A Comparative Genomic Analysis of Hydrocarbon Synthesis in *Desulfovibrio* sp.

Submitted by Peggy Marie Madeleine Dousseaud to the University of Exeter
as a thesis for the degree of
Doctor of Philosophy in Biological Sciences
In April 2018

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: ……..

Abstract

To fulfil global energy demand and to mitigate economical, geopolitical and ecological challenges associated with fossil fuel utilisation, the energy sector is moving towards greater use of sustainable and environmentally friendly energy sources, including biofuels. The ideal transport biofuel would be hydrocarbons that are identical to fossil petroleum. However, to date characterised hydrocarbon biosynthetic pathways include a decarbonylation or decarboxylation reaction, which involves the loss of one carbon resulting in odd-numbered carbon chain hydrocarbons. This carbon loss decreases carbon efficiency for alkane production, which reduces microbial fuel economic competitiveness. Therefore, it is key that new pathways for alkane production are identified.

The sulphate-reducing bacteria genus *Desulfovibrio* was previously reported to synthesise even-numbered carbon chain alkanes, which suggests an alternative route for alkane production without carbon loss. This investigation aimed to verify *Desulfovibrio* alkane biosynthesis and characterise the possible synthetic pathway. Ten *Desulfovibrio* strains, representing seven species, were screened for alkane synthesis using isotopically labelled growth media. The ability to produce alkanes within the *Desulfovibrio* genus was confirmed and was shown to be strain-specific under a set of culture conditions. The biogenic alkanes detected were octadecane (C18), nonadecane (C19) and eicosane (C20), with a predominance of even-numbered carbon chain alkanes. Fatty acid analysis of *Desulfovibrio* strains showed an alkane biosynthetic pathway was unlikely to involve a decarbonylation or decarboxylation step. A novel hypothesis was therefore proposed that alkane biosynthesis by *Desulfovibrio* follows a metabolic route, which has not previously been characterised, involving a series of reduction reactions from the fatty acid pool.

The characterisation of the putative *Desulfovibrio* hydrogenation pathway for alkane biosynthesis was undertaken *via* a target-directed genome mining approach. The genomic DNA of nine *Desulfovibrio* spp. was purified, sequenced, *de novo* assembled and annotated. Seven of these nine genomes are unpublished to date. No homologs of previously characterised alkane biosynthetic enzymes from bacteria were *in silico* identified in the genomes and proteomes of alkane producing *Desulfovibrio* spp., suggesting that *Desulfovibrio* alkane biosynthetic pathway is likely to be catalysed by currently uncharacterised enzymes.

The 16S rRNA-based phylogeny of *Desulfovibrio* spp. supported the hypothesis that the *Desulfovibrio* alkane biosynthetic pathway was acquired by a common ancestral strain *via* horizontal gene transfer. The ability of *Desulfovibrio* to produce alkanes was therefore hypothesised to be due to the presence of recruited genes encoding enzymes involved in alkane synthesis. A comparative genomic analysis intersecting six-alkane producing and four non-alkane producing *Desulfovibrio* genomes resulted in the *in silico* identification of 33 hypothetical proteins considered with high confidence to be exclusive to alkane producing *Desulfovibrio* strains. A novel hypothetical *Desulfovibrio* alkane biosynthetic pathway was proposed involving a V-type ATPase, an uncharacterised protein, named as a putative reductase in this investigation, and a putative methyltransferase, which were predicted to be exclusive to alkane producing *Desulfovibrio* spp. The inorganic phosphates resulting from the ATP hydrolysis catalysed by the V-type ATPase would be involved in a reaction with fatty alcohols to form alkyl phosphates, which are putative activated intermediates required for the hydrogenation route from fatty alcohols to alkanes. The putative reductase and the methyltransferase, predicted to share similar structural features with known alkane-binding proteins, would subsequently reduce alkyl phosphates to alkanes and to *iso*-alkanes respectively. Empirical investigation of the candidate molecular basis function in *Desulfovibrio* alkane biosynthesis was undertaken. The *Desulfovibrio* alkane biosynthetic pathway remains to be fully characterised.

Acknowledgements

To everyone who made this Ph.D. possible, I wish to sincerely thank you.

I wish to express my extreme gratitude to my supervisors, Prof. John Love, Dr Stephen Michell and Prof. Rob Lee for giving me the opportunity to accomplish this Ph.D., for their guidance and supervision throughout this project. I am also immensely grateful to Dr Sara Burton for her invaluable support, belief and encouragement throughout the time writing this thesis. Furthermore, I would like to thank Royal Dutch Shell, notably Dr Jeremy Shears and Dr David Parker, for funding this research project.

I also would like to thank everyone who contributed to the work of this research project. During my Ph.D., I have been unduly fortunate to meet Prof. Judy Wall (University of Missouri) and to work within her laboratory for one week. I wish to thank everyone in the Wall laboratory for their very warm welcome and knowledge transfer about the genus *Desulfovibrio*. I am extremely grateful to Dr. Deborah Salmon for her technical training on gas chromatography and mass spectrometry handling and data analysis, her belief and friendship. I am thankful to Exeter Sequencing Service for their assistance with DNA sequencing and data analysis, namely Dr Karen Moore, Miss Audrey Farbos and Mr Paul O'Neill. I also would like to thank Dr Jens Benninghoff (Unilever) for his assistance in protein analysis.

I wish to thank everyone in the Exeter Microbial Biofuels Group and present occupants of the Mezzanine laboratory especially to Dr Richard Tennant, Dr Christine Sambles, Dr Chloe Singleton, Dr Anja Nenninger and Ms Monica Ayine. I am extremely grateful for their help, endless encouragement and friendship. I am particularly grateful to Mr Jamie Gilman for always having offset the "seesaw" throughout this Ph.D. I have undoubtedly enjoyed working with them all.

I also would like to thank my friends for their continuous support and cheer, especially Miss Constance de Taillac, Miss Audrey Farbos, Miss Candice François, Miss Chloé Laborde, Dr Anke Lange and Dr Thomas Monfeuga.

Finally, I would like to thank my family to whom I dedicate this thesis. Despite the distance, their boundless love, belief and support were essential for me; they got me to where I am now. Un immense merci.

Table of Contents

List of Figures

List of Tables

Definition

Atom-% excess	A measure of the abundance of a stable isotope, in percent of atoms i given isotope

List of Abbreviations

A m^{-2}	Ampere per square metre
ADO	Aldehyde deformylating oxygenase
AMP	Adenosine monophosphate
AAR	Acyl-ACP reductase
ATCC	American Type Culture Collection
ATP	Adenosine triphosphate
BLAST	Basic Local Alignment Search Tool
boe	Barrel of oil equivalent
bp	Base pair
CAR	Carboxylic acid reductase
COG	Clusters of Orthologous Group
CPU	Central processing unit
DAP	2,6-diaminopimelic acid
DCM	Dichloromethane
ΔG^0	Gibbs free energy
DNA	Deoxyribonucleic acid
DSM	Deutsche Sammlung von Mikroorganisme (German Collection of Microorganism)
$E^{0'}$	Standard electrode potential
EIC	Extracted ion chromatogram
FAP	Fatty acid photodecarboxylase
FAR	Fatty acid reductase
fmol	Femtomole
GC-MS	Gas Chromatography- Mass Spectrometry
GHz	Gigahertz
g l^{-1}	Gramme per Litre
g mol^{-1}	Gramme per mole
h	Hour
HMM	Hidden Markov Model
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kb	Kilobase
kDa	Kilodalton
kJ mol^{-1}	Kilojoule per mole
kV cm^{-1}	Kilovolt per centimetre

LB	Luria-Bertani
m	Metre
m/z	Mass to charge ratio
mA	Milliampere
Mb	Megabase pair
MEC	Microbial electrolysis cell
mg l^{-1}	Milligramme per Litre
min	Minute
ml	Millilitre
ml min^{-1}	Millilitre per minute
mM	Millimolar
mRNA	Messenger ribonucleic acid
ms	Milliseconde
mV	Millivolt
NAD^+/NADH	Nicotinamide adenine dinucleotide oxidised/reduced
NCIMB	National Collections of Industrial, Marine and Food Bacteria
ng	Nanogramme
ng μl^{-1}	Nanogramme per microlitre
nm	Nanometre
nmol	Nanomole
OD	Optical density
Ols	Olefin synthase
Pa	Pascal
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
pH	Potential hydrogen
ppm	Part-per-million
PVDF	Polyvinylidene difluoride
RAM	Random-access memory
RBS	Ribisome binding site
RNA	Ribonucleic acid
rpm	Rotation per minute
rRNA	Ribosomal ribonucleic acid
s	Seconde
S	Svedberg unit
SDS-PAGE	Sodium dodecyl sulfate - Polyacrilamide gel electrophoresis
sp., spp.	Species
SRO	Sulphate-reducing organism
U	Unit
μg	Microgramme
μg ml^{-1}	Microgramme per millilitre
μl	Microlitre
μm	Micrometre

*µ*M	Micromolar
UV	Ultra violet
V	Volt
V-type ATPase	Vacuolar-type ATPase
v/v	Volume per volume
w/v	Weight per volume
x g	Times gravity

1. Introduction

1.1. Biofuels; Sustainable and Environmental - Friendly Alternative Sources of Energy

In 2016, world energy consumption reached 90.8 million barrels of oil equivalent (boe), with crude oil remaining the most consumed fuel accounting for 33.3 % of world energy consumption (British Petroleum, 2017). Due to population and economic growth, global energy demand is forecast to increase by 28 %, rising up to 126.9 billion boe, with 103.5 million barrels of crude oil per day in 2040 (International Energy Agency, 2016). Crude oil is a mixture of linear alkanes, branched alkanes, cyclic alkanes, alkenes and aromatic hydrocarbons (Battin-Leclerc & Blurock, 2011). Generated by fractional distillation of crude oil, transportation fuels differ by the ratio and the molecular weight of hydrocarbon constituents. Gasoline, used in spark-ignition internal combustion engines, contains between 40 % and 60 % volume per volume (v/v) of linear, branched and cyclic hydrocarbons and between 20 % and 40 % (v/v) of aromatic hydrocarbons (Lee *et al.,* 2008). In gasoline, the hydrocarbons range from C_4 to C_{12} (Sawyer, 1993). Diesel, used in compression-ignition internal combustion engines, possesses approximately 75 % (v/v) of linear, branched and cyclic hydrocarbons and 25 % (v/v) of aromatic hydrocarbons (Lee *et al.,* 2008). In Diesel, the carbon chain-length distribution is comprised from C_9 to C_{23} (Lee *et al.,* 2008). The eponymous “jet fuel”, used in continuous combustion “jet” engines, is mainly composed of linear, branched and cyclic hydrocarbons with a maximum of 25 % (v/v) of aromatic hydrocarbons (Kallio *et al.,* 2014). The carbon number of jet fuel hydrocarbons varies from 6 to 16 (Kallio *et al.,* 2014). Although variable, the composition of retail transport fuels has to comply with accredited physical and chemical fuel properties. In the European Union, the BS EN 228 and BS EN 590 standards respectively cover gasoline and Diesel quality. Civil jet fuel production conforms with the international standard specification ASTM D1655.

Crude oil, coal and natural gas are classified as fossil fuels. Synthesised from the remains of living material over geological time, fossil fuels are a finite and geographically localised resource, often causing political instability in petroleum-producing regions such as the Arctic, Middle East and Venezuela. Fossil fuel combustion emits greenhouse gasses into the atmosphere. Global carbon dioxide (CO_2) emissions from fossil fuel utilisation reached 33.4 billion metric tons of CO_2 in 2016 and is forecast to rise by 34 %

to 43.2 billion metric tons of CO_2 in 2040 (International Energy Agency, 2016). Increases in atmospheric CO_2 concentration largely contribute to global warming and ocean acidification (Matthews *et al.*, 2009; Doney *et al.*, 2009). Since the Industrial revolution, the global average of combined land and ocean surface temperature rose by 0.85 °C and ocean acidity increased by 26 % (Intergovernmental Panel on Climate Change, 2014). Climate change is linked to extreme weather events and rise in sea level (Rosenzweig *et al.*, 2001; Intergovernmental Panel on Climate Change, 2014). Climate change also poses a threat to food security (Wheeler & Von Braun, 2013) and biodiversity (Bellard *et al.,* 2014). The need to reduce global CO_2 emissions was accredited with the adoption by 190 countries of a legally binding agreement to keep global warming below 2 °C, at the United Nations Framework Convention on Climate Change Conference (COP21) in 2015. Thus, continued exploitation of fossil fuels to supply increasing energy demand raises economical, geopolitical and ecological challenges.

To fulfil the global demand in energy and to mitigate challenges associated with fossil fuel utilisation, the global energy sector is in transition towards sustainable and environmental-friendly energy sources, including biofuels. The term “biofuels” refers to fuels produced from biomass feedstock. Primary biofuels are unprocessed biomass, such as wood pellets. Secondary biofuels are solid, liquid or gaseous fuels derived from processed biomass, for instance charcoal, ethanol, biodiesel or hydrogen (H_2). Since 2005, the global production of liquid biofuels for transportation rose by 14.1 %, with 563 million boe produced in 2016 (British Petroleum, 2017). To support this low-carbon transition and to achieve their climate agreement pledges, countries decreed policies to shift towards biofuels in the transportation sector. The European Renewable Energy Directive (2009/28/EC) mandates European Union countries to substitute at least 10 % of their transport fuels with biofuels by 2020. The US Energy Independence and Security Act of 2007 established a Renewable Fuel Standard 2 (RFS2) mandating 36 billion gallons of biofuels to be consumed per year by 2022. International collaborations have also been initiated for the development of sustainable low-carbon alternative fuels in transport. The Biofuture Platform, proposed by the government of Brazil, brings together 20 countries, organizations, academia and the private sector to facilitate the dialogue and work for the development of biofuels in transport. The Below50 project from the World Business Council for Sustainable Development aims to promote the market for biofuels that emit at least 50 % fewer greenhouse gasses than fossil fuels.

1.2. Current Liquid Fuel Alternatives for Transportation; Additive Biofuels

1.2.1. Conventional Biofuels

Conventional biofuels refer to ethanol and biodiesel produced from feedstocks that could be used as food and/or feed. Ethanol is produced by fermentation of starch or sugars produced by crops such as *Zea mays* (maize), *Saccharum officinarum* (sugarcane) and *Manihot esculenta* (cassava), followed by distillation (Kang *et al.,* 2014). Biodiesel is produced by the transesterification of predominantly plant oils or to a lesser extent of animal fat with primary alcohol to form fatty acid alkyl-esters (Agarwal, 2007). Biodiesel raw materials by preponderance include the oils from *Glycine max* (soya bean), *Brassica napus* (rapeseed), *Arecaceae* (palm tree), animal fat (from beef, chicken or pork), *Zea mays* (maize) and *Helianthus* (sunflower) (Souza *et al.*, 2017). Conventional biofuel production is well-established and implemented at a commercial scale. Ethanol has been produced on an industrial scale in Brazil for over 40 years due to the 'ProAlcool' program (Goldemberg *et al.*, 2004). Since 2016, Brazil counts 383 biorefineries for production of conventional ethanol (USDA Foreign Agricultural Service, 2016). European Union is the largest producer of conventional biodiesel, with approximatively 200 biorefineries in 2016 (USDA Foreign Agricultural Service, 2017). In total, conventional biofuels contributed to approximatively 4 % of the world transport fuels in 2016 (International Energy Agency, 2017).

Ethanol can be blended as an anti-knock additive with gasoline in existing spark ignition engines at a maximal concentration of 27.5 %, according to the current legislations (International Renewable Energy Agency, 2015). Blending gasoline with ethanol offers a higher energy efficiency than pure gasoline, due to the high octane number (a standard measure of the fuel anti-knock properties) of ethanol, which allows high compression ratios in spark ignition engines. However, ethanol use is limited due to its high hygroscopicity, which causes the corrosion of production infrastructures and fossil fuel engines (Rabinovitch-Deere *et al.*, 2013). Ethanol can be combusted as fuel only in designed engines known as flex-fuel engines. Moreover, ethanol energy density is only 70 % of gasoline energy density (International Renewable Energy Agency, 2016).

Current legislations allow biodiesel to be blended with petroleum-based Diesel at a concentration up to 7 % in existing compression ignition engines (International Renewable Energy Agency, 2016). In comparison to mineral Diesel fuel, biodiesel has a similar energy density and a higher cetane number (a standard measure of the Diesel

fuel ignition value; Balat & Balat, 2010). However, combustion of biodiesel has been shown to produce higher toxic nitrogen oxide emissions than combustion of mineral Diesel fuel. Other limiting factors to the use of biodiesel in compression ignition engines are the poor oxidative stability of biodiesel, which increases biodiesel viscosity once exposed to oxygen, and the poor cold flow proprieties, which induce the formation of solid crystals in the fuel tank in cold areas (Lanjekar & Deshmukh, 2016).

The conversion of food crops to biofuels has brought about direct competition over land and water use for food production (Mohr & Raman, 2013). Consequently, global food prices have risen (Graham-Rowe, 2011). Moreover, the profitability of fuel crops has led to the intensification of fuel crop plantations, reducing habitat complexity and causing loss of biodiversity (Prescott *et al.,* 2015). In Malaysia, a large part of the Borneo's jungle has been substituted by palm plantation (Graham-Rowe, 2011; Kircher, 2015). The intensive plantations of fuel crops also require intensive chemical inputs, including fertilisers and pesticides. Intensive application of fertilisers and pesticides offsets the environmental benefits of biofuels, by polluting water and increasing nitrogen oxide emissions from nitrogen-based fertilisers (Elobeid *et al.*, 2013). Another concern of conventional biofuel production relates to the net greenhouse gas emission reduction. In May 2018, a legislative proposal was adopted by the European Parliament for a better management of land use, land use change and forestry (LULUCF) sector to mitigate the climate change, amending Regulation (EU) No 525/2013 and Decision No 529/2013/EU. This new legislation aims to promote the "no debit rule", where greenhouse gas emissions from land use are entirely compensated by removal of CO_2 from the atmosphere through action in the sector, such as planting new trees.

1.2.2. Advanced Biofuels

The cumulative concerns over conventional biofuels have stimulated the development of advanced biofuels. Advanced biofuels are produced from non edible biomass, mainly lignocellulosic material (Ho *et al.,* 2014). Lignocellulose is a constituent of the plant cell wall and is comprised of cellulose (β 1-4 linked chains of glucose molecules), hemicellulose (various 5-6 carbon sugars such as arabinose, xylose, galactose and glucose) and lignin (polymers of three phenol alcohols). The lignocellulosic polysaccharides are converted into biofuels by either a bio-chemical or a thermo-chemical process. The bio-chemical process involves either chemical or bio-chemical saccharification and fermentation, leading to the production of ethanol or

butanol (Ho *et al.,* 2014). Butanol has a similar energy density to gasoline and a higher octane number than gasoline (Ndaba *et al.*, 2015). According to the European Union standard (EN 228), butanol can be blended with gasoline up to 15 %. The thermo-chemical process involves gasification or pyrolysis (Naik *et al.,* 2010) leading to different types of bio-oils and syngas such as methanol, dimethyl ether (DME) and refined Fischer-Tropsch liquids (FTL) (Ho *et al.,* 2014). Methanol and DME, with an energy density 50 % lower than gasoline, can be blended at low concentration with gasoline, up to 3 % according to the European Union standard (EN 228). This blending limit is mainly due to human toxicity and corrosive proprieties of methanol and DME. FTL fuels have a similar energy density to Diesel and can be blended with jet-fuel up to a concentration of 50 % (International Renewable Energy Agency, 2016).

Another potential feedstock for advanced biofuels are microalgae. Microalgae, such as *Chlorella* spp., synthesise a substantial quantity of various lipids. Depending on the strain, microalgae average lipid content varies between 1 % to 70 % of dry biomass (Mata *et al.,* 2010). For example, the relative lipid abundance of the green algae *Botryococcus braunii* can constitute up to 62 % of its dry mass (Metzger & Largeau, 2005). Microalgal lipids can be either used to produce fatty acid alkyl-ester biodiesel by transesterification, or hydrogenated then upgraded by catalytic deoxygenation to produce hydrotreated vegetable oil biodiesel. Hydrotreated vegetable oil biodiesel has a similar energy density to Diesel and can be blended with jet-fuel up to a concentration of 50 % (International Renewable Energy Agency, 2016).

Advanced biofuels are at various stages of commercial development. The commercialisation of advanced biofuels is mainly hampered by feedstock availability, which can be affected by climatic seasons, and by production technology performance and cost effectiveness. For example, production of microalgal biodiesel is not yet economically viable, due to the low biomass concentration and costly downstream purification and conversion processes. To increase biomass growth and therefore to achieve optimal oil yields, microalgae are grown in closed bioreactors, requiring high cost of construction and maintenance. Moreover, harvesting, concentrating and dewatering microalgal biomass require high inputs of fossil fuel derived energy, consequently generating CO_2 emissions, and are costly (Medipally *et al.,* 2014). Downstream conversions of microbial lipids into hydrotreated vegetable oil biodiesel involve costly hydrogenation conditions (high pressure and temperature) and expensive hydrogen quantities, limiting the commercial development of hydrotreated vegetable oil biodiesel (Vásquez *et al.*, 2017). Commercialisation of a cost-competitive lignocellulosic

butanol compared to fossil fuels is impeded by low butanol yield of the production technology. The low butanol yield is partly due to butanol toxicity to the fermentation organism, and the acetate and ethanol formation, fermentation by-products. Additionally, butanol recovery from low concentration yield requires high energy inputs and is consequently costly (Ndaba *et al.*, 2015). Lignocellulosic ethanol and biomethanol advanced biofuels have been recently commercialised. Beta Renewables S.p.A. (Tortona, Italy) produces ethanol from agricultural residues (wheat straw and rice straw) and from *Arundo donax* (reed) at commercial scale using the enzymatic conversion process PROESA™ technology. Enerkem Alberta Biofuels (Edmonton, Canada) produces lignocellulosic ethanol and biomethanol by gasification of solid household wastes. Another advanced biofuel, a liquid oil known as "Biocrude" has been commercialised by Ensyn Technologies Inc. (Ottawa, Canada). Biocrude is produced from wood-derived feedstock using the patented fast pyrolysis process RTP® technology. Biocrude can be subsequently upgraded to transport biofuels.

Hence, conventional biofuels and current advanced biofuels are essentially additives due to their chemical and physical proprieties, and/or their production technologies. The ideal transport fuel alternative would be a substitute for fossil fuel, compatible with existing production infrastructures and vehicle engines, with similar energy density to fossil fuels and posing no threat to global food security; *i.e.* biologically produced hydrocarbons that are identical to fossil petroleum (Howard *et al.,* 2013). Biosynthetic pathways of petroleum replicate hydrocarbons, such as alkanes and alkenes, therefore represent an attractive target for the development of superior, advanced biofuels.

1.3. Biogenic Alka(e)nes; Advanced Biofuel Alternatives for Fossil Fuel Substitution in Transportation

Alka(e)nes biosynthesis is widespread is nature, with different pathways having evolved independently across the phylogeny. Presently, seven alka(e)ne anabolic pathways have been fully characterised. The alkane biosynthetic pathway in *Arabidopsis thaliana* is as yet incomplete and still under investigation.

1.3.1. Pathways for Alkane and Alkene Biosynthesis

Two alka(e)ne biosynthetic pathways have been characterised in cyanobacteria: the decarbonylation pathway (Figure 1.1A) and the α-olefin synthase (Ols) pathway (Figure 1.1B). *Synechococcus elogatus* PCC7942 and *Anabaena* sp. strain PCC7120 synthesise heptadecanes and pentadecanes *via* a decarbonylation pathway. Fatty acyl-acyl carrier proteins (acyl-ACPs) are reduced by an acyl-ACP reductase (AAR) to fatty aldehydes, which are in turn decarbonylated to alkanes by an aldehyde deformylating oxygenase (ADO) (Schrimer *et al.,* 2010; Li *et al.,* 2012). The cyanobacterium *Synechococcus* sp. strain PCC 7002 produces C_{19} alkenes from fatty acyl-ACPs by an elongation-decarboxylation mechanism catalysed by Ols (Mendez-Perez *et al.,* 2011).

In *Pseudomonas*, two enzyme families, with UndA and UndB as representative enzymes, have been characterised to catalyse the oxidative decarboxylation of medium chain (C_{10}-C_{14}) fatty acids to their corresponding (C_9-C_{13}) terminal alkenes (Figure 1.1C). UndA is a nonheme iron dependent oxidase (Rui *et al.*, 2014) and UndB is a membrane-bound desaturase-like enzyme (Rui *et al.*, 2015). The bacteria *Jeotgalicoccus* spp. also synthesise terminal olefin C_{19} alkenes *via* decarboxylation of free fatty acids catalysed by a P_{450} fatty acid decarboxylase, $OleT_{JE}$ (Rude *et al.,* 2011; Figure 1.1D).

Recently, the hydrocarbon anabolic pathway in the microalga *Chlorella variabilis* NC64A has been elucidated. *Chlorella variabilis* NC64A produce heptadecenes, heptadecanes and pentadecanes *via* a decarboxylation reaction, catalysed by a fatty acid photodecarboxylase (FAP) in response to blue light (Sorigué *et al.*, 2017; Figure 1.1E).

The OleABCD protein families were characterised in *Micrococcus luteus* to catalyse a head-to-head condensation of two fatty acyl-CoAs to produce long-chain alkenes (Beller *et al.,* 2010; Figure 1.1F). The OleABCD protein families were classified within thiolase, α/β-hydrolase, AMP-dependent ligase/synthetase, and short-chain dehydrogenase superfamilies respectively, and were shown to be present in a variety of bacteria (Sukovich *et al.,* 2010). OleA has been the most studied and catalysed the Claisen condensation of fatty acid derivatives to β-ketoacids which then undergo spontaneous decarbonylation to produce ketones (Frias *et al.,* 2011).

The biosynthetic pathway of very long chain cuticular hydrocarbons (C_{21} to C_{37}) in *Drosophila melanogaster* involves a reduction step catalysed by an acyl-CoA

reductase, which converts very long chain fatty acyl-CoAs into fatty aldehydes. In turn, very long chain fatty aldehydes are decarbonylated into alkanes or alkenes by an insect specific P_{450} oxidative decarbonylase of the CYP4G family (Reed *et al.*, 1994; Qiu *et al.*, 2012; Figure 1.1G).

A pathway for the synthesis of very long chain alkanes (C_{20} to C_{36}), major components of the cuticular waxes, in *Arabidopsis thaliana* has been proposed. Very long chain fatty acyl-CoAs could be reduced to aldehydes by a very-long-chain-acyl-CoA reductase (CER3), which are subsequently decarbonylated to very long chain alkanes by a very-long-chain-aldehyde decarbonylase (CER1). CER3 and CER1 could act synergistically as a hetero-dimer to convert fatty acyl-coA into alkanes, using the cytochrome b5 as redox co-factor (Millar *et al.*, 1999; Bernard *et al.*, 2012; Figure 1.1H).

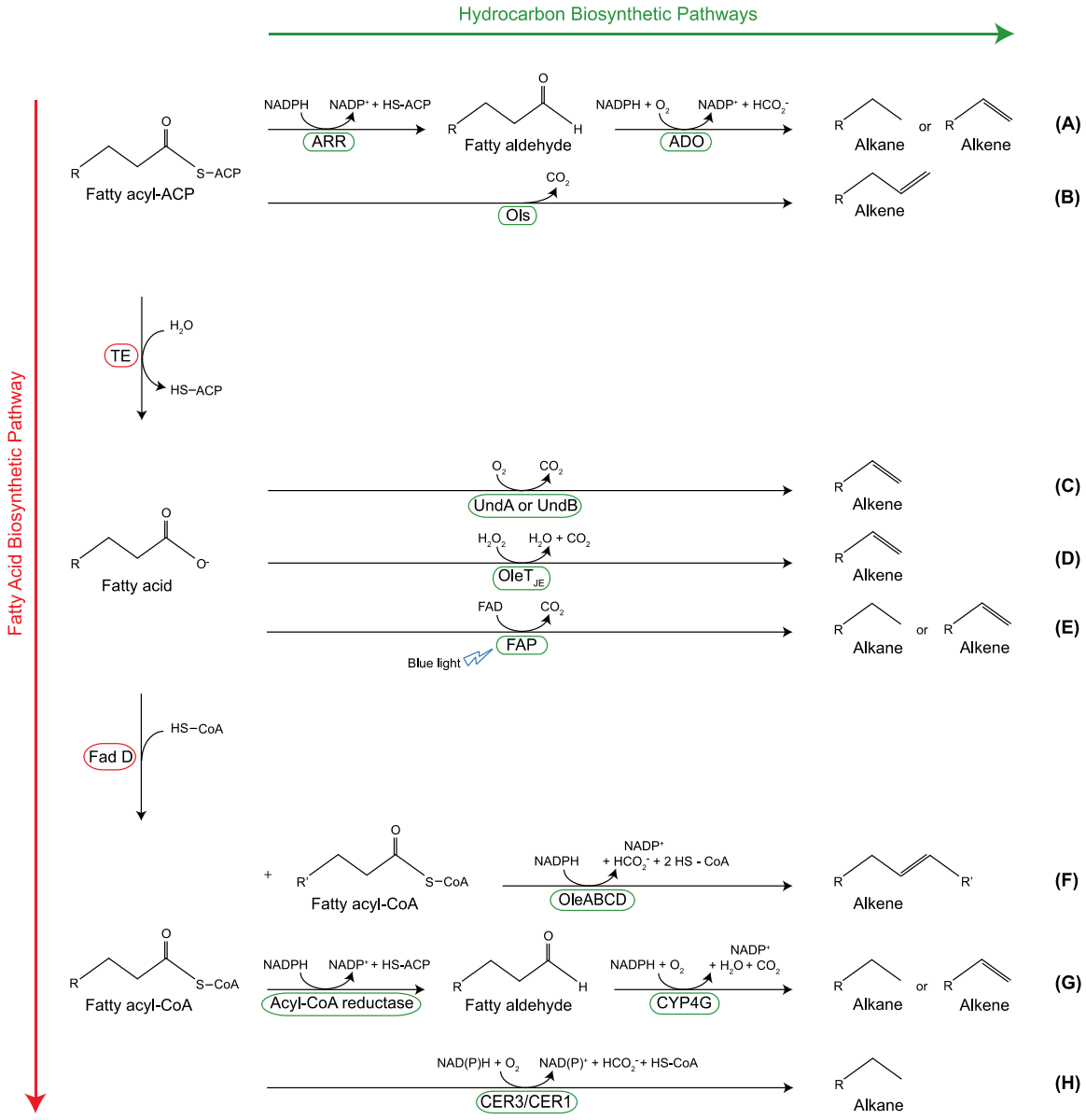


Figure 1.1. Pathways for alkane and alkene biosynthesis

Biogenic alkanes and alkenes derive from either fatty acyl-ACP, fatty acids or fatty acyl-CoA [on the left; thio-esterase (TE) and long-chain-fatty-acid-CoA ligase (Fad D)]. The alka(e)ne biosynthetic pathways characterised and under characterisation to date are shown: (A), the fatty acyl-ACP conversion to terminal alka(e)nes catalysed by acyl-ACP reductase (AAR) and aldehyde deformylating oxygenase (ADO; Schirmer *et al.,* 2010); (B), the fatty acyl-ACP elongation-decarboxylation to terminal alkenes catalysed by α-olefin synthase (Ols; Mendez-Perez *et al.,* 2011); (C), the fatty acid oxidative decarboxylation to terminal alkenes catalysed by either UndA or UndB decarboxylase (Rui *et al.*, 2014; Rui *et al.*, 2015); (D), the fatty acid decarboxylation to terminal alkenes catalysed by $OleT_{JE}$ P_{450} decarboxylase (Rude *et al.,* 2011); (E), the fatty acid decarboxylation to terminal alka(e)nes catalysed by fatty acid photodecarboxylase (FAP) in response to blue light (Sorigué *et al.*, 2017) ; (F), the long chain alkene synthesis *via* a head-to-head condensation of fatty acyl-CoA, catalysed by Ole ABCD (Frias *et al.,* 2011); (G) the long chain fatty acyl-CoA conversion to terminal alka(e)nes catalysed by acyl-CoA reductase and the insect CYP4G oxidative decarbonylase (Reed *et al.*, 1994); (H), the proposed decarbonylation pathway of long chain fatty acyl-CoA to terminal alkanes catalysed by the enzymatic complex CER1/CER3 (Bernard *et al.*, 2012).

1.3.2. Hindrances for direct fossil fuel substitution by biogenic hydrocarbons

There are some limits to the direct fossil fuel replacement by biogenic hydrocarbons. Biogenic hydrocarbon yields are not sufficient for industrial exploitation (Fu *et al.*, 2015). Moreover, anabolic hydrocarbon pathways produce a narrow range of hydrocarbons. For direct fossil fuel replacement, a hydrocarbon mixture of varied chain length and branched alkanes and alkenes is required (Howard *et al.,* 2013). Progress in metabolic engineering and synthetic biology have opened the possibility to overcome these limits in favour of viable biofuel production (Keasling, 2008; Lee *et al.,* 2008; Connor & Liao, 2009; Peralta-Yahya *et al.,* 2012; Howard *et al.,* 2013; Lee *et al.,* 2015). Metabolic engineering enables alteration of the metabolic pathways of an organism towards the production of a desired chemical at a desirable rate (Keasling, 2008; Lee *et al.,* 2012). Metabolic engineering requires a library of biological elements such as promoters, ribonucleic acids, proteins, *etc*. Synthetic biology is focused on the design and construction of these biological elements (Connor & Atsumi, 2010). The main challenge of the metabolic engineering is to maximise metabolic flux to desired chemicals without impairing cell viability, resulting in decreased productivity (Keasling, 2008; Janßen & Steinbüchel, 2014).

Synthetic pathways for the production of petroleum replica alka(e)nes have already been engineered and implemented in bacteria (Howard *et al.*, 2013; Crépin *et al.*, 2016), cyanobacteria (Peramuna *et al.*, 2015), yeasts (Zhou *et al.*, 2016) and fungi (Sinha *et al.*, 2017). One of the principal strategies in metabolic engineering for production of petroleum replica alka(e)nes is to optimise the fatty acid metabolic pathway in order to tailor and boost hydrocarbon production, as all characterised alka(e)nes biosynthetic pathways derive from the fatty acid metabolic pathway. Choi *et al.* altered *Escherichia coli* fatty acid pathway towards the synthesis of short chain fatty acids by deletion of the *fadE* gene to block the β-oxidation, by deletion of the *fadR* gene to indirectly enhance fatty acid biosynthesis, and by expression of a modified thioesterase to enhance fatty acyl-ACPs conversion to free fatty acids (Choi & Lee, 2013). The co-expression of *Clostridium acetobutylicum* fatty acyl-CoA reductase and *A. thaliana* fatty aldehyde decarbonylase (CER1) in the fatty acid metabolism engineered *E. coli* resulted in the production of 580.8 mg l^{-1} short chain alkanes (Choi & Lee, 2013). Presently, it is the highest titer of biogenic hydrocarbons achieved by an engineered *E. coli* (Wang & Zhu, 2018). Microorganisms with a highly efficient fatty acid metabolism are therefore potential microbial platforms for production of petroleum replica alka(e)nes. Recently, a

strain of *Aureobasidium melanogenum* was found to produce and secrete 38.24 g l^{-1} of crude heavy oils in a medium containing inulin, after 5 days in batch fermentation. The heterologous expression of an inulinase gene from *Kluyveromyces marximus* enabled the *A. melanogenum* strain to produce 43.0 g l^{-1} of extracellular crude heavy oils in a medium containing inulin, after 5 days in batch fermentation. Wild type and mutant *A. melanogenum* extracellular crude heavy oils is composed over 80 % of long chain alkanes (Xin *et al.*, 2017). However, it is important to highlight that the unit of measurement (g l^{-1}) used for alkane production provides insufficient information for industrial exploitation. An appropriate unit would be the amount of alkane moles produced per time unit, per biomass unit and per catalyst unit.

A relevant observation is that all characterised alka(e)ne biosynthetic pathways include a decarbonylation or decarboxylation step, which involves the loss of one carbon resulting in odd-numbered carbon chain hydrocarbons. The carbon loss decreases carbon efficiency for alkane production. A higher carbon efficiency where all the carbons are converted to alkanes would, as a direct consequence, increase microbial fuel yields up to 5 %. Moreover, this would enable microbial fuels to become economically competitive and fulfil the global retail fuel demand (Lee *et al.,* 2015). Therefore, it is key that new pathways for alkane production are identified.

The detection of even-numbered carbon chain alkanes produced by *D. desulfuricans* suggests an alternative route for alkane production without carbon loss. Sulphate-reducing bacteria were first reported to produce hydrocarbons by the notable founders of microbial ecology Jankowski and ZoBell (1944). Re-visited several times, the accumulated literature proposes that *D. desulfuricans* synthesise a wide range of intracellular and extracellular alkanes that possess carbon-chain lengths from C_{11} to C_{35} (Oppenheimer, 1965; Davis, 1968; Bagaeva & Chernova, 1994).

1.4. The Desulfovibrio Genus

Desulfovibrio (*Desulfovibrionaceae*, class *Deltaproteobacteria*) is a genus of gram-negative, mainly rod-shaped bacteria, approximatively 2.3 *μ*m in length and 0.7 *μ*m in diameter (Postgate, 1984; Figure 1.2). *Desulfovibrio* spp. do not produce spores (Postgate, 1984). *Desulfovibrio* bacteria contain a pigment, the desulfoviridin, which fluoresces red in alkaline pH and blue-green in acid pH under an exciting wavelength of 365 nm (Lee & Peck, 1971; Laue *et al.,* 2001).

1.4.1. Physiology

Desulfovibrio spp. belong to the sulphate-reducing organism (SRO) group, a highly diverse morphological and metabolical group of eubacteria and archaea. SRO are capable of anaerobic respiration by dissimilatory sulphate reduction. Sulphate is used as a terminal electron acceptor for dissimilation of H_2 or various organic compounds, resulting in release of energy to support growth (Postgate, 1984). SRO were also found to grow fermentatively in syntrophy with other micro-organisms through metabolic cooperation. The anaerobic respiration and the syntrophic fermentation are controlled by two distinct energy metabolisms (Walker *et al.,* 2009). In response to environment changes, SRO are able to switch between these two energy metabolisms. Energy metabolism in SRO shows high complexity and plasticity, involving many electron transport pathways from different carbon sources (Muyzer & Stams, 2008).

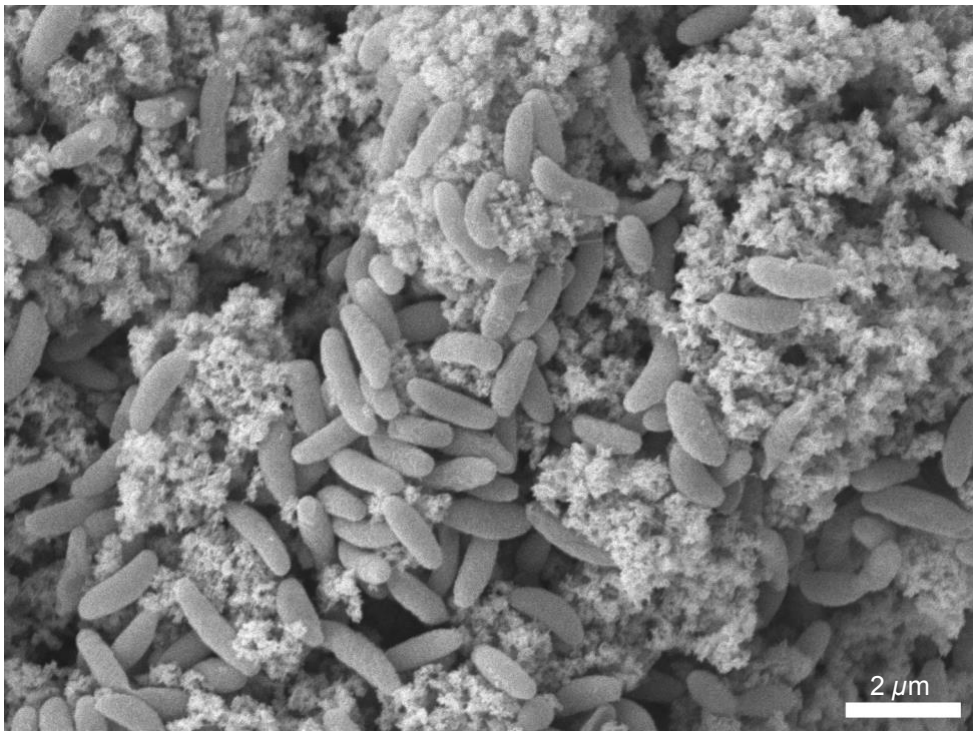


Figure 1.2. Scanning electron micrograph of *Desulfovibrio desulfuricans subsp. desulfuricans* strain California27.137.58326 (NCIMB 8326)

D. desulfuricans subsp. desulfuricans NCIMB 8326 was grown in planktonic cultures to stationary phase. Culture samples were chemically fixed with aldehydes, post-fixed with osmium tetroxide, dehydrated in ethanol, critical point dried and sputter coated with gold and palladium, prior visualisation under scanning electron microscope (*cf.* Chapter 2 - Materials and Methods; Section 2.2). The rods are *D. desulfuricans* cells, surrounded by iron sulphide precipitate.

Respiratory metabolism

Desulfovibrio spp. reduce sulphate in the cytoplasm under the control of soluble enzymes. Sulphate is a thermodynamically stable compound, with a very low redox potential ($E^{0'}$ = - 516 mV). The sulphate redox potential is too negative to allow sulphate reduction by intracellular electron carriers such as ferrodoxin or reduced nicotinamide adenine dinucleotide (NADH) present in sulphate reducers. Therefore, sulphate needs to be activated prior to be reduced. The activation step, catalysed by a sulphate adenyltransferase, consumes two adenosine triphosphate (ATP) molecules and produces adenosine phosphosulphate (APS) (Akagi & Campbell, 1962; Equation 1). An APS reductase then reduces APS to sulphite and adenosine monophosphate (AMP) (Peck Jr, 1961; Equation 2). Finally the sulphite is reduced to sulphide by a dissimilatory sulphite reductase (Kobayashi *et al.,* 1972; Equation 3). Overall, the sulphate reduction consumes eight moles of electron with eight and a half moles of proton (Keller & Wall, 2011; Equation 4).

$$SO_4^{2-} + ATP^{4-} + H^+ \rightarrow APS^{2-} + PPi \quad (1)$$

$$APS^{2-} + 2\ e^- + 0.5\ H^+ \rightarrow 0.5\ HSO_3^- + 0.5\ SO_3^{2-} + AMP^{2-} \quad (2)$$

$$0.5\ HSO_3^- + 0.5\ SO_3^{2-} + 6\ e^- + 7\ H^+ \rightarrow 0.5\ HS^- + 0.5\ H_2S + 3\ H_2O \quad (3)$$

$$SO_4^{2-} + ATP^4 + 8\ e^- + 8.5\ H^+ \rightarrow 0.5\ HS^- + 0.5\ H_2S + AMP^{2-} + 3\ H_2O + PPi \quad (4)$$

Although predominantly sulphate reducers (Plugge *et al.,* 2011), *Desulfovibrio* spp. can use other electron acceptors. *Desulfovibrio* spp. can perform dissimilatory reduction of other sulphur compounds to sulphide, such as sulphite (Findley & Akagi, 1969) and thiosulphate (Haschke & Campbell, 1971). *Desulfovibrio* spp. can also use nitrate and nitrite as electron acceptors (Seitz & Cypionka, 1986; Moura *et al.,* 1997). Moreover, *Desulfovibrio* spp. can reduce iron (Park *et al.,* 2008), uranyl (Lovley & Phillips, 1992), pertechnetate (Lloyd *et al.,* 1999), selenite (Tucker *et al.,* 1998), molybdate (Tucker *et al.,* 1998), chromate (Lovley & Phillips, 1994) and arsenate (Macy *et al.,* 2000). Some *Desulfovibrio* spp. are capable of aerobic respiration (Sigalevich & Cohen, 2000). *D. desulfuricans* ATCC 27774 can grow in the presence of 18 % O_2, which is approximatively the atmospheric oxygen content (Lobo *et al.,* 2007). However, aerobic respiration is more a protective ability against oxidative stress and leads to a poor or absent growth (Dolla *et al.,* 2006). Organic compounds can also act as terminal electron

acceptors for *Desulfovibrio* spp., such as fumarate (Miller & Wakerley, 1966) and sulphonates (Lie *et al.,* 1996).

Desulfovibrio spp. incompletely oxidise a broad range of substrates (Grossman & Postgate, 1953; Postgate, 1984) including H_2 (Brandis & Thauer, 1981), sugars (Ollivier *et al.,* 1988), amino acids (Baena *et al.,* 1998), formate (Jansen *et al.,* 1984; Martins *et al.,* 2016), alkanes (Novelli & ZoBell, 1944; Rosenfeld, 1947) or short-chain alcohols (Bryant *et al.,* 1977; Nanninga & Gottschal, 1987). In the citric acid cycle, the oxidation of succinate to fumarate has a redox potential ($E^{0'}$ = + 32 mV) 100 mV more positive than the redox potential of the activated form sulphate reduction ($E^{0'}$ = - 68 mV) (Thauer *et al.,* 1989). Therefore, *Desulfovibrio* spp. cannot completely oxidise substrates to CO_2 and release acetate as an end-product of the oxidation (Postgate, 1984). *Desulfovibrio* spp. preferentially oxidise lactate and pyruvate (Macpherson & Miller, 1963). Oxidation of lactate or pyruvate results in one ATP molecule synthesis by substrate-level phosphorylation. Oxidation of two moles of lactate or pyruvate is thus required to balance out the reduction of one mole of sulphate (Peck, 1960), following these equations :

Electron donor	Equation
Pyruvate	$2\ CH_3COCOO^- + SO_4^{2-} \rightarrow 2\ CH_3COO^- + 2HCO_3^- + 0.5\ HS^- + 0.5\ H_2S$
Lactate	$2\ CH_3CHOHCOO^- + SO_4^{2-} \rightarrow 2\ CH_3COO^- + 2HCO_3^- + 0.5\ HS^- + 0.5\ H^+ + 0.5\ H_2S$

As the two ATP molecules produced by substrate-level phosphorylation during substrate oxidation are used for sulphate activation, *Desulfovibrio* spp. require another ATP source for growth. Energy to support growth is provided *via* electron transfer generated by cycling of metabolites. These metabolites, such as H_2, formate and carbon monoxide (CO), are by-products of the anaerobic respiration. The nature of the electron carriers varies depending on the electron acceptors, the carbon sources and the organism (Carepo *et al.,* 2002; Pereira *et al.,* 2011; Martins *et al.,* 2016).

The observation of a H_2 burst in an early stage of *Desulfovibrio* growth on lactate and sulphate, along with the evidence of two hydrogenases and a cytochrome C_3 in the *Desulfovibrio* genus, led to the proposal of an electron transport pathway called the

hydrogen cycling model (Tsuji & Yagi, 1980; Odom & Peck Jr, 1981; Figure 1.3A). According to the hydrogen cycling model, electrons and protons generated by substrate oxidation are reduced to H_2 by cytoplasmic hydrogenases. The resulting H_2 diffuses into the periplasm, where periplasmic hydrogenases oxidise H_2 into protons and electrons. Electrons are directed to the periplasmic cytochrome C_3 and then transferred through transmembrane protein complexes to the cytoplasm to reduce sulphate, generating a membrane potential. The membrane potential and the proton gradient activate an ATP synthase transmembrane complex, which drives ATP synthesis. Experimental evidences support the hydrogen cycling model. The presence of H_2 was detected and measured in *Desulfovibrio* cells growing on pyruvate (Peck *et al.,* 1987). The reduction in expression level by antisense messenger ribonucleic acid (mRNA) of a specific hydrogenase in *D. vulgaris* Hildenborough induces decreased growth (Van den Berg *et al.,* 1991).

However, the hydrogen cycling model remains controversial. Hydrogen synthesis from lactate oxidation was found to be energetically unfavourable (Pankhania *et al.,* 1988). Moreover, although putative cytoplasmic hydrogenases were identified in the genome of *D. vulgaris* Hildenborough (Heidelberg *et al.,* 2004), supporting the hydrogen cycling model, cytoplasmic hydrogenases are not conserved within the genus *Desulfovibrio* (Hauser *et al.,* 2011). The finding that, exogenous H_2 does not competitively inhibit substrate oxidation in *Desulfovibrio* spp., led to a new electron pathway proposal in which H_2 is not an obligatory intermediate (Lupton *et al.,* 1984). This alternative proposal, called the trace H_2 transformation model, postulates a direct transport of electrons from the donor to the acceptor, without the participation of H_2 (Figure 1.3B). Trace H_2 is produced and consumed only for controlling the redox state of the electron carriers during *Desulfovibrio* growth on organic compounds plus sulphate (Lupton *et al.,* 1984). Both experimentally proved, the H_2 cycling and the trace H_2 transformation models were assembled into a unified model (Noguera *et al.,* 1998). According to this unified model, the simultaneous functioning of the two electron transfer pathways ensures *Desulfovibrio* sulphatic growth on lactate. The model estimated 48 % of the electrons form H_2 before being used by sulphate while 52 % are directly transferred to sulphate (Noguera *et al.,* 1998).

In addition to the cycling of H_2, formate cycling (Figure 1.3C) and CO cycling (Figure 1.3D) are potential mechanisms for energy conservation. In the cytoplasm, formate can be produced by a pyruvate-formate lyase, which converts pyruvate and coenzyme A to acetyl-CoA and formate (Voordouw, 2002). Formate diffuse through the

membrane and periplasmic formate dehydrogenases generate H_2 from formate (Pereira *et al.,* 2011). H_2 is oxidised into protons and electrons by periplasmic hydrogenases. Electrons are directed to the periplasmic cytochrome C_3 and then transferred through transmembrane protein complexes to the cytoplasm to reduce sulphate, generating a membrane potential. The membrane potential and the proton gradient activate an ATP synthase transmembrane complex, which drives ATP synthesis. Evidence that formate cycling contributes to *D. vulgaris* Hildenborough energy production was found by observation of the reduced growth of a *D. vulgaris* mutant lacking formate dehydrogenase (da Silva *et al.,* 2013). Production of CO during substrate degradation by *D. vulgaris*, in particular when pyruvate is the substrate, was also detected. A CO cycling model was proposed by Voordow (2002). A cytoplasmic CO dehydrogenase (CODH) and a CO-dependent hydrogenase convert CO in CO_2 and H_2 and then transfer CO_2 and H_2 into the periplasm. H_2 is oxidised into protons and electrons by periplasmic hydrogenases. Electrons are directed to the periplasmic cytochrome C_3 and then transferred through transmembrane protein complexes to the cytoplasm to reduce sulphate, generating a membrane potential. The membrane potential and the proton gradient activate an ATP synthase transmembrane complex, which drives ATP synthesis (Voordouw, 2002).

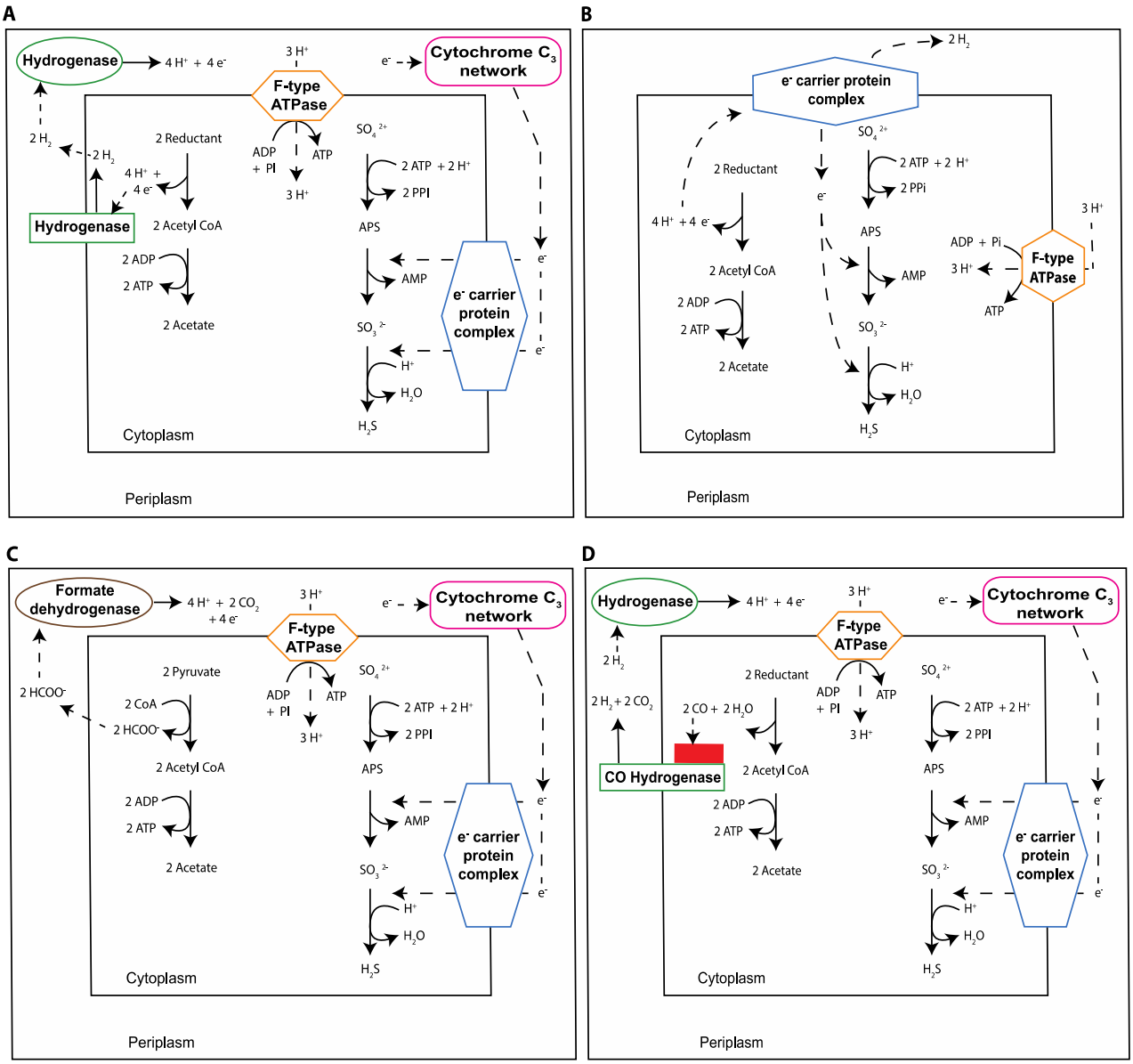


Figure 1.3. Chemiosmotic energy conservation mechanisms in *D. vulgaris* strain Hildenborough

Different pathways have been proposed for electron transfer sustaining ATP production in *D. vulgaris*. (A), H_2 cycling model suggests that H_2 is transferred from cytoplasm to periplasm where it is oxidised into protons and electrons by action of cytoplasmic and periplasmic hydrogenases. (B), The trace H_2 transformation model was proposed in which electrons are directly transferred from the donor to the acceptor, without participation of reduced metabolites. (C), Formate cycling model proposes that formate diffuses from the cytoplasm to the periplasm where it is oxidised by a formate dehydrogenase into CO_2, protons and electrons. (D), CO cycling involves a cytoplasmic CO dehydrogenase (red square) and a CO-dependent hydrogenase which convert CO in CO_2 and H_2 into the periplasm. A periplasmic hydrogenase then oxidise H_2 into protons and electrons. *D. vulgaris* energy conservation shows a high complexity and flexibility.

Syntrophic metabolism

Syntrophic metabolism is an "obligatory mutualistic metabolism" which is established by microbial partners, designated as syntrophs, that combine their metabolism in order to survive with minimal energy resources (Morris *et al.,* 2013). In the absence of an inorganic electron acceptor as sulphate, *Desulfovibrio* spp. hydrolyse and ferment organic compounds to acetate, formate, CO_2 and H_2. Fermentation of organic compounds is thermodynamically unfavourable, unless the partial pressure of H_2, waste product of fermentation, is maintained low, inferior to $10e^{-1}$ Pa. *Desulfovibrio* spp. therefore require a syntrophic association with an electron-accepting partner which maintains low hydrogen concentration (McInerney *et al.*, 2008). The first syntrophy associating *Desulfovibrio* spp. was discovered with an hydrogenotrophic methanogen archaeon *Methanobacterium* strain MOH in low sulphate medium (Bryant *et al.,* 1977). In absence of sulphate, *D*. *vulgaris* strain Hildenborough completely ferment lactate (Equation 1) or ethanol producing acetate, H_2, and CO_2. Subsequently, *Methanobacterium* strain MOH use H_2 to reduce CO_2 to methane (Equation 2). Overall for one mole of lactate consumed, 0.5 mole of methane is produced (Walker *et al.,* 2009; Equation 3).

Lactate fermentation by *Desulfovibrio*:

$$2\ CH_3CHOHCOO^- + 2\ H_2O \rightarrow 2\ CH_3COO^- + CO_2 + 4\ H_2 \quad (1)$$

Methanogenesis by methanogen archae:

$$4\ H_2 + CO_2 \rightarrow CH_4 + 2\ H_2O \quad (2)$$

Syntrophy with lactate:

$$2\ CH_3CHOHCOO^- \rightarrow 2\ CH_3COO^- + CH_4 \quad (3)$$

In environments rich in sulphate, anaerobic respiration is preferentially adopted by *Desulfovibrio* spp., as anaerobic respiration produced higher energy yields than methanogenesis. Thus, in environments rich in sulphate, *Desulfovibrio* spp. tend to out-compete methanogens for acetate, H_2 and CO_2. However, *Desulfovibrio* syntrophic growth has also been observed in sulphate-reducing environments. In environments rich in sulphate, *Desulfovibrio* spp. are the electron-accepting partner in syntrophic association and reduce sulphate. In co-culture with a methanogenic archaeon, such as *Methanosarcina barkeri* or *Methanosaeta concilii*, *D. vulgaris* consume H_2 produced by

the methanogen fermenting acetate (Phelps *et al.,* 1985; Ozuolmez *et al.,* 2015). *D. alaskensis* were observed to use H_2 and formate resulting from oxidation of fatty acids by *Syntrophomonas wolfei*, an anaerobic bacterium (Krumholz *et al.,* 2015). *Desulfovibrio* spp. were also found to grow in syntrophy with phototrophic green sulphur bacteria, involving exchange of sulphur compounds (Biebl & Pfennig, 1978).

Syntrophic metabolism proceeds near to thermodynamic equilibrium whereby the net free energy available is small (ΔG^0 < - 20 kJ mol^{-1} of substrate) and shared by all the syntrophs (Jackson & McInerney, 2002; Adams *et al.,* 2006). Despite a small free energy available, syntrophic interactions involving *Desulfovibrio* spp. are widespread in the microbial world, suggesting that *Desulfovibrio* syntrophic lifestyle is predominant over the sulphidogenic lifestyle (Sieber *et al.,* 2014).

1.4.2. Habitats

The high complexity and plasticity of *Desulfovibrio* energy metabolism may in part explain the diversity of their ecological niches. *Desulfovibrio* spp. are ubiquitous, inhabiting environments characterised by wide ranges of salinity, temperature and pH. *Desulfovibrio* spp. have been detected in fresh waters, for example the Lake Constance in central Europe (Bak & Pfennig, 1991), and in saline and hypersaline waters or sediments, including sediments from the Great Salt Lake in Utah (U.S.A) (Kjeldsen *et al.,* 2007). Although *Desulfovibrio* spp. are predominately mesophiles (Postgate, 1984), *D. frigidus* and *D. ferrireducens*, isolated from fjord sediments of Svalbard, are able to grow at - 2 °C (Vandieken *et al.,* 2006). Thermophilic *Desulfovibrio* spp. have been isolated such as *D. thermophilus*, capable of growth at a temperature of 85 °C (Rozanova & Khudiakova, 1974). However, thermophilic *Desulfovibrio* spp. have been reclassified to a new genus; *Thermodesulfovibrio* (Henry *et al.,* 1994). *Desulfovibrio* spp. have been detected in alkaline (Sasi Jyothsna *et al.,* 2008), neutrophilic (Haouari *et al.,* 2006) and acidic (Mago *et al.,* 2004) environments. Mainly found in anoxic environments, *Desulfovibrio* spp. were considered to be strict anaerobes. However, some *Desulfovibrio* spp., such as *D. aerotolerans*, have been isolated from micro-oxic environments, especially at the oxic-anoxic interfaces of microbial sediments (Sass *et al.,* 1997; Mogensen *et al.,* 2005). Furthermore, *Desulfovibrio* spp. are also natural constituents of the healthy intestinal flora in animals (Howard & Hungate, 1976; Gebhart *et al.,* 1993) and in humans (Warren *et al.,* 2005).

1.4.3. Significance

Due to their physiology, *Desulfovibrio* spp. play an important role in the biological carbon and sulphur cycles. *Desulfovibrio* spp. reduce sulphate and release sulphide ions, which can be then incorporated into proteins. In marine sediments, *Desulfovibrio* spp. participate to the complete mineralisation of carbon compounds with other microorganisms (Jørgensen, 1982). Beyond these ecological roles, *Desulfovibrio* spp. have various impacts on human health and on the environment, which are a consequence of their highly flexible energy metabolism and their ability to produce hydrogen sulphide.

Human health impacts

Human diseases due to *Desulfovibrio* infections are rare. To date, inflammatory bowel diseases (Loubinoux *et al.,* 2002; Rooks *et al.,* 2014; Earley *et al.,* 2015), liver abscesses (Tee *et al.,* 1996; Koyano *et al.,* 2015), brain abscesses (Lozniewski *et al.,* 1999), periodontitis (Loubinoux *et al.,* 2002; Dzierzewicz *et al.,* 2010) and blood stream infections (Goldstein *et al.,* 2003; Liderot *et al.,* 2010) have been diagnosed due to *Desulfovibrio* infections and four species have been characterised as potential opportunistic human pathogens. An appropriate antibacterial treatment against *Desulfovibrio* infections has not been yet established (Earley *et al.,* 2015). *Desulfovibrio* spp. are resistant to a wide range of antibiotics, partly as the hydrogen sulphide produced by *Desulfovibrio* spp. is a defence mechanism against antimicrobials (Shatalin *et al.*, 2011).

Desulfovibrio might also play a role in autism. The sulphur metabolism of autistic subjects was investigated. Autistic subjects had lower levels of plasma sulphate and excreted higher levels of sulphur compounds in their urine than control subjects (Waring & Klovrza, 2000). A more recent study proved that the intestinal flora of autistic patients statistically contains more *Desulfovibrio* bacteria than the intestinal flora of control subjects (Finegold, 2011).

Additionally, hydrogen sulphide, released by *Desulfovibrio* spp., is toxic and odorous. Exposure to hydrogen sulphide causes diverse health disorders in particular respiratory diseases (Duong *et al.,* 2001; Christia-Lotter *et al.,* 2007; Lim *et al.,* 2016).

Environmental impacts

The ability of *Desulfovibrio* spp. to degrade hydrocarbons has been exploited industrially for bioremediation of BTEX (benzene, toluene, ethylbenzene and xylene), which are toxic compounds found in contaminated soils from the oil industry (Farhadian *et al.,* 2008). Moreover, due to their ability to produce hydrogen sulphide, *Desulfovibrio* spp. have been industrially used for bioremediation of toxic metals present in industrial wastes and sewage waters (Pol *et al.,* 1998; Kiran *et al.,* 2016). Toxic metals (Me), such as cobalt, copper, nickel, cadmium and zinc, precipitate as metal sulphides in presence of hydrogen sulphide, according to the following equation:

$$HS^- + Me^{2+} \rightarrow MeS + H^+$$

Metal sulphides have low solubility, allowing their recovery and reuse from contaminated solutions (Fu & Wang, 2011). Bioremediation by *Desulfovibrio* spp. has the advantage to concomitantly remove sulphate and sulphide, which are also toxic metals. Currently, two bioremediation processes mediated by *Desulfovibrio* spp. have been patented: Thiopaq O&G Technology by Paquell BV (The Netherlands), a joint venture of Shell and Paques, and BioSulphide® by BioteQ, (Canada) (Hussain *et al.,* 2016). The Thiopaq® O&G process is a biodesulphurisation process of sour biogas streams with sulphur recovery. Biogas is a mixture of different gases, mainly methane and CO_2, which is produced from anaerobic degradation of organic compounds by bacteria. Biogas is considered sour if it contains more than 4 ppm H_2S by volume (Fidler *et al.*, 2003). BioSulphide® process purifies waste waters of toxic metals using hydrogen sulphide produced by *Desulfovibrio* spp.

Additionally, *Desulfovibrio* spp. were used in microbial electrolysis cells (MEC) for the production of renewable energy such as electricity and H_2 (Cooney *et al.,* 1996; Tatsumi *et al.*, 1999; Lojou *et al.,* 2002). In MEC, microorganisms oxidise a substrate to CO_2, protons and electrons in the anode chamber. The anode is the reducing electrode which supplies electrons. Electrons are transferred from the anode to the cathode, the oxidising electrode which accepts electrons, either by direct or indirect extracellular electron transfer. The electron flow between anode and cathode generates an electric current. Protons produced from substrate oxidation are transported through an electrolyte to the cathode chamber. In the cathode chamber, the protons can produce either H_2 by reacting with electrons, or other biofuels, such as methane, ethanol and biodiesel, by reacting with another micro-organism (Badwal *et al.,* 2014). *Desulfovibrio* spp. were used as bio-anode in H_2/O_2 MEC. In a H_2/O_2 MEC, H_2 is oxidised by a hydrogenase into protons

and electrons at the anode. At the cathode, protons are converted by an oxygenase into water (de Poulpiquet *et al.*, 2014). The first H_2/O_2 biofuel cell associated *D. vulgaris* cells at the anode and an O_2-oxidising enzyme at the cathode. The microbial biofuel cell achieved an open circuit potential of 1.17 V for at least 2 h (Tsujimura *et al.,* 2001). Moreover, the use of *Desulfovibrio* spp. as bio-anode demonstrated the possibility to associate removal of heavy metals with energy generation. In a MEC anode enriched with *Desulfovibrio* spp. and containing low metal concentration (up to 20 mg l^{-1}) of Cu^{2+}, more than 98 % of Cu^{2+} was removed and a stable current of 0.48 V was generated (Miran *et al.*, 2017). *Desulfovibrio* spp. were also used as bio-cathode for production H_2. *Desulfovibrio* strain G11 inoculated at the cathode of a MEC led to a current development from 0.17 to 0.76 A m^{-2} in 9 days and hydrogen production was observed (Croese *et al.,* 2011). One of the main challenges for any MEC using biocatalysts is to maintain viable and active the biological constituents of the system. To overcome growth inhibition by biogenic hydrogen sulphide, purified hydrogenases from *Desulfovibrio* spp. were successfully used as biocatalysts instead of using *Desulfovibrio* cells. However, these enzymes out of their physiological cell environment lose stability and so, lose activity (Lojou, 2011). Immobilisation of *Desulfovibrio* cells as biocatalysts was successful and more stable than using purified enzymes (Lojou *et al.,* 2002). Bioelectrochemical technologies for energy recovery represent an expending research area, as their applications show a great potential for renewable energy production (Head & Gray, 2016).

Desulfovibrio spp. are also associated with negative economic impacts, particularly affecting the oil industry. *Desulfovibrio* biosulphidogenesis causes the “souring” of crude oil (Youssef *et al.,* 2009). Moreover, hydrogen sulphide is corrosive, leading to the corrosion of metal infrastructures such as drilling and pipeline equipments (Hamilton, 1998; Enning & Garrelfs, 2014). Another negative impact for the oil industry is the ability of *Desulfovibrio* spp. to anaerobically oxidise hydrocarbons (Davis & Yarbrough, 1966; Widdel *et al.,* 2010; Cui *et al*., 2015).

Desulfovibrio spp. also affect anaerobic digestion treatment of agro-industrial wastes. Anaerobic digestion is an efficient waste treatment process inducing biogas production. Anaerobic digestion involves various microorganisms, with the last step performed by methanogens to produce methane (Gunaseelan, 1997). Sulphate being a common constituent of industrial wastes, *Desulfovibrio* spp. may out-compete methanogens for substrates during the anaerobic digestion. In consequence, methane production is reduced or inhibited. Moreover, the sulphide produced by *Desulfovibrio* spp. and the other toxic metals, such as cobalt, copper, nickel, cadmium and zinc, are noxious

to methanogens (Chen *et al.,* 2008). Biogenic sulphide and toxic metals in wastes can therefore affect methanogenesis and energy recovery. However, biogenic sulphide produced by *Desulfovibrio* spp. can be used as a metal detoxification agent to optimise energy recovery (Paulo *et al.,* 2015). A right balance must be therefore established between sulphate reduction and methanogenesis in anaerobic digestion to optimise the toxic metal remediation by biogenic sulphide.

1.4.4. Omics Approaches for Holistic Analyses of *Desulfovibrio*

Desulfovibrio negative impacts on human health and on the oil industry, and their valuable applications in biotechnology for renewable energy production and bioremediation sparked a strong interest for genetic and metabolic characterisation of *Desulfovibrio* spp.

Genomics

D. vulgaris strain Hildenborough was the first *Desulfovibrio* strain to have its genome sequenced and assembled (Heidelberg *et al.,* 2004). Subsequently, the genome of 63 *Desulfovibrio* strains, representing 33 spp., have been sequenced, according to the GenBank database in 2017. Among these sequenced genomes, 14 genomes of 10 *Desulfovibrio* spp., are completely assembled. *Desulfovibrio* genomes range in size between 2.6 megabase pairs (Mb) and 5.7 Mb, and have a GC content varying from 41.8 % to 68 %, according to the GenBank database in 2017.

Transcriptomics and Proteomics

Global transcriptomic and proteomic analyses of *Desulfovibrio* spp. enabled understanding of their energy metabolism and identification of the underlying molecular basis. *D. vulgaris* strain Hildenborough growth on different electron donor and acceptor compounds or under stress conditions showed changes in transcriptional profile, highlighting the complexity and flexibility of *Desulfovibrio* energy metabolism (Chhabra *et al.,* 2006; He *et al*., 2006; Mukhopadhyay *et al.,* 2006; Pereira *et al.,* 2008). Moreover, transcriptomic analyses of *D. alaskensis* strain G20 shed light that energy metabolism in *Desulfovibrio* is species-specific (Keller *et al.,* 2014). The enzymes involved in energy metabolism differ for each *Desulfovibrio* spp.

Desulfovibrio syntrophic metabolism was also studied using transcriptomic and proteomic approaches. Several transcriptomic analyses of *D. vulgaris* in syntrophic association with a methanogen archaeon came to the same conclusions. A set of three functionally unknown genes and a five-gene cluster encoding lipoproteins and membrane-bound proteins are involved in syntrophic metabolism (Walker *et al.*, 2009; Plugge *et al.*, 2010). A phylogenetic study showed that the presence of the three functionally unknown gene set in *D. vulgaris* was the result of a horizontal gene transfer from a methanogen to *D. vulgaris* (Scholten *et al.,* 2007). The five gene cluster has orthologs in four *Desulfovibrio* strains (Plugge *et al.*, 2010). However, transcriptional profiles of *D. vulgaris* strain Hildenborough and *D. alaskensis* strain G20 both growing in syntrophy with an archaeon methanogen were highly different, suggesting absence of conserved genes sustaining syntrophic metabolism in *Desulfovibrio* spp. (Meyer *et al.*, 2013).

Molecular Biology

The development of molecular biology methods for *Desulfovibrio* engineering has been hampered by challenges encountered in microbiological manipulation (Bender *et al.*, 2006). Additionally, the natural resistance of *Desulfovibrio* spp. to commonly used antibiotics impeded the expansion of molecular biology tools (Postgate, 1984). Annotation of the *D. vulgaris* Hildenborough genome predicted several multi-drug exporters (Payne *et al.,* 2004). However, the genetic manipulation of *Desulfovibrio* spp. has been improved, notably for the model organism *D. vulgaris* Hildenborough. *Desulfovibrio* mutagenesis has been successfully performed by generalised transduction (Rapp & Wall, 1987), transposition (Wall *et al.*, 1996; Groh *et al.*, 2005) conjugation (Powell *et al.,* 1989; Argyle *et al.,* 1992) and electroporation (Rousset *et al.*, 1991; Aubert *et al.*, 1998; Casalot *et al.*, 2002; Broco *et al.,* 2005; Parks *et al.*, 2013).

Targeted gene mutagenesis methods were developed in *Desulfovibrio* spp. for elucidation of gene function. Targeted gene mutagenesis enables disruption of gene activity or deletion of genes of interest. In *D. vulgaris* Hildenborough, the expression of a periplasmic hydrogenase gene was reduced using an antisense mRNA (Van den Berg *et al.,* 1991). In *D. alaskensis* G20, the gene encoding the tetraheme cytochrome C_3 was interrupted by insertion of a suicide plasmid *via* a single homologous recombination event. However, genome arrangements restoring a wild-type phenotype for the gene of interest and retaining the selection marker were observed (Rapp-Giles *et al.,* 2000).

Targeted gene mutagenesis by gene deletion in *Desulfovibrio* spp. was first reported by Rousset *et al* (1991). Substitution of a hydrogenase coding gene from *D. fructosovorans* by a kanamycin resistance cassette from a suicide plasmid occurred by double homologous recombination events (Rousset *et al.,* 1991). Gene deletion by one-step allelic exchange mutagenesis was thereafter successfully performed in *D. gigas* ATCC 19364 (Broco *et al.,* 2005; da Silva *et al.,* 2015), *D. vulgaris* strain Hildenborough (Bender *et al.,* 2007; Keller *et al.,* 2009; Chhabra *et al.,* 2011) and *D. desulfuricans* ND132 (Parks *et al.,* 2013).

Although sequential deletion of three hydrogenase genes was performed in *D. fructosovorans* resulting in a strain carrying three antibiotic resistance markers (Casalot *et al.,* 2002), the narrow range of selection marker suitable to *Desulfovibrio* spp. promoted the development of two-step markerless allelic exchange methods. Fu and Voordouw achieved the deletion of an oxygen sensor coding gene in *D. vulgaris* Hildenborough using a two-step allelic exchange protocol, involving a chloramphenicol resistance marker and the *sacB* gene from *Bacillus subtilis*, as counter-selectable marker (Fu & Voordouw, 1997). However, *D. vulgaris* resistance to sucrose was found to depend on medium composition, initial culture density, and the time of exposure (Fu & Voordouw, 1998). Moreover, the sucrose resistance of 50 % of the transformed *D. vulgaris* cells was due to inactivation of the *sacB* gene through the transposon IS*D1* insertion in the gene sequence, instead of removal of the *sacB* gene through a second recombination event (Fu & Voordouw, 1998). Another two-step markerless allelic exchange method was developed using a spectinomycin resistance marker and the endogenous gene *upp* as counter-selectable marker. As a prerequisite, the *upp* gene needs to be deleted in the strain to be engineered (Keller *et al.,* 2009). This markerless allelic exchange method was successfully used for *D. vulgaris* Hildenborough engineering (Korte *et al.*, 2014; De León *et al.,* 2017).

Heterologous expressions of *Desulfovibrio* genes were also performed and enabled structural and functional characterisation of proteins. *Desulfovibrio* genes were heterologously expressed in different *Desulfovibrio* spp. hosts (Voordouw *et al.,* 1990; Tan *et al.,* 1994; Aubert *et al.,* 1997; Aubert *et al.,* 1998). The first attempts of *Desulfovibrio* gene expression in an organism of differrent genus were unsuccessful resulting in accumulation of apoproteins due to the lack of iron-sulphur clusters (Voordouw *et al.,* 1987; Van Dongen *et al.,* 1988). These observations suggested that the heterologous expression of *Desulfovibrio* genes require a related host providing a homologous background to *Desulfovibrio* spp. However, Cannac *et al.* proved that the

photosynthetic bacterium *Rhodobacter sphaeroides* transformed with the gene encoding the cytochrome c_3 from *D. vulgaris* Hildenborough produced the cytochrome c_3 holoform (Cannac *et al.*, 1991). The possibility to heterologously express *Desulfovibrio* genes in a host from a different genus was thereafter confirmed notably in *E. coli* (Pollock & Voordouw, 1994, da Costa *et al.*, 2000, Kitamura *et al.*, 2004) and in *Shewanella onedensis* MR-1 (Ozawa *et al.*, 2000)

1.5. Alternative Hypothetical Pathways for Alkane Biosynthesis in Desulfovibrio

A hypothetical pathway for even-numbered carbon chain alkane synthesis by *D. desulfuricans* is that *D. desulfuricans* has an unusual fatty acid metabolism producing odd-chain fatty acids which are subsequently decarboxylated to even-chain, terminal alkanes. This hypothesis was proposed by Bagaeva based on isotopic labelling experiments (Bagaeva, 1998 reviewed in Ladygina *et al*., 2006; Figure 1.4). Bagaeva's hypothetical pathway involves the reduction of acetate and formate to aldehyde. Acetate is produced by incomplete oxidation of lactate during anaerobic respiration and by conversion of formate. Formate is generated from CO_2 reduction catalysed by a formate dehydrogenase.

Acetaldehyde and formaldehyde undergo aldol condensations resulting in carbon chain elongation, and subsequent oxidations producing even and odd carbon chain free fatty acids. Bagaeva observed that ^{14}C derived from methyl groups of labelled acetate was preferentially incorporated into alkanes than ^{14}C from labelled carboxylic groups and concluded that hydrocarbons are produced by decarboxylation of free fatty acids. No further biochemical nor genetic studies were performed to confirm this hypothesis.

However, the type of *Desulfovibrio* synthesised alkanes consistently reported in the literature (C_{18}-C_{35}, n-alkanes; Oppenheimer, 1965; Davis, 1968; Bagaeva & Chernova, 1994) is typical of "white mineral oil", a commonly used lubricant present in most manufactured items. These previous analyses may be criticised as they failed to distinguish any metabolic hydrocarbon production from potential contamination with white oil, leading to unreliable data regarding if *Desulfovibrio* spp. produce hydrocarbons and which type of alkane is produced.

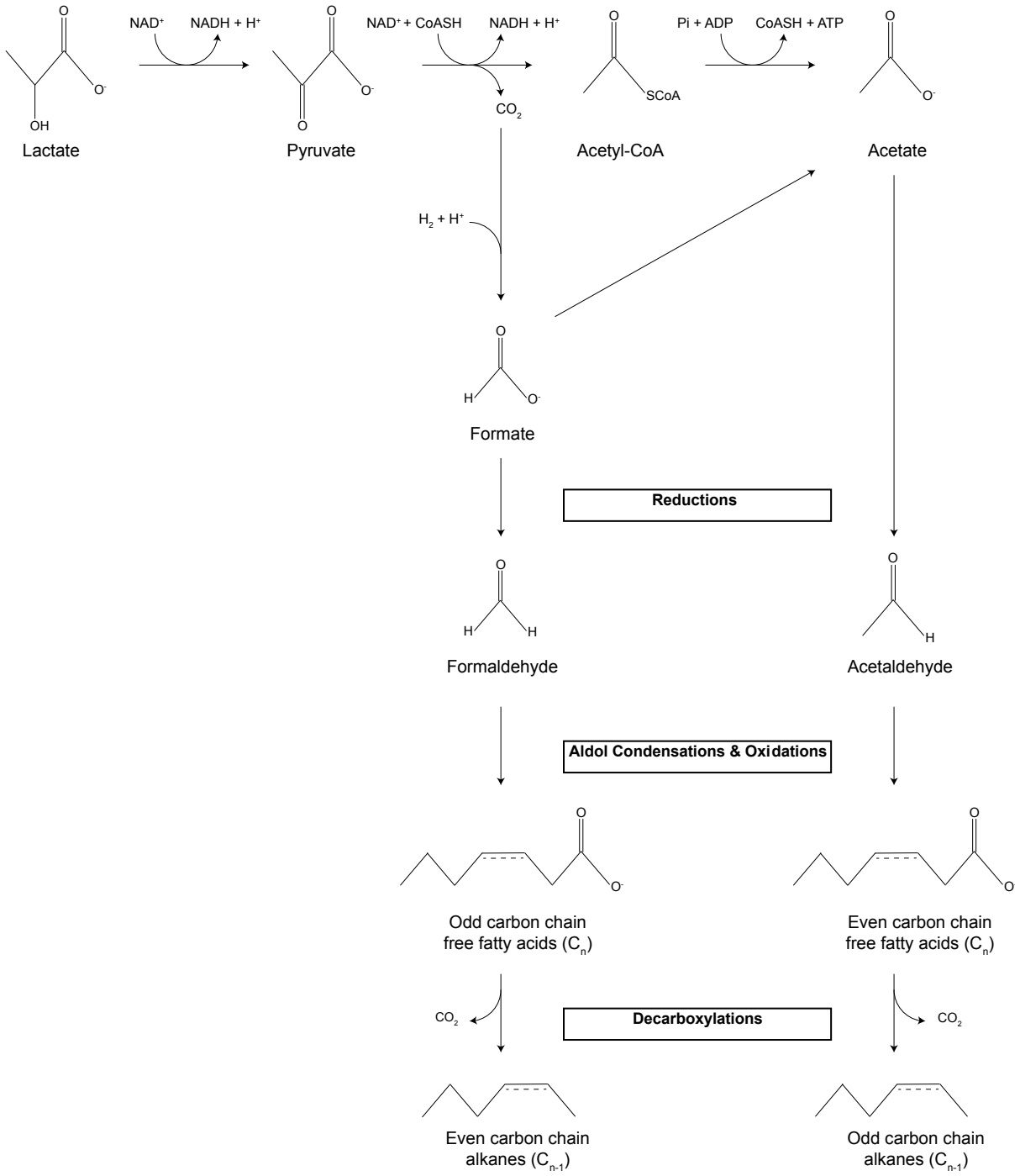


Figure 1.4. Hypothetical *Desulfovibrio* hydrocarbon biosynthetic pathway *via* decarboxylation of fatty acid intermediates

Based on isotopic labelling experiments, Bagaeva proposed a pathway of hydrocarbon production *via* odd or even carbon chain fatty acid decarboxylation. The precursor molecule of odd carbon chain fatty acids is formate, generated from CO_2 reduction catalysed by a formate dehydrogenase. The precursor molecule of even carbon chain fatty acids is acetate, produced by lactate incomplete oxidation and by formate conversion.

An alternative hypothesis to explain even-numbered carbon chain alkane synthesis by *D. desulfuricans* does not involve a decarboxylation step, but relies on a possible series of reduction reactions from fatty acids. A similar hypothesis was suggested after observation of even-numbered carbon chain alkane synthesis by *Vibrio furnissii* M1 (Park *et al.,* 2001). Park proposed the direct reduction of fatty alcohols to the corresponding alkanes by hydrogenation (Park, 2005; Figure 1.5). However, extensive genomic and biochemical studies proved the absence of alkane production in *V. furnissii* M1, revoking the existence of the reductive pathway (Wackett *et al.,* 2007) and potentially supressing research impetus into similar pathways in different systems such as *Desulfovibrio*.

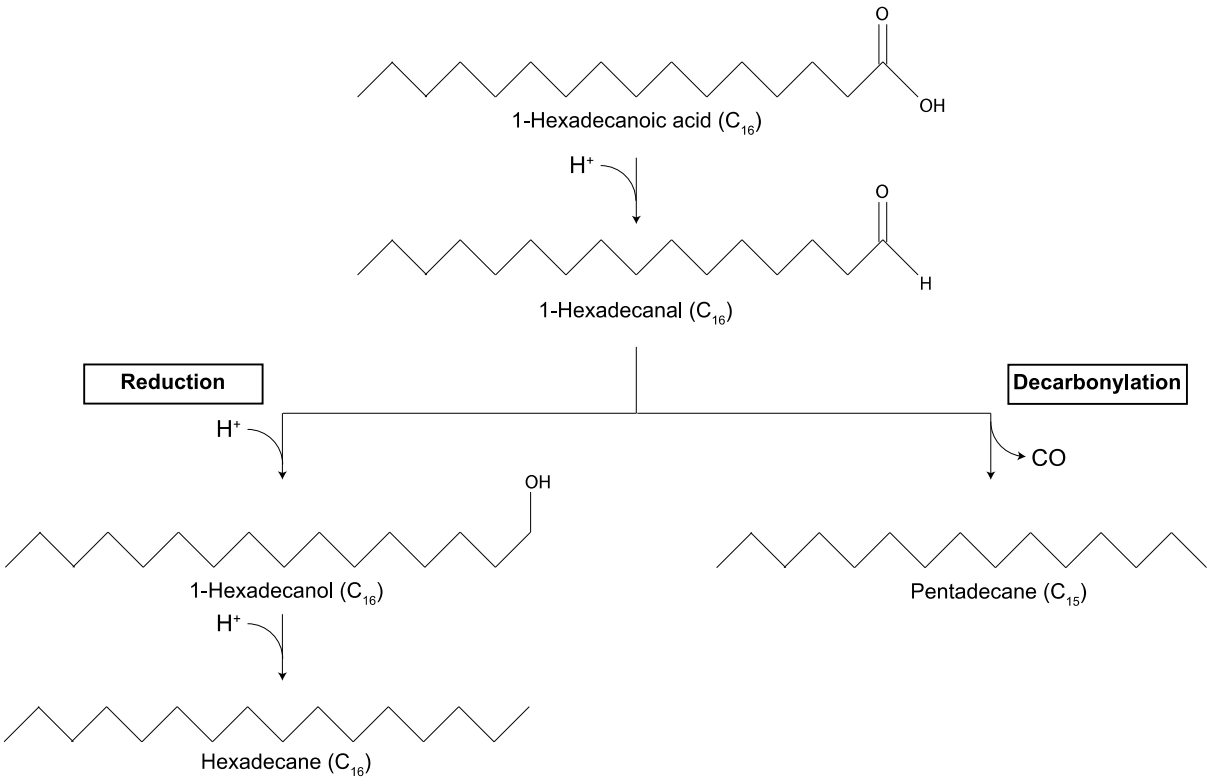


Figure 1.5. Hypothetical hydrocarbon biosynthetic pathway in *V. furnissii* M1, *via* reduction of fatty alcohol intermediates and decarbonylation of fatty aldehyde intermediates

The co-existence of a decarbonylation route and a reductive route from fatty acids was proposed by Park to explain odd and even-numbered carbon chain alkane synthesis by *Vibrio furnissii* M1 (Park, 2005). The fatty 1-hexadecanoic acid is reduced to 1-hexadecanal, which is then either decarbonylated to pentadecane or reduced to 1-hexadecanol. The 1-hexadecanol alcohol is subsequently reduced to hexadecane.

1.6. Hypothesis and Project Aims

This project focused on the characterisation of alkane biosynthetic pathway(s) in *Desulfovibrio*. *Desulfovibrio* spp. are not suited to industrial scale alkane production due to slow growth, requiring carefully monitored anaerobic conditions and a complex growth medium. However, genes which encode enzymes involved in *Desulfovibrio* alkane synthesis pathway(s) are of interest for production of replica petroleum by a suitably engineered host.

In this project, we hypothesised that the synthesis of even-numbered carbon chain alkanes by *D. desulfuricans* would occur either (Figure 1.6A) by decarbonylation of the odd-numbered carbon chain fatty acids or (Figure 1.6B) by a series of reduction reactions from even-numbered carbon chain fatty acids.

To address this hypothesis and to characterise *Desulfovibrio* alkane biosynthetic pathway(s), the aims of this investigation were:

1. Verification of hydrocarbon biosynthesis in the *Desulfovibrio* genus, using a method which enables the unambiguous distinction between biogenic hydrocarbons and non-metabolically derived hydrocarbons.

2. Structural characterisation of the biogenic alkanes and quantification of the productivity for total alkane synthesis by *Desulfovibrio* spp.

3. Determination of the fatty acid content of the *Desulfovibrio* spp. involved in this study.

4. *In silico* identification of the molecular basis involved in *Desulfovibrio* alkane biosynthesis through target-directed genome mining.

5. Empirical verification of the molecular components identified *in silico* to be putatively involved in *Desulfovibrio* alkane biosynthesis.

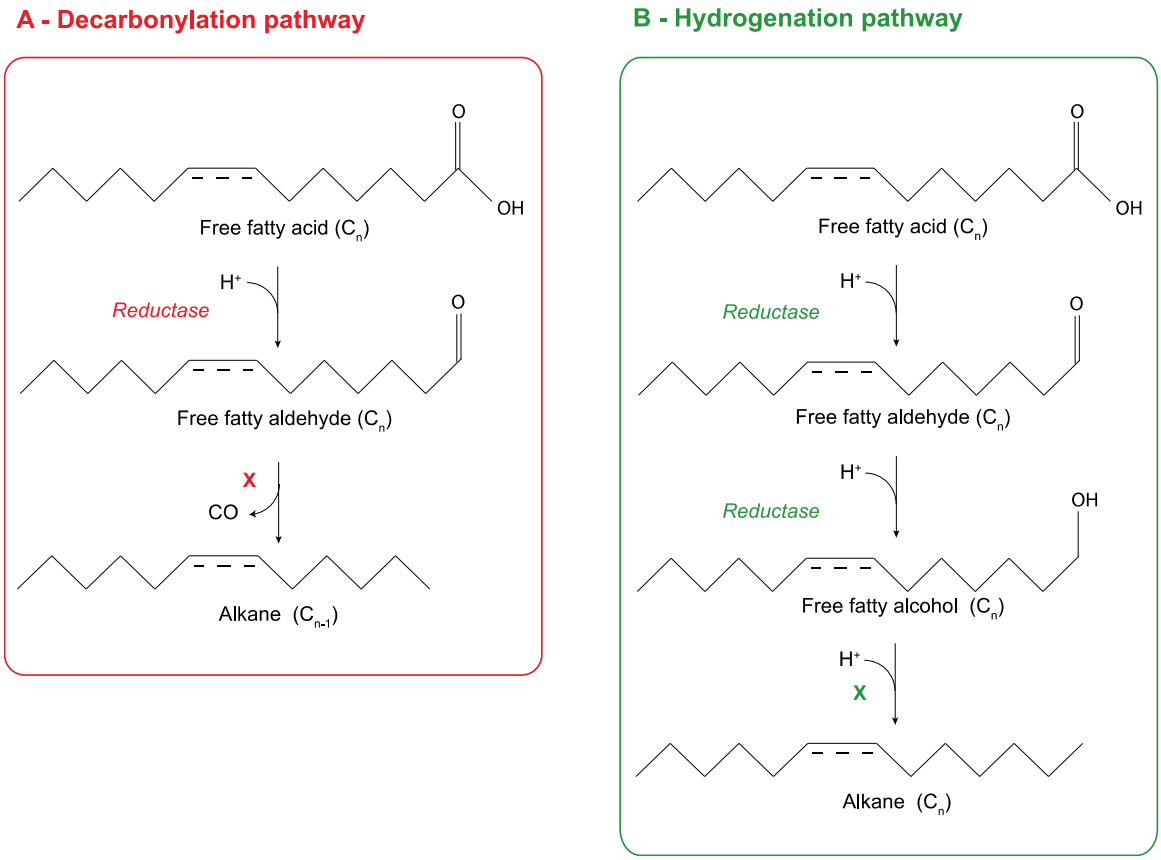


Figure 1.6. Hypothetical pathways for even-numbered carbon chain alkane biosynthesis in *Desulfovibrio*

Hypothetical pathways for even-numbered carbon chain alkane biosynthesis in *D. desulfuricans* would follow either (A; pathway enclosed in red) a decarbonylation route from odd-numbered carbon chain fatty acids or (B; pathway enclosed in green) a reductive hydrogenation pathway from even-numbered carbon chain fatty acids. X represents unknown enzymes.

2. Materials and Methods

2.1. Strains, Media, Culture and Preservation Conditions

2.1.1. Strains

Desulfovibrio strains used in this study are listed in table 2.1. *Desulfovibrio* strains were obtained as lyophilised cultures and rehydrated following culture collection's instructions.

Table 2.1. *Desulfovibrio* strains used in this study

Accession number	***Desulfovibrio* strain, environmental source and physiological characteristics**	Source
NCIMB 8326	*D. desulfuricans subsp. desulfuricans* California27.137.5; Top soil in Long Beach Oil Field, California (U.S.A.); Halophile, 30 °C	NCIMB Ltd, Aberdeen, Scotland
NCIMB 9332	*D. gigas;* Fresh water, Etang de Berre, Marseille, France; 30 °C	NCIMB Ltd, Aberdeen, Scotland
DSM 10636	*D. gabonensis* SEBR 2840; Water sample from oil pipeline, Gabon; Halophile; 30 °C	NCIMB Ltd, Aberdeen, Scotland
DSM 16681	*D. paquesii* SB1; Sulfidogenic sludge of a full-scale synthesis-gas-fed bioreactor treating wastewater from a zinc smelter, Budel-Dorplein, The Netherlands; 37 °C	DSMZ, Braunschweig, Germany
NCIMB 8338	*D. desulfuricans subsp. desulfuricans* CubaHC29.130.4; Estuary well; Halophile, 30 °C	NCIMB Ltd, Aberdeen, Scotland
DSM 18311	*D. marinus* E-2; Marine sediments, Sfax, Tunisia; Halophile; 37 °C	DSMZ, Braunschweig, Germany
NCIMB 8307	*D. desulfuricans subsp. desulfuricans* Essex6; Tar and sand mixture surrounding corroded gas main in waterlogged clay, South Essex, U.K.; 30 °C	NCIMB Ltd, Aberdeen, Scotland
NCIMB 12906	*D. alcoholivorans;* Mud from sewage plant, Gottingen, Germany; 37 °C	NCIMB Ltd, Aberdeen, Scotland
DSM 4370	*D. giganteus* STg; Gut of the soil-feeding termite *Cubitermes sp.*, Congo; Halophile; 37 °C	DSMZ, Braunschweig, Germany
	D. vulgaris subsp. vulgaris Hildenborough; Wealden Clay, Hildenborough, Kent, U.K.; 30 °C	Judy Wall Laboratory, University of Missouri, USA

NEB 5-alpha *E. coli* (New England Biolabs Ltd., Hitchin, UK) was used for plasmid storage and amplification. BL21 Star™ (DE3) *E. coli* (ThermoFisher Scientific Inc., Horsham, UK) was used for recombinant protein expression. WM3064 *E. coli* (provided by Judy Wall laboratory, University of Missouri) was used as conjugative donor strain for plasmid transfer between *E. coli* and *Desulfovibrio* spp.

Table 2.2. Genotypic characteristics of *E. coli* used in this study

***E. coli* strain**	**Genotype**	**Source**
NEB 5-alpha	*fhu*A2 Δ*(arg*F-*lacZ)* U169 *pho*A *gln*V44 80 *(lacZ)*M15 *gyr*A96 *rec*A1 *rel*A1 *end*A1 *thi*-1 *hsd*R17	New England Biolabs, Hitchin, UK
BL21 Star™ (DE3)	F- *omp*T *hsd*SB *(r_B-, m_B-) gal dcm rne*131 (DE3)	ThermoFisher Scientific Inc., Horsham, UK
WM3064	*thrB*1004 *pro thi rpsL hsd*S *lacZ*Δ*M15* RP4-1360 Δ(*araBAD)567* Δ*dapA*1341::[*erm pir*]	Judy Wall Laboratory, University of Missouri, USA

2.1.2. Media and Culture Conditions

Unless otherwise stated, all chemicals were purchased from ThermoFisher scientific Inc. and Sigma-Aldrich (Gillingham, UK). Solutions were prepared using 18 $M\Omega cm^3$ milli-Q water (Merck-Millipore, Feltham, UK).

Desulfovibrio were cultured in an anaerobic growth chamber (Whitley A35 Workstation, Don Whitley Scientific Ltd, Shipley, United Kingdom) at 37 °C under an anaerobic atmosphere of 80 % N_2, 10 % CO_2 and 10 % H_2, in sodium lactate medium (60 mM sodium lactate, 30 mM sodium sulphate, 8 mM $MgCl_2$, 30 mM Tris (pH 7.4), 20 mM NH_4Cl, 2 mM K_2HPO_4, 2 mM NaH_2PO_4, 600 *μ*M $CaCl_2$, 120 *μ*M Ethylenediaminetetraacetic acid (EDTA), 60 *μ*M $FeCl_2$, 0.1 % (w/v) yeast extract, 0.1 % (v/v) Thauers vitamin solution and 0.6 % (v/v) trace element solution). The Thauers vitamin solution; 490 *μ*M pyridoxine HCl, 410 *μ*M nicotinic acid, 360 *μ*M *p*-aminobenzoic acid, 240 *μ*M lipoic acid, 210 *μ*M DL pantothenic acid, 150 *μ*M thiamine HCl, 130 *μ*M riboflavin, 80 *μ*M biotin, 50 *μ*M folic acid, 14.3 mM choline chloride and 10 *μ*M vitamin B12. The trace element solution; 2.5 mM $MnCl_2$, 1.47 mM $ZnCl_2$, 1.26 mM $CoCl_2$, 380 *μ*M $NiSO_4$, 320 *μ*M H_3BO_3, 210 *μ*M Na_2MoO_4, 35 *μ*M Na_2SeO_3, 24 *μ*M Na_2WO_4 and 11.7 *μ*M $CuCl_2$. The reductant sodium thioglycolate was added to a final concentration of 1.2

mM. For halophile *Desulfovibrio* strains, 2.5 % (w/v) NaCl was added to sodium lactate medium. The pH was adjusted to 7.2 prior to autoclaving. For solidified medium, 1.5 % (w/v) agar was added to sodium lactate medium. For *Desulfovibrio* plating, cells were spread on empty petri dishes and covered with molten medium, in the anaerobic chamber. For transformant screening, geneticin G418 dissolved in water was added to sodium lactate medium to a final concentration of 400 μg ml^{-1}. For alkane biosynthesis screening, *Desulfovibrio* were cultured in sodium lactate medium containing 10 % (v/v) deuterium oxide (D_2O; 99.8 atom % D, NMR grade, ACROS Organics, Geel, Belgium).

NEB 5-alpha and BL21 Star™ (DE3) *E. coli* were cultured aerobically or anaerobically at 37 °C in Luria-Bertani (LB) medium (10 g l^{-1} tryptone, 10 g l^{-1} NaCl, and 5 g l^{-1} yeast extract). WM3064 *E. coli* was cultured aerobically at 37 °C in LC medium (10 g l^{-1} tryptone, 5 g l^{-1} NaCl, and 5 g l^{-1} yeast extract), supplemented with 0.3 mM 2,6-diaminopimelic acid (DAP). For solidified medium, 1.5 % (w/v) agar was added to LB or LC medium. For transformant screening, kanamycin dissolved in water was added to LB or LC medium to a final concentration of 50 μg ml^{-1}.

2.1.3. Strain Preservation Conditions

Equal volumes of microbial culture and a 50 % (v/v) glycerol solution were mixed before being snap-frozen in liquid nitrogen and stored at - 80 °C.

2.2. Scanning Electron Microscopy

D. desulfuricans NCIMB 8326 cells from a planktonic culture were harvested by centrifugation at 8,000 x *g* for 10 min. Pellets were fixed in 100 mM phosphate buffer (pH 7.2) containing 3 % (v/v) glutaraldehyde and 4 % (v/v) paraformaldehyde, for 1 h at room temperature. Pellets were washed three times for 5 min in 100 mM phosphate buffer (pH 7.2), prior to being post-fixed overnight in 1 % (v/v) osmium tetroxide diluted in deionised water. Pellets were then washed five times in 0.2 μm filtered deionised water and dehydrated by successively resuspending them in 30 % (v/v), 50 % (v/v), 75 % (v/v), 90 % (v/v) and twice in anhydrous ethanol for 10 min. Cells were filtered through a 0.2 μm black polycarbonate filter paper and dried for 4 h using critical point drying. Cells were

sputter-coated with 10 nm of gold and palladium (80:20) using a turbomolecular-pumped coating system Q150T ES from Quorum Technologies Ltd. (Ashford, U.K.), prior to being observed under a JSM-6390LV Scanning Electron Microscope from JEOL Ltd. (Tokyo, Japan).

2.3. Hydrocarbon and Fatty Acid Extraction and Analysis

2.3.1. *Desulfovibrio* Culture Conditions for Metabolism Screening

Liquid cultures of wild type *Desulfovibrio* were prepared by inoculating a single colony into 5 ml sodium lactate medium. Cultures were incubated anaerobically for 3-6 days at 37 °C. Post incubation, an entire 5 ml culture was used to inoculate 35 ml sodium lactate medium. 40 ml liquid cultures were anaerobically incubated for 10 days at 37 °C, prior to cellular organic compound extraction.

2.3.2. Dichloromethane Extraction of Cellular Organic Compounds

Bacterial cultures were transferred into Oak Ridge High-Speed Centrifuge polytetrafluoroethylene (PTFE) tubes (Nalgene™, Rochester, USA) previously cleaned with anhydrous dichloromethane (DCM; ≥ 99.9 % stabilised, PESTINORM® for capillary GC analysis, VWR International Ltd., Lutterworth, UK). Cultures were centrifuged at 16,000 x *g* for 30 min at 4 °C and supernatants were discarded. Pellets were washed three times in 25 ml sterile phosphate-buffered saline (PBS; 0.01 mM Na_2PO_4 $7H_2O$, 3mM KCl, 140 mM NaCl; pH 7.4). Pellets were then snap-frozen in liquid nitrogen before being uncapped and left overnight in a CoolSafe 4L ScanVac freeze-dryer (LaboGene ApS, Lynge, Denmark). Once the pellets were freeze-dried, 500 *µ*l DCM spiked with 10 *µ*M β-cyclocitral (an internal standard) was added to the pellets using a glass syringe. The samples were then placed in a 45 kHz ultrasonic water bath (Ultrasonic cleaner, VWR) for 45 min. For negative controls, extraction blanks were performed on uninoculated growth medium.

2.3.3. Preparation of Cellular Organic Compound Extracts for Gas Chromatography – Mass Spectrometry (GC-MS) Analysis

After sonication, samples were filtered using a glass syringe. The syringe needle was replaced with 4 mm Millex® syringe filters [polyvinylidene difluoride (PVDF) membrane, 0.45 *µ*m pore size; Whatman, Maidstone, UK]. Between each sample, the glass syringe was cleaned with anhydrous DCM (six times), anhydrous ethyl acetate (six times) and anhydrous DCM (six times). Filtered samples were transferred into glass vials (300 *µ*l screw thread vial with fused insert; ThermoFisher Scientific Inc.), prior to being screened for hydrocarbons by GC-MS. A sample of 500 *µ*l DCM spiked with 10 *µ*M β-cyclocitral was also sonicated and filtered prior to being analysed by GC-MS.

For fatty acid analysis, 100 *µ*l filtered organic compound extracts and 100 *µ*l DCM spiked with 10 *µ*M β-cyclocitral (negative control) were derivatised. Filtered organic compound extract samples and the negative control sample were uncapped and dried under vacuum using a Standard EZ-2 Series evaporator with the program “Low BP” (Genevac, Ipswich, UK). 20 *µ*l of 240 mM methoxyamine hydrochloride dissolved in anhydrous pyridine was added to the dried samples using a glass syringe. Samples were incubated at 37 °C for 2 h with the lid on. 40 *µ*l of anhydrous N-Methyl-N-(trimethylsilyl) trifluoroacetamide was then added to the samples using a glass syringe. Samples were incubated at 37 °C for 30 min with the lid on, prior to being screened for fatty acids by GC-MS. A sample containing only the derivatisation agents [methoxyamine hydrochloride and N-Methyl-N-(trimethylsilyl) trifluoroacetamide] was also prepared and analysed by GC-MS.

Glass syringes were rinsed with anhydrous DCM (three times), a 10 % (v/v) methanol and 1 % (v/v) glacial acetic acid solution (three times) and a 50 % (v/v) methanol solution (three times).

2.3.4. Preparation of Organic Compound Standards

Standard solutions of 40 mg l^{-1} C_8 - C_{20} and C_{21} - C_{40} alkanes (Sigma-Aldrich) were used as retention time standards for alkanes. A 5-point calibration curve of C_8 - C_{20} and C_{21} - C_{40} alkane solutions diluted to final concentrations of $5x10^{-3}$ mg ml^{-1}, $2.5 x10^{-3}$

mg ml^{-1}, 1.25x10^{-3} mg ml^{-1}, 0.625 x10^{-3} mg ml^{-1} and 0.3125x10^{-3} mg ml^{-1} in DCM spiked with 10 μM β-cyclocitral was performed for alkane quantification.

A solution of tetradecanoic acid ($C_{14:0}$), pentadecanoic acid ($C_{15:0}$), hexadecanoic acid ($C_{16:0}$), heptadecanoic acid ($C_{17:0}$), octadecanoic acid ($C_{18:0}$), nonadecanoic acid ($C_{19:0}$), eicosanoic acid ($C_{20:0}$) and heneicosanoic acid ($C_{21:0}$) dissolved to a final concentration of 10 μM in DCM spiked with 10 μM β-cyclocitral was prepared, derivatised and used as retention time standards for fatty acids.

2.3.5. Gas Chromatography – Quadrupole-Time Of Flight Mass Spectrometer (GC/Q-TOF/MS) Parameters

Analyses were performed using a 7200 series GC(7890A)/Q-TOF/MS system from Agilent Technologies (Santa Clara, USA). 5 μl of samples and standards were injected into a non-deactivated, baffled glass liner with a 12:1 split ratio (14.448 ml min^{-1} split flow). The inlet temperature was maintained at 250 °C with a 3.0 ml min^{-1} septum purge flow. A Zebron semi-volatiles (Phenomenex, Torrance, USA) column (30 m x 250 μm x 0.25 μm) was maintained at a constant helium flow rate of 1.5 ml min^{-1}. The temperature gradient of the GC was initially held for 4 min at 70 °C, then was ramped at a rate of 15 °C min^{-1} until 310 °C was achieved and held for 6 min. Data were analysed using Agilent technologies MassHunter Qualitative Analysis software (version B.07.00), MassHunter Q-TOF Quantitative Analysis software (B.08.00) and the National Institute of Standards and Technology (NIST) Mass Spectral Library (version 11) for identification of unknown compounds. Data visualisation and statistical analyses were performed using the software Prism (version 7; GraphPad Software Inc., San Diego, USA). Data were statistically analysed using a two-tailed paired t-test at the 95% confidence level. It was assumed that the data followed a Gaussian distribution.

2.4. Whole Genome Purification and Sequencing

2.4.1. *Desulfovibrio* Culture Conditions for Genomic Deoxyribonucleic Acid (DNA) Purification

Liquid cultures of wild type *D. desulfuricans* NCIMB 8326, *D. desulfuricans* NCIMB 8338, *D. desulfuricans* NCIMB 8307, *D. gigas* NCIMB 9332, *D. alcoholivorans* NCIMB 12906, *D. gabonensis* NCIMB 10636, *D. paquesii* DSM 16681, *D. marinus* DSM 18311 and *D. giganteus* DSM 4370 were prepared by inoculating a single colony into 10 ml of sodium lactate medium. Cultures were incubated anaerobically for 7 days at 37 °C, prior to genomic DNA purification. As a negative control, uninoculated sodium lactate medium was incubated anaerobically for 7 days at 37 °C, prior to genomic DNA purification.

2.4.2. Genomic DNA Purification

A 1.5 ml sample of each Desulfovibrio liquid culture was centrifuged for 5 min at 5,900 x g. Desulfovibrio genomic DNA was then purified from pellets using the QiAamp DNA Mini Kit (Qiagen, Hilden, Germany), following to the manufacturer's instructions until the DNA elution step. Purified Desulfovibrio genomic DNA was eluted into 80 µl of nuclease-free water (Thermo-Fisher Scientific Inc.) and stored at - 20 C until required.

2.4.3. Analysis of Purified Genomic DNA by Agarose Gel Electrophoresis

Agarose gels were prepared by dissolving 0.5 % (w/v) agarose in TAE buffer (40 mM Tris-Base, 20 mM glacial acetic acid and 1 mM EDTA; pH 8.6), supplemented with 0.01 % (v/v) SYBR® safe DNA gel stain (ThermoFisher scientific Inc.). Purified genomic DNA was mixed with DNA loading dye (ThermoFisher scientific Inc.), at a volume ratio of 6:1 respectively. Prepared solutions and a GeneRuler 1 kb DNA Ladder (ThermoFisher scientific Inc.) were loaded onto an agarose gel. Electrophoresis was performed at 130 V, 400 mA for 45 min in TAE buffer, using a PerfectBlue™ Horizontal Minigelsystems (PEQLAB Ltd., Fareham, UK). DNA was visualised using a UV trans-illuminator (UVP, California, U.S.A.).

2.4.4. Quantification of Purified Genomic DNA

Purified genomic DNA was quantified using the Qubit® 2.0 Fluorometer with a Qubit® dsDNA High Sensibility Assay Kit (Thermo-Fisher Scientific Inc.). A Qubit working solution was prepared by diluting the fluorescent DNA dye 1:200 in the Qubit® dsDNA high sensibility buffer. The Qubit working solution was added to 3 μl of purified genomic DNA and to 10 μl of provided standards with known concentration, to a final volume of 200 μl. Samples and standards were vortexed and incubated for 2 min at room temperature, prior to being analysed using Qubit® 2.0 Fluorometer. DNA standards with known concentrations were analysed before the purified genomic DNA samples, allowing accurate DNA quantification.

2.4.5. Preparation of Genomic DNA Sequencing Libraries

D. desulfuricans NCIMB 8326, *D. gigas* NCIMB 9332, *D. gabonensis* NCIMB 10636, *D. marinus* DSM 18311 and *D. giganteus* DSM 4370 genomic DNA were sent to Shell Technology Centre for library preparation. Sequencing libraries for *D. paquesii* DSM 16681, *D. desulfuricans* NCIMB 8338, *D. desulfuricans* NCIMB 8307, *D. alcoholivorans* NCIMB 12906 genomes and for the DNA purified from uninoculated sodium lactate medium (negative control) were prepared at Exeter University.

Purified genomic DNA was prepared for sequencing into libraries of 300 base paired-end fragments following the instructions of Nextera XT DNA Library Preparation Kit (Illumina Inc., San Diego, USA). Genomic DNA was first fragmented and ligated to sequencing adapters, in one step called tagmentation. Tagmented DNA fragment were then amplified by polymerase chain reaction (PCR) with primers hybridising adapter sequences. PCR primers had an overhang composed of full adapter sequences, sequencing index and complementary sequence to the flow cell oligo. Sequencing library DNA fragments flanked by adapters, index and region complementary to the flow cell oligo were finally purified to remove PCR reagents and short fragments.

2.4.6. Quality Control of Genomic DNA Sequencing Libraries

The quality of genomic DNA sequencing libraries was assessed using a 2200 TapeStation (Agilent Technologies, Santa Clara, USA) with D1000 ScreenTape (Agilent Technologies), following the manufacturer's instructions.

2.4.7. Genomic DNA Sequencing

An equimolar pool of the respective DNA sequencing libraries was sequenced using an Illumina MiSeq™ at Shell Technology Centre (Houston, USA),

2.5. Bioinformatic processing

Sequence data was accessed and all subsequent analyses was performed using a local server containing 32 3.1 GHz CPUs and 256 Gb RAM. The system was installed with Fedora version 2.1 Linux operating system.

Throughout this section, exemplar command lines are written with the appropriate methods using the `courier font`.

2.5.1. Quality Analysis of the Sequencing Paired-End Reads and Genome *De novo* Assembly

The quality of the sequencing raw reads was assessed using FastQC (version 0.10.1; http://www.bioinformatics.babraham.ac.uk/projects/fastqc; Andrews, 2010).

Any DNA read bases with a Phred score below 20 and the sequencing adapters from the 3' end of reads were removed using TrimGalore (version 0.3.3; Kruegar, 2014). The parameter "`paired`" of TrimGalore was also applied to the raw reads to remove paired-end reads if at least one read sequence was shorter than the length threshold of 20 base pairs (bp).

Trimmed paired-end reads were merged using the program Fast Length Adjustment of Short Reads (FLASH; version 1.2.7), with default parameters (Magoč & Salzberg, 2011). Non-overlapping reads were stored in separated files.

Trimmed and "flashed" reads were assembled *de novo* using SPAdes (version 3.8.0) with the parameter "`careful`" (Bankevich *et al.,* 2012). The "`careful`"

parameter was used to reduce the number of mismatches. The program was run using 16 CPUs.

2.5.2. Genome *De novo* Assembly Quality Assessment

The quality of the *Desulfovibrio* genome assemblies was evaluated using Quality Assessment Tool for Genome Assemblies (QUAST, version 3.0) using default parameters (Gurevich *et al.*, 2013). The program was run using 16 CPUs.

The genome per-base coverage was determined using the alignment software tool Bowtie2 (version 2.1.0; Langmead *et al.*, 2009). *De novo* assembled *Desulfovibrio* genomes were indexed using the `bowtie-build` command line. Raw paired-end reads used for building a genome *de novo* assembly were then aligned to the genome assembly, using the command line `Bowtie2`. An alignment was defined as "valid" if the insert size between the alignment of a paired-end read was comprised between 1 bp to 100,000 bp. Bowtie2 was also programmed to report up to 5 valid alignments per paired-end reads. Resultant valid alignments were provided in a SAM (Sequence Alignment Map) file. The program was run using 16 CPUs.

Indexing –

```
bowtie-build denovo_assembly_x.fasta denovo_assembly_x_index
```

Aligning –

```
bowtie2 -k 5 -I 1 -X 100000 –x denovo_assembly_x_index -1
x_raw_data_read1.fastq -2 x_raw_data_read2.fastq -S
coverage_x.sam  -p 16
```

The SAM file generated for each *Desulfovibrio* genome assembly was sorted and indexed into a BAM format file using SAMtools (version 1.3.1; Li *et al.*, 2009). Quality control analysis of the BAM files was performed using the application Qualimap 2 (version 2.2), to determine the per-base coverage for each genome (García-Alcalde *et al.*, 2012; Okonechnikov *et al.*, 2015).

The accuracy of the *Desulfovibrio* genome assemblies was determined using the program Recognition of Errors in Assemblies using Paired Reads (REAPR; version 1.0.15; Hunt *et al.*, 2013), for all the *Desulfovibrio* genome assemblies in this study. First, a perfect and uniquely mapping read coverage was generated for each assembly, using

the command line `perfectmap`. Second, the flashed paired-end reads used for building a genome assembly were mapped against the genome assembly using the command line `smaltmap`. Resultant alignments were provided in a sorted and indexed BAM file. The BAM file was screened for anomalies in coverage patterns of an assembly by the reads used to build the assembly, using the command line `pipeline`.

Mapping –

```
reapr perfectmap denovo_assembly_x.fa x_raw_read1.fastq
x_raw_read2.fastq 300 x_reapr_perfectmap.bam

reapr smaltmap denovo_assembly_x.fa x_flashed_read1.fastq
x_flashed_read2.fastq x_reapr_smaltmap.bam
```

Screening for assembly errors –

```
reapr pipeline denovo_assembly_x.fa x_reapr_smaltmap.bam
reapr_stat_x_assembly x_reapr_perfectmap.bam
```

For the *Desulfovibrio* genome with a sequence available in GenBank database (Benson *et al.*, 2013), the accuracy of the *Desulfovibrio* genome assemblies was also assessed using the wrapper dnadiff from NUCmer (version 1.3; Kurtz *et al.*, 2004),.

Genome assemblies were annotated using Prokka (Seemann, 2014). The parameter "`rfam`" was applied to annotate the non-coding RNAs. The program was run using 16 CPUs.

The completeness of the *Desulfovibrio* genome assemblies was estimated using Benchmarking Universal Single-Copy Orthologs (BUSCO; version 1.22) with the BUSCO lineage dataset for bacteria (Simão *et al.*, 2015). The program was run using 16 CPUs.

2.5.3. Alkane Producing *Desulfovibrio* Screening for Characterised Alka(e)ne Biosynthetic Enzymes

A nucleotide BLAST (Basic Local Alignment Search Tool) database was built with each alkane producing *Desulfovibrio* assembled genome, using the command line `makeblastdb` from the software BLAST+ (version 2.2.28; Altschup *et al.*, 1990;

Camacho *et al.,* 2009). Characterised alka(e)ne biosynthetic enzyme sequences were aligned to the constructed databases, using the command `tblastn`, with an E-value threshold of $10e^{-5}$.

Characterised alka(e)ne biosynthetic enzymes were screened for protein domains in the profile-Hidden Markov Models (HMM) database Pfam (Sonnhammer *et al.,* 1997), using the command line `hmmscan` from the software package HMMER 3.1b1 (http://hmmer.org). Alkane producing *Desulfovibrio* proteomes were scanned for the protein domain HMM profiles of the characterised alka(e)ne biosynthetic enzymes, using the program `hmmsearch` from the software package HMMER 3.1b1, with an E-value threshold equal or superior to $10e^{-5}$.

2.5.4. Phylogenetic Analysis

The 16S rRNA gene sequence of *Desulfovibrio* type strains and *Desulfobacter postgatei* DSM 2034 type strain was extracted from the Ribosomal Database Project (RDP; release 11.4; https://rdp.cme.msu.edu). Alignment of the 16S rRNA gene sequences was performed using the multi-sequence alignment program MAFFT (version 7.245) with the parameter "`auto`" (Katoh *et al.*, 2002). A phylogenetic tree was constructed by the approximately maximum likelihood method using the program FastTree (version 1.9.0) with the parameter "`nt`" (Price *et al.*, 2010). The tree was graphically represented using FigTree (version 1.4.2; http://tree.bio.ed.ac.uk/software/figtree/).

2.5.5. Genomic Comparison through a Sequence based Approach

Desulfovibrio genomes generated in this study were aligned to the model organism *D. vulgaris* strain Hildenborough genome from GenBank (AE017285.1) using the program `blastn` from the software BLAST+ (version 2.2.28), with an E-value threshold of $10e^{-5}$. Visualisation of the *Desulfovibrio* genomic alignment was generated by the tool BLAST Ring Image Generator (BRIG; version 0.95; Alikhan *et al.*, 2011). BRIG was configured to display the alignments with a minimum percentage identity of 70 %.

2.5.6. Genomic Comparison through a Gene Content based Approach

Two *Desulfovibrio* pan-genomes were independently generated using the platform ANalysis and VIsualization platform for 'Omics data (Anvi'o; version 2.1.1; Eren *et al.*, 2015) and the software *get_homologues* (version 2.0; Contreras-Moreira & Vinuesa, 2013).

Desulfovibrio pan-genome generated by the platform Anvi'o

The FASTA files of the *de novo* assembled *Desulfovibrio* genomes from this study and the *D. vulgaris* strain Hildenborough genome (GenBank AE017285.1) were used as input data for Anvi'o pan-genomic analysis. The FASTA files were screened for open reading frames using the program PROkaryotic DYnamic Programming Gene finding ALgorithm (Prodigal; Hyatt *et al.* 2010), by the command line `anvi-script-FASTA-to-contigs-db`. Identified proteins were then annotated using the Clusters of Orthologous Groups of proteins (COGs) database (Tatusov *et al.*, 2000), by the program `anvi-run-ncbi-cogs`. The program was run using 16 CPUs. *Desulfovibrio* annotated genomes were gathered and stored into a specific Anvi'o database, using the command line `anvi-gen-genomes-storage`.

To generate *Desulfovibrio* pan-genome, identified protein sequences were aligned to each other for sequence similarity searching, using the software Double Index AlignMent Of Next-generation sequencing Data (DIAMOND; version 0.7.9; Buchfink *et al.*, 2015) with the parameter "`sensitive`". Proteins were then clustered into homologous group using Markov CLuster (MCL) algorithm (van Dongen & Abreu-Goodger, 2012), with an inflation value of 7. The Anvi'o default maxbit score threshold was set to 0.5. The MCL algorithm was configured to report all the predicted clusters, including those which occurred in only one genome. Finally, protein clusters were hierarchically organised by measuring Euclidean distance between clusters and using the agglomerative Ward linkage clustering method.

Anvi'o pan-genome generation –

```
anvi-pan-genome -g desulfovibrio-GENOMES.h5 -J
desulfovibrio_pangenome --num-threads 16 --sensitive --mcl-
inflation 7
```

Desulfovibrio pan-genome generated by the software get_homolgues

The GenBank files of the Prokka annotated *Desulfovibrio* genomes from this study and the *D. vulgaris* strain Hildenborough genome (GenBank AE017285.1) were used as input data for *get_homologues* pan-genomic analysis. The GenBank files were aligned to each other for sequence similarity searching, using BLAST+ (version 2.2.28). A minimum pairwise alignment coverage of 75% and an E-value of $10e^{-5}$ were imposed to the protein sequence similarity searching. Protein and nucleotide sequences were then clustered independently using COGtriangles (Kristensen *et al.*, 2010) and OrthoMCL (Li *et al.,* 2003) algorithms. COGtriangles and OrthoMCL algorithms were configured to report all the predicted clusters, including those which occurred in only one genome.

Sequence similarity searching and COGtriangles clustering –

```
get_homologues.pl -d annotated_desulfovibrio_genomes -t 0 -G
```

Sequence similarity searching and OrthoMCL clustering –

```
get_homologues.pl -d annotated_desulfovibrio_genomes -t 0 -M
```

The two sets of protein clusters identified independently by COGtriangles and OrthoMCL algorithms were then intersected, in order to generate a consensus pan-genome matrix. The parameter "`m`" and "`T`" were used to produce intersection pan-genome matrices and a parsimony-based pan-genomic tree respectively.

```
compare_clusters.pl -o cluster_set_intersection -m -T
-d COGtriangles_clusters,OrthoMCL_clusters
```

The composition of *Desulfovibrio* pan-genome by *get_homologues* was determined using the command line `parse_pangenome_matrix.pl` with the parameter "`s`".

```
parse_pangenome_matrix.pl -m pangenome_matrix_t0.tab -s
```

2.5.7. Identification of Protein Clusters Exclusive to Alkane Producing *Desulfovibrio*

Anvi'o *Desulfovibrio* pan-genome was displayed through the Anvi'o interactive interface, which enabled the collection of protein clusters exclusively present in alkane producing strains into a "bin" named "PC shared only by alkane producing strains". Information about the protein clusters gathered into the bin "PC shared only by alkane producing strains" were accessed by the command line `anvi-summarize`.

To identify protein clusters exclusively present in alkane producing strains from the *get_homologues Desulfovibrio* pan-genome, two text files were generated one regrouping the alkane producing *Desulfovibrio* GenBank files, the second regrouping the non-alkane producing *Desulfovibrio* GenBank files. The protein clusters present in alkane producing *Desulfovibrio* spp. and absent in non-alkane producing *Desulfovibrio* spp. were identified using the command line `parse_pangenome_matrix.pl`.

```
parse_pangenome_matrix.pl –m pangenome_matrix_t0.tab
-A alkane_producing_desulfovibrio.txt –B
non_alkane_producing_desulfovibrio.txt –g
```

To screen protein clusters predicted to be exclusively present in alkane producing *Desulfovibrio* spp. for false positive clusters, protein sequences of each cluster were aligned using the command line `linsi` from MAFFT. A HMM profile was built from the protein sequence alignment using the program `hmmbuild` from the HMMER 3.1b1. A single consensus sequence from the HMM profile was then generated using the program `hmmemit` from HMMER 3.1b1, with the parameter "`c`". The parameter "`c`" allows the generation of a single consensus sequence by selecting the maximum probability residue at each position in the protein sequence. The consensus sequence of each protein cluster exclusively present in alkane producing *Desulfovibrio* spp. was used as a query for a `blastp` search against the non-alkane producing *Desulfovibrio* proteomes, with an E-value threshold of $10e^{-5}$.

To establish a list of protein clusters considered exclusive to alkane producing *Desulfovibrio* spp., a protein BLAST database was built with the consensus sequence of protein clusters predicted to be exclusive to alkane producing *Desulfovibrio* spp. by the program Anvi'o, using the command line `makeblastdb` from BLAST+. The consensus

sequence of protein clusters predicted to be exclusive to alkane producing *Desulfovibrio* spp. by the program *get_homologues* was used as a query for a `blastp` search against the Anvi'o protein cluster consensus sequence database.

2.5.8. Identification of Protein Clusters Potentially Involved in Alkane Production

To verify protein cluster annotation, protein cluster consensus sequences were used as a query for a `blastp` search against the protein database UniProtKB (http://www.uniprot.org; Bateman *et al.*, 2017) and for a `rpsblast` (version 2.2.15) search against the COG database. The `blastp` and the `rpsblast` searches were programmed with an E-value threshold of $10e^{-5}$. Moreover, protein cluster HMM profiles were used as a query for a `hmmsearch` search against UniProtKB. The E-value threshold was configured at $10e^{-5}$.

The tertiary structure of protein clusters exclusive to alkane producing *Desulfovibrio* spp. were predicted from their consensus sequence using the program RaptorX (http://raptorx.uchicago.edu; Källberg *et al.,* 2012).

The RCSB Protein Data Bank (http://www.rcsb.org/; Berman *et al.*, 2000) was used to identify the Structural Classification of Proteins (SCOP; Murzin *et al.*, 1995) fold category of the best template protein structure used by RaptorX.

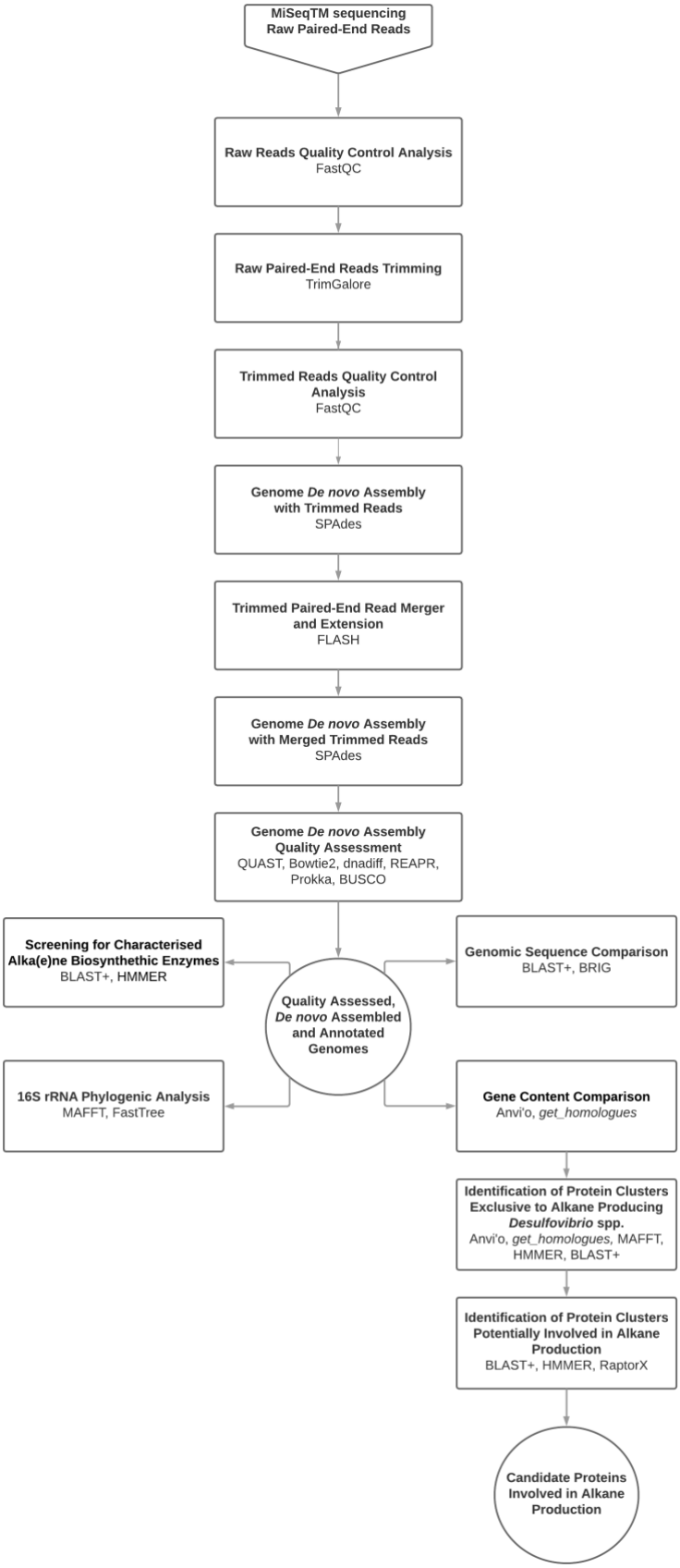


Figure 2.1. Flowchart of the bioinformatic processes performed for elucidation of *Desulfovibrio* alkane biosynthetic pathway Flowchart rendered in Lucidchart (http://www.lucidchart.com)

2.6. Molecular Biology Methods

2.6.1. DNA Digestion by Restriction Enzymes

Restriction enzyme DNA digestions were performed with 100 ng to 1 μg DNA fragment, 10 U FastDigest enzymes (Thermo-Fisher Scientific Inc.), 1-fold FastDigest buffer (Thermo-Fisher Scientific Inc.), in a final volume of 20 μl nuclease-free water (Thermo-Fisher Scientific Inc.). Digestion mixtures were incubated for 20 min at 37 ºC, then for 20 min at 65 ºC for enzyme inactivation. Digested DNA fragments were visualised on an appropriate percentage (w/v) agarose gel, as described in section 2.4.3.

2.6.2. DNA Ligation

Ligation reactions were catalysed by 5 Weiss U T4 DNA Ligase (Thermo-Fisher Scientific Inc.) in a final volume of 20 μl nuclease-free water (Thermo-Fisher Scientific Inc.) containing 3-fold molar excess of DNA fragment insert over 25 ng of DNA destination plasmid, 1-fold T4 DNA Ligase buffer (Thermo-Fisher Scientific Inc.) and 500 μM ATP. Ligation mixtures were incubated at room temperature for 1 h.

2.6.3. *Desulfovibrio* Sensitivity Assay to Geneticin G418

200 μl of wild type *Desulfovibrio* liquid cultures inoculated with a single colony were cultured on solid sodium lactate medium, supplemented with geneticin G418 to a final concentration of 400 μg ml^{-1}, 500 μg ml^{-1} and 600 μg ml^{-1}. 200 μl of wild type *Desulfovibrio* liquid cultures were also cultured on solid sodium lactate medium without geneticin G418, as positive controls. Plates were incubated anaerobically for 7 days at 37 °C.

2.6.4. Preparation of Competent *E. coli* Cells

E. coli cells were made competent by a modified protocol of Hanahan chemical treatment (Hanahan, 1985). Liquid cultures of *E. coli* were prepared by inoculating a single colony into 5 ml LB medium for NEB 5-alpha and BL21 Star™ (DE3) *E. coli* or LC medium supplemented with 0.3 mM DAP for WM3064 *E. coli*. Cultures were incubated

aerobically at 37 °C overnight with 220 rpm agitation. 400 µl of *E. coli* overnight cultures were used to inoculate 40 ml LB medium for NEB 5-alpha and BL21 Star™ (DE3) *E. coli* or LC medium supplemented with 0.3 mM DAP for WM3064 *E. coli*. Cultures were incubated aerobically at 37 °C with 220 rpm agitation until achieving an optical density of 0.4 at 600 nm and were then centrifuged for 10 min at 2,300 x *g* and at 4 °C. Bacterial pellets were resuspended into 8 ml ice-cold TF-1 buffer (15 % (w/v) glycerol, 100 mM KCl, 30 mM CH_3CO_2K, 10 mM $CaCl_2.2H_2O$, 50 mM $Cl_2Mn.4H_2O$; pH 6.4) and incubated for 15 min on ice. *E. coli* cells suspended in TF-1 buffer were then centrifuged for 10 min at 2,300 x *g* at 4 °C and supernatants were discarded. Pellets were resuspended into 4 ml ice-cold TF-2 buffer (15 % (w/v) glycerol, 10 mM KCl, 75 mM $CaCl_2.2H_2O$, 10 mM MOPS buffer; pH 6.8). 100 µl aliquots of *E. coli* cells suspended in TF-2 buffer were sampled and stored at - 80 °C until required.

2.6.5. Bacterial Transformations

E. coli transformation by heat-shock

Competent *E. coli* cells were transformed with desired plasmid DNA using a heat-shock protocol. 100 µl aliquots of competent cells suspended in TF-2 buffer were mixed with 100 ng of desired plasmid DNA. The bacterial cell and plasmid mixtures were incubated for 40 min on ice, followed by 2 min at 42 °C and subsequently 5 min on ice. 250 µl of LB medium pre-warmed at 37 °C was added to the mixtures. The mixtures were incubated aerobically for 1 h at 37 °C with 220 rpm agitation. NEB 5-alpha and BL21 Star™ (DE3) *E. coli* transformants were then screened by plating 100 µl cell suspension on solid LB medium supplemented with 50 µg ml^{-1} kanamycin. WM3064 *E. coli* transformants were then screened by plating 100 µl cell suspension on solid LC medium supplemented with 0.3 mM DAP and 50 µg ml^{-1} kanamycin. Plates were incubated aerobically, overnight at 37 ºC in a static incubator.

Desulfovibrio transformation by electroporation

The protocols followed for *Desulfovibrio* transformation by electroporation were based on the protocol provided by Judy Wall laboratory (University of Missouri). *Desulfovibrio* cells from 50 ml liquid cultures, inoculated with a single colony and incubated anaerobically at 37 ºC to early stationary growth phase, were made electro-competent by five washes in 50 ml chilled, sterile 30 mM Tris-HCl buffer (pH 7.2). After the last wash, electro-competent *Desulfovibrio* cells were resuspended into 500 µl chilled,

sterile 30 mM Tris-HCl buffer (pH 7.2). 50 µl aliquots of electro-competent *Desulfovibrio* cells were mixed with different amounts of desired DNA plasmid (0.1 µg, 0.5 µg, 0.75 µg, 1 µg, 5 µg and 10 µg). The *Desulfovibrio* cell and plasmid mixtures were transferred to chilled 1 mm-gapped electroporation cuvettes (Molecular BioProducts Inc., San Diego, USA). *Desulfovibrio* electroporation was performed using a Gene Pulser model 165-2078 coupled with a Gene Controller model 165-2098 (Bio-Rad, Watford, UK). Various field strengths (5 kV cm^{-1}, 7.5 kV cm^{-1}, 10 kV cm^{-1}, 12.5 kV cm^{-1}, 15 kV cm^{-1}, 17.5 kV cm^{-1} and 20 kV cm^{-1}) were applied to the cuvettes by adjusting the pulse voltage. The resistance and the capacitance of the electroporator were set up at 25 µF and 200 Ω to 400 Ω respectively. After pulsing, electroporated *Desulfovibrio* cells were incubated overnight in 1 ml sodium lactate medium under anaerobic atmosphere at 37 °C. *Desulfovibrio* transformants were then screened by plating 200 µl to 700 µl cell suspension on solid sodium lactate medium supplemented with 400 µg ml^{-1} geneticin G418. The plates were incubated anaerobically for 7 days at 37 °C.

Desulfovibrio transformation by conjugation with MW3064 E. coli

The protocol followed for *Desulfovibrio* transformation by conjugation with MW3064 *E. coli* was based on the protocol provided by Judy Wall laboratory (University of Missouri). Liquid cultures of wild type *Desulfovibrio* were prepared by inoculating a single colony into 5 ml sodium lactate medium. After 3-6 days incubation at 37 °C under anaerobic atmosphere, samples of 1.5 ml *Desulfovibrio* cultures were centrifuged for 8 min at 5,900 x *g* and pellets were resuspended into 100 µl of sodium lactate medium. Liquid cultures of MW3064 *E. coli* transformant (containing the plasmid DNA of interest) were prepared by inoculating a single colony into 5 ml LC medium supplemented with 0.3 mM DAP and 50 µg ml^{-1} kanamycin, and were incubated aerobically overnight at 37 °C with 220 rpm agitation. 200 µl of *E. coli* overnight cultures were used to inoculate 5 ml LC medium supplemented with 0.3 mM DAP and 50 µg ml^{-1} kanamycin. Cultures were incubated aerobically for 4 h at 37 ºC and 100 rpm agitation. Samples of 700 µl *E. coli* subcultures were then centrifuged for 8 min at 5,900 x *g*. *E. coli* pellets were resuspended with the 100 µl *Desulfovibrio* cell resuspension. The mating mixes were spotted onto sterile membrane filters (0.45 µm pore size; Whatman), placed on solid sodium lactate medium. Plates were incubated anaerobically overnight at 37 ºC. Membrane filters were then soaked into 5 ml sodium lactate medium and were incubated anaerobically for 4 h at 34 ºC and 100 rpm agitation. *Desulfovibrio* transformants were then screened by plating 200 µl cell suspension on solid sodium lactate medium supplemented with 400 µg ml^{-1} geneticin G418. The plates were incubated anaerobically for 4-7 days at 37 °C.

2.6.6. Plasmid Purification

Plasmids were purified from 10 ml cultures incubated aerobically overnight at 37 °C for *E. coli* transformants and incubated anaerobically for 3 days at 37 °C for *D. vulgaris* Hildenborough transformants. Plasmid purifications were performed using a GeneJet Plasmid Miniprep Kit (Thermo-Fisher Scientific Inc.), following the manufacturer's instructions until the plasmid elution step. Purified plasmids were eluted into 30 *µ*l of nuclease-free water (Thermo-Fisher Scientific Inc.) and stored at - 20 °C until required.

2.6.7. Verification of the Plasmid pPD3 Transfer into *D. vulgaris* Hildenborough

Purified plasmids from *D. vulgaris* Hildenborough transformants were digested with EcoRI and PstI restriction enzymes, as described in section 2.6.1. The purified plasmids, *D. vulgaris* Hildenborough genomic DNA and the plasmid pPD3 were also used as template for PCR using the primers Verif_PCR_pPD3_F and Verif_PCR_pPD3_R (Table 2.3). The following PCR mixture was prepared: 1-fold Q5 Buffer (New England BioLabs Ltd.), 200 *µ*M deoxynucleoside triphosphates (New England BioLabs Ltd.), 0.5 *µ*M Verif_PCR_pPD3_F, 0.5 *µ*M Verif_PCR_pPD3_R, 1 U Q5 DNA polymerase (New England BioLabs Ltd.) and 100 ng purified plasmid or genomic DNA in a final volume of 50 *µ*l nuclease–free water (Thermo-Fisher Scientific Inc.). Reaction conditions started with 30 s DNA denaturation at 98 °C followed by 30 cycles consisting of 10 s at 98 °C, 30 s at 64 °C and 1 min at 72 °C, ending with a 5 min extension at 72 °C. Amplified DNA fragments were visualized on an appropriate percentage (w/v) agarose gel, as described in section 2.4.3.

Table 2.3. PCR primers used for verification of pPD3 transfer into *D. vulgaris* Hildenborough

Primers	Sequence (5' - 3')
Verif_PCR_pPD3_F	CTTTGGCTGTAGGTGCTAGG
Verif_PCR_pPD3_R	CGTCGGTGAGCCAGAGTTTC

2.6.8. Plasmid Quantification

Purified plasmids were quantified using the Qubit® 2.0 Fluorometer with a Qubit® dsDNA Broad Range Assay Kit (Thermo-Fisher Scientific Inc.), using the protocol described in section 2.4.4.

2.6.9. Verification of Plasmid Constructs and Sequences

Plasmid constructs were assessed by restriction enzyme digestions, as described in section 2.6.1. Plasmid sequence were verified by Sanger sequencing, performed by Source Bioscience (Nottingham, UK). Primers used for plasmid sequence verification are listed in table 2.4. Sequencing results were evaluated for coverage and sequence similarity using Clone Manager Professional edition (version 9; Scientific & Educational Software, Denver, USA)

Table 2.4. Primers used for plasmid sequence verification

Insert sequence verification of pEX1K3-Reductase and pEX1K3-Methyltransferase	
Primer name	Primer sequence (5' – 3')
Veri_pEX1K3_F	GCCACCTGACGTCTAAGAAA
Insert sequence verification of pPD3-Reductase and pPD3-Methyltransferase	
Primer name	Primer sequence (5' – 3')
Veri_pPD3_F	TACATCACCGACGAGCAAGG
Insert sequence verification of pEX1K3-Reductase, pEX1K3-Methyltransferase, pPD3-Reductase and pPD3-Methyltransferase	
Primer name	Primer sequence (5' – 3')
Veri _R	ATTACCGCCTTTGAGTGAGC
Whole pPD3 sequence verification	
Primer name	Primer sequence (5' – 3')
pPD3_seq_10F	AAAAGGGCAAGGTGTCACCA
pPD3_seq_148R	GCCTTTGAGTGAGCTGATAC
pPD3_seq_464F	GTATCTCAGTTCGGTGTAGG
pPD3_seq_639R	CATACCTCGCTCTGCTAATC
pPD3_seq_1021F	CCGTCAAGTCAGCGTAATGC
pPD3_seq_1186R	CCTCGGTGAGTTTTCTCCTT
pPD3_seq_1511F	AACACTGCCAGCGCATCAAC
pPD3_seq_1672R	CTGGCTGACGGAATTTATGC
pPD3_seq_1982F	TGAGACACAACGTGGCTTTG
pPD3_seq_2113R	GCCTCGTGATACGCCTATTT
pPD3_seq_2546F	CGGCGTAGCGATAACGAAGC
pPD3_seq_2670R	GGGAAGCCTGTTCTAATAGC
pPD3_seq_3038F	ACTTGGCACTGGAGGTTGTC
pPD3_seq_3137R	ACAGCAGCGACGCCAAGAAC
pPD3_seq_3503F	GTTTCAGATCGGGATGGAAG
pPD3_seq_3664R	TATCCGGCAAGAGGTACAAG
pPD3_seq_3968F	TGGTCACCCAAAGGCTACAC
pPD3_seq_4149R	CGCGCGCTTCTTCTTTTCTT
pPD3_seq_4496F	CTTGTCGACTTCCCAATTCC
pPD3_seq_4687R	CTGACGCCGTTGGATACACC
pPD3_seq_4965F	TCGGCCAGGGCTACAAAATC
pPD3_seq_5150R	CCTGCTTCTCTTCGATCTTC

2.7. Plasmid Vectors; Design and Construction

2.7.1. Plasmids pEX1K3-Reductase and pEX1K3-Methyltransferase, *E. coli* Expression Vectors

The coding sequences of a putative reductase and a putative methyltransferase from *D. desulfuricans* NCIMB 8326 were separately cloned downstream of the T7 promoter sequence, fused to the lactose operon operator gene and a *E. coli* ribosome-binding site (RBS) sequences, and upstream of the T7 terminator sequence into the plasmid pEX1K3-ccdB, using a modular cloning strategy (Weber *et al.*, 2011; Figures 2.2 and 2.3). The modular cloning strategy allows the ordered assembly of DNA fragments into a destination vector in one reaction. As a prerequisite, the DNA fragments to be cloned were flanked with defined sequences at the 5' terminus, called prefixes, and at the 3' terminus, called suffixes. Prefix and suffix sequences flanked to a DNA fragment were determined by the cloning position of the DNA fragment into the final vector. All prefix and suffix sequences contained a restriction site for the BsaI enzyme. Once cut with BsaI, prefix and suffix sequences had compatible cohesive extremities which allowed the ordered cloning of DNA fragments into a destination vector. Another prerequisite of the modular cloning strategy was the presence of the *ccdB* gene flanked by BsaI restriction sites in the destination vector. Successful cloning results in the removal of the *ccdB* gene from the destination vector, which ensures the growth of bacterial cells transformed with a vector of correct construction.

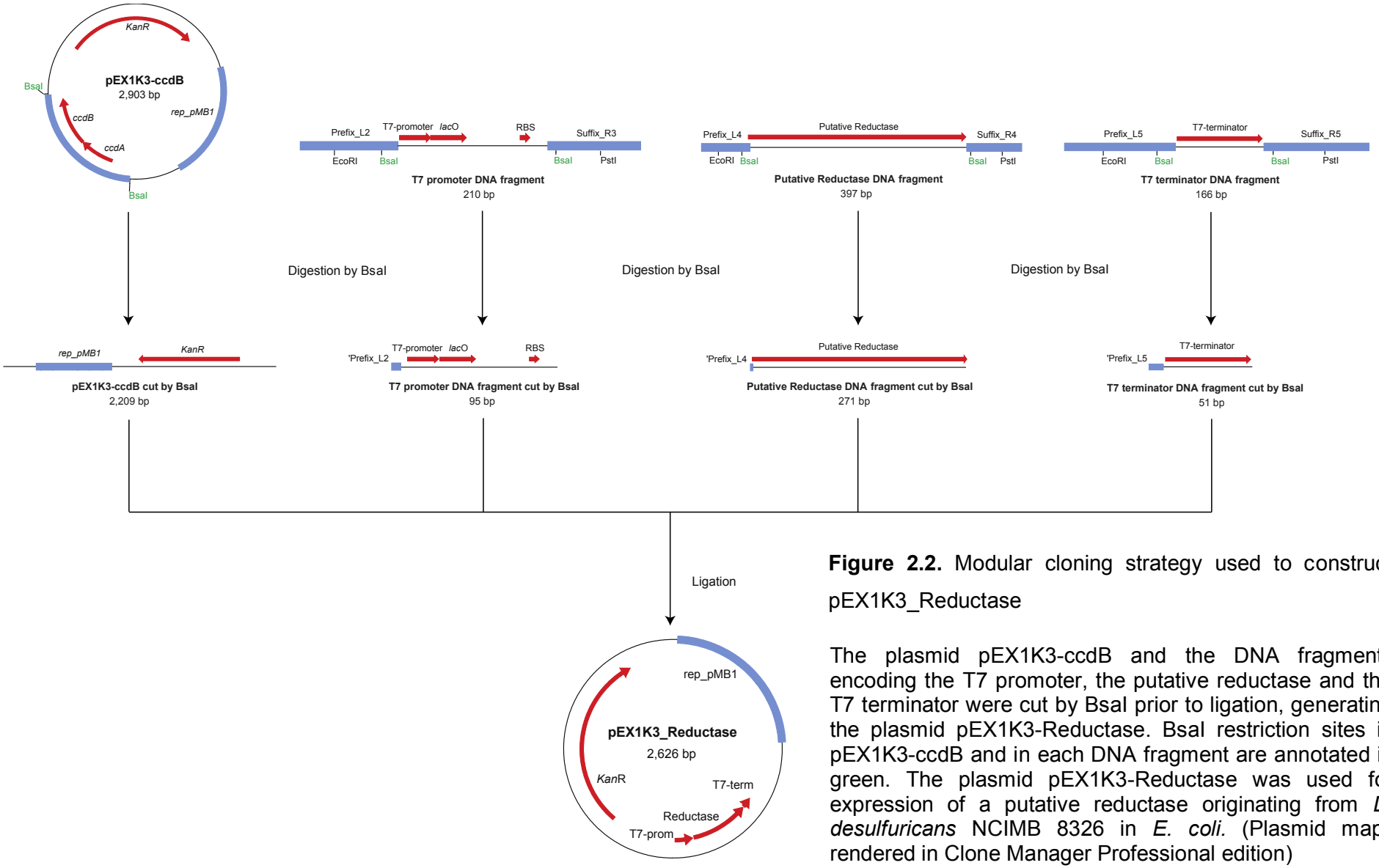


Figure 2.2. Modular cloning strategy used to construct pEX1K3_Reductase

The plasmid pEX1K3-ccdB and the DNA fragments encoding the T7 promoter, the putative reductase and the T7 terminator were cut by BsaI prior to ligation, generating the plasmid pEX1K3-Reductase. BsaI restriction sites in pEX1K3-ccdB and in each DNA fragment are annotated in green. The plasmid pEX1K3-Reductase was used for expression of a putative reductase originating from *D. desulfuricans* NCIMB 8326 in *E. coli.* (Plasmid maps rendered in Clone Manager Professional edition)

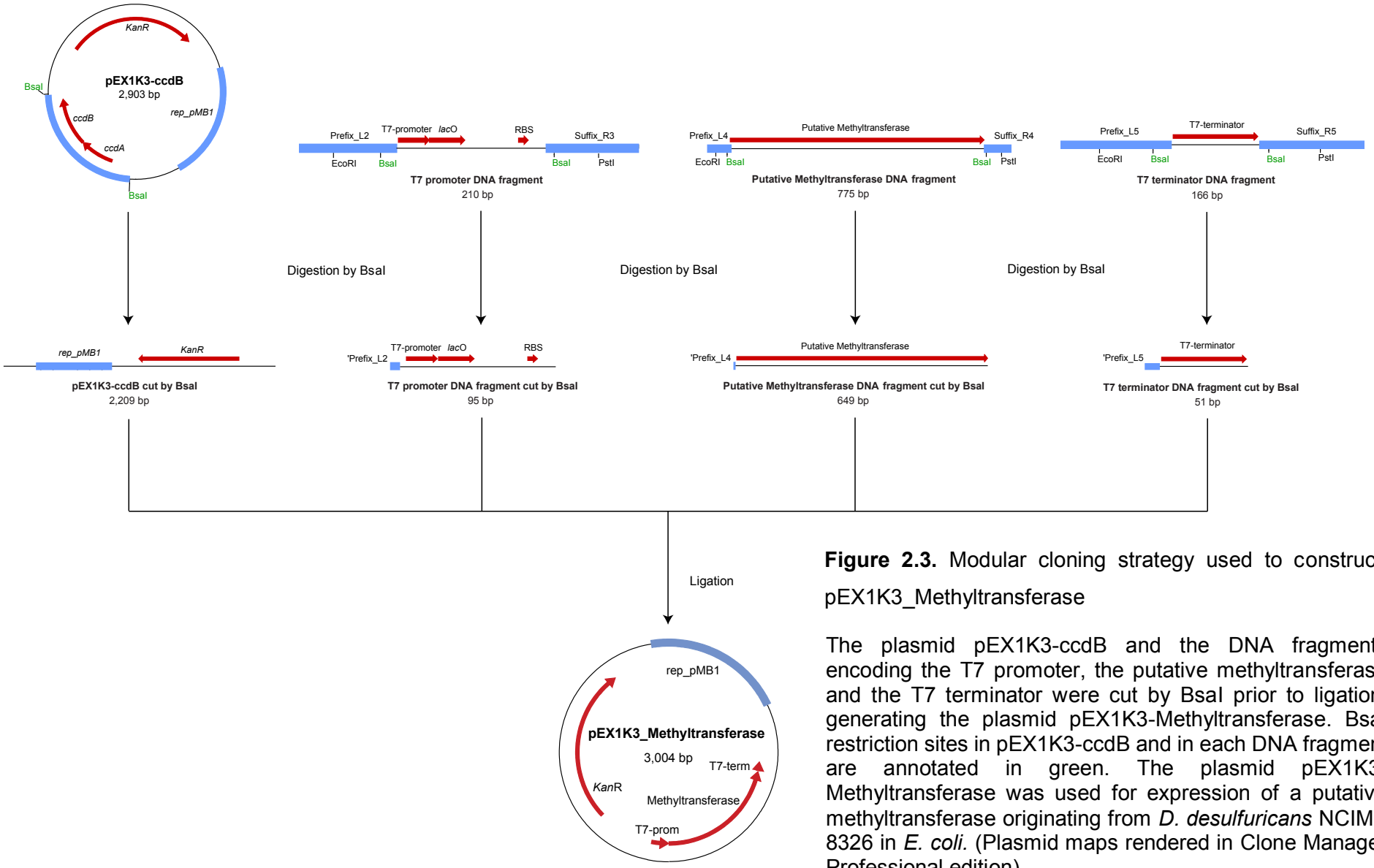


Figure 2.3. Modular cloning strategy used to construct pEX1K3_Methyltransferase

The plasmid pEX1K3-ccdB and the DNA fragments encoding the T7 promoter, the putative methyltransferase and the T7 terminator were cut by BsaI prior to ligation, generating the plasmid pEX1K3-Methyltransferase. BsaI restriction sites in pEX1K3-ccdB and in each DNA fragment are annotated in green. The plasmid pEX1K3-Methyltransferase was used for expression of a putative methyltransferase originating from *D. desulfuricans* NCIMB 8326 in *E. coli.* (Plasmid maps rendered in Clone Manager Professional edition)

The DNA coding sequences of the putative reductase and methyltransferase were extracted from the *de novo* assembled *D. desulfuricans* NCIMB 8326 genome within this study. The DNA sequence encoding a poly-histidine tag "Nterm-MAGSHHHHHGS-Cterm" was added to the DNA coding sequences of the putative reductase and methyltransferase. The histidine-tagged putative reductase, the histidine-tagged putative methyltransferase, the T7 promoter and the T7 terminator DNA fragments flanked with prefix and suffix sequences were synthesised by Twist Biosciences (San Francisco, USA; Table 2.5).

The DNA fragments received from Twist Biosciences were resuspended into nuclease-free water (Thermo-Fisher Scientific Inc.) to a final concentration of 10 ng μl^{-1}. Modular cloning reactions were performed with 20 fmol of desired DNA fragments, 20 fmol of pEX1K3-ccdB, 1-fold FastDigest buffer, 10 U FastDigest BsaI, 5 Weiss U T4 DNA Ligase and 500 μM ATP in a final volume of 20 μl nuclease-free water (Thermo-Fisher Scientific Inc.). Molecular cloning reaction conditions consisted of 37 °C for 2 min and 22 °C for 2 min for 50 cycles and ended with 37 °C for 5 min, 22 °C for 2 min and 65 °C for 10 min. The resulting pEX1K3-Reductase and pEX1K3-Methyltransferase plasmid constructs were verified as described in section 2.6.8.

Table 2.5. Synthesised DNA fragments for construction of pEX1K3-Reductase and pEX1K3-Methyltransferase

DNA Fragment	Sequence (5' – 3')	Size
Putative Reductase	CAGGAAACAGCTATGACCATGGAATTCGCGGCCGCTTCTAGAGACTCTGTGGTCTCAAATGGCTGGGTCGCACCACCATCATCATCA CGGTAGCGTGCTTTCGGATCAAGCTCTCGAATCTCTCAGGCAGAAGCTGGAAGAGTGCCACGAGTGGCCGTGCCAGTATATGTTCAA GTTCATTGTACCGCATGGCCAGAGCCACCAGCTCTGCGCTGTGTTGGAGATGATGCCCAAAAGCGAGCGTGCTTCCAGCTCCGGCAA GTACGTGAGCCTCACCATTGAAGAGCATATGTCCTCGCCCGAGGAAGTGGTCATGGTGTATCAGAAGGCGTCCACCGTGCCGGGAGT GCTCGCGTTGTGACTGCTGAGACCAGTGGCTCCAGACGAAGTTACTAGTAGCGGCCGCTGCAGGACTGGCCGTCGTTTTACA	430 bp
Putative Methyltransferase	CAGGAAACAGCTATGACCATGGAATTCGCGGCCGCTTCTAGAGACTCTGTGGTCTCAAATGGCTGGGTCGCACCACCATCATCATCA CGGTAGCAATGACTATCCCAAACGCCTCAACCTCGGCAGTGGCAAGCTTTTCAAGGATTCTTACCTGAATGTGGATTACGCCCCCTTC TGGCGGCCGGACGTGGTGGCCGATCTCAATGAGCCCTTTCCACCCCACGAGGTGGCGCAGACGGCCCGCTTCGGCGAAATCCGCC TCCAACCTGGAACTTTTGAAGAAATCGTGGCCCGCGACGTGCTGGAGCACATTCCCAATCTGGTGACATGCATGAGTTCGTGCCTGT CTTTGCTGGCTCCTGGCGGCTGCATGCACATCATCGTCCCGTACGACCTCTCCCACGGCGCCTGGCAGGATCCCACCCACGTACGC GCATTCAATGAGCGCAGCTGGACCTATTATACGGAATGGTATTGGTACCTTGGCTGGCAGGAGGCGCGGTTCGAACTGACTCGGCTG GAGTTCGTGCACTCCGAGCTGGGCGAAAAACTCGCGGCTGAGGGCGTGGACAAGGATACCATCCTGCGCACGCCCCGCGCTGTGG ACGCCATGGAGGTAACGCTGGAAAAAGTTCTGCTTACCGACAAAGAGCGCCATTACGTGCAGAAATACTTGAAAAATCCGCTGACGC CGCCGGACGATTGTTTCCCAGGCGCAGCGACGCCCGACTGCGACTGACTGCTGAGACCAGTGGCTCCAGACGAAGTTACTAGTAGC GGCCGCTGCAGGACTGGCCGTCGTTTTACA	808 bp
T7 promoter	GCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTACAG GAAACAGCTATGACCATGGAATTCGCGGCCGCTTCTAGAGACTCTGTGGTCTCATACTTAATACGACTCACTATAGGGGAATTGTGAGC GGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGAATGTGAGACCACGAAGTTACTAGTAGCGG CCGCTGCAGGACTGGCCGTCGTTTTACACCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCC AGGAACCGTAAAAAGGCCGC	373 bp
T7 terminator	GCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTACAGG AAACAGCTATGACCATGGAATTCGCGGCCGCTTCTAGAGACTCTGTGGTCTCACTGCTAGCATAACCCCTTGGGGCCTCTAAACGGGTC TTGAGGGGTTTTTTGTTCGTGAGACCACGAAGTTACTAGTAGCGGCCGCTGCAGGACTGGCCGTCGTTTTACACCACAGAATCAGGGGA TAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGC	329 bp

Prefix and suffix sequences required for the modular cloning strategy used to construct pEX1K3-Reductase and pEX1K3-Methyltransferase are shown in blue and in red respectively.

2.7.2. Plasmid pEC-K-VtATPase, *E. coli* Expression Vector

The V-type ATPase operon coding sequence was extracted from *de novo* assembled *D. desulfuricans* NCIMB 8326 genome within this study. The 5'-ACCTAAAGAGGAGAAAA-3' *E. coli* RBS sequence was added upstream of each operon subunit coding sequence. The DNA sequence encoding a poly-histidine tag "Nterm-MAGSHHHHHGS-Cterm" was also added to the coding sequence of one subunit of the operon. The modified V-type ATPase operon DNA sequence was synthesised and cloned into the expression vector pJ431 by ATUM (Newark, USA), generating the plasmid pEC-K-VtATPase (Figure 2.4).

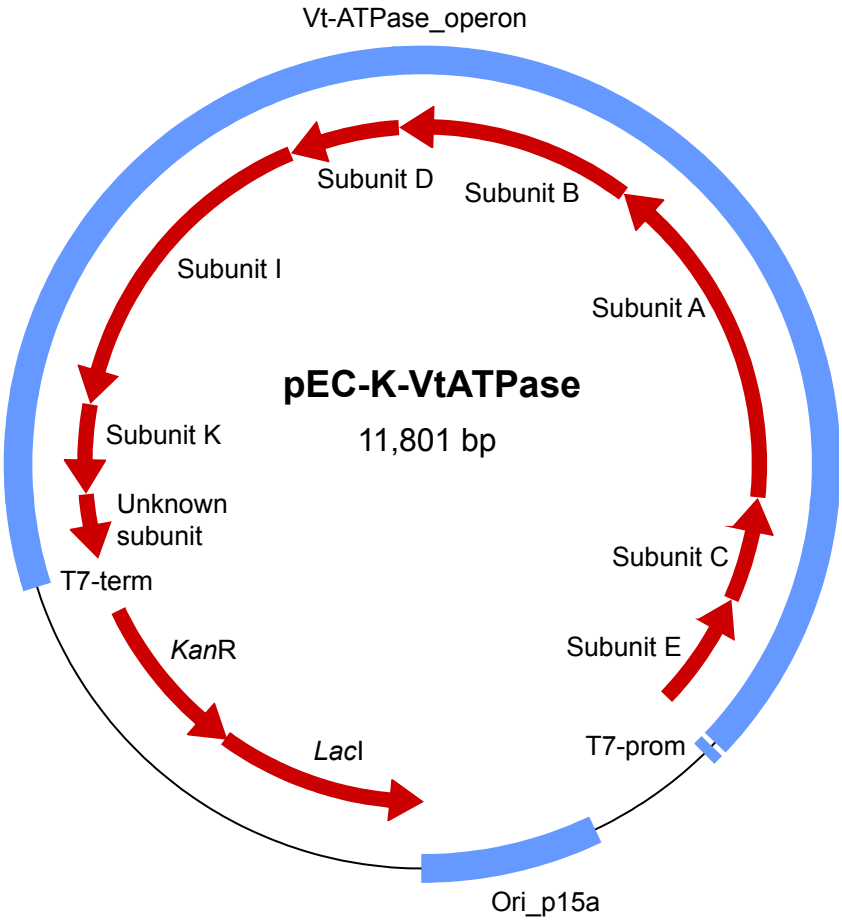


Figure 2.4. Plasmid map of pEC-K-VtATPase

The plasmid pEC-K-VtATPase synthesised by ATUM was used for expression of the V-type ATPase originating from *D. desulfuricans* NCIMB 8326 in *E. coli*. The expression of the histidine-tagged V-type ATPase was under control of the T7 promoter. (Plasmid map rendered in Clone Manager Professional edition)

2.7.3. Plasmid pPD3, Shuttle Vector between *E. coli* and *Desulfovibrio*

The *rfp* gene (BioBrick part BBa-J05540-RFP) was excised from the plasmid pSB1K3-RFP (Table 2.6) by PstI and EcoRI restriction enzyme digestion. The EcoRI/PstI cut fragment containing the *E. coli* replicon pMB1 and a kanamycin resistance gene (BioBrick part BBa_P1003) from pSB1K3-RFP was then ligated with a 55 bp multiple cloning site sequence, previously cut with the same restriction endonucleases, generating the plasmid pPD1. The multiple cloning site sequence was generated by annealing 50 μM of MCS-FF and 50 μM MCS-R, two single strand nucleic acid oligomers (Table 2.7). Annealing conditions started with 2 min at 95 °C, followed by a temperature gradient from 90 °C to 35 °C decreasing 5 °C every 15 s.

The oriT DNA region from the plasmid pS797 (Table 2.6) was amplified by PCR using the primers pS797-oriT_F and pS797-oriT_R (Table 2.7). The PCR mixture was 1-fold Q5 buffer, 200 μM deoxynucleoside triphosphates, 0.5 μM pS797-oriT_F, 0.5 μM pS797-oriT_R, 1 U Q5 DNA polymerase and 100 ng pS797 plasmid in a final volume of 50 μl nuclease–free water (Thermo-Fisher Scientific Inc.). PCR reaction conditions started with 30 s DNA denaturation at 98 °C, followed by 30 cycles consisting of 10 s at 98 °C, 30 s at 58 °C and 30 s at 72 °C, ending with a 5 min extension at 72 °C. Amplified oriT DNA region was purified from the PCR mixture using a Wizard® SV Gel and PCR Clean-Up System kit (Promega, Madison, USA), following the manufacturer's instructions. The pS797-oriT_F and pS797-oriT_R primers were designed with overhang sequences containing a XbaI restriction site. The plasmid pPD1 and the amplified oriT DNA region were digested by XbaI, prior to ligation generating the plasmid pPD2.

The pBG1 region from the plasmid pMO719 (Table 2.7) was excised by EcoRI digestion and ligated into pPD2, previously cut by the same restriction endonuclease. The resulting pPD3 plasmid construct (Figure 2.5) was verified as described in section 2.6.8.

Table 2.6. Plasmids used for construction of the shuttle vector pPD3

Plasmid	Relevant characteristics	Source
pSB1K3-RFP	BioBrick cloning vector pSB1K3 containing the BioBrick part BBa-J05540-RFP encoding a red fluorescent protein and *E. coli* replicon pMB1; Kan^R	Exeter Microbial Biofuels Group (Exeter, UK)
pPD1	pSB1K3-RFP excised of BBa-J05540-RFP and containing a multiple cloning site, Kan^R	This study
pS797	*Geobacillus* shuttle vector containing an oriT region; Amp^R, Kan^R	Exeter Microbial Biofuels Group (Exeter, UK)
pPD2	pPD1 containing an oriT region; Kan^R	This study
pMO719	*Desulfovibrio* vector containing *D. desulfuricans* G100A cryptic plasmid pBG1; $Spec^R$	Judy Wall Laboratory (Missouri, USA)

Table 2.7. Primers used for construction of the shuttle vector pPD3

Primers	Overhang sequence (5' - 3')	Short primer sequence (5' - 3')	Use
MCS-F		CCTGCAGCGGCCGCTACTAGTACGAA AGAGACCGGTCTCTAGTACTCTAGAAG CGGCCGCGAATTCCATC	Generation of a multiple cloning site DNA fragment
MCS-R		GATGGAATTCGCGGCCGCTTCTAGAGT ACTAGAGACCGGTCTCTTTCGTACTAGT AGCGGCCGCTGCAGG	Generation of a multiple cloning site DNA fragment
pS797-oriT_F	AGTCTAGACTGGATCCTCGGTA CCCGGTCT	TGCCTTGCTCGTCGGTGATG	Amplification of the oriT region sequence from pS797
pS797-oriT_R	CATCTAGAGTGAGCTCTTGTCGA CTTCCCA	ATTCCACATTGCAATAATAG	Amplification of the oriT region sequence from pS797

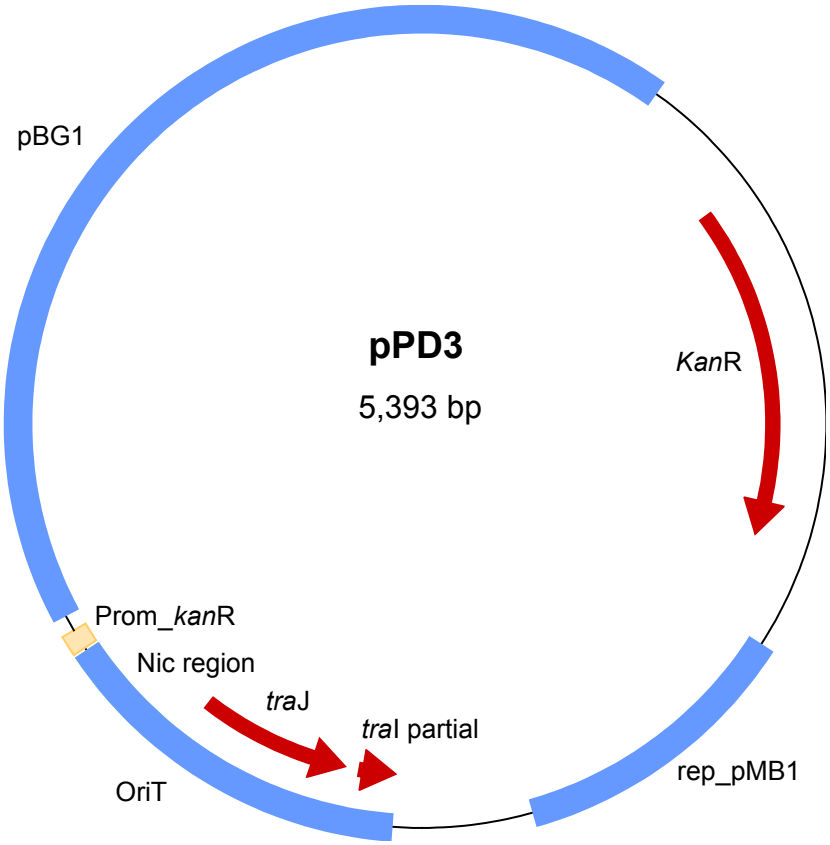


Figure 2.5. Plasmid map of pPD3

The plasmid pPD3 was designed and constructed for gene transfer by transformation and conjugation into *Desulfovibrio*. The plasmid pPD3 notably contains an origin of replication for *E. coli* (rep_pMB1), an origin of replication for *Desulfovibrio* (pBG1) and an oriT region required for plasmid conjugal transfer. (Plasmid map rendered in Clone Manager Professional edition)

2.7.4. Plasmids pPD3-Reductase and pPD3-Methyltransferase, *Desulfovibrio* Expression Vectors

The histidine-tagged putative reductase and methyltransferase coding sequences (Table 2.5) were separately cloned downstream of a kanamycin resistance gene promoter and RBS sequences, and upstream of a kanamycin resistance gene terminator sequence into the plasmid pEX1K3-ccdB, using the modular cloning strategy as described in section 2.7.1. The kanamycin resistance gene promoter, RBS and terminator sequences were extracted from the plasmid pPD3 and were synthesised flanked with prefix and suffix sequences, by Twist Biosciences (Table 2.8).

The histidine-tagged putative reductase and methyltransferase DNA fragments, both flanked by the kanamycin resistance gene promoter, RBS and terminator sequences as results of the modular cloning strategy, were excised from pEX1K3 by ScaI and PstI restriction enzyme digestion. The ScaI/PstI cut DNA fragments were then ligated into pPD3, previously cut by the same restriction endonucleases. The resulting pPD3-Reductase and pPD3-Methyltransferase plasmid constructs (Figure 2.6 A and B) were verified as described in section 2.6.8.

Table 2.8. Synthesised DNA fragments for construction of pPD3-Reductase and pPD3-Methyltransferase

DNA Fragment	Sequence (5' – 3')	Size
Kanamycin promoter-RBS	CAGGAAACAGCTATGACCATGGAATTCGGATCCTGACTCTGTGGTCTCATACTAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGATGATATATTTTTATCTTGTGCAATGTAACATCAGAGATTTTGAGACACAACGTGGCTTTGTTGAATAAATCGAACTTTTGCTGAGTTGAAGGATCAGAATGTGAGACCACGAAGTTATCTAGACTGCAGGACTGGCCGTCGTTTTACA	252 bp
Kanamycin terminator	CAGGAAACAGCTATGACCATGGAATTCGCGGCCGCTCCTAAGGGACTCTGTGGTCTCACTGCTGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAATTCGTGAGACCACGAAGTTACTAGTAGCGGCCGCTGCAGGACTGGCCGTCGTTTTACA	165 bp

Prefix and suffix sequences required for the modular cloning strategy used to construct pPD3-Reductase and pPD3-Methyltransferase are shown in blue and in red respectively.

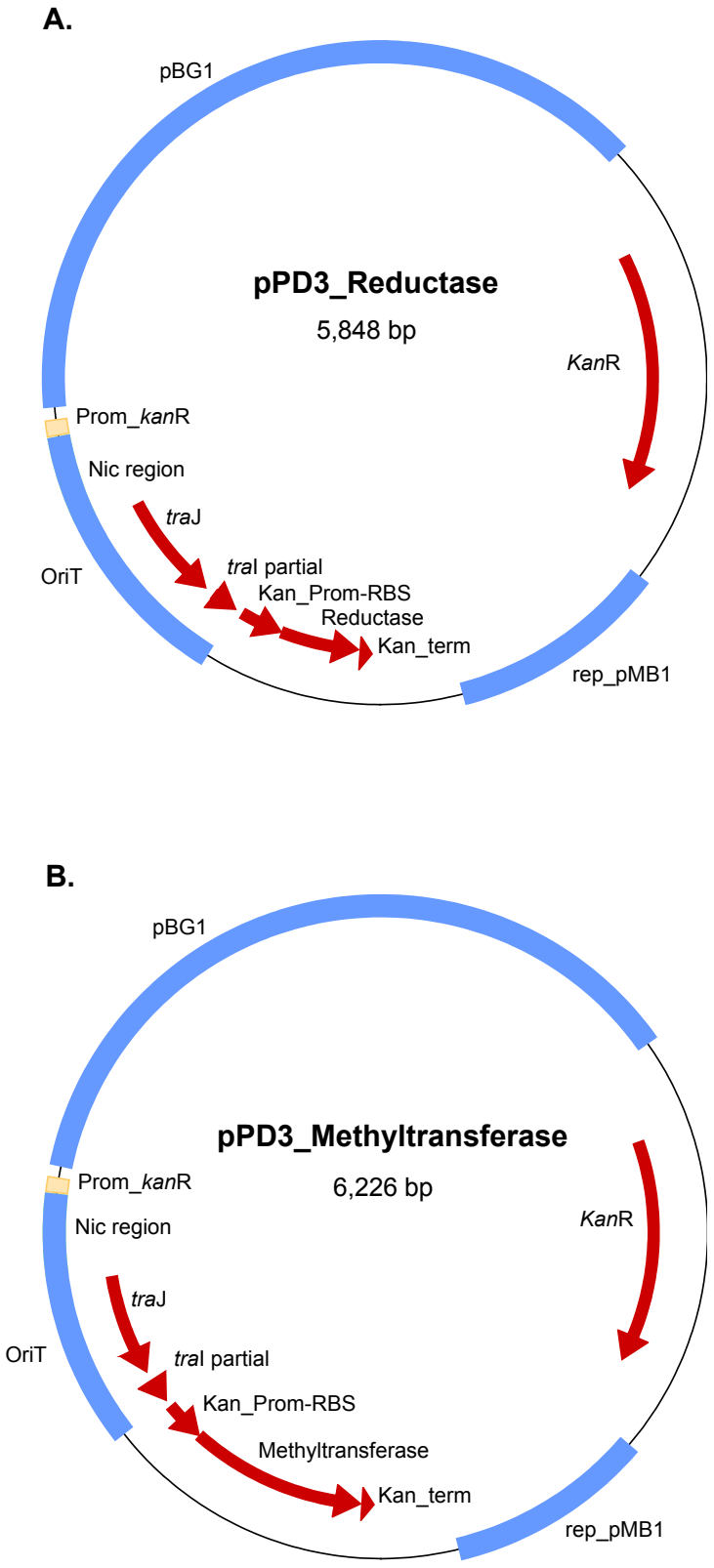


Figure 2.6. Plasmid maps of pPD3-Reductase and pPD3-Methyltransferase

The plasmids pPD3-Reductase (A) and pPD3-Methyltransferase (B) were used for expression of the putative reductase and methyltransferase originating from *D. desulfuricans* NCIMB 8326 in *D. vulgaris* strain Hildenborough. (Plasmid maps rendered in Clone Manager Professional edition)

2.7.5. Plasmid pPD3-VtATPase, *Desulfovibrio* Expression Vector

The pPD3-VtATPase expression vector was designed to carry the histidine-tagged V-type ATPase operon sequence from the plasmid pEC-K-VtATPase, which was cloned into the vector pPD3, downstream of the kanamycin resistance gene promoter and RBS fragment (Table 2.8) and upstream of the kanamycin resistance gene terminator fragment (Table 2.8).

The histidine-tagged V-type ATPase operon sequence was excised from pEC-K-VtATPase by XbaI and Bsu36I restriction enzyme digestion. The kanamycin resistance gene promoter and RBS fragment was cut with BamHI and XbaI restriction enzymes. The kanamycin resistance gene terminator fragment was cut with Bsu36I and PstI restriction enzymes. The plasmid pPD3 was cut with BamHI and PstI restriction enzymes. The ligation of the four digested DNA fragments was performed with different fragment molar ratios (V-type ATPase operon/Kanamycin promoter-RBS/Kanamycin terminator/pPD3; 1/3/3/1; 1/4/4/1; 1/5/5/1).

2.8. Functional Verification of Candidate Genes for Alkane Synthesis in Desulfovibrio spp.

2.8.1. Heterologous Expression of Candidate Genes in *D. vulgaris* Hildenborough

Liquid cultures of *D. vulgaris* transformants (containing the plasmid DNA of interest) were prepared by inoculating a single colony into 5 ml sodium lactate medium supplemented with 400 μg ml^{-1} geneticin G418. Cultures were anaerobically incubated for 3 days at 37 °C. Post-incubation, the entire 5 ml cultures were used to inoculate 45 ml sodium lactate medium supplemented with 400 μg ml^{-1} geneticin G418 and 1 % (v/v) 10 mM tetradecyl-d_{29}-phosphate solution. The tetradecyl-d_{29}-phosphate (provided by David Leys laboratory, University of Manchester) was previously dissolved in anhydrous ethanol to a final concentration of 10 mM. After 10 days incubation at 37 °C under anaerobic atmosphere, 50 ml liquid cultures were then analysed for alkane production by GC-MS as described in section 2.3.1 and for protein content as described in section 2.9.

2.8.2. Enrichment of *D. desulfuricans* NCIMB 8326 Growth Medium with Deuterated Alkyl Phosphates

Liquid cultures of wild-type *D. desulfuricans* NCIMB 8326 were prepared by inoculating a single colony into 5 ml sodium lactate medium. Cultures were anaerobically incubated for 5 days at 37 °C. Post-incubation, the entire 5 ml cultures were used to inoculate 45 ml sodium lactate medium supplemented with 1 % (v/v) of either octadecyl-1,1-d_2-phosphate solution, hexadecyl-d_{33}-phosphate solution or tetradecyl-d_{29}-phosphate solution. The octadecyl-1,1-d_2-phosphate and hexadecyl-d_{33}-phosphate (provided by David Leys laboratory, University of Manchester) were previously dissolved in anhydrous ethanol to a final concentration of 1 mM. After 10 days incubation at 37 °C under anaerobic atmosphere, 50 ml liquid cultures were then analysed for alkane production by GC-MS as described in section 2.3.1. Total protein amount from the cultures was quantified as described in section 2.9.2.

2.8.3. *Desulfovibrio* Alkane Biosynthesis Sensitivity Assay to the Dark

Liquid cultures of wild-type *D. desulfuricans* NCIMB 8326 were prepared by inoculating a single colony into 5 ml sodium lactate medium. Cultures were anaerobically incubated for 5 days at 37 °C either in the diurnal daylight or in the dark (cultures entirely wrapped with aluminium foil). Post-incubation, the entire 5 ml cultures were used to inoculate 45 ml sodium lactate medium. After 10 days incubation at 37 °C under anaerobic atmosphere either in the diurnal daylight or in the dark (cultures entirely wrapped with aluminium foil), 50 ml liquid cultures were analysed for alkane production by GC-MS as described in section 2.3.1. Total protein amount from the cultures was quantified as described in section 2.9.2.

2.9. Functional Verification of Candidate Genes for Alkane Synthesis in E. coli

2.9.1. *In vivo* Protein Function Assays

Liquid cultures of BL21 Star™ (DE3) *E. coli* transformants (containing the plasmid DNA of interest) were prepared by inoculating a single colony into 5 ml LB medium

supplemented with 50 μg ml^{-1} kanamycin. Cultures were incubated aerobically or anaerobically at 37 °C overnight with 220 rpm agitation. 500 μl of *E. coli* overnight cultures were used to inoculate 50 ml LB medium supplemented with 50 μg ml^{-1} kanamycin and 1 % (v/v) 10 mM tetradecyl-d_{29}-phosphate solution. Cultures were incubated aerobically or anaerobically at 37 °C with 220 rpm agitation until achieving an optical density value of 0.7 at 600 nm for aerobic cultures and 0.2 at 600 nm for anaerobic cultures. Cultures were then supplemented with 200 μM isopropyl β-D-1-thiogalactopyranoside (IPTG), for induction of gene expression controlled by the T7 promoter. After induction, cultures were incubated aerobically or anaerobically for 4h at 37 °C with 220 rpm agitation. Cultures were then analysed for alkane production by GC-MS as described in section 2.3.1 and for protein content as described in section 2.9.

2.9.2. *In vitro* Protein Function Assays

E. coli Culture Conditions

Liquid cultures of BL21 Star™ (DE3) *E. coli* transformants (containing the plasmid DNA of interest) were prepared by inoculating a single colony into 5 ml LB medium supplemented with 50 μg ml^{-1} kanamycin. Cultures were incubated aerobically at 37 °C overnight with 220 rpm agitation. 1 ml of *E. coli* overnight cultures were used to inoculate 100 ml LB medium supplemented with 50 μg ml^{-1} kanamycin. Cultures were incubated aerobically at 37 °C with 220 rpm agitation until achieving an optical density value of 0.7 at 600 nm. Cultures were then supplemented with 1 mM IPTG for induction of gene expression controlled by the T7 promoter. After induction, cultures were incubated aerobically at 37 °C for 4 h with 220 rpm agitation.

Cell Lysis

Bacterial cells were harvested by centrifugation at 4,700 x *g* for 30 min at 4 °C and supernatants were discarded. Pellets were resuspended into 5 ml chilled, sterile PBS. Bacterial cells were lysed with 10 bursts of 10 μm amplitude and 30 s using a Soniprep 150 Ultrasonic Disintegrator (MSE Ltd., London, UK). Bursts were alternated with 90 s rest on ice, to prevent overheating of the samples. After lysis, 500 μl aliquots of lysed bacterial cells were sampled and analysed for alkane production by GC-MS and for protein content, as described in section 2.9. For alkane production analysis, samples were snap-frozen in liquid nitrogen, freeze-dried overnight and resuspended into 200 μl

DCM spiked with 10 *µ*M β-cyclocitral, prior to being screened for alkane production by GC-MS as described in section 2.3.1.

Function Protein Assay

Tetradecyl-d_{29}-phosphate dissolved in anhydrous ethanol was added to the lysed bacterial cells to a final concentration of 100 *µ*M. Lysed bacterial cells supplemented with 100 *µ*M tetradecyl-d_{29}-phosphate were incubated for 48h at 37 °C. 500 *µ*l aliquots of lysed bacterial cells supplemented with 100 *µ*M tetradecyl-d_{29}-phosphate were sampled at the beginning of the incubation, after 24 h and 48 h incubation. Samples were snap-frozen in liquid nitrogen, freeze-dried overnight and resuspended into 200 *µ*l DCM spiked with 10 *µ*M β-cyclocitral, prior to being screened for alkane production by GC-MS as described in section 2.3.1.

2.10. Protein Analysis

2.10.1. Cell Lysis

Barring *in vitro* protein function assays where bacterial cells were lysed by sonication, bacterial cells were chemically lysed using BugBuster Protein Extraction Reagent (Merck Millipore).

E. coli cells were harvested by centrifugation for 30 min at 4,700 x *g* and at 4 °C. *Desulfovibrio* cells were harvested for 30 min at 16,000 x *g* at 4 °C. Supernatants were discarded and pellets were resuspended into BugBuster Protein Extraction Reagent, according to the manufacturer's instructions. Resuspended pellets were incubated for 20 min at room temperature with 100 rpm agitation, prior to being centrifuged for 30 min at 16,000 x *g* and at 4 °C.

2.10.2. Protein Quantification

Protein quantifications were performed using the Qubit® 2.0 Fluorometer with a Qubit® Protein Assay Kit (Thermo-Fisher Scientific Inc.). A Qubit working solution was prepared by diluting the fluorescent protein dye 1:200 in the Qubit® protein buffer. The Qubit working solution was added to 3 *µ*l of protein samples and to 10 *µ*l of provided

standards with known concentration, to a final volume of 200 µl. Samples and standards were vortexed and incubated for 15 min at room temperature, prior to being analysed using Qubit® 2.0 Fluorometer. Protein standards with known concentrations were analysed before the study samples, allowing accurate protein quantification.

2.10.3. Sodium Dodecyl Sulfate - PolyAcrilamide Gel Electrophoresis (SDS-PAGE)

Protein samples and a purified histidine-tagged esterase (positive control for SDS-PAGE and Western Blot) were diluted in Laemmli loading buffer (125 mM Tris-HCl (pH 6.8), 10 mM EDTA (pH 6.8), 20 % (w/v) glycerol, 0.1 % (w/v) SDS, 10 % (v/v) β-mercaptoethanol, 0.05 % (w/v) bromophenol blue; Laemmli, 1970) for protein denaturation, prior to being boiled at 100 °C for 10 min and then cooled on ice for 1 min. Denaturated protein samples and a SeeBlue™ Plus2 Pre-stained Protein Standard (ThermoFisher scientific Inc.) were loaded onto a 4-20 % gradient Bis-Tris ExpressPlus™ polyacrylamide gels (GeneScript, Piscataway, USA). Electrophoresis was performed at 80 V, 400 mA for 10 min and then at 130 V, 400 mA for 50 min, in a MOPS running buffer (50 mM Tris-Base, 50 mM MOPS, 1 mM EDTA and 0.1 % (w/v) SDS; pH 7.7), using a mini-PROTEAN® Tetra cell (Bio-Rad).

Protein bands were stained by soaking SDS-PAGE gels into a Coomassie blue G solution (40 % (v/v) anhydrous ethanol, 1.74 M glacial acetic acid, 0.1 % (w/v) Coomassie blue G) for 30 min on a shaking platform. SDS-PAGE gels were then washed several times into deionised water to remove the unbound stain.

2.10.4. Western Blot

The Western blot procedure was performed from SDS-PAGE gels, with unstained protein bands. Unstained protein bands from SDS-PAGE gels were transferred to a 0.45 µm PVDF membrane (Invitrolon™ PVDF/Filter Paper Sandwich, ThermoFisher scientific Inc.), using a Pierce™ G2 Fast Blotter (ThermoFisher scientific Inc.). Before blotting, SDS-PAGE gels were soaked in Pierce™ 1-Step Transfer Buffer (ThermoFisher scientific Inc.) and were placed on the top of PVDF membranes, previously soaked in anhydrous methanol. SDS-PAGE gels and PVDF membranes were placed inside a Pierce™ G2 Fast Blotter between four Pierce Western Blotting Filter

Papers (ThermoFisher scientific Inc.), also previously soaked in Pierce™ 1-Step Transfer Buffer. The blotting was performed with a current of 25 V and 1.3 A for 15 min.

After the blotting, PVDF membranes were probed with an anti-Histidine tag primary antibody raised in mouse, THE™ Anti-His mAb (GeneScript), and with an anti-mouse secondary antibody raised in goat and conjugated to the IRDye® 680RD fluorophore (LI-COR, Lincoln, USA), using an iBind Western Device (ThermoFisher scientific Inc.). For antibody probing, PVDF membranes were soaked in 5 ml of iBind mix, prepared by diluting 6 ml iBind™ 5-fold buffer and 300 *µ*l iBind™ 100-fold additives (from an iBind™ Solution Kit, ThermoFisher scientific Inc.) into 23.7 ml deionised water. The PVDF membranes were then placed on an iBind™ card (ThermoFisher scientific Inc.) pre-equilibrated with 6 ml of iBind mix, in an iBind Western Device. An iBind Western Device automatically performs membrane blocking, washes and antibody incubations by sequential lateral flow capillary diffusion of the following solutions:

1 - 2 ml primary antibody solution (1 *u*g THE™ Anti-His mAb in 2 ml iBind mix)
2 - 2 ml iBind mix
3 - 2 ml secondary antibody solution (2 *u*g IRDye® 680RD Goat-anti-Mouse Antibody in 2 ml iBind mix)
4 - 6 ml iBind mix

After the above solutions were diffused across the surface of the PVDF membranes, the IRDye 680RD fluorophore was excited at 676 nm and fluorescence was detected at 700 nm using an Odyssey CLx Imaging System (LI-COR) for visualisation of histidine-tagged protein bands.

2.10.5. Online Tools used for Protein Expression Analysis

The molecular weight of the proteins of interest was determined by translating their DNA sequence into protein sequence using the Translate tool from ExPASy Bioinformatics Resource Portal (http://www.expasy.org). The protein sequences were then used as query for the ProtParam tool from ExPASy Bioinformatics Resource Portal. The ProtParam tool computes various physical and chemical protein parameters, including the molecular weight, from protein sequences.

The *E. coli* Codon Usage Analyzer 2.1 online tool was used to determine the frequency with which each codon of a DNA coding sequence of interest is used by *E. coli*. (http://www.faculty.ucr.edu/~mmaduro/codonusage/usage.htm)

3. *Desulfovibrio* Metabolism Screening for Biogenic Hydrocarbons and Fatty Acids

3.1. Introduction and Abstract

Sulphate-reducing bacteria from sediments were first reported to produce up to 1 % of dry mass as hydrocarbons (Jankowski & ZoBell, 1944; Oppenheimer, 1965). Further investigations into hydrocarbon production by sulphate-reducing bacteria led to the identification of the bacterium *D. desulfuricans*. *D. desulfuricans subsp. desulfuricans* California27.137.5 (NCIMB 8326) was reported to produce odd and even C_{11}-C_{35} carbon chain paraffinic hydrocarbons, with a predominance of C_{25}-C_{35} alkanes, up to 2.25 % of dry biomass (Davis, 1968). Another strain, *D. desulfuricans subsp. desulfuricans* Essex6 (NCIMB 8307) was also reported to synthesise intracellular hydrocarbons, mainly composed of long-chain C_{25}–C_{35} alkanes, and extracellular hydrocarbons, aliphatic normal and *iso*-forms C_{14}–C_{25} alkanes, up to 30 mg l^{-1} (Bagaeva & Chernova, 1994; Bagaeva, 2000).

Hydrocarbon synthesis within the genus *Desulfovibrio* was verified in this study by cultivating *Desulfovibrio* spp. in isotopically labelled growth medium. Metabolites synthesised by an organism growing in isotopically labelled medium will be labelled by integration of stable isotopes in their structure, while compounds of non-biogenic origin are not. Thus cultivating *Desulfovibrio* in isotopically labelled growth medium enabled the differentiation of biogenic hydrocarbons from non-metabolically derived hydrocarbons. These isotope labelling experiments therefore allowed the verification of hydrocarbon biosynthesis by *Desulfovibrio* spp. and the establishment of a definitive profile of biogenic alkanes.

In this study, hydrocarbon biosynthesis by *D. desulfuricans* was initially verified within three strains. *D. desulfuricans* NCIMB 8326 and *D. desulfuricans* NCIMB 8338 were proven to be capable of hydrocarbon biosynthesis. However, hydrocarbon biosynthesis by *D. desulfuricans* NCIMB 8307 was not confirmed in this study. Hydrocarbon biosynthesis screening was then extended to seven additional *Desulfovibrio* species. *D. gabonensis* DSM 10636, *D. gigas* NCIMB 9332, *D. marinus* DSM 18311 and *D. paquesii* DSM 16681 were shown to be capable of hydrocarbon

biosynthesis. However, hydrocarbon biosynthesis by *D. vulgaris* strain Hildenborough, *D. giganteus* DSM 4370 and *D. alcoholivorans* NCIMB 12906 was not detected.

Moreover, GC-MS spectra of the hydrocarbon producing *Desulfovibrio* strains revealed that *Desulfovibrio* biogenic hydrocarbons are composed of octadecane (C_{18}), nonadecane (C_{19}) and eicosane (C_{20}), with higher quantity of even numbered carbon chain alkanes. To date, all characterised alkane biosynthetic pathways derive from the fatty acid metabolic pathway (*cf*. Chapter 1 – Introduction; Figure 1.1). It was therefore postulated that either *Desulfovibrio* had an unusual fatty acid metabolism producing odd-chain fatty acids which were subsequently decarboxylated to even-chain, terminal alkanes; or fatty acid metabolism in *Desulfovibrio* was typical of all other organisms producing even-chain fatty acids but, with a significant modification hence, alkane production did not occur via decarbonylation or decarboxylation.

The biogenic fatty acid content of *Desulfovibrio* spp. involved in this study was analysed. The fatty acid analysis indicated that *Desulfovibrio* spp. do not produce either saturated carbon chain nonadecanoic (C_{19}) acid or saturated carbon chain heneicosanoic acid (C_{21}). With the assumption that fatty acids are the metabolite precursors of biogenic alkanes in *Desulfovibrio*, the biosynthesis of octadecanes (C_{18}) and eicosanes (C_{20}) was unlikely to involve a decarbonylation or decarboxylation step as identified in all previously characterised alkane biosynthetic pathways. This deduction engendered a novel hypothesis that alkane production by *Desulfovibrio* involves a series of reduction reactions from fatty acids.

3.2. Desulfovibrio Metabolism Screening for Biogenic Hydrocarbons

3.2.1. Alkane Standard Mass Spectra and Retention Times

Standard solutions of C_8 - C_{20} and C_{21} - C_{40} alkanes were analysed by CG-MS for mass spectrum fragmentation pattern and retention time determination. The mass spectrum of the alkane standards ionised by electron impact ionisation showed a fragmentation pattern with ion peak clusters, evenly separated by 14 mass units (Figure 3.1A). Each ion peak corresponds to an alkyl radical, which differ by the loss of a CH_2 group (corresponding to 14 mass units). The alkane standard fragmentation pattern was composed in abundance of butyl ($C_4H_9^+$), pentyl ($C_5H_{11}^+$) and hexyl ($C_6H_{13}^+$), with a mass to charge ratio (m/z) of 57.07, 71.09 and 85.10 respectively. The mass spectrum of organic compounds extracted from *Desulfovibrio* were screened for the same fragmentation pattern to the alkane standard mass spectrum, in order to detect alkanes within organic compound extracts.

Alkanes detected within organic compound extracts were identified by their retention time. The retention time of a compound is the time required for the compound to migrate through the gas chromatography column from its injection into the column to its elution. The longer the carbon chain of an alkane, the more time the alkane interacts with the stationary phase and therefore the longer the alkane migration through the column (Figure 3.1B). Alkane standard solutions and samples were spiked with β-cyclocitral, as an internal standard.

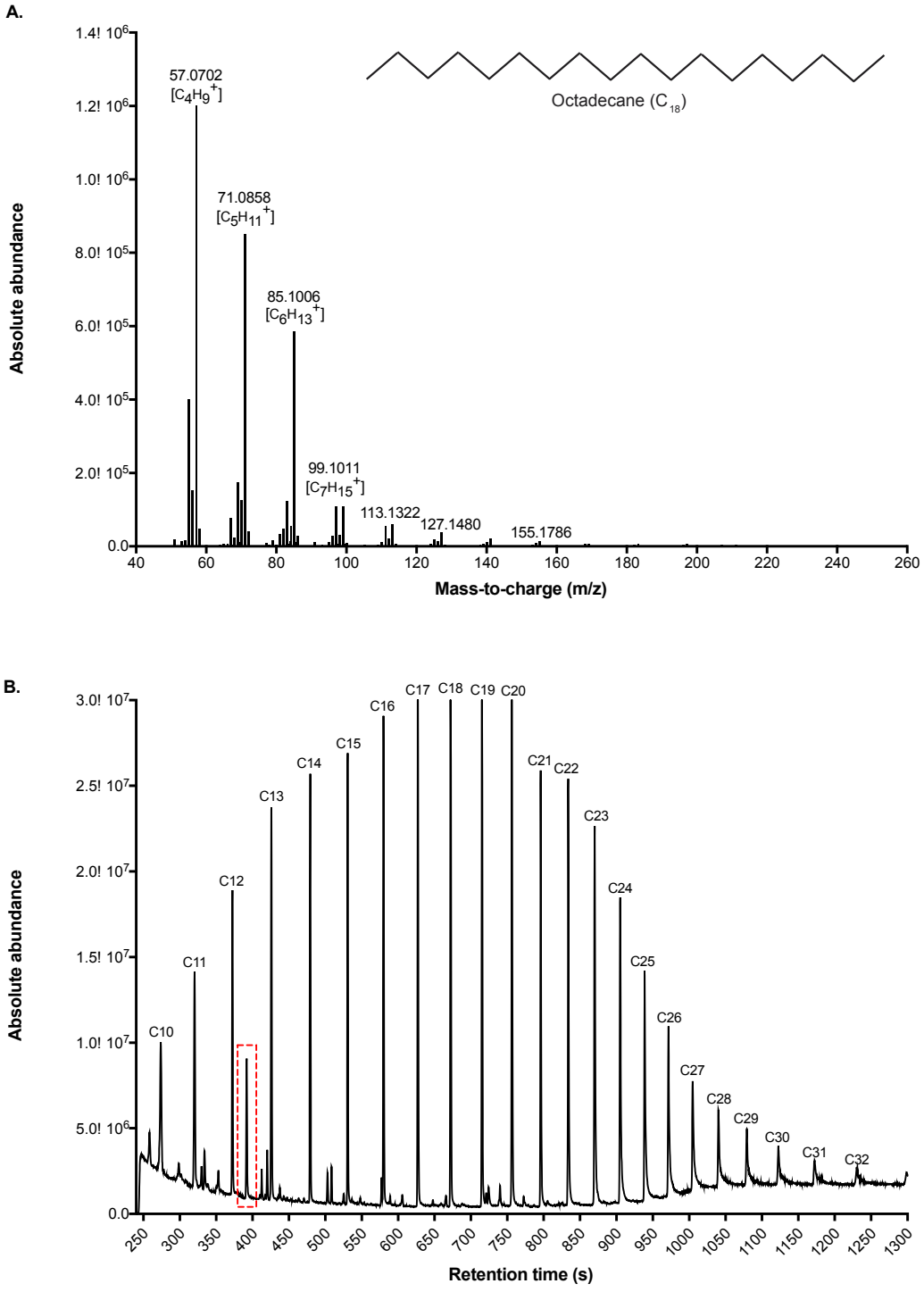


Figure 3.1. Mass spectrum of the octadecane (C_{18}) standard and total ion chromatogram of C_{10}-C_{32} alkane standards

Standard solutions of alkanes spiked with β-cyclocitral were analysed by CG-MS for alkane mass spectrum fragmentation pattern and retention time determination. The mass spectrum of the octadecane (C_{18}) standard, with a retention time of 673.3 s, showed the fragmentation pattern of alkanes ionised by electron impact ionisation (A). The total ion chromatogram of C_{10}-C_{32} alkane standards allowed the determination of the retention time of each alkane (B). The peak corresponding to the β-cyclocitral (internal standard) is enclosed by a red dashed box.

3.2.2. Stable Isotope Labelling Experiments for Alkane Biosynthesis Screening and Biogenic Alkane Profile Determination in *Desulfovibrio*

Ten *Desulfovibrio* strains were cultured for 10 days in unlabelled sodium lactate medium and in deuterated sodium lactate medium, prior to alkane biosynthesis screening. Cultivating *Desulfovibrio* in deuterated sodium lactate medium allowed incorporation of deuterium in nascent biogenic alkanes. Deuterium is one of the stable isotopes of hydrogen. Stable isotopes possess the same number of protons, so have the same physical and chemical properties, but they differ in mass due to a different number of neutrons. Deuterium possesses one neutron making it one mass unit heavier than hydrogen which does not have neutron. Moreover, stable isotopes are naturally present in organic compounds (Urey, 1948). Therefore labelling metabolites with heavy stable isotopes changes metabolite mass and isotopic ratio (Lehmann, 2017), and renders them detectable.

The most abundant alkyl fragment generated from unlabelled alkane ionisation is the butyl ($C_4H_9^+$), with a molecular mass of 57 g mol^{-1}. As deuterium is naturally present, deuterated butyl $C_4H_8D_1^+$ is observed in an unlabelled alkane mass spectrum, at the m/z 58. Incorporation of deuterium in nascent alkanes synthesised by *Desulfovibrio* cultivated in deuterated medium increases the abundance of the deuterated butyl fragment in the mass spectrum. Therefore, the alkanes detected by GC-MS were considered biogenic when the ratio of the deuterated butyl m/z 58 ion abundance over the butyl m/z 57 ion abundance of alkanes detected within the organic compound extracts of *Desulfovibrio* grown in deuterated medium was significantly higher than the same isotopic ratio of alkanes detected within the organic compound extracts of *Desulfovibrio* grown in unlabelled medium.

After 10 days incubation, organic compounds from *Desulfovibrio* were extracted into DCM, an organic solvent in which hydrocarbons are soluble. After GC-MS analysis of the organic compounds extracted, the 57 m/z ion and the 58 m/z ion chromatograms were extracted from the total ion chromatogram of the organic compound extracts and were integrated. Peak areas of 57 m/z ion and 58 m/z ion fragments of each detected alkane were determined and the isotopic ratios of 58 m/z ion to 57 m/z ion peak areas were calculated. The significance of the observed difference between the ratios of 58 m/z ion to 57 m/z ion peak areas of alkanes detected within the organic compound extracts of *Desulfovibrio* grown in deuterated medium and of *Desulfovibrio* grown in

unlabelled medium was statistically evaluated by t-test, with the assumption that peak area ratio values followed a normal distribution.

Screening of D. desulfuricans strains for alkane production

Alkane biosynthesis within the genus *Desulfovibrio* was reported in *D. desulfuricans* (Davis 1968; Bagaeva & Chernova, 1994). Therefore, three candidate *D. desulfuricans* strains: *D. desulfuricans* NCIMB 8326 (Figure 3.2), *D. desulfuricans* NCIMB 8338 (Figure 3.3) *and D. desulfuricans* NCIMB 8307 (Figure 3.4) were initially screened for alkane production.

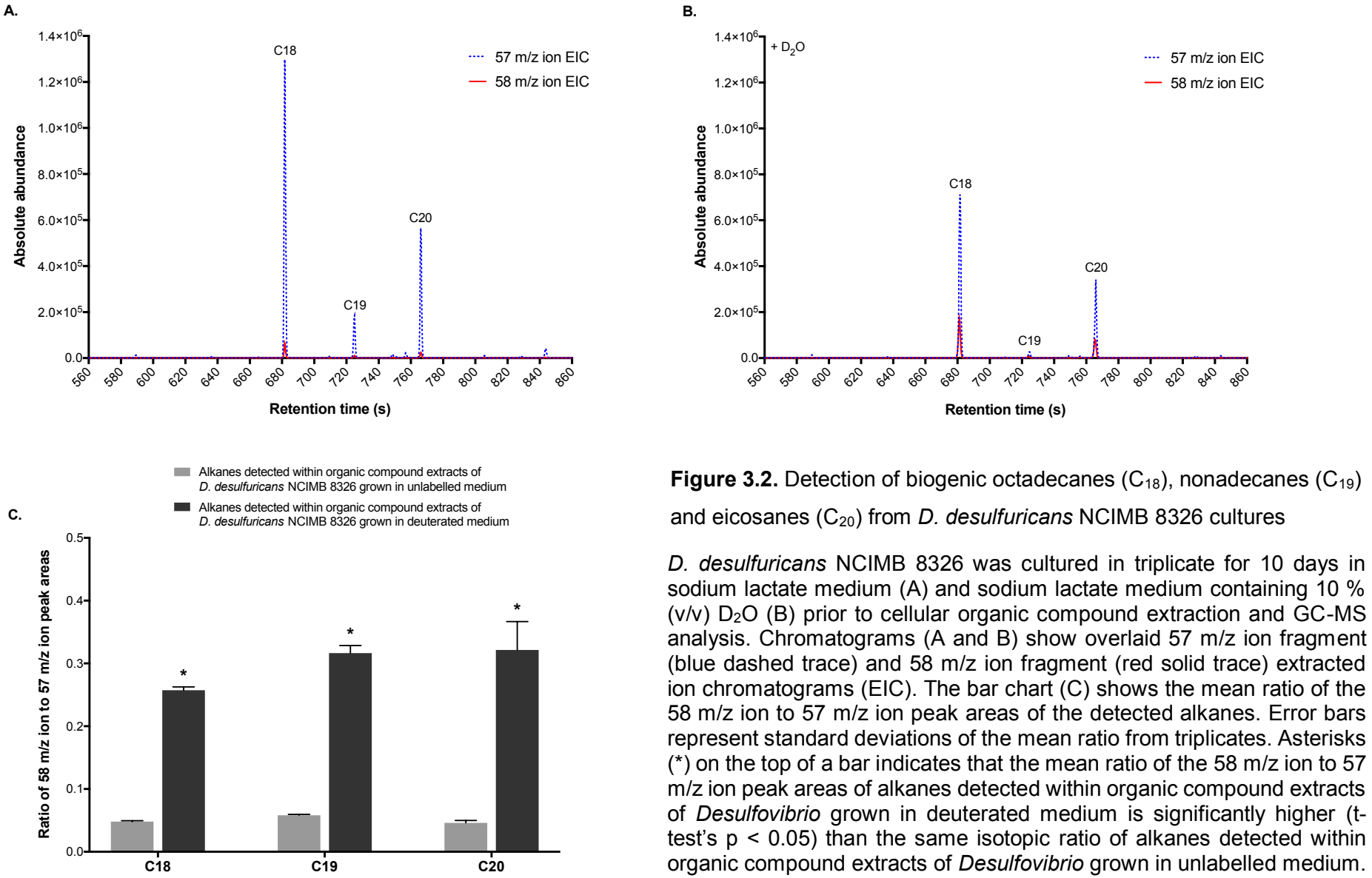


Figure 3.2. Detection of biogenic octadecanes (C_{18}), nonadecanes (C_{19}) and eicosanes (C_{20}) from *D. desulfuricans* NCIMB 8326 cultures

D. desulfuricans NCIMB 8326 was cultured in triplicate for 10 days in sodium lactate medium (A) and sodium lactate medium containing 10 % (v/v) D_2O (B) prior to cellular organic compound extraction and GC-MS analysis. Chromatograms (A and B) show overlaid 57 m/z ion fragment (blue dashed trace) and 58 m/z ion fragment (red solid trace) extracted ion chromatograms (EIC). The bar chart (C) shows the mean ratio of the 58 m/z ion to 57 m/z ion peak areas of the detected alkanes. Error bars represent standard deviations of the mean ratio from triplicates. Asterisks (*) on the top of a bar indicates that the mean ratio of the 58 m/z ion to 57 m/z ion peak areas of alkanes detected within organic compound extracts of *Desulfovibrio* grown in deuterated medium is significantly higher (t-test's $p < 0.05$) than the same isotopic ratio of alkanes detected within organic compound extracts of *Desulfovibrio* grown in unlabelled medium.

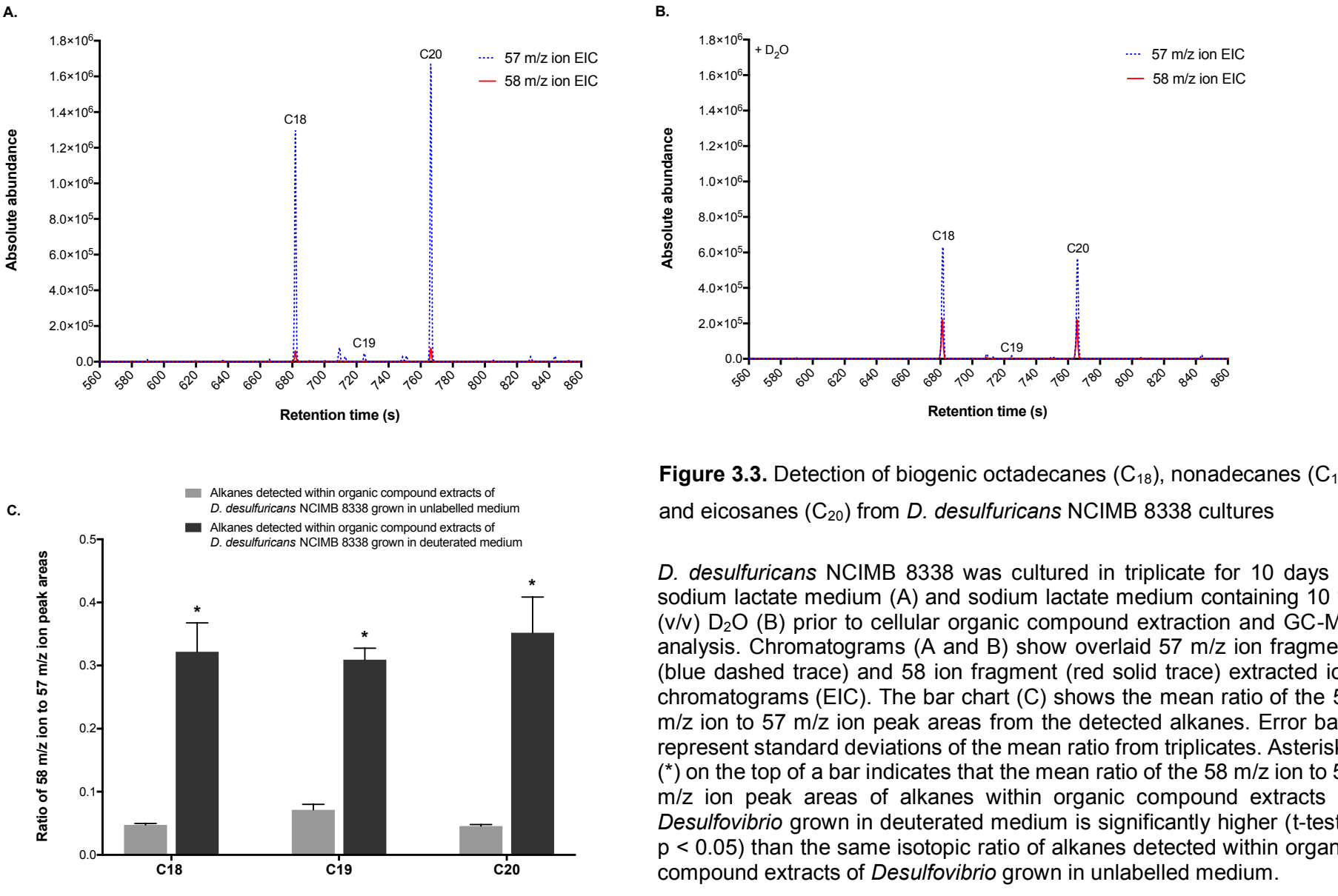


Figure 3.3. Detection of biogenic octadecanes (C_{18}), nonadecanes (C_{19}) and eicosanes (C_{20}) from *D. desulfuricans* NCIMB 8338 cultures

D. desulfuricans NCIMB 8338 was cultured in triplicate for 10 days in sodium lactate medium (A) and sodium lactate medium containing 10 % (v/v) D_2O (B) prior to cellular organic compound extraction and GC-MS analysis. Chromatograms (A and B) show overlaid 57 m/z ion fragment (blue dashed trace) and 58 ion fragment (red solid trace) extracted ion chromatograms (EIC). The bar chart (C) shows the mean ratio of the 58 m/z ion to 57 m/z ion peak areas from the detected alkanes. Error bars represent standard deviations of the mean ratio from triplicates. Asterisks (*) on the top of a bar indicates that the mean ratio of the 58 m/z ion to 57 m/z ion peak areas of alkanes within organic compound extracts of *Desulfovibrio* grown in deuterated medium is significantly higher (t-test's $p < 0.05$) than the same isotopic ratio of alkanes detected within organic compound extracts of *Desulfovibrio* grown in unlabelled medium.

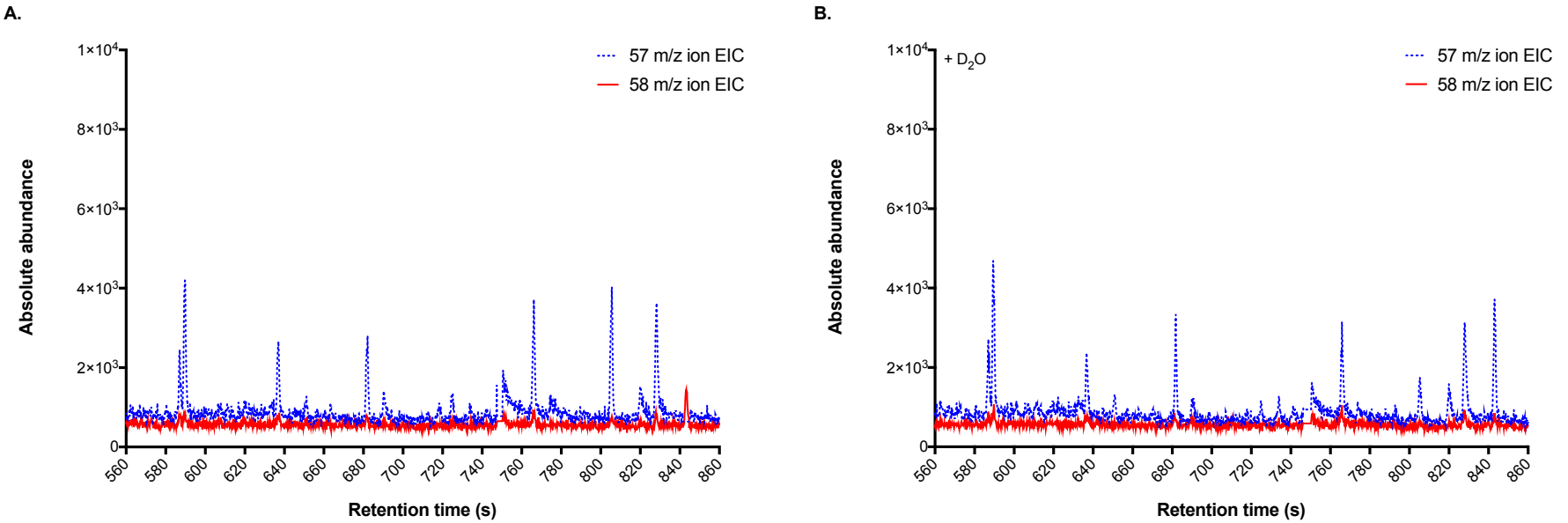


Figure 3.4. Undetectability of alkanes from *D. desulfuricans* NCIMB 8307 cultures

D. desulfuricans NCIMB 8307 was cultured in triplicates for 10 days in sodium lactate medium (A) and sodium lactate medium containing 10 % (v/v) D_2O (B) prior to cellular organic compound extraction and GC-MS analysis. Chromatograms (A and B) show overlaid 57 m/z ion fragment (blue dashed trace) and 58 ion fragment (red solid trace) extracted ion chromatograms (EIC).

Octadecanes (C_{18}), nonadecane (C_{19}) and eicosane (C_{20}) were detected by GC-MS within the organic compound extracts of *D. desulfuricans* NCIMB 8326 and *D. desulfuricans* NCIMB 8338. T-test results indicated that the mean ratio of 58 m/z ion to 57 m/z ion peak areas of alkanes detected within the organic compound extracts of *D. desulfuricans* NCIMB 8326 and *D. desulfuricans* NCIMB 8338 grown in deuterated medium were significantly higher than the same isotopic ratio of alkanes detected within the organic compound extracts of *D. desulfuricans* NCIMB 8326 and from *D. desulfuricans* NCIMB 8338 grown in unlabelled medium. Therefore, *D. desulfuricans* NCIMB 8326 and *D. desulfuricans* NCIMB 8338 were deduced to be capable of octadecane, nonadecane and eicosane synthesis, under these culture conditions.

However, no alkanes were detected by GC-MS within the organic compound extracts of *D. desulfuricans* NCIMB 8307, suggesting that *D. desulfuricans* NCIMB 8307 does not synthesise alkanes, under these culture conditions.

Screening of additional Desulfovibrio species for alkane production

Hydrocarbon synthesis screening was subsequently extended to seven additional *Desulfovibrio* species: *D. gabonensis* DSM 10636 (Figure 3.5), *D. gigas* NCIMB 9332 (Figure 3.6), *D. marinus* DSM 18311 (Figure 3.7), *D. paquesii* DSM 16681 (Figure 3.8), *D. vulgaris* strain Hildenborough (Figure 3.9), *D. giganteus* DSM 4370 (Figure 3.10) and *D. alcoholivorans* NCIMB 12906 (Figure 3.11).

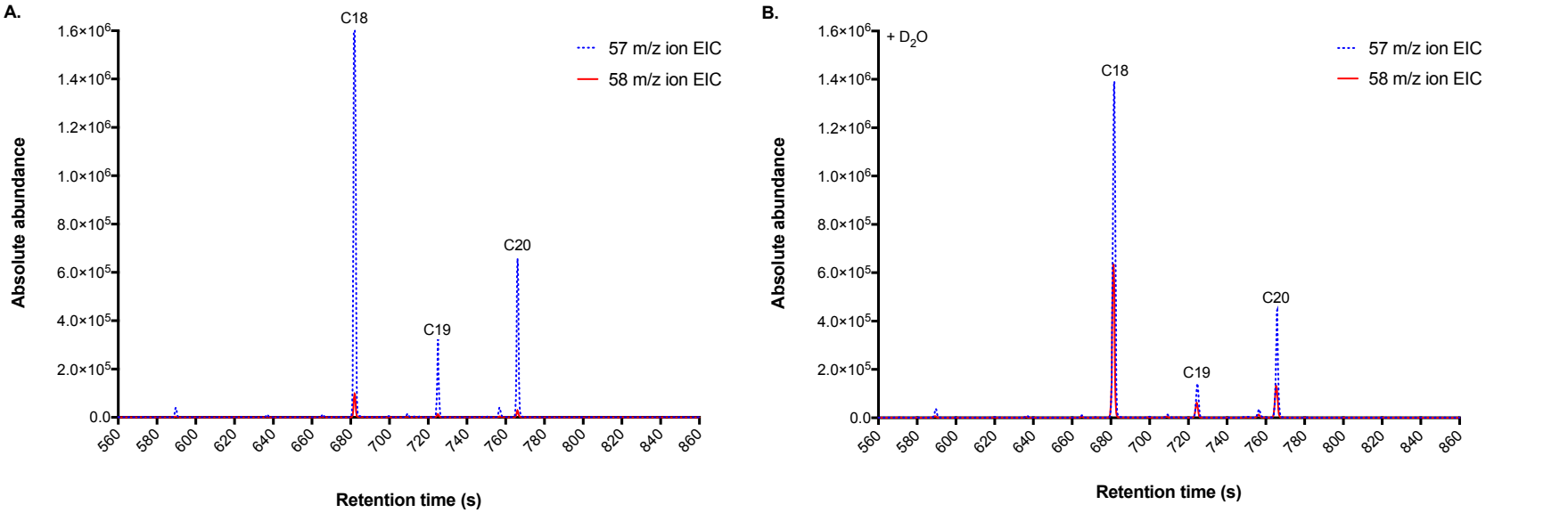


Figure 3.5. Detection of biogenic octadecanes (C_{18}), nonadecanes (C_{19}) and eicosanes (C_{20}) from *D. gabonensis* DSM 10636 cultures

D. gabonensis DSM 10636 was cultured in triplicates for 10 days in sodium lactate medium (A) and sodium lactate medium containing 10 % (v/v) D_2O (B) prior to cellular organic compound extraction and GC-MS analysis. Chromatograms (A and B) show overlaid 57 m/z ion fragment (blue dashed trace) and 58 ion fragment (red solid trace) extracted ion chromatograms (EIC). The bar chart (C) shows the mean ratio of the 58 m/z ion to 57 m/z ion peak areas from the detected alkanes. Error bars represent standard deviations of the mean ratio from triplicates. Asterisks (*) on the top of a bar indicates that the mean ratio of the 58 m/z ion to 57 m/z ion peak areas of alkanes detected within organic compound extracts of *Desulfovibrio* grown in deuterated medium is significantly higher (t-test's $p < 0.05$) than the same isotopic ratio of alkanes detected within organic compound extracts of *Desulfovibrio* grown in unlabelled medium.

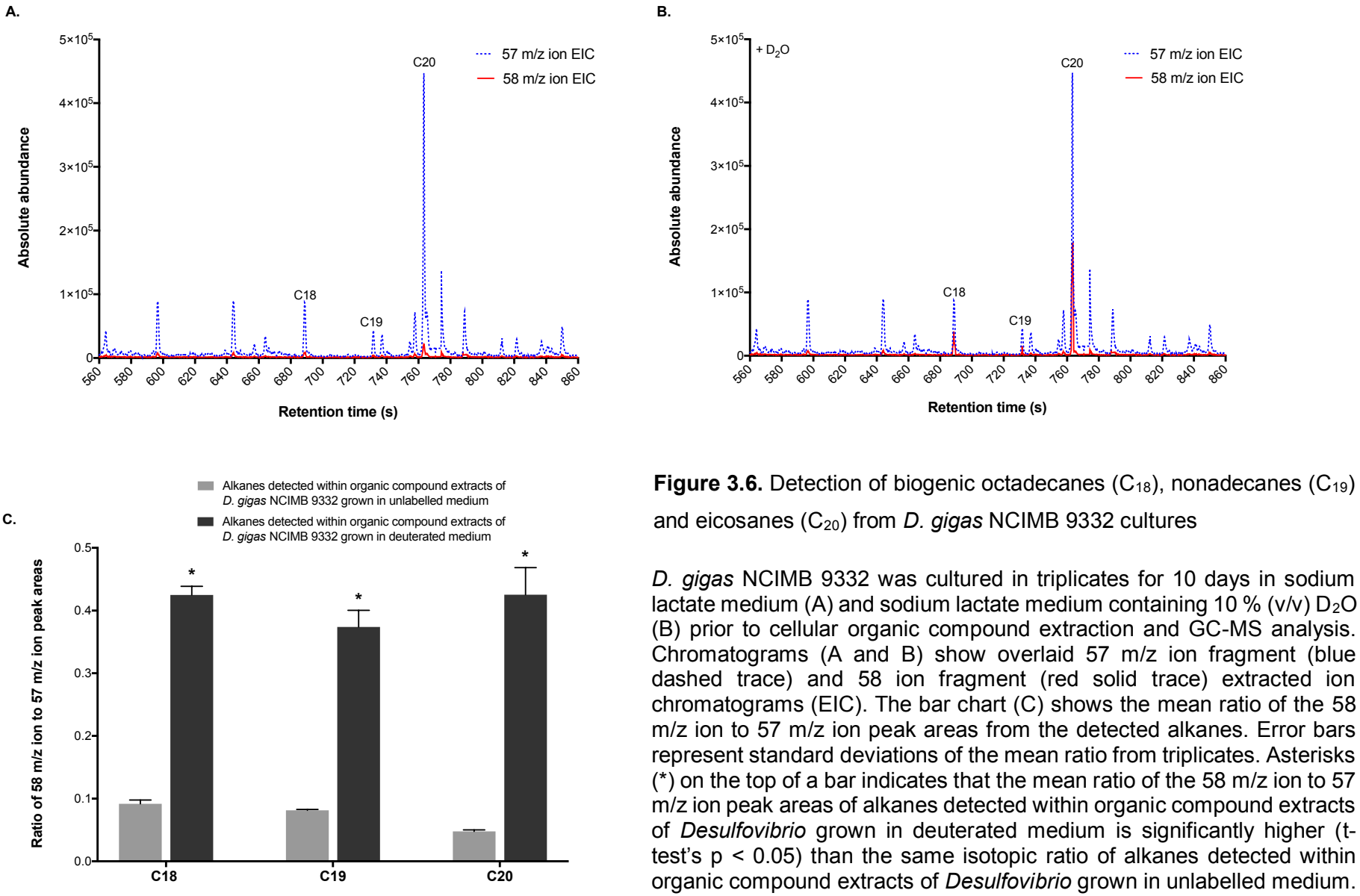


Figure 3.6. Detection of biogenic octadecanes (C_{18}), nonadecanes (C_{19}) and eicosanes (C_{20}) from *D. gigas* NCIMB 9332 cultures

D. gigas NCIMB 9332 was cultured in triplicates for 10 days in sodium lactate medium (A) and sodium lactate medium containing 10 % (v/v) D_2O (B) prior to cellular organic compound extraction and GC-MS analysis. Chromatograms (A and B) show overlaid 57 m/z ion fragment (blue dashed trace) and 58 ion fragment (red solid trace) extracted ion chromatograms (EIC). The bar chart (C) shows the mean ratio of the 58 m/z ion to 57 m/z ion peak areas from the detected alkanes. Error bars represent standard deviations of the mean ratio from triplicates. Asterisks (*) on the top of a bar indicates that the mean ratio of the 58 m/z ion to 57 m/z ion peak areas of alkanes detected within organic compound extracts of *Desulfovibrio* grown in deuterated medium is significantly higher (t-test's $p < 0.05$) than the same isotopic ratio of alkanes detected within organic compound extracts of *Desulfovibrio* grown in unlabelled medium.

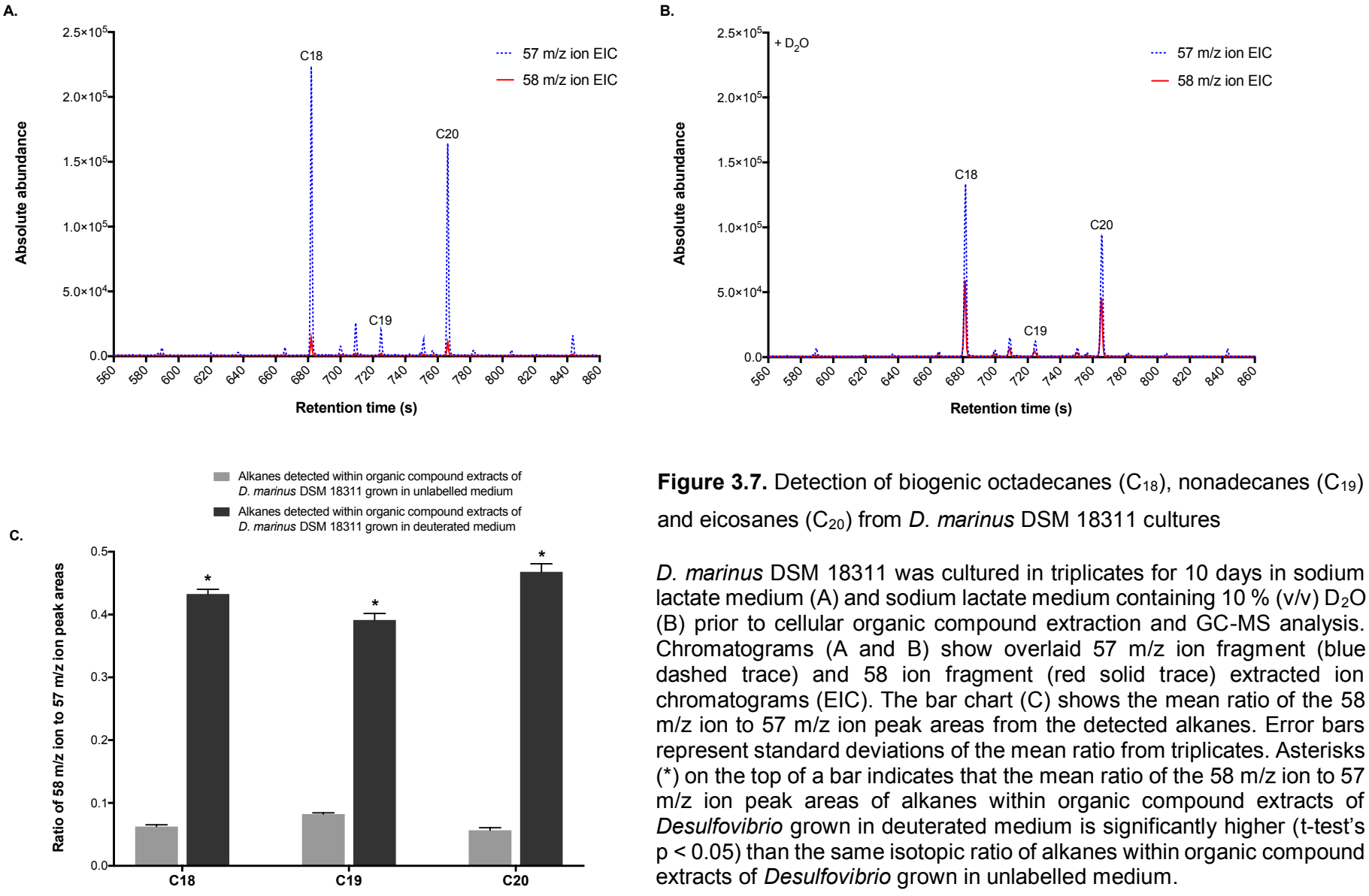


Figure 3.7. Detection of biogenic octadecanes (C_{18}), nonadecanes (C_{19}) and eicosanes (C_{20}) from *D. marinus* DSM 18311 cultures

D. marinus DSM 18311 was cultured in triplicates for 10 days in sodium lactate medium (A) and sodium lactate medium containing 10 % (v/v) D_2O (B) prior to cellular organic compound extraction and GC-MS analysis. Chromatograms (A and B) show overlaid 57 m/z ion fragment (blue dashed trace) and 58 ion fragment (red solid trace) extracted ion chromatograms (EIC). The bar chart (C) shows the mean ratio of the 58 m/z ion to 57 m/z ion peak areas from the detected alkanes. Error bars represent standard deviations of the mean ratio from triplicates. Asterisks (*) on the top of a bar indicates that the mean ratio of the 58 m/z ion to 57 m/z ion peak areas of alkanes within organic compound extracts of *Desulfovibrio* grown in deuterated medium is significantly higher (t-test's $p < 0.05$) than the same isotopic ratio of alkanes within organic compound extracts of *Desulfovibrio* grown in unlabelled medium.

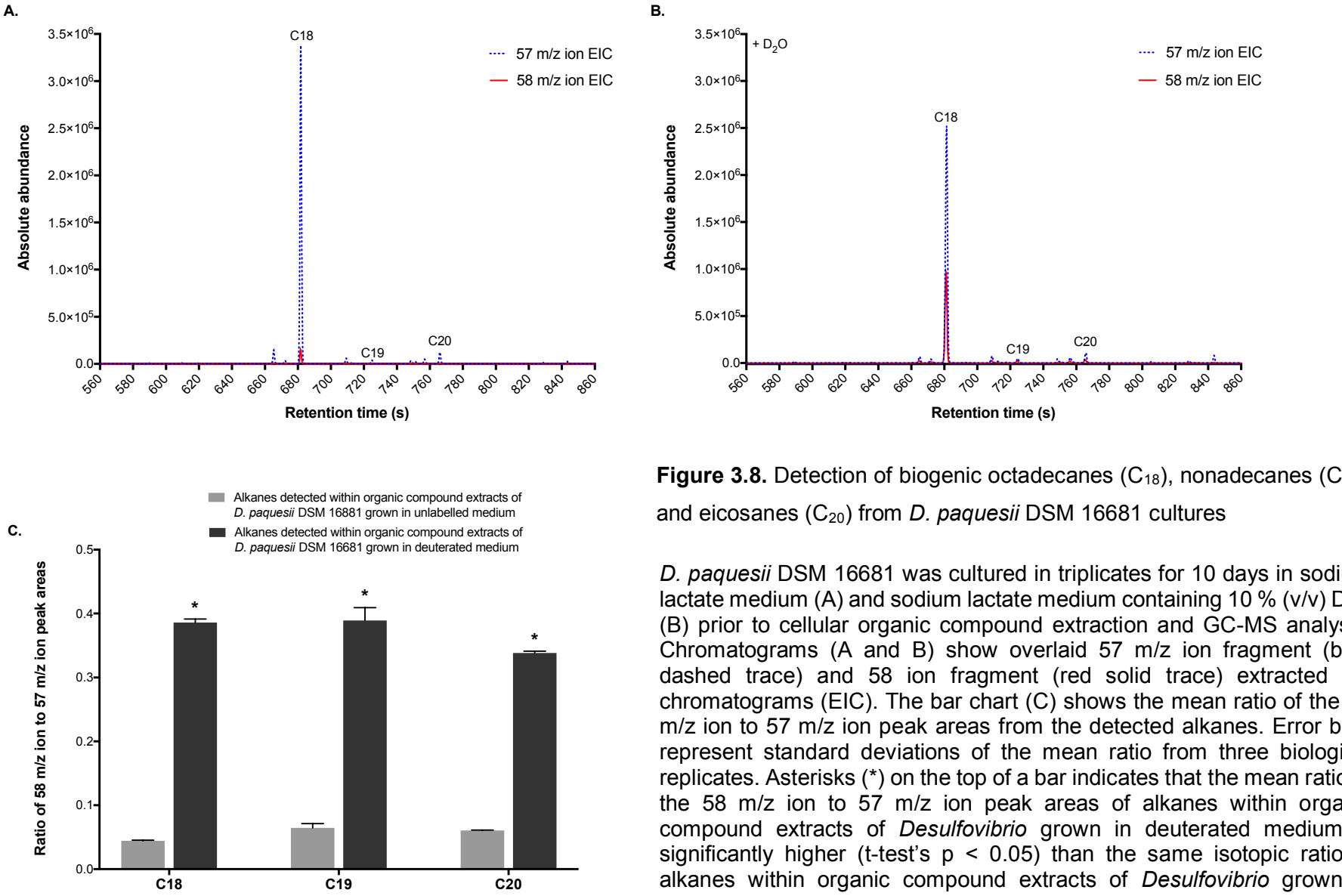


Figure 3.8. Detection of biogenic octadecanes (C_{18}), nonadecanes (C_{19}) and eicosanes (C_{20}) from *D. paquesii* DSM 16681 cultures

D. paquesii DSM 16681 was cultured in triplicates for 10 days in sodium lactate medium (A) and sodium lactate medium containing 10 % (v/v) D_2O (B) prior to cellular organic compound extraction and GC-MS analysis. Chromatograms (A and B) show overlaid 57 m/z ion fragment (blue dashed trace) and 58 ion fragment (red solid trace) extracted ion chromatograms (EIC). The bar chart (C) shows the mean ratio of the 58 m/z ion to 57 m/z ion peak areas from the detected alkanes. Error bars represent standard deviations of the mean ratio from three biological replicates. Asterisks (*) on the top of a bar indicates that the mean ratio of the 58 m/z ion to 57 m/z ion peak areas of alkanes within organic compound extracts of *Desulfovibrio* grown in deuterated medium is significantly higher (t-test's $p < 0.05$) than the same isotopic ratio of alkanes within organic compound extracts of *Desulfovibrio* grown in unlabelled medium.

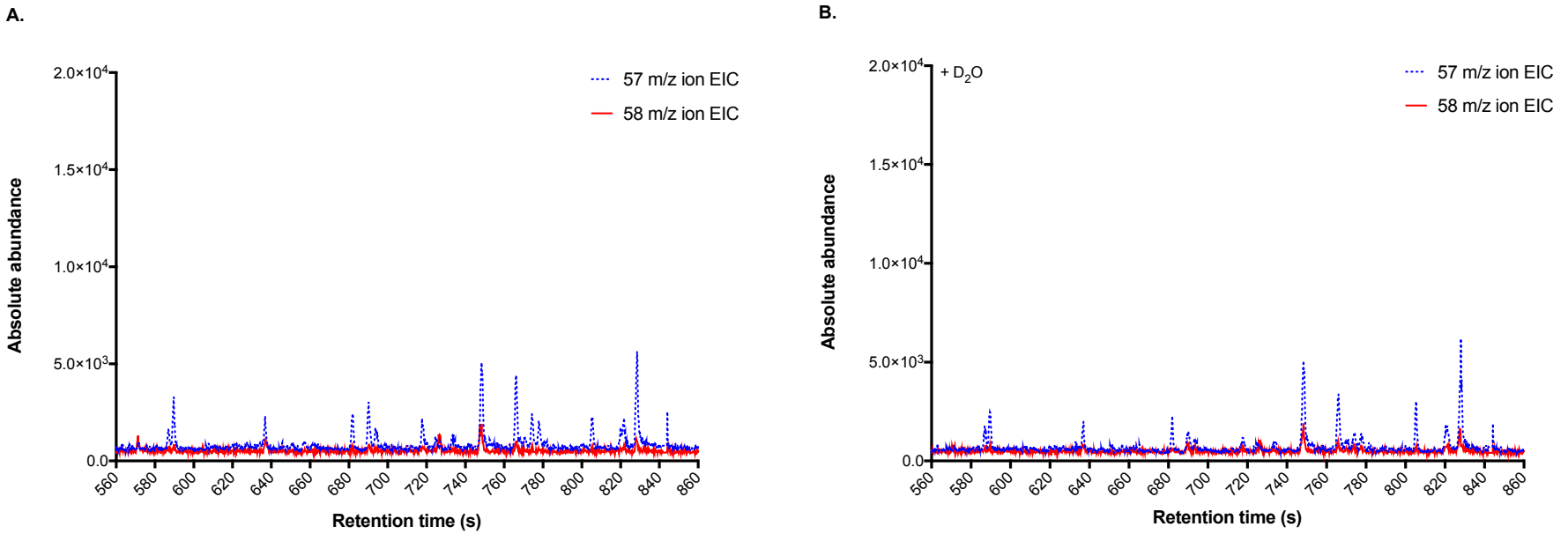
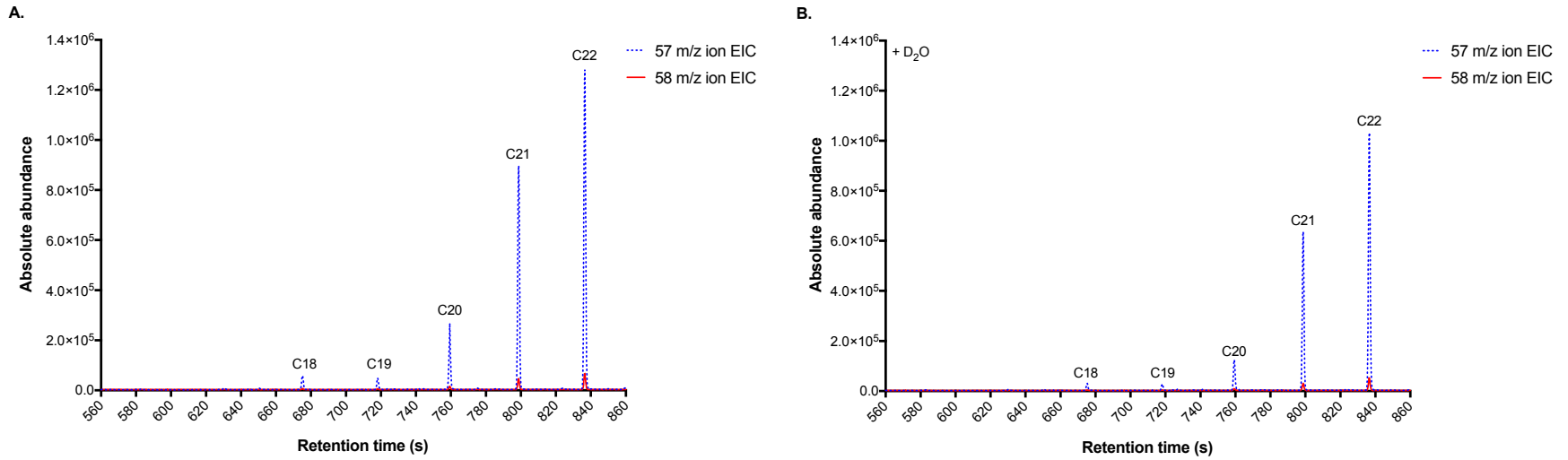


Figure 3.9. Undetectability of alkanes from *D. vulgaris* strain Hildenborough cultures

D. vulgaris strain Hildenborough was cultured in triplicates for 10 days in sodium lactate medium (A) and sodium lactate medium containing 10 % (v/v) D_2O (B) prior to cellular organic compound extraction and GC-MS analysis. Chromatograms (A and B) show overlaid 57 m/z ion fragment (blue dashed trace) and 58 ion fragment (red solid trace) extracted ion chromatograms (EIC).



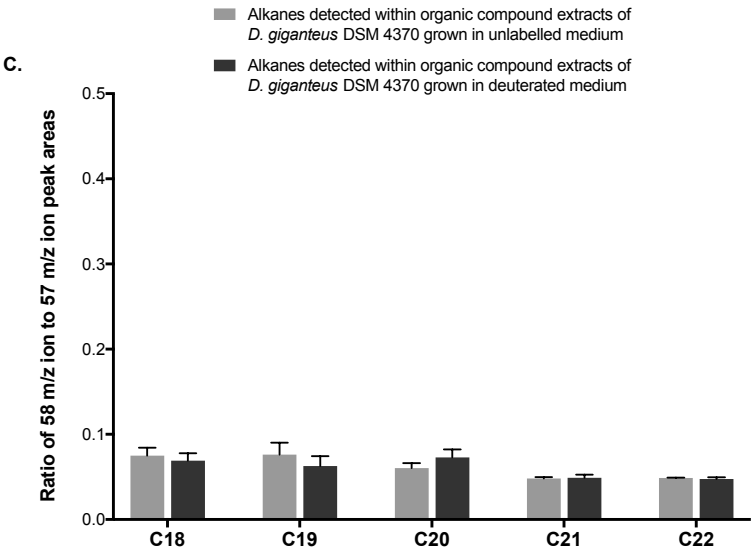


Figure 3.10. Detection of non-biogenic alkanes (C_{18}-C_{22}) from *D. giganteus* DSM 4370 cultures

D. giganteus DSM 4370 was cultured in triplicates for 10 days in sodium lactate medium (A) and sodium lactate medium containing 10 % (v/v) D_2O (B) prior to cellular organic compound extraction and GC-MS analysis. Chromatograms (A and B) show overlaid 57 m/z ion fragment (blue dashed trace) and 58 ion fragment (red solid trace) extracted ion chromatograms (EIC). The bar chart (C) shows the mean ratio of the 58 m/z ion to 57 m/z ion peak areas from the alkanes detected within organic compound extracts of *Desulfovibrio* grown in unlabelled and deuterated medium. Error bars represent standard deviations of the mean ratio from triplicates.

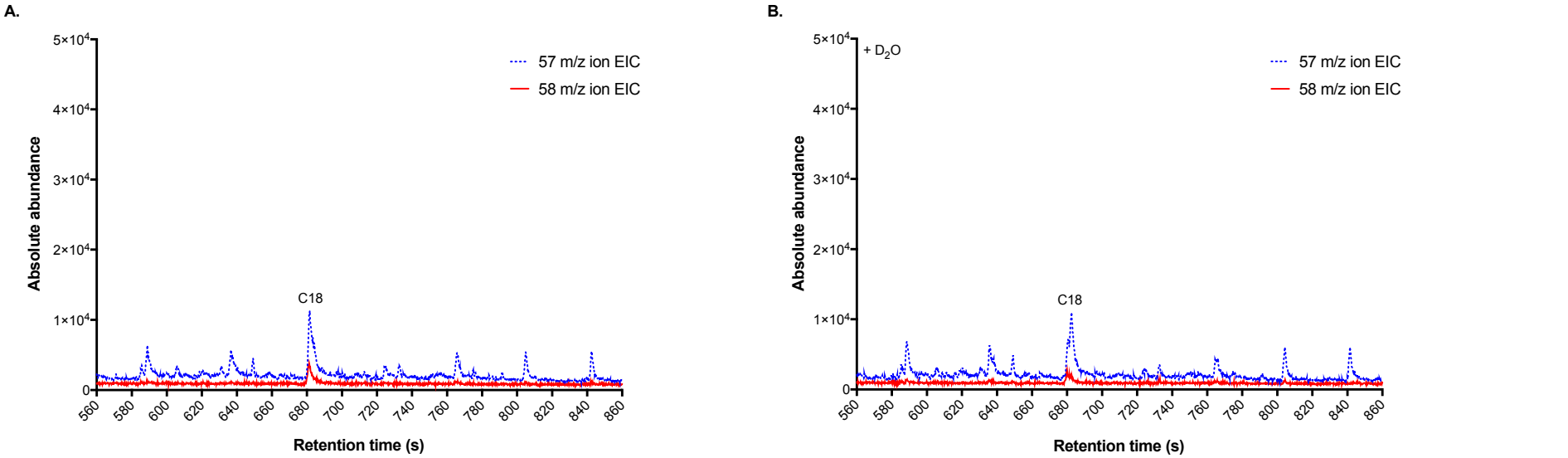


Figure 3.11. Detection of non-biogenic octadecanes from *D. alcoholivorans* NCIMB 12906 cultures

D. alcoholivorans NCIMB 12906 was cultured in triplicates for 10 days in sodium lactate medium (A) and sodium lactate medium containing 10 % (v/v) D_2O (B) prior to cellular organic compound extraction and GC-MS analysis. Chromatograms (A and B) show overlaid 57 m/z ion fragment (blue dashed trace) and 58 ion fragment (red solid trace) extracted ion chromatograms (EIC). The bar chart (C) shows the mean ratio of the 58 m/z ion to 57 m/z ion peak areas from the alkanes detected within organic compound extracts of *Desulfovibrio* grown in unlabelled and deuterated medium. Error bars represent standard deviations of the mean ratio from triplicates.

Octadecanes (C_{18}), nonadecanes (C_{19}) and eicosanes (C_{20}) were detected by GC-MS within the organic compound extracts of *D. gabonensis* DSM 10636, *D. gigas* NCIMB 9332, *D. marinus* DSM 18311 and *D. paquesii* DSM 16681. T-test results indicated that the mean ratio of 58 m/z ion to 57 m/z ion peak areas of alkanes detected within the organic compound extracts of *D. gabonensis* DSM 10636, *D. gigas* NCIMB 9332, *D. marinus* DSM 18311 and *D. paquesii* DSM 16681 grown in deuterated medium were significantly higher than the same isotopic ratio of alkanes detected within the organic compound extracts of these strains grown in unlabelled medium. Therefore, *D. gabonensis* DSM 10636, *D. gigas* NCIMB 9332, *D. marinus* DSM 18311, *D. paquesii* DSM 16681 were considered to be capable of octadecane, nonadecane and eicosane synthesis, under these culture conditions.

In contrast, no alkanes were detected by GC-MS within the organic compound extracts of *D. vulgaris* strain Hildenborough, suggesting that *D. vulgaris* strain Hildenborough does not synthesise alkanes, under these culture conditions.

Alkanes from C_{18} to C_{22} chain length were detected by GC-MS within the organic compound extracts of *D. giganteus* DSM 4370 and *D. alcoholivorans* NCIMB 12906. However, t-test results indicated the mean ratio of 58 m/z ion to 57 m/z ion peak areas of alkanes detected within the organic compound extracts of *D. giganteus* DSM 4370 and *D. alcoholivorans* NCIMB 12906 grown in deuterated medium were not significantly higher (t-test's $p > 0.05$) than the same isotopic ratio of alkanes detected within the organic compound extracts of these strains grown in unlabelled medium. Therefore, alkanes detected by GC-MS in these organic compound extracts did not have a biogenic origin. *D. giganteus* DSM 4370 and *D. alcoholivorans* NCIMB 12906 were thus deduced incapable to synthesise alkanes, under these culture conditions.

Table 3.1. Summary of alkane production by the *Desulfovibrio* spp. screened in this study

***Desulfovibrio* species and strain**	**Culture collection catalogue number**	**Detection of alkane production**	**Biogenic alkanes detected**
D. desulfuricans subsp. desulfuricans California27.137.5	NCIMB 8326	Yes	Octadecane (C_{18}) Nonadecane (C_{19}) Eicosane (C_{20})
D. desulfuricans subsp. desulfuricans CubaHC29.130.4	NCIMB 8338	Yes	Octadecane (C_{18}) Nonadecane (C_{19}) Eicosane (C_{20})
D. gabonensis SEBR 2840	DSM 10636	Yes	Octadecane (C_{18}) Nonadecane (C_{19}) Eicosane (C_{20})
D. gigas	NCIMB 9332	Yes	Octadecane (C_{18}) Nonadecane (C_{19}) Eicosane (C_{20})
D. marinus E-2	DSM 18311	Yes	Octadecane (C_{18}) Nonadecane (C_{19}) Eicosane (C_{20})
D. paquesii SB1	DSM 16681	Yes	Octadecane (C_{18}) Nonadecane (C_{19}) Eicosane (C_{20})
D. desulfuricans subsp. desulfuricans Essex6	NCIMB 8307	No	-
D. vulgaris Hildenborough	-	No	-
D. giganteus STg	DSM 4370	No	-
D. alcoholivorans	NCIMB 12906	No	-

3.2.3. Quantification of Biogenic Alkanes

In order to quantify alkanes produced by *Desulfovibrio* spp., calibration curves of alkane concentration in mg ml^{-1} to the peak area was performed for octadecane, nonadecane and eicosane (Figure 3.12).

For each alkane producing *Desulfovibrio* strain, the concentration in mg ml^{-1} of biogenic alkanes was determined according to their peak areas using the calibration curves. The mass concentration was then converted to a molar concentration. The amount of biogenic alkane moles extracted from 40 ml cultures was calculated and normalised to the amount of total protein of the cultures used for the organic compound extraction (Figure 3.13).

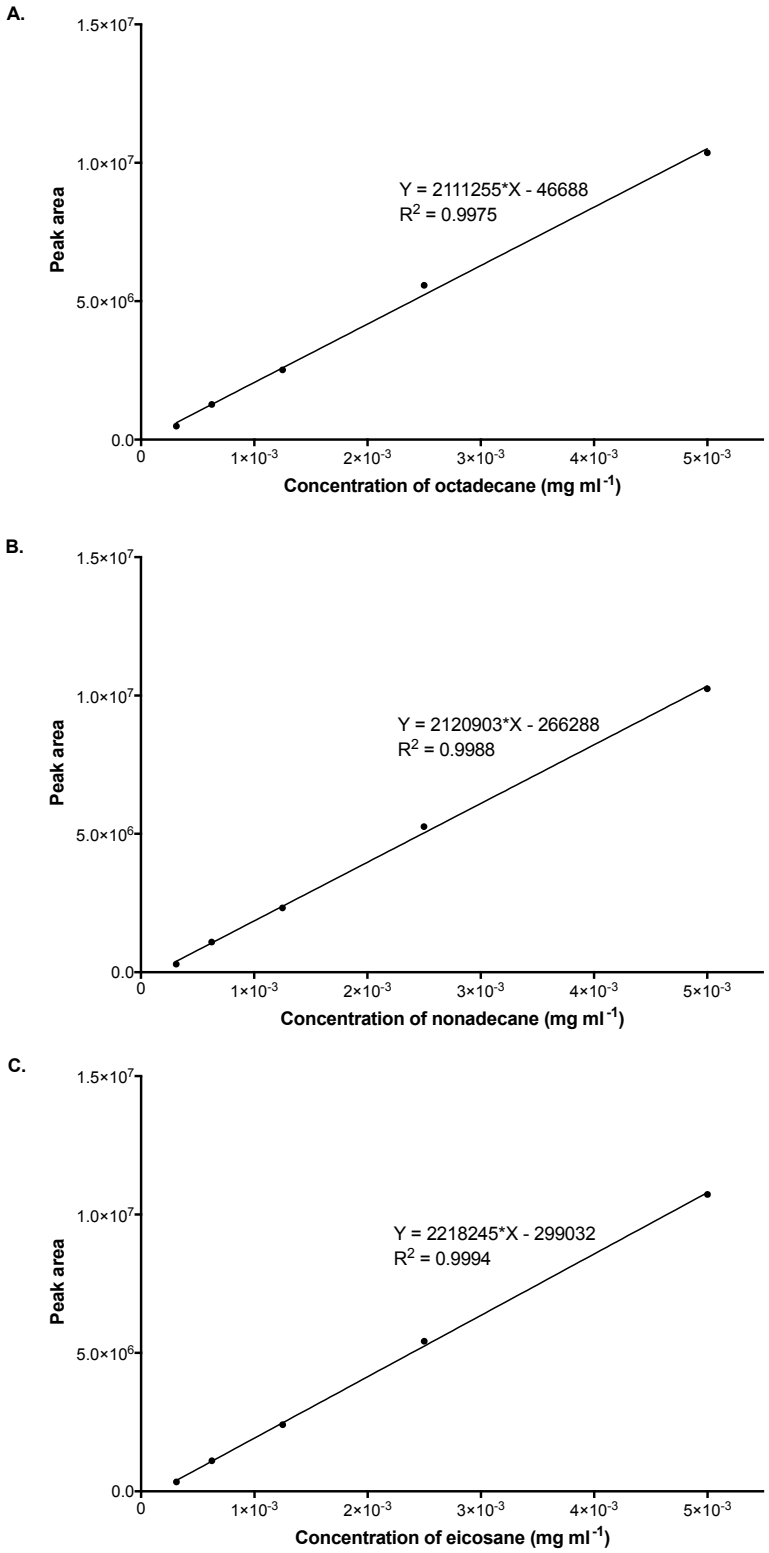


Figure 3.12. Octadecane, nonadecane and eicosane calibration curves of concentration (mg ml^{-1}) to peak area

Solutions of 5 $x10^{-3}$ mg ml^{-1}, 2.5 $x10^{-3}$ mg ml^{-1}, 1.25$x10^{-3}$ mg ml^{-1}, 0.625$x10^{-3}$ mg ml^{-1} and 0.3125 $x10^{-3}$ mg ml^{-1} octadecane (A), nonadecane (B) and eicosane (C) diluted in DCM were analysed by GC-MS. The alkane peak areas were extracted and plotted against alkane concentration.

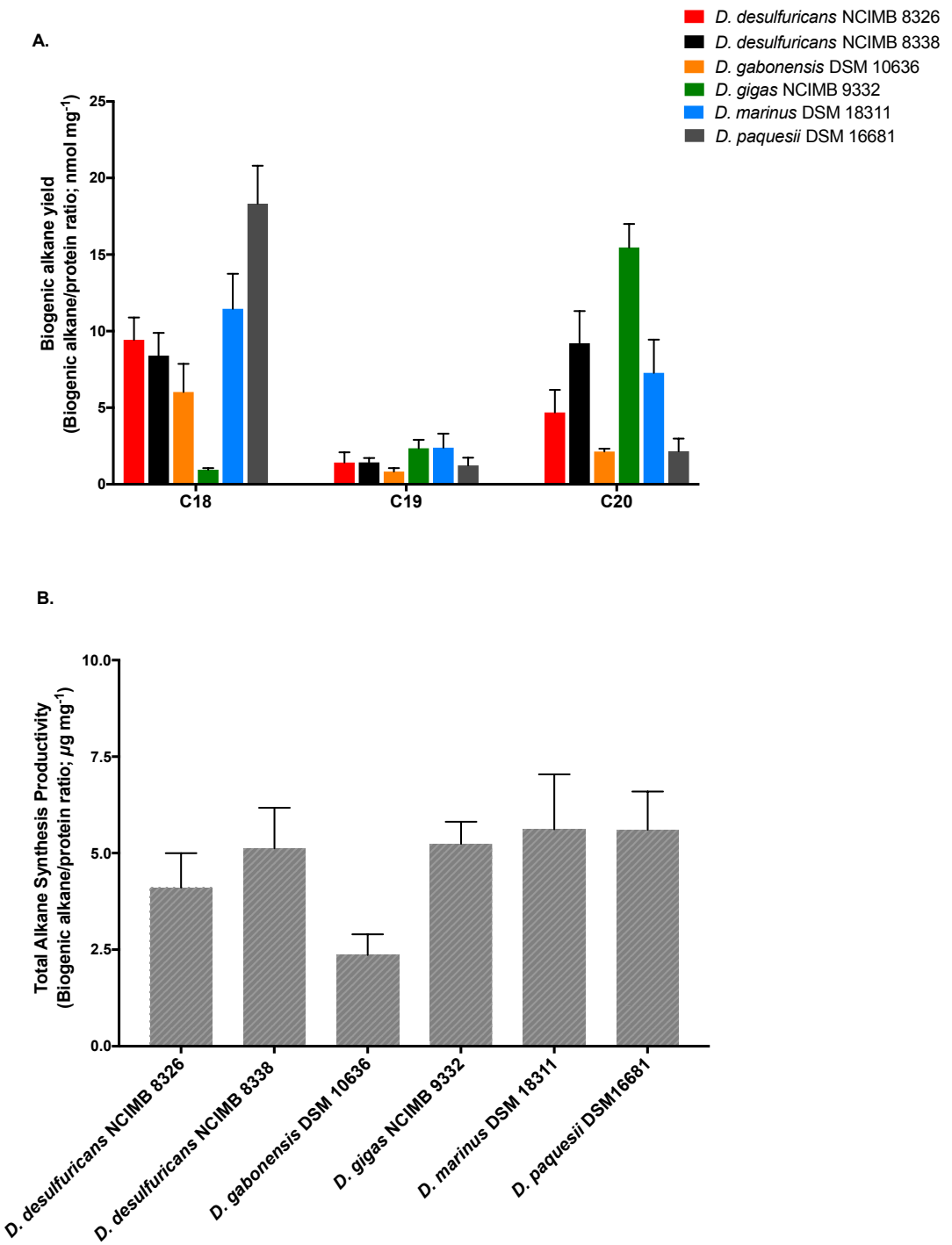


Figure 3.13. Biogenic alkane yields and total alkane synthesis productivity by alkane producing *Desulfovibrio* strains

Biogenic alkanes were extracted from 40 ml *Desulfovibrio* cultures incubated for 10 days under an anaerobic atmosphere of 80 % N_2, 10 % CO_2 and 10 % H_2, and analysed by GC-MS. Biogenic alkanes were quantified using MassHunter Q-TOF Quantitative Analysis software (Agilent Technologies). The quantified amount of biogenic alkanes was normalised to the amount of total proteins extracted from the initial 40 ml cultures, to determine the yield of each biogenic alkane (A) and the total alkane synthesis productivity of each *Desulfovibrio* strains (B). Error bars represent standard deviations of the biogenic alkane yield or the total alkane synthesis productivity from triplicates.

Within this study, alkane producing *Desulfovibrio* strains predominantly synthesised octadecanes (C_{18}) and eicosanes (C_{20}). The highest yield of octadecane was produced by *D. paquesii* DSM 16681 with 18.33 nmol of octadecane produced on average per mg of total proteins in 10 days. The highest yield of eicosane was produced by *D. gigas* NCIMB 9332 with 15.46 nmol of eicosane produced on average per mg of total proteins in 10 days. The highest productivity for total alkane synthesis was achieved by *D. marinus* DSM 18311 with 5.61 *μ*g of total alkanes produced on average per mg of total proteins.

3.3. Desulfovibrio Metabolism Screening for Biogenic Fatty Acids

3.3.1. Fatty acid Standard Mass Spectra and Retention Times

Standard solutions of 10 *μ*M tetradecanoic acid ($C_{14:0}$), pentadecanoic acid ($C_{15:0}$), hexadecanoic acid ($C_{16:0}$), heptadecanoic acid ($C_{17:0}$), octadecanoic acid ($C_{18:0}$), nonadecanoic acid ($C_{20:0}$) and heneicosanoic acid ($C_{21:0}$) were derivatised by silylation, prior to being analysed by GC-MS for mass spectrum fragmentation pattern (Figure 3.14A) and retention time determination (Figure 3.14B).

The mass spectrum of a fatty acid trimethylsilyl ester derivative showed a fragmentation pattern characterised by the ion fragments of 75.03 m/z (corresponding to the oxygenised trimethylsilyl group, $C_2H_7OSi^+$), 117.03 m/z ($C_4H_9O_2Si+$) and 129.04 m/z ($C_5H_9O_2Si^+$). From the ion peak cluster of 129.04 m/z, the ion peak clusters are evenly separated by 14 mass units, corresponding to the loss of a CH_2 group from the fatty acid trimethylsilyl ester derivative. The mass spectrum of organic compounds extracted from *Desulfovibrio* and derivatised by silylation were screened for the same fragmentation pattern to the fatty acid trimethylsilyl ester derivative mass spectrum, in order to detect fatty acids within organic compound extracts.

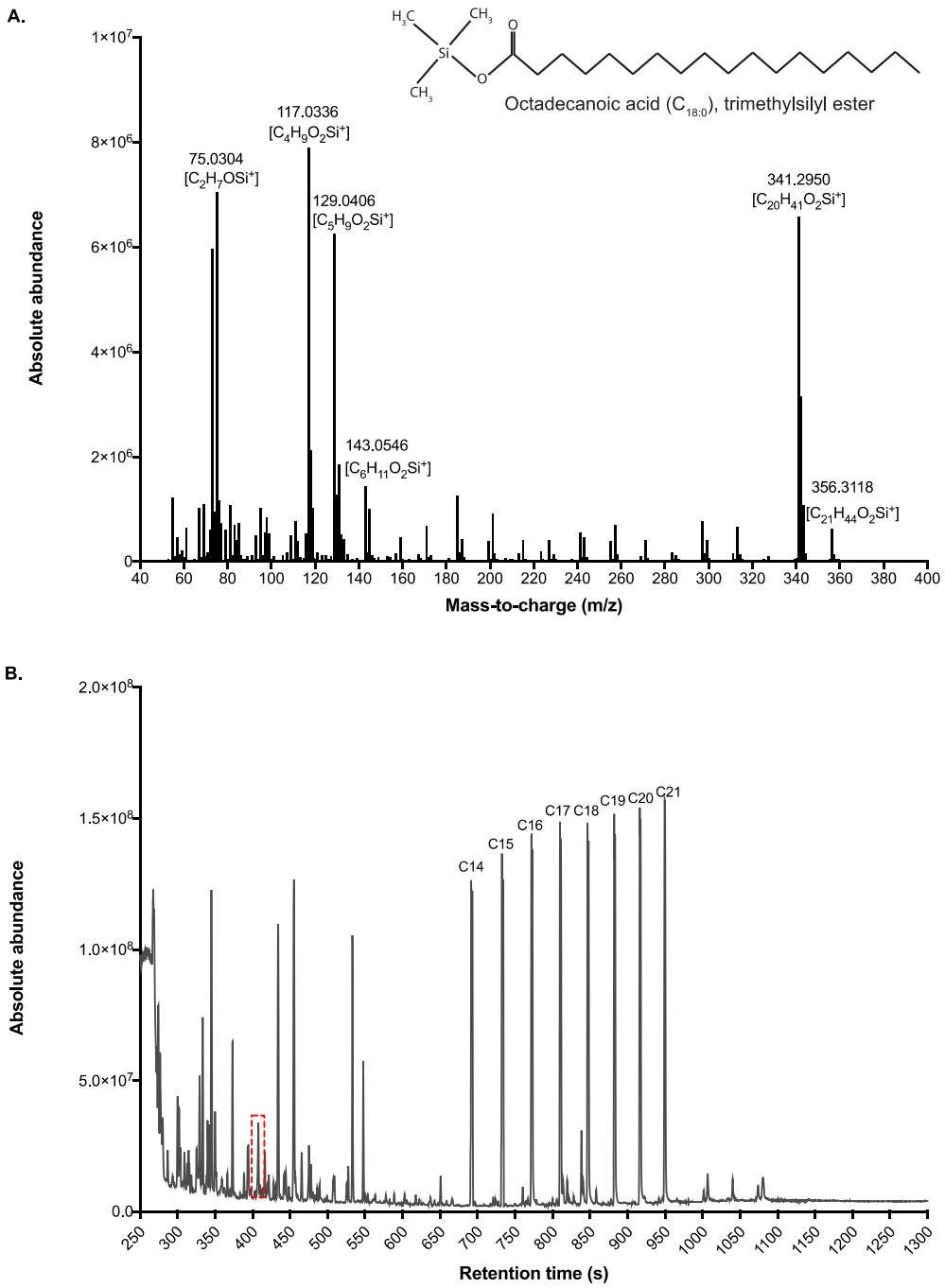


Figure 3.14. Mass spectrum of the octadecanoic acid ($C_{18:0}$), trimethylsilyl ester standard and total ion chromatograms of fatty acid, trimethylsilyl ester derivative standards

Standard solutions of $C_{14:0}$-$C_{21:0}$ fatty acid, spiked with β-cyclocitral, were derivatised by silylation and analysed by CG-MS for mass spectrum fragmentation pattern and retention time determination. The mass spectrum of the octadecanoic acid ($C_{18:0}$), trimethylsilyl ester standard, with a retention time of 849.2 s, showed the fragmentation pattern of fatty acid trimethylsilyl ester derivatives ionised by electron impact ionisation (A). The total ion chromatogram of $C_{14:0}$-$C_{21:0}$ fatty acid trimethylsilyl ester derivative standards allowed the determination of the retention time of each fatty acid (B). The peak corresponding to the β-cyclocitral (internal standard) is enclosed by a red dashed box.

3.3.2. *Desulfovibrio* Fatty Acid Content

Biogenic fatty acids from *D. desulfuricans* NCIMB 8326, *D. desulfuricans* NCIMB 8338, *D. gabonensis* DSM 10636, *D. gigas* NCIMB 9332, *D. marinus* DSM 18311, *D. paquesii* DSM 16681, *D. desulfuricans* NCIMB 8307, *D. vulgaris* strain Hildenborough, *D. giganteus* DSM 4370 and *D. alcoholivorans* NCIMB 12906 were extracted in DCM, derivatised by silylation and analysed by GC-MS. The peak area of the m/z 143.0546 ion fragment from each detected fatty acid was extracted, integrated and normalised to the total protein amount of the cultures used for organic compound extraction. The proportion of each fatty acid was then calculated (Figure 3.15).

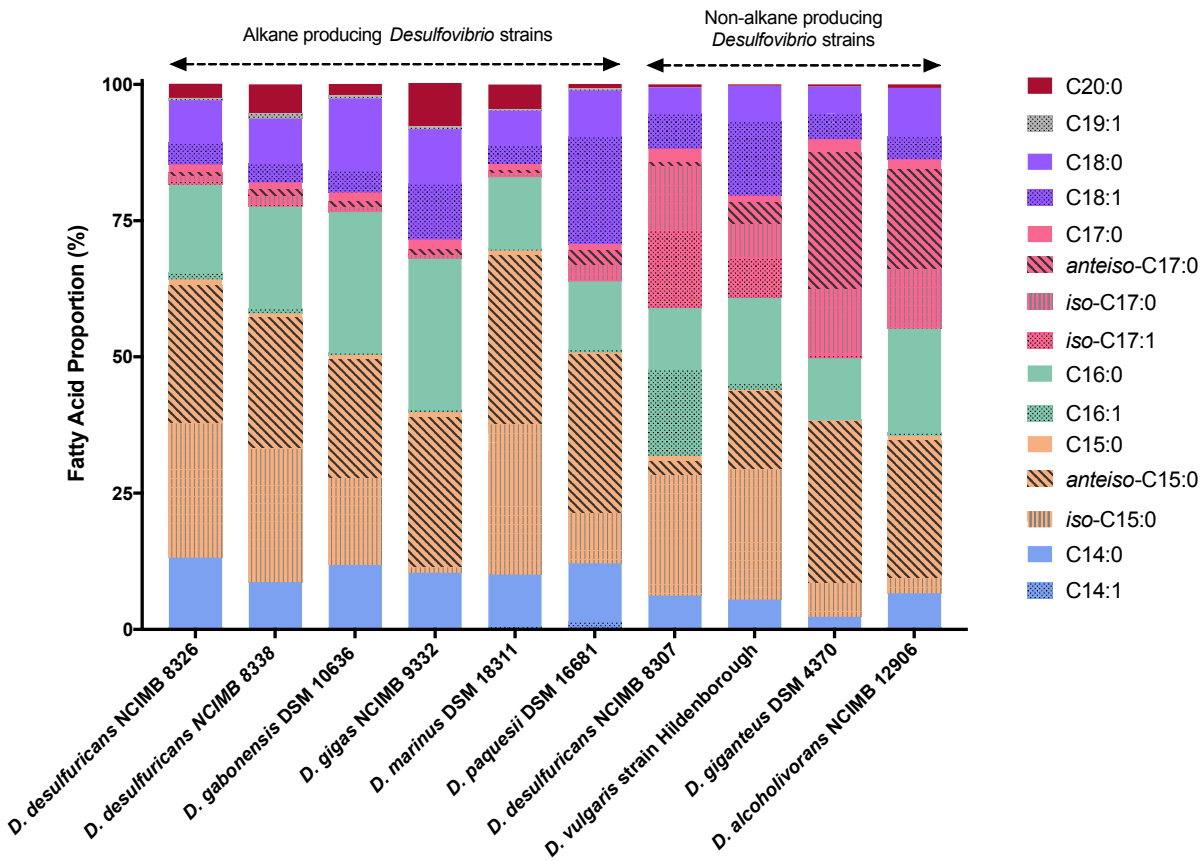


Figure 3.15. Fatty acid content of the *Desulfovibrio* spp. screened in this study

Mean proportion of each detected fatty acid from the alkane producing strains (*D. desulfuricans* NCIMB 8326, *D. desulfuricans* NCIMB 8338, *D. gabonensis* DSM 10636, *D. gigas* NCIMB 9332, *D. marinus* DSM 18311, *D. paquesii* DSM 16681) and the non-alkane producing strains (*D. desulfuricans* NCIMB 8307, *D. vulgaris* strain Hildenborough, *D. giganteus* DSM 4370, *D. alcoholivorans* NCIMB 12906) is shown. Mean proportions were calculated from triplicates (error bars not shown). The fatty acids detected by GC-MS from *Desulfovibrio* spp. were: C14:1, unsaturated carbon chain tetradecenoic acid; C14:0, saturated straight carbon chain tetradecanoic acid; *iso*-C15:0, saturated *iso*-branched carbon chain pentadecanoic acid; *anteiso*-C15:0, saturated *anteiso*-branched carbon chain pentadecanoic acid; C15:0, saturated straight carbon chain pentadecanoic acid; C16:1, unsaturated carbon chain hexadecenoic acid; C16:0, saturated straight carbon chain hexadecanoic acid; *iso*-C17:1, unsaturated *iso*-branched carbon chain heptadecenoic acid; *iso*-C17:0, saturated *iso*-branched carbon chain heptadecanoic acid; *anteiso*-C17:0, saturated *anteiso*-branched carbon chain heptadecanoic acid; C17:0, saturated straight carbon chain heptadecanoic acid; C18:1, unsaturated carbon chain octadecenoic acid; C18:0, saturated straight carbon chain octadecanoic acid; C19:1; unsaturated carbon chain nonadecenoic acid, C20:0, saturated straight carbon chain eicosanoic acid.

According to the GC-MS analysis, the predominant fatty acids synthesised by the *Desulfovibrio* spp. within this study were saturated branched carbon chain pentadecanoic acid ($C_{15:0}$) and heptadecanoic acid ($C_{17:0}$), saturated straight carbon chain hexadecanoic acid ($C_{16:0}$) and unsaturated carbon chain octadecenoic acid ($C_{18:1}$). Interestingly, the biogenic fatty acids between each *Desulfovibrio* spp. within this study showed differences in structure and in quantity of fatty acids. Moreover, the fatty acid content of alkanes producing *Desulfovibrio* differs from the fatty acid content of non-alkane producing strains. For example, saturated straight carbon chain eicosanoic acid ($C_{20:0}$) was detected in higher quantity within the organic compounds extracts from alkane producing strains than from the non-alkane producing strains. Conversely, non-alkane producing strains were shown to produce branched carbon chain heptadecanoic acid ($C_{17:0}$) in higher quantity than alkane producing strains. Furthermore, unsaturated carbon chain nonadecenoic acid ($C_{19:1}$) was exclusively detected within the organic compounds extracts from the alkane producing *Desulfovibrio* strains.

3.4. Discussion

Alkane biosynthesis within the genus *Desulfovibrio* was confirmed in this study, and interestingly, was shown to be strain-specific under one set of conditions. *D. desulfuricans* NCIMB 8326, *D. desulfuricans* NCIMB 8338, *D. gabonensis* DSM 10636, *D. gigas* NCIMB 9332, *D. marinus* DSM 18311 and *D. paquesii* DSM 16681 were proven to be capable of alkane synthesis. However, no biogenic alkanes were detected by GC-MS within *D. desulfuricans* NCIMB8307, *D. vulgaris* strain Hildenborough, *D. giganteus* DSM 4370 and *D. alcoholivorans* NCIMB 12906 organic compound extracts, suggesting that these strains do not synthesise alkanes, under these culture conditions. This is contrary to previous papers reporting *D. desulfuricans* NCIMB 8307 as an alkane producer (Bagaeva & Chernova, 1994; Bagaeva & Beliaeva, 2000; Bagaeva, 2000).

The ability to produce alkanes by *Desulfovibrio* spp. was determined by GC-MS analysis and thus relied on the detectability of the GC-MS for biogenic alkanes. Consequently, the deduced inability of alkane production by *Desulfovibrio* spp. could be due to biogenic alkane concentrations being below the level of detection of the GC-MS. However, advances in GC-MS enable to achieve low limits of detection (Fialkov *et al.*, 2007). In this study, the lowest quantity of C_8-C_{20} alkane strandards analysed by the GC-MS was 312.5 pg $\mu l^{-1}$and alkanes were detected by GC-MS. The lowest biogenic alkane

concentration previously reported to be produced by *D. desulfuricans* NCIMB 8307 was determined at 8.3 ng μl^{-1} (Bagaeva, 2000), suggesting that the GC-MS method used in this study would have detected biogenic alkanes within *D. desulfuricans* NCIMB 8307 organic compound extracts.

As the ability to produce alkanes is not ubiquitous within the genus *Desulfovibrio*, it can be inferred that alkanes are non-essential for *Desulfovibrio* growth and thus are considered secondary metabolites (Demain & Fang, 2000). Production of secondary metabolites can be influenced by the culture conditions (Ruiz *et al.*, 2010), which may explain the absence of biogenic alkane detection by GC-MS within *Desulfovibrio* organic compound extracts. *D. desulfuricans* NCIMB 8307 was shown to produce hydrocarbons when grown on lactate, pyruvate and ethanol carbon sources (Bagaeva & Beliaeva, 2000). In this study, *Desulfovibrio* strains were cultured on sodium lactate based medium, suggesting that the undetected alkane biosynthesis by *Desulfovibrio* was unlikely due to the carbon source. It has also been proven that the composition of the anaerobic atmosphere affected the hydrocarbon synthesis productivity of *D. desulfuricans* NCIMB 8307. Under an atmosphere composed of CO_2 and H_2 (1:19), hydrocarbon synthesis productivity of *D. desulfuricans* NCIMB 8307 was determined at 0.11 mg of hydrocarbons produced on average per mg of total proteins, while under an atmosphere composed of CO_2 and H_2 (1:1), the productivity reduced to 0.07 mg of hydrocarbons produced on average per mg of total proteins (Bagaeva, 2000). In this study, *Desulfovibrio* strains were cultured under an anaerobic atmosphere composed of CO_2 and H_2 (1:1), which may explain the low biogenic alkane yields but not the absence of biogenic alkane detection by the GC-MS within *Desulfovibrio* organic compound extracts.

In this study, alkane producing *Desulfovibrio* strains were shown to synthesise octadecane (C_{18}), nonadecane (C_{19}) and eicosane (C_{20}), with higher quantity of even - numbered carbon chain alkanes. Thus, the range of *Desulfovibrio* biogenic alkanes was much narrowed compared to previous estimates (Oppenheimer, 1965; Davis, 1968; Bagaeva & Chernova, 1994), which could have confounded non-metabolically derived hydrocarbons for biogenic alkanes. The use of isotopically labelled growth medium coupled with high - resolution mass spectrometry was proven successful for the discrimination of metabolites from compounds of non-biogenic origin. Metabolite profiling of *Synechococcus* sp. PCC 7002 was performed analysing extracts from unlabelled and labelled cyanobacterial cultures. Approximatively, 90 % of the compounds in *Synechococcus* cell extracts and growth media detected by mass spectrometry were not labelled and so classified as contamination (Baran *et al.,* 2010). In this study, alkanes

from C_{18} to C_{22} chain length detected within the organic compounds extracts of *D. giganteus* DSM 4370 and *D. alcoholivorans* NCIMB 12906 grown in deuterated sodium lactate medium were demonstrated to be unlabelled and thus deduced to do not have a biogenic origin. The detection of non-biogenic derived alkanes within organic compound extracts of *Desulfovibrio* may result from contamination during the microbial culture preparation or/and during the extraction process of cellular organic compounds.

The observation that all alkane producing *Desulfovibrio* strains showed the same biogenic alkane profile suggested that these strains share a unique alkane biosynthetic pathway. As all characterised alkane biosynthetic pathways to date derive from the fatty acid metabolic pathway (*cf*. Chapter 1 – Introduction; Figure 1.1), the fatty acid content of the *Desulfovibrio* spp. involved in this study was analysed by GC-MS. The biogenic fatty acids were extracted from the same cultures used to screen *Desulfovibrio* spp. for alkane production, as it has been reported that changes in culture conditions affect the fatty acid content (Taylor & Parkes, 1983; Vainshtein *et al.*, 1992).

The cellular fatty acid composition of *Desulfovibrio* spp. has been studied most notably for taxonomy (Makula & Finnerty, 1974; Ueki & Suto, 1979; Edlund *et al.,* 1985) and pathogenicity (Wolny *et al.,* 2011; Lodowska *et al.,* 2012). Due to fatty acid occurrence and prevalence in *Desulfovibrio* spp., the rare *iso*-branched saturated pentadecanoic acid (*iso*-$C_{15:0}$) and unsaturated heptadecenoic acid (*iso*-$C_{17:1}$) were designated as successful biomarkers for identification of new species (Boon *et al.,* 1977; Van Houten *et al.,* 2009; Cao *et al.,* 2016). Except *D. gigas* NCIMB 9332 for which *iso*-$C_{17:1}$ fatty acid was not detected within organic compound extracts, the *Desulfovibrio* spp. involved in this study were shown to produce *iso*-$C_{15:0}$ and *iso*-$C_{17:1}$. However, the absence of *iso*-$C_{17:1}$ fatty acid production by *D. gigas* NCIMB 9332 was previously reported (Ueki & Suto, 1979; Vainshtein *et al.*, 1992). These observations endorsed the reliability of the fatty acid analysis within this study.

Although classified among the single genus *Desulfovibrio*, the fatty acid content between *Desulfovibrio* spp. disclose heterogeneity, as previously reported and also as observed in this study (Vainshtein *et al.,* 1992; Dzierzewicz *et al.*, 1996). This heterogeneity illustrates the metabolic diversity among the *Desulfovibrio* genus. This could explain why not all the *Desulfovibrio* strains are capable of alkane synthesis. The fatty acid content of alkane producing *Desulfovibrio* was noticed to differ from the fatty acid content of non-alkane producing strains. With the assumption that the fatty acids are the metabolite precursors of alkanes in *Desulfovibrio*, the undetectability of biogenic

alkane by the GC-MS in *Desulfovibrio* cultures may be explained by the limited quantity or the absence of fatty acid precursors. Furthermore, this fatty acid heterogeneity suggested taxonomic disorder within the *Desulfovibrio* genus. *D. desulfuricans* NCIMB 8326 and *D. desulfuricans* NCIMB 8338 showed a similar fatty acid content, which differed from *D. desulfuricans* NCIMB 8307. *D. desulfuricans* NCIMB 8326 and *D. desulfuricans* NCIMB 8338 fatty acid content was composed of 25.3 % and 24.1 % on average of *ante*iso-pentadecanoic acid ($C_{15:0}$) respectively, and 1.2 % and 1.0 % on average of unsaturated hexadecenoic acid ($C_{16:1}$) respectively. Conversely, the fatty acid content of *D. desulfuricans* NCIMB 8307 contained 2.6 % on average of *ante*iso-pentadecanoic acid ($C_{15:0}$) and 15.8 % on average of unsaturated hexadecenoic acid ($C_{16:1}$). *Desulfovibrio* taxonomy has previously been reconsidered a number of times (Ueki & Suto, 1979; Voordouw, 1995; Morais-Silva *et al.*, 2014). Novel genera such as *Pseudodesulfovibrio* or *Halodesulfovibrio* have been proposed leading to the reclassification of some *Desulfovibrio* species (Cao *et al.,* 2016; Shivani *et al.*, 2017).

Importantly, the *Desulfovibrio* fatty acid content analysis showed that neither saturated carbon chain nonadecanoic ($C_{19:0}$) acid nor saturated carbon chain heneicosanoic acid ($C_{21:0}$) were detected by GC-MS, suggesting that *Desulfovibrio* spp. do not produce these two fatty acids. With the assumption that fatty acids are the metabolite precursors of alkanes in *Desulfovibrio,* the synthesis of octadecanes (C_{18}) or eicosanes (C_{20}) is unlikely to occur by decarbonylation or decarboxylation. This deduction supported the redaction of the hypothetical *Desulfovibrio* hydrocarbon biosynthetic pathway proposed by Bagaeva (Bagaeva, 1998 reviewed in Ladygina *et al*., 2006), which involves a decarboxylation step (*cf.* Chapter 1 - Introduction; Figure 1.4). Therefore, a novel hypothesis was proposed that alkane biosynthesis by *Desulfovibrio* involves a series of reduction reactions from fatty acids (Figure 3.16). In addition, the synthesis of nonadecanes was hypothesised to be the result of a methyl addition to octadecanes.

Consistent with this hypothetical alkane biosynthetic pathway in *Desulfovibrio*, if the proportion of saturated octadecanoic acid ($C_{18:0}$) is higher than saturated eicosanoic acid ($C_{20:0}$), higher quantity of octadecanes (C_{18}) and nonadecanes (C_{19}) than eicosanes (C_{20}) would be detected. This trend was observed for 5 of the 6 alkane producing *Desulfovibrio* strains investigated, and was particularly evident for *D. paquesii* DSM 16681 (Figure 3.17).

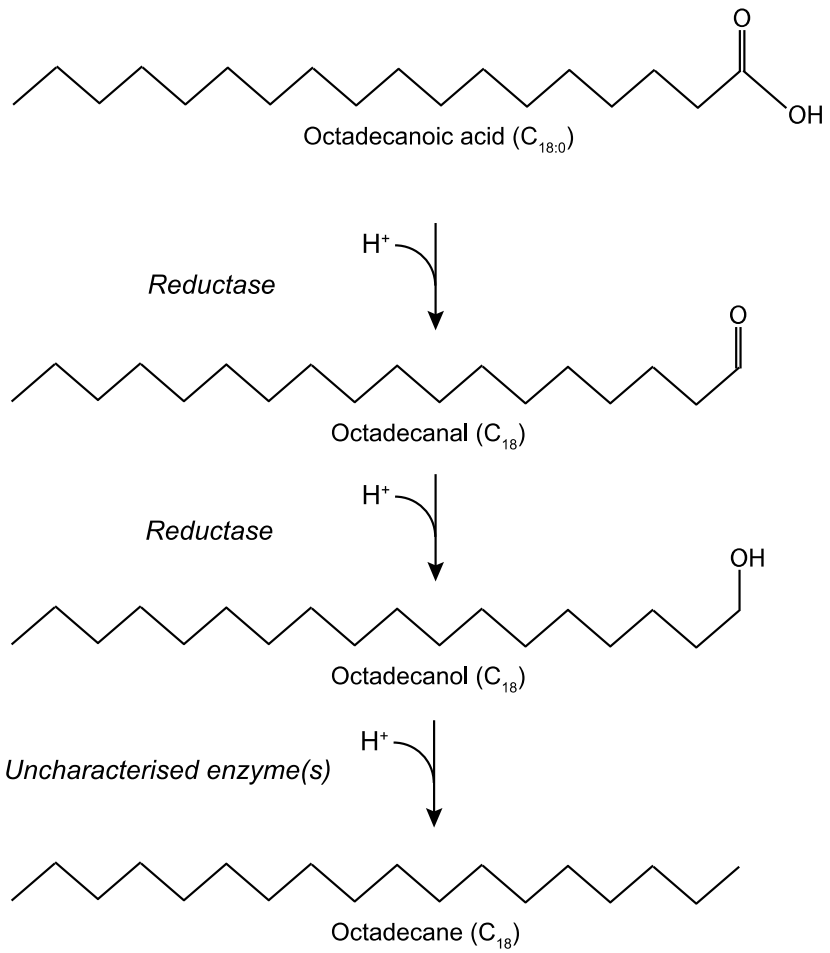


Figure 3.16. Hypothetical alkane biosynthetic pathway in *Desulfovibrio* involving a series of reduction reactions from fatty acids

The hypothetical *Desulfovibrio* alkane biosynthetic pathway of this study would follow a reductive hydrogenation route from even-numbered carbon chain fatty acids, *e.g.* the hypothetical biosynthetic pathway of octadecanes (C_{18}) from saturated straight carbon chain octadecanoic acids ($C_{18:0}$).

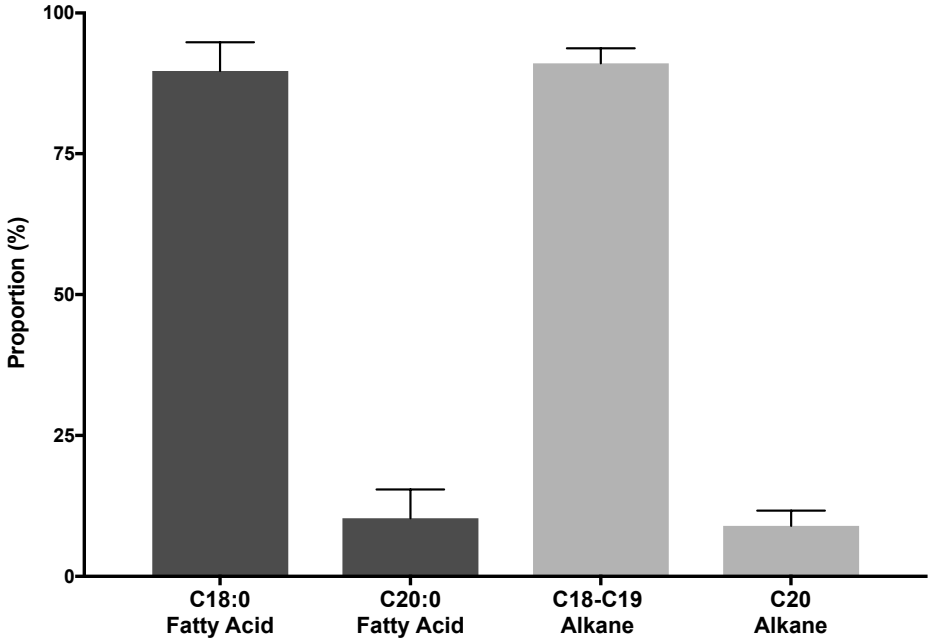


Figure 3.17. Proportion of biogenic saturated octadecanoic acid ($C_{18:0}$), saturated eicosanoic acid ($C_{20:0}$), octadecane (C_{18}), nonadecane (C_{19}) and eicosane (C_{20}) produced by *D. paquesii* DSM 16681

Mean proportion of saturated octadecanoic acid ($C_{18:0}$), saturated eicosanoic acid ($C_{20:0}$), octadecane (C_{18}), nonadecane (C_{19}) and eicosane (C_{20}) produced by the alkane producing *D. paquesii* DSM 16681 is shown. Mean proportions were calculated from triplicates (error bars shown).

This hypothetical and novel alkane biosynthetic pathway without carbon loss sparks a strong interest for petroleum replica production with higher carbon efficiency compared to the currently characterised alkane biosynthetic pathways. However, *Desulfovibrio* productivity to produce alkanes was assessed between 2.36 and 5.6 *µ*g of total alkanes produced on average per mg of total proteins, when cultured on a sodium lactate medium under an anaerobic atmosphere composed of CO_2 and H_2 (1:1) for 10 days. Consequently, native alkane yields from wild type *Desulfovibrio* spp. are too low to be cost competitive with petroleum derived fuels. Moreover, the complex growth medium, the slow growth and the anaerobic condition requirement of *Desulfovibrio* spp. impede their exploitation in an industrial context. The characterisation of *Desulfovibrio* alkane biosynthetic pathway is therefore necessary to enable the implementation of the pathway into another, more suitable host for an industrial exploitation. The characterisation of *Desulfovibrio* alkane biosynthetic pathway would also allow the optimisation of the pathway expression for higher alkane yields.

4. Comparative Genomic Analysis for *in silico* Identification of Candidate Molecular Basis involved in *Desulfovibrio* Alkane Biosynthesis

4.1. Introduction and Abstract

In this study, the characterisation of the putative hydrogenation pathway for alkane biosynthesis in *Desulfovibrio* was undertaken *via* a target-directed genome mining approach. To this end, the genomic DNA from nine *Desulfovibrio* strains was purified, sequenced, assembled *de novo* and annotated. Since the first method to sequence nucleic acids was developed (Sanger & Coulson, 1975), sequencing techniques and technologies have improved considerably both in rapidity and efficiency. Current sequencing technologies known as next generation sequencing technologies are ultra-high throughput, low-cost and widely available. Moreover, advances in bioinformatics techniques have enabled assembly algorithms and annotation software to be more accurate and user-friendly (Heather & Chain, 2016). The genomic DNA sequenced, assembled *de novo* and annotated in this study was purified from *D. desulfuricans* NCIMB 8326, *D. desulfuricans* NCIMB 8338, *D. gabonensis* NCIMB 10636, *D. gigas* NCIMB 9332, *D. marinus* DSM 18311, *D. paquesii* DSM 16681, *D. desulfuricans* NCIMB 8307, *D. giganteus* DSM 4370 and *D. alcoholivorans* NCIMB 12906.

The genomic DNA from alkane producing *Desulfovibrio* strains were first screened for alka(e)ne biosynthetic enzymes previously characterised from bacteria. No homologs of the characterised alka(e)ne biosynthetic enzymes was identified within the genomes of the alkane producing *Desulfovibrio* strains. This implied that alkane biosynthesis in *Desulfovibrio* was likely to be catalysed by enzymes which have not yet been reported.

The phylogenetic distribution of the *Desulfovibrio* spp. within this study based on 16S rRNA gene sequence was subsequently established. This phylogenetic distribution supported the hypothesis that the ability of *Desulfovibrio* spp. to produce alkanes was acquired by a common ancestral strain *via* horizontal gene transfer. Therefore, a novel hypothesis was proposed that the ability of *Desulfovibrio* spp. to

produce alkanes is due to the presence of genes encoding enzymes involved in alkane synthesis.

To identify the genes encoding alkane biosynthetic enzymes, a comparative genomic analysis of six alkane producing and four non-alkane producing *Desulfovibrio* genomes was subsequently undertaken. Comparative genomic analysis compares two or more genomes of different species or strains to discern the similarities and differences between the genomes (Wei *et al.,* 2002). Comparative genomic approaches have proven successful in providing insights into evolution (Bansal & Meyer, 2002), environmental adaptability (Lau *et al.*, 2011), pathogenicity and in elucidating metabolic pathways (Van Lanen & Shen, 2006). Characterisation of the cyanobacterial alkane biosynthetic pathway was previously achieved through a comparative genomic study that intersected genomes of alkane producing and non-alkane producing cyanobacterial strains (Schrimer *et al.,* 2010). Similarly, Parks *et al.* discovered the genetic basis of bacterial mercury methylation, identifying two genes that encode a corrinoid and iron-sulfur proteins in six known mercury methylating *Desulfovibrio* species but absent in non mercury methylating *Desulfovibrio* spp. (Parks *et al.,* 2013).

In this study, pan-genomes of the nine *de novo* assembled *Desulfovibrio* genomes within in this study, plus the model organism *D. vulgaris* strain Hildenborough genome were generated. The pan-genome of a species or strain cohort is defined by a core genome compiling genes shared by all genomes from the cohort and a dispensable genome constituted of partially shared and strain or species-specific genes (Tettelin *et al.,* 2005). Pan-genome analyses provided a framework for identifying hypothetical genes exclusively present in alkane producing *Desulfovibrio* strains. To target the hypothetical genes encoding enzymes potentially involved in alkane production among the hypothetical genes exclusively present in alkane producing *Desulfovibrio* strains, genes were annotated according to both protein sequence similarity and protein domain predictions. Moreover, hypothetical proteins exclusively present in alkane producing *Desulfovibrio* were screened for structural similarities with alkane-binding proteins. A list of 33 hypothetical proteins considered with high confidence to be exclusive to alkane producing *Desulfovibrio* strains was established. Two of the 33 hypothetical proteins were likely to share structural similarities with alkane-binding proteins.

4.2. Desulfovibrio Genome Sequencing

4.2.1. Genomic DNA Purification and Sequencing Paired-End Library Preparations

The genomic DNA from *D. desulfuricans* NCIMB 8326, *D. desulfuricans* NCIMB 8338, *D. gabonensis* NCIMB 10636, *D. gigas* NCIMB 9332, *D. marinus* DSM 18311, *D. paquesii* DSM 16681, *D. desulfuricans* NCIMB 8307, *D. giganteus* DSM 4370, *D. alcoholivorans* NCIMB 12906 and from an uninoculated sodium lactate medium sample (used as a negative control) was purified and quantified. *D. desulfuricans* NCIMB 8326, *D. gabonensis* NCIMB 10636, *D. gigas* NCIMB 9332, *D. marinus* DSM 18311 and *D. giganteus* DSM 4370 genomic DNA were sent to Shell Technology Centre for library preparation and sequencing. Genomic DNA library for *D. desulfuricans* NCIMB 8338, *D. paquesii* DSM 16681, *D. desulfuricans* NCIMB 8307, *D. alcoholivorans* NCIMB 12906 and for the uninoculated *Desulfovibrio* medium was prepared and validated by TapeStation (Agilent Technologies) at Exeter University, prior to being sent to Shell Technology Centre for sequencing (Table 4.1). DNA purified from uninoculated sodium lactate medium was sequenced in order to determine background from the DNA purification kit and the genomic DNA library kit.

Table 4.1. Quantification of purified DNA and prepared sequencing DNA libraries

Desulfovibrio **species and strain**	**Culture collection catalogue number**	**Purified DNA concentration (ng μl^{-1})**	**Average DNA fragment size (bp) between 150 bp and 1,000 bp**	**DNA fragment molarity (nmol l^{-1}) between 150 bp and 1,000 bp**
D. desulfuricans subsp. desulfuricans California27.137.5	NCIMB 8326	10.4	#	#
D. desulfuricans subsp. desulfuricans CubaHC29.130.4	NCIMB 8338	11.3	803	10.7
D. gabonensis SEBR 2840	DSM 10636	16.5	#	#
D. gigas	NCIMB 9332	11.6	#	#
D. marinus E-2	DSM 18311	17.4	#	#
D. paquesii SB1	DSM 16681	21.8	864	8.99
D. desulfuricans subsp. desulfuricans Essex6	NCIMB 8307	28.7	842	8.38
D. giganteus STg	DSM 4370	21.4	#	#
D. alcoholivorans	NCIMB 12906	23.5	775	9.05
Medium (negative control)	-	Too low to be detected	963	0.74

The concentration of purified genomic DNA for each of the sequenced *Desulfovibrio* strains and the negative control was determined using the Qubit® 2.0 Fluorometer (ThermoFisher Scientific Inc.). Prepared genomic DNA libraries were validated by TapeStation (Agilent Technologies). Libraries of the samples highlighted in grey were prepared at Shell Technology Centre, (#; validation data not provided).

4.2.2. Base Quality Analysis of the Sequencing Paired-End Reads

Desulfovibrio genomes and the DNA purified from uninoculated sodium lactate medium were sequenced using 300 bp paired-end fragments. The resulting DNA sequences, known as reads, were quality assessed using the program FastQC (Figure 4.1). FastQC calculates a Phred quality score for each nucleotide in the read sequence. A Phred score is logarithmically correlated to the probability that a base is called incorrectly by the sequencer. Higher Phred score values correspond to lower probabilities of an incorrect base call. The read DNA sequences were trimmed with a Phred score threshold of 20, using the program TrimGalore. Therefore, the bases with a probability of an incorrect call lower than 1 in 100 times were removed from the read sequences. Additionally, sequencing adapters from the 3' end of reads were removed using TrimGalore, to avoid mis-assembly of the genome. During DNA sequencing library preparation, genomic DNA was fragmented and adapters were ligated to 5' and 3' ends of the DNA fragments. Sequencing by synthesis is initiated at the read primer, annealed to DNA fragment complementary to the 5' adapter sequence and continues for 300 cycles. If the DNA fragment is shorter than 300 cycles, the 3' adapter sequence may be sequenced as part of the read, either fully or partly. Presence of 3' adapter sequence in reads might cause genome mis-assembly. After trimming, the read DNA sequences were again quality assessed using FastQC (Figure 4.1).

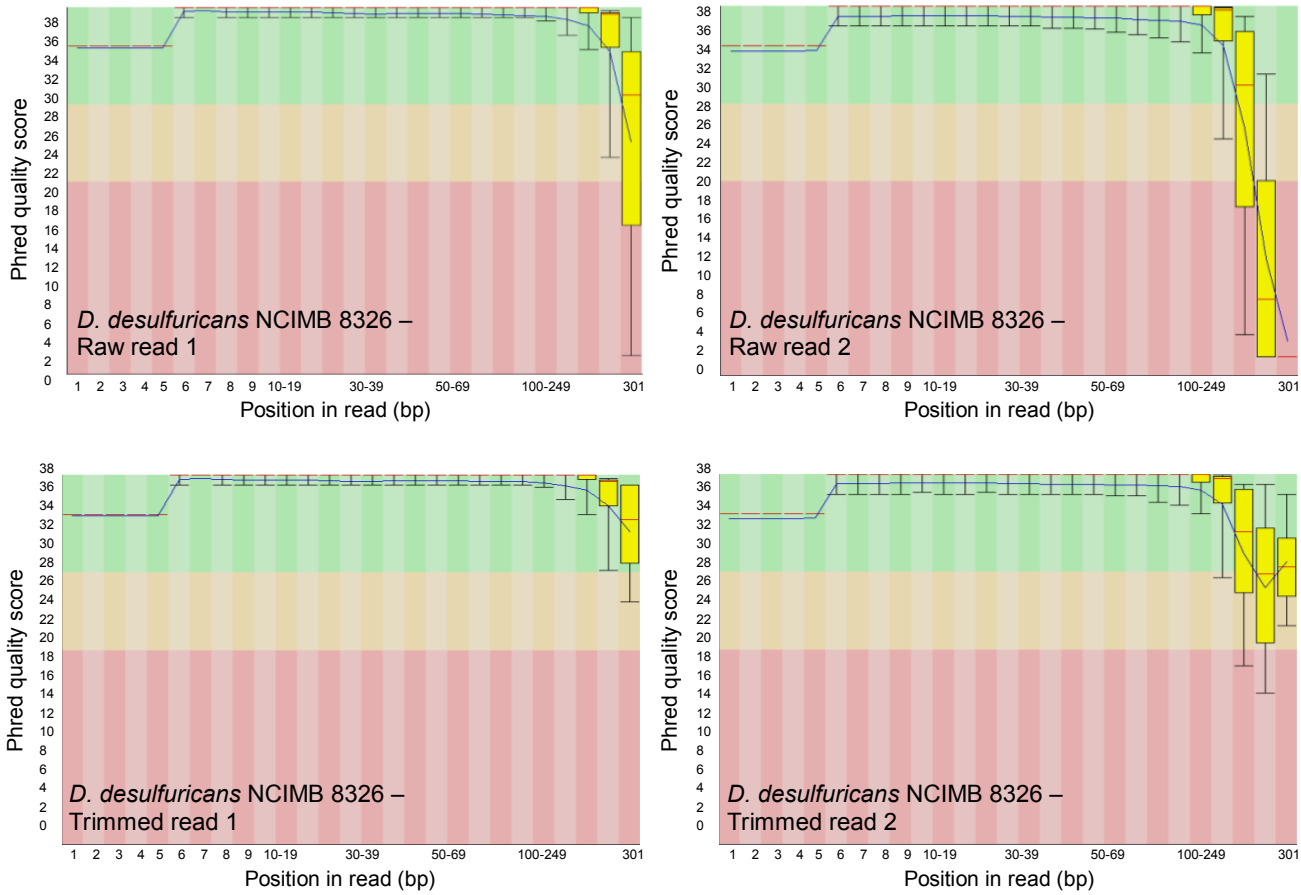


Figure 4.1. Quality of raw and trimmed DNA sequence reads from *D. desulfuricans* NCIMB 8326

D. desulfuricans NCIMB 8326 genome sample, 2,124,680 reads and 639,528,680 bases were quality assessed using FastQC. For the first set of raw reads (Read 1), 18,989,989 bases were removed from 699,352 reads as their quality Phred score was below the threshold of 20. For the second set of raw reads (read 2), 95,437,053 bases were removed from 566,446 reads as their quality Phred score was below the threshold of 20.

The parameter ‘paired’ of the program TrimGalore was also applied to the raw reads, to ensure having a set of paired-end reads with an appropriate length for genome assembly after quality and adapter trimming. Quality and adapter trimming can result in read length below 20 bp or in the trimming of the entire read sequence. The parameter ‘paired’ enables removing paired reads if at least one read sequence is shorter than the length threshold of 20 bp. The number of reads remaining after trimming and the percentage of raw reads lost are shown in table 4.2.

Table 4.2. Numbers of raw data reads from *Desulfovibrio* genome sequencing and reads after trimming with a length threshold of 20 bp

Desulfovibrio **species and strain**	**Culture collection catalogue number**	**Raw data reads**	**Reads after trimming**	**Percentage of read lost (%)**
D. desulfuricans subsp. desulfuricans California27.137.5	NCIMB 8326	2,124,680	2,111,705	0.61
D. desulfuricans subsp. desulfuricans CubaHC29.130.4	NCIMB 8338	1,183,191	1,180,334	0.24
D. gabonensis SEBR 2840	DSM 10636	3,357,530	3,202,146	4.63
D. gigas	NCIMB 9332	2,718,467	2,524,482	7.14
D. marinus E-2	DSM 18311	7,098,330	7,073,009	0.36
D. paquesii SB1	DSM 16681	1,220,384	1,217,652	0.22
D. desulfuricans subsp. desulfuricans Essex6	NCIMB 8307	1,209,013	1,205,832	0.26
D. giganteus STg	DSM 4370	8,916,204	8,890,565	0.29
D. alcoholivorans	NCIMB 12906	1,442,212	1,438,969	0.22
Medium (negative control)	-	57,600	56,676	1.6

Raw reads data from *Desulfovibrio* genomes and from the DNA purified from uninoculated sodium lactate medium were trimmed using the program TrimGalore with the parameter “paired”.

4.2.3. Genome *De novo* Assembly

Desulfovibrio genomes and the DNA purified from uninoculated sodium lactate medium were assembled *de novo* by the genome assembler SPAdes. The software SPAdes was chosen as it has been designed to assemble *de novo* reads from single cell datasets (Bankevich *et al.,* 2012). *De novo* assemblies of *Desulfovibrio* genomes were initially performed with trimmed DNA sequence reads (Table 4.3).

Table 4.3. SPAdes *de novo* assembly metrics for *Desulfovibrio* genome assemblies using trimmed reads as input data

***Desulfovibrio* species and strain**	**Culture collection catalogue number**	**Number of trimmed reads used for the assembly**	**Number of scaffolds**	**Total length (bp)**	**Maximum length (bp)**	**N50 (bp)**	**N90 (bp)**
D. desulfuricans subsp. desulfuricans California27.137.5	NCIMB 8326	2,111,705	199	4,480,823	1,377,393	899,711	183,021
D. desulfuricans subsp. desulfuricans CubaHC29.130.4	NCIMB 8338	1,180,334	880	5,303,499	518,274	246,947	21,739
D. gabonensis SEBR 2840	DSM 10636	3,202,146	393	4,600,288	1,070,174	671,878	70,299
D. gigas	NCIMB 9332	2,524,482	698	4,231,142	418,950	115,783	11,099
D. marinus E-2	DSM 18311	7,073,009	743	5,256,405	1,066,256	461,377	77,463
D. paquesii SB1	DSM 16681	1,217,652	329	4,438,662	524,972	136,413	34,000
D. desulfuricans subsp. desulfuricans Essex6	NCIMB 8307	1,205,832	205	3,478,983	727,925	220,725	61,982
D. giganteus STg	DSM 4370	8,890,565	694	5,679,305	429,673	82,244	15,107
D. alcoholivorans	NCIMB 12906	1,438,969	286	5,026,872	306,109	87,560	20,984
Medium (negative control)	-	56,676	3568	1,827,991	6,668	482	393

De novo assemblies of *Desulfovibrio* genomes were assessed by the number of scaffolds, the estimated total length of the genome ('Total length'), the length of the longest scaffold ('Maximum length'), the values of N50 and N90. The N50 value is the length of the shortest scaffold such that 50 % of the genome is contained in scaffolds of N50 or longer. The N90 value is the length of the shortest scaffold such that 90 % of the genome is contained in scaffolds of N90 or longer.

The *Desulfovibrio* genome assemblies by SPAdes using trimmed DNA sequence reads were highly fragmented compared to publically available bacterial genomes. Despite achieving a high coverage, the short reads of 300 bp do not span long repeat regions, which consequently generates gaps in the assembly and causes read assembly errors (Ricker *et al.*, 2012). Longer reads would cover long repeat regions to overcome fragmented and incorrect genome assembly (Treangen & Salzberg, 2013). Pairs of trimmed DNA sequences which had overlapping regions were merged using the program FLASH, to generate single longer reads. For example, 1,997,265 reads, i.e. 94.6 % of the 2,111,705 trimmed reads from *D. desulfuricans* NCIMB 8326 genome were merged and extended. *Desulfovibrio* genomes and the DNA purified from uninoculated sodium lactate medium were subsequently assembled *de novo* from "flashed" trimmed DNA sequence reads using the software SPAdes (Table 4.4).

Table 4.4. SPAdes *de novo* assembly metrics for *Desulfovibrio* genome assemblies using "flashed" trimmed reads as input data

***Desulfovibrio* species and strains**	**Culture collection catalogue number**	**Number of trimmed reads used for the assembly**	**Number of trimmed reads merged by FLASH**	**Number of scaffolds**	**Total length (bp)**	**Maximum length (bp)**	**N50 (bp)**	**N90 (bp)**
D. desulfuricans subsp. desulfuricans California27.137.5	NCIMB 8326	2,111,705	1,997,265	21	4,396,671	1,377,393	899,711	183,491
D. desulfuricans subsp. desulfuricans CubaHC29.130.4	NCIMB 8338	1,180,334	629,658	189	4,999,065	518,274	333,004	64,257
D. gabonensis SEBR 2840	DSM 10636	3,202,146	3,046,668	44	4,431,658	1,070,329	671,878	114,740
D. gigas	NCIMB 9332	2,524,482	2,412,663	117	3,944,774	403,377	115,094	28,476
D. marinus E-2	DSM 18311	7,073,009	4,591,209	69	4,911,909	1,066,256	497,016	153,458
D. paquesii SB1	DSM 16681	1,217,652	780,823	106	4,339,072	520,356	134,799	39,120
D. desulfuricans subsp. desulfuricans Essex6	NCIMB 8307	1,205,832	397,426	52	3,411,800	727,925	220,725	68,047
D. giganteus STg	DSM 4370	8,890,565	6,136,565	163	5,412,799	429,673	89,263	25,687
D. alcoholivorans	NCIMB 12906	1,438,969	932,465	127	4,959,721	382,603	124,077	23,442
Medium (negative control)	-	56,676	53,841	644	516,465	3,973	767	638

De novo assemblies of *Desulfovibrio* genomes were assessed by the number of scaffolds, the estimated total length of the genome ('Total length'), the length of the longest scaffold ('Maximum length'), the values of N50 and N90. The N50 value is the length of the shortest scaffold such that 50 % of the genome is contained in scaffolds of N50 or longer. The N90 value is the length of the shortest scaffold such that 90 % of the genome is contained in scaffolds of N90 or longer.

Using longer, merged DNA sequence reads for *Desulfovibrio* genome assemblies improved the contiguity of the assemblies. For example, *D. desulfuricans* NCIMB 8326 genome was fragmented into 199 scaffolds when assembled with trimmed DNA sequence reads, whereas it was only fragmented into 21 scaffolds when assembled with "flashed" trimmed DNA sequence reads.

4.2.4. *De novo* Assembly Quality Assessment

The quality of *Desulfovibrio* genome assemblies generated with "flashed" trimmed reads was assessed using the program QUAST. QUAST evaluates a series of metrics used for assembly quality assessment (Gurevich *et al.,* 2013). The program QUAST evaluated a total genome length ranging between 3.3 Mb and 5.3 Mb and a GC content varying from 45 % to 65 % for the *Desulfovibrio* assemblies within this study. According to the GenBank database in 2017, *Desulfovibrio* genomes range in size between 2.6 Mb and 5.7 Mb, and have a GC content varying from 41.8 % to 68 %. The total genome length and the GC content evaluated by QUAST for the *Desulfovibrio* assemblies therefore fell within the expected genome size and GC content ranges for the genus, suggesting no contamination. Moreover, the GC content distribution determined by QUAST for the *Desulfovibrio* assembled genomes followed only one Gaussian curve, suggesting no genome contamination.

De novo assembled Desulfovibrio genome contiguity

The quality of a genome assembly is partially assessed by the contiguity of the assembly. The higher an assembly contiguity is, the less the assembly is fragmented. Therefore, the probability that the assembly contains the same number of base pairs as the real biological genome is higher. The N50 is one of the metrics evaluating the contiguity of an assembly. Longer N50 lengths represent higher assembly contiguity.

N50 values of the constructed *Desulfovibrio* assemblies within this study were compared to the N50 values of the 30 *Desulfovibrio* assemblies at a level of contigs or scaffolds available in GenBank (Figure 4.2). *D. desulfuricans* NCIMB 8326 assembly within this study showed the longest N50 length of 899,711 bp. Four of the nine *Desulfovibrio* assemblies within this study showed a higher contiguity than 60 % of the other *Desulfovibrio* assemblies from this study and GenBank. None of the N50 length

of the *Desulfovibrio* assemblies from this study fell within the lowest 20 % of the N50 value for *Desulfovibrio* assemblies.

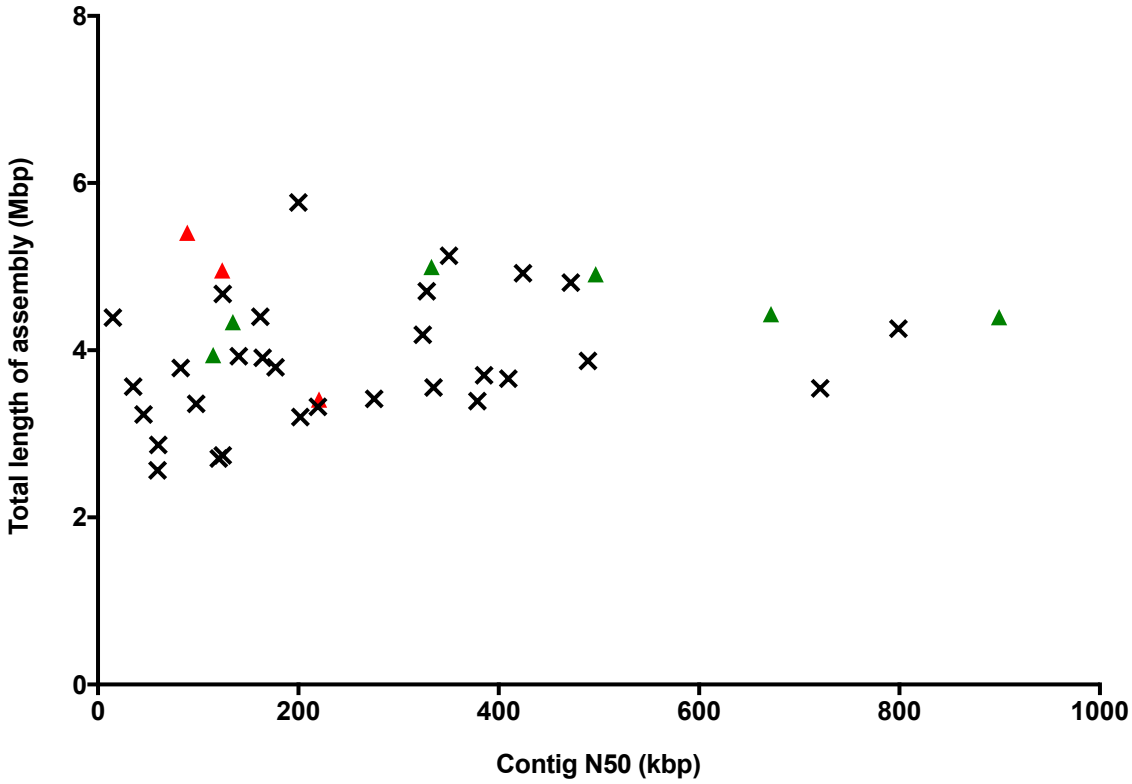


Figure 4.2. Assembly lengths and contiguity levels of *Desulfovibrio* genome assemblies

The *Desulfovibrio* genome assemblies generated in this study are displayed by a triangle. A green triangle represents the genome assembly of an alkane producing *Desulfovibrio* strain. A red triangle represents the genome assembly of a non-alkane producing *Desulfovibrio* strain. The *Desulfovibrio* genome assemblies at a level of contigs or scaffolds from GenBank are displayed by a black cross. The total lengths and N50 values were calculated using the program QUAST.

De novo assembled Desulfovibrio genome accuracy

Another parameter assessing the quality of an assembly is its accuracy. An accurate genome assembly has identical bases in the identical order to the real biological genome. The genome per-base coverage provides an estimation of the accuracy of an assembled genome. The genome per-base coverage is the average number of times that a base in a genome is sequenced. Analysing the per-base coverage of *Desulfovibrio* genome (Table 4.5), a base was sequenced at least 68 times on average ($68 \leq x \leq 678$), which gave a high degree of confidence about the correctness of read sequences used for the *Desulfovibrio* assemblies.

Table 4.5. Per-base coverage of *Desulfovibrio* genomes

Desulfovibrio **species and strains**	**Culture collection catalogue number**	**Per-base coverage**
D. desulfuricans subsp. desulfuricans California27.137.5	NCIMB 8326	240.5
D. desulfuricans subsp. desulfuricans CubaHC29.130.4	NCIMB 8338	71.4
D. gabonensis SEBR 2840	DSM 10636	316.4
D. gigas	NCIMB 9332	231.7
D. marinus E-2	DSM 18311	565.7
D. paquesii SB1	DSM 16681	68.3
D. desulfuricans subsp. desulfuricans Essex6	NCIMB 8307	112.8
D. giganteus STg	DSM 4370	678.3
D. alcoholivorans	NCIMB 12906	79.1

Assessment of the per-base coverage of *Desulfovibrio* genomes was based on the alignment of raw paired-end reads used for a genome assembly to the assembly, using the alignment software Bowtie 2. The software Qualimap 2 was then used to calculate the per-base coverage for each genome.

D. gigas NCIMB 9332 and *D. desulfuricans* NCIMB 8307 have published genomes in GenBank. The accuracy of assembled *D. gigas* NCIMB 9332 and *D. desulfuricans* NCIMB 8307 genomes within this study was therefore assessed by comparison to the published genomes, using the wrapper dnadiff from NUCmer. The wrapper dnadiff aligns two assemblies of a same strain (Kurtz *et al.*, 2004). Two incomplete genome assemblies and one plasmid sequence are available in GenBank for *D. gigas* NCIMB 9332. The assembled *D. gigas* NCIMB 9332 genome sequence within this study was aligned to the published genome sequence with the highest level of assembly (GenBank accession number CP006585.1; Table 4.6).

The assembled *D. gigas* NCIMB 9332 genome from this study aligned to the entire published genome sequence (GenBank CP006585.1), with 93.37 % identity. 25 scaffolds of the assembled *D. gigas* NCIMB 9332 genome from this study did not align to the published genome sequence with the highest level of assembly (GenBank CP006585.1). These 25 scaffolds were then aligned to the scaffolds of the second published assembly (GenBank assembly accession GCA_000429285.1) and to the plasmid sequence (GenBank accession number CP006586.1) published for *D. gigas* NCIMB 9332. Only 2 of the 25 scaffolds of assembled *D. gigas* NCIMB 9332 genome aligned to the scaffolds of the second published assembly (GenBank GCA_000429285.1) and none aligned to the plasmid sequence (GenBank CP006586.1). Therefore, in this study, 23 hypothetical scaffolds, previously unreported and representing 21,029 bp, were found to complement *D. gigas* NCIMB 9332 genome.

Table 4.6. dnadiff metrics from the alignment of the assembled *D. gigas* NCIMB 9332 genome sequence from this study to the published *D. gigas* NCIMB 9332 genome sequence (GenBank CP006585.1)

dnadiff statistics	**Genome reference: *D. gigas* NCIMB 9332 (GenBank CP006585.1)**	**Genome query: *D. gigas* NCIMB 9332 from this study**
Total number of input sequences	1	117
Number of input sequences with at least one alignment	1 (100.00 %)	92 (78.63 %)
Number of input sequences with no alignment	0 (0.00 %)	25 (21.73 %)
Total number of bases in the input sequences	3,693,999	3,944,774
Total number of bases contained within an alignment	3,669,038 (99.32 %)	3,683,356 (93.37 %)
Total number of unaligned bases	24,961 (0.68 %)	261,418 (6.63 %)

The assembled *D. desulfuricans* NCIMB 8307 genome sequence from this study was aligned to the published *D. desulfuricans* NCIMB 8307 genome sequence (GenBank accession number ATUZ00000000.1; Table 4.7).

The assembled *D. desulfuricans* NCIMB 8307 genome from this study aligned to the entire published genome sequence, with 99.49 % identity. 20 scaffolds of assembled *D. desulfuricans* NCIMB 8307 genome from this study did not align to the published genome sequence. The published genome of *D. desulfuricans* NCIMB 8307 is incomplete. Therefore, in this study, 20 hypothetical scaffolds, previously unreported and representing 17,319 bp, were found to complement *D. desulfuricans* NCIMB 8307 genome.

Table 4.7. dnadiff metrics from the alignment of the assembled *D. desulfuricans* NCIMB 8307 genome sequence from this study to the published *D. desulfuricans* NCIMB 8307 genome sequence (GenBank ATUZ00000000.1)

dnadiff statistics	**Genome reference: *D. desulfuricans* DSM 642 (GenBank ATUZ00000000.1)**	**Genome query: *D. desulfuricans* NCIMB 8307 from this study**
Total number of input sequences	20	52
Number of input sequences with at least one alignment	20 (100.00%)	32 (61.54%)
Number of input sequences with no alignment	0 (0.00%)	20 (38.46 %)
Total number of bases in the input sequences	3,391,683	3,411,800
Total number of bases contained within an alignment	3,391,572 (100.00 %)	3,394,481 (99.49 %)
Total number of unaligned bases	111 (0 %)	17,319 (0.51 %)

With the exception of *D. gigas* NCIMB 9332 and *D. desulfuricans* NCIMB 8307, the other *Desulfovibrio* strains within this study do not have a published genome for comparison. To evaluate the accuracy of the *Desulfovibrio* genome assemblies without the requirement of a published genome, the program REAPR was used. REAPR assesses the accuracy of an assembly by analysing the fragment coverage and the fragment coverage distribution (FCD) at each base of an assembly sequence. Initially, REAPR maps paired-end reads to the genome assembly. REAPR identifies a fragment coverage error for a base when less than 5 reads align the base perfectly (*i.e.* along the entire read length) and uniquely (*i.e.* at only one position in the genome assembly). REAPR identifies a FCD error for a base when the difference between the observed FCD and the ideal FCD for a base is higher than the FCD error threshold calculated by REAPR for the assembly. As output, REAPR provides the percentage of error free bases of the analysed assembly, corresponding to regions of the assembly that are extremely likely to be correct (Hunt *et al.*, 2013). According to the REAPR analysis, *de novo* assembled *Desulfovibrio* genomes within this study were found to be at least 91.63 % correct.

De novo assembled Desulfovibrio genome completeness

Finally, the completeness of the *Desulfovibrio* genome assemblies within this study was evaluated. A genome assembly is defined as complete if it contains all of the genes present in the real biological genome. The completeness evaluation of the assembled *Desulfovibrio* genomes was critical in this study, as the aim was to identify the genic sequences encoding the enzymes involved in *Desulfovibrio* alkane biosynthetic pathway. Firstly, assembled *Desulfovibrio* genomes were annotated using the software Prokka. The numbers of predicted genes and proteins from the assembled *Desulfovibrio* genomes within this study and from the complete *Desulfovibrio* genomes available in GenBank were compared (Figure 4.3).

The numbers of predicted genes and proteins from the assembled *Desulfovibrio* genomes within this study were similar to the number of predicted genes and proteins from the complete *Desulfovibrio* genomes available in GenBank. This observation suggested that the *Desulfovibrio* genome assemblies involved in this study were complete. Moreover, this observation endorsed the absence of contamination in *Desulfovibrio* genome assemblies.

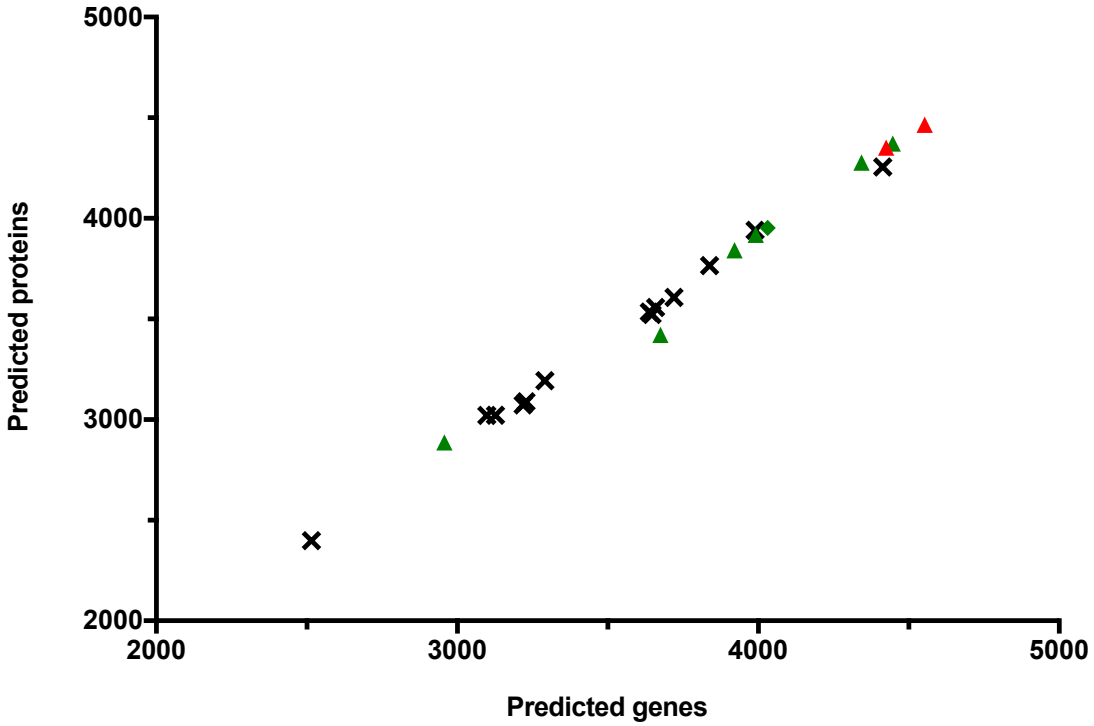


Figure 4.3. Numbers of predicted proteins and genes in assembled *Desulfovibrio* genomes

The assembled *Desulfovibrio* genomes from this study are displayed by a triangle. A green triangle represents the assembled genome of an alkane producing *Desulfovibrio* strain. A red triangle represents the assembled genome of a non-alkane producing *Desulfovibrio* strain. The complete *Desulfovibrio* genomes from GenBank are displayed by a black cross.

The tool BUSCO was secondly used to confirm the completeness of the *Desulfovibrio* assemblies. BUSCO provides quantitative assessment of the completeness of genome assemblies in terms of gene content. For prokaryotic genomes, BUSCO screens the genome assemblies for a set of 40 genes, which were selected from the orthologs database OrthoDB and designated as universal markers to be single-copy orthologs in at least 90 % of the species for the clade considered (Simão *et al.*, 2015). The BUSCO metrics for *Desulfovibrio* genome assemblies are shown in table 4.8.

The totality of the BUSCO universal marker genes were entirely recovered for all the *Desulfovibrio* genome assemblies within this study, indicating that the assemblies were likely to be complete. The maximum number of BUSCO universal marker genes that occurred complete in duplicate copies was one per genome assembly.

Table 4.8. BUSCO metrics for *Desulfovibrio* genome assemblies

Desulfovibrio **species and strains**	**Culture collection catalogue number**	**"Complete" universal marker genes**	**"Complete" universal marker genes in single copy**	**"Complete" universal marker genes in multiple copy**
D. desulfuricans subsp. desulfuricans California27.137.5	NCIMB 8326	40	39	1
D. desulfuricans subsp. desulfuricans CubaHC29.130.4	NCIMB 8338	40	39	1
D. gabonensis SEBR 2840	DSM 10636	40	39	1
D. gigas	NCIMB 9332	40	40	0
D. marinus E-2	DSM 18311	40	39	1
D. paquesii SB1	DSM 16681	40	40	0
D. desulfuricans subsp. desulfuricans Essex6	NCIMB 8307	40	39	1
D. giganteus STg	DSM 4370	40	40	0
D. alcoholivorans	NCIMB 12906	40	40	0

The recovered genes defined as 'complete' have a sequence length within two standard deviations of the BUSCO group mean length (i.e. within 95 % expectation).

In summary, *de novo* assembled *Desulfovibrio* genomes from this study disclosed minimal contamination, correct contiguity and accuracy. Moreover, *de novo* assembled *Desulfovibrio* genomes from this study were likely to be complete, which is a decisive assembly quality for this investigation.

4.3. Bioinformatics Analysis and Pathway Mining

4.3.1. Alkane Producing *Desulfovibrio* Screening for Characterised Alka(e)ne Biosynthetic Enzymes from Bacteria

To identify the molecular basis of *Desulfovibrio* alkane production, alkane producing *Desulfovibrio* strains were screened for alka(e)ne biosynthetic enzymes previously characterised from bacteria.

Screening for characterised alka(e)ne biosynthetic enzymes by sequence similarity

Bacterial alka(e)ne biosynthetic enzyme sequences were aligned to the alkane producing *Desulfovibrio* assembled genome sequences. Alignment of bacterial alka(e)ne biosynthetic enzyme sequences to the alkane producing *Desulfovibrio* assembled genomes was chosen over the alignment of characterised alka(e)ne biosynthetic enzyme sequences to predicted proteins by Prokka from alkane producing *Desulfovibrio*. Consequently, the search for characterised alka(e)ne biosynthetic enzymes in the alkane producing *Desulfovibrio* did not rely on the program Prokka predictions. According to the blast metrics of the best alignment for each alka(e)ne biosynthetic enzymes (Table 4.9), none of the translated characterised alka(e)ne biosynthetic enzyme sequences aligned to the alkane producing *Desulfovibrio* genomes with an E-value of 0 and 100% identity. These stringent E-value and percentage of identity were chosen to identify any identical copies of the bacterial alka(e)ne biosynthetic enzymes in alkane producing *Desulfovibrio*. Reliable prediction of functional similarities between proteins in different organisms is based on identification of orthologues, defined as the same protein in different organisms (Pearson, 2013). Alignment results of the bacterial alka(e)ne biosynthetic enzymes to the alkane producing *Desulfovibrio* assembled genomes showed that none of these enzymes have an identical copy within the alkane producing *Desulfovibrio*.

Table 4.9. Greatest similarity of the characterised alka(e)ne biosynthetic enzyme sequence to alkane producing *Desulfovibrio* genomes

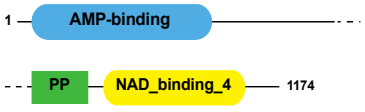
Alka(e)ne biosynthetic enzyme sequence	Alkane producing *Desulfovibrio* genome database	E-value	Bit score	Sequence percentage identity (%)
Acyl-ACP reductase (AAR)	No hit	-	-	-
LuxC from the protein complex fatty acid reductase (FAR)	*D. gabonensis* NCIMB 10636	$9e^{-15}$	74.7	24.65
LuxD from the protein complex fatty acid reductase (FAR)	No hit	-	-	-
LuxE from the protein complex fatty acid reductase (FAR)	No hit	-	-	-
OleA from the protein complex OleABCD	*D. marinus* DSM 18311	$5e^{-15}$	74.3	26.69
	D. desulfuricans NCIMB 8338	$5e^{-15}$	74.3	26.69
OleB from the protein complex OleABCD	*D. gabonensis* NCIMB 10636	$2e^{-07}$	50.1	32.02
OleC from the protein complex OleABCD	*D. gabonensis* NCIMB 10636	$2e^{-16}$	81.3	23.03
OleD from the protein complex OleABCD	*D. marinus* DSM 18311	$2e^{-13}$	69.7	27.62
Carboxylic acid reductase (CAR)	*D. desulfuricans* NCIMB 8326	$3e^{-15}$	78.6	22.51
Aldehyde deformylating oxygenase (ADO)	No hit	-	-	-
Fatty-acid Decarboxylase $OleT_{JE}$ P450 enzyme	No hit	-	-	-
Olefin Synthase (Ols)	*D. gigas* NCIMB 9332	$2e^{-118}$	420	34.16
UndA	No hit	-	-	-
UndB	No hit	-	-	-

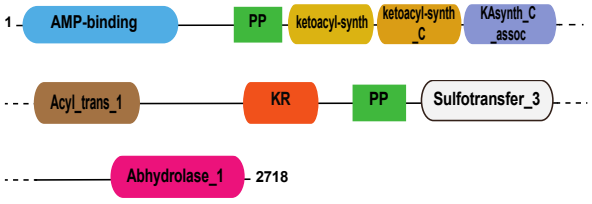
Bacterial enzymes that have previously been characterised in alka(e)ne biosynthetic pathways were reverse-translated and aligned to the alkane producing *Desulfovibrio* genomes generated in this study. The lowest E-value hit for each alka(e)ne biosynthetic enzyme is reported.

Screening for characterised alka(e)ne biosynthetic enzymes by protein domain homology

Inferring functional similarities can also be based on structural similarities, such as identical protein domain content, shared by proteins (Forslund & Sonnhammer, 2008). The previously characterised bacterial alka(e)ne biosynthetic enzymes were therefore screened for protein domains. Except the acyl-ACP reductase (AAR) in which no protein domain was predicted, the enzymes contained one to ten protein domains (Table 4.10).

Table 4.10. Bacterial alka(e)ne biosynthetic enzyme predicted protein domains

Bacterial alka(e)ne biosynthetic enzymes	Predicted domain	Predicted domain description	Domain independent E-value
Acyl-ACP reductase (AAR) 	No hit	-	-
LuxC from the protein complex fatty acid reductase (FAR) 	LuxC (PF05893)	Acyl-CoA reductase, LuxC	$4.0e^{-140}$
LuxD from the protein complex fatty acid reductase (FAR) 	Acyl transferase 2 (PF02273)	Acyl transferase	$4.6e^{-167}$
LuxE from the protein complex fatty acid reductase (FAR) 	LuxE (PF04443)	Acyl-protein synthetase, LuxE	$4.6e^{-50}$
OleA from the protein complex OleABCD 	Thiolase_N (PF00108)	Thiolase, N-terminal domain	$2.4e^{-06}$
	ACP_syn_III_C (PF08541)	3-Oxoacyl-[acyl-carrier-protein (ACP)] synthase III C terminal	$1.6e^{-16}$
OleB from the protein complex OleABCD 	Abhydrolase_1 (PF00561)	Alpha/beta hydrolase fold	$7.6e^{-22}$
OleC from the protein complex OleABCD 	AMP-binding (PF00501)	AMP-binding enzyme	$8.3e^{-56}$
OleD from the protein complex OleABCD 	3Beta_HSD (PF01073)	3-beta hydroxysteroid dehydrogenase/ isomerase family	$2.7e^{-57}$
Carboxylic acid reductase (CAR) 	AMP-binding (PF00501)	AMP-binding enzyme	$8.9e^{-65}$
	PP-binding (PF00550)	Phosphopantetheine attachment site	$3.1e^{-10}$
	NAD_binding_4 (PF07993)	Male sterility protein	$6.2e^{-54}$

Aldehyde Deformylating Oxygenase (ADO) 	Ald_deCOase (PF11266)	Long-chain fatty aldehyde decarbonylase	$1.3e^{-121}$
Fatty-acid Decarboxylase $OleT_{JE}$ P450 enzyme 	p450 (PF00067)	Cytochrome P450	$3.5e^{-15}$
Olefin Synthase (Ols) 	AMP-binding (PF00501)	AMP-binding enzyme	$1.7e^{-86}$
	PP-binding (PF00550)	Phosphopantetheine attachment site	$6.9e^{-14}$
	ketoacyl-synt (PF00109)	Beta-ketoacyl synthase, N-terminal domain	$6.4e^{-85}$
	ketoacyl-synt_C (PF02801)	Beta-ketoacyl synthase, C-terminal domain	$6.9e^{-46}$
	KAsynt_C_asoc (PF16197)	Ketoacyl-synthetase C-terminal extension	$2.1e^{-14}$
	Acyl_transf_1 (PF00698)	Acyl transferase domain	$7.0e^{-46}$
	KR (PF08659)	Ketoreductase	$2.6e^{-56}$
	PP-binding (PF00550)	Phosphopantetheine attachment site	$3.6e^{-11}$
	Sulfotransfer_3 (PF13469)	Sulfotransferase family	$3.9e^{-35}$
	Abhydrolase_1 (PF00561)	Alpha/beta hydrolase fold	$7.1e^{-24}$
UndA 	Haem_oxygenas_2 (PF14518)	Iron-containing redox enzyme	$1.1e^{-12}$
UndB 	FA-desaturase (PF00487)	Fatty cid desaturase	$6.1e^{-18}$

The bacterial alka(e)ne biosynthetic enzymes were screened for domains by alignment to the profile-HMM database Pfam. Domain independent E-values are reported for each predicted domain. Domain independent E-values estimate the probability that a sequence matches one specific domain in the database search, hypothesising that the sequence only hits this specific domain.

Alkane producing *Desulfovibrio* proteomes, generated in this study from annotation of the assembled genomes, were screened for the Hidden Markov model (HMM) profile of the 20 protein domains identified within the characterised alka(e)ne biosynthetic enzymes. HMM profiles are statistical models, scoring for residues and for evolutionary events such as deletion and insertion at any position in a sequence. Therefore, the HMM profile of a sequence informs which residues are the most likely and their degree of conservation for each position in the sequence (Eddy, 1996). HMM profiles enable identification of highly conserved sequence domains which are likely to be crucial to the functionality of the molecule. Screening alkane producing *Desulfovibrio* proteomes for HMM profiles of protein domains found in characterised alka(e)ne biosynthetic enzymes would identify proteins from alkane producing *Desulfovibrio* proteomes that are most likely to share similarity in function with the characterised alka(e)ne biosynthetic enzymes (Appendix 3).

Homologous *Ald_deCOase* (PF11266) and *p450* (PF00067) domains were not detected in alkane producing *Desulfovibrio* proteomes, implying that alkane producing *Desulfovibrio* do not have proteins with similar function to neither ADO nor $OleT_{JE}$ P_{450}.

Homologous *LuxC* (PF05893) and *LuxE* (PF04443) domains were identified in *D. gabonensis* NCIMB 10636 and *D. desulfuricans* NCIMB 8326. However, the *acyl_transferase_2* (PF02273) protein domain found in LuxD was not detected in alkane producing *Desulfovibrio* proteomes. The subunit LuxD was proven to be a crucial component for the FAR complex (Howard *et al.,* 2013). Therefore, a protein complex with similar function to the FAR complex is unlikely to be present in alkane producing *Desulfovibrio*.

Homologous CAR and OleABCD protein domains were identified in all the alkane producing *Desulfovibrio* proteomes. However, the protein domains identified for CAR and for the OleABCD protein complex have a generic catalytic function and therefore are present in wide range of enzymes. Identification of proteins from alkane producing *Desulfovibrio* proteomes that are most likely to share similarity in function with either CAR or the OleABCD protein complex by protein domain prediction is thus irrelevant.

Except *KAsynt_C_assoc* (PF16197) and *Sulfotransfer_3* (PF13469) domains, protein domains constituting Ols have homologs in all the alkane producing *Desulfovibrio* proteomes. The sulfotransferase domain (*Sulfotransfer_3* (PF13469) is

one of the specific protein domains of Ols (Coates *et al.*, 2014). This protein domain is predicted to catalyse the activation of the hydroxyl group of an acyl-substrate via sulphonation, which is required for the final reactions to α-olefins (Mendez-perez *et al.*, 2011). Thus the alkane biosynthesis in *Desulfovibrio* is not catalysed by an enzyme with similar function to Ols.

Homologous *Haem_oxygenas_2 (PF14518)* domain was detected only in *D. gigas* NCIMB 9332 and homologous *FA-desaturase (PF00487)* domain was not detected in alkane producing *Desulfovibrio* proteomes, implying that alkane producing *Desulfovibrio* do not have proteins with similar function to neither UndA nor UndB.

In summary, alkane biosynthesis in *Desulfovibrio* is likely to be catalysed by enzymes which have not yet been reported, according to existing data.

4.3.2. Phylogenetic Distribution of the Alkane Producing *Desulfovibrio* strains

Phylogenetic distribution of the alkane producing *Desulfovibrio* strains was established based on 16S rRNA gene sequence. The 16S rRNA gene sequence from the annotated *Desulfovibrio* genomes within this study was identified. Only one 16S rRNA gene sequence was predicted for each *Desulfovibrio* genome assembly, suggesting no genome contamination. The 16S rRNA phylogenetic tree for the genus *Desulfovibrio* was generated with alignment of the 16S rRNA gene sequences of the 9 *Desulfovibrio* strains within this study, 59 *Desulfovibrio* type strains and the *Desulfobacter postgatei* DSM 2034 type strain from the Ribosomal Database Project (RDP) (Figure 4.4). The outgroup taxon *Desulfobacter postgatei* DSM 2034 type strain was used as root for the phylogenetic tree. The phylogenetic tree was built using the maximum likelihood method.

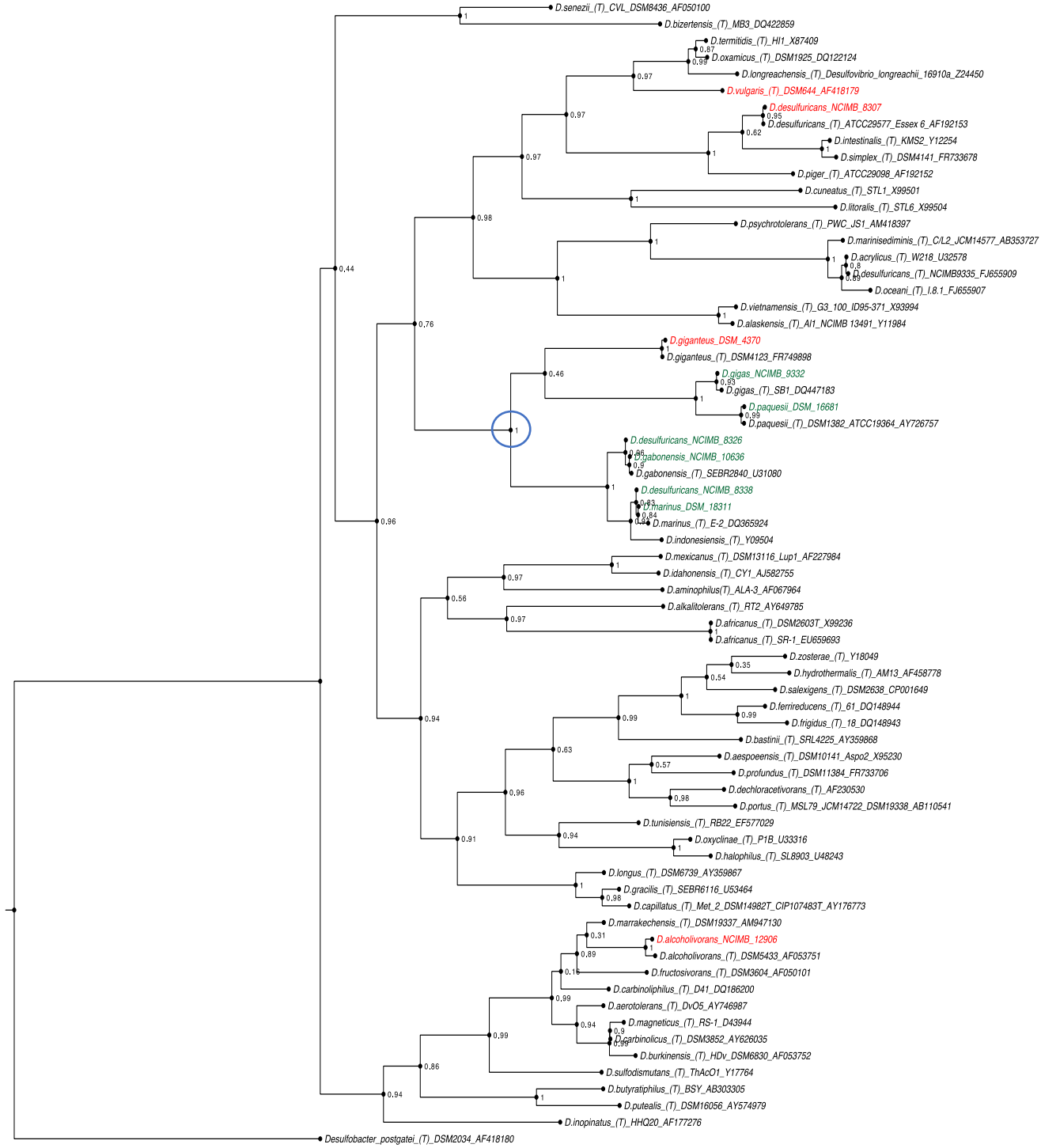


Figure 4.4. Phylogenetic tree of the genus *Desulfovibrio* based on 16S rRNA gene sequence

The tree was generated using the alignment of 16S rRNA gene sequences from 59 *Desulfovibrio* type strains, the 9 *Desulfovibrio* strains within the study and *Desulfobacter postgatei* DSM 2034 type strain used as an outgroup taxon. Type strains are indicated by "(T)." *Desulfovibrio* strains in red do not produce alkanes. *Desulfovibrio* strains in green synthesise alkanes. The blue circle identifies the *Desulfovibrio* ancestral strain of all the alkane producing strains. The tree was constructed using the approximately maximum likelihood method. Numbers at nodes are bootstrap values.

Interestingly all the *Desulfovibrio* strains proven to be capable of alkane production in this study were part of the same clade. In this clade, only one *Desulfovibrio* strain does not synthesise alkane, *D. giganteus* DSM 4370. Thus, the ability to produce alkanes may have been acquired by a *Desulfovibrio* ancestral strain *via* horizontal gene transfer. Therefore, a novel hypothesis was proposed that the ability of *Desulfovibrio* to produce alkanes is due to the presence of genes encoding enzymes catalysing alkane synthesis.

4.3.3. *Desulfovibrio* Genomic comparison

To substantiate this novel hypothesis, a whole-genome comparative analysis was performed to estimate the genomic diversity between the alkane producing and non-alkane producing *Desulfovibrio* genomes.

Genomic comparison through a sequence based approach

The similarity of *Desulfovibrio* genome sequences were firstly evaluated. *Desulfovibrio* genomes generated in this study were aligned to the model organism *D. vulgaris* strain Hildenborough genome (GenBank AE017285.1) using the tool BLAST with an E-value threshold of $10e^{-5}$. To visualise similarity of the assembled *Desulfovibrio* genomes, the alignments with a minimum percentage identity of 70 % were displayed (Figure 4.5).

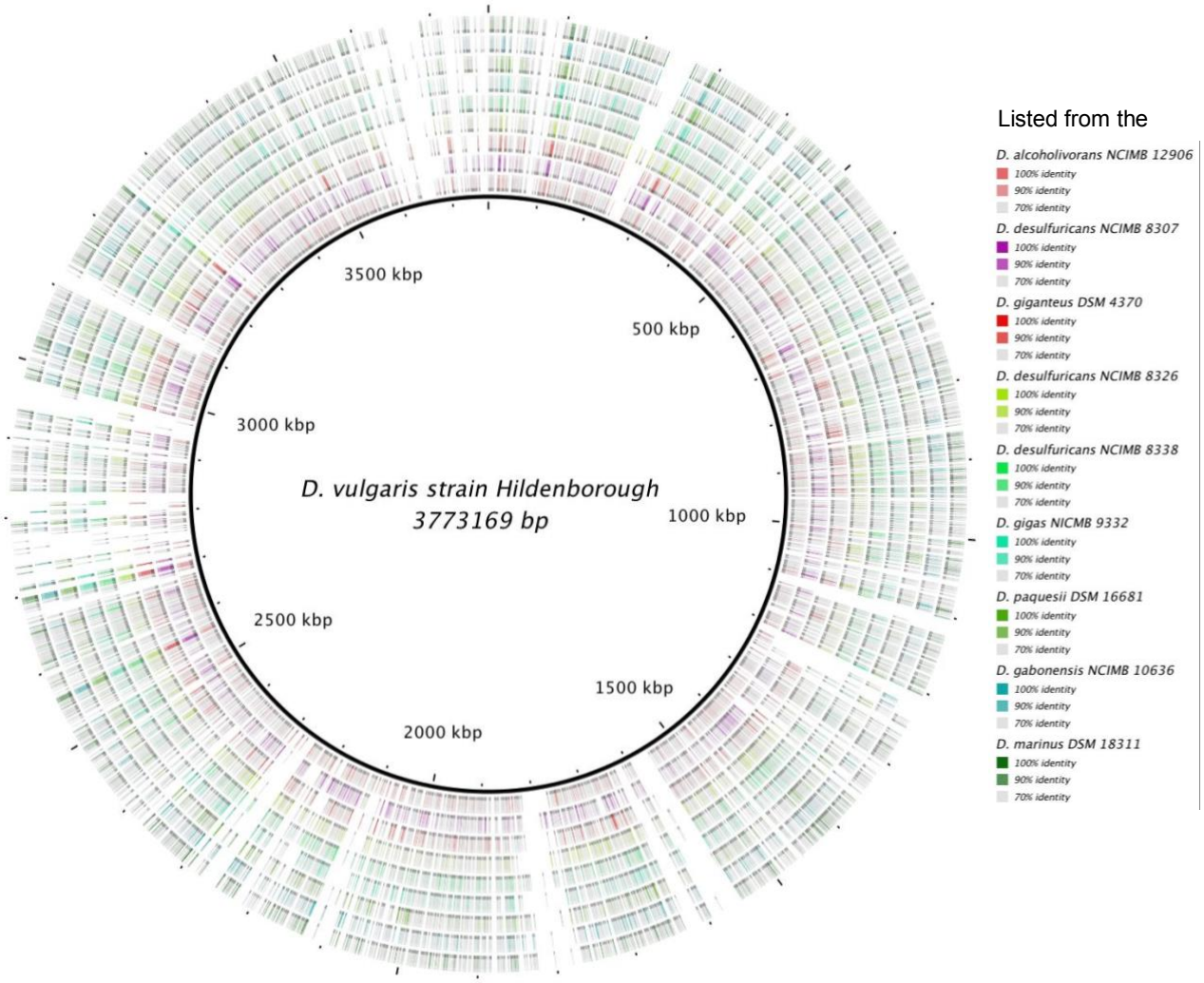


Figure 4.5. Representation of the *de novo* Assembled *Desulfovibrio* genome alignments to the genome of *D. vulgaris* strain Hildenborough, model organism for the genus

Assembled *Desulfovibrio* genomes from this study were aligned to the model organism *D. vulgaris* strain Hildenborough genome using BLAST. The visualisation of this genomic comparison was generated by the tool BRIG. The black inside track represents the *D. vulgaris* strain Hildenborough genome, which is surrounded by nine concentric rings representing the assembled *Desulfovibrio* genomes from this study. *Desulfovibrio* genomes with a red hue do not produce alkanes and *Desulfovibrio* genomes with a green hue are capable of alkane synthesis. Only the genomic regions of assembled *Desulfovibrio* genomes which aligned to *D. vulgaris* strain Hildenborough genome with an E-value equal or greater than $10e^{-5}$ and with a minimum of 70 % identity are coloured. The colour intensity in each ring represents the BLAST match identity. The white genomic regions represent regions with a BLAST output below the E-value and percentage identity thresholds.

The substantial number of “white” and pale coloured genomic regions in figure 4.5 showed that *Desulfovibrio* genomes are substantially different to the model organism *D. vulgaris* strain Hildenborough genome. Moreover, comparing the *Desulfovibrio* genome alignments to the *D. vulgaris* strain Hildenborough genome, *Desulfovibrio* genomes differently aligned to the reference genome, suggesting that *Desulfovibrio* genomes have strong variations.

Genomic Comparison through a Gene Content based Approach

The gene content of *Desulfovibrio* genomes was secondly investigated through pan-genomic analyses. Pan-genomes of the nine assembled *Desulfovibrio* genomes plus the model organism genome for the genus, *D. vulgaris* strain Hildenborough, were established using two different approaches (Figure 4.6). *Desulfovibrio* pan-genome generated by Anvi’o relied on the sequencing and *de novo* assembly quality of *Desulfovibrio* genomes within his study. *Desulfovibrio* pan-genome generated by *get_homologues* relied on the sequencing, the *de novo* assembly and the annotation quality of *Desulfovibrio* genomes within this study.

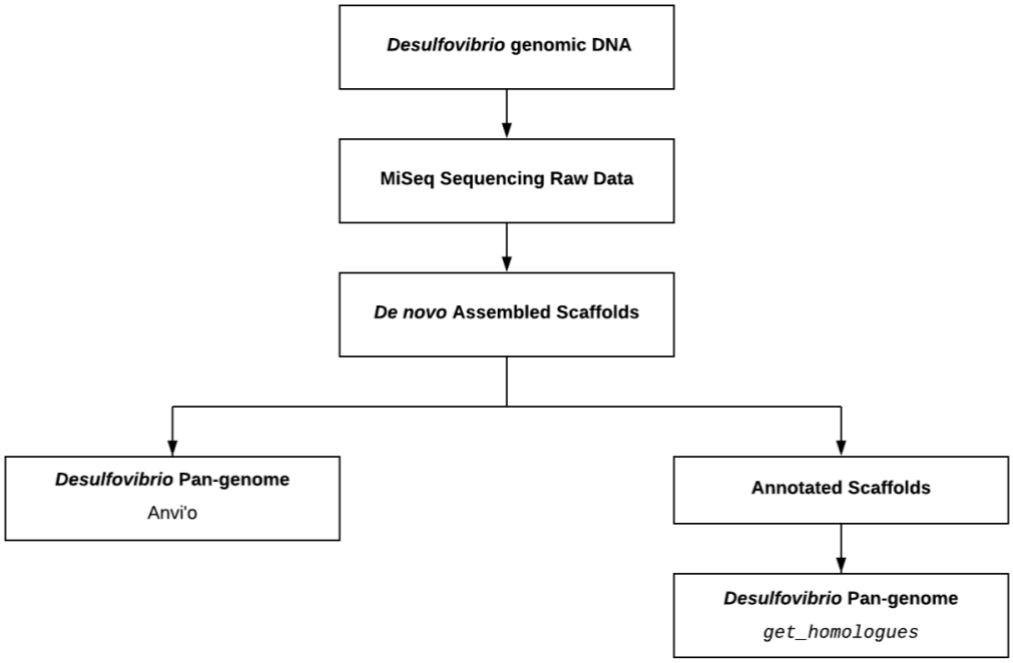


Figure 4.6. Flowchart of *Desulfovibrio* pan-genome generation

Pan-genomes of ten *Desulfovibrio* strains were generated from the *de novo* assembled scaffolds using the platform Anvi'o and from the annotated scaffolds using the software get_homologues. (Flowchart rendered in Lucidchart)

Desulfovibrio pan-genome generated by the platform Anvi'o

Pan-genomic analysis of the ten *Desulfovibrio* genomes was firstly performed using the platform Anvi'o. The *de novo* assembled *Desulfovibrio* genomes from this study and *D. vulgaris* strain Hildenborough genome from GenBank (AE017285.1) were used as input data for Anvi'o pan-genomic analysis.

Anvi'o pan-genomic analysis started with identification of open reading frames in *Desulfovibrio* assembled genomes, using the program Prodigal. Prodigal program reported 39,618 putative proteins in the ten *Desulfovibrio* genomes. The putative proteins were then annotated using the Clusters of Orthologous Groups (COG) of proteins database. To generate a pan-genome, the 39,618 protein sequences were aligned to each other for sequence similarity searching, using the software DIAMOND. Proteins were clustered into 16,574 homologous groups using MCL algorithm, with an inflation value of 7, as closely related genomes are being compared. The *Desulfovibrio* pan-genome was therefore composed of 16,574 detected protein clusters, including those occurring in only one genome.

Presence or absence of the 16,574 protein clusters in each *Desulfovibrio* strain was investigated. Protein clusters common to all the *Desulfovibrio* strains of the study were gathered into a bin named "*Desulfovibrio* core genome" (Figure 4.7). *Desulfovibrio* core genome contained 560 protein clusters, which corresponded to 3.4 % of the pan-genome. This illustrated *Desulfovibrio* genomic variations between strains.

A parsimony pan-genomic tree was also built showing the phylogenetic relationship between the *Desulfovibrio* strains based on their protein cluster content (Figure 4.7, top right). Interestingly, the non-alkane producing *D. desulfuricans* NCIMB 8307 has the closest protein cluster content to the protein cluster content of alkane producing strains, according to Anvi'o pan-genomic tree. In the *Desulfovibrio* 16S rRNA phylogenetic tree, *D. giganteus* DSM 4370 is inferred to be the closest non-alkane producing strain to the alkane producing strains. Phylogeny between prokaryotes of the same genus is unlikely to be established based on a single gene sequence variation, such as the 16S rRNA gene, notably due to horizontal gene transfer events (Snel *et al.*, 1999). Pan-genomic trees based on the protein/gene content is a more accurate approach to estimate phylogenetic relationship between prokaryotes of the same genus.

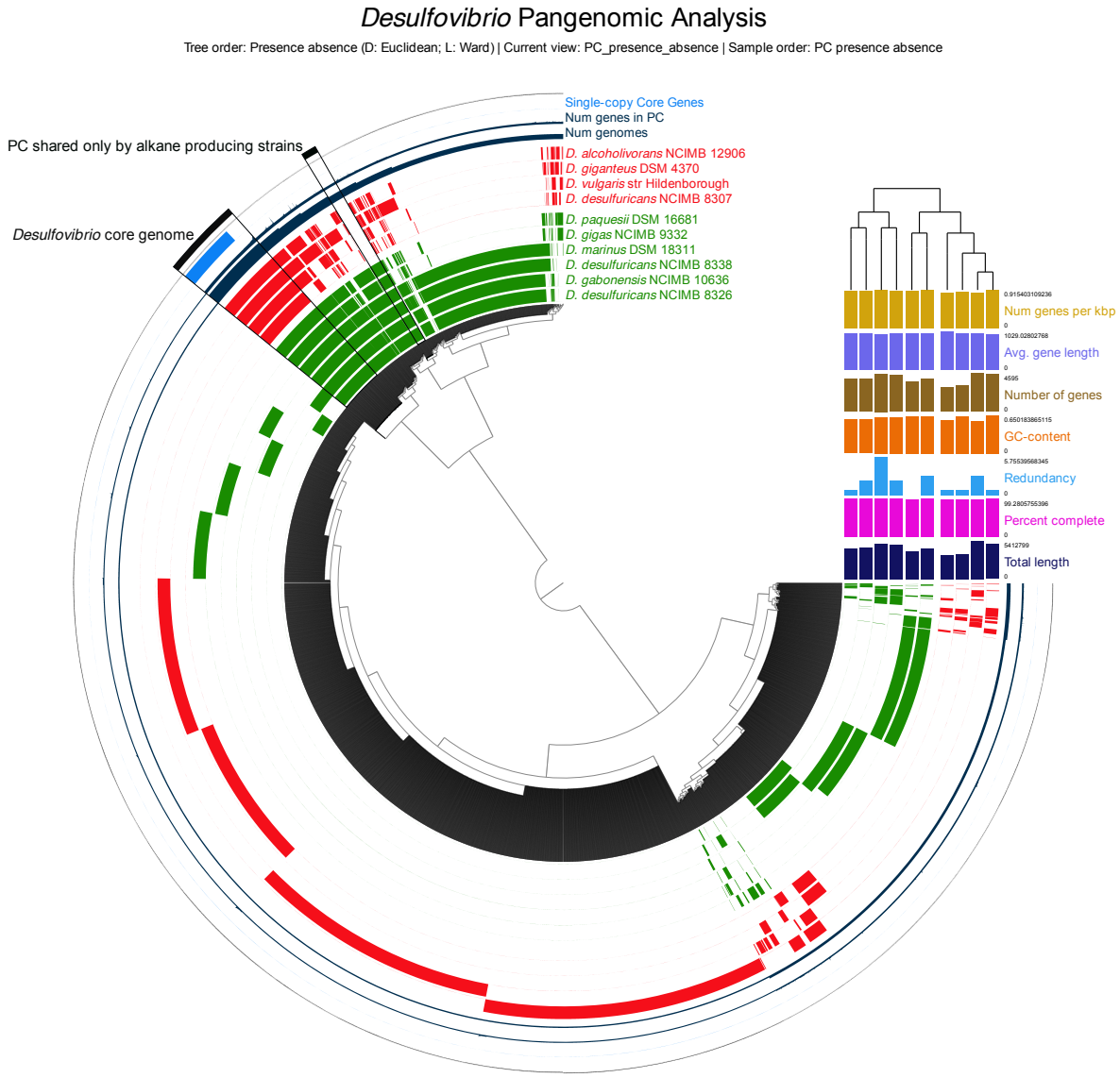


Figure 4.7. Visualisation of Anvi'o pan-genomic analysis of 10 *Desulfovibrio* genomes based on the presence and absence of 16,574 protein clusters

Ten *Desulfovibrio* genomes were involved in the pan-genomic analysis using the platform Anvi'o. *Desulfovibrio* genomes in red do not produce alkanes. *Desulfovibrio* genomes in green synthesise alkanes. A parsimony pan-genomic tree of the *Desulfovibrio* strains involved in the pan-genome was built (top right of the figure). Protein clusters (PC) present in all the *Desulfovibrio* genomes involved in the pan-genome and PC exclusively present in alkane producing strain genomes were collected into the "*Desulfovibrio* core genome" bin and the "PC shared only by alkane producing strains" bin respectively.

Desulfovibrio pan-genome *generated by the software get_homologues*

Secondly, pan-genomic analysis of the ten *Desulfovibrio* genomes was performed using the software *get_homologues*. The annotated *Desulfovibrio* genomes from this study and the annotated *D. vulgaris* strain Hildenborough genome from GenBank (AE017285.1) were used as input data. *get_homologues* pan-genomic analysis of a group of related organisms started with sequence similarity searching, using the software BLAST+. Then, two clustering algorithms OrthoMCL and COGtriangles identified 12,836 and 13,277 protein clusters respectively in the ten *Desulfovibrio* strains, including those occurring in only one strain. Intersecting OrthoMCL clusters and COGtriangles clusters, 11,184 clusters were detected by both algorithms (Figure 4.8).

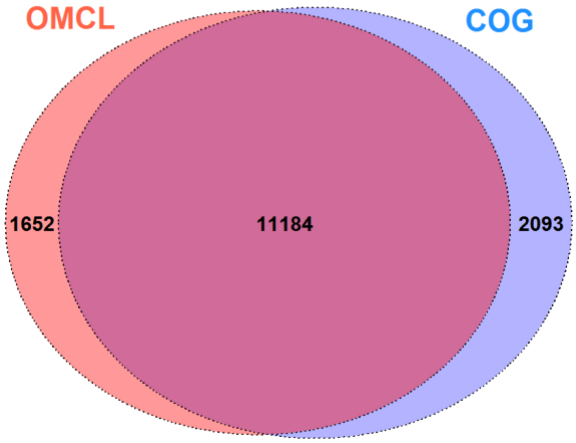


Figure 4.8. Venn diagram of protein clusters from 10 *Desulfovibrio* strains, identified by OrthoMCL and COGtriangles algorithms

The genomes from ten *Desulfovibrio* strains were involved in a pan-genomic analysis using the software *get_homologues*. Protein sequences were clustered into 12,836 orthologous groups using OrthoMCL algorithm and into 13,277 orthologous groups using COGtriangles algorithm. Intersecting OrthoMCL clusters and COGtriangles clusters, 11,184 clusters were detected by both algorithms.

Desulfovibrio pan-genome generated by *get_homologues* therefore comprised the 11,184 protein clusters detected by both OrthoMCL and COGtriangles algorithms. Presence or absence of the 11,184 proteins in each individual *Desulfovibrio* strain was investigated. The core genome of *Desulfovibrio* pan-genome generated by *get_homologues* was composed of 872 protein clusters, which corresponded to 7.8 % of the pan-genome (Figure 4.9). This illustrated *Desulfovibrio* genomic variations between strains.

A parsimony pangenomic tree was also built from the *Desulfovibrio* pan-genome generated by *get_homologues* (Figure 4.10). In the generated *get_homologues* pan-genomic tree, the non-alkane producing *D. giganteus* DSM 4370 has the closest gene content to the gene content of alkane producing strains. The *get_homologues* pan-genomic tree corroborates the 16S rRNA phylogenetic tree for the *Desulfovibrio* genus and differs from the Anvi'o pan-genomic tree. Depending on homology criteria and clustering method used, different orthologous clusters are detected among organisms and so different gene family content based phylogeny for these organisms are predicted (Hughes *et al.*, 2005).

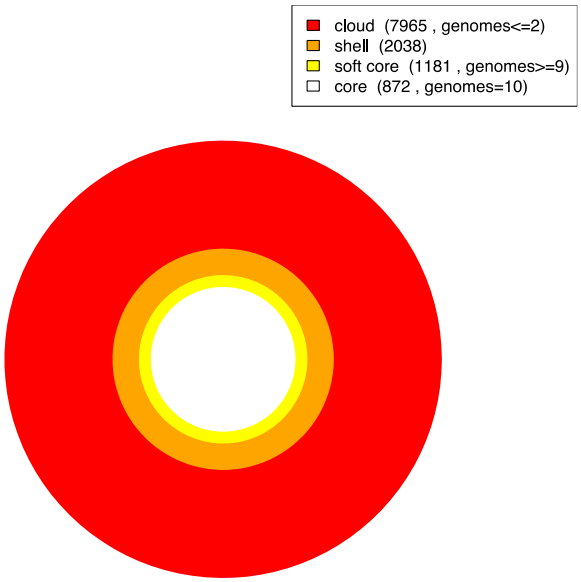


Figure 4.9. Partition of the *Desulfovibrio* pan-genome into cloud, shell, soft-core and core genomes, according to the *get_homologues* analysis

The *Desulfovibrio* pan-genome generated by *get_homologues* comprised 11,184 protein clusters. The core genome, which contained clusters present in all *Desulfovibrio* genomes, consisted of 872 protein clusters. The soft-core genome contained 1,181 protein clusters which were present in at least 9 *Desulfovibrio* genomes. 7,965 protein clusters were present in at most two *Desulfovibrio* genomes and constituted the cloud genome. The remaining 2,038 protein clusters were present in several *Desulfovibrio* genomes and composed the shell genome.

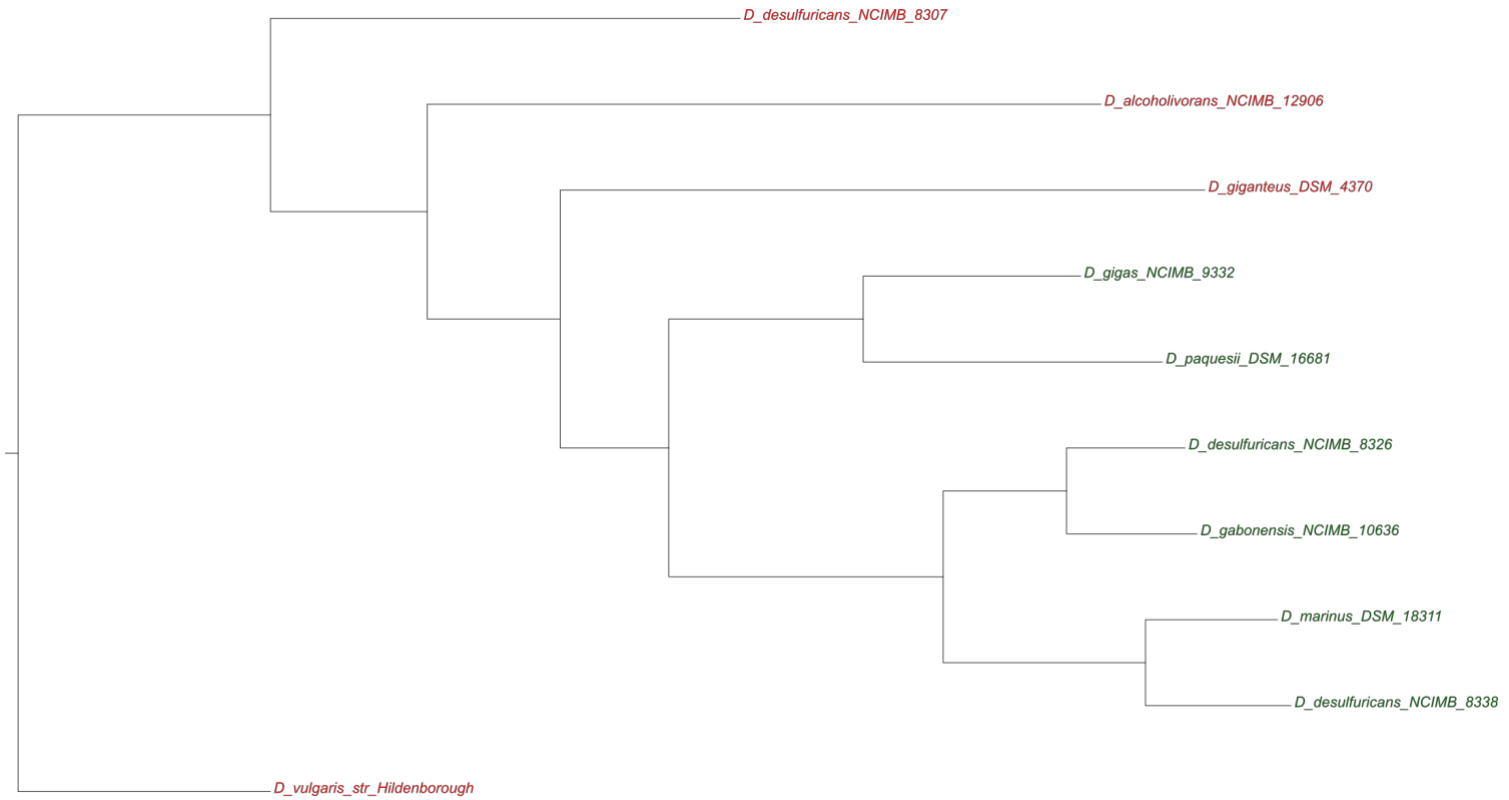


Figure 4.10. Parsimony pan-genomic tree for 10 *Desulfovibrio* strains based on presence and absence of 11,184 protein clusters, generated by the software *get_homologues*

Genomes from ten *Desulfovibrio* strains were involved in a pan-genomic analysis using the software *get_homologues*. From Prokka annotated *Desulfovibrio* genomes, 11,184 orthologous protein clusters were detected by both OrthoMCL and COGtriangles algorithms. Phylogenetic relatedness between strains was determined based on presence and absence of the 11,184 orthologous proteins in their genome. *Desulfovibrio* strains in red do not produce alkanes. *Desulfovibrio* strains in green synthesise alkanes.

4.3.4. Identification of Proteins Clusters Exclusively Present in Alkane Producing *Desulfovibrio*

Consistent with the hypothesis that alkane producing *Desulfovibrio* strains share orthologous genes encoding alkane biosynthetic enzymes, protein clusters exclusively present in alkane producing strains were identified.

Protein clusters predicted to be exclusive to alkane producing Desulfovibrio strains by the platform Anvi'o

Protein clusters exclusively shared by alkane producing *Desulfovibrio* spp. were collected into a bin named "PC shared only by alkane producing strains" (Figure 4.7). This bin included 107 protein clusters (Appendix 4).

The 107 protein clusters were checked for false positive clusters, which were those identified as exclusively present in alkane producing *Desulfovibrio* strains but which are present, but not identified, in non-alkane producing strains. To screen for false positive clusters, the consensus sequence of each of the 107 protein clusters identified to be exclusive to alkane producing strains was aligned to the non-alkane producing *Desulfovibrio* proteomes. Ten protein cluster consensus sequences aligned to proteins from non-alkane producing *Desulfovibrio* proteomes, with an E-value of 0 and a highest reported percentage of identity of 67.0 % (Appendix 4).

Protein clusters predicted to be exclusive to alkane producing Desulfovibrio strains by the software get_homologues

The software *get_homologues* offers the possibility to identify protein clusters present in a defined group of strains which are absent in a second defined group of strains. In this study, the alkane producing *Desulfovibrio* genomes were gathered into one group, divided from the non-alkane producing strains composing another group of strains. The software *get_homologues* identified 104 protein clusters as exclusive to alkane producing strains (Appendix 5).

These 104 protein clusters were also checked for false positive clusters. The consensus sequence of each of the 104 protein clusters exclusively present in alkane

producing strains was aligned to the non-alkane producing *Desulfovibrio* proteomes. Only one protein cluster consensus sequence aligned to proteins from non-alkane producing *Desulfovibrio* proteomes, with an E-value of 0 and a percentage identity of 49.9 % (Appendix 5).

Establishment of a list of protein clusters exclusive to alkane producing Desulfovibrio

Generation of *Desulfovibrio* pan-genomes by the programs Anvi'o and *get_homologues* followed distinct workflows, using different type of input data, sequence alignment algorithms and homology clustering algorithms (Figure 4.11).

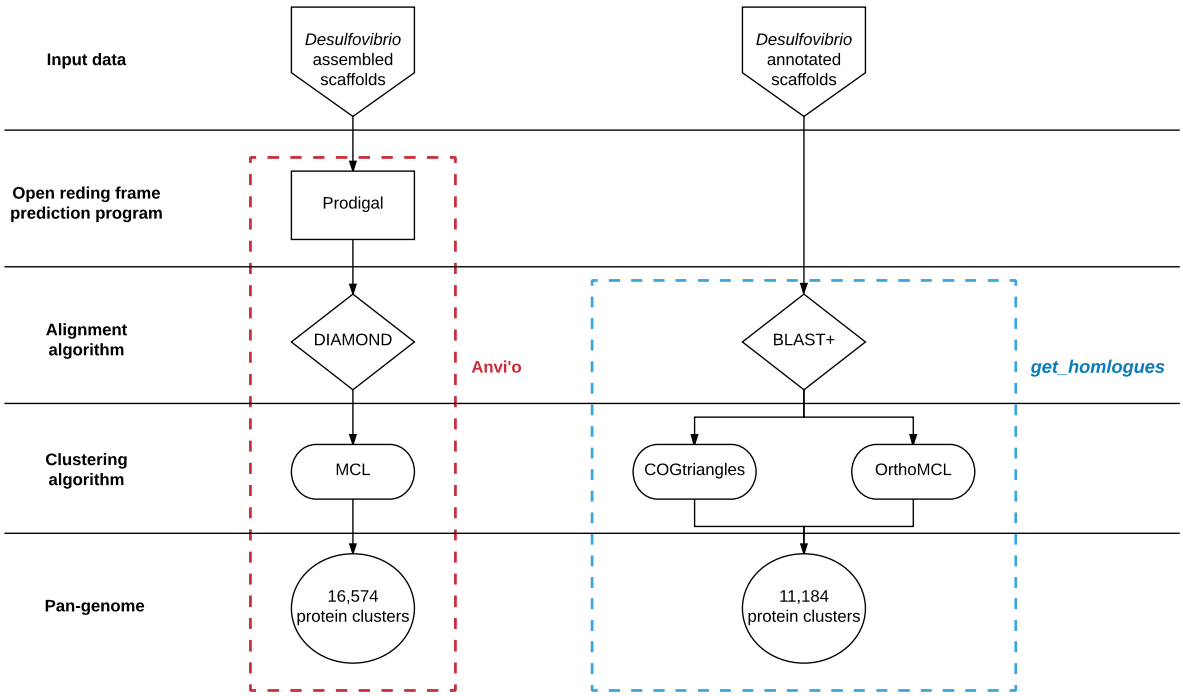


Figure 4.11. Comparison of Anvi'o and *get_homologues* workflows for *Desulfovibrio* pan-genome generation

Desulfovibrio pan-genomes were generated using two distinct workflows from the platform Anvi'o and from the software *get_homolgues*. The Anvi'o workflow started with identification of open reading frames in genome sequences, using the program Prodigal. Proteins were then aligned to each other for sequence similarity searching using the algorithm DIAMOND. Proteins were clustered in 16,574 orthologous groups using the MCL algorithm. The g*et_homologues* software used BLAST+ to align proteins to each other for sequence similarity searching. Proteins were then clustered independently using COGtriangles and OrthoMCL algorithms. Both clustering algorithms identified 11,184 protein clusters in common. (Flowchart rendered in Lucidchart)

Pan-genome establishment by two distinct workflows assigned confidence that protein clusters detected by both workflows are present in proteomes of strains under consideration. To establish a list of protein clusters that can be considered exclusive to alkane producing *Desulfovibrio* with confidence, the two sets of protein clusters predicted exclusive to alkane producing strains by Anvi'o and *get_homologues* programs were intersected.

Thirty-three protein cluster consensus sequences from *get_homologues* aligned to the protein cluster consensus sequences from Anvi'o with a maximal E-value of $7e^{-54}$ and a percentage of identity ranging from 90 % to 100 %. The 33 protein clusters were ordered into their predicted COG functional category (Table 4.11).

Table 4.11. Protein clusters predicted exclusive to alkane producing *Desulfovibrio* strains by both Anvi'o and *get_homologues* pan-genomic tools

Protein clusters (PC) predicted exclusive to alkane producing strains by *get_homologues*	PC predicted exclusive to alkane producing strains by Anvi'o	E-value	Bit score	Percentage identity (%)	PC consensus sequence function
Inorganic ion transport and metabolism					
PC_3696	PC_00001312	0	1,002	100	Na (+)/H (+) antiporter subunit A
PC_3699	PC_00001334	0	553	100	Na (+)/H (+) antiporter subunit B
PC_3700	PC_00001296	$7e^{-54}$	158	100	Putative Na (+)/H (+) antiporter subunit B
PC_3698	PC_00001341	$1e^{-130}$	362	100	Na (+)/H (+) antiporter subunit C
PC_3697	PC_00001457	0	1,005	100	Na (+)/H (+) antiporter subunit D
PC_3702	PC_00001298	$1e^{-74}$	213	100	Na (+)/H (+) antiporter subunit F
PC_3701	PC_00001345	$1e^{-88}$	250	100	Na (+)/H (+) antiporter subunit G
PC_2663	PC_00001463	0	1,200	100	Putative sodium-dependent transporter
PC_1132	PC_00001062	$4e^{-178}$	492	93.5	Cation transporter
Energy production and conversion					
PC_1363	PC_00001526	0	1,309	100	V-type ATPase subunit A
PC_1362	PC_00001492	0	890	100	V-type ATPase subunit B
PC_1361	PC_00001396	$2e^{-141}$	390	100	V-type ATPase subunit D
PC_1359	PC_00001318	$8e^{-109}$	304	100	V-type ATPase subunit K
PC_3837	PC_00001379	0	1,867	99.9	Trimethylamine-N-oxide reductase (cytochrome c)
PC_3431	PC_00001310	$3e^{-97}$	272	100	Neelaredoxin
PC_3457	PC_00001535	$3e^{-169}$	463	100	(2Fe-2S) Ferredoxin

Protein clusters (PC) predicted exclusive to alkane producing strains by `get_homologues`	`Anvi'o` PC predicted exclusive to alkane producing strains	E-value	Bit score	Percentage identity (%)	PC consensus sequence function
Signal transduction mechanisms					
PC_1702	PC_00001372	0	1,350	99.7	Blue-light-activated integral membrane sensor
PC_1228	PC_00001376	$4e^{-91}$	257	98.5	Putative two-component response regulator
PC_1852	PC_00001316	$3e^{-105}$	294	100	Universal stress protein
PC_2878	PC_00001375	0	657	100	Diguanylate cyclase (GGDEF) domain-containing protein
PC_1838	PC_00001319	$5e^{-76}$	217	100	Anti-sigma factor antagonist
Amino acid transport and metabolism					
PC_2848	PC_00001417	0	581	100	ABC-type transporter, integral membrane subunit
PC_2847	PC_00001382	$2e^{-180}$	494	100	ABC-type transporter, integral membrane subunit
Mobilome: prophages, transposons					
PC_0120	PC_00001470	0	915	100	Phage terminase large subunit
Cell wall/membrane/envelope biogenesis					
PC_0998	PC_00001464	$2e^{-88}$	249	100	Diacylglycerol kinase
PC_1773	PC_00001484	0	1,253	100	Capsular polysaccharide biosynthesis protein
Post-translational modification, protein turnover, chaperones					
PC_0937	PC_00001410	0	618	100	Paraslipin
Defence mechanisms					
PC_1024	PC_00001518	0	890	100	MATE family efflux transporter
General functional prediction only					
PC_0064	PC_00001501	$2e^{-132}$	366	100	Chloramphenicol acetyltransferase

Protein clusters (PC) predicted exclusive to alkane producing strains by `get_homologues`	`Anvi'o` PC predicted exclusive to alkane producing strains	E-value	Bit score	Percentage identity (%)	PC consensus sequence function
General functional prediction only					
PC_3584	PC_00001454	$1e^{-159}$	437	99.5	Type 12 methyltransferase
Unknown function					
PC_3405	PC_00001333	$4e^{-58}$	169	93.2	Uncharacterised protein
PC_3239	PC_00001342	$2e^{-70}$	223	100	Uncharacterised protein
PC_2182	PC_00001299	$4e^{-103}$	288	100	Uncharacterised protein

To establish the list of protein clusters that were predicted exclusive to alkane producing *Desulfovibrio* strains by both Anvi'o and *get_homologues* programs, consensus sequences from protein clusters detected by *get_homologues* were used as a query for a blastp search against the Anvi'o protein cluster consensus sequences.

4.3.5. Identification of Protein Clusters Potentially Involved in Alkane Production

Further investigations were carried out on the sets of 107 and 104 protein clusters detected to be exclusive to alkane producing *Desulfovibrio* strains by Anvi'o and *get_homologues* respectively, in order to identify proteins potentially involved in alkane production.

Identification of protein clusters potentially involved in alkane production by function prediction

Protein clusters generated by Anvi'o were annotated at the beginning of pan-genomic analysis by sequence similarity searching in the protein database COG. The annotation of the protein clusters generated by *get_homologues* was established prior to the pan-genomic analysis by the software Prokka. Analysing the protein cluster annotation, none of the protein clusters exclusive to alkane producing *Desulfovibrio* strains had a predicted function evidently involved in alkane production. Moreover, not all protein clusters had a predicted function.

The protein cluster annotation was checked by sequence similarity prediction. The annotation of protein clusters from Anvi'o was verified by sequence similarity searching in UniProtKB, a protein database different to the COG database used for annotation. The annotation of protein clusters from *get_homologues* was verified by sequence similarity searching in COG database, a protein database different to UniProtKB, RefSeq, Pfam and TIGRFAMs databases used by the annotation tool Prokka. Protein cluster annotation was also checked according to protein domain prediction. After annotation verification, none of the protein clusters exclusive to alkane producing *Desulfovibrio* strains had a predicted function evidently involved in alkane production. Moreover, 13 protein clusters from Anvi'o and 29 protein clusters from *get_homologues* still had an unknown function (Appendix 4 and 5).

To estimate the functional diversity among the protein clusters predicted exclusive to alkane producing *Desulfovibrio* strains, the COG functional categories annotating protein clusters from Anvi'o and *get_homologues* were listed and their prediction frequency was enumerated (Figure 4.12). Among the 107 protein clusters predicted exclusive to alkane producing strains by Anvi'o, 94 protein clusters were

assigned into 21 out of the 26 COG functional categories. Among the 104 protein clusters predicted exclusive to alkane producing strains by *get_homologues*, 77 protein clusters were assigned into 17 out of the 26 COG functional categories. The frequency median value of the COG functional categories annotating those protein clusters from either Anvi'o or *get_homologues* was 4. Therefore, protein clusters predicted exclusive to alkane producing *Desulfovibrio* strains were diverse, involved in a wide range of functions.

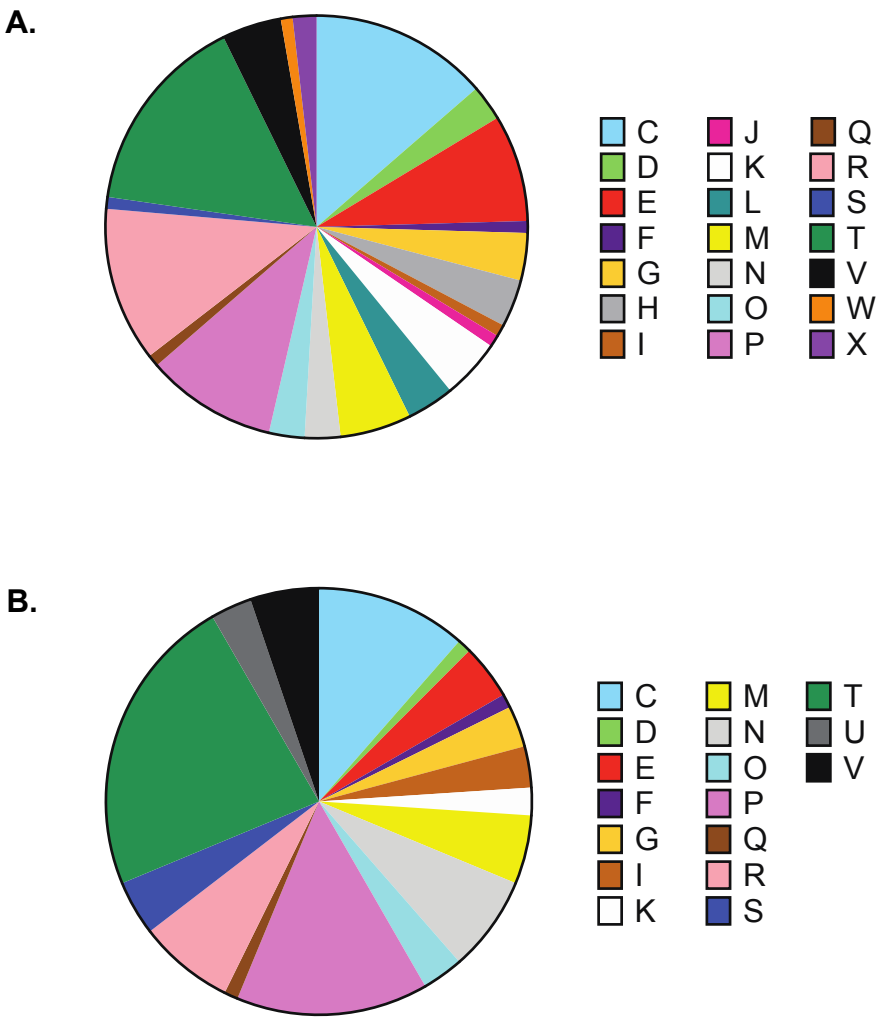


Figure 4.12. Proportion of the COG functional categories annotating protein clusters exclusively detected in alkane producing *Desulfovibrio* strains

The COG functional categories annotating protein clusters detected to be exclusively present in alkane producing *Desulfovibrio* strains by Anvi'o (A) and by *get_homologues* (B) were listed and their proportion was calculated. Designation of COG functional categories: C, Energy production and conversion; D, Cell cycle control, cell division and chromosome partitioning; E, Amino acid transport and metabolism; F, Nucleotide transport and metabolism; G, Carbohydrate transport and metabolism; H, Coenzyme transport and metabolism; I, Lipid transport and metabolism; J, Translation, including ribosome structure and biogenesis; K, Transcription; L, Replication, recombination and repair; M, Cell wall/membrane/envelop biogenesis; N, Cell motility; O, Post-translational modification, protein turnover, chaperones; P, Inorganic ion transport and metabolism; Q, Secondary metabolites biosynthesis, transport and catabolism; R, General functional prediction only; S, Function unknown; T, Signal transduction mechanisms; V, Defence mechanisms, W, Extracellular structures, X, Mobilome: prophages, transposons.

Identification of protein clusters potentially involved in alkane production by structure prediction

Protein clusters sharing similar structural features with alkane-binding proteins were identified among the protein clusters predicted exclusive to alkane producing *Desulfovibrio* strains. A previous study analysed the structure of proteins interacting with long-chain (at least 10 carbons) alkanes and listed the name of the SCOP fold category of 407 alkane-binding proteins (Park *et al.*, 2015). The SCOP database hierarchically classifies proteins of known structure, based on evolutionary relationships and structural similarities. Structure proteins gathered in the same SCOP fold category share similar structural features (Murzin *et al.*, 1995). The tertiary structure of the protein clusters exclusive to alkane producing *Desulfovibrio* strains were therefore modelled and screened for SCOP fold category previously identified in alkane-binding proteins.

The tertiary structure of protein clusters exclusive to alkane producing *Desulfovibrio* were predicted from their consensus sequence using the program RaptorX. Tertiary structure prediction by RaptorX uses template-based modelling methods which compares target sequence or target sequence-profile to identify homologous proteins or proteins sharing common structural folds. The characterised structure of identified proteins serves as template for modelling the tertiary structure of a protein target (Peng & Xu, 2011; Ma *et al.*, 2013). A tertiary structure was predicted for the 107 protein clusters and the 104 protein clusters predicted exclusive to alkane producing *Desulfovibrio* strains by Anvi'o and *get_homologues* respectively.

For each tertiary structure prediction, the Global Distance Test (GDT) or the un-normalised GDT (uGDT) value and the p-value RaptorX metrics were analysed to evaluate the quality of the prediction. uGDT is a more accurate metric than GDT for structure prediction of proteins of at least 100 amino acids in length. The prediction from RaptorX modelling method is reliable when the p-value is below $10e^{-4}$ (Ma *et al.*, 2013). The p-value of a predicted tertiary structure is the probability that the predicted structure, which was modelled from a template considered as the best for the protein, is less accurate than a set of randomly-generated structures. For a p-value below $10e^{-4}$, predicted structures have a GDT or an uGDT value equal or greater than 50. A predicted structure with a GDT or an uGDT value equal or greater than 50 shares the same structural features at least at the fold level, with the best template structure used for the prediction (Ma *et al.*, 2013).

Among the 107 predicted tertiary structures for the protein clusters detected exclusive to alkane producing *Desulfovibrio* strains by Anvi'o, 89 structure predictions had a p-value below $10e^{-4}$. For those 89 structure predictions, the best template protein structure used for the prediction was checked for domain annotation from the SCOP database. Only 23 of the best template protein structures had a SCOP domain annotation. Focusing on the SCOP fold category, 8 of the best template protein structures were classified into a SCOP fold category previously identified in alkane-binding proteins. Therefore, 8 protein clusters predicted exclusive to alkane producing *Desulfovibrio* strains by Anvi'o likely share similar structural features with alkane-binding proteins (Figure 4.13).

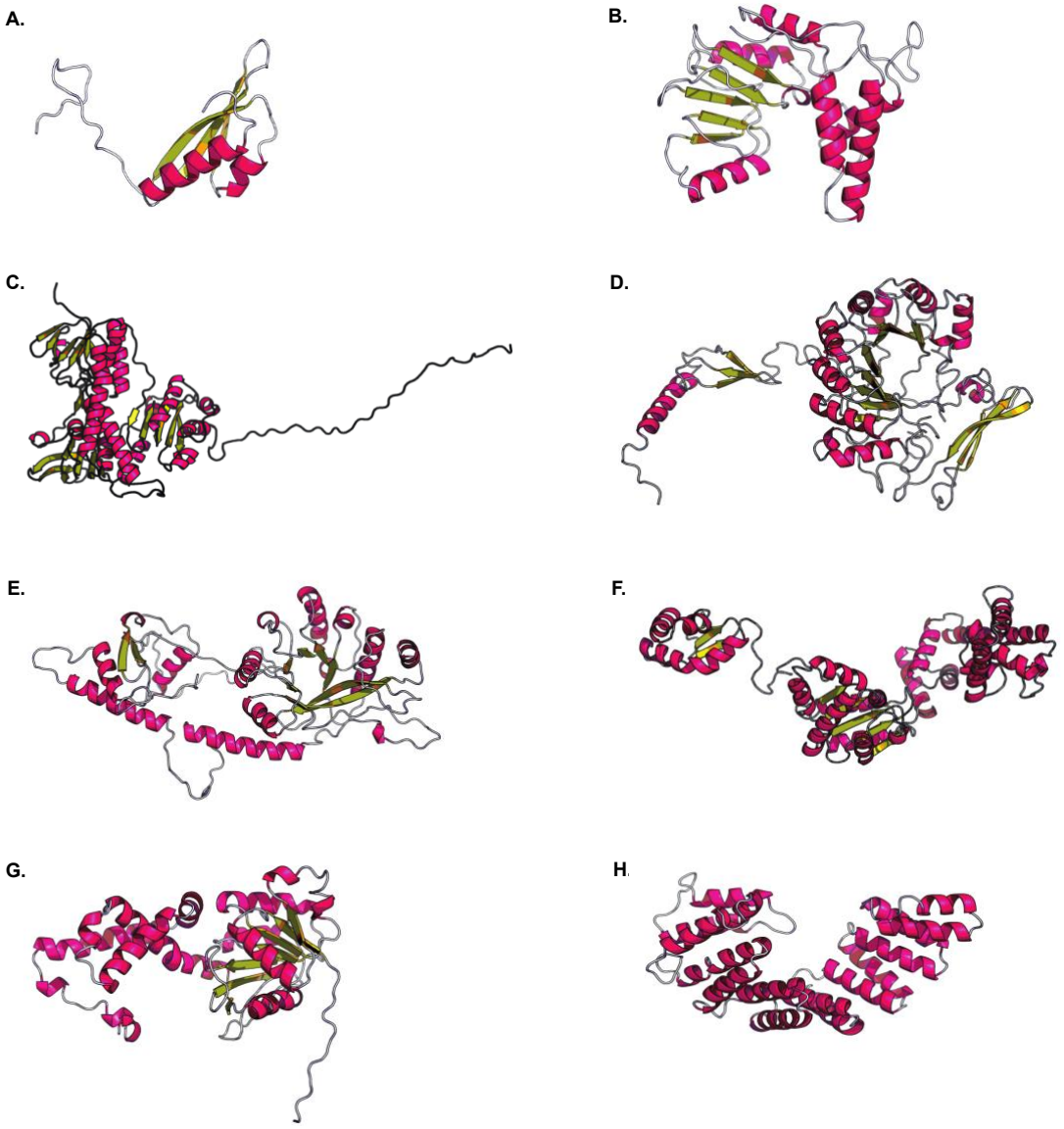


Figure 4.13. Predicted tertiary structures of Anvi'o protein clusters, which likely share similar structural features with alkane-binding proteins

The tertiary structures of Anvi'o protein clusters shown here were predicted by RaptorX from template protein structures that share the same SCOP fold category as the tertiary structures of alkane-binding proteins. Protein structures gathered in the same SCOP fold category share similar structural features. A, PC_00001333 predicted tertiary structure (p-value = $2.62e^{-5}$; GDT = 67; SCOP fold classification: Ferredoxin-like); B, PC_00001454 predicted tertiary structure (p-value = $1.04e^{-5}$; uGDT = 100; SCOP fold classification: S-adenosyl-L-methionine-dependent methyltransferases); C, PC_00001521 predicted tertiary structure (p-value = $6.09e^{-15}$; uGDT = 319; SCOP fold classification: FAD/NAD(P)-binding domain); D, PC_00001409 predicted tertiary structure (p-value = $1.91e^{-6}$; uGDT = 205; SCOP fold classification: TIM β/α-barrel); E, PC_00001403 predicted tertiary structure (p-value = $1.98e^{-4}$; uGDT = 120; SCOP fold classification: TIM β/α-barrel); F, PC_00001377 predicted tertiary structure (p-value =; uGDT =; SCOP fold classification: DNA/RNA-binding 3-helical bundle); G, PC_00001495 predicted tertiary structure (p-value = $3.92e^{-5}$; uGDT = 152; SCOP fold classification: 6-phosphogluconate dehydrogenase C-terminal domain-like); H, PC_00001384 predicted tertiary structure (p-value = $8.92e^{-6}$; uGDT = 113; SCOP fold classification: α/α-superhelix).

Among the 104 tertiary structure predictions for the protein clusters detected exclusive to alkane producing *Desulfovibrio* strains by *get_homologues*, 73 structures predictions had a p-value below $10e^{-4}$. For those 73 predictions, the best template protein structure used for structure prediction was checked for domain annotation from the SCOP database. Only 14 of the best template protein structures had a SCOP domain annotation, with 6 classified into a SCOP fold category previously identified in alkane-binding proteins. Therefore, 6 protein clusters predicted exclusive to alkane producing *Desulfovibrio* strains by *get_homologues* likely share similar structural features with alkane-binding proteins (Figure 4.14).

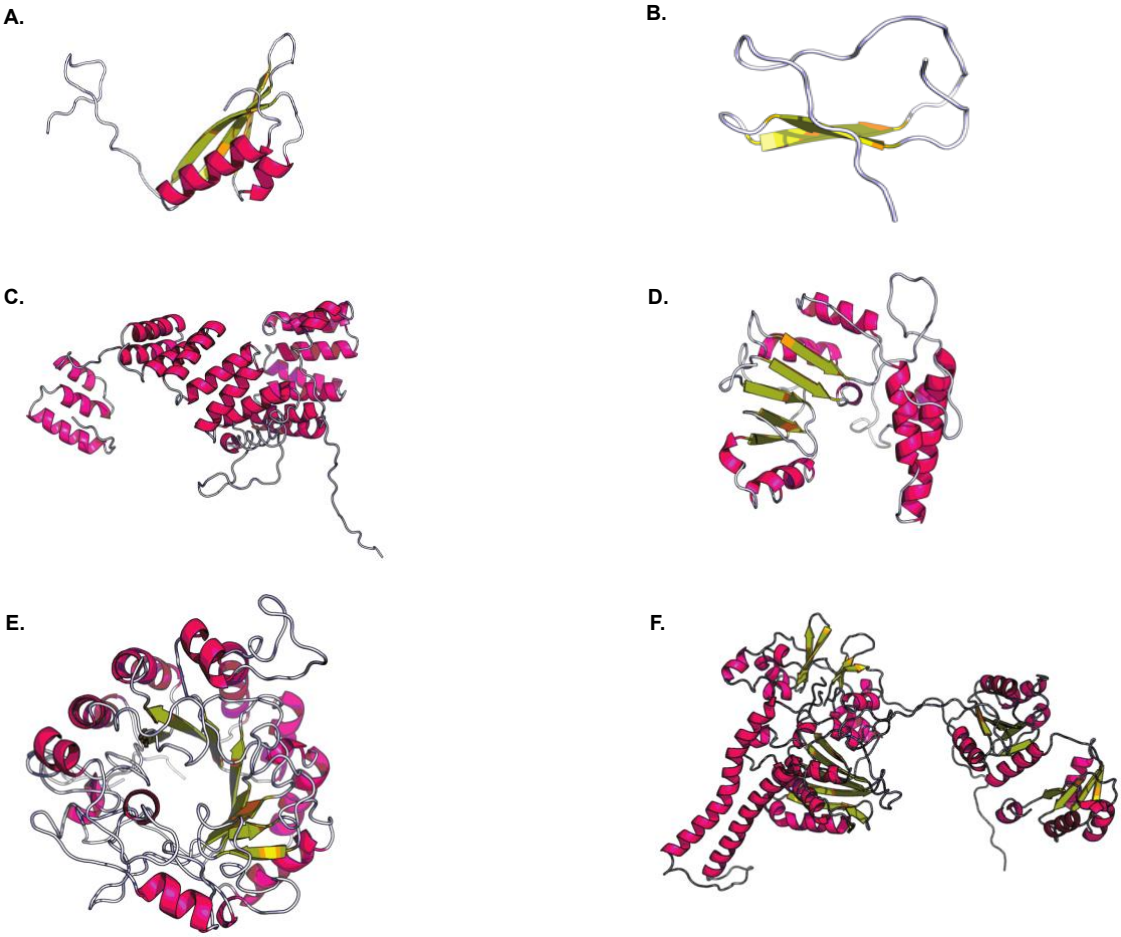


Figure 4.14. Predicted tertiary structures of *get_homologues* protein clusters, likely sharing similar structural features with alkane-binding proteins

The predicted tertiary structures of *get_homologues* protein clusters shown here were predicted by RaptorX from template protein structures that share the same SCOP fold category as the tertiary structures of alkane-binding proteins. Protein structures gathered in the same SCOP fold category share similar structural features. A, PC_3405 predicted tertiary structure (p-value = $2.97e^{-5}$; GDT = 67; SCOP fold classification: Ferredoxin-like); B, PC_0149 predicted tertiary structure (p-value = $4.93e^{-4}$; uGDT = 74; SCOP fold classification: Immunoglobin-like β-sandwich); C, PC_0678 predicted tertiary structure (p-value = $1.87e^{-6}$; uGDT = 148; SCOP fold classification: α/α-superhelix); D, PC_3584 predicted tertiary structure (p-value = $1.90e^{-5}$; uGDT = 90; SCOP fold classification: S-adenosyl-L-methionine-dependent methyltransferases); E, PC_2088 predicted tertiary structure (p-value = $2.00e^{-5}$; uGDT = 132; SCOP fold classification: TIM β/α-barrel); F, PC_3173 predicted tertiary structure (p-value = $1.40e^{-5}$; uGDT = 401; SCOP fold classification: Ferredoxin-like).

Interestingly, two protein clusters (Anvi'o PC_00001333; *get_homologues* PC_3405 and Anvi'o PC_00001454; *get_homologues* PC_3584) identified by both Anvi'o and *get_homologues* pan-genomic tools to be exclusive to alkane producing strains were likely to share similar structural features with alkane-binding proteins.

4.4. Discussion

In this study, the genome of *D. desulfuricans* NCIMB 8326, *D. desulfuricans* NCIMB 8338, *D. gabonensis* NCIMB 10636, *D. gigas* NCIMB 9332, *D. marinus* DSM 18311, *D. paquesii* DSM 16681, *D. desulfuricans* NCIMB 8307, *D. giganteus* DSM 4370 and *D. alcoholivorans* NCIMB 12906 were sequenced using Illumina paired-end short read sequencing technology, assembled *de novo* and compared. The comparative genomic analysis was therefore based on the sequencing and *de novo* assembly quality of the *Desulfovibrio* genomes. Although *Desulfovibrio* assemblies from this study show a correct size, contiguity, accuracy and completeness, genomes were fragmented and so incomplete. A high number of scaffolds is mostly due to the presence of repeated elements in the genome (Ricker *et al.*, 2012). Genome *de novo* assembly with short reads is a computational challenge, notably as short reads cannot span longer repetitive genomic regions. Therefore, genome *de novo* assembly with short reads can result in incomplete and misassembled genomes (Treangen and Salzberg, 2013). Long read sequencing technologies have now been developed such as the single-molecule real-time sequencing produced by Pacific Biosciences (PacBio sequencing) and the single-molecule Nanopore MinION® sequencing from Oxford Nanopore Technologies Ltd. Long read sequencing technologies enable the generation of complete genome assembly. In a previous study, *Clostridium autoethanogenum* DSM 10061 genome was assembled *de novo* in at least 22 scaffolds using short read sequencing technologies and in only a single scaffold covering all the repetitive genomic regions using reads from PacBio sequencing (Brown *et al.*, 2014). In a separate study, *Acinetobacter baylyi* ADP1 genome was sequenced using the MinION® technology and assembled *de novo* into a single scaffold, covering 99.8 % of the reference genome with an identity greater than 99.98 % (Madoui *et al.*, 2015). Therefore, it is likely that the access to complete *Desulfovibrio* genomes using long read sequencing technology would have enhanced the comparative genomic analysis accuracy.

However, despite sequencing advancements, comparative genomic analyses still rely on prediction of genomic sequences. Final genome assemblies depend on the assembly strategy applied by the software assembler chosen and the parameters implemented to the assembly process (Ekblom & Wolf, 2014). Assembly of the same raw sequencing data by different assemblers result in various models of the existing genomic sequence (Studholme, 2016).

According to the 16S rRNA phylogenetic tree generated in this study, all the alkane producing *Desulfovibrio* strains are clustered within a single clade. The 16S rRNA phylogenetic tree of the study disclosed the same relationships between *Desulfovibrio* spp. to published previous 16S rRNA phylogenetic trees for the genus (Gilmour *et al.*, 2011; da Silva *et al.*, 2015). Moreover, the *Desulfovibrio* strains involved in this study and their type strain recorded in RDP database are at the same location in the 16S rRNA phylogenetic tree, with a bootstrap value superior to 0.8. These observations illustrate the robustness of the 16S rRNA phylogenetic tree generated in this study. Interestingly, *D. desulfuricans* NCIMB 8307 is not positioned in the same clade than the other *D. desulfuricans* strains. As previously discussed in chapter 3 (*cf.* Chapter 3 - Metabolism Screening; Section 3.4), *Desulfovibrio* taxonomy is disordered. The 16S rRNA gene of *D. desulfuricans* strain ND132 was shown to be 98 % to 99 % identical to *D. dechloracetivorans* 16S rRNA gene, while it is 88 % to 90 % similar to those of *D. desulfuricans* strains (Brown *et al.*, 2011).

The comparative genomic analysis of the ten *Desulfovibrio* strains within this study revealed great genomic diversity. Both *Desulfovibrio* pan-genomes generated within this study had a core genome gathering less than 10 % of the pan-genome genes, indicating *Desulfovibrio* pan-genomes are open. An open pan-genome means that the number of genes increases by addition of a new genome to the pan-genome (Tettelin *et al.*, 2008). Regarding the hypothesis that the ability of *Desulfovibrio* to produce alkanes is due to the presence of specific genes, the genomic diversity of the ten *Desulfovibrio* strains was an advantage for identification of the genes encoding alkane biosynthetic enzymes *via* a comparative genomic approach. The orthologous genes between the alkane producing strains represented a small proportion of the genomes. Between 2,957 and 4,553 genes were predicted for each *Desulfovibrio* strain involved in the pan-genomic analyses. Only 107 and 104 proteins clusters were predicted to be exclusive to alkane producing *Desulfovibrio* strains by the platform Anvi'o and the software *get_homologues* respectively.

It is important to point out that the investigation of this study, to identify of the molecular basis of *Desulfovibrio* alkane production, was led on the assumption that *D. giganteus* DSM 4370 lost the specific set of genes encoding enzymes involved in alkane synthesis. *D. giganteus* DSM 4370 is the non-alkane producing strain clustered in the same clade than the alkane producing strains on the 16S rRNA phylogenic tree. By consequence, the protein clusters exclusively present in alkane producing strains were considered. However, the undetected alkane production of *D. giganteus* DSM 4370 could be due either to the deficient expression of the specific set of genes involved in alkane production or to the presence of non-coding mutations in these genes. Therefore, it is possible that the genes encoding alkane biosynthetic enzymes were absent from the lists of candidate molecular basis generated in this study.

According to this study, 33 protein clusters were considered as being exclusive to alkane producing *Desulfovibrio* with confidence. Among these 33 protein clusters, subunits of a Vacuolar-type ATPase were identified. According to the syntenic information provided by *get_homologues* software, the gene operon for the V-type ATPase contains open reading frames in the order of K, I, D, B, A, C and E subunits. The subunit G and the subunit F characterised in other bacterial V-type ATPase have not been predicted in alkane producing genomes (Takase *et al.,* 1994; Yokoyama *et al.,* 2000). Based on the schematic model of *Enterococcus hirae* V-type ATPase (Saijo *et al.*, 2011), a schematic structural model of the V-type ATPase was drawn with the identified subunits in alkane producing genomes (Figure 4.15).

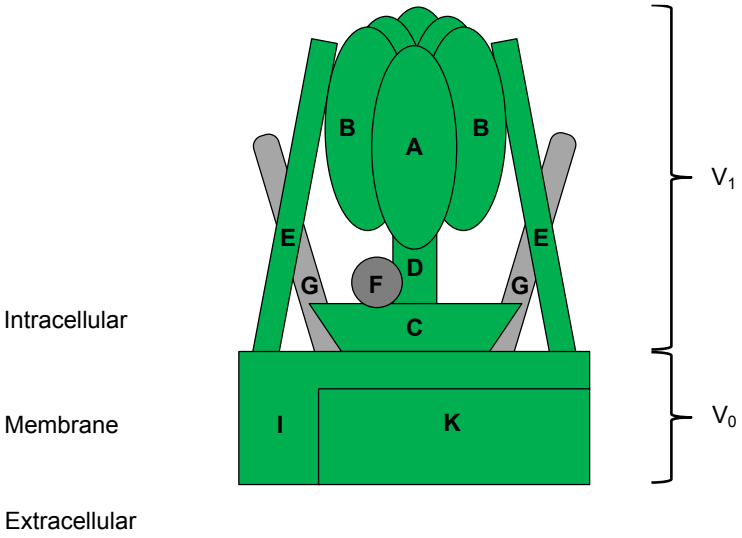


Figure 4.15. Schematic model of the putative V-type ATPase exclusively present in alkane producing *Desulfovibrio* strains, taken and adapted from Saijo *et al.* (2011)

From genome annotation, a V-type ATPase is present in the alkane producing *Desulfovibrio* strains and are composed of seven subunits (coloured in green). Alternating arrangement of A and B subunits constitute the V_1 ATP catalytic domain. The subunit I and K form the V_0 ion-translocating membrane domain. V_1 and V_0 domains are connected by central and peripheral stalks composed of C, D and E subunits. The schematic model of V-type ATPase present in the alkane producing *Desulfovibrio* was adapted from the schematic model of *E. hirae* V-ATPase (Saijo *et al.*, 2011). Subunits coloured in grey were not predicted in the alkane producing *Desulfovibrio* genomes.

Comparing the alkane producing *D. gigas* genome to other *Desulfovibrio* genomes, neither of which belonged to the alkane producing clade on the 16S rRNA phylogenetic tree, Morais-Silva *et al.* reported that a V-type ATPase were exclusively present in *D. gigas* genome (Morais-Silva *et al.,* 2014). This observation endorsed the reliability of the comparative genomic analysis within this study.

According to the hypothetical *Desulfovibrio* alkane biosynthetic pathway within this study, alkanes are produced by hydrogenation of fatty alcohols. Direct hydrogenation of fatty alcohols is unlikely, as it requires the dissociation of the carbon – oxygen covalent bond. Oxygen atoms being more electronegative than carbon atoms, the carbon-oxygen bond is strongly polarised towards oxygen. Therefore, the reductive route leading to alkanes from fatty alcohols presumably involves activated intermediates. V-type ATPases function as primary proton pumps, coupling transfer of protons or cations across the cytoplasmic membrane with ATP hydrolysis (Lolkema *et al.,* 2003). Inorganic phosphates resulting from ATP hydrolysis could be involved in a reaction with fatty alcohols to form alkyl phosphates. The oxygen-phosphate bond would weaken the oxygen-carbon bond. Alkyl phosphates could be the activated intermediates required for the hydrogenation route from fatty alcohols to alkanes (Figure 4.16).

In the list of protein clusters considered exclusive to alkane producing *Desulfovibrio* strains, two clusters, a type 12 methyltransferase (*get_homologues* PC_3584; Anvi'o PC_00001454) and an uncharacterised protein (*get_homologues* PC_3405; Anvi'o PC_00001333), were predicted to share similar structural features with alkane-binding proteins. These predicted structural characteristics led to the hypothesis that the uncharacterised protein (*get_homologues* PC_3405; Anvi'o PC_00001333) could catalyse the hydrogenolysis of alkyl phosphates to alkanes. As for the type 12 methyltransferase (*get_homologues* PC_3584; Anvi'o PC_00001454), it could catalyse the hydrogenolysis of alkyl phosphates and transfer a methyl group to form *iso*-alkanes (Figure 4.16).

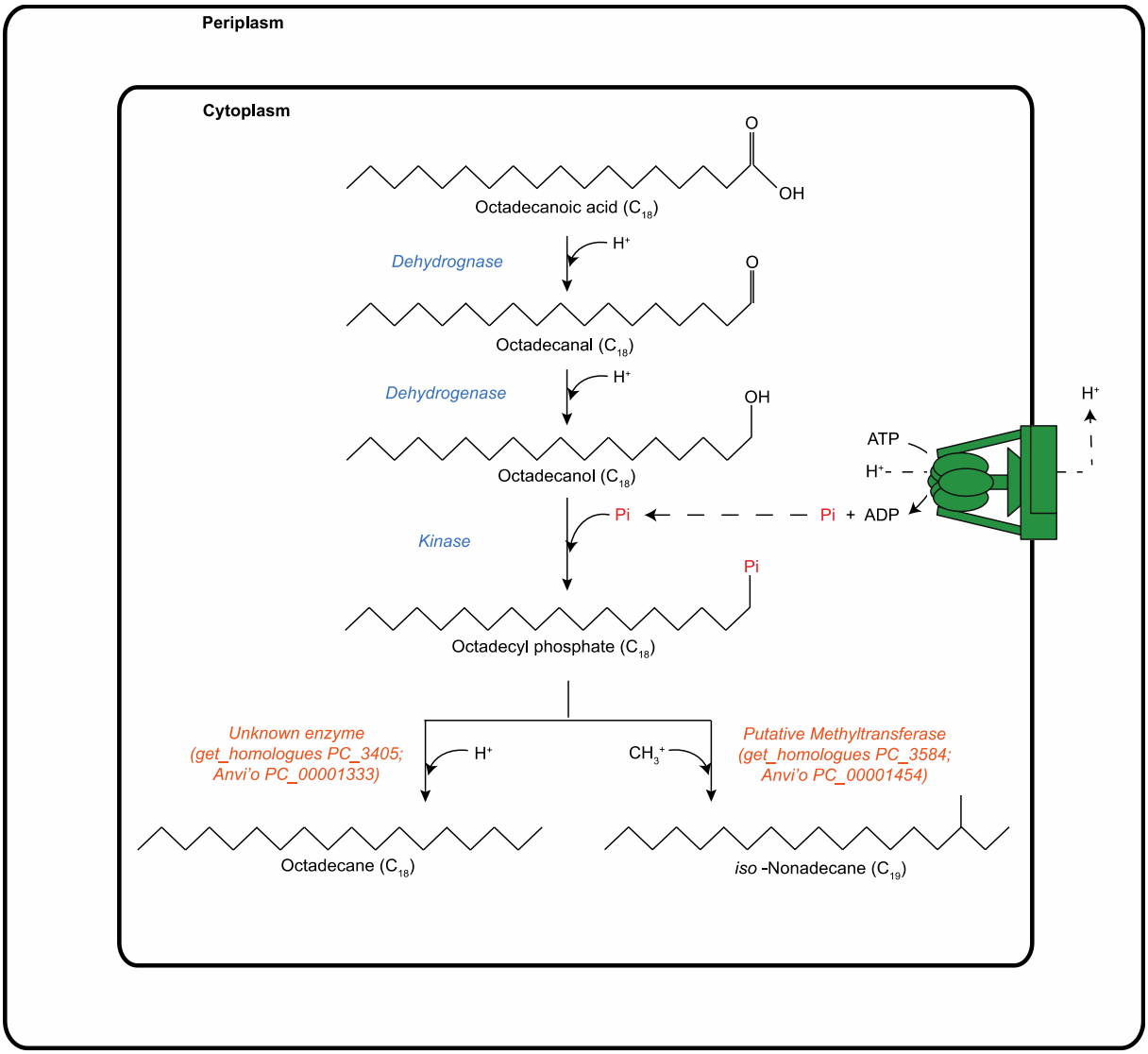


Figure 4.16. Hypothetical pathway for alkane production in *Desulfovibrio,* deduced *in silico* within this study

Alkane production in *Desulfovibrio* would follow a reductive hydrogenation pathway from even-numbered carbon chain fatty acids to even-numbered carbon chain alkane, via alkyl phosphate intermediates. Fatty alcohols generated by reductions from fatty acids would be phosphorylated into their corresponding alkyl phosphates, with a phosphate group provided from ATP hydrolysis catalysed by a V-type ATPase. The alkyl phosphates would be subsequently reduced into the alkanes or *iso*-alkanes.

5. Empirical Approaches for Investigating the Putative Function of the Candidate Molecular Basis for *Desulfovibrio* Alkane Biosynthesis

5.1. Introduction and Abstract

The functional verification of the molecular components, previously identified *in silico* to be putatively involved in the *Desulfovibrio* alkane biosynthetic pathway, was empirically investigated in this chapter. The first approach considered for functional verification aimed to delete the alkane synthesis encoding candidate genes in *Desulfovibrio* spp. To this end, a plasmid vector (pPD3) was designed and constructed to genetically manipulate *Desulfovibrio* spp. Among the mechanisms of genetic transfer in bacteria, transformation by electroporation and conjugation of the alkane producing *Desulfovibrio* strains were investigated for optimisation. As the non-alkane producing *D. vulgaris* Hildenborough was proven several times to be transformable (Van den Berg *et al*., 1989; Fu & Voordouw, 1997; Keller *et al.*, 2009; Chhabra *et al.*, 2011; Vita *et al.*, 2015), *D. vulgaris* Hildenborough transformations by electroporation and conjugations were also performed as a positive control. Gene transfer into an alkane producing *Desulfovibrio* strain by transformation or by conjugation remained unsuccessful within this study, hindering the "knock-out" of candidate genes for alkane production. However, *D. vulgaris* strain Hildenborough transformation by conjugation was successful, which offered the possibility for heterologous expression of candidate genes in a *Desulfovibrio* sp. The approach considered for functional gene verification subsequently progressed towards heterologous expression of the candidate genes in *D. vulgaris* Hildenborough and *E. coli*.

The hypothetical *Desulfovibrio* alkane biosynthetic pathway deduced from the *in silico* genomic comparison within this study (*cf.* Chapter 4 – Comparative Genomic Analysis; Figure 4.16) involved a V-type ATPase, which by catalysing ATP hydrolysis would provide inorganic phosphates for the putative formation of alkyl phosphates from fatty alcohols. *D. vulgaris* Hildenborough and *E. coli* expressing the V-type ATPase operon originating from the alkane producing *D. desulfuricans* NCIMB 8326 would be therefore screened for alkyl phosphate production. Heterologous expression of the V-type ATPase operon in *D. vulgaris* Hildenborough was hampered by the abortive

construction of the designed expression vector. Heterologous expression of the V-type ATPase operon was attempted in *E. coli* hosting the *Photorhabdus luminescens* FAR complex. The expression of the *P. luminescens* FAR complex in *E. coli* was previously reported to increase the amount of cellular fatty acids (Howard *et al.*, 2013). Expression of the V-type ATPase operon in *E. coli* hosting the *P. luminescens* FAR complex proved unsuccessful.

An alternative strategy to verify if alkyl phosphates were the metabolite precursors in *Desulfovibrio* alkane biosynthesis was adopted. This consisted of supplementing the alkane producing *D. desulfuricans* NCIMB 8326 medium with deuterated alkyl phosphates. Isotopically labelled precursors will label its downstream metabolites, as identified by mass spectrometry (Lehmann, 2016). Here was postulated that production of nascent deuterated alkanes by *D. desulfuricans* NCIMB 8326 cultivated in sodium lactate medium supplemented with deuterated alkyl phosphates would confirm that alkyl phosphates were indeed the metabolite precursors in *Desulfovibrio* alkane biosynthesis. However, no nascent deuterated alkanes were detected by GC-MS within the organic compound extracts from *D. desulfuricans* NCIMB 8326 cultivated in sodium lactate medium supplemented with deuterated alkyl phosphates.

The hypothetical *Desulfovibrio* alkane biosynthetic pathway deduced from the *in silico* genomic comparison within this study (*cf.* Chapter 4 – Comparative Genomic Analysis; Figure 4.16) also involved an uncharacterised protein, named as a putative reductase in this study, and a putative methyltransferase. The putative reductase and methyltransferase would catalyse the synthesis of alkanes and *iso*-alkanes respectively from alkyl phosphates. Heterologous expression of the putative reductase and methyltransferase genes were undertaken in *D. vulgaris* Hildenborough and *E. coli*. Expression vectors for *Desulfovibrio* and *E. coli*, carrying the putative histidine-tagged reductase and methyltransferase genes separately, were successfully constructed and validated by sequencing. Expression of either the putative reductase or methyltransferase gene in *D. vulgaris* Hildenborough was unsuccessful. However, expression of the putative reductase and methyltransferase gene in *E. coli* cultivated under aerobic conditions was achieved. Organic compounds from *E. coli* which expressed either the putative reductase or methyltransferase gene, cultured in LB medium supplemented with deuterated alkyl phosphates under aerobic conditions, were screened by GC-MS for unlabelled alkanes and deuterated alkanes. Neither

unlabelled alkanes nor deuterated alkanes were detected by GC-MS within these organic compound extracts.

The putative reductase and methyltransferase genes of interest originated from *D. desulfuricans* NCIMB 8326, an anaerobe. Therefore, the putative reductase and methyltransferase are potentially sensitive to oxygen. *E. coli* hosting either the putative reductase or methyltransferase gene were thus cultured under anaerobic conditions in LB medium supplemented with deuterated alkyl phosphates. Only the expression of the putative methyltransferase gene in *E. coli* cultured under anaerobic conditions was observed. Organic compounds from *E. coli* which expressed the putative methyltransferase gene, cultured in LB medium supplemented with deuterated alkyl phosphates under anaerobic conditions were screened for unlabelled alkanes and deuterated alkanes. Neither unlabelled alkanes nor deuterated alkanes were detected by GC-MS within these organic compound extracts.

The deuterated alkyl phosphates are composed of a non-polar long alkyl group with a polar phosphate extremity. Due to these structural characteristics, it is probable that the alkyl phosphates interact with the phospholipid of *E. coli* membranes. Therefore, it is uncertain that the deuterated alkyl phosphates penetrated *E. coli* cells. To bypass the bilayer phospholipid barriers, the putative reductase and methyltransferase genes were firstly expressed in *E. coli* cultured under aerobic conditions, prior to protein extraction. Only the expression of the putative methyltransferase gene in *E. coli* was observed. Deuterated alkyl phosphates were subsequently added to the whole cell protein extracts. The whole cell protein extracts were screened for unlabelled alkanes and deuterated alkanes after 24 h and 48 h incubation at 37 °C. No unlabelled alkanes and deuterated alkanes were detected by GC-MS either after 24 h or after 48 h incubation within the whole cell protein extracts including the putative methyltransferase.

Recently, a new alkane biosynthetic pathway catalysed by a fatty acid photodecarboxylase was reported to be light sensitive. The fatty acid photodecarboxylase was demonstrated to be active in blue light while its enzymatic activity stopped in the dark or red light (Sorigué *et al.*, 2017). Among the 33 protein clusters considered with confidence as exclusive to alkane producing *Desulfovibrio* (*cf.* Chapter 4 – Comparative Genomic Analysis; Table 4.11), a protein annotated as a blue-light-activated sensor protein is listed. A novel hypothesis was therefore proposed that *Desulfovibrio* alkane biosynthesis was affected by the dark. To verify this novel

hypothesis, the alkane producing *D. desulfuricans* NCIMB 8326 was cultivated either in the daylight or in the dark prior to being screened for alkane production. *D. desulfuricans* NCIMB 8326 was proven to synthesise alkanes in the same quantity in the dark or in the light.

5.2. Development of Molecular Biology Tools and Methods for Desulfovibrio Engineering

5.2.1. Design and Construction of the Plasmid pPD3, Vector for *Desulfovibrio* Engineering

To genetically manipulate alkane producing *Desulfovibrio* strains, a plasmid vector (pPD3) was designed and constructed for gene transfer into *Desulfovibrio* spp. by transformation and conjugation.

Kanamycin resistance gene as a selectable marker for gene transfer into alkane producing Desulfovibrio strains

The choice of a selectable marker for gene transfer into *Desulfovibrio* was based on the established geneticin sensitivity of *D. vulgaris* strain Hildenborough. *D. vulgaris* Hildenborough was proven sensitive to geneticin at 400 μg ml^{-1} (Bender *et al.*, 2006). Alkane producing *Desulfovibrio* strains were thus screened for geneticin sensitivity at 400 μg ml^{-1}, 500 μg ml^{-1} and 600 μg ml^{-1}. No evidence of *Desulfovibrio* growth was observed on sodium lactate medium plate supplemented with any of these three concentrations of geneticin, after 7 days of incubation, whilst *Desulfovibrio* spp. grew on sodium lactate medium without antibiotic. The alkane producing *Desulfovibrio* strains within this study were thus sensitive to geneticin at a concentration of 400 μg ml^{-1} and above. A kanamycin resistance gene conferring geneticin resistance was consequently chosen as a selectable marker for positive selection of *Desulfovibrio* cells which have undergone gene transfer. Kanamycin resistance genes code an aminoglycoside phosphotransferase, which inactivates aminoglycoside antibiotics, such as geneticin (Shaw *et al.*, 1993).

Cloning strategies for construction of the plasmid pPD3

The plasmid pSB1K3-RFP provided the kanamycin resistance gene (BioBrick part BBa_P1003) and the replicon pMB1 for plasmid replication in *E. coli*. pSB1K3-RFP also contains the gene coding the red fluorescent protein (RFP). The first step of pPD3 construction was the substitution of the *rfp* gene in pSB1K3-RFP (BioBrick part BBa-

J05540-RFP) with a synthesised multiple cloning site, generating the plasmid pPD1 (Figure 5.1).

An 'oriT' region, containing the Nic region, the *traJ* gene and a truncated *tra*I gene, required for plasmid conjugal transfer, was amplified from the plasmid pS797 and cloned into pPD1, generating the mobilisable plasmid pPD2 (Figure 5.2).

Finally, the cryptic plasmid pBG1 of *D. desulfuricans* G100A was excised from pMO719 and cloned into pPD2 to allow plasmid replication in *Desulfovibrio*. The resulting shuttle plasmid pPD3 was validated by sequencing (Figure 5.3).

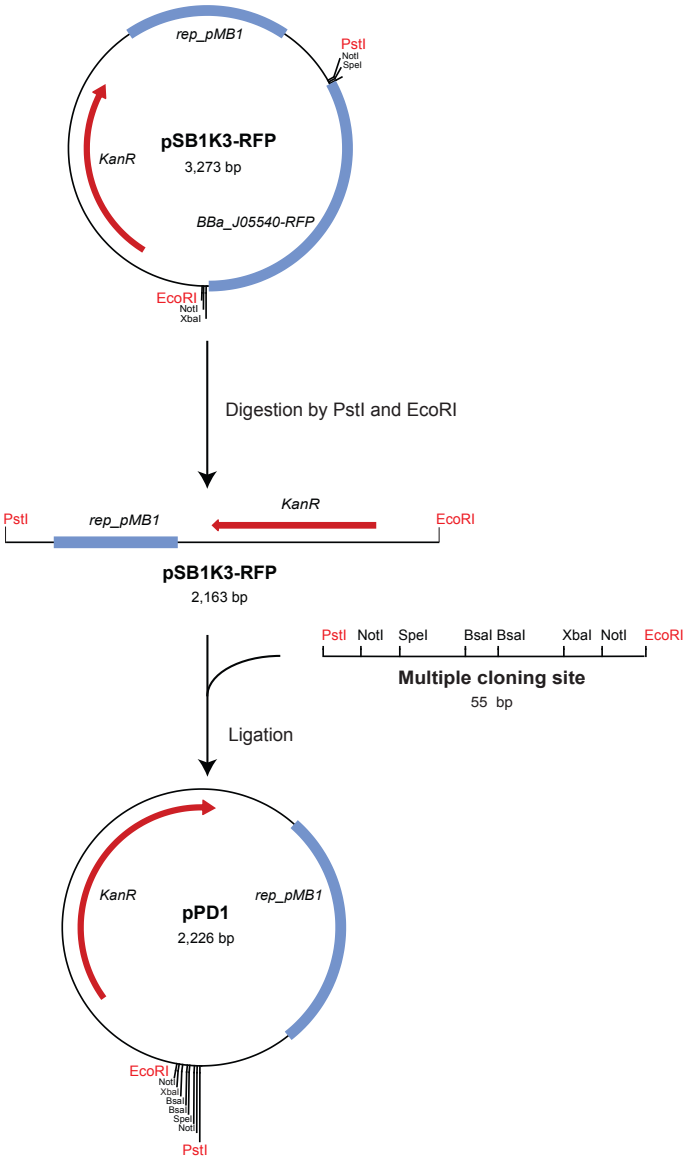


Figure 5.1. Cloning strategy used for construction of the plasmid pPD1

The plasmid pPD1 was constructed by insertion of an EcoRI-PstI cut multiple cloning site into pSB1K3-RFP cut with the same restriction endonucleases. EcoRI and PstI restriction sites of the multiple cloning site, pSB1K3-RFP and pPD1 are annotated in red. The multiple cloning site was generated by annealing MCS-F and MCS-R, two 70 bp single strand nucleic acid oligomers.

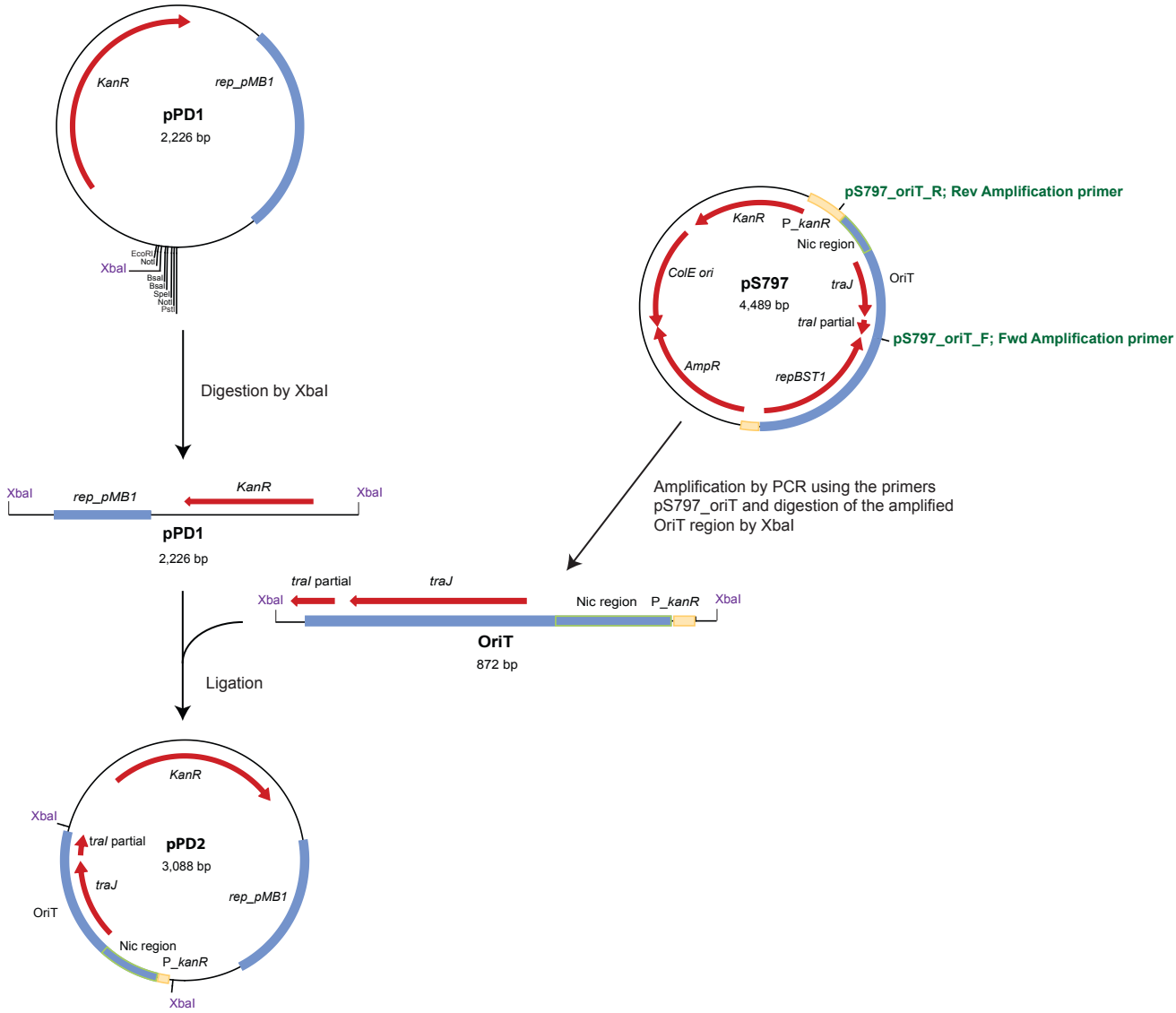


Figure 5.2. Cloning strategy used for construction of the mobilisable plasmid pPD2

The mobilisable plasmid pPD2 was constructed by ligation of the oriT region from pS797, into pPD1. The oriT region was amplified from pS797 by PCR using the primers pS797-oriT_F and pS797-oriT_R (primers written in green). The primers pS797_oriT were designed with flanking regions containing an XbaI restriction site. pPD1 and the amplified region containing oriT were then cut by XbaI, prior to ligation. XbaI restriction sites of pPD1, the amplified region containing oriT and pPD2 are annotated in violet.

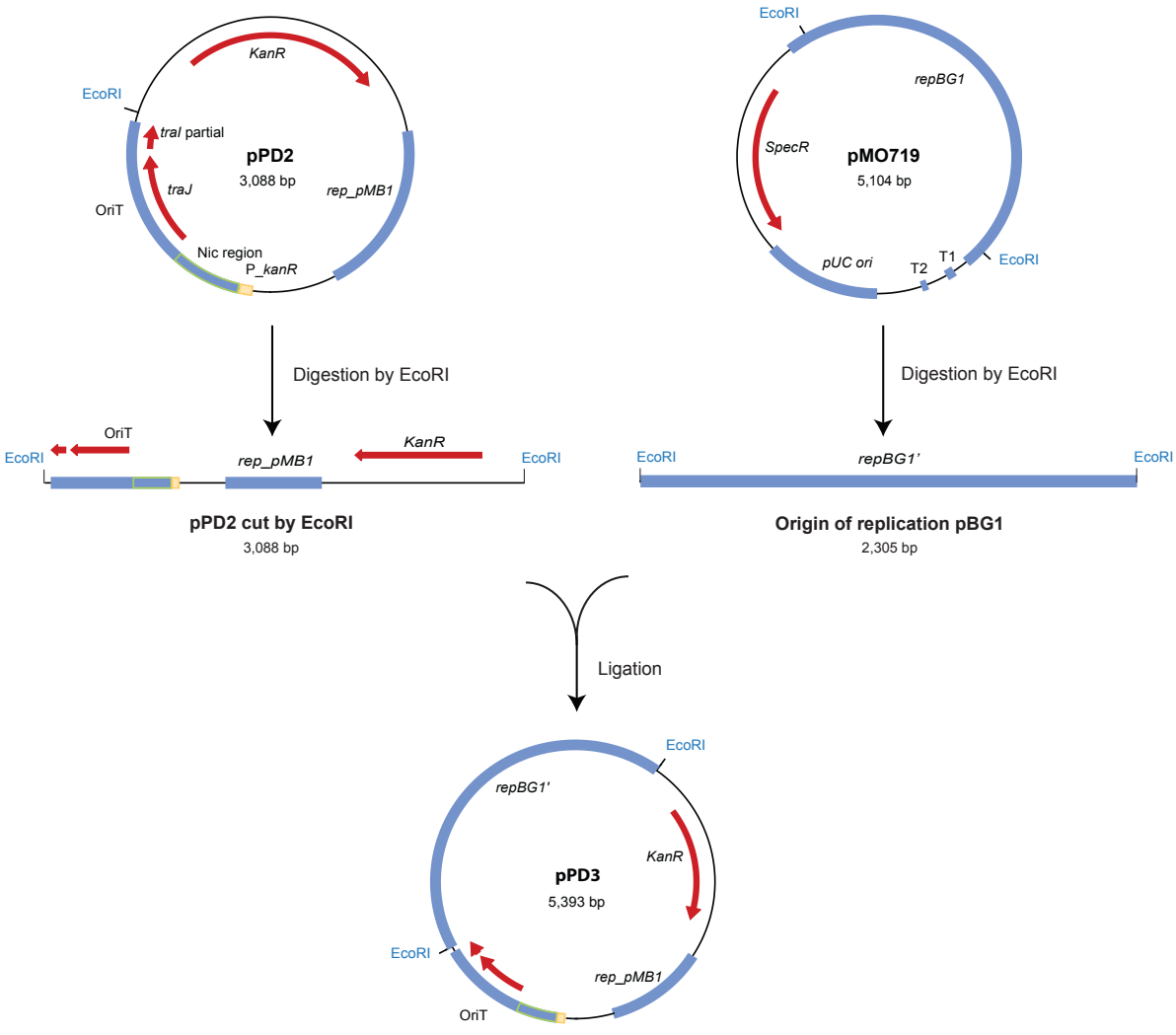


Figure 5.3. Cloning strategy used for construction of the shuttle plasmid pPD3

The shuttle plasmid pPD3 was generated by ligation of the *Desulfovibrio* plasmid pBG1 into pPD2. pBG1 was excised from pMO719 by EcoRI digestion and ligated into pPD2, previously cut by EcoRI. EcoRI restriction sites of pPD2, pMO719 and pPD3 are annotated in blue.

5.2.2. Optimisation of a Genetic Transfer Protocol for Alkane Producing *Desulfovibrio Engineering*

Desulfovibrio transformation by electroporation

Desulfovibrio cells were made electro-competent by 2 washes in 30 mM Tris-HCl buffer at pH 7.2, a protocol successfully used with *D. vulgaris* Hildenborough (Chhabra *et al.*, 2011). Arcing occurred during electroporation of *D. desulfuricans* NCIMB 8326, *D. desulfuricans* NCIMB 8338, *D. gabonensis* DSM 10636 and *D. marinus* DSM 18311. Contrary to *D. vulgaris* Hildenborough which grows in non-salt supplemented sodium lactate medium, *D. desulfuricans* NCIMB 8326, *D. desulfuricans* NCIMB 8338, *D. gabonensis* DSM 10636 and *D. marinus* DSM 18311 were cultured in a sodium lactate medium supplemented with 2.5 % (w/v) NaCl. Arcing can be due to a high concentration of salts in the electroporation suspension. To reduce the chance of arcing, *Desulfovibrio* cells were therefore washed up to five times using 30 mM Tris-HCl buffer at pH 7.2. After five washes, no electrical discharge was observed during *Desulfovibrio* electroporation.

Desulfovibrio electroporation was performed with different amounts (0.1 *µ*g, 0.5 *µ*g, 0.75 *µ*g, 1 *µ*g, 5 *µ*g and 10 *µ*g) of plasmid pPD3 added to the electro-competent cells and with different field strengths (5 kV cm^{-1}, 7.5 kV cm^{-1}, 10 kV cm^{-1}, 12.5 kV cm^{-1}, 15 kV cm^{-1}, 17.5 kV cm^{-1} and 20 kV cm^{-1}) applied to the electro-competent cells. The resistance and the capacitance of the electroporator were respectively set up at 25 *µ*F and 200 Ω to 400 Ω for a pulse length between 5 ms to 10 ms. For each set of conditions applied and for each *Desulfovibrio* strains, no microbial growth was observed on sodium lactate medium supplemented with geneticin, whilst *Desulfovibrio* colonies were observed only on medium without antibiotic. *Desulfovibrio* electroporation remained unsuccessful.

Desulfovibrio conjugation with E. coli strain MW3064

Desulfovibrio conjugation was performed with the conjugative strain WM3064 *E. coli*. WM3064 *E. coli* is auxotrophic for DAP, which allowed the counter-selection against *E. coli* cell donors by depriving the medium of DAP after conjugation with *Desulfovibrio* spp.

Three days after the mating event, transformed *D. vulgaris* Hildenborough colonies were observed on selective agar containing geneticin. The black colour of the colonies indicated that the colonies belong to *Desulfovibrio* genus. Moreover, the selective medium supplemented with 400 μg.ml^{-1} of geneticin and deprived of DAP ensured the prevention of the growth of WM3064 *E. coli* containing pPD3. The negative control, the wild-type *D. vulgaris* Hildenborough cells failed to grow on selective agar.

Liquid cultures of sodium lactate medium supplemented with 400 μg.ml^{-1} geneticin and deprived of DAP were inoculated with a *D. vulgaris* Hildenborough transformant single colony. After 3 days of incubation, plasmids were purified from *D. vulgaris* Hildenborough transformant cultures. Plasmids purified from *D. vulgaris* Hildenborough transformants were used as template for PCR using primers specific to the plasmid pPD3 (Figure 5.4). A 1.6 kb fragment was amplified from the plasmid pPD3 and from the plasmids of *D. vulgaris* Hildenborough transformants (Figure 5.5). No DNA was amplified from *D. vulgaris* Hildenborough genomic DNA, used as negative control (Figure 5.5). Moreover, plasmids purified from *D. vulgaris* Hildenborough transformants were digested by EcoRI and by PstI to allow fragment sizing. Plasmids purified from *D. vulgaris* Hildenborough transformants showed the same digestion fragment profiles compared to pPD3 digested by the same endonucleases (Figure 5.6); confirming that the transfer of pPD3 from *E. coli* WM3064 to *D. vulgaris* Hildenborough was successful.

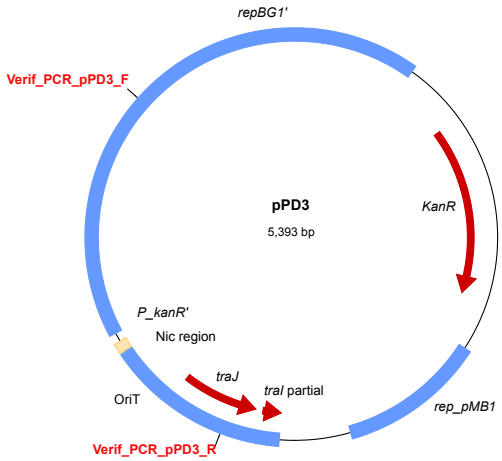


Figure 5.4. Map of the shuttle vector pPD3 with the primers used for verification of pPD3 transfer into *D. vulgaris* Hildenborough

Position of the primers Verif_PCR_pPD3_F and Verif_PCR_pPD3_R (written in red) are shown on the plasmid map.

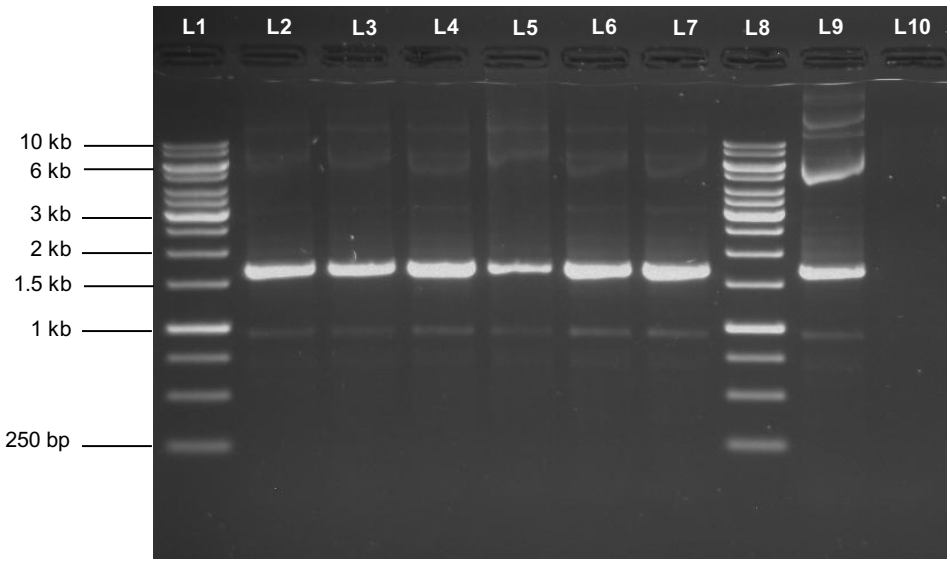


Figure 5.5. PCR amplification profiles of plasmids purified from *D. vulgaris* Hildenborough transformants, the plasmid pPD3 and *D. vulgaris* Hildenborough genomic DNA

Plasmids purified from *D. vulgaris* Hildenborough transformants, the plasmid pPD3 and *D. vulgaris* Hildenborough genomic DNA were used as template for PCR using the primers Verif_PCR_pPD3_F and Verif_PCR_pPD3_R. Amplified DNA fragments were separated on a 1 % agarose gel by electrophoresis for 1 h at 130 V. A 1.6 kb DNA fragment is amplified by PCR from plasmids purified from *D. vulgaris* Hildenborough transformants (L2 to L7), and from pPD3 (used as positive control; L9). No DNA amplification was observed from *D. vulgaris* Hildenborough genomic DNA, used as negative control (L10). GeneRuler 1 kb DNA ladder (ThermoFisher Scientific; L1 and L8).

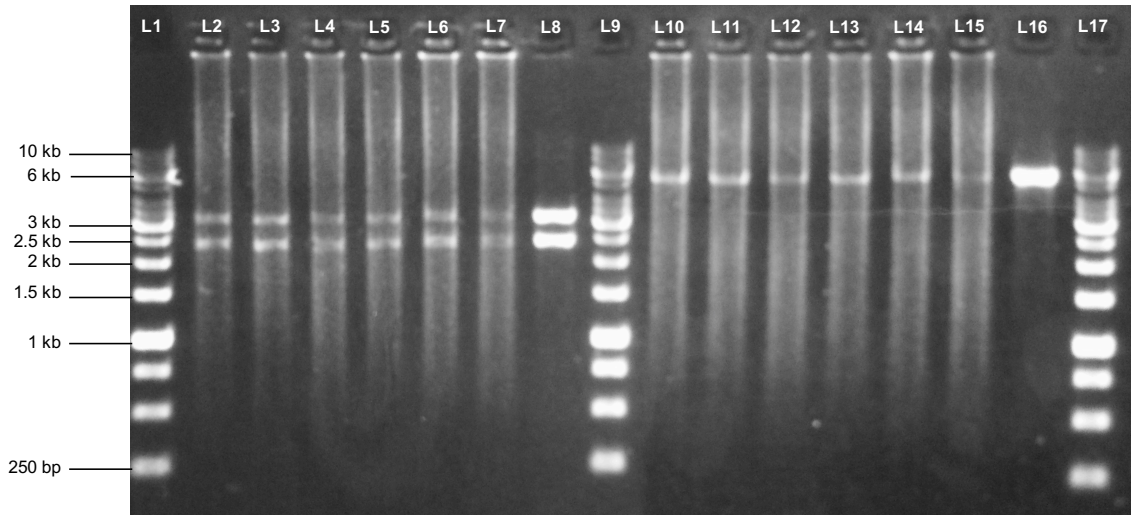


Figure 5.6. EcoRI and PstI Digestion profiles of plasmids purified from *D. vulgaris* Hildenborough transformants and the plasmid pPD3

Plasmids purified from *D. vulgaris* Hildenborough transformants and the plasmid pPD3 were digested by EcoRI and PstI. Digested products were separated on a 1 % agarose gel by electrophoresis for 1 h at 130 V. EcoRI digestion of plasmids purified from *D. vulgaris* Hildenborough transformants resulted in two DNA fragments of approximatively 3,000 bp and 2,000 bp (L2 to L7). Digestion of pPD3 by EcoRI cuts the plasmid into two DNA fragments of 3,088 bp and 2,305 bp (L8). PstI digestion of plasmids purified from the *D. vulgaris* Hildenborough transformants resulted in one DNA fragment, with a size comprised between 5,000 bp and 6,000 bp (L10 to L15). PstI digestion of pPD3 linearises the plasmid of 5,393 bp (L16). GeneRuler 1 kb DNA ladder (ThermoFisher Scientific; L1, L9 and L17).

One week after the mating event, no bacterial growth of alkane producing *Desulfovibrio* strains which had undergone the conjugation was observed on medium supplemented with geneticin. In contrast, *D*esulfovibrio colonies were observed on agar without geneticin. *Desulfovibrio* conjugation with WM3064 *E. coli* was repeated. Alkane producing strain transformation by conjugation remained unsuccessful whilst *D. vulgaris* Hildenborough transformation was successfully repeated.

In summary, gene transfer into the alkane producing *Desulfovibrio* strains was not achieved within this study, which hindered the "knock out" strategy for functional verification of candidate genes in an alkane producing *Desulfovibrio* strain. However, the successful *D. vulgaris* strain Hildenborough transformation by conjugation offered the possibility for heterologous expression of candidate genes in a *Desulfovibrio* sp. Therefore, the approach considered for functional verification of candidate genes progressed toward heterologous expression of the candidate genes in *D. vulgaris* strain Hildenborough and *E. coli*.

5.3. Empirical Investigation of the V-type ATPase Function in Desulfovibrio Alkane Biosynthesis

According to the hypothetical *Desulfovibrio* alkane biosynthetic pathway deduced from the *in silico* genomic comparison (*cf.* Chapter 4 – Comparative Genomic Analysis; Figure 4.16), a V-type ATPase would catalyse ATP hydrolysis and the resulting inorganic phosphates would be involved in a reaction with fatty alcohols to form alkyl phosphates (Figure 5.7). Alkyl phosphates were hypothesised to be the activated metabolite intermediates required for the hydrogenation route from fatty alcohols to alkanes.

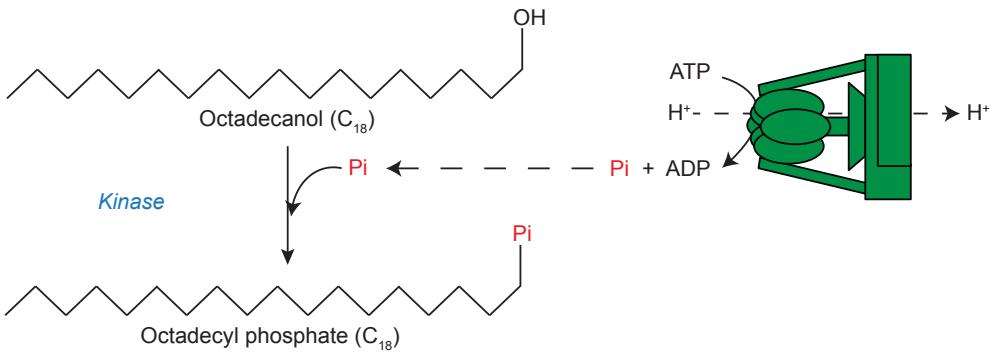


Figure 5.7. Hypothetical role of a V-type ATPase in *Desulfovibrio* alkane biosynthesis

To verify the putative role of the V-type ATPase in *Desulfovibrio* alkane biosynthesis, heterologous expression of the V-type ATPase operon originating from the alkane producing *D. desulfuricans* NCIMB 8326 was attempted in *D. vulgaris* Hildenborough and *E. coli* to screen the transformants for alkyl phosphate production.

5.3.1. Heterologous Expression of the V-type ATPase in *D. vulgaris* Hildenborough

Heterologous expression of the V-type ATPase in *D. vulgaris* Hildenborough was hampered by the abortive construction of the pPD3-VtATPase expression vector. The cloning strategy involved the ligation of four DNA fragments of various sizes. Thus it proved problematic to achieve an optimal insert:vector molar ratio for this ligation.

5.3.2. Heterologous Expression of the V-type ATPase in *E. coli* hosting the *P. luminescens* FAR

The plasmid pEC-K-VtATPase, carrying the histidine-tagged V-type ATPase operon, was transferred into BL21 Star (DE3) *E. coli* previously transformed with the plasmid pCED. The plasmid pCED carries the genes *luxC*, *luxE*, and *luxD* from *P. luminescens* encoding a FAR complex. Expression of the histidine-tagged V-type ATPase operon in *E. coli* hosting the *P. luminescens* FAR complex was induced by addition of IPTG in triplicate cultures, incubated under aerobic conditions. The V-type ATPase operon expression in *E. coli* was verified by SDS-PAGE and Western blotting (Figure 5.8). The V-type ATPase operon subunit, with the N-terminal fused to a poly-histidine tag, was predicted to have a molecular weight of 12.95 kDa using the ProtParam online tool. No fluorescent band was detected on the Western Blot of the whole cell protein extracts from *E. coli* transformed with pCED and pEC-K-VtATPase. Absence of fluorescent signal was not due to a defective Western blot procedure as a fluorescent signal was detected for the positive control. This observation suggested that the V-type ATPase was not produced by *E. coli* transformed with the plasmids pCED and pEC-K-VtATPase, cultured under these conditions.

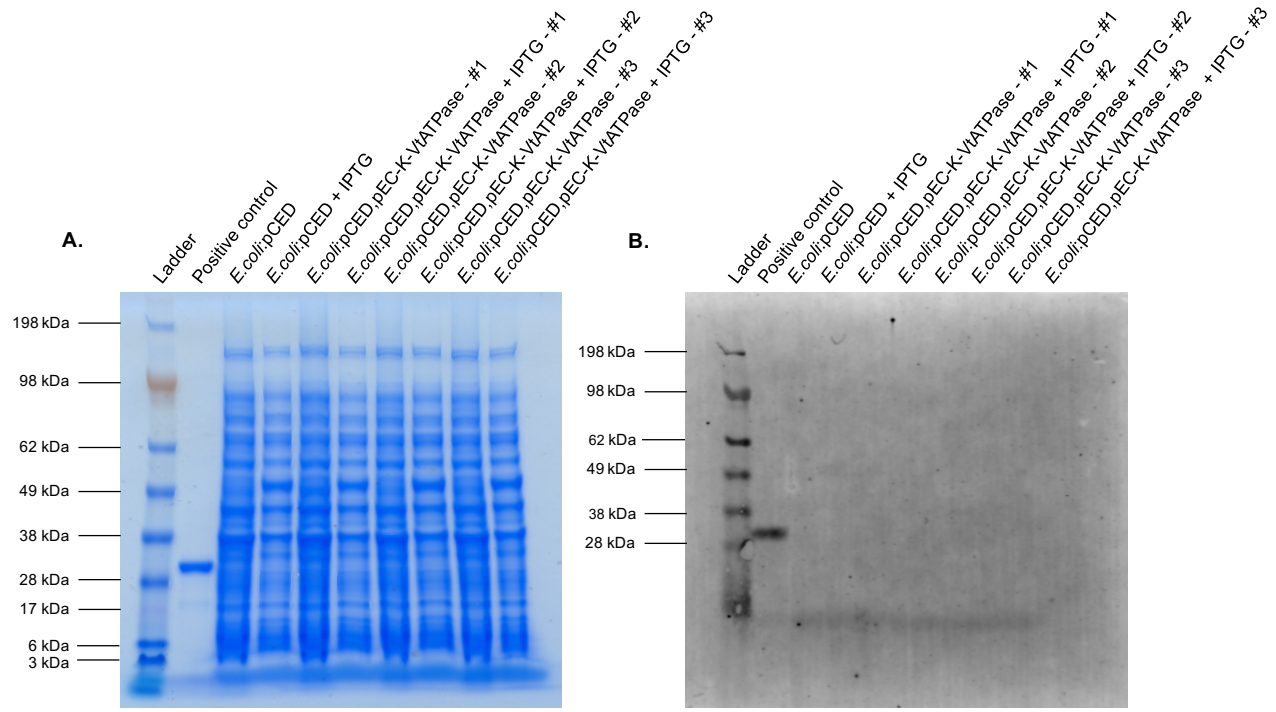


Figure 5.8. Expression profiles of the V-type ATPase in *E. coli*

E. coli transformed with pCED (*E.coli*:pCED) and *E. coli* transformed with pCED and pEC-K-VtATPase (*E.coli*:pCED,pEC-K-VtATPase) were aerobically cultured in triplicate (#1, #2, #3) in LB medium at 37°C, shaking at 220 rpm to an OD of 0.7, at 600 nm. Plasmid gene expression was then induced by addition of 200 μM IPTG where appropriate. Cultures were incubated for another 4 h, prior to cell lysis. Proteins from whole cell extracts were separated by SDS-PAGE on a 4-20 % gradient Bis-Tris polyacrylamide gel, for 45 min, at 140 V in 1-fold MOPS running buffer, followed by Coomassie Blue staining (A). The positive control used was a purified histidine-tagged esterase (31.1 kDa). SeeBlue™ Plus2 Pre-stained Protein Standard (ThermoFisher Scientific) was used as protein ladder. Proteins were transferred from an unstained polyacrylamide gel to a PVDF membrane for Western blot analysis (B). The blot was probed with a mouse anti-His Tag primary antibody and with a goat anti-mouse secondary antibody conjugated to the IRDye 680RD fluorophore. The IRDye 680RD fluorophore was excited at 676 nm and fluorescence detected at 700 nm.

5.3.3. Enrichment of the Alkane Producing *D. desulfuricans* NCIMB 8326 Medium with Deuterated Alkyl Phosphates

Deuterated octadecyl-1,1-d_2-phosphate, hexadecyl-d_{33}-phosphate and tetradecyl-d_{29}-phosphate were chosen to supplement *Desulfovibrio* sodium lactate medium. According to the hypothetical *Desulfovibrio* alkane biosynthetic pathway*, D. desulfuricans* NCIMB 8326 cultured in sodium lactate medium supplemented with either octadecyl-1,1-d_2-phosphate or hexadecyl-d_{33}-phosphate or tetradecyl-d_{29}-phosphate would produce nascent octadecane-d_2, hexadecane-d_{33} and tetradecane-d_{29} respectively (Figure 5.9).

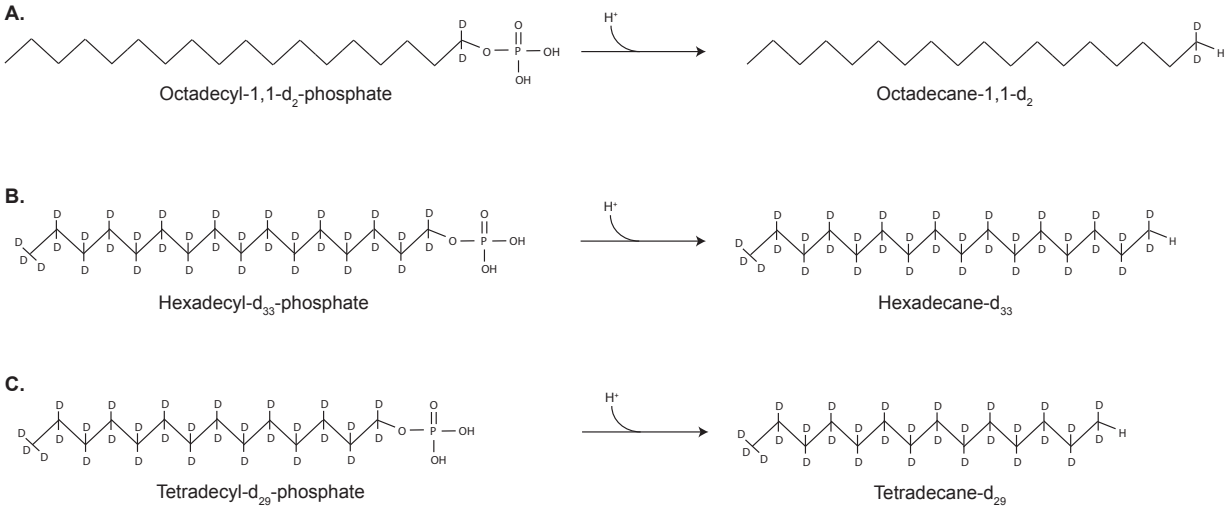


Figure 5.9. Hypothetical reduction reactions from octadecyl-1,1-d_2-phosphate, hexadecyl-d_{33}-phosphate and tetradecyl-d_{29}-phosphate to octadecane-1,1-d_2, hexadecane-d_{33} and tetradecane-d_{29}

According to the hypothetical Desulfovibrio alkane pathway within this study, octadecyl-1,1-d_2-phosphate, hexadecyl-d_{33}-phosphate and tetradecyl-d_{29}-phosphate would be reduced to octadecane-1,1-d_2 (A), hexadecane-d_{33} (B) and tetradecane-d_{29} (C) respectively.

Dissolution of octadecyl-1,1-d_2-phosphate, hexadecyl-d_{33}-phosphate and tetradecyl-d_{29}-phosphate in anhydrous ethanol

To supplement *Desulfovibrio* sodium lactate medium with deuterated alkyl phosphates, dissolution of octadecyl-1,1-d_2-phosphate, hexadecyl-d_{33}-phosphate and tetradecyl-d_{29}-phosphate in water was attempted at different concentrations (60 mM, 50 mM, 40 mM, 30 mM, 20 mM, 10 mM, 5 mM and 1 mM). The deuterated alkyl phosphates were insoluble in water, probably due to their non-polar long alkyl group. To increase the deuterated alkyl phosphate solubility in water, cyclodextrins were used as complexing agents. β-cyclodextrins and δ-cyclodextrins were chosen for their cavity size which can incorporate large molecules. Solutions of β-cyclodextrins or δ-cyclodextrins to a final concentration of 50 mM, 100 mM or 150 mM mixed with deuterated alkyl phosphates to a final concentration of 50 mM, 40 mM, 30 mM, 20 mM, 10 mM, 5 mM and 1 mM were prepared in water. Deuterated alkyl phosphates in presence of cyclodextrins did not dissolve in water.

Attempts to dissolve octadecyl-1,1-d_2-phosphate, hexadecyl-d_{33}-phosphate and tetradecyl-d_{29}-phosphate in organic solvents such as acetone, diethyl ether, anhydrous methanol and anhydrous ethanol were carried out at different concentrations. The deuterated alkyl phosphates were soluble only in anhydrous ethanol to a final concentration of 1 mM for octadecyl-1,1-d_2-phosphate and hexadecyl-d_{33}-phosphate, and to a final concentration of 10 mM for tetradecyl-d_{29}-phosphate.

Screening of D. desulfuricans NCIMB 8326 cultured in sodium lactate medium enriched with deuterated alkyl phosphates solution for deuterated Alkanes

Organic compounds from *D. desulfuricans* NCIMB 8326 cultivated in sodium lactate medium in the presence of 10 *μ*M octadecyl-1,1-d_2-phosphate were screened for nascent octadecane-1,1-d_2.

The mass spectrum of unlabelled alkanes is composed by preponderance of butyl, pentyl and hexyl. As octadecane-1,1-d_2 carries two deuterium atoms, the mass spectrum of octadecane-1,1-d_2 is composed by preponderance of deuterated butyl ($C_4H_7D_2^+$), pentyl ($C_5H_9D_2^+$) and hexyl ($C_6H_{11}D_2^+$), with a molecular mass of 59 g mol^{-1}, 73 g mol^{-1} and 87 g mol^{-1} respectively. Organic compounds from *D. desulfuricans* NCIMB 8326 grown in sodium lactate medium supplemented with 1 % (v/v) octadecyl-1,1-d_2-phosphate solution were thus screened by GC-MS for 59 m/z ion, 73 m/z ion and 87 m/z ion fragments, in order of preponderance. One compound (other than octadecyl-1,1-d_2-phosphate) eluted at 673.3 s ionised into 59 m/z ion, 73 m/z ion and 87 m/z ion fragments by preponderance (Figure 5.10A).

Importantly, no compounds eluted at 673.3 s were detected within the octadecyl-1,1-d_2-phosphate solution. The compound eluted at 673.3 s from the organic compounds of *D. desulfuricans* NCIMB 8326 cultivated in sodium lactate medium supplemented with octadecyl-1,1-d_2-phosphate solution was therefore not due to contamination derived from the octadecyl-1,1-d_2-phosphate solution (Figure 5.10B).

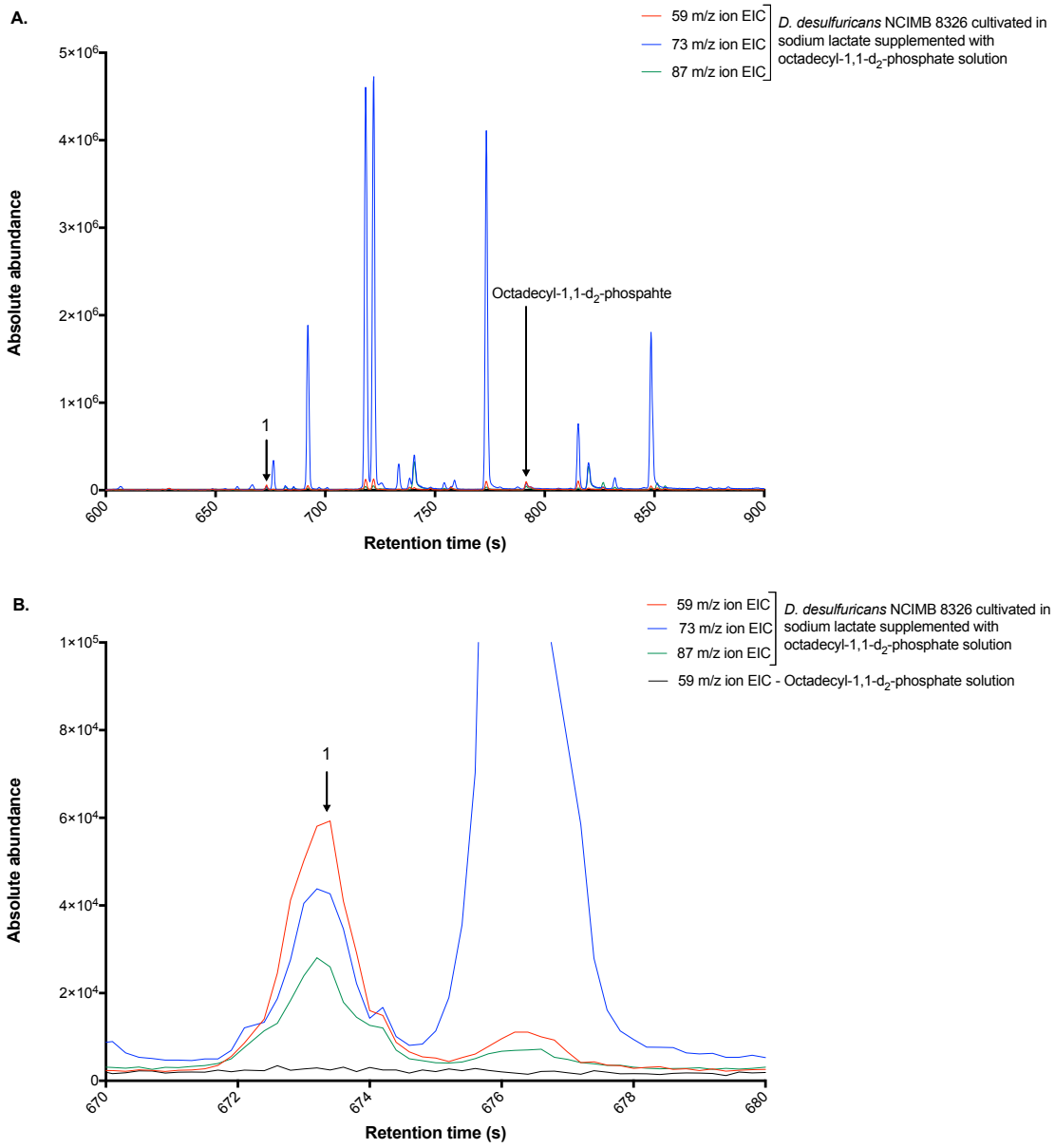


Figure 5.10. Detection of a putative deuterated alkane within the organic compounds of *D. desulfuricans* NCIMB 8326 cultivated in sodium lactate medium supplemented with octadecyl-1,1-d_2-phosphate solution

Organic compounds from *D. desulfuricans* NCIMB 8326 cultivated in sodium lactate medium supplemented with 1 % (v/v) 1 mM octadecyl-1,1-d_2-phosphate solution were screened for deuterated alkanes by GC-MS. The extracted 59 m/z, 73 m/z and 87 m/z ion chromatograms (EIC) of organic compounds from *D. desulfuricans* NCIMB 8326 revealed that one compound (1; indicated by an arrow) ionised into 59 m/z, 73 m/z and 87 m/z ion fragments by order of preponderance (A). The octadecyl-1,1-d_2-phosphate solution was also analysed by GC-MS. The superimposition of the 59 m/z EIC of the octadecyl-1,1-d_2-phosphate solution on the EIC of organic compounds from *D. desulfuricans* NCIMB 8326 revealed that this putative deuterated alkane eluted at 673.3 s was not present within the octadecyl-1,1-d_2-phosphate solution (B).

The mass spectrum of the compound eluted at 673.3 s detected within the organic compounds of *D. desulfuricans* NCIMB 8326 cultivated in sodium lactate medium supplemented with octadecyl-1,1-d_2-phosphate solution showed the fragmentation pattern of an unlabelled alkane. According to the retention time, unlabelled alkanes eluted at 673.3 s correspond to octadecanes. As deuterium is naturally present, deuterated butyl ($C_4H_7D_2^+$), pentyl ($C_5H_9D_2^+$) and hexyl ($C_6H_{11}D_2^+$) can be observed in the mass spectrum of unlabelled alkanes. Therefore to determinate if *D. desulfuricans* NCIMB 8326 cultivated in sodium lactate medium supplemented with octadecyl-1,1-d_2-phosphate solution synthesised nascent octadecane-1,1-d_2, the 59 m/z ion to 57 m/z ion peak area ratio of octadecanes detected within the organic compounds of *D. desulfuricans* NCIMB 8326 cultivated in the presence of octadecyl-1,1-d_2-phosphate was calculated and compared the same isotopic ratio of octadecanes detected within the organic compounds of *D. desulfuricans* NCIMB 8326 cultivated in the absence of octadecyl-1,1-d_2-phosphate (Figure 5.11). Biosynthesis of nascent octadecane-1,1-d_2 by *D. desulfuricans* NCIMB 8326 would be confirmed if the 59 m/z ion to mz57 ion peak area ratio of octadecanes detected within the organic compounds of *D. desulfuricans* NCIMB 8326 cultivated in the presence of octadecyl-1,1-d_2-phosphate was significantly higher than the same isotopic ratio of octadecanes detected within the organic compounds of *D. desulfuricans* NCIMB 8326 cultivated in the absence of octadecyl-1,1-d_2-phosphate.

The 59 m/z ion to 57 m/z ion peak area ratios of octadecanes detected within the organic compounds of *D. desulfuricans* NCIMB 8326 cultivated either in the absence or in presence of octadecyl-1,1-d_2-phosphate were not significantly different (t-test's $p > 0.05$). Therefore, biosynthesis of nascent octadecane-1,1-d_2 by *D. desulfuricans* NCIMB 8326 cultivated in sodium lactate medium supplemented with octadecyl-1,1-d_2-phosphate solution was not detected by GC-MS under these conditions.

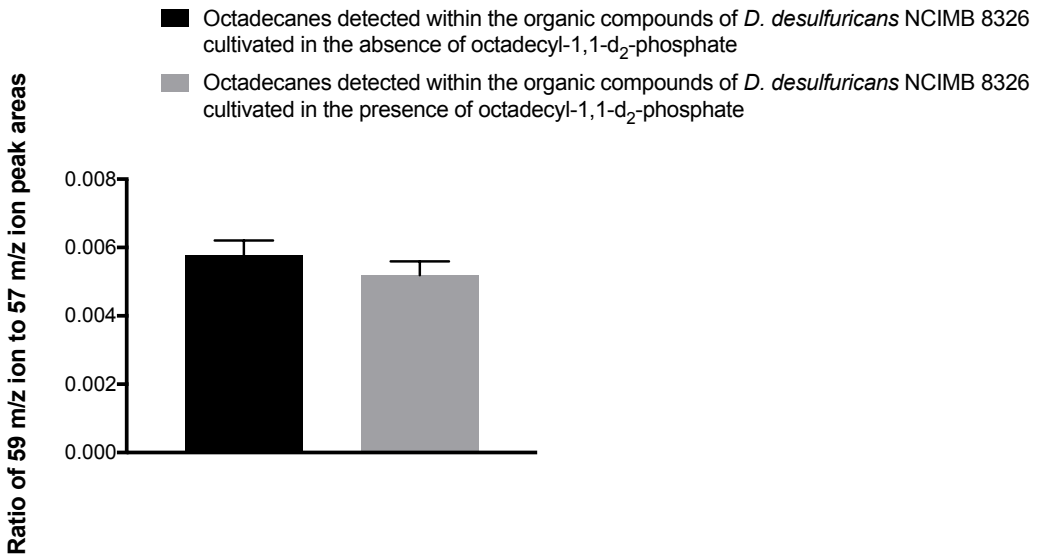


Figure 5.11. 59 m/z to 57 m/z ion peak area ratios of octadecanes detected within the organic compounds of *D. desulfuricans* NCIMB 8326 cultivated in the absence or presence of octadecyl-1,1-d_2-phosphate

D. desulfuricans NCIMB 8326 was cultivated in sodium lactate medium in the absence or presence of 1 % (v/v) 1 mM octadecyl-1,1-d_2-phosphate solution. After 7-day incubation, organic compounds from *D. desulfuricans* NCIMB 8326 cells were screened for deuterated alkanes by GC-MS. The 59 m/z ion to 57 m/z ion peak area ratio of octadecanes detected within the organic compounds of *D. desulfuricans* NCIMB 8326 cultivated in the presence of octadecyl-1,1-d_2-phosphate was calculated and compared to the same isotopic ratio of octadecanes detected from *D. desulfuricans* NCIMB 8326 cultivated in the absence of octadecyl-1,1-d_2-phosphate. Error bars represent standard deviations of the mean ratio from triplicates.

Organic compounds from *D. desulfuricans* NCIMB 8326 cultivated in sodium lactate medium in the presence of either 10 μM hexadecyl-d_{33}-phosphate or 100 μM tetradecyl-d_{29}-phosphate were screened for nascent hexadecane-d_{33} and tetradecane-d_{29}.

Hexadecane-d_{33} and tetradecane-d_{29} are fully deuterated. The mass spectrum of fully deuterated alkanes is composed by preponderance of deuterated butyl ($C_4D_9^+$), pentyl ($C_5D_{11}^+$) and hexyl ($C_6D_{13}^+$), with a molecular mass of 66 g mol^{-1}, 82 g mol^{-1} and 98 g mol^{-1} respectively. Organic compounds from *D. desulfuricans* NCIMB 8326 grown in sodium lactate medium supplemented with either hexadecyl-d_{33}-phosphate solution or tetradecyl-d_{29}-phosphate solution were thus screened by GC-MS for 66 m/z, 82 m/z and 98 m/z ion fragments, in order of preponderance.

Two compounds (other than hexadecyl-d_{33}-phosphate), eluted at 519.4 s and 728.3 s, within the organic compounds of *D. desulfuricans* NCIMB 8326 grown in sodium lactate medium supplemented with hexadecyl-d_{33}-phosphate solution ionised into 66 m/z, 82 m/z and 98 m/z ion fragments by preponderance (Figure 5.12A). Only the mass spectrum of the compound eluted at 519.4 s revealed a fragmentation pattern of a deuterated alkane, suggesting that this compound was a deuterated alkane.

However, by superimposing the extracted 66 m/z ion chromatogram of the hexadecyl-d_{33}-phosphate solution (Figure 5.12B) and after confirmation of the mass spectral similarity, it appeared that the presumed deuterated alkane eluted at 519.4 s from the organic compounds of *D. desulfuricans* NCIMB 8326 grown in sodium lactate medium supplemented with hexadecyl-d_{33}-phosphate solution was also detected within the hexadecyl-d_{33}-phosphate solution. Therefore, this presumed deuterated alkane might have been due to contamination derived from the hexadecyl-d_{33}-phosphate solution, rather to be a nascent biogenic deuterated alkane produced by *D. desulfuricans* NCIMB 8326.

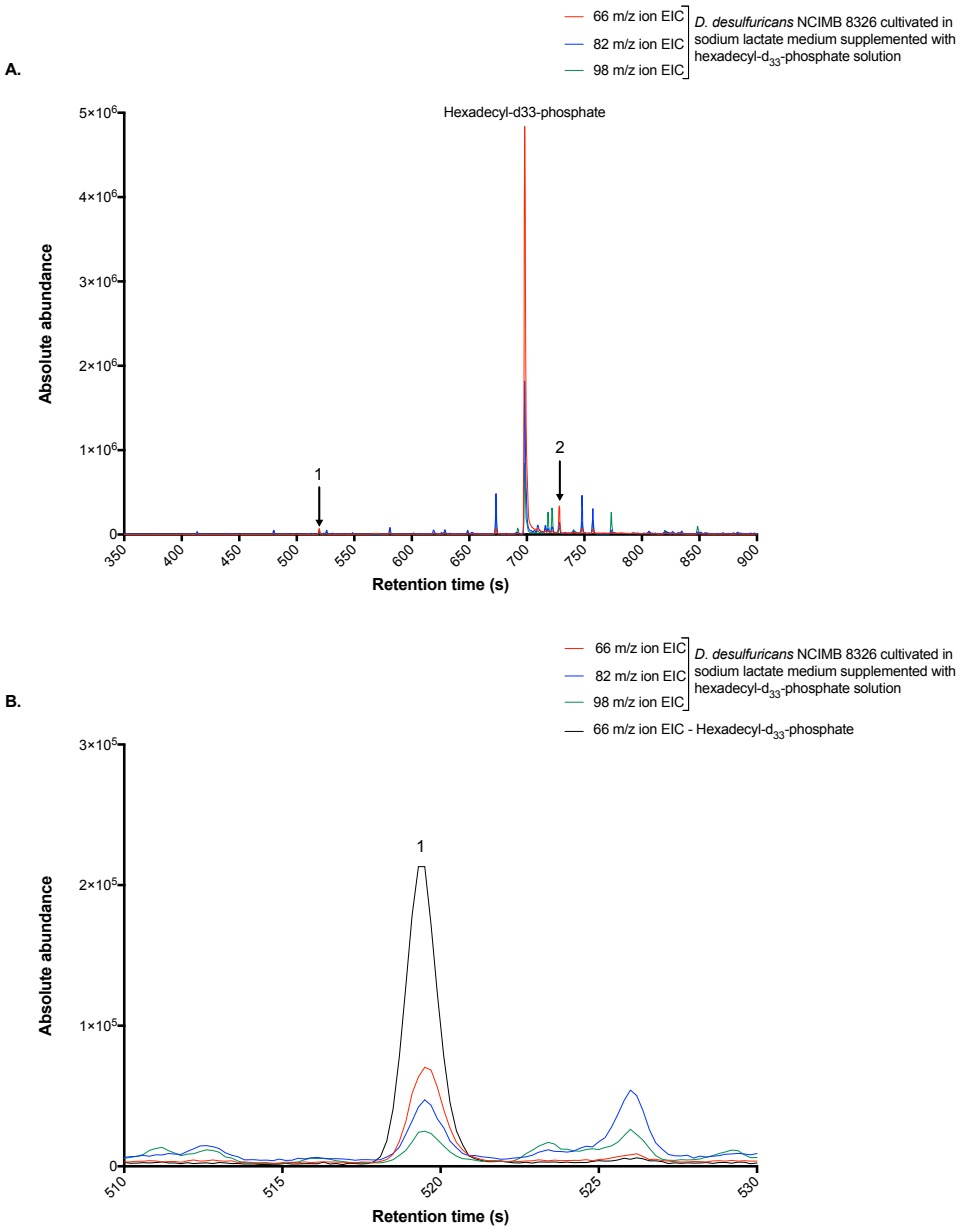


Figure 5.12. Detection of a non-biogenic fully deuterated alkane within the organic compounds of *D. desulfuricans* NCIMB 8326 cultivated in sodium lactate medium supplemented with hexadecyl-d_{33}-phosphate solution

Organic compounds from *D. desulfuricans* NCIMB 8326 cultivated in sodium lactate medium supplemented with 1 % (v/v) 1 mM hexadecyl-d_{33}-phosphate solution were screened for fully deuterated alkanes by GC-MS. The extracted 66 m/z, 82 m/z and 98 m/z ion chromatograms (EIC) of organic compounds from *D. desulfuricans* NCIMB 8326 revealed that two compounds (1 and 2; indicated by an arrow) fragmented into 66 m/z, 82 m/z and 98 m/z ion fragments by order of preponderance (A). The mass spectrum of the compound eluted at 519.4 s disclosed a fully deuterated alkane fragmentation pattern. The hexadecyl-d_{33}-phosphate solution was also analysed by GC-MS. The superimposition of the 66 m/z EIC of the hexadecyl-d_{33}-phosphate solution on the EIC of organic compounds from *D. desulfuricans* NCIMB 8326 revealed that this presumed fully deuterated alkane was also detected within the hexadecyl-d_{33}-phosphate solution (B).

One compound (other than tetradecyl-d_{29}-phosphate), eluted at 416.0 s, within the organic compounds of *D. desulfuricans* NCIMB 8326 grown in sodium lactate medium supplemented with tetradecyl-d_{29}-phosphate solution ionised into 66 m/z, 82 m/z and 98 m/z ion fragments by preponderance (Figure 5.13A). The mass spectrum of this compound showed a fragmentation pattern of a deuterated alkane, suggesting that this compound was a deuterated alkane.

However, by superimposing the extracted 66 m/z ion chromatogram of the tetradecyl-d_{29}-phosphate solution (Figure 5.13A) and after confirmation of the mass spectral similarity, it appeared that the presumed deuterated alkane eluted at 416.0 s from the organic compounds of *D. desulfuricans* NCIMB 8326 cultivated in sodium lactate medium supplemented with tetradecyl-d_{29}-phosphate solution was also detected within the tetradecyl-d_{29}-phosphate solution. Therefore, this presumed deuterated alkane might have been due to contamination derived from the tetradecyl-d_{29}-phosphate solution, rather to be a nascent biogenic deuterated alkane produced by *D. desulfuricans* NCIMB 8326.

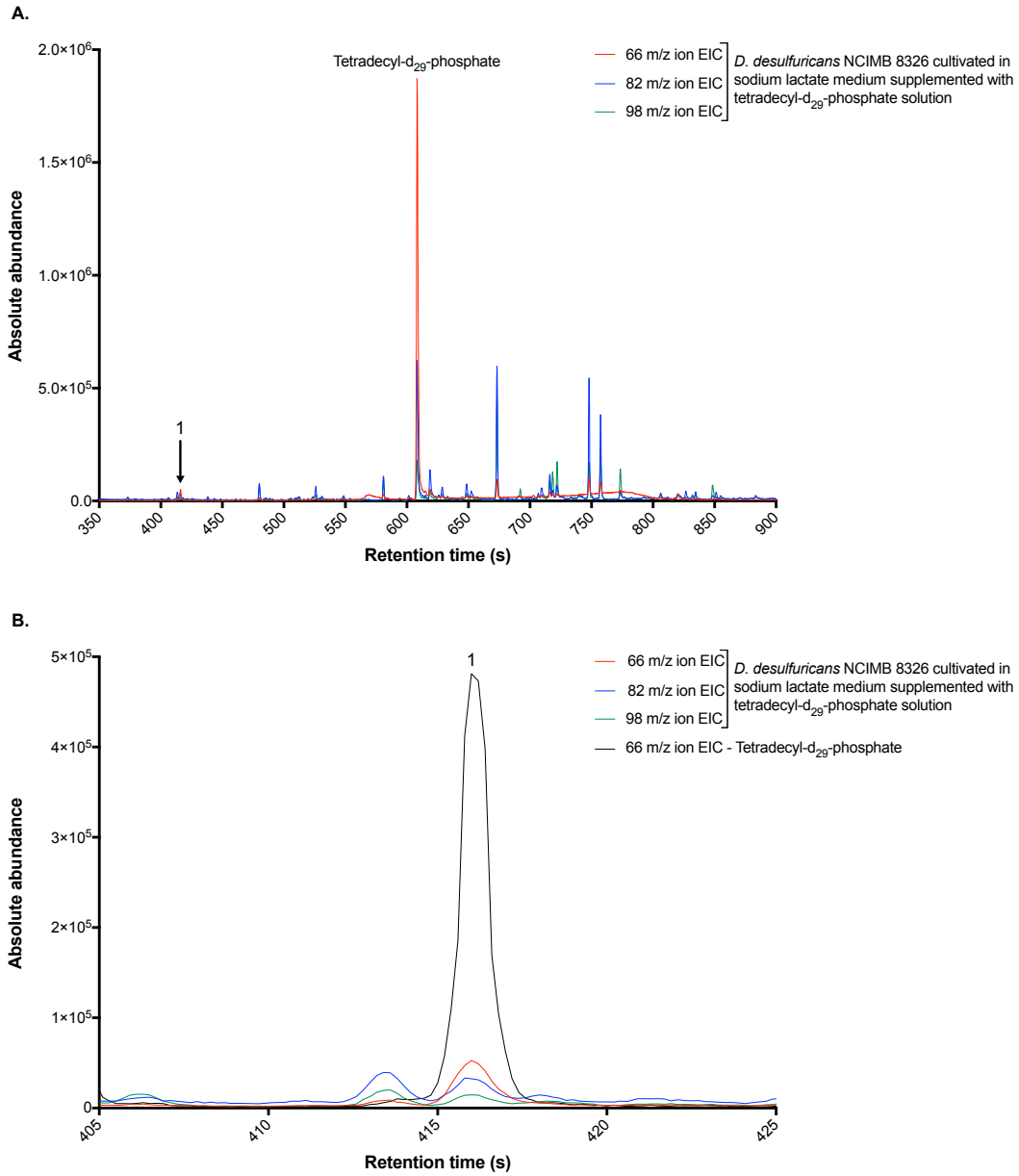


Figure 5.13. Detection of a non-biogenic fully deuterated alkane within the organic compounds of *D. desulfuricans* NCIMB 8326 cultivated in sodium lactate medium supplemented with tetradecyl-d_{29}-phosphate solution

Organic compounds from *D. desulfuricans* NCIMB 8326 cultivated in sodium lactate medium supplemented with 1 % (v/v) 10 mM tetradecyl-d_{29}-phosphate solution were screened for fully deuterated alkanes by GC-MS. The extracted 66 m/z, 82 m/z and 98 m/z ion chromatograms (EIC) of organic compounds from *D. desulfuricans* NCIMB 8326 revealed that one compound (1; indicated by an arrow) fragmented into 66 m/z, 82 m/z and 98 m/z ion fragments by order of preponderance (A). The mass spectrum of this compound eluted at 416.0 s disclosed a fully deuterated alkane fragmentation pattern. The tetradecyl-d_{29}-phosphate solution was also analysed by GC-MS. The superimposition of the 66 m/z EIC of the tetracecyl-d_{29}-phosphate solution on the EIC of organic compounds from *D. desulfuricans* NCIMB 8326 revealed that this presumed fully deuterated alkane was also detected within the tetracecyl-d_{29}-phosphate solution (B).

5.4. Empirical Investigation of a Putative Reductase and a Putative Methyltransferase Functions in Desulfovibrio Alkane Biosynthesis

According to the hypothetical *Desulfovibrio* alkane biosynthetic pathway deduced from the *in silico* genomic comparison (*cf.* Chapter 4 – Comparative Genomic Analysis; Figure 4.16), alkyl phosphates are reduced to alkanes by a putative reductase and by a putative methyltransferase to *iso*-alkanes (Figure 5.14).

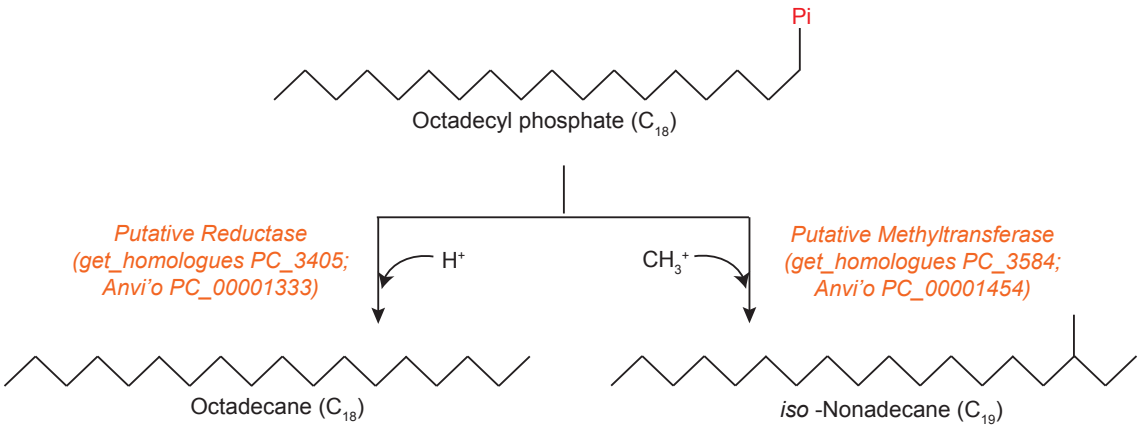


Figure 5.14. Hypothetical reduction reactions catalysed a putative reductase and a putative methyltransferase in *Desulfovibrio* alkane biosynthesis

To verify the hypothetical function of the putative reductase and methyltransferase in *Desulfovibrio* alkane biosynthesis, heterologous expression of the putative reductase and methyltransferase genes originating from the alkane producing *D. desulfuricans* NCIMB 8326 were investigated in *D. vulgaris* Hildenborough and *E. coli*. *D. vulgaris* Hildenborough and *E. coli* expressing either the putative reductase or methyltransferase gene were cultivated in medium supplemented with 100 *μ*M tetradecyl-d_{29}-phosphate, prior to being screened for fully deuterated and unlabelled alkane production.

5.4.1. Heterologous Expression of the Putative Reductase and the Putative Methyltransferase in *D. vulgaris* Hildenborough

The *Desulfovibrio* expression vectors pPD3-Reductase and pPD3-Methyltransferase, carrying the putative histidine-tagged reductase and methyltransferase gene respectively, were successfully constructed and validated by sequencing. *D. vulgaris* Hildenborough was subsequently transformed by conjugation with pPD3, pPD3-Reductase and pPD3-Methyltransferase separately.

Wild type *D. vulgaris* Hildenborough, *D. vulgaris* Hildenborough transformed with pPD3, *D. vulgaris* Hildenborough transformed with pPD3-Reductase and *D. vulgaris* Hildenborough transformed with pPD3-Methyltransferase were cultured in triplicate in sodium lactate medium supplemented with 1 % (v/v) 10 mM tetradecyl-d_{29}-phosphate solution under anaerobic conditions. Expression of the putative histidine-tagged reductase and methyltransferase genes was verified by SDS-PAGE and Western blotting (Figure 5.15). The online tool ProtParam predicted a molecular weight of 11.3 kDa for the putative histidine-tagged reductase and a molecular weight of 25.9 kDa for the putative histidine-tagged methyltransferase. No fluorescent band was detected on the Western Blot of the whole cell protein extracts from *D. vulgaris* Hildenborough transformed with either pPD3-Reductase or pPD3-Methyltransferase. Absence of fluorescent signal was not due to a defective Western blot procedure as a fluorescent signal was detected for the positive control. This observation suggested that neither the putative reductase nor the putative methyltransferase were produced by *D. vulgaris* Hildenborough.

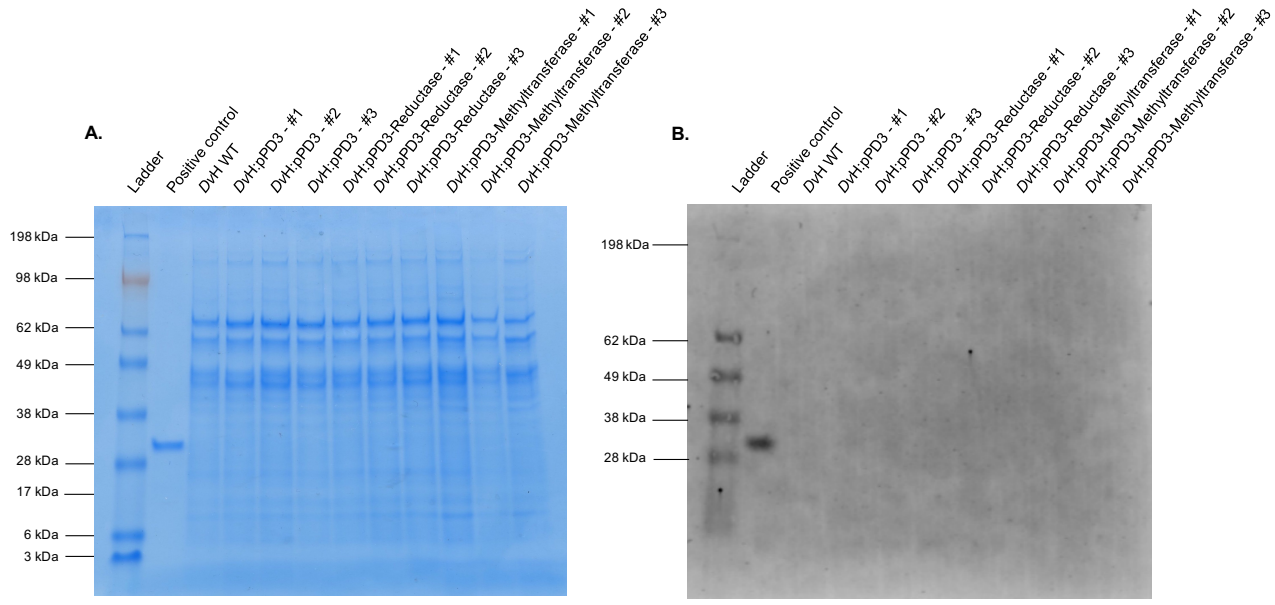


Figure 5.15. Expression profiles of the putative reductase and methyltransferase in *D. vulgaris* Hildenborough

Wild type *D. vulgaris* Hildenborough (*Dv*H WT), *D. vulgaris* Hildenborough transformed with pPD3 (*Dv*H:pPD3), *D. vulgaris* Hildenborough transformed with pPD3-Reductase (*Dv*H:pPD3-Reductase) and *D. vulgaris* Hildenborough transformed with pPD3-Methyltransferase (*Dv*H:pPD3-Methyltransferase) were cultured in triplicate (#1, #2, #3) in sodium lactate medium supplemented with 1 % (v/v) 10 mM tetradecyl-d_{29}-phosphate solution. After 7-day incubation at 37 °C under an anaerobic atmosphere of 80 % N_2, 10 % CO_2 and 10 % H_2, cells were lysed and proteins from whole cell extracts were separated by SDS-PAGE on a 4-20 % gradient Bis-Tris polyacrylamide gel, for 45 min, at 140 V in 1-fold MOPS running buffer, followed by Coomassie Blue staining (A). The positive control used is a purified histidine-tagged esterase (31.1 kDa). SeeBlue™ Plus2 Pre-stained Protein Standard (ThermoFisher Scientific) was used as protein ladder. Proteins were transferred from an unstained polyacrylamide gel to a PVDF membrane for Western blot analysis (B). The blot was probed with a mouse anti-His Tag primary antibody and with a goat anti-mouse secondary antibody conjugated to the IRDye 680RD fluorophore. The IRDye 680RD fluorophore was excited at 676 nm and fluorescence detected at 700 nm.

5.4.2. Heterologous Expression of the Putative Reductase in *E. coli*

The *E. coli* expression vector pEX1K3-Reductase, carrying the putative histidine-tagged reductase gene, was successfully constructed and validated by sequencing. BL21 Star (DE3) *E. coli* was subsequently transformed with pEX1K3 and pEX1K3-Reductase separately. Expression of the gene encoding the histidine-tagged putative reductase in *E. coli* was induced by addition of IPTG in triplicate cultures, incubated under either aerobic or anaerobic conditions.

A fluorescent band below 17 kDa was detected on the Western blot of two of the triplicate whole cell protein extracts from *E. coli* transformed with pEX1K3-Reductase, cultured under aerobic conditions (Figure 5.16). Therefore, the putative reductase gene was successfully produced in two of the triplicate cultures, incubated under aerobic conditions.

No fluorescent signal was detected on the Western blot of the whole cell protein extracts from *E. coli* transformed with pEX1K3-Reductase, cultured under anaerobic conditions (Figure 5.17). Absence of fluorescent signal was not due to a defective Western blot experiment as a fluorescent signal was detected for the positive control, suggesting that the putative reductase was not produced by *E. coli* transformed with pEX1K3-Reductase, cultured under anaerobic conditions.

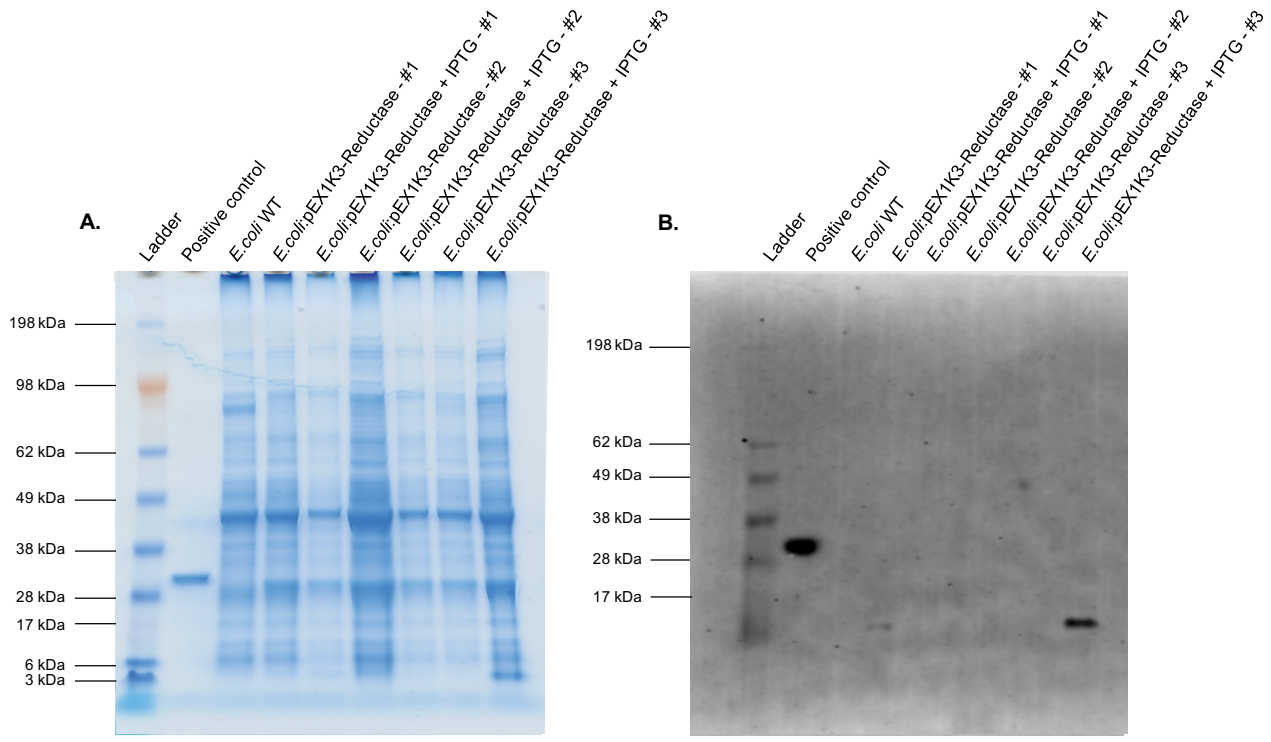


Figure 5.16. Expression profiles of the putative reductase in *E. coli*, cultured under aerobic conditions

Wild type *E. coli* (*E.coli* WT) and *E. coli* transformed with pEX1K3-Reductase (*E.coli*:pEX1K3-Reductase) were aerobically cultured in triplicate (#1, #2, #3) in LB medium supplemented with 1 % (v/v) 10 mM tetradecyl-d_{29}-phosphate solution and incubated at 37°C, shaking at 220 rpm to an OD of 0.7, at 600 nm. Plasmid gene expression was then induced by addition of 200 μM IPTG where appropriate. Cultures were incubated for another 4 h, prior to cell lysis. Proteins from whole cell extracts were separated by SDS-PAGE on a 4-20 % gradient Bis-Tris polyacrylamide gel, for 45 min, at 140 V in 1-fold MOPS running buffer, followed by Coomassie Blue staining (A). The positive control used is a purified His-tag esterase (31.1 kDa). SeeBlue™ Plus2 Pre-stained Protein Standard (ThermoFisher Scientific) was used as protein ladder. Proteins were transferred from an unstained polyacrylamide gel to a PVDF membrane for Western blot analysis (B). The blot was probed with a mouse anti-His Tag primary antibody and with a goat anti-mouse secondary antibody conjugated to the IRDye 680RD fluorophore. The IRDye 680RD fluorophore was excited at 676 nm and fluorescence detected at 700 nm.

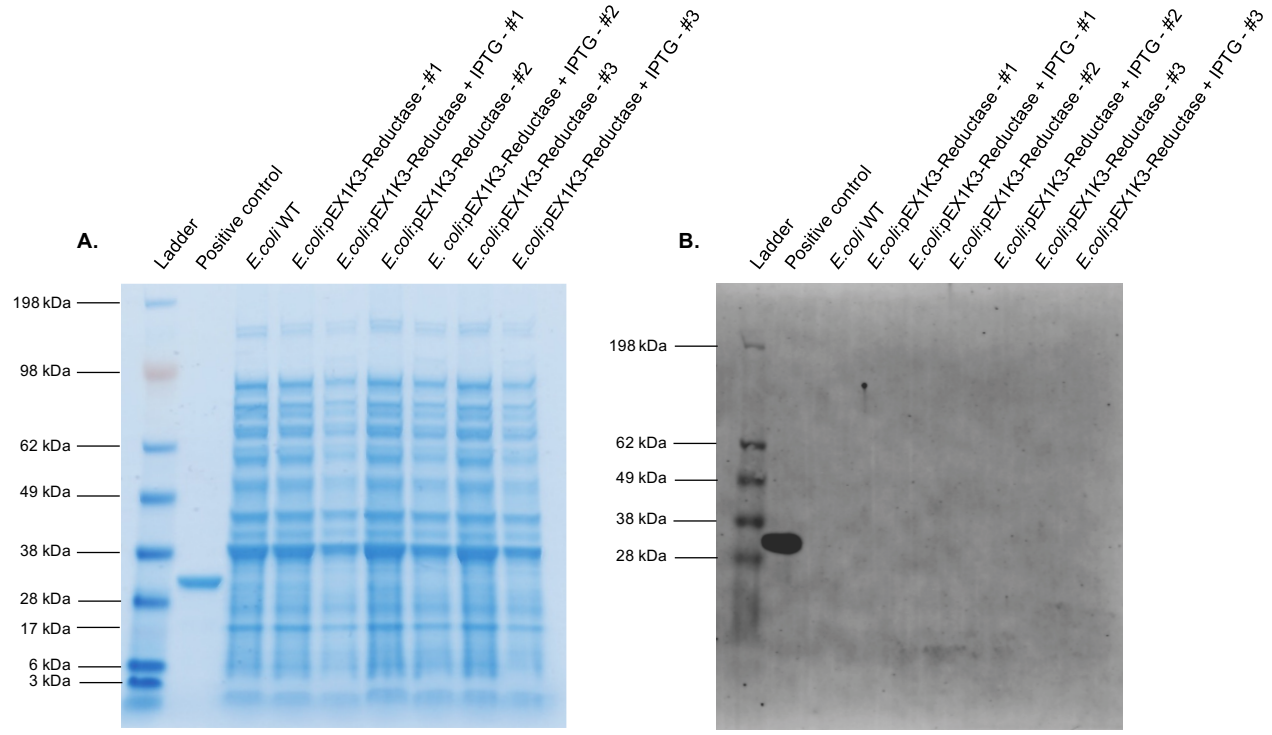


Figure 5.17. Expression profiles of the putative reductase in *E. coli*, cultured under anaerobic conditions

Wild type *E. coli* (*E.coli* WT) and *E. coli* transformed with pEX1K3-Reductase (*E.coli:* pEX1K3-Reductase) were anaerobically cultured in triplicate (#1, #2, #3) in LB medium supplemented with 1 % (v/v) 10 mM tetradecyl-d_{29}-phosphate solution and incubated at 37°C, shaking at 220 rpm to an OD of 0.2, at 600 nm. Plasmid gene expression was then induced by addition of 200 *µ*M IPTG where appropriate. Cultures were incubated for another 4 h, prior to cell lysis. Proteins from whole cell extracts were separated by SDS-PAGE on a 4-20 % gradient Bis-Tris polyacrylamide gel, for 45 min, at 140 V in 1-fold MOPS running buffer, followed by Coomassie Blue staining (A). The positive control used is a purified His-tag esterase (31.1 kDa). SeeBlue™ Plus2 Pre-stained Protein Standard (ThermoFisher Scientific) was used as protein ladder. Proteins were transferred from an unstained polyacrylamide gel to a PVDF membrane for Western blot analysis (B). The blot was probed with a mouse anti-His Tag primary antibody and with a goat anti-mouse secondary antibody conjugated to the IRDye 680RD fluorophore. The IRDye 680RD fluorophore was excited at 676 nm and fluorescence detected at 700 nm.

Organic compounds from *E. coli* which produced the putative reductase, cultured in presence of 100 μM tetradecyl-d_{29}-phosphate under aerobic conditions were screened for fully deuterated and unlabelled alkanes by GC-MS.

No compounds (other than the tetradecyl-d_{29}-phosphate) ionised into 66 m/z, 82 m/z and 98 m/z ion fragments by order of preponderance were detected by GC-MS within the organic compounds of *E. coli* (Figure 5.18A). Therefore, no fully deuterated alkanes were detected by GC-MS within the organic compounds of *E. coli*.

The mass spectra of four compounds (shown by an arrow; Figure 5.18B) revealed an unlabelled alkane fragmentation pattern and were detected within the organic compounds of *E. coli*. These four presumed unlabelled alkanes were also detected within the tetradecyl-d_{29}-phosphate solution. Therefore, these four presumed unlabelled alkanes may have been due to contamination derived from the tetradecyl-d_{29}-phosphate solution, rather to be biogenic unlabelled alkanes produced by *E. coli*.

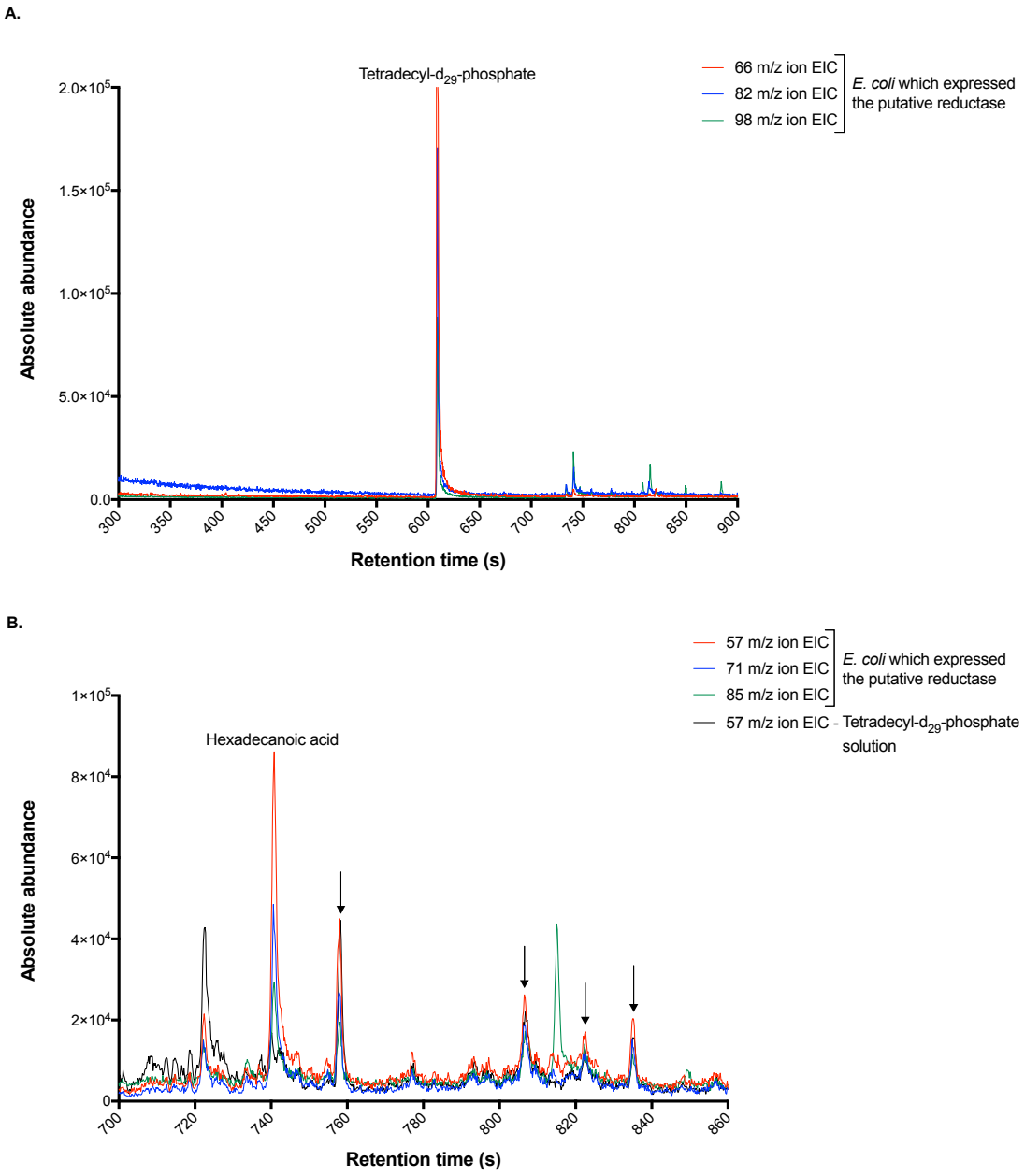


Figure 5.18. Undetectability of fully deuterated alkanes and detection of non-biogenic alkanes within the organic compounds of *E. coli* which expressed the putative reductase

Organic compounds from *E. coli* which expressed the putative reductase, cultured in presence of 100 μM tetradecyl-d_{29}-phosphate under aerobic conditions, and the tetradecyl-d_{29}-phosphate solution were screened by GC-MS for fully deuterated alkanes and unlabelled alkanes. The extracted 66 m/z, 82 m/z and 98 m/z ion chromatograms (EIC) of organic compounds from *E. coli* revealed that no fully deuterated alkanes were detected (A). The 57 m/z, 71 m/z and 85 m/z EIC of organic compounds from *E. coli* superimposed to the 57 m/z EIC of the tetracecyl-d_{29}-phosphate solution revealed that the detected compounds with a mass spectrum disclosing an unlabelled alkane fragmentation pattern (indicated by an arrow) were also detected within the tetradecyl-d_{29}-phosphate solution (B).

5.4.3. Heterologous Expression of the Putative Methyltransferase in *E. coli*

The *E. coli* expression vector pEX1K3-Methyltransferase, carrying the putative histidine-tagged methyltransferase gene, was successfully constructed and validated by sequencing. BL21 Star (DE3) *E. coli* was subsequently transformed with pEX1K3 and pEX1K3-Methyltransferase separately. Expression of the gene encoding the putative histidine-tagged methyltransferase in *E. coli* was induced by addition of IPTG in triplicate cultures, incubated under either aerobic or anaerobic conditions. In one of the anaerobic triplicate cultures, *E. coli* growth, monitored by optical density measurement, followed an aerobic growth curve; indicating that the culture was not under anaerobic conditions and therefore was removed from the experiment.

A fluorescent band below 28 kDa was detected on the Western blot of the whole cell protein extracts from *E. coli* transformed with pEXK3-Methyltransferase cultures, which were supplemented with IPTG and incubated under aerobic conditions (Figure 5.19) or anaerobic conditions (Figure 5.20). The putative methyltransferase gene was therefore successfully produced by *E. coli*, cultured under aerobic and anaerobic conditions.

A weak fluorescent band below 28 kDa was also detected on the Western blot of the whole cell protein extracts from *E. coli* transformed with pEXK3-Methyltransferase cultures which were not supplemented with IPTG and incubated under aerobic conditions (Figure 5.19) or anaerobic conditions (Figure 5.20). The production of the putative methyltransferase gene in *E. coli* cultures which were not supplemented with IPTG suggested that the T7 promoter was leaky.

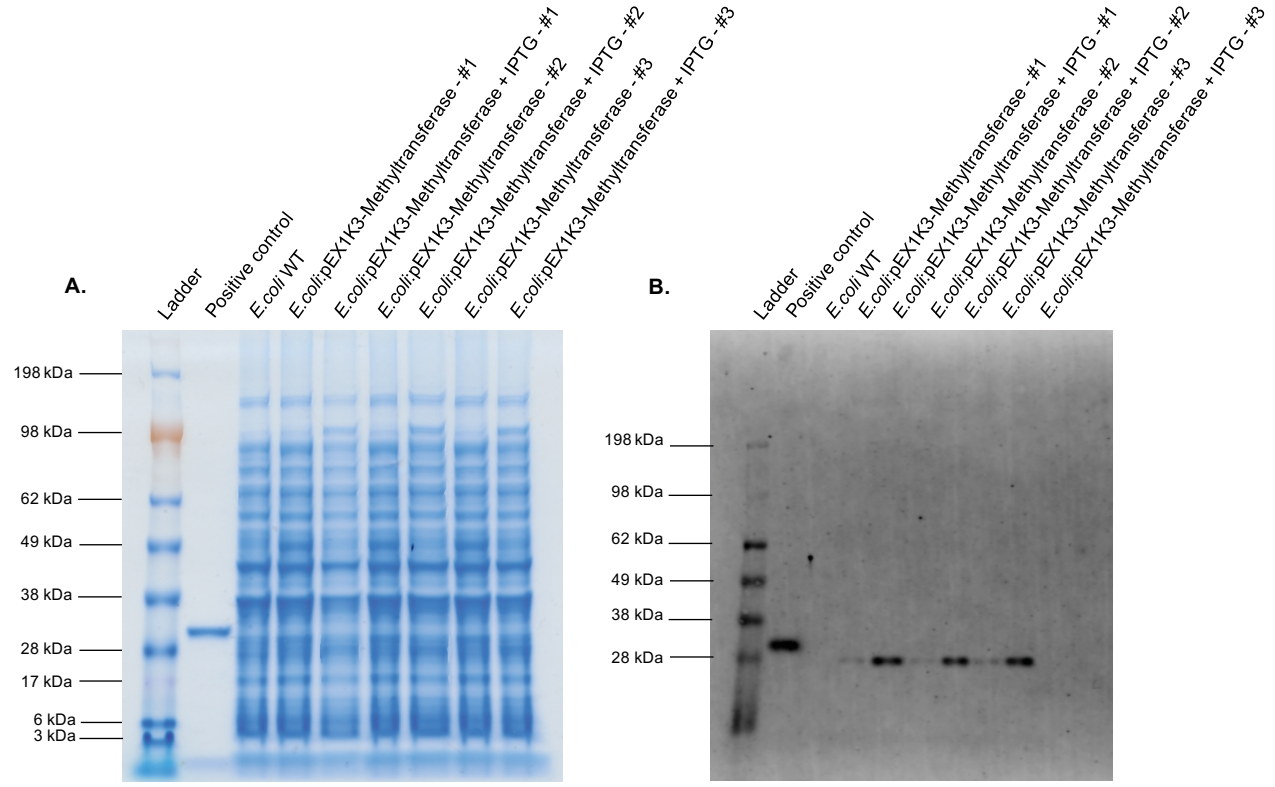


Figure 5.19. Expression profiles of the putative methyltransferase in *E. coli*, cultured under aerobic conditions

Wild type *E. coli* (*E.coli* WT) and *E. coli* transformed with pEX1K3-Methyltransferase (*E.coli:* pEX1K3-Methyltransferase) were aerobically cultured in triplicate (#1, #2, #3) in LB medium supplemented with 1 % (v/v) 10 mM tetradecyl-d_{29}-phosphate solution and incubated at 37°C, shaking at 220 rpm to an OD of 0.7, at 600 nm. Plasmid gene expression was then induced by addition of 200 μM IPTG where appropriate. Cultures were incubated for another 4 h, prior to cell lysis. Proteins from whole cell extracts were separated by SDS-PAGE on a 4-20 % gradient Bis-Tris polyacrylamide gel, for 45 min at 140 V in 1-fold MOPS running buffer, followed by Coomassie Blue staining (A). The positive control used is a purified His-tag esterase (31.1 kDa). SeeBlue™ Plus2 Pre-stained Protein Standard (ThermoFisher Scientific) was used as protein ladder. Proteins were transferred from an unstained polyacrylamide gel to a PVDF membrane for Western blot analysis (B). The blot was probed with a mouse anti-His Tag primary antibody and with a goat anti-mouse secondary antibody conjugated to the IRDye 680RD fluorophore. The IRDye 680RD fluorophore was excited at 676 nm and fluorescence detected at 700 nm.

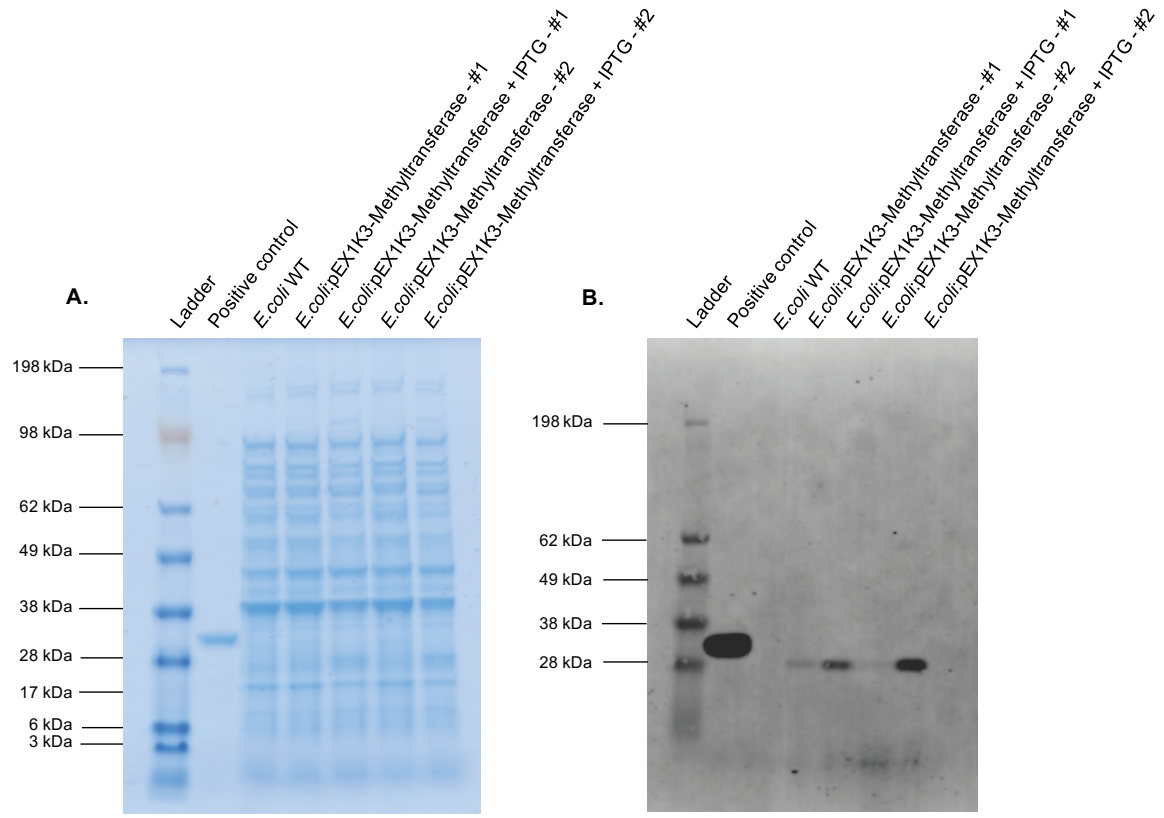


Figure 5.20. Expression profiles of the putative methyltransferase in *E. coli*, cultured under anaerobic conditions

Wild type *E. coli* (*E.coli* WT) and *E. coli* transformed with pEX1K3-Methyltransferase (*E.coli:* pEX1K3-Methyltransferase) were anaerobically cultured in duplicate (#1, #2) in LB medium supplemented with 1 % (v/v) 10 mM tetradecyl-d_{29}-phosphate solution and incubated at 37°C, shaking at 220 rpm to an OD of 0.2, at 600 nm. Plasmid gene expression was then induced by addition of 200 *μ*M IPTG where appropriate. Cultures were incubated for another 4 h, prior to cell lysis. Proteins from whole cell extracts were separated by SDS-PAGE on a 4-20 % gradient Bis-Tris polyacrylamide gel, for 45 min at 140 V in 1-fold MOPS running buffer, followed by Coomassie Blue staining (A). The positive control used is a purified His-tag esterase (31.1 kDa). SeeBlue™ Plus2 Pre-stained Protein Standard (ThermoFisher Scientific) was used as protein ladder. Proteins were transferred from an unstained polyacrylamide gel to a PVDF membrane for Western blot analysis (B). The blot was probed with a mouse anti-His Tag primary antibody and with a goat anti-mouse secondary antibody conjugated to the IRDye 680RD fluorophore. The IRDye 680RD fluorophore was excited at 676 nm and fluorescence detected at 700 nm.

Organic compounds from *E. coli* which produced the putative methyltransferase, cultured in presence of 100 *μ*M tetradecyl-d_{29}-phosphate under either aerobic or anaerobic conditions were screened for fully deuterated and unlabelled alkanes by GC-MS. The same GC-MS profiles were observed for the organic compounds from *E. coli* cultured under either aerobic or anaerobic conditions.

No compounds (other than the tetradecyl-d_{29}-phosphate) ionised into 66 m/z, 82 m/z and 98 m/z ion fragments by order of preponderance were detected by GC-MS within the organic compounds of *E. coli* (Figure 5.21A). Therefore, no fully deuterated alkanes were detected by GC-MS within the organic compounds of *E. coli*.

The mass spectra of four compounds (shown by an arrow; Figure 5.21B) revealed an unlabelled alkane fragmentation pattern and were detected within the organic compounds of *E. coli*. These four presumed unlabelled alkanes were also detected within the tetradecyl-d_{29}-phosphate solution. Therefore, these four presumed unlabelled alkanes may have been due to contamination derived from the tetradecyl-d_{29}-phosphate solution, rather to be biogenic unlabelled alkanes produced by *E. coli*. This conclusion was endorsed by the detection of the same presumed unlabelled alkanes within the organic compounds of *E. coli* which expressed the putative reductase.

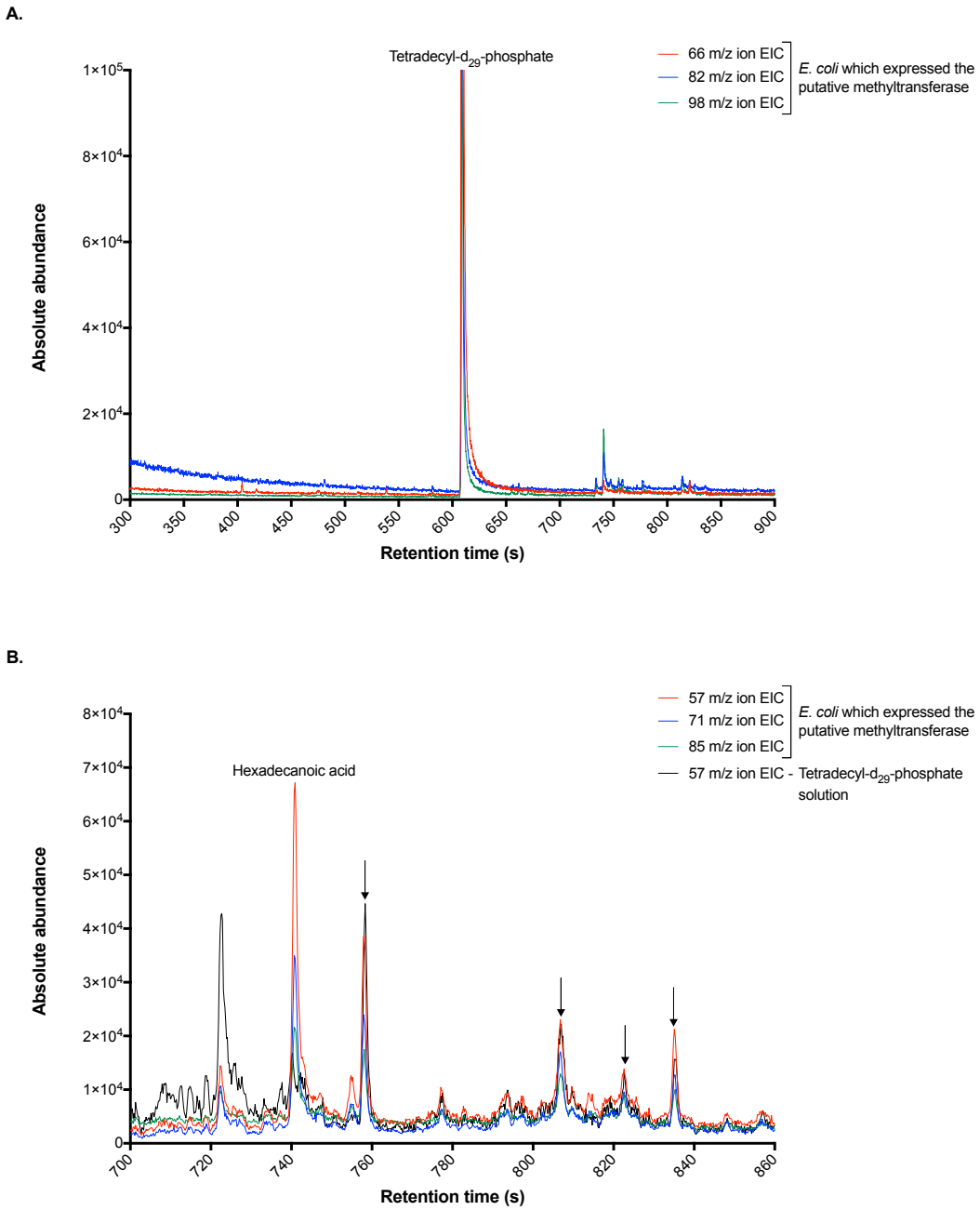


Figure 5.21. Undetectability of fully deuterated alkanes and detection of non-biogenic alkanes within the organic compounds of *E. coli* which expressed the putative methyltransferase

Organic compounds from *E. coli* which expressed the putative methyltransferase, cultured in presence of 100 μM tetradecyl-d_{29}-phosphate under either aerobic or anaerobic conditions, and the tetradecyl-d_{29}-phosphate solution were screened by GC-MS for deuterated alkanes and unlabelled alkanes. The extracted 66 m/z, 82 m/z and 98 m/z ion chromatograms (EIC) of organic compounds from *E. coli* revealed that no fully deuterated alkanes were detected (A). The 57 m/z, 71 m/z and 85 m/z EIC of organic compounds from *E. coli* superimposed to the 57 m/z EIC of the tetracecyl-d_{29}-phosphate solution revealed that the detected compounds with a mass spectrum disclosing an unlabelled alkane fragmentation pattern (indicated by an arrow) were also detected within the tetradecyl-d_{29}-phosphate solution (B).

5.4.4. *In vitro* Assay for Investigation of the Putative Reductase and the Putative Methyltransferase Functions

Although tetradecyl-d_{29}-phosphate was detectable on the GC-MS spectrum of cell pellets washed 3 times with PBS, it is uncertain that *E. coli* cells incorporated tetradecyl-d_{29}-phosphate due to their structural characteristics. Therefore, an *in vitro* functional verification of the putative reductase and methyltransferase was attempted.

Expression of the genes encoding the histidine-tagged putative reductase and methyltransferase in *E. coli* was induced by addition of IPTG in triplicate cultures, incubated under aerobic conditions. The putative histidine-tagged reductase and methyltransferase gene expression in *E. coli* was determined by SDS-PAGE and Western blot (Figure 5.22). No fluorescent band below 17 kDa was detected on the Western blot of the whole cell protein extracts from *E. coli* transformed with pEX1K3-Reductase, suggesting that the putative reductase was not produced by *E. coli*. However, a fluorescent band below 28 kDa was detected on the Western blot of the whole cell protein extracts from *E. coli* transformed with pEXK3-Methyltransferase, suggesting that the putative methyltransferase was successfully produced by *E. coli*.

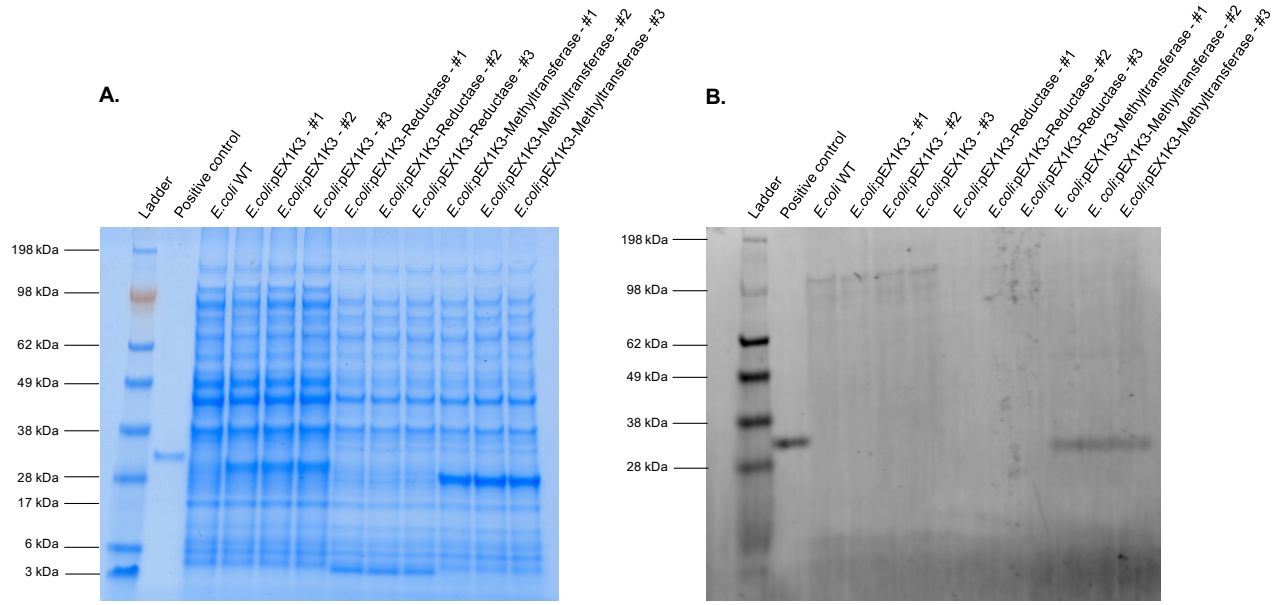


Figure 5.22. Expression profiles of the putative reductase and methyltransferase in *E. coli* for *in vitro* protein function assay

Wild type *E. coli* (*E.coli* WT), *E. coli* transformed with pEX1K3 (*E.coli:*pEX1K3), *E. coli* transformed with pEX1K3-Reductase (*E.coli:*pEX1K3-Reductase) and *E. coli* transformed with pEX1K3-Methyltransferase (*E.coli:*pEX1K3-Methyltransferase) were aerobically cultured in triplicate (#1, #2, #3) in LB medium at 37°C, shaking at 220 rpm to an OD of 0.7, at 600 nm. Plasmid gene expression was then induced by addition of 1 mM IPTG. Cultures were incubated for another 4 h, prior to cell lysis. Proteins from whole cell extracts were separated by SDS-PAGE on a 4-20 % gradient Bis-Tris polyacrylamide gel, for 45 min at 140 V in 1-fold MOPS running buffer, followed by Coomassie Blue staining (A). The positive control used is a purified His-tag esterase (31.1 kDa). SeeBlue™ Plus2 Pre-stained Protein Standard (ThermoFisher Scientific) was used as protein ladder. Proteins were transferred from an unstained polyacrylamide gel to a PVDF membrane for Western blot analysis (B). The blot was probed with a mouse anti-His Tag primary antibody and with a goat anti-mouse secondary antibody conjugated to the IRDye 680RD fluorophore. The IRDye 680RD fluorophore was excited at 676 nm and fluorescence detected at 700 nm.

After cell lysis, 100 μM tetradecyl-d_{29}-phosphate was added to the whole cell protein extracts which included the putative methyltransferase. The protein extracts were then screened for fully deuterated and unlabelled alkanes by GC-MS at the beginning of the incubation, after 24 h and 48 h incubation at 37 °C.

No compounds (other than the tetradecyl-d_{29}-phosphate) ionised into 66 m/z ion, 82 m/z ion and 98 m/z ion fragments (by order of preponderance) were detected by GC-MS within the protein extracts at the beginning of the incubation (Figure 5.23A), after 24 h (Figure 5.23B) and 48 h (Figure 5.23C) incubation. Therefore, no fully deuterated alkanes were detected by GC-MS within protein extracts including the putative methyltransferase.

Five compounds ionised into 57 m/z ion, 71 m/z ion and 85 m/z ion fragments (by order of preponderance) were detected by GC-MS within the protein extracts after 24 h (Figure 5.24B) and 48 h (Figure 5.24C). However, the mass spectra of the five compounds did not reveal an unlabelled alkane fragmentation pattern, suggesting that no unlabelled alkanes were detected by GC-MS within the protein extracts including the putative methyltransferase. Four of the five compounds were identified by mass spectra similarities to the mass bank database NIST11 to be fatty acids (with a similarity score above 85). One of the compounds remained unknown.

The presence of the putative methyltransferase in the whole cell protein extracts was noted to be potentially linked to the increase in fatty acid quantity. However, the experiment was also performed with wild type *E. coli* and *E. coli* hosting the vector pEX1K3, the destination vector of the putative methyltransferase gene. The increase in fatty acid quantity throughout the incubation was observed in the whole cell protein extracts from wild type *E. coli* (Appendix 6) and *E. coli* hosting the vector pEX1K3 (Appendix 7).

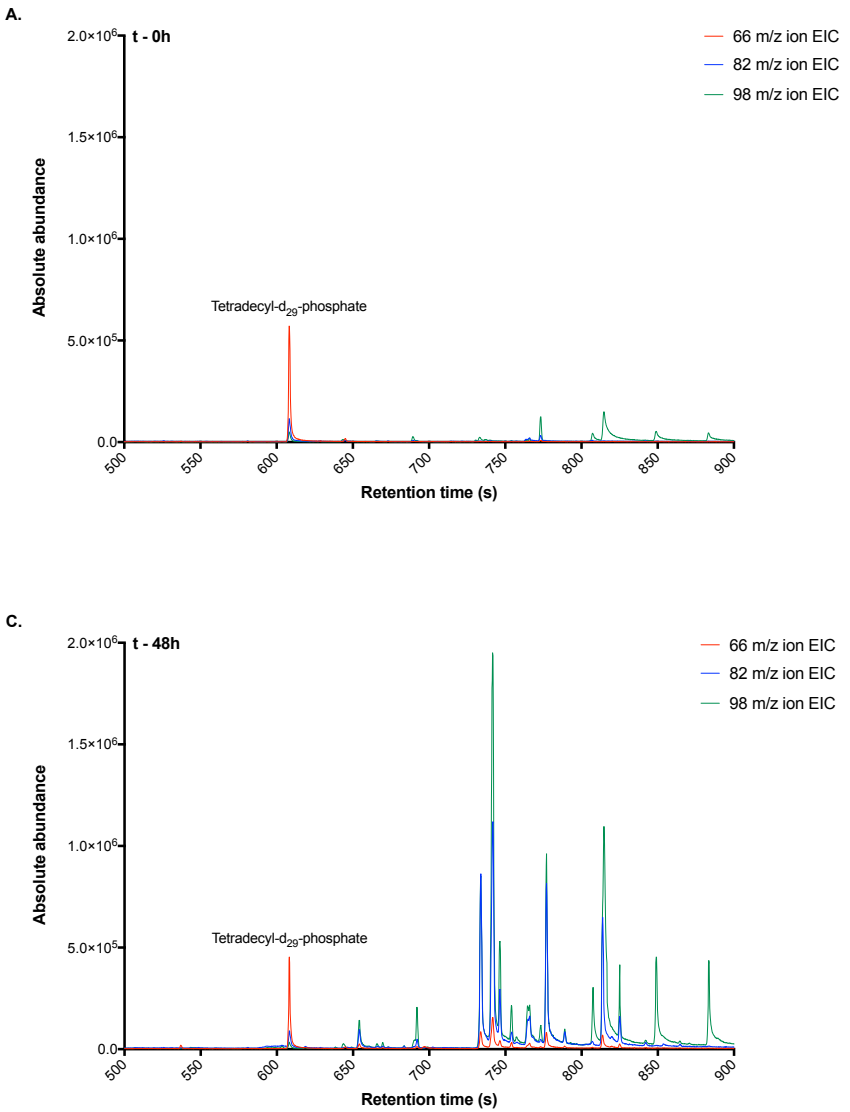


Figure 5.23. Undetectability of fully deuterated alkanes from whole cell protein extracts including the putative methyltransferase and supplemented with 100 μM tetradecyl-d_{29}-phosphate throughout 48 h incubation

Proteins from *E. coli* which expressed the putative methyltransferase were extracted and the whole cell protein extracts were spiked with 100 μM tetradecyl-d_{29}-phosphate prior to being incubated for 48 h at 37 °C. The extracted 66 m/z, 82 m/z and 98 m/z ion chromatograms (EIC) from whole cell protein extracts at the beginning of the incubation (A), after 24 h (B) and 48 h (C) incubation revealed that no fully deuterated alkanes were detected.

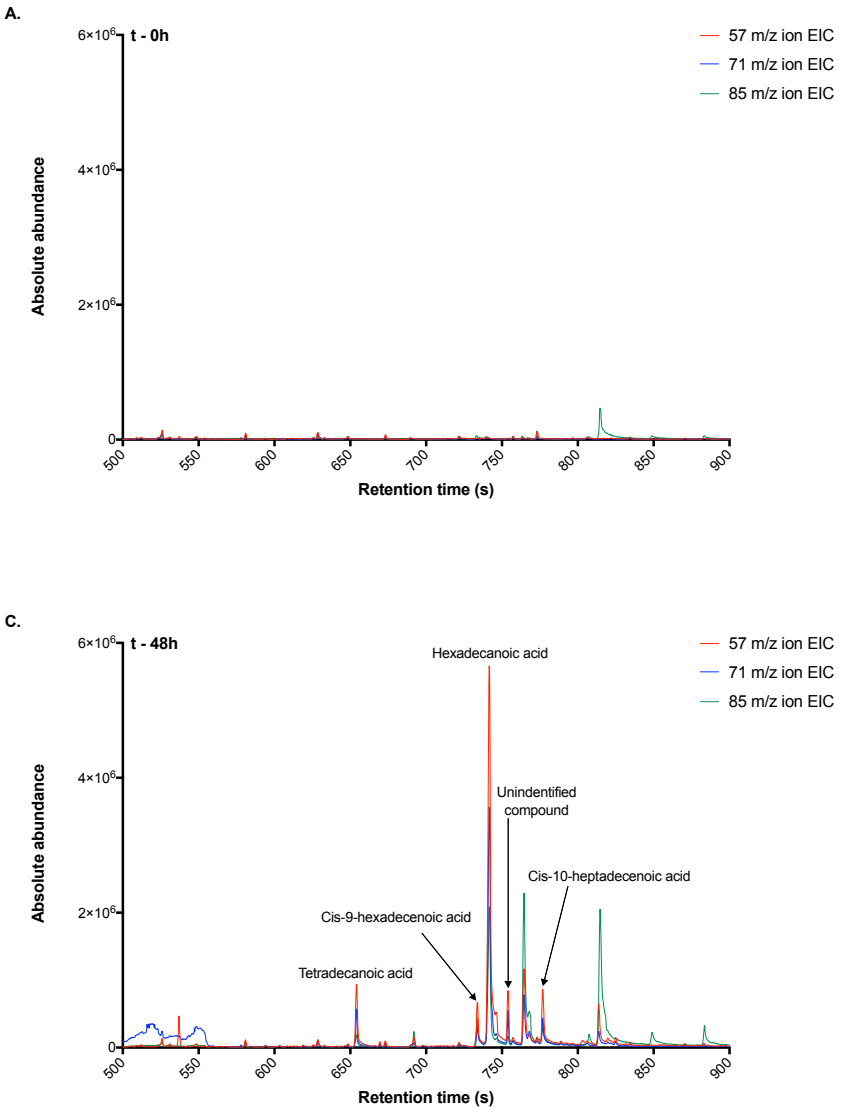


Figure 5.24. Undetectability of unlabelled alkanes from whole cell protein extracts including the putative methyltransferase and supplemented with 100 μM tetradecyl-d_{29}-phosphate throughout 48 h incubation

Proteins from *E. coli* which expressed the putative methyltransferase were extracted and the whole cell protein extracts were spiked with 100 μM tetradecyl-d_{29}-phosphate prior to being incubated for 48 h at 37 °C. The extracted 57 m/z, 71 m/z and 85 m/z ion chromatograms (EIC) from whole cell protein extracts at the beginning of the incubation (A), after 24 h (B) and 48 h (C) incubation revealed that no unlabelled alkanes were detected.

5.5. Empirical Investigation of the Desulfovibrio Alkane Biosynthesis Sensitivity to the Dark

D. desulfuricans NCIMB 8326 was cultivated in sodium lactate medium either in the diurnal daylight or in the dark, at 37 °C under an anaerobic atmosphere of 80 % N_2, 10 % CO_2 and 10 % H_2. After 10 days incubation, *D. desulfuricans* NCIMB 8326 cells were screened for alkanes by GC-MS. Octadecane, nonadecane and eicosane were detected by GC-MS within the organic compounds of *D. desulfuricans* NCIMB 8326 cultivated in the diurnal daylight or in the dark (Figure 5.25A). Moreover, the amounts of alkanes produced by *D. desulfuricans* NCIMB 8326 cultivated either in the diurnal daylight or in the dark were not significantly different (t-test's $p > 0.05$) (Figure 5.25B). Therefore, *Desulfovibrio* alkane biosynthesis was not affected by the dark.

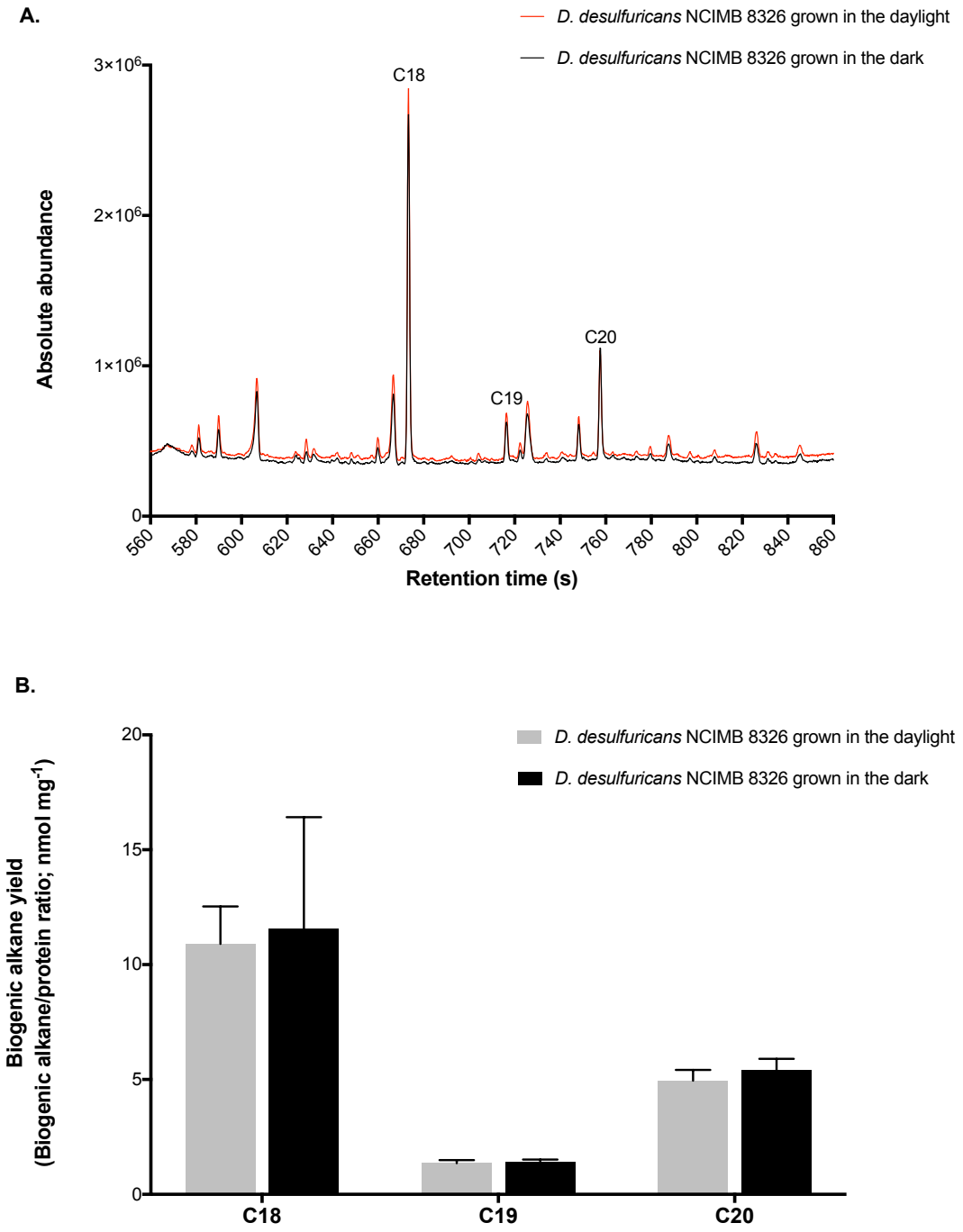


Figure 5.25. Detection of alkane biosynthesis and biogenic alkane yields from *D. desulfovibrio* NCIMB 8326 grown in the light and in the dark

D. desulfuricans NCIMB 8326 was cultivated in sodium lactate medium either in the diurnal daylight or in the dark, at 37 °C under an anaerobic atmosphere of 80 % N_2, 10 % CO_2 and 10 % H_2, prior to being screened for alkanes by GC-MS. Octadecane, nonadecane and eicosane were detected within the organic compounds from both *D. desulfovibrio* NCIMB 8326 grown in the daylight and in the dark (A). Biogenic alkanes were quantified using MassHunter Q-TOF Quantitative Analysis software and the calculated amount of biogenic alkane was normalised to the amount of total proteins extracted from the cultures (B). Error bars represent standard deviations of the biogenic alkane yield from triplicates.

5.6. Discussion

The strategies adopted within this study for the empirical verification of the hypothetical candidate gene function in *Desulfovibrio* alkane biosynthesis were not conclusive and therefore, the *Desulfovibrio* alkane biosynthetic pathway remains to be characterised. The empirical verification was hindered by the challenges in genetically manipulating the alkane producing *Desulfovibrio* spp. These challenges obstructed the route to candidate gene functional verification using a reverse genetic approach. Among the alkane producing *Desulfovibrio* spp., *D. gigas* NCIMB 9332 was previously reported to be transformable by electroporation (Broco *et al.*, 2005; Morais-Silva *et al.*, 2013; da Silva *et al.,* 2015). Within this study, difficulties were encountered for microbiological manipulations of *D. gigas* NCIMB 9332 including the culturing and revival of lyophilised cultures.

Functional gene verification was therefore attempted by heterologous expression of the candidate genes in the non-alkane producing strain *D. vulgaris* Hildenborough and *E. coli*. The V-type ATPase operon expression in *D. vulgaris* Hildenborough was not achieved due to a deficient cloning strategy for the plasmid construction. A more adapted cloning strategy than the conventional cloning method, involving restriction enzymes to generate cohesive fragment extremities, may help to achieve the ligation of four fragments, *e.g.* the Hot Fusion method (Fu *et al.*, 2014) or the OEPR (Overlap Extension PCR and Recombination *in vivo*) method (Liu *et al.*, 2017).

The V-type ATPase operon expression in *E. coli* was not confirmed by Western blotting. Comparing the SDS-PAGE protein profiles of *E. coli* hosting or not the V-type ATPase, no differences in quantity or in intensity of protein bands were observed, suggesting that V-type ATPase operon was not expressed or was undetectable by Western blotting. In order to increase the expression of the V-type ATPase operon, several strategies previously reported to enhance recombinant protein expression could be investigated, including a better host choice or host engineering (Xiao *et al.*, 2014), a better expression vector choice or expression vector optimisation (Rosano & Ceccarelli, 2014) and alternative culture conditions (Chhetri *et al.*, 2015). Moreover, the use of a polycistronic vector designed for protein complex expression, such as pST44 (Tan *et al.*, 2005), would offer the possibility to individually express each subunit of the protein complex, using the monocistronic expression vectors required to build the

polycistronic expression vector. Individual expression of each subunit enables the verification of the subunit coding sequence integrity. An additional factor likely to be impairing the recombinant protein expression is the codon usage bias (Zhou *et al.,* 2016). The codon usage analysis of the V-type ATPase subunit, with the N-terminal fused to a poly-histidine tag, revealed that two codons poorly used by *E. coli* precede the poly-histidine tag in the mRNA sequence. These two poorly used codons could have triggered the stop of the V-type ATPase operon translation before the poly-histidine, which may explain the absence of a fluorescent band on the Western blot. Thus, codon optimisation of the V-type ATPase coding sequence for *E. coli* might promote the recombinant V-type ATPase expression.

Enrichment of the alkane producing *D. desulfuricans* NCIMB 8326 growth medium with isotopically labelled alkyl phosphates, the predicted precursors to biogenic alkanes was also explored. Stable isotope labelling of predicted precursors has been successfully used to elucidate novel pathways (Chokkathukalam *et al.,* 2014). For example, identifying the labelled metabolites in *Cyanothece* sp. ATCC 51142 cultured in presence of 2-^{13}C glycerol enabled the discovery of a threonine-independent route for isoleucine synthesis (Wu *et al.,* 2010). *D. desulfuricans* NCIMB 8326 was previously proven to be capable of octadecane synthesis. Therefore, the octadecyl-1,1-d_2-phosphate was chosen to supplement *D. desulfuricans* NCIMB 8326 growth medium. No nascent deuterated alkanes synthesised by *D. desulfuricans* NCIMB 8326 cultivated in the presence of octadecyl-1,1-d_2-phosphate were detected by GC-MS.

Octadecyl-1,1-d_2-phosphate was found to be only dissolvable in anhydrous ethanol to a final concentration of 1 mM. The dissolution of octadecyl-1,1-d_2-phosphate in anhydrous ethanol limited the growth medium enrichment to 1 % (v/v) octadecyl-1,1-d_2-phosphate solution, to avoid any lethal effect of ethanol on the micro-organisms. According to the *Desulfovibrio* alkane biosynthetic pathway within this study, octadecyl-1,1-d_2-phosphate would be reduced into octadecane-1,1-d_2. The putative nascent octadecane-1,1-d_2 synthesised by *D. desulfuricans* NCIMB 8326 cultured in medium supplemented with 1 % (v/v) octadecyl-1,1-d_2-phosphate solution would be mixed with the biogenic unlabelled octadecanes. The presumed isotopic enrichment of octadecanes would therefore be diluted and lower than 1 atom-% excess. As deuterium is naturally present, it was probable that this presumed isotopic enrichment of octadecanes was below the natural octadecane deuterium enrichment.

Isotopic enrichments of below 1 atom-% excess were previously proven successful. The majority of the studies involving isotopic enrichments of below 1 atom-% excess used the gas chromatography-combustion-isotope ratio mass spectrometry (GC-C-IRMS) technology. This technology enables the detection of isotopic enrichment of target metabolites as low as 0.0001 % (Guo *et al.*, 1997; Heinzle *et al.*, 2008). The detectability of a conventional GC-MS platform, used within this study, is less accurate for very low isotopic enrichments (Zabielski *et al.,* 2013). Strategies were developed to overcome the GC-MS relative detectability for very low isotopic enrichments, including the use of fully isotopically labelled precursors (Patterson *et al.*, 1997), gas chromatography optimisation for better compound separation and the development of mathematical approaches (Mass isotopomer distribution analysis) coupled to software for in-depth data processing (Hellerstein & Neese, 1992; Krämer *et al.*, 2018). Subsequently, the fully deuterated teratrdecyl-d_{29}-phosphate and hexadecyl-d_{33}-phosphate were chosen to supplement *D. desulfuricans* NCIMB 8326 growth medium. Although *D. desulfuricans* NCIMB 8326 was not proven to be capable of tetradecane or hexadecane synthesis, *Desulfovibrio* biogenic alkanes were showed to be of different carbon chain length. Thus, the assumption was made that the substrate specificity of the enzyme catalysing alkane synthesis in *Desulfovibrio* is broad.

An additional advantage of *Desulfovibrio* growth medium enrichment with either teratrdecyl-d_{29}-phosphate or hexadecyl-d_{33}-phosphate was that the presumed isotopic enrichment of alkanes would not be diluted by the presence of biogenic unlabelled tetradecanes or hexadecanes. Furthermore, fully deuterated compounds have a different retention time on the total ion chromatogram compared to their correspondent unlabelled compounds (Lehmann, 2016). As compounds interact through hydrogen bonding with the chromatographic column, substitution of hydrogens by deuterium atoms in a molecule reduces the retention time of the molecule in the chromatographic column. Therefore, the presumed isotopic enrichment of alkanes would not be diluted by the presence of contaminant unlabelled tetradecanes or hexadecanes.

No nascent deuterated alkanes synthesised by *D. desulfuricans* NCIMB 8326 cultivated in sodium lactate enriched to 1 % (v/v) of either teratrdecyl-d_{29}-phosphate or hexadecyl-d_{33}-phosphate solution were detected by GC-MS in this study. This suggested that the alkyl phosphates were not the precursors of alkanes in the *Desulfovibrio* alkane biosynthetic pathway. However, it was probable that the deuterated alkyl phosphates did not penetrate *D. desulfuricans* NCIMB 8326 cells due to structural characteristics.

Heterologous expression of the candidate putative reductase and methyltransferase genes for *Desulfovibrio* alkane biosynthesis were attempted in *D. vulgaris* Hildenborough and *E. coli*. The expression of these two candidate genes in *D. vulgaris* Hildenborough were not detected by Western blotting. Comparing the SDS-PAGE protein profiles of wild type *D. vulgaris* and *D. vulgaris* hosting either the putative reductase or the putative methyltransferase gene, no differences in quantity or in intensity of protein bands were observed. This suggested that the putative reductase and methyltransferase were not expressed or were undetectable by Western blotting. The coding sequence of the putative reductase and methyltransferase genes were cloned downstream of the kanamycin promoter. The kanamycin promoter was proven to be functional in *D.* vulgaris, as *D. vulgaris* transformed with the plasmid pPD3 is resistant to geneticin. However, the kanamycin promoter might have induced an expression level that was too low to be detectable. The choice of a native promoter from *Desulfovibrio* spp. may enhance the recombinant protein expression in *D. vulgaris* (Aubert *et al.*, 1998; Rousset *et al.*, 1998). The expression of the putative reductase and methyltransferase genes was achieved in *E. coli.* The expression of the putative reductase was not confirmed by Western blotting at each attempt. This could be due to incomplete denaturation of the reductase resulting in the inaccessibility of the poly-histidine tag for the Western blot anti-bodies.

The putative reductase and methyltransferase were hypothesised to catalyse the hydrogenolysis of alkyl phosphates into alkanes and *iso*-alkanes respectively. Therefore, *E. coli* expressing either the putative reductase or methyltransferase gene were cultivated in LB medium supplemented with 100 *μ*M tetradecyl-d_{29}-phosphate. The fully deuterated teratrdecyl-d_{29}-phosphate was chosen over the hexadecyl-d_{33}-phosphate, as the teratrdecyl-d_{29}-phosphate had a higher solubility in anhydrous ethanol than the hexadecyl-d_{33}-phosphate. Growth medium enrichment with a higher quantity of isotopically labelled compounds maximised the chances of detecting isotopic enrichment of the target metabolites. Neither deuterated alkanes or unlabelled alkanes produced by *E. coli* which expressed either the putative reductase or methyltransferase gene were detected by GC-MS. Moreover, neither deuterated alkanes or unlabelled alkanes were detected by GC-MS within *E. coli* whole cell protein extracts including the methyltransferase and spiked with the tetradecyl-d_{29}-phosphate solution, after 48 h incubation. These results suggested that the putative reductase and methyltransferase do not catalyse the hydrogenolysis of alkyl phosphates into alkanes. However, the heterologously expressed putative reductase and methyltransferase by *E. coli* were potentially inactive. This presumptive inactivity may be due to protein misfolding, or due

to erroneous or lack of post-transcriptional modifications of the proteins operated by *E. coli* (Derman *et al.*, 1993; González-Montalbán *et al.*, 2007). The presumptive inactivity of the reductase and methyltransferase may also be due to the lack of suitable co-factor, which may be only present in *Desulfovibrio* spp. (Akhtar & Jones, 2014). The presence of the poly-histidine tag in the protein sequence offers the possibility to purify the reductase and methyltransferase and therefore another perspective for functional investigation.

The final strategy adopted to characterise the *Desulfovibrio* alkane biosynthetic pathway was to determine the dark sensitivity of the pathway. *Desulfovibrio* alkane biosynthesis was not proven to be affected by the dark in this study. This deduction revokes the hypothetical involvement of light activated proteins in *Desulfovibrio* alkane biosynthesis. Moreover, it suggested that no homologs of the fatty acid photodecarboxylase from *C. variabilis* NC64A (Sorigué *et al.*, 2017) were present in *D. desulfuricans* NCIMB 8326. Therefore, *Desulfovibrio* alkane biosynthesis likely follows a novel metabolic route, which remains to be characterised.

6. General Discussion

6.1. Strain-Specific Alkane Biosynthesis within the genus Desulfovibrio

Hydrocarbon synthesis by the sulphate-reducing bacterium *D. desulfuricans* was re-visited several times since the first report from the notable founders of microbial ecology Jankowski and ZoBell (1944). The accumulated literature proposes that *D. desulfuricans* synthesises a wide range of intracellular and extracellular alkanes that possess carbon-chain lengths from C_{11} to C_{35} (Oppenheimer, 1965; Davis, 1968; Bagaeva & Chernova, 1994; Bagaeva, 2000). However, the *Desulfovibrio* synthesised alkanes consistently reported are common contaminants from manufactured items and culture media. In this study, ten *Desulfovibrio* strains, representing seven species, were cultivated in deuterated growth medium in order to discriminate biogenic alkanes from non-metabolically derived hydrocarbons. Using isotopically labelled growth medium coupled with high - resolution mass spectrometry enabled the unambiguous confirmation of alkane biosynthesis by *Desulfovibrio* spp. Interestingly, the ability to produce alkane was shown to be strain-specific. *D. desulfuricans* NCIMB 8326, *D. desulfuricans* NCIMB 8338, *D. gabonensis* DSM 10636, *D. gigas* NCIMB 9332, *D. marinus* DSM 18311 and *D. paquesii* DSM 16681 were proven to be capable of octadecane (C_{18}), nonadecane (C_{19}) and eicosane (C_{20}) synthesis, with predominance of even-numbered carbon chain alkanes. Alkane biosynthesis by *D. desulfuricans* NCIMB 8307, *D. vulgaris* Hildenborough, *D. giganteus* DSM 4370 and *D. alcoholivorans* NCIMB 12906 was not detected under the culture conditions used within this study.

The ability to produce alkanes only within a restricted clade of the *Desulfovibrio* genus raises the question of the physiological function of alkanes therein. Intracellular hydrocarbons were previously observed to be components of microbial cell membranes. It was subsequently hypothesised that hydrocarbons play a protective role against adverse environments (Bagaeva & Zinurova, 2004; Ladygina *et al.*, 2006). As the ability to produce alkanes is not ubiquitous within the genus *Desulfovibrio*, alkanes were deduced to be non-essential for growth and thus alkanes were considered secondary metabolites. Secondary metabolites serve survival functions for the micro-organisms which produce them (Demain & Fang, 2000). Recent studies on cyanobacterial alkane production reported that intracellular alkane concentration increases in response to nitrogen deficiency and salt accumulation (Kageyama *et al.*, 2015). Furthermore, alkanes

produced by *Synechococcus* sp. PCC7002 and *Synechocystis* sp. PCC6803 were proven to promote cyanobacterial growth at low temperature (Mendez-Perez *et al.*, 2014; Berla *et al*., 2015). *Desulfovibrio* long and saturated carbon chain alkanes, detected by GC-MS in this study, would decrease the fluidity of the cell membranes and could protect the cells against environmental stresses. It was observed that the divergent salinity tolerance of alkane producing *Desulfovibrio* spp. controverts a putative lifeguarding role of alkanes in response to salt stress.

Alkanes were reported to be one of the carbon sources oxidised by *Desulfovibrio* for growth, as discussed in chapter 1 (*cf.* Chapter 1 – Introduction; section 1.4.1; Novelli & ZoBell, 1944; Rosenfeld, 1947). *D. desulfuricans* NCIMB 8326 was demonstrated to be capable of octadecane mineralisation (Davis & Yarbrough, 1966). Alkane biosynthesis by *Desulfovibrio* could then be an indirect regulation mechanism to maintain growth and so, survival. On this assumption, the absence of biogenic alkane detection by GC-MS within *D. desulfuricans* NCIMB 8307, *D. vulgaris* Hildenborough, *D. giganteus* DSM 4370 and *D. alcoholivorans* NCIMB 12906 organic compound extracts could be explained by a possible metabolic cycle of hydrocarbons. These strains could have produced alkanes under the studied set of culture conditions and then used the biogenic alkanes as source of carbon for their growth.

More widely, *Desulfovibrio* alkanes could also be considered to contribute to the biological hydrocarbon cycle. *Desulfovibrio* spp. were previously assigned to play an important role in the biological carbon and sulphur cycles (Jørgensen, 1982). Moreover, cyanobacterial hydrocarbons were demonstrated to significantly contribute to the hydrocarbon cycle in the ocean. This contribution was further extrapolated to freshwater, marine and terrestrial environments due to the widespread distribution of cyanobacteria (Lea-Smith *et al.*, 2015).

6.2. A Novel Metabolic Route for Alkane Biosynthesis in Desulfovibrio

In this study, the hypothesis was proposed that alkane producing *Desulfovibrio* spp. share a unique alkane biosynthetic pathway, based on the GC-MS detection of identical biogenic alkanes from each of the species investigated. The structural nature of biogenic hydrocarbons is characteristic of the biosynthetic pathway. Structural characterisation of biogenic hydrocarbons allowed the determination of which

hydrocarbon biosynthetic pathway is present in cyanobacterial strains (Coates *et al.*, 2014).

To date all the characterised alkane biosynthetic pathways derive from the fatty acid metabolic pathway. The *Desulfovibrio* alkane biosynthetic pathway was thus assumed to derive from the fatty acid metabolic pathway. The fatty acid content analysis of the *Desulfovibrio* spp. involved in this study revealed that neither saturated carbon chain nonadecanoic acid ($C_{19:0}$) nor saturated carbon chain heneicosanoic acid ($C_{21:0}$) were detected by GC-MS, suggesting that *Desulfovibrio* spp. do not produce these two fatty acids. The biosynthesis of octadecanes (C_{18}) and eicosanes (C_{20}) by *Desulfovibrio* spp. was thus considered unlikely to involve a decarbonylation or decarboxylation step, as identified in all previously characterised alkane biosynthetic pathways. This deduction engendered the hypothesis that alkane production by *Desulfovibrio* spp. involves a series of reduction reactions from fatty acids. *Desulfovibrio* alkane biosynthetic pathway may therefore follow a novel metabolic route which has not previously been characterised.

This hypothesis of a novel reductive pathway for *Desulfovibrio* alkane biosynthesis was endorsed by the absence of characterised alkane biosynthetic enzyme homologs in the alkane producing *Desulfovibrio* spp. In this study, alkane producing genomes and proteomes were screened for homologs of the previously characterised alkane biosynthetic enzymes from bacteria using sequence similarity and protein domain homology. Recently, a new alkane synthase namely fatty acid photodecarboxylase was discovered and characterised to be activated by the light (Sorigué *et al.*, 2017). Thus, the dark sensitivity of *Desulfovibrio* alkane biosynthesis was also investigated. However, *Desulfovibrio* alkane biosynthesis was proven to be unaffected by the dark, suggesting that no fatty acid photodecarboxylase homologs were present in alkane producing *Desulfovibrio* spp. This observation suggested that the *Desulfovibrio* alkane biosynthetic pathway is likely to be catalysed by currently uncharacterised enzymes.

6.3. Acquisition of the Alkane Biosynthetic Pathway in Desulfovibrio via Horizontal Gene Transfer

The phylogenetic distribution of *Desulfovibrio* strains with the ability to produce alkanes based on 16S rRNA gene sequence was established. The 16S rRNA-based

phylogeny revealed that the alkane producing *Desulfovibrio* strains are clustered within a single clade. A novel hypothesis was thus proposed, that the *Desulfovibrio* alkane biosynthetic pathway was acquired by a common ancestral strain *via* horizontal gene transfer. This hypothesis relied on relatively weak support since only ten *Desulfovibrio* strains were investigated for alkane biosynthesis in this study and the *Desulfovibrio* genus encompasses a high diversity of micro-organisms. Furthermore, *Desulfovibrio* taxonomy is frequently revised, as previously discussed in chapters 3 (*cf.* Chapter 3 - Metabolism Screening; Section 3.4) and 4 (cf. chapter 4 - Comparative Genomic Analysis; Section 4.4). Additionally, this hypothesis was based on the 16S rRNA-based phylogeny of this study. As discussed in chapter 4 (cf. chapter 4 - Comparative Genomic Analysis; Section 4.4), phylogeny between prokaryotes of the same genus is unlikely to be established based on a single gene sequence variation, notably due to horizontal gene transfer events (Snel *et al.*, 1999; Doolittle, 1999). However, horizontal gene transfer events were previously demonstrated to be the genesis of hydrocarbon biosynthetic pathway acquisition in other organisms, based on 16S rRNA-based phylogeny analysis. The OLS pathway may have been acquired *via* horizontal gene transfer in the cyanobacteria *Leptolyngbya* sp. PCC 6406 (Coates *et al.*, 2014). Additionally, horizontal gene transfer events played a role in acquisition of genes encoding the isopentenyl diphosphate synthesis among bacteria. Isopentenyl diphosphate is the precursor metabolite of isoprenoids (Boucher & Doolittle, 2000).

Statistical analyses of the GC content or the codon usage of alkane producing *Desulfovibrio* genomes might substantiate the hypothesis that the *Desulfovibrio* alkane biosynthetic pathway was acquired *via* horizontal gene transfer (Garcia-Vallvé *et al.*, 2000; Ravenhall *et al.*, 2015). *Desulfovibrio magneticus* strain RS-1 genes encoding the magnetosome biosynthesis and obtained by horizontal gene transfer from magnetotactic bacteria revealed a significant lower GC content than the genome average GC content (Nakazawa *et al.*, 2009).

Horizontal gene transfer events were proven to be a source of environmental adaptation (Koonin *et al.*, 2001; Polz *et al.*, 2013). *D. vulgaris* Hildenborough genes encoding recombinases, transposases, rubredoxin:oxygen oxidoreductase-1 and hybrid cluster protein-1 and acquired by horizontal gene transfer were demonstrated to promote cell survival under oxygen and nitrite stress (Johnston *et al.,* 2008). With the assumption that *Desulfovibrio* alkane biosynthetic pathway was acquired *via* horizontal gene transfer, alkane synthesis could be an adaptive trait in adverse environments; corroborating the putative protective role of alkanes.

6.4. A Comparative Genomic Strategy for Desulfovibrio Alkane Biosynthetic Pathway Characterisation

With the assumption that the *Desulfovibrio* alkane biosynthetic pathway was acquired *via* horizontal gene transfer, the hypothesis was proposed that the ability of *Desulfovibrio* to produce alkanes is due to the presence of genes encoding enzymes involved in alkane synthesis. This study subsequently progressed towards a comparative genomic analysis of alkane producing and non-alkane producing *Desulfovibrio* spp. through the generation of pan-genomes. Pan-genome analyses provide a framework for comparative genomic analysis (Tettelin *et al.*, 2008; Rouli *et al.*, 2015). Pan-genomes studies were previously proven to be powerful target-directed genome mining tools for the investigation of metabolic networks (Vieira *et al.*, 2011; Mosquera-Rendón *et al.*, 2016) and the discovery of strain-specific characteristics (Auria *et al.*, 2010; Brüggemann *et al.*, 2018).

The *Desulfovibrio* strains involved in the comparative genomic analysis revealed great genomic diversity overall. This genomic diversity was an advantage for this study in that the number of genes exclusively present in the alkane producing strains was found to be a small proportion of the genomes. Thus, it allowed the targeted consideration of the alkane synthetic enzymes. Each of the two software programs that were used for pan-genome generation identified more than 100 proteins that were exclusively present in alkane producing *Desulfovibrio* spp. Of these proteins, only 33 were identified by both programs. Pan-genomic studies rely on the type of input data, the sequence alignment, the homology clustering and the annotation algorithms and parameters (Vernikos *et al.*, 2015). By isolating those 33 proteins that were identified by both pan-genomic software, which implement distinct methods, confidence in the accuracy of protein identification was increased. Moreover, the determination of the proteins that were identified by both pan-genomic software programs enabled a reduction in the number of candidate proteins exclusively present in alkane producing proteomes.

From the 33 proteins predicted to be exclusive to alkane producing *Desulfovibrio* spp., a novel hypothetical *Desulfovibrio* alkane biosynthetic pathway was proposed, which involved a V-type ATPase, an uncharacterised protein, named as a putative reductase in this study, and a putative methyltransferase. The inorganic phosphates resulting from the ATP hydrolysis catalysed by the V-type ATPase was hypothesised to could be involved in a reaction with fatty alcohols to form alkyl phosphates. The oxygen-phosphate bond of alkyl phosphates would weaken the oxygen-carbon bond of fatty

alcohols. Therefore, the alkyl phosphates would be the activated intermediates required for the hydrogenation route from fatty alcohols to alkanes. The putative reductase and the methyltransferase, predicted to share similar structural features with known alkane-binding proteins, would then reduce alkyl phosphates to alkanes and to *iso*-alkanes respectively.

6.5. Further Considerations and Studies for Desulfovibrio Alkane Biosynthetic Pathway Characterisation

The *Desulfovibrio* alkane biosynthetic pathway remains to be fully characterised. Characterisation of the hypothetical *Desulfovibrio* alkane biosynthetic pathway within this study was partly hampered by the defective alkane producing strain transformation methods used. Optimisation of the transformation protocol would allow the deletion of candidate genes for alkane production and consequently the empirical verification of their function. Additionally, the candidate enzymes originating from *D. desulfuricans* NCIMB 8326 could have expressed as inclusion bodies in *E. coli*, which would have impeded the substrate access to the enzymes, and therefore any activity assay. Recombinant proteins heterogeneously expressed in host organisms are usually instable and insoluble, resulting in the formation of inclusion bodies (Rosano & Ceccarelli, 2014). The separate analysis of the crude extract and the membrane fraction protein contents from the organism expressing the candidate enzymes would allow verification that the candidate enzymes are trapped into inclusion bodies. Recombinant proteins trapped into inclusion bodies could be then purified. Furthermore, purification of the candidate enzymes could allow their crystallisation, another perspective for functional investigation. Further biochemical studies could also be performed for the functional characterisation of the candidate enzymes. Both the putative reductase and the methyltransferase were predicted to catalyse the reduction of alkyl phosphates to alkanes. A colorimetric assay using a dye such as methylene blue or resazurin, which change colour once reduced, could provide preliminary verification of the putative reductive function of these two candidate enzymes. Besides, the use of an isotopically labelled fatty acid precursor, such as octadecanoic acid, to supplement alkane producing *Desulfovibrio* medium might confirm the hypothetical reductive route for *Desulfovibrio* alkane biosynthesis from fatty acids. The microalgae hydrocarbon biosynthetic pathway was proven to derive from fatty

acid metabolic pathway, by labelling the microalgae growth medium with deuterated palmitic acid (Sorigué *et al.*, 2016).

Interestingly, among the 33 proteins predicted to be exclusive to alkane producing *Desulfovibrio* spp., a phage terminase large subunit is listed. Horizontal gene transfers from bacteriophage to bacteria are an important source of genome plasticity in prokaryotes (Canchaya *et al.*, 2003). Walker *et al*. compared the genomes of two closely related *D. vulgaris* strains and observed that the greatest genomic differences result in number and type of prophages (Walker *et al.*, 2009). Genomic basis encoding *Desulfovibrio* alkane biosynthetic pathway could be prophage related genes. Analysis of alkane producing *Desulfovibrio* genomes for prophage sequence identification and annotation could provide information towards the characterisation of the *Desulfovibrio* alkane biosynthetic pathway.

However, the metabolic diversity between strains does not systematically correlate with the genomic diversity. The metabolic diversity can be explained by diversity in enzyme expression regulation and activity (Maslov *et al.*, 2009). Three transcription factors were proven to confer nitric oxide and nitrite resistance to *D. desulfuricans* 27774 (Cadby *et al.*, 2017). Moreover, the same phenotype exhibited by different strains, such as antibiotic sensitivity, is not necessarily governed by the same genetic basis (van Opijnen *et al.*, 2016). Therefore, additional holistic approaches such as transcriptomics and/or proteomics approaches would contribute to the elucidation of *Desulfovibrio* alkane biosynthetic pathway.

Taking in to account the great genomic diversity between *Desulfovibrio* spp., the transcriptomics and/or proteomics studies would be more informative if only one alkane producing strain is considered. The gene expression level and the proteins produced by this strain cultured in different conditions, either affecting or enhancing the alkane synthesis, could be compared. As discussed in chapter 3 (*cf.* Chapter 3 - Metabolism Screening; Section 3.4), alkane biosynthesis by *Desulfovibrio* was proven to be affected by the composition of the anaerobic atmosphere (Bagaeva, 2000). Therefore, further studies investigating the effects of physiological conditions or growth stage on *Desulfovibrio* alkane content would be useful.

The *Desulfovibrio* alkane biosynthetic reductive pathway without carbon loss hypothesised in this study sparks a strong interest for petroleum replica production with higher carbon efficiency, compared to the currently characterised hydrocarbon

biosynthetic pathways. Characterisation of the genetic basis for *Desulfovibrio* alkane biosynthesis would enable the implementation of the pathway into another, more suitable host for an industrial exploitation than *Desulfovibrio* spp. The *Desulfovibrio* alkane biosynthetic pathway characterisation would also allow the optimisation of the pathway expression for higher alkane yields, since native alkane yields from wild type *Desulfovibrio* are too low to be cost competitive with petroleum derived fuels. In this study, *Desulfovibrio* productivity for total alkane synthesis was assessed to be 4.7 *μ*g of total alkanes produced per mg of total proteins. Moreover, *Desulfovibrio* biogenic alkanes were showed to be of different carbon chain length. Thus, *Desulfovibrio* alkane synthetic enzymes might have a broad substrate specificity, which gives a potential for alkane synthesis from a variety of fatty acids in the bacterial host.

7. Conclusion

This project aimed to characterise the hydrocarbon biosynthetic pathway(s) in the sulphate-reducing bacteria genus *Desulfovibrio*. To this end, the hydrocarbon biosynthesis in the *Desulfovibrio* genus was confirmed. Ten *Desulfovibrio* strains, representing seven species, were screened for alkane biosynthesis, using isotopically labelled medium coupled to high - resolution mass spectrometry. *D. desulfuricans* NCIMB 8326, *D. desulfuricans* NCIMB 8338, *D. gabonensis* DSM 10636, *D. gigas* NCIMB 9332, *D. marinus* DSM 18311 and *D. paquesii* DSM 16681 were proven to be capable of octadecane (C_{18}), nonadecane (C_{19}) and eicosane (C_{20}) synthesis, with higher quantity of even numbered carbon chain alkanes synthesis. The productivity for total alkane synthesis by *Desulfovibrio* spp. was determined as 4.7 *μ*g total alkanes produced per mg of total proteins. However, alkane biosynthesis by *D. desulfuricans* NCIMB 8307, *D. vulgaris* Hildenborough, *D. giganteus* DSM 4370 and *D. alcoholivorans* NCIMB 12906 was not detected.

All the alkane producing *Desulfovibrio* strains in this study synthesise the same alkanes, suggesting that these strains shared a unique alkane biosynthetic pathway. The fatty acid content analysis of the ten *Desulfovibrio* spp. revealed that alkane biosynthesis was unlikely to involve a decarbonylation or decarboxylation step, as identified in all previously characterised alkane biosynthetic pathways. On that basis, we hypothesised that the *Desulfovibrio* alkane biosynthetic pathway may therefore follow a novel metabolic route, which has not previously been characterised, involving a series of reduction reactions from fatty acids.

This project then focused on the *in silico* identification of candidate genes for *Desulfovibrio* alkane biosynthesis through target-directed genome mining. The genomic DNA from nine *Desulfovibrio* strains was purified, sequenced, assembled *de novo* and annotated. The genome of *D. vulgaris* strain Hildenborough was previously completely sequenced and publically available. No homologs of the previously characterised alkane biosynthetic enzymes from bacteria were *in silico* identified in the genomes and proteomes of alkane producing *Desulfovibrio* spp., suggesting that *Desulfovibrio* alkane biosynthetic pathway is likely to be catalysed by currently uncharacterised enzymes. This observation endorses the hypothesis that *Desulfovibrio* alkane biosynthetic pathway may follow a novel metabolic route.

The 16S rRNA-based phylogeny of *Desulfovibrio* spp. supported the hypothesis that *Desulfovibrio* alkane biosynthetic pathway was acquired by a common ancestral strain *via* horizontal gene transfer. The ability of *Desulfovibrio* to produce alkanes was therefore hypothesised to be due to the presence of recruited genes encoding enzymes involved in alkane synthesis. This project subsequently progressed towards a comparative genomic analysis of six alkane producing and four non-alkane producing *Desulfovibrio* genomes. Thirty-three proteins were predicted *in silico* as being exclusive to alkane producing *Desulfovibrio* spp. A novel hypothetical *Desulfovibrio* alkane biosynthetic pathway was proposed involving a V-type ATPase, an uncharacterised protein, named as a putative reductase in this study, and a putative methyltransferase, which were predicted to be exclusive to alkane producing *Desulfovibrio* spp. The inorganic phosphates resulting from the ATP hydrolysis catalysed by the V-type ATPase would be involved in a reaction with fatty alcohols to form alkyl phosphates, putative activated intermediates required for the hydrogenation route from fatty alcohols to alkanes. The putative reductase and the methyltransferase, predicted to share similar structural features with known alkane-binding proteins, would subsequently reduce alkyl phosphates to alkanes and to *iso*-alkanes respectively.

Several strategies were adopted for empirical verification of the candidate genes for *Desulfovibrio* alkane biosynthesis. Molecular tools for engineering *Desulfovibrio* were designed and constructed. Development of an efficient and reliable transformation method for the alkane producing *Desulfovibrio* was exhaustively trialled but without sufficient success. However, *D. vulgaris* strain Hildenborough, proven to be non-alkane producer, was successfully transformed and used subsequently as heterologous host for genes candidate for alkane synthesis. Heterologous expression of candidate genes for *Desulfovibrio* alkane biosynthesis was also performed in *E. coli*. Additionally, isotopically labelled alkyl phosphates, the predicted metabolite precursors of alkanes, were used to supplement alkane producing strain medium. The final strategy adopted to characterise the *Desulfovibrio* alkane biosynthetic pathway was to determine the dark sensitivity of the pathway. However, the *Desulfovibrio* alkane biosynthetic pathway remains to be fully characterised.

Bibliography

Adams, C.J., Redmond, M.C. & Valentine, D.L. (2006) Pure-Culture Growth of Fermentative Bacteria, Facilitated by H_2 Removal: Bioenergetics and H_2 Production. *Applied and Environmental microbiology*, **72**; 1079–1085.

Agarwal, A.K. (2007) Biofuels (alcohols and biodiesel) applications as fuels for internal combustion engines. *Progress in Energy and Combustion Science*, **33**; 233–271.

Akagi, J.M. & Campbell, L.L. (1962) Studies on thermophilic sulfate-reducing bacteria III. Adenosine Triphosphate-sulfurylase of *Clostridium nigrificans* and *Desulfovibrio desulfuricans*. *Journal of Bacteriology*, **84**; 1194–1201.

Akhtar, M.K. & Jones, P.R. (2014) Cofactor Engineering for Enhancing the Flux of Metabolic Pathways. *Frontiers in Bioengineering and Biotechnology*, **2**; 1–6.

Alikhan, N., Petty, N.K., Zakour, N.L. Ben & Beatson, S.A. (2011) BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics*, **12**; 402.

Altschup, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) Basic Local Alignment Search Tool. *Journal of Molecular Biology*, **215**; 403–410.

Andrews, S. (2010) FastQC - A quality control tool for high throughput sequence data.

Argyle, J.L., Rapp-Giles, B.J. & Wall, J.D. (1992) Plasmid Transfer By Conjugation In *Desulfovibrio desulfuricans*. *Federation of European Microbiological Societies Microbiology Letters*, **94**; 255–262.

Aubert, C., Leroy, G., Bruschi, M., Wall, J.D. & Dolla, A. (1997) A Single Mutation in the Heme 4 Environment of *Desulfovibrio desulfuricans* Norway Cytochrome c_3 (*M*r 26,000) Greatly Affects the Molecule Reactivity. *The Journal of Biological Chemistry*, **272**; 15128–15134.

Aubert, C., Lojou, E., Bianco, P., Rousset, M., Durand, M., Bruschi, M. & Dolla, A. (1998) The *Desulfuromonas acetoxidans* Triheme Cytochrome c_7 Produced in *Desulfovibrio desulfuricans* Retains Its Metal Reductase Activity. *Applied and Environmental microbiology*, **64**; 1308–1312.

Auria, G.D., Jiménez-Hernández, N., Peris-Bondia, F., Moya, A. & Latorre, A. (2010) *Legionella pneumophila* pangenome reveals strain-specific virulence factors. *BMC Genomics*, **11**; 1–13.

Badwal, S.P.S., Giddey, S.S., Munnings, C., Bhatt, A.I. & Hollenkamp, A.F. (2014) Emerging electrochemical energy conversion and storage technologies. *Frontiers in Chemistry*, **2**; 1–28.

Baena, S., Fardeau, M.L., Labat, M., Ollivier, B., Garcia, J.L. & Patel, B.K. (1998) *Desulfovibrio aminophilus* sp. nov., a novel amino acid degrading and sulfate reducing bacterium from an anaerobic dairy wastewater lagoon. *Systematic and Applied Microbiology*, **21**; 498–504.

Bagaeva, T.V. (2000) Effects of composition of the gaseous phase on the formation of hydrocarbons in *Desulfovibrio desulfuricans*. *Applied Biochemistry and Microbiology*, **36**; 165–168.

Bagaeva, T.V. & Beliaeva, M.I. (2000) Effect of various organic compounds on the growth and hydrocarbon production by sulfur-reducing bacteria. *Izvestiia Akademii Nauk. Seriia Biologicheskaia,* **May**-**June**; 382–386.

Bagaeva, T.V. & Chernova, T.G. (1994) Comparative characteristics of extracellular and intracellular hydrocarbons of *Desulfovibrio desulfuricans. Biochemistry (Moscow)*, **59**; 23–24.

Bagaeva, T.V. & Zinurova, E.E. (2004) Comparative characterization of extracellular and intracellular hydrocarbons of *Clostridium pasteurianum*. *Biochemistry (Moscow)*, **69**; 427–428.

Bak, F. & Pfennig, N. (1991) Sulfate-reducing bacteria in littoral sediment of Lake Constance. *FEMS Microbiology Ecology*, **8**; 43–52.

Balat, M. & Balat, H. (2010) Progress in biodiesel processing. *Applied Energy*, **87**; 1815–1835.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A. V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A. & Pevzner, P.A. (2012) SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*, **19**; 455–477.

Bansal, A.K. & Meyer, T.E. (2002) Evolutionary Analysis by Whole-Genome Comparisons. *Journal of bacteriology*, **184**; 2260–2272.

Baran, R., Bowen, B.P., Bouskill, N.J., Brodie, E.L., Yannone, S.M. & Northen, T.R. (2010) Metabolite Identification in *Synechococcus* sp. PCC 7002 Using Untargeted Stable Isotope Assisted Metabolite Profiling. *Analytical Chemistry*, **82**; 9034–9042.

Bateman, A., Martin, M.J., O'Donovan, C., Magrane, M., Alpi, E., Antunes, R., Bely, B., Bingley, M., Bonilla, C., Britto, R., Bursteinas, B., Bye-AJee, H., Cowley, A., Silva, A. Da, Giorgi, M. De, Dogan, T., Fazzini, F., Castro, L.G., Figueira, L., Garmiri, P., Georghiou, G., Gonzalez, D., Hatton-Ellis, E., Li, W., Liu, W., Lopez, R., Luo, J., Lussi, Y., MacDougall, A., Nightingale, A., Palka, B., Pichler, K., Poggioli, D., Pundir, S., Pureza, L., Qi, G., Rosanoff, S., Saidi, R., Sawford, T., Shypitsyna, A., Speretta, E., Turner, E., Tyagi, N., Volynkin, V., Wardell, T., Warner, K., Watkins, X., Zaru, R., Zellner, H., Xenarios, I., Bougueleret, L., Bridge, A., Poux, S., Redaschi, N., Aimo, L., ArgoudPuy, G., Auchincloss, A., Axelsen, K., Bansal, P., Baratin, D., Blatter, M.C., Boeckmann, B., Bolleman, J., Boutet, E., Breuza, L., Casal-Casas, C., Castro, E. De, Coudert, E., Cuche, B., Doche, M., Dornevil, D., Duvaud, S., Estreicher, A., Famiglietti, L., Feuermann, M., Gasteiger, E., Gehant, S., Gerritsen, V., Gos, A., Gruaz-Gumowski, N., Hinz, U., Hulo, C., Jungo, F., Keller, G., Lara, V., Lemercier, P., Lieberherr, D., Lombardot, T., Martin, X., Masson, P., Morgat, A., Neto, T., Nouspikel, N., Paesano, S., Pedruzzi, I., Pilbout, S., Pozzato, M., Pruess, M., Rivoire, C., Roechert, B., Schneider, M., Sigrist, C., Sonesson, K., Staehli, S., Stutz, A., Sundaram, S., Tognolli, M., Verbregue, L., Veuthey, A.L., Wu, C.H., Arighi, C.N., Arminski, L., Chen, C., Chen, Y., Garavelli, J.S., Huang, H., Laiho, K., McGarvey, P., Natale, D.A., Ross, K., Vinayaka, C.R., Wang, Q., Wang, Y., Yeh, L.S. & Zhang, J. (2017) UniProt: The universal protein knowledgebase. *Nucleic Acids Research*, **45**; D158–D169.

Battin-Leclerc, F. & Blurock, E. (2011) Towards cleaner combustion engines through groundbreaking detailed chemical kinetic models. *Chemical Society Review*, **40**; 4762–4782.

Bellard, C., Bertelsmeier, C., Leadley, P., Thuiller, W. & Courchamp, F. (2014) Impacts of Climate Change on the Future of Biodiversity. *Ecology Letters*, **15**; 365–377.

Beller, H.R., Goh, E.B. & Keasling, J.D. (2010) Genes involved in long-chain alkene biosynthesis in *Micrococcus luteus*. *Applied and Environmental Microbiology*, **76**; 1212–1223.

Bender, K.S., Cheyen, H. & Wall, J.D. (2006) Analysing the Metabolic Capabilities of *Desulfovibrio* Species through Genetic Manipulation. *Biotechnology and Genetic Engineering Reviews*, **23**; 157–174.

Bender, K.S., Yen, H.C.B., Hemme, C.L., Yang, Z., He, Z., He, Q., Zhou, J., Huang, K.H., Alm, E.J., Hazen, T.C., Arkin, A.P. & Wall, J.D. (2007) Analysis of a ferric uptake regulator (Fur) mutant of *Desulfovibrio vulgaris* Hildenborough. *Applied and Environmental Microbiology*, **73**; 5389–5400.

Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. & Sayers, E.W. (2013) GenBank. *Nucleic Acids Research*, **41**; 36–42.

Berla, B.M., Saha, R., Maranas, C.D. & Pakrasi, H.B. (2015) Cyanobacterial Alkanes Modulate Photosynthetic Cyclic Electron Flow to Assist Growth under Cold Stress. *Scientific Reports*, **5**; 1–12.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. & Bourne, P.E. (2000) The Protein Data Bank. *Nucleic acids research*, **28**; 235–242.

Bernard, A., Domergue, F., Pascal, S., Jetter, R., Renne, C., Faure, J.-D., Haslam, R.P., Napier, J. a., Lessire, R. & Joubes, J. (2012) Reconstitution of Plant Alkane Biosynthesis in Yeast Demonstrates That *Arabidopsis* ECERIFERUM1 and ECERIFERUM3 Are Core Components of a Very-Long-Chain Alkane Synthesis Complex. *The Plant Cell*, **24**; 3106–3118.

Biebl, H. & Pfennig, N. (1978) Growth yields of green sulfur bacteria in mixed cultures with sulfur and sulfate reducing bacteria. *Archives of Microbiology*, **117**; 9–16.

Boon, J.J., Deleeuw, J.W., Vanderhoek, G.J. & Vosjan, J.H. (1977) Significance and Taxonomic Value of Iso and Anteiso Monoenoic Fatty-Acids and Branched β-Hydroxy Acids in *Desulfovibrio desulfuricans. Journal of Bacteriology*, **129**; 1183–1191.

Boucher, Y. & Doolittle, W.F. (2000) The role of lateral gene transfer in the evolution of isoprenoid biosynthesis pathways. *Molecular Microbiology*, **37**; 703–716.

Brandis, A. & Thauer, R.K. (1981) Growth of *Desulfovibrio* species on Hydrogen and Sulphate as Sole Energy Source. *Microbiology*, **126**; 249–252.

British Petroleum. (2017) *Statistical Review of World Energy 2017*.

Broco, M., Rousset, M., Oliveira, S. & Rodrigues-Pousada, C. (2005) Deletion of flavoredoxin gene in *Desulfovibrio gigas* reveals its participation in thiosulfate reduction. *FEBS Letters*, **579**; 4803–4807.

Brown, S.D., Nagaraju, S., Utturkar, S., Tissera, S. De, Segovia, S., Mitchell, W., Land, M.L., Dassanayake, A. & Köpke, M. (2014) Comparison of single-molecule sequencing and hybrid approaches for finishing the genome of *Clostridium autoethanogenum* and analysis of CRISPR systems in industrial relevant *Clostridia*. *Biotechnology for Biofuels*, **7**; 1–18.

Brown, S.D., Wall, J.D., Kucken, A.M., Gilmour, C.C., Podar, M., Brandt, C.C., Teshima, H., Detter, J.C., Han, C.S., Land, M.L., Lucas, S., Han, J., Pennacchio, L., Nolan, M., Pitluck, S., Woyke, T., Goodwin, L., Palumbo, A.V. & Elias, D.A. (2011) Genome Sequence of the Mercury-Methylating and Pleomorphic *Desulfovibrio africanus* Strain Walvis Bay. *Journal of bacteriology*, **193**; 4037–4038.

Brüggemann, H., Jensen, A., Nazipi, S., Aslan, H., Meyer, R.L., Poehlein, A., Brzuszkiewicz, E., Al-Zeer, M.A., Brinkmann, V. & Söderquist, B. (2018) Pan-genome analysis of the genus *Finegoldia* identifies two distinct clades, strain-specific heterogeneity, and putative virulence factors. *Scientific Reports*, **8**; 1–11.

Bryant, M.P., Campbell, L.L., Reddy, C.A. & Crabill, M.R. (1977) Growth of *Desulfovibrio* in lactate or ethanol media low in sulfate in association with H_2-utilizing methanogenic bacteria. *Applied and Environmental Microbiology*, **33**; 1162–1169.

Buchfink, B., Xie, C. & Huson, D.H. (2015) Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, **12**; 59–60.

Cadby, I.T., Faulkner, M., Cheneby, J., Long, J., Helden, J. Van, Dolla, A. & Cole, J.A. (2017) Coordinated response of the *Desulfovibrio desulfuricans* 27774 transcriptome to nitrate, nitrite and nitric oxide. *Scientific Reports*, **7**; 1–16.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. & Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**; 421.

Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillmann, M.L. & Brüssow, H. (2003) Phage as agents of lateral gene transfer. *Current Opinion in Microbiology*, **6**; 417–424.

Cannac, V., Caffrey, M.S., Voordouw, G. & Cusanovich, M.A. (1991) Expression of the gene encoding cytochrome c_3 from the sulfate-reducing bacterium *Desulfovibrio vulgaris* in the purple photosynthetic bacterium *Rhodobacter sphaeroides*. *Archives of biochemistry and biophysics*, **286**; 629–632.

Cao, J., Gayet, N., Zeng, X., Shao, Z., Jebbar, M. & Alain, K. (2016) *Pseudodesulfovibrio indicus* gen. nov., sp. nov., a piezophilic sulfate-reducing bacterium from the Indian Ocean and reclassification of four species of the genus *Desulfovibrio*. *International Journal of Systematic and Evolutionary Microbiology*, **66**; 3904–3911.

Carepo, M., Baptista, J.F., Pamplona, A., Fauque, G., Moura, J.J.G. & Reis, M.A.M. (2002) Hydrogen metabolism in *Desulfovibrio desulfuricans* strain New Jersey (NCIMB 8313) - Comparative study with *D. vulgaris* and *D. gigas* species. *Anaerobe*, **8**; 325–332.

Casalot, L., de Luca, G., Dermoun, Z., Rousset, M. & de Philip, P. (2002) Evidence for a fourth hydrogenase in *Desulfovibrio fructosovorans*. *Journal of Bacteriology*, **184**; 853–856.

Chen, Y., Cheng, J.J. & Creamer, K.S. (2008) Inhibition of anaerobic digestion process: A review. *Bioresource Technology*, **99**; 4044–4064.

Chhabra, S.R., He, Q., Huang, K.H., Gaucher, S.P., Alm, E.J., He, Z., Hadi, M.Z., Hazen, T.C., Wall, J.D., Zhou, J., Arkin, A.P. & Singh, A.K. (2006) Global Analysis of Heat Shock Response in *Desulfovibrio vulgaris* Hildenborough. *Society*, **188**; 1817–1828.

Chhabra, S.R., Butland, G., Elias, D.A., Chandonia, J.M., Fok, O.Y., Juba, T.R., Gorur, A., Allen, S., Leung, C.M., Keller, K.L., Reveco, S., Zane, G.M., Semkiw, E., Prathapam, R., Gold, B., Singer, M., Ouellet, M., Szakal, E.D., Jorgens, D., Price, M.N., Witkowska, H.E., Beller, H.R., Arkin, A.P., Hazen, T.C., Biggin, M.D., Auer, M., Wall, J.D. & Keasling, J.D. (2011) Generalized schemes for high-throughput manipulation of the *Desulfovibrio vulgaris* genome. *Applied and Environmental Microbiology*, **77**; 7595–7604.

Chhetri, G., Kalita, P. & Tripathi, T. (2015) An efficient protocol to enhance recombinant protein expression using ethanol in *Escherichia coli. MethodsX*, **2**; 385–391.

Choi, Y.J. & Lee, S.Y. (2013) Microbial production of short-chain alkanes. *Nature*, **502**; 571–574.

Chokkathukalam, A., Kim, D.H., Barrett, M.P., Breitling, R. & Creek, D.J. (2014) Stable isotope-labeling studies in metabolomics: new insights into structure and dynamics of metabolic networks. *Bioanalysis*, **6**; 511–524.

Christia-Lotter, A., Bartoli, C., Piercecchi-Marti, M.D., Demory, D., Pelissier-Alicot, A.L., Sanvoisin, A. & Leonetti, G. (2007) Fatal occupational inhalation of hydrogen sulfide. *Forensic Science International*, **169**; 206–209.

Coates, R.C., Podell, S., Korobeynikov, A., Lapidus, A., Pevzner, P., Sherman, D.H., Allen, E.E., Gerwick, L. & Gerwick, W.H. (2014) Characterization of cyanobacterial hydrocarbon composition and distribution of biosynthetic pathways. *PLoS ONE*, **9**.

Connor, M.R. & Atsumi, S. (2010) Synthetic biology guides biofuel production. *Journal of Biomedicine and Biotechnology*, **2010**; 1-9.

Connor, M.R. & Liao, J.C. (2009) Microbial production of advanced transportation fuels in non-natural hosts. *Current Opinion in Biotechnology*, **20**; 307–15.

Contreras-Moreira, B. & Vinuesa, P. (2013) GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Applied and Environmental Microbiology*, **79**; 7696–7701.

Cooney, M.J., Roschi, E., Marison, I.W., Comninellis, C. & Von Stockar, U. (1996) Physiologic studies with the sulfate-reducing bacterium *Desulfovibrio desulfuricans*: Evaluation for use in a biofuel cell. *Enzyme and Microbial Technology*, **18**; 358–365.

Crépin, L., Lombard, E. & Guillouet, S.E. (2016) Metabolic engineering of *Cupriavidus necator* for heterotrophic and autotrophic alka(e)ne production. *Metabolic Engineering*, **37**; 92–101.

Croese, E., Pereira, M.A., Euverink, G.J.W., Stams, A.J.M. & Geelhoed, J.S. (2011) Analysis of the microbial community of the biocathode of a hydrogen-producing microbial electrolysis cell. *Applied Microbiology and Biotechnology*, **92**; 1083–1093.

Cui, M., Ma, A., Qi, H., Zhuang, X. & Zhuang, G. (2015) Anaerobic oxidation of methane: An “active” microbial process. *MicrobiologyOpen*, **4**; 1–11.

da Costa, P.N., Conte, C. & Saraiva, L.M. (2000) Expression of a *Desulfovibrio* tetraheme cytochrome *c* in *Escherichia coli*. *Biochemical and Biophysical Research Communications*, **268**; 688–691.

da Silva, S.M., Voordouw, J., Leitão, C., Martins, M., Voordouw, G. & Pereira, I.A.C. (2013) Function of formate dehydrogenases in *Desulfovibrio vulgaris* Hildenborough energy metabolism. *Microbiology*, **159**; 1760–1769.

da Silva, S.M., Amaral, C., Neves, S.S., Santos, C., Pimentel, C. & Rodrigues-Pousada, C. (2015) An HcpR paralog of *Desulfovibrio gigas* provides protection against nitrosative stress. *FEBS Open Bio*, **5**; 594–604.

Davis, J.B. (1968) Paraffinic hydrocarbons in the sulfate-reducing bacterium *Desulfovibrio desulfuricans*. *Chemical Geology*, **3**; 155–160.

Davis, J.B. & Yarbrough, H.F. (1966) Anaerobic oxidation of hydrocarbons by *Desulfovibrio desulfuricans*. *Chemical Geology*, **1**; 137–144.

de León, K.B., Zane, G.M., Trotter, V. V., Krantz, G.P., Arkin, A.P., Butland, G.P., Walian, P.J., Fields, M.W. & Wall, J.D. (2017) Unintended laboratory-driven evolution reveals genetic requirements for biofilm formation by *Desulfovibrio vulgaris* Hildenborough. *mBio*, **8**; 1–16.

de Poulpiquet, A., Ranava, D., Monsalve, K., Giudici-Orticoni, M.T. & Lojou, E. (2014) Biohydrogen for a New Generation of H2/O2 Biofuel Cells: A Sustainable Energy Perspective. *ChemElectroChem Reviews*, **1**; 1724–1750.

Demain, A.L. & Fang, A. (2000) The Natural Functions of Secondary Metabolites. *Advances in Biochemical Engineering/Biotechnology*, **69**; 1–39.

Derman, A., Prinz, W., Belin, D. & Beckwith, J. (1993) Mutations that allow disulfide bond formation in the cytoplasm of *Escherichia coli*. *Science*, **262**; 1744–1747.

Dolla, A., Fournier, M. & Dermoun, Z. (2006) Oxygen defense in sulfate-reducing bacteria. *Journal of Biotechnology*, **126**; 87–100.

Doney, S.C., Fabry, V.J., Feely, R.A. & Kleypas, J.A. (2009) Ocean Acidification: The Other CO_2 Problem. *Annual Review of Marine Science*, **1**; 169–192.

Doolittle, W.F. (1999) Phylogenetic Classification and the Universal Tree. *Science*, **284**; 2124–2129.

Duong, T.X., Suruda, A.J. & Maier, L.A. (2001) Interstitial fibrosis following hydrogen sulfide exposure. *American Journal of Industrial Medecine*, **40**; 221–224.

Dzierzewicz, Z., Cwalina, B., Kurkiewicz, S., Chodurek, E. & Wilczok, T. (1996) Intraspecies variability of cellular fatty acids among soil and intestinal strains of *Desulfovibrio desulfuricans*. *Applied and Environmental Microbiology*, **62**; 3360–3365.

Dzierzewicz, Z., Szczerba, J., Lodowska, J., Wolny, D., Gruchlik, A., Orchel, A. & Weglarz, L. (2010) The role of *Desulfovibrio desulfuricans* lipopolysaccharides in modulation of periodontal inflammation through stimulation of human gingival fibroblasts. *Archives of Oral Biology*, **55**; 515–522.

Earley, H., Lennon, G., Balfe, A., Kilcoyne, M., Clyne, M., Joshi, L., Carrington, S., Martin, S.T., Coffey, J.C., Winter, D.C. & O'Connell, P.R. (2015) A preliminary study examining the binding capacity of *Akkermansia muciniphila* and *Desulfovibrio* spp., to colonic mucin in health and ulcerative colitis. *PLoS ONE*, **10**.

Eddy, S.R. (1996) Hidden Markov models. *Current Opinion in Structural Biology*, **6**; 361–365.

Edlund, A., Nichols, P.D., Roffey, R. & White, D.C. (1985) Extractable and lipopolysaccharide fatty acid and hydroxy acid profiles from *Desulfovibrio* species. *Journal of Lipid Research*, **26**; 982–988.

Ekblom, R. & Wolf, J.B.W. (2014) A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications*, **7**; 1026–1042.

Elobeid, A., Carriquiry, M.A., Dumortier, J., Rosas, F., Mulik, K., Fabiosa, J., Hayes, D.J. & Babcock, B.A. (2013) Biofuel Expansion, Fertilizer Use, and GHG Emissions: Unintended Consequences of Mitigation Policies. *Economics Research International*, **2013**; 1–12.

Enning, D. & Garrelfs, J. (2014) Corrosion of iron by sulfate-reducing bacteria: New views of an old problem. *Applied and Environmental Microbiology*, **80**; 1226–1236.

Eren, A.M., Esen, C., Quince, C., Vineis, J.H., Morrison, H.G., Sogin, M.L. & Delmont, T.O. (2015) Anvi ’o : an advanced analysis and visualization platform for ‘ omics data. *PeerJ*, **3**.

Farhadian, M., Vachelard, C., Duchez, D. & Larroche, C. (2008) *In situ* bioremediation of monoaromatic pollutants in groundwater: A review. *Bioresource Technology*, **99**; 5296–5308.

Fialkov, A.B., Steiner, U., Lehotay, S.J. & Amirav, A. (2007) Sensitivity and noise in GC-MS: Achieving low limits of detection for difficult analytes. *International Journal of Mass Spectrometry*, **260**; 31–48.

Fidler, B., Sublette, K., Jenneman, G. & Bala, G. (2003) A Novel Approach to Hydrogen Sulfide Removal From Natural Gas. *Society of Petroleum Engineers (SPE)*, **81203**.

Findley, J.E. & Akagi, J.M. (1969) Evidence for Thiosulfate Formation during Sulfite Reduction by *Desulfovibrio vulgaris*. *Biochemical and Biophysical Research Communications*, **36**; 266–271.

Finegold, S.M. (2011) *Desulfovibrio* species are potentially important in regressive autism. *Medical Hypotheses*, **77**; 270–274.

Forslund, K. & Sonnhammer, E.L.L. (2008) Predicting protein function from domain content. *Bioinformatics*, **24**; 1681–1687.

Frias, J.A., Richman, J.E., Erickson, J.S. & Wackett, L.P. (2011) Purification and characterization of OleA from *Xanthomonas campestris* and demonstration of a non-decarboxylative claisen condensation reaction. *Journal of Biological Chemistry*, **286**; 10930–10938.

Fu, C., Donovan, W.P., Shikapwashya-Hasser, O., Ye, X. & Cole, R.H. (2014) Hot fusion: An efficient method to clone multiple DNA fragments as well as inverted repeats without ligase. *PLoS ONE*, **9**.

Fu, F. & Wang, Q. (2011) Removal of heavy metal ions from wastewaters: A review. *Journal of Environmental Management*, **92**; 407–418.

Fu, R. & Voordouw, G. (1997) Targeted gene-replacement mutagenesis of *dcrA*, encoding an oxygen sensor of the sulfate-reducing bacterium *Desulfovibrio vulgaris* Hildenborough. *Microbiology*, **143**; 1815–1826.

Fu, R. & Voordouw, G. (1998) IS*D1*, an insertion element from the sulfate-reducing bacterium *Desulfovibrio vulgaris* Hildenborough: Structure, transposition, and distribution. *Applied and Environmental Microbiology*, **64**; 53–61.

Fu, W.J., Chi, Z., Ma, Z.C., Zhou, H.X., Liu, G.L., Lee, C.F. & Chi, Z.M. (2015) Hydrocarbons, the advanced biofuels produced by different organisms, the evidence that alkanes in petroleum can be renewable. *Applied Microbiology and Biotechnology*, **99**; 7481–7494.

García-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L.M., Götz, S., Tarazona, S., Dopazo, J., Meyer, T.F. & Conesa, A. (2012) Qualimap: Evaluating next-generation sequencing alignment data. *Bioinformatics*, **28**; 2678–2679.

García-Vallvé, S., Romeu, A. & Palau, J. (2000) Horizontal gene transfer in bacterial and archaeal complete genomes. *Genome research*, **10**; 1719–1725.

Gebhart, C.J., Barns, S.M., McOrist, S., Lin, G.F. & Lawson, G.H. (1993) Ileal symbiont intracellularis, an obligate intracellular bacterium of porcine intestines showing a relationship to *Desulfovibrio* species. *International Journal of Systematic Bacteriology*, **43**; 533–538.

Gilmour, C.C., Elias, D.A., Kucken, A.M., Brown, S.D., Palumbo, A. V., Schadt, C.W. & Wall, J.D. (2011) Sulfate-reducing bacterium *Desulfovibrio desulfuricans* ND132 as a model for understanding bacterial mercury methylation. *Applied and Environmental Microbiology*, **77**; 3938–3951.

Goldemberg, J., Coelho, S.T., Nastari, P.M. & Lucon, O. (2004) Ethanol learning curve - The Brazilian experience. *Biomass and Bioenergy*, **26**; 301–304.

Goldstein, E.J.C., Citron, D.M., Peraino, V.A. & Cross, S.A. (2003) *Desulfovibrio desulfuricans* bacteremia and review of human *Desulfovibrio* infections. *Journal of Clinical Microbiology*, **41**; 2752–2754.

González-Montalbán, N., García-Fruitós, E. & Villaverde, A. (2007) Recombinant protein solubility - does more mean better? *Nature biotechnology*, **25**; 718–720.

Graham-Rowe, D. (2011) Agriculture: Beyond food versus fuel. *Nature*, **474**; 6–8.

Groh, J.L., Luo, Q., Ballard, J.D. & Krumholz, L.R. (2005) A method adapting microarray technology for signature-tagged mutagenesis of *Desulfovibrio desulfuricans* G20 and *Shewanella oneidensis* MR-1 in anaerobic sediment survival experiments. *Applied and Environmental Microbiology*, **71**; 7064–7074.

Grossman, J.P. & Postgate, J.R. (1953) Cultivation of sulphate-reducing bacteria. *Nature*, **171**; 600–602.

Gunaseelan, V.N. (1997) Anaerobic digestion of biomass for methane production: A review. *Biomass and Bioenergy*, **13**; 83–114.

Guo, Z., Nielsen, S., Burguera, B. & Jensen, M.D. (1997) Free fatty acid turnover measured using ultralow doses of [U-^{13}C]palmitate. *Journal of Lipid Research*, **38**; 1888–1895.

Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. (2013) QUAST : quality assessment tool for genome assemblies. *Bioinformatics*, **29**; 1072–1075.

Hamilton, W.A. (1998) Bioenergetics of sulphate-reducing bacteria in relation to their environmental impact. *Biodegradation*, **9**; 201–212.

Hanahan, D. (1985) Techniques for transformation of E. coli. In *DNA Cloning: A Practical Approach* (ed. Glover, D.M.). IRL Press, Oxford, pp. 109–135.

Haouari, O., Fardeau, M.L., Casalot, L., Tholozan, J.L., Hamdi, M. & Ollivier, B. (2006) Isolated of sulfate-reducing bacteria from Tunisian marine sediments and description of *Desulfovibrio bizertensis* sp. nov. *International Journal of Systematic and Evolutionary Microbiology*, **56**; 2909–2913.

Haschke, R.H. & Campbell, L.L. (1971) Thiosulfate reductase of *Desulfovibrio vulgaris*. *Journal of Bacteriology*, **106**; 603–607.

Hauser, L.J., Land, M.L., Brown, S.D., Larimer, F., Keller, K.L., Rapp-Giles, B.J., Price, M.N., Lin, M., Bruce, D.C., Detter, J.C., Tapia, R., Han, C.S., Goodwin, L.A., Cheng, J.F., Pitluck, S., Copeland, A., Lucas, S., Nolan, M., Lapidus, A.L., Palumbo, A. V. & Wall, J.D. (2011) Complete genome sequence and updated annotation of *Desulfovibrio alaskensis* G20. *Journal of Bacteriology*, **193**; 4268–4269.

He, Q., Huang, K.H., He, Z., Alm, E.J., Fields, M.W., Hazen, T.C., Arkin, A.P., Wall, J.D. & Zhou, J. (2006) Energetic consequences of nitrite stress in *Desulfovibrio vulgaris* Hildenborough, inferred from global transcriptional analysis. *Applied and Environmental Microbiology*, **72**; 4370–4381.

Head, I.M. & Gray, N.D. (2016) Microbial Biotechnology 2020; microbiology of fossil fuel resources. *Microbial Biotechnology*, **9**; 626–634.

Heather, J.M. & Chain, B. (2016) The sequence of sequencers: The history of sequencing DNA. *Genomics*, **107**; 1–8.

Heidelberg, J.F., Seshadri, R., Haveman, S.A., Hemme, C.L., Paulsen, I.T., Kolonay, J.F., Eisen, J.A., Ward, N., Methe, B., Brinkac, L.M., Daugherty, S.C., Deboy, R.T., Dodson, R.J., Durkin, A.S., Madupu, R., Nelson, W.C., Sullivan, S.A., Fouts, D., Haft, D.H., Selengut, J., Peterson, J.D., Davidsen, T.M., Zafar, N., Zhou, L., Radune, D., Dimitrov, G., Hance, M., Tran, K., Khouri, H., Gill, J., Utterback, T.R., Feldblyum, T. V, Wall, J.D., Voordouw, G. & Fraser, C.M. (2004) The genome sequence of the anaerobic, sulfate-reducing bacterium *Desulfovibrio vulgaris* Hildenborough. *Nature Biotechnology*, **22**; 554–559.

Heinzle, E., Yuan, Y., Kumar, S., Wittmann, C., Gehre, M., Richnow, H.H., Wehrung, P., Adam, P. & Albrecht, P. (2008) Analysis of ^{13}C labeling enrichment in microbial culture applying metabolic tracer experiments using gas chromatography-combustion-isotope ratio mass spectrometry. *Analytical Biochemistry*, **380**; 202–210.

Hellerstein, M.K. & Neese, R.A. (1992) Mass isotopomer distribution analysis: a technique for measuring biosynthesis and turnover of polymers. *The American Journal of Physiology*, **263**; E988–E1001.

Henry, E.A., Devereux, R., Maki, J.S., Gilmour, C.C., Woese, C.R., Mandelco, L., Schauder, R., Remsen, C.C. & Mitchell, R. (1994) Characterization of a new thermophilic sulfate-reducing bacterium *Thermodesulfovibrio yellowstonii*, gen. nov. and sp. nov.: its phylogenetic relationship to *Thermodesulfobacterium commune* and their origins deep within the bacterial domain. *Archives of Microbiology*, **161**; 62–69.

Ho, D.P., Ngo, H.H. & Guo, W. (2014) A mini review on renewable sources for biofuel. *Bioresource Technology*, **169**; 742–749.

Howard, B.H. & Hungate, R.E. (1976) *Desulfovibrio* of the sheep rumen. *Applied and Environmental Microbiology*, **32**; 598–602.

Howard, T.P., Middelhaufe, S., Moore, K., Edner, C., Kolak, D., Taylor, G., Parker, D., Lee, R., Smirnoff, N., Aves, S. & Love, J. (2013) Synthesis of customized petroleum-replica fuel molecules by targeted modification of free fatty acid pools in *Escherichia coli. Proceedings of the National Academy of Sciences*, **110**; 1–6.

Hughes, A.L., Ekollu, V., Friedman, R. & Rose, J.R. (2005) Gene Family Content-Based Phylogeny of Prokaryotes: The Effect of Criteria for Inferring Homology. *Systematic Biology*, **54**; 268–276.

Hunt, M., Kikuchi, T., Sanders, M., Newbold, C., Berriman, M. & Otto, T.D. (2013) REAPR : a universal tool for genome assembly evaluation. *Genome Biology*, **14**; R47.

Hussain, A., Hasan, A., Javid, A. & Iqbal, J. (2016) Exploited application of sulfate-reducing bacteria for concomitant treatment of metallic and non-metallic wastes : a mini review. *3 Biotech*, **6**; 1–10.

Hyatt, D., Chen, G.L., LoCascio, P.F., Land, M.L., Larimer, F.W. & Hauser, L.J. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, **11**; 119.

Intergovernmental Panel on Climate Change. (2014) *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.

International Energy Agency. (2016) *International Energy Outlook 2016 with Projections to 2040*.

International Energy Agency. (2017) *Tracking Clean Energy Progress 2017*.

International Renewable Energy Agency. (2015) *Renewable Energy Policy Brief Brazil*.

International Renewable Energy Agency. (2016) *Innovation Outlook Advanced Liquid Biofuels*.

Jackson, B.E. & McInerney, M.J. (2002) Anaerobic microbial metabolism can proceed close to thermodynamic limits. *Nature*, **415**; 454–456.

Jankowski, G.J. & ZoBell, C.E. (1944) Hydrocarbon production by sulphate-reducing bacteria. *Journal of Biochemistry*, **47**; 447.

Jansen, K., Thauer, R.K., Widdel, F. & Fuchs, G. (1984) Carbon assimilation pathways in sulfate reducing bacteria. Formate, carbon dioxide, carbon monoxide, and acetate assimilation by *Desulfovibrio baarsii*. *Archives of Microbiology*, **138**; 257–262.

Janßen, H.J. & Steinbüchel, A. (2014) Fatty acid synthesis in *Escherichia coli* and its applications towards the production of fatty acid based biofuels. *Biotechnology for Biofuels*, **7**.

Johnston, S., Lin, S., Lee, P., Caffrey, S.M., Wildschut, J., Voordouw, J.K.,da Silva, S.M., Pereira, I.A.C. & Voordouw, G. (2008) A genomic island of the sulfate-reducing bacterium *Desulfovibrio vulgaris* Hildenborough promotes survival under stress conditions while decreasing the efficiency of anaerobic growth. *Environmental Microbiology*, **11**; 981–991.

Jørgensen, B.B. (1982) Mineralization of organic matter in the sea bed – the role of sulphate reduction. *Nature*, **296**; 643–645.

Kageyama, H., Waditee-Sirisattha, R., Sirisattha, S., Tanaka, Y., Mahakhant, A. & Takabe, T. (2015) Improved Alkane Production in Nitrogen-Fixing and Halotolerant Cyanobacteria *via* Abiotic Stresses and Genetic Manipulation of Alkane Synthetic Genes. *Current Microbiology*, **71**; 115–120.

Källberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H. & Xu, J. (2012) Template-based protein structure modeling using the RaptorX web server. *Nature Protocols*, **7**; 1511–1522.

Kallio, P., Pásztor, A., Akhtar, M.K. & Jones, P.R. (2014) Renewable jet fuel. *Current Opinion in Biotechnology*, **26**; 50–55.

Kang, Q., Appels, L., Baeyens, J., Dewil, R. & Tan, T. (2014) Energy-Efficient Production of Cassava-Based Bio-Ethanol. *Advances in Bioscience and Biotechnology*, **5**; 925–939.

Katoh, K., Misawa, K., Kuma, K. & Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research*, **30**; 3059–3066.

Keasling, J.D. (2008) Synthetic Biology for Synthetic Chemistry. *ACS Chemical Biology*, **3**; 64–76.

Keller, K.L., Bender, K.S. & Wall, J.D. (2009) Development of a markerless genetic exchange system for *Desulfovibrio vulgaris* Hildenborough and its use in generating a strain with increased transformation efficiency. *Applied and Environmental Microbiology*, **75**; 7682–7691.

Keller, K.L., Rapp-Giles, B.J., Semkiw, E.S., Porat, I., Brown, S.D. & Wall, J.D. (2014) New model for electron flow for sulfate reduction in *Desulfovibrio alaskensis* G20. *Applied and Environmental Microbiology*, **80**; 855–68.

Keller, K.L. & Wall, J.D. (2011) Genetics and molecular biology of the electron flow for sulfate respiration in *Desulfovibrio*. *Frontiers in Microbiology*, **2**; 135.

Kiran, M.G., Pakshirajan, K. & Das, G. (2016) Heavy metal removal from multicomponent system by sulfate reducing bacteria: mechanism and cell surface characterization. *Journal of Hazardous Materials*, **324**; 62–70.

Kircher, M. (2015) Sustainability of biofuels and renewable chemicals production from biomass. *Current Opinion in Chemical Biology*, **29**; 26–31.

Kitamura, M., Takayama, Y., Kojima, S., Kohno, K., Ogata, H., Higuchi, Y. & Inoue, H. (2004) Cloning and expression of the enolase gene from *Desulfovibrio vulgaris* (Miyazaki F). *Biochimica et Biophysica Acta*, **1676**; 172–181.

Kjeldsen, K.U., Loy, A., Jakobsen, T.F., Thomsen, T.R., Wagner, M. & Ingvorsen, K. (2007) Diversity of sulfate-reducing bacteria from an extreme hypersaline sediment, Great Salt Lake (Utah). *FEMS Microbiology Ecology*, **60**; 287–298.

Kobayashi, K., Takahashi, E. & Ishimoto, M. (1972) Biochemical studies on sulfate-reducing bacteria. XI. Purification and some properties of sulfite reductase, desulfoviridin. *Journal of Biochemistry*, **72**; 879–887.

Koonin, E. V, Makarova, K.S. & Aravind, L. (2001) Horizontal Gene Transfer in Prokaryotes: Quantfication and Classification. *Annual Review of Microbiology*, **55**; 709–742.

Korte, H.L., Fels, S.R., Christensen, G.A., Price, M.N., Kuehl, J. V., Zane, G.M., Deutschbauer, A.M., Arkin, A.P. & Wall, J.D. (2014) Genetic basis for nitrate resistance in *Desulfovibrio* strains. *Frontiers in Microbiology*, **5**; 1–12.

Koyano, S., Tatsuno, K., Okazaki, M., Ohkusu, K., Sasaki, T., Saito, R., Okugawa, S. & Moriya, K. (2015) A Case of Liver Abscess with *Desulfovibrio desulfuricans* Bacteremia. *Case Reports in Infectious Diseases*, **2015**; 1–4.

Krämer, L., Jäger, C., Trezzi, J.-P., Jacobs, D. & Hiller, K. (2018) Quantification of Stable Isotope Traces Close to Natural Enrichment in Human Plasma Metabolites Using Gas Chromatography-Mass Spectrometry. *Metabolites*, **8**.

Kristensen, D.M., Kannan, L., Coleman, M.K., Wolf, Y.I., Sorokin, A., Koonin, E.V. & Mushegian, A. (2010) A low-polynomial algorithm for assembling clusters of orthologous groups from intergenomic symmetric best matches. *Bioinformatics*, **26**; 1481–1487.

Kruegar, F. (2014) Babraham Bioinformatics - Trim Galore !

Krumholz, L.R., Bradstock, P., Sheik, C.S., Diao, Y., Gazioglu, O., Gorby, Y. & McInerney, M.J. (2015) Syntrophic Growth of *Desulfovibrio alaskensis* Requires Genes for H_2 and Formate Metabolism as Well as Those for Flagellum and Biofilm Formation. *Applied and Environmental Microbiology*, **81**; 2339–2348.

Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C. & Salzberg, S.L. (2004) Versatile and open software for comparing large genomes. *Genome Biology*, **5**; R12.

Ladygina, N., Dedyukhina, E.G. & Vainshtein, M.B. (2006) A review on microbial synthesis of hydrocarbons. *Process Biochemistry*, **41**; 1001–1014.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**; 680–685.

Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, **10**; R25.

Lanjekar, R.D. & Deshmukh, D. (2016) A review of the effect of the composition of biodiesel on NOx emission, oxidative stability and cold flow properties. *Renewable and Sustainable Energy Reviews*, **54**; 1401–1411.

Lau, S.K.P., Fan, R.Y.Y., Ho, T.C.C., Wong, G.K.M., Tsang, A.K.L., Teng, J.L.L., Chen, W., Watt, R.M., Curreem, S.O.T., Tse, H., Yuen, K. & Woo, P.C.Y. (2011) Environmental adaptability and stress tolerance of *Laribacter hongkongensis* : a genome-wide analysis. *Cell & Bioscience*, **1**; 22.

Laue, H., Friedrich, M., Ruff, J. & Alasdair, M. (2001) Dissimilatory Sulfite Reductase (Desulfoviridin) of the Taurine-Degrading , Non-Sulfate-Reducing Bacterium *Bilophila wadsworthia* RZATAU Contains a Fused DsrB-DsrD Subunit. *Journal of Bacteriology*, **183**; 1727–1733.

Lea-Smith, D.J., Biller, S.J., Davey, M.P., Cotton, C.A.R., Perez Sepulveda, B.M., Turchyn, A. V., Scanlan, D.J., Smith, A.G., Chisholm, S.W. & Howe, C.J. (2015) Contribution of cyanobacterial alkane production to the ocean hydrocarbon cycle. *Proceedings of the National Academy of Sciences*, **112**; 13591–13596.

Lee, J.P. & Peck, H.D. (1971) Purification of the enzyme reducing bisulfite to trithionate from *Desulfovibrio gigas* and its identification as desulfoviridin. *Biochemical and Biophysical Research Communications*, **45**; 583–589.

Lee, J.W., Na, D., Park, J.M., Lee, J., Choi, S. & Lee, S.Y. (2012) Systems metabolic engineering of microorganisms for natural and non-natural chemicals. *Nature Chemical Biology*, **8**; 536–546.

Lee, S.K., Chou, H., Ham, T.S., Lee, T.S. & Keasling, J.D. (2008) Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. *Current Opinion in Biotechnology*, **19**; 556–563.

Lee, S.Y., Kim, H.M. & Seungwoo, C. (2015) Metabolic engineering for the production of hydrocarbon fuels. *Current Opinion in Biotechnology*, **33**; 15–22.

Lehmann, W.D. (2017) A Timeline of Stable Isotopes and Mass Spectrometry in the Life Sciences. *Mass Spectrometry Reviews*, **36**; 58–85.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. & Durbin, R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**; 2078–2079.

Li, N., Chang, W.C., Warui, D.M., Booker, S.J., Krebs, C. & Bollinger, J.M. (2012) Evidence for only oxygenative cleavage of aldehydes to alk(a/e)nes and formate by cyanobacterial aldehyde decarbonylases. *Biochemistry*, **51**; 7908–7916.

Liderot, K., Larsson, M., Boräng, S. & Özenci, V. (2010) Polymicrobial bloodstream infection with *Eggerthella lenta* and *Desulfovibrio desulfuricans*. *Journal of Clinical Microbiology*, **48**; 3810–3812.

Lie, T.J., Pitta, T., Leadbetter, E.R., Godchaux, W. & Leadbetter, J.R. (1996) Sulfonates: Novel electron acceptors in anaerobic respiration. *Archives of Microbiology*, **166**; 204–210.

Lim, E., Mbowe, O., Lee, A.S.W. & Davis, J. (2016) Effect of environmental exposure to hydrogen sulfide on central nervous system and respiratory function: a systematic review of human studies. *International Journal of Occupational and Environmental Health*, **22**; 80–90.

Liu, C.J., Jiang, H., Wu, L., Zhu, L.Y., Meng, E. & Zhang, D.Y. (2017) OEPR cloning: An efficient and seamless cloning strategy for large- and multi-fragments. *Scientific Reports*, **7**; 2–8.

Lloyd, J.R., Ridley, J., Khizniak, T., Lyalikova, N.N. & Macaskie, L.E. (1999) Reduction of technetium by *Desulfovibrio desulfuricans*: Biocatalyst characterization and use in a flowthrough bioreactor. *Applied and Environmental Microbiology*, **65**; 2691–2696.

Lobo, S.A.L., Melo, A.M.P., Carita, J.N., Teixeira, M. & Saraiva, L.M. (2007) The anaerobe *Desulfovibrio desulfuricans* ATCC 27774 grows at nearly atmospheric oxygen levels. *FEBS Letters*, **581**; 433–436.

Lodowska, J., Wolny, D., Jaworska-Kik, M., Kurkiewicz, S., Dzierżewicz, Z. & Węglarz, L. (2012) The chemical composition of endotoxin isolated from intestinal strain of *Desulfovibrio desulfuricans*. *The Scientific World Journal*, **2012**; ID 647352.

Lojou, E., Durand, M.C., Dolla, A. & Bianco, P. (2002) Hydrogenase Activity Control at *Desulfovibrio vulgaris* Cell-Coated Carbon Electrodes: Biochemical and Chemical Factors Influencing the Mediated Bioelectrocatalysis. *Electroanalysis*, **14**; 914–922.

Lojou, E. (2011) Hydrogenases as catalysts for fuel cells: Strategies for efficient immobilization at electrode interfaces. *Electrochimica Acta*, **56**; 10385–10397.

Lolkema, J.S., Chaban, Y. & Boekema, E.J. (2003) Subunit Composition, Structure, and Distribution of Bacterial V-Type ATPases. *Journal of Bioenergetics and Biomembranes*, **35**; 323–335.

Loubinoux, J., Bronowicki, J.P., Pereira, I.A.C., Mougenel, J.L. & Faou, A.E. (2002) Sulfate-reducing bacteria in human feces and their association with inflammatory bowel diseases. *FEMS Microbiology Ecology*, **40**; 107–112.

Lovley, D.R. & Phillips, E.J.P. (1992) Reduction of Uranium By *Desulfovibrio desulfuricans*. *Applied and Environmental Microbiology*, **58**; 850–856.

Lovley, D.R. & Phillips, E.J.P. (1994) Reduction of chromate by *Desulfovibrio vulgaris* and its c3 cytochrome. *Applied and Environmental Microbiology*, **60**; 726–728.

Lozniewski, A., Maurer, P., Schuhmacher, H., Carlier, J.P. & Mory, F. (1999) First isolation of *Desulfovibrio* species as part of a polymicrobial infection froma brain abscess. *European Journal of Clinical Microbiology and Infectious Diseases*, **18**; 602–603.

Lupton, F.S., Conrad, R. & Zeikus, J.G. (1984) Physiological function of hydrogen metabolism during growth of sulfidogenic bacteria on organic substrates. *Journal of Bacteriology*, **159**; 843–849.

Ma, J., Wang, S., Zhao, F. & Xu, J. (2013) Protein threading using context-specific alignment potential. *Bioinformatics*, **29**.

Macpherson, R. & Miller, J.D.A. (1963) Nutritional Studies on *Desulfovibrio desulfuricans* using Chemically Defined Media. *Journal of General Microbiology*, **31**; 365–373.

Macy, J.M., Santini, J.M., Pauling, B. V, O'Neill, a H. & Sly, L.I. (2000) Two new arsenate/sulfate-reducing bacteria: mechanisms of arsenate reduction. *Archives of Microbiology*, **173**; 49–57.

Madoui, M., Engelen, S., Cruaud, C., Belser, C., Bertrand, L., Alberti, A., Lemainque, A., Wincker, P. & Aury, J. (2015) Genome assembly using Nanopore-guided long and error-free DNA reads. *BMC Genomics*, **16**; 1–11.

Mago, M., Basso, O., Tardy-Jacquenod, C. & Caumette, P. (2004) *Desulfovibrio bastinii* sp. nov. and *Desulfovibrio gracilis* sp. nov., moderately halophilic, sulfate-reducing bacteria isolated from deep subsurface oilfield water. *International Journal of Systematic and Evolutionary Microbiology*, **54**; 1693–1697.

Magoč, T. & Salzberg, S.L. (2011) FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, **27**; 2957–2963.

Makula, R.A. & Finnerty, W.R. (1974) Phospholipid composition of *Desulfovibrio* species. *Journal of Bacteriology*, **120**; 1279–1283.

Martins, M., Mourato, C., Morais-Silva, F.O., Rodrigues-Pousada, C., Voordouw, G., Wall, J.D. & Pereira, I.A.C. (2016) Electron transfer pathways of formate-driven H2 production in *Desulfovibrio*. *Applied Microbiology and Biotechnology*, **100**; 8135–8146.

Maslov, S., Krishna, S., Pang, T.Y. & Sneppen, K. (2009) Toolbox model of evolution of prokaryotic metabolic networks and their regulation. *Proceedings of the National Academy of Sciences*, **106**; 9743–9748.

Mata, T.M., Martins, A.A. & Caetano, N.S. (2010) Microalgae for biodiesel production and other applications: A review. *Renewable and Sustainable Energy Reviews*, **14**; 217–232.

Matthews, D.H., Gillett, N.P., Stott, P.A. & Zickfeld, K. (2009) The proportionality of global warming to cumulative carbon emissions. *Nature*, **459**; 829–832.

McInerney, M.J., Struchtemeyer, C.G., Sieber, J., Mouttaki, H., Stams, A.J.M., Schink, B., Rohlin, L. & Gunsalus, R.P. (2008) Physiology, ecology, phylogeny, and genomics of microorganisms capable of syntrophic metabolism. *Annals of the New York Academy of Sciences*, **1125**; 58–72.

Medipally, S.R., Yusoff, F.M., Banerjee, S. & Shariff, M. (2014) Microalgae as sustainable renewable energy feedstock for biofuel production. *BioMed Research International*, **2015**.

Mendez-Perez, D., Begemann, M.B. & Pfleger, B.F. (2011) Modular Synthase-Encoding Gene Involved in α-Olefin Biosynthesis in *Synechococcus* sp. Strain PCC 7002. *Applied and Environmental Microbiology*, **77**; 4264–4267.

Mendez-Perez, D., Herman, N.A. & Pfleger, B.F. (2014) A Desaturase Gene Involved in the Formation of 1,14-Nonadecadiene in *Synechococcus* sp. Strain PCC 7002. *Applied and Environmental Microbiology*, **80**; 6073–6079.

Metzger, P. & Largeau, C. (2005) *Botryococcus braunii*: A rich source for hydrocarbons and related ether lipids. *Applied Microbiology and Biotechnology*, **66**; 486–496.

Meyer, B., Kuehl, J., Deutschbauer, A.M., Price, M.N., Arkin, A.P. & Stahl, D.A. (2013) Variation among *Desulfovibrio* species in electron transfer systems used for syntrophic growth. *Journal of Bacteriology*, **195**; 990–1004.

Millar, A.A., Clemens, S., Zachgo, S., Giblin, E.M., Taylor, D.C. & Kunst, L. (1999) *CUT1*, an *Arabidopsis* gene required for cuticular wax biosynthesis and pollen fertility, encodes a very-long-chain fatty acid condensing enzyme. *The Plant cell*, **11**; 2115–2127.

Miller, J.D. & Wakerley, D.S. (1966) Growth of sulphate-reducing bacteria by fumarate dismutation. *Journal of General Microbiology*, **43**; 101–107.

Miran, W., Jang, J., Nawaz, M., Shahzad, A., Jeong, S.E., Jeon, C.O. & Lee, D.S. (2017) Mixed sulfate-reducing bacteria-enriched microbial fuel cells for the treatment of wastewater containing copper. *Chemosphere*, **189**; 134–142.

Mogensen, G.L., Kjeldsen, K.U. & Ingvorsen, K. (2005) *Desulfovibrio aerotolerans* sp. nov., an oxygen tolerant sulphate-reducing bacterium isolated from activated sludge. *Anaerobe*, **11**; 339–349.

Mohr, A. & Raman, S. (2013) Lessons from first generation biofuels and implications for the sustainability appraisal of second generation biofuels. *Energy Policy*, **63**; 114–122.

Morais-Silva, F.O., Rezende, A.M., Pimentel, C., Santos, C.I., Clemente, C., Varela-Raposo, A., Resende, D.M., da Silva, S.M., de Oliveira, L.M., Matos, M., Costa, D.A., Flores, O., Ruiz, J.C. & Rodrigues-Pousada, C. (2014) Genome sequence of the model sulfate reducer *Desulfovibrio gigas*: A comparative analysis within the *Desulfovibrio* genus. *MicrobiologyOpen*, **3**; 513–530.

Morais-Silva, F.O., Santos, C.I., Rodrigues, R., Pereira, I.A.C. & Rodrigues-Pousada, C. (2013) Roles of HynAB and Ech, the only two hydrogenases found in the model sulfate reducer *Desulfovibrio gigas*. *Journal of Bacteriology*, **195**; 4753–4760.

Morris, B.E.L., Henneberger, R., Huber, H. & Moissl-Eichinger, C. (2013) Microbial syntrophy: Interaction for the common good. *FEMS Microbiology Reviews*, **37**; 384–406.

Mosquera-Rendón, J., Rada-Bravo, A.M., Cárdenas-Brito, S., Corredor, M., Restrepo-Pineda, E. & Benítez-Páez, A. (2016) Pangenome-wide and molecular evolution analyses of the *Pseudomonas aeruginosa* species. *BMC Genomics*, **17**; 1–14.

Moura, I., Bursakov, S., Costa, C. & Moura, J.J. (1997) Nitrate and nitrite utilization in sulfate-reducing bacteria. *Anaerobe*, **3**; 279–290.

Mukhopadhyay, A., He, Z., Alm, E.J., Arkin, A.P., Baidoo, E.E., Borglin, S.C., Chen, W., Hazen, T.C., He, Q., Holman, H.Y., Huang, K., Huang, R., Joyner, D.C., Katz, N., Keller, M., Oeller, P., Redding, A., Sun, J., Wall, J., Wei, J., Yang, Z., Yen, H.C., Zhou, J. & Keasling, J.D. (2006) Salt stress in *Desulfovibrio vulgaris* hildenborough: An integrated genomics approach. *Journal of Bacteriology*, **188**; 4068–4078.

Murzin, A.G., Brenner, S.E., Hubbard, T. & Chothia, C. (1995) SCOP: A structural classification of proteins database for the investigation of sequences and structures. *Journal of Molecular Biology*, **247**; 536–540.

Muyzer, G. & Stams, A.J.M. (2008) The ecology and biotechnology of sulphate-reducing bacteria. *Nature Reviews Microbiology*, **6**; 441–454.

Naik, S.N., Goud, V. V., Rout, P.K. & Dalai, A.K. (2010) Production of first and second generation biofuels: A comprehensive review. *Renewable and Sustainable Energy Reviews*, **14**; 578–597.

Nakazawa, H., Arakaki, A., Narita-yamada, S., Yashiro, I., Jinno, K., Aoki, N., Tsuruyama, A., Okamura, Y., Tanikawa, S., Fujita, N., Takeyama, H. & Matsunaga, T. (2009) Whole genome sequence of *Desulfovibrio magneticus* strain RS-1 revealed common gene clusters in magnetotactic bacteria. *Genome Research*, **19**; 1801–1808.

Nanninga, H.J. & Gottschal, J.C. (1987) Properties of *Desulfovibrio carbinolicus* sp. nov. and Other Sulfate-Reducing Bacteria Isolated from an Anaerobic-Purification Plant, *Applied and Environmental Microbiology,* **53**; 802–809.

Ndaba, B., Chiyanzu, I. & Marx, S. (2015) n-Butanol derived from biochemical and chemical routes: A review. *Biotechnology Reports*, **8**; 1–9.

Noguera, D.R., Brusseau, G.A., Rittmann, B.E. & Stahl, D.A. (1998) A unified model describing the role of hydrogen in the growth of *Desulfovibrio vulgaris* under different environmental conditions. *Biotechnology and Bioengineering*, **59**; 732–46.

Novelli, G.D. & ZoBell, C.E. (1944) Assimilation of petroleum hydrocarbons by sulfate-reducing bacteria. *Journal of Bacteriology*, **47**; 447-448

Odom, J.M. & Peck Jr, H.D. (1981) Hydrogen cycling as a general mechanism for energy coupling in the sulfate-reducing bacteria, *Desulfovibrio* sp. *FEMS Microbiology Letters*, **12**; 47–50.

Okonechnikov, K., Conesa, A. & García-Alcalde, F. (2015) Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*, **32**; 292–294.

Ollivier, B., Cord-Ruwisch, R., Hatchikian, E.C. & Garcia, J.L. (1988) Characterization of *Desulfovibrio fructosovorans* sp. nov. *Archives of Microbiology*, **149**; 447–450.

Oppenheimer, C.H. (1965) Bacterial production of hydrocarbon-like materials. *Zeitschrift fur Allgemeine Mikrobiologie,* **5**; 284–307.

Ozawa, K., Tsapin, A.I., Nealson, K.H., Cusanovich, M.A. & Akutsu, H. (2000) Expression of a tetraheme protein, *Desulfovibrio vulgaris* Miyazaki F cytochrome c_3, in *Shewanella oneidensis* MR-1. *Applied and Environmental Microbiology*, **66**; 4168–4171.

Ozuolmez, D., Na, H., Lever, M.A., Kjeldsen, K.U., Jørgensen, B.B. & Plugge, C.M. (2015) Methanogenic archaea and sulfate reducing bacteria co-cultured on acetate: Teamwork or coexistence? *Frontiers in Microbiology*, **6**; 1–12.

Pankhania, I.P., Spormann, A.M., Hamilton, W.A. & Thauer, R.K. (1988) Lactate conversion to acetate, CO_2 and H_2 in cell suspensions of *Desulfovibrio vulgaris* (Marburg): indications for the involvement of an energy driven reaction. *Archives of Microbiology*, **150**; 26–31.

Park, H.S., Lin, S. & Voordouw, G. (2008) Ferric iron reduction by *Desulfovibrio vulgaris* Hildenborough wild type and energy metabolism mutants. *Antonie van Leeuwenhoek*, **93**; 79–85.

Park, J., Pham, H. V., Mogensen, K., Solling, T.I., Bennetzen, M.V. & Houk, K.N. (2015) Hydrocarbon Binding by Proteins: Structures of Protein Binding Sites for $\geq C_{10}$ Linear Alkanes or Long-Chain Alkyl and Alkenyl Groups. *Journal of Organic Chemistry*, **80**; 997–1005.

Park, M.O. (2005) New Pathway for Long-Chain n-Alkane Synthesis via 1-Alcohol in *Vibrio furnissii* M1. *Journal of Bacteriology*, **187**; 12–16.

Park, M.O., Tanabe, M., Hirata, K. & Miyamoto, K. (2001) Isolation and characterization of a bacterium that produces hydrocarbons extracellularly which are equivalent to light oil. *Applied Microbiology and Biotechnology*, **56**; 448–452.

Parks, J.M., Johs, A., Podar, M., Bridou, R., Hurt, R.A., Smith, S.D., Tomanicek, S.J., Qian, Y., Brown, S.D., Brandt, C.C., Palumbo, A. V, Smith, J.C., Wall, J.D., Elias, D.A. & Liang, L. (2013) The genetic basis for bacterial mercury methylation. *Science*, **339**; 1332–1335.

Patterson, B.W., Zhang, X.J., Chen, Y., Klein, S. & Wolfe, R.R. (1997) Measurement of very low stable isotope enrichments by gas chromatography/mass spectrometry: application to measurement of muscle protein synthesis. *Metabolism*, **46**; 943–948.

Paulo, L.M., Stams, A.J.M. & Sousa, D.Z. (2015) Methanogens, sulphate and heavy metals: a complex system. *Reviews in Environmental Science and Biotechnology*, **14**; 537–553.

Payne, R.B., Hemme, C.L. & Wall, J.D. (2004) A new frontier in genomic research. *World Pipelines*, **4**; 53–55.

Pearson, W.R. (2013) An Introduction to Sequence Similarity (“ Homology ”) Searching. *Current Protocols in Bioinformatics*, **42**.

Peck, H.D. (1960) Evidence for oxidative phosphorylation during the reduction of sulfate with hydrogen by *Desulfovibrio desulfuricans. The Journal of Biological Chemistry*, **235**; 2734–2738.

Peck, H.D., LeGall, J., Lespinat, P.A., Berlier, Y. & Fauque, G. (1987) A direct demonstration of hydrogen cycling by *Desulfovibrio vulgaris* employing membrane-inlet mass spectrometry. *FEMS Microbiology Letters*, **40**; 295–299.

Peck Jr, H.D. (1961) Enzymatic basis for assimilatory and dissimilatory sulfate reduction. *Journal of Bacteriology*, **82**; 933–939.

Peng, J. & Xu, J. (2011) RaptorX: Exploiting structure information for protein alignment by statistical inference. *Proteins: Structure, Function and Bioinformatics*, **79**; 161–171.

Peralta-Yahya, P.P., Zhang, F., del Cardayre, S.B. & Keasling, J.D. (2012) Microbial engineering for the production of advanced biofuels. *Nature*, **488**; 320–328.

Peramuna, A., Morton, R. & Summers, M. (2015) Enhancing Alkane Production in Cyanobacterial Lipid Droplets: A Model Platform for Industrially Relevant Compound Production. *Life*, **5**; 1111–1126.

Pereira, I.A.C., Ramos, A.R., Grein, F., Marques, M.C., da Silva, S.M. & Venceslau, S.S. (2011) A comparative genomic analysis of energy metabolism in sulfate reducing bacteria and archaea. *Frontiers in Microbiology*, **2**; 1–22.

Pereira, P.M., He, Q., Valente, F.M.A., Xavier, A. V., Zhou, J., Pereira, I.A.C. & Louro, R.O. (2008) Energy metabolism in *Desulfovibrio vulgaris* Hildenborough: Insights from transcriptome analysis. *Antonie van Leeuwenhoek*, **93**; 347–362.

Phelps, T.J., Conrad, R. & Zeikus, J.G. (1985) Sulfate-dependent interspecies H_2 transfer between *Methnosarcina barkeri* and *Desulfovibrio vulgaris* during coculture metabolism of acetate or methanol. *Applied and Environmental Microbiology*, **50**; 589–594.

Plugge, C.M., Scholten, J.C.M., Culley, D.E., Nie, L., Brockman, F.J. & Zhang, W. (2010) Global transcriptomics analysis of the *Desulfovibrio vulgaris* change from syntrophic growth with *Methanosarcina barkeri* to sulfidogenic metabolism. *Microbiology*, **156**; 2746–2756.

Plugge, C.M., Zhang, W., Scholten, J.C.M. & Stams, A.J.M. (2011) Metabolic flexibility of sulfate-reducing bacteria. *Frontiers in Microbiology*, **2**; 1–8.

Pol, L.W.H., Lens, P.N.L., Stams, A.J.M. & Lettinga, G. (1998) Anaerobic treatment of sulphate-rich wastewaters. *Biodegradation*, **9**; 213–224.

Pollock, W.B.R. & Voordouw, G. (1994) Aerobic expression of the *cyf* gene encoding cytochrome c-553 from *Desulfovibrio vulgaris* Hildenborough in *Escherichia coli. Microbiology*, **140**; 879–887.

Polz, M.F., Alm, E.J. & Hanage, W.P. (2013) Horizontal gene transfer and the evolution of bacterial and archaeal population structure. *Trends in Genetics*, **29**; 170–175.

Postgate, J.R. (1984) The Sulphate-reducing Bacteria. *Cambridge University Press.*

Powell, B., Mergeay, M. & Christofi, N. (1989) Transfer of broad host-range plasmids to sulfate-reducing bacteria. *FEMS Microbiology Letters*, **59**; 269–274.

Prescott, G.W., Edwards, D.P. & Foster, W. A. (2015) Retaining biodiversity in intensive farmland: epiphyte removal in oil palm plantations does not affect yield. *Ecology and Evolution*, **5**; 1944–1954.

Price, M.N., Dehal, P.S. & Arkin, A.P. (2010) FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS ONE*, **5**.

Qiu, Y., Tittiger, C., Wicker-Thomas, C., Goff, G. Le, Young, S., Wajnberg, E., Fricaux, T., Taquet, N., Blomquist, G.J. & Feyereisen, R. (2012) An insect-specific P450 oxidative decarbonylase for cuticular hydrocarbon biosynthesis. *Proceedings of the National Academy of Sciences*, **109**; 14858–14863.

Rabinovitch-Deere, C.A., Oliver, J.W.K., Rodriguez, G.M. & Atsumi, S. (2013) Synthetic biology and metabolic engineering approaches to produce biofuels. *Chemical Reviews*, **113**; 4611–32.

Rapp-Giles, B.J., Casalot, L., English, R.S., Ringbauer, J.A., Dolla, A. & Wall, J.D. (2000) Cytochrome c_3 mutants of *Desulfovibrio desulfuricans*. *Applied and Environmental Microbiology*, **66**; 671–677.

Rapp, B.J. & Wall, J.D. (1987) Genetic transfer in *Desulfovibrio desulfuricans*. *Proceedings of the National Academy of Sciences of the United States of America*, **84**; 9128–9130.

Ravenhall, M., Škunca, N., Lassalle, F. & Dessimoz, C. (2015) Inferring Horizontal Gene Transfer. *PLoS Computational Biology*, **11**.

Reed, J.R., Vanderwel, D., Choi, S., Pomonis, J.G., Reitz, R.C. & Blomquist, G.J. (1994) Unusual mechanism of hydrocarbon formation in the housefly: cytochrome P450 converts aldehyde to the sex pheromone component (*Z*)-9-tricosene and CO_2. *Proceedings of the National Academy of Sciences of the United States of America*, **91**; 10000–10004.

Ricker, N., Qian, H. & Fulthorpe, R.R. (2012) The limitations of draft assemblies for understanding prokaryotic adaptation and evolution. *Genomics*, **100**; 167–175.

Rooks, M.G., Veiga, P., Wardwell-scott, L.H., Tickle, T., Segata, N., Michaud, M., Gallini, C.A., Beal, C., Van Hylckama-Vlieg, J.E.T., Ballal, S.A., Morgan, X.C., Glickman, J.N., Gevers, D., Huttenhower, C. & Garrett, W.S. (2014) Gut microbiome composition and function in experimental colitis during active disease and treatment-induced remission. *The ISME Journal*, **8**; 1403–1417.

Rosano, G.L. & Ceccarelli, E.A. (2014) Recombinant protein expression in *Escherichia coli*: Advances and challenges. *Frontiers in Microbiology*, **5**; 1–17.

Rosenfeld, W.D. (1947) Anaerobic Oxidation of Hydrocarbons By Sulfate-Reducing Bacteria. *Journal of Bacteriology*, **54**; 664–665.

Rosenzweig, C., Iglesias, A., Yang, X.B., Epstein, P.R. & Chivian, E. (2001) Climate change and extreme weather events. *Global Change & Human Health*, **2**; 90–104.

Rouli, L., Merhej, V., Fournier, P. & Raoult, D. (2015) The bacterial pangenome as a new tool for analysing pathogenic bacteria. *New Microbes and New Infections*, **7**; 72–85.

Rousset, M., Dermoun, Z., Chippaux, M. & Bélaich, J.P. (1991) Marker exchange mutagenesis of the hydN genes in *Desulfovibrio fructosovorans*. *Molecular Microbiology*, **5**; 1735–1740.

Rousset, M., Magro, V., Forget, N., Guigliarelli, B., Belaich, J.P. & Hatchikian, E.C. (1998) Heterologous expression of the *Desulfovibrio gigas* [NiFe] hydrogenase in *Desulfovibrio fructosovorans* MR400. *Journal of Bacteriology*, **180**; 4982–4986.

Rozanova, E.P. & Khudiakova, A.I. (1974) A new non-spore forming thermophilic organism, reducing sulfates, *Desulfovibrio thermophilus* nov. sp. *Mikrobiologiia*, **43**; 1069–1075.

Rude, M.A., Baron, T.S., Brubaker, S., Alibhai, M., del Cardayre, S.B. & Schirmer, A. (2011) Terminal Olefin (1-Alkene) Biosynthesis by a Novel P450 Fatty Acid Decarboxylase from *Jeotgalicoccus* Species. *Applied and Environmental Microbiology*, **77**; 1718–1727.

Rui, Z., Li, X., Zhu, X., Liu, J., Domigan, B., Barr, I., Cate, J.H.D. & Zhang, W. (2014) Microbial biosynthesis of medium-chain 1-alkenes by a nonheme iron oxidase. *Proceedings of the National Academy of Sciences*, **111**; 18237–18242.

Rui, Z., Harris, N.C., Zhu, X., Huang, W. & Zhang, W. (2015) Discovery of a Family of Desaturase-Like Enzymes for 1-Alkene Biosynthesis. *ACS Catalysis*, **5**; 7091–7094.

Ruiz, B., Chávez, A., Forero, A., García-Huante, Y., Romero, A., Sánchez, M., Rocha, D., Sánchez, B., Rodríguez-Sanoja, R., Sánchez, S. & Langley, E. (2010) Production of microbial secondary metabolites: Regulation by the carbon source. *Critical Reviews in Microbiology*, **36**; 146–167.

Saijo, S., Arai, S., Hossain, K.M.M., Yamato, I., Suzuki, K., Kakinuma, Y., Ishizuka-Katsura, Y., Ohsawa, N., Terada, T., Shirouzu, M., Yokoyama, S., Iwata, S. & Murata, T. (2011) Crystal structure of the central axis DF complex of the prokaryotic V-ATPase. *Proceedings of the National Academy of Sciences*, **108**; 19955–19960.

Sanger, F. & Coulson, A.R. (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology*, **94**; 441–448.

Sasi Jyothsna, T.S., Sasikala, C. & Ramana, C. V. (2008) *Desulfovibrio psychrotolerans* sp. nov., a psychrotolerant and moderately alkaliphilic sulfate-reducing deltaproteobacterium from the Himalayas. *International Journal of Systematic and Evolutionary Microbiology*, **58**; 821–825.

Sass, H., Cypionka, H. & Babenzien, H.D. (1997) Vertical distribution of sulfate-reducing bacteria at the oxic-anoxic interface in sediments of the oligotrophic Lake Stechlin. *FEMS Microbiology Ecology*, **22**; 245–255.

Sawyer, R.F. (1993) Trends in auto emissions and gasoline composition. *Environmental Health Perspectives*, **101**; 5–12.

Scholten, J.C., Culley, D.E., Brockman, F.J., Wu, G. & Zhang, W. (2007) Evolution of the syntrophic interaction between *Desulfovibrio vulgaris* and *Methanosarcina barkeri*: Involvement of an ancient horizontal gene transfer. *Biochemical and Biophysical Research Communications*, **352**; 48–54.

Schrimer, A., Rude, M.A., Li, X., Popova, E. & del Cardayre, S.B. (2010) Microbial Biosynthesis of Alkanes. *Science*, **329**; 559–562.

Seemann, T. (2014) Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*, **30**; 2068–2069.

Seitz, H.J. & Cypionka, H. (1986) Chemolithotrophic Growth of *Desulfovibrio desulfuricans* with Hydrogen coupled to Ammonification of Nitrate or Nitrite. *Archives of Microbiology*, **146**; 63–67.

Shatalin, K., Shatalina, E., Mironov, A. & Nudler, E. (2011) H_2S: A Universal Defense Against Antibiotics in Bacteria. *Science*, **334**; 986–990.

Shaw, K.J., Rather, P.N., Hare, R.S. & Miller, G.H. (1993) Molecular genetics of aminoglycoside resistance genes and familial relationships of the aminoglycoside-modifying enzymes. *Microbiological Reviews*, **57**; 138–163.

Shivani, Y., Subhash, Y., Sasikala, C. & Ramana, C.V. (2017) *Halodesulfovibrio spirochaetisodalis* gen. nov. sp. nov. and reclassification of four *Desulfovibrio* spp. *International Journal of Systematic and evolutionary microbiology*, **67**; 87–93.

Sieber, J.R., Le, H.M. & McInerney, M.J. (2014) The importance of hydrogen and formate transfer for syntrophic fatty, aromatic and alicyclic metabolism. *Environmental Microbiology*, **16**; 177–188.

Sigalevich, P. & Cohen, Y. (2000) Oxygen-dependent growth of the sulfate-reducing bacterium *Desulfovibrio oxyclinae* in coculture with *Marinobacter* sp. strain MB in an aerated sulfate-depleted chemostat. *Applied and Environmental Microbiology*, **66**; 5019–5023.

Silby, M.W., Winstanley, C., Godfrey, S.A.C., Levy, S.B. & Jackson, R.W. (2011) *Pseudomonas* genomes : diverse and adaptable. *FEMS Microbiology Reviews*, **35**; 652–680.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E. V & Zdobnov, E.M. (2015) BUSCO : assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, **31**; 3210–3212.

Sinha, M., Weyda, I., Sørensen, A., Bruno, K.S. & Ahring, B.K. (2017) Alkane biosynthesis by *Aspergillus carbonarius* ITEM 5010 through heterologous expression of *Synechococcus elongatus* acyl-ACP/CoA reductase and aldehyde deformylating oxygenase genes. *AMB Express*, **7**.

Snel, B., Bork, P. & Huynen, A.H. (1999) Genome phylogeny based on gene content. *Nature Genetics*, **21**; 108–110.

Sonnhammer, E.L., Eddy, S.R. & Durbin, R. (1997) Pfam: A comprehensive database of protein domain families based on seed alignments. *Proteins*, **28**; 405–420.

Sorigué, D., Légeret, B., Cuiné, S., Morales, P., Mirabella, B., Guédeney, G., Li-Beisson, Y., Jetter, R., Peltier, G. & Beisson, F. (2016) Microalgae synthesize hydrocarbons from long-chain fatty acids via a light-dependent pathway. *Plant Physiology*, **171**; 2393–2405.

Sorigué, D., Légeret, B., Cuiné, S., Blangy, S., Moulin, S., Billon, E., Richaud, P., Brugière, S., Couté, Y., Nurizzo, D., Müller, P., Brettel, K., Pignol, D., Arnoux, P., Li-beisson, Y., Peltier, G. & Beisson, F. (2017) An algal photoenzyme converts fatty acids to hydrocarbons. *Science*, **357**; 903–907.

Souza, S.P., Seabra, J.E.A. & Nogueira, L.A.H. (2017) Feedstocks for biodiesel production: Brazilian and global perspectives. *Biofuels*.

Studholme, D.J. (2016) Genome Update. Let the consumer beware : *Streptomyces* genome sequence quality. *Microbial Biotechnology*, **9**; 3–7.

Sukovich, D.J., Seffernick, J.L., Richman, J.E., Gralnick, J.A. & Wackett, L.P. (2010) Widespread head-to-head hydrocarbon biosynthesis in bacteria and role of OleA. *Applied and Environmental Microbiology*, **76**; 3850–3862.

Takase, K., Kakinuma, S., Yamato, I., Konishi, K., Igarashi, K. & Kakinuma, Y. (1994) Sequencing and characterization of the *ntp* gene cluster for vacuolar-type Na^+-translocating ATPase of *Enterococcus hirae*. *Journal of Biological Chemistry*, **269**; 11037–11044.

Tan, J., Soriano, A., Lui, S.M. & Cowan, J.A. (1994) Functional Expression and Characterization of the Assimilatory-Type Sulfite Reductase from *Desulfovibrio vulgaris* (Hildenborough). *Archives of biochemistry and biophysics*, **312**; 516–523.

Tan, S., Kern, R.C. & Selleck, W. (2005) The pST44 polycistronic expression system for producing protein complexes in *Escherichia coli*. *Protein Expression and Purification*, **40**; 385–395.

Tatsumi, H., Takagi, K., Fujita, M., Kano, K. & Ikeda, T. (1999) Electrochemical study of reversible hydrogenase reaction of *Desulfovibrio vulgaris* cells with methyl viologen as an electron carrier. *Analytical Chemistry*, **71**; 1753–1759.

Tatusov, R.L., Galperin, M.Y., Natale, D.A. & Koonin, E.V. (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research*, **28**; 33–36.

Taylor, J. & Parkes, R.J. (1983) The Cellular Fatty Acids of the Sulphate-reducing Bacteria, *Desulfobacter* sp., *Desulfobulbus* sp. and *Desulfovibrio desulfuricans*. *Journal of General Microbiology*, **129**; 3303–3309.

Tee, W., Dyall-Smith, M., Woods, W. & Eisen, D. (1996) Probable new species of *Desulfovibrio* isolated from a pyogenic liver abscess. *Journal of Clinical Microbiology*, **34**; 1760–1764.

Tettelin, H., Riley, D., Cattuto, C. & Medini, D. (2008) Comparative genomics: the bacterial pan-genome. *Current Opinion in Microbiology*, **11**; 472–477.

Thauer, R.K., Möller-Zinkhan, D. & Spormann, A.M. (1989) Biochemistry of acetate catabolism in anaerobic chemotrophic bacteria. *Annual Review of Microbiology*, **43**; 43–67.

Treangen, T.J. & Salzberg, S.L. (2013) Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nature Reviews Genetics*, **13**; 36–46.

Tsuji, K. & Yagi, T. (1980) Significance of Hydrogen Burst from Growing Cultures of *Desulfovibrio vulgaris*, Miyazaki, and the role of Hydrogenase and Cytochrome C_3 in Energy Production System. *Archives of Microbiology*, **125**; 35–42.

Tsujimura, S., Fujita, M., Tatsumi, H., Kano, K. & Ikeda, T. (2001) Bioelectrocatalysis-based dihydrogen/dioxygen fuel cell operating at physiological pH. *Physical Chemistry Chemical Physics*, **3**; 1331–1335.

Tucker, M.D., Barton, L.L. & Thomson, B.M. (1998) Reduction of Cr, Mo, Se and U by *Desulfovibrio desulfuricans* immobilized in polyacrylamide gels. *Journal of Industrial Microbiology & Biotechnology*, **20**; 13–19.

Ueki, A. & Suto, T. (1979) Cellular fatty acid composition of sulfate-reducing bacteria. *Journal of General and Applied Microbiology*, **25**; 185–196.

Urey, H.C. (1948) Oxygen Isotopes in Nature and in the Laboratory. *Science*, **108**; 489–496.

U.S. Energy Information Administration. (2018) *What drives crude oil prices?*

USDA Foreign Agricultural Service. (2016) *GAIN Report: Brazil Biofuels annual report*.

USDA Foreign Agricultural Service. (2017) *GAIN Report: EU Biofuels Annual Report*.

Vainshtein, M., Hippe, H. & Kroppenstedt, R.M. (1992) Cellular Fatty Acid Composition of *Desulfovibrio* Species and Its Use in Classification of Sulfate-reducing Bacteria. *Systematic and Applied Microbiology*, **15**; 554–566.

Van den Berg, W.A.M., Stokkermans, J.P.W.G. & Van Dongen, W.M.A.M. (1989) Development of a plasmid transfer system for the anaerobic sulfate reducer, *Desulfovibrio vulgaris*. *Journal of Biotechnology*, **12**; 173–184.

Van den Berg, W.A.M., Van Dongen, W.M.A.M. & Veeger, C. (1991) Reduction of the amount of periplasmic hydrogenase in *Desulfovibrio vulgaris* (Hildenborough) with antisense RNA: direct evidence for an important role of this hydrogenase in lactate metabolism. *Journal of Bacteriology*, **173**; 3688–3694.

Van Dongen, S. & Abreu-Goodger, C. (2012) Using MCL to extract clusters from networks. *Methods in Molecular Biology*, **804**; 281–295.

Van Dongen, W., Hagen, W., Van den Berg, W. & Veeger, C. (1988) Evidence for an unusual mechanism of membrane translocation of the periplasmic hydrogenase *Desulfovibrio vulgaris* (Hildenborough), as derived from expression in *Escherichia coli*. *FEMS Microbiology Letters*, **50**; 5–9.

Van Houten Bernd, H.G.W., Meulepas, R.J.W.,Van Doesburg, W., Smidt, H., Muyzer, G. & Stams, A.J.M. (2009) *Desulfovibrio paquesii* sp. nov., a hydrogenotrophic sulfate-reducing bacterium isolated from a synthesis-gas-fed bioreactor treating zinc- and sulfate-rich wastewater. *International Journal of Systematic and Evolutionary Microbiology*, **59**; 229–233.

Van Lanen, S.G. & Shen, B. (2006) Microbial genomics for the improvement of natural product discovery. *Current Opinion in Microbiology*, **9**; 252–260.

Van Opijnen, T., Dedrick, S. & Bento, J. (2016) Strain Dependent Genetic Networks for Antibiotic-Sensitivity in a Bacterial Pathogen with a Large Pan-Genome. *PLoS Pathogens*, **12**.

Vandieken, V., Knoblauch, C. & Jørgensen, B.B. (2006) *Desulfovibrio frigidus* sp. nov. and *Desulfovibrio ferrireducens* sp. nov., psychrotolerant bacteria isolated from Arctic fjord sediments (Svalbard) with the ability to reduce Fe(III). *International Journal of Systematic and Evolutionary Microbiology*, **56**; 681–685.

Vásquez, M. C., Silva, E. E. & Castillo, E. F. (2017) Hydrotreatment of vegetable oils: A review of the technologies and its developments for jet biofuel production. *Biomass and Bioenergy*, **105**; 197-206

Vernikos, G., Medini, D., Riley, D.R. & Tettelin, H. (2015) Ten years of pan-genome analyses. *Current Opinion in Microbiology*, **23**; 148–154.

Vieira, G., Sabarly, V., Bourguignon, P.Y., Durot, M., Fèvre, F. Le, Mornico, D., Vallenet, D., Bouvet, O., Denamur, E., Schachter, V. & Médigue, C. (2011) Core and panmetabolism in *Escherichia coli*. *Journal of Bacteriology*, **193**; 1461–1472.

Vita, N., Valette, O., Brasseur, G., Lignon, S., Denis, Y., Ansaldi, M., Dolla, A. & Pieulle, L. (2015) The primary pathway for lactate oxidation in *Desulfovibrio vulgaris*. *Frontiers in Microbiology*, **6**; 1–15.

Voordouw, G. (1995) Minireview The Genus *Desulfovibrio* : The Centennial. *Applied and Environmental Microbiology*, **61**; 2813–2819.

Voordouw, G., Hagen, W.R., Krüse-Wolters, M., Van Berkel-Arts, A. & Veeger, C. (1987) Purification and characterization of *Desulfovibrio vulgaris* (Hildenborough) hydrogenase expressed in *Escherichia coli*. *European Journal of Biochemistry*, **162**; 31–36.

Voordouw, G., Pollock, W.B., Bruschi, M., Guerlesquin, F., Rapp-Giles, B.J. & Wall, J.D. (1990) Functional expression of *Desulfovibrio vulgaris* Hildenborough cytochrome c_3 in *Desulfovibrio desulfuricans* G200 after conjugational gene transfer from *Escherichia coli. Journal of Bacteriology*, **172**; 6122–6126.

Voordouw, G. (2002) Carbon Monoxide Cycling by *Desulfovibrio vulgaris* Hildenborough. *Society*, **184**; 5903–5911.

Wackett, L.P., Frias, J.A., Seffernick, J.L., Sukovich, D.J. & Cameron, S.M. (2007) Genomic and biochemical studies demonstrating the absence of an alkane-producing phenotype in *Vibrio furnissii* M1. *Applied and Environmental Microbiology*, **73**; 7192–7198.

Walker, C.B., He, Z., Yang, Z.K., Ringbauer, J.A., He, Q., Zhou, J., Voordouw, G., Wall, J.D., Arkin, A.P., Hazen, T.C., Stolyar, S. & Stahl, D.A. (2009) The electron transfer system of syntrophically grown *Desulfovibrio vulgaris*. *Journal of Bacteriology*, **191**; 5793–5801.

Walker, C.B., Stolyar, S., Chivian, D., Pinel, N., Gabster, J.A., Dehal, P.S., He, Z., Yang, Z.K., Yen, H.C.B., Zhou, J., Wall, J.D., Hazen, T.C., Arkin, A.P. & Stahl, D.A. (2009) Contribution of mobile genetic elements to *Desulfovibrio vulgaris* genome plasticity. *Environmental Microbiology*, **11**; 2244–2252.

Wall, J.D., Murnan, T., Argyle, J., English, R.S. & Rapp-Giles, B.J. (1996) Transposon mutagenesis in *Desulfovibrio desulfuricans*: Development of a random mutagenesis tool from Tn7. *Applied and Environmental Microbiology*, **62**; 3762–3767.

Wang, J. & Zhu, K. (2018) Microbial production of alka(e)ne biofuels. *Current Opinion in Biotechnology*, **50**; 11–18.

Waring, R.H. & Klovrza, L.V. (2000) Sulphur metabolism in autism. *Journal of Nutritional and Environmental Medecine*, **10**; 25–32.

Warren, Y.A., Citron, D.M., Merriam, C.V., Goldstein, J.C. & Goldstein, E.J.C. (2005) Biochemical Differentiation and Comparison of *Desulfovibrio* Species and Other Phenotypically Similar Genera. *Journal of Clinical Microbiology.*, **43**; 4041–4045.

Weber, E., Engler, C., Gruetzner, R., Werner, S. & Marillonnet, S. (2011) A modular cloning system for standardized assembly of multigene constructs. *PLoS ONE*, **6**.

Wei, L., Liu, Y., Dubchak, I., Shon, J. & Park, J. (2002) Comparative genomics approaches to study organism similarities and differences. *Journal of Biomedical Informatics*, **35**; 142–150.

Wheeler, T. & Von Braun, J. (2013) Climate Change Impacts on Global Food Security. *Science*, **341**; 508–513.

Widdel, F., Knittel, K. & Galushko, A. (2010) Anaerobic Hydrocarbon-Degrading Microorganisms: An Overview. In *Handbook of Hydrocarbon and Lipid Microbiology* (ed. by Springer). Berlin Heidelberg, pp. 1997–2021.

Wolny, D., Lodowska, J., Jaworska-Kik, M., Kurkiewicz, S., Węglarz, L. & Dzierzewicz, Z. (2011) Chemical composition of *Desulfovibrio desulfuricans* lipid A. *Archives of Microbiology*, **193**; 15–21.

Wu, B., Zhang, B., Feng, X., Rubens, J.R., Huang, R., Hicks, L.M., Pakrasi, H.B. & Tang, Y.J. (2010) Alternative isoleucine synthesis pathway in cyanobacterial species. *Microbiology*, **156**; 596–602.

Xiao, S., Shiloach, J. & Betenbaugh, M.J. (2014) Engineering cells to improve protein expression. *Current Opinion in Structural Biology*, **26**; 32–38.

Xin, F.H., Zhang, Y., Xue, S.J., Chi, Z., Liu, G.L., Hu, Z. & Chi, Z.M. (2017) Heavy oils (mainly alkanes) over-production from inulin by *Aureobasidium melanogenum* 9-1 and its transformant 88 carrying an inulinase gene. *Renewable Energy*, **105**; 561–568.

Yokoyama, K., Ohkuma, S., Taguchi, H., Yasunaga, T., Wakabayashi, T. & Yoshida, M. (2000) V-Type H^+-ATPase/Synthase from a Thermophilic Eubacterium, *Thermus thermophilus*. *Biochemistry*, **275**; 13955–13961.

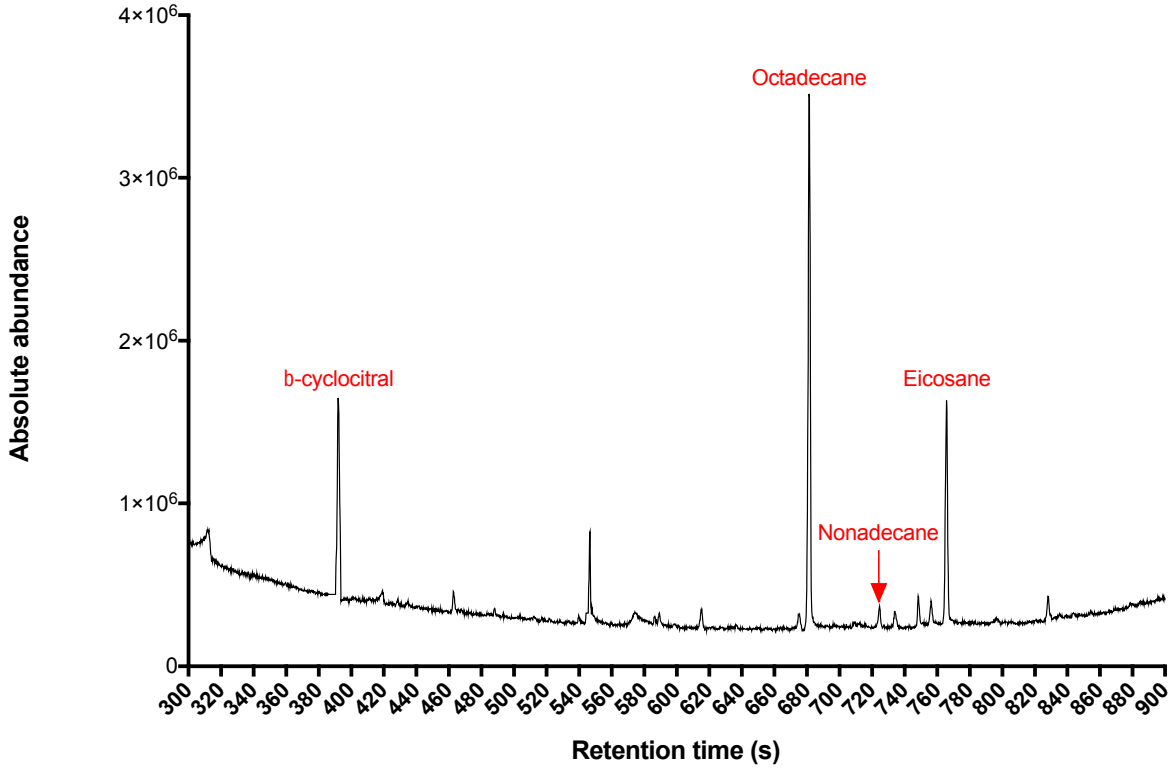
Youssef, N., Elshahed, M.S. & McInerney, M.J. (2009) Microbial processes in oil fields: culprits, problems, and opportunities. *Advances in Applied Microbiology*, **66**; 141–251.

Zabielski, P., Ford, G.C., Mai Person, X., Jaleel, A., Dewey, J.D. & Sreekumaran Nair, K. (2013) Comparison of different mass spectrometry techniques in the measurement of L-[ring-$^{13}C_6$]phenylalanine incorporation into mixed muscle proteins. *Journal of Mass Spectrometry*, **48**; 269–275.

Zhou, Y.J., Buijs, N.A., Zhu, Z., Gómez, D.O., Boonsombuti, A., Siewers, V. & Nielsen, J. (2016) Harnessing Yeast Peroxisomes for Biosynthesis of Fatty-Acid-Derived Biofuels and Chemicals with Relieved Side-Pathway Competition. *Journal of the American Chemical Society*, **138**; 15368–15377.

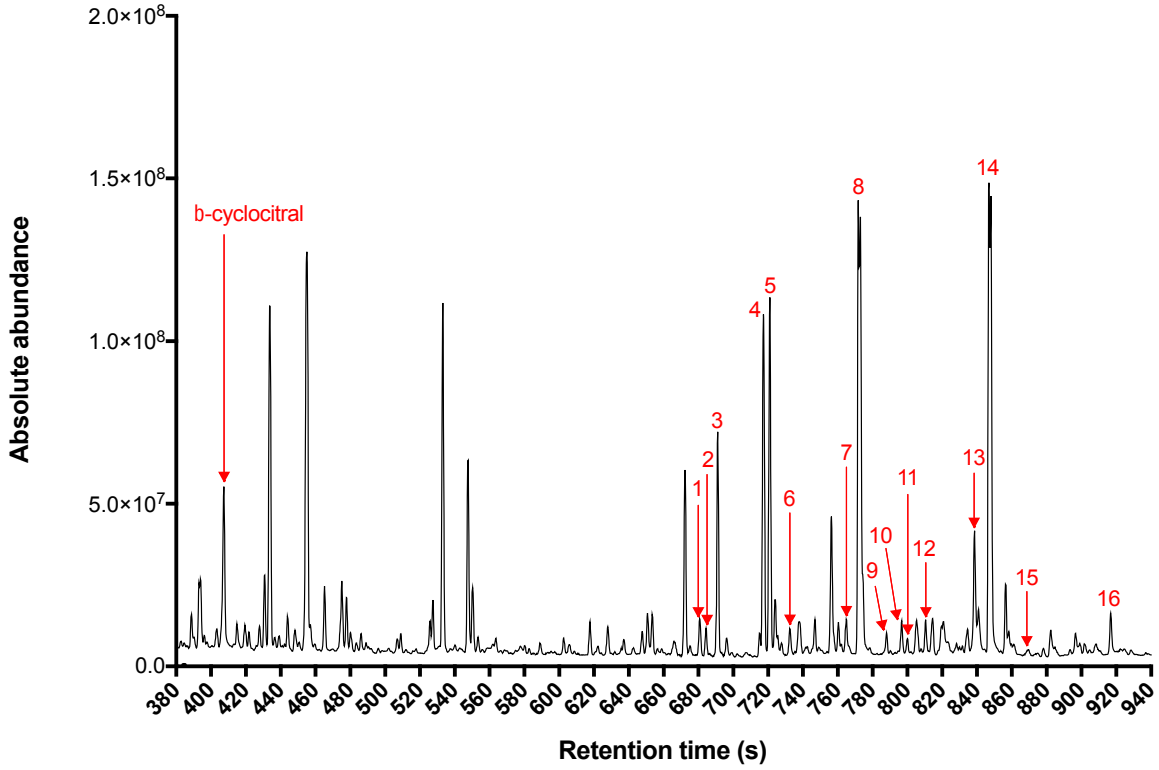
Zhou, Z., Dang, Y., Zhou, M., Li, L., Yu, C., Fu, J., Chen, S. & Liu, Y. (2016) Codon usage is an important determinant of gene expression levels largely through its effects on transcription. *Proceedings of the National Academy of Sciences*, **113**; E6117–E6125.

Appendices



Appendix 1. Total ion chromatogram of cellular organic compounds extracted from *D. desulfuricans* NCIMB 8326

D. desulfuricans NCIMB 8326 was cultured in a sodium lactate medium under an anaerobic atmosphere of 80 % N_2, 10 % CO_2 and 10 % H_2. After 7-day incubation, cellular organic compounds were extracted in DCM and analysed by GC-MS. The total ion chromatogram of organic compounds from the *D. desulfuricans* NCIMB 8326 were screened for alkanes. The organic compound extracts were spiked with 10 μM of β-cyclocitral (internal standard).



Appendix 2. Total ion chromatogram of cellular organic compounds derivatised by silylation extracted from *D. desulfuricans* NCIMB 8326

D. desulfuricans NCIMB 8326 was cultured in a sodium lactate medium under an anaerobic atmosphere of 80 % N_2, 10 % CO_2 and 10 % H_2. After 10-day incubation, cellular organic compounds were extracted in DCM, derivatised by silylation and analysed by GC-MS. The total ion chromatogram of derivatised organic compounds from the *D. desulfuricans* NCIMB 8326 were screened for fatty acids: 1 and 2, unsaturated tetradecenoic acid ($C_{14:1}$); 3, saturated tetradecanoic acid ($C_{14:0}$), 4; saturated *iso*-pentadecanoic acid ($C_{15:0}$); 5, saturated *anteiso*-pentadecanoic acid ($C_{15:0}$); 6, saturated pentadecanoic acid ($C_{15:0}$); 7, unsaturated hexadecenoic acid ($C_{16:1}$); 8, saturated hexadecanoic acid ($C_{16:0}$); 9, unsaturated *iso*-heptadecenoic acid ($C_{17:1}$); 10, saturated *iso*-heptadecanoic acid ($C_{17:0}$); 11, saturated *anteiso*-heptadecanoic acid ($C_{17:0}$); 12, saturated heptadecanoic acid ($C_{17:0}$); 13, unsaturated octadecenoic acid ($C_{18:1}$); 14, saturated octadecanoic acid ($C_{18:0}$); 15; unsaturated nonadecenoic acid ($C_{19:1}$), 16; saturated eicosanoic acid ($C_{20:0}$). The organic compound extract was spiked with 10 μM of β-cyclocitral (internal standard).

Appendix 3. Greatest homology of protein domains from characterised alka(e)ne biosynthetic enzymes identified in alkane producing *Desulfovibrio* proteomes

Protein domains identified in characterised Alka(e)ne biosynthetic enzymes	**Alkane producing *Desulfovibrio* proteome strain**	**Sequence E-value**	**Sequence bit score**	**Sequence bias**	**Description of the protein best hit**
LuxC (PF05893)	*D. gabonensis* NCIMB 10636	$1.8e^{-68}$	229.3	0.0	Acyl-CoA reductase (LuxC)
	D. desulfuricans NCIMB 8326	$1.7e^{-68}$	229.4	0.0	Acyl-CoA reductase (LuxC)
Acyl_transf_2 (PF02273)	No hit	-	-	-	-
LuxE (PF04443)	*D. gabonensis* NCIMB 10636	$6.4e^{-14}$	49.5	0.0	Acyl-protein synthetase (LuxE)
	D. desulfuricans NCIMB 8326	$2.2e^{-13}$	47.8	0.0	Acyl-protein synthetase (LuxE)
AMP-binding (PF00501)	*D. gabonensis* NCIMB 10636	$8.4e^{-118}$	391.4	0.0	Long chain fatty acid-CoA ligase
	D. desulfuricans NCIMB 8326	$6e^{-118}$	391.9	0.0	Long chain fatty acid-CoA ligase
	D. gigas NCIMB 9332	$7.1e^{-101}$	335.5	0.0	Acetyl-CoA synthetase
	D. desulfuricans NCIMB 8338	$3.9e^{-117}$	389.4	0.0	Long chain fatty acid-CoA ligase
	D. paquesii DSM 16681	$3.3e^{-102}$	340.0	0.0	Acetyl-CoA synthetase
	D. marinus DSM 18311	$2e^{-117}$	390.3	0.0	Long chain fatty acid-CoA ligase
PP-binding (PF00550)	*D. gabonensis* NCIMB 10636	$7.5e^{-15}$	53.3	1.7	Acyl carrier protein
	D. desulfuricans NCIMB 8326	$7.8e^{-15}$	52.9	1.9	Acyl carrier protein
	D. gigas NCIMB 9332	$2.7e^{-16}$	57.4	1.2	Acyl carrier protein
	D. desulfuricans NCIMB 8338	$3.4e^{-14}$	51.0	1.4	Acyl carrier protein
	D. paquesii DSM 16681	$2.8e^{-16}$	57.5	1.0	Acyl carrier protein
	D. marinus DSM 18311	$3.3e^{-14}$	51.0	1.4	Acyl carrier protein

NAD_binding_4 (PF07993)	*D. gabonensis* NCIMB 10636	$2.7e^{-16}$	57.3	0.0	dTDP-glucose 4,6-dehydratase 2
	D. desulfuricans NCIMB 8326	$3.1e^{-16}$	57.2	0.0	dTDP-glucose 4,6-dehydratase 2
	D. gigas NCIMB 9332	$4.2e^{-12}$	43.3	0.1	UDP-glucose 4-epimerase
	D. desulfuricans NCIMB 8338	$1.7e^{-15}$	54.9	0.0	dTDP-glucose 4,6-dehydratase
	D. paquesii DSM 16681	$3.6e^{-14}$	50.3	0.0	dTDP-glucose 4,6-dehydratase
	D. marinus DSM 18311	$1.7e^{-15}$	54.9	0.0	dTDP-glucose 4,6-dehydratase
Ald_deCOase (PF11266)	No hit	-	-	-	-
p450 (PF00067)	No hit	-	-	-	-
Thiolase_N (PF00108)	*D. gabonensis* NCIMB 10636	$1.1e^{-7}$	29.4	0	3-oxoacyl-[acyl-carrier-protein] synthase 2
	D. desulfuricans NCIMB 8326	$7.8e^{-8}$	29.8	0.2	3-oxoacyl-[acyl-carrier-protein] synthase 2
	D. gigas NCIMB 9332	$7.1e^{-8}$	29.8	1.2	3-oxoacyl-[acyl-carrier-protein] synthase 3
	D. desulfuricans NCIMB 8338	$8.6e^{-8}$	29.8	0.2	3-oxoacyl-[acyl-carrier-protein] synthase 2
	D. paquesii DSM 16681	$6.2e^{-9}$	33.4	1.1	3-oxoacyl-[acyl-carrier-protein] synthase 3
	D. marinus DSM 18311	$8.3e^{-8}$	29.9	0.2	3-oxoacyl-[acyl-carrier-protein] synthase 2
ACP_syn_III_C (PF08541)	*D. gabonensis* NCIMB 10636	$1.4e^{-35}$	119.2	1.1	3-oxoacyl-[acyl-carrier-protein] synthase 3
	D. desulfuricans NCIMB 8326	$2.8e^{-35}$	118.2	1.2	3-oxoacyl-[acyl-carrier-protein] synthase 3
	D. gigas NCIMB 9332	$4.4e^{-36}$	120.6	0.1	3-oxoacyl-[acyl-carrier-protein] synthase 3
	D. desulfuricans NCIMB 8338	$4.2e^{-35}$	117.8	0.3	3-oxoacyl-[acyl-carrier-protein] synthase 3
	D. paquesii DSM 16681	$5e^{-35}$	117.4	0.0	3-oxoacyl-[acyl-carrier-protein] synthase 3
	D. marinus DSM 18311	$4.1e^{-35}$	117.8	0.3	3-oxoacyl-[acyl-carrier-protein] synthase 3

Abhydrolase_1 (PF00561)	*D. gabonensis* NCIMB 10636	$1.1e^{-34}$	118.1	0.0	Homoserine O-acetyltransferase
	D. desulfuricans NCIMB 8326	$1.8e^{-35}$	120.7	0.0	Homoserine O-acetyltransferase
	D. gigas NCIMB 9332	$9.4e^{-30}$	101.7	0.1	Homoserine O-acetyltransferase
	D. desulfuricans NCIMB 8338	$1.1e^{-34}$	118.2	0.0	Homoserine O-acetyltransferase
	D. paquesii DSM 16681	$2.4e^{-33}$	113.7	0.1	Homoserine O-acetyltransferase
	D. marinus DSM 18311	$1.4e^{-35}$	117.8	0.0	Homoserine O-acetyltransferase
3Beta_HSD (PF01073)	*D. gabonensis* NCIMB 10636	$4.7e^{-22}$	76.1	0.0	dTDP-glucose 4,6-dehydratase
	D. desulfuricans NCIMB 8326	$2.4e^{-22}$	77.1	0.0	dTDP-glucose 4,6-dehydratase
	D. gigas NCIMB 9332	$3.6e^{-22}$	76.3	0.0	UDP-glucose 4-epimerase
	D. desulfuricans NCIMB 8338	$2.2e^{-21}$	74.1	0.0	UDP-glucose 4-epimerase
	D. paquesii DSM 16681	$5.9e^{-23}$	79.0	0.0	UDP-glucose 4-epimerase
	D. marinus DSM 18311	$2.2e^{-21}$	74.1	0.0	UDP-glucose 4-epimerase
ketoacyl-synt (PF00109)	*D. gabonensis* NCIMB 10636	$1.9e^{-61}$	205.7	0.1	3-oxoacyl-[acyl-carrier-protein] synthase 2
	D. desulfuricans NCIMB 8326	$1.9e^{-61}$	205.7	0.1	3-oxoacyl-[acyl-carrier-protein] synthase 2
	D. gigas NCIMB 9332	$5.8e^{-82}$	272.7	0.0	Erythronolide synthase, modules 3 and 4
	D. desulfuricans NCIMB 8338	$3.3e^{-61}$	205.1	0.1	3-oxoacyl-[acyl-carrier-protein] synthase 2
	D. paquesii DSM 16681	$1.4e^{-59}$	199.6	0.2	3-oxoacyl-[acyl-carrier-protein] synthase 2
	D. marinus DSM 18311	$3.2e^{-61}$	205.1	0.1	3-oxoacyl-[acyl-carrier-protein] synthase 2

ketoacyl-synt_C (PF02801)	*D. gabonensis* NCIMB 10636	$1.8e^{-37}$	125.9	0.5	3-oxoacyl-[acyl-carrier-protein] synthase 2
	D. desulfuricans NCIMB 8326	$1.7e^{-37}$	125.9	0.5	3-oxoacyl-[acyl-carrier-protein] synthase 2
	D. gigas NCIMB 9332	$3.1e^{-39}$	131.3	0.1	3-oxoacyl-[acyl-carrier-protein] synthase 2
	D. desulfuricans NCIMB 8338	$2.6e^{-37}$	125.5	0.3	3-oxoacyl-[acyl-carrier-protein] synthase 2
	D. paquesii DSM 16681	$2.1e^{-36}$	122.4	0.0	3-oxoacyl-[acyl-carrier-protein] synthase 2
	D. marinus DSM 18311	$1.7e^{-37}$	126.1	0.4	3-oxoacyl-[acyl-carrier-protein] synthase 2
KAsynt_C_assoc (PF16197)	No hit	-	-	-	-
Acyl_transf_1 (PF00698)	*D. gabonensis* NCIMB 10636	$2.6e^{-35}$	120.3	0.7	Malonyl CoA acyl carrier protein transacylase
	D. desulfuricans NCIMB 8326	$2.7e^{-35}$	120.2	0.5	Malonyl CoA acyl carrier protein transacylase
	D. gigas NCIMB 9332	$6.1e^{-65}$	217.4	0.0	Erythronolide synthase, modules 3 and 4
	D. desulfuricans NCIMB 8338	$1.2e^{-37}$	128.1	0.4	Malonyl CoA acyl carrier protein transacylase
	D. paquesii DSM 16681	$1.2e^{-36}$	124.7	0.5	Malonyl CoA acyl carrier protein transacylase
	D. marinus DSM 18311	$2.5e^{-38}$	130.3	0.2	Malonyl CoA acyl carrier protein transacylase
KR (PF08659)	*D. gabonensis* NCIMB 10636	$3.5e^{-22}$	77.0	0.5	3-oxoacyl-[acyl-carrier-protein] reductase FabG
	D. desulfuricans NCIMB 8326	$1.7e^{-22}$	78.1	0.9	3-oxoacyl-[acyl-carrier-protein] reductase FabG
	D. gigas NCIMB 9332	$1.4e^{-22}$	215.0	3.3	Phthiocerol synthesis polyketide synthase type I PpsC
	D. desulfuricans NCIMB 8338	$5.3e^{-22}$	76.6	2.3	3-oxoacyl-[acyl-carrier-protein] reductase FabG
	D. paquesii DSM 16681	$5.9e^{-22}$	86	0.4	Putative oxidoreductase
	D. marinus DSM 18311	$1.5e^{-21}$	75.1	2.3	3-oxoacyl-[acyl-carrier-protein] reductase FabG

Sulfotransfer_3 (PF13469)	*D. gabonensis* NCIMB 10636	$2.2e^{-13}$	49.0	1.6	Sulfotransferase domain protein
	D. desulfuricans NCIMB 8326	$7.2e^{-19}$	66.9	0.4	Sulfotransferase domain protein
	D. gigas NCIMB 9332	$1.1e^{-14}$	53.0	0.1	Hypothetical protein
	D. desulfuricans NCIMB 8338	$3.9e^{-13}$	48.3	0.4	Sulfotransferase domain protein
	D. marinus DSM 18311	$2.4e^{-18}$	65.3	0.0	Hypothetical protein
Haem_ oxygenas_2 (PF14518)	*D. gigas* NCIMB 9332	$1.1e^{-5}$	23	0.6	Heme oxygenase
FA-desaturase (PF00487)	No hit	-	-	-	-

Alkane producing *Desulfovibrio* proteomes were screened for 20-protein domain HMM profiles identified in the characterised alka(ene) biosynthetic enzymes, using the program hmmsearch with an E-value threshold equal or superior to $10e^{-5}$. The lowest E-value hit for each protein domain found in each *Desulfovibrio* proteome is reported.

Appendix 4. List of protein clusters predicted to be exclusively present in alkane producing *Desulfovibrio* strains by the pan-genomic analysis tool Anvi'o

Protein clusters	Anvi'o protein cluster annotation	False positive protein clusters	Protein cluster annotation according to sequence similarity against UniProt database	Protein cluster annotation according to domain similarity against UniProt database
PC_00001333	Putative lipoic acid-binding regulatory protein	No hit	Uncharacterised protein	No hit
PC_00001454	2-polyprenyl-3-methyl-5-hydroxy-6-metoxy-1,4-benzoquinol methylase	No hit	Methyltransferase protein	No hit
PC_00001521	Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide dehydrogenase (E3) component or related enzyme	No significant hit	Putative FAD-dependent pyridine nucleotide-disulphide oidoreductase	Mercuric reductase; Dihydrolipoyl dehydrogenase
PC_00001409	Uncharacterized Fe-S cluster-containing enzyme, radical SAM superfamily	No significant hit	Putative radical SAM domain protein	Uncharacterised protein
PC_00001403	c-di-GMP-related signal transduction protein, contains EAL and HDOD domains	No significant hit	Putative metal-dependent hydrolase HDOD	Uncharacterised protein
PC_00001377	Chromosomal replication initiation ATPase DnaA	No significant hit	Chromosomal replication initiator protein DnaA	Chromosomal replication initiator protein DnaA
PC_00001495	Prephenate dehydrogenase	No significant hit	Prephenate dehydrogenase	Prephenate dehydrogenase

PC_00001384	**Tetratricopeptide (TPR) repeat**	**No significant hit**	**TPR repeat protein**	**TPR repeat protein**
PC_00001345	Multisubunit Na+/H+ antiporter, MnhG subunit	No hit	Putative Na+/H+ antiporter subunit	Na (+)/ H (+) antiporter Subunit G1
PC_00001341	Multisubunit Na+/H+ antiporter, MnhC subunit	No hit	Putative NADH-ubiquinone Oxidoreductase chain 4L	Na (+)/ H (+) antiporter Subunit C1
PC_00001526	Archaeal/vacuolar-type H+-ATPase catalytic subunit A/Vma1	No significant hit	V-type ATPase alpha chain	V-type ATPase alpha chain
PC_00001463	Di- and tricarboxylate transporter; Di- and tricarboxylate transporter	No significant hit	Putative sodium-dependent transporter	Uncharacterised transporter
PC_00001464	Diacylglycerol kinase	No hit	Diacylglycerol kinase	No hit
PC_00001445	Hemerythrin	No significant hit	Hemerythrin	Bacteriohemerythrin
PC_00001422	MFS-type transporter involved in bile tolerance, Atg22 family	No significant hit	MFS transporter	Uncharacterised MFS-type transporter
PC_00001404	3D (Asp-Asp-Asp) domain	No significant hit	Uncharacterised protein	No hit
PC_00001484	NDP-sugar epimerase, includes UDP-GlcNAc-inverting 4,6-dehydratase FlaA1 and capsular polysaccharide biosynthesis protein EpsC	No significant hit	Polysaccharide biosynthesis protein	Capsular polysaccharide biosynthesis protein CapD
PC_00001296	Uncharacterized MnhB-related membrane protein	No hit	Uncharacterised protein	No hit

PC_00001298	Multisubunit Na+/H+ antiporter, MnhF subunit	No hit	Na (+)/H (+) antiporter subunit F	Na+/H+ antiporter subunit F
PC_00001396	Archaeal/vacuolar-type H+-ATPase subunit D/Vma8	No hit	V-type ATPase subunit D	V-type ATPase subunit D
PC_00001518	Na+-driven multidrug efflux pump	No significant hit	MATE efflux family transporter	Multidrug export protein MepA
PC_00001305	Hypothetical protein	No significant hit	Cytochrome c3	Cytochrome C3
PC_00001356	Hypothetical protein	No significant hit	Oxidoreductase	No hit
PC_00001534	Predicted carbamoyl transferase, NodU family	No significant hit	Membrane protein	No hit
PC_00001334	Multisubunit Na+/H+ antiporter, MnhB subunit; Multisubunit Na+/H+ antiporter, MnhB subunit	No hit	Na+/H+ antiporter subunit B	Na+/H+ antiporter subunit B
PC_00001470	Phage terminase large subunit	No hit	Phage terminase large subunit	No hit
PC_00001316	Nucleotide-binding universal stress protein, UspA family	No significant hit	Universal stress protein	Universal stress protein
PC_00001314	DnaJ-class molecular chaperone with C-terminal Zn finger domain	No significant hit	Di-tetraheme cytochrome C3	Cytochrome c3

PC_00001310	Desulfoferrodoxin, superoxide reductase-like (SORL) domain	No hit	Desulfoferrodoxin	Superoxide reductase
PC_00001318	FoF1-type ATP synthase, membrane subunit c/Archaeal/vacuolar-type H+-ATPase, subunit K	No hit	V-type ATPase subunit K	No hit
PC_00001319	Anti- sigma regulatory factor (antagonist of anti-sigma factor)	No significant hit	Anti-sigma factor antagonist	Anti-sigma factor antagonist
PC_00001457	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	No significant hit	Na (+)/H (+) antiporter subunit D	NADH-quinone Oxidoreductase subunit N
PC_00001492	Archaeal/vacuolar-type H+-ATPase subunit B/Vma2	No significant hit	V-type ATPase subunit beta	V-type ATPase subunit beta
PC_00001062	Divalent metal cation (Fe/Co/Zn/Cd) transporter	No hit	Cation transporter	Metal tolerance protein
PC_00001544	CheY chemotaxis protein or a CheY-like REC (receiver) domain	No significant hit	Uncharacterised protein	No hit
PC_00001389	RNA-splicing ligase RtcB, repairs tRNA damage	No significant hit	Cytoplasmic protein RtcB	tRNA-splicing ligase RtcB
PC_00001501	Acetyltransferase (isoleucine patch superfamily)	No significant hit	Putative chloramphenicol transferase	Acetyltransferase
PC_00001520	ATP-dependent protease Clp, ATPase subunit	No significant hit	Putative ATPase	ATP-dependent Clp protease ATP-binding subunit ClpX

PC_00001329	Cell division protein ZapA, inhibits GTPase activity of FtsZ	No significant hit	Cell division protein ZapA	Hypothetical domain
PC_00001323	Nucleotide-binding universal stress protein, UspA family	No significant hit	Universal stress protein	Universal stress protein
PC_00001322	Phage repressor protein C, contains Cro/C1-type HTH and peptisase s24 domains	No significant hit	Putative phage repressor	No hit
PC_00001465	CO or xanthine dehydrogenase, Mo-binding subunit	No significant hit	Putative aerobic-type CO dehydrogenase, large Subunit CoxL/CutL –like protein	Xanthine dehydrogenase Molybdenum-binding subunit
PC_00001467	PAS domain; Signal transduction histidine kinase, nitrate/nitrite-specific; Predicted signal transduction protein containing a membrane domain, an EAL and a GGDEF domain	No significant hit	Sensory box protein	Uncharacterised signalling protein
PC_00001442	ABC-type oligopeptide transport system, periplasmic component	No significant hit	Family 5 extracellular solute-binding protein	Putative ABC transporter peptide-binding protein
PC_00001423	Multidrug efflux pump subunit AcrA (membrane-fusion protein)	No significant hit	Efflux transporter MFP subunit	Macrolide export protein MacA

PC_00001408	Cell division protein YceG, involved in septum cleavage	No significant hit	Endolytic murein transglycosylase	Cell division protein YceG
PC_00001405	Predicted phosphoesterase, related to the Icc protein	No significant hit	Metallophosphoesterase	Uncharacterised protein
PC_00001417	ABC-type spermidine/putrescine transport system, permease component II	No significant hit	Binding protein dependent transport systems inner membrane protein	Probable sulphate transport system permease protein
PC_00001231	Cu/Ag efflux pump CusA; Cu/Ag efflux pump CusA	No significant hit	Acriflavin resistence protein	Cation efflux system protein CuSA
PC_00001482	Spore maturation protein CgeB	No significant hit	Uncharacterised protein	No hit
PC_00001480	Na+-driven multidrug efflux pump	No significant hit	Putative multidrug efflux transporter	Multidrug resistant protein
PC_00001398	ABC-type Fe3+/spermidine/putrescine transport systems, ATPase components	No significant hit	Fe(3+)-transporting ATPase	Spermidine/putrescine import ATP-binding protein
PC_00001395	Permease of the drug/metabolite transporter (DMT) superfamily	No significant hit	Putative permease of drug/ metabolite transporter (DMT) superfamily	Uncharacterised transporter protein
PC_00001370	DNA-binding response regulator, OmpR family, contains REC and winged-helix (wHTH) domain; CheY chemotaxis protein or a CheY-like REC (receiver) domain; K+-sensing histidine kinase KdpD	No significant hit	Sensor histidine kinase/ response regulator	Hybrid signal transduction histidine kinase

PC_00001372	DNA-binding transcriptional response regulator, NtrC family, contains REC, AAA-type ATPase, and a Fis-type DNA-binding domains; Signal transduction histidine kinase regulating C4-dicarboxylate transport system; Signal transduction histidine kinase involved in nitrogen fixation and metabolism regulation	No significant hit	Putative integral membrane sensor hybrid histidine kinase	Sensor histidine kinase protein
PC_00001375	Two-component response regulator, PleD family, consists of two REC domains and a diguanylate cyclase (GGDEF) domain	No significant hit	Diguanylate cyclase (GGDEF) domain containing protein	Reponse regulator PleD
PC_00001376	DNA-binding response regulator, OmpR family, contains REC and winged-helix (wHTH) domain	No significant hit	Putative two-component response regulator	Chemotaxis protein
PC_00001516	Glycosyltransferase involved in cell wall bisynthesis	No significant hit	Glycosyl transferase group 1	Glycosyltransferase
PC_00001510	4-amino-4-deoxy-L-arabinose transferase or related glycosyltransferase of PMT family	No significant hit	PMT family Glycosyltransferase, 4-amino-4-deoxy-L-arabinose transferase	No hit
PC_00001359	NAD(P)H-nitrite reductase, large subunit	No significant hit	Nitrite sulphite reductase 4Fe-4S region	Sulphite reductase
PC_00001352	ABC-type antimicrobial peptide transport system, permease component	No significant hit	ABC-type antimicrobial peptide transport system, permease component	Macrolide export ATP-binding/permease protein MacB

PC_00001350	Formate hydrogenylase subunit 6/NADH:ubiquinone oxidoreductase 23 kD subunit (chain I)	No significant hit	EchF, belonging to the H(+) or Na(+)-translocating nadh dehydrogenase (ndh) family	NADH-quinone Oxidoreductase subunit I
PC_00001535	Aerobic-type carbon monoxide dehydrogenase, small subunit, CoxS/CutS family	No significant hit	Putative xanthine dehydrogenase iron-sulphur subunit	Carbon monoxide dehydrogenase
PC_00001532	Flagellar capping protein FliD	No significant hit	Flagellar hook-associated protein 2	Flagellar hook-associated protein 2
PC_00001338	Peptidoglycan/xylan/chitin deacetylase, PgdA/CDA1 family	No significant hit	Glycosyl transferase 1	No hit
PC_00001331	Predicted unusual protein kinase regulating ubiquinone biosynthesis, AarF/ABC1/UbiB family	No significant hit	ABC-1 domain containing protein	Probable protein kinase
PC_00001471	HD-like signal output (HDOD) domain, no enzymatic activity	No significant hit	Putative metal dependent phosphorylase	No hit
PC_00001477	CO or xanthine dehydrogenase, FAD-binding subunit	No significant hit	Aerobic-type carbon Monoxide dehydrogenase, middle subunit CoxM/CutM-like protein	4-hydroxybenzoyl-CoA reductase subunit beta
PC_00001315	Formamidopyrimidine-DNA glycosylase	No significant hit	Formanidopyrimidine-DNA glycosylase	Formanidopyrimidine-DNA glycosylase
PC_00001312	Formate hydrogenlyase subunit 3/Multisubunit Na+/H+ antiporter, MnhD subunit	No significant hit	Putative NADH-ubiquinone/ plastoquinone (Complex I), subunit	Na+/H+ antiporter subunit A

PC_00001439	Glyoxylase or a related metal-dependent hydrolase, beta-lactamase superfamily II	No significant hit	Beta-lactamase domain protein	N-acyl homoserine lactonase
PC_00001410	Regulator of protease activity HflC, stomatin/prohibitin superfamily	No significant hit	Stomatin/prohibitin-family membrane protease subunit ybbk	Regulator of protease
PC_00001227	DNA-binding transcriptional response regulator, NtrC family, contains REC, AAA-type ATPase, and a Fis-type DNA-binding domains	No significant hit	Two component sigma-54 specific transcriptional regulator	Transcriptional regulatory protein
PC_00001491	Permease of the drug/metabolite transporter (DMT) superfamily	No significant hit	Permease of the drug/ metabolite transporter (DMT) superfamily	No hit
PC_00001382	ABC-type spermidine/putrescine transport system, permease component II	No significant hit	ABC-type transporter, integral membrane subunit	ABC transporter, permease protein
PC_00001497	Hydroxyethylthiazole kinase, sugar kinase family	No significant hit	Hydroxyethylthiazole kinase	Hydroxyethylthiazole kinase
PC_00001364	Thiamine monophosphate synthase	No significant hit	Thiamine-phosphate synthase	Thiamine-phosphate synthase
PC_00001547	AraC-type DNA-binding domain and AraC-containing proteins	No significant hit	Transcriptional regulator, AraC family	HTH-type transcriptional activator
PC_00001545	Glyoxylase or a related metal-dependent hydrolase, beta-lactamase superfamily II; Rhodanese-related sulfurtransferase	No significant hit	Metallo-beta lactamase	Beta-lactamase hydrolase-like protein

PC_00001431	5-enolpyruvylshikimate-3-phosphate synthase	No significant hit	3-phosphoshikimate 1-carboxyvinyltransferase	3-phosphoshikimate 1-carboxyvinyltransferase
PC_00001430	Enamine deaminase RidA, house cleaning of reactive enamine intermediates, YjgF/YER057c/UK114 family	No significant hit	Endoribonuclease L-PSP	Enamine deaminase
PC_00001367	Cytidine deaminase; ABC-type amino acid transport/signal transduction system, periplasmic component/domain	No significant hit	ABC transporter periplasmic protein	Cytidine deaminase
PC_00001505	Magnesium-transporting ATPase (P-type)	No significant hit	P-type ATPase, HAD Superfamily, subfamily IC	Cation transporting ATPase
PC_00001503	Polyferredoxin	No significant hit	Iron sulphur binding protein	Uncharacterised ferredoxin
PC_00001440	Hypothetical protein	No hit	Uncharacterised protein	No hit
PC_00001343	Hypothetical protein	No significant hit	Uncharacterised protein	No hit
PC_00001342	Hypothetical protein	No hit	Uncharacterised protein	No hit
PC_00001462	Hypothetical protein	No significant hit	Uncharacterised protein	No hit
PC_00001294	Hypothetical protein	No significant hit	Uncharacterised protein	No hit
PC_00001299	Hypothetical protein	No hit	Uncharacterised protein	No hit
PC_00001335	Hypothetical protein	No hit	Uncharacterised protein	No hit
PC_00001220	Hypothetical protein	No significant hit	Major capsid protein	No hit

PC_00001499	Hypothetical protein	No hit	Uncharacterised protein	No hit
PC_00001363	Hypothetical protein	No significant hit	Uncharacterised protein	No hit
PC_00001406	Hypothetical protein	No significant hit	Uncharacterised protein	No hit
PC_00001238	Hypothetical protein	No significant hit	Uncharacterised protein	No hit
PC_00001303	Hypothetical protein	No significant hit	Uncharacterised protein	No hit
PC_00001379	Anaerobic selenocysteine-containing dehydrogenase	Alignment with a protein from *D. alcoholivorans* NCIMB 12906 (bit score of 644, 49.9 % identity, E-value of 0)	Trimethylamine-N-oxide reductase	Oxidoreductase
PC_00001397	Excinuclease UvrABC, nuclease subunit	Alignment with a protein from *D. desulfuricans* NCIMB 8307 (bit score of 652, 51.4 % identity, E-value of 0) Alignment with a protein from *D. giganteus* DSM 4370 (bit score of 734, 58.0 % identity, E-value of 0) Alignment with a protein from *D. vulgaris* str. Hildenborough (bit score of 648, 52.5 % identity, E-value of 0)	UvrABC system protein C	UvrABC system protein
PC_00001433	Tfp pilus assembly protein, ATPase PilM	Alignment with a protein from *D. giganteus* DSM 4370 (bit score of 578, 46.6 % identity, E-value of 0)	Uncharacterised protein	No hit

PC_00001434	Multidrug efflux pump subunit AcrB	Alignment with a protein from *D. alcoholivorans* NCIMB 12906 (bit score of 1,282, 62.5 % identity E-value of 0) Alignment with a protein from *D. desulfuricans* NCIMB 8307 (bit score of 898, 43.9 % identity, E-value of 0) Alignment with a protein from *D. giganteus* DSM 4370 (bit score of 1,450, 67.0 % identity E-value of 0) Alignment with a protein from *D. vulgaris* str. Hildenborough (bit score of 1,017, 50.3 % identity E-value of 0)	Hydrophobe/amphiphile efflux-1 (HAE1) family transporter	Efflux pump membrane transporter
PC_00001418	Methyl-accepting chemotaxis protein; PAS domain	Alignment with a protein from *D. desulfuricans* NCIMB 8307 (bit score of 779, 48.9 % identity, E-value of 0) Alignment with a protein from *D. vulgaris* str. Hildenborough (bit score of 740, 49.0 % identity, E-value of 0)	Methyl-accepting chemotaxis sensory transducer with Pas/Pac sensor	Methyl-accepting chemotaxis protein
PC_00001415	Single-stranded DNA-specific exonuclease, DHH superfamily, may be involved in archaeal DNA replication initiation	Alignment with a protein from *D. alcoholivorans* NCIMB 12906 (bit score of 547, 53.1 % identity, E-value of 0) Alignment with a protein from *D. giganteus* DSM 4370 (bit score of 624, 57.0 % identity, E-value of 0) Alignment with a protein from *D. vulgaris* str. Hildenborough (bit score of 592, 55.4 % identity, E-value of 0)	Single stranded DNA specific exonuclease RecJ	Single stranded DNA specific exonuclease RecJ
PC_00001489	Fe2+ transport system protein B; Fe2+ transport system protein B	Alignment with a protein from *D. giganteus* DSM 4370 (bit score of 800, 50.5 % identity, E-value of 0)	Ferrous iron transport protein B	Ferrous iron transport protein B

PC_00001474	TolB amino-terminal domain (function unknown)	Alignment with a protein from *D. giganteus* DSM 4370 (bit score of 525, 46.8 % identity, E-value of 0)	Putative FG-GAP repeat Family protein	No hit
PC_00001485	Outer membrane protein assembly factor BamA	Alignment with a protein from *D. alcoholivorans* NCIMB 12906 (bit score of 837, 49.2 % identity, E-value of 0) Alignment with a protein from *D. desulfuricans* NCIMB 8307 (bit score of 652, 41.3 % identity, E-value of 0) Alignment with a protein from *D. giganteus* DSM 4370 (bit score of 999, 53.9 % identity, E-value of 0) Alignment with a protein from *D. vulgaris* str. Hildenborough (bit score of 704, 48.5 % identity, E-value of 0)	Outer membrane protein assembly factor BamA	Outer membrane protein assembly factor BamA
PC_00001533	Predicted ABC-type transport system involved in lysophospholipase L1 biosynthesis, permease component	Alignment with a protein from *D. alcoholivorans* NCIMB 12906 (bit score of 561, 44.0 % identity, E-value of 0) Alignment with a protein from *D. giganteus* DSM 4370 (bit score of 807, 52.0 % identity, E-value of 0)	Glycosyl transferase 1	ABC transporter permease

The Anvi'o pan-genomic analysis predicted 107 protein clusters that were exclusively present in alkane producing *Desulfovibrio* strains. The nine protein clusters with a predicted structure likely sharing similar structural features with alkane-binding proteins are written in bold (top of the table). Protein clusters with an unknown function after annotation verification are highlighted in grey. Protein clusters identified to be exclusively present in alkane producing strains but which the consensus sequence aligned to non-alkane producing *Desulfovibrio* protein sequences with an E-value of 0 are written in red (bottom of the table). Alignments with an E-value inferior to $10e^{-5}$ are defined as "no significant hit".

Appendix 5. List of protein clusters predicted to be exclusively present in alkane producing *Desulfovibrio* strains by the pan-genomic analysis tool *get_homologues*

Protein clusters	`Prokka` protein cluster annotation	False positive protein clusters	Protein cluster annotation according to domain similarity against UniProt database
PC_0149	**Desulforedoxin**	**No significant hit**	**Desulforedoxin**
PC_0678	**Lipoprotein NlpI**	**No significant hit**	**TPR repeat protein**
PC_3584	**Putative type 12 methyltransferase**	**No significant hit**	**No hit**
PC_2088	**Nitrogenase cofactor biosynthesis protein NifB**	**No significant hit**	**Nitrogenase cofactor biosynthesis protein NifB**
PC_3173	**Histidine kinase TodS 2**	**No significant hit**	**Sensor histidine kinase**
PC_3405	**Hypothetical protein**	**No hit**	**No hit**
PC_0120	Phage terminase large subunit	No hit	No hit
PC_0749	Von Willebrand factor typeA domain protein	No significant hit	Uncharacterised protein
PC_998	Prokaryotic diacylglycerol kinase	No hit	No hit
PC_1024	MATE family efflux transporter	No significant hit	Multidrug export protein MepA

PC_1025	PilZ domain protein	No hit	No hit
PC_1050	Hydrolase CocE/NonD family protein	No hit	Serine esterase
PC_1079	Laminin G domain protein	No significant hit	No hit
PC_1132	Cation transporter FieF	No significant hit	Zinc transporter
PC_1200	Racemase	No hit	Alanine racemase
PC_1209	CheW-like domain protein, chemotaxis protein	No hit	Chemotaxis protein CheW
PC_1211	Purine-binding chemotaxis protein	No significant hit	Chemotaxis CheW protein
PC_1359	V-type ATP synthase subunit K	No hit	No hit
PC_1360	V-type ATP synthase subunit I	No hit	V-type ATPase subunit I
PC_1361	V-type ATP synthase subunit D	No hit	V-type ATP Synthase subunit D
PC_1362	NtpB, V-type ATPase subunit B	No significant hit	V-type ATPase subunit B
PC_1363	NtpA, V-type ATPase subunit A	No significant hit	V-type ATPase subunit A
PC_1364	V-type ATPase subunit C	No hit	No hit
PC_1365	V-type ATP synthase subunit E	No hit	V-type ATPase subunit E
PC_1409	Divergent AAA domain protein	No hit	No hit

PC_1712	FeoA, Iron transport protein	No significant hit	Ferrous iron transport protein A
PC_1852	UspA, Universal stress protein	No significant hit	Universal stress protein
PC_2096	CyaA 2, Adenylate cyclase	No significant hit	Adenylate cyclase
PC_2228	YmfD 2, Major Facilitator Superfamily transporter	No significant hit	Arabinose-proton symporter
PC_2346	Cation diffusion facilitator family transporter	No hit	No hit
PC_2528	Maa, Maltose acetyltransferase	No significant hit	Acetyltransferase
PC_2663	Citrate transporter	No significant hit	Uncharacterised transporter
PC_3257	Acyl-coenzyme A:6-aminopenicillanic acid acyl-transferase	No hit	Acyl-coenzyme A:6-aminopenicillanic acid acyl-transferase
PC_3317	SufS, Aminotransferase class V	No significant hit	Aminotransferase class V
PC_3347	Patatin-like_ phospholipase	No hit	No hit
PC_3431	Nlr, Superoxide dismutase	No hit	Neelaredoxin
PC_3588	Putative Rubrerythrin	No hit	No hit
PC_3692	FtrB, Transcriptional activator	No hit	Uncharacterised protein
PC_3697	MnhD1, Na (+)/H (+) antiporter subunit D	No significant hit	NADH-quinone oxidoreductase subunit N

PC_3698	Putative monovalent cation/H+ antiporter subunit C	No hit	Na (+)/H (+) antiporter subunit C1
PC_3699	MrpB, Na (+)/H (+) antiporter subunit B	No hit	Na (+)/H (+) antiporter subunit B
PC_3700	Putative monovalent cation/H+ antiporter subunit B	No hit	No hit
PC_3701	MnhG1, Na (+)/H (+) antiporter Subunit G	No hit	Na (+)/H (+) antiporter Subunit G1
PC_3702	Putative monovalent cation/H+ antiporter subunit F	No hit	Na (+)/H (+) antiporter Subunit F
PC_3703	MnhE1, Na (+)/H (+) antiporter subunit E	No hit	Na (+)/H (+) antiporter Subunit E1
PC_0064	Cat 1, Putative chloramphenicol acetyltransferase	No significant hit	Chloramphenicol acetyltransferase
PC_0071	Polysaccharide deacetylase	No significant hit	No hit
PC_0096	ZraS 1, Pas/Pac sensor signal transduction histidine kinase	No significant hit	Sensor histidine kinase
PC_0581	CusB, RND family efflux transporter MFP subunit	No significant hit	Cation efflux system protein
PC_0703	Soj 1, Chromosome partitioning protein	No significant hit	Putative replication protein
PC_0718	Polymer-forming cytoskeletal	No significant hit	Polymer-forming cytoskeletal

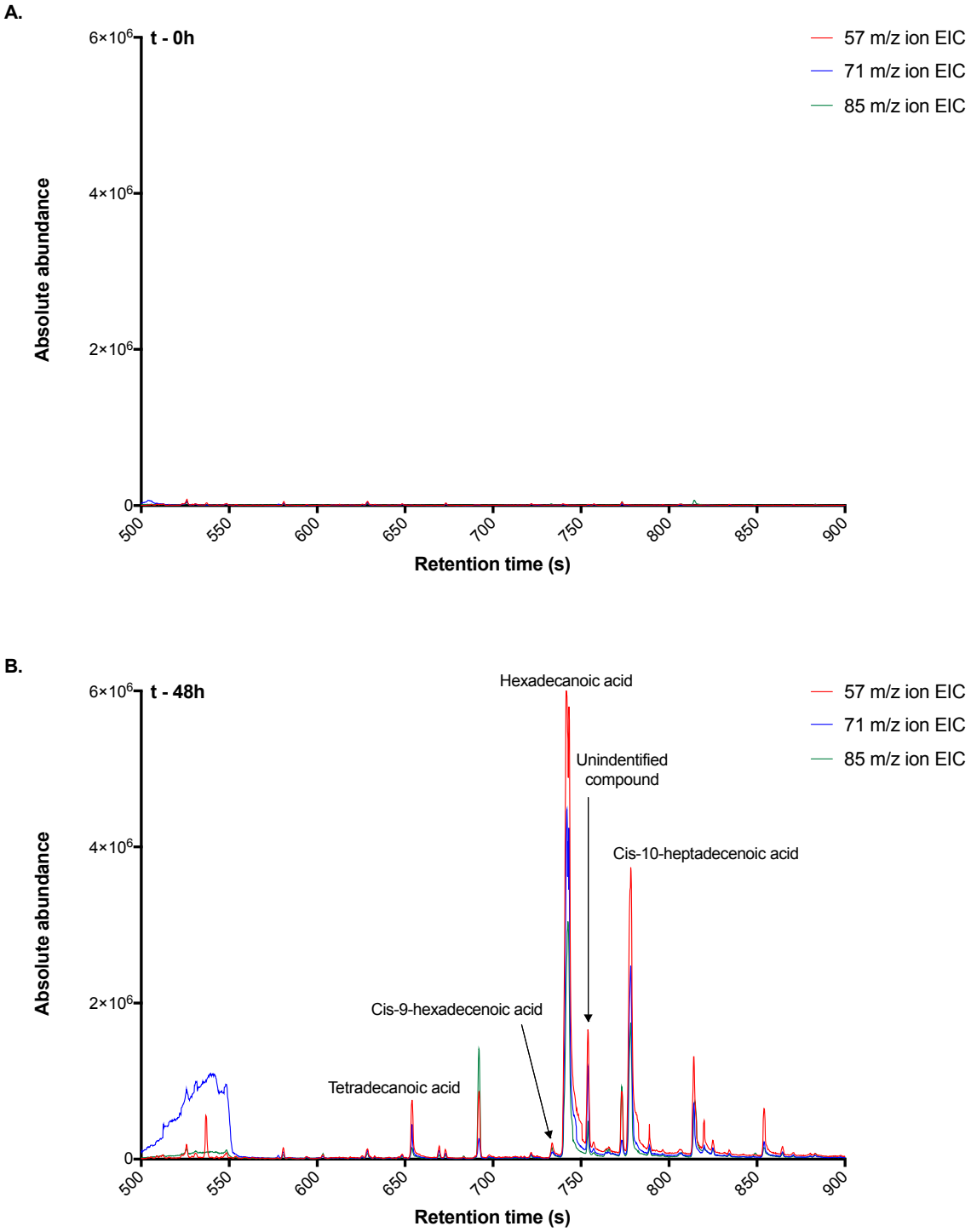
PC_0937	HflC 1, Putative stomatin/prohibitin-family membrane protease	No significant hit	FtsH protease regulator HflK
PC_0965	CbdC, Putative oxidoreductase FAD/NAD(P)-binding domain protein	No significant hit	Flavohemoprotein
PC_1205	ArcB 4, Multi-sensor hybrid histidine kinase	No significant hit	Signal transduction histidine protein kinase
PC_1206	CheB3 2, Chemotaxis response regulator protein-glutamate methylesterase	No significant hit	Chemotaxis response regulator protein-glutamate methylesterase
PC_1208	FrzCD, Methyl-accepting chemotaxis sensory transducer	No significant hit	Methyl-accepting chemotaxis chemotaxis protein
PC_1210	CheR2 2, Chemotaxis protein Methyltransferase	No significant hit	Chemotaxis protein Methyltransferase
PC_1228	ArlR 2, Putative two-component response regulator	No significant hit	Chemotaxis protein
PC_1262	Proteolipid membrane potential modulator	No significant hit	Proteolipid membrane potential modulator
PC_1702	Blue-light-activated protein	No significant hit	Sensor protein
PC_1773	PglF, Polysaccharide biosynthesis protein	No significant hit	Capsular polysaccharide biosynthesis protein
PC_1838	BtrV 3, Anti-sigma factor antagonist	No significant hit	Anti-sigma factor antagonist
PC_2015	ExbB 2, Biopolymer transporter protein	No significant hit	Biopolymer transporter protein

PC_2436	Cupin domain protein	No significant hit	Uncharacterised protein
PC_2565	PaaE 1, Oxidoreductase FAD/NAD(P) binding domain-containing protein	No significant hit	No hit
PC_2847	Spermidine-putrescine ABC transporter membrane protein	No significant hit	Transport system permease protein
PC_2848	PotH, Binding-protein-dependent transport system, inner membrane subunit	No significant hit	Probable sulphate transport system permease protein
PC_2878	PleD 2, Diguanylate cyclase (GGDEF) domain-containing protein	No significant hit	Response regulator
PC_2898	MepS, NLP/P60 protein	No significant hit	Gamma-DL-glutamyl hydrolase
PC_3085	Signal transduction histidine kinase regulating citrate/ malate metabolism	No significant hit	No hit
PC_3172	GAF domain protein	No significant hit	No hit
PC_3416	Long-chain-fatty-acid CoA ligase FadD15	No significant hit	Long chain fatty acid CoA ligase
PC_3457	HcrC, Xanthine dehydrogenase yagT iron-sulphur-binding subunit	No significant hit	Xanthine dehydrogenase yagT iron-sulphur-binding subunit
PC_3693	MrpD, Na (+)/H (+) antiporter subunit D	No significant hit	NADH-quinone Oxidoreductase chain N
PC_3696	HyfB, Na (+)/H (+) antiporter subunit A	No significant hit	NADH-quinone oxidoreductase subunit N

PC_1245	Hypothetical protein	No hit	No hit
PC_0063	Hypothetical protein	No hit	No hit
PC_0079	Hypothetical protein	No significant hit	No hit
PC_0104	Hypothetical protein	No hit	No hit
PC_0209	Hypothetical protein	No hit	No hit
PC_0311	Hypothetical protein	No hit	No hit
PC_0409	Hypothetical protein	No hit	No hit
PC_0449	Hypothetical protein	No hit	No hit
PC_0483	Hypothetical protein	No significant hit	No hit
PC_0657	Hypothetical protein	No hit	No hit
PC_0658	Hypothetical protein	No hit	No hit
PC_0938	Hypothetical protein	No hit	No hit
PC_1015	Hypothetical protein	No significant hit	No hit
PC_1159	Hypothetical protein	No hit	No hit
PC_1160	Hypothetical protein	No hit	No hit
PC_1218	Hypothetical protein	No hit	No hit
PC_1238	Hypothetical protein	No significant hit	No hit
PC_1406	Hypothetical protein	No hit	No hit

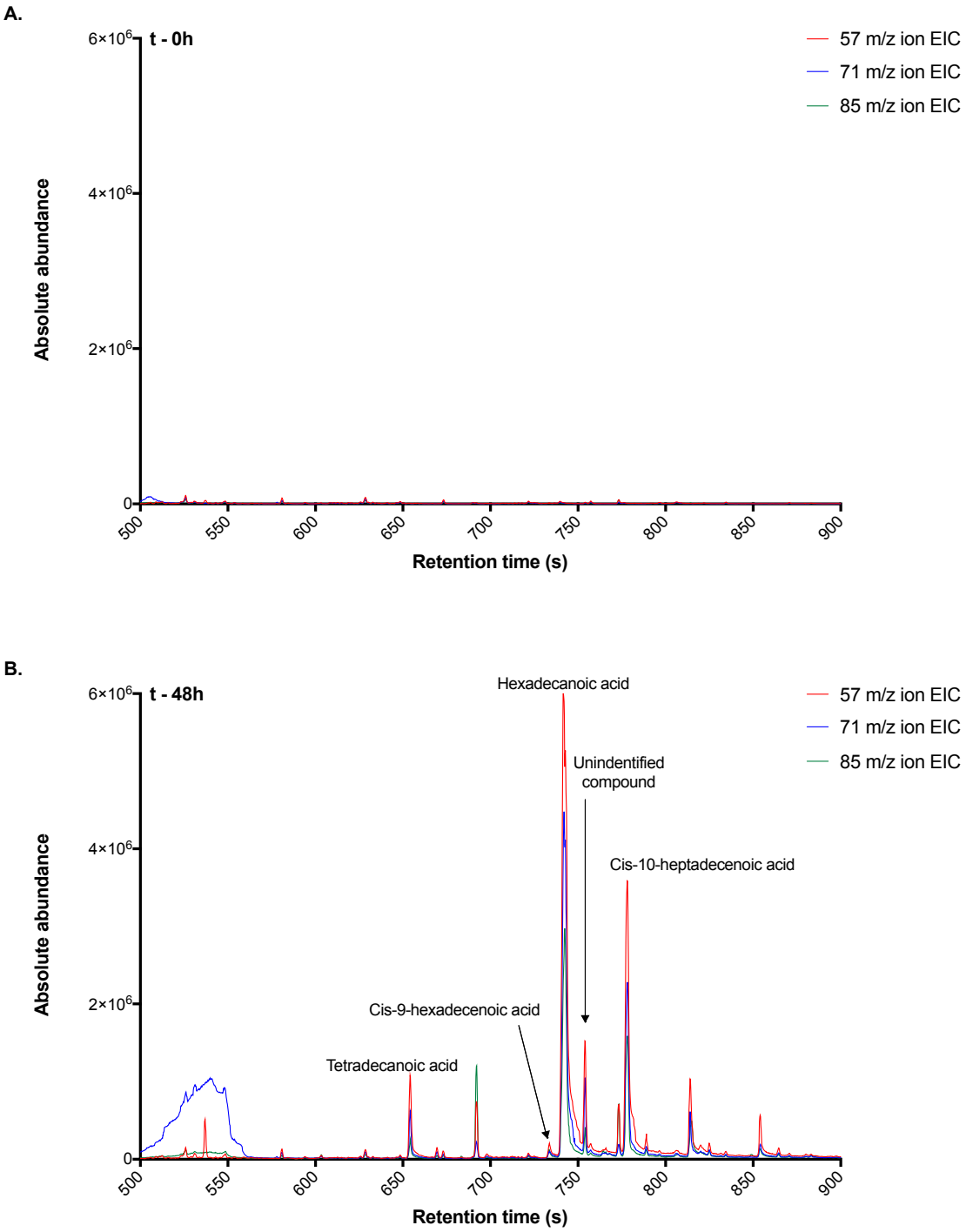
PC_1850	Hypothetical protein	No hit	No hit
PC_2182	Hypothetical protein	No hit	No hit
PC_2404	Hypothetical protein	No significant hit	No hit
PC_2887	Hypothetical protein	No hit	No hit
PC_2949	Hypothetical protein	No hit	No hit
PC_3232	Hypothetical protein	No hit	No hit
PC_3239	Hypothetical protein	No hit	No hit
PC_0308	Hypothetical protein	No significant hit	No hit
PC_0920	Hypothetical protein	No significant hit	No hit
PC_3354	Hypothetical protein	No significant hit	No hit
PC_3837	TorZ, Trimethylamine-N-oxide reductase	Alignment with a protein sequence from D. alcoholivorans NCIMB 12906 (bit score of 644, 49.9 % identity, E-value of 0)	Oxidoreductase

The *get_homologues* pan-genomic analysis predicted 104 protein clusters that were exclusively present in alkane producing *Desulfovibrio* strains. The six protein clusters with a predicted structure likely sharing similar structural features with alkane-binding proteins are written in bold (top of the table). Protein clusters with an unknown function after annotation verification are highlighted in grey. Protein clusters identified to be exclusively present in alkane producing strains but which the consensus sequence aligned to non-alkane producing *Desulfovibrio* protein sequences with an E-value of 0 are written in red (bottom of the table). Alignments with an E-value inferior to $10e^{-5}$ are defined as “no significant hit.



Appendix 6. Detection of an increase in fatty acid quantity within the whole cell protein extracts from wild type *E. coli*, supplemented with 100 μM tetradecyl-d_{29}-phosphate and incubated for 48 h

Proteins from *E. coli* aerobically cultured to the exponential growth phase were extracted and whole cell protein extracts were spiked with 100 μM tetradecyl-d_{29}-phosphate, prior to being incubated for 48 h at 37 °C. Whole cell protein extracts were analysed by GC-MS at the beginning of the incubation (A) and 48 h (B) incubation. The extracted 57 m/z, 71 m/z and 85 m/z ion chromatograms (EIC) from whole cell protein extracts at the beginning of the incubation and at 48 h incubation revealed an increase in fatty acid quantity within the whole cell protein extracts throughout the incubation.



Appendix 7. Detection of an increase in fatty acid quantity within the whole cell protein extracts from *E. coli* transformed with the plasmid pEX1K3, supplemented with 100 μM tetradecyl-d_{29}-phosphate and incubated for 48 h

Proteins from *E. coli* transformed with pEX1K3 which was aerobically cultured to the exponential growth phase were extracted and whole cell protein extracts were spiked with 100 μM tetradecyl-d_{29}-phosphate, prior to being incubated for 48 h at 37 °C. Whole cell protein extracts were analysed by GC-MS at the beginning of the incubation (A) and 48 h (B) incubation. The extracted 57 m/z, 71 m/z and 85 m/z ion chromatograms (EIC) from whole cell protein extracts at the beginning of the incubation and at 48 h incubation revealed an increase in fatty acid quantity within the whole cell protein extracts throughout the incubation.