



HAL
open science

Clostridium beijerinckii DSM 6423, une souche plateforme émergente pour la bioproduction de solvants

Rémi Hocq

► **To cite this version:**

Rémi Hocq. Clostridium beijerinckii DSM 6423, une souche plateforme émergente pour la bioproduction de solvants. Biotechnologie. Institut agronomique, vétérinaire et forestier de France, 2019. Français. NNT : 2019IAVF0026 . tel-04784315

HAL Id: tel-04784315

<https://pastel.hal.science/tel-04784315v1>

Submitted on 15 Nov 2024

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

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

NNT° : 2019 IAVF 0026

THESE DE DOCTORAT

préparée à l'Institut des sciences et industries du vivant et de l'environnement (AgroParisTech)

pour obtenir le grade de

Docteur de l'Institut agronomique vétérinaire et forestier de France

Spécialité : Biotechnologies

École doctorale n° 581

Agriculture, alimentation, biologie, environnement et santé (ABIES)

par

Rémi HOCQ

***Clostridium beijerinckii* DSM 6423,**

une souche plateforme émergente pour la bioproduction de solvants

Directeur de thèse : Nicolas LOPES FERREIRA
Co-encadrement de la thèse : François WASELS

Thèse présentée et soutenue à Rueil-Malmaison, le 16 Décembre 2019 :

Composition du jury :

M. Bruno DUPUY, Directeur de Recherche, Institut Pasteur
Mme Olga SOUTOURINA, Professeure, Université Paris-Sud
Mme Chantal TARDIF, Professeure, Université Aix-Marseille
Mme Laurence ETTWILLER, Chercheuse, New England Biolabs
M. François WASELS, Chercheur, IFP Energies Nouvelles
M. Nicolas LOPES FERREIRA, Chercheur, IFP Energies Nouvelles

Président
Rapporteur
Rapporteur
Examinatrice
Co-encadrant
Directeur de thèse

Remerciements

J'aimerais tout d'abord remercier les membres de mon jury de thèse pour avoir accepté d'évaluer mon travail, et IFPEN pour m'avoir donné l'opportunité de le réaliser au sein de son département Biotechnologie.

Je voudrais bien sûr remercier plus particulièrement mes encadrants, Nicolas et François, pour avoir dirigé mes travaux de thèse tout en me laissant une grande liberté dans mes recherches. Nicolas, ne change rien, je ne sais pas si, dans mon cas, j'aurais pu avoir un meilleur directeur de thèse. François, merci pour tous tes conseils, et pour les quelques verres écoulés après de dures journées de labeur, toujours dans le sérieux et la modération les plus totaux bien entendu !

J'aimerais également remercier Gwladys pour tout ce qu'elle a fait pour ma thèse (c'est-à-dire beaucoup). Sans elle, ma thèse n'aurait sans doute pas été aussi réussie. Tu es vraiment une technicienne exceptionnelle, j'espère que tu en as conscience et que tu continueras à travailler comme ça, efficacement et dans la bonne humeur !

Mes remerciements vont également à Maxime, un stagiaire lui aussi exceptionnel, et dont la persévérance a permis quelques avancées majeures dans ma thèse. Maxime, ça a été vraiment un plaisir de t'encadrer, j'espère que la suite pour toi te réservera plein de bonnes choses et je ne doute pas que d'un point de vue professionnel tu continues sur cette bonne lancée.

A Sandra et Angélique, merci pour toute votre aide et vos conseils en fermentation. Grâce à vous, je n'aurai rien fait exploser, je me serai seulement contenté de faire déborder... A Benji, merci pour les nombreux coups de main ! Et, avec un peu d'avance, félicitation monsieur l'ingénieur ! Je te souhaite le meilleur pour la suite. A Frédérique, merci pour nos nombreuses conversations, tu me manqueras ! A Juju, si ton travail est dans l'ombre, il n'en demeure pas moins essentiel pour les recherches menées au département en général et pour ma thèse en particulier, et je t'en remercie.

J'aimerais également remercier les autres thésards et post-docs, membres fidèles de "la pause-café de 16h en bas du bâtiment", pour les cafés bien sûr, et pour les quelques soirées mémorables ! Bonne chance à tous ceux qui restent (vous en aurez besoin !).

Plus généralement, je remercie tous les membres de ce département pour leur accueil et leur gentillesse. Je ne sais pas encore ce que mon chemin professionnel me réserve, mais j'espère sincèrement que j'aurai la chance de me retrouver dans une équipe avec une aussi bonne ambiance de travail à l'avenir.

Merci également à mes parents, ma sœur et mes amis qui m'ont soutenu (ou supporté, selon) pendant ces 3 années. Merci à Virginia, enfin, pour tout.

Abréviations

ABE: Acétone-Butanol-Ethanol

Ack: acétate kinase

ACP : analyse en composantes principales

Acr : anti-CRISPR

ADC: acétoacetate décarboxylase

ADN : acide désoxyribonucléique

AdhE1,2: alcool/aldéhyde déshydrogénase 1, 2

ARN : acide ribonucléique

ARNcr: ARN CRISPR

ARNcrtr: ARN CRISPR trans-encodé

ARNg: ARN guide

ARNm: ARN messager

ARNnc: ARN non codant

ARNr: ARN ribosomal

ARNt: ARN de transfert

aTc : anhydrotetracycline

ATCC : *American Type Culture Collection*

ATP : adénosine tri-phosphate

BCD : butyryl-CoA déshydrogénase

BCS : synthèse du butyryl-CoA (opéron)

BdhA,B,C: butanol déshydrogénase A,B,C

Buk: butyrate kinase

C. XXX: Clostridium XXX

C2/3/4/5/6: molécules à 2/3/4/5/6 carbones

Cas9 : *CRISPR associated protein 9*

catB : amphénicol acétyltransférase B

CoA : coenzyme A

CRISPR: *clustered regularly interspaced palindromic repeats*

Crt : crotonase

DO₆₀₀ : densité optique à 600 nm

DSM: *Deutsche Sammlung von Mikroorganismen*

EMP : Embden-Meyerhof-Parnas, voie de la glycolyse

Fd : ferrédoxine

FRT : Site de reconnaissance de la recombinaison Flp

GC : chromatographie en phase gazeuse

Hbd : β -hydroxybutyryl-CoA déshydrogénase

HPLC: chromatographie en phase liquide à haute performance

IBE : Isopropanol-Butanol-Ethanol

IEP : intron-encoded protein

InterA : intergénique antisens

InterS : intergénique sens

IntraA : intragénique antisens

IntraS : intragénique sens

kb: kilobase

LB: *lysogeny broth*

LC-MS/MS: liquid chromatography coupled with tandem mass spectrometry

NADH/NAD⁺ : Nicotinamide adenine dinucleotide, forme réduite / oxydée

NADPH/NADP⁺ : Nicotinamide adenine dinucleotide phosphate, forme réduite / oxydée

NCIMB: *National Collection of Industrial, food, and Marine Bacteria*

nt: nucléotide

pb : paire de bases

ORF: cadre ouvert de lecture

Pfor : pyruvate-ferrédoxine oxydoréductase

Pta : phosphate transacétylase

Ptb : phosphate transbutyrylase

RCC : répression catabolique du carbone

RNase : ribonucléase

RPM : lecture par million

s-Adh : alcool déshydrogénase secondaire

SPT : système phosphotransférase

Thl: thiolase

TSS : site d'initiation de la transcription

udk : uridine kinase

udp : uridine phosphatase

upp : uridine phosphoribosyltransférase

UTR : région non traduite

Sommaire

1	Analyse Bibliographique	1
1.1	Vers la production de propylène biosourcé	1
1.1.1	Contexte	1
1.1.2	Présentation des <i>Clostridia</i> solvantogènes	6
1.1.3	La fermentation A/IBE, du XIX ^{ème} siècle à aujourd'hui	8
1.2	Production d'isopropanol par fermentation IBE	11
1.2.1	Caractéristiques fermentaires A/IBE	11
1.2.2	Fermentation de substrats industriels : état de l'art chez Clostridium	16
1.3	Metabolisme des Clostridia A/IBE	23
1.3.1	Généralités	24
1.3.2	Du substrat carboné au pyruvate	26
1.3.3	Du pyruvate aux solvants	32
1.3.4	La solvantogenèse, un processus influencé par de nombreux facteurs	40
1.4	Régulation du métabolisme A/IBE	44
1.4.1	Les omiques, pour une meilleure compréhension du métabolisme A/IBE	44
1.4.2	L'analyse des TSS, un nouvel outil d'analyse de la régulation de l'expression..	48
1.5	Optimisation génétique des microorganismes solvantogènes de genre Clostridium	56
1.5.1	Généralités	56
1.5.2	Les outils classiques de modification du génome	57
1.5.3	La révolution CRISPR	63
2	Problématiques de la thèse	69
3	Partie I : Développement d'un outil d'édition ciblée du génome basé sur CRISPR-Cas9	73
3.1	Résumé des travaux publiés	73
3.2	Valorisation scientifique (Article)	79
3.3	Protection intellectuelle (Brevets)	82
4	Partie II : Rôle de σ^{54} dans le métabolisme de <i>Clostridium beijerinckii</i>	85
4.1	Résumé des travaux publiés	85
4.2	Valorisation scientifique (article)	88
5	Partie III : Analyse comparative des TSS dans trois <i>Clostridia</i> solvantogènes	89
5.1	Introduction	89
5.2	Matériel et Méthodes	92
5.2.1	Souches, milieu et conditions de culture	92

5.2.2	Extraction des ARN	92
5.2.3	Préparation des bibliothèques RNA-seq	92
5.2.4	Préparation des bibliothèques eCAPP-seq	93
5.2.5	Traitement des données eCAPP-seq	93
5.2.6	Visualisation des données	94
5.2.7	Northern blot	94
5.3	Résultats	95
5.3.1	La normalisation par l'expression améliore la détection des TSS	95
5.3.2	Le eCAPP-seq identifie des milliers de TSS pour trois <i>Clostridia</i>	99
5.3.3	Le eCAPP-seq révèle un contrôle différencié de la conversion de l'acétoacétyl-CoA entre <i>C. beijerinckii</i> et <i>C. acetobutylicum</i>	101
5.4	Discussion	103
5.5	Figures supplémentaires	107
6	Partie IV : Métabolisme sur substrat 1G, approches fermentaire et multi-omique	113
6.1	Introduction	113
6.2	Matériel et Méthodes	115
6.2.1	Milieux de culture	115
6.2.2	Fermentations	116
6.2.3	Analytique	117
6.2.4	Extractions de matériel biologique	118
6.2.5	Transcriptomique	119
6.2.6	Protéomique	119
6.2.7	Visualisation des données	121
6.3	Résultats	121
6.3.1	Choix du substrat 1G	121
6.3.2	Séparation des phases acidogènes et solvantogènes en fermentation continue 122	
6.3.3	Plan d'expériences et analyse fermentaire	124
6.3.4	Analyse omique préliminaire	128
6.4	Discussion	143
6.5	Figures supplémentaires	148
7	Discussion générale	175
8	References	181
9	Annexe	195

Liste des figures

Figure 1: Taxonomie des <i>Clostridia</i> solvantogènes A/IBE d'après Poehlein <i>et al.</i>	7
Figure 2 : Représentation schématique d'un bioréacteur en mode continu, tel qu'utilisé dans cette étude.	12
Figure 3 : Evolution typique d'une fermentation en bioréacteur en mode <i>batch</i> pour la souche DSM 6423.	14
Figure 4: Structure des deux composés glycosidiques de l'amidon : l'amylose et l'amylopectine.	18
Figure 5: Cycle de vie des <i>Clostridia</i>	26
Figure 6: Représentation schématique de l'import du glucose, fructose et saccharose et de leur métabolisation par la voie EMP chez <i>C. beijerinckii</i>	30
Figure 7 : Représentation schématique des voies majeures du métabolisme A/IBE.	33
Figure 8 : Procédure de préparation des bibliothèques pour le <i>Capp-switch sequencing</i>	54
Figure 9 : Analyse transcriptomique des motifs impliqués dans l'expression différentielle des TSS pour chaque substrat testé, chez <i>C. phytofermantans</i>	55
Figure 10: Description schématique de l'utilisation de la méthode ClosTron pour l'inactivation de gène chez les <i>Clostridia</i>	59
Figure 11 : Stratégie d'échange allélique développée par Al-Hinai <i>et al.</i>	61
Figure 12 : Description schématique du fonctionnement du système CRISPR chez <i>S. pyogenes in vivo</i>	63
Figure 13 : Description schématique du fonctionnement du système CRISPR-Cas9 pour l'édition de génome eucaryote.	65
Figure 14 : Description schématique des différentes stratégies employées dans cette étude pour modifier génétiquement la souche DSM 6423.	74
Figure 15 : Préparation des bibliothèques et analyse de données eCAPP-seq.	91
Figure 16 : Le eCAPP-seq améliore l'identification des TSS chez <i>C. beijerinckii</i> DSM 6423.	96
Figure 17 : Exemples de TSS retrouvés chez <i>C. beijerinckii</i> DSM 6423 pour a. un gène fortement exprimé et b. un groupe de gènes faiblement exprimés.	98
Figure 18 : Caractéristiques transcriptomiques générales des jeux de données eCAPP-seq de <i>C. beijerinckii</i> DSM 6423, <i>C. beijerinckii</i> NCIMB 8052 et <i>C. acetobutylicum</i> ATCC 824.	99
Figure 19 : Analyse Northern Blot des structures transcriptionnelle des gènes de synthèse du butyryl-CoA (BCS).	102
Figure 20 : Profils de productions de solvants par la souche <i>C. beijerinckii</i> DSM 6423 sur substrats 1G en fioles serties, points finaux.	122
Figure 21 : Titres cumulés des acides et des solvants dans les différents échantillons.	125
Figure 22 : Rendements cumulés en acides et en solvants dans les différents échantillons. ..	125

Figure 23 : Analyse en composantes principales : représentation des variables selon les deux premières composantes de la variance (cercle des corrélations).	126
Figure 24 : Analyse en composantes principales : représentation des individus selon les deux premières composantes de la variance.	127
Figure 25 : Séparation des protéines totales sur gel SDS-PAGE 10%.	129
Figure 26 : Analyse en composantes principales : représentation des individus passés en protéomique selon les deux premières composantes de la variance.	130
Figure 27 : Comparaison acidogène / solvantogène, <i>volcano plots</i>	131
Figure 28 : Comparaison acidogène / solvantogène, protéines communes dans chaque jeu de données.	132
Figure 29 : Comparaison acidogène / solvantogène, <i>heatmap</i> des catégories « Division cellulaire » et « Motilité ».	134
Figure 30 : Comparaison acidogène / solvantogène, <i>heatmap</i> de la catégorie « Métabolisme AIBE ».	136
Figure 31 : Alignement <i>blastP</i> entre CIBE_3469 et <i>sudB</i> de <i>P. furiosus</i>	138
Figure 32 : Comparaison acidogène / solvantogène, <i>heatmaps</i> des catégories « Choc thermique » et « Sporulation ».	140
Figure 33 : Comparaison substrat à substrat, <i>heatmaps</i> des catégories « Transport et métabolisme du carbone » et « Transport et métabolisme des molécules inorganiques ».	141

Liste des Tableaux

Tableau 1 : Composition moyenne des constituants principaux des mélasses de canne et de betterave (Ölbrich, 1963).	21
Tableau 2: Liste non exhaustive des substrats carbonés métabolisables par <i>C. acetobutylicum</i> et <i>C. beijerinckii</i> , basée sur les travaux de Keis <i>et al.</i>	28
Tableau 3: Principales voies de consommation de sources carbonées (hors glucose, fructose, saccharose).	31
Tableau 4 : Liste de facteurs impliqués dans la régulation de l'initiation de la transcription de gènes importants pour le cycle de vie des <i>Clostridia</i> solvantogènes.	49

Liste des Figures Supplémentaires

Figure Supplémentaire 1 : Tableau récapitulatif des données issues du séquençage eCAPP-seq et RNA-seq.	107
Figure Supplémentaire 2 : Tableau récapitulatif des amorces utilisées pour obtenir les sondes du <i>northern blot</i>.	108
Figure Supplémentaire 3 : Distribution des expression RNA-seq pour les gènes liés aux TSS identifiés.	108
Figure Supplémentaire 4 : Tableau récapitulatif du nombre et de la classification des TSS dans chaque souche.	109
Figure Supplémentaire 5 : Exemples de gènes présentant un TSS antisens dans les trois souches.	110
Figure Supplémentaire 6 : Nuage de points représentant, pour la souche DSM 6423, la longueur du 5'UTR d'un gène par rapport à son niveau d'expression en RNA-seq.	111
Figure Supplémentaire 7 : Certificat d'analyses réalisé par France Mélasse, donnant la composition de la mélasse de betterave utilisée dans cette étude.	149
Figure Supplémentaire 8 : Dosage de la biomasse et des métabolites dans les échantillons biologiques utilisés dans cette étude.	150
Figure Supplémentaire 9 : Catégorisation manuelle de 657 protéines détectées en protéomique.	151

ANALYSE BIBLIOGRAPHIQUE

1.1 VERS LA PRODUCTION DE PROPYLENE BIOSOURCE

1.1.1 Contexte

Depuis plus d'un siècle, la production d'énergie repose essentiellement sur les ressources fossiles. En 2014, plus de 80% de l'approvisionnement en énergie primaire à l'échelle planétaire est obtenu à partir de ces ressources, partagées entre le charbon (29%), le pétrole (31%) et le gaz naturel (21%)¹. Cette répartition cache toutefois des disparités importantes selon les secteurs. Le domaine des transports par exemple, représentant en France plus d'un tiers des dépenses énergétiques nationales, montre encore une dépendance forte au pétrole². En 2017, les carburants alternatifs à l'essence et au gazole pétroliers ne représentaient que 7,7 % des carburants consommés avec un taux de pénétration en stagnation.

Dans le cadre de la transition énergétique, la recherche de solutions alternatives à la consommation des ressources fossiles reste donc une priorité. Elles doivent notamment permettre de limiter les rejets en dioxyde de carbone, principal gaz à effet de serre massivement rejeté par l'homme et responsable d'un réchauffement climatique global³ qui aura des répercussions potentiellement graves⁴. Le protocole de Kyoto⁵, et plus récemment les accords de Paris⁶, traduisent une volonté commune de la majorité des pays d'évoluer vers un modèle à faible impact carbone, excluant à long terme l'utilisation des ressources fossiles et promouvant des alternatives « vertes ».

Parallèlement à cette problématique écologique, la question de la disponibilité des ressources fossiles en elles-mêmes est une autre source d'inquiétudes. La théorie du pic pétrolier postule que l'évolution de la production de pétrole au cours du temps suit une courbe dite « de Hubbert », qui passe nécessairement par un pic⁷. Ce dernier a déjà été observé pour certains pays⁸ et illustre une répartition géographique hétérogène de cette richesse. L'Europe est par exemple très dépendante des pays producteurs, plus de 50% de l'énergie consommée étant issue de produits d'importation⁹. En 2013, l'Agence Internationale de l'Energie estimait que le Moyen-Orient serait seul à pouvoir répondre à partir de 2020 à l'augmentation de la demande mondiale en pétrole¹⁰. Le développement d'un mix énergétique plus équilibré entre

énergies fossiles et renouvelables est essentiel pour à terme limiter cette dépendance vis-à-vis des pays producteurs, mais la réalisation d'un tel objectif demande de stimuler l'innovation dans la quête de sources d'énergies alternatives et locales.

Parmi les différentes solutions explorées pour limiter la dépendance aux ressources fossiles du secteur des transports, la valorisation de la biomasse végétale a suscité un intérêt considérable. De par sa composition, riche en carbone, cette biomasse peut être utilisée facilement comme combustible après bio-raffinage¹¹. Sa production nécessite une fixation importante de dioxyde de carbone, et son utilisation contribue ainsi à réduire la quantité de ce gaz à effet de serre dans l'atmosphère¹². Cette nouvelle filière est aujourd'hui encouragée par les acteurs publics (par exemple : objectif européen de 10% du pouvoir calorifique inférieur issus de sources renouvelables dans les transports d'ici à 2020¹³), notamment par le biais de mesures fiscales incitatives, comme la taxe TGAP mise en place par la loi de finances de 2005 en France. Grâce à ce type de mesures, une première filière de biocarburants issues des organes de réserve de plantes vivrières (grains de maïs, de blé ou de colza, racines de betterave) a pu prendre une part de marché conséquente – presque 4% en 2015, soit 79 Mtep¹⁴ - dans le marché des carburants routiers. Le carburant SP95-E10, contenant jusqu'à 10 % en volume d'éthanol, est ainsi devenu le premier carburant consommé par les Français dans les véhicules essence avec une part de marché de près de 43 % en 2018².

Mais cette première génération de biocarburants (1G) suscite un certain nombre de controverses, notamment vis-à-vis de la compétition avec l'industrie agroalimentaire pour les terres arables¹⁵. Face à cette polémique, les efforts de recherche se tournent sur l'utilisation des éléments de structure des plantes. La biomasse lignocellulosique, dite de « seconde génération » ou 2G, peut être obtenue à partir de déchets des industries agroalimentaires et forestières mais également de terrains non exploités pour la production agroalimentaire. Cette ressource, disponible en quantités compatibles avec des procédés industriels, présente toutefois des caractéristiques qui la rendent difficile d'utilisation. Sa structure enchevêtrée, basée sur de multiples liaisons covalentes entre ses différents constituants (cellulose/hémicellulose/lignine) rend coûteux le bioprocédé de production d'éthanol-carburant par voie fermentaire à cause du prétraitement physico-chimique aujourd'hui nécessaire à l'hydrolyse enzymatique de la cellulose. Ceci multiplie *in fine* par deux le

coût de production par rapport à un éthanol-carburant de première génération¹⁶. Les biocarburants de deuxième génération, pour être concurrentiels, doivent donc à l'heure actuelle être supportés par des mesures fiscales adaptées, ce qui met une fois de plus en exergue le rôle important des acteurs publics pour promouvoir ce type d'énergie.

Outre les carburants, un très grand nombre de molécules chimiques carbonées hautement valorisables sont issues du raffinage du pétrole ou du gaz naturel¹⁷. Depuis plus de 100 ans, la pétrochimie repose sur le raffinage de quelques molécules « plateformes », dont l'utilisation combinée permet d'obtenir un vaste répertoire de produits dérivés. Avec plus de 145 Mt de production annuelle mondiale en 2016, l'éthylène, précurseur notable du polyéthylène et du polystyrène, deux matières plastiques d'importance, en est un parfait exemple. Il est aujourd'hui quasi-exclusivement obtenu par vapocraquage des paraffines simples (C2-C4) du gaz naturel ou du naphta, un produit issu de la distillation du pétrole¹⁸. La production pétrochimique globale, incluant celle de l'éthylène, s'élevait en 2016 à plus de 450 Mt de produits¹⁹. Avec une croissance moyenne de plus de 4% par an, il s'agit d'un marché non négligeable, qui est toutefois soumis aux mêmes contraintes environnementales (émission de gaz à effet de serre) et économiques (disponibilité des sources et dépendance énergétique) que les carburants issus des ressources fossiles.

Pour ces raisons, et de la même façon que pour les biocarburants, de nouvelles filières responsables, écologiques, locales émergent via une chimie « verte »²⁰, qui prend notamment en compte :

- le recyclage (e.g. procédés de traitement/réutilisation de plastique usagé de type Carbios²¹) ;
- la valorisation de coproduits industriels. La société Lanzatech²² utilise par exemple un procédé pour convertir le gaz issu d'aciéries ou d'industries chimiques en butadiène, utilisé pour la production de caoutchouc synthétique et de nylon;
- l'utilisation de la biomasse végétale pour produire des produits traditionnellement obtenus par pétrochimie. Genomatica²³ propose par exemple un procédé fermentaire pour la conversion de la biomasse en 2,3-butanediol, dont les dérivés incluent des applications dans les industries automobile, textile et chimique. Par une voie similaire, Global Bioenergies²⁴ développe un procédé

de bioproduction d'isobutène, produit chimique de plateforme utilisé notamment pour la synthèse de plastiques et caoutchoucs.

Alcène simple à trois carbones produit majoritairement par voie pétrochimique, le propylène (ou propène) pourrait également bénéficier d'une production biosourcée. Deuxième synthon chimique produit dans le monde (100 Mt en 2016), avec une demande mondiale toujours soutenue (4%/an)²⁵, le propylène constitue une cible d'intérêt pour la chimie verte. Cette molécule est en effet essentielle pour la synthèse de nombreux dérivés valorisables, allant du polypropylène, matière plastique qui constitue plus de 60% de la demande en propylène, à l'acrylonitrile nécessaire à la fabrication des fibres acryliques dédiées à l'industrie textile, en passant par l'oxyde de propylène, utilisé pour la synthèse de mousses et d'élastomères.

La particularité du propylène est qu'il ne bénéficie pas d'une ligne de production dédiée dans les procédés traditionnels de la pétrochimie. En effet, c'est un sous-produit de deux procédés de raffinage : le vapocraquage destiné à la production d'éthylène ou le craquage catalytique en lit fluidisé (FCC) permettant la production d'essence²⁵. Les quantités produites varient en fonction de la charge employée (éthylène, éthane, naphta ou gazole). Mais, depuis 2009, un changement de charge s'opère dans les vapocraqueurs, l'éthane extrait des gaz de schistes substituant peu à peu le naphta. Cette modification a réduit la production de deux coproduits du raffinage du naphta, en l'occurrence le butadiène (facteur 6) et le propylène (facteur 20). Ceci limite les capacités de production mondiale et, dans un contexte de forte demande, conduit au développement d'unités dédiées à la production de ces oléfines²⁶. Diverses voies basées sur des ressources fossiles sont ainsi envisagées, et chacune d'entre elles peut être adaptée pour utiliser de la biomasse²⁷. Parmi les options déjà explorées figurent la déshydrogénation du propane, la voie « *Coal-To-Olefins* » (CTO) et la métathèse éthylène-butène. Mais incorporer le traitement de la biomasse dans ces procédés multiplie les étapes et peut sensiblement augmenter le coût final de production, ce qui pose la question de la compétitivité du produit.

Dans une adaptation du procédé de métathèse éthylène-butène, Braskem propose par exemple d'obtenir l'éthylène par une première étape de déshydratation de bioéthanol obtenu par fermentation, de dimériser l'éthylène ainsi produit pour former du butylène, puis de faire réagir butylène et éthylène par métathèse croisée pour obtenir le propylène²⁸. La faisabilité de ce type d'approche dépend de multiples

paramètres (coût de la matière première, prétraitement, rendements et sélectivités des étapes de conversion etc.) et la complexité de tels procédés suggère qu'une alternative plus directe serait préférable.

L'isopropanol (également alcool isopropylique, 2-propanol), alcool secondaire à trois carbones, est traditionnellement obtenu par hydratation du propylène²⁵. La réaction opposée est toutefois également possible par catalyse hétérogène, avec un taux de conversion et une sélectivité élevés pour le propylène, ainsi qu'un coût de production faible (basse température, purification simple car produits et catalyseur ne se mélangent pas)^{29,30}. Pour le biopropylène, l'intérêt de la technique est toutefois tributaire de la faisabilité d'un procédé direct de production d'isopropanol biosourcé.

Plusieurs souches bactériennes du genre *Clostridium* produisent naturellement par voie fermentaire de l'acétone, de l'isopropanol, du n-butanol* et de l'éthanol³¹. La fermentation de ces souches, appelées A/IBE en fonction des solvants qu'elles produisent, a constitué dans la première moitié du XX^{ème} siècle le deuxième procédé biologique industriel en termes de taille et de volume de produits, uniquement surpassé par la fermentation éthanologique^{32,33}. Le potentiel des souches solvantogènes de *Clostridium* pourrait ainsi être exploité pour produire, à partir des sucres issus de la biomasse végétale, de l'isopropanol, qui servirait alors de charge au procédé de catalyse hétérogène permettant sa conversion en biopropylène.

* Par soucis de simplicité, les occurrences du mot « butanol » dans la suite désignent toutes du n-butanol

1.1.2 Présentation des *Clostridia* solvantogènes

Les *Clostridia* appartiennent au groupe des Firmicutes et forment un genre assez riche dans le règne bactérien (plus de 200 espèces différentes selon la taxonomie proposée par la LPSN³⁴). Au niveau morphologique, ces bactéries sont proches de celles du genre *Bacillus* et se caractérisent par une forme de bâtonnet arrondi à l'état végétatif. La différence majeure entre les bactéries du genre *Clostridium* et celles du genre *Bacillus* se situent au niveau de leur métabolisme (aérobie pour les *Bacilli*, anaérobie pour les *Clostridia*).

Parmi les bactéries, la classification « *Clostridium* » repose sur quatre critères³⁵:

- un métabolisme anaérobie strict, susmentionné ;
- la capacité à former des endospores, formes de résistance ;
- l'incapacité à réduire les sulfates ;
- la présence d'une paroi cellulaire de type Gram-positif.

Les *Clostridia* sont un groupe d'importance pour la communauté scientifique pour deux raisons. La première est qu'un certain nombre de maladies graves (par exemple le tétanos ou le botulisme) sont dues à des infections de membres pathogènes de cette famille³⁶. La deuxième est la possibilité d'utiliser des souches dites « solvantogènes » en biotechnologie. Ces *Clostridia*, non pathogènes, ont naturellement la capacité de transformer une variété importante de sucres pour produire des espèces chimiques d'intérêt, et notamment de l'acétone, du butanol, et de l'éthanol³². On parle alors de fermentation ABE, et de fermentation IBE pour certaines souches capables de réduire en proportion variable l'acétone en isopropanol^{31,37} à l'aide d'une alcool déshydrogénase secondaire (s-ADH) dépendante du NADPH^{38,39}.

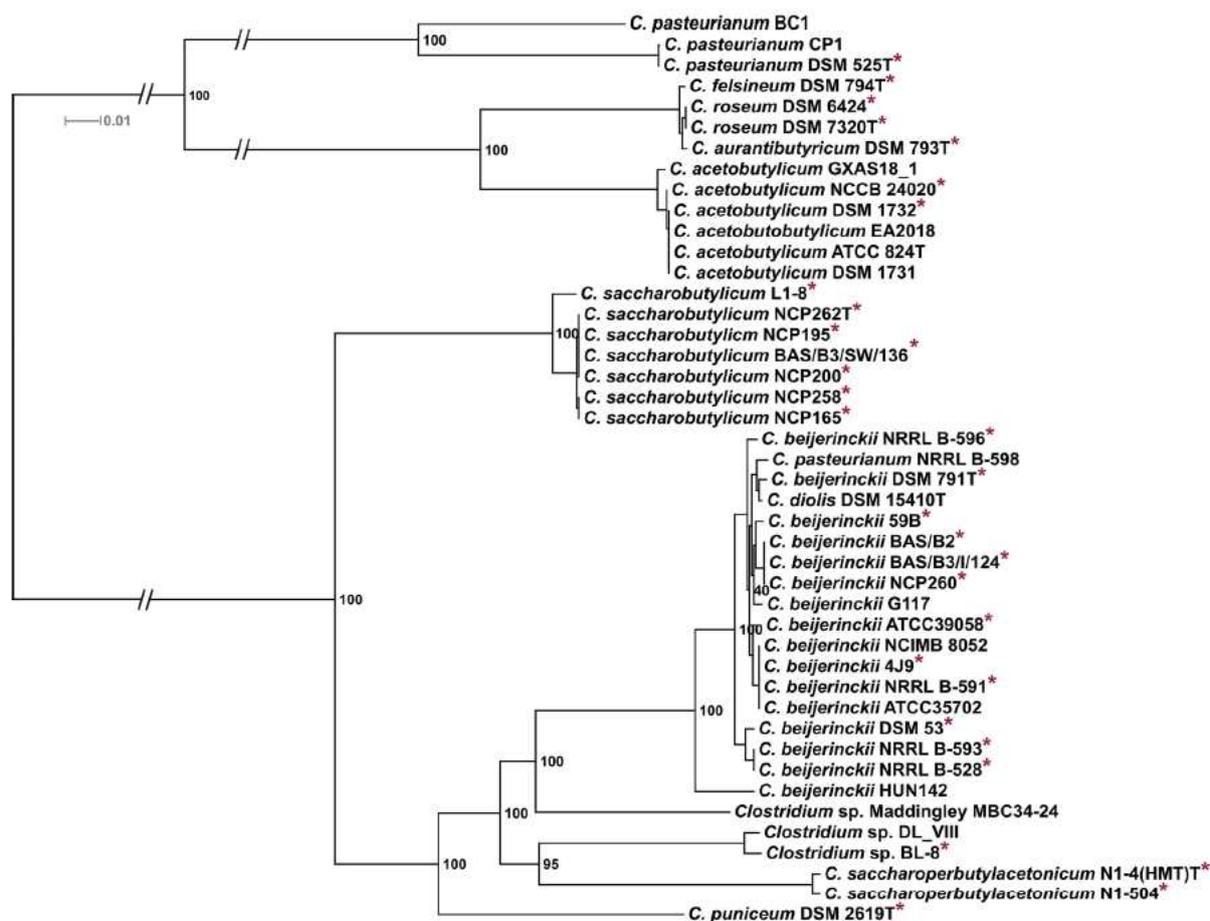


Figure 1: Taxonomie des *Clostridia* solvantogènes A/IBE d'après Poehlein et al.⁴⁰

Les espèces de *Clostridia* solvantogènes présentent des similarités phénotypiques importantes, ce qui rendait difficile leur classification avant l'avènement des techniques de séquençage modernes⁴¹. Avec l'aide de ces nouveaux outils, et notamment de la comparaison de séquences d'ARNr 16S^{42,43}, la taxonomie a évolué pour comprendre quatre espèces solvantogènes majeures: *C. acetobutylicum*, *C. saccharoperbutylacetonicum*, *C. saccharobutylicum* et *C. beijerinckii*. Plus récemment, l'analyse comparative des génomes complets de 30 souches solvantogènes, réparties en 10 espèces différentes, a permis de séparer les *Clostridia* solvantogènes en deux clades principaux (Figure 1)⁴⁰. Ces deux groupes séparent notamment les espèces modèles que sont *C. acetobutylicum* et *C. beijerinckii*, qui *in fine* forment le référentiel de cette taxonomie. Ces deux espèces sont d'ailleurs les plus étudiées dans la littérature, avec comme références les souches ABE *C. acetobutylicum* ATCC 824 et *C. beijerinckii* NCIMB 8052. Les travaux réalisés sur ces souches portent majoritairement sur l'amélioration de leur tolérance au butanol et de leur capacité de

production de ce composé chimique, qui présente de multiples avantages vis-à-vis de l'éthanol dans le cadre d'une utilisation en tant que biocarburant⁴⁴.

Les souches de *Clostridium* naturellement capables d'effectuer une fermentation IBE sont peu nombreuses et moins performantes que les souches ABE. De ce fait, les souches IBE restent encore peu décrites. A la lumière des données bibliographiques, la souche *C. beijerinckii* DSM 6423 (ou NRRL B-593) peut être considérée comme la référence pour la fermentation IBE^{31,37,38,45-51}. Son génome, assez proche de la souche NCIMB 8052⁴⁰, a récemment été complètement séquencé dans une étude de Máté de Gerando et collègues⁵². La description du génome de cette bactérie, combinée à une première étude transcriptomique également réalisée dans cette étude, a permis d'approfondir la compréhension des mécanismes métaboliques qui régissent la production d'isopropanol dans cet organisme. Ceci constitue un premier pas vers l'amélioration de ses performances fermentaires.

1.1.3 La fermentation A/IBE, du XIX^{ème} siècle à aujourd'hui

Pasteur identifiait en 1861 et 1862 pour la première fois le mécanisme fermentaire d'un organisme qu'il nommait alors « vibron butyrique », dont les produits principaux étaient l'acide butyrique et le butanol^{53,54}. Rapidement, Albert Fitz⁵⁵ et Martinus Beijerinck⁵⁶ confirmèrent ces observations en décrivant des organismes disposant de propriétés similaires : *Bacillus butylicus*, *Granulobacter butylicus* et *Granulobacter saccharobutyricus*.

Mais ce n'est qu'au début du XX^{ème} siècle que l'idée d'industrialiser un procédé de fermentation basé sur l'utilisation de ces souches bactériennes germa, poussée par une explosion de la demande en caoutchoucs synthétiques, productibles à partir de butadiène. La synthèse de butadiène était possible à partir de 3-méthylbutan-1-ol, et l'entreprise Strange & Graham Ltd. se lança dans l'isolement d'une souche bactérienne productrice de ce composé⁵⁷. Aucune souche ne put être isolée, mais au cours de la démarche, Chaïm Weizmann, membre de l'équipe de recherche de l'entreprise, breveta l'utilisation d'une souche bactérienne pour la fermentation ABE (Acétone-Butanol-Ethanol)⁵⁸. Cette souche, initialement appelée BY et identifiée comme appartenant à l'espèce *Clostridium acetobutylicum* quelques années plus tard⁵⁹, présentait notamment des performances fermentaires exceptionnelles sur

amidon⁶⁰, une forme de stockage des sucres végétaux, essentiellement produit dans l'industrie céréalière. Le développement d'un procédé industriel mettant en œuvre la fermentation ABE fut très rapide car lié aux besoins en acétone pour la production de cordite, un explosif largement utilisé au cours de la Première Guerre Mondiale. Par la suite, c'est son intérêt pour la production de butanol qui permit de pérenniser les unités industrielles⁵⁷. La demande venait notamment de l'industrie automobile qui, en pleine expansion, était consommatrice d'acétate de butyle. Ce dérivé du butanol est un solvant idéal pour la production de vernis nitrocellulosiques à séchage rapide, utilisés en finition sur la carrosserie⁶¹. En 1927, la production de solvants par fermentation ABE atteignait ainsi plus de cent tonnes par jour aux Etats-Unis⁶².

Au début des années 30, le développement de la synthèse d'acétone par voie pétrochimique soulevait déjà des inquiétudes pour l'industrie ABE³². En réponse, de nouvelles souches furent isolées, afin, dans un premier temps, d'améliorer la fermentation de l'amidon, et par la suite d'utiliser d'autres sources de carbone comme matière première. A cette époque, la mélasse, un sous-produit de l'industrie sucrière, était un substrat de choix, riche en sucres simples, et dont le prix par rapport à l'amidon était très avantageux^{32,62,63}. Avec les souches adaptées, la mélasse devint rapidement la matière première la plus utilisée dans le monde. Après la Seconde Guerre Mondiale, elle commença toutefois à être utilisée comme supplément pour l'alimentation animale, et vit de fait son prix augmenter. Parallèlement à cela, le développement de l'industrie pétrochimique permit d'obtenir à moindre coût du butanol, et le procédé ABE périclita. Entre les années 1950 et 1960, la quasi-totalité des usines implantées aux Etats-Unis et en Grande-Bretagne ferma⁶⁴.

Le contexte actuel (cf. sous-partie 1.1.1) propice aux énergies alternatives, explique le regain d'intérêt de la communauté scientifique pour la fermentation ABE, qui vise en particulier la production de biobutanol⁶⁵. Cette espèce chimique pourrait aujourd'hui être utilisée comme biocarburant avancé car elle dispose d'avantages certains vis-à-vis de l'éthanol, comme par exemple son pouvoir calorifique inférieur plus élevé, sa corrosivité moindre pour les moteurs et sa faible volatilité⁶⁶. La plupart des efforts de recherche visent donc à améliorer la production de butanol selon deux axes principaux :

- L'obtention de souches génétiquement modifiées permettant par exemple de fermenter une variété plus importante de sources de carbone (comme le

mélange hexose/pentose extrait de la biomasse lignocellulosique, substrat 2G⁶⁷, pour plus de détails sur les substrats voir sous-partie 1.2.2) ou pour favoriser la production de butanol vis-à-vis de l'acétone⁶⁸⁻⁷¹ ;

- L'amélioration des procédés fermentaires traditionnellement associés à *Clostridium*, par exemple pour éviter l'accumulation des produits fermentaires toxiques dans le milieu de culture. Plus particulièrement, la toxicité du butanol constitue une barrière importante dans le développement d'un procédé compétitif pour sa production. Différentes approches, comme la fermentation continue^{49,72} ou l'extraction en ligne des alcools par strippage au gaz (*gas stripping*^{73,74}), permettent d'atténuer ce problème.

Ces approches rationnelles sont également applicables à la fermentation IBE. Comme mentionné précédemment, peu de *Clostridia* sont naturellement capables de produire de l'isopropanol et les données disponibles à leur sujet sont limitées. De fait, les travaux sur la fermentation IBE suivent deux tendances principales : l'étude de la fermentation IBE naturelle, notamment dans la souche de référence *C. beijerinckii* DSM 6423, et l'ingénierie génétique de souches ABE modèles (e.g. *C. acetobutylicum* ATCC 824) pour les transformer en organismes producteurs d'IBE. Cette deuxième tendance a été favorisée⁷⁵⁻⁷⁸, en partie parce que les souches naturelles IBE ne pouvaient pas, antérieurement aux résultats obtenus dans cette thèse, être modifiées génétiquement de manière dirigée. Les seules méthodes disponibles pour la modification génétique de la souche DSM 6423, ou de toute autre souche IBE, consistaient en l'altération aléatoire de son génome à l'aide de composés mutagènes⁴⁷. Des travaux du département Biotechnologie d'IFP Energies nouvelles, ont toutefois montré, peu avant le démarrage de la thèse, que la souche *C. beijerinckii* DSM 6423 pouvait être transformée, ce qui a constitué une étape importante pour envisager son amélioration génétique afin d'améliorer ses performances fermentaires, qui à l'heure actuelle sont plus faibles que celles des meilleures souches ABE.

Outre la détermination de la séquence génomique de la souche *C. beijerinckii* DSM 6423⁵², une thèse réalisée précédemment à IFPEN a permis de caractériser son métabolisme sur un milieu modèle comportant du glucose et très peu d'impuretés, notamment via des analyses transcriptomiques⁵². Si une telle étude est absolument nécessaire pour établir un premier modèle de fonctionnement de la souche, son comportement sur milieu industriel restait encore à élucider. Pour être industrialisable

la souche doit notamment être capable de métaboliser efficacement des substrats peu coûteux et abondants. Les espèces A/IBE du genre *Clostridium* sont connues pour leur capacité à utiliser une variété importante de sources carbonées, et la sélection des plus appropriées est une étape clé dans le processus de mise au point d'un procédé fermentaire basé sur l'utilisation de la souche DSM 6423.

1.2 PRODUCTION D'ISOPROPANOL PAR FERMENTATION IBE

1.2.1 Caractéristiques fermentaires A/IBE

1.2.1.1 Bilans fermentaires

Les performances fermentaires de cultures A/IBE employant des souches de *Clostridium* sont évaluées selon quatre critères principaux, liés à la production de solvants :

- 1) le rendement en solvants vis-à-vis de la quantité de carbone consommée par la bactérie. Il est exprimé généralement en grammes de solvants par gramme de source carbonée. Le rendement de conversion d'une fermentation A/IBE type est généralement situé autour de 35%, proche du rendement théorique maximal⁷⁹.
- 2) le titre de solvants, c'est-à-dire la concentration à un instant t de l'ensemble des solvants dans le milieu de culture. Le titre maximal (c'est-à-dire final) est généralement utilisé comme outil de comparaison, et change largement en fonction des souches et des conditions expérimentales (typiquement entre 5 et 30 g/L).
- 3) la productivité (en grammes de solvants produits par litre et par heure) qui est essentiellement utilisée pour comparer des fermentations réalisées selon un mode continu (cf. sous-partie suivante). Dans ce type de conduite fermentaire, la notion de temps est importante et les productivités classiquement rencontrées sont comprises entre 0,2 et 0,7 g/L/h pour une fermentation continue employant des souches non améliorées.
- 4) le ratio entre solvants (A/B/E ou I/B/E) qui correspond à la composition des solvants produits dans le milieu de culture. Même s'ils peuvent varier en fonction des souches utilisées et des conditions de culture, les ratios butanol:acétone/isopropanol:éthanol obtenus expérimentalement sont généralement de 6:3:1. Ce paramètre est particulièrement étudié dans les

stratégies génétiques qui visent à favoriser la sélectivité de la fermentation pour un de ces solvants (dans la plupart des cas le butanol).

1.2.1.2 Modes de culture

La fermentation *batch*, grâce à sa facilité d'utilisation, est le mode de fermentation le plus utilisé dans l'industrie. Le procédé consiste à simplement laisser croître un microorganisme dans une enceinte contrôlée, sans apport de substrat au cours de la fermentation. Le procédé *fed-batch* est similaire, avec toutefois un apport d'une source concentrée de substrat pendant la fermentation.

La fermentation continue (Figure 2) est plus compliquée à mettre en place à l'échelle industrielle. Du milieu de culture contenant le substrat carboné est apporté en permanence par une pompe dans le réacteur de fermentation. Cet apport est couplé à un soutirage d'une même quantité de milieu fermenté. Les principaux paramètres de ce type de conduite sont le débit d'apport et de soutirage du milieu de culture, et le volume du fermenteur. Le rapport entre ces deux quantités permet d'obtenir le taux de dilution, exprimé en h^{-1} .

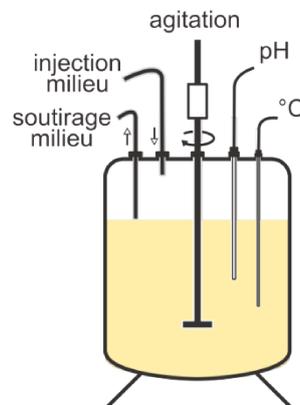


Figure 2 : Représentation schématique d'un bioréacteur en mode continu, tel qu'utilisé dans cette étude.

Ce type de fermentation, contrairement aux deux précédentes, peut être opéré pendant des temps très longs, ce qui permet au niveau industriel d'éviter des procédures répétées de nettoyage / stérilisation / ensemencement du bioréacteur et de réduire la toxicité des produits de fermentation via l'apport continu de milieu non fermenté. Par ailleurs, le procédé peut être couplé à une immobilisation des cellules

sur une phase solide, à l'intérieur du bioréacteur. La phase solide, sur laquelle les bactéries sont fixées pour former un biofilm, reste dans le bioréacteur et permet d'augmenter le temps de résidence des cellules dans le milieu. Ceci se traduit par des améliorations critiques de la productivité. Survase *et al.* multiplie par exemple par plus de dix (de 0,47 à 5,52 g/L/h) la productivité d'une fermentation IBE avec la souche *C. beijerinckii* DSM 6423 en immobilisant les cellules sur de la pulpe de bois⁴⁹. Toutefois, outre les difficultés d'extrapolation industrielle de ce type de procédé, la fermentation continue présente des risques nettement accrus de contaminations par d'autres microorganismes ou par des bactériophages.

Si les performances et les critères importants des trois modes de culture possibles pour la fermentation ABE ont été largement décrits dans la littérature, la pertinence de favoriser l'un d'entre eux en particulier pour une utilisation industrielle fait encore débat. Dans une étude comparative des trois modes de fermentation chez *C. acetobutylicum* ATCC 824 avec contrôle du pH, Li *et al.* soulignent par exemple l'amélioration des performances de la souche *C. acetobutylicum* ATCC 824 en fermentation continue⁷².

Par ailleurs, sur un aspect plus fondamental, la fermentation continue est un mode de conduite de choix pour étudier les différents états physiologiques rencontrés au cours d'une fermentation A/IBE (cf. paragraphe ci-dessous). Vasconcelos *et al.*, en fixant un taux de dilution assez faible, en jouant sur le milieu et en contrôlant le pH, utilisent la fermentation continue pour fixer le métabolisme bactérien dans un état donné (notamment acidogénèse, solvantogénèse, cf. paragraphe ci-dessous), ce qui permet d'étudier chacun de ces états avec une population homogène et de façon reproductible⁸⁰.

1.2.1.3 Déroulement type d'une fermentation batch A/IBE

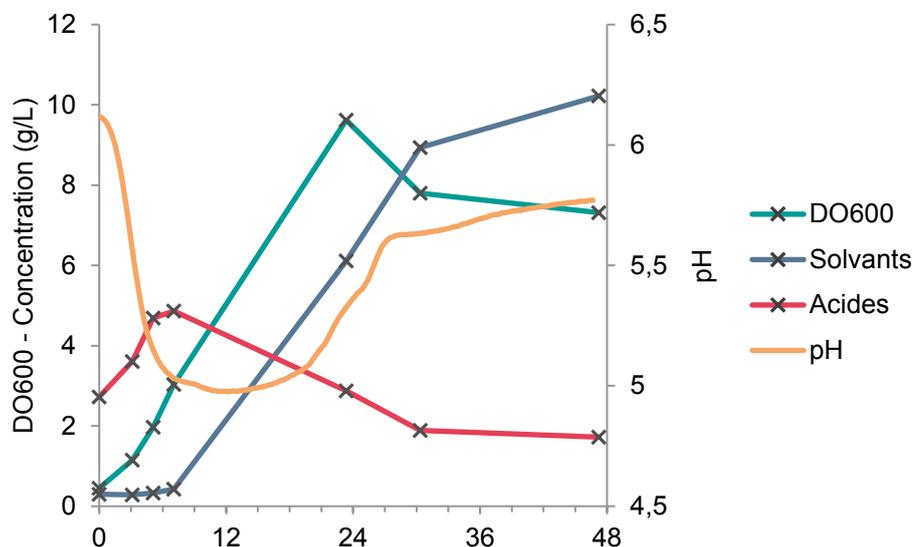


Figure 3 : Evolution typique d'une fermentation en bioréacteur en mode *batch* pour la souche DSM 6423. Fermentation réalisée en milieu dit « de Gapes »⁸¹ avec 70 g/L de glucose et 2,9 g/L d'acétate d'ammonium. Sont représentés la biomasse (en valeur de densité optique à 600 nm), les solvants (éthanol, acétone, isopropanol, butanol), les acides (acétique, butyrique), et le pH.

De par leur métabolisme particulier (cf. sous-partie 1.3), les *Clostridia* A/IBE fermentent les sucres en anaérobiose en suivant un comportement biphasique. Dans une fermentation *batch* typique (Figure 3) la première phase, qualifiée d'acidogène, est caractérisée par une forte production d'acides organiques (acétate, butyrate), ce qui se traduit par une baisse graduelle du pH dans le bioréacteur. Le pH diminue jusqu'à un niveau critique pour la physiologie du microorganisme, qui reconsume les acides et produit le mélange A/IBE dans une deuxième phase : la solvantogénèse. Le seuil de pH pour lequel cette modification s'opère varie légèrement en fonction des conditions expérimentales et de la souche concernée, mais se situe généralement aux alentours de 5. Ce changement, appelé revirement métabolique (*metabolic switch*), provoque une remontée du pH, typiquement jusqu'à une autre valeur seuil (proche de 5,5), à laquelle une nouvelle phase acidogène s'ensuivra. Les phases alternent ainsi jusqu'à la fin de la fermentation. Une fermentation *batch* typique s'arrête entre 24 et 72 heures après l'ensemencement, après une simple ou double alternance de phases acidogènes et solvantogènes.

Le pH n'est toutefois pas le seul facteur qui permet de suivre en ligne l'évolution d'une fermentation. Le volume et le débit de gaz produit par la fermentation, ainsi que sa composition en dihydrogène et en dioxyde de carbone, sont extrêmement informatifs pour suivre le déroulement du procédé. En phase acidogène, la culture se caractérise par un fort débit de gaz, qui diminue progressivement, et par une forte proportion de dihydrogène dans la mixture gazeuse. La transition s'effectue à débit faible, puis la production reprend avec celle des solvants, jusqu'à atteindre un débit plus important encore que la production initiale. Les gaz comportent alors une proportion majoritaire de dioxyde de carbone (environ 60-70%). Le débit diminue alors progressivement jusqu'à la fin de la fermentation.

1.2.1.4 Limitations principales des fermentations A/BE

La faculté des souches solvantogènes de *Clostridium* à produire des solvants n'est pas un caractère stable : une dégénérescence réversible ou irréversible de cette capacité est en effet possible³². Si ces mécanismes de dégénérescence ne sont pas encore parfaitement compris, il semblerait toutefois que la propagation sérielle des cellules en phase végétative provoque, directement ou indirectement, la perte de la capacité à produire des solvants. De fait, il est devenu habituel de conserver les souches solvantogènes sous forme de spores, et de ne démarrer des cultures qu'après activation de ces dernières par choc thermique (et élimination concomitante des cellules végétatives). Le passage par cet état physiologique particulier permettrait de conserver la capacité de production de solvants pour des fermentations en mode *batch*. Toutefois, même en utilisant des spores, la dégénérescence peut également être un problème dans les procédés de culture continue, qui peuvent s'étaler sur de longues périodes.

Le principal facteur limitant reste la toxicité des produits, en particulier du butanol. Des concentrations relativement faibles de butanol comprises entre 7 et 13 g/L inhibent la croissance de moitié chez *C. acetobutylicum* ATCC 824⁸². Au-delà, la croissance bactérienne est stoppée. Cette toxicité limite ainsi les capacités fermentaires de ces souches. A titre d'exemple, la souche génétiquement modifiée de *C. beijerinckii* BA101 détient un titre record en butanol : 21 g/L, ce qui reste peu élevé, notamment au regard des fermentations éthanoliques pouvant dépasser les 100 g/L⁸³.

D'un point de vue procédés, des concentrations aussi faibles impliquent des coûts de purification du mélange A/IBE très importants⁶².

La limite physiologique de toxicité n'est jamais atteinte pour les autres solvants, le butanol étant le principal produit de ces fermentations et le plus toxique. En effet, l'acétone et l'éthanol inhibent la croissance de *C. acetobutylicum* ATCC 824 de 50% aux environs de 40 g/L, et l'inhibition devient complète vers 70 g/L⁸². L'isopropanol semble avoir une toxicité légèrement plus forte, avec une inhibition de croissance observée sur boîte pour *C. beijerinckii* DSM 6423 à une concentration de 35 g/L⁴⁷. Ces observations suggèrent qu'une amélioration du titre en isopropanol est possible si la résistance au butanol est améliorée, ou si la production de butanol est limitée, voire éliminée. Un mélange IBE plus concentré en isopropanol en sortie de fermenteur permettrait ainsi de diminuer les coûts de séparation, tout en améliorant le rendement en isopropanol.

D'autres optimisations peuvent considérablement améliorer ces performances, notamment via le soutirage des alcools. Un procédé continu permet l'utilisation de différentes méthodes d'extraction des solvants⁸⁴ (liquide-liquide, perstraction, distillation sur membrane, *gas stripping* notamment), ce qui favorise leur production en levant partiellement leurs effets inhibiteurs.

1.2.2 Fermentation de substrats industriels : état de l'art chez Clostridium

1.2.2.1 Généralités

Les substrats carbonés industriels utilisés pour des procédés fermentaires peuvent être classés en deux familles principales, ou « générations ». Le classement des substrats par catégories de générations différentes provient initialement de l'industrie des biocarburants, dans laquelle cette distinction est directement reliée aux impacts écologiques et humains de l'utilisation de ces substrats.

Comme décrit précédemment, l'utilisation de ressources 1G a soulevé des critiques, notamment parce qu'elle détourne l'utilisation de terres arables et de ressources comestibles pour l'industrie de l'énergie, avec comme conséquence une augmentation des tensions sur les prix de la nourriture. Les substrats 1G généralement utilisés pour produire du bioéthanol sont classiquement issus des industries sucrières et amidonnières. Ils sont par conséquent majoritairement produits à partir de canne à

sucré, de betterave, de maïs et de blé. Leur fermentation par des levures industrielles (du genre *Saccharomyces*) est aujourd'hui très performante et ne nécessite pas de prétraitement physico-chimique.

Les substrats de deuxième génération (2G) sont issus de plantes non consommables, ou des éléments de structure des plantes destinées à l'alimentation, ce qui permet de limiter les compétitions d'usages qui concernaient les ressources 1G. Leur utilisation n'est toutefois pas encore généralisée car des recherches sont toujours en cours pour améliorer leurs performances. L'optimisation de l'étape de prétraitement physico-chimique, permettant la libération du polymère de glucose (cellulose), est ainsi cruciale pour limiter son coût tout en permettant l'obtention de substrats facilement fermentescibles. En effet, en fonction de sa sévérité, différents inhibiteurs notamment dérivés des sucres peuvent être générés et impacter les performances fermentaires des microorganismes.

Un des principaux avantages des espèces solvantogènes du genre *Clostridium* pour la fermentation industrielle réside dans leur capacité à métaboliser des substrats industriels variés. Sous diverses formes, la canne à sucre^{85,86}, le maïs⁸⁷⁻⁸⁹, le manioc^{88,90}, le sorgho⁸⁸, le topinambour⁹¹, mais aussi des déchets agricoles^{92,93}, des coproduits de l'industrie fromagère^{94,95}, du glycérol de l'industrie des biodiesels^{50,96} ainsi qu'un certain nombre de substrats ligno-cellulosiques de seconde génération^{74,90,97-105} ont tous été testés avec succès.

1.2.2.2 Substrats 1G

De par leur large disponibilité, leur faible coût et leur haute fermentescibilité, amidons et mélasses (coproduits de la fabrication du sucre de canne et du sucre de betterave) ont successivement servi de matière première pour la fermentation ABE, avant que cette dernière ne disparaisse. L'évolution d'un substrat à l'autre, bien que dictée par des considérations économiques fortes, a été progressive, en particulier en raison de l'utilisation qui peut être faite de l'amidon dans de nombreux secteurs dont celui de l'agroalimentaire¹⁰⁶. A l'origine, les mélasses étaient bien moins onéreuses, ce qui a incité à isoler de nouvelles souches, capables de produire un mélange ABE à partir d'une composition carbonée très différente de celle de l'amidon³². En raison de l'augmentation progressive de leur utilisation pour l'alimentation animale, leur coût a

toutefois augmenté au sortir de la Seconde Guerre Mondiale, ce qui, combiné à l'essor de la pétrochimie, a précipité la fermeture des usines ABE dans les années 50-60³². Dans le contexte de la transition écologique, l'utilisation de ces deux types de substrats de première génération pour la fermentation A/IBE pourrait être remise au goût du jour.

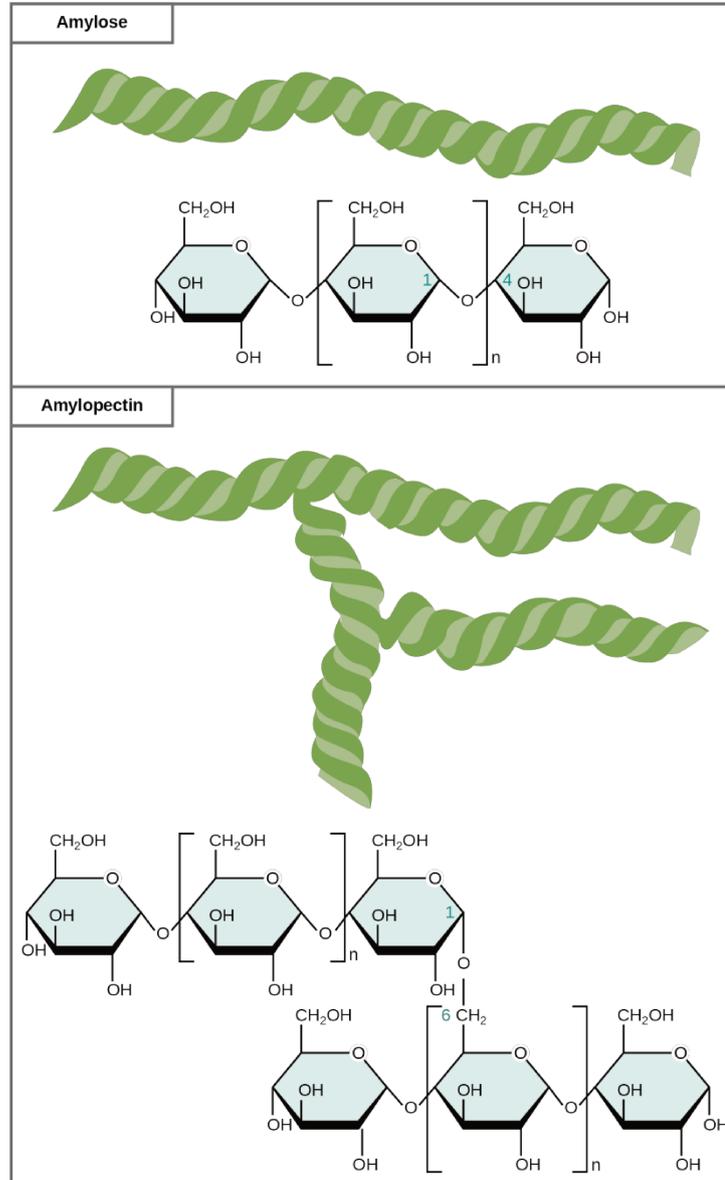


Figure 4: Structure des deux composés glycosidiques de l'amidon : l'amylose et l'amylopectine¹⁰⁹.

L'amidon est une forme de stockage du carbone chez les plantes, constituée de deux molécules insolubles principales : l'amylose et l'amylopectine (Figure 4). L'amylose est un polymère linéaire contenant des unités d'anhydroglucose liées en α -1,4. L'amylopectine est formée par l'agrégation de plusieurs molécules d'amylose par le biais de liaisons α -1,6 entre des résidus glucose. De ce fait, l'amylopectine présente

un poids moléculaire nettement plus élevé que l'amylose (facteur d'environ 10^3), avec de légères variations de taille en fonction du végétal utilisé. Ce dernier joue également sur le ratio entre les deux formes de stockage, le maïs cireux ne contenant par exemple que de l'amylopectine^{107,108}. La structure hautement branchée de l'amidon nécessite l'action concomitante de plusieurs types d'enzymes pour libérer l'intégralité des monomères de glucose et ainsi permettre la fermentation.

Les *Clostridia* solvantogènes produisent un tel cocktail d'enzymes saccharolytiques, constitué d' α - et β -amylases, d' α -glucosidases, de gluco-amylases et de pullulanases^{110,111}, et qui, pour certaines espèces, permettent une conversion très efficace de substrats amidonnés⁸⁸. L'inconvénient principal de l'amidon est sa difficulté d'utilisation dans le cadre d'une fermentation continue : l'amidon, insoluble, peut être difficile à pomper, notamment s'il a été gélatinisé au préalable (c'est-à-dire chauffé dans de l'eau, comme par exemple au cours d'une stérilisation par autoclavage). Cet obstacle peut toutefois être contourné en ajoutant une étape préliminaire de liquéfaction enzymatique, permettant de produire un sirop de glucose compatible avec un système de fermentation continue à l'échelle laboratoire⁸⁷.

Les amidons commerciaux ne contiennent que très peu d'impuretés potentiellement inhibitrices (moins de 1% du poids sec, sous forme de protéines, lipides et ions métalliques), ce qui est un avantage conséquent pour la fermentation bactérienne. Ce n'est pas le cas des mélasses, dont la composition complexe est susceptible d'inhiber les performances fermentaires¹¹². Sous-produit du raffinage du saccharose la mélasse contient, outre une quantité non négligeable de sucres, une proportion importante d'autres substances. Parmi celles-ci figurent des composés organiques azotés (ammoniaque, nitrates, amides, bêtaïnes, acides aminés, protéines), des acides organiques (acide lactique issu de la dégradation des sucres et panel d'acides présents en plus faible quantités) et des sels minéraux (Tableau 1). Ces sels peuvent représenter plus de 10% de la matière sèche des mélasses, ce qui, même dilué, peut potentiellement inhiber la croissance cellulaire¹¹³. Par exemple, les performances fermentaires de la souche DSM 13864 de *C. saccharobutylicum* sont fortement affectées par l'incorporation de mélasse de canne pour des concentrations en sucres supérieures à 6%⁸⁶. Lorsque la fermentation ABE était menée à l'échelle industrielle, les mélasses utilisées étaient diluées à 6,5% en sucres au maximum, pour des raisons similaires^{32,112}. Par ailleurs, les mélasses présentent une flore microbienne

importante, qui peut également affecter la fermentation en l'absence de stérilisation préalable. L'étude des mélasses pour la fermentation des levures a toutefois montré que l'autoclavage en lui-même pouvait aussi affecter négativement les propriétés fermentaires du produit¹¹². Enfin, les mélasses ne sont pas des substances homogènes et contiennent un taux élevé de particules, pouvant dépasser les 10^6 par millilitre. Pour la fermentation des levures, ces particules ont un effet toxique sur la croissance et le procédé de fermentation sous-jacent requiert une étape préliminaire de filtration de la mélasse, ce qui met en exergue la question du prétraitement. Pour une fermentation ABE chez *C. saccharobutylicum*, Ni et collègues ont préalablement traité en conditions acides leur mélasse de canne, avant de la centrifuger pour éliminer les particules en suspension⁸⁶. Il ne semble pas toutefois qu'un tel prétraitement fut réalisé dans la fermentation ABE historique¹¹². Malgré tous ces inconvénients, certains constituants des mélasses peuvent avoir un impact positif sur la croissance cellulaire, et notamment :

- les composés azotés (particulièrement les acides aminés) ;
- les vitamines, présentes à l'état de trace (thiamine, riboflavine, pyridoxine, niacine, acide pantothénique, acide folique, biotine) ;
- les sels (phosphate, calcium, potassium, magnésium ainsi que des ions métalliques présents à l'état de trace, comme le zinc ou le cuivre)¹¹².

La nature de la mélasse (c'est-à-dire canne à sucre ou betterave) et le procédé d'extraction utilisé influent grandement sur leur composition, tant au niveau des impuretés que de la nature et de la quantité des sucres présents. La mélasse de betterave contient essentiellement du saccharose, ainsi qu'une faible proportion de glucose, de fructose et de raffinose (trimère galactose-glucose-fructose). En revanche, la mélasse de canne contient plus de sucres simples, plus facilement métabolisables, avec un ratio massique saccharose-glucose-fructose proche de 2-1-1¹¹² (valeur moyenne, susceptible à d'importantes variations). Cette répartition hétérogène des sucres peut toutefois impliquer des problèmes pour leur consommation par les microorganismes. Ces derniers ont en effet généralement un ordre de préférence pour la consommation des sucres : les sucres primaires (comme par exemple le glucose) sont favorisés par rapport aux sucres secondaires. La régulation génétique de ce mécanisme, appelé répression catabolique du carbone (RCC), a été étudiée en détails dans divers organismes modèles (pour une publication assez complète sur le sujet,

voir Görke et Stülke, 2008¹¹⁴). Chez *C. beijerinckii*, la RCC ne semble pas intervenir directement pour l'utilisation concomitante d'un mélange glucose/saccharose puisque l'expression des gènes liés à la consommation du saccharose n'est pas affectée¹¹⁵. Malgré cela, la vitesse de consommation du saccharose diminue en présence de glucose, et les mécanismes mis en jeu dans ce phénomène régulateur n'ont pas été décrits.

Composé	Mélasse de betterave (%)		Mélasse de canne (%)	
Eau	16,5		20,0	
Molécules organiques				
	Saccharose	51,0	32,0	
Sucres	Glucose	0,5	53,0	14,0
	Fructose	0,5		16,0
	Raffinose	1,0		
Autres : molécules azotées, acides libres et liés, substances gélatineuses solubles		19,0		10,0
Molécules inorganiques				
	SiO ₂	0,1	0,5	
	K ₂ O	3,9	3,5	
	CaO	0,26	1,5	
	MgO	0,16	0,1	
	P ₂ O ₅	0,06	0,2	
	Na ₂ O	1,3	≈ 11,5	≈ 8,0
	Fe ₂ O ₃	0,02	0,2	
	Al ₂ O ₃	0,07		
	Soude et résidus carbonés (ex: CO ₂)	3,5		
	Résidus soufrés (ex: SO ₃)	0,55	1,6	
	Chlorures	1,6	0,4	
		100		100

Tableau 1 : Composition moyenne des constituants principaux des mélasses de canne et de betterave (Ólbrich, 1963¹¹²).

1.2.2.3 *Substrats 2G*

Comme mentionné précédemment, et malgré leurs qualités fermentaires, amidons et mélasses sont des substrats dits de première génération et rentrent en compétition avec l'alimentation humaine et/ou animale. L'utilisation alternative de biomasses ligno-cellulosiques, ressource abondante, abordable, et ne présentant pas les mêmes limitations que les sources de première génération¹⁵, a donc été envisagée pour la production de biocarburants (éthanol par des levures ou butanol via la fermentation ABE).

La biomasse ligno-cellulosique est constituée de trois composantes principales, dont la quantité relative varie en fonction de la source : la cellulose, l'hémicellulose et la lignine¹¹⁶. Ces composés adoptent une structure complexe dans les plantes, ce qui rend leur hydrolyse directe peu efficace¹¹⁷. Dans une optique de production de bioéthanol de seconde génération, il est possible d'améliorer les performances hydrolytiques via l'utilisation de prétraitements (physique, chimique et/ou thermochimiques) pour rendre accessible les composants individuels de cette biomasse et plus particulièrement la cellulose afin d'en extraire le glucose par hydrolyse enzymatique de manière efficace. Par la suite, les monosaccharides libérés sont fermentés par des souches de *S. cerevisiae*. Pour une fermentation A/IBE, ces étapes de prétraitement et d'hydrolyse enzymatique sont également indispensables puisque les souches A/IBE sont incapables de métaboliser la cellulose et l'hémicellulose.

Les études initiales réalisées pour la fermentation éthanolique de la biomasse ligno-cellulosique ont montré que la nature du prétraitement était cruciale pour la suite du procédé¹¹⁶⁻¹¹⁸, puisqu'il peut entraîner l'apparition de composés de dégradation toxiques pour les microorganismes. Ezeji et collègues ont étudié l'effet de certains de ces composés chez *C. beijerinckii* BA 101 (dérivée de la souche NCIMB 8052)⁹². De façon surprenante, dans cette souche, furfural et 5-hydroxyméthyl furfural semblent avoir un effet positif sur la croissance cellulaire et la production de solvants. D'autres substances, également co-produites lors du prétraitement de la biomasse, affectent au contraire très largement ses capacités fermentaires (acides férulique et coumarique, notamment).

La répression catabolique du carbone est également une limitation d'importance, mais cette dernière semble propre aux souches de *C. acetobutylicum*,

qui préfèrent le glucose au xylose^{99,119}. La plupart des études fermentaires réalisées sur la biomasse ligno-cellulosique ont été réalisées avec des souches de *C. beijerinckii*, et toutes relèvent une co-utilisation du xylose^{74,90,97,100,101,103–105}. Ce constat, associé par ailleurs à la variété de substrats ligno-cellulosiques fermentés avec succès dans ces études, suggère que les souches de *C. beijerinckii* seraient adaptées pour des fermentations 2G.

1.3 METABOLISME DES CLOSTRIDIA A/IBE

Le potentiel de l'utilisation industrielle des souches solvantogènes du genre *Clostridium* pour répondre à certaines des problématiques environnementales et économiques modernes explique le regain d'intérêt des vingt dernières années à leur égard. Malgré cette tendance et les efforts de recherche fournis, les adaptations industrielles peinent encore à se développer. Les souches A/IBE présentent des limitations, notamment liées à la toxicité de leurs produits, et l'approfondissement des connaissances - actuellement loin d'être exhaustives - liées au fonctionnement biologique de ces souches et à leur métabolisme s'avère crucial pour leur amélioration. Une poignée de souches modèles ont été largement étudiées par la communauté scientifique pour caractériser le métabolisme particulier des *Clostridia* A/IBE. Ainsi, la grande majorité des données publiées concerne *C. acetobutylicum* ATCC 824¹²⁰, et dans une moindre mesure *C. beijerinckii* NCIMB 8052, plus proche de la souche DSM 6423. *C. acetobutylicum* et *C. beijerinckii* sont des espèces qui présentent une physiologie comparable, notamment du point de vue fermentaire. Cette caractéristique, associée à la bibliographie assez pauvre dévolue à *C. beijerinckii* DSM 6423, justifie l'incorporation des résultats scientifiques obtenus chez *C. acetobutylicum* dans les parties qui vont suivre, ainsi que l'hypothèse inhérente que ces derniers sont applicables à la souche DSM 6423.

Toutefois, il convient de rappeler que malgré leurs similitudes, ces deux espèces sont clairement distinctes d'un point de vue phylogénétique⁴⁰. Sur boîte de Petri, les souches de *C. acetobutylicum* et de *C. beijerinckii* présentent un phénotype parfaitement distinguable (morphologie des colonies bactériennes, vitesse de croissance à température optimale). D'un point de vue génétique, les espèces *C. beijerinckii* et *C. acetobutylicum* sont également très différentes. Les souches

appartenant au même clade que *C. acetobutylicum* disposent d'environ 4000 gènes, contre 5000-6000 pour les espèces plus proches de *C. beijerinckii*, ce qui suggère d'ailleurs une complexité génétique plus importante dans ces dernières⁴⁰. Outre la différence de taille, les souches *C. beijerinckii* NCIMB 8052 et *C. acetobutylicum* ATCC 824 ne présentent qu'une similarité très limitée des séquences d'ADN de leur chromosome respectif¹²¹. Les gènes liés à la production de solvants sont arrangés de façon très différente (un mégaplasmide contient la plupart des gènes liés à la solvantogénèse chez *C. acetobutylicum* ATCC 824, ce qui n'est pas le cas chez *C. beijerinckii* NCIMB 8052). Ces éléments suggèrent que certains résultats scientifiques décrits chez *C. acetobutylicum* ne soient pas vérifiables chez *C. beijerinckii*.

1.3.1 Généralités

Métabolisme, définition: « ensemble des processus complexes et incessants de transformation de matière et d'énergie par la cellule ou l'organisme, au cours des phénomènes d'édification et de dégradation organiques (anabolisme et catabolisme) »¹²²

En tant qu'anaérobies, les *Clostridia* solvantogènes ne peuvent utiliser la phosphorylation oxydative pour générer de l'énergie sous forme d'ATP. A la place, les organismes A/IBE ont mis en place un mécanisme fermentaire biphasique très particulier. A l'état végétatif, l'ATP produit est issu d'une part des voies de dégradation des substrats carbonés, et d'autre part de la conversion de l'acétyl-CoA et du butyryl-CoA, produits du métabolisme intermédiaire, en acide acétique et butyrique, respectivement. Cette période de production d'acides est appelée « acidogénèse » et constitue la voie préférentielle de génération d'énergie lors de la phase de croissance exponentielle. Chez *C. acetobutylicum* ATCC 824, l'accumulation d'acides organiques provoque une baisse significative du pH extracellulaire, alors que le pH intracellulaire, que la cellule essaye de maintenir à la neutralité, varie moins. En conséquence, le gradient de pH transmembranaire (ΔpH) augmente^{123,124}. Dans une souche dérivée de *C. pasteurianum* ATCC 6013, Kell *et al.* montrent en 1981 que les acides organiques sont capables de traverser la membrane cellulaire majoritairement sous leur forme indissociée¹²⁵. La variation du ΔpH qui en découle se traduit par une accumulation des

acides organiques dans la bactérie, avec des concentrations en acétate et butyrate atteignant jusqu'à 13 et 7 fois la valeur de la concentration extracellulaire, respectivement¹²⁶. L'accumulation d'acides favorise la diminution du pH intracellulaire, source de stress considérable pour la bactérie: la disruption du gradient de pH transmembranaire finit en effet par provoquer l'arrêt des fonctions métaboliques principales et, à terme, la mort cellulaire³².

En réponse à ce phénomène d'acidification, la morphologie des cellules de *Clostridium* évolue pour adopter une forme de cigare gonflé, appelée « forme clostridiale », en accumulant progressivement du granuloze, une forme de stockage du glucose¹²⁷. Ce changement morphologique s'accompagne d'un phénomène de reconsommation des acides qui sont réduits successivement en aldéhydes puis en alcools. Le mécanisme réactionnel provoque également la production d'acétone, converti dans certaines souches en isopropanol sous l'action d'une alcool secondaire déshydrogénase (s-ADH)^{38,39}. Pendant cette phase de la fermentation, appelée « solvantogénèse », la croissance bactérienne est très réduite, principalement parce que les voies de production des solvants sont moins favorables que celles des acides pour la génération d'ATP, ce qui réduit considérablement la quantité d'énergie utilisable pour produire de la biomasse.

Si les deux phases métaboliques distinctes sont bien caractérisées, la régulation du passage de l'une à l'autre, phénomène appelé « revirement métabolique », n'est toujours pas parfaitement élucidée. Un certain nombre de facteurs semblent être impliqués, comme le pH extracellulaire, la concentration de la source de carbone, le ratio NADH/NAD⁺ ou le dépassement d'une quantité seuil d'acides acétique et/ou butyrique intracellulaire³². Parallèlement à ce revirement métabolique, un deuxième phénomène biologique de défense s'initialise: la sporulation. Ce processus, extrêmement complexe, met en scène un ballet de facteurs génétiques impliqués dans une succession de phases sporulatoires distinctes¹²⁸. La bactérie, en utilisant les réserves de granuloze qu'elle a accumulées comme source d'énergie, finalise en fin de phase stationnaire les étapes nécessaires à la formation d'une spore. Cet état physiologique se caractérise par une dormance métabolique (c'est-à-dire un arrêt des fonctions métaboliques) et par une résistance exceptionnelle à toute une variété de stress (par exemple : acidité du milieu, présence d'oxygène, température basse ou élevée, exposition aux UV)¹²⁹. Lorsque les conditions

environnementales redeviennent favorables, une spore est capable de se réactiver : la germination produit ainsi une cellule végétative, ce qui boucle par là-même le cycle de vie des *Clostridia* solvantogènes A/IBE (Figure 5).

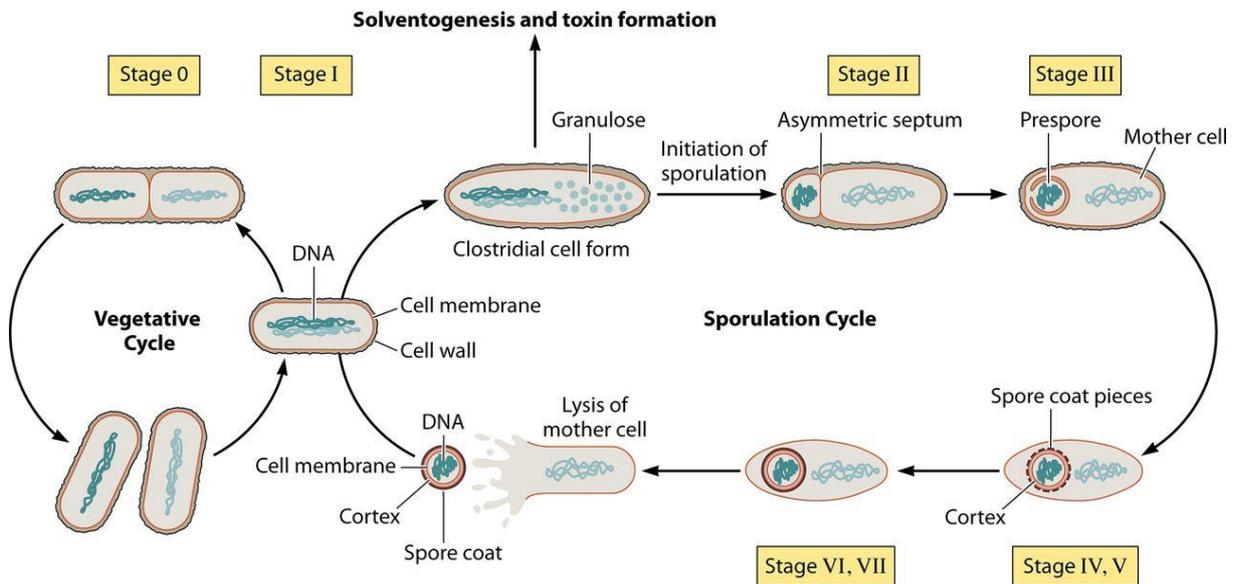


Figure 5: Cycle de vie des Clostridia¹²⁸.

La suite de cette partie de l'introduction vise à décrire les voies métaboliques majeures mises en place par les *Clostridia* solvantogènes au cours de leur cycle de vie cellulaire. Leur compréhension est en effet cruciale pour 1) choisir des substrats carbonés adaptés dans une optique industrielle, 2) orienter le métabolisme dans le sens de production des solvants, 3) déterminer des cibles génétiques pour améliorer les performances (rendement, sélectivité) fermentaires.

1.3.2 Du substrat carboné au pyruvate

Les espèces solvantogènes du genre *Clostridium* sont capables de fermenter une variété importante de mono- et polysaccharides⁴³ (Tableau 2), caractéristique très avantageuse pour la fermentation ABE puisqu'elle rend possible l'utilisation de nombreux substrats, issus de l'agriculture, de l'industrie ou de déchets^{74,92,98,99}. Toutes les voies mises en jeu dans la dégradation des sucres permettent d'aboutir à la synthèse de pyruvate, molécule de base pour la suite du métabolisme du carbone¹³⁰.

Au cours de cette thèse, les résultats ont été obtenus en utilisant un panel de sucres limité (glucose, fructose, saccharose). Pour la suite, seules les voies métaboliques permettant leur consommation seront détaillées.

Analyse Bibliographique

Espèce	<i>C. acetobutylicum</i>	<i>C. beijerinckii</i>
Nbr de souches testées	7	16
Hexoses		
Glucose	++	++
Fructose	++	++
Mannose	++	++
Pentoses		
Xylose	++	++
Arabinose	++	++
Ribose	--	+/-
Rhamnose	-	+ (f)
Hétéroholosides		
Methyl-glucopyranoside	++	++
Amygdaline	++	++
Salicyline	++	++
Diholosides		
Saccharose	++	++
Cellobiose	++	++
Lactose	++	++
Maltose	++	++
Trehalose	-	++
Turanose	+ (f)	++
Mélibiose	-	+
Triholosides		
Raffinose	++	++
Melezitose	-	++
Polyholosides		
Dextrine	++	++
Glycogen	++	+
Amidon	++	++
Pectine	++	++
Inuline	+	++
Polyols		
Glycerol	++ (f)	++ (f)
Arabitol	--	+
Dulcitol	--	+
Inositol	--	++
Mannitol	++	++
Sorbitol	+/-	++

Tableau 2: Liste non exhaustive des substrats carbonés métabolisables par *C. acetobutylicum* et *C. beijerinckii*, basée sur les travaux de Keis et al⁴³.

Pour chaque espèce, plusieurs souches ont été testées, et la notation indique la proportion de souches capables de fermenter chaque substrat. ++ : toutes les souches; + : la plupart des souches; +/- : 40-60% des souches ; - : une minorité des souches ; -- : aucune souche ; (f) indique une métabolisation faible du substrat.

1.3.2.1 Voies de dégradation : glucose, fructose & saccharose

Le glucose, le fructose et le saccharose sont classiquement métabolisés par la voie d'Embden-Meyerhof-Parnas (EMP, glycolyse). Les différentes étapes, en partant de chacun de ces substrats, sont représentées dans la Figure 6. L'utilisation de ces substrats commence avec leur importation dans la cellule. Chez les *Clostridia* solvantogènes, la majeure partie des sucres sont transportés dans la cellule via le système phosphotransférase (SPT)^{115,131}. Ce système se distingue des autres par sa double activité : l'hydrolyse d'une molécule d'ATP permet l'import dans la cellule du sucre concerné ainsi que sa phosphorylation simultanée, ce qui permet de réaliser la première étape de la voie EMP. Le transport SPT est également central pour les cascades de régulation génétique de la RCC chez les solvantogènes, l'état de phosphorylation du transporteur jouant un rôle dans la répression ou l'activation des gènes contrôlés par la RCC^{115,131}. Chez *C. beijerinckii*, le fructose et le glucose sont importés de cette manière¹³², bien que ce dernier puisse également être récupéré par un autre système de transport (symport avec un proton¹³³). Le saccharose, un disaccharide constitué d'un monomère de glucose et d'un monomère de fructose reliés par une liaison α (1 \rightarrow 2) β , est également importé directement par un système SPT, avant d'être clivé dans la cellule et de rejoindre les voies classiques de la glycolyse^{134,135}.

Chaque molécule de glucose dégradée permet de générer deux molécules de pyruvate, deux molécules de NADH, ainsi que deux molécules d'ATP. En tant qu'organismes anaérobies, les *Clostridia* solvantogènes ne disposent pas de la chaîne de transport d'électrons qui leur permettrait, outre de générer des molécules d'ATP supplémentaires par molécule de glucose consommée, de réoxyder le NADH en NAD⁺. L'équilibre redox, sous la forme du ratio NADH/NAD⁺, doit en conséquence être contrôlé d'une autre manière, et ce contrôle est fondamental pour le métabolisme de la bactérie (cf. sous-partie 1.3.4).

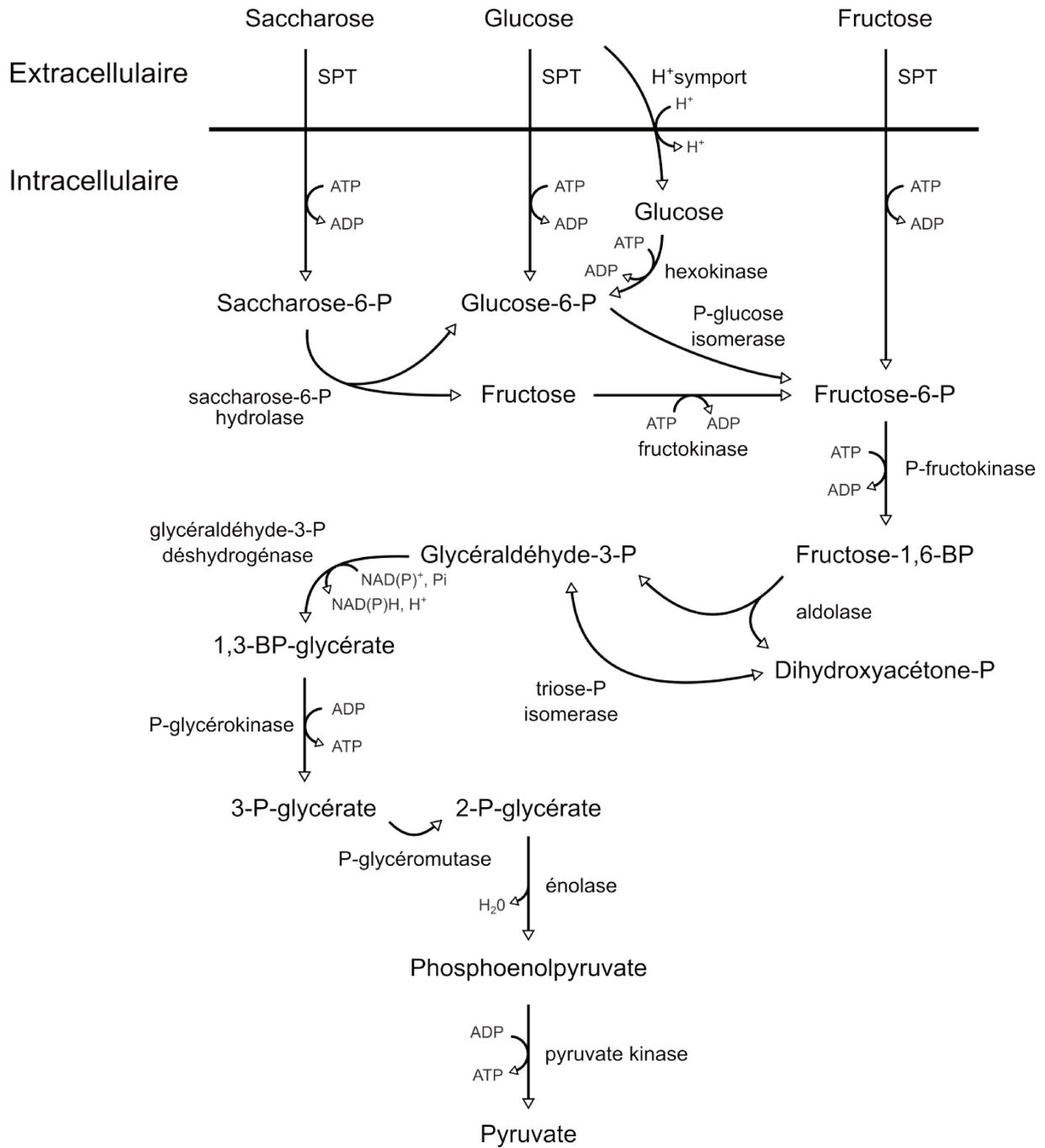


Figure 6: Représentation schématique de l'import du glucose, fructose et saccharose et de leur métabolisation par la voie EMP chez *C. beijerinckii*.

P : phospho ou phosphate. BP : bisphospho ou bisphosphate

1.3.2.2 Voies de dégradation : autres substrats carbonés simples

Le Tableau 3 récapitule sommairement les principales voies hétérotrophiques connues pour d'autres sources carbonées dans les *Clostridia* solvantogènes. Récemment, la mixotrophie de *C. beijerinckii* a été mise en évidence¹³⁶, avec une fixation de carbone et d'hydrogène inorganique observée en présence de substrats carbonés conventionnels. La contribution des voies putatives inhérentes à ce type de fixation est toutefois mineure par rapport à la consommation totale du carbone par la cellule.

Voie	Source carbonée modèle	Espèce utilisée pour la caractérisation	Référence
Voie du tagatose-6-P	Galactose	<i>C. acetobutylicum</i>	Servinsky et al. ¹¹¹
Voie de Leloir	Galactose	<i>C. acetobutylicum</i>	Servinsky et al. ¹¹¹
Voie des pentoses phosphates	Xylose	<i>C. acetobutylicum</i>	Volesky et al. ¹³⁷ , Gu et al. ¹³⁸
Voie des phosphocétolases	Xylose	<i>C. acetobutylicum</i>	Liu et al. ¹³⁹
Voie du glycérol	Glycérol	<i>C. acetobutylicum</i>	Vasconcelos et al. ⁸⁰
Voie du propanediol	Glycérol	<i>C. beijerinckii</i>	Gungormunsler et al. ⁵⁰
Voie du rhamnose	Rhamnose	<i>C. beijerinckii</i>	Diallo et al. ¹⁴⁰

Tableau 3: Principales voies de consommation de sources carbonées (hors glucose, fructose, saccharose).

1.3.2.3 Voies de dégradation des polysaccharides

Le métabolisme des polysaccharides dépend essentiellement de la nature des monomères dont ils sont constitués, du type de liaisons glucosidiques qui les relie, ainsi que de la quantité de ces monomères par molécule de polysaccharide. La plupart des disaccharides d'importance biologique (saccharose, maltose, lactose, cellobiose) peuvent être directement importés dans la cellule, avant d'être clivés dans leurs monomères respectifs, qui seront eux-mêmes pris en charge par les voies métaboliques présentées ci-dessus¹¹¹. Les polysaccharides de plus grand poids moléculaire ne peuvent toutefois traverser la barrière membranaire. Pour répondre à ce problème, les *Clostridia* A/IBE sont capables d'exporter dans le milieu extracellulaire un certain nombre d'enzymes dont le rôle est de réduire la taille des

polymères de sucre. Les sucres simples (c'est-à-dire mono- et disaccharides) ainsi obtenus peuvent alors être importés et métabolisés^{111,141}.

Toutefois, malgré une activité cellulolytique faible relevée chez *C. acetobutylicum* ATCC 824 lorsque cette souche est cultivée sur certaines sources carbonées¹⁴², celle-ci n'est pas suffisante pour permettre aux souches A/IBE d'être cultivées sur substrat cellulosique / hémicellulosique. D'autres souches solvantogènes en sont capables, comme *C. cellulolyticum* et *C. cellulovorans*, par le biais d'un complexe protéique multimérique excrété dans la matrice extracellulaire, le cellulosome¹⁴³. De façon intéressante, des homologues pour la plupart des gènes du cellulosome sont présents chez *C. acetobutylicum* ATCC 824, et la complémentation des gènes existants par génie génétique, dans le but de restaurer cette activité, aurait une grande valeur scientifique puisqu'elle permettrait potentiellement d'utiliser la biomasse lignocellulosique directement (c'est-à-dire sans prétraitement thermo-chimique / enzymatique)¹⁴⁴.

1.3.3 Du pyruvate aux solvants

Toutes les voies métaboliques décrites par la suite sont représentées dans la Figure 7.

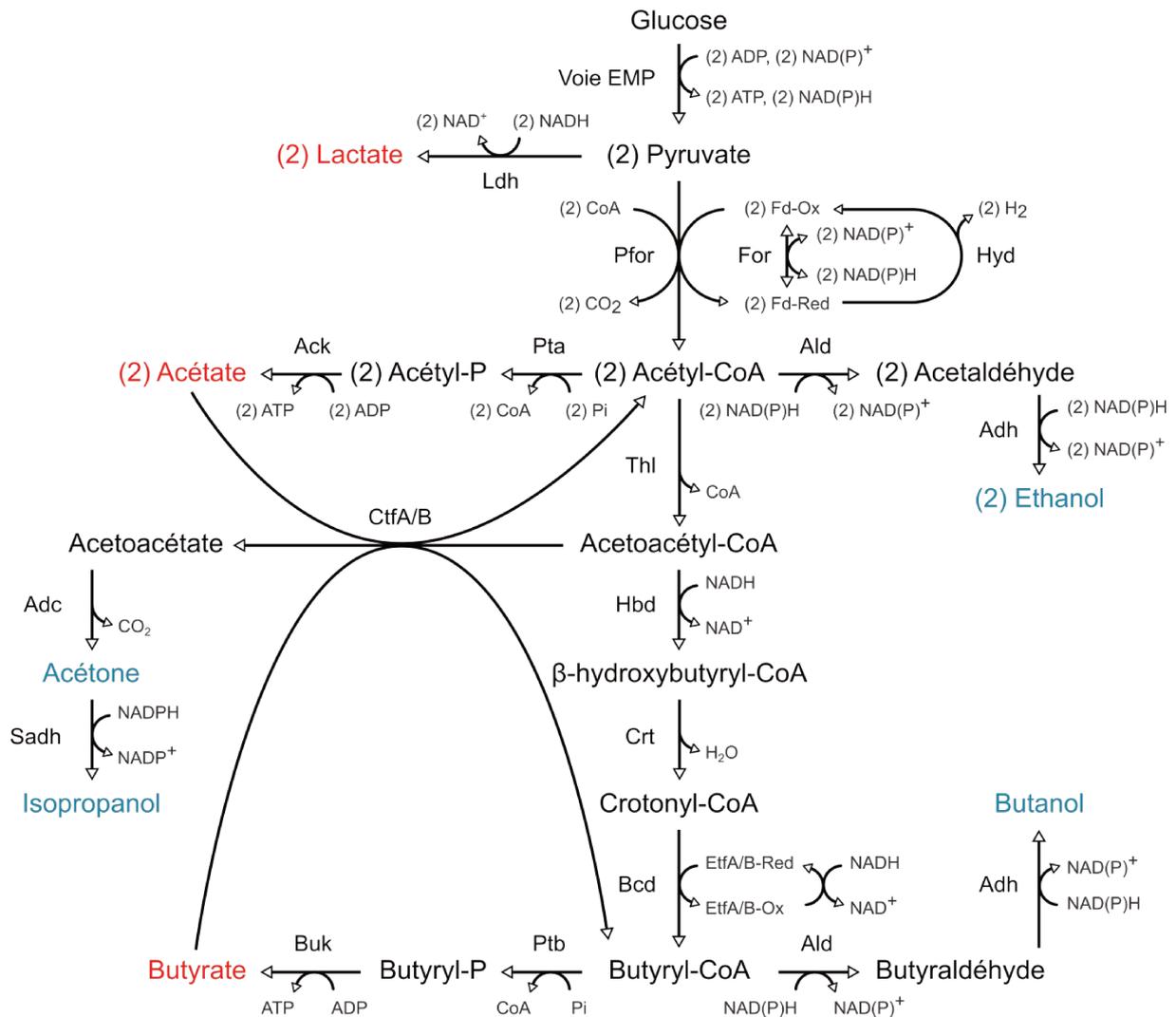


Figure 7 : Représentation schématique des voies majeures du métabolisme A/BE.

Les produits finaux majeurs du métabolisme sont indiqués en couleur (rouge : acides ; bleu : solvants). La voie de décarboxylation du pyruvate et la voie du 2,3-butanediol ne sont pas représentées. Abréviations non communes, ordre alphabétique : Ack : acétate kinase ; Adc : acétoacétate décarboxylase ; Adh : alcool déshydrogénase ; Ald ; aldéhyde déshydrogénase ; Bcd : butyryl-CoA déshydrogénase ; Buk : butyrate kinase ; CtfA/B : CoA transférases A et B ; Crt : crotonase ; EtfA/B : *electron tranfert subunits A & B* ; Fd : ferrédoxine ; For : ferrédoxine oxydoréductase ; Hbd : hydroxybutyryl déshydrogénase ; Hyd : hydrogénase ; Ldh : lactate déshydrogénase ; Pfor : pyruvate ferrédoxine oxydoréductase ; Pta : phosphotransacétylase ; Ptb : phosphotransbutyrylase ; Thl : thiolase.

1.3.3.1 Voies de synthèse de l'acétyl-CoA et du butyryl-CoA

1.3.3.1.1 Du pyruvate à l'acétyl-CoA

Le pyruvate, produit de la dégradation des sources carbonées, est oxydé par une pyruvate-ferrédoxine oxydoréductase (Pfor) en acétyl-CoA selon la réaction suivante :



La décarboxylation du pyruvate s'accompagne donc de la génération de CO₂ et de ferrédoxine (Fd) réduite. La ferrédoxine oxydée, cofacteur qui permet ici le transfert d'électrons, doit toutefois être régénérée pour assurer le renouvellement de la réaction. En fonction de l'état physiologique de la cellule, deux voies principales peuvent être empruntées pour ce faire. Pendant la phase d'acidogénèse, la ferrédoxine est principalement oxydée par l'action d'une hydrogénase, HydA, ce qui se traduit également par la production de dihydrogène. L'action des NAD(P)H-ferrédoxine oxydoréductases est favorisée pendant la solvantogénèse, et permet la production simultanée de NAD(P)H et de ferrédoxine oxydée³².

1.3.3.1.2 De l'acétyl-CoA au butyryl-CoA

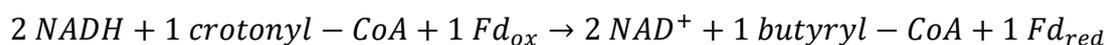
L'acétyl-CoA généré par l'oxydation du pyruvate est au centre du métabolisme branché des souches A/IBE. Cette molécule est impliquée dans trois voies métaboliques majeures : la voie de synthèse de l'acétate, celle de l'éthanol et celle du butyryl-CoA. Cette dernière voie est active à la fois pendant l'acidogénèse et pendant la solvantogénèse - bien que son activité soit fluctuante et soit maximale pendant la solvantogénèse - et se caractérise par l'action successive de quatre enzymes : une thiolase, une β-hydroxybutyryl-CoA déshydrogénase, une crotonase, et une butyryl-CoA déshydrogénase¹⁴⁵⁻¹⁴⁷. Boynton et collègues ont montré que chez *C. acetobutylicum* ATCC 824, les trois derniers gènes sont organisés en opéron (opéron BCS, qui comprend également les gènes *etfA* et *etfB*)¹⁴⁸. La régulation transcriptionnelle spécifique de cet opéron, ainsi que celle du gène de la thiolase, semblerait dépendre directement du ratio NADH/NAD (cf. sous-partie 1.3.4)¹⁴⁹.

Au niveau moléculaire, la thiolase catalyse la réaction de condensation de deux molécules d'acétyl-CoA en acétoacétyl-CoA, avec relargage d'une molécule de CoA.

Cette enzyme, première de la voie, est en compétition avec les voies de synthèse de l'acétate et de l'éthanol pour l'acétyl-CoA et joue *de facto* un rôle important dans la répartition du flux de carbone entre molécules à deux carbones (C2 : acétate, butyrate) et à trois/quatre carbones (C3 : acétone, isopropanol ; C4 : butyrate, butanol)¹⁴⁵. Par ailleurs, pendant la phase acidogène la production d'ATP dépendante de la voie du butyrate est deux fois moins importante que celle dépendant de la production d'acétate. L'activité de la thiolase doit donc être parfaitement régulée pour optimiser le rendement énergétique de la cellule. Dans cette optique, le ratio CoA/acétyl-CoA a été suggéré comme ayant un rôle majeur, dans la mesure où la thiolase est inhibée par le CoA pour des concentrations de l'ordre du micromolaire, et où la concentration d'acétyl-CoA dans les conditions physiologiques n'est pas suffisante pour saturer l'enzyme^{145,150}. Il semblerait par ailleurs que le gène codant la thiolase soit essentiel chez *C. acetobutylicum*¹⁵¹. Des travaux récents ont toutefois montré que les gènes *thl* et *hydA* étaient inactivables simultanément, ce qui permet d'obtenir un phénotype assez particulier : les C3-C4 sont complètement éliminés du métabolisme et la production d'hydrogène est interrompue¹⁵².

L'acétoacétyl-CoA est ensuite pris en charge par la β -hydroxybutyryl-CoA déshydrogénase (Hbd) pour produire du 3-hydroxybutyryl-CoA. Pour ce faire, cette réaction de réduction consomme également du NADH^{148,153}. Hbd est en compétition avec la CoA transférase (de la voie de synthèse de l'acétone / isopropanol) pour l'acétoacétate et est de fait impliqué dans la répartition C3/C4 du carbone. Lehmann et Lütke-Eversloh ont étudié l'effet de la délétion du gène qui code cette enzyme chez *C. acetobutylicum* ATCC 824¹⁵⁴. De façon intéressante, le phénotype observé est similaire à celui de la double délétion *thl/hydA* : les C4 sont éliminés, la production de dihydrogène et d'acétone est très fortement perturbée, et le flux de carbone redirigé presque intégralement vers les C2.

La crotonase et la butyryl-CoA déshydrogénase (Bcd) catalysent les deux dernières étapes de la bioconversion de l'acétyl-CoA en butyryl-CoA : la déshydratation du β -hydroxybutyryl-CoA en crotonyl-CoA puis sa réduction en butyryl-CoA. Li *et al* ont purifié l'enzyme Bcd chez *C. kluyveri* et montrent que cette étape de réduction, si elle consomme du NADH, est également couplée à une réduction concomitante de la ferrédoxine selon le bilan suivant¹⁵⁵ :



Pour ce faire, Bcd est impliqué dans un complexe protéique qui comprend également les protéines EtfA et EtfB et qui permet de coupler successivement la réduction de la ferrédoxine (exergonique) à la réduction du crotonyl-CoA (endergonique). En fonction de l'état physiologique de la cellule (tout comme pour la Pfor), la ferrédoxine réduite peut être prise en charge pour régénérer son équivalent oxydé par l'hydrogénase HydA ou par une NAD(P)H-ferrédoxine oxydoréductase. En fonction de la voie suivie par la ferrédoxine réduite, le bilan réactionnel est modifié en conséquence (production de H₂ ou de NAD(P)⁺, respectivement).

1.3.3.2 Voies de synthèse des acides

1.3.3.2.1 De l'acétyl-CoA à l'acétate

La production d'acétate permet aux *Clostridia* solvantogènes de générer de l'énergie en phase acidogène, avec un bilan en ATP par unité de glucose doublé par rapport à la glycolyse seule.

La bioconversion de l'acétyl-CoA en acétate commence par la phosphorylation de la phosphate acétyltransférase en acétyl-phosphate. Cette réaction consomme du phosphate inorganique et libère le CoA en parallèle. Sous l'action d'une acétate kinase, le phosphate de l'acétyl-phosphate est ensuite transféré sur un ADP pour produire d'une part l'ATP, et d'autre part l'acétate en lui-même.

Cette voie est surtout active pendant la phase acidogène, et son activité est fortement diminuée - jusqu'à 90% de baisse d'activité - en phase solvantogène^{146,156}. Boynton et collègues relèvent une organisation en opéron des deux gènes impliqués dans cette voie (*pta*, *ack*) et suggèrent une régulation commune, transcriptionnelle, qui expliquerait ce phénomène¹⁵⁶. Toutefois, Winzer *et al.* montrent par *northern blot* chez *C. acetobutylicum* DSM 1731 que *ack* est transcrit tout au long de la fermentation, ce qui suppose au contraire une régulation post-transcriptionnelle de l'activité des deux gènes de l'opéron¹⁵⁷.

La voie de l'acétate, malgré son rôle pour la production d'énergie chez *Clostridium*, n'est pas essentielle pour les souches A/IBE. En effet, cette voie a été perturbée à plusieurs reprises chez *C. acetobutylicum*, avec différentes méthodes, et des phénotypes obtenus assez variés^{69,71,151,158}. Si la diversité des phénotypes

(baisse ou non de la production d'acétate) est difficile à expliquer, ces études illustrent toutes la flexibilité des *Clostridia* solvantogènes, qui, grâce à leur métabolisme branché, peuvent efficacement rediriger le flux de carbone en réponse à une perturbation métabolique (vers le même produit grâce à une redondance génétique, ou vers d'autres produits comme le butanol ou l'éthanol).

1.3.3.2.2 Du butyryl-CoA au butyrate

La voie de synthèse du butyrate est très similaire à celle de l'acétate. Une phosphate transbutyrylase (Ptb) catalyse la phosphorylation du butyryl-CoA, et le butyryl-phosphate est ensuite pris en charge par une butyrate kinase (Buk), qui génère, outre le butyrate, une molécule d'ATP. A cause de la condensation par la thiolase de deux molécules d'acétyl-CoA, cette voie produit deux fois moins d'ATP que la voie de l'acétate. Elle permet toutefois, en passant par la voie de synthèse du butyryl-CoA, de consommer du NADH, qui est sur-produit en phase acidogène.

Bien que les gènes *ptb* et *buk* soient organisés en opéron¹⁵⁹, leurs produits ont un comportement enzymatique distinct qui suggère une régulation post-transcriptionnelle de leur activité¹⁴⁶. Chez *C. acetobutylicum* ATCC 824, Ptb et Buk présentent un pic d'activité enzymatique en phase acidogène, qui décroît largement en solvantogénèse.

Similairement à la voie de l'acétate, la voie du butyrate a été perturbée à plusieurs reprises chez *C. acetobutylicum*, avec toutefois une plus grande homogénéité dans les résultats (baisse systématique de la production de butyrate et augmentation de la production de solvants)^{69,160,161}. Les résultats les plus récents, obtenus par une technique de génie génétique plus perfectionnée, suggèrent que *buk* ne peut être inactivé sans que *ptb* ne le soit, probablement à cause de la toxicité d'une accumulation trop importante de butyryl-phosphate¹⁶¹. Dans le mutant $\Delta buk\Delta ptb$ résultant de cette étude, un nouveau métabolite est détecté en phase acidogène (2-hydroxy-valérate) et la production de butanol est augmentée de 58%. Encore une fois, ces résultats soulignent la flexibilité métabolique des souches A/IBE.

1.3.3.3 Voies de synthèse des solvants

1.3.3.3.1 Réassimilation des acides et production d'acétone

En fin de phase acidogène, l'accumulation des acides entraîne la mise en place d'un mécanisme de défense qui provoque leur réassimilation. Pour ce faire, la voie la plus simple consiste à emprunter le chemin inverse des voies de synthèse des acides. Les travaux d'Hartmanis et de Gatenbeck prouvent la réversibilité de la réaction catalysée par la butyrate kinase (Buk) en conditions physiologiques, avec un pic d'activité dans cette direction détecté en début de phase solvantogène¹⁴⁶. Wiesenborn *et al.* démontrent un mécanisme de régulation intéressant pour l'enzyme Ptb, dont l'activité dans les deux directions réactionnelles possibles est affectée différemment par le pH¹⁶². Une baisse du pH entraîne une baisse drastique de l'activité enzymatique dans le sens de formation du butyrate, mais affecte sensiblement moins la réaction inverse. Ceci pourrait expliquer la consommation du butyryl-phosphate par cette enzyme à faible pH intracellulaire (c'est-à-dire inférieur à 6, condition vérifiée en fin de phase acidogène¹²⁴). Plus récemment, des études de flux métabolique¹⁶³⁻¹⁶⁵ réalisés chez *C. acetobutylicum* ATCC 824, corroborées par des observations de Lehmann *et al.*⁷¹, ont montré que cette voie était partiellement empruntée pour la reconsommation du butyrate, mais pas pour l'acétate.

Cette voie, énergétiquement peu intéressante car consommatrice d'ATP, est identifiée dès 1984 comme ne pouvant expliquer seule le phénomène de reconsommation des acides¹⁶⁶. L'action d'une CoA transférase s'avère nécessaire : cette dernière catalyse le transfert de CoA de l'acétoacétyl-CoA aux acides organiques (acétate, butyrate ou autres), ce qui permet de générer les acyl-CoA correspondants tout en produisant de l'acétoacétate^{166,167}. L'acétoacétate est ensuite réduit en acétone par une acétoacétate décarboxylase (Adc), avec production de CO₂. Comparativement au mécanisme précédent, celui-là est avantageux : la CoA transférase ne consomme pas directement d'ATP, ce qui favorise grandement son utilisation. Wiesenborn et collègues ont purifié la CoA transférase de *C. acetobutylicum* ATCC 824 et identifié une structure hétérotétramérique, contenant deux sous-unités protéiques α et deux sous-unités β ¹⁶⁷. Cary *et al.* ont identifié les gènes associés à ces sous-unités : *ctfA* et *ctfB*, tous deux présents sur le mégaplasmide pSOL de *C. acetobutylicum* ATCC 824¹⁶⁸. Ces deux gènes sont organisés en opéron avec *adhE1*, qui code une

alcool/aldéhyde déshydrogénase bifonctionnelle impliquée dans la synthèse de butanol et d'éthanol¹⁶⁹.

Différentes stratégies ont visé à perturber l'expression des gènes *ctfA/B* et *adc* avec des effets assez similaires chez *C. acetobutylicum*^{71,151,170,171} : en diminuant ou en éliminant la synthèse d'acétone, la réassimilation des solvants est impactée, ce qui se traduit par une baisse de la production de solvants. A l'inverse, lorsque la voie de l'acétone est renforcée via la construction d'un opéron synthétique surexprimant les gènes *ctfA*, *ctfB* et *adc*, un quasi-doublement du titre en acétone est mesuré par rapport à la souche sauvage¹⁷². Cette amélioration se traduit également par la production simultanée d'une plus grande quantité d'éthanol et de butanol, probablement due à la réutilisation d'une plus grande quantité d'acides. Tous ces résultats démontrent à nouveau la difficulté de modifier une voie sans en perturber une autre dans un métabolisme branché et dynamique comme celui des *Clostridia* A/IBE.

1.3.3.3.2 Production des alcools

Tout comme les voies de synthèse des acides, les voies de production de butanol et d'éthanol sont branchées respectivement au niveau du butyryl-CoA et de l'acétyl-CoA. Les deux voies sont similaires et catalysées par les mêmes enzymes : une première étape de réduction permet d'obtenir les aldéhydes correspondants, et une seconde achève le processus en convertissant les aldéhydes en alcools³².

Contrairement à la plupart des flux métaboliques décrits jusqu'à présent, plusieurs enzymes sont impliquées pour chaque réaction et catalysent à la fois la bioconversion de l'acétyl-CoA et du butyryl-CoA. Chez *C. acetobutylicum*, cinq enzymes contribuent majoritairement à la voie de production des alcools : deux alcool / aldéhyde déshydrogénases bifonctionnelles (AdhE1, AdhE2) et trois alcool déshydrogénases (BdhA, BdhB, BdhC). La contribution de chacune d'entre elles dans les flux de carbone est complexe et n'a été élucidée que récemment chez *C. acetobutylicum* ATCC 824^{173,174}. Par une approche biochimique et omique, Yoo *et al.* ont quantifié le nombre de protéines et les métabolites (butanol, éthanol notamment) présents dans différentes conditions expérimentales. Leurs analyses permettent d'évaluer l'implication de chacune des protéines dans le flux de carbone pour différents états métaboliques¹⁷⁴. Ce type d'analyse n'a toutefois jamais été réalisé chez *C.*

beijerinckii, et l'implication des nombreuses alcool / aldéhydes déshydrogénases encodées dans les génomes des souches de cette espèce ne peut, en l'état, qu'être suggérée par de précédentes études transcriptomiques^{52,175,176}. En effet, dans les transcriptomes correspondants (cinétique au cours de fermentations de type *batch* sur milieu modèle), seule une aldéhyde déshydrogénase et 3 alcool déshydrogénases sont exprimées en phase solvantogène, ce qui suggère que, dans ces conditions expérimentales, les protéines correspondantes sont les actrices majeures de la production d'alcools.

Chez *C. acetobutylicum*, la production d'alcools peut également se faire par une autre voie. La voie fermentaire éthanolique classique (via une pyruvate decarboxylase) est en effet fonctionnelle dans la souche ATCC 824¹⁷³. L'analyse comparée des génomes de souches de ces deux espèces semble toutefois indiquer que cette voie est absente chez *C. beijerinckii*.

Dans les souches IBE, un alcool supplémentaire est également produit: la voie de l'acétone est complétée par une alcool secondaire déshydrogénase (s-ADH) dépendante du NADPH qui produit l'isopropanol, et qui est également capable de réduire le butyraldéhyde^{38,39}. Chez *C. beijerinckii* DSM 6423 le gène qui l'encode forme un opéron avec deux autres gènes⁵². Leur produit sont une hydrogénase putative¹⁷⁷, ainsi qu'un analogue de SudA, une sulfure déshydrogénase putative.

La souche DSM 6423 n'ayant, avant les travaux réalisés au cours de cette thèse, jamais été modifiée génétiquement par une approche ciblée, l'étude de la s-ADH s'est pour l'instant limitée à son expression hétérologue chez *C. acetobutylicum* ATCC 824. L'insertion de ce gène a permis à plusieurs reprises d'obtenir des souches IBE performantes⁷⁵⁻⁷⁸, avec une conversion presque parfaite de l'acétone et des titres en isopropanol compris entre 5 et 8 g.L⁻¹.

1.3.4 La solvantogénèse, un processus influencé par de nombreux facteurs

Dans une optique de production d'alcools supérieurs par voie A/IBE, la compréhension et la maîtrise des voies solvantogènes et de leur régulation sont des prérequis cruciaux. Si les voies métaboliques inhérentes sont caractérisées, la façon dont elles sont induites n'est pas entièrement comprise. En particulier, les phénomènes moléculaires qui engendrent le revêtement métabolique, étape charnière

du cycle de vie clostridien entre acidogénèse et solvantogénèse, ne sont pas parfaitement élucidés. Il semblerait en effet qu'une multitude de facteurs connus et inconnus influencent collectivement l'expression des voies solvantogènes.

A l'échelle de l'organisme, l'environnement direct joue par exemple un rôle déterminant. Des facteurs simples, liés au mode de culture, comme le pH, la nature des nutriments et leur disponibilité, la présence d'oxygène, de métabolites, l'agitation ou la température peuvent significativement affecter les voies A/IBE³². Très tôt, certains auteurs ont avancé le rôle de la concentration intracellulaire en butyrate^{69,164} ou en butyryl-phosphate^{124,178}, produits du métabolisme, comme déclencheurs de la solvantogénèse. Ces hypothèses ont toutefois été mises à mal par les inactivations des gènes de la voie du butyrate¹⁶¹, qui se traduisent par une amélioration de la production d'alcools.

Par ailleurs, en manipulant les conditions environnementales, un métabolisme solvantogène peut être induit et maintenu pendant de longues périodes. Il est par exemple possible de maintenir en fermentation continue *C. acetobutylicum* ATCC 824 en phase acidogène, solvantogène, ou alcologène (production majeure de butanol et d'éthanol, sans acétone), en régulant le pH de la culture et en jouant sur l'équilibre redox (en utilisant du glycérol, qui est un substrat plus réduit que le glucose)^{80,174,179}.

L'équilibre redox intracellulaire est également une source majeure de régulation de la solvantogénèse. Pour de nombreuses étapes, le métabolisme du carbone dans les souches A/IBE est en effet inextricablement lié aux réactions de transfert d'électrons. Les organismes aérobies sont capables d'utiliser les nucléotides réduits produits pendant la glycolyse pour générer de l'énergie, tout en maintenant l'équilibre redox. Ce n'est pas le cas des souches A/IBE qui utilisent leur métabolisme particulier pour arriver à cet équilibre. L'appréciation de cette relation intime entre les flux de carbone et d'électrons n'est pas nouvelle : Jones et Wood décrivent dès 1986 la ferrédoxine comme la « molécule pivot » autour de laquelle tout le métabolisme A/IBE est articulé³².

La ferrédoxine de *C. pasteurianum* ATCC 6013 est une petite protéine de 55 acides aminés qui comporte deux noyaux fer/soufre constitués de quatre atomes de chaque¹⁸⁰. Cette protéine, caractérisée par un très faible potentiel d'oxydoréduction¹⁸¹ ($E_0 = -400$ mV), est un partenaire réactionnel idéal pour les réactions d'oxydoréduction, ce qui lui vaut d'être particulièrement bien conservée parmi les

organismes anaérobies. La ferrédoxine est, entre autres, impliquée dans la réaction de décarboxylation oxydative du pyruvate en acétyl-CoA, et a donc un rôle central dans le métabolisme A/IBE¹⁸².

En acidogénèse, c'est-à-dire dans un état de forte disponibilité en NAD(P)H, les électrons captés par la ferrédoxine pendant l'oxydation du pyruvate sont transférés à une hydrogénase, HydA, qui réduit des protons en dihydrogène. Le dihydrogène s'évapore, emportant par là-même l'excès d'énergie réductrice produit par la bactérie en phase acidogène. Pendant la solvantogénèse, le NAD(P)H est consommé par les voies de synthèse des solvants, en plus grande quantité que la glycolyse n'en fournit. Les NAD(P)-ferrédoxine oxydoréductases sont alors un moyen pour la bactérie de régénérer la ferrédoxine oxydée tout en produisant du NAD(P)H. En absence de glucose-6-phosphate déshydrogénase (une enzyme productrice de NADPH) chez la plupart des *Clostridia*, les NADPH-ferrédoxine oxydoréductase ont été pressenties comme les actrices principales de la production du NADPH cellulaire¹⁸³, ce qui a été vérifié récemment chez *C. acetobutylicum*¹⁸⁴.

C. acetobutylicum et *C. beijerinckii* ne sont par ailleurs pas tout à fait similaires vis-à-vis de leur gestion de l'équilibre redox. En effet, si *C. beijerinckii* dispose de plusieurs hydrogénases permettant l'évacuation de l'excès de pouvoir réducteur, son génome contient également tous les gènes nécessaires à l'expression d'un système de transport d'électron de type Rnf (pour *Rhodobacter nitrogen fixation*)¹⁸⁵. Chez *C. ljungdahlii*, un *Clostridium* acétogène, ce système est utilisé pour régénérer la ferrédoxine oxydée, tout en produisant du NADH et en exportant un proton hors de la cellule¹⁸⁶. Le gradient de protons peut alors être utilisé pour générer de l'énergie sous la forme d'ATP via une ATP synthase F₀F₁. Bien que ce système n'ait pas été investigué chez *C. beijerinckii*, sa forte conservation suggère une fonction similaire dans cette espèce¹⁸⁷, et pourrait donc être utilisé en conjonction avec l'hydrogénase pour gérer le flux d'électrons générés par la glycolyse.

En phase solvantogène, le flux électronique est redirigé et s'accompagne d'une baisse d'activité de l'hydrogénase, et donc d'une baisse de la production de dihydrogène. L'activité de la NADH-ferrédoxine oxydoréductase (ou du système Rnf) peut également être dirigée dans le sens de formation du NAD⁺^{110,186}. En phase acidogène, la cellule peut ainsi régénérer du NAD⁺ tout en produisant de la ferrédoxine réduite. L'excès de pouvoir réducteur généré par la glycolyse peut ainsi être éliminé

via le passage des électrons de la ferrédoxine réduite au dihydrogène sous l'action de l'hydrogénase. Ce système permet également de produire de la ferrédoxine réduite pour l'activité de la butyryl-CoA déshydrogénase, autre enzyme importante qui utilise la ferrédoxine comme partenaire.

Le contrôle précis de la répartition du flux d'électrons par la ferrédoxine permet de réguler finement le pouvoir redox de la cellule. Elle est par conséquent essentielle dans l'adaptation du métabolisme aux conditions environnementales, comme mentionné précédemment avec le cas du glycérol. En modifiant ces conditions environnementales, il est donc possible de perturber le flux électronique et ainsi d'engendrer une modification du flux de carbone. La plupart des stratégies expérimentales visant le transport électronique ciblent l'hydrogénase, dont l'activité peut être facilement inhibée par une multitude de facteurs (e.g. limitation en fer¹⁸⁸, injection de monoxyde de carbone^{189,190}, ajout d'agents réducteurs comme le méthylviologène¹⁹¹ qui peut remplacer la ferrédoxine dans le transport d'électrons et les réactions d'oxydoréduction). La diminution de l'activité de l'hydrogénase provoque une concentration accrue de NAD(P)H dans le milieu, et la cellule élimine cet excès en favorisant les voies de synthèse des solvants en lieu et place des acides.

L'équilibre redox peut également directement intervenir dans des voies de régulation transcriptionnelle. La protéine Rex a, en particulier, été proposée comme régulatrice de l'expression de l'opéron de synthèse du butyryl-CoA (cf. sous-partie 1.3.3)¹⁴⁹. Le modèle proposé est le suivant : lorsque la concentration relative en NADH est élevée, la protéine Rex se fixe au NADH. Lorsque cette concentration diminue, Rex se libère et se fixe sur des séquences en amont des gènes *thl* (thiolase), *crt* (crotonase, premier gène de l'opéron) et *adhE2* (alcool déshydrogénase, qui catalyse la formation de butanol et d'éthanol) et réprime leur transcription. En inactivant ce gène, Wietzke et Bahl observent une augmentation conséquente (multipliée par plus de 3) de l'activité enzymatique spécifique de la thiolase, de la crotonase et des alcools déshydrogénases dépendantes du NADH, couplée à une augmentation de la production de butanol, d'éthanol, au détriment de la production d'H₂ et d'acétone¹⁴⁹.

Tous ces résultats suggèrent qu'un contrôle multi-échelle est exercé par le pouvoir réducteur sur les voies métaboliques A/IBE, et plus particulièrement sur la solvantogénèse. L'initiation de cette dernière a également été associée de longue date à un autre processus biologique majeur : la sporulation. En effet, dans certaines

conditions, les souches solvantogènes peuvent perdre simultanément toute faculté de production de solvants et de sporulation : on parle de dégénérescence¹⁶⁹. Ce phénomène aboutit à un arrêt de la culture précoce et à une mort cellulaire importante. La dégénérescence est un phénomène conservé chez les souches A/IBE, même si les mécanismes sous-jacents sont fortement dépendants des espèces considérées¹⁶⁹ – en particulier entre *C. beijerinckii* et *C. acetobutylicum*. Cette corrélation entre les deux processus biologiques est corroborée par le rôle crucial de Spo0A, le facteur de sporulation principal chez les *Clostridia*, dont la délétion élimine à la fois la production de solvants et la différenciation en spores^{192,193}. Cette interconnexion entre les processus biologiques de sporulation et de solvantogénèse a toutefois été remise en question à plusieurs reprises. Tracy et collègues, en étudiant l'effet de la délétion de facteurs de sporulation chez *C. acetobutylicum* ATCC 824, ont récemment réussi à découpler les deux phénomènes biologiques¹⁹⁴. L'inactivation de *sigG*, notamment, permet d'obtenir une souche non sporulante, mais capable de produire normalement des solvants. Il semblerait donc que ces deux processus biologiques soient bien distincts, et que les régulateurs communs de la solvantogénèse et de la sporulation se situent au niveau de leur initiation.

Malgré les avancées récentes, les mécanismes de régulation permettant l'initiation et le maintien de la solvantogénèse restent à détailler. Il semblerait en effet que les processus de contrôle mis en place par les *Clostridia* solvantogènes soient multiples et probablement interconnectés. Analyser ce type de régulation nécessite d'avoir une vue globale du fonctionnement de la cellule, ce qui est facilité par les méthodes d'analyse à haut-débit modernes.

1.4 REGULATION DU METABOLISME A/IBE

1.4.1 Les omiques, pour une meilleure compréhension du métabolisme A/IBE

L'utilisation des techniques d'analyse à haut débit modernes (les « omiques ») pour les *Clostridia* solvantogènes remonte au début des années 2000. Le séquençage du génome de la souche *C. acetobutylicum* ATCC 824 en 2001¹³⁰ et celui de *C. beijerinckii* NCIMB 8052 en 2007¹⁹⁵ ont permis de considérablement améliorer notre compréhension génétique de ces deux espèces. Les informations essentielles

recueillies par analyse génomique comparative (c'est-à-dire répartition, longueur, nature des gènes présents sur le génome) ont permis l'utilisation des autres analyses omiques émergentes.

Les plus populaires sont incontestablement les techniques transcriptomiques, notamment le *DNA microarray* et plus récemment le RNA-seq. Ces techniques, en proposant d'identifier puis de quantifier les ARNm d'un microorganisme dans des conditions données, permettent d'attribuer une valeur d'expression génétique pour chaque gène. En mesurant la variation d'expression entre deux conditions expérimentales différentes, une analyse globale de la relation entre ces conditions et l'expression génétique peut ainsi être établie.

Chez *C. beijerinckii* NCIMB 8052, la première étude transcriptomique publiée visait à décrypter deux phénomènes biologiques différents : la transition entre acidogénèse et solvantogénèse au cours d'une fermentation en mode *batch*, et le phénotype particulier du mutant BA101, souche hyper-productrice de butanol¹⁹⁶. L'analyse du *DNA microarray*, constitué d'environ 500 gènes, a mis en relation l'évolution du transcriptome au cours de la croissance du microorganisme avec sa physiologie. Par exemple, les gènes liés à la voie de synthèse des acides (*pta*, *ack*, *buk*) sont surexprimés en début de fermentation, avant d'être rapidement réprimés au fur et à mesure que la culture progresse vers l'état solvantogène. La transition entre les deux états est marquée par la surexpression des gènes liés aux voies de production des solvants (*ald*, *ctfA*, *ctfB*, *adc*) et à l'initiation de la sporulation (*spo0A*, *sigF*). Par la suite, d'autres gènes intervenant dans la cascade de régulation de la sporulation sont successivement activés. L'analyse de la différence entre le mutant BA101 et la souche sauvage a révélé une relation entre le phénotype hyper-producteur de la souche et un affaiblissement de la cascade de régulation de la sporulation et du système SPT¹⁹⁶.

Par la suite et grâce à l'évolution des technologies, une analyse similaire a pu être réalisée en RNA-seq (non limitante en termes de nombre de gènes considérés) sur cette même souche et par la même équipe^{175,176}. Outre une validation des résultats obtenus en *microarray*, le RNA-seq a permis d'identifier ou de confirmer un certain nombre de structures opéroniques d'importance, d'affiner l'annotation du génome en ajoutant ou en modifiant certains gènes, et dans un contexte génétique comprenant plus d'une vingtaine d'alcools déshydrogénases putatives différentes, d'identifier les

trois copies ayant un rôle prépondérant dans la production de solvants chez cette souche.

En identifiant des gènes clés du métabolisme, ce type d'analyse peut donc sensiblement simplifier une stratégie rationnelle d'ingénierie métabolique. Récemment, Seo *et al.* ont appliqué une stratégie omique (génome, transcriptome) pour analyser la souche *C. beijerinckii* BA105, une souche incapable de reconommer efficacement les acides qu'elle produit et sur-consommatrice de glucose. L'analyse a révélé des mutations sur un site de fixation Spo0A dans le promoteur d'un des six gènes codant putativement AbrB, un régulateur transcriptionnel d'importance chez *C. acetobutylicum* impliqué dans la transition entre phase exponentielle et phase stationnaire de croissance, ainsi qu'une mutation ponctuelle de la glucokinase (qui induit la substitution d'une valine en glycine en position 303), dont l'altération de l'activité peut jouer sur les mécanismes de RCC. D'un point de vue transcriptomique, la mutation en amont d'*abrB* se traduit par la surexpression du gène. Bien que le rôle exact des deux mutations n'ait pas été élucidé, le transcriptome est largement remanié et se caractérise par une modification d'expression importante des gènes régulés par AbrB chez *C. acetobutylicum* (expression forte et constante des voies de production des acides, et répression des voies de synthèse des solvants, notamment).

Le séquençage du génome de la souche *C. beijerinckii* DSM 6423 a permis d'identifier une structure génétique unique⁵². Outre son chromosome principal, assez semblable à celui de la souche NCIMB 8052, le génome de *C. beijeirinckii* DSM 6423 contient pas moins de trois éléments génétiques mobiles : deux plasmides et un bactériophage linéaire, qui n'ont toutefois pas de fonction claire identifiée. L'analyse génomique a permis d'identifier plus de 6000 gènes dont la régulation transcriptomique a ensuite été étudiée au cours d'une fermentation *batch*, similairement à ce qui a été fait sur la souche NCIMB 8052^{175,176}. L'analyse de l'évolution du transcriptome au cours de la fermentation a permis d'identifier des structures opéroniques similaires à celles de la souche 8052. L'analyse relève également des différences intéressantes avec la souche NCIMB 8052. En particulier, la comparaison suggère une régulation différente des gènes liés à la solvantogénèse, avec notamment une expression stable de l'opéron homologue de *ald-ctfA-ctfB-adc* (surexprimé seulement en début de solvantogénèse chez la souche 8052), et la présence d'un deuxième opéron *ctfA-ctfB-adc* dont l'expression est, elle, régulée. Ces différences peuvent potentiellement

conduire à des disparités en termes de comportement fermentaire, dont la plus évidente est la synthèse d'isopropanol plutôt que d'acétone.

Les outils transcriptomiques seuls ne sont cependant pas suffisants pour apprécier l'intégralité des phénomènes de régulations qui peuvent se dérouler dans un organisme vivant¹⁹⁷. Chez les bactéries, divers processus biologiques d'importance, comme la RCC¹¹⁴ ou la sporulation¹²⁸ pour les Firmicutes, ont été identifiés comme dépendants fortement de régulations post-transcriptionnelles. Parmi les autres stratégies à haut-débit envisageables, la protéomique permet d'étudier les régulations métaboliques à l'échelle protéique, en identifiant et quantifiant les protéines présentes à un instant t , ainsi qu'en fournissant une cartographie des modifications post-traductionnelles dans la cellule.

Mao *et al.* ont caractérisé dans deux études successives les protéomes cytoplasmiques et membranaires de *C. acetobutylicum* DSM 1731 et d'un mutant résistant au butanol (Rh8) par une approche d'électrophorèse sur gel de polyacrylamide (PAGE) couplée à une analyse par spectrométrie de masse MALDI-TOF^{198,199}. Leur étude cytoplasmique confirme le rôle des protéines chaperonnes dans la tolérance au butanol (GroEL/ES, notamment), et leur étude membranaire le rôle de certaines protéines impliquées dans la production de composés (peptidoglycane, lipides) de la membrane cellulaire. En effet, en modifiant la composition de leur membrane, les *Clostridia* solvantogènes pourraient compenser l'augmentation délétère de la fluidité de la bicouche lipidique induite par la présence de solvants²⁰⁰.

Malgré les informations importantes que l'étude du protéome peut fournir, cette approche est moins fréquente que la transcriptomique, probablement en raison d'une mise en œuvre plus lourde. Des techniques plus rapides permettent dorénavant d'augmenter sensiblement la vitesse d'analyse, en soumettant directement un échantillon protéique trypsinisé à un système LC-MS/MS (c'est-à-dire sans passer par des gels d'électrophorèse type PAGE). L'analyse des spectres de masse des peptides, après digestion par la trypsine, permet d'identifier les protéines présentes initialement dans le l'échantillon biologique. Cette technique a notamment été utilisée par l'équipe du Pr. Soucaille pour identifier et quantifier le protéome de *C. acetobutylicum* ATCC 824 dans trois états physiologiques différents¹⁷⁴. Dans cette étude, les transcriptomes ont également été établis, et l'intégration des deux types de données omiques a servi à l'amélioration d'un modèle métabolique à l'échelle du génome (GSM ou GSMM) du

fonctionnement de *C. acetobutylicum*. Idéalement, un tel modèle permet de simuler le comportement de l'organisme étudié *in silico*, et ainsi de prédire son comportement *in vivo* dans diverses conditions expérimentales. Même si à l'heure actuelle la précision de ces modèles est toujours limitée, leur développement, enrichi notamment par l'intégration de données multi-omiques, devrait certainement contribuer à l'amélioration de leur pouvoir prédictif^{201,202}, avec à terme un enjeu crucial pour l'ingénierie métabolique.

1.4.2 L'analyse des TSS, un nouvel outil d'analyse de la régulation de l'expression

Plusieurs techniques dérivées du RNA-seq ont permis, dans la dernière décennie, de s'intéresser plus particulièrement à l'initiation de la transcription et à sa régulation chez les bactéries.

1.4.2.1 Introduction

L'initiation de la transcription est une étape dont la régulation est essentielle pour le contrôle de l'expression génétique. De nombreux facteurs influent sur le positionnement des ARN polymérases le long du chromosome, et modulent ainsi leur capacité à démarrer la transcription.

Quels sont ces facteurs moléculaires majeurs qui participent à la régulation de l'expression génétique via la reconnaissance de sites spécifiques (TSS pour *transcription start sites*) par l'ARN polymérase ?

La littérature scientifique est, à ce sujet, très riche. Un article de synthèse récent de Browning et Busby²⁰³ brosse un tableau exhaustif des connaissances en la matière. Chez les procaryotes, deux processus biologiques majeurs sont principalement impliqués : la régulation par facteur sigma et via des facteurs de transcription. Le Tableau 4 présente quelques voies de régulation importantes liées à ces types de facteur chez *C. acetobutylicum* et *C. beijerinckii*.

Type	Nom	Fonction(s) principale(s)	Organisme solvantogène	Référence
Facteur sigma alternatif	σ^E	Sporulation, solvantogénèse	<i>C. acetobutylicum</i>	Tracy et al. 2011 ¹⁹⁴
	σ^F	Sporulation, solvantogénèse	<i>C. acetobutylicum</i>	Jones et al. 2011 ²⁰⁴
	σ^G	Sporulation	<i>C. acetobutylicum</i>	Tracy et al. 2011 ¹⁹⁴
	σ^H	Sporulation, solvantogénèse ?	<i>C. acetobutylicum</i>	Jones et al. 2008 ²⁰⁵
	σ^K	Sporulation, solvantogénèse	<i>C. acetobutylicum</i>	Al Hinai et al. 2014 ²⁰⁶
	σ^{54}	Conso. sucres alternatifs	<i>C. acetobutylicum</i>	Nie et al. 2016 ²⁰⁷
Facteur de transcription	CcpA	Conso. sucres alternatifs, sporulation, solvantogénèse	<i>C. acetobutylicum</i>	Ren et al. 2012 ²⁰⁸
	Spo0A	Sporulation, solvantogénèse	<i>C. beijerinckii</i>	Ravagnagni et al. 2002 ²⁰⁹
	AbrB	Sporulation, solvantogénèse	<i>C. acetobutylicum</i>	Scotcher et al. 2005 ²¹⁰ Xue et al. 2016 ²¹¹
	PerR	Réponse au stress oxydatif	<i>C. acetobutylicum</i>	Hillman et al. 2009 ²¹²
	Rex	Equilibre redox	<i>C. acetobutylicum</i>	Wietzke et Bahl 2012 ¹⁴⁹

Tableau 4 : Liste de facteurs impliqués dans la régulation de l'initiation de la transcription de gènes importants pour le cycle de vie des *Clostridia* solvantogènes.

1.4.2.1.1 Facteurs sigma

La première voie de régulation consiste à diriger l'ARN polymérase vers un sous-ensemble de promoteurs donnés²⁰³. Le cœur de cette enzyme est constitué de plusieurs sous-unités protéiques : deux sous-unités α , une sous-unité β et une β' , une sous-unité ω . Ce cœur ne peut reconnaître l'ADN et initier la transcription qu'en formant une holoenzyme en se liant à une dernière sous-unité, le facteur σ . Chez les bactéries, une multitude de facteurs sigma co-existent et permettent de former différentes holoenzymes capables de reconnaître des promoteurs spécifiques et de transcrire les gènes qu'ils contrôlent. Un facteur sigma prédominant (ou de ménage, c'est-à-dire fortement exprimé dans la plupart des conditions expérimentales, comme par exemple σ^{70} chez *E. coli* ou σ^A chez les *Clostridia*) permet de transcrire la majorité des gènes, qui présentent une combinaison de séquences en amont du TSS reconnues par l'holoenzyme formée avec ce facteur sigma. Cette combinaison de

séquences définit la force du promoteur pour chaque gène en jouant sur l'affinité ADN-holoenzyme. Cette dernière reconnaît en effet préférentiellement une séquence particulière, appelée séquence consensus, et tout changement vis-à-vis de cette dernière se traduit par une diminution plus ou moins importante de sa reconnaissance par l'holoenzyme.

D'autres facteurs sigma, dits alternatifs, peuvent remplacer une proportion significative des facteurs sigma de ménage et former de nouvelles holoenzymes, capables de reconnaître des séquences nucléiques différentes. Ce type de régulation permet d'orienter la machinerie transcriptionnelle vers un sous-ensemble de gènes particuliers, généralement dans le but de répondre à un stimulus environnemental. Un tel contrôle est souvent affiné par un arsenal de moyens permettant de réguler précisément la concentration cellulaire et/ou l'activité de chaque facteur sigma alternatif (comme par exemple en précipitant la dégradation d'un facteur sigma donné, ou encore en le séquestrant avec une autre protéine de type « anti-sigma »). Chez les *Clostridia* solvantogènes, la régulation par facteur sigma a été principalement documentée pour caractériser les processus de sporulation¹²⁸ (cf. Tableau 4). La sporulation étant liée à la solvantogénèse (cf. sous-partie 1.3.4), cette régulation concerne également la production de solvants. L'activation de la sporulation est orchestrée en cascade par une variété de facteurs sigma et de facteur de transcriptions, qui permettent à chacune des différentes étapes de se dérouler correctement. Si la solvantogénèse est affectée par certains facteurs de sporulation, aucun facteur sigma régulant spécifiquement la solvantogénèse n'a toutefois été identifié.

La régulation par facteur sigma concerne également la consommation de sucres chez *C. acetobutylicum*²⁰⁷. La consommation du cellobiose est ainsi spécifiquement contrôlée par σ^{54} (également σ^L), un facteur sigma alternatif très particulier, qui n'a aucune homologie de séquence avec l'ensemble des autres facteurs sigma²¹³ (qui présentent des homologies structurales avec le facteur de ménage σ^{70}). Ce facteur se distingue par un mode de fonctionnement atypique : l'holoenzyme comprenant σ^{54} est incapable de démarrer la transcription seule, même après s'être fixée à son promoteur cible. Un activateur transcriptionnel (EBP, pour *Enhancer Binding Protein*) est nécessaire pour fournir l'activité ATPase nécessaire à l'initiation de la transcription (transition d'un état conformationnel fermé à ouvert pour

l'holoenzyme). Cet activateur reconnaît généralement une séquence d'ADN en amont du site de fixation de l'holoenzyme, ce qui permet l'activation de cette dernière. L'activité de chaque EBP (faculté à se lier à l'ADN et/ou à hydrolyser l'ATP) est elle-même régulée, généralement pour permettre une réponse ciblée à un stimulus environnemental.

En résumé, la régulation par facteur sigma est un mode privilégié de régulation de l'expression génétique chez les bactéries. Chez les solvantogènes, cette voie régule des processus biologiques importants (sporulation, solvantogenèse, assimilation de sucres alternatifs), bien qu'accessoires.

1.4.2.1.2 Facteurs de transcription

La reconnaissance du promoteur bactérien par l'ARN polymérase peut être significativement modulée par des facteurs de transcription²⁰³. Ces facteurs se lient à l'ADN dans la région promotrice (autour du promoteur, voire directement sur le promoteur) et activent ou répriment l'initiation de la transcription (on parle respectivement d'activateurs et de répresseurs). Cette régulation suit différents modes de fonctionnement ; la répression par encombrement stérique ou l'activation par augmentation de l'affinité pour le promoteur avec un des domaines de l'ARN polymérase sont les modes d'action les plus courants. Chez les solvantogènes, quelques facteurs de transcription ont été caractérisés (cf. Tableau 4). Comme pour les facteurs sigma, la sporulation, la solvantogenèse et la RCC ont été majoritairement associés à ces voies de régulation. Il n'est toutefois pas exclu (et même probable) que l'identification de ces procédés biologiques découle d'un parti pris dans la recherche sur les bactéries solvantogènes (c'est-à-dire que la majeure partie de la recherche se concentre sur ces procédés), et que de nombreux autres mécanismes moléculaires puissent être régulés au stade transcriptionnel par des facteurs de transcription et/ou des facteurs sigma non caractérisés.

1.4.2.2 Méthodes de caractérisation à haut débit des TSS

Le RNA-seq est une technique relativement facile d'utilisation et très informative, ce qui a largement contribué à sa rapide diffusion dans la communauté

scientifique²¹⁴. De nombreuses méthodes de séquençage dérivées du RNA-seq ont également vu le jour, et permettent par exemple d'étudier en détail les interactions ARN-protéines (CLIP-seq), l'état traductionnel de la cellule (Ribo-seq) ou encore la vitesse de synthèse et de dégradation des ARN *in vivo* (4sU-seq). Cette adaptabilité de la procédure de préparation des bibliothèques a permis, dans les dix dernières années, de développer plusieurs approches ciblant spécifiquement les sites d'initiation de la transcription bactérienne.

Sharma et collègues ont été les pionniers de cette exploration en développant le dRNA-seq (*differential* RNA-seq) pour une analyse des TSS chez *Helicobacter pylori*²¹⁵. Cette méthode se base sur une particularité des ARNm prokaryotes : le premier nucléotide de leurs extrémités 5' est un nucléotide triphosphate, ce qui les différencie de la plupart des autres ARN, et en particulier les ARN ribosomiaux et de transfert, qui, après maturation, commencent par un nucléotide monophosphate. Par ailleurs, les ARNm dégradés en 5' ne possèdent plus cette extrémité 5' triphosphate. Ainsi, ces ARN peuvent être spécifiquement digérés par une exonucléase spécifique des extrémités 5' monophosphates (TEX, pour *terminal exonuclease*), ce qui permet d'enrichir la bibliothèque en ARNm intacts possédant des extrémités 5' triphosphates dans le but de les séquencer. Chez *H. pylori*, l'identification des TSS suggère l'existence de multiples formes de régulation autrement inaccessibles par l'utilisation d'un RNA-seq conventionnel. L'analyse des TSS a permis de dévoiler une organisation opéronique complexe (87,5% des gènes regroupés en opérons), ainsi qu'une dynamique étonnante de la machinerie d'initiation transcriptionnelle, qui, dans des conditions expérimentales différentes, favorise un positionnement différentiel des TSS pris en charge par l'ARN polymérase. Cette flexibilité semble être un mécanisme important d'adaptation du microorganisme à l'environnement. A pH faible (c'est-à-dire en mimant les conditions de vie du microorganisme, dans l'estomac), les auteurs observent par exemple l'activation de TSS situés au sein d'opérons, permettant de sur-exprimer certains gènes liés à la pathogénicité de cet organisme. Par ailleurs, presque la moitié des TSS sont détectés dans le sens opposé des gènes avec lesquels ils ont été associés. Cela suggère une importance particulière de la transcription antisens pour la régulation de la transcription, avec un impact potentiel sur la dégradation des ARN double brins (dus à la complémentarité sens/antisens), rendue possible par la RNase

III de cet organisme. Suite à cette première étude, le dRNA-seq a été réutilisé de nombreuses fois, pour plusieurs espèces bactériennes²¹⁶.

Malgré une relative facilité de mise en œuvre, le dRNA-seq souffre toutefois d'un inconvénient technique. L'action de la TEX est en effet incomplète, et l'identification des TSS doit de fait passer par un double séquençage, avec et sans traitement TEX (TEX⁺ et TEX⁻), et par une mesure de l'enrichissement des lectures mesurées dans les librairies TEX⁺. Cette limitation a conduit au développement d'une nouvelle technique, le *Cappable-seq*, qui, au lieu de dégrader les ARN monophosphates qui sont majoritaires, vise à spécifiquement purifier les ARN triphosphates²¹⁷. Cette purification passe par l'utilisation d'une enzyme (VCE, pour *Vaccinia Capping Enzyme*) qui incorpore sur les ARN triphosphate un résidu guanidine modifié avec une desthiobiotine. Cette propriété permet de purifier sélectivement ces ARN par interaction desthiobiotine/streptavidine, et se traduit par l'obtention de librairies pour lesquelles une condition contrôle (similaire en termes d'utilisation à la librairie TEX⁻) n'est pas nécessaire. Chez *E. coli*, cette technique a permis d'identifier plus de 16000 TSS, dont plus de 6000 inconnus jusqu'alors. Ettwiller et collègues apportent également la preuve de concept du *Cappable-seq* en déterminant directement la position de plusieurs centaines de TSS dans un microbiote intestinal de souris.

Plus récemment, Boutard *et al.* ont adapté le *Cappable-seq* pour simplifier la procédure de préparation des librairies²¹⁸. L'innovation réside principalement dans l'utilisation d'une transcriptase reverse capable de faire du *template switching*²¹⁹, ce qui permet d'éviter une étape de ligation d'une des amorces universelles de séquençage, et par conséquent de produire l'ADNc directement sur les billes de streptavidine (sans éluer les ARN, Figure 8). Les auteurs nomment cette approche *Capp-switch sequencing*, et l'appliquent à *Clostridium phytofermantans*, une espèce solvantogène particulière, incapable de produire des alcools supérieurs mais qui dispose d'un arsenal protéique adapté à la digestion de biomasse lignocellulosique (substrat de type 2G).

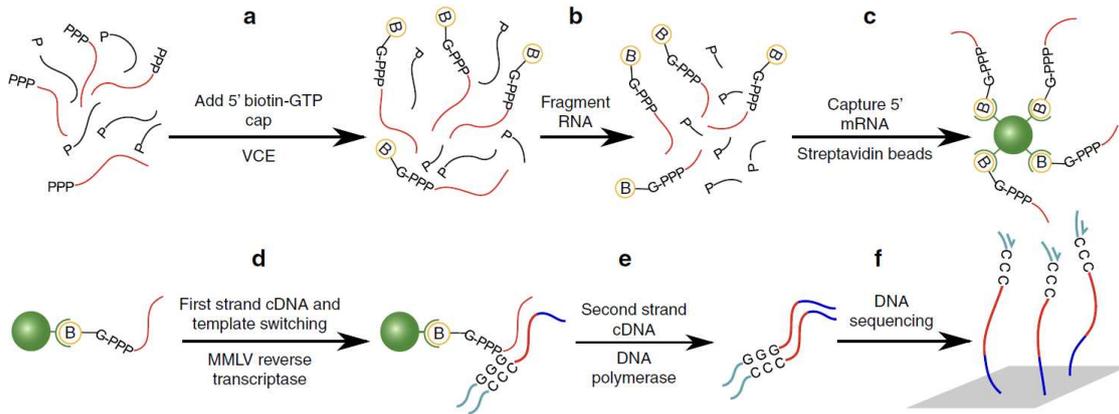


Figure 8 : Procédure de préparation des bibliothèques pour le *Capp-switch sequencing*²¹⁸.

Parmi les résultats principaux de l'étude, il s'avère que la répartition des TSS à l'échelle du transcriptome est considérablement modulée par la source de carbone, qui peut faire varier la position, la force, et le nombre de TSS pour un gène donné. C'est particulièrement le cas pour les gènes directement impliqués dans le métabolisme des sources carbonées concernées, qui en général se retrouvent également sur-exprimés dans ces conditions.

Contrairement à *H. pylori*, il semblerait que la transcription antisens ne soit pas un mécanisme de régulation aussi important chez les *Clostridia*, avec moins d'1% des TSS assignés à une initiation antisens. En revanche, les quelques événements antisens détectés par la méthode ne sont probablement pas anecdotiques. Un TSS antisens est par exemple identifié dans le 5'UTR du gène *spo0A*, et qui s'oppose également à la transcription du gène *spoIVB*, un autre facteur de sporulation. Ces TSS ne sont détectés que lorsque la souche est cultivée sur des polysaccharides, ce qui suggère qu'en présence d'une quantité abondante de sucres simples, ces gènes, fortement impliqués dans le processus de sporulation, sont réprimés par ce mécanisme transcriptionnel.

Si la transcription antisens est limitée, il semblerait que la transcription sens intragénique ait un rôle plus prononcé. Plus de 50% des TSS détectés appartiennent à cette catégorie et, bien que chacun contienne en moyenne moins de lectures qu'un TSS canonique, il est tout à fait possible qu'une transcription permissive ait un rôle important à jouer dans la régulation transcriptionnelle (production d'isoformes protéiques différentes, protection de l'ARNm, ou autre). En analysant les motifs

associés à ce type de TSS, les auteurs montrent que la transcription permissive serait due à un motif riche en AT.

L'analyse des motifs permet ainsi d'identifier, au moins en partie, les mécanismes transcriptionnels qui favorisent l'initiation de la transcription à un site plutôt qu'à un autre. La séquence du promoteur, et en particulier de certaines régions de ce dernier, est en effet déterminante pour l'expression²⁰³. L'analyse de séquences communes à des catégories de gènes permet d'imaginer une régulation de l'expression de ces gènes via un même facteur transcriptionnel (Figure 9). Dans le cas de *C. phytofermantans*, les motifs identifiés en amont ou en aval des TSS identifiés uniquement sur un substrat donné révèlent un mécanisme de régulation dépendant de la source de carbone. Les TSS dépendants de la présence d'acide galacturonique et de l'homogalacturonane présentent par exemple dans leur voisinage un motif très proche du motif reconnu par la protéine CcpA de *B. subtilis*, impliquée dans la RCC. Les auteurs ont ainsi identifié un gène codant un régulateur transcriptionnel (Cphy2742) contenant trois de ces motifs, qui voit son expression augmenter en présence de ces substrats, ce qui suggère un rôle possiblement important pour la dégradation des deux sources carbonées correspondantes.

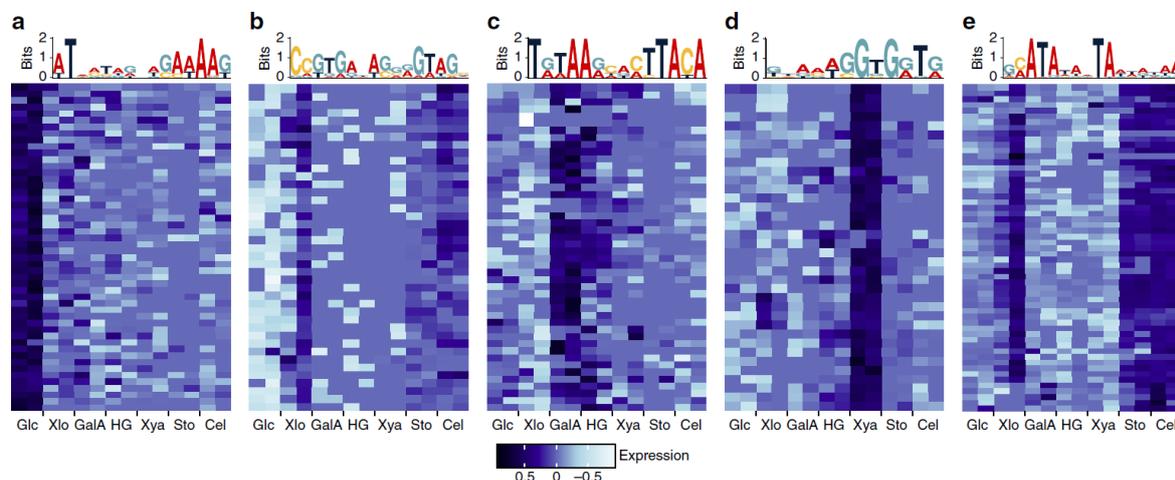


Figure 9 : Analyse transcriptomique des motifs impliqués dans l'expression différentielle des TSS pour chaque substrat testé, chez *C. phytofermantans*²¹⁸.
 Glc : glucose ; Xlo : xylose ; GalA : acide galacturonique ; HG : homogalacturonane (pectine de pomme) ; Xya : xylane (bouleau) ; Sto : canne de maïs ; Cel : cellulose.

Comme le montre la découverte de nombreux éléments non codants dans *H. pylori*²¹⁵, l'analyse des TSS permet également de corriger l'annotation du génome, en ajoutant des caractéristiques génomiques jusque-là non identifiées, ou en mettant à

jour l'annotation automatique des gènes. Chez *C. phytofermantans*, l'exploration des TSS a également permis de suggérer la présence de plus de 700 nouvelles unités de transcription, dont 60 ont été validées par une analyse protéomique. Cet aspect exploratoire, combiné à l'appréciation des mécanismes de régulation transcriptionnelle que ce type de méthode permet, fait du *Cappable-seq* et du *Capp-switch sequencing* des outils particulièrement performants pour le décryptage génétique d'espèces peu étudiées. Par ailleurs, ces méthodes peuvent être couplées à des techniques de séquençage permettant le séquençage de long fragments, ce qui permet, en plus d'identifier les sites d'initiation de la transcription, de valider ou découvrir de nouvelles structures opéroniques et de déterminer les sites de terminaison de la transcription²²⁰.

1.5 OPTIMISATION GENETIQUE DES MICROORGANISMES SOLVANTOGENES DE GENRE CLOSTRIDIUM

1.5.1 Généralités

L'intérêt renouvelé pour les souches solvantogènes A/IBE dans les deux dernières décennies a permis de faire émerger des pistes nouvelles pour améliorer génétiquement ces organismes, et en particulier les souches de *C. acetobutylicum*. Si cet organisme n'est toujours pas aussi facilement manipulable que *E. coli* et *B. subtilis*, il bénéficie désormais d'une boîte à outils génétiques variés, qui se base sur un certain nombre de plasmides navettes, de marqueurs de sélection et de contre-sélection, de promoteurs inductibles et de gènes rapporteurs²²¹.

Les difficultés principales de l'utilisation des solvantogènes comme châssis biologique pour le génie métabolique sont liées à la présence de barrières physiques et génétiques importantes (systèmes complexes de restriction/modification, forte activité des nucléases cellulaires, membranes cellulaires externes épaisses et peu pénétrables, faible propension à la recombinaison homologe). L'entrée de matériel génétique dans la bactérie est une des étapes essentielles aux procédures de mutagenèse ciblée (cf. sous-partie suivante), et elle n'est maîtrisée que pour un nombre limité de souches solvantogènes²²¹.

Le système de restriction/modification (RM) endogène de *C. acetobutylicum* ATCC 824 a été contourné il a plusieurs décennies dans une étude rapportée par Mermelstein *et al.*²²² La méthylation de l'ADN exogène par la méthylase du phage de *B. subtilis* ϕ 3T1 s'est révélée être un élément clé pour son introduction dans *C. acetobutylicum* ATCC 824. Au contraire, l'ADN ne semble pas requérir de prétraitement particulier pour sa transformation chez *C. beijerinckii* NCIMB 8052. Ce n'est cependant pas le cas pour d'autres souches de la même espèce. En particulier, une étude récente²²³ a démontré que la transformation de *C. beijerinckii* NRRL B-598 (anciennement désignée de manière erronée comme *C. pasteurianum*²²⁴) nécessite l'utilisation d'un ADN non méthylé, qui ne peut donc pas être préparé à partir des cellules d'*E. coli* couramment utilisées en construction génétique et qui réalisent des méthylations de type Dam et Dcm. Avant le début de ces travaux de thèse, une étude interne IFPEN s'est inspirée de ce résultat et l'a reproduit pour la souche *C. beijerinckii* DSM 6423. Il s'agissait là d'une étape préliminaire cruciale puisqu'elle permet d'envisager une approche génétique pour améliorer cet organisme.

1.5.2 Les outils classiques de modification du génome

1.5.2.1 Mutagénèse aléatoire

La mutagénèse aléatoire s'attache à la compréhension d'un phénotype par l'étude de son génotype. Classiquement, une librairie de mutants est générée par exposition d'un organisme biologique à des agents mutagènes, et une phase importante de sélection permet d'obtenir des clones présentant un phénotype souhaité. L'analyse de leur génome permet d'en déduire les gènes dont les mutations sont potentiellement responsables du phénotype observé.

Cette méthode présente l'avantage non négligeable pour les espèces solvantogènes du genre *Clostridium* de ne pas nécessiter d'outil génétique particulier. Les premières études fructueuses de ce type d'approche chez *C. beijerinckii* remontent aux années 80. La souche BA101 hyper-productrice de butanol a été obtenue par exposition de *C. beijerinckii* NCIMB 8052 à un agent chimique mutagène, le N-méthyl-N'-N-nitrosoguanidine (NTG), puis à la sélection de mutants en présence d'un analogue non métabolisable du glucose (2-deoxyglucose)²²⁵. Bien que les mutations responsables n'aient pas été caractérisées, l'amélioration de la souche

serait due, au moins en partie, à l'amélioration des activités acétoacétate décarboxylase, CoA transférase et butyraldéhyde déshydrogénase du métabolisme central⁶⁸.

Outre la nécessité de cribler un nombre important de clones, la principale limitation de ce type d'approche aléatoire réside dans le décalage entre le phénotype désiré – généralement un accroissement de la production de solvants – et la pression de sélection exerçable. Ce décalage induit une difficulté pour l'utilisateur dans le choix critique de la pression de sélection, qui ne conduira pas nécessairement au résultat escompté. Une étude récente illustre parfaitement ce phénomène : en couplant l'utilisation de NTG au *genome shuffling* (une technique permettant l'échange de patrimoines génétiques de deux souches différentes, voire de deux espèces différentes) chez *C. beijerinckii* DSM 6423, associée à une sélection sur des concentrations toxiques d'isopropanol, Máté de Gerando *et al.* ont obtenu des clones fortement résistants à leur agent sélectif, mais avec une productivité en isopropanol à peine augmentée⁴⁷. Ce résultat est probablement dû au fait que les mécanismes de tolérance aux solvants et les mécanismes de leur production ne sont pas forcément liés^{83,224}, ainsi qu'au fait que la synthèse d'isopropanol n'est probablement pas limitante pour la production de solvants dans cet organisme.

A l'inverse, les auteurs décrivent l'obtention d'un clone résistant à l'allyl alcool (clone AA), qui présente une capacité de production de solvants fortement diminuée. L'allyl alcool est un produit qui peut servir de substrat aux alcool déshydrogénases, qui l'utilisent pour produire de l'acroléine, une molécule hautement toxique. La sélection par mutagenèse aléatoire se traduit par l'obtention de clones défectifs en activité alcool déshydrogénase²²⁶, ce qui explique le phénotype observé. Dans cette étude, la caractérisation génétique du mutant n'a toutefois pas été faite, ce qui aurait pu permettre de déterminer les facteurs moléculaires responsables de ce phénotype chez *C. beijerinckii*.

1.5.2.2 Mutagenèse ciblée

La mutagenèse ciblée, ou dirigée, consiste à modifier de manière précise le génome, dans le but d'obtenir un phénotype désiré.

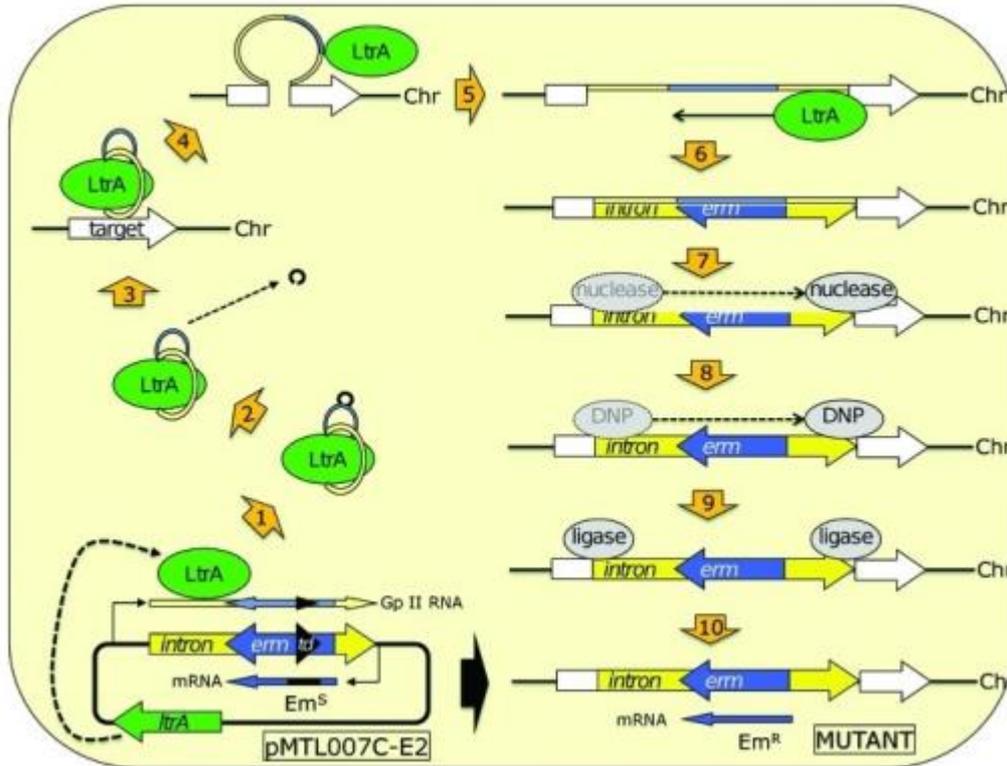


Figure 10: Description schématique de l'utilisation de la méthode Clostron pour l'inactivation de gène chez les *Clostridia*²³⁴.

1 : formation du complexe intron / IEP, à partir d'un plasmide, qui ne confère pas de résistance à l'érythromycine
 2 : auto-épissage de l'intron de groupe 1 présent dans le gène de résistance à l'érythromycine, et qui inactive ce dernier.
 3 : reconnaissance du locus génomique à inactiver par l'intron.
 4-10 : insertion de l'intron dans la cible génétique, inactivation du gène et expression correcte du gène de résistance à l'érythromycine, permettant la sélection du mutant.

Historiquement, l'utilisation d'introns mobiles de groupe II a été une première approche largement utilisée chez les solvantogènes modèles et qui a permis, en intégrant une séquence parasite à un locus précis, d'inactiver un gène cible. Le fonctionnement *in vivo* de ces introns mobiles de groupe II a été longuement explicité²²⁷. Succinctement, il s'agit d'ARN doués d'activité catalytique, capables de s'insérer dans une séquence d'ADN précise par un mécanisme de rétro-transposition, et d'être épissés après transcription, rétablissant ainsi la fonction du gène dans lequel l'intron s'était inséré. La séquence d'insertion est à la fois dépendante de la complémentarité d'une partie de la séquence de l'intron avec le site cible et de l'expression d'une protéine exogène, dont l'ORF est contenue dans l'intron. Cette protéine, l'*Intron-Encoded Protein* (IEP), permet à la fois l'insertion de l'intron et son épissage post-transcriptionnel. Pour un usage en génie génétique, l'ORF codant l'IEP peut être retiré de l'intron et placé sur le plasmide à un autre locus. Après rétrotransposition dans le gène ciblé, la perte du plasmide induit l'incapacité pour la

bactérie d'exciser l'intron, ce qui aboutit à une interruption permanente du gène visé, l'intron contenant des codons stop prématurés dans tous les cadres de lecture. Cette méthode d'inactivation de gènes, appelée commercialement *TargeTron*, a été utilisée avec succès chez plusieurs solvantogènes^{67,70,228-233}.

La principale limitation du *TargeTron* réside dans la variabilité des fréquences d'insertion dans le génome des *Clostridia*, qui dépend largement du locus visé. Ce système a été amélioré par Heap *et al.* pour inclure dans l'intron un ORF codant pour un gène de résistance à l'érythromycine, lui-même interrompu par un intron mobile de groupe I²³⁵ (Figure 10). Ce type d'intron est capable de s'auto-épisser, et ce uniquement s'il est transcrit dans une orientation donnée. Cette orientation est incorrecte sur le plasmide contenant l'intron, qui ne confère donc pas de résistance à l'érythromycine, et devient correcte après *retrohoming*. L'évènement d'insertion de l'intron dans le génome devient par ce système facilement sélectionnable sur milieu supplémenté en érythromycine. La méthode a par la suite été améliorée pour inclure un système permettant l'excision du marqueur de sélection du chromosome, et permettre ainsi d'inactiver successivement et rapidement plusieurs gènes dans la même souche²³⁶. Malgré ces efforts, la technique *TargeTron* et ses dérivées restent limitées sur plusieurs points²²¹ et ce système est essentiellement resté cantonné à l'inactivation de gène (même si Heap *et al.* arrivent à démontrer une insertion maximale de 1 kb chez *C. sporogenes*²³⁶).

Pour contourner ces restrictions, quelques équipes de recherche se sont tournées vers l'échange allélique par recombinaison homologue, qui permet de modifier le génome « proprement » en réalisant les insertions / délétions souhaitées à l'acide nucléique près. La base du système consiste à apporter, sous la forme d'un plasmide instable, les modifications à apporter au gène ciblé, entouré de zones homologues aux séquences naturellement présentes autour de ce gène. Grâce à la machinerie cellulaire endogène, un premier croisement par recombinaison homologue entre une région identique entre le plasmide et le génome peut alors s'effectuer, qui aboutit à l'intégration dudit plasmide dans le génome. Cet évènement est facilement sélectionnable par l'ajout d'un marqueur de résistance dans le plasmide. Ce dernier étant faiblement répliqué, la résistance n'est exprimée de façon stable qu'après intégration. Le deuxième évènement de recombinaison homologue permet l'excision du plasmide. En fonction des régions d'homologie impliquées, la souche obtenue aura

un génotype sauvage ou muté. Dans le deuxième cas, une modification stable est obtenue, dans un génome dépourvu de marqueur de sélection. Ce deuxième cas peut cependant être difficile à isoler, de sorte que l'obtention d'un clone correct peut prendre énormément de temps, voire ne pas aboutir du tout²²¹.

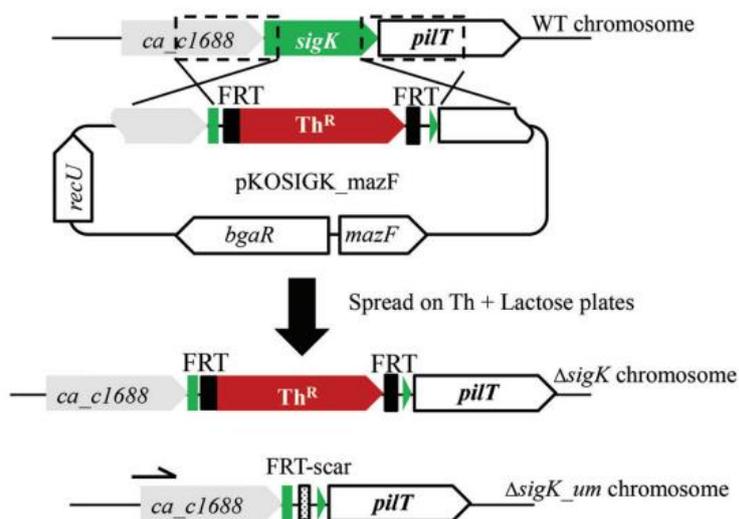


Figure 11 : **Stratégie d'échange allélique développée par Al-Hinai et al.**²³⁷

La délétion du gène *sigK* est réalisée dans cet exemple. Des zones d'homologies en amont et en aval du gène de résistance au thiamphénicol (*Th^R*), lui-même entouré de sites FRT, permettent l'échange allélique. La sélection est effectuée en présence de thiamphénicol et de lactose, simultanément ou successivement. La présence de lactose induit l'expression du gène *mazF*, hautement toxique, ce qui permet de sélectionner l'absence de plasmide, sous forme libre ou intégrée dans le génome par simple recombinaison homologue. Le gène *Th^R* est ensuite enlevé dans une seconde phase par l'expression exogène de la *Flp* recombinase, qui ce faisant laisse une cicatrice dans le génome.

Pour surmonter cette difficulté, il n'est pas rare d'utiliser un voire plusieurs marqueur(s) de contre-sélection. Dans une méthode brevetée en 2006²³⁸, deux gènes de résistance sont utilisés, dont un des deux (*catP*, résistance aux amphénicol) est placé entre les deux bras d'homologie, et entouré de séquences FRT (*Flp recognition target*). L'insertion du plasmide dans le chromosome via recombinaison homologue peut être sélectionnée en recherchant les clones résistants au thiamphénicol. Cette méthode est couplée à l'utilisation d'un marqueur de contre-sélection, le gène *upp* (uridine phosphoribosyltransférase). Ce gène, endogène de *C. acetobutylicum* ATCC 824, confère une sensibilité au 5-fluorouracile (5-FU), en permettant sa métabolisation et transformation en analogues nucléotidiques hautement toxiques. En plaçant *upp* sur le plasmide, en dehors de la région contenant les zones d'homologie et le gène *catP*, l'utilisation de 5-FU permet ainsi de sélectionner rapidement les cellules qui ont perdu le plasmide dans une souche Δupp , insensible au 5-FU en l'absence du plasmide. Le

gène de résistance au chloramphénicol peut ensuite être excisé par l'utilisation de la recombinase Flp. Similairement, Al-Hinai *et al.* ont développé un système utilisant une toxine inductible au lactose comme marqueur de contre-sélection, avec également une recombinaison Flp pour enlever le gène de résistance au thiamphénicol²³⁷ (Figure 11).

Malgré la flexibilité que ce type de stratégie apporte par rapport aux introns mobiles de groupe II, les méthodes développées par ces approches ne sont pour autant pas parfaites. L'utilisation de la recombinase Flp laisse un des deux sites FRT dans le génome (environ 80 bases), une cicatrice qui peut notamment avoir un effet polaire. Yoo *et al.* ont récemment rapporté que la présence de telles cicatrices FRT dans le génome perturbait probablement la transcription, en provoquant un arrêt prématuré de l'ARN polymérase¹⁷³. Le gène *adhE1*, qu'ils tentaient de supprimer, est le premier gène de l'opéron *sol* qui comprend également les gènes *ctfA* et *ctfB*. Ainsi, les auteurs ont observé que l'expression des gènes successifs au sein de l'opéron était perturbée, même après excision du marqueur de sélection. Ils attribuent cet effet à la structure secondaire du site FRT, qui, en adoptant une conformation de type tige-boucle, empêcherait l'ARN polymérase de transcrire efficacement les deux gènes suivants de l'opéron.

Ce problème potentiel de cicatrice résiduelle, laissée par le génie génétique, a été abordé par une équipe qui a réussi à utiliser avec succès l'échange allélique pour générer des mutants sans cicatrices ni marqueurs (méthode ACE²³⁹). Les auteurs exploitent un gène particulier, le gène *pyrE*, qui chez *C. acetobutylicum* est nécessaire pour synthétiser l'uracile. La délétion de ce marqueur induit une auxotrophie pour l'uracile, mais provoque également une résistance à l'acide 5-fluoroacétique, dont la métabolisation dans la souche sauvage via PyrE est toxique. En alternant l'état de *pyrE* (actif vs inactif), Heap *et al.* démontrent le potentiel de leur méthode en insérant séquentiellement, sans marqueur ni cicatrice, plus de 40 kb d'ADN du phage lambda, ce qui constitue la plus grosse insertion effectuée à ce jour dans cette espèce. Leur méthode, qui exploite pour l'expression de PyrE des promoteurs endogènes du génome, est toutefois tributaire d'une expression forte du gène ciblé, et ne permet de faire d'insertions que sur un nombre limité de loci dans le génome (c'est-à-dire à la suite de promoteurs endogènes).

Malgré l'aide apportée par la contre-sélection, le processus d'échange allélique standard, à multiples étapes, reste long et laborieux, et laisse dans la plupart des cas

des traces dans l'organisme modifié qui peuvent potentiellement avoir un effet polaire. Par ailleurs, la détermination d'un gène efficace, endogène ou exogène, de contre-sélection peut se révéler une tâche difficile, qui doit être validée pour toute nouvelle souche²⁴⁰. Pour toutes ces raisons les méthodes basées sur l'échange allélique bénéficient grandement des améliorations apportées par une technique génétique récente, rapide, et quasi-universelle : CRISPR-Cas9.

1.5.3 La révolution CRISPR

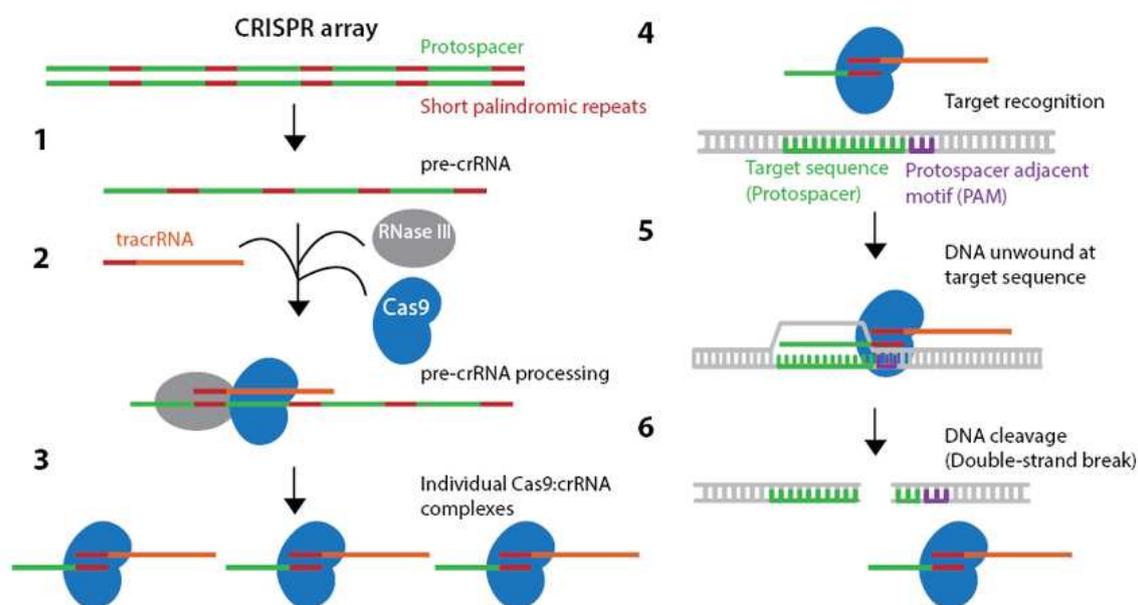


Figure 12 : **Description schématique du fonctionnement du système CRISPR chez *S. pyogenes* in vivo²⁴².** L'ADN viral (protospacer), entrecoupé de séquences répétées (short palindromic repeats) est stocké dans le génome à un locus particulier, le CRISPR array. Après transcription de ce locus (1), le pré-ARNcr qui en résulte subit une maturation qui fait intervenir l'ARNcrtr, la RNase III et Cas9 (2), qui permet, pour chaque protospacer, de former un complexe Cas9-ARNcr-ARNcrtr (3). La reconnaissance de la cible (c'est-à-dire un bactériophage) est réalisée par association de bases entre le protospacer et l'ADN viral. Cette reconnaissance est dépendante de la présence d'une séquence adjacente (PAM) (4). L'ADN reconnu est ensuite clivé (5,6).

Le système CRISPR (*clustered regularly interspaced palindromic repeats*) est *in vivo* impliqué dans l'immunité acquise bactérienne, et plus particulièrement dans la guerre incessante qui fait rage entre bactériophages et bactéries²⁴¹. Pendant une infection virale, un ensemble de protéines Cas (*CRISPR associated*) catalysent l'intégration dans le génome de fragments d'ADN étranger, et utilisent cette information

pour reconnaître et dégrader par la suite toute séquence comportant le ou les motif(s) intégré(s). Chez *Streptococcus pyogenes*, la nucléase Cas9 est l'acteur final de ce système, qui s'associe avec deux molécules d'ARN pour cliver, et ainsi inactiver, l'ADN viral. La première, appelée ARNcr (ARN CRISPR), comporte la séquence (*protospacer*) qui permet la reconnaissance par complémentarité de base avec l'ADN étranger, ainsi qu'une autre séquence nécessaire au déroulement de l'ADN chromosomique pour permettre cette reconnaissance (le *protospacer adjacent motif*, PAM). La deuxième, l'ARNcrtr (ARN CRISPR trans-encodé) est nécessaire pour la maturation de l'ARNcr et a également un rôle structural important pour le maintien de l'interaction ARNcr / Cas9 (Figure 12).

Jinek *et al.* établissent une modification simple qui facilite grandement l'utilisation de cette outil²⁴³. Ils fusionnent l'ARNcr et l'ARNcrtr en une seule molécule, l'ARN guide (ARNg) et démontrent que cet ARN guide est suffisant pour amener Cas9 à cliver sa cible *in vitro* avec une grande efficacité. Le potentiel de la méthode tient principalement en sa reprogrammation facile : en changeant la séquence de l'ARNg, la cible génétique est modifiée en conséquence. Très rapidement, Mali *et al.*²⁴⁴ et Cong *et al.*²⁴⁵ réutilisent ces travaux pour éditer des génomes de cellules de mammifères. Dans un système eucaryote, deux stratégies sont principalement utilisées, qui exploitent toutes les deux les mécanismes de réparation cellulaire mis en place par l'organisme :

- la réparation NHEJ (*non-homologous end joining*) qui répare l'ADN ciblé en raboutant les deux morceaux clivés. Cette réparation n'est pas parfaite, et peut aboutir à des erreurs. La force du système est qu'elle sélectionne l'erreur puisqu'en cas de réparation correcte, la séquence d'ADN reconnue par le complexe Cas9/gRNA est restaurée, ce qui provoque une nouvelle cassure. Le cycle se répète ainsi jusqu'à l'apparition de mutations, sous la forme d'insertions ou de délétions, qui provoquent dans la majorité des cas l'inactivation du gène ciblé.
- le processus de réparation HDR (*homology directed repair*) qui restaure l'intégrité d'une cassure d'ADN en utilisant l'autre allèle du gène concerné. Cette matrice de réparation, dans le cas d'une modification génétique réalisée avec CRISPR-Cas9, restaure également la cible de découpage du système, qui est à nouveau clivée. Dans ce cas de figure, l'utilisateur doit amener concomitamment une matrice de réparation comportant des zones d'homologies, ainsi que les modifications

désirées, comprenant également une altération de la séquence reconnue par Cas9. Ceci permet de ne pas se limiter à l'inactivation de gène, car il est possible d'apporter toute sorte de modifications dans le génome de la cible (Figure 13).

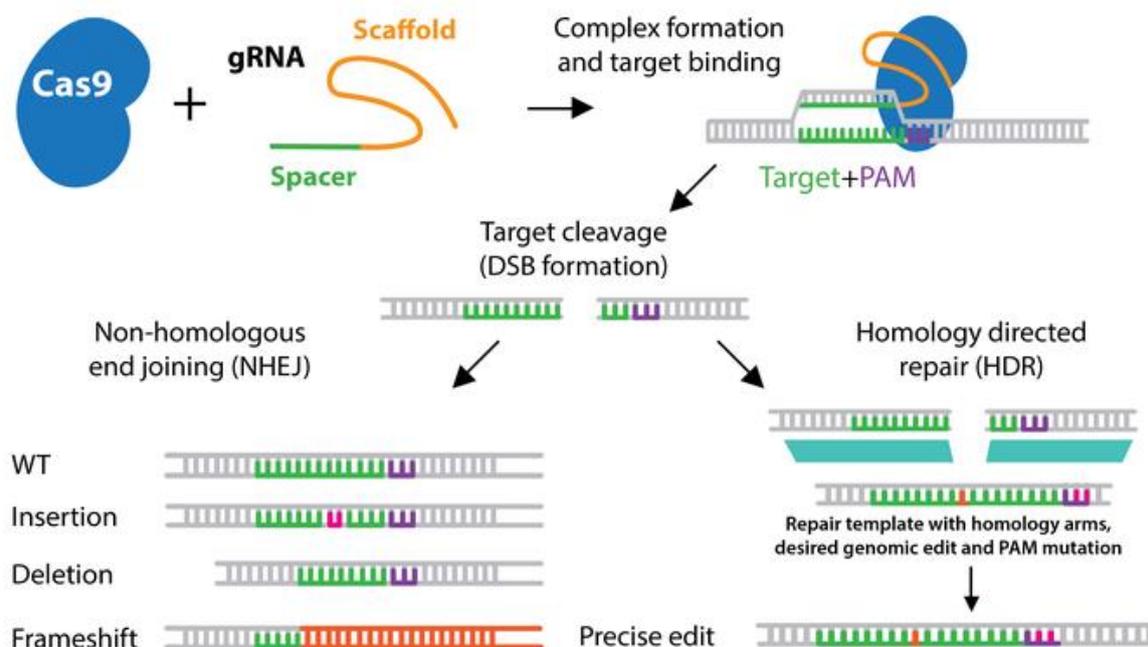


Figure 13 : Description schématique du fonctionnement du système CRISPR-Cas9 pour l'édition de génome eucaryote²⁴².

L'expression exogène de la protéine Cas9 et de l'ARNg chimérique, programmé pour cibler un locus précis, induit le clivage de l'ADN à ce locus dans la cellule hôte. Cette dernière peut mettre alors en place deux processus de réparation, qui à terme permettent d'arrêter l'action de Cas9 en faisant disparaître son site de reconnaissance. La voie NHEJ génère des mutations qui aboutissent généralement à l'inactivation du gène. La voie HDR, en présence d'une matrice de réparation exogène, peut également être utilisée pour apporter, outre une altération du site de reconnaissance, des modifications désirées.

Si le système NHEJ semble être déficient, ou absent, dans la majorité des bactéries^{246,247}, *Clostridium* compris, il n'y a à l'heure actuelle pas de consensus sur la faculté des bactéries à déclencher une voie HDR en réponse à la cassure double brin générée par Cas9. En réalité, dans la majorité des cas, l'activité catalytique de Cas9 est hautement toxique chez les procaryotes²⁴⁷⁻²⁵². Cette toxicité serait liée d'une part à l'absence ou la déficience du système NHEJ, et d'autre part à l'efficacité du système CRISPR, qui empêcherait une réparation efficace par recombinaison homologue (avec le chromosome sœur généré pendant la division cellulaire, qui est lui aussi très rapidement clivé²⁴⁶). Pour tirer parti du système CRISPR, Jiang *et al.* utilisent cette toxicité chez *Streptococcus pneumoniae* et *E. coli* en se servant de CRISPR comme d'un marqueur de contre-sélection pour l'échange allélique, en s'aidant respectivement

de machineries de recombinaison endogène et exogène efficaces²⁵¹. Cette stratégie permet, sans intégrer dans le génome ni marqueur de sélection, ni marqueur de contre-sélection, d'obtenir directement une souche bactérienne génétiquement modifiée, sans autre modification que celles désirées.

Chez les *Clostridia* solvantogènes, des outils basés sur CRISPR ont ainsi été développés, selon cette stratégie générale, pour modifier les gènes des souches étudiées, en ajouter ou en supprimer^{123,247,253–257}. Au cours de deux études successives, Wang *et al.* démontrent la faisabilité du système CRISPR/Cas9 pour l'édition de génome appliqué à *C. beijerinckii* NCIMB 8052, malgré la faible efficacité de recombinaison homologue de la souche, sans utiliser de système de réparation exogène^{254,255}. Pour contrecarrer la toxicité du complexe Cas9/ARNg, les auteurs placent le gène *Cas9* sous le contrôle d'un promoteur inductible au lactose. Ils parviennent ce faisant à supprimer 1,5 kb dans les gènes *pta* et *ack*, à insérer dans ce locus le gène *adhE* (1,6 kb), et à réaliser une mutation ponctuelle dans l'ORF du gène *pta*. Toutefois, ils observent également un fort taux d'intégration de leur plasmide dans leurs clones positifs (c'est-à-dire qui ont réalisé l'échange allélique complet, avec en plus une intégration du vecteur par croisement simple), évènement qu'ils arrivent à supprimer dans leurs clones par l'utilisation d'un ARNg ciblant le plasmide.

Mais ce premier outil CRISPR est limité par la taille des ajouts envisageables car tous les éléments sont amenés via un seul plasmide et l'intégration dans un seul vecteur de *cas9*, du marqueur de sélection, de l'ARNg, et de la matrice de réparation peut rapidement se montrer limitante en termes de taille. Récemment, Wasels et collègues démontrent chez *C. acetobutylicum* ATCC 824 que cette barrière peut être nettement atténuée en utilisant un système à deux plasmides compatibles, l'un contenant *cas9* sous le contrôle d'un promoteur inductible à l'anhydrotétracycline, et l'autre contenant l'ARNg et la matrice de réparation²⁵⁶. Avec ce système, les auteurs parviennent à insérer 3,6 kb dans le génome en une étape, sans utiliser le procédé de sous-cultures successives proposé par Wang *et al.*²⁵⁵.

Des résultats récents ont montré que ce système pouvait encore être perfectionné. En effet, malgré l'inductibilité des éléments CRISPR, une baisse significative des efficacités de transformation est observée par rapport aux plasmides vides. Cette observation traduit une fuite probable des promoteurs inductibles, suffisante pour provoquer une expression des gènes CRISPR, fortement cytotoxiques.

Le nouveau système mis en place chez *C. acetobutylicum* prévoit l'utilisation d'une protéine anti-CRISPR, AcrIIA4²⁵⁸, dont l'expression contrôlée permet d'augmenter significativement les efficacités de transformation. Cette protéine, identifiée chez un prophage de *Listeria monocytogenes*, peut se lier au complexe Cas9-ARNg et totalement inhiber son activité à plusieurs niveaux^{259,260}. Dans le cadre de l'outil génétique, l'induction de l'expression de cette protéine pendant la phase de transformation permet simplement de contrecarrer l'effet toxique de la transcription permissive des éléments CRISPR.

Au commencement des travaux de cette thèse, ce type de système à deux plasmides ne pouvait être employé chez *C. beijerinckii* DSM 6423 dans la mesure où cette souche est naturellement résistante au thiamphénicol, un des deux marqueurs les plus utilisés chez les *Clostridia* solvantogènes. Des travaux réalisés à IFPEN, antérieurs à cette thèse, démontrent cependant que l'activité catalytique de Cas9 est fonctionnelle dans cette souche, puisqu'elle permet d'éliminer des éléments extrachromosomiques (plasmide pNF1, bactériophage ϕ 6423) avec un système à un seul plasmide (qui confère une résistance à l'érythromycine). Considérant ces résultats préliminaires, et les résultats obtenus par Wang *et al.* sur une souche relativement proche²⁵⁵, il est probable qu'un outil CRISPR soit adaptable à la souche *C. beijerinckii* DSM 6423.

2 PROBLEMATIQUES DE LA THESE

Une part importante de composés chimiques intermédiaires est issue du raffinage du pétrole. La pénurie programmée de cette ressource, ainsi que l'impact environnemental croissant de son extraction et de sa transformation, poussent la communauté internationale à envisager et développer des alternatives « vertes » à la pétrochimie²⁰. Le propylène, traditionnellement obtenu par voie fossile, représente un marché important dans l'industrie chimique, notamment pour la production de polypropylène.

Le développement d'un procédé biosourcé de production de propylène permettrait de diversifier l'offre et de limiter la dépendance du propylène aux voies fossiles. L'option retenue par IFPEN consiste à produire, par fermentation de sucres de première ou seconde génération, un mélange d'alcools contenant de l'isopropanol. Ce procédé fermentaire implique l'utilisation de souches de *Clostridium* solvantogènes capables de réaliser des fermentations de type IBE (isopropanol-butanol-éthanol). L'isopropanol obtenu par cette voie peut être converti en propylène, via une étape supplémentaire de déshydratation.

Cependant, les souches solvantogènes IBE naturelles sont peu nombreuses, peu décrites et peu performantes. Grâce à une première thèse IFPEN, la souche *C. beijerinckii* DSM 6423 (également NRRL B-593) a été décrite plus en profondeur. Cette souche réduit l'acétone, normalement produit par la vaste majorité des *Clostridia* solvantogènes, en isopropanol sous l'action d'une alcool déshydrogénase dépendante du NADPH (s-ADH)^{38,39}. Des travaux récents du département Biotechnologie d'IFPEN ont permis de séquencer son génome et de cartographier l'évolution de son transcriptome au cours d'une fermentation en bioréacteur sur milieu modèle⁵². Cette thèse se place dans la continuité de ces recherches et vise à faire de *C. beijerinckii* DSM 6423 une souche plateforme, caractérisée et facilement modifiable, pour pouvoir, à terme, la rendre adaptée à une bioproduction industrielle de solvants.

Dans cette optique, quatre stratégies de recherche principales ont été suivies. La première consistait à développer un outil génétique permettant de modifier la souche DSM 6423 de manière efficace. En effet, même si une première étape avait

été franchie avant le début de ma thèse avec l'établissement d'un protocole de transformation de ce microorganisme, beaucoup restait à faire pour développer un outil génétique dédié. Par ailleurs, un outil génétique performant basé sur la technologie CRISPR-Cas9 avait été développé en parallèle par l'équipe pour une autre souche, *C. acetobutylicum* ATCC 824²⁵⁶. L'objectif principal de ce premier volet était donc d'adapter cet outil à la souche *C. beijerinckii* DSM 6423.

Deuxièmement, un mutant de *C. beijerinckii* DSM 6423 incapable de produire de l'isopropanol semblait intéressant à étudier afin d'identifier le/les facteur(s) moléculaire(s) responsable(s) de son phénotype. Cette souche a été obtenue précédemment par mutagenèse aléatoire en présence d'allyl alcool, un composé toxique pour les cellules présentant une activité alcool déshydrogénase⁴⁷. L'analyse en profondeur des mutations présentes dans cette souche incapable de produire des solvants pouvait potentiellement permettre d'identifier de nouveaux facteurs ou nouvelles voies de régulations impliquées dans le métabolisme solvantogène de ce producteur naturel d'isopropanol et éventuellement d'espèces / souches proches.

Troisièmement, une approche de type omique a été entreprise pour comparer la souche DSM 6423 aux microorganismes modèles de la fermentation ABE, *C. acetobutylicum* ATCC 824 et *C. beijerinckii* NCIMB 8052. Ces trois souches présentent une physiologie proche, bien que l'analyse de leurs génomes respectifs ait révélé des disparités (notamment entre les deux espèces, *C. beijerinckii* et *C. acetobutylicum*). Une méthode dérivée du RNA-seq, le *Capp-switch seq*²¹⁸, a été utilisée pour déterminer les sites d'initiation de la transcription (TSS) dans chacune de ces souches. L'identification des TSS le long d'un génome permet notamment de révéler la dynamique particulière des machineries d'initiation transcriptionnelle, et une étude comparative inter-souche permettrait d'identifier des mécanismes de régulations communs ou divergents dans ces souches solvantogènes de référence.

La caractérisation des TSS complète les travaux de Máté de Gérando et al.⁵², en approfondissant nos connaissances sur la souche en conditions modèles (*batch*, glucose). Mais ces dernières ne reflètent pas la réalité industrielle : le dernier volet de cette thèse a donc consisté à caractériser les capacités fermentaires de la souche sur des substrats utilisables à large échelle, en fermentation continue. Après sélection d'un substrat, la mélasse de betterave, parmi une variété de sources carbonées, un procédé de fermentation continu a été développé pour séparer les phases

métaboliques caractéristiques des *Clostridia* solvantogènes (acidogenèse, solvantogenèse). Ce procédé fermentaire a ensuite été couplé à une analyse multi-omique (protéomique, transcriptomique, TSS) de manière à mettre en lumière les mécanismes métaboliques et génétiques spécifiquement mis en place en présence de substrats complexes.

3 PARTIE I : DEVELOPPEMENT D'UN OUTIL D'EDITION CIBLEE DU GENOME BASE SUR CRISPR-Cas9

Cette partie couvre des travaux publiés dans la revue *Methods*²⁶¹ et dans deux brevets déposés à l'INPI^{262,263} :

Diallo, M.*, Hocq, R.* et al. **Adaptation and application of a two-plasmid inducible CRISPR-Cas9 system in *Clostridium beijerinckii***. *Methods*. 2019. doi: 10.1016/j.jymeth.2019.07.022. *Contributeurs égaux.

Hocq et al 2018 Brevet FR18/73.492 : **Bactéries *Clostridium* génétiquement modifiées, préparation et utilisations de celles-ci** (date de dépôt : 20/12/2018)

Hocq et al 2019 Brevet FR19/05.485 : **Outil génétique optimisé pour modifier les bactéries du genre *Clostridium*** (date de dépôt : 24/05/2019)

3.1 RESUME DES TRAVAUX PUBLIES

Comme mentionné en introduction, l'intérêt des souches solvantogènes du genre *Clostridium* provient de leur capacité naturelle à produire des alcools d'intérêt industriel. Leurs performances fermentaires restent toutefois en deçà des attentes. Des améliorations concernant les conduites fermentaires et les souches bactériennes employées sont donc nécessaires pour qu'un procédé économiquement viable aboutisse.

L'ingénierie métabolique peut améliorer les souches utilisées dans de tels procédés fermentaires, mais est tributaire d'outils génétiques performants. Le développement d'outils d'édition du génome basés sur la technologie CRISPR-Cas9 a permis de faciliter considérablement les efforts en génie génétique, en permettant de modifier avec précision le génome de quantité d'organismes avec une grande efficacité. A ce jour, seul un nombre restreint de souches modèles de *Clostridia* solvantogènes ont été modifiées avec des outils basés sur CRISPR (cf. sous-partie 1.5.3). Rien n'a pour l'instant pu être expérimenté pour la souche modèle de la

fermentation de type IBE, *C. beijerinckii* DSM 6423. Une partie de mes travaux de thèse a donc consisté à développer un outil génétique performant pour cette souche (Figure 14).

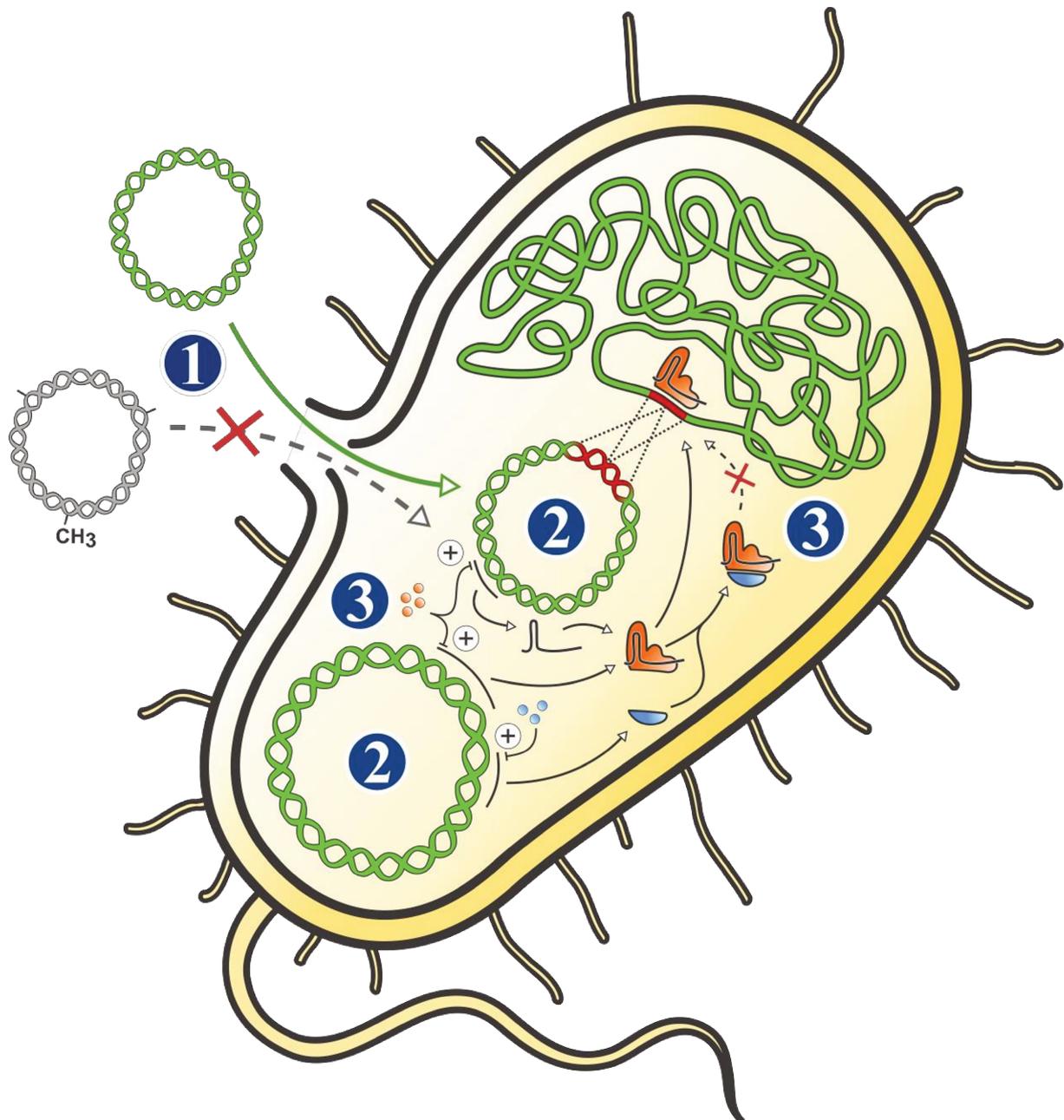


Figure 14 : Description schématique des différentes stratégies employées dans cette étude pour modifier génétiquement la souche DSM 6423.

1. L'utilisation de plasmide Dam⁻ Dcm⁻ permet de transformer la souche. **2.** Un système à deux plasmides est utilisé. Le premier plasmide encode des éléments constants (notamment cas9 et le gène codant la protéine anti-CRISPR AcrIIA4). Le deuxième plasmide comprend les éléments variables (ARNg, matrice de réparation), et doit être modifié pour chaque cible génétique. **3.** Le système repose sur l'utilisation de gènes inducibles, pour contrôler la toxicité latente de Cas9. cas9 est sous le contrôle d'un promoteur inducible à l'anhydrotétracycline, acrlIA4 est inducible au lactose. AcrIIA4, un inhibiteur de la nucléase Cas9, est surexprimée au moment de la transformation de manière à obtenir des colonies. Cas9 n'est exprimé qu'au moment de réaliser la modification désirée.

Les premières études dédiées à ces souches de *Clostridium* illustrent certaines difficultés d'utilisation de CRISPR-Cas9 liées à la toxicité de Cas9 ou du complexe Cas9-gRNA. En particulier, les faibles efficacités de recombinaison homologue et d'efficacités de transformation limitent le nombre de transformants obtenus après électroporation. La transformabilité de la souche *C. beijerinckii* DSM 6423 n'a été démontrée que peu de temps avant le début de ces travaux, avec des efficacités de transformation très faibles. Les premières expériences conduites dans le cadre de cette thèse ont confirmé ces résultats, et permis d'obtenir des transformants en utilisant des plasmides vides préparés dans une souche d'*E. coli dam⁻ dcm⁻* (1-5 CFU/μg). Au vu des faibles efficacités de transformation observées, il était alors nécessaire d'utiliser de grandes quantités d'ADN ($\approx 20 \mu\text{g}$) afin d'obtenir un nombre de colonies conséquent (≈ 100 CFU pour un plasmide sans insert), dans l'optique de pouvoir utiliser des plasmides CRISPR, dont les efficacités de transformations sont plus limitées, et donc de pouvoir modifier la souche.

Comme première preuve de concept, la suppression avec un complexe Cas9-gRNA du plasmide naturel pNF2 a été réalisée. Ce plasmide, l'un des 3 éléments génétiques mobiles naturels de la souche *C. beijerinckii* DSM 6423⁵², s'avère interférer avec la transformation d'éléments génétiques exogènes. En effet, sa perte a permis d'améliorer grandement les efficacités de transformation (entre 20 et 2000 fois, selon le plasmide). Ce type d'amélioration, obtenue en éliminant un plasmide naturel, a déjà été reporté dans la littérature pour une autre souche solvantogène²⁶⁴, avec un effet toutefois moins important.

L'effet drastique de cette modification pourrait être dû au nombre de copies relativement élevé (10-15 copies estimées par séquençage Illumina) du pNF2 dans la souche sauvage. Ceci pourrait entraîner une forme de compétition entre sa réplication et celle de plasmides exogènes. Alternativement, la nature des 3 gènes inconnus codés par le plasmide (le dernier coderait une protéine impliquée dans la réplication du plasmide) pourrait également impacter d'une façon ou d'une autre la procédure de transformation. Des recherches complémentaires sont nécessaires pour (in)valider l'une ou l'autre de ces hypothèses.

Par ailleurs, le plasmide pNF2 présente des particularités intéressantes. Ainsi, sa relative petite taille (≈ 4 kb) et sa stabilité naturelle dans la souche nous ont conduits à tenter de l'utiliser comme outil génétique. Pour ce faire, la partie prédite comme étant

impliquée dans la réplication du plasmide naturel a été assemblée avec l'origine de réplication pour *E. coli* et le gène de résistance à l'ampicilline du plasmide pUC19, ainsi que divers gènes de résistance utilisables chez *Clostridium* (*ermB*, *catP*, *aad9* conférant respectivement une résistance à l'érythromycine, au thiamphénicol et à la spectinomycine). Nous avons ainsi obtenu plusieurs plasmides, désignée pNF3E/C/S, qui présentaient de très bonnes efficacités de transformation chez *C. beijerinckii* DSM 6423 Δ pNF2 (entre 1500 et 2500 CFU/ μ g, selon le gène de résistance). Nous avons par la suite confirmé le potentiel de cet outil en démontrant que ces plasmides pouvaient être transformés, avec des efficacités variables, dans *C. beijerinckii* NCIMB 8052, *C. acetobutylicum* ATCC 824, *C. ljungdahlii* DSM 13528 et *B. subtilis* 168. Considérant le large spectre d'hôtes de ces plasmides, ces derniers pourraient être utilisés pour diverses applications chez les Firmicutes.

L'amélioration des efficacités de transformation obtenues dans la souche DSM 6423 est significative, et devrait permettre d'accélérer les procédures de génie génétique. Dans cette optique, le second volet de notre stratégie de développement d'un outil génétique dans la souche DSM 6423 a consisté à valider un système CRISPR-Cas9 à deux plasmides, équivalent à celui décrit chez *C. acetobutylicum* ATCC 824 par notre groupe²⁵⁶. Cette stratégie présente quelques avantages intéressants par rapport aux outils à plasmide unique développés jusqu'alors dans d'autres solvantogènes. Utiliser deux plasmides permet de séparer les éléments du système CRISPR, ce qui peut être bénéfique lors de l'étape de clonage, *E. coli* étant également sensible à la toxicité des complexes Cas9-gRNA. Par ailleurs, les clonages sont facilités par l'utilisation de plasmides plus petits. Enfin, le gain de place obtenu peut être utilisé pour apporter d'autres éléments utiles (comme des recombinaisons, par exemple²⁶⁵), ou pour réaliser des insertions plus importantes, particulièrement indiquées dans une optique de génie métabolique.

L'inconvénient majeur d'un tel système est qu'il nécessite l'utilisation de deux origines de réplication et de deux marqueurs de sélection compatibles. Des tests de croissance en présence d'antibiotiques ont permis de déterminer que *C. beijerinckii* DSM 6423 est sensible à l'érythromycine et à la spectinomycine, mais résistante au thiamphénicol. Cependant, nous n'avons réussi à obtenir de transformants stables qu'en utilisant le gène *ermB* de *C. difficile*, qui confère une résistance à l'érythromycine. Après de multiples essais infructueux, nous n'avons pas été en

mesures de valider l'utilisation du gène *aad9* conférant la résistance à la spectinomycine dans cette souche. N'ayant à disposition qu'un marqueur de résistance, une analyse du matériel génétique de la souche DSM 6423 par la plateforme d'analyse MAGE²⁶⁶ a abouti à l'identification d'un gène *catB* (CIBE_3859, amphénicol acétyltransférase), prédit comme étant le seul élément génétique pouvant conférer une résistance au thiamphénicol dans le génome. La délétion de ce gène à l'aide d'un système CRISPR-Cas n'utilisant qu'un plasmide unique a rendu la souche sensible au thiamphénicol. Le gène *catP* de *C. perfringens* employé dans l'outil appliqué à *C. acetobutylicum* ATCC 824 a alors pu être utilisé comme marqueur de sélection. Ainsi, le plasmide pEC750C (origine pIP404) portant le gène *catP* a été introduit avec succès dans la souche DSM 6423 Δ *catB* contenant le plasmide exogène pFW01 (origine repH, résistance à l'érythromycine), validant la compatibilité entre les origines de répllication utilisées.

La dernière stratégie mise en œuvre dans cette étude a visé à améliorer la procédure de génie génétique en contrôlant précisément l'activité de Cas9, de manière à réduire fortement, voire à abolir, sa toxicité. En effet, malgré les améliorations susmentionnées (amélioration des efficacités de transformation, système double plasmide), la toxicité des éléments CRISPR pouvait dans certains cas poser problème pour leur introduction dans la souche DSM 6423 Δ *catB*, malgré l'utilisation d'un système inductible pour l'expression de *cas9* et du gRNA²⁵⁶. En présence d'un plasmide contrôlant l'expression de *cas9*, nous n'avons en effet jamais été en mesure d'obtenir des clones après transformation d'un deuxième plasmide (nécessaire pour l'expression d'un gRNA ciblant le gène *upp*, et pour la matrice de réparation permettant l'éviction de ce gène), et ce même après élimination du pNF2. Ces résultats peuvent s'expliquer par un niveau basal de transcription (en absence d'inducteur) des éléments CRISPR, suffisamment toxique pour contre-sélectionner les clones contenant les deux plasmides.

Cette hypothèse suggère qu'une forme de contrôle complémentaire pourrait être utile pour améliorer l'efficacité du processus. Etant confronté au même problème chez *C. acetobutylicum* ATCC 824, notre groupe a développé un système CRISPR/anti-CRISPR (cf. manuscrit pour publication en annexe), qui permet d'inhiber l'activité nucléolytique de Cas9-gRNA au niveau protéique. Cet ajout, basé sur une expression inductible d'*acrIIA4* codant une protéine de défense contre le système

Cas9 et isolée d'un prophage de *Listeria monocytogenes*, augmente significativement les efficacités de transformation des plasmides CRISPR, en les rendant similaires aux efficacités de transformation des plasmides vides. Au premier essai avec ce système chez *C. beijerinckii* DSM 6423 $\Delta catB$, 28 clones portant les plasmides nécessaires à l'élimination du gène *upp* ont ainsi été obtenus sur milieu sélectif, en présence de lactose, l'inducteur de l'expression d'*acrIIA4*. Après induction du système CRISPR, dans un milieu sans lactose, tous les clones testés (10) se sont avérés être positifs, ce qui apporte la preuve de concept du système.

En revanche, l'éviction du gène *upp* (CIBE_0562, codant une uridine phosphoribosyltransférase) n'a pas eu d'impact sur le phénotype. Son inactivation chez *C. acetobutylicum* ATCC 824 provoque en effet une résistance au 5-fluorouracile (5-FU), un analogue nucléosidique toxique²³⁸. La sensibilité persistante au 5-FU dans la souche Δupp pourrait être due à la présence d'un système uridine kinase/phosphorylase (CIBE_1796 & CIBE_3715) dans son génome, absent chez *C. acetobutylicum*. Chez l'homme, le 5-FU est utilisé pour le traitement de certains cancers et son métabolisme est caractérisé²⁶⁷. L'une des voies de métabolisation passe par une uridine phosphatase et une uridine kinase, ce qui expliquerait pourquoi dans la souche Δupp , qui présente une voie métabolique analogue, ce produit est encore fortement toxique. Pour valider cette hypothèse, une délétion complémentaire de *udk* et/ou de *udp* dans la souche Δupp serait nécessaire.

En conclusion, les diverses stratégies entreprises dans cette étude ont permis de mettre au point un outil génétique performant, adapté à la souche DSM 6423, qui a permis une édition ciblée de son génome pour la première fois. Plusieurs phénotypes intéressants ont été obtenus (sensibilité au thiamphénicol, efficacité de transformation accrue) qui devraient permettre de faciliter grandement le génie génétique de la souche. Par ailleurs, les différentes approches menées ici ne sont pour la plupart pas spécifiques de *C. beijerinckii* DSM 6423, et pourraient également être appliquées à d'autres microorganismes atypiques.

3.2 VALORISATION SCIENTIFIQUE (ARTICLE)

Le travail de développement d'un outil génétique performant dans la souche *C. beijerinckii* DSM 6423 a pu être en partie valorisé au travers d'un article scientifique, fruit d'une collaboration entre notre laboratoire et le laboratoire Food & Biobased Research à Wageningen (Pays-Bas) pour valider l'utilisation de l'outil dans la souche de référence NCIMB 8052. Dans cette publication, nos deux équipes décrivent l'utilisation d'un système CRISPR inductible, à deux plasmides, dans les deux souches de *C. beijerinckii*. Notre contribution à l'article a trait à toutes les expériences réalisées dans la souche DSM 6423, et notamment :

- Description de la transformation de la souche DSM 6423 ;
- Adaptation du système à deux plasmides de *C. acetobutylicum* à la souche DSM 6423 ;
- Obtention pour la première fois de mutants de la souche DSM 6423 par une approche ciblée ;
- Caractérisation de deux nouveaux phénotypes ($\Delta catB$, $\Delta pNF2$).



Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Adaptation and application of a two-plasmid inducible CRISPR-Cas9 system in *Clostridium beijerinckii*

Mamou Diallo^{a,*}, Rémi Hocq^{b,1}, Florent Collas^a, Gwladys Chartier^b, François Wasels^b,
Hani Surya Wijaya^a, Marc W.T. Werten^a, Emil J.H. Wolbert^a, Servé W.M. Kengen^c,
John van der Oost^c, Nicolas Lopes Ferreira^{b,*}, Ana M. López-Contreras^a

^a Bioconversion Group, Wageningen Food and Biobased Research, Bornse Weilanden 9, 6708WG Wageningen, the Netherlands

^b IFP Energies nouvelles, Biotechnology Department, 1 et 4 avenue de Bois-Préau, 92852 Rueil-Malmaison, France

^c Laboratory of Microbiology, Wageningen University, Stippeneng 4, 6708 WE Wageningen, the Netherlands

ARTICLE INFO

Keywords:

Clostridium beijerinckii

CRISPR-Cas9

Nuclease

Genome editing

ABSTRACT

Recent developments in CRISPR technologies have opened new possibilities for improving genome editing tools dedicated to the *Clostridium* genus. In this study we adapted a two-plasmid tool based on this technology to enable scarless modification of the genome of two reference strains of *Clostridium beijerinckii* producing an Acetone/Butanol/Ethanol (ABE) or an Isopropanol/Butanol/Ethanol (IBE) mix of solvents. In the NCIMB 8052 ABE-producing strain, inactivation of the SpoIIIE sporulation factor encoding gene resulted in sporulation-deficient mutants, and this phenotype was reverted by complementing the mutant strain with a functional *spoIIIE* gene. Furthermore, the fungal cellulase-encoding *celA* gene was inserted into the *C. beijerinckii* NCIMB 8052 chromosome, resulting in mutants with endoglucanase activity. A similar two-plasmid approach was next used to edit the genome of the natural IBE-producing strain *C. beijerinckii* DSM 6423, which has never been genetically engineered before. Firstly, the *catB* gene conferring thiamphenicol resistance was deleted to make this strain compatible with our dual-plasmid editing system. As a proof of concept, our dual-plasmid system was then used in *C. beijerinckii* DSM 6423 $\Delta catB$ to remove the endogenous pNF2 plasmid, which led to a sharp increase of transformation efficiencies.

1. Introduction

Solventogenic *Clostridia* are mesophilic, rod shaped gram positive, anaerobic and sporogenic bacteria belonging to the *Firmicutes* phylum. Model species such as *Clostridium acetobutylicum* and *Clostridium beijerinckii* are able to produce a wide range of valuable chemicals among which acetone/isopropanol, butanol and ethanol during the A/IBE fermentation [1–5]. Moreover, *Clostridium* species have been reported to produce other chemicals such as acetoin, 2,3-butanediol [2,6], or 1,3-propanediol [7–9]. However, the A/IBE fermentations suffer from the low end-products final concentrations mainly related to their toxicity [10].

Over the last decades, many efforts were made to improve the solvent tolerance and productivity of solventogenic *Clostridia* by genetic

engineering. Forward genetics (i.e. random mutagenesis) have been mostly used to study molecular mechanisms in *Clostridia* [11,12], because targeted mutagenesis has long been impeded by the lack of efficient genetic tools for the *Clostridium* genus. Since the adaptation of the TargeTron technology [13] to the *Clostridia* genus in 2007 and the creation of a genetic tool called ClosTron [14,15], targeted approaches have been greatly facilitated. This technology allows gene disruption through the precise insertion of a mobile group II intron into the genome. However, polar effects on flanking genes have been reported, making the subsequent analysis of mutant phenotypes difficult [16]. Moreover, considering that this method does not enable large genetic insertions, other tools were created to alleviate these bottlenecks. In particular, strategies using counter-selection markers have been developed to select rare allelic exchange events [17–19]. In 2013, Al-

Abbreviations: ABE, Acetone Butanol Ethanol; CRISPR, Clustered Regularly Interspaced Short Palindromic Repeats; Cas, CRISPR-associated protein; Cas9n, Cas9 nickase; DAE, Double Allelic Exchange; DSB, Double-Strand Break; gDNA, genomic DNA; gRNA, guide RNA; HR, Homologous Recombination; NHEJ, Non-Homologous End Joining; PAM, Protospacer Adjacent Motif; PCR, Polymerase Chain Reaction; SAE, Simple Allelic Exchange

* Corresponding authors.

E-mail addresses: mamou.diallo@wur.nl (M. Diallo), nicolas.lopes-ferreira@ifpen.fr (N.L. Ferreira).

¹ Mamou Diallo and Rémi Hocq contributed equally to this work.

<https://doi.org/10.1016/j.ymeth.2019.07.022>

Received 14 March 2019; Received in revised form 22 July 2019; Accepted 25 July 2019

1046-2023/ © 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Please cite this article as: Mamou Diallo, et al., Methods, <https://doi.org/10.1016/j.ymeth.2019.07.022>

Table 1
Strains and plasmids used in this study.

Strains and plasmids	Relevant information	Reference or source
Clostridia		
<i>Clostridium beijerinckii</i> NCIMB 8052	ABE producer, sensitive to erythromycin (25–50 µg/mL) and spectinomycin (650 µg/mL)	[46]
<i>Clostridium beijerinckii</i> DSM 6423	IBE producer, <i>adh</i> gene, sensitive to erythromycin (25–50 µg/mL), resistant to thiamphenicol (15 µg/mL)	[3]
<i>Clostridium difficile</i> str. 630	Gene <i>xytR</i> and promoter of the gene <i>xytB</i>	[29]
Other microbial strains		
<i>E. coli</i> XL1-blue	endA1 gyrA96(nalR) thi-1 recA1 relA1 lac glnV44 F ⁺ ::Tn10 proAB + lacIq Δ(lacZ)M15] hsdR17(rK- mK +)	Agilent
<i>E. coli</i> NEB 10-beta	Δ(ara-leu) 7697 araD139 fhuA ΔlacX74 galK16 galE15 e14-φ80dlacZΔM15 recA1 relA1 endA1 nupG rpsL (Str ^R) rph spoT1 Δ(mrr-hsdRMS-mcrBC)	NEB
<i>E. coli</i> INV110	F ⁺ {traΔ36 proAB lacI ^q lacΔM15} rpsL (Str ^R) thr leu endA thi-1 lacY galK galT ara tonA tsx dam dcm supE44 Δ(lac-proAB) Δ(mcrCmrr)102::Tn10(Tet ^R)	Invitrogen
<i>S. cerevisiae</i> W303	MATa/MATα {leu2-3,112 trp1-1 can1-100 ura3-1 ade2-1 his3-11,15} [phi +]	ATCC 200060
<i>C. beijerinckii</i> NCIMB 8052 genome editing plasmids		
pMTL500E	Ap ^R ; Em ^R ; colE1; pAMβ1	[35]
pSEC500E	Ap ^R ; Em ^R ; colE1; pAMβ1; 2µ; URA3	This study
pYC100S	Sp ^R ; colE1; pCB102; 2µ; URA3	This study
pWUR100S (pS)	Sp ^R ; colE1; pCB102	This study
pE_X_cas9	pSEC500E derivative with <i>cas9</i> (inducible <i>xytB</i> promoter) insertion	This study
pS_X ^R KR	pWUR100S derivative with <i>Kan^R</i> ; chimera reprogrammable gRNA (inducible <i>xytB</i> promoter) and MCS	This study
pS_X ^R _celA_S1	pS_X ^R KR derivative with cassette “ <i>celA</i> ” and gRNA <i>celA_S1</i> (PAM sequence located 36 bp downstream of the <i>hbd</i> gene (cbei_0325) stop codon) insertion	This study
pS_X ^R ThlPcelA_S1	pS_X ^R _celA_S1 derivative with the <i>thl</i> promoter from cbei_0411	This study
pS_X ^R _ΔspoIIE_S1	pS_X ^R KR derivative with cassette “Δ <i>spoIIE</i> ” and gRNA <i>spoIIE_S1</i> (PAM located 571 bp downstream the start codon of <i>spoIIE</i>) insertion	This study
pWUR3	pMTL500E derivative with <i>celA</i> (constitutive promoter <i>eglA</i> and its signal peptide) insertion	[41]
<i>SpoIIE</i> complementation plasmid		
pSpoIIE	pWUR100S derivative with the <i>spoIIE</i> (cbei_0097)	This study
<i>C. beijerinckii</i> DSM 6423 genome editing plasmids		
pFW01	Em ^R ; colE1; pCB102	Wasels et al. (2017)
pCas9 _{ind}	pFW01 derivative with <i>cas9</i> (inducible Pcm-tetO2/1 promoter) and <i>tetR</i> (miniP _{thl} promoter) insertions	Wasels et al. (2017)
pEC750C	Tm ^R ; colE1; pIP404	Wasels et al. (2017)
pCas9 _{ind} -ΔcatB_S1	pCas9 _{ind} derivative with cassette “Δ <i>catB</i> ” and gRNA <i>catB_S1</i> (PAM sequence located 76 bp upstream of the <i>catB</i> gene (CIBE_3859) stop codon (inducible Pcm-2tetO1 promoter)) insertion	This study
pGRNA-pNF2_S1	pEC750C derivative with gRNA <i>pNF2_S1</i> (PAM sequence located 285 bp downstream of the CIBE_p20002 gene start/ stop codon (inducible Pcm-2tetO1 promoter)) insertion	This study

Ap, ampicillin; Em, erythromycin; Sp, spectinomycin; Tm, thiamphenicol; Kan, kanamycin; Tet, tetracycline.

Hinai et al. described a markerless genome editing strategy using the MazF toxin from *Escherichia coli* as a counter selection marker [17]. The methodology enabled the deletion of several genes and the insertion of a 3.6 kb DNA fragment leaving a FRT scar in the modified chromosome. Ehsaan et al. developed the ACE system (Allele Coupled Exchange) using the *pyrE* gene, encoding the pyrimidine biosynthesis enzyme “orotate phosphoribosyltransferase”, as a positive and negative selection marker [18,20]. However, these methods are time consuming as they require several selection and subculturing steps. The cells in which the plasmid has been integrated into the genome via a single crossover event are selected and then subcultured in selective media to further allow the allelic exchange. This lengthens the engineering process and favors the occurrence of spontaneous mutations.

CRISPR-Cas9 technology for genetic modification of eukaryotic and prokaryotic cells provides an alternative to these genome editing methods [21]. Recently, several publications reported the use of CRISPR-Cas9 tools for scarless genome editing in *Clostridium cellulolyticum* [22], *C. acetobutylicum* [23] and *C. beijerinckii* [23–26]. Most groups reported difficulties when the native Cas9 nuclease from *S. pyogenes* was constitutively expressed. Indeed, the double strand break (DSB) generated by Cas9 is reported to be lethal because of the absence of non-homologous end-joining (NHEJ) mechanisms in *Clostridia* [22,27]. Therefore, the Cas9 nuclease was often replaced by a Cas9 nickase [23,25] in which the RuvC domain was mutated (D10A) [28]. This methodology enabled the insertion of DNA fragments up to 1.7 kb after serial transfers [22]. Alternatively, Wang et al. [24] overcame the DSB lethality by placing the *cas9* gene under the control of a lactose inducible promoter [29] enabling deletions, insertions or single-nucleotide modifications (SNM) within the genome of *C. beijerinckii*. All

the systems described above are based on a “single vector strategy”. Such strategies are however limited due to the length of the *cas9* gene (4.1 kb), which restricts the cargo capacity of the transformation vectors and therefore limits the size of potential genome insertions.

Approaches using two separate plasmids for the sequential introduction of the *cas9* gene and the gRNA sequence into a targeted strain have been recently described in the solventogenic *C. acetobutylicum* [30] and in the acetogen *C. autoethanogenum* [31]. In the case of *C. acetobutylicum*, the system enabled precise genome editing with an efficiency of 100%. Indeed, the system enabled several modifications ranging from a single base pair substitution to the genomic insertion of a 3.6 kb DNA fragment. In the case of *C. autoethanogenum*, the system enabled a partial deletion of an alcohol dehydrogenase gene and the deletion of a 2,3-butanediol dehydrogenase gene (1 kb) with an efficiency of 50%. Although being powerful, the two-plasmid editing strategy requires two functional selection markers, which might not always be available for some strains.

In this study, we adapted our recently-described dual-plasmid CRISPR-Cas9 genetic tool [30] to edit *C. beijerinckii* strains. In the ABE model strain *C. beijerinckii* NCIMB 8052, the *cas9* gene and the gRNA expression cassette were placed on two distinct plasmids. Both *cas9* and gRNA were placed under the control of a xylose inducible promoter [32] to enable a tight control of nuclease activity. The homologous DNA repair template, required for allelic exchange, was placed on the same plasmid as the gRNA expression cassette. The lethal effect of DSB was taken as a strong asset to use the CRISPR system as a counter-selection against unedited cells. This system enabled the scarless deletion of 2.379 kb within the *spoIIE* gene (cbei_0097) as well as the insertion of the *celA* gene from *Neocallimastix patriciarum* in *C. beijerinckii* NCIMB

8052 with 100% and 75% efficiency, respectively.

The applicability of this two-plasmid approach was also investigated in the IBE producing *C. beijerinckii* DSM 6423 strain, for which successful genetic modification have never been reported so far. Being only able to use the erythromycin resistance gene as a selection marker, a preliminary strategy involving a single-plasmid CRISPR-Cas9 tool was undertaken to inactivate the CIBE_3859 gene (*catB*), encoding a putative amphenicol acetyltransferase. The Δ *catB* mutant was sensitive to thiamphenicol and therefore compatible with a two-plasmid editing approach, which was subsequently used to delete the endogenous pNF2 plasmid as a proof of concept. Interestingly, the removal of this mobile genetic element resulted in a drastically increased transformation efficiency, paving the way to further genome editing of this natural isopropanol producer.

2. Materials and methods

2.1. Strains and culture condition

Relevant characteristics of the bacterial strains and plasmids used in this study are listed in Table 1. *C. beijerinckii* NCIMB 8052 and DSM 6423 are laboratory strains. The *spoIIE* mutant was stored as vegetative cells in mCGM with 20% glycerol at -80°C . All the other strains were kept as spore suspension. Except for fermentation assays, liquid cultures of *C. beijerinckii* NCIMB 8052 and related mutants were grown in mCGM containing per liter: yeast extract, 5 g; KH_2PO_4 , 0.75 g; K_2HPO_4 , 0.75 g; asparagine- H_2O , 2 g; $(\text{NH}_4)_2\text{SO}_4$, 2 g; cysteine, 0.5 g; $\text{MgSO}_4 \cdot 7 \text{H}_2\text{O}$, 0.4 g; $\text{MnSO}_4 \cdot \text{H}_2\text{O}$, 0.01 g; $\text{FeSO}_4 \cdot 7 \text{H}_2\text{O}$, 0.01 g; glucose, 10 g. These microorganisms were grown and selected on mCGM agar containing per liter: yeast extract, 1 g; tryptone, 2 g; KH_2PO_4 , 0.5 g; K_2HPO_4 , 1 g; $(\text{NH}_4)_2\text{SO}_4$, 2 g; $\text{MgSO}_4 \cdot 7 \text{H}_2\text{O}$, 0.1 g; $\text{MnSO}_4 \cdot \text{H}_2\text{O}$, 0.01 g; $\text{FeSO}_4 \cdot 7 \text{H}_2\text{O}$, 0.015 g; CaCl_2 , 0.01 g; CoCl_2 , 0.002 g; CaCl_2 , 0.002 g; glucose, 50 g and agar 12 g. *C. beijerinckii* DSM 6423 and related mutants were cultivated in liquid 2YTG (per liter: tryptone, 16 g; yeast extract, 10 g; NaCl, 5 g; glucose, 20 g, pH 5.2) and grown on solid 2YTG (same as above, with neutral pH, 15 g/L agar and 5 g/L glucose). *Clostridium* mutants were selected in culture medium containing 50 $\mu\text{g}/\text{mL}$ of erythromycin (Duchefa), 650 $\mu\text{g}/\text{mL}$ spectinomycin (Duchefa) and/or 15 $\mu\text{g}/\text{mL}$ thiamphenicol (Sigma-Aldrich), as needed. Liquid media were made anaerobic by flushing with nitrogen gas or by incubation for at least 4 h in an anaerobic chamber. Cultivation was performed at 37°C , without shaking, and anaerobically in (i) an anaerobic chamber or (ii) in glass serum vials as described above. *E. coli* and *S. cerevisiae* strains were stored as vegetative cells in 15% glycerol solution at -80°C . *E. coli* was grown on LB medium, supplemented with 100 $\mu\text{g}/\text{mL}$ of ampicillin (Duchefa), 100 $\mu\text{g}/\text{mL}$ spectinomycin (Duchefa), 25 $\mu\text{g}/\text{mL}$ chloramphenicol (Sigma-Aldrich) and/or 500 $\mu\text{g}/\text{mL}$ erythromycin (Duchefa).

2.2. Plasmid construction and transformation

Except where noted, plasmids and DNA templates used for *C. beijerinckii* NCIMB 8052 CRISPR system were generated as yeast/*E. coli*/*Clostridium* shuttle vector in the yeast *S. cerevisiae* by homologous recombination as described previously [34]. Yeasts were transformed by the lithium acetate method [35]. Competent cells were transformed with linear plasmids and insert(s) produced by PCR. Primers are listed in supplementary file S1. Each primer contained a 30 bp 5'-tail specific to another DNA fragment to allow homologous recombination. Final yeast/*E. coli*/*Clostridium* shuttle vectors obtained from yeast were digested by AvrII and self-ligated using the T4-DNA ligase (NEB) to evict the yeast replication origin (2microns) and selection marker (URA3 gene).

The *SpoIIE* complementation plasmid was constructed using the Circular Polymerase Extension Cloning [36]. The pS shuttle vector was digested by SphI and XhoI enzymes for linearization and the insert was

amplified by PCR on genomic DNA from *C. beijerinckii* NCIMB 8052 using the primers Cbei0097prom_F and Cbei0097_R.

The plasmid pS_XR_celA_S1 was constructed (with the promoter of the *thl* gene, cbei_3630, driving expression of *celA*) via homologous recombination in *S. cerevisiae* using the primers listed in supplementary file S1. Unfortunately, non-silent mutations were found in the *celA* gene. In the meantime, transcriptome analysis [37] showed that another copy of the *thl* gene, cbei_0411 was more expressed and thus its promoter would be a stronger promoter than cbei_3630 and a better choice. Therefore, to simultaneously address both matters, a BsaI/XcmI fragment encompassing the relevant regions was replaced with an accordingly designed synthetic BsaI/XcmI fragment (Genscript). The resulting vector, pS_XRThlPcelA_S1, was sequence-verified and uses the promoter of the *thl* gene (cbei_0411) to drive the expression of the *celA* gene.

Plasmids were introduced in *C. beijerinckii* NCIMB 8052 and related recombinant strains prepared according to the protocol described previously [38].

To construct pCas9_{ind}- Δ *catB*_S1, a DNA cassette containing the Δ *catB* template DNA as well as the gRNA targeting the *C. beijerinckii* DSM 6423 *catB* locus (anhydrotetracycline inducible promoter) was synthesized by Eurofin Genomics and cloned in the pEX-A258 commercial vector. This cassette was amplified by PCR (Δ *catB*_fwd and Δ *catB*_rev primers) and cloned at the XhoI restriction site in pCas9_{ind} [30].

To construct pGRNA-pNF2, a gRNA expression cassette composed of the Pcm-2tetO1 promoter [39] and a 20-nt guiding sequence targeting pNF2 fused with the chimeric gRNA sequence [28], pNF2-gRNA, was synthesized (Eurofins Genomics, sequence available in supplementary file S1), and cloned into the SacI-digested pEC750C.

Prior to electroporation, plasmids were transformed in a *dam*⁻ *dcm*⁻ *E. coli* strain (INV110, Invitrogen). *Dam*⁻ *Dcm*⁻ DNA was then used to transform *C. beijerinckii* DSM 6423 as described previously [40], with a few modifications outlined thereafter. 2YTG 20 g/L glucose was used for liquid cultures. Mid-exponential phase (OD₆₀₀ = 0.6–0.8) cells were used and 5–20 μg DNA was electroporated with a single pulse (100 Ω , 25 μF , 1.4 kV) in 2 mm-gap cuvettes.

2.3. Reprogramming the gRNA protospacer in pS_X^RK^R

The CRISPR-Cas9 target sequences were selected after a manual search in the target region for a 25 bp spacer sequence flanked in its 3' end by a 5'-NGG-3' motif named protospacer adjacent motif (PAM). The spacer was designed by the annealing of two primers; (i) the primer "Target_Forw" consisting in the spacer sequence to which the motif "5'-GAGG-3'" was added at its 5' end and (ii) the primer "Target_Rev" consisting in the reverse complement of the spacer sequence to which the motif "5'-AAAC-3'" was added at its 5' end. Both primers were mixed (1 μM each), heated at 95°C for 10 min and cooled down at room temperature. The primer dimer was then inserted in pS_X^RK^R by Golden Gate cloning using the BsaI enzyme (NEB). Ligation and digestion were performed concomitantly for 1 h at 37°C .

2.4. Selection of mutant strains

C. beijerinckii NCIMB 8052 recombinant strains harboring the two plasmids of the CRISPR-Cas9 system were grown at 37°C in 25 mL liquid mCGM medium containing 0 (without induction) or 40 g/L xylose (pre-induction) for 24 h–3 days to induce the CRISPR-Cas9 system. Once the optical density of the culture containing xylose reached 1.5 MAU at 600 nm, these cultures were concentrated in 100 μL of fresh mCGM media. 50 μL of the cell suspension were spread on mCGM selective plates supplemented 40 g/L xylose, 50 $\mu\text{g}/\text{mL}$ erythromycin and 650 $\mu\text{g}/\text{mL}$ spectinomycin and incubated for 24 h–3 days to obtain single colonies. Edited cells were screened by colony PCR to detect single and double allelic exchange profiles.

C. beijerinckii DSM 6423 recombinant colonies bearing a single- or dual-plasmid genome editing system were treated as described previously [30], with a 200 ng/mL anhydrotetracycline (Sigma-Aldrich) induction.

2.5. Fermentation assays

Fermentations were performed in CM2 medium [41] which contains per liter: yeast extract, 5 g; KH_2PO_4 , 1 g; K_2HPO_4 , 0.76 g; NH_4Ac , 3 g; p-aminobenzoic acid, 0.1 g; $\text{MgSO}_4 \cdot 7 \text{H}_2\text{O}$, 1 g; and $\text{FeSO}_4 \cdot 7 \text{H}_2\text{O}$, 0.5 g, glucose, 60 g. 50 mL of media were inoculated with 250 μL of a heat shocked fresh spore suspension and incubated overnight at 35 °C. The overnight cultures were inoculated (5% (v/v)) in fresh culture medium and incubated at 35 °C for 72 h. The concentrations of metabolites were determined in clear culture supernatants after removal of microbial biomass by centrifugation. Glucose, acetate, butyrate, lactate, acetoin, acetone, butanol and ethanol were determined by HPLC as described previously [2,42] using valeric acid (30 mM) as an internal standard.

2.6. Microscopy

Phase contrast microscopy (Olympus BX51; Olympus, Tokyo, Japan) was used to determine the presence of endospores or pre-spore stages. Native bacteria were observed by phase contrast microscopy at $\times 1000$ magnifications.

3. Results

3.1. Genome editing strategy

The *cas9* gene was placed on a different plasmid from the gRNA expression cassette and the editing template so that (1) cloning was facilitated (i.e. with the use of smaller plasmids) and (2) space was available for large editing templates. Since previous CRISPR-Cas9-related studies in *Clostridia* hypothesized that DSBs catalyzed by Cas9 are likely to severely impair transformation efficiency [22,43], both the *cas9* gene located on pE_X_cas9 and the gRNA expression cassette located on pS_XKR were placed under the control of the inducible promoter Pxy1B from *C. difficile* str. 630 [32] (Supplementary file S2). To complete our 2-plasmid inducible system, the *xylR* regulatory gene was further cloned into the plasmid pS_XKR yielding pS_X^RKR.

Editing templates and gRNA expression cassettes were thus inserted into the pS_X^RKR plasmid (Fig. 1a). Each editing template contains two homology sequences (HS) of at least 500 bp, named HS1 and HS2. For genomic deletion, the two HS flank the region that has to be removed, which contains the CRISPR target site. Genomic insertions were designed by placing the sequence of interest between HS1 and HS2. This configuration allows 3 different allelic exchange events between the genome and the plasmid containing the DNA template (Supplementary file S3): two single-crossover allelic exchanges (SAE) and one double-crossover allelic exchange (DAE). The latter exchange is the only one which results in genomic DNA without the CRISPR-Cas9 target, which enables the survival of the transformants upon Cas9 induction.

The pS_X^RKR-based targeting plasmids were then introduced into a *C. beijerinckii* NCIMB 8052 recombinant strain already transformed with the pE_X_cas9 plasmid (Fig. 1b–c).

Following plasmid selection, xylose was used to induce the expression of both *cas9* and the gRNA. This critical step enables the selection of correctly edited cells, as, in the absence of non-homologous repair systems, Cas9 targets the bacterial chromosome, resulting in cell death. The final recombinant strains were next obtained by curing the pS_X^RKR-based plasmid and the pE_X_cas9 plasmid by serial sub-culturing in mCGM media lacking both antibiotics (Fig. 1d–e). Alternatively, selectively curing the targeting plasmid allows to directly reuse the modified cells for another round of genome editing (Fig. 1g).

To evaluate the robustness of this new genome-editing approach for

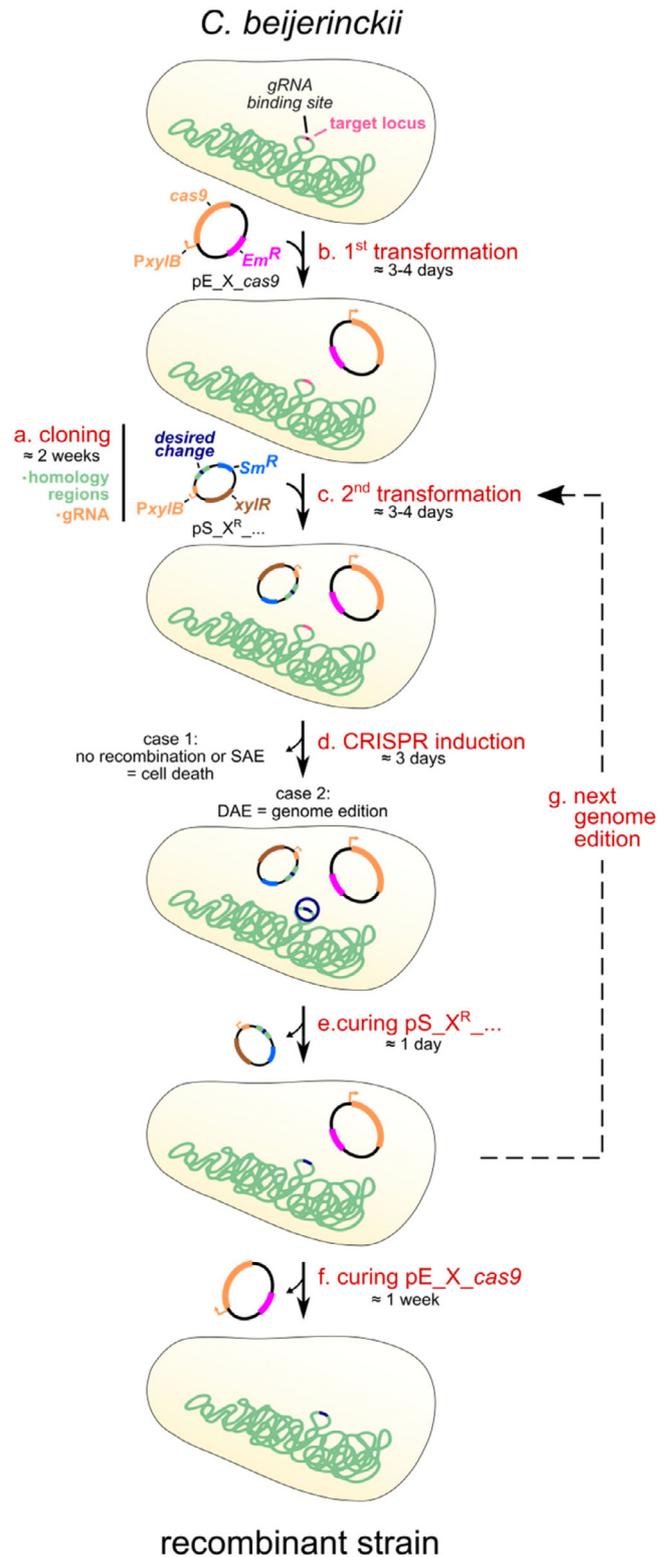


Fig. 1. Dual-plasmid inducible CRISPR-Cas9 strategy for *C. beijerinckii* NCIMB 8052 genome editing. a. Preliminary cloning of homology sequences and gRNA protospacer in pS_X^RKR plasmid to obtain a targeting plasmid. b,c. Transformation of (b) pE_X_cas9 in the WT strain and (c) of the targeting plasmid in the latter strain. d. Induction of *cas9* and gRNA expression to yield an edited strain. e,f,g. Curing of the targeting plasmid (e) and pE_X_cas9 plasmid (f) yields a markerless genome edited mutant. Alternatively (g), the targeting plasmid can be specifically removed to promote another round of genome editing. gRNA: guide RNA; Em^R: erythromycin resistance marker; Sm^R: spectinomycin resistance marker.

Table 2
Chromosomal modifications performed in this study.

Strain	DNA template	Size of the sequences (base pairs)				Screening results ^a			
		Deleted	Inserted	HS1	HS2	DAE	WT	DAE & WT	“.”
NCIMB 8052	<i>ΔspoIIE</i>	2424	0	741	1020	7	0	0	1
	<i>celA</i>	19	1667	1001	1017	4	0	0	1
DSM 6423	<i>ΔcatB</i>	653	8	400	400	3	0	0	1

HS1 and HS2 Homologous sequence; WT: wild type; DAE: Double allelic exchange; SAE: Single allelic exchange; “.”: no amplification.

^a Screening was performed on colonies growing with the highest xylose/aTc concentrations.

modification of *C. beijerinckii* NCIMB 8052, the system was used for: (i) a gene deletion and (ii) a gene insertion. The modifications performed in this study and their characteristics are summarized in Table 2.

3.2. Deletion of the *spoIIE* gene, characterization of the edited strains and complementation with a functional *spoIIE* gene

The *spoIIE* gene (cbei_0097) encoding a phosphatase involved in the sporulation regulation was targeted first. In model organisms *Bacillus subtilis* and *C. acetobutylicum*, SpoIIE is involved in the first steps (Stage II) of the regulation the sporulation mechanisms cycle [44,45]. It enables the activation of the sporulation-specific sigma factor σ^F and is necessary to complete asymmetric division. In *C. acetobutylicum*, SpoIIE inactivation does not impair solvent production [46,47]. The pS_X Δ *spoIIE*_S1 plasmid contains the “*ΔspoIIE*” DNA editing template (Table 1 and Fig. 2a) and the “*spoIIE*_S1” gRNA. HS1 and HS2 templates were designed to remove 2379 bp within the coding sequence of the *spoIIE* gene. Without induction of the CRISPR-Cas9 machinery, PCR screening showed that the full pS_X Δ *spoIIE*_S1 plasmid frequently integrated into the host genome by SAE mechanisms either with HS1 or with HS2. However, no DAE event was observed (data not shown). Overnight xylose induction in the *C. beijerinckii* (pE_X Δ *cas9*, pS_X Δ *spoIIE*_S1) strain negatively affected its growth, suggesting a toxic effect of Cas9 activity (Fig. 2c). This overnight culture was then plated on mCGM agar containing xylose (Fig. 2d), which permitted the selection of correctly edited *ΔspoIIE* cells (7/7 DAE events, Fig. 2e and Table 2). Microscopic analysis showed that the *ΔspoIIE* cells were more elongated than the wild-type cells and did not form viable endospores (Fig. 3a-b). Similarly to what was observed in *C. acetobutylicum*, mutant strains were still able to produce solvents (Supplementary file S4). To test whether the asporogeneous phenotype was effectively caused by the disruption of the *spoIIE* gene, complementation with a plasmidic copy of the *spoIIE* gene under the control of its native promoter was performed. As expected, sporulation was restored in the complemented mutant (Fig. 3a-b).

3.3. Insertion of the *celA* gene from *Neocallimastix patriciarum*

The *celA* gene from the fungus *Neocallimastix patriciarum* [48,49], encoding a cellobiohydrolase, was inserted downstream of the *hbd* gene (cbei_0325). The pS_XR Δ *celA*_S1 plasmid, harboring the “*celA*” DNA template (Table 1 and Fig. 2b) and the “*celA*_S1” gRNA, was introduced in strain NCIMB 8052 (pE_X Δ *cas9*). The *celA* template was designed to enable the insertion of the *celA* gene under the control of the thiolase promoter from *C. beijerinckii* (1667 bp) and the deletion of a 19 bp sequence located 36 base pairs downstream of the stop codon of the *hbd* gene (cbei_0325).

Similarly to what was observed for *spoIIE* inactivation, only SAE events were detected by PCR in the transformants containing pE_X Δ *cas9* and pS_XR Δ *celA*_S1, prior to xylose induction. Following xylose induction, 3 out of the 4 xylose-resistant colonies tested were correctly edited (Fig. 2e and Table 2).

The resulting *celA* mutants displayed cellulolytic activity on mCGM plates containing 1 g/L CMC (carboxymethylcellulose), similarly to a

control *C. beijerinckii* strain carrying the *celA* transcription unit on a plasmid (pWUR3, [48]) (Fig. 3b).

3.4. Adaptation of the dual-plasmid genome editing strategy to *C. beijerinckii* DSM 6423

The advantages provided by a two-plasmid CRISPR-Cas9 genome editing system (i.e. scarless modifications, high efficiency, relatively high DNA cargo capacity) prompted us to develop a similar system in the poorly studied *C. beijerinckii* DSM 6423 strain. A first obstacle was the lack of knowledge on the transformability of this natural IBE producer. This hurdle was overcome by applying a DNA demethylation strategy prior to electroporation, as described by Kolek et al. [33] in the closely related *C. beijerinckii* NRRL B-598 strain (Supplementary file S7). This treatment enabled the generation of transformants containing pFW01-derived vectors [30,50] (Supplementary file 7). However, vectors bearing the spectinomycin resistance marker could not be introduced successfully in this strain (data not shown). This was a major bottleneck for the use of a two-plasmid editing strategy since this microorganism was also resistant to thiamphenicol, an antibiotic often used for selection in clostridial recombinant systems [51].

To alleviate this second obstacle, a single-plasmid CRISPR-based approach was developed to disrupt the *catB* gene (CIBE_3859), predicted to encode an amphenicol acetyltransferase probably involved in thiamphenicol resistance. pE_X Δ *cas9* plasmid was not chosen as a backbone since this already large vector (\approx 13 kb, Supplementary file S2) does not carry the *xylR* gene necessary for *cas9* inducibility. Instead, the 8874 bp pCas9 Δ plasmid previously used for genome editing in *C. acetobutylicum* ATCC 824 [30] was engineered to include an anhydrotetracycline-inducible gRNA expression cassette targeting *catB* as well as homology sequences designed to delete 99% of the *catB* gene (Fig. 4a, Table 2). The resulting pCas9 Δ *catB*_S1 vector was introduced in *C. beijerinckii* DSM 6423, and mutants with the expected *catB* deletion were obtained after anhydrotetracycline (aTc) induction (Fig. 4b).

As anticipated, these *ΔcatB* mutants were sensitive to thiamphenicol (Fig. 4c) and allowed the maintenance of the pEC750C vector in this modified strain, which carries the *catP* marker from *C. perfringens* [52]. Besides, this vector could be introduced in a strain already bearing a pFW01-based plasmid. In order to test whether a dual-plasmid CRISPR-Cas9 editing approach could be applied to the *ΔcatB* mutant, we constructed a pEC750C-based plasmid carrying an aTc-inducible gRNA cassette targeting the endogenous pNF2 plasmid (pGRNA-pNF2_S1). The pNF2 plasmid [53] was discovered during the reconstruction of *C. beijerinckii* DSM 6423 genome along with other extra-chromosomal genetic elements (i.e. another plasmid and a linear phage). The pNF2 plasmid is the smallest of these elements (4282 bp) and bears 4 putative genes, 3 of which have no predicted function, and the last one is predicted to encode a potential plasmid replication protein. The pGRNA-pNF2_S1 vector was subsequently introduced in *C. beijerinckii* DSM 6423 *ΔcatB* (pCas9 Δ) which allowed the highly efficient selection of Δ pNF2 mutants (Fig. 4d-e). Fermentation assays further revealed that the deletion of Δ pNF2 did not affect alcohol production (Supplementary file S8). However, pNF2 removal resulted in increased transformation

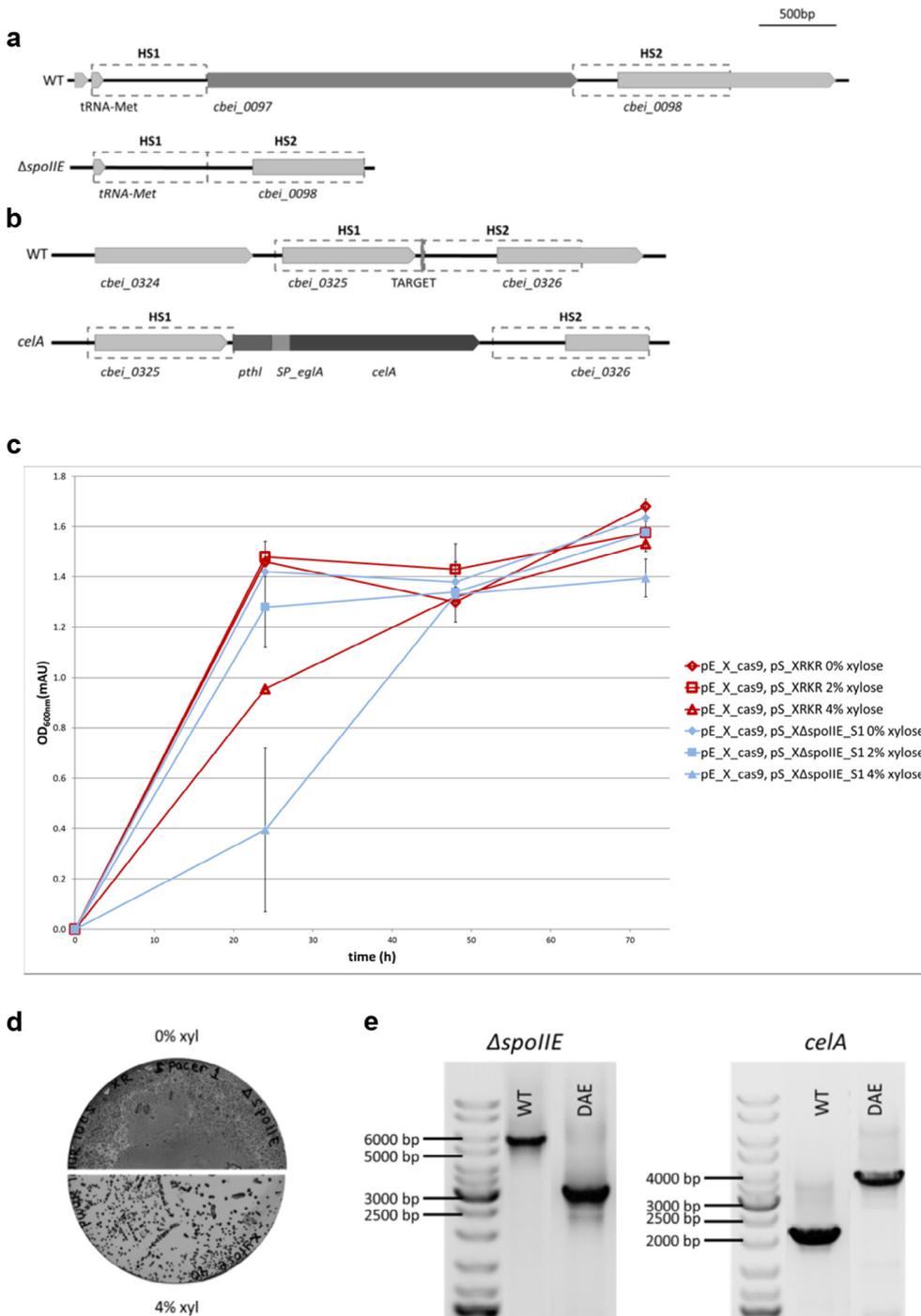


Fig. 2. Genome editing of the *C. beijerinckii* NCIMB 8052 strain: design and selection. a, b. Targeted loci and their relative DNA templates. (a) Disruption of *spoIIE*. (b) Insertion of the *celA* gene from *N. patriarum*. c, d, e. Selection and screening of edited strain. (c) Growth of *C. beijerinckii* NCIMB 8052 pE_X_cas9 transformed with the empty gRNA plasmid pS_XRKR or the *spoIIE* disruption plasmid pS_X $\Delta spoIIE$ _S1. The transformants were grown in duplicates in mCGM supplemented with erythromycin and spectinomycin and with 0–4% of xylose. The error bars represent the error of the replicates (d) Selection of *spoIIE* on mCGM agar plate containing 0 or 4% xylose (w/v) inoculated with overnight pre-induced cultures. (e) Growing xylose-resistant colonies were subsequently screened by colony PCR. The screening for *spoIIE* gene disruption was performed using primers cbei_0096_F/cbei_0098_R. Expected sizes: 5462 bp (WT) and 3083 bp (DAE). The screening of the insertion of the *celA* gene was performed using primers cbei_0325_F/cbei_0326_R. Expected sizes: 2083 bp (WT) and 3607 bp (DAE). HS1 and HS2: homologous sequence 1 and 2; *SP_eglA*: signal peptide *eglA*; *pthI*: promoter of *C. beijerinckii* thiolase gene (cbei_3630); WT: wild type; DAE: double allelic exchange (edited cells).

efficiencies (from 20 to 2000 fold depending on the plasmid, Fig. 4f).

In summary, we successfully adapted the two-plasmid CRISPR-Cas9 editing tool to *C. beijerinckii* DSM 6423 and created successively two platform strains. The first modification made the strain sensitive to thiamphenicol, which allowed us to use two different resistance markers. This mutant strain was further engineered by removing pNF2 which allowed us to obtain a second strain, with higher transformation efficiencies.

4. Discussion

In this study we describe the successful adaptation of a CRISPR/Cas9 genetic tool in the solvent-producing *C. beijerinckii*. The two-

plasmid system, in which the *cas9* gene was cloned on a first plasmid - the nuclease plasmid - while the gRNA and the DNA template were cloned on a second plasmid - the editing plasmid, was successfully used to edit the genome of two reference strains of *C. beijerinckii*. Such a method benefits its user on several points. Firstly, the cloning procedure is facilitated because the toxic genes of the CRISPR-Cas9 machinery are localized on separate genetic elements. Therefore, it is easier to construct plasmids for a given genetic modification, which will also be significantly smaller, likely resulting in higher transformation efficiencies in *Clostridia*.

Secondly, our methodology helps reducing the time spent on the traditionally laborious genetic engineering steps. The strain bearing the nuclease plasmid can indeed be stored at -80°C and used multiple

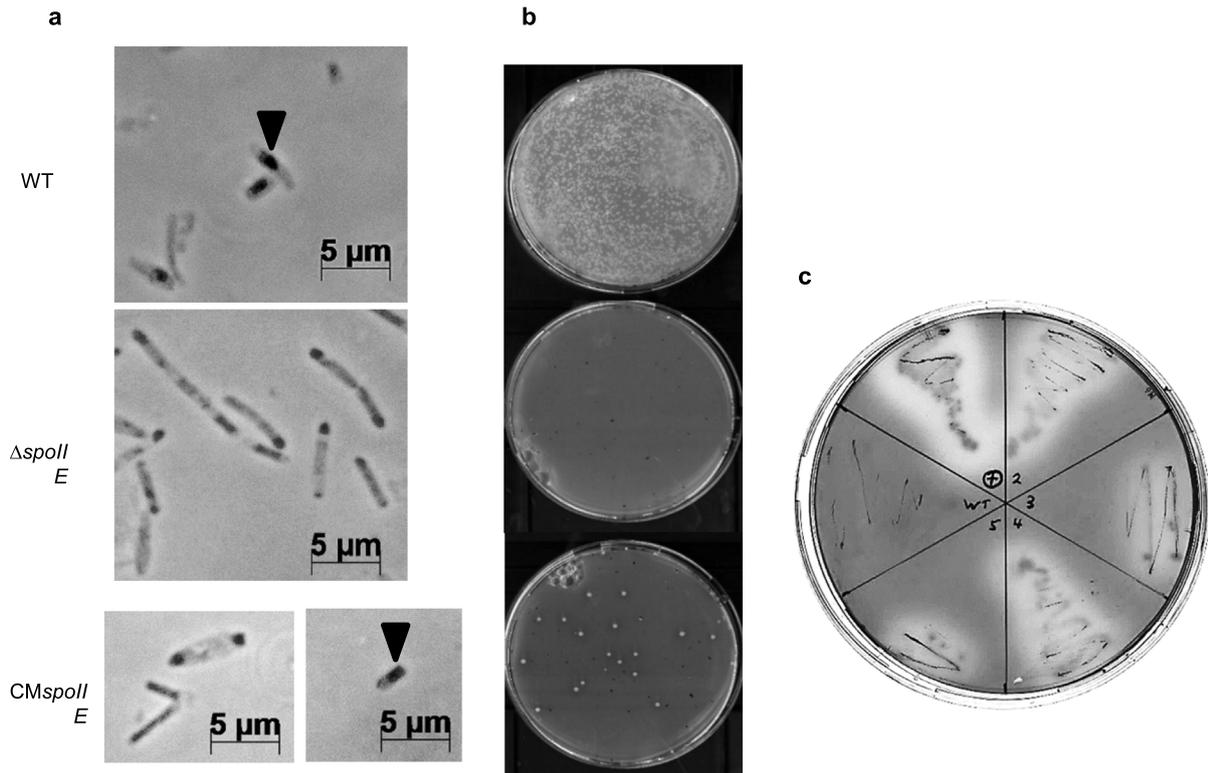


Fig. 3. Novel phenotypes in genetically modified *C. beijerinckii* NCIMB 8052 mutants. a. Light microscopy pictures at 73 h of fermentation of *C. beijerinckii* NCIMB 8052 wt, $\Delta spoII E$, $CMspoII E$ (complemented mutant) strains. Spores are marked by a triangle. b. Growth of *C. beijerinckii* NCIMB 8052 wt, $\Delta spoII E$, $CMspoII E$ on CM2 agar plates after heat shocking 200 μ L of a 48-hour-old culture. c. Detection of the cellulase activity of *C. beijerinckii* celA mutants. WT: *C. beijerinckii* wild type, 2–5: clone number (CelA mutants after plasmid curing), + : positive control (non-integrated self-replicating CelA plasmid) the halo surrounding the cells are caused by the degradation of carboxymethylcellulose.

times, for several genetic constructions. Similarly, the editing plasmid can be specifically cured after a genome editing experiment, permitting multiple successive rounds of genome editing. This editing plasmid was besides designed to fasten the cloning process by incorporating genetic elements allowing a quick and easy reprogramming of the gRNA. In our hands, CRISPR-Cas9-based counter-selection also proved to be extremely efficient with most clones – if not all – edited after the induction step, which also contributes to shorten the mutant screening process.

Lastly, the system displays a good genetic cargo capacity (up to 3.5 kb insertions with *C. acetobutylicum* [30]), which makes it suited for large genome modifications required for metabolic engineering. The additional space provided by the second plasmid also gives more freedom for the design of DNA templates and notably enables the use of longer homologous sequence to favor allelic exchanges [54].

In the CRISPR system described in this study, an inducible Cas9 nuclease [24,31] was chosen over the Cas9 nickase previously used by other groups [22,23]. In those studies, edited cells were obtained at lower frequencies than in studies when the Cas9 nuclease was used. This can be explained by a lower toxicity of a single nick in the genome, rather than a DSB. Since the Cas9 nickase is less toxic, and therefore less selective, its use increases the probability of having a mixed population, especially when the desired mutation negatively impacts the fitness of the bacterium, or when the editing events are rare (e.g. insertion of large DNA fragments). As an example, the disruption of *adc* in *C. acetobutylicum* or *C. beijerinckii* using the nickase led to the isolation of 1 and 3 correctly edited colonies out of 16 colonies tested, respectively, whereas other mutations (*xylR*, *cbei_3923*) were obtained in almost all of the tested colonies [23].

In order to use Cas9, the inducible promoter P_{xylB} from *C. difficile* and the *xylR* gene, whose product codes for the associated regulatory

protein [32], were used to control the expression of both *cas9* and the gRNA. The lethal effect of the DSB resulting from their expression following xylose induction was then used as a strong counter selection marker. In this study, this approach allowed the isolation of correctly edited cells at an efficiency close to 100%. Moreover, unlike what is reported in studies using the Cas9 nickase, serial transfers of recombinant strains were not required to isolate clones that undergone large fragment insertions in their genome. All mutants generated in this study were therefore isolated within one week starting from the strain *C. beijerinckii* NCIMB 8052 (pE_X_{cas9}). Although the system is mediated by two plasmids, their removal required only a few serial transfers making the system as efficient as those previously described. The easy clearance of the plasmids carrying the gRNA and DNA template and the absence of selection marker in the final strain thus enable multiple rounds of genome editing.

In this study, we deleted a 2.379 kb DNA fragment within the *spoII E* gene (*cbei_0097*), and inserted the 1.68 kb *N. patriciarum celA* gene. The disruption of the *spoII E* gene, described here for the first time in *C. beijerinckii*, resulted in asporogenous mutant strains. These results are in accordance with studies in earlier *Bacillus subtilis* [55] and other solventogenic *Clostridia* [46,56], which showed that a reduced *spoII E* expression coincided with hampered sporulation and that *spoII E* disruption prevents sporulation.

The *spoII E* mutants generated in this study kept the ability to produce solvents at similar levels to the wild-type strain (Supplementary file S4). These observations indicate that this gene is necessary for sporulation but does not play a role in solvent production, contrary to *spo0A* for example. This characteristic has also been observed in *spoII E* negative mutants generated in *C. acetobutylicum* [46,47]. In contrast to the *spo0A* homologues in *Clostridia*, the *spoII E* gene is not well conserved between clostridial species. Indeed, while *C. acetobutylicum*

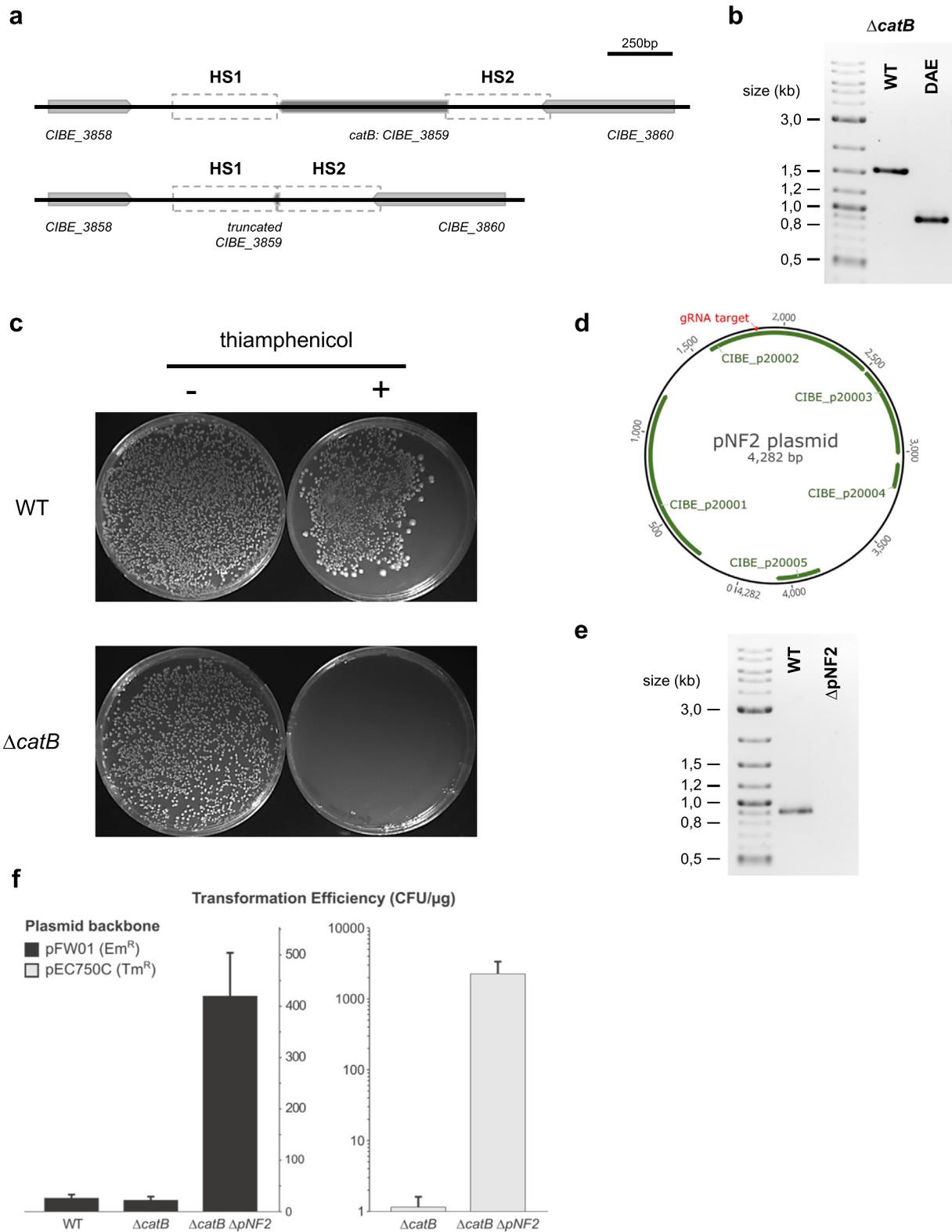


Fig. 4. Adaptation of the two-plasmid CRISPR-Cas9 system to *C. beijerinckii* DSM 6423. **a.** *catB* (CIBE_3859) genome deletion design by a single-plasmid strategy. **b.** Verification of *catB* deletion by PCR with primers RH76 and RH77, which amplify a genome region encompassing the homology sequences (expected sizes: 1553 (WT) and 900 ($\Delta catB$) bp). **c.** Thiamphenicol resistance assay on the resulting $\Delta catB$ strain. **d.** Map of the natural pNF2 plasmid from *C. beijerinckii* DSM 6423. The locus targeted by the pGRNA-pNF2.S1 vector (containing the thiamphenicol resistance gene *catP*) is shown in red. **e.** Verification of the deletion of the pNF2 plasmid with a dual-plasmid CRISPR-Cas9 strategy. Primers pNF2_fwd and pNF2_rev were used for PCR (expected size: 907 bp (WT); no amplification is expected from a $\Delta catB \Delta pNF2$ strain). **f.** Transformation efficiencies in the wild-type, the $\Delta catB$ and the $\Delta catB \Delta pNF2$ strain. Errors bars represent the s.e.m. of duplicate experiments. Em^R: erythromycin resistance gene Tm^R: thiamphenicol resistance gene. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

spoIIE gene has only 36% identity with *C. beijerinckii spoIIE*, *spo0A* homologues have as much as 74% sequence identity. We observed differences at the phenotypic level between *C. acetobutylicum* and *C. beijerinckii spoIIE* mutants. Indeed *spoIIE* mutants in *C. beijerinckii* have two dark spots at their poles (Fig. 3), corresponding to the forespore condensed DNA, that are absent in *C. acetobutylicum spoIIE* mutants but were observed in *C. acetobutylicum sigE* mutants [57]. Moreover *C. beijerinckii spoIIE* mutant cells display an asymmetric septum and sometimes two septa at each pole of the cell, which is similar to the morphology of *B. subtilis spoIIE* mutants [44,58]. We can thus hypothesize that even though SpoIIE is necessary for the correct completion of the stage II of the sporulation cycle in *C. beijerinckii* and *C. acetobutylicum*, it is not needed for asymmetric septum formation in *C. beijerinckii* contrary to what is observed in *C. acetobutylicum*.

The insertion of the “*thl_p-celA*” transcriptional unit yielded mutant strains which showed cellulase activity on plate assays, not observed with the WT strain. The CelA enzyme from the anaerobic fungus *N. patriciarum* is a cellulase that contains an active domain belonging to the Family 6 of glycosyl hydrolases. This enzyme has been expressed before in clostridial strains using a plasmid vector, either alone or as part of a mini-cellulosome [48,59]. This study shows for the first time the insertion of a cellulase gene into the chromosome of a solventogenic strain and its functional expression. This opens novel possibilities for the direct fermentation of sugar polymers like cellulose into ABE or IBE mixtures.

The success of our genome editing strategy prompted us to apply it also to another strain. The atypical *C. beijerinckii* DSM 6423 is genetically very close to the NCIMB 8052 model strain [60], but its modification by targeted genome editing techniques has however never been reported, in spite of the microorganism value as being one of the few natural IBE-producing strains. Taking into account this strain characteristics, we successfully adapted our dual-plasmid inducible CRISPR-Cas9 approach. Indeed, some adjustments were found necessary to (1) circumvent the restriction-modification system of the strain (using Dam⁻ Dcm⁻ DNA, similarly to the NRRL B-598 strain [33]) and (2) find an alternative to the spectinomycin resistance gene, which proved to be malfunctioning in our hands. This bottleneck was specifically tackled by using a single-plasmid inducible CRISPR-Cas9 strategy to disrupt a putative amphenicol resistance gene (CIBE_3859, encoding the *catB* gene). The resulting strain was demonstrated to be sensitive to thiamphenicol, and the *catP* resistance marker (from *C. perfringens* [52], used extensively in *C. acetobutylicum* [51]) could subsequently be used in this mutant. This mutant was further exploited to delete the pNF2 natural plasmid [53] with a dual-plasmid inducible CRISPR-Cas9 system, underlining the compatibility of the $\Delta catB$ strain with such a system. The curation of the pNF2 plasmid was preferred to the modification of chromosomal genes because it required the introduction of a small editing plasmid without editing template. Indeed the poor transformation efficiency of the DSM 6423 strain severely impedes our efforts. For the disruption of the *catB* gene for example, several transformation experiments were required to introduce the single-plasmid inducible CRISPR-Cas9 and only one was successful, yielding 1 transformant when using 20 μ g DNA. This problem underlines the need for better transformation efficiencies in *C. beijerinckii* DSM 6423 for future genome editing experiments.

CRISPR-Cas9-based removal of non-essential genetic elements has already been pioneered in solventogenic *Clostridia*, with notable effects on the primary/secondary metabolism and/or on exogenous plasmid transformation efficiencies [30,61]. If the product pattern and quantities of the resulting $\Delta catB$ $\Delta pNF2$ strain were not strongly affected, the mutant interestingly showed much higher transformation efficiencies, most likely because the endogenous pNF2 plasmid interferes with the replication of non-natural plasmids introduced into the cell. As mentioned above, this improvement is critical as transformation efficiencies are normally extremely low in the wild-type genetic background (c.a. 4 CFU/ μ g with an empty plasmid – Supplementary file S7), which will

likely considerably help constructing future strains. The rise of the CRISPR/Cas9 technology in the field of *Clostridium* genetics has been a real breakthrough. The relatively simple generation of scarless mutants in a short period of time will hopefully facilitate the research into the physiology of solvent production and lead to a better understanding of the regulatory mechanisms involved in gene expression. The dual-plasmid system described in this study should overcome issues reported previously (higher transformation efficiency, flexibility for the design of DNA templates, convenience of the cloning procedure, rapid mutant selection, cargo capacity). We showed here that this advantageous system, initially developed for *C. acetobutylicum*, could be simply adapted to work as efficiently in a model microorganism from another species. We also demonstrated that this approach could be expanded to atypical microorganisms, recalcitrant to genetic manipulations, by following simple steps (i.e. finding the correct DNA methylation pattern, removing antibiotic resistance genes, removing mobile genetic elements), which suggests that similar strategies are likely to function in other prokaryotes.

Funding

This research was supported by the European Union Marie Skłodowska Curie Innovative Training Networks (ITN) Clospore – (contract number 642068) and the European Union’s Horizon 2020 Macrofuels project (contract number 654010).

Declaration of Competing Interest

The authors declare no financial or commercial conflict of interest.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jymeth.2019.07.022>.

References

- [1] S.A. Survase, G. Jurgens, A. van Heiningen, T. Granström, Continuous production of isopropanol and butanol using *Clostridium beijerinckii* DSM 6423, Appl. Microbiol. Biotechnol. 91 (2011) 1305–1313, <https://doi.org/10.1007/s00253-011-3322-3>.
- [2] F. Collas, W. Kuit, B. Clement, R.R. Marchal, A.M. Lopez-Contreras, F.F. Monot, et al., Simultaneous production of isopropanol, butanol, ethanol and 2,3-butanediol by *Clostridium acetobutylicum* ATCC 824 engineered strains, AMB Exp. 2 (2012) 45, <https://doi.org/10.1186/2191-0855-2-45>.
- [3] H.A. George, J.L. Johnson, W.E. Moore, L.V. Holdeman, J.S. Chen, Acetone isopropanol, and butanol production by *Clostridium beijerinckii* (syn. *Clostridium butyricum*) and *Clostridium aurantibutyricum*, Appl. Environ. Microbiol. 45 (1983) 1160–1163.
- [4] D.T. Jones, D.R. Woods, Acetone-butanol fermentation revisited, Microbiol. Rev. 50 (1986) 484–524.
- [5] P. Dürre, Fermentative butanol production, Ann. N. Y. Acad. Sci. 1125 (2008) 353–362, <https://doi.org/10.1196/annals.1419.009>.
- [6] M.a.J. Siemerink, W. Kuit, A.M.L. Contreras, G. Eggink, J. van der Oost, S.W.M. Kengen, D-2,3-butanediol production due to heterologous expression of an acetoin reductase in *Clostridium acetobutylicum*, Appl. Environ. Microbiol. 77 (2011) 2582–2588, <https://doi.org/10.1128/AEM.01616-10>.
- [7] C.W. Forsberg, Production of 1,3-propanediol from glycerol by *Clostridium acetobutylicum* and other clostridium species, Appl. Environ. Microbiol. 53 (1987) 639–643.
- [8] M. González-Pajuelo, J.C. Andrade, I. Vasconcelos, Production of 1,3-propanediol by *Clostridium butyricum* VPI 3266 in continuous cultures with high yield and productivity, J. Ind. Microbiol. Biotechnol. 32 (2005) 391–396, <https://doi.org/10.1007/s10295-005-0012-0>.
- [9] B. Otte, E. Grunwaldt, O. Mahmoud, S. Jennewein, Genome shuffling in *Clostridium diolis* DSM 15410 for improved 1,3-propanediol production, Appl. Environ. Microbiol. 75 (2009) 7610–7616, <https://doi.org/10.1128/AEM.01774-09>.
- [10] P. Patakova, J. Kolek, K. Sedlar, P. Koscova, B. Branska, K. Kupkova, et al., Comparative analysis of high butanol tolerance and production in clostridia, Biotechnol. Adv. (2017), <https://doi.org/10.1016/j.biotechadv.2017.12.004>.
- [11] M. Hermann, F. Fayolle, R. Marchal, L. Podvin, M. Sebald, J.P. Vandecasteele, Isolation and characterization of butanol-resistant mutants of *Clostridium acetobutylicum*, Appl. Environ. Microbiol. 50 (1985) 1238–1243.
- [12] H.M. de Gérando, F. Fayolle-Guichard, L. Rudant, S.K. Millah, F. Monot, N.L. Ferreira, et al., Improving isopropanol tolerance and production of *Clostridium beijerinckii* DSM 6423 by random mutagenesis and genome shuffling, Appl. Microbiol. Biotechnol. 100 (2016) 5427–5436, <https://doi.org/10.1007/s00253->

- 016-7302-5.
- [13] M. Karberg, H. Guo, J. Zhong, R. Coon, J. Perutka, A.M. Lambowitz, Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria, *Nat. Biotechnol.* 19 (2001) 1162–1167, <https://doi.org/10.1038/nbt1201-1162>.
- [14] J.T. Heap, S.A. Kuehne, M. Ehsaan, S.T. Cartman, C.M. Cooksley, J.C. Scott, et al., The Clostron: mutagenesis in *Clostridium* refined and streamlined, *J. Microbiol. Methods* 80 (2010) 49–55, <https://doi.org/10.1016/j.mimet.2009.10.018>.
- [15] J.T. Heap, O.J. Pennington, S.T. Cartman, G.P. Carter, N.P. Minton, The Clostron: a universal gene knock-out system for the genus *Clostridium*, *J. Microbiol. Meth.* 70 (2007) 452–464, <https://doi.org/10.1016/j.mimet.2007.05.021>.
- [16] S.A. Rodriguez, J.-J. Yu, G. Davis, B.P. Arulananandam, K.E. Klose, Targeted inactivation of *Francisella tularensis* genes by group II introns, *Appl. Environ. Microbiol.* 74 (2008) 2619–2626, <https://doi.org/10.1128/AEM.02905-07>.
- [17] M.A. Al-Hinai, A.G. Fast, E.T. Papoutsakis, Novel system for efficient isolation of *Clostridium* double-crossover allelic exchange mutants enabling markerless chromosomal gene deletions and DNA integration, *Appl. Environ. Microbiol.* 78 (2012) 8112–8121, <https://doi.org/10.1128/AEM.02214-12>.
- [18] M. Ehsaan, W. Kuit, Y. Zhang, S.T. Cartman, J.T. Heap, K. Winzer, et al., Mutant generation by allelic exchange and genome resequencing of the biobutanol organism *Clostridium acetobutylicum* ATCC 824, *Biotechnol. Biofuels* 9 (2016) 4, <https://doi.org/10.1186/s13068-015-0410-0>.
- [19] C. Foulquier, C.-N. Huang, N.-P.-T. Nguyen, A. Thiel, T. Wilding-Steel, J. Soula, et al., An efficient method for markerless mutant generation by allelic exchange in *Clostridium acetobutylicum* and *Clostridium saccharobutylicum* using suicide vectors, *Biotechnol. Biofuels* 12 (2019) 31, <https://doi.org/10.1186/s13068-019-1364-4>.
- [20] J.T. Heap, M. Ehsaan, C.M. Cooksley, Y.K. Ng, S.T. Cartman, K. Winzer, et al., Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker, *Nucl. Acids Res.* 40 (2012), <https://doi.org/10.1093/nar/gkr1321> e59–e59.
- [21] I. Mougialakos, E.F. Bosma, J. Ganguly, J. van der Oost, R. van Kranenburg, Hijacking CRISPR-Cas for high-throughput bacterial metabolic engineering: advances and prospects, *Curr. Opin. Biotechnol.* 50 (2018) 146–157, <https://doi.org/10.1016/j.copbio.2018.01.002>.
- [22] T. Xu, Y. Li, Z. Shi, C.L. Hemme, Y. Li, Y. Zhu, et al., Efficient genome editing in *Clostridium cellulolyticum* via CRISPR-Cas9 nickase, *Appl. Environ. Microbiol.* 81 (2015) 4423–4431, <https://doi.org/10.1128/AEM.00873-15>.
- [23] Q. Li, J. Chen, N.P. Minton, Y. Zhang, Z. Wen, J. Liu, et al., CRISPR-based genome editing and expression control systems in *Clostridium acetobutylicum* and *Clostridium beijerinckii*, *Biotechnol. J.* (2016), <https://doi.org/10.1002/biot.201600053>.
- [24] Y. Wang, Z.-T. Zhang, S.-O. Seo, P. Lynn, T. Lu, Y.-S. Jin, et al., Bacterial genome editing with CRISPR-Cas9: deletion, integration, single nucleotide modification, and desirable “clean” mutant selection in *Clostridium beijerinckii* as an example, *ACS Synth. Biol.* 5 (2016) 721–732, <https://doi.org/10.1021/acssynbio.6b00060>.
- [25] Y. Wang, Z.-T. Zhang, S.-O. Seo, K. Choi, T. Lu, Y.-S. Jin, et al., Markerless chromosomal gene deletion in *Clostridium beijerinckii* using CRISPR-Cas9 system, *J. Biotechnol.* 200 (2015) 1–5, <https://doi.org/10.1016/j.jbiotec.2015.02.005>.
- [26] Q. Li, F.M. Seys, N.P. Minton, J. Yang, Y. Jiang, W. Jiang, et al., CRISPR-Cas9 D10A nickase-assisted base editing in solvent producer *Clostridium beijerinckii*, *Biotechnol. Bioeng.* (2019), <https://doi.org/10.1002/bit.26949>.
- [27] T.E. Wilson, L.M. Topper, P.L. Palmos, Non-homologous end-joining: bacteria join the chromosome breakdance, *Trends Biochem. Sci.* 28 (2003) 62–66, [https://doi.org/10.1016/S0968-0004\(03\)00005-7](https://doi.org/10.1016/S0968-0004(03)00005-7).
- [28] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J.A. Doudna, E. Charpentier, A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity, *Science* 337 (2012) 816–821, <https://doi.org/10.1126/science.1225829>.
- [29] A.H. Hartman, H. Liu, S.B. Melville, Construction and characterization of a lactose-inducible promoter system for controlled gene expression in *Clostridium perfringens*, *Appl. Environ. Microbiol.* 77 (2011) 471–478, <https://doi.org/10.1128/AEM.01536-10>.
- [30] F. Wasels, J. Jean-Marie, F. Collas, A.M. López-Contreras, N. Lopes Ferreira, A two-plasmid inducible CRISPR/Cas9 genome editing tool for *Clostridium acetobutylicum*, *J. Microbiol. Meth.* 140 (2017) 5–11, <https://doi.org/10.1016/j.mimet.2017.06.010>.
- [31] S. Nagaraju, N.K. Davies, D.J.F. Walker, M. Köpke, S.D. Simpson, Genome editing of *Clostridium autoethanogenum* using CRISPR/Cas9, *Biotechnol. Biofuels* 9 (2016) 219, <https://doi.org/10.1186/s13068-016-0638-3>.
- [32] H. Nariya, S. Miyata, T. Kuwahara, A. Okabe, Development and characterization of a xylose-inducible gene expression system for *Clostridium perfringens*, *Appl. Environ. Microbiol.* 77 (2011) 8439–8441, <https://doi.org/10.1128/AEM.05668-11>.
- [33] J. Kolek, K. Sedlar, I. Provaznik, P. Patakova, Dam and Dem methylations prevent gene transfer into *Clostridium pasteurianum* NRRL B-598: development of methods for electrotransformation, conjugation, and sonoporation, *Biotechnol. Biofuels* 9 (2016) 14, <https://doi.org/10.1186/s13068-016-0436-y>.
- [34] H.V. Colot, G. Park, G.E. Turner, C. Ringelberg, C.M. Crew, L. Litvinkova, et al., A high-throughput gene knock-out procedure for *Neurospora* reveals functions for multiple transcription factors, *Proc. Natl. Acad. Sci.* 103 (2006) 10352–10357, <https://doi.org/10.1073/pnas.0601456103>.
- [35] R. Agatep, R.D. Kirkpatrick, D.L. Parchaliuk, R.A. Woods, R.D. Gietz, Transformation of *Saccharomyces cerevisiae* by the lithium acetate/single-stranded carrier DNA/polyethylene glycol protocol, *Tech Tips Online* 3 (1998) 133–137, [https://doi.org/10.1016/S1366-2120\(08\)70121-1](https://doi.org/10.1016/S1366-2120(08)70121-1).
- [36] J. Quan, J. Tian, Circular polymerase extension cloning of complex gene libraries and pathways, *PLoS ONE* 4 (2009) e6441, <https://doi.org/10.1371/journal.pone.0006441>.
- [37] Y. Wang, X. Li, Y. Mao, H.P. Blaschek, Genome-wide dynamic transcriptional profiling in *Clostridium beijerinckii* NCIMB 8052 using single-nucleotide resolution RNA-Seq, *BMC Genom.* 13 (2012) 102, <https://doi.org/10.1186/1471-2164-13-102>.
- [38] J.D. Outram, M. Loughlin, T.J. Swinfield, J.K. Brehm, D.E. Thompson, N.P. Minton, Introduction of plasmids into whole cells of *Clostridium acetobutylicum* by electroporation, *FEMS Microbiol. Lett.* 56 (1988) 83–88, <https://doi.org/10.1111/j.1574-6968.1988.tb03154.x>.
- [39] H. Dong, W. Tao, Y. Zhang, Y. Li, Development of an anhydrotetracycline-inducible gene expression system for solvent-producing *Clostridium acetobutylicum*: a useful tool for strain engineering, *Metab. Eng.* 14 (2012) 59–67, <https://doi.org/10.1016/j.jmben.2011.10.004>.
- [40] L.D. Mermelstein, E.T. Papoutsakis, In vivo methylation in *Escherichia coli* by the *Bacillus subtilis* phage phi 3T I methyltransferase to protect plasmids from restriction upon transformation of *Clostridium acetobutylicum* ATCC 824, *Appl. Environ. Microbiol.* 59 (1993) 1077–1081.
- [41] O. Mutschlechner, H. Swoboda, J.R. Gapes, Continuous two-stage ABE-fermentation using *Clostridium beijerinckii* NRRL B592 operating with a growth rate in the first stage vessel close to its maximal value, *J. Mol. Microbiol. Biotechnol.* 2 (2000) 101–105.
- [42] W. Kuit, N.P. Minton, A.M. López-Contreras, G. Eggink, Disruption of the acetate kinase (ack) gene of *Clostridium acetobutylicum* results in delayed acetate production, *Appl. Microbiol. Biotechnol.* 94 (2012) 729–741, <https://doi.org/10.1007/s00253-011-3848-4>.
- [43] J. Zhang, W. Zong, W. Hong, Z.-T. Zhang, Y. Wang, Exploiting endogenous CRISPR-Cas system for multiplex genome editing in *Clostridium tyrobutyricum* and engineer the strain for high-level butanol production, *Metab. Eng.* (2018), <https://doi.org/10.1016/j.jmben.2018.03.007>.
- [44] I. Barák, J. Behari, G. Olmedo, P. Guzmán, D.P. Brown, E. Castro, et al., Structure and function of the *Bacillus* SpoIIIE protein and its localization to sites of sporulation septum assembly, *Mol. Microbiol.* 19 (1996) 1047–1060, <https://doi.org/10.1046/j.1365-2958.1996.433963.x>.
- [45] M.A. Al-Hinai, S.W. Jones, E.T. Papoutsakis, The *Clostridium* sporulation programs: diversity and preservation of endospore differentiation, *Microbiol. Mol. Biol. Rev.* 79 (2015) 19–37, <https://doi.org/10.1128/MMBR.00025-14>.
- [46] M.C. Scotcher, G.N. Bennett, SpoIIIE regulates sporulation but does not directly affect solventogenesis in *Clostridium acetobutylicum* ATCC 824, *J. Bacteriol.* 187 (2005) 1930–1936, <https://doi.org/10.1128/JB.187.6.1930-1936.2005>.
- [47] C. Bi, S.W. Jones, D.R. Hess, B.P. Tracy, E.T. Papoutsakis, SpoIIIE is necessary for asymmetric division, sporulation, and expression of sigmaF, sigmaE, and sigmaG but does not control solvent production in *Clostridium acetobutylicum* ATCC 824, *J. Bacteriol.* 193 (2011) 5130–5137, <https://doi.org/10.1128/JB.05474-11>.
- [48] A.M. Lopez-Contreras, H. Smidt, W.M. de Vos, H. Mooibroek, P.A.M. Claassen, J. van der Oost, *Clostridium beijerinckii* cells expressing *Neocallimastix patriciarum* glycoside hydrolases show enhanced lichenan utilization and solvent production, *Appl. Environ. Microbiol.* 67 (2002) 5127–5133, <https://doi.org/10.1128/aem.67.11.5127-5133.2001>.
- [49] S. Denman, G.P. Xue, B. Patel, Characterization of a *Neocallimastix patriciarum* cellulase cDNA (celA) homologous to *Trichoderma reesei* cellobiohydrolase II, *Appl. Environ. Microbiol.* 62 (1996) 1889–1896.
- [50] R. Hocq, M. Bouilloux-Lafont, N.L. Ferreira, F. Wasels, $\sigma 54$ (σL) plays a central role in carbon metabolism in the industrially relevant *Clostridium beijerinckii*, *BioRxiv*, 2018, pp. 505099, <https://doi.org/10.1101/505099>.
- [51] K. Schwarz, Y. Zhang, W. Kuit, M. Ehsaan, K. Kovács, K. Winzer, et al., New tools for the genetic modification of industrial Clostridia, *Direct Microb. Convers. Biomass to Adv. Biofuels*, Elsevier, 2015, pp. 241–289, <https://doi.org/10.1016/B978-0-444-59592-8.00013-0>.
- [52] C. Steffen, H. Matzura, Nucleotide sequence analysis and expression studies of a chloramphenicol-acetyltransferase-coding gene from *Clostridium perfringens*, *Gene* 75 (1989) 349–354, [https://doi.org/10.1016/0378-1119\(89\)90282-5](https://doi.org/10.1016/0378-1119(89)90282-5).
- [53] H. Máté de Gérand, F. Wasels, A. Bisson, B. Clement, F. Bidard, E. Jourdir, et al., Genome and transcriptome of the natural isopropanol producer *Clostridium beijerinckii* DSM6423, *BMC Genom.* 19 (2018) 242, <https://doi.org/10.1186/s12864-018-4636-7>.
- [54] J.T. Heap, M. Ehsaan, C.M. Cooksley, Y.-K.Y.-K. Ng, S.T. Cartman, K. Winzer, et al., Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker, *Nucleic Acids Res.* 40 (2012) e59, <https://doi.org/10.1093/nar/gkr1321>.
- [55] I. Barák, P. Youngman, SpoIIIE mutants of *Bacillus subtilis* comprise two distinct phenotypic classes consistent with a dual functional role for the SpoIIIE protein, *J. Bacteriol.* 178 (1996) 4984–4989, <https://doi.org/10.1128/JB.178.16.4984-4989.1996>.
- [56] J. Kolek, M. Diallo, M. Vasyukivska, B. Branska, K. Sedlar, A.M. López-Contreras, et al., Comparison of expression of key sporulation, solventogenic and acetogenic genes in *C. beijerinckii* NRRL B-598 and its mutant strain overexpressing spo0A, *Appl. Microbiol. Biotechnol.* 101 (2017) 8279–8291, <https://doi.org/10.1007/s00253-017-8555-3>.
- [57] B.P. Tracy, S.W. Jones, E.T. Papoutsakis, Inactivation of σE and σG in *Clostridium acetobutylicum* illuminates their roles in clostridial-cell-form biogenesis, granulose synthesis, solventogenesis, and spore morphogenesis, *J. Bacteriol.* 193 (2011) 1414–1426, <https://doi.org/10.1128/JB.01380-10>.
- [58] K. Carniol, S. Ben-Yehuda, N. King, R. Losick, Genetic dissection of the sporulation protein SpoIIIE and its role in asymmetric division in *Bacillus subtilis*, *J. Bacteriol.* 187 (2005) 3511–3520, <https://doi.org/10.1128/JB.187.10.3511-3520.2005>.
- [59] F. Mingardon, A. Chanal, A.M. Lopez-Contreras, C. Dray, E.A. Bayer, H.-P. Fierobe, Incorporation of fungal cellulases in bacterial minicelluloses yields viable, synergistically acting cellulolytic complexes, *Appl. Environ. Microbiol.* 73 (2007) 3822–3832, <https://doi.org/10.1128/AEM.00398-07>.
- [60] A. Poehelein, J.D.M. Solano, S.K. Flitsch, P. Krabben, K. Winzer, S.J. Reid, et al., Microbial solvent formation revisited by comparative genome analysis, *Biotechnol. Biofuels* 10 (2017) 58, <https://doi.org/10.1186/s13068-017-0742-z>.
- [61] Y. Gu, J. Feng, Z.T. Zhang, S. Wang, L. Guo, Y. Wang, et al., Curing the endogenous megaplasmid in *Clostridium saccharoperbutylacetonicum* N1–4 (HMT) using CRISPR-Cas9 and preliminary investigation of the role of the plasmid for the strain metabolism, *Fuel* 236 (2019) 1559–1566, <https://doi.org/10.1016/j.fuel.2018.09.030>.

Construction of pE_Xcas9

P1_xyIB_F TATTACGCCAGCTGGCGAAAGGGGGATG AAATAAAAGTTATTTTGCATTGACAAAG
 Cas9_xyIB_F GTACTTTTATCCAT CCCTCCTTGAATGCCACTTCATAACTATCG
 xyIB_cas9_F GGCATTCAAGGAGGG ATGGATAAAAAGTACAGTATTGGTCTAG
 16-P2_cas9R GCTGGCCTTTTGCTCACATGTTT TTTCTAAAAAATTCTTAATCTCCGCCTAGTTGAC

Construction of pS_X^RKR

P1_xyIR_R TATTACGCCAGCTGGCGAAAGGGGGATG AAATAAAAGTTATTTTGCATTGACAAAG
 bsaI_K_XyIB_R CATAACACCCCTTGTAGGTCTC CCTCCTTGAATGCCACTTCATAAC
 bsaI_X_kan_F CATAACACCCCTTGTAGGTCTC CCTCCTTGAATGCCACTTCATAAC
 bsaI_R_Kan_R CTAGCTCTAAAACCGAGACC CCAATTCTGATTAGAAAAACTC
 bsaI_traCRNA_F AGAATTGGGGTCTCGGTTTTAGA GCTAGAAAATAGCAAGTAAAATAAG
 MCS_Term_R AGATCTCGAGGCCTGCAGGCATGCAAGCTTGAATTTTTTAAAAAATAACTCTGTAG

Construction of pS_X^RKR_celA

P1_hbd_F AAAAAATCAAGCTTGCATGCCTGCAGGCCATTTTAGCTAAGGATATCTATAATG
 0326_celA_R AGTTATTTATATATTTCCAI GATTACGAATTCGAGCTCGGTACCC
 celA_0326_F CGAATTCGTAATCATG GAAATATATAAATACTACTAGATTACATAAATAAG
 P2_0326_R CGCGTCCATGGAGATCTCGAGGCCTGCAGG AGATGCTCCTGTCTCCTGTATTTT
 CelA_S1_Forw GGAG TATGGAAAACTCAAGTTTATGCTT
 CelA_S1_Rev AAAC AAGCATAAACTTGAGTTTTTCCATA

Construction of pS_X^R_ΔspolIE

P1_Trna_F TATTTTTAAAAAATTCAGCTTGCATGC TCGCGGGTGGAGCAGTTGGTAGC
 16_0098_TRNA_R TTTTGAAACTACTAC TTCCAAATAATGTAACCTGATTATAGCATC
 16_TRNA_cbei_0098_F TTACATTATTTGGAA GTAGTAGTTTCAAAGTTTATTTAGATAGC
 P2_0098_R TACGAATTCGAGCTCGGTACCCGGGGATCC ATAAACCAGCGGATAATTTTTCAAGG
 spolIE_S1_Forw GGAG ATTAATACTAATTACTTCTTTTCTT
 spolIE_S1_Rev AAAC AAGAAAAGAAGTAATTAGTATTAAT

Construction of pSpolIE

Cbei0097prom_F GACGTTGTAAAACGACGGCCAGTGCCAAGC ACGAATTTTGGTCACTTAAGCCAAG
 Cbei0097_R AGTCGACGTCACGCGTCCATGGAGATCTCG CTTATTGCATGATAATCTTAGCTATC

Construction of pCas9_{ind}-ΔcatB

ΔcatB_fwd TGTTATGGATTATAAGCGGCTCGAG GACGTCAAACCATGTTAATCATTGC
 ΔcatB_rev AATCTATCACTGATAGGGACTCGAG CAATTTCCACAAAGAATTCGCTAGC

Construction of pCas9_{ind}-ΔpNF2

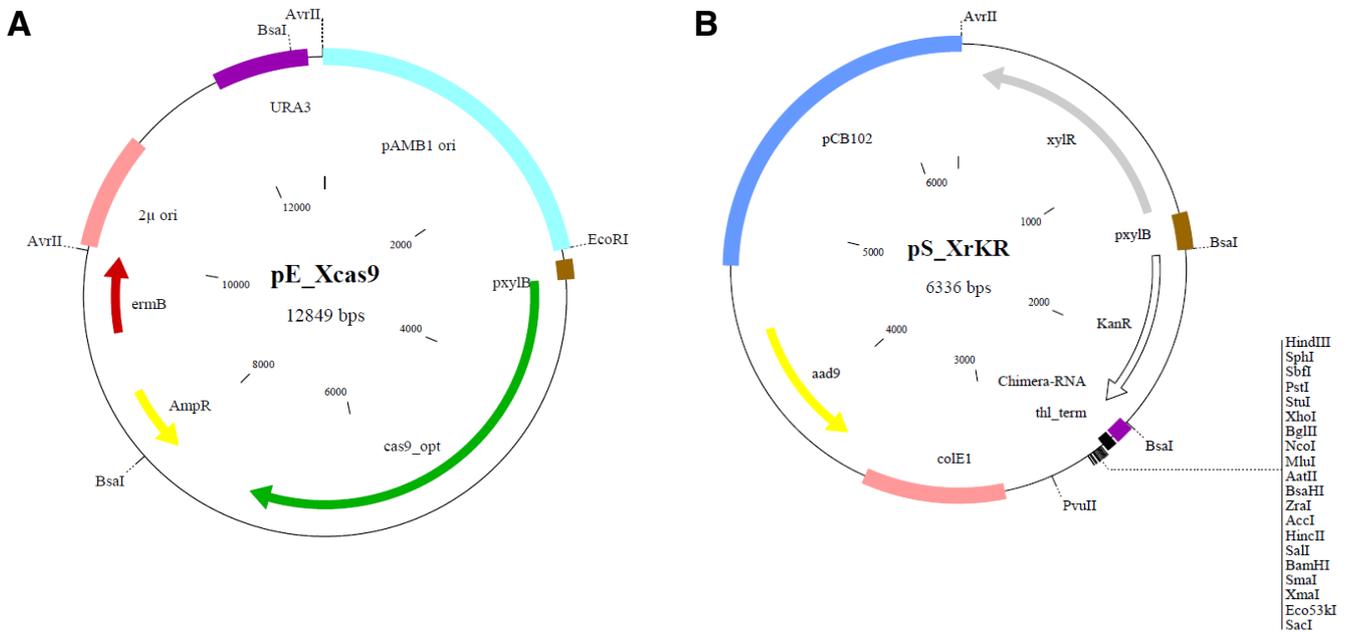
pNF2-gRNA GAGCTCACTCTATCATTGATAGAGTTTGAACCTCTATCATTGATAGAGTATAATATCTTTGTTTCATTTAAG
 CCATCTACTAAACAAGTTTTAGAGCTAGAAATAGCAAGTAAAATAAGGCTAGTCCGTTATCAACTTGAA
 AAAGTGGCACCGAGTCGGTCTTTTTTTGAAGCTTGAGCTC

Primer used for screening

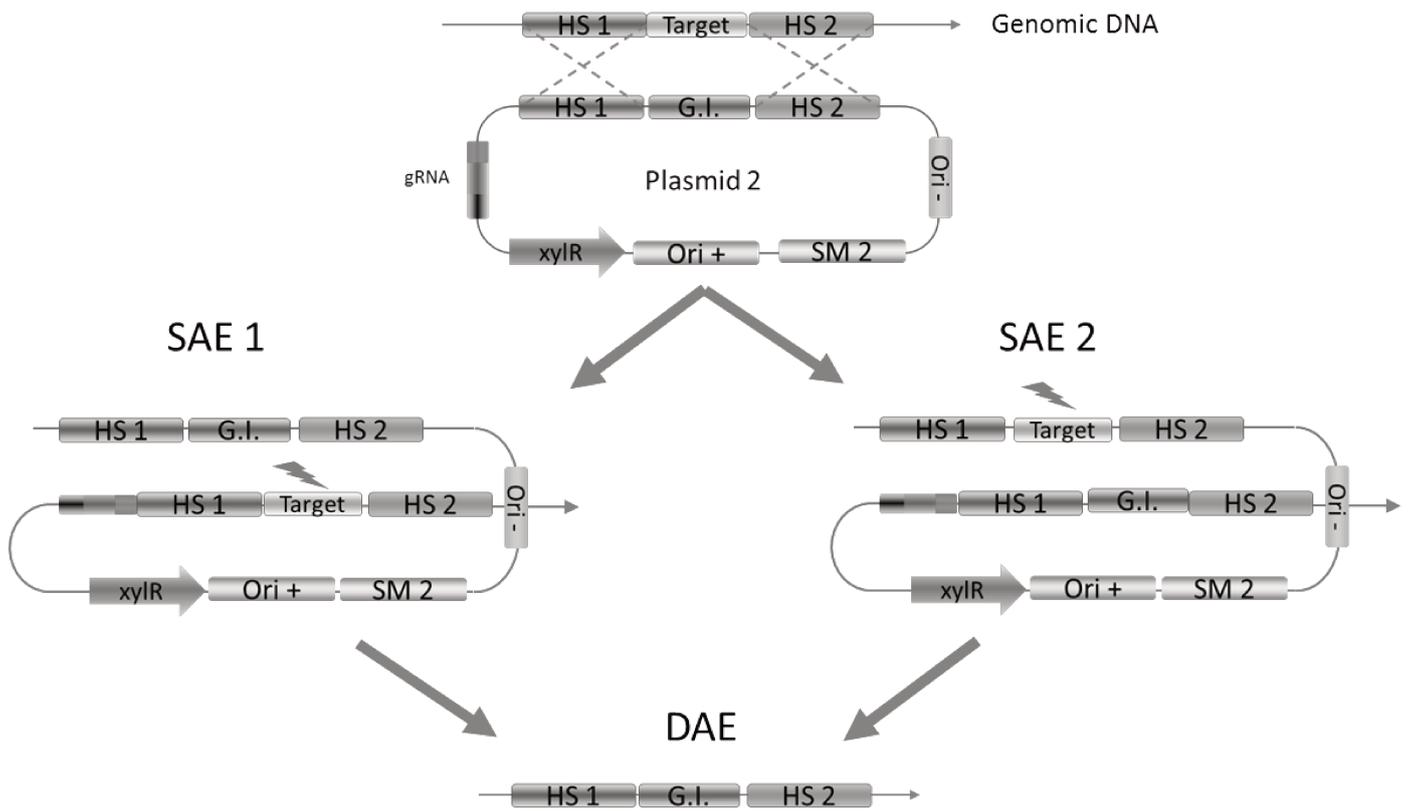
Cbei_325_F (*celA*) ATAAGTAACAATTAGATAATTATGAAGTTAATCCTTAG
 Cbei_326_R (*celA*) CATTGCTTTTCAGGTCTTCTTTTGCTG
 Cbei_0096_F (*spolIE*) CGAAGATATTATGTCTAAGTTTCTA
 Cbei_0098_R (*spolIE*) CATTACATTCCATACAATATTTATTGTATAAACCAGC
 RH076 (*catB*) CATATAATAAAAGGAAACCTCTTGATCG
 RH077 (*catB*) ATTGCCAGCCTAACACTTGG
 pNF2_fwd (pNF2) GGGCGCACTTATACACCACC
 pNF2_rev (pNF2) TGCTACGCACCCCTAAAGG

Bold: restriction sites (**BsaI**); Underlined: non-specific tail;

**Supplementary file S1: list of primers used
 in this study**



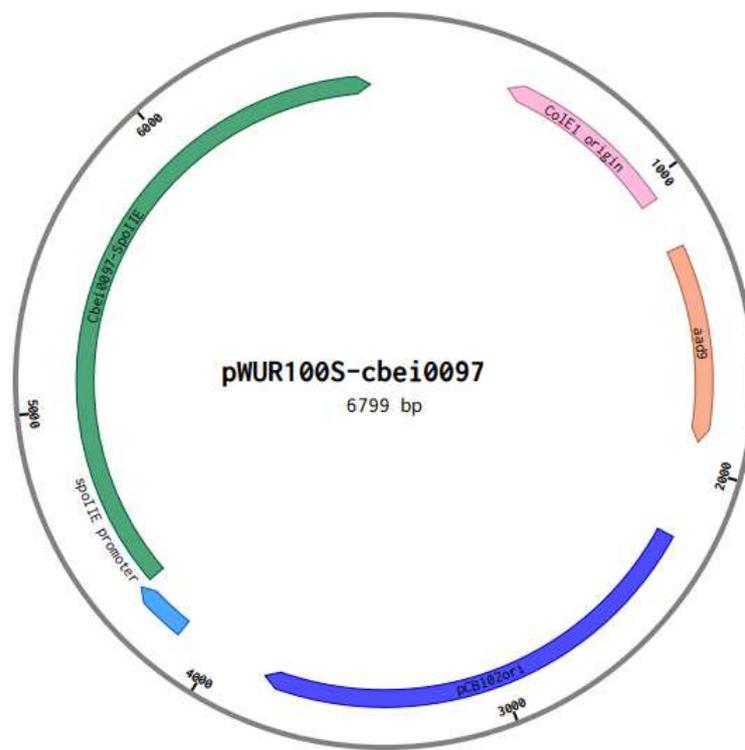
Supplementary file S2 : Maps of the plasmids pE_X_cas9 (A) and pS_X^RKR (B). pAMB β 1 replication origin; xylBp: promoter of the operon xylAB from *C. difficile*; cas9_{opt}: codon optimized *cas9* gene; colE1, gram negative replicon; *amp^R*: gene resistance ampicillin; *ermB*: gene resistance erythromycin; 2 μ : yeast replicon 2 μ ; *xylR*: gene of the regulation protein of the pxyIB; *kan^R*: kanamycin resistance gene; Chimera: minimum sequence for gRNA (26); thl^T: terminator of the gene *thl* (*cbei_3630*). MCS: multi-cloning site; *aad9*: spectinomycin resistance gene, pCB102: replication origin.



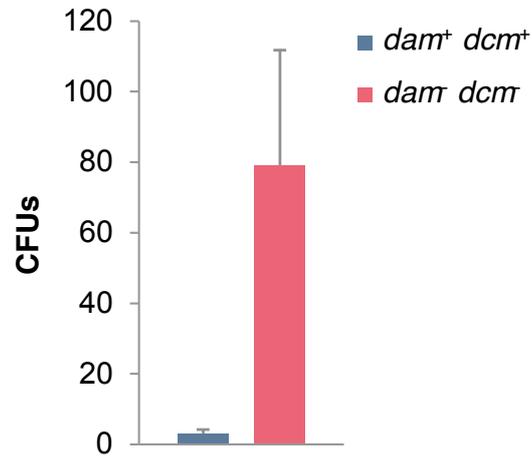
Supplementary file S3 : Allelic exchange events between the genomic DNA and editing DNA template. In our two plasmid system, the plasmid 2 harbours the gRNA directed against the target on the genomic DNA, the regulator of the promoter *xyIB* from *C. difficile* sp. 630 and the editing DNA template consisting in the 2 sequences flanking the target in the genome (HS1 and HS2, Homologous Sequences) and a gene of interest (G.I.) inserted between the two homologous sequences. SAE: Single Allelic Exchange and DAE Double Allelic Exchange. In cells with a SAE, the target site is still present in the genome whereas is disappeared in the DAE cells. Strain harbouring SAE will then be sensitive to the activated CRISPR-Cas9 system

	concentration (g.L ⁻¹)					
	Acetate	Butyrate	Acetone	Butanol	Ethanol	Total ABE
wild-type	0.4 ± 0.0	0.1 ± 0.0	2.0 ± 0.5	7.7 ± 0.6	n.d	9.7 ± 0.6
ΔspolIE	0.4 ± 0.0	0.8 ± 0.0	2.7 ± 0.3	7.0 ± 0.1	n.d	9.7 ± 0.3

Supplementary file S2: Final products concentration in g.L⁻¹ after a 72 h-fermentation experiments in 100 mL bottles of *C. beijerinckii* NCIMB 8052 wild type and ΔspolIE strain. Values are given ± SEM of biological duplicates. Fermentations were performed in CM2 medium with 50 g glucose.



Supplementary file S6: Maps of the pSpolIE plasmids: *colE1*, Gram negative replicon, *aad9*: spectinomycin resistance gene, pCB102: Gram positive replication origin.



Supplementary file S7: Comparison of transformation efficiencies in *C. beijerinckii* DSM 6423. 20 μ g of the pCas9_{ind} was used with and without *E. coli* Dam and Dcm methylations in triplicate experiments. Error bars represent the standard error of the mean (s.e.m.) of these replicates.

	concentration (g.L ⁻¹)				
	ethanol	acetone	isopropanol	butanol	total
wild-type	0.1 ± 0.0	0.2 ± 0.0	4.1 ± 0.1	7.0 ± 0.1	11.5 ± 0.1
ΔcatB	0.1 ± 0.0	0.1 ± 0.0	3.9 ± 0.1	6.8 ± 0.2	10.9 ± 0.3
ΔcatB ΔpNF2	0.1 ± 0.1	0.1 ± 0.0	4.5 ± 0.1	7.8 ± 0.5	12.5 ± 0.5

Supplementary file S8 : Final solvent concentration of 72h-fermentation experiments of *C. beijerinckii* DSM 6423 and derivative strains. Values are given ± SEM of biological triplicates. Fermentations were performed in Gapes medium (per litre: 2.5 g yeast extract, 1 g KH₂PO₄, 0.6 g K₂HPO₄, 1 g MgSO₄ 7H₂O, 6.6 mg FeSO₄ 7H₂O, 100 mg p-aminobenzoic acid, 2.9 g ammonium acetate, 60 g glucose).

3.3 PROTECTION INTELLECTUELLE (BREVETS)

L'originalité et le caractère inventif des outils génétiques développés au cours de cette thèse a permis le dépôt de deux brevets de perfectionnement. Le premier couvre l'obtention d'une souche $\Delta catB$ et son utilisation pour un système CRISPR-Cas9 à deux plasmides précédemment brevetés.

Son application est exemplifiée par :

- L'inactivation du gène *catB* par une approche à plasmide unique ;
- L'éviction du plasmide pNF2 (sans caractérisation du phénotype) ;
- L'obtention d'un mutant Δupp avec le système CRISPR/anti-CRISPR.

Le second brevet protège et décrit le phénotype de la souche *C. beijerinckii* DSM 6423 ou ses dérivées, dépourvu du plasmide pNF2 (souches $\Delta pNF2$), ainsi que la suite de plasmides basés sur le pNF2 (pNF3C, pNF3E, pNF3S). Son application est exemplifiée par :

- Les efficacités de transformation de différents plasmides dans la souche pNF2 ;
- Les efficacités de transformation des différents pNF3, dans les souches DSM 6423 et NCIMB 8052, mais également chez *C. acetobutylicum*, *C. ljungdahlii* et *B. subtilis*.

Abrégé

BACTERIES CLOSTRIDIUM GENETIQUEMENT MODIFIEES, PREPARATION ET UTILISATIONS DE CELLES-CI

La présente invention concerne la modification génétique de bactéries du genre *Clostridium*, typiquement de bactéries solvantogènes du genre *Clostridium*, en particulier de bactéries possédant à l'état sauvage un gène codant une amphénicol-O-acetyltransférase. Elle concerne ainsi des méthodes, outils et kits permettant une telle modification génétique, en particulier l'élimination ou la modification d'une séquence codant ou contrôlant la transcription d'une amphénicol-O-acetyltransférase, les bactéries génétiquement modifiées obtenues et leurs utilisations, en particulier pour produire un solvant, de préférence à l'échelle industrielle.

Revendications

- [Revendication 1] Acide nucléique reconnaissant i) une séquence codant une amphénicol-O-acetyltransférase, ii) une séquence contrôlant la transcription d'une séquence codant une amphénicol-O-acetyltransférase, ou iii) une séquence flanquant la séquence codant une amphénicol-O-acetyltransférase, au sein du génome d'une bactérie du genre *Clostridium*.
- [Revendication 2] Acide nucléique selon la revendication 1, caractérisé en ce que la séquence reconnue est le gène *catB* de séquence SEQ ID NO : 18 ou une séquence identique à au moins 70% à celle-ci, ou une séquence comprenant entre 1 et 1000 nucléotides, ou paire de séquences dont chaque séquence comprend entre 1 et 1000 nucléotides, flanquant la séquence SEQ ID NO : 18 ou une séquence identique à au moins 70% à celle-ci.
- [Revendication 3] Acide nucléique selon la revendication 1 ou 2, caractérisé en ce que ledit acide nucléique est sélectionné parmi une cassette d'expression et un vecteur, de préférence un plasmide.
- [Revendication 4] Acide nucléique selon l'une quelconque des revendications 1 à 3, caractérisé en ce que l'acide nucléique comprend un ARN guide (ARNg) et/ou une matrice de modification.
- [Revendication 5] Acide nucléique selon l'une des revendications 1 à 4, caractérisé en ce que la bactérie *Clostridium* est une bactérie capable de produire de l'isopropanol à l'état sauvage.
- [Revendication 6] Acide nucléique selon l'une des revendications 1 à 5, caractérisé en ce que la bactérie *Clostridium* est une bactérie *C.beijerinckii* dont le sous-clade est sélectionnée parmi DSM 6423, LMG 7814, LMG 7815, NRRL B-593, NCCB 27006 et un sous-clade présentant au moins 95% d'identité avec la souche DSM6423.
- [Revendication 7] Acide nucléique selon l'une des revendications 3 à 6, caractérisé en ce qu'il s'agit du plasmide pCas9ind- Δ *catB* de séquence SEQ ID NO : 21 ou du plasmide pCas9ind-gRNA_*catB* de séquence SEQ ID NO : 38.
- [Revendication 8] Utilisation d'un acide nucléique selon l'une des revendications 3 à 7 pour transformer et/ou modifier génétiquement une bactérie *Clostridium* capable de produire de l'isopropanol à l'état sauvage.
- [Revendication 9] Utilisation d'un acide nucléique selon l'une des revendications 3 à 7, ledit acide nucléique ne présentant pas de méthylation aux niveaux des motifs reconnus par les méthyltransférases de type Dam et Dcm, pour

transformer un sous-clade *C. beijerinckii* sélectionné parmi DSM 6423, LMG 7814, LMG 7815, NRRL B-593, NCCB 27006 et un sous-clade présentant au moins 95%, de préférence 97%, d'identité avec la souche DSM 6423.

- [Revendication 10] Procédé pour transformer, et de préférence modifier génétiquement, une bactérie du genre *Clostridium* à l'aide d'un outil de modification génétique, caractérisé en ce qu'il comprend une étape de transformation de la bactérie par introduction dans ladite bactérie d'un acide nucléique selon l'une des revendications 1 à 7.
- [Revendication 11] Procédé selon la revendication 10, caractérisé en ce que la bactérie est transformée à l'aide d'un outil CRISPR utilisant une enzyme responsable de la coupure d'au moins un brin de la séquence cible codant ou contrôlant la transcription d'une amphénicol-O-acetyltransférase.
- [Revendication 12] Bactérie du genre *Clostridium* modifiée génétiquement obtenu(e) à l'aide du procédé selon la revendication 10 ou 11.
- [Revendication 13] Bactérie *C. beijerinckii* DSM6423 $\Delta catB$ déposée sous le numéro LMG P-31151 ou une version génétiquement modifiée de celle-ci.
- [Revendication 14] Utilisation de la bactérie modifiée génétiquement selon la revendication 12, ou de la bactérie *C. beijerinckii* DSM6423 $\Delta catB$ déposée sous le numéro LMG P-31151 ou d'une version génétiquement modifiée de celle-ci selon revendication 13, pour produire un solvant, de préférence l'isopropanol, ou un mélange de solvants, de préférence à l'échelle industrielle.
- [Revendication 15] Kit comprenant (i) un acide nucléique selon l'une des revendications 3 à 7 et (ii) au moins un outil sélectionné parmi les éléments d'un outil de modification génétique ; un acide nucléique en tant qu'ARNg ; un acide nucléique en tant que matrice de réparation ; au moins une paire d'amorces ; et un inducteur permettant l'expression d'une protéine codée par ledit outil.

Abrégé

OUTIL GENETIQUE OPTIMISÉ POUR MODIFIER LES BACTERIES

La présente invention concerne la transformation et la modification génétique de bactéries appartenant au phylum des Firmicutes. Elle concerne ainsi des méthodes, outils et kits permettant une telle modification génétique, impliquant en particulier une séquence d'acide nucléique utilisée pour faciliter la transformation de la bactérie, ladite séquence comprenant i) tout ou partie de la séquence SEQ ID NO : 126 et ii) une séquence permettant la modification du matériel génétique d'une bactérie et/ou l'expression au sein de ladite bactérie d'une séquence d'ADN partiellement ou totalement absente du matériel génétique présent au sein de la version sauvage de ladite bactérie. La description vise aussi les bactéries génétiquement modifiées obtenues et leurs utilisations, en particulier pour produire un solvant, de préférence à l'échelle industrielle.

Revendications

- [Revendication 1] Acide nucléique comprenant i) tout ou partie de la séquence SEQ ID NO : 126 et ii) une séquence permettant la modification du matériel génétique d'une bactérie et/ou l'expression au sein de ladite bactérie d'une séquence d'ADN partiellement ou totalement absente du matériel génétique présent au sein de la version sauvage de ladite bactérie.
- [Revendication 2] Acide nucléique selon la revendication 1, caractérisé en ce que la séquence permettant la modification du matériel génétique de la bactérie est une matrice de modification permettant, par un mécanisme de recombinaison homologue, le remplacement d'une portion du matériel génétique de la bactérie par une séquence d'intérêt.
- [Revendication 3] Acide nucléique selon la revendication 1 ou 2, caractérisé en ce que l'acide nucléique comprend en outre iii) une séquence codant une endonucléase d'ADN et/ou iv) un ou plusieurs ARN guides (ARNg), chaque ARNg comprenant une structure ARN de fixation à l'endonucléase d'ADN et une séquence complémentaire de la portion ciblée du matériel génétique de la bactérie.
- [Revendication 4] Acide nucléique selon l'une quelconque des revendications 1 à 3, caractérisé en ce que ledit acide nucléique est sélectionné parmi une cassette d'expression et un vecteur, et est de préférence un plasmide, par exemple un plasmide ayant une séquence sélectionnée parmi SEQ ID NO : 119, SEQ ID NO : 123, SEQ ID NO : 124 et SEQ ID NO : 125.
- [Revendication 5] Outil génétique pour transformer et modifier génétiquement une bactérie, caractérisé en ce que ledit outil génétique comprend au moins :
- un premier acide nucléique codant au moins une endonucléase d'ADN, dans lequel la séquence codant l'endonucléase d'ADN est placée sous le contrôle d'un promoteur, et
 - un deuxième acide nucléique tel que décrit dans l'une des quelconque revendications 1 à 4, et

en ce que au moins l'un desdits acide nucléique de l'outil génétique comprend en outre de préférence une séquence codant une protéine anti-CRISPR placée sous le contrôle d'un promoteur inductible, ou en ce que l'outil génétique comprend en outre un troisième acide nucléique codant une protéine anti-CRISPR placée sous le contrôle d'un promoteur inductible.

- [Revendication 6] Outil génétique selon la revendication 5, caractérisé en ce que le premier acide nucléique code en outre un ou plusieurs ARN guides (ARNg) ou en ce que l'outil génétique comprend en outre un ou plusieurs ARNg.
- [Revendication 7] Utilisation d'un acide nucléique tel que décrit dans l'une quelconque des revendications 1 à 4 ou d'un outil génétique tel que décrit dans l'une des revendications 5 ou 6, pour transformer et éventuellement modifier génétiquement une bactérie.
- [Revendication 8] Procédé pour transformer, et de préférence modifier génétiquement, une bactérie à l'aide d'un outil de modification génétique, caractérisé en ce qu'il comprend une étape de transformation de la bactérie par introduction dans ladite bactérie d'un acide nucléique selon l'une quelconque des revendications 1 à 4.
- [Revendication 9] Acide nucléique selon l'une quelconque des revendications 1 à 4, outil génétique selon la revendication 5 ou 6, utilisation selon la revendication 7, ou procédé selon la revendication 8, caractérisé(e) en ce que la bactérie appartient au phylum des Firmicutes, et appartient de préférence au genre *Clostridium*, au genre *Bacillus* ou au genre *Lactobacillus*.
- [Revendication 10] Acide nucléique, outil génétique, utilisation ou procédé selon la revendication 9, caractérisé en ce que la bactérie est une bactérie du genre *Clostridium*, de préférence une bactérie solvantogène sélectionnée parmi *C. acetobutylicum*, *C. cellulolyticum*, *C. phytofermentans*, *C. beijerinckii*, *C. saccharobutylicum*, *C. saccharoperbutylacetonicum*, *C. sporogenes*, *C. butyricum*, *C. aurantibutyricum*, *C. tyrobutyricum* ou une bactérie acétogène sélectionnée parmi *C. aceticum*, *C. thermoaceticum*, *C. ljungdahlii*, *C. autoethanogenum*, *C. difficile*, *C. scatologenes* et *C. carboxydivorans*.
- [Revendication 11] Acide nucléique, outil génétique, utilisation ou procédé selon la revendication 9 ou 10, caractérisé en ce que la bactérie est une bactérie *C. beijerinckii* dépourvue du plasmide pNF2, de préférence un sous-clade sélectionné parmi DSM 6423, LMG 7814, LMG 7815, NRRL B-593, NCCB 27006 et un sous-clade présentant au moins 95%, de préférence 97%, d'identité avec la souche DSM 6423.
- [Revendication 12] Acide nucléique, outil génétique, utilisation ou procédé selon la revendication 10 ou 11, caractérisé en ce que la bactérie est la bactérie *C. beijerinckii* IFP963 $\Delta catB \Delta pNF2$ déposée sous le numéro LMG P-31277 ou une version génétiquement modifiée de celle-ci.
- [Revendication 13] Kit pour transformer et de préférence modifier génétiquement une

bactérie ou pour produire au moins un solvant à l'aide d'une bactérie, ledit kit comprenant un acide nucléique tel que décrit dans l'une quelconque des revendications 1 à 4 et au moins un inducteur adapté au promoteur inductible de l'expression de la protéine anti-CRISPR sélectionné utilisé au sein de l'outil génétique tel que décrit dans l'une des revendications 5 ou 6.

[Revendication 14] Utilisation d'un acide nucléique tel que décrit dans l'une quelconque des revendications 1 à 4, de l'outil génétique selon l'une des revendications 5 ou 6, du procédé selon la revendication 8 ou d'un kit selon la revendication 13, pour permettre la production d'un solvant ou d'un mélange de solvants à l'échelle industrielle, de préférence d'acétone, de butanol, d'éthanol, d'isopropanol ou d'un mélange de ceux-ci, typiquement d'un mélange isopropanol/butanol.

[Revendication 15] Bactérie *C. beijerinckii* IFP963 $\Delta catB$ $\Delta pNF2$ déposée sous le numéro LMG P-31277 ou une version génétiquement modifiée de celle-ci.

4 PARTIE II : ROLE DE σ^{54} DANS LE METABOLISME DE *CLOSTRIDIUM BEIJERINCKII*

Le développement d'outils génétiques performants est une étape clé pour l'obtention de souches solvantogènes compatibles avec l'industrie. La modification raisonnée d'un génome bactérien nécessite toutefois également de connaître son fonctionnement, pour pouvoir déterminer avec justesse les modifications à apporter. Cette partie de la thèse s'inscrit dans cette logique en caractérisant un facteur moléculaire important pour la solvantogénèse et pour l'assimilation de sucres alternatifs : σ^{54} . Les travaux décrits dans la suite ont été publiés dans la revue *Scientific Reports*²⁶⁸ :

Hocq, R. et al. σ^{54} (σ^L) plays a central role in carbon metabolism in the industrially relevant *C. beijerinckii*. *Scientific reports* 9.1 (2019): 7228.

4.1 RESUME DES TRAVAUX PUBLIES

La solvantogénèse, processus biologique qui caractérise des souches de *Clostridium* d'intérêt industriel, est bien décrite puisque l'utilisation industrielle de tels microorganismes remonte au début du XXème siècle. Cependant, les facteurs régulant la solvantogénèse au niveau moléculaire restent malgré tout méconnus, en particulier chez *C. beijerinckii*. Cet aspect revêt une importance particulière, puisqu'une caractérisation plus poussée des voies de régulation pourrait permettre d'apporter à des pistes d'amélioration des performances fermentaires modestes des souches sauvages.

Dans une étude antérieure⁴⁷, notre groupe a rapporté l'obtention d'un mutant de la souche *C. beijerinckii* DSM 6423 qui présentait une déficience importante de sa production de solvants (incapacité à produire de l'isopropanol, capacité limitée à produire du butanol). Ce mutant, nommé AA, a été obtenu par mutagénèse aléatoire

en présence d'alcool allylique, un composé qui conduit à une molécule toxique, l'acroléine, lorsqu'elle est métabolisée par l'action d'une alcool déshydrogénase.

Afin de caractériser les facteurs moléculaires responsables de ce phénotype particulier, le génome de cette souche a été séquencé, permettant d'identifier 15 mutations ponctuelles au sein de son génome. Parmi celles-ci, une mutation du gène CIBE_0767 (*sigL*), qui encode le facteur sigma 54 (σ^{54}), a retenu notre attention. En effet, les facteurs sigma font partie d'un système de régulation largement adopté par les procaryotes pour réguler l'initiation de la transcription, et donc l'expression génétique (pour plus de détails, cf. sous-partie 1.4.2.2). La mutation S \rightarrow F, en position 366, affecte le motif HTH (*helix-turn-helix*), une séquence protéique hautement conservée et impliquée dans la reconnaissance du promoteur (position -12)²⁶⁹. Par ailleurs, le motif reconnu par l'ARN polymérase complexée avec le facteur sigma 54 (décrit dans la littérature²⁷⁰) est présent pour plusieurs souches de *C. beijerinckii* en amont de gènes clés du métabolisme solvantogène (alcool déshydrogénases, notamment). Tous ces éléments faisaient donc de σ^{54} un candidat idéal pour une validation fonctionnelle.

Nous avons donc vérifié l'impact fonctionnel de ce gène en complétant la souche AA avec un plasmide qui permet l'expression d'un σ^{54} sauvage. La complémentation restaure effectivement un phénotype sauvage. L'effet de la complémentation sur l'expression de plusieurs alcools déshydrogénases a également été vérifié par qPCR. Par ailleurs, l'analyse des promoteurs contrôlés par σ^{54} laissait sous-entendre qu'un nombre potentiellement important de gènes liés au métabolisme des sucres étaient régulés par ce gène. Des essais sur boîte de Petri en milieu minimum ont permis de démontrer que la souche AA avait également une déficience pour l'assimilation du lactose et du cellobiose. Après complémentation, ce phénotype disparaît également.

Ces éléments permettaient de démontrer que la mutation ponctuelle du gène *sigL* était fortement impliquée dans les voies de régulation du métabolisme carboné de la souche DSM 6423, sans toutefois caractériser l'effet exact de la mutation pour σ^{54} . En sachant que cette mutation devait affecter la reconnaissance du promoteur, nous avons formulé l'hypothèse que cette mutation devait inactiver la protéine. Pour la vérifier, nous avons supprimé le gène *sigL* dans *C. beijerinckii* NCIMB 8052 (nommé $\Delta sigL$) par la méthode CRISPR-Cas9, ce qui devait permettre par ailleurs de vérifier

nos résultats préliminaires dans la souche modèle de *C. beijerinckii*. Comme attendu, le mutant $\Delta sigL$ obtenu présentait un phénotype similaire à celui de la souche AA (solvantogénèse et consommation de sucres affectées). La complémentation avec le gène sauvage restaure la production d'alcools et la consommation de lactose, alors que la complémentation avec le gène *sigL* dérivé de la souche AA n'a pas d'effet, ce qui permet de démontrer sans ambiguïté que la mutation détectée dans la souche AA inactive σ^{54} .

Cette étude souligne le rôle crucial joué par σ^{54} pour contrôler le métabolisme du carbone de l'espèce *C. beijerinckii*. Ces résultats sont notables puisque, dans un procédé fermentaire destiné à produire des biocarburants/biochimiques, la faculté du microorganisme à assimiler les sucres et à les transformer en alcools est fondamentale. La caractérisation de la dépendance des voies vis-à-vis de σ^{54} présentée dans cet article constitue une première étape importante, mais qui apporte également de nouvelles questions. Quels sont les stimuli environnementaux qui aboutissent à une activation de l'expression des gènes contrôlés par σ^{54} ? Quels sont les acteurs de cette relation signal-expression *in vivo* ?

Le fonctionnement particulier du facteur σ^{54} pourrait apporter de premiers éléments de réponse. En effet, σ^{54} est distinct de tous les autres facteurs sigma d'un point de vue évolutif²¹³. Son expression seule ne suffit en effet pas à initier la transcription d'un promoteur cible. La présence d'une deuxième protéine, appelée *Enhancer Binding Protein* (EBP), est en effet absolument nécessaire pour fournir l'activité ATPase requise pour provoquer la transition du complexe ARN polymérase – ADN d'un état conformationnel fermé à un état ouvert, et ainsi initier la transcription²⁷¹. Dans la plupart des cas, l'expression de σ^{54} est stable et les EBP sont en réalité les vrais effecteurs de la régulation. Ils sont généralement constitués d'un domaine ATPase, d'un domaine de liaison à l'ADN et d'un domaine de transduction du signal, qui leur permet de détecter un stimulus environnemental, qui *in fine* module l'activité de l'EBP. Si de nombreux EBP sont encodés dans le génome de *C. beijerinckii*, leur mode de fonctionnement (gènes contrôlés, signaux perçus/conduits) reste encore à élucider.

Depuis la publication de cet article, un autre groupe s'est intéressé aux gènes régulés par σ^{54} dans l'ordre des *Clostridiales*²⁷². Leur analyse bioinformatique multi-espèce a permis d'identifier les motifs reconnus par les EBP en amont des gènes

importants de la solvantogénèse, entre autres. Une analyse des domaines protéiques présents dans ces EBP suggère que, chez *Clostridium beijerinckii*, la majorité des signaux reconnus sont intracellulaires et que leur activité serait modulée par une liaison avec des ligands (comme par exemple des métabolites). Sans toutefois vérifier expérimentalement ces résultats, cette nouvelle étude apporte des éléments supplémentaires pour la compréhension des mécanismes régulant la solvantogénèse chez *C. beijerinckii*.

4.2 VALORISATION SCIENTIFIQUE (ARTICLE)

SCIENTIFIC REPORTS

OPEN

σ^{54} (σ^L) plays a central role in carbon metabolism in the industrially relevant *Clostridium beijerinckii*

Rémi Hocq, Maxime Bouilloux-Lafont, Nicolas Lopes Ferreira & François Wasels

The solventogenic *C. beijerinckii* DSM 6423, a microorganism that naturally produces isopropanol and butanol, was previously modified by random mutagenesis. In this work, one of the resulting mutants was characterized. This strain, selected with allyl alcohol and designated as the AA mutant, shows a dominant production of acids, a severely diminished butanol synthesis capacity, and produces acetone instead of isopropanol. Interestingly, this solvent-deficient strain was also found to have a limited consumption of two carbohydrates and to be still able to form spores, highlighting its particular phenotype. Sequencing of the AA mutant revealed point mutations in several genes including CIBE_0767 (*sigL*), which encodes the σ^{54} sigma factor. Complementation with wild-type *sigL* fully restored solvent production and sugar assimilation and RT-qPCR analyses revealed its transcriptional control of several genes related to solventogenesis, demonstrating the central role of σ^{54} in *C. beijerinckii* DSM 6423. Comparative genomics analysis suggested that this function is conserved at the species level, and this hypothesis was further confirmed through the deletion of *sigL* in the model strain *C. beijerinckii* NCIMB 8052.

In the context of worldwide energy transition, research for alternatives to fossil fuels has become a priority. In particular, the replacement of petrochemistry by a low carbon emission industry has been a major challenge as our global consumption of petrochemicals keeps on increasing¹. The valorization of plant biomass to synthesize ethanol by microbial fermentation has already been pioneered for biofuel production² and could therefore be applied to bio-based chemistry³.

A few strains from the *Clostridium* genus are naturally able to produce isopropanol and butanol^{4,5}, two compounds that could be used as biochemical and biofuel, respectively. However, those organisms are not producing these metabolites in quantities compatible with an economically viable industrial process⁶. However, with the increasing availability of efficient genetic tools in *Clostridia*⁷, metabolic engineering approaches could be undertaken to enhance solvent productivity. *Clostridium beijerinckii* DSM 6423 (NRRL B-593) is the only natural isopropanol-butanol producing strain whose genome and transcriptome have been investigated⁸. It may therefore be the best candidate for genetic engineering, although its particular physiology is still poorly understood. For this purpose, gaining additional knowledge on metabolism regulation in this strain would greatly benefit future optimization efforts. In particular, identifying the molecular effectors controlling solvent production may provide valuable insights to define adequate genetic engineering strategies.

As no genetic toolbox was available for this particular strain, Máté de Gerando and coworkers performed random mutagenesis coupled with genome shuffling to increase isopropanol productivity by selecting isopropanol-tolerant strains⁹. In this work, random mutagenesis followed by allyl alcohol selection also generated an interesting mutant, further referred to as AA mutant. This strain mainly produces acids, shows no isopropanol production and a strongly attenuated butanol synthesis capacity. These results are consistent with those obtained in *Clostridium acetobutylicum* DSM 1792, in which mutants obtained in the presence of allyl alcohol - precursor of the highly toxic acrolein molecule in the reaction catalyzed by alcohol dehydrogenases - permitted the selection of butanol-deficient strains¹⁰. Nevertheless, in both cases the key mutated genes causing these phenotypes have not been clearly identified.

IFP Energies nouvelles, 1 et 4 avenue de Bois-Préau, 92852, Rueil-Malmaison, France. Correspondence and requests for materials should be addressed to F.W. (email: francois.wasels@ifpen.fr)

In this study, we demonstrate that the solvent production deficiency in the AA mutant is due to a point mutation in the CIBE_0767 (*sigL*) gene, which encodes the transcriptional regulator σ^{54} (also referred to as σ^L). Similarly to other sigma factors, σ^{54} regulates genetic expression by incorporating the RNA polymerase complex and binding to specific promoter sequences, thus enabling selective transcription of a subset of genes^{11–13}. The importance of this sigma factor for solventogenesis was further confirmed at the species level by deleting the corresponding gene in the *C. beijerinckii* model strain (i.e. NCIMB 8052). In addition to regulating alcohol synthesis pathways, our experiments revealed that this sigma factor also controls utilization of alternative carbon sources, such as lactose and cellobiose, and is not required to complete the sporulation process, making it a central and specific controller of carbon metabolism in *C. beijerinckii*.

Results

The *C. beijerinckii* AA mutant metabolism is orientated towards acid production. Phenotypic characterization of a solventogenic *Clostridium* is usually assessed by the identification of fermentation products. We therefore performed triplicate fermentation assays of both *C. beijerinckii* DSM 6423 wild-type and AA strains to compare their product pattern after 48 hours in Gapes medium.

When compared to the wild-type strain, the AA mutant showed a very distinctive behavior. Firstly, growth was strongly impacted with a ca. 2-fold decrease in OD₆₀₀ units after 48 hours of fermentation in Gapes medium, influencing end-products concentrations and glucose consumption (Fig. 1, Supplementary File S1).

Although butanol was still produced, the central metabolism of the AA strain shifted towards acid formation, with notably a strong increase in butyrate concentration resulting in a final solvent:acid ratio of 1:3, compared to 6:1 in the wild-type strain (Fig. 1c).

The phenotype of the AA mutant, mainly producing butyrate, may result from the coupled inactivation of one or several gene(s) involved in isopropanol and butanol production.

The AA mutant is still able to complete the sporulation process. Complete or partial loss of solventogenesis in the *C. beijerinckii* species is often associated with a degeneration phenomenon (i.e. loss of the capacity to form spores in addition to the loss of solvent production)¹⁴. Given its phenotype, we therefore decided to investigate the sporulation ability of the AA mutant.

A comparative assay confirmed that sporulation ability was conserved, despite an approximate 10/100-fold decrease in the number of spores in comparison to the wild-type strain (Supplementary File S1). However, given the limited growth of the AA strain, this decrease may not necessarily be linked to a genetic deficiency but could instead result from a diminished ability to accumulate population density.

Importantly, demonstration of the sporulation ability in the AA strain makes its phenotype unrelated to the degeneration phenomenon, which suggests that its genome contains solventogenic-specific mutations.

A point mutation in the *sigL* gene is predicted to impede solvent formation and sugar consumption. We next adopted a forward genetic approach to identify the gene(s) responsible for the AA phenotype. Whole-genome sequencing followed by read mapping on the wild-type genome allowed the identification of 13 SNPs (Table 1) along the chromosome of the strain. Among the affected genes, CIBE_0767 particularly retained our attention as P-BLAST analysis against Swissprot database revealed high similarities with putative RNA polymerase σ^{54} factors (encoded by the *sigL* gene) and notably the one from *Clostridium kluyveri*¹⁵.

In the AA strain, the SNP detected in the *sigL* gene causes a serine to phenylalanine substitution at position 366. The σ^{54} transcriptional factor has already been well described in several bacterial species and contains three major domains¹² (Fig. 2a). Region I (RI) is involved in core RNA polymerase, enhancer, and DNA binding. Region II (RII) is acidic and poorly conserved. Region III (RIII) is divided in several conserved sub-regions interacting with core RNA polymerase (CBD) and DNA at the -12 consensus motif (HTH) and the -24 motif (RpoN). *In silico* analysis of the *sigL* gene revealed that the amino acid substitution (S366F) is localized in the HTH motif, at an extremely conserved position (Fig. 2b). Given the significant difference in terms of steric hindrance, polarity and hydrophobicity resulting from a phenylalanine to serine substitution, this mutation may drastically impede the -12-motif recognition, yielding a partial or total inactivation of σ^{54} .

We subsequently looked at the predicted σ^{54} -regulon along the chromosome of *C. beijerinckii* DSM 6423. The σ^{54} binding motif consensus (TGGCANNNNNNTTGCW), based on the study of validated σ^{54} promoters by Barrios and colleagues¹³, was found at 57 genomic sites. Among them, 35 were localized up to a few hundred base pairs upstream of a coding sequence (Supplementary File S2). In particular, two genes coding for strongly expressed⁸ alcohol dehydrogenases (CIBE_2050 and CIBE_2622, which convert acetaldehyde and butyraldehyde into ethanol and butanol, respectively)^{16,17} as well as the secondary alcohol dehydrogenase s-ADH encoding gene (CIBE_3470, whose product converts acetone to isopropanol)^{16,18,19} were identified (Fig. 2c). Although 2,3-butanediol is usually not detected in our culture conditions, the only gene predicted to encode a 2,3-butanediol dehydrogenase (CIBE_1696, similar to the one described by Raedts *et al.*)²⁰ was identified as being part of the regulon.

Unlike other sigma factors, σ^{54} -bound RNA polymerase requires the ATPase activity of an adjacent Enhancer Binding Protein (EBP) to start transcription initiation^{21,22}. Importantly, genes encoding σ^{54} EBPs were found in the direct vicinity of CIBE_2622, CIBE_3470 and CIBE_1696, which strengthens the hypothesis of a σ^{54} -driven transcriptional control of their expression.

In order to investigate the functionality of the σ^{54} -regulon, we used the synteny tool from MaGe platform (16) to compare the position of σ^{54} binding sites in *C. beijerinckii* DSM 6423 with two other *C. beijerinckii* model strains (*C. beijerinckii* NCIMB 8052 and the recently reannotated NRRL B-598)^{23,24}. Apart from the s-ADH encoding gene which drives isopropanol production and is only present in the DSM 6423 strain, these putative σ^{54} -dependent transcriptional units appeared to be conserved in these acetone-butanol producing strains.

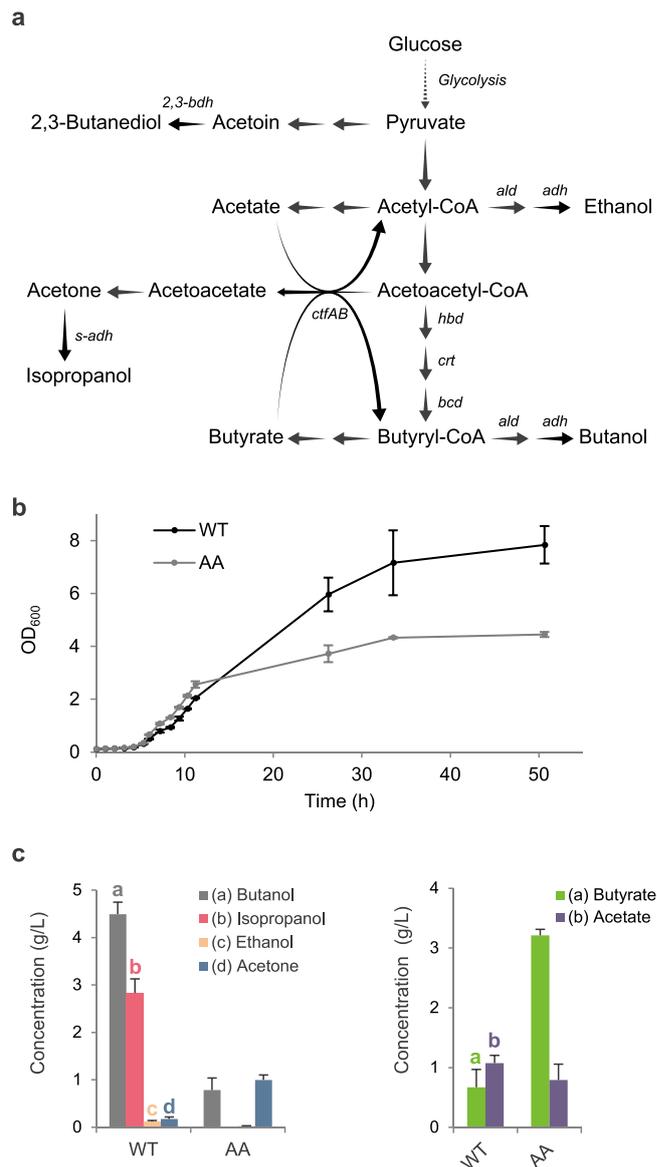


Figure 1. Comparative phenotypic analysis of *C. beijerinckii* DSM 6423 wild-type and AA strains. **(a)** Simplified central metabolism of *C. beijerinckii* DSM 6423. **(b)** Growth kinetics comparison between *C. beijerinckii* DSM 6423 (WT) and the AA mutant. **(c)** Final solvent and acid concentrations measured after 48 h of fermentation in Gapes medium for wild-type *C. beijerinckii* DSM 6423 and the AA mutant. Error bars indicate the standard deviation of triplicate experiments.

Besides, downregulation of the butanol and secondary alcohol dehydrogenases linked to σ^{54} potential inactivation would match the AA phenotype, which gives another hint that *sigL* mutation may be significant.

Interestingly, other genes predicted to be part of the σ^{54} -regulon are mainly associated with sugar uptake (mainly PTS-based transporters, for a review in solventogenic *Clostridia*, see Mitchell, 2015)²⁵ and metabolism (Fig. 2d, Supplementary File S2). Lactose and cellobiose transport and metabolism are also predicted to be regulated by σ^{54} . These operons contain genes encoding a phosphotransferase system (PTS) for sugar transport and a hydrolase (β -galactosidase CIBE_0885; β -glucosidase CIBE_1708). Interestingly, a highly expressed and conserved operon encoding a dihydroxyacetone phosphate kinase complex (*dhaKLM*; CIBE_2581-2582-2583) also appears to be controlled by σ^{54} . This enzyme can be involved in glycerol metabolism, or directly in the central carbon pathway when associated to 6P-fructose aldolases^{26,27}. Notably, two 6P-fructose aldolases (CIBE_0334 and CIBE_0411), one of which being well expressed in the available RNA-seq dataset⁸, are annotated in the DSM 6423 genome. Lastly, parts of the pentose phosphate pathway also seem to be regulated by σ^{54} , with two conserved similar operons encoding transketolases, ribose/ribulose phosphate epimerases and isomerases (Supplementary File S2). Importantly, some of these putative σ^{54} -dependent transcriptional units appeared to be well conserved in the NCIMB 8052 and NRRL-B598 strains (Fig. 2d).

Locus tag	Description	Change	Genome position	CDS Position	Amino Acid Change	Protein Effect
CIBE_0366	Conserved protein of unknown function	C -> T	390,092	2,215	Q -> X	Truncation
CIBE_0767	Putative RNA polymerase sigma-54 factor	C -> T	770,679	1,097	S -> F	Substitution
CIBE_1279	Conserved exported protein of unknown function	C -> T	1,347,126	31	none	None
CIBE_1303	Conserved protein of unknown function	A -> G	1,373,758	218	Y -> C	Substitution
N/A ^a	Intergenic	G -> A	1,691,364	N/A ^a	N/A ^a	N/A ^a
CIBE_2357	Binding-protein-dependent transport systems inner membrane component	G -> A	2,397,512	614	G -> D	Substitution
CIBE_2358	NMT1/THI5-like protein	G -> A	2,397,959	138	None	None
CIBE_2486	Protein of unknown function	C -> T	2,535,980	103	P -> S	Substitution
CIBE_2488	Protein of unknown function	C -> T	2,537,804	157	P -> S	Substitution
CIBE_3408	Putative hydrolase subunit antagonist of KipI	C -> T	3,476,677	206	T -> I	Substitution
CIBE_3475	Histidine kinase	C -> A	3,572,491	732	R -> S	Substitution
CIBE_4586	Dihydroxy-acid dehydratase	G -> T	4,675,905	1,159	V -> F	Substitution
N/A ^a	Intergenic	C -> T	4,931,169	N/A ^a	N/A ^a	N/A ^a
CIBE_5453	Short chain acyl-CoA transferase: fused alpha subunit; beta subunit	A -> G	5,622,646	919	S -> P	Substitution
CIBE_5584	Putative ABC transporter (ATP-binding protein)	G -> A	5,774,626	146	R -> K	Substitution

Table 1. Mutations detected in the AA strain. ^aNot applicable.

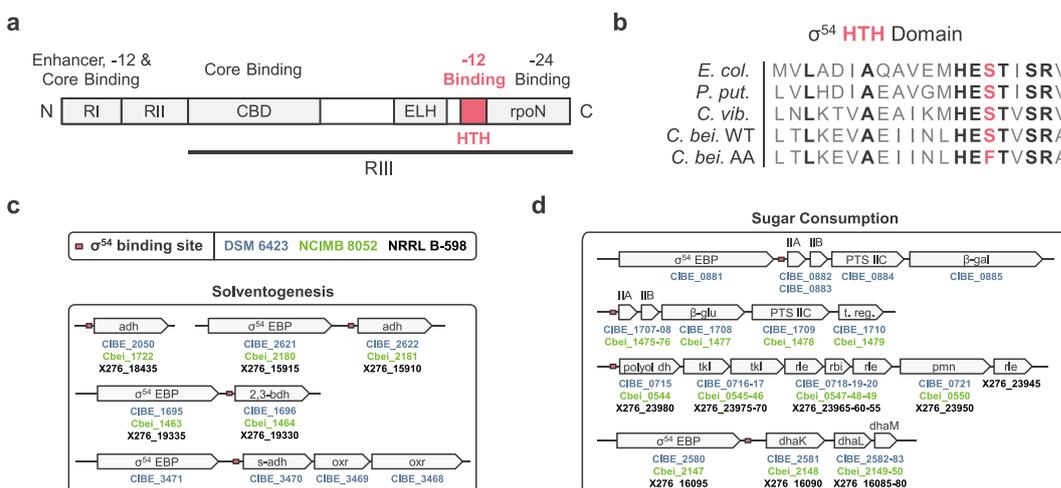


Figure 2. Structural position of σ^{54} mutation in *C. beijerinckii* DSM 6423 AA strain and predictions for this sigma factor role in several model strains. **(a)** Structural organization of the σ^{54} sigma factor. RI-III: region I to III. CBD: core binding domain. ELH: extra-long α -helix. HTH: helix turn helix motif. rpoN: RNA polymerase factor N domain. **(b)** Localization of the S336F mutation identified in the AA strain. *E. col.*: *Escherichia coli*; *P. put.*: *Pseudomonas putida*; *C. vib.*: *Caulobacter vibrioides*; *C. bei.*: *Clostridium beijerinckii*. **(c,d)** In silico predictions of **(c)** solvent production and **(d)** sugar metabolism regulation by σ^{54} in several *C. beijerinckii* strains. Transcriptional units involved in these pathway bear an upstream σ^{54} binding site (consensus: TGGCANNNNNTTGCW). *adh*: alcohol dehydrogenase; *2,3-bdh*: 2,3-butanediol dehydrogenase; *s-adh*: secondary alcohol dehydrogenase; *oxr*: oxidoreductase. PTS: phosphotransferase system; β -gal: β -galactosidase; β -glu: β -glucosidase; *t. reg.*: transcriptional regulator; *dhaK/L/M*: dihydroxyacetone kinase subunits K/L/M; *polyol dh*: polyol dehydrogenase; *tkl*: transketolase; *rle*: ribulose phosphate epimerase; *rbi*: ribose phosphate isomerase; *pmm*: phosphomannomutase.

In summary, analysis of the AA genotype revealed an amino acid substitution potentially inactivating a protein predicted not only to regulate solventogenesis but also other central metabolic pathways such as sugar consumption in several *C. beijerinckii* strains.

Complementation with wild-type *sigL* gene fully restores solventogenesis and sugar metabolism in the AA strain. As several SNPs were detected in the genome of the AA mutant, complementation assays were required to characterize the corresponding phenotype. Considering the predicted crucial role of the putative σ^{54} -encoding gene (*sigL*, CIBE_0767), we decided to clone the wild-type coding sequence of *sigL* under the control of its endogenous promoter in the pFW01 vector, yielding the pFW- σ^{54} plasmid, and to introduce

it into both wild-type and AA strains. Fermentation analysis revealed that both WT and AA pFW- σ^{54} strains displayed a wild-type solventogenic metabolism (Fig. 3). Indeed, normal acid and solvent production (Fig. 3a), as well as glucose consumption and biomass production (Supplementary File S3) were fully restored in the AA pFW- σ^{54} strain when compared to the wild-type strain containing the empty vector. Interestingly, the WT pFW- σ^{54} recombinant strain did not exhibit any major change when compared to the wild-type microorganism.

These results confirmed that CIBE_0767, encoding the σ^{54} transcriptional factor, is drastically influencing the regulation of the central metabolism of the AA strain in our conditions. We therefore kept on investigating its impact on other predicted metabolic pathways.

Plate assays on minimal medium were used to analyze carbon sources usage. In our conditions, no effect after 24 or 48 hours was seen on bacterial growth on fructose, mannose, xylose, arabinose, sucrose, glycerol, inulin, starch or mannitol (data not shown). In contrast, growth was impeded with cellobiose and lactose in the AA control strain (empty vector), and restored when complemented with wild-type σ^{54} (Fig. 3b). However, quantification of this inhibition in liquid medium revealed unexpected results as growth was not altered on 2YT 20 g/L cellobiose medium (data not shown) whereas lactose did exert a strong effect on bacterial growth. After 24 h of fermentation the AA pFW01 strain had grown similarly on 2YT 20 g/L lactose than on the 2YT control condition containing no supplementary carbon source (Fig. 3c). HPLC analysis confirmed that almost no lactose was consumed by the AA strain in this time frame. On the other hand, normal growth on lactose was observed with the complementation strain with a consumption of ca. 8 g/L of lactose in 24 h. The fermentation was carried out for 72 more hours, and end-point analysis revealed that lactose uptake inhibition was only partial (Supplementary File S4).

Transcriptional control of the predicted σ^{54} regulon was further assessed by performing RT-qPCR on the WT/AA strains, with and without *sigL* complementation (Fig. 3d). Expression of CIBE_2050 (*adh-1*), CIBE_2622 (*adh-2*) and CIBE_3470 (*s-adh*) was decreased (c.a. 2.5–5 fold) in the AA strain and complementation resulted in a drastic increase of their expression (≈ 10 times), congruent with the strong expression of *sigL* on a multi-copy plasmid. This stronger expression however did not result in a higher titer of solvents as outlined before. Alcohol synthesis may thus be limited by another factor (e.g. butanol tolerance or another rate-limiting step in the solventogenic pathways) in *C. beijerinckii* DSM 6423.

The *sigL* expression level was interestingly much lower in the WT strain than in the AA mutant, which could be linked to a potential feedback regulation.

As expected, expression of the predicted σ^{54} -independent *ald* gene (CIBE_4606, first gene of the *sol* operon)^{8,28} did not vary among these strains. Similarly, expression of two σ^{54} EBPs (CIBE_2621, CIBE_3471) was not affected, indicating that their expression is not driven by σ^{54} . This result was also anticipated, since these genes are not expected to be regulated by σ^{54} .

Altogether, these results indicate that the S366F substitution in the σ^{54} is likely the unique modification in the AA strain that impacts solvent production and sugar assimilation. Moreover, our results suggest that σ^{54} is a master regulator of these metabolic pathways in *C. beijerinckii* DSM 6423.

Overexpression of the acid-reassimilation pathway does not restore a wild-type phenotype in the AA strain. The AA phenotype could be linked to two possible mechanisms: a deficient alcohol dehydrogenase activity in the cell and/or a defect in the acid uptake pathway.

To verify whether acid uptake might be impaired in the AA strain, we designed an experiment in which the acid reassimilation pathway would be overexpressed in the AA mutant. In order to do this, a pFW-FC06 plasmid was constructed to overexpress the *s-adh*, *ctfA*, and *ctfB* genes in an operonic structure. This operon was previously used in an overexpression plasmid^{29–31} or integrated in the chromosome³² to convert the model acetone-butanol producing organism *C. acetobutylicum* ATCC 824 into a performant isopropanol-butanol producer.

Introduction of the pFW-FC06 vector slightly increased solvent production and glucose consumption in the AA strain, compared to the introduction of an empty vector (Supplementary File S5). However, even if acetone was fully converted to isopropanol, the recombinant AA strain containing pFW-FC06 still accumulated butyrate at levels comparable to the AA mutant. The latter observation suggests that, consistently with *in silico* predictions, a CtfA/B complex is still functional in the AA mutant (i.e. overexpression of *ctfA/B* does not have much effect on acid levels, similarly to the wild-type strain). The main metabolism roadblock thus probably comes from the aforementioned deficient butanol synthesis pathway. Indeed, in the AA strain, this result would be expected if cyclic conversion of butyryl-CoA to butyrate, and back to butyryl-CoA via the CtfA/B enzyme, was occurring.

In summary, this experiment gives an additional insight on the molecular mechanisms triggering solventogenesis in *C. beijerinckii*: by regulating alcohol dehydrogenase expression, σ^{54} -mediated transcription may be able to modulate the balance between acidogenesis and solventogenesis.

CRISPR/Cas9-mediated deletion of *sigL* in *C. beijerinckii* NCIMB 8052 results in an AA-like phenotype. Complementation with wild-type σ^{54} in the AA strain shed light on *sigL* role in solventogenesis and sugar uptake. However, given the many mutations found along the genome of the AA mutant, we aimed at confirming that the observed phenotype was the unique effect of *sigL* inactivation by deleting it in a wild-type strain. The model strain *C. beijerinckii* NCIMB 8052 was chosen as a chassis for this modification since, unlike the DSM 6423 strain, markerless genome editing techniques have already been demonstrated to be effective in this microorganism^{33–35}. More particularly, genetic modifications using CRISPR/Cas9-based tools were recently made available to the *Clostridium* community: this technology comes with several advantages (i.e. genome editing overall efficiency, simplicity and rapidity), particularly interesting for genetic engineering of the notably difficult-to-modify clostridial strains, as reviewed recently by Xue *et al.*³⁶

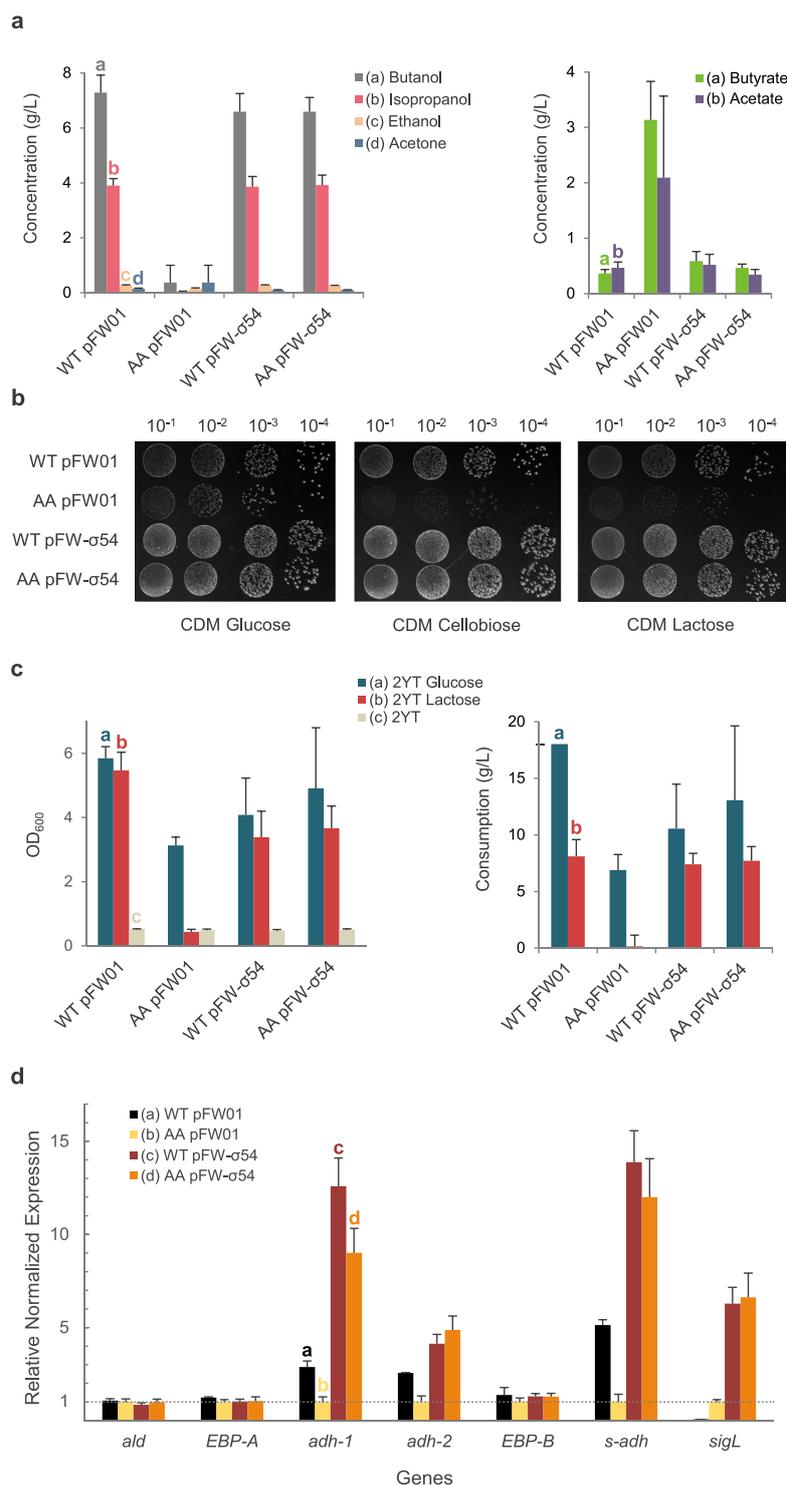


Figure 3. Phenotypic analysis of *C. beijerinckii* DSM 6423 σ^{54} complementation in the AA strain.

(a) Fermentation products after 72 hours in Gapes 60 g/L glucose medium of *C. beijerinckii* DSM 6423 wild-type (WT), AA strains containing an empty plasmid (pFW01) or the complementation plasmid (pFW01- σ^{54}). (b) Sugar utilization assay on CDM plates containing 20 g/L glucose, cellobiose or lactose. Pictures were taken after 24 h of incubation on glucose and cellobiose, and 48 h on lactose. (c) Sugar utilization assay in liquid 2YT with 20 g/L glucose (2YTG), lactose (2YTTL) or without added carbon source (2YT). Biomass (given as a measure of optical density at 600 nm, OD₆₀₀) and carbon consumption are measured 24 h after the beginning of the fermentation. (d) RT-qPCR assay measuring the stationary-phase expression of several genes (*ald*: CIBE_4606; *EBP-A*: CIBE_2621; *adh-1*: CIBE_2050; *adh-2*: CIBE_2622; *EBP-B*: CIBE_3471; *s-adh*: CIBE_3470; *sigL*: CIBE_0767). Values are given relative to the AA pFW01 expression value. Error bars indicate the standard deviation of triplicate experiments, except for qPCR data that were obtained in duplicates.

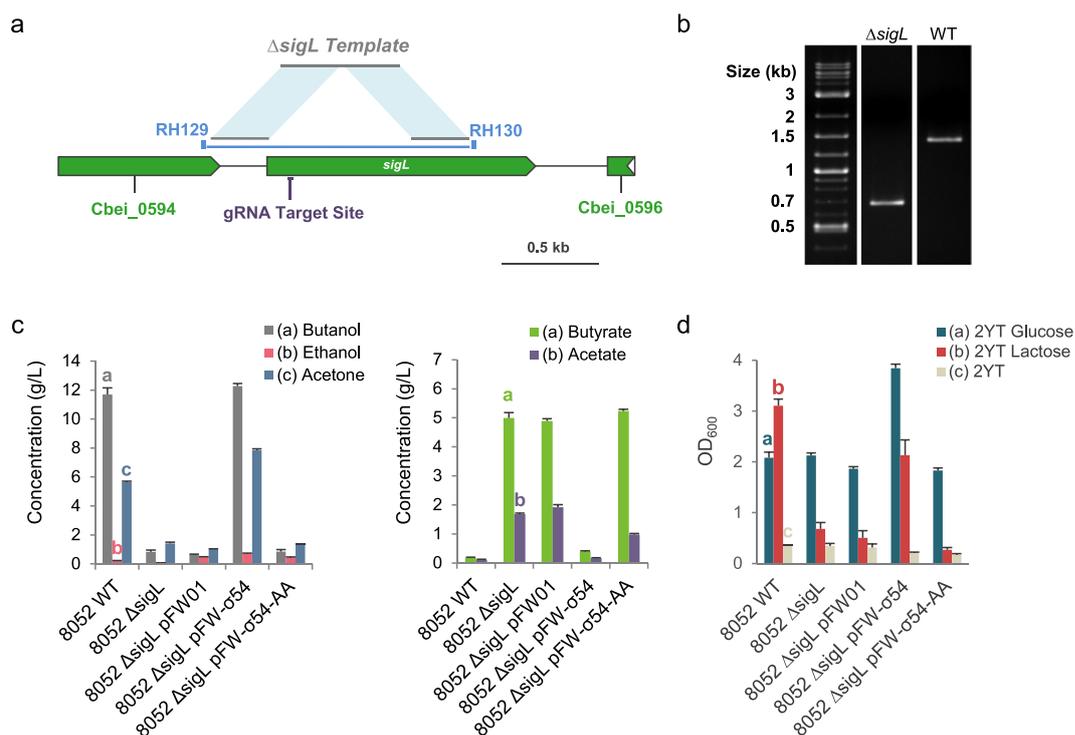


Figure 4. Design of CRISPR/Cas9-mediated σ^{54} deletion in *C. beijerinckii* NCIMB 8052 and phenotypic comparison of wild-type and mutant σ^{54} complementation. **(a)** *sigL* (Cbei_0595) genomic region with CRISPR/Cas9 genome editing design. **(b)** Verification by colony PCR of *sigL* deletion with primers RH129 and RH130 encompassing the deleted region. Expected band sizes are 1406 (wild-type) and 670 ($\Delta sigL$) bp. **(c)** Comparative analysis of fermentation capacities of *C. beijerinckii* NCIMB 8052 wild-type, $\Delta sigL$ and complemented strains. **(d)** Growth comparison of *C. beijerinckii* NCIMB 8052 wild-type, $\Delta sigL$ and complemented strains after 24 hours of fermentation in 2YT, 2YTG and 2YTL liquid media. gRNA: guide RNA. Error bars indicate the standard deviation of triplicate experiments.

A genome editing strategy based on the one described by Wasels *et al.*³² was undertaken. Briefly, an anhydrotetracycline-inducible system allows the selection of double allelic exchange events between the clostridial DNA and a user-defined editing template, by using the Cas9/gRNA complex to target the unmodified bacterial chromosome. Double strand breaks thus only occur in the wild-type or single allelic exchange cases, resulting in cell death.

A plasmid based on the pCas9_{ind} vector was constructed to comprise, in addition to the inducible *cas9* gene, an anhydrotetracycline-inducible gRNA expression cassette and a gene editing template designed to delete most of the *sigL* gene (i.e. gene truncation results in a 13 AA protein, Fig. 4a). Introduction of this plasmid in *C. beijerinckii* NCIMB 8052 and induction of the CRISPR/Cas9 machinery permitted the selection of $\Delta sigL$ clones (Fig. 4b, Supplementary File S7).

One of those $\Delta sigL$ clones was subsequently complemented with the pFW- $\sigma 54$ and pFW- $\sigma 54$ -AA plasmids, the latter allowing the expression of the S336F-mutated σ^{54} .

Fermentation assays revealed that solvent production was drastically altered in the $\Delta sigL$ strain, and that a wild-type phenotype could only be restored with pFW- $\sigma 54$ (Fig. 4c, Supplementary File S7). Complementation with pFW- $\sigma 54$ -AA notably appeared to have no effect on acid reassimilation and solvent production when compared to the $\Delta sigL$ empty vector control strain. Besides, all of those strains displayed a resistance to heat shock (cells scraped from one-week-old plates, 80 °C for 10 min, data not shown), highlighting their retained capacity to correctly form spores. Lastly, $\Delta sigL$ mutant growth appeared to be highly inhibited in lactose-based medium, similarly to the AA strain (Fig. 4d). Cellobiose uptake however did not seem to be impacted in this strain (data not shown).

In addition to confirming our hypothesis on the role of σ^{54} for solventogenesis and sugar utilization, deleting *sigL* in the NCIMB 8052 model strain allowed us to expand our conclusions regarding this transcriptional factor to other microorganisms belonging to the *C. beijerinckii* species.

Discussion

In this study, the allyl alcohol mutant generated from *C. beijerinckii* DSM 6423 by Máté de Gerando *et al.*⁹ was genetically characterized by a forward genetic approach. Complementation assays demonstrated that the AA phenotype is due to a unique point mutation in the CIBE_0767 *sigL* gene encoding the σ^{54} transcriptional factor. This mutation leads to the substitution of a highly conserved serine in the HTH domain by a phenylalanine residue (S336F). Given the central role of the HTH domain in DNA recognition^{37,38}, this mutation is likely to

strongly impede the σ^{54} -RNA polymerase complex binding to promoter sequences and thus to inhibit transcription of the σ^{54} -regulon. If the effect of this particular mutation on σ^{54} functionality has not been described so far, Coppard and Merrick³⁸ performed targeted mutagenesis of the corresponding serine of the HTH domain of *K. pneumoniae*. Its substitution by most residues (especially large ones like lysine or tryptophane) completely shut down σ^{54} -mediated transcription, suggesting this serine is crucial for a functional activity, which supports our hypothesis that σ^{54} is inactivated in the AA strain.

The impact of the S336F mutation in the σ^{54} sequence was experimentally investigated by introducing expression plasmids of either the wild-type or the S336F version of σ^{54} in a $\Delta sigL$ mutant of the model strain *C. beijerinckii* NCIMB 8052. This mutant displayed a phenotype similar to the AA strain, solventogenesis being severely impaired, which resulted in a butyrate-orientated metabolism. Complementation with pFW- σ^{54} fully restored solventogenesis, while the introduction of the pFW- σ^{54} -AA vector in the $\Delta sigL$ strain had no effect. These experiments overall demonstrated that the S336F mutation completely inactivates σ^{54} and are thus consistent with our previous results.

The deleterious effects of allyl alcohol on alcohol-producing microorganisms (yeast^{39–41}, *E. coli*⁴²) and its tendency to generate spontaneous mutations have been reported decades ago. Reverse activity of alcohol dehydrogenases catalyzes the formation of the highly toxic acrolein molecule from allyl alcohol, enabling the selection of alcohol dehydrogenase deficient mutants. In *Clostridia*, similar work was pioneered by Dürre and colleagues¹⁰, who obtained *C. acetobutylicum* mutants unable to produce butanol. Interestingly, butyraldehyde dehydrogenase activity was drastically reduced in these strains. Given the *in silico* σ^{54} -targeted promoter predictions and the expression pattern of the *ald* gene in the WT and AA strains, σ^{54} apparently does not control the butyraldehyde dehydrogenase genes transcription in the DSM 6423 strain. The supposed inactivation of σ^{54} in the AA mutant rather seems to lead to a decreased expression of two major alcohol dehydrogenases genes, resulting in a phenotype close to what was observed in the *C. acetobutylicum* corresponding mutants.

Acetate and butyrate are normally produced during the acidogenic phase and are further consumed during solventogenesis⁴³. In the AA mutant, increased acid production results in a substantial drop of the pH which may prematurely cause cell growth to abort. qPCR data revealed that some normally highly expressed alcohol dehydrogenase genes were underexpressed in the AA strain, which suggests a link between the increase of acid production and the partial loss in butanol production capacity. The significant upregulation of those genes transcription when *sigL* was overexpressed in both the WT and AA strains further consolidates this hypothesis.

Isopropanol production was also abolished in the AA strain but its precursor - acetone - was still detected. As expected, expression of the sole secondary alcohol dehydrogenase (s-ADH, encoded by the CIBE_3470 gene), which catalyzes the acetone-to-isopropanol reaction, was highly affected in this strain. Being linked to acid consumption, acetone formation indicates that butyrate and acetate uptake still occurs. In the DSM 6423 wild-type strain, the resulting acetyl-CoA and butyryl-CoA are further reduced into acetaldehyde/butyraldehyde and then ethanol/butanol. Because alcohol synthesis is drastically reduced in the AA strain, this implies that most of the acetyl-CoA and butyryl-CoA are converted back to acetate and butyrate first by the CtfA/B CoA-transferase^{44,45} and then by phosphotransbutyrylase/acetylase - butyrate/acetate kinase pathway^{46–51}. This assumption would be supported by the increased (acetone + isopropanol):butanol ratio in the AA strain (close to 1:1, versus 1:3 in the wild-type). Indeed, this ratio is expected to be higher if butyryl-CoA is preferentially used to synthesize butyrate instead of butanol.

Among acidic products, butyrate synthesis seems to be preferred over acetate production in the AA strain. Acetate was indeed consumed in most - but not all - fermentation experiments (i.e. final concentration in the fermentation broth less than the 2.4 g/L contained in the Gapes medium) involving the AA mutant or the NCIMB 8052 $\Delta sigL$ strain, which also supports the hypothesis that acid uptake does occur. This also suggests that the butyryl-CoA synthesis operon (i.e. genes linking acetyl-CoA and butyryl-CoA, namely: *hbd*, *crt*, *bcd*, *etfA*, *etfB*^{52–54}), is still constitutively expressed in the AA mutant, similarly to the WT strain⁸. Besides, overexpression of the acid reassimilation pathway (CtfA/B, s-ADH) in the AA strain did not increase acid consumption. In accordance with *in silico* predictions, our data thus strongly suggest that σ^{54} main mode of action consists in driving the regulation of solventogenesis by modulating alcohol dehydrogenases expression and not by interfering on the bacterial acid uptake capacity.

σ^{54} was also shown to control sugar uptake and usage, consistently with what Nie *et al.* predicted for the *Clostridium* genus⁵⁵. In particular, growth was limited on cellobiose and lactose in the AA strain, though cellobiose inhibition could only be observed on minimal medium. This may be explained by the existence of another cellobiose assimilation pathway in the DSM 6423 genome. In the NCIMB 8052 strain, *sigL* deletion was demonstrated to impact lactose uptake, confirming that σ^{54} is involved in sugar uptake and/or metabolism in *C. beijerinckii*.

σ^{54} belongs to a unique category of sigma factors, being evolutionary distinct to the σ^{70} family, which comprises all of the other sigma factors^{56,57}. This difference implies a very different mode of action for the σ^{54} -RNA polymerase complex: unlike its σ^{70} counterpart, the holoenzyme is unable to initiate transcription without the help of an enhancer binding protein (EBP)²². EBPs are often acting several hundred of base pairs upstream of the σ^{54} -holoenzyme deposition sites and contain a central AAA+ domain. The latter provides, through ATP hydrolysis, the energy necessary for structural rearrangements within the σ^{54} -bound holoenzyme, which in turn allows transcription initiation. EBP control on σ^{54} -dependant transcription is mainly exerted by a signal-sensing domain, which upon specific environmental stimuli activates or represses its ATPase activity²². This enhancer-dependent system offers a very tight control on gene expression, and σ^{54} regulons are often associated with various biological processes that require a stringent control (e.g. virulence^{58–60} or biofilm formation^{61–63}). Multiple σ^{54} EBPs were found in *C. beijerinckii* genomes and are thus likely to be the primary effectors of σ^{54} -mediated regulation. Indeed, RNA-seq analysis in the *C. beijerinckii* DSM 6423 wild-type strain⁸ shows that *sigL* is not differentially expressed over the time-course of a batch fermentation, whereas its target solventogenic genes (CIBE_2050, CIBE_2622,

Strains and plasmids	Relevant characteristics	Source or reference
<i>Clostridium beijerinckii</i>		
DSM 6423	Isopropanol-butanol producer	DSMZ
AA	Allyl alcohol-resistant strain obtained by random mutagenesis	Máté de Gerando <i>et al.</i> (2016)
NCIMB 8052	Acetone-butanol producer	NCIMB
NCIMB 8052 $\Delta sigL$	Markerless deletion of <i>sigL</i>	This study
<i>Escherichia coli</i>		
NEB 10- β	Cloning strain	NEB
INV110	<i>dam</i> ⁻ <i>dcm</i> ⁻ strain	Invitrogen
Plasmids		
pFW01	<i>Em</i> ^R ; ColE1 origin; pCB102 origin	Wasels <i>et al.</i> (2017)
pEC750C	<i>Cm</i> ^R , ColE1 origin, pIP404 origin	Wasels <i>et al.</i> (2017)
pCas9 _{con}	pFW01 derivative with <i>cas9</i> (constitutive minipTh1 promoter)	Wasels <i>et al.</i> (2017)
pCas9 _{ind}	pFW01 derivative with <i>cas9</i> (inducible TetO2-1 promoter)	Wasels <i>et al.</i> (2017)
pFC06	pEC750C derivative with <i>sadh ctfA ctfB</i> (constitutive <i>thl</i> promoter) insertions	Collas <i>et al.</i> (2014)
pFW- σ^{54}	pFW01 derivative with <i>sigL</i> insertion (endogenous promoter)	This study
pFW- σ^{54} -AA	pFW- σ^{54} bearing the S336F mutation from the AA strain	This study
pCas9- $\Delta sigL$	pCas9 _{ind} derivative with $\Delta sigL$ homology regions and gRNA (inducible TetO2-1 promoter)	This study
pFW-FC06	pFW01 derivative with <i>sadh ctfA ctfB</i> (constitutive <i>thl</i> promoter) insertions	This study

Table 2. Strains and plasmids used in this study. Em, erythromycin; Cm, chloramphenicol.

CIBE_3470) are among the top upregulated genes during the transition from acidogenesis to solventogenesis, which suggests that activation of their transcription is mediated by other factors (i.e. the enhancer binding proteins). RT-qPCR data also suggest that, at least in some cases, expression of *sigL* is regulated, which might provide an additional mean to regulate alcohol synthesis and/or sugar uptake. A well-known example of EBP-related regulation is the σ^{54} -driven transcription of the levanase operon in *B. subtilis*, which encodes an enzymatic and transport machinery allowing the degradation and consumption of fructose polymers^{64,65}. The enhancer binding protein LevR was shown to mediate the activation and repression of the expression of this operon by a complex interplay with the PTS system⁶⁶. In *C. beijerinckii*, the exact stimuli recognized by EBPs and underlying regulatory mechanisms that allow transcription of σ^{54} target genes nonetheless still remain to be clarified.

In bacteria, the housekeeping sigma factor (e.g. σ^{70} in *E. coli*) integrates RNA polymerase and drives transcription of most housekeeping genes. However, bacterial genomes encode multiple alternative sigma factors which direct RNA polymerase to different promoters, providing a simple but broadly exploited strategy to regulate genetic expression^{67,68}. In solventogenic *Clostridia*, alternative sigma factors have mainly been highlighted for their crucial role in the regulation of the multi-stages sporulation cascade^{69–71}. In *C. beijerinckii*, no sporulation-specific targets could be found in the predicted σ^{54} regulons, and σ^{54} inactivation did not impede the sporulation cascade, which underlines its role as a carbon-specific regulator in this bacteria.

In summary, we described σ^{54} as a master regulator of solventogenic pathways at the species level in *C. beijerinckii*. This transcriptional factor was also found to control sugar consumption and is therefore an essential controller of carbon metabolism in *C. beijerinckii*.

Material and Methods

Strains, media and culture conditions. Strains and plasmids used in this study are presented in Table 2.

Clostridium beijerinckii was grown in liquid 2YTG (per liter: 16 g tryptone, 10 g yeast extract, 5 g NaCl, 20 g glucose, pH 5.2). Solid media was prepared with 15 g/L agar and less glucose (5 g/L). Strains were cultivated in an anaerobic chamber (Bactron) at 34 °C without shaking.

Escherichia coli was grown in liquid or solid LB media, in aerobic conditions at 37 °C with 180 rpm agitation when necessary.

Whole-genome sequencing. Genomic DNA of the AA strain was purified with the GenElute Bacterial Genomic DNA Kit (Sigma-Aldrich) and was subsequently sequenced on an Illumina MiSeq sequencer (2 × 250 paired-end reads). Mutations were detected by read mapping against the *C. beijerinckii* DSM 6423 genome⁸ using Geneious R10⁷².

Predictions of σ^{54} -regulons. Genomes of the *C. beijerinckii* DSM 6423, NRRL B-598 and NCIMB 8052 strains were scanned for the σ^{54} consensus motif (TGGCANNNNNTTGCW¹³, no mismatches allowed) using Geneious R10⁷². Genes were considered potentially regulated by σ^{54} if motifs were found up to 500 bp upstream their coding sequences. Predicted candidate genes/operons were then compared in the three strains using the synteny tool from MaGe platform⁷³.

Fermentation assays and analytical methods. For fermentation assays, modified Gapes medium⁷⁴ was used. This medium contains per liter: 2.5 g yeast extract, 1 g KH₂PO₄, 0.6 g K₂HPO₄, 1 g MgSO₄, 7 H₂O, 6.6 mg FeSO₄, 7 H₂O, 0.1 g para-aminobenzoic acid, 2.9 g ammonium acetate, 60 or 80 g glucose. The medium was supplemented with erythromycin (20 µg/mL) when appropriate.

Fermentations were performed in triplicate. For each biological replicate, several clones isolated in 2YTG agar plates were used to inoculate 5 mL Gapes medium. These precultures were carried out at 34 °C overnight in the anaerobic chamber. Serum flasks containing 20 mL Gapes medium were then inoculated with 2 mL of preculture and sealed with rubber stoppers. A pressure relief valve system was punctured through the rubber stoppers to prevent overpressure, and the serum bottles were incubated 48–72 hours at 34 °C with agitation outside of the anaerobic chamber.

After OD₆₀₀ (UV-1800 spectrophotometer, Shimadzu) was measured, samples from the fermentation were centrifuged 5 min at 8000 g. The supernatant was diluted with an internal standard (1-propanol: final concentration 0.5 g/L). Metabolites concentrations in the supernatant were quantified by chromatography.

Gas chromatography (Porabond-Q column from Agilent technologies, 25 m length, 0.32 internal diameter, 5 µm film thickness coupled to a flame ionization detector) was performed to determine solvent concentrations. Helium was used as a carrier gas at a flowrate of 1.6 mL/min. Column was gradually heated from 50 °C to 250 °C in a 30 min run.

Acid concentrations were quantified by HPLC (Aminex HPX-87H from Biorad coupled to a Spectra System RI 150 refractometer and a Waters 2487 Dual λ UV detector set at 210 nm). 0.01 M sulfuric acid mobile phase was used at a flowrate of 0.6 mL/min. Column temperature was set at 60 °C.

Residual sugar quantities were determined by HPLC using an Aminex HXP-87P (Bio-Rad) coupled to a Varian 350 RI refractometer for detection. Water was used as a mobile phase at a flowrate of 0.4 mL/min. When quantifying sugar concentration for 2YT-based samples, a Micro-Guard De-Ashing Refill cartridge (Bio-Rad) was added to the chromatographic system for prior desalting of the samples. Column temperature was set at 80 °C.

Plasmid construction and transformation. All primers used in this study are listed in Supplementary File S6.

The *sigL* gene (CIBE_0767) from *C. beijerinckii* DSM 6423 together with its endogenous promoter were amplified by PCR with primers RH077 and RH078. The pFW01 backbone as well as the thiolase terminator were amplified by PCR from plasmid pCas9_{con}³² with primers RH086 and RH087. Those two amplicons were subsequently assembled by Gibson assembly (NEB)⁷⁵ to yield pFW-σ54. From this plasmid, pFW-σ54-AA was obtained by site-directed mutagenesis with primers RH136 and RH137.

The *sadh* gene (CIBE_3470) from *C. beijerinckii* DSM 6423 and the *ctfA* and *ctfB* genes from *C. acetobutylicum* ATCC 824 were amplified by PCR with primers MKz01 and MKz02 from plasmid pFC006²⁹. The resulting PCR product was cloned between the KpnI and SalI restriction sites to obtain the pFW-FC06 vector.

For the deletion of *sigL* in *C. beijerinckii* NCIMB 8052, two homology regions and an anhydrotetracycline-inducible gRNA expressing cassette were first synthesized by BaseClear in a pUC57 plasmid. Homology regions were designed to delete a 736 bp fragment in the *sigL* gene, yielding a truncated ORF (13 amino acid residues instead of 463). This cassette was amplified with primers RH125 and RH126 and subsequently cloned into the pCas9_{ind} vector³² at the XhoI restriction site. The gRNA protospacer (designed using Genious R10)⁷² was then introduced in the resulting vector by Golden Gate assembly at the BsaI restriction site with primers guide_{sigL}_fwd and guide_{sigL}_rev, yielding pCas9-Δ*sigL*.

The plasmids were then isolated and electroporated into *C. beijerinckii* as previously described^{76,77}. *dam*⁺ *dcm*⁺ DNA and *dam*⁻ *dcm*⁻ DNA were used for *C. beijerinckii* NCIMB 8052 and DSM 6423, respectively.

qPCR assays. Starting from fresh precultures, 20 mL bacterial cultures were carried out anaerobically in duplicates for 16 hours at 34 °C and 120 rpm, starting with an initial OD₆₀₀ value of ≈0.1. Ten mL samples were stabilized on ice by addition of a cold stop solution (ethanol:acid phenol at 1:18 ratio), centrifuged, and the cell pellets were frozen at -80 °C.

TRI reagent (Ambion) was further used to lyse the cells. RNAs were then extracted with the Direct-zol RNA purification kit (Zymo Research). After DNase treatment (Turbo DNase, Invitrogen), RNAs were column-purified with the RNA Clean & Concentrator 25 kit (Zymo Research).

For reverse transcription, 100 ng or RNAs were used in 20 µL reactions (iScript cDNA Synthesis Kit, Bio-Rad). Of the resulting cDNAs, 0.5 µL were then used for qPCR (SsoAdvanced Universal SYBR Green Supermix, Bio-Rad) along with adequate controls (no RT controls, positive controls on gDNA, no template controls). Primers are given with their corresponding qPCR efficiency in Supplementary File S6.

qPCR data were analyzed with the ΔΔCq method, using *gapC* (CIBE_0769) as a housekeeping gene control and taking into account reaction efficiencies.

CRISPR/Cas9 genome edition. We designed a CRISPR/Cas9 approach to inactivate *sigL* in the NCIMB 8052 strain. An inducible system based on anhydrotetracycline (aTc) addition was used, based on the pCas9_{ind} vector previously used in *C. acetobutylicum* ATCC 824³². Homology regions for genome editing were designed to delete most of the *sigL* gene, resulting in a severely truncated protein (13 amino acid residues, versus 463 for the full gene; Fig. 4a). Following insertion of those homology sequences and of an anhydrotetracycline inducible gRNA cassette, the resulting pCas9-Δ*sigL* plasmid was introduced into *C. beijerinckii* NCIMB 8052 by electroporation. Δ*sigL* mutants were obtained similarly to the method described by Wasels *et al.*³². Briefly, transformants resuspended in liquid 2YTG (pH 5.2, 20 g/L glucose) and serially diluted for spotting onto 2YTG plates

containing erythromycin (20 µg/mL) and anhydrotetracycline (50 ng/mL). Isolated colonies were tested by colony PCR with primers RH129 and RH130, encompassing the homology regions in the genome. Positive mutants were subsequently cured by streaking them twice on 2YTG plates supplemented with 50 ng/mL anhydrotetracycline. Plasmid presence in isolated colonies was next tested on 2YTG plates containing erythromycin (20 µg/mL), and erythromycin-sensitive clones were selected for further analysis.

Sporulation assays. Sporulation assays were based on survival to heat shock, as similarly described⁷¹. Briefly, triplicates precultures were carried out overnight in 2YTG medium. The following day, serum bottles containing 20 mL 2YTG (15 g/L glucose, pH 6.8) were inoculated with 2 mL of preculture and subsequently sealed. After 5 days at 34 °C without agitation, 1 mL of culture was heat shocked (10 min, 80 °C) and serially diluted. 5 µL spots were made for each dilution on 2YTG plates, which were then incubated overnight at 34 °C.

Sugar utilization assays. For liquid sugar utilization tests, liquid 2YTG (pH 6.8) was used, as well as 2YTL (pH 6.8, 20 g/L lactose) and 2YT (pH 6.8, without purified carbon source). Precultures were carried out overnight from fresh colonies in 2YTG medium (pH 6.8) containing 5 g/L glucose and 20 µg/mL erythromycin at several dilutions (up to 10⁻⁴ dilution factor). The following day, fresh precultures were used to inoculate 20 mL (AA strain; serum bottles)/5 mL (Δ sigL strain; 24 deep-well plates) of 2YTG, 2YTL or 2YT media with or without (as needed) erythromycin 20 µg/mL at 0.01 units of OD₆₀₀. Samples were taken after 24 and 96 hours, and sugar concentration was quantified as described above.

Minimal medium (CDM) was used to visualize sugar uptake inhibition by the AA strain. This medium, described by Vasconcelos *et al.*⁷⁸, was supplemented with 15 g/L agar and contained 20 g/L glucose, lactose or cellobiose. Single colonies picked on CDM glucose (5 g/L) plates were serially diluted in water. Spots were made with 5 µL of the 10⁻¹, 10⁻², 10⁻³ and 10⁻⁴ dilutions on minimal medium plates containing glucose, cellobiose or lactose (20 g/L). Plates were visualized after 24 h (glucose, cellobiose) or 48 h (lactose).

Data Availability

Raw reads corresponding to the genome sequencing of the AA mutant have been deposited to the Sequence Read Archive (SRA) database (<https://www.ncbi.nlm.nih.gov/sra>) with the BioProject accession number PRJ-NA530065.

References

1. IHS Chemical. Petrochemical Industry Overview. Chemical Economics Handbook (2017).
2. Naik, S. N., Goud, V. V., Rout, P. K. & Dalai, A. K. Production of first and second generation biofuels. A comprehensive review. *Renewable and Sustainable Energy Reviews* **14**, 578–597, <https://doi.org/10.1016/j.rser.2009.10.003> (2010).
3. Weusthuis, R. A., Aarts, J. M. & Sanders, J. P. M. From biofuel to bioproduct: is bioethanol a suitable fermentation feedstock for synthesis of bulk chemicals? *Biofuels, Bioproducts and Biorefining* **5**, 486–494 (2011).
4. Chen, J.-S. & Hiu, S. F. Acetone-butanol-isopropanol production by *Clostridium beijerinckii* (synonym, *Clostridium butylicum*). *Biotechnology Letters* **8**, 371–376, <https://doi.org/10.1007/BF01040869> (1986).
5. Zhang, C., Li, T. & He, J. Characterization and genome analysis of a butanol-isopropanol-producing *Clostridium beijerinckii* strain BGS1. *Biotechnology for Biofuels* **11**, 280, <https://doi.org/10.1186/s13068-018-1274-x> (2018).
6. Grisales Diaz, V. H. & Olivar Tost, G. Energy efficiency of acetone, butanol, and ethanol (ABE) recovery by heat-integrated distillation. *Bioprocess and Biosystems Engineering* **41**, 395–405, <https://doi.org/10.1007/s00449-017-1874-z> (2018).
7. Pyne, M. E., Bruder, M., Moo-Young, M., Chung, D. A. & Chou, C. P. Technical guide for genetic advancement of underdeveloped and intractable *Clostridium*. *Biotechnology advances* **32**, 623–641, <https://doi.org/10.1016/j.biotechadv.2014.04.003> (2014).
8. Máté de Gérando, H. *et al.* Genome and transcriptome of the natural isopropanol producer *Clostridium beijerinckii* DSM6423. *BMC genomics* **19**, 242 (2018).
9. Máté de Gérando, H. *et al.* Improving isopropanol tolerance and production of *Clostridium beijerinckii* DSM 6423 by random mutagenesis and genome shuffling. *Applied Microbiology and Biotechnology* **100**, 5427–5436, <https://doi.org/10.1007/s00253-016-7302-5> (2016).
10. Dürre, P., Kuhn, A. & Gottschalk, G. Treatment with allyl alcohol selects specifically for mutants of *Clostridium acetobutylicum* defective in butanol synthesis. *FEMS Microbiology Letters* **36**, 77–81 (1986).
11. Buck, M., Gallegos, M.-T., Studholme, D. J., Guo, Y. & Gralla, J. D. The Bacterial Enhancer-Dependent ζ 54(ζ N) Transcription Factor. *Journal of Bacteriology* **182**, 4129–4136, <https://doi.org/10.1128/JB.182.15.4129-4136.2000> (2000).
12. Zhang, N., Darbari, V. C., Glyde, R., Zhang, X. & Buck, M. The bacterial enhancer-dependent RNA polymerase. *Biochemical Journal* **473**, 3741–3753 (2016).
13. Barrios, H., Valderrama, B. & Morett, E. Compilation and analysis of sigma(54)-dependent promoter sequences. *Nucleic Acids Research* **27**, 4305–4313 (1999).
14. Kashket, E. R. & Cao, Z.-Y. Clostridial strain degeneration. *FEMS microbiology reviews* **17**, 307–315 (1995).
15. Söhling, B. & Gottschalk, G. Molecular analysis of the anaerobic succinate degradation pathway in *Clostridium kluyveri*. *Journal of Bacteriology* **178**, 871–880 (1996).
16. Hiu, S. F., Zhu, C.-X., Yan, R.-T. & Chen, J.-S. Butanol-ethanol dehydrogenase and butanol-ethanol-isopropanol dehydrogenase: different alcohol dehydrogenases in two strains of *Clostridium beijerinckii* (*Clostridium butylicum*). *Applied and Environmental Microbiology* **53**, 697–703 (1987).
17. Welch, R. W., Rudolph, F. B. & Papoutsakis, E. T. Purification and characterization of the NADH-dependent butanol dehydrogenase from *Clostridium acetobutylicum* (ATCC 824). *Archives of Biochemistry and Biophysics* **273**, 309–318, [https://doi.org/10.1016/0003-9861\(89\)90489-X](https://doi.org/10.1016/0003-9861(89)90489-X) (1989).
18. Ismaiel, A. A., Zhu, C. X., Colby, G. D. & Chen, J. S. Purification and characterization of a primary-secondary alcohol dehydrogenase from two strains of *Clostridium beijerinckii*. *Journal of Bacteriology* **175**, 5097–5105, <https://doi.org/10.1128/jb.175.16.5097-5105.1993> (1993).
19. Korkhin, Y. *et al.* NADP-dependent bacterial alcohol dehydrogenases. Crystal structure, cofactor-binding and cofactor specificity of the ADHs of *Clostridium beijerinckii* and *Thermoanaerobacter brockii* 11 Edited by R. Huber. *Journal of Molecular Biology* **278**, 967–981, <https://doi.org/10.1006/jmbi.1998.1750> (1998).
20. Raedts, J., Siemerink, M. A. J., Levisson, M., van der Oost, J. & Kengen, S. W. M. Molecular characterization of an NADPH-dependent acetoin reductase/2,3-butanediol dehydrogenase from *Clostridium beijerinckii* NCIMB 8052. *Applied and Environmental Microbiology* **80**, 2011–2020, <https://doi.org/10.1128/AEM.04007-13> (2014).

21. Rappas, M., Bose, D. & Zhang, X. Bacterial enhancer-binding proteins: unlocking σ^{54} -dependent gene transcription. *Current Opinion in Structural Biology* **17**, 110–116 (2007).
22. Bush, M. & Dixon, R. The Role of Bacterial Enhancer Binding Proteins as Specialized Activators of σ^{54} -Dependent Transcription. *Microbiol. Mol. Biol. Rev.* **76**, 497, <https://doi.org/10.1128/MMBR.00006-12> (2012).
23. Sedlar, K. *et al.* Complete genome sequence of *Clostridium pasteurianum* NRRL B-598, a non-type strain producing butanol. *Journal of biotechnology* **214**, 113–114, <https://doi.org/10.1016/j.jbiotec.2015.09.022> (2015).
24. Sedlar, K., Kolek, J., Provaznik, I. & Patakova, P. Reclassification of non-type strain *Clostridium pasteurianum* NRRL B-598 as *Clostridium beijerinckii* NRRL B-598. *Journal of biotechnology* **244**, 1–3 (2017).
25. Mitchell, W. J. The Phosphotransferase System in Solventogenic Clostridia. *Journal of molecular microbiology and biotechnology* **25**, 129–142, <https://doi.org/10.1159/000375125> (2015).
26. Schürmann, M. & Sprenger, G. A. Fructose-6-phosphate Aldolase Is a Novel Class I Aldolase from *Escherichia coli* and Is Related to a Novel Group of Bacterial Transaldolases.
27. Johnson, E. A., Burke, S. K., Forage, R. G. & Lin, E. C. Purification and properties of dihydroxyacetone kinase from *Klebsiella pneumoniae*. *Journal of Bacteriology* **160**, 55–60 (1984).
28. Cornillot, E., Nair, R. V., Papoutsakis, E. T. & Soucaille, P. The genes for butanol and acetone formation in *Clostridium acetobutylicum* ATCC 824 reside on a large plasmid whose loss leads to degeneration of the strain. *J Bacteriol* **179**, <https://doi.org/10.1128/jb.179.17.5442-5447.1997> (1997).
29. Collas, F. *et al.* Simultaneous production of isopropanol, butanol, ethanol and 2, 3-butanediol by *Clostridium acetobutylicum* ATCC 824 engineered strains. *Amb Express* **2**, 45 (2012).
30. Dusseau, S., Croux, C., Soucaille, P. & Meynial-Salles, I. Metabolic engineering of *Clostridium acetobutylicum* ATCC 824 for the high-yield production of a biofuel composed of an isopropanol/butanol/ethanol mixture. *Metab Eng* **18**, <https://doi.org/10.1016/j.ymben.2013.03.003> (2013).
31. Lee, J. *et al.* Metabolic Engineering of *Clostridium acetobutylicum* ATCC 824 for Isopropanol-Butanol-Ethanol Fermentation. *Applied and Environmental Microbiology* **78**, 1416–1423, <https://doi.org/10.1128/AEM.06382-11> (2012).
32. Wasels, F., Jean-Marie, J., Collas, F., López-Contreras, A. M. & Ferreira, N. L. A two-plasmid inducible CRISPR/Cas9 genome editing tool for *Clostridium acetobutylicum*. *Journal of microbiological methods* (2017).
33. Wang, Y. *et al.* Bacterial genome editing with CRISPR-Cas9: deletion, integration, single nucleotide modification, and desirable “clean” mutant selection in *Clostridium beijerinckii* as an example. *ACS synthetic biology* **5**, 721–732 (2016).
34. Li, Q. *et al.* CRISPR-based genome editing and expression control systems in *Clostridium acetobutylicum* and *Clostridium beijerinckii*. *Biotechnology journal* **11**, 961–972 (2016).
35. Wang, Y. *et al.* Markerless chromosomal gene deletion in *Clostridium beijerinckii* using CRISPR/Cas9 system. *Journal of biotechnology* **200**, 1–5 (2015).
36. Xue, C., Zhao, J., Chen, L., Yang, S.-T. & Bai, F. Recent advances and state-of-the-art strategies in strain and process engineering for biobutanol production by *Clostridium acetobutylicum*. *Biotechnology advances* **35**, 310–322, <https://doi.org/10.1016/j.biotechadv.2017.01.007> (2017).
37. Merrick, M. & Chambers, S. The helix-turn-helix motif of sigma 54 is involved in recognition of the -13 promoter region. *Journal of Bacteriology* **174**, 7221–7226 (1992).
38. Coppard, J. R. & Merrick, M. J. Cassette mutagenesis implicates a helix-turn-helix motif in promoter recognition by the novel RNA polymerase sigma factor sigma 54. *Mol Microbiol* **5**, 1309–1317, <https://doi.org/10.1111/j.1365-2958.1991.tb00777.x> (1991).
39. Lutstorf, U. & Megnet, R. Multiple forms of alcohol dehydrogenase in *Saccharomyces cerevisiae*. *Archives of Biochemistry and Biophysics* **126**, 933–944, [https://doi.org/10.1016/0003-9861\(68\)90487-6](https://doi.org/10.1016/0003-9861(68)90487-6) (1968).
40. Ciriacy, M. Genetics of alcohol dehydrogenase in *Saccharomyces cerevisiae*. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **29**, 315–325, [https://doi.org/10.1016/0027-5107\(75\)90053-6](https://doi.org/10.1016/0027-5107(75)90053-6) (1975).
41. Rando, R. R. Allyl alcohol-induced irreversible inhibition of yeast alcohol dehydrogenase. *Biochemical Pharmacology* **23**, 2328–2331, [https://doi.org/10.1016/0006-2952\(74\)90563-2](https://doi.org/10.1016/0006-2952(74)90563-2) (1974).
42. Lorowitz, W. & Clark, D. *Escherichia coli* mutants with a temperature-sensitive alcohol dehydrogenase. *J. Bacteriol.* **152**, 935 (1982).
43. Jones, D. T. & Woods, D. R. Acetone-butanol fermentation revisited. *Microbiological Reviews* **50**, 484–524 (1986).
44. Wiesenborn, D. P., Rudolph, F. B. & Papoutsakis, E. T. Coenzyme A transferase from *Clostridium acetobutylicum* ATCC 824 and its role in the uptake of acids. *Applied and Environmental Microbiology* **55**, 323–329 (1989).
45. Cary, J. W., Petersen, D. J., Papoutsakis, E. T. & Bennett, G. N. Cloning and expression of *Clostridium acetobutylicum* ATCC 824 acetoacetyl-coenzyme A: acetate/butyrate: coenzyme A-transferase in *Escherichia coli*. *Applied and Environmental Microbiology* **56**, 1576–1583 (1990).
46. Winzer, K., Lorenz, K. & Dürre, P. Acetate kinase from *Clostridium acetobutylicum*: a highly specific enzyme that is actively transcribed during acidogenesis and solventogenesis. *Microbiology* **143**, 3279–3286 (1997).
47. Boynton, Z. L., Bennett, G. N. & Rudolph, F. B. Cloning, sequencing, and expression of genes encoding phosphotransacetylase and acetate kinase from *Clostridium acetobutylicum* ATCC 824. *Appl Environ Microbiol* **62** (1996).
48. Cary, J. W., Petersen, D. J., Papoutsakis, E. T. & Bennett, G. N. Cloning and expression of *Clostridium acetobutylicum* phosphotransbutyrylase and butyrate kinase genes in *Escherichia coli*. *Journal of Bacteriology* **170**, 4613–4618 (1988).
49. Hartmanis, M. G. Butyrate kinase from *Clostridium acetobutylicum*. *Journal of Biological Chemistry* **262**, 617–621 (1987).
50. Wiesenborn, D. P., Rudolph, F. B. & Papoutsakis, E. T. Phosphotransbutyrylase from *Clostridium acetobutylicum* ATCC 824 and its role in acidogenesis. *Applied and Environmental Microbiology* **55**, 317–322 (1989).
51. Thompson, D. K. & Chen, J. S. Purification and properties of an acetoacetyl coenzyme A-reacting phosphotransbutyrylase from *Clostridium beijerinckii* (“*Clostridium butylicum*”) NRRL B593. *Applied and Environmental Microbiology* **56**, 607–613 (1990).
52. Boynton, Z. L., Bennet, G. N. & Rudolph, F. B. Cloning, sequencing, and expression of clustered genes encoding beta-hydroxybutyryl-coenzyme A (CoA) dehydrogenase, crotonase, and butyryl-CoA dehydrogenase from *Clostridium acetobutylicum* ATCC 824. *Journal of Bacteriology* **178**, 3015–3024 (1996).
53. Colby, G. D. & Chen, J. S. Purification and properties of 3-hydroxybutyryl-coenzyme A dehydrogenase from *Clostridium beijerinckii* (“*Clostridium butylicum*”) NRRL B593. *Applied and Environmental Microbiology* **58**, 3297–3302 (1992).
54. 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*. *Journal of Bacteriology* **190**, 843–850, <https://doi.org/10.1128/JB.01417-07> (2008).
55. Nie, X. *et al.* PTS regulation domain-containing transcriptional activator CelR and sigma factor σ^{54} control cellobiose utilization in *Clostridium acetobutylicum*. *Molecular Microbiology* **100**, 289–302, <https://doi.org/10.1111/mmi.13316> (2016).
56. Merrick, M. J. In a class of its own - the RNA polymerase sigma factor σ^{54} (σ^N). *Mol Microbiol* **10**, 903–909, <https://doi.org/10.1111/j.1365-2958.1993.tb00961.x> (1993).
57. Paget, M. S. B. & Helmann, J. D. The σ^{70} family of sigma factors. *Genome biology* **4**, 203 (2003).
58. Jovanovic, M. *et al.* Regulation of the co-evolved HrpR and HrpS AAA+ proteins required for *Pseudomonas syringae* pathogenicity. *Nature communications* **2**, 177 EP – (2011).
59. Bernard, C. S., Brunet, Y. R., Gavioli, M., Llobès, R. & Cascales, E. Regulation of Type VI Secretion Gene Clusters by σ^{54} and Cognate Enhancer Binding Proteins. *Journal of Bacteriology* **193**, 2158–2167, <https://doi.org/10.1128/JB.00029-11> (2011).

60. Fisher, M. A. *et al.* *Borrelia burgdorferi* σ^{54} is required for mammalian infection and vector transmission but not for tick colonization. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 5162–5167, <https://doi.org/10.1073/pnas.0408536102> (2005).
61. Visick, K. L. An intricate network of regulators controls biofilm formation and colonization by *Vibrio fischeri*. *Molecular Microbiology* **74**, 782–789, <https://doi.org/10.1111/j.1365-2958.2009.06899.x> (2009).
62. Wolfe, A. J., Millikan, D. S., Campbell, J. M. & Visick, K. L. *Vibrio fischeri* σ^{54} Controls Motility, Biofilm Formation, Luminescence, and Colonization. *Applied and Environmental Microbiology* **70**, 2520–2524, <https://doi.org/10.1128/AEM.70.4.2520-2524.2004> (2004).
63. Saldías, M. S., Lamothe, J., Wu, R. & Valvano, M. A. *Burkholderia cenocepacia* requires the RpoN Sigma Factor for Biofilm Formation and Intracellular Trafficking within Macrophages. *Infection and immunity* **76**, 1059–1067, <https://doi.org/10.1128/IAI.01167-07> (2008).
64. Martin-Verstraete, I., Débarbouillé, M., Klier, A. & Rapoport, G. Levanase operon of *Bacillus subtilis* includes a fructose-specific phosphotransferase system regulating the expression of the operon. *Journal of Molecular Biology* **214**, 657–671, [https://doi.org/10.1016/0022-2836\(90\)90284-S](https://doi.org/10.1016/0022-2836(90)90284-S) (1990).
65. Martin, I., Debarbouille, M., Ferrari, E., Klier, A. & Rapoport, G. Characterization of the levanase gene of *Bacillus subtilis* which shows homology to yeast invertase. *Molecular & general genetics: MGG* **208**, 177–184 (1987).
66. Martin-Verstraete, I. *et al.* Antagonistic effects of dual PTS-catalysed phosphorylation on the *Bacillus subtilis* transcriptional activator LevR. *Mol Microbiol* **28**, 293–303, <https://doi.org/10.1046/j.1365-2958.1998.00781.x> (1998).
67. Browning, D. F. & Busby, S. J. W. The regulation of bacterial transcription initiation. *Nature Reviews Microbiology* **2**, 57 EP – (2004).
68. Browning, D. F. & Busby, S. J. W. Local and global regulation of transcription initiation in bacteria. *Nature Reviews Microbiology* **14**, 638 EP – (2016).
69. Al-Hinai, M. A., Jones, S. W. & Papoutsakis, E. T. σ^K of *Clostridium acetobutylicum* Is the First Known Sporulation-Specific Sigma Factor with Two Developmentally Separated Roles, One Early and One Late in Sporulation. *Journal of Bacteriology* **196**, 287–299, <https://doi.org/10.1128/JB.01103-13> (2014).
70. Jones, S. W., Tracy, B. P., Gaida, S. M. & Papoutsakis, E. T. Inactivation of σ^F in *Clostridium acetobutylicum* ATCC 824 blocks sporulation prior to asymmetric division, abolishes σ^F - and σ^G -protein expression, but does not block solvent formation. *Journal of Bacteriology*, <https://doi.org/10.1128/JB.00088-11> (2011).
71. Tracy, B. P., Jones, S. W. & Papoutsakis, E. T. Inactivation of σ^E and σ^G in *Clostridium acetobutylicum* illuminates their roles in clostridial-cell-form biogenesis, granulose synthesis, solventogenesis, and spore morphogenesis. *Journal of Bacteriology* **193**, 1414–1426 (2011).
72. Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).
73. Vallenet, D. *et al.* MicroScope—an integrated microbial resource for the curation and comparative analysis of genomic and metabolic data. *Nucleic Acids Research* **41**, D636–D647 (2012).
74. Gapes, J. R., Nimcevic, D. & Friedl, A. Long-Term Continuous Cultivation of *Clostridium beijerinckii* in a Two-Stage Chemostat with On-Line Solvent Removal. *Applied and Environmental Microbiology* **62**, 3210–3219 (1996).
75. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6**, 343–345 (2009).
76. Mermelstein, L. D. & Papoutsakis, E. T. *In vivo* methylation in *Escherichia coli* by the *Bacillus subtilis* phage phi 3T I methyltransferase to protect plasmids from restriction upon transformation of *Clostridium acetobutylicum* ATCC 824. *Applied and Environmental Microbiology* **59**, 1077–1081 (1993).
77. Kolek, J., Sedlar, K., Provaznik, I. & Patakova, P. Dam and Dcm methylations prevent gene transfer into *Clostridium pasteurianum* NRRL B-598: development of methods for electrotransformation, conjugation, and sonoporation. *Biotechnology for Biofuels* **9**, 14 (2016).
78. Vasconcelos, I., Girbal, L. & Soucaille, P. Regulation of carbon and electron flow in *Clostridium acetobutylicum* grown in chemostat culture at neutral pH on mixtures of glucose and glycerol. *J Bacteriol* **176** (1994).

Acknowledgements

We are grateful to Michelle Kuntz for constructing the pFW-FC06 plasmid and to Gwladys Chartier for technical assistance.

Author Contributions

R.H., N.L.F. and F.W. designed the study. R.H. performed experiments on *C. beijerinckii* DSM 6423 and AA strains. M.B.-L. performed experiments regarding the *C. beijerinckii* NCIMB 8052 strain. R.H., M.B.-L., N.L.F. and F.W. wrote the paper.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-43822-2>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

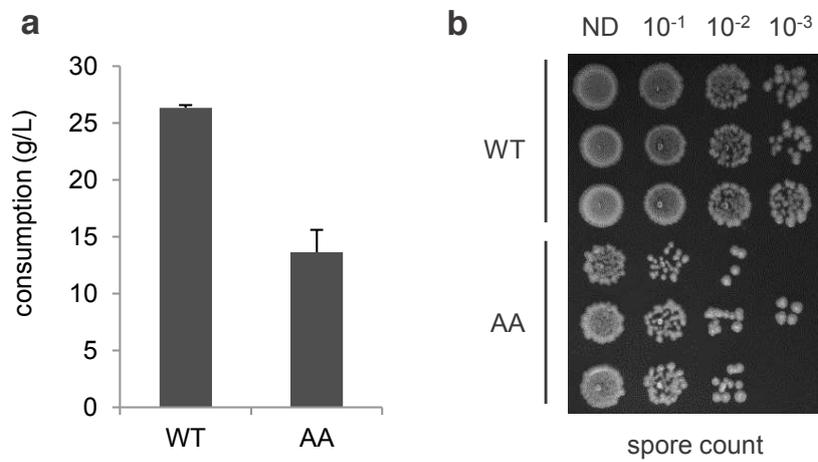
© The Author(s) 2019

**σ^{54} (σ^L) plays a central role in carbon metabolism in the industrially relevant
*Clostridium beijerinckii***

Rémi Hocq^a, Maxime Bouilloux-Lafont^a, Nicolas Lopes Ferreira^a and François Wasels^{a 1}
^aIFP Energies nouvelles, 1 et 4 avenue de Bois-Préau, 92852 Rueil-Malmaison, France

¹Corresponding author: François Wasels (francois.wasels@ifpen.fr)

Supplementary files



S1 : Comparative phenotypic analysis of *C. beijerinckii* DSM 6423 wild-type and AA strains, supplementary data. a. Final glucose consumption measured after 48h of fermentation in Gapes medium for wild-type *C. beijerinckii* DSM 6423 and the AA mutant. Biomass is given as a measure of optical density at 600 nm (OD_{600}). Error bars indicate the standard deviation of triplicate experiments. b. Sporulation assay on 2YTG plates comparing triplicate heat-shocked cultures of the wild-type and AA strains.

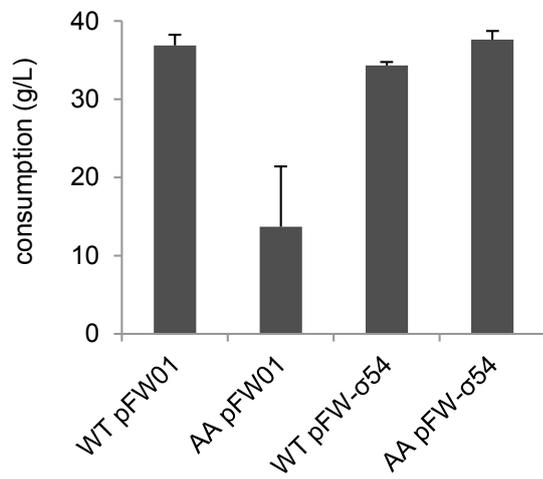
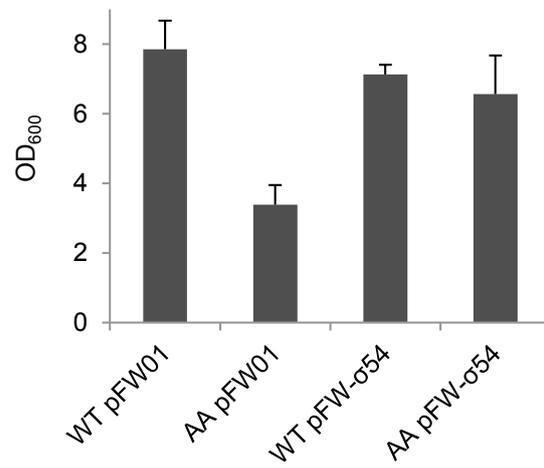
Supplementary file S1
(Wasels)

Locus tag	Automatic Annotation	Comment
CIBE_0239	Sigma-54 factor, interaction domain-containing protein	putative operon involved in PP pathway
CIBE_0709	PTS sugar transporter	putative operon involved in fructose/galactitol transport
CIBE_0715	Iditol 2-dehydrogenase	putative operon involved in PP pathway
CIBE_0882	phosphotransferase system (PTS) lichenan-specific enzyme IIA component	putative operon involved in lactose metabolism/transport
CIBE_1109	Phage-rlike protein	
CIBE_1140	chitosanase	
CIBE_1144	PTS lactose transporter subunit IIC	putative PTS transport encoding operon
CIBE_1149	PTS system fructose subfamily IIA component	putative sugar transport/metabolism encoding operon
CIBE_1157	PTS mannose transporter subunit IIA	putative operon involved in mannose metabolism/transport
CIBE_1696	acetoin reductase/2,3-butanediol dehydrogenase	
CIBE_1707	phosphotransferase system (PTS) lichenan-specific enzyme IIA component	putative operon involved in cellobiose metabolism/transport
CIBE_2050	NADPH-dependent butanol dehydrogenase	butanol metabolism
CIBE_2524	Acyl carrier protein	putative operon involved in fatty acid metabolism
CIBE_2531	4-hydroxybutyryl-CoA dehydratase/vinylacetyl-CoA-Delta-isomerase	putative operon involved in 4-hydroxybutyrate conversion to crotonyl-CoA
CIBE_2581	dihydroxyacetone kinase, N-terminal domain	putative operon yielding dihydroxyacetone kinase
CIBE_2622	NADPH-dependent butanol dehydrogenase	butanol metabolism
CIBE_2943	PTS fructose transporter subunit IIA	
CIBE_3081	ketopantoate hydroxymethyltransferase	putative operon involved in coenzyme A metabolism
CIBE_3134	conserved protein of unknown function	
CIBE_3200	cytidine deaminase	
CIBE_3470	NADP-dependent isopropanol dehydrogenase	putative operon, isopropanol metabolism
CIBE_3479	Methyl-accepting chemotaxis protein McpA	
CIBE_3483	ABC transporter substrate-binding protein	putative operon with CoA transferase
CIBE_3485	Citrate transporter	
CIBE_4059	Acetate CoA-transferase YdiF	putative operon
CIBE_4264	putative PTS system fructose subfamily IIA component	putative operon involved in glycosaminoglycan metabolism
CIBE_4670	fragment of fused mannose-specific PTS enzymes: IIA component ; IIB component (part 1)	mannose metabolism
CIBE_4671	putative Transcriptional regulatory protein LevR	
CIBE_5166	Flagellar hook-associated protein FlgK	
CIBE_5246	conserved protein of unknown function	
CIBE_5470	PTS system fructose subfamily IIA component	putative operon involved in fructosamine metabolism
CIBE_5548	phosphotransferase system (PTS) lichenan-specific enzyme IIA component	putative operon involved in sugar transport
CIBE_5598	phosphotransferase system (PTS) lichenan-specific enzyme IIA component	putative operon involved in sugar transport
CIBE_5606	Membrane spanning protein	

PP, pentose phosphate

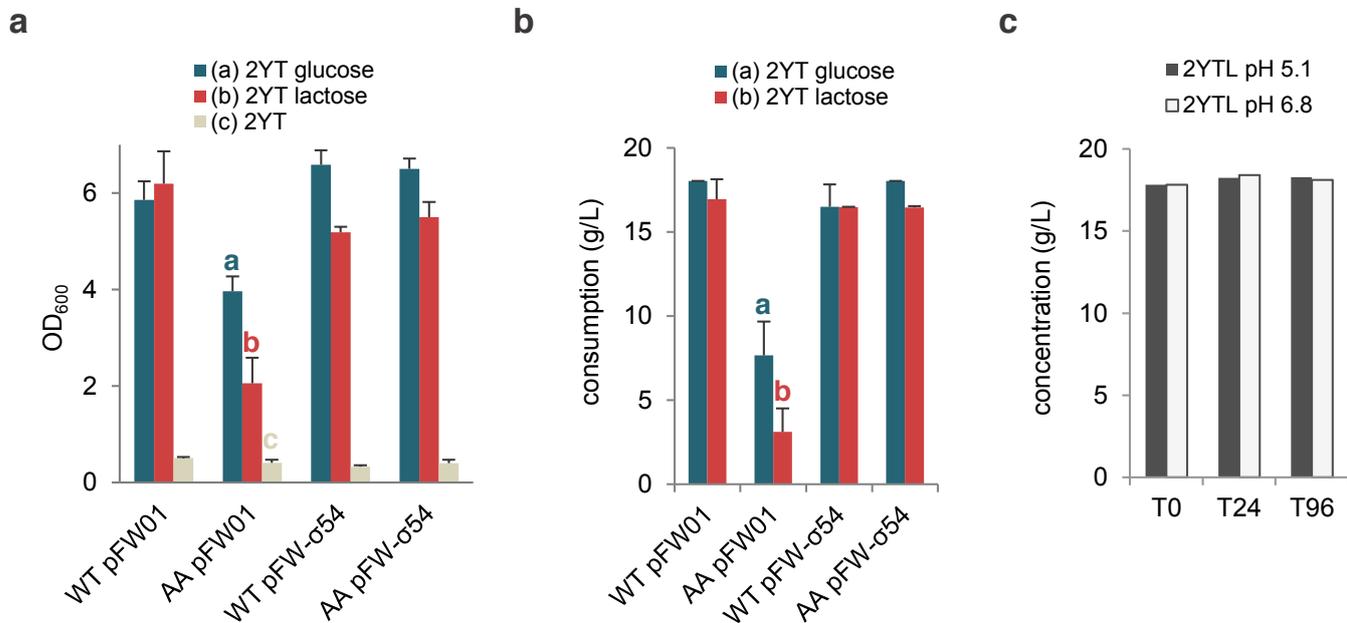
S2 : List of *C. beijerinckii* DSM 6423 genes predicted to be controlled by σ^{54} . Predictions are based on the presence of the σ^{54} consensus motif (TGGCANNNNNTTGCW) in the correct orientation up to 500 bp upstream of a given coding sequence.

Supplementary file S2
(Wasels)

a**b**

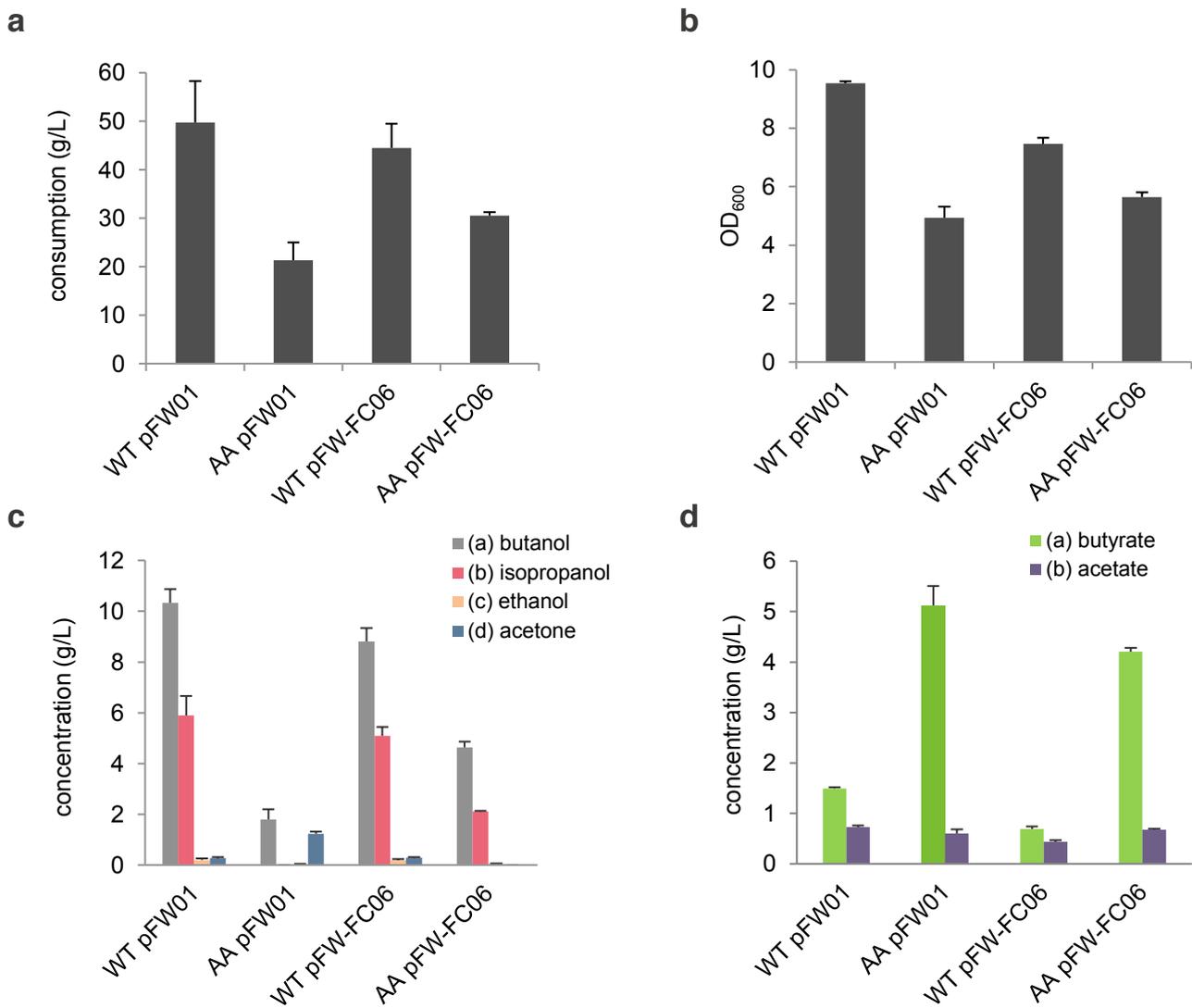
S3 : Phenotypic analysis of *C. beijerinckii* DSM 6423 σ^{54} complementation in the AA strain, supplementary data 1. a,b, Final (a) glucose consumption and (b) biomass measured after 48h of fermentation in Gapes medium for wild-type *C. beijerinckii* DSM 6423 (WT) and the AA mutant containing an empty plasmid (pFW01) or the complementation plasmid (pFW01- σ^{54}). Biomass is given as a measure of optical density at 600 nm (OD₆₀₀). Error bars indicate the standard deviation of triplicate experiments.

Supplementary file S3
(Wasels)



S4: Phenotypic analysis of *C. beijerinckii* DSM 6423 σ^{54} complementation in the AA strain, supplementary data 2. a,b, Sugar utilization assay in liquid 2YT with 20 g/L glucose, lactose or without added carbon source. Biomass (optical density at 600 nm, OD₆₀₀) (a) and carbon consumption (b) are measured 96h after the beginning of the fermentation. c. Lactose does not degrade during a 96-hour period in a 2YT-based medium. Error bars indicate the standard deviation of triplicate experiments.

Supplementary file S4
(Wasels)



S5: Comparative fermentation assays in *C. beijerinckii* DSM 6423 wild-type and AA strains overexpressing *s-adh* and *ctfAB*. a,b,c,d Final (a) glucose consumption, (b) biomass, (c) solvent and (d) acid concentrations measured after 72h of fermentation in Gapes medium for wild-type *C. beijerinckii* DSM 6423 (WT) and the AA mutant containing an empty plasmid (pFW01) or the pFW-FC06 plasmid. Biomass is given as a measure of optical density at 600 nm (OD₆₀₀). Error bars indicate the standard deviation of triplicate experiments.

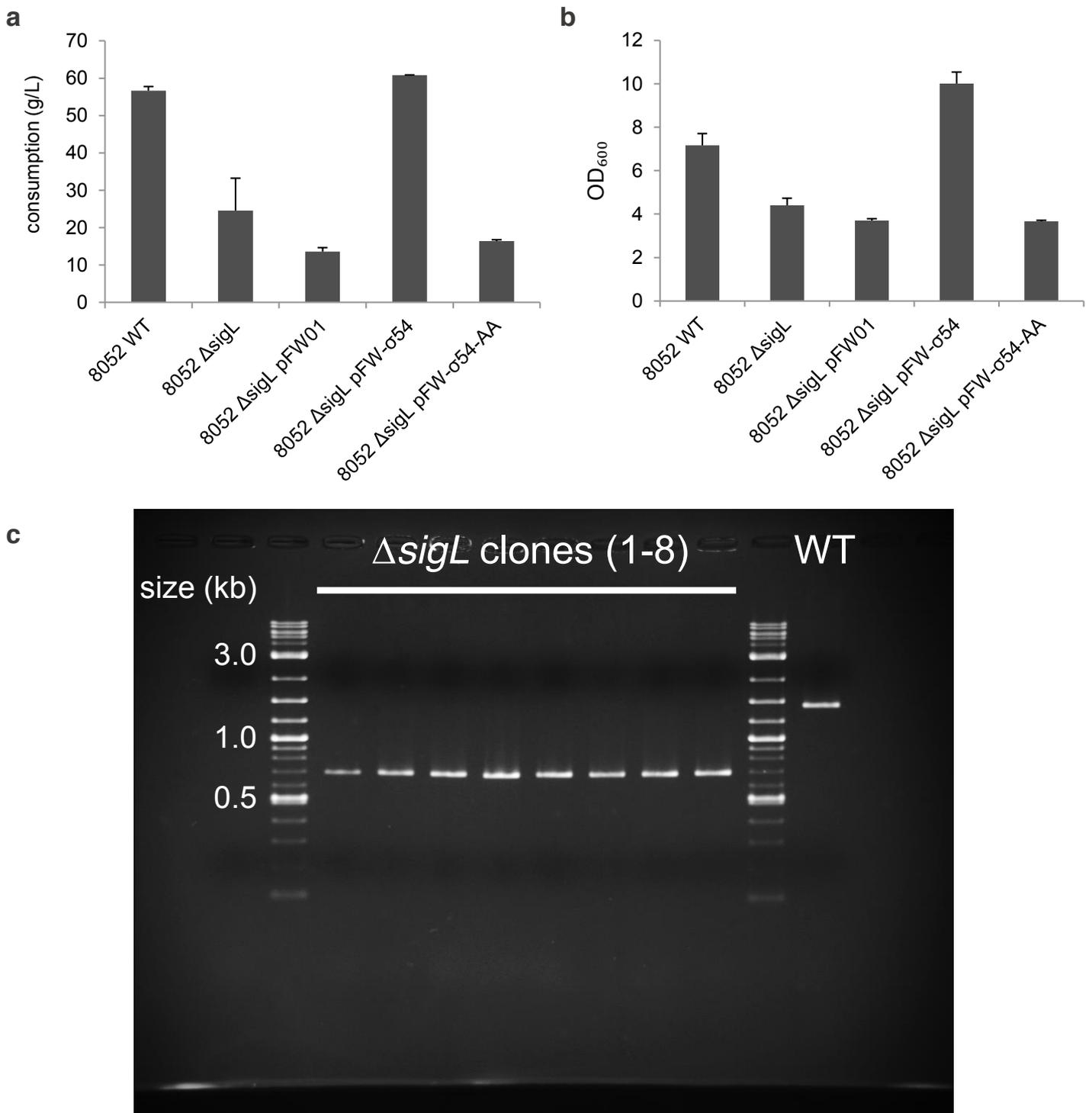
Supplementary file S5
(Wasels)

Name	Sequence (5' → 3')
Construction of pFW-σ54	
RH077	ACTGTTCTCGAGGACCTGGATGCTGTGGATTAG
RH078	TGACACGTCGACTTATAATCTCTTCCTTGCTGAAGATG
RH086	CAGCAAGGAAGAGATTATAAGTCGACAAAGTATTGTTAAAAATAACTCTG
RH087	AAAACTAATCCACAGCATCCAGGTCCTCGAGCCGCTTATAATCCATAAC
Construction of pFW-σ54-AA	
RH136	TCAATCTTCATGAATTTACCGTAAGTAGAGC
RH137	GCTCTACTTACGGTAAATTCATGAAGATTGA
Construction of pΔsigL	
RH125	TTGTTATGGATTATAAGCGGCTCGAG
RH126	CAATCTATCACTGATAGGGACTCGAG
guide_sigL_fwd	TCATATTATATTCTGATAATTACA
guide_sigL_rev	AAACTGTAATTATCAGAATATAAT
Verification of sigL deletion	
RH129	GGTGTTAACTTTATTGAATGTCC
RH130	GCAACTTCTTTTAATGTCAAAGG
Construction of pFW-FC06	
Mkz01	AAAAAAGGTACCCAAGTATTTTTTATTACATTAATATAGTTAAA
Mkz02	AAAAAAGTCGACTGTTTCATAGTATTTCTTTCT

Target	Name	Sequence (5' → 3')	qPCR efficiency (%)
<i>sigL</i>	0767_F	AGGGTATTTGGAGATATCAACACGA	106,2
	0767_R	GCAATCAACGACTTAACAGATGCT	
<i>gapC</i>	0769_F	AGGGTATTTGGAGATATCAACACGA	104,6
	0769_R	TGGTTCCAATGCCTGTACCA	
<i>adh-1</i>	2050_F	GGAGTAGAACCAGATCCATCAGT	105,2
	2050_R	GTGATCCTCCACCCATAGCAA	
<i>EBP-A</i>	2621_F	GGATGGGACCTTTACAGGAGG	105,3
	2621_R	CGGCATTTCCCAATTTCTGT	
<i>adh-2</i>	2622_F	ACAGTAATGAAAGGCGCAGAAG	106,3
	2622_R	TCCACATAGCCTTTGCAGCA	
<i>s-adh</i>	3470_F	AATTGGCATTGGAGCTGTTGG	105,7
	3470_R	CTCAACACAAATCGGCCTGC	
<i>EBP-B</i>	3471_F	TTGTGGCAATAAACTGTGGCG	105,2
	3471_R	CACCCTCCTCTTTTAGCTCCC	
<i>ald</i>	4606_F	CTCCATATGGCGTTATAGGTGCA	108,1
	4606_R	ACTACAGCATTTCAGCAGCT	

S6 : List of primers used in this study.

Supplementary file S6
(Wasels)



S7: CRISPR/Cas9-mediated deletion of σ^{54} in *C. beijerinckii* NCIMB 8052, complementation and comparative fermentation assays, supplementary data. a,b, Final (a) glucose consumption and (b) biomass measured after 72h of fermentation in Gapes medium for wild-type (WT) *C. beijerinckii* NCIMB 8052 and the $\Delta sigL$ mutant containing an empty plasmid (pFW01), the pFW- σ^{54} plasmid or the pFW- σ^{54} -AA plasmid. Biomass is given as a measure of optical density at 600 nm (OD₆₀₀). Error bars indicate the standard deviation of triplicate experiments. c. Verification by colony PCR of *sigL* deletion with primers RH129 and RH130 encompassing the deleted region, full agarose gel corresponding to figure 4b. Expected band sizes are 1406 (wild-type) and 670 ($\Delta sigL$) bp.

Supplementary file S7
(Wasels)

5 PARTIE III : ANALYSE COMPARATIVE DES TSS DANS TROIS *CLOSTRIDIA SOLVANTOGENES*

L'étude spécifique de la sous-unité alternative σ^{54} de l'ARN polymérase chez *C. beijerinckii* illustre parfaitement l'importance des phénomènes moléculaires qui contribuent au démarrage de la transcription d'un gène bactérien. Mais, étant donné le peu de littérature dédié à cette espèce, il est aussi important d'acquérir des données plus globales sur la régulation de l'expression des gènes. Dans cette partie de la thèse, les événements d'initiation de la transcription ont été étudiés à l'échelle du génome pour *C. beijerinckii* DSM 6423 ainsi que pour deux autres souches solvantogènes modèles.

5.1 INTRODUCTION

La régulation de l'expression génétique – concept-clé en biologie moléculaire – peut subvenir de multiples façons, à plusieurs niveaux et détermine *in fine* la réaction cellulaire vis-à-vis de stimuli environnementaux. Chez les bactéries, la quantité de transcrits est par exemple finement contrôlée par de nombreux facteurs qui influencent dynamiquement la vie des ARNm, de leur synthèse à leur dégradation. L'un des mécanismes de contrôle transcriptionnel principaux est la régulation précise du choix et de la dynamique d'utilisation du/des site(s) d'initiation de la transcription d'un gène donné²⁰³. Dans cette optique, trouver des méthodes pour mesurer la dynamique de choix et d'utilisation des sites d'initiation de la transcription (*Transcription Start Site*, TSS) peut fournir des informations pertinentes pour la compréhension de la régulation de la transcription bactérienne.

Avec l'avènement des méthodes de séquençage à haut débit, la détermination des TSS procaryotiques a été grandement facilitée par des approches dérivées de la technologie RNA-seq. Ces méthodes tirent toutes parti de la présence caractéristique d'un groupement triphosphate en 5' du premier nucléotide des ARN non maturés (comme les ARN messagers). Le dRNA-seq (*differential* RNA-seq) utilise par exemple un traitement à l'exonucléase terminale pour enlever les ARN maturés (comme les ARNr et ARNt) avant l'étape de séquençage²¹⁶. Les premiers travaux utilisant cette

méthode suggéraient toutefois qu'elle n'enrichissait que partiellement en transcrits 5' triphosphates, et nécessitait donc le séquençage d'une librairie de comparaison pour laquelle le traitement enzymatique n'avait pas été réalisé.

Des approches plus récentes ont démontré que ce besoin d'expérience contrôle pouvait être contourné en purifiant spécifiquement les transcrits 5' triphosphates^{217,218}. Dans cette méthode, les ARN sont sélectivement ciblés par une enzyme (*Vaccinia capping enzyme*, VCE), qui ajoute en 5' une coiffe guanosine modifiée avec une desthiobiotine (*Cappable-seq*) ou une biotine (*Capp-Switch seq*), ce qui permet leur capture par affinité avec de la streptavidine. Dans cette partie de la thèse, nous avons utilisé, en collaboration avec le Génoscope, l'approche *Capp-Switch seq* pour cartographier, à l'échelle du génome, les TSS de trois souches solvantogènes du genre *Clostridium*. Nous montrons tout d'abord que les jeux de données *Capp-Switch seq* générés dans cette étude sont fortement biaisés et favorisent systématiquement la détection de TSS pour des gènes très exprimés (jusqu'à plusieurs centaines de TSS secondaires pour un seul gène). Nous avons ensuite utilisé cette information pour corriger le pipeline d'analyse en incorporant une étape additionnelle. Cette dernière consiste à normaliser les valeurs d'expression des TSS par l'expression RNA-seq des gènes auxquels ils sont associés (Figure 15). Ce faisant, le nombre de TSS secondaires détectés pour les gènes surexprimés diminue drastiquement, et, à l'inverse, plusieurs milliers de TSS primaires ont pu être détectés pour des gènes peu ou moyennement exprimés.

Avec cette étape additionnelle, notre nouvelle stratégie (nommée eCAPP-seq pour *enhanced Capp-Switch seq*) a pu identifier un total de 12 124 TSS en phase exponentielle pour trois *Clostridia* (*C. acetobutylicum* ATCC 824, *C. beijerinckii* NCIMB 8052, *C. beijerinckii* DSM 6423). Si les analyses à l'échelle du transcriptome sont encore en cours, nous avons validé notre approche en comparant expérimentalement les TSS liés à l'opéron de synthèse du butyryl-CoA (opéron BCS¹⁴⁸) identifiés chez *C. beijerinckii* et *C. acetobutylicum* par *northern blot*.

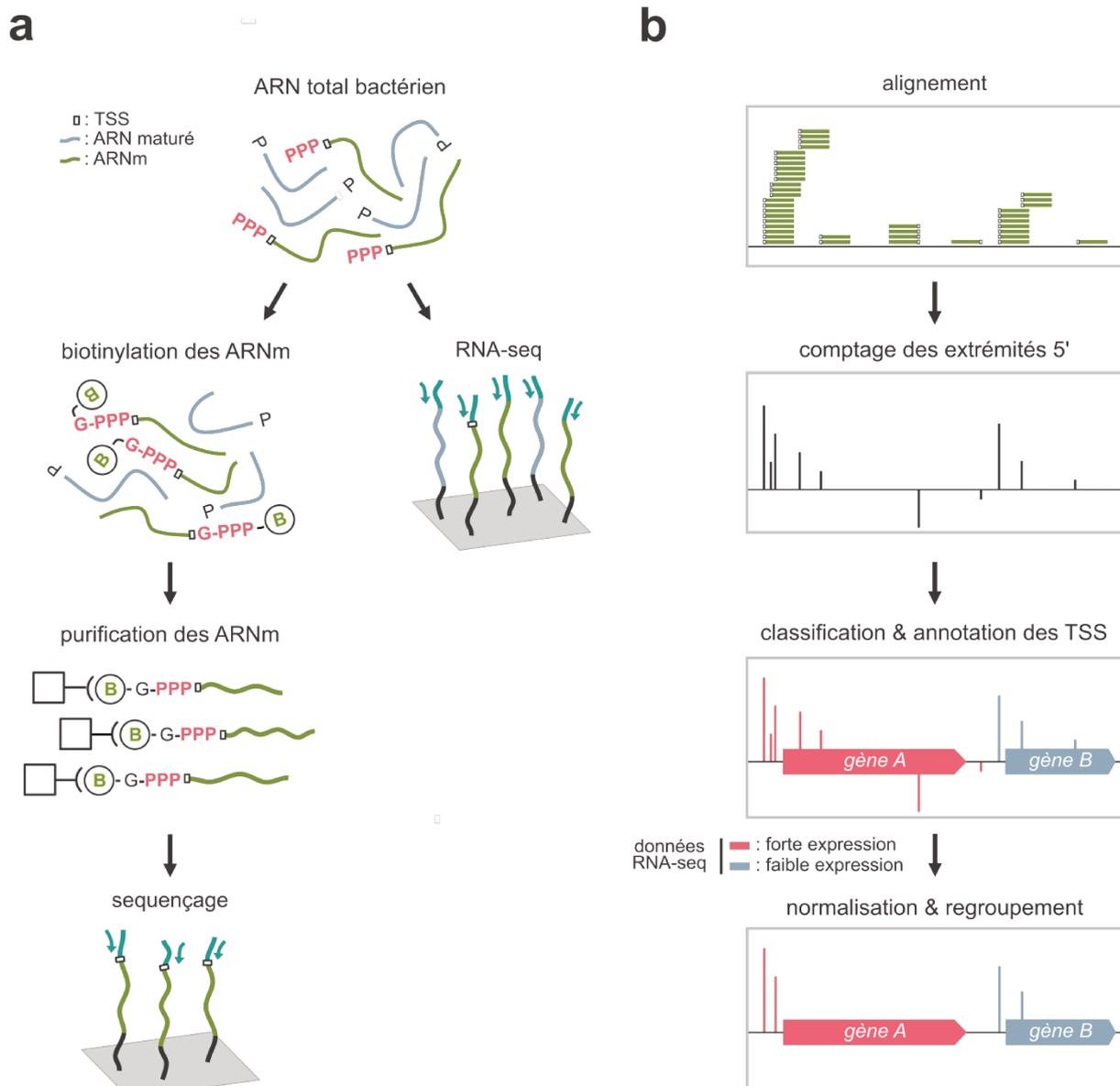


Figure 15 : **Préparation des bibliothèques et analyse de données eCAPP-seq.**

a. A partir d'un échantillon d'ARN bactériens, deux banques eCAPP-seq et RNA-seq sont constituées. Pour la première, les ARNm 5' triphosphates sont tout d'abord purifiés spécifiquement par affinité biotine/streptavidine. Après préparation des bibliothèques suivant une méthode basée sur le template-switching, les banques eCAPP-seq (et RNA-seq) sont séquençées sur un séquenceur Illumina. **b.** Le pipeline d'analyse suit plusieurs étapes. Après alignement des lectures sur le génome, l'extrémité 5' des lectures est comptée pour identifier des TSS putatifs. Chaque TSS est ensuite associé avec le gène le plus proche, et classifié en fonction de sa position vis-à-vis de ce gène. Les données d'expression RNA-seq sont utilisées pour normaliser l'expression des TSS, puis les TSS proches (fenêtre glissante < 6 pb) sont regroupés. Enfin, un seuil est appliqué sur les valeurs d'expression des TSS pour obtenir les jeux de données finaux.

5.2 MATÉRIEL ET MÉTHODES

5.2.1 Souches, milieu et conditions de culture

Les souches bactériennes (*C. acetobutylicum* ATCC 824, *C. beijerinckii* NCIMB 8052 et DSM 6423) sont cultivées en anaérobiose à 34°C en milieu 2YTG liquide (par litre : 16 g de tryptone, 10 g d'extrait de levure, 5 g de NaCl, 20 g de glucose, pH 5,2). Le milieu solide est préparé de manière similaire, avec 15 g/L d'agar, 5 g/L de glucose et sans ajustement du pH. Pour les cultures destinées aux expériences eCAPP-seq, 10 mL de culture en duplicats furent récoltés en phase semi-logarithmique ($DO_{600} \approx 0,5$) et immédiatement stabilisés sur glace avec une solution d'équilibrage froide (1,25 mL d'une mixture phénol acide : éthanol en proportions 1 : 18). Après centrifugation, les culots cellulaires ont été conservés à -80°C jusqu'à l'étape d'extraction des ARN.

5.2.2 Extraction des ARN

Les culots congelés ont été resuspendus dans du réactif TRIzol (Invitrogen). Suite à la lyse cellulaire, du chloroforme a été ajouté et la phase aqueuse obtenue après mélange / centrifugation a été récupérée. Après précipitation à l'isopropanol, l'ARN total a ensuite été traité avec la DNase Turbo (Invitrogen) puis purifié (*RNA Clean & Concentrator-5*, protocole de purification des ARN de plus de 200 pb, Zymo Research). Les ARN ribosomaux ont ensuite été déplétés avec le kit RiboZero (Illumina), et les ARN obtenus stockés à -80°C.

5.2.3 Préparation des bibliothèques RNA-seq

Les bibliothèques RNA-seq sont préparées avec un protocole de *template switching*. Après déplétion des ARNr, la MMLV reverse transcriptase (SMARTScribe, Takara) est utilisée pour obtenir les ADNc, avec le mix d'amorces *SMART stranded N6* (Takara). Les ADNc sont purifiés à l'aide de billes magnétiques (AMPure Beads, Beckman Coulter). La bibliothèque est ensuite obtenue par PCR en utilisant les billes comme matrice (polymérase SeqAmp, *Universal Forward PCR primer*, *Indexed Reverse PCR primer*, Takara). Après purification sur billes magnétiques, les bibliothèques sont séquencées sur un séquenceur MiSeq (Illumina).

5.2.4 Préparation des bibliothèques eCAPP-seq

Le protocole de préparation des bibliothèques *Capp-Switch seq* a été utilisé²¹⁸. Rapidement, une coiffe biotinyllée est d'abord ajoutée à l'extrémité 5' triphosphate des ARN en utilisant l'enzyme VCE. Après fragmentation, les ARN biotinyllés sont purifiés avec des billes magnétiques couplées à de la streptavidine. La synthèse des ADNc est réalisée directement sur ces billes avec la méthode dite de *template switching* en utilisant la reverse transcriptase SMARTscribe MMLV RT (Takara). Les ADNc, après élution, sont utilisés comme matrice pour une PCR (amorces *Universal Forward PCR primer* et *indexed Reverse PCR primer*, Takara), et les bibliothèques obtenues sont séquencées sur un séquenceur MiSeq (Illumina).

5.2.5 Traitement des données eCAPP-seq

Les trois premières paires de bases des lectures sens des jeux de données eCAPP-seq sont d'abord supprimées car celles-ci dérivent de la méthode *template switching* utilisée pendant la préparation des bibliothèques. Les lectures eCAPP-seq et RNA-seq sont ensuite alignées avec Geneious R10 sur les génomes adéquats (*C. acetobutylicum* ATCC 824 : GCA_000008765.1; *C. beijerinckii* NCIMB 8052 : GCA_000016965.1; *C. beijerinckii* DSM 6423 : GCA_900010805.1). 95-99% des lectures ont été alignées à des positions uniques, pour un total compris entre 0,73 million (réplicat 2 ATCC 824 eCAPP-seq) et 2,9 million (réplicat 1 NCIMB 8052 RNA-seq) lectures par échantillon (Figure Supplémentaire 1). Les données eCAPP-seq ont ensuite été traitées par une succession de scripts Perl. Les TSS ont d'abord été identifiés en comptant le nombre de lectures sens démarrant à chaque position du génome. Les annotations du génome ont ensuite été utilisées pour associer les TSS avec les gènes les plus proches et pour les classer en 4 sous-catégories²¹⁸ : InterS (TSS intergénique avec un gène dans la même orientation), InterA (TSS intergénique avec un gène dans l'orientation opposée), IntraS (TSS intragénique dans un gène avec la même orientation), IntraA (TSS intragénique dans un gène avec l'orientation opposée). Les réplicats sont ensuite comparés et les positions détectées dans un seul des deux jeux sont éliminées. Les annotations des gènes ont ensuite été utilisées pour associer les TSS aux jeux de données RNA-seq. Les jeux de données normalisés ont

été obtenus en divisant le nombre de lectures comptées pour chaque TSS par la moyenne (deux réplicats pour les souches DSM 6423 et NCIMB 8052, un réplikat pour la souche ATCC 824) des valeurs d'expression RNA-seq (en transcrits par million) du gène associé, puis en normalisant toutes les valeurs obtenues pour un total de 1 million de « lectures normalisées ». Les positions des TSS proches sont ensuite regroupées dans les jeux de données non normalisés et normalisés en ne gardant que les positions avec le plus haut nombre de lectures dans des fenêtres glissantes de 5 pb. Ces positions sont ensuite filtrées avec un seuil à 10 ou 25 (jeu de donnée à forte confiance) lectures par million de lectures.

5.2.6 Visualisation des données

Les distributions, graphes de corrélation et diagramme de Venn ont été réalisés avec R grâce aux *packages* ggplot2 et VennDiagram.

5.2.7 Northern blot

10 µg d'ARN dénaturé par traitement thermique par puits sont séparés sur un gel d'agarose 1,2% dénaturant (formaldéhyde 6,6 %). Après électrophorèse, les ARN sont transférés toute une nuit par capillarité avec du tampon SSC 10X (1X : 150 mM NaCl, 15 mM citrate de sodium, pH 7) sur une membrane de nitrocellulose (Hybond N+, GE Healthcare). Les sondes ont été obtenues en deux étapes. Tout d'abord, des fragments d'environ 300-400 pb ont été amplifiés pour chaque gène ciblé par PCR sur de l'ADN génomique. Puis ces produits de PCR ont eux-mêmes servi comme matrice pour une PCR unidirectionnelle (amplification de ≈ 250-350 pb du brin antisens, avec incorporation de α -³²P dATP – Perkin Elmer). Les amorces utilisées pour chacune de ces PCR sont données dans la Figure Supplémentaire 2. Les ARN ont ensuite été fixés de façon covalente sur les membranes de nitrocellulose par exposition aux UV, puis pré-bloqués 1 h à 65 °C avec de l'ADN de sperme de saumon dans du tampon de Church (0,5 M phosphate de sodium, pH 7, 7% SDS, 1 mM EDTA). Les sondes radioactives ont été ajoutées par la suite pour une hybridation de 6 h à 65 °C. Après deux lavages dans du tampon SSC 2X 0,5% SDS et un dans du tampon SSC 0,1X

0,5% SDS à 65°C, le signal radioactif est révélé par exposition sur un écran phosphor (GE Healthcare) et analysé sur un imageur Typhoon.

5.3 RÉSULTATS

5.3.1 La normalisation par l'expression améliore la détection des TSS

Pour identifier les TSS dans les génomes de *C. acetobutylicum* ATCC 824, *C. beijerinckii* DSM 6423 et *C. beijerinckii* NCIMB 8052, nous avons dans un premier temps réalisé un *Capp-Switch seq* avec de l'ARN obtenu de cultures en phase semi-exponentielle. Nous avons rapidement remarqué que les TSS identifiés par cette méthode avaient tendance à s'accumuler sur certains gènes, et en particulier sur ceux qui étaient décrits comme fortement exprimés dans d'autres études^{52,176}.

Nous avons donc supposé que le seuil final utilisé pour filtrer les TSS, formulé en termes de lectures par million (RPM, 10 RPM pour un *Capp-Switch seq*²¹⁸), devait être ajusté dynamiquement pour tenir compte de l'expression génétique local en aval de chaque TSS. Nous avons donc réalisé des séquençages de type RNA-seq sur les mêmes échantillons d'ARN qui avaient été utilisés pour préparer les bibliothèques de *Capp-Switch seq* (Figure 15a). Nous avons ensuite normalisé la force de chaque TSS (en RPM) par la valeur de l'expression (en transcrits par million, TPM) du gène le plus proche (Figure 15b). Du fait que chaque TSS identifié est associé à une valeur de RPM normalisée qui dépend de l'expression génétique, cette méthode, que nous avons nommée eCAPP-seq, nous a permis de garder en pratique un seuil en RPM normalisé invariable et simple à la fin du pipeline d'analyse des données (10 RPM normalisés, ou 25 pour une plus haute confiance).

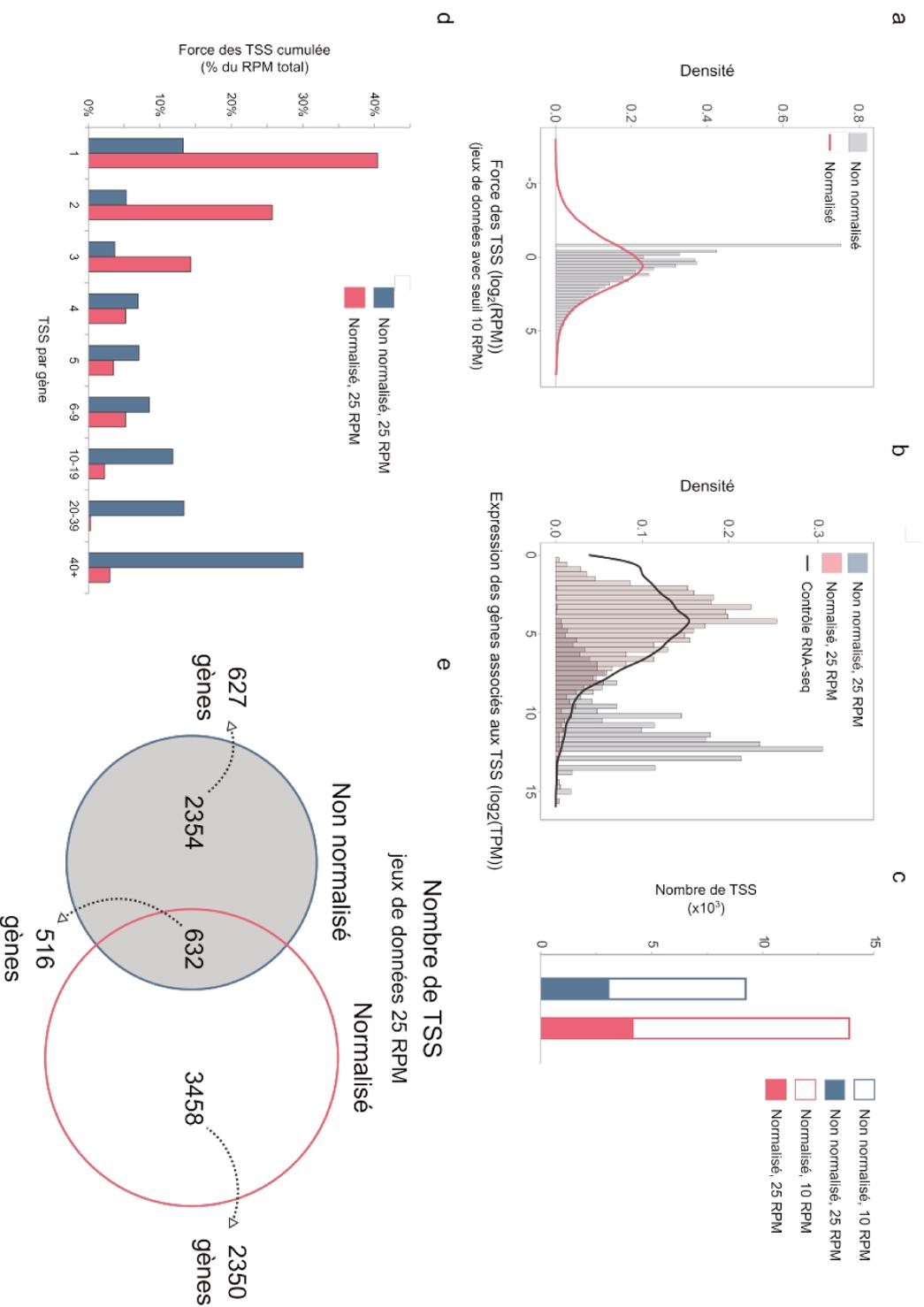


Figure 16 : Le eCAPP-seq améliore l'identification des TSS chez *C. beijerinckii* DSM 6423.

a. Distribution de l'expression des TSS avec/sans normalisation **b.** Distribution des expression RNA-seq de i) tous les gènes (ligne noire) ii) le sous-ensemble de gènes pour lesquels des TSS ont été trouvés dans les jeux de données normalisés (bleu) et non normalisés (rouge). **c.** Nombre de TSS trouvés dans les jeux de données normalisés et non normalisés. Les jeux de données à haute confiance sont obtenus avec une valeur seuil de 25 RPM (au lieu de 10). **d.** Nombre de TSS par gène (pour lesquels des TSS sont détectés). La valeur représentée correspond au nombre de lectures tombant dans chaque catégorie (en pourcentage de la somme totale des RPM). **e.** Diagramme de Venn montrant des TSS trouvés pour chaque jeu de données à haute confiance et le nombre de gènes associés.

Nous avons ensuite comparé pour chaque souche les jeux de données avec et sans normalisation (Figure 16 et Figure Supplémentaire 3). Le fait d'ajuster la force des TSS relativement à l'expression génétique locale a radicalement changé la distribution du jeu de données (Figure 16a). En comparant la distribution de l'expression des sous-ensembles de gènes comportant au moins un TSS, cela s'est traduit par un changement significatif. En effet, les TSS sont majoritairement détectés dans les jeux de données non normalisés pour des gènes très exprimés (> 1000 TPM), et cette distribution change radicalement après normalisation pour se rapprocher de la distribution représentative de l'ensemble des jeux de données RNA-seq (Figure 16b). Cela suggère que notre nouvelle stratégie détecte les TSS plus uniformément le long d'un génome bactérien qu'auparavant. De surcroît, la normalisation permet aussi d'obtenir un plus grand nombre de TSS (Figure 16c), ce qui est également avantageux.

Nous avons ensuite testé notre hypothèse selon laquelle, dans le premier jeu de données, les TSS avaient tendance à s'accumuler pour un petit nombre de gènes (Figure 16d). Comme attendu, sans normalisation, une proportion élevée des lectures (> 70%) contribue à la détection de TSS associés à des gènes qui en comportent plus de 4, avec un maximum de 244 TSS pour un seul gène (Figure 17). A l'inverse, la normalisation par l'expression réduit ce nombre à un petit peu moins de 15%, ce qui indique 1) qu'un nombre globalement plus élevé de gènes sont identifiés avec au moins un TSS après normalisation et 2) que l'étape de normalisation nettoie le jeu de données en éliminant une proportion importante de TSS secondaires qui pourraient être liés à une transcription permissive ou à un artefact lié à la technique de préparation des bibliothèques.

Enfin, nous avons également comparé les TSS avec/sans normalisation (Figure 16e). La plupart sont spécifiques de chaque jeu de données, ce qui souligne l'impact d'une étape additionnelle de normalisation sur les TSS identifiés.

Partie III : Analyse comparative des TSS dans trois *Clostridia* solvantogènes

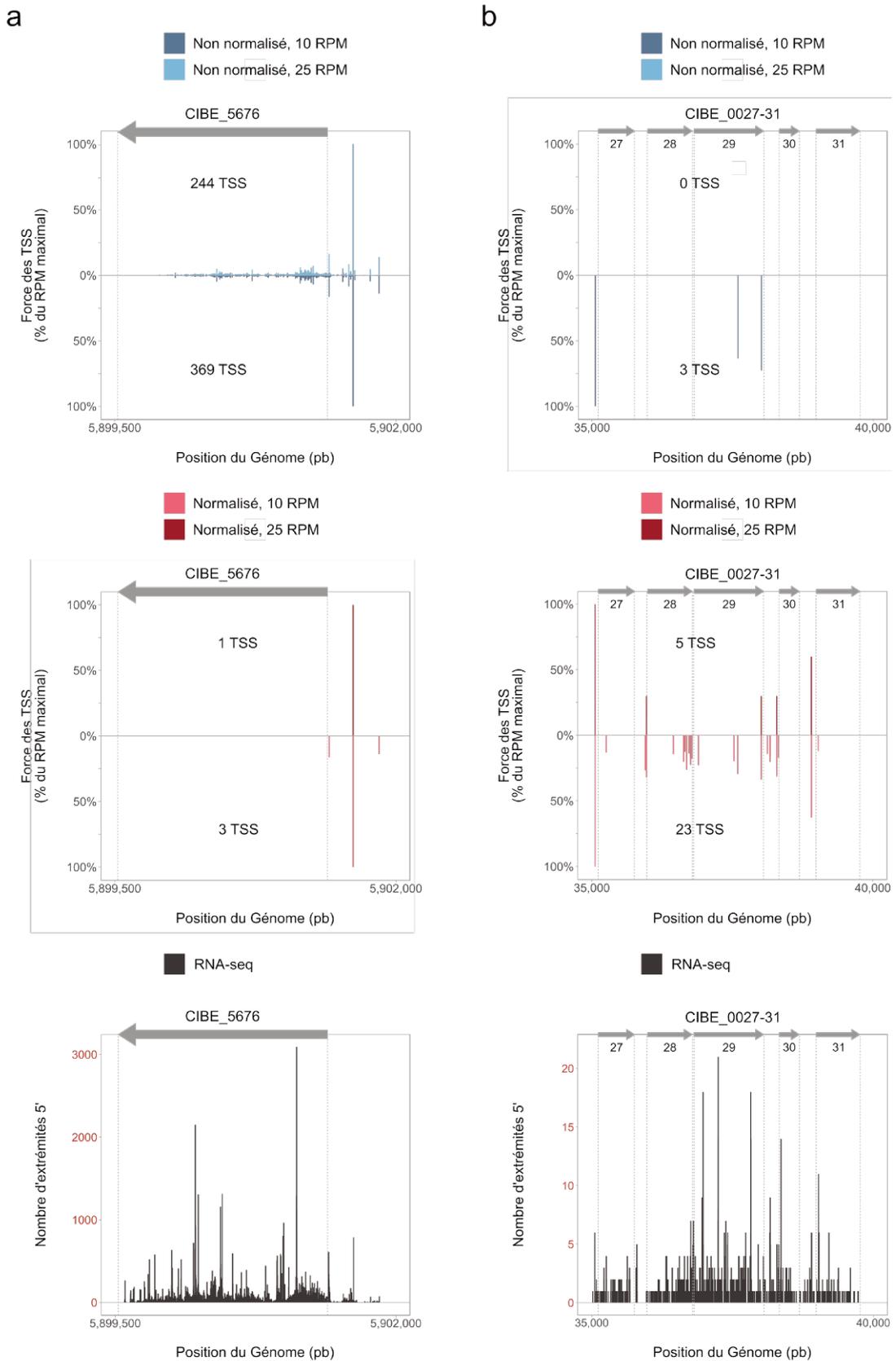
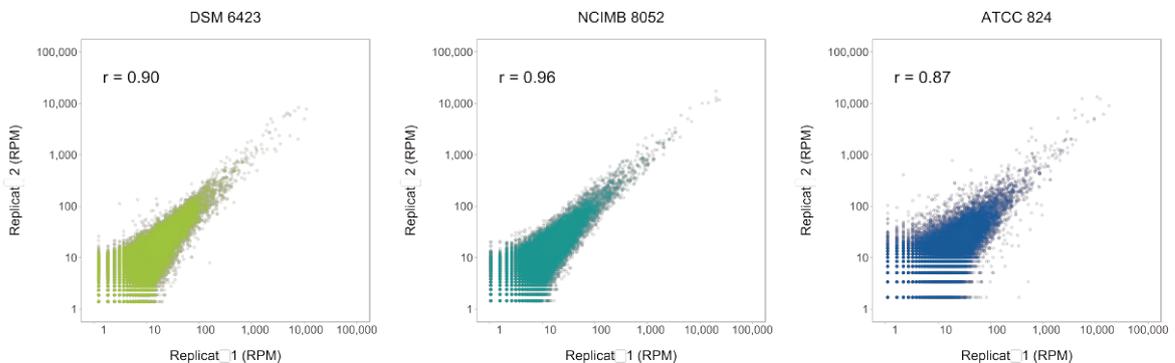


Figure 17 : Exemples de TSS retrouvés chez *C. beijerinckii* DSM 6423 pour **a. un gène fortement exprimé et b. un groupe de gènes faiblement exprimés.** Jeux de données RNA-seq (extrémités 5' représentées) en noir, normalisés en rouge, non normalisés en bleu.

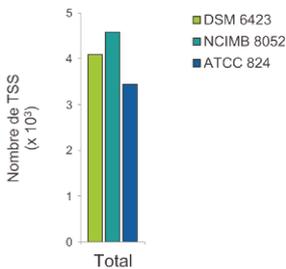
En résumé, la normalisation par l'expression apportée par notre approche eCAPP-seq permet une détection plus uniforme des TSS le long des génomes bactériens. Les jeux de données non normalisés ont tendance à identifier des TSS pour des gènes très exprimés ne représentant qu'une fraction des gènes exprimés. La normalisation améliore la qualité de la détection de manière indépendante de l'expression la détection des TSS et augmente également le nombre de gènes pour lesquels au moins un TSS est détecté.

5.3.2 Le eCAPP-seq identifie des milliers de TSS pour trois *Clostridia*

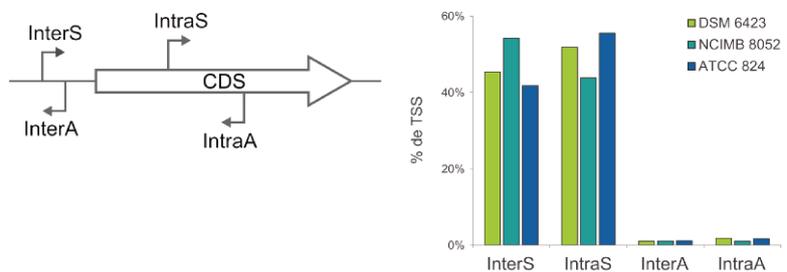
a



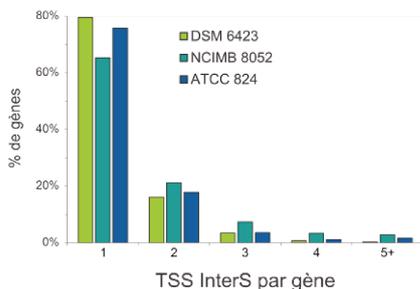
b



c



d



e

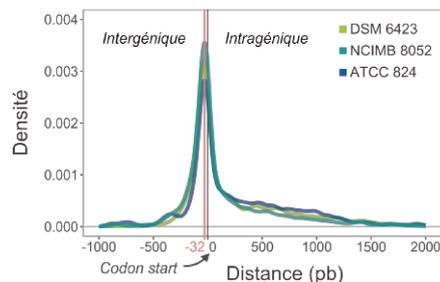


Figure 18 : **Caractéristiques transcriptomiques générales des jeux de données eCAPP-seq de *C. beijerinckii* DSM 6423, *C. beijerinckii* NCIMB 8052 et *C. acetobutylicum* ATCC 824.**

a. Analyse en duplicats des TSS trouvés pour chaque set de réplicats (avant l'étape de normalisation). **b.** Nombre de TSS retrouvés pour chaque souche, avec un seuil à 25 RPM. **c.** Classification des TSS en 4 catégories : intergénique sens (InterS) ; intragénique sens (IntraS), intergénique antisens (InterA), intragénique antisens (IntraA). **d.** Nombre de TSS InterS par gène pour chaque souche. Les valeurs correspondent au pourcentage de gènes pour lesquels au moins un TSS a été détecté. **e.** Pour les TSS sens, distribution des distances séparant les TSS du codon start.

Après avoir optimisé le pipeline d'identification des TSS, nous nous sommes concentrés sur les données obtenues pour chaque souche (Figure 18). Pour chaque duplicat, une corrélation significative est obtenue (corrélation de Pearson > 0,87, Figure 18a). Le processus d'analyse identifie *in fine* respectivement 4090, 4593 et 3441 TSS dans *C. beijerinckii* DSM 6423, *C. beijerinckii* NCIMB 8052 et *C. acetobutylicum* ATCC 824, avec le seuil à haute confiance (> 25 RPM normalisés, Figure 18b). En abaissant ce seuil à 10 RPM, le nombre de TSS augmente jusqu'à environ 13000 pour les deux souches de *C. beijerinckii* et plus de 9000 pour *C. acetobutylicum*. Toutefois, ces jeux de données n'ont pas été utilisés dans les analyses suivantes pour limiter le taux de découverte de faux positifs.

Les TSS ont ensuite été classifiés en 4 sous-catégories (Figure 18c) : InterS (TSS intergénique avec un gène dans la même orientation), InterA (TSS intergénique avec un gène dans l'orientation opposée), IntraS (TSS intragénique dans un gène avec la même orientation, IntraA (TSS intragénique dans un gène avec l'orientation opposée). Dans les trois microorganismes, la répartition des TSS est relativement similaire, avec la majorité des TSS identifiés en orientation sens (97-98%). Une forte proportion de TSS intragéniques est également observée (InterS : 40-55 % ; IntraS : 40-55%), mais qui n'est toutefois pas surprenante vis-à-vis d'autres études^{218,273}, qui suggèrent que cette abondance est due à un phénomène de régulation ou plus simplement à une transcription permissive. Toutefois, même si les proportions de TSS InterS et IntraS sont similaires, la plupart des lectures contribuent au signal InterS (entre 56 et 65% du nombre de lectures total, Figure Supplémentaire 4). Ces TSS intergéniques, canoniques, sont retrouvés en amont de 1468 gènes (DSM 6423, 23% des gènes), 1525 gènes (NCIMB 8052, 29%) et 976 (ATCC 824, 25%). Pour la plupart de ces gènes, nos données révèlent que, dans nos conditions expérimentales, leur transcription est contrôlée par un unique TSS (dans 65-80% des cas, selon la souche ; Figure 18d). Moins de 3% des TSS sont observés en orientation antisens, ce qui est cohérent avec les premiers résultats obtenus chez *C. phytofermantans* (Boutard et al.²¹⁸ ; Figure 18c). Nous observons toutefois également que ces TSS pourraient avoir d'importantes fonctions biologiques (

Figure Supplémentaire 5). En effet, des événements d'initiation de la transcription en orientation antisens sont par exemple détectés pour des gènes impliqués dans le contrôle de la transcription, de l'équilibre rédox et de l'assimilation

des sucres, ce qui suggère que la transcription antisens pourrait être un moyen de réguler l'expression des enzymes impliquées dans ces importants processus cellulaires *in vivo*.

La longueur des 5' UTR excède rarement 200 pb dans les trois souches (Figure 18e), sans corrélation entre la longueur du 5'UTR et de l'expression génétique (Figure Supplémentaire 6). La plupart des 5' UTR comportent entre 0 et 100 pb (avec un maximum de TSS commençant à 32 pb du codon *start*), avec $\approx 2-3\%$ des transcrits catégorisés comme *leaderless* (transcrits ne comportant pas de RBS, pour cette analyse tous les transcrits avec un 5' UTR < 6 pb), ce qui suggère que, même si elle est commune dans d'autres bactéries²⁷⁴, la transcription *leaderless* semble se produire à des niveaux relativement faibles dans les *Clostridia* solvantogènes.

5.3.3 Le eCAPP-seq révèle un contrôle différencié de la conversion de l'acétoacétyl-CoA entre *C. beijerinckii* et *C. acetobutylicum*

Si l'analyse des caractéristiques des jeux de données montre des tendances similaires pour les trois souches à l'échelle du transcriptome, les données révèlent également des différences intéressantes entre espèces, par exemple au niveau de l'opéron contenant les gènes impliqués dans la synthèse du butyryl-CoA. Les *Clostridia* solvantogènes ont un métabolisme atypique qui leur permet de convertir de multiples sources de carbone en un panel de composés à hautes valeurs ajoutées (dont notamment le butanol, l'éthanol, l'acétone, l'isopropanol, cf. sous-partie 1.3). Les voies métaboliques relatives sont toutes centrées autour de l'acétyl-CoA (Figure 19a), qui sert de brique élémentaire pour la synthèse de ces métabolites, en étant notamment condensé en acétoacétyl-CoA par la thiolase (Thl) avant d'être pris en charge dans la voie dite « BCS » (pour *butyryl-CoA synthesis*, cf. sous-partie 1.3.3). Celle-ci implique 5 enzymes (Hbd, Crt, Bcd, EtfA, EtfB) dont les gènes forment un opéron chez *C. acetobutylicum*¹⁴⁸.

Bien que plusieurs analyses RNA-seq aient suggéré que cette organisation était probablement conservée chez *C. beijerinckii*^{52,175,187}, notre analyse comparative des TSS indique que ce n'est pas le cas (Figure 19b). En effet, si le TSS en amont de *crt* est bien conservé, un second TSS, fortement utilisé, est observé en amont de *hbd*, le dernier gène de l'opéron, qui code la première enzyme de la voie métabolique.

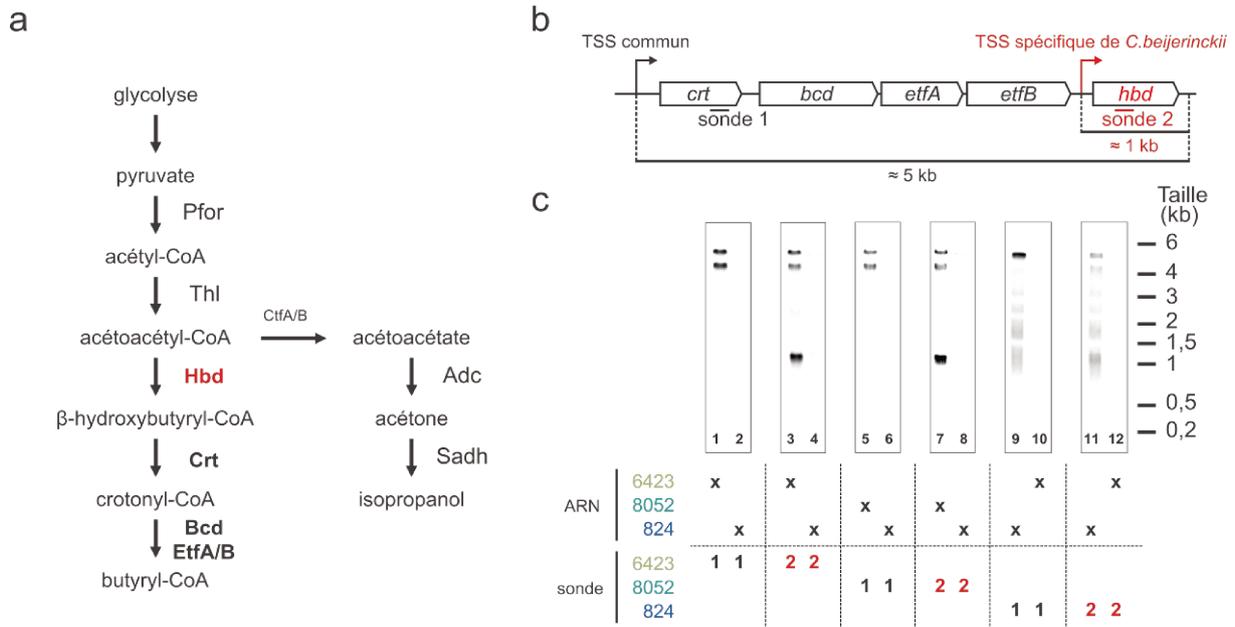


Figure 19 : **Analyse Northern Blot des structures transcriptionnelle des gènes de synthèse du butyryl-CoA (BCS).**

a. Voies métaboliques orientant le métabolisme solvantogène vers la formation d'acétone/isopropanol ou vers la synthèse de butanol. **b.** TSS retrouvés pour chaque espèce le long des gènes BCS. Les sondes utilisées pour les Northern Blots sont montrées pour les gènes *crt* et *hbd*. **c.** Northern Blot anti-*crt* et anti-*hbd* réalisés pour chaque souche.

Pour conforter ces données *in silico*, une analyse de taille et le nombre de transcrit de cet opéron a été menée par *northern blot* (Figure 19c). Pour chaque souche, des sondes d'ADN simple brin radiomarquées ciblant *crt* ou *hbd* ont été hybridées à des ARN bactériens transférés sur membrane de nitrocellulose. Les résultats montrent clairement un transcrit *hbd* (≈ 1 kb) spécifique aux souches de *C. beijerinckii*. Les Northern Blot anti-*hbd* révèlent également un signal de haut poids moléculaire (entre 4 et 6 kb) également observé sur les *northern blots* anti-*crt*, ce qui suggère que *hbd* est également transcrit au sein de l'opéron BCS originel. De façon surprenante, le signal de haut poids moléculaire est constitué de deux transcrits distincts, détectés dans toutes les expériences sur les souches de *C. beijerinckii*. Ces deux transcrits contiennent donc à la fois les sites de liaison des sondes anti-*hbd* et anti-*crt*, ce qui suggère peut-être 1) l'existence d'un TSS secondaire non détecté par notre approche, 2) une forme de maturation transcriptionnelle qui raccourcirait une partie des transcrits BCS ou 3) plus simplement un biais expérimental quelconque. Nous n'avons toutefois pas été en mesure de vérifier ces hypothèses.

En résumé, une approche expérimentale par *northern blot* a permis de confirmer une observation faite à l'aide de l'analyse eCAPP-seq en confirmant la présence d'un TSS alternatif permettant la transcription indépendante de *hbd* chez *C.*

beijerinckii. Nos résultats indiquent également que, comme chez *C. acetobutylicum*, *hbd* est également transcrit à partir du promoteur de *crt*, ce qui suggère une régulation fine de l'expression de *hbd* dans cette espèce.

5.4 DISCUSSION

Dans cette partie de la thèse, une démarche plus globale d'analyse des mécanismes de régulation transcriptionnelle a été entreprise au travers de la détermination des sites d'initiation de la transcription à l'échelle du génome dans trois souches de *Clostridia* solvantogènes pour lesquelles ce type de données n'avait jusqu'alors pas été documenté. Par ailleurs, l'outil développé précédemment et utilisé pour étudier *C. phytofermentans*²¹⁸ a été amélioré via une étape de normalisation des TSS dans une tentative de mieux refléter la réalité biologique étudiée.

Ce protocole a été mis au point après que notre tentative initiale, basée sur une préparation de librairie de type *Capp-Switch seq*, ait révélé un biais probable lié à l'expression des gènes dans les données. Nous avons donc réalisé un RNA-seq avec les mêmes ARN pour vérifier et confirmer cette hypothèse, puis utilisé les valeurs d'expression issues du RNA-seq pour normaliser les données *Capp-Switch seq*. Cette étape additionnelle permet ainsi d'améliorer les résultats tant du point de vue global (indépendance vis-à-vis de l'expression, nombre de TSS par gène fortement réduit) que d'un point de vue local (cf. les exemples de la Figure 17), ce qui souligne l'importance d'utiliser une méthode de normalisation pour traiter les données *Capp-Switch*.

Cette amélioration constatée pourrait s'expliquer de façon biologique et/ou technique. Pour les gènes très exprimés, la présence d'un promoteur fort pourrait provoquer, à faible fréquence, une initiation de la transcription à des sites plus ou moins proches du TSS primaire. Une autre explication serait celle d'une purification imparfaite : malgré la forte affinité biotine-streptavidine, il est possible qu'une quantité variable d'ARN non biotinylé soit co-purifiée. Pour les gènes très exprimés, cette quantité est plus importante, ce qui se traduit par une augmentation des signaux considérés comme des TSS par l'algorithme de détection.

Ces deux hypothèses ne sont pas mutuellement exclusives. Mais nous avons estimé que dans un cas comme dans l'autre une méthode permettant de prendre en

compte l'expression des gènes dans les algorithmes de détection des TSS serait bénéfique. En effet, dans le premier cas, les TSS liés à une transcription permissive sont a priori peu intéressants car probablement peu fonctionnels. Ils ne représenteraient en effet qu'une minorité des transcrits, et leur rôle, s'ils en ont un, serait difficile à déterminer. Dans le second cas, une partie des TSS identifiés ne reflèterait pas de réalité biologique, ce qui est d'autant plus problématique. Enfin, il est également probable que les TSS des gènes peu voire moyennement exprimés ne soient pas correctement détectés.

Ces résultats soulignent donc la nécessité d'un second séquençage pour traiter des données de type *Capp-Switch seq*. Dans leurs travaux originaux²¹⁷, Ettwiller et collègues s'étaient déjà interrogés sur l'intérêt d'un séquençage contrôle, similaire à ce qui était réalisé pour une analyse dRNA-seq, et en avaient conclu que ce contrôle n'était pas nécessaire dans leurs conditions, puisqu'il ne permettait d'éliminer qu'une faible minorité de TSS.

Plusieurs hypothèses peuvent expliquer l'opposition entre nos conclusions respectives. Tout d'abord, l'utilisation faite de notre librairie contrôle est sensiblement différente de ce qui a été réalisé par le laboratoire de Ettwiller. En effet, les données contrôles sont utilisées dans le cadre du eCAPP-seq à une échelle relativement large, celle du transcrit, puisque l'ensemble des lectures alignées sur un gène donné sont prises en compte pour déterminer son niveau d'expression en RNA-seq. A l'inverse, le contrôle utilisé dans le cadre du développement du *Cappable-seq* n'est utilisé que de manière très locale, puisque pour chaque TSS, c'est un enrichissement local qui est calculé par rapport au nombre de lectures commençant à la même position génomique dans la librairie contrôle. Cette différence est peut-être à mettre en relation avec une différence technique significative liée au protocole de préparation des librairies. En effet, les protocoles de préparation de librairie traditionnels pour séquençage Illumina ont tendance à générer des données avec une couverture inégale le long des gènes, et, en particulier, moins fournie vers les extrémités 5' des transcrits (c'est-à-dire les positions les plus proches des TSS attendus)^{219,275}. Considérant que la purification liée à l'enzyme VCE enrichit à l'inverse les extrémités 5', il est probable que la comparaison des deux jeux de données soit biaisée à l'échelle locale. Alternativement, les différences observées entre *Cappable-seq* et *Capp-Switch*

seq par rapport à l'utilisation d'une librairie contrôle pourraient venir d'un biais technique lié aux différences en termes de protocole expérimental.

Après cette étape d'amélioration du *Capp-Switch seq*, nous avons cherché à valider certains des résultats obtenus expérimentalement. Pour ce faire, nous nous sommes intéressés à l'opéron de synthèse du butyryl-CoA, qui présente une différence notable entre les espèces *C. beijerinckii* et *C. acetobutylicum*. L'opéron BCS, identifié chez *C. acetobutylicum* il y a presque 25 ans¹⁴⁸, est un opéron crucial du métabolisme solvantogène puisqu'il permet la production de molécules à 4 carbones. Les données non normalisées pour cet opéron, très exprimé dans les trois souches, indiquent des dizaines de TSS différents. Après normalisation, un nombre très limité subsiste. Si l'analyse eCAPP-seq valide l'organisation en opéron chez *C. acetobutylicum* (1 seul TSS), elle suggère à l'inverse une structure opéronique / sous-opéronique chez *C. beijerinckii* (2 TSS), avec une régulation distincte pour le gène codant la première enzyme de la voie métabolique, *hbd*. Cette structure a pu être vérifiée par Northern Blot, qui malgré la présence d'un transcrit énigmatique, démontre que le nombre de TSS primaires est très limité. Pour *C. beijerinckii*, l'implication est importante puisqu'elle signifie que la transcription de *hbd* est contrôlée indépendamment, ce qui pourrait être important pour régler finement la conversion de l'acétoacétyl-CoA en butanol ou en isopropanol.

Les TSS identifiés par notre approche eCAPP-seq partagent des propriétés générales à l'échelle du transcriptome (proportions d'InterS, InterA, IntraS et IntraA ; longueur des 5'UTR). Des analyses sont encore en cours pour développer cet aspect de l'étude, en particulier via la détermination de la conservation des sites d'initiation de la transcription inter-espèces. Une telle étude serait probablement très informative puisqu'elle permettrait de comparer des souches physiologiquement proches, avec des traits phénotypiques très conservés (notamment le métabolisme solvantogène et la sporulation) mais qui diffèrent génétiquement⁴⁰. La taille de chaque génome est un exemple révélateur de cette hétérogénéité (environ 4000, 5000 et 6000 pour *C. acetobutylicum* ATCC 824, *C. beijerinckii* NCIMB 8052 et *C. beijerinckii* DSM 6423). Cela se traduit par des différences métaboliques parfois importantes : *C. acetobutylicum*, par exemple, gère l'équilibre rédox de façon différente de *C. beijerinckii*²⁷⁶. Même entre les deux souches de *C. beijerinckii*, le métabolisme primaire diffère puisque NCIMB 8052 ne peut produire d'isopropanol, au contraire de DSM 6423

qui utilise cette voie pour régénérer du pouvoir oxydant (NADP⁺), ce qui suggère également des capacités différentes permettant de gérer la balance redox cellulaire. De plus, des recherches récentes ont montré que la souche DSM 6423 possède 3 éléments génétiques mobiles (2 plasmides, 1 phage linéaire), à comparer à la souche NCIMB 8052 qui n'en comporte aucun⁵², ce qui suggère un chemin évolutif particulier. Malgré ces différences, ces 3 souches ont conservé une capacité commune à produire des solvants à 3 et 4 atomes de carbones. L'étude de la conservation des réseaux génétiques de régulation communs ou propres à ces souches peut potentiellement apporter des informations nouvelles sur la régulation des voies solvantogènes, qui ne sont pas encore entièrement comprises malgré le fait qu'elles soient la principale raison de l'utilisation de ces microorganismes²¹¹.

Afin d'étudier la conservation des TSS entre ces souches, une méthode permettant de comparer les séquences entourant les TSS de gènes homologues est en cours de développement en collaboration avec le département de Mathématiques Appliquées d'IFPEN. Intuitivement, une faible dérive nucléotidique due à une pression de sélection forte au cours de l'évolution est attendue pour les promoteurs (c'est-à-dire la séquence immédiatement en amont d'un TSS identifié) de gènes ayant un impact fonctionnel important, et qui devraient donc être conservés. A l'inverse, les TSS non conservés impliquent une pression sélective plus faible, pouvant être liée à une fonction accessoire ou peu importante, ou encore à un chemin évolutif différent entre les microorganismes étudiés. Cette approche est similaire à celle de Shao *et al.* qui, en comparant la conservation des TSS pour de multiples espèces du genre *Shewanella*, ont par exemple identifié de nombreux TSS non canoniques fortement conservés (pour environ 30% des intragéniques sens, notamment), avec, probablement, un rôle biologique important vis-à-vis des gènes concernés²⁷³.

En conclusion, bien que l'analyse transcriptomique globale permettant de comparer les TSS de chaque souche ne soit pas terminée à l'heure de la rédaction de ce manuscrit, les résultats obtenus pourront être valorisés, en particulier dans la mesure où très peu de TSS étaient connus pour les souches solvantogènes majeures avant cette étude. La régulation différentielle de l'expression du gène *hbd* est un nouvel exemple des différences importantes qui existent entre *C. acetobutylicum* et *C. beijerinckii*. L'optimisation du protocole de détection des TSS pourrait être utilisée par la communauté scientifique pour d'autres souches bactériennes, dans la mesure où

l'implémentation dans le protocole *Capp-Switch seq*, déjà établie, est directe et relativement simple.

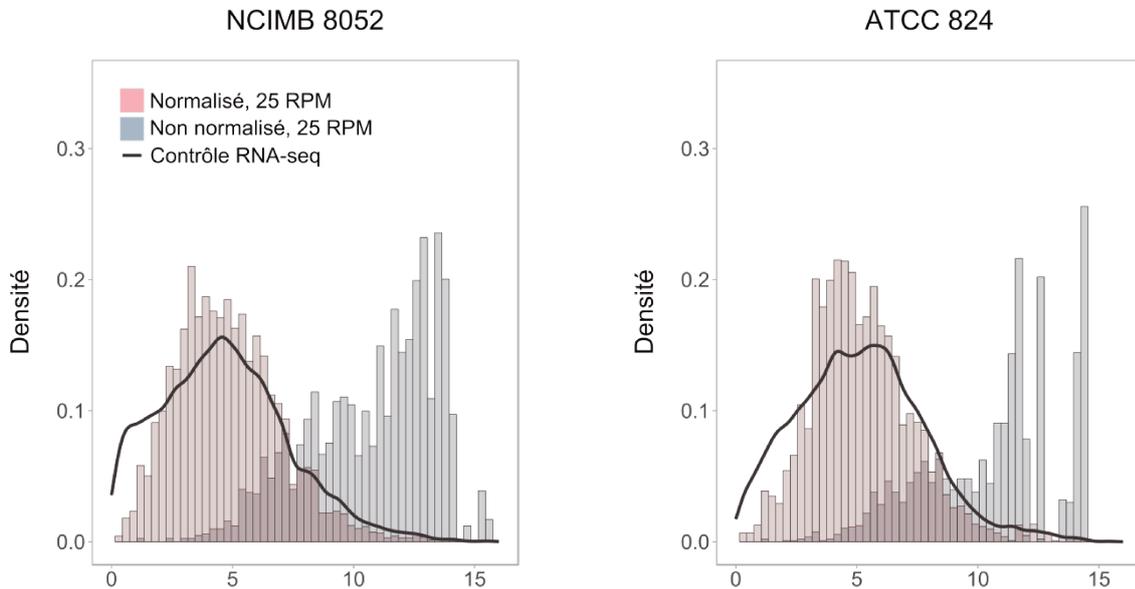
5.5 FIGURES SUPPLEMENTAIRES

Réplicat	Lectures totales	Alignées	Alignées (%)
eCAPP-seq			
6423 A	2 058 616	2 039 006	99,0%
6423 B	2 596 025	2 556 181	98,5%
8052 A	2 310 835	2 261 155	97,9%
8052 B	2 527 066	2 484 830	98,3%
824 A	2 450 525	2 411 469	98,4%
824 B	738 486	732 417	99,2%
RNA-seq			
6423 A	3 008 134	2 861 724	95,1%
6423 B	2 327 162	2 244 603	96,5%
8052 A	2 954 602	2 945 853	99,7%
8052 B	3 059 460	2 848 880	93,1%
824 B	2 559 372	2 466 919	96,4%

Figure Supplémentaire 1 : **Tableau récapitulatif des données issues du séquençage eCAPP-seq et RNA-seq.**

Souche	Gène	Amorce sens	Amorce antisens
Etape 1 : amplification des locus cibles			
DSM 6423	<i>crt</i>	TCCTGGTCCAACACTATTCTTGCT	ACTGGTGCTGGTGAGAAATCTT
DSM 6423	<i>hbd</i>	TCCAGCTCCAGTAATGAAGCTT	ATTTTCTTCCAAGCCATCCAGC
NCIMB 8052	<i>crt</i>	GCAAGTTTGAAGGAGGTCC	GTTTCCTAAAATACCAAATCTTTACC
NCIMB 8052	<i>hbd</i>	TGAAGCTGCAATCGAAAACATGA	ACAACGAATCCTGGAGCTTCTG
ATCC 824	<i>crt</i>	AGCTGTCATTGCATCCTTTTGA	GTCAACCAGAAGTAGGTCTCGG
ATCC 824	<i>hbd</i>	GAATTTCCGGAACAGTTGACCT	CAACAAATCCTGGTGCTTCTGC
Etape 2 : génération des sondes simple brin			
DSM 6423	<i>crt</i>	X	AGTCCTGCTTCTGGTTGACC
DSM 6423	<i>hbd</i>	X	CTCTGTACTTGTTATCACCTGTTTC
NCIMB 8052	<i>crt</i>	X	CAGCAACAAAAGATTTCTCACC
NCIMB 8052	<i>hbd</i>	X	CTTCTGCAACTTCTACTGGTTC
ATCC 824	<i>crt</i>	X	CCAAATGCTTCTGATTCAAATGC
ATCC 824	<i>hbd</i>	X	CCTCTTATTACCTCTACAAGCTTC

Figure Supplémentaire 2 : Tableau récapitulatif des amorces utilisées pour obtenir les sondes du *northern blot*.



Expression des gènes associés aux TSS ($\log_2(\text{TPM})$) Expression des gènes associés aux TSS ($\log_2(\text{TPM})$)
 Figure Supplémentaire 3 : **Distribution des expressions RNA-seq pour les gènes liés aux TSS identifiés.**
 Sont représentés en rouge et bleu les jeux de données normalisés et non normalisés (sous-ensemble de gènes).
 La ligne noire correspond à la distribution RNA-seq totale (tous les gènes). A gauche : souche NCIMB 8052. A droite : souche ATCC 824.

Classe	DSM 6423		NCIMB 8052		ATCC 824	
	TSS	TSS (%)	TSS	TSS (%)	TSS	TSS (%)
Total	4090	100%	4583	100%	3441	100%
InterS	1855	45,4%	2482	54,2%	1436	41,7%
IntraS	2120	51,8%	2006	43,8%	1910	55,5%
InterA	41	1,0%	48	1,0%	38	1,1%
IntraA	74	1,8%	47	1,0%	57	1,7%

Figure Supplémentaire 4 : **Tableau récapitulatif du nombre et de la classification des TSS dans chaque souche.**

Les jeux de données 25 RPM sont utilisés.

Souche	Locus tag	Classe	Annotation du gène
DSM 6423	CIBE_0497	IntraA	Arsenical pump membrane protein
DSM 6423	CIBE_0903	InterA	FMN reductase
DSM 6423	CIBE_1015	IntraA	transposase
DSM 6423	CIBE_1015	IntraA	transposase
DSM 6423	CIBE_1455	IntraA	rny
DSM 6423	CIBE_2133	IntraA	dapH
DSM 6423	CIBE_2434	IntraA	nifD
DSM 6423	CIBE_3408	InterA	kipA
DSM 6423	CIBE_3638	InterA	transposase
DSM 6423	CIBE_4230	InterA	4Fe-4S ferredoxin, iron-sulfur binding domain protein
DSM 6423	CIBE_4290	InterA	transposase
DSM 6423	CIBE_4290	InterA	transposase
DSM 6423	CIBE_4294	IntraA	4Fe-4S ferredoxin, iron-sulfur binding domain protein
DSM 6423	CIBE_4301	InterA	Nitroreductase
DSM 6423	CIBE_4636	InterA	Phosphoenolpyruvate synthase
DSM 6423	CIBE_6118	IntraA	transposase
DSM 6423	CIBE_p0007	InterA	putative resolvase
NCIMB 8052	CBEI_RS02635	IntraA	2-dehydro-3-deoxy-phosphogluconate aldolase
NCIMB 8052	CBEI_RS06420	IntraA	rny
NCIMB 8052	CBEI_RS07045	IntraA	XRE family transcriptional regulator
NCIMB 8052	CBEI_RS07045	IntraA	XRE family transcriptional regulator
NCIMB 8052	CBEI_RS07045	IntraA	XRE family transcriptional regulator
NCIMB 8052	CBEI_RS09320	IntraA	dapD
NCIMB 8052	CBEI_RS10410	IntraA	nitrogenase component I subunit alpha
NCIMB 8052	CBEI_RS10985	IntraA	sensor histidine kinase
NCIMB 8052	CBEI_RS10985	IntraA	sensor histidine kinase
NCIMB 8052	CBEI_RS16660	IntraA	hydF

Partie III : Analyse comparative des TSS dans trois Clostridia solvantogènes

NCIMB 8052	CBEI_RS18415	InterA	nitroreductase
NCIMB 8052	CBEI_RS18755	IntraA	NAD(P)-dependent oxidoreductase
NCIMB 8052	CBEI_RS18760	IntraA	transcriptional regulator
NCIMB 8052	CBEI_RS18975	IntraA	FAD:protein FMN transferase
NCIMB 8052	CBEI_RS19550	IntraA	two-component system sensor histidine kinase AtoS
NCIMB 8052	CBEI_RS20560	IntraA	hybrid sensor histidine kinase/response regulator
NCIMB 8052	CBEI_RS22195	InterA	ABC-2 transporter permease
NCIMB 8052	CBEI_RS25195	InterA	MFS transporter
NCIMB 8052	CBEI_RS25195	InterA	MFS transporter
NCIMB 8052	CBEI_RS25520	InterA	sigma-70 family RNA polymerase sigma factor
NCIMB 8052	CBEI_RS26125	IntraA	chromosome partitioning protein ParB
NCIMB 8052	CBEI_RS26130	IntraA	ParA family protein
NCIMB 8052	CBEI_RS26130	IntraA	ParA family protein
ATCC 824	CA_C0204	IntraA	sortase
ATCC 824	CA_C0385	IntraA	beta-glucosidase
ATCC 824	CA_C0393	IntraA	KDG operon repressor
ATCC 824	CA_C0533	InterA	maltose-6'-phosphate glucosidase
ATCC 824	CA_C0955	IntraA	ATP-dependent Zn protease
ATCC 824	CA_C1558	InterA	ABC transporter ATPase
ATCC 824	CA_C1559	InterA	transcriptional regulator
ATCC 824	CA_C1568	IntraA	diguanylate cyclase/phosphodiesterase
ATCC 824	CA_C1872	IntraA	replication initiation protein REP
ATCC 824	CA_C1947	InterA	phage transcriptional regulator
ATCC 824	CA_C2438	IntraA	phosphatase
ATCC 824	CA_C2659	IntraA	stress-responsive transcriptional regulator PspC
ATCC 824	CA_C2955	InterA	lactose phosphotransferase system repressor lacR
ATCC 824	CA_C3331	InterA	alanine racemase
ATCC 824	CA_C3466	IntraA	LysR family transcriptional regulator
ATCC 824	CA_C3616	InterA	ABC transporter ATPase
ATCC 824	CA_C3687	InterA	AcrR family transcriptional regulator
ATCC 824	CA_C3738	IntraA	ribonuclease P
ATCC 824	CA_Ct026	IntraA	tRNA-Pro
ATCC 824	CA_P0013	IntraA	FAD dependent dehydrogenase
ATCC 824	CA_P0017	InterA	spore germination protein, GRKB
ATCC 824	CA_P0043	IntraA	membrane flavodoxin oxidoreductase

Figure Supplémentaire 5 : **Exemples de gènes présentant un TSS antisens dans les trois souches.**
Les jeux de données 25 RPM sont utilisés.

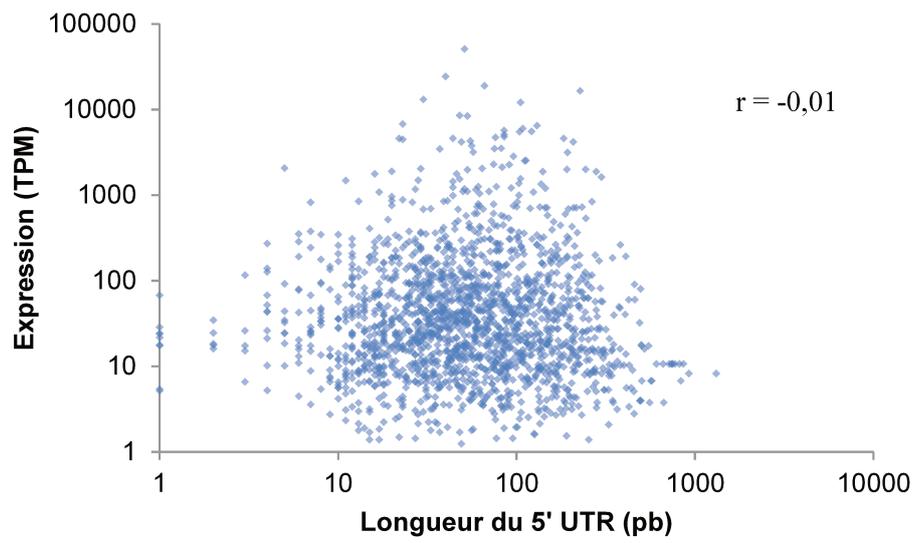


Figure Supplémentaire 6 : Nuage de points représentant, pour la souche DSM 6423, la longueur du 5'UTR d'un gène par rapport à son niveau d'expression en RNA-seq. Les jeux de données 25 RPM sont utilisés.

6 PARTIE IV : METABOLISME SUR SUBSTRAT 1G, APPROCHES FERMENTAIRE ET MULTI-OMIQUE

L'identification des TSS, tout comme les analyses transcriptomiques publiées par Máté de Gérando et collègues⁵², a été réalisé pour des cultures réalisées dans un milieu modèle, avec du glucose comme seule source de carbone, au cours d'une fermentation de type *batch*. Si les informations obtenues par ces méthodes sont importantes pour la compréhension de la souche DSM 6423, elles ne permettent pas de modéliser correctement son comportement dans des conditions proches de l'industrie. Dans cette partie de la thèse, nous avons mis à profit l'arsenal d'outils d'analyses omiques récemment complété par l'approche eCAPP-seq pour comprendre comment certains paramètres d'une fermentation industrielle utilisant un procédé continu et un substrat 1G, affectent au niveau moléculaire et métabolique le fonctionnement de cette souche.

6.1 INTRODUCTION

Dans une optique de production de propylène biosourcé, la caractérisation de la souche *C. beijerinckii* DSM 6423 permettrait de mieux comprendre le métabolisme de ce microorganisme naturellement capable de produire de l'isopropanol, un alcool facilement transformable en propylène.

Malgré cet intérêt, peu d'études se sont intéressées à cette souche jusqu'à présent, principalement à cause de ses performances en termes de production de solvants (titre maximal faible). Avec la publication récente de son génome⁵² et le développement des premiers outils d'édition génomique au cours de cette thèse, une optimisation par ingénierie métabolique est désormais envisageable, afin par exemple d'améliorer ses capacités de production d'isopropanol.

Ce travail d'ingénierie métabolique est d'autant plus facilité que la quantité de données biologiques caractérisant la souche sauvage est désormais importante. Si les premiers transcriptomes obtenus précédemment⁵² constituent une étape cruciale dans

la compréhension du microorganisme, les conditions expérimentales dans lesquelles ils ont été obtenus sont très éloignées de celles qu'on emploierait dans l'industrie. En particulier, le milieu de culture et le type de fermentation sont deux variables importantes pour la viabilité économique du procédé envisagé.

Pour la fermentation éthanolique à grande échelle, les principaux substrats industriels utilisés à l'heure actuelle sont des substrats de première génération (1G, correspondant aux éléments de réserves des plantes) basés sur la canne à sucre et le maïs. Historiquement, la fermentation ABE industrielle a également utilisé ces sources carbonées³². Les amidons de maïs et de blé étaient en effet très utilisés au début du XX^{ème} siècle, et leur utilisation fut, jusqu'aux années 50-60, progressivement supplantée par des mélasses (canne, betterave), moins onéreuses. Ces dernières, riches en saccharose, ont favorisé l'utilisation d'espèces saccharolytiques comme *C. beijerinckii* au détriment de souches de *C. acetobutylicum*¹²¹. Pour *C. beijerinckii* DSM 6423, l'utilisation de certains substrats 1G a été expérimentée^{50,277}, sans aller toutefois plus loin que l'analyse des performances fermentaires.

Indépendamment du substrat, un procédé continu serait le procédé le plus approprié, économiquement parlant, pour une fermentation industrielle A/IBE moderne⁷². Ce type de fermentation revêt plusieurs avantages, et permet, entre autres, de limiter la toxicité des métabolites en diluant en permanence le milieu de culture. Chez *C. beijerinckii* DSM 6423, plusieurs études rapportent, sur substrats carbonés synthétiques, des productivités en mixture A/IBE comprises entre $\approx 0,5$ et $\approx 5,5$ g.L.h⁻¹ selon la mise en œuvre^{48,49,51} (augmentations notables si immobilisation des cellules dans le réacteur, soutirage continu des solvants par *gas stripping*). Similairement à l'étude de substrats industriels, ces travaux se concentrent principalement sur les performances fermentaires.

En effet, le microorganisme étant peu connu, aucune étude n'a pour l'instant cherché à coupler fermentations en conditions pseudo-industrielles et analyses métaboliques basées sur des techniques à haut-débit modernes, de manière à identifier les facteurs moléculaires d'importance pour la mise en œuvre de tels procédés. Dans cette thèse, de premières expérimentations ont donc été réalisées pour 1) mettre au point des conditions fermentaires plus proches de celles de l'industrie et 2) collecter et analyser des données transcriptomiques, protéomiques et métaboliques dans ces conditions.

Ces travaux se sont initialement inspirés des conditions expérimentales publiées par le groupe du Pr. Soucaille¹⁷⁴. La fermentation continue peut en effet également être utilisée comme un outil de recherche fondamentale pour l'étude des solvantogènes, en permettant une séparation stable des phases métaboliques (acidogène, solvantogène, alcoologène) chez *C. acetobutylicum*. Après des modifications substantielles de ces conditions opératoires, des phases acidogènes et solvantogènes stables ont été obtenues chez *C. beijerinckii* DSM 6423 en fermentation continue, ce qui a permis de conduire des fermentations sur substrat modèle (glucose, saccharose) et 1G (mélasse de betterave, riche en saccharose).

Dans un deuxième temps, la mise en œuvre de ces fermentations a été couplée à une analyse multi-paramètres : métabolites principaux, transcriptomes (RNA-seq) et protéomes (LC-MS/MS). Pour certaines de ces conditions, le eCapp-seq, développé dans la partie précédente, a été appliqué de manière à caractériser la dynamique d'utilisation des TSS entre les différentes conditions explorées.

6.2 MATERIEL ET METHODES

6.2.1 Milieux de culture

6.2.1.1 Milieux modèles

Le milieu utilisé dans les tests initiaux pour la séparation des phases acidogènes et solvantogènes est un milieu minimum⁸⁰, contenant par litre : KH₂PO₄ 0,1 g, KCl 0,65 g, MgSO₄, 7 H₂O 0,2 g, FeSO₄, 7 H₂O 28 mg, acide para-aminobenzoïque 8 mg, NH₄Cl 1,5 g, biotine 0,04 mg, antimousse (Biospumex 36K, Pmc Ouvrie) 50 mg, glucose ou saccharose 260 ou 130 mmoles.

Le milieu de référence employé pour les autres expérimentations est un milieu riche, adapté d'une publication de Gapes *et al.*⁸¹, qui contient par litre : extrait de levure 2,5 g, KH₂PO₄ 1 g, K₂HPO₄ 0,6 g, MgSO₄, 7 H₂O 1 g, FeSO₄, 7 H₂O, 6,6 mg, acide para-aminobenzoïque 0,1 g, acétate d'ammonium 0 ou 2,9 g, antimousse (Biospumex 36K, Pmc Ouvrie) 50 mg, glucose ou saccharose 260 ou 130 mmoles.

Le milieu est ensuite tamponné à pH 3,5 si utilisé comme alimentation d'une fermentation continue, avant d'être stérilisé par filtration.

6.2.1.2 Incorporation et prétraitement de la mélasse de betterave

L'incorporation de substrat industriel est réalisée par substitution du glucose (260 mM) dans le milieu par une quantité équivalente de carbone mesuré dans la mélasse de betterave (dilution à 130 mM de saccharose, soit 9,95% massique). La mélasse doit par ailleurs être pré-traitée avant d'être utilisable en fermentation continue. Une première dilution à 260 mM en saccharose est réalisée, et le milieu est centrifugé deux fois 30 min à 17000 g à 30°C, en ne gardant à chaque fois que le surnageant. La mélasse est alors supplémentée avec le reste du milieu pour atteindre 130 mM en saccharose, avant d'être filtrée progressivement (diamètre des pores : 8-12 µm, puis 2-3 µm, puis 1,6 µm). La dernière filtration, à 0,2 µm, permet de stériliser le milieu.

6.2.2 Fermentations

6.2.2.1 En fioles serties

Pour chaque fermentation, 36 mL de milieu de référence ou industriel sont réduits au moins 24h en hotte anaérobie à 34°C. L'ensemencement est réalisé avec 4 mL d'une préculture réalisée sur le milieu de référence pendant 16 à 24h sans agitation en hotte anaérobie.

Les fioles, une fois serties, sont placées dans un incubateur à 34°C, sans agitation. Après 5h de culture, un système de valve anti-retour (*Evaporation control valves*, Vap-Lock) est installé pour prévenir la surpression du gaz issu de la fermentation. Une agitation magnétique suffisante pour remettre les solutions hétérogènes en suspension (300 rpm) est alors employée, et la culture est incubée dans ces conditions pendant 4 jours.

6.2.2.2 En bioréacteur

Les fermentations en bioréacteur sont réalisées dans un système *Sartorius Biostat Q* pilotant un à deux bioréacteurs de 500 mL. Le pH est contrôlé par ajout de base (NH₄ 5,5 M) et/ou d'acide (H₂SO₄ 3 M). Le système d'alimentation/soustraction de milieu pour la fermentation continue est piloté par une unité de pompage Scilog

Chemtech (Parker), qui peut transférer à un débit massique continu le milieu de fermentation dans le/les bioréacteur(s). Le soutirage du même volume de milieu fermenté est opérée par une sonde de niveau, branchée sur la même pompe.

Après montage, les bioréacteurs sont remplis avec 180 mL d'eau et stérilisés par autoclavage. 180 mL de milieu Gapes concentré 2X (glucose ou saccharose) stérilisés séparément sont ensuite ajoutés. Avant inoculation, le milieu est réduit pendant une nuit par injection d'azote, et équilibré à 34°C. L'ensemencement est réalisé à partir de 40 mL de préculture, cultivée sur le milieu de référence pendant 16 à 24h sans agitation en hotte anaérobie. La fermentation est alors opérée en mode *batch* pendant 12 à 24h, puis en mode continu pour le restant de l'expérimentation (typiquement entre 300 et 500 heures de culture). 3 à 5 volumes de milieu de culture sont passés dans le bioréacteur avant le premier échantillonnage pour stabiliser la culture, puis 2 à 3 entre chaque prélèvement.

6.2.3 Analytique

6.2.3.1 Mesure de la biomasse

La biomasse bactérienne est estimée par une mesure de l'absorbance à 600 nm à l'aide d'un spectrophotomètre (Shimadzu, UV-1800). Une normalisation par rapport à la valeur initiale est réalisée pour minimiser les écarts d'absorbance dus au milieu utilisé.

6.2.3.2 Dosages des substrats et des métabolites

Après prélèvement, les échantillons biologiques sont centrifugés (4000 rpm, 5 min). Le surnageant est dilué 20 fois dans de l'eau contenant un standard interne (1-propanol, exactement environ 0,60 g/L), stérilisé par filtration et dosé sur différents systèmes chromatographiques. La quantification sur chacun de ces systèmes est réalisée à l'aide du standard interne (solvants) et de standards externes (solvants, sucres, acides) préparés en interne.

6.2.3.2.1 Dosage des solvants

Les solvants sont séparés par GC sur une colonne CP-PoraBOND-Q (Varian) de 25 m, diamètre interne 0,32 mm, et une épaisseur de film de 5 μm , chauffée graduellement de 50 à 250 °C. L'hélium est utilisé comme gaz vecteur à un débit de 1,6 mL/min. La détection est réalisée à l'aide d'un détecteur d'ionisation de flamme.

6.2.3.2.2 Dosage des sucres

Les sucres sont séparés par HPLC sur une colonne Aminex HXP-87P (BioRad), chauffée à 80°C. L'eau déminéralisée est utilisée comme éluant, à un débit de 0,4 mL/min. Un réfractomètre (Varian 350 RI) permet la détection.

6.2.3.2.3 Dosage des acides

Les acides sont séparés par HPLC sur une colonne Aminex HPX-87H (BioRad) chauffée à 60°C. De l'acide sulfurique 0,01 M est utilisé comme éluant, avec un débit de 0,6 mL/min. Une double détection est utilisée : un réfractomètre (Spectra System RI 150) est couplé à un détecteur UV réglé à 210 nm (Waters 2487 Dual λ). Cette chaîne peut également être utilisée pour mesurer les concentrations de glucose et des différents solvants.

6.2.4 Extractions de matériel biologique

Les prélèvements (5 mL par extraction) sont immédiatement mis sur glace. Une solution « stop » est ensuite ajoutée (625 μL d'une solution d'éthanol et de phénol acide en proportions 19:1) pour arrêter les processus biologiques enzymatiques (transcription, traduction, dégradation etc.) dans les prélèvements destinés aux analyses à haut débit. Un autre prélèvement est utilisé directement pour la quantification des métabolites et des sucres, ainsi que les mesures de DO_{600} et de pH.

En vue des extractions d'ARN et de protéines, les échantillons concernés sont centrifugés 5 min à 5000 g et 4°C puis les culots sont immédiatement congelés à -80°C.

6.2.4.1 Extraction d'ARN

Les culots cellulaires sont repris dans du TRIzol (Invitrogen). Après lyse et centrifugation, le surnageant est traité avec du chloroforme. La phase aqueuse est récupérée et est utilisée pour une précipitation des acide nucléiques à l'isopropanol. Après centrifugation et lavage à l'éthanol 80%, les ARN sont repris dans de l'eau et traités à la DNase Turbo (Invitrogen). Après cette étape, le kit *RNA Clean & Concentrator –25* (Zymo Research) est utilisé pour purifier les ARN de tailles supérieures à 200 nt. La qualité des ARN est vérifiée sur une puce d'un BioAnalyzer (RNA 6000 Nano kit, Agilent) avant congélation et envoi au Génoscope (groupe d'Andrew Tolonen) pour la préparation des bibliothèques de RNA-seq et de eCAPP-seq.

6.2.4.2 Extraction des protéines

Les protéines contenues dans un culot cellulaire sont extraites par une méthode de lyse avec des détergents doux (*B-PER Complete Bacterial Protein Extraction Reagent*, ThermoScientific, supplémenté en PMSF 10 mM et en EDTA 1 mM), selon les instructions du fabricant. La qualité des protéines obtenues est estimée sur gel SDS-PAGE (10% Mini-PROTEAN TGX Precast Protein Gels, Bio-Rad, utilisation selon les recommandations du fabricant). Les protéines sont congelées puis envoyées directement à la plateforme de protéomique de l'Institut Jacques Monod.

6.2.5 Transcriptomique

Les bibliothèques ont été préparées et envoyées au séquençage similairement à ce qui est décrit dans la Partie III (cf. sous-parties 5.2.3 et 5.2.4).

6.2.6 Protéomique

6.2.6.1 Acquisition LC-MS/MS

Les protéines des lysats ont d'abord été précipitées à l'acétone à -20°C pendant 4 heures. 5 µg de protéines ont été incubés une nuit à 37°C avec 20 µL de carbonate d'ammonium 25 mM contenant de la trypsine (12,5 µg/mL ; Promega). Les peptides

obtenus ont été dessalés (*ZipTip μ -C18 Pipette Tips* ; Millipore) puis analysés en triplicats biologiques sur un spectromètre de masse (*Orbitrap Tribrid Fusion* ; ThermoScientific) couplé à une colonne chromatographique (*Nano-LC Proxeon 1200*, ThermoScientific) équipée avec une source ionisante *EASY-Spray* (ThermoScientific). Les peptides sont séparés par chromatographie avec les paramètres suivants : précolonne *Acclaim PepMap100 C18* (2 cm, 75 μ m i.d., 3 μ m, 100 Å), colonne *Pepmap-RSLC Proxeon C18* (50 cm, 75 μ m i.d., 2 μ m, 100 Å), débit à 300 nL/min, gradient de 95% de solvant A (eau, 0,1 % acide formique) à 40% de solvant B (80 % acétonitrile, 0,1 % acide formique), sur une période de 120 minutes. Les colonnes sont régénérées pendant 20 min, pour un temps d'expérimentation total de 140 minutes. Les peptides sont analysés dans une cellule *Orbitrap*, en mode *full ion scan*, à une résolution de 120 000 à une *m/z* de 200, avec une distribution de masses comprises entre 375 et 1550 *m/z* et une cible AGC de 4×10^5 . Les fragments sont obtenus par activation via une dissociation induite par collision haute avec une énergie de collision de 27 %, et une fenêtre d'isolation de quadrupole de 1.6 Da. Les données MS/MS ont été acquises dans le piège à ions en mode *Top-Speed* pendant 3 secondes, avec une cible AGC de 1×10^4 et une exclusion dynamique de 60 secondes. Les données MS/MS du précurseur le plus intense sont d'abord acquises. Les peptides avec un nombre de charges indéterminé ou monochargés ont été exclus de l'acquisition. Le temps maximum d'accumulation d'ions a été réglé sur 100 ms pour l'acquisition MS et 35 ms pour l'acquisition MS/MS.

6.2.6.2 Traitement des données LC-MS/MS

Les fichiers bruts ont été analysés en utilisant 1) le moteur de recherche *Mascot* couplé au logiciel *Proteome Discoverer 2.2* (Thermo Fisher Scientific) pour l'identification des peptides et 2) le logiciel *Progenesis Q1* (Waters) avec deux bases de données protéiques (une pour le chromosome de la souche DSM 6423, et l'autre pour les éléments génétiques mobiles) pour l'identification et la quantification des protéines.

Sur le logiciel *Proteome Discoverer 2.2*, plusieurs modifications post-traductionnelles ont été recherchées : oxydation des méthionines, acétylation des

protéines en N-terminal, phosphorylation des sérines, des thréonines et des tyrosines, et butyrylation des lysines.

6.2.7 Visualisation des données

Les analyses en composantes principales, diagramme de Venn, *heatmaps* et *volcano plots* ont été réalisés avec R grâce aux *packages* FactoMineR, factoextra, VennDiagram, gplots, RColorBrewer, ggplot2.

6.3 RESULTATS

6.3.1 Choix du substrat 1G

Dans une étape préliminaire de ce travail de thèse, plusieurs substrats de première génération ont été testés avec la souche DSM 6423 en fermentation *batch*, en fiole. Les essais de fermentation de 2 sirops issus de l'hydrolyse de l'amidon, de l'amidon de blé et de maïs, et de la mélasse de canne et de betterave sont présentés ci-dessous (Figure 20).

Tous sont fermentescibles et permettent d'obtenir des titres finaux en solvants compris entre 8 et 12 g/L. Partant de ce constat, les sirops et les amidons ont été éliminés pour se focaliser sur les substrats les moins coûteux ou les moins valorisés aujourd'hui par la filière industrielle. Les amidons, de surcroît, sont difficilement compatibles avec un système de fermentation continu à l'échelle laboratoire car peu ou pas pompables sans prétraitement. La mélasse de betterave a finalement été choisie pour deux raisons. La première est que sa composition en sucres (principalement du saccharose, Figure Supplémentaire 7) diffère le plus des conditions déjà explorées par Máté de Gérando *et al*⁵². La deuxième est que la mélasse de betterave semble affecter les performances de la souche, ce qui pourrait se traduire par des changements intéressants à l'échelle moléculaire, accessibles grâce à l'approche protéomique / transcriptomique.

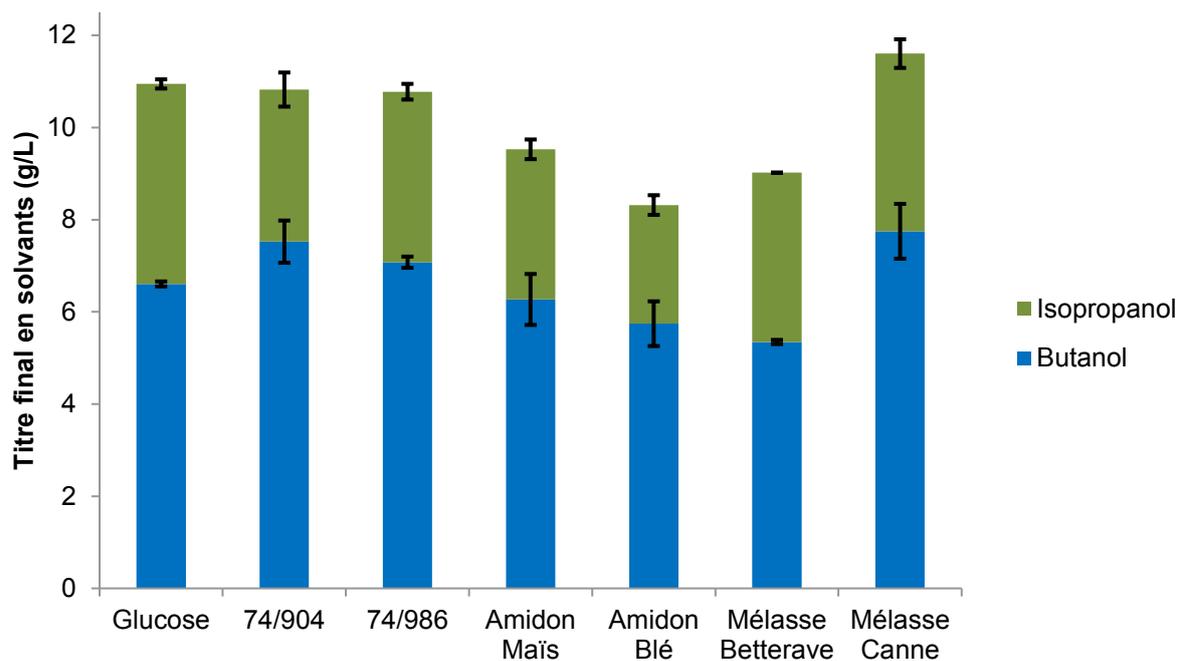


Figure 20 : Profils de productions de solvants par la souche *C. beijerinckii* DSM 6423 sur substrats 1G en fioles serties, points finaux.

Fermentations réalisées en duplicat. Les barres d'erreur correspondent à l'erreur standard de la moyenne.

6.3.2 Séparation des phases acidogènes et solvantogènes en fermentation continue

Les conditions fermentaires développées chez *C. acetobutylicum*¹⁷⁴ ont servi de point de départ pour mettre au point une approche similaire pour *C. beijerinckii*. Chez *C. acetobutylicum*, des phases acidogènes et solvantogènes stables peuvent être obtenues en chemostat (limitation phosphate en milieu minimum) à $D = 0,05 \text{ h}^{-1}$, en jouant sur la valeur du pH consigne (solvantogénèse à pH 4,5, acidogénèse à pH 6,3). Les tests initiaux ont démontré que cette combinaison de facteurs ne fonctionnait pas pour la souche DSM 6423, en particulier pour la phase solvantogène. De nombreuses améliorations incrémentales ont été nécessaires pour arriver à un résultat satisfaisant, c'est-à-dire un ensemble de conditions expérimentales qui permettent, indépendamment du substrat, de mesurer un pourcentage massique d'acides (lactique, acétique, butyrique) dans les produits proches de 100% en phase acidogène, et un pourcentage massique de solvants dans les produits proche de 100%

en phase solvantogène (butanol, isopropanol, acétone, éthanol). Ces améliorations sont décrites ci-après.

En utilisant un taux de dilution $D = 0,045 \text{ h}^{-1}$ en acidogénèse, les conditions testées permettaient déjà d'obtenir 89% d'acides dans le panel des produits. Les travaux initiaux ont donc essentiellement visé à améliorer la phase solvantogène. Le pH dans cette phase a tout d'abord été augmenté à 5,4. De 4,5 à 5,4, le pH n'avait que peu ou pas d'effet sur la distribution des produits (environ 50% de solvants), mais avait un impact important sur la consommation de sucres, sur la production de biomasse, et donc sur le titre en produits.

Malgré une amélioration du titre en solvants, le taux de dilution était trop élevé pour permettre une croissance normale sur la mélasse de betterave, probablement à cause de sa concentration élevée en sels, se traduisant par un taux de croissance limité. Le taux de dilution a alors été réduit à $0,015 \text{ h}^{-1}$, permettant d'observer une croissance bactérienne, mais également d'augmenter le pourcentage en solvants à 80%.

Pour améliorer encore les conditions expérimentales en solvantogénèse, de l'acétate (sous la forme d'acétate de sodium à 3,3 g/L), a été ajouté au milieu de culture contenant la mélasse. La présence d'acétate, un des produits majeurs du métabolisme acidogène, stimule l'expression des voies solvantogènes et provoque son assimilation chez *C. acetobutylicum*²⁷⁸. Dans ces conditions expérimentales, la souche ne produit plus que des solvants.

Ces améliorations auraient pu être suffisantes, mais au cours de la validation finale sur milieu synthétique (glucose), les performances de la souche se sont montrées instables, avec une baisse de 30 à 40 % de solvants en l'espace de quelques jours. Ce phénomène n'était pas visible en présence de mélasse, ce qui suggère que la mélasse, en tant que produit riche, apporte également des éléments importants pour le métabolisme solvantogène.

Partant de cette hypothèse, nous avons remplacé le milieu minimum par un milieu riche (milieu Gapes, contenant 2,5 g/L d'extrait de levure, avec 2,9 g/L ou 0 g/L d'acétate d'ammonium en fonction de la phase métabolique), ce qui a corrigé le problème d'instabilité, tout en apportant de nouveaux problèmes.

La phase acidogène, en particulier, était moins performante en termes de distribution des produits (81% en acides). Les conditions expérimentales ont donc été optimisées sur milieu riche, en augmentant à la fois le pH (7,0) et le taux de dilution ($D = 0,100 \text{ h}^{-1}$), ce qui a permis d'augmenter ce pourcentage à environ 99%.

Pour la phase solvantogène, le taux de dilution a également été augmenté ($D = 0,025 \text{ h}^{-1}$), essentiellement par commodité expérimentale, ce qui a permis d'obtenir entre 85 et 96% de solvants, tout en étant relativement stable sur une période de 400 heures environ.

En résumé, deux conditions expérimentales permettant de fixer le métabolisme en acidogénèse ou solvantogénèse ont été développées pour *C. beijerinckii* DSM 6423. Ces conditions sont compatibles avec la mélasse de betterave et sont caractérisées en fermentation continue par les paramètres suivants :

- Acidogénèse : milieu Gapes, $D = 0,025 \text{ h}^{-1}$, pH = 7,0 ;
- Solvantogénèse : milieu Gapes, $D = 0,100 \text{ h}^{-1}$, pH = 5,4, acétate d'ammonium 2,9 g/L.

6.3.3 Plan d'expériences et analyse fermentaire

En plus de la phase métabolique, l'autre paramètre variable dans le plan d'expérience était la nature du substrat carboné. Pour ce dernier, nous avons défini une condition contrôle avec un sucre modèle caractérisé : le glucose. La mélasse de betterave a également été choisie pour les raisons explicitées dans le point 6.3.1. La mélasse de betterave étant essentiellement riche en saccharose, nous avons rajouté une condition saccharose, qui doit permettre de découpler l'effet lié au sucre et l'effet du reste des composés de la mélasse sur l'expression génétique.

Les échantillonnages ont été réalisés en quadruplicat, afin de prévenir des erreurs expérimentales et de pouvoir analyser trois échantillons pour chaque condition lors de l'analyse omique. Les résultats en termes de titre et de rendement en produits (par rapport à la quantité de sucres consommés) sont respectivement montrés dans la Figure 21 et la Figure 22. Le tableau de valeurs correspondantes est rapporté en Figure Supplémentaire 8.

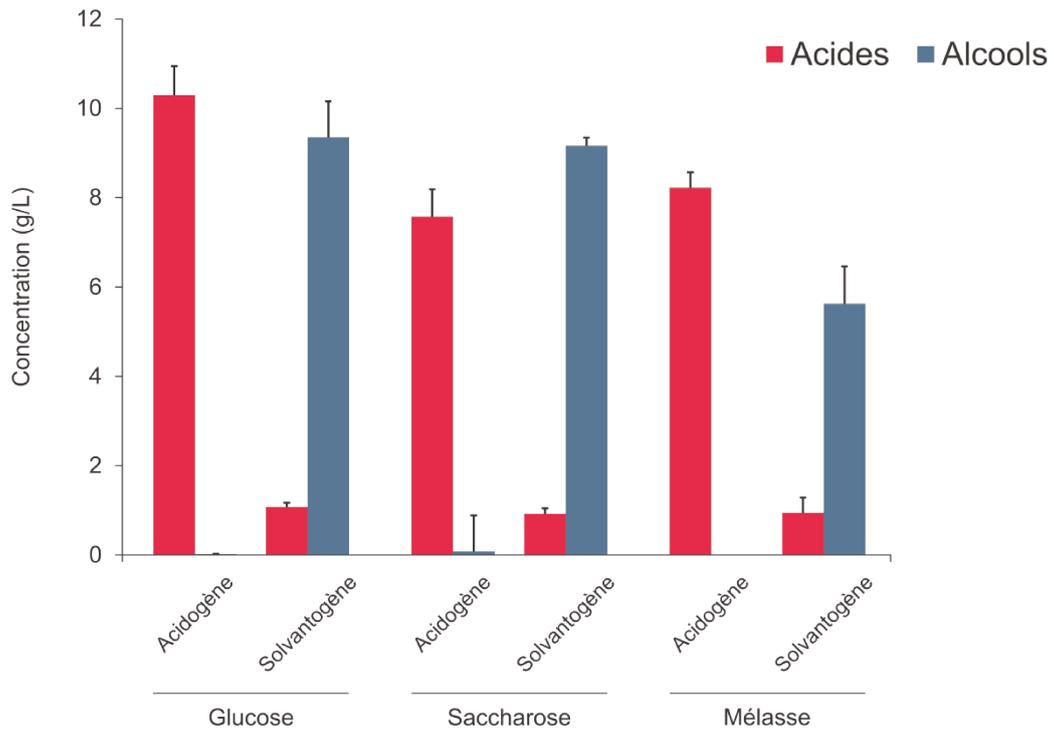


Figure 21 : **Titres cumulés des acides et des solvants dans les différents échantillons.** Les barres d'erreur représentent l'erreur standard de la moyenne de 4 réplicats biologiques.

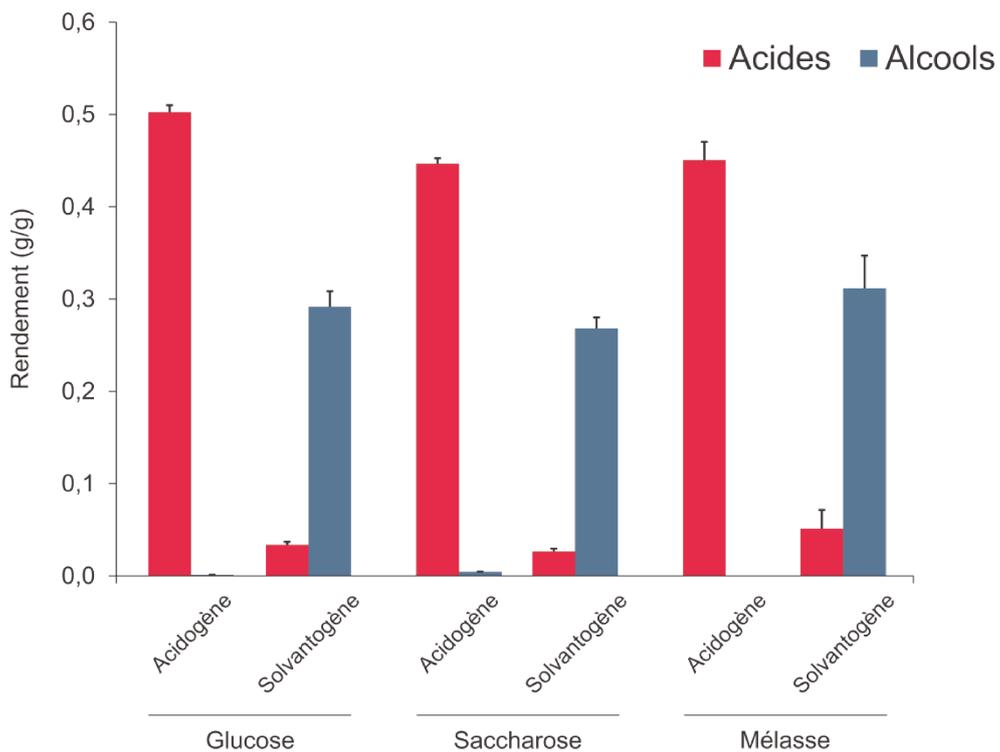


Figure 22 : **Rendements cumulés en acides et en solvants dans les différents échantillons.** Les barres d'erreur représentent l'erreur standard de la moyenne de 4 réplicats biologiques.

Afin de vérifier la qualité de l'expérimentation, une analyse de la variance, basée sur les différentes grandeurs mesurées (dosage de chaque métabolite, biomasse), a également été réalisée (Figure 23 et Figure 24).

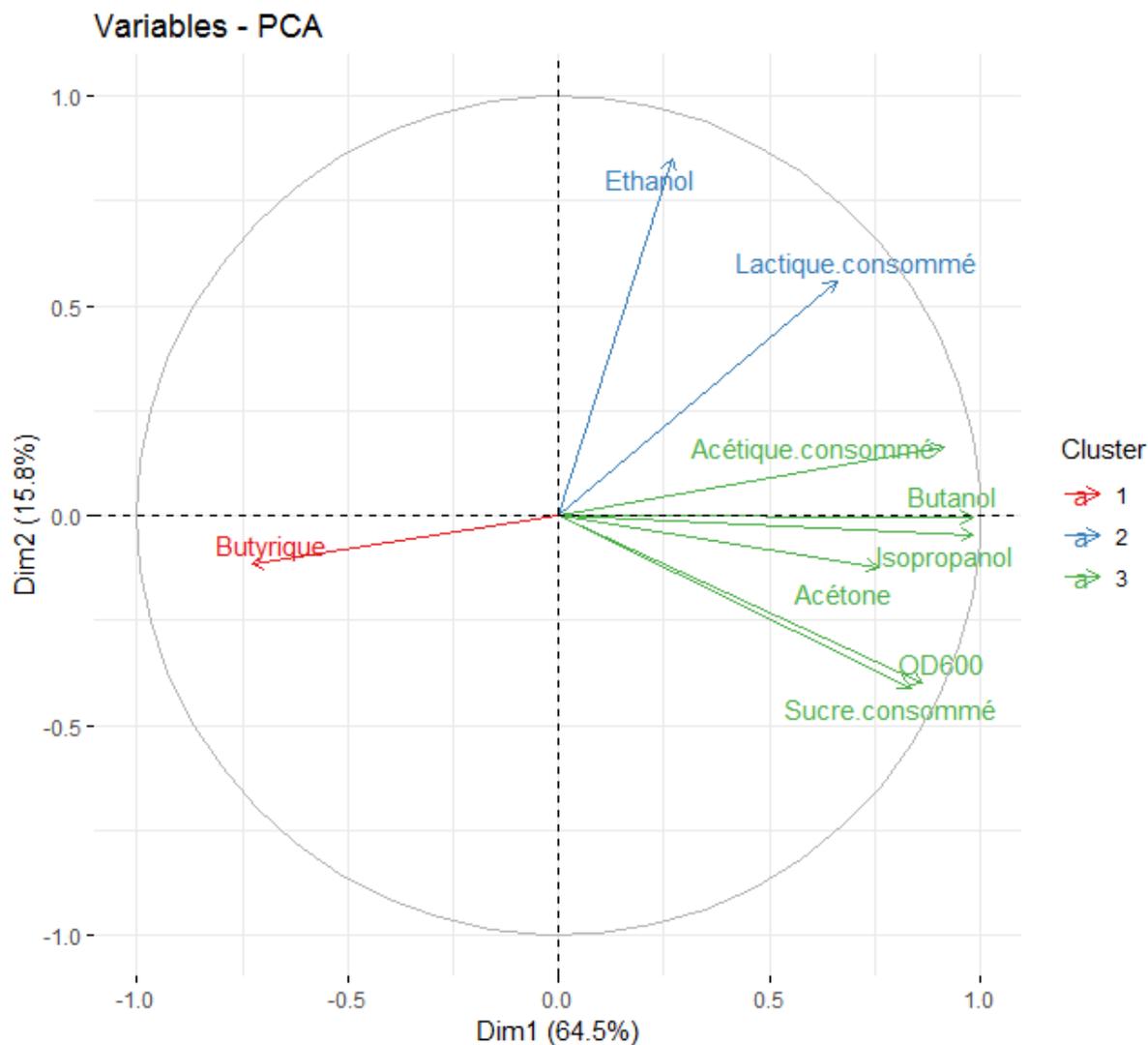


Figure 23 : Analyse en composantes principales : représentation des variables selon les deux premières composantes de la variance (cercle des corrélations).

La couleur regroupe les variables dont la variation est positivement corrélée.

PCA - analyse individuelle

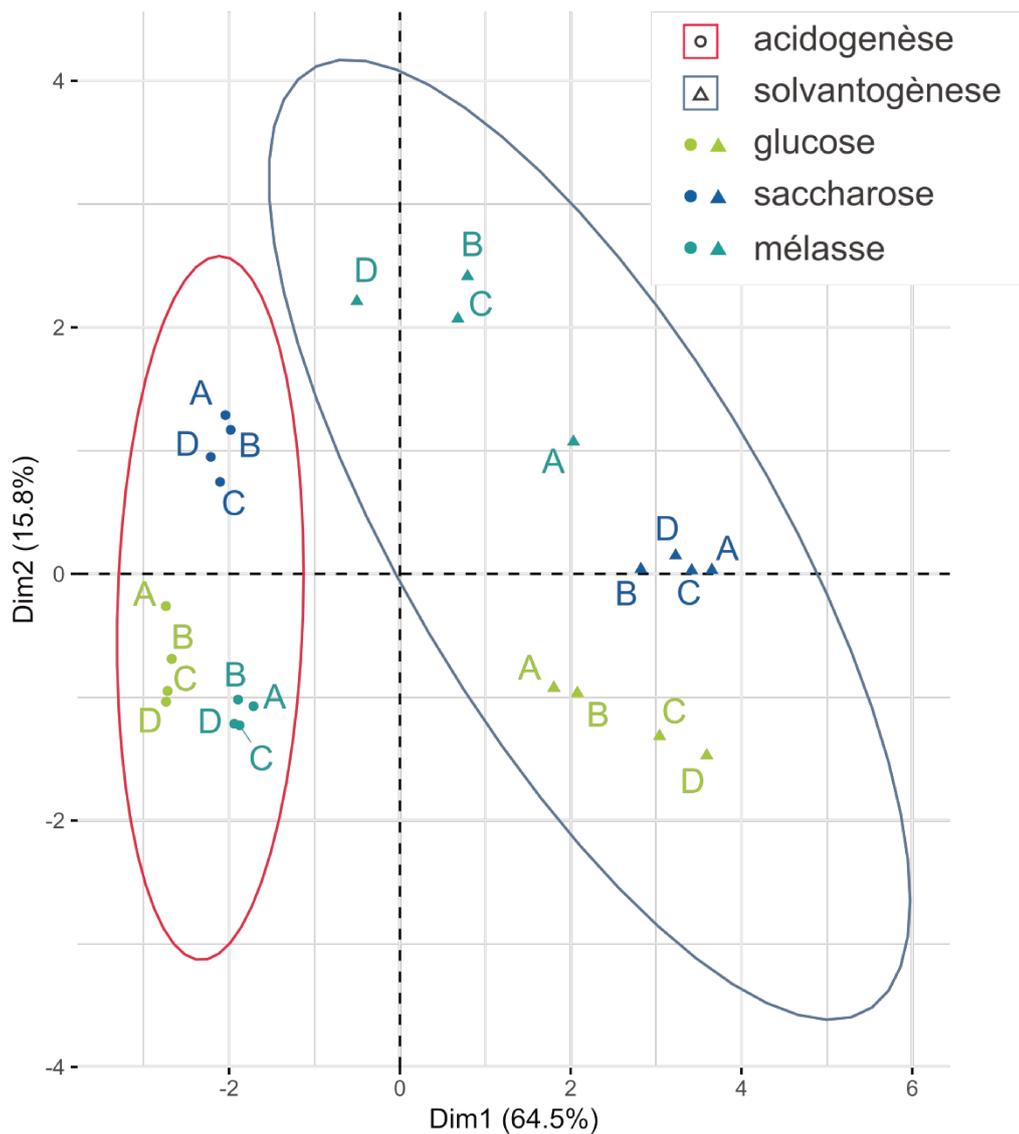


Figure 24 : Analyse en composantes principales : représentation des individus selon les deux premières composantes de la variance.

L'axe 1 de l'ACP mesure essentiellement les variations liées aux phases métaboliques (dosages de la plupart des métabolites, biomasse). L'axe 2 prend essentiellement en compte deux variables : la concentration en éthanol et en acide lactique. De l'acide lactique (et dans une moindre mesure, de l'acide acétique) est présent initialement dans la mélasse de betterave, ce qui se traduit par une consommation de ce substrat en phase solvantogène. Parallèlement, les données indiquent que l'éthanol n'a été détecté que dans un nombre limité de conditions (saccharose acidogène et solvantogène, ainsi que mélasse acidogène). Par ailleurs, les variables de consommation des sucres et de mesure de la biomasse ont également une composante non négligeable sur l'axe 2, et ces deux paramètres inter-corrélés,

s'ils dépendent principalement de la phase métabolique, varient également significativement entre les substrats (par exemple, doublement de la DO_{600} et de la consommation de sucres entre la condition glucose solvantogène et la condition mélasse solvantogène). Dans une certaine mesure, on peut donc considérer que l'axe 2 sépare les échantillons en fonction de la source de carbone.

L'ensemble de ces résultats est globalement satisfaisant, avec un regroupement des différents réplicats sur les deux axes de l'ACP, des rendements en acides / solvants similaires selon les substrats, et une forte représentation des acides et solvants en phases acidogène et solvantogène, respectivement. Ce constat suggère que la qualité du plan d'expérience est suffisante pour réaliser une analyse omique complémentaire à cette description physiologique.

6.3.4 Analyse omique préliminaire

6.3.4.1 Stratégies d'analyse

L'un des inconvénients des conditions de fermentation continue explorées dans cette étude est que, pour une partie des expérimentations (en phase solvantogène, notamment), les cellules sont en phase quasi-stationnaire, ce qui se traduit 1) par des ARNs de mauvaise qualité (*RNA Integrity Number* 6-7, en moyenne) et 2) par une proportion d'ARNm dans les ARN totaux probablement moins importante qu'en phase semi-logarithmique.

Ces défauts ont ralenti les efforts de préparation des bibliothèques (de RNA-seq, et, dans une plus grande mesure, d'eCAPP-seq), qui devraient toutefois être prochainement séquencées.

A l'inverse, les protéines semblaient être d'excellente qualité (exemple Figure 25). Les échantillons ont donc pu être, sans problème technique, envoyés à la plateforme de protéomique de l'Institut Jacques Monod pour analyse quantitative.

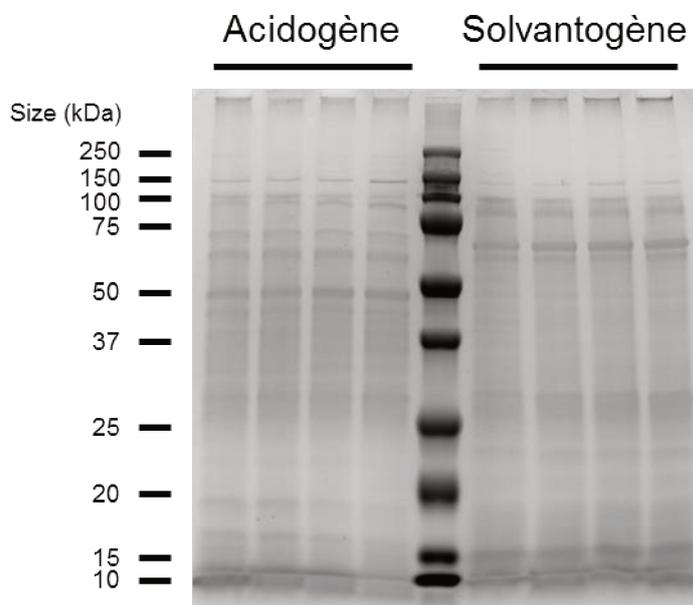


Figure 25 : **Séparation des protéines totales sur gel SDS-PAGE 10%.**
Les 4 réplicats pour la condition glucose sont représentés en acidogenèse et en solvantogenèse.

Pour toutes les conditions explorées, 1739 protéines ont été identifiées et quantifiées par l'analyse LC-MS/MS. Ces protéines sont principalement cytosoliques, et correspondent donc à la fraction étudiée. Quelques protéines membranaires sont également visibles, et leur présence dans la fraction cytosolique est probablement due à la méthode d'extraction utilisée (lyse cellulaire en présence de détergents doux, susceptibles de solubiliser certaines protéines membranaires).

L'analyse en composante principale (Figure 26) suggère l'existence de 4 groupes principaux pour les individus. Le premier regroupe toute les conditions solvantogènes, les trois autres regroupent séparément les conditions acidogènes des trois substrats. Cette répartition nous a poussé à faire deux types d'analyses comparatives : une comparaison acidogène / solvantogène pour chaque substrat, et une comparaison substrat à substrat pour les conditions acidogènes.

PCA - analyse individuelle

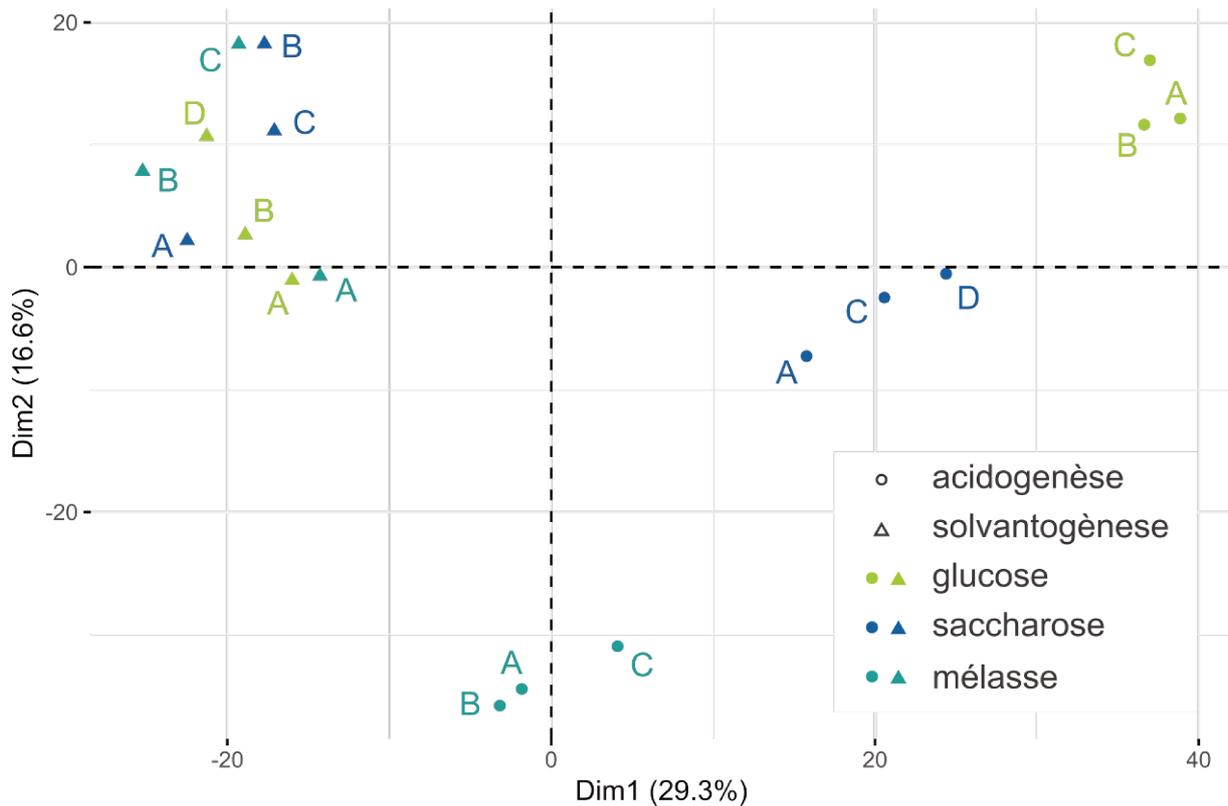


Figure 26 : **Analyse en composantes principales : représentation des individus passés en protéomique selon les deux premières composantes de la variance.** La numérotation des individus est la même que celle de la Figure 24. Les variables utilisées pour l'analyse sont les valeurs d'abondance pour chaque protéine identifiée.

6.3.4.2 Comparaison acidogène / solvantogène

6.3.4.2.1 Tendances générales

Les comparaisons acidogénèse/solvantogénèse par *volcano plots* indiquent un comportement dépendant de la source carbonée, avec un nombre de protéines enrichies dans chaque état physiologique assez variable (Figure 27). Cette analyse est cohérente avec l'ACP, et souligne le besoin de prendre en compte la source carbonée dans cette comparaison. Comme suggéré par l'ACP, les jeux de données glucose et saccharose sont plus proches l'un de l'autre avec une corrélation de Pearson calculée sur un ratio (*fold change*) entre conditions solvantogène et acidogène de 0,70, contre 0,49 et 0,48 pour les jeux de données glucose / mélasse et saccharose / mélasse, respectivement. Les jeux de données glucose et saccharose pourraient donc être regroupés sous un label « milieu modèle », et la mélasse dans une catégorie à part (« milieu industriel »), ce qui est cohérent avec la composition très différente entre ces types de milieu.

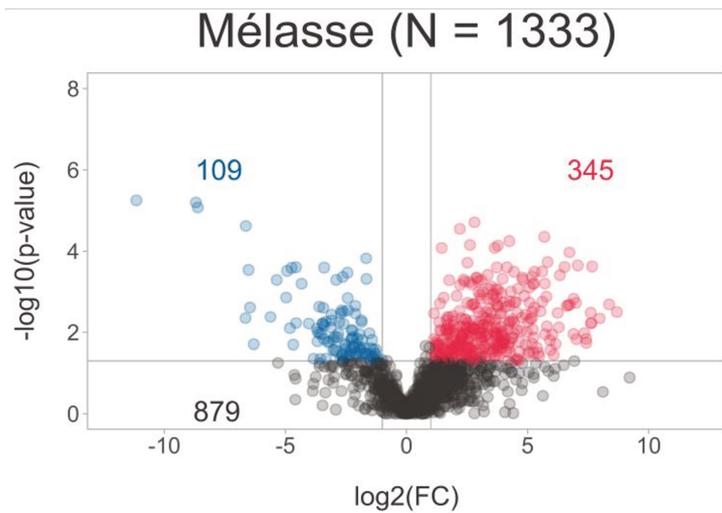
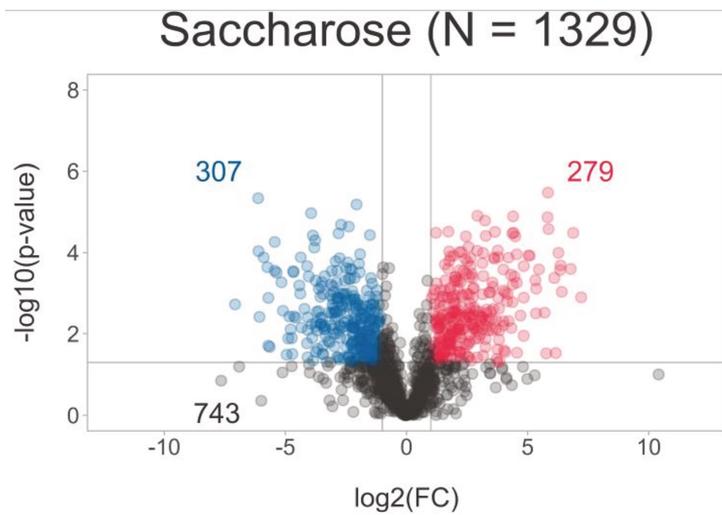
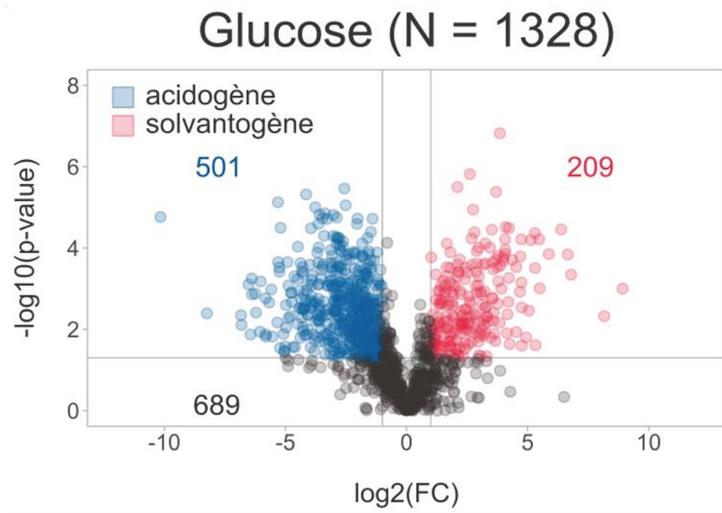


Figure 27 : **Comparaison acidogène / solvantogène, volcano plots.**
 Pour chaque condition, seules les protéines avec un minimum de 2 peptides identifiés ont été conservées. Une

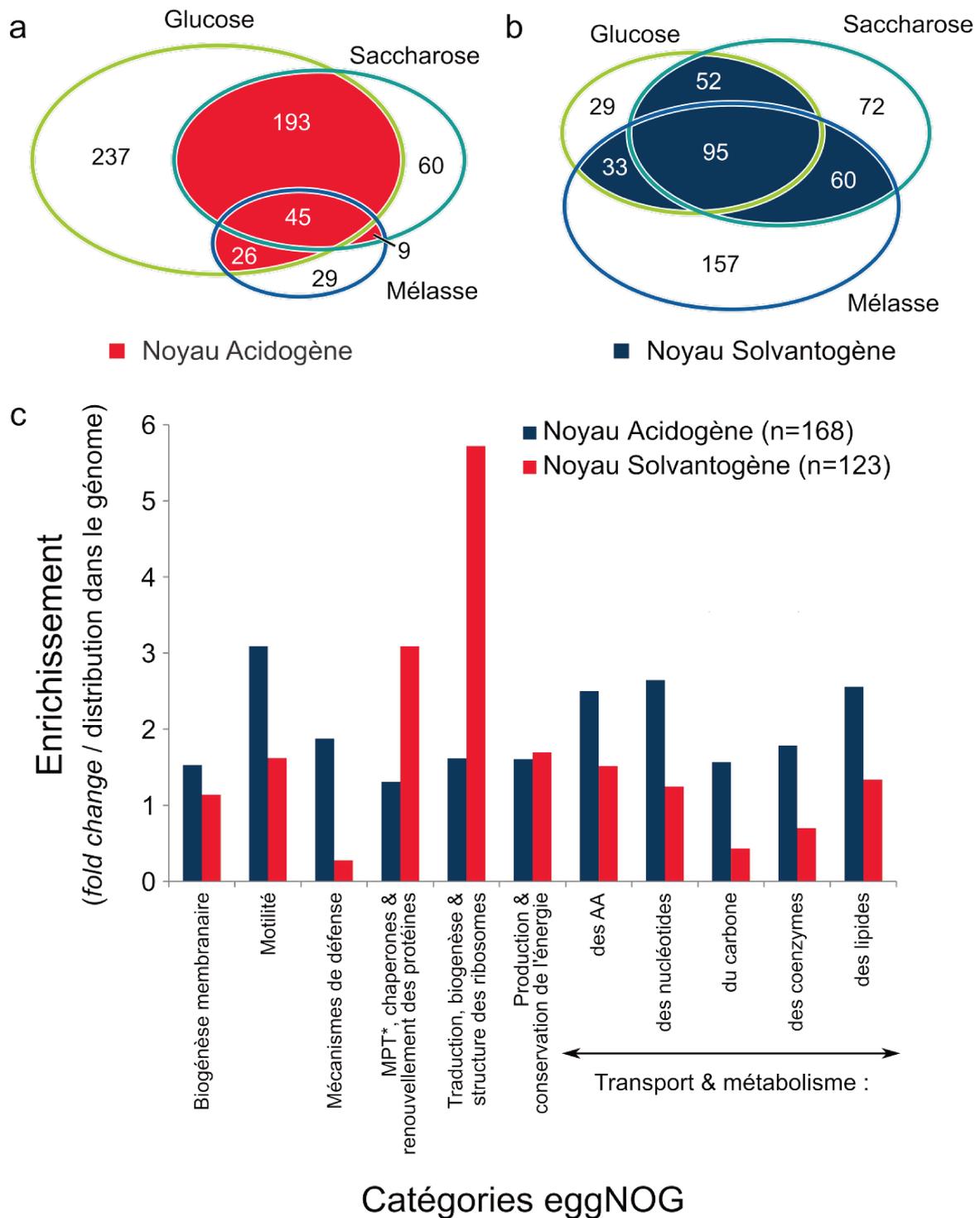


Figure 28 : **Comparaison acidogène / solvantogène, protéines communes dans chaque jeu de données.**
a,b. Diagrammes de Venn représentant les protéines communes enrichies significativement en phase acidogène (**a**) et en phase solvantogène (**b**). Les gènes dont les produits sont présents dans au moins deux jeux de données sont considérés comme faisant partie d'un noyau de gènes spécifiques de la phase métabolique considéré. **c.** Enrichissement eggNOG réalisé sur les noyaux déterminés en (a,b). Seuls les groupes d'au moins 5 gènes avec un enrichissement supérieur à 1,5 dans une des deux catégories sont considérés. *MPT : modification post-traductionnelle.

Les enrichissements en protéines pour chaque phase métabolique et pour chaque substrat ont ensuite été comparés, de manière à obtenir un groupe minimal de protéines enrichies pour au moins 2 des 3 substrats (Figure 28ab). L'enrichissement eggNOG réalisé sur ces groupes de gènes révèle des informations intéressantes (Figure 28c). En solvantogénèse, deux catégories majeures sont représentées. La première (« modifications post-traductionnelles, chaperones et renouvellement des protéines ») s'avère avant tout liée à un enrichissement des chaperones et protéines de choc thermique (Figure Supplémentaire 9). Ce résultat est attendu, puisque ces protéines sont impliquées dans des mécanismes de résistance aux solvants chez les souches qui en sont productrices⁸³. De façon plus surprenante, la deuxième catégorie (« traduction, structure et biogénèse des ribosomes ») regroupe essentiellement des protéines ribosomales, presque toutes enrichies pour les trois substrats.

En phase acidogène, les catégories eggNOG les plus enrichies ont trait au métabolisme, à la motilité, et aux mécanismes de défense (Figure 28). Si les deux premières sont attendues (la motilité par rapport à la phase métabolique¹⁸⁷, le métabolisme par rapport au taux de croissance élevé), la dernière est plus surprenante. Après un examen manuel, la prédiction eggNOG pour les protéines impliquées dans cette catégorie est peu convaincante (contrairement aux autres catégories). Pour la suite, une catégorisation fonctionnelle a donc été réalisée manuellement pour analyser les données. Dans le cas des mécanismes de défense, l'enrichissement en acidogénèse n'est pas clair et plutôt dépendant du substrat (Figure Supplémentaire 9). Une exception est toutefois notable, celle des gènes CIBE_3326 et CIBE_3328, qui codent respectivement une enzyme de restriction et une enzyme de modification de l'ADN. Outre leur régulation différentielle, la seule identification de ces gènes est potentiellement intéressante, puisqu'il s'agit des deux seuls gènes de R/M détectés par la MS, ce qui suggère qu'il s'agit du système R/M principal de la souche. Cette information pourrait éventuellement être utilisée pour améliorer les efficacités de transformation de la souche.

6.3.4.2.2 Motilité

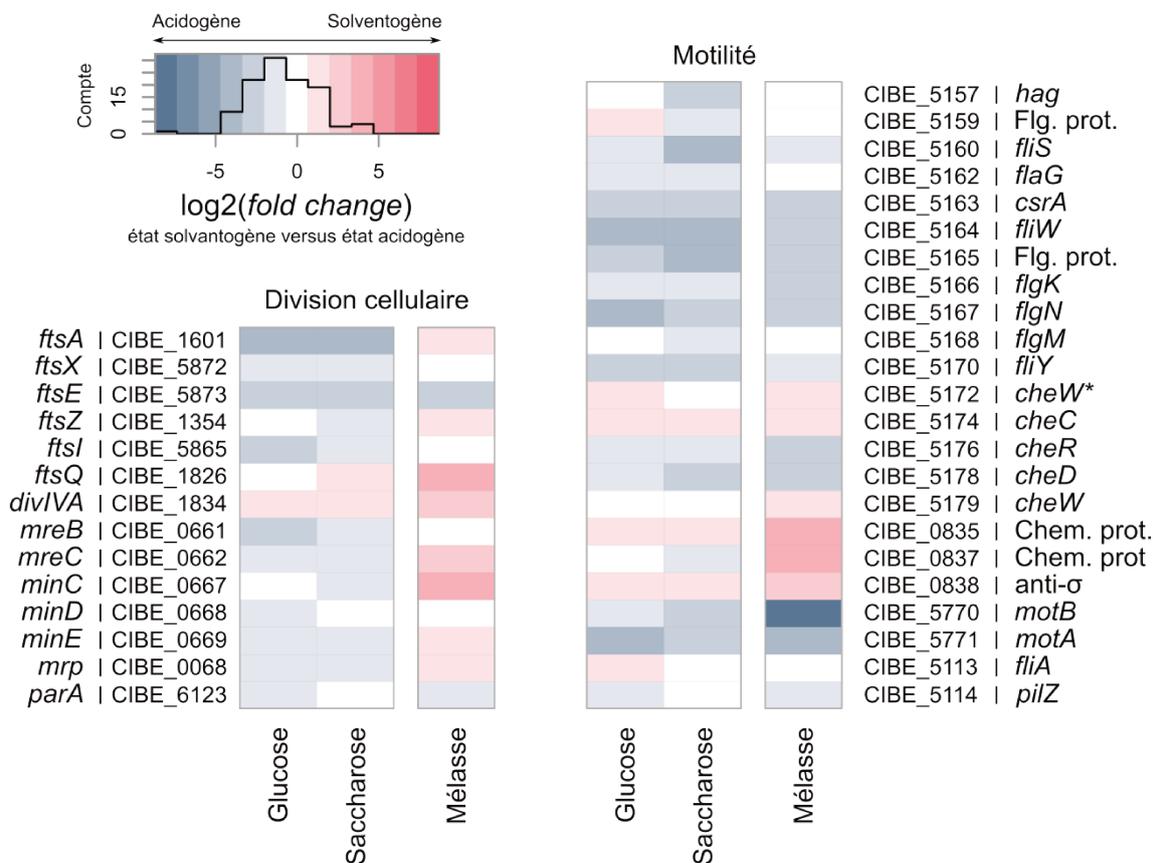


Figure 29 : Comparaison acidogène / solvantogène, heatmap des catégories « Division cellulaire » et « Motilité ».

Pour les gènes dupliqués, l'astérisque * indique la/les protéine(s) majoritaire(s). Flg. prot. : protéine de flagelle, Chem. prot. : protéine de chimiotactie.

La catégorisation fonctionnelle manuelle confirme la plupart des tendances communes relevées par l'enrichissement eggNOG, tout en soulignant les différences entre substrats (Figure Supplémentaire 9). La motilité est par exemple associée à la phase acidogène, et cela quel que soit le substrat (Figure 29). En revanche, les protéines liées à la division cellulaire montrent un comportement dépendant du substrat (Figure 29). Avec un taux de dilution du milieu plus élevé en phase acidogène, une augmentation de l'expression des protéines impliquées dans la division cellulaire et la réplication de l'ADN (Figure Supplémentaire 9) est attendue, mais n'est réellement observée que pour les conditions en milieu modèle. A l'inverse, la division cellulaire semble être affectée négativement en phase acidogène sur mélasse, c'est-à-dire dans la condition pour laquelle la vitesse d'introduction de milieu frais est maximale. Cela

suggère un effet inhibiteur fort, probablement dû à la présence de molécules toxiques, qui pourraient être éliminées plus facilement en phase solvantogène, quand le taux de dilution est moins élevé. Cette tendance est visible également dans les comparaisons substrat à substrat (cf. sous-partie 6.3.4.3).

6.3.4.2.3 Métabolisme A/IBE

Des tendances intéressantes sont également visibles au niveau du métabolisme A/IBE (Figure 30). Une différence marquée est encore visible entre substrats modèles et industriel, avec une induction de la plupart des gènes liés au métabolisme A/IBE sur mélasse, probablement pour la raison susmentionnée (inhibition en phase acidogène).

La régulation est globalement plus fine dans les conditions modèles. La plupart des protéines liées au métabolisme acidogène sont plus exprimées en phase solvantogène, ou gardent à minima un niveau d'expression constant. Cela suggère que la différence en termes de métabolites produits dépend de la présence des protéines impliquées dans la solvantogénèse qui sont, effectivement, bien plus exprimées dans ces conditions. De façon assez intéressante, l'évolution de Ptb et Buk est différente malgré l'organisation des gènes correspondants en opéron (voir Cary et coll.²⁷⁹ chez *C. acetobutylicum*, également visible sur les données eCAPP-seq, cf. partie III), ce qui implique probablement un contrôle post-transcriptionnel différent.

La voie de synthèse du butyryl-CoA est également légèrement dérégulée en phase solvantogène, à l'exception notable de Hbd qui présente une régulation différenciée entre les conditions glucose et saccharose, indépendante des produits des autres gènes de l'opéron BCS. Cette observation est cohérente avec les résultats obtenus dans la Partie III, c'est-à-dire que, chez *C. beijerinckii*, *hbd* est sous le contrôle de son propre promoteur. Pour les autres gènes de l'opéron, le profil de régulation est positivement corrélée à celui du produit du gène *rex*, un régulateur transcriptionnel sensible au pouvoir réducteur, et qui réprime chez *C. acetobutylicum* l'expression de gènes impliqués dans la synthèse de butanol, en particulier l'opéron BCS et le gène de la thiolase¹⁴⁹.

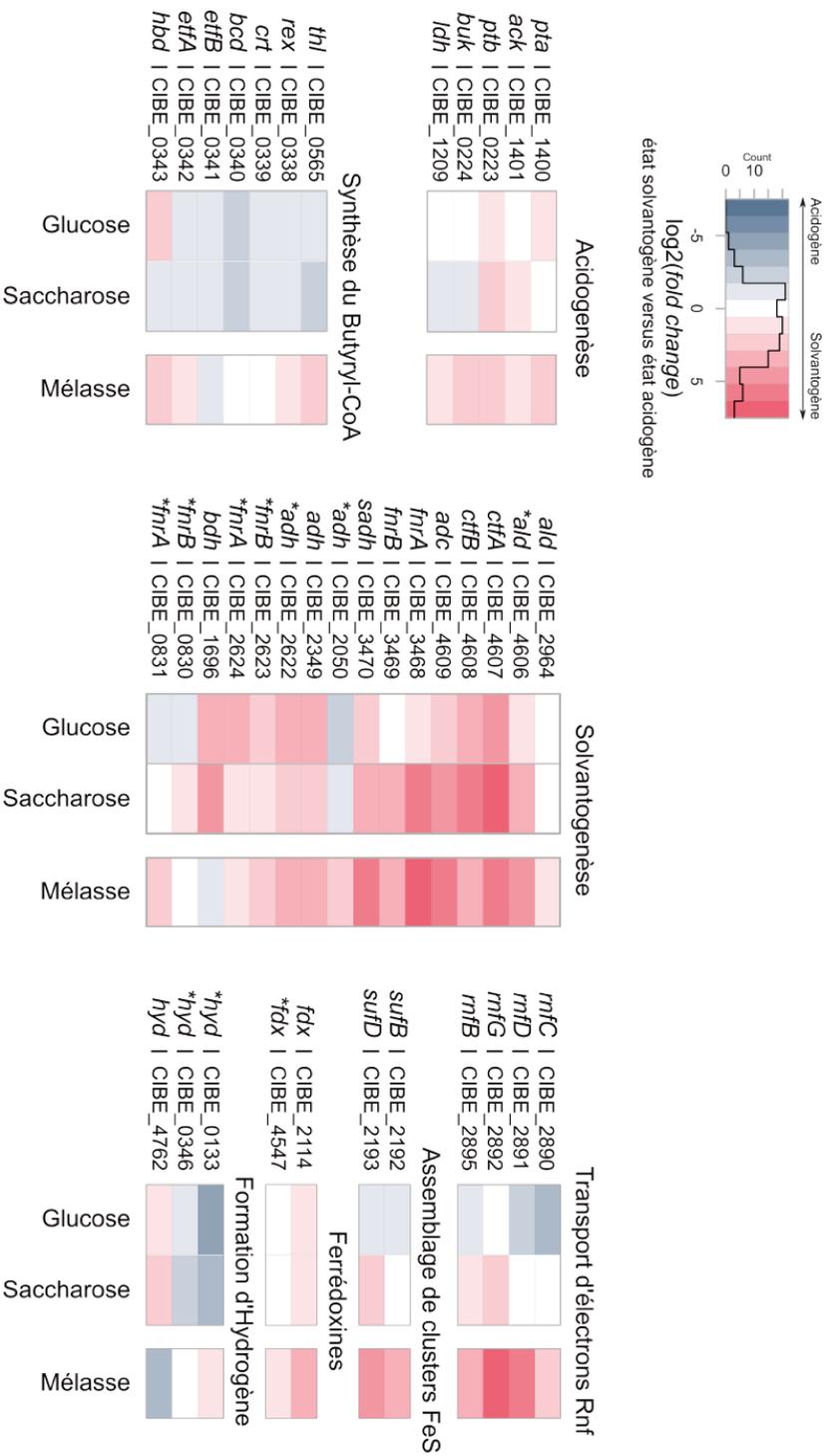


Figure 30 : Comparaison acidogène / solvantogène, heatmap de la catégorie « Métabolisme AIBE ». Pour les gènes dupliqués, l'astérisque * indique la/les protéine(s) majoritaire(s). bdh indique ici un gène encodant une enzyme de type 2,3-butanediole déshydrogénase.

Les voies solvantogènes sont chez les *Clostridia* souvent sujettes à des redondances, ce qui pose la question des implications de chaque paralogue dans le phénotype final¹⁷³. Avec les mesures des abondances absolues pour chaque protéine, la protéomique permet d'avoir une idée plus précise de cette répartition*. Ainsi, les gènes ici définis comme majeurs codant l'aldéhyde déshydrogénase (CIBE_4606), la ferrédoxine (CIBE_4547), les hydrogénases (CIBE_0133 et CIBE_0346, toutes deux spécifiques de la phase acidogène) et les alcool déshydrogénases (CIBE_2050 et CIBE_2622) sont identifiées dans nos conditions expérimentales. De façon surprenante, ces deux enzymes présentent un profil de régulation différencié, qui suggère une participation différente dans l'activité alcool déshydrogénase globale de la souche.

La caractéristique principale de la souche DSM 6423 est la présence d'une alcool déshydrogénase secondaire (CIBE_3470) surreprésentée en phase solvantogène. L'organisation du gène *s-adh* correspondant en opéron avec un gène codant une sulfure déshydrogénase putative (*sudA*, CIBE_3468) et un gène dont le produit a une fonction inconnue (CIBE_3469) a été préalablement décrite⁵². Dans le cadre de cette étude, nous avons aligné la séquence protéique de CIBE_3469 avec *sudB* (Uniprot Q8U194) de *Pyrococcus furiosus* (Figure 31). Malgré l'éloignement phylogénétique entre les souches, l'alignement indique que CIBE_3469 et *sudB* sont relativement proches.

Par ailleurs, le complexe SudAB a été caractérisé chez *P. furiosus* et présente notamment une forte activité NADPH-ferrédoxine oxydoréductase²⁸⁰. Du fait de l'organisation en opéron des gènes CIBE_3468/3469/3470 et du fait que l'alcool déshydrogénase secondaire soit dépendante du NADPH, nous pouvons supposer que l'activité NADPH-ferrédoxine oxydoréductase est probablement liée à celle de s-ADH et proposons de réannoter respectivement les gènes CIBE_3468 et CIBE_3469 en *fnrA* et *fnrB*. De façon assez intéressante, des homologues de ces deux gènes sont également présents dans la souche DSM 6423 (CIBE_2623/2624 et CIBE_0830/0831), et l'abondance protéique des produits de ces deux gènes est nettement plus forte que celle des produits des gènes CIBE_3468/3469. Les protéines

* par soucis d'économie d'espace, les valeurs des abondances absolues ne sont pas rapportées dans cette thèse

codées par CIBE_2623/2624 sont probablement associées à l'alcool déshydrogénase majeure CIBE_2622, également prédite comme étant NADPH-dépendante (tout comme l'est l'autre alcool déshydrogénase majeure, CIBE_2050).

```

Query: sp|Q8U194|SUDHB_PYRFU Sulfide dehydrogenase subunit beta OS=Pyrococcus
furiosus (strain ATCC 43587 / DSM 3638 / JCM 8422 / Vc1) OX=186497 GN=sudB
PE=1 SV=1 Query ID: lcl|Query_234039 Length: 278

>CIBE_3469
Sequence ID: Query_234041 Length: 296
Range 1: 1 to 281

Score:281 bits(719), Expect:1e-99,
Method:Compositional matrix adjust.,
Identities:148/281(53%), Positives:189/281(67%), Gaps:9/281(3%)

Query 1 MFKILRKERLAPGINLFEIESPRIAKHAKPGQFVMIRLHEKGERIPLTIADVDISKGSIT 60
M+KI+ K+ L I +IE+PR+AK AKPGQF++I+ EKGERIPLTIAD D KG++T
Sbjct 1 MYKIVSKKELTNNIFSMIDIEAPRVAKSAPGQFIIIKNDEKGERIPLTIADYDQEKGTVT 60

Query 61 IVAQEVGKTTRELGTYEAGDYILDVLGPLGKPSHIDYFG-----TVVMIGGGVGVAEI 113
IV Q VGK I++L + GD++ D +GPLG PS + + + GGVG A +
Sbjct 61 IVFQTVGKGTKQLAAFNEGDHVDVFGPLGVPSEFIHEDIEELKKNFIFVAGGVGAAPV 120

Query 114 YPVAKAMKEKGNVISILGFRTKDLVFWEDKLRSVSDEVIVTTNDGSYGMKGFTTHALQK 173
YP K M E G V ILG R KDL+ +E+KL++ + + VTT+DGSY KG + L++
Sbjct 121 YPQVKWMHEHGIADVILGSRNKDLLIYEKLNAAAGNLYVTTDDGSYEFKGTGSDMLKE 180

Query 174 LIE-EGRKIDLVHAVGPAIMMKAVAELTKPYGIKTVASLNPIMVDGTGMCGACRVTVGGE 232
L+ +G+K D +GP IMMK + LTK GI T SLNPIMVDGTGMCGACRVTVGGE
Sbjct 181 LVNNQGKKYDHAIIGPMIMMKFTSMLTKELGIPTTVSLNPIMVDGTGMCGACRVTVGGE 240

Query 233 VKFACVDGPEFDAHLVDWDQLMNRLAYYRDLE-KISLEKWE 272
VKFACVDGPEFD HLV++D+ M R A Y+ E K LE E
Sbjct 241 VKFACVDGPEFDGHLVNYDESMRRQAMYKTEEGKAQLEVEE 281
    
```

Figure 31 : Alignement *blastP* entre CIBE_3469 et *sudB* de *P. furiosus*.

Alignement réalisé sur la plateforme en ligne du NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) avec les paramètres par défaut.

6.3.4.2.4 Stress et sporulation

Les processus solvantogènes sont souvent associés à différents procédés biologiques, comme la production de protéines chaperones et d'adaptation aux chocs thermiques, qui ont de longue date été associées à des mécanismes de résistance aux solvants⁸³, et à la sporulation²⁰⁹. Dans le cadre d'une fermentation continue, le nombre de spores évolue rapidement pour atteindre un niveau nul en quelques jours, malgré le fait qu'une partie des gènes liés à la sporulation soient induits dans ces

conditions²⁸¹. Nous nous sommes donc intéressés à ces différentes voies dans le cadre de cette étude protéomique (Figure 32).

Les données révèlent que la phase solvantogène n'induit l'expression que d'une quantité limitée de protéines dites de « choc thermique » en conditions modèles, avec les plus gros effets relevés pour GroESL et Hsp18 (toutes deux impliquées dans la résistance au stress lié au passage en phase solvantogène²⁸⁶⁻²⁸⁹. A l'inverse, sur substrat industriel, la plupart des protéines de choc thermique sont induites en phase solvantogène, ce qui pourrait être lié à un stress dû à la composition du milieu. Similairement, sur ce milieu industriel, la plupart des protéines de sporulation retrouvées dans le jeu de données sont surreprésentées en solvantogénèse, ce qui contraste avec les milieux modèles. Il convient toutefois de noter un phénomène similaire concernant les protéines de sporulation de phase II à V pour les conditions saccharose et mélasse, et qui pourrait donc être dû à la présence de saccharose, sucre commun à ces deux milieux.

Les histidine kinases sont des protéines impliquées dans divers procédés biologiques chez les Firmicutes, et qui incluent notamment la sporulation²⁸² et la répression catabolique du carbone (RCC)¹¹⁴. Dans nos conditions expérimentales, une seule histidine kinase est détectée pour toutes les conditions (CIBE_0761). Etant donné son profil de régulation (plutôt présente en phase acidogène, et plus particulièrement sur glucose cf. Figure Supplémentaire 9) et la position du gène correspondant sur le chromosome (proche des gènes CIBE_0769 à CIBE_0774 impliqués dans la glycolyse), il est probable que cette dernière soit impliquée dans les mécanismes de RCC plutôt que dans la sporulation.

Partie IV : Métabolisme sur substrat 1G, approches fermentaire et multi-omique

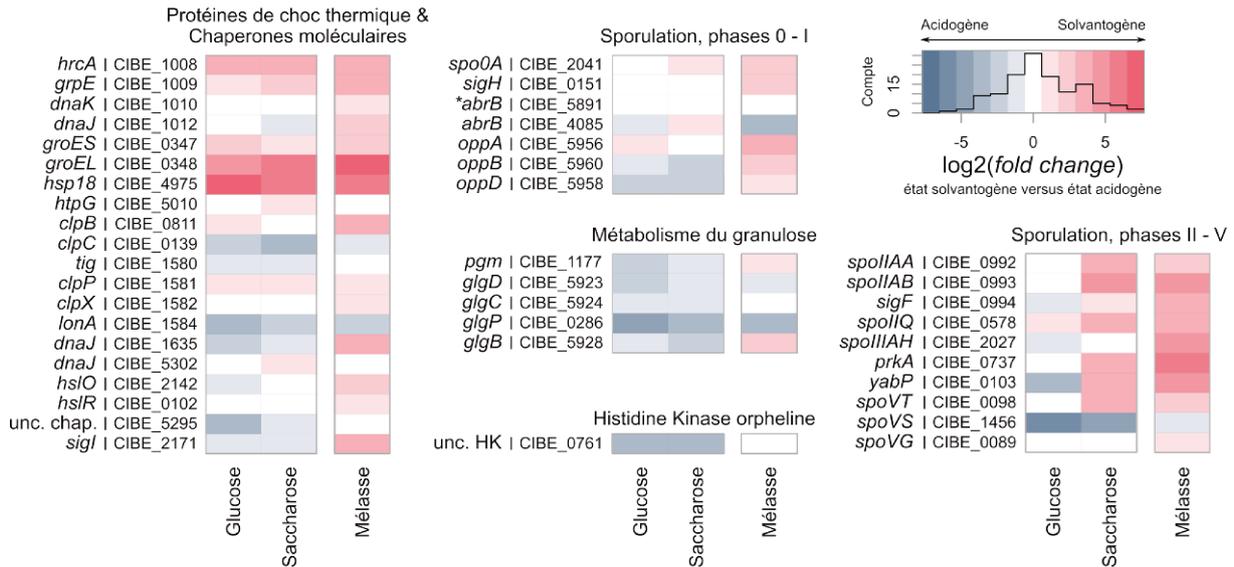


Figure 32 : Comparaison acidogène / solvantogène, *heatmaps* des catégories « Choc thermique » et « Sporulation ».

Pour les gènes dupliqués, l'astérisque * indique la/les protéine(s) majoritaire(s). unc. chap. : chaperone non caractérisée ; unc. HK : histidine kinase non caractérisée.

6.3.4.3 Comparaison substrat à substrat

Pour l'analyse des données, une deuxième approche consiste à comparer les substrats entre eux. Pour cette étude préliminaire, comme susmentionné, nous n'avons considéré que les échantillons obtenus en phase acidogène. Considérant la nature des sources carbonées, une première comparaison peut être effectuée entre les voies d'import et de métabolisation du carbone (Figure 33).

Les enzymes liées à la glycolyse présentent un profil de régulation substrat-dépendant assez intéressant. Entre substrats modèles, les différences concernent principalement les protéines qui constituent la voie EMP. Sur mélasse, la présence des protéines intervenant dans la glycolyse est, comme pour beaucoup d'autres voies (Figure Supplémentaire 9), fortement diminuée, à l'exception particulièrement notable d'une fructose bisphosphate aldolase alternative codée par le gène CIBE_3600, très peu représentée sur substrats modèles. Ces différences pourraient aboutir dans les conditions explorées à l'accumulation de certains intermédiaires de la glycolyse, avec hypothétiquement un effet sur la RCC^{283,284} ou d'autres voies de signalisation²⁸⁵.

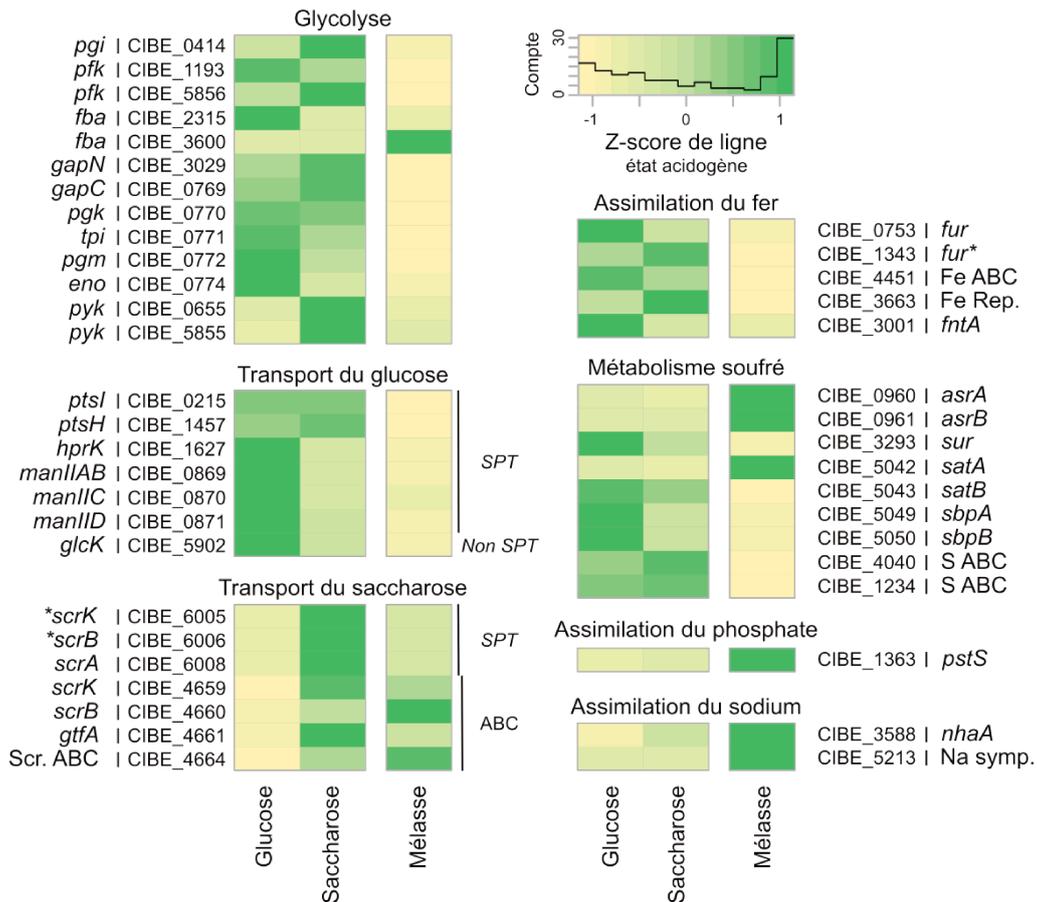


Figure 33 : Comparaison substrat à substrat, **heatmaps des catégories « Transport et métabolisme du carbone » et « Transport et métabolisme des molécules inorganiques »**. Pour les gènes dupliqués, l'astérisque * indique la/les protéine(s) majoritaire(s). Scr. ABC : sous-unité du système de transport ABC pour le sucrose. Fe ABC : sous-unité du système de transport ABC spécifique du fer. Fe Rep. : Répresseur dépendant du fer. S ABC : sous-unité du système de transport ABC spécifique du soufre. Na symp. : symporteur de sodium.

L'effet du substrat sur les voies de transport du glucose et du saccharose est également clair. La présence de saccharose (condition saccharose ou mélasse) inhibe fortement les voies d'import du glucose (par SPT : gènes CIBE_0869 - CIBE_0870 - CIBE_0871 ; une voie non SPT, non caractérisée dans cette étude, doit également exister de par la présence d'une glucokinase CIBE_5902). A l'inverse, la présence de glucose résulte en une diminution de la quantité de protéines prédites comme étant liées à l'import de saccharose. Les gènes codant ces dernières sont répartis en deux sous-catégories. La première correspond à un opéron (CIBE_6005- CIBE_6008) permettant l'assimilation du saccharose par un SPT, avec des gènes homologues à ceux décrits chez *C. beijerinckii* NCIMB 8052¹³⁵. Le gène *scrA* encode une sous-unité SPT EIIBC spécifique du saccharose, qui permet d'importer et de phosphoryler simultanément le sucre. Le gène *scrB* code quant à lui une saccharose-6P hydrolase,

tandis que le produit de *scrK* est une fructokinase permettant de catalyser la phosphorylation du fructose obtenu après clivage du saccharose-6P. Après analyse du protéome de la souche, une deuxième voie semble également importante pour le transport de saccharose. Ainsi, le gène *gffA* (CIBE_4661) encode une saccharose phosphorylase qui peut cliver et phosphoryler simultanément une molécule de glucose en présence d'ATP. La présence de cette enzyme suggère la présence d'une voie indépendante d'un SPT, qui ne fait entrer le saccharose dans la cellule que sous une forme phosphorylée. Les gènes suivants (CIBE_4662-64) sont prédits comme codant les sous-unités d'un transporteur ABC spécifique lié aux carbohydrates. Le produit de CIBE_4664 est détecté par la MS et fortement induit sur saccharose et mélasse. Par ailleurs, les gènes précédant *gffA*, CIBE_4659 et CIBE_4660 sont respectivement des paralogues de *scrK* et *scrB*. Une voie ABC (codée par l'opéron CIBE_4659-CIBE_4664) est donc probablement impliquée dans le transport du saccharose. En présence de mélasse, les produits des gènes de base du SPT sont également sous-représentés (Ptsi, Ptsh, Hprk). Dans ces conditions, l'import de saccharose par une voie SPT est défavorisé, ce qui se traduit dans le protéome par une présence plus importante des composants du système d'import ABC (produits des gènes CIBE_4659- CIBE_4664) aux dépens du système SPT (produits des gènes CIBE_6005 - CIBE_6008).

Une autre différence compositionnelle d'importance entre les substrats modèles et industriel consiste en la présence importante de sels minéraux (10% du poids d'une mélasse, avant dilution dans le milieu de culture). Nous nous sommes donc intéressés aux voies de transport et de métabolisme de différents sels inorganiques (Figure 33). La composition du milieu semble en effet impacter l'assimilation de ces sels. L'assimilation de fer semble réprimée sur mélasse de betterave, alors que celles de phosphate et de sodium sont à l'inverse induites. Ces régulations antagonistes sont difficiles à expliquer, mais reflètent probablement les besoins différentiels de la bactérie sur ce type de milieu. Pour le soufre, la majeure partie des voies d'assimilation sont réprimées, et l'opéron putatif CIBE_0960-0961 encodant les sous-unités AsrA et AsrB d'une sulfite réductase sont fortement surexprimés. Cette enzyme pourrait participer à un processus de détoxification du milieu riche en sulfites (130 mg/kg dans la mélasse avant dilution, Figure Supplémentaire 7), en transformant une partie du soufre minéral en soufre gazeux (H₂S).

6.4 DISCUSSION

Pour l'industrialisation d'un procédé AIBE, et en particulier dans l'optique de produire de l'isopropanol biosourcé, des progrès significatifs doivent encore être faits pour l'amélioration du comportement fermentaire de la souche DSM 6423. Ceci passe par une meilleure compréhension de son fonctionnement en général et particulièrement dans des conditions proches de l'industrie. Dans cette partie de la thèse, nous nous sommes intéressés à l'impact du substrat carboné sur sa physiologie et son métabolisme. Pour ce faire, nous avons tout d'abord testé les capacités du microorganisme à fermenter une sélection limitée de substrats utilisables industriellement. La souche a démontré sa capacité à fermenter chacun d'entre eux, ce qui confirme son potentiel pour une application industrielle^{50,277}. A l'issue de cette étape, nous avons sélectionné la mélasse de betterave notamment pour des raisons de prix et de compatibilité avec un procédé de fermentation continue.

La fermentation continue est un second paramètre industriel d'importance pour un procédé AIBE, puisqu'elle permet de diminuer l'impact de la toxicité des produits tout en limitant les opérations de vidange / nettoyage / remplissage du bioréacteur. Dans le cadre d'une étude omique, la fermentation continue est également avantageuse car elle permet, idéalement²⁸¹ :

- de stabiliser le métabolisme branché de la souche dans une configuration particulière (cf. chez *C. acetobutylicum*^{80,174}). Une séparation en fermentation continue des métabolismes acidogène / solvantogène n'a, à notre connaissance, jamais été rapportée pour une souche de *C. beijerinckii*.
- de garder le microorganisme dans un environnement stable, c'est-à-dire un milieu dont la composition varie peu au cours du temps, et donc peu susceptible de faire varier les données omiques.
- en imposant un taux de croissance constant, de limiter les changements d'expression de gènes liés à la croissance.
- de s'affranchir de la persistance des transcrits et protéines en cours de fermentation. Comme les cellules doivent toujours se diviser et donc former de nouveaux ARNm / protéines, le renouvellement entre les phases acidogène et solvantogène, qui normalement se suivent en condition *batch*, peut être complet.

En d'autres mots, en absence d'information sur la stabilité de ces molécules, la fermentation continue permet de s'assurer que l'expression des gènes mesurée ne provient pas d'une autre phase de croissance.

- dans le cas des fermentations de solvantogènes, de découpler les processus de solvantogénèse et de sporulation, puisque les spores sont graduellement éliminées en cours de procédé au profit de cellules en division active.

Dans nos conditions expérimentales, le basculement physiologique entre phases acidogène et solvantogène repose sur la manipulation de différents paramètres (pH, présence d'acétate, taux de dilution) et permet de définir les deux métabolismes de façon claire. Un procédé fermentaire indépendant du taux de dilution et donc du taux de croissance était désirable pour une des raisons susmentionnées, mais n'a pas pu être établi. Il s'agit d'une limitation de notre procédé, puisque les données comparatives acidogène / solvantogène vont prendre en compte des gènes liés partiellement ou totalement au taux de croissance. Pour la comparaison substrat à substrat, cette limitation ne s'applique toutefois pas.

Comme attendu, l'analyse préliminaire des données protéomiques révèle des tendances globales fortement dépendantes de la source carbonée. L'effet principal de la mélasse de betterave semble être, en phase acidogène, de provoquer une inhibition significative de l'expression de gènes liés à la croissance, visible sur de nombreux points. En solvantogénèse, en raison d'un taux de dilution 4 fois plus faible, la souche a plus de temps pour détoxifier et homogénéiser le milieu, ce qui pourrait expliquer le rapprochement des conditions solvantogènes entre elles. Alternativement, la différence entre substrats modèles et industriel pourrait être liée à un biais technique, possiblement liée à l'étape de normalisation des données LC-MS/MS. Les données de transcriptomique, une fois qu'elles auront été obtenues, seront importantes pour valider ou invalider le processus de traitement de données de protéomique.

En attendant la transcriptomique, l'étude protéomique a permis de révéler pour les deux conditions modèles différentes structures et régulations génétiques d'intérêt. L'opéron saccharose (CIBE_6005-08), identifié précédemment chez *C. beijerinckii* NCIMB 8052 (Cbei_5009-12)¹³⁵, semble également important pour la souche DSM 6423. Dans le modèle proposé pour le transport et la métabolisation du saccharose dans la souche NCIMB 8052, le sucre est importé et phosphorylé directement par un SPT (ScrA), puis le saccharose-6P est clivé par ScrB en glucose-6P et en fructose.

Pour les auteurs, l'activité fructokinase de ScrK permet d'obtenir du fructose-6P qui rejoint la voie EMP, tout comme le glucose-6P. Toutefois, nos données montrent que la condition saccharose provoque la surexpression d'une 1-phosphofruktokinase (Fruk, codée par CIBE_2186, Figure Supplémentaire 9) avec un FC > 12 par rapport à la condition glucose. Il est donc possible que la fructokinase *scrK* produise du fructose-1P, pris en charge ensuite par FruK avant de rejoindre la glycolyse. Il est également possible que les voies de consommation du saccharose et du fructose (opéron fructose putatif : CIBE_2185-87) soient toutes deux induites par la présence de saccharose.

Par ailleurs, l'analyse de nos données révèle une différence significative avec la souche NCIMB 8052. Si l'inactivation de l'opéron saccharose provoque chez cette dernière une auxotrophie pour le saccharose¹³⁴, il est peu probable que cela survienne chez *C. beijerinckii* DSM 6423. En effet, cette dernière semble disposer d'un îlot génétique absent de la souche NCIMB 8052, dont l'expression est fortement induite sur saccharose / mélasse et qui contient des gènes probablement impliqués dans l'import et la métabolisation du saccharose via un transporteur ABC (CIBE_4659-4664, incluant potentiellement également CIBE_4665 qui code un régulateur transcriptionnel). Cette voie est d'autant plus intéressante qu'il semblerait que la consommation du saccharose en milieu modèle et industriel implique différemment les deux structures métaboliques.

Les régulations mises en place entre acidogénèse et solvantogénèse sont, pour les milieux modèles, similaires à celles dévoilées par Máté de Gérando et collègues⁵² en *batch* (induction des voies solvantogènes et des chaperones, répression des gènes liés à la motilité, notamment). Quelques différences importantes doivent toutefois être soulignées. Les gènes liés à la sporulation ne sont globalement pas induits sur glucose en solvantogénèse. Ce résultat est attendu puisque, dans nos conditions, la solvantogénèse n'est pas un état stationnaire et rejoint des observations précédentes faites chez *C. acetobutylicum*²⁸¹. En revanche, cela ne semble pas être le cas sur saccharose, ce qui suggère que les voies de signalisation liées aux sucres (comme par exemple la RCC) pourraient être corrélées chez *C. beijerinckii* à certains gènes contrôlant le processus de sporulation.

L'augmentation massive du nombre de protéines ribosomales en phase solvantogène est également remarquable. Le taux de croissance fixé par

l'expérimentation est 4 fois inférieur en phase solvantogène par rapport à la phase acidogène, ce qui contraste avec une telle augmentation. Après une inspection manuelle, il semblerait que si l'abondance relative en protéines ribosomales augmente bel et bien, celle des protéines liées à la transcription (sous-unités de l'ARN polymérase notamment) et des facteurs de traduction (impliqués dans son initiation, son élongation et sa terminaison) diminue (Figure Supplémentaire 9). Cela suggère, malgré la hausse probable du nombre de ribosomes, une baisse de l'activité traductionnelle, et la présence d'une proportion de ribosomes inactifs plus importante. La surreprésentation de Hpf (*hibernation promoting factor*, responsable de la dimérisation et inactivation des ribosomes²⁹⁰) et d'une protéine inhibitrice de la traduction (CIBE_v1_3881, cf. Figure Supplémentaire 9) conforte cette hypothèse.

Les deux alcool déshydrogénases majeures (CIBE_2050, CIBE_2622), en conditions modèles, présentent également un profil de régulation antagoniste intéressant, et différent de ce qui avait été rapporté précédemment. CIBE_2050 est largement présente en phase acidogène, et est sous-exprimée en phase solvantogène. A l'inverse, CIBE_2622 est très peu exprimée en phase acidogène, et largement induite en phase solvantogène, pour atteindre un niveau d'expression similaire. Cela suggère que, si la production résiduelle d'alcools en phase acidogène est quasi-exclusivement liée à CIBE_2050, les deux enzymes se partagent équitablement la tâche en phase solvantogène. Cette régulation différentielle est intéressante puisque, comme montré dans la Partie II, les deux gènes codants ces enzymes sont contrôlés par σ^{54} , ce qui suggère que deux voies de régulation séparées, liées à différents EBP, coexistent, ou bien que des facteurs post-transcriptionnels interviennent.

Ces deux enzymes, tout comme s-Adh, sont prédites comme dépendantes du NADPH, ce qui suggère qu'en phase solvantogène une partie importante du pouvoir réducteur doit être fournie sous la forme de NADPH à la cellule pour pouvoir fonctionner correctement. Pour régénérer le NADPH, plusieurs moyens peuvent être employés. La glycéraldéhyde-3P déshydrogénase alternative GapN permet notamment d'en produire dans le cadre de la glycolyse, mais son abondance relative vis-à-vis de la glycéraldéhyde-3P déshydrogénase canonique GapC (Figure Supplémentaire 9) suggère que son implication est mineure dans ce processus. Alternativement, la ferrédoxine réduite, obtenue principalement pendant l'étape de

décarboxylation oxydative du pyruvate, peut être utilisée pour produire du H₂ via une hydrogénase ou du NAD(P)H par l'action de NAD(P)H-ferrédoxine oxydoréductase. L'identification de trois hétérodimères SudAB, probablement responsables de l'activité NADPH-ferrédoxine oxydoréductase, est importante car elle permet de faire le lien entre la production d'alcool et la répartition du pouvoir réducteur dans la cellule. Pour compléter ces résultats, il conviendrait de vérifier la spécificité pour le NADH/NADPH de chacune des enzymes SudAB et des différentes alcool déshydrogénases, et également de déterminer cette spécificité pour l'aldéhyde déshydrogénase majeure, CIBE_4606.

En conclusion, cette partie de la thèse a généré des données importantes pour affiner notre compréhension des mécanismes de régulation acidogène / solvantogène chez *C. beijerinckii* DSM 6423 et pour comprendre l'effet de substrats carbonés modèles et industriel sur le métabolisme de la souche. Les données RNA-seq et eCAPP-seq devraient permettre de compléter ces informations pour, à terme, dresser un tableau détaillé du métabolisme A/IBE dans nos conditions expérimentales.

6.5 FIGURES SUPPLEMENTAIRES



Département Industrie
27-29, rue Chateaubriand - 75008 Paris, France
Tel. : +33 (0)1 42 99 00 40 - Fax : +33 (0)1 45 63 39 21
lm@melasse.net

Département Nutrition Animale
11 bis, Boulevard Robert Schuman - CS 80467 - 62020 Arras Cedex, France
Tel. : +33 (0)3 21 23 34 13 - Fax : +33 (0)3 21 23 90 00
lmna@melasse.net

RESULTAT D'ANALYSE

N° 60005063

IFP ENERGIES NOUVELLES
92852 RUEIL-MALMAISON

Produit : MELASSE DE BETTERAVE

N° Lot :

Transport :

Date livraison : 05/03/2018

Origine/Destination:

Dépôt : LABORATOIRE

Emplacement :

Tonnage : 0,000

Date d'analyse : 21/03/2018

Edité le : 23/03/2018

Référence : Mélasse de betterave FI501S0711

Commentaire :

ANALYSES	RESULTATS	UNITES
BRIX C (DOUBLE DILUTION)	75,200	°
CENDRES BRUTES (550 °C)	8,800	%
CHLORURES	0,400	%
DENSITE à 20° C	1,360	
ETHANOL	0,011	%
HUMIDITE FISCHER	28,100	%
POTASSIUM	3,600	%
POTASSIUM EXPRIME EN K2O	4,300	%
SUCRES TOTAUX (LUFF-SCHOORL) EXP. EN SACCHAROSE	45,900	%
MATIERE ORGANIQUE	62,200	%
MATIERES AZOTEES TOTALES (KJELDAHL VIA DUMAS)	8,900	%
MANGANESE	7,800	mg/kg
AZOTE TOTAL	1,400	%

Visa Labo

Visa Siège

Lefebvre

Résultats donnés à titre confidentiel



Part of the United Molasses Group Limited

FRANCE MELASSES - S.A. au capital de 2 481 488 € - www.francemelasses.net - www.umgroup.com
Siège social : 27-29, rue Chateaubriand - 75008 PARIS - SIREN : 591 920 251 - RCS PARIS - TVA N° FR01 591 920 251

RESULTAT D'ANALYSE

N° 60005063

IFP ENERGIES NOUVELLES
92852 RUEIL-MALMAISON

Produit : MELASSE DE BETTERAVE

N° Lot :

Transport :

Date livraison : 05/03/2018

Référence : Mélasse de betterave FI501S0711

Commentaire :

Origine/Destination:

Dépôt : LABORATOIRE

Emplacement :

Tonnage : 0,000

Date d'analyse : 21/03/2018

Edité le : 23/03/2018

ANALYSES	RESULTATS	UNITES
SODIUM	1,300	%
SODIUM EN NA2O	1,800	%
PHOSPHORE	0,020	%
PHOSPHORE EXPRIME EN P2O5	0,010	%
SULFITES EXP. SO2	130,000	mg/kg
VISCOSITE à 20 °C	815,000	cPs

Visa Labo

Lefebvre

Visa Siège

Résultats donnés à titre confidentiel



Part of the United Molasses Group Limited

FRANCE MELASSES - S.A. au capital de 2 481 488 € - www.francemelasses.net - www.umgroup.com
Siège social : 27-29, rue Chateaubriand - 75008 PARIS - SIREN : 591 920 251 - RCS PARIS - TVA N° FR01 591 920 251

Figure Supplémentaire 7 : **Certificat d'analyses réalisé par France Mélasse, donnant la composition de la mélasse de betterave utilisée dans cette étude.**

Partie IV : Métabolisme sur substrat 1G, approches fermentaire et multi-omique

		Biomasse		Alcools (g/L)					Acides (g/L)				Sucres (g/L)				
		Réplicat	OD ₆₀₀	Ethanol	Acétone	Isopropanol	Butanol	Total produit	Lactique (consommé)	Acétique (consommé)	Butyrique	Total produit	Glucose (consommé)	Fructose (consommé)	Saccharose (consommé)	Total consommé	
Glucose	Acidogène	A	2,2	0,0	0,0	0,0	0,0	0,0	-1,9	-2,0	4,7	8,6	17,8	/	/	17,8	
		B	3,3	0,0	0,0	0,0	0,0	0,0	-2,0	-2,4	5,6	10,1	20,0	/	/	20,0	
		C	3,6	0,0	0,0	0,0	0,0	0,0	0,0	-2,3	-2,6	6,3	11,2	21,9	/	/	21,9
		D	3,2	0,0	0,0	0,0	0,0	0,0	0,0	-3,5	-2,4	5,5	11,4	22,1	/	/	22,1
	Solvantogène	A	6,6	0,0	0,0	2,5	4,9	7,4	0,0	1,0	1,0	1,0	29,9	/	/	29,9	
		B	6,4	0,0	0,0	3,4	5,5	9,0	0,0	0,9	1,1	1,1	30,8	/	/	30,8	
		C	6,4	0,0	0,1	3,8	6,0	9,9	0,0	0,9	1,3	1,3	32,9	/	/	32,9	
		D	7,0	0,0	0,1	4,4	6,7	11,2	0,0	1,0	0,8	0,8	34,1	/	/	34,1	
Saccharose	Acidogène	A	3,1	0,1	0,0	0,0	0,0	0,1	-2,0	-1,3	3,1	6,3	-0,7	0,7	14,3	14,3	
		B	3,2	0,1	0,0	0,0	0,0	0,1	-2,4	-1,1	3,2	6,7	-0,2	1,4	14,1	15,2	
		C	3,3	0,1	0,0	0,0	0,0	0,1	-2,8	-1,5	4,5	8,8	1,8	1,4	16,7	19,9	
		D	3,0	0,1	0,0	0,0	0,0	0,1	-2,4	-1,6	4,4	8,5	-0,1	1,4	17,0	18,3	
	Solvantogène	A	7,5	0,1	0,1	3,8	5,7	9,6	0,0	1,4	0,6	0,6	1,6	1,4	28,8	31,8	
		B	7,6	0,1	0,0	3,6	5,0	8,8	0,0	1,2	1,1	1,1	1,6	1,4	32,4	35,4	
		C	6,7	0,1	0,1	3,8	5,0	9,0	0,0	1,2	0,8	0,8	1,6	1,4	31,6	34,6	
		D	5,8	0,1	0,1	3,8	5,3	9,2	0,0	1,1	1,1	1,1	1,6	1,4	32,3	35,3	
Mélasse Betterave	Acidogène	A	4,1	0,0	0,0	0,0	0,0	0,0	-5,8	-0,7	0,9	7,5	0,7	2,2	16,0	18,9	
		B	3,8	0,0	0,0	0,0	0,0	0,0	-6,3	-0,7	0,8	7,8	0,8	2,2	14,2	17,2	
		C	4,3	0,0	0,0	0,0	0,0	0,0	0,0	-6,2	-0,9	1,5	8,7	0,8	2,2	15,6	18,7
		D	4,6	0,0	0,0	0,0	0,0	0,0	0,0	-5,6	-1,1	2,2	9,0	0,7	2,2	15,4	18,3
	Solvantogène	A	5,3	0,1	0,1	2,0	5,4	7,6	2,2	-0,3	1,0	1,3	2,9	3,8	17,6	24,3	
		B	3,4	0,1	0,0	2,1	3,4	5,7	2,5	1,4	0,2	0,2	2,9	1,5	10,0	14,5	
		C	3,4	0,1	0,0	2,1	3,4	5,7	1,8	0,9	0,5	0,5	2,9	1,2	13,6	17,7	
		D	2,3	0,1	0,0	1,1	2,3	3,5	1,7	0,0	1,7	1,7	2,9	1,3	11,9	16,2	

Figure Supplémentaire 8 : Dosage de la biomasse et des métabolites dans les échantillons biologiques utilisés dans cette étude. Les réplicats utilisés pour les analyses omiques sont indiqués en rouge.

Figure Supplémentaire 9 : Catégorisation manuelle de 657 protéines détectées en protéomique.

Catégorie	Sous-catégorie	Gène	Locus tag	Description	Nombre de peptides	Glucose		Saccharose		Mélasse	
						log2(solventogène / acidogène)	Z-score (acidogène)	log2(solventogène / acidogène)	Z-score (acidogène)	log2(solventogène / acidogène)	Z-score (acidogène)
Métabolisme A/IBE	Acidogénèse	<i>ldh</i>	CIBE_v1_1209	ID:29050106 ldh L-lactate dehydrogenase	7	-0,2	0,59	-0,6	0,56	1,4	-1,15
		<i>pta</i>	CIBE_v1_1400	ID:29050297 pta Phosphate acetyltransferase	17	1,0	0,25	0,4	0,85	2,2	-1,10
		<i>ack</i>	CIBE_v1_1401	ID:29050298 ackA acetate kinase	25	-0,4	1,13	1,5	-0,34	1,3	-0,78
		<i>ptb</i>	CIBE_v1_0223	ID:29049120 ptb phosphate butyryl coenzyme A transferase	20	1,4	0,55	1,8	0,60	1,9	-1,15
		<i>buk</i>	CIBE_v1_0224	ID:29049121 buk branched-chain fatty-acid kinase	23	0,0	0,87	-0,6	0,22	2,7	-1,09
	Formation de Butyryl-CoA	<i>thl</i>	CIBE_v1_0565	ID:29049462 mmgA degradative acetoacetyl-CoA thiolase	32	-1,6	0,16	-1,8	0,91	2,4	-1,07
		<i>rex</i>	CIBE_v1_0338	ID:29049235 rex Redox-sensing transcriptional repressor Rex	8	-1,0	0,11	-0,8	0,94	1,1	-1,05
		<i>crt</i>	CIBE_v1_0339	ID:29049236 yngF putative Methylglutaconyl-CoA hydratase	21	-1,4	0,84	-0,7	0,27	-0,1	-1,11
		<i>bcd</i>	CIBE_v1_0340	ID:29049237 acdA acyl-CoA dehydrogenase	22	-1,9	0,92	-1,9	0,14	0,3	-1,06
		<i>etfB</i>	CIBE_v1_0341	ID:29049238 etfB Electron transfer flavoprotein subunit beta	22	-1,7	0,53	-1,2	0,62	-0,6	-1,15
	Solventogénèse	<i>etfA</i>	CIBE_v1_0342	ID:29049239 etfA electron transfer flavoprotein (alpha subunit)	18	-1,4	0,58	-1,2	0,57	0,6	-1,15
		<i>hbd</i>	CIBE_v1_0343	ID:29049240 mmgB 3-hydroxybutyryl-CoA dehydrogenase	23	1,8	-0,35	-0,8	1,13	2,2	-0,78
		<i>ald</i>	CIBE_v1_2964	ID:29051861 Aldehyde dehydrogenase	2	-0,1	-0,02	-0,1	1,01	0,8	-0,99
		<i>ald</i>	CIBE_v1_4606	ID:29053503 eutE Ethanolamine utilization protein EutE	14	1,6	1,15	4,0	-0,57	4,6	-0,58
		<i>ctfA</i>	CIBE_v1_4607	ID:29053504 yodS putative aminoacylate CoA-transferase	6	4,5	1,15	6,4	-0,59	5,8	-0,56
		<i>ctfB</i>	CIBE_v1_4608	ID:29053505 yodR putative acyloate-acetoacetate CoA-transferase	8	3,3	0,72	6,1	-1,14	3,1	0,42
		<i>adc</i>	CIBE_v1_4609	ID:29053506 adcJ Acetoacetate decarboxylase	15	2,2	1,15	5,1	-0,54	5,7	-0,61
		<i>fnrA</i>	CIBE_v1_3468	ID:29052365 sudA Sulfide dehydrogenase subunit alpha	1	1,1	1,15	5,4	-0,55	7,5	-0,61
		<i>fnrB</i>	CIBE_v1_3469	ID:29052366 conserved protein of unknown function	1	0,3	1,15	3,3	-0,49	3,9	-0,66
		<i>sadh</i>	CIBE_v1_3470	ID:29052367 adh NADP-dependent isopropanol dehydrogenase	7	1,8	1,15	3,1	-0,52	5,8	-0,63
Solventogénèse	<i>adh</i>	CIBE_v1_2050	ID:29050947 adh NADPH-dependent butanol dehydrogenase	23	-1,9	0,89	-0,7	0,19	2,2	-1,08	
	<i>adh</i>	CIBE_v1_2349	ID:29051246 Iron-containing alcohol dehydrogenase	5	3,0	-0,64	2,4	1,15	3,9	-0,52	
	<i>adh</i>	CIBE_v1_2622	ID:29051519 adh NADPH-dependent butanol dehydrogenase	12	3,2	-0,96	2,6	1,04	3,0	-0,08	
	<i>fnrB</i>	CIBE_v1_2623	ID:29051520 conserved protein of unknown function	9	2,8	-1,10	1,0	0,85	2,0	0,25	
	<i>fnrA</i>	CIBE_v1_2624	ID:29051521 sudA Sulfide dehydrogenase subunit alpha	3	3,2	-1,14	1,1	0,44	0,6	0,70	
	<i>2,3-bdh</i>	CIBE_v1_1696	ID:29050593 bdhA acetoin reductase/2,3-butanediol dehydrogenase	4	3,3	-0,53	4,4	-0,63	-1,1	1,15	
	<i>fnrB</i>	CIBE_v1_0830	ID:29049727 conserved protein of unknown function	15	-0,7	1,11	1,1	-0,83	0,1	-0,29	
	<i>fnrA</i>	CIBE_v1_0831	ID:29049728 sudA Sulfide dehydrogenase subunit alpha	19	-0,7	0,97	0,1	0,07	1,9	-1,03	

Partie IV : Métabolisme sur substrat 1G, approches fermentaire et multi-omique

Métabolisme A/IBE	Transport d'électron Rnf	<i>mfc</i>	CIBE_v1_2890	ID:29051787 rnfC Electron transport complex subunit RnfC	5	-2,9	1,15	-0,5	-0,55	2,1	-0,60	
		<i>mfd</i>	CIBE_v1_2891	ID:29051788 Electron transport complex, RnfABCDGE type, D subunit	2	-2,2	1,15	0,0	-0,49	5,6	-0,66	
		<i>mfg</i>	CIBE_v1_2892	ID:29051789 Electron transport complex, RnfABCDGE type, G subunit	7	0,2	1,15	2,7	-0,48	7,1	-0,67	
	Assemblage de cluster FeS	<i>mfb</i>	CIBE_v1_2895	ID:29051792 Protein RnfB	1	-0,9	1,09	0,6	-0,21	3,4	-0,88	
		<i>sufB</i>	CIBE_v1_2192	ID:29051089 sufB FeS cluster formation protein	1	-1,1	1,15	-0,5	-0,50	3,1	-0,65	
		<i>sufD</i>	CIBE_v1_2193	ID:29051090 FeS assembly protein SufD	2	-0,6	0,85	2,4	0,26	4,8	-1,10	
	Formation d'hydrogène	<i>hyd</i>	CIBE_v1_2114	ID:29051011 Hydrogenase, Fe-only	17	-1,7	0,36	-2,3	0,77	-0,1	-1,13	
		<i>hyd</i>	CIBE_v1_4547	ID:29053444 hndD NADP-reducing hydrogenase subunit HndC	2	1,5	-0,60	1,8	-0,56	-3,0	1,15	
		<i>hyd</i>	CIBE_v1_0346	ID:29049243 Ferredoxin hydrogenase	12	-4,3	0,84	-3,6	0,27	1,2	-1,11	
		<i>fdx</i>	CIBE_v1_4762	ID:29053659 Ferredoxin	1	1,6	0,77	1,6	0,36	2,9	-1,13	
		<i>fdx</i>	CIBE_v1_0133	ID:29049030 Metal-binding protein of ferredoxin fold	3	0,4	0,29	0,2	0,82	1,6	-1,11	
	Croissance mixotrophique	Voie de Wood-Ljungdahl	<i>codh</i>	CIBE_v1_6073	ID:29054970 cooS Carbon monoxide dehydrogenase	1	3,3	-1,12	1,3	0,79	4,6	0,33
			<i>fdh</i>	CIBE_v1_4552	ID:29053449 fdhF Formate dehydrogenase H	4	0,0	-0,56	3,3	-0,59	-3,2	1,15
			<i>fhs</i>	CIBE_v1_0111	ID:29049008 fhs Formate-tetrahydrofolate ligase	7	-0,1	0,89	0,0	0,19	2,7	-1,08
			<i>folD</i>	CIBE_v1_2030	ID:29050927 folD methylenetetrahydrofolate dehydrogenase; methenyltetrahydrofolate cyclohydrolase	4	-2,8	0,97	-2,2	0,05	-1,4	-1,03
<i>pflB</i>			CIBE_v1_1204	ID:29050101 pflB pyruvate formate lyase I	3	-2,6	0,34	-2,2	0,78	-0,1	-1,13	
<i>pflA</i>			CIBE_v1_1205	ID:29050102 pflA pyruvate formate lyase activating enzyme 1	9	-1,6	1,05	-0,3	-0,10	0,7	-0,95	
<i>pflB</i>			CIBE_v1_1206	ID:29050103 pflB pyruvate formate lyase I	19	-1,5	0,61	0,1	0,55	-0,7	-1,15	
<i>pfor</i>			CIBE_v1_5186	ID:29054083 nifJ Pyruvate-flavodoxin oxidoreductase	58	-0,8	0,77	-1,0	0,36	0,8	-1,13	
Synthèse de carbonate		<i>cdh</i>	CIBE_v1_5326	ID:29054223 yvdA putative carbonic anhydrase	8	1,9	0,11	1,7	0,94	0,8	-1,05	
Assimilation du carbone	Glycolyse	<i>pgi</i>	CIBE_v1_0414	ID:29049311 pgii glucose-6-phosphate isomerase	33	-1,2	-0,09	-2,2	1,04	0,1	-0,95	
		<i>pfk</i>	CIBE_v1_1193	ID:29050090 pfkA 6-phosphofructokinase 2	6	-0,7	0,89	0,2	0,19	1,2	-1,08	
		<i>pfk</i>	CIBE_v1_5856	ID:29054753 pfkA 6-phosphofructokinase	16	-1,7	0,04	-3,0	0,98	-0,3	-1,02	
		<i>fba</i>	CIBE_v1_2315	ID:29051212 fbaA fructose-1,6-bisphosphate aldolase	20	-2,7	1,15	-1,6	-0,52	0,2	-0,63	
		<i>fba</i>	CIBE_v1_3600	ID:29052497 fba Fructose-bisphosphate aldolase class 1	5	1,9	-0,58	-0,1	-0,58	-8,6	1,15	
		<i>gapN</i>	CIBE_v1_3029	ID:29051926 gapN NADP-dependent glyceraldehyde-3-phosphate dehydrogenase	8	0,4	0,11	0,3	0,94	1,4	-1,05	

		<i>gapC</i>	CIBE_v1_0769	ID:29049666 gapA glyceraldehyde-3-phosphate dehydrogenase	27	-0,1	0,33	-0,9	0,80	0,5	-1,12	
		<i>pgk</i>	CIBE_v1_0770	ID:29049667 pgk phosphoglycerate kinase	29	-1,2	0,68	-0,7	0,47	0,0	-1,15	
		<i>tpi</i>	CIBE_v1_0771	ID:29049668 tpiA triose phosphate isomerase	19	0,9	0,89	0,1	0,20	4,2	-1,08	
		<i>pgm</i>	CIBE_v1_0772	ID:29049669 pgm phosphoglycerate mutase	30	-1,2	1,00	-0,4	0,01	1,7	-1,00	
		<i>eno</i>	CIBE_v1_0774	ID:29049671 eno enolase	24	2,0	1,11	4,0	-0,28	3,8	-0,83	
		<i>pyk</i>	CIBE_v1_0655	ID:29049552 pykF Pyruvate kinase	5	0,2	-0,46	0,4	1,15	0,3	-0,68	
		<i>pyk</i>	CIBE_v1_5855	ID:29054752 pyk pyruvate kinase	23	-0,6	-0,67	-1,2	1,15	-0,3	-0,48	
	Métabolisme du glucose via SPT	<i>ptsl</i>	CIBE_v1_0215	ID:29049112 ptsl phosphotransferase system (PTS) enzyme I	23	-1,4	0,60	-1,3	0,56	1,1	-1,15	
		<i>ptsh</i>	CIBE_v1_1457	ID:29050354 ptsh histidine-containing phosphocarrier protein of the phosphotransferase system (PTS) (HPr protein)	6	1,7	0,33	0,5	0,79	3,0	-1,12	
		<i>hprK</i>	CIBE_v1_1627	ID:29050524 hprK HPr kinase/phosphorylase	2	-1,7	1,11	-0,9	-0,29	-0,3	-0,82	
		<i>manIIAB</i>	CIBE_v1_0869	ID:29049766 manX fused mannose-specific PTS enzymes: IIA component ; IIB component	20	-2,8	1,12	-1,1	-0,30	1,1	-0,81	
		<i>manIIC</i>	CIBE_v1_0870	ID:29049767 levF phosphotransferase system (PTS) fructose-specific enzyme IIC component	2	-1,6	1,14	0,0	-0,43	4,6	-0,71	
		<i>manIID</i>	CIBE_v1_0871	ID:29049768 levG phosphotransferase system (PTS) fructose-specific enzyme IID component	13	-1,1	1,08	-0,8	-0,20	2,3	-0,89	
Assimilation du carbone	Métabolisme du glucose via non SPT	<i>glcK</i>	CIBE_v1_5902	ID:29054799 Glucokinase	9	-1,1	1,10	-0,6	-0,24	2,0	-0,86	
		Métabolisme du saccharose via SPT	<i>scrK</i>	CIBE_v1_6005	ID:29054902 ydjE Uncharacterized sugar kinase YdjE	9	-1,1	-0,75	-0,7	1,14	0,2	-0,39
			<i>scrB</i>	CIBE_v1_6006	ID:29054903 Sucrose-6-phosphate hydrolase	3	-2,8	-0,71	-4,4	1,14	-1,6	-0,43
			<i>scrA</i>	CIBE_v1_6008	ID:29054905 scrA PTS system sucrose-specific EIIBC component [Includes: Sucrose-specific phosphotransferase enzyme IIB component ; Sucrose permease IIC component]	3	-2,3	-0,79	-2,0	1,12	0,0	-0,33
		Métabolisme du saccharose via ABC	<i>scrK</i>	CIBE_v1_4659	ID:29053556 ydjE Uncharacterized sugar kinase YdjE	6	-2,5	-1,04	-0,9	0,95	-0,8	0,09
			<i>scrB</i>	CIBE_v1_4660	ID:29053557 Sucrose-6-phosphate hydrolase	4	-5,2	-0,96	-1,7	-0,07	-3,6	1,03
			<i>gtfA</i>	CIBE_v1_4661	ID:29053558 gtfA Sucrose phosphorylase	3	-5,0	-0,87	-4,6	1,09	-2,5	-0,22
			X	CIBE_v1_4664	ID:29053561 ABC-type sugar transport system, periplasmic component	13	-1,6	-1,09	-2,0	0,20	0,9	0,88
		Métabolisme du fructose via SPT	<i>fruR</i>	CIBE_v1_2185	ID:29051082 fruR transcriptional regulator (DeoR family)	3	-1,4	-0,51	-0,9	1,15	1,9	-0,65
			<i>fruK</i>	CIBE_v1_2186	ID:29051083 fruK fructose-1-phosphate kinase	14	-0,5	-0,67	-2,5	1,15	1,2	-0,47
			<i>fruA</i>	CIBE_v1_2187	ID:29051084 fruA phosphotransferase system (PTS) fructose-specific enzyme IABC component	6	-1,6	-0,71	-3,4	1,14	-0,5	-0,43

Partie IV : Métabolisme sur substrat 1G, approches fermentaire et multi-omique

	<i>hrcA</i>	CIBE_v1_1008	ID:29049905 hrcA Heat-inducible transcription repressor HrcA	2	3,1	-1,08	3,7	0,90	3,5	0,18
	<i>grpE</i>	CIBE_v1_1009	ID:29049906 grpE Protein GrpE	11	1,6	0,93	2,4	0,13	3,0	-1,06
	<i>dnaK</i>	CIBE_v1_1010	ID:29049907 dnaK molecular chaperone	33	0,2	1,11	-0,1	-0,29	1,6	-0,83
	<i>dnaJ</i>	CIBE_v1_1012	ID:29049909 dnaJ co-factor of molecular chaperone	10	0,1	0,01	-0,8	0,99	2,9	-1,01
	<i>groES</i>	CIBE_v1_0347	ID:29049244 groES chaperonin small subunit	6	2,9	-0,87	1,6	-0,22	1,9	1,09
	<i>groEL</i>	CIBE_v1_0348	ID:29049245 groEL chaperonin large subunit	78	5,1	0,97	5,8	0,07	7,7	-1,03
	<i>htpG</i>	CIBE_v1_5010	ID:29053907 htpG Chaperone protein HtpG	12	-0,3	-0,08	1,1	-0,96	0,1	1,04
	<i>hsp18</i>	CIBE_v1_4975	ID:29053872 hsp1 18 kDa heat shock protein	6	6,7	-0,73	5,4	-0,41	5,5	1,14
	<i>clpB</i>	CIBE_v1_0811	ID:29049708 clpB protein disaggregation chaperone	8	1,1	-0,11	0,3	1,05	4,0	-0,94
	<i>clpC</i>	CIBE_v1_0139	ID:29049036 clpC class III stress response-related ATPase, AAA+ superfamily	12	-2,0	-0,48	-3,9	1,15	-1,0	-0,67
	<i>clpP</i>	CIBE_v1_1581	ID:29050478 clpP ATP-dependent Clp protease proteolytic subunit	6	0,7	0,26	1,0	0,84	0,7	-1,10
	<i>clpX</i>	CIBE_v1_1582	ID:29050479 clpX protein unfolding ATPase required for presentation of proteins to proteases	14	-0,2	0,55	-0,3	0,61	1,0	-1,15
	<i>tig</i>	CIBE_v1_1580	ID:29050477 tig prolyl isomerase (trigger factor)	28	-0,7	-0,41	-1,5	1,14	0,3	-0,73
Réponse au stress	<i>dnaJ</i>	CIBE_v1_1635	ID:29050532 Heat shock protein DnaJ domain protein	2	-2,4	0,69	-1,2	0,46	3,0	-1,15
	<i>dnaJ</i>	CIBE_v1_5302	ID:29054199 Heat shock protein DnaJ domain protein	3	0,0	0,31	0,6	0,81	0,1	-1,12
	<i>hslO</i>	CIBE_v1_2142	ID:29051039 hslO 33 kDa chaperonin	6	-1,6	0,87	0,2	0,23	1,8	-1,09
	<i>hslR</i>	CIBE_v1_0102	ID:29048999 hslR ribosomal RNA binding protein involved in 50S recycling; heat shock protein	3	-0,1	0,43	0,2	0,71	0,9	-1,14
	<i>X</i>	CIBE_v1_5295	ID:29054192 Molecular chaperone	3	-3,7	1,13	-1,3	-0,35	0,2	-0,78
	<i>sigI</i>	CIBE_v1_2171	ID:29051068 RNA polymerase, sigma 28 subunit, FliA/WhiG family	1	-1,1	0,74	-1,2	0,40	3,2	-1,14
	<i>lonA</i>	CIBE_v1_1584	ID:29050481 lonA class III heat-shock ATP-dependent LonA protease	3	-3,8	1,08	-2,3	-0,89	-2,0	-0,20
	<i>cbeRI</i>	CIBE_v1_3326	ID:29052223 hsdR HsdR	5	-4,1	-0,15	-2,3	-0,92	-6,6	1,07
	<i>cbeMI</i>	CIBE_v1_3328	ID:29052225 Type I restriction-modification system DNA methylase	5	-1,0	0,85	-1,1	0,26	0,1	-1,10
	<i>bla</i>	CIBE_v1_1194	ID:29050091 putative Beta-lactamase	12	-0,6	-0,28	-2,2	1,11	1,1	-0,83
Mécanismes de défense et de virulence	<i>bla</i>	CIBE_v1_2005	ID:29050902 Beta-lactamase domain protein	3	-1,2	0,38	-1,9	0,76	0,9	-1,13
	<i>bla</i>	CIBE_v1_3294	ID:29052191 Beta-lactamase domain protein	8	1,5	0,67	1,6	0,48	4,4	-1,15
	<i>bla</i>	CIBE_v1_4012	ID:29052909 Beta-lactamase domain protein	6	-1,9	1,13	-0,3	-0,35	0,2	-0,78
	<i>bla</i>	CIBE_v1_6075	ID:29054972 Beta-lactamase domain protein	28	1,9	0,91	2,9	0,16	6,2	-1,07
	<i>cloA</i>	CIBE_v2_p0012	ID:29048874 closticin	21	0,2	1,13	2,2	-0,38	3,4	-0,76

	<i>cloB</i>	CIBE_v2_p0013	ID:29048875 putative bacteriocin transport accessory protein	2	-0,6	1,14	1,1	-0,42	3,6	-0,72	
	<i>cloC</i>	CIBE_v2_p0004	ID:29048866 putative peptidase	17	0,4	1,15	3,0	-0,50	4,9	-0,65	
	<i>fldH</i>	CIBE_v1_1739	ID:29050636 fldH Phenyllactate dehydrogenase	2	2,5	-0,73	2,2	-0,41	-0,3	1,14	
	<i>lcn</i>	CIBE_v1_2453	ID:29051350 Lactococcin	6	1,2	1,02	1,8	-0,03	3,0	-0,98	
	<i>lexA</i>	CIBE_v1_3014	ID:29051911 lexA transcriptional repressor of the SOS regulon	1	-1,1	0,67	-0,7	0,48	-1,4	-1,15	
	<i>recA</i>	CIBE_v1_1454	ID:29050351 recA multifunctional SOS repair factor	3	0,1	-1,14	-2,6	0,73	0,4	0,41	
	<i>ruvA</i>	CIBE_v1_1766	ID:29050663 ruvA Holliday junction ATP-dependent DNA helicase RuvA	4	-2,8	1,13	-0,2	-0,35	0,4	-0,78	
Réponse SOS et réparation de l'ADN	<i>ligA</i>	CIBE_v1_0425	ID:29049322 ligA DNA ligase	2	-2,2	-1,15	-1,7	0,59	-3,5	0,56	
	<i>exoA</i>	CIBE_v1_2787	ID:29051684 exoA apurinic/aprimidinic endonuclease	6	-1,5	0,95	-1,1	0,09	-0,7	-1,04	
	<i>lig</i>	CIBE_v1_5255	ID:29054152 DNA ligase III	1	-0,1	-0,35	-0,7	1,13	1,0	-0,78	
	<i>uvrA</i>	CIBE_v1_5868	ID:29054765 uvrA excinuclease ABC (subunit A)	5	-2,8	0,56	-1,6	0,60	-2,4	-1,15	
	<i>recR</i>	CIBE_v1_5967	ID:29054864 recR DNA repair and recombination protein	3	-6,0	1,05	-4,7	-0,11	-3,3	-0,94	
	<i>recQ</i>	CIBE_v1_2168	ID:29051065 ATP-dependent DNA helicase RecQ	1	-1,0	-0,55	0,9	-0,60	-5,4	1,15	
	<i>asp23</i>	CIBE_v1_2028	ID:29050925 conserved protein of unknown function	5	-1,3	0,80	-2,0	0,32	0,9	-1,12	
Réponse au stress	X	CIBE_v1_0529	ID:29049426 2-alkenal reductase	7	2,2	0,28	1,6	0,83	3,8	-1,11	
	<i>nmo</i>	CIBE_v1_0637	ID:29049534 2-nitropropane dioxygenase, NPD	5	-2,1	1,04	-1,0	-0,10	0,3	-0,95	
	<i>ntr</i>	CIBE_v1_1216	ID:29050113 NAD(P)H nitroreductase	4	-1,7	0,45	-2,0	0,70	2,0	-1,15	
	<i>nmo</i>	CIBE_v1_1268	ID:29050165 2-nitropropane dioxygenase, NPD	8	-2,0	0,44	-1,3	0,70	-0,5	-1,15	
	<i>ntr</i>	CIBE_v1_1704	ID:29050601 Nitroreductase family protein	7	0,6	1,14	0,9	-0,43	2,4	-0,71	
	X	CIBE_v1_1730	ID:29050627 Carboxymuconolactone decarboxylase	4	-2,2	1,12	-2,2	-0,32	0,6	-0,80	
	<i>ntr</i>	CIBE_v1_2461	ID:29051358 Nitroreductase	8	3,6	-0,60	4,4	-0,56	1,5	1,15	
	<i>ntr</i>	CIBE_v1_2756	ID:29051653 Nitroreductase	5	-1,6	0,40	-2,4	0,74	1,4	-1,14	
	Détoxification	<i>nfrAB</i>	CIBE_v1_3295	ID:29052192 nfrAB NADPH-FMN oxidoreductase (nitroreductase)	10	-2,0	1,13	0,4	-0,37	2,1	-0,76
		<i>ntr</i>	CIBE_v1_3438	ID:29052335 Nitroreductase	3	-1,6	0,28	-2,6	0,83	0,2	-1,11
<i>ntr</i>		CIBE_v1_3864	ID:29052761 Nitroreductase	2	-2,1	0,71	-1,6	0,44	1,8	-1,14	
<i>ntr</i>		CIBE_v1_4301	ID:29053198 Nitroreductase	2	-1,2	0,98	-0,4	0,04	2,6	-1,02	
<i>ntr</i>		CIBE_v1_4354	ID:29053251 Nitroreductase	3	-3,3	0,89	-3,5	0,19	-1,1	-1,08	
<i>ntr</i>		CIBE_v1_4707	ID:29053604 Nitroreductase	5	-0,7	0,48	-0,6	0,67	1,7	-1,15	
<i>ntr</i>		CIBE_v1_5227	ID:29054124 Nitroreductase	11	0,1	0,70	-0,3	0,45	3,1	-1,15	
<i>ntr</i>		CIBE_v1_5429	ID:29054326 Nitroreductase	2	1,1	-0,54	0,9	-0,61	1,6	1,15	
<i>glo</i>		CIBE_v1_4727	ID:29053624 Glyoxalase	3	2,1	1,08	1,9	-0,19	4,8	-0,89	
<i>glo</i>		CIBE_v1_1498	ID:29050395 Glyoxalase	7	-3,2	0,99	-2,8	0,02	-1,1	-1,01	
	<i>gloA</i>	CIBE_v1_2311	ID:29051208 gloA putative lactoylglutathione lyase	6	0,6	0,61	0,8	0,54	3,2	-1,15	

Partie IV : Métabolisme sur substrat 1G, approches fermentaire et multi-omique

Réponse au stress	Détoxification	<i>glo</i>	CIBE_v1_2593	ID:29051490 Glyoxalase/bleomycin resistance protein/dioxygenase	4	1,1	-0,11	0,7	1,05	2,2	-0,94	
		<i>grx</i>	CIBE_v1_0141	ID:29049038 Uncharacterized glutaredoxin-like 8.6 kDa protein in rubredoxin operon	6	1,7	-0,69	0,9	-0,46	1,1	1,15	
		<i>rbr</i>	CIBE_v1_0289	ID:29049186 rbr Rubrerythrin	9	-2,9	0,00	-3,4	1,00	0,5	-1,00	
		<i>hme</i>	CIBE_v1_0627	ID:29049524 Bacteriohemerythrin	4	1,2	-0,12	0,8	1,05	1,4	-0,94	
		<i>rbr</i>	CIBE_v1_0740	ID:29049637 rbr Rubrerythrin	2	0,7	1,13	6,2	-0,76	5,9	-0,38	
		<i>bcp</i>	CIBE_v1_0942	ID:29049839 bcp putative peroxiredoxin bcp	4	-0,6	1,03	0,5	-0,07	1,0	-0,96	
		<i>trxA</i>	CIBE_v1_0953	ID:29049850 trxA Thioredoxin	7	2,2	-0,95	1,3	1,04	1,7	-0,08	
		<i>trxB</i>	CIBE_v1_0955	ID:29049852 Thioredoxin reductase	7	1,2	-0,57	-0,7	1,15	0,6	-0,59	
		<i>ohrR</i>	CIBE_v1_0956	ID:29049853 ohrR transcriptional regulator sensing organic peroxides	5	-0,5	-0,17	-0,9	1,07	-0,1	-0,90	
		<i>msrA</i>	CIBE_v1_1559	ID:29050456 msrA Peptide methionine sulfoxide reductase MsrA	2	3,8	0,46	2,7	0,69	7,0	-1,15	
		<i>perR</i>	CIBE_v1_1590	ID:29050487 perR Transcriptional regulator PerR	2	-2,6	0,88	-2,2	0,21	0,0	-1,09	
		<i>rbr3A</i>	CIBE_v1_1637	ID:29050534 rbr3A Reverse rubrerythrin-1	3	1,3	-1,04	-1,1	0,95	1,2	0,09	
		<i>sodA</i>	CIBE_v1_2199	ID:29051096 sodA Superoxide dismutase [Mn/Fe]	1	0,6	1,14	3,1	-0,40	5,5	-0,74	
		<i>ahpC</i>	CIBE_v1_2575	ID:29051472 ahpC alkyl hydroperoxide reductase (small subunit)	3	0,5	-0,38	-1,2	1,13	2,5	-0,76	
		Stress oxydatif	<i>hme</i>	CIBE_v1_2761	ID:29051658 Bacteriohemerythrin	5	0,6	0,04	0,2	0,98	3,0	-1,02
			<i>msrB</i>	CIBE_v1_2902	ID:29051799 msrB peptide methionine R-sulfoxide reductase	7	1,2	0,69	1,7	0,45	3,4	-1,15
			<i>rbr3A</i>	CIBE_v1_2934	ID:29051831 rbr3A Reverse rubrerythrin-1	7	0,9	-0,50	-0,7	1,15	2,8	-0,65
			<i>tpx</i>	CIBE_v1_3182	ID:29052079 tpx putative peroxiredoxin	5	3,4	1,07	3,3	-0,17	3,7	-0,90
			<i>trxB</i>	CIBE_v1_3183	ID:29052080 trxB Thioredoxin reductase	10	1,4	0,20	0,5	0,89	2,2	-1,08
			<i>rbr</i>	CIBE_v1_3635	ID:29052532 Rubrerythrin	5	0,7	1,05	0,9	-0,12	1,4	-0,94
			<i>ahpD</i>	CIBE_v1_3886	ID:29052783 Alkylhydroperoxidase AhpD family core domain protein	2	-6,2	1,15	-3,1	-0,52	-0,6	-0,63
			<i>trxB</i>	CIBE_v1_4036	ID:29052933 Thioredoxin reductase	6	-1,5	0,25	-1,6	0,85	1,3	-1,10
			<i>trxA</i>	CIBE_v1_4037	ID:29052934 Thioredoxin domain	3	-0,7	0,78	-0,5	0,35	0,6	-1,13
			<i>dfx</i>	CIBE_v1_4056	ID:29052953 dfx Desulfoferrodoxin	4	1,4	1,13	2,4	-0,34	4,3	-0,79
			<i>nox</i>	CIBE_v1_4850	ID:29053747 nox NADH oxidase	8	-1,8	-0,32	-2,9	1,12	-0,3	-0,80
			<i>fprA</i>	CIBE_v1_5384	ID:29054281 fprA Flavo-diiron protein FprA1	9	-1,1	1,06	-0,3	-0,14	1,5	-0,92
			<i>hme</i>	CIBE_v1_5755	ID:29054652 Hemerythrin	2	-0,9	-0,62	-5,7	1,15	-2,2	-0,53
		<i>hme</i>	CIBE_v1_5757	ID:29054654 Hemerythrin	6	-2,3	-0,64	-5,4	1,15	-2,3	-0,51	
		<i>dkgB</i>	CIBE_v1_4472	ID:29053369 dkgB 2,5-diketo-D-gluconic acid reductase B	15	-3,6	1,13	-2,6	-0,37	-0,5	-0,76	
	Réponse stringente	<i>relA</i>	CIBE_v1_1563	ID:29050460 RelA/SpoT domain protein	1	-1,3	1,13	0,8	-0,34	2,2	-0,78	
		<i>relA</i>	CIBE_v1_5508	ID:29054405 RelA/SpoT domain protein	2	1,1	0,07	1,0	0,96	2,2	-1,04	
		<i>codY</i>	CIBE_v1_1425	ID:29050322 codY transcriptional regulator, GTP and BCAA-dependent	1	-2,6	0,89	-1,5	0,19	-2,1	-1,08	

Réponse au stress	Carence en carbone	<i>csrA</i>	CIBE_v1_5163	ID:29054060 <i>csrA</i> carbon storage regulator	2	-2,4	-0,32	-2,5	1,12	-2,1	-0,80
		<i>cstA</i>	CIBE_v1_0725	ID:29049622 <i>cstA</i> carbon starvation-induced membrane protein	2	-4,8	1,15	1,3	-0,51	4,2	-0,64
		<i>hpf</i>	CIBE_v1_0591	ID:29049488 <i>hpf</i> ribosome-associated sigma 54 modulation protein; ribosome dimerisation factor	2	3,0	-0,66	0,9	1,15	4,6	-0,49
	Système toxine-antitoxine	<i>nodA</i>	CIBE_v1_5877	ID:29054774 <i>nodA</i> endoribonuclease toxin	1	-0,9	-0,89	-2,8	-0,20	-2,8	1,08
		<i>nodAI</i>	CIBE_v1_5878	ID:29054775 conserved protein of unknown function	1	-3,1	-1,15	-3,2	0,48	-1,2	0,67
	<i>Quorum sensing</i>	<i>luxS</i>	CIBE_v1_1483	ID:29050380 <i>luxS</i> S-ribosylhomocysteine lyase	10	0,6	1,06	1,8	-0,14	2,5	-0,92

Motilité	<i>hag</i>	CIBE_v1_5157	ID:29054054 <i>hag</i> flagellin protein	10	-0,4	-1,13	-2,2	0,76	-0,2	0,37
	<i>X</i>	CIBE_v1_5159	ID:29054056 putative Flagellar hook-associated protein 2	10	0,9	-0,76	-1,3	1,13	0,2	-0,37
	<i>fliS</i>	CIBE_v1_5160	ID:29054057 <i>fliS</i> Flagellar protein FlIS	6	-1,3	-0,65	-3,7	1,15	-1,4	-0,50
	<i>X</i>	CIBE_v1_5162	ID:29054059 FlaG protein	4	-1,2	-0,16	-2,0	1,07	-0,6	-0,91
	<i>csrA</i>	CIBE_v1_5163	ID:29054060 <i>csrA</i> carbon storage regulator	2	-2,4	-0,32	-2,5	1,12	-2,1	-0,80
	<i>fliW</i>	CIBE_v1_5164	ID:29054061 <i>fliW</i> assembly factor of the flagellum	8	-4,2	1,08	-3,4	-0,19	-2,9	-0,89
	<i>X</i>	CIBE_v1_5165	ID:29054062 Flagellar hook-associated protein 3	4	-2,6	-1,10	-3,7	0,25	-2,7	0,85
	<i>flgK</i>	CIBE_v1_5166	ID:29054063 <i>flgK</i> Flagellar hook-associated protein FlgK	13	-0,9	-0,82	-1,7	-0,29	-3,0	1,11
	<i>flgN</i>	CIBE_v1_5167	ID:29054064 FlgN family protein	1	-3,8	0,96	-2,2	0,07	-2,4	-1,03
	<i>flgM</i>	CIBE_v1_5168	ID:29054065 Anti-sigma-28 factor FlgM	3	0,2	-0,57	-0,8	1,15	-0,6	-0,58
	<i>fliY</i>	CIBE_v1_5170	ID:29054067 <i>fliY</i> flagellar motor switching and energizing phosphatase	15	-2,4	0,70	-2,7	0,44	-1,3	-1,15
	<i>cheW</i>	CIBE_v1_5172	ID:29054069 putative CheW protein	7	0,7	-0,10	0,4	1,05	1,6	-0,94
	<i>cheC</i>	CIBE_v1_5174	ID:29054071 CheC, inhibitor of MCP methylation	2	0,7	0,98	1,4	0,03	1,2	-1,01
	<i>cheR</i>	CIBE_v1_5176	ID:29054073 <i>cheR</i> Chemotaxis protein methyltransferase	6	-1,7	0,29	-1,2	0,82	-3,0	-1,11
	<i>cheD</i>	CIBE_v1_5178	ID:29054075 <i>cheD</i> sequence specific deamidase required for methylation of methyl-accepting chemotaxis proteins (MCPs) by CheR	4	-1,8	-0,53	-2,4	1,15	-2,1	-0,62
	<i>cheW</i>	CIBE_v1_5179	ID:29054076 putative CheW protein	8	0,3	-0,08	-0,3	1,04	1,2	-0,96
	<i>cheW</i>	CIBE_v1_5761	ID:29054658 putative CheW protein	1	-1,3	-0,62	-5,6	1,15	0,2	-0,53
	<i>cheC</i>	CIBE_v1_5977	ID:29054874 CheC domain protein	7	-0,5	0,04	-1,5	0,98	1,1	-1,02
	<i>spo0A</i>	CIBE_v1_2041	ID:29050938 <i>spo0A</i> response regulator	12	-0,3	0,70	0,8	0,44	2,3	-1,14
	<i>fliA</i>	CIBE_v1_5113	ID:29054010 RNA polymerase sigma factor	3	0,7	-0,93	0,6	-0,12	-0,1	1,06
	<i>pilZ</i>	CIBE_v1_5114	ID:29054011 Type IV pilus assembly PilZ	2	-1,5	1,15	-0,6	-0,62	-1,2	-0,53
	<i>fliL</i>	CIBE_v1_5122	ID:29054019 Flagellar basal body-associated protein FlIL	3	-2,6	0,61	-2,9	0,54	-0,2	-1,15
	<i>X</i>	CIBE_v1_5123	ID:29054020 <i>ylzI</i> putative flagellar protein	4	-0,2	-0,14	-1,3	1,06	1,3	-0,92
	<i>flgE</i>	CIBE_v1_5124	ID:29054021 conserved protein of unknown function	7	-1,3	-1,09	-3,2	0,87	-2,3	0,23

Partie IV : Métabolisme sur substrat 1G, approches fermentaire et multi-omique

Motilité	<i>flgD</i>	CIBE_v1_5126	ID:29054023 Flagellar hook capping protein (fragment)	2	-0,4	-0,09	-1,9	1,04	0,2	-0,95
	<i>X</i>	CIBE_v1_5127	ID:29054024 Flagellar hook-length control protein	5	-1,9	-0,53	-3,5	1,15	-2,6	-0,62
	<i>fliG</i>	CIBE_v1_5131	ID:29054028 fliG flagellar motor switching and energizing component	3	2,0	-0,67	-0,2	1,15	3,0	-0,48
	<i>fliF</i>	CIBE_v1_5132	ID:29054029 Flagellar M-ring protein	5	-0,8	-0,20	-1,1	1,08	0,6	-0,89
	<i>cheV</i>	CIBE_v1_0917	ID:29049814 cheV Chemotaxis protein CheV	11	-3,1	1,11	-1,1	-0,27	-1,2	-0,84
	<i>X</i>	CIBE_v1_1606	ID:29050503 Methyl-accepting chemotaxis sensory transducer	2	0,9	-0,21	-0,7	1,09	3,0	-0,88
	<i>X</i>	CIBE_v1_4594	ID:29053491 Methyl-accepting chemotaxis sensory transducer	13	0,8	-0,30	-0,4	1,11	1,4	-0,82
	<i>cheW</i>	CIBE_v1_4595	ID:29053492 GCN5 family acetyltransferase	2	-0,6	1,11	-0,2	-0,29	1,5	-0,82
	<i>X</i>	CIBE_v1_3540	ID:29052437 Methyl-accepting chemotaxis sensory transducer	12	0,2	0,81	0,4	0,30	2,2	-1,12
	<i>motB</i>	CIBE_v1_5770	ID:29054667 motB motility protein B; MotB component of the stator flagellum complex	4	-1,1	-0,58	-3,1	-0,57	-8,7	1,15
	<i>motA</i>	CIBE_v1_5771	ID:29054668 motA motility protein A; MotA component of the stator flagellum complex	5	-3,6	-0,46	-2,9	-0,69	-3,4	1,15
	<i>mcp</i>	CIBE_v1_0279	ID:29049176 Methyl-accepting chemotaxis sensory transducer	4	0,9	0,44	3,8	-1,14	2,6	0,71
	<i>mcp</i>	CIBE_v1_0835	ID:29049732 Chemotaxis protein	19	1,9	0,13	1,5	0,93	3,8	-1,06
	<i>mcp</i>	CIBE_v1_0837	ID:29049734 Chemotaxis protein	16	-0,5	0,41	-1,3	0,73	3,6	-1,14
	<i>X</i>	CIBE_v1_0838	ID:29049735 Anti-sigma factor antagonist	2	1,6	-0,14	1,2	1,06	2,6	-0,93
	<i>mcp</i>	CIBE_v1_2305	ID:29051202 Methyl-accepting chemotaxis protein	9	1,2	-0,80	0,5	1,12	0,6	-0,32
	<i>mcp</i>	CIBE_v1_2627	ID:29051524 conserved exported protein of unknown function	13	0,6	-0,74	0,9	-0,40	1,4	1,14
	<i>mcp</i>	CIBE_v1_4848	ID:29053745 Methyl-accepting chemotaxis sensory transducer	14	-2,0	1,12	-0,8	-0,33	2,9	-0,79
	<i>mcp</i>	CIBE_v1_3299	ID:29052196 conserved protein of unknown function	8	-0,9	-0,23	-0,4	1,10	2,2	-0,86
	<i>mcp</i>	CIBE_v1_1059	ID:29049956 Methyl-accepting chemotaxis protein	6	3,0	-0,86	0,4	1,10	3,7	-0,24
	<i>mcp</i>	CIBE_v1_1062	ID:29049959 Methyl-accepting chemotaxis protein	21	1,3	-0,59	-0,6	1,15	2,1	-0,57
	<i>mcp</i>	CIBE_v1_3612	ID:29052509 Chemotaxis sensory transducer	27	-1,5	1,15	1,1	-0,52	3,3	-0,63
	<i>mcp</i>	CIBE_v1_3705	ID:29052602 Methyl-accepting chemotaxis sensory transducer	17	1,8	-0,39	0,1	1,14	3,8	-0,75
	<i>mcp</i>	CIBE_v1_5253	ID:29054150 conserved exported protein of unknown function	12	-1,9	1,15	1,4	-0,49	3,7	-0,66
	<i>mcp</i>	CIBE_v1_5510	ID:29054407 Methyl-accepting chemotaxis sensory transducer	19	1,6	0,06	0,4	0,97	2,7	-1,03
	<i>mcp</i>	CIBE_v1_0835	ID:29049732 Chemotaxis protein	19	1,9	0,13	1,5	0,93	3,8	-1,06
	<i>mcp</i>	CIBE_v1_6080	ID:29054977 Chemotaxis protein	11	-4,4	0,16	-4,2	0,91	-0,4	-1,07
	<i>mcp</i>	CIBE_v1_4382	ID:29053279 Methyl-accepting chemotaxis sensory transducer	13	-0,6	-0,87	-0,5	-0,22	-0,7	1,09
	<i>mcp</i>	CIBE_v1_5637	ID:29054534 Methyl-accepting chemotaxis sensory transducer	10	0,1	-0,34	1,4	-0,78	-0,1	1,13

	<i>fabH</i>	CIBE_v1_1267	ID:29050164 <i>fabH</i> beta-ketoacyl-acyl carrier protein synthase III 1	4	-2,9	0,73	-2,1	0,42	-0,6	-1,14
	<i>fabK</i>	CIBE_v1_1268	ID:29050165 2-nitropropane dioxygenase, NPD	8	-2,0	0,44	-1,3	0,70	-0,5	-1,15
	<i>fabD</i>	CIBE_v1_1269	ID:29050166 <i>fabD</i> malonyl CoA:acyl carrier protein transacylase	5	-1,0	0,91	0,4	0,16	-0,7	-1,07
	<i>fabG</i>	CIBE_v1_0624	ID:29049521 <i>kduD</i> 2-keto-3-deoxygluconate oxidoreductase	8	1,5	1,02	3,3	-0,05	2,6	-0,98
	<i>fabF</i>	CIBE_v1_1271	ID:29050168 <i>fabF</i> beta-ketoacyl-acyl carrier protein synthase II	17	-1,8	-0,67	-1,8	-0,48	-3,1	1,15
Synthèse d'acides gras	<i>accB</i>	CIBE_v1_1272	ID:29050169 Acetyl-CoA carboxylase, biotin carboxyl carrier protein	4	-2,3	0,77	-1,5	0,36	-0,1	-1,13
	<i>fabZ</i>	CIBE_v1_1273	ID:29050170 <i>fabZ</i> beta-hydroxyacyl-[acyl carrier protein] dehydratase	4	2,4	0,39	1,1	0,75	1,9	-1,14
	<i>accC</i>	CIBE_v1_1274	ID:29050171 <i>accC</i> acetyl-CoA carboxylase subunit (biotin carboxylase subunit)	7	-1,7	1,11	0,1	-0,30	-1,7	-0,82
	<i>accD</i>	CIBE_v1_1275	ID:29050172 <i>accD</i> acetyl-CoA carboxylase (carboxyltransferase beta subunit)	3	-5,3	1,12	-3,1	-0,33	-1,3	-0,79
	<i>accA</i>	CIBE_v1_1276	ID:29050173 <i>accA</i> acetyl-CoA carboxylase (carboxyltransferase alpha subunit)	6	-3,3	1,13	-1,9	-0,35	0,7	-0,78
	<i>acpP</i>	CIBE_v1_1405	ID:29050302 <i>acpA</i> acyl carrier protein	1	1,9	0,11	1,5	0,94	2,8	-1,05
Biogénèse de la membrane	<i>glmS</i>	CIBE_v1_0267	ID:29049164 <i>glmS</i> Glutamine-fructose-6-phosphate aminotransferase [isomerizing]	2	-2,2	0,53	-3,2	0,63	1,1	-1,15
	<i>glmM</i>	CIBE_v1_0230	ID:29049127 <i>glmM</i> phosphoglucosamine mutase	12	-0,8	0,61	-0,4	0,54	1,6	-1,15
	<i>glmU</i>	CIBE_v1_0090	ID:29048987 <i>glmU</i> bifunctional glucosamine-1-phosphate N-acetyltransferase/UDP-N-acetylglucosamine pyrophosphorylase	5	-2,7	1,04	-1,6	-0,08	1,1	-0,96
	<i>murAB</i>	CIBE_v1_6083	ID:29054980 <i>murAB</i> UDP-N-acetylglucosamine 1-carboxyvinyltransferase	5	-0,4	1,09	0,5	-0,22	1,0	-0,87
	<i>murB</i>	CIBE_v1_5863	ID:29054760 <i>murB</i> UDP-N-acetylenoylpyruvoylglucosamine reductase	2	-2,5	1,15	-1,6	-0,70	-1,7	-0,44
	<i>murC</i>	CIBE_v1_0085	ID:29048982 <i>murC</i> UDP-N-acetylmuramate-L-alanine ligase	5	-3,1	1,14	-1,2	-0,40	-0,2	-0,74
	<i>murD</i>	CIBE_v1_0118	ID:29049015 <i>murD</i> UDP-N-acetylmuramoylalanine-D-glutamate ligase	8	-3,0	1,13	-0,7	-0,36	-1,0	-0,77
	<i>murE</i>	CIBE_v1_1822	ID:29050719 <i>murE</i> UDP-N-acetylmuramoylalanine-D-glutamate-2,6-diaminopimelate ligase	9	-3,0	1,15	-2,8	-0,58	-2,9	-0,57
	<i>murF</i>	CIBE_v1_1823	ID:29050720 UDP-N-acetylmuramoyl-tripeptide-D-alanyl-D-alanine ligase	15	-2,8	0,99	-2,4	0,02	-1,6	-1,01
	<i>murI</i>	CIBE_v1_0036	ID:29048933 <i>murI</i> Glutamate racemase	2	-2,0	0,65	-1,1	0,50	0,3	-1,15
	<i>nagBA</i>	CIBE_v1_5473	ID:29054370 <i>nagBA</i> N-acetylglucosamine-6-phosphate isomerase	8	-2,9	1,13	-1,7	-0,35	0,5	-0,78
	<i>nagA</i>	CIBE_v1_5475	ID:29054372 <i>nagA</i> N-acetylglucosamine-6-phosphate deacetylase	6	-2,0	0,81	-1,9	0,31	2,0	-1,12
	<i>ddl</i>	CIBE_v1_0752	ID:29049649 <i>ddl</i> D-alanine-D-alanine ligase	9	-1,1	0,79	-1,2	0,33	1,6	-1,12
	<i>ddc</i>	CIBE_v1_1131	ID:29050028 D-Ala-D-Ala carboxypeptidase	8	-3,7	1,15	-2,4	-0,67	-1,0	-0,48
	<i>lysM</i>	CIBE_v1_0543	ID:29049440 Peptidoglycan-binding LysM	6	-0,2	1,10	2,0	-0,85	1,4	-0,25
<i>pbpA</i>	CIBE_v1_5093	ID:29053990 <i>pbpA</i> Penicillin-binding protein 1A	3	-1,8	1,12	-1,1	-0,32	0,0	-0,80	

Partie IV : Métabolisme sur substrat 1G, approches fermentaire et multi-omique

Biosynthèse du peptidoglycane	<i>pbpC</i>	CIBE_v1_2312	ID:29051209 pbpC Penicillin-binding protein 3	7	-3,3	1,10	-0,9	-0,25	-0,8	-0,85	
	<i>gtaB</i>	CIBE_v1_1299	ID:29050196 gtaB UTP-glucose-1-phosphate uridylyltransferase	5	-0,8	-0,66	-5,6	1,15	-0,9	-0,49	
	<i>gtaB</i>	CIBE_v1_5525	ID:29054422 gtaB UTP-glucose-1-phosphate uridylyltransferase	18	-3,6	0,94	-3,0	0,11	-0,6	-1,05	
	Biosynthèse des acides teichoïques	<i>ywqF</i>	CIBE_v1_5651	ID:29054548 ywqF UDP-glucose dehydrogenase	1	5,6	-0,70	-2,6	1,15	2,4	-0,45
		<i>mnaA</i>	CIBE_v1_0564	ID:29049461 mnaA UDP-N-acetylmannosamine 2-epimerase	5	-2,8	-0,36	-2,3	-0,77	-2,7	1,13
		<i>mnaA</i>	CIBE_v1_5535	ID:29054432 UDP-N-acetylglucosamine 2-epimerase homolog	7	-3,0	1,15	-1,8	-0,65	-2,6	-0,51
		<i>dltC</i>	CIBE_v1_5198	ID:29054095 dltC D-alanine--poly(phosphoribitol) ligase subunit 2	2	2,3	0,55	2,2	0,60	3,7	-1,15
		<i>rmlA</i>	CIBE_v1_1289	ID:29050186 rmlA glucose-1-phosphate thymidyltransferase	4	0,8	-0,43	-2,2	1,14	1,6	-0,72
	Biosynthesis de TDP-L-rhamnose	<i>rmlC</i>	CIBE_v1_1290	ID:29050187 rmlC dTDP-4-deoxyrhamnose-3,5-epimerase	11	0,6	-0,17	-1,1	1,07	1,2	-0,91
		<i>rmlD</i>	CIBE_v1_1291	ID:29050188 dTDP-4-deoxyrhamnose reductase	6	3,9	1,07	3,4	-0,91	3,7	-0,15
		<i>rmlB</i>	CIBE_v1_1293	ID:29050190 spsJ dTDP-glucose 4,6-dehydratase	4	0,0	-0,20	-2,6	1,08	2,2	-0,89
		<i>rmlC</i>	CIBE_v1_5144	ID:29054041 rmlC dTDP-4-deoxyrhamnose-3,5-epimerase	3	-4,3	0,33	-4,6	0,80	-3,1	-1,12
		<i>rmlB</i>	CIBE_v1_5147	ID:29054044 spsJ dTDP-glucose 4,6-dehydratase	4	-2,7	-1,11	-2,9	0,26	-2,5	0,84
		<i>dgk</i>	CIBE_v1_0744	ID:29049641 Diacylglycerol kinase, catalytic region	13	-0,9	0,94	-0,4	0,11	1,1	-1,05
	Biogénèse de la membrane		CIBE_v1_0061	ID:29048958 3D domain protein	2	2,5	1,14	4,7	-0,43	4,3	-0,72
			CIBE_v1_0480	ID:29049377 protein of unknown function	2	2,9	1,08	3,7	-0,18	5,7	-0,90
			CIBE_v1_0488	ID:29049385 Glycoside hydrolase, family 25	1	1,3	-1,15	0,9	0,58	1,0	0,58
			CIBE_v1_3350	ID:29052247 Beta and gamma crystallin	5	1,3	1,15	3,7	-0,50	6,6	-0,65
			CIBE_v1_4341	ID:29053238 Sugar-binding protein	1	8,7	-0,61	7,7	-0,55	1,7	1,15
			CIBE_v1_5267	ID:29054164 YD repeat protein	24	3,1	0,41	3,1	0,73	3,3	-1,14
		CIBE_v1_5575	ID:29054472 Glycoside hydrolase	2	3,3	-0,22	4,3	-0,87	3,4	1,09	
Protéines membranaires solvantogènes inconnues			CIBE_v1_5615	ID:29054512 Cell wall hydrolase	6	1,4	-0,06	1,2	1,03	1,0	-0,97
			CIBE_v1_5616	ID:29054513 Cell wall hydrolase	2	1,9	-1,10	1,2	0,25	1,2	0,85
			CIBE_v1_5628	ID:29054525 putative cell wall binding repeat-containing protein	35	4,4	-0,65	3,4	1,15	5,2	-0,50
			CIBE_v1_5630	ID:29054527 putative cell wall binding repeat-containing protein	26	4,7	-0,58	-0,8	1,15	3,8	-0,58
			CIBE_v1_5631	ID:29054528 putative cell wall binding repeat-containing protein	50	3,8	-0,58	0,5	1,15	3,1	-0,58
			CIBE_v1_5676	ID:29054573 putative cell wall binding repeat-containing protein	40	4,2	-0,40	2,7	1,14	5,2	-0,74
			CIBE_v1_5682	ID:29054579 Cell wall-binding protein	19	3,4	1,14	4,0	-0,70	3,4	-0,44
			CIBE_v1_5894	ID:29054791 Glycoside hydrolase	10	2,2	-0,43	1,8	1,14	2,7	-0,71
		CIBE_v1_2159	ID:29051056 putative cell wall binding repeat-containing protein	13	3,0	1,13	4,5	-0,78	4,6	-0,35	
	CIBE_v1_3306	ID:29052203 Cell wall-binding protein	2	1,8	0,38	1,9	0,75	6,0	-1,13		

Biogénèse de la membrane	Protéines membranaires solvantogènes inconnues	CIBE_v1_4995	ID:29053892 tcsA CD4+ T-cell-stimulating antigen	9	1,0	0,10	0,3	0,95	2,6	-1,05
		CIBE_v1_5638	ID:29054535 Cell wall-binding protein	16	3,5	-0,41	3,8	-0,73	1,3	1,14
		CIBE_v1_5678	ID:29054575 putative cell wall binding repeat-containing protein	12	1,2	-0,01	0,8	1,01	3,3	-0,99
		CIBE_v1_5681	ID:29054578 Cell wall-binding protein	22	1,1	1,15	2,3	-0,51	3,5	-0,65
		CIBE_v1_5683	ID:29054580 putative cell wall binding repeat-containing protein	34	3,5	1,15	4,8	-0,59	6,2	-0,57
		CIBE_v2_p0003	ID:29048865 putative cell wall binding protein	16	2,5	1,05	3,2	-0,10	2,3	-0,95

Réplication	Division cellulaire	<i>ftsA</i>	CIBE_v1_1601	ID:29050498 Cell division protein FtsA	11	-3,6	0,30	-3,4	0,82	0,8	-1,12	
		<i>ftsE</i>	CIBE_v1_5873	ID:29054770 ftsE cell-division ABC transporter (ATP-binding protein)	5	-3,2	1,15	-2,3	-0,45	-2,8	-0,69	
		<i>ftsX</i>	CIBE_v1_5872	ID:29054769 Cell division protein FtsX	3	-1,9	1,12	-0,7	-0,31	0,2	-0,81	
		<i>ftsZ</i>	CIBE_v1_1354	ID:29050251 ftsZ cell-division initiation protein	9	-0,6	0,61	-0,7	0,54	0,9	-1,15	
		<i>ftsI</i>	CIBE_v1_5865	ID:29054762 Penicillin-binding protein	8	-2,3	0,48	-1,8	0,67	-0,2	-1,15	
		<i>ftsQ</i>	CIBE_v1_1826	ID:29050723 Polypeptide-transport-associated domain protein, FtsQ-type	2	-0,5	0,71	0,7	0,44	4,0	-1,14	
		<i>zapA</i>	CIBE_v1_1134	ID:29050031 conserved protein of unknown function	5	-4,6	1,06	-2,2	-0,92	-1,7	-0,14	
		<i>divIVA</i>	CIBE_v1_1834	ID:29050731 Cell division protein DivIVA	8	1,0	0,15	0,9	0,92	2,6	-1,06	
		<i>mreB</i>	CIBE_v1_0661	ID:29049558 mreB cell-shape determining protein	7	-2,1	1,15	-1,2	-0,59	-0,6	-0,57	
		<i>mreC</i>	CIBE_v1_0662	ID:29049559 Cell shape-determining protein MreC	3	-1,7	0,92	-1,3	0,14	2,3	-1,06	
		<i>minC</i>	CIBE_v1_0667	ID:29049564 minC putative septum site-determining protein MinC	1	-0,3	-0,36	-1,7	1,13	3,4	-0,77	
		<i>minD</i>	CIBE_v1_0668	ID:29049565 minD ATPase activator of MinC	7	-1,4	0,79	-0,2	-1,12	-0,5	0,33	
		<i>minE</i>	CIBE_v1_0669	ID:29049566 minE Cell division topological specificity factor	3	-1,5	0,60	-1,6	0,56	1,0	-1,15	
		<i>mrp</i>	CIBE_v1_0068	ID:29048965 mrp Protein mrp homolog	10	-1,6	0,25	-2,0	0,85	1,2	-1,10	
		<i>parA</i>	CIBE_v1_6123	ID:29055020 parA chromosome partitioning protein; transcriptional regulator	3	-1,8	1,06	-0,4	-0,13	-0,8	-0,93	
		Réplication de l'ADN	<i>dnaE</i>	CIBE_v1_5858	ID:29054755 dnaE DNA polymerase III subunit alpha	1	0,4	-0,18	1,4	-0,90	-0,9	1,08
			<i>dnaN</i>	CIBE_v1_0002	ID:29048899 DNA polymerase III subunit beta	13	-1,2	0,87	0,0	0,22	0,0	-1,09
			<i>dnaQ</i>	CIBE_v1_1208	ID:29050105 Exonuclease, RNase T and DNA polymerase III	5	-3,8	1,15	-1,0	-0,46	-1,4	-0,69
<i>gyrB</i>	CIBE_v1_0006		ID:29048903 gyrB DNA gyrase (subunit B)	14	-2,1	0,02	-2,5	0,99	0,9	-1,01		
<i>gyrA</i>	CIBE_v1_0007		ID:29048904 gyrA DNA gyrase (subunit A)	9	-2,3	0,47	-1,7	0,68	0,6	-1,15		
<i>ligA</i>	CIBE_v1_0425		ID:29049322 ligA DNA ligase	2	-2,2	-1,15	-1,7	0,59	-3,5	0,56		
<i>polA</i>	CIBE_v1_0979		ID:29049876 polA DNA polymerase I	2	-3,3	0,58	-1,1	0,58	1,7	-1,15		
<i>topA</i>	CIBE_v1_1424	ID:29050321 topA DNA topoisomerase I	4	0,1	0,12	0,2	0,93	1,2	-1,06			

Partie IV : Métabolisme sur substrat 1G, approches fermentaire et multi-omique

Réplication	Réplication de l'ADN	<i>topB</i>	CIBE_v1_2217	ID:29051114 topB DNA topoisomerase III	1	-0,6	-0,53	0,4	-0,62	-2,1	1,15	
		<i>lig</i>	CIBE_v1_5255	ID:29054152 DNA ligase III	1	-0,1	-0,35	-0,7	1,13	1,0	-0,78	
		<i>hbs</i>	CIBE_v1_0101	ID:29048998 hbs non-specific DNA-binding protein HBSu; signal recognition particle-like (SRP) component	7	1,4	-0,07	0,7	1,03	2,9	-0,96	
		<i>lig</i>	CIBE_v1_1071	ID:29049968 NAD-dependent DNA ligase	5	-1,4	1,15	0,8	-0,60	-0,1	-0,56	
Sporulation	Sporulation, étapes 0 et I	<i>spo0A</i>	CIBE_v1_2041	ID:29050938 spo0A response regulator	12	-0,3	0,70	0,8	0,44	2,3	-1,14	
		<i>sigH</i>	CIBE_v1_0151	ID:29049048 sigH RNA polymerase sigma-30 factor (sigma(H))	1	-0,3	0,23	-0,4	0,86	1,9	-1,10	
		<i>abrB</i>	CIBE_v1_5891	ID:29054788 Transcriptional regulator, AbrB family	3	-0,3	0,09	0,0	0,95	-0,2	-1,04	
		<i>abrB</i>	CIBE_v1_4085	ID:29052982 Transcriptional regulator, AbrB family	4	-1,2	-0,35	1,4	-0,78	-3,9	1,13	
		<i>oppA</i>	CIBE_v1_5956	ID:29054853 oppA Oligopeptide-binding protein OppA	27	0,9	0,95	0,4	0,10	3,0	-1,05	
		<i>oppB</i>	CIBE_v1_5960	ID:29054857 oppB oligopeptide ABC transporter (permease)	2	-1,7	1,03	-1,9	-0,05	2,0	-0,97	
		<i>oppD</i>	CIBE_v1_5958	ID:29054855 oppD oligopeptide ABC transporter (ATP-binding protein)	4	-2,3	1,15	-2,4	-0,55	0,7	-0,61	
		Sporulation, étapes II à V	<i>spolIAB</i>	CIBE_v1_0993	ID:29049890 spolIAB anti-sigma F factor and serine kinase	5	-0,2	1,15	4,2	-0,60	4,4	-0,55
			<i>sigF</i>	CIBE_v1_0994	ID:29049891 sigF RNA polymerase sigma factor (sigma-F)	2	-0,8	1,12	1,2	-0,31	3,6	-0,80
			<i>spolIQ</i>	CIBE_v1_0578	ID:29049475 Peptidase M23B	4	0,9	1,01	4,1	-0,99	4,0	-0,03
	<i>spolIIAH</i>		CIBE_v1_2027	ID:29050924 Stage III sporulation protein AH	4	-1,3	0,99	0,0	0,01	5,0	-1,01	
	<i>prk</i>		CIBE_v1_0737	ID:29049634 prkA serine protein kinase	2	-0,4	1,15	4,0	-0,67	6,1	-0,48	
	<i>yabP</i>		CIBE_v1_0103	ID:29049000 YabP family protein	1	-3,9	1,14	3,2	-0,73	5,3	-0,41	
	<i>spoVS</i>		CIBE_v1_1456	ID:29050353 spoVS regulator required for stage V sporulation	2	-6,4	0,01	-4,4	1,00	-1,7	-1,00	
	<i>spoVG</i>		CIBE_v1_0089	ID:29048986 spoVG regulator required for spore cortex synthesis	9	0,0	0,90	-0,2	0,18	0,7	-1,08	
	<i>spoVT</i>		CIBE_v1_0098	ID:29048995 spoVT transcriptional regulator	1	0,5	0,99	3,9	-1,01	2,9	0,02	
	Métabolisme du granuloze		<i>pgm</i>	CIBE_v1_1177	ID:29050074 pgcA alpha-phosphoglucomutase	25	-2,3	1,04	-1,3	-0,08	1,1	-0,96
		<i>glgD</i>	CIBE_v1_5923	ID:29054820 Glucose-1-phosphate adenylyltransferase	6	-2,7	0,59	-1,4	0,56	-1,5	-1,15	
		<i>glgC</i>	CIBE_v1_5924	ID:29054821 glgC glucose-1-phosphate adenylyltransferase (ADP-glucose pyrophosphorylase) subunit alpha	6	-1,6	0,70	-0,9	0,44	0,0	-1,15	
		<i>glgP</i>	CIBE_v1_0286	ID:29049183 malP Maltodextrin phosphorylase	4	-4,6	0,81	-3,7	0,31	-3,0	-1,12	
<i>glgB</i>		CIBE_v1_5928	ID:29054825 glgB 1,4-alpha-glucan branching enzyme GlgB	6	-1,6	0,19	-2,2	0,89	2,2	-1,08		
HK orpheline	<i>X</i>	CIBE_v1_0761	ID:29049658 Histidine kinase	8	-3,6	0,92	-3,0	0,14	-0,2	-1,06		

Métabolisme secondaire	Biosynthèse de polyamines	<i>nspC</i>	CIBE_v1_5306	ID:29054203 nspC Carboxynorspermidine/carboxyspermidine decarboxylase	2	-5,8	1,08	-4,7	-0,18	-1,7	-0,90
		<i>nspS</i>	CIBE_v1_5307	ID:29054204 Carboxynorspermidine synthase	13	-3,4	1,15	-1,4	-0,50	-0,5	-0,66
		<i>speE</i>	CIBE_v1_5308	ID:29054205 speE Spermidine synthase	6	-3,8	1,01	-3,5	-0,03	-2,5	-0,98
		<i>speA</i>	CIBE_v1_5309	ID:29054206 speA Arginine decarboxylase	4	-1,7	1,13	-4,4	-0,34	0,8	-0,78
		<i>aguA</i>	CIBE_v1_2337	ID:29051234 aguA putative agmatine deiminase	5	-2,2	1,13	0,0	-0,36	1,2	-0,77
	Biosynthèse d'acide dipicolinique	<i>dapB</i>	CIBE_v1_2137	ID:29051034 dapB 4-hydroxy-tetrahydopicolinate reductase	11	-0,2	0,99	0,5	0,02	0,5	-1,01
		<i>dapA</i>	CIBE_v1_2138	ID:29051035 dapA dihydroadiponate synthase	5	-0,9	-0,32	-1,2	1,12	0,5	-0,80
		<i>asd</i>	CIBE_v1_2139	ID:29051036 asd aspartate-semialdehyde dehydrogenase	25	-2,5	1,06	-1,6	-0,13	-0,9	-0,93
		<i>dapA</i>	CIBE_v1_2138	ID:29051035 dapA dihydroadiponate synthase	5	-0,9	-0,32	-1,2	1,12	0,5	-0,80

Biosynthèse de cofacteurs	Formation d'ATP/GTP	<i>adk</i>	CIBE_v1_0190	ID:29049087 adk adenylate kinase	11	-1,9	1,01	-0,6	-0,02	0,2	-0,99	
		<i>ndk</i>	CIBE_v1_5096	ID:29053993 ndk Nucleoside diphosphate kinase	4	-2,0	1,15	1,1	-0,67	-0,9	-0,48	
		<i>gmk</i>	CIBE_v1_1377	ID:29050274 gmk guanylate kinase	5	-1,4	1,13	0,2	-0,38	0,9	-0,75	
		<i>atpF</i>	CIBE_v1_0569	ID:29049466 atpF ATP synthase subunit b	6	-0,2	0,55	-0,4	0,61	0,1	-1,15	
		<i>atpH</i>	CIBE_v1_0570	ID:29049467 atpH ATP synthase subunit delta	11	0,3	0,23	0,1	0,86	0,3	-1,10	
		<i>atpA</i>	CIBE_v1_0571	ID:29049468 atpA ATP synthase (subunit alpha, component F1)	21	0,0	0,82	0,1	0,29	-0,2	-1,11	
		<i>atpG</i>	CIBE_v1_0572	ID:29049469 atpG ATP synthase gamma chain	10	-1,7	0,56	-1,7	0,60	1,0	-1,15	
		<i>atpD</i>	CIBE_v1_0573	ID:29049470 atpD ATP synthase (subunit beta, component F1)	28	0,6	0,98	0,8	0,04	0,5	-1,02	
		<i>atpC</i>	CIBE_v1_0574	ID:29049471 atpC ATP synthase epsilon chain	3	0,6	-0,76	0,7	1,13	-0,1	-0,37	
		<i>nadD</i>	CIBE_v1_0684	ID:29049581 nadD putative nicotinate-nucleotide adenyltransferase	4	-3,5	0,50	-3,4	0,66	-2,2	-1,15	
	Biosynthèse de NAD(P) ⁺	<i>nadC</i>	CIBE_v1_0972	ID:29049869 nadC nicotinate-nucleotide pyrophosphorylase	3	1,2	1,12	4,0	-0,81	-3,0	-0,30	
		<i>ppnK</i>	CIBE_v1_2036	ID:29050933 ppnK putative inorganic polyphosphate/ATP-NAD kinase	1	-0,6	0,48	-0,7	-1,15	-1,2	0,67	
		<i>pncB</i>	CIBE_v1_1593	ID:29050490 pncB Nicotinate phosphoribosyltransferase	15	2,9	0,92	2,7	0,14	3,2	-1,06	
		<i>pncB</i>	CIBE_v1_2500	ID:29051397 Nicotinate phosphoribosyltransferase	1	-4,5	1,14	-1,1	-0,42	1,6	-0,72	
		<i>nadA</i>	CIBE_v1_0970	ID:29049867 nadA Quinolinate synthase A	1	1,4	0,79	5,6	0,34	4,0	-1,12	
		<i>nadE</i>	CIBE_v1_1241	ID:29050138 NAD ⁺ synthetase	3	-3,3	1,12	-1,8	-0,31	0,7	-0,81	
		Biosynthèse de thiamine	<i>thiF</i>	CIBE_v1_0813	ID:29049710 Thiamine biosynthesis protein ThiF	3	-3,5	0,93	-3,0	0,13	0,6	-1,06
			<i>thiG</i>	CIBE_v1_1754	ID:29050651 thiG hydroxyethylthiazole phosphate synthetase (thiamine biosynthesis)	6	1,4	0,71	1,6	0,44	1,8	-1,14
	<i>thiH</i>		CIBE_v1_1755	ID:29050652 Thiazole biosynthesis protein ThiH	2	-1,8	1,13	-0,6	-0,36	-0,9	-0,77	
	<i>apbE</i>		CIBE_v1_4449	ID:29053346 Thiamine biosynthesis protein ApbE	2	0,6	1,12	2,3	-0,31	7,3	-0,81	

Partie IV : Métabolisme sur substrat 1G, approches fermentaire et multi-omique

Biosynthèse de thiamine	<i>thiH</i>	CIBE_v1_5341	ID:29054238 thiH Thiamine biosynthesis protein ThiH	7	-2,9	1,12	-2,6	-0,30	-1,1	-0,82		
	<i>tpg</i>	CIBE_v1_1390	ID:29050287 Thiamine pyrophosphokinase	1	-5,8	1,14	-0,3	-0,72	-2,1	-0,42		
	<i>thiE</i>	CIBE_v1_5391	ID:29054288 thiE thiamine-phosphate pyrophosphorylase	1	-1,9	-0,81	-2,1	1,12	-3,8	-0,31		
	<i>thiM</i>	CIBE_v1_5392	ID:29054289 thiM hydroxyethylthiazole kinase	4	-1,2	0,98	-0,9	0,03	-0,2	-1,02		
	<i>thiC</i>	CIBE_v1_5393	ID:29054290 thiC Phosphomethylpyrimidine synthase	6	2,2	-0,17	1,5	1,07	-0,1	-0,91		
	<i>thiJ</i>	CIBE_v1_5190	ID:29054087 4-methyl-5(B-hydroxyethyl)-thiazole monophosphate biosynthesis protein	8	1,1	1,12	2,7	-0,30	3,7	-0,81		
Biosynthèse de biotine	<i>X</i>	CIBE_v1_1210	ID:29050107 Biotin--acetyl-CoA-carboxylase ligase	5	-2,8	-0,75	-1,8	1,14	-1,4	-0,39		
	<i>bioB</i>	CIBE_v1_2327	ID:29051224 bioB Biotin synthase	1	-3,8	1,15	-0,1	-0,65	-2,8	-0,50		
Biosynthesis de panthoténate	<i>panC</i>	CIBE_v1_3080	ID:29051977 panC pantothenate synthetase	6	-2,5	1,05	0,4	-0,11	1,2	-0,94		
	<i>panB</i>	CIBE_v1_3081	ID:29051978 panB ketopantoate hydroxymethyltransferase	13	2,2	0,73	2,8	0,41	4,6	-1,14		
	<i>coaD</i>	CIBE_v1_1396	ID:29050293 coaD phosphopantetheine adenyltransferase	2	-2,3	0,94	-1,1	0,11	1,1	-1,05		
	<i>panD</i>	CIBE_v1_3079	ID:29051976 panD aspartate 1-decarboxylase	8	-0,5	-0,99	-1,3	-0,03	-0,5	1,01		
Biosynthèse de cofacteurs	Biosynthèse de cobalamine	<i>cbiH</i>	CIBE_v1_1506	ID:29050403 cbiH Cobalt-precorrin-3B C(17)-methyltransferase	4	-1,2	-0,36	-2,1	1,13	-1,3	-0,77	
		<i>btuR</i>	CIBE_v1_1514	ID:29050411 Cobinamide adenosyltransferase	5	-2,2	0,97	-1,4	0,05	0,4	-1,02	
		<i>cbiC</i>	CIBE_v1_2730	ID:29051627 cbiC Cobalt-precorrin-8X methylmutase	2	-6,2	0,47	-4,8	-1,15	-6,5	0,68	
		<i>cbiF</i>	CIBE_v1_2735	ID:29051632 cbiF putative cobalt-precorrin-4 C(11)-methyltransferase	7	-2,1	1,15	0,0	-0,49	0,7	-0,66	
		<i>cobW</i>	CIBE_v1_2904	ID:29051801 Cobalamin biosynthesis protein CobW	2	0,3	1,11	1,8	-0,27	2,7	-0,83	
		<i>cobB</i>	CIBE_v1_5900	ID:29054797 cobB NAD-dependent protein deacetylase	3	-1,1	1,09	-0,6	-0,88	-1,1	-0,20	
		<i>cbiT</i>	CIBE_v1_2733	ID:29051630 cbiT putative cobalt-precorrin-6Y C(15)-methyltransferase [decarboxylating]	1	-0,5	1,15	2,4	-0,53	-1,0	-0,63	
		<i>cobQ</i>	CIBE_v1_1523	ID:29050420 cobQ Cobyric acid synthase	2	0,2	0,37	-0,1	0,77	1,0	-1,13	
		<i>ribF</i>	CIBE_v1_1444	ID:29050341 Riboflavin biosynthesis protein RibF	3	-4,9	1,15	-2,2	-0,45	-0,8	-0,69	
		Biosynthèse de riboflavine	<i>ribD</i>	CIBE_v1_1462	ID:29050359 ribD fused diaminohydroxyphosphoribosylaminopyrimidine deaminase; 5-amino-6-(5-phosphoribosylamino) uracil reductase	5	-0,5	-0,27	-1,7	1,11	0,7	-0,84
			<i>ribE</i>	CIBE_v1_1463	ID:29050360 ribE riboflavin synthase (alpha subunit)	5	-1,7	-0,36	-2,8	1,13	1,4	-0,77
<i>ribAB</i>	CIBE_v1_1464		ID:29050361 ribAB fused GTP cyclohydrolase II and 3,4-dihydroxy-2-butanone 4-phosphate synthase	5	-3,1	-0,20	-2,6	1,08	0,1	-0,89		
<i>ribH</i>	CIBE_v1_1465		ID:29050362 ribH 6,7-dimethyl-8-ribityllumazine synthase, beta subunit	7	-0,8	-0,19	-1,3	1,08	1,9	-0,89		
<i>ribD</i>	CIBE_v1_4303		ID:29053200 Riboflavin biosynthesis protein RibD	2	0,1	1,05	2,0	-0,11	1,3	-0,94		
Biosynthèse de pyridoxamine	<i>pdxH</i>	CIBE_v1_2649	ID:29051546 Pyridoxamine 5'-phosphate oxidase-like	4	1,8	0,85	3,1	-1,10	3,0	0,26		
	<i>pdxH</i>	CIBE_v1_2692	ID:29051589 Pyridoxamine 5'-phosphate oxidase-related	7	-1,9	1,06	-1,4	-0,14	2,3	-0,92		
	<i>pdxT</i>	CIBE_v1_6054	ID:29054951 pdxT pyridoxal 5'-phosphate synthase complex, glutamine amidotransferase subunit PdxT	1	2,3	-1,01	1,0	0,99	-1,5	0,01		
	<i>pdxS</i>	CIBE_v1_6055	ID:29054952 pdxS pyridoxal 5'-phosphate synthase complex, synthase subunit	11	1,2	1,04	2,6	-0,09	1,7	-0,95		

Assimilation de sels minéraux	Fer	<i>fur</i>	CIBE_v1_0753	ID:29049650 Ferric uptake regulator, Fur family	2	-4,0	1,09	-1,2	-0,21	0,7	-0,88
		<i>fur</i>	CIBE_v1_1343	ID:29050240 fur Ferric uptake regulation protein	4	0,3	0,22	0,5	0,87	1,7	-1,09
		<i>ftnA</i>	CIBE_v1_3001	ID:29051898 ftnA Ferritin	7	-0,1	1,13	2,5	-0,37	3,1	-0,77
		X	CIBE_v1_4451	ID:29053348 Iron ABC transporter	4	2,2	0,86	3,7	0,24	5,2	-1,10
		X	CIBE_v1_3663	ID:29052560 Iron dependent repressor	1	-0,6	-0,03	-0,4	1,01	2,6	-0,99
	Soufre	<i>asrA</i>	CIBE_v1_0960	ID:29049857 asrA Anaerobic sulfite reductase subunit A	2	-1,3	-0,52	-3,3	-0,63	0,7	1,15
		<i>asrB</i>	CIBE_v1_0961	ID:29049858 asrB Anaerobic sulfite reductase subunit B	2	6,5	-0,60	-3,6	-0,56	0,1	1,15
		X	CIBE_v1_3293	ID:29052190 Sulfite reductase, assimilatory-type	8	-2,8	1,04	-2,9	-0,07	-1,3	-0,96
		X	CIBE_v1_1234	ID:29050131 putative sulfonate/nitrate transport system substrate-binding protein	5	-0,6	0,52	-1,6	0,64	2,2	-1,15
		X	CIBE_v1_4040	ID:29052937 Aliphatic sulfonates family ABC transporter, periplasmic ligand-binding protein	20	0,1	0,27	-0,5	0,84	1,7	-1,11
		X	CIBE_v1_5042	ID:29053939 Sulfate adenyltransferase	12	-1,8	-0,52	-0,3	-0,63	-3,8	1,15
		<i>cysD</i>	CIBE_v1_5043	ID:29053940 cysD sulfate adenyltransferase, subunit 2	10	-2,1	0,83	-1,5	0,29	-0,1	-1,11
		<i>sbo</i>	CIBE_v1_5049	ID:29053946 sbp sulfate transporter subunit ; periplasmic-binding component of ABC superfamily	10	-1,1	1,04	-0,6	-0,09	1,0	-0,95
	Nitrate	<i>sbp</i>	CIBE_v1_5050	ID:29053947 sbp sulfate transporter subunit ; periplasmic-binding component of ABC superfamily	14	0,5	1,04	1,4	-0,09	4,8	-0,95
		X	CIBE_v1_1234	ID:29050131 putative sulfonate/nitrate transport system substrate-binding protein	5	-0,6	0,52	-1,6	0,64	2,2	-1,15
	Phosphate	X	CIBE_v1_1013	ID:29049910 Nitrite and sulphite reductase 4Fe-4S region	16	-2,1	0,99	-0,9	0,01	-0,1	-1,01
		<i>pstS</i>	CIBE_v1_1363	ID:29050260 pstS phosphate ABC transporter	14	4,0	-0,65	3,5	-0,50	1,6	1,15
	Sodium	<i>nhaA</i>	CIBE_v1_3588	ID:29052485 nhaA Na(+)/H(+) antiporter NhaA 1	1	0,9	-0,85	0,5	-0,25	-0,4	1,10
		X	CIBE_v1_5213	ID:29054110 Sodium:dicarboxylate symporter	3	-2,6	-0,58	-2,8	-0,57	-0,5	1,15
	Molybdenum	<i>modA</i>	CIBE_v1_2424	ID:29051321 modA molybdate-binding lipoprotein	8	-3,2	1,12	1,0	-0,33	1,6	-0,79

Partie IV : Métabolisme sur substrat 1G, approches fermentaire et multi-omique

	X	CIBE_v1_b0001	ID:29048876 protein of unknown function	1	4,2	-0,31	5,7	-0,81	1,6	1,12
	X	CIBE_v1_b0003	ID:29048878 Glycosyl hydrolase family protein	3	-0,3	1,03	0,0	-0,05	1,5	-0,97
	X	CIBE_v1_b0007	ID:29048882 protein of unknown function	6	5,3	-0,83	4,0	-0,29	2,3	1,11
	X	CIBE_v1_b0008	ID:29048883 conserved protein of unknown function	12	4,0	-1,08	2,6	0,18	1,7	0,90
	X	CIBE_v1_b0010	ID:29048885 protein of unknown function	3	8,2	-0,94	5,8	-0,11	4,2	1,05
	X	CIBE_v1_b0012	ID:29048887 conserved protein of unknown function	1	0,7	1,14	2,0	-0,43	3,1	-0,71
Phage linéaire Φ6423	X	CIBE_v1_b0013	ID:29048888 protein of unknown function	7	4,7	-0,58	3,8	-0,57	-2,4	1,15
	X	CIBE_v1_b0014	ID:29048889 putative Major head protein	16	5,1	-0,67	4,2	-0,48	4,3	1,15
	X	CIBE_v1_b0015	ID:29048890 protein of unknown function	13	-1,1	0,84	-0,5	0,27	0,4	-1,11
	X	CIBE_v1_b0016	ID:29048891 putative connector protein	8	1,9	-0,40	2,1	-0,73	2,5	1,14
	X	CIBE_v1_b0017	ID:29048892 protein of unknown function	6	3,1	-0,42	3,6	-0,72	0,4	1,14
Eléments génétiques mobiles	X	CIBE_v1_b0018	ID:29048893 protein of unknown function	11	4,7	-0,60	4,4	-0,56	2,9	1,15
	X	CIBE_v1_b0019	ID:29048894 protein of unknown function	14	3,9	1,13	4,5	-0,78	5,3	-0,35
	X	CIBE_v1_b0020	ID:29048895 protein of unknown function	3	6,4	-0,61	6,3	-0,54	1,4	1,15
	X	CIBE_v2_p0002	ID:29048864 hypothetical protein	1	-4,3	-0,52	-4,8	1,15	-0,6	-0,63
	X	CIBE_v2_p0003	ID:29048865 putative cell wall binding protein	16	2,5	1,05	3,2	-0,10	2,3	-0,95
	<i>cloC</i>	CIBE_v2_p0004	ID:29048866 putative peptidase	17	0,4	1,15	3,0	-0,50	4,9	-0,65
Plasmide pNF1	X	CIBE_v2_p0005	ID:29048867 exported protein of unknown function	2	2,6	0,91	3,0	-1,07	3,6	0,16
	X	CIBE_v2_p0007	ID:29048869 hypothetical protein	1	-1,6	1,10	-0,6	-0,24	-0,1	-0,86
	<i>cloA</i>	CIBE_v2_p0012	ID:29048874 closticin	21	0,2	1,13	2,2	-0,38	3,4	-0,76
	<i>cloB</i>	CIBE_v2_p0013	ID:29048875 putative bacteriocin transport accessory protein	2	-0,6	1,14	1,1	-0,42	3,6	-0,72
Plasmide pNF2	X	CIBE_v2_p20001	ID:42905720 hypothetical protein	8	-5,3	1,15	-1,9	-0,66	0,4	-0,49
	X	CIBE_v2_p20004	ID:42905723 hypothetical protein	7	0,7	1,15	3,3	-0,50	4,5	-0,66

Transcription	<i>rpoA</i>	CIBE_v1_0198	ID:29049095 rpoA RNA polymerase (alpha subunit)	9	-1,6	1,09	-0,7	-0,21	0,3	-0,88	
	<i>rpoB</i>	CIBE_v1_0160	ID:29049057 rpoB RNA polymerase (beta subunit)	29	-3,3	1,15	-0,8	-0,47	-0,7	-0,68	
	<i>rpoC</i>	CIBE_v1_0161	ID:29049058 rpoC RNA polymerase (beta' subunit)	19	-2,6	0,88	-0,8	-1,09	-1,0	0,20	
	<i>rpoZ</i>	CIBE_v1_1378	ID:29050275 rpoZ DNA-directed RNA polymerase subunit omega	6	-2,8	1,08	-1,5	-0,18	0,1	-0,90	
	<i>sigA</i>	CIBE_v1_1039	ID:29049936 sigA RNA polymerase major sigma-43 factor (sigma-A)	12	-0,7	0,22	-1,0	0,87	1,5	-1,09	
	<i>rho</i>	CIBE_v1_0553	ID:29049450 rho Transcription termination factor Rho	1	-3,0	1,14	-3,8	-0,43	0,6	-0,71	
	<i>nusA</i>	CIBE_v1_1437	ID:29050334 nusA transcription translation coupling factor involved in Rho-dependent transcription termination	20	-1,6	0,94	-0,7	0,12	-0,1	-1,05	
	<i>nusB</i>	CIBE_v1_2029	ID:29050926 nusB N utilization substance protein B homolog	9	0,2	0,87	0,4	0,22	2,9	-1,09	
	<i>nusG</i>	CIBE_v1_0155	ID:29049052 nusG transcription antitermination factor	13	-1,5	-0,36	-1,4	1,13	-1,0	-0,77	
	<i>greA</i>	CIBE_v1_0114	ID:29049011 greA transcription elongation factor	15	2,5	0,49	2,2	0,66	4,4	-1,15	
	<i>greA</i>	CIBE_v1_0912	ID:29049809 greA Transcription elongation factor GreA	2	0,1	1,15	1,0	-0,69	0,7	-0,46	
Systèmes d'expression génétique	<i>rplK</i>	CIBE_v1_0156	ID:29049053 rplK ribosomal protein L11 (BL11)	6	-1,0	0,98	-0,6	0,04	1,5	-1,02	
	<i>rplA</i>	CIBE_v1_0157	ID:29049054 rplA ribosomal protein L1 (BL1)	11	1,7	-0,53	0,0	1,15	2,5	-0,63	
	<i>rplJ</i>	CIBE_v1_0158	ID:29049055 rplJ ribosomal protein L10 (BL5)	4	-0,7	0,45	-1,2	0,69	2,3	-1,15	
	<i>rplL</i>	CIBE_v1_0159	ID:29049056 rplL ribosomal protein L12 (BL9)	9	-0,3	-0,26	-0,9	1,10	0,9	-0,85	
	<i>rpsL</i>	CIBE_v1_0162	ID:29049059 rpsL ribosomal protein S12 (BS12)	1	5,3	-0,37	1,6	1,13	5,9	-0,77	
	<i>rpsG</i>	CIBE_v1_0163	ID:29049060 rpsG ribosomal protein S7 (BS7)	13	2,9	-0,75	2,0	-0,39	0,8	1,14	
	<i>rpsJ</i>	CIBE_v1_0166	ID:29049063 rpsJ ribosomal protein S10 (BS13)	6	1,0	1,04	1,7	-0,08	2,6	-0,96	
	<i>rplC</i>	CIBE_v1_0167	ID:29049064 rplC ribosomal protein L3 (BL3)	14	2,7	-0,16	1,9	1,07	4,1	-0,91	
	<i>rplD</i>	CIBE_v1_0168	ID:29049065 rplD ribosomal protein L4	9	3,2	-0,18	2,0	1,08	6,3	-0,90	
	<i>rplW</i>	CIBE_v1_0169	ID:29049066 rplW ribosomal protein L23	8	2,5	-0,50	0,3	1,15	4,8	-0,65	
	Protéines ribosomales	<i>rplB</i>	CIBE_v1_0170	ID:29049067 rplB ribosomal protein L2 (BL2)	17	2,4	0,65	2,2	0,51	4,3	-1,15
		<i>rpsS</i>	CIBE_v1_0172	ID:29049069 rpsS ribosomal protein S19 (BS19)	8	2,7	-0,50	1,6	1,15	2,8	-0,65
		<i>rplV</i>	CIBE_v1_0173	ID:29049070 rplV ribosomal protein L22 (BL17)	3	3,8	-0,44	1,4	1,14	6,5	-0,70
		<i>rpsC</i>	CIBE_v1_0174	ID:29049071 rpsC ribosomal protein S3 (BS3)	15	4,1	-0,45	2,6	1,15	4,4	-0,69
		<i>rplP</i>	CIBE_v1_0175	ID:29049072 rplP ribosomal protein L16	4	2,9	-0,24	1,8	1,10	5,2	-0,86
		<i>rpmC</i>	CIBE_v1_0176	ID:29049073 rpmC ribosomal protein L29	3	2,4	-0,65	-0,6	1,15	2,4	-0,50
		<i>rpsQ</i>	CIBE_v1_0177	ID:29049074 rpsQ ribosomal protein S17 (BS16)	2	4,1	-0,46	1,4	1,15	8,7	-0,68
		<i>rplNA</i>	CIBE_v1_0178	ID:29049075 rplNA ribosomal protein L14	7	3,3	0,37	2,8	0,76	6,7	-1,13
		<i>rplX</i>	CIBE_v1_0179	ID:29049076 rplX ribosomal protein L24 (BL23)	1	3,1	-0,36	1,2	1,13	6,9	-0,77
		<i>rplE</i>	CIBE_v1_0180	ID:29049077 rplE ribosomal protein L5 (BL6)	19	1,5	-0,34	0,3	1,13	2,8	-0,78
<i>rpsH</i>	CIBE_v1_0183	ID:29049080 rpsH ribosomal protein S8 (BS8)	8	2,1	-0,62	0,0	1,15	1,7	-0,54		
	<i>rplF</i>	CIBE_v1_0184	ID:29049081 rplF ribosomal protein L6 (BL8)	12	3,0	-0,40	0,9	1,14	6,9	-0,74	

Partie IV : Métabolisme sur substrat 1G, approches fermentaire et multi-omique

Systèmes d'expression génétique	Protéines ribosomales	<i>rplR</i>	CIBE_v1_0185	ID:29049082 rplR ribosomal protein L18	3	2,3	-0,05	1,3	1,02	3,7	-0,97
		<i>rpsE</i>	CIBE_v1_0186	ID:29049083 rpsE ribosomal protein S5	13	4,2	-0,56	0,8	1,15	4,8	-0,59
		<i>rpmD</i>	CIBE_v1_0187	ID:29049084 rpmD ribosomal protein L30 (BL27)	5	1,9	-0,74	-0,5	1,14	1,6	-0,40
		<i>rplO</i>	CIBE_v1_0188	ID:29049085 rplO ribosomal protein L15	8	3,6	-0,44	1,5	1,14	5,3	-0,70
		<i>rpsM</i>	CIBE_v1_0195	ID:29049092 rpsM ribosomal protein S13	8	2,8	0,56	2,7	0,59	4,4	-1,15
		<i>rpsK</i>	CIBE_v1_0196	ID:29049093 rpsK ribosomal protein S11 (BS11)	5	1,8	-0,30	0,6	1,12	3,0	-0,82
		<i>rpsD</i>	CIBE_v1_0197	ID:29049094 rpsD 30S ribosomal subunit protein S4	7	3,5	-0,36	2,5	1,13	3,9	-0,77
		<i>rplQ</i>	CIBE_v1_0199	ID:29049096 rplQ ribosomal protein L17 (BL15)	4	2,6	-0,26	0,9	1,10	5,8	-0,85
		<i>rplM</i>	CIBE_v1_0204	ID:29049101 rplM ribosomal protein L13	11	2,1	-0,21	1,2	1,09	3,5	-0,88
		<i>rpsI</i>	CIBE_v1_0205	ID:29049102 rpsI ribosomal protein S9	4	2,8	-0,01	2,3	1,01	4,0	-0,99
		<i>rplU</i>	CIBE_v1_0679	ID:29049576 rplU ribosomal protein L21 (BL20)	4	2,3	-0,53	1,2	1,15	3,3	-0,62
		<i>rpmA</i>	CIBE_v1_0681	ID:29049578 rpmA ribosomal protein L27 (BL24)	2	3,3	-0,53	0,7	1,15	4,1	-0,62
		<i>rpsT</i>	CIBE_v1_1001	ID:29049898 rpsT ribosomal protein S20 (BS20)	1	3,7	-0,45	1,2	1,15	6,5	-0,69
		<i>prmA</i>	CIBE_v1_1016	ID:29049913 prmA ribosomal protein L11 methyltransferase	4	1,4	0,15	1,7	-1,06	0,8	0,92
		<i>rplY</i>	CIBE_v1_1176	ID:29050073 rplY 50S ribosomal protein L25	10	5,5	-0,81	5,0	1,12	5,6	-0,31
		<i>rpmB</i>	CIBE_v1_1391	ID:29050288 rpmB ribosomal protein L28	1	8,8	-0,57	1,9	1,15	10,6	-0,58
		<i>rpsP</i>	CIBE_v1_1413	ID:29050310 rpsP ribosomal protein S16 (BS17)	6	3,7	-0,57	-0,3	1,15	4,2	-0,59
		<i>rplS</i>	CIBE_v1_1417	ID:29050314 rplS ribosomal protein L19	8	2,1	0,58	2,0	0,58	5,3	-1,15
		<i>rpsB</i>	CIBE_v1_1426	ID:29050323 rpsB ribosomal protein S2	9	1,2	-1,14	1,0	0,43	0,6	0,72
		<i>X</i>	CIBE_v1_1439	ID:29050336 Ribosomal protein L7Ae/L30e/S12e/Gadd45	4	-1,6	0,12	-1,2	0,93	-0,5	-1,06
		<i>rpsO</i>	CIBE_v1_1445	ID:29050342 rpsO ribosomal protein S15 (BS18)	6	3,9	-0,61	1,3	1,15	2,9	-0,55
		<i>rimO</i>	CIBE_v1_1452	ID:29050349 rimO Ribosomal protein S12 methylthiotransferase RimO	1	-1,3	0,96	1,1	0,08	1,4	-1,04
		<i>rplT</i>	CIBE_v1_1810	ID:29050707 rplT ribosomal protein L20	7	2,0	-1,02	1,1	0,98	2,3	0,04
		<i>rpsU</i>	CIBE_v1_3174	ID:29052071 rpsU ribosomal protein S21	1	1,8	0,47	2,3	0,67	4,1	-1,15
		<i>rplI</i>	CIBE_v1_6104	ID:29055001 rplI ribosomal protein L9	9	1,2	-0,55	-0,1	1,15	0,4	-0,61
		<i>rpsR</i>	CIBE_v1_6107	ID:29055004 rpsR ribosomal protein S18	2	3,2	-0,44	1,0	1,15	5,1	-0,70
<i>rpsF</i>	CIBE_v1_6109	ID:29055006 rpsF ribosomal protein S6 (BS9)	6	1,2	-1,06	-0,2	0,93	-0,1	0,13		
<i>rpmH</i>	CIBE_v1_6132	ID:29055029 rpmH ribosomal protein L34	4	5,3	-0,50	2,0	1,15	7,6	-0,65		
Maturation et modification des ARNr	<i>rnjA</i>	CIBE_v1_1344	ID:29050241 rnjA ribonuclease J1	13	1,2	-0,61	0,7	-0,55	-1,0	1,15	
	<i>mc</i>	CIBE_v1_1406	ID:29050303 mc ribonuclease III	4	1,3	1,14	2,8	-0,44	1,1	-0,70	
	<i>rluB</i>	CIBE_v1_2301	ID:29051198 rluB pseudouridine synthase	1	2,2	0,77	2,4	0,36	6,4	-1,13	
	<i>rsmF</i>	CIBE_v1_1340	ID:29050237 rsmF pre-16S ribosomal RNA maturation enzyme	5	-1,0	0,80	-0,2	0,32	1,1	-1,12	
	<i>rsmH</i>	CIBE_v1_1817	ID:29050714 rsmH 16S rRNA m4C1402 methyltransferase	4	-1,4	0,27	-1,7	0,83	0,1	-1,11	
<i>rsmG</i>	CIBE_v1_6125	ID:29055022 rsmG 7-methylguanosine methyltransferase	3	-3,1	1,01	-1,7	-0,02	-0,1	-0,99		

Maturation et modification des ARNr	X	CIBE_v1_1384	ID:29050281 16S rRNA methyltransferase	5	-1,9	0,91	-0,6	-1,07	-1,3	0,16	
	<i>rbfA</i>	CIBE_v1_1441	ID:29050338 rbfA pre-ribosomal (17S) RNA binding factor A	6	-0,3	0,28	-0,7	0,83	0,3	-1,11	
	<i>infA</i>	CIBE_v1_0193	ID:29049090 infA initiation factor IF-1	3	-1,3	1,03	-0,9	-0,06	0,7	-0,97	
	<i>infB</i>	CIBE_v1_1440	ID:29050337 infB initiation factor IF-2	18	-1,4	0,77	-0,8	0,36	0,0	-1,13	
	<i>infC</i>	CIBE_v1_1808	ID:29050705 infC initiation factor IF-3	3	3,1	-0,55	-0,2	1,15	3,6	-0,60	
	<i>tufA</i>	CIBE_v1_0152	ID:29049049 tufA elongation factor Tu	43	-4,5	1,15	-1,9	-0,45	-1,3	-0,70	
	<i>fusA</i>	CIBE_v1_0164	ID:29049061 fusA elongation factor G	44	-3,6	0,02	-3,1	-1,01	-4,3	0,99	
	<i>fusA</i>	CIBE_v1_3340	ID:29052237 Elongation factor G, domain IV	5	-3,6	1,15	-0,9	-0,50	-0,9	-0,65	
	<i>lepA</i>	CIBE_v1_1005	ID:29049902 lepA ribosomal elongation factor, GTPase	2	-3,2	1,11	0,4	-0,28	-0,7	-0,83	
	<i>tsf</i>	CIBE_v1_1427	ID:29050324 tsf elongation factor Ts	7	-0,8	-0,02	-1,3	1,01	-0,1	-0,99	
	<i>efp</i>	CIBE_v1_2018	ID:29050915 efp elongation factor P	12	-1,7	0,76	-1,0	0,37	0,0	-1,13	
	Facteurs de traduction	<i>prfA</i>	CIBE_v1_0558	ID:29049455 prfA peptide chain release factor 1	17	-1,5	0,36	-1,5	0,77	-0,8	-1,13
		<i>prfB</i>	CIBE_v1_0593	ID:29049490 prfB Peptide chain release factor 2	7	-1,5	0,77	-0,5	0,36	-1,0	-1,13
		<i>prfC</i>	CIBE_v1_5627	ID:29054524 prfC Peptide chain release factor 3	5	-5,6	1,01	-1,7	-0,03	0,0	-0,98
<i>defA</i>		CIBE_v1_1381	ID:29050278 defA peptide deformylase	8	0,0	-0,24	-0,4	1,10	1,4	-0,86	
<i>def</i>		CIBE_v1_4187	ID:29053084 def Peptide deformylase 2	2	0,0	1,10	2,4	-0,24	6,1	-0,86	
<i>def</i>		CIBE_v1_4224	ID:29053121 def Peptide deformylase	3	0,3	0,76	0,4	0,38	2,0	-1,13	
Systèmes d'expression génétique		<i>mapA</i>	CIBE_v1_2167	ID:29051064 Methionine aminopeptidase A	9	-1,1	0,81	-0,5	0,31	1,5	-1,12
	<i>mapA</i>	CIBE_v1_0191	ID:29049088 mapA methionine aminopeptidase	6	-1,6	0,11	-2,1	0,94	0,5	-1,05	
	<i>mapA</i>	CIBE_v1_4313	ID:29053210 Methionine aminopeptidase A	5	-1,2	0,98	-0,4	0,05	3,4	-1,02	
	<i>hpf</i>	CIBE_v1_0591	ID:29049488 hpf ribosome-associated sigma 54 modulation protein; ribosome dimerisation factor	2	3,0	-0,66	0,9	1,15	4,6	-0,49	
	X	CIBE_v1_3881	ID:29052778 Translation initiation inhibitor	1	2,3	0,77	2,4	0,36	6,7	-1,13	
	<i>ileS</i>	CIBE_v1_0049	ID:29048946 ileS Isoleucine--tRNA ligase	5	-0,4	0,86	-0,3	0,23	-0,8	-1,10	
	<i>metS</i>	CIBE_v1_0059	ID:29048956 metS methionyl-tRNA synthetase	3	-1,2	0,49	-0,1	0,66	2,8	-1,15	
	<i>asnS</i>	CIBE_v1_0083	ID:29048980 asnS asparaginyl tRNA synthetase	16	-0,1	0,72	0,2	0,43	0,9	-1,14	
	<i>lysS</i>	CIBE_v1_0115	ID:29049012 lysS lysyl-tRNA synthetase	20	-1,9	-0,02	-1,8	1,01	0,8	-0,99	
	<i>glyQS</i>	CIBE_v1_0117	ID:29049014 glyQS Glycine--tRNA ligase	10	-0,7	0,31	-0,7	0,81	0,4	-1,12	
Synthèse d'ARNt	<i>proS</i>	CIBE_v1_0146	ID:29049043 proS prolyl-tRNA synthetase	15	-0,9	0,81	-0,4	0,31	-1,6	-1,12	
	<i>cysS</i>	CIBE_v1_0147	ID:29049044 cysS cysteinyl-tRNA synthetase	12	-1,7	0,89	-1,8	0,19	-0,4	-1,08	
	<i>thrS</i>	CIBE_v1_0295	ID:29049192 thrS Threonine--tRNA ligase	10	-0,7	1,01	-0,3	-0,02	1,7	-0,99	
	<i>thrZ</i>	CIBE_v1_3846	ID:29052743 thrZ threonyl-tRNA synthetase	4	-0,5	1,15	0,4	-0,69	0,0	-0,46	
	<i>trpS</i>	CIBE_v1_0976	ID:29049873 trpS Tryptophan--tRNA ligase	7	0,0	0,23	0,4	0,86	1,7	-1,10	
	<i>argS</i>	CIBE_v1_1214	ID:29050111 argS Arginine--tRNA ligase	7	-1,3	1,01	0,2	-0,03	0,3	-0,99	
	<i>alaS</i>	CIBE_v1_1338	ID:29050235 alaS alanyl-tRNA synthetase	12	-2,0	0,92	-1,9	0,14	0,1	-1,06	

Partie IV : Métabolisme sur substrat 1G, approches fermentaire et multi-omique

Systèmes d'expression génétique	Synthèse d'ARNt	<i>gltX</i>	CIBE_v1_1648	ID:29050545 <i>gltX</i> Glutamate--tRNA ligase	14	-1,3	1,00	-0,9	0,00	0,6	-1,00
		<i>serS</i>	CIBE_v1_1655	ID:29050552 <i>serS</i> Serine--tRNA ligase	11	-0,4	0,48	-0,7	0,67	0,3	-1,15
		<i>hisS</i>	CIBE_v1_1781	ID:29050678 <i>hisS</i> histidyl-tRNA synthetase	6	-0,5	0,92	0,7	0,14	1,8	-1,06
		<i>aspS</i>	CIBE_v1_1782	ID:29050679 <i>aspS</i> aspartyl-tRNA synthetase	32	-2,1	1,09	-0,6	-0,87	-2,1	-0,22
		<i>aspS</i>	CIBE_v1_2562	ID:29051459 <i>aspS</i> Aspartate--tRNA ligase	5	-0,7	1,00	0,3	-1,00	-0,3	0,00
		<i>pheS</i>	CIBE_v1_1813	ID:29050710 <i>pheS</i> phenylalanyl-tRNA synthetase (alpha subunit)	2	3,0	-0,62	2,1	-0,53	0,2	1,15
		<i>pheT</i>	CIBE_v1_1814	ID:29050711 <i>pheT</i> Phenylalanine--tRNA ligase beta subunit	6	0,6	0,78	0,8	0,35	1,1	-1,13
		<i>valS</i>	CIBE_v1_2124	ID:29051021 <i>valS</i> valyl-tRNA synthetase	4	-0,8	-0,57	-0,6	-0,58	-5,0	1,15
		<i>glnS</i>	CIBE_v1_1650	ID:29050547 <i>glnS</i> glutamyl-tRNA synthetase	8	0,0	1,08	1,0	-0,18	1,2	-0,90
		<i>gatC</i>	CIBE_v1_2563	ID:29051460 <i>gatC</i> Glutamyl-tRNA(Gln) amidotransferase subunit C 2	2	-1,0	1,14	1,1	-0,41	1,1	-0,73
		<i>gatA</i>	CIBE_v1_2564	ID:29051461 <i>gatA</i> glutamyl-tRNA(Gln) amidotransferase (subunit A)	9	-1,4	1,12	-0,3	-0,33	1,6	-0,80
		<i>gatB</i>	CIBE_v1_2565	ID:29051462 <i>gatB</i> glutamyl-tRNA(Gln) amidotransferase (subunit B)	5	-2,0	0,82	-1,3	0,29	0,6	-1,11
		<i>leuS</i>	CIBE_v1_5495	ID:29054392 <i>leuS</i> Leucine--tRNA ligase	6	-1,9	1,00	-0,9	0,00	0,3	-1,00
<i>tyrS</i>	CIBE_v1_5521	ID:29054418 <i>tyrS</i> tyrosyl-tRNA synthetase	8	-0,9	0,57	0,6	0,59	0,4	-1,15		

Métabolisme des acides aminés	Famille du glutamate (glu, gln, arg, pro)	X	CIBE_v1_1185	ID:29050082 Formiminotransferase-cyclodeaminase [Clostridium beijerinckii DSM6423]	3	-5,8	1,15	-2,7	-0,53	-0,4	-0,62
		<i>gltB</i>	CIBE_v1_5062	ID:29053959 <i>gltB</i> glutamate synthase (small subunit) [Clostridium beijerinckii DSM6423]	21	-0,1	0,20	-0,4	0,88	1,4	-1,09
		<i>gltA</i>	CIBE_v1_5063	ID:29053960 <i>gltA</i> glutamate synthase (large subunit) [Clostridium beijerinckii DSM6423]	35	-0,6	0,89	0,0	0,19	0,9	-1,08
		<i>aspB</i>	CIBE_v1_5183	ID:29054080 <i>aspB</i> putative aspartate aminotransferase [Clostridium beijerinckii DSM6423]	12	-1,9	0,02	-0,3	0,99	-2,3	-1,01
		<i>yugH</i>	CIBE_v1_4998	ID:29053895 <i>yugH</i> putative aspartate aminotransferase [Clostridium beijerinckii DSM6423]	3	-0,9	1,07	0,4	-0,16	-0,5	-0,91
		<i>yhdR</i>	CIBE_v1_5725	ID:29054622 <i>yhdR</i> putative aspartate aminotransferase YhdR [Clostridium beijerinckii DSM6423]	1	0,5	1,13	2,2	-0,35	3,0	-0,78
		<i>glnA</i>	CIBE_v1_5065	ID:29053962 <i>glnA</i> glutamine synthetase [Clostridium beijerinckii DSM6423]	14	0,4	0,36	0,8	0,77	2,0	-1,13
		<i>glnA</i>	CIBE_v1_0604	ID:29049501 <i>glnA</i> Glutamine synthetase [Clostridium beijerinckii DSM6423]	20	-2,7	1,01	-1,7	-0,02	1,2	-0,99
		<i>pabA</i>	CIBE_v1_4972	ID:29053869 <i>pabA</i> 4-amino-4-deoxychorismate synthase; anthranilate synthase (subunit II) [Clostridium beijerinckii DSM6423]	1	0,2	1,15	2,7	-0,62	3,6	-0,53
		<i>proB</i>	CIBE_v1_0028	ID:29048925 <i>proB</i> Glutamate 5-kinase [Clostridium beijerinckii DSM6423]	4	-1,2	1,11	1,3	-0,28	4,0	-0,83
		<i>proA</i>	CIBE_v1_0029	ID:29048926 <i>proA</i> gamma-glutamyl phosphate reductase	7	-1,4	-0,61	-1,0	-0,55	-3,2	1,15
		<i>argI</i>	CIBE_v1_2886	ID:29051783 <i>argI</i> ornithine carbamoyltransferase 1 [Clostridium beijerinckii DSM6423]	5	-0,9	1,02	0,3	-0,05	1,7	-0,97
		<i>argG</i>	CIBE_v1_5421	ID:29054318 <i>argG</i> argininosuccinate synthase [Clostridium beijerinckii DSM6423]	7	0,5	-0,57	-0,4	1,15	1,1	-0,59

Famille du glutamate (glu, gln, arg, pro)	<i>argC</i>	CIBE_v1_5423	ID:29054320 <i>argC</i>] N-acetylglutamate gamma-semialdehyde dehydrogenase [Clostridium beijerinckii DSM6423]	4	0,0	1,08	0,6	-0,18	0,6	-0,90	
	<i>argJ</i>	CIBE_v1_5424	ID:29054321 <i>argJ</i>] ornithine acetyltransferase; amino-acid acetyltransferase [Clostridium beijerinckii DSM6423]	4	0,4	1,05	0,7	-0,12	1,8	-0,94	
	<i>argD</i>	CIBE_v1_5426	ID:29054323 <i>argD</i>] N-acetylornithine aminotransferase [Clostridium beijerinckii DSM6423]	1	-3,1	1,15	1,9	-0,58	1,1	-0,58	
	<i>argF</i>	CIBE_v1_5427	ID:29054324 <i>argF</i>] ornithine carbamoyltransferase [Clostridium beijerinckii DSM6423]	3	-0,9	1,15	-0,3	-0,45	0,0	-0,70	
Famille de la sérine (ser, gly, cys)	X	CIBE_v1_5955	ID:29054852 3-phosphoglycerate dehydrogenase [Clostridium beijerinckii DSM6423]	17	-1,1	0,70	-1,6	0,44	0,9	-1,15	
	<i>serC</i>	CIBE_v1_0082	ID:29048979 <i>serC</i>] Phosphoserine aminotransferase [Clostridium beijerinckii DSM6423]	32	-0,4	-0,39	-1,5	1,14	0,2	-0,75	
	X	CIBE_v1_0922	ID:29049819 Serine--pyruvate transaminase [Clostridium beijerinckii DSM6423]	9	-0,3	-0,31	-1,7	1,12	3,1	-0,81	
	<i>cysK</i>	CIBE_v1_0748	ID:29049645 <i>cysK</i>] cysteine synthase [Clostridium beijerinckii DSM6423]	19	-2,2	0,86	-1,6	0,24	0,9	-1,10	
	<i>cysK</i>	CIBE_v1_5226	ID:29054123 <i>cysK</i>] cysteine synthase [Clostridium beijerinckii DSM6423]	21	-0,5	0,62	-0,8	0,53	1,2	-1,15	
	<i>glyA</i>	CIBE_v1_2212	ID:29051109 <i>glyA</i>] serine hydroxymethyltransferase [Clostridium beijerinckii DSM6423]	21	0,6	1,00	2,4	0,00	3,1	-1,00	
	X	CIBE_v1_5004	ID:29053901 Threonine aldolase [Clostridium beijerinckii DSM6423]	10	-2,3	0,76	-2,7	0,37	-0,7	-1,13	
Métabolisme des acides aminés	Famille du pyruvate (ala, val, leu)	<i>leuA</i>	CIBE_v1_0075	ID:29048972 <i>leuA</i>] 2-isopropylmalate synthase 2 [Clostridium beijerinckii DSM6423]	9	2,1	-0,14	0,8	1,06	1,8	-0,92
		<i>leuB</i>	CIBE_v1_0236	ID:29049133 <i>leuB</i>] 3-isopropylmalate dehydrogenase [Clostridium beijerinckii DSM6423]	7	0,3	1,11	2,3	-0,30	2,2	-0,82
		<i>leuC</i>	CIBE_v1_0234	ID:29049131 <i>leuC</i>] 3-isopropylmalate dehydratase large subunit [Clostridium beijerinckii DSM6423]	10	-0,3	1,10	1,9	-0,24	1,0	-0,86
		<i>leuD</i>	CIBE_v1_0235	ID:29049132 <i>leuD</i>] 3-isopropylmalate dehydratase small subunit [Clostridium beijerinckii DSM6423]	4	-0,1	1,06	2,0	-0,15	0,1	-0,92
		X	CIBE_v1_1799	ID:29050696 Threonine dehydratase	9	-2,2	1,12	-0,8	-0,30	0,9	-0,82
		<i>ilvD</i>	CIBE_v1_1743	ID:29050640 <i>ilvD</i>] dihydroxy-acid dehydratase [Clostridium beijerinckii DSM6423]	14	0,8	1,04	2,1	-0,09	1,3	-0,95
		<i>ilvB</i>	CIBE_v1_0237	ID:29049134 <i>ilvB</i>] acetolactate synthase (acetohydroxy-acid synthase) (large subunit) [Clostridium beijerinckii DSM6423]	1	-1,3	1,15	4,2	-0,45	0,0	-0,69
		<i>ilvC</i>	CIBE_v1_0238	ID:29049135 <i>ilvC</i>] acetohydroxy-acid isomeroeductase [Clostridium beijerinckii DSM6423]	16	-0,3	1,13	1,4	-0,34	3,7	-0,78
		<i>ilvH</i>	CIBE_v1_0232	ID:29049129 <i>ilvH</i>] acetolactate synthase (acetohydroxy-acid synthase) (small subunit) [Clostridium beijerinckii DSM6423]	3	0,4	-0,54	1,9	-0,62	-4,6	1,15
		<i>ilvE</i>	CIBE_v1_1240	ID:29050137 <i>ilvE</i>] Branched-chain-amino-acid aminotransferase [Clostridium beijerinckii DSM6423]	18	-2,1	1,10	-0,6	-0,23	0,8	-0,86
Famille des aromatiques (phe, tyr, trp)	X	CIBE_v1_5481	ID:29054378 Prephenate dehydrogenase [Clostridium beijerinckii DSM6423]	4	-2,2	0,11	-2,3	-1,05	-2,6	0,94	
	<i>aroF</i>	CIBE_v1_5482	ID:29054379 <i>aroF</i>] Phospho-2-dehydro-3-deoxyheptonate aldolase [Clostridium beijerinckii DSM6423]	11	-2,9	-0,13	-1,9	1,06	-3,7	-0,93	
	<i>aroQ</i>	CIBE_v1_5483	ID:29054380 <i>aroQ</i>] 3-dehydroquinate dehydratase [Clostridium beijerinckii DSM6423]	9	-3,0	1,14	-1,7	-0,40	-1,2	-0,74	
	<i>aroK</i>	CIBE_v1_5484	ID:29054381 <i>aroK</i>] Shikimate kinase [Clostridium beijerinckii DSM6423]	2	-4,7	1,15	-2,8	-0,51	-2,2	-0,64	
	<i>aroE</i>	CIBE_v1_5485	ID:29054382 <i>aroE</i>] Shikimate dehydrogenase [Clostridium beijerinckii DSM6423]	11	-3,5	1,12	-3,1	-0,33	-1,7	-0,80	
	X	CIBE_v1_5486	ID:29054383 Bifunctional P-protein, chorismate mutase/prephenate dehydratase [Clostridium beijerinckii DSM6423]	7	-5,2	1,15	-3,9	-0,55	-2,1	-0,60	
	<i>aroC</i>	CIBE_v1_5487	ID:29054384 <i>aroC</i>] Chorismate synthase [Clostridium beijerinckii DSM6423]	11	-2,1	-0,06	-4,1	1,03	-2,5	-0,97	
<i>aroA</i>	CIBE_v1_5488	ID:29054385 <i>aroA</i>] 3-phosphoshikimate 1-carboxyvinyltransferase [Clostridium beijerinckii DSM6423]	10	-4,1	0,44	-3,3	-1,14	-3,5	0,70		

Partie IV : Métabolisme sur substrat 1G, approches fermentaire et multi-omique

Famille des aromatiques (phe, tyr, trp)	<i>aroB</i>	CIBE_v1_5489	ID:29054386 <i>aroB</i>] 3-dehydroquinate synthase [Clostridium bejerinckii DSM6423]	16	-3,0	1,14	-1,4	-0,41	0,6	-0,73
	<i>aroF</i>	CIBE_v1_5490	ID:29054387 <i>aroF</i>] Phospho-2-dehydro-3-deoxyheptonate aldolase [Clostridium bejerinckii DSM6423]	14	-1,1	1,12	-2,3	-0,30	-3,1	-0,82
	<i>X</i>	CIBE_v1_2878	ID:29051775 Phospho-2-dehydro-3-deoxyheptonate aldolase [Clostridium bejerinckii DSM6423]	13	-0,7	-0,66	-1,3	1,15	0,2	-0,49
	<i>trpE</i>	CIBE_v1_2078	ID:29050975 <i>trpE</i>] Anthranilate synthase component 1 [Clostridium bejerinckii DSM6423]	10	0,0	1,10	2,1	-0,25	3,4	-0,85
	<i>trpD</i>	CIBE_v1_2080	ID:29050977 <i>trpD</i>] anthranilate phosphoribosyltransferase [Clostridium bejerinckii DSM6423]	2	-1,6	0,74	-1,1	0,40	1,1	-1,14
	<i>trpC</i>	CIBE_v1_2081	ID:29050978 <i>trpC</i>] Indole-3-glycerol phosphate synthase [Clostridium bejerinckii DSM6423]	4	3,6	1,15	6,9	-0,55	7,0	-0,61
	<i>trpF</i>	CIBE_v1_2082	ID:29050979 <i>trpF</i>] N-(5'-phosphoribosyl)anthranilate isomerase [Clostridium bejerinckii DSM6423]	3	0,2	1,06	1,1	-0,13	2,5	-0,93
	<i>trpB</i>	CIBE_v1_2083	ID:29050980 <i>trpB</i>] tryptophan synthase (beta subunit) [Clostridium bejerinckii DSM6423]	15	0,8	0,94	1,9	0,11	2,3	-1,05
	<i>trpA</i>	CIBE_v1_2084	ID:29050981 <i>trpA</i>] tryptophan synthase (alpha subunit) [Clostridium bejerinckii DSM6423]	4	-0,4	1,15	3,5	-0,51	3,6	-0,64
Métabolisme des acides aminés	<i>X</i>	CIBE_v1_5067	ID:29053964 Aspartokinase [Clostridium bejerinckii DSM6423]	7	-3,0	-0,59	-2,4	1,15	-2,6	-0,57
	<i>hom</i>	CIBE_v1_1787	ID:29050684 <i>hom</i>] Homoserine dehydrogenase [Clostridium bejerinckii DSM6423]	8	-0,4	0,53	0,6	0,62	4,3	-1,15
	<i>X</i>	CIBE_v1_1546	ID:29050443 Threonine synthase [Clostridium bejerinckii DSM6423]	12	-2,1	0,34	-0,9	0,78	-0,3	-1,13
	<i>metAA</i>	CIBE_v1_1530	ID:29050427 <i>metAA</i>] homoserine O-acetyltransferase [Clostridium bejerinckii DSM6423]	19	-2,9	1,07	-1,3	-0,15	0,0	-0,92
	<i>X</i>	CIBE_v1_0943	ID:29049840 O-acetylhomoserine aminocarboxypropyltransferase [Clostridium bejerinckii DSM6423]	21	-3,6	0,79	-3,2	0,34	-0,9	-1,13
	<i>metC</i>	CIBE_v1_0261	ID:29049158 <i>metC</i>] cystathionine beta-lyase [Clostridium bejerinckii DSM6423]	9	-1,5	-0,19	-1,9	1,08	-0,7	-0,89
	<i>yxjH</i>	CIBE_v1_2751	ID:29051648 <i>yxjH</i>] putative methyl-tetrahydrofolate methyltransferase [Clostridium bejerinckii DSM6423]	6	1,1	-0,29	2,0	-0,83	-0,9	1,11
	<i>metH</i>	CIBE_v1_3684	ID:29052581 <i>metH</i>] homocysteine-N5-methyltetrahydrofolate transmethylase, B12-dependent [Clostridium bejerinckii DSM6423]	4	-1,2	0,45	-1,2	0,70	-1,0	-1,15
	<i>lysA</i>	CIBE_v1_5716	ID:29054613 <i>lysA</i>] diaminopimelate decarboxylase [Clostridium bejerinckii DSM6423]	8	-3,9	1,15	-0,7	-0,54	0,7	-0,61
	<i>dapA</i>	CIBE_v1_5697	ID:29054594 <i>dapA</i>] 4-hydroxy-tetrahydrodipicolinate synthase [Clostridium bejerinckii DSM6423]	19	-1,0	0,91	-0,3	0,16	2,0	-1,07
	<i>dapB</i>	CIBE_v1_2137	ID:29051034 <i>dapB</i>] 4-hydroxy-tetrahydrodipicolinate reductase [Clostridium bejerinckii DSM6423]	11	-0,2	0,99	0,5	0,02	0,5	-1,01
	<i>dapA</i>	CIBE_v1_2138	ID:29051035 <i>dapA</i>] dihydrodipicolinate synthase [Clostridium bejerinckii DSM6423]	5	-0,9	-0,32	-1,2	1,12	0,5	-0,80
<i>asd</i>	CIBE_v1_2139	ID:29051036 <i>asd</i>] aspartate-semialdehyde dehydrogenase [Clostridium bejerinckii DSM6423]	25	-2,5	1,06	-1,6	-0,13	-0,9	-0,93	
<i>asd</i>	CIBE_v1_0689	ID:29049586 <i>asd</i>] Aspartate-semialdehyde dehydrogenase [Clostridium bejerinckii DSM6423]	2	-1,0	1,09	1,7	-0,87	2,8	-0,22	
<i>ansB</i>	CIBE_v1_2830	ID:29051727 <i>ansB</i>] L-asparaginase 2 [Clostridium bejerinckii DSM6423]	8	-2,8	0,99	-2,5	0,03	-1,4	-1,01	
Biosynthèse de l'histidine	<i>hisB</i>	CIBE_v1_1572	ID:29050469 <i>hisB</i>] imidazoleglycerol-phosphate dehydratase [Clostridium bejerinckii DSM6423]	1	1,8	-0,46	2,4	-0,69	-0,9	1,15
	<i>hisH</i>	CIBE_v1_1573	ID:29050470 <i>hisH</i>] amidotransferase (glutaminase) [Clostridium bejerinckii DSM6423]	3	1,3	1,07	1,1	-0,90	1,7	-0,17
	<i>hisA</i>	CIBE_v1_1574	ID:29050471 <i>hisA</i>] phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase [Clostridium bejerinckii DSM6423]	6	0,9	1,01	1,7	-0,02	0,3	-0,99
	<i>hisF</i>	CIBE_v1_1575	ID:29050472 <i>hisF</i>] imidazole glycerol phosphate synthase subunit [Clostridium bejerinckii DSM6423]	5	0,5	0,97	0,7	0,05	2,2	-1,02

Biosynthèse de l'histidine	<i>hisI</i>	CIBE_v1_1576	ID:29050473 hisI fragment of phosphoribosyl-AMP cyclohydrolase; phosphoribosyl-ATP pyrophosphohydrolase (part 1) [Clostridium beijerinckii DSM6423]	1	-0,3	1,02	0,7	-0,04	-1,0	-0,98	
	<i>hisI</i>	CIBE_v1_1577	ID:29050474 hisI fragment of phosphoribosyl-AMP cyclohydrolase; phosphoribosyl-ATP pyrophosphohydrolase (part 2) [Clostridium beijerinckii DSM6423]	1	2,0	0,69	2,1	0,46	1,9	-1,15	
	X	CIBE_v1_1167	ID:29050064 Histidinol phosphate phosphatase [Clostridium beijerinckii DSM6423]	4	-5,2	1,15	-1,9	-0,50	2,7	-0,65	
	<i>hisZ</i>	CIBE_v1_1568	ID:29050465 hisZ ATP phosphoribosyltransferase regulatory subunit [Clostridium beijerinckii DSM6423]	10	2,6	0,66	3,3	0,49	-0,3	-1,15	
	<i>hisD</i>	CIBE_v1_1570	ID:29050467 hisD histidinol dehydrogenase [Clostridium beijerinckii DSM6423]	17	1,7	1,00	2,6	0,00	1,3	-1,00	
	<i>hisC</i>	CIBE_v1_2155	ID:29051052 hisC Histidinol-phosphate aminotransferase [Clostridium beijerinckii DSM6423]	3	-1,6	-0,22	0,6	-0,87	-3,1	1,09	
	<i>glnH</i>	CIBE_v1_2826	ID:29051723 glnH glutamine transporter subunit ; periplasmic binding component of ABC superfamily [Clostridium beijerinckii DSM6423]	14	-1,6	0,74	-1,0	0,40	2,0	-1,14	
	<i>glnH</i>	CIBE_v1_2827	ID:29051724 glnH glutamine transporter subunit ; periplasmic binding component of ABC superfamily [Clostridium beijerinckii DSM6423]	7	-1,4	0,53	-1,9	0,62	1,2	-1,15	
	<i>artR</i>	CIBE_v1_5708	ID:29054605 artR high affinity arginine ABC transporter (ATP-binding protein) [Clostridium beijerinckii DSM6423]	1	-1,6	1,15	-0,7	-0,52	-1,4	-0,63	
	<i>artQ</i>	CIBE_v1_5709	ID:29054606 artQ high affinity arginine ABC transporter (permease) [Clostridium beijerinckii DSM6423]	3	-1,2	1,12	0,2	-0,81	-0,1	-0,30	
Métabolisme des acides aminés	<i>artR</i>	CIBE_v1_2958	ID:29051855 artR high affinity arginine ABC transporter (ATP-binding protein) [Clostridium beijerinckii DSM6423]	3	-2,7	1,15	-0,5	-0,55	1,7	-0,61	
	<i>artR</i>	CIBE_v1_2829	ID:29051726 artR high affinity arginine ABC transporter (ATP-binding protein) [Clostridium beijerinckii DSM6423]	6	-3,6	1,03	-1,9	-0,07	-0,7	-0,96	
	X	CIBE_v1_2103	ID:29051000 Branched-chain amino acid ABC transporter substrate-binding protein [Clostridium beijerinckii DSM6423]	11	-0,8	-0,92	-0,8	1,07	-0,9	-0,15	
	<i>livF</i>	CIBE_v1_2107	ID:29051004 livF leucine/isoleucine/valine transporter subunit ; ATP-binding component of ABC superfamily [Clostridium beijerinckii DSM6423]	8	-5,7	0,91	-3,7	0,16	-3,0	-1,07	
	X	CIBE_v1_2108	ID:29051005 Branched-chain amino acid ABC transporter substrate-binding protein [Clostridium beijerinckii DSM6423]	7	-1,4	-1,06	-1,9	0,14	-1,8	0,92	
	Transporteurs	<i>livF</i>	CIBE_v1_6061	ID:29054958 livF leucine/isoleucine/valine transporter subunit ; ATP-binding component of ABC superfamily [Clostridium beijerinckii DSM6423]	2	-1,4	0,96	0,2	0,08	0,9	-1,04
		X	CIBE_v1_2957	ID:29051854 Polar amino acid ABC transporter, inner membrane subunit [Clostridium beijerinckii DSM6423]	1	-3,5	0,79	-4,4	0,33	-0,5	-1,12
		X	CIBE_v1_2959	ID:29051856 Amino acid ABC transporter substrate-binding protein [Clostridium beijerinckii DSM6423]	20	-0,5	-0,06	-0,8	1,03	1,8	-0,97
		X	CIBE_v1_2828	ID:29051725 Amino acid ABC transporter permease [Clostridium beijerinckii DSM6423]	5	-4,3	1,08	-3,4	-0,19	-0,1	-0,89
		X	CIBE_v1_3409	ID:29052306 ABC transporter permease [Clostridium beijerinckii DSM6423]	19	-0,3	1,11	-0,6	-0,29	3,8	-0,82
<i>opuCA</i>		CIBE_v1_3410	ID:29052307 opuCA glycine betaine/carnitine/choline/choline sulfate ABC transporter (ATP-binding protein) [Clostridium beijerinckii DSM6423]	11	-1,8	1,10	-1,0	-0,24	0,6	-0,86	
X		CIBE_v1_4245	ID:29053142 Glycine/betaine ABC transporter substrate-binding protein [Clostridium beijerinckii DSM6423]	5	0,6	-0,25	0,3	1,10	3,7	-0,85	
X		CIBE_v1_5220	ID:29054117 ABC transporter substrate-binding protein [Clostridium beijerinckii DSM6423]	1	-0,3	-0,49	-3,4	1,15	4,1	-0,66	
<i>metN</i>		CIBE_v1_0802	ID:29049699 metN methionine ABC transporter (ATP-binding protein) [Clostridium beijerinckii DSM6423]	5	-3,5	1,11	-1,9	-0,27	0,2	-0,84	
<i>metP</i>		CIBE_v1_0803	ID:29049700 metP methionine ABC transporter, permease component [Clostridium beijerinckii DSM6423]	2	-2,0	1,05	-0,8	-0,12	1,5	-0,93	
<i>metQ</i>	CIBE_v1_0804	ID:29049701 metQ methionine ABC transporter, substrate binding lipoprotein [Clostridium beijerinckii DSM6423]	12	-0,3	0,29	-0,6	0,82	1,8	-1,11		

7 DISCUSSION GENERALE

Les travaux de ma thèse s'inscrivent dans un projet plus global d'IFPEN visant à mettre au point, à l'aide d'une souche de *Clostridium* solvantogène, un bioprocédé de production de solvants en général et d'isopropanol en particulier. Différents obstacles limitent le développement de ce procédé, et plusieurs sont liés aux performances des souches naturelles (faible sélectivité pour l'isopropanol, titres peu élevés, toxicité des produits). Les solutions pour pallier ces contraintes reposent sur deux axes principaux : une amélioration du bioprocédé et/ou du microorganisme lui-même.

C. beijerinckii DSM 6423 est un candidat prometteur pour l'établissement d'un bioprocédé IBE, grâce notamment à sa capacité naturelle rare à produire ce mélange de solvants. Mais l'amélioration de ses performances est freinée par i) le peu de connaissances que nous avons à son sujet et ii) l'impossibilité de modifier la souche par une approche dirigée. Ces deux points ont été abordés au cours de ma thèse par diverses approches.

Une partie de mes travaux a donc consisté à caractériser *C. beijerinckii* DSM 6423 de différentes manières. La solvantogénèse est un processus biologique connu de longue date, mais sa régulation n'est que partiellement élucidée²¹¹. La description du rôle du facteur σ^{54} sur l'expression de plusieurs voies métaboliques dont celles impliquées dans la solvantogénèse chez *C. beijerinckii*, apporte ainsi de nouveaux éléments de compréhension. D'autres questions à ce sujet sont désormais ouvertes : est-ce que la régulation de l'expression par σ^{54} intervient dans le passage d'une phase acidogène à une phase solvantogène ? Si oui, quelles sont les voies de signalisation impliquées ? Y répondre demandera une étude plus approfondie passant notamment par la caractérisation des EBP impliqués, pour tenter de lever le voile entourant la régulation des voies solvantogènes pour cette espèce, ainsi que pour d'autres solvantogènes. Depuis la publication de nos résultats, une autre équipe s'est en effet intéressée à la distribution des régulons associés à σ^{54} parmi les Clostridiales, confirmant nos résultats et montrant que le rôle de σ^{54} pour la production des alcools était conservé pour une variété d'espèces²⁷².

L'impact crucial du facteur σ^{54} sur le métabolisme central de *C. beijerinckii* souligne le rôle clé de la régulation du processus d'initiation de la transcription pour l'expression génétique chez les procaryotes. Dans ce contexte, déterminer les sites d'initiation de la transcription et, de fait les séquences promotrices de certains gènes d'intérêt peut apporter des informations importantes pour mieux appréhender les réseaux de régulation génétique. Grâce à une méthode dérivée du *Capp-Switch sequencing*²¹⁸, des cartes définissant à l'échelle du génome les sites d'initiation de la transcription pour *C. beijerinckii* DSM 6423 ainsi que pour deux autres souches solvantogènes ont été établies au cours de ma thèse. Si l'analyse comparative de la conservation des promoteurs dans ces trois microorganismes est encore en cours, les résultats préliminaires mettent en lumière des similarités claires entre ces derniers, avec toutefois quelques différences significatives. En particulier, la structure opéronique des gènes de synthèse du butyryl-CoA, longtemps considérée comme conservée entre *C. acetobutylicum* et *C. beijerinckii*, se révèle en réalité sensiblement différente chez ce dernier. La structure opéronique / sous-opéronique dont dépend l'expression de *hbd* est en effet claire, avec des implications possibles sur la régulation d'une répartition C3-C4 des solvants produits chez *C. beijerinckii*. Cette information pourrait être utilisée pour, par exemple, moduler le niveau d'expression de *hbd* et ainsi potentiellement augmenter la sélectivité pour l'isopropanol par génie génétique.

La structure génétique différentielle de l'opéron de synthèse du butyryl-CoA est un exemple pertinent des apports que peuvent fournir les analyses omiques. L'acquisition massive de données bioinformatiques qu'elles permettent est devenue incontournable pour caractériser un organisme. Dans cette optique, les données obtenues sur les TSS ont donc été complétées par une série d'expérimentations visant à mieux comprendre les régulations mises en place entre les phases acidogènes et solvantogènes et entre différentes sources carbonées. Pour cela, un procédé de fermentation continue distinguant ces deux phases a été employé pour la première fois chez *C. beijerinckii* DSM 6423. Permettant ainsi de discriminer efficacement les deux phases métaboliques et de les coupler à une série d'analyses omiques (TSS, RNA-seq, LC-MS/MS). Si une partie de ces analyses est encore en cours, la protéomique révèle de premiers résultats intéressants. L'acquisition des données sur substrat modèle (glucose, saccharose) confirme dans l'ensemble les changements d'expression acidogène/solvantogène obtenus avec les analyses *batch* précédentes⁵²,

tout en mettant en lumière quelques différences (dans le métabolisme central et les voies liées à la sporulation, notamment). De plus, les analyses préliminaires suggèrent que les voies de consommation du saccharose sont sensiblement différentes de celles de la souche NCIMB 8052, ce qui souligne une nouvelle différence entre ces deux microorganismes. L'analyse du schéma d'expression protéique en conditions modèles nous a par ailleurs permis de tirer de premières conclusions vis-à-vis de l'effet d'une source 1G, la mélasse de betterave, sur le métabolisme. A fort taux de dilution (i.e. acidogénèse), la toxicité du substrat provoque probablement une inhibition de croissance, alors qu'à faible taux de dilution (i.e. solvantogénèse), la souche dispose de plus de temps pour détoxifier le milieu et se rapproche ainsi des conditions modèles. Maintenant que les conditions fermentaires ont été fixées, une telle analyse, relativement simple à mettre en œuvre, pourrait être répétée pour une multitude de sources carbonées différentes, modèles ou pas, sous réserve d'une compatibilité avec la fermentation continue (c'est-à-dire pour des substrats liquéfiables).

Parallèlement à tous ces efforts de caractérisation de la souche, une deuxième partie de ma thèse a consisté à adapter pour cette dernière un outil de génie génétique initialement développé pour *C. acetobutylicum*²⁵⁶. Grâce à l'emploi de plasmides Dam⁻ Dcm⁻ et de la technologie CRISPR-Cas9, la souche DSM 6423 a pu être modifiée pour la première fois de manière dirigée. L'efficacité de la méthode mise au point a également été significativement améliorée en limitant l'activité cytotoxique de Cas9, qui réduit sensiblement les efficacités de transformation, par le biais d'une protéine anti-CRISPR, selon un modèle établi à IFPEN pour *C. acetobutylicum* (manuscrit en annexe). Plusieurs phénotypes intéressants ont ainsi pu être obtenus ($\Delta catB$, sensibilité au thiamphénicol ; $\Delta pNF2$, augmentation des efficacités de transformation), qui pourront être utilisés dans la suite pour construire une souche performante. Des vecteurs portant une nouvelle origine de répllication, compatibles avec plusieurs espèces Gram positives, ont également été développés. Ces derniers présentent une efficacité de transformation élevée dans la souche $\Delta catB \Delta pNF2$, ce qui devrait également contribuer à l'utilisation de ce microorganisme comme plateforme pour de futures modifications.

Les approches développées dans cette thèse (caractérisation phénotypique, données omiques, outil génétique), participent donc à l'établissement de *C. beijerinckii* DSM 6423 comme nouvel organisme de choix pour étudier ou réaliser industriellement

des fermentations AIBE. Le socle de connaissances sur cette souche s'est consolidé et, associé à notre faculté nouvelle à la modifier itérativement, ceci devrait faciliter l'étude et l'amélioration de ce microorganisme.

Ces nouvelles perspectives sont encourageantes mais également nécessaires car, à terme, le procédé envisagé prévoit d'employer des substrats 2G, dont la nature entraîne de nouvelles complications (toxicité du substrat, mélange de sucres à 5 et 6 carbones, nécessité d'extraire les sucres par des enzymes cellulolytiques etc.) et donc de nouveaux défis pour cette souche. La possibilité de modifier le génome à volonté est très intéressante pour pallier ces problèmes, mais les outils d'édition du génome développés dans cette thèse sont toutefois limités par une mise en œuvre relativement lente. Pour chaque modification à réaliser, un certain nombre d'étapes fastidieuses (clonage, déméthylation dans la souche *dam⁻ dcm⁻*, relativement instables, en particulier) doivent être réalisées avec succès.

Il serait donc utile de développer ou d'appliquer, en complément des outils validés au cours de cette thèse, des techniques plus rapides dans un but de *screening*, même si ces dernières sont moins précises. Plusieurs approches pourraient être adoptées. Les techniques de mutagenèse aléatoire seraient par exemple très adaptées pour la production de souches résistantes aux solvants ou aux inhibiteurs de croissance présents dans les substrats 1G / 2G, puisque ces derniers peuvent directement être utilisés comme pression de sélection. Ce type d'approche a déjà été utilisé avec succès pour obtenir, avec un mutagène chimique, des souches dérivées de la souche DSM 6423 résistantes à l'isopropanol⁴⁷, et pourrait être réemployé pour développer une résistance plus importante au butanol, la molécule la plus toxique du mélange IBE (et donc la plus limitante). Couplé à un séquençage des souches obtenues, cette approche pourrait permettre de déterminer des cibles dont l'altération permettrait une meilleure résistance aux solvants, et qui pourraient être reproduites dans une autre souche à l'aide de CRISPR-Cas9. Alternativement, une souche à haute fréquence de mutations pourrait être générée, ce qui permettrait de se passer des mutagènes chimiques. Le gène *mutS*, par exemple, est un gène important pour la réparation de l'ADN par la voie *mismatch repair* (MMR), et sa délétion entraîne une augmentation drastique du nombre de mutations provoquées à chaque génération chez *E. coli*²⁹¹ et *B. subtilis*²⁹².

Outre une application en mutagenèse aléatoire, la voie MMR est intéressante car son inactivation totale ou partielle peut être réalisée pour provoquer des modifications à volonté dans les génomes bactériens à l'aide d'oligonucléotides^{293,294}. Si ce type d'approche dirigée est rapide (pas de clonage à effectuer entre les modifications), il requiert également l'utilisation d'une recombinaise d'ADN simple brin compatible avec le microorganisme d'intérêt. Pour les *Clostridia*, peu de recherches ont encore été effectuées en ce sens. Il semblerait toutefois que la recombinaise RecT de *Clostridium perfringens* puisse réaliser ce type de modification génétique, bien qu'à des fréquences nettement plus basses que le système λ -Red chez *E. coli*²⁶⁵. Pour appliquer ce type d'approches chez *C. beijerinckii*, de premiers efforts pour identifier une recombinaise fonctionnelle seraient donc nécessaires.

Alternativement, d'autres approches probablement plus faciles à mettre en œuvre pourraient être envisagées. Plusieurs stratégies basées sur des systèmes CRISPR permettent par exemple de limiter l'expression d'un gène, le plus connu (et déjà testé chez *C. beijerinckii* NCIMB 8052) étant dCas9^{295,296}. Le système Cas13, qui cible et dégrade spécifiquement un ou plusieurs ARNm, pourrait également être utilisé²⁹⁷. Ces approches permettent d'étudier rapidement la fonction d'un ou de plusieurs gènes, ce qui, considérant la simplicité de ces systèmes (clonage simple d'un ARNg pour chaque cible) pourrait être utilisé en première approche avant de passer à des techniques ciblées d'édition du génome. Le système Cas3, récemment adapté pour *P. aeruginosa*²⁹⁸, permet d'éliminer des portions de génome de l'ordre de la kilobase ou de la dizaine de kilobases, et peut être facilement reprogrammé, et cela sans être tributaire d'un mécanisme de recombinaison homologue. Un tel système pourrait, chez *C. beijerinckii*, être utilisé pour obtenir des mutants rapidement, en n'étant toutefois limité en l'état qu'à la suppression de gènes préexistants.

En conclusion, les travaux décrits dans cette thèse ouvrent la voie pour l'utilisation de souches dérivées de *C. beijerinckii* DSM 6423 pour une utilisation industrielle. A cette fin, de nombreux défis doivent encore être relevés, et leur résolution s'appuiera sans doute sur une modification raisonnée de ce microorganisme, permise à la fois par mon travail de thèse et par les potentielles nouvelles innovations à venir.

8 REFERENCES

1. CEA. Mémento sur l'énergie (2016).
2. IFPEN. Tableau de bord des biocarburants. Available at <https://www.ifpenergiesnouvelles.fr/article/tableau-bord-biocarburants-2019> (2019).
3. Commissariat général au développement durable. Les émissions de CO₂ dues à la combustion d'énergie dans le monde en 2012 (2015).
4. IPCC, Geneva, Switzerland. IPCC, 2014: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)] (2014).
5. United Nations Framework Convention on Climate Change. Kyoto Protocol ; 3rd Conference of the Parties, United Nations (1997).
6. United Nations Framework Convention on Climate Change. Paris Agreement ; 21st Conference of the Parties, United Nations (2015).
7. Hubbert, M. K. in *Nuclear Energy and the Fossil Fuels* (1956), 9-11, 21-22.
8. Laherrère, J. Future of Oil Supplies. Paper presented at the Seminar Center of Energy Conversion Zurich (2003).
9. Eurostat. EU Energy in Figures. Statistical Pocketbook 2015, European Commission (2016).
10. International Energy Agency. World Energy Outlook (2013).
11. Parajuli, R. *et al.* Biorefining in the prevailing energy and materials crisis. A review of sustainable pathways for biorefinery value chains and sustainability assessment methodologies. *Renewable and Sustainable Energy Reviews* **43**, 244–263; 10.1016/j.rser.2014.11.041 (2015).
12. United Nations Environment Programme. Assessing Biofuels (2009).
13. Parlement Européen. Directive 2009/28/CE (2009).
14. IFP Energies Nouvelles. Tableau de bord biocarburants 2017 (2017).
15. HLPE on Food Security and Nutrition. Biofuels and food security. Committee on World Food Security (2013).
16. Alonso, D. M., Bond, J. Q. & Dumesic, J. A. Catalytic conversion of biomass to biofuels. *Green Chem.* **12**, 1493; 10.1039/c004654j (2010).
17. Matar, S. & Hatch, L. F. *Chemistry of petrochemical processes*. 2nd ed. (Gulf Professional Pub, Boston, 2001).
18. IHS Chemical. Ethylene. Chemical Economics Handbook (2017).
19. IHS Chemical. Petrochemical Industry Overview. Chemical Economics Handbook (2017).
20. Anastas, P. T. & Warner, J. C. *Green chemistry. Theory and practice* (Oxford University Press, Oxford England, New York, 1998).
21. Carbios. Website homepage. Available at <https://carbios.fr/>.
22. Lanzatech. Website homepage. Available at <http://www.lanzatech.com/>.
23. Genomatica. Website homepage. Available at <https://www.genomatica.com/>.
24. Global Bioenergies. Website homepage. Available at <http://www.global-bioenergies.com/>.
25. IHS Chemical. Propylene. Chemical Economics Handbook (2017).
26. Bricker, J. *History and State of the Art of Ethane/Propane Dehydrogenation Catalysis. Presented at the Board on Chemical Sciences and Technology Workshop on The Changing Landscape of Hydrocarbon Feedstocks for Chemical Production: Implications for Catalysis*. Washington, DC, March 7–9, 2016 (2016).
27. Vasiliadou, E. S. & Lemonidou, A. A. Production of Biopropylene Using Biomass-Derived Sources; 10.1002/9781119951438.eibc2454 (2016).
28. Cision PRWeb. BioPlastek 2012: Does Bio-Polypropylene Make Economic Sense? Available at <http://www.prweb.com/releases/2012/2/prweb9194258.htm> (2012).

References

29. Larmier K. Transformations de l'isopropanol sur solides aluminiques : une approche mixte expérimentale / modélisation multi-échelle.
30. Ren, H. *et al.* Selective hydrodeoxygenation of biomass-derived oxygenates to unsaturated hydrocarbons using molybdenum carbide catalysts. *ChemSusChem* **6**, 798–801; 10.1002/cssc.201200991 (2013).
31. Chen, J.-S. & Hiu, S. F. Acetone-butanol-isopropanol production by *Clostridium beijerinckii* (synonym, *Clostridium butylicum*). *Biotechnology Letters* **8**, 371–376; 10.1007/BF01040869 (1986).
32. Jones, D. T. & Woods, D. R. Acetone-butanol fermentation revisited. *Microbiological Reviews* **50**, 484–524 (1986).
33. Moon, H. G. *et al.* One hundred years of clostridial butanol fermentation. *FEMS Microbiology Letters* **363**, fnw001-fnw001; 10.1093/femsle/fnw001 (2016).
34. List of Prokaryotic names with Standing in Nomenclature. *Clostridium* Genus. Available at <http://www.bacterio.net/clostridium.html> (2013).
35. Stevens, D. L., Bryant, A. E. & Carroll, K. in *Manual of Clinical Microbiology, Eleventh Edition* (American Society of Microbiology 2015), pp. 940–966.
36. Wells, C. L. & Wilkins, T. D. in *Medical Microbiology*, edited by Baron S (1996).
37. George, H. A., Johnson, J. L., Moore, W. E. C., Holdeman, L. V. & Chen, J. S. Acetone, Isopropanol, and Butanol Production by *Clostridium beijerinckii* (syn. *Clostridium butylicum*) and *Clostridium aurantibutyricum*. *Applied and Environmental Microbiology* **45**, 1160–1163 (1983).
38. Ismaiel, A. A., Zhu, C. X., Colby, G. D. & Chen, J. S. Purification and characterization of a primary-secondary alcohol dehydrogenase from two strains of *Clostridium beijerinckii*. *Journal of Bacteriology* **175**, 5097–5105; 10.1128/jb.175.16.5097-5105.1993 (1993).
39. Hiu, S. F., Zhu, C.-X., Yan, R.-T. & Chen, J.-S. Butanol-ethanol dehydrogenase and butanol-ethanol-isopropanol dehydrogenase: different alcohol dehydrogenases in two strains of *Clostridium beijerinckii* (*Clostridium butylicum*). *Applied and Environmental Microbiology* **53**, 697–703 (1987).
40. Poehlein, A. *et al.* Microbial solvent formation revisited by comparative genome analysis. *Biotechnology for Biofuels* **10**, 58; 10.1186/s13068-017-0742-z (2017).
41. Rogers, P., Chen, J.-S. & Zidwick, M. in *The prokaryotes*. 3rd edition, Vol. 1, edited by Dworkin M (Springer, New York, USA, 2006). 3rd edition, Vol. 1, pp. 672–755.
42. KEIS, S., BENNETT, C. F., WARD, V. K. & JONES, D. T. Taxonomy and Phylogeny of Industrial Solvent-Producing Clostridia. *International Journal of Systematic and Evolutionary Microbiology* **45**, 693–705; 10.1099/00207713-45-4-693 (1995).
43. Keis, S., Shaheen, R. & Jones, D. T. Emended descriptions of *Clostridium acetobutylicum* and *Clostridium beijerinckii*, and descriptions of *Clostridium saccharoperbutylacetonicum* sp. nov. and *Clostridium saccharobutylicum* sp. nov. *International Journal of Systematic and Evolutionary Microbiology* **51**, 2095–2103; 10.1099/00207713-51-6-2095 (2001).
44. Bankar, S. B., Survase, S. A., Ojamo, H. & Granström, T. Biobutanol. The outlook of an academic and industrialist. *RSC Adv.* **3**, 24734; 10.1039/c3ra43011a (2013).
45. Colby, G. D. & Chen, J. S. Purification and properties of 3-hydroxybutyryl-coenzyme A dehydrogenase from *Clostridium beijerinckii* ("*Clostridium butylicum*") NRRL B593. *Applied and Environmental Microbiology* **58**, 3297–3302 (1992).
46. Thompson, D. K. & Chen, J. S. Purification and properties of an acetoacetyl coenzyme A-reacting phosphotransbutyrylase from *Clostridium beijerinckii* ("*Clostridium butylicum*") NRRL B593. *Applied and Environmental Microbiology* **56**, 607–613 (1990).
47. Máté de Gérando, H. *et al.* Improving isopropanol tolerance and production of *Clostridium beijerinckii* DSM 6423 by random mutagenesis and genome shuffling. *Applied Microbiology and Biotechnology* **100**, 5427–5436; 10.1007/s00253-016-7302-5 (2016).
48. Survase, S. A., van Heiningen, A. & Granström, T. Wood pulp as an immobilization matrix for the continuous production of isopropanol and butanol. *Journal of Industrial Microbiology & Biotechnology* **40**, 209–215; 10.1007/s10295-012-1219-5 (2013).
49. Survase, S. A., Jurgens, G., van Heiningen, A. & Granström, T. Continuous production of isopropanol and butanol using *Clostridium beijerinckii* DSM 6423. *Applied Microbiology and Biotechnology* **91**, 1305–1313; 10.1007/s00253-011-3322-3 (2011).

50. Gungormusler, M., Gonen, C. & Azbar, N. Continuous production of 1,3-propanediol using raw glycerol with immobilized *Clostridium beijerinckii* NRRL B-593 in comparison to suspended culture. *Bioprocess and Biosystems Engineering* **34**, 727–733; 10.1007/s00449-011-0522-2 (2011).
51. Vrije, T. de, Budde, M., van der Wal, H., Am Claassen, P. & López-Contreras, A. M. “In situ” removal of isopropanol, butanol and ethanol from fermentation broth by gas stripping. *Bioresource technology* **137**, 153–159 (2013).
52. Máté de Gérando, H. *et al.* Genome and transcriptome of the natural isopropanol producer *Clostridium beijerinckii* DSM6423. *BMC genomics* **19**, 242 (2018).
53. Pasteur, L. Animalcules infusoires vivant sans gaz oxygène libre et déterminant des fermentations. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences* **52:344-347** (1861).
54. Pasteur, L. Quelques résultats nouveaux relatifs aux fermentations acétiques et butyriques. *Bulletin de la Société de Chimie de Paris* **52-53** (1862).
55. Fitz, A. Über die Gährung des Glycerins. *European Journal of Inorganic Chemistry* **9**, 1348–1352 (1876).
56. Beijerinck, M. W. Über die butylalkoholgärung und das butyl ferment. *Verhandel Akad Wetenschappen Amsterdam*, 3–51 (1893).
57. Dürre, P. New insights and novel developments in clostridial acetone/butanol/isopropanol fermentation. *Applied Microbiology and Biotechnology* **49**, 639–648; 10.1007/s002530051226 (1998).
58. Weizmann, C. Improvements in the bacterial fermentation of carbohydrates and in bacterial cultures for the same. *British patent 4845* (1915).
59. McCoy, E., Fred, E. B., Peterson, W. H. & Hastings, E. G. A cultural study of the acetone butyl alcohol organism. *The Journal of Infectious Diseases*, 457–483 (1926).
60. Gabriel, C. L. Butanol fermentation process. *Industrial & engineering chemistry* **20**, 1063–1067 (1928).
61. Gabriel, C. L. & Crawford, F. M. Development of the butyl-acetonic fermentation industry. *Industrial & engineering chemistry* **22**, 1163–1165 (1930).
62. Wang, Y., Janssen, H. & Blaschek, H. P. Fermentative biobutanol production: an old topic with remarkable recent advances. *Bioprocessing of Renewable Resources to Commodity Bioproducts*, 227–260 (2014).
63. Sauer, M. Industrial production of acetone and butanol by fermentation—100 years later. *FEMS Microbiology Letters* **363** (2016).
64. Rose, A. H. Industrial microbiology. *Industrial microbiology* (1961).
65. Lutke-Eversloh, T. & Bahl, H. Metabolic engineering of *Clostridium acetobutylicum*: recent advances to improve butanol production. *Curr Opin Biotechnol* **22**; 10.1016/j.copbio.2011.01.011 (2011).
66. Dürre, P. Fermentative butanol production. Bulk chemical and biofuel. *Annals of the New York Academy of Sciences* **1125**, 353–362; 10.1196/annals.1419.009 (2008).
67. Ren, C. *et al.* Identification and inactivation of pleiotropic regulator CcpA to eliminate glucose repression of xylose utilization in *Clostridium acetobutylicum*. *Metabolic engineering* **12**, 446–454 (2010).
68. Qureshi, N. & Blaschek, H. P. Recent advances in ABE fermentation: hyper-butanol producing *Clostridium beijerinckii* BA101. *Journal of Industrial Microbiology and Biotechnology* **27**, 287–291 (2001).
69. Green, E. M. *et al.* Genetic manipulation of acid formation pathways by gene inactivation in *Clostridium acetobutylicum* ATCC 824. *Microbiology* **142**, 2079–2086 (1996).
70. Jiang, Y. *et al.* Disruption of the acetoacetate decarboxylase gene in solvent-producing *Clostridium acetobutylicum* increases the butanol ratio. *Metabolic engineering* **11**, 284–291 (2009).
71. Lehmann, D. *et al.* Modifying the product pattern of *Clostridium acetobutylicum*. *Applied Microbiology and Biotechnology* **94**, 743–754 (2012).
72. Li, S.-Y., Srivastava, R., Suib, S. L., Li, Y. & Parnas, R. S. Performance of batch, fed-batch, and continuous A–B–E fermentation with pH-control. *Bioresource technology* **102**, 4241–4250 (2011).
73. Zheng, Y.-N. *et al.* Problems with the microbial production of butanol. *Journal of Industrial Microbiology & Biotechnology* **36**, 1127–1138 (2009).

References

74. Qureshi, N., Saha, B. C., Hector, R. E., Hughes, S. R. & Cotta, M. A. Butanol production from wheat straw by simultaneous saccharification and fermentation using *Clostridium beijerinckii*: Part I—Batch fermentation. *Biomass and Bioenergy* **32**, 168–175 (2008).
75. Lee, J. *et al.* Metabolic Engineering of *Clostridium acetobutylicum* ATCC 824 for Isopropanol-Butanol-Ethanol Fermentation. *Applied and Environmental Microbiology* **78**, 1416–1423; 10.1128/AEM.06382-11 (2012).
76. Dusseaux, S., Croux, C., Soucaille, P. & Meynial-Salles, I. Metabolic engineering of *Clostridium acetobutylicum* ATCC 824 for the high-yield production of a biofuel composed of an isopropanol/butanol/ethanol mixture. *Metab Eng* **18**; 10.1016/j.ymben.2013.03.003 (2013).
77. Collas, F. *et al.* Simultaneous production of isopropanol, butanol, ethanol and 2, 3-butanediol by *Clostridium acetobutylicum* ATCC 824 engineered strains. *Amb Express* **2**, 45 (2012).
78. Dai, Z. *et al.* Introducing a single secondary alcohol dehydrogenase into butanol-tolerant *Clostridium acetobutylicum* Rh8 switches ABE fermentation to high level IBE fermentation. *Biotechnology for Biofuels* **5**, 44 (2012).
79. Yerushalmi, L., Volesky, B., Leung, W. K. & Neufeld, R. J. Variations of solvent yield in acetone-butanol fermentation. *Applied Microbiology and Biotechnology* **18**, 279–286; 10.1007/BF00500492 (1983).
80. Vasconcelos, I., Girbal, L. & Soucaille, P. Regulation of carbon and electron flow in *Clostridium acetobutylicum* grown in chemostat culture at neutral pH on mixtures of glucose and glycerol. *J Bacteriol* **176** (1994).
81. Gapes, J. R., Nimcevic, D. & Friedl, A. Long-Term Continuous Cultivation of *Clostridium beijerinckii* in a Two-Stage Chemostat with On-Line Solvent Removal. *Applied and Environmental Microbiology* **62**, 3210–3219 (1996).
82. Costa, J. M. & Moreira, A. R. (ACS Publications 1983).
83. Patakova, P. *et al.* Comparative analysis of high butanol tolerance and production in *Clostridia*. *Biotechnology advances* (2017).
84. Kujawska, A., Kujawski, J., Bryjak, M. & Kujawski, W. ABE fermentation products recovery methods—A review. *Renewable and Sustainable Energy Reviews* **48**, 648–661; 10.1016/j.rser.2015.04.028 (2015).
85. Jiang, W., Zhao, J., Wang, Z. & Yang, S.-T. Stable high-titer n-butanol production from sucrose and sugarcane juice by *Clostridium acetobutylicum* JB200 in repeated batch fermentations. *Bioresource technology* **163**, 172–179 (2014).
86. Ni, Y., Wang, Y. & Sun, Z. Butanol production from cane molasses by *Clostridium saccharobutylicum* DSM 13864: batch and semicontinuous fermentation. *Applied Biochemistry and Biotechnology* **166**, 1896–1907 (2012).
87. Ezeji, T. C., Qureshi, N. & Blaschek, H. P. Production of acetone butanol (AB) from liquefied corn starch, a commercial substrate, using *Clostridium beijerinckii* coupled with product recovery by gas stripping. *Journal of Industrial Microbiology & Biotechnology* **34**, 771–777 (2007).
88. Thang, V. H., Kanda, K. & Kobayashi, G. Production of acetone–butanol–ethanol (ABE) in direct fermentation of cassava by *Clostridium saccharoperbutylacetonicum* N1-4. *Applied Biochemistry and Biotechnology* **161**, 157–170 (2010).
89. Parekh, M., Formanek, J. & Blaschek, H. P. Development of a cost-effective glucose-corn steep medium for production of butanol by *Clostridium beijerinckii*. *Journal of Industrial Microbiology and Biotechnology* **21**, 187–191 (1998).
90. Lu, C., Zhao, J., Yang, S.-T. & Wei, D. Fed-batch fermentation for n-butanol production from cassava bagasse hydrolysate in a fibrous bed bioreactor with continuous gas stripping. *Bioresource technology* **104**, 380–387 (2012).
91. Marchal, R., Blanchet, D. & Vandecasteele, J. P. Industrial optimization of acetone-butanol fermentation: a study of the utilization of Jerusalem artichokes. *Applied Microbiology and Biotechnology* **23**, 92–98 (1985).
92. Ezeji, T., Qureshi, N. & Blaschek, H. P. Butanol production from agricultural residues. Impact of degradation products on *Clostridium beijerinckii* growth and butanol fermentation. *Biotechnology and bioengineering* **97**, 1460–1469; 10.1002/bit.21373 (2007).
93. Survase, S. A., Sklavounos, E., van Heiningen, A. & Granström, T. Market refused vegetables as a supplement for improved acetone–butanol–ethanol production by *Clostridium acetobutylicum* DSM 792. *Industrial crops and products* **45**, 349–354 (2013).
94. Raganati, F. *et al.* Butanol production by bioconversion of cheese whey in a continuous packed bed reactor. *Bioresource technology* **138**, 259–265 (2013).
95. Foda, M. I., Dong, H. & Li, Y. Study the suitability of cheese whey for bio-butanol production by *Clostridia*. *J Am Sci* **6**, 39–46 (2010).

96. Xin, F. *et al.* Enhanced biobutanol production with high yield from crude glycerol by acetone uncoupled *Clostridium* sp. strain CT7. *Bioresource technology* **244**, 575–581; 10.1016/j.biortech.2017.08.002 (2017).
97. Bellido, C., Loureiro Pinto, M., Coca, M., González-Benito, G. & García-Cubero, M. T. Acetone-butanol-ethanol (ABE) production by *Clostridium beijerinckii* from wheat straw hydrolysates. Efficient use of penta and hexa carbohydrates. *Bioresource technology* **167**, 198–205; 10.1016/j.biortech.2014.06.020 (2014).
98. Ezeji, T. & Blaschek, H. P. Fermentation of dried distillers' grains and solubles (DDGS) hydrolysates to solvents and value-added products by solventogenic *clostridia*. *Bioresource technology* **99**, 5232–5242 (2008).
99. Qureshi, N., Li, X.-L., Hughes, S., Saha, B. C. & Cotta, M. A. Butanol production from corn fiber xylan using *Clostridium acetobutylicum*. *Biotechnology Progress* **22**, 673–680 (2006).
100. Qureshi, N. *et al.* Butanol production by *Clostridium beijerinckii*. Part I: use of acid and enzyme hydrolyzed corn fiber. *Bioresource technology* **99**, 5915–5922 (2008).
101. Qureshi, N. *et al.* Production of butanol (a biofuel) from agricultural residues: Part II—Use of corn stover and switchgrass hydrolysates. *Biomass and Bioenergy* **34**, 566–571 (2010).
102. Chen, W.-H., Chen, Y.-C. & Lin, J.-G. Evaluation of biobutanol production from non-pretreated rice straw hydrolysate under non-sterile environmental conditions. *Bioresource technology* **135**, 262–268 (2013).
103. Lu, C., Dong, J. & Yang, S.-T. Butanol production from wood pulping hydrolysate in an integrated fermentation–gas stripping process. *Bioresource technology* **143**, 467–475 (2013).
104. Qureshi, N., Saha, B. C., Dien, B., Hector, R. E. & Cotta, M. A. Production of butanol (a biofuel) from agricultural residues: Part I—Use of barley straw hydrolysate. *Biomass and Bioenergy* **34**, 559–565 (2010).
105. Qureshi, N. *et al.* Process integration for simultaneous saccharification, fermentation, and recovery (SSFR). Production of butanol from corn stover using *Clostridium beijerinckii* P260. *Bioresource technology* **154**, 222–228; 10.1016/j.biortech.2013.11.080 (2014).
106. Ellis, R. P. *et al.* Starch production and industrial use. *Journal of the Science of Food and Agriculture* **77**, 289–311; 10.1002/(SICI)1097-0010(199807)77:3<289::AID-JSFA38>3.0.CO;2-D (1998).
107. Zobel, H. F. Molecules to granules: a comprehensive starch review. *Starch-Stärke* **40**, 44–50 (1988).
108. Swinkels, J. J.M. Composition and properties of commercial native starches. *Starch-Stärke* **37**, 1–5 (1985).
109. Clark, M. A., Choi, J. H. & Douglas, M. *Biology 2e* (OpenStax College, Rice University, Houston, Texas, 2018-).
110. Mitchell, W. J. in *Advances in microbial physiology* 39, edited by R.K. Poole (Academic, San Diego, 1998), pp. 31–130.
111. Servinsky, M. D., Kiel, J. T., Dupuy, N. F. & Sund, C. J. Transcriptional analysis of differential carbohydrate utilization by *Clostridium acetobutylicum*. *Microbiology* **156**; 10.1099/mic.0.037085-0 (2010).
112. Ölbrich, H. The Molasses Book. *Fermentation Technologist, Institute für Zuckerindustrie, Berlin (Germany)* (1963).
113. Qureshi, N., Lolas, A. & Blaschek, H. P. Soy molasses as fermentation substrate for production of butanol using *Clostridium beijerinckii* BA101. *Journal of Industrial Microbiology & Biotechnology* **26**, 290–295; 10.1038/sj.jim.7000131 (2001).
114. Görke, B. & Stülke, J. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nature Reviews Microbiology* **6**, 613 EP - (2008).
115. Mitchell, W. J. The Phosphotransferase System in Solventogenic *Clostridia*. *Journal of molecular microbiology and biotechnology* **25**, 129–142; 10.1159/000375125 (2015).
116. Balat, M. Production of bioethanol from lignocellulosic materials via the biochemical pathway: a review. *Energy conversion and management* **52**, 858–875 (2011).
117. Taherzadeh, J. M. & Karimi, K. *Pretreatment of Lignocellulosic Wastes to Improve Ethanol and Biogas Production: A Review* (2008).
118. Taherzadeh, M. J. & Karimi, K. Enzymatic-based hydrolysis processes for ethanol from lignocellulosic materials: A review. *BioResources* **2**, 707–738 (2007).
119. Ounine, K., Petitdemange, H., Raval, G. & Gay, R. Regulation and butanol inhibition of D-xylose and D-glucose uptake in *Clostridium acetobutylicum*. *Applied and Environmental Microbiology* **49**, 874–878 (1985).

References

120. Patakova, P., Linhova, M., Rychtera, M., Paulova, L. & Melzoch, K. Novel and neglected issues of acetone-butanol-ethanol (ABE) fermentation by *clostridia*. *Clostridium* metabolic diversity, tools for process mapping and continuous fermentation systems. *Biotechnology advances* **31**, 58–67; 10.1016/j.biotechadv.2012.01.010 (2013).
121. Johnson, J. L., Toth, J., Santiwatanakul, S. & Chen, J.-S. Cultures of “*Clostridium acetobutylicum*” from various collections comprise *Clostridium acetobutylicum*, *Clostridium beijerinckii*, and two other distinct types based on DNA-DNA reassociation. *International Journal of Systematic and Evolutionary Microbiology* **47**, 420–424 (1997).
122. Dictionnaire Larousse. Métabolisme. Définition. Available at <http://www.larousse.fr/dictionnaires/francais/m%C3%A9tabolisme/50809> (2017).
123. Huang, H. *et al.* CRISPR/Cas9-Based Efficient Genome Editing in *Clostridium ljungdahlii*, an Autotrophic Gas-Fermenting Bacterium. *ACS synthetic biology* **5**, 1355–1361; 10.1021/acssynbio.6b00044 (2016).
124. Gottwald, M. & Gottschalk, G. The internal pH of *Clostridium acetobutylicum* and its effect on the shift from acid to solvent formation. *Archives of microbiology* **143**, 42–46 (1985).
125. Kell, D. B., Peck, M. W., Rodger, G. & Morris, J. G. On the permeability to weak acids and bases of the cytoplasmic membrane of *Clostridium pasteurianum*. *Biochemical and biophysical research communications* **99**, 81–88 (1981).
126. Huang, L., Forsberg, C. W. & Gibbins, L. N. Influence of external pH and fermentation products on *Clostridium acetobutylicum* intracellular pH and cellular distribution of fermentation products. *Applied and Environmental Microbiology* **51**, 1230–1234 (1986).
127. Jones, D. T. *et al.* Solvent production and morphological changes in *Clostridium acetobutylicum*. *Applied and Environmental Microbiology* **43**, 1434–1439 (1982).
128. Al-Hinai, M. A., Jones, S. W. & Papoutsakis, E. T. The *Clostridium* sporulation programs: diversity and preservation of endospore differentiation. *Microbiol Mol Biol Rev* **79**; 10.1128/MMBR.00025-14 (2015).
129. Setlow, P. in *Bacterial Stress Responses, Second Edition* (American Society of Microbiology 2011), pp. 319–332.
130. Nölling, J. *et al.* Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium *Clostridium acetobutylicum*. *Journal of Bacteriology* **183**, 4823–4838; 10.1128/JB.183.16.4823-4838.2001 (2001).
131. Mitchell, W. J. Sugar uptake by the solventogenic *clostridia*. *World Journal of Microbiology and Biotechnology* **32**, 32; 10.1007/s11274-015-1981-4 (2016).
132. Mitchell, W. J. Carbohydrate Uptake and Utilization by *Clostridium beijerinckii* NCIMB 8052. *Anaerobe* **2**, 379–384 (1996).
133. Lee, J., Mitchell, W. J., Tangney, M. & Blaschek, H. P. Evidence for the presence of an alternative glucose transport system in *Clostridium beijerinckii* NCIMB 8052 and the solvent-hyperproducing mutant BA101. *Applied and Environmental Microbiology* **71**, 3384–3387 (2005).
134. Tangney, M., Rouse, C., Yazdanian, M. & Mitchell, W. J. Note: sucrose transport and metabolism in *Clostridium beijerinckii* NCIMB 8052. *Journal of applied microbiology* **84**, 914 (1998).
135. Reid, S. J., Rafudeen, M. S. & Leat, N. G. The genes controlling sucrose utilization in *Clostridium beijerinckii* NCIMB 8052 constitute an operon. *Microbiology* **145**, 1461–1472 (1999).
136. Sandoval-Espinola, W. J., Chinn, M. S., Thon, M. R. & Bruno-Bárcena, J. M. Evidence of mixotrophic carbon-capture by n-butanol-producer *Clostridium beijerinckii*. *Scientific Reports* **7**, 12759; 10.1038/s41598-017-12962-8 (2017).
137. Volesky, B. & Szczesny, T. Bacterial conversion of pentose sugars to acetone and butanol. *Pentoses and Lignin*, 101–118 (1983).
138. Gu, Y. *et al.* Reconstruction of xylose utilization pathway and regulons in Firmicutes. *BMC genomics* **11**, 255 (2010).
139. Liu, L. *et al.* Phosphoketolase pathway for xylose catabolism in *Clostridium acetobutylicum* revealed by ¹³C metabolic flux analysis. *Journal of Bacteriology* **194**, 5413–5422 (2012).
140. Diallo, M. *et al.* l-Rhamnose Metabolism in *Clostridium beijerinckii* Strain DSM 6423. *Applied and Environmental Microbiology* **85**; 10.1128/AEM.02656-18 (2019).
141. Paquet, V., Croux, C., Goma, G. & Soucaille, P. Purification and characterization of the extracellular alpha-amylase from *Clostridium acetobutylicum* ATCC 824. *Applied and Environmental Microbiology* **57**, 212–218 (1991).
142. López-Contreras, A. M. *et al.* Substrate-induced production and secretion of cellulases by *Clostridium acetobutylicum*. *Applied and Environmental Microbiology* **70**, 5238–5243 (2004).

143. Tamaru, Y. The *Clostridium cellulovorans* cellulosome: an enzyme complex with plant cell wall degrading activity. *The Chemical Record* **1**, 24–32 (2001).
144. Papoutsakis, E. T. Engineering solventogenic *clostridia*. *Current opinion in biotechnology* **19**, 420–429 (2008).
145. Wiesenborn, D. P., Rudolph, F. B. & Papoutsakis, E. T. Thiolase from *Clostridium acetobutylicum* ATCC 824 and its role in the synthesis of acids and solvents. *Applied and Environmental Microbiology* **54**, 2717–2722 (1988).
146. Hartmanis, M. G. N. & Gatenbeck, S. Intermediary metabolism in *Clostridium acetobutylicum*: levels of enzymes involved in the formation of acetate and butyrate. *Applied and Environmental Microbiology* **47**, 1277–1283 (1984).
147. Yan, R.-T., Zhu, C.-X., Golemboski, C. & Chen, J.-S. Expression of solvent-forming enzymes and onset of solvent production in batch cultures of *Clostridium beijerinckii* (“*Clostridium butylicum*”). *Applied and Environmental Microbiology* **54**, 642–648 (1988).
148. Boynton, Z. L., Bennet, G. N. & Rudolph, F. B. Cloning, sequencing, and expression of clustered genes encoding beta-hydroxybutyryl-coenzyme A (CoA) dehydrogenase, crotonase, and butyryl-CoA dehydrogenase from *Clostridium acetobutylicum* ATCC 824. *Journal of Bacteriology* **178**, 3015–3024 (1996).
149. Wietzke, M. & Bahl, H. The redox-sensing protein Rex, a transcriptional regulator of solventogenesis in *Clostridium acetobutylicum*. *Appl Microbiol Biotechnol* **96**; 10.1007/s00253-012-4112-2 (2012).
150. Boynton, Z. L., Bennett, G. N. & Rudolph, F. B. Intracellular concentrations of coenzyme A and its derivatives from *Clostridium acetobutylicum* ATCC 824 and their roles in enzyme regulation. *Applied and Environmental Microbiology* **60**, 39–44 (1994).
151. Cooksley, C. M. *et al.* Targeted mutagenesis of the *Clostridium acetobutylicum* acetone-butanol-ethanol fermentation pathway. *Metab Eng* **14**; 10.1016/j.ymben.2012.09.001 (2012).
152. Soucaille, P., Percheron, B., Croux, C. & Meynial-Salles, I. *Clostridium acetobutylicum* strains unable to produce hydrogen and useful for the continuous production of chemicals and fuels (Google Patents, 2015).
153. Sommer, B., Garbe, D., Schrepfer, P. & Brück, T. Characterization of a highly thermostable β -hydroxybutyryl CoA dehydrogenase from *Clostridium acetobutylicum* ATCC 824. *Journal of Molecular Catalysis B: Enzymatic* **98**, 138–144 (2013).
154. Lehmann, D. & Lutke-Eversloh, T. Switching *Clostridium acetobutylicum* to an ethanol producer by disruption of the butyrate/butanol fermentative pathway. *Metab Eng* **13**; 10.1016/j.ymben.2011.04.006 (2011).
155. 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*. *Journal of Bacteriology* **190**, 843–850; 10.1128/JB.01417-07 (2008).
156. Boynton, Z. L., Bennett, G. N. & Rudolph, F. B. Cloning, sequencing, and expression of genes encoding phosphotransacetylase and acetate kinase from *Clostridium acetobutylicum* ATCC 824. *Appl Environ Microbiol* **62** (1996).
157. Winzer, K., Lorenz, K. & Dürre, P. Acetate kinase from *Clostridium acetobutylicum*: a highly specific enzyme that is actively transcribed during acidogenesis and solventogenesis. *Microbiology* **143**, 3279–3286 (1997).
158. Kuit, W., Minton, N. P., López-Contreras, A. M. & Eggink, G. Disruption of the acetate kinase (ack) gene of *Clostridium acetobutylicum* results in delayed acetate production. *Applied Microbiology and Biotechnology* **94**, 729–741 (2012).
159. Walter, K. A., Nair, R. V., Cary, J. W., Bennett, G. N. & Papoutsakis, E. T. Sequence and arrangement of two genes of the butyrate-synthesis pathway of *Clostridium acetobutylicum* ATCC 824. *Gene* **134**, 107–111 (1993).
160. Lehmann, D., Radomski, N. & Lütke-Eversloh, T. New insights into the butyric acid metabolism of *Clostridium acetobutylicum*. *Applied Microbiology and Biotechnology* **96**, 1325–1339 (2012).
161. Yoo, M., Croux, C., Meynial-Salles, I. & Soucaille, P. Metabolic flexibility of a butyrate pathway mutant of *Clostridium acetobutylicum*. *Metabolic engineering* **40**, 138–147 (2017).
162. Wiesenborn, D. P., Rudolph, F. B. & Papoutsakis, E. T. Phosphotransbutyrylase from *Clostridium acetobutylicum* ATCC 824 and its role in acidogenesis. *Applied and Environmental Microbiology* **55**, 317–322 (1989).
163. Desai, R. P., Harris, L. M., Welker, N. E. & Papoutsakis, E. T. Metabolic flux analysis elucidates the importance of the acid-formation pathways in regulating solvent production by *Clostridium acetobutylicum*. *Metabolic engineering* **1**, 206–213; 10.1006/mben.1999.0118 (1999).
164. Harris, L. M., Desai, R. P., Welker, N. E. & Papoutsakis, E. T. Characterization of recombinant strains of the *Clostridium acetobutylicum* butyrate kinase inactivation mutant: need for new phenomenological models for solventogenesis and butanol inhibition? *Biotechnology and bioengineering* **67**, 1–11 (2000).

References

165. Harris, L. M., Blank, L., Desai, R. P., Welker, N. E. & Papoutsakis, E. T. Fermentation characterization and flux analysis of recombinant strains of *Clostridium acetobutylicum* with an inactivated solR gene. *Journal of Industrial Microbiology and Biotechnology* **27**, 322–328; 10.1038/sj.jim.7000191 (2001).
166. Hartmanis, M. G. N., Klason, T. & Gatenbeck, S. Uptake and activation of acetate and butyrate in *Clostridium acetobutylicum*. *Applied Microbiology and Biotechnology* **20**, 66–71 (1984).
167. Wiesenborn, D. P., Rudolph, F. B. & Papoutsakis, E. T. Coenzyme A transferase from *Clostridium acetobutylicum* ATCC 824 and its role in the uptake of acids. *Applied and Environmental Microbiology* **55**, 323–329 (1989).
168. Cary, J. W., Petersen, D. J., Papoutsakis, E. T. & Bennett, G. N. Cloning and expression of *Clostridium acetobutylicum* ATCC 824 acetoacetyl-coenzyme A: acetate/butyrate: coenzyme A-transferase in *Escherichia coli*. *Applied and Environmental Microbiology* **56**, 1576–1583 (1990).
169. Cornillot, E., Nair, R. V., Papoutsakis, E. T. & Soucaille, P. The genes for butanol and acetone formation in *Clostridium acetobutylicum* ATCC 824 reside on a large plasmid whose loss leads to degeneration of the strain. *J Bacteriol* **179**; 10.1128/jb.179.17.5442-5447.1997 (1997).
170. Tummala, S. B., Junne, S. G. & Papoutsakis, E. T. Antisense RNA downregulation of coenzyme A transferase combined with alcohol-aldehyde dehydrogenase overexpression leads to predominantly alcohologenic *Clostridium acetobutylicum* fermentations. *Journal of Bacteriology* **185**, 3644–3653 (2003).
171. Tummala, S. B., Welker, N. E. & Papoutsakis, E. T. Design of antisense RNA constructs for downregulation of the acetone formation pathway of *Clostridium acetobutylicum*. *Journal of Bacteriology* **185**, 1923–1934 (2003).
172. Mermelstein, L. D., Papoutsakis, E. T., Petersen, D. J. & Bennett, G. N. Metabolic engineering of *Clostridium acetobutylicum* ATCC 824 for increased solvent production by enhancement of acetone formation enzyme activities using a synthetic acetone operon. *Biotechnology and bioengineering* **42**, 1053–1060 (1993).
173. Yoo, M., Croux, C., Meynial-Salles, I. & Soucaille, P. Elucidation of the roles of adhE1 and adhE2 in the primary metabolism of *Clostridium acetobutylicum* by combining in-frame gene deletion and a quantitative system-scale approach. *Biotechnology for Biofuels* **9**, 92; 10.1186/s13068-016-0507-0 (2016).
174. Yoo M, Bestel-Corre G, Croux C, Riviere A, Meynial-Salles I, Soucaille P. A Quantitative System-Scale Characterization of the Metabolism of *Clostridium acetobutylicum*. *mBio* (2015).
175. Wang, Y., Li, X., Mao, Y. & Blaschek, H. P. Single-nucleotide resolution analysis of the transcriptome structure of *Clostridium beijerinckii* NCIMB 8052 using RNA-Seq. *BMC genomics* **12**, 479 (2011).
176. Wang, Y., Li, X., Mao, Y. & Blaschek, H. P. Genome-wide dynamic transcriptional profiling in *Clostridium beijerinckii* NCIMB 8052 using single-nucleotide resolution RNA-Seq. *BMC genomics* **13**, 102 (2012).
177. Lee, S. Y., Jang, Y. S., Lee, J. M., Im, J. A. & Song, H. *Recombinant mutant microorganisms having increased ability to produce alcohols and method of producing alcohols using the same* (Google Patents, 2012), <https://www.google.com/patents/US20120301936>.
178. Zhao, Y., Tomas, C. A., Rudolph, F. B., Papoutsakis, E. T. & Bennett, G. N. Intracellular butyryl phosphate and acetyl phosphate concentrations in *Clostridium acetobutylicum* and their implications for solvent formation. *Applied and Environmental Microbiology* **71**, 530–537 (2005).
179. Girbal, L. & Soucaille, P. Regulation of solvent production in *Clostridium acetobutylicum*. *Trends Biotechnol* **16**; 10.1016/S0167-7799(97)01141-4 (1998).
180. Bertini, I. *et al.* Solution structure of the oxidized 24Fe-4S ferredoxin from *Clostridium pasteurianum*. *European journal of biochemistry* **232**, 192–205 (1995).
181. Jungermann, K., Thauer, R. K., Leimenstoll, G. & Decker, K. Function of reduced pyridine nucleotide-ferredoxin oxidoreductases in saccharolytic *Clostridia*. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **305**, 268–280 (1973).
182. Charon, M.-H., Volbeda, A., Chabriere, E., Pieulle, L. & Fontecilla-Camps, J. C. Structure and electron transfer mechanism of pyruvate. Ferredoxin oxidoreductase. *Current Opinion in Structural Biology* **9**, 663–669; 10.1016/S0959-440X(99)00027-5 (1999).
183. Rogers, P. in *Advances in applied microbiology*, edited by A. I. Laskin (Academic Press, San Diego, 1986), pp. 1–60.
184. Soucaille, P., Meynial-Salles, I., Foulquier, C. & Riviere, A. *New polypeptide having ferredoxin-nadp+ reductase activity, polynucleotide encoding the same and uses thereof* (2018).

185. 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. *Cellular and Molecular Life Sciences* **68**, 613–634; 10.1007/s00018-010-0555-8 (2011).
186. Tremblay, P.-L., Zhang, T., Dar, S. A., Leang, C. & Lovley, D. R. The Rnf complex of *Clostridium ljungdahlii* is a proton-translocating ferredoxin:NAD⁺ oxidoreductase essential for autotrophic growth. *mBio* **4**, e00406-12; 10.1128/mBio.00406-12 (2012).
187. Patakova, P. *et al.* Acidogenesis, solventogenesis, metabolic stress response and life cycle changes in *Clostridium beijerinckii* NRRL B-598 at the transcriptomic level. *Scientific Reports* **9**, 1371; 10.1038/s41598-018-37679-0 (2019).
188. Am Junelles, Janati-Idrissi, R., Petitdemange, H. & Gay, R. Iron effect on acetone-butanol fermentation. *Current Microbiology* **17**, 299–303 (1988).
189. Kim, B. H., Bellows, P., Datta, R. & Zeikus, J. G. Control of carbon and electron flow in *Clostridium acetobutylicum* fermentations: utilization of carbon monoxide to inhibit hydrogen production and to enhance butanol yields. *Applied and Environmental Microbiology* **48**, 764–770 (1984).
190. Datta, R. & Zeikus, J. G. Modulation of acetone-butanol-ethanol fermentation by carbon monoxide and organic acids. *Applied and Environmental Microbiology* **49**, 522–529 (1985).
191. Peguin, S. & Soucaille, P. Modulation of carbon and electron flow in *Clostridium acetobutylicum* by iron limitation and methyl viologen addition. *Applied and Environmental Microbiology* **61**, 403–405 (1995).
192. Paredes, C. J., Alsaker, K. V. & Papoutsakis, E. T. A comparative genomic view of clostridial sporulation and physiology. *Nature Reviews Microbiology* **3**, 969–978 (2005).
193. Harris, L. M., Welker, N. E. & Papoutsakis, E. T. Northern, morphological, and fermentation analysis of spo0A inactivation and overexpression in *Clostridium acetobutylicum* ATCC 824. *J Bacteriol* **184**; 10.1128/JB.184.13.3586-3597.2002 (2002).
194. Tracy, B. P., Jones, S. W. & Papoutsakis, E. T. Inactivation of σ^E and σ^G in *Clostridium acetobutylicum* illuminates their roles in clostridial-cell-form biogenesis, granulose synthesis, solventogenesis, and spore morphogenesis. *Journal of Bacteriology* **193**, 1414–1426 (2011).
195. U.S. Department of Energy Joint Genome Institute. *Clostridium beijerinckii* NCIMB 8052. Project ID 3634512. Available at <http://genome.jgi.doe.gov/clobe/clobe.home.html> (2007).
196. Shi, Z. & Blaschek, H. P. Transcriptional analysis of *Clostridium beijerinckii* NCIMB 8052 and the hyper-butanol-producing mutant BA101 during the shift from acidogenesis to solventogenesis. *Applied and Environmental Microbiology* **74**, 7709–7714 (2008).
197. Zhang, W., Li, F. & Nie, L. Integrating multiple 'omics' analysis for microbial biology: application and methodologies. *Microbiology* **156**, 287–301 (2010).
198. Mao, S. *et al.* Proteome reference map and comparative proteomic analysis between a wild type *Clostridium acetobutylicum* DSM 1731 and its mutant with enhanced butanol tolerance and butanol yield. *Journal of proteome research* **9**, 3046–3061 (2010).
199. Mao, S. *et al.* Comparative analysis on the membrane proteome of *Clostridium acetobutylicum* wild type strain and its butanol-tolerant mutant. *Molecular BioSystems* **7**, 1660–1677 (2011).
200. Borden, J. R. & Papoutsakis, E. T. Dynamics of genomic-library enrichment and identification of solvent tolerance genes for *Clostridium acetobutylicum*. *Applied and Environmental Microbiology* **73**, 3061–3068 (2007).
201. Blazeck, J. & Alper, H. Systems metabolic engineering: Genome-scale models and beyond. *Biotechnology journal* **5**, 647–659 (2010).
202. Xu, C. *et al.* Genome-scale metabolic model in guiding metabolic engineering of microbial improvement. *Applied Microbiology and Biotechnology* **97**, 519–539 (2013).
203. Browning, D. F. & Busby, S. J. W. Local and global regulation of transcription initiation in bacteria. *Nature Reviews Microbiology* **14**, 638 EP - (2016).
204. Jones, S. W., Tracy, B. P., Gaida, S. M. & Papoutsakis, E. T. Inactivation of σ^F in *Clostridium acetobutylicum* ATCC 824 blocks sporulation prior to asymmetric division, abolishes σ^E - and σ^G -protein expression, but does not block solvent formation. *Journal of Bacteriology*; 10.1128/JB.00088-11 (2011).

References

205. Jones, S. W. *et al.* The transcriptional program underlying the physiology of clostridial sporulation. *Genome biology* **9**, R114; 10.1186/gb-2008-9-7-r114 (2008).
206. Al-Hinai, M. A., Jones, S. W. & Papoutsakis, E. T. σ^k of *Clostridium acetobutylicum* Is the First Known Sporulation-Specific Sigma Factor with Two Developmentally Separated Roles, One Early and One Late in Sporulation. *Journal of Bacteriology* **196**, 287–299; 10.1128/JB.01103-13 (2014).
207. Nie, X. *et al.* PTS regulation domain-containing transcriptional activator CelR and sigma factor $\sigma(54)$ control cellobiose utilization in *Clostridium acetobutylicum*. *Molecular Microbiology* **100**, 289–302; 10.1111/mmi.13316 (2016).
208. Ren, C. *et al.* Pleiotropic functions of catabolite control protein CcpA in Butanol-producing *Clostridium acetobutylicum*. *BMC genomics* **13**, 349; 10.1186/1471-2164-13-349 (2012).
209. Ravagnani, A. *et al.* Spo0A directly controls the switch from acid to solvent production in solvent-forming clostridia. *Mol Microbiol* **37**, 1172–1185; 10.1046/j.1365-2958.2000.02071.x (2000).
210. Scotcher, M. C., Rudolph, F. B. & Bennett, G. N. Expression of abrB310 and SinR, and effects of decreased abrB310 expression on the transition from acidogenesis to solventogenesis, in *Clostridium acetobutylicum* ATCC 824. *Applied and Environmental Microbiology* **71**, 1987–1995; 10.1128/AEM.71.4.1987-1995.2005 (2005).
211. Xue, Q. *et al.* Roles of three AbrBs in regulating two-phase *Clostridium acetobutylicum* fermentation. *Applied Microbiology and Biotechnology* **100**, 9081–9089; 10.1007/s00253-016-7638-x (2016).
212. Hillmann, F. *et al.* The role of PerR in O₂-affected gene expression of *Clostridium acetobutylicum*. *J Bacteriol* **191**; 10.1128/JB.00351-09 (2009).
213. Merrick, M. J. In a class of its own - the RNA polymerase sigma factor σ^{54} (σ^N). *Mol Microbiol* **10**, 903–909; 10.1111/j.1365-2958.1993.tb00961.x (1993).
214. Hrdlickova, R., Toloue, M. & Tian, B. RNA-Seq methods for transcriptome analysis. *Wiley interdisciplinary reviews. RNA* **8**; 10.1002/wrna.1364 (2017).
215. Sharma, C. M. *et al.* The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* **464**, 250–255; 10.1038/nature08756 (2010).
216. Sharma, C. M. & Vogel, J. Differential RNA-seq. The approach behind and the biological insight gained. *Current opinion in microbiology* **19**, 97–105; 10.1016/j.mib.2014.06.010 (2014).
217. Ettwiller, L., Buswell, J., Yigit, E. & Schildkraut, I. A novel enrichment strategy reveals unprecedented number of novel transcription start sites at single base resolution in a model prokaryote and the gut microbiome. *BMC genomics* **17**, 199; 10.1186/s12864-016-2539-z (2016).
218. Boutard, M. *et al.* Global repositioning of transcription start sites in a plant-fermenting bacterium. *Nature communications* **7**, 13783; 10.1038/ncomms13783 (2016).
219. Zhu, Y. Y., Machleder, E. M., Chenchik, A., Li, R. & Siebert, P. D. Reverse Transcriptase Template Switching. A SMART™ Approach for Full-Length cDNA Library Construction. *BioTechniques* **30**, 892–897; 10.2144/01304pf02 (2001).
220. Yan, B., Boitano, M., Clark, T. A. & Ettwiller, L. SMRT-Cappable-seq reveals complex operon variants in bacteria. *Nature communications* **9**, 3676; 10.1038/s41467-018-05997-6 (2018).
221. Schwarz, K. *et al.* in *Direct Microbial Conversion of Biomass to Advanced Biofuels* (Elsevier2015), pp. 241–289.
222. Mermelstein, L. D. & Papoutsakis, E. T. *In vivo* methylation in *Escherichia coli* by the *Bacillus subtilis* phage phi 3T I methyltransferase to protect plasmids from restriction upon transformation of *Clostridium acetobutylicum* ATCC 824. *Applied and Environmental Microbiology* **59**, 1077–1081 (1993).
223. Kolek, J., Sedlar, K., Provaznik, I. & Patakova, P. Dam and Dcm methylations prevent gene transfer into *Clostridium pasteurianum* NRRL B-598: development of methods for electrotransformation, conjugation, and sonoporation. *Biotechnology for Biofuels* **9**, 14 (2016).
224. Sedlar, K., Kolek, J., Provaznik, I. & Patakova, P. Reclassification of non-type strain *Clostridium pasteurianum* NRRL B-598 as *Clostridium beijerinckii* NRRL B-598. *Journal of biotechnology* **244**, 1–3 (2017).
225. Annous, B. A. & Blaschek, H. P. Isolation and characterization of *Clostridium acetobutylicum* mutants with enhanced amyolytic activity. *Applied and Environmental Microbiology* **57**, 2544–2548 (1991).
226. Dürre, P., Kuhn, A. & Gottschalk, G. Treatment with allyl alcohol selects specifically for mutants of *Clostridium acetobutylicum* defective in butanol synthesis. *FEMS Microbiology Letters* **36**, 77–81 (1986).

227. Lambowitz, A. M. & Zimmerly, S. Mobile group II introns. *Annu. Rev. Genet.* **38**, 1–35 (2004).
228. Shao, L. *et al.* Targeted gene disruption by use of a group II intron (targetron) vector in *Clostridium acetobutylicum*. *Cell research* **17**, 963 (2007).
229. Dong, H., Zhang, Y., Dai, Z. & Li, Y. Engineering *Clostridium* strain to accept unmethylated DNA. *PLoS One* **5**, e9038 (2010).
230. Cui, G.-z. *et al.* Targeted gene engineering in *Clostridium cellulolyticum* H10 without methylation. *Journal of microbiological methods* **89**, 201–208 (2012).
231. Tolonen, A. C., Chilaka, A. C. & Church, G. M. Targeted gene inactivation in *Clostridium phytofermentans* shows that cellulose degradation requires the family 9 hydrolase Cphy3367. *Molecular Microbiology* **74**, 1300–1313; 10.1111/j.1365-2958.2009.06890.x (2009).
232. Xiao, H. *et al.* Metabolic engineering of D-xylose pathway in *Clostridium beijerinckii* to optimize solvent production from xylose mother liquid. *Metabolic engineering* **14**, 569–578; 10.1016/j.ymben.2012.05.003 (2012).
233. Zhang, L. *et al.* Ribulokinase and transcriptional regulation of arabinose metabolism in *Clostridium acetobutylicum*. *Journal of Bacteriology* **194**, 1055–1064 (2012).
234. Kuehne, S. A. & Minton, N. P. ClosTron-mediated engineering of *Clostridium*. *Bioengineered* **3**, 247–254 (2012).
235. Heap, J. T., Pennington, O. J., Cartman, S. T., Carter, G. P. & Minton, N. P. The ClosTron: a universal gene knock-out system for the genus *Clostridium*. *J Microbiol Methods* **70**; 10.1016/j.mimet.2007.05.021 (2007).
236. Heap, J. T. *et al.* The ClosTron: mutagenesis in *Clostridium* refined and streamlined. *Journal of microbiological methods* **80**, 49–55 (2010).
237. Al-Hinai, M. A., Fast, A. G. & Papoutsakis, E. T. Novel system for efficient isolation of *Clostridium* double-crossover allelic exchange mutants enabling markerless chromosomal gene deletions and DNA integration. *Applied and Environmental Microbiology* **78**, 8112–8121 (2012).
238. Soucaille, P., Figge, R. & Croux, C. Process for chromosomal integration and DNA sequence replacement in *Clostridia*. 2006. *International Patent W 2008*.
239. Heap, J. T. *et al.* Integration of DNA into bacterial chromosomes from plasmids without a counter-selection marker. *Nucleic Acids Research* **40**, e59–e59 (2012).
240. Reyrat, J.-M., Pelicic, V., Gicquel, B. & Rappuoli, R. Counterselectable markers: untapped tools for bacterial genetics and pathogenesis. *Infection and immunity* **66**, 4011–4017 (1998).
241. Jackson, S. A. *et al.* CRISPR-Cas. Adapting to change. *Science (New York, N.Y.)* **356**; 10.1126/science.aal5056 (2017).
242. Addgene. CRISPR History and Development for Genome Engineering. Available at <https://www.addgene.org/crispr/history/>.
243. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
244. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
245. Le Cong *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
246. Cui, L. & Bikard, D. Consequences of Cas9 cleavage in the chromosome of *Escherichia coli*. *Nucleic Acids Research* **44**, 4243–4251; 10.1093/nar/gkw223 (2016).
247. Xu, T. *et al.* Efficient Genome Editing in *Clostridium cellulolyticum* via CRISPR-Cas9 Nickase. *Applied and Environmental Microbiology* **81**, 4423–4431; 10.1128/AEM.00873-15 (2015).
248. Huang, H., Zheng, G., Jiang, W., Hu, H. & Lu, Y. One-step high-efficiency CRISPR/Cas9-mediated genome editing in *Streptomyces*. *Acta biochimica et biophysica Sinica* **47**, 231–243 (2015).
249. Tong, Y., Charusanti, P., Zhang, L., Weber, T. & Lee, S. Y. CRISPR-Cas9 based engineering of actinomycetal genomes. *ACS synthetic biology* **4**, 1020–1029 (2015).
250. Cobb, R. E., Wang, Y. & Zhao, H. High-efficiency multiplex genome editing of *Streptomyces* species using an engineered CRISPR/Cas system. *ACS synthetic biology* **4**, 723–728 (2014).
251. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature biotechnology* **31**, 233–239 (2013).
252. Oh, J.-H. & van Pijkeren, J.-P. CRISPR–Cas9-assisted recombineering in *Lactobacillus reuteri*. *Nucleic Acids Research* **42**, e131–e131 (2014).

References

253. Pyne, M. E., Bruder, M. R., Moo-Young, M., Chung, D. A. & Chou, C. P. Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in *Clostridium* **6**, 25666 EP - (2016).
254. Wang, Y. *et al.* Markerless chromosomal gene deletion in *Clostridium beijerinckii* using CRISPR/Cas9 system. *Journal of biotechnology* **200**, 1–5 (2015).
255. Wang, Y. *et al.* Bacterial genome editing with CRISPR-Cas9: deletion, Integration, single nucleotide modification, and desirable “clean” mutant selection in *Clostridium beijerinckii* as an example. *ACS synthetic biology* **5**, 721–732 (2016).
256. Wasels, F., Jean-Marie, J., Collas, F., López-Contreras, A. M. & Ferreira, N. L. A two-plasmid inducible CRISPR/Cas9 genome editing tool for *Clostridium acetobutylicum*. *Journal of microbiological methods* (2017).
257. Li, Q. *et al.* CRISPR-based genome editing and expression control systems in *Clostridium acetobutylicum* and *Clostridium beijerinckii*. *Biotechnology journal* **11**, 961–972 (2016).
258. Rauch, B. J. *et al.* Inhibition of CRISPR-Cas9 with bacteriophage proteins. *Cell* **168**, 150–158. e10 (2017).
259. Yang, H. & Patel, D. J. Inhibition mechanism of an anti-CRISPR suppressor AcrIIA4 targeting SpyCas9. *Molecular Cell* **67**, 117–127. e5 (2017).
260. Dong, D. *et al.* Structural basis of CRISPR-SpyCas9 inhibition by an anti-CRISPR protein. *Nature* **546**, 436–439; 10.1038/nature22377 (2017).
261. Diallo, M. *et al.* Adaptation and application of a two-plasmid inducible CRISPR-Cas9 system in *Clostridium beijerinckii*. *Methods*; 10.1016/j.ymeth.2019.07.022 (2019).
262. Hocq, R., Chartier, G., Wasels, F. & Lopes Ferreira, N. *Brevet FR18/73.492 Bactéries Clostridium génétiquement modifiées, préparation et utilisations de celles-ci (date de dépôt : 20/12/2018)* ,
263. Hocq, R., Wasels, F. & Lopes Ferreira, N. *Brevet FR19/05.485 : Outil génétique optimisé pour modifier les bactéries du genre Clostridium (date de dépôt : 24/05/2019)* ,
264. Gu, Y. *et al.* Curing the endogenous megaplasmid in *Clostridium saccharoperbutylacetonicum* N1-4 (HMT) using CRISPR-Cas9 and preliminary investigation of the role of the plasmid for the strain metabolism. *Fuel* **236**, 1559–1566; 10.1016/j.fuel.2018.09.030 (2019).
265. Dong, H., Tao, W., Gong, F., Li, Y. & Zhang, Y. A functional recT gene for recombineering of *Clostridium*. *Journal of biotechnology* **173**, 65–67; 10.1016/j.jbiotec.2013.12.011 (2014).
266. Vallenet, D. *et al.* MicroScope—an integrated microbial resource for the curation and comparative analysis of genomic and metabolic data. *Nucleic Acids Research* **41**, D636–D647 (2012).
267. Miura, K. *et al.* 5-fu metabolism in cancer and orally-administrable 5-fu drugs. *Cancers (Basel)* **2**, 1717–1730; 10.3390/cancers2031717 (2010).
268. Hocq, R., Bouilloux-Lafont, M., Lopes Ferreira, N. & Wasels, F. σ 54 (σ L) plays a central role in carbon metabolism in the industrially relevant *Clostridium beijerinckii*. *Scientific Reports* **9**, 7228; 10.1038/s41598-019-43822-2 (2019).
269. Merrick, M. & Chambers, S. The helix-turn-helix motif of sigma 54 is involved in recognition of the -13 promoter region. *Journal of Bacteriology* **174**, 7221–7226 (1992).
270. Barrios, H., Valderrama, B. & Morett, E. Compilation and analysis of sigma(54)-dependent promoter sequences. *Nucleic Acids Research* **27**, 4305–4313 (1999).
271. Wigneshweraraj, S. *et al.* Modus operandi of the bacterial RNA polymerase containing the sigma54 promoter-specificity factor. *Molecular Microbiology* **68**, 538–546; 10.1111/j.1365-2958.2008.06181.x (2008).
272. Nie, X., Dong, W. & Yang, C. Genomic reconstruction of σ 54 regulons in *Clostridiales*. *BMC genomics* **20**, 565; 10.1186/s12864-019-5918-4 (2019).
273. Shao, W., Price, M. N., Deutschbauer, A. M., Romine, M. F. & Arkin, A. P. Conservation of Transcription Start Sites within Genes across a Bacterial Genus. *mBio* **5**; 10.1128/mBio.01398-14 (2014).
274. Cortes, T. *et al.* Genome-wide mapping of transcriptional start sites defines an extensive leaderless transcriptome in *Mycobacterium tuberculosis*. *Cell reports* **5**, 1121–1131; 10.1016/j.celrep.2013.10.031 (2013).
275. Wellenreuther, R., Schupp, I., Poustka, A., Wiemann, S. & Consortium, T. G. c. SMART amplification combined with cDNA size fractionation in order to obtain large full-length clones. *BMC genomics* **5**, 36; 10.1186/1471-2164-5-36 (2004).

276. Siemerink, M. A. J., Schwarz, K., Kuit, W. & Kengen, S. W. M. in *Systems biology of Clostridium*, edited by P. Dürre (Imperial College Press, London, 2014).
277. Rochón, E., Cebreiros, F., Ferrari, M. D. & Lareo, C. Isopropanol-butanol production from sugarcane and sugarcane-sweet sorghum juices by *Clostridium beijerinckii* DSM 6423. *Biomass and Bioenergy* **128**, 105331; 10.1016/j.biombioe.2019.105331 (2019).
278. Alsaker, K. V., Paredes, C. & Papoutsakis, E. T. Metabolite stress and tolerance in the production of biofuels and chemicals: gene-expression-based systems analysis of butanol, butyrate, and acetate stresses in the anaerobe *Clostridium acetobutylicum*. *Biotechnol Bioeng* **105** (2010).
279. Cary, J. W., Petersen, D. J., Papoutsakis, E. T. & Bennett, G. N. Cloning and expression of *Clostridium acetobutylicum* phosphotransbutyrylase and butyrate kinase genes in *Escherichia coli*. *Journal of Bacteriology* **170**, 4613–4618 (1988).
280. Ma, K. & Adams, M. W. Ferredoxin:NADP oxidoreductase from *Pyrococcus furiosus*. *Methods in enzymology* **334**, 40–45; 10.1016/s0076-6879(01)34456-7 (2001).
281. Grimmler, C. *et al.* Genome-wide gene expression analysis of the switch between acidogenesis and solventogenesis in continuous cultures of *Clostridium acetobutylicum*. *J Mol Microbiol Biotechnol* **20**; 10.1159/000320973 (2011).
282. Jiang, M., Shao, W., Perego, M. & Hoch, J. A. Multiple histidine kinases regulate entry into stationary phase and sporulation in *Bacillus subtilis*. *Molecular Microbiology* **38**, 535–542; 10.1046/j.1365-2958.2000.02148.x (2000).
283. Deutscher, J., Kuster, E., Bergstedt, U., Charrier, V. & Hillen, W. Protein kinase-dependent HPr/CcpA interaction links glycolytic activity to carbon catabolite repression in gram-positive bacteria. *Mol Microbiol* **15**, 1049–1053; 10.1111/j.1365-2958.1995.tb02280.x (1995).
284. Jault, J. M. *et al.* The HPr kinase from *Bacillus subtilis* is a homo-oligomeric enzyme which exhibits strong positive cooperativity for nucleotide and fructose 1,6-bisphosphate binding. *Journal of Biological Chemistry* **275**, 1773–1780; 10.1074/jbc.275.3.1773 (2000).
285. Freier, D. & Gottschalk, G. I(+)-lactate dehydrogenase of *Clostridium acetobutylicum* is activated by fructose-1,6-bisphosphate. *FEMS Microbiology Letters* **43**, 229–233; 10.1111/j.1574-6968.1987.tb02128.x (1987).
286. Sauer, U. & Dürre, P. Sequence and molecular characterization of a DNA region encoding a small heat shock protein of *Clostridium acetobutylicum*. *Journal of Bacteriology* **175**, 3394–3400; 10.1128/jb.175.11.3394-3400.1993 (1993).
287. Tomas, C. A., Welker, N. E. & Papoutsakis, E. T. Overexpression of groESL in *Clostridium acetobutylicum* results in increased solvent production and tolerance, prolonged metabolism, and changes in the cell's transcriptional program. *Applied and Environmental Microbiology* **69**, 4951–4965; 10.1128/aem.69.8.4951-4965.2003 (2003).
288. Bahl, H., Müller, H., Behrens, S., Joseph, H. & Narberhaus, F. Expression of heat shock genes in *Clostridium acetobutylicum*. *FEMS microbiology reviews* **17**, 341–348; 10.1111/j.1574-6976.1995.tb00217.x (1995).
289. Pich, A., Narberhaus, F. & Bahl, H. Induction of heat shock proteins during initiation of solvent formation in *Clostridium acetobutylicum*. *Applied Microbiology and Biotechnology* **33**, 697–704; 10.1007/BF00604941 (1990).
290. Akanuma, G. *et al.* Ribosome dimerization is essential for the efficient regrowth of *Bacillus subtilis*. *Microbiology (Reading, England)* **162**, 448–458; 10.1099/mic.0.000234 (2016).
291. Horst, J.-P., Wu, T.-h. & Marinus, M. G. *Escherichia coli* mutator genes. *Trends in Microbiology* **7**, 29–36; 10.1016/S0966-842X(98)01424-3 (1999).
292. Ginetti, F., Perego, M., Albertini, A. M. & Galizzi, A. *Bacillus subtilis* mutS mutL operon. Identification, nucleotide sequence and mutagenesis. *Microbiology (Reading, England)* **142** (Pt 8), 2021–2029; 10.1099/13500872-142-8-2021 (1996).
293. Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *nature* **460**, 894–898; 10.1038/nature08187 (2009).
294. Nyerges, Á. *et al.* A highly precise and portable genome engineering method allows comparison of mutational effects across bacterial species. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 2502–2507; 10.1073/pnas.1520040113 (2016).
295. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* **152**, 1173–1183; 10.1016/j.cell.2013.02.022 (2013).
296. Wang, Y. *et al.* Gene transcription repression in *Clostridium beijerinckii* using CRISPR-dCas9. *Biotechnology and bioengineering* **113**, 2739–2743; 10.1002/bit.26020 (2016).

References

297. Abudayyeh, O. O. *et al.* RNA targeting with CRISPR-Cas13. *Nature* **550**, 280–284; 10.1038/nature24049 (2017).
298. Csorgo, B. *Development of CRISPR-Cas3 for large-scale genome manipulation* (New Approaches and Concepts in Microbiology Conference EMBL Heidelberg, 10/07/19).

9 ANNEXE

Le manuscrit inséré dans cette annexe est une version actualisée (mais non terminée) d'un document qui sera prochainement soumis à publication.

Rapidement, ce dernier décrit chez *Clostridium acetobutylicum* le système CRISPR/anti-CRISPR également utilisé au cours de cette thèse chez *C. beijerinckii* DSM 6423. Ce système permet d'obtenir systématiquement des colonies après transformation, en inhibant de façon contrôlée l'activité cytotoxique de Cas9. Le manuscrit décrit également l'utilisation de ce système pour supprimer de manière « propre » (sans cicatrice dans le chromosome) les gènes codant les différentes butanol déshydrogénases dépendantes du NADPH (BdhA, BdhB, BdhC) et les phénotypes correspondants (suppressions simples, doubles et triple). Ces phénotypes sont validés *in fine* par une complémentation via une série de plasmides.

1 **BdhA, BdhB and BdhC Contributions to Alcohol Production in *Clostridium***
2 ***acetobutylicum* DSM 792**

3

4

5

6 François Wasels,^a# Gwladys Chartier,^a Rémi Hocq,^a Nicolas Lopes Ferreira^a

7

8

9 ^aBiotechnology Department, IFP Energies nouvelles, Rueil-Malmaison, France

10

11

12 **Running title:** Bdh contribution to solvent production

13

14 **Keywords:** *Clostridium acetobutylicum*, alcohol dehydrogenases, metabolic
15 engineering, CRISPR-Cas9, Anti-CRISPR

16

17

18 # Address correspondence to François Wasels, francois.wasels@ifpen.fr

19 **ABSTRACT**

20 Although *Clostridium acetobutylicum* is the model organism for the study of ABE
21 fermentation, its study has long been impeded by the lack of efficient genetic tools for
22 genome editing. In particular, contribution of the six alcohol dehydrogenases
23 encoded by the reference strain has mostly been studied with the generation of
24 single gene deletion strains. In this study, we improved a recently developed
25 CRISPR-Cas9 tool by using an anti-CRISPR protein encoding gene, which allowed
26 us to realize modifications that we failed to achieve with the early version of our
27 genetic tool. In particular, we deleted the genes coding for NADPH-dependent
28 alcohol dehydrogenases BdhA, BdhB and BdhC. Since our tool allowed us to realize
29 iterative deletions , we were able to construct mutants with multiple gene deletions
30 and to better characterize the role of the corresponding enzymes. Although those
31 genes have been shown not to be major for alcohol synthesis, we show here that
32 simultaneous inactivation significantly affects solvent production. Complementation
33 experiments confirmed the key role of these enzymes. Those results are of interest
34 for the metabolic engineering of this industrially relevant organism, in particular for
35 the carbon flux reorientation towards industrially relevant products.

36

37 **IMPORTANCE**

38 Efficient methods for genetic manipulation of *C. acetobutylicum* are needed in order
39 to better understand the physiology of this relevant microorganism, and to construct
40 strains compatible with the industrial requirements. The genetic tool described in this
41 study allows relatively fast, precise, efficient, markerless and iterative modifications in
42 the genome of this bacterium, and potentially in other bacterial species. As an

43 example of application, markerless mutants in which up to three genes coding for
44 alcohol dehydrogenases are inactivated have been constructed and characterized
45 through fermentation assays.

46 INTRODUCTION

47 *Clostridium acetobutylicum* has been established as the model organism for the
48 study of ABE (Acetone - Butanol - Ethanol) fermentation. The need of alternatives to
49 petrochemistry has renewed the interest for this microorganism in the last decades.
50 Since *C. acetobutylicum* is mainly producing alcohols, enzymes implicated in their
51 production have been thoroughly studied.

52 In total, seven genes encoding enzymes with predicted alcohol dehydrogenase
53 activity have been identified in the genome of *C. acetobutylicum* DSM 792 (1). Genes
54 CA_P0162 and CA_P0035 code for AdhE1 and AdhE2, respectively. Those enzymes
55 contain both aldehyde dehydrogenase and alcohol dehydrogenase domains,
56 although AdhE1 only retains aldehyde dehydrogenase activity (2). The five other
57 identified genes (CA_C3298, CA_C3299, CA_C3375, CA_C3392, CA_P0059) code
58 for alcohol dehydrogenases (1). Out of them, three enzymes (BdhA, BdhB and BdhC,
59 encoded by CA_C3299, CA_C3298, and CA_C3392, respectively) have been shown
60 to be NADPH-dependent, unlike AdhE2 which is NADH-dependent (2). Mutants of
61 strains ATCC 824 and DSM 1731 have been constructed previously in two different
62 studies (3, 4), in which the three corresponding genes have been inactivated
63 separately with ClosTron-based approaches (5), which showed that those genes
64 have no or little contribution to solvent production. However, to date, there has been
65 no study about the concomitant deletion of several copies of *bdh* genes.

66 In fact, difficulties to genetically modify this bacterium, mainly because of poor
67 transformability and low homologous recombination (HR) efficiencies, have impeded
68 a better understanding of its metabolism and physiology. To overcome this issue,
69 genetic tools dedicated to the *Clostridium* genus in general and to *C. acetobutylicum*

70 in particular have been developed in the recent years (6). Lately, several research
71 groups working on *Clostridium* have developed tools based on CRISPR-Cas (7–12).
72 Briefly, the Cas nuclease is guided to a target site by a so-called guide RNA (gRNA),
73 where it will generate a double strand break (DSB). For most bacteria, including
74 those belonging to the *Clostridium* genus, those DSBs are lethal when occurring in
75 an essential nucleic acid, such as the chromosome, and are repaired through HR
76 when a template is available. Providing a template altering the target site allows the
77 bacteria to repair its genome and to prevent another DSB. The strength of the
78 strategy is that it constrains the bacteria to modify the target site with double allelic
79 exchange by using the editing template to survive. In some cases (10), endogenous
80 CRISPR-Cas systems have been exploited, but since none of them has been
81 identified in *C. acetobutylicum* strains so far, the heterologous Cas9 from
82 *Streptococcus pyogenes* is, to date, the most commonly used nuclease in this
83 bacterium.

84 CRISPR-Cas9 gene editing in *Clostridium* requires three elements: *cas9*, a gRNA
85 expression cassette, and an editing template. Given the low transformation and HR
86 frequencies in *Clostridium*, the use of constitutive promoters for the expression of
87 *cas9* and gRNA is generally not possible. In fact, only a few publications reported on
88 the successful generation of mutants when using constitutively expressed nucleases
89 and gRNA. Overall, most groups reported on the toxicity of Cas9, and the need to
90 decrease it. Some research groups are using variants of Cas9 that catalyze a single
91 strand break (SSB) instead of a DSB, which can be repaired by the cell and is
92 therefore less toxic (8, 13). Ultimately, the cell will be protected against the nickase
93 action only if it modifies the target site, through HR with the provided editing
94 template. On one hand, it is therefore easier to obtain transformants. On the other

95 hand, mutants are generally isolated at lower frequencies and after sub-culturing,
96 because of the less selective action of the nickase (6). Another strategy relies on the
97 use of inducible promoters to drive the expression of *cas9* and/or the gRNA (9, 11,
98 12). Once transformants in which *cas9* is repressed have been obtained, it is
99 possible to trigger the expression of the ribonucleoprotein complex through the
100 addition of the inducer, and to select edited cells.

101 Our group recently developed an inducible CRISPR-Cas9 tool for *Clostridium* based
102 on the use of two plasmids (14). In this system, the *cas9* gene is located on a first
103 plasmid, pCas9_{ind}, under the control of Pcm-tetO2/1, together with the *tetR* gene
104 whose product represses the transcription of *cas9* in the absence of
105 anhydrotetracycline (aTc) (15). While this first plasmid is invariable, the second
106 plasmid, generically designated as pGRNA, is target-specific and contains a gRNA
107 expression cassette and the editing template. Both pCas9_{ind} and pGRNA plasmids
108 are introduced into the cell sequentially, and once transformants have been obtained,
109 they are spotted on aTc-containing plates to trigger the expression of *cas9* and select
110 mutants.

111 Although this system has been successfully used to target several genes in the
112 genome of *C. acetobutylicum*, we report here on the difficulties encountered to obtain
113 transformants containing both pCas9 and pGRNA in some cases, most likely
114 because of a basal leaky expression in the absence of inducer of both the nuclease
115 and the gRNA. To improve our tool and overcome this problem, we modified pCas9_{ind}
116 by cloning a gene encoding an anti-Cas9 protein, AcrIIA4 (16), under the control of a
117 lactose-inducible promoter (17), allowing us to specifically control Cas9 nuclease
118 activity and thus to increase the transformation efficiency of pGRNA plasmids.

119 As a proof-of-concept and a mean to deepen our knowledge about the metabolism of
120 the industrially relevant *C. acetobutylicum* DSM 792 (also known as *C.*
121 *acetobutylicum* ATCC 824), we used this improved tool to create mutants in which
122 one or several genes coding for alcohol dehydrogenases were inactivated. In
123 particular, we targeted the NADPH-dependent enzymes, namely BdhA, BdhB and
124 BdhC. Therefore, mutants were generated in which one, two, or all of the
125 corresponding genes were deleted, and were further characterized in batch
126 fermentation assays. Subsequently, mutants displaying the stronger effects were
127 complemented with plasmids carrying each of these three *bdh* copies to confirm the
128 phenotypes obtained and better determine the contribution of each gene to alcohol
129 production.

130 RESULTS

131 Improvement of a CRISPR-Cas9 genetic tool

132 In order to determine the contribution of BdhA, BdhB and BdhC for butanol
133 production in *C. acetobutylicum* DSM 792, we first sought to create mutant strains
134 using a CRISPR-Cas9 genetic tool developed recently in our laboratory (14). First of
135 all, we implemented a protocol for the rapid replacement of the gRNA sequence in
136 the targeting plasmids. A cassette containing the aTc-inducible promoter Pcm-2tetO1
137 (15) and the gRNA chimeric sequence (18), separated by a 38 nt cassette containing
138 two BsaI sites was cloned into pEC750C, yielding pGRNA_{ind}. Upon restriction with
139 BsaI, overhang extremities were generated, allowing the insertion of two hybridized
140 single-strand oligonucleotides with extremities compatible with the linearized
141 pGRNA_{ind}, in order to generate a new target specific plasmid (Fig. 1A). This simple
142 modification allowed us to generate new pGRNA plasmids faster than through the
143 cloning of a synthesized complete gRNA expression cassette, as previously done in
144 our laboratory (14). We used this strategy to construct plasmids targeting *bdhA*, *bdhB*
145 and *bdhC*, designated as pGRNA-*bdhA*, pGRNA-*bdhB* and pGRNA-*bdhC*,
146 respectively.

147 The next step consisted in the cloning of SOE PCR products that can be used as
148 editing template for the reparation of the DSB created through the action of the Cas9-
149 gRNA complex (Fig. 1B). Editing templates were constructed in order to delete most
150 of the targeted genes and cloned into the appropriate vectors, yielding pGRNA-
151 Δ *bdhA*, pGRNA- Δ *bdhB* and pGRNA- Δ *bdhC*. Since *bdhA* and *bdhB* are located next
152 to each other in the DSM 792 chromosome, one single editing template was

153 designed for the deletion of both genes at once and cloned into pGRNA-bdhB to
154 obtain pGRNA- Δ *bdhA* Δ *bdhB*.

155 Those plasmids were then introduced in *C. acetobutylicum* DSM 792 containing
156 pCas9_{ind} (further referred to as DSM 792 (pCas9_{ind})). Notably, most of the assays
157 performed to introduce plasmids pGRNA- Δ *bdhB* or pGRNA- Δ *bdhA* Δ *bdhB* into this
158 recombinant strain failed, as no transformants were obtained. Assuming that this
159 result might be due to a toxic leaky expression of the *cas9* gene and the gRNA,
160 which would prevent the generation of transformants, we modified the pCas9_{ind}
161 plasmid by inserting both the *acrIIA4* gene under the control of P_{bgal} and the
162 associated transcriptional repressor encoding gene *bgaR* (17, 19), yielding pCas9_{acr}.
163 This modification allows the precise control of Cas9 activity, in addition to the aTc-
164 mediated transcriptional activation of the nuclease encoding gene. The procedure by
165 which pCas9_{acr} can be used in combination with pGRNA plasmids to create multiple
166 mutations in the genome of DSM 792 is shown in Fig. 1C.

167 Transformation of pGRNA plasmids into DSM 792 (pCas9_{ind}) and DSM 792
168 (pCas9_{acr}) was further assessed, using lactose-supplemented plates for the selection
169 of transformants. Transformation efficiencies are shown in Fig. 2. Although plasmids
170 pGRNA- Δ *bdhA* and pGRNA- Δ *bdhC* were introduced at similar rates in DSM 792
171 (pCas9_{ind}) or DSM 792 (pCas9_{acr}), the generation of transformants containing
172 pGRNA- Δ *bdhB* and pGRNA- Δ *bdhA* Δ *bdhB* was greatly improved when using a strain
173 containing pCas9_{acr} instead of pCas9_{ind}. Therefore, all further experiments were
174 performed using DSM 792 (pCas9_{acr}).

175 **Mutants construction**

176 Once transformants containing pCas9_{acr} and either pGRNA- $\Delta bdhA$, pGRNA- $\Delta bdhB$,
177 pGRNA- $\Delta bdhC$ or pGRNA- $\Delta bdhA\Delta bdhB$ were obtained, expression of *cas9* was
178 induced as previously described (14). For each modification, at least three aTc-
179 resistant colonies obtained from independent transformants were analyzed and
180 confirmed by PCR, yielding the corresponding mutants that we stored for further
181 experiments. The editing efficiency for every single modification was 100 % (data not
182 shown). To generate DSM 792 $\Delta bdhA\Delta bdhC$, DSM 792 $\Delta bdhB\Delta bdhC$ and DSM 792
183 $\Delta bdhA\Delta bdhB\Delta bdhC$, one DSM 792 $\Delta bdhC$ mutant was randomly chosen (i.e. the
184 first mutant obtained), streaked on 2YTG plates containing only erythromycin until
185 pGRNA- $\Delta bdhC$ was cured, and further transformed with plasmids pGRNA- $\Delta bdhA$,
186 pGRNA- $\Delta bdhB$ or pGRNA- $\Delta bdhA\Delta bdhB$, respectively. Once transformants were
187 obtained, expression of *cas9* was induced and aTc-resistant colonies obtained from
188 independent transformants were analyzed by PCR, confirmed, and saved for further
189 experiments. The editing efficiency for this new round of modification was 100 %
190 (data not shown). PCR analyzes of *bdhA-bdhB* and *bdhC* loci in each different
191 mutant are shown in [Fig. 3](#).

192 **Mutants characterization**

193 Those mutants were then characterized by batch fermentation performed in flasks.
194 For each type of mutation (i.e. deletion of one, two, or the three butanol
195 dehydrogenase encoding genes), the three randomly chosen independent mutants
196 selected were assessed. Final fermentation products are shown in [Table 2](#). No
197 significant difference was observed for the single deletion mutants compared to the
198 wild-type strain, nor for the double mutants $\Delta bdhA\Delta bdhC$ and $\Delta bdhB\Delta bdhC$. On the
199 opposite, final fermentation products concentrations were strongly affected in
200 mutants $\Delta bdhA\Delta bdhB$ and $\Delta bdhA\Delta bdhB\Delta bdhC$, with lower solvent titers and higher

201 acids concentrations, resulting in lower final pH consistent with the so-called acid
202 crash. Although ethanol and butanol production decrease in those mutants could
203 have been anticipated, it was interesting to observe that acetone production was also
204 significantly diminished. Despite this decrease, acetone became the main
205 fermentation products in terms of solvents. Hypothesizing that the inability of the
206 mutants to efficiently convert acids into solvents led to an irreversible pH decrease
207 and acid crashes, we performed pH-controlled batch fermentations of the wild-type
208 strain as well as a $\Delta bdhA\Delta bdhB\Delta bdhC$ mutant in bioreactor (Table 3). Although
209 solvent production of the mutant was increased compared to flask fermentations, the
210 final solvent concentrations were still strongly reduced compared to the wild-type
211 strain. The major difference observed compared to flask fermentation was the strong
212 accumulation of lactate that even exceeded butyrate and acetate concentrations.

213 **Mutants complementation**

214 To confirm that the phenotypes observed were due to the deletion of *bdh* genes, we
215 complemented $\Delta bdhA\Delta bdhB$ and $\Delta bdhA\Delta bdhB\Delta bdhC$ mutants with plasmids
216 allowing the expression of each gene separately, under the control of its native
217 promoter. As a control, the wild-type strain was also transformed with those
218 plasmids. For each complementation (i.e. each plasmid introduced in each strain),
219 three randomly chosen transformants were analyzed through batch fermentation
220 assays performed in flasks as previously done, and final fermentation products are
221 presented in Table 4. In the wild-type strain, overexpression of each gene had no
222 significant effect compared to the presence of the empty control vector. However, the
223 two $\Delta bdhA\Delta bdhB$ and $\Delta bdhA\Delta bdhB\Delta bdhC$ mutants assessed were successfully
224 complemented with any of the *bdh* copy, and still undergone acid crash when

225 transformed with an empty vector. In fact, even the phenotype of the $\Delta bdhA\Delta bdhB$
226 mutant was reverted by the overexpression of *bdhC*.

227 **DISCUSSION**

228 Efficient genetic tools are required to better understand biology, in particular in the
229 case of *C. acetobutylicum*. For this purpose, several dedicated tools have been
230 developed in the recent years, including those based on CRISPR-Cas approaches
231 whose proved their efficiency. Although powerful, CRISPR-Cas-based genetic tools
232 also have drawbacks. In *Clostridium*, the main one is the toxicity of the system, which
233 makes it difficult to use because of the low transformation and HR frequencies
234 observed in these bacteria. In some cases, Cas9n has been used instead of Cas9
235 because of the too high toxicity of the later, even when placing the corresponding
236 gene under the control of an inducible promoter. However, the use of nickases can
237 be laborious, for the very same reason, i.e. because it does not allow to eliminate all
238 unmodified cells. In other words, introduction of the CRISPR-Cas machinery is
239 facilitated when Cas9n is employed, but mutant selection is more efficient with the
240 use of a native Cas9.

241 In this study, we intended to use a two-plasmid CRISPR-Cas9 tool recently
242 developed to inactivate *bdhA*, *bdhB* and *bdhC*, the three genes identified in the
243 genome of *C. acetobutylicum* DSM 792 that are coding for NADPH-dependent
244 alcohol dehydrogenases. We first failed to generate DSM 792 $\Delta bdhB$ and DSM 792
245 $\Delta bdhA\Delta bdhB$ mutants, because we did not manage to obtain transformants
246 containing both pCas9_{ind} and the corresponding pGRNA plasmids. Although it cannot
247 be ruled out that transformation efficiency of pGRNA plasmids is affected by the
248 editing template, it is intuitively most likely linked to the gRNA strength. In fact, we
249 observed in our laboratory that transformation efficiency varies when using pGRNA
250 plasmids containing different protospacers, independently from the editing template
251 used (data not shown). In the example reported in this study, we did not manage to

252 obtain transformants of DSM 792 (pCas9_{ind}) with pGRNA- $\Delta bdhB$, until the last assay
253 that was done to compare transformation efficiencies of DSM 792 (pCas9_{ind}) and
254 DSM 792 (pCas9_{acr}), for which one experiment out of four independent replicates
255 allowed us to obtain three transformants. Similarly, no transformant was ever
256 obtained when introducing pGRNA- $\Delta bdhA\Delta bdhB$ in DSM 792 (pCas9_{ind}). Obviously,
257 a faster option to circumvent this problem would have been to change the gRNA
258 sequence targeting *bdhB*, especially since we facilitated the retargeting of pGRNA
259 plasmids compared to the early version of the tool (Fig. 1A). However, we already
260 faced this toxicity problem with a few other targeting plasmids in our laboratory (data
261 not reported), and we therefore wanted to modify our tool so that it would work with
262 any pGRNA plasmid created, whatever the gRNA sequence is. To do so, we lowered
263 the toxicity of the nuclease by adding a second regulation of its action through the
264 use of an anti-CRISPR protein. Interestingly, the transformation efficiency of pGRNA-
265 $\Delta bdhA$ and pGRNA- $\Delta bdhC$ was not improved, suggesting that placing *cas9* under
266 the control of an inducible promoter was sufficient when using those gRNA
267 sequences. However, transformation efficiencies with pGRNA- $\Delta bdhB$ and pGRNA-
268 $\Delta bdhA\Delta bdhB$ were greatly increased when DSM 792 (pCas9_{acr}) was used,
269 demonstrating that the modification of the plasmid effectively improved our tool. To
270 the best of our knowledge, anti-CRISPR proteins have been successfully used to
271 realize CRISPR-based genome editing in human cells (20) and in the archaea
272 *Sulfolobus islandicus* (21), and this is the first report of their use in bacteria.
273 Combining two types of activity control, we managed to design a strategy which
274 allows fast, reliable and efficient genome editing. It allows different insertion, deletion,
275 single nucleotide modification without the presence of any unwanted scar or antibiotic

276 resistance marker in the genome, and can therefore be used to perform several
277 rounds of modification.

278 The genes targeted in this study have been inactivated in previous studies with the
279 use of the Clostron genetic tool (5), with slightly different outcomes (3, 4). In the first
280 of these studies, *bdhA* and *bdhB* were successfully inactivated but the corresponding
281 strains *Cac-bdhA459s::CT* and *Cac-bdhB475a::CT* were not significantly different
282 from the wild-type strain, while deletion of *bdhC* was not investigated (3). In a more
283 recent study (4), BdhB was shown to be relevant for alcohol production, since the
284 corresponding mutant had lower titers in butanol and ethanol. Interestingly, BdhC
285 (referred to as YqhD) was shown to be involved in ethanol production, following the
286 observation that a $\Delta yqhD$ mutant had lower final ethanol titer, but was not affected in
287 terms of butanol production (4). Finally, deletion of BdhA had no significant effect, as
288 reported in the first study (3, 4).

289 Despite the fact that the Clostron has revolutionized the genetics of *Clostridium* by
290 allowing fast inactivation of genes in bacteria which were difficult to modify, it suffers
291 from some limitations. First, it has been shown that the disruption of a gene rather
292 than its clean deletion can trigger polar effects, which could explain the differences
293 observed between those two studies concerning the importance of BdhB. Moreover,
294 another limitation of the Clostron is that the inactivation of a gene requires its
295 disruption with an intron that contains an antibiotic resistance marker, which therefore
296 limits the number of modifications than can be done in the chromosome. For this
297 reason, the two studies cited above (3, 4) focused on the creation of single gene
298 deletion strains. Using our improved CRISPR-Cas9 genetic tool, it was possible to
299 create clean mutants in which several genes were inactivated. Moreover, these
300 mutants are more compatible with industrial applications, since their genome does

301 not contain antibiotic resistance marker. Finally, the efficiency of the tool allowed to
302 generate numerous independent mutants, providing a higher confidence in the
303 results obtained since they represent the mean of biological replicates instead of
304 technical replicates.

305 Results obtained in this study suggest that none of the *bdh* gene is essential for
306 solvent production. Indeed, every single deletion mutant retained the ability to
307 efficiently produce solvents in the conditions tested. Therefore, it can be
308 hypothesized that the two remaining *bdh* copies allow the synthesis of sufficient Bdh
309 enzymes in the cell for the efficient production of solvents. This result is in
310 accordance with most of the conclusions drawn from previous studies, although the
311 lower ethanol production in $\Delta bdhC$ and the diminished production of both ethanol and
312 butanol in $\Delta bdhB$ was not observed, unlike what was shown in one of those studies
313 (4). On the opposite, we observed that the deletion of several *bdh* copies could have
314 strong effects on the solvent production capacities of *C. acetobutylicum*. In particular,
315 *bdhA* and *bdhC* apparently have stronger impact than *bdhB*, since $\Delta bdhA\Delta bdhB$ is
316 the only double mutant with significantly affected fermentation products pattern
317 compared to the wild-type strain. These results are in accordance with data reporting
318 that BdhB and BdhA are more expressed than BdhC (2), and therefore are likely to
319 have a major role. As intuited, the triple mutant $\Delta bdhA\Delta bdhB\Delta bdhC$ was even more
320 heavily affected than the $\Delta bdhA\Delta bdhB$ mutant, although neither butanol nor ethanol
321 production was abolished, suggesting that other enzymes are still capable of
322 compensating the lower alcohol dehydrogenase activity in the mutant, most likely
323 AdhE2 and/or the products of CA_C3375 and CA_P0059. Notably, fermentations
324 performed in pH-controlled bioreactors did not allow to prevent acid accumulation,
325 showing the mutant incapacity to reduce acids. Instead, the triple mutant produced

326 large amounts of lactate in these conditions, most likely to regenerate its NAD(P)⁺
327 pool which is not produced by Bdh enzymes anymore. Interestingly, although final
328 solvent titers were lower than for the wild-type strain, a classical ratio
329 ethanol:acetone:butanol was obtained, unlike what was observed in flask
330 fermentation assays.

331 Complementation assays showed that any *bdh* copy is capable of reverting the acid
332 crash phenotype of $\Delta bdhA\Delta bdhB$ and $\Delta bdhA\Delta bdhB\Delta bdhC$ mutants in non pH-
333 controlled flasks fermentations. Although those genes have been introduced in the
334 cells under the control of their native promoter, the plasmid copy number probably
335 allows their overexpression. In fact, it results in an efficient fermentation of
336 $\Delta bdhA\Delta bdhB$ even when supplemented with *bdhC*, which is still present in the
337 genome of the mutant.

338 In summary, we improved a CRISPR-Cas9 genetic tool dedicated to the *Clostridium*
339 genus through the tunable control of Cas9 activity with the anti-CRISPR protein
340 AcrIIA4. As a proof-of-concept for this tool, we investigated the importance of the
341 three genes coding for NADPH-dependent alcohol dehydrogenases identified in the
342 genome of *C. acetobutylicum* DSM 792. Notably, while previous studies only focused
343 on single gene deletion that did not allow to figure out the major role of these
344 enzymes, our tool allowed us to construct multiple deletion mutants that showed the
345 major role of *bdh* genes for solvent production. In addition to the interest of these
346 results in terms of metabolic engineering, the tool developed in this study will allow to
347 step up the construction of clean mutants in order to enhance our understanding of
348 the molecular physiology of this industrially relevant micro-organism.

349 MATERIALS AND METHODS

350 Bacterial strains, plasmids and culture conditions

351 Relevant characteristics of the bacterial strains and plasmids used in this study are
352 listed in [Table 1](#). *C. acetobutylicum* was grown anaerobically at 34 °C in liquid 2YTG
353 medium (16 g L⁻¹ Bacto tryptone, 10 g L⁻¹ yeast extract, 4 g L⁻¹ NaCl and 5 g L⁻¹
354 glucose) or on solid 2YTG with 1.5% agar, supplemented with 40 µg mL⁻¹
355 erythromycin and/or 15 µg mL⁻¹ thiamphenicol if necessary. *Escherichia coli* was
356 grown aerobically at 37 °C and 200 rpm in liquid LB medium or solid LB with 1.5%
357 agar supplemented with erythromycin (500 µg mL⁻¹ for solid media and 100 µg mL⁻¹
358 for liquid media), chloramphenicol (25 µg mL⁻¹ for solid media and 12.5 µg mL⁻¹ for
359 liquid media), or tetracycline (20 µg mL⁻¹) if necessary.

360

361 DNA manipulation and plasmid construction

362 All enzymes used for DNA modification were purchased from New England Biolabs
363 (NEB). The Q5[®] High-Fidelity Polymerase (NEB) was used for all PCR amplifications,
364 except for colony PCRs, which were performed using OneTaq[®] DNA polymerase
365 (NEB). Genomic DNA (gDNA) was extracted from *C. acetobutylicum* using the
366 ChargeSwitch gDNA Mini Bacteria Kit (ThermoFisher Scientific). Plasmids and PCR
367 product purification kits were purchased from Qiagen. All primers used in this study
368 are listed in Supplementary [Table S1](#).

369 A fragment containing *bgaR* and a codon optimized *acrIIA4* under the control of
370 P_{bgal} was synthesized and cloned into the pEX-K4 vector (Eurofins Genomics). The
371 whole cassette was amplified with primers P01 and P02, digested with SacI and

372 cloned into pCas9_{ind} linearized with the same restriction enzyme, yielding pCas9_{acr}. A
373 fragment containing the Pcm-2tetO1 promoter fused with the chimeric gRNA
374 sequence, separated by a 38 nt spacer containing two BsaI restriction sites, was
375 synthesized and cloned into the pEX-K4 vector (Eurofins Genomics). This cassette
376 was amplified with primers P01 and P02, digested with SacI and cloned into
377 pEC750C, yielding pGRNA_{ind}. Sequences of the synthesized fragments are listed in
378 the Supplementary Data.

379 The gRNA target sites were chosen using Geneious R10.2
380 (<https://www.geneious.com>), which provide for each CRISPR-Cas9 potential target
381 site activity scores (ranging from 0 to 1) predicted with an algorithm (22), and
382 specificity scoring, i.e. the probability that the gRNA will not target another location in
383 the genome. The pGRNA-bdhA, pGRNA-bdhB and pGRNA-bdhC plasmids were
384 constructed by ligating into the dephosphorylated BsaI-digested pGRNA_{ind} the
385 hybridization product of 5'P-modified oligonucleotides couples P03-P04, P05-P06
386 and P07-P08, respectively. Briefly, hybridization was performed in a thermocycler
387 according to the following protocol: 500 nmole of each primer were mixed in a 10 μ l
388 reaction containing 1 mM Tris-HCl, denatured at 95°C during 30 s, and further
389 hybridized by decreasing the temperature by 1°C every 5 s 70 times.

390 The pGRNA- Δ *bdhA* plasmid was constructed by cloning the amplicon obtained using
391 a splicing by overhang extension (SOE) PCR with primer pairs P09/P10 and P11/P12
392 on DSM 792 gDNA, into pGRNA-bdhA digested with BamHI and Sall. The pGRNA-
393 Δ *bdhB* plasmid was constructed by cloning the SOE PCR product obtained using
394 primer pairs P13/P14 and P15/P16 on DSM 792 gDNA, into pGRNA-bdhB digested
395 with BamHI and Sall. Similarly, the pGRNA- Δ *bdhA* Δ *bdhB* plasmid was constructed
396 by cloning the SOE PCR product obtained using primer pairs P13/P17 and P18/P19

397 on DSM 792 gDNA, into the same BamHI/Sall double-digested pGRNA-bdhB. The
398 pGRNA- Δ *bdhC* plasmid was constructed by cloning the SOE PCR product obtained
399 using primer pairs P20/P21 and P22/P23 on DSM 792 gDNA, into pGRNA-bdhC
400 digested with XhoI and XbaI.

401 For complementation assays, plasmids pFW01-*bdhA*, pFW01-*bdhB* and pFW01-
402 *bdhC* were constructed by cloning the PCR products obtained with primer pairs
403 P24/P25, P26/P27 and P28/P29 into pFW01 digested with XhoI and Sall.

404

405 **Transformation, induction of *cas9* expression and mutant confirmation**

406 Plasmids were transformed into chemically competent NEB 10-beta competent *E.*
407 *coli* cells containing pAN2 (5) for DNA methylation. The vectors were then isolated
408 and electroporated into *C. acetobutylicum* DSM 792 as previously described (23).
409 Cells of DSM 792 containing either pCas9_{ind} or pCas9_{acr} were further transformed
410 with pEC750C and derivative plasmids (i.e. plasmids containing a gRNA expression
411 cassette with or without editing template) and selected on 2YTG solid media
412 supplemented with erythromycin, thiamphenicol and 40 mM lactose. Expression of
413 *cas9* was induced in transformants by exposure to aTc as described before (14),
414 except that inducer concentration was 1 $\mu\text{g mL}^{-1}$. Mutants were confirmed by
415 amplification of the *bdhA-bdhB* and *bdhC* loci performed on gDNA using primers
416 pairs P30/P31 and P32/P33, respectively.

417

418 **Fermentation assays**

419 Flasks batch fermentations were performed in modified Gapes medium (Gapes et al.,
420 1996) containing 2.5 g L⁻¹ yeast extract, 1 g L⁻¹ KH₂PO₄, 0.6 g L⁻¹ K₂HPO₄, 1 g L⁻¹
421 MgSO₄ · 7H₂O, 6.6 mg L⁻¹ FeSO₄ · 7H₂O, 0.1 g L⁻¹ 4-aminobenzoic acid, 5.8 g L⁻¹
422 CH₃COONH₄ and 80 g L⁻¹ glucose. Erythromycin (40 µg mL⁻¹) was used for
423 complementation assays. Fresh colonies were used to inoculate 5 mL Gapes media.
424 After 72 h of growth in an anaerobic chamber at 34°C without agitation, 1 mL of these
425 precultures was inoculated into 100-mL flasks containing 19 mL of the same liquid
426 medium. The flasks were sealed with rubber stoppers, and a pressure relief valve
427 system was punctured through the rubber stoppers to prevent overpressure. Flasks
428 were further incubated outside of the anaerobic chamber for 72 h at 34°C with
429 agitation at 120 rpm.

430 Bioreactor batch fermentations were performed in Gapes medium (Gapes et al.,
431 1996). Fresh colonies were used to inoculate 5 mL Gapes media. After 72 h of
432 growth in an anaerobic chamber at 34°C without agitation, 1 mL of these precultures
433 was inoculated into a 100-mL Erlenmeyer containing 40 mL of the same liquid
434 medium and incubated for another 24 h in the anaerobic chamber. This second
435 preculture was then used to inoculate a bioreactor containing 360 mL of Gapes
436 medium.

437 The fermentation samples were centrifuged at 5,000 g for 5 min and the supernatant
438 was diluted with an internal standard (0.5 g L⁻¹ propanol). The concentrations of the
439 solvents produced were determined by gas chromatography on a PoraBOND-Q
440 column (25 m length, 0.32 mm internal diameter, 0.5 µm film thickness; Agilent
441 Technologies) equipped with a flame ionization detector. Helium was used as carrier
442 gas, at a flow rate of 1.6 mL min⁻¹. Column was gradually heated from 50°C to 250°C
443 in a 30 min run.

444 Acids concentrations were quantified by HPLC (Aminex HPX-87H from Biorad
445 coupled to a Spectra System RI 150 refractometer and a Waters 2487 Dual λ UV
446 detector set at 210 nm). 0.01 M sulfuric acid mobile phase was used at a flowrate of
447 0.6 mL min⁻¹. Column temperature was set to 60°C.

448 Residual sugar quantities were determined by HPLC using an Aminex HXP-87P (Bio-
449 Rad) coupled to a Varian 350 RI refractometer for detection. Water was used as a
450 mobile phase at a flowrate of 0.4 mL/min.

451 **ACKNOWLEDGMENTS**

452 This research received no specific grant from any funding agency in the public,
453 commercial, or not-for-profit sectors.

454 **TABLE 1** Bacterial strains and plasmids used in this study

Bacterial strain or plasmid	Relevant Characteristics	Source or reference
Strains		
<i>C. acetobutylicum</i> DSM 792	Wild-type	DSMZ
<i>E. coli</i> NEB 10-beta		NEB
Plasmids		
pAN2	<i>tetA</i> , Φ 3T I gene, p15A origin	(5, 23)
pFW01	<i>ermB</i> , ColE1 origin, pCB102 origin	
pCas9 _{ind}	<i>ermB</i> , ColE1 origin, pCB102 origin, <i>cas9</i> (Pcm-tetO2/1 promoter), <i>tetR</i>	(14)
pCas9 _{acr}	pCas9 _{ind} derivative with <i>acrIIA4</i> (P _{gal} promoter) and <i>bgaR</i> insertions	This study
pEC750C	<i>catP</i> , ColE1 origin, pIP404 origin	(14)
pGRNA _{ind}	pEC750C derivative with gRNA expression cassette (Pcm-2tetO1 promoter) insertion	This study
pGRNA-bdhA	pGRNA _{ind} derivative targeting <i>bdhA</i>	This study
pGRNA- Δ <i>bdhA</i>	pGRNA-bdhA derivative with Δ <i>bdhA</i> editing template insertion	This study
pGRNA-bdhB	pGRNA _{ind} derivative targeting <i>bdhB</i>	This study
pGRNA- Δ <i>bdhB</i>	pGRNA-bdhB derivative with Δ <i>bdhB</i> editing template insertion	This study
pGRNA- Δ <i>bdhA</i> Δ <i>bdhB</i>	pGRNA-bdhB derivative with Δ <i>bdhA</i> Δ <i>bdhB</i> editing template insertion	This study
pGRNA-bdhC	pGRNA _{ind} derivative targeting <i>bdhC</i>	This study
pGRNA- Δ <i>bdhC</i>	pGRNA-bdhC derivative with Δ <i>bdhC</i> editing template insertion	This study
pFW01- <i>bdhA</i>	pFW01 derivative with <i>bdhA</i> insertion	This study
pFW01- <i>bdhB</i>	pFW01 derivative with <i>bdhB</i> insertion	This study
pFW01- <i>bdhC</i>	pFW01 derivative with <i>bdhC</i> insertion	This study

455

456 **TABLE 2** Fermentation products of the mutants analyzed in this study. Results shown are average values \pm standard deviations
 457 from at least three independent technical replicates. For mutants assays, values are from biological replicates.

Strain	Products (g/L) ^a								Glucose consumption (g/L)	Final pH
	Solvents				Acids					
	Ethanol	Acetone	Butanol	Total	Acetate	Butyrate	Lactate	Total		
WT	1.3 \pm 0.1	9.0 \pm 0.5	12.4 \pm 1.4	22.8 \pm 2.0	2.6 \pm 0.1	2.4 \pm 0.1	ND	5.0 \pm 0.2	69.2 \pm 0.2	5.1 \pm 0.0
$\Delta bdhA$	1.2 \pm 0.1	9.1 \pm 0.4	11.6 \pm 0.4	22.0 \pm 0.1	2.4 \pm 0.1	2.3 \pm 0.1	ND	4.7 \pm 0.0	69.4 \pm 0.2	5.2 \pm 0.0
$\Delta bdhB$	0.9 \pm 0.0	9.2 \pm 0.2	12.1 \pm 0.6	22.2 \pm 0.6	2.2 \pm 0.1	2.7 \pm 0.1	ND	4.9 \pm 0.2	71.0 \pm 2.4	5.3 \pm 0.0
$\Delta bdhC$	1.3 \pm 0.2	9.3 \pm 0.3	12.8 \pm 1.0	23.4 \pm 1.4	2.6 \pm 0.4	2.1 \pm 0.5	ND	4.7 \pm 0.9	71.4 \pm 2.6	5.3 \pm 0.1
$\Delta bdhA\Delta bdhB$	0.5 \pm 0.0	3.8 \pm 0.6	3.2 \pm 0.7	7.4 \pm 1.4	4.5 \pm 0.4	5.0 \pm 0.2	0.2 \pm 0.1	9.7 \pm 0.4	35.9 \pm 2.2	4.8 \pm 0.0
$\Delta bdhA\Delta bdhC$	1.1 \pm 0.2	9.5 \pm 0.5	13.3 \pm 0.1	23.9 \pm 0.4	2.2 \pm 0.2	2.8 \pm 0.4	ND	5.0 \pm 0.2	72.0 \pm 0.6	5.2 \pm 0.0
$\Delta bdhB\Delta bdhC$	1.0 \pm 0.1	9.3 \pm 1.1	11.6 \pm 0.7	21.9 \pm 1.6	2.0 \pm 0.2	2.9 \pm 0.5	ND	4.9 \pm 0.2	73.5 \pm 1.8	5.3 \pm 0.0
$\Delta bdhA\Delta bdhB\Delta bdhC$	0.4 \pm 0.0	2.5 \pm 0.2	1.3 \pm 0.2	4.2 \pm 0.3	5.3 \pm 0.3	5.9 \pm 0.2	0.2 \pm 0.1	11.4 \pm 0.5	26.7 \pm 1.8	4.7 \pm 0.0

^a ND, Not Detected

458

459 **TABLE 3** Fermentation products in bioreactors

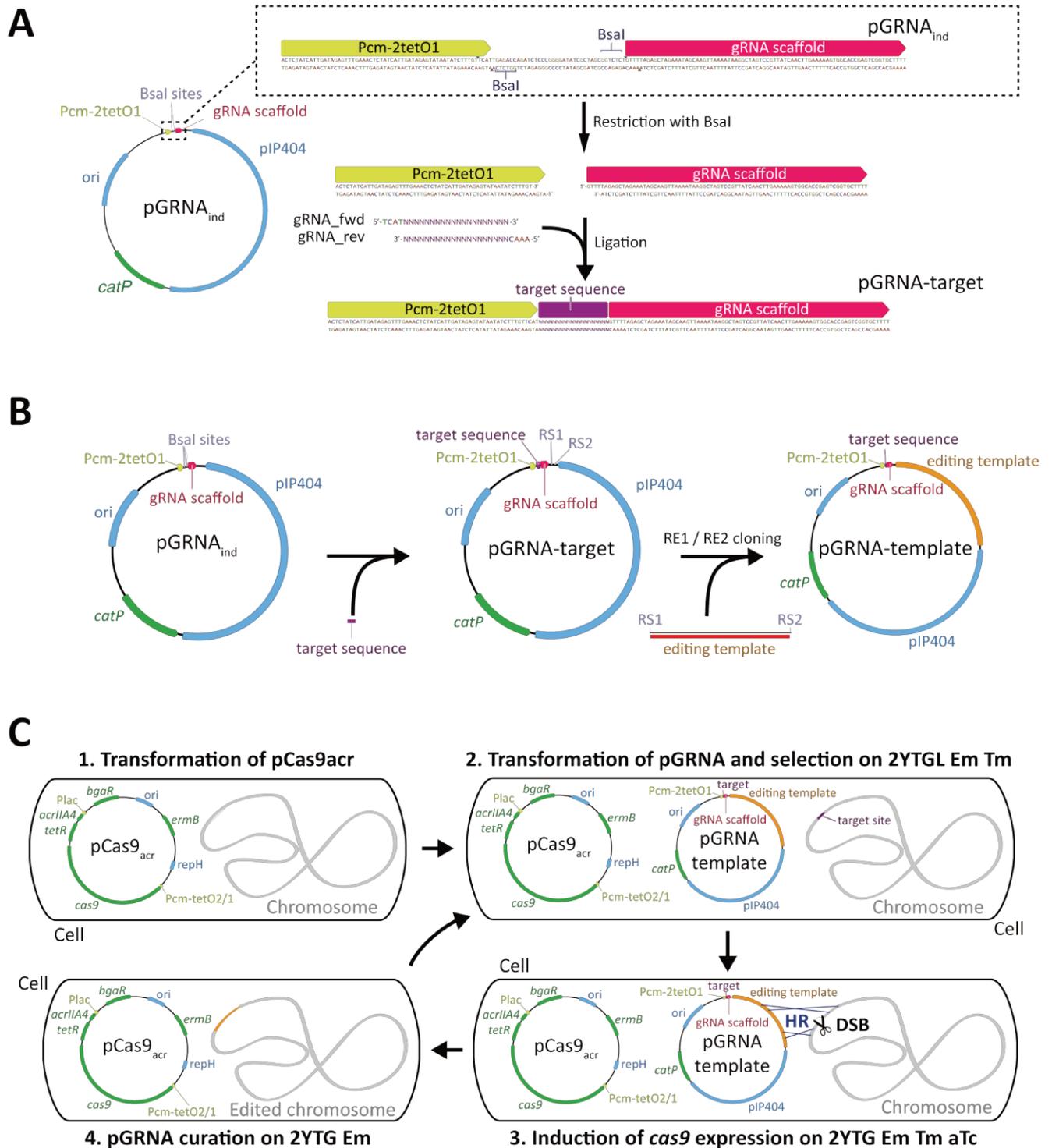
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492

Strain	Products (g/L)								Glucose ⁴⁶² consumption (g/L)
	Solvents				Acids				
	Ethanol	Acetone	Butanol	Total	Acetate	Butyrate	Lactate	Total	
WT	1.1	4.7	13.4	19.1	2.8	2.7	0.5	6.0	
<i>ΔbdhAΔbdhBΔbdhC</i>	0.3	1.6	3.7	5.6	5.1	5.1	6.6	16.8	

464 **TABLE 4** Complementation assays. Results shown are average values \pm standard deviations from at least three independent
 465 biological replicates.

	Ethanol	Acétone	Isopropanol	Butanol	Solvants					
DSM 792 pFW01	1,8	0,1	8,9	0,3	0,0	0,0	15,5	1,0	26,2	1,4
DSM 792 pFW01-bdhA										
DSM 792 pFW01-bdhB	1,6	0,2	8,4	0,8	0,2	0,0	14,0	2,1	24,2	3,0
DSM 792 pFW01-bdhC	1,8	0,3	9,6	0,4	0,0	0,0	16,2	0,5	27,7	0,9
DSM 792 Δ bdhA Δ bdhB pFW01	0,7	0,0	2,6	0,5	0,0	0,0	1,8	0,6	5,1	1,1
DSM 792 Δ bdhA Δ bdhB pFW01- bdhA	1,9	0,1	9,6	0,1	0,0	0,0	17,3	0,7	28,9	0,7
DSM 792 Δ bdhA Δ bdhB pFW01- bdhB	1,8	0,0	8,8	0,3	0,4	0,2	16,4	0,4	27,4	0,2
DSM 792 Δ bdhA Δ bdhB pFW01- bdhC	1,4	0,1	10,2	0,3	0,0	0,0	15,5	0,2	27,1	0,5
DSM 792 Δ bdhA Δ bdhB Δ bdhC pFW01	0,5	0,0	2,3	0,2	0,0	0,0	0,7	0,2	3,6	0,4
DSM 792 Δ bdhA Δ bdhB Δ bdhC pFW01-bdhA	1,4	0,1	9,6	0,2	0,0	0,0	15,2	0,7	26,2	0,9
DSM 792 Δ bdhA Δ bdhB Δ bdhC pFW01-bdhB	1,7	0,2	8,7	0,5	0,1	0,1	16,8	0,6	27,2	1,4
DSM 792 Δ bdhA Δ bdhB Δ bdhC pFW01-bdhC	1,1	0,0	9,9	0,3	0,0	0,0	14,6	0,0	25,6	0,3

466



468

469

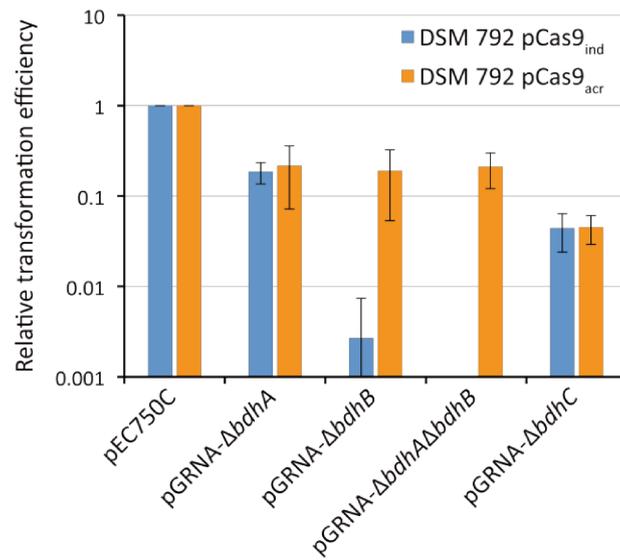
470

471

472

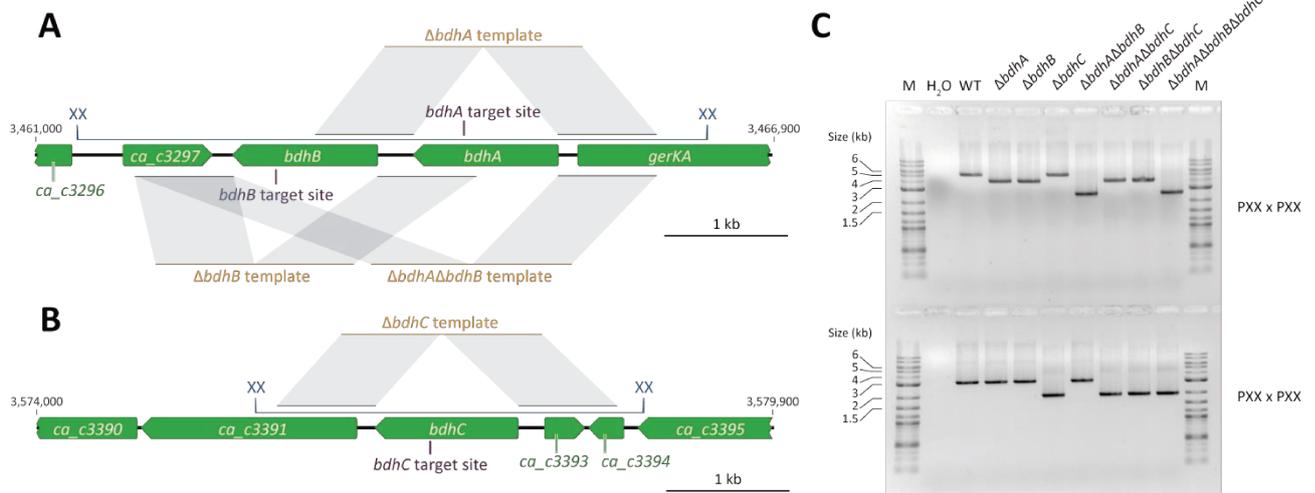
FIG 1 Easy reprogramming of the two plasmid CRISPR-Cas9 genetic tool for multiple genome edition. (A) Retargeting of the pGRNA plasmid through the cloning of two hybridized primers between the aTc-inducible promoter Pcm-2tetO1 and the gRNA scaffold. (B) Insertion of the editing template within the pGRNA plasmid by restriction

473 enzyme-based cloning. (C) Iterative genome modification. Step 1, the strain is
474 transformed with pCas9_{acr} and transformants are selected on erythromycin-
475 containing media. Step 2, the resulting strain is transformed with a pGRNA editing
476 plasmid and transformants are selected on media containing erythromycin,
477 thiamphenicol and lactose to activate the transcription of *acrIIA4*. Step 3, expression
478 of *cas9* is triggered on media containing erythromycin, thiamphenicol and aTc, but no
479 lactose. Step 4, once aTc-resistant colonies have been confirmed, pGRNA is curated
480 on media containing erythromycin and no thiamphenicol, yielding cells ready for a
481 new round of modification. ori, replication origin for *E. coli*; pip404 and repH,
482 compatible replication origins for *C. acetobutylicum*; *catP*,
483 thiamphenicol/chloramphenicol-resistance gene; *ermB*, erythromycin-resistance
484 gene; RS, restriction site; RE, restriction enzyme; DSB, double strand break, HR,
485 homologous recombination.



486

487 **FIG 2** Relative transformation efficiencies of pGRNA-template plasmids in DSM 792
 488 containing pCas9ind or pCas9acr. Results shown are average values \pm standard
 489 deviations from at least four independent replicates for each transformation.



490

491 **FIG 3** PCR analyses of representative deletion mutants constructed in this study. (A

492 and B) *bdh* loci from DSM 792. Genes are shown in green, gRNA target sites are

493 indicated in violet, homology sequenced between gDNA and the editing templates

494 are figured in grey. (A) *bdhA*- *bdhB* locus (B) *bdhC* locus. (C) Amplification of the

495 *bdhA*-*bdhB* and *bdhC* loci. Amplification with primers pair P30-P31 yields a, a, and a

496 bp product in WT, $\Delta bdhA$ and $\Delta bdhA\Delta bdhB$ strains, respectively. Amplification with

497 primers pair P32-P33 yields a, and a bp product in WT and $\Delta bdhC$ strains,

498 respectively. Lane M, 2-log DNA ladder (0.1 to 10 kb; NEB); Lane H₂O, amplification

499 on water.

500 References

- 501 1. NNilling J, Breton G, Omelchenko MV, Makarova KS, Zeng Q, Gibson R, Lee
502 HM, Dubois J, Qiu D, Hitti J, Wolf YI, Tatusov RL, Sabathe F, Doucette-Stamm
503 L, Soucaille P, Daly MJ, Bennett GN, Koonin EV, Smith DR. 2001. Genome
504 sequence and comparative analysis of the solvent-producing bacterium
505 *Clostridium acetobutylicum*. *J Bacteriol* 183:4823–4838.
506 doi:10.1128/JB.183.16.4823-4838.2001.
- 507 2. Yoo M, Bestel-Corre G, Croux C, Riviere A, Meynial-Salles I, Soucaille P. 2015.
508 A Quantitative System-Scale Characterization of the Metabolism of *Clostridium*
509 *acetobutylicum*. *MBio* 6:e01808-15. doi:10.1128/mBio.01808-15.
- 510 3. Cooksley CM, Zhang Y, Wang H, Redl S, Winzer K, Minton NP. 2012. Targeted
511 mutagenesis of the *Clostridium acetobutylicum* acetone-butanol-ethanol
512 fermentation pathway. *Metab Eng* 14:630–641.
513 doi:10.1016/j.ymben.2012.09.001.
- 514 4. Dai Z, Dong H, Zhang Y, Li Y. 2016. Elucidating the contributions of multiple
515 aldehyde/alcohol dehydrogenases to butanol and ethanol production in
516 *Clostridium acetobutylicum*. *Sci Rep* 6:28189. doi:10.1038/srep28189.
- 517 5. Heap JT, Pennington OJ, Cartman ST, Carter GP, Minton NP. 2007. The
518 Clostron. A universal gene knock-out system for the genus *Clostridium*. *Journal*
519 *of Microbiological Methods* 70:452–464. doi:10.1016/j.mimet.2007.05.021.
- 520 6. Joseph RC, Kim NM, Sandoval NR. 2018. Recent Developments of the Synthetic
521 Biology Toolkit for *Clostridium*. *Frontiers in Microbiology* 9:154.
522 doi:10.3389/fmicb.2018.00154.
- 523 7. Wang Y, Zhang Z-T, Seo S-O, Choi K, Lu T, Jin Y-S, Blaschek HP. 2015.
524 Markerless chromosomal gene deletion in *Clostridium beijerinckii* using CRISPR-
525 Cas9 system. *J Biotechnol* 200:10mal gene deletionjbiotec.2015.02.005.
- 526 8. Xu T, Li Y, Shi Z, Hemme CL, Li Y, Zhu Y, van Nostrand JD, He Z, Zhou J. 2015.
527 Efficient Genome Editing in *Clostridium cellulolyticum* via CRISPR-Cas9
528 Nickase. *Appl Environ Microbiol* 81:4423323.1016/j.mimet.2007.05.021.1..
- 529 9. Nagaraju S, Davies NK, Walker DJF, KKu Y, van Nostrand JD, He Genome
530 editing of *Clostridium autoethanogenum* using CRISPR-Cas9. *Biotechnol*
531 *Biofuels* 9:219. doi:10.1186/s13068-016-0638-3.
- 532 10. Pyne ME, Bruder MR, Moo-Young M, Chung DA, Chou CP. 2016. Harnessing
533 heterologous and endogenous CRISPR-Cas machineries for efficient markerless
534 genome editing in *Clostridium*. *Sci Rep* 6:25666. doi:10.1038/srep25666.
- 535 11. Wang Y, Zhang Z-T, Seo S-O, Lynn P, Lu T, Jin Y-S, Blaschek HP. 2016.
536 Bacterial Genome Editing with CRISPR-Cas9. Deletion, Integration, Single
537 Nucleotide Modification, and Desirable "Clean" Mutant Selection in *Clostridium*
538 *beijerinckii* as an Example. *ACS Synth Biol* 5:721–732.
539 doi:10.1021/acssynbio.6b00060.
- 540 12. Wang S, Dong S, Wang P, Tao Y, Wang Y. 2017. Genome Editing in *Clostridium*
541 *saccharoperbutylaceticum* N1-4 with the CRISPR-Cas9 System. *Appl Environ*
542 *Microbiol* 83. doi:10.1128/AEM.00233-17.
- 543 13. Li Q, Chen J, Minton NP, Zhang Y, Wen Z, Liu J, Yang H, Zeng Z, Ren X, Yang
544 J, Gu Y, Jiang W, Jiang Y, Yang S. 2016. CRISPR-based genome editing and
545 expression control systems in *Clostridium acetobutylicum* and *Clostridium*
546 *beijerinckii*. *Biotechnol J* 11:961n NP, Zhang Y, Wen Z, Liu J, Yang 14.
547 Wasels, F., Jean-Marie, J., Collas, F., LL, LJ, Yang H, Zeng Z, Ren X, Yang J,
548 Gu 2017. A two-plasmid inducible CRISPR-Cas9 genome editing tool for
549 *Clostridium acetobutylicum*. *Journal of Microbiological Methods* 140:5ological

550 Methods 140:5Jiang Y, Yang S.15. Dong H, Tao W, Zhang Y, Li Y. 2012.
551 Development of an anhydrotetracycline-inducible gene expression system for
552 solvent-producing *Clostridium acetobutylicum*. A useful tool for strain
553 engineering. *Metab Eng* 14:59–67. doi:10.1016/j.ymben.2011.10.004.
554 16. Rauch BJ, Silvis MR, Hultquist JF, Waters CS, McGregor MJ, Krogan NJ,
555 Bondy-Denomy J. 2017. Inhibition of CRISPR-Cas9 with Bacteriophage Proteins.
556 *Cell* 168:150-158.e10. doi:10.1016/j.cell.2016.12.009.
557 17. Hartman AH, Liu H, Melville SB. 2011. Construction and characterization of a
558 lactose-inducible promoter system for controlled gene expression in *Clostridium*
559 *perfringens*. *Appl Environ Microbiol* 77:471–478. doi:10.1128/AEM.01536-10.
560 18. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A
561 programmable dual-RNA-guided DNA endonuclease in adaptive bacterial
562 immunity. *Science* 337:816ski K, Fonfara I, Hauer M, Doudna
563 19. Al-Hinai MA, Fast AG, Papoutsakis ET. 2012. Novel system for efficient isolation
564 of *Clostridium* double-crossover allelic exchange mutants enabling markerless
565 chromosomal gene deletions and DNA integration. *Appl Environ Microbiol*
566 78:8112–8121. doi:10.1128/AEM.02214-12.
567 20. Shin J, Jiang F, Liu J-J, Bray NL, Rauch BJ, Baik SH, Nogales E, Bondy-
568 Denomy J, Corn JE, Doudna JA. 2017. Disabling Cas9 by an anti-CRISPR DNA
569 mimic. *Sci Adv* 3:e1701620. doi:10.1126/sciadv.1701620.
570 21. Mayo-Muu-Mung F, Liu J-J, Bray NL, Rauch BJ, Baik SH, Nogales E, Bondy-
571 Denomy J, Corn JE, Doudna JA. 2017. Disabling Cas9 by an anti-CRISPR DNA
572 mimic. *Sci Adv* 3:e1701620. doi:10.1126/sciadv.1701620. doi:10.11222.
573 Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I,
574 Sullender M, Ebert BL, Xavier RJ, Root DE. 2014. Rational design of highly
575 active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol*
576 32:12622262.36-10.mben.2011.10.004.
577 23. Mermelstein LD, Papoutsakis ET. 1993. In vivo methylation in *Escherichia coli* by
578 the *Bacillus subtilis* phage phi 3T I methyltransferase to protect plasmids from
579 restriction upon transformation of *Clostridium acetobutylicum* ATCC 824. *Appl*
580 *Environ Microbiol* 59:1077–1081.
581

Title : *Clostridium beijerinckii* DSM 6423, an emerging platform strain for solvent bioproduction

Keywords : *Clostridium*, metabolism, solvents, fermentation, omics, genome editing

Abstract : *Clostridium beijerinckii* DSM 6423 is an atypical bacterial strain which can ferment a wide array of substrates and naturally produce a solvent mixture composed of isopropanol, butanol and ethanol (IBE). Despite its potential, the prospect of using this microorganism for a biobased isopropanol synthesis process remains limited. In particular, its poor selectivity to this alcohol and its weak tolerance to the solvent mixture reduce isopropanol recoverability, although these drawbacks might be alleviated by targeted genome editing approaches. For this matter, developing genetic tools, that were until this work inexistent for this bacterium, and increasing our knowledge of its metabolism are important steps. In this context, a first IFPEN thesis did the ground work by drawing its genetic map and isolating isopropanol-tolerant mutants.

Based on this context, my thesis first aimed at developing efficient targeted genetic tools for this relevant strain. Use of Dam⁻ Dcm⁻ plasmids permitted for the first time the successful transformation of this microorganism, enabling the development of a reusable, scarless, genome editing tool based on the CRISPR-Cas9 technology. As a proof of concept, the 2 genes CIBE_0562 (*upp*) and CIBE_3859 (*catB*) were inactivated. This second modification made the strain sensitive to thiamphenicol and allowed the use of a dual-plasmid CRISPR system that had previously been engineered in IFPEN to modify the genome of a reference solventogenic species: *Clostridium acetobutylicum*. After the deletion of the pNF2 endogenous plasmid, transformation efficiencies were drastically increased, resulting in a $\Delta catB \Delta pNF2$ platform strain that may be used in the future to construct an efficient strain for isopropanol production.

In addition to the development of these genetic tools, my thesis also aimed at characterizing the microorganism by different approaches. Firstly, we studied the impact of several model or industrial carbon sources on the strain performances and on its metabolism. Focusing on three substrates (glucose, sucrose, beetroot molasse), a continuous fermentation process was next designed to isolate the two IBE-characteristic physiological states and coupled to a multi-omic analysis. Proteomics preliminary results highlight the severe impacts of industrial substrates on growth.

Secondly, the Capp-Switch sequencing transcriptomic approach was used to map transcription start sites (TSS) at the genome scale for the DSM 6423 strain and two other reference solventogenic strains. A comparative TSS analysis notably revealed a differential regulation of the butyryl-CoA synthesis operon between the *C. beijerinckii* and *C. acetobutylicum* species.

Although initially not planned at the beginning of this work, the sequencing of the AA strain, one of the mutants that were obtained in the first thesis and that is unable to produce isopropanol, as well as a targeted genome approach reproducing its phenotype in a *C. beijerinckii* model strain, also revealed the crucial role of σ^{54} in controlling solvent synthesis and alternative sugar uptake pathways.

To conclude, by deepening our knowledge on *C. beijerinckii* DSM 6423 and by allowing targeted genome editing approaches, my thesis is involved in a series of work conducted at IFPEN that aim at making this strain an IBE fermentation model organism, in order to use it for the bioproduction of isopropanol at the industrial scale.

Titre : *Clostridium beijerinckii* DSM 6423, une souche plateforme émergente pour la bioproduction de solvants

Mots-clés : *Clostridium*, métabolisme, solvants, fermentation, omiques, édition du génome

Résumé : *Clostridium beijerinckii* DSM 6423 est une souche bactérienne atypique capable de fermenter une variété de substrats et de produire naturellement un mélange de solvants composé d'isopropanol, de butanol et d'éthanol (IBE). Malgré ce potentiel, des limitations subsistent pour son utilisation dans un procédé de synthèse d'isopropanol biosourcé. En particulier, sa faible sélectivité envers cet alcool et sa faible tolérance au mélange de solvants réduisent la quantité de produits récupérables. Ces défauts pourraient être corrigés par des modifications ciblées du génome de cette souche, mais ceci passe par le développement d'outils génétiques, jusqu'à présent inexistant pour cette bactérie, et une meilleure compréhension de son métabolisme. Dans ce contexte, une première thèse réalisée à IFPEN a permis d'établir sa carte génétique et d'isoler de premiers mutants présentant une tolérance accrue à l'isopropanol.

Dans la continuité de ces premiers résultats, les travaux de thèse présentés dans ce manuscrit se sont d'abord attachés à développer des outils de génétique ciblée performants. L'utilisation de plasmides Dam⁻Dcm⁻ a permis pour la première fois de transformer ce microorganisme avec succès, ouvrant la voie au développement d'un outil basé sur la technologie CRISPR-Cas9, qui révolutionne l'ingénierie génétique en modifiant les génomes de manière itérative et sans laisser de trace. Comme preuve de concept, les 2 gènes CIBE_0562 (*upp*) et CIBE_3859 (*catB*) ont été inactivés. Cette deuxième modification a rendu ce micro-organisme sensible au thiamphénicol et compatible avec un système CRISPR à deux plasmides développé précédemment à IFPEN pour modifier le génome d'une souche solvantogène de référence, *Clostridium acetobutylicum*. Après suppression du plasmide endogène pNF2, les efficacités de transformation de la souche ont été drastiquement accrues, nous permettant ainsi d'obtenir une souche plateforme $\Delta catB \Delta pNF2$ qui pourra constituer le point de départ de la construction d'une souche performante pour la production d'isopropanol.

Parallèlement au développement de ces outils génétiques, ma thèse s'est également consacrée à la caractérisation du microorganisme par plusieurs approches. La première a consisté à déterminer l'impact de différents substrats, modèles ou industriels, sur les performances et le métabolisme de la souche. Après sélection de trois substrats (glucose, saccharose, mélasse de betterave), un procédé de fermentation continue a été mis au point pour isoler les deux états physiologiques caractéristiques d'une fermentation IBE, puis couplé à une analyse multi-omiques. Les résultats préliminaires, s'appuyant sur les données obtenues par protéomique, soulignent l'impact très important de l'utilisation de substrats industriels sur la croissance.

Deuxièmement, une approche transcriptomique, le *Capp-Switch sequencing*, a été employée pour cartographier, à l'échelle du génome, l'ensemble des sites d'initiation de la transcription (TSS) pour la souche DSM 6423 et deux autres souches solvantogènes de référence. Une analyse comparative des TSS a notamment révélé une régulation différentielle de l'opéron de synthèse du butyryl-CoA entre les espèces *C. beijerinckii* et *C. acetobutylicum*.

Bien que non prévu au début de ce travail, le séquençage d'un mutant obtenu lors de travaux précédents et incapable de produire de l'isopropanol (souche AA), puis la reconstruction de son phénotype dans une souche modèle de *C. beijerinckii*, a par ailleurs révélé le rôle fondamental du facteur σ^{54} pour le contrôle des voies de synthèse des solvants et d'assimilation de sucres alternatifs.

En conclusion, en approfondissant nos connaissances sur *C. beijerinckii* DSM 6423 et en permettant des modifications ciblées de son génome, ma thèse s'inscrit dans un ensemble de travaux menés à IFPEN visant à faire de cette souche un organisme modèle pour la fermentation IBE, permettant à terme la bioproduction d'isopropanol à échelle industrielle.