Systems Biology for Power-to-Gas Applications with *Methanothermobacter* spp.

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

B.Sc. Isabella Casini
aus Poggibonsi, Italien

Tübingen
2022

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	11.11.2022
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter:	Prof. Dr. Largus T. Angenent
2. Berichterstatter:	Prof. Dr. Boris Maček

I dedicate my dissertation to the establishment of an environmentally, socially, and economically sustainable world in which one's demographics and identity do not hinder their access to education or the possibility to pursue their passions.

Summary

The growing demand for power production with the concurrent rise in atmospheric greenhouse gas concentrations requires new solutions for the energy and waste management sectors. Biotechnological approaches, such as the implementation of living microbes or enzymes, are emerging as solutions. Microbes are attractive biocatalysts for both the storage of renewable electric power and the conversion of waste streams into valorized products. The power-to-gas platform converts excess renewable electric power by electrolyzing water to produce molecular hydrogen and oxygen. The energy can be stored as the hydrogen gas or further converted into other products, such as methane. Methane, which is the main component in natural gas, can be combusted and used in the natural gas infrastructure. Hydrogenotrophic methanogens are gas-fermenting microbes that can be used as biocatalysts to convert this hydrogen gas into methane (and biomass) *via* methanogenesis. Carbon dioxide from the atmosphere or from combustion processes with non-fossil fuels should ideally be utilized as the carbon source for these microbes. The company Electrochaea GmbH has commercialized this process using a thermophilic *Methanothermobacter* sp. Methane production *via* hydrogenotrophic methanogenesis is comprised of a methyl branch that fixes and reduces carbon dioxide and a carbonyl branch that initiates biomass production. Furthermore, hydrogenotrophic methanogenesis produces a low ATP-to-methane yield, and thus, these microbes are often nominated to live at the thermodynamic limit of life. A few of these thermophiles have been used to study hydrogenotrophic methanogenesis. However, knowledge of these microbes is lacking in most other metabolic pathways. It would be beneficial to understand their metabolisms better to fully harness these microbes for biotechnological applications.

In the first half of this dissertation, I compare the metabolism of three *Methanothermobacter* spp.: **1)** *Methanothermobacter thermautotrophicus* ΔH; **2)** *Methanothermobacter thermautotrophicus* Z-245; and **3)** *Methanothermobacter marburgensis* Marburg. I employed systems biology, including transcriptomics and

proteomics, and genome-scale metabolic modeling together with continuous bioreactor runs to investigate these microbes. I also cultivated *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$, which is a strain that was recently genetically modified in our group to grow on formate. I made new insights regarding the different mechanisms of anabolic formate production in the three *Methanothermobacter* spp. Further, I studied the effect on the growth of the microbe and energy costs from the formate dehydrogenase cassette, which is required for growth on formate.

In the second half of this dissertation, I investigated the model acetogen *Clostridium ljungdahlii*. Acetogenic bacteria convert carbon monoxide in addition to molecular hydrogen and carbon dioxide into biomass, acetate, and ethanol *via* the Wood-Ljungdahl pathway (WLP). Similar to methanogenesis, the WLP consists of a methyl and a carbonyl branch. While the carbonyl branch is conserved between the two pathways, the methyl branch is not. Furthermore, like methanogens, acetogens live at the thermodynamic limit of life. There is no net ATP yield during the WLP from carbon dioxide and molecular hydrogen. Thus, understanding the energy metabolism of acetogens is particularly important for autotrophic growth, and the role of the energy-conserving RNF-complex is crucial. Despite energetic limitations, the company LanzaTech has already commercialized the use of acetogens for autotrophic ethanol production (*via* the WLP). LanzaTech upgrades waste carbon (carbon monoxide and carbon dioxide) gas streams to produce the ethanol, which can be used as a (bio)fuel. In this dissertation, I performed computational analyses on CRISPR systems that were developed and employed in our lab to investigate the roles of genes and redirect carbon flux in the central carbon metabolism of *C. ljungdahlii*. I implemented *in-silico* tools for estimating the potential of a novel CRISPR-Cas9-based base-editing system, and the prominence and importance within acetogenic bacteria of the *rseC*, that is a potential transcriptional regulator for the RNF-complex.

Kurzfassung

Der wachsende Bedarf an Energieerzeugung bei gleichzeitigem Anstieg der Treibhausgaskonzentration in der Atmosphäre erfordert neue Lösungen für den Energie- und Abfallwirtschaftssektor. Biotechnologische Ansätze, wie der Einsatz lebender Mikroben oder Enzyme, könnte hierfür als Lösung dienen. Mikroben sind attraktive Biokatalysatoren sowohl für die Speicherung von erneuerbarer elektrischer Energie als auch für die Umwandlung von Abfallströmen in höherwertige Chemikalien. Die *Power-to-Gas*-Plattform wandelt überschüssige erneuerbare elektrische Energie durch Elektrolyse von Wasser, in molekularen Wasserstoff und Sauerstoff um. Die Energie kann dann als Wasserstoffgas gespeichert oder in andere Produkte, wie z. B. Methan, umgewandelt werden. Methan, der Hauptbestandteil von Erdgas, kann verbrannt und in der Erdgasinfrastruktur verwendet werden. Hydrogenotrophe Methanogene sind gasfermentierende Mikroben, die als Biokatalysatoren eingesetzt werden können, um dieses Wasserstoffgas über die Methanogenese in Methan (und Biomasse) umzuwandeln. Kohlendioxid aus der Atmosphäre oder aus Verbrennungsprozessen mit nicht fossilen Brennstoffen sollte idealerweise als Kohlenstoffquelle für diese Mikroben genutzt werden. Die Firma Electrochaea GmbH hat dieses Verfahren unter Verwendung einer thermophilen *Methanothermobacter* sp. kommerzialisiert. Die Produktion von Methan durch hydrogenotrophe Methanogenese besteht aus einem Methylzweig, der Kohlendioxid bindet und reduziert, und einem Carbonylzweig, der die Biomasseproduktion einleitet. Darüber hinaus erzeugt die hydrogenotrophe Methanogenese eine geringe ATP-zu-Methan-Ausbeute, so dass diese Mikroben an der thermodynamischen Grenze des Lebens leben. Einige dieser thermophilen Mikroben wurden zur Untersuchung der hydrogenotrophen Methanogenese eingesetzt. Zu den meisten anderen Stoffwechselwegen in diesen Mikroben fehlt es jedoch an Wissen. Es wäre von Vorteil, ihre Stoffwechselvorgänge besser zu verstehen, um diese Mikroben für biotechnologische Anwendungen voll nutzbar zu machen.

In der ersten Hälfte dieser Dissertation habe ich den Stoffwechsel von drei

Methanothermobacter spp. verglichen: **1)** *Methanothermobacter thermautotrophicus* ΔH; **2)** *Methanothermobacter thermautotrophicus* Z-245; und **3)** *Methanothermobacter marburgensis* Marburg. Zur Untersuchung dieser Mikroben habe ich Systembiologie eingesetzt, die Transkriptomics und Proteomics, und Stoffwechselmodellierung auf Genomebene in Verbindung mit kontinuierlichen Bioreaktorstudien umfasst. Ich kultivierte auch *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$, einen Stamm, der kürzlich in unserer Gruppe genetisch verändert wurde, um auf Formiat zu wachsen. Ich habe neue Erkenntnisse über die unterschiedlichen Mechanismen der anabolen Formiatproduktion in den drei *Methanothermobacter* spp. gewonnen. Außerdem untersuchte ich die Auswirkungen auf das Wachstum der Mikrobe und die Energiekosten der Formiat-Dehydrogenase-Kassette, die für das Wachstum auf Formiat erforderlich ist.

In der zweiten Hälfte dieser Dissertation untersuchte ich das acetogene Bakterium *Clostridium ljungdahlii*. Acetogene Bakterien wandeln über den Wood-Ljungdahl-Weg (WLP) neben molekularem Wasserstoff und Kohlendioxid auch Kohlenmonoxid in Biomasse, Acetat und Ethanol um. Wie die Methanogenese besteht auch der WLP aus einem Methyl- und einem Carbonylzweig. Während der Carbonylzweig zwischen den beiden Stoffwechselwegen konserviert ist, ist der Methylzweig nicht konserviert. Außerdem leben Acetogene wie Methanogene an der thermodynamischen Grenze des Lebens. Zudem gibt es keine Netto-ATP-Ausbeute vom WLP aus Kohlendioxid und Wasserstoff. Daher ist das Verständnis des Energiestoffwechsels von Acetogenen besonders wichtig für das autotrophe Wachstum. Außerdem ist die Rolle des energieerhaltenden RNF-Komplexes entscheidend. Trotz energetischer Einschränkungen hat das Unternehmen LanzaTech die Nutzung von Acetogenen für die autotrophe Ethanolproduktion (über den WLP) bereits kommerzialisiert. LanzaTech veredelt Abfallkohlenstoff (Kohlenmonoxid und Kohlendioxid), um Ethanol zu produzieren, das als (Bio-)Kraftstoff verwendet werden kann. In dieser Dissertation führte ich computergestützte Analysen an CRISPR-Systemen durch, die in unserem Labor entwickelt und eingesetzt wurden, um die Rolle von Genen zu untersuchen und den Kohlenstoffffluss im zentralen Kohlenstoffstoffwechsel von *C. ljungdahlii* umzulenken. Ich habe *In-silico*-Tools implementiert, um das Potenzial eines neuartigen CRISPR-Cas9-basierten *Base-Editing*-Systems und die Bedeutung des *rseC*, eines potenziellen Transkriptionsregulators für den RNF-Komplex, in acetogenen Bakterien abzuschätzen.

Acknowledgments

To begin, I would like to acknowledge my two advisors, Prof. Dr. Lars Angenent and Dr. Bastian Molitor. Lars not only accepted me into his lab but promoted my pursuit of the PhD in Germany directly after my Bachelor's degree in the USA. Both Lars and Bastian gave me the opportunity to switch my focus to a relatively different field than what my background was. They permitted and facilitated my attending courses, workshops, and research stays to gain the knowledge and collaborations that I would need to complete this project. I am thankful for the generous financial support that I received to pursue such a large and expensive project with state-of-the-art methods. Both took time to meet with me to discuss and advise on my project, and to edit and proofread my reports, manuscripts, conference abstracts, posters, presentation slides, and of course this dissertation. I am especially grateful for the support they gave me in writing the two DAAD grants that I was awarded and discussions about my future plans as an engineer and scientist.

Second, I would like to thank the rest of my committee, Prof. Dr. Andreas Dräger and Prof. Dr. Boris Maček. Prof. Dräger advised me in and out of his classroom and providing me ideas and support regarding the metabolic modeling. He was a supportive collaborator for work, which laid outside of my PhD, and a strong advocate for me attending conferences relevant to my field. I would like to thank Prof. Maček for early-on discussions regarding proteomic techniques that would be used in this work and reviewing this written dissertation. I would like to thank Prof. Angenent, Prof. Dräger, Prof. Maček, and Dr. Molitor for their participation in my oral defense.

My project was made possible by the collaborations that literally spanned the globe. The opportunity to collaborate with the Marcellin Group at the Australian Institute for Bioengineering and Nanotechnology (AIBN) at the University of Queensland in Brisbane Australia was formative. I am thankful for the opportunity that Prof. Dr. Esteban Marcellin gave me to join his group in 2019. I am immensely thankful for Dr. Timothy McCubbin's support with modeling, analysis of fermentation and proteomics

data, and integration of these data into the model. He was supportive during my stay in the Marcellin group but also after my return to Germany, flexibly meeting me virtually at odd hours to account for time zone differences. I am grateful to Dr. Ricardo Axayacatl (Axa) Gonzalez-García who first supervised me at AIBN and taught me the basics of bioreactor work and omics methods. I am thankful to James Heffernan who introduced me to gas fermentation and omics methods with anaerobes and provided endless advice about bioreactors, but who also made the extra effort as a friend to show me around Brisbane and the surrounding areas.

In addition, I am appreciative of the collaborations with the Ley Lab at the Max Planck Institute (MPI) for Biology Tübingen. Prof. Dr. Ruth Ley generously hosted and supported my bioreactor work in her lab. I am thankful to Dr. Nicholas Youngblut, who initiated the setup of the bioreactor system and who always responded promptly to any need. I am very grateful to the additional staff that facilitated my work at the MPI including: Karin Klein and Ursula Schach for their help with administrative work in and out of lab; Sophie Maisch for her help ordering endless parts for the bioreactors; and Athina Iliopoulou and Katharina Vernali for their flexibility sharing and their help operating the large autoclave, even at odd hours. I am infinitely grateful for my friendship with Dr. Sofia Esquivel-Elizondo. She guided me professionally and personally with an abundance of patience and empathy in and out of the lab over the last few years. I am thankful for the opportunity to have learned from her and be able to share some of my knowledge for her project.

Moreover, I would like to thank all the lab members of the Environmental Biotechnology group. I especially am thankful to Dr. Akanksha Mishra, Dr. Christian Klask, and Dr. Christian Fink who were the only PhD students when I joined and made me feel welcome in the lab, the institute, Tübingen, and Germany. I am grateful not only for their help introducing me to the wet-lab world of biology and reactors but for the countless hours chatting, eating, and adventuring around Tübingen. I am thankful for my other fellow PhD colleagues whom I have the privilege of calling also friends, especially: Sarah Schulz, Patrick Schweizer, Nils Rohbohm, Caroline Schlaiß, Angelia (Han) Wang, Ginés Martínez Cano, Andrés Ortiz Ardila, Laura Rovira-Alsina, and Lucas Mühling. I am thankful for the PostDocs, who provided guidance scientifically in lab and personally during beer hours. I am grateful for Nicolai Kreitli for his help with the setup and clean-up of the bioreactors systems. I appreciate Dr. Pengfei Xia and Dr. Christian Klask who both approached me with the ideas and needs for bioinformatical

analyses, from which additional collaborations were born. I am also grateful to Daria Evseeva who enthusiastically and promptly answered questions and performed bioinformatic analyses.

Further, I would also like to thank colleagues from AIBN including Dr. Michael Macdonald, Rosemary Gillane, Dr. Vishnuvardhan Mahamkali, and Anna Cameron. I would like to thank Dr. Albane Ruaud, Jaime Leonardo Moreno, Nasim Farahani Zayas, Andrea Borbon, Liam Fitzstevens, Michael Bell, Yihua Liu, Marco Podobnik, Dr. Stacey Heaver, Dr. Daphne Welter, Dr. Sergio LaTorre, and Bilal Hammoud from the MPI for their friendship. Additionally, I am thankful to Irinia Droste-Borel from the Proteome Center Tübingen, who was always willing to answer my questions to the best of her abilities and coordinate my proteomics work in a timely fashion. Further, I am grateful to Silvia Morini for a friendship that blossomed from classwork and collaborations with Prof. Dr. Andreas Dräger. To Jorge Morato García and Albane (again) for both the fun times as flatmates and scientific discussions over drinks. I am grateful to Alex Warsewa who has encouraged and supported me the last months of the dissertation. Further, I am grateful to Clarissa Wehrli and Anton Zollbrecht for the friendship that began in Australia but continued back to Europe.

Additionally, I would like to thank my family, who facilitated and gave me the privileges I have today to pursue my interests. I thank my parents, Shannon Berg and Roberto Casini, for supporting me throughout my life and my academic career, and my siblings, Giorgio, Sofia, and Emma who remind me of the energy and influence of younger generations. I thank the rest of my family in Italy and the United States for their words of encouragement, and their hospitality when visiting. I am grateful to my previous supervisor, Prof. Dr. Denise McKahn who encouraged me in my scientific journey, and to my friends from university and before, including, but not limited to: Kate Shambaugh, Cassiopeia Lee, Preeti Reddy Dasari, Lissette Ramirez, and Haley Manchester who have been supportive regardless of the time zone they are in.

Finally, I want also acknowledge the generous funding from the Alexander von Humboldt Foundation in the framework of the Alexander von Humboldt Professorship, which was awarded to Prof. Angenent. I further would like to thank the Deutscher Akademischer Austauschdienst (German Academic Exchange Service) for their Kurzstipendium für Doktoranden, which enabled my collaborative trip to Australia.

Thank you all for the support during this experience!

Contents

List of Figures

List of Tables

List of Publications

a) **Published journal articles**

Klask, C.M., Jäger, B., Casini, I., Angenent, L.T. and Molitor, B., 2022. Genetic evidence reveals the indispensable role of the *rseC* gene for autotrophy and the importance of a functional electron balance for nitrate reduction in *Clostridium ljungdahlii*. *Frontiers in Microbiology*, **13**:887578.

Heffernan, J.K., Valgepea, K., de Souza Pinto Lemgruber, R., Casini, I., Plan, M., Tappel, R., Simpson, S.D., Köpke, M., Nielsen, L.K., Marcellin, E. 2020. Enhancing CO_2-valorization using *Clostridium autoethanogenum* for sustainable fuel and chemicals production. *Frontiers in Bioengineering and Biotechnology*, **8**,204.

Xia, P.-F., Casini, I., Schulz, S., Klask, C.-M., Angenent, L.T., Molitor, B. 2020. Reprogramming acetogenic bacteria with CRISPR-targeted base editing *via* deamination. *ACS Synthetic Biology*, **9**(8), 2162–2171.

b) **Preliminary manuscripts**

Casini, I., McCubbin, T., Esquivel-Elizondo, S., Luque, G., Evseeva, D., Fink, C., Beblawy, S., Youngblut, N.D., Huson, D.H., Dräger, A., Ley, R.E., Marcellin, E., Angenent, L.T., Molitor, B. Genome-scale metabolic modeling of *Methanothermobacter* spp. provides an integrated systems biology platform for power-to-gas technology. *In preparation*.

Casini, I., McCubbin, T., Luque, G., Ley, R.E., Marcellin, E., Angenent, L.T., Molitor, B. Episomal formate dehydrogenase gene expression in genetically engineered *Methanothermobacter thermautotrophicus* ΔH enables formate consumption in continuous fermentation at reduced non-growth-associated maintenance energies. *In preparation*.

Esquivel-Elizondo, S., Casini, I., Luque, G., Bağci, C., Youngblut, N.D., Angenent, L.T., Ley, R.E.. Characterization of the syntrophy between *Christensenella minuta* and *Methanobrevibacter smithii* under continuous growth conditions. *In preparation*.

Angenent, L.T., Casini, I., Schröder, U., Harnisch, F., Molitor, B. Electric-power storage into chemical energy by combining or integrating electrochemistry and biology. *In preparation*.

Symbols

Molecular species are found in the **Abbreviations** section below.

α	threshold value used to determine statistical significance
$biomass_{aq\_out}$	moles of aqueous biomass exiting a system
$C_{acetate}$	concentration of acetate at a specified time point
$C_{ethanol}$	concentration of ethanol at a specified time point
$C_{fructose}$	concentration of fructose at a specified time point
C_{in}	moles of carbon entering a system
C_{out}	moles of carbon exiting a system
CH_{4,gas_out}	moles of gaseous methane exiting a system
CO_{2,aq_out}	moles of aqueous carbon dioxide exiting a system
CO_{2,gas_in}	moles of gaseous carbon dioxide entering a system
CO_{2,gas_out}	moles of gaseous carbon dioxide exiting a system
D	number of distal nucleotides
$\frac{dx}{dt}$	change in mass of a metabolite (numerator) in time (denominator)
ΔG	Gibbs free energy
E	protein concentration
E_{in}	moles of electrons (*via* degree of reduction) entering a system
E_{out}	moles of electrons (*via* degree of reduction) exiting a system
k_{cat}	enzyme turnover rate
m	number of matched nucleotides in distal sites
$\dot{n}$	molar flow rate
n	generations (microbial growth)
N	number of nucleotides that should be identical to the target sequence

$Naformate_{in}$	moles of sodium formate entering a system
OD_0	initial optical density (time = 0)
OD_t	optical density at a transfer
P	pressure
$\dot{V}$	volumetric flow rate
R	molar gas constant
T	temperature
ρ	connectivity
μ	growth rate
π	participation number
S	stoichiometric matrix
υ	flux vector
$\cup$	union
$\cap$	intersection

Abbreviations

2D	two-dimensional
A	adenine
acetyl-CoA	acetyl coenzyme A
ACS	acetyl-CoA synthetase
AD	anaerobic digester
ADP	adenosine diphosphate
AID	activation -induced cytosine deaminase
AMP	adenosine monophosphate
ANOVA	analysis of variance
Arg	arginine
Asn	asparagine
ATP	adenosine triphosphate
ATPM	ATP hydrolysis reaction (NGAM)
BES	bioelectrochemical system
BiGG	biochemical, genomic, and genetic
bp	base pair(s)
BLASTp	protein basic local alignment search tool
C	carbon
C	cytosine
C1	one-carbon compounds
cDNA	complementary deoxyribonucleic acid
CDS	coding DNA sequence
CH_4	methane
CO	carbon monoxide

CO_2	carbon dioxide
CoA	coenzyme A
CoB	coenzyme B
COBRA	constraint-based reconstruction and analysis
COBRApy	constraint-based reconstruction and analysis in python
COG	clusters of orthologous gene
CoM	coenzyme M
CoM-S-S-CoB	CoM-CoB heterodisulfide
CRISPR	clustered regularly interspaced short palindromic repeats
CTAB	cetyltrimethylammounium bromide
DDA	data dependent acquisition
DE	differential expression
DH	*Methanothermobacter thermautotrophicus* ΔH
DNA	deoxyribonucleic acid
DR	direct repeats
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen
E-matrix	expression matrix
EMF	elementary flux modes
EX_	prefix for the exchange (pseudo) reactions
F-	formyl group
F_{420}	coenzyme F_{420} or 8-hydroxy-5-deazaflavin (oxidized)
$F_{420}H_2$	coenzyme F_{420} or 8-hydroxy-5-deazaflavin (reduced)
FASTA	fast-all
FBA	flux balance analysis
fdh/FDH	formate dehydrogenase
Fd_{ox}	ferredoxin (oxidized)
Fd_{red}	ferredoxin (reduced)
FIA	flow injection analysis
Fmd	molybdenum-dependent formylmethanofuran dehydrogenase
FMFR	formyl-methanofuran

Frh	F_{420}-reducing Ni-Fe hydrogenase
FVA	flux variability analysis
Fwd	tungsten-dependent formyl-MFR dehydrogenase
G	guanine
GAM	growth-associated maintenance energy
GC	gas chromatograph
gDNA	genomic deoxyribonucleic acid
GECKO	GEM with enzymatic constraints using kinetic and omics
GEM	genome-scale metabolic model
GIMME(p)	gene inactivity moderated by metabolism and expression (by proteome)
Gln	glutamine
Gly	glycine
GmbH	Gesellschaft mit beschränkter Haftung
gRNA	guide RNA
H	hydrogen
H^+	proton
H_2	molecular hydrogen
H_2O	water
H_4MPT	tetrahydromethanopterin
HCOOH	formic acid
Hmd	MeH_4MPT hydrogenase/H_2-forming methylene-H_4MPT dehydrogenase
HPLC	high pressure liquid chromatography
HRT	hydraulic retention time
iBAQ	intensity-based absolute quantification
IEF	isoelectric focusing
Ile	isoleucine
iMAT	integrative metabolic analysis tool
IOMA	integrative omics metabolic analysis

jQMM	JBEI quantitative metabolic modeling library
K	biomass correlation coefficient
K	lysine
kb	kilobases
KEGG	Kyoto encyclopedia of genes and genomes
LB medium	Luria-Bertani medium
LC	liquid chromatogrpahy
log_2FC	log_2 fold change
LPS	lipopolysaccharide
LVA tag	leucine-valine-alanine tag
Lys	lysine
M-	methyl group
Me-	methenyl
MADE	metabolic adjustment by differential expression
MBA	model-building algorithm
MCMC	Markov chain Monte Carlo
Mcr	methyl-coenzyme M reductase isoenzyme I
MeH_4MPT	methenyl-tetrahydromethanopterin
Me-model	metabolic and expression model
MEMOTE	metabolic model test
MER	methane evolution rate
Met	methionine
METRADE	metabolic and transcriptomics adaptation estimator
MFR	methanofuran
MH_4MPT	methyl-tetrahydromethanopterin
MIRIAM	minimum information requested in the annotation of biochemical model
MM	*Methanothermobacter marburgensis* Marburg
mRNA	messenger ribonucleic acid
Mrt	methyl-coenzyme M reductase isoenzyme II

MS	mass spectrometry
MS/MS	tandem MS
MS-media	minimal salts media
Mtd	MyH_4MPT dehydrogenase
Mtr	MH_4MPT coenzyme M methyltransferase
Mvh	F_{420}-non-reducing Ni-Fe hydrogenase
My	methylene
MyH_4MPT	methylene-tetrahydromethanopterin
N_2	molecular nitrogen
Na^+	sodium ion
NaCOOH	sodium formate
NAD^+	nicotinamide adenine dinucleotide (oxidized)
NADH	nicotinamide adenine dinucleotide (reduced)
$NADP^+$	nicotinamide adenine dinucleotide phosphate (oxidized)
NADPH	nicotinamide adenine dinucleotide phosphate (reduced)
NaOH	sodium hydroxide
NCBI	National Center for Biotechnology Information
ND/n.d.	not detected/not detectable
NG	no gene
NGAM	non-growth-associated maintenance energy
NGS	next-generation sequencing
NH_4^+	ammonium
Ni-Fe	nickel iron
NS/n.s.	not significant
NTA	nitrilotriacetic acid
O_2	molecular oxygen
OD_{600}	optical density at 600 nm
ORF	open reading frame
ORP	oxidation-reduction potential
P	proline

PacBio	Pacific Biosciences
PAM	protospacer adjacent motif
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PCT	Proteome Center Tübingen
Pfl	pyruvate formate-lyase
Phe	phenylalanine
Pro	proline
PROM	probabilistic regulation of metabolism
PtG	power-to-gas
PtX	power-to-x
qRT-PCR	quantitative reverse transcription PCR
R	arginine
RAST	rapid annotations using subsystems technology
RCM	reinforced clostridial medium
RM	restriction/modification
RNA	ribonucleic acid
RNF	*Rhodobacter* nitrogen fixation complex
rpm	revolutions per minute
rRNA	ribosomal ribonucleic acid
S-matrix/S	stoichiometric matrix
SBML	systems biology markup language
SBO	systems biology ontology
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SH	sulfhydryl
SIP	stable isotope probing
SMRT	single-molecule, real-time
sp	species
SPOT	simplified pearson correlation with transcriptomic data
spp	species (plural)

SWATH-MS	sequential windowed acquisition of all theoretical fragment ion mass spectra
syngas	synthesis gas
T	thymine
TIGER	toolbox for integrating genome-scale metabolism, expression, and regulation
TPM	transcripts per million
Trp	tryptophan
URI	uniform resource identifier
UGI	uracil glycosylase inhibitor
UV	ultra violet
vs.	versus
WLP	Wood-Ljungdahl pathway
WT	wild type
XIC	extracted ion chromatograms
ZZ	*Methanothermobacter thermautotrophicus* Z-245

Chapter 1

1.1 Motivation and objectives

The anthropogenic contribution to climate change is no longer deniable (Hegerl *et al.*, 2007). While the Earth does experience natural heating and cooling cycles, the massive rise in greenhouse gases after the start of the industrial revolution greatly affects the Earth's atmospheric composition (Hegerl *et al.*, 2007). The consequence of not addressing this phenomenon, which is amplified by an increasing world population, is predicted to cause a multitude of problems. This includes but is not limited to: **1)** more intense weather (stronger storms and extreme temperatures); **2)** rising sea levels; **3)** increased sea acidity; and **4)** loss of biodiversity (Parncutt, 2019). These ramifications are predicted to have a direct impact on the quality of human life and to cause an estimated one billion premature human deaths over the next centuries (Parncutt, 2019). We are not only facing an ecological crisis but a global ethical crisis that can only be addressed when actions are taken worldwide. The Paris Agreement of 2015 was a recent effort in which the United Nations outlined the 17 Sustainable Development Goals that highlighted the necessity for both sustainable and equitable development (United Nations, 2015).

In the United States, carbon dioxide (CO_2) accounts for 80 weight-% of the greenhouse gases and is primarily produced in the transportation, electricity, and industrial sectors (United States Environmental Protection Agency, 2021). The combustion of fossil fuels across these sectors is the primary culprit for these high levels. Biotechnology in which microbes (and sometimes enzymes) are applied as biocatalysts can help: **1)** limit future carbon dioxide generation; and **2)** reduce already above-ground carbon dioxide. Above-ground carbon dioxide refers to the carbon dioxide that is no longer in the form of a fossil fuel, for example, carbon dioxide from

the atmosphere and from combustion processes (Köpke and Simpson, 2020). The demand of fossil fuels can be reduced through renewable power sources, such as wind and solar. However, their intermittency in production over time requires a means by which excess renewable power can be stored for later consumption. The power-to-x platform refers to the conversion of electric power into another product (x). Typically, excess electric power generated from renewable power sources is used. The power-to-x platform addresses the need to diminish carbon dioxide production, because it reduces future carbon dioxide emissions by employing excess renewable electric power to electrolyze water into molecular hydrogen (H_2) and molecular oxygen (O_2). The electric power is stored in molecular hydrogen as chemical bonds and can be harnessed directly or upgraded to another gaseous or liquid product. In the case of gaseous products, this process is called power-to-gas. In most countries, there are limited options for direct molecular hydrogen utilization, particularly because the current fossil fuel infrastructures cannot typically handle more than 10 vol-% molecular hydrogen (Dolci *et al.*, 2019; Rusmanis *et al.*, 2019). One option is to convert the molecular hydrogen together with carbon dioxide into methane (CH_4) for which the existing natural gas infrastructure can be employed immediately. This is favorable for storage and distribution and further addresses the need to recycle existing above-ground carbon dioxide. The upgrading of molecular hydrogen into methane can be achieved (thermo)chemically or biologically. The chemical process, or the Sabatier process, employs metal catalysts for the conversion. Alternatively, the biological process, known as biomethanation, is carried out with biocatalysts. The work presented in this dissertation focuses on biomethanation. Biomethanation harnesses methanogens, microbes that natively consume molecular hydrogen and carbon dioxide and produce methane through the metabolic pathway of methanogenesis. This enables simultaneous upgrading of molecular hydrogen gas and reduction of above-ground carbon dioxide (Martin *et al.*, 2013). The combustion of methane produces carbon dioxide, but the possibility to recycle this carbon dioxide for methane generation *via* biomethanation exists and aligns with a circular carbon economy. The company Electrochaea GmbH is commercializing this process and has already constructed and tested various pilot plants including at grid scale (1.5 MW) (Rusmanis *et al.*, 2019).

As an alternative to methane, acetate and ethanol can be produced with acetogenic bacteria (acetogens). Acetogens can grow autotrophically by fixing C1 gases *via* the ancient, but energetically efficient, Wood-Ljungdahl pathway (WLP), a pathway with

various parallels to methanogenesis (Drake *et al.*, 2008). The WLP produces acetate, which is a common intermediary metabolite that can also be implemented as a feedstock for other microbes (Molitor *et al.*, 2019), and ethanol that can be combusted directly as a fuel source. Acetogens can grow on energy-rich carbon monoxide (CO) in addition to molecular hydrogen and carbon dioxide. While the lower reduction potential of carbon monoxide promotes more reduced products (ethanol), excess carbon monoxide is inhibitory to their hydrogenases (Thauer *et al.*, 1974; Pavan *et al.*, 2022). Recently, Heffernan *et al.* (2020) demonstrated that supplying carbon monoxide in low concentrations (2%) to molecular hydrogen and carbon dioxide, reduced the inhibitory effects and increased ethanol production compared to molecular hydrogen and carbon dioxide alone. The combination of the three gases is known as synthesis gas (syngas), and certain industrial waste gas streams resemble various syngas compositions. The company LanzaTech has already commercialized the reuse of industrial waste gas streams to produce ethanol and is expanding to include jet fuel and other high-value chemicals.

My work presented in this dissertation focuses on integrating metabolic engineering and other systems biology techniques to study microbes, specifically methanogens, for biotechnological solutions related to climate change. Understanding the metabolism of a microbe is the first step to identifying a potential biocatalyst, and is favorable before modifying or optimizing the metabolism. The first part of my dissertation focuses on employing methanogens, *Methanothermobacter* spp. for the power-to-gas platform. I built and ran genome-scale metabolic models (GEMs) for three closely related methanogens. I cultured the microbes in continuous bioreactors and sampled for fermentation, transcriptomics, and proteomics data. I analyzed and integrated these data into the GEMs to elucidate differences between the microbes. Furthermore, I tested a genetically modified plasmid-carrying *Methanothermobacter* sp. (Fink *et al.*, 2021) and compared the performance of this strain under different growth conditions to the wild-type strains. I found that the plasmid was stable and added the intended metabolic function (the ability to consume formate as a sole substrate), while not hindering the performance of the microbe compared to the wild type. I modeled the modifications and observed a decrease in the non-growth-associated maintenance (NGAM) energy, which is linked to higher biomass production rates. The ability to genetically modify the methanogens and to model the outcome of those modifications to the metabolism provides the potential to widen the spectrum of substrates and to generate high-value

products.

In the second half of my dissertation (Chapters 5 and 6), I supported the work of my colleagues who aimed to achieve similar goals with acetogens, in particular with the model acetogen *Clostridium ljungdahlii*. As the focus of my dissertation is on methanogens, I have not included an extensive literature review in Chapter 2 on acetogens. The Chapters 5 and 6, with the work on acetogens, provide their own introduction with relevant literature. In this paragraph, I first list peer-reviewed review papers that the reader may find useful for a general understanding of acetogens for biotechnological applications: **1)** acetogenic metabolism for C1 utilization (Katsyv and Müller, 2020); **2)** industrial applications with acetogens (Pavan *et al.*, 2022); **3)** genetic tools for acetogens (Bourgade *et al.*, 2021); and **4)** analytical tools for acetogens (Heffernan *et al.*, 2022). The genetic tools available for *Clostridium* are more developed than for *Methanothermobacter*. In particular, the clustered regularly interspaced short palindromic repeats (CRISPR)/Cas systems for *Clostridium* have allowed more extensive and diverse genetic manipulation. The collaborative work presented in my dissertation aimed to investigate the acetogenic metabolism with these methods. I provided support through bioinformatical analyses *via* computational tools that I designed and built. The first tool (fundamentally a string-searching algorithm) included a genome-scale algorithm that evaluated the performance and identified potential editing targets for the new base-editing CRISPR-dCas9 system. My tool highlighted the applicability of the new CRISPR system for modifying microbes. The second tool employed the preexisting protein Basic Local Alignment Search Tool (BLASTp) (Camacho *et al.*, 2009; Altschul *et al.*, 1990, 1997) to identify homologs for the *Rhodobacter* nitrogen fixation (RNF) complex and its predicted transcriptional regulator (*rseC*) from *C. ljungdahlii* in other acetogens. My tool semi-automated this investigation, which began with a genome sequence availability search and download, and finally the homolog inquiry in BLASTp. The repeated presence of genes across phenotypically related microbes may help explain the function and importance of those genes. More details on the contents of each chapter can be found in the subsection below.

1.2 Organization and summary of chapters

Chapter 2 presents a literature review that covers the following topics: **1)** the power-to-gas platform; **2)** methanogenic bacteria and methanogenesis; **3)** metabolic modeling methods; and **4)** additional system biology methods.

Chapter 3 contains five major contributions: **1)** the sequencing and resequencing of three methanogenic archaea; **2)** the reconstruction of genome-scale metabolic models (GEMs) for the three microbes; **3)** the cultivation on molecular hydrogen and carbon dioxide of the methanogens in continuous bioreactors; **4)** the multi-omics analyses from samples, which were taken during the aforesaid bioreactor runs; and **5)** the integration of the wet-lab experimental data into the GEMs. In Chapter 3, three methanogens, *Methanothermobacter thermautotrophicus* ΔH, *Methanothermobacter thermautotrophicus* Z-245, and *Methanothermobacter marburgensis* Marburg, are compared using the data previously described at pan-genome, pan-transcriptome, and pan-proteome levels. The comparison demonstrated that *M. thermautotrophicus* ΔH was able to consume more molecular hydrogen and carbon dioxide and produce more methane per unit biomass than the other microbes. Conversely, *M. marburgensis* Marburg had the highest biomass production rate of the three. Nevertheless, I found significant differences in the transcriptome and proteome of the three microbes. There were fewer differentially expressed genes and proteins between *M. thermautotrophicus* ΔH and *M. thermautotrophicus* Z-245 than to *M. marburgensis* Marburg. This supports the phylogenic relationship between the microbes. The metabolic models in combination with the proteomics data suggested that the three microbes used three different anabolic formate production mechanisms. *M. thermautotrophicus* ΔH appears to prefer the pyruvate formate-lyase (*pfl*), *M. thermautotrophicus* Z-245 the formate dehydrogenase cassette (*fdhCAB*), and *M. marburgensis* Marburg the cassette-less formate dehydrogenase (*fdhAB*). Large steady-state replicated datasets are not common. Thus, these results also can serve as a rich data source for future research.

Chapter 4 builds on the work in Chapter 3. I applied the same systems biology methods, but compared the growth behavior of a new strain, *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$, which contains a plasmid that carried the *fdhCAB* (Fink *et al.*, 2021), to *M. thermautotrophicus* ΔH wild type and *M. thermautotrophicus* Z-245 wild type. *M. thermautotrophicus* ΔH wild type is unable to utilize only formate as an energy and carbon source (unlike *M. thermautotrophicus* Z-245), due to the missing

fdhCAB genes (Nölling and Reeve, 1997). This chapter provides the results of the comparison between *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$, which was grown on molecular hydrogen and carbon dioxide or on sodium formate (Na-formate), and between *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ and *M. thermautotrophicus* Z-245, which were both grown on sodium formate. The results indicated that *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ performed comparably to *M. thermautotrophicus* Z-245 during growth on sodium formate, but with significantly different molecular hydrogen production rates. Furthermore, I compared *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ with *M. thermautotrophicus* Z-245 and *M. marburgensis* Marburg during growth on molecular hydrogen and carbon dioxide. For this comparison, I included the growth data from Chapter 3. The performance of the three strains was comparable. For growth on molecular hydrogen and carbon dioxide, the non-growth-associated maintenance energy of *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ decreased, which was consistent with the higher biomass to methane carbon yields observed. These non-growth-associated maintenance values were similar to those of *M. thermautotrophicus* Z-245 and *M. marburgensis* Marburg in Chapter 3. For growth on sodium formate, the non-growth-associated maintenance of *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ and *M. thermautotrophicus* Z-245 were also similar. The omics analysis revealed the impact of the constitutive promoter, which drives the expression of the *fdhCAB* genes on the plasmid, and which leads to significantly higher abundances of transcripts. However, the protein abundances were not different, indicating potential post-transcriptional regulation. Further, the advantages and disadvantages of using sodium formate or formic acid as a substrate are discussed. This chapter describes the first bioreactor work with a genetically modified *M. thermautotrophicus* strain. This is an essential step forward in the evaluation of genetically modified methanogens for power-to-x applications, and one of the few quantitative continuous bioreactor studies with methanogens grown on formate.

Chapter 5 contains the published manuscript that describes the novel CRISPR/Cas9 base-editing tool that was harnessed for the genetic engineering of *C. ljungdahlii* (Xia *et al.*, 2020). Dr. Peng-Fei Xia developed this base-editing tool that employs homology-directed repair of single nucleotides on both strands of DNA *via* a novel CRISPR/Cas9. This markerless techniques enables a cytosine-to-thymine substitution without leaving any traces, such as selective markers. The base-editing tool was

implemented to generate nonsense mutations *via* the creation of premature STOP codons for genes in the central carbon metabolism of *C. ljungdahlii*. The bifunctional aldehyde-alcohol dehydrogenase isozymes (*adhE1* and *adhE2*) and the aldehyde:ferredoxin oxidoreductase isozymes (*aor1* and *aor2*), were targeted in attempts to simultaneously increase acetate production and decrease ethanol production. The goal of the bioinformatics was to assess the potential of the base-editing tool, and to find the frequency of possible cytosine-to-thymine substitutions in the adenine-thymine-rich (A-T-rich) genome of *C. ljungdahlii*. I designed a genome-scale algorithm that searched the genome for the protospacer NGG and then identified cytosines that are located in the editable window. The algorithm then returned the resulting mutation type and whether the mutation would occur in a coding DNA sequence (CDS). I found that 99.83% of CDSs could be edited with this tool of which 81.36% could result in nonsense mutations. The tool can easily be adapted by changing the genome sequence, protospacer sequence, and the editing window. My tool highlighted the potential ability of the base-editing tool for genome engineering even in A-T-rich microbes.

Chapter 6 is the published manuscript for which Dr. Christian-Marco Klask investigated the membrane-bound transhydrogenase RNF complex, the putative RNF regulator (*rseC*), and the gene cluster that encodes for a putative nitrate reductase of *C. ljungdahlii*. These genes are critical to the energy metabolism in *C. ljungdahlii* and particularly its ability to grow autotrophically (Emerson *et al.*, 2019; Tremblay *et al.*, 2012). Dr. Klask generated various deletion strains utilizing the CRISPR/Cas12a system that he developed. He cultivated the resulting strains in batch experiments under heterotrophic and autotrophic conditions, while comparing ammonium and nitrate as nitrogen sources. The wet lab experiments supported the hypothesis that under autotrophic conditions the *rseC* is essential and it acts as a positive transcriptional regulator for the RNF-cluster encoding genes. I contributed to this work by conducting a bioinformatical analysis of 47 different acetogens to investigate the prominence and role of potential *rseC* genes. I wrote scripts that searched in the National Center for Biotechnology Information (NCBI) database for the complete genome sequences (in GenBank format) of acetogenic bacteria that had been included in a previously published article (Bengelsdorf *et al.*, 2018). My scripts then generated the protein FASTA files and compared them to *rseC* and RNF cluster protein sequences in *C. ljungdahlii* DSM 13528 (CLJU_c11350-CLJU_c11410) using BLASTp (Camacho

et al., 2009; Altschul *et al.*, 1990, 1997). Four genes from the Ech-gene cluster from *Thermoanaerobacter kivui* (TKV_c19720, TKV_c19710, TKV_c19690, TKV_c19740) were also applied as queries to genomes. The program analyzed, whether: **1)** a potential *rseC* gene was found; **2)** there was an RNF cluster found (based on the presence and proximity of the *rnfC* and *rnfD* subunits); **3)** the *rseC* gene was flanking the *rnfC* gene; and **4)** the presence of four Ech-gene cluster subunits could be predicted (*ech2A1*, *ech2A2*, *ech2B*, *echE2*). My analysis of acetogenic bacteria highlighted that there is, though, not invariably: **1)** a widespread presence of *rseC* genes; **2)** a typical vicinity of the *rseC* genes to the RNF complexes; and **3)** an infrequent presence of *rseC* when there is only an Ech-complex but no RNF-complex. These finds supporting the important role of the *rseC* in acetogenic metabolism.

Chapter 7 summarizes the major conclusions of the previous chapters and highlights how metabolic engineering and other systems biology techniques can be exploited to better understand and alter the metabolism of microbes in biotechnological applications. A section is also included with suggested future work.

Chapter 2

2.1 The power-to-gas platform

The power-to-gas platform is the principle of converting excess renewable energy (*e.g.*, wind, solar) to a gaseous energy source that can be utilized at a later time (Bailera *et al.*, 2017; Götz *et al.*, 2016). It is well known that the typical peak production of renewable energy is out of phase with the typical peak demand times, which requires a way to store the energy such that it can be accessed once again. There are four primary ways of storing energy: **1)** electrically (*e.g.*, capacitors); **2)** mechanically (*e.g.*, pumped hydroelectric); **3)** (electro)chemically (*e.g.*, conventional batteries); and **4)** thermally (*e.g.*, phase change materials) (Chen *et al.*, 2009). While traditional storage mechanisms serve their purpose, power-to-gas is superior in regards to smaller discharge losses over time and the potential for larger storage capacities (Sterner and Specht, 2021). As a method of chemical storage, power-to-gas stores the electrical energy in a gaseous chemical form, such as H_2 and CH_4, which facilitates the long-term storage and distribution of that energy. The generation of H_2 and CH_4 are sometimes referred to as power-to-hydrogen and power-to-methane, respectively (Bailera *et al.*, 2017; Götz *et al.*, 2016). The CH_4 can be generated by combining above-ground CO_2 with H_2, which is produced from the electrolysis of water (using excess renewable electric power). While the H_2 infrastructure is lacking, CH_4 is already harnessed as an energy source (it is the main component in natural gas), which means that the infrastructure required for the storage, distribution, and consumption is already in place in many communities. Therefore, CH_4 that is produced by power-to-gas makes it a cleaner fuel than fossil fuels (Guerra *et al.*, 2018; Schiebahn *et al.*, 2015).

Choosing a source of CO_2 that aligns with a circular carbon economy improves the sustainability of power-to-gas. Some of these potential CO_2 sources are: **1)** biogas from

anaerobic digestion, which contains up to ∼40% carbon dioxide (Martin *et al.*, 2013); **2)** exhaust gas from non-fossil power plants, such as renewable natural gas and biomass; **3)** off-gases from other industrial processes that are difficult to be decarbonized entirely, such as cement and steel production (Angenent *et al.*, 2017; Götz *et al.*, 2016; Schiebahn *et al.*, 2015); and **4)** direct air carbon capture (Keith *et al.*, 2018; Yuan *et al.*, 2016). CH_4 can also be produced *via* thermo-chemical processes, such as the hydrocarbon forming Sabatier process. However, these typically require high temperature ($>$ 200 °C) and pressures ($>$ 1 MPa), as well as a metal catalyst (iron, nickel, cobalt, ruthenium, *etc*.) that is often sensitive to gas impurities (Dry, 2002; Leonzio, 2016; Van Der Laan and Beenackers, 1999). These sensitivities limit the ability to directly convert the sustainable CO_2 sources listed above. Alternatively, H_2 and (impure) CO_2 can be fed to microbes, such as hydrogenotrophic methanogens, which naturally consume these compounds as substrates to produce CH_4 in a process termed "biomethanation" (**Equation (2.1)**) (Guerra *et al.*, 2018; Leonzio, 2016; Rachbauer *et al.*, 2017). This process is already commercialized by the company Electrochaea GmbH.

$$4\,H_2 + CO_2 \rightarrow CH_4 + 2\,H_2O \tag{2.1}$$

While power-to-gas specifies gaseous final products, power-to-x is a broader process that encompasses the conversion of excess renewable electric power to gaseous or liquid products. The process of generating liquid products is also sometimes nominated power-to-fuel (Sterner and Specht, 2021). As for CH_4 generation, in liquid fuel production, CO_2 and H_2 can be fed to biocatalysts to generate valorized liquid products. For example, acetogenic bacteria can produce acetate and ethanol, and the latter of which can be utilized as a biofuel (Köpke and Simpson, 2020).

2.2 *Methanothermobacter* spp.

This dissertation focuses on species in the *Methanothermobacter* genus, specifically *Methanothermobacter thermautotrophicus* ΔH (previously classified *Methanobacterium thermoautotrophicum* strain ΔH), *Methanothermobacter thermautotrophicus* Z-245 (formerly known as *Methanobacterium thermoformicicum*), and *Methanothermobacter marburgensis* (previously classified as *Methanobacterium thermoautotrophicum* strain

Marburg) (Wasserfallen *et al.*, 2000). Their detailed taxonomy is as follows: **Domain**, Archaea; **Kingdom and Phylum**, Euryarchaeota; **Superclass**, Methanomada group; **Class**, Methanobacteria; **Order**, Methanobacteriales; **Family**, Methanobacteriaceae; **Genus**, *Methanothermobacter*; **Species**, *Methanothermobacter thermautotrophicus* or *Methanothermobacter maburgensis*; **Strain**, ΔH and Z-245 for *Methanothermobacter thermautotrophicus*, and Marburg for *Methanothermobacter marburgensis* (Adam *et al.*, 2017; Petitjean *et al.*, 2015).

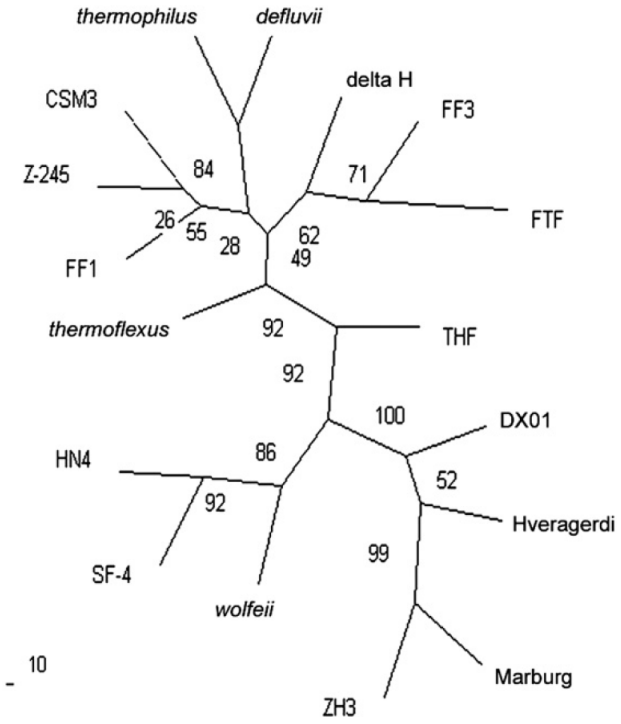

Figure 2.1: Phylogenetic tree of *Methanothermobacter* spp. based on 16S rRNA (Ding *et al.*, 2010).

For *M. thermautotrophicus* alone, 18 strains were already identified, 12 of which can be ordered from DSMZ (Ding *et al.*, 2010). The genetic variation that is found between strains, is depicted in the phylogenetic tree based on 16S rRNA (**Figure 2.1**). However, smaller distances between strains are not always consistent with phenotypical similarities. Despite clear phenotypic differences (**Section 2.2.2**), *M. thermautotrophicus* ΔH and Z-245 are more closely related to each other than to *M. Marburgensis* Marburg (**Figure 2.1**, (Ding *et al.*, 2010)).

Both *M. thermautotrophicus* ΔH and *M. marburgensis* Marburg were isolated from

anaerobic digester sludge and their genomes fully sequenced and assembled (**Table 2.1**). For the last 40 years, both species have served as model microbes for hydrogenotrophic methanogenesis. This has resulted in an abundance of literature especially on the hydrogenotrophic methanogenesis pathway. Kaster *et al.* (2011b) already extensively compared the two microbes, with a focus on the methanogenesis pathway. However, this attention was missing for *M. thermautotrophicus* Z-245 and other metabolic pathways in *Methanothermobacter* spp., resulting in still more than 500 hypothetical proteins in the genomes of these microbes (PATRIC, 2014). Both *M. thermautotrophicus* Z-245 and *M. marburgensis* Marburg have plasmids, with lengths of 11 kb and 4.4 kb, respectively (Liesegang *et al.*, 2010; Nölling *et al.*, 1992) (**Table 2.1**). The plasmid of *M. thermautotrophicus* Z-245 (pFZ1) contains elements of a restriction modification (RM) system, including a methyltranferase, which is likely involved in the methylation of DNA (Boone, 2015; Nölling *et al.*, 1992). The cryptic plasmid of *M. marburgensis* Marburg (pME2001) has also been sequenced (Bokranz *et al.*, 1990) and was recently employed as the backbone of a shuttle vector system for genetic modifications with *M. thermautotrophicus* ΔH (Fink *et al.*, 2021).

2.2.1 Growth conditions

The three microbes can all grow on H_2 and CO_2 *via* the Wolfe Cycle (**Section 2.3**) (Thauer, 2012) under similar optimal conditions, without the supplementation of any organic compounds (Lyu *et al.*, 2018) (**Table 2.1**). *M. thermautotrophicus* ΔH and *M. marburgensis* Marburg can both grow limitedly on CO, which is toxic to most hydrogenases (Diender et al., 2015), though the latter microbe doing so better than the former (Diender *et al.*, 2016). Further *M. thermautotrophicus* Z-245 can utilize formate as its sole energy and carbon source, likely due to the complete F_{420}-dependent formate dehydrogenase cassette (**Section 2.3**) (Nölling and Reeve, 1997). Additionally, *M. marburgensis* Marburg is capable of uptaking and incorporating into biomass a wide variety of compounds in measurable quantities: **1)** formate (Tanner *et al.*, 1989); **2)** acetate (Fuchs *et al.*, 1978a; Oberlies *et al.*, 1980); **3)** pyruvate (Hüster and Thauer, 1983); **4)** succinate (Diekert *et al.*, 1980b); **5)** fumarate (Fuchs *et al.*, 1978b); **6)** 5-aminolevulinate (Diekert *et al.*, 1980a); **7)** and phenylacetic acid (Tersteegen *et al.*, 1997).

Table 2.1 Comparision of *Methanothermobacter* species studied in this work.

	***M. thermautotrophicus* ΔH**	***M. thermautotrophicus* Z-245**	***M. marburgensis* Marburg**
Isolation Location	AD [1] sludge, Urbana, Illinois, USA [2]	Isolated by [3]	AD sludge, Marburg, DE [4]
Genome Length	1.751 Mb [5]	Scaffold level from metagenome [6]	1.639 Mb [7]
Model Organism	✓	X	✓
Optimal pH (range)	7.0-8.0, (6.0-8.8) [8]	No literature value found	6.8-7.4, (5.0-8.0) [8]
Optimal Temp (°C)	65-70 [8]	55 [9]	65 [8]
Max μ (h^{-1}) [10]	0.23-0.3 [11]	No literature value found	0.69 [12]
Plasmid (length)	X	pFZ1 (11 Mb) [13]	pME2001 (4.4 Mb) [13]
Substrates	H_2/CO_2, very limited CO	H_2/CO_2, formate	H_2/CO_2, limited CO

[1] AD, anaerobic digester
[2] (Zeikus and Wolfe, 1972)
[3] (Zhilina and Ilarionov, 1984)
[4] (Fuchs *et al.*, 1978a)
[5] (Smith *et al.*, 1997)
[6] (Rinke *et al.*, 2021)
[7] (Liesegang *et al.*, 2010)
[8] (Wasserfallen *et al.*, 2000; Zabranska and Pokorna, 2017)
[9] (Yamamoto *et al.*, 1989; Zhilina and Ilarionov, 1984)
[10] μ, growth rate
[11] (Seely *et al.*, 1983; de Poorter *et al.*, 2007; Rittmann *et al.*, 2015)
[12] (Schönheit *et al.*, 1980)
[13] (Luo *et al.*, 2001)

2.2.2 Formate

The ability to utilize formate as the sole energy and carbon source for methanogenesis (**Equation (2.2)**) is limited in *Methanothermobacter* to strains that harbor the formate dehydrogenase cassette. Six *Methanothermobacter* spp. were already identified with this capability (Wasserfallen *et al.*, 2000), and this number is increasing as new microbes are isolated (Hassa *et al.*, 2019). The formate dehydrogenase cassette is an operon comprised of a formate transporter subunit (*fdhC*) and two additional subunits (*fdhA* and *fdhB*) found in that order (Nölling and Reeve, 1997; Nölling *et al.*, 1993; Schauer *et al.*, 1986; Tanner *et al.*, 1989; White and Ferry, 1992). Formate is a crucial intermediate for biomass synthesis, specifically in purine biosynthesis and DNA production (Kaster *et al.*, 2011b). *M. thermautotrophicus* ΔH and *M. marburgensis* lack the formate dehydrogenase cassette, however, both have putative *fdhA* and *fdhB* (MTH1139 and MTH1140 and MTBMA_c15220 and MTBMA_15230, respectively) that were predicted for formate production. *M. thermautotrophicus* ΔH is also predicted to have the main subunit of the pyruvate formate-lyase (*pfl*) and its activating enzyme (MTH346 and MTH345, respectively); however, it was not proposed to be responsible for formate generation (Kaster *et al.*, 2011b). The formate dehydrogenase cassette of *M. thermautotrophicus* Z-245, which permits it to grow on formate, was recently added to a shuttle-vector plasmid system (pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$) and harnessed to genetically modify *M. thermautotrophicus* ΔH enabling the modified microbe to also grow on formate (Fink *et al.*, 2021).

$$4\,\mathrm{HCOOH} \rightarrow \mathrm{CH_4} + 3\,\mathrm{CO_2} + 2\,\mathrm{H_2O} \tag{2.2}$$

2.2.3 Biotechnological advantages

The microbes present some characteristics that are opportune for biotechnological applications. As discussed previously, the microbes grow anaerobically on a minimal salt medium which facilitates both production of the medium and reduces the likelihood of microbial contamination. Despite the classification of strict anaerobes, the methanogens exhibit some tolerance to oxygen (Martin *et al.*, 2013; Kiener and Leisinger, 1983; Kiener *et al.*, 1988). The sum reaction for methanogenesis (**Equation (2.1)**), has an overall change in Gibb's free energy (ΔG°') of -131 kJ/mol (the sum of

two exergonic and six endergonic reactions), resulting in methanogenesis as an exothermic process (more details in **Section 2.3**). The thermophilic nature of the microbes lowers the cooling requirement for growing the microbes and additionally assists in preventing microbial contamination (Martin *et al.*, 2013). These thermophiles also have relatively short doubling times compared to other methanogens (particularly mesophiles) (Lyu *et al.*, 2018), and can produce CH_4 at high rates (with low byproduct production) (Martin *et al.*, 2013). It is also worth noting that *Methanothermobacter thermautotrophicus* Hveragerdi is being implemented commercially for power-to-gas applications by the company Electrochaea (Martin *et al.*, 2013).

2.3 The Wolfe Cycle of hydrogenotrophic methanogenesis

Hydrogenotrophic methanogenesis begins with CO_2, which is subsequently reduced four times using H_2 (**Figure 2.2**). First, CO_2 is reduced to formyl-methanofuran (FMFR) consuming the first H_2 molecule (which was used to reduce ferredoxin) and producing the first water (H_2O) molecule (step 1 in **Figure 2.2**). The enzyme FMFR dehydrogenase (Fwd/Fmd) carries out this reaction. The formyl group is then transferred from MFR to another coenzyme tetrahydromethanopterin (H_4MPT) by FMFR/H_4MPT formyl-transferase (Ftr, step 2 in **Figure 2.2**). Methenyl-H_4MPT (MeH_4MPT) cyclohydrolase (Mch) then dehydrates the FH_4MPT, removing the second water molecule and producing MeH_4MPT (step 3 in **Figure 2.2**). MeH_4MPT is then subsequently reduced twice by either coenzyme F_{420} (steps 4a and 6 in **Figure 2.2**), which itself is reduced by two H_2 molecules (Frh, steps 5 and 7 in **Figure 2.2**) or first directly by H_2 and then subsequently by coenzyme F_{420} (steps 4b and 6 in **Figure 2.2**). This produces the intermediates methylene-H_4MPT (MyH_4MPT) by the enzyme MyH_4MPT dehydrogenase (Mtd, step 4a in **Figure 2.2**) or the H_2-forming methylene-H_4MPT dehydrogenase (Hmd, step 4b in **Figure 2.2**) and methyl-H_4MPT (MH_4MPT) by the enzyme MyH_4MPT reductase (Mer). The activity of the Mtd and Hmd together (Mtd/Hmd cycle) may be used to circumvent nickel limitations that prevent the formation of the Frh that is nickel dependent (Hendrickson and Leigh, 2008; Kaster *et al.*, 2011b). By the enzyme MH_4MPT/CoM M-transferase (Mtr), the methyl group is transferred to sulfhydryl coenzyme M (CoM-SH) to produce M-S-CoM in the

first exergonic reaction of the cycle, with simultaneous generation of a sodium ion (Na^+) gradient over the membrane (step 8 in **Figure 2.2**). The Na^+ gradient is required to power adenosine 5'-triphosphate (ATP) production *via* the ATP synthase (Kaster *et al.*, 2011b). The M-S-CoM is then reduced to CH_4 with the simultaneous oxidation of coenzyme B (CoB-SH) and CoM-SH to form the CoM-CoB heterodisuflide complex (CoM-S-S-CoB) by CoM reductase (Mcr or Mrt). Simultaneously, the Hdr complex with the last H_2 molecule reduces the CoM-S-S-CoB to form CoM-SH and CoB-SH, and reduces ferredoxin (to be oxidized in the first step of the cycle) in the second exergonic reaction of the cycle (Mvh and Hdr, step 10 in **Figure 2.2**) (Costa and Leigh, 2014; Kaster *et al.*, 2011b,a; Thauer, 2012; Zabranska and Pokorna, 2017). This bifurcation step only produces enough ferredoxin for CH_4 (and ATP) production, and anabolic functions need to be supplemented with ferredoxin produced from the energy-converting [NiFe]-hydrogenases (Eha and Ehb, step 12 in **Figure 2.2**) (Lie *et al.*, 2012). Alternatively, hydrogenotrophic methanogenesis can begin with formate. In this case, the formate dehydrogenase cassette simultaneously converts the formate into CO_2 and produces a reduced coenzyme F_{420} (step 11 in **Figure 2.2**). The CO_2 is then reduced as previously describes, while the reduced coenzyme F_{420} is oxidized to produce H_2 (required by other enzymes in the Wolfe cycle), likely *via* the HMD/MTD cycle or the Frh (steps 4a and 4b or 5 and 7 in **Figure 2.2**) (Hendrickson and Leigh, 2008).

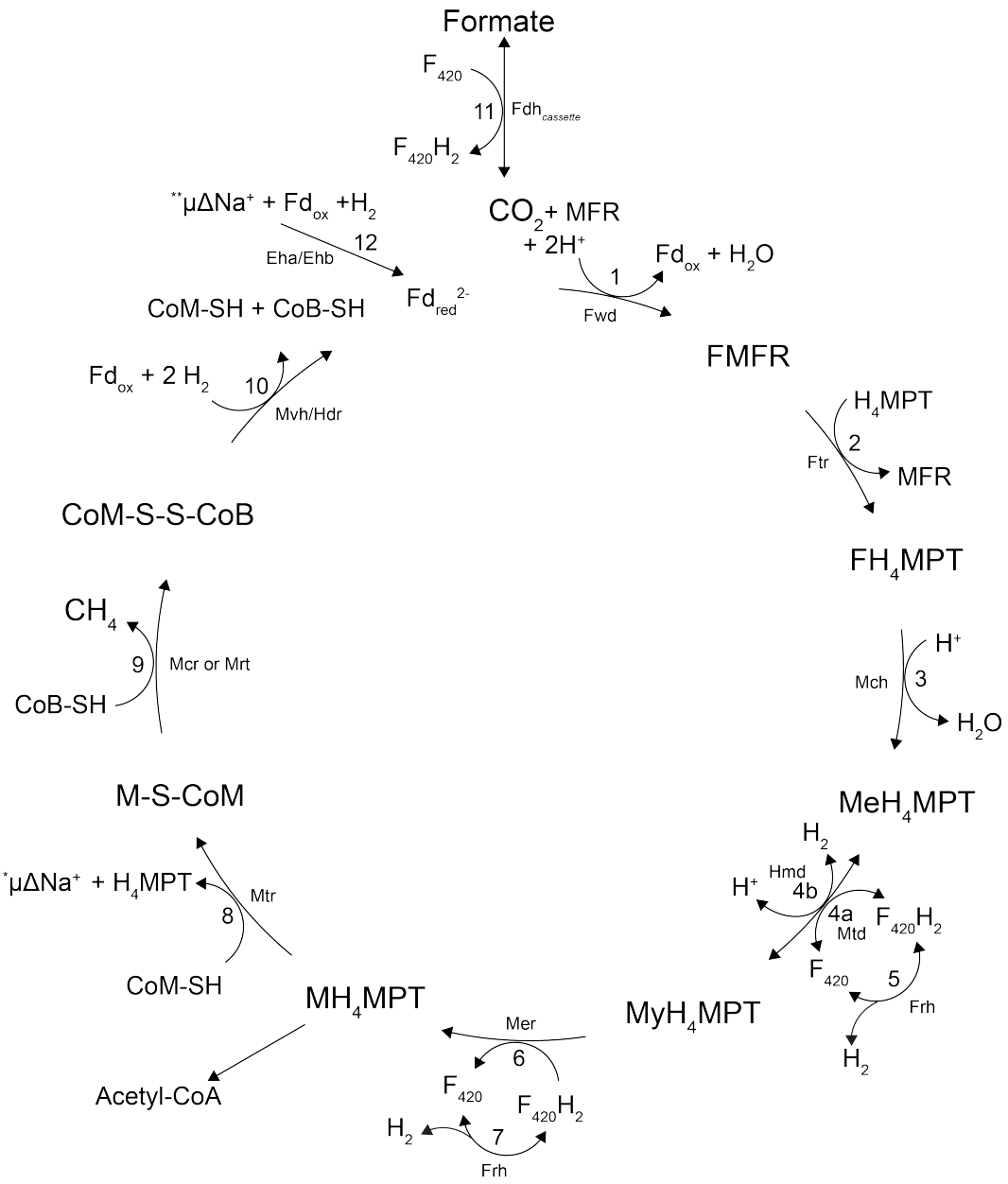

Figure 2.2: The Wolfe Cycle adapted from Thauer (2012). MFR, methanofuran; F, formyl; H_4MPT, tetrahydromethanopterin; Me, methenyl; My, methylene; M, methyl; CoM, coenzyme M; CoB, coenzyme B; CoM-S-S-CoB, CoM-CoB heterodisulfide, CoA, Coenzyme A. $Fdh_{cassette}$, formate dehydrogenase cassette; Fwd, FMFR dehydrogenase; Ftr, FMFR/H_4MPT formyltransferase; Mch, MeH_4MPT cyclohydrolase; Frh, F_{420}-reducing hydrogenase; Mtd, MyH_4MPT dehydrogenase; Hmd, H_2-forming MyH_4MPT dehydrogenase; Mer, MyH_4MPT reductase; Mtr, MH_4MPT/CoM methyltransferase; Mcr or Mrt, CoM reductase; Mvh/Hdr, F_{420}-non reducing hydrogenase with the heterodisuflide reductase; Eha/Ehb, energy-converting [NiFe]-hydrogenases; *, sodium ion export; **, sodium ion anaplerotic import.

2.4 Energy metabolism differences in methanogens

The energy metabolism of the order of Methanobacteriales (and Methanomicrobiales, Methanococcales, and Methanopyrales) differs from that of the order Methanosarcinales and Methanocellales, which have cytochromes and methanophenazines, which are proteins and cofactors involved in oxidation and reduction reactions, respectively (Beifuss *et al.*, 2000; Jussofie and Gottschalk, 1986; Kaster *et al.*, 2011b). In cytochrome- and methanophenazine-lacking methanogens, the exergonic reduction of the CoM-S-S-CoB complex provides the energy for the endergonic reduction of ferredoxin that is used earlier, forming the electron bifurcation process of the Wolfe cycle (Schlegel and Müller, 2013; Thauer, 2012). In these methanogens, an ion-motive force for ATP generation is only generated with the Mtr reaction (Goyal et al., 2016). However, the two membrane bound energy-converting [NiFe]-hydrogenase (Eha and Ehb) and the Na^+/proton (H+) antiporter also influence the Na^+ gradient (Kaster et al., 2011a). Conversely, the cytochrome- and methanophenazine-containing microbes use these ion carriers with a few additional enzymes (membrane-bound hydrogenase (Vho) and the membrane-bound energy-converting hydrogenase (Ech)) to reduce the ferredoxin that is needed for the reduction of CO_2 to FMFR, and to generate an ion-motive force by using this membrane-coupled electron transfer chain in addition to the Mtr reaction Schlegel and Müller (2013).

2.5 Cultivation of *Methanothermobacter*

The three thermophiles grow optimally under similar conditions (**Table 2.1**). The three microbes can produce all of their necessary amino acids and other biomass precursors, and thus, only require a minimal salts (MS) media and an H_2:CO_2 (optimally 4:1) gas mixture (Zeikus and Wolfe, 1972). The composition of one possible medium that can be adopted is listed below (**Table 2.2**) (Gerhard *et al.*, 1993). This was the backbone media composition for the experiments conducted in this dissertation, and alterations are specified in the specific chapters. Varying growth conditions is necessary to better understand the metabolisms of the microbes. Past research has studied the cultivation of microbes; some example studies are discussed below.

Table 2.2 Standard liquid MS-media composition (per liter of media).

Substance long name	Formula	Quantity
millipore water	H_2O	950 ml[1]
sodium chloride	NaCl	0.45 g
sodium hydrogen carbonate	$NaHCO_3$	6.00 g
di-potassium hydrogen phosphate	K_2HPO_4	0.17 g
ammonium chloride	NH_4Cl	1.86 g
magnesium chloride	MgCl x 6 H_2O	0.08 g
calcium chloride	$CaCl_2$ x 2 H_2O	0.06 g
potassium di-hydrogen phosphate	KH_2PO_4	0.23 g
ammonium nickel sulfate (0.2 % w/v)	$(NH_4)2Ni(SO_4)2$	1 mL
iron III chloride (0.2 % w/v)	$FeCl_3$ x 5 H_2O	1 mL
resazurin (0.025 %)	Resazurin	4 mL
trace elements (10x)	Trace elements[2]	1 mL
hydrochloric acid (4 M)	HCl	1 mL
l-cysteine hydrochloride	$C_3H_7NO_2S \cdot HCl$	0.5 g

[1] adjusted to 1000 mL after adding the rest of the constitutents
[2] with nitrilotriacetic acid (NTA) and trisodium nitrilotriacetate as chelating agents (Martin *et al.*, 2013)

2.5.1 Oxygen response

Methanogens are classified as strict or obligate anaerobes due to their limited ability to remove O_2 radicals and the inactivation of some enzymes when exposed to O_2. However, this O_2 sensitivity varies between species and depending on cultivation conditions. Particularly at non-growth conditions, *M. marburgensis* Marburg is less sensitive to O_2 than other methanogens (Kiener and Leisinger, 1983). It is hypothesized that methanogens that possess the reduced F_{420}-dependent oxidase (*fprA* or *fpaA*, genes MTH1350/MTBMA_c17400) are better able to handle oxidative stress (Kiener and Leisinger, 1983; Karr, 2010). The *fprA* is found in an operon that is predicted to encode two other enzymes related to oxidative stress, a putative rubrerythrin (MTH1351/MTBMA_c17410) and a putative rubrerythrin/rubredoxin (MTH1352/MTBMA_c17420), along with a putative transcriptional regulator, MsvR (MTH1349/MTBMA_c17390) (Karr, 2010; Lyu and Lu, 2018). In *Methanosarcina acetivorans*, the MsvR acts as a negative regulator preventing transcription except under oxic conditions (Sheehan *et al.*, 2015). Another oxidative response enzyme is the superoxide dismutase (*sod*) (MTH160/MTBMA_c06110) (Meile *et al.*, 1995). Lastly, to protect and prevent the progression of methanogenesis (potentially also in response to

H_2 limitations), as a hypothesized signaling mechanism, the reduced F_{420} is (reversibly) converted to F_{390} (Hausinger *et al.*, 1985; Vermeij *et al.*, 1997).

2.5.2 Cultivation experiments with *M. thermautotrophicus* ΔH

M. thermautotrophicus ΔH is one of the most studied methanogens, particularly for investigating the genes or enzymes related to hydrogenotrophic methanogenesis. De Poorter *et al.* (2003) conducted continous bioreactor cultivation experiments of *M. thermautotrophicus* ΔH with the aim to understand the bioenergetics of the synthesis of FMFR and the reduction of CoM-S-S-CoB. The authors operated a 3.0 L fermenter with 1.1 L culturing volume (stirred at a speed of 1500 rpm), which was maintained at a pH of 7.0 and a temperature of 65°C. Growth conditions were as follows: **1)** dilution rates from 0.06 to 0.3 (h^{-1}); **2)** gassing rates (H_2/CO_2, 80/20% v/v) from 100 to 400 ($mL{\cdot}min^{-1}$); and **3)** hydrogen partial pressures from 0.015 to 0.500 (bar). The measured growth properties measured included optical density at 600 nm (OD_{600}) and specific CH_4 production ($mol{\cdot}g^{-1}{\cdot}h^{-1}$). Measured bioenergetic parameters included membrane potentials (mV), intracellular pH, the proton-motive force (mV), and the sodium ion-motive force (mV) (de Poorter *et al.*, 2003).

Detailed growth curves (lag, exponential, and linear phases) regarding the effect of H_2 partial pressure are presented by de Poorter *et al.* (2007). The authors concluded that at low H_2 partial pressures, the growth and CH_4 production become fully coupled, which is indicated by diminishing specific growth yields (biomass produced per mole of CH_4) (de Poorter *et al.*, 2007). Both methanogenesis and biomass growth require H_2. This shared dependence creates the coupling effect. Similarly, Pihl *et al.* (1994) discussed the transcription of the *mrt vs.* the *mcr* at different phases of growth. The authors found that during early growth phase (which could mean a larger abundance of H_2), the Mrt is present, while at later stages of cultivation (where H_2 could be more limited), the Mcr was present and in much larger quantities than the Mrt had been. It is to be noted that this experiment was not run until steady state was reached.

Kato *et al.* (2008) varied growth conditions of *M. thermautotrophicus* ΔH and conducted experiments on the whole transcriptome and on selected major enzymes involved in methanogenesis with microarrays and qRT-PCR, respectively. They studied: **1)** low (40°C) and high (80°C) temperatures; **2)** low (6.0) and high (8.5) pH; **3)** H_2 limitation; **4)** no agitation (static); **5)** an oxidative environment; and **6)** an

ammonium-rich environment (Kato *et al.*, 2008). They found four major conclusions: **1)** *M. thermautotrophicus* ΔH does not have a universal stress response mechanism (*e.g.*, in response to oxidative stress); **2)** some of the genes involved in methanogenesis (*mtd* (step 4 in **Figure 2.2**), *mer* (step 6 in **Figure 2.2**), *frh* (steps 5 and 7 in **Figure 2.2**), and *mcr* (step 9 in **Figure 2.2**)) were upregulated under H_2-limited conditions; **3)** some surface cell structures were modified under different environmental conditions such as high pH; and **4)** in some stress conditions, the CO_2 assimilation systems (that require energy) were suppressed (Kato *et al.*, 2008).

Yoshinaga *et al.* (2015) showed that *M. thermautotrophicus* ΔH varies its cell membrane lipid composition under different environmental conditions, particularly limited H_2, potassium, and phosphate. The authors conducted experiments in fed-batch bioreactor of 65 L with a 50 L culturing volume maintained at a pH of 7.5 and 65°C, and gassed at 2 bar (H_2/CO_2, 80/20% v/v or N_2/CO_2, 80/20% v/v for H_2 limited conditions by volume). The authors found that there was a shift from phospholipids to glycolipids that was controlled by different lipid regulatory mechanisms. A few of these mechanisms were found in the study and varied, depending on which nutrient was limited (Yoshinaga *et al.*, 2015).

The effect of redox potential on CH_4 production in *M. thermautotrophicus* (strain not specified) was tested using a bioelectrochemical system (BES) (Hirano *et al.*, 2013). The BES was comprised of a single-chamber electrolysis cell and one electrode, which was set at different voltages to control the redox potential of the system. A model quinone (AQDS) was used as an electron mediator, which was either oxidized or reduced electrochemically. A redox potential of less than -0.5 V is difficult to achieve with traditional reducing agents. With the BES, the authors were able to test potentials as low as -0.8 V. They also tested redox potentials of 0.4 V, -0.1 V, and -0.4 V. Hirano *et al.* (2013) found that the highest cell density and most CH_4 produced per cell was achieved at a redox potential of -0.8 V, which is negative enough to produce H_2, and differs from the -0.2 V to -0.4 V that had previously been documented for methanogenesis. While CH_4 production improved linearly with a lower redox potential, the cell densities did not, rather they showed an upside down bell curve correlation (Hirano *et al.*, 2013). The improved CH_4 production and microbial growth observed at a lower redox potential, highlights the potential of BESs for applications with methanogens.

2.5.3 Cultivation experiments with *M. thermautotrophicus* Z-245

Continuous bioreactor cultivation experiments with *M. thermautotrophicus* Z-245 are limited, with most cultivations related to either its formate dehydrogenase cassette or plasmid (pFZ1). Nölling *et al.* (1991) cultured seven formate utilizing strains in total, including *M. thermautotrophicus* Z-245. *M. thermautotrophicus* ΔH and *M. marburgensis* Marburg were also cultured in this study. *M. thermautotrophicus* Z-245 along with two other strains, FTF and THF, had extrachromosomal DNA found (Nölling et al., 1991). Nölling and Reeve (1997) also cultivated *M. thermautotrophicus* Z-245 in continuous bioreactors both under H_2/CO_2, and Na-formate, while targeting the transcription of formate dehydrogenase cassette. They reached a few conclusions: **1)** the formate dehydrogenase cassette is almost always transcribed in *M. thermautotrophicus* Z-245 regardless of the substrate supplied; **2)** methanogenesis proceeds *via* the same enzymes even from formate; **3)** the *mcr* is the only (methanogenesis) gene detected *via* transcription under growth-halting H_2 limitation (Nölling and Reeve, 1997).

2.5.4 Cultivation experiments with *M. marburgensis* Marburg

Medium dilution rates and H_2/CO_2 gas inflow rates were altered in experiments that were performed with *M. marburgensis* (Rittmann *et al.*, 2012). The study showed that biomass concentration decreased and specific CH_4 production increased with increasing dilution rates. This indicates that at a lower liquid dilution rate, uptake and production rates were volumetrically the highest, thus increasing the efficiency. Increasing the gas inflow rates did not increase the biomass concentration nor the specific CH_4 production. It was concluded that specific CH_4 production was indeed a function of medium dilution rates (Rittmann *et al.*, 2012). Similarly, also using *M. marburgensis* Marburg, Seifert *et al.* (2014) tested the effect of bioreactor pressure, H_2 partial pressure, and inlet gas flow rate on the total CH_4 evolution rate (MER), the volumetric CH_4 content in the off-gas, and the specific CH_4 production. They also found that the gas-liquid mass transfer was the limiting factor. Generally, increasing the H_2 partial pressure improved the CH_4 production. Increasing the bioreactor pressure did not make much difference alone, and increasing the inlet gas flow rates improved the overall quantity of CH_4 produced, but not the efficiency of the process (there was less CH_4 volumetrically in the off-gas). However, the two conditions together resulted in better quantity and quality of the gaseous product

(Seifert *et al.*, 2014).

Bernacchi *et al.* (2014) cultivated *M. marburgensis* Marburg in continuous bioreactors (2 L and 10 L) with the objective of identifying parameters that would influence the ratio of biomass growth rate to CH_4 production. The aim was to change the biomass growth without harming the CH_4 production. This was done simultaneously, conducting dynamic experiments that were designed by implementing fractional factorial design with a multilinear regression and analysis of variance (ANOVA). Eight different parameters were controlled in 18 bioreactor runs. Medium dilution rate and ammonium concentration could be altered to change the biomass growth without affecting the CH_4 evolution rate (Bernacchi *et al.*, 2014). In a follow-up study, the authors developed a workflow to increase the production of CH_4 in a bioreactor using *M. marburgensis* Marburg (Bernacchi *et al.*, 2016).

2.5.5 Cultivation experiments with other *Methanothermobacter* spp.

Past research using *M. thermautotrophicus* Hveragerdi, showed that the main limiting factor in both CH_4 production and biomass growth was the H_2 gas transfer rate, which was determined by temperature, headspace pressure, mixing rate, and concentration of CO_2 in the headspace. The high temperature required by *M. thermautotrophicus* resulted in a low H_2 solubility. By increasing the headspace pressure and mixing rate, and decreasing the concentration of CO_2 (not past the 4:1 stoichiometric ratio of H_2 to CO_2), it was possible to increase the CH_4 production (Martin *et al.*, 2013; Schill *et al.*, 1999). Schill *et al.* (1996) found that in continuous bioreactors (using *M. thermautotrophicus* Hveragerdi), dilution rates, as well as the gas influent rates, were vital to both the increased consumption of substrate gas and the production of CH_4. However, with increased dilution rates, biomass concentration decreased (Schill *et al.*, 1996). Kato *et al.* (2014) ran a coculture of *M. thermautotrophicus* TM and *Thermacetogenium phaeum*, where syntrophically acetate was converted to H_2 and CO_2 by the latter and then those products to CH_4 by the former. They showed that a higher CO_2 concentration limited that process in terms of overall CH_4 production (Kato *et al.*, 2014).

2.6 Genome-scale metabolic models (GEMs)

Genome-scale metabolic models (GEMs) are mathematical models that represent the metabolic functions of a microbe as a network of metabolites (nodes) linked by reactions. Reactions are associated with the genes by means of the enzyme (coded by that gene) that would catalyze that respective reaction. There are two ways of constructing GEMs, the more common bottom-up approach, which is described and applied in this work, and the top-down approach, which employs a large amount of experimental data with statistical methods to create a model (Palsson, 2015). The bottom-up GEMs are built using a COnstraint-Based Reconstruction and Analysis (COBRA) approach, which is able to represent the genotype-phenotype relationship while employing physiochemical constraints (Lewis *et al.*, 2012). Thus, GEMs are a source of Biochemical, Genomic, and Genetic (BiGG) data, which can be harnessed for metabolic engineering (especially *in-silico* planning of genetic modifications), analysis of biological network properties and interspecies interactions, and the prediction of cellular phenotypes and evolutionary processes (Chubukov *et al.*, 2016; Palsson, 2015). The organisms that already have been modeled range from microbes to unicellular eukaryotes to humans (Nilsson *et al.*, 2017b). Recent examples for application of GEMs are found across disciplines from pharmaceuticals and drug development (Apaolaza *et al.*, 2018; Dunphy and Papin, 2018), to cancer research (Masoudi-Nejad and Asgari, 2015; Nilsson and Nielsen, 2017; Resendis-Antonio *et al.*, 2010), to alternative energy generation (Marcellin *et al.*, 2016; Zuñiga *et al.*, 2016), and chemical production (Chen *et al.*, 2017; Gustavsson and Lee, 2016; Pinu *et al.*, 2018; Valgepea *et al.*, 2017a).

The construction of the models is labor and time-intensive, with that of *Escherichia coli* illustrating this perfectly. Its first GEM was created in 2000 (Edwards and Palsson, 2000), but each year, new publications emerge with the latest updates based on new understandings of the microbe and its metabolism (Monk *et al.*, 2017). The published models can be found on a webpage sponsored by the University of Minho (Cardoso, 2018). However, most organisms are not given that great level of attention. From **Figure 2.3**, and **Figure 2.4** taken from the publications Kim *et al.* (2017) and Gu *et al.* (2018), respectively, it is possible to see the ongoing uneven distribution of GEM reconstructions throughout the different domains of life, specifically an under-representation for Archaea (Kim *et al.*, 2012, 2017; Oberhardt *et al.*, 2009). Currently, only 127 of 6239 (2%) of GEMs are archaeal, and of those, for only nine microbes were manually curated (**Figure**

2.4) (Gu *et al.*, 2018). Thor *et al.* (2017) provides a more detailed comparison of the different manually curated models for archaea. Over one thousand manually curated biological models, of which 179 are GEMs, are hosted at the BioModels repository by the European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI) (`https://www.ebi.ac.uk/biomodels/`, 03/2022). However, no manually curated GEM exists for *M. thermautotrophicus* ΔH, *M. thermautotrophicus* Z-245, nor *M. marburgensis* Marburg.

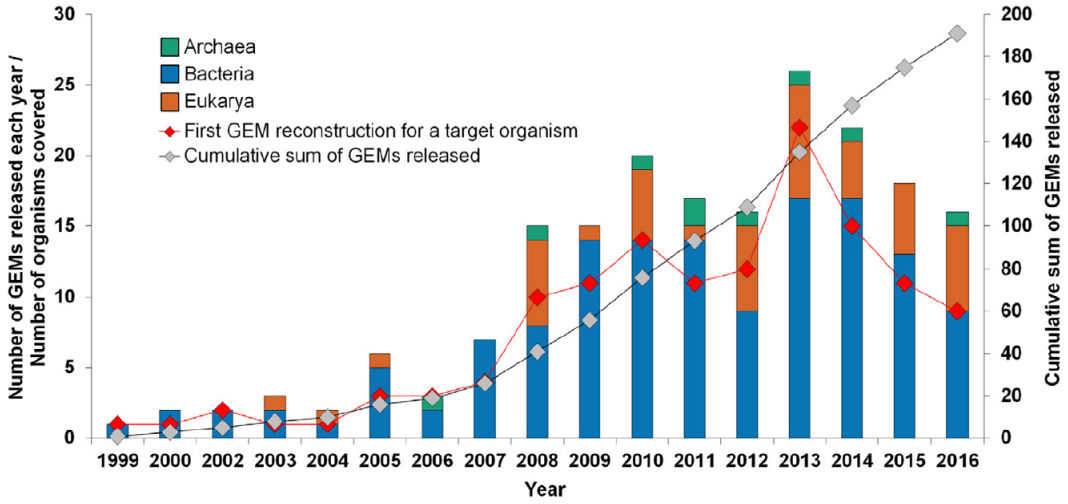

Figure 2.3: The total number of GEMs constructed by domain of life taken from Kim *et al.* (2017).

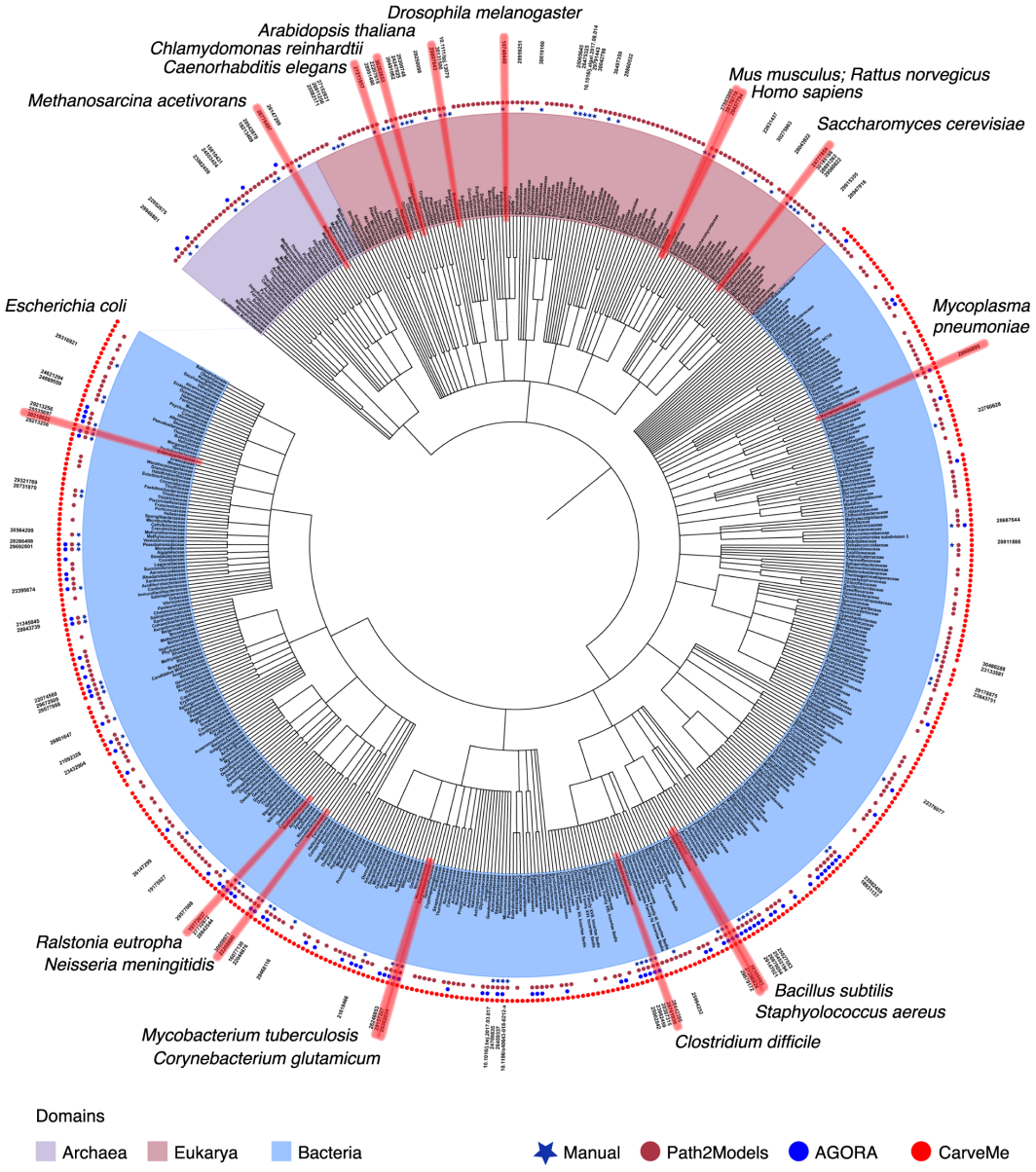

Figure 2.4: The distribution of GEMs constructed by domain of life and method of construction taken from Figure one of Gu *et al.* (2018).

Of the ten archaea species for which models have been built (meaning the genome sequences are available), only one is a Crenarchaeota, while the rest are Euryarchaeota. Of these Euryarchaeota, seven are methanogens: **1)** *Methanosarcina acetivorans* (Benedict *et al.*, 2012; Kumar *et al.*, 2011; Nazem-Bokaee *et al.*, 2016; Peterson *et al.*, 2016); **2)** *Methanosarcina mazei* (Robertson *et al.*, 2005); **3)** *Methanosarcina barkeri* (Feist *et al.*, 2006; Gonnerman *et al.*, 2013); **4)** *Methanospirillum hungatei* (Hamilton

et al., 2015); **5)** *Methanococcus maripaludis* S2 (Goyal *et al.*, 2014); **6)** *Methanocaldococcus jannaschii* (Tsoka *et al.*, 2004); and **7)** *Methanobrevibacter smithii* (Shoaie *et al.*, 2013; Thor *et al.*, 2017). Additionally, a model for *Methanoperedens nitroreducens*, a methanotroph or a methane consumer, was recently released (He *et al.*, 2022). The methanotroph's methane utilizing pathways are almost identical to those producing methane in *Methanosarcina*; however, they operate in the inverse direction (He *et al.*, 2022). The genome sequences and the models of these eight microbes may serve as a good comparison to those of *M. thermautotrophicus* ΔH, *M. thermautotrophicus* Z-245 and *M. marburgensis* Marburg. Of the 12 *M. thermautotrophicus* strains (DSMZ, 2018), only ΔH has a completed published sequence (Smith *et al.*, 1997). *M. thermautotrophicus* Z-245 has only a scaffold-level genome sequence derived from metagenomics (Rinke *et al.*, 2021). The strains still missing are: Hveragerdi (Icelandic - DSM3590) (Butsch and Bachofen, 1981), CaT2 (Kosaka *et al.*, 2014), ZH3 (Stettler *et al.*, 1995), HMT-1 (Kadam and Ranade, 1992), SF-4 (Yamamoto *et al.*, 1989), FTF (Touzel *et al.*, 1988), YTB (Zeikus *et al.*, 1980), JW501 and JW510 (Wiegel, 1990), TM (Hattori *et al.*, 2000), and THF (Wasserfallen *et al.*, 2000; Zinder and Koch, 1984). The genome of *M. marburgensis* Marburg (Fuchs *et al.*, 1978a), has also been sequenced, and the genome sequence compared to *M. thermautotrophicus* ΔH (Kaster *et al.*, 2011b; Liesegang *et al.*, 2010).

A high-quality model requires high-quality data, and GEMs are not excluded from this. Our limited understanding of organisms is aggravated by the challenges involved in measuring the components that could be implemented for constraining and validating the models. Not all metabolic pathways are known nor may they be studied in the organisms of interest. This limits the breadth of the GEMs. Further, the details of those pathways and enzymes involved may be unknown. This limits the depth of the GEMs. However, methods such as omics (genomics, transcriptomics, proteomics, and metabolomics) have improved our ability to analyze the metabolism of cells, and can be applied as constraints after the GEM reconstruction.

2.7 GEM reconstruction

As mentioned earlier, the reconstruction process of a GEM is long and detailed. The publication by Thiele and Palsson (2010) outlined the 96 steps that have become the

standard protocol for constructing these models (Thiele and Palsson, 2010). The four major stages are listed below. Multiple reviews have been published that describe the tools that can be used for the construction process (Copeland *et al.*, 2012; Faria *et al.*, 2018; Mendoza *et al.*, 2019).

1. Automated construction of a draft model
2. Manual refinement of the draft model
3. Conversion of the model into a mathematical model
4. Quantitative evaluation and refinement of the model (*e.g.*, through validation using fermentation data)

The COBRA approach becomes appropriate when the data that is typically needed for constraining a model is either limited or unavailable. Initially, the conservation of mass and energy laws are observed to constrain the solution space of the model. This is done after the conversion of the model to a mathematical model (Orth *et al.*, 2010). The conversion to a mathematical model is executed by the formation of a stoichiometric matrix (S-matrix, S). This matrix has the reactions in the network as the columns and the metabolites as the rows, where the values indicate the stoichiometric value of a given metabolite in a given reaction. The literature-defined sign convention is as follows: a value less than zero indicates it is being consumed; thus, it is a reactant; a value greater than zero indicates it is being produced; thus, it is a product. The S-matrix is considered a sparse matrix because most of the values are zero. The connectivity (ρ_i) is used to describe in how many reactions a metabolite (i) is present. This number can be found by summing the number of times that a column for a given metabolite is non-zero. Similarly, for the reactions, the participation number (π_i) is the number of metabolites that participate in a given reaction (i). It is found by summing the number of times a reaction has a non-zero stoichiometric value. **Equation (2.3)** shows the product of the S-matrix and the flux vector (υ), a vector with all the fluxes of all the reactions in the network), which is the derivative of the concentration vector. The derivative of the concentration vector is the same as the dynamic mass balance ($\frac{dx}{dt}$) (Palsson, 2015). The initial constraint is assuming steady-state conditions, meaning that the rate of change of the mass of each metabolite is constant with time; $\frac{dx}{dt} = 0$ (Orth *et al.*, 2010). Incorporation of the omics data is then adopted to validate the model and constrain it further, as described further below (Marcellin and Nielsen, 2018).

$$\frac{dx}{dt} = S\upsilon \tag{2.3}$$

The biomass reaction is one of the most important reactions in the model. This reaction is heavily debated, and a recent publication demonstrated the problems that are caused by using a reaction that is not organism-specific (Xavier *et al.*, 2017). Given that different biomasses have different compositions, and thus metabolite requirements, the availability of metabolites for other pathways in the network changes, altering the results of modeling such as flux balance analysis (FBA). There is the chance that pathways are active/inactive or present/not present (because they are required for the biomass reaction) in the network that should not be, which once again will change the results of the simulations. Further, it is critical that the biomass is defined as having a molecular weight of 1 g/mmol to allow fair comparison between different conditions and different models. To achieve this molecular weight, it is likely that scaling will need to be done from the experimentally measured biomass composition (Chan *et al.*, 2017).

The biomass reaction should contain components of protein (comprised of amino acids), RNA (comprised of nucleic acids), DNA (comprised of nucleic acids), lipids (comprised of fatty acids or isoprene units), lipopolysaccharide (LPS, constructed from a lipid and polysaccharide), peptidoglycan (murein) or pseudomurein, glycogen (polysaccharide), polyamines, and everything left (Thiele and Palsson, 2010). In the case of *M. thermautotrophicus* ΔH, *M. thermautotrophicus* Z-245 and *M. marburgensis* Marburg, an unusual cell wall composed of pseudomurein, not murein as in bacteria, is found. Further, the cell membrane in the three microbes, like in all archaea, differs in composition and structure from those of bacteria. The membrane lipids of archaea have an ether rather than ester bond and isoprene units rather than fatty acids (Koga, 2011; Koga and Morii, 2007). Lastly, the isoprenoid hydrocarbon side chains are linked to a *sn*-glycerol-1-phosphate backbone rather than a *sn*-glycerol-3-phosphate backbone (Jain *et al.*, 2014). This means that special attention needs to be given when constructing the model to find pathways that build the correct precursors for pseudomurein and the isoprenoid chains (Steenbakkers *et al.*, 2006; Visweswaran *et al.*, 2011), particularly because those compositions have been found to change under different environmental conditions (Yoshinaga *et al.*, 2015). Duboc *et al.* (1995) analyzed the biomass composition (given in percent mass) of *M. thermautotrophicus* Hveragerdi for standard growth conditions (batch (1.5 L), 60°C, pH of 6.8, H_2/CO_2

80/20% v/v gas ratio) after the exponential growth phase in terms of macromolecules: protein (60%), carbohydrates (could not be determined), RNA (6%); elemental composition: C (42.01%), H (5.89%), N (10.88%), S (0.49%), P (3.71%), and ash (17.35%); and ion fractions: K (4.31%), Mg (0.23%), Na (0.97%), P (3.7%), Ca (0.964 x 10^{-6} %), B (0.80 x 10^{-6} %), Co (0.15 x 10^{-6} %), Cr (0.20 x 10^{-6} %), Cu (0.13 x 10^{-6} %), Fe (8.20 x 10^{-6} %), Mn (0.02 x 10^{-6} %), Mo (0.05 x 10^{-6} %), Ni (1.80 x 10^{-6} %), and Zn (0.60 x 10^{-6} %) (Duboc *et al.*, 1995). However, this publication does not specify all the macromolecules required for modeling, nor the biomass composition at different growth phases or conditions.

Since the publication of the Thiele and Palsson (2010) protocol, advances have been made in the field of GEM reconstructions. In particular, as modeling has become more popular, attempts have been made to standardize not only the procedure but also the format of models, especially making them more accessible to individuals that were not involved in their construction (Tsigkinopoulou *et al.*, 2017). The Minimum Information Requested In the Annotation of biochemical Models (MIRIAM) comprises guidelines regarding: **1)** the language the models are written in; **2)** a reference (*e.g.*, publication) that is biologically relevant to the model; and **3)** a set of inputs and results that are reproducible (Le Novère *et al.*, 2005). Though the Thiele and Palsson (2010) protocol did specify to include different database identifiers, more recently, to be MIRIAM compliant requires the inclusion of Uniform Resource Identifiers (URI) for reactions and metabolites using the MIRIAM identifiers database (`http://identifiers.org`). These URIs help to overcome the difficulty of having few standards for naming reactions and compounds. The identifiers can be used to link the same object named differently across different databases Juty *et al.* (2011). The BiGG knowledge base is also working to standardize modeling nomenclature (particularly reaction and metabolite IDs). This knowledge base contains GEMs that have fixed reaction and metabolite IDs (also searchable within the knowledge base) (King *et al.*, 2015a). This standardization of IDs means that tools can be written and employed across models. Tools like ModelPolisher become available when a model is written with the BiGG IDs. This tool will check the model compliance and add MIRIAM compliant identifiers (Römer *et al.*, 2016). Tools, such as MEtabolic MOdel Test (MEMOTE), have been developed to test the quality of the model after construction, biochemically (*e.g.*, for mass and charge balance in reactions) but also for MIRIAM-compliance (*e.g.*, for the inclusion of metabolite identifiers from various databases and Systems Biology

Ontology (SBO) terms) (Lieven *et al.*, 2020).

2.8 GEM analytical techniques

The ability to genetically modify organisms and alter their metabolic pathways to produce substances of interest has led to the concept of (microbial) "cell factories," typically placed in the field of metabolic engineering (Nielsen, 2001). There are a plethora of methods for using and analyzing the constructed GEM, including some biased and some unbiased. Already the 2012 review by Lewis *et al.* (2012) listed almost 150 methods and 25 software tools (**Figure 2.5**). These methods covered elementary flux modes (EFM), strain design, extreme pathway (ExPa, identification), community modeling, flux balance analysis (FBA), gap-filling, gene/reaction perturbations, objective function approximation, omics constraints, regulation, sampling, thermodynamics, and the stoichiometric matrix (Lewis *et al.*, 2012; Zomorrodi *et al.*, 2012). Every couple of years, reviews are released, here are some more recent ones that describe updates to the previously developed tools and workflows (Abarca, 2017; Chen and Nielsen, 2013; Fong, 2014; Hosseini and Marashi, 2017; Long *et al.*, 2015; King *et al.*, 2015c; Suthers *et al.*, 2021).

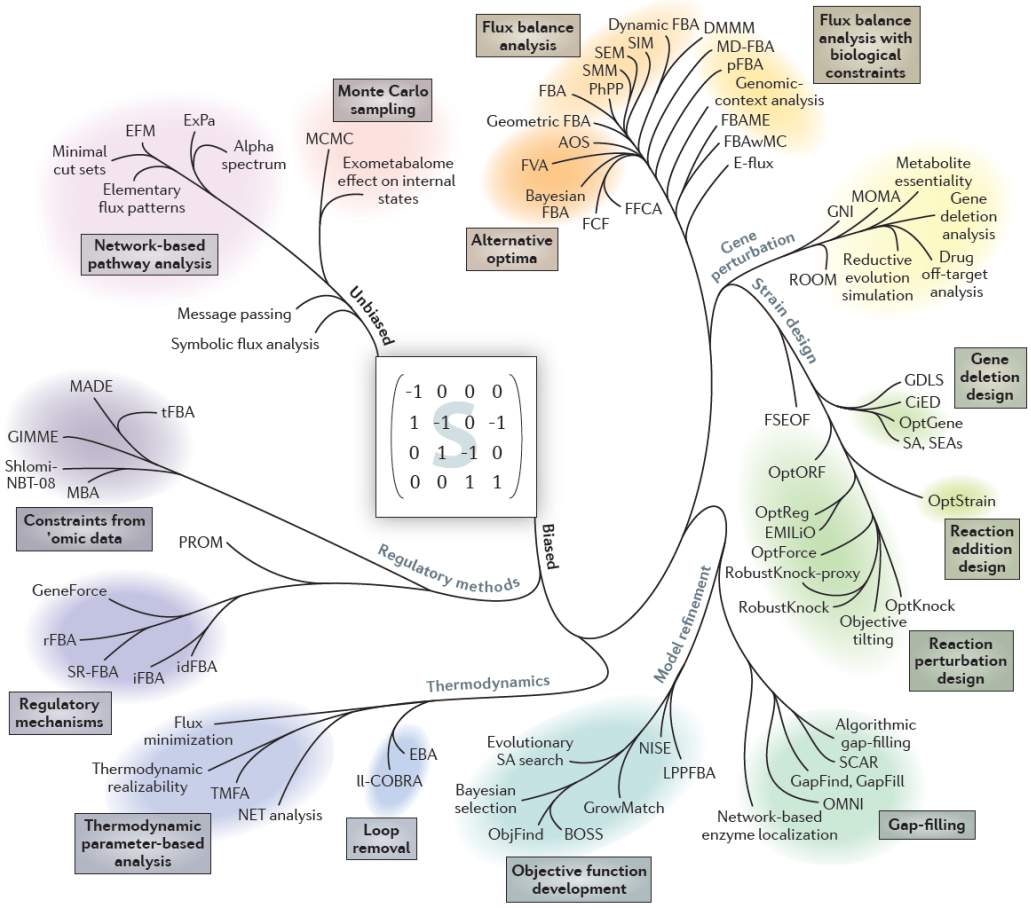

Figure 2.5: Constraint-based modeling methods taken from Lewis *et al.* (2012).

FBA is not only used for analysis after GEM construction but also during the GEM refinement process. It maintains the assumption that the system is operating at steady-state conditions, meaning that the dynamic mass balance ($\frac{dx}{dt}$) is not changing in time, and thus, equals zero (**Equation (2.3)**) (Orth *et al.*, 2010; Palsson, 2015). FBA requires an objective function, which introduces a bias into the optimization problem. The most common objective function is the maximization of the biomass (production or export/ exchange) reaction. However, other examples include: **1)** maximization of ATP yield; **2)** minimization of overall intracellular flux; and **3)** minimization of ATP production (Palsson, 2015; Schuetz *et al.*, 2007). Respecting the constraints of the model and the objective function, FBA solves for the fluxes through each reaction (Orth *et al.*, 2010). There are other unbiased techniques that also predict flux through a network, such as the Markov-Chain Monte Carlo (MCMC) sampling. However, the unbiased methods are

greatly outnumbered by biased methods (O'Brien *et al.*, 2015; Samal and Martin, 2011; Samal *et al.*, 2010; Schellenberger *et al.*, 2012).

2.9 Omics data collection

The omics field of study can be depicted as a pyramid with genomics at the base, followed by transcriptomics, proteomics, and lastly metabolomics at the top. Genomics is the study of the genes on a genome, and transcriptomics is the expression profile of that genome (the RNA), otherwise known as the transcriptome. Both genomics and transcriptomics require sequencing methods for which there are a variety of techniques available. More detailed comparisons for those techniques can be found in the following reviews (Mardis, 2017; Stark *et al.*, 2019). Briefly, over the last 15 years, next-generation sequencing (NGS) has pushed high-throughput analyses both for long-read and short-read sequencing methods. Some of the most common methods, and the ones relevant to this dissertation, are Pacific Biosciences (PacBio) and Illumina sequencing, used in this case for long-read genome sequencing and short-read RNA sequencing, respectively. PacBio is a single-molecule, real-time (SMRT) sequencing technology that converts double-stranded DNA into single single-stranded circular DNA molecules (by adding adapters) that are replicated in individual wells of the SMRT® Cell using polymerases to produce fluorescently labeled DNA strands between 1-50 kilobases (kb) long (Stark *et al.*, 2019). Instead, the Illumina technique first slices and tags DNA strands with adapters, then primers adding indexes and oligos (synthetic single-stranded chain of nucleotides). These oligos have complementary sets in the channels of the flow cells, there the DNA is able to bind. The fixed DNA undergoes a series of complementation and amplification *via* polymerases that also insert fluorescently labeled nucleotides. This occurs for both the forward and reverse strands (Illumina, 2013). Both methods can be employed for DNA and RNA sequencing. When sequencing RNA, the RNA must first be converted into complementary DNA (cDNA) (Stark *et al.*, 2019). Bioinformatic analyses are needed to align the read fragments, and in typical transcriptomics studies, a normalization step to "housekeeping" genes is done (Abrams *et al.*, 2019).

Further, proteomics and metabolomics are additional methods adopted to better understand cellular metabolism and constrain metabolic models. While similar

techniques are used for proteomics and metabolomics, the two approaches differ in the molecules they measure and thus, the information that can be attained. Proteomics investigates the entirety of the proteins available at a given time point. The proteins identified can qualitatively or quantitatively provide insight to potential cellular functions and metabolisms (Richter *et al.*, 2016; Valgepea *et al.*, 2017b). Metabolomics comprises the analyses of both intracellular and extracellular metabolites. Metabolites include, but are not limited to: **1)** sugars; **2)** amino acids; and **3)** other small molecules that are intermediates in the central metabolism. Metabolites are the result of the activity of proteins. Therefore, metabolomics help to elucidate details of metabolic pathways by measuring metabolite concentrations during different fermentation conditions. Knowing metabolite concentrations (even relative) is helpful to predict which metabolic pathway might be active under specific environmental conditions or where bottlenecks in the metabolism occur. The following paragraph briefly outlines some of the techniques and instruments that can be utilized.

Gel electrophoresis, isotope labeling, and label-free quantification with mass spectrometry are techniques that are used for omics studies. Different gel electrophoresis techniques can be adopted to separate molecules of interest, including isoelectric focusing (IEF), sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), and two-dimensional (2D) gels as a combination of IEF and SDS-PAGE. In IEF, the molecules will separate based on their isoelectric point (charge). In SDS-PAGE, the polypeptides are treated with SDS first, which evenly charges molecule. This allows the separation of protein molecules by size. Lastly, in 2D IEF-SDS combination gels, polypeptides can be separated simultaneously by charge and size (Resing and Ahn, 2005).

Mass spectrometry (MS) can be employed for qualitative (identification) and quantitative analysis of both proteins and metabolites. Liquid chromatography (LC) coupled with tandem MS (MS/MS) is often the MS technique adopted to identify proteins or metabolites from complex mixtures. Gel spots from gel electrophoresis can be analyzed by LC-MS for protein molecule identification from less complex mixtures. For quantitative analysis, both labeled and unlabeled samples can be compared. Labeling methods are discussed below.

Stable isotope labeling can be employed, potentially in conjunction with stable isotope probing (SIP), where the flux of isotope-labeled metabolites is traced throughout a pathway of interest. Isotopes, such as hydrogen (H *vs.* ^{2}H), carbon (^{12}C *vs.*

^{13}C), and oxygen (^{16}O *vs.* ^{18}O), can be utilized to distinguish between samples. Dimethyl labeling is commonly employed in which two methyl functional groups are added to a molecule (*e.g.*, peptide digests from complex proteome samples). The methyl groups can have three different weights depending on the hydrogen and carbon isotopes used. Typically, to properly distinguish between molecules, a weight difference of at least four Daltons is required (Boersema *et al.*, 2009). Labeling strategies can be applied during growth or after protein fragmentation. The proteomes from two different growth conditions can be compared when the proteins have been differentially labeled with isotopes during growth. Alternatively, peptides derived from the proteomes of two different growth conditions can be labeled after extraction, for example, with dimethyl labeling.

Label-free quantification is performed by coupling MS with computational calculations. Unlike stable isotope labeling, label-free quantification does not add a label to distinguish between samples, and thus, requires that samples are analyzed individually. Extracted ion chromatograms (XIC)-based label-free strategy paired with the MaxQuant software (Max Planck Institute of Biochemistry, Planegg, Germany) or Proteome Discover™ (Thermo Fischer Scientific, Waltham, USA) can be employed for the quantification. The Sequential Windowed Acquisition of All Theoretical Fragment Ion Mass Spectra (SWATH-MS) software may also be used (Huang *et al.*, 2015).

While transcriptomics, proteomics, and metabolomics each provide slightly different information, theoretically, metabolomics could provide the highest resolution image. However, given limitations in sampling, measuring, and analysis, the ability to accurate quantify the metabolome is still challenging, especially when using oxygen-sensitive microbes (Heffernan *et al.*, 2022). Thus, through the adoption of these omics, more knowledge regarding the use of DNA, RNA, proteins, and metabolites in cellular metabolism can be elucidated, which allows predicting an outcome before attempting to alter a metabolic pathway.

2.10 GEM omics data integration

As previously described, omics are comprised of broad and rich data sets. There are efforts to generate omics with high-throughput methods, and this produces large quantities of complex data. This data can be capitalized on further by integrating it into

GEMs. After the construction of the GEM, the model will require a set of constraints that were defined through the refinement process. However, omics data will help to further reduce the solution space that was generated during the modeling process, and this improves the ability to model what is observed in reality (Blazier and Papin, 2012). Individually, each omics technique presents its own limitations. However, integrated together in a multi-level omics approach, uncertainty can be reduced, and new information gained (Zhang *et al.*, 2010). There are currently numerous algorithms available for integrating omics data into GEMs, and these vary depending on the type of omics data available (Blazier and Papin, 2012; Chae *et al.*, 2017; Saha *et al.*, 2014; Schmidt *et al.*, 2013b). For **transcriptomics**, or gene expression data alone, the following tools have been developed: Gene Inactivity Moderated by Metabolism and Expression (GIMME) (Becker and Palsson, 2008), the Integrative Metabolic Analysis Tool (iMAT) (Shlomi *et al.*, 2008; Zur *et al.*, 2010), Metabolic Adjustment by Differential Expression (MADE) (Jensen and Papin, 2010), Probabilistic Regulation Of Metabolism (PROM) (Chandrasekaran and Price, 2010), Toolbox for Integrating Genome-scale Metabolism, Expression, and Regulation (TIGER) (Jensen *et al.*, 2011), E-Flux2, Simplified Pearson cOrrelation with Transcriptomic data (SPOT) (Kim *et al.*, 2015), the Lee *et al.*, method (Lee *et al.*, 2012), and the HPCOF method (Zhang *et al.*, 2017). For **proteomics**, or protein expression data alone, the following tools have been developed: GECKO (Sánchez *et al.*, 2017) and GIMMEp (Bordbar *et al.*, 2012). MEtabolic and TRanscriptomics Adaptation Estimator (METRADE) requires transcriptomics data and codon usage data (Angione and Lió, 2015). NExt (Martínez and Nielsen, 2014) and GIM3E (Schmidt *et al.*, 2013a) both utilize transcriptomics and metabolomics data, with the former also requiring thermodynamics data. Integrative Omics Metabolic Analysis (IOMA) incorporates **proteomics** and **metabolomics** data (Yizhak *et al.*, 2010). While for **fluxomics** (specifically ^{13}C data), there is the JBEI Quantitative Metabolic Modeling library (jQMM) (Birkel *et al.*, 2017). Lastly, Model-Building Algorithm (MBA) integrates **transcriptomics**, **proteomics**, **metabolomics**, and **phenotypic** data (Jerby *et al.*, 2010).

The need for more data to constrain the models is not satisfied with multiple omics techniques applied independently. Rather, a recent review by Nilsson *et al.* (2017a) called for the use of the kinetome, or a library of kinetic parameters that span the genome of an organism (Nilsson *et al.*, 2017a). The publication gave three strategies towards acquiring the kinetome because, as of now, it has been challenging. One of the mentioned strategies

calculates the enzymes turnover rate (k_{cat}, these values can also be measured with some difficulty) by dividing the metabolic flux or υ (*e.g.*, found by using fluxomics (Nielsen, 2003)) by the protein concentration or E (quantified by proteomics). This is shown mathematically in **Equation (2.4)**. The collection of the k_{cat} values, or the kcatome, can be employed in other types of models (**Section 2.11**).

$$E \geq \frac{\upsilon}{k_{cat}} \tag{2.4}$$

2.11 Next-level GEMs

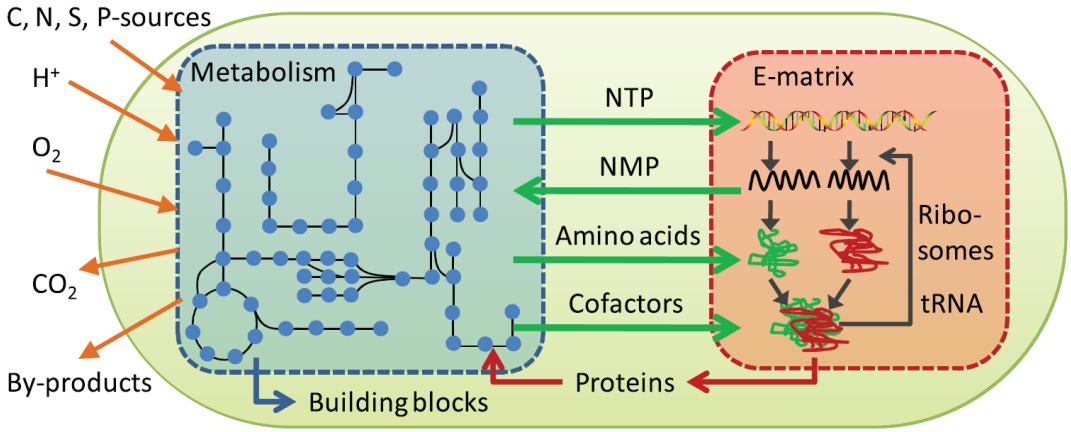

Figure 2.6: The interaction between the S-matrix and the E-matrix taken from Thiele *et al.* (2012).

The next level of a GEM (sometimes termed M-models for clarity by O'Brien and Palsson (2015)) is a metabolism and macromolecule synthesis model. This model nominated the metabolic and expression model (ME-model) merges metabolism with the protein expression (Thiele *et al.*, 2012; Fang *et al.*, 2020). ME-models have the typical S-matrix with the metabolic information, but they also contain an expression matrix (E-matrix) of mRNA and proteins (**Figure 2.6**). The E-matrix is a massive matrix that contains a high count of reactions and components. This count is often much larger than the S-matrix, and in the cases found in Thiele *et al.* (2009, 2012), that difference was more than an order of magnitude.

Proteins have varying costs in terms of the quantity of resources required to produce them. The costs are calculated by analyzing the genome sequence for each protein and

counting how many nucleotides and amino acids are required for the DNA, RNA, and protein itself (Thiele *et al.*, 2009). The ME-models account for this cost to produce the protein and compare it to the benefit received when using it over another protein, with this information stored in the E-matrix. The model is further constrained by enzyme catalytic constraints, the k_{cat} values, which can be found using metabolomics and proteomics data as described earlier. This strategy helps define the transcriptome and proteome states, or gene expression and protein abundance, respectively. The treatment of the biomass reaction differs between the two model types. In GEMs, the flux through the biomass reaction (representing the growth of the organism) is an input. This is not the case for ME-models, where the biomass reaction is one of the outputs (O'Brien and Palsson, 2015). The last major difference between model types, which was added more recently, is the ability to better map the compartmentalization of proteins through protein translocation (Liu *et al.*, 2014). Localization of proteins within cells allows for a spatial representation and analysis of the cell. Nevertheless, the first step to making these more complex models is the construction of a GEM.

There are a plethora of tools to build GEMs (Mendoza *et al.*, 2019), but without thorough curation, the model will not predict experimental results, thus rendering the model less impactful. As more applications are discovered, and the demand for models increases, there is more pressure to produce high-quality models more quickly (Machado *et al.*, 2018). Thus, consideration for the quality of a model over the quantity of models produced (and the economic feasibility) may also play a significant role in the future, making quality control tools more important (Lieven *et al.*, 2020). A more carefully built model can result not only in (partially) curated genome annotations, which are sources of information that can be exploited for biological work, such as genetic modifications, but also in the prediction and verification of biological mechanisms, using techniques described in **Section 2.8**. These results can inform research, such as product optimization (*e.g.*, CH_4 production) and alternative substrate usage (*e.g.*, formate), as well as inter-organism comparisons at a genomic level. GEMs provide the opportunity to attempt *in-silico* experiments in a fraction of the time required in the wet lab, potentially saving time and money.

Chapter 3

Casini, I., McCubbin, T., Esquivel-Elizondo, S., Luque, G.G., Evseeva, D., Fink, C., Beblawy, S., Youngblut, N.D., Huson, D.H., Dräger, A., Ley, R.E., Marcellin, E., Angenent, L.T., Molitor, B. Genome-scale metabolic modeling of *Methanothermobacter* spp. provides an integrated systems-biology platform for power-to-gas technology. *Preliminary Manuscript*.

3.1 Author contributions

Bastian Molitor (B.M.) and Largus T. Angenent (L.T.A.) initiated the work. Isabella Casini (I.C.), B.M,. and L.T.A. designed the experiments. Christian Fink (C.F.), Sebastian Beblawy (S.B.), I.C., and Daria Evseeva (D.E.) carried out the genome sequencing, and D.E. assembled the genomes under supervision of Daniel H. Huson (D.H.H.). I.C. built the models with the assistance of Tim McCubbin (T.M.), Andreas Dräger (A.D.), and Esteban Marcellin (E.M.). Ruth E. Ley (R.E.L.) provided the bioreactor system. I.C. operated the bioreactors with the help of Sofia Esquivel-Elizondo (S.E.-E.) and Nicholas D. Youngblut (N.D.Y.). I.C. adapted the RNA and protein extraction methods with the help of S.E.-E., and E.M.. I.C. analyzed the omics data with the help of T.M. and Guillermo G. Luque (G.G.L.). I.C. analyzed the rest of the data and modelled with the help of T.M and B.M.. L.T.A. and B.M. supervised the project. I.C. produced all the figures and tables. I.C. and B.M. wrote the manuscript, while all edited the paper and approved the final version.

3.2 Abstract

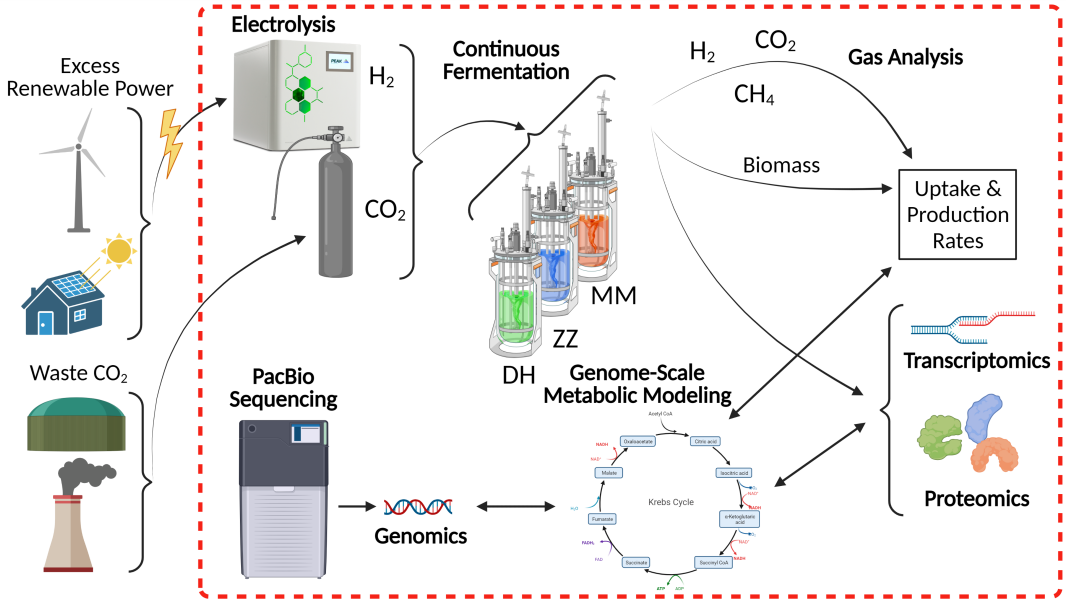

Figure 3.1: Graphical abstract created with BioRender.com

Methanogenesis is the metabolism in which methanogenic archaea (methanogens) generate cellular energy for their growth while producing methane. Hydrogenotrophic methanogens thrive on carbon dioxide and molecular hydrogen as sole carbon and energy sources. Thermophilic and hydrogenotrophic *Methanothermobacter* spp. have been adopted as model microbes to study the physiology and biochemistry of (hydrogenotrophic) methanogenesis for many decades. In recent years, *Methanothermobacter* spp. have been recognized as robust biocatalysts for a circular carbon economy and are now applied in power-to-gas technology on a large scale. While many details on the physiology and biochemical equipment of *Methanothermobacter* spp. are known, relevant questions remain on the interplay of the metabolism, and genetic tools have only recently been established in our lab. Here, we generated the first manually curated genome-scale metabolic reconstruction for three *Methanothermobacter* spp.. We investigated differences in growth performance and gas consumption/production in quadruplicate bioreactors. We found the highest methane production rate for *Methanothermobacter thermautotrophicus* ΔH, while *Methanothermobacter marburgensis* Marburg reached the highest biomass concentration. From the bioreactors at steady state, we collected statistically reliable

transcriptomics and proteomics data sets, which we integrated with our genome-scale metabolic models. The implementation of an umbrella model that contains combined reactions from all three microbes allowed us to perform an interspecies comparison of the omics data sets. While the observed differences in the growth behavior cannot be fully explained with the omics data, the comparison enabled us to identify important differences in growth-related pathways, such as in the production of formate for anabolism.

3.3 Significance

Renewable energy sources (*e.g.*, wind and solar) provide carbon-free electric power, but their intermittency and offset between peak production and peak demand generate the need to store this electric power. Furthermore, these technologies alone do not satisfy the demand for carbon-based commodities. Power-to-gas technology provides a means to store intermittent renewable electric power with concomitant carbon dioxide recycling into chemical energy, such as in the form of methane, on a centralized and decentralized scale. This is particularly important to establish equitable energy strategies for all countries, as is highlighted by the United Nations Sustainable Development Goals. With this work, we provide an integrated systems biology platform to optimize biological power-to-gas technology and to formulate strategies for the production of value-added products with *Methanothermobacter* spp..

3.4 Introduction

Emission of greenhouse gases, primarily carbon dioxide, from fossil fuels to satisfy electric power demands and commodity production is the main driver of anthropogenic climate change. Solutions to mitigate the devastating effects of greenhouse gas emissions worldwide, and to decarbonize the energy and industrial sectors are of urgent need. Societies must reshape in a way that allows the efficient implementation of: **1)** renewable electric power to replace fossil sources for public and industrial energy demands; and **2)** carbon dioxide as feedstock instead of pollution for the production of commodities within a circular carbon economy. Power-to-gas technologies are implemented to convert excess renewable electric power *via* electrolysis of water into

dioxygen and molecular hydrogen. In an additional methanation step, the molecular hydrogen can be combined with carbon dioxide with this gas mixture being processed into methane. Potential carbon dioxide sources that would align with a sustainable circular carbon economy include: **1)** biogas from anaerobic digestion, which contains up to ∼40% carbon dioxide (Martin *et al.*, 2013); **2)** exhaust gas from non-fossil power plants, such as renewable natural gas and biomass; **3)** off-gases from other industrial processes that are difficult to be decarbonized entirely, such as cement and steel production (Angenent *et al.*, 2017; Götz *et al.*, 2016; Schiebahn *et al.*, 2015); and **4)** direct air carbon capture (Yuan *et al.*, 2016; Keith *et al.*, 2018). While power-to-gas can provide carbon-free molecular hydrogen, the "hydrogen-infrastructure" is not well-established, and the addition of molecular hydrogen into the existing natural gas grid is typically limited to under 10% v/v depending on the location (Dolci *et al.*, 2019). However, some studies have considered higher ranges (Melaina *et al.*, 2013; Hall *et al.*, 2021; Zhao *et al.*, 2020). After an additional methanation step, the final product is methane, which resembles the main constituent of fossil natural gas. This methane can be injected into the natural gas grid infrastructure, which is already in place in many communities, for storage, distribution, and consumption purposes, with virtually unlimited capacities. Therefore, methane from power-to-gas has the potential to replace fossil natural gas in the natural gas grid as a decarbonization strategy and for communities to become more independent in securing energy demands (Schiebahn *et al.*, 2015; Guerra *et al.*, 2018).

The methanation step in power-to-gas can be performed *via* thermo-chemical processes, such as the hydrocarbon-forming Sabatier process (Müller *et al.*, 2013). However, these processes typically require high temperature ($> 200^{\circ}$C) and pressures ($>$ 1 MPa), as well as a metal catalyst (*e.g.*, iron, nickel, cobalt, ruthenium) that is sensitive to gas impurities (Dry, 2002; Leonzio, 2016; Van Der Laan and Beenackers, 1999). Alternatively, the methanation step can be performed biologically with microbes, known as hydrogenotrophic methanogenic archaea (methanogens) as biocatalysts, which naturally metabolize carbon dioxide and molecular hydrogen to produce methane (**Equation (3.1)**) (Guerra *et al.*, 2018; Leonzio, 2016; Rachbauer *et al.*, 2017). Thermophilic methanogenic species of the genus *Methanothermobacter* have been adopted for this purpose on a large scale (Pfeifer *et al.*, 2021).

$$4\,H_2 + CO_2 \rightarrow CH_4 + 2\,H_2O \tag{3.1}$$

Over the last 40 years, *Methanothermobacter thermautotrophicus* ΔH (formerly *Methanobacterium thermoautotrophicum* ΔH) and *Methanothermobacter marburgensis* Marburg (formerly *Methanobacterium thermoautotrophicum* Marburg) have served as model microbes to study hydrogenotrophic methanogenesis (Wasserfallen *et al.*, 2000). This resulted in an abundance of literature on their core metabolism, methanogenesis, for example, in an extensive comparative genome study (Kaster *et al.*, 2011b). Nevertheless, there are still more than 500 hypothetical proteins and many pathways much less studied than methanogenesis present in these microbes (PATRIC, 2014). Another related species, *Methanothermobacter thermautotrophicus* Z-245 (formerly *Methanobacterium thermoformicicum* Z-245, (Wasserfallen *et al.*, 2000)) has primarily been studied for its formate dehydrogenase genes (Nölling and Reeve, 1997) and its plasmid (pFZ1) (Nölling *et al.*, 1991). All three microbes grow on carbon dioxide and molecular hydrogen, but *M. thermautotrophicus* Z-245 also utilizes formate as a growth substrate due to the presence of a catabolic formate dehydrogenase (Nölling and Reeve, 1997). In addition, *M. thermautotrophicus* ΔH and *M. marburgensis* Marburg both limitedly utilize carbon monoxide (Diender *et al.*, 2016). The three microbes have similar optimal growth temperatures (∼65°C), pHs (∼7), and no requirements for any organic compounds for growth (Lyu *et al.*, 2018). Further, their relatively short doubling time compared to other methanogens (particularly mesophilic species) (Lyu *et al.*, 2018) promotes their use for biotechnological applications. The exothermic nature of hydrogenotrophic methanogenesis renders thermophilic species especially suitable because less energy is required for cooling the bioreactor (Martin *et al.*, 2013). Few studies looked at the full transcriptome or full proteome of *M. thermautotrophicus* ΔH (Kato *et al.*, 2008; Christendat *et al.*, 2000; Liu *et al.*, 2019) or *M. marburgensis* Marburg (Diender *et al.*, 2016), but none looked at *M. thermautotrophicus* Z-245. Further, no studies compared the methanogens using pan-transcriptomics and pan-proteomics (in general there are few pan-proteome studies that compare directly across closely related species (Broadbent *et al.*, 2016; Murugaiyan *et al.*, 2020)), and a few studies collected samples during steady state from bioreactors (de Poorter *et al.*, 2007, 2003; Nölling and Reeve, 1997; Afting *et al.*, 2000).

Genome-scale metabolic models (GEMs) are mathematical models that represent the metabolic capacities of a microbe as a network of metabolites (nodes) linked by reactions. Reactions are associated to (a set of) genes by means of the enzyme (encoded by these genes) that catalyze that respective reaction. GEMs assume steady-state

conditions and conservation of energy and mass. Over the last 20 years, thousands of GEMs have been reconstructed across all three domains of life; however, archaea remain to be underrepresented. Only 127 of 6239 (2%) are archaeal GEMs, and of those, only nine are manually curated (Gu *et al.*, 2019). Many tools have been developed to accelerate the model-building process, mainly through automation, as highlighted elsewhere (Mendoza *et al.*, 2019). However, many of these tools rely on genome annotations, biochemical data, and other models from databases (Thiele and Palsson, 2010), which are more limited and less certain for archaea. The applications of GEMs and related models have been extensively reviewed (Suthers *et al.*, 2021). Briefly, GEMs can be utilized to predict phenotypes and find metabolic bottlenecks for optimization (including through gene deletions or insertions). Furthermore, GEMs provide a backbone for integrating cultivation data, omics data (genomics, transcriptomics, proteomics, metabolomics, and fluxomics), kinetics data, and thermodynamics data, through a variety of methods from flux balance analysis (FBA) to machine learning (Suthers *et al.*, 2021). To assess the differences and potential advantages for bioprocessing between the three different model *Methanothermobacter* spp., we (re)sequenced the genomes, compared them in a pan-genome fashion, reconstructed high-quality manually-curated GEMs, operated quadruplicate continous bioreactors to achieve steady-state conditions, collected transcriptomics and proteomics data sets from these bioreactors, and integrated these data sets into the GEMs. We compare the three microbes and demonstrate how a systems-biology approach can also be applied for interspecies comparisons while providing a rich collection of multi-level data that will be a valuable resource for the scientific community, for example, to help filling the knowledge gaps in archaeal metabolism.

3.5 Materials and Methods

3.5.1 Microbial strains and medium composition

M. thermautotrophicus ΔH (DSM 1053), *M. thermautotrophicus* Z-245 (DSM 3720), and *M. marburgensis* Marburg (DSM 2133) were obtained from the DSMZ (Braunschweig, Germany) and were cultivated essentially as described in Fink *et al.* (2021), with the only variation that the atmosphere of the anaerobic chamber (UniLab Pro Eco, MBraun, Garching, Germany) contained N_2/CO_2, 95/5% v/v. Preparation of

batch and continuous media for bioreactor runs was adjusted from the mineral media of Balch *et al.* (1979) and Martin *et al.* (2013). The mineral medium contained (per liter): nitrilotriacetic acid (NTA), 0.096 g; trisodium nitrilotriacetate, 0.275 g; sodium chloride, 0.675 g; di-potassium hydrogen phosphate, 0.255 g; ammonium chloride, 2.006 g; magnesium chloride hexahydrate, 0.12 g; calcium chloride dihydrate, 0.090 g; potassium di-hydrogen phosphate, 0.345; ammonium nickel sulfate, 1.5 mL (0.2% w/v); iron(II)chloride tetrahydrate, 1.5 mL (0.2% w/v); resazurin indicator solution, 4 mL (0.025% w/v); and trace element solution, 1.5 mL. The trace element solution was prepared 10-fold as stated by (Balch *et al.*, 1979) but with 2 g/L NTA and the pH adjusted to 6.0 with 5 M potassium hydroxide. In the continuous media, 0.02 mL/L of Anti Foam SE-15 (Sigma-Aldrich, Merck, Darmstadt, Germany) was supplemented. All media were prepared with Millipore water (18.2 MΩ·cm). The medium was autoclaved in glass Schott bottles with butyl septa with Masterflex® L/S Norprene Food-Grade Tubing, L/S 14 tubing (Cole-Parmer GmbH, Wertheim, Germany) inserted or directly in the bioreactor vessels (**Materials and Methods, 3.5.8**).

3.5.2 Genome sequencing

M. thermautotrophicus ΔH, *M. thermautotrophicus* Z-245, and *M. marburgensis* Marburg were grown in mineral medium overnight. The entire biomass (50 mL) was collected by centrifugation at 3170 x g and room temperature for 20 min (5920 R Eppendorf, Hamburg, Germany) and the genomic DNA was extracted using a phenol-chloroform extraction method. In brief, to the biomass pellet, 500 μL of cetyltrimethylammounium bromide Buffer (CTAB, made according to Cold Spring Harbor Protocols (2009), however without the Polyvinylpyrrolidone) were added and the pellet resuspended. The mixture was transferred to sterile 2 mL bead beating tubes (containing 500 μL of 0.1 μm BeadBeater® zirconia beads, Carl Roth, Karlsruhe, Germany) and vortexed (Vortex-Genie® 2, VWR International GmbH, Darmstadt, Germany) at 2700 min^{-1} for 5 sec on then 1 sec off repeatedly for 1 min, after which, the tube was placed on ice. 500 μL of ROTI®Phenol/Chloroform/Isoamyl alcohol (ratio of 25:24:1, Carl Roth, Karlsruhe, Germany) was added to tube, the tube inverted and then centrifuged at 4°C and 16000 x g for 10 min (5424, Eppendorf, Hamburg, Germany). The top layer was transferred (using wide orifice pipet tips) to a phase lock tube (prepared ahead by adding 2 mm^3 of vacuum grease Dow Corning® (VWR

International GmbH, Darmstadt, Germany) into 2 mL tubes, which were centrifuged until 9391 x g were reached, and then autoclaved to sterilize). Another 500 μL of ROTI®Phenol/Chloroform/Isoamyl alcohol was added, the tube was inverted to mix, and centrifugation was performed as in the previous step. The supernatant was transferred to new tubes, 500 μL of Chloroform/Isoamyl alcohol (24:1, VWR International GmbH, Darmstadt, Germany) added, the tube was inverted to mix, and centrifugation was performed as in the previous step. The supernatant was added to a fresh tube, and the gDNA was precipitated by adding 0.1 volumes of cold 3 M sodium acetate and 2 volumes of ice-cold absolute ethanol, inverting the tube to mix, and then incubating it overnight at -20°C. The following day, the tube was centrifuged again at 4°C and 16000 x g for 10 min, the supernatant removed and the gDNA washed with 300 μL of ice-cold 70% v/v ethanol. Centrifugation was performed as in the previous step, the supernatant removed, and the tube air-dried for approximately 1.5 h at 50°C in a ThermoMixer® C (Eppendorf, Hamburg, Germany). The pellet was resuspend with 44 μL elution buffer (10 mM Tris·Cl, pH 8.5 and nuclease free water (New England Biolabs, Ipswich, United States)) and allowed to rest for an hour to resolve. To remove RNA, 1 μL of Bovine Ribonuclease A (VWR International GmbH, Darmstadt, Germany) was added and the tube was allowed to rest for 30 min. The quality of the gDNA was then checked using the Femto Pulse System (1.0.0.32, Agilent, Santa Clara, United States) according to manufacturer's instructions (using the Genomic DNA 165 kb Kit (Agilent, Santa Clara, United States) and a 70 min separation time). Library preparation was conducted with the SMRTbell® Express Template Preparation Kit (PN 101-397-100 Version 3, January 2018, Pacific Biosciences, Menlo Park, United States) as specified in the manufacturer's instructions. The genomes were then sequenced using the Sequel I System (Pacific Biosciences, Menlo Park, United States).

3.5.3 Genome assembly

De-novo genome assembly (including the plasmid when applicable) was conducted and methylation patterns were deciphered from the PacBio sequencing results. The utilized pipeline included seven main aspects: **1)** contamination control; **2)** assembly by Canu (which includes error correction, trimming and assembly); **3)** first polishing; **4)** circularizing; **5)** second polishing; **6)** annotation by Rapid Annotations using Subsystems Technology (RAST); and **7)** frame-shift correction. The final annotation

was performed by National Center for Biotechnology Information (NCBI) upon genome sequence submission. Details on the applied programs can be found in **Table S1**.

3.5.4 Genome comparisons

Protein FASTA files were generated using the GenBank files from NCBI and the BioPython version 1.77 (Cock *et al.*, 2009). Genome-wide comparisons were performed with protein Basic Local Alignment Search Tool (BLASTp+) version 2.10.0 locally with an Expect value (e-value) cut-off of 0.001 (Camacho *et al.*, 2009; Altschul *et al.*, 1997, 1990). The new *M. thermautotrophicus* ΔH and *M. marburgensis* Marburg sequences were compared to the previously published ones, accession numbers NC_000916.1 (Smith *et al.*, 1997) and NC_014408.1 (Liesegang *et al.*, 2010), respectively. Some gene-wise annotations were completed with the online BLASTp tool (Madden, 2003).

3.5.5 Clusters of orthologous genes (COG) functional annotation

The COG functional annotation was performed on the protein FASTA file for each microbe with the script cdd2cog v0.2 from the bac-genomics-scripts according to the author's instructions (Leimbach, 2016). A modification to the code as suggested in issue #14 (`https://github.com/aleimba/bac-genomics-scripts/issues/14`, accessed 11/2021) was applied to allow the use of the COG2020 database (Galperin *et al.*, 2021).

3.5.6 Genome-scale metabolic model reconstruction

An umbrella model for *M. thermautotrophicus* ΔH, *M. thermautotrophicus* Z-245 and *M. marburgensis* Marburg, which included all reactions from the three microbes, was built in Microsoft® Excel® (Microsoft 365 MSO, Version 2202, Washington, United States), following the protocol established by Thiele and Palsson (2010). The genome sequences, assemblies, and annotations were used as the backbone for the umbrella model. Reactions and pathways were added based on data from KEGG (Ogata *et al.*, 1999), ModelSEED (Seaver *et al.*, 2021), UniProt (Consortium, 2019), Brenda (Jeske *et al.*, 2019), BioCyc (Karp *et al.*, 2019), MetaCyc (Caspi *et al.*, 2020), BIGG (King

et al., 2015a), and NCBI (National Center for Biotechnology Information (NCBI)[Internet], 1988). Metabolite protonation was determined (for pH 7.0) using the commandline tool cxcalc and MarvinSketch 18.8.0, Chemaxon (`http://www.chemaxon.com`). When possible, reactions and genes were verified with literature. Genes for which no genus-specific evidence was found were BLAST searched to genes of species with stronger evidence. These BLAST results are specified in the comment section of the reconstruction. Over 790 references were cited for which the organism and type of evidence were recorded using Evidence & Conclusion Ontology (categories: biochemical, genetic, physiological, sequence, modeling, and no data), which were then used to determine the confidence of each reaction for each microbe (Thiele and Palsson, 2010). Published GEMs from the following microbes were used to gap fill and validate pathways in the Methanothermobacter umbrella model: **1)** Methanosarcina acetivorans: iVS941 (Kumar *et al.*, 2011), iMB745 (Benedict *et al.*, 2012) iMAC868a (Nazem-Bokaee *et al.*, 2016), iST807 (Peterson *et al.*, 2016); **2)** Methanosarcina barkeri: iAF692 (Feist *et al.*, 2006), iMG746 (Gonnerman *et al.*, 2013); **3)** Methanospirillum hungatei, iMhu428 (Hamilton *et al.*, 2015); **4)** Methanocaldococcus maripaludis S2, iMM518 (Goyal *et al.*, 2014); **5)** Methanobrevibacter smithii, iMsi385 (Shoaie *et al.*, 2013); **6)** Methanocaldococcus jannaschii, iTS436 (Selkov *et al.*, 1997). The umbrella model was converted to three strain-specific GEMs written in SBML Level 3 Version 1 and verified in MEMOTE 0.13.0 (Lieven *et al.*, 2020) (the GEMs were constrained with the experimental data that were adjusted with the maximum likelihood estimates). The GEMs aimed to be MIRIAM-compliant, including metabolite and reaction annotations with Compact Identifiers for various databases (gene annotations were not available given the newly annotated genomes, however the old gene annotations can be found in **Data S1**) (Le Novère *et al.*, 2005). Directionality of the reactions were determined using thermodynamics-based flux analysis (TFVA) (Mahamkali *et al.*, 2021) and the ModelSEED database (Henry *et al.*, 2010). Otherwise, directionality was set to reversible, except for reactions that caused loops with ATP (*e.g.*, with ATP losing a phosphate group (Thiele and Palsson, 2010)) or between redox carriers (Marcellin *et al.*, 2016). To verify the three strain-specific GEMs, flux balance analyses (FBAs; with maximization of biomass as an objective function) were run using semi-constrained models, to confirm the GEMs' ability to grow under certain growth conditions (**Data S2**). The FBAs were run using COBRApy version 0.22.1 (Ebrahim

et al., 2013) with various constraints of the GEMs (**Data S3**).

3.5.7 Biomass composition determination and maintenance energies

The biomass composition for the three microbes in the GEMs is assumed to be the same (**Data S1**). Briefly, the fraction for each of the molecules that make up the biomass was found by averaging those taken from previously published methanogen GEMs (Thor *et al.*, 2017), and the incorporation of empirically found data for *M. thermautotrophicus* ΔH (**Data S1**, reaction ID, BIOMASS_X, where X is DH, ZZ, or MM). The elemental composition of the biomass, $CH_{1.681}O_{0.418}N_{0.222}S_{0.004}$ (molecular weight of 23.502 g/mol) was taken from Duboc *et al.* (1995). Growth-associated maintenance (GAM) energy is included in the biomass macromolecule synthesis reactions. Non-growth-associated ATP maintenance costs (NGAM) are represented by using the ATP hydrolysis reaction (ATPM).

3.5.8 Bioreactor setup and operating conditions

Continuous cultures were carried out in the BioXplorer 100 bioreactor platform controlled with the WinISO ver. 2.3.149.1 software (H.E.L., London, England). Each bioreactor was equipped with temperature, pH (part number Z001013510), and ORP (part number Z061013510) sensors (I&L Biosystems GmbH, Königswinter, Germany); a 0.15 μm sparging stone; a magnetic coupled stirring system; three peristaltic pumps for media feed-in, base feed-in (1 N NaOH was used for pH control), and effluent-out; a mass-flow controller (Red-y smart min; Vögtlin, Muttenz, Switzerland) to control the inlet gas flow rate; a condenser for the exhaust-gas line; and a separate sampling and inoculum port (fitted with a rubber butyl stopper). The bioreactors were fitted with Masterflex® L/S Norprene Food-Grade Tubing, L/S 14 (Cole-Parmer GmbH, Wertheim-Mondfel, Germany) except for the gas inlet lines which were fitted with Masterflex® C-Flex ULTRA tubing, L/S 16 (Cole Parmer, Wertheim-Mondfel, Germany). The upstream gas mixture was set using Bronkhorst EL-Flow® Prestige mass flow controllers (Bronkhorst Deutschland Nord GmbH, Kamen, Germany) and mixed in a doubled-ended cylinder (Swagelok® Stuttgart, Reutlingen, Germany). The exhaust gas flow rate was measured offline using a MilliGascounter MGC-1 V3.4 PMMA (Dr.-Ing. RITTER Apparatebau GmbH & Co. KG, Bochum, Germany). The

pH and ORP sensors, the pumps, and the MFCs were calibrated before the run. The bioreactors were filled with the mineral medium and autoclaved for one hour at 121°C. Afterwards, the bioreactors were connected to the bioreactor platform and the temperature was set to 65°C with agitation at 700 rpm. Then, the bioreactors were sparged through sterile filters Minisart® HY (0.2 μm pore size; Satorius AG, Göttingen, Germany) for two h with H_2/CO_2 (80/20% v/v) at a gas flow rate of 10 mL/min. Before inoculation, the mineral medium was reduced with sterile anaerobic L-cysteine-HCl (0.5 g/L) and disodium sulfide nonahydrate (0.3 g/L), and the pH control set to 7.3. Each bioreactor was inoculated with either 4 mL (*M. thermautotrophicus* ΔH and *M. thermautotrophicus* Z-245) or 3.6 mL (*M. marburgensis* Marburg) of preculture grown in serum bottles to an OD_{600} of 0.35-0.36. The bioreactors were operated in batch mode for one day until an OD_{600} of approximately 1.00 was reached at which point continuous mode was started with a medium feed at a dilution rate of 0.83 d^{-1}. The first continuous operating period was conducted for approximately 12 days (**Figure A.4**), after which the bioreactors were emptied except for about 3-5 mL, which were used as inoculum for a second period with a starting OD_{600} of 0.2-0.3. When an OD_{600} of approximately 1.00 was reached, the bioreactors were switched into continuous mode again with a medium feed at a dilution rate of 1.11 d^{-1}. Steady state was reached after three hydraulic retention times (HRT) (Jensen, 2001), which was 2.7 days in our setup. After an additional 3 HRTs (day 6.8), the samples for the transcriptomics, proteomics, and gram cell-dry weight determination were taken.

During both periods, daily samples of OD_{600}, pH, exhaust gas flow rate, and inlet and exhaust gas composition were taken. For the liquid culture samples, 1 mL of dead volume was first removed, then 1 mL was used for OD_{600} and pH measurements. To measure pH, the samples were maintained at 65°C (Thermomixer 460-0223, Eppendorf, Hamburg, Germany) and measured within 1 min of taking the sample (FiveEasy™ Plus pH/mV Benchtop meter with the micro pH electrode LE422 (Mettler-Toledo GmbH, Gießen, Germany) calibrated at 25°C and set to 65°C).

3.5.9 Cross-contamination check

The bioreactors were checked for cross-contamination by polymerase chain reaction (PCR) using the Thermo Scientific™ Phire Plant Direct PCR Master Mix (Thermo Fischer Scientific, Dreieich, Germany) and custom primers (**Table S2**). DNA was

extracted by boiling 300 μL of the bioreactor culture (with an approximate OD_{600} of 1.0) at 100°C for 10 minutes (Thermomixer 460-0223, Eppendorf, Hamburg, Germany); 1 μL was directly taken as template for the PCR reaction. Primers were used at a 10 μM concentration. The PCR cycle was carried out 28 times in an Mastercycler® pro S (Eppendorf, Hamburg, Germany). The results of the reactions were separated and visualized with gel electrophoresis (1% w/v agarose and SYBR™ Safe DNA Gel Stain (Thermo Fischer Scientific, Dreieich, Germany) and Gel Doc™ XR+ visualizer (Bio-Rad, Feldkirchen, Germany).

3.5.10 Fermentation gas analysis

A 490 Micro Gas Chromatograph (microGC; Agilent, Santa Clara, United States) fitted with a multi-valve port system (Teckso GmbH, Neukirchen-Vluyn, Germany) was used to analyze the inlet and outlet gas compositions. The microGC was equipped with two columns, the Molecular Sieve 5A PLOT 0.25 mm, 10 m (Agilent, Santa Clara, United States) that used Argon as a carrier gas to measure H_2, O_2, N_2, CH_4 and CO and the PoraPLOT Q PLOT, 0.25 mm, 10 m (Agilent, Santa Clara, United States) that used Helium as a carrier gas to measure CO_2, N_2 combined with O_2, and H_2S. The microGC was calibrated before the run using six calibration levels (**Table S3**). Each level was sampled for four replicates and the average was taken as the calibration point. The total method lasted 180 seconds with a sample time of 20 seconds. The injector and sample line temperature were set at of 110°C, and column temperatures and pressures at 60°C and 150 kPa, respectively.

3.5.11 Biomass concentration analysis

The biomass correlation coefficient (K in g/L/OD_{600}) as defined in Valgepea *et al.* (2017a) was found by sampling 60 mL from each bioreactor (6 Falcon™ tubes of 10 mL) at the end of each steady-state period. The samples were centrifuged, the supernatant removed, and two Falcon™ tubes combined with 0.5 mL of Millipore water in pre-weighed glass vials (548-0028; VWR International GmbH, Darmstadt, Germany), resulting in three technical replicates per bioreactor. The vials were dried at 200 mbar (absolute pressure) and 80°C for three days, and the weight of the biomass was recorded. The slope of the measured biomass weight to the respective OD_{600} was

taken and divided by the volume (0.02 L), to give the following K values (g/L/OD_{600}): *M. thermautotrophicus* ΔH, 0.320; *M. thermautotrophicus* Z-245, 0.305; and *M. marburgensis* Marburg, 0.275.

3.5.12 Carbon balance calculation

Initial carbon balances using molar flow rates ($\dot{n}$, mmol/h) were calculated according to conservation of mass (**Equation (3.2)**) using the ideal gas law (**Equation (3.3)**) with the following assumptions: the ideal gas constant of 8.3144 L·kPa/K/mol, a temperature (T) of 25°C (298.15 K), and a pressure (P) of 1 atm (101.325 kPa). The volumetric flow rate ($\dot{V}$, L/h) was found by multiplying the gas concentration (the fraction measured by the MicroGC) by the measured gas flow rate. This was done for the CO_2 of the inlet gas (CO_{2,gas_in}), and the CO_2 (CO_{2,gas_out}) and CH_4 (CH_{4,gas_out}) of the outlet gas. In addition, to account for the soluble CO_2 (CO_{2,aq_out}) that was removed from the bioreactors in the effluent, Henry's law was used. To find the quantities of soluble bicarbonate and carbonate in the medium, the acid dissociation constants (and pKa) were found using the equations provided by Prieto and Millero (2002) (**Data S4**). Salinity of the medium (4.565 ppt), required for the equations in Prieto and Millero (2002), was estimated by summing the amount of cation and anions of the salts added to medium. The estimated concentrations of solubilized and removed carbon species were subtracted from the calculated gas uptake values. Lastly, to calculate the biomass concentration in the bioreactor ($biomass_{aq\_outlet}$), the biomass correlation coefficient (K in g/L/OD_{600}) was divided by the measured bioreactor OD_{600} and volume. The resulting g_{CDW} was divided by the molecular weight (23.502 g/mol) (Duboc *et al.*, 1995) and multiplied by the measured dilution rate of the bioreactors (h^{-1}). Given inconsistencies in carbon balances in our measurements, a gross measurement error analysis was performed following the macroscopic balance method described in (Wang and Stephanopoulos, 1983) (**Data S3**). Gas and biomass data at time points with suspected gross measurement error were dropped for the carbon balance calculations.

$$\begin{aligned} 0 &= C_{in} - C_{out} \\ 0 &= CO_{2,gas\_in} - (CO_{2,gas\_out} + CO_{2,aq\_out} + CH_{4,gas\_out} + biomass_{aq\_out}) \end{aligned} \tag{3.2}$$

$$\dot{n} = \frac{P\dot{V}}{RT} \tag{3.3}$$

3.5.13 Normalized product distribution

Product distributions were calculated using CH_4 and biomass data (mmol/h) for bioreactors and time points that did not have suspected gross measurement error. The quantity of CH_4 or biomass was divided by the sum of the CH_4 and biomass, and then multiplied by 100 resulting in the normalized product distribution percentage. An analysis of variance (ANOVA) test was conducted in Excel® using the ANOVA: Single Factor ($\alpha = 0.05$) function of the Data Analysis Addin to analyze statistically significant differences between the product distribution ratios of the three different microbes.

3.5.14 Interspecies comparison

All gas and biomass data points without suspected gross measurement error were used for the comparative analysis (Independent T-tests) between the three microbes, expect for one bioreactor replicate for *M. thermautotrophicus* Z-245 that washed out during the fermentation. The T-tests (two-tailed and confidence internals of 95%) were carried out in Python 3.6.13 using the Researchpy package Version 0.3.2 and the ttest function (Bryant, 2018). Three-way comparisons were performed in Excel® using the ANOVA: Single Factor ($\alpha = 0.05$) function of the Data Analysis Addin. The maximum likelihood estimates of the measurements for time points without gross measurement error were calculated (Wang and Stephanopoulos, 1983), resulting in closed carbon balances for those time points (assuming 90% confidence intervals for the test function). These adjusted values were then averaged across time points for each microbe and used to constrain the GEMs (**Data S4; Materials and Methods, 3.5.12, 3.5.20**).

3.5.15 Transcriptomics

RNA sample preparation

When the steady state was reached, four-times 9 mL (quadruplicates) of bioreactor sample was placed into 5 mL of prechilled RNAlater® (overnight at4°C). The samples were stored overnight at 4°C and then frozen at -20°C until RNA isolation. The samples

were thawed on ice and centrifuged at 4°C and 4100 x g for 10 min (Multifuge X3R TX-1000, Fischer Scientific, USA), the supernatant discarded, and the samples resuspended in 800 μL of RNase free water. The 800 μL were mixed with 950 μL of saturated phenol (Sigma-Aldrich, Merck, Germany) and 115 μL of a lysis solution in Lysing Matrix B (MP Biomedicals Germany GmbH, Eschwege, Germany). The lysis solution contained, sodium acetate (20 mM pH 5.2), sodium dodecyl sulfate (SDS; 0.5% v/v), ethylenediaminetetraacetic acid (EDTA; 1 mM), and DNase/RNase-free distilled water (DI; Invitrogen, Thermo Fischer Scientific, Germany). The cells were homogenized for one cycle (5 x 40 s at 6 m/s and 20 s off) in a FastPrep-24^{TM} 5G bead-beater (MP Biomedicals Germany GmbH, Eschwege, Germany) and then centrifuged at room temperature and 21130 x g for 10 min (5424 Eppendorf, Hamburg, Germany). The top layer was placed into 600 μL of phenol-chloroform-isoamyl alcohol (ROTI® Aqua-P/C/I, Carl Roth, Karlsruhe, Germany) and centrifuged as in the previous step. This step was repeated twice. To precipitate the RNA, the top layer was added to 1.2 mL of ethanol (undenatured absolute, SERVA Electrophoresis GmbH, Heidelberg, Germany) and left overnight at -80°C. The following day, the tubes were centrifuged at 4°C and 21130 x g for 10 min, and then the supernatant removed. The RNA pellet was washed with 1 mL of 75% v/v ethanol solution and centrifuged as in the previous step. The supernatant was pipetted off, and the pellet resuspended with 53 μL of DNase/RNase-free distilled water. The RNA was cleaned (which included DNA depletion) and concentrated using the RNA Clean & Concentrator kit (Zymo Research, Irvine, CA, United States), according to the manufacturer's instructions. The cleaning and concentrating protocol was repeated four times. The quantity (6-16 μg) and quality (RNA integrity index $>$ 9) of the purified RNA was measured with a 2100 Agilent Bioanalyzer (Agilent Technologies, Santa Clara, United States) using the Agilent RNA 6000 Nano Kit (Agilent Technologies, Santa Clara, United States) before freezing the samples at -80°C until sequencing.

RNA sequencing

The library preparation was performed using 100 ng of RNA and the TruSeq^{TM} Stranded Total RNA Kit with Ribo-Zero^{TM} Plus (Illumina, San Diego, United States). Pair-ended sequencing was performed using the NovaSeq^{TM} 6000 with the Flow Cell Type 2 x 100 bp (Illumina, San Diego, United States). Demultiplexing of the sequences was performed

with Illumina bcl2fastq (2.20) software.

Raw sequencing data analysis

Sequencing reads were quality filtered and trimmed with BBMap version 38.93 (Bushnell, 2014) and FastX Toolkit version 0.0.14 (Gordon and Hannon, 2017) with specific parameters for paired-end reads. Reads from each bioreactor culture were mapped to the reference genomes of *M. thermautotrophicus* ΔH (CP064324), *M. thermautotrophicus* Z-245 (CP064336 and CP064337) and *M. marburgensis* Marburg (CP069376 and CP069377) using BWA-MEM version 0.7.17 (Li, 2013). Mapped reads were assigned to genomic features using the FeatureCounts program from the Subread package version 2.0.1 (Liao *et al.*, 2014). Additionally, an alignment-free quantification of gene features was performed using Salmon (Patro *et al.*, 2017). A pairwise differential expression analysis was performed using DESeq2 version 1.32.0 (Love *et al.*, 2014) with the homologous genes present in the pan-genome database (**Materials and Methods, 3.5.17**) and ConsensusDE version 1.10.0 (Waardenberg and Field, 2019). This process was orchestrated using SnakeMake version 6.8.0 (Mölder *et al.*, 2021) to achieve reproducibility (**Data S3**). The output files from Salmon were used for analysis of each microbe individually, and only genes with at least 2 non-zero transcript values were considered. The averages of each gene were then used.

3.5.16 Label-free proteomics

Protein sample preparation

The proteomics sampling and sample preparation method from (Valgepea *et al.*, 2018) was used with the following modifications. Screw-cap 2-mL tubes with lysing matrix B (MP Biomedicals Germany GmbH, Eschwege, Germany) were used in the homogenizer (FastPrep-24^{TM} 5G, MP Biomedicals Germany GmbH, Eschwege, Germany). For cell lysis the one cycle of "bead beating" as described for transcriptomics sample preparation was repeated three times. The following protein precipitation method from the Proteome Center Tübingen (PCT) at the University of Tübingen was applied. The supernatant was transferred to a 15-mL Falcon^{TM} tube with 4 mL and 0.5 mL of ice-cold 100% acetone and 100% methanol, respectively, and precipitated overnight at -20°C. The following day, the tubes were centrifuged at 2200 x g and 4°C for 20 minutes (Multifuge X3R

TX-1000, Thermo Fischer Scientific, Waltham, United States). The supernatant was removed, and the pellet washed with 1 mL of ice-cold acetone (80/20% v/v in water) and centrifuged as in the previous step. The supernatant was removed, and the protein pellet was left to air dry on ice for 15 min before freezing at -20°C.

Protein measurement

Pellets were resuspended in a denaturation buffer (6 M urea, 2 M thiourea in 10 mM Tris pH 8.0) and 20 μg of protein were subjected to tryptic in-solution digestion (0.2 μg trypsin). The samples were run for an LC-MS/MS analysis on an Easy-nLC^{TM} 1200 (Thermo Fisher Scientific, Dreieich, Germany) coupled to a Orbitrap Exploris^{TM} 480 mass spectrometer (Thermo Fisher Scientific, Dreieich, Germany). The LC-MS analysis was performed as described previously (Fagbadebo *et al.*, 2021) with the following exception: a 113 min linear gradient was used to separate peptides on the LC-MS.

Protein analysis with proteome discoverer

Raw data dependent acquisition (DDA) data from LC-MS/MS analysis was performed using Thermo Fisher Proteome Discover^{TM} software (version 2.5.0.400, Thermo Fisher Scientific, Scoresby, Australia). Files were processed using two different methods. For intensity based absolute quantification (iBAQ) analysis, each strain was processed individually through the precursor label-free quantification workflow with their corresponding FASTA files and then converted to iBAQ values as described in Schwanhäusser *et al.* (2011). For differential expression analysis, all files were processed together in the one study using the custom pan-genome FASTA file (Materials and Methods, 3.5.17). Both methods used the same workflows and only differed by their choice of FASTA file. Briefly, spectra were searched against the respective databases with carbamidomethylation set as a fixed modification and up to 3 methionine oxidations. Acetylation, methionine loss, and acetylation after methionine loss at the N-terminus were also permitted. The mass tolerance of precursors and fragments were 10 ppm and 0.02 Da respectively. The minimum and maximum peptide lengths were six and 144 amino acids respectively. Two missed cleavages were allowed per peptide. Proteins were filtered to a 1% FDR cut-off threshold. For quantification, only peptides unique to a protein group were used. All other parameters were kept at their default settings.

3.5.17 Pan-genome differential expression database creation

$$(ZZ \cup MM) \cap DH \tag{3.4}$$

With the *M. thermautotrophicus* ΔH genes as a reference, an intersection of homologous genes/proteins based on the best BLASTp+ hits from *M. thermautotrophicus* Z-245, and *M. marburgensis* Marburg were used (**Equation (3.4)**). *M. thermautotrophicus* ΔH and *M. thermautotrophicus* Z-245 had 81 additional homologous genes/proteins, which were not considered when *M. marburgensis* Marburg was being analyzed, while *M. thermautotrophicus* ΔH and *M. marburgensis* Marburg had 11. This pan-genome of homologous genes groups can be found in **Data S5**.

To simulate *in-silico* tryptic digestion in Proteome Discoverer^{TM}, a protein FASTA file that represented the mapping across gene groups was required. To account for genetic heterogeneity that leads to variation in protein sequences across microbes and not favor one microbe, peptide sequences from the other microbes had to be added to the gene-group sets (Broadbent *et al.*, 2016). To retrieve all peptides for each microbe, an *in-silico* tryptic digestion of the protein FASTA files for *M. thermautotrophicus* ΔH, *M. thermautotrophicus* Z-245, and *M. marburgensis* Marburg was performed using Rapid Peptides Generator version 1.2.4 (Maillet, 2020). A new enzyme was defined to cleave after lysine (K) or arginine (R) except if proline (P) follows: cleaving rule as (K or R,) with the exception (K or R,)(P). Then, using the *M. thermautotrophicus* ΔH protein FASTA file, unique peptides (with a length greater than six amino acids) from homologous proteins of *M. thermautotrophicus* Z-245 and *M. marburgensis* Marburg that were not already in the *M. thermautotrophicus* ΔH protein were added (**Data S3**). This was achieved by adding these peptides to the FASTA file for the respective protein. Importantly, these unique peptides were added to the protein before the C-terminal peptide, which guaranteed the following computational steps to recognize the combined protein correctly. If *M. thermautotrophicus* Z-245 and *M. marburgensis* Marburg shared a C-terminus different from the one of *M. thermautotrophicus* ΔH, then that C-terminus was additionally amended as the C-terminus for the new combined protein, otherwise the C-terminus from *M. thermautotrophicus* ΔH was kept. This new protein FASTA file was then used for the proteomics analysis (**Materials and Methods, 3.5.16**).

3.5.18 Eha/Ehb ratio determination

The ratio of the energy-converting hydrogenases Eha and Ehb subunits was found by summing the abundances (either transcripts or relative protein abundances) of the subunits detected for each energy-converting hydrogenase and dividing those sums by the number of subunits found. This was done to normalize against subunits that were not found. The normalized abundance of Eha was then divided by that of Ehb. This ratio was then averaged across the replicates (bioreactor samples) of each microbe. An ANOVA test was conducted in Excel® using the ANOVA: Single Factor ($\alpha = 0.05$) function of the Data Analysis Addin to test if there was a statistically significant difference between the Eha/Ehb ratios for the three different microbes (**Data S5**).

3.5.19 Methanogenesis relative abundances

Gap-filled relative abundances were found for the different protein (complexes) involved in the Wolfe cycle of methanogenesis. The iBAQ values of each subunit were summed and divided by the number of detected subunits (the maximum number of subunits missing per microbe per protein complex was one). This sum was then multiplied by the theoretical number of subunits that should be found (effectively assigning the non-detected subunit the average iBAQ value for the complex) to produce the total protein complex normalized iBAQ value. This value was then divided by the sum of all iBAQ values in the proteome and multiplied by 100, resulting in relative (normalized) protein abundance percentages (**Data S5**).

3.5.20 Integrating fermentation data in the GEMs

To validate the three strain-specific GEMs, experimental fermentation data was used to constrain the models for the FBAs (with biomass maximization as the objective function). The data did not consist of time points with suspected gross measurement error and was adjusted with the maximum likelihood estimates. These FBAs were run using COBRApy version 0.22.1 (Ebrahim *et al.*, 2013) (**Data S3**).

From the GEMs, constrained with adjusted experimental values, GIMME (Becker and Palsson, 2008) with the transcriptomics and proteomics data were applied to create reduced models (thresholds set at the lower quartile). This was done in the COBRA Toolbox v.3.1 (Heirendt *et al.*, 2019) using MATLAB (R2018b) and the Gurobi

Optimizer v.9.0.1 (Gurobi Optimization, LLC, 2020). Also in the COBRA Toolbox, FBAs (with and without loops allowed) were run on the resulting reduced models, using either the maximization of biomass exchange (EX_Biomass_e) or the maximization of ATP dissipation (ATPM) as the objective function (**Data S3**).

3.5.21 Data availability

The **genome** sequences, genome annotations, and methylation patterns were deposited to NCBI (National Center for Biotechnology Information (NCBI)[Internet], 1988), and after release can be found under Bioproject ID PRJNA674001 (**Table S4**). The **transcriptomics** data (gene expression data) will be deposited to the NCBI's Gene Expression Omnibus. The **proteomics** DDA data will be deposited to the ProteomeXchange Consortium *via* the PRIDE partner repository (Perez-Riverol *et al.*, 2019; Deutsch *et al.*, 2016, 2020). The **modeling** related files will be deposited in an Open Modeling EXchange format (OMEX) (Bergmann *et al.*, 2014) to BioModels (Glont *et al.*, 2018; Malik-Sheriff *et al.*, 2020) and currently can be found in **Data S1** at `https://github.com/isacasini/Dissertation_Casini/tree/main/Chapter_3/DataS1`. The GEM was provided in the following formats: Excel® model (heavily commented), .xml, .json, and .mat. Additional files include: the benchmarking reports (.html) from MEMOTE (Lieven *et al.*, 2020), the map files for visualization (.json) with Escher (King *et al.*, 2015b), the gas fermentation data used to constrain the model (as Excel® files), and the individual microbe and differential expression analyses data for both the transcriptomics and proteomics, which was used to constrain the model (as Excel® files). Additional scripts and programs are also available on GitHub at the following repository link: `https://github.com/isacasini/Dissertation_Casini/tree/main/Chapter_3/DataS3`.

3.6 Results and discussion

3.6.1 Updated genome sequences for *Methanothermobacter* spp. provide the basis for high-quality genome-scale metabolic reconstructions

Methanothermobacter spp. are considered promising biocatalysts for power-to-gas applications (Pfeifer *et al.*, 2021). Here, we compared three different *Methanothermobacter* spp. with an integrated systems-biology approach: **1)** *M. thermautotrophicus* ΔH; **2)** *M. thermautotrophicus* Z-245; and **3)** *M. marburgensis* Marburg (Wasserfallen *et al.*, 2000). For *M. thermautotrophicus* ΔH and *M. marburgensis* Marburg, genome sequences were available (Smith *et al.*, 1997; Liesegang *et al.*, 2010), but for *M. thermautotrophicus* Z-245, the genome had not been fully sequenced to date.

First, we conducted independent de-novo sequencing and genome assembly for *M. thermautotrophicus* ΔH, *M. thermautotrophicus* Z-245, and *M. marburgensis* Marburg with long-read sequencing technology (**Table 3.1**, **Table S4, Materials and Methods, 3.5.2, 3.5.3**). This provided high-quality genome sequences including the species-specific methylation pattern of the three microbes (**Text A.1.1, Table S5**). In our genome annotation, we assigned COG functional annotations to 1501/1796, 1505/1804, and 1440/1730 genes for *M. thermautotrophicus* ΔH, *M. thermautotrophicus* Z-245, and *M. marburgensis* Marburg, respectively (**Table S6**). We identified several differences between the old and our new sequences for *M. thermautotrophicus* ΔH and *M. marburgensis* Marburg (**Table 3.1**), which we considered for the GEM reconstruction as described below. These differences are likely due to advances in sequencing methods and in assembly and annotation algorithms over the last few decades (*M. thermautotrophicus* ΔH was sequenced in 1997 (Smith *et al.*, 1997) and *M. marburgensis* Marburg in 2010 (Liesegang *et al.*, 2010)). For sequencing, we used strains directly from the DSMZ without excessive subculturing (**Materials and Methods, 3.5.1**). However, we cannot exclude differences due to lab-culture adaptation.

Specifically, our genome annotation led to changes in genes with implications on metabolic functions (**Table S7**). For example, the two homologs of adenosine monophosphate (AMP)-forming acetyl-CoA synthetase (ACS) were split into two open reading frames (ORF) in the old *M. thermautotrophicus* ΔH genome annotation

(MTH217-MTH216 and MTH1603-MTH1604) (**Table 3.2**). In our new annotation, both of the two sets of genes map to only one ORF (ISG35_00975 and ISG35_07685). In the first set, there remains an in-frame stop codon located in the middle of the gene (at amino acid number 559), which is noted with the comment "internal stop" in the GenBank file (in total, four genes in the genome have this note). This stop codon likely renders this ACS non-functional in *M. thermautotrophicus* ΔH, as hypothesized before by Ingram-Smith and Smith (2007). However, in the second set, the new gene annotation (ISG35_07685) does not contain an in-frame stop codon anymore, also confirming the error in the old genome sequence that Ingram-Smith and Smith (2007) had found. While Ingram-Smith and Smith (2007) were unable to characterize this second ACS, our omics data revealed that it was both transcribed and translated and is assigned to a reaction in the GEM (**Data S5, S4**). Furthermore, a putative acetate transporter is annotated in both the old (MTH215) and new (ISG35_07680) sequence, and we find this gene to be transcribed and translated (**Table 3.2**). It remains to be elucidated experimentally, what implications these findings have for the acetate metabolism. Overall, the genome sequences with the updated annotations provided a high-quality basis for the construction of our GEMs.

Table 3.1 Summary of sequenced *Methanothemobacter* spp. genomes.

Microbe	Plasmid	Length (bps)	GC %	Gene Count	CDS	Previous Sequences Accensions	Difference with Old Sequence	Similarities with Old Sequence
M. thermautotrophicus ΔH	N/A	1,751,429	49.56	1844	1796	NC_000916.1[1]	448 SN[4] 247 IB[5] 91 O2N[6] 92 N2O[7]	1733 CDS
M. thermautotrophicus Z-245	pFZ1	1,758,79 11,0155[8]	49.46 42.515[8]	1840 125[8]	1792 125[8]	GCA_013330715.16[2,9]	N/A	N/A
M. marburgensis Marburg	pM2001 (pMTBMA4)	1,634,705 4,4415[8]	48.65 45.405[8]	1774 55[8]	1725 55[8]	NC_014408.1[3]	68 SN 30 IB 25 O2N 66 N2O	1675 CDS

[1] (Smith *et al.*, 1997)
[2] (Rinke *et al.*, 2021)
[3] (Liesegang *et al.*, 2010)
[4] SN, single nucleotides
[5] IB, indel bases
[6] O2N, old genes that do not map to any new genes
[7] N2O, new genes that do not map to any old genes
[8] indicates values for the plasmid
[9] sequence was from a metagenome and the assembly only to scaffold level

Table 3.2 New and old acetyl-CoA synthetase/acetate-CoA ligase genes. The comparison of the acetate transporter and acetyl-CoA synthetase/acetate-CoA ligase genes in the old and new genome sequences, and their presence in the transcriptomics and proteomics data of *M. thermautotrophicus* ΔH. The position in the top (abundance range) genes or proteins is also listed.

Old gene	New gene	Function	Present transcriptomics (place within the top genes)	Present proteomic (place within the top proteins)	Comments
MTH215	ISG35_00970	Acetate Transporter	Yes, gene 278	Yes (218)	Reaction ACt2r
MTH216	ISG35_00975	Acetyl-CoA synthetase (ACS)/acetate-CoA ligase	Yes (237)	No	Unknown if the transcripts are only from the half of the gene before the stop codon
MTH217	ISG35_00975	Acetyl-CoA synthetase (ACS)/acetate-CoA ligase	Yes (237)	No	Unknown if the transcripts are only from the half of the gene before the stop codon
MTH1603	ISG35_07685	Acetyl-CoA synthetase (ACS)/acetate-CoA ligase	Yes (779)	Yes (726)	Reaction ACS in the GEM
MTH1604	ISG35_07685	Acetyl-CoA synthetase (ACS)/acetate-CoA ligase	Yes (779)	Yes (726)	Reaction ACS in the GEM

3.6.2 GEM reconstruction results in the highest genome coverage of current methanogen models

We intended to compare physiological differences between the three microbes on a systems biological level. Therefore, based on our *de-novo* genome sequences and the respective genome annotations, we reconstructed GEMs. To streamline a comparison between our three model microbes, we constructed an umbrella model that integrates the GEM reconstruction for all three microbes. This umbrella model consists of 618 reactions (including 46 exchange reactions, 56 transport reactions, and seven biomass reactions, with 11 transport and 11 exchange reactions that act as pseudo reactions for orphan metabolites), 555 metabolites, and 509-526 genes (depending on the microbe) (**Data S1**). This reflects 29.3% (1796, *M. thermautotrophicus* ΔH), 28.8% (1804, *M. thermautotrophicus* Z-245), and 29.4% (1730, *M. marburgensis* Marburg) of the protein-coding genes, which is the highest coverage for archaea and specifically methanogens compared to published GEMs (Thor *et al.*, 2017). Based on the umbrella model, we derived strain-specific GEMs for each of the three microbes, which we used for further modeling purposes as described below (**Data S1**). We tested the GEMs in MEMOTE and each achieved a score of 85% (**Data S1**)(Lieven *et al.*, 2020), and produced metabolic pathway maps to visualize the GEMs using Escher (**Data S1**) (King *et al.*, 2015b).

3.6.3 *M. thermautotrophicus* ΔH performs significantly different from *M. thermautotrophicus* Z-245 and *M. marburgensis* Marburg based on fermentation data

To compare the metabolism of the three microbes and to produce statistically relevant data to constrain the GEMs, we operated quadruplicate chemostat bioreactors (N=4) with the three microbes in pure culture under the same growth conditions in one experimental run. For *M. thermautotrophicus* Z-245, we discarded one replicate (N=3) due to a pump malfunction in this bioreactor and wash-out of the cells (**Materials and Methods, 3.5.8**). We performed PCR analyses with strain-specific primer pairs, which excluded cross-contamination between the bioreactors (**Figure A.1, Materials and Methods, 3.5.9**). Under steady-state growth conditions, *M. thermautotrophicus* ΔH had consumption rates for molecular hydrogen of 234.36 ± 40.76 mmol/g_{CDW}/h and carbon

dioxide of 58.18 ± 11.49 mmol/g_{CDW}/h, with a production of 52.50 ± 8.53 mmol/g_{CDW}/h methane and 0.03 ± 0.01 g_{CDW} biomass (**Figure 3.2A, 3.2B**, **Table S8**). This was a significantly higher consumption rate for molecular hydrogen (1.59-fold and 1.72-fold) and carbon dioxide (1.69-fold and 1.72-fold), as well as production rate for methane (1.62-fold and 1.82-fold), compared to both *M. thermautotrophicus* Z-245 and *M. marburgensis* Marburg (**Figure 3.2A, 3.2B**, **Data S4**). In contrast, biomass concentration of 0.04 ± 0.01 g_{CDW} was significantly (1.36-fold) higher for *M. marburgensis* Marburg compared to the other two microbes (**Figure 3.2A, 3.2B**, **Data S4**). This resulted in differences in the normalized distribution of products with the highest methane-to-biomass ratio of 96.4 ± 0.57 for *M. thermautotrophicus* ΔH and the lowest of 93.5 ± 0.98 for *M. marburgensis* Marburg under our experimental conditions (**Figure 3.2C**).

We did not attempt to maximize methane production rates or achieve high methane partial pressures in the product gas in our experiments, but instead, to produce statistically sound replicate data sets under identical experimental conditions for the three microbes. We here applied a volume gas per volume bioreactor per minute (vvm) of 0.08, while others on average applied a much higher vvm of 2.38 (Pfeifer *et al.*, 2021). Past studies have shown that there is a tradeoff between achieving high methane production rates and high methane partial pressure in the product gas, and have worked on improving these metrics by altering fermentation conditions (Pfeifer *et al.*, 2021). Thus, it is important to consider the fermentation conditions of our study when interpreting the interspecies comparison.

3.6.4 Interspecies comparison of multi-level omics under steady-state growth conditions reveals different gene-expression patterns

To assess the bioreactor experiments in more depth, we performed transcriptomics and proteomics analyses during steady state for all replicate bioreactors (**Data S5**). We achieved high reproducibility of the replicates for both transcriptomics and proteomics, and high coverage of the transcriptome (94.18-99.77%) and proteome (78.29-79.82%) with each microbe (**Table 3.3A**). The high coverage of both the transcriptome and the proteome indicates that the density of coding sequences in the genomes of the microbes

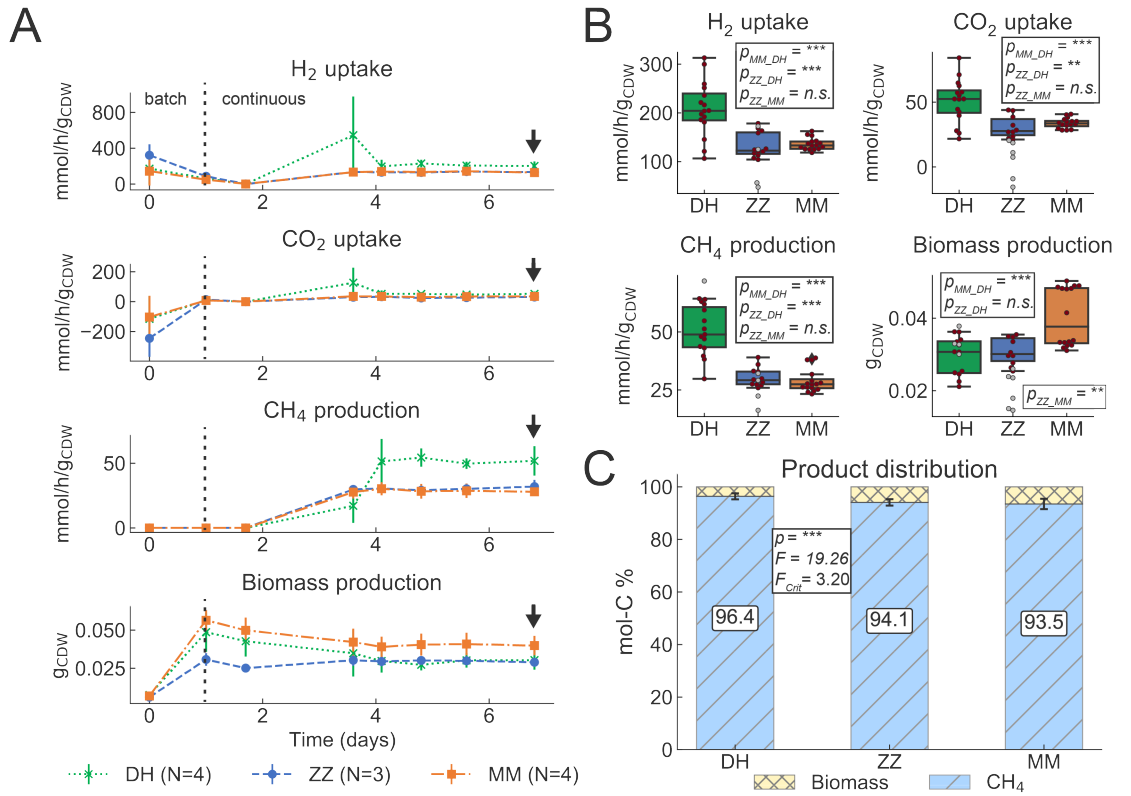

Figure 3.2: Fermentation data from bioreactors with *M. thermautotrophicus* ΔH (DH), *M. thermautotrophicus* Z-245 (ZZ), and *M. marburgensis* Marburg (MM). (A) Gas consumption (H_2 and CO_2 uptake), and CH_4 and biomass production data from quadruplicate (DH and MM) and triplicate (ZZ) bioreactors for the fermentation period of 7 days. CO_2 gas data at time point zero is negative due to estimated soluble carbon species adjustments. Data for further analyses (transcriptomics, proteomics) were taken on day 7 as indicated by arrows. **(B)** Average gas consumption (H_2 and CO_2 uptake), and CH_4 and biomass production data during steady-state period (days 4 to 7) For statistical analysis in pair-wise comparisons with T-test, data points without suspected gross measurement error (red circles) were included, data points with suspected gross measurement error (gray circles), were excluded (**Materials and Methods, 3.5.12**). **(C)** Average normalized product distribution, including statistical analysis by ANOVA (**Materials and Methods, 3.5.13**. DH, *M. thermautotrophicus* ΔH; ZZ, *M. thermautotrophicus* Z-245; MM, *M. marburgensis* Marburg; ***, p <0.0001; **, p <0.001; n.s., not significant (p >0.05); F, F value; F_{crit}, F critical value.

is further reflected on the transcript and protein level. The numbers of identified proteins in our study are similar to those found for *M. thermautotrophicus* ΔH in a previous study by (Liu *et al.*, 2019). We did not find a linear correlation between the differential expression analysis of the transcriptome and proteome on an individual species level (**Figure A.2**), which is an observation made before with other microbes (Becker and Wittmann, 2018).

To compare the omics data sets between microbes, we merged the underlying genomes into pan-transcriptome and pan-proteome databases (gene groups, **Data S5**), respectively (**Materials and Methods, 3.5.17**). This allowed us to perform differential expression analyses for the pairwise comparison of *M. marburgensis* Marburg versus *M. thermautotrophicus* ΔH (MM/DH), *M. thermautotrophicus* Z-245 versus *M. thermautotrophicus* ΔH (ZZ/DH), and *M. thermautotrophicus* Z-245 versus *M. marburgensis* Marburg (ZZ/MM). We found many differentially expressed genes in the pan-transcriptome, but much fewer differentially abundant proteins in the pan-proteome comparison, and in both omics sets, fewer differences between *M. thermautotrophicus* ΔH and *M. thermautotrophicus* Z-245 than to *M. marburgensis* Marburg (**Table 3.3B, Figure 3.3C-G**, **Table S9, Data S5**). In addition, the 100 most highly expressed genes on the transcript level share few overlaps (**Figure 3.3A**), while we found a high overlap of the 100 most abundant proteins between microbes (**Figure 3.3B**). This was also reflected in the pathway distribution when mapping these highly expressed genes or abundant proteins to the GEMs (**Table S10, Table S11**). Overall, even though the conditions in the bioreactor experiments were identical for all three closely related species, the transcriptome and proteome composition between the microbes revealed significant differences.

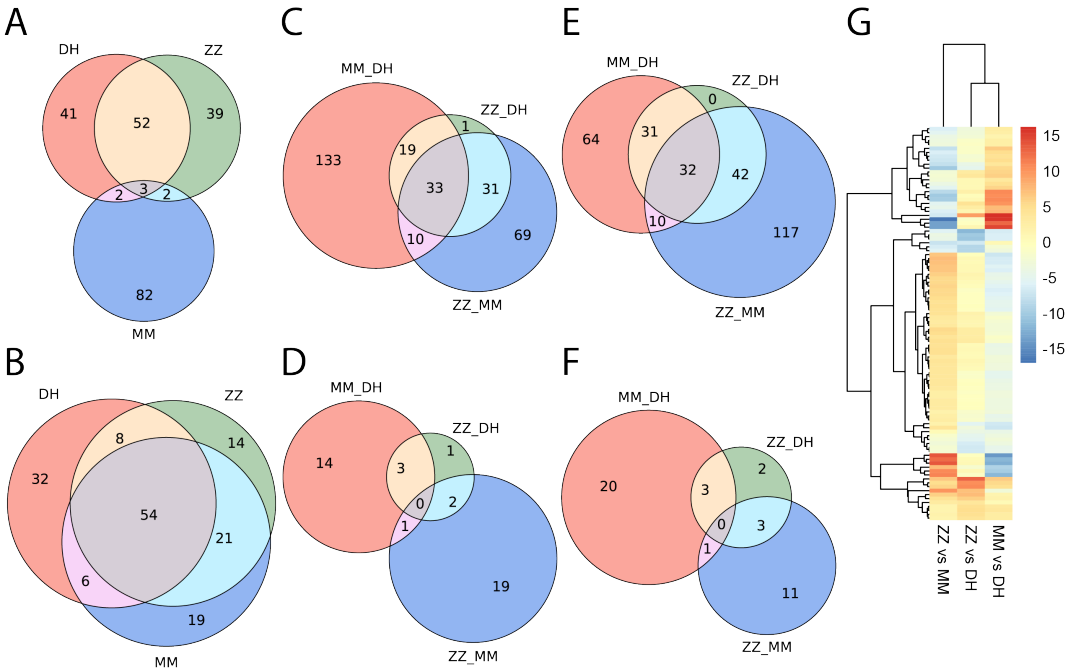

Figure 3.3: Overview of transcriptomics and proteomics interspecies comparison. **(A)** Comparison of the top 100 transcribed genes (by average number of transcripts per million). **(B)** Comparison of the top 100 abundant proteins (by average iBAQ value). Transcripts/proteins that do not map to a gene group are counted only once collectively, hence there are less than 100 genes total in A and B. **(C)** Upregulated transcripts; **(D)** more abundant proteins; **(E)** downregulated transcripts; **(F)** less abundant proteins, respectively, from the pairwise differential expression analysis that map to reactions in the GEMs. **(G)** Heatmap of $\log_2$FC in transcriptomes top 100. DH, *M. thermautotrophicus* ΔH; ZZ, *M. thermautotrophicus* Z-245; MM, *M. marburgensis* Marburg.

Table 3.3 Metrics of the transcriptomics and proteomics for the individual analyses and the differential expression analyses for the three microbes. Statistical significance was determined by an adjusted P-value ≤ 0.05.

A) Individual analyses				
Omics	**Metric**	**DH**[1]	**ZZ**[2]	**MM**[3]
Transcriptomics	*Million reads*	26.3 ($\pm$ 2.0)	27.4 ($\pm$ 0.8)	26.0 ($\pm$ 1.1)
	Number of CDSs found (%)	1697/1796 (94.48%)	1699/1804 (94.18%)	1726/1730 (99.77%)
Proteomics	*Number of proteins found*	1406/1796 (78.29%)	1440/1804 (79.82%)	1374/1730 (79.42%)
B) Differential expression analyses				
Omics	**Metric**	**MM vs DH**	**ZZ vs DH**	**ZZ vs MM**
Transcriptomics	*DE*[4] *genes (range of* log_2FC*)*	1482 (-13.1 to 17.8)	816 (-17.0 to 13.8)	1498 (-14.1 to 16.7)
	DE genes with $\lvert log_2FC \rvert \geq 2$	975	265	1001
	DE genes that map to a reaction in the GEM	469	249	484
Proteomics	*DE proteins (range of* log_2FC*)*	121 (-6.6 to 6.6)	91 (-6.6 to 6.6)	93 (-6.0 to 5.9)
	DE proteins with $\lvert log_2FC \rvert \geq 2$	103	59	35
	DE proteins that map to a reaction in the GEM	43	17	38

[1] DH, *M. thermautotrophicus* ΔH
[2] ZZ, *M. thermautotrophicus* Z-245
[3] MM, *M. marburgensis* Marburg
[4] DE, differential expression

3.6.5 Significant differences in growth behavior are not mirrored in differential expression of methanogenesis pathway genes

To gain deeper insight into possible explanations for the differences in the growth behavior, we took a closer look at methanogenesis as the core metabolic feature of methanogens. The differential transcriptomic analysis generally indicated that for *M. marburgensis* Marburg, genes of the Wolfe Cycle (*i.e.*, methanogenesis) (Thauer, 2012) were upregulated compared to the other two microbes (**Figure 3.4, Data S5**). There were, however, some genes that were downregulated instead, including F_{420}-non-reducing Ni-Fe hydrogenase (Mvh), molybdenum-dependent formylmethanofuran dehydrogenase (Fmd), and several subunits of the F_{420} -reducing Ni-Fe hydrogenase (Frh), methyl-tetrahydromethanopterin (H_4MPT) coenzyme M methyltransferase (Mtr), and methyl-coenzyme M reductase isoenzyme II (Mrt). Nonetheless, these differences on transcriptome level did not translate to the proteome level in the same degree, as only a handful of the proteins were differentially abundant between the three microbes (**Figure 3.4**, **Data S5**). Also, the relative abundances of the different protein subunits in each protein complex, as well as the relative abundance of proteins involved in methanogenesis compared to the entire proteome were found to be similar between the three microbes (**Table S12, Data S5**). The most abundant methanogenesis-related proteins were Mtr, methyl-coenzyme M reductase isoenzyme I (Mcr), and tungsten-dependent formyl-MFR dehydrogenase (Fwd) for *M. thermautotrophicus* ΔH, *M. thermautotrophicus* Z-245, and *M. marburgensis* Marburg, respectively. All three microbes had higher abundances of Mcr to Mrt, Fwd to Fmd, and F_{420} -dependent methylene-H_4MPT dehydrogenase (Mtd) to H_2-forming methylene-H_4MPT dehydrogenase (Hmd). In addition, all three microbes had similar ratios of the energy-converting hydrogenases Eha and Ehb on proteome level (**Table S13**). Therefore, differences in the abundance of methanogenesis-related proteins did not explain differences in methane production between the microbes.

The aforementioned higher biomass production rate (0.01 $\pm$ 0.002 gCDW/L/h) for *M. marburgensis* Marburg compared to *M. thermautotrophicus* ΔH (1.40-fold) and M. thermautotrophicus Z-245 (1.39-fold), was not affected (P-value 0.428) between the replicates by a dioxygen intrusion into one of the bioreactors (**Figure 3.2B-C**, **Figure A.3**). However, this dioxygen intrusion led to an elevated oxidative-stress response, which we were able to detect in the transcriptome and proteome of *M. marburgensis*

Marburg (**Text A.1.2**). When considering the first few reactions that branch off from methanogenesis towards biomass growth (*e.g.*, ACS, CODH/CODHr2, POR2, HMGCOAS, PC), we still did not find any obvious reasons for the different growth behaviors (**Data S5**). Thus, we looked at metabolic functions that are related to biomass growth further downstream in the metabolism. We applied the GEMs and found that the higher biomass production rates in *M. marburgensis* Marburg was partially supported when considering the expression of metabolic genes that are related to biomass growth. Compared to *M. thermautotrophicus* ΔH, 199 genes were upregulated and 132 genes were downregulated, while compared to *M. thermautotrophicus* Z-245 198 genes were upregulated and 163 genes were downregulated in *M. marburgensis* Marburg, respectively (**Table S14**). The pathways with the most differentially expressed genes (up and down) include: **1)** purine biosynthesis, **2)** methanogenesis cofactor biosynthesis pathways (adenosylcobalamin and methanopterin); and **3)** central carbon metabolism (TCA, gluconeogenesis, and mixed sugar metabolism) (**Table S14**).

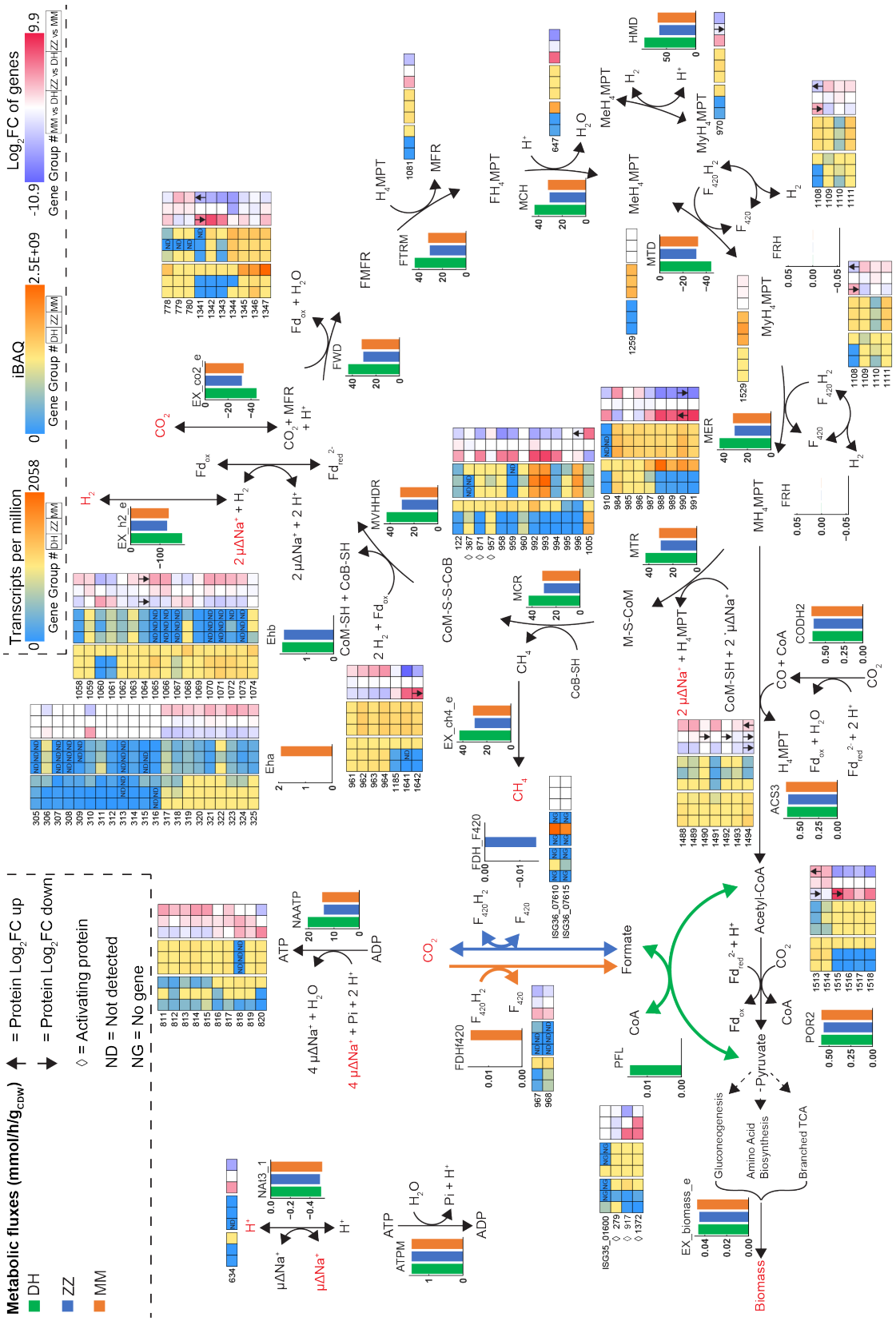

Figure 3.4: (Caption on the next page.)

Figure 3.4: (Previous page.) **Branched Wolfe Cycle adapted from Thauer (2012) with other reactions involved in the energy metabolism.** Fluxes are from the proteomics reduced model, constrained with experimental data that was adjusted for gross measurement error. Transcripts per million (TPM) for genes, and iBAQ values for proteins. Log_2 fold change (log_2FC) for differentially expressed genes and proteins. Gene group is used as ID for the omics. For the PFL reaction, only *M. thermautotrophicus* ΔH has the gene, thus the gene ID is used (ISG35_01600). For the FDH_F420, only *M. thermautotrophicus* Z-245 has the formate dehydrogenase cassette, thus the gene IDs are used (ISG36_07610 and ISG36_07615). Red text refers to metabolites that are exchanged across the membrane. **Microbes**: DH, *M. thermautotrophicus* ΔH; MM, *M. marburgensis* Marburg; ZZ, *M. thermautotrophicus* Z-245. **Compounds**: CH_4, methane, CO, carbon monoxide; CoA, Coenzyme A; CoB, coenzyme B; CoM, coenzyme M; CoM-S-S-CoB, CoM-CoB heterodisulfide; CO_2, carbon dioxide; F, formyl; Fdox/rd, ferredoxin oxidized/reduced; H_2, hydrogen; H^+, proton; H_4MPT, tetrahydromethanopterin; M, methyl; Me, methenyl; MFR, methanofuran; My, methylene; Na^+, sodium ion. **Reactions/Enzymes**: ATPM, ATP maintenance (pseudo reaction) ; CODHr2, CO dehydrogenase/acetyl-CoA synthase; Eha/Ehb, energy converting hydrogenases; EX_biomass_e, biomass exchange (pseudo reaction); EX_ch4_e, CH_4 exchange (pseudo reaction) ; EX_co2_e, CO_2 exchange (pseudo reaction); EX_h2_e, H_2 exchange (pseudo reaction); FDHf420, F_{420}-dependent formate dehydrogenase; FDH_F420, F_{420}-dependent formate dehydrogenase cassette; FRH, F_{420}-reducing hydrogenase; FTRM, FMFR/H_4MPT formyltransferase; FWD, FMFR dehydrogenase (tungsten- and molybdenum-dependent isozymes); HMD, MeH_4MPT hydrogenase; MCH, MeH_4MPT cyclohydrolase; MCR, MCoM reductase (I and II); MER, MyH_4MPT reductase; MTD, MyH_4MPT dehydrogenase; MTR, MH_4MPT/CoM methyltransferase; MVHHDR, F_{420}-non-reducing hydrogenase with the heterodisulfide reductase; NAATP, ATP synthase; Nat3_1, Na^+/H^+ antiporter; PFL, pyruvate formate-lyase; POR2, pyruvate synthase. **Other**: ◇, activating protein; ND, not detected; NG, no gene.

3.6.6 Integration of experimental data into GEMs results in feasible solutions for flux balance analyses and reveals differences in the generation of formate for biomass growth

With the steady-state fermentation data from the bioreactor experiment (consumption rates of carbon dioxide and molecular hydrogen, and production rates of methane and biomass), we constrained the GEMs in flux balance analyses (FBA, **Materials and Methods, 3.5.20**). To correct for experimental errors, we used gross measurement adjusted values for flux balance analyses (**Data S4, S5, Materials and Methods, 3.5.12**). The GEM found a feasible solution with a 0.1%, 0.1%, and 0.5% deviation of the reaction flux bounds from the calculated fluxes for *M. thermautotrophicus* ΔH, *M. thermautotrophicus* Z-245, and *M. marburgensis* Marburg, respectively (**Data S2, S5**). The maximum fluxes of the non-growth-associated ATP maintenance (reaction ATPM) varied between the three microbes. The non-growth-associated maintenance represents the dissipation of ATP (Valgepea *et al.*, 2018), and was highest for *M. thermautotrophicus* ΔH, compared to *M. thermautotrophicus* Z-245, and *M.*

marburgensis Marburg (**Table 3.4**). The higher the non-growth-associated maintenance, the more methane (and ATP) needs to be produced per biomass unit. The interspecies variation in product distribution that we found in our experiments agrees with the variations in the modeled maximum non-growth-associated maintenance (**Table 3.4; Figure 3.2D**). We further constrained the GEMs with the transcriptomics and proteomics data, which resulted in reduced GEMs, and performed a set of simulations comparing the unconstrained and reduced GEMs (**Figure 3.4; Data S2, Materials and Methods, 3.5.20**). With this, we were able to have a closer look at exemplified pathways, such as the conversion of methenyl-H_4MPT to methylene-H_4MPT, which did not allow to resolve the exact flux catalyzed by different enzymes for these reactions (**Text A.1.3**).

With our pan-genome comparison, we found the formate acetyltransferase/pyruvate-formate lyase (Pfl, reaction PFL) in *M. thermautotrophicus* ΔH (MTH346/ISG35_1600), but neither in *M. marburgensis* Marburg (Kaster *et al.*, 2011b; Sawers and Watson, 1998) nor in *M. thermautotrophicus* Z-245. This enzyme catalyzes the CoA-dependent reversible conversion of pyruvate to formate (for modeling purposes, the reaction is used irreversibly due to thermodynamics considerations, **Materials and Methods, 3.5.6**). We found several putative Pfl-activating enzymes in the three microbes (**Text A.1.4**). However, given the missing pyruvate-formate lyase, we assumed that *M. thermautotrophicus* Z-245 and *M. marburgensis* Marburg do not utilize the PFL reaction, and thus, no genes are included for those reactions in the GEMs.

Theoretically, *M. thermautotrophicus* ΔH can use a formate dehydrogenase (Fdh, ISG35_05415 and ISG35_05410, gene groups 967 and 968) in addition to the pyruvate-formate lyase to synthesize formate for biomass growth. The formate dehydrogenase was hypothesized to be the relevant reaction for formate production before (Kaster *et al.*, 2011b). However, reducing the GEM with proteomics data resulted in the pyruvate-formate lyase as the preferred formate production route, because formate dehydrogenase is not detected with proteomics, thus eliminating the corresponding FDHf420 reaction (**Figure 3.4; Data S2**). The pyruvate-formate lyase is also upregulated compared to the formate dehydrogenase in the transcriptomics data. This indicates that *M. thermautotrophicus* ΔH primarily produces formate for biomass growth *via* the pyruvate-formate lyase; however, this hypothesis will need to be confirmed experimentally.

In contrast to both *M. thermautotrophicus* ΔH and *M. marburgensis* Marburg, *M. thermautotrophicus* Z-245 has two functioning sets of formate dehydrogenases, one of which is encoded in the operon *fdhCAB* (*i.e.*, the formate dehydrogenase cassette), which includes a putative formate transporter (FdhC), and which is responsible for utilization of formate as an alternative carbon and electron source (Nölling and Reeve, 1997). Nölling and Reeve (1997) found upregulation of expression of the formate dehydrogenase cassette when cells were limited by molecular hydrogen availability. Interestingly, the formate dehydrogenase cassette was also highly abundant in our transcriptome and proteome, despite a high molecular hydrogen availability (**Figure 3.4; Data S5**). The second set of formate dehydrogenase genes (gene groups 967 and 968) does not include a putative transporter-encoding gene, and is homologous to the formate dehydrogenase found in *M. thermautotrophicus* ΔH and *M. marburgensis* Marburg. Thus, the latter may be involved in formate production for biomass growth (reaction FDHf420 for *M. thermautotrophicus* ΔH and *M. marburgensis* Marburg). However, in *M. thermautotrophicus* Z-245 the transcript levels of this formate dehydrogenase were less abundant compared to the formate dehydrogenase cassette and was not detected with proteomics. Simulations with *M. thermautotrophicus* Z-245 primarily predicted the use of the FDH_F420 (the formate dehydrogenase cassette) in the direction of formate production rather than FDHf420 (including the omics constrained runs except for the loopless FBA with transcriptomics) (**Figure 3.4; Data S2**).

M. marburgensis Marburg neither encodes the pyruvate-formate lyase nor the formate dehydrogenase cassette. Thus, it is likely that formate is only produced from the formate dehydrogenase of gene groups 967 and 968. This was also hypothesized before, and supported by a formate auxotrophic strain of *M. marburgensis* Marburg, which did not exhibit formate dehydrogenase activity anymore (Tanner *et al.*, 1989). All simulations (including the omics constrained ones) predicted the use of the FDHf420 reaction (**Data S2**). The two subunits were upregulated in the transcriptome by 2.4-4.9 $\log_2$FC compared to the other two microbes (**Figure 3.4 and Data S5**). However, only one subunit was found in the proteome with low abundance (gene group 967: 0.000132%). This is orders of magnitude lower compared to the formate dehydrogenase cassette in *M. thermautotrophicus* Z-245 (FdhB: 0.254%, FdhA: 0.191%, FdhC: 0.126%). Because *M. marburgensis* Marburg reached higher biomass concentrations (**Figure 3.2C**), these findings raise questions on how *M. marburgensis*

Marburg actually produces formate for biomass growth. Potentially, formate production is achieved by other yet unconsidered means. In case the low-abundant formate dehydrogenase is responsible for formate production, questions on the kinetic properties of the formate dehydrogenase remain. While more experimentation will be required to resolve this, it demonstrates how our integrated systems biology approach can be utilized to identify differences in closely related species with implications on the development of bioprocessing and metabolic engineering strategies.

Table 3.4 ATPM flux ranges for different constraints. Range considers the ATPM flux with the objective function of biomass production maximization to ATPM maximization.

DH[1]	**ZZ**[2]	**MM**[3]
Model[4]: 1.5-17.85	Model: 1.5-10.73	Model: 1.5-12.20
Model trans[5]: 1.5-17.46	Model trans: 1.5-10.48	Model trans: 11.97-12.19
Model prot[6]: 1.5-17.55	Model prot: 1.5-10.46	Model prot: 1.5-11.90

[1] DH, *M. thermautotrophicus* ΔH
[2] ZZ, *M. thermautotrophicus* Z-245
[3] MM, *M. marburgensis* Marburg
[4] Model, model only constrained with experimental data
[5] Model trans, model additionally constraint with the transcriptomics data
[6] Model prot, model additionally constraint with the proteomics data

3.7 Summary and outlook

Power-to-gas provides a means for storing excess renewable electric power while simultaneously capturing carbon dioxide. Methanogens can act as biocatalysts to generate methane. Although methanogenesis is well studied, there are still knowledge gaps surrounding methanogenic metabolism, which limit their ability to be harnessed for more versatile biotechnological applications.

In this study, we compared *M. thermautotrophicus* ΔH, *M. thermautotrophicus* Z-245, and *M. marburgensis* Marburg at a genomic, transcriptomic, and proteomic level to identify differences in their metabolism. While the genomes encode over 1600 genes that appear homologous, most of the genes are differentially expressed on transcriptomic level. Nevertheless, the proteome did not exhibit nearly as many differences, though in both omics datasets, fewer differences were found between the two *M. thermautotrophicus* species compared to *M. marburgensis* Marburg, which is consistent with their phylogeny (Wasserfallen *et al.*, 2000). Under the conditions of our

bioreactor experiments, *M. thermautotrophicus* ΔH had a higher specific methane production rate compared to the other microbes, while *M. marburgensis* Marburg reached higher biomass concentrations. From our modeling results (with the maximization of the ATPM reaction), it was possible to intuit that *M. thermautotrophicus* ΔH has the highest and *M. marburgensis* Marburg the lowest non-growth-associated maintenance. As ATP is the main energy currency in the cell, this finding plays a crucial role in adopting and optimizing *Methanothermobacter* spp. as cell factories for different purposes (Chen and Nielsen, 2019). From the omics data, we identified that the three microbes may use three different enzymes to produce formate for biomass growth, such as for purine biosynthesis (White, 1997; Wei *et al.*, 2015; Brown *et al.*, 2011). These findings have to be taken into account for further optimization strategies for commercial applications, and our GEMs can now be utilized to investigate further pathways under various growth conditions.

Understanding the metabolism of a microbe is especially important when trying to rewire it for biotechnological purposes (*e.g.*, use of different substrates or production of value-added compounds). Recently, we developed a genetic system for *M. thermautotrophicus* ΔH demonstrating gain-of-function modifications with a shuttle-vector system (Fink *et al.*, 2021). We added the formate dehydrogenase cassette from *M. thermautotrophicus* Z-245 to *M. thermautotrophicus* ΔH, which permitted growth on formate as the sole carbon and energy source (natively *M. thermautotrophicus* ΔH only utilizes carbon dioxide and molecular hydrogen) (Fink *et al.*, 2021). The ability to combine the GEM with this genetic system will be a massive step towards power-to-x applications with thermophilic methanogens, because now the verified and validated GEM can be applied to predict phenotypes.

Methanothermobacter spp. have been already implemented for power-to-gas applications. Understanding the metabolic strengths and limitations of different strains will help to select the most appropriate strain for a given application. The high-quality (umbrella) model, and multi-replicated steady-state gas fermentation and omics data is a rich resource that can be used for a plethora of follow-up analyses and to develop metabolic engineering strategies.

3.8 Acknowledgments

This work was supported by the Humboldt foundation in the framework of the Humboldt professorship, which was awarded to L.T.A., and by the Max Planck Institute for Developmental Biology (R.E.L. and L.T.A.). I.C. would like to acknowledge support from the German Academic Exchange Service (DAAD) through the DAAD *Kurzzeitstipendien für Doktoranden*. The authors would like to thank Luis Antoniotti and Jürgen Barth from the Max Planck Institute for Biology Tübingen workshop for their help with modifying the bioreactor system. The authors would like to thank Nicolai Kreitli, Lucas Mühling, Nils Rohbohm, and Andrés Ortíz Ardilla for helpful input and support with experimentation. The authors would like to acknowledge the assistance of the Genome Center of the Max Planck Institute for Biology Tübingen, particularly Christa Lanz, for their assistance in the PacBio sequencing, and the Proteome Center Tübingen (PCT) for their support with the proteomics. The photo of the electrolyzer in the graphical abstract is courtesy of Peak Scientific.

3.9 Supplementary information

Supplementary information: supplementary texts and figures are found in in **Appendix A**. Details and the GitHub link for the additional supporting material (modeling, fermentation data, omics data and scripts) can also be found in **Appendix A**.

Chapter 4

Casini, I., McCubbin, T., Luque, G., Ley, R.E., Marcellin, E., Angenent, L.T., Molitor, B. Episomal formate dehydrogenase gene expression in genetically engineered *Methanothermobacter thermautotrophicus* ΔH enables formate consumption in continuous fermentation at reduced non-growth-associated maintenance energies. *Preliminary Manuscript*.

4.1 Author contributions

Bastian Molitor (B.M.) and Largus T. Angenent (L.T.A.) initiated the work. Isabella Casini (I.C.), B.M., and L.T.A. designed the experiments. Ruth E. Ley (R.E.L.) provided the bioreactor system and the space. I.C. operated the bioreactors. I.C. extracted the RNA and proteins. I.C. analyzed the omics data with the help of Timothy McCubbin (T.M.) and Guillermo Luque (G.G.L.). I.C. analyzed the rest of the data and modeled with the help of T.M. and Esteban Marcellin (E.M.). L.T.A. and B.M. supervised the project. I.C. produced all the figures and tables. I.C. and B.M. wrote the manuscript, while all edited the paper and approved the final version.

4.2 Abstract

Methane is produced biologically by methanogenic archaea that convert substrates, such as molecular hydrogen and carbon dioxide, acetate, and methylated compounds. Methane is a greenhouse gas and plays an important role in the global carbon cycle. Anthropogenic activities, such as extraction and combustion of fossil fuels and agriculture, have led to an increase in atmospheric methane levels. Although methane is a harmful greenhouse gas when released into the atmosphere, it is also the primary

component of natural gas. While fossil natural gas utilization needs to be avoided, methane can be sustainably generated with the power-to-gas platform, including biomethanation with methanogenic archaea. A circular carbon economy in which one-carbon compounds, such as carbon dioxide and methane, are recycled would enable the sustainable use of existing carbon-based energy infrastructure and systems. *Methanothermobacter* spp. are already being employed as biocatalysts for sustainable methane production at industrial scales. We recently constructed a genome-scale metabolic model (GEM) for *Methanothermobacter thermautotrophicus* ΔH and developed a genetic system that allowed us to expand the metabolism of *M. thermautotrophicus* ΔH to grow solely on formate. Here, we investigated the genetically modified *M. thermautotrophicus* ΔH strain (*M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$) that encoded the F_{420}-dependent formate dehydrogenase from *Methanothermobacter thermautotrophicus* Z-245 on formate in continuous bioreactors. We found comparable uptake and production rates to the native formate user, *M. thermautotrophicus* Z-245. By integrating fermentation, transcriptomics, and proteomics data into the GEM, we observed a decrease in non-growth-associated maintenance energy with formate compared to molecular hydrogen and carbon dioxide. This was linked to the increase in the product ratio of biomass to methane. Understanding the (energy) metabolism through systems biology, including metabolic modeling techniques is integral for biotechnology, especially when using synthetic biology.

4.3 Significance

Methanothermobacter spp. are already used as biocatalysts in the power-to-gas platform, specifically for methane production from molecular hydrogen and carbon dioxide. Power-to-gas provides a way to overcome the intermittent production of renewable electric power from fluctuating sources, such as solar and wind, by allowing to store excess electric power. The recent establishment of a genetic system for *Methanothermobacter* spp. offers an opportunity to expand the substrate and product spectrum of the microbe, thus, advancing its potential in biotechnology (*e.g.*, power-to-x). Here, we use systems biology, including metabolic modeling to demonstrate the stable consumption of formate in a genetically modified

Methanothermobacter thermautotrophicus ΔH strain with a plasmid that encodes for a F_{420}-dependent formate dehydrogenase and leads to lower non-growth-associated maintenance energies. Our findings have implications for the adaptation of methanogens for biotechnological applications.

4.4 Introduction

A rapid industrialization, which was driven by fossil fuels, has contributed considerably to a rise in atmospheric greenhouse gas concentrations, particularly carbon dioxide (CO_2), leading to anthropogenic climate change. The increase in atmospheric carbon dioxide concentrations has led to numerous detectable problems, including, but not limited to, rising global temperatures, acidification of oceans, rising sea levels, loss of biodiversity, and more severe weather (Ciscar *et al.*, 2011). These physical consequences also have direct (negative) economic implications, which can be quantified using metrics such as gross domestic product (Ciscar *et al.*, 2011). Simultaneously, the increasing population and associated energy demands require alternative and renewable power supplies. Nevertheless, traditional renewable power sources are intermittent, and there is often a discrepancy between peak production and peak demand periods. This necessitates the storage of excess electric power for later use. The power-to-gas platform with biomethanation addresses the need to store electric power and sequester carbon dioxide (Martin *et al.*, 2013). First, molecular hydrogen (H_2) is generated by electrolyzing water. Second, methane (CH_4) is produced by combining the molecular hydrogen with various carbon dioxide sources, which may include biogas from anaerobic digestion or industrial waste gas streams (Angenent *et al.*, 2017; Götz *et al.*, 2016; Schiebahn *et al.*, 2015). Methane is the primary component in natural gas, for which there is an existing and well-established infrastructure. However, presently this infrastructure can only tolerate, at best, a maximum of 10 vol% of molecular hydrogen as a fuel (Rusmanis *et al.*, 2019). Biomethanation with methanogens is advantageous over thermo-chemical conversion because: **1)** it does not require high-purity sulfur-free gases that can poison the metal catalysts (Dry, 2002; Leonzio, 2016), instead the typical sulfur compounds present, such as H_2S, can be used as sulfur sources by the methanogens (Liu *et al.*, 2012); **2)** it can be switched off and on again by adjusting the temperature of operation; **3)** it does

not require high temperatures and pressures (Müller *et al.*, 2013; Dry, 2002; Leonzio, 2016); and **4)** it does not require rare and expensive metal catalysts (Van Der Laan and Beenackers, 1999).

Hydrogenotrophic methanogens, specifically *Methanothermobacter* spp., are already industrially backed biocatalysts for biomethanation, which is because of their fast doubling time and high methane evolution rates (Rusmanis *et al.*, 2019; Martin *et al.*, 2013). *Methanothermobacter thermautotrophicus* ΔH (formerly *Methanobacterium thermoautotrophicum* ΔH) has been a model microbe for studying methanogenesis since the 1970s. Nevertheless, a genetic system was only recently developed in our group (Fink *et al.*, 2021), now opening the possibility to expand the substrate and product spectrum for biotechnology such as power-to-x. Hydrogenotrophic methanogens, such as *M. thermautotrophicus* ΔH, grow with molecular hydrogen and carbon dioxide as their energy and carbon sources (**Equation (4.1)**). However, related species and strains, such as *Methanothermobacter thermautotrophicus* Z-245 and *Methanothermobacter wolfeii*, are also capable of using solely formate (formic acid) for both energy and carbon sources, which is due to their F_{420}-dependent formate dehydrogenase (*fdhCAB*, EC 1.17.98.3) (**Equation (4.2)**) (Wasserfallen *et al.*, 2000). Previously, we used a shuttle vector that encoded the F_{420}-dependent formate dehydrogenase cassette from *M. thermautotrophicus* Z-245 to amend the capability of using solely formate as a substrate to *M. thermautotrophicus* ΔH (Fink *et al.*, 2021).

$$4\,H_2 + CO_2 \rightarrow CH_4 + 2\,H_2O \tag{4.1}$$

$$4\,HCOOH \rightarrow CH_4 + 3\,CO_2 + 2\,H_2O \tag{4.2}$$

Formate has been identified to be one of the most promising liquid one-carbon substrates (along with methanol) in terms of energetic efficiency, even under anaerobic conditions (Claassens *et al.*, 2019). Nevertheless, few studies performed quantitative cultivations on solely formate and none in a minimal media without complex additives (such as vitamins). Further, the calculations in Claassens *et al.* (2019) were based on only two studies with methanogens and four with acetogenic bacteria. Although Costa *et al.* (2013a) performed continuous bioreactor fermentations on sodium formate (sodium formate) with *Methanococcus maripaludis*, they began with a batch fermentation on gaseous substrates (molecular hydrogen and carbon dioxide) to reach a

high biomass concentration. When they switched to a continuous fermentation for growth on formate, all gassing was halted. Other studies cultivated methanogens on formate but focused on: **1)** the role of formate as a metabolite and substrate during hydrogenotrophic methanogenesis *via* genetic manipulations and protein interactions (Costa *et al.*, 2013b, 2010); and **2)** the transcription of the F_{420}-dependent formate dehydrogenase for growth on molecular hydrogen and carbon dioxide and sodium formate (Nölling and Reeve, 1997).

Formic acid, or the salt form, formate, was already conceptualized as a means to store molecular hydrogen *via* the (reversible) hydrogenation of carbon dioxide in a non-gaseous state (Álvarez *et al.*, 2017). The liquid or solid form facilitates the storage and transport of the compound and energy (molecular hydrogen). The review by Álvarez *et al.* (2017) provides an extensive summary and assessment of the current thermochemical methods, including catalysts types that can be used to produce formic acid or formate. Nevertheless, thermochemical systems are often sensitive to contamination or poisoning, may be constructed of rare or expensive material, and may require high temperatures (Roy *et al.*, 2018). However, there are more sustainable ways to produce formate, including by electrochemical, photochemical, and biological processes (Cotton *et al.*, 2020). It is an attractive prospect to use sustainably produced formate as an intermediate and substrate for other biotechnological process that can mitigate both the rise in greenhouse gases and the energy crisis.

Genome-scale metabolic models (GEMs) mathematically represent the stoichiometric metabolism (S-matrix) of organisms and the genes responsible for those metabolic functions (gene-protein-reaction relationships), assuming conservation of mass and energy at steady-state conditions (Maarleveld *et al.*, 2013). These models are used as the scaffold for integrating cultivation, omics, regulatory, and thermodynamics data (Suthers *et al.*, 2021; Mahamkali *et al.*, 2021). They are the base of more complex models such as metabolism and expression models (Fang *et al.*, 2020), and kinetic models (Saa and Nielsen, 2017). They are adopted to: **1)** model cultivation conditions (*e.g.*, difficult-to-measure intracellular fluxes); **2)** predict phenotypes (*e.g.*, genetic modifications, synthetic lethality); and **3)** to comprehend the metabolism (*e.g.*, determine gene essentiality) of an organism (Henry *et al.*, 2010; Zhang and Hua, 2016). These *in-silico* techniques help to edit and optimize metabolisms for applications in many sectors such as biotechnology and alternative energy generation (Marcellin *et al.*, 2016). The increase in tools for automated construction of GEMs has facilitated the

models own generation (Mendoza *et al.*, 2019). However, most of these tools do not replace an essential manual curation step that ensures a high-quality model (Henry *et al.*, 2010; Thiele and Palsson, 2010). Archaeal models only represent a small fraction of the GEMs that have been built and manually curated, with 2% and 0.14%, respectively. However, we recently released high-quality GEMs for different *Methanothermobacter* spp. and we use these models (with some modifications) for the modeling in this work (**Chapter 3**). The substrate use for methanogen models varies, depending on the methanogen species, and has included: **1)** carbon dioxide and molecular hydrogen; **2)** methylated substrates (*e.g.*, methanol); **3)** acetate; and **4)** carbon monoxide (Thor *et al.*, 2017). Seldom has formate been considered as a substrate. Hamilton *et al.* (2015) modeled formate co-use (and co-culture) with *Methanospirillum hungatei* that was grown on molecular hydrogen, carbon dioxide, and acetate. Goyal *et al.* (2014) modeled the co-use of formate and molecular hydrogen with *Methanococcus maripaludis*. Lastly, Richards *et al.* (2016) used a flux variability analysis (FVA) constrained with formate as the sole substrate in *Methanococcus maripaludis*.

Here, we assessed the performance of the modified strain, *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$, which is able to utilize formate as the sole growth substrate (Fink *et al.*, 2021). For this, we constrained our GEMs with fermentation, transcriptomics, and proteomics data, which we collected from quadruplicate continuous bioreactors in steady state. We compared at a pan-transcriptomics and pan-proteomics level the growth of *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ on molecular hydrogen and carbon dioxide *vs.* on sodium formate, *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ *vs.* *M. thermautotrophicus* Z-245, and *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ *vs. M. thermautotrophicus* ΔH wild type.

We applied systems biology, including metabolic modeling methods to highlight potential changes in methanogenic metabolism after genetic modifications. We evaluated the potential of (sodium) formate as a substrate for methanogenesis and an intermediate to store excess renewable electric power and carbon dioxide.

4.5 Materials and Methods

4.5.1 Microbial strains and medium composition

M. thermautotrophicus ΔH (DSM 1053) and *M. thermautotrophicus* Z-245 (DSM 3720) were obtained from the DSMZ (Braunschweig, Germany). *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ was previously generated in our lab by Fink *et al.* (2021). The microbes were cultivated essentially as described in **Chapter 3**. Briefly, the serum bottles were incubated at 60°C with shaking at 150 RMP (Lab Companion ISS-7100R; Jeio Tech, Republic of Korea). Growth was typically observed within 16 h for growth on H_2/CO_2 (80/20% v/v) and 36 h for growth on Na-formate. Both growth conditions reached the late exponential phase at an OD_{600} of approximately 0.2 at which point the bottles were removed from the incubator.

Preparation of batch and continuous media for the bioreactor runs was as described in **Chapter 3**, with the following adjustments. Serum bottles for batch cultivation on Na-formate were sparged with N_2/CO_2 (80/20 %, v/v), and 100 mM Na-formate was added after autoclaving. All serum bottles were supplemented with 0.001 mM sodium selenate and 0.01 mM sodium molybdate dihydrate. Continuous media was supplemented with 0.0015 mM sodium selenate and 0.015 mM sodium molybdate dihydrate. For the bioreactors with Na-formate, the continuous media contained 355 $\pm$ 5 mM. A concentrated Na-formate solution was prepared and sterilely added to the continuous media after autoclaving, sparging, and reducing the media.

4.5.2 Biomass composition determination and maintenance energies

The biomass composition, growth-associated maintenance (GAM), and non-growth-associated ATP maintenance costs (NGAM) were determined as described in **Chapter 3**.

4.5.3 Bioreactor setup and operating conditions

The bioreactor setup was as described in **Chapter 3** except that bioreactors that were grown on Na-formate were pH adjusted using 1 N HCl instead of 1 N NaOH and were sparged with N_2/CO_2 (80/20 %, v/v) (rather than H2/CO_2 (80/20 %, v/v)) at a gas flow rate of 10 mL min^{-1}. As previously described, each reactor was inoculated with 6 mL of

preculture that was grown in serum bottles (OD_{600} ~0.2). The reactors were operated in batch mode until an OD_{600} of approximately 1.0 for reactors that were grown on H_2/CO_2 and 0.1-0.15 for reactors that were grown on Na-formate was reached. At this point, the continuous mode was started with a media feed at a dilution rate of ~1.0 d^{-1} (this rate was achieved by ramping over several days for the reactors that grown on Na-formate). A steady-state condition is defined as three hydraulic retention times (HRT) (Jensen, 2001) and was 2.7 days. After reaching steady state, the samples for the transcriptomics, proteomics, and gram cell-dry weight were taken. Daily sampling of OD_{600}, pH, exhaust gas flow rate, and inlet and exhaust gas composition were performed as described in **Chapter 3**. After steady state was reached, the reactor effluent was sampled daily for Na-formate and other carboxylates and alcohols (**Materials and Methods, 4.5.7**).

4.5.4 Cross-contamination check

The reactors were checked for cross-contamination as was described in **Chapter 3**. The *M. thermautotrophicus* Z-245 forward and reverse primers were used to check for *M. thermautotrophicus* Z-245, while both the *M. thermautotrophicus* ΔH forward and reverse and the *M. marburgensis* Marburg forward and reverse primers used were to detect *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ (**Table S15**).

4.5.5 Fermentation gas analysis

The gas analysis was performed as described in **Chapter 3** but with adjusted calibration levels (**Table S16**).

4.5.6 Biomass concentration analysis

The biomass correlation coefficient was calculated as described in **Chapter 3**. The following K values (g/L/OD_{600}) were found: *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ (H_2/CO_2), 0.31; *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ (Na-formate), 0.89; and *M. thermautotrophicus* Z-245 (Na-formate), 0.95.

4.5.7 Sodium formate concentration measurements

Sodium formate concentrations were analyzed *via* a high-pressure liquid chromatography (HPLC) (SIL-40C, Shimadzu Europa, Duisburg, Germany) system that was equipped with an Aminex HPX-87H column (300 by 7.8 mm; Bio-Rad, CA, USA) and a refractive index detector (RID-20A). A 5 mM sulfuric acid solution was used as the eluent, with a flow rate of 0.6 mL min^{-1} and a sample run time of 30-60 min. The oven temperature was set to 60°C, while the sample rack of the attached autosampler to 4°C. For HPLC sample preparation, all culture samples (0.5 mL ± 0.1 mL) were filtered using 0.22 μm filters (ROTILABO^{TM} PVDF, 13 mm, Carl Roth, Karlsruhe, Germany). Na-formate calibration curves were prepared with concentrations ranging from 0.5-10 mM and 20-400 mM. Additionally, we sampled for the following compounds, though none were detected: carboxylates (acetate, *n*-butyrate, *n*-caproate, ethyl *n*-valerate, isobutyrate, isovalerate, propionate, succinate, and valerate), and alcohols (ethanol, *n*-butanol, isopropanol, and methanol).

4.5.8 Calculating production rates through carbon and electron balances

For the reactor runs on H_2/CO_2, the carbon balances were performed as described in **Chapter 3**, with an adjusted salinity value of 4.568 ppt rather than 4.565 ppt (**Chapter 3**). However, given larger inconsistencies in carbon and electron balances with the raw data for bioreactor runs on Na-formate, a different method was applied. No Na-formate was detected in the liquid media during steady state, and thus it is assumed that the entire Na-formate from the feed was consumed. However, the quantity of 1 M HCl solution that was added to the bioreactors for pH control was not insignificant (unlike the quantity of added NaOH for fermentations on H_2/CO_2). The feed rate of the pH control pumps was determined, and a dilution factor was calculated. Approximately 21.88 ± 2.93% of the liquid feed into the reactors came from the acid feed, thus lowering the effective Na-formate feed concentration to 291.38 ± 10.55 mM (**Table S17**). Further, the method in **Chapter 3**, which we used to estimate dissolved carbon species, is very sensitive to salinity. The increased salt concentrations from the Na-formate (new value estimated at 17.80 ± 0.44 ppt) was not sufficient for closing both carbon and electron balances. It is also noted that variation in pH was more

frequent due to the production of NaOH with Na-formate consumption and at times there was a noticeable shift in the online pH probe measurement that required a manual offset after offline pH measurements. The dissolved CO_2 values are quite sensitive to changes in pH. Thus, to solve for the quantities of produced methane and CO_2, **Equations (4.3) and (4.4)**, which represent the carbon (C) and electron (E, using degree of reduction) balances, respectively, were solved as a system of equations. $Naformate_{aq\_in}$, is the Na-formate substrate; CO_{2,gas_out}, CO_{2,aq_out}, CO_{2,gas_aq_out} are the gaseous, soluble, and the sum of the gaseous and soluble CO_2 products, respectively; CH_{4,gas_out} is the gaseous CH_4 product; H_{2,gas_out} is the gaseous H_2 product; and $biomass_{aq\_outlet}$ is the biomass product in the reactor (liquid) effluent. The biomass coefficient in **Equation (4.4)** was calculated assuming a molecular formula of $CH_{1.681}O_{0.481}N_{0.222}$, as described in (Michael and Kargi, 2002).

$$\begin{aligned} 0 &= C_{in} - C_{out} \\ 0 &= Naformate_{aq\_in} - (CO_{2,gas\_out} + CO_{2,aq\_out} + CH_{4,gas\_out} + biomass_{aq\_out}) \\ 0 &= Naformate_{aq\_in} - (CO_{2,gas\_aq\_out} + CH_{4,gas\_out} + biomass_{aq\_out}) \end{aligned} \tag{4.3}$$

$$\begin{aligned} 0 &= E_{in} - E_{out} \\ 0 &= 2 \cdot Naformate_{in} - (8 \cdot CH_{4,gas\_out} + 4.053 \cdot biomass_{aq\_out} + 2 \cdot H_{2,gas\_out}) \end{aligned} \tag{4.4}$$

4.5.9 Normalized product distribution

Product distributions were as described in **Chapter 3**. The CO_2 production was also accounted for during runs with Na-formate. A single factor analysis of variance (ANOVA) test was conducted in Excel® to analyze statistically significant differences between the product distribution ratios of *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ and *M. thermautotrophicus* Z-245 that were grown on Na-formate. A second ANOVA was conducted between the *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ that was grown on H_2/CO_2 from this work with *M. thermautotrophicus* ΔH, *M. thermautotrophicus* Z-245, and *M. marburgensis* Marburg that were grown on H_2/CO_2 from **Chapter 3**.

4.5.10 Energy efficiency

Energy efficiencies under anaerobic conditions were calculated as described in Equation 1 of Claassens *et al.* (2019), using the heat of combustions provided in Supplementary Data 1 of Claassens *et al.* (2019). H_2: -260 kJ/mol; HCOOH: -245 kJ/mol; CH_4: -860 kJ/mol.

4.5.11 Interspecies comparison

To compare the species, Independent T-tests were used as described in **Chapter 3**. The first bioreactor replicate of *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ that was grown on Na-formate stopped spinning and thus, the gas and biomass data points were not included in the T-tests. The rest of the gas and biomass data points were used for the comparative analysis between *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ and *M. thermautotrophicus* Z-245 that were grown on Na-formate. A second set of T-tests was conducted between the *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ that was grown on H_2/CO_2 from this work with *M. thermautotrophicus* ΔH, *M. thermautotrophicus* Z-245, and *M. marburgensis* Marburg that were grown on H_2/CO_2 from **Chapter 3**.

4.5.12 Transcriptomics

RNA sample preparation

RNA extraction, purification, and quantity (2.8-8.3 μg) and quality (RNA integrity index $>$ 7.9) checks were performed as described in **Chapter 3**.

RNA sequencing

RNA sequencing was performed as described in **Chapter 3**.

Raw sequencing data analysis

The RNA-seq reads processing was performed as described in **Chapter 3** with only *M. thermautotrophicus* ΔH (CP064324) and *M. thermautotrophicus* Z-245 (CP064336 and CP064337). The *M. thermautotrophicus* ΔH genome files were modified before use by adding the plasmid pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ (Fink *et al.*, 2021). *M.*

thermautotrophicus ΔH pMVS1111A:P$_{hmtB}$-*fdh*$_{Z\text{-}245}$ that was grown on H_2/CO_2 (DH_H) and *M. thermautotrophicus* ΔH pMVS1111A:P$_{hmtB}$-*fdh*$_{Z\text{-}245}$ that was grown on Na-formate (DH_F) were compared, and *M. thermautotrophicus* ΔH pMVS1111A:P$_{hmtB}$-*fdh*$_{Z\text{-}245}$ that was grown on Na-formate (DH_F) and *M. thermautotrophicus* Z-245 (ZZ_F) that was grown on Na-formate were compared.

4.5.13 Label-free proteomics

Protein sample preparation

Protein extraction and purification were performed as described in **Chapter 3**, except the biomass pellet was washed with PBS adjusted to pH 2.0 (rather than 7.0).

Protein measurement

Proteins were measured as described in **Chapter 3**.

Protein analysis with proteome discoverer

Raw data dependent acquisition (DDA) data from LC-MS/MS were processed as described in **Chapter 3** with the following modifications. A protein FASTA file for *M. thermautotrophicus* ΔH pMVS1111A:P$_{hmtB}$-*fdh*$_{Z\text{-}245}$ was made by adding the protein sequences from the pMVS1111A:P$_{hmtB}$-*fdh*$_{Z\text{-}245}$ plasmid to the protein FASTA file of *M. thermautotrophicus* ΔH. For the differential expression analysis, the new pan-genome was used (**Materials and Methods, 4.5.14**), which included only *M. thermautotrophicus* ΔH pMVS1111A:P$_{hmtB}$-*fdh*$_{Z\text{-}245}$ and *M. thermautotrophicus* Z-245.

4.5.14 Pan-genome differential expression database creation

The pan-genome was created as described in **Chapter 3**, except using only the protein FASTA files of *M. thermautotrophicus* ΔH pMVS1111A:P$_{hmtB}$-*fdh*$_{Z\text{-}245}$ and *M. thermautotrophicus* Z-245 (1750 homologous CDS). For the proteomics analysis, a protein FASTA file of the pan-genome was required and created in the same way as described in **Chapter 3**, using only the protein FASTA files for *M. thermautotrophicus* ΔH pMVS1111A:P$_{hmtB}$-*fdh*$_{Z\text{-}245}$ and *M. thermautotrophicus* Z-245.

4.5.15 Eha/Ehb ratio determination

The Eha/Ehb ratios were determined as in **Chapter 3**.

4.5.16 Methanogenesis relative abundances

Methanogenesis relative abundances were determined as in **Chapter 3**.

4.5.17 Genome-scale metabolic modeling

The *M. thermautotrophicus* ΔH and *M. thermautotrophicus* Z-245 models reconstructed in **Chapter 3** were used. The *M. thermautotrophicus* ΔH model was adapted for *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ by adding the F_{420}-dependent formate dehydrogenase cassette genes to the FDH_F420 reaction, thus activating the reaction (**Data S1**). This adjustment was performed using COBRApy version 0.22.1 (Ebrahim *et al.*, 2013). The integration of the fermentation data into the models was performed as in **Chapter 3** (**Data S1**).

4.5.18 Data availability

The transcriptomics data (gene expression data) will be deposited to the NCBI's Gene Expression Omnibus. The proteomics DDA data will be deposited to the ProteomeXchange Consortium *via* the PRIDE partner repository (Perez-Riverol *et al.*, 2019; Deutsch *et al.*, 2016, 2020). The adjusted modeling files, including the modified *M. thermautotrophicus* ΔH model, the gas fermentation data used to constrain the model (as Excel® files), and the individual microbe and differential expression analyses data for both the transcriptomics and proteomics, which were used to constrain the model (as Excel® files) and can be found in a GitHub repository along with additional scripts and programs at the following link: `https://github.com/isacasini/Dissertation_Casini/tree/main/Chapter_4`.

4.6 Results and discussion

4.6.1 Comparable production rates for growth on sodium formate with the two strains

The ability of *M. thermautotrophicus* ΔH pMVS1111A:P$_{hmtB}$-*fdh*$_{Z\text{-}245}$ to grow on formate was compared to the closely related native formate-user, *M. thermautotrophicus* Z-245, by operating pure culture continuous bioreactors (**Figure B.1**, **Materials and Methods, 4.5.4**), which were fed with 355 ± 5 mM sodium formate (**Figure 4.1A**, **Materials and Methods, 4.5.3**). The production rates of methane and carbon dioxide were calculated through carbon balances (**Materials and Methods, 4.5.5**). The two microbes did not have significantly different sodium formate uptake rates, or methane, carbon dioxide, and biomass production rates (**Figure 4.1B**, **Data S2**). However, the molecular hydrogen production rate of *M. thermautotrophicus* ΔH pMVS1111A:P$_{hmtB}$-*fdh*$_{Z\text{-}245}$ was significantly higher than that of *M. thermautotrophicus* Z-245, albeit both were low. The gas data, thus, indicated that there is no significant disadvantage to harboring the plasmid for growth on sodium formate.

4.6.2 No detrimental consequence from constitutive expression of the formate dehydrogenase in *M. thermautotrophicus* ΔH pMVS1111A:P$_{hmtB}$-*fdh*$_{Z\text{-}245}$ that was grown on sodium formate

The constitutive promoter on the pMVS1111A:P$_{hmtB}$-*fdh*$_{Z\text{-}245}$ precipitated the continuous replication and transcription of the episomal formate dehydrogenase cassette, even under non-selective conditions. The plasmid itself is relatively small (12.8 kb) compared to the average 119 kb found in Euryarcheoata (Shintani *et al.*, 2015), and contains an additional 14 genes that were effectively added to *M. thermautotrophicus* ΔH pMVS1111A:P$_{hmtB}$-*fdh*$_{Z\text{-}245}$. All the genes were detected in the transcriptomics except for ORF3_Luo and ORF4_Luo, even in the cultures that were grown on molecular hydrogen and carbon dioxide. These two open reading frames correspond to the ORF3 and ORF4 from Luo *et al.* (2001). The former contains helix–turn–helix and ATP/GTP binding motifs that are likely involved in plasmid replication, and latter has an unknown function (Luo *et al.*, 2001). From the proteomics analysis, the gene products for ORF3_Luo, *KanR* (a kanamycin resistance cassette),

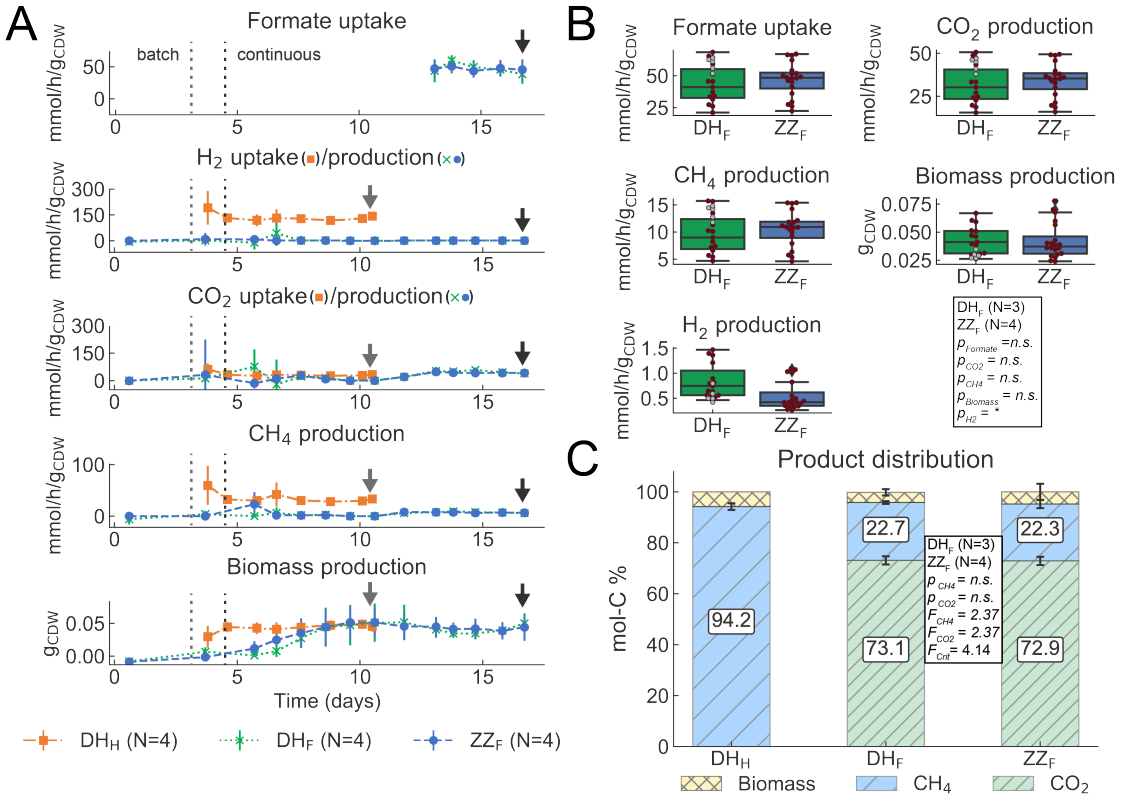

Figure 4.1: Fermentation data from continuous bioreactors with *M. thermautotrophicus* ΔH pMVS1111A:P$_{hmtB}$-*fdh*$_{Z\text{-}245}$ (DH) and *M. thermautotrophicus* Z-245 (ZZ). (A) Gas consumption (H_2 and CO_2 uptake - orange line DH_H) or Na-formate consumption (green and blue lines, DH_F and ZZ_F), H_2 and CO_2 production rates (green and blue lines, DH_F and ZZ_F), and CH_4 and biomass production data from quadruplicate bioreactors for the fermentation period of 11 (DH_H) and 17 (DH_F and ZZ_F) days. Data for further analyses (transcriptomics, proteomics) were taken on days 11 (DH_H) and 17 (DH_F and ZZ_F) as indicated by the gray and black arrows, respectively. Formate uptake rates were only determined once steady state was reached and all formate provided was consumed. **(B)** Average Na-formate consumption and H_2, CO_2, CH_4, and biomass production data during the steady-state period (days 13 to 17) for DH_F and ZZ_F. For statistical analysis in pair-wise comparisons with T-test, all data points were included (red circles) except for the data points from the first bioreactor replicate of DH_F which stopped spinning (gray circles) (**Materials and Methods**, **4.5.11**). **(C)** Average normalized product distribution, including statistical analysis by ANOVA for DH_F and ZZ_F (**Materials and Methods, 3.5.13**. CH_4, methane; CO_2, carbon dioxide; H_2, molecular hydrogen; DH, *M. thermautotrophicus* ΔH pMVS1111A:P$_{hmtB}$-*fdh*$_{Z\text{-}245}$; ZZ, *M. thermautotrophicus* Z-245; F (subscript), Na-formate as substrate; H (subscript), H_2 and CO_2 as substrates; *, $p < 0.05$; n.s., not significant ($p > 0.05$); F, F value; F_{crit}, F critical value.

orf3 (an opening read frame with an unknown function (Nölling and Reeve, 1997)), *fdhB*, *fdhA*, and *fdhC* were detected under both growth conditions. ORF5_Luo (ORF5 from (Luo *et al.*, 2001)), which is a putative excisionase, was additionally found for growth on molecular hydrogen and carbon dioxide. For *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$, the transcription of the formate dehydrogenase cassette was upregulated two orders of magnitude for growth on sodium formate compared to molecular hydrogen and carbon dioxide (**Figure 4.2**, **Data S2**). Further, the protein abundances (based on iBAQ) were one and two orders of magnitude higher for subunits *fdhAB* and *fdhC*, respectively (**Figure 4.2**, **Data S2**). The lack of regulation also resulted in higher transcription of the formate dehydrogenase cassette between *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$. and *M. thermautotrophicus* Z-245 that were grown on sodium formate. However, this difference was not found on the proteome level (**Figure 4.2**, **Data S2**).

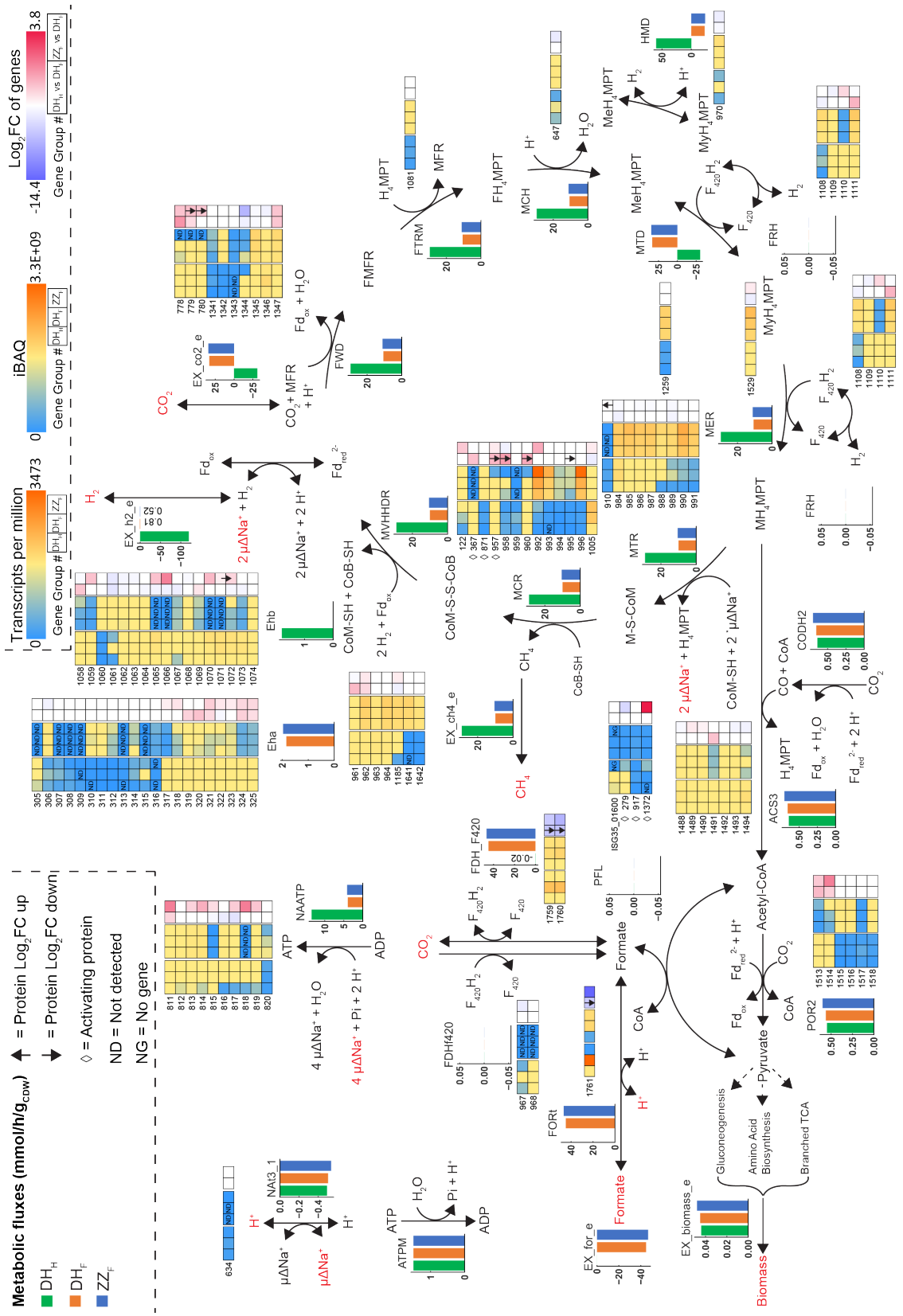

Figure 4.2: (Caption on the next page.)

Figure 4.2: (Previous page.) **Branched Wolfe Cycle adapted from Thauer (2012) with other reactions involved in the energy metabolism.** Fluxes are from the proteomics reduced model, which were constrained with experimental data. Units used are transcripts per million (TPM) for genes, and iBAQ values for proteins. Log_2 fold change (log_2FC) for differentially expressed genes and proteins. The Gene group is used as ID for the omics. For the PFL reaction, only *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ has the gene, thus, the gene ID is used (ISG35_01600). Red text refers to metabolites that are exchanged across the membrane. For reactions Nat3_1, EX_h2_e, and FDH_F420, the (non-zero) fluxes are too small to be identified, and text has been added with the value. **Microbes**: DH, *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$; ZZ, *M. thermautotrophicus* Z-245. **Growth conditions (subscripts)**: H, growth on H_2/CO_2; F, growth on sodium formate. **Compounds**: CH_4, methane, CO, carbon monoxide; CoA, Coenzyme A; CoB, coenzyme B; CoM, coenzyme M; CoM-S-S-CoB, CoM-CoB heterodisulfide; CO_2, carbon dioxide; F, formyl; Fdox/rd, ferredoxin oxidized/ reduced; H_2, hydrogen; H^+, proton; H_4MPT, tetrahydromethanopterin; M, methyl; Me, methenyl; MFR, methanofuran; My, methylene; Na^+, sodium ion. **Reactions/Enzymes**: ATPM, ATP maintenance (pseudo reaction) ; CODHr2, CO dehydrogenase/acetyl-CoA synthase; Eha/Ehb, energy converting hydrogenases; EX_biomass_e, biomass exchange (pseudo reaction); EX_ch4_e, CH_4 exchange (pseudo reaction) ; EX_co2_e, CO_2 exchange (pseudo reaction); EX_h2_e, H_2 exchange (pseudo reaction); FDHf420, F_{420}-dependent formate dehydrogenase; FDH_F420, F_{420}-dependent formate dehydrogenase cassette; FRH, F_{420}-reducing hydrogenase; FTRM, FMFR/H_4MPT formyltransferase; FWD, FMFR dehydrogenase (tungsten- and molybdenum-dependent isozymes); HMD, MeH_4MPT hydrogenase; MCH, MeH_4MPT cyclohydrolase; MCR, MCoM reductase (I and II); MER, MyH_4MPT reductase; MTD, MyH_4MPT dehydrogenase; MTR, MH_4MPT/CoM methyltransferase; MVHHDR, F_{420}-non-reducing hydrogenase with the heterodisulfide reductase; NAATP, ATP synthase; Nat3_1, Na^+/H^+ antiporter; PFL, pyruvate formate-lyase; POR2, pyruvate synthase. **Other**: ◇, activating protein; ND, not detected; NG, no gene.

4.6.3 Multi-level omics reveals differential gene expression and protein abundance patterns between species and conditions

We used transcriptomics and proteomics analyses to study the steady-state bioreactor experiments. We achieved high reproducibility of the replicates for both transcriptomics and proteomics and high coverage of the transcriptome (87.2-95.4%) and proteome (77.7-78.3%) with each microbe (**Table 4.1**, **Data S3**). Further, the lack of a linear correlation between the differential expression analysis of the transcriptome and proteome by species was once again identified (**Figure B.2**, **Chapter 3**).

Table 4.1 Metrics of the transcriptomics and proteomics for the individual analyses and the differential expression analyses for the two microbes under varying growth conditions. Statistical significance was determined by an adjusted P-value ≤ 0.05.

A) Individual analyses				
Omics	**Metric**	**DH_H[1]**	**DH_F[2]**	**ZZ_F[3]**
Transcriptomics	*Million reads*	66.6 ($\pm$ 5.2)	67.4 ($\pm$ 6.4)	76.4 ($\pm$ 9.5)
	Number of CDSs found (%)	1580/1812 (87.20%)	1765/1812 (94.41%)	1721/1804 (95.40%)
Proteomics	*Number of proteins found*	1415/1812 (78.10%)	1407/1812 (77.65%)	1413/1804 (78.33%)
B) Differential expression analyses				
Omics	**Metric**	**DH_H *vs.* DH_F**	**ZZ_F *vs.* DH_F**	
Transcriptomics	*DE*[4] *genes (range of log_2FC)*	796 (-5.82 to 3.90)	1056 (-15.86 to 15.26)	
	DE genes with $\lvert log_2FC \rvert \geq 2$	136	265	
	DE genes that map to a reaction in the GEM	219	314	
Proteomics	*DE proteins (range of log_2FC)*	59 (-6.64 to 6.64)	65 (-6.64 to 6.64)	
	DE proteins with $\lvert log_2FC \rvert \geq 2$	10	43	
	DE proteins that map to a reaction in the GEM	15	20	

[1] DH_H, *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ that was grown on H_2 and CO_2
[2] DH_F, *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ that was grown on Na-formate
[3] ZZ_F, *M. thermautotrophicus* Z-245 that was grown on Na-formate
[4] DE, differential expression

We compared the omics data sets by microbes and growth conditions using the same methods described in **Chapter 3**. Briefly, the genomes were merged to generate a new pan-transcriptome and pan-proteome and then used for the pairwise comparison of: **1)** *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ that was grown on molecular hydrogen and carbon dioxide *vs.* the microbe that was grown on sodium formate (DH_H/DH_F); and **2)** *M. thermautotrophicus* Z-245 *vs.* *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ that were both grown on sodium formate (ZZ_F/DH_F, **Materials and Methods**, **4.5.12** and **4.5.13**). More differentially expressed genes were identified in the pan-transcriptome than differentially abundant proteins in the pan-proteome (**Table 4.1**, **Figure 4.3**).

Most of the proteins in methanogenesis had similar average relative abundances (compared to the whole proteome, **Table S18**, **Data S3**). However, the methyl-tetrahydromethanopterin coenzyme M methyltransferase (Mtr) was more abundant (12.6-14.5 *vs.* 8.0-9.6) in this work than in the earlier experiment by **Chapter 3**. Further, the methyl-coenzyme M reductase isoenzyme I (Mcr) was more abundant than the methyl-coenzyme M reductase isoenzyme II (Mrt) in both this work and that of **Chapter 3**. However, the abundance for the Mcr in *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ and *M. thermautotrophicus* Z-245 that was grown on sodium formate was higher than that of the microbes that were grown on molecular hydrogen and carbon dioxide (8.57 *vs.* 6.79 and 6.29 for the former, and 15.15 *vs.* 9.99 for the latter, **Table S19**).

We analyzed the sodium ion (Na^+) transporting energy-converting hydrogenase (Eha/Ehb) ratios as previously performed in **Chapter 3** (**Table S19**, **Materials and Methods**, **4.5.15**). The Eha was proven essential for growth as it supplements the required ferredoxin for biomass synthesis (Lie *et al.*, 2012). There was an increase in the relative transcript and proteome abundances between the former and the present work. The ratio differences between *M. thermautotrophicus* Z-245 that was grown on sodium formate (this work) and molecular hydrogen and carbon dioxide (**Chapter 3**) were not significant (Table S20). However, the ratios for *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ were significantly different across substrates (this work) and compared to the *M. thermautotrophicus* ΔH wild type that was grown on molecular hydrogen and carbon dioxide from **Chapter 3** (**Table S20**). This increase in the relative abundance of Eha supports an increase in biomass to methane production ratios found in *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ (**Figure B.3**).

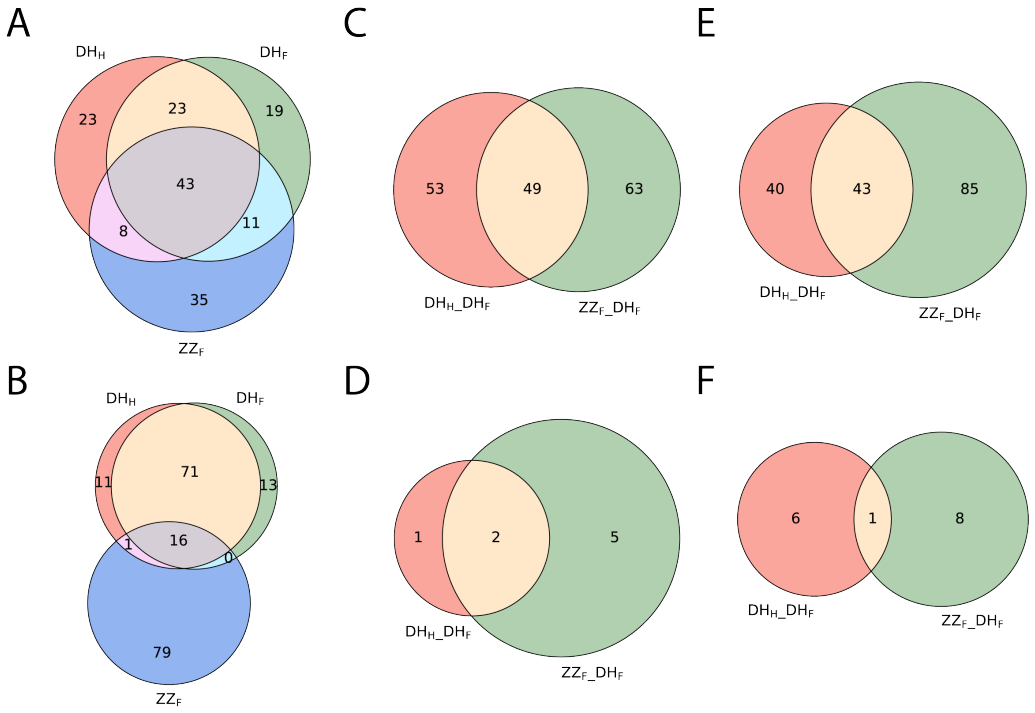

Figure 4.3: Overview of transcriptomics and proteomics interspecies comparison. **(A)** Comparison of the top 100 transcribed genes (by average number of transcripts per million). **(B)** Comparison of the top 100 abundant proteins (by average iBAQ value). Transcripts/proteins that do not map to a gene group are counted only once collectively, hence there are less than 100 genes total in A and B. **(C)** Upregulated transcripts; **(D)** more abundant proteins; **(E)** downregulated transcripts; **(F)** less abundant proteins, respectively, from the pairwise differential expression analysis that map to reactions in the GEMs. DH, *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$; ZZ, *M. thermautotrophicus* Z-245; F (subscript), Na-formate as substrate; H (subscript), H_2 and CO_2 as substrates.

4.6.4 Multi-level data and lower modeled non-growth-associated maintenance energy indicate an advantage of the formate dehydrogenase for growth

We previously hypothesized that *M. thermautotrophicus* ΔH wild type produces formate for biomass using the pyruvate formate-lyase (Pfl) rather than its formate dehydrogenase (FdhAB, not the cassette) (**Chapter 3**). On the contrary, but consistent with the results of **Chapter 3**, only the formate dehydrogenase cassette was found to be active in *M. thermautotrophicus* Z-245 for formate production when grown on molecular hydrogen and carbon dioxide (**Chapter 3**) and for formate consumption when grown on sodium formate (this chapter, **Data S3**). Further, *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ changed its formate metabolism to mirror that of *M. thermautotrophicus* Z-245. It altered its formate production mechanism from the pyruvate formate-lyase to the formate dehydrogenase cassette when grown on molecular hydrogen and carbon dioxide, and it used the cassette exclusively for formate consumption when grown on sodium formate. While the *pfl* was transcribed and translated under both growth conditions, the relative abundances (TPM and iBAQ, respectively) were less than those of the formate dehydrogenase cassette. The production of formate by the formate dehydrogenase cassette for growth on molecular hydrogen and carbon dioxide, may have created a regulatory effect on the *pfl*. Additionally, the presence of formate as a growth substrate would have also removed the necessity of the pyruvate formate-lyase for formate generation. Further, the native *fdhAB* was also transcribed under both growth conditions, but no proteins were detected (**Figure 4.2**, **Data S3**).

Fink *et al.* (2021) found that the plasmid was stable for 21-28 generations under both non-selective and selective (Na-formate) batch growth conditions. Under continuous cultivations on molecular hydrogen and carbon dioxide (non-selective), after 16 generations, the plasmid was still detected, and on sodium formate, the plasmid was still found after 26-27 generations (**Figure B.1**). However, detection of the plasmid *via* PCR does not confirm the presence of the plasmid in the cells. Additional experiments would be required to know this for certain and to determine the plasmid copy number. The formate dehydrogenase encoded on the plasmid was favored even under non-selective conditions. This raises the question whether the formate dehydrogenase provided a larger metabolic benefit than a disadvantage, even for growth on molecular hydrogen

and carbon dioxide. Albeit, thus far, we do not know of any regulatory mechanism for the formate dehydrogenases.

While only molecular hydrogen production rates (of the gas data) were significantly different between *M. thermautotrophicus* ΔH pMVS1111A:P$_{hmtB}$-*fdh*$_{Z\text{-}245}$ and *M. thermautotrophicus* Z-245 when cultivated on sodium formate, there may be some differences between *M. thermautotrophicus* ΔH wild type and *M. thermautotrophicus* ΔH pMVS1111A:P$_{hmtB}$-*fdh*$_{Z\text{-}245}$ that was grown on molecular hydrogen and carbon dioxide (**Figure B.3**). It is not optimal to compare the results of this experiment with the one in **Chapter 3**, particularly given the addition of 0.0015 mM sodium selenate and 0.015 mM sodium molybdate dihydrate to the continuous media in this work; nevertheless, the other cultivation conditions were maintained constant (*e.g.*, gassing rates (10 ml/min) and concentrations (H_2/CO_2 80/20 %v/v), and dilution rates (0.046 $\pm$ 0.003-0.005)). The T-test analysis (**Figure B.3B**) indicated that the growth of *M. thermautotrophicus* ΔH pMVS1111A:P$_{hmtB}$-*fdh*$_{Z\text{-}245}$ resembles more that of *M. thermautotrophicus* Z-245 and *M. marburgensis* Marburg rather than the unmodified *M. thermautotrophicus* ΔH. The adenosine triphosphate (ATP) dissipation or the non-growth-associated maintenance energy (reaction ATPM) estimated through the flux balance analyses (FBA, **Data S4**, **Materials and Methods**, **4.5.17**) also resembles that of *M. thermautotrophicus* Z-245 and *M. marburgensis* Marburg (**Table 4.2**). The modeled ATPM flux is lower than *M. thermautotrophicus* ΔH wild type, which is unusual given the added cost of producing, transcribing, and translating the plasmid.

Further, based on FBA results, the modeled non-growth-associated maintenance energy was even lower for growth on sodium formate (**Tables 4.2**, **Data S4**). This means that more energy was directed into biomass production compared to methane generation. This is favorable from an evolutionary growth perspective. However, this may be a disadvantage if methane is the target product. The addition of molecular nitrogen (N_2, instead of molecular hydrogen) for growth on formate did not appear to alter the proteome and downstream metabolism, thus, also not providing a benefit (**Text B.1.1**, **Table S21**). While these results may indicate a growth advantage to having the formate dehydrogenase cassette, additional experiments will be required to ascertain these differences in maintenance costs and the impact of the plasmid on the metabolism of the microbes.

Table 4.2 ATPM flux ranges for different constraints. Range considers the ATPM flux with the objective function of biomass production maximization to ATPM maximization.

DH^{1}_{H}	DH^{2}_{F}	ZZ^{3}_{F}
Model[4]: 1.5-11.99	Model: 1.5-1.50	Model: 1.5-3.43
Model trans[5]: 1.5-11.44	Model trans: 1.5-1.5	Model trans: 1.5-3.43
Model prot[6]: 1.5-11.99	Model prot: 1.5-1.5	Model prot: 1.5-3.43

[1] DH_H, *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ that was grown on H_2 and CO_2
[2] DH_F, *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ that was grown on Na-formate
[3] ZZ_F, *M. thermautotrophicus* Z-245 that was grown on Na-formate
[4] Model, model only constrained with experimental data
[5] Model trans, model additionally constraint with the transcriptomics data
[6] Model prot, model additionally constraint with the proteomics data

4.6.5 Formate as a substrate is theoretically advantageous for electron but not carbon yield

Compared to gaseous substrates, formate provides the following benefits (the disadvantages are discussed later): as a solid or liquid, it is easily stored and transported and has a higher solubility than molecular hydrogen and carbon dioxide (Claassens *et al.*, 2019). Stoichiometrically, the theoretical product (methane) to substrate yield ratios are the same (0.25) for both, growth on molecular hydrogen and carbon dioxide and on formate (formic acid) (**Equations (4.1)** and **(4.2)**). However, the lower heat of combustion of formic acid compared to molecular hydrogen (245 kJ/mol *vs.* 260 kJ/mol) leads to a higher theoretical energy efficiency for methane production from formic acid (87.75% *vs.* 82.69%) (**Materials and Methods**, **4.5.10**). Claassens *et al.* (2019) calculated energy efficiencies using published experimental values for methanogenic growth on molecular hydrogen and carbon dioxide (79-86%) and formic acid (81-86%). It is important to note that not all the experiments were performed using minimal media, and some included supplements such as vitamin solutions.

Regardless of the differences in the theoretical energy efficiency, the theoretical ATP yield, 0.5 mol/methane, is the same for growth on formate and molecular hydrogen and carbon dioxide (Schlegel and Müller, 2013; Thauer *et al.*, 2008). Nevertheless, when calculating this ATP yield from experimental constrained FBAs (fluxes of reactions NAATP/EX_ch4_e), the resulting values are less than 0.5, with those for formate often lower (**Table 4.3**). During methanogenesis, the Mtr produces the sodium ion (Na^+) gradient that is used to power the ATP synthase. However, the Na^+ gradient is also

consumed by the Eha/Ehb to produce reduced ferredoxin. The reduced ferredoxin mediates and provides electrons for important metabolic reactions including reactions necessary for biomass synthesis. When more biomass is produced more ATP is consumed. Thus, the higher the biomass to methane yield (mol-C %), the lower the ATP to methane yield (**Table 4.3**).

Table 4.3 Energy efficiencies, ATP yields (from proteome constrained FBAs, Data S2 Chapter 3, Data S4 Chapter 4), and their statistical differences from on ANOVA based on substrate

Energy Efficiency					
	DH^1_H		**DH^2_F**	**ZZ^3_F**	
Product Yield (Y)	0.24		0.223	0.225	
Energy Efficiency (E)[4]	0.7939		0.7828	0.7898	
ATP Ratios					
Ratio	**DH_H**	**DH_F**	**ZZ_F**	**DH**[5]	**ZZ**[6]
ATP/H_2	0.112[7] (0.112[8], 0.110[9])	N/A	N/A	0.116 (0.116, 0.115)	0.112 (0.112, 0.111)
ATP/(H_2)*4	0.450 (0.450, 0.441)	N/A	N/A	0.464 (0.464, 0.460)	0.448 (0.448, 0.443)
ATP/CO_2	0.436 (0.436, 0.427)	N/A	N/A	0.454 (0.454, 0.450)	0.434 (0.434, 0.429)
ATP/Formate	N/A	0.088 (0.088, 0.088)	0.088 (0.090, 0.090)	N/A	N/A
ATP/CH_4	0.465 (0.465, 0.457)	0.395 (0.396, 0.395)	0.393 (0.399, 0.398)	0.475 (0.475, 0.471)	0.463 (0.463, 0.458)
Biomass yield mol-C % [10]	5.8	15.6	17.7	3.6	5.9
ANOVA: H_2/CO_2 *vs.* Formate [11]					
FBA objective function	**Biomass max**		**ATP max**	**Biomass max loopless**	
P-value	0.0006		0.0007	0.0017	

[1] DH_H, *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ that was grown on H_2 and CO_2
[2] DH_F, *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ that was grown on Na-formate
[3] ZZ_F, *M. thermautotrophicus* Z-245 that was grown on Na-formate
[4] E = Y*(heat of combustion CH_4/(heat of combustion H_2 or Na-formate)) CH_4: 860 kJ/mol; H_2: 260 kJ/mol; Na-formate: 245 kJ/mol)
[5] DH, *M. thermautotrophicus* ΔH that was grown on H_2 and CO_2 from **Chapter 3**
[6] ZZ, *M. thermautotrophicus* Z-245 that was grown on H_2 and CO_2 from **Chapter 3**
[7] First value objective function: maximization of biomass
[8] Second value objective function: maximization of ATPM
[9] Third value objective function: maximization of biomass with a loopless
[10] Calculated as: biomass/(biomass + CH_4) using the same values as depicted in the product distributions (biomass and CH_4) in **Figure 4.1C** and **Figure B.3**
[11] Comparison with ANOVA of the ATP/CH_4 ratios for H_2 and CO_2 *vs.* Na-formate at different objective functions, **Data S4**

Despite the advantages of working with formate as a substrate, some physiochemical technicalities can make it a challenging substrate. Formate as the uncharged acid, formic acid, is more suitable for transport across the membrane. However, it requires a low pH

to be the dominant species (pKa 3.75). This limits the ability to use formic acid as a substrate for many microbes, including *M. thermautotrophicus*, which has a pH range of 6.0-8.8 (Wasserfallen *et al.*, 2000). Sodium formate is highly soluble in water with a solubility of 97.2 g/100 mL (14 M) at 20°C (PubChem). The ability to reach high concentrations facilitates its storage and feeding as a substrate in continuous media for bioreactors. Nevertheless, the consumption of one mole of formate produces one mole of base (*e.g.*, NaOH, **Equation (4.5)**, (Costa *et al.*, 2013a)). This increase in pH requires the addition of acid, which would also contribute to the dilution of the system. The quantity of acid is not negligible and accounted for 25% v/v of the continuous liquid flow through the reactor using 1 N HCl (**Materials and Methods**, **4.5.8**). Nevertheless, high concentrations of formate are not possible to use because of toxicity starting at 100 mM, likely due to the need to transport a proton across the membrane along with the substrate (or the protonated substrate) (Claassens *et al.*, 2019), though some studies have used higher concentrations (200 mM, (Fink *et al.*, 2021; Müller *et al.*, 2021), 380 mM (Costa *et al.*, 2013a; Long *et al.*, 2017). Transporting protons affects the sodium ion gradient that controls the ATP synthase (**Text B.1.3**). *M. thermautotrophicus* ΔH typically maintains an intracellular of 0.3 pH units below the extracellular pH (Jarrell and Sprott, 1981). This creates a proton gradient that supports the transport of protons into the cell, which in turn, through the Na^+/H^+ antiporter, simultaneously pumps out sodium ions. Lastly, both formic acid and formate create a salty broth which complicates and limits bioreactor effluent recycling opportunities. Cultivation on molecular hydrogen and carbon dioxide does not have this disadvantage.

$$4\,NaCOOH + 2\,H_2O \rightarrow CH_4 + 3\,CO_2 + 4\,NaOH \tag{4.5}$$

Using molecular hydrogen and carbon dioxide as substrates permits the optimal stoichiometric ratio to produce methane (four molecular hydrogen and one carbon dioxide). However, when using formate (or formic acid), the electron to carbon ratio is predetermined (1:1), which is not ideal for methane production, and thus produces three moles of carbon dioxide (in excess). While it is easier to use all the electrons (given the aforementioned reasons), carbon dioxide is released, potentially making molecular hydrogen and carbon dioxide a better substrate choice when both electrons and carbon yields are important. The challenges of gas-liquid mass transfer can be overcome, as demonstrated with the 1.5 MW demonstration plant in Denmark by Electrochaea

GmbH. Using a nine-meter tall continuous stirred-tank reactor with a gas retention time of three min and pressurized to 4-9 bar (gauge), a methane content of 99% in the off-gas was achieved (Rusmanis *et al.*, 2019). Nevertheless, supplementation of molecular hydrogen to the formate could reoptimize the electron to carbon ratio and has already been modeled as a co-substrate (Goyal *et al.*, 2014). In conclusion, formate is advantageous compared to molecular hydrogen and carbon dioxide in regards to its ability to be stored, transported and dissolved, and its electron accessibility. However, physio-chemical nature of formate creates some biological challenges that may not make it such a practical substrate. Further, increased biomass compared to methane production on formate is not ideal when targeting methane production. Thus, the ability to optimize gas ratios, increase gas-liquid mass transfer rates, prolong gas retention times, and reach higher methane to biomass ratios, may be sufficient to favor molecular hydrogen and carbon dioxide as substrates for industrial-scale biomethanation.

4.7 Summary and outlook

the conversion of power to gas can proceed directly from excess renewable electric power to molecular hydrogen and then methane *via* the hydrogenation of carbon dioxide. However, formate can also be generated as an intermediate compound, which allows both the energy and carbon to be more easily stored and transported in a liquid or solid state. This attribute means that on-site electrical to chemical energy conversion (molecular hydrogen production) can have a more flexible final product (methane gas) generation step off-site and even at a different time point. In this study, we investigated how a genetically modified *M. thermautotrophicus* ΔH (*M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$) that is able to use formate as a sole substrate performs on both its native substrates (molecular hydrogen and carbon dioxide) and formate. We compare the performance to a closely related native formate user, *M. thermautotrophicus* Z-245, and to *M. thermautotrophicus* ΔH wild type. We implemented genomics, transcriptomics, and proteomics from samples that we extracted at steady state from quadruplicate continuous bioreactors. With our cultivation conditions, we identified few disadvantages in performance of the different strains. *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ performed almost identically to *M. thermautotrophicus* Z-245 for growth on formate, and the

transcriptomics and proteomics revealed no more differences than were found between the two wild-type strains in **Chapter 3**. We detected that on the transcriptomics level, but not the proteomics level, the substrate had a more relevant influence on the top 100 abundant genes and proteins than the identity of the strain.

Further, consistent with previous findings, we confirmed the stability of the plasmid that harbors the F_{420}-dependent formate dehydrogenase cassette in *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$, even under non-selective growth conditions (molecular hydrogen and carbon dioxide). However, using the omics data, we could infer not only the stability but also the preference for the heterologous formate dehydrogenase in the anabolism, even for growth on molecular hydrogen and carbon dioxide. We found in our modeling a decrease in the non-growth-associated maintenance energy by constraining the GEMs with the gas fermentation and omics data. This supported the postulated preference for the formate dehydrogenase. A decrease in non-growth-associated maintenance energy (even under cultivations with molecular hydrogen and carbon dioxide) indicates a higher biomass to methane ratio, which is provides a growth advantage for the microbe.

A solid comprehension of microbial energy and central carbon metabolism is beneficial for efficiently and effectively exploit microbes for biotechnological purposes. While some methanogens, including *M. thermautotrophicus* ΔH, have been well studied during the last 50 years to investigate methanogenesis, our understanding of these microbes is still far behind others, particularly microbes that are used as chassis for biotechnological applications. The recent release of the GEMs for three thermophilic *Methanothermobacter* spp. (**Chapter 3**), in conjunction with the recent establishment of a genetic system for *M. thermautotrophicus* ΔH (Fink *et al.*, 2021) is forging the way for the use of these thermophiles in biotechnology beyond power-to-gas system. The work presented in this paper is the first example merging the results of those two studies and another extensive set of multi-level omics data, which the scientific community can continue to use for subsequent analyses.

4.8 Acknowledgments

This work was supported by the Humboldt Foundation in the framework of the Humboldt professorship, which was awarded to Largus T. Angenent (L.T.A.), and by

the Max Planck Institute for Biology Tübingen (Ruth E. Ley (R.E.L.) and L.T.A.). The authors would like to thank Nicolai Kreitli and Lucas Mühling for helpful input and support with experimentation. The authors would like to acknowledge the assistance of the Proteome Center Tübingen (PCT) for their support with the proteomics.

4.9 Supplementary information

Supplementary information: supplementary texts and figures are found in in **Appendix B**. Details and the GitHub link for the additional supporting material (modeling, fermentation data, omics data and scripts) can also be found in **Appendix B**.

Chapter 5

Reproduced and adapted with permission from: Xia, P.F., Casini, I., Schulz, S., Klask, C.M., Angenent, L.T., and Molitor, B. (2020). Reprogramming acetogenic bacteria with CRISPR-targeted base editing *via* deamination. *ACS Synth. Biol.* 9(8), 2162-2171.

5.1 Author contributions

Peng-Fei Xia (P.F.X.) designed the experiments, performed all genetic engineering, and evaluated the data. **Isabella Casini (I.C.) performed the computational work, including the design, writing, and execution of the program used (`https://github.com/isacasini/SNV_Xia_et_al_2020`). I.C. produced the data for Figure 5.3, Table C.3, and the online supplemental Excel® Data set `https://pubs.acs.org/doi/10.1021/acssynbio.0c00226`.** Christian-Marco Klask (C.M.K.) and Sarah Schulz (S.S.) performed the cultivation experiments and analyzed those samples. Largus T. Angenent (L.T.A.) and Bastian Molitor (B.M.) supervised the work. P.F.X. and B.M. wrote the manuscript. All authors edited and revised the written text before submission.

5.2 Abstract

Acetogenic bacteria are rising in popularity as chassis microbes in biotechnology due to their capability of converting inorganic one-carbon (C1) gases to organic chemicals. To fully uncover the potential of acetogenic bacteria, synthetic-biology tools are imperative to either engineer designed functions or to interrogate the physiology. Here, we report a genome-editing tool at a one-nucleotide resolution, namely base editing, for acetogenic

bacteria based on CRISPR-targeted deamination. This tool combines nuclease deactivated Cas9 with activation-induced cytidine deaminase to enable cytosine-to-thymine substitution without DNA cleavage, homology-directed repair, and donor DNA, which are generally the bottlenecks for applying conventional CRISPR-Cas systems in bacteria. We designed and validated a modularized base-editing tool in the model acetogenic bacterium *Clostridium ljungdahlii*. The editing principles were investigated, and an *in-silico* analysis revealed the capability of base editing across the genome and the potential for off-target events. Moreover, genes related to acetate and ethanol production were disrupted individually by installing premature STOP codons to reprogram carbon flux towards improved acetate production. This resulted in engineered *C. ljungdahlii* strains with the desired phenotypes and stable genotypes. Our base-editing tool promotes the application and research in acetogenic bacteria and provides a blueprint to upgrade CRISPR-Cas-based genome editing in bacteria in general.

5.3 Introduction

Global climate change is challenging the future of human societies, resulting in the need for a sustainable food supply and greener synthesis of fuels and chemicals. One possible solution is by applying biotechnology to convert inorganic one-carbon (C1) gases, such as carbon dioxide (CO_2) and carbon monoxide (CO), into protein, biofuels, and commodity chemicals (Köpke *et al.*, 2010; Richter *et al.*, 2016; Ueki *et al.*, 2014; Molitor *et al.*, 2019). Both gases are already available in large quantities, including in synthesis gas (syngas) and industrial waste gases (Molitor *et al.*, 2016a). Many studies have found that the model acetogenic bacterium *Clostridium ljungdahlii* can convert these gases with hydrogen gas (H2) into mainly acetate and ethanol *via*the Wood-Ljungdahl pathway (Köpke *et al.*, 2010; Müller, 2019). LanzaTech, Inc. (Skokie, IL, USA) has already completed the industrial scale-up by utilizing a closely related acetogenic bacterium (*Clostridium autoethanogenum*). However, synthetic biology and metabolic engineering are imperative to improve the productivity further and to expand the product spectrum (Köpke *et al.*, 2010; Ueki *et al.*, 2014; Banerjee *et al.*, 2014; Leang *et al.*, 2013). Currently, the lack of efficient genome-editing tools delays the progress at the molecular level to optimize acetogenic bacteria for biotechnology.

Clustered regularly interspaced short palindromic repeats (CRISPR)-Cas-based genome editing is a RNA programmable, precise, and robust approach for gene perturbation, and has been applied in a plethora of living organisms (Knott and Doudna, 2018), revolutionizing science. Recently, CRISPR-Cas systems were also adapted to be functional in acetogenic bacteria for gene deletion, insertion, and regulation. To date, different CRISPR-Cas systems (*e.g.*, Cas9 and Cas12a) have been established in *C. ljungdahlii* (Huang *et al.*, 2016, 2019; Zhao *et al.*, 2019), *Eubacterium limosum* (Shin *et al.*, 2019), and *C. autoethanogenum* (Nagaraju *et al.*, 2016). For these systems, first, the Cas protein (*e.g.*, Cas9 from *Streptococcus pyogenes*) is targeted to a highly specific site on the genome by a guide RNA (gRNA). Then, the Cas protein cleaves the genomic DNA at this site and introduces a double-strand break. To survive, the cell has to repair the double-strand break through DNA repairing mechanisms such as homology-directed repair or non-homologous end joining (Selle and Barrangou, 2015). For homology-directed repair, a donor DNA has to be provided as a template, which has to contain homologies to the genome on both sides of the double-strand break. Depending on the design of the donor DNA, it is possible to generate a variety of desired mutations such as point mutations, gene deletions, and gene insertions. The homology-directed repair of the double-strand break results in genome editing at the target site without leaving a selective marker (*i.e.*, antibiotic resistance gene) and scar.

However, CRISPR-Cas-based genome editing is generally challenging in bacteria, because the Cas nuclease is often toxic to bacteria, and bacteria typically lack efficient homology-directed repair or non-homologous end joining machineries to repair the double-strand break (Vento *et al.*, 2019). Therefore, it is essential that a sufficient number of cells receive the CRISPR-Cas system to ensure that enough cells survive a DNA cleavage by undergoing the inefficient homology-directed repair process with donor DNAs (Selle and Barrangou, 2015). This renders the CRISPR-Cas system even more difficult for acetogenic bacteria, which are typically recalcitrant to receiving foreign DNA (Huang *et al.*, 2016; Zhao *et al.*, 2019; Molitor *et al.*, 2016b). Consequently, the process of cleavage-and-repairing, which is typically considered the important advantage of conventional CRISPR-Cas systems, becomes a bottleneck to perform CRISPR-Cas-based genome editing in acetogenic bacteria.

Lately, a new CRISPR-Cas-based genome-editing tool, namely base editing, was developed by combining a CRISPR-Cas system with a deamination system to achieve genome editing at a one-nucleotide resolution without the necessity for DNA cleavage,

homology-directed repair, and donor DNA (Rees and Liu, 2018; Gaudelli *et al.*, 2017; Komor *et al.*, 2016; Nishida *et al.*, 2016). By creating a fusion of a nuclease impaired Cas protein (*i.e.*, nuclease deactivated Cas9, dCas9) and a deaminase, this tool generates cytosine (C) to thymine (T) substitutions with cytidine deaminase (**Figure 5.1A**), or adenine (A) to guanine (G) substitutions with adenosine deaminase (Rees and Liu, 2018; Molla and Yang, 2019). Base editing provides distinctive advantages for genome editing in acetogenic bacteria by circumventing the bottlenecks of conventional CRISPR-Cas systems in bacteria. One such advantage is that the required DNA-uptake ability of acetogenic bacteria is minimized. Despite a few principal demonstrations in other bacteria (Wang *et al.*, 2018; Chen *et al.*, 2018; Gu *et al.*, 2018; Tong *et al.*, 2019; Li *et al.*, 2019; Banno *et al.*, 2018) the potential of base editing in acetogenic bacteria has not yet been unraveled.

Here, we developed a modularized base-editing tool for acetogenic bacteria by coupling dCas9 from *S. pyogenes* with activation-induced cytidine deaminase from the sea lamprey *Petromyzon marinus* (Banno *et al.*, 2018). Efficient base editing was validated, and the editing principles were investigated in the model acetogenic bacterium *C. ljungdahlii*. Genome-scale *in-silico* analysis revealed the capability of our base-editing tool and the potential for off-target events. As a first application, we employed base editing to reprogram the distribution of the carbon flux from acetyl-CoA to acetate and ethanol during heterotrophic and autotrophic fermentation, linking designed single-nucleotide variations with industrially relevant bacteria. Our base-editing tool will promote the research and application of C1 utilization with acetogenic bacteria, and more generally, provides an example for upgrading CRISPR-Cas-based genome-editing tools in bacteria.

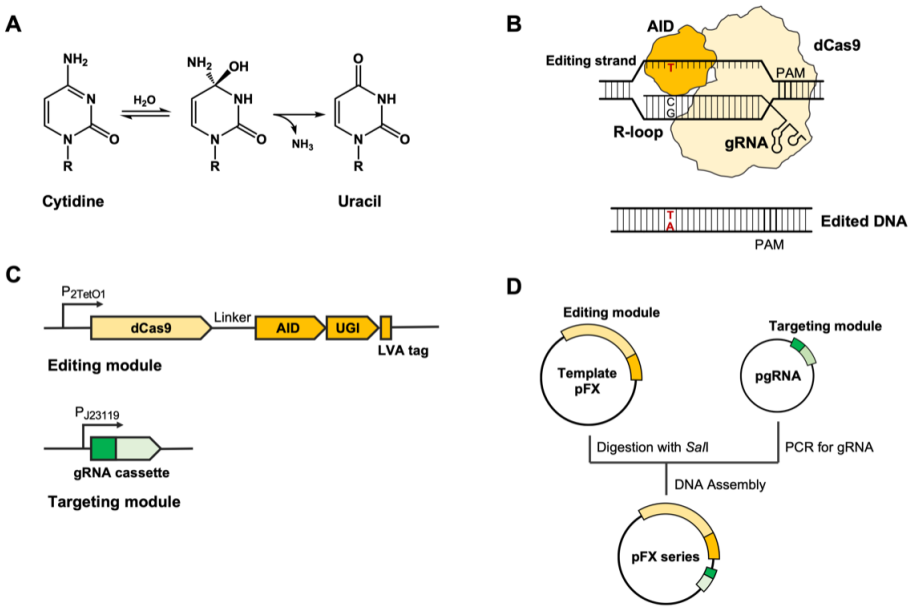

Figure 5.1: Design of base-editing tool in *C. ljungdahlii*. **(A)** Chemistry of deamination process converting cytidine to uracil. **(B)** Mechanism of base editing. Targeted by a gRNA, dCas9 binds to the target DNA and forms an R-loop. Activation-induced cytosine deaminase deaminates the Cs in the single strand DNA in the R-loop (Editing strand), resulting in C-to-T single-nucleotide variations in the genome. **(C)** The editing module consists of dCas9, activation-induced cytosine deaminase, uracil glycosylase inhibitor, and Leu-Val-Ala tag under the control of an inducible tetR-Ptet system, and the targeting module contains the gRNA cassette under the control of the constitutive PJ23119 promoter. **(D)** Modularized strategy to generate editing plasmid series. To generate an editing plasmid, an inverse PCR is employed to generate the gRNA using pgRNA as a template. Template pFX plasmid is digested with SalI and then assembled with the amplified gRNA cassette, resulting in the editing plasmid (pFX series). PAM, protospacer adjacent motif; AID, activation-induced cytosine deaminase; UGI, uracil glycosylase inhibitor; and LVA tag, Leu-Val-Ala tag.)

5.4 Materials and Methods

5.4.1 Strains and media

C. ljungdahlii DSM13528 was used as the wild-type strain, and all *C. ljungdahlii* strains that were used in this study are listed in **Table C.4**. Reinforced clostridial medium (RCM) was employed for general cultivation of *C. ljungdahlii* (Molitor *et al.*, 2016b). Modified PETC medium (ATCC 1754 medium) was used for fermentation experiments with 2 g/L (10 mM) of 2-(N-morpholino)ethanesulfonic acid buffering the medium instead of $NaHCO_3$. For heterotrophic fermentation, 5 g/L (27.8 mM) of fructose were

added in 250 mL serum bottles with 100 mL of medium as the carbon source. For autotrophic fermentation, an H_2/CO_2 mixture (80/20 vol-%, 1.5 bar) was added to the headspace of 1 L cultivation bottles (Pressure plus Duran bottle, VWR) with 100 mL of medium. The headspace was refilled to 1.5 bar every 24 h during fermentation. Clarithromycin (5 μg/L) was added when necessary. The manipulation of *C. ljungdahlii* was performed in an anaerobic chamber (UNILab Pro Eco, MBraun, Germany) with an O_2 level below 10 ppm. The cultivation was performed at 37°C. *E. coli* TOP10 (Invitrogen) was used for general cloning and gene manipulation. *E. coli* was cultivated in Luria-Bertani (LB) medium. Ampicillin (100 μg/L), spectinomycin (60 μg/L), and erythromycin (400 μg/L on plates and 250 μg/L for liquid medium) were used to select and maintain plasmids in *E. coli*.

5.4.2 Plasmid construction

The plasmids used in this study are summarized in **Table C.5**. pMTLdCas9 was a generous gift from Gregory Stephanopoulos (Woolston *et al.*, 2018), pTargetF (Addgene plasmid # 62226) was a gift from Sheng Yang (Jiang *et al.*, 2015), and pScI_dCas-CDA-UL (Addgene plasmid # 108551) was a gift from Akihiko Kondo (Banno *et al.*, 2018). The pFX template was constructed by flanking the fragment consisting of activation-induced cytidine deaminase, uracil glycosylase inhibitor, and Leu-Val-Ala tag from pScI_dCas-CDA-UL with dCas9 on pMTLdCas9. The pgRNA01 plasmid was generated *via*inverse PCR using pTargetF as a template to recover the original P_{J23119} promoter and create the gRNA01 cassette, generating pgRNA01 (**Figure C.1**). Then, the pgRNA plasmid series containing the designed gRNA cassettes were constructed *via*inverse PCR using pgRNA01 as a template. Based on the pFX and pgRNA plasmid series, a modularized method was employed to construct the pFX plasmid series for base editing. First, the pFX template was digested with *SalI* and the gRNA cassette was amplified by primers EBT-PFX-88 and -89. Second, the pFX plasmid series was generated by combining these two parts *via* Gibson assembly (New England labs, NEB) (**Figure 5.1D**). Plasmids were methylated *via* co-transformation with pANA1, which carries ϕ3tI methyltransferase before transformation of *C. ljungdahlii* with the respective plasmid (Molitor *et al.*, 2016b). PCRs were performed using Q5 polymerase (NEB). The primers (Integrated DNA Technologies, BVBA, Leuven, Belgium) for plasmid construction are summarized in **Table C.6**.

5.4.3 Transformation of *C. ljungdahlii*

Transformation of *C. ljungdahlii* was following a modified protocol previously established (Molitor *et al.*, 2016b). Briefly, RCM was inoculated and transferred twice with *C. ljungdahlii* at a 1:100 dilution. When the OD_{600} in the second culture reached 0.2 - 0.4, 12 - 16 mL of culture were harvested by stepwise centrifugation at 10,000 rpm for 1.5 min (mySPIN™12 Centrifuge, Thermo Fisher Scientific). The cells were then washed with ice-cold glycerol (10 vol-%) for three times and resuspended in 200 μL of 10 vol-% glycerol as electrocompetent cells. The methylated plasmids (2 μg) were mixed with the competent cells individually for electroporation (Micropulse system, 2.5 kV, 600 Ω and 25 μF, Bio-Rad) in 2-mm cuvettes. After electroporation, the cells were immediately transferred to pre-warmed RCM for a 12 - 18 h recovery. Subsequently, RCM medium with 5 μg/L of clarithromycin was inoculated with the cells, and the outgrowth cultures were used directly for base editing.

5.4.4 Base editing in *C. ljungdahlii*

The methylated pFX plasmid was transformed into *C. ljungdahlii* according to the protocol above. After electroporation, fresh RCM with 5 μg/L of clarithromycin and 100 ng/mL of anhydrotetracycline was inoculated (1:10 dilution) with the outgrowth culture (RCM with only clarithromycin) in a Hungate-type anaerobic culture tube and incubated for 18 - 24 h to induce base editing. After the induction, the culture was mixed with molten RCM agar (1.0 weight-%) with clarithromycin and poured into petri dishes. Single colonies were picked for further analysis. First, a colony PCR was conducted using Phire Plant Direct PCR Master Mix (Thermo Fisher Scientific) to amplify the fragment containing the target DNA sequence. Second, the fragment was sequenced to verify the single-nucleotide variations. Primers for amplifying the edited DNA fragments and sequencing are listed in **Table C.7**.

5.4.5 Plasmid curing

To cure the cells from the plasmid, a colony with the designed single-nucleotide variations was used to re-inoculate fresh RCM without clarithromycin and transferred for a second time. The second culture was poured with RCM agar (1.0 weight-%) at different dilutions. Subsequently, single colonies were picked to determine the loss of

plasmids using colony PCR. If the colony PCR showed no PCR signal, the colony was further tested in RCM and RCM with clarithromycin to identify the curation from the plasmid. A colony that: 1) carries the designed single-nucleotide variations; 2) has no PCR signal for the editing plasmid; and 3) fails to grow in selective medium, was regarded as an edited strain for subsequent stability evaluation and fermentation experiments.

5.4.6 Serial transfer experiments

To test the stability of single-nucleotide variations in the edited strains, the obtained QX3, QX4, QX5, and QX6 strains were used to inoculate RCM from single colonies and transferred into fresh RCM at a 1:100 dilution after the OD_{600} reached late exponential or stationary phase (OD_{600} 1.3 to 2.0). The single-nucleotide variations were tested after each transfer and the sequencing results after the 10^{th} transfer are shown as a demonstration of the stability. The number of generations (n) was calculated by the following equation:

$$n = \frac{log(OD_t) - log(OD_0)}{log2} \tag{5.1}$$

where OD_t is the OD_{600} before each transfer and OD_0 is the initial OD_{600} after each transfer. The total number of generations is the sum of the generation numbers of each transfer.

5.4.7 Fermentation experiments

Fresh PETC medium was inoculated with *C. ljungdahlii* strains (from RCM cultures) and adapted to this medium in two transfers. Cells from the second preculture were harvested at late exponential phase and washed before inoculation. For heterotrophic fermentation, the wild-type, QX3, and QX4 strains were tested, and for autotrophic fermentation, the wild-type, QX5, and QX6 strains were tested. For all experiments, the initial OD_{600} was adjusted to 0.1 and samples were taken at different time intervals to analyze the growth and products. The fermentation products, including acetate, ethanol, and the substrate fructose, were measured using HPLC (LC20, Shimazu, Japan) with a RID detector and Aminex HPX-87H column (oven temperature 65°C) using 5 mM

H2SO4 as elution solvent (0.6 mL/min).

The acetate and ethanol yields in the heterotrophic fermentations were calculated as following:

$$\text{Yield (mol/mol)} = \frac{C_{Acetate/Ethanol,t}}{C_{Frutose,0} - C_{Frutose,t}} \tag{5.2}$$

Where $C_{Acetate/Ethanol,t}$ is the concentration of acetate or ethanol at 107.50 h; $C_{Fructose,0}$ is the concentration of fructose at time 0; and $C_{Fructose,t}$ is the concentration of fructose at 107.50 h.

For autotrophic fermentation, the consumed CO_2 was first calculated by adding up the carbon in biomass ($C_5H_8O_2N$), acetate, and ethanol. 100 mL of a *C. ljungdahlii* culture with an OD_{600} of 1.0 equals 24.4 mg of biomass in dry cell weight (Klask *et al.*, 2020), which corresponds to 0.214 mmol of carbon. Then the yield was calculated by:

$$\text{Yield (mol/mol)} = \frac{C_{Acetate/Ethanol,t}}{\text{consumed } CO_2 \text{ per culture volume}} \tag{5.3}$$

Where $C_{Acetate/Ethanol,t}$ is the concentration of acetate or ethanol at 137.75 h.

5.4.8 Genome-scale algorithm design

We developed an algorithm to identify all editable sites at genome-scale to identify possible single-nucleotide variations and mutations at translation level. The algorithm was coded in Python (Pycharm 2018.1 with a virtual environment), and the commented scripts and necessary files have been uploaded to GitHub (`https://github.com/isacasini/SNV_Xia_et_al_2020`). In brief, the algorithm first reads the genomic DNA of *C. ljungdahlii* (NCBI GenBank, Access No. CP001666.1), as well as an additional file with the necessary genomic information (*i.e.*, NCBI identifier, NCBI annotations, *etc*.), and it identifies all protospacers by locating the protospacer adjacent motifs (nucleotides NGG) as potential editable sites. According to the editing principles of our base editing system, the algorithm finds Cs in the identified protospacers between position -11 and -19. If a C (or several Cs) is found, it considers the site as an editable site and changes C to T, returning the single-nucleotide variations. If the editable site is located in a coding region, the algorithm also returns the resulting changes in amino acid. Additionally, the algorithm returns information about premature STOP codons in the first 70% of a gene. Another

analysis based on the hot-spot editing window (position -16 to -19) was also conducted by adjusting the parameters of the algorithm. All editable sites and analysis are summarized in the **Datasets (C.1)** and **Table C.2**.

5.4.9 Off-target event evaluation

Off-target events are a major concern of CRISPR-Cas-based genome editing tools, including base editing. We calculated the number of potential off-target sites by the following equations (Banno *et al.*, 2018):

$$\text{Correction factor} = 1 - 0.88^4 \tag{5.4}$$

$$\text{Appearance rate} = \frac{D!}{4^m \times m! \times (D-m)!} \times \frac{1}{4^N} \times \frac{1}{4^2} \times \text{Correction factor} \tag{5.5}$$

$$\text{Potential off-target sites} = \text{Genome size} \times \text{Appearance rate} \times 2 \tag{5.6}$$

Where D is the number of distal nucleotides, m is the number of matched nucleotides in distal sites, N is the number of nucleotides that should be identical to the target sequence. The correction factor indicates the probability of at least one C being located in the hot-spot editing window, which is determined under the assumption that most likely only Cs in the hot-spot editing window (position -16 to -19) can be edited by our base-editing tool, and by the A-T content (77%) of *C. ljungdahlii*. For the current design, we set D to 12, m to 5, and N to 8 (Banno *et al.*, 2018). The genome size of *C. ljungdahlii* is 4.63 Mb.

5.5 Results

5.5.1 Design of a modularized base-editing tool for *C. ljungdahlii*

For our base-editing tool, we constructed a fusion of a dCas9 with an activation-induced cytidine deaminase. We selected the dCas9 (D10A and H840A) from *S. pyogenes* and combined it with the activation-induced cytidine deaminase from *P. marinus* to

minimize the toxicity of the Cas nuclease and to obtain promising deamination performance (Banno *et al.*, 2018). The dCas9, together with a gRNA, serves as a navigator to target a specific DNA sequence (a protospacer), which has to be located next to a protospacer adjacent motif (PAM), on the genome. When binding to the target an R-loop is formed (**Figure 5.1B**). Cytidine deaminase then converts cytidine to uracil *via*a deamination process (**Figure 5.1A**). In the following replication or repair of the DNA, the cell reads uridine as T, which results in a C-to-T single nucleotide variation (Banno *et al.*, 2018). The deamination occurs on the single-strand DNA (editing strand) in the R-Loop and changes C to T in a defined editing window (**Figure 5.1B**) (Rees and Liu, 2018; Banno *et al.*, 2018). Based on previous reports, an uracil glycosylase inhibitor was fused to an activation-induced cytidine deaminase to increase the editing efficiency and to prevent the excision of uracil on the editing strand. Furthermore, a fusion to a Leu-Val-Ala protein degradation tag was added, which has been reported to lead to an overall lower amount of the fusion protein in the cell, to minimize the potential toxicity of dCas9 and uracil glycosylase inhibitor (Banno *et al.*, 2018). We used a loose linker of 363 bp (121 amino acids) for our base-editing tool to expand the editing window and increase the editing efficiency (Banno *et al.*, 2018; Tan *et al.*, 2019). Finally, we employed a tetracycline repressor-promoter (*tetR*-P_{tet}) system, which is inducible with anhydrotetracycline in *C. ljungdahlii*, for the regulated expression of our base-editing tool (**Figure 5.1C**) (Woolston *et al.*, 2018).

To introduce our base-editing tool to *C. ljungdahlii*, we designed a modularized plasmid system, which contains: 1) a template plasmid (pFX) that carries the editing module, consisting of dCas9, activation-induced cytidine deaminase, uracil glycosylase inhibitor, and Leu-Val-Ala tag under the control of the inducible *tetR*-P_{tet} system; and 2) a helper plasmid (pgRNA01) for the streamlined generation of the targeting modules (gRNAs, **Figure 5.1C**). This modularization allows the use of inverse PCR on the helper plasmid to exchange the protospacer in the gRNA for a specific genome target site. The protospacer is driven by the PJ23119 promoter and is flanked with the *S. pyogenes* Cas9 scaffold to form the new targeting module (**Figure C.1**). Afterwards, the targeting module can be assembled with the pFX plasmid, creating the editing plasmid (**Figure 5.1D**). Finally, *C. ljungdahlii* can be transformed with the methylated editing plasmids to mediate base editing.

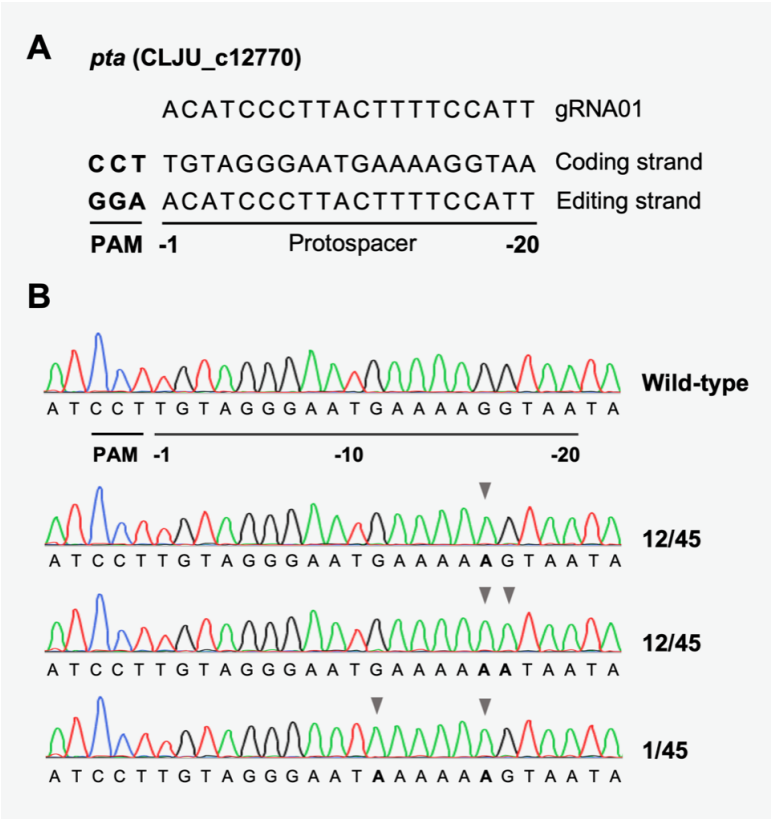

Figure 5.2: Design and validation of base editing on *pta* by gRNA01. (A) Genome sequence of editing location. The gRNA01 is complimentary to the coding strand, while deamination occurs on the editing strand (protospacer). The position of nucleotides in the protospacer were counted from the first nucleotide adjacent to the protospacer adjacent motif (position -1). **(B)** Sequencing results to validate base editing using *pta* as a target. The protospacer adjacent motif is displayed on the complimentary sequence. Arrows indicate identified single-nucleotide variations. PAM, protospacer adjacent motif.

5.5.2 Validation of base editing in *C. ljungdahlii*

To validate our system, *pta* (CLJU_c12770) from *C. ljungdahlii*, which codes for the phosphotransacetylase, was selected as a first target (gRNA01, **Table C.1, Figure 5.2A**). We discovered efficient conversion from C to T on the editing strand, leading to a G-to-A single-nucleotide variation in the coding strand (**Figure 5.2A**, **Figure 5.2B**). In total, 45 colonies from 5 individual rounds of base editing were picked to analyze the frequency and editing window (**Table C.2**). Three clean editing patterns were identified (**Figure 5.2B**). The highest editing frequency was found for position -16 of the protospacer in all three patterns (counting the site adjacent to the protospacer adjacent motif as position -1). Twelve out of 45 colonies showed a single mutation at this

position -16, 12 colonies showed double mutations at positions -16 and -17, and 1 colony showed a double mutation at positions -12 and -16 (**Figure 5.2B**). We also identified 2 colonies with mutations at position -11 (**Figure 5.2B**) and -2 (**Figure C.2**), respectively, while the latter one (position -2) was a colony with mixed signals at position -17 (**Figure C.2**). Importantly, this finding did not considerably influence the chance to select colonies with the desired single-nucleotide variations in only one round of selection.

We further interrogated the editing principles of the base-editing tool. First, we targeted another site in *pta* (gRNA06, **Table C.1**). In this case, only 1 out of 8 colonies was found to be edited at position -7 (**Figure C.3A**), which suggests a lower editing efficiency at position -7. Accordingly, although editing is possible in a wide editing window, the editing efficiency of our base-editing tool might be lower between position -2 to -11. Second, our base-editing tool converts C to T on the editing strand, which indicates that it only deaminates C(s) in the protospacer (Rees and Liu, 2018; Nishida *et al.*, 2016; Banno *et al.*, 2018). Therefore, we designed two gRNAs with no Cs in the protospacers (gRNA02 and gRNA07, **Table C.1**) to examine the base-editing mechanism further, because we hypothesized that in this case no base editing would occur. Importantly, this experiment can hardly be done in non-A-T-rich bacteria in which protospacers without a C are much less abundant. As anticipated, we did not observe single-nucleotide variations in any colony (**Figure C.3B**). Third, others have demonstrated that activation-induced cytidine deaminase-mediated base editing showed a hot-spot editing window of five nucleotides starting from the opposite end of the protospacer adjacent motif, and that the window shifts depending on the length of protospacers(Rees and Liu, 2018; Banno *et al.*, 2018). We discovered that our base-editing tool did not lead to mutations of the C at position -20 with neither a 20- or a 22-nucleotide protospacer (gRNA05 and gRNA15, **Table C.1**). This suggests that the hot-spot editing window with high editing efficiency in *C. ljungdahlii* starts from position -19 (**Figure C.3C**). However, more targets have to be investigated to determine whether editing is possible at position -20. To conclude, we observed that: 1) The C(s) within position -2 to -19 of the protospacer are editable; 2) a hot-spot editing window of our base-editing tool is located approximately between position -16 to -19; and 3) the editing efficiency is low in the range of position -2 to -11.

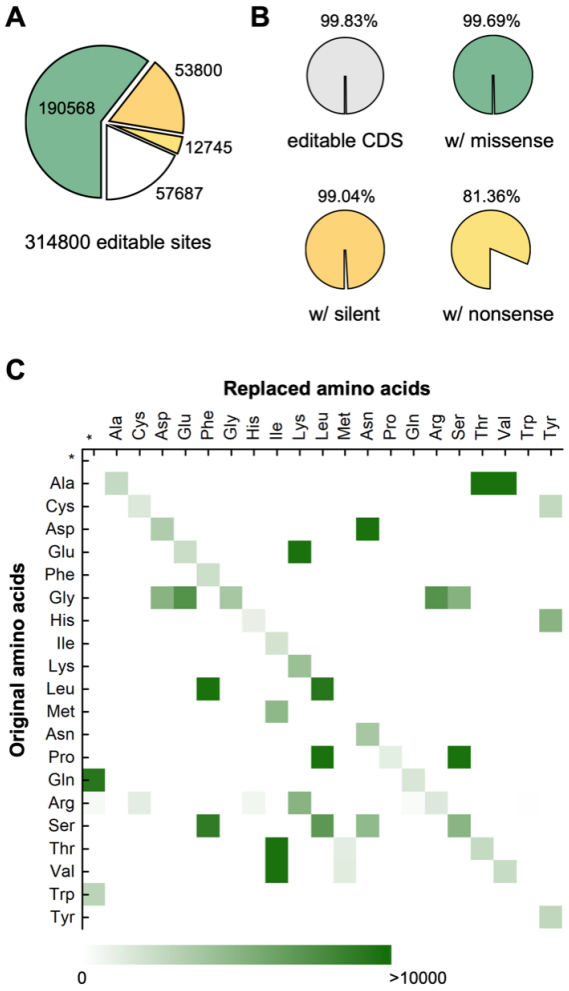

Figure 5.3: *In-silico* evaluation of base-editing capability in *C. ljungdahlii*. **(A)** Pie chart with the numbers of sites that can be edited on a genome scale. Green indicates missense mutations, orange indicates silent mutations, yellow indicates nonsense mutations, and white indicates single-nucleotide variations that are not located in coding regions. **(B)** Pie charts with the percentages of genes that can be edited in generating different kinds of mutations in coding regions. **(C)** Amino acid replacement matrix generated by base editing. The green squares indicate the possible mutations with a lighter green color indicating fewer and a darker green color indicating more possible mutations on genome-scale.

5.5.3 *In-silico* evaluation of base-editing capability on genome-scale

One doubt for applying base editing in acetogenic bacteria might be a questionable editing capability, because dCas9 from *S. pyogenes* recognizes the nucleotides NGG as the protospacer adjacent motif and *C. ljungdahlii* is an A-T-rich bacterium (77% A-T) (NCBI GenBank Access No. CP001666.1). To investigate the editing capability of our

base-editing tool, we developed a genome-scale algorithm. The algorithm reads all possible protospacer adjacent motifs from the genome sequence, and then identifies Cs on the editing strand and converts those Cs to Ts within the editing window. This returns mutations at translational level and identifies the genome-wide capability to generate missense mutations, silent mutations, and nonsense mutations (by installing premature STOP codons) (**Dataset 1, C.1**). To avoid an overestimation of the capability, we defined position -19 to -11 as the editing window, which combined our results and previous reports (Rees and Liu, 2018). The editing outside this window might still have been possible but with low efficiency. We found that 314,800 sites could be potentially edited among which 257,133 sites are located in coding regions (**Figure 5.3A**). These editable sites involve 99.83% (4,177 out of 4,184) of all genes. Only 7 genes that encode short hypothetic proteins (25 to 50 amino acids) cannot be edited (**Table C.2**). We found that: 1) 99.69% (4,171 out of 4,184) genes can be edited to have missense mutations; 2) 99.04% (4,144 out of 4,184) genes can be edited to have silent mutations; and 3) 81.36% (3,404 out of 4,184) genes can be inactivated/truncated by installing premature STOP codons (Kuscu *et al.*, 2017) (**Figure 5.3B, Table C.3**). Moreover, we found that 71.92% (3009 out of 4184) genes can be inactivated by installing premature STOP codons within the first 70% of the coding sequence (**Table C.3**). We also identified the editable sites, which can be edited with high efficiency by using the hot-spot editing window (position -16 to -19) to give a full evaluation of the capability of our base-editing tool (**Table C.3** and **Dataset 2**, (C.1)). These *in-silico* results demonstrate a great capability of base editing even in an A-T-rich bacterium such as *C. ljungdahlii*.

On a genome scale, we discovered that, except for Trp and Met codons, all other amino acid codons can be edited to lead to silent mutations without changing the amino acid, while 15 out of 20 amino acids (excluding Phe, Ile, Lys, Asn, and Trp) can be changed to another amino acid (missense mutation) by changing the codon *via* single-nucleotide variations (**Figure 5.3C**). Importantly, Gln, Arg, and Trp codons can be replaced to STOP codons. By changing CAA to TAA or CAG to TAG, 9,329 Gln codons can be changed to STOP codons. Arg (427 sites) also shows potential to be mutated to a STOP codon by converting CGA to TGA. Trp (2,989 sites) offers a different strategy to install premature STOP codons, because base editing changes CCA to TTA, TCA, or CTA on the editing strand and results in TAA, TGA, or TAG on the coding strand. Compared with a predicted amino acid replacement matrix in G-C-rich

Streptomyces species, only 3 out of 32 different amino acid replacement routines were not identified for *C. ljungdahlii* (Pro to Phe, Gly to Lys, and Gly to Asn) (**Figure 5.3C**), which probably results from the A-T-rich genome (Tong *et al.*, 2019).

Moreover, we calculated the potential for off-target events. We investigated off-target sites, which partially match the target sequence, with a modified equation from a previous report in which a similar design of the base-editing tool was used for *E. coli* (Banno *et al.*, 2018). Based on this previous report, eight nucleotides in an off-target DNA sequence have to be identical to the gRNA sequence from position -1 to -8, and additional five nucleotides have to match between position -9 to -20 to lead to an off-target event. Besides these parameters, we assumed that editing in a potential off-target site would most likely occur when: 1) the editing site is located in the hot-spot editing window (position -16 to -19); and 2) only when C(s) are present in the hot-spot editing window. Moreover, we added a correction factor for *C. ljungdahlii* due to the A-T-rich genome, which results in low abundance of C(s) in the hot-spot editing window. We found that off-target sites occur approximately 2.7 times per random gRNA in *C. ljungdahlii*, which is much lower than in *E. coli* (8 times) (Banno *et al.*, 2018). According to the genome-scale evaluation, approximately 1.8 potential off-target sites would be located in a coding region, and might lead to missense/nonsense mutations if these sites were edited. Based on these numbers, the probability of actual off-target events can be considered to be very low. However, we want to acknowledge that there is a risk to have off-target events when using base-editing tools.

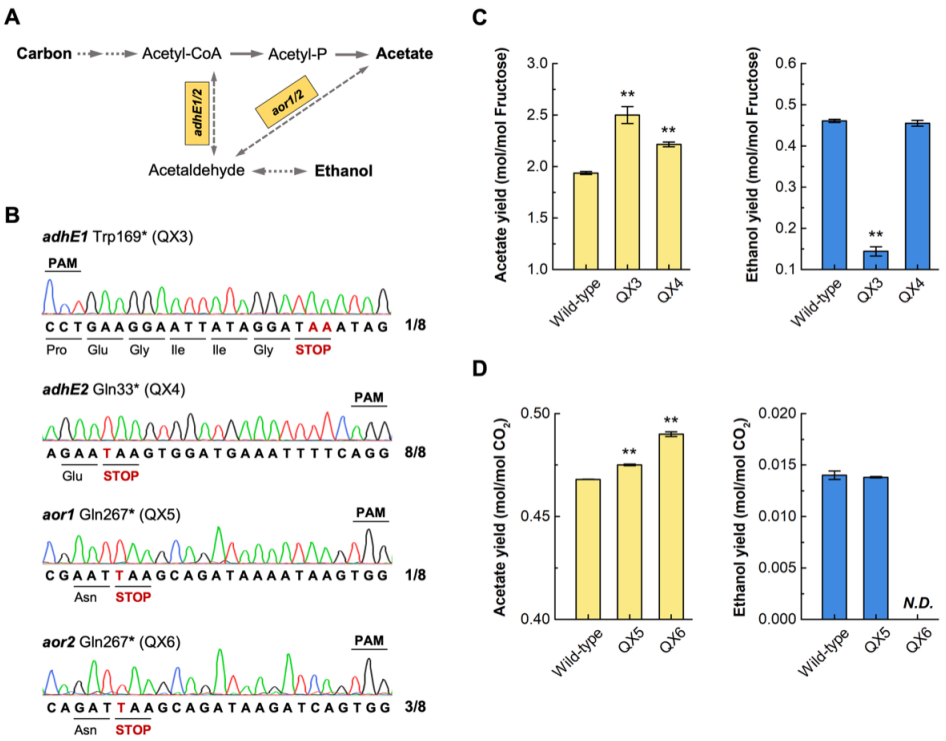

Figure 5.4: Reprogramming carbon flux towards improved acetate production. **(A)** Metabolic pathway from acetyl-CoA to acetate and ethanol. Under heterotrophic conditions, ADHE1/2 (encoded by *adhE1/2*) convert acetyl-CoA to ethanol *via* acetaldehyde as an intermediate, while under autotrophic conditions, AOR1/2 (encoded by *aor1/2*) convert acetate to acetaldehyde. **(B)** Validation of premature STOP codons in the four edited strains (QX3, QX4, QX5, and QX6). The edited sequences and amino acids are shown in the protospacer region. **(C)** Acetate and ethanol yields of wild-type, QX3, and QX4 under heterotrophic conditions with 5 g/L (27.8 mM) of fructose as the carbon source. **(D)** Acetate and ethanol yields of wild-type, QX5, and QX6 under autotrophic conditions with a gas mixture of H_2/CO_2 (80/20 vol-%, 1.5 bar) as the substrate. The fermentation experiments were conducted in triplicate (N=3), and the error bars indicate the standard deviations. The differences in acetate yield and ethanol yield were verified by *t*-test with a $P < 0.05$ as a significant difference and a $P < 0.001$ as a highly significant difference (**).

5.5.4 Reprogramming carbon flux by installing premature STOP codons

To further demonstrate the application potential of our base-editing tool, we disrupted genes involved in ethanol production in *C. ljungdahlii* to reprogram the carbon flux for improved acetate production as a first application. To achieve this, we targeted four genes in two different metabolic pathways individually. First, we targeted *adhE1* (CLJU_c16510) and *adhE2* (CLJU_c16520), which encode isoenzymes of the

bifunctional aldehyde-alcohol dehydrogenase. This enzyme converts acetyl-CoA to ethanol *via* acetaldehyde as an intermediate under heterotrophic conditions (**Figure 5.4A**) (Richter *et al.*, 2016; Leang *et al.*, 2013). The premature STOP codons were successfully installed in *adhE1* (STOP after 19.4% of the coding region with gRNA10, **Table C.1**) and *adhE2* (STOP after 3.8% of the coding region with gRNA11, **Table C.1**), respectively, generating strains QX3 (*adhE1* Trp169*) and QX4 (*adhE2* Gln33*) (**Figure 5.4B**). Second, we targeted *aor1* (CLJU_c20110) and *aor2* (CLJU_c20210), which encode isoenzymes of the aldehyde:ferredoxin oxidoreductase. This enzyme converts acetate (in the form of undissociated acetic acid) to acetaldehyde under autotrophic conditions (**Figure 5.4A**) (Richter *et al.*, 2016). Accordingly, we inactivated *aor1* (STOP after 44.0% of the coding region) and *aor2* (STOP after 44.0% of the coding region) by installing STOP codons with gRNA19 and gRNA14 (**Table C.1**), generating strains QX5 (*aor1* Gln267*) and QX6 (*aor2* Gln267*), respectively (**Figure 5.4B**).

We first tested the stability of all four strains by serial transfer experiments. We confirmed the correct genotype (single-nucleotide variations) at the edited location after 10 transfers with more than 65 generations, indicating that our base-editing tool resulted in stable genotypes (**Figure C.4**). Next, we investigated the physiology with bottle experiments. For heterotrophic conditions, QX3 and QX4 showed growth defects compared to the wild-type strain, and did not consume all of the provided fructose (**Figure C.5A** and **Figure C.5B**). However, both QX3 and QX4 achieved higher final acetate yields, because at similar final biomass and acetate concentrations, these strains achieved lower final ethanol concentrations (**Figure 5.4C, Figure C.5C** and **Figure C.5D**). Especially for QX3, we observed a 28.9% higher acetate yield and a 68.6% reduced ethanol yield compared to the wild-type strain (**Figure 5.4C**), which is in agreement with a previous report on an *adhE1* deletion in *C. ljungdahlii* (Leang *et al.*, 2013). For autotrophic conditions, we also found growth defects for QX5 and QX6 compared to the wild-type strain, and less overall substrate (H_2/CO_2) consumption (**Figure C.6A** and **Figure C.6B**). QX5 showed wild-type-like patterns in the yield of acetate and ethanol, however, at an overall lower absolute level of final concentrations (**Figure 5.4D, Figure C.6C**, and **Figure C.6D**). QX6 showed a slight, but significant, increase of 4.6% in the final acetate yield (0.490 mol acetate/mol consumed CO_2) compared to the wild-type strain (0.468 mol acetate/mol consumed CO_2), while ethanol was below our detection limit (**Figure 5.4D, Figure C.6C**). The higher acetate yield

likely resulted from a redistribution of carbon from biomass and/or ethanol to acetate production. The final biomass and acetate concentrations were considerably lower compared to the wild-type strain (**Figure C.6A** and **Figure C.6D**), while we further acknowledge that this increase (4.6%) in acetate yield by QX6, although significant, is only marginal. However, this increase brings the acetate yield (0.490 mol acetate/mol consumed CO_2) closer to the theoretical limit (0.500 mol acetate/mol consumed CO_2). In addition, we calculated the acetate and ethanol yields based on the dry cell weight under either heterotrophic (**Figure C.7A** and **Figure C.7B**) or autotrophic conditions (**Figure C.7C** and **Figure C.7D**). We found that the yields per dry cell weight of the edited strains shared similar variation patterns with those calculated based on consumed carbon (**Figure C.7** and **Figure 5.4**). Except for QX5, which showed an increased ethanol yield per dry cell weight, but had a similar ethanol yield per consumed carbon compared to the wild-type strain.

These results are in agreement with the results from previous studies with *C. ljungdahlii* and the closely related acetogenic bacterium *C. autoethanogenum* in which full knock-out strains of the respective genes were investigated (Leang *et al.*, 2013; Liew *et al.*, 2017). Therefore, although we cannot eliminate the possibility for off-target events, the probability that our fermentation results are caused by off-target events is very low. We found that a single-gene inactivation by introducing a premature STOP codon (*adhE1* and *aor2*) could be enough to generate strains with higher acetate yield and lower ethanol yield under either heterotrophic or autotrophic conditions. To further optimize the metabolism, it is also possible to obtain multigene inactivation in one strain with multiplex base editing by employing established assembly approaches to generate gRNA arrays with protospacers targeting different genes as described by others (Banno *et al.*, 2018; Woolston *et al.*, 2018).

5.6 Discussion

5.6.1 An expanded synthetic biology toolkit for acetogenic bacteria

We developed a base-editing tool for the model acetogenic bacterium *C. ljungdahlii* and enabled genome editing at a one-nucleotide resolution without DNA cleavage, homology-directed repair, and donor DNA. Base editing bypasses the general bottlenecks of applying CRISPR-Cas systems in bacteria, which include the toxicity of

Cas nucleases and inefficient DNA repairing mechanisms. It also lowers the requirement of transformation efficiency in *C. ljungdahlii* compared to conventional CRISPR-Cas-based genome editing (Huang *et al.*, 2016; Zhao *et al.*, 2019; Vento *et al.*, 2019). We discovered a great capability of our base-editing tool. With base editing, we can: **1)** install STOP codons to 3,404 genes in *C. ljungdahlii* to reprogram the metabolisms directly; **2)** generate silent mutations in 4,144 genes to interrogate codon preference; and **3)** replace amino acids *via* missense mutations in 4,171 genes to perform protein research and engineering *in vivo*. Moreover, we observed that desired single-nucleotide variations could be obtained in a single round of selection. Only in two cases, we observed both wild-type and edited signals in one colony. In a previous study, mixed populations have been reported as an issue for base editing in *C. berjerinkii*, making a second round of selection necessary (Li *et al.*, 2019). Notably, we observed high precision single-nucleotide variations in *C. ljungdahlii* with limited bystander nucleotide substitutions (undesired single-nucleotide variations within the editing window on the editing strand) (**Figure 5.4B**). This is an advantage returned by the A-T-rich genome of *C. ljungdahlii*, which naturally overcomes bystander base editing with limited Cs in a target sequence and leads to precise base editing. We designed a modularized system to enable fast generation of the base-editing plasmid series. The employed plasmid backbone, replicon for clostridia, antibiotic resistances markers, and the dCas9 protein have been separately demonstrated to be functional in various species in the order Clostridiales, including *Acetobacterium woodii* (dCas9 has not yet been validated) (Beck *et al.*, 2020; Hoffmeister *et al.*, 2016), *Eubacterium limosum* (Shin *et al.*, 2019), and *C. autoethanogenum* (Nagaraju *et al.*, 2016). Accordingly, the system could be easily generalized in acetogenic bacteria, which mainly belong to the order Clostridiales.

5.6.2 Linking base editing with microbial C1 utilization

Base editing was first invented to revert single-nucleotide mutations related to human diseases (Komor *et al.*, 2016). Despite an increasing utilization in medicine and agriculture (Rees and Liu, 2018; Molla and Yang, 2019), only a few reports validated bacterial base-editing principles, especially for bacterial pathogens (Wang *et al.*, 2018; Chen *et al.*, 2018; Gu *et al.*, 2018). Furthermore, only few reports exist for biotechnologically relevant bacteria, and these do not demonstrate a specific

biotechnological application (Tong *et al.*, 2019; Li *et al.*, 2019; Banno *et al.*, 2018). Presumably, base editing in bacteria might be hindered by a low editing capability in relevant genes or low stability of resulting single-nucleotide variations, which would not be favorable for industrial biotechnology. To overcome this presumption, we inactivated four genes related to ethanol production in *C. ljungdahlii* as a first application, with the goal to increase the acetate yield. This would improve the production of certain platform chemicals that require acetate as an intermediate (Nevin *et al.*, 2011; Nangle *et al.*, 2017). For instance, our acetate-producing strain can be considered to further improve the two-stage bioprocess for single-cell protein production from C1 gases, with acetate as the carbon-fixing intermediate product before being fed to aerobic yeasts, especially for industrial gases that contain CO (Molitor *et al.*, 2019). Importantly, single-nucleotide variations generated by base editing are clean mutations in the genome, which may also occur in natural evolution. Thus, our base-editing tool, in principle, provides a unique venue to engineer industrially relevant bacteria without creating genetically modified organisms (GMOs). However, this advantage is not recognized in the legislation of all countries, and especially in Europe, CRISPR-Cas-based genome editing is often *per se* considered to generate GMOs, irrespective of the outcome of the editing.

5.6.3 Limitations and perspectives for base editing in A-T-rich bacteria

Not surprisingly, base editing has its limitations in A-T-rich bacteria. First, the editing sites are still limited in A-T-rich genomes, because of the protospacer adjacent motif (nucleotides NGG) that is recognized by dCas9 from *S. pyogenes*. Despite a large number of editable sites, not the entire genome can be covered. To overcome this, possible strategies include using a dCas protein with a different protospacer adjacent motif such as Cas12a with nucleotides TTTV as a protospacer adjacent motif (Li *et al.*, 2018) or xCas9 with nucleotides NG as a protospacer adjacent motif (Hu *et al.*, 2018). Second, base editing is site-specific, and not all sites following the editing principles can be edited (Molla and Yang, 2019). Evidently, when we tried to introduce a STOP codon at Gln235 in *aor1* with gRNA13, no colonies with the expected single-nucleotide variations were identified (data not shown), while we obtained 2 out of 8 colonies with Gln237* replacement by using gRNA19. A different deaminase may be necessary to

circumvent this limitation. Third, base editing, although in low probability, arouses off-target events in bacteria, which should not be underestimated. Possible ways to minimize editing at off-target sites would include: **1)** optimizing the gRNAs based on a pre-off-target evaluation; and **2)** employing a CRISPR-Cas system with higher specificity (Zuo *et al.*, 2020). Finally, base editing intrinsically cannot insert DNA fragments into the genome. Yet, it offers a new angle to edit the genome with CRISPR-Cas systems without DNA cleavage. Starting from this perspective, a recent report demonstrated DNA insertion into the genome without cutting the DNA by coupling a CRISPR-Cas system to a reverse-transcriptase (Anzalone *et al.*, 2019).

In summary, we established an efficient base-editing tool for gene manipulation in acetogenic bacteria. Further, we demonstrated the use of this cutting-edge genome-editing tool in C1 utilization with the industrially relevant acetogenic bacterium *C. ljungdahlii*. Our strategy provides an example for upgrading bacterial genome-editing tools with CRISPR systems in general, especially for bacteria that are sensitive to heterologously expressed Cas nucleases (*e.g.*, cyanobacteria41) and those with limited capability of receiving foreign DNA.

5.7 Acknowledgements

This work was supported by the Alexander von Humboldt Foundation in the framework of the Alexander von Humboldt Professorship Largus T. Angenent (L.T.A.) and the Humboldt Research Fellowship for Postdoctoral Researchers Peng-Fei Xia (P.-F.X.). Sarah Schulz (S.S.) thanks the support from the German Federal Environmental Foundation (Deutsche Bundesstiftung Umwelt, DBU). Bastian Molitor (B.M.) and L.T.A are grateful to funding from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany´s Excellence Strategy – EXC 2124 – 390838134. Finally, L.T.A. is supported as a Max Planck Fellow by the Max Planck Institute for Developmental Biology.

5.8 Supplemental information

Supplemental information: summary of gRNAs and protospacers; summary of the base-editing results and *in-silico* evaluation; strains, plasmids, and primers used in this study;

figures supporting the main results. These data can be found in **Appendix C**.

Chapter 6

Reproduced and adapted with permission from: Klask, C-M., Jäger, B., Casini, I., Angenent, L.T., Molitor, B. (2022), Genetic evidence reveals the indispensable role of the *rseC* gene for autotrophy and the importance of a functional electron balance for nitrate reduction in *Clostridium ljungdahlii*. *Frontiers in Microbiology*, 13:887578.

6.1 Author contributions

Christian-Marco Klask (C.M.K.) and Bastian Molitor (B.M.) designed the experiments and wrote the manuscript. C.M.K. performed the genetic work, conducted the growth experiments, analyzed the metabolites, performed the *in-silico* research for selected model microbes, and analyzed the experimental data. C.M.K. and Benedikt Jäger (B.J.) performed the qRT-PCR experiments. **Isabella Casini (I.C.) performed the Python-based *in-silico* analysis of sequenced genomes of acetogenic bacteria. I.C. generated Table D.7.** Largus T. Angenent (L.T.A.) and B.M. supervised the work. All authors edited the manuscript and approved the final version.

6.2 Abstract

For *Clostridium ljungdahlii*, the RNF complex plays a key role for energy conversion from gaseous substrates such as hydrogen and carbon dioxide. In a previous study, a disruption of RNF-complex genes led to the loss of autotrophy, while heterotrophy was still possible *via* glycolysis. Furthermore, it was shown that the energy limitation during autotrophy could be lifted by nitrate supplementation, which resulted in an elevated cellular growth and ATP yield. Here, we used CRISPR-Cas12a to delete: **1)** the RNF complex-encoding gene cluster *rnfCDGEAB*; **2)** the putative RNF regulator gene *rseC*;

and **3)** a gene cluster that encodes for a putative nitrate reductase. The deletion of either *rnfCDGEAB* or *rseC* resulted in a complete loss of autotrophy, which could be restored by plasmid-based complementation of the deleted genes. We observed a transcriptional repression of the RNF-gene cluster in the *rseC*-deletion strain during autotrophy and investigated the distribution of the *rseC* gene among acetogenic bacteria. To examine nitrate reduction and its connection to the RNF complex, we compared autotrophic and heterotrophic growth of our three deletion strains with either ammonium or nitrate. The *rnfCDGEAB*- and *rseC*-deletion strains failed to reduce nitrate as a metabolic activity in non-growing cultures during autotrophy but not during heterotrophy. In contrast, the nitrate reductase deletion strain was able to grow in all tested conditions but lost the ability to reduce nitrate. Our findings highlight the important role of the *rseC* gene for autotrophy, and in addition, contribute to understand the connection of nitrate reduction to energy metabolism.

6.3 Introduction

Acetogenic bacteria (*i.e.*, acetogens), such as *Clostridium ljungdahlii*, maintain autotrophic growth with mixtures of the gaseous substrates carbon dioxide, carbon monoxide, and hydrogen as carbon and energy sources (Drake *et al.*, 2008; Katsyv and Müller, 2020). The pathway that allows carbon fixation for autotrophic growth in acetogens is the Wood-Ljungdahl pathway (Wood *et al.*, 1986; Ljungdahl, 2009). Overall, the Wood-Ljungdahl pathway is considered the most energy-efficient pathway for carbon fixation that exists in nature (Fast and Papoutsakis, 2012; Song *et al.*, 2020). In the Wood-Ljungdahl pathway, two molecules of carbon dioxide are reduced to one carbonyl group and one methyl group, which are then combined with coenzyme A to the central metabolite acetyl-coenzyme A (Ljungdhal, 1986). The electrons for these reductions can be derived from the oxidation of hydrogen or carbon monoxide, while carbon monoxide can also enter the pathway directly to provide the carbonyl group (Wood, 1991). For carbon fixation, acetyl-coenzyme A is channeled into the anabolism for cellular growth (Ragsdale and Pierce, 2008). For energy conservation, acetyl-coenzyme A is converted to acetate, which generates cellular energy by substrate level phosphorylation (Schuchmann and Müller, 2014). One mole of ATP is generated per mole of acetate that is produced. However, in the first step of the pathway, after

carbon dioxide was reduced to formate, one mole of ATP is invested to activate the formate to formyl-tetrahydrofolate (Wood *et al.*, 1986; Ljungdahl, 2009). Thus, the energy balance of the Wood-Ljungdahl pathway alone is net zero (Schuchmann and Müller, 2014). All required cellular energy for the anabolism of the microbes during autotrophy is generated *via* membrane-coupled phosphorylation (Katsyv and Müller, 2020). In *C. ljungdahlii*, the membrane-bound transhydrogenase *Rhodobacter* nitrogen fixation (RNF) complex (Schmehl *et al.*, 1993; Biegel *et al.*, 2011) utilizes two electrons from the oxidation of reduced ferredoxin to reduce NAD^+ to NADH, while simultaneously one proton is translocated across the membrane (Tremblay *et al.*, 2012; Schuchmann and Müller, 2014). A proton-dependent F_1F_O ATPase then consumes the chemiosmotic proton gradient to generate ATP (Köpke *et al.*, 2010; Al-Bassam *et al.*, 2018). In the presence of carbon dioxide and hydrogen, theoretically, *C. ljungdahlii* can generate a maximum of 0.63 moles ATP per mole acetate for the anabolism *via* membrane-coupled phosphorylation. Thus, the conservation of cellular energy during autotrophy occurs at the thermodynamic limit of life (Schuchmann and Müller, 2014).

For *C. ljungdahlii*, the RNF complex is encoded by the RNF-gene cluster *rnfCDGEAB*. Although the RNF complex plays an essential role for energy conservation during autotrophy in *C. ljungdahlii* (Tremblay *et al.*, 2012), fundamental knowledge about the regulation and gene expression control of the encoding RNF-gene cluster is missing. Transcriptome studies with *C. ljungdahlii* revealed that the RNF complex is under strict gene expression control and strongly induced during autotrophy (Tan *et al.*, 2013; Al-Bassam *et al.*, 2018). The regulatory mechanisms behind this remain unknown. However, the small gene *rseC*, which is located directly upstream of *rnfC* in *C. ljungdahlii*, is also highly expressed during autotrophy and follows the expression profile of *rnfC* (Al-Bassam *et al.*, 2018). The gene *rseC* is annotated to contain the conserved protein domain family RseC_MucC (pfam04246) (Köpke *et al.*, 2010). The domain family RseC_MucC is found in positive transcriptional regulators in other microbes. The one representative, RseC, was found to be involved in the oxidative stress response in *Escherichia coli* (De Las Peñas *et al.*, 1997; Missiakas *et al.*, 1997; Koo *et al.*, 2003), and in thiamine synthesis in *Salmonella typhimurium* (Beck *et al.*, 1997). The other representative, MucC, was found to be involved in the regulation of the alginate formation of *Azotobacter vinelandii* (Martinez-Salazar *et al.*, 1996) and *Pseudomonas aeruginosa* (Boucher *et al.*, 1997). Others identified a transcription start site for *C. ljungdahlii*, which is located upstream of the *rseC* gene, and a putative

terminator sequence, which is located between *rseC* and *rnfC*. This indicates that *rseC* is expressed as an individual transcript apart from the RNF-gene cluster transcripts (Al-Bassam *et al.*, 2018). Altogether, this led to the assumption that the *rseC* gene product is closely linked to the RNF complex, and could be important for the regulation of autotrophy in *C. ljungdahlii* (Köpke *et al.*, 2010; Al-Bassam *et al.*, 2018).

While autotrophy in acetogens results in low cellular energy yields, Emerson *et al.* (2019) reported that *C. ljungdahlii* is able to couple the reduction of nitrate to the generation of ATP during growth with carbon dioxide and hydrogen. This relieved the energy limitation during autotrophy and resulted in a significantly higher biomass yield (Emerson *et al.*, 2019). We confirmed this in a bioreactor study, and biomass yields were considerably higher with nitrate, but resulted in stochastic crashes of the continuous bioreactor cultures (Klask *et al.*, 2020). Emerson *et al.* (2019) proposed that electrons, which are required for nitrate reduction, are provided by NADH. One route to regenerate NADH is by the RNF complex where reduced ferredoxin is consumed (Biegel *et al.*, 2011), which would link nitrate reduction to the energy metabolism. It was assumed that nitrate reduction is accelerating the RNF-complex activity, and thus the generation of ATP (Emerson *et al.*, 2019). This way, the co-utilization of carbon dioxide and nitrate with hydrogen was suggested to yield up to 1.5 ATP through the concerted action of the RNF complex and the ATPase (Emerson *et al.*, 2019). This would be a 2.4-fold increase in ATP yield compared to the ATP yield with carbon dioxide and hydrogen alone (Schuchmann and Müller, 2014).

To investigate the autotrophy in *C. ljungdahlii* with respect to regulatory aspects and the interplay with nitrate reduction, we addressed three main questions: 1) Is the *rseC* gene involved in the regulation of the RNF-gene cluster?; 2) Is nitrate reduction dependent on a functional RNF complex?; and 3) Is nitrate reduction abolished by the deletion of the nitrate reductase that is annotated in the genome of *C. ljungdahlii*?

6.4 Materials and Methods

6.4.1 Bacterial strains and growth

Escherichia coli TOP10 (Invitrogen), *E. coli* EPI300 (Lucigen), and *E. coli* HB101 pKR2013 (DSM 5599) were grown at 37°C in Luria Broth (LB) medium containing (per liter): 5 g NaCl; 10 g peptone; and 5 g yeast extract. *C. ljungdahlii* ATCC13528

was generally cultivated in anaerobic Rich Clostridial Medium (RCM) containing per liter: 5 g fructose; 3 g yeast extract; 10 g meat extract; 10 g peptone; 5 g NaCl; 1 g soluble starch; 3 g sodium acetate; 0.5 g L-cysteine HCl; and 4 mL resazurin-solution (0.025 vol-%). For growth experiments with *C. ljungdahlii*, standard PETC medium (Klask et al., 2020) was used containing (per liter): 1 g yeast extract; 1.0 g NH_4Cl; 0.1 g KCl; 0.2 g $MgSO_4x7\ H_2O$; 0.8 g NaCl; 0.1 g KH_2PO_4; 0.02 g $CaCl_2x2\ H_2O$; 4 mL resazurin-solution (0.025 vol-%); 10 mL trace element solution (TE, 100x); 10 mL Wolfe's vitamin solution (100x); 10 mL reducing agent (100x); and 20 mL of fructose/2-(N-morpholino)ethanesulfonic acid (MES) solution (50x). TE was prepared as 100x stock solution containing (per liter): 2 g nitrilotriacetic acid (NTA); 1 g $MnSO_4xH_2O$; 0.8 g $Fe(SO_4)_2(NH_4Cl)2x6\ H_2O$; 0.2 g $CoCl_2x6\ H_2O$; 0.0002 g $ZnSO_4x7\ H_2O$; 0.2 g $CuCl_2x2\ H_2O$; 0.02 g $NiCl_2x6\ H_2O$; 0.02 g $Na_2MoO_4x2\ H_2O$; 0.02 g Na_2SeO_4; and 0.02 g Na_2WO_4. The pH of the TE was adjusted to 6.0 after adding NTA. The solution was autoclaved and stored at 4°C. Wolfe's vitamin solution was prepared aerobically containing (per liter): 2 mg biotin; 2 mg folic acid; 10 mg pyridoxine-hydrochloride; 5 mg thiamin-HCl; 5 mg riboflavin; 5 mg nicotinic acid; 5 mg calcium pantothenate; 5 mg p-aminobenzoic acid; 5 mg lipoic acid; and 0.1 mg cobalamin. The vitamin solution was sterilized using a sterile filter (0.2 μm), sparged with N_2 through a sterile filter, and stored at 4°C. The 50x fructose/MES solution contained (per 100 mL): 25 g fructose; and 10 g MES. The pH was adjusted to 6.0 by adding KOH. For autotrophic experiments, fructose was omitted. In nitrate experiments, ammonium chloride was replaced with sodium nitrate ($NaNO_3$) in the equal molar amount (=18.7 mM). The reducing agent solution contained (per 100 mL): 0.9 g NaCl and 4 g L-cysteine HCl and was prepared with anaerobic water under anaerobic conditions. The reducing agent was stored at room temperature. For solid LB medium, 1.5 weight-% agar was added. For solid RCM or PETC medium 1.0-2.0 weight-% agar was added. For conjugation of *C. ljungdahlii* cells (see below) a modified PETC medium (PETC+5gS) was used containing additionally (per liter): 5 g peptone and 5 g meat extract.

Liquid *E. coli* cultures and autotrophic *C. ljungdahlii* cultures were agitated at 150 revolutions per minute (rpm) (Lab Companion Incubater Shaker ISS-7100R, Jeio Tech). Heterotrophic cultures of *C. ljungdahlii* and LB plates with *E. coli* cells were incubated without shaking (Incubator IN260, Memmert). Anaerobic work was performed in an anaerobic chamber (Glovebox-System UNIlab Pro, MBraun) with an N_2 (100 vol-%)

atmosphere. However, *C. ljungdahlii* cultures in bottles were transferred at the bench with sterile syringes and needles. Before each transfer between serum bottles, we flamed the top of the rubber stopper with ethanol (70 vol-%) at a Bunsen burner. All plating work with *C. ljungdahlii* was performed in the anaerobic chamber with a maximum of 5 parts per million (ppm) oxygen in the atmosphere. All plating work with *E. coli* was carried out in a lamina flow bench (Hera Safe KS18, Thermo Fischer Scientific). Antibiotics (see below) were added to maintain plasmid stability in recombinant cultures of *E. coli* and *C. ljungdahlii*.

6.4.2 Antibiotics

Chloramphenicol (30 μg/mL), ampicillin (100 μg/mL), kanamycin (50 μg/mL), and trimethoprim (10 μg/ml) were applied to maintain plasmids in *E. coli* strains, while thiamphenicol (5 mg/mL) was used for recombinant strains of *C. ljungdahlii*. Trimethoprim was dissolved in DMSO (100 vol-%). Thiamphenicol was prepared as aerobic stock solution (25 mg/mL) in DMSO (100 vol-%) and diluted with sterile water (1:10) before use. The diluted thiamphenicol solution (2.5 mg/mL) was transferred into a sterile 1 mL syringe. 100 μL of this solution was used to add to a 50 mL RCM or PETC medium (final concentration of 5 μg/mL). The use of DMSO over ethanol as solvent for thiamphenicol prevented the addition of external ethanol, which is a metabolite, to cultures of *C. ljungdahlii*. All antibiotic stock solutions were stored at -20°C.

6.4.3 General cloning and gene manipulation

The broad-host shuttle-vector system pMTL80000 (Heap *et al.*, 2009) was used for all cloning steps. All generated plasmids of this study (**Table D.1**) were cloned with restriction endonucleases and T4 ligase from NEB (New England Biolabs) or Gibson assembly (NEBuilder® HiFi DNA Assembly, New England Biolabs). PCR work was carried out with primers provided by IDT (Integrated DNA Technologies) (**Table D.2**) and with a proof-reading Q5® High-Fidelity DNA Polymerase (New England Biolabs) according to the manufacture's guidelines. Genomic DNA (gDNA) was purified from 2 mL of exponential cultures of *C. ljungdahlii* with the NucleoSpin Tissue Mini kit (Macherey-Nagel) and used as PCR-template. Notably, instead of performing harsh cell

disruption according to the manufacture's recommendation, we applied a 6x10 sec vortex interval during the procedure. All PCR steps were performed with Q5® High-Fidelity DNA Polymerase (New England Biolabs) and primers provided by IDT (Integrated DNA Technologies) (**Table D.2**). PCR products were purified with QIAquick PCR Purification Kit (Qiagen).

Design and generation of clustered regularly interspaced short palindromic repeats-FnCas12a plasmids for gene deletion

The broad-host plasmid pMTL83152 (Heap *et al.*, 2009) was used as backbone (**Table D.1**). The gene *Fncas12a* of *Francisella novicida* (Zetsche *et al.*, 2015) was obtained from plasmid pY001 (Addgene #69973) and amplified with primers cas12a_fwd_BamHI and cas12a_rv_NcoI generating *BamHI* and *NcoI* restriction sites for a subsequent restriction cloning to generate pMTL83152_*Fn*Cas12a. Two homology-directed repair arms (HDR1/HDR2) each with a size of 1000-1200 bp, which flank the targeted gene, were individually amplified with HDR_upst_fwd/rv and HDR_dwst_fwd/rv primers generating an overlap of 25-40 bp to each other. The fragments were purified, and 50-100 ng of both fragments were used as template for a subsequent fusion PCR using HDR_upst_fwdOv and HDR_dwst_rvOv primers, which generated new overlaps at 5' and 3' (fusion fragment HDR1/2). An crRNA array was synthesized and cloned as minigene into plasmid pUC19 by IDT (Integrated DNA Technologies) (**Table D.3**). The crRNA array sequence contained the mini-promoter P4 (5'-TTGACAAATTTATTTTTTAAAGTTAAAATTAAGTTG-3') (Xu *et al.*, 2015), the FnCas12a-specific directed repeats (DR) sequence (5'-TAATTTCTACTGTTGTAG AT-3') (Zetsche *et al.*, 2015), 1-2 sgRNA for the targeted gene(s) (Pam sequence TTV for target RNF and TTTV for target *rseC* and *nar*), and the rrbn-T1-terminator (Orosz *et al.*, 1991). The crRNA array fragment was amplified with primers minigene_crRNA_fwd/rv creating overhangs to the fused HDR1/2 fragment and the plasmid backbone. For gene targets with a size >2 kb, such as *rnfCDGEAB* and *nar*, two sgRNA (and two DRs) were used in the same crRNA array (**Table D.3**). For the assembly reaction (Gibson Assembly Ultra Kit, Synthetic Genomics), the plasmid pMTL83152_*Fncas12a* was first digested using *BbvCI* and CIP (New England Biolabs) for 3h at 37°C, purified by PCR-clean, and then mixed with the purified fused HDR1/2 fragment and the crRNA array fragment. Using electrocompetent *E. coli* EPI300 cells

(TransforMax^{TM}, Lucigen) and electroporation for transformation highly increased cloning efficiency for the CRISPR-Cas12a constructs in *E. coli*. For inducible Cas12a expression, the P_{thl} module was replaced with the tetR-O1 promoter module ($P_{tetR-O1}$) (Woolston *et al.*, 2018) using restriction sites *SbfI* and *BamHI*, for all generated CRISPR-Cas12a plasmids.

Generation of overexpression and complementation plasmids

The *rnfCDGEAB* gene cluster (CLJU_c11360-410) and a 213-bp sequence located upstream of *rnfC*, which contains the putative native promoter sequence (P_{nat}), were amplified as one fragment using primers rnfCDGEAB+213bp_fwd and rnfCDGEAB_rv. The *rseC* gene (CLJU_c11350) was amplified using primers *rseC*_fwd and *rseC*_rv. The gene cluster CLJU_c23710-30, here referred to as *nar*, was amplified as one fragment using primers nar-full_fwd and nar-full_rv. All PCR products were purified with the QIAquick PCR Purification kit (Qiagen). Subsequently, the purified fragments were ligated into pMinit2.0 (New England Biolabs) and used for transformation of $CaCl_2$-competent *E. coli* TOP10 cells (Sambrook and Russell, 2006). Next, the plasmid DNA was digested using the restriction sites determined by the used PCR primers and the fragment was cloned into the pMTL83151 plasmid generating pMTL83151_P_{nat}_*rnfCDGEAB* or into the pMTL83152 plasmid generating pMTL83152_*rseC* and pMTL83152_*nar*. Subsequently, all cloned fragments were verified again with test-digestion of the plasmid DNA and Sanger sequencing to exclude mutations in the gene sequences.

6.4.4 Screening for correct plasmid DNA and genome editing

For screening and continuous purity control of our *C. ljungdahlii* strains (**Table D.4**), we performed PCRs from culture samples or from purified DNA with the Phire Plant Direct PCR Master Mix (Thermo Fischer Scientific). *E. coli* colonies grown on selective LB plates after receiving plasmid constructs were analyzed for the correctly assembled plasmids using the Phire Plant Direct PCR Master Mix (Thermo Fischer Scientific). A small amount of recombinant *E. coli* cell material was directly transferred to the reaction mix. For *C. ljungdahlii* cells, 0.5-1 mL culture sample was harvested by centrifugation for 3 min at 13806 rpm (Centrifuge 5424, FA-45-24-11, Eppendorf) and resuspended in 100-500 μL 10 mM NaOH depending on the size of the cell pellet.

Subsequently, cell suspensions were boiled for 10 min at 98°C. The hot reaction tubes were incubated on ice for 1 min and quickly vortexed before they served as a DNA template. In general, we used 20 μL PCR master mix, which consisted of 10 μL Phire Plant Mix 2x, 0.8 μL of each primer, 1 μL cell lysate sample, and 7.4 μL nuclease-free water. The PCR reaction was carried out according to the manufacture's guidelines. We generally used the primers tra60bp_fwd and repH_401bp_rv or repH_643bp_rv for these control PCRs (**Table D.2**), because they bind to the plasmid backbone of every pMTL plasmid used in this study. Verification of gene deletion in the genome of *C. ljungdahlii* was performed with "outside" primers, which bound upstream and downstream of the used homology-directed repair arms (HDR1/2) on the genomic DNA (**Table D.2**). In addition, we performed test-digestion of the generated plasmids with restriction enzymes (New England Biolabs), and analyzed the fragment pattern *via* gel electrophoresis. The final plasmid sequence was verified by Sanger sequencing. Plasmid DNA was purified from *E. coli* with self-made purification buffers (described below). Correct plasmid DNA was then purified with the QIAprep Spin Miniprep kit (Qiagen) *prior* to further use.

6.4.5 A fast method for plasmid purification from *E. coli* without use of a commercial kit

For screening of successfully transformed *E. coli* cells we used a time- and money-saving protocol to purify plasmid DNA from multiple samples without using a commercial kit, which is a modified alkaline lysis protocol adapted from (Sambrook, 1989). All centrifugation steps were performed at 13806 rpm for 5 min (Centrifuge 5424, FA-45-24-11, Eppendorf). Recombinant *E. coli* cells were grown overnight in 5 mL selective liquid LB at 37°C and 150 rpm. 1.5-3 mL cell suspension were harvested in 1.5 mL reaction tubes. The supernatant was discarded, and the pellet was resuspended by vortexing in 150 μL P1-buffer (50 mM Tris, 10 mM EDTA, 100 μg/mL RNAseA, pH 8.0 with HCl). Cells were lysed in 150 μL P2-buffer (200 mM NaOH, 1 vol-% SDS) and inverted five times. Proteins were precipitated by adding 250 μL P3-buffer (2.55 M Na-acetate, pH 4.8 with acetic acid). The samples were inverted five times and centrifuged. Subsequently, 500 μL of the supernatant were transferred into new 1.5 mL tubes and mixed with 500 μL isopropanol. The samples were quickly vortexed and centrifuged again. Afterwards, the supernatant was discarded. At this step,

the precipitated and non-visible plasmid-DNA pellet remained on the bottom of the tube. The pellet was washed twice with ice-cold ethanol (70 vol-%) omitting resuspending the DNA. After the second washing, the supernatant was discarded completely and the remaining ethanol was first removed by snapping the tube on a piece of clean paper towel and then through drying at 50-65°C for 10 min. The dried pellet was resuspended in 30 μL elution buffer (Tris/EDTA, pH 7.2) or deionized water. Purified plasmid-DNA with a concentration of 250-500 ng/μL was clean enough for subsequent cloning steps and test-digestion, however, an additional clean-up with the QIAquick PCR Purification Kit (Qiagen) was carried out when a Sanger sequencing reaction was necessary. P1-buffer needed to be stored at 4°C to maintain RNAse activity for up to 3 months. P2- and P3-buffer were stored at room temperature.

6.4.6 A modified conjugation protocol for *C. ljungdahlii*

This protocol was adapted and modified according to Mock *et al.* (2015). Cells of *C. ljungdahlii* were grown in RCM overnight to mid exponential growth until an OD_{600} of 0.4-0.8 was reached (NanoPhotometer® NP80, Implen). *E. coli* HB101 pKR2013 (DSM 5599) harboring the desired CRISPR-Cas12a-plasmid was grown as pre-culture in selective 5 mL LB medium overnight. The plasmid pKR2013 contains essential genes to mediate conjugation and a kanamycin resistance cassette. 1-2 mL of the *E. coli* cells were used to inoculate 10 mL selective LB medium in 50 mL baffled flask and cultivated until mid-exponential growth (OD_{600} 0.5-1.0). Subsequently, the *E. coli* culture was cooled to 4°C and 2 mL were transferred into sterile 2 mL reaction tubes. The *C. ljungdahlii* culture was kept at room temperature until use. Inside the anaerobic chamber, *E. coli* cells were centrifuged softly at 2900 rpm (mySpin™ 12 mini centrifuge, Thermo Fischer Scientific) to protect pili, and washed once with sterile and anaerobic 0.1 M phosphate buffered saline (PBS) at pH 6.0. Afterwards, the washed pellet was resuspended gently in 100-150 μL cell suspension of *C. ljungdahlii* and directly transferred to well-dried RCM-agar plates (2 vol-% agar). Spot-mating was carried out at 37°C inside the anaerobic chamber overnight. After 8-24 h the spot was resuspended with anaerobic PBS (pH 6.0) and centrifuged at 10000 rpm (mySpinTM 12 mini centrifuge, Thermo Fischer Scientific) for 3 min. The supernatant was discarded, and the pellet was resuspended in the remaining volume of the tube. Subsequently, 100 μL of the cell suspension was plated onto selective PETC+5gS-agar plates, which

contained 5 g/L of peptone and 5 g/L meat extract to support growth. Selective agar plates should not be older than 2-3 days. Thiamphenicol was added for plasmid selectivity. Trimethoprim (10 mg/mL) was added to counter-select against *E. coli*. Growth was obtained after 4-5 days at 37°C inside the anaerobic chamber. *C. ljungdahlii* colonies were transferred into Hungate tubes containing 5 mL RCM with the respective antibiotics. A successful transformation of *C. ljungdahlii* with the correct plasmid was confirmed as follow: **1)** growth in selective RCM with a characteristic pH decrease due to acetogenesis; **2)** control PCRs with primers for plasmid specific fragments; and **3)** plasmid purification from the culture and re-transformation into *E. coli* TOP10 cells.

6.4.7 Electroporation of *C. ljungdahlii* cells

Electroporation of *C. ljungdahlii* cells was performed as previously reported (Xia *et al.*, 2020) and applied for all non-CRISPR-based plasmids. Single colonies growing on selective plates were verified by PCR analyses and by re-transformation *E. coli* with plasmid DNA, which was extracted from *C. ljungdahlii*.

6.4.8 Growth experiments with *C. ljungdahlii*

In general, all recombinant *C. ljungdahlii* strains were pre-grown in 50 mL RCM in 100 mL serum bottles for 24-48 h. Subsequently, 2 mL cell suspension were used to inoculate 50 mL PETC medium in 100 mL serum bottles. This PETC pre-culture was cultivated for 40-48h at 37°C until mid-exponential growth phase at OD_{600} of 0.5-1.0. Afterwards, cells were transferred anaerobically into 50 mL reaction tubes, which were equilibrated for 3-5 days inside the anaerobic chamber. Cell harvest was performed outside the anaerobic chamber at 3700 rpm for 12 min (Centrifuge 5920R, S-4x1000, Eppendorf) at room temperature. After the centrifugation, the tubes were transferred back immediately into the anaerobic chamber, to keep the time at aerobic conditions at a minimum. Inside the anaerobic chamber, the supernatant was discarded, and the pellet was resuspended in fresh PETC medium to adjust to an OD_{600} of 5-10. The concentrated cell suspension was then transferred into sterile and anaerobic 10 mL Hungate tubes, sealed carefully, and used to inoculate main cultures outside of the anaerobic chamber. 1 mL of the cell suspension was used to inoculate 100 mL PETC

main cultures. For heterotrophic growth experiment, 240 mL serum bottles were used. Autotrophic growth experiments were performed in 1000 mL Duran pressure plus bottles (Schott), to provide a high medium-to-headspace ratio. The Duran pressure plus bottles were sealed with butyl stoppers and a GL45 ring cap. Before inoculation of autotrophic cultures, the N_2 headspace was replaced with a sterile gas mixture consisting of H_2/CO_2 (80/20 vol-%). Each bottle contained 0.5 bar overpressure. All cultures were cultivated in biological triplicates as batch cultures. The gas headspace was not refilled during the experiments. However, for the strain *C. ljungdahlii* pMTL83151_P_{nat}_*rnfCDGEAB* and the control strain *C. ljungdahlii* pMTL83151 we refilled the headspace during this experiment with the same gas mixture to 0.5 bar overpressure at time points 44.5 h, 73.5 h, and 148.5 h (**Results, 6.5.4**). We did not measure the headspace gas composition during our experiments. Culture samples of 3 mL were taken at the bench and used for: **1)** OD_{600} measurement; **2)** pH measurement; **3)** HPLC analyses (acetate and ethanol); and **4)** FIA analyses (nitrate, nitrite, and ammonium). All culture samples were stored at -20°C until use. OD_{600} samples were diluted with medium or PBS buffer when the absorbance was > 0.5. We applied a two-tailed student's t-test for all cultivation data. All p-values (P) below 0.001 indicate high significance and are given as ≤ 0.001.

6.4.9 HPLC analyses

HPLC analyzes were performed as described before (Klask *et al.*, 2020). In addition, all frozen supernatant samples were thawed at 30°C for 10 min and 300 rpm, vortexed briefly, and centrifuged for 3 min at 13806 rpm (Centrifuge 5424, FA-45-24-11, Eppendorf) before use. All HPLC samples were randomized.

6.4.10 Measurement of nitrate, nitrite, and ammonium

Nitrate and nitrite concentrations were measured in a FIA continuous-flow analyzer system (AA3 HR AutoAnalyzer System, Seal Analytical GmbH, Germany) as described before (Klueglein *et al.*, 2014). Briefly, nitrate is reduced to nitrite with hydrazine and then reacts with sulfanilamide and NEDD (N-1-Naphthylethylenediamine di-HCl) to form a pink complex, which can be quantified photo-metrically at 550 nm. The protocol follows the DIN 38405/ISO 13395

standard methods. Ammonium concentrations were measured in the same system but with salicylate and dichloroisocyanuric acid forming a blue complex that is measured at 660 nm instead. The protocol was following DIN 38406/ISO 11732 standard methods. Culture samples of *C. ljungdahlii* were treated as explained above for HPLC preparation. However, we prepared 1:50 dilution in 1 mL with deionized water *prior* to the FIA analyses. Standards for nitrate, nitrite, and ammonium were measured before and during the analyses for a standard curve and to minimize drift effects. Nitrate concentrations of each sample were calculated by the difference of the amount of nitrite measured with and without the *prior* reduction by hydrazine.

6.4.11 Growth experiment for RNA extraction from *C. ljungdahlii*

For the expression analyses, we grew the strains *C. ljungdahlii* WT, *C. ljungdahlii* ΔRNF, and *C. ljungdahlii* ΔrseC under autotrophic and heterotrophic conditions as described above. The cultivation medium was PETC with ammonium as nitrogen source. Pre-cultures were grown in heterotrophic medium for 48 h. Next, the cells were transferred into the anaerobic chamber and harvested for 12 min at 25°C and 3700 rpm (Centrifuge 5920 R, S-4x1000, Eppendorf) outside of the anaerobic chamber. The supernatant was discarded under anaerobic conditions and the pellet was resuspended in fresh medium of the main cultures. The start OD_{600} for autotrophic main cultures was 0.2, while it was 0.15 for heterotrophic conditions. The cultures were cultivated at 37°C. 10 mL culture samples were taken after 3 h and 20 h. The samples were immediately cooled on ice and centrifuged for 12 min at 4°C and 3700 rpm (Centrifuge 5920 R, S-4x1000, Eppendorf). The cell pellets were stored at -20°C until RNA extraction.

RNA was purified from *C. ljungdahlii* with the RNeasy Mini Kit (Qiagen) as described before (Liu et al., 2013). For the RNA extraction, we used 2·108 cells, which was approximately 10 mL of a *C. ljungdahlii* culture at OD_{600} 0.2. The cell lysis was performed in the lysis buffer of the kit with 50 mg glass beads (0.1 mm silica spheres, MP Biomedicals) in a bead beater (5G-FastPrep, MP Biomedicals) for 2x 60s at 9 m/s. RNA samples were eluted in 30 μL nuclease-free water. After the extraction procedure, an additional DNase I digest (RNase free Kit, Thermo-Scientific) was performed to remove potential DNA contamination. Elimination of genomic DNA was confirmed with PCR analyses and gel electrophoresis. cDNA synthesis was performed with the QuantiTect Reverse Transcriptase Kit (Qiagen) according to the manufacturer´s

instructions. We used 500 ng RNA as template for each reaction. cDNA was stored at -20°C until further use.

6.4.12 qRT-PCR analyses

All qRT-PCR analyses were performed in a Quantstudio 3 Thermocycler (Applied Biosystems, Thermo Scientific). The PCR reaction mix contained 10 μL SYBR Green Master mix (Thermo Scientific), 1 μL of a fwd and rv qRT-PCR primer (final concentration 500 nM) (**Table D.2**), and 1 μL ($\sim$5 ng) cDNA template. We used the rho gene as reference gene, which was described before as suitable candidate for qRT-PCR experiments with *C. ljungdahlii* (Liu *et al.*, 2013). We added RNA controls to further exclude gDNA contamination in our samples. All qRT-PCR reactions were performed in technical triplicate according to the manufacturer's instructions. We set the Ct threshold to 0.1. The fold change in gene expression between the samples was determined with the $2^{-\Delta\Delta Ct}$ method as described before (Livak and Schmittgen, 2001). We examined the PCR efficiency of our qPCR master mix by using plasmid DNA containing the sequence of each tested gene in a series of dilutions (10^{-1}, 10^{-2}, $5{\cdot}10^{-3}$, 10^{-3}, $5{\cdot}10^{-4}$, 10^{-4}). The slopes were ranging from 0.04-0.09 for the RNF-gene cluster genes and 0.17 for *rseC*, and were thus, close to zero, which proofs that the efficiencies are similar and the $2^{-\Delta\Delta Ct}$ can be used for interpretation of the qRT-PCR data (Livak and Schmittgen, 2001). We applied a two-tailed Student's t-test based on our ΔCt values for each gene to analyze the significance of our samples in comparison to the wild type.

6.4.13 Strain preservation

Cultures of *C. ljungdahlii* were stored at -80°C. For this, cultures were grown in RCM until late exponential growth phase (OD_{600} 0.8-1.2) at 37°C for 36-48 h. The cells were transferred into anaerobic 50 mL reaction tubes inside the anaerobic chamber and harvested outside of the anaerobic chamber for 12 min at 4°C and 3700 rpm (Centrifuge 5920 R, S-4x1000, Eppendorf). The supernatant was discarded inside the anaerobic chamber and the pellet was resuspended in fresh RCM medium to an OD_{600} of 5-10. 2 mL of the cell suspension was transferred into 10 mL serum bottles, which were previously filled with 2 mL of 25-50 vol-% anaerobic and autoclaved glycerol. The serum bottles were briefly vortexed outside the anaerobic chamber, incubated on ice for

10-15 min and subsequently frozen at -80°C. For inoculation of a new RCM culture, a single serum bottle was quickly thawed up under rinsing water and 1-2 mL of the cell suspension was immediately transferred with a syringe into the medium bottle. Cultures of *E. coli* were stored at -80°C in sterile screw-cap tubes filled with 25-50 vol-% glycerol.

6.5 Results

6.5.1 A full deletion of the RNF complex confirmed its indispensable role for autotrophy in *C. ljungdahlii*

First, we achieved a full deletion of the RNF-gene cluster in *C. ljungdahlii* with a clustered regularly interspaced short palindromic repeats (CRISPR)-associated protein Cas12a (CRISPR-Cas12a) system, which we implemented and used to generate all deletion strains in this study (**Figure 6.1A, Text D.1.1**). We confirmed the identity of the *C. ljungdahlii* ΔRNF strain (**Figure 6.1B, Text D.1.2**), and compared the growth of *C. ljungdahlii* wild-type (WT) to the growth of *C. ljungdahlii* ΔRNF. We performed growth experiments with carbon dioxide and hydrogen (autotrophy) and with fructose (heterotrophy), while we added equimolar amounts of either ammonium or nitrate as nitrogen source to the medium for both autotrophy and heterotrophy (**Materials and Methods, 6.4, Figure 6.2, D.1**). As expected, we observed growth for *C. ljungdahlii* WT in all growth experiments (**Figure 6.2A, D.1A**). However, the nitrogen source had a distinct influence on the growth rate, final OD_{600}, fermentation product spectrum, and pH (**Table 6.1, Figure 6.2B, Text D.1.3, Figure D.1B**). We found that nitrate reduction occurred rapidly in our growth experiments (**Figure 6.2F, D.1F**). *C. ljungdahlii* WT utilized all provided nitrate within 53 h of cultivation with carbon dioxide and hydrogen (**Figure 6.2F**) and within 47 h of cultivation with fructose (**Figure D.1F**). The ammonium concentrations increased concomitant with decreasing nitrate concentrations when nitrate was provided in the medium (**Figure 6.2E, Figure D.1E**). Noteworthy, we also observed an increase in the ammonium concentration when ammonium was provided as the nitrogen source during autotrophy (**Figure 6.2F**). We had added a small amount of yeast extract (0.1 weight-%) in all cultivation conditions (**Material and Methods, 6.4**). We found this increase in the ammonium concentration

in all our cultivation experiments, and we argued that the additional ammonium was a by-product of the fermentation of the added yeast extract in our cultivation medium (**Materials and Methods, 6.4, Text D.1.3**). We did not measure any nitrite as an intermediate of the nitrate reduction pathway (discussed below).

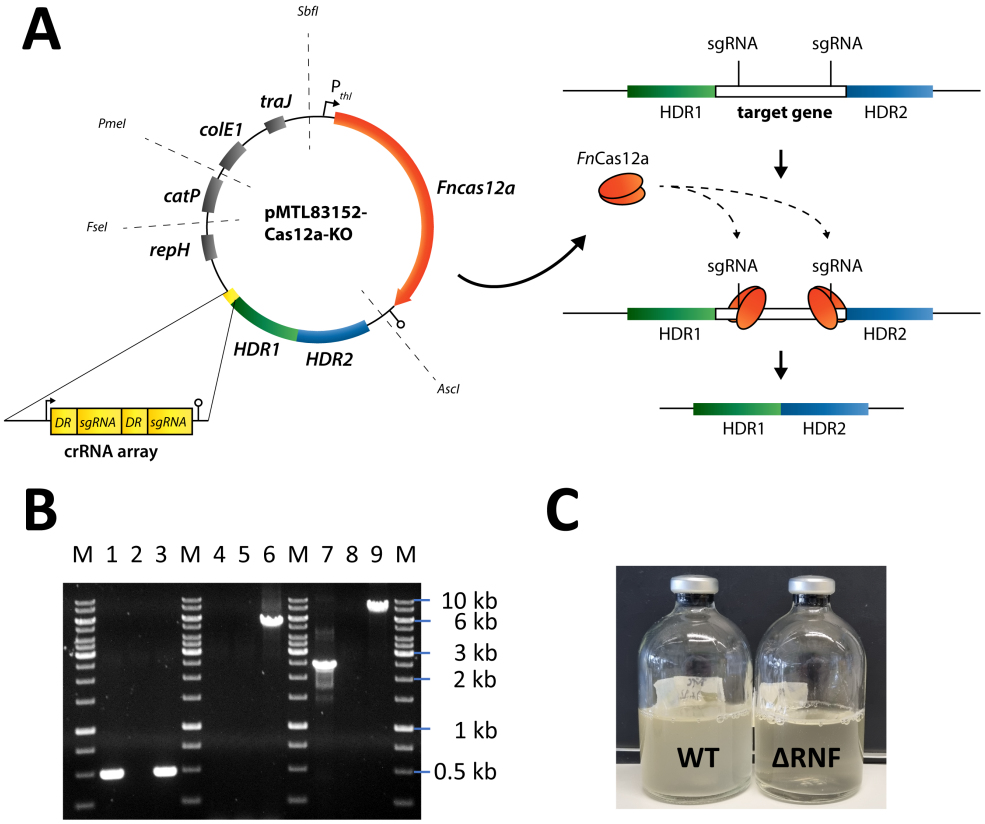

Figure 6.1: CRISPR-Cas12a-mediated *rnfCDGEAB* gene cluster deletion in *C. ljungdahlii*. **(A)** modular CRISPR-Cas12a system established in the pMTL80000 shuttle-vector system (Heap *et al.*, 2009). The final CRISPR-Cas12a plasmid for deletion of *rnfCDGEAB* contained the *Fncas12a* gene, homology-directed repair arms (HDRs), and a specific crRNA array comprising two directed repeats (DRs) and two sgRNA, which targeted the *rnfC* and *rnfB* genes. **(B)** agarose gel with PCR-samples for the *fdhA* fragment (WT: 501 bp, deletion strain: 501 bp), *rnfCDGEAB* fragment (WT: 5047 bp, deletion strain: no fragment), and for a fragment that was amplified with primers that bind ∼1250 bp upstream and downstream of the *rnfCDGEAB* gene cluster locus (WT: 7550 bp, deletion strain: 2503 bp). DNA-template: gDNA of *C. ljungdahlii* ΔRNF (lane 1, 4, and 7); gDNA of *C. ljungdahlii* WT (lane 3, 6, and 9); and water (lane 2, 5, 8). M: Generuler^{TM} 1 kb DNA ladder. **(C)** growth of the wild type (WT) and reduced growth of the deletion strain (ΔRNF) with fructose in PETC medium. HDR1/2, homology-directed repair arm flanking the targeted gene; crRNA array, sequence containing *FnCas12a*-specific DRs and sgRNAs; sgRNA, guide RNA; *repH*, Gram-positive origin of replication; *catP*, antibiotic resistant cassette against chloramphenicol/thiamphenicol; *colE1*, Gram-negative origin of replication; *traJ*, conjugation gene; P_{thl}, promoter sequence of the thiolase gene in *Clostridium acetobutylicum*; *AscI, FseI, PmeI,* and *SbfI* are unique-cutting restriction sites, which were preserved during the cloning to maintain the modular functionality of the plasmid backbone.

Table 6.1 Performance of all tested *C. ljungdahlii* strains in autotrophic batch cultivation experiments. Cultures were grown with carbon dioxide and hydrogen (autotrophy) in PETC medium, which contained either ammonium or nitrate as nitrogen source. A gas atmosphere of H_2/CO_2 (80/20 vol-%) with 0.5 bar overpressure was applied. Growth was not detected for any culture of *C. ljungdahlii* ΔRNF or *C. ljungdahlii* ΔrseC. Data is represented as mean values from biological triplicates ± standard deviation. WT, *C. ljungdahlii* wild type; ΔRNF, *C. ljungdahlii* with deleted *rnfCDGEAB* gene cluster; ΔrseC, *C. ljungdahlii* with deleted rseC gene; and Δnar, *C. ljungdahlii* with deleted nitrate reductase gene cluster. CO_2, carbon dioxide; and H_2, hydrogen. Given in percentage is the difference in performance in comparison to the wild type with the same nitrogen source.

strain	**nitrogen source**	**growth rate (μ in h)**[a]	**maximum OD_{600} value**	**maximum acetate concentration (mM)**	**maximum ethanol concentration (mM)**
WT	ammonium	0.024±0.002	0.56±0.01	59.5±1.8	1.9±0.4
WT	nitrate	0.072±0.004	1.00±0.06	50.1±2.1	8.0±1.6
ΔRNF	ammonium	-	-	5.7±3.0	n.d.[b]
ΔRNF	nitrate	-	-	2.3±1.1	n.d.[b]
Δ*rseC*	ammonium	-	-	2.0±0.5	n.d.[b]
Δ*rseC*	nitrate	-	-	1.9±0.1	n.d.[b]
Δ*nar*	ammonium	0.018±0.001 (-24%, *)	0.44±0.01 (-21%,***)	44.8±0.2 (-25%, ***)	3.3±0.2 (+79%, *)
Δ*nar*	nitrate	0.017±0.003 (-76%, ***)	0.44±0.01 (-55%,***)	41.9±1.9 (16%,*)	2.9±0.4 (-64%,*)

[a] μ values were calculated based on the individual OD_{600} values of each triplicate in the exponential growth phase.
[b] n.d., not detectable
[c] n.s., not significant ($P > 0.05$)
*, significant ($P \leq 0.05$)
**, significant ($P \leq 0.01$)
***, significant ($P \leq 0.001$)

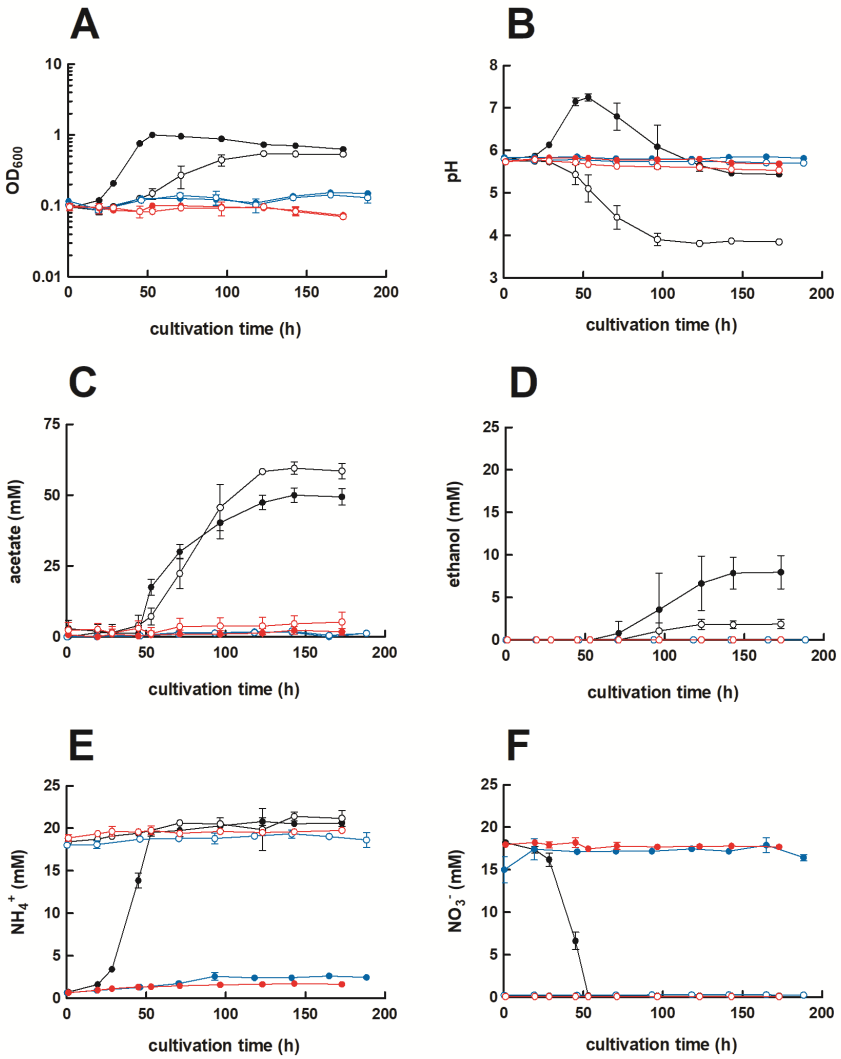

Figure 6.2: Cultivation of *C. ljungdahlii* WT, *C. ljungdahlii* ΔRNF, and *C. ljungdahlii* Δ*rseC* in nitrate- or ammonium-containing medium with H_2 and CO_2. Cultures of *C. ljungdahlii* strain WT (●, ○), ΔRNF (●, ○), and ΔrseC (●, ○) were grown in 100 mL PETC medium in 1 L bottles at 37°C and 150 rpm. The headspace consisted of H_2 and CO_2 (80/20 vol-%) and was set to 0.5 bar overpressure. The medium contained either 18.7 mM nitrate (NO_3^-) (filled circles) or 18.7 mM ammonium (NH_4^+) (open circles) as nitrogen source. The cultivation times were 173 h for cultures of *C. ljungdahlii* WT and *C. ljungdahlii* ΔRNF and 186 h for cultures of *C. ljungdahlii* Δ*rseC*. All cultures were grown in biological triplicates, data is given as mean values, with error bars indicating the standard deviation. **(A)** growth; **(B)** pH-behavior; **(C)** acetate concentrations; **(D)** ethanol concentration; **(E)** ammonium concentration; and **(F)** nitrate concentrations. WT, wild type; ΔRNF, RNF-gene cluster deletion; Δ*rseC*, *rseC* gene deletion.

6.5.2 The deletion of the RNF complex influenced nitrate reduction during heterotrophy

For the *C. ljungdahlii* ΔRNF strain, heterotrophic growth with fructose was still possible but notably reduced (**Figure 6.1C, D.1A**). Compared to the wild type, we observed significant reduction for *C. ljungdahlii* ΔRNF with ammonium and nitrate, respectively, of the growth rates (-34%, -42%), maximum OD_{600} values (-53%, -56%), and maximum acetate concentrations (-32%, -42%) (**Table 6.1, D.5**). The maximum ethanol concentration was significantly reduced with ammonium (-41%), and ethanol was not produced at all with nitrate (**Table 6.1, Figure D.1C, D.1D**). During heterotrophy, *C. ljungdahlii* ΔRNF was able to utilize nitrate but considerably slower than the wild type (**Figure D.1F**). At the end of the cultivation, cultures of *C. ljungdahlii* ΔRNF had only consumed 49% of the provided nitrate (**Figure D.1F**). Overall, we observed a halt in growth and metabolic activity for cultures of *C. ljungdahlii* ΔRNF with fructose after 47 h of cultivation in nitrate-containing medium and after 56 h of cultivation in ammonium-containing medium (**Figure D.1**). Fructose concentrations at the end of the cultivation remained at a concentration of 8.0-9.7 mM, which is still 30-35% of the initially provided concentration (**Figure D.1G**). The pH did not increase during heterotrophy with nitrate in *C. ljungdahlii* ΔRNF, but instead slowly decreased until the end of the cultivation (**Figure D.1B**). Notably, the final pH for heterotrophic cultures of *C. ljungdahlii* ΔRNF with nitrate was still higher compared to cultures with ammonium (**Figure D.1B**).

6.5.3 The *rseC* gene is essential for autotrophy in *C. ljungdahlii*

Next, we investigated the role of the small putative regulator gene *rseC* (CLJU_c11350). We applied our CRISPR-Cas12a system to delete the *rseC* gene from the genome (**Figure D.2A**). We performed growth experiments with the generated *C. ljungdahlii* Δ*rseC* strain under the same conditions as for the *C. ljungdahlii* WT and ΔRNF strains. Cultures of *C. ljungdahlii* Δ*rseC* did not grow with carbon dioxide and hydrogen, neither with ammonium nor with nitrate, during a total cultivation time of 189 h (**Figure 6.2**). Non-growing cultures for this strain did not accumulate notable concentrations of acetate or ethanol during the cultivation time (**Figure 6.2C, 6.2D**). Furthermore, we did not observe nitrate reduction or a remarkable change in pH as a metabolic activity of

non-growing cultures for this strain during autotrophy (**Figure 6.2B, 6.2E, 6.2F**).

However, heterotrophic growth of *C. ljungdahlii* Δ*rseC* was possible, and in contrast to *C. ljungdahlii* ΔRNF, the impact was less pronounced for growth with ammonium but limited to some extent with nitrate (**Figure D.1A**). In comparison to the wild type with ammonium, only the maximum ethanol concentration (-29%) was significantly reduced (**Table D.5, Figure D.1C**). In contrast, with nitrate, the heterotrophic growth rates (-34%), maximum OD_{600} (-30%), and maximum ethanol concentrations (-42%) were all significantly reduced, while the maximum acetate concentration was not (**Table D.5, Figure D.1D**). Nitrate reduction was not restricted during heterotrophy in *C. ljungdahlii* Δ*rseC* (**Figure D.1E, D.1F**). Indeed, we observed a rapid utilization of all supplied nitrate within 60 h of cultivation, which is similar to the observations that we had made for the wild type (**Figure D.1F**). Thus, *rseC* seems to be involved in regulating the expression of the RNF-gene cluster during autotrophy, but not during heterotrophy. However, the exact impact on gene expression of the RNF-gene cluster cannot be deduced from these findings.

6.5.4 Plasmid-based complementation relieved the phenotypes of the *C. ljungdahlii* ΔRNF and Δ*rseC* strains

We questioned whether the wild-type phenotype, particularly with respect to autotrophy, can be restored by plasmid-based gene complementation in the *C. ljungdahlii* ΔRNF and *C. ljungdahlii* Δ*rseC* strains. Therefore, we generated the plasmid-carrying strains *C. ljungdahlii* ΔRNF pMTL83151_P_{nat}_*rnfCDGEAB* and *C. ljungdahlii* Δ*rseC* pMTL83152_rseC. The plasmids encode the RNF-gene cluster under the control of the native promoter region upstream of the *rnfC* gene from the genome (P_{nat}) in pMTL83151_P_{nat}_*rnfCDGEAB* and the *rseC* gene under the control of the constitutive thiolase promoter (P_{thl}) in pMTL83152_*rseC*, respectively. We investigated the complementation strains in ammonium-containing medium with carbon dioxide and hydrogen for growth (**Figure D.3**). Indeed, the plasmid-based expression of the deleted genes relieved the phenotype and enabled autotrophy with carbon dioxide and hydrogen for both strains (**Table 6.2, Figure D.3**). The control strains that carried an empty plasmid failed to grow autotrophically, as we had already observed for the non-complemented deletion strains.

For *C. ljungdahlii* ΔRNF pMTL83151_P_{nat}_rnfCDGEAB we found a significantly

reduced maximum OD_{600} (-71%), and maximum acetate concentration (-22%) (**Table 6.1, Figure D.3A, D.3B, D.3C**), while the maximum ethanol concentration was similar to the wild type (**Figure D.3D**). Furthermore, the complemented strain had a prolonged lag phase of 71 h (**Figure D.3A**). Notably, the medium for the complementation experiments always contained antibiotics, which generally caused a slightly negative impact on growth of plasmid-carrying *C. ljungdahlii* strains such as just described for the *C. ljungdahlii* ΔRNF pMTL83151_P_{nat}_*rnfCDGEAB* strain. Surprisingly, this was not the case for the *C. ljungdahlii* Δ*rseC* pMTL83152_*rseC* strain. Instead of a prolonged lag phase, we observed a slightly shortened lag phase for this strain (**Figure 6.1A, 6.3A**). Compared to the wild type, this strain reached a slightly but significantly increased maximum OD_{600} (+17%) and maximum acetate concentration (+6%) (**Table 6.2, Figure D.3A**). However, this strain did not produce any detectable ethanol during the cultivation (**Table 6.2, Figure D.3D**). Furthermore, the pH value did not show any notable change (**Figure D.3B**). The differences in the performance of our complementation strains might be explained by the plasmid copy number and the differences in the used promoter sequences for the complementation plasmids, while this hypothesis awaits further experimentation (**Text D.1.4**).

Table 6.2 Performance of the plasmid-based complemented deletion strains of *C. ljungdahlii* in autotrophic batch cultivation experiments. Cultures were grown with carbon dioxide and hydrogen (autotrophy) in PETC medium, which contained either ammonium or nitrate as nitrogen source. A gas atmosphere of H_2/CO_2 (80/20 vol-%) with 0.5 bar overpressure was applied. Data is represented as mean values from biological triplicates ± standard deviation. The WT data from **Table 6.1** are shown again for comparison. WT, wild type; ΔRNF, deletion of the *rnfCDGEAB* gene cluster; Δ*rseC*, deletion of the *rseC* gene; Δ*nar*, deletion of the nitrate reductase gene cluster; ΔRNF compl., complementation strain *C. ljungdahlii* pMTL83151_P_n*at*_*rnfCDGEAB*; Δ*rseC* compl., complementation strain *C. ljungdahlii* pMTL83152_*rseC*; and Δ*nar* compl., complementation strain *C. ljungdahlii* pMTL83152_*nar*. Given in percentage is the difference in performance in comparison to the wild type with the same nitrogen source.

strain	nitrogen source	growth rate (μ in h)[a]	maximum OD_{600} value	maximum acetate concentration (mM)	maximum ethanol concentration (mM)
WT	ammonium	0.024±0.002	0.56±0.01	59.5±1.8	1.9±0.4
Δ RNF compl.	ammonium	0.024±0.001 (-3%, **)	0.40±0.03 (-29%,**)	46.7±3.7 (-22%, **)	2.2±0.2 (+18%, n.s.[c])
Δ *rseC* compl.	ammonium	0.022±0.002 (-8%, n.s.[c])	0.66±0.03 (+17%, *)	63.2±0.2 (+6%, *)	n.d.[b]
WT	nitrate	0.072± 0.004	1.00±0.06	50.1±2.1	8.0±1.6
Δ *nar* compl.	nitrate	0.054± 0.001 (-26%, **)	1.54±0.03 (+54%, ***)	41.7±2.5 (-17%, *)	3.4±0.5 (-58%, *)

[a] μ values were calculated based on the individual OD_{600} values of each triplicate in the exponential growth phase.
[b] n.d., not detectable
[c] n.s., not significant ($P > 0.05$)
[*] significant ($P \leq 0.05$)
[**] significant ($P \leq 0.01$)
[***] significant ($P \leq 0.001$)

6.5.5 The gene expression profiles of *rnf* genes and the *rseC* gene in the deletion strains revealed regulatory effects

We further investigated the activating or repressing function on the gene expression of the RNF-gene cluster by RseC. For this, we performed qRT-PCR analyses to investigate the individual expression profiles of the genes *rnfC*, *rnfD*, *rnfG*, *rnfE*, *rnfA*, *rnfB*, and *rseC* in the *C. ljungdahlii* Δ*rseC* strain. We included the *C. ljungdahlii* ΔRNF and *C. ljungdahlii* WT strains as controls (**Figure D.5, Materials and Methods**). We analyzed samples after 3 h and 20 h to investigate the transcriptomic response after inoculating the autotrophic and heterotrophic main cultures from heterotrophic pre-cultures. During the cultivation of the six main cultures (three strains, two conditions), *C. ljungdahlii*

WT grew during autotrophy and heterotrophy, while *C. ljungdahlii* ΔRNF and *C. ljungdahlii* Δ*rseC* only grew during heterotrophy (**Figure D.5C, D.5D**). Thus, we exposed non-growing cells of *C. ljungdahlii* ΔRNF and *C. ljungdahlii* Δ*rseC* to an autotrophic environment and collected samples after 3 h and 20 h. We argued that this would result in a transcriptomic response, even though the cultures did not show a difference in OD_{600} over the 20-h incubation time of this experiment (*C. ljungdahlii* ΔRNF, 0.20±0.02 and *C. ljungdahlii* Δ*rseC*, 0.22±0.02, respectively), but remained at the inoculation OD_{600}. In comparison, for the wild type, 1-2 generations can be expected within 20 h of cultivation during autotrophy, which resulted in an increase of the OD_{600} from 0.28 to 0.39 (**Figure D.5C**).

The qRT-PCR results in this paragraph are given as $\log_2$ (fold change ($\log_2$FC) in gene expression), where we discuss a value of $\leq$ -2 (0.25-fold) as a biologically relevant downregulation, and a value of $\geq$ +2 (4-fold) as a biologically relevant upregulation (**Figure 6.3**). We did not measure any expression signals for any of the deleted RNF genes in the *C. ljungdahlii* ΔRNF strain and for the deleted *rseC* gene in the *C. ljungdahlii* Δ*rseC* strain. We found that all RNF-gene cluster genes (except for *rnfE*, -1.8±0.2) were downregulated (ranging from -2.2±0.1 [*rnfD*] to -4.7±0.1 [*rnfC*]) in the *C. ljungdahlii* Δ*rseC* strain, when exposing non-growing cells of this strain to hydrogen and carbon dioxide (**Figure 6.3**). We observed a similar pattern of downregulation for the 3-h and 20-h samples of the *C. ljungdahlii* Δ*rseC* strain (**Figure 6.3A, 6.3C**). In the heterotrophic samples, the RNF-gene cluster genes were not downregulated in the 3-h samples (ranging from -1.0±0.1 [*rnfC*] to -1.8±0.1 [*rnfE*]). After 20 h of cultivation time during heterotrophy, we even observed a pattern in which none of the genes had a hllog2 (fold change in gene expression) of $\leq$ -1 or $\geq$ +1. In the *C. ljungdahlii* ΔRNF strain as a control, we found that rseC expression was upregulated in the 3-h samples during autotrophy (+2.4±0.04) but not during heterotrophy (+1.3±0.1) (**Figure 6.3B**). In the 20-h samples both conditions were not different (**Figure 6.3D**). For the wild type, all genes (except for *rnfD* and *rnfE* in the 3-h sample) were upregulated during autotrophy when compared to heterotrophy for the 3-h samples (ranging from +2.3±0.03 [*rnfB*] to +5.4±0.1 [*rseC*]), and for the 20-h samples (ranging from +2.8±0.04 [*rnfB*] to +3.8±0.1 [*rnfE*]), respectively (**Figure D.5**). Thus, from our qPCR results we concluded that RseC likely has a positive regulatory impact on the RNF-gene cluster during autotrophy, but not during heterotrophy, while we cannot rule out that this regulation was an indirect effect, which involves further regulatory elements, from these

experiments.

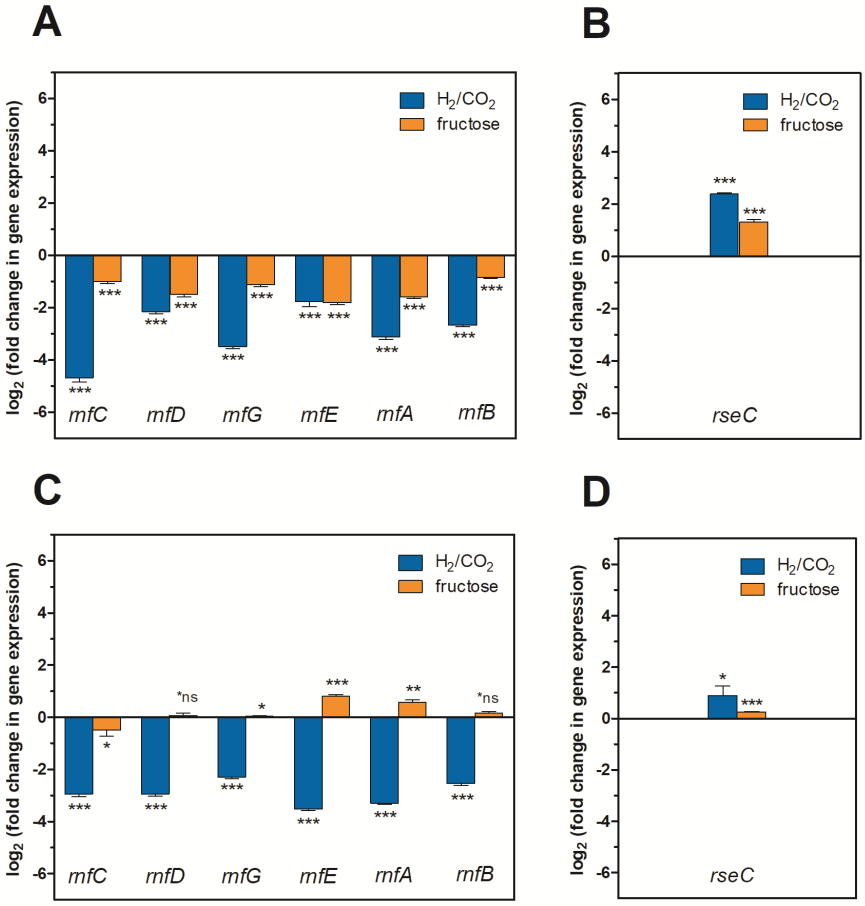

Figure 6.3: Gene expression change of the *rnfCDGEAB* cluster genes and the *rseC* gene in the ΔRNF and Δ*rseC* deletion strains. **(A)** gene expression change for the genes *rnfC, rnfD, rnfG, rnfE, rnfA*, and *rnfB* in strain *C. ljungdahlii* Δ*rseC* after 3h cultivation time; **(B)** gene expression change for the gene *rseC* in strain *C. ljungdahlii* ΔRNF after 3 h cultivation time; **(C)** gene expression change for the genes *rnfC, rnfD, rnfG, rnfE, rnfA,* and *rnfB* in strain *C. ljungdahlii* Δ*rseC* after 20 h cultivation time; and **(D)** gene expression change for the gene *rseC* in strain *C. ljungdahlii* ΔRNF after 20 h cultivation time. RNA samples were purified from cultures that were cultivated either autotrophically with hydrogen and carbon dioxide (blue bars) or heterotrophically with fructose (orange bars). cDNA was synthesized from the purified RNA samples and used as template for qRT-PCR analyses. The individual gene expression profiles of each gene was calculated using the wild-type strain as reference, which was grown under the same conditions. The *rho* gene was used as "housekeeping" gene. The fold change in gene expression was determined with the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001). ***, $P \leq 0.001$; **, $P \leq 0.01$; *, $P \leq 0.05$; *ns, not significant ($P > 0.05$). We defined $\log_2$FC $\leq$ -2 as downregulated genes and $\geq$ +2 as upregulated genes.

6.5.6 The *rseC* gene is abundantly found among acetogens

Based on our previous findings, we investigated whether *rseC* genes are present in the genomes of selected other microbes, including the most prominent model acetogens. We searched for putative *rseC* genes in the RNF-complex gene-containing genomes of the model acetogens *C. ljungdahlii*, *Clostridium autoethanogenum*, *A. woodii*, *Eubacterium limosum*, *Clostridium carboxidovorans*, and the model non-acetogen *Clostridium kluyveri* (**Figure 6.4, Table 6.3**). Indeed, we found putative *rseC* genes for all these candidates. We looked at the genomic location and distance of the *rseC* gene to the RNF-gene cluster (**Figure 6.4, Table 6.3**). We noticed that the *rseC* gene was located directly upstream of the RNF complex gene cluster in *C. ljungdahlii* (CLJU_c11350), *C. autoethanogenum* (CAETHG_3225), *C. carboxidovorans* (Ccar_25725), and *C. kluyveri* (CKL_1263). However, the *rseC* gene in *A. woodii* (Awo_C21740) and *E. limosum* (B2M23_08890) was not in direct genetic vicinity of the RNF-gene cluster (**Figure 6.4, Table 6.3**). In addition, we identified a second gene with homologies to *rseC* in *C. carboxidovorans* (Cca_07835) and *C. kluyveri* (CKL_2767), but neither RNF-complex genes nor other genes that are involved in the autotrophic metabolism, such as the genes for the Wood-Ljungdahl pathway, are located in the direct vicinity of these second *rseC* homologs (**Figure 6.4, Table 6.3**). Furthermore, the *rseC* protein sequence seems to be highly conserved in the microbes that contain an RNF-gene cluster (**Figure D.6, Text D.1.6**).

In contrast, we did not find a putative *rseC* gene when we searched the genomes of two further model acetogens *Moorella thermoacetica* and *Thermoanaerobacter kivui*, which possess an energy-converting hydrogenase (Ech) complex instead of an RNF complex (Hess *et al.*, 2014). Notably, we also identified a putative *rseC* gene in the non-acetogenic bacteria *R. capsulatus*, which is the microbe in which the RNF complex was first described (Schmehl *et al.*, 1993), and in *Thermotoga maritima* (Schmehl *et al.*, 1993). The *rseC* gene in *R. capsulatus* is located upstream of *rnfF* instead of *rnfC*, which is separated by five genes (**Figure 6.4**). In *T. maritima*, RNF genes were not found next to the putative *rseC* gene (THEMA_1487) (**Figure 6.4**). Also *E. coli* possesses one *rseC* gene that is organized in the *rseABC* gene cluster, which plays an important role in the SoxR-mediated oxidative stress response as described elsewhere (**Figure 6.4, Text D.1.5**) (Koo *et al.*, 2003).

These findings led to the question, whether we can find an acetogen that contains an

RNF-complex but no *rseC* gene. To answer this question, we performed a more extensive and automated genome search. We based this search on a collection of 47 out of 61 acetogens with a fully sequenced genome from Bengelsdorf *et al.* (2018), which are stored in the German Collection of Microorganisms and Cell Cultures (Deutsche Sammlung von Mikroorganismen und Zellkulturen, DSMZ) and American Type Culture Collection (ATCC). In this automated genome search, we confirmed our initial findings, and in total identified 30 acetogens that contained potential RNF-complex genes (**Table D.7**). Seven of these acetogens (*Acetitomaculum ruminis* DSM 5522, *Blautia hydrogenotrophica* DSM 10507, *Blautia schinkii*, *Marvinbryantia formatexigens* DSM 14469, *Oxobacter pfennigii*, *Terrisporobacter mayombei*, and *Treponema primitia* ZAS-2) did not show evidence for an *rseC* gene. We identified several acetogens that contained more than one potential *rseC* gene (**Table D.7**). Interestingly, we found three acetogens (*Acetobacterium fimetarium*, Clostridium magnum DSM 2767, and *Oxobacter pfennigii*), which have both potential RNF-complex genes and Ech-complex genes in their genomes, while of those *Oxobacter pfennigii* was not found to contain an *rseC* gene as mentioned above (**Table D.7**).

Overall, we found with our genome search that not all investigated acetogens that contain RNF genes have putative *rseC* genes in their genomes. Some acetogens have several putative *rseC* homologs, and some of these putative *rseC* genes are associated with the RNF genes, while others are not. We did not identify an acetogen that contained an *rseC* gene and exclusively an Ech-complex but no RNF-complex (**Table 6.3, D.7**). The reasons for these observations and the regulatory effects in acetogens with and without an *rseC* gene will be important future research questions.

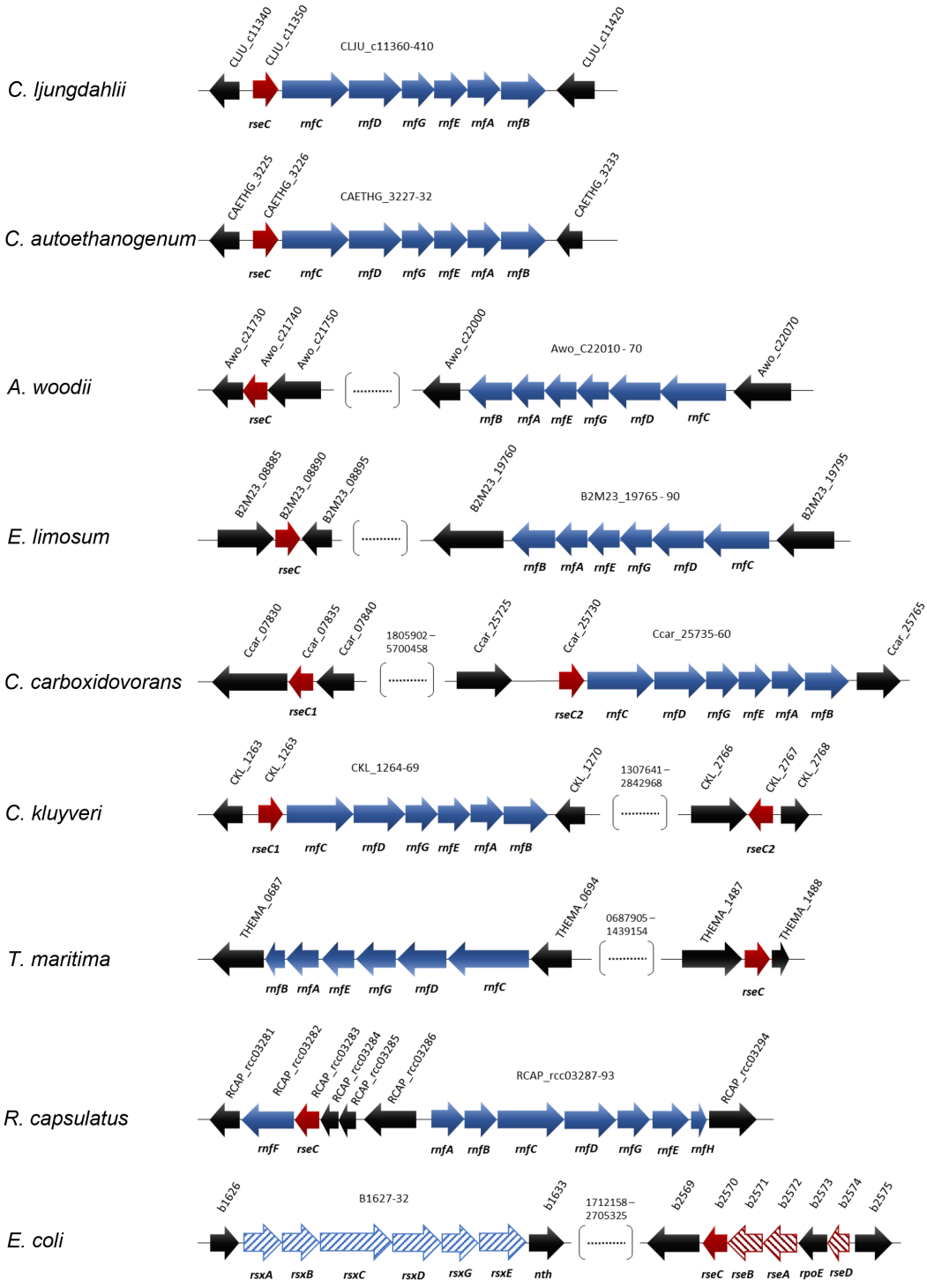

Figure 6.4: (Caption on the next page.)

Figure 6.4: (Previous page.) **Location and orientation of *rseC* genes in model microbes that possess RNF complex gene clusters.** The conserved protein domain RseC_MucC (pfam04246) was identified in the *rseC* protein sequence of *C. ljungdahlii* and used to search for putative *rseC* genes in the genome of *C. autoethanogenum*, *A. woodii*, *E. limosum*, *C. carboxidovorans*, *C. kluyveri*, *T. maritima*, *R. capsulatus*, and *E. coli*. All sequence analyses and gene arrangements were adapted from the JGI platform and the NCBI database (03/2021). The type strains are listed in **Table 6.3**. In red, putative *rseC* genes; in red pattern fill, *rseC*-associated genes in *E. coli*; in blue, RNF-complex gene cluster; in blue pattern fill, *rsx* genes, which are homologous to the *rnf* genes in *R. capsulatus*.

Table 6.3 Distribution of *rseC* genes in model microbes.

Microbe[a]	Amount of *rseC* genes[b]	RNF or Ech	*rseC* associated with RNF genes	Gene locus
Clostridium ljungdahlii	1	RNF[c]	yes	CJLU_c11350
Clostridium autoethanogenum	1	RNF[c]	yes	CAETHG_3226
Clostridium carboxidovorans	2	RNF[c]	yes, one of them	Ccar_07835, Ccar_025730
Clostridium kluyveri	2	RNF[c]	yes, one of them	CKL_1263, CKL_2767
Eubacterium limosum	1	RNF[d]	no	B2M23_08890
Acetobacterium woodii	1	RNF[d]	no	Awo_c21740
Thermotoga maritima	1	RNF[d]	no	THEMA_1487
Moorella thermoacetica	0	Ech	no	-
Thermoanaerobacter kivui	0	Ech	no	-
Rhodobacter capsulatus	1	RNF[e]	yes	RCAP_rcc03283
Escherichia coli	1	Rsx[f]	no, but with Rsx	b2570

[a] The type strains were: *C. ljungdahlii* DSM13528; *C. autoethanogenum* DSM10061; *C. carboxidovorans* P7; *C. kluyveri* DSM555; *E. limosum* ATCC8486; *A. woodii* DSM1030; *T. maritima* DSM3109; *M. thermoacetica* ATCC39073; *T. kivui* DSM2030; *R. capsulatus* SB1003; and *E. coli* K-12.
[b] The pfam domain pfam04426 was used to search for putative *rseC* genes in each genome.
[c] The RNF complex uses (or is supposed to use) protons.
[d] The RNF complex uses (or is supposed to use) sodium ions.
[e] The RNF complex either uses protons or sodium ions. Experimental data are missing.
[f] Rsx is encoded by *rsxABCDGE* and is homologous to the RNF-gene cluster in *R. capsulatus*.

6.5.7 The *nar* gene cluster encodes a functional nitrate reductase in *C. ljungdahlii*

We had found that nitrate reduction during heterotrophy is impacted for the *C. ljungdahlii* ΔRNF strain but not the *C. ljungdahlii* Δ*rseC* strain. Thus, we aimed to explore nitrate metabolism and the interplay with the RNF complex further. For *C. ljungdahlii*, it was postulated that nitrate is reduced by nitrate reductase to nitrite and, subsequently, converted *via*nitrite reductase and hydroxylamine reductase into ammonium, and the involved genes were predicted in the genome (Köpke *et al.*, 2010; Nagarajan *et al.*, 2013). Emerson *et al.* (2019) had found that in the presence of nitrate the expression level of the genes that encode the putative nitrate reductase (CLJU_c23710-30) were significantly increased. The three genes are annotated as nitrate reductase NADH oxidase subunit (CLJU_c23710), nitrate reductase electron transfer subunit (CLJU_c23720), and nitrate reductase catalytic subunit (CLJU_c23730) (Köpke *et al.*, 2010). We refer to these three genes (CLJU_c23710-30) as the *nar* gene cluster. We verified the absence of the *nar* gene cluster from the genome of the *C. ljungdahlii* Δ*nar* strain, after mediating the deletion with our CRISPR-Cas12a system (**Figure D.2B**). This strain was able to grow during autotrophy and heterotrophy, but had completely lost the ability to reduce nitrate under both conditions (**Figure 6.5F, D.7F**). Consequently, we observed similar growth and pH behavior for cultures of *C. ljungdahlii* Δ*nar* during autotrophy with either ammonium or nitrate (**Figure 6.5A, 6.6B, D.7A, D.7B**). Enhanced autotrophic growth in nitrate-containing medium when compared to ammonium-containing medium, such as with the wild-type strain, was not detected (**Figure 6.5A**).

However, we still observed differences in the growth when compared to the wild type. For ammonium and nitrate conditions, respectively, growth rates (-24%, -76%), maximum OD_{600} (-21%, -55%), and maximum acetate concentrations (-25%, -16%) during autotrophy of *C. ljungdahlii* Δ*nar* were significantly reduced (**Table 6.1, Figure 6.5A**). Instead, the maximum ethanol concentrations were significantly increased (+79%) for ammonium, but significantly decreased (-64%) for nitrate conditions, respectively (**Figure 6.5D, Table 6.1**). A pH increase as a consequence of ammonium production from nitrate reduction, such as observed for the wild type, was not observed in cultures of *C. ljungdahlii* Δ*nar* (**Figure 6.5B**).

For heterotrophic cultures with ammonium, we did not observe significant differences

in growth behavior and acetate production to the wild type. Only the maximum ethanol concentration was significantly increased (+45%) (**Table D.5, Figure D.7**). With fructose and nitrate, the maximum observed OD_{600} (-32%) and maximum acetate concentration (-34%) were significantly reduced, while the maximum ethanol concentration (+234%) was significantly increased (**Table D.5, Figure D.7D**). The provided fructose was only consumed completely by *C. ljungdahlii* Δ*nar* in ammonium-containing but not in nitrate-containing medium (**Figure D.7G**).

Finally, we confirmed that the complementation of *C. ljungdahlii* Δ*nar* with the plasmid pMTL83152_*nar*, which encodes the *nar* gene cluster under the expression control of the constitutive P_{thl} promoter, enabled the *C. ljungdahlii* Δ*nar* pMTL83152_*nar* strain to utilize nitrate under autotrophic conditions again, while this was not possible in an empty plasmid control strain (**Figure D.8**). The nitrate cultures of *C. ljungdahlii* Δ*nar* pMTL83152_*nar* had significantly different growth rates (-26%), maximum OD_{600} (+54%), as well as maximum acetate (-17%) and ethanol (-57%) concentrations in comparison to the wild type when growing with nitrate (**Table 6.2, Figure D.8**). In summary, we revealed that the expression of the *nar* gene cluster led to the only functional nitrate reductase in *C. ljungdahlii* under the tested conditions.

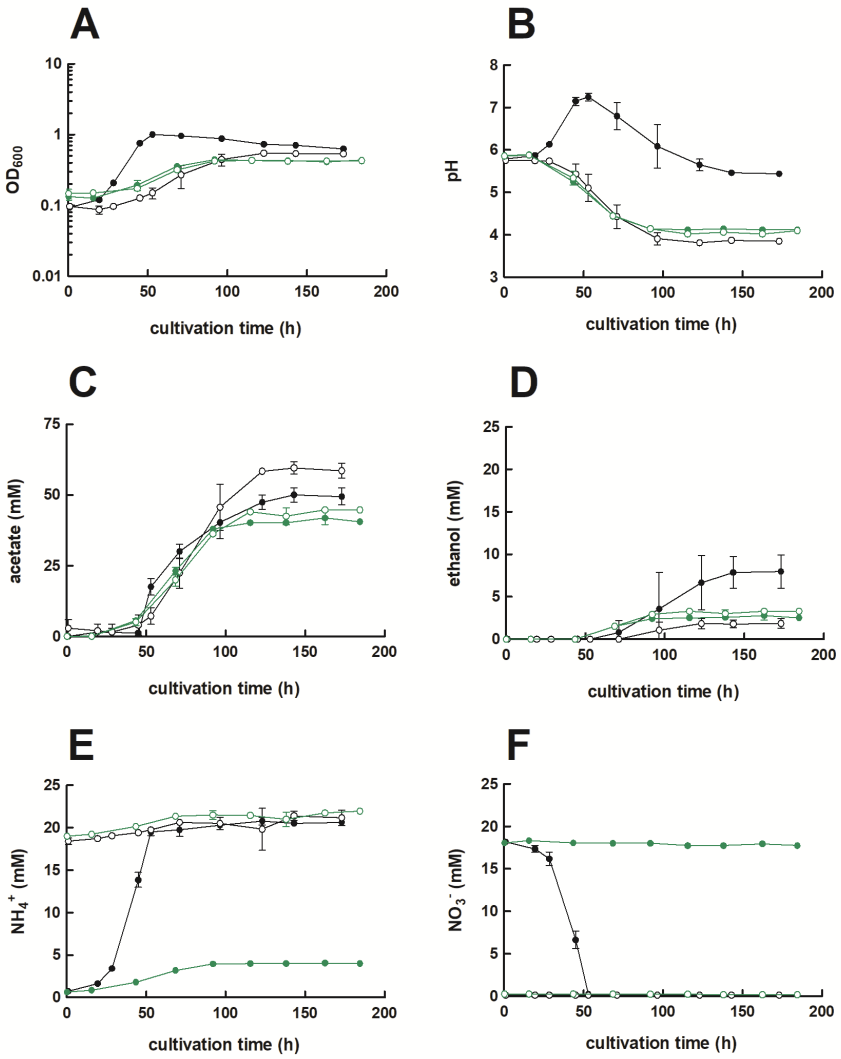

Figure 6.5: Growth, pH behavior, nitrate reduction of *C. ljungdahlii* Δ*nar* with H_2 and CO_2. Cultures were grown in 100 mL PETC medium in 1 L bottles at 37°C and 150 rpm for 185 h. The headspace consisted of H_2 and CO_2 (80/20 vol-%) and was set to 0.5 bar overpressure. The medium contained either 18.7 mM nitrate (NO_3^-) (●) or 18.7 mM ammonium (NH_4^+) (○) as nitrogen source. The *C. ljungdahlii* WT data (●, ○) from Fig. 6.1 is given for comparison. All cultures were grown in biological triplicates, data is given as mean values, with error bars indicating the standard deviation. **(A)** growth; **(B)** pH-behavior; **(C)** acetate concentrations; **(D)** ethanol concentration; **(E)** ammonium concentration; and **(F)**, nitrate concentrations. Δ*nar*, deletion of nitrate reductase gene cluster; rpm, revolutions per minute; CO_2, carbon dioxide; and H_2, hydrogen.

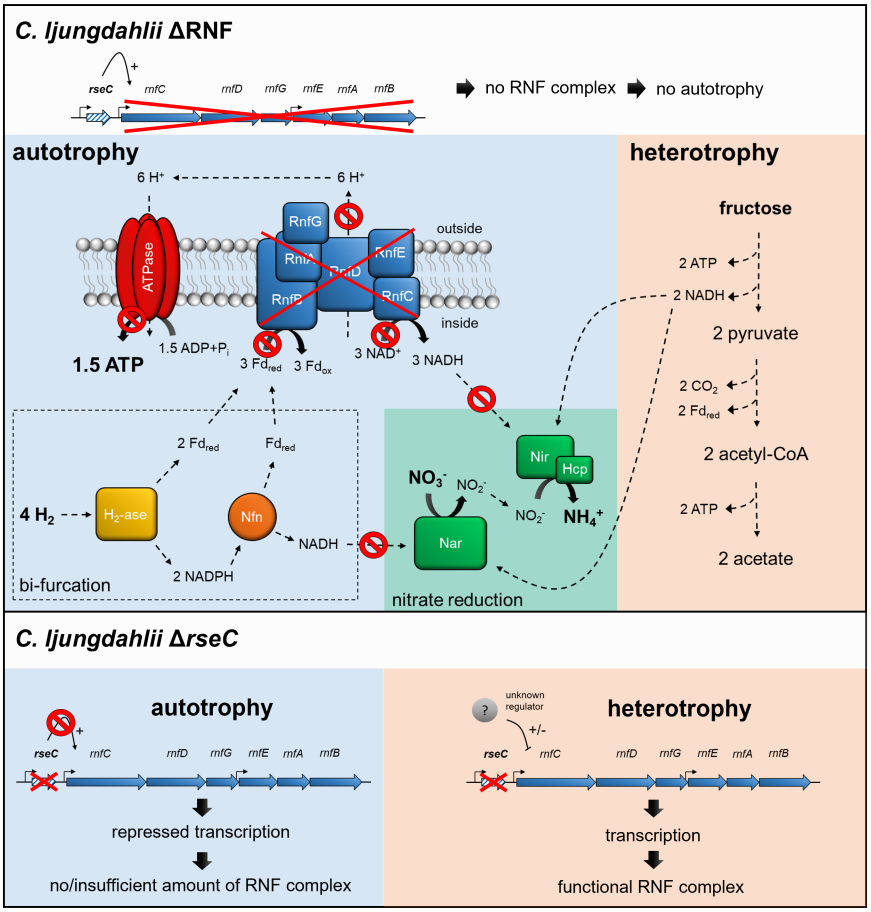

Figure 6.6: Schematic model of RNF-gene regulation and nitrate reduction in the deletion strains *C. ljungdahlii* ΔRNF and *C. ljungdahlii* Δ*rseC* during autotrophy and heterotrophy. In both deletion strains, nitrate reduction is not possible in non-growing cells during autotrophy with carbon dioxide and hydrogen due to the lack of a functional RNF complex, and thus the missing regeneration of reducing equivalents such as NADH. On the contrary, nitrate reduction can proceed in *C. ljungdahlii* ΔRNF during heterotrophy with NADH, which is provided by glycolysis of fructose. In *C. ljungdahlii* Δ*rseC*, the RNF complex genes are repressed during autotrophy but not during heterotrophy, which indicates a further unknown regulation mechanism during heterotrophy. Thus, a functional RNF complex is formed, and nitrate reduction can proceed such as proposed for the wild type. Abbreviations: H_2, hydrogen; H^+, proton, CO_2, carbon dioxide; NO_3^-, nitrate; NO_2^-, nitrite; NH_4^+, ammonium; ATP, adenosine triphosphate; ADP + Pi, adenosine diphosphate + phosphate; $Fd_{red/ox}$, reduced/oxidized ferredoxin; NADH/NAD^+, reduced/oxidize nicotinamide adenine dinucleotide; NADPH/$NADP^+$, reduced/oxidized nicotinamide adenine dinucleotide phosphate; RnfCDGEAB, RNF-complex subunits; Nar, nitrate reductase; Nir, nitrite reductase; Hcp, hydroxylamine reductase; H_2-ase, bifurcating hydrogenase/lyase; Nfn, bifurcating transhydrogenase; e^-, electron; ΔRNF, *C. ljungdahlii* ΔRNF; and Δ*rseC*, *C. ljungdahlii* Δ*rseC*. The model was adapted from Emerson *et al.* (2019).

6.6 Discussion

6.6.1 A functional RNF complex is essential for autotrophy but not for heterotrophy in *C. ljungdahlii*

Here, we provided further insight into the autotrophy of *C. ljungdahlii* and the connection to nitrate metabolism. With the strain *C. ljungdahlii* ΔRNF, we confirmed previous work by Tremblay *et al.* (2012) that the absence of the RNF complex leads to a complete loss of autotrophy in *C. ljungdahlii*. Unlike in the previous study by Tremblay *et al.* (2012), our strain provides a stable genotype that cannot revert back to the wild-type genotype, which can be used to further study the energy conservation principles in this acetogen (**Figure 6.1B, 6.2**). Heterotrophic growth in our deletion strain was still possible, but considerably reduced when compared to the wild type (**Figure 6.1C, D.1**), which was expected from the work by Tremblay *et al.* (2012). While we did not measure the difference in the headspace gas composition during heterotrophy for *C. ljungdahlii* ΔRNF and wild type, we argue that *C. ljungdahlii* ΔRNF lost the ability to fixate the carbon dioxide that is released during glycolysis, which is the defining feature of acetogens (Drake *et al.*, 2008; Schuchmann and Müller, 2014). Thus, even though the Wood-Ljungdahl pathway was still present, this strain was not able to balance the electrons in the metabolism sufficiently to drive the Wood-Ljungdahl pathway and/or in other metabolic pathways, which resulted in a remaining 30% of fructose that was not consumed until the end of the batch cultivation (**Figure D.1**). However, further research is required to confirm this hypothesis. The RNF deletion in *A. woodii* did also lead to reduced acetate production during heterotrophy, but the strain reached similar OD_{600} values compared to the *A. woodii* wild type (Westphal *et al.*, 2018). In comparison to *C. ljungdahlii*, the RNF complex of *A. woodii* uses sodium ions instead of protons to generate the chemiosmotic gradient, which is then consumed by a sodium-dependent F_1F_O ATPase to generate ATP (Biegel and Müller, 2010; Hess *et al.*, 2013). Overall, this further confirms the meticulous differences in the energy conservation and redox balancing in different acetogens (Katsyv and Müller, 2020), which have to be considered to apply acetogens for biotechnological purposes.

6.6.2 RseC is a regulator of the RNF complex genes and plays a critical role during autotrophy

We further investigated the regulation of the RNF-gene cluster by the putative regulator RseC. The *rseC* gene is known to encode a transcriptional regulator in other microbes such as *E. coli* and *S. typhimurium* (**Text D.1.5**) (Beck *et al.*, 1997; De Las Peñas *et al.*, 1997; Yura and Nakahigashi, 1999; Koo *et al.*, 2003). Our results demonstrated that RseC played a critical role for the formation of a functional RNF complex in *C. ljungdahlii* (**Table 6.2, Figure 6.2**). A deletion of the *rseC* gene led to the complete loss of autotrophy (**Figure 6.2**). With our qRT-PCR analyses, we confirmed that RseC, indeed, had a positive regulatory effect on the expression of the RNF-gene cluster during autotrophy. Our results indicate that RseC is essential for the activation of RNF-gene cluster expression during autotrophy, but not during heterotrophy, while we cannot rule out that this activation is mediated by other modulating activities such as secondary regulators (**Figure 6.3A, 6.3C, Figure 6.6, Figure D.5**). Al-Bassam *et al.* (2018) identified a transcription start site upstream of *rseC* and found that *rseC* was poorly translated under heterotrophic conditions similar to *rnfC* in *C. ljungdahlii*. Two transcription start sites were identified within the RNF-complex gene cluster *rnfCDGEAB*, which could be preceded by potential promoter-binding sites for RseC (Al-Bassam *et al.*, 2018). We can speculate that RseC most likely binds to the transcription start site, which is located upstream of the *rnfC* gene, because the entire RNF-gene cluster was downregulated during autotrophy and heterotrophy after 3 h of cultivation, (**Figure 6.3A**). Based on our qRT-PCR results from the heterotrophic samples after 20 h of cultivation, it can be further speculated that the second transcription start site, which is located upstream of *rnfE* (Al-Bassam et al., 2018), might be a second binding site for RseC. We see a slight upregulation of the three genes *rnfE*, *rnfA*, and *rnfB* in these samples for the *rseC* deletion compared to the wild-type strain. Thus, we argue that RseC could act as a negative regulator to modulate RNF-gene cluster expression during heterotrophy by repressing the genes *rnfE*, *rnfA*, and *rnfB* (**Figure 6.3C**). The plasmid-based complementation of *rseC* in the *C. ljungdahlii* Δ*rseC* strain (and overexpression in the wild-type background) re-enabled growth with carbon dioxide and hydrogen, and reduced the lag phase during the transition from heterotrophy to autotrophy (**Table 6.2, Figure D.3, Figure D.4**). This further argues for a function of RseC as a positive regulator of the RNF-gene cluster.

The RseC protein in *E. coli* seems to contain two transmembrane domains with the C-terminal end being located in the cytoplasm (Daley *et al.*, 2005). Thus, another possible function of RseC could be the modulation of protein-protein interactions with the RNF complex, because for all the RseC homologs that we investigated here, two transmembrane helices were predicted (**Table D.6**). For instance, RseC could stabilize the RNF complex in the membrane, which is required for the electron translocation mechanism or the interaction with other cytoplasmic proteins during autotrophy, but not during heterotrophy. This could explain why a lack of the RseC protein is not leading to the same reduced heterotrophy as observed for the *C. ljungdahlii* ΔRNF strain. The upregulation of the *rseC* gene in the *C. ljungdahlii* ΔRNF strain indicates that the *rseC* gene itself is also under further transcriptional control (**Figure 6.3B, 6.3D**). An upregulation of the *rseC* gene in all samples might indicate the effort of the cells to induce RNF-gene cluster expression further, which is apparently not possible in the *C. ljungdahlii* ΔRNF strain. Thus, potentially additional direct or indirect regulatory effects are mediated by RseC. This could be, for example, the regulation of further genes as a positive or negative regulator, which in turn could have an effect on the functionality of the RNF complex. Importantly, others had identified a TetR-family protein that is involved as an alternative sigma-factor in the regulation of autotrophy *vs.* heterotrophy in *C. autoethanogenum* (de Souza Pinto Lemgruber *et al.*, 2019). de Souza Pinto Lemgruber *et al.* (2019) had identified a promoter sequence upstream of the *rseC* gene in *C. autoethanogenum* (CAETHG_3226) that is likely recognized by this alternative sigma-factor. The same sequence motif is found upstream of the *rseC* gene (CJLU_c11350) in *C. ljungdahlii*. However, the function of the TetR-family protein has not yet been investigated in *C. ljungdahlii*. Therefore, protein-DNA binding experiments should be performed to investigate the binding ability of RseC to putative promoter regions, as well as the regulation of the *rseC* gene itself, in future experiments. In addition, the study of the subcellular localization of RseC will be required to unravel the regulatory functions of RseC in *C. ljungdahlii* and other acetogens with an RNF complex in more detail.

6.6.3 Nitrate reduction does not require a functional RNF complex but benefits from a correct electron balance

Furthermore, we investigated the nitrate metabolism in *C. ljungdahlii*. We confirmed that the genes CLJU_c23710-30 encode the functional subunits of the only nitrate reductase under the tested conditions for *C. ljungdahlii* (**Figure 6.5, D.7, D.8**). In the presence of nitrate, the wild type quickly utilized all nitrate even though we had found in all our growth experiments that a sufficient amount of nitrogen-source was covered by the added yeast extract (**Figure 6.2F, D.1**). Thus, nitrate reduction in *C. ljungdahlii* is mainly used for energy conversion, and therefore must be of a dissimilatory function (Emerson *et al.*, 2019). However, *C. ljungdahlii* neither possesses genes for cytochromes nor for the biosynthesis of ubiquinone, which limits the generation of a chemiosmotic gradient to the RNF complex (Köpke *et al.*, 2010). The nitrate reductase in *C. ljungdahlii* is most likely located in the cytosol rather than associated with the membrane, as one would expect from dissimilatory nitrate reductases in bacteria (Zumft, 1997; Köpke *et al.*, 2010; Nagarajan *et al.*, 2013). This type of nitrate reduction was described as fermentative nitrate reduction and was already observed for other microbes, but is less understood than the assimilatory and dissimilatory nitrate reduction in the microbial world (Hall, 1973; Hasan and Hall, 1975; Seifritz *et al.*, 1993; Emerson *et al.*, 2019). Contrarily, nitrate reduction does not function as an independent energy-conserving pathway, because autotrophic growth is not possible even with the addition of nitrate (**Figure 6.2**). The stoichiometry for nitrate reduction in *C. ljungdahlii* is proposed as follows: $4H_2 + 2H^+ + NO_3^- + 1.5ADP + 1.5Pi \rightleftharpoons 4H_2O + NH_4^+ + 1.5ATP$ with $\Delta rG'^0$ = -150 kJ/mol H_2 (Thauer *et al.*, 1977; Emerson *et al.*, 2019). This mechanism would require electron bifurcation from the hydrogenases and the activity of the RNF complex, but would then provide ATP completely independent of the Wood-Ljungdahl pathway (or more general, independent of the carbon metabolism) (Buckel and Thauer, 2018; Emerson *et al.*, 2019). Thus, we hypothesized that nitrate reduction in *C. ljungdahlii* requires a functional RNF complex for a correct electron balance. Indeed, non-growing cells of both *C. ljungdahlii* ΔRNF and *C. ljungdahlii* Δ*rseC* were not able to reduce nitrate during autotrophy (**Figure 6.2F**). Further biochemical investigations on the enzyme activity under these conditions will have to provide more detailed insight on this hypothesis. However, nitrate reduction still proceeded in both deletion strains

during heterotrophy (**Figure D.1F**). In *C. ljungdahlii* ΔRNF a functional RNF complex was not present during heterotrophy because the RNF-complex encoding genes were deleted, but the required reducing equivalents for nitrate reduction were likely provided by glycolysis (**Figure 6.6**). In contrast, in *C. ljungdahlii* Δ*rseC*, nitrate reduction was not impacted during heterotrophy, because the RNF complex genes were deregulated (expressed) under these conditions and a functional RNF complex was formed (**Figure 6.3, 6.6**). It remains to be answered whether there is a direct interplay between the nitrate reductase and the RNF complex, and whether this interplay is different during heterotrophy and autotrophy.

6.6.4 The electron balance in the deletion strains is impacted beyond nitrate reduction

In general, the reduced growth indicated that *C. ljungdahlii* ΔRNF was not able to balance the electrons from glycolysis efficiently during heterotrophy. This led to the reduction in biomass and acetate production, while ethanol production was completely absent in heterotrophic cultures of *C. ljungdahlii* ΔRNF. This indicates that reducing power for a further reduction of acetate was not available (**Table D.5, Figure D.1**). In the batch experiments of Emerson *et al.* (2019), *C. ljungdahlii* WT did not produce considerable amounts of ethanol when growing with nitrate (and carbon dioxide and hydrogen). When *C. ljungdahlii* WT was cultivated in pH-controlled bioreactors under continuous conditions, enhanced biomass and increased ethanol production rates were observed (Klask *et al.*, 2020). This observation could not be fully explained yet, but it was assumed that electrons are used concomitantly for the reduction of nitrate and for the reduction of acetate. This distribution of electrons changed in the absence of the nitrate reductase in the *C. ljungdahlii* Δ*nar* strain and higher maximum ethanol concentrations were observed already in batch conditions. Thus, the overall electron balance between fermentation products, besides the loss of nitrate reduction activity, is impacted by the deletion of the *nar* gene cluster. A possible explanation for increased ethanol production of the *C. ljungdahlii* Δ*nar* strain during heterotrophy could be that more reducing equivalents (*e.g.*, NADH, NADPH) are available to the alcohol dehydrogenases (ADHs), which would be used to reduce nitrate in the wild type, and which predominantly catalyze the reduction of acetyl-CoA to ethanol during heterotrophy but not during autotrophy in *C. ljungdahlii* (Schuchmann and Müller,

2014; Richter *et al.*, 2016; Liew *et al.*, 2017). Consequently, less acetate should be produced by the *C. ljungdahlii* Δ*nar* strain. Indeed, the *C. ljungdahlii* Δ*nar* strain produced only 66% ($P \leq 0.001$) of the acetate concentration that was measured for the wild type, when growing under heterotrophic conditions with nitrate (**Table D.1, Figure D.1C**). Furthermore, this would argue for NADH or NADPH as the electron donor of the nitrate reductase as proposed by Emerson *et al.* (2019). On the contrary, we observed increased ethanol concentrations also in the presence of ammonium instead of nitrate as nitrogen source. This might be due to an involvement of the nitrate reductase in other processes, but could also be due to genetic polar effects in the *nar* deletion strain. In addition, nitrate reduction could be regulated and controlled differently during autotrophy and heterotrophy. It remains elusive, how the change in the distribution of electrons affects other NADH-dependent metabolic pathways in more detail. While further research, such as the investigation of intracellular NAD(H)/NADP(H) levels, is required to understand the regulatory mechanisms during autotrophy and the mechanism of energy conservation during nitrate reduction, with this work, we provide a deeper insight into the autotrophic metabolism and nitrate reduction in *C. ljungdahlii*.

6.7 Acknowledgements

The plasmid pMTL2tet01gusA was kindly provided by Dr. Gregory Stephanopoulos (Department of Chemical Engineering, Massachusetts Institute of Technology). We thank Dr. PengFei Xia (Environmental Biotechnology Group, University of Tübingen) for his advice for the CRISPR design and Franziska Schädler (Geomicrobiology and Microbial Ecology, University of Tübingen) for her support with the nitrate and ammonium measurements. We thank Nicole Smith (Environmental Biotechnology Group, University of Tübingen) for her support in medium preparations.

6.8 Supplementary information

Supplementary information: summary of gRNAs and protospacers; summary of the base-editing results and *in-silico* evaluation; strains, plasmids, and primers used in this study; figures supporting the main results. These data can be found in **Appendix D**.

Chapter 7

7.1 Closing summary

The results in this dissertation contribute to the implementation of one-carbon (C1) utilizing microbes for environmental biotechnology, including renewable power storage. This dissertation illustrates the advantages of integrating both wet- and dry-laboratory methods for systems biology including metabolic engineering, as well as the strength of interdisciplinary and collaborative work and research. In Chapters 3 and 4, we studied methanogens, which are biocatalysts for biomethanation in power-to-gas technology. In Chapter 3, we built and constrained genome-scale metabolic models (GEMs) using cultivation data, transcriptomics, and proteomics. The combination of these methods bolstered the comparison of three methanogens, which were *M. thermautotrophicus* ΔH, *M. thermautotrophicus* Z-245, and *M. marburgensis* Marburg, elucidating a potential difference in their anabolic formate production metabolism. In Chapter 4, we examined the advantages and disadvantages of sodium formate as a substrate and performed the first continuous bioreactor cultivation of a genetically modified *Methanothermobacter* species. We observed that *M. thermautotrophicus* ΔH, which was modified to carry a catabolic F_{420}-dependent formate dehydrogenase cassette, is not hindered from the genetic modifications, particularly in terms of biomass production. Using the metabolic models, a potential decrease in non-growth associated maintenance energy is identified for formate dehydrogenase cassette containing microbes and for growth on formate. A lower non-growth associated maintenance energy means that there a cell may have more energy available for other functions, particularly growth. From an evolutionary perspective, more biomass is advantageous. However, from a biomethanation perspective, more biomass results in less carbon towards the desired product (methane).

Clostridium autoethanogenum is a microbe that is already employed for commercial

bioethanol production from syngas. *Clostridum ljungdahlii* is closely related to this industrially relevant microbe. In Chapters 5 and 6, I assisted *via* bioinformatics methods to study *Clostridum ljungdahlii*. In Chapter 5, I implemented a program that identified potential editing sites of a CRISPR system and the result of the mutations based on the genome sequence. I determined that, despite the A-T rich genome of *C. ljungdahlii*, there were over 314 thousand possible editing sites on the genome, which would allow for the editing of 99.8% of the CDSs of which 81.4% could be non-sense mutations, demonstrating the power of the CRISPR-Cas9 system. In Chapter 6, I semi-automatedly compared 47 acetogen genomes for the presence of the *rseC* gene and the RNF- and Ech-gene clusters. The analysis identified the presence of the *rseC* in 30 acetogens, of which the majority had the RNF-gene clusters with flanking *rseC*. This supports the hypothesis that the *rseC* is crucial for autotrophic growth that requires the RNF-gene complex.

7.2 Recommendations for future work

7.2.1 Updated and expanded applications of the GEMs

In Chapter 3, the construction of the GEMs for the three *Methanothermobacter* microbes was described, and in Chapters 3 and 4 the models were employed. Nevertheless, GEMs are never truly ever finished. This is demonstrated with the model organism *E. coli*, whose models have been updated frequently for almost 20 years as new knowledge is gained. The three *Methanothermobacter* spp. are less studied than *E. coli*, which is one of the model microbes used for systems biology and metabolic engineering. With more than 500 hypothetical genes, there is a large capacity for new studies to be conducted and for the GEMs to be updated. For example, a recent study by Baumann *et al.* (2022), investigated the lipid composition in *M. marburgensis* Marburg. This lipid composition can be included in the next iteration of the *M. marburgensis* Marburg GEM. Generally, the biomass component of the GEMs could be improved by conducting both elemental and macroscopic component analysis. The former could generate a more accurate molecular formula (which is needed to normalize uptake and production rates), and the latter could improve the biomass reaction. It is important to note that the biomass composition is likely to change under different environmental growth conditions (Schulz *et al.*, 2021). While the amount of carbon that reaches the

biomass component of *Methanothermobacter* may be small (Martin *et al.*, 2013), studies have argued the importance of the biomass reaction for accurately modeling metabolisms (Chan *et al.*, 2017).

Deep learning tools could be implemented to not only elucidate the functions of hypothetical genes, but also to identify (new) transcription factors (TFs). DeepTFactor (Kim *et al.*, 2021) and PredicTF (Oliveira Monteiro *et al.*, 2022) are tools designed to find TFs for prokaryotes and eukaryotes or only bacteria, respectively. Regulation in archaea has different features than in bacteria, thus, it is important that archaea are also considered in these tools. TFs are important when building global regulatory networks, and various methods exist to construct these models (Imam *et al.*, 2015; Zhang *et al.*, 2022; Song *et al.*, 2017; Feist *et al.*, 2009; Faria *et al.*, 2014). Recently, Prathiviraj and Chellapandi (2020) generated an intial genome-scale regulatory network for *M. thermautotrophicus* ΔH. It is important to note that this model is purely based on computational and little curated data, and thus it should only serve as preliminary model. Their model could be: **1)** updated with new TFs detection algorithms; **2)** applied to the other methanogenic microbes; and **3)** integrated with the GEMs and transcriptomics data presented in this dissertation. Particularly, the transcriptomics data of (the genetically modified) *M. thermautotrophicus* ΔH grown on different substrates (Chapter 4) should provide more insight regarding regulational changes in the presence or absence of molecular hydrogen (Shalvarjian and Nayak, 2021; Zhang *et al.*, 2022).

When more biochemical data, in particular protein kinetics data, would become available, the ability to operate more advanced models, such as kinetic models or ME-models, would arise. Nevertheless, it may be possible to estimate some of the kinetic properties, such as the k_{cat} values, through machine and deep learning methods (Heckmann *et al.*, 2018; Kroll *et al.*, 2021). However, it is important not to neglect the effect of the training data sets implemented on the outcome of these learning algorithms. Further, it is crucial to acknowledge that the under-representation of archaea in these studies and thus, in the potential data sets may lead to a larger biases in the outcomes. Additional modeling strategies to consider may include constraining the models using: **1)** protein allocation (Nilsson *et al.*, 2017a; Jahn *et al.*, 2021); and **2)** transcriptomics and proteomics data implementing alternative algorithms (*e.g.*, SPOT, PROM, iMAT, GIMMEp) such as those discussed in Chapter 2 (**Section 2.10**). Further, the ability to genetically modify *Methanothermobacter* opens the possibility to adopt the following strain engineering approaches to predict *in-sicilo* gene deletions and

insertions: **1)** the OptGene (Patil *et al.*, 2005), OptKnock (Burgard *et al.*, 2003), and OptForce (Ranganathan *et al.*, 2010) algorithms housed in the COBRA Toolbox (Heirendt *et al.*, 2019); **2)** the (stand-alone) OptCouple algorithm (Jensen *et al.*, 2019); and **3)** the OptFlux tool (Rocha *et al.*, 2010).

7.2.2 GEMs to improve power-to-gas and power-to-x

This next section will broadly discuss some challenges in manipulating the metabolism of *Methanothermobacter* spp. and how the previously mentioned *in-silico* strain engineering approaches can be used to improve power-to-gas (methane production) and power-to-x (alternative product production). It is important to note that it may be difficult to increase methane production. Even with wild-type *Methanothermobacter* spp. it has already been possible to obtain upward of 95% substrate to methane conversion (accounting *via* carbon conversion) (Buan, 2018; Martin *et al.*, 2013). A decrease in carbon and energy diverted to biomass would be required and may be able to be addressed using gene knockouts or media supplementation. While the high methane production rate is advantageous when methane is the target product, it is problematic to target other metabolites (native or not). ATP production is coupled with methane production *via* the sodium ion gradient built during methanogenesis that powers the ATP synthase. This signifies that it will be difficult to divert substrate carbon away from methanogenesis without affecting ATP production and thus, the overall health and growth of the microbe. If carbon could be diverted towards biomass, the sodium ion gradient would still be more strained. The sodium ion gradient is also consumed to power the energy-converting hydrogenases that supply reduced ferredoxin for anaplerotic functions (*e.g.*, for biomass synthesis). Further, biomass synthesis branches off from methanogenesis at the same point at which the sodium ion gradient is built (at the methyl-tetrahydromethanopterin/Coenzyme M methyltransferase). Therefore, the majority of all other metabolic functions of the microbe are after this point. Especially with most carbon currently flowing into methane, it will be difficult to reach large fluxes or production rates for metabolites in other pathways. Thus, to be able to generate a significant amount of a product besides methane: **1)** substrate carbon needs to be diverted from methane and likely first into acetyl-CoA (from methyl-tetrahydromethanopterin); and **2)** the sodium ion gradient needs to be regenerated to allow for ATP production to continue. *In-silico* strain engineering tools

may be able to help identify solutions to these points.

The OptGene, OptKnock, and OptForce all permit the user to optimize for the (over)production of a chosen metabolite using GEMs. In the case of power-to-gas, methane would be selected as the target metabolite. OptGene and OptKnock focus on potential gene or reaction knockouts that would lead to an increased methane production while still maintaining biomass growth. The OptFlux software aims to reach similar conclusions. Alternatively, OptForce also considers up- and downregulations of genes by up- and downregulating reaction fluxes. The linear metabolism of *Methanothermobacter* spp. may limit potential genetic modifications, particularly knockouts, that can be applied to maximize a chosen product. Thus, the consideration of up- and downregulation is attractive alternative. Further, OptCouple not only accounts for gene knockouts but also medium supplementation and gene insertions. Medium supplementation, for an example amino acid addition, has been applied in various studies already (Valgepea *et al.*, 2017a; Jensen *et al.*, 2019). Shadow price analyses can be used to help determine what supplements may be beneficial for producing a target. Nevertheless, the additional cost of any supplement is an economic consideration, particularly for scale-up. However, the likelihood of overproducing a native product during power-to-gas with *Methanothermobacter* spp. is unlikely and thus gene insertions to broaden metabolic functionality become attractive.

When a nonnative product is of interest, metabolic functions or heterologous reactions need to be added to extend a native pathway or append a nonnative pathway. Typically, a target compound is first selected. The selection could be based on economic relevance and include products or precursors for other industrial processes. Then, *in-silico* tools, for example, OptStrain (Pharkya *et al.*, 2004) or the pathway search platform by Chatsurachai *et al.* (2012), can be used to design pathways to produce that target. With the GEM as a backbone, the different tools apply varying constraints, such as pathway length and thermodynamic feasibility, to design pathways that aim to produce the target compound (Saa *et al.*, 2019). After *in-silico* reaction and gene insertions to the GEM, previously discussed modeling methods (*e.g.*, FBA, FVA) can be exploited to trial the effect of the modifications on the microbe. The manipulation of a metabolism is not often simple and is more challenging for *Methanothermobacter* spp. given the previously mentioned (energetic) constraints.

7.2.3 Confirmation of omics hypotheses through biochemistry and fluorescence

In Chapter 3, we proposed that the three *Methanothermobacter* species utilized different enzymes to produce anabolic formate for biomass. Confirmation of these results would require more biochemical analyses. These may include enzyme assays to check the *in-vitro* functionality of the different formate producing enzymes and to measure kinetic parameters (*e.g.*, V_{max} and K_m). In addition to kinetic parameters, it could be useful to run absolute quantification targeted proteomics to find the absolute abundance (rather than relative abundance that was used in this dissertation) of the formate producing enzymes. The targeted proteomics methods, such as isotopic-labeled absolute quantification (AQUA) or protein standard absolute quantification (PSAQ) for peptide or proteins targets, respectively, are more expensive but could provide an opportunity to better elucidate the biodynamics occurring in the different species (or under different growth conditions, for example, Chapter 4 for growth on formate) (Lindemann *et al.*, 2017).

In Chapter 4, we predicted that the formate dehydrogenase cassette lowers the non-growth associated maintenance energy costs for *Methanothermobacter* and that non-growth associated maintenance energy decreases further for growth on sodium formate. Confirmation of this hypothesis would be required in the wet-lab, which would be possible as both the non-growth associated maintenance energy and the growth associated maintenance energy can be experimentally measured through steady-state cultivations at different growth rates (Thiele and Palsson, 2010). Further, it could be interesting to see the localization of the formate dehydrogenase cassette that was added *via* genetic modifications to *M. thermautotrophicus* ΔH and whether it shifts position under different growth conditions (*e.g.*, specifically molecular hydrogen and carbon dioxide *vs.* sodium formate). Fluorescence reporter systems are often implemented for protein localization (Feilmeier *et al.*, 2000). The ability to add fluorescent proteins (FPs) requires a genetic system, which was recently developed and employed to create the strain cultivated in Chapter 4. However, most reporter systems, such as MCherry and the green fluorescent protein (GFP), are not suitable for anaerobic applications (Streett *et al.*, 2019). In recent years, more efforts have been made to develop FPs that can be cultivated anaerobically (some of the studies using *C. ljungdahlii*) such as fluorescence activating and absorption-shifting tag (FAST) (Streett *et al.*, 2019),

HaloTag, and SNAP-tag (Charubin *et al.*, 2020; Ozbakir *et al.*, 2019). Still, unlike *C. ljungdahlii*, the *Methanothermobacter* species are also thermophilic, which means they would require a thermostable FP. Some thermostable and oxygen-independent FPs are the flavin-binding FPs that are derived from light-oxygen-voltage (LOV) domains (Nazarenko *et al.*, 2019; Wingen *et al.*, 2017). However, these do not fluoresce as strongly and are often overpowered by the native fluorescence at 480 nm (from the cofactor F_{420} that absorbs at 420 nm). In the preprint by Hernandez and Costa (2022), they successfully adopted both FAST and splitFAST (which fluoresces only when two proteins interact) for two different substrates, which were molecular hydrogen and formate. They utilized the splitFAST on one of formate dehydrogenases in *Methanococcus maripaludis* and employed FAST for protein localization for two proteins: **1)** the FruA (one subunit of the F_{420}-reducing hydrogenase); and **2)** the FlaI (the ATPase of the archaeal flagellum) (Hernandez and Costa, 2022). Thus, there still remains the prospect to investigate the localization of formate dehydrogenases, and in particular the genetically modified formate dehydrogenase cassette with the constitutive promoter and on differing substrates. Lastly, *M. maripaludis* is mesophilic, thus, as far as the authors know, a thermostable anaerobic fluorescence reporter system still has not been used in thermophilic methanogenic archaea.

7.2.4 Improving transcriptomics and proteomics analyses

In both Chapters 3 and 4, differential expression analyses for both transcriptomics and proteomics were conducted. Analyses were performed between the three *Methanothermobacter* species, and with the modified *M. thermautotrophicus* ΔH. Further comparisons were performed under varying substrate conditions (molecular hydrogen and carbon dioxide as well as sodium formate). Nevertheless, given time constraints, the results of this differential expression analysis were not extensively analyzed manually and could benefit from further examination. In addition to more manual curation, machine learning methods, such as support vector machines, could be employed to help determine key genes between microbes or between growth conditions (Piles *et al.*, 2019). However, the current data set (four replicates per condition and microbe) is small for machine learning techniques.

In both Chapters 3 and 4, a poor correlation between the transcriptome and proteome was observed. This could indicate the importance and influence of post-transcriptional

regulation or post-translational modifications. Bathke *et al.* (2019) investigated small RNA (sRNA) to better understand potential post-transcriptional regulation. The authors found differences in the transcriptome and proteome at different stages of (batch) growth as a result of changes in the environment conditions (nutrient pools decreases and product pools increases). However, the omics samples in this dissertation were taken during steady state, thus, changing environmental conditions should not be a considerable factor. Nevertheless, the differential expression analysis between microbes may help highlight regulational and/or post-translational differences between the species under the same growth conditions. In a recent review, Chung *et al.* (2021) highlighted algorithms and methods to "integrate transcriptional regulatory networks (discussed in **Section 7.2.1**), post-translational modifications, epigenetics, protein–protein interactions and protein stability, allostery, and signaling networks" into GEMs. As noted before, most of these algorithms were designed, verified, and validated using *E. coli* and *S. cerevisiae* and may not be as conclusive with the lesser-studied methanogens.

7.2.5 Confirmation and expansion of *in-silico* acetogenic analyses

In Chapter 5, the genome-scale algorithm predicted *in-silico* the large potential of the base-editing tool. Two editing windows were considered: **1)** a more liberal one, between positions -11 to -19 (bases upstream of the protospacer adjustment motif, PAM); and **2)** a more conservative one, the hot-spot, between positions -16 and -19. The editing windows were determined based on a limited number of genes (5) and 11 gRNAs. While the efficiency and the reach of the tool were concluded to be promising, follow-up (CRISPR) experiments would further support that conclusion.

In Chapter 6, an analysis of the *rseC* genes that were found in 47 acetogenic bacteria was conducted. Bengelsdorf *et al.* (2018) published a list of 60 acetogens. Of these 60, 47 had publicly available complete genome sequences and were selected for the *in-silico* analysis. Nevertheless, a more extensive bioinformatics search could be conducted. Acetogens are understood to use the reductive acetyl-CoA or Wood-Ljungdahl-Pathway (WLP) for both biomass production (*via* acetyl-CoA) and energy conservation (Ragsdale and Pierce, 2008). A more extensive search through all available genomes for this signature acetogenic pathway, could result in more species considered. Multiple sequence alignment of the WLP proteins could be conducted in

MUSCLE (Edgar, 2004) (rather than in BLASTp) and then phylogenetic trees, as described in Esquivel-Elizondo *et al.* (2021), could be built. A phylogenetic study of the vicinity of the *rseC* to the RNA-gene cluster could support (or disprove) synteny for these genomic regions and could further help elucidate the regulatory role of the *rseC* as well as the regulatory mechanisms in acetogens that can autotrophically grow without an *rseC*.

7.2.6 Improvement of experimental bioreactor setup

As the carbon balances and the gross measurement error analyses indicated in Chapters 3 and 4, there was significant (non-random) error in the bioreactor system setup. For gas fermentations with molecular hydrogen and carbon dioxide, these errors could be overcome by estimating the soluble carbon species *via* extrapolated pKa values (Prieto and Millero, 2002) (**Section 3.5.12**). However, this method was not sufficient to adjust for the error found for the fermentations with formate. As sodium formate was consumed, NaOH as produced (Costa *et al.*, 2013a). This decreased the stability of the pH in the bioreactors. Further, the high concentration of sodium formate (as a substrate) and HCl (as a pH control) increased the salt concentrations in the bioreactors. The variable pH in conjunction with the high salinities made it more difficult to estimate the pKas using the aforementioned method by Prieto and Millero (2002). Further, the empirical data employed to estimate the pKas, was typically taken from natural bodies of water (*e.g.*, oceans or lakes). Unfortunately, our high cultivation temperatures (65°C) were not in these ranges. In the future, salinity (potentially *via* conductivity) and the dissolved carbon dioxide could be empirically measured in the experimental setup.

Additionally, the micro GC, operated for on-line gas measurements, experienced drift over time (for which Agilent recommends a monthly re-calibration). To monitor drift and correct uptake rates, the substrate gas composition was measured through the course of the experiments. However, the substrate gas composition did not contain methane, eliminating its ability to be applied as a calibration gas (mixture). The first improvement would be to implement a (on-line) calibration gas mixture that contains all probable detectable gases. The second improvement would be to adopt an inert gas (*e.g.*, argon) as an internal standard within the calibration gas mixture as used in Valgepea *et al.* (2017b, 2018). Nevertheless, these solutions also present some challenges. The micro GC utilizes both helium and argon as carrier gases, thus, using argon (though relatively abundant

and inexpensive) as an internal standard is not possible. Nitrogen is an inexpensive inert gas; however, **1)** it was used as a gaseous substrate under sodium formate cultivation conditions; **2)** could be assimilated by some microbes (including *Methanothermobacter* spp.) as a nitrogen source; and **3)** is abundant in the air, some of which inevitably enters the bioreactors ($\sim$ 1%). These factors make nitrogen an unsuitable gas for an internal standard. Further, both improvements require premixed gas tanks, which under some conditions may require explosive gases (*e.g.*, molecular hydrogen) and/or expensive or rare gases (an inert gas that is not argon, helium, nor nitrogen). The ability to store large quantities of gas is not trivial and was not possible given the facilities for the experiments conducted in Chapters 3 and 4.

Lastly, the small volumes ($\sim$130 mL, bioreactor volume of 200 mL) of the bioreactors operated in Chapters 3 and 4 require less (continuous) media for operation, which saves both time and money. Nevertheless, this means when sampling the bioreactors during a fermentation, the sample volumes need to be minimized as not to disturb the steady state. Further, the small overall volume yields less biomass, which can provide difficulties when sampling for the omics and for calculating the correlation between OD_{600} and g_{CDW}. More replicates with larger volumes would likely provide less variability than was experienced with the current method; however this type of setup is frequently infeasible.

7.3 Implications

The capitalization of C1-fermenting microbes as biocatalysts for power-to-gas and power-to-x applications is promoted by this dissertation. The work with methanogens including: **1)** the manually curated GEMs; **2)** the continuous autotrophic bioreactor fermentation data sets; and **3)** the quadruplicate transcriptomics and proteomics data sampled during steady state, provide the foundation that when coupled with the genetic system, can expand the product spectrum of *Methanothermobacter* spp. from power-to-gas to power-to-x *via* metabolic engineering. Furthermore, the *in-silico* tools developed for *C. ljungdahlii* continue to broaden our understanding of its energy metabolism, which is required for its optimization and strain engineering for bioethanol production. In summary, both microbes serve to: **1)** mitigate greenhouse gas emission by utilization of carbon waste streams as substrates; **2)** store excess renewable power *via* the consumption of molecular hydrogen; and **3)** provide alternative energy supplies

that can be integrated smoothly into the current energy infrastructure.

Appendix A

Supplementary information for Chapter 3:

Supplementary text and figures are included in this chapter, the supplementary tables can be accessed at this link: https://github.com/isacasini/Dissertation_Casini/blob/main/Supplementary_table_Ch3_4.xlsx, and the rest of the supplementary data can be accessed at the following link: https://github.com/isacasini/Dissertation_Casini/tree/main/Chapter_3

- Supplementary text A.1.1 to A.1.4
- Supplementary figures A.1 to A.4
- Supplementary tables S1 to S14 (Excel® workbook)
- Data S1: Genome-scale metabolic modeling files
  - ☐ File 1: GEM Excel® workbook
  - ☐ File 2: Biomass equation calculation Excel®
  - ☐ Files 3-5: .xml model files for DH, ZZ, MM (CobraPy/Python work)
  - ☐ Files 6-8: .json model files for DH, ZZ, MM (Escher work)
  - ☐ Files 9-11: .mat model files for DH, ZZ, MM (COBRA Toolbox work)
  - ☐ Files 12-14: .html MEMOTE report files for DH, ZZ, MM
  - ☐ Files 15-16: .json Escher map files for DH, ZZ, MM
  - ☐ Files 17-19: .xml model files for DH, ZZ, MM BOUND with fermentation period 2 data (COBRA Toolbox work)
- Data S2: Excel® workbook with FBA simulations with GEMs

- ☐ Sheet 1: Constraint definitions for sheets 2-4
- ☐ Sheets 2-4: FBA results of constraining the GEMs on test conditions for DH, ZZ, MM
- ☐ Sheets 5-7: FBA results of constraining GEMs on fermentation data (gross measurement adjusted) normal/Gimme Transcriptomics/Gimme Proteomics, for DH, ZZ, MM

- Data S3: Relevant scripts
  - ☐ Script 1: Gross measurement error analysis .ipynb
  - ☐ Script 2: Modeling 1 (CobraPy) used for model validation (tests the different conditions) .ipynb
  - ☐ Script 3: Modeling 2 (CobraPy) used for model verification (constrains with fermentation data -adjusted with gross measurement error) .ipynb
  - ☐ Script 4: Transcriptomics differential expression notebook in R .ipynb
  - ☐ Script 5: Pan-proteome maker (with iBAQ) .ipynb
  - ☐ Script 6: MATLAB file to run FBAs through GIMME (COBRA Toolbox) .mat

- Data S4: Excel® workbook with fermentation data
  - ☐ Sheets 1-12: Time series data of OD_{600}, g_{CDW}, H_2, CO_2, and CH_4 production for all reactors
  - ☐ Sheets 13-15: Summary time series data of OD_{600}, g_{CDW} and dilution rates of all reactors
  - ☐ Sheets 16-23: Averaged steady-state time points for H_2 and CO_2 uptake rates, and CH_4 production rates in mmol/h and mmol/g_{CDW}/h or mmol-C/L/h
  - ☐ Sheets 24: Media salinity calculation
  - ☐ Sheets 25: Dissolved carbon species calculation
  - ☐ Sheet 26: Carbon balances: **1)** before adjustment any adjustments; **2)** after accounting for dissolved carbon species; and **3)** after dropping time points with estimated gross measurement error

- ☐ Sheet 27: Steady-state data after gross measurement analysis (constraints for the GEMs)

- Data S5: Excel® workbook with clean omics data
  - ☐ Sheets 1-3: Transcriptomics individual DH, ZZ, MM
  - ☐ Sheets 4-6: Transcriptomics differential expression MM *vs.* DH, ZZ *vs.* DH, ZZ *vs.* MM
  - ☐ Sheets 7-9: Proteomics individual DH, ZZ, MM
  - ☐ Sheets 10-12: Proteomics differential expression MM *vs.* DH, ZZ *vs.* DH, ZZ *vs.* MM
  - ☐ Sheet 13: Relative abundances of Wolfe Cycle transcripts (TPM)
  - ☐ Sheet 14: Relative abundances of Wolfe Cycle proteins (iBAQ)
  - ☐ Sheet 15: Relative abundances of Eha/Ehb transcripts (TPM)/proteins (iBAQ)
  - ☐ Sheet 16: Pan-genome

A.1 Supplementary text

A.1.1 Methylation patterns of the microbes

The methylation patterns of the three microbes were elucidated by single-molecule real-time long-read sequencing (**Materials and methods, 3.5.2**). The microbes exhibited N^6-methyladenine (m^6A) and N^4-methylcytosine (m^4C) methylation in three distinct methylation patterns (**Table S5**). The more closely related *M. thermautotrophicus* ΔH and *M. thermautotrophicus* Z-245 shared one sequence motif (with a m^6A modification) that exhibited features of a Type I restriction/modification system (Murray, 2000) (**Table S5**). The two m^6A methylation patterns from *M. marburgensis* Marburg was more distinct but also characteristic for a Type I restriction modification system (**Table S5**). The genomes of all three microbes encode putative components of a Type I restriction/modification system (**Table S5**), which could be responsible for the methylation of the DNA (Loenen *et al.*, 2014; Roberts *et al.*, 2015). *M. marburgensis* Marburg had an additional m^4C modification, which resembles a

Type-II modification, with putative methyl-transferase genes encoded in the genome (**Table S5**). *M. thermautotrophicus* Z-245 had an additional motif with a m^4C modification for which the responsible Type II restriction/modification methyltransferase is located on the plasmid (pFZ1) (Nölling *et al.*, 1992), assigning a known function to the plasmid. While *M. marburgensis* Marburg also harbors a plasmid (pME2001), its function is still unknown but does not obvious restriction/modification systems.

A.1.2 Multi-level omics under steady-state growth conditions indicate elevated oxidative stress response

We did not find obvious reasons for the different growth behaviors when analyzing methanogenesis gene products. In the multi-level omics analysis, we had found several genes and gene products that are potentially involved in oxidative stress response and were highest on average in *M. marburgensis* Marburg compared to the other two microbes. These included, in the transcriptomics (though not on the proteomics level), the F_{390} synthase (gene group 1317) that protects cofactor F_{420} (and halts methanogenesis) by converting it into the oxygen-protected form, cofactor F_{390} (Vermeij *et al.*, 1996), and in the proteomics (though not on the transcriptomics level), three enzymes (gene groups 128, 186, and 1155) that are annotated as F_{420} H_2-dependent oxidases that convert dioxygen into water. While we measured dioxygen in the range of $0.069 \pm 0.085\%$ v/v in the exhaust gas of all bioreactors, indeed, we had found significantly higher dioxygen levels of 0.315 ± 0.098 (P-value 7.38E-77) in the exhaust gas of one of the bioreactors with *M. marburgensis* Marburg (**Figure A.3**). Thus, our results confirmed our observation of increased dioxygen levels in an increased level of oxidative stress in the averaged transcriptome and proteome in *M. marburgensis* Marburg.

A.1.3 Conversion of methenyl-H_4MPT to methylene-H_4MPT may not be able to be accurately predicted in the GEM despite integration of multi-omics data

In all three microbes, we found high abundance of the enzymes F_{420}-dependent methylene-H_4MPT dehydrogenase (Mtd), H_2-forming methylene-H_4MPT

dehydrogenase (Hmd), and F_{420}-reducing Ni-Fe hydrogenase (Frh), which are required to catalyze the conversion of methenyl-H_4MPT to methylene-H_4MPT (**Figure 3.4**). Frh generates reduced cofactor F_{420}, and thus, provides reducing equivalents for Mtd under standard conditions. During nickel limitation, the conversion of methenyl-H_4MPT to methylene-H_4MPT is catalyzed by Mtd together with Hmd instead, also known as the Hmd-Mtd cycle (Hendrickson and Leigh, 2008). Under these conditions, in addition, the complex of Mtd and Hmd can replace Frh to provide reduced cofactor F_{420} by catalyzing the reaction in reverse (Kaster *et al.*, 2011b; Hendrickson and Leigh, 2008). While the Frh was found to be experimentally reversible in vitro (Kaster *et al.*, 2011b), we assumed irreversibility for our modeling in the direction of reduced cofactor F_{420} production, because molecular hydrogen was present in access in our experiments. The presence of all three proteins (including Hmd) in the transcriptomics and proteomics data indicates that there might be a slight nickel-limitation, potentially occurring after high biomass concentrations were reached (**Data S2**), which resulted in the partial bypassing of Frh in our modeling results. Alternatively, the presence of Hmd could be an example of the evolutionary tradeoff between rapid response time to changing environmental conditions (*e.g.*, limited nickel) and lowering de-novo protein synthesis costs (Wessely *et al.*, 2011). However, despite constraining the GEMs with the omics data, we did not find a solution that uses all three reactions simultaneously (unless Frh was set to work reversibly, which generated a loop with the other two reactions). The flux balance analyses predicted the use of Hmd with twice the flux and Mtd working in reverse for all simulations, except for the reduced *M. thermautotrophicus* Z-245 GEM for which Mtd and Frh are both active in the forward direction (**Data S2**).

A.1.4 Predicted pyruvate formate-lyase-activating enzymes are present and highly upregulated in *M. thermautotrophicus* Z-245 and *M. marburgensis* Marburg despite the missing pyruvate formate-lyase main subunit.

We only identified pyruvate formate-lyase in *M. thermautotrophicus* ΔH. An additional activating radical SAM enzyme, pyruvate formate-lyase 2 activating enzyme (gene group 279) that is present in *M. thermautotrophicus* ΔH (MTH345/ISG35_1595) (Sawers and Watson, 1998; Kaster *et al.*, 2011b), also had low confidence BLASTp+

hits for *M. thermautotrophicus* Z-245 and *M. marburgensis* Marburg. In addition, *M. thermautotrophicus* ΔH had low confidence annotations for two additional Pfl-activating enzymes (gene groups 917 and 1372), which also contain an annotated radical SAM core domain (Uniprot, MTH1069 and MTH1586) (Consortium, 2019). In turn, these two putative Pfl-activating enzymes had low confidence BLASTp+ hits for *M. thermautotrophicus* Z-245 and *M. marburgensis* Marburg. The differential transcriptomics analysis showed upregulation of the gene that encodes the activating protein in gene group 279 in *M. thermautotrophicus* ΔH compared to the other two microbes (**Data S5**). However, the opposite was found for the other putative activating enzymes (gene groups 917 and 1372), which were highly upregulated in *M. thermautotrophicus* Z-245 ($\log_2$FC of 5) and *M. marburgensis* Marburg ($\log_2$FC of 6). The upregulation of these gene groups may indicate their importance to other glycyl free radical (*e.g.*, SAM) dependent enzymes whose function is still unknown (**Data S5**) (Sawers and Watson, 1998).

A.2 Supplementary figures

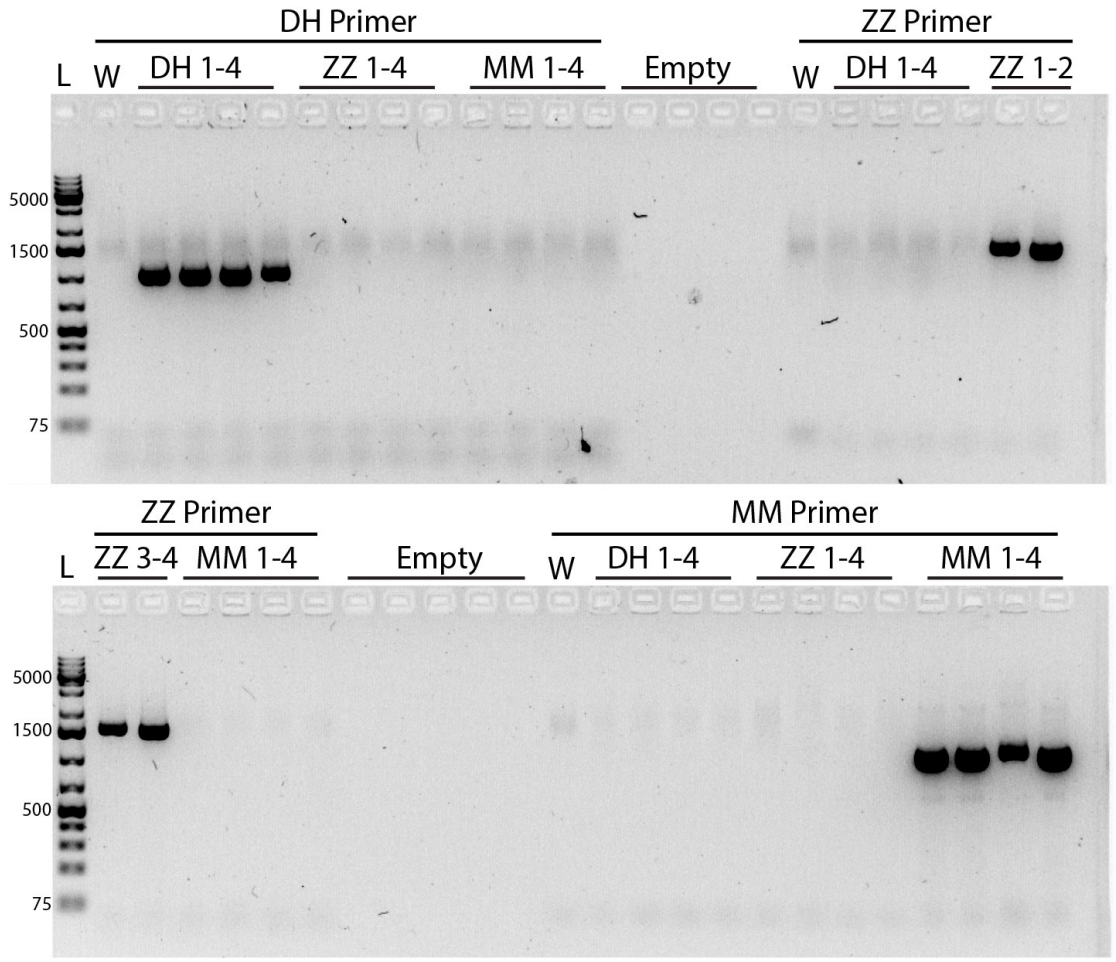

Figure A.1: Check for cross-contamination in the bioreactors during Period 2. L, ladder (1k bp); W, Millipore water; DH, *M. thermautotrophicus* ΔH; ZZ, *M. thermautotrophicus* Z-245; MM, *M. marburgensis* Marburg. Primers are listed in **Table S2**.

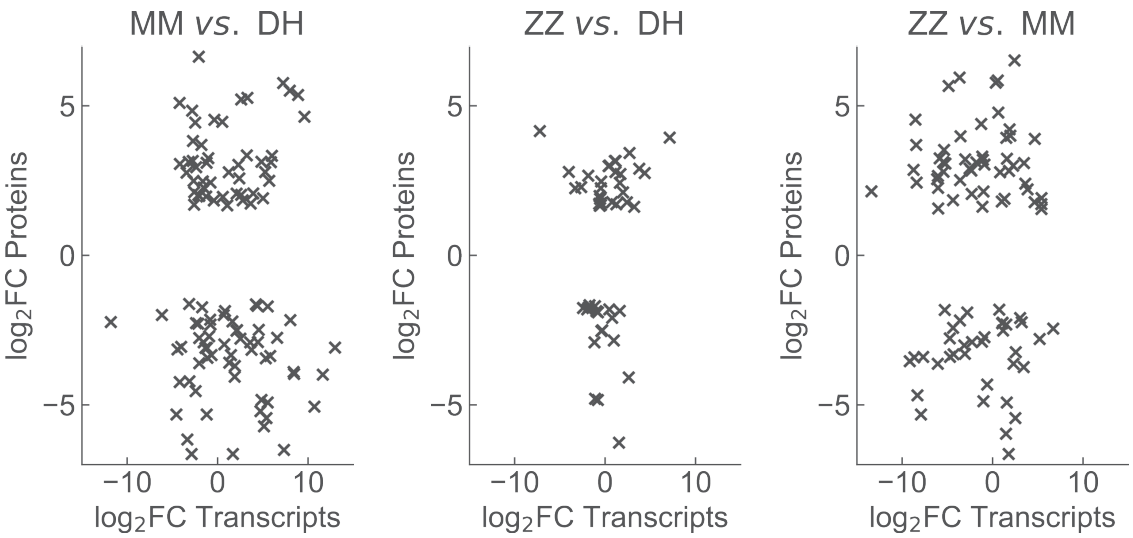

Figure A.2: Significant (adjusted P-value $\leq$ 0.05) differentially expressed proteins *vs.* differentially expressed genes (transcripts) for the 3 pairwise comparisons. DH, *M. thermautotrophicus* ΔH; ZZ, *M. thermautotrophicus* Z-245; MM, *M. marburgensis* Marburg.

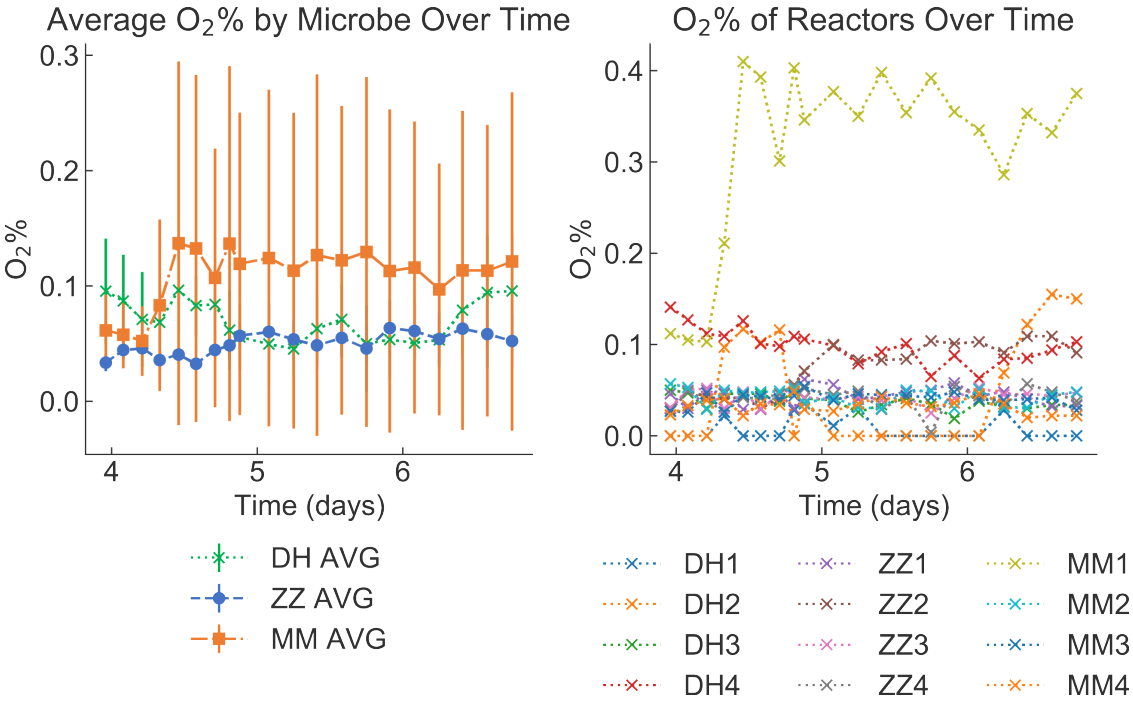

Figure A.3: Dioxygen (O_2) levels (% v/v) in the bioreactor exhaust lines during Period 2, averaged by microbe and by each bioreactor replicate. DH, *M. thermautotrophicus* ΔH; ZZ, *M. thermautotrophicus* Z-245; MM, *M. marburgensis* Marburg.

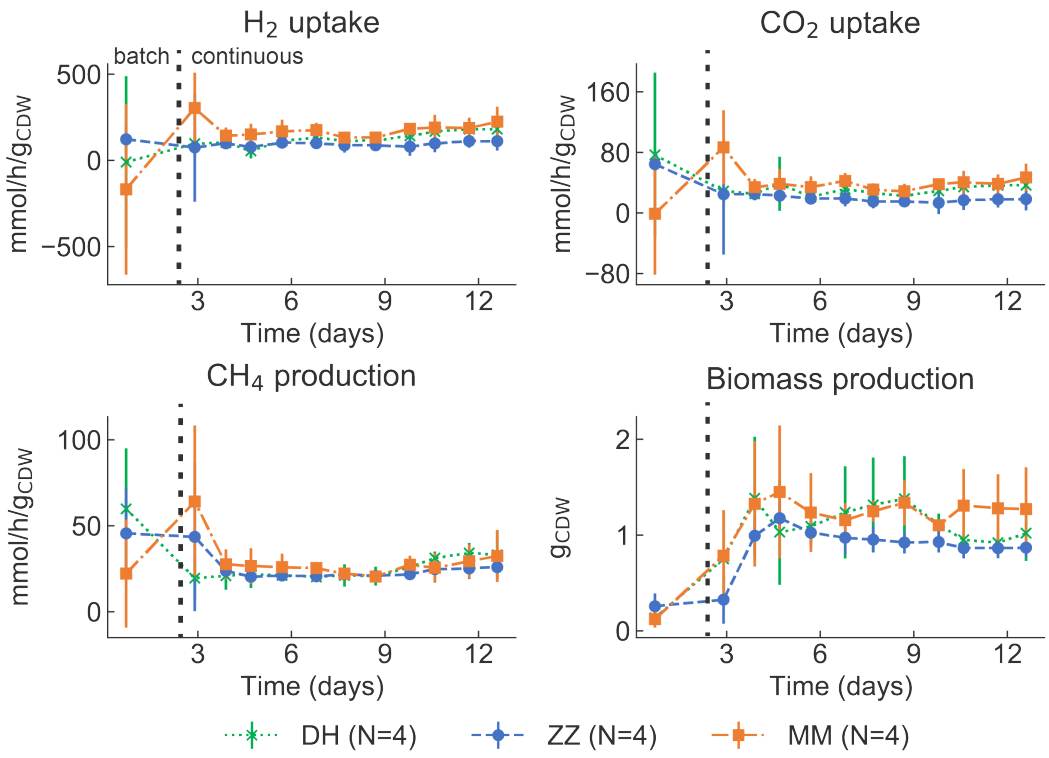

Figure A.4: Gas consumption (H_2 and CO_2 uptake), and CH_4 and biomass production data from quadruplicate bioreactors for the fermentation period of 13 days (Period 1). DH, *M. thermautotrophicus* ΔH; ZZ, *M. thermautotrophicus* Z-245; MM, *M. marburgensis* Marburg.

Appendix B

Supplementary text and figures are included in this chapter, the supplementary tables can be accessed at this link: https://github.com/isacasini/Dissertation_Casini/blob/main/Supplementary_table_Ch3_4.xlsx, and the rest of the supplementary data can be accessed at the following link: https://github.com/isacasini/Dissertation_Casini/tree/main/Chapter_4

- Supplementary text B.1.1 to B.1.3
- Supplementary figures B.1 to B.3
- Supplementary tables S15 to S21 (Excel® workbook)
- Data S1: GitHub repository with relevant scripts and files https://github.com/isacasini/Dissertation_Casini/tree/main/Chapter_4/DataS1
  - ☐ Script 1 (File 1): Modeling 2 (CobraPy) used for model verification (constrains with fermentation data), includes adjusting the DH model for genetically modified DH) .ipynb
  - ☐ Script 2 (File 2): MATLAB file to run FBAs through GIMME (COBRA Toolbox) .mat
  - ☐ Files 3-8 updated model files for the DH_H, DH_F, ZZ_F .xml and .mat
- Data S2: Excel® workbook with fermentation data
  - ☐ Sheets 1-12: Time series data of OD_{600}, g_{CDW}, H_2, CO_2, CH_4, for all reactors
  - ☐ Sheets 13-15: Summary time series data of OD_{600}, g_{CDW} and dilution rates of all reactors
  - ☐ Sheets 16-19: Averaged steady-state time points for Na-formate, H_2, CO_2, CH_4 in mmol/g_{CDW}/h

- ☐ Sheet 20: Steady-state data used to constrain the GEMs

- Data S3: Excel® workbook with clean omics data
  - ☐ Sheets 1-3: Transcriptomics individual DH_H, DH_F, ZZ_F
  - ☐ Sheets 4-5: Transcriptomics differential expression DH_H *vs.* DH_F, ZZ_F *vs.* DH_F
  - ☐ Sheets 6-8: Proteomics individual DH_H, DH_F, ZZ_F
  - ☐ Sheets 9-10: Proteomics differential expression DH_H *vs.* DH_F, ZZ_F *vs.* DH_F
  - ☐ Sheet 11: Relative abundances of Wolfe Cycle transcripts (TPM)
  - ☐ Sheet 12: Relative abundances of Wolfe Cycle proteins (iBAQ)
  - ☐ Sheet 13: Relative abundances of Eha/Ehb transcripts (TPM)/proteins (iBAQ)
  - ☐ Sheet 14: Pan-genome (DH_H/DH_F,ZZ_F)

- Data S4: Excel® workbook with FBA simulations with GEMs
  - ☐ Sheets 1-3: FBA results of constraining GEMs on fermentation data (gross measurement adjusted) normal/Gimme Transcriptomics/Gimme Proteomics, for DH_H, DH_F, ZZ_F
  - ☐ Sheets 4: Calculation of the ATP ratios using proteomics constrained FBA.

B.1 Supplementary text

B.1.1 Sparging with molecular nitrogen does not affect the nitrogen metabolism when ammonium is present

While ammonium chloride was supplied as the nitrogen source, for the fermentations on sodium formate we sparged with molecular nitrogen thus we investigated the nitrogen metabolism here. The genomes of both microbes contain the nitrogen fixation (*nif*) operon that encodes a nitrogenase NifDK, a nitrogenase reductase NifH1, a global nitrogen regulator NrpR, as well as NifEN that are involved the iron-molybdenum cofactor biosynthesis, and $glnb_i$ and $glnb_ii$ that are two regulatory binding sites (**Table S21**) (Goyal *et al.*, 2016; Lie *et al.*, 2022). Although, in both continuous bioreactor

growth conditions, ammonium chloride was the intended nitrogen source, for growth on sodium formate, molecular nitrogen (N_2) was also provided to maintain a 20 vol-% carbon dioxide (CO_2) partial pressure (**Materials and Methods**, 4.5.3), which presented a potential additional nitrogen source. Nevertheless, the nitrogen fixation (reaction NIT1b2) requires 16 mol of ATP and 4 mol of reduced ferredoxin, which each transfer two electrons per mole of molecular nitrogen, making this process energy intensive and typically less favored when ammonium is available (Poudel *et al.*, 2018).

Lie and Leigh (2003) found that NrpR acts as a repressor to the *nif* operon and to the glutamine synthase (*glnA*). Ammonia, alanine, and molecular nitrogen are signaling compounds that produce full repression, partial repression, and derepression of the *nif*, respectively, *via* 2-oxoglutarate (Lie *et al.*, 2005). 2-oxoglutarate serves as an inducer to the *nif* operon by binding to the NrpR and decreasing the binding affinity of the NrpR to the operators of the *nif* operon (nifOR) (Goyal *et al.*, 2016). When ammonia is present, 2-oxoglutarate is consumed to produce glutamate and thus, does not bind to the NrpR (Goyal *et al.*, 2016). When alanine is used as a nitrogen source, intermediate repression is observed (Lie and Leigh, 2002). As with ammonia, alanine with 2-oxoglutarate can produce glutamate and pyruvate. When molecular nitrogen is used as the nitrogen source, the microbes experience a nitrogen deficiency (Goyal *et al.*, 2016). 2-oxoglutarate is not as consumed and binds to NrpR that then hinders the binding of the NrpR to the nifOR and allows the transcription of the *nif*.

A decrease in NrpR would also indicate the derepression of the nif genes, promoting nitrogen fixation (Lie *et al.*, 2005). In the differential expression analysis for *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ on molecular hydrogen and carbon dioxide, and sodium formate conditions, the NrpR (Gene group 1357) was downregulated, which is consistent with the upregulation of the primary nitrogenase subunits H1 and D (subunit K was not detected, **Table S21**). The cultures grown on molecular hydrogen and carbon dioxide reached higher biomass concentrations, which may have resulted in a larger nitrogen consumption. Lower nitrogen levels may have triggered the downregulation of the NrpR in attempts to increase the transcription of the *nif* operon in preparation for an alternative nitrogen source (molecular nitrogen). However, to avoid nitrogen limitations, a concentration of 2.00 g/L (ammonium chloride) was used. This was determined to not be limiting based on preliminary experiments (not shown) that tested a range of 0.19 g/L (Fink *et al.*, 2021) to 2.6 g/L. Alternatively, the larger excess of ammonia in cultures grown on sodium formate, may

have promoted the transcription of NrpR for the repression of the *nif* despite having molecular nitrogen also available. Nevertheless, no significant difference between the two growth conditions was found in the proteomics, and further experiments would be required to determine if: **1)** Excess ammonia concentrations in the bioreactor vary between the growth conditions; and **2)** any molecular nitrogen is being utilized as a nitrogen source (**Table S21**).

However, we found significant differences in expression of the *nif* genes and the NrpR abundance between *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ and *M. thermautotrophicus* Z-245 for growth on sodium formate (this work) and between *M. thermautotrophicus* ΔH wild type and *M. thermautotrophicus* Z-245 molecular hydrogen and carbon dioxide (**Chapter 3**, **Table S21**). The first three subunits of the *nif* genes (Gene groups 1348-1350) were downregulated, and the other subunits (Gene groups 1351-1354 and 1357) upregulated under both growth conditions. This suggests that the additional molecular nitrogen in our conditions did not change the metabolism significantly, but there may be differences in the regulation of the nitrogen metabolism in the two microbes. Nevertheless, as before, none of these differences were identified in the proteomics (**Table S21**).

B.1.2 Predicted switch in methylene-tetrahydromethanopterin production mechanism on formate compared to on molecular hydrogen and carbon dioxide

Growth on sodium formate through formate dehydrogenase, generates reduced cofactor F_{420}. The FBAs for both *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ and *M. thermautotrophicus* Z-245 grown on sodium formate were constrained using uptake and production rates alone or additionally with transcriptomics data. Both sets of constraints predicted the formation of methylene-tetrahydromethanopterin (MyH4MPT) *via* the methenyl (Me)-H4MPT hydrogenase (model reaction HMD) and the oxidation of reduced F_{420} *via* the F_{420}-reducing hydrogenase (model reaction FRH) working in the reverse direction (**Data S3, S4**). Constraining the FBAs further with proteomics data blocked flux through the FRH, and instead, forced flux through the F_{420}-dependent MyH4MPT dehydrogenase (model reaction MTD) for MyH4MPT synthesis (in excess). After, the HMD worked in reverse to produce MeH4MPT and molecular

hydrogen. The molecular hydrogen was then used as a substrate for the F_{420}-non-reducing hydrogenase with the heterodisulfide reductase complex (model reaction MVHHDR) and the energy converting hydrogenases (model reactions Eha and Ehb) (**Figure 4.2**). During growth on molecular hydrogen and carbon dioxide, the MTD and HMD operated in the opposite direction and the FRH is once again blocked (**Figure 4.2**) consistent with the results in Chapter 3.

B.1.3 Modeling sodium formate transport in the genome-scale metabolic model

The independent transport of a (negatively) charged compound (*e.g.*, formate) across the membrane of the cell is unlikely, even with a (hypothesized) transporter (*fdhC*) given that it would alter the overall charge of the cell. It is likely that a cation (a proton, H^+, or a sodium ion, Na^+) is simultaneously transported. The formate transport was modeled this way in the genome-scale metabolic models (model reaction FORT, with a proton). This also meant that the pseudo proton exchange reaction (model reaction EX_h_e) must have allowed the transport of extracellular protons into the cell (Richards *et al.*, 2016). The sodium ion gradient between the extracellular and intracellular environments is critical to the energy metabolism of the microbes as is associated with the proton gradient. The microbes possess a Na^+/H^+ antiporter that is hypothesized to help control the pH of the cell by exchanging extra- and intra-cellular sodium ions and protons (Kaster *et al.*, 2011b). If sodium ion had been selected as the cation for the formate transport reaction, then this antiporter would have been responsible for transporting out again the sodium ion that had been brought into the cell with the formate. Simultaneously, extracellular protons, would be transported into the cell, which would have once again required the same flux through EX_h_e.

B.2 Supplementary figures

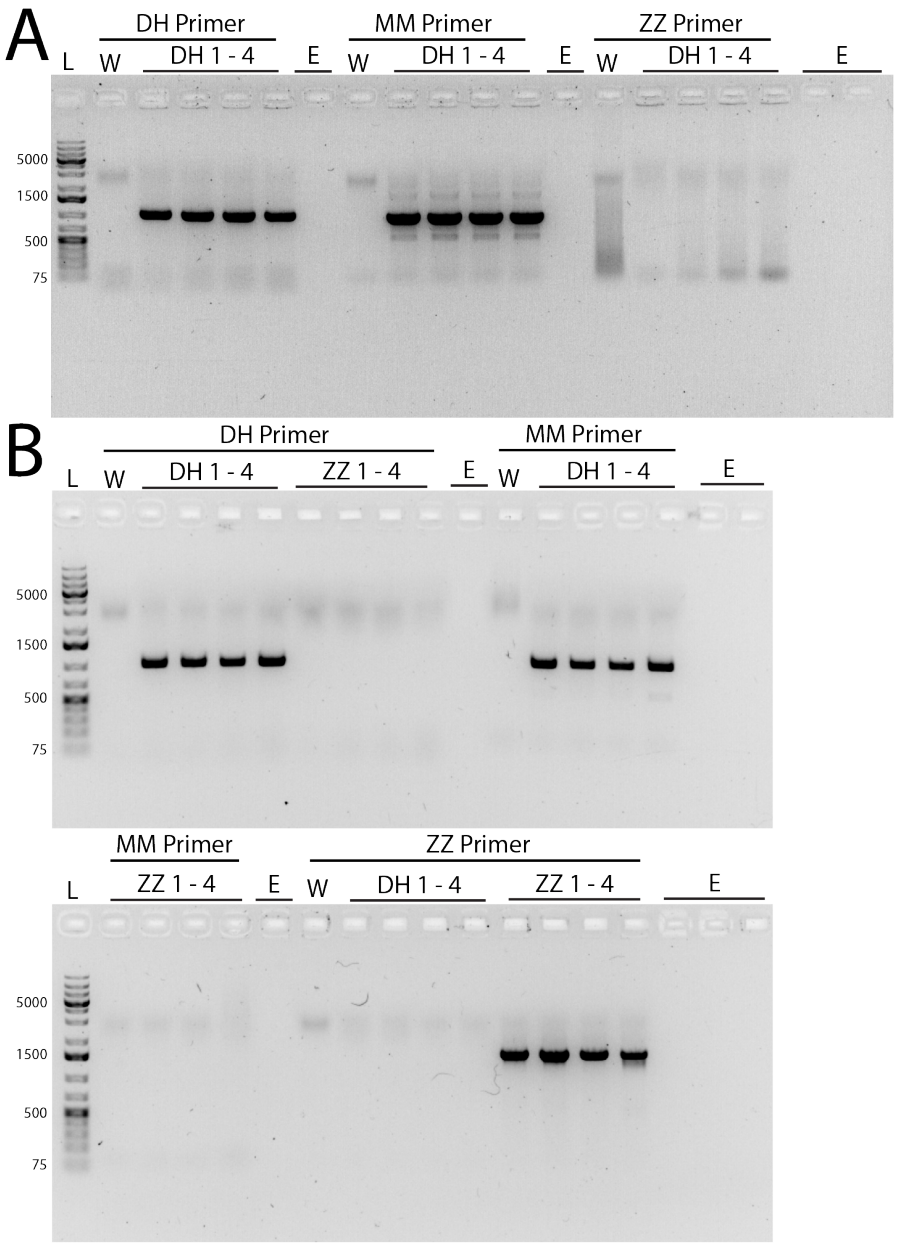

Figure B.1: (Caption on the next page.)

Figure B.1: (Previous page.) **Check for cross-contamination in the bioreactors during steady state. (A)** *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ cultivated on molecular hydrogen and carbon dioxide; **(B)** *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$ and *M. thermautotrophicus* Z-245 cultivated on sodium formate. E, empty lane; L, ladder (1k bp); W, Millipore water; DH, *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$; ZZ, *M. thermautotrophicus* Z-245; MM, *M. marburgensis* Marburg. Primers are listed in **Table S15**.

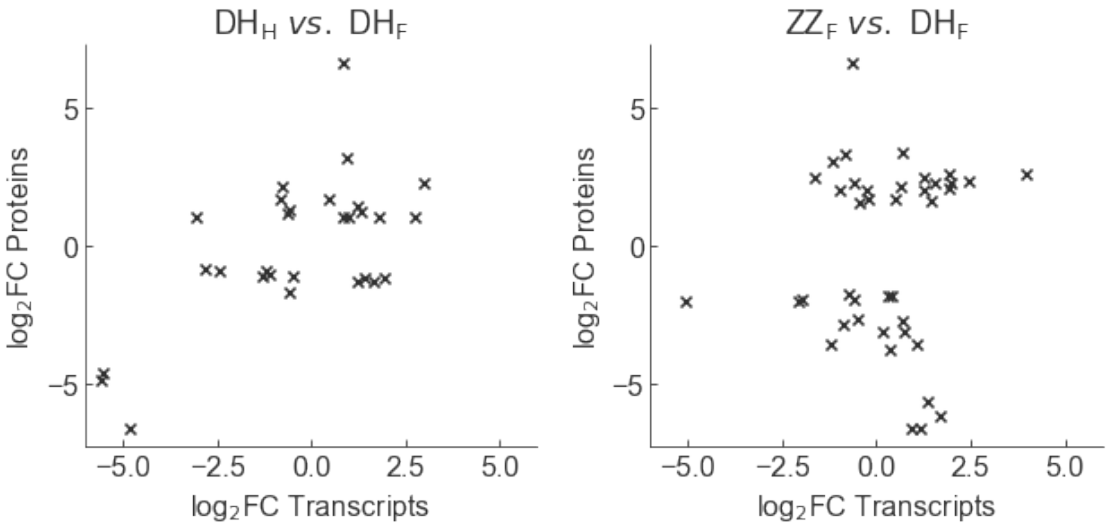

Figure B.2: Significant (adjusted P-value $\leq$ 0.05) differentially expressed proteins *vs.* differentially expressed genes (transcripts) for the 3 pairwise comparisons. DH, *M. thermautotrophicus* ΔH pMVS1111A:P_{hmtB}-$fdh_{Z\text{-}245}$; ZZ, *M. thermautotrophicus* Z-245; F, sodium formate as substrate; H, molecular hydrogen and carbon dioxide as substrate.

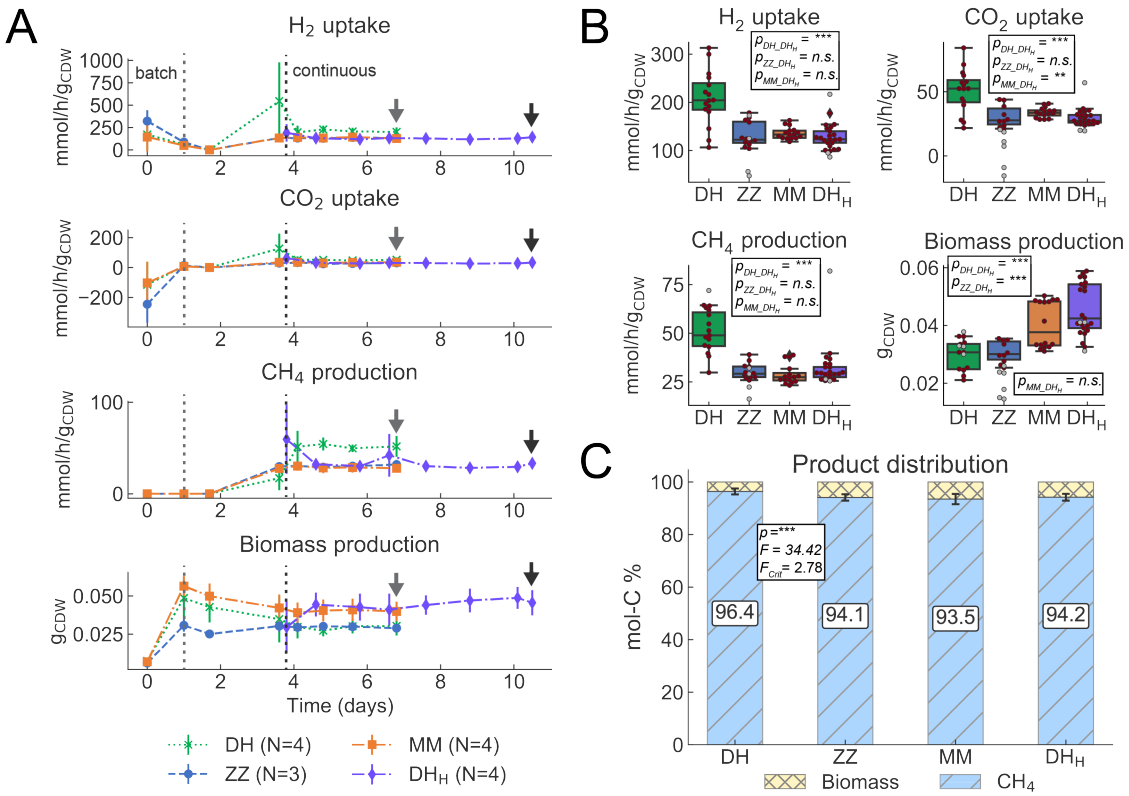

Figure B.3: Fermentation data from continuous bioreactors with *M. thermautotrophicus* ΔH pMVS1111A:P$_{hmtB}$-*fdh*$_{Z\text{-}245}$ grown on molecular hydrogen and carbon dioxide (DH$_H$), with the previous data of *M. thermautotrophicus* ΔH (DH), *M. thermautotrophicus* Z-245 (ZZ), and *M. marburgensis* Marburg (MM) also grown on molecular hydrogen and carbon dioxide from from Chapter 3. **(A)** Gas consumption (H_2 and CO_2 uptake), and CH_4 and biomass production data from quadruplicate (DH, MM, and DH$_H$) and triplicate (ZZ) bioreactors for the fermentation period of 7 (DH, ZZ, MM) or 11 (DH$_H$) days. Data for further analyses (transcriptomics, proteomics) were taken on day 7 as indicated by gray arrows (DH, ZZ, MM) and day 11 as indicated by the black arrow (DH$_H$). **(B)** Average gas consumption (H_2 and CO_2 uptake), and CH_4 and biomass production data during steady-state period (days 4 to 7 (DH, ZZ, MM) and days 5 to 11 (DH$_H$)). For statistical analysis in pair-wise comparisons with t-test, data points without suspected gross measurement error (red circles) were included, data points with suspected gross measurement error (gray circles), were excluded (**Materials and methods, 3.5.12**). **(C)** Average normalized product distribution, including statistical analysis by ANOVA (**Materials and methods, 3.5.13**. CH_4, methane; CO_2, carbon dioxide; H_2, molecular hydrogen; DH, *M. thermautotrophicus* ΔH wild type; DH$_H$, *M. thermautotrophicus* ΔH pMVS1111A:P$_{hmtB}$-*fdh*$_{Z\text{-}245}$; ZZ, *M. thermautotrophicus* Z-245; MM, *M. marburgensis* Marburg; ***, p <0.0001; **, p <0.001; n.s., not significant (p >0.05); F, F value; F$_{crit}$, F critical value.

Appendix C

Supplementary information for Chapter 5.

- Supplementary data sets 1 to 2 in Section C.1
- Supplementary figures C.1 to C.7
- Supplementary tables C.1 to C.7

C.1 Supplementary datasets

The supplementary data sets 1 (PDF) and 2 (Excel® workbook) can be downloaded at the following link: `https://pubs.acs.org/doi/10.1021/acssynbio.0c00226?goto=supporting-info` Immediate download data set 2 is also possible at: `https://ndownloader.figstatic.com/files/23907312`

C.2 Supplementary figures

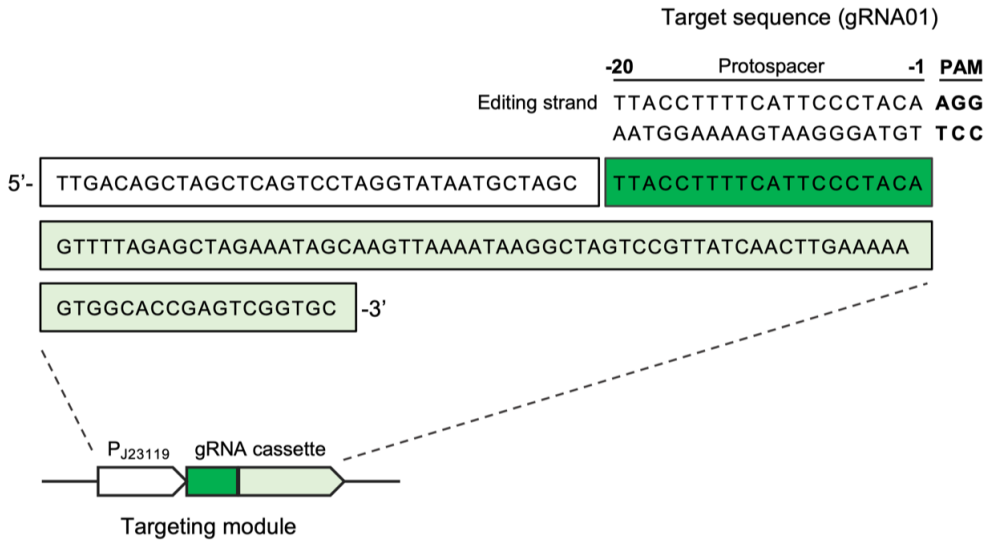

Figure C.1: Scheme and sequence of gRNA cassette. Taking gRNA01 as an example, the sequences of PJ23119 promoter (white), gRNA01 (green), and gRNA scaffold of dCas9 from *S. pyogenes* (light green) are shown. For each line, the sequence reads from left to right and does not indicate the structure. Both strands of the target sequence are displayed. The gRNA01 will bind the complimentary strand of the editing strand, and potentially edits the Cs on the editing strand within the editing window. The nucleotide directly adjacent to the protospacer adjacent motifs (PAM) is counted as position -1 and the starting position.

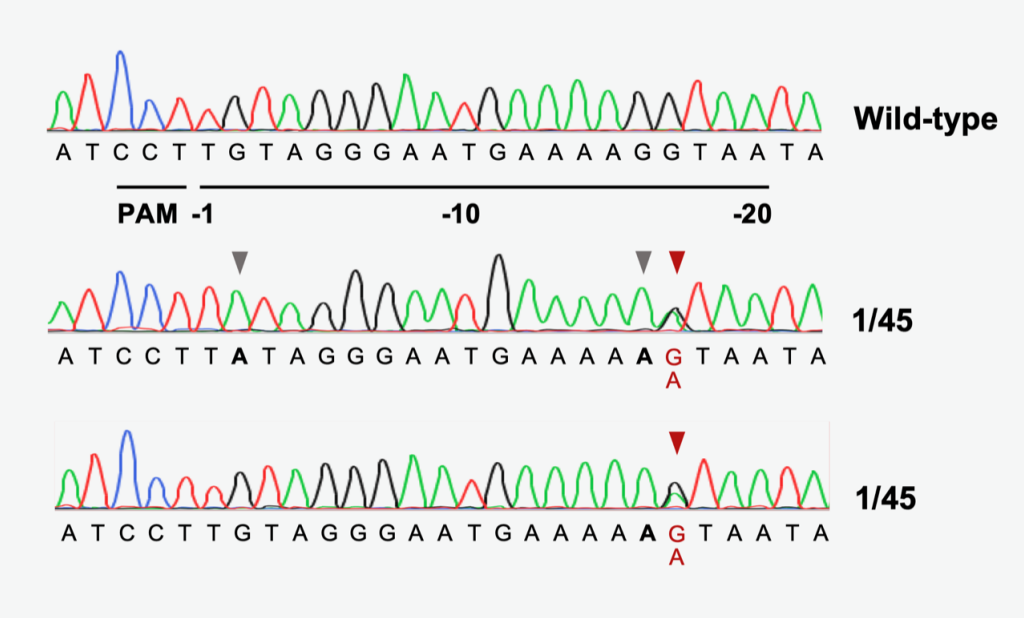

Figure C.2: Mixed signals in the two colonies when editing *pta* using gRNA01. Grey arrows indicate clean mutations, and red arrows indicate mixed signals of G and A. PAM, protospacer adjacent motifs.

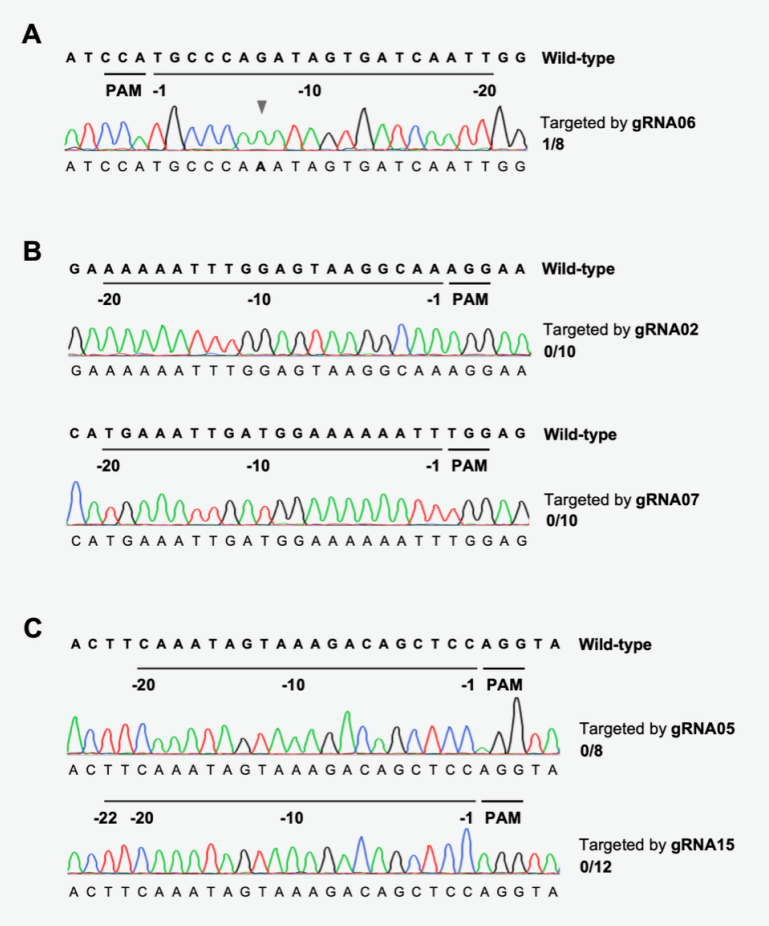

Figure C.3: Editing principles. **(A)** Base editing in *pta* with gRNA06. Grey arrow indicates successful editing. **(B)** Evaluation of base editing on target sequences without Cs using gRNA02 and gRNA07. **(C)** Base editing with different length of gRNAs (gRNA05 and gRNA15) for editing the C at position -20. PAM, protospacer adjacent motifs.

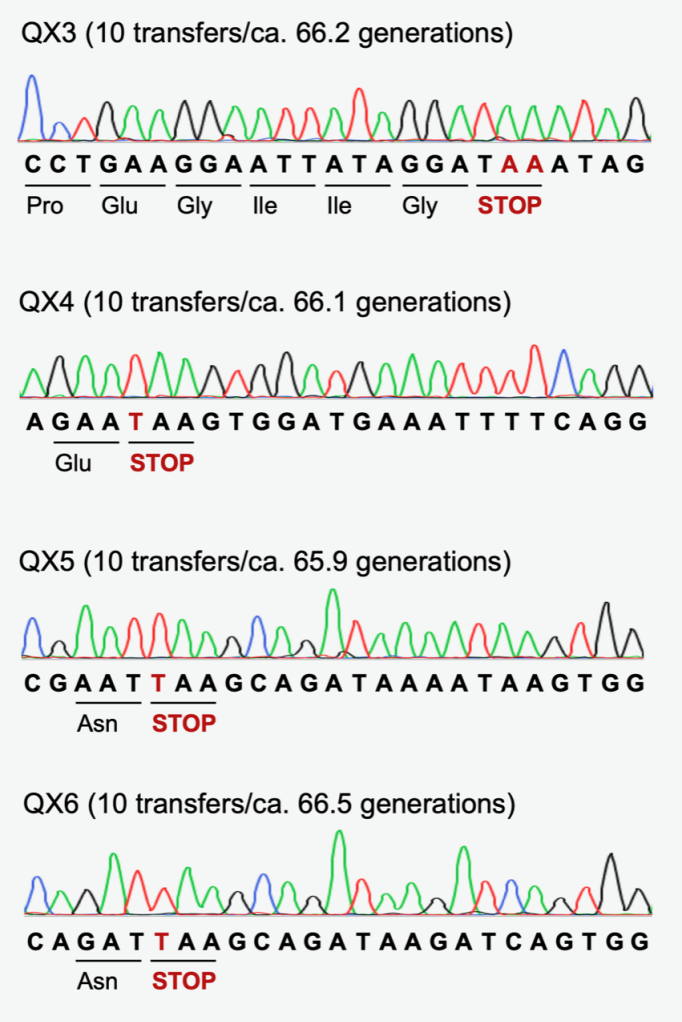

Figure C.4: Stability test of four edited strains. QX3, QX4, QX5, and QX6 were transferred 10 times in RCM and the single-nucleotide variations were sequenced to check stability. The sequencing results of the 10^{th} transfer are shown here.

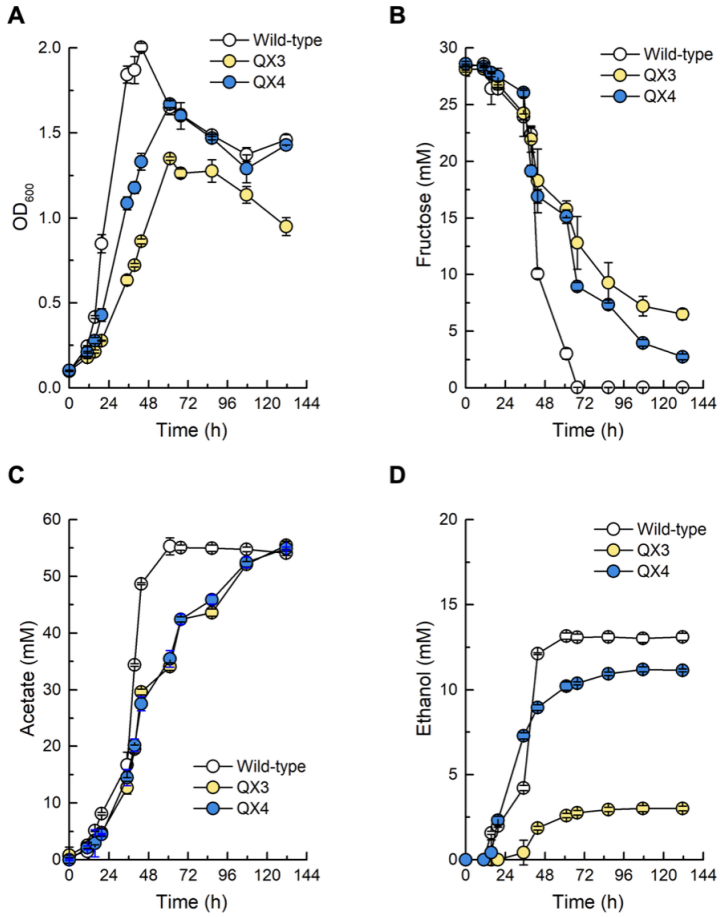

Figure C.5: Fermentation performances of wild-type, QX3 (*adhE1* Trp169*), and QX4 (*adhE2* Gln33*) under heterotrophic conditions with 5 g/L (27.8 mM) of fructose as the carbon source. **(A)** Growth, **(B)** concentration of fructose, **(C)** concentrations of acetate, and **(D)** concentration of ethanol in the culture during fermentation.

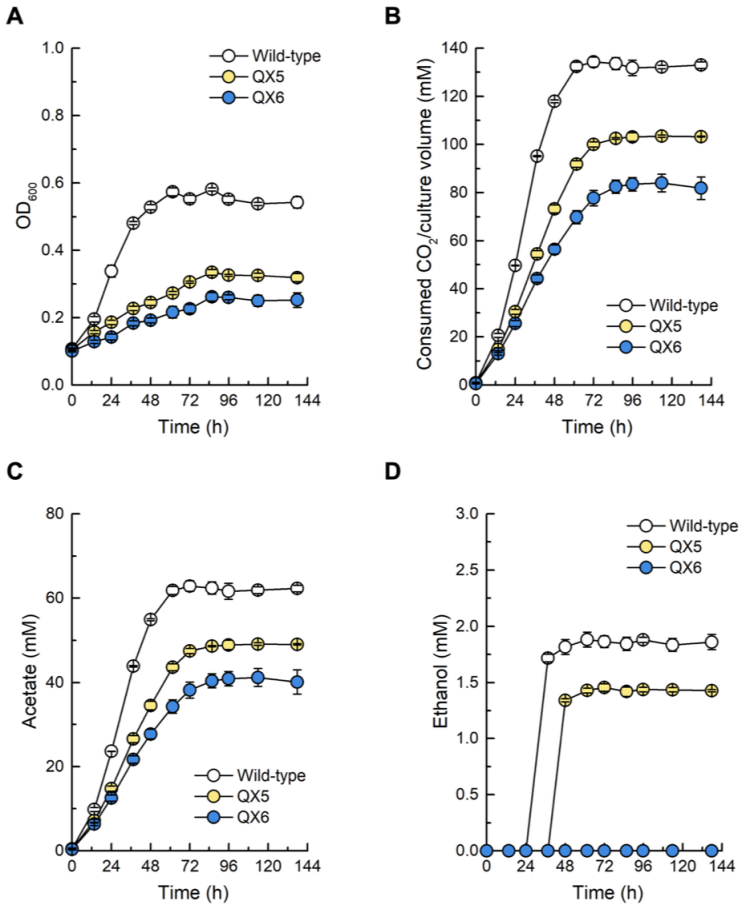

Figure C.6: Fermentation performance of wild-type, QX5 (*aor1* Gln267*), and QX6 (*aor2* Gln267*) under autotrophic conditions with a gas mixture (H_2/CO_2, 80/20 vol-%, 1.5 bar) as the substrate. **(A)** Growth, **(B)** consumed CO_2 per culture volume, **(C)** concentrations of acetate, and **(D)** concentration of ethanol in the culture during fermentation.

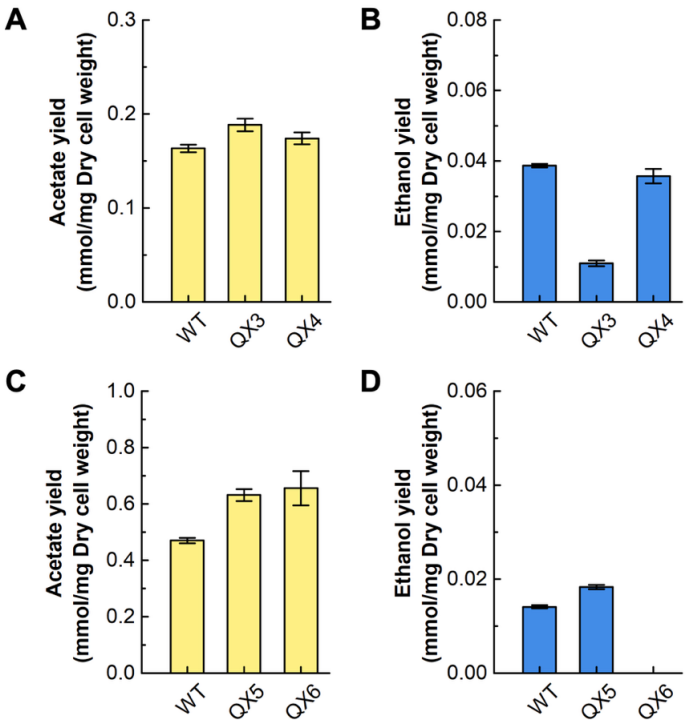

Figure C.7: Acetate and ethanol yields per dry cell weight. Acetate **(A)** and ethanol **(B)** yields of wild-type, QX3, and QX4 under heterotrophic conditions with 5 g/L (27.8 mM) of fructose as the carbon source. Acetate **(C)** and ethanol **(D)** yields of wild-type, QX5, and QX6 under autotrophic conditions with a gas mixture of H_2/CO_2 (80/20 vol-%, 1.5 bar) as the substrate. The fermentation experiments were conducted in triplicate (N=3), and the error bars indicate the standard deviations.

C.3 Supplementary tables

Table C.1 gRNA and protospacer adjacent motifs (PAM) sequences used in this study.

gRNA	Targets[a]	Strand[b]	PAM	gRNA sequence
gRNA01	*pta*	N	AGG	TTACCTTTTCATTCCCTACA
gRNA02	*pta*	C	AGG	AAAAATTTGGAGTAAGGCAA
gRNA05	*pta*	C	AGG	CAAATAGTAAAGACAGCTCC
gRNA06	*pta*	N	TGG	AATTGATCACTATCTGGGCA
gRNA07	*pta*	C	TGG	TGAAATTGATGGAAAAAATT
gRNA10	*adhE1*	N	AGG	ATCCATCCTATAATTCCTTC
gRNA11	*adhE2*	C	AGG	AACAAGTGGATGAAATTTTC
gRNA13	*aor1*	C	TGG	GGTCAGGGAATGCCAACTTA
gRNA14	*aor2*	C	TGG	GATCAAGCAGATAAGATCAG
gRNA15	*pta*	C	AGG	TTCAAATAGTAAAGACAGCTCC
gRNA19	*aor1*	C	TGG	AATCAAGCAGATAAAATAAG

[a] *pta* (CLJU_c12770), *adhE1* (CLJU_c16510), *adhE2* (CLJU_c16520), *aor1* (CLJU_c20110), and *aor2* (CLJU_c20210).
[b] N indicates non-coding strand, and C indicates coding strand.

Table C.2 Summary of base editing in pta using gRNA01.

No.	Screened colonies	Edited colonies	Efficiency	Editing site
1	*8*	8	100%	-16 (6); -16 and -17 (2)
2	*8*	4	50%	-16 (3); -16 and -17 (1)
3	*16*	9	56%	-16 (2); -16 and -17 (5); -16 and -11 (1)
4	*11*	4	36%	-16 (2); -16 and -17 (2)
5	*2*	2	100%	-16 and -17 (1); -2, -16 and -17 (1)
Total	*45*	27 (25)[b]	60% (55%)[c]	

[a] Negative numbers indicate the editing sites, and numbers in brackets indicate the number of colonies with the corresponding edit. Two colonies with mixed signals are included in the numbers here, but not in the numbers in the main text.
[b] Number in brackets indicates the edited colonies with pure signals.
[c] Number in brackets indicates the editing efficiency calculated with clean edited colonies only.

Table C.3 Summary of the *in-silico* evaluation of our base-editing tool.

Editing Type	Editing window (-11 to -19)		Editing window (-16 to -19)	
	Editable sites	**Editable genes**	**Editable sites**	**Editable genes**
Missense mutation	190568	4171	98162	4159
Silent mutation	66545	4144	22393	3896
Nonsense mutation	12745	3404	6149	2657
Nonsense mutation (70%)[a]	-	3009	-	2203
Non-editable[b]	-	7	-	21

[a] Nonsense mutation (70%) indicates that a premature STOP codon can be installed within the first 70% of the coding sequence of a target gene.
[b] Details for non-editable genes can be found in **Dataset 1 and 2** (C.1).

Table C.4 *Clostridium ljungdahlii* strains used in this study.

Strains	Description	References
DSM13528	*Clostridium ljungdahlii* type strain	DSM13528
QX3	DSM13528, *adhE1* Trp169*	This study
QX4	DSM13528, *adhE2* Gln33*	This study
QX5	DSM13528, *aor1* Gln267*	This study
QX6	DSM13528, *aor2* Gln267*	This study

Table C.5 Plasmids used in this study.

Name	Description	References
pANA1	p15A, *bla*, Φ3tI methyltransferase	Molitor *et al.* (2016b)
pMTLdSpCas9	pMTL82254, P_{2tetO1}:*dcas9*	Woolston *et al.* (2018)
pTargetF	ColE1, *smR*	Jiang *et al.* (2015)
pScI_dCas9-CDA-UL	pSC101, *bla*, *dcas9*-PmCDA1-UGI-LVA	Banno *et al.* (2018)
pgRNA01	pTargetF, P_{J23119}:gRNA01	This study
pgRNA02	pgRNA01, P_{J23119}:gRNA02	This study
pgRNA05	pgRNA01, P_{J23119}:gRNA05	This study
pgRNA06	pgRNA01, P_{J23119}:gRNA06	This study
pgRNA07	pgRNA01, P_{J23119}:gRNA07	This study
pgRNA10	pgRNA01, P_{J23119}:gRNA10	This study
pgRNA11	pgRNA01, P_{J23119}:gRNA11	This study
pgRNA13	pgRNA01, P_{J23119}:gRNA13	This study
pgRNA14	pgRNA01, P_{J23119}:gRNA14	This study
pgRNA15	pgRNA01, P_{J23119}:gRNA15	This study
pgRNA19	pgRNA01, P_{J23119}:gRNA19	This study
pFX	pMTLdSpCas9, dcas9 fused with PmCDA-UGI-LVA tag	This study
pFX01	pFX, P_{J23119}:gRNA01	This study
pFX02	pFX, P_{J23119}:gRNA02	This study
pFX05	pFX, P_{J23119}:gRNA05	This study
pFX06	pFX, P_{J23119}:gRNA06	This study
pFX07	pFX, P_{J23119}:gRNA07	This study
pFX10	pFX, P_{J23119}:gRNA10	This study
pFX11	pFX, P_{J23119}:gRNA11	This study
pFX13	pFX, P_{J23119}:gRNA13	This study
pFX14	pFX, P_{J23119}:gRNA14	This study
pFX15	pFX, P_{J23119}:gRNA15	This study
pFX19	pFX, P_{J23119}:gRNA19	This study

Table C.6 Primers used for plasmid construction

Primer	Sequence	Note
EBT-PFX-080	GATTTGAGTCAGCTAGGAGGTGACGGTGGAGGAGGTTCTGGAGG	pFX series
EBT-PFX-086	ATGCCTGGAGATCCTTACTCGAGTTATGCAACCAGTCCTAGCATCTTG	
EBT-PFX-081	CCTCCAGAACCTCCTCCACCGTCACCTCCTAGCTGACTCAAATC	

Continued on next page

Table C.6 – continued from previous page

Primer	Sequence	Note
EBT-PFX-087	CAAGATGCTAGGACTGGTTGCATAACTCGAGTAAGGATCTCC AGGCAT	
EBT-PFX-088	GGCTCACCTTCGGGTGGGCCTTTCTGCGTTACCGCATATGCTG GATCCTT	
EBT-PFX-089	ACGTTGTAAAACGACGGCCAGTGCCGAGCTCTGCAGGTCGAC TCTAGAGAAT	
EBT-PFX-090	TTACCTTTTCATTCCCTACAGTTTTAGAGCTAGAAATAGC	gRNA01
EBT-PFX-091	TGTAGGGAATGAAAAGGTAAGCTAGCATTATACCTAGGAC	
EBT-PFX-133	AAAAATTTGGAGTAAGGCAAGTTTTAGAGCTAGAAATAGC	gRNA02
EBT-PFX-134	TTGCCTTACTCCAAATTTTTGCTAGCATTATACCTAGGAC	
EBT-PFX-139	CAAATAGTAAAGACAGCTCCGTTTTAGAGCTAGAAATAGC	gRNA05
EBT-PFX-140	GGAGCTGTCTTTACTATTTGGCTAGCATTATACCTAGGAC	
EBT-PFX-141	AATTGATCACTATCTGGGCAGTTTTAGAGCTAGAAATAGC	gRNA06
EBT-PFX-142	TGCCCAGATAGTGATCAATTGCTAGCATTATACCTAGGAC	
EBT-PFX-143	TGAAATTGATGGAAAAAATTGTTTTAGAGCTAGAAATAGC	gRNA07
EBT-PFX-144	AATTTTTTCCATCAATTTCAGCTAGCATTATACCTAGGAC	
EBT-PFX-153	ATCCATCCTATAATTCCTTCGTTTTAGAGCTAGAAATAGC	gRNA10
EBT-PFX-154	GAAGGAATTATAGGATGGATGCTAGCATTATACCTAGGAC	
EBT-PFX-155	AACAAGTGGATGAAATTTTCGTTTTAGAGCTAGAAATAGC	gRNA11
EBT-PFX-156	GAAAATTTCATCCACTTGTTGCTAGCATTATACCTAGGAC	
EBT-PFX-178	GGTCAGGGAATGCCAACTTAGTTTTAGAGCTAGAAATAGC	gRNA13
EBT-PFX-179	TAAGTTGGCATTCCCTGACCGCTAGCATTATACCTAGGAC	
EBT-PFX-180	GATCAAGCAGATAAGATCAGGTTTTAGAGCTAGAAATAGC	gRNA14
EBT-PFX-181	CTGATCTTATCTGCTTGATCGCTAGCATTATACCTAGGAC	
EBT-PFX-182	TTCAAATAGTAAAGACAGCTCCGTTTTAGAGCTAGAAATAGC	gRNA15
EBT-PFX-183	GGAGCTGTCTTTACTATTTGAAGCTAGCATTATACCTAGGAC	
EBT-PFX-201	AATCAAGCAGATAAAATAAGGTTTTAGAGCTAGAAATAGC	gRNA19
EBT-PFX-202	CTTATTTTATCTGCTTGATTGCTAGCATTATACCTAGGAC	

Table C.7 Primers used for the verification of base editing.

Primer	Sequence	Note
EBT-PFX-131	AGGATAGGACATACCCTGTG	Verification of *pta* editing
EBT-PFX-132	CATCTACAGACATGCCTGTTC	
EBT-PFX-130	GGAGTAAGGCAAAGGAAGAC	Sequencing for *pta*
EBT-PFX-166	TCCCCAATTTAGCATACTAGGC	Verification of *adhE1* editing
EBT-PFX-167	CACATATGCCTCCAGTGCAT	
EBT-PFX-168	TTACTGACTGCTCTGAGGCA	Sequencing for *adhE1*
EBT-PFX-169	ATGCACTGGAGGCATATGTG	Verification of *adhE2* editing
EBT-PFX-170	GTGCAACTCCAAGACTACCAT	
EBT-PFX-171	AGGAGCACCAGCTTTAACTG	Sequencing for *adhE2*
EBT-PFX-163	TGAAGAAGCGCTTCAAGTTC	Verification of aor1 editing
EBT-PFX-164	CTGCCTCTAATAGTGAATCTGC	
EBT-PFX-200	CGTTGGTGCAGTTATGGGAT	Sequencing for *aor1*
EBT-PFX-187	CTAAGGCAATGGGGATTGGA	Verification of *aor2* editing
EBT-PFX-188	AGTTTCCACCTCCTTAGGCTA	
EBT-PFX-189	GGGAGCAGAATTCAAAGCAG	Sequencing for *aor2*

Appendix D

Supplementary information for Chapter 6.

- Supplementary text D.1.1 to D.1.6
- Supplementary figures D.1 to D.8
- Supplementary tables D.1 to D.7

D.1 Supplementary text

D.1.1 Implementing a CRISPR-Cas12a system for *C. ljungdahlii*

For *C. ljungdahlii*, the first CRISPR-based gene-deletion system was implemented with CRISPR-Cas9 (Huang *et al.*, 2016). Later, a CRISPR-Cas12a system was successfully realized in *C. ljungdahlii* (Zhao *et al.*, 2019). Cas12a, also known as Cpf1, is an alternative Cas-type nuclease, which uses an AT-rich protospacer adjacent motif sequence instead of GC-rich protospacer adjacent motif sequences, which are preferred by Cas9. Since genomes of acetogens are in general AT rich (Drake *et al.*, 2008), the utilization of a CRISPR-Cas12a system offers more potential genome-editing sites compared to a CRISPR-Cas9 system. In addition, Cas12a cleaves the targeted DNA in a staggered pattern, which is postulated to increase the efficiency of DNA-repair mechanisms or might allow gene insertion through non-homologous end joining (Fagerlund *et al.*, 2015; Zetsche *et al.*, 2015; Bayat *et al.*, 2018).

We implemented a CRISPR-Cas12a system in the shuttle-vector system pMTL80000 (Heap *et al.*, 2009) (**Figure 6.1A**). We chose the constitutive thiolase promoter P_{thl} (Heap *et al.*, 2009) and the anhydrotetracycline-inducible promoter $P_{tetR-O1}$ (Dong *et al.*, 2012; Woolston *et al.*, 2018), to investigate whether the expression of the

Cas12a-nuclease gene itself leads to poor transformation efficiency and genome edit rates as previously reported (Huang *et al.*, 2016). While cloning of the Cas12a gene was readily achieved in *E. coli*, we required several assembling attempts to form the final CRISPR-Cas12a plasmids, which contained the homology-directed repair arms and single guide RNAs. Once the plasmids were generated, we did not observe a noticeable difference between growth of the respective *E. coli* strain or an *E. coli* strain that harbors an empty pMTL plasmid. We also did not see any difference between recombinant *C. ljungdahlii* strains that harbor the control plasmids pMTL83152_Cas12a and pMTL83151_*tetR-O1*_Cas12a, which both lack homologous repair DNA and sgRNAs. However, transfer of the final CRISPR-Cas12a plasmids into *C. ljungdahlii* cells could only be achieved by using conjugation instead of electroporation. Overall, we found successful genome edited cells by using the plasmids pMTL83152_Cas12a-RNF, pMTL83152_Cas12a-rseC, and pMTL83152_Cas12a-nar. All CRISPR-Cas12a plasmids that harbored the $P_{tetR-O1}$ promoter instead of P_{thl} were successfully transferred into *C. ljungdahlii* cells, but genome edits were not detectable after induction with anhydrotetracycline and screening several colonies.

D.1.2 Confirmation of strains

We purified gDNA from the generated strain *C. ljungdahlii* ΔRNF and used it for PCR analyses (Material and Methods, 6.4). A *rnfCDGEAB* fragment could only be amplified when using gDNA from *C. ljungdahlii* WT but not with gDNA of the deletion strain (**Figure 6.1B**). In addition, the deletion strain showed the expected shortened fragment when using primers that bound outside of the RNF-gene cluster (**Figure 6.1B**). Sanger sequencing of this fragment confirmed the precise deletion of *rnfCDGEAB* from the genome. We performed similar PCR screening experiments with the *C. ljungdahlii* Δ*rseC* and Δ*nar* strain to confirm the genome edits (**Figure D.2**). After generating the *C. ljungdahlii* Δ*rseC* strain and verifying the successful genome edit, several transfers in non-selective RCM and a subsequent cultivation at 42°C for 72 h were required to cure the strain from the pMTL83152_Cas12a-rseC plasmid. The genome edit remained stable during this procedure. In contrast, the plasmid curing was achieved quickly for the *C. ljungdahlii* ΔRNF and Δ*nar* strains after several transfers in non-selective RCM and subsequent isolation of single colonies on non-selective RCM plates.

D.1.3 Growth of *C. ljungdahlii* WT with nitrate or ammonium as nitrogen source

During autotrophy with ammonium, *C. ljungdahlii* WT showed a growth rate of 0.024±0.002 h^{-1}, reached its maximum OD_{600} at 0.56±0.01, and produced 59.5±1.8 mM acetate and 1.9±0.4 mM ethanol (**Table 6.1, Figure 6.2**). When using nitrate instead, the growth rate was 0.072±0.004 h^{-1} and the maximum OD_{600} increased to 1.00±0.06 (**Table 6.1, Figure 6.2A**). This is a significant increase of 198% ($P \leq 0.001$) and 79% ($P \leq 0.001$) in comparison to ammonium conditions, respectively. In contrast, the maximum acetate concentration decreased to 50.1±2.1 mM, which is a significant reduction of 16% (P = 0.008), and maximum ethanol concentrations increased to 8.0±1.6 mM corresponding to a significant increase of 327% (P = 0.007) when using nitrate as nitrogen source (**Table 6.1, Figure 6.2C, 6.2D**). Notably, the ethanol production of nitrate grown cells only started after all nitrate was consumed at a cultivation time of 47 h. During heterotrophy, we observed that *C. ljungdahlii* WT performed slightly better in ammonium-containing medium in terms of growth and production of acetate and ethanol (**Table D.1, Figure D.1**). With nitrate and fructose, the growth rate decreased by 7% (P = 0.04), maximum OD_{600} values dropped by 10% (P = 0.03), and maximum concentrations for acetate were significantly reduced by 17% ($P \leq 0.001$) and for ethanol by 53% ($P \leq 0.001$) in comparison to growth with fructose and ammonium (**Table D.1**). During autotrophy, we observed that the pH of the medium initially increased up to pH 7.31, before it decreased again until the end of the cultivation (**Figure 6.2B**). The pH of ammonium cultures only decreased during the cultivation. This pH effect was not observed during heterotrophy, but the pH decreased slower for those cultures growing with nitrate (**Figure D.1B**). Notably, during heterotrophy, a halt in growth, pH decrease, and metabolic activity was observed after 55.5 h of cultivation when the fructose pool had been consumed completely (**Figure D.1**). In all our experiments, we found that a sufficient amount of nitrogen was supplied already by the yeast extract (1 g/L) that we had added, because further ammonium accumulated and was not consumed, even when only ammonium was provided as the nitrogen source. Similar observations were already reported by Emerson *et al.* (2019).

D.1.4 Effects of promoter strength and plasmid copy number on the plasmid-based complementation of the gene deletion strains

Our complementation plasmids contained the Gram-positive origin of replication pCB102, which was originally derived from *Clostridium butyricum* (Heap *et al.*, 2009). However, neither the exact replication mechanism nor the plasmid copy number are known. We speculated that at least several copies of our plasmid were present in our complementation strains. A negative impact on the growth due to a higher copy number of the *rnfCDGEAB* gene cluster in our complementation strain *C. ljungdahlii* pMTL83151_P_{nat}_*rnfCDGEAB* might explain the difference in maximum OD_{600} values and acetate concentrations (**Table 6.1, Figure D.3**). The difference between the *rnf*-complementation and the *rseC*-complementation strain might be explained by the different promoters that we used (non-characterized P_{nat} *vs.* constitutive P_{thl}).

We observed a growth stimulating effect in the *C. ljungdahlii* Δ*rseC* pMTL83152_*rseC* strain, even though we typically find plasmid-carrying strains to perform worse, which we argued is due to the cellular burden that is generated by the expression of plasmid-encoded functions and the addition of antibiotics. To investigate whether overexpression of *rseC* in the wild-type strain increases autotrophic growth further, we generated the *C. ljungdahlii* pMTL83152_*rseC* strain. This strain carries the complementation plasmid with the constitutive P_{thl} promoter in the wild-type background. During autotrophy with carbon dioxide and hydrogen in ammonium-containing medium of the *C. ljungdahlii* pMTL83152_*rseC* strain, the maximum acetate concentration was significantly reduced compared to the wild type and ethanol was not produced (**Figure D.4C, D.4D**). Overall, this strain performed similar when compared to the *C. ljungdahlii* Δ*rseC* pMTL83152_*rseC* strain.

We also attempted to generate a plasmid that carries the *rnfCDGEAB* gene cluster under the control of a constitutive promoter. However, any attempts to generate a fusion of the constitutive promoter P_{thl} with the *rnfCDGEAB* gene cluster failed already during the cloning steps in *E. coli*. Thus, for the expression of *rnfCDGEAB* in the wild type, we also used the native P_{nat} promoter sequence, which most likely is under the same expression control as the genomic copy of the RNF-gene cluster. Indeed, the cultivation of *C. ljungdahlii* pMTL83151_P_{nat}_*rnfCDGEAB* did not show any notable impact on growth and product formation when compared to the control strain that carried an empty plasmid, and the performance was similar to the complementation strain *C. ljungdahlii*

ΔRNF pMTL83151_P_{nat}_*rnfCDGEAB* (**Figure D.4**).

D.1.5 The role of *rseC* genes in non-acetogens

The *rseC* gene was already described as an important factor in the regulation of the oxidative stress response in E. coli, which is mediated by SoxR and the *rsxABCDGE* genes that share homology to the *rnf* genes in *R. capsulatus* (Koo *et al.*, 2003). In *E. coli*, the *rseC* gene is organized in the *rseD-rpoE-rseABC* operon, but located separately from the *rsxABCDGE* genes (De Las Peñas *et al.*, 1997; Missiakas *et al.*, 1997; Koo *et al.*, 2003). The proteins that are encoded by the *rseD-rpoE-rseABC* operon integrate signals from the redox state of SoxR, which senses the cellular levels of the oxidants superoxide and nitric oxide, and which is reduced by the membrane-bound complex Rsx. Thus, it is assumed that electrons from NADH are channeled through the Rsx complex and are transferred onto SoxR directly or indirectly to regenerate a reduced state after oxidation by oxidants (Koo *et al.*, 2003; Biegel *et al.*, 2011). The proposed system would be similar to the electron translocation of the RNF complex (Biegel *et al.*, 2011). It is assumed that the RseC protein is responsible to regulate the expression of SoxR by repressing its own regulator gene *soxS* (Koo *et al.*, 2003). In *E. coli*, a Tn*10*-transposon insertion mutant of *rseC* showed increased expression levels of the gene *soxS* (Koo *et al.*, 2003). The level of *soxS* transcript was found to be responsible for the redox state of SoxR, and higher levels of *soxS* mRNA indicated higher oxidation rates of SoxR (Ding and Demple, 1997). Therefore, it is assumed that the function of RseC in *E. coli* is likely to keep the level of *soxS* transcript low, which then keeps SoxR in its reduced form (Koo *et al.*, 2003). However, whether RseC interacts or interferes either with the Rsx complex or *soxS* transcript/SoxR is not understood, and neither is the function as transcription regulator (Koo *et al.*, 2003). Koo *et al.* (2003) reported that *rseC* of *E. coli* shares homology to the N-terminal half of the *rnfF* gene in *R. capsulatus*, and thus postulated a regulatory function of *rseC* in the nitrogen fixation. In our *in-silico* analysis, we found that *rnfF* and *rseC* in *R. capsulatus* are likely separated genes (**Figure 6.4**). This does not exclude the potential role of *rseC* in the nitrogen fixation regulation, but it questions the homology between *rseC* and *rnf* genes. In addition, the *rnfF* gene is not part of the RNF complex gene cluster in genomes of acetogens.

D.1.6 Conservation of the RseC amino-acid sequence

The conservation of the RseC amino-acid sequence was between 59% and 100% for *C. ljungdahlii*, *C. autoethanogenum*, *C. carboxidovorans*, and *C. kluyveri*, which is a high similarity (**Figure D.6**). In addition, the amino-acid sequence length is nearly identical with 138 amino acids (*C. ljungdahlii*, *C. autoethanogenum*, and *C. carboxidovorans*) and 137 amino acids (*C. kluyveri*), respectively. The second RseC homolog from *C. carboxidovorans* and *C. kluyveri* shared an identity of 65% with each other, but only between 25% and 49% to all other RseC proteins (**Figure D.6**). The RseC from *A. woodii* and *E. limosum* shared a similarity of 57% with each other, and only of 34% to 35% with the RseC proteins that are encoded directly upstream of the RNF-gene clusters (**Figure 6.4, Figure D.6**). The RseC proteins from *R. capsulatus* and *E. coli* have the same amino-acid sequence length (159 amino acids), but shared low similarities to each other (31%) as well as to the RseC proteins from the other microbes (18-34%) (**Figure D.6**). The similarity of the RseC protein from *C. ljungdahlii* and *R. capsulatus* was only 23%, while it was 36% for the RseC protein from *C. ljungdahlii* in comparison to the RseC protein from *E. coli*.

D.2 Supplementary figures

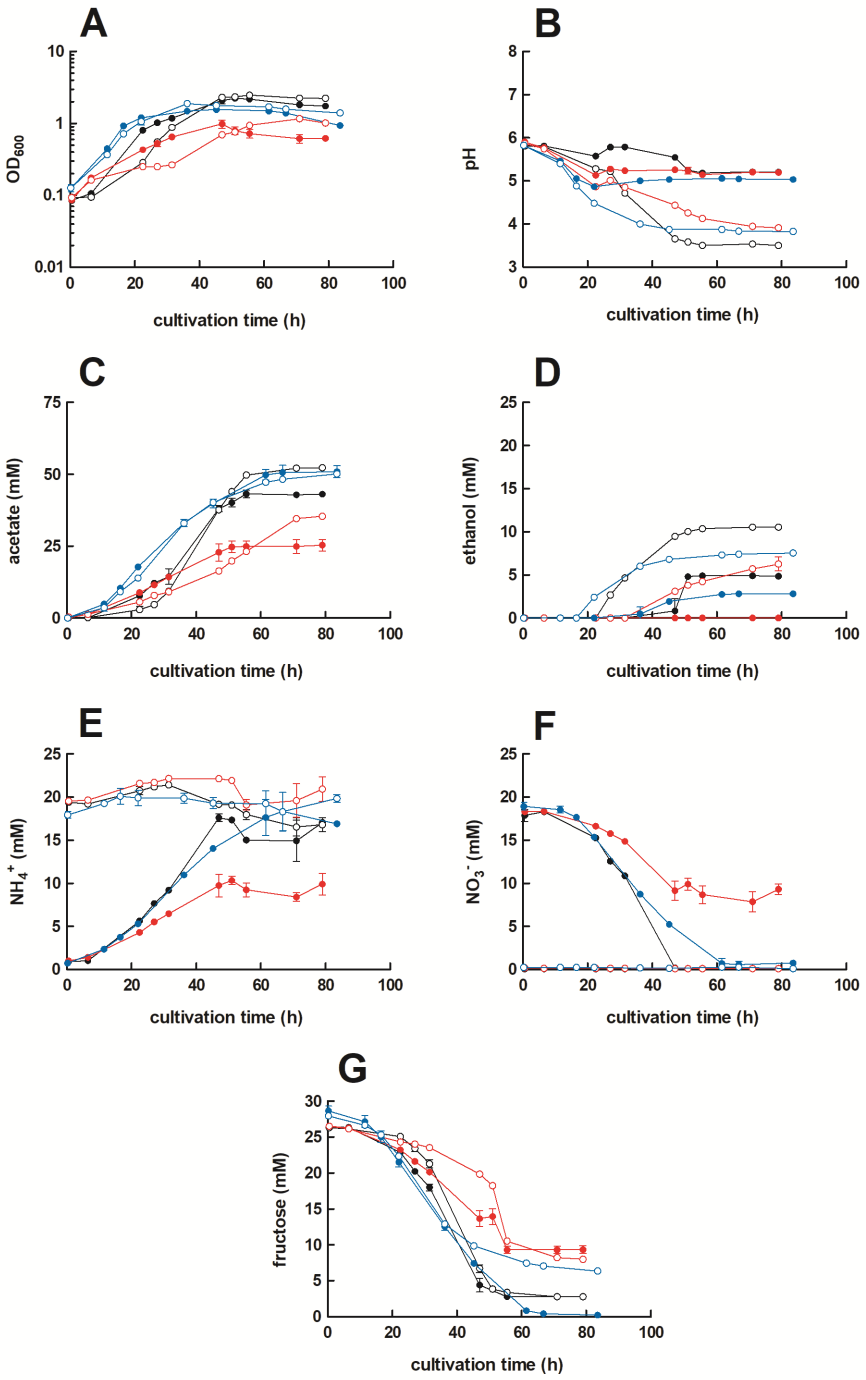

Figure D.1: (Continued on the following page.)

Figure D.1: (Previous page.) **Heterotrophic growth and metabolic products of *C. ljungdahlii* WT, ΔRNF, and Δ*rseC*.** Cultures of *C. ljungdahlii WT* (●, ○), ΔRNF (●, ○), and Δ*rseC* (●, ○) were grown in 100 mL PETC medium in 240 mL bottles at 37°C. The headspace consisted of N_2 (100 vol-%). Fructose (5 g/L) was added as carbon source. The medium contained either 18.7 mM nitrate (NO_3^-) (filled circles) or 18.7 mM ammonium (NH_4^+) (open circles) as nitrogen source. The cultivation times were 79 h for the WT and ΔRNF strain, and 84 h for the Δ*rseC* strain. All cultures were grown in biological triplicates, data is given as mean values, with error bars indicating the standard deviation. **(A)** growth; **(B)** pH-behavior; **(C)** acetate concentrations; **(D)** ethanol concentration; **(E)** ammonium concentration; and **(F)** nitrate concentrations. WT, wild type; ΔRNF, RNF-gene cluster deletion; Δ*rseC*, *rseC* gene deletion.

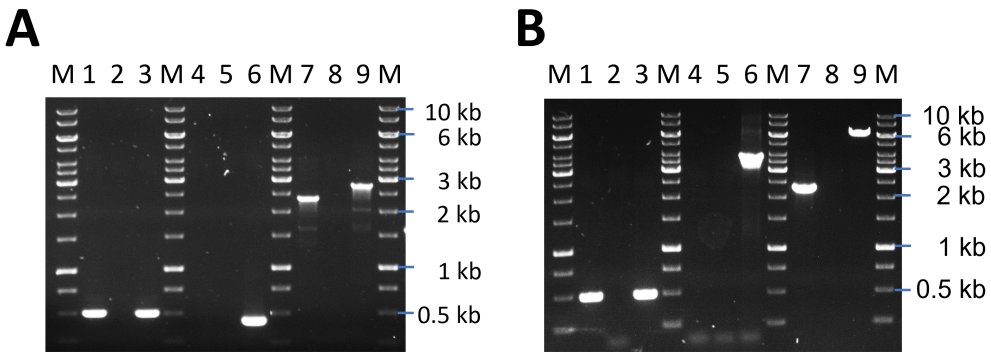

Figure D.2: CRISPR-Cas12a-mediated *rseC* gene and *nar* gene cluster deletion in *C. ljungdahlii*. **(A)** verification of the *rseC* gene deletion. PCR-samples for the *fdhA* fragment (WT: 501 bp, deletion strain: 501 bp), *rseC* fragment (WT: 417 bp, deletion strain: no fragment), and for a fragment that was amplified with primers that bind 1104 bp upstream and 1208 bp downstream of the *rseC* gene locus (WT: 2755 bp, deletion strain: 2338 bp). DNA-template: gDNA of *C. ljungdahlii* Δ*rseC* (lane A1, A4, and A7); gDNA of *C. ljungdahlii* WT (lane A3, A6, and A9); and water (lane A2, A5, A8). **(B)** verification of the *nar* gene cluster deletion PCR samples for the *fdhA* fragment (WT: 501 bp, deletion strain: 501 bp), *nar* fragment (WT: 3739 bp, deletion strain: no fragment), and for a fragment that was amplified with primers that bind 1137 bp upstream and 1110 bp downstream of the *nar* gene cluster locus (WT: 5986 bp, deletion strain: 2247 bp). DNA-template: gDNA of *C. ljungdahlii* Δ*nar* (lane B1, B4, and B7); gDNA of *C. ljungdahlii* WT (lane B3, B6, and B9); and water (lane B2, B5, B8). M: Generuler^{TM} 1 kb.

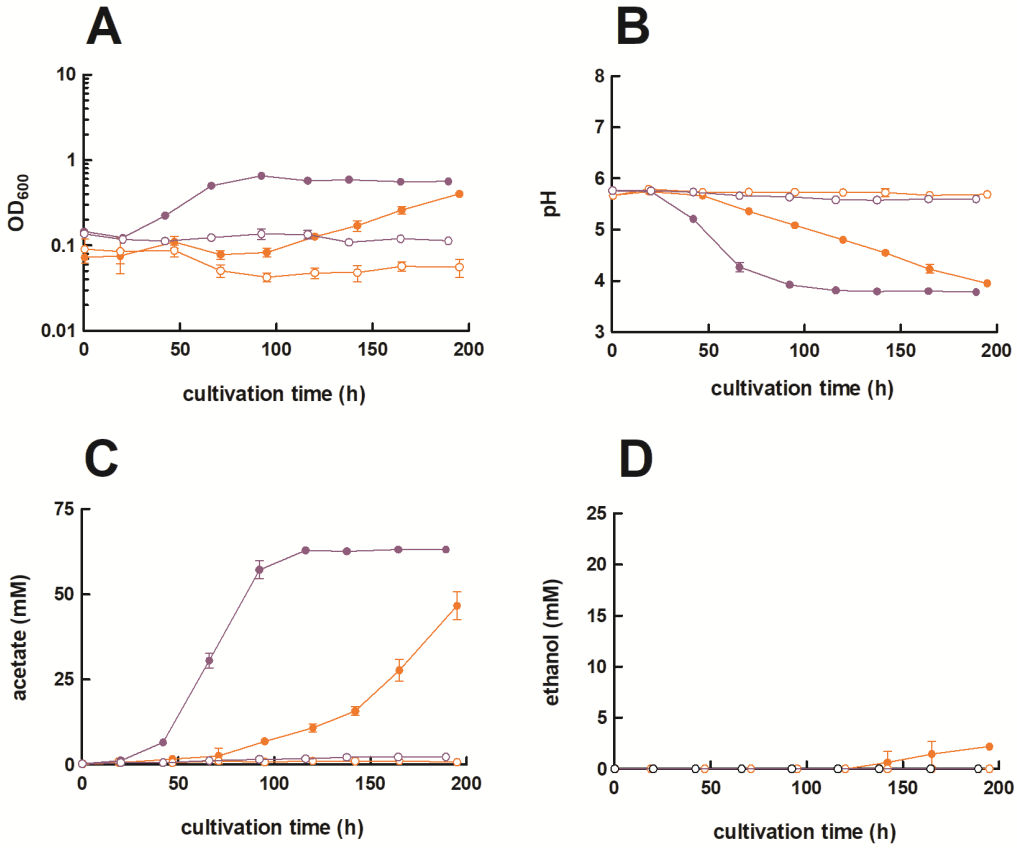

Figure D.3: Growth and pH behavior of plasmid-based complementation of *C. ljungdahlii* ΔRNF and *C. ljungdahlii* Δ*rseC* with H_2 and CO_2. Cultures were grown in 100 mL PETC medium in 1 L bottles at 37°C and 150 rpm for 195 h and 189 h, respectively. The headspace consisted of H_2 and CO_2 (80/20 vol-%) and was set to 0.5 bar overpressure. Only 18.7 mM ammonium (NH_4^+) but no nitrate was added to the medium. All cultures were grown in biological triplicates, data is given as mean values, with error bars indicating the standard deviation. The *C. ljungdahlii* WT data (○) from Fig. 1 is given for comparison. **(A)**, growth and **(B)**, pH-behavior, **(C)**, acetate concentration, and **(D)**, ethanol concentration of the *C. ljungdahlii* ΔRNF and Δ*rseC* strains. (●) *C. ljungdahlii* ΔRNF pMTL83151_P_{nat}_*rnfCDGEAB*; (○) *C. ljungdahlii* ΔRNF pMTL83151; (●) *C. ljungdahlii* Δ*rseC* pMTL83152_*rseC*; and (○) *C. ljungdahlii* Δ*rseC* pMTL83152. ΔRNF, *rnfCDGEAB* gene cluster deletion; Δ*rseC*, deletion of *rseC*; P_{nat}, native promoter sequence upstream of *rnfC*; P_{thl}, promoter of the thiolase gene in *C. acetobutylicum*; rpm, revolutions per minute; CO_2, carbon dioxide; and H_2, hydrogen.

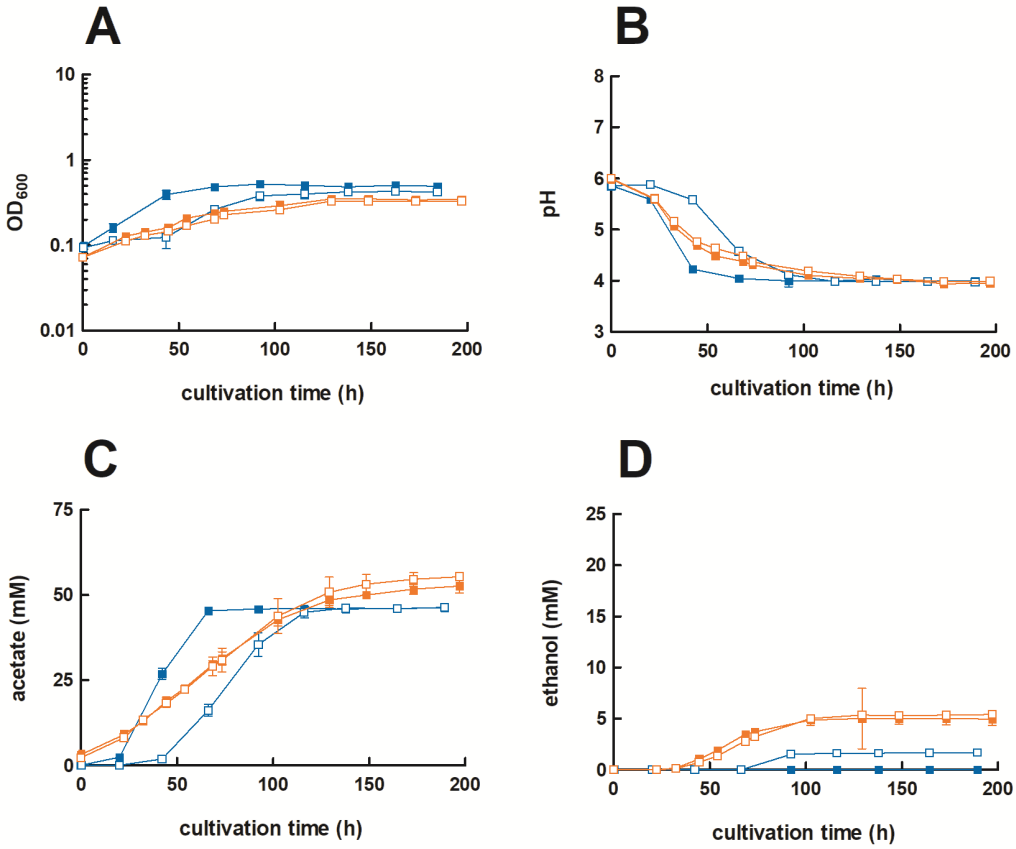

Figure D.4: Autotrophic growth and metabolic products of the overexpression strains *C. ljungdahlii* pMTL83151_P_{nat}_*rnfCDGEAB* and *C. ljungdahlii* pMTL83152_*rseC*. Cultures were grown in 100 mL PETC medium in 1 L bottles at 37°C and 150 rpm. The headspace consisted of H_2 and CO_2 (80/20 vol-%) and was set to 0.5 bar overpressure. For the strain *C. ljungdahlii* pMTL83151_P_{nat}_*rnfCDGEAB* and the control strain *C. ljungdahlii* pMTL83151 we refilled the headspace during this experiment with the same gas mixture to 0.5 bar overpressure at time points 44.5 h, 73.5 h, and 148.5 h. The medium contained 18.7 mM ammonium as nitrogen source. Thiamphenicol (5 μg/mL) was used for selection. All cultures were grown in biological triplicates, data is given as mean values, with error bars indicating the standard deviation. The cultivation time was 185 h and 197 h for *C. ljungdahlii* pMTL83151_P_{nat}_*rnfCDGEAB* and *C. ljungdahlii* pMTL83152_*rseC*, respectively. (■) *C. ljungdahlii* pMTL83151_P_{nat}_*rnfCDGEAB*; (□) *C. ljungdahlii* pMTL83151 (empty plasmid); (■) *C. ljungdahlii* pMTL83152_*rseC*; (□) *C. ljungdahlii* pMTL83152 (empty plasmid). The *C. ljungdahlii* WT data (○) from **Figure 5.1** is given for comparison. **(A)** growth; **(B)** pH-behavior; **(C)** acetate concentrations; and **(D)** ethanol concentration. rpm, revolutions per minute; CO_2, carbon dioxide; and H_2, hydrogen.

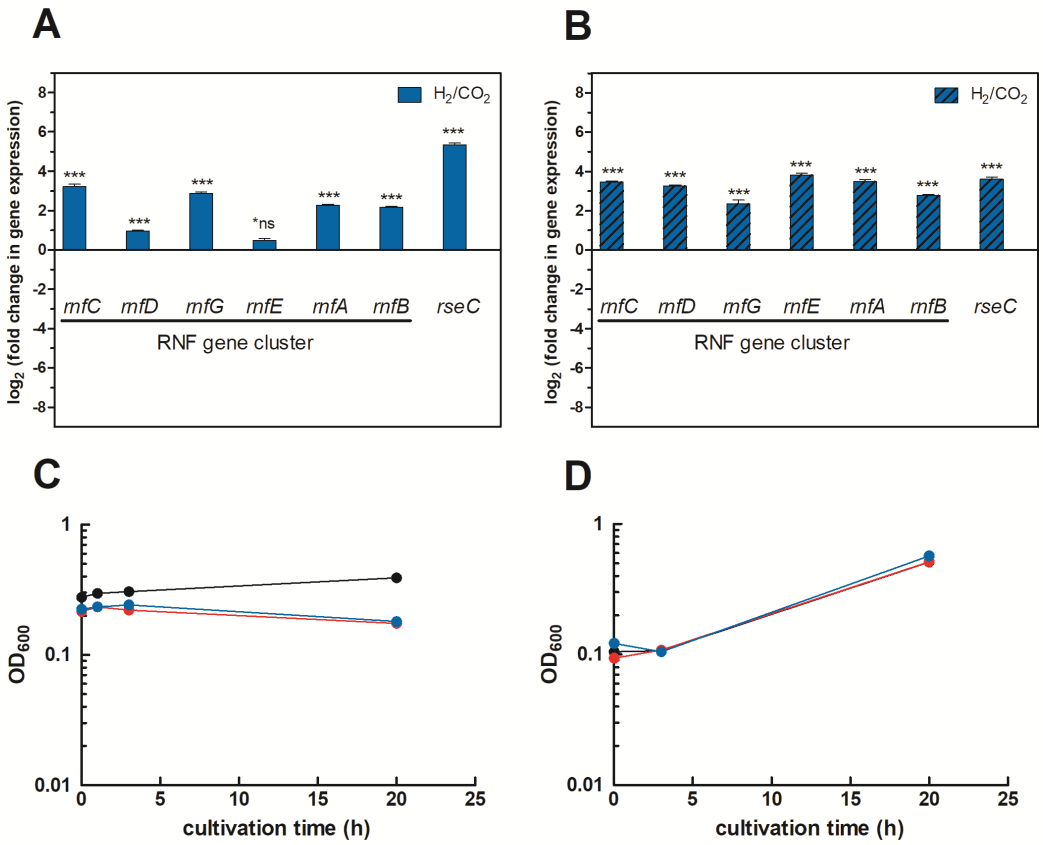

Figure D.5: Gene expression change of the *rnfCDGEAB* cluster genes and the *rseC* gene in the wild-type strain from heterotrophy to autotrophy and growth curves of the cultivation experiments for the qPCR analyses. **(A)** gene expression change after 3 h cultivation time; **(B)** gene expression change after 20 h cultivation time; **(C)** growth of *C. ljungdahlii* WT (●), ΔRNF (●), and Δ*rseC* (●) during autotrophic conditions; D, growth of *C. ljungdahlii* WT (●), ΔRNF (●), and Δ*rseC* (●) during heterotrophic conditions. RNA samples were purified from cultures that were cultivated either autotrophically with hydrogen and carbon dioxide or heterotrophically with fructose. Cells of the deletion mutants *C. ljungdahlii* ΔRNF and *C. ljungdahlii* Δ*rseC* were not growing under autotrophic conditions and remained at their inoculation OD_{600} of 0.2±0.02 and 0.22±0.02, respectively. cDNA was synthesized from the purified RNA samples and used as template for qRT-PCR analyses. The *rho* gene was used as "housekeeping" gene. The fold change in gene expression was determined with the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001). ***, P $\leq$ 0.001; and *ns, not significant (P $>$ 0.5). We defined $\log_2$FC $\leq$ -2 as downregulated genes and $\geq$ +2 as upregulated genes.

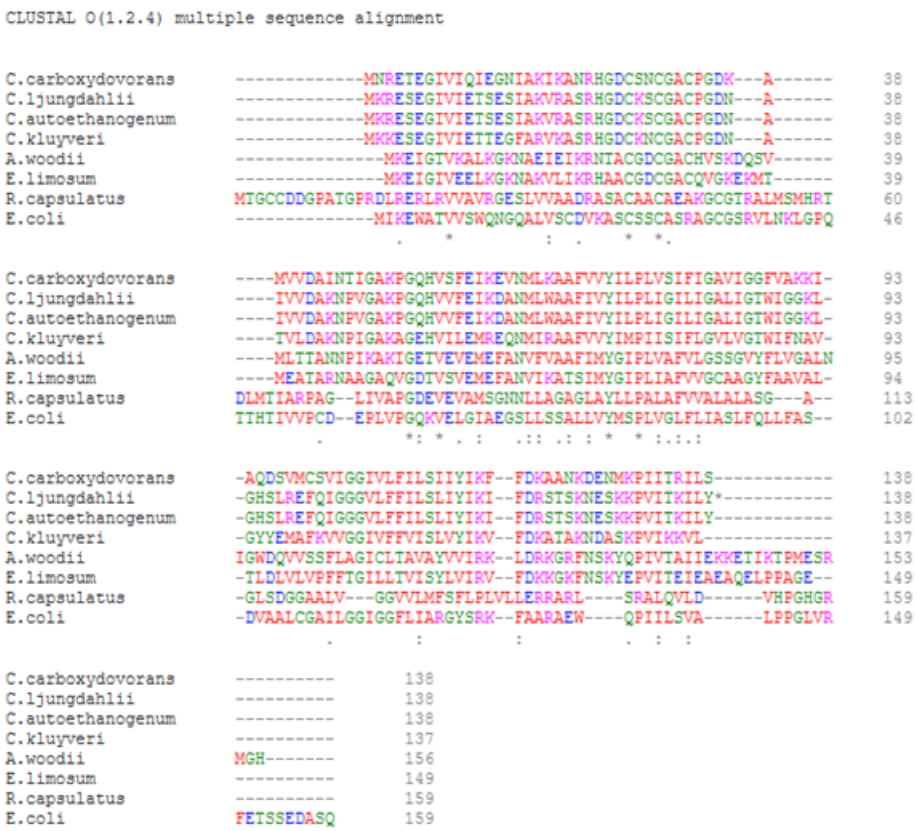

Figure D.6: Multiple sequence alignment of RseC amino-acid sequence using CLUSTAL Omega. The symbols indicate low similarity (.), high similarity (:), and identical amino acids (*) between the amino acid sequences. Similar colors indicate similar amino acids. The type strains were *C. ljungdahlii* DSM13528; *C. autoethanogenum* DSM10061; *C. carboxidovorans* P7; *C. kluyveri* DSM555; *E. limosum* ATCC8486; *A. woodii* DSM1030; *R. capsulatus* SB1003; and *E. coli* K-12. Clustal omega version 1.2.4. with default settings was used for the analysis (`https://www.ebi.ac.uk/Tools/msa/clustalo/`, 05/2021). All listed RseC proteins are predicted to contain two transmembrane helices (`https://services.healthtech.dtu.dk/service.php?TMHMM-2.0`, (Möller *et al.*, 2001)).

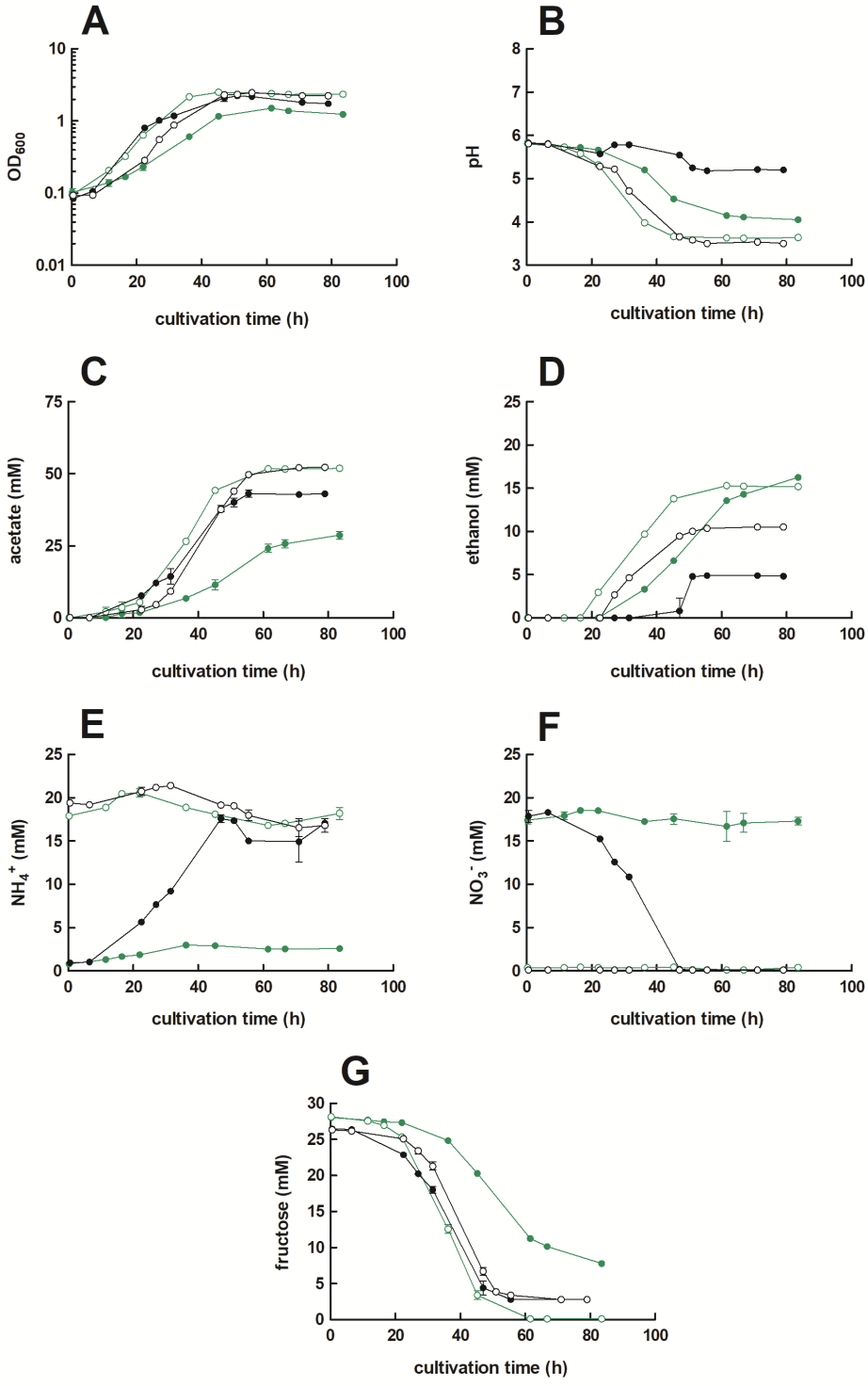

Figure D.7: (Caption on the next page.)

Figure D.7: (Previous page.) **Heterotrophic growth and metabolic products of *C. ljungdahlii* Δ*nar*.** Cultures were grown in 100 mL PETC medium in 240 mL bottles at 37°C. Fructose (5 g/L) was added as carbon source. The headspace consisted of N_2 (100 vol-%). The medium contained either 18.7 mM nitrate (●) or 18.7 mM ammonium (○) as nitrogen source. All cultures were grown in biological triplicates, data is given as mean values, with error bars indicating the standard deviation. The cultivation times was 84 h. The *C. ljungdahlii* WT data (●, ○) from **Figure D.1** is given for comparison. **(A)** growth; **(B)** pH-behavior; **(C)** acetate concentrations; **(D)** ethanol concentration; **(E)** ammonium concentration; **(F)** nitrate concentrations; and **(G)** fructose concentrations. Δ*nar*, deletion of the nitrate reductase genes.

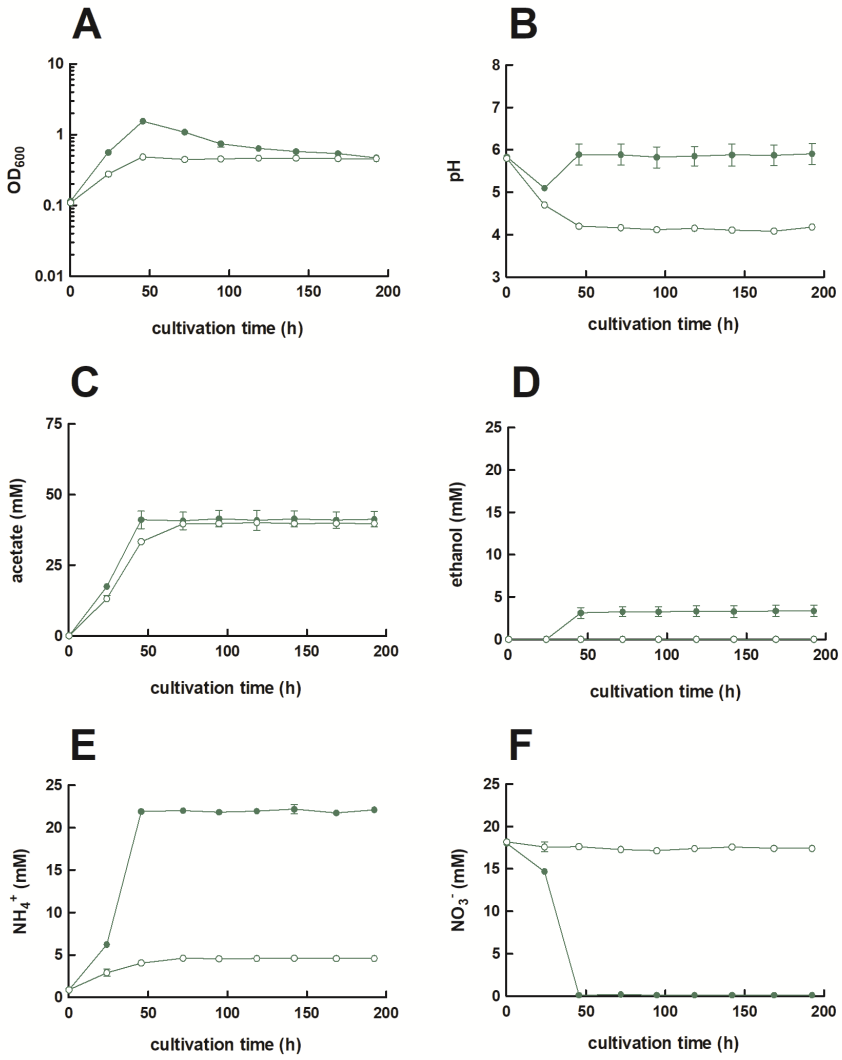

Figure D.8: (Caption on the next page.)

Figure D.8: (Previous page.) **Autotrophic growth and metabolic products of plasmid-based complemented strain *C. ljungdahlii* Δ*nar* pMTL83152_*nar***. Cultures were grown in 100 mL PETC medium in 1 L bottles at 37∘C and 150 rpm. The headspace consisted of H_2 and CO_2 (80/20 vol%) and was set to 0.5 bar overpressure. The medium contained 18.7 mM nitrate (NO_3^-) but no ammonium (NH_4^+) as nitrogen source. Thiamphenicol (5 μg/mL) was used for selection. All cultures were grown in biological triplicates, data is given as mean values, with error bars indicating the standard deviation. The cultivation times was 192.5 h. (•) *C. ljungdahlii* Δ*nar* pMTL83152_*nar*; (○) *C. ljungdahlii* Δ*nar* pMTL83152 (empty plasmid); **(A)** growth, **(B)** pH-behavior; **(C)** acetate concentrations; **(D)** ethanol concentration; **(E)** ammonium concentration; and **(F)** nitrate concentrations. Δ*nar*, gene deletion of the nitrate reductase genes; rpm, revolutions per minute; CO_2, carbon dioxide; and H_2, hydrogen.

D.3 Supplementary tables

Table D.1 Plasmids used in this study.

Plasmids	Function	Source
pMTL83151	shuttle-vector	(Heap *et al.*, 2009)
pMTL83152	shuttle-vector with constitutive thiolase promoter P_{thl}	(Heap *et al.*, 2009)
pMTL2tetO1gusA	pMTL82254 with pminithl:tetR-O1 and p2tetO1:gusA	(Woolston *et al.*, 2018)
pMTL8315tet	shuttle-vector with inducible promoter system *tetR-O1*	this study
pMTL83151_P_{nat}_*rnfCDGEAB*	overexpression of *rnfCDGEAB* through native promoter P_{nat}	this study
pMTL83152_*rseC*	overexpression of *rseC* through constitutive promoter P_{thl}	this study
pMTL83152_*nar*	overexpression of *nar* through constitutive promoter P_{thl}	this study
pY001_FnCpf1(Cas12a)	expression of FnCas12a	(Zetsche *et al.*, 2015), Addgene 69973
pMTL83152_FnCas12a	constitutive expression of FnCas12a through P_{thl}	this study
pMTL83152_FnCas12a_Δ*rseC*	constitutive expression of FnCas12a through P_{thl}, constitutive expression of a single sgRNA targeting *rseC* on the genome, fused repair HDR1/2 fragment for homologous recombination and marker-less gene deletion	this study
pMTL83152_FnCas12a_Δ*nar*	constitutive expression of FnCas12a through P_{thl}, constitutive expression of two sgRNA targeting *nar* on the genome, fused repair HDR1/2 fragment for homologous recombination and marker-less gene deletion	this study

Continued on next page

Table D.1 – continued from previous page

Plasmids	Function	Source
pMTL83152_FnCas12a_Δ*rnfCDGEAB*	constitutive expression of FnCas12a through P_{thl}, constitutive expression of two sgRNA targeting *rnfCDGEAB* on the genome, fused repair HDR1/2 fragment for homologous recombination and marker-less gene deletion	this study
pMTL8315tet_FnCas12a	inducible expression of FnCas12a through tetR-O1 promoter system	this study
pMTL8315tet_FnCas12a_Δ*rseC*	inducible expression of FnCas12a through tetR-O1 promoter system, constitutive expression of a single sgRNA targeting *rseC* on the genome, fused repair HDR1/2 fragment for homologous recombination and marker-less gene deletion	this study
pMTL8315tet_FnCas12a_Δ*nar*	inducible expression of FnCas12a through tetR-O1 promoter system, constitutive expression of two sgRNA targeting *nar* on the genome, fused repair HDR1/2 fragment for homologous recombination and marker-less gene deletion	this study
pMTL8315te_FnCas12a_Δ*rnfCDGEAB*	inducible expression of FnCas12a through tetR-O1 promoter system, constitutive expression of two sgRNA targeting *rnfCDGEAB* on the genome, fused repair HDR1/2 fragment for homologous recombination and marker-less gene deletion	this study

Table D.2 Primers used in this study.

Primer	Sequence	Function
rnfCDGEAB+213bp_fw_BamHI	GGATCCGTAATTTGTGTACAAACTTTAATTAATGGAGAGAC	Amplification of the RNF complex gene cluster (CLJU_c11360-410) + and promoter sequence (P_{nat})
rnfCDGEAB_rv_NcoI	CCATGGTTATGAATTTGCAGCAGCTTCATTCTTG	Amplification of the RNF complex gene cluster (CLJU_c11360-410) + and promoter sequence (P_{nat})
rseC_fwd_BamHI	GGATCCAGGAGGTTAAGAATGAAAAGAGAATCGGAGGGTATTG	Amplification of the putative RNF regulator gene *rseC* (CLJU_c11350)
rseC_rv_NcoI	CCATGGTCAATACAATATCTTTGTGATTACTGGC	Amplification of the putative RNF regulator gene *rseC* (CLJU_c11350)
nar-full_fwd_BamHI	GGCAGCTTACCGGGATCCAGGAGGTTAAGAATGAATTACGTGGAAGTAAAACAATCAAC	Amplification of a gene cluster (CLJU_c23710-30) encoding for a nitrate reductase
nar-full_rv_NcoI	GCACGGTCGTCGCCATGGTTAAAAAGTATACTCTAAATTTTCCTTTATATTAAAAAAGTC	Amplification of a gene cluster (CLJU_c23710-30) encoding for a nitrate reductase
Seq1_RNF_744bp_fwd	GGAAAATTCAGACAAGGTAGTTGC	Sanger sequencing of the *rnfCDGEAB* fragment
Seq2_RNF _1502bp_fwd	CAGAAAATAGAGCTGCAGGTGAAAG	Sanger sequencing of the *rnfCDGEAB* fragment
Seq3_RNF _2268bp_fwd	CTGGCAGATTCCAGTAGTAATGATTG	Sanger sequencing of the *rnfCDGEAB* fragment
Seq4_RNF _3000bp_fwd	GGGACAGTTTAAGGATAAAAAGGCAG	Sanger sequencing of the *rnfCDGEAB* fragment
Seq5_RNF_3787bp_fwd	GCAAATGGAGGTGAAGCATAATG	Sanger sequencing of the *rnfCDGEAB* fragment
Seq6_RNF _4502bp_fwd	GTGAATCCACTTGTAGACTTAGTAGAAG	Sanger sequencing of the *rnfCDGEAB* fragment

Continued on next page

Table D.2 – continued from previous page

Primer	Sequence	Function
Seq7_RNF_5047bp_rv	TTATGAATTTGCAGCAGCTTCATTCTTG	Sanger sequencing of the *rnfCDGEAB* fragment
Seq1 _nar_456bp_rv	GCACCTCCTTATACTCTAAAAGATTTTG	Sanger sequencing of the *nar* fragment
Seq2 _nar_610bp_fwd	CTGTTTCAGATTTTCTCGGGTCAATTG	Sanger sequencing of the *nar* fragment
Seq3 _nar_1059bp_rv	CCAAAGCATAGAGAAGAAATTGC	Sanger sequencing of the *nar* fragment
Seq4 _nar_1186bp_fwd	CCCACAATGCCTTAATTTCTCCG	Sanger sequencing of the *nar* fragment
Seq5 _nar_1677bp_rv	GTAAAGCTCATTTATGAAGATGCAGCC	Sanger sequencing of the *nar* fragment
Seq6 _nar_1798bp_fwd	CCCTAGTTCTAGTCTGGGTATGC	Sanger sequencing of the *nar* fragment
Seq7 _nar_2303bp_rv	CCAGATACCGGTATTGTAGAGTACG	Sanger sequencing of the *nar* fragment
Seq8 _nar_2476bp_fwd	CATCTAGCTACACACTGCGG	Sanger sequencing of the *nar* fragment
Seq9 _nar_2902bp_rv	GATGCACAAAAAATAAAGGATGCAGC	Sanger sequencing of the *nar* fragment
Seq10 _nar_3086bp_fwd	CTTCATATCTGCCTGCTGCA	Sanger sequencing of the *nar* fragment
Seq11 _nar_3385bp_rv	GGAATTGTAGCAGCTAGTAATATGGC	Sanger sequencing of the *nar* fragment
tetR-O1_fwd_SbfI	CCTGCAGGATAAAAAAATTGTAGATAAATTTTATAAAATAG	Amplification of the inducible promoter system tetR-O1
tetR-O1_rv_BamHI	GGATCCTATTTCAAATTCAAGTTTATCGCTCTAATGAAC	Amplification of the inducible promoter system tetR-O1
repH _401bp_rv	CTCTAACGGCTTGATGTGTTGG	Primer binding in the backbone of pMTL83151 and pMTL83152 upstream of *repH*

Continued on next page

Table D.2 – continued from previous page

Primer	Sequence	Function
fdhA_fwd	AGTGCAGCGTATTCGTAAGG	Amplification of a 501 bp fragment of the *fdhA* gene in C. ljungdahlii
fdhA_rv	TAATGAGCCACGTCGTGTTG	Amplification of a 501 bp fragment of the *fdhA* gene in C. ljungdahlii
repH_643bp_rv	GCACTGTTATGCCTTTTGACTATCAC	Primer binding in the backbone of pMTL83151 and pMTL83152 upstream of *repH*
traJ_60bp_fw	CATGCGCTCCATCAAGAAGAG	Primer binding in the backbone of pMTL83151 and pMTL83152 downstream of *traJ*
rnfC_250bp_rv	CTCCTATATCTACAACCTTTCCAGAAGTAG	Primer binding 250 bp upstream of *rnfC*, which was used for Sanger sequencing and PCR screening
cas12a_fwd_BamHI	GGTACCGGATCCATGTCAATTTATCAAGAATTTGTTAATA	Amplification of *Fncas12a*
cas12a_rv_NcoI	GGTACCCCATGGTTAGTTATTCCTATTCTGCAC	Amplification of *Fncas12a*
Seq1_cas12a	CACAGATATAGATGAGGCG	Sanger sequencing of *Fncas12a*
Seq2_cas12a	GCTTCTGGAGCTTTGTCT	Sanger sequencing of *Fncas12a*
Seq3_cas12a	GTAGTTACAACGATGCAAAG	Sanger sequencing of *Fncas12a*
Seq4_cas12a	CCGCTGTACCAATAACAC	Sanger sequencing of *Fncas12a*
Seq5_cas12a	GGCTAATGGTTGGGATAA	Sanger sequencing of *Fncas12a*
Seq6_cas12a	CTTATTCATCACACCCAG	Sanger sequencing of *Fncas12a*

Continued on next page

Table D.2 – continued from previous page

Primer	Sequence	Function
Seq7_cas12a	CAAGATGTGGTTTATAAGC	Sanger sequencing of *Fncas12a*
Seq8_cas12a	CCTCTTTAGCTGGGTGAGTG	Sanger sequencing of *Fncas12a*
Seq9_cas12a	CAAGGTAGAGAAGCAGGTC	Sanger sequencing of *Fncas12a*
Seq10_cas12a	GCTCTAAGCACTCCCCCAG	Sanger sequencing of *Fncas12a*
Seq11_cas12a	GCTAAGCTAACTAGTGTC	Sanger sequencing of *Fncas12a*
Seq12_cas12a	CCATTTACATCTGCTACTGG	Sanger sequencing of *Fncas12a*
HDR_rnfB_fwdOv	TGTAAAAATTATTGAAAGAG GTGTTTAAGATGGCAGTGGA GCAAAGCTT	Amplification of homology-directed repair arm downstream of *rnfB* with overhang to the homology-directed repair arm upstream of *rnfC*
HDR_rnfB_rv	ATGTAAAGGGTTCACATAAA ATAGCTGT	Amplification of homology-directed repair arm downstream of *rnfB*
HDR_rnfC_fwdOv	CAAGTTGAAAAATTTAATAA AAAAATAAGTGGCTTGAAAT CAATAGTTAACGCAATAG	Amplification of homology-directed repair arm upstream of rnfC with overhang to the *Fncas12a* sequence
HDR_rnfC_fwd	GGCTTGAAATCAATAGTTAA CGCAATAG	Amplification of homology-directed repair arm upstream of *rnfC* without overhang
HDR_rnfC_rvOV	TCAGCAAATTTAAGCTTTGC TCCACTGCCATCTTAAACAC CTCTTTCAATAATTTTTACAG C	Amplification of homology-directed repair arm upstream of rnfC with overhang to the homology-directed repair arm downstream of *rnfB*
Seq_HDR_rnfB_881bp_fwd	GACCTGGTTCGGATATCCAT CC	Sanger sequencing of the HDR_*rnfB* fragment

Continued on next page

Table D.2 – continued from previous page

Primer	Sequence	Function
minigene_crRNA_RNF_fwd	TTTATGTGAACCCTTTACATTTGACAAATT	Amplification of crRNA array consisting of 22-bp overhang to HDR_*rnfB* , p4-promoter, direct repeats, sgRNA (TTA), and rrnB-T1 terminator for genome target *rnfCDGEAB*
minigene_crRNA_all_rv	GTTGGTAGCTTAATATATAAGAATAAAACGAAAGG	Amplification of crRNA array consisting of 22-bp overhang to pMTL83152-Cas12a, p4-promoter, direct repeats, sgRNA (TTA), and rrnB-T1 terminator for genome target *rnfCDGEAB*, *rseC*, and *nar*
outside_RNF_HDR_dwst_rv	GCATGGGAGTGTTAATATGAAAAAAGGG	Verification of *rnfCDGEAB* deletion
outside_RNF_HDR_upst_fwd	GGAGGCTATTAAGGGACCGT	Verification of *rnfCDGEAB* deletion
HDR_rseC_dwst_fwdOv	CGCTAACAAATAATAGGAGGTGTATTATGTAATTTGTGTACAAACTTTAATTAATGGAGAGAC	Amplification of a homology-directed repair arm downstream of *rseC* with 28-bp overlap to HDR_*rseC*_upst
HDR_rseC_dwst_rv	TAGTTGTAACCCTCTGTATAAGTGGAATTC	Amplification of a homology-directed repair arm downstream of *rseC*
HDR_rseC_upst_fwd	CTCATTGAAGTATATGTTAATGGCAGAAAAAAAGTTC	Amplification of a homology-directed repair arm upstream of *rseC*
HDR_rseC_upst_fwdOv	CAAGTTGAAAAATTTAATAAAAAAATAAGTCTCATTGAAGTATATGTTAATGGCAGAAAAAAAGTTC	Amplification of a homology-directed repair arm upstream of *rseC* with 30-bp overlap to pMTL83152-Cas12a
HDR_rseC_upst_rvOv	TCCATTAATTAAAGTTTGTACACAAATTACATAATACACCTCCTATTATTTGTTAGCGTTTTC	Amplification of a homology-directed repair arm upstream of *rseC* with 30-bp overlap to fragment HDR_*rseC*_dwst

Continued on next page

Table D.2 – continued from previous page

Primer	Sequence	Function
minigene_crRNA_rseC_fwd	CTTATACAGAGGGTTACAACTATTGACAAATT	Amplification of crRNA array consisting of 22-bp overhang to HDR_*rseC*_dwst, p4-promoter, direct repeats, sgRNA (TTA), and rrnB-T1 terminator for genome target *rseC*
outside_rseC_HDRdwst_rv	CCCATCATAGGTCCACCTGAAA	Verification of *rseC* deletion
outside_rseC_HDRupst_fwd	CGAGCTGAAGGTTGTAAAAATATCCG	Verification of *rseC* deletion
seq_rseC_145bpupst_fwd	GAAGGTAATACTGTTCAATATCGATACAGA	Verification of *rseC* deletion
HDR_nar_dwst_fwdOv	TCTTTTTCATAAATTTAGAGTATACTTTCTCCACTTCTCAATATTTTTTACTGAAAATAC	Amplification of a homology-directed repair arm downstream of *nar* with overhang to HDR_*nar*_upst
HDR_nar_dwst_rv	TTGGAATGACAGGACTCTATATAGTTATGG	Amplification of a homology-directed repair arm downstream
HDR_nar_upst_fwd	TACAACCTCTGTTAGTACTGCTGATATTACATC	Amplification of a homology-directed repair arm upstream of *nar*
HDR_nar_upst_fwdOv	CAAGTTGAAAAATTTAATAAAAAAATAAGTTACAACCTCTGTTAGTACTGCTGATATTACATC	Amplification of a homology-directed repair arm upstream of *nar* with overhang to *Fncas12a*
HDR_nar_upst_rvOv	GTATTTTCAGTAAAAAATATTGAGAAGTGGAGAAAGTATACTCTAAATTTATGAAAAAGAATTTTA	Amplification of a homology-directed repair arm upstream of *nar* with overhang to HDR_*nar*_dwst

Continued on next page

Table D.2 – continued from previous page

Primer	Sequence	Function
minigene_nar_fwd	ATATAGAGTCCTGTCATTCCAATTGACAAATT	Amplification of crRNA array consisting of 22-bp overhang to HDR_*nar*_dwst, p4-promoter, direct repeats, sgRNA (TTA), and rrnB-T1 terminator for genome target *nar*
seq_nar_95bp_dwst_fwd	CCGGATAACCTTTAGTGGGAAGT	Verification of *nar* deletion
seq_nar_132bp_upst_rv	GCGCCATAATTCAAGGGGAT	Verification of *nar* deletion
outside_nar_HDRdwst_rv	GGGTTGACGTAGATGGAGGAAG	Verification of *nar* deletion
outside_nar_HDRupst_fwd	CCTTTAAGCTTCCACCATTTGCC	Verification of *nar* deletion
qPCR_rseC_fwd	GCTAGTAGACACGGAGATTG	Amplification of a 142 bp fragment from *rseC*
qPCR_rseC_rv	CTGCCCATAACATATTTGC	Amplification of a 142 bp fragment from *rseC*
qPCR_rnfC_fwd	GCACCTATACCAGATAAGGT	Amplification of a 160 bp fragment from *rnfC*
qPCR_rnfC_rv	CCTTTCCAGAAGTAGATGCAT	Amplification of a 160 bp fragment from *rnfC*
qPCR_rnfD_fwd	CCTCATGTTCGTTGTGATG	Amplification of a 157 bp fragment from *rnfD*
qPCR_rnfD_rv	CAAAGTACTCCGTAACTACAGC	Amplification of a 157 bp fragment from *rnfD*
qPCR_rnfG_fwd	CATCACCAGTAGCAGCG	Amplification of a 156 bp fragment from *rnfG*
qPCR_rnfG_rv	CTGCAGGTACAACATATGC	Amplification of a 156 bp fragment from *rnfG*
qPCR_rnfE_fwd	TGTGTCCAGCACTGGC	Amplification of a 138 bp fragment from *rnfE*
qPCR_rnfE_rv	CAGGGACACGTACCTTAG	Amplification of a 138 bp fragment from *rnfE*

Continued on next page

Table D.2 – continued from previous page

Primer	Sequence	Function
qPCR_rnfA_fwd	GCATCTGTAGGTATGGGTATG	Amplification of a 136 bp fragment from rnfA
qPCR_rnfA_rv	CAATAAGAAGTACAAAAACTACCG	Amplification of a 136 bp fragment from rnfA
qPCR_rnfB_fwd	GCAATGGAAGTGAATCCAC	Amplification of a 155 bp fragment from *rnfB*
qPCR_rnfB_rv	GCTGCTTTTCCAGGTAC	Amplification of a 155 bp fragment from *rnfB*
qPCR_rho_fwd	GGACTCTTTCAGGAGGACTA	Amplification of a 243 bp fragment from *rho*
qPCR_rho_rv	ATACATCTATGGCAGGGAAT	Amplification of a 243 bp fragment from *rho*

Table D.3 Synthesized mini genes that contain crRNA arrays for this study. Gene synthesis was performed by IDT (Integrated DNA Technologies). Each mini gene contains 20-22-bp overhang to the pMTL-backbone and to the fused homology-directed repair arms. Directed-repeats sequence of 20 bp (underlined). sgRNA with TTV PAM for the RNF complex gene cluster deletion and with TTTV PAM for the *nar* and *rseC* deletion (bold). Two sgRNAs were used to target RNF and *nar*.

Name	Sequence (3'→ 5')
minigene_crRNA-RNF	TTTATGTGAACCCTTTACATTTGACAAATTTATTTTTTAAAGTTAAAATTAAGTTGTAATTTCTACTGTTGTAGAT**AAAAGTTTTCGAGGTGGAGTACA**TAATTTCTACTGTTGTAGAT**CAACAGCAGAGCAAGAATGAAGC**ATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATTCTTATATATTAAGCTACCAAC
minigene_crRNA-rseC	CTTATACAGAGGGTTACAACTATTGACAAATTTATTTTTTAAAGTTAAAATTAAGTTGTAATTTCTACTGTTGTAGAT**ATAGATCTACAAGCAAAAATGAG**ATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATTCTTATATATTAAGCTACCAAC
minigene_crRNA-nar	ATATAGAGTCCTGTCATTCCAATTGACAAATTTATTTTTTAAAGTTAAAATTAAGTTGTAATTTCTACTGTTGTAGAT**TATTTCTTGTTTATAGCTTTCAT**TAATTTCTACTGTTGTAGAT**TACAGCAAAATCCATCATTTACC**ATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATTCTTATATATTAAGCTACCAAC

Table D.4 Used and generated *C. ljungdahlii* strains in this study.

Clostridial strain	Plasmid	Phenotype	
		Heterotrophic	Autotrophic
C. ljungdahlii DSM13528	-	yes	yes
C. ljungdahlii DSM13528	pMTL83151	yes	yes
C. ljungdahlii DSM13528	pMTL83151_PtetR-O1	yes	yes
C. ljungdahlii DSM13528	pMTL83152	yes	yes
C. ljungdahlii DSM13528	pMTL83151_P_{nat}_*rnfCDGEAB*	yes	yes
C. ljungdahlii DSM13528	pMTL83152_*rseC*	yes	yes
C. ljungdahlii DSM13528	pMTL83151_*nar*	yes	yes
C. ljungdahlii ΔRNF[a]	-	yes (reduced)	no
C. ljungdahlii ΔRNF[a]	pMTL83151	yes (reduced)	no
C. ljungdahlii ΔRNF[a]	pMTL83151_P_{nat}_*rnfCDGEAB*	yes	yes
C. ljungdahlii Δ*rseC*	-	yes	no
C. ljungdahlii Δ*rseC*	pMTL83152	yes	no
C. ljungdahlii Δ*rseC*	pMTL83152 _*rseC*	yes	yes
C. ljungdahlii Δ*nar*	-	yes	yes
C. ljungdahlii Δ*nar*	pMTL83152	yes	yes
C. ljungdahlii Δ*nar*	pMTL83152_*nar*	yes	yes

[a] ΔRNF = Δ*rnfCDGEAB*

Table D.5 Performance of all tested *C. ljungdahlii* strains in heterotrophic batch cultivation experiments. Cultures were grown with fructose (5 g/L) in PETC medium, which contained either ammonium or nitrate as nitrogen source. All growth experiments were performed under anaerobic conditions. Data is given as mean values ± standard deviation from biological triplicates. WT, *C. ljungdahlii* wild type; ΔRNF, *C. ljungdahlii* with deleted *rnfCDGEAB* gene cluster; Δ*rseC*, *C. ljungdahlii* with deleted *rseC* gene; and Δ*nar*, *C. ljungdahlii* with deleted nitrate reductase gene cluster. Given in percentage is the difference in performance in comparison to the wild type with the same nitrogen source.

Strain	Nitrogen source	Growth rate (μ in h)[a]	Maximum OD_{600} value	Maximum acetate concentration (mM)	Maximum ethanol concentration (mM)
WT	ammonium	0.079±0.002	2.49±0.03	52.3±0.7	10.6±0.1
WT	nitrate	0.073±0.002	2.24±0.10	43.6±0.6	5.0±0.1
ΔRNF	ammonium	0.052±0.003 (-34%, ***)	1.16±0.03 (-53%, ***)	35.4±0.5 (-32%, ***)	6.3±0.8 (-32%, n.s.[c])
ΔRNF	nitrate	0.042±0.003 (-42%, ***)	0.98±0.10 (-56%, ***)	25.4±1.7 (-42%, ***)	n.d.[b]
ΔrseC	ammonium	0.084±0.002 (6%, n.s.[c])	1.90±0.15 (-31%, ***)	50.1±0.3 (-4%, n.s.[c])	7.5±0.1 (-29%, ***)
ΔrseC	nitrate	0.048±0.002 (-34%, ***)	1.58±0.03 (-30%, ***)	50.9±1.7 (-3%, n.s.[c])	2.9±0.1 (-42%, ***)
Δnar	ammonium	0.071±0.002 (-11%, **)	2.35±0.04 (-6%, n.s.[c])	51.9±0.9 (-1%, n.s.[c])	15.3±0.1 (+44%, ***)
Δnar	nitrate	0.067±0.001 (-9%, n.s.[c])	1.51±0.03 (-32%, ***)	28.7±1.1 (-34%, ***)	16.6±0.2 (+234%, ***)

[a] μ values were calculated based on the individual OD_{600} values of each triplicate in the exponential growth phase.
[b] n.d., not detectable
[c] n.s., not significant ($P > 0.05$)
[*] significant ($P \leq 0.05$)
[**] significant ($P \leq 0.01$)
[***] significant ($P \leq 0.001$)

Table D.6 RseC peptide sequences and amount of predicted transmembrane helices. Putative transmembrane helices were predicted with the TMHMM-2.0 tool (`https://services.healthtech.dtu.dk/service.php?TMHMM-2.0`) based on the Rsec peptide sequence.

Microbe	RseC peptide sequence	Predicted transmembrane helices
C. ljungdahlii	MKRESEGIVIETSESIAKVRASRHGDCKSCGACPGDN AIVVDAKNPVGAKPGQHVVFEIKDANMLWAAFIVY ILPLIGILIGALIGTWIGGKLGHSLREFQIGGGVLFFILS LIYIKIFDRSTSKNESKKPVITKILY	2
C. autoethanogenum	MKRESEGIVIETSESIAKVRASRHGDCKSCGACPGDN AIVVDAKNPVGAKPGQHVVFEIKDANMLWAAFIVY ILPLIGILIGALIGTWIGGKLGHSLREFQIGGGVLFFILS LIYIKIFDRSTSKNESKKPVITKILY	2
C. carboxidovorans	MNRETEGIVIQIEGNIAKIKANRHGDCSNCGACPGDK AMVVDAINTIGAKPGQHVSFEIKEVNMLKAAFVVYI LPLVSIFIGAVIGGFVAKKIAQDSVMCSVIGGIVLFILS IIYIKFFDKAANKDENMKPIITRILS	2
C. kluyveri	MKKESEGIVIETTEGFARVKASRHGDCKNCGACPGD NATVLDAKNPIGAKAGEHVILEMREQNMIRAAFVV YIMPIISIFLGVLVGTWIFNAVGYYEMAFKVVGGIVF FVISLVYIKVFDKATAKNDASKPVIKKVL	2
E. limosum	MKEIGIVEELKGKNAKVLIKRHAACGDCGACQVGK EKMTMEATARNAAGAQVGDTVSVEMEFANVIKATS IMYGIPLIAFVVGCAAGYFAAVALTLDLVLVPFFTGI LLTVISYLVIRVFDKKGKFNSKYEPVITEIEAEAQELP PAGE	2
A. woodii	MKEIGTVKALKGKNAEIEIKRNTACGDCGACHVSKD QSVMLTTANNPIKAKIGETVEVEMEFANVFVAAFIM YGIPLVAFVLGSSGVYFLVGALNIGWDQVVSSFLAGI CLTAVAYVVIRKLDRKGRFNSKYQPIVTAIIEKKETIK TPMESRMGH	2
R. capsulatus	MTGCCDDGPATGPRDLRERLRVVAVRGESLVVAAD RASACAACAEAKGCGTRALMSMHRTDLMTIARPAG LIVAPGDEVEVAMSGNNLLAGAGLAYLLPALAFVV ALALASGAGLSDGGAALVGGVVLMFSFLPLVLLERR ARLSRALQVLDVHPGHGR	2
E. coli	MIKEWATVVSWQNGQALVSCDVKASCSSCASRAGC GSRVLNKLGPQTTHTIVVPCDEPLVPGQKVELGIAEG SLLSSALLVYMSPLVGLFLIASLFQLLFASDVAALCG AILGGIGGFLIARGYSRKFAARAEWQPIILSVALPPGL VRFETSSEDASQ	2

Table D.7 Distribution of potential *rseC* genes in genomes of acetogens. Given the availability of full genomes in NCBI, 47 of the 61 acetogenic bacteria listed in Table 6.2 from Bengelsdorf *et al.* (2018) were considered in this analysis. Using the NCBI Datasets API, ncbi-datasets-pylib 12.15.0, the GenBank files of these 47 acetogens were retrieved and used to generate protein FASTA files. Using BLASTp (BioPython 1.79), the *rseC* and RNF-gene cluster from *Clostridium ljungdahlii* DSM 13528 (CLJU_c11350- CLJU_c11410), and four subunits from the Ech-gene cluster, *ech2A1, ech2A2, ech2B, echE2,* from *Thermoanaerobacter kivui* (TKV_c19720, TKV_c19710, TKV_c19690, TKV_C19740) were used as the queries to the 47 genomes. An expect value (E-value) of $10e^{-10}$ was used as a threshold for finding potential gene matches (`https://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/650/_E_value.html`). Y, yes, present; N, no, not present; ND, no detected; and ID, identity.

Acetogenic bacterium[a]	**Assembly ID**	***rseC* gene ID**[b]	***rnfC* gene ID**	***rnf*-gene cluster (ABCDEG)**[c]	***rseC* gene flanking *rnfC* gene**[d]	***ech2A1* gene ID**[e]	***ech2A2* gene ID**[e]	***ech2B* gene ID**[e]	***echE2* gene ID**[e]
Acetitomaculum ruminis DSM 5522	GCF_900112085.1	---	['BM153_RS11735']	['Y']	---	---	---	---	---
Acetoanaerobium noterae	GCF_900168025.1	['B5X47_RS10270','B5X47_RS12780']	['B5X47_RS09760']	['Y']	[['N'],['N']]	['B5X47_RS08535']	['B5X47_RS08535']	---	---
Acetobacterium bakii	GCF_900235925.1	['DXY11_RS15255']	['DXY11_RS15080']	['Y']	[['N']]	['DXY11_RS11030','DXY11_RS11045']	['DXY11_RS11045','DXY11_RS11030']	---	['DXY11_RS11050']

Continued on next page

Table D.7 – continued from previous page

Acetogenic bacterium[a]	Assembly ID	*rseC* gene ID[b]	*rnfC* gene ID	*rnf*-gene cluster (ABCDEG)[c]	*rseC* gene flanking *rnfC* gene[d]	*ech2A1* gene ID[e]	*ech2A2* gene ID[e]	*ech2B* gene ID[e]	*echE2* gene ID[e]
Acetobacterium dehalogenans DSM 11527	GCF_000472665.1	['A3KS_RS0106250','A3KS_RS0116425','A3KS_RS0114380']	['A3KS_RS0106245','A3KS_RS0114035']	['Y','Y']	[['Y','N'],['N','N'],['N','N']]	['A3KS_RS0109555']	['A3KS_RS0109555']	---	---
Acetobacterium fimetarium	GCF_014284475.1	['GH808_RS08595','GH808_RS03350']	['GH808_RS08600','GH808_RS10855']	['Y','Y']	[['Y','N'],['N','N']]	['GH808_RS06185','GH808_RS06170']	['GH808_RS06185','GH808_RS06170']	['GH808_RS06180']	['GH808_RS06165']
Acetobacterium malicum	GCA_014284495.1	['GH811_03365','GH811_03590','GH811_03750','GH811_03865','GH811_01580']	['GH811_03360','GH811_01420']	['Y','Y']	[['Y','N'],['N','N'],['N','N'],['N','N'],['N','N']]	['GH811_14760']	['GH811_14760']	---	---

Continued on next page

Table D.7 – continued from previous page

Acetogenic bacterium[a]	Assembly ID	*rseC* gene ID[b]	*rnfC* gene ID	*rnf*-gene cluster (ABCDEG)[c]	*rseC* gene flanking *rnfC* gene[d]	*ech2A1* gene ID[e]	*ech2A2* gene ID[e]	*ech2B* gene ID[e]	*echE2* gene ID[e]
Acetobacterium paludosum	GCF_008086595.1	['FZC41_RS05030']	['FZC41_RS05025','FZC41_RS08725']	['Y','Y']	[['Y','N']]	---	---	---	---
Acetobacterium tundrae	GCF_008086615.1	['FZC40_RS07445','FZC40_RS13050']	['FZC40_RS07440','FZC40_RS09950']	['Y','Y']	[['Y','N'],['N','N']]	---	---	---	---
Acetobacterium wieringae	GCF_008107585.1	['FXB42_RS13260','FXB42_RS01650']	['FXB42_RS13255','FXB42_RS01495']	['Y','Y']	[['Y','N'],['N','N']]	['FXB42_RS10035']	['FXB42_RS10035']	---	---
Acetobacterium woodii DSM 1030	GCA_000247605.1	['Awo_c21740']	['Awo_c22060']	['Y']	[['N']]	---	---	---	---
Acetohalobium arabaticum DSM 5501	GCA_000144695.1	['Acear_0297']	['Acear_0533','Acear_0362']	['Y','Y']	[['N','N']]	['Acear_1098','Acear_1099','Acear_1097']	['Acear_1099','Acear_1097','Acear_1098']	---	---

Continued on next page

Table D.7 – continued from previous page

Acetogenic bacterium[a]	**Assembly ID**	***rseC* gene ID[b]**	***rnfC* gene ID**	***rnf*-gene cluster (ABCDEG)[c]**	***rseC* gene flanking *rnfC* gene[d]**	***ech2A1* gene ID[e]**	***ech2A2* gene ID[e]**	***ech2B* gene ID[e]**	***echE2* gene ID[e]**
Acetonema longum DSM 6540	GCF_000219125.1	---	---	['ND','*']	---	['ALO_RS00265','ALO_RS08220','ALO_RS08205','ALO_RS00270','ALO_RS00275']	['ALO_RS00275','ALO_RS08220','ALO_RS00270','ALO_RS08205','ALO_RS00265']	['ALO_RS00245','ALO_RS08215']	['ALO_RS00240','ALO_RS08200','ALO_RS09345']
Alkalibaculum bacchi	GCF_003317055.1	['DES36_RS03130']	['DES36_RS00630']	['Y']	[['N']]	---	---	---	---
Blautia hydrogenotrophica DSM 10507	GCF_001404935.1	---	['ARA85_RS00980']	['Y']	---	---	---	---	---
Blautia schinkii	GCF_013304825.1	---	['HFM85_RS08490']	['Y']	---	---	---	---	---

Continued on next page

Table D.7 – continued from previous page

Acetogenic bacterium[a]	Assembly ID	*rseC* gene ID[b]	*rnfC* gene ID	*rnf*-gene cluster (ABCDEG)[c]	*rseC* gene flanking *rnfC* gene[d]	*ech2A1* gene ID[e]	*ech2A2* gene ID[e]	*ech2B* gene ID[e]	*echE2* gene ID[e]
Calderihabitans maritimus	GCF_002207765.1	---	---	['ND','*']	---	['KKC1_RS08925','KKC1_RS14735','KKC1_RS06640','KKC1_RS06640','KKC1_RS01160','KKC1_RS08930','KKC1_RS01175','KKC1_RS08935','KKC1_RS14740','KKC1_RS14745']	['KKC1_RS06640','KKC1_RS06640','KKC1_RS08935','KKC1_RS01175','KKC1_RS14735','KKC1_RS14745','KKC1_RS08930','KKC1_RS01160']	['KKC1_RS06645','KKC1_RS08905','KKC1_RS14720','KKC1_RS01165']	['KKC1_RS06665','KKC1_RS01180','KKC1_RS08900','KKC1_RS14710','KKC1_RS10630']

Continued on next page

Table D.7 – continued from previous page

Acetogenic bacterium[a]	**Assembly ID**	***rseC* gene ID[b]**	***rnfC* gene ID**	***rnf*-gene cluster (ABCDEG)[c]**	***rseC* gene flanking *rnfC* gene[d]**	***ech2A1* gene ID[e]**	***ech2A2* gene ID[e]**	***ech2B* gene ID[e]**	***echE2* gene ID[e]**
Carboxydothermus ferrireducens DSM 11255	GCF_000427565.1	---	---	['ND','*']	---	['CARFE_RS0112825','CARFE_RS0112830','CARFE_RS0112835']	['CARFE_RS0112835','CARFE_RS0112825','CARFE_RS0112830']	['CARFE_RS0112805']	['CARFE_RS0112800']
Carboxydothermus hydrogenoformans	GCA_000012865.1	---	---	['ND','*']	---	['CHY_1417','CHY_1832','CHY_1832','CHY_1832','CHY_1416','CHY_1415']	['CHY_1415','CHY_1832','CHY_1832','CHY_1417','CHY_1416']	['CHY_1831','CHY_1421']	['CHY_1827','CHY_1422']
Carboxydothermus pertinax	GCF_001950255.1	---	---	['ND','*']	---	['cpu_RS11170','cpu_RS01910','cpu_RS01910','cpu_RS01910','cpu_RS11165','cpu_RS11160']	['cpu_RS11160','cpu_RS11170','cpu_RS01910','cpu_RS01910']	['cpu_RS01915','cpu_RS11190']	['cpu_RS01935','cpu_RS11195']

Continued on next page

Table D.7 – continued from previous page

Acetogenic bacterium[a]	Assembly ID	*rseC* gene ID[b]	*rnfC* gene ID	*rnf*-gene cluster (ABCDEG)[c]	*rseC* gene flanking *rnfC* gene[d]	*ech2A1* gene ID[e]	*ech2A2* gene ID[e]	*ech2B* gene ID[e]	*echE2* gene ID[e]
Clostridium aceticum	GCA_001042715.1	['CACET_c17430']	['CACET_c16320']	['Y']	[['N']]	['CACET_c33260','CACET_c33270','CACET_c33250','CACET_c29680']	['CACET_c33250','CACET_c33260','CACET_c33270','CACET_c29680']	---	---
Clostridium autoethanogenum DSM 10061	GCA_000484505.1	['CAETHG_3226']	['CAETHG_3227']	['Y']	[['Y']]	---	---	---	---
Clostridium carboxidivorans	GCA_001038625.1	['Ccar_25730','Ccar_07835']	['Ccar_25735']	['Y']	[['Y'],['N']]	['Ccar_06750','Ccar_06735']	['Ccar_06750','Ccar_06735']	---	['Ccar_06755']
Clostridium coskatii	GCA_001675205.1	['CLCOS_05690']	['CLCOS_05700']	['Y']	[['Y']]	---	---	---	---
Clostridium drakei	GCA_003096175.1	['B9W14_04695','B9W14_12340']	['B9W14_04700']	['Y']	[['Y'],['N']]	['B9W14_11370','B9W14_11385']	['B9W14_11385','B9W14_11370']	---	['B9W14_11390']

Continued on next page

Table D.7 – continued from previous page

Acetogenic bacterium[a]	Assembly ID	*rseC* gene ID[b]	*rnfC* gene ID	*rnf*-gene cluster (ABCDEG)[c]	*rseC* gene flanking *rnfC* gene[d]	*ech2A1* gene ID[e]	*ech2A2* gene ID[e]	*ech2B* gene ID[e]	*echE2* gene ID[e]
Clostridium formicaceticum	GCA_002080475.1	['CLFO_18400','CLFO_12270']	['CLFO_17180']	['Y']	[['N'],['N']]	['CLFO_36480','CLFO_36490','CLFO_36470']	['CLFO_36470','CLFO_36480','CLFO_36490']	---	---
Clostridium kluyveri DSM 555	GCA_000016505.1	['CKL_1263','CKL_2767']	['CKL_1264']	['Y']	[['Y'],['N']]	---	---	---	---
Clostridium ljungdahlii DSM 13528	GCA_000143685.1	['CLJU_c11350']	['CLJU_c11360']	['Y']	[['Y']]	---	---	---	---
Clostridium magnum DSM 2767	GCF_900129955.1	['BUC18_RS08450','BUC18_RS02475']	['BUC18_RS08445','BUC18_RS02470']	['Y','Y']	[['Y','N'],['N','Y']]	['BUC18_RS02805','BUC18_RS02790']	['BUC18_RS02790']	['BUC18_RS02800']	['BUC18_RS02785']
Clostridium ragsdalei P11	GCF_001675165.1	['CLRAG_RS05955']	['CLRAG_RS05950']	['Y']	[['Y']]	---	---	---	---
Clostridium scatologenes	GCA_000968375.1	['CSCA_2973','CSCA_1298']	['CSCA_2972']	['Y']	[['Y'],['N']]	['CSCA_1581','CSCA_1578']	['CSCA_1578','CSCA_1581']	---	['CSCA_1577']
Eubacterium aggregans	GCF_900107815.1	['BLW33_RS05705']	['BLW33_RS04770']	['Y']	[['N']]	---	---	---	---

Continued on next page

Table D.7 – continued from previous page

Acetogenic bacterium[a]	Assembly ID	*rseC* gene ID[b]	*rnfC* gene ID	*rnf*-gene cluster (ABCDEG)[c]	*rseC* gene flanking *rnfC* gene[d]	*ech2A1* gene ID[e]	*ech2A2* gene ID[e]	*ech2B* gene ID[e]	*echE2* gene ID[e]
Eubacterium limosum	GCA_000807675.2	['B2M23_08890']	['B2M23_19790']	['Y']	[['N']]	---	---	---	---
Marvinbryantia formatexigens DSM 14469	GCF_900102475.1	---	['BLR58_RS17940']	['Y']	---	['BLR58_RS16225','BLR58_RS16230','BLR58_RS16220','BLR58_RS16215']	['BLR58_RS16215','BLR58_RS16220','BLR58_RS16230']	---	---
Moorella mulderi DSM 14980	GCF_001594015.1	---	---	['ND','*']	---	['MOMUL_RS08150','MOMUL_RS08155','MOMUL_RS08160']	['MOMUL_RS08160','MOMUL_RS08155','MOMUL_RS08150']	['MOMUL_RS08130']	['MOMUL_RS08125']

Continued on next page

Table D.7 – continued from previous page

Acetogenic bacterium[a]	Assembly ID	*rseC* gene ID[b]	*rnfC* gene ID	*rnf*-gene cluster (ABCDEG)[c]	*rseC* gene flanking *rnfC* gene[d]	*ech2A1* gene ID[e]	*ech2A2* gene ID[e]	*ech2B* gene ID[e]	*echE2* gene ID[e]
Moorella thermoacetica	GCF_001874605.1	---	---	['ND','*']	---	['MTJW_RS04625','MTJW_RS11345','MTJW_RS11365','MTJW_RS04630','MTJW_RS11350','MTJW_RS04635']	['MTJW_RS04635','MTJW_RS11365','MTJW_RS11350','MTJW_RS04630','MTJW_RS11345','MTJW_RS04625']	['MTJW_RS04605','MTJW_RS11360']	['MTJW_RS11340','MTJW_RS04600']
Oxobacter pfennigii	GCF_001317355.1	---	['OXPF_RS13305']	['Y']	---	['OXPF_RS01035','OXPF_RS20260','OXPF_RS20250','OXPF_RS20255','OXPF_RS01030','OXPF_RS01025']	['OXPF_RS20260','OXPF_RS01025','OXPF_RS20255','OXPF_RS20250','OXPF_RS01035','OXPF_RS01030']	['OXPF_RS01055']	['OXPF_RS01060']

Continued on next page

Table D.7 – continued from previous page

Acetogenic bacterium[a]	Assembly ID	*rseC* gene ID[b]	*rnfC* gene ID	*rnf*-gene cluster (ABCDEG)[c]	*rseC* gene flanking *rnfC* gene[d]	*ech2A1* gene ID[e]	*ech2A2* gene ID[e]	*ech2B* gene ID[e]	*echE2* gene ID[e]
Rhodobacter capsulatus SB 1003	GCF_000021865.1	---	['RCAP_RS16250']	['Y','*']	---	['RCAP_RS10560','RCAP_RS07605','RCAP_RS07610','RCAP_RS07615']	['RCAP_RS10560','RCAP_RS07610','RCAP_RS07615']	['RCAP_RS07575']	['RCAP_RS07530']
Sporomusa acidovorans DSM 3132	GCF_900101845.1	['BLR65_RS14630','BLR65_RS04290','BLR65_RS03790']	['BLR65_RS14625']	['Y']	[['Y'],['N'],['N']]	['BLR65_RS15875','BLR65_RS15880']	---	---	---
Sporomusa malonica	GCF_900176355.1	['B9A22_RS13135','B9A22_RS03705']	['B9A22_RS13130']	['Y']	[['Y'],['N']]	['B9A22_RS12830','B9A22_RS12815']	['B9A22_RS12830','B9A22_RS12815']	---	['B9A22_RS12810']
Sporomusa ovata DSM 2662	GCF_000445445.1	['SOV_RS03895']	['SOV_RS03900']	['Y']	[['Y']]	['SOV_RS04180','SOV_RS04195']	---	---	['SOV_RS04200']

Continued on next page

Table D.7 – continued from previous page

Acetogenic bacterium[a]	**Assembly ID**	***rseC* gene ID[b]**	***rnfC* gene ID**	***rnf*-gene cluster (ABCDEG)[c]**	***rseC* gene flanking *rnfC* gene[d]**	***ech2A1* gene ID[e]**	***ech2A2* gene ID[e]**	***ech2B* gene ID[e]**	***echE2* gene ID[e]**
Sporomusa silvacetica DSM 10669	GCF_002257705.1	['SPSIL_RS12000','SPSIL_RS09415']	['SPSIL_RS11995']	['Y']	[['Y'],['N']]	['SPSIL_RS11675','SPSIL_RS11660']	['SPSIL_RS11675']	---	['SPSIL_RS11655']
Sporomusa sphaeroides DSM 2875	GCF_900042765.1	['SSPH_RS15205']	['SSPH_RS15210']	['Y']	[['Y']]	['SSPH_RS02485','SSPH_RS02470']	['SSPH_RS02470']	---	['SSPH_RS02465']
Sporomusa termitida	GCA_007641255.1	['SPTER_14240']	['SPTER_14250']	['Y']	[['Y']]	['SPTER_39500','SPTER_39470']	['SPTER_39470']	---	['SPTER_39460']
Terrisporobacter mayombei	GCF_014333445.1	---	['H9L25_RS04125']	['Y']	---	---	---	---	---
Thermacetogenium phaeum DSM 12270	GCA_000305935.1	---	---	['ND','*']	---	['Tph_c26280','Tph_c21360','Tph_c26310']	['Tph_c26280','Tph_c26310','Tph_c21360']	['Tph_c26290','Tph_c21350']	['Tph_c26330','Tph_c21320']

Continued on next page

Table D.7 – continued from previous page

Acetogenic bacterium[a]	Assembly ID	*rseC* gene ID[b]	*rnfC* gene ID	*rnf*-gene cluster (ABCDEG)[c]	*rseC* gene flanking *rnfC* gene[d]	*ech2A1* gene ID[e]	*ech2A2* gene ID[e]	*ech2B* gene ID[e]	*echE2* gene ID[e]
Thermoanaerobacter kivui	GCA_000763575.1	---	---	['ND','*']	---	['TKV_c19720','TKV_c01230','TKV_c19710']	['TKV_c19710','TKV_c01230','TKV_c19720']	['TKV_c19690','TKV_c01240']	['TKV_c19740','TKV_c01310']
Treponema primitia ZAS-2	GCF_000214375.1	---	['TREPR_RS17540','TREPR_RS10980']	['Y','Y']	---	---	---	---	---

[a] All acetogens with available full genome sequences were selected from Table 2 in Bengelsdorf *et al.* (2018).

[b] Potential *rseC* genes were screened in each genome by protein sequence comparison to the RseC of *C. ljungdahlii* (CLJU_C11350).

[c] The presence of (potential) RNF-gene clusters was defined by potential RnfC and RnfD protein sequence within +/- 500 bp of each other. The RnfC (CLJU_c11360) and RnfD (CLJU_c11370) protein sequences of *C. ljungdahlii* were used as reference sequences.

[d] Potential *rseC* genes were screened for being located +/- 500 bp of the identified *rnfC* gene.

[e] The presence of (potential) Ech-gene clusters was screened using the protein sequences of Ech2A1, Ech2A2, Ech2B, EchE2, from *T. kivui* (TKV_c19720, TKV_c19710, TKV_c19690, TKV_C19740) as references.

[*] Indicates that a potential G subunit of the RNF-gene cluster is missing (other potential subunits ABCDE were found).

Bibliography

Abarca, M. (2017). *Modeling metabolic networks and their environment interaction*. Thesis, Université de Nantes.

Abrams, Z. B., Johnson, T. S., Huang, K., Payne, P. R., and Coombes, K. (2019). A protocol to evaluate RNA sequencing normalization methods. *BMC Bioinformatics*, **20**(24), 1–7.

Adam, P. S., Borrel, G., Brochier-Armanet, C., and Gribaldo, S. (2017). The growing tree of archaea: New perspectives on their diversity, evolution and ecology. *The ISME Journal*, **11**(11), 2407.

Afting, C., Kremmer, E., Brucker, C., Hochheimer, A., and Thauer, R. K. (2000). Regulation of the synthesis of H_2-forming methylenetetrahydromethanopterin dehydrogenase (Hmd) and of HmdII and HmdIII in *Methanothermobacter marburgensis*. *Archives of Microbiology*, **174**(4), 225–232.

Al-Bassam, M. M., Kim, J.-N., Zaramela, L. S., Kellman, B. P., Zuniga, C., Wozniak, J. M., Gonzalez, D. J., and Zengler, K. (2018). Optimization of carbon and energy utilization through differential translational efficiency. *Nature Communications*, **9**(1), 1–13.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, **215**(3), 403–410.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, **25**(17), 3389–3402.

Álvarez, A., Bansode, A., Urakawa, A., Bavykina, A. V., Wezendonk, T. A., Makkee, M., Gascon, J., and Kapteijn, F. (2017). Challenges in the greener production

of formates/formic acid, methanol, and DME by heterogeneously catalyzed CO_2 hydrogenation processes. *Chemical Reviews*, **117**(14), 9804–9838.

Angenent, L. T., Usack, J. G., Xu, J., Hafenbradl, D., Posmanik, R., and Tester, J. W. (2017). Integrating electrochemical, biological, physical, and thermochemical process units to expand the applicability of anaerobic digestion. *Bioresource Technology*.

Angione, C. and Lió, P. (2015). Predictive analytics of environmental adaptability in multi-omic network models. *Scientific Reports*, **5**, 15147.

Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., Chen, P. J., Wilson, C., Newby, G. A., Raguram, A., *et al.* (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, **576**(7785), 149–157.

Apaolaza, I., José-Eneriz, E. S., Agirre, X., Prósper, F., and Planes, F. J. (2018). COBRA methods and metabolic drug targets in cancer. *Molecular & Cellular Oncology*, **5**(1), e1389672.

Bailera, M., Lisbona, P., Romeo, L. M., and Espatolero, S. (2017). Power to gas projects review: Lab, pilot and demo plants for storing renewable energy and CO_2. *Renewable and Sustainable Energy Reviews*, **69**, 292–312.

Balch, W., Fox, G. E., Magrum, L. J., Woese, C. R., and Wolfe, R. (1979). Methanogens: Reevaluation of a unique biological group. *Microbiological Reviews*, **43**(2), 260–296.

Banerjee, A., Leang, C., Ueki, T., Nevin, K. P., and Lovley, D. R. (2014). Lactose-inducible system for metabolic engineering of *Clostridium ljungdahlii*. *Applied and Environmental Microbiology*, **80**(8), 2410–2416.

Banno, S., Nishida, K., Arazoe, T., Mitsunobu, H., and Kondo, A. (2018). Deaminase-mediated multiplex genome editing in *Escherichia coli*. *Nature Microbiology*, **3**(4), 423–429.

Bathke, J., Konzer, A., Remes, B., McIntosh, M., and Klug, G. (2019). Comparative analyses of the variation of the transcriptome and proteome of *Rhodobacter sphaeroides* throughout growth. *BMC genomics*, **20**(1), 1–13.

Baumann, L. M., Taubner, R.-S., Oláh, K., Rohrweber, A.-C., Schuster, B., Birgel, D., and Rittmann, S. K.-M. (2022). Quantitative analysis of core lipid production in *Methanothermobacter marburgensis* at different scales. *Bioengineering*, **9**(4), 169.

Bayat, H., Modarressi, M. H., and Rahimpour, A. (2018). The conspicuity of CRISPR-Cpf1 system as a significant breakthrough in genome editing. *Current Microbiology*, **75**(1), 107–115.

Beck, B. J., Connolly, L. E., De Las Peñas, A., and Downs, D. M. (1997). Evidence that *rseC*, a gene in the *rpoE* cluster, has a role in thiamine synthesis in *Salmonella typhimurium*. *Journal of Bacteriology*, **179**(20), 6504–6508.

Beck, M. H., Flaiz, M., Bengelsdorf, F. R., and Dürre, P. (2020). Induced heterologous expression of the arginine deiminase pathway promotes growth advantages in the strict anaerobe *Acetobacterium woodii*. *Applied Microbiology and Biotechnology*, **104**(2), 687–699.

Becker, J. and Wittmann, C. (2018). From systems biology to metabolically engineered cells—an omics perspective on the development of industrial microbes. *Current Opinion in Microbiology*, **45**, 180–188.

Becker, S. A. and Palsson, B. Ø. (2008). Context-specific metabolic networks are consistent with experiments. *PLoS Computational Biology*, **4**(5), e1000082.

Beifuss, U., Tietze, M., Bäumer, S., and Deppenmeier, U. (2000). Methanophenazine: Structure, total synthesis, and function of a new cofactor from methanogenic Archaea. *Angewandte Chemie International Edition*, **39**(14), 2470–2472.

Benedict, M. N., Gonnerman, M. C., Metcalf, W. W., and Price, N. D. (2012). Genome-scale metabolic reconstruction and hypothesis testing in the methanogenic archaeon *Methanosarcina acetivorans* C2A. *Journal of Bacteriology*, **194**(4), 855–865.

Bengelsdorf, F. R., Beck, M. H., Erz, C., Hoffmeister, S., Karl, M. M., Riegler, P., Wirth, S., Poehlein, A., Weuster-Botz, D., and Duerre, P. (2018). Bacterial anaerobic synthesis gas (syngas) and CO_2 + H_2 fermentation. In *Advances in Applied Microbiology*, volume 103, pages 143–221. Elsevier.

Bergmann, F. T., Adams, R., Moodie, S., Cooper, J., Glont, M., Golebiewski, M., Hucka, M., Laibe, C., Miller, A. K., Nickerson, D. P., *et al.* (2014). COMBINE archive and OMEX format: One file to share all information to reproduce a modeling project. *BMC Bioinformatics*, **15**(1), 1–9.

Bernacchi, S., Rittmann, S., Seifert, A. H., Krajete, A., and Herwig, C. (2014). Experimental methods for screening parameters influencing the growth to product yield (Y (x/CH_4)) of a biological methane production (BMP) process performed with *Methanothermobacter marburgensis*. *AIMS bioengineering*, **1**(2), 72–86.

Bernacchi, S., Krajete, A., and Herwig, C. (2016). Experimental workflow for developing a feed forward strategy to control biomass growth and exploit maximum specific methane productivity of *Methanothermobacter marburgensis* in a biological methane production process (BMPP). *AIMS Microbiology*, **2**(3), 262–277.

Biegel, E. and Müller, V. (2010). Bacterial Na^+-translocating ferredoxin: NAD^+ oxidoreductase. *Proceedings of the National Academy of Sciences*, **107**(42), 18138–18142.

Biegel, E., Schmidt, S., González, J. M., and Müller, V. (2011). Biochemistry, evolution and physiological function of the Rnf complex, a novel ion-motive electron transport complex in prokaryotes. *Cellular and Molecular Life Sciences*, **68**(4), 613–634.

Birkel, G. W., Ghosh, A., Kumar, V. S., Weaver, D., Ando, D., Backman, T. W., Arkin, A. P., Keasling, J. D., and Martín, H. G. (2017). The JBEI quantitative metabolic modeling library (jQMM): A Python library for modeling microbial metabolism. *BMC Bioinformatics*, **18**(1), 205.

Blazier, A. S. and Papin, J. A. (2012). Integration of expression data in genome-scale metabolic network reconstructions. *Frontiers in Physiology*, **3**, 299.

Boersema, P. J., Raijmakers, R., Lemeer, S., Mohammed, S., and Heck, A. J. (2009). Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nature Protocols*, **4**(4), 484.

Bokranz, M., Klein, A., and Meile, L. (1990). Complete nucleotide sequence of plasmid pME2001 of *Methanobacterium thermoautotrophicum* (Marburg). *Nucleic Acids Research*, **18**(2), 363.

Boone, D. R. (2015). *Methanothermobacter*. *Bergey's Manual of Systematics of Archaea and Bacteria*, pages 1–8.

Bordbar, A., Mo, M. L., Nakayasu, E. S., Schrimpe-Rutledge, A. C., Kim, Y., Metz, T. O., Jones, M. B., Frank, B. C., Smith, R. D., Peterson, S. N., and Palsson, B. Ø. (2012). Model-driven multi-omic data analysis elucidates metabolic immunomodulators of macrophage activation. *Molecular Systems Biology*, **8**(1), 558.

Boucher, J., Schurr, M., Yu, H., Rowen, D., and Deretic, V. (1997). *Pseudomonas aeruginosa* in cystic fibrosis: Role of *mucC* in the regulation of alginate production and stress sensitivity. *Microbiology*, **143**(11), 3473–3480.

Bourgade, B., Minton, N. P., and Islam, M. A. (2021). Genetic and metabolic engineering challenges of C1-gas fermenting acetogenic chassis organisms. *FEMS Microbiology Reviews*, **45**(2), fuab008.

Broadbent, J. A., Broszczak, D. A., Tennakoon, I. U., and Huygens, F. (2016). Pan-proteomics, a concept for unifying quantitative proteome measurements when comparing closely-related bacterial strains. *Expert Review of Proteomics*, **13**(4), 355–365.

Brown, A. M., Hoopes, S. L., White, R. H., and Sarisky, C. A. (2011). Purine biosynthesis in Archaea: Variations on a theme. *Biology Direct*, **6**(1), 1–21.

Bryant, C. (2018). ResearchPy.

Buan, N. R. (2018). Methanogens: Pushing the boundaries of biology. *Emerging Topics in Life Sciences*, **2**(4), 629–646.

Buckel, W. and Thauer, R. K. (2018). Flavin-based electron bifurcation, ferredoxin, flavodoxin, and anaerobic respiration with protons (Ech) or NAD^+ (Rnf) as electron acceptors: A historical review. *Frontiers in Microbiology*, **9**, 401.

Burgard, A. P., Pharkya, P., and Maranas, C. D. (2003). Optknock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnology and bioengineering*, **84**(6), 647–657.

Bushnell, B. (2014). BMap: A fast, accurate, splice-aware aligner. Technical report, Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States).

Butsch, B. and Bachofen, R. (1981). Temperature studies on methanogenic bacteria in Icelandic hot springs. *Research Institute Nedri As, Hveragerdi, Iceland*, **36**, 21–42.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, **10**(1), 1–9.

Cardoso, J. (2018). Genome-scale Metabolic Models. `http://darwin.di.uminho.pt/models/models`. Accessed: 2022-05-25.

Caspi, R., Billington, R., Keseler, I. M., Kothari, A., Krummenacker, M., Midford, P. E., Ong, W. K., Paley, S., Subhraveti, P., and Karp, P. D. (2020). The MetaCyc database of metabolic pathways and enzymes-a 2019 update. *Nucleic Acids Research*, **48**(D1), D445–D453.

Chae, T. U., Choi, S. Y., Kim, J. W., Ko, Y.-S., and Lee, S. Y. (2017). Recent advances in systems metabolic engineering tools and strategies. *Current Opinion in Biotechnology*, **47**, 67–82.

Chan, S. H., Cai, J., Wang, L., Simons-Senftle, M. N., and Maranas, C. D. (2017). Standardizing biomass reactions and ensuring complete mass balance in genome-scale metabolic models. *Bioinformatics*, **33**(22), 3603–3609.

Chandrasekaran, S. and Price, N. D. (2010). Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences*, **107**(41), 17845–17850.

Charubin, K., Streett, H., and Papoutsakis, E. T. (2020). Development of strong anaerobic fluorescent reporters for *Clostridium acetobutylicum* and *Clostridium ljungdahlii* using HaloTag and SNAP-tag proteins. *Applied and Environmental Microbiology*, **86**(20), e01271–20.

Chatsurachai, S., Furusawa, C., and Shimizu, H. (2012). An *in silico* platform for the design of heterologous pathways in nonnative metabolite production. *BMC bioinformatics*, **13**(1), 1–11.

Chen, H., Cong, T. N., Yang, W., Tan, C., Li, Y., and Ding, Y. (2009). Progress in electrical energy storage system: A critical review. *Progress in Natural Science*, **19**(3), 291–312.

Chen, W., Zhang, Y., Zhang, Y., Pi, Y., Gu, T., Song, L., Wang, Y., and Ji, Q. (2018). CRISPR/Cas9-based genome editing in *Pseudomonas aeruginosa* and cytidine deaminase-mediated base editing in *Pseudomonas* species. *IScience*, **6**, 222–231.

Chen, X., Gao, C., Guo, L., Hu, G., Luo, Q., Liu, J., Nielsen, J., Chen, J., and Liu, L. (2017). DCEO biotechnology: Tools to design, construct, evaluate, and optimize the metabolic pathway for biosynthesis of chemicals. *Chemical Reviews*, **118**(1), 4–72.

Chen, Y. and Nielsen, J. (2013). Advances in metabolic pathway and strain engineering paving the way for sustainable production of chemical building blocks. *Current Opinion in Biotechnology*, **24**(6), 965–972.

Chen, Y. and Nielsen, J. (2019). Energy metabolism controls phenotypes by protein efficiency and allocation. *Proceedings of the National Academy of Sciences*, **116**(35), 17592–17597.

Christendat, D., Yee, A., Dharamsi, A., Kluger, Y., Savchenko, A., Cort, J. R., Booth, V., Mackereth, C. D., Saridakis, V., Ekiel, I., *et al.* (2000). Structural proteomics of an archaeon. *Nature Structural Biology*, **7**(10), 903–909.

Chubukov, V., Mukhopadhyay, A., Petzold, C. J., Keasling, J. D., and Martín, H. G. (2016). Synthetic and systems biology for microbial production of commodity chemicals. *NPJ Systems Biology and Applications*, **2**, 16009.

Chung, C. H., Lin, D.-W., Eames, A., and Chandrasekaran, S. (2021). Next-generation genome-scale metabolic modeling through integration of regulatory mechanisms. *Metabolites*, **11**(9), 606.

Ciscar, J.-C., Iglesias, A., Feyen, L., Szabó, L., Van Regemorter, D., Amelung, B., Nicholls, R., Watkiss, P., Christensen, O. B., Dankers, R., *et al.* (2011). Physical and economic consequences of climate change in Europe. *Proceedings of the National Academy of Sciences*, **108**(7), 2678–2683.

Claassens, N. J., Cotton, C. A., Kopljar, D., and Bar-Even, A. (2019). Making quantitative sense of electromicrobial production. *Nature Catalysis*, **2**(5), 437–447.

Cock, P. J., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., *et al.* (2009). Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, **25**(11), 1422–1423.

Cold Spring Harbor Protocols (2009). CTAB extraction buffer.

Consortium, U. (2019). Uniprot: A worldwide hub of protein knowledge. *Nucleic Acids Research*, **47**(D1), D506–D515.

Copeland, W. B., Bartley, B. A., Chandran, D., Galdzicki, M., Kim, K. H., Sleight, S. C., Maranas, C. D., and Sauro, H. M. (2012). Computational tools for metabolic engineering. *Metabolic Engineering*, **14**(3), 270–280.

Costa, K. C. and Leigh, J. A. (2014). Metabolic versatility in methanogens. *Current Opinion in Biotechnology*, **29**, 70–75.

Costa, K. C., Wong, P. M., Wang, T., Lie, T. J., Dodsworth, J. A., Swanson, I., Burn, J. A., Hackett, M., and Leigh, J. A. (2010). Protein complexing in a methanogen suggests electron bifurcation and electron delivery from formate to heterodisulfide reductase. *Proceedings of the National Academy of Sciences*, **107**(24), 11050–11055.

Costa, K. C., Yoon, S. H., Pan, M., Burn, J. A., Baliga, N. S., and Leigh, J. A. (2013a). Effects of H_2 and formate on growth yield and regulation of methanogenesis in *Methanococcus maripaludis*. *Journal of Bacteriology*, **195**(7), 1456–1462.

Costa, K. C., Lie, T. J., Jacobs, M. A., and Leigh, J. A. (2013b). H_2-independent growth of the hydrogenotrophic methanogen *Methanococcus maripaludis*. *MBio*, **4**(2), e00062–13.

Cotton, C. A., Claassens, N. J., Benito-Vaquerizo, S., and Bar-Even, A. (2020). Renewable methanol and formate as microbial feedstocks. *Current Opinion in Biotechnology*, **62**, 168–180.

Daley, D. O., Rapp, M., Granseth, E., Melén, K., Drew, D., and Von Heijne, G. (2005). Global topology analysis of the *Escherichia coli* inner membrane proteome. *Science*, **308**(5726), 1321–1323.

De Las Peñas, A., Connolly, L., and Gross, C. A. (1997). The σe-mediated response to extracytoplasmic stress in *Escherichia coli* is transduced by RseA and RseB, two negative regulators of σe. *Molecular Microbiology*, **24**(2), 373–385.

de Poorter, L. M., Geerts, W. G., Theuvenet, A. P., and Keltjens, J. T. (2003). Bioenergetics of the formyl-methanofuran dehydrogenase and heterodisulfide reductase reactions in *Methanothermobacter thermautotrophicus*. *European Journal of Biochemistry*, **270**(1), 66–75.

de Poorter, L. M., Geerts, W. J., and Keltjens, J. T. (2007). Coupling of *Methanothermobacter thermautotrophicus* methane formation and growth in fed-batch and continuous cultures under different H_2 gassing regimens. *Applied and Environmental Microbiology*, **73**(3), 740–749.

de Souza Pinto Lemgruber, R., Valgepea, K., Gonzalez Garcia, R. A., De Bakker, C., Palfreyman, R. W., Tappel, R., Köpke, M., Simpson, S. D., Nielsen, L. K., and Marcellin, E. (2019). A TetR-family protein (CAETHG_0459) activates transcription from a new promoter motif associated with essential genes for autotrophic growth in acetogens. *Frontiers in Microbiology*, page 2549.

Deutsch, E. W., Csordas, A., Sun, Z., Jarnuczak, A., Perez-Riverol, Y., Ternent, T., Campbell, D. S., Bernal-Llinares, M., Okuda, S., Kawano, S., *et al.* (2016). The ProteomeXchange consortium in 2017: Supporting the cultural change in proteomics public data deposition. *Nucleic Acids Research*, page gkw936.

Deutsch, E. W., Bandeira, N., Sharma, V., Perez-Riverol, Y., Carver, J. J., Kundu, D. J., García-Seisdedos, D., Jarnuczak, A. F., Hewapathirana, S., Pullman, B. S., *et al.* (2020). The ProteomeXchange consortium in 2020: Enabling 'big data'approaches in proteomics. *Nucleic Acids Research*, **48**(D1), D1145–D1152.

Diekert, G., Jaenchen, R., and Thauer, R. K. (1980a). Biosynthetic evidence for a nickel tetrapyrrole structure of factor F_{430} from *Methanobacterium thermoautotrophicum*. *FEBS Letters*, **119**(1), 118–120.

Diekert, G., Gilles, H.-H., Jaenchen, R., and Thauer, R. K. (1980b). Incorporation of 8 succinate per mol nickel into factors F_{430} by *Methanobacterium thermoautotrophicum*. *Archives of Microbiology*, **128**(2), 256–262.

Diender, M., Pereira, R., Wessels, H. J., Stams, A. J., and Sousa, D. Z. (2016). Proteomic analysis of the hydrogen and carbon monoxide metabolism of *Methanothermobacter marburgensis*. *Frontiers in Microbiology*, **7**, 1049.

Ding, H. and Demple, B. (1997). *In vivo* kinetics of a redox-regulated transcriptional switch. *Proceedings of the National Academy of Sciences*, **94**(16), 8445–8449.

Ding, X., Yang, W.-J., Min, H., Peng, X.-T., Zhou, H.-Y., and Lu, Z.-M. (2010). Isolation and characterization of a new strain of *Methanothermobacter marburgens*is DX01 from hot springs in China. *Anaerobe*, **16**(1), 54–59.

Dolci, F., Thomas, D., Hilliard, S., Guerra, C. F., Hancke, R., Ito, H., Jegoux, M., Kreeft, G., Leaver, J., Newborough, M., *et al.* (2019). Incentives and legal barriers for power-to-hydrogen pathways: An international snapshot. *International Journal of Hydrogen Energy*, **44**(23), 11394–11401.

Dong, H., Tao, W., Zhang, Y., and Li, Y. (2012). Development of an anhydrotetracycline-inducible gene expression system for solvent-producing *Clostridium acetobutylicum*: A useful tool for strain engineering. *Metabolic Engineering*, **14**(1), 59–67.

Drake, H. L., Gößner, A. S., and Daniel, S. L. (2008). Old acetogens, new light. *Annals of the New York Academy of Sciences*, **1125**(1), 100–128.

Dry, M. E. (2002). The Fischer–Tropsch process: 1950–2000. *Catalysis Today*, **71**(3-4), 227–241.

Duboc, P., Schill, N., Menoud, L., Van Gulik, W., and Von Stockar, U. (1995). Measurements of sulfur, phosphorus and other ions in microbial biomass: Influence on correct determination of elemental composition and degree of reduction. *Journal of Biotechnology*, **43**(2), 145–158.

Dunphy, L. J. and Papin, J. A. (2018). Biomedical applications of genome-scale metabolic network reconstructions of human pathogens. *Current Opinion in Biotechnology*, **51**, 70–79.

Ebrahim, A., Lerman, J. A., Palsson, B. Ø., and Hyduke, D. R. (2013). COBRApy: Constraints-based reconstruction and analysis for Python. *BMC Systems Biology*, **7**(1), 1–6.

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**(5), 1792–1797.

Edwards, J. and Palsson, B. Ø. (2000). The *Escherichia coli* MG1655 *in silico* metabolic genotype: Its definition, characteristics, and capabilities. *Proceedings of the National Academy of Sciences*, **97**(10), 5528–5533.

Emerson, D. F., Woolston, B. M., Liu, N., Donnelly, M., Currie, D. H., and Stephanopoulos, G. (2019). Enhancing hydrogen-dependent growth of and carbon dioxide fixation by *Clostridium ljungdahlii* through nitrate supplementation. *Biotechnology and Bioengineering*, **116**(2), 294–306.

Esquivel-Elizondo, S., Bağcı, C., Temovska, M., Jeon, B. S., Bessarab, I., Williams, R. B., Huson, D. H., and Angenent, L. T. (2021). The isolate *Caproiciproducens* sp. 7D4C2 produces *n*-caproate at mildly acidic conditions from hexoses: Genome and rBOX comparison with related strains and chain-elongating bacteria. *Frontiers in Microbiology*, **11**, 3335.

Fagbadebo, F. O., Kaiser, P. D., Zittlau, K., Bartlick, N., Wagner, T. R., Froehlich, T., Jarjour, G., Nueske, S., Scholz, A., Traenkle, B., *et al.* (2021). A nanobody-based toolset to monitor and modify the mitochondrial GTPase Miro1. *bioRxiv*.

Fagerlund, R. D., Staals, R. H., and Fineran, P. C. (2015). The Cpf1 CRISPR-Cas protein expands genome-editing tools. *Genome Biology*, **16**(1), 1–3.

Fang, X., Lloyd, C. J., and Palsson, B. Ø. (2020). Reconstructing organisms *in silico*: Genome-scale models and their emerging applications. *Nature Reviews Microbiology*, **18**(12), 731–743.

Faria, J. P., Overbeek, R., Xia, F., Rocha, M., Rocha, I., and Henry, C. S. (2014). Genome-scale bacterial transcriptional regulatory networks: Reconstruction and integrated analysis with metabolic models. *Briefings in Bioinformatics*, **15**(4), 592–611.

Faria, J. P., Rocha, M., Rocha, I., and Henry, C. S. (2018). Methods for automated genome-scale metabolic model reconstruction. *Biochemical Society Transactions*, page BST20170246.

Fast, A. G. and Papoutsakis, E. T. (2012). Stoichiometric and energetic analyses of non-photosynthetic CO_2-fixation pathways to support synthetic biology strategies for production of fuels and chemicals. *Current Opinion in Chemical Engineering*, **1**(4), 380–395.

Feilmeier, B. J., Iseminger, G., Schroeder, D., Webber, H., and Phillips, G. J. (2000). Green fluorescent protein functions as a reporter for protein localization in *Escherichia coli*. *Journal of Bacteriology*, **182**(14), 4068–4076.

Feist, A. M., Scholten, J. C., Palsson, B. Ø., Brockman, F. J., and Ideker, T. (2006). Modeling methanogenesis with a genome-scale metabolic reconstruction of *Methanosarcina barkeri*. *Molecular Systems Biology*, **2**(1).

Feist, A. M., Herrgård, M. J., Thiele, I., Reed, J. L., and Palsson, B. Ø. (2009). Reconstruction of biochemical networks in microorganisms. *Nature Reviews Microbiology*, **7**(2), 129–143.

Fink, C., Beblawy, S., Enkerlin, A. M., Mühling, L., Angenent, L. T., and Molitor, B. (2021). A shuttle-vector system allows heterologous gene expression in the thermophilic methanogen *Methanothermobacter thermautotrophicus* ΔH. *Mbio*, **12**(6), e02766–21.

Fong, S. S. (2014). Computational approaches to metabolic engineering utilizing systems biology and synthetic biology. *Computational and Structural Biotechnology Journal*, **11**(18), 28–34.

Fuchs, G., Stupperich, E., and Thauer, R. K. (1978a). Acetate assimilation and the synthesis of alanine, aspartate and glutamate in *Methanobacterium thermoautotrophicum*. *Archives of Microbiology*, **117**(1), 61–66.

Fuchs, G., Stupperich, E., and Thauer, R. K. (1978b). Function of fumarate reductase in methanogenic bacteria (*Methanobacterium*). *Archives of Microbiology*, **119**(2), 215–218.

Galperin, M. Y., Wolf, Y. I., Makarova, K. S., Vera Alvarez, R., Landsman, D., and Koonin, E. V. (2021). COG database update: Focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Research*, **49**(D1), D274–D281.

Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., and Liu, D. R. (2017). Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature*, **551**(7681), 464–471.

Gerhard, E., Butsch, B. M., Marison, I. W., and von Stockar, U. (1993). Improved growth and methane production conditions for *Methanobacterium thermoautotrophicum*. *Applied Microbiology and Biotechnology*, **40**(2-3), 432–437.

Glont, M., Nguyen, T. V. N., Graesslin, M., Hälke, R., Ali, R., Schramm, J., Wimalaratne, S. M., Kothamachu, V. B., Rodriguez, N., Swat, M. J., Eils, J., Eils, R., Laibe, C., Malik-Sheriff, R. S., Chelliah, V., Le Novère, N., and Hermjakob, H. (2018). Biomodels: Expanding horizons to include more modelling approaches and formats. *Nucleic Acids Research*, **46**(D1), D1248–D1253.

Gonnerman, M. C., Benedict, M. N., Feist, A. M., Metcalf, W. W., and Price, N. D. (2013). Genomically and biochemically accurate metabolic reconstruction of *Methanosarcina barkeri* Fusaro, iMG746. *Biotechnology Journal*, **8**(9), 1070–1079.

Gordon, A. and Hannon, G. (2017). Fastx-toolkit. FASTQ/A short-reads pre-processing tools. 2010. Unpublished available online at: `http://hannonlab.cshl.edu/fastx_toolkit`. Accessed: 2022-03-16.

Götz, M., Lefebvre, J., Mörs, F., Koch, A. M., Graf, F., Bajohr, S., Reimert, R., and Kolb, T. (2016). Renewable power-to-gas: A technological and economic review. *Renewable Energy*, **85**, 1371–1390.

Goyal, N., Widiastuti, H., Karimi, I., and Zhou, Z. (2014). A genome-scale metabolic model of *Methanococcus maripaludis* S2 for CO_2 capture and conversion to methane. *Molecular BioSystems*, **10**(5), 1043–1054.

Goyal, N., Zhou, Z., and Karimi, I. A. (2016). Metabolic processes of *Methanococcus maripaludis* and potential applications. *Microbial Cell Factories*, **15**(1), 1–19.

Gu, C., Kim, G. B., Kim, W. J., Kim, H. U., and Lee, S. Y. (2019). Current status and applications of genome-scale metabolic models. *Genome Biology*, **20**(1), 1–18.

Gu, T., Zhao, S., Pi, Y., Chen, W., Chen, C., Liu, Q., Li, M., Han, D., and Ji, Q. (2018). Highly efficient base editing in *Staphylococcus aureus* using an engineered CRISPR RNA-guided cytidine deaminase. *Chemical Science*, **9**(12), 3248–3253.

Guerra, L., Rossi, S., Rodrigues, J., Gomes, J., Puna, J., and Santos, M. T. (2018). Methane production by a combined Sabatier reaction/water electrolysis process. *Journal of Environmental Chemical Engineering*, **6**(1), 671–676.

Gurobi Optimization, LLC (2020). Gurobi optimizer reference manual. Gurobi Optimization Inc.

Gustavsson, M. and Lee, S. Y. (2016). Prospects of microbial cell factories developed through systems metabolic engineering. *Microbial Biotechnology*, **9**(5), 610–617.

Hall, J. B. (1973). The occurrence of nitrate on the early earth and its role in the evolution of the prokaryotes. *Space Life Sciences*, **4**(1), 204–213.

Hall, J. E., Hooker, P., and Jeffrey, K. E. (2021). Gas detection of hydrogen/natural gas blends in the gas industry. *International Journal of Hydrogen Energy*, **46**(23), 12555–12565.

Hamilton, J. J., Contreras, M. C., and Reed, J. L. (2015). Thermodynamics and H_2 transfer in a methanogenic, syntrophic community. *PLoS Computational Biology*, **11**(7), e1004364.

Hasan, S. and Hall, J. (1975). The physiological function of nitrate reduction in *Clostridium perfringens*. *Microbiology*, **87**(1), 120–128.

Hassa, J., Wibberg, D., Maus, I., Pühler, A., and Schlüter, A. (2019). Genome analyses and genome-centered metatranscriptomics of *Methanothermobacter wolfeii* strain SIV6, isolated from a thermophilic production-scale biogas fermenter. *Microorganisms*, **8**(1), 13.

Hattori, S., Kamagata, Y., Hanada, S., and Shoun, H. (2000). *Thermacetogenium phaeum* gen. nov., sp. nov., a strictly anaerobic, thermophilic, syntrophic acetate-oxidizing

bacterium. *International Journal of Systematic and Evolutionary Microbiology*, **50**(4), 1601–1609.

Hausinger, R. P., Orme-Johnson, W. H., and Walsh, C. (1985). Factor 390 chromophores: Phosphodiester between AMP or GMP and methanogenic factor 420. *Biochemistry*, **24**(7), 1629–1633.

He, B., Cai, C., McCubbin, T., Muriel, J. C., Sonnenschein, N., Hu, S., Yuan, Z., and Marcellin, E. (2022). A genome-scale metabolic model of *Methanoperedens nitroreducens*: Assessing bioenergetics and thermodynamic feasibility. *Metabolites*, **12**(4), 314.

Heap, J. T., Pennington, O. J., Cartman, S. T., and Minton, N. P. (2009). A modular system for *Clostridium* shuttle plasmids. *Journal of Microbiological Methods*, **78**(1), 79–85.

Heckmann, D., Lloyd, C. J., Mih, N., Ha, Y., Zielinski, D. C., Haiman, Z. B., Desouki, A. A., Lercher, M. J., and Palsson, B. Ø. (2018). Machine learning applied to enzyme turnover numbers reveals protein structural correlates and improves metabolic models. *Nature Communications*, **9**(1), 1–10.

Heffernan, J. K., Valgepea, K., de Souza Pinto Lemgruber, R., Casini, I., Plan, M., Tappel, R., Simpson, S. D., Köpke, M., Nielsen, L. K., and Marcellin, E. (2020). Enhancing CO_2-valorization using *Clostridium autoethanogenum* for sustainable fuel and chemicals production. *Frontiers in Bioengineering and Biotechnology*, **8**, 204.

Heffernan, J. K., Mahamkali, V., Valgepea, K., Marcellin, E., and Nielsen, L. K. (2022). Analytical tools for unravelling the metabolism of gas-fermenting *Clostridia*. *Current Opinion in Biotechnology*, **75**, 102700.

Hegerl, G. C., Zwiers, F. W., Braconnot, P., Gillett, N. P., Luo, Y., Orsini, J. A. M., Nicholls, N., Penner, J. E., Stott, P. A., Allen, M., *et al.* (2007). Understanding and attributing climate change.

Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S. N., Richelle, A., Heinken, A., Haraldsdóttir, H. S., Wachowiak, J., Keating, S. M., Vlasov, V., *et al.* (2019). Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v. 3.0. *Nature Protocols*, **14**(3), 639–702.

Hendrickson, E. L. and Leigh, J. A. (2008). Roles of coenzyme F_{420}-reducing hydrogenases and hydrogen-and F_{420}-dependent methylenetetrahydromethanopterin dehydrogenases in reduction of F_{420} and production of hydrogen during methanogenesis. *Journal of Bacteriology*, **190**(14), 4818–4821.

Henry, C. S., DeJongh, M., Best, A. A., Frybarger, P. M., Linsay, B., and Stevens, R. L. (2010). High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nature Biotechnology*, **28**(9), 977–982.

Hernandez, E. and Costa, K. (2022). The fluorescence-activating and absorption-shifting tag (FAST) enables live-cell fluorescence imaging of *Methanococcus maripaludis*. *bioRxiv*.

Hess, V., Schuchmann, K., and Müller, V. (2013). The ferredoxin: NAD^+ oxidoreductase (Rnf) from the acetogen *Acetobacterium woodii* requires Na^+ and is reversibly coupled to the membrane potential. *Journal of Biological Chemistry*, **288**(44), 31496–31502.

Hess, V., Poehlein, A., Weghoff, M. C., Daniel, R., and Müller, V. (2014). A genome-guided analysis of energy conservation in the thermophilic, cytochrome-free acetogenic bacterium *Thermoanaerobacter kivui*. *BMC Genomics*, **15**(1), 1–14.

Hirano, S., Matsumoto, N., Morita, M., Sasaki, K., and Ohmura, N. (2013). Electrochemical control of redox potential affects methanogenesis of the hydrogenotrophic methanogen *Methanothermobacter thermautotrophicus*. *Letters in Applied Microbiology*, **56**(5), 315–321.

Hoffmeister, S., Gerdom, M., Bengelsdorf, F. R., Linder, S., Flüchter, S., Öztürk, H., Blümke, W., May, A., Fischer, R.-J., Bahl, H., *et al.* (2016). Acetone production with metabolically engineered strains of *Acetobacterium woodii*. *Metabolic Engineering*, **36**, 37–47.

Hosseini, Z. and Marashi, S.-A. (2017). Discovering missing reactions of metabolic networks by using gene co-expression data. *Scientific Reports*, **7**, 41774.

Hu, J. H., Miller, S. M., Geurts, M. H., Tang, W., Chen, L., Sun, N., Zeina, C. M., Gao, X., Rees, H. A., Lin, Z., *et al.* (2018). Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature*, **556**(7699), 57–63.

Huang, H., Chai, C., Li, N., Rowe, P., Minton, N. P., Yang, S., Jiang, W., and Gu, Y. (2016). CRISPR/Cas9-based efficient genome editing in *Clostridium ljungdahlii*, an autotrophic gas-fermenting bacterium. *ACS Synthetic Biology*, **5**(12), 1355–1361.

Huang, H., Chai, C., Yang, S., Jiang, W., and Gu, Y. (2019). Phage serine integrase-mediated genome engineering for efficient expression of chemical biosynthetic pathway in gas-fermenting *Clostridium ljungdahlii*. *Metabolic Engineering*, **52**, 293–302.

Huang, Q., Yang, L., Luo, J., Guo, L., Wang, Z., Yang, X., Jin, W., Fang, Y., Ye, J., and Shan, B. (2015). SWATH enables precise label-free quantification on proteome scale. *Proteomics*, **15**(7), 1215–1223.

Hüster, R. and Thauer, R. K. (1983). Pyruvate assimilation by *Methanobacterium thermoautotrophicum*. *FEMS Microbiology Letters*, **19**(2-3), 207–209.

Illumina (2013). Illumina Sequencing Technology. `https://www.youtube.com/watch?v=womKfikWlxM&ab_channel=Illumina`. Accessed: 2022-03-16.

Imam, S., Noguera, D. R., and Donohue, T. J. (2015). An integrated approach to reconstructing genome-scale transcriptional regulatory networks. *PLoS Computational Biology*, **11**(2), e1004103.

Ingram-Smith, C. and Smith, K. S. (2007). AMP-forming acetyl-CoA synthetases in Archaea show unexpected diversity in substrate utilization. *Archaea*, **2**(2), 95–107.

Jahn, M., Crang, N., Janasch, M., Hober, A., Forsström, B., Kimler, K., Mattausch, A., Chen, Q., Asplund-Samuelsson, J., and Hudson, E. P. (2021). Protein allocation and utilization in the versatile chemolithoautotroph *Cupriavidus necator*. *Elife*, **10**, e69019.

Jain, S., Caforio, A., and Driessen, A. J. (2014). Biosynthesis of archaeal membrane ether lipids. *Frontiers in Microbiology*, **5**, 641.

Jarrell, K. F. and Sprott, G. D. (1981). The transmembrane electrical potential and intracellular pH in methanogenic bacteria. *Canadian Journal of Microbiology*, **27**(7), 720–728.

Jensen, J. N. (2001). Approach to steady state in completely mixed flow reactors. *Journal of Environmental Engineering*, **127**(1), 13–18.

Jensen, K., Broeken, V., Hansen, A. S. L., Sonnenschein, N., and Herrgård, M. J. (2019). OptCouple: Joint simulation of gene knockouts, insertions and medium modifications for prediction of growth-coupled strain designs. *Metabolic Engineering Communications*, **8**, e00087.

Jensen, P. A. and Papin, J. A. (2010). Functional integration of a metabolic network model and expression data without arbitrary thresholding. *Bioinformatics*, **27**(4), 541–547.

Jensen, P. A., Lutz, K. A., and Papin, J. A. (2011). TIGER: Toolbox for integrating genome-scale metabolic models, expression data, and transcriptional regulatory networks. *BMC Systems Biology*, **5**(1), 147.

Jerby, L., Shlomi, T., and Ruppin, E. (2010). Computational reconstruction of tissue-specific metabolic models: Application to human liver metabolism. *Molecular Systems Biology*, **6**(1), 401.

Jeske, L., Placzek, S., Schomburg, I., Chang, A., and Schomburg, D. (2019). BRENDA in 2019: A European ELIXIR core data resource. *Nucleic Acids Research*, **47**(D1), D542–D549.

Jiang, Y., Chen, B., Duan, C., Sun, B., Yang, J., and Yang, S. (2015). Multigene editing in the *Escherichia coli* genome *via* the CRISPR-Cas9 system. *Applied and Environmental Microbiology*, **81**(7), 2506–2514.

Jussofie, A. and Gottschalk, G. (1986). Further studies on the distribution of cytochromes in methanogenic bacteria. *FEMS Microbiology Letters*, **37**(1), 15–18.

Juty, N., Le Novere, N., and Laibe, C. (2011). Identifiers.org and MIRIAM registry: Community resources to provide persistent identification. *Nucleic Acids Research*, **40**(D1), D580–D586.

Kadam, P. and Ranade, D. (1992). Isolation of *Methanobacterium formicicum* and *Methanosarcina barkeri* from a creek near bombay. *Tropical Ecology*, **33**(2), 200–204.

Karp, P. D., Billington, R., Caspi, R., Fulcher, C. A., Latendresse, M., Kothari, A., Keseler, I. M., Krummenacker, M., Midford, P. E., Ong, Q., *et al.* (2019). The BioCyc collection of microbial genomes and metabolic pathways. *Briefings in Bioinformatics*, **20**(4), 1085–1093.

Karr, E. A. (2010). The methanogen-specific transcription factor MsvR regulates the *fpaA-rlp-rub* oxidative stress operon adjacent to *msvR* in *Methanothermobacter thermautotrophicus*. *Journal of Bacteriology*, **192**(22), 5914–5922.

Kaster, A.-K., Moll, J., Parey, K., and Thauer, R. K. (2011a). Coupling of ferredoxin and heterodisulfide reduction *via* electron bifurcation in hydrogenotrophic methanogenic Archaea. *Proceedings of the National Academy of Sciences*, **108**(7), 2981–2986.

Kaster, A.-K., Goenrich, M., Seedorf, H., Liesegang, H., Wollherr, A., Gottschalk, G., and Thauer, R. K. (2011b). More than 200 genes required for methane formation from H_2 and CO_2 and energy conservation are present in *Methanothermobacter marburgensis* and *Methanothermobacter thermautotrophicus*. *Archaea*, **2011**.

Kato, S., Kosaka, T., and Watanabe, K. (2008). Comparative transcriptome analysis of responses of *Methanothermobacter thermautotrophicus* to different environmental stimuli. *Environmental Microbiology*, **10**(4), 893–905.

Kato, S., Yoshida, R., Yamaguchi, T., Sato, T., Yumoto, I., and Kamagata, Y. (2014). The effects of elevated CO_2 concentration on competitive interaction between aceticlastic and syntrophic methanogenesis in a model microbial consortium. *Frontiers in Microbiology*, **5**, 575.

Katsyv, A. and Müller, V. (2020). Overcoming energetic barriers in acetogenic C1 conversion. *Frontiers in Bioengineering and Biotechnology*, page 1420.

Keith, D. W., Holmes, G., Angelo, D. S., and Heidel, K. (2018). A process for capturing CO_2 from the atmosphere. *Joule*, **2**(8), 1573–1594.

Kiener, A. and Leisinger, T. (1983). Oxygen sensitivity of methanogenic bacteria. *Systematic and Applied Microbiology*, **4**(3), 305–312.

Kiener, A., Orme-Johnson, W. H., and Walsh, C. T. (1988). Reversible conversion of coenzyme F_{420} to the 8-OH-AMP and 8-OH-GMP esters, F_{390}-A and F_{390}-

G, on oxygen exposure and reestablishment of anaerobiosis in *Methanobacterium thermoautotrophicum*. *Archives of Microbiology*, **150**(3), 249–253.

Kim, B., Kim, W. J., Kim, D. I., and Lee, S. Y. (2015). Applications of genome-scale metabolic network model in metabolic engineering. *Journal of Industrial Microbiology & Biotechnology*, **42**(3), 339–348.

Kim, G. B., Gao, Y., Palsson, B. Ø., and Lee, S. Y. (2021). DeepTFactor: A deep learning-based tool for the prediction of transcription factors. *Proceedings of the National Academy of Sciences*, **118**(2), e2021171118.

Kim, T. Y., Sohn, S. B., Kim, Y. B., Kim, W. J., and Lee, S. Y. (2012). Recent advances in reconstruction and applications of genome-scale metabolic models. *Current Opinion in Biotechnology*, **23**(4), 617–623.

Kim, W. J., Kim, H. U., and Lee, S. Y. (2017). Current state and applications of microbial genome-scale metabolic models. *Current Opinion in Systems Biology*, **2**, 10–18.

King, Z. A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J. A., Ebrahim, A., Palsson, B. Ø., and Lewis, N. E. (2015a). BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research*, **44**(D1), D515–D522.

King, Z. A., Dräger, A., Ebrahim, A., Sonnenschein, N., Lewis, N. E., and Palsson, B. Ø. (2015b). Escher: A web application for building, sharing, and embedding data-rich visualizations of biological pathways. *PLoS Computational Biology*, **11**(8), e1004321.

King, Z. A., Lloyd, C. J., Feist, A. M., and Palsson, B. Ø. (2015c). Next-generation genome-scale models for metabolic engineering. *Current Opinion in Biotechnology*, **35**, 23–29.

Klask, C.-M., Kliem-Kuster, N., Molitor, B., and Angenent, L. T. (2020). Nitrate feed improves growth and ethanol production of *Clostridium ljungdahlii* with CO_2 and H_2, but results in stochastic inhibition events. *Frontiers in Microbiology*, **11**, 724.

Klueglein, N., Zeitvogel, F., Stierhof, Y.-D., Floetenmeyer, M., Konhauser, K. O., Kappler, A., and Obst, M. (2014). Potential role of nitrite for abiotic Fe (II)oxidation

and cell encrustation during nitrate reduction by denitrifying bacteria. *Applied and Environmental Microbiology*, **80**(3), 1051–1061.

Knott, G. J. and Doudna, J. A. (2018). CRISPR-Cas guides the future of genetic engineering. *Science*, **361**(6405), 866–869.

Koga, Y. (2011). Early evolution of membrane lipids: How did the lipid divide occur? *Journal of Molecular Evolution*, **72**(3), 274–282.

Koga, Y. and Morii, H. (2007). Biosynthesis of ether-type polar lipids in Archaea and evolutionary considerations. *Microbiology and Molecular Biology Reviews*, **71**(1), 97–120.

Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., and Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, **533**(7603), 420–424.

Koo, M.-S., Lee, J.-H., Rah, S.-Y., Yeo, W.-S., Lee, J.-W., Lee, K.-L., Koh, Y.-S., Kang, S.-O., and Roe, J.-H. (2003). A reducing system of the superoxide sensor SoxR in *Escherichia coli*. *The EMBO journal*, **22**(11), 2614–2622.

Köpke, M. and Simpson, S. D. (2020). Pollution to products: Recycling of 'above ground'carbon by gas fermentation. *Current Opinion in Biotechnology*, **65**, 180–189.

Köpke, M., Held, C., Hujer, S., Liesegang, H., Wiezer, A., Wollherr, A., Ehrenreich, A., Liebl, W., Gottschalk, G., and Dürre, P. (2010). *Clostridium ljungdahlii* represents a microbial production platform based on syngas. *Proceedings of the National Academy of Sciences*, **107**(29), 13087–13092.

Kosaka, T., Toh, H., Fujiyama, A., Sakaki, Y., Watanabe, K., Meng, X., Hanada, S., and Toyoda, A. (2014). Physiological and genetic basis for self-aggregation of a thermophilic hydrogenotrophic methanogen, *Methanothermobacter* strain CaT2. *Environmental Microbiology Reports*, **6**(3), 268–277.

Kroll, A., Engqvist, M. K., Heckmann, D., and Lercher, M. J. (2021). Deep learning allows genome-scale prediction of Michaelis constants from structural features. *PLoS Biology*, **19**(10), e3001402.

Kumar, V. S., Ferry, J. G., and Maranas, C. D. (2011). Metabolic reconstruction of the archaeon methanogen *Methanosarcina acetivorans*. *BMC Systems Biology*, **5**(1), 28.

Kuscu, C., Parlak, M., Tufan, T., Yang, J., Szlachta, K., Wei, X., Mammadov, R., and Adli, M. (2017). CRISPR-STOP: Gene silencing through base-editing-induced nonsense mutations. *Nature Methods*, **14**(7), 710–712.

Le Novère, N., Finney, A., Hucka, M., Bhalla, U. S., Campagne, F., Collado-Vides, J., Crampin, E. J., Halstead, M., Klipp, E., and Mendes, P. (2005). Minimum information requested in the annotation of biochemical models (MIRIAM). *Nature Biotechnology*, **23**(12), 1509.

Leang, C., Ueki, T., Nevin, K. P., and Lovley, D. R. (2013). A genetic system for *Clostridium ljungdahlii*: A chassis for autotrophic production of biocommodities and a model homoacetogen. *Applied and Environmental Microbiology*, **79**(4), 1102–1109.

Lee, D., Smallbone, K., Dunn, W. B., Murabito, E., Winder, C. L., Kell, D. B., Mendes, P., and Swainston, N. (2012). Improving metabolic flux predictions using absolute gene expression data. *BMC Systems Biology*, **6**(1), 1–9.

Leimbach, A. (2016). bac-genomics-scripts: Bovine *E. coli* mastitis comparative genomics edition. *Genome Announcements*.

Leonzio, G. (2016). Process analysis of biological Sabatier reaction for bio-methane production. *Chemical Engineering Journal*, **290**, 490–498.

Lewis, N. E., Nagarajan, H., and Palsson, B. Ø. (2012). Constraining the metabolic genotype–phenotype relationship using a phylogeny of *in silico* methods. *Nature Reviews Microbiology*, **10**(4), 291.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.

Li, Q., Seys, F. M., Minton, N. P., Yang, J., Jiang, Y., Jiang, W., and Yang, S. (2019). CRISPR–Cas9D10A nickase-assisted base editing in the solvent producer *Clostridium beijerinckii*. *Biotechnology and Bioengineering*, **116**(6), 1475–1483.

Li, X., Wang, Y., Liu, Y., Yang, B., Wang, X., Wei, J., Lu, Z., Zhang, Y., Wu, J., Huang, X., *et al.* (2018). Base editing with a Cpf1–cytidine deaminase fusion. *Nature Biotechnology*, **36**(4), 324–327.

Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**(7), 923–930.

Lie, T. J. and Leigh, J. A. (2002). Regulatory response of *Methanococcus maripaludis* to alanine, an intermediate nitrogen source. *Journal of Bacteriology*, **184**(19), 5301–5306.

Lie, T. J. and Leigh, J. A. (2003). A novel repressor of *nif* and *glnA* expression in the methanogenic archaeon *Methanococcus maripaludis*. *Molecular Microbiology*, **47**(1), 235–246.

Lie, T. J., Wood, G. E., and Leigh, J. A. (2005). Regulation of *nif* expression in *Methanococcus maripaludis*: Roles of the euryarchaeal repressor NrpR, 2-oxoglutarate, and two operators. *Journal of Biological Chemistry*, **280**(7), 5236–5241.

Lie, T. J., Costa, K. C., Lupa, B., Korpole, S., Whitman, W. B., and Leigh, J. A. (2012). Essential anaplerotic role for the energy-converting hydrogenase Eha in hydrogenotrophic methanogenesis. *Proceedings of the National Academy of Sciences*, **109**(38), 15473–15478.

Lie, T. J., Kuo, Y. P., Leite, M., Costa, K. C., Harwood, C. S., and Leigh, J. A. (2022). A genetic study of Nif-associated genes in a hyperthermophilic methanogen. *Microbiology Spectrum*, **10**(1), e02093–21.

Liesegang, H., Kaster, A.-K., Wiezer, A., Goenrich, M., Wollherr, A., Seedorf, H., Gottschalk, G., and Thauer, R. K. (2010). Complete genome sequence of *Methanothermobacter marburgensis*, a methanoarchaeon model organism. *Journal of Bacteriology*, **192**(21), 5850–5851.

Lieven, C., Beber, M. E., Olivier, B. G., Bergmann, F. T., Ataman, M., Babaei, P., Bartell, J. A., Blank, L. M., Chauhan, S., Correia, K., *et al.* (2020). MEMOTE for standardized genome-scale metabolic model testing. *Nature Biotechnology*, **38**(3), 272–276.

Liew, F., Henstra, A. M., Köpke, M., Winzer, K., Simpson, S. D., and Minton, N. P. (2017). Metabolic engineering of *Clostridium autoethanogenum* for selective alcohol production. *Metabolic Engineering*, **40**, 104–114.

Lindemann, C., Thomanek, N., Hundt, F., Lerari, T., Meyer, H. E., Wolters, D., and Marcus, K. (2017). Strategies in relative and absolute quantitative mass spectrometry based proteomics. *Biological Chemistry*, **398**(5-6), 687–699.

Liu, C., Mao, L., Zheng, X., Yuan, J., Hu, B., Cai, Y., Xie, H., Peng, X., and Ding, X. (2019). Comparative proteomic analysis of *Methanothermobacter thermautotrophicus* reveals methane formation from H_2 and CO_2 under different temperature conditions. *Microbiologyopen*, **8**(5), e00715.

Liu, J., Tan, Y., Yang, X., Chen, X., and Li, F. (2013). Evaluation of *Clostridium ljungdahlii* DSM 13528 reference genes in gene expression studies by qRT-PCR. *Journal of Bioscience and Bioengineering*, **116**(4), 460–464.

Liu, J. K., O'Brien, E. J., Lerman, J. A., Zengler, K., Palsson, B. Ø., and Feist, A. M. (2014). Reconstruction and modeling protein translocation and compartmentalization in *Escherichia coli* at the genome-scale. *BMC Systems Biology*, **8**(1), 110.

Liu, Y., Beer, L. L., and Whitman, W. B. (2012). Methanogens: A window into ancient sulfur metabolism. *Trends in microbiology*, **20**(5), 251–258.

Livak, K. J. and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods*, **25**(4), 402–408.

Ljungdahl, L. G. (2009). A life with acetogens, thermophiles, and cellulolytic anaerobes. *Annual Review of Microbiology*, **63**, 1–25.

Ljungdhal, L. (1986). The autotrophic pathway of acetate synthesis in acetogenic bacteria. *Annual Reviews in Microbiology*, **40**(1), 415–450.

Loenen, W. A., Dryden, D. T., Raleigh, E. A., and Wilson, G. G. (2014). Type I restriction enzymes and their relatives. *Nucleic Acids Research*, **42**(1), 20–44.

Long, F., Wang, L., Lupa, B., and Whitman, W. B. (2017). A flexible system for cultivation of *Methanococcus* and other formate-utilizing methanogens. *Archaea*, **2017**.

Long, M. R., Ong, W. K., and Reed, J. L. (2015). Computational methods in metabolic engineering for strain design. *Current Opinion in Biotechnology*, **34**, 135–141.

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, **15**(12), 1–21.

Luo, Y., Leisinger, T., and Wasserfallen, A. (2001). Comparative sequence analysis of plasmids pME2001 and pME2200 of *Methanothermobacter marburgensis* strains Marburg and ZH3. *Plasmid*, **45**(1), 18–30.

Lyu, Z. and Lu, Y. (2018). Metabolic shift at the class level sheds light on adaptation of methanogens to oxidative environments. *The ISME journal*, **12**(2), 411–423.

Lyu, Z., Liu, Y., Stams, A., and Sousa, D. (2018). Diversity and taxonomy of methanogens. *Biogenesis of Hydrocarbons. Handbook of Hydrocarbon and Lipid Microbiology*, pages 1–59.

Maarleveld, T. R., Khandelwal, R. A., Olivier, B. G., Teusink, B., and Bruggeman, F. J. (2013). Basic concepts and principles of stoichiometric modeling of metabolic networks. *Biotechnology Journal*, **8**(9), 997–1008.

Machado, D., Andrejev, S., Tramontano, M., and Patil, K. R. (2018). Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *bioRxiv*, page 223198.

Madden, T. (2003). The BLAST sequence analysis tool. *The NCBI Handbook*.

Mahamkali, V., McCubbin, T., Beber, M. E., Noor, E., Marcellin, E., and Nielsen, L. K. (2021). multiTFA: A Python package for multi-variate thermodynamics-based flux analysis. *Bioinformatics*, **37**(18), 3064–3066.

Maillet, N. (2020). Rapid Peptides Generator: Fast and efficient *in silico* protein digestion. *NAR Genomics and Bioinformatics*, **2**(1), lqz004.

Malik-Sheriff, R. S., Glont, M., Nguyen, T. V. N., Tiwari, K., Roberts, M. G., Xavier, A., Vu, M. T., Men, J., Maire, M., Kananathan, S., Fairbanks, E. L., Meyer, J. P., Arankalle, C., Varusai, T. M., Knight-Schrijver, V., Li, L., Dueñas-Roca, C., Dass, G., Keating, S. M., Park, Y. M., Buso, N., Rodriguez, N., Hucka, M., and Hermjakob,

H. (2020). BioModels — 15 years of sharing computational models in life science. *Nucleic Acids Research*, **48**(D1), D407–D415. gkz1055.

Marcellin, E. and Nielsen, L. K. (2018). Advances in analytical tools for high throughput strain engineering. *Current Opinion in Biotechnology*, **54**, 33–40.

Marcellin, E., Behrendorff, J. B., Nagaraju, S., DeTissera, S., Segovia, S., Palfreyman, R. W., Daniell, J., Licona-Cassani, C., Quek, L.-e., and Speight, R. (2016). Low carbon fuels and commodity chemicals from waste gases–systematic approach to understand energy metabolism in a model acetogen. *Green Chemistry*, **18**(10), 3020–3028.

Mardis, E. R. (2017). DNA sequencing technologies: 2006–2016. *Nature Protocols*, **12**(2), 213–218.

Martin, M. R., Fornero, J. J., Stark, R., Mets, L., and Angenent, L. T. (2013). A single-culture bioprocess of *Methanothermobacter thermautotrophicus* to upgrade digester biogas by CO_2-to-CH_4 conversion with H_2. *Archaea*, **2013**.

Martínez, V. S. and Nielsen, L. K. (2014). *NExt: Integration of thermodynamic constraints and metabolomics data into a metabolic network*, pages 65–78. Springer.

Martinez-Salazar, J. M., Moreno, S., Najera, R., Boucher, J. C., Espín, G., Soberon-Chavez, G., and Deretic, V. (1996). Characterization of the genes coding for the putative sigma factor AlgU and its regulators MucA, MucB, MucC, and MucD in *Azotobacter vinelandii* and evaluation of their roles in alginate biosynthesis. *Journal of Bacteriology*, **178**(7), 1800–1808.

Masoudi-Nejad, A. and Asgari, Y. (2015). Metabolic cancer biology: Structural-based analysis of cancer as a metabolic disease, new sights and opportunities for disease treatment. In *Seminars in Cancer Biology*, volume 30, pages 21–29. Elsevier.

Meile, L., Fischer, K., and Leisinger, T. (1995). Characterization of the superoxide dismutase gene and its upstream region from *Methanobacterium thermoautotrophicum* Marburg. *FEMS Microbiology Letters*, **128**(3), 247–253.

Melaina, M. W., Antonia, O., and Penev, M. (2013). Blending hydrogen into natural gas pipeline networks: A review of key issues. *NREL*.

Mendoza, S. N., Olivier, B. G., Molenaar, D., and Teusink, B. (2019). A systematic assessment of current genome-scale metabolic reconstruction tools. *bioRxiv*, page 558411.

Michael, L. S. and Kargi, F. (2002). *Bioprocess Engineering: Basic Concepts*. Prentice-Hall International, Upper Saddle River, NJ, USA.

Missiakas, D., Mayer, M. P., Lemaire, M., Georgopoulos, C., and Raina, S. (1997). Modulation of the *Escherichia coli*σ^E (RpoE) heat-shock transcription-factor activity by the RseA, RseB and RseC proteins. *Molecular microbiology*, **24**(2), 355–371.

Mock, J., Zheng, Y., Mueller, A. P., Ly, S., Tran, L., Segovia, S., Nagaraju, S., Köpke, M., Dürre, P., and Thauer, R. K. (2015). Energy conservation associated with ethanol formation from H_2 and CO_2 in *Clostridium autoethanogenum* involving electron bifurcation. *Journal of Bacteriology*, **197**(18), 2965–2980.

Mölder, F., Jablonski, K. P., Letcher, B., Hall, M. B., Tomkins-Tinch, C. H., Sochat, V., Forster, J., Lee, S., Twardziok, S. O., Kanitz, A., *et al.* (2021). Sustainable data analysis with Snakemake. *F1000Research*, **10**.

Molitor, B., Richter, H., Martin, M. E., Jensen, R. O., Juminaga, A., Mihalcea, C., and Angenent, L. T. (2016a). Carbon recovery by fermentation of CO-rich off gases–turning steel mills into biorefineries. *Bioresource Technology*, **215**, 386–396.

Molitor, B., Kirchner, K., Henrich, A. W., Schmitz, S., and Rosenbaum, M. A. (2016b). Expanding the molecular toolkit for the homoacetogen *Clostridium ljungdahlii*. *Scientific Reports*, **6**(1), 1–10.

Molitor, B., Mishra, A., and Angenent, L. T. (2019). Power-to-protein: Converting renewable electric power and carbon dioxide into single cell protein with a two-stage bioprocess. *Energy & Environmental Science*, **12**(12), 3515–3521.

Molla, K. A. and Yang, Y. (2019). CRISPR/Cas-mediated base editing: Technical considerations and practical applications. *Trends in Biotechnology*, **37**(10), 1121–1142.

Möller, S., Croning, M. D., and Apweiler, R. (2001). Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics*, **17**(7), 646–653.

Monk, J. M., Lloyd, C. J., Brunk, E., Mih, N., Sastry, A., King, Z., Takeuchi, R., Nomura, W., Zhang, Z., and Mori, H. (2017). iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nature Biotechnology*, **35**(10), 904.

Müller, A. L., Gu, W., Patsalo, V., Deutzmann, J. S., Williamson, J. R., and Spormann, A. M. (2021). An alternative resource allocation strategy in the chemolithoautotrophic archaeon *Methanococcus maripaludis*. *Proceedings of the National Academy of Sciences*, **118**(16).

Müller, K., Städter, M., Rachow, F., Hoffmannbeck, D., and Schmeißer, D. (2013). Sabatier-based CO_2-methanation by catalytic conversion. *Environmental Earth Sciences*, **70**(8), 3771–3778.

Müller, V. (2019). New horizons in acetogenic conversion of one-carbon substrates and biological hydrogen storage. *Trends in Biotechnology*, **37**(12), 1344–1354.

Murray, N. E. (2000). Type I restriction systems: Sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiology and Molecular Biology Reviews*, **64**(2), 412–434.

Murugaiyan, J., Eravci, M., Weise, C., Roesler, U., Sprague, L. D., Neubauer, H., and Wareth, G. (2020). Pan-proteomic analysis and elucidation of protein abundance among the closely related brucella species, brucella abortus and brucella melitensis. *Biomolecules*, **10**(6), 836.

Nagarajan, H., Sahin, M., Nogales, J., Latif, H., Lovley, D. R., Ebrahim, A., and Zengler, K. (2013). Characterizing acetogenic metabolism using a genome-scale metabolic reconstruction of *Clostridium ljungdahlii*. *Microbial Cell Factories*, **12**(1), 1–13.

Nagaraju, S., Davies, N. K., Walker, D. J. F., Köpke, M., and Simpson, S. D. (2016). Genome editing of *Clostridium autoethanogenum* using CRISPR/Cas9. *Biotechnology for Biofuels*, **9**(1), 1–8.

Nangle, S. N., Sakimoto, K. K., Silver, P. A., and Nocera, D. G. (2017). Biological-inorganic hybrid systems as a generalized platform for chemical production. *Current Opinion in Chemical Biology*, **41**, 107–113.

National Center for Biotechnology Information (NCBI)[Internet] (1988). National Center for Biotechnology Information. `https://www.nasa.gov/nh/pluto-the-other-red-planet`.

Nazarenko, V. V., Remeeva, A., Yudenko, A., Kovalev, K., Dubenko, A., Goncharov, I. M., Kuzmichev, P., Rogachev, A. V., Buslaev, P., Borshchevskiy, V., *et al.* (2019). A thermostable flavin-based fluorescent protein from *Chloroflexus aggregans*: A framework for ultra-high resolution structural studies. *Photochemical & Photobiological Sciences*, **18**(7), 1793–1805.

Nazem-Bokaee, H., Gopalakrishnan, S., Ferry, J. G., Wood, T. K., and Maranas, C. D. (2016). Assessing methanotrophy and carbon fixation for biofuel production by *Methanosarcina acetivorans*. *Microbial Cell Factories*, **15**(1), 10.

Nevin, K. P., Hensley, S. A., Franks, A. E., Summers, Z. M., Ou, J., Woodard, T. L., Snoeyenbos-West, O. L., and Lovley, D. R. (2011). Electrosynthesis of organic compounds from carbon dioxide is catalyzed by a diversity of acetogenic microorganisms. *Applied and Environmental Microbiology*, **77**(9), 2882–2886.

Nielsen, J. (2001). Metabolic engineering. *Applied Microbiology and Biotechnology*, **55**(3), 263–283.

Nielsen, J. (2003). It is all about metabolic fluxes. *Journal of Bacteriology*, **185**(24), 7031–7035.

Nilsson, A. and Nielsen, J. (2017). Genome scale metabolic modeling of cancer. *Metabolic Engineering*, **43**, 103–112.

Nilsson, A., Nielsen, J., and Palsson, B. Ø. (2017a). Metabolic models of protein allocation call for the kinetome. *Cell Systems*, **5**(6), 538–541.

Nilsson, A., Mardinoglu, A., and Nielsen, J. (2017b). Predicting growth of the healthy infant using a genome scale metabolic model. *NPJ Systems Biology and Applications*, **3**(1), 3.

Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K. Y., *et al.* (2016). Targeted nucleotide editing using

hybrid prokaryotic and vertebrate adaptive immune systems. *Science*, **353**(6305), aaf8729.

Nölling, J. and Reeve, J. N. (1997). Growth-and substrate-dependent transcription of the formate dehydrogenase (*fdhcab*) operon in *Methanobacterium thermoformicicum* Z-245. *Journal of Bacteriology*, **179**(3), 899–908.

Nölling, J., Frijlink, M., and de Vos, W. M. (1991). Isolation and characterization of plasmids from different strains of methanobacterium thermoformicicum. *Microbiology*, **137**(8), 1981–1986.

Nölling, J., Eeden, F. J. v., Eggen, R. I., and de Vos, W. M. (1992). Modular organization of related archaeal plasmids encoding different restriction-modification systems in methanobacterium thermoformicicum. *Nucleic Acids Research*, **20**(24), 6501–6507.

Nölling, J., Hahn, D., Ludwig, W., and De Vos, W. M. (1993). Phylogenetic analysis of thermophilic *Methanobacterium* sp.: Evidence for a formate-utilizing ancestor. *Systematic and Applied Microbiology*, **16**(2), 208–215.

Oberhardt, M. A., Palsson, B. Ø., and Papin, J. A. (2009). Applications of genome-scale metabolic reconstructions. *Molecular Systems Biology*, **5**(1), 320.

Oberlies, G., Fuchs, G., and Thauer, R. K. (1980). Acetate thiokinase and the assimilation of acetate in *Methanobacterium thermoautotrophicum*. *Archives of Microbiology*, **128**(2), 248–252.

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. (1999). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, **27**(1), 29–34.

Oliveira Monteiro, L. M., Saraiva, J. P., Brizola Toscan, R., Stadler, P. F., Silva-Rocha, R., and Nunes da Rocha, U. (2022). PredicTF: Prediction of bacterial transcription factors in complex microbial communities using deep learning. *Environmental Microbiome*, **17**(1), 1–11.

Orosz, A., BOROS, I., and VENETIANER, P. (1991). Analysis of the complex transcription termination region of the *Escherichia coli rrnB* gene. *European Journal of Biochemistry*, **201**(3), 653–659.

Orth, J. D., Thiele, I., and Palsson, B. Ø. (2010). What is flux balance analysis? *Nature Biotechnology*, **28**(3), 245–248.

Ozbakir, H. F., Anderson, N. T., Fan, K.-C., and Mukherjee, A. (2019). Beyond the green fluorescent protein: Biomolecular reporters for anaerobic and deep-tissue imaging. *Bioconjugate Chemistry*, **31**(2), 293–302.

O'Brien, E. J. and Palsson, B. Ø. (2015). Computing the functional proteome: Recent progress and future prospects for genome-scale models. *Current Opinion in Biotechnology*, **34**, 125–134.

O'Brien, E. J., Monk, J. M., and Palsson, B. Ø. (2015). Using genome-scale models to predict biological capabilities. *Cell*, **161**(5), 971–987.

Palsson, B. Ø. (2015). *Systems Biology: Constraint-Based Reconstruction and Analysis*. Cambridge University Press.

Parncutt, R. (2019). The human cost of anthropogenic global warming: Semi-quantitative prediction and the 1,000-tonne rule. *Frontiers in Psychology*, page 2323.

Patil, K. R., Rocha, I., Förster, J., and Nielsen, J. (2005). Evolutionary programming as a platform for *in silico* metabolic engineering. *BMC bioinformatics*, **6**(1), 1–12.

PATRIC (2014). `https://www.patricbrc.org/view/Genome/187420.15`.

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, **14**(4), 417–419.

Pavan, M., Reinmets, K., Garg, S., Mueller, A. P., Marcellin, E., Köpke, M., and Valgepea, K. (2022). Advances in systems metabolic engineering of autotrophic carbon oxide-fixing biocatalysts towards a circular economy. *Metabolic Engineering*.

Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D. J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., *et al.* (2019). The pride database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Research*, **47**(D1), D442–D450.

Peterson, J. R., Thor, S., Kohler, L., Kohler, P. R., Metcalf, W. W., and Luthey-Schulten, Z. (2016). Genome-wide gene expression and RNA half-life measurements allow predictions of regulation and metabolic behavior in *Methanosarcina acetivorans*. *BMC Genomics*, **17**(1), 924.

Petitjean, C., Deschamps, P., López-García, P., Moreira, D., and Brochier-Armanet, C. (2015). Extending the conserved phylogenetic core of Archaea disentangles the evolution of the third domain of life. *Molecular Biology and Evolution*, **32**(5), 1242–1254.

Pfeifer, K., Ergal, İ., Koller, M., Basen, M., Schuster, B., and Simon, K.-M. R. (2021). Archaea biotechnology. *Biotechnology Advances*, **47**, 107668.

Pharkya, P., Burgard, A. P., and Maranas, C. D. (2004). Optstrain: A computational framework for redesign of microbial production systems. *Genome research*, **14**(11), 2367–2376.

Pihl, T. D., Sharma, S., and Reeve, J. N. (1994). Growth phase-dependent transcription of the genes that encode the two methyl coenzyme m reductase isoenzymes and N^5-methyltetrahydromethanopterin: Coenzyme M methyltransferase in *Methanobacterium thermoautotrophicum* ΔH. *Journal of Bacteriology*, **176**(20), 6384–6391.

Piles, M., Fernandez-Lozano, C., Velasco-Galilea, M., González-Rodríguez, O., Sánchez, J. P., Torrallardona, D., Ballester, M., and Quintanilla, R. (2019). Machine learning applied to transcriptomic data to identify genes associated with feed efficiency in pigs. *Genetics Selection Evolution*, **51**(1), 1–15.

Pinu, F. R., Granucci, N., Daniell, J., Han, T.-L., Carneiro, S., Rocha, I., Nielsen, J., and Villas-Boas, S. G. (2018). Metabolite secretion in microorganisms: The theory of metabolic overflow put to the test. *Metabolomics*, **14**(4), 43.

Poudel, S., Colman, D. R., Fixen, K. R., Ledbetter, R. N., Zheng, Y., Pence, N., Seefeldt, L. C., Peters, J. W., Harwood, C. S., and Boyd, E. S. (2018). Electron transfer to nitrogenase in different genomic and metabolic backgrounds. *Journal of Bacteriology*, **200**(10), e00757–17.

Prathiviraj, R. and Chellapandi, P. (2020). Modeling a global regulatory network of *Methanothermobacter thermautotrophicus* strain ΔH. *Network Modeling Analysis in Health Informatics and Bioinformatics*, **9**(1), 1–15.

Prieto, F. J. M. and Millero, F. J. (2002). The values of $pK_1 + pK_2$ for the dissociation of carbonic acid in seawater. *Geochimica et Cosmochimica Acta*, **66**(14), 2529–2540.

Rachbauer, L., Beyer, R., Bochmann, G., and Fuchs, W. (2017). Characteristics of adapted hydrogenotrophic community during biomethanation. *Science of the Total Environment*, **595**, 912–919.

Ragsdale, S. W. and Pierce, E. (2008). Acetogenesis and the Wood–Ljungdahl pathway of CO_2 fixation. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, **1784**(12), 1873–1898.

Ranganathan, S., Suthers, P. F., and Maranas, C. D. (2010). OptForce: An optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Computational Biology*, **6**(4), e1000744.

Rees, H. A. and Liu, D. R. (2018). Base editing: Precision chemistry on the genome and transcriptome of living cells. *Nature Reviews Genetics*, **19**(12), 770–788.

Resendis-Antonio, O., Checa, A., and Encarnación, S. (2010). Modeling core metabolism in cancer cells: Surveying the topology underlying the Warburg effect. *PloS One*, **5**(8), e12383.

Resing, K. A. and Ahn, N. G. (2005). Proteomics strategies for protein identification. *FEBS Letters*, **579**(4), 885–889.

Richards, M. A., Lie, T. J., Zhang, J., Ragsdale, S. W., Leigh, J. A., and Price, N. D. (2016). Exploring hydrogenotrophic methanogenesis: A genome scale metabolic reconstruction of *Methanococcus maripaludis*. *Journal of Bacteriology*, **198**(24), 3379–3390.

Richter, H., Molitor, B., Wei, H., Chen, W., Aristilde, L., and Angenent, L. (2016). Ethanol production in syngas-fermenting *Clostridium ljungdahlii* is controlled by thermodynamics rather than by enzyme expression. *Energy & Environmental Science*, **9**(7), 2392–2399.

Rinke, C., Chuvochina, M., Mussig, A. J., Chaumeil, P.-A., Davín, A. A., Waite, D. W., Whitman, W. B., Parks, D. H., and Hugenholtz, P. (2021). A standardized archaeal taxonomy for the genome taxonomy database. *Nature Microbiology*, **6**(7), 946–959.

Rittmann, S., Seifert, A., and Herwig, C. (2012). Quantitative analysis of media dilution rate effects on *Methanothermobacter marburgensis* grown in continuous culture on H_2 and CO_2. *Biomass and Bioenergy*, **36**, 293–301.

Rittmann, S., Seifert, A., and Herwig, C. (2015). Essential prerequisites for successful bioprocess development of biological CH_4 production from CO_2 and H_2. *Critical Reviews in Biotechnology*, **35**(2), 141–151.

Roberts, R. J., Vincze, T., Posfai, J., and Macelis, D. (2015). REBASE—a database for DNA restriction and modification: Enzymes, genes and genomes. *Nucleic Acids Research*, **43**(D1), D298–D299.

Robertson, C. E., Harris, J. K., Spear, J. R., and Pace, N. R. (2005). Phylogenetic diversity and ecology of environmental archaea. *Current Opinion in Microbiology*, **8**(6), 638–642.

Rocha, I., Maia, P., Evangelista, P., Vilaça, P., Soares, S., Pinto, J. P., Nielsen, J., Patil, K. R., Ferreira, E. C., and Rocha, M. (2010). OptFlux: An open-source software platform for *in silico* metabolic engineering. *BMC systems Biology*, **4**(1), 1–12.

Römer, M., Eichner, J., Dräger, A., Wrzodek, C., Wrzodek, F., and Zell, A. (2016). ZBIT bioinformatics toolbox: A web-platform for systems biology and expression data analysis. *PloS One*, **11**(2), e0149263.

Roy, S., Cherevotan, A., and Peter, S. C. (2018). Thermochemical CO_2 hydrogenation to single carbon products: Scientific and technological challenges. *ACS Energy Letters*, **3**(8), 1938–1966.

Rusmanis, D., O'Shea, R., Wall, D. M., and Murphy, J. D. (2019). Biological hydrogen methanation systems–an overview of design and efficiency. *Bioengineered*, **10**(1), 604–634.

Saa, P. A. and Nielsen, L. K. (2017). Formulation, construction and analysis of kinetic models of metabolism: A review of modelling frameworks. *Biotechnology Advances*, **35**(8), 981–1003.

Saa, P. A., Cortés, M. P., López, J., Bustos, D., Maass, A., and Agosin, E. (2019). Expanding metabolic capabilities using novel pathway designs: Computational tools and case studies. *Biotechnology Journal*, **14**(9), 1800734.

Saha, R., Chowdhury, A., and Maranas, C. D. (2014). Recent advances in the reconstruction of metabolic models and integration of omics data. *Current Opinion in Biotechnology*, **29**, 39–45.

Samal, A. and Martin, O. C. (2011). Randomizing genome-scale metabolic networks. *Plos One*, **6**(7), e22295.

Samal, A., Matias Rodrigues, J. F., Jost, J., Martin, O. C., and Wagner, A. (2010). Genotype networks in metabolic reaction spaces. *BMC Systems Biology*, **4**(1), 30.

Sambrook, J. (1989). Molecular cloning: A laboratory manual. *Synthetic oligonucleotides*.

Sambrook, J. and Russell, D. W. (2006). Preparation and transformation of competent *E. coli* using calcium chloride. *Cold Spring Harbor Protocols*, **2006**(1), pdb–prot3932.

Sánchez, B. J., Zhang, C., Nilsson, A., Lahtvee, P., Kerkhoven, E. J., and Nielsen, J. (2017). Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Molecular Systems Biology*, **13**(8), 935.

Sawers, G. and Watson, G. (1998). A glycyl radical solution: Oxygen-dependent interconversion of pyruvate formate-lyase. *Molecular Microbiology*, **29**(4), 945–954.

Schauer, N. L., Ferry, J. G., Honek, J. F., Orme-Johnson, W. H., and Walsh, C. (1986). Mechanistic studies of the coenzyme F_{420}-reducing formate dehydrogenase from *Methanobacterium formicicum*. *Biochemistry*, **25**(22), 7163–7168.

Schellenberger, J., Zielinski, D. C., Choi, W., Madireddi, S., Portnoy, V., Scott, D. A., Reed, J. L., and Osterman, A. L. (2012). Predicting outcomes of steady-state ^{13}C isotope tracing experiments using Monte Carlo sampling. *BMC Systems Biology*, **6**(1), 9.

Schiebahn, S., Grube, T., Robinius, M., Tietze, V., Kumar, B., and Stolten, D. (2015). Power to gas: Technological overview, systems analysis and economic assessment for

a case study in Germany. *International Journal of Hydrogen Energy*, **40**(12), 4285–4294.

Schill, N., Van Gulik, W., Voisard, D., and Von Stockar, U. (1996). Continuous cultures limited by a gaseous substrate: Development of a simple, unstructured mathematical model and experimental verification with *Methanobacterium thermoautotrophicum*. *Biotechnology and Bioengineering*, **51**(6), 645–658.

Schill, N. A., Liu, J., and Stockar, U. v. (1999). Thermodynamic analysis of growth of *Methanobacterium thermoautotrophicum*. *Biotechnology and Bioengineering*, **64**(1), 74–81.

Schlegel, K. and Müller, V. (2013). Evolution of Na^+ and H^+ bioenergetics in methanogenic Archaea. *Biochemical Society Transactions*, **41**(1).

Schmehl, M., Jahn, A., Meyer zu Vilsendorf, A., Hennecke, S., Masepohl, B., Schuppler, M., Marxer, M., Oelze, J., and Klipp, W. (1993). Identification of a new class of nitrogen fixation genes in *Rhodobacter capsalatus*: A putative membrane complex involved in electron transport to nitrogenase. *Molecular and General Genetics MGG*, **241**(5), 602–615.

Schmidt, B. J., Ebrahim, A., Metz, T. O., Adkins, J. N., Palsson, B. Ø., and Hyduke, D. R. (2013a). GIM^{3}E: Condition-specific models of cellular metabolism developed from metabolomics and expression data. *Bioinformatics*, **29**(22), 2900–2908.

Schmidt, B. J., Papin, J. A., and Musante, C. J. (2013b). Mechanistic systems modeling to guide drug discovery and development. *Drug Discovery Today*, **18**(3-4), 116–127.

Schönheit, P., Moll, J., and Thauer, R. K. (1980). Growth parameters (K_s, μ_{max}, Y_s) of *Methanobacterium thermoautotrophicum*. *Archives of Microbiology*, **127**(1), 59–65.

Schuchmann, K. and Müller, V. (2014). Autotrophy at the thermodynamic limit of life: A model for energy conservation in acetogenic bacteria. *Nature Reviews Microbiology*, **12**(12), 809–821.

Schuetz, R., Kuepfer, L., and Sauer, U. (2007). Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Molecular Systems Biology*, **3**(1), 119.

Schulz, C., Kumelj, T., Karlsen, E., and Almaas, E. (2021). Genome-scale metabolic modelling when changes in environmental conditions affect biomass composition. *PLoS Computational Biology*, **17**(5), e1008528.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature*, **473**(7347), 337–342.

Seaver, S. M., Liu, F., Zhang, Q., Jeffryes, J., Faria, J. P., Edirisinghe, J. N., Mundy, M., Chia, N., Noor, E., Beber, M. E., *et al.* (2021). The ModelSEED Biochemistry Database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes. *Nucleic Acids Research*, **49**(D1), D575–D588.

Seely, R. J., Krueger, R. D., and Fahrney, D. E. (1983). Cyclic-2, 3-diphosphoglycerate levels in *Methanobacterium thermoautotrophicum* reflect inorganic phosphate availability. *Biochemical and Biophysical Research Communications*, **116**(3), 1125–1128.

Seifert, A., Rittmann, S., and Herwig, C. (2014). Analysis of process related factors to increase volumetric productivity and quality of biomethane with *Methanothermobacter marburgensis*. *Applied Energy*, **132**, 155–162.

Seifritz, C., Daniel, S. L., Gößner, A., and Drake, H. L. (1993). Nitrate as a preferred electron sink for the acetogen *Clostridium thermoaceticum*. *Journal of Bacteriology*, **175**(24), 8008–8013.

Selkov, E., Maltsev, N., Olsen, G. J., Overbeek, R., Whitman, W. B., *et al.* (1997). A reconstruction of the metabolism of *Methanococcus jannaschii* from sequence data. *Gene*, **197**(1-2), GC11–GC26.

Selle, K. and Barrangou, R. (2015). Harnessing CRISPR–Cas systems for bacterial genome editing. *Trends in microbiology*, **23**(4), 225–232.

Shalvarjian, K. E. and Nayak, D. D. (2021). Transcriptional regulation of methanogenic metabolism in Archaea. *Current Opinion in Microbiology*, **60**, 8–15.

Sheehan, R., McCarver, A. C., Isom, C. E., Karr, E. A., and Lessner, D. J. (2015). The *Methanosarcina acetivorans* thioredoxin system activates DNA binding of the redox-sensitive transcriptional regulator MsvR. *Journal of Industrial Microbiology and Biotechnology*, **42**(6), 965–969.

Shin, J., Kang, S., Song, Y., Jin, S., Lee, J. S., Lee, J.-K., Kim, D. R., Kim, S. C., Cho, S., and Cho, B.-K. (2019). Genome engineering of *Eubacterium limosum* using expanded genetic tools and the CRISPR-Cas9 system. *ACS Synthetic Biology*, **8**(9), 2059–2068.

Shintani, M., Sanchez, Z. K., and Kimbara, K. (2015). Genomics of microbial plasmids: Classification and identification based on replication and transfer systems and host taxonomy. *Frontiers in Microbiology*, **6**, 242.

Shlomi, T., Cabili, M. N., Herrgård, M. J., Palsson, B. Ø., and Ruppin, E. (2008). Network-based prediction of human tissue-specific metabolism. *Nature Biotechnology*, **26**(9), 1003.

Shoaie, S., Karlsson, F., Mardinoglu, A., Nookaew, I., Bordel, S., and Nielsen, J. (2013). Understanding the interactions between bacteria in the human gut through metabolic modeling. *Scientific Reports*, **3**, 2532.

Smith, D. R., Doucette-Stamm, L., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., and Gilbert, K. (1997). Complete genome sequence of *Methanobacterium thermoautotrophicum* ΔH: Functional analysis and comparative genomics. *Journal of Bacteriology*, **179**(22), 7135–7155.

Song, Q., Grene, R., Heath, L. S., and Li, S. (2017). Identification of regulatory modules in genome scale transcription regulatory networks. *BMC Systems Biology*, **11**(1), 1–18.

Song, Y., Lee, J. S., Shin, J., Lee, G. M., Jin, S., Kang, S., Lee, J.-K., Kim, D. R., Lee, E. Y., Kim, S. C., *et al.* (2020). Functional cooperation of the glycine synthase-reductase and Wood–Ljungdahl pathways for autotrophic growth of *Clostridium drakei*. *Proceedings of the National Academy of Sciences*, **117**(13), 7516–7523.

Stark, R., Grzelak, M., and Hadfield, J. (2019). RNA sequencing: The teenage years. *Nature Reviews Genetics*, **20**(11), 631–656.

Steenbakkers, P. J., Geerts, W. J., Ayman-Oz, N. A., and Keltjens, J. T. (2006). Identification of pseudomurein cell wall binding domains. *Molecular Microbiology*, **62**(6), 1618–1630.

Sterner, M. and Specht, M. (2021). Power-to-gas and power-to-x—the history and results of developing a new storage concept. *Energies*, **14**(20), 6594.

Stettler, R., Pfister, P., and Leisinger, T. (1995). Characterization of a plasmid carried by *Methanobacterium thermoautotrophicum* ZH3, a methanogen closely related to *Methanobacterium thermoautotrophicum* Marburg. *Systematic and Applied Microbiology*, **17**(4), 484–491.

Streett, H. E., Kalis, K. M., and Papoutsakis, E. T. (2019). A strongly fluorescing anaerobic reporter and protein-tagging system for *Clostridium* organisms based on the fluorescence-activating and absorption-shifting tag protein (FAST). *Applied and Environmental Microbiology*, **85**(14), e00622–19.

Suthers, P. F., Foster, C. J., Sarkar, D., Wang, L., and Maranas, C. D. (2021). Recent advances in constraint and machine learning-based metabolic modeling by leveraging stoichiometric balances, thermodynamic feasibility and kinetic law formalisms. *Metabolic Engineering*, **63**, 13–33.

Tan, J., Zhang, F., Karcher, D., and Bock, R. (2019). Engineering of high-precision base editors for site-specific single nucleotide replacement. *Nature Communications*, **10**(1), 1–10.

Tan, Y., Liu, J., Chen, X., Zheng, H., and Li, F. (2013). RNA-seq-based comparative transcriptome analysis of the syngas-utilizing bacterium *Clostridium ljungdahlii* DSM 13528 grown autotrophically and heterotrophically. *Molecular BioSystems*, **9**(11), 2775–2784.

Tanner, R. S., McINERNEY, M. J., and Nagle, D. (1989). Formate auxotroph of *Methanobacterium thermoautotrophicum* Marburg. *Journal of Bacteriology*, **171**(12), 6534–6538.

Tersteegen, A., Linder, D., Thauer, R. K., and Hedderich, R. (1997). Structures and functions of four anabolic 2-oxoacid oxidoreductases in *Methanobacterium thermoautotrophicum*. *European Journal of Biochemistry*, **244**(3), 862–868.

Thauer, R. K. (2012). The Wolfe cycle comes full circle. *Proceedings of the National Academy of Sciences*, **109**(38), 15084–15085.

Thauer, R. K., Käfer, B., Jungermann, K., and Zähringer, M. (1974). The reaction of the iron-sulfur protein hydrogenase with carbon monoxide. *European Journal of Biochemistry*, **42**(2), 447–452.

Thauer, R. K., Jungermann, K., and Decker, K. (1977). Energy conservation in chemotrophic anaerobic bacteria. *Bacteriological Reviews*, **41**(1), 100–180.

Thauer, R. K., Kaster, A.-K., Seedorf, H., Buckel, W., and Hedderich, R. (2008). Methanogenic Archaea: Ecologically relevant differences in energy conservation. *Nature Reviews Microbiology*, **6**(8), 579–591.

Thiele, I. and Palsson, B. Ø. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*, **5**(1), 93.

Thiele, I., Jamshidi, N., Fleming, R. M., and Palsson, B. Ø. (2009). Genome-scale reconstruction of *Escherichia coli*'s transcriptional and translational machinery: A knowledge base, its mathematical formulation, and its functional characterization. *PLoS Computational Biology*, **5**(3), e1000312.

Thiele, I., Fleming, R. M., Que, R., Bordbar, A., Diep, D., and Palsson, B. Ø. (2012). Multiscale modeling of metabolism and macromolecular synthesis in *E. coli* and its application to the evolution of codon usage. *PloS One*, **7**(9), e45635.

Thor, S., Peterson, J. R., and Luthey-Schulten, Z. (2017). Genome-scale metabolic modeling of Archaea lends insight into diversity of metabolic function. *Archaea*, **2017**.

Tong, Y., Whitford, C. M., Robertsen, H. L., Blin, K., JØrgensen, T. S., Klitgaard, A. K., Gren, T., Jiang, X., Weber, T., and Lee, S. Y. (2019). Highly efficient DSB-free base editing for streptomycetes with CRISPR-BEST. *Proceedings of the National Academy of Sciences*, **116**(41), 20366–20375.

Touzel, J. P., Petroff, D., Maestrojuan, G. M., Prensier, G., and Albagnac, G. (1988). Isolation and characterization of a thermophilic *Methanobacterium* able to use formate, the strain FTF. *Archives of Microbiology*, **149**(4), 291–296.

Tremblay, P.-L., Zhang, T., Dar, S. A., Leang, C., and Lovley, D. R. (2012). The Rnf complex of *Clostridium ljungdahlii* is a proton-translocating ferredoxin: NAD^+ oxidoreductase essential for autotrophic growth. *MBio*, **4**(1), e00406–12.

Tsigkinopoulou, A., Baker, S. M., and Breitling, R. (2017). Respectful modeling: Addressing uncertainty in dynamic system models for molecular biology. *Trends in Biotechnology*, **35**(6), 518–529.

Tsoka, S., Simon, D., and Ouzounis, C. A. (2004). Automated metabolic reconstruction for *Methanococcus jannaschii*. *Archaea*, **1**(4), 223–229.

Ueki, T., Nevin, K. P., Woodard, T. L., and Lovley, D. R. (2014). Converting carbon dioxide to butyrate with an engineered strain of *Clostridium ljungdahlii*. *MBio*, **5**(5), e01636–14.

United Nations, editor (2015). *Framework Convention on Climate Change*, volume 21 of *21st Conference of the Parties*, Paris, France. United Nations.

United States Environmental Protection Agency (2021). Overview of greenhouse gases. `https://www.epa.gov/ghgemissions/overview-greenhouse-gases`. Accessed: 2022-03-08.

Valgepea, K., Loi, K. Q., Behrendorff, J. B., de SP Lemgruber, R., Plan, M., Hodson, M. P., Köpke, M., Nielsen, L. K., and Marcellin, E. (2017a). Arginine deiminase pathway provides ATP and boosts growth of the gas-fermenting acetogen *Clostridium autoethanogenum*. *Metabolic Engineering*, **41**, 202–211.

Valgepea, K., Lemgruber, R. d. S. P., Meaghan, K., Palfreyman, R. W., Abdalla, T., Heijstra, B. D., Behrendorff, J. B., Tappel, R., Köpke, M., Simpson, S. D., Nielsen, L. K., and Marcellin, E. (2017b). Maintenance of ATP homeostasis triggers metabolic shifts in gas-fermenting acetogens. *Cell Systems*, **4**(5), 505–515. e5.

Valgepea, K., de Souza Pinto Lemgruber, R., Abdalla, T., Binos, S., Takemori, N., Takemori, A., Tanaka, Y., Tappel, R., Köpke, M., Simpson, S. D., *et al.* (2018). H_2 drives metabolic rearrangements in gas-fermenting *Clostridium autoethanogenum*. *Biotechnology for Biofuels*, **11**(1), 1–15.

Van Der Laan, G. P. and Beenackers, A. (1999). Kinetics and selectivity of the Fischer–Tropsch synthesis: A literature review. *Catalysis Reviews*, **41**(3-4), 255–318.

Vento, J. M., Crook, N., and Beisel, C. L. (2019). Barriers to genome editing with CRISPR in bacteria. *Journal of Industrial Microbiology and Biotechnology*, **46**(9-10), 1327–1341.

Vermeij, P., van der Steen, R., Keltjens, J. T., Vogels, G. D., and Leisinger, T. (1996). Coenzyme F_{390} synthetase from *Methanobacterium thermoautotrophicum* Marburg belongs to the superfamily of adenylate-forming enzymes. *Journal of Bacteriology*, **178**(2), 505–510.

Vermeij, P., Pennings, J., Maassen, S. M., Keltjens, J. T., and Vogels, G. D. (1997). Cellular levels of factor 390 and methanogenic enzymes during growth of *Methanobacterium thermoautotrophicum* ΔH. *Journal of Bacteriology*, **179**(21), 6640–6648.

Visweswaran, G. R. R., Dijkstra, B. W., and Kok, J. (2011). Murein and pseudomurein cell wall binding domains of bacteria and archaea—a comparative view. *Applied Microbiology and Biotechnology*, **92**(5), 921–928.

Waardenberg, A. J. and Field, M. A. (2019). consensusDE: An R package for assessing consensus of multiple RNA-seq algorithms with RUV correction. *PeerJ*, **7**, e8206.

Wang, N. S. and Stephanopoulos, G. (1983). Application of macroscopic balances to the identification of gross measurement errors. *Biotechnology and Bioengineering*, **25**(9), 2177–2208.

Wang, Y., Wang, S., Chen, W., Song, L., Zhang, Y., Shen, Z., Yu, F., Li, M., and Ji, Q. (2018). CRISPR-Cas9 and CRISPR-assisted cytidine deaminase enable precise and efficient genome editing in *Klebsiella pneumoniae*. *Applied and Environmental Microbiology*, **84**(23), e01834–18.

Wasserfallen, A., Nölling, J., Pfister, P., Reeve, J., and De Macario, E. C. (2000). Phylogenetic analysis of 18 thermophilic *Methanobacterium* isolates supports the proposals to create a new genus, *Methanothermobacter* gen. nov., and to reclassify several isolates in three species, *Methanothermobacter thermautotrophicus*

comb. nov., *Methanothermobacter wolfeii* comb. nov., and *Methanothermobacter marburgensis* sp. nov. *International Journal of Systematic and Evolutionary Microbiology*, **50**(1), 43–53.

Wei, Y., Li, B., Prakash, D., Ferry, J. G., Elliott, S. J., and Stubbe, J. (2015). A ferredoxin disulfide reductase delivers electrons to the *Methanosarcina barkeri* class III ribonucleotide reductase. *Biochemistry*, **54**(47), 7019–7028.

Wessely, F., Bartl, M., Guthke, R., Li, P., Schuster, S., and Kaleta, C. (2011). Optimal regulatory strategies for metabolic pathways in *Escherichia coli* depending on protein costs. *Molecular Systems Biology*, **7**(1), 515.

Westphal, L., Wiechmann, A., Baker, J., Minton, N. P., and Müller, V. (2018). The Rnf complex is an energy-coupled transhydrogenase essential to reversibly link cellular NADH and ferredoxin pools in the acetogen *Acetobacterium woodii*. *Journal of Bacteriology*, **200**(21), e00357–18.

White, R. H. (1997). Purine biosynthesis in the domain Archaea without folates or modified folates. *Journal of Bacteriology*, **179**(10), 3374–3377.

White, W. B. and Ferry, J. G. (1992). Identification of formate dehydrogenase-specific mRNA species and nucleotide sequence of the *fdhC* gene of *Methanobacterium formicicum*. *Journal of Bacteriology*, **174**(15), 4997–5004.

Wiegel, J. (1990). Temperature spans for growth: Hypothesis and discussion. *FEMS Microbiology Letters*, **75**(2-3), 155–169.

Wingen, M., Jaeger, K.-E., Gensch, T., and Drepper, T. (2017). Novel thermostable flavin-binding fluorescent proteins from thermophilic organisms. *Photochemistry and Photobiology*, **93**(3), 849–856.

Wood, H. G. (1991). Life with CO or CO_2 and H_2 as a source of carbon and energy. *The FASEB Journal*, **5**(2), 156–163.

Wood, H. G., Ragsdale, S. W., and Pezacka, E. (1986). The acetyl-CoA pathway of autotrophic growth. *FEMS Microbiology Reviews*, **2**(4), 345–362.

Woolston, B. M., Emerson, D. F., Currie, D. H., and Stephanopoulos, G. (2018). Rediverting carbon flux in *Clostridium ljungdahlii* using CRISPR interference (CRISPRi). *Metabolic Engineering*, **48**, 243–253.

Xavier, J. C., Patil, K. R., and Rocha, I. (2017). Integration of biomass formulations of genome-scale metabolic models with experimental data reveals universally essential cofactors in prokaryotes. *Metabolic Engineering*, **39**, 200–208.

Xia, P.-F., Casini, I., Schulz, S., Klask, C.-M., Angenent, L. T., and Molitor, B. (2020). Reprogramming acetogenic bacteria with CRISPR-targeted base editing *via* deamination. *ACS Synthetic Biology*, **9**(8), 2162–2171.

Xu, T., Li, Y., Shi, Z., Hemme, C. L., Li, Y., Zhu, Y., Van Nostrand, J. D., He, Z., and Zhou, J. (2015). Efficient genome editing in *Clostridium cellulolyticum via* CRISPR-Cas9 nickase. *Applied and Environmental Microbiology*, **81**(13), 4423–4431.

Yamamoto, K., Tachibana, A., Dhavises, G., Tanaka, T., Taniguchi, M., and Oi, S. (1989). Characterization of a thermophilic formate-utilizing methanogen, *Methanobacterium thermoformicicum* strain SF-4. *Agricultural and Biological Chemistry*, **53**(2), 533–534.

Yizhak, K., Benyamini, T., Liebermeister, W., Ruppin, E., and Shlomi, T. (2010). Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model. *Bioinformatics*, **26**(12), i255–i260.

Yoshinaga, M. Y., Gagen, E. J., Wörmer, L., Broda, N. K., Meador, T. B., Wendt, J., Thomm, M., and Hinrichs, K.-U. (2015). *Methanothermobacter thermautotrophicus* modulates its membrane lipids in response to hydrogen and nutrient availability. *Frontiers in Microbiology*, **6**.

Yuan, Z., Eden, M. R., and Gani, R. (2016). Toward the development and deployment of large-scale carbon dioxide capture and conversion processes. *Industrial & Engineering Chemistry Research*, **55**(12), 3383–3419.

Yura, T. and Nakahigashi, K. (1999). Regulation of the heat-shock response. *Current Opinion in Microbiology*, **2**(2), 153–158.

Zabranska, J. and Pokorna, D. (2017). Bioconversion of carbon dioxide to methane using hydrogen and hydrogenotrophic methanogens. *Biotechnology Advances*.

Zeikus, J. and Wolfe, R. (1972). *Methanobacterium thermoautotrophicus* sp. n., an anaerobic, autotrophic, extreme thermophile. *Journal of Bacteriology*, **109**(2), 707–713.

Zeikus, J., Ben-Bassat, A., and Hegge, P. (1980). Microbiology of methanogenesis in thermal, volcanic environments. *Journal of Bacteriology*, **143**(1), 432–440.

Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., Volz, S. E., Joung, J., Van Der Oost, J., Regev, A., *et al.* (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, **163**(3), 759–771.

Zhang, C. and Hua, Q. (2016). Applications of genome-scale metabolic models in biotechnology and systems medicine. *Frontiers in Physiology*, **6**, 413.

Zhang, S., Knaack, S., and Roy, S. (2022). Enabling studies of genome-scale regulatory network evolution in large phylogenies with MRTLE. In *Yeast Functional Genomics*, pages 439–455. Springer.

Zhang, S.-W., Gou, W.-L., and Li, Y. (2017). Prediction of metabolic fluxes from gene expression data with huber penalty convex optimization function. *Molecular BioSystems*, **13**(5), 901–909.

Zhang, W., Li, F., and Nie, L. (2010). Integrating multiple 'omics' analysis for microbial biology: Application and methodologies. *Microbiology*, **156**(2), 287–301.

Zhao, R., Liu, Y., Zhang, H., Chai, C., Wang, J., Jiang, W., and Gu, Y. (2019). CRISPR-Cas12a-mediated gene deletion and regulation in *Clostridium ljungdahlii* and its application in carbon flux redirection in synthesis gas fermentation. *ACS Synthetic Biology*, **8**(10), 2270–2279.

Zhao, Y., McDonell, V., and Samuelsen, S. (2020). Assessment of the combustion performance of a room furnace operating on pipeline natural gas mixed with simulated biogas or hydrogen. *International Journal of Hydrogen Energy*, **45**(19), 11368–11379.

Zhilina, T. and Ilarionov, S. (1984). Isolation and comparative characteristics of methanogenic bacteria assimilating formate with the description of *Methanobacterium thermoformicicum* sp. nov. *Mikrobiologiya.*, **53**(5), 785–790.

Zinder, S. H. and Koch, M. (1984). Non-aceticlastic methanogenesis from acetate: Acetate oxidation by a thermophilic syntrophic coculture. *Archives of Microbiology*, **138**(3), 263–272.

Zomorrodi, A. R., Suthers, P. F., Ranganathan, S., and Maranas, C. D. (2012). Mathematical optimization applications in metabolic networks. *Metabolic Engineering*, **14**(6), 672–686.

Zumft, W. G. (1997). Cell biology and molecular basis of denitrification. *Microbiology and Molecular Biology Reviews*, **61**(4), 533–616.

Zuñiga, C., Li, C.-T., Huelsman, T., Levering, J., Zielinski, D. C., McConnell, B. O., Long, C. P., Knoshaug, E. P., Guarnieri, M. T., and Antoniewicz, M. R. (2016). Genome-scale metabolic model for the green alga *Chlorella vulgaris* UTEX 395 accurately predicts phenotypes under autotrophic, heterotrophic, and mixotrophic growth conditions. *Plant Physiology*, page pp. 00593.2016.

Zuo, E., Sun, Y., Yuan, T., He, B., Zhou, C., Ying, W., Liu, J., Wei, W., Zeng, R., Li, Y., *et al.* (2020). A rationally engineered cytosine base editor retains high on-target activity while reducing both DNA and RNA off-target effects. *Nature Methods*, **17**(6), 600–604.

Zur, H., Ruppin, E., and Shlomi, T. (2010). iMAT: An integrative metabolic analysis tool. *Bioinformatics*, **26**(24), 3140–3142.