Investigating *Clostridium saccharoperbutylacetonicum* as a potential chassis for the production of geranyl pyrophosphate-derived biofuels

Thesis for the degree of MRes in Microbiology

School of Biosciences
Faculty of Sciences
University of Kent

Stephen Nicholas Free

2020

Declaration

I confirm that no part of this thesis has been submitted in support of an application for any degree or qualification at either the University of Kent, any other university or higher education learning institution.

Stephen Free

2020

Abstract

This study focuses on *Clostridium saccharoperbutylacetonicum*'s potential in industrial biofuel biosynthesis, owing to *Clostridium*'s established use in butanol, acetone and ethanol production and this species' relative aerotolerance and natural adaptation to utilise a variety of sugars present in industrial feedstocks. Initial experiments aimed to express bacterial microcompartments in *C. saccharoperbutylacetonicum* faced cloning difficulties, precluding the completion of this research. Thereafter, the focus switched to the possibility of using this organism for the production of the biofuels alpha-pinene, linalool and R-limonene, all of which require the metabolite geranyl pyrophosphate as a precursor.

Preliminary genome searches were promising, with genes encoding putative enzymes of the MEP/DOXP pathway linking glycolysis and geranyl pyrophosphate production found. Toxicity of the aforementioned biofuels was then investigated using *E. coli*, where each exhibited significant toxicity, suggesting that a more desirable chassis for their production is desirable for industrial-scale biosynthesis.

To further investigate the putative genes discovered, a combination of homology analyses and structural modelling was used to compare them to known pathway enzymes. Close structural homology was found to bona-fide *E. coli* K-12 enzymes, with substrate docking analysis on both our *Clostridium* models and *E. coli* X-ray crystal structures providing very similar results.

This work suggests that *C. saccharoperbutylacetonicum* may be capable of geranyl pyrophosphate production, and is therefore a good chassis for the introduction of single recombinant enzymes that can produce alpha-pinene, linalool and R-limonene using geranyl pyrophosphate as a substrate.

Acknowledgements

I can consider myself lucky to have received great emotional support from my family and partner – who has also been there to provide opinions on my work.

I would like to thank my Supervisor Mark Shepherd for welcoming me into his laboratory group and for making sure his laboratory was a safe space for us to continue with some limited work during the challenging times faced towards the end of this degree. I am also thankful for his helpful guidance on this project, always finding time to discuss experiments with me.

I would like to thank the other members of the Shepherd Lab group: C. Webster, L Holmes, T. Monaghan, C. Ribeiro, S. Fussaini and O. Martinez Retegi for their continual support and creation of a friendly and safe workspace.

I am grateful to G. Robinson for attending our weekly laboratory meetings and providing a secondary professional perspective on our work, continually showing great insight. S. Blackburn from the Robinson group laid down much of the groundwork for my experiments with *Clostridium* and she was a great help in bringing me up to speed with the techniques required; for this I am extremely grateful.

I would lastly like to thank the excellent Technical Laboratory staff at the University of Kent, particularly R. Akerman, who was always around to help us with any issues with a smile.

Table of Contents

List of Figures

List of Tables

List of Supplementary Material

1.0 Introduction:

1.1 The need for renewable fuels:

Fossil fuels are a finite non-renewable source of energy which produce compounds such as Carbon Dioxide which are damaging to our planet. As such, economically viable renewable energy sources are in ever-increasing demand globally as governments and companies seek to establish themselves as having the people and the planet's best interests in mind.

Electrical energy is often seen as the way forward for many applications, however battery technology is unable to keep up with modern day needs, resulting in poor range when used in vehicles, high cost and heavy weight. It must also be considered as to how the electricity itself has originally been sourced, as only through renewable generation does it truly provide the alternative to fossil fuel sought after.

Biofuels however encompasses a wide variety of energy-rich compounds obtained from renewable biomass feedstocks (1) and do not suffer the same drawbacks as electrical energy – notably functioning as a fuel additive or replacement in existing engine technology with little or no adjustment needed. As such, biofuel could provide a more reasonable alternative, especially in applications where range and weight may be crucial – such as the civilian aeronautical industry (2).

Waste by-products of crop plants are often used as a feedstock due to their wide availability. However, the amount of land required to meet the current fossil fuel requirements far exceeds the supply of crop feedstock available (3) and attempting to meet these demands would have a severe negative impact on food production. Cellulose is often used as a base for biofuel due to its abundancy, however these

"1st Generation" (1G) biofuels are expensive and energy-intensive to produce due to the complexity of cellulose and its hydrolysis (4). As such, biofuels produced using these feedstocks are often only used as additives to fossil fuels. Due to the vast swathes of arable land required for this feedstock and its resulting impact upon not only the food industry but also ecosystems, 1G biofuels have been met with many restrictions and bans from leading institutions such as the EU (5). As such, microbial feedstocks may be a more viable carbon source to meet the current demands for fossil fuel and help limit their harmful effects on the planet. Bacteria along with other microorganisms produce 2nd Generation (2G) biofuels from cellulosic energy crops and lignocellulosic waste as opposed to the food crops used in 1st Generation biofuel production (6). Microalgae are used in the production of 3rd Generation (3G) biofuels from solar radiation. Both 2G and 3G biofuels are more energy-efficient than their 1G counterparts, however require more advanced technology and are often smaller in scale (7).

Despite the ability of Algae to produce biofuels from sunlight (8), bacterial feedstocks may be the more viable option currently due to the issues with managing large-scale Microalgal bioreactors (9) (10).

1.2 Clostridial species used for biofuel production:

Clostridial species are of particular interest in the field of biofuel production due to their wide diversity in metabolic potential, though their anaerobic nature adds complexity to their study and use. Some species such as *C. saccharoperbutylacetonicum* are useful in industrial applications due to the extensive genetic tools available and their capacity to produce a wide range of extremely useful metabolites, such as alcohols and industrially relevant acids such as butyric acid – a common food flavouring. The genetic research thus far carried out, however, has only reached so far as identifying a diverse range of proteolytic

enzymes, yet failed to fully elucidate many of the metabolic pathways they are involved in.

Biofuel-producing *Clostridium* species tend to fall into one of two phylogenetic clades (11) and are often non-pathogenic to humans. However, their use in biofuel biogenesis was greatly reduced in the early 20th century as petrochemical methods became preferred due to reduced cost and availability of fossil fuels globally (12). However, in the current climate, production of these same fuels from renewable resources is beginning to become preferential. This once again opens an avenue for research into the production of biofuels by *Clostridium* and, with renewed interest and funding, the process may be refined into more commercially viable processes.

An ideal *Clostridium* species for use in industrial biofuel production may be *C. acetobutylicum* (13) due to the vast number of studies, metabolomic and genetic profiling which exist, or the lesser known species *C. saccharoperbutylacetonicum* due to its greater tolerance for oxygen exposure. It is noteworthy that a few species (*C. carnis*, *histolyticum*, and *tertium*) can be subcultured for aerotolerance (14), however none of these species have been shown to produce high yields of conventional *Clostridium* biofuels such as acetone, butanol or ethanol. Alternatively, coculture of *C. acetobutylicum* with *Bacillus subtilis* was utilised to maintain an anaerobic medium and remove the costly need for N_2 purging (15).

It may also be possible to create a Clostridial feedstock which consumes Carbon Dioxide as a substrate for butanol production, as acetogenic strains have been shown to fix CO_2 into acetyl-CoA (16), with some even shown to produce butanol from Carbon Monoxide, although these feedstocks typically result in significant decreases in biomass (17).

1.3 ABE fermentation by Clostridial species:

Clostridial biofuel synthesis occurs via ABE fermentation, named after its main products – Acetone, Butanol and Ethanol. Lignocellulosic crop waste, such as Corn Stover (i.e. the waste material left in the field after harvesting of corn crops), has been a popular feedstock in recent history due to its abundance and low cost, and there is high demand to expand the repertoire of biofuels which can be produced from these raw materials in large, stable and reproducible quantities (18) (19). As seen in **Figure 1**, this lignocellulosic biomass first requires a variety of pre-treatments (20) to break down the lignin, which inhibits hydrolytic enzymes from reaching the hemicellulose and cellulose fibres and initiating hydrolysis into glucose and xylose (21). The availability of the biomass used must therefore be weighed against the lignin content, as biomass with a higher percentage content of lignin will require more lengthy and costly pre-treatments before they can be used as a feedstock.

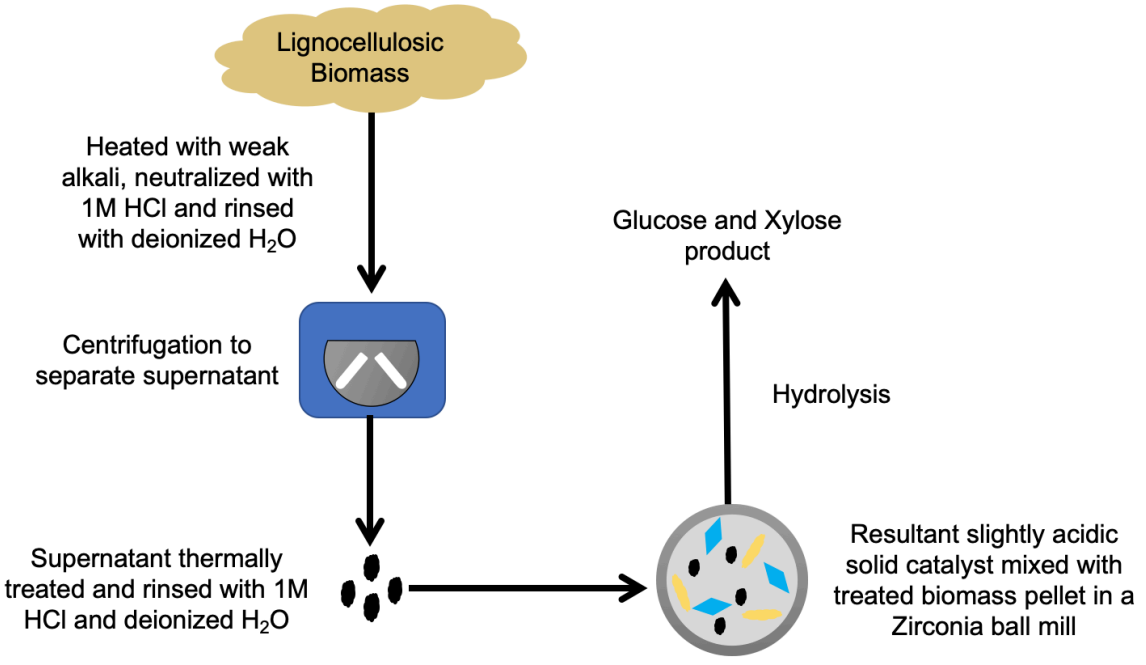


Figure 1: Lignocellulosic biomass pre-treatment for use in ABE fermentation
An example of pre-treatment of lignocellulosic biomass. These processes can be optimised to achieve high glucose yields of approximately 70%.

Hexose sugars are then converted into acetyl-CoA during glycolysis, and xylose (a pentose sugar) is channelled into glycolysis via the pentose phosphate pathway. The resultant acetyl CoA is then used in the acidogenic and solventogenic pathways, detailed below, for the production of acetone, butanol and ethanol.

1.3.1 Acidogenesis

Clostridium species will first enter acidogenesis, the first of two phases. This occurs during exponential cell growth and, during this phase, organic acids – chiefly Acetic and Butyric acid – are produced from acetyl-CoA, ATP is produced (causing exponential cell growth), and alcohol titres are significantly reduced.

The production of these acids reduces the internal pH and eventually reaches a "pH breakpoint" (22). This mechanism is not yet fully understood, however it is believed to trigger the upregulation of solventogenic genes via the master regulator Spo0A which is phosphorylated once the pH breakpoint is reached (23) and the *sol* operon, which may occur simultaneously or independently of sporulation (24).

1.3.2 Solventogenesis

Solventogenesis is a mechanism via which *Clostridium* can avoid acid stress caused by excess acidogenesis and shows a tendency for upregulation once cell growth has slowed. Solventogenesis tends to be induced at pH <4.5, however this varies between different species of *Clostridium* (25). During this phase the organic acids produced by the acidogenic phase are converted to the solvents acetone, butanol and ethanol (26).

The balance between the two phases (**Figure 2**) must be carefully regulated by the cells as the solvents produced are also toxic to the cell (e.g. butanol disrupts the cell membrane).

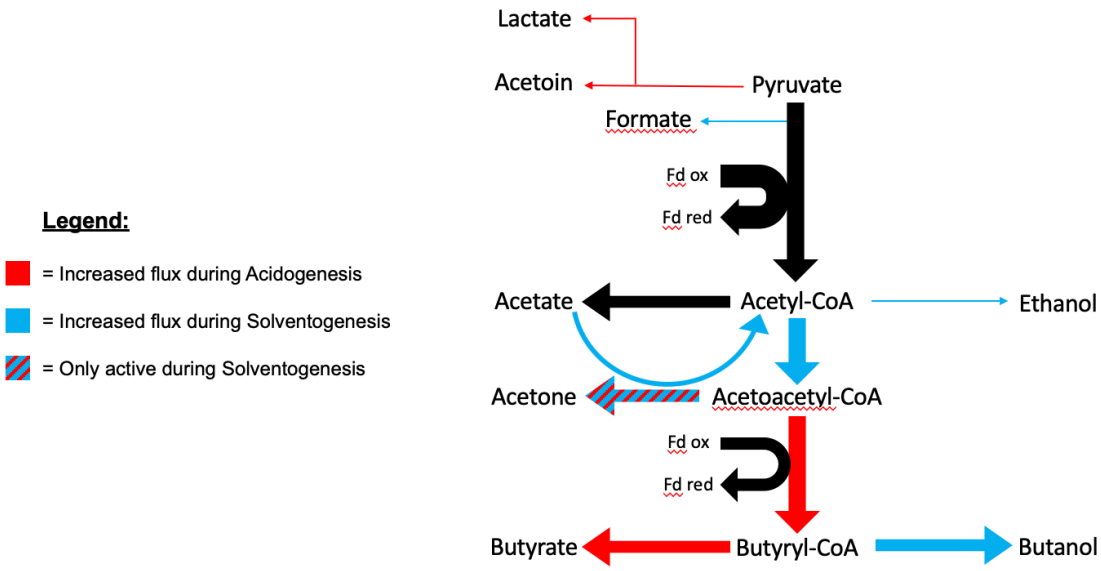


Figure 2: Metabolic flux during switch between Acidogenesis and Solventogenesis
Illustration of the main target pathways affected by the switch between Acidogenesis and Solventogenesis in Clostridia.

1.4 Bacterial Microcompartments and their potential applications in biofuel production:

Bacterial Microcompartments (BMCs) are proteinaceous shells similar in shape to viral capsids (**Figure 3**), two examples being PDU and EUT BMCs which are used in the metabolism of Propanediol and Ethanolamine, respectively. In future, BMCs may be useful for the engineering of strains to rapidly produce biofuels as metabolic intermediates that are toxic to the cell may be sequestered within the BMC (27). Also, substrate channelling via the co-localisation of pathway enzymes may diminish the concentration of toxic intermediates and result in an overall improvement of pathway flux and higher biofuel yields.

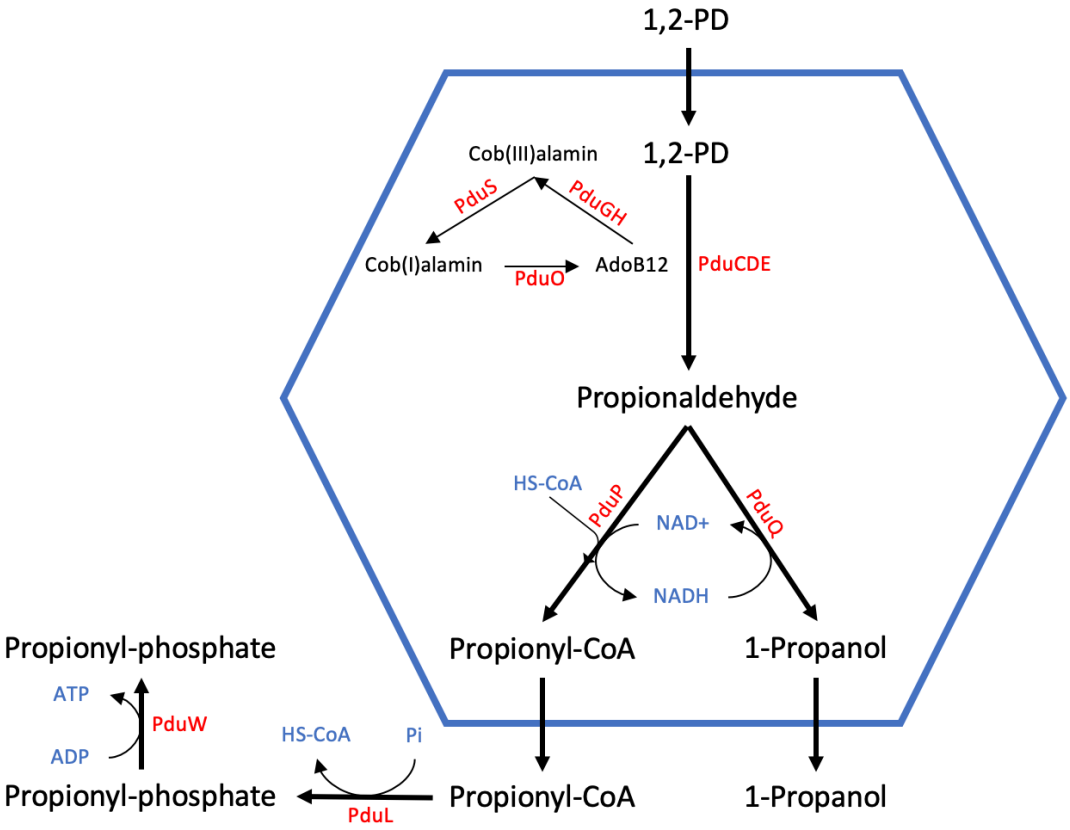


Figure 3: Overview of enzymatic reactions within a PDU BMC
A cross-sectional representation of an Propanediol-Utilising (PDU) bacterial microcompartment, complete with its metabolic pathway.1,2-PD represents 1,2-Propanediol.

BMCs are formed from a large group of proteins, each with specific functions (**Table 1**).The icosahedral cage of BMCs is constructed of three main classes of shell proteins. Cyclical hexagons made from a class of major shell proteins known as BMC-H form the main facets of this cage, containing small central pores (28). BMC-P is a class of minor shell proteins – such as PduN – which cap the vertices of the BMC structure and are crucial for the formation of Propanediol Utilisation (PDU) microcompartments, with its deletion preventing cage formation (29). The third class of BMC proteins, BMC-T, is responsible for the assembly of hexagonal trimer groups with a large central pore, acting as shell permeability modulators (28) (30).

Table 1: *pdu* operon gene functions
Functions of each protein present in the *pdu* operon of *C. freundii*.

***Citrobacter freundii* Propanediol Utilisation Proteins**	
PduA	Hexamer Facet Component
PduB	Hexamer Facet Component
PduC	Ado-B_{12}-dependent Diol dehydratase large subunit
PduD	Ado-B_{12}-dependent Diol dehydratase medium subunit
PduE	Ado-B_{12}-dependent Diol dehydratase small subunit
PduF	Propanediol diffusion facilitator
PduG	Diol dehydratase reactivation protein
PduH	Diol dehydratase reactivation protein
PduJ	Hexamer Facet Component
PduK	Hexamer Facet Component
PduL	Phosphotransacylase
PduM	Unknown
PduN	Pentameric Vertices
PduO	Cobalamin adenosyltransferase
PduP	CoA-dependent propionaldehyde dehydrogenase
PduQ	1-Propanol dehydrogenase
PduS	Cobalamin reductase
PduT	Hexamer Facet Component
PduU	Hexamer Facet Component
PduV	Unknown
PduW	Propionate kinase
PduX	L-threonine kinase

The possibility of BMC expression being linked to acidogenesis in *Clostridium kluyveri* has been postulated due to the proximity of BMC genes (*eutM/L* orthologues – themselves orthologues of *pduA/B*) to NAD-dependent dehydrogenases (31). Furthermore, BMCs have been linked to solventogenesis in

Vibrio furnissii by the discovery of functional alcohol dehydrogenase genes flanked by *pduJ/ccmO* orthologues (32). This suggests that there may be some precedent for the use of microcompartments to boost industrial biogenesis of butanol.

1.5 Metabolic intermediates in *C. Saccharoperbutylacetonicum* as potential biofuel precursors:

Of specific interest was the pathways that produce geranyl pyrophosphate (GPP), a precursor to several biofuels including geraniol, alpha-pinene, linalool and R-limonene. Linalool is a precursor to jet fuel that is in high demand and would benefit greatly from a bacterial chassis capable of producing it in high yields (33). Other products such as Limonene and Geraniol are also in continuous high demand due to their use in fragrance manufacture, and also have potential uses as components of fuels, although current production processes do not produce high yields (34) (35). Cyclic and bicyclic monoterpenoids also feature within this pathway, such as alpha-Pinene. These are much more fluid at lower temperatures and can therefore make them ideal as components in biodiesel fuel used at lower temperatures (36).

The production of these biofuels is tied closely to the biosynthesis of isoprenoids, a varied class of molecules crucial to cellular survival in bacteria, which attracts great interest in the pathway enzymes as targets for antimicrobials to kill pathogenic bacteria (37). With this in mind, it is surprising that little research has been carried out on the Terpenoid pathways, particularly in *Clostridium*.

GPP can be synthesised via the Mevalonate (MVA) or Non-Mevalonate MEP/DOXP (less commonly known as DXP) pathways (**Figure 4**). The ispH enzyme within the MEP/DOXP pathway functions to produce both isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) (38) (39), allowing for this pathway to

produce GPP without depending on the co-functionality of the MVA pathway. The MVA pathway has been shown to be less efficient at producing GPP from Glucose than the MEP/DOXP pathway (40) and yields a ratio of 3 moles IPP for every 7 moles of DMAPP, whereas the MEP/DOXP pathway yields a 5:1 ratio. However, it should be noted that the MEP/DOXP pathway requires an extra mole of NADH and two moles of ATP, requiring more energy and cellular reducing power (40).

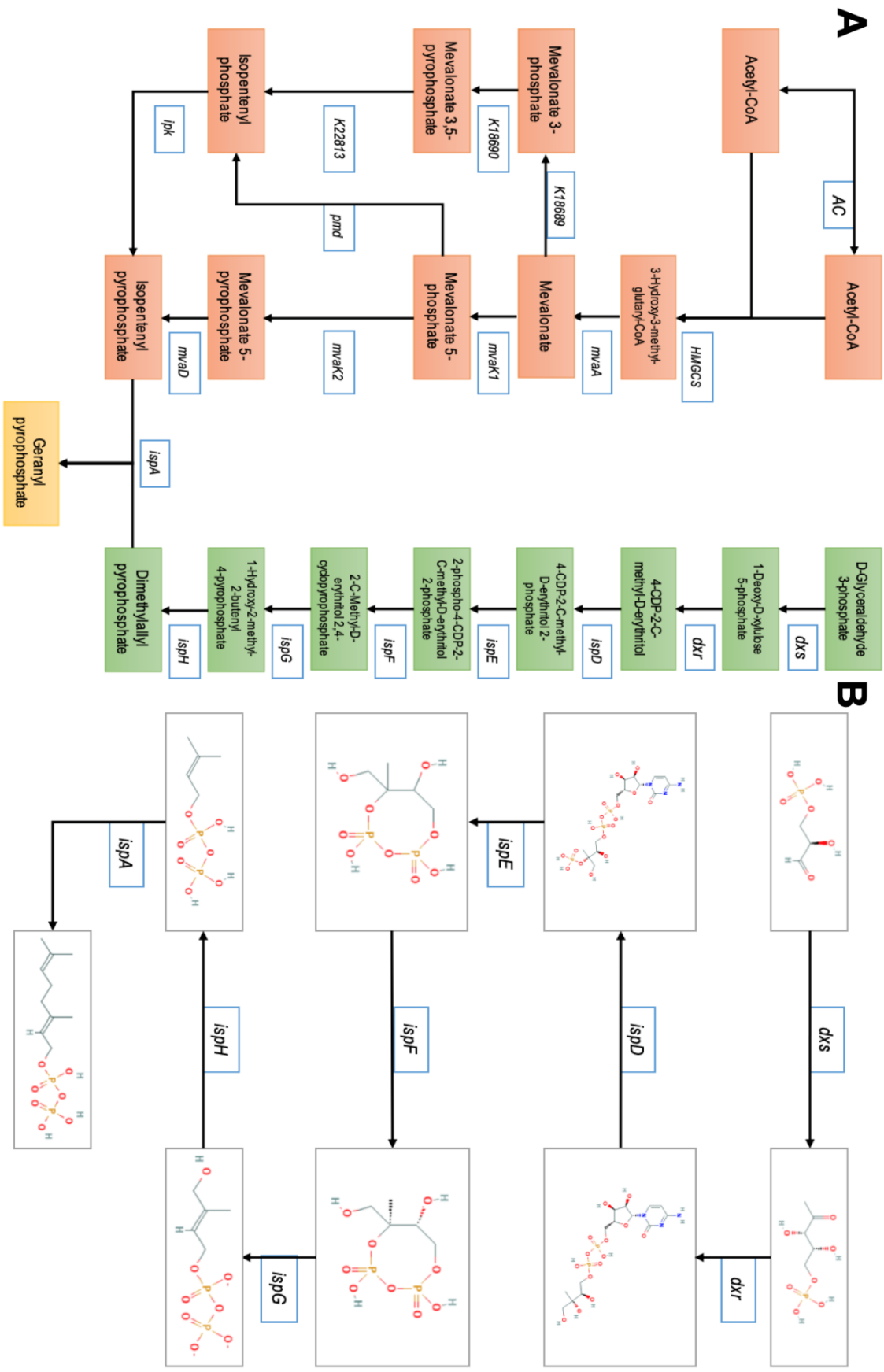


Figure 4: Production of GPP by the MVA and MEP/DOXP pathways
(A) The MVA and MEP/DOXP pathways which convert acetyl-CoA from glycolysis into Geranyl-PP – subsequently feeding into many crucial metabolic pathways such as ubiquinone and, crucially, monoterpenoid biosynthesis. **(B)** Structural representation of the MEP/DOXP pathway.

As the enzyme ispA (**Figure 4**) requires a 1:1 ratio of each of these compounds to produce GPP, this would suggest that the MVA pathway is better suited to the production of certain biofuels and precursors found downstream in the Monoterpenoid Biosynthesis pathway. However, the expression of the MVA pathway in eubacteria is rarer than the MEP/DOXP pathway, although exceptions are noted, with both pathways having been observed within Listeria monocytogenes (41) and an unclassified species of Streptomyces known as Strain CL190 (42).

Yu V. Ershov hypothesised that, despite most heterotrophic microorganisms tending towards use of the MVA pathway, anaerobic heterotrophs such as *C. saccharoperbutylacetonicum* are more likely to utilise the MEP/DOXP pathway as overall energy efficiency is more important to their viability than the slight carbon losses involved in the MVA pathway (43). This is further facilitated by their capability to regenerate pyridine nucleotides through mechanisms favoured by a greater availability of $NAD(P)^-$, generated by the consumption of H^+ at the dxs, IspD and IspG-mediated steps, however it should be noted that NADPH is generated by dxr and H^+ by dxr and IspH.

C. ljungdahlii is known to possess the MEP/DOXP pathway endogenously (44) (45), though evidence is severely lacking for the activity of this pathway within Clostridia. Mevalonate synthesis has been induced in the anaerobe *C. ljungdahlii* via transformation of the *E. faecalis* genes *mvaE and mvaS*, among other heterologous genes to complete the lower mevalonate pathway (45), however little is known overall regarding the MEP/DOXP pathway within Clostridial species and far less within *C. saccharoperbutylacetonicum*. Indeed, there is no literature on the MEP/DOXP pathway in *C. saccharoperbutylacetonicum*, so the current project sought to investigate this gap in knowledge further.

1.6 Aims and experimental strategy

The initial aim of this work was to establish a cloning strategy for the introduction of PDU BMC shell protein-encoding genes into *C. saccharoperbutylacetonicum*, and to demonstrate their expression via the use of a translational fusion of *pduA* linked to *mCherry* (46).

The next main aim was to establish a protocol via which the bacterial hosts could be assessed for this tolerance to a range of biofuels that are produced from geranyl pyrophosphate.

The final aim was to use homology approaches and structural modelling to determine whether there is an intact MEP/DOXP pathway in *C. saccharoperbutylacetonicum.* Putative MEP/DOXP genes found in *C. saccharoperbutylacetonicum* would be compared to bona-fide *E. coli* counterparts in order to ascertain whether they are likely to catalyse the same reactions.

2.0 Materials and Methods:

2.1 Media:

Rich Clostridia nutrient Medium (RCM) consists of 33.0 g Oxoid RCM powder, dissolved in MiliQ H_2O to a total volume of 1 L and then autoclaved.

2xGCM consists of 5.0 g Yeast Extract, 0.75 g Dipotassium Phosphate, 0.75 g Monopotassium phosphate, 0.4 g Magnesium sulphate, 0.01 g Iron sulphate, 0.01 g Manganese sulphate, 1 g sodium chloride, 2 g ammonium sulphate and 2 g Asparagine. These were dissolved into MiliQ H_2O to a total volume of 500 mL. This medium was then autoclaved.

1xGCM was prepared by adding 500 mL of autoclaved 10% Glucose solution to 500 mL autoclaved 2xGCM Medium.

Electroporation Buffer with Salt (EPB_S) consists of 5.13 g Sucrose (300 mM), 101.65 mg Magnesium Chloride hexahydrate (10 mM), 30.36 mg Monosodium Phosphate monohydrate (4.4 mM), 8.04 mg Sodium Phosphate dibasic heptahydrate (0.6 mM).

Electroporation Buffer without Salt (EPB_NS) consists of 5.13 g Sucrose (300 mM), 30.36 mg Monosodium Phosphate monohydrate (4.4 mM), 8.04 mg Sodium Phosphate dibasic heptahydrate (0.6 mM).

The Sucrose and MgCl (Sucrose alone for EPB_NS) were dissolved in MiliQ H_2O to a total volume of 25 mL and autoclaved. The phosphate stocks were dissolved in MiliQ H_2O to a total volume of 25 mL and filter sterilised through 0.2 µm pores. The two stocks were then combined to create 50 mL stocks of EPB_S/EPB_NS.

2.2 *E. coli* Growth Curves:

A colony grown on an agar plate was picked and placed into a universal tube containing sterile RCM to grow overnight in a shaken water bath at 37 °C, 180 RPM. 3 mL of overnight culture was then added to 27 mL autoclaved RCM in an anaerobic serum bottle (1:10) for each repeat and grown in a 37 °C static incubator for the duration of the experiment, only briefly removed for sampling and biofuel addition.

2.3 Colony PCR:

Single *E. coli* colonies were picked from agar plates grown overnight at 37 °C and then suspended in a pre-assembled 2X NEB Q5 PCR reaction mix. These were then placed into a Biometra T3000 Thermocylcer following variations of the protocol: [94 °C 5 min, (94 °C 15 s, ANN 30 s, 72 °C EXT)x35, 72 °C 5 min, 10 °C ∞] Where ANN = primer annealing temperature and EXT = extension time.

Primers:

Table 2: Table of primers used during PCR
Primers used to attempt cloning of *pduABJKNU* from a PlysS vector into a PMTL 8000 series vector. Primers 1-2 were designed to produce *pduABJKNU* PCR fragments with overhangs which could be used in Gibson assembly. Primers 3-4 were designed for PMTL PCR vector amplification. Primers 5-6 were designed for *pduABJKNU* PCR, producing fragments with overhangs containing restriction sites for subsequent conventional cloning.

Primer No.	Oligonucleotide Name	Sequence (5'-3')	Scale [µMol]	Purification
1	mcpdu_F_83353	TTTTTAAGGAGGTGTGTTACATATGG TGAGCAAGGGCGAGGAGGATA	0.025	Standard Desalting
2	mcpdu_R_83353.84422	GCAGGCTTCTTATTTTTATGCTAGCTT ATGTCCGGGTGATGGGACAGGCG	0.025	Standard Desalting
3	374	GCTAGCATAAAAATAAGAAGCCTGC	0.025	Standard Desalting
4	375	CATATGTAACACACCTCCTTAAAAATTAC	0.025	Standard Desalting
5	mcherry_AseI_F	CCCATTAATATGGTGAGCAAGGGCGAG GAGGATAACATG	0.025	Standard Desalting
6	pduU_NheI_R	CCCGCTAGCTTATGTCCGGGTGATGGG ACAGGCG	0.025	Standard Desalting

Vector Miniprep:

QIAprep® Spin PCR Miniprep Kit and Protocol used, with room temperature MiliQ H_2O used in place of elution buffer (Buffer EB) with an extra spin cycle due to subsequent need for electroporation and to maximise DNA yields.

PCR purification:

QIAquick® PCR Purification Kit and Protocol used, with room temperature MiliQ H_2O used in place of elution buffer (Buffer EB) with an extra spin cycle due to subsequent need for electroporation and to maximise DNA yields.

Transformation:

5 mL of an overnight culture of *C. saccharoperbutylacetonicum* was inoculated into 45 mL 1xGCM within a sealed anoxic serum bottle. This was grown at 32 °C until it reached an OD_{600} of over 1.2. This was then placed on ice for 3 mins, with 20 mL decanted into each of two 50 mL Falcon tubes inside the anoxic chamber.

The Falcon tubes were then centrifuged at 4000xg at 4°C for 10 mins, transferred into the anoxic chamber, supernatant discarded and the pellets resuspended in 20 mL EPB_S.

The centrifugation step was then repeated, with the resultant pellets this time resuspended together in 1 mL EPB_NS inside the anoxic chamber.

Electroporation:

200 µL of the cell resuspension was then added to two Electroporation Cuvettes and sealed before removal from the anoxic chamber. One electroporation Cuvette contained 1000 ng of DNA (3.20 µL) and the other contained the equivalent volume

of MiliQ H_2O to act as a negative control. Each was chilled on ice for 5 mins before electroporation at 1.5 kV, after which 1 mL 1xGCM was added to each cuvette to allow the cells to recover overnight within the anoxic chamber at 32 °C.

2.4 Structural Modelling:

E. coli X-ray crystal structures were chosen from Uniprot (47) based on experimental evidence being available at the protein level for the proteins Dxs (48), Dxr (49), IspD (50), IspE (51), IspF (52), IspG & IspH (53) and IspA (54). Amino acid sequences for *C. saccharoperbutylacetonicum* protein genes were inputted into the Phyre2 web server (55) and modelled against the top matching bona-fide counterpart protein. This created a .pdb file which was opened in Pymol (56).

2.5 Substrate docking:

Molecular graphics and analyses were performed with UCSF Chimera developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311 (57). The docking programme was performed using Autodock Vina's (58) Chimera plugin and the binding sites captured chosen by their hybrid (empirical and knowledge-based) scoring system (58). Hydrogen bonding was also noted where possible as a measure of binding strength.

Docking was performed on both the predicted *C. saccharoperbutylacetonicum* models and the *E. coli* X-Ray crystal structures to act as a control. *Protein* models were coloured in straw and ligands coloured by element (Carbon – green, Hydrogen – white, Nitrogen – blue, Oxygen – red, Sulphur – orange).

3 Results:

3.1 Cloning by PCR and Restriction Digests:

Initially, we attempted to clone *mCherryPduA-U* (46) into the *Clostridium* shuttle vector pmtl83353 (59) using Gibson Assembly. This would act as a proof of concept that recombinant BMCs could be expressed in *C. saccharoperbutylacetonicum*. Subsequent overexpression of the microcompartments in *C. saccharoperbutylacetonicum* with the mCherry tag would then allow visualisation via fluorescence microscopy. To this end, primers were designed to amplify the *pduA-U* insert with overhangs complementary to our shuttle vectors, and linear fragments of the shuttle vectors were amplified that could be compatible for Gibson assembly with the *pduA-U* insert. Amplification of the *pduA-U* insert generated a low yield of PCR product.

Modification of the annealing temperature and extension time did not improve this, and the development of a touchdown PCR programme also failed to increase the yield. It was not possible to amplify cloning fragments for the two shuttle vectors, so restriction enzymes were used to digest the shuttle vectors with NdeI and NheI as a contingency plan to generate the desired cloning fragments. The *pduA-U* insert however contained no enzymatic restriction site at the *mCherry* end, so a new mcherry_AseI_F primer was designed to amplify this fragment with a new AseI restriction site that would produce a sticky end compatible with the NdeI sticky ends found in our shuttle vectors. The reverse primer for the *pduA-U* insert contained an NheI restriction site. Analysis of the AseI/NheI *pduA-U* fragment (generated by PCR) and the pMTL83353 NdeI/NheI fragment (generated by restriction digestion) is shown in **Figure 5**.

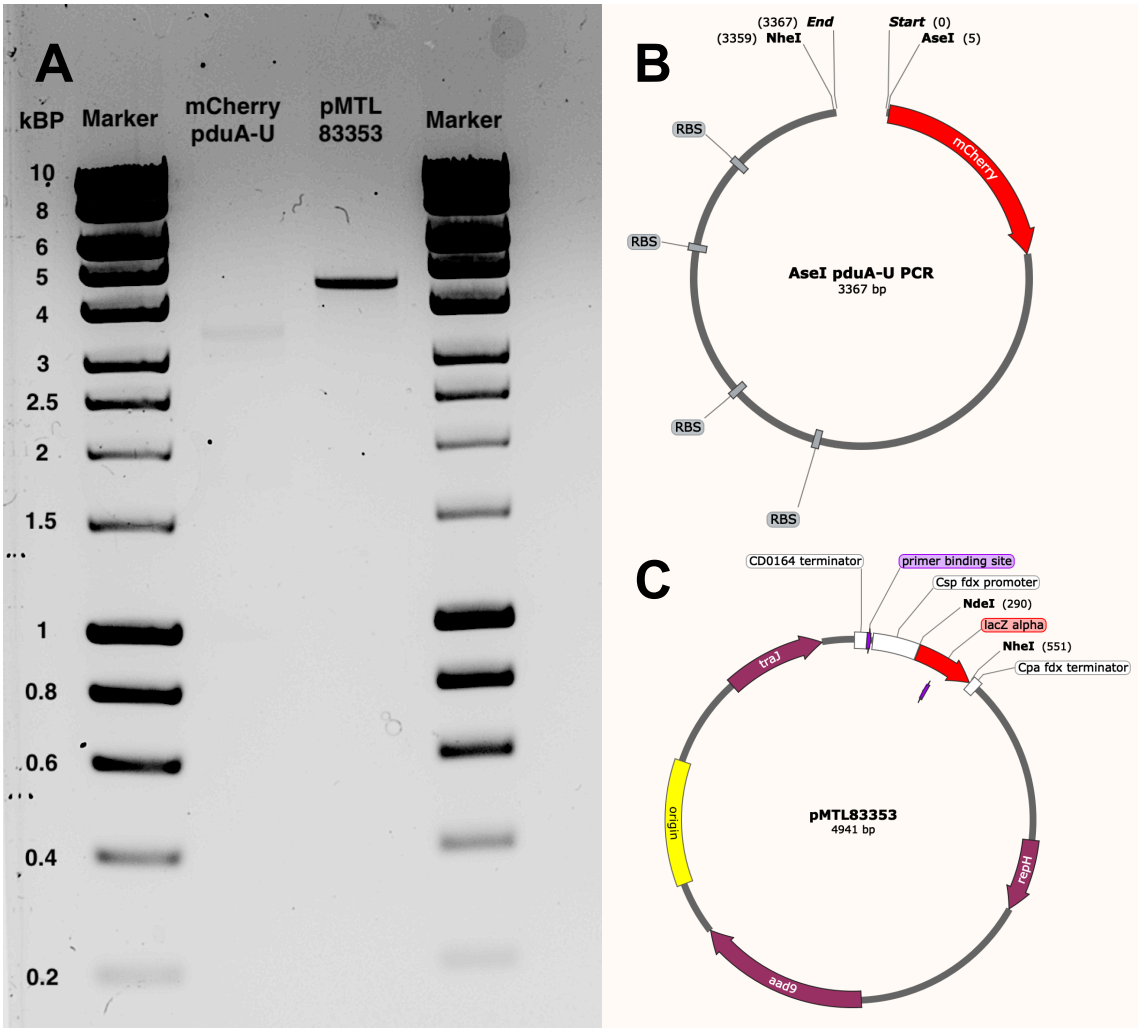


Figure 5: Results of *pduABJKNU*/pMTL Vector PCR experiments
(A) The desired insert band at 3.367 kBp (centre-left) remains faint. Subsequent repeats sequentially eluting more PCR samples did little to improve this result. Our double digest of pMTL83353 with NdeI and NheI-HF however, produces a clear band at 4.941 kBp. **(B)** Vector map of our desired *mCherryPduA-U* PCR product to each end of which our primers mcherry_AseI_F and pduU_NheI_R would anneal (**Table 2**). **(C)** Vector map of our desired pMTL83353 fragment, with the fragment containing LacZ alpha, removed during the double digest by incubation with NdeI and NheI, highlighted in bright red.

Though the restriction digest of our shuttle vector produced clear bands, once again the PCR amplification of the *pduA-U* insert did not produce a yield of DNA to provide sufficient digested insert for feasible ligation, even after sequential elution and gel extraction.

3.2 Establishing a biofuel susceptibility assay using *E. coli*:

As the BMC cloning work was unsuccessful, the project changed focus to investigating the toxicity of various biofuel towards bacteria. Preliminary analysis of the *C. saccharoperbutylacetonicum* genome revealed the presence of putative pathway enzymes for the production of geranyl pyrophosphate, a precursor for the biofuels geraniol, alpha-pinene, linalool and R-limonene. Further analysis of these pathway enzymes is performed in the next section, but this section deals with the ability of bacteria to tolerate these biofuels. The initial aim was to perform biofuel tolerance work for both *E. coli* and *C. saccharoperbutylacetonicum*, but time restrictions precluded the completion of this work for *C. saccharoperbutylacetonicum*. Hence, this section reports toxicity data for *E. coli* that can be used for future comparisons with *C. saccharoperbutylacetonicum*.

A range of industrially useful GPP-derived biofuels (at 5 % v/v final concentration) were tested for their toxicity against *E. coli* using growth curves. Cells were grown anaerobically in 30 mL serum bottles containing sterile RCM to allow for comparison to a future repeat using *C. saccharoperbutylacetonicum*. Biofuels were added when the cells had reached an OD_{600} of ~0.300, and the OD_{600} was then measured every hour (**Figure 6**). These data indicate that, whilst all biofuels displayed significant toxicity compared to the negative control, the GPP-derived biofuels (alpha-pinene and R-limonene) display this effect even more so.

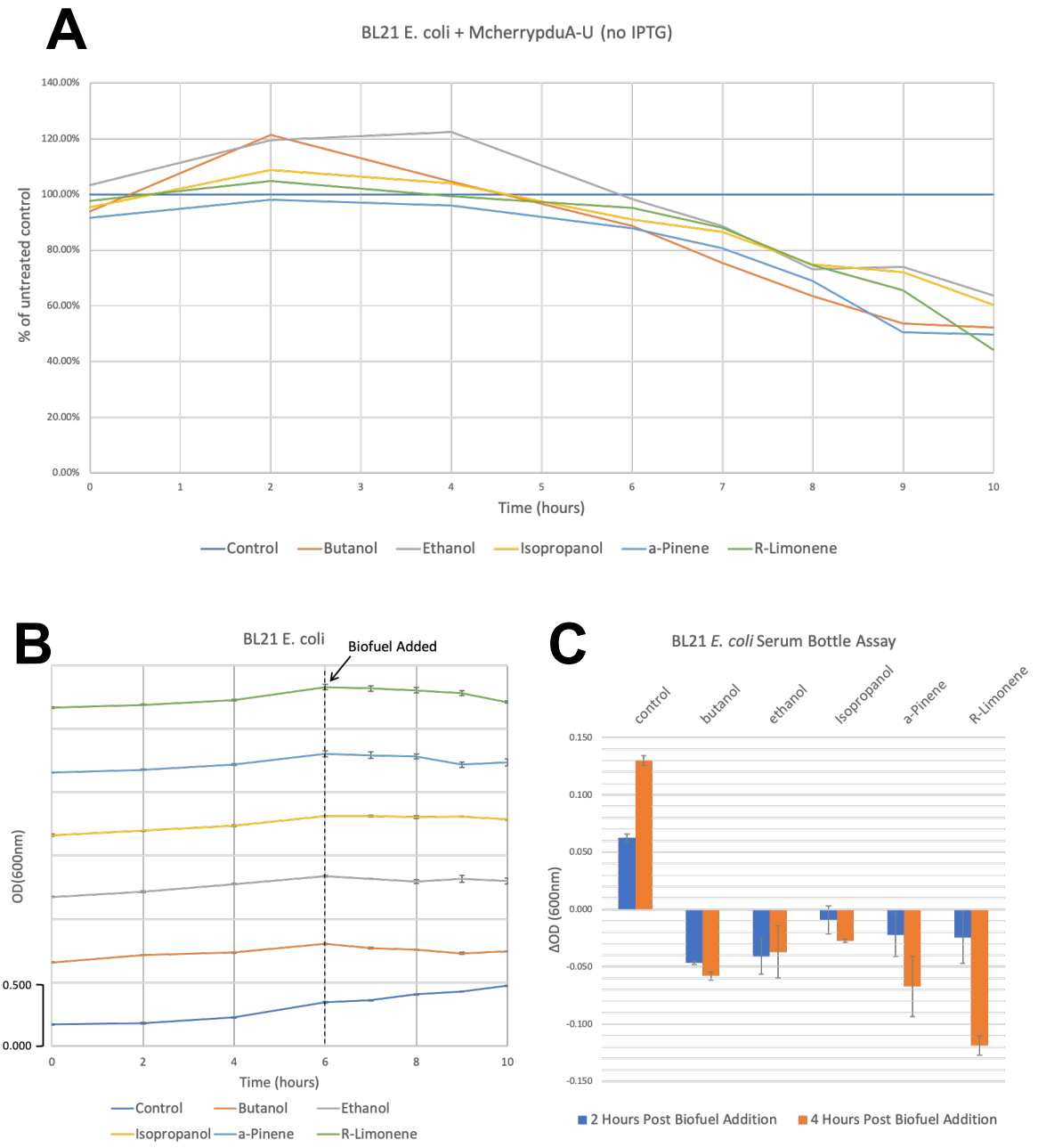


Figure 6: BL21 *E. coli* biofuel media-exposure growth assay
(A) Growth curves for our *E. coli* BL21 grown in serum bottles with different biofuels at concentration of 5 % v/v. Growth data is represented as a percentage of the untreated control. **(B)** Raw OD (600 nm) data used for **A** with offset curves and error bars. Each gridline on the y-axis represents 0.500 nm as is seen in the scale bar found in the bottom-left. **(C)** The mean change in Optical Density (measured at 600 nm) seen after 2 and 4 hours of exposure to the biofuels. Measurements were recorded in triplicate and averaged. Error bars show standard deviation.

Since growth curves provide a measure of both bacteriostatic and bactericidal effects, viability assays were then undertaken to assess biofuel killing after four hours of exposure to concentrations of biofuel across a four log range. For this experiment, three of the GPP-derived biofuels (alpha-pinene, linalool and R-limonene) were analysed alongside three well-studied alcohols that have been

previously produced in *Clostridium* (butanol, ethanol and isopropanol). The physical properties of these biofuels are shown in **Table 3**.

Table 3: Chemical properties of biofuels used in media-exposure and viability assays
Solubility data for all alcohols/biofuels used in the viability assays. LogP represents the log of the biofuels' partition coefficients.

Compound	Density (g/ml)	MW (g/mol)	LogP	Solubility (g/l)	Soluble percentage of 5% v/v (%)
Butanol	0.81	74.12	0.88	68	100
Ethanol	0.789	46.07	-0.31	approx. 100	100
Isopropanol	0.785	60.1	0.05	approx. 100	100
alpha-Pinene	0.858	136.23	4.83	2.5×10^{-3}	4×10^{-5}
R-Limonene	0.842	136.23	4.57	Insoluble	0
Linalool	0.868	154.25	2.97	1.59	2.5

5% v/v = 50ml/l = approx. 62.5g/l

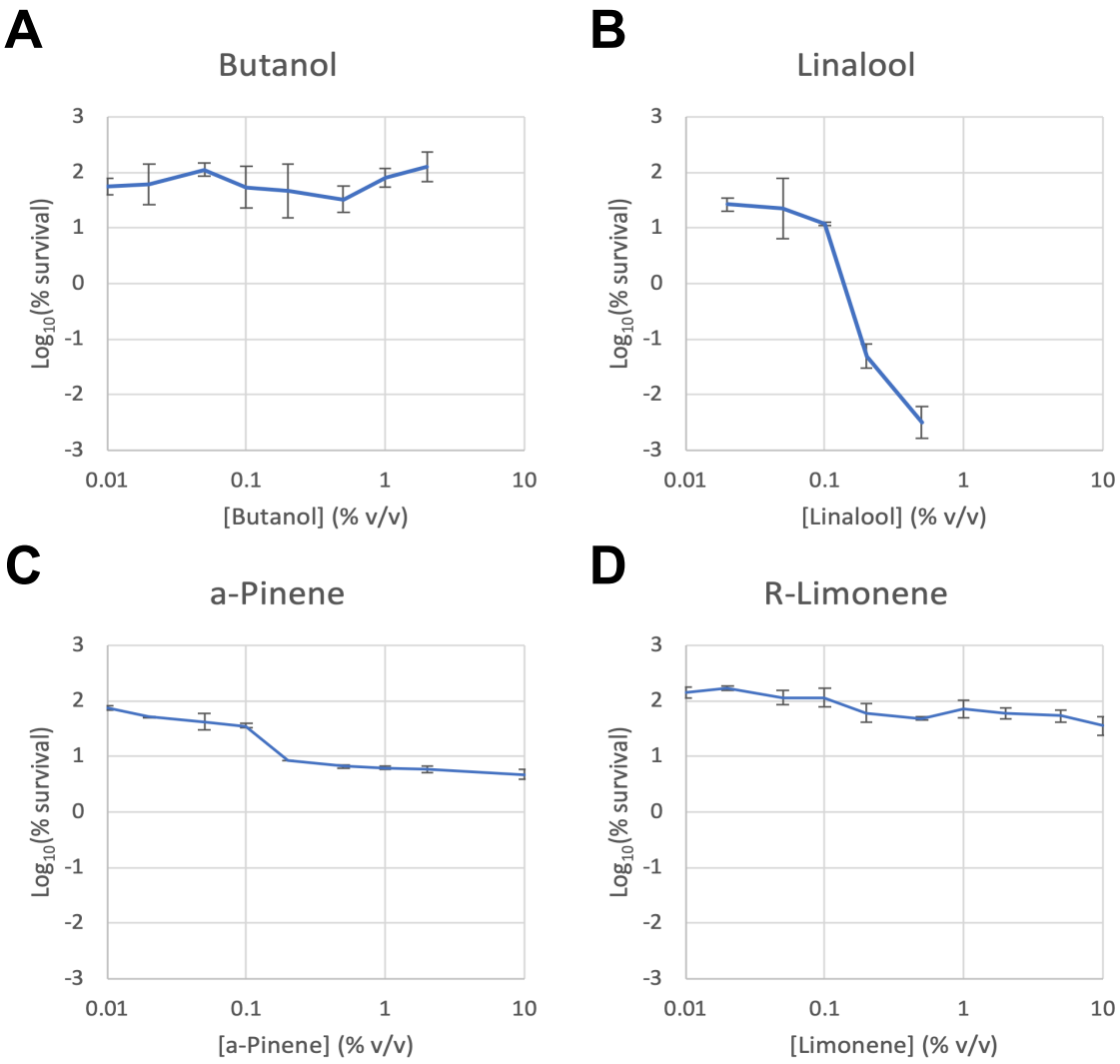


Figure 7: Double log graphs for BL21 *E. coli* biofuel viability assay
Double Log graphs illustrating the results of an *E. coli* viability assay with four different biofuels – **(A)** Butanol shows very little toxicity before suddenly preventing colony growth. **(B)** Linalool shows high levels of toxicity to *E. coli*, preventing colony formation at lower concentrations than our other three biofuels tested. **(C)** alpha-Pinene demonstrates moderate toxicity with a typical reverse sigmoid graph. **(D)** R-Limonene shows some toxicity with a less clear reverse sigmoid trend than is seen in **C**. Measurements were taken from two biological repeats in triplicate technical repeats.

The viability data shown in **Figure 7** shows that Linalool is significantly more toxic to *E. coli* BL21 than other biofuels, with no growth observed above a concentration of 0.5 % v/v (**Figure 7B**). Butanol was not toxic up to concentrations of 2 % v/v, above which no growth could be observed. Alpha-pinene (**Figure 7C**) and R-limonene (**Figure 7D**) both show significantly higher toxicity to *E. coli* than butanol, with alpha-pinene more toxic at lower concentrations than R-limonene, however they do not have as significant an effect on cell death as linalool. Images of colony plates produced can be found in **Appendix A-1**.

3.3 MEP/DOXP Pathway Enzyme Structural Modelling:

In order to identify genes encoding putative enzymes of the MEP/DOXP pathway, BLASTp searches were undertaken on the *C. saccharoperbutylacetonicum* genome using bona-fide MEP/DOXP pathway enzymes from *E. coli* K-12 as input sequences. Multiple sequence alignments showed that the identified protein sequences found therein share a high level of sequence similarity to the *E. coli* genes (**Appendix A-2**). ispH demonstrated a long mismatched region in the second half of its protein sequence. The second, unmatching, half of the ispH protein was put therefore through an additional BLASTp search to attempt to elucidate its function. From this search, a match was found with *E. coli* 30S ribosomal protein S1. This is reasonable as ispH can form a 4-hydroxy-3-methylbut-2-enyl diphosphate reductase/30S Ribosomal Protein bifunctional enzyme. This region was omitted from subsequent modelling.

Table 4: Multiple sequence alignment results for MEP/DOXP genes found in *C. saccharoperbutylacetonicum*

Multiple sequence alignment results between the homology-inferred protein sequences found in the genome mine of *C. saccharoperbutylacetonicum* and bona-fide enzymes from *E. coli* K12. The E-value represents the chance of receiving a random hit in the BLASTp results.

Enzyme	Sequence identity (%)	E Value
dxs	28.94	3e-89
dxr	41.39	6e-107
IspD	30.26	3e-31
IspE	28.67	2e-29
IspF	57.24	2e-59
IspG	49.13	3e-126
IspH	29.55	1e-40
IspA	43.25	1e-58

To gain greater insights into whether the putative pathways enzymes from *C. saccharoperbutylacetonicum* are likely to catalyse the same reactions as the *E. coli* enzymes, structural modelling was performed using the Phyre2 server (55). All

structural models were compared to X-ray crystal structures of the bona-fide *E. coli* enzymes (**Figure 8**). Very close structural homology can be observed in these models between the *E. coli* and *C. saccharoperbutylacetonicum* structures, and where available the cofactors from the crystal structures fitted well in the structural models. However, the *C. saccharoperbutylacetonicum* structural models for *dxs* and *ispD* have loops that cover the active sites.

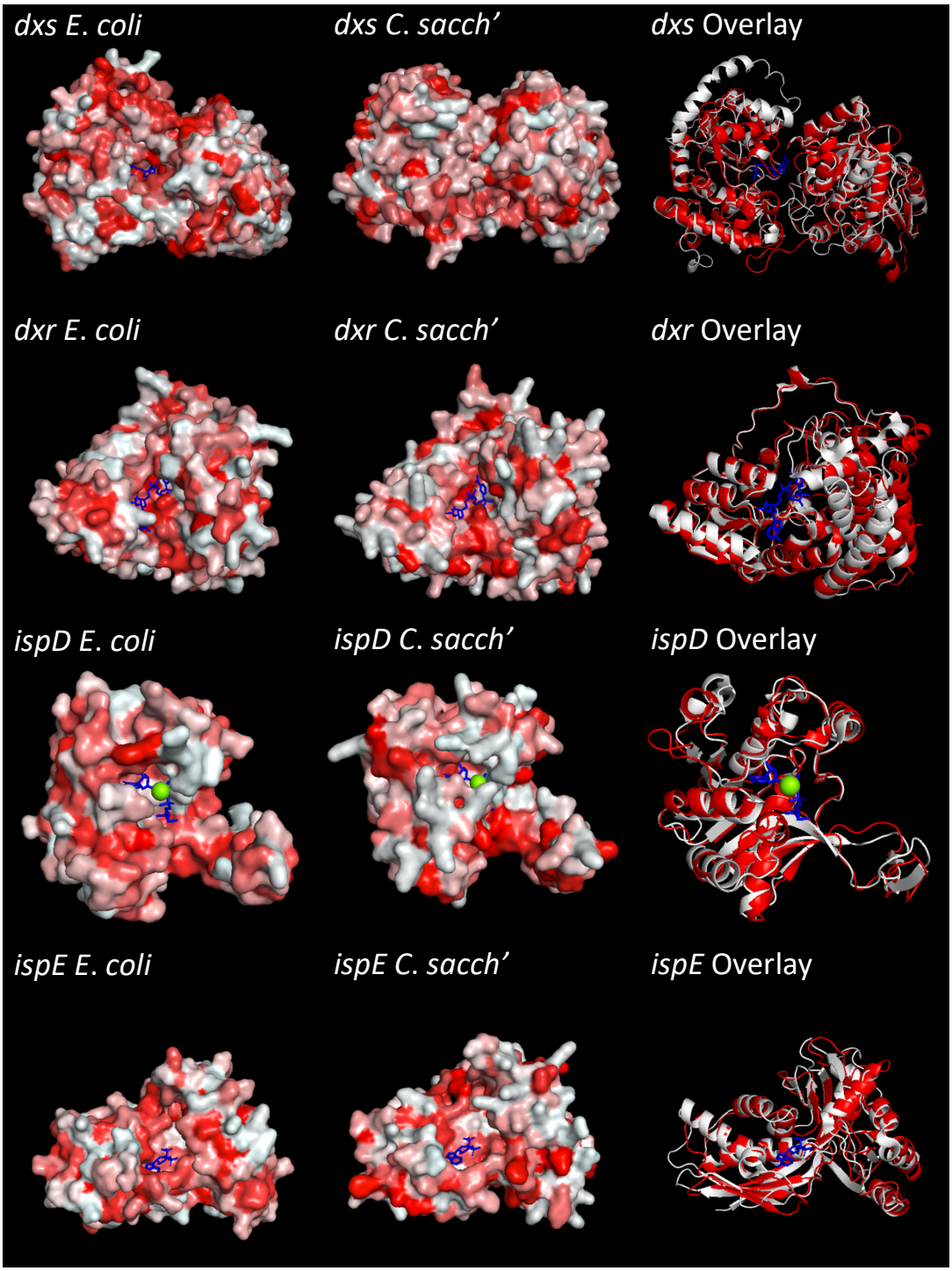


Figure 8: Tertiary structure models predicted from MEP/DOXP genes found in *C. saccharoperbutylacetonicum*

Comparison of structural models for *C. saccharoperbutylacetonicum* MEP/DOXP pathways enzymes with crystal structures from bona-fide pathway enzymes from *E. coli*. (**Left)** *E. coli* X-Ray Crystallography structures visualised in Pymol. **(Center)** *C. saccharoperbutylacetonicum* models were made using the Phyre2 software, modelled against the closest-matching bona-fide enzyme entry. Surface models were coloured using the "color_h" hydrophobicity script in Pymol. **(Right)** Cartoon backbone overlay of *E. coli* in **white** *and C. saccharoperbutylacetonicum* in **red**. Substrates, where available, are represented in **blue** and corresponding cofactors are **coloured by element**. (55)

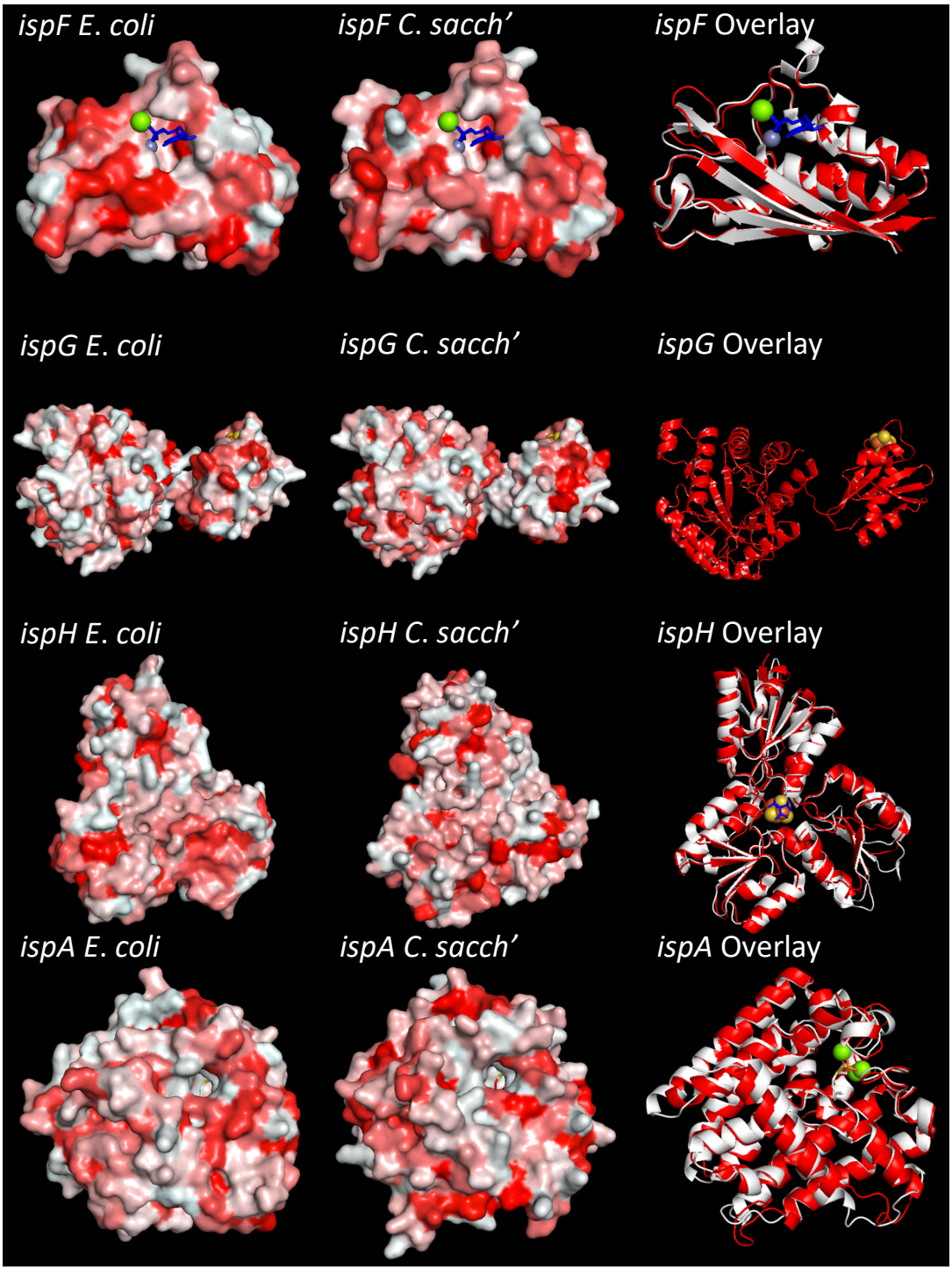


Figure 8 continued: Tertiary structure models predicted from MEP/DOXP genes found in *C. saccharoperbutylacetonicum*

Comparison of structural models for *C. saccharoperbutylacetonicum* MEP/DOXP pathways enzymes with crystal structures from bona-fide pathway enzymes from *E. coli*. (**Left)** *E. coli* X-Ray Crystallography structures visualised in Pymol. **(Center)** *C. saccharoperbutylacetonicum* models were made using the Phyre2 software, modelled against the closest-matching bona-fide enzyme entry. Surface models were coloured using the "color_h" hydrophobicity script in Pymol. **(Right)** Cartoon backbone overlay of *E. coli* in **white** *and C. saccharoperbutylacetonicum* in **red**. Substrates, where available, are represented in **blue** and corresponding cofactors are **coloured by element**. (55)

3.4 Docking substrates into structural models for putative MEP/DOXP pathway enzymes from *C. saccharoperbutylacetonicum*:

To further explore the possibility that the enzymes encoded within the genome of *C. saccharoperbutylacetonicum* are fully functional, the docking software AutoDock Vina (58) was used to dock pathway intermediates (**Figure 4**) to the structural models produced in **Figure 7** in addition to the X-ray crystallography structures from *E. coli* to allow comparison.

Figure 9 suggests that close similarity is found in the binding pocket locations of both *E. coli* and *C. saccharoperbutylacetonicum*, with the predicted binding sites matching up closely between them. *Dxs, Dxr and IspH* were opened in Pymol (56) to allow for a clear cutaway, visualising the substrate binding pocket. Only part of IspG is depicted to allow for a clearer view of the ligand. Full Versions of which are available in **Appendix A-3**. Although small variation can be observed between the two sets of data, it is clear that much of the binding pocket's integrity appears to be preserved in our predicted protein structure models.

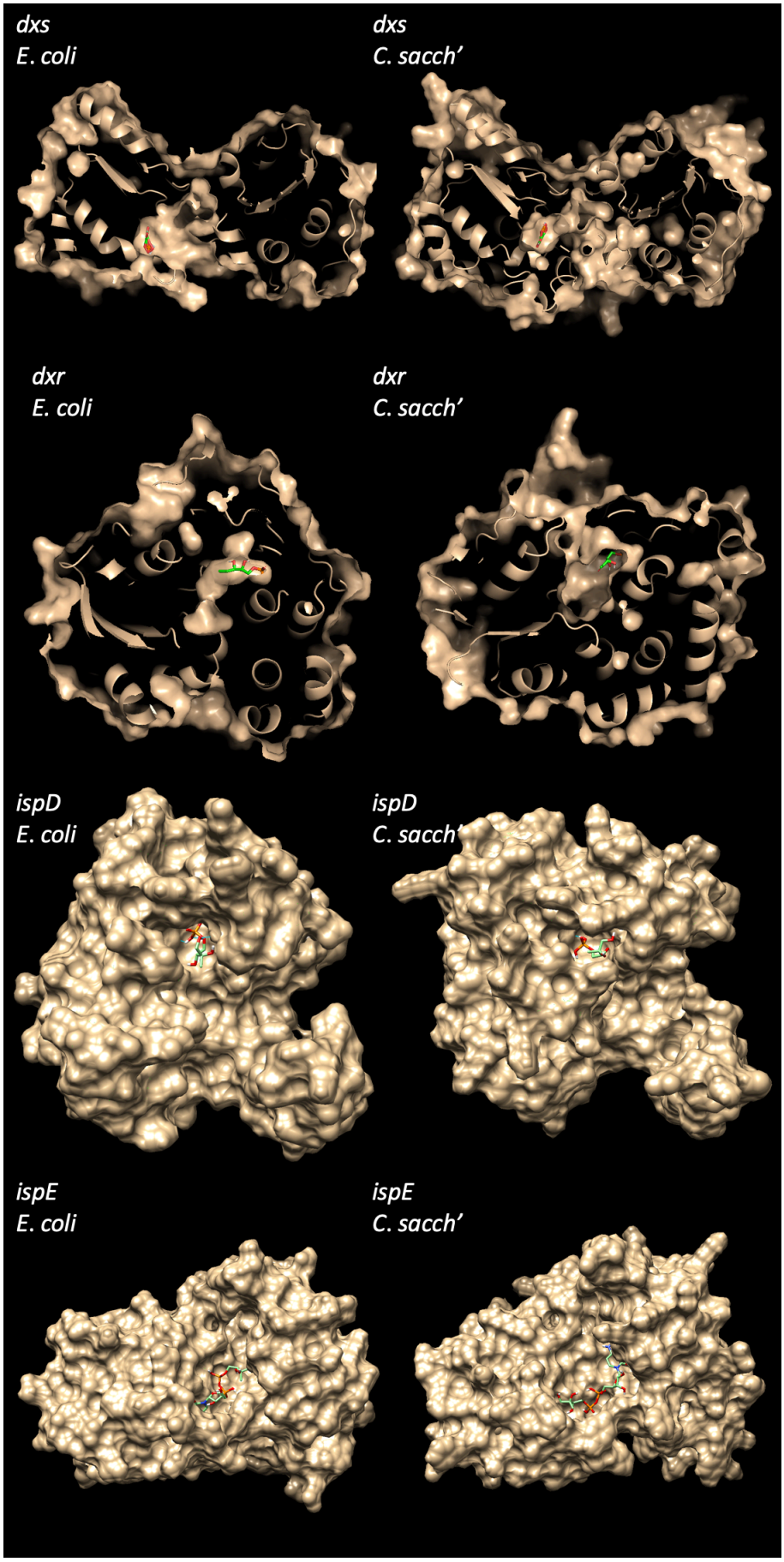


Figure 9: Substrate docking predictions for MEP/DOXP genes found in *C. saccharoperbutylacetonicum*

A selection of the clearest images produced from running the MEP/DOXP enzymes through docking software with their putative substrates. The *E. coli* crystal structures are shown on the **Left** and *C. saccharoperbutylacetonicum* structural models are shown on the **Right**. (57).

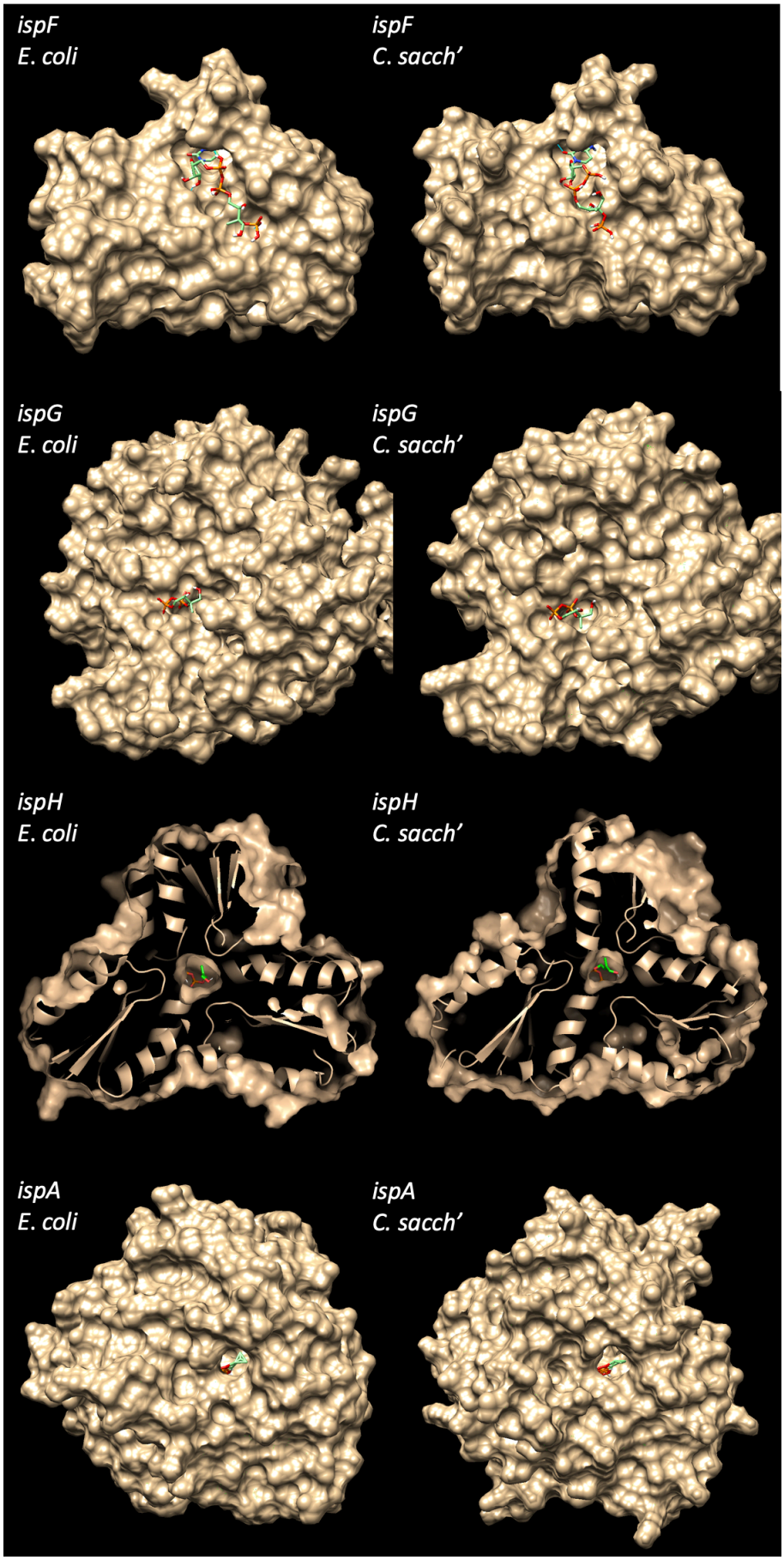


Figure 9 continued: Substrate docking predictions for MEP/DOXP genes found in *C. saccharoperbutylacetonicum*

A selection of the clearest images produced from running the MEP/DOXP enzymes through docking software with their putative substrates. The *E. coli* crystal structures are shown on the **Left** and *C. saccharoperbutylacetonicum* structural models are shown on the **Right**. (57).

4.0 Discussion

4.1 Design of a novel *mCherryPduA-U* shuttle vector:

The initial aim of this project was to clone the BMC-encoding operon *mCherryPduA-U* (46) into the shuttle vector pMTL83353 (59) for expression work in *C. saccharoperbutylacetonicum.* Unfortunately, it was not possible to produce enough *mCherryPduA-U* insert for cloning into pMTL83353. Future work with alternative cloning strategies will be undertaken and, once this vector has been constructed, the planned work on visualising the expression of a recombinant BMC in *C. saccharoperbutylacetonicum* can be performed using fluorescence or TEM microscopy. Subsequent experiments might include the tagging of a variety of biofuel pathways enzymes to the microcompartment with the aim of improving biofuel yield through sequestration of toxic intermediates and improving pathway flux via substrate channelling between pathway enzymes. Such approaches have been successfully performed in the past for ethanol production in *E. coli*, where BMCs have been used to protect the cells from toxic acetaldehyde intermediates (60).

4.2 Tolerance of bacteria to biofuels:

If one is to use *C. saccharoperbutylacetonicum* as a scaffold to produce various biofuels, then it is important to gain an insight into the toxicity of these compounds towards bacteria. Initial growth experiments with *E. coli* (**Figure 6**) clearly shows the toxic nature of biofuels, with each displaying toxicity towards *E. coli* when compared to the negative control. Alpha-Pinene and R-Limonene diminish the OD significantly at a final concentration of 5 % v/v (**Figure 6**), which might suggest that these biofuels are lysing the cells. The growth data in **Figure 6** largely agree with the viability assays in **Figure 7**, although Alpha-Pinene elicits greater killing compared to R-Limonene. However, this may be explained by the insolubility of R-Limonene

(**Table 4**) impairing its bactericidal effect when employed at a small scale. In future, it would be sensible to repeat the viability experiments with Butanol (**Figure 7A**) to more accurately define a dose response.

Previous work by Dunlop et al. has investigated the toxicity of various biofuels and biofuel precursors on *E. coli* with the aim of testing the impact of efflux pumps on bacterial survival (61). These findings showed that Butanol and Isopentanol both gradually reduced the normalised OD_{600} to a similar degree with increasing concentration, whilst Limonene had a significantly stronger effect. However, this study reported that much higher concentrations of alpha-Pinene were required to exhibit a significant effect on *E. coli* OD_{600} compared to the current work. Indeed, herein the growth data showed that alpha-Pinene exhibited a similar toxicity to R-Limonene, although the viability data did confirm that alpha-Pinene was the most toxic. Further research should be carried out to investigate whether this is due to differences in the resistance of *E. coli* strains used in both studies.

The overall toxicity of most of these biofuels to *E. coli* shows that it would be of great benefit to be able to use a bacterial chassis that is more tolerant to biofuels. Since *Clostridium* strains have evolved to produce toxic products such as butanol, it is anticipated that this genus may be naturally more tolerant to some of the biofuels in the current study, and future growth and viability assays can be used to test this hypothesis. Given the versatility of this organism in terms of feedstock utilisation from waste streams and established processes for growth on an industrial scale, this approach for the production of a variety of biofuels may have great potential for scale up.

4.3 *C. saccharoperbutylacetonicum* MEP/DOXP Enzyme Structure Prediction and Substrate Docking:

The close homology in the backbone structure of the MEP/DOXP pathway enzymes found in the *C. saccharoperbutylacetonicum* genome compared to the bona-fide *E. coli* enzymes (**Figure 8**) strongly suggests that *C. saccharoperbutylacetonicum* is able to produce the biofuel precursor geranyl pyrophosphate. Future studies should aim to express these enzymes within *C. saccharoperbutylacetonicum* to confirm these results are valid and not affected during cytoplasmic protein folding.

The substrate docking work (**Figure 9**) further supports that these enzymes may indeed be functional, owing to the similarity in hydrogen bonding to the substrates and the similarity in the conformation of their docking into the active sites when compared with the X-ray crystal structures of the bona-fide *E. coli* enzymes. This can be further analysed in future by examining exactly which residues are involved in substrate binding. In addition, experimental laboratory work should be undertaken on *C. saccharoperbutylacetonicum* to examine gene and protein expression.

Furthermore, metabolite analysis using mass spectrometry would be performed to determine unequivocally whether geranyl pyrophosphate is produced as it is the key metabolite in the production of alpha-pinene, linalool and R-limonene. If this work is successful, genes encoding terminal enzymes of alpha-pinene, linalool and R-limonene synthesis could be cloned into pMTL shuttle vectors and introduced into *C. saccharoperbutylacetonicum*. Subsequent metabolomics analysis of pathway flux and biofuel yield/titre would then provide useful insights into the suitability of *C. saccharoperbutylacetonicum* as a chassis for these pathways.

Bibliography

1. **Ali, M, et al.** 16 - The use of crop residues for biofuel production. *Construction, Biomedical, and other Industrial Applications.* 2019, pp. 369-395.
2. **Viswanathan, V, Sripad, S and Fredericks, WL.** Why aren't there electric airplanes yet? *The Conversation.* [Online] Nov 27, 2018. [Cited: 12 15, 2019.]
3. **Hill, J, et al.** Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. *PNAS.* 2006, Vol. 103, 30, pp. 11206-11210.
4. **Carere, CR, et al.** Third Generation Biofuels via Direct Cellulose Fermentation. *International journal of molecular sciences.* 2008, Vol. 9, 7, pp. 1342–1360.
5. **Aro, EM.** From first generation biofuels to advanced solar biofuels. *Ambio.* 2016, Vol. 45, Suppl 1, pp. 24-31.
6. **Mohr, A, Raman, S.** Lessons from first generation biofuels and implications for the sustainability appraisal of second generation biofuels. *Energy Policy.* 2013, Vol. 63, pp. 114-122.
7. **Lackner, M.** 3rd-Generation Biofuels: Bacteria and Algae as Sustainable Producers and Converters. *Handbook of Climate Change Mitigation and Adaptation.* 2015. pp. 1-32.
8. **Medipally, SR, et al.** Microalgae as Sustainable Renewable Energy Feedstock for Biofuel Production. *Biomed Research International.* 2015, 519513.
9. **Khan, MI, Shin, JH and Kim, JD.** The promising future of microalgae: current status, challenges, and optimization of a sustainable and renewable industry for biofuels, feed, and other products. *Microbial Cell Factories.* 2018, Vol. 17, p. 36.
10. **Patil, V, Tran, KQ and Giselrød, HR.** Towards sustainable production of biofuels from microalgae. *International journal of molecular sciences.* 2008, Vol. 9, 7, pp. 1188–1195.
11. **Sun, C, et al.** Genomic comparison of *Clostridium* species with the potential of utilizing red algal biomass for biobutanol production. *Biotechnology For Biofuels.* 2018, Vol. 11, 42.
12. **Samul, D, et al.** Beneficial and harmful roles of bacteria from the *Clostridium* genus. *Acta Biochimica Polonica.* 2013, Vol. 60, 4, pp. 515-521.
13. **Shao, M and Chen, H.** Feasibility of acetone–butanol–ethanol (ABE) fermentation from Amorphophallus konjac waste by *Clostridium acetobutylicum* ATCC 824. *Process Biochemistry.* 2015, Vol. 50, 8, pp. 1301-1307.
14. **Wells, CL and Wilkins, TD.** Chapter 18: Clostridia: Sporeforming Anaerobic *Bacilli*. [book auth.] S Baron. *Medical Microbiology.* 4th Edition. s.l. : Galveston (TX): University of Texas Medical Branch at Galveston, 1996. Chapter 18.
15. **Abd-Alla MH, El-Enany AWE.** Production of acetone-butanol-ethanol from spoilage date palm (Phoenix dactylifera L.) fruits by mixed culture of *Clostridium acetobutylicum* and *Bacillus subtilis. Biomass and Bioenergy.* 2012, Vol. 42, pp. 172-178.
16. **Liew, F, et al.** Insights into CO2 Fixation Pathway of *Clostridium autoethanogenum* by Targeted Mutagenesis. *mBio.* 2016, Vol. 7, 3, pp. e00427-16.
17. **Fernández-Naveira, A, et al.** Carbon monoxide bioconversion to butanol-ethanol by *Clostridium carboxidivorans*: kinetics and toxicity of alcohols. *Applied Microbiology and Biotechnology.* 2016, Vol. 100, pp. 4231-4240.
18. **Raud, M et al.** Potentials and challenges in lignocellulosic biofuel production technology. *Renewable and Sustainable Energy Reviews.* 2019, Vol. 111, pp. 44-56.
19. **Wi, S.G et al.** Lignocellulose conversion for biofuel: a new pretreatment greatly improves downstream biocatalytic hydrolysis of various lignocellulosic materials. *Biotechnology for Biofuels.* 2015, Vol. 8. Article 228.
20. **Chaturvedi, V., Verma, P.** An overview of key pretreatment processes employed for bioconversion of lignocellulosic biomass into biofuels and value added products. *3 Biotech.* 2013, pp. 415–431.

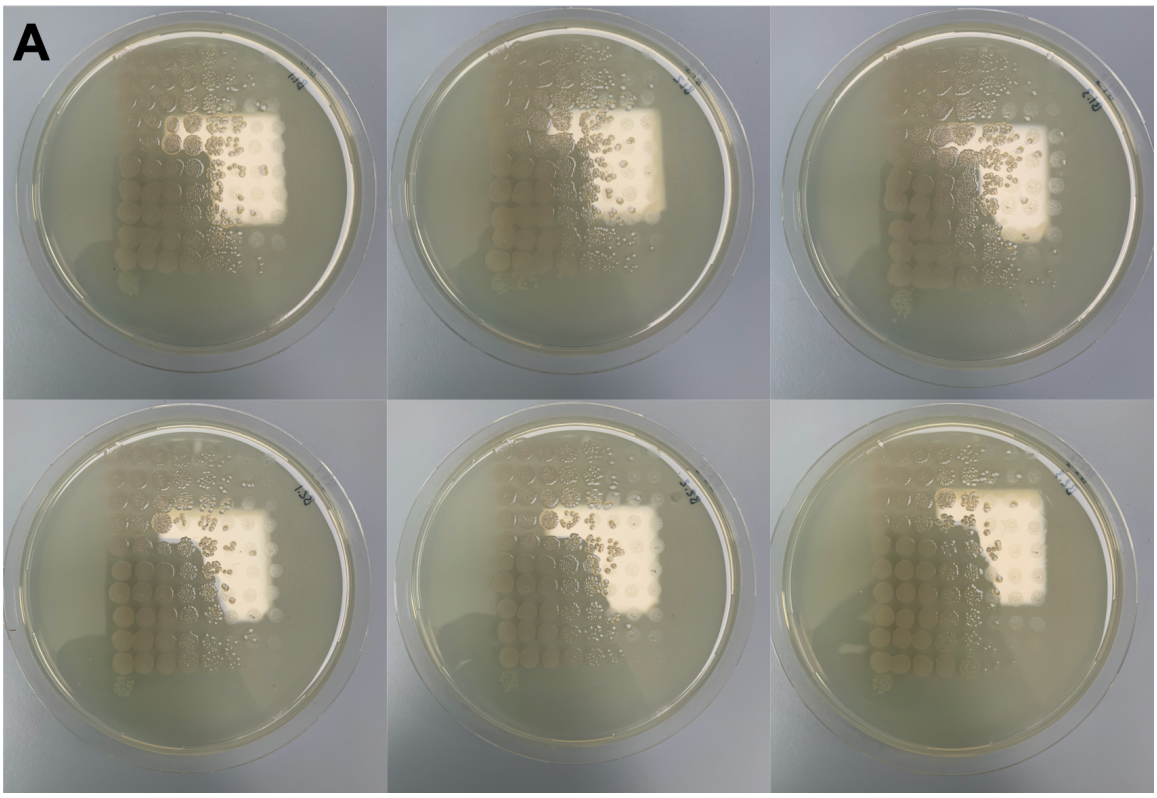
21. **Przybysz Buzała, K et al.** Conversion of various types of lignocellulosic biomass to fermentable sugars using kraft pulping and enzymatic hydrolysis. *Wood Science and Technology.* 2017, Vol. 51, pp. 873–885.
22. **Patakova, P, et al.** Acidogenesis, solventogenesis, metabolic stress response and life cycle changes in *Clostridium beijerinckii* NRRL B-598 at the transcriptomic level. *Scientific Reports.* 2019, Vol. 9, p. 1371.
23. **Sandoval, NR, et al.** Whole-genome sequence of an evolved *Clostridium pasteurianum* strain reveals Spo0A deficiency responsible for increased butanol production and superior growth. *Biotechnology for Biofuels.* 2015, Vol. 8, p. 227.
24. **Thorn, GJ, King, JR and Jabbari, S.** pH-induced gene regulation of solvent production by *Clostridium acetobutylicum* in continuous culture: Parameter estimation and sporulation modelling. *Mathematical Biosciences.* 2013, Vol. 241, 2, pp. 149-166.
25. **Visioli, LJ et al.** Recent advances on biobutanol production. *Sustainable Chemical Processes.* 2014, Vol. 2, p. Article number: 15.
26. **Peabody, GL and Kao, KC.** Recent progress in biobutanol tolerance in microbial systems with an emphasis on *Clostridium*. *Microbiology Letters.* 2016, Vol. 363, 5.
27. **Axen SD, Erbilgin O, Kerfeld CA.** A Taxonomy of Bacterial Microcompartment Loci Constructed by a Novel Scoring Method. *PLoS Computational Biology.* 2014, Vol. 10, 10, p. e1003898.
28. **Krupovic M, Koonin EV.** Cellular origin of the viral capsid-like bacterial microcompartments. *Biology Direct.* 2017, Vol. 12, p. 25.
29. **Parsons, JB et al.** Synthesis of Empty Bacterial Microcompartments, Directed Organelle Protein Incorporation, and Evidence of Filament- Associated Organelle Movement. *Molecular Cell.* 2010, Vol. 38, 2, pp. 305-315.
30. **Thompson, MC et al.** Identification of a Unique Fe-S Cluster Binding Site in a Glycyl-Radical Type Microcompartment Shell Protein. *Journal of Molecular Biology.* 2014, Vol. 426, 19, pp. 3287-3304.
31. **Seedorf H, Fricke WF, Veith B, et al.** The genome of *Clostridium kluyveri*, a strict anaerobe with unique metabolic features. *PNAS.* 2008, Vol. 105, 6, pp. 2128–2133.
32. **Wackett, LP, et al.** Genomic and Biochemical Studies Demonstrating the Absence of an Alkane-Producing Phenotype in *Vibrio furnissii* M1. *Applied and Environmental Microbiology.* 2007, Vol. 73, 22, pp. 7192–7198.
33. **Mendez-Perez, D. et al.** Production of jet fuel precursor monoterpenoids from engineered *Escherichia coli*. *Biotechnology and Bioengineering.* 2017, Vol. 114, 8, pp. 1703-1712.
34. **Jongedijk, E. et al.** Biotechnological production of limonene in microorganisms . *Applied Microbiology and Biotechnology.* 2016, Vol. 100, pp. 2927–2938.
35. **Hellier, P. et al.** Combustion and emissions characterization of terpenes with a view to their biological production in *cyanobacteria*. *Fuel.* 2013, Vol. 111, pp. 670-688 .
36. **Pahima, E. et al.** Computational design of biofuels from terpenes and terpenoids. *Sustainable Energy & Fuels.* 2019, Vol. 3, 2, pp. 457-466.
37. **Heuston, S. et al.** Isoprenoid biosynthesis in bacterial pathogens. *Microbiology.* 2012, Vol. 158, 6, pp. 1389–1401.
38. **Zhao, L. et al.** Methylerythritol Phosphate Pathway of Isoprenoid Biosynthesis. *Annual Review of Biochemistry.* 2013, Vol. 82, pp. 497-530.
39. **Rohdich, F. et al.** Studies on the nonmevalonate terpene biosynthetic pathway: Metabolic role of IspH (LytB) protein. *PNAS.* 2002, Vol. 99, 3, pp. 1158–1163.
40. **Fu-Xing, N. et al.** Metabolic engineering for the microbial production of isoprenoids: Carotenoids and isoprenoid-based biofuels. *Synthetic and Systems Biotechnology.* 2017, Vol. 2, 3, pp. 167-175.

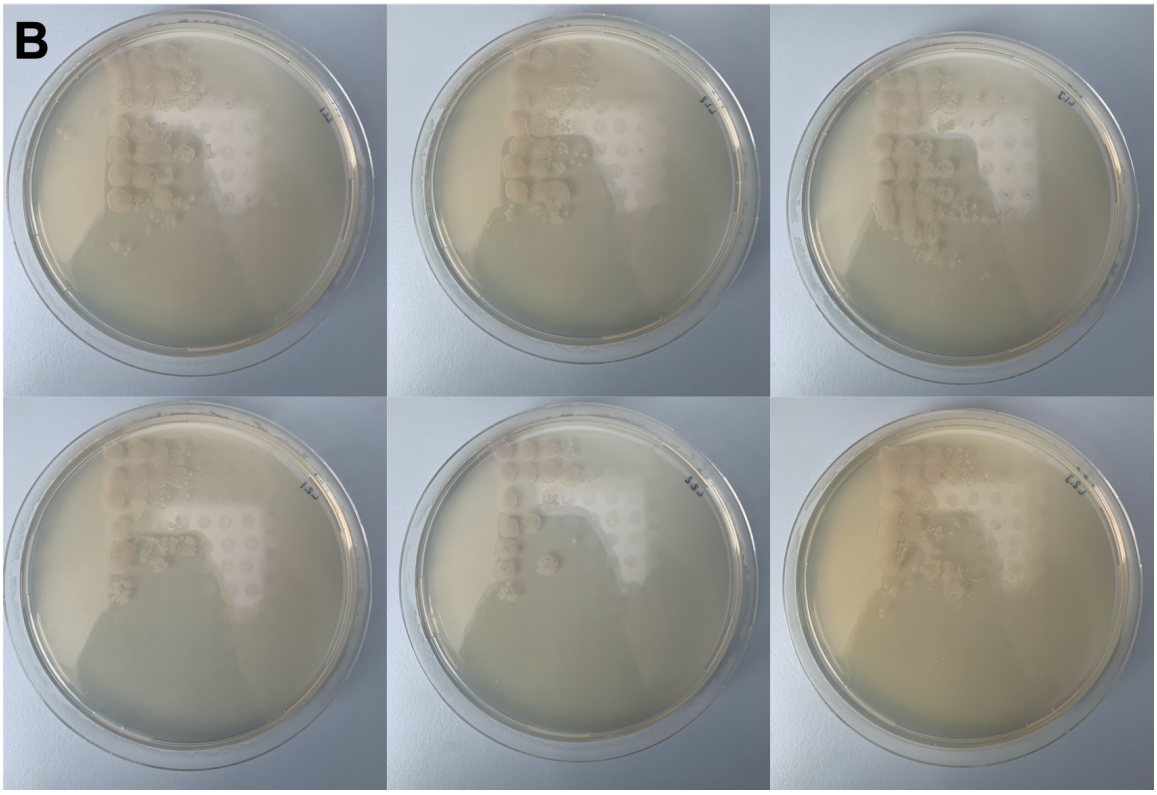
41. **Begley, M et al.** The interplay between classical and alternative isoprenoid biosynthesis controls γδ T cell bioactivity of *Listeria monocytogenes*. *FEBS Letters.* 2004, Vol. 561, 1-3, pp. 99-104.
42. **Kuzuyama, T et al.** Cloning and Characterization of 1-Deoxy-d-Xylulose 5-Phosphate Synthase from *Streptomyces sp.* Strain CL190, Which Uses both the Mevalonate and Nonmevalonate Pathways for Isopentenyl Diphosphate Biosynthesis. *Journal of Bacteriology.* 2000, Vol. 182, 4, pp. 891-897.
43. **Ershov, Y.** 2-C-Methylerythritol Phosphate Pathway of Isoprenoid Biosynthesis as a Target in Identifying New Antibiotics, Herbicides, and Immunomodulators: A Review. *Applied Biochemistry and Microbiology.* 2007, Vol. 43, 2, pp. 115-138.
44. **Nagarajan, H. et al.** Characterizing acetogenic metabolism using a genome-scale metabolic reconstruction of *Clostridium ljungdahlii*. *Microbial Cell Factories.* 2013, Vol. 12, p. 118.
45. **Diner, BA. et al.** Synthesis of Heterologous Mevalonic Acid Pathway Enzymes in *Clostridium ljungdahlii* for the Conversion of Fructose and of Syngas to Mevalonate and Isoprene. *Applied and Environmental Microbiology.* 2018, Vol. 84, 1, pp. e01723-17.
46. **Uddin, I et al.** A Generic Self-Assembly Process in Microcompartments and Synthetic Protein Nanotubes. *Small.* 2018, Vol. 14, (19), p. 1704020.
47. **Consortium, The UniProt.** UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 2019, Vol. 47, pp. D506-515.
48. **Xiang, S et al.** Crystal structure of 1-deoxy-D-xylulose 5-phosphate synthase, a crucial enzyme for isoprenoids biosynthesis. *The Journal of Biological Chemistry.* 2007, Vol. 282, 4, pp. 2676-82.
49. **Kuzuyama, T et al.** Characterization of 1-deoxy-D-xylulose 5-phosphate reductoisomerase, an enzyme involved in isopentenyl diphosphate biosynthesis, and identification of its catalytic amino acid residues. *The Journal of Biological Chemistry.* 2000, Vol. 275, 26, pp. 19928-32.
50. **Rohdich, F et al.** Cytidine 5'-triphosphate-dependent biosynthesis of isoprenoids: YgbP protein of *Escherichia coli* catalyzes the formation of 4-diphosphocytidyl-2-C-methylerythritol. *PNAS.* 1999, Vol. 96, 21, pp. 11758-63.
51. **Lüttgen, H et al.** Biosynthesis of terpenoids: YchB protein of *Escherichia coli* phosphorylates the 2-hydroxy group of 4-diphosphocytidyl-2C-methyl-d-erythritol. *PNAS.* 2000, Vol. 97, 3, pp. 1062–1067.
52. **Herz, S et al.** Biosynthesis of terpenoids: YgbB protein converts 4-diphosphocytidyl-2C-methyl-D-erythritol 2-phosphate to 2C-methyl-D-erythritol 2,4-cyclodiphosphate. *PNAS.* 2000, Vol. 97, 6, pp. 2486-90.
53. **Rohdich, F et al.** The deoxyxylulose phosphate pathway of isoprenoid biosynthesis: Studies on the mechanisms of the reactions catalyzed by IspG and IspH protein. *PNAS.* 2003, Vol. 100, 4, pp. 1586–1591.
54. **Cheon Lee, P et al.** Directed evolution of *Escherichia coli* farnesyl diphosphate synthase (IspA) reveals novel structural determinants of chain length specificity. *Metabolic Engineering.* 2005, Vol. 7, 1, pp. 18-26.
55. **Kelley, LA et al.** Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols.* 2015, Vol. 10, pp. 845-858.
56. **Schrodinger, LLC.** The PyMOL Molecular Graphics System, Version 2.4.0. 2010.
57. **Pettersen, EF et al.** UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem.* 2004, Vol. 25, 13, pp. 1605-12.
58. **Trott O, Olson AJ.** AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *Journal of Computational Chemistry.* 2010, Vol. 31, pp. 455-461.

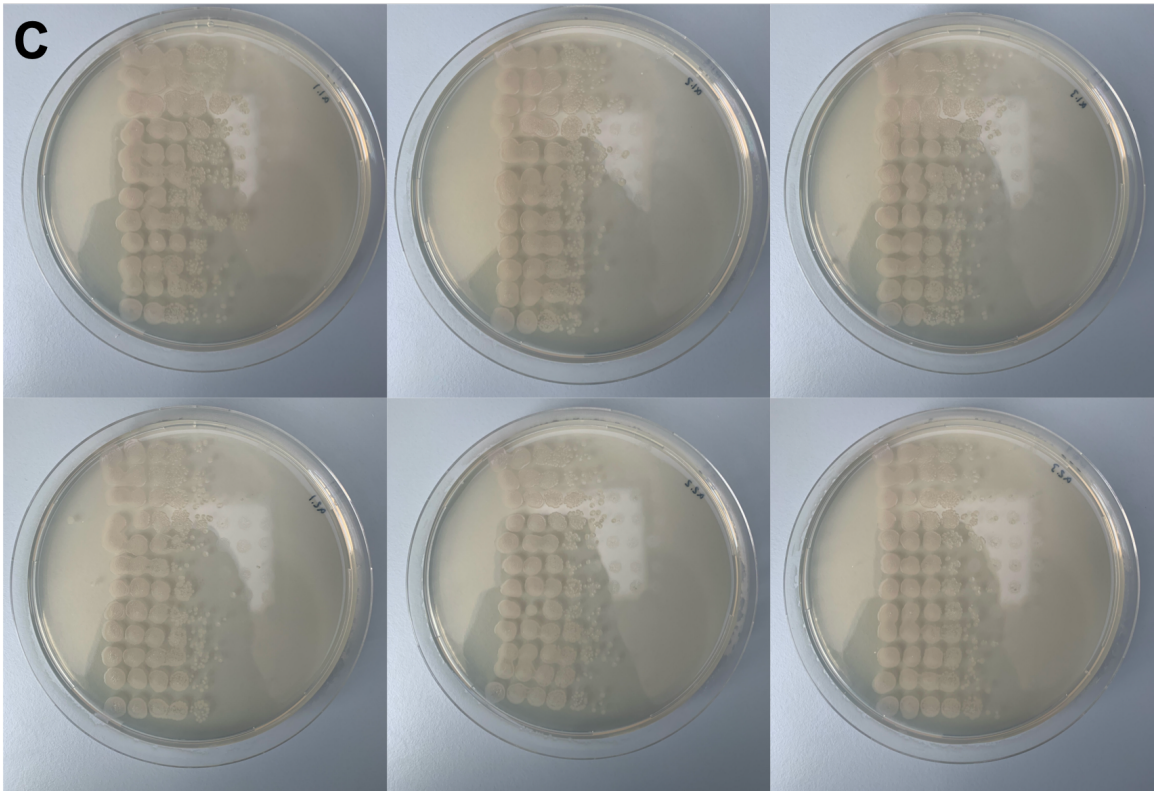
59. **Heap, JT et al.** A modular system for *Clostridium* shuttle plasmids. *Journal of Microbiological Methods.* 2009, Vol. 78, 1, pp. 79-85.
60. **Lawrence, AD et al.** Solution structure of a bacterial microcompartment targeting peptide and its application in the construction of an ethanol bioreactor. *ACS Synthetic Biology.* 2014, Vol. 3, 7, pp. 454-465.
61. **Dunlop MJ, et al.** Engineering microbial biofuel tolerance and export using efflux pumps. *Molecular Systems Biology.* 2011, Vol. 7, p. 487.
62. **Liew, F, et al.** Metabolic engineering of *Clostridium autoethanogenum* for selective alcohol production. *Metabolic Engineering.* 2017, Vol. 40, pp. 104-114.
63. **Abdelaal, AS. et al.** Genetic improvement of n-butanol tolerance in *Escherichia coli* by heterologous overexpression of groESL operon from *Clostridium acetobutylicum*. *3 Biotech.* 2015, Vol. 5, 4, pp. 401–410.

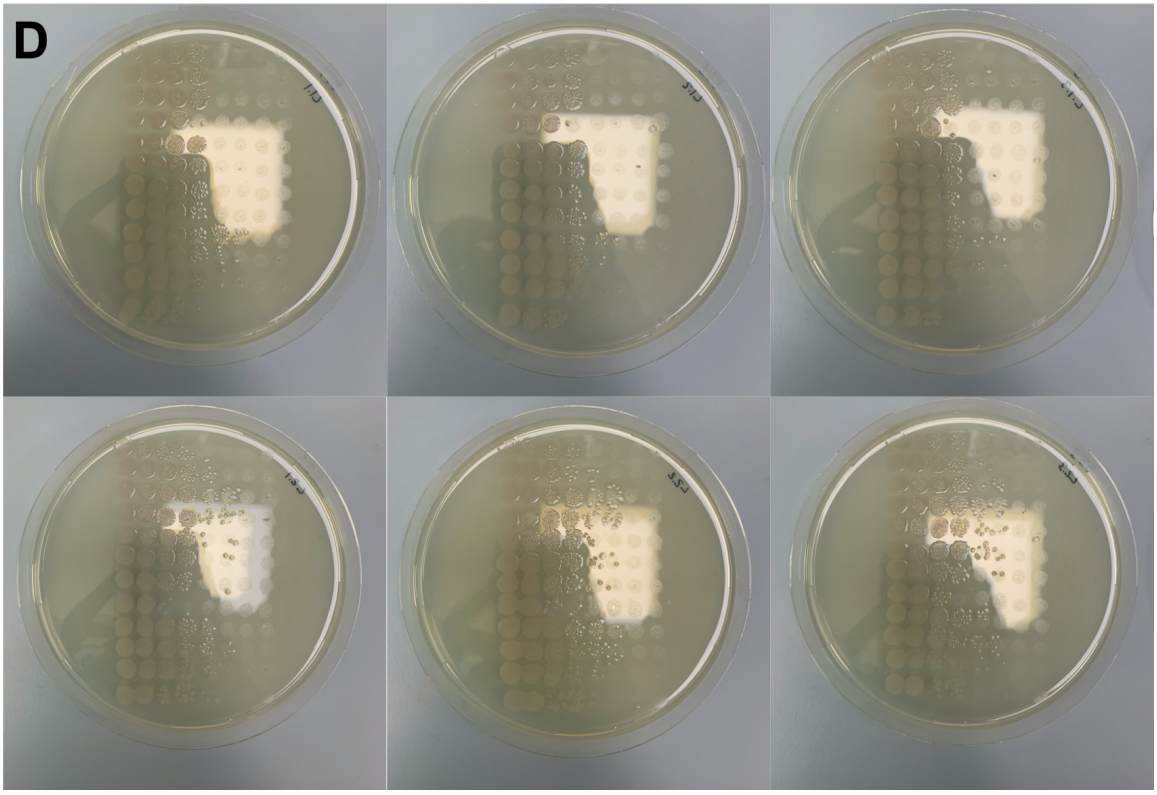
Appendix

Appendix A-1: *E. coli* viability assay agar plates exposed to (**A**) Butanol. (**B**) Linalool. (**C**) Alpha-pinene. (**D**) R-Limonene.






C


D

Appendix A-2: Multiple sequence alignments between *C. saccharoperbutylacetonicum* (**Query_10001**) and *E. coli* K12 (**Query_1002**) protein sequences.

dxs

```
Query_10001   1   MF--------LEKINSPMDIKKLKIEELKVLAYETRAALINKISNAGGHSGPNLGMVEMSVAMHYVFDSPKDKFVFDVSH  72
Query_10002   1   MSFDIAKYPTLALVDSTQELRLLPKESLPKLCDELRRYLLDSVSRSSGHFASGLGTVELTVALHYVYNTPFDQLIWDVGH  80

Query_10001   73  QSYPHKILTGRKEAFTDPAHFADVSGYTNPLESEHDFFTIGHTSTSVSLALGLAKGRDLKKEEENIIAIIGDGSLSGGEA  152
Query_10002   81  QAYPHKILTGRRDKIGTIRQKGGLHPFPWRGESEYDVLSVGHSSTSISAGIGIAVAAEKEGKNRRTVCVIGDGAITAGMA  160

Query_10001   153 LEALDYAGEYNNNLIIIVNDNDQSIAENHGGI----------YKTLRK---------------LRETNG-----ASSE  200
Query_10002   161 FEAMNHAGDIRPDMLVILNDNEMSISENVGALNNHLAQLLSGKLYSSLREGGKKVFSGVPPIKELLKRTEEHIKGMVVPG  240

Query_10001   201 NMFTAFGLDYRYLEEGHDVAKLVEFFESVKGIDHPIVLHIHTIKGKGLSYAEANHEGWHSGGPFNVEDGS-PKG-GGWVS  278
Query_10002   241 TLFEELGFNYIGPVDGHDVLGLITTLKNMRDLKGPQFLHIMTKKGRGYEPAEKDPITFHAVPKFDPSSGCLPKSSGGLPS  320

Query_10001   279 DNTVF-ESLKELLDNNDKAVVLNAGTPMGLGFVPGVREEYVERGQFIDVGIAEENAVAMSSGIAKNGGTPVFGTFAPFFQ  357
Query_10002   321 YSKIFGDWLCETAAKDNKLMAITPAMREGSGMVEFSRKF---PDRYFDVAIAEQHAVTFAAGLAIGGYKPIVAIYSTFLQ  397

Query_10001   358 RTYDQMSHDLCLNDNPATFLVLMPGVYGMNTNTHIALCDIQMFAHIPNLIYLAPTSKEEYQQMFKFATTQKEHPVGIRVP  437
Query_10002   398 RAYDQVLHDVAIQKLPVLFAIDRAGIVGADGQTHQGAFDLSYLRCIPEMVIMTPSDENECRQMLYTGYHYNDGPSAVRYP  477

Query_10001   438 GMLSVTGVPDTTDYSDYNKNKVEKQGSKVAIIAVGPMLQMAIDTEKKLKEETGMEITVINPRFLTGLDEALLNSLKTDHQ  517
Query_10002   478 RGNAVGVELTPLEKLPIGKGIVKRRGEKLAILNFGTLMPEAAKVAESL------NATLVDMRFVKPLDEALILEMAASHE  551

Query_10001   518 LVITLEDGELIGGYGQNVASFYGD--STMKVKNVGISKEFHTDFNPVELLAENGMSVEKLTAYIKENVNE  585
Query_10002   552 ALVTVEENAIMGGAGSGVNEVLMAHRKPVPVLNIGLPDFFIPQGTQEEMRAELGLDAAGMEAKIKAWLA-  620
```

dxr

```
Query_10001   1   MKKLSILGVTGSIGTQTLDVIRKSNGELKLIGVTANTSVKKIVDIIEEFKPAYVAMMDSKSAIEVKKHCKETNKKIEVLD  80
Query_10002   1   MKQLTILGSTGSIGCSTLDVVRHNPEHFRVVALVAGKNVTRMVEQCLEFSPRYAVMDDEASAKLLKTMLQQQGSRTEVLS  80

Query_10001   81  GIEGLNKIASLDEIDIVVTSVVGMIGLEPTMKAIEAKKDIALANKETLVVAGEMVMKAAKKSNVRILPVDSEHSAIFQSL  160
Query_10002   81  GQQAACDMAALEDVDQVMAAIVGAAGLLPTLAAIRAGKTILLANKESLVTCGRLFMDAVKQSKAQLLPVDSEHNAIFQSL  160

Query_10001   161 RGNDI----------NSLRKIILTASGGPFRGKKISELQEIKVKDALKHPKWNMGRKISIDSATLMNKGLEVIEAHWLFD  230
Query_10002   161 PQPIQHNLGYADLEQNGVVSILLTGSGGPFRETPLRDLATMTPDQACRHPNWSMGRKISVDSATMMNKGLEYIEARWLFN  240

Query_10001   231 CDYDNIQVLVHPQSIVHSMVEYTDGSIIAQLGAQDMRLPIQYALNYEERKNLVAETIDFYEISKLTFEKPDIATFKSLDL  310
Query_10002   241 ASASQMEVLIHPQSVIHSMVRYQDGSVLAQLGEPDMRTPIAHTMAWPNRVNSGVKPLDFCKLSALTFAAPDYDRYPCLKL  320

Query_10001   311 SFKAGKIGGLMPTILNGANEAAVDLFLNEKIKFLEIPNIIEKCMEVFKKEVSKELNLNNIIELDKTVKDYVVNNAVN---  387
Query_10002   321 AMEAFEQGQAATTALNAANEITVAAFLAQQIRFTDIAALNLSVLE--KMDMREPQCVDDVLSVDANAREVARKEVMRLAS  398
```

ispD

```
Query_10001   1   M------VSAIVLAGGRGKRMGCSKSKQYIELEGKPILYYTLKQFIDNKYIDKIVLVVPEDEKEYCKNEVLDKYGLIIHK  74
Query_10002   1   MATTHLDVCAVVPAAGFGRRMQTECPKQYLSIGNQTILEHSVHALLAHPRVKRVVIAISPGD-SRFAQLPLANHPQI--T  77

Query_10001   75  IVVGGKERQDSVYNALDQLNGSDIVLIHDGARPFVSQRVINDAIKYTRI-YGAAAPGVKPKDTIKIRDE-KNFSVNTLER  152
Query_10002   78  VVDGGDERADSVLAGLKAAGDAQWVLVHDAARPCLHQDDLARLLALSETSRTGGILAAPVRDTMKRAEPGKNAIAHTVDR  157

Query_10001   153 DSLVAIQTPQAFDFNLIYECHKKVKKSKIVVTDDTAVVENFGNKVYIYEGDYTNIKITTPEDLVFAKHFARMV------  225
Query_10002   158 NGLWHALTPQFFPRELLHDCLTRALNEGATITDEASALEYCGFHPQLVEGRADNIKVTRPEDLALAEFYLTRTIHQENT  236
```

ispE

```
Query_10001   1   MKIK--AYAKINIALDVVGKREDGYHLLRMIMQTIDLYDIIEVKKSDFGINIKCNKHYVPTDERNLAYKAAKLFKETYSI  78
Query_10002   1   MRTQWPSPAKLNLFLYITGQRADGYHTLQTLFQFLDYGDTISIELRDDGDIRLLTPVEGVEHEDNLIVRAARLLMKTAAD  80

Query_10001   79  S------DGVDIDITKYIPVSAGLAGGSTDGAAVLKLMNKIFNINASDDDLKALGLKLGADVPYCINGGTALCEGIGEKI  152
Query_10002   81  SGRLPTGSGANISIDKRLPMGGGLGGGSSNAATVLVALNHLWQCGLSMDELAEMGLTLGADVPVFVRGHAAFAEGVGEIL  160

Query_10001   153 TELKRFKDKILVLIKPPFGVSTKEVYKEFDLSKVIIHPKIEELIKNMEKDDIYFVANNMKNLLENVTLRKHRIISRIKED  232
Query_10002   161 TPVDP-PEKWYLVAHPGVSIPTPVIFKDPELPRNTPKRSIETLLKC----------EFSNDCEVIARKRFREVDAVLS-  227

Query_10001   233 TTIDGAIGTMMSGSGPTVFAFFDDMLKAQTCYDNMKKKYKDVFITRTI----------  280
Query_10002   228 -WLLEYAPSRLTGTGACVFAEFDTESEARQVLEQAPE-WLNGFVAKGANLSPLHRAML  283
```

ispF

```
Query_10001  1    MRVGLGYDVHRLVENRKLILGGVEIPYEKGLLGHSDADVLLHAIMDSLLGACALGDIGKHFPDTDSKFSGISSLLLLKET  80
Query_10002  1    MRIGHGFDVHAFGGEGPIIIGGVRIPYEKGLLAHSDGDVALHALTDALLGAAALGDIGKLFPDTDPAFKGADSRELLREA  80

Query_10001  81   GKLIFEAGYAVNNIDATIIAQKPKMLPHIQNMRENISEALNIDLDKINIKATTEEGLGFTGEMLGISAQSIASV-ENIK  158
Query_10002  81   WRRIQAKGYTLGNVDVTIIAQAPKMLPHIPQMRVFIAEDLGCHMDDVNVKATTTEKLGFTGRGEGIACEAVALLIKATK  159
```

ispG

```
Query_10001  1    ------MERRKSRKVKVGNIYVGGNAPITVQSMTNTPTKDVEATLKQIQQLFDAGCDIVRCAVLDMEDAEKLGEITKKSP  74
Query_10002  1    MHNQAPIQRRKSTRIYVGNVPIGDGAPIAVQSMTNTRTTDVEATVNQIKALERVGADIVRVSVPTMDAAEAFKLIKQQVN  80

Query_10001  75   IPVVADIHFDYKLALKAIENGVSALRINPGNIGSVDRIKILKEACVEKQIPIRIGVNSGSLEKDILEKYGMPTAEGLVES  154
Query_10002  81   VPLVADIHFDYRIALKVAEYGVDCLRINPGNIGNEERIRMVVDCARDKNIPIRIGVNAGSLEKDLQEKYGEPTPQALLES  160

Query_10001  155  ALRHVKILEDLDFYDIVISIKSSNVQMMIDCYRMIAEKCDYPLHLGVTEAGTIQRGTIKSSIGIGTLLAEGIGDTIRVSL  234
Query_10002  161  AMRHVDHLDRLNFDQFKVSVKASDVFLAVESYRLLAKQIDQPLHLGITEAGGARSGAVKSAIGLGLLLSEGIGDTLRVSL  240

Query_10001  235  TTDPVQEIKVGTEILRTLGLKKKGVEFVSCPTCGRTQINLIKIAEEVEKRLENHKKDIKVAVMGCVVNGPGEAREADIGI  314
Query_10002  241  AADPVEEIKVGFDILKSLRIRSRGINFIACPTCSRQEFDVIGTVNALEQRLEDIITPMDVSIIGCVVNGPGEALVSTLGV  320

Query_10001  315  AGGIGEGIIFKKGQIIKKVKEEELVDALMKEIENL----------------  349
Query_10002  321  TGGNKKSGLYEDGVRKDRLDNNDMIDQLEARIRAKASQLDEARRIDVQQVEK  372
```

ispH

```
Query_10001  1    MSDVLLARNAGFCFGVQRAVDEALRIQKEYNKKIYTLGPLIHNNDVVKFLENNNIYSIEFEEVDKLSVGDVIVVRSHGVS  80
Query_10002  1    -MQILLANPRGFCAGVDRAISIVENALAIYGAPIYVRHEVVHNRYVVDSLRERGAIFIE--QISEVPDGAILIFSAHGVS  77

Query_10001  81   EMILDYLENKQFTVINATCPFVTNIQKKVKKYSKEGYHIVILGDKNHPEVIGINGWCNN---KAIITKDGTFEEQIP---  154
Query_10002  78   QAVRNEAKSRDLTVFDATCPLVTKVHMEVARASRRGEESILIGHAGHPEVEGTMGQYSNPEGGMYLVESPDDVWKLTVKN  157

Query_10001  155  -QKVCLVSQTTEKMKNWENTLVNL-SRVKELL--AFNTICSATEVRQKSTNQISKEADAMIVIGGKNSSNTTKLYQIAKD  230
Query_10002  158  EEKLSFMTQTTLSVDDTSDVIDALRKRFPKIVGPRKDDICYATTNRQEAVRALAEQAEVVLVVGSKNSSNSNRLAELAQR  237

Query_10001  231  NCINTIHIENFKELPKDFINKNIKKIGVTAGASTPDWIIREVLDTMNT----------ENNVNTENQLDLMNQNDIKV  298
Query_10002  238  MGKRAFLIDDAKDIQEEWV-KEVKCVGVTAGASAPDILVQNVVARLQQLGGGEAIPLEGREENIVFEVPKEL--RVDIRE  314

Query_10001  299  IIGTEVEEKVISKNNDGIVIEMNGYGIDGIIPYNELTTKGDSKSLADSINVGDLIKAKVIKLQNNDGFVVLSRLEYEREE  378
Query_10002  315  VD----------------------------------------------------------------------------  316

Query_10001  379  AYKELEKLFTEGTIFEVNIKEAREKGLVSDYKGVRIFIPASQIDIRFIQNKDEFVGKNLEVKLIDFSLGKPVKVVASRRV  458
Query_10002       ----------------------------------------------------------------------------

Query_10001  459  LLEEAKNAEEAKALENINIGDVVKGEVKRFTNFGAFININGIDGLLHLSQISWNHIKNAGDVLKEGQIIDVKVIDLDKET  538
Query_10002       ----------------------------------------------------------------------------

Query_10001  539  KKLSLSMKELMPKPWDNAKEKYPEESIVLGKVVRLNDFGAFVELEPGVDGLVHISKISHNRIEHPSEVLKIGQEIKAKIL  618
Query_10002       ----------------------------------------------------------------------------

Query_10001  619  GVDEENKRISLSIKDV  634
Query_10002       ----------------
```

ispA

```
Query_10001  1    MEM--EINNLKQDIDSYLKDYFINKGTYNKVIYDSCSYSLNIGGKRIRPILLALTYYIYKENYNKVIPMAAAIEMIHTYS  78
Query_10002  1    MDFPQQLEACVKQANQALSRFIAPLPFQNTPVVETMQYGALLGGKRLRPFLVYATGHMFGVSTNTLDAPAAAVECIHAYS  80

Query_10001  79   LIHDDLPCMDDDDLRRGQPTNHIKFQENIAVLAGDALLNEAMIIMMKFA--LENKENSLLAAHEIAAAAGAEGMIGGQVV  156
Query_10002  81   LIHDDLPAMDDDDLRRGLPTCHVKFGEANAILAGDALQTLAFSILSDADMPEVSDRDRISMISELASASGIAGMCGGQAL  160

Query_10001  157  DIVSEGKEISKDELEYMHAKKTGALIKASIVAGAILADAP-KDDLEMFEKYGEKLGLVFQIKDDILDVVGDKNKLGKNTH  235
Query_10002  161  DLDAEGKHVPLDALERIHRHKTGALIRAAVRLGALSAGDKGRRALPVLDKYAESIGLAFQVQDDILDVVGDTATLGKRQG  240

Query_10001  236  ADEEHNKTNFISVFGLEKCIDLCESLTEECITILNNL---SVNAEYLVNLTYNLLNREN  291
Query_10002  241  ADQQLGKSTYPALLGLEQARKKARDLIDDARQSLKQLAEQSLDTSALEALADYIIQRNK  299
```

ispH mismatching protein sequence

```
Query_10001  1    --------------------GTEVEEKVISKNNDGIVIEMNGYGIDGIIPYNEL-TTKGDSKSLADSINVGDLIKAKVIK  59
Query_10002  1    MTESFAQLFEESLKEIETRPGSIVRGVVVAIDKDVVLVDA-GLKSESAIPAEQFKNAQGELE-----IQVGDEVDVALDA  74

Query_10001  60   LQNNDGFVVLSRLEYEREEAYKELEKLFTEGTIFEVNIKEAREKGLVSDYKGVRIFIPASQIDIRFIQNKDEFVGKNLEV  139
Query_10002  75   VEDGFGETLLSREKAKRHEAWITLEKAYEDAETVTGVINGKVKGGFTVELNGIRAFLPGSLVDVRPVRDTLHLEGKELEF  154

Query_10001  140  KLIDFSLGKPVKVVASRRVLLEEAKNAEEAKALENINIGDVVKGEVKRFTNFGAFININGIDGLLHLSQISWNHIKNAGD  219
Query_10002  155  KVIKLDQ-KRNNVVVSRRAVIESENSAERDQLLENLQEGMEVKGIVKNLTDYGAFVDLGGVDGLLHITDMAWKRVKHPSE  233

Query_10001  220  VLKEGQIIDVKVIDLDKETKKLSLSMKELMPKPWDNAKEKYPEESIVLGKVVRLNDFGAFVELEPGVDGLVHISKISHNR  299
Query_10002  234  IVNVGDEITVKVLKFDRERTRVSLGLKQLGEDPWVAIAKRYPEGTKLTGRVTNLTDYGCFVEIEEGVEGLVHVSEMDWTN  313

Query_10001  300  IE-HPSEVLKIGQEIKAKILGVDEENKRISLSIKDV----------------------------------------------  334
Query_10002  314  KNIHPSKVVNVGDVVEVMVLDIDEERRRISLGLKQCKANPWQQFAETHNKGDRVEGKIKSITDFGIFIGLDGGIDGLVHL  393

Query_10001       ----------------------------------------------------------------------------
Query_10002  394  SDISWNVAGEEAVREYKKGDEIAAVVLQVDAERERISLGVKQLAEDPFNNWVALNKKGAIVTGKVTAVDAKGATVELADG  473

Query_10001       ----------------------------------------------------------------------------
Query_10002  474  VEGYLRASEASRDRVEDATLVLSVGDEVEAKFTGVDRKNRAISLSVRAKDEADEKDAIATVNKQEDANFSNNAMAEAFKA  553

Query_10001       ----
Query_10002  554  AKGE  557
```

Appendix A-3: IspG Autodock Vina full enzyme images for (**A**) *E. coli* and (**B**) *C. saccharoperbutylacetonicum*.