 (2011).

58. Wang, J., Lin, M., Xu, M. & Yang, S.-T. Anaerobic Fermentation for Production of Carboxylic Acids as Bulk Chemicals from Renewable Biomass. *Adv. Biochem. Eng. Biotechnol.* **156**, 323–361 (2016).

59. Kroschwitz, J. I. & Howe-Grant, M. *Kirk-Othmer encyclopedia of chemical technology*. (John Wiley, 1991).

60. Dischert, W., Raynaud, C., Dumon-Seignovert, L. & Dumoulin, N. Microorganisms with improved 1,3-propanediol and butyric acid production. (2020).

61. Song, Y. & Cho, B.-K. Draft Genome Sequence of Chemolithoautotrophic Acetogenic Butanol-Producing *Eubacterium limosum* ATCC 8486. *Genome Announc.* **3**, (2015).

62. Lynd, L. H. & Zeikus, J. G. Metabolism of H2-CO2, methanol, and glucose by *Butyribacterium methylotrophicum*. *J. Bacteriol.* **153**, 1415–1423 (1983).

63. Dürre, P. Butanol formation from gaseous substrates. *FEMS Microbiol. Lett.* **363**, fnw040 (2016).

64. Flaiz, M., Ludwig, G., Bengelsdorf, F. R. & Dürre, P. Production of the biocommodities butanol and acetone from methanol with fluorescent FAST-tagged proteins using metabolically engineered strains of *Eubacterium limosum*. *Biotechnol. Biofuels* **14**, 117 (2021).

65. Sanford, P. A. & Woolston, B. M. Synthetic or natural? Metabolic engineering for assimilation and valorization of methanol. *Curr. Opin. Biotechnol.* **74**, 171–179 (2022).

66. Ragsdale, S. W. Pyruvate Ferredoxin Oxidoreductase and Its Radical Intermediate. *Chem. Rev.* **103**, 2333–2346 (2003).

67. Poehlein, A. *et al.* An Ancient Pathway Combining Carbon Dioxide Fixation with the Generation and Utilization of a Sodium Ion Gradient for ATP Synthesis. *PLOS ONE* **7**, e33439 (2012).

68. Drake, H. L., Gössner, A. S. & Daniel, S. L. Old acetogens, new light. *Ann. N. Y. Acad. Sci.* **1125**, 100–128 (2008).

69. Schirwitz, K., Schmidt, A. & Lamzin, V. S. High-resolution structures of formate dehydrogenase from *Candida boidinii*. *Protein Sci. Publ. Protein Soc.* **16**, 1146–1156 (2007).

70. Dietrich, H. M., Kremp, F., Öppinger, C., Ribaric, L. & Müller, V. Biochemistry of methanol-dependent acetogenesis in *Eubacterium callanderi* KIST612. *Environ. Microbiol.* **23**, 4505–4517 (2021).

71. Öppinger, C., Kremp, F. & Müller, V. Is reduced ferredoxin the physiological electron donor for MetVF-type methylenetetrahydrofolate reductases in acetogenesis? A hypothesis. *Int. Microbiol.* **25**, 75–88 (2022).

72. Doukov, T. I., Iverson, T. M., Seravalli, J., Ragsdale, S. W. & Drennan, C. L. A Ni-Fe-Cu center in a bifunctional carbon monoxide dehydrogenase/acetyl-CoA synthase. *Science* **298**, 567–572 (2002).

73. Can, M., Armstrong, F. A. & Ragsdale, S. W. Structure, Function, and Mechanism of the Nickel Metalloenzymes, CO Dehydrogenase, and Acetyl-CoA Synthase. *Chem. Rev.* **114**, 4149–4174 (2014).

74. Ragsdale, S. W., Lindahl, P. A. & Münck, E. Mössbauer, EPR, and optical studies of the corrinoid/iron-sulfur protein involved in the synthesis of acetyl coenzyme A by *Clostridium thermoaceticum*. *J. Biol. Chem.* **262**, 14289–14297 (1987).

75. Loubière, P., Gros, E., Paquet, V. & Lindley, N. D. Kinetics and physiological implications of the growth behaviour of *Eubacterium limosum* on glucose/methanol mixtures. *Microbiology* **138**, 979–985 (1992).

76. Li, F. *et al.* Coupled Ferredoxin and Crotonyl Coenzyme A (CoA) Reduction with NADH Catalyzed by the Butyryl-CoA Dehydrogenase/Etf Complex from *Clostridium kluyveri*. *J. Bacteriol.* **190**, 843–850 (2008).

77. Jeong, J. *et al.* Energy Conservation Model Based on Genomic and Experimental Analyses of a Carbon Monoxide-Utilizing, Butyrate-Forming Acetogen, *Eubacterium limosum* KIST612. *Appl. Environ. Microbiol.* **81**, 4782–4790 (2015).

78. Zhang, J. *et al.* Metabolic engineering of *Clostridium tyrobutyricum* for n-butanol production from sugarcane juice. *Appl. Microbiol. Biotechnol.* **101**, 4327–4337 (2017).

79. Louis, P. *et al.* Restricted Distribution of the Butyrate Kinase Pathway among Butyrate-Producing Bacteria from the Human Colon. *J. Bacteriol.* **186**, 2099–2106 (2004).

80. Drake, H. L. *Acetogenesis*. (Springer US, 1994). doi:10.1007/978-1-4615-1777-1.

81. Martin, D. R., Misra, A. & Drake, H. L. Dissimilation of Carbon Monoxide to Acetic Acid by Glucose-Limited Cultures of *Clostridium thermoaceticum*. *Appl. Environ. Microbiol.* **49**, 1412–1417 (1985).

82. Braus-Stromeyer, S. A., Schnappauf, G., Braus, G. H., Gössner, A. S. & Drake, H. L. Carbonic anhydrase in *Acetobacterium woodii* and other acetogenic bacteria. *J. Bacteriol.* **179**, 7197–7200 (1997).

83. Le Bloas, P., Guilbert, N., Loubière, P. & Lindley, N. D. Growth inhibition and pyruvate overflow during glucose metabolism of *Eubacterium limosum* are related to a limited capacity to reassimilate CO2 by the acetyl-CoA pathway. *Microbiology* **139**, 1861–1868 (1993).

84. Biegel, E., Schmidt, S., González, J. M. & Müller, V. Biochemistry, evolution and physiological function of the Rnf complex, a novel ion-motive electron transport complex in prokaryotes. *Cell. Mol. Life Sci.* **68**, 613–634 (2011).

85. Imkamp, F., Biegel, E., Jayamani, E., Buckel, W. & Müller, V. Dissection of the Caffeate Respiratory Chain in the Acetogen *Acetobacterium woodii*: Identification of an Rnf-Type NADH Dehydrogenase as a Potential Coupling Site. *J. Bacteriol.* **189**, 8145–8153 (2007).

86. Hess, V., Schuchmann, K. & Müller, V. The Ferredoxin:NAD+ Oxidoreductase (Rnf) from the Acetogen *Acetobacterium woodi*i Requires Na+ and Is Reversibly Coupled to the Membrane Potential. *J. Biol. Chem.* **288**, 31496–31502 (2013).

87. Buckel, W. & Thauer, R. K. Energy conservation via electron bifurcating ferredoxin reduction and proton/Na+ translocating ferredoxin oxidation. *Biochim. Biophys. Acta BBA - Bioenerg.* **1827**, 94–113 (2013).

88. Song, Y. *et al.* Determination of the Genome and Primary Transcriptome of Syngas Fermenting *Eubacterium limosum* ATCC 8486. *Sci. Rep.* **7**, (2017).

89. Westphal, L., Wiechmann, A., Baker, J., Minton, N. P. & Müller, V. The Rnf Complex Is an Energy-Coupled Transhydrogenase Essential To Reversibly Link Cellular NADH and Ferredoxin Pools in the Acetogen *Acetobacterium woodii*. *J. Bacteriol.* **200**, e00357-18 (2018).

90. Köpke, M. *et al. Clostridium ljungdahlii* represents a microbial production platform based on syngas. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 13087–13092 (2010).

91. Schmehl, M. *et al.* Identification of a new class of nitrogen fixation genes in *Rhodobacter capsulatus*: a putative membrane complex involved in electron transport to nitrogenase. *Mol. Gen. Genet. MGG* **241**, 602–615 (1993).

92. Kuhns, M. *et al.* The Rnf complex from the acetogenic bacterium *Acetobacterium woodii*: Purification and characterization of RnfC and RnfB. *Biochim. Biophys. Acta BBA - Bioenerg.* **1861**, 148263 (2020).

93. Backiel, J. *et al.* Covalent Binding of Flavins to RnfG and RnfD in the Rnf Complex from *Vibrio cholerae*. *Biochemistry* **47**, 11273–11284 (2008).

94. Biegel, E., Schmidt, S. & Müller, V. Genetic, immunological and biochemical evidence for a Rnf complex in the acetogen *Acetobacterium woodii*. *Environ. Microbiol.* **11**, 1438–1443 (2009).

95. Westphal, L., Wiechmann, A., Baker, J., Minton, N. P. & Müller, V. The Rnf Complex Is an Energy-Coupled Transhydrogenase Essential To Reversibly Link Cellular NADH and Ferredoxin Pools in the Acetogen *Acetobacterium woodii*. *J. Bacteriol.* **200**, e00357-18 (2018).

96. Hedderich, R. & Forzi, L. Energy-converting [NiFe] hydrogenases: more than just $H_2$ activation. *J. Mol. Microbiol. Biotechnol.* **10**, 92–104 (2005).

97. Schoelmerich, M. C. & Müller, V. Energy conservation by a hydrogenase-dependent chemiosmotic mechanism in an ancient metabolic pathway. *Proc. Natl. Acad. Sci.* **116**, 6329–6334 (2019).

98. Meuer, J., Bartoschek, S., Koch, J., Künkel, A. & Hedderich, R. Purification and catalytic properties of Ech hydrogenase from *Methanosarcina barkeri*. *Eur. J. Biochem.* **265**, 325–335 (1999).

99. Litty, D. & Müller, V. A $Na^+$ $A_1A_0$ ATP synthase with a V-type *c* subunit in a mesophilic bacterium. *FEBS J.* **287**, 3012–3023 (2020).

100. Müller, V. & Grüber, G. ATP synthases: structure, function and evolution of unique energy converters. *Cell. Mol. Life Sci. CMLS* **60**, 474–494 (2003).

101. Martin, W. F. Hydrogen, metals, bifurcating electrons, and proton gradients: The early evolution of biological energy conservation. *FEBS Lett.* **586**, 485–493 (2012).

102. Thauer, R. K., Jungermann, K. & Decker, K. Energy conservation in chemotrophic anaerobic bacteria. *Bacteriol. Rev*, 100-180 (1977).

103. Mitchell, P. The protonmotive Q cycle: a general formulation. *FEBS Lett.* **59**, 137–139 (1975).

104. Chowdhury, N. P. *et al.* Studies on the Mechanism of Electron Bifurcation Catalyzed by Electron Transferring Flavoprotein (Etf) and Butyryl-CoA Dehydrogenase (Bcd) of *Acidaminococcus fermentans*. *J. Biol. Chem.* **289**, 5145–5157 (2014).

105. Schut, G. J. & Adams, M. W. W. The Iron-Hydrogenase of *Thermotoga maritima* Utilizes Ferredoxin and NADH Synergistically: a New Perspective on Anaerobic Hydrogen Production. *J. Bacteriol.* **191**, 4451–4457 (2009).

106. Elbehti, A., Brasseur, G. & Lemesle-Meunier, D. First Evidence for Existence of an Uphill Electron Transfer through the bc1 and NADH-Q Oxidoreductase Complexes of the Acidophilic Obligate Chemolithotrophic Ferrous Ion-Oxidizing Bacterium *Thiobacillus ferrooxidans*. *J. Bacteriol.* **182**, 3602–3606 (2000).

107. Schuchmann, K. & Müller, V. A bacterial electron-bifurcating hydrogenase. *J. Biol. Chem.* **287**, 31165–31171 (2012).

108. Weghoff, M. C., Bertsch, J. & Müller, V. A novel mode of lactate metabolism in strictly anaerobic bacteria. *Environ. Microbiol.* **17**, 670–677 (2015).

109. Müller, V., Chowdhury, N. P. & Basen, M. Electron Bifurcation: A Long-Hidden Energy-Coupling Mechanism. *Annu. Rev. Microbiol.* **72**, 331–353 (2018).

110. Nicolet, Y., Cavazza, C. & Fontecilla-Camps, J. C. Fe-only hydrogenases: structure, function and evolution. *J. Inorg. Biochem.* **91**, 1–8 (2002).

111. Albracht, S. P. Nickel hydrogenases: in search of the active site. *Biochim. Biophys. Acta* **1188**, 167–204 (1994).

112. Vignais, P. M. & Billoud, B. Occurrence, classification, and biological function of hydrogenases: an overview. *Chem. Rev.* **107**, 4206–4272 (2007).

113. Fontecilla-Camps, J. C. & Ragsdale, S. W. Nickel–Iron–Sulfur Active Sites: Hydrogenase and Co Dehydrogenase. in *Advances in Inorganic Chemistry* (ed. Sykes, A. G.) vol. 47 283–333 (Academic Press, 1999).

114. Volbeda, A. *et al.* Crystal structure of the nickel–iron hydrogenase from *Desulfovibrio gigas*. *Nature* **373**, 580–587 (1995).

115. Hedderich, R. Energy-converting [NiFe] hydrogenases from archaea and extremophiles: ancestors of complex I. *J. Bioenerg. Biomembr.* **36**, 65–75 (2004).

116. Wang, S., Huang, H., Kahnt, J. & Thauer, R. K. A Reversible Electron-Bifurcating Ferredoxin- and NAD-Dependent [FeFe]-Hydrogenase (HydABC) in *Moorella thermoacetica*. *J. Bacteriol.* **195**, 1267–1275 (2013).

117. Kremp, F. & Müller, V. Methanol and methyl group conversion in acetogenic bacteria: biochemistry, physiology and application. *FEMS Microbiol. Rev.* **45**, fuaa040 (2021).

118. Loubière, P. & Lindley, N. D. The use of acetate as an additional co-substrate improves methylotrophi growth of the acetogenic anaerobe *Eubacterium limosum* when CO2 fixation is rate-limiting. *Microbiology* **137**, 2247–2251 (1991).

119. Hyduke, D. R., Lewis, N. E. & Palsson, B. Ø. Analysis of omics data with genome-scale models of metabolism. *Mol. Biosyst.* **9**, 167–174 (2013).

120. Lerman, J. A. *et al.* In silico method for modelling metabolism and gene product expression at genome scale. *Nat. Commun.* **3**, (2012).

121. Norman, R. O. J. *et al.* Genome-scale model of *C. autoethanogenum* reveals optimal bioprocess conditions for high-value chemical production from carbon monoxide. *Eng. Biol.* **3**, 32–40 (2019).

122. Song, Y. *et al.* Genome-scale analysis of syngas fermenting acetogenic bacteria reveals the translational regulation for its autotrophic growth. *BMC Genomics* **19**, (2018).

123. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S. & Weissman, J. S. Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223 (2009).

124. Kim, J.-Y. *et al.* Methanol supply speeds up synthesis gas fermentation by methylotrophic-acetogenic bacterium, *Eubacterium limosum* KIST612. *Bioresour. Technol.* **321**, 124521 (2021).

125. Valgepea, K. *et al.* Absolute Proteome Quantification in the Gas-Fermenting Acetogen *Clostridium autoethanogenum*. *mSystems* **7**, e00026-22 (2022).

126. Gorochowski, T. E. *et al.* Absolute quantification of translational regulation and burden using combined sequencing approaches. *Mol. Syst. Biol.* **15**, (2019).

127. Bartholomäus, A. *et al.* Bacteria differently regulate mRNA abundance to specifically respond to various stresses. *Philos. Transact. A Math. Phys. Eng. Sci.* **374**, (2016).

128. Gorochowski, T. E. *et al.* Genetic circuit characterization and debugging using RNA-seq. *Mol. Syst. Biol.* **13**, (2017).

129. Dwidar, M., Park, J.-Y., Mitchell, R. J. & Sang, B.-I. The Future of Butyric Acid in Industry. *Sci. World J.* **2012**, 471417 (2012).

130. Karlson, B., Bellavitis, C. & France, N. Commercializing LanzaTech, from waste to fuel: An effectuation case. *J. Manag. Organ.* **27**, 175–196 (2021).

131. Bertsch, J. & Müller, V. Bioenergetic constraints for conversion of syngas to biofuels in acetogenic bacteria. *Biotechnol. Biofuels* **8**, 210 (2015).

132. Giuliano, A., Freda, C. & Catizzone, E. Techno-Economic Assessment of Bio-Syngas Production for Methanol Synthesis: A Focus on the Water–Gas Shift and Carbon Capture Sections. *Bioengineering* **7**, 70 (2020).

133. Cotton, C. A., Claassens, N. J., Benito-Vaquerizo, S. & Bar-Even, A. Renewable methanol and formate as microbial feedstocks. *Curr. Opin. Biotechnol.* **62**, 168–180 (2020).

134. Pacaud, S., Loubière, P., Goma, G. & Lindley, N. D. Organic acid production during methylotrophic growth of *Eubacterium limosum* B2: displacement towards increased butyric acid yields by supplementing with acetate. *Appl. Microbiol. Biotechnol.* **23**, 330–335 (1986).

135. Loubiere, P., Pacaud, S., Goma, G. & Lindley, N. D. The effect of formate on the acidogenic fermentation of methanol by *Eubacterium limosum*. *J. Gen. Appl. Microbiol.* **33**, 463–470 (1987).

136. Loubiere, P., Goma, G. & Lindley, N. D. A non-passive mechanism of butyrate excretion operates during acidogenic fermentation of methanol by *Eubacterium limosum*. *Antonie Van Leeuwenhoek* **57**, 83–89 (1990).

137. Marinos, G., Kaleta, C. & Waschina, S. Defining the nutritional input for genome-scale metabolic models: A roadmap. *PLoS ONE* **15**, e0236890 (2020).

138. Richter, H., Loftus, S. E. & Angenent, L. T. Integrating syngas fermentation with the carboxylate platform and yeast fermentation to reduce medium cost and improve biofuel productivity. *Environ. Technol.* **34**, 1983–1994 (2013).

139. Dragosits, M. & Mattanovich, D. Adaptive laboratory evolution – principles and applications for biotechnology. *Microb. Cell Factories* **12**, 64 (2013).

140. Kang, S. *et al.* Adaptive Laboratory Evolution of *Eubacterium limosum* ATCC 8486 on Carbon Monoxide. *Front. Microbiol.* **11**, 402 (2020).

141. Erickson, L. E. Biomass elemental composition and energy content. *Biotechnol. Bioeng.* **22**, 451–456 (1980).

142. Darling, A. E., Mau, B. & Perna, N. T. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. *PLOS ONE* **5**, e11147 (2010).

143. Doellinger, J., Schneider, A., Hoeller, M. & Lasch, P. Sample Preparation by Easy Extraction and Digestion (SPEED) - A Universal, Rapid, and Detergent-free Protocol for Proteomics Based on Acid Extraction. *Mol. Cell. Proteomics* **19**, 209–222 (2020).

144. Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. DIA-NN: Neural networks and interference correction enable deep proteome coverage in high throughput. *Nat. Methods* **17**, 41–44 (2020).

145. Roh, H. *et al.* Complete genome sequence of a carbon monoxide-utilizing acetogen, *Eubacterium limosum* KIST612. *J. Bacteriol.* **193**, 307–308 (2011).

146. Weirather, J. L. *et al.* Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Research* **6**, 100 (2017).

147. Hu, L. *et al.* Transgenerational Epigenetic Inheritance Under Environmental Stress by Genome-Wide DNA Methylation Profiling in *Cyanobacterium. Front. Microbiol.* **9**, 1479 (2018).

148. Jensen, T. Ø. *et al.* Genome-wide systematic identification of methyltransferase recognition and modification patterns. *Nat. Commun.* **10**, 3311 (2019).

149. Choe, D. *et al.* Adaptive laboratory evolution of a genome-reduced *Escherichia coli*. *Nat. Commun.* **10**, 935 (2019).

150. Espinosa, M. I. *et al.* Adaptive laboratory evolution of native methanol assimilation in *Saccharomyces cerevisiae*. *Nat. Commun.* **11**, 5564 (2020).

151. Har, J. R. G., Agee, A., Bennett, R. K., Papoutsakis, E. T. & Antoniewicz, M. R. Adaptive laboratory evolution of methylotrophic *Escherichia coli* enables synthesis of all amino acids from methanol-derived carbon. *Appl. Microbiol. Biotechnol.* **105**, 869–876 (2021).

152. Wang, Y. *et al.* Adaptive laboratory evolution enhances methanol tolerance and conversion in engineered *Corynebacterium glutamicum*. *Commun. Biol.* **3**, 1–15 (2020).

153. Mariotto, C., Loubière, P., Goma, G. & Lindley, N. D. Influence of various reducing agents on methylotrophic growth and organic acid production of *Eubacterium limosum*. *Appl. Microbiol. Biotechnol.* **32**, 193–198 (1989).

154. Litty, D. & Müller, V. A Na+ A1AO ATP synthase with a V-type c subunit in a mesophilic bacterium. *FEBS J.* **287,** 3012-3023 (2020).

155. Coskun, U. *et al.* Cross-talk in the A1-ATPase from *Methanosarcina mazei* Go1 due to nucleotide binding. *J. Biol. Chem.* **277**, 17327–17333 (2002).

156. Cossio, P. *et al.* Bayesian inference of rotor ring stoichiometry from electron microscopy images of archaeal ATP synthase. *Microscopy* **67**, 266–273 (2018).

157. Vonck, J., Pisa, K. Y., Morgner, N., Brutschy, B. & Müller, V. Three-dimensional Structure of A1A0 ATP Synthase from the Hyperthermophilic Archaeon *Pyrococcus furiosus* by Electron Microscopy. *J. Biol. Chem.* **284**, 10110–10119 (2009).

158. Kremp, F. & Müller, V. Methanol and methyl group conversion in acetogenic bacteria: biochemistry, physiology and application. *FEMS Microbiol. Rev.* **45**, fuaa040 (2021).

159. Bar-Even, A., Flamholz, A., Noor, E. & Milo, R. Thermodynamic constraints shape the structure of carbon fixation pathways. *Biochim. Biophys. Acta* **1817**, 1646–1659 (2012).

160. Furdui, C. & Ragsdale, S. W. The Role of Pyruvate Ferredoxin Oxidoreductase in Pyruvate Synthesis during Autotrophic Growth by the Wood-Ljungdahl Pathway. *J. Biol. Chem.* **275**, 28494–28499 (2000).

161. Pinto, R., Harrison, J. S., Hsu, T., Jacobs, W. R. & Leyh, T. S. Sulfite Reduction in *Mycobacteria*. *J. Bacteriol.* **189**, 6714–6722 (2007).

162. Kitabatake, M., So, M. W., Tumbula, D. L. & Söll, D. Cysteine Biosynthesis Pathway in the Archaeon *Methanosarcina barkeri* Encoded by Acquired Bacterial Genes? *J. Bacteriol.* **182**, 143–145 (2000).

163. Hallenbeck, P. C., Clark, M. A. & Barrett, E. L. Characterization of anaerobic sulfite reduction by *Salmonella typhimurium* and purification of the anaerobically induced sulfite reductase. *J. Bacteriol.* **171**, 3008–3015 (1989).

164. Nakatani, T. *et al.* Enhancement of thioredoxin/glutaredoxin-mediated L-cysteine synthesis from S-sulfocysteine increases L-cysteine production in *Escherichia coli*. *Microb. Cell Factories* **11**, 62 (2012).

165. Paul, L. & Krzycki, J. A. Sequence and transcript analysis of a novel *Methanosarcina barkeri* methyltransferase II homolog and its associated corrinoid protein homologous to methionine synthase. *J. Bacteriol.* **178**, 6599–6607 (1996).

166. Aitken, S. M., Lodha, P. H. & Morneau, D. J. K. The enzymes of the transsulfuration pathways: Active-site characterizations. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **1814**, 1511–1517 (2011).

167. Sampson, E. M. & Bobik, T. A. Microcompartments for B12-Dependent 1,2-Propanediol Degradation Provide Protection from DNA and Cellular Damage by a Reactive Metabolic Intermediate. *J. Bacteriol.* **190**, 2966–2971 (2008).

168. Liu, G., Jiang, Y.-M., Liu, Y.-C., Han, L.-L. & Feng, H. A novel DNA methylation motif identified in *Bacillus pumilus* BA06 and possible roles in the regulation of gene expression. *Appl. Microbiol. Biotechnol.* **104**, 3445–3457 (2020).

169. Nguyen, N.-P.-T., Raynaud, C., Meynial-Salles, I. & Soucaille, P. Reviving the Weizmann process for commercial n-butanol production. *Nat. Commun.* **9**, 3682 (2018).

170. Smith, C. J. *et al.* Current fossil fuel infrastructure does not yet commit us to 1.5 °C warming. *Nat. Commun.* **10**, 101 (2019).

171. Cui, X. & Kær, S. K. Thermodynamic Analyses of a Moderate-Temperature Process of Carbon Dioxide Hydrogenation to Methanol via Reverse Water–Gas Shift with In Situ Water Removal. *Ind. Eng. Chem. Res.* **58**, 10559–10569 (2019).

172. Simon Araya, S. *et al.* A Review of The Methanol Economy: The Fuel Cell Route. *Energies* **13**, 596 (2020).

173. Davidi, D. *et al.* Global characterization of in vivo enzyme catalytic rates and their correspondence to *in vitro* kcat measurements. *Proc. Natl. Acad. Sci.* **113**, 3401–3406 (2016).

174. Pregnon, G., Minton, N. P. & Soucaille, P. Genome Sequence of *Eubacterium limosum* B2 and Evolution for Growth on a Mineral Medium with Methanol and CO2 as Sole Carbon Sources. *Microorganisms* **10**, 1790 (2022).

175. Loferer-Krößbacher, M., Klima, J. & Psenner, R. Determination of Bacterial Cell Dry Mass by Transmission Electron Microscopy and Densitometric Image Analysis. *Appl. Environ. Microbiol.* **64**, 688–694 (1998).

176. Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. & Ralser, M. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat. Methods* **17**, 41–44 (2020).

177. Stead, M. B. *et al.* RNAsnap™: a rapid, quantitative and inexpensive, method for isolating total RNA from bacteria. *Nucleic Acids Res.* **40**, e156 (2012).

178. Moffitt, J. R., Pandey, S., Boettiger, A. N., Wang, S. & Zhuang, X. Spatial organization shapes the turnover of a bacterial transcriptome. *eLife* **5**, e13065 (2016).

179. Owens, N. D. L. *et al.* Measuring Absolute RNA Copy Numbers at High Temporal Resolution Reveals Transcriptome Kinetics In Development. *Cell Rep.* **14**, 632–647 (2016).

180. Beber, M. E. *et al.* eQuilibrator 3.0: a database solution for thermodynamic constant estimation. *Nucleic Acids Res.* **50**, D603–D609 (2022).

181. Wiechmann, A., Ciurus, S., Oswald, F., Seiler, V. N. & Müller, V. It does not always take two to tango: "Syntrophy" via hydrogen cycling in one bacterial cell. *ISME J.* **14**, 1561–1570 (2020).

182. Bremer, H. & Dennis, P. P. Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates. *EcoSal Plus* (2008) doi:10.1128/ecosal.5.2.3.

183. Piehowski, P. D. *et al.* Sources of Technical Variability in Quantitative LC–MS Proteomics: Human Brain Tissue Sample Analysis. *J. Proteome Res.* **12**, 2128–2137 (2013).

184. Litty, D. & Müller, V. A Na+ A1AO ATP synthase with a V-type c subunit in a mesophilic bacterium. *FEBS J.* **287**, 3012–3023 (2020).

185. Hanegraaf, P. P. & Muller, E. B. The dynamics of the macromolecular composition of biomass. *J. Theor. Biol.* **212**, 237–251 (2001).

186. Yoo, M. *et al.* A Quantitative System-Scale Characterization of the Metabolism of *Clostridium acetobutylicum*. *mBio* **6**, e01808-15 (2015).

187. Bobik, T. A., Havemann, G. D., Busch, R. J., Williams, D. S. & Aldrich, H. C. The Propanediol Utilization (pdu) Operon of *Salmonella enterica* Serovar Typhimurium LT2 Includes Genes Necessary for Formation of Polyhedral Organelles Involved in Coenzyme B12-Dependent 1,2-Propanediol Degradation. *J. Bacteriol.* **181**, 5967–5975 (1999).

188. Bobik, T. A. Polyhedral organelles compartmenting bacterial metabolic processes. *Appl. Microbiol. Biotechnol.* **70**, 517–525 (2006).

189. Chowdhury, N. P., Alberti, L., Linder, M. & Müller, V. Exploring Bacterial Microcompartments in the Acetogenic Bacterium *Acetobacterium woodii*. *Front. Microbiol.* **11**, 593467 (2020).

190. Petit, E. *et al.* Involvement of a Bacterial Microcompartment in the Metabolism of Fucose and Rhamnose by *Clostridium phytofermentans*. *PLoS ONE* **8**, e54337 (2013).

191. Shively, J. M., Ball, F., Brown, D. H. & Saunders, R. E. Functional organelles in prokaryotes: polyhedral inclusions (carboxysomes) of *Thiobacillus neapolitanus*. *Science* **182**, 584–586 (1973).

192. Feng, Y. *et al.* Comparative genomics and proteomics of *Eubacterium maltosivorans*: functional identification of trimethylamine methyltransferases and bacterial microcompartments in a human intestinal bacterium with a versatile lifestyle. *Environ. Microbiol.* **n/a**, (2021).

193. Hanegraaf, P. P. F. & Muller, E. B. The Dynamics of the Macromolecular Composition of Biomass. *J. Theor. Biol.* **212**, 237–251 (2001).

194. Valgepea, K., Adamberg, K., Seiman, A. & Vilu, R. *Escherichia coli* achieves faster growth by increasing catalytic and translation rates of proteins. *Mol. Biosyst.* **9**, 2344 (2013).

195. Bar-Even, A. *et al.* The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. *Biochemistry* **50**, 4402–4410 (2011).

196. Nielsen, C. F., Lange, L. & Meyer, A. S. Classification and enzyme kinetics of formate dehydrogenases for biomanufacturing via CO2 utilization. *Biotechnol. Adv.* **37**, 107408 (2019).

197. Vees, C. A., Neuendorf, C. S. & Pflügl, S. Towards continuous industrial bioprocessing with solventogenic and acetogenic *clostridia*: challenges, progress and perspectives. *J. Ind. Microbiol. Biotechnol.* **47**, 753–787 (2020).

198. Liu, Y., Zhang, Z., Jiang, W. & Gu, Y. Protein acetylation-mediated cross regulation of acetic acid and ethanol synthesis in the gas-fermenting *Clostridium ljungdahlii*. *J. Biol. Chem.* **298**, (2022).

# Appendix

## 1. Total RNA extraction protocol

### RNA extraction and addition of spike-in mixes protocol for absolute quantification of transcript number per cell

This protocol aims to provide efficient method allowing the absolute quantification of transcript by integration of two sets of RNA spikes-in during total RNA extraction: **Designed RNA spike-in** and **ERCC spike in mix**.

### A. Solutions for absolute mRNA extraction (RNA-snap method) and spikes-in used for quantification

**RNA solution** (to be prepared fresh before each RNA extraction). Make sure to use DEPC water or 121°C 50 minutes sterilized MilliQ water for dilutions.

| |
|---|
| Formamide (RNA grade) 95% |
| SDS 0,025 % |
| EDTA 18mM |
| 2-mercaptoethanol 1% |

**Zirconium beads for Gram positive cells** (Sigma : Z763764-50EA : BeadBug™ prefilled tubes, 2mL capacity : with ~300 µl of 0,1 mm zirconium beads)

**Two sets of RNA spike-in mix**

**Designed RNA spike-in mix (6 transcripts)**
One set of RNA spike-in mix was designed following the composition and concentrations of the spike-in mix described by Moffitt et al. RNA spike-in (sequences from phiX174 genome) synthesized by SynbioTech. The preparation of designed RNA spike-mix is detailed below.

**ERCC spike in (92 transcripts)**
Preparation of the second spike-in mix set is described below.

**RNA loading dye:**

| |
|---|
| Formamide (RNA grade) 47,5% |
| SDS 0,01 % |
| Bromophenol Blue 0,01 % |
| Xylene Cyanol 0,005 % |
| EDTA 0,5 mM |

**10✕ DNase buffer** – Commercial solution: Dnase I Buffer (NEB, ref: B0303S)

Buffer composition (1X) :

| | |
|---|---|
| $MgCl_2$ | 2.5 mM |
| $CaCl_2$ | 0.5 mM |

| | |
|---|---|
| Tris-HCl | 10 mM |
| pH | 7,6 |

**EDTA**, 0,5M, pH 8.0, Molecular Biology Grade, DEPC-Treated (Sigma, ref: 324506 - 100mL)

**RNA Clean & Concentrator-5** (Zymo - 50 preps kit)

**B-Extraction protocol**

1. Take ~$9.10^8$ cells (~1 ml of 4 $OD_{620}$ culture) and centrifuge at 16000g - 30 seconds – at room temperature and remove supernatant by aspiration. At this stage the pellet can be rapidly frozen in liquid nitrogen and stored at -80°C or directly used in step 2
2. Resuspend the pellet with 500 µL **RNA solution** and homogenize.
3. **Glucose condition only**: divide the sample volume by two.
4. Transfer the solution in BeadBug™ prefilled tubes. **Transfer 250µL for glucose solution and the totality of the cells for methanol condition.**
5. Beat the cells using FastPrep with 3 cycles of 30 seconds at maximum speed: 6.5 $m.s^{-1}$.
6. Transfer the content (cells in solution + beads) in 1,5 mL Eppendorf tube using 1mL tip.
7. Add 25µL of **designed RNA spike-in mix** and homogenize
8. Incubate at 95°C for 7 minutes
9. Centrifuge at 16000g – 10 minutes – at room temperature and transfer the supernatant carefully to a fresh tube. At this stage the RNA extract can be stored at -20°C for at least one year.
10. Add 226 µL of DEPC water to 45,6 µL of supernatant.
11. Add 35 µl of **10✕ DNase buffer** and 5 µL TURBO DNase, mix gently and incubate at 37°C for 30 min.
12. Remove the tube and add 35 µL of EDTA 50 mM, mix well.
13. Warm the sample at 75°C for 10 minutes and stock RNA at -80°C.
14. Check RNA quality and DNase efficacy with 2% gel electrophoresis (after addition of RNA loading dye 2X to 10 µL of total RNA).
15. Transfer 300 µL of total RNA to 2mL Eppendorf tube for in-column purification (Zymo)
16. Add 600 µL of of RNA Binding Solution (RNA Clean and Concentrator kit Zymo) to the total RNA and mix well by pipetting.
17. Add 900 µL of pure ethanol to the mix and mix well.
18. Transfer the mixture to the Zymo-Spin Column and centrifuge at 16000g for 30 seconds and discard the flow-through. Repeat the step until all RNA sample is loaded in the column.
19. Add 400 µL of RNA Prep buffer in the column and centrifuge at 16000g for 30 seconds, discard the flow-through.
20. Add 600 µL of Wash Buffer in the column, centrifuge at 16000g for 30 seconds and discard the flow-through.
21. Repeat the step 19.
22. Change the collection tube and centrifuge the Zymo spin-column at 16000g for 2 minutes. Discard the collection tube and transfer the Zymo spin-column in a fresh Eppendorf 1,5 ml tube.
23. Add 10 µL of DEPC mQ water and wait for 5 minutes. Centrifuge at 16000g for 30 seconds.

24. Repeat the step 22 and keep the tube in ice.
25. Transfer the Zymo spin-column to a new 1,5 mL Eppendorf tube and add 5µL of DEPC mQ water. Centrifuge at 16000g for 30 seconds. Use this second elution to check if RNA had remained in the column by measuring the concentration by Qubit measurement.
26. Measure the RNA concentration after purification and concentration by Qubit measurement and nanodrop. Keep in mind that RNA concentration close to 100ng/µL might be underestimated by Qubit measurement. Measure the concentration on the Nanodrop for double checking.
27. RNA can be stored at -80°C for at least one month.

**Table S1 :** Dilution table for spike-in mix preparation (Synbiotech)

| Spike-in | Spike-in target concentration in the mix (mol/L) (10-fold inferior compared to Moffitt protocol) | **Final concentration of each spike before pooling** | Molar concentration before first dilution | **Dilution 1** (Dilution factor) | **Dilution 2** (Dilution factor) | **Dilution 3** (Dilution factor) | **Final dilution of each spike before pooling** | Number of spike-in molecules (SynbioTech) added per µg of RNA (glucose chemostat) | Number of spike-in molecules (SynbioTech) added per µg of RNA (methanol chemostat) | Number of spike-in molecules added per µg of RNA (Moffitt experiment) |
|---|---|---|---|---|---|---|---|---|---|---|
| spkH | 1,6E-14 | 9,6E-14 | 7,439E-07 | **75** | **50** (50µL of dilution 1 in 1950 µL formamide) | **20** (50µL of dilution 2 in 950 µL formamide) | **24 µL** of spike-in H dilution 3 in **976 µl** of formamide | 6,88E+03 | 9,64E+03 | 8,92E+03 |
| spkA | 8E-13 | 4,8E-12 | 5,340E-07 | **75** | **40** (50µL of dilution 1 in 1950 µL formamide) | - | **27 µL** of spike-in A dilution 2 in **973 µL** of formamide | 3,44E+05 | 4,82E+05 | 4,46E+05 |
| spkF | 4E-12 | 2,4E-11 | 6,104E-07 | **75** | **40** (50µL of dilution 1 in 1950 µL formamide) | - | **59 µL** of spike-in F dilution 2 in **941 µL** of formamide | 1,72E+06 | 2,41E+06 | 2,23E+06 |
| spkD | 2E-11 | 1,2E-10 | 1,215E-06 | **75** | **10** (100 µL of dilution 1 in 900 µL formamide) | - | **74 µL** of spike-in D dilution 2 in **926 µL** of formamide | 8,6E+06 | 1,20E+07 | 1,12E+07 |
| spkG | 1E-10 | 6,00E-10 | 1,408E-06 | **75** | - | - | **32 µL** of spike-in G dilution 1 in **968 µL** of formamide | 4,30E+07 | 6,02E+07 | 5,58E+07 |
| spkK | 5E-10 | 3,00E-09 | 2,507E-06 | **75** | - | - | **90 µL** of spike-in K dilution 1 in **910 µL** of formamide | 2,15E+08 | 3,01E+08 | 2,79E+08 |

**Table S1**: Details on ERCC spike-in concentration and number of molecule per µg of total RNA.

| ERCC ID subgroup | concentration in Mix 1 (attomoles/ul) | Size (nt) | MW (g/mol) | Spike-in concentration after dilution (attomoles/ul) | Spike-in concentration in total RNA mixture before ribodepletion (attomoles/ul) | Number of spike-in molecules added per 1µg of RNA (*E. limosum* B2 experiment) | Number of spike-in molecules added per 1µg of RNA (Owens et al.) |
|---|---|---|---|---|---|---|---|
| A | 7500 | 523 | 167067 | 25 | 2,5 | 3,01E+07 | 3,24E+07 |
| A | 0,11444092 | 1136 | 366884 | 0,00038147 | 3,8147E-05 | 4,59E+02 | 4,95E+02 |
| A | 29,296875 | 644 | 207392 | 0,09765625 | 0,009765625 | 1,18E+05 | 1,27E+05 |
| A | 3,66210938 | 1130 | 364127 | 0,012207031 | 0,001220703 | 1,47E+04 | 1,58E+04 |
| A | 1,83105469 | 2022 | 651363 | 0,006103516 | 0,000610352 | 7,35E+03 | 7,92E+03 |
| A | 58,59375 | 1023 | 328348 | 0,1953125 | 0,01953125 | 2,35E+05 | 2,53E+05 |
| A | 0,02861023 | 1022 | 325166 | 9,53674E-05 | 9,53674E-06 | 1,15E+02 | 1,24E+02 |
| A | 7,32421875 | 844 | 271169 | 0,024414063 | 0,002441406 | 2,94E+04 | 3,17E+04 |
| A | 234,375 | 1124 | 361558 | 0,78125 | 0,078125 | 9,41E+05 | 1,01E+06 |
| A | 117,1875 | 521 | 166158 | 0,390625 | 0,0390625 | 4,70E+05 | 5,07E+05 |
| A | 0,45776367 | 523 | 167040 | 0,001525879 | 0,000152588 | 1,84E+03 | 1,98E+03 |
| A | 937,5 | 1022 | 328268 | 3,125 | 0,3125 | 3,76E+06 | 4,06E+06 |
| A | 468,75 | 1991 | 639815 | 1,5625 | 0,15625 | 1,88E+06 | 2,03E+06 |
| A | 0,22888184 | 1022 | 324755 | 0,000762939 | 7,62939E-05 | 9,19E+02 | 9,90E+02 |
| A | 30000 | 1059 | 342111 | 100 | 10 | 1,20E+08 | 1,30E+08 |
| A | 117,1875 | 771 | 248124 | 0,390625 | 0,0390625 | 4,70E+05 | 5,07E+05 |
| A | 1,83105469 | 274 | 88449 | 0,006103516 | 0,000610352 | 7,35E+03 | 7,92E+03 |
| A | 1875 | 1033 | 333206 | 6,25 | 0,625 | 7,53E+06 | 8,11E+06 |
| A | 29,296875 | 538 | 173255 | 0,09765625 | 0,009765625 | 1,18E+05 | 1,27E+05 |
| A | 0,91552734 | 1023 | 330968 | 0,003051758 | 0,000305176 | 3,68E+03 | 3,96E+03 |
| A | 7,32421875 | 537 | 173168 | 0,024414063 | 0,002441406 | 2,94E+04 | 3,17E+04 |
| A | 0,45776367 | 494 | 159050 | 0,001525879 | 0,000152588 | 1,84E+03 | 1,98E+03 |
| A | 14,6484375 | 1023 | 330306 | 0,048828125 | 0,004882813 | 5,88E+04 | 6,34E+04 |
| B | 937,5 | 984 | 316428 | 3,125 | 0,3125 | 3,76E+06 | 4,06E+06 |
| B | 58,59375 | 1994 | 640770 | 0,1953125 | 0,01953125 | 2,35E+05 | 2,53E+05 |
| B | 1,83105469 | 1138 | 365574 | 0,006103516 | 0,000610352 | 7,35E+03 | 7,92E+03 |
| B | 7,32421875 | 1019 | 327983 | 0,024414063 | 0,002441406 | 2,94E+04 | 3,17E+04 |
| B | 117,1875 | 1130 | 364220 | 0,390625 | 0,0390625 | 4,70E+05 | 5,07E+05 |
| B | 468,75 | 1023 | 325593 | 1,5625 | 0,15625 | 1,88E+06 | 2,03E+06 |
| B | 58,59375 | 274 | 88211 | 0,1953125 | 0,01953125 | 2,35E+05 | 2,53E+05 |
| B | 29,296875 | 1023 | 327814 | 0,09765625 | 0,009765625 | 1,18E+05 | 1,27E+05 |
| B | 234,375 | 523 | 168046 | 0,78125 | 0,078125 | 9,41E+05 | 1,01E+06 |
| B | 3,66210938 | 644 | 207300 | 0,012207031 | 0,001220703 | 1,47E+04 | 1,58E+04 |
| B | 0,91552734 | 603 | 193808 | 0,003051758 | 0,000305176 | 3,68E+03 | 3,96E+03 |
| B | 0,01430512 | 1023 | 325285 | 4,76837E-05 | 4,76837E-06 | 5,74E+01 | 6,19E+01 |
| B | 15000 | 1107 | 356408 | 50 | 5 | 6,02E+07 | 6,49E+07 |
| B | 0,22888184 | 2022 | 647199 | 0,000762939 | 7,62939E-05 | 9,19E+02 | 9,90E+02 |
| B | 0,91552734 | 536 | 172776 | 0,003051758 | 0,000305176 | 3,68E+03 | 3,96E+03 |
| B | 0,05722046 | 1136 | 365599 | 0,000190735 | 1,90735E-05 | 2,30E+02 | 2,48E+02 |
| B | 14,6484375 | 1118 | 359287 | 0,048828125 | 0,004882813 | 5,88E+04 | 6,34E+04 |
| B | 0,11444092 | 1024 | 328403 | 0,00038147 | 3,8147E-05 | 4,59E+02 | 4,95E+02 |
| B | 0,22888184 | 493 | 158941 | 0,000762939 | 7,62939E-05 | 9,19E+02 | 9,90E+02 |
| B | 14,6484375 | 494 | 159762 | 0,048828125 | 0,004882813 | 5,88E+04 | 6,34E+04 |
| B | 3,66210938 | 743 | 238975 | 0,012207031 | 0,001220703 | 1,47E+04 | 1,58E+04 |
| B | 0,45776367 | 1027 | 330577 | 0,001525879 | 0,000152588 | 1,84E+03 | 1,98E+03 |
| B | 3750 | 505 | 162873 | 12,5 | 1,25 | 1,51E+07 | 1,62E+07 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| C | 0,11444092 | 994 | 320107 | 0,00038147 | 3,8147E-05 | 4,59E+02 | 4,95E+02 |
| C | 0,22888184 | 844 | 271530 | 0,000762939 | 7,62939E-05 | 9,19E+02 | 9,90E+02 |
| C | 0,22888184 | 536 | 172979 | 0,000762939 | 7,62939E-05 | 9,19E+02 | 9,90E+02 |
| C | 3,66210938 | 740 | 237170 | 0,012207031 | 0,001220703 | 1,47E+04 | 1,58E+04 |
| C | 0,91552734 | 744 | 239586 | 0,003051758 | 0,000305176 | 3,68E+03 | 3,96E+03 |
| C | 117,1875 | 1156 | 372189 | 0,390625 | 0,0390625 | 4,70E+05 | 5,07E+05 |
| C | 14,6484375 | 274 | 88821 | 0,048828125 | 0,004882813 | 5,88E+04 | 6,34E+04 |
| C | 0,01430512 | 1021 | 328131 | 4,76837E-05 | 4,76837E-06 | 5,74E+01 | 6,19E+01 |
| C | 1,83105469 | 1136 | 366390 | 0,006103516 | 0,000610352 | 7,35E+03 | 7,92E+03 |
| C | 58,59375 | 642 | 205965 | 0,1953125 | 0,01953125 | 2,35E+05 | 2,53E+05 |
| C | 15000 | 522 | 167391 | 50 | 5 | 6,02E+07 | 6,49E+07 |
| C | 234,375 | 642 | 206286 | 0,78125 | 0,078125 | 9,41E+05 | 1,01E+06 |
| C | 29,296875 | 994 | 320289 | 0,09765625 | 0,009765625 | 1,18E+05 | 1,27E+05 |
| C | 0,05722046 | 1143 | 368811 | 0,000190735 | 1,90735E-05 | 2,30E+02 | 2,48E+02 |
| C | 14,6484375 | 1350 | 434247 | 0,048828125 | 0,004882813 | 5,88E+04 | 6,34E+04 |
| C | 468,75 | 994 | 319203 | 1,5625 | 0,15625 | 1,88E+06 | 2,03E+06 |
| C | 3750 | 840 | 270543 | 12,5 | 1,25 | 1,51E+07 | 1,62E+07 |
| C | 0,91552734 | 536 | 172456 | 0,003051758 | 0,000305176 | 3,68E+03 | 3,96E+03 |
| C | 3,66210938 | 784 | 251552 | 0,012207031 | 0,001220703 | 1,47E+04 | 1,58E+04 |
| C | 937,5 | 1042 | 336023 | 3,125 | 0,3125 | 3,76E+06 | 4,06E+06 |
| C | 7,32421875 | 1019 | 328479 | 0,024414063 | 0,002441406 | 2,94E+04 | 3,17E+04 |
| C | 58,59375 | 523 | 166260 | 0,1953125 | 0,01953125 | 2,35E+05 | 2,53E+05 |
| C | 0,45776367 | 1022 | 324602 | 0,001525879 | 0,000152588 | 1,84E+03 | 1,98E+03 |
| D | 15000 | 1061 | 341005 | 50 | 5 | 6,02E+07 | 6,49E+07 |
| D | 937,5 | 1023 | 327373 | 3,125 | 0,3125 | 3,76E+06 | 4,06E+06 |
| D | 0,91552734 | 808 | 261261 | 0,003051758 | 0,000305176 | 3,68E+03 | 3,96E+03 |
| D | 3,66210938 | 1957 | 631239 | 0,012207031 | 0,001220703 | 1,47E+04 | 1,58E+04 |
| D | 234,375 | 751 | 241026 | 0,78125 | 0,078125 | 9,41E+05 | 1,01E+06 |
| D | 0,22888184 | 1122 | 362520 | 0,000762939 | 7,62939E-05 | 9,19E+02 | 9,90E+02 |
| D | 468,75 | 1023 | 329965 | 1,5625 | 0,15625 | 1,88E+06 | 2,03E+06 |
| D | 3750 | 522 | 167939 | 12,5 | 1,25 | 1,51E+07 | 1,62E+07 |
| D | 0,01430512 | 992 | 319955 | 4,76837E-05 | 4,76837E-06 | 5,74E+01 | 6,19E+01 |
| D | 14,6484375 | 525 | 168601 | 0,048828125 | 0,004882813 | 5,88E+04 | 6,34E+04 |
| D | 0,05722046 | 1136 | 366296 | 0,000190735 | 1,90735E-05 | 2,30E+02 | 2,48E+02 |
| D | 1,83105469 | 1137 | 366506 | 0,006103516 | 0,000610352 | 7,35E+03 | 7,92E+03 |
| D | 3,66210938 | 273 | 87549 | 0,012207031 | 0,001220703 | 1,47E+04 | 1,58E+04 |
| D | 29,296875 | 993 | 319938 | 0,09765625 | 0,009765625 | 1,18E+05 | 1,27E+05 |
| D | 58,59375 | 644 | 207606 | 0,1953125 | 0,01953125 | 2,35E+05 | 2,53E+05 |
| D | 0,22888184 | 534 | 172174 | 0,000762939 | 7,62939E-05 | 9,19E+02 | 9,90E+02 |
| D | 0,11444092 | 1020 | 328476 | 0,00038147 | 3,8147E-05 | 4,59E+02 | 4,95E+02 |
| D | 117,1875 | 1136 | 364774 | 0,390625 | 0,0390625 | 4,70E+05 | 5,07E+05 |
| D | 0,91552734 | 537 | 173069 | 0,003051758 | 0,000305176 | 3,68E+03 | 3,96E+03 |
| D | 7,32421875 | 743 | 239284 | 0,024414063 | 0,002441406 | 2,94E+04 | 3,17E+04 |
| D | 14,6484375 | 543 | 174799 | 0,048828125 | 0,004882813 | 5,88E+04 | 6,34E+04 |
| D | 58,59375 | 872 | 279634 | 0,1953125 | 0,01953125 | 2,35E+05 | 2,53E+05 |
| D | 0,45776367 | 1024 | 326242 | 0,001525879 | 0,000152588 | 1,84E+03 | 1,98E+03 |

*microorganisms*

MDPI

*Article*

# Genome Sequence of *Eubacterium limosum* B2 and Evolution for Growth on a Mineral Medium with Methanol and $CO_2$ as Sole Carbon Sources

**Guillaume Pregnon [1], Nigel P. Minton [2] and Philippe Soucaille [1,2,*]**

[1] INSA, UPS, INP, Toulouse Biotechnology Institute (TBI), Université de Toulouse, 31400 Toulouse, France
[2] BBSRC/EPSRC Synthetic Biology Research Centre (SBRC), School of Life Sciences, University Park, The University of Nottingham, Nottingham NG7 2RD, UK
* Correspondence: soucaille@insa-toulouse.fr; Tel.: +33-(0)-561-559-452

**Abstract:** *Eubacterium limosum* is an acetogen that can produce butyrate along with acetate as the main fermentation end-product from methanol, a promising C1 feedstock. Although physiological characterization of *E. limosum* B2 during methylotrophy was previously performed, the strain was cultured in a semi-defined medium, limiting the scope for further metabolic insights. Here, we sequenced the complete genome of the native strain and performed adaptive laboratory evolution to sustain growth on methanol mineral medium. The evolved population significantly improved its maximal growth rate by 3.45-fold. Furthermore, three clones from the evolved population were isolated on methanol mineral medium without cysteine by the addition of sodium thiosulfate. To identify mutations related to growth improvement, the whole genomes of wild-type *E. limosum* B2, the 10th, 25th, 50th, and 75th generations, and the three clones were sequenced. We explored the total proteomes of the native and the best evolved clone (n°2) and noticed significant differences in proteins involved in gluconeogenesis, anaplerotic reactions, and sulphate metabolism. Furthermore, a homologous recombination was found in subunit S of the type I restriction-modification system between both strains, changing the structure of the subunit, its sequence recognition and the methylome of the evolved clone. Taken together, the genomic, proteomic and methylomic data suggest a possible epigenetic mechanism of metabolic regulation.

**Keywords:** *Eubacterium limosum*; acetogen; adaptive laboratory evolution; genome sequence; methanol; butyric acid; mineral medium; proteomics



**Citation:** Pregnon, G.; Minton, N.P.; Soucaille, P. Genome Sequence of *Eubacterium limosum* B2 and Evolution for Growth on a Mineral Medium with Methanol and $CO_2$ as Sole Carbon Sources. *Microorganisms* **2022**, *10*, 1790. https://doi.org/10.3390/microorganisms10091790

Academic Editors: Laurent Dufossé and Juan M. Gonzalez

Received: 20 July 2022
Accepted: 1 September 2022
Published: 5 September 2022



## 1. Introduction

Strictly anaerobic acetogenic bacteria are known to fix one carbon (C1) substrates, such as synthesis gas (syngas) $CO_2$ + $H_2$ or CO, using the methyl and carbonyl branches of the Wood–Ljungdahl pathway (WLP). Syngas metabolization consists of the reduction of two molecules of $CO_2$ with $H_2$ as an electron donor or by incorporation of CO in the carbonyl branch by the carbon monoxide dehydrogenase/acetyl-CoA enzyme and results in the production of one molecule of acetyl-CoA [1]. This central intermediate is mainly used to produce ATP and acetate. However, depending on the acetogen, various molecules can be produced from acetyl-CoA, such as ethanol, or more complex products, including lactate, butyrate, butanol or 2,3-butanediol [2,3]. Conversion of C1 compounds to four-carbon products, such as butyrate, by microorganisms has received much attention due to its potential application in numerous fields, including the energy, cosmetics, and pharmaceutical industries [4]. Moreover, butyryl-CoA, an intermediate in the butyrate pathway, can be reduced to n-butanol, an attractive molecule that can be used both as a platform chemical and a biofuel [5]. Driven by growing ecological interest, syngas exploitation by bacteria is booming with the development of the first factories of industrial scale [6]. However, gas substrate utilization faces several technical challenges, such as

storage and gas–liquid mass transfer, decreasing bacterial productivity [7]. Methanol constitutes an attractive C1 carbon source due to its wide availability throughout the world, easy production process, storability, and liquid state to avoid mass transfer issues [8,9]. The acetogen *Eubacterium limosum* can metabolize methanol and remains one of the few bacterial species able to convert it to butyrate with the coproduction of acetate [10]. Although numerous physiological studies of *E. limosum* B2 have been performed in the past [10–14], its genome sequence is still undetermined, and the first biochemical characterization of anaerobic methanol metabolism, including the energetic aspect, was performed in the model methylotrophic bacterium *Acetobacterium woodii* [15], which does not produce C4 compounds. Two molecules of methanol are converted to two molecules of methyl-THF by a methyltransferase system. One molecule is oxidized into $CO_2$ through the reverse use of the methyl branch to produce the reducing equivalents needed to then reduce $CO_2$ to CO, while the second molecule is used with CO by the ACS/CODH complex to produce acetyl-CoA. This method of methanol assimilation through the WLP was also demonstrated for *Eubacterium callanderi* KIST 612, formerly named *E. limosum* KIST612 [16]. Moreover, the first evidence of butanol production by *E. limosum* ATCC 8486 from methanol and formate used as cosubstrates was recently obtained, reinforcing the interest in this bacterial species [17]. Methanol metabolism by *E. limosum* B2 was characterized in semi-defined medium containing yeast extract (YE) and cysteine [10]. To the best of our knowledge, no study has been performed on any *E. limosum* strain to develop a mineral medium for growth on methanol as a carbon and energy source. A strain that can grow on a mineral medium will have several advantages for in-depth metabolic characterization using flux balance analysis and genome-scale models [18]. The semi-defined medium previously used for *E. limosum* B2 contained YE but also cysteine, as the strain was shown to be an auxotroph for cysteine [10,19]. Synthetic medium free from YE is also of economic interest, as YE is an expansive supplement prohibiting favourable economic scale-up [20]. Growth improvement can be easily achieved by adaptive laboratory evolution (ALE), which consists of culturing a microorganism for several generations in restrictive medium to naturally improve its growth parameters [21]. Adaptive laboratory evolution has already been performed on *E. limosum* ATCC 8486 using CO as a carbon and energy source, improving its growth rate 1.4-fold [22]. Here, we aimed to deepen our knowledge of the *E. limosum* B2 strain at the genome level and to develop a mineral medium allowing good growth on methanol of an evolved strain for further performing systems biology analysis in chemostat culture. For this purpose, we (i) performed de novo whole-genome sequencing by single-molecule real-time technology (SMRT), (ii) developed a defined medium allowing the strain to grow with methanol as a carbon and energy source without the addition of YE but with cysteine, (iii) improved the growth performance of the strain on this medium by ALE, (iv) developed a mineral medium allowing the adapted strain to grow with methanol as a carbon and energy source without the addition of cysteine, (v) studied genomic modifications that occurred during the evolution process and (vi) performed a comparative study of the total proteome between an isolated clone from the evolved population and the wild-type strain.

## 2. Materials and Methods

### *2.1. Bacterial Strain and Growth Conditions*

The *E. limosum* B2 strain used in this study has been studied in our laboratory since 1985 and is stored at −80 °C. Culture growth was performed in serum bottles (in 25 mL of culture medium) under a strict anaerobic nitrogen ($N_2$) atmosphere and incubated at 37 °C at pH 7.4. The wild-type strain was cultured in rich medium for anaerobic bacteria (M187), which comprised the following components: tryptone 30 g/L, yeast extract 20 g/L, glucose 5 g/L, and cysteine hydrochloride 0.5 g/L. The adapted strain in the methanol substrate was cultured in synthetic medium composed of the following components: NaCl 0.6 g/L, $NH_4Cl$ 1 g/L, $MgSO_4$-$7H_2O$ 0.12 g/L, $CaCl_2$ 0.08 g/L, $KH_2PO_4$ 1 g/L, vitamin solution 100X (biotin 20 mg/L, pantothenic acid 50 mg/L, lipoic acid 50 mg/L), sodium thiosulfate pentahydrate 4 mM, potassium carbonate 1 g/L, titanium citrate 2 mM as a reducing agent,

and trace element solution 100X ($MnCl_2$ 0.1 g/L, $CoCl_2$-$2H_2O$ 0.1 g/L, $ZnCl_2$ 0.1 g/L, $CuCl_2$ 0.02 g/L, $H_3BO_3$ 0.01 g/L, Na molybdate 0.01 g/L, $Na_2SeO_3$ 0.017 g/L, $NiSO_4$-$6H_2O$ 0.026 g/L, $FeSO_4$,$6H_2O$ 1 g/L and nitrilotriacetic acid 12.8 g/L). Resazurin was added at $1.10^{-4}$% (*wt*/*wt*) as a redox potential indicator. Glucose (6 g/L) and methanol (6.4 g/L) were the two substrates tested, corresponding to 200 mM carbon.

Isolated clones from the methanol-adapted population were cultured in solid synthetic medium with glucose as the substrate but without cysteine. The product concentrations for solid synthetic medium elaboration remained the same as those for liquid synthetic medium supplemented with 15 g/L of agarose. After solidification, plates were placed in an anaerobic chamber overnight before streaking the population to ensure balance of the anaerobic atmosphere on the solid medium (90% $N_2$, 5% $CO_2$ and 5% $H_2$).

### *2.2. Adaptive Laboratory Evolution*

The ALE was performed from the first generation of a yeast extract-free adapted *E. limosum* B2 strain previously obtained in the laboratory in synthetic medium with 200 mM methanol as the substrate (unpublished data). A 5% volume of culture was transferred to fresh medium every three generations. The population was cultured over 100 generations.

### *2.3. Clone Isolation on Methanol Mineral Medium without Cysteine*

Single clones from the adapted population (over 100 generations) were isolated in solid glucose mineral medium without cysteine in an anaerobic chamber at 37 °C. Cysteine was replaced by 4 mM sodium thiosulfate in the synthetic medium. Three clones were cultivated in liquid methanol mineral medium without cysteine to follow and compare the growth profile from the 10th, 25th, 50th and 75th generations adapted in MMM with cysteine. Cysteine was replaced by 4 mM sodium thiosulfate.

### *2.4. Analytical Methods*

Culture density was measured by spectrophotometry at 620 nm (OD620) against a water blank. Growth rates were calculated from these measurements. Substrates and products were measured using a high-performance liquid chromatography instrument (Agilent 1200 series, Massy, France) equipped with a refractive index detector and Aminex HPX 87 H 300 × 7.8 mm. The acid $H_2SO_4$ (0.5 mM) was used as the mobile phase and elution was performed at 48 °C. The biomass formula $C_4H_7O_2N_{0.6}$ was used to convert biomass values into molar cell carbon concentration (mC) [23]. Acetate and butyrate production yields were calculated in millimolar carbon from the product per 100 mM carbon from methanol (mC/100 mC MetOH).

### *2.5. Genomic DNA Extraction*

For de novo sequencing, the wild-type strain of *E. limosum* B2 was cultured in rich medium for anaerobic bacteria (M187). Genomic DNA of harvested cells in mid-exponential phase from 1 mL culture was carried out with the Wizard® HMW DNA extraction kit (Promega, Madison, WI, USA) according to the manufacturer's instructions with the following modifications. The lysis step with lysozyme treatment was extended to 2 h to optimize the DNA extraction yield and genomic DNA was resuspended in 10 mM Tris-HCl (pH 8.5) solution. The DNA quantification was performed spectrophotometrically by a Nanodrop 2000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA) and fluorometrically using a Qubit™ instrument (Thermo Fisher Scientific, Wilmington, DE, USA). Quality assessment was performed by (i) gel electrophoresis, (ii) the A260/A280 and A260/230 ratios (>1.9) and (iii) DNA concentration comparison by Qubit and Nanodrop. The difference between the results from both instruments had to be less than 2 for validation.

For genomic comparison of the evolved population through generations, whole-genomic DNA of the wild-type strain, the 10th, 25th, 50th, and 75th generations and three isolated clones was extracted using the same kit as that used for the WT strain for de novo sequencing, measured by Nanodrop and stored at −20 °C until sequencing.

### *2.6. Whole Genome Sequencing*

Single-molecule real-time long read sequencing of the *E. limosum* B2 wild-type strain was performed at the Gentyane Sequencing Platform (Clermont-Ferrand, France, https://gentyane.clermont.inrae.fr, accessed on 25 November 2020) with a PacBio Sequel II Sequencer (Pacific Biosciences, Menlo Park, CA, USA). The SMRTBell library was prepared using a SMRTbell Express 2 Template prep kit (Pacific Biosciences, Menlo Park, CA, USA), following the "procedure and checklist -preparing Multiplexed Microbial Libraries using SMRTbell Express Template prep kit 2.0" protocol. Genomic DNA (1 µg) of each strain was sheared using g-tubes (Covaris, England), generating DNA fragments of approximately 10 kb. A Fragment Analyzer (Agilent Technologies, Santa Clara, CA, USA) assay was used to assess the fragment size distribution. Sheared genomic DNA was carried into the enzymatic reactions to remove the single-strand overhangs and to repair any damage that may be present on the DNA backbone. An A-tailing reaction followed by overhang adapter ligation was conducted to generate SMRTbell templates. The sample was then purified with 0.45X AMPure PB beads to obtain the final libraries of approximately 10 kb. The SMRTBell libraries were inspected for quality and quantified on a Fragment Analyzer (Agilent Technologies, Santa Clara, CA, USA) and with a Qubit fluorimeter with the Qubit dsDNA HS reagent Assay kit (Life Technologies, Eugene, OR, USA). A ready-to-sequence SMRTBell Polymerase Complex was created using a Binding Kit 2.0 (Pacific Biosciences, Menlo Park, CA, USA) and the primer V4, and the diffusion loading protocol was used according to the manufacturer's instructions. The PacBio Sequel instrument was programmed to load a 100 pM library and sequenced in CLR mode on a PacBio SMRTcell 8M, with a Sequencing Plate 2.0 (Pacific Biosciences, Menlo Park, CA, USA) and 2 h of pre-extension time and acquiring one movie of 15 h per SMRTcell.

Whole-genomic DNA of the wild-type strain, the 10th, 25th, 50th, and 75th generations and the three isolated clones were sequenced with the Ion S5 system on the Get-Biopuces sequencing platform (Toulouse, France, https://get.genotoul.fr, accessed on 20 March 2021).

### *2.7. Functional Annotation and Genome Analysis*

Gene prediction and functional annotation of the wild-type strain of *E. limosum* B2 were performed using the NCBI Prokaryotic Genome Annotation Pipeline PGAP [24]. The genome sequence can be found under the accession number CP097376.1. Whole-genome sequence alignment of *E. limosum* B2 against *E. limosum* ATCC 8486 was performed using ProgressiveMauve software (v2.4.0) [25].

### *2.8. Methylome and Type I Restriction-Modification System Analysis*

For analysis of the whole DNA methylation level, methylation of adenine on the 6th carbon (m6a) and methylation of cytosine on the 4th carbon (m4c) were determined using the ipdSummary and Motifmaker tools included in the SMRT Link software (v10.1). The protein sequence alignment for the analysis of the type I RM system was performed using ClustalW in Ugene software (v40.1). Conserved domain analysis was performed using the NCBI conserved domain search tool.

### *2.9. Protein Isolation and Total Proteome Analysis*

Wild-type *E. limosum* B2 and clone 2 were cultured in methanol mineral medium containing 0.5 g/L yeast extract in duplicate. Cells were harvested at mid-exponential growth phase, and sample volumes were adjusted to normalize the cell number to $10^9$ for each condition. The samples were then centrifuged at 16,000× *g* for 30 s, and the supernatants were discarded. Cell pellets were washed twice with 1 mL of phosphate-buffered saline solution. After discarding the supernatant, the cell pellets were frozen in liquid nitrogen and stored at −20 °C. Total proteins were isolated using the SPEED extraction method [26]. Each pellet was resuspended in 150 µL of trifluoroacetic acid (TFA, Supelco MS Grade, Darmstadt, Germany), warmed at 70 °C for 3 min and centrifuged at 16,000× *g* for 1 min. A total of 130 µL of supernatant was carefully transferred to a fresh 1.5 mL Eppendorf tube. Fourteen microliters of supernatant were taken for each sample and diluted in 56 µL 75% TFA to measure the protein concentration by the Bradford assay. Each sample was then neutralized with 1.5 mL of TRIS 2 M, prepared in LC–MS grade water (Supelco), transferred to glass vials and freeze dried. Samples were sent to the John van Geest Cancer Research Center at Nottingham, resolubilized, purified and injected into a TripleTOF 6600 mass spectrometer (Sciex). Two injections per sample were performed, giving a technical duplicate for each sample. Sequential window acquisition of all theoretical spectra (SWATH) and additional data treatment using DIA-NN (https://github.com/vdemichev/DiaNN, accessed on 1 April 2022) were performed to analyse the two groups of samples [27].

## 3. Results

### *3.1. De Novo Sequencing, Annotation and Analysis of the E. limosum B2 Genome*

Although the genomes of several *E. limosum* strains have been sequenced, the genome sequence of *E. limosum* B2, the most well-studied strain from a physiological point of view, remains unknown [28–30]. De novo sequencing was performed on the WT strain using single molecule real-time technology (PacBio). A total of 76,620 long reads were used to produce two linear contigs of 4.06 Mb and 358 kb for a total genome size of 4.424 Mb. The main issue in obtaining one circular contig was the presence of a prophage, present in two copies in the genome and localized at one end of each of the contigs. A complete circular assembly of 4,421,327 base pairs was created using long-range PCR associated with Sanger sequencing. Based on large reads, PacBio sequencing technology can be prone to errors [31]. To correct possible errors from the sequencer, the circular genome sequence was compared to the whole genome sequence of the *E. limosum* B2 WT strain obtained by the IonTorrent S5 sequencer. One nucleotide insertion was observed in the PacBio sequence, localized in an *ompD* homologous gene, which was not confirmed by Ion S5 sequencing.

The genome sequence of *E. limosum* B2 has a G+C content of 47.2% and contains 4192 genes, 51 tRNA genes, 16 rRNA genes including 5 23S, 16S and 5S rRNA operons, a supplementary 5S rRNA gene, 1 tmRNA and two clusters of CRISPR genes, similar to the *E. limosum* SA11 and KIST612 strains [28,29]. The genome of *E. limosum* B2 showed 99.837% sequence identity with the *E. limosum* ATCC 8486 strain.

### *3.2. Genomic Comparison of E. limosum B2 vs. E. limosum ATCC 8486*

Differences between genes and intergenic regions were identified (Table 1).

**Table 1.** Genomic differences between the *Eubacterium limosum* B2 and ATCC 8486 strains.

| Locus Tag ATCC8486 | Locus Tag B2 | Position B2 | Position ATCC8486 | Gene | Nucleotide Variation ATCC8486/B2 | Type | AA Change | Description |
|---|---|---|---|---|---|---|---|---|
| B2M23_RS00035 | M5595_17445 | 3667592 | 4431 | - | G1265T | SNP | $Ala^{421}Glu$ | Phage terminase |
| B2M23_RS00620 | M5595_16850 | 3547040 | 124983 | - | A514G | SNP | $Tyr^{171}His$ | Acyl ACP thioesterase |
| B2M23_RS00770 | M5595_16695 | 3512058 | 159965 | - | C1065T | SNP | $Val^{355}Val$ | Transcriptional activator |
| Intergenic region | Intergenic region | 2725372 | 946652 | - | G/A | SNP | - | - |
| Intergenic region | Intergenic region | 3264417 | 407607 | - | A/- | DEL | - | - |
| B2M23_RS06095 | M5595_11305 | 2338369 | 1333655 | *cheY* | A187G | SNP | $Arg^{63}Gly$ | Response regulator |
| B2M23_RS08865 | M5595_08530 | 1755671 | 1916353 | *aspS* | G1004A | SNP | $Gly^{335}Glu$ | Aspartate–tRNA ligase |
| B2M23_RS08970 | M5595_08425 | 1735036 | 1936988 | - | G1007T | SNP | $Thr^{336}Asp$ | Bacitracin export permease |
| Intergenic region | Intergenic region | 1442364 | 2231325 | - | A/C | SNP | - | Lysin riboswitch |
| B2M23_RS11215 | M5595_20825 | 4406797 | 2387123 | - | -/G | INS | - | Hypothetical protein |
| Intergenic region | M5595_20830 | 4406887 | 2387213 | - | -/A | INS | - | Hypothetical protein |
| Intergenic region | M5595_20830 | 4406916 | 2387242 | - | -/T | INS | - | Hypothetical protein |
| B2M23_RS13215 | M5595_19385 | 4089680 | 2815304 | - | G234T | SNP | $Met^{78}Ile$ | Mobilization protein |
| Intergenic region | M5595_00360 | 73507 | 3183052 | - | -/C | INS | - | Restriction endonuclease subunit S |
| Intergenic region | M5595_00360 | 73556 | 3183101 | - | -/TG | INS | - | Restriction endonuclease subunit S |
| Intergenic region | M5595_00360 | 73990 | 3183516 | - | -/G | INS | - | Restriction endonuclease subunit S |
| Intergenic region | Intergenic region | 367145 | 3476696 | - | T/- | DEL | - | - |
| B2M23_RS17820 | M5595_03295 | 694242 | 3803795 | *ktrA* | A/- | DEL | | TrkA family potassium uptake protein |
| B2M23_RS18850 | M5595_04320 | 887882 | 3997435 | *pucK* | C/T | SNP | $Ile^{116}Thr$ | Uric acid permease PucK/Xanthine permease |
| B2M23_RS18850 | M5595_04320 | 887885 | 3997438 | *pucK* | A/G | SNP | $Gly^{117}Asp$ | Uric acid permease PucK/Xanthine permease |
| B2M23_RS19910 | M5595_05415 | 1121402 | 4230955 | *pbpF* | T/C | SNP | $Ser^{459}Pro$ | Transglycosylase domain-containing protein |

In addition to large genomic rearrangement, as we describe later, a total of 21 small genomic differences, including 12 single nucleotide polymorphisms (SNPs), six nucleotide insertions and three single nucleotide deletions, were observed for the *E. limosum* B2 genome compared to the *E. limosum* ATCC 8486 chromosome. Differences occurred mainly in genes coding for regulators, such as a potential response regulator (M5595_11305) introducing an $Arg^{63}Gly$ substitution in the protein or a transcriptional activator (M5595_16695) introducing a silent mutation. A total of four genetic differences were detected in genes coding for a prophage operon, one in a terminase encoding gene (M5595_17445) leading to an $Ala^{421}Glu$ substitution in the protein, two in intergenic regions between genes of unknown function and one in a gene coding for a hypothetical protein. Furthermore, three insertions were noticed in a gene of *E. limosum* B2 encoding a potential S subunit (M5595_00360) of a type I restriction endonuclease. Other genomic differences were localized in genes encoding transporters, such as a uric acid permease-encoding gene (M5595_04320) or a bacitracin export permease-encoding gene (M5595_08425). Additionally, deletion of an insertion sequence (IS) was detected in the genome of *E. limosum* B2 compared to *E. limosum* ATCC 8486. This IS was present in six and five copies in *E. limosum* ATCC 8486 and B2 strains, respectively. The missing IS in *E. limosum* B2 was located upstream of a sodium/phosphate transporter-encoding gene, reducing the size of the gene by 42 nucleotides in the *E. limosum* ATCC 8486 gene according to the annotation. Whole genome alignment using ProgressiveMauve software also revealed genome reorganization of five locally colinear blocks (LCBs) with reverse complement orientation (Figure 1).

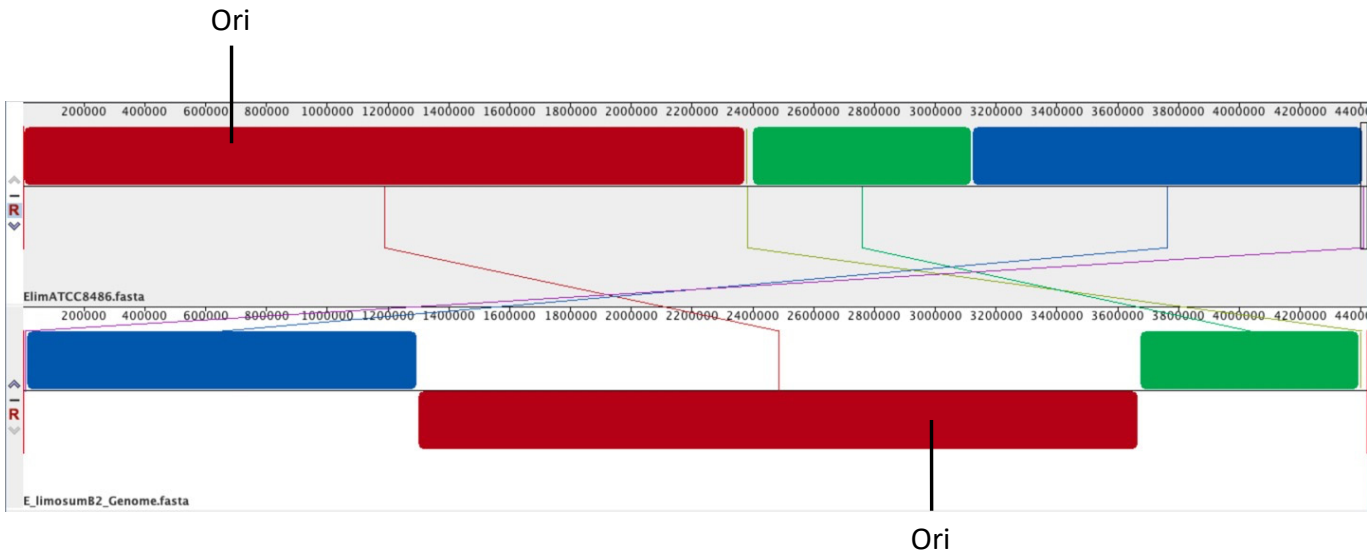



**Figure 1.** Whole genome alignment of *E. limosum* B2 (bottom) to *E. limosum* ATCC 8486 (top) using ProgressiveMauve software [25]. Ori, origin of replication at genomic position 689,833 for *E. limosum* ATCC8486 and 2,982,134 for *E. limosum* B2.

### *3.3. Whole DNA Methylation Level*

The DNA methylation of the chromosome is an epigenetic modification that plays a crucial role in the response to environmental stresses, genome structure rearrangements and gene regulation [32,33]. To explore the methylome of *E. limosum* B2, a kinetic data analysis was performed on the SMRT sequence obtained on the PacBio system to identify the locations of m6A and m4C methylated nucleotides. The five restriction-modification (RM) systems identified in *E. limosum* ATCC 8486, including one type I RM and four type II RMs, were also present in the *E. limosum* B2 strain. Methylome data from *E. limosum* B2 showed only three different motifs of the m6A type for a total of 3974 methylated motifs out of 8444 probable motifs. The *E. limosum* B2 strain had a modification frequency of 0.96 modifications/kb (Table 2).

**Table 2.** Summary of methylome data from the *E. limosum* B2 and the evolved clone 2 strains.

| Strain | Motif | Predicted RM Type | Modified Position | Methylation Type | Fraction (%) | Number of Methylated Motifs Detected | Number of Total Motifs Detected |
|---|---|---|---|---|---|---|---|
| *E. limosum* B2 | GCGRAG | II | 5 | m6A | 67.6 | 2058 | 3042 |
| | CAAAAAR | II | 6 | m6A | 37.3 | 1623 | 4350 |
| | CNNTAYNNNNNTCC | I | 5 | m6A | 27.9 | 293 | 1052 |
| *E. limosum* B2 clone 2 | GCGRAG | II | 5 | m6A | 81 | 2464 | 3042 |
| | CAAAAAR | II | 6 | m6A | 83.7 | 3642 | 4350 |
| | CNNTAYNNNNGTG | I | 5 | m6A | 79.8 | 576 | 1052 |

The GCGRA$^{m6}$G motif was most prevalent, with 2058 occurrences among a total of 3042 sites. A GCGRA $^{m6}$G-like motif, 'GCGCA $^{m6}$G', was identified in *Rhuminococcus flavefaciens* FD-1 according to the REbase databank. This motif site is recognized by a type II RM (RflFIII), sharing 66% similarity to Class I SAM-dependent methyltransferase from *Eubacterium* species but only 21% similarity with N-6 DNA methylase from *E. limosum* B2.

Compared to the methylome of a close metabolic neighbour, *A. woodii,* almost all the detected motifs carried a methyl group, while the motif fraction that contained methylated nucleotides was lower (67.6%) for *E. limosum* B2 [34]. Furthermore, the number of methylated nucleotides expressed per kb of DNA was calculated to be 0.96 modifications/kb for *A. woodii*. The CAAAAA $^{m6}$R methylated motif present in the *E. limosum* B2 genome was also detected in *A. woodii* [33]. This adenine in position 6 in the motif is methylated by a type II methyltransferase, annotated N-6 DNA methylase (Awo_c14460), sharing 48.92% similarity to the N-6 DNA methylase of *E. limosum* B2 (M5595_07210). A total of 27.9% of CNNTAYNNNNNTCC carry a modification of their adenine at positions 5. Although the associated enzyme remains unknown, these nonpalindromic bipartite motifs are likely recognized by a Type I RM system and most likely encoded by the M5595_00345 to M5595_00360 genes.

### *3.4. Strain Adaptation on Methanol-Defined Medium without Yeast Extract*

The first step for growth optimization on methanol substrate was to adapt the strain on a YE-free defined medium. The WT strain, initially cultivated in M187-rich medium, was inoculated in methanol-defined medium without yeast extract to obtain a strain that could grow in YE-depleted medium. Two subcultures in freshly defined medium were performed to ensure YE from the original rich medium culture was at least diluted 400-fold and played a minor role in the YE-depleted medium. The cell density of the WT strain in methanol-defined medium was compared to that in methanol-rich medium as a control (Figure 2).

After a 3-day lag phase, a population started to grow, showing a maximum $OD_{620}$ of 0.670 ± 0.05 and a maximum specific growth rate (μmax) of 0.017 ± 0.002 $h^{-1}$. The WT strain cultivated on M187 rich medium showed a maximum $OD_{620}$ of 2.165 ± 0.05 with a μmax of 0.049 ± 0.003 $h^{-1}$. The growth rate in methanol-defined medium decreased by 65% compared to that in the rich medium, but the WT strain showed its ability to overcome YE depletion. Obtaining a strain that can grow without YE represents a strong asset to precisely determine metabolic fluxes from different carbon sources using a genome-scale model. However, the specific growth rate of 0.017 $h^{-1}$ remained low compared to the maximum growth rate of 0.05 $h^{-1}$ measured for the *E. limosum* B2 strain in semi-defined medium supplemented with YE 0.5 $g{\cdot}L^{-1}$ YE [10]. The next step of the study was to improve the physiological characteristics of the strain in methanol-defined medium by ALE.

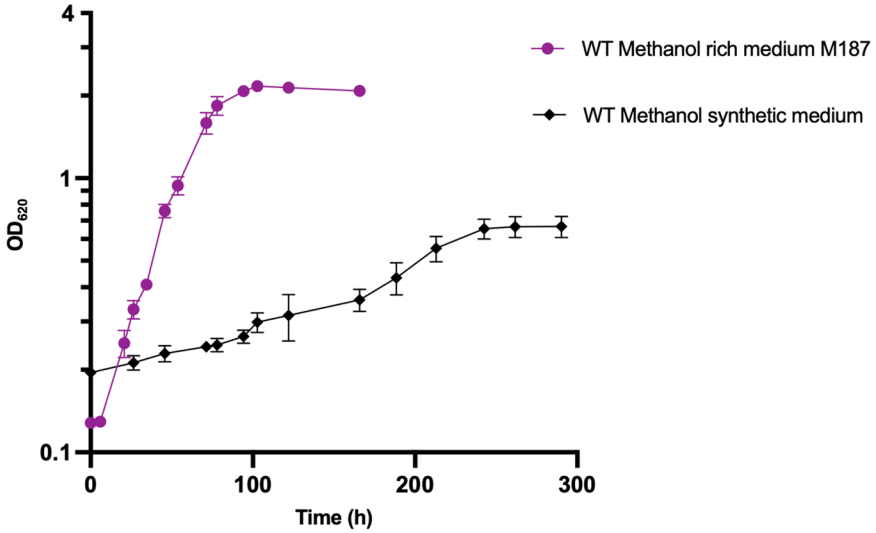



**Figure 2.** Growth profile of *E. limosum* B2 on methanol-rich and synthetic medium without YE. Cultures were performed in triplicate and errors bars indicates standard deviation.

### *3.5. Adaptative Laboratory Evolution of E. limosum B2 in a Methanol-Defined Medium*

The ALE on *E. limosum* ATCC 8486 was reported on carbon monoxide in semi-defined medium supplemented with 2 $g \cdot L^{-1}$ YE. The results showed a notable improvement in the physiological characteristics of the *E. limosum* ATCC 8486 strain, with a maximal optical density (620 nm) and a growth rate that increased 2.14-fold and 1.44-fold [35], respectively. To the best of our knowledge, no ALE has been performed using methanol as a substrate with *Eubacterium* species. Few studies have focused on ALE to improve growth on methanol. Such an experiment was performed on the natural methylotroph *Saccharomyces cerevisiae*, improving the final biomass by 44% after 230 generations [36]. Additionally, ALE on methanol was performed for synthetic methylotrophs such as *Escherichia coli* or *Corynebacterium glutamicum* to improve their tolerance [37,38] to methanol. The ALE was performed on *E. limosum* B2 to obtain a strain with the best growth performance on methanol-defined medium. In total, the strain was cultured for more than 100 generations with a subculture performed every three generations. The ALE showed a clear improvement in the maximal growth rate (Figure 3A).

The lag phase was gradually reduced as the strain was subcultured, from approximately 70 h for the first generation to less than 5 h for the evolved population. The growth rate progressively increased over generations, with a μmax at $0.022 \pm 0.001$ $h^{-1}$ for the 10th generation, $0.047 \pm 0.0084$ $h^{-1}$ for the 25th generation, and $0.057 \pm 0.0026$ $h^{-1}$ for the 50th generation, reaching a maximal value of $0.076 \pm 0.0019$ $h^{-1}$ for the 75th generation. Beyond this last generation, no significant physiological growth improvement was noticed, suggesting that the population reached its maximal potential. In comparison, a maximum growth rate of 0.07 $h^{-1}$ was observed for *E. limosum* B2 in semi-defined methanol (200 mM) medium supplemented with sodium acetate at 12.7 mM [10].

To further investigate the physiologic parameters of the strain on methanol-defined medium, product measurements were systematically performed before each subculture, and values are presented for the 10th, 25th, 50th and 75th generations (Figure 3B). Significant changes in carbon distribution were observed for acetate production yield between generations, with a significant yield increase of approximately 80% observed between the 10th and 50th generations, from $45 \pm 4$ to $77 \pm 5$ mC acetate/100 mC MetOH. The mean acetate yield decreased by approximately 20% between the 50th and 75th generations, reaching $50.5 \pm 4.5$. In contrast, the butyrate production yield showed a significant decrease of approximately 40% between the 10th and 50th generations from $67 \pm 2$ to $39.5 \pm 0.5$ mC butyrate/100 mC MetOH. For the 75th generation, the mean butyrate yield showed an increase to $55.5 \pm 7.5$ mC butyrate/100 mC MetOH, approximately 40% that

for to the 50th generation. The carbon flux toward acetate tended to increase from the 10th to 50th generation, while the reverse phenomenon was observed for butyrate production. Carbon from methanol tended to have an equimolar distribution between acetate and butyrate for the 75th generation. Globally, the ALE did not drastically affect the carbon balance between the two products. Carbon incorporation into biomass rapidly increased from the 10th generation (7.5 ± 0.5) to reach 28 ± 1 mC biomass/100 mC MetOH for the 75th generation.

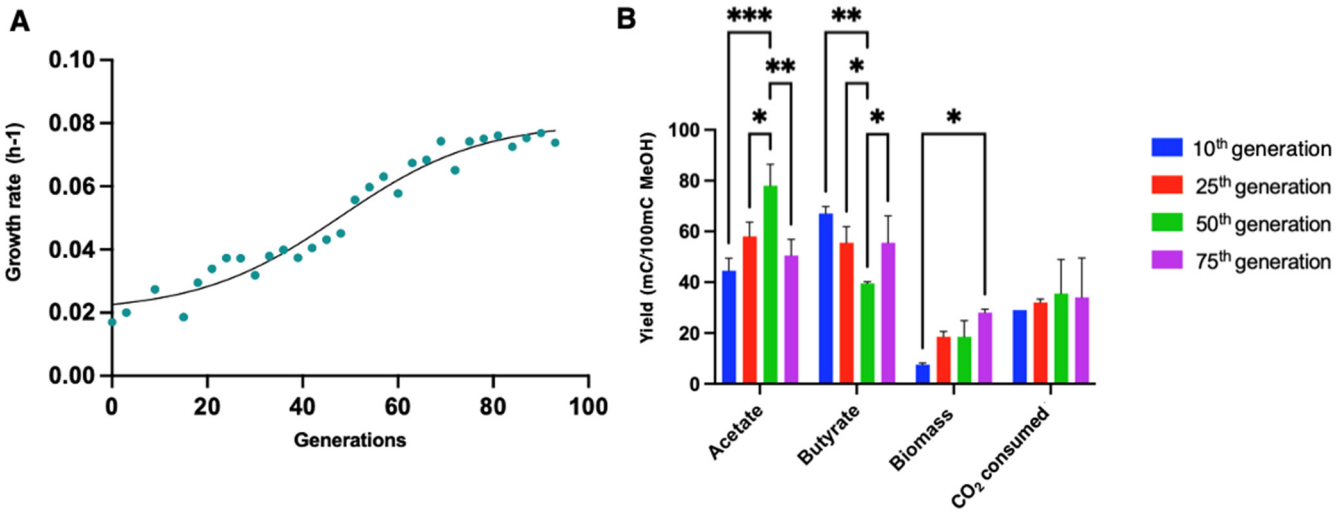



**Figure 3.** Evolution of physiologic characteristics during ALE on methanol methylotrophic growth of *E. limosum* B2. (**A**) Evolution of the maximum growth rate for each generation during ALE. (**B**) Carbon yields obtained from consumed methanol to acetate, butyrate, biomass, and $CO_2$. Yields are expressed in millimolar of carbon (mC) per 100 mC of methanol consumed. Cultures were performed in triplicate and error bars indicates standard deviation. Asterisk indicates that the results were significantly different from the control according to the one-sided Student's *t*-test: *: $p$ value $\leq 0.05$, **: $p$ value $\leq 0.01$, ***: $p$ value $\leq 0.001$.

According to the statistical analysis method applied, no significant change was noticed over generations for the $CO_2$ consumed, which is a necessary cosubstrate for methylotrophic growth of *E. limosum* B2 [10]. In general, more carbon from the substrate was dedicated to biomass production through ALE, while the $CO_2$ consumption rate remained constant. The evolved population on synthetic methanol medium tended to have an equimolar carbon distribution from methanol and $CO_2$ to produce acetate and butyrate.

### *3.6. Isolation of Individual Clones Growing on Methanol Mineral Medium*

The next step of the study was to isolate a strain growing on a methanol-defined medium in the absence of cysteine, i.e., a methanol mineral medium (MMM). To obtain a strain able to grow on a defined medium without cysteine and with methanol as substrate, cells from the advanced evolved population (75th generation) were first spread on a solid glucose mineral medium (SGMM) as no colonies were obtained when spread on a solid MMM (SMMM) supplemented or not with cysteine. This phenomenon was probably due to $CO_2$ accessibility issues (although cultures were performed in an anaerobic chamber with 5% $CO_2$) along with possible methanol evaporation. After growth on SGMM, three colonies were selected and cultured in MMM medium at 37 °C. However, after three subcultures, clone growth declined progressively, presumably from cysteine deficiency. Although *E. limosum* B2 was not documented as auxotrophic for cysteine, alternative sources of sulphur were tested to overcome this presumed auxotrophy [39]. Sodium dithionite, sodium metabisulfite and sodium thiosulfate were added separately to the MMM at 1, 2 and 4 mM. Clones grew on all sulphur sources, with a better maximal biomass observed at 4 mM sodium dithionite. Although sodium dithionite slightly improved growth perfor-

mance compared to sodium thiosulfate at 4 mM, this last sulphur source was chosen for two reasons: (i) it allows similar growth performance compared to a methanol-defined medium with cysteine and, (ii) sodium dithionite is also a reducing agent. Therefore, we chose to dissociate the source of sulphur from the source of the reducing agent (titanium citrate at 2 mM) to obtain more reproducible fermentation data.

### *3.7. Growth Profile of Adapted Clones in Liquid MMM*

Growth profiles of the three isolated clones were determined in liquid MMM supplemented with sodium thiosulfate. Optical density monitoring of the three clones showed similar growth profiles, with a maximum OD of 3.69 ± 0.43 for clone 1, 4.03 ± 0.08 for clone 2 and 3.44 ± 0.37 for clone 3. The maximum growth rate was 0.054 ± 0.008 $h^{-1}$ for clone 1, 0.060 ± 0.003 $h^{-1}$ for clone 2 and 0.047 ± 0.004 $h^{-1}$ for clone 3 (Figure 4A). Globally, clone 2 showed the best growth performance compared to the two others.

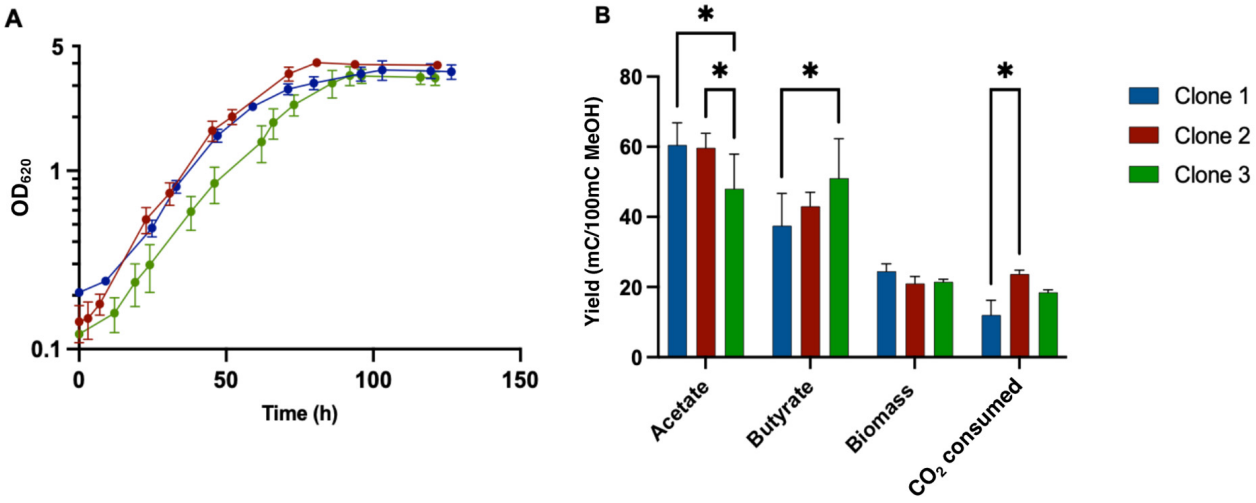



**Figure 4.** Physiological study of isolated clones from evolved populations growing on YE and cysteine-free synthetic medium with 200 mM methanol. (**A**) Growth profile monitoring by measuring the optical density (620 nm). (**B**) Carbon balance from methanol to products, biomass and $CO_2$ consumed. Cultures were performed in triplicate and error bars indicates standard deviation. Asterisk indicates that the results were significantly different from the control according to the one-sided Student's *t*-test: *: *p* value $\leq 0.05$.

Analysis of the carbon balance for the three clones showed that the acetate yield of clone 3, with a value of 48 ± 6, was significantly lower than those of clones 1 and 2, with values of 60.5 ± 4.5 and 59.7 ± 4.16 mC acetate/100 mC MetOH, respectively (Figure 4B). The butyrate yield of clone 1, with a value of 37.5 ± 6.5, was significantly lower than those of Clones 2 and 3, with values of 43 ± 4 and 51 ± 8 mC butyrate/100 mC MetOH, respectively. No significant difference was observed among the clones for biomass yield. However, $CO_2$ consumption was significantly higher for clone 2, with a value of 23.27 ± 1.15 mC $CO_2$ consumed/100 mC MetOH, compared to clone 1 and clone 3, with values of 12 ± 3 mC and 18.5 ± 0.5 $CO_2$ consumed/100 mC MetOH, respectively.

Culture of the wild-type strain showed no significant growth on MMM with sodium thiosulfate addition, indicating that genetic modification/regulation could be responsible for the ability of the strain to grow on MMM. Whole-genome sequencing of the 10th, 25th, 50th and 75th generations from the evolved population and the 3 isolated clones was performed to potentially identify mutated genes responsible for (i) the improved growth on MMM plus cysteine during ALE and (ii) the potential ability of the isolated clones to grow on MMM supplemented with thiosulfate.

### *3.8. Mutation Profiles of the ALE Lineage by Whole Genome Resequencing*

To better understand the underlying mechanisms of growth improvement on MMM supplemented with cysteine, the genome sequences of the evolved populations after the 10th, 25th, 50th, and 75th generations were determined. Furthermore, the genomes of the three selected clones growing on liquid MMM supplemented with thiosulfate were also sequenced (Table 3).

A total of 28 genetic modifications, including two conserved mutations, were identified by comparison with the wild-type strain sequence. Surprisingly, no mutation occurred in genes directly involved in methanol metabolism or the WLP. Two conserved one-base insertions spaced by only 69 bp were found in all generations and clones (except for one insertion that was not observed in clone 3) at the end and the beginning of two adjacent genes (coding for hypothetical proteins) and part of a large prophage operon. Such conserved mutations over generations suggest a response by genomic adaptation of the strain to the restrictive growth conditions on methanol-defined medium without YE. However, understanding the role of insertions in genes coding for hypothetical proteins remains difficult. An ALE experiment performed to improve the growth of the *Sporomusa ovata* acetogen on a liquid methanol mineral medium also showed similar insertions in genes belonging to a prophage operon [40]. In addition to the mutations in genes from a prophage operon, the structure of type I RM encoding genes of the three adapted clones was different compared to the WT strain (Figure 5).

The system is composed of one R subunit, one M subunit, one integrase and three S subunits. Among the S subunits annotated, two subunits seem truncated (M5595_00355 and M5595_00360), while one was complete (M5595_00345). We found four repeated sequences, including two repeated sequences in the complete S subunit and two other sequences in truncated S subunits in the WT and adapted strains. Interestingly, we observed a homologous recombination event without mutation in these S subunits between the adapted and WT strains, giving a complete S subunit in the truncated S subunit region and vice versa. The apparent complete S subunit encoding the gene sequence was 1149 bp for the WT strain versus 1236 bp for the clones 1, 2 and 3, indicating a different protein sequence. After alignment of protein sequences of both S subunits of WT and adapted strains, the structure appeared to be different at the C-terminal side of the protein (Figure 6). Furthermore, the search for conserved domains indicated two target recognition domains (TRDs), the first located in the 101–192 residue interval, which corresponds to the modified region, and the second in the 236–396 residue interval. The sequence analysis of the S subunit from the adapted strains strongly suggests the recognition of a different DNA sequence. This was confirmed by the determination of the methylome of the clone 2 (Table 2) that non only present a different recognition sequence for the type I RM (CNNTAYNNNNGTG instead of CNNTAYNNNNNTCC) but also a much higher proportion of CAAAAA $^{m6}$R methylated motifs (83.7 versus 37.3%).

Eleven mutations were detected in the 10th generation compared to the WT strain, with eight mutations at 100% frequency and three other mutations below 58.9% frequency. The only SNP detected in a gene coding for a metabolic enzyme was in the gene encoding the butyryl-CoA: acetate CoA transferase, but this mutation was not detected in the next generations, indicating that it was not needed to improve growth on methanol. Globally, the mutations found in early generations declined for more advanced evolved populations. Furthermore, the three isolated clones showed only a few mutations compared to the WT strain and shared no mutation in common except the ones in the prophage operon and the reorganization of the S subunit encoding genes of the type I RM.

**Table 3.** List of mutations observed for the 10th, 25th, 50th, and 75th generations and Clones 1, 2 and 3 during ALE against the WT strain of *E. limosum* B2. Blue indicates a high rate of mutation among the considered population, while red indicates a low rate of mutation. The numbers represent the percentage of the population carrying genetic changes compared to the WT strain.

| Locus Tag *E. limosum* B2 | Position | WT | Variant | Type | 10th | 25th | 50th | 75th | C1 | C2 | C3 | Gene | Gene in 5′ of Intergenic Region | Locus tag 5′ Gene Intergenic Region | Gene in 3′ of Intergenic Region | Locus tag 3′ Gene Intergenic Region |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| M5595_20825 | 4406858 | - | T | INS | 100 | 100 | 100 | 100 | 100 | 100 | 100 | Hypothetical protein | - | - | - | - |
| M5595_20830 | 4406927 | - | T | INS | 100 | 100 | 100 | 100 | 100 | 100 | | Hypothetical protein | - | - | - | - |
| M5595_00020 | 4683 | G | A | SNP | 100 | | 100 | 100 | 100 | | 100 | Hypothetical protein | - | - | - | - |
| M5595_04930 | 1017540 | C | A | SNP | 100 | | | | | 100 | 100 | Hypothetical protein | - | - | - | - |
| M5595_17525 | 3682586 | A | C | SNP | 36.2 | | 22 | | | | | Hypothetical protein | - | - | - | - |
| - | 1193634 | A | - | DEL | 58.9 | | | | | | | Intergenic region | Ketol-acid reductoisomerase (3′ -> 5′) | | Alpha-amylase family glycosyl hydrolase (3′ -> 5′) | |
| M5595_16535 | 3470810 | G | A | SNP | 100 | | | | | | | Sugar binding transcriptionnal regulator | - | - | - | - |
| M5595_12615 | 2631029 | T | C | SNP | 100 | | | | | | | HAD-IA family hydrolase | - | - | - | - |
| M5595_17525 | 3682598 | C | T | SNP | 31.8 | | | | | | | Hypothetical protein | - | - | - | - |
| - | 4262268 | C | - | DEL | 100 | | | | | | | Intergenic region | Threonine synthase (3′ -> 5′) | M5595_20065 | Homoserine dehydrogenase (5′ -> 3′) | M5595_20070 |
| M5595_06115 | 1274807 | T | C | SNP | 100 | | | | | | | Butyryl-CoA:acetate CoA-transferase | - | - | - | - |
| M5595_11550 | 2401587 | C | T | SNP | | 75.5 | | | | | | Prepilin peptidase | - | - | - | - |
| M5595_08805 | 1810120 | T | - | DEL | | | 95.9 | | | | | Hypothetical protein | - | - | - | - |
| - | 2743930 | A | - | DEL | | | 43.4 | | | | | Intergenic region | Helix-turn-helix transcriptional regulator (5′ -> 3′) | M5595_13160 | Lysozyme family protein (5′ -> 3′) | M5595_13165 |
| M5595_15025 | 3141664 | AT | - | DEL | | | 37 | | | | | Arsenate reductase family protein | - | - | - | - |
| M5595_10115 | 2083561 | A | - | DEL | | | 62.1 | | | | | Glycosyltransferase family 2 protein | - | - | - | - |
| - | 3378354 | T | C | SNP | | | | 70.2 | | | | Intergenic region | Rrf2 family transcriptional regulator (3′ -> 5′) | M5595_16095 | FtsX-like permease family protein (5′ -> 3′) | M5595_16100 |
| - | 2164707 | T | G | SNP | | | | 70.2 | | | | Intergenic region (riboswitch) | Leucine-rich repeat domain-containing protein (5′ -> 3′) | M5595_10385 | Energy-coupling factor ABC transporter permease (3′ -> 5′) | M5595_10390 |
| M5595_09145 | 1871776 | AACTG | - | DEL | | | | 25 | | | | Cysteine synthase A | - | - | - | - |
| - | 4051265 | T | - | DEL | | | | 88.2 | | | | Intergenic region | HPr family phosphocarrier protein (5′ -> 3′) | M5595_19190 | Aminodeoxychorismate/anthranilate synthase component II (5′ -> 3′) | M5595_19195 |
| M5595_03625 | 755658 | C | A | SNP | | | | 35.2 | | | | Peptidoglycan DD-metalloendopeptidase | - | - | - | - |
| - | 825607 | T | - | DEL | | | | 46.7 | | | | Intergenic region | PucR family transcriptional regulator (5′ -> 3′) | M5595_03980 | DUF917 domain-containing protein (5′ -> 3′) | M5595_03985 |
| M5595_00050 | 9361 | T | - | DEL | | | | 93.3 | | | | Hypothetical protein | - | - | - | - |
| - | 370141 | A | - | DEL | | | | 50 | 100 | 100 | | Intergenic region | ParB/RepB/Spo0J family partition protein (5′ -> 3′) | M5595_01730 | Hypothetical protein (5′ -> 3′) | M5595_01735 |
| M5595_05545 | 1145373 | T | C | SNP | | | | | | 100 | | Nucleotide sugar dehydrogenase | - | - | - | - |
| M5595_10115 | 2082950 | A | - | DEL | | | | | | | 100 | Glycosyltransferase family 2 protein | - | - | - | - |
| - | 3881186 | T | C | SNP | | | | 25 | | | | Intergenic region | tRNA-Phe (5′ -> 3′) | M5595_18430 | Cobalamin-dependent protein (5′ -> 3′) | M5595_18435 |
| M5595_00360 | 73656 | - | T | INS | | | | | | | 100 | Restriction endonuclease subunit S | - | - | - | - |

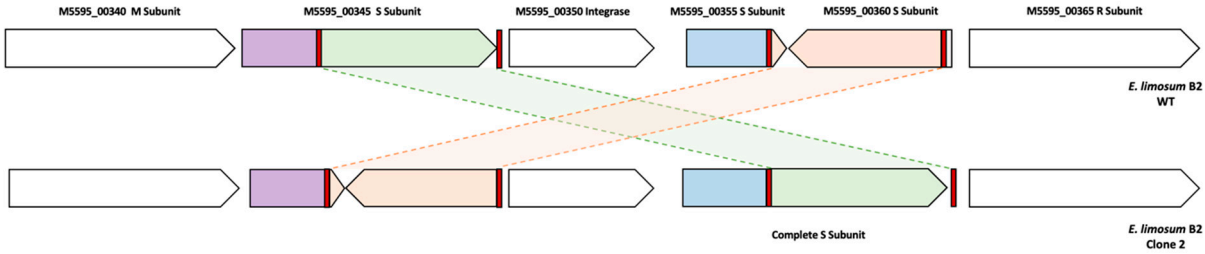



**Figure 5.** Restriction-modification system type I organization in *E. limosum* B2 WT and clone 2. Repeated sequences are highlighted in red while homologous recombination is represented by interchange of green (714 bp) and orange (642 bp) genomic parts. The organization of this region in clones 1 and 3 was the same as clone 2.

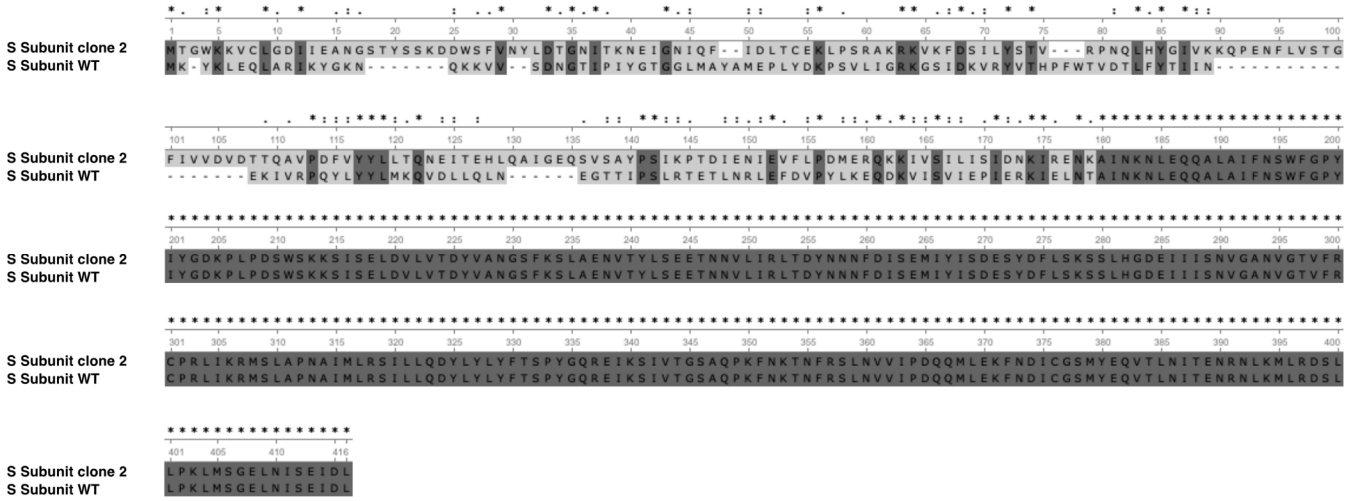



**Figure 6.** Protein sequence alignment of the apparent complete S subunit of RM type I system for WT and clone 2 strains. * indicates conserved aminoacids. Clones 1, 2 and 3 produce the same S subunit of RM type I system.

Whole genome sequencing for the 10th, 25th, 50th, and 75th generations and isolated clones during the ALE showed no evidence of genetic modification in genes involved in methanol oxidation or in genes coding for the WLP and enzymes for energy conservation. However, the WT strain could grow in liquid methanol mineral medium without YE, while the evolved clone 2 could grow if sodium thiosulfate was added. To characterize the effect of the modified methylome on gene expression and also better understand the adaptation mechanisms on synthetic medium, the complete proteome profile was performed for WT and clone 2 strains.

### *3.9. Proteomic Analysis of the WT and Evolved Clone 2*

#### 3.9.1. Global Analysis of the Proteomes

The total proteome of *E. limosum* B2 was compared between the WT strain and clone 2 on MMM supplemented with 0.5 g/L YE to sustain the growth of the WT strain, as it failed to grow on MMM supplemented with sulphur as the only mineral source. The optimal cell density remained similar for both strains, with an $OD_{620}$ of 4.0. However, the specific growth rate of clone 2 reached 0.084 ± 0.005 $h^{-1,}$ compared to 0.035 ± 0.001 $h^{-1}$ for the WT strain, improving the maximal growth rate 2.4-fold for the adapted strain compared to the WT strain (Figure 7). Although the MMM was supplemented with YE, the clone 2 strain showed a significant improvement in growth compared to the native strain.

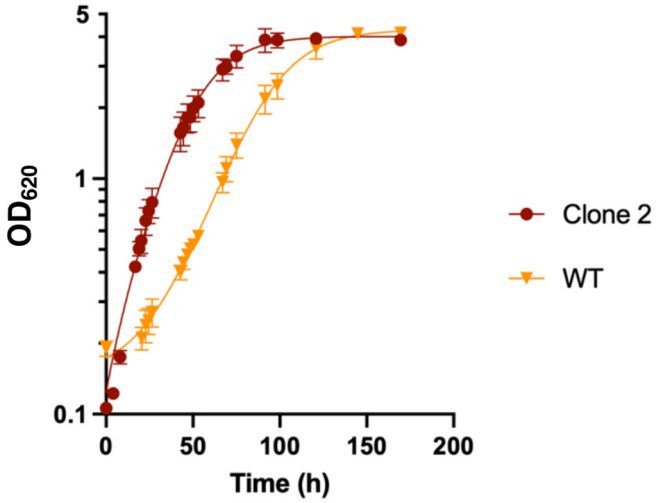



**Figure 7.** Growth profile of the WT and clone 2 strains in methanol synthetic medium supplemented with 0.5 $g \cdot L^{-1}$ YE. Cultures were performed in triplicate and error bars indicates standard deviation.

To evaluate the efficiency of total proteome extraction, the abundance of ribosomal proteins was analysed as a representative of cytosolic proteins, while the extraction of membrane-bound proteins was assessed by examining the proteins of the ATP synthase complex. A total of 46 ribosomal proteins out of 51 were identified for both conditions, indicating efficient extraction using the SPEED method with an average coefficient of variation of 11.5% for the clone 2 strain and 35% for WT (Figure 8A). The variability among cytosolic proteins between replicates for the native strain was superior to that for the adapted strain. This difference might be explained by a different biomass composition between strains. After cell lysis in trifluoroacetic acid (TFA), a larger amount of precipitated compound was observed for the native strain compared to the adapted strain. Based on the ratio of the total ribosomal intensity value to the total proteome intensity value for both conditions, the relative number of ribosomes per cell was 2.57-fold higher in the clone 2 strain than in the native strain. This value was closely related to the specific growth rate difference of both strains, as the clone 2 strain showed a specific growth rate 2.4-fold higher than that of the WT strain.

Regarding the efficiency of the total protein extraction method for membrane-bound proteins, seven out of nine constitutive proteins of the ATP synthase complex were identified. The ATP synthase in *E. callanderi* KIST 612 was identified to be a unique combination of a $Na^+$ $A_1A_0$ ATP synthase with a V-type *c* subunit organized in an operon of nine genes [41]. A similar operon was found in the genome of *E. limosum* B2, and seven out of nine constitutive proteins of the ATPase operon were identified by mass spectrometry (Figure 8B). The ATPase complex is organized as an $A_1A_0$ ATP synthase with a catalytic head constituted by A and B subunits, a central stalk comprising the cytoplasmic subunits C, D and F and a rotor located in the membrane including eight to ten *c* and *a* subunits that are stoichiometrically equivalent. In addition, A-type ATP synthases contain subunits H and E [42]. The precise stoichiometry for the ATP synthase of *E. limosum* B2 is currently unknown and requires further investigation using imaging technology to determine the structure of the complex [43]. For example, the stoichiometry of *Pyrococcus furiosus* was established with the following formula: $A_3B_3CDE_2FH_2ac_{10}$ [44]. Based on the previous stochiometric formula, the intensity value measured for the different subunits of the ATP synthase complex did not match. A clear difference in the proportion of total proteome intensity was observed between the *a* and *c* subunits, the C and D subunits, and the A and B subunits. The extraction method applied to the samples did not allow total membrane-bound protein extraction but allowed the identification of a sufficient proportion of membranous proteins, as numerous transporters were identified through this extraction method coupled to mass spectrometry.

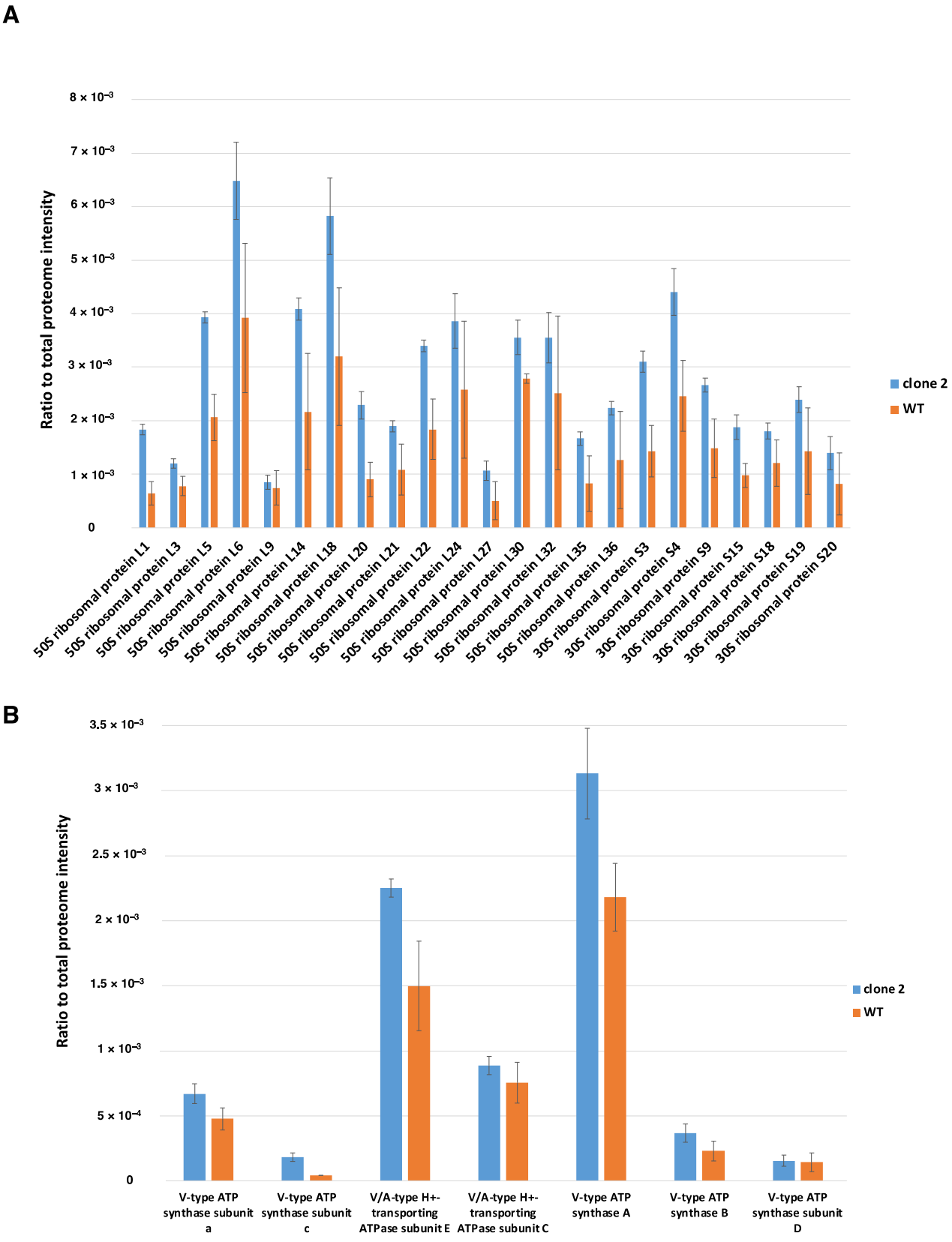



**Figure 8.** Assessment of the total proteome extraction. (**A**) Assessment of the total cytosolic protein extraction by comparison of all ribosomal proteins extracted. (**B**) Assessment of total proteome extraction on membrane bound proteins, example of ATPase complex. Data expressed in proportion of total proteome intensity. Blue bars correspond to clone 2 strain, and orange bars correspond to the WT strain.

By combining replicates of both strain cultures, a total of 1256 different proteins were detected, with 1197 proteins identified for the clone 2 strain and 1015 proteins for the WT. A difference of 182 supplementary proteins was observed for the clone 2 strain, suggesting that the mobilization of higher protein diversity led to better growth performance on methanol synthetic medium. The complete list of differentially produced proteins is available in Table S1.

The distribution of differentially produced proteins between the clone 2 strain and native strain (One-sided Student's *t*-test, $p$ value < 0.05) showed over- and underproduced proteins with more than or less than 1 $Log_2$ fold change. The global proteome showed substantial differences between both conditions, with a total of 91 significantly differentially produced proteins with a fold change (FC) value >2 or <−2 and a $p$ value < 0.05. A total of 52 proteins were overproduced, while 39 proteins were underproduced in the clone 2 strain compared to the WT strain (Figure 9). Proteins from various metabolic functions and pathways were identified as differentially expressed between the conditions, suggesting metabolic flux reorganization between the two strains.

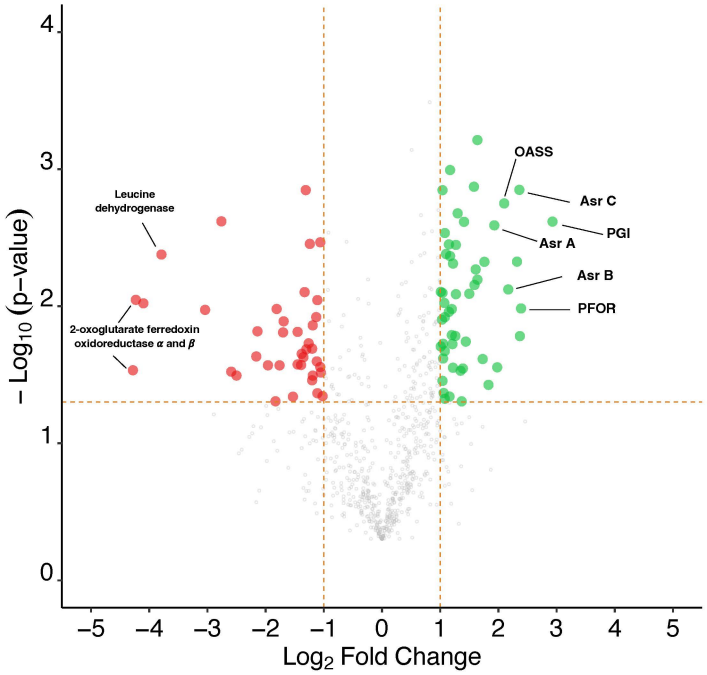



**Figure 9.** Volcano plot representing the results of total proteome analysis between the clone 2 strain against the WT strain. Green dots show significantly overproduced proteins, while red dots show significantly underproduced proteins. Horizontal dotted line represents the applied significance threshold of a one-sided student's *t*-test $p$ value of 0.05. Vertical dotted lines represent the applied threshold of a $log_2$ fold change of >1 and <−1, corresponding to an absolute fold change >2 or <−2. Asr, anaerobic sulphite reductase; OASS, O-acetyl-homoserine sulphhydrylase; PFOR, pyruvate ferredoxin oxidoreductase; PGI, glucose-6-phosphate isomerase.

### 3.9.2. Analysis of the Methanol Oxidation and Wood–Ljungdahl Pathways

Methanol oxidation has been well characterized in acetogens with evidence of reverse use of the methyl branch of WLP in *E. callanderi* KIST612, a closely related species, to sustain reducing equivalent equilibrium [16]. Additionally, energy production and reducing equivalent balance were reported to be ensured by the $Na^+$-dependent Rnf complex associated with the ATP synthase complex, creating a proton motive force by importing $Na^+$ ions to oxidize NADH and reduce Fd [15,16]. Before entering the WLP, methanol is oxidized and coupled to a tetrahydrofolate (THF) molecule by a methanol-specific methyltransferase complex composed of three proteins containing two catalytic domains [45]. Methanol corrinoid methyltransferase (MTI) transfers the methyl group of methanol to the central

cobalt atom of corrinoid protein (CoP), followed by the transfer of the methyl group to a THF molecule by methyltetrahydrofolate cobalamin methyltransferase (MTII). No mutation occurred in the central metabolism coupled to the energy conservation system, and as expected, the proteomic analysis revealed a high abundance of proteins involved in methanol oxidation and the WLP for both conditions, but although the level of all these proteins was higher in the evolved clone, only two were above the 2-fold difference that we set as a threshold (Figure 10). Carbon monoxide dehydrogenase (CooS) was significantly more abundant in the adapted strain, showing an FC value of 2.22, as well as the thiolase, FC value of 2.08, involved in butyrogenesis. Globally, proteins of the CODH/ACS complex in the carbonyl branch, which are involved in $CO_2$ fixation into acetyl-CoA, and proteins of the butyrogenesis pathway were slightly overproduced compared to proteins involved in the methyl branch and acetogenesis.

#### 3.9.3. Analysis of Gluconeogenesis and Anaplerotic Reactions

When glucose is not available during autotrophic or methylotrophic conditions, acetyl-CoA, $CO_2$, ATP and reducing power are used in gluconeogenic reactions to produce all the intermediates needed for nucleic acid, amino acid, cell wall and essential cofactor biosynthesis [46]. The first step is catalysed by the pyruvate-ferredoxin oxidoreductase (PFOR) A0A317RSE4 to produce pyruvate [47] from acetyl-CoA, reduced ferredoxin and $CO_2$ (Figure 11).

Interestingly, substantial overproduction of PFOR was observed, showing an FC value of 5.23, and this was among the most abundant differentially produced proteins detected. Another gluconeogenic enzyme, glucose-6-phosphate isomerase (PGI), was the most overproduced protein, with an FC value of 7.60. Glucose 6-phosphate was demonstrated to represent a key metabolite for peptidoglycan as well as cell wall polysaccharide synthesis. Such upregulation of both the PFOR- and PGI-encoding genes must be a solution developed by the evolved strain to improve its growth capacity in synthetic medium. The carbon recovered in biomass for the clone 2 strain represented approximately 20% of the total carbon metabolized during methylotrophic fermentation against approximately 6.5% for the 10th generation population (Figures 4 and 5).

The pyruvate carboxylase (PC), which catalyses the carboxylation of pyruvate into oxaloacetate, the first metabolic intermediary of the TCA cycle, was underproduced (FC value −2.29). Furthermore, the 2-oxoglutarate ferredoxin oxidoreductase subunits alpha and beta (KGOR) were the most highly underproduced proteins (FC values −17.11 and −19.47, respectively). As isocitrate dehydrogenase was upregulated (FC 3.11), this might allow a high flux of alpha-ketoglutarate production that could further be used for glutamate production.

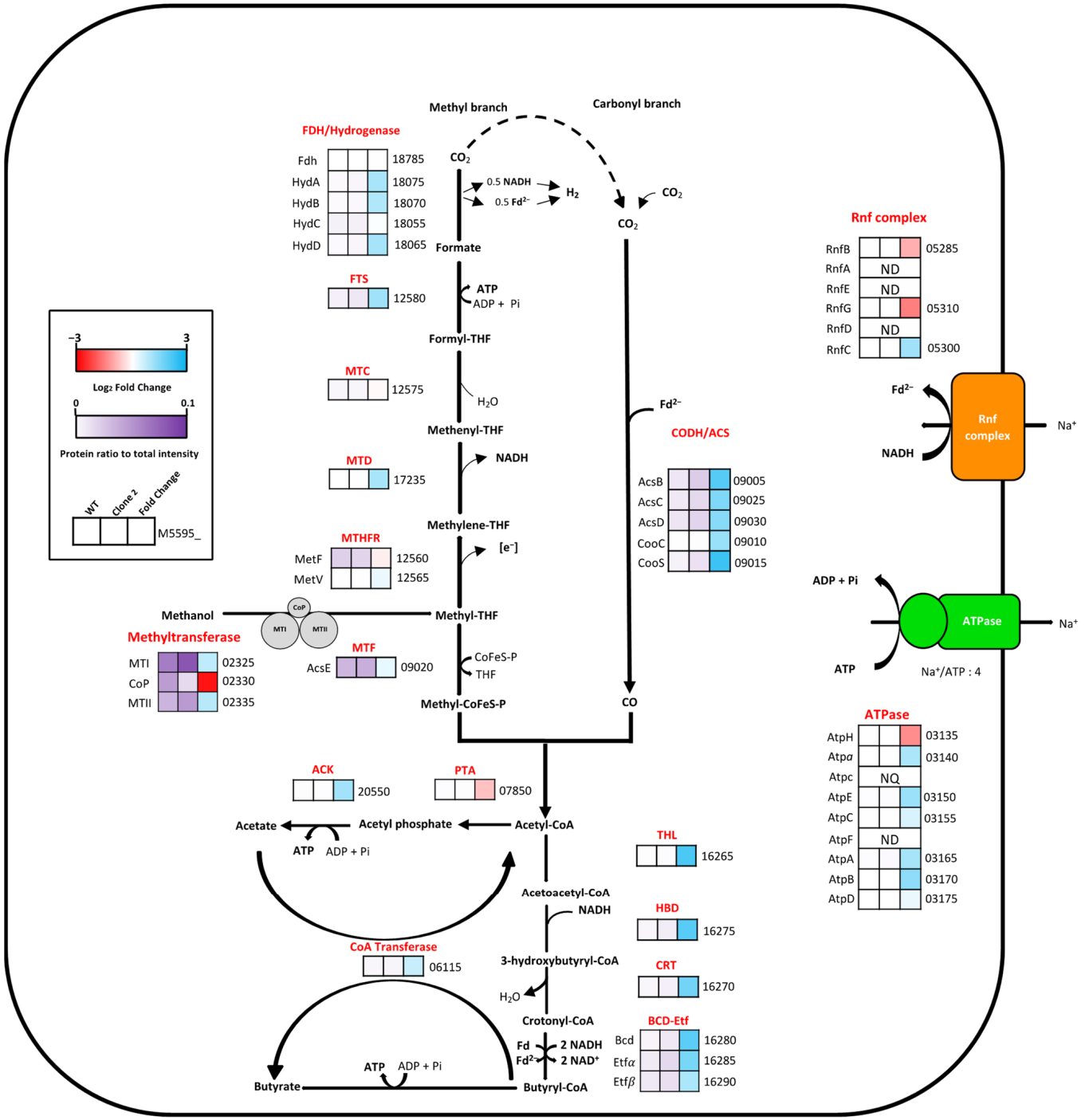



**Figure 10.** Comparative enzyme abundance in the central carbon metabolism between the WT and the adapted strain clone 2 in methanol synthetic medium. The first and second boxes correspond to the relative protein abundance expressed as a proportion of total intensity value of the WT and clone 2 strains, respectively. The third box corresponds to the $Log_2$ fold change of protein abundance of the clone 2 strain relative to the WT strain. The number indicated at the right of each box corresponds to the *E. limosum* B2 locus tag, starting with M5595; [$e^-$], undefined reducing equivalent; NQ, not quantified; ND, not detected. List of proteins is available in Table S1.

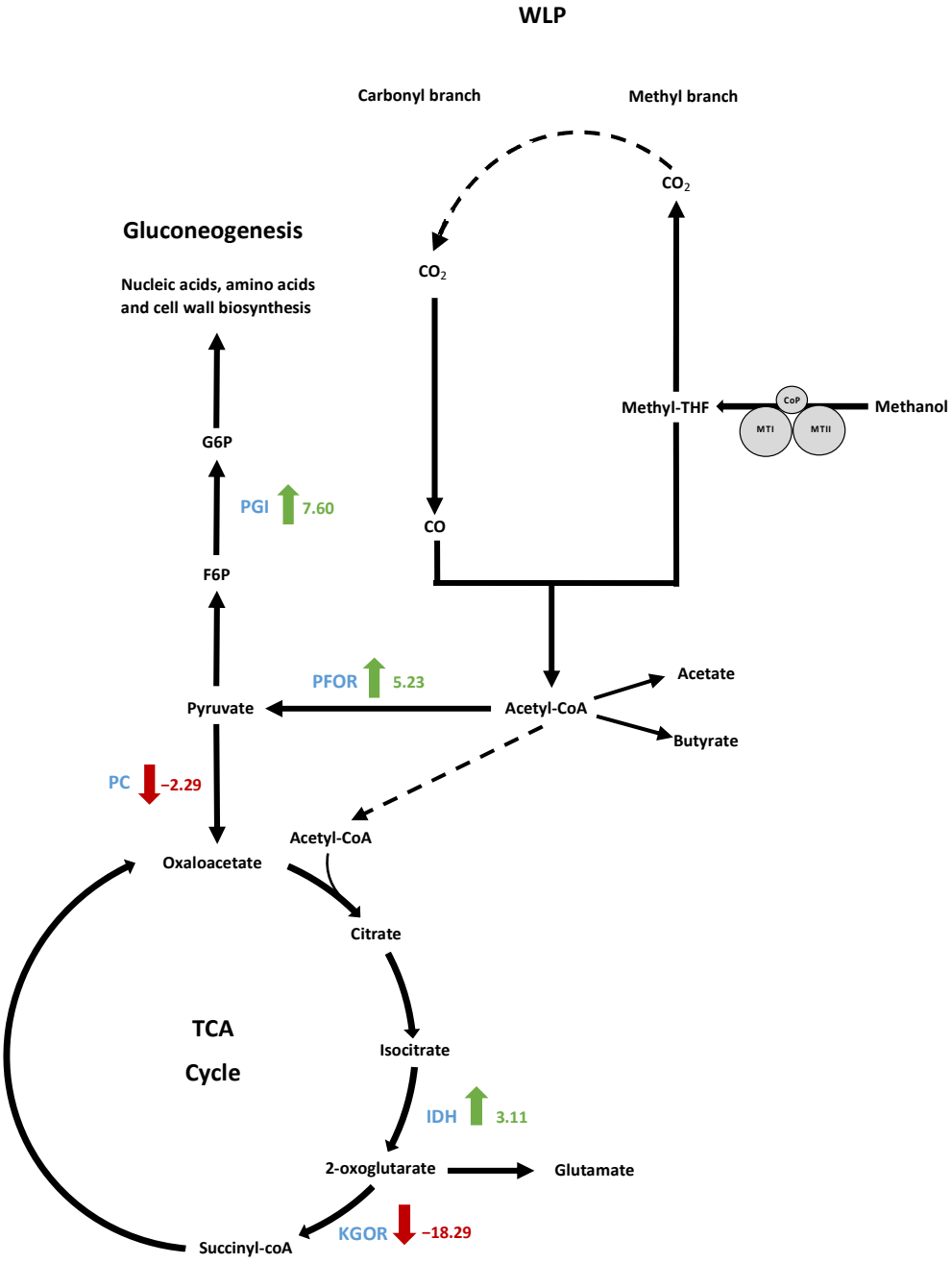



**Figure 11.** Schematic pathway showing significatively differentially produced proteins in clone 2 strain compared to WT strain in gluconeogenesis and tricarboxylic acid cycle alongside WLP. Pathways were simplified to show only proteins of interest. Green arrows indicate protein overproduction and red arrows indicate underproduction in clone 2 strain compared to WT strain. The scores indicate the fold change values. IDH, isocitrate dehydrogenase; KGOR, 2-oxoglutarate ferredoxin oxidoreductase; PC, pyruvate carboxylase; PFOR, pyruvate ferredoxin oxidoreductase; PGI, glucose-6-phosphate isomerase.

#### 3.9.4. Sulphate Metabolism for the Biosynthesis of Sulphur-Containing Amino Acids

The clone 2 strain can grow on methanol mineral medium with the addition of sodium thiosulfate instead of cysteine, which is required for the WT strain. Interestingly, anaerobic sulphite reductase subunits A, B and C (A0A317RV53, A0A317RVB5 and A0A317RTV7) were among the most statistically overproduced proteins in the adapted clone 2, showing an average FC value of 4.47. In mycobacteria, anaerobic sulphite reductase has previously been demonstrated to be essential for growth on sulphate or sulphite as the sole source

of sulphur [48]. Anaerobic sulphite reductase catalyses sulphide ($S^{2-}$) production from sulphite ($SO_3^{2-}$). Cysteine is then synthesized from sulphide and serine by a two-step enzymatic pathway [49]. First, serine is acetylated by a serine transacetylase (CysE) to form O-acetylserine, and second, cysteine is produced from O-acetylserine and sulphide via a reaction catalysed by a cysteine synthase (CysK). Furthermore, CysK (A0A317RWJ0) was overexpressed in the clone 2 strain with an FC value of 1.87. There remains some uncertainty regarding the sulphite supply by thiosulfate. Sulphite can be produced either by the reduction of sulphate ($SO_4^{2-}$) or by the reduction of thiosulfate by a thiosulfate reductase [50]. However, no thiosulfate reductase-encoding gene was found in the genome of *E. limosum* B2 or among the detected proteins, and $SO_4^{2-}$ from iron sulphate did not sustain the growth of the strain in synthetic medium. Where does the sulphite come from if sulphate is not the sulphur source? Evidence of cysteine production with thiosulfate was reported in *E. coli* [51]: L-serine can be acetylated to O-acetylserine by a serine acetyltransferase, and the product can be further converted with the addition of thiosulfate to S-sulphocysteine by an O-acetyl-homoserine sulphhydrylase OASS (A0A317RPB8). Then, L-cysteine can be synthesized from S-sulphocysteine by thioredoxin/glutaredoxin, liberating sulphite, which would be recycled for further cysteine production. The OASS was highly overproduced in the adapted strain (FC 4.30), suggesting the utilization of this pathway to produce cysteine in the adapted clone 2 (Figure 12).

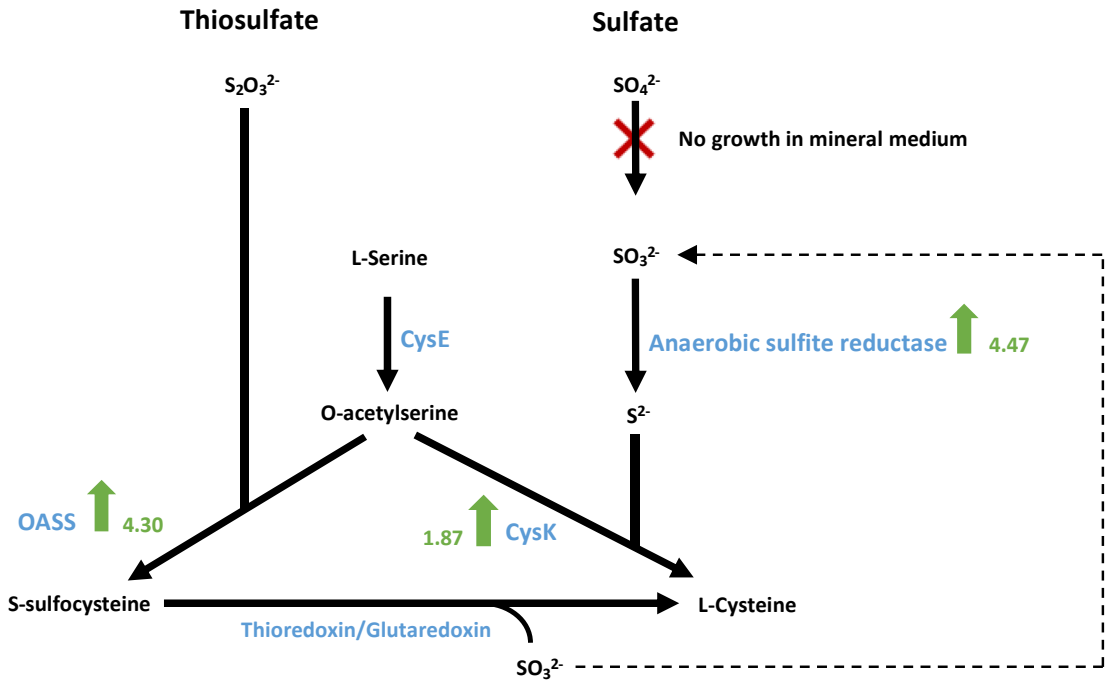



**Figure 12.** Theoretic pathway for sulphur assimilation and cysteine biosynthesis in *E. limosum* B2 in mineral medium. Green arrows indicate protein overproduction in clone 2 strain compared to WT. The scores indicate the fold change values. CysE, serine transacetylase; CysK, cysteine synthase; OASS, O-acetyl-homoserine.

Interestingly, 5-methyltetrahydrofolate-homocysteine methyltransferase, also known as methionine synthase, was the second strongest underproduced protein (FC −3.56) among amino acid biosynthesis-related proteins. This enzyme was shown to have methionine synthase activity [52]. Methionine synthase is involved in the last step of methionine biosynthesis, using homocysteine and methyl-cobalamin as cofactors. Decreased production of methionine synthase could suggest the preservation of homocysteine to produce cysteine via the transsulfuration pathway [53]. In addition to cysteine production through L-serine transformation and thiosulfate utilization, this pathway was probably used by the clone 2 strain to produce cysteine. This function likely emerged during the evolution of

the strain, as cystathionine gamma-synthase (A0A317RYT5) and cystathionine beta-lyase (A0A317RVG5), two key enzymes of the transsulfuration pathway, were only detected in the clone 2 strain.

#### 3.9.5. Other Differentially Produced Proteins

Because all ALE processes were performed without YE, the evolved strain strongly regulated its metabolism to produce the required amino acids. While proteins involved in amino acid biosynthesis were found to be slightly overproduced, others were strongly underproduced, such as leucine dehydrogenase (A0A1H0LY56), with an FC value of −13.81.

Surprisingly, several enzymes involved in propanediol utilization (*pdu* operon) were strongly underproduced in clone 2, indicating an important function in the native strain. The enzyme BMC domain-containing protein (A0A0U3FC86) which is involved in the bacterial microcompartment for 1,2 propanediol (1,2 PDO) or glycerol metabolism exhibits significant underproduction (FC −6.76). The enclosed proteinaceous structure metabolizes 1,2-propanediol to propanol and propionate or to produce 3-hydroxypropionate from glycerol. Encapsulation is necessary to limit cytosolic exposure to propionaldehyde, a toxic metabolic intermediary [54]. The presence of cytosolic microcompartments was described for *E. maltosivorans* with 1,2 PDO or betaine as carbon sources. Furthermore, BMC structures were also reported on *A. woodii*, a close relation, with 1,2 PDO and 2,3 butanediol as carbon sources, but no BMC was detected with methanol as a carbon source. The presence of proteins from the *pdu* operon at high intensity for the native strain remains enigmatic, as no propanol or 3-hydroxypropionate was detected by HPLC.

## 4. Discussion

In this study, the genome of the most studied (from a physiological point of view) strain of *E. limosum* was sequenced. The *E. limosum* B2 genome shows 99.837% sequence identity with that of the *E. limosum* ATCC 8486 strain. Although the two genomes are very similar, they exhibit chromosomal rearrangement in prophage operons, insertion sequences and 5 colinear blocks (LCBs) with a reverse complement orientation. Furthermore, several differences were found in genes coding for regulators and an S subunit-encoding gene of a type I RM system.

*Eubacterium limosum* B2 required unknown components in the complex media for growth on methanol [10]. To overcome this problem, we applied an adaptive evolution approach to develop an optimized strain of *E. limosum* B2 capable of growth on a methanol-defined medium without yeast extract. By this approach, the growth rate was enhanced 3.45-fold in 75 generations. From the evolved population, three clones were isolated and grown on methanol mineral medium without cysteine (which was replaced by sodium thiosulfate). Among the three isolated clones, the clone 2 strain showed the best growth characteristics, showing a maximal growth rate 2.72-fold higher than that of the 10th generation population cultured in methanol-defined medium supplemented with cysteine. Genome sequencing of the WT, 10th, 25th, 50th, and 75th generations and three evolved clones showed no specific mutation in central carbon metabolism or conserved mutation over generations except in a large phage operon and in the S subunit encoding genes of a type I RM system. Proteomics data, on the other hand, showed important differences in the proteomic compositions of the WT and clone 2. First, overproduction was observed for the enzymes of both (i) the carbonyl branch of the WLP and (ii) the butyrogenesis pathway. This should lead to an improvement in $CO_2$ fixation ability by both (i) an enhancement in reduced ferredoxin production by the electron bifurcating complex Bcd-Etf in the butyrogenesis pathway and (ii) by an increase in the level if CODH that also requires reduced $Fd^{2-}$ as a source of electrons.

Better growth of clone 2 could also be the result of higher production of enzymes involved in the synthesis of key building blocks of the cells, such as the PFOR involved in pyruvate production from acetyl-CoA, PGI responsible for glucose-6-P production from fructose-6-P or IDH involved in a-keto-glutarate production needed for glutamate production. The ability of clone 2 to grow on a methanol mineral medium in the absence

of cysteine could be the result of a combination of higher production of (i) the anaerobic sulphite reductase and (ii) the enzymes of the transsulphuration pathway allowing cysteine biosynthesis from thiosulfate. The considerable modification of the proteomic profile of the evolved clone 2 was associated to changes of its methylome profile suggesting that it is the result of an epigenetic phenomenon. The rearranged S subunit of the type I RM in the evolved clones, recognize a different motif (CNNTAYNNNNGTG instead of CNNTAYNNNNNTCC) and a higher proportion of these motifs are methylated. Similarly, a much higher proportion of type II motifs (CAAAAA $^{m6}$R) are also methylated (83.7 versus 37.3% in wild type). These modifications in the methylome profile are probably responsible in the large change in gene expression. A similar phenomenon was previously demonstrated in *Bacillus pumilus* when the type I RM-encoding genes were deleted [55]. This study highlighted the metabolic/genetic adaptation of the *E. limosum* B2 strain growing on methanol-defined medium and showed the importance of improving our knowledge of cell adaptation mechanisms in a constrained environment. The evolved clone 2 that grows on a mineral medium should allow us to develop and use a genome-scale model of *E. limosum* B2 for both flux analysis and rational metabolic engineering of *E. limosum* B2, as previously performed with *Clostridium acetobutylicum* [56]. Furthermore, to better characterize the metabolism of this strain, it can now be used in chemostat cultures in mineral medium at a constant specific growth rate using different carbon sources to perform a complete systems biology analysis.

**Supplementary Materials:** The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/microorganisms10091790/s1. Table S1: Complete list of proteins detected by mass spectrometry for *E. limosum* B2 WT and clone 2 strains.

**Author Contributions:** Conceptualization, G.P. and P.S.; methodology, G.P. and P.S.; validation, G.P. and P.S.; resources, P.S.; data curation, G.P. and P.S.; writing—original draft preparation, G.P.; writing—review and editing, N.P.M. and P.S.; supervision, P.S.; project administration, P.S.; funding acquisition, P.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Era CoBiotech program, grant number ANR-17-COBI-0002 and BBSRC BB/T010630/1.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Methylome, raw reads from whole-genome sequencing and raw proteomics data are available upon request. The annotated genome sequence of *E. limosum* B2 WT was deposited in Genbank (NCBI) under the accession number CP097376.1.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Wood, H.G.; Ragsdale, S.W.; Pezacka, E. The Acetyl-CoA Pathway of Autotrophic Growth. *FEMS Microbiol. Rev.* **1986**, *2*, 345–362. [CrossRef]
2. Bengelsdorf, F.R.; Straub, M.; Dürre, P. Bacterial Synthesis Gas (Syngas) Fermentation. *Environ. Technol.* **2013**, *34*, 1639–1651. [CrossRef]
3. Henstra, A.M.; Sipma, J.; Rinzema, A.; Stams, A.J.M. Microbiology of Synthesis Gas Fermentation for Biofuel Production. *Curr. Opin. Biotechnol.* **2007**, *18*, 200–206. [CrossRef]
4. Dwidar, M.; Park, J.-Y.; Mitchell, R.J.; Sang, B.-I. The Future of Butyric Acid in Industry. *Sci. World J.* **2012**, *2012*, 471417. [CrossRef]
5. Dürre, P. Butanol Formation from Gaseous Substrates. *FEMS Microbiol. Lett.* **2016**, *363*, fnw040. [CrossRef]
6. Karlson, B.; Bellavitis, C.; France, N. Commercializing Lanza Tech, from Waste to Fuel: An Effectuation Case. *J. Manag. Organ.* **2021**, *27*, 175–196. [CrossRef]
7. Bertsch, J.; Müller, V. Bioenergetic Constraints for Conversion of Syngas to Biofuels in Acetogenic Bacteria. *Biotechnol. Biofuels* **2015**, *8*, 210. [CrossRef] [PubMed]
8. Cotton, C.A.; Claassens, N.J.; Benito-Vaquerizo, S.; Bar-Even, A. Renewable Methanol and Formate as Microbial Feedstocks. *Curr. Opin. Biotechnol.* **2020**, *62*, 168–180. [CrossRef]
9. Giuliano, A.; Freda, C.; Catizzone, E. Techno-Economic Assessment of Bio-Syngas Production for Methanol Synthesis: A Focus on the Water–Gas Shift and Carbon Capture Sections. *Bioengineering* **2020**, *7*, 70. [CrossRef]

10. Pacaud, S.; Loubiere, P.; Goma, G. Methanol Metabolism by *Eubacterium limosum* B2: Effects of pH and Carbon Dioxide on Growth and Organic Acid Production. *Curr. Microbiol.* **1985**, *12*, 245–250. [CrossRef]
11. Loubière, P.; Gros, E.; Paquet, V.; Lindley, N.D. Kinetics and Physiological Implications of the Growth Behaviour of *Eubacterium limosum* on Glucose/Methanol Mixtures. *Microbiology* **1992**, *138*, 979–985. [CrossRef]
12. Loubiere, P.; Pacaud, S.; Goma, G.; Lindley, N.D. The Effect of Formate on the Acidogenic Fermentation of Methanol by *Eubacterium limosum*. *J. Gen. Appl. Microbiol.* **1987**, *33*, 463–470. [CrossRef]
13. Pacaud, S.; Loubiere, P.; Goma, G.; Lindley, N.D. Effects of Various Organic Acid Supplements on Growth Rates of *Eubacterium limosum* B2 on Methanol. *Appl. Microbiol. Biotechnol.* **1986**, *24*, 75–78. [CrossRef]
14. Loubiere, P.; Goma, G.; Lindley, N.D. A Non-Passive Mechanism of Butyrate Excretion Operates during Acidogenic Fermentation of Methanol by *Eubacterium limosum*. *Antonie Van Leeuwenhoek* **1990**, *57*, 83–89. [CrossRef]
15. Kremp, F.; Poehlein, A.; Daniel, R.; Müller, V. Methanol Metabolism in the Acetogenic Bacterium *Acetobacterium Woodii*. *Environ. Microbiol.* **2018**, *20*, 4369–4384. [CrossRef]
16. Dietrich, H.M.; Kremp, F.; Öppinger, C.; Ribaric, L.; Müller, V. Biochemistry of Methanol-Dependent Acetogenesis in *Eubacterium callanderi* KIST612. *Environ. Microbiol.* **2021**, *23*, 4505–4517. [CrossRef]
17. Wood, J.C.; Marcellin, E.; Plan, M.R.; Virdis, B. High Methanol-to-Formate Ratios Induce Butanol Production in *Eubacterium limosum*. *Microb. Biotechnol.* **2021**, *15*, 1542–1549. [CrossRef]
18. Marinos, G.; Kaleta, C.; Waschina, S. Defining the Nutritional Input for Genome-Scale Metabolic Models: A Roadmap. *PLoS ONE* **2020**, *15*, e0236890. [CrossRef]
19. Genthner, B.R.; Davis, C.L.; Bryant, M.P. Features of Rumen and Sewage Sludge Strains of *Eubacterium limosum*, a Methanol- and $H_2$-$CO_2$-Utilizing Species. *Appl. Environ. Microbiol.* **1981**, *42*, 12–19. [CrossRef]
20. Richter, H.; Loftus, S.E.; Angenent, L.T. Integrating Syngas Fermentation with the Carboxylate Platform and Yeast Fermentation to Reduce Medium Cost and Improve Biofuel Productivity. *Environ. Technol.* **2013**, *34*, 1983–1994. [CrossRef]
21. Dragosits, M.; Mattanovich, D. Adaptive Laboratory Evolution–Principles and Applications for Biotechnology. *Microb. Cell Factories* **2013**, *12*, 64. [CrossRef]
22. Kang, S.; Song, Y.; Jin, S.; Shin, J.; Bae, J.; Kim, D.R.; Lee, J.-K.; Kim, S.; Cho, S.; Cho, B.K. Adaptive Laboratory Evolution of *Eubacterium limosum* ATCC 8486 on Carbon Monoxide. *Front. Microbiol.* **2020**, *11*, 402. [CrossRef] [PubMed]
23. Erickson, L.E. Biomass Elemental Composition and Energy Content. *Biotechnol. Bioeng.* **1980**, *22*, 451–456. [CrossRef]
24. Tatusova, T.; DiCuccio, M.; Badretdin, A.; Chetvernin, V.; Nawrocki, E.P.; Zaslavsky, L.; Lomsadze, A.; Pruitt, K.D.; Borodovsky, M.; Ostell, J. NCBI Prokaryotic Genome Annotation Pipeline. *Nucleic Acids Res.* **2016**, *44*, 6614–6624. [CrossRef]
25. Darling, A.E.; Mau, B.; Perna, N.T. ProgressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement. *PLoS ONE* **2010**, *5*, e11147. [CrossRef]
26. Doellinger, J.; Schneider, A.; Hoeller, M.; Lasch, P. Sample Preparation by Easy Extraction and Digestion (SPEED)-A Universal, Rapid, and Detergent-Free Protocol for Proteomics Based on Acid Extraction. *Mol. Cell. Proteom.* **2020**, *19*, 209–222. [CrossRef]
27. Demichev, V.; Messner, C.B.; Vernardis, S.I.; Lilley, K.S.; Ralser, M. DIA-NN: Neural Networks and Interference Correction Enable Deep Proteome Coverage in High Throughput. *Nat. Methods* **2020**, *17*, 41–44. [CrossRef]
28. Kelly, W.J.; Henderson, G.; Pacheco, D.M.; Li, D.; Reilly, K.; Naylor, G.E.; Janssen, P.H.; Attwood, G.T.; Altermann, E.; Leahy, S.C. The Complete Genome Sequence of *Eubacterium limosum* SA11, a Metabolically Versatile Rumen Acetogen. *Stand. Genom. Sci.* **2016**, *11*, 26. [CrossRef]
29. Roh, H.; Ko, H.-J.; Kim, D.; Choi, D.G.; Park, S.; Kim, S.; Chang, I.S.; Choi, I.-G. Complete Genome Sequence of a Carbon Monoxide-Utilizing Acetogen, *Eubacterium limosum* KIST612. *J. Bacteriol.* **2011**, *193*, 307–308. [CrossRef]
30. Song, Y.; Shin, J.; Jeong, Y.; Jin, S.; Lee, J.-K.; Kim, D.R.; Kim, S.C.; Cho, S.; Cho, B.-K. Determination of the Genome and Primary Transcriptome of Syngas Fermenting *Eubacterium limosum* ATCC 8486. *Sci. Rep.* **2017**, *7*, 13694. [CrossRef]
31. Weirather, J.L.; de Cesare, M.; Wang, Y.; Piazza, P.; Sebastiano, V.; Wang, X.-J.; Buck, D.; Au, K.F. Comprehensive Comparison of Pacific Biosciences and Oxford Nanopore Technologies and Their Applications to Transcriptome Analysis. *F1000Research* **2017**, *6*, 100. [CrossRef] [PubMed]
32. Hu, L.; Xiao, P.; Jiang, Y.; Dong, M.; Chen, Z.; Li, H.; Hu, Z.; Lei, A.; Wang, J. Transgenerational Epigenetic Inheritance Under Environmental Stress by Genome-Wide DNA Methylation Profiling in Cyanobacterium. *Front. Microbiol.* **2018**, *9*, 1479. [CrossRef] [PubMed]
33. Blow, M.J.; Clark, T.A.; Daum, C.G.; Deutschbauer, A.M.; Fomenkov, A.; Fries, R.; Froula, J.; Kang, D.D.; Malmstrom, R.R.; Morgan, R.D.; et al. The Epigenomic Landscape of Prokaryotes. *PLoS Genet.* **2016**, *12*, e1005854. [CrossRef]
34. Jensen, T.Ø.; Tellgren-Roth, C.; Redl, S.; Maury, J.; Jacobsen, S.A.B.; Pedersen, L.E.; Nielsen, A.T. Genome-Wide Systematic Identification of Methyltransferase Recognition and Modification Patterns. *Nat. Commun.* **2019**, *10*, 3311. [CrossRef] [PubMed]
35. Choe, D.; Lee, J.H.; Yoo, M.; Hwang, S.; Sung, B.H.; Cho, S.; Palsson, B.; Kim, S.C.; Cho, B.-K. Adaptive Laboratory Evolution of a Genome-Reduced *Escherichia coli*. *Nat. Commun.* **2019**, *10*, 935. [CrossRef]
36. Espinosa, M.I.; Gonzalez-Garcia, R.A.; Valgepea, K.; Plan, M.R.; Scott, C.; Pretorius, I.S.; Marcellin, E.; Paulsen, I.T.; Williams, T.C. Adaptive Laboratory Evolution of Native Methanol Assimilation in *Saccharomyces Cerevisiae*. *Nat. Commun.* **2020**, *11*, 5564. [CrossRef]
37. Har, J.R.G.; Agee, A.; Bennett, R.K.; Papoutsakis, E.T.; Antoniewicz, M.R. Adaptive Laboratory Evolution of Methylotrophic *Escherichia coli* Enables Synthesis of All Amino Acids from Methanol-Derived Carbon. *Appl. Microbiol. Biotechnol.* **2021**, *105*, 869–876. [CrossRef]

38. Wang, Y.; Fan, L.; Tuyishime, P.; Liu, J.; Zhang, K.; Gao, N.; Zhang, Z.; Ni, X.; Feng, J.; Yuan, Q.; et al. Adaptive Laboratory Evolution Enhances Methanol Tolerance and Conversion in Engineered *Corynebacterium glutamicum*. *Commun. Biol.* **2020**, *3*, 1–15. [CrossRef]
39. Mariotto, C.; Loubière, P.; Goma, G.; Lindley, N.D. Influence of Various Reducing Agents on Methylotrophic Growth and Organic Acid Production of *Eubacterium limosum*. *Appl. Microbiol. Biotechnol.* **1989**, *32*, 193–198. [CrossRef]
40. Tremblay, P.-L.; Höglund, D.; Koza, A.; Bonde, I.; Zhang, T. Adaptation of the Autotrophic Acetogen *Sporomusa ovata* to Methanol Accelerates the Conversion of $CO_2$ to Organic Products. *Sci. Rep.* **2015**, *5*, 16168. [CrossRef]
41. Litty, D.; Müller, V. A Na+ A1AO ATP Synthase with a V-Type c Subunit in a Mesophilic Bacterium. *FEBS J.* **2020**, *287*, 3012–3023. [CrossRef] [PubMed]
42. Coskun, U.; Grüber, G.; Koch, M.H.J.; Godovac-Zimmermann, J.; Lemker, T.; Müller, V. Cross-Talk in the A1-ATPase from *Methanosarcina mazei* Gö1 Due to Nucleotide Binding. *J. Biol. Chem.* **2002**, *277*, 17327–17333. [CrossRef] [PubMed]
43. Cossio, P.; Allegretti, M.; Mayer, F.; Müller, V.; Vonck, J.; Hummer, G. Bayesian Inference of Rotor Ring Stoichiometry from Electron Microscopy Images of Archaeal ATP Synthase. *Microscopy* **2018**, *67*, 266–273. [CrossRef] [PubMed]
44. Vonck, J.; Pisa, K.Y.; Morgner, N.; Brutschy, B.; Müller, V. Three-Dimensional Structure of A1A0 ATP Synthase from the Hyperthermophilic Archaeon *Pyrococcus furiosus* by Electron Microscopy. *J. Biol. Chem.* **2009**, *284*, 10110–10119. [CrossRef]
45. Kremp, F.; Müller, V. Methanol and Methyl Group Conversion in Acetogenic Bacteria: Biochemistry, Physiology and Application. *FEMS Microbiol. Rev.* **2021**, *45*, fuaa040. [CrossRef]
46. Bar-Even, A.; Flamholz, A.; Noor, E.; Milo, R. Thermodynamic Constraints Shape the Structure of Carbon Fixation Pathways. *Biochim. Biophys. Acta* **2012**, *1817*, 1646–1659. [CrossRef]
47. Furdui, C.; Ragsdale, S.W. The Role of Pyruvate Ferredoxin Oxidoreductase in Pyruvate Synthesis during Autotrophic Growth by the Wood-Ljungdahl Pathway. *J. Biol. Chem.* **2000**, *275*, 28494–28499. [CrossRef]
48. Pinto, R.; Harrison, J.S.; Hsu, T.; Jacobs, W.R.; Leyh, T.S. Sulfite Reduction in Mycobacteria. *J. Bacteriol.* **2007**, *189*, 6714–6722. [CrossRef]
49. Kitabatake, M.; So, M.W.; Tumbula, D.L.; Söll, D. Cysteine Biosynthesis Pathway in the Archaeon *Methanosarcina barkeri* Encoded by Acquired Bacterial Genes? *J. Bacteriol.* **2000**, *182*, 143–145. [CrossRef]
50. Hallenbeck, P.C.; Clark, M.A.; Barrett, E.L. Characterization of Anaerobic Sulfite Reduction by *Salmonella typhimurium* and Purification of the Anaerobically Induced Sulfite Reductase. *J. Bacteriol.* **1989**, *171*, 3008–3015. [CrossRef]
51. Nakatani, T.; Ohtsu, I.; Nonaka, G.; Wiriyathanawudhiwong, N.; Morigasaki, S.; Takagi, H. Enhancement of Thioredoxin/Glutaredoxin-Mediated L-Cysteine Synthesis from S-Sulfocysteine Increases L-Cysteine Production in *Escherichia coli*. *Microb. Cell Factories* **2012**, *11*, 62. [CrossRef] [PubMed]
52. Paul, L.; Krzycki, J.A. Sequence and Transcript Analysis of a Novel *Methanosarcina barkeri* Methyltransferase II Homolog and Its Associated Corrinoid Protein Homologous to Methionine Synthase. *J. Bacteriol.* **1996**, *178*, 6599–6607. [CrossRef] [PubMed]
53. Aitken, S.M.; Lodha, P.H.; Morneau, D.J.K. The Enzymes of the Transsulfuration Pathways: Active-Site Characterizations. *Biochim. Biophys. Acta BBA-Proteins Proteom.* **2011**, *1814*, 1511–1517. [CrossRef] [PubMed]
54. Sampson, E.M.; Bobik, T.A. Microcompartments for $B_{12}$-Dependent 1,2-Propanediol Degradation Provide Protection from DNA and Cellular Damage by a Reactive Metabolic Intermediate. *J. Bacteriol.* **2008**, *190*, 2966–2971. [CrossRef]
55. Liu, G.; Jiang, Y.-M.; Liu, Y.-C.; Han, L.-L.; Feng, H. A Novel DNA Methylation Motif Identified in *Bacillus pumilus* BA06 and Possible Roles in the Regulation of Gene Expression. *Appl. Microbiol. Biotechnol.* **2020**, *104*, 3445–3457. [CrossRef]
56. Nguyen, N.-P.-T.; Raynaud, C.; Meynial-Salles, I.; Soucaille, P. Reviving the Weizmann Process for Commercial N-Butanol Production. *Nat. Commun.* **2018**, *9*, 3682. [CrossRef]