Developing a proteomics pipeline for characterisation of a synthetic microbial co-culture

Mengxun Shi

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

First Supervisor: Dr Jagroop Pandhal

Second Supervisor: Pro Mark Dickman

Department of Chemical and Biological Engineering

Faculty of Engineering

The University of Sheffield

Submitted 12th December 2023

Abstract

Inspired by microbial communities can undertake more complex tasks than pure cell cultures and can achieve continuous production of valuable products, an increasingly number of attempts have been made to grow artificial or synthetic microbial communities composed of designated species. However, detailed understanding of their communications and stability is often insufficient. A detailed understanding of the composition and dynamics of microbial communities is necessary to gain fundamental insights into the interactions among species.

In this project, a synthetic microbial co-culture system was constructed using genetically engineered strains of photosynthetic cyanobacterium *Synechococcus elongatus* cscB/SPS and nitrogen-fixing bacterium *Azotobacter vinelandii* ΔnifL, in which the sucrose produced from *S. elongatus* cscB/SPS would be the carbon source for *A. vinelandii* ΔnifL and the ammonium produced from *A. vinelandii* ΔnifL would be the nitrogen source for *S. elongatus* cscB/SPS. This synthetic microbial co-culture can convert sunlight, carbondioxide, and atmospheric nitrogen into desired bio-product polyhydroxybutyrate (PHB).

Quantitative proteomics can provide valuable insights into how microbial strains adapt to changing conditions. However, current workflows and methodologies are not suitable for simple artificial co-culture systems where strain ratios are dynamic. Therefore, a mass spectrometry-based label-free shotgun proteomics workflow was established for the analysis of the composition and changes of the synthetic microbial co-culture proteome. Furthermore, a new normalization method was proposed, named “LFQRatio”, to reflect the relative contributions of the two distinct cell types emerging from the cell ratio changes during co-cultivation and minimise the impact of cell number changes for proteome quantification in microbial co-culture.

Lastly, comparative proteomics was performed to analyse the protein changes of *A. vinelandii* ΔnifL and *S. elongatus* cscB/SPS in the synthetic co-culture, showing higher relative abundance of proteins in nitrogen fixation and photosynthesis pathways, which is evidence of the cross-feeding between the two members. In addition, co-culture proteomics changes over time were studied, demonstrating some interactions between co-cultures, including general carbon metabolism, two-component system, bacterial chemotaxis, etc. The proteomics results reveal potential targets for optimisation to keep the synthetic microbial co-culture healthier and longer-lived.

Acknowledgements

I would like to express my deepest appreciation to my supervisor Dr Jagroop Pandhal for his guidance and support during my project, I know it is a big challenge for both of us. I would also like to thank my sponsors China Scholarship Council (CSC), and the University of Sheffield for giving me the opportunity to do my PhD project under great environment.

I am sincerely thankful to all of the amazing and lovely researchers in CBE that have provided me with support and advice during my PhD, particularly Dr Caroline Evens – thank you for teaching me how to do LC-MS/MS and proteomics data analysis, Dr Esther Karunakaran – thank you for giving me advice with my project, Kasia Okurowska Emery – thank you for teaching me the flow cytometer, and Syed Mohammad – thank you for helping me with the GC! I would also like to thank James Grinham, Kirsty L Franklin and Sarah Whittle.

Many thanks to all the members in the Pandhal group: Juliano, Alaa, Wan, Josie, Fikayo, Jose, Helen, Zongting, Ali, Charlotte, and Joanna. It has been a great experience to work with them.

I would like to acknowledge Professor Daniel Ducat at the Michigan State University and Dr Leonardo Curatti at INBIOTEC, who kindly gifted me the strains for my project. Special thanks to Professor Xiaoxia Nina Lin at the University of Michigan for her advice of synthetic co-culture and Dr Josselin Noirel for his help with the R script and modelling.

Finally, I would like to thank my family for all the supports, especially my fiancé Minglin Cheng (soon to be husband), who has kept my motivation and given me the emotional support. I couldn't have completed my PhD without them!

Declaration

I declare that this thesis is my own work and has been written by me. This work has not been previously presented for or any other degree. I am aware of the University's Guidance on the Use of Unfair Means (https://www.sheffield.ac.uk/new-students/unfair-means).

Contents

List of Figures

List of Tables

List of abbreviations

Abbreviation	Definition
2-OG	2-oxoglutarate
3-PG	3-phosphoglycerate
BU	Bottom-Up
CA	Carbonic Anhydrase
CAP	Catabolite Gene Activator Protein
CaCl2	Calcium Chloride
CB	Calvin-Benson
CC	Cellular Component
CCM	CO_2 Concentration Mechanism
CFUs	Colony-Forming Units
CID	Collision-Induced Dissociation
Cm	Chloramphenicol
CscB	Sucrose permease
CO_2	Carbon Dioxide
Cyt	Cytochrome
DTT	Dithiothreitol
ECD	Electron Capture Dissociation
ESI	Electrospray Ionisation
ETD	Electron Transfer Dissociation
FC	Fold Change
FCM	Flow Cytometer

FDR	False Discovery Rate
FNR	Ferredoxin-$NADP^+$ Reductase
Fru-6-P	Fructose-6-Phosphate
FTMS	Fourier Transform Mass Spectrometry
GC	Gas Chromatography
Gln	Glutamine
Glu	Glutamate
GO	Gene Ontology
GOGAT	Glutamate Synthase
GS	Glutamine Synthetase
HPLC	High Performance Liquid Chromatography
iBAQ	Intensity Based Absolute Quantification
ICAT	Isotope-Coded Affinity Tag
ICPL	Isotope-Coded Protein Label
ICR	Ion Cyclotron Resonance
IDA	Iodoacetic Acid
IMAC	Immobilized Metal Ion Affinity Chromatography
ITMS	Ion Trap Mass Spectrometry
iTRAQ	Isobaric Tag for Absolute and Relative Quantitation
KEGG	Kyoto Encyclopedia of Genes and Genomes
Km	Kanamycin
LFQ	Label-Free Quantification
LHCs	Light-Harvesting Complexes

MALDI	Matrix-Assisted Laser Desorption/Ionisation
MD	Middle-Down
MF	Molecular Function
MOPS	3-(N-morpholino) Propanesulfonic Acid
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
NADPH	Nicotinamide Adenine Dinucleotide Phosphate
Nar	Nitrate Reductase
Nir	Nitrite Reductase
Nrt	Nitrate/Nitrite Transporter
O_2	Oxygen
OD	Optical Density
PC	Platocyanin
PCA	Principal Component Analysis
PCP	Pentachlorophenol
PGA	3-Phosphoglyceric Acid
PHA	Polyhydroxyalkanoate
PHB	Polyhydroxybutyrate
pI	Isoelectric point
PQ	Plastoquinone
PSI	Photosystem I
PSII	Photosystem II
PSMs	Peptide Spectrum Matches

PTMs	Post-Translational Modifications
PVA	Polyvinyl Alcohol
ROS	Reactive Oxygen Species
rpm	Revolutions Per Minute
RuBisCO	Ribulose-1,5-bisphosphate carboxylase/oxygenase
SGR	Specific Growth Rate
SILAC	Stable Isotope Labelling by Amino Acids In Cell Culture
SDS	Sodium Dodecyl Sulphate
SDS-PAGE	Sodium Dodecyl Sulphate – Polyacrylamide Gel Electrophoresis
SPS	Sucrose Phosphate Synthase
TAP	Tandem Affinity Purification
TCA	Tricarboxylic Acid
TD	Top-Down
TFA	Trifluoroacetic Acid
TMT	Tandem Mass Tag
TOF	Time-of-Flight
Urt	Urea Transporter
v/v	Volume/ Volume
wt	Wild type
w/w	Weight/Weight

Chapter 1 Literature Review

1.1 Introduction to microbial communities

Microbial communities, ubiquitous in nature, are groupings of two or more microorganisms that live in the same environment. The interactions existing in the microbial community systems can help the members adapt to changes in the environment and can resist the invasion of harmful strains. Examples include the complexes of algae and fungi – lichens (Aschenbrenner et al., 2016), marine ecotype systems of marine cyanobacteria and related heterotrophic bacteria (Kearney et al., 2021), and soil symbiotic systems composed of nitrogen-fixing Rhizobia (Trabelsi et al., 2012), etc. Microbial communities have been used in industrial processes for years, e.g., fermentation and wastewater treatment (Kouzuma and Watanabe, 2014; Jiang et al., 2017). However, getting them to do more complex tasks in biomanufacturing or resource recovery is more of a challenge, and therefore the rapid development of synthetic biology and omics tools have helped in characterising the system and the design space for artificial consortia.

Microbial consortia, for example, containing bacteria, fungi and/or microalgae can perform more complicated activities and withstand more changeable environments than individual populations due to their distinct features (Brenner et al., 2008). Members in microbial communities can interact with each other via exchange of specialised chemical signals or metabolite transactions. Moreover, labour division among different individuals makes it valuable in bioproduction (Hays et al., 2015). Eiteman *et al.*, for example, developed two strains of *Escherichia coli* system; one that only takes glucose and the other only consumes xylose, to more efficiently metabolise the substrates (Eiteman et al., 2008). Zhou et al. created a synthetic pathway for the production of acetylated diol paclitaxel precursor, a drug utilised to treat a range of cancers but rarely seen in nature, in the co-culture of *Saccharomyces cerevisiae* and *E. coli*, neither of them can manufacture paclitaxel precursors on their own (Zhou et al., 2015). What's more, when the feedstock is tangled and frequently varies in composition, such as in wastewater treatment, microbial communities are usually more effective (Cydzik-Kwiatkowska and Zielińska, 2016).

1.1.1 Interactions between microbial communities

Many kinds of interactions can occur among microbial communities, through either positive or negative effects (Figure 1-1). Members of the communities rely on each other to survive is named 'mutualism' (or cooperation). Similarly, 'protocooperation' refers to the scenarios where members of microbial communities can benefit from each other, but not obligatory (Tshikantwa et al., 2018). In some communities, one member can affect the well-being of the other but is not affected itself; This positive one-way interaction is known as 'commensalism', whereas negative one-way interaction is known as 'amensalism.' (Roell et al., 2019). In 'parasitism' microbial interaction, one organism benefits over the other, and bacteriophages are the commonly known examples (Tshikantwa et al., 2018). A typical example of a negative effect is 'predation', where a predator feeds its prey, which usually happens during the perturbation of the ecological niche. Further, 'competition' describes the scenarios where one species competes for limited nutrients or energy with others (Kong et al., 2018).

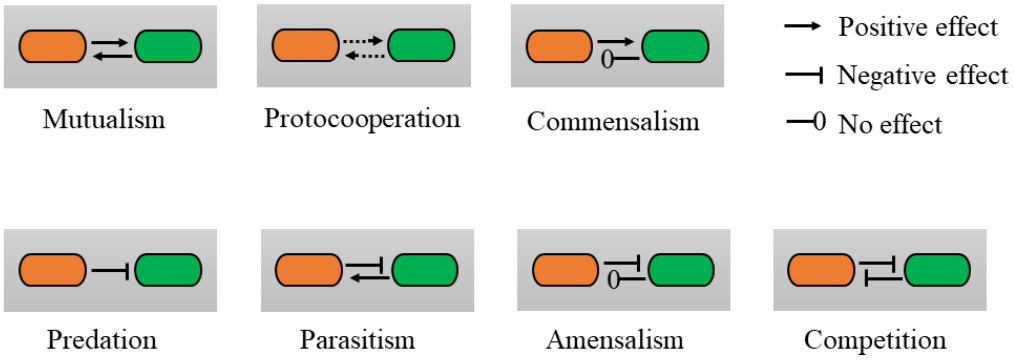


Figure 1-1 Illustration of different interactions in microbial communities

Microbial community members can interact with each other in a variety of ways, including physical contact, metabolic exchange, and chemical signalling (Lindemann et al., 2016). For example, Benomar *et al.* demonstrated that physical contact causes emergent features in two anaerobic bacteria *Clostridium acetobutylicum* and *Desulfovibrio vulgaris*, by co-culturing them with or without dialysis (Benomar et al., 2015). Metabolic exchange is particularly prevalent in bacteria-microalgae symbiosis,

because many microalgae require external micronutrients, such as vitamin B_{12} (cobalamin) to manufacture essential nutrients (Kazamia et al., 2012). Many investigations on vitamin B12 and photosynthate exchange in microalgae and bacterial co-cultures have been conducted, including *Thalassiosira purpureum* with *Halomonas* sp. (Croft et al., 2005), *Lobomonas rostrata* with *Mesorhizobium loti* (Grant et al., 2014; Helliwell et al., 2018), and *Chlamydomonas nivalis* with *Mesorhizobium loti* co-cultures (Kazamia et al., 2012). In addition to nutrient exchange, chemical signalling interactions also aid in coordinating reactions between members of the microbial communities. Chemical interaction is the process where one member of a microbial community produces a compound, such as small signalling molecules, which can be recognised and responded by other cells (Duan et al., 2009). For example, the mycorrhiza helper bacterium Streptomyces strain AcH 505 can produce a fungal growth stimulating compound (auxofuran) and fungal inhibitory molecules (WS-5995 B and C-antibiotics), which improves the mycelial growth of ectomycorrhizal fungi and inhibit the growth of phytopathogenic fungi (Riedlinger et al., 2006). Another notable type of chemical action is the production of molecules that are inhibitory to other microorganisms or the host, such as antimicrobials and toxins (Kenny and Balskus, 2018). It should be emphasised that the interactions between the microbes are species-specific (Fukami et al., 1997; Park et al., 2008).

1.1.2 Microbial community engineering principles

1.1.2.1 Intercellular interaction engineering

As previously stated, bacteria interact in a variety of ways, including mutualism, commensalism, amensalism, and competition. Members of communities may compete for the same nutrients while also exchanging their products. Thus, understanding how to split activities in the microbial population to produce optimal behaviour is thus a major challenge of designing synthetic microbial communities (Johns et al., 2016). Co-culture of photosynthetic microalgae or cyanobacteria with bacteria is a promising strategy, as there are some noticeable beneficial pathways from the heterotrophs presented, including cross-feeding of TCA cycle intermediates, increasing dissolved carbon dioxide (CO_2), mitigation of reactive oxygen species (ROS), increasing availability of essential metal ions, and secretion of soluble vitamins. Many attempts have been made to develop a mutualism or symbiotic connection between

microorganisms. Christie-Oleza *et al.*, for example, built a phototroph-heterotroph mutualistic relationship between *Roseobacter pomeroyi* and *Synechococcus* sp. WH7803, demonstrating that while *Synechococcus* is primarily focused on photosynthesis and carbon fixation, it relies on heterotrophic bacteria *R. pomeroyi* to remineralize organic matter leakage, enabling nutrients to keep flowing in the mutualism system (Christie-Oleza et al., 2017). And adding nitrogen-fixation bacteria *Azotobacter chroococcum* to *Chlamydomonas reinhardtii* increased the lipid synthesis from 65.99 mg/L to 387.76 mg/L (Xu et al., 2018).

1.1.2.2 Spatial partitioning construction

Although different synthetic microbial communities have been examined in mixed co-culture, certain microbial communities are spatially isolated, including metabolic or physiological incompatibility, to increase bioproductivity or boost product separation. There are several methods for spatial separation of microorganisms, including biofilm, dialysis, encapsulation, microfluid, etc. Weiss et al., for example, co-cultured alginate-encapsulated *S. elongatus* with *Halomonas boliviensis* to create PHB, which can increase sucrose productivity by raising sucrose export rates and allowing the isolation of heterotrophic biomass (Weiss et al., 2017). This approach was also employed in study by Therien's et al. (2014), who co-cultivated *Chlamydomonas reinhardii* with alginate encapsulated *Synechococcus* sp. PCC 7002 to address the issue overgrowth by *Synechococcus*. (Therien et al., 2014). Other approaches for reducing physical contact have also been investigated, including biofilm (Brenner et al., 2008; Brenner and Arnold, 2011) and dialysis (Benomar et al., 2015).

1.1.2.3 Community robustness maintaining

Resistance and resilience are two essential characteristics to consider in engineered microbial communities because a loss of engineering function can result in cheating phenotypes that consume public resources without contributing anything over time (Allison and Martiny, 2009; Johns et al., 2016). Attempts to circumvent the challenge of cheating phenotypes have been made, e.g., employing nutrient-limited synthetic medium (Asfahl and Schuster, 2017). Furthermore, the invasion of exotic species poses a risk to the community's resilience. Although microbial communities offer an edge in terms of resistance to alien species as compared to pure cultures, it is still a concern that must be overlooked when constructing synthetic microbial communities. To some

extent, using severe cultivation conditions, such as high salt pressure and high temperature, can help to resist the invasion of foreign species. Another issue to consider is genetic evolution, because biological components always change due to selection, mutation and gene flow (horizontal gene transfer) (Pandhal and Noirel, 2014). Cells in the consortium will evolve, which might be useful or destructive, but we normally prevent evolution to manage the possible negative impacts of evolution.

1.1.3 Microbial community applications

1.1.3.1 Human health

Microbial communities have received extensive attention for their important impact on human health. Specifically, the human microbiome, the community of microbes living in human body, plays critical roles in human physiology, immunity, development, and nutrition (Eng and Borenstein, 2019). For example, the gut microbiome takes part in immune system development, modulating elements such as lymphoid structure development and T cell differentiation (Kabat et al., 2014). Butyrate generated by microbes aids to colonic health by serving as a vital energy source and anti-inflammatory agent (Donohoe et al., 2011).

1.1.3.2 Bio-manufacturing

Microbial monocultures are the foundation of the majority of existing bio-industrial productions, but they are always vulnerable to a variety of conditions, such as contamination by foreign bacteria (Weiss et al., 2017). Microbial communities are more resistant to alien species than monocultures, hence they have been utilised to produce goods such as lipid (Xu et al., 2018) and biofuels (Unnithan et al., 2014). Zhou et al. divided acetylated diol paclitaxel precursor synthetic pathway into two microorganisms, *Saccharomyces cerevisiae* and *Escherichia coli*, neither of which could produce paclitaxel precursor alone, providing a good idea for bioproduction in microbial community through the method for separation of spatial pathway module (Zhou et al., 2015). Zuroff et al. demonstrated co-culture of *S. cerevisiae* cdt-1 and *Clostridium phytofermentans* to manufacture ethanol up to 22 g/L under semi-aerobic conditions (Zuroff et al., 2013).

1.1.3.3 Wastewater treatment and environmental remediation

It is a promising study area on the use of microbial communities to address complicated chemical waste polluted locations, such as dye-stained water bodies and toxic heavy metal pollution (Ghosh et al., 2016). Different types of microbial communities can reduce complicated waste chemical pollutants more effectively than a single microbe (Waghmode et al., 2011). This might be because certain species cannot survive in poisonous environments, but they can survive when harmful substances are broken down by another species. *Sphingobium chlorophenolicum*, for example, is an excellent strain for degrading pentachlorophenol (PCP), a very harmful environmental contaminant, but it is Hg^{2+} sensitive, while *Ralstonia metallidurans* is a Hg^{2+} reducer. Kim et al. created a model in which *S. chlorophenolicum* was placed in the core layer was shielded from the exterior by *R. metalliduran*, allowing them to remove PCP and Hg^{2+} simultaneously (Kim et al., 2011). Negative interactions, such as alga-bacterium and alga-fungus bioflocculation, have been utilised to rapidly reduce the density of hazardous algal cells (Sun et al., 2018). According to one study, a bioflocculant generated by *Solibaccilus silvestris* could successfully flocculate *Nannochloropsis* sp., a free-floating marine microalgae, without the need of extra coagulants. This bioflocculant is significant since it can be regenerated and avoids secondary contamination (Wan et al., 2013; Alam et al., 2016).

1.2 Cyanobacteria and *Synechococcus elongatus*

1.2.1 History and evolution

Cyanobacteria (Figure 1-2), also called blue-green algae, are a unique phylum of prokaryotes with the ability of photosynthesis, which played a crucial role in the evolution of the biosphere and in the formation of atmospheric oxygen (Esteves-Ferreira et al., 2018). At least 1.05 billion years ago, oxygenic photosynthesis spread throughout several eukaryotic clades, resulting in various forms of algae and, subsequently, plants. The first endosymbiosis of a cyanobacterium within a unicellular eukaryote, as well as subsequent higher-order endosymbiotic processes, were responsible for this significant evolutionary stride (Demoulin et al., 2019). There are over 2000 different species of cyanobacteria in 150 genera, with a broad variety of

shapes and sizes (Vincent, 2009). Nowadays, cyanobacteria are still one of the most important primary producers on Earth (Ting et al., 2002).

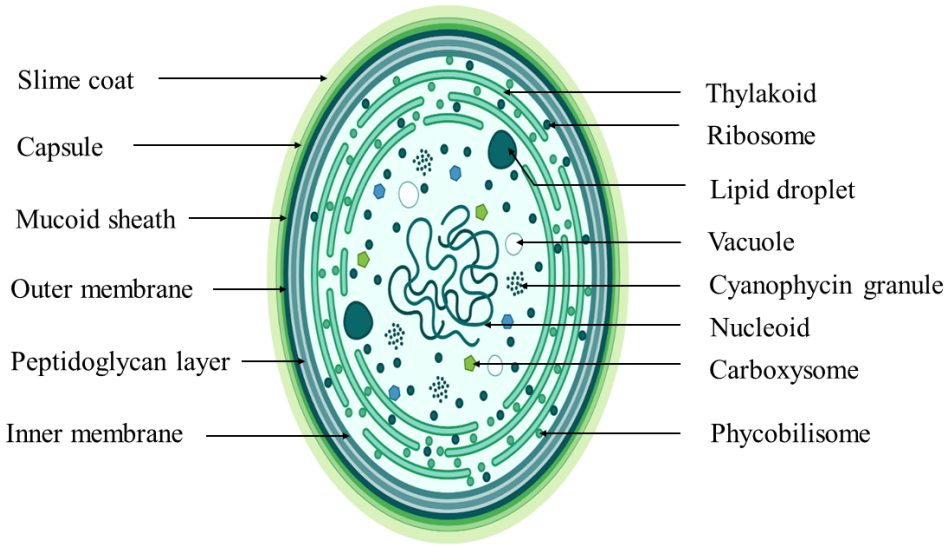


Figure 1-2 Schematic diagram of cyanobacteria showing the major organelles. Image generated from BioRender.

1.2.2 Carbon assimilation

Carbon fixation is the process through which organisms transform inorganic carbon (carbon dioxide) into organic substrates through the Calvin-Benson cycle (CB cycle).

1.2.2.1 CO_2 concentration mechanism (CCM)

Usually, the incorporation of inorganic carbon in the initial steps of CO_2 fixation in the CB cycle remains limited. Spatial confinement of CO_2 in the CCM is an effective strategy for promoting the availability of inorganic carbon (So and Espie, 1998). The CCM is composed of three key components {Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO), carbonic anhydrase (CA), and carboxysome}, which can introduce atmospheric carbon into the CB cycle (Figure 1-3) (Zhang et al., 2017).

RuBisCO is the key and rate-limiting enzyme for carbon fixation, which is selective for two substrates: CO_2 and O_2. One function of RuBisCO is to react with CO_2 to produce 3-phosphoglycerate (3PG) and the other is to initiate C2 photorespiration to produce

toxic 2-phosphoglycolate and release CO_2 (Andersson, 2008; Zhang et al., 2017). Therefore, the poor specificity of RuBisCO for CO_2 compared to O2 is a major factor in the slow fixation of CO_2 (Carmo-Silva et al., 2015). There have been many attempts on RuBisCO, e.g., improving its specificity towards CO_2 in *Synechococcus* PCC 7002 (Cai et al., 2014) or overexpression of RuBisCO in *Synechocystis* 6803 (Liang and Lindblad, 2017), which have been proved to increase the CO_2 fixation. CA also affects the catalytic efficiency of RuBisCO by mediating the hydration and dehydration of CO_2 and HCO_3^- to alter the CO_2 concentration within the carboxysome (Zhang et al., 2017). Almost all cyanobacteria use the carboxysome, a bacterial compartment filled with RuBisCO. Carboxysomes is important for co-localising RuBisCO and CA (Hickman et al., 2013). The presence of the cytosolic CA greatly obstructs carbon fixation rate and facilitates the diffusion of CO_2 out of the cell. Therefore, the CA must be located in the carboxysome for an effective CCM, which allows for the concentration of CO_2 near RuBisCO for carbon fixation to CB cycle (Zhang et al., 2017).

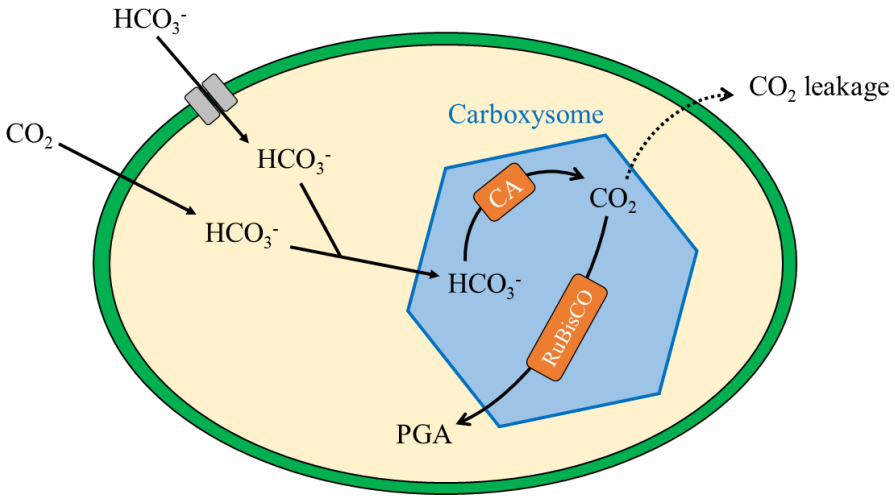


Figure 1-3 Schematic diagram of the cyanobacterial CO_2 concentrating mechanism (CCM). CO_2, carbon dioxide; HCO_3^-, bicarbonate; CA, carbonic anhydrase; RuBisCO, Ribulose-1,5-bisphosphate carboxylase/oxygenase; PGA, 3-Phosphoglyceric acid.

1.2.2.2 Oxygenic photosynthesis

Oxygenic photosynthesis is the process that occurs in plants, algae and cyanobacteria in which the solar energy is converted into chemical energy. It is divided into two steps: the light reaction and the dark reaction (Figure 1-4).

The complete photosynthetic reaction in cyanobacteria needs four main protein supercomplexes, including photosystem I (PSI, plastocyanin-ferredoxin oxidoreductase), photosystem II (PSII, water-plastoquinone oxidoreductase), cytochrome b_6f (plastoquinone-plastocyanin oxidoreductase), and F-ATPase (proton-motive force-driven ATP synthase) (Gao et al., 2018). The solar energy is captured by peripheral light-harvesting complexes (LHCs) and transferred to PSII (P680) and PSI (P700). PSII absorbs electrons by splitting water molecules into molecular oxygen, producing high-energy electrons which are transferred to the quinone pool (Grabolle and Dau, 2005; Johnson, 2016). Afterwards, the electrons are transferred to the thylakoid-embedded cytochrome b_6f complex, which oxidises plastoquinone and reduces plastocyanin (Heinz et al., 2016). Then PSI oxidises plastocyanin and reduce $NADP^+$ to NADPH with ferredoxin-$NADP^+$ reductase (FNR) (Figure 1-3) (Brettel and Leibl, 2001; Gao et al., 2018). The protons generated during the processes are used by F-ATPase to produce ATP, which is as the energy supplier for the dark reaction together with NADPH (Pfannschmidt 2003).

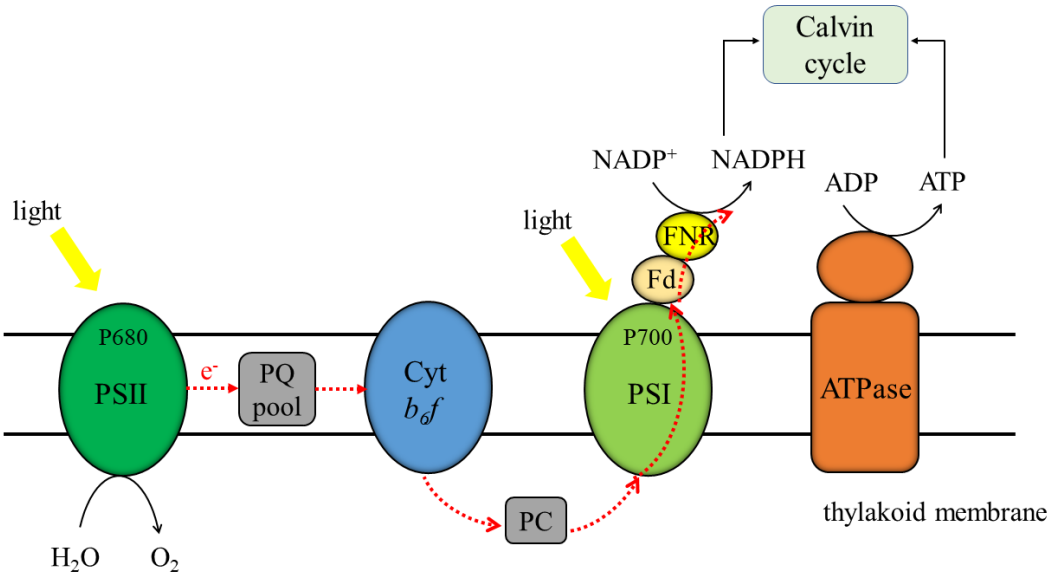


Figure 1-4 Schematic representation of the photosynthetic embedded in the thylakoid

membranes of cyanobacterial cells. Light energy is converted into chemical energy (NADPH and ATP) through four main protein supercomplexes, which is then used for carbon fixation. PSII, photosystem II; PQ, plastoquinone; Cyt, cytochrome; PC, plastocyanin; PSI, photosystem I; ATPase

1.2.3 Nitrogen assimilation and nitrogen control

1.2.3.1 Nitrogen assimilation

Cyanobacteria can use various types of nitrogen source, including inorganic ions such as nitrate and ammonium, as well as organic compounds such as amino acids (arginine, glutamine, and glutamate), urea, and nitrogen-containing bases. Some filamentous cyanobacteria can also assimilate atmospheric nitrogen in the heterocyst through nitrogen fixation pathway (Flores and Herrero, 2005). The preferred nitrogen source for cyanobacteria is ammonium, while glutamine is also a good nitrogen source for many bacteria (Herrero et al., 2001).

The entry of nitrogenous compounds into the cell, usually at low concentrations (e.g., below 1 μM) in the environment, occurs via permeases located in the cytoplasmic membrane. The different nitrogen sources are absorbed via ABC (ATP-binding cassette)-type permeases (Figure 1-5). Nrt permease, encoded by *nrtA*, *nrtB*, *nrtC*, and *nrtD*, can actively transport both nitrate and nitrite (Valladares et al., 2002). Urt, encoded by *urtA*, *urtB*, *urtC*, *urtD*, and *urtE*, is a high-affinity urea transporter, which can take up urea at lower concentrations (Esteves-Ferreira et al., 2018). And ammonium is transported by a secondary transporter called Amt (Vázquez-Bermüdez et al., 2002). Then the nitrogen sources are metabolised to ammonium through different enzymes, such as ferredoxin-dependent nitrate and nitrite reductases and Ni^{2+} dependent urease for nitrate and urea, respectively, which is then incorporated into the carbon skeletons via the glutamine synthetase-glutamate synthase (GS/GOGAT) pathway (Flores and Herrero, 2005).

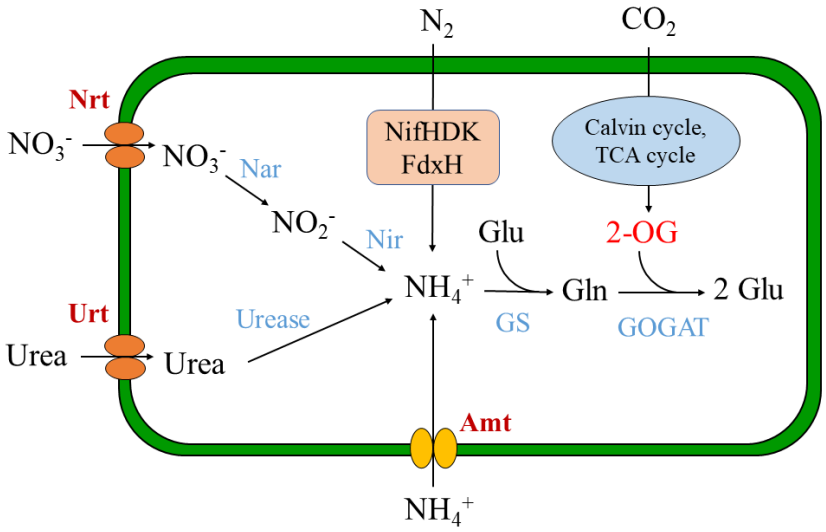


Figure 1-5 Different nitrogen assimilation pathways in cyanobacteria. Adapted from ref (Flores and Herrero, 2005). Different types of nitrogen are uptake by transporters and incorporated into the carbon skeletons via the GS/GOGAT pathway. Nrt, ABC-type nitrate/nitrite transporter; Urt, ABC-type urea transporter; Amt, ammonium transporter; NifHDK, nitrogenase complex; FdxH, heterocyst-specific ferredoxin; Nar, nitrate reductase; Nir, nitrite reductase; GS, glutamine synthetase; GOGAT, glutamate synthase; Glu, glutamate; Gln, glutamine; 2-OG, 2-oxoglutarate.

1.2.3.2 Nitrogen control

Nitrogen control is a common biological process among microorganisms, where cells tend to utilise more easily assimilated source of nitrogen when it is available and repress other nitrogen assimilation pathways (Herrero et al., 2001). For example, the presence of ammonium in the medium leads to a decrease in the abundance of nitrogen assimilating enzymes and inhibits the activity of the combined nitrogen transport system (Forchhammer, 2004; Esteves-Ferreira et al., 2018). There are three main nitrogen control systems for microorganisms, including NtrB-NtrC two-component system of enteric bacteria and some other proteobacteria (Merrick and Edwards, 1995), GATA family cofactor transcription cofactors of fungi and yeast (Marzluf, 1997), and a novel nitrogen control system NtcA found in *Bacillus subtilis* and cyanobacteria (Herrero et al., 2001).

Nitrogen metabolism control in cyanobacteria depends mainly on the nitrogen-control transcription factor NtcA. The NtcA transcriptional regulator belongs to bacterial transcription factors and has a similar structure of CAP (catabolite gene activator protein), a representative transcriptional factor in *Escherichia coli* (Wisén et al., 2004). NtcA is autoregulatory and its activity is influenced by 2-oxoglutarate (2-OG) and a signal transduction protein P_{II} (*glnB* gene product) (Flores and Herrero, 2005; Esteves-Ferreira et al., 2018). The P_{II} protein is a 2-OG sensor and can interact with 2-OG depending on 2-OG concentration, leading to the conformational changes of PII and releasing the PII interacting X (PipX) from PII to interact with NtcA (Llácer et al., 2010; Esteves-Ferreira et al., 2018). As described above, nitrogen control in cyanobacteria is achieved through P_{II} sensing 2-OG concentration, which acts as an important member of the carbon skeleton for incorporation nitrogen into organic materials. Therefore, nitrogen control is also an important component of carbon/nitrogen homeostasis in cyanobacterial cells.

1.2.4 Application of cyanobacteria

Cyanobacteria are one of the most promising organisms for capturing carbon dioxide, which is capable of utilising photosynthetic system to absorb CO_2 and convert it into bioproducts (Zhang et al., 2017). In addition, new metabolic pathways can be created through metabolic engineering to redirect carbon into desired chemical production. Alcohols, such as ethanol (Dexter and Fu, 2009), isopropanol (Kusakabe et al., 2013), and 1-butanol (Lan and Liao, 2011), are the most studied chemicals produced in the cyanobacterial strains *Synechocystis* sp. 6803 and *Synechococcus elongatus* 7942.

Photosynthetic cyanobacteria, capable of producing sucrose, are also good carbon source supply candidates for microbial communities. Carbon sink has been engineered to enhance the photochemical efficiency in cyanobacteria. A good example is the increase of sucrose production in *S. elongatus* PCC 7942 by expressing heterologous sucrose permease gene cscB from *Escherichia coli*, which allows sucrose export in high salt concentrations (Ducat et al., 2012; Hays and Ducat, 2015). Sucrose phosphate synthase (SPS) is one of the key enzymes during sucrose synthesis and overexpression of SPS increased sucrose productivity by 93% compared to the wild-type *S. elongatus* PCC 7942 (Abramson et al., 2016; Duan et al., 2016). Due to the good capability of

sucrose production, engineered *S. elongatus* PCC 7942 has been used as a carbon supplier to co-cultured with heterotrophs, such as Bacillus subtilis, Escherichia coli, and Saccharomyces cerevisiae (Hays et al., 2017).

1.2.5 *Synechococcus elongatus* PCC 7942

Synechococcus elongatus PCC 7942, also called *Anacystis nidulans* R2, is a unicellular, freshwater cyanobacterium. It is a rod-shaped non-nitrogen-fixing cyanobacterium about 2 μm in length that is commonly used as a model strain (Smith and Francis, 2016a). *Synechococcus elongatus* PCC 7942, with a genome of approximately 2.7 Mb, was the first cyanobacterium to be shown a reliable transformation by exogenous addition of DNA (Shestakov and Ti, 1970). Over the decades, many genetic tools for *S. elongatus* have been created and applied to other transformable cyanobacteria (Santos-Merino et al., 2019).

As one of the cyanobacteria, *S. elongatus* PCC 7942 has the capacity to produce chemicals and biofuels, including acetone (Chwa et al., 2016), glycerol (Hirokawa et al., 2017), fatty acids (Santos-Merino et al., 2018), etc. But current research on the synthesis of sucrose by *S. elongatus* has become more attractive as described above since sucrose can supply direct carbon source for many species (Abramson et al., 2016; Hays et al., 2017).

1.3 Azotobacter vinelandii

1.3.1 Introduction to *A. vinelandii*

Azotobacter vinelandii is a widespread soil bacterium belonging to the *Pseudomonas* genus that can fix nitrogen (Baars et al., 2016). It is a rod-shaped gram-negative bacterium and about 2um in length, with a single circular genome of approximately 5.4 Mb, and has been studied for over 100 years (Setubal et al., 2009; Smith and Francis, 2016a; Noar and Bruno-Bárcena, 2018). *Azotobacter vinelandii* is an obligate aerobic bacterium that differs it from many other nitrogen fixing microorganisms that demand low oxygen concentrations to prevent nitrogenases from oxygen damage (Villa et al., 2014; Barney et al., 2015). As an established and genetically tractable model organism,

Azotobacter vinelandii has been used for research of nitrogen fixation and siderophore production (Baars et al., 2016).

1.3.2 Nitrogen fixation and nitrogenase

Nitrogen is a crucial element of microorganisms and is especially important in the synthesis of cellular components like nucleotides and proteins. Biological nitrogen fixation (BNF) is the process that microorganisms reduce molecular nitrogen (N_2) from atmosphere to ammonium or other nitrogen-containing compounds (Alleman et al., 2021). There are many different types of microbes with the capacity to fix nitrogen. One major type is symbiotic nitrogen-fixing microorganisms, such as *Azospirillum* sp., *Rhizobium* sp., and *Frankia* sp., which usually grow with plants. The others are free-living microorganisms, including *Azotobacter* sp. and some of the *Clostridium* sp., *Bacillus* sp., *Pseudomonas* sp., *Klebsiella* sp., and Cyanobacteria (Yan et al., 2008; Sivasakthi et al., 2017).

Nitrogenase, which is encoded by *nif* genes, is the essential enzyme involved in the nitrogen fixation process. It is a highly conserved metalloenzyme and can be divided into three classes according to the different types of Fe-S clusters: iron-iron (FeFe), vanadium-iron (VFe) and molybdenum-iron (MoFe) nitrogenases (Kraepiel et al., 2009; Kuypers et al., 2018). All nitrogenase enzymes are composed of two components. Component I, encoded by, nifDK, vnfDGK or anfDGK, contains the active site for nitrogen reduction with Mo, V or Fe in the active centre, respectively. Component II, encoded by nifH, vnfH or anfH, combines ATP hydrolysis with interprotein electron transfer (Zehr et al., 2003). Because the Mo-containing enzyme has been the most thoroughly researched, nifH, which encodes electron transfer proteins, is used to identify nitrogen-fixing microbes in the environment (Kuypers et al., 2018).

A. vinelandii is an aerobic soil bacterium that can convert N_2 to ammonium by using any of the three different types of nitrogenases. FeFe and VFe nitrogenases are suppressed by Mo, whereas Mo-containing nitrogenase allows for the expression in medium containing Mo. FeFe nitrogenase is expressed in a medium lacking both Mo and V, while VFe nitrogenase is expressed in Mo-deficient medium containing V (Maynard et al., 1994).

1.3.3 Siderophores

Siderophores are small molecules produced under trace metals limitation environments, particularly iron (Fe), to help deliver these metals into organisms for their growth and production (McRose et al., 2018). *A. vinelandii* is a nitrogen-fixing bacterium, which encodes Fe-, Mo-, and V- nitrogenase. Siderophores produced in *A. vinelandii* are crucial for the acquisition of nitrogenase co-factors that are needed for nitrogen fixation.

Different types of siderophores have been identified in gram-negative bacteria (Cornelis et al., 2009; Khasheii et al., 2021). *Azotobacter* can produce two types of siderophores in response to Fe limitation (Figure 1-6). One of the types is pyoverdine siderophores, e.g., azotobactin, which exhibit specific absorbance and fluorescence at 380 and 500 nm, respectively (Menhart et al., 1991; Yoneyama et al., 2011). The other type is catechol siderophores, including azotochelin, aminochelin, protochelin, and 2,3-dihydroxybenzoic acid (DHBA), which have a specific absorbance at 310 nm (Bellenger et al., 2007). According to the research by McRose et al. (2017), the production of all siderophores increased under Fe-limitation, while only catechol siderophore production increased under Mo-limitation, with protochelin exhibiting the highest response (McRose et al., 2017). This siderophore uptake system enables *A. vinelandii* to absorb trace elements in a variety of metal-limited environments and avoid harmful metals like tungsten (W), which have no biological benefit in this organism (Wichard et al., 2009).

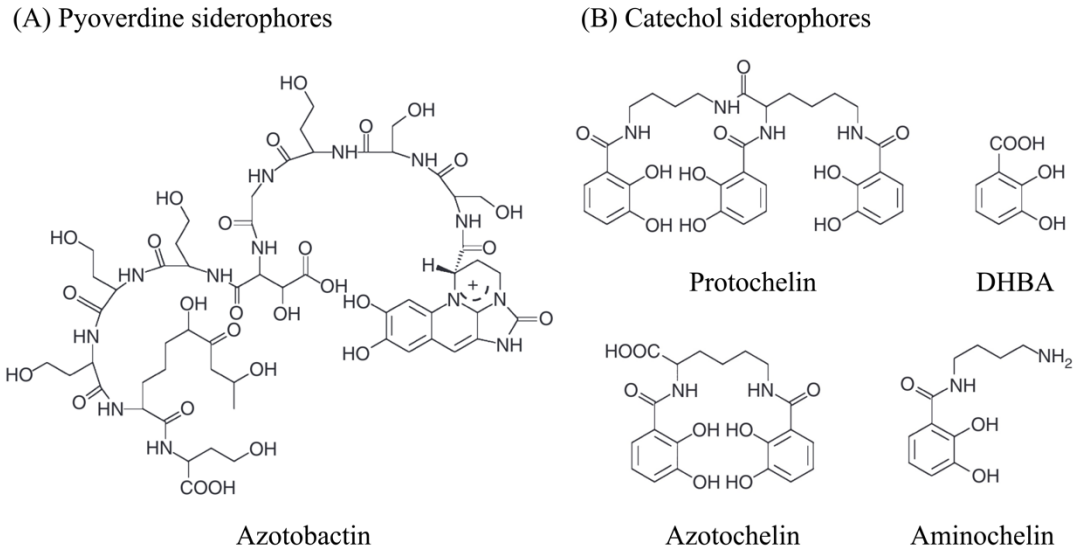


Figure 1-6 Two types of siderophores produced in *A. vinelandii*. (A) Pyoverdine siderophores, with specific absorbance and fluorescence at 380 and 500 nm, respectively. (B) Catechol siderophores, with specific absorbance at 310 nm (Yoneyama et al., 2011).

1.3.4 Applications of *A. vinelandii*

A. vinelandii, as one of the important aerobic free-living nitrogen-fixing bacteria, has a good effect on plant nutrition and biological soil fertility (Aasfar et al., 2021). It is also a good model organism, which has been used to research aerobic nitrogen fixation, respiration, microbial physiology, hydrogen generation and absorption, and other enzyme kinetics (Noar and Bruno-Bárcena, 2018). Several research focusing on the nitrogen fixation capability of *A. vinelandii* have been undertaken using major enzymes from the ammonium production and transport system. For example, Bali et al. discovered a rise in ammonia when the *nifL* gene was damaged but the promoters remained intact (Bali et al., 1992). Based on their findings, Barney et al. introduced a kanamycin cassette upstream of the SmaI gene, resulting in a spontaneous mutation in the Nif+ phenotype and enhanced nitrogen release (Barney et al., 2015). In addition, because of the genetically tractable traits and natural competence, *A. vinelandii* is also of good interest to produce multiple products, e.g., hydrogen (Noar et al., 2015), alginate (Mærk et al., 2020), and polyhydroxybutyrate (Yoneyama et al., 2015).

Moreover, the ability of ammonium production allows it as a good nitrogen provider in mutualistic or symbiotic conditions. Apart from ammonium, siderophores produced from *A. vinelandii* have also been verified as a source of nitrogen to support the growth of some green algae, such as *Neochloris oleoabundans* and *Scenedesmus* sp. BA032 (Villa et al., 2014).

1.4 Techniques for microbial interaction analysis

The systemic complexity of synthetic microbial consortia increases the challenge of studying their interactions. However, with the development of system biology and advancement in genomics, transcriptomics, proteomics, and metabolomics techniques, it has become more feasible to examine the interactions among different members in the microbial communities.

1.4.1 16s rRNA sequencing

For decades, the 16S rRNA gene has been a cornerstone of sequence-based bacterial study. The 16S rRNA gene is a highly conserved constituent of the transcriptional machinery of all DNA-based living forms, making it ideal as a target gene for DNA sequencing in samples including thousands of different species (Johnson et al., 2019). Universal PCR primers can be constructed to target the conserved sections of the 16S gene, allowing the gene to be amplified in a broad variety of microorganisms from a single sample. 16S rRNA gene sequence analysis can be employed to identify badly documented, poorly isolated, or phenotypically abnormal strains, and can lead to the identification of new pathogens and non-cultured bacteria (Clarridge, 2004). It has been extensively used to identify, classify, and quantify microorganisms in complex biological mixes such as environmental communities (Boaro et al., 2014; Koo et al., 2017) and human microbiome (Ames et al., 2017; Huang et al., 2020).

1.4.2 Transcriptomics

Transcriptomics is the study of the transcriptome, which is the collection of all

ribonucleic acid (RNA) in a sample (a cell, tissue, or organ) at any particular moment (Saliba et al., 2014). RNA supports various tasks within the cell, and analysing the transcriptome reveals how genes operate and whether proteins are created as predicted. Transcriptomics technique has been applied to the analysis of interactions between some microbial cocultures. For example, transcriptomics of a co-culture between *Pseudo-nitzschia multiseries* and a *Sulfitobacter* species revealed an up-regulation of cell division in the diatom, promoting by the hormone indole 3-acetic acid production in the bacterium and an up-regulation of tryptophan biosynthesis genes in the diatom in the meanwhile (Amin et al., 2015). However, its application is limited by the fact that mRNA abundance usually provides little information about protein activity and is not a substitute for detailed functional analysis of genes (Feder and Walser, 2005).

1.4.3 Modelling

It is difficult to recognise microbial communities and their interactions sometimes due to current technical limitations in detecting possible metabolite exchange fluxes between various species in microbial communities (Hanemaaijer et al., 2017; Ang et al., 2018). Built models may be simulated in this regard to assist address knowledge gaps and produce tested hypotheses for designing new trials. There have been attempts at mathematical models and numerical simulation to uncover how the stability and productivity of the synthetic consortia are affected by the system structure and parametrisation (Johns et al., 2016; Di and Yang, 2019). Other community models have also been created to investigate microbial interactions in a variety of ecological niches, including the host microbiome (Greenblum et al., 2012), wastewater treatment (Palatsi et al., 2010), biogeochemistry and bioremediation (Hamilton et al., 2015), and synthetic complexes (Ye et al., 2014).

1.5 Proteomics

The term ‘proteome’ was first coined in the 1990s by Marc Wilkins to describe the entire protein in an organism (Wilkins, 2009). Proteomics is the large-scale study and analysis of proteomes with the goal of determining protein activities in the biological environment (Landels et al., 2015). Proteomics may be classified into two types based

on their purpose: discovery proteomics and focused proteomics. Targeted proteomics focuses on particular proteins, allowing for great sensitivity, quantitative precision, and consistency in detecting target proteins. Peptides are formed by proteolytic cleavage of the targeted protein in targeted proteomics studies, and a selective reaction tracking assay has been used to quantify these peptides by LC-MS/MS (Chen and Liu, 2019). Discovery proteomics, also called shotgun proteomics, is a protein profiling technique. It is mostly utilised for global proteomic screening and proteome comparisons between cells under various physiological circumstances (Guo, 2015). Discovery proteomics is no longer focused on finding specific protein targets, and the broad spectrum of proteins that may be discovered and quantified is a crucial indicator in this sort of investigation.

1.5.1 Proteomic analysis strategies

To date, there are three main approaches to identify and characterise proteins, i.e., bottom-up (BU), top-down (TD), and middle-down (MD) (Figure 1-7).

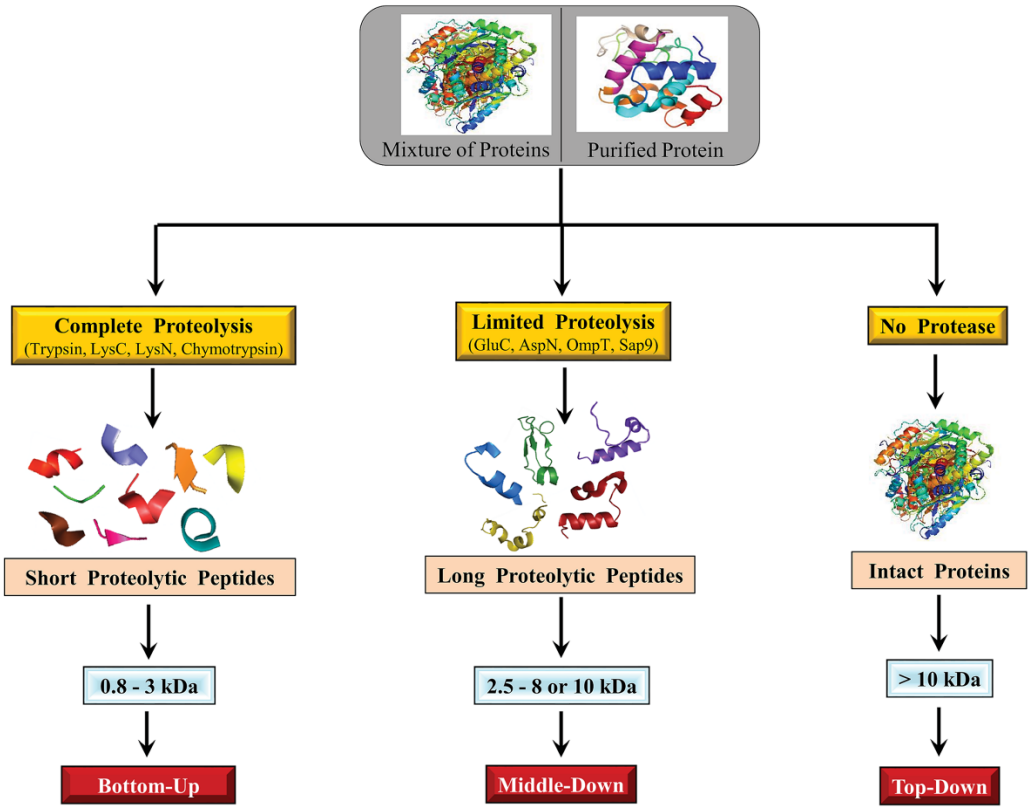


Figure 1-7 Proteomics strategies shows different proteomic analysis approaches: bottom-up, middle-down, and top-down(Pandeswari and Sabareesh, 2019). The bottom-up method focuses on proteolytic peptides. The middle-down method analyses longer proteolytic peptides. The top-down method analyses the intact proteins.

1.5.1.1 Bottom-up approach

The bottom-up approach, also called shotgun proteomics, is the most widely used method. It refers to the identification and characterisation of proteins by analysis of proteolysis peptides (Zhang et al., 2013). A typical bottom-up proteomics workflow generally involves five steps (Figure 1-8): a) protein extraction and isolation from cell lysate or tissue; b) protein digestion by proteases; c) peptides fractionation and ionisation; d) mass spectra generation; e) MS/MS spectra generation (Aebersold and Mann, 2003). Protein identification is achieved by comparing the peptide spectra generated from a mass spectrometer with theoretical peptides digested on the protein database *in silico*.

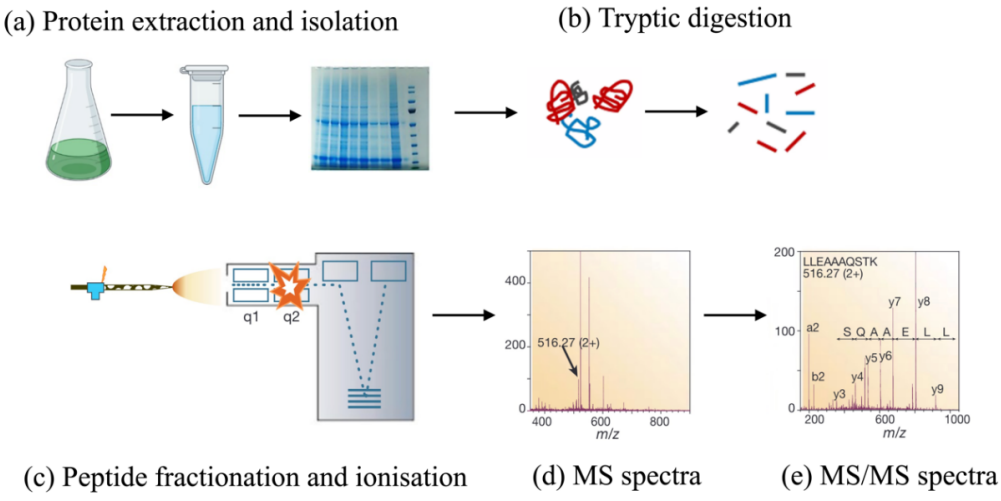


Figure 1-8 Typical bottom-up proteomics workflow. A standard bottom-up proteomics experiment consists of five stages. a) protein extraction from cell lysate or tissue samples, usually followed by a step for protein quality evaluation and separation, e.g., one-dimensional gel electrophoresis; b) protein digestion by proteases, usually by trypsin, producing peptides with C-terminal protonated amino acids, which is better in subsequent peptide sequencing; c) peptides fractionation and ionisation. The peptides

are separated by high-pressure liquid chromatography (LC) in very fine capillaries and eluted into an electrospray ion source where they are charged; d) mass spectra generation; and e) MS/MS spectra generation. Adapted from ref (Aebersold and Mann, 2003).

In the global proteomics analysis, the extraction and separation of proteins from their chemical and physical interactions with other biomolecules in specific cellular sub-compartments have become a vital step (Zhang et al., 2013). The addition of detergent can greatly promote protein solubility in the solution, especially for hydrophobic proteins. Depending on their ionic properties, detergents can be classified as ionic detergents, non-ionic detergents and zwitterionic detergents (Arachea et al., 2012). Ionic detergents, such as sodium dodecyl sulfate (SDS), are highly denaturing agents with the special property of separating proteins into their monomeric form and are therefore commonly used for western blot and molecular weight analysis. Non-ionic detergents, such as Triton X-100, are less effective as protein denaturing agents and are often used for protein-protein interactions. Zwitterionic detergents, such as CHAPS, which have both negatively and positively charged heads, are more effective than non-ionic detergents in protein-protein interactions, with less protein denaturation than ionic detergents. However, most surfactants are detrimental to the sensitivity of MS peptides, and it is important to remove them in the later step (Zhang et al., 2013).

The enzymatic digestion of proteins into specific peptides is one of the key steps in obtaining accurate proteomic data. The specificity of different proteolytic enzymes for cleaving the amide bonds between individual protein residues varies (Gundry et al., 2010). A commonly used protease is trypsin, which cleaves on the carboxyl side of arginine and lysine. It is possible to produce more specialised peptides for various uses by the combination of various proteases. For instance, combining highly selective and non-selective proteases enhances protein coverage and post-translational modifications for samples with minimal complexity, such as protein complexes. Combining highly selective proteases enhances protein and proteome coverage and sensitivity for complicated proteomic samples by producing complementary peptides (Zhang et al., 2013).

The two most prevalent methods for ionising proteins or peptides for mass

spectrometric analysis are electrospray ionisation (ESI) and matrix-assisted laser desorption/ionisation (MALDI) (Aebersold and Mann, 2003). ESI is a gentle ionisation method that is widely utilised to generate gas phase ions of thermally labile massive supramoleculars without fragmentation (Banerjee and Mazumdar, 2012). It ionised the analytes out of solution, thus tandem mass spectrometry can be established by coupling to liquid-based separation methods such as chromatography and electrophoresis. In MALDI, the analyte is incorporated in an acidic matrix that can absorb UV light heavily. Short laser pulses cause a portion of the matrix to quickly heat up and evaporate/ionise with the analyte (Dreisewerd, 2003; Nadler et al., 2017). However, the low time efficiency of instruments has limited the development of MALDI mass spectrometry for complicated proteomics (Nadler et al., 2017).

1.5.1.2 Top-down approach

The top-down approach, in contrast, focuses on intact proteins that are up to 200 kDa (Han et al., 2006). The intact protein ions are introduced into the gas phase by electrospray ionization, followed by fragmentation, usually by collision-induced dissociation (CID) or by electron capture dissociation (ECD) or electron transfer dissociation (ETD) in a mass spectrometer, resulting in mass protein and fragment ions for protein identification (Timp and Timp, 2020). Compared to BU approach, the top-down proteomics approach is more useful for post-translational modifications (PTMs) and protein isoforms determination. However, it has significant limitations due to difficulties with protein fractionation, protein ionization and fragmentation in the gas phase and is about 100-fold less sensitive than BU approach (Zhang et al., 2013). With the increase of protein molecular weight (*Mw*), the fragmentation efficiency of intact proteins decreases due to the complexity of the tertiary structure. Therefore, most top-down applications focus on proteins smaller than 70 kDa, with only a few applicable to larger proteins (>100 kDa) (Siuti and Kelleher, 2008; Timp and Timp, 2020).

1.5.1.3 Middle-down approach

Middle-down proteomics is an emerging approach that combines the methodologies and instrumentation of BU and TD approaches to overcome their drawbacks. Like BU approach, MD approach also analyse proteolytic peptides, cleaved with special

proteases such as OmpT, Sap9, and IdeS (Wu et al., 2013; Pandeswari and Sabareesh, 2019). These proteases result in larger proteolytic peptides, with *Mw* of 2.5-10 kDa. Thus, the number of proteolytic peptides generated by middle-down proteomics would be less than by BU approach, which reduce the complexity of the samples. In the meantime, there is a greater probability of detecting more unique peptides by an MD approach due to longer peptides (Pandeswari and Sabareesh, 2019). In addition, large peptide fragment generation leads to similar advantages as TD proteomics, such as further insight into PTMs, without the challenges of analysing entire proteins (Zhang et al., 2013).

1.5.2 Mass spectrometer

A mass spectrometer is an instrument used to determine the mass-to-charge ratio (m/z) of gas-phase ions. It usually consists of three modules: a) An ion source, which generates ions from gas phase sample molecules (or moves ions in solution into the gas phase in the case of ESI); b) A mass analyzer, which uses electromagnetic fields to sort the ions by their masses; c) A detector, which provides measurements indicating the abundance of ions with different mass-to-charge ratios (Patel et al., 2012). There are many different types of mass spectrometers that achieve accurate mass analysis of samples, including quadrupole, ion trap, time-of-flight (TOF), and ion cyclotron resonance (ICR). Different forms of mass spectrometers are usually connected in series to achieve complementary advantages, e.g., Q-TOF and Q-Exactive mass spectrometer.

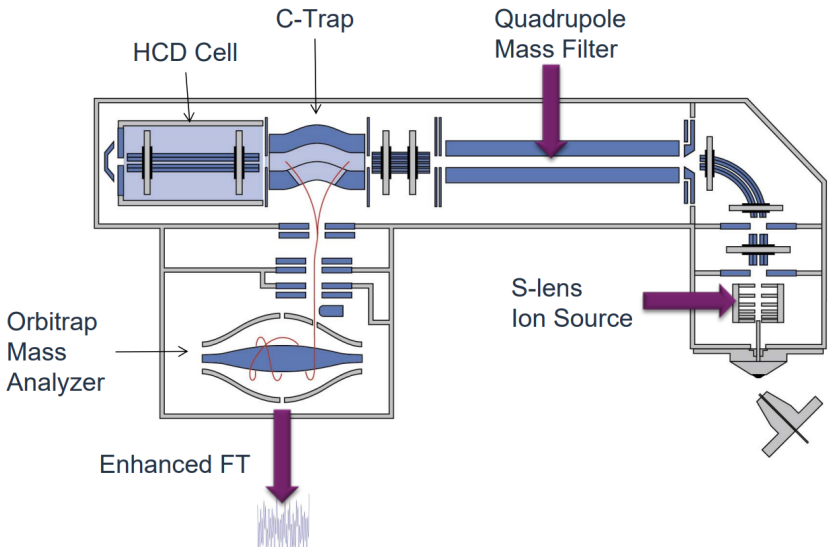


Figure 1-9. Schematic example of the Q-Exactive Hybrid Quadrupole-Orbitrap mass spectrometer (http://www.thermoscientific.com/). It is composed of a few components, including ion source, stacked-ring ion guide (S-lens), quadrupole mass filter, curved linear trap (C-trap), Higher Energy Collisional Dissociation (HCD) cell, and Orbitrap mass analyzer. Various ways can be used to introduce samples into the ion source. Ions are sent from the source to the quadrupole through the injection flatapole. The quadrupole rod assembly functions as an ion transmission device, with the ability to filter transmitted ions based on their mass-to-charge ratios. To get mass spectra, the ions are transported into the C-Trap and subsequently injected into the Orbitrap mass analyzer. Ions are also transported through the C-Trap into the HCD cell to conduct MS/MS investigations using the quadrupole mass filter.

1.5.3 Mass spectrometry-based proteomics

Quantitative proteomics is a strong method for discovery and target proteomic research to comprehend the overall proteome dynamics in cells, tissues or organisms. Proteomic quantification can be accomplished using either relative or absolute quantitative approaches. Relative protein quantification techniques, such as stable isotope labelling by amino acids in cell culture (SILAC), isobaric tags for relative and absolute quantification (iTRAQ), isotope-coded affinity tags (ICAT), isotope-coded protein label (ICPL) and isobaric tags, couldcompare the relative quantity of proteins or peptides in samples and characterise proteomic kinetics in cells. While absolute

measurement of intracellular proteome concentrations is essential for a thorough knowledge of organism metabolism and mathematical modelling in systems biology (Arike et al., 2012). There are three main techniques for quantitative proteomic profiling, including 2D-PAGE, stable isotope labeling, and label-free approaches (Figure 1-10). Among them, label-free and label-based approaches are the most popular mass spectrometry-based proteomics because of their efficiency and high throughput.

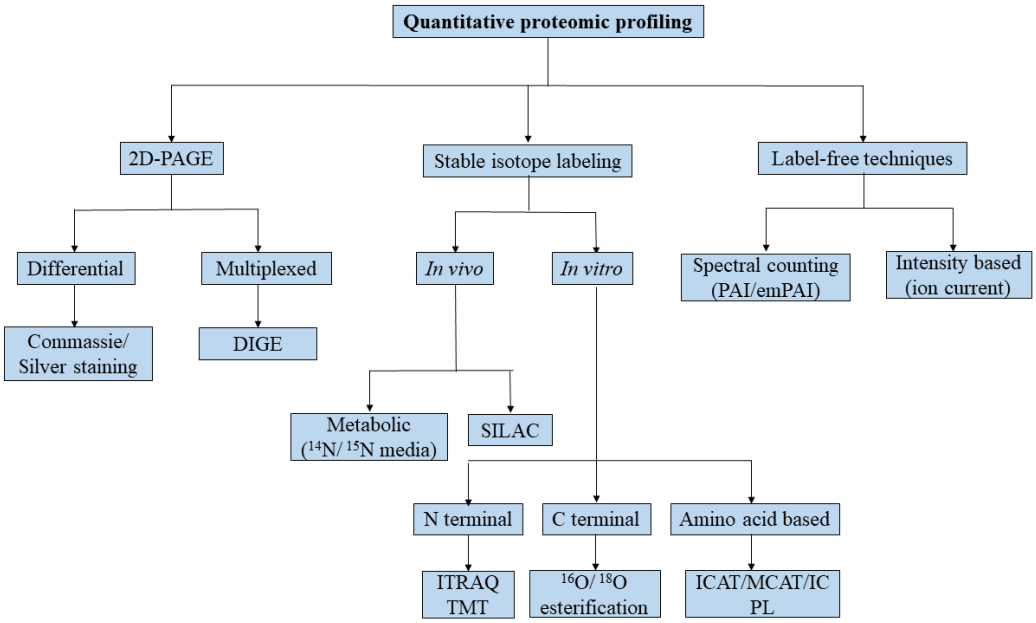


Figure 1-10 Different methods of quantitative proteomic profiling. Adapted from ref (Yan and Chen, 2005).

1.5.3.1 Label-free proteomics

Label-free proteome quantification approach provides straightforward method for large-scale analysis of biological samples, which has gained more acceptance in many research areas due to the less cost and less time demand (Anand et al., 2017). According to the quantification methods, label-free quantification can be divided into spectral counting-based and ion intensity-based quantification. Spectral counting-based method quantifies relative protein abundance by comparing the number of detected MS/MS spectra (e.g., spectral counts, peptide spectrum matches {PSMs}, etc) from the same protein in each of the numerous LC-MS/MS datasets (Zhu et al., 2010). Whereas this approach has the benefit of being simple and capable of quantifying proteins for which no peptides are detected in one condition, it has become somewhat outdated for MS-

based quantification since accuracy can be a problem, especially when comparing minor variations in abundance (Goeminne et al., 2018). Intensity-based method refers to protein quantification using the spectral area under the peaks in MS or MS/MS spectra, typically such as intensity, iBAQ intensity (intensity Based Absolute Quantification), and LFQ (label-free quantitation) intensity, which makes it more accurate than spectral counting-based method (Zhu et al., 2010). IBAQ technique was verified for the greatest correlation between biological replicates and the lowest change in ribosomal protein abundances in *E. coli* (Arike et al., 2012). Label-free proteome quantification technique is widely used because of the simultaneous identification and quantification of proteins without the time-consuming and expensive procedure of adding stable isotopes into samples, and it is applicable to samples from any source (Li et al., 2012). Label-free techniques, on the other hand, offer restricted quantification performance in respect of precision, accuracy, and reproducibility since samples are prepared and tested individually.

1.5.3.2 Label-based proteomics

In label-based strategies, samples are differentially labelled, pooled, and then subjected to mass spectrometry analysis and quantification, which minimises the predicted differences when treating the samples separately (Bantscheff et al., 2007; Anand et al., 2017). The most frequently utilised labelling methods include metabolic labelling, chemical labelling, and enzymatic labelling. Metabolic labelling was initially achieved by culturing bacteria in controlled ^{15}N-enriched medium (^{15}N labelling) (Oda et al., 1999). It became more popular after the SILAC approach developed in 2002 (Ong et al., 2002). However, many biological systems, such as natural microbial communities, are not susceptible to effective metabolic labelling (Verberkmoes et al., 2009). To address this, chemical or enzymatic techniques for labelling proteins or peptides with distinct isotopic tags have been devised. Extracted proteins, for example, can be labelled with ICAT following cell lysis (Gygi et al., 1999). Another two isobaric chemical labelling techniques, tandem mass tag (TMT) and iTRAQ, have recently gained popularity for quantitative proteomics (Thompson et al., 2003; Ross et al., 2004). Following proteolysis, samples are labelled with various isotopic variations of TMT or iTRAQ before being mixed for LC-MS/MS analysis. Both two tags have three functional components: a reporter ion group, a mass normalisation group, and an amine-

reactive group (Li et al., 2012). The quantification depends on the intensity of reporter ions, which are generated and detected at MS^2 level (Megger et al., 2014). Enzymatic labelling can also achieve in vitro, which has been taken place during proteolytic digestion or after digestion in a second incubation stage with the protease (Bantscheff et al., 2007). However, the fact that complete labelling is rarely obtained and that various peptides integrate the label at varying rates hampers data processing is a practical drawback, which limits its application (Ramos-Fernández et al., 2007).

1.5.4 Proteomics applications in biological research

1.5.4.1 Profiling of protein expression

Nowadays, the large quantity of information produced by traditional techniques based on MS-based methodologies has demonstrated that proteomics has emerged as one of the most efficient methods to conduct biological study. The most important goals of proteomics are to map all of the proteins in a biological system and to understand how they re-regulate spatiotemporal interference (Altelaar et al., 2013). It is also the foundation for effective microorganism co-culture. Helliwell et al., for example, used comparative proteomics to demonstrate the fundamental alterations in the metabolism of *Lobomonas rostrata* when co-cultured with *Mesorhizobium loti* strain (Helliwell et al., 2018).

1.5.4.2 Protein-protein interactions

Proteins often interact with one another in multiprotein complexes of varying composition (Altelaar et al., 2013). It is crucial to understand the distinct and dynamic proteome, and this necessitates the creation of a physical interaction chart. Many investigations on protein-protein interactions have been conducted since the 1989 publication of the yeast two-hybrid system (Sardiu and Washburn, 2011). Affinity purification-mass spectrometry (APMS), which uses purified proteins from cell extracts, is the most often used approach for examining protein-protein interactions. After purification, mass spectrometry will be used to determine the protein attached to the pure protein (Sardiu and Washburn, 2011). This method may be used to determine protein interactions within cells. A variety of affinity purification methods have been used, the most common one is tandem affinity purification (TAP) (Völkel et al., 2010).

1.5.4.3 Post-translational modifications (PTMs) analysis

The process of covalently processing translated proteins, such as acetylation, glycosylation, phosphorylation, and ubiquitination, is referred to as post-translational modification. Nowadays, the study of PTMs utilising proteomics methods is gaining interest, particularly in phosphorylation, the most ubiquitous and extensive alteration in organisms. Wei et al., for example, utilise dimethyl-labeled, $Ti4^{+}$-immobilized metal ion affinity chromatography (IMAC) proteomics on *Dunaliella salina* palmella formation and discovered 151 salinity response proteins and 35 salinity-responsive phosphoproteins involved in a variety of signalling and metabolic pathways on palmella formation (Wei et al., 2017).

1.5.5 Proteome bioinformatics

Bioinformatics has become an increasingly significant component of proteomics, with the generation of more complex data sets based on comprehensive mass spectrometry-based proteomics. A typical workflow for a systematic approach to identify MS data contains a few steps, including protein database generation, protein profiling, and functional annotation (Figure 1-11). Some recently created bioinformatics pipelines and software programmes have begun to aid in the standardisation and simplification of proteomics analysis, shifting complicated and powerful approaches from professionals to users. Examples are ProteinPilot (https://sciex.com/products/software/proteinpilot-software) (Jagtap et al., 2012), Mascot (http://www.matrixscience.com/) (Palmblad et al., 2007), Galaxy (https://usegalaxy.eu/) (Jagtap et al., 2015), MaxQuant (https://www.maxquant.org/) (Tyanova et al., 2016), and MetaProteomeAnalyzer (http://www.mpa.ovgu.de/) (Heyer et al., 2019),.

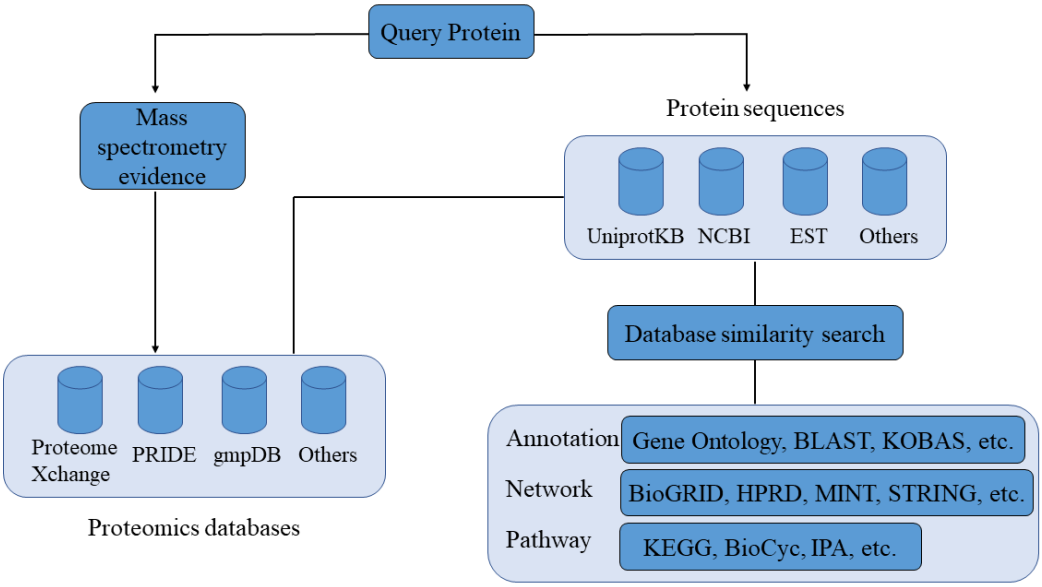


Figure 1-11 Overall workflow for mass spectrometry-based data analysis and proteome functional annotation.

1.6 Thesis motivation

1.6.1 Strains selection

Considering carbon and nitrogen are the two most costly nutrients during microbial cultivation and bioproduction, carbon and nitrogen fixing bacteria provide good candidates for artificial microbial consortia. Therefore, the photosynthetic cyanobacterium *S. elongatus* cscB/SPS reported in Abramson *et al.* 2016, which is a derivative of *S. elongatus* PCC 7942 by integrating sucrose permease (cscB) from *E. coli* and overexpressing sucrose phosphate synthase (SPS) from *Synechocystis* sp. PCC 6803 (Abramson et al., 2016), and the nitrogen fixation bacterium *A. vinelandii* ΔnifL which is a derivate of *A. vinelandii* DJ by knockout of the *nifL* gene, a repressor of the nitrogenase (Ortiz-Marquez et al., 2012) were chosen in this project.

1.6.2 Aims and objectives

The aim in this work is to construct a synthetic microbial community based on the photosynthetic cyanobacterium *S. elongatus* cscB/SPS and the nitrogen fixation bacterium *A. vinelandii* ΔnifL. In this system, sucrose produced by *S. elongatus*

cscB/SPS is the carbon source for *A. vinelandii* ΔnifL and ammonia produced by *A. vinelandii* ΔnifL is the nitrogen source for *S. elongatus* cscB/SPS, allowing the microbial community to grow without the requirement of organic sources of carbon and nitrogen nutrients in the media. This designed microbial community also has the potential function of PHB production (Chapter 3). In addition, different proteomics quantification methods were evaluated to find a best approach for the synthetic microbial co-culture (Chapter 4). And an approach for performing proteomics experiments and data analysis to investigate interactions among the designed microbial community was developed (Chapter 5). The proteomics results can be used for the optimisation of the co-culture and the proteomics-based approach can also be used for the construction of other synthetic microbial communities.

Chapter 2 Materials and methods

2.1 *Synechococcus elongatus* cscB/SPS

2.1.1 Strain and growth conditions

S. elongatus cscB/SPS used in this project is a derivative of the wild-type (wt) *S. elongatus* PCC 7942 and has the ability of continuous sucrose export. Briefly, sucrose permease (*cscB*) cloned from *E. coli* genomic DNA (ATCC 700927) was integrated into neutral site 3 of the genome with chloramphenicol resistance and sucrose phosphate synthase (*SPS*) was overexpressed by the plasmid pAM1579 (Addgene ID 40240) with kanamycin resistance (Ducat et al., 2012; Abramson et al., 2016). Both genes are under the control of an IPTG inducible promoter; hence, IPTG was added to promote the transcription of *cscB* and *SPS*.

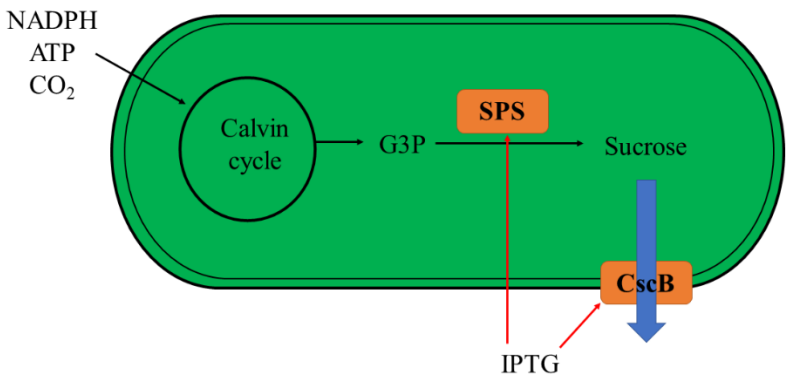


Figure 2-1 Overview of the engineered *S. elongatus* cscB/SPS strain. Cartoon schematics showed the improvement sucrose production and secretion by expression of SPS and CscB induced by IPTG. SPS: sucrose phosphate synthase; CscB: sucrose permease.

The strain was kindly gifted by Professor Daniel Ducat on agar plate. First, one colony was inoculated into BG-11 medium without antibiotics overnight to recover the cells and then 25 μg/mL chloramphenicol and 50 μg/mL kanamycin were added for selection. Precultures were screened by spotting on LB agar plates without antibiotics to check contamination. Stocks were maintained at -80 degrees with micro-organism preservation beads in cryovial (Technical Service Consultants Ltd). Unless stated otherwise, *S. elongatus* cscB/SPS was cultivated in BG11 medium (Appendix A) at 30 ℃ with agitation at 120 revolutions per minute (rpm) and light intensity at 120 $\mu E \cdot m^{-2} \cdot s^{-1}$. Antibiotics were added to achieve the following final

concentrations: kanamycin (Km), 50 μg/mL and chloramphenicol (Cm), 25 μg/mL. 0.5 mM IPTG was added to induce the *cscB* and *SPS* expression. All works were performed in a laminar flow hood under sterile conditions. Growth was monitored using spectrophotometer at an absorbance wavelength (λ) of 750 nanometres (nm) or flow cytometer (details shown at section 2.6).

2.1.2 Sucrose quantification

Sucrose production in *S. elongatus* cscB/SPS was triggered by addition of 0.5 mM IPTG. Sucrose concentration in the culture was determined using the Sucrose/ D-Glucose Assay Kit (K-SUCGL, Megazyme). The method depends on the production of a red colored quinoneimine dye compound by the reaction of glucose oxidase and peroxidase at pH 7.4 using *p*-hydroxybenzoic acid and 4-aminoantipyrine, before which sucrose is hydrolysed by the enzyme β-fructosidase to D-glucose and D-fructose. The color intensity is directly proportional to the sucrose concentration.

Briefly, 0.2 mL of cell culture supernatant was mixed with 0.2 mL acetate and 0.2 mL β-fructosidase solution respectively, which were then cultured at 50 ℃ for 20 mins. Add 3 mL GOPOD reagent to each tube and incubate them at 50 ℃ for 20 mins. Distilled water was as reagent blank and treated with the same procedures. Measure all samples at 510 nm against the reagent blank. 1 mg/mL D-Glucose standard solution was measured every time with three replicates as the control for calculation of the factor to convert from absorbance to D-Glucose amount.

2.2 *Azotobacter vinelandii* Δ nifL

2.2.1 Strain and growth conditions

A. vinelandii ΔnifL used in this project is a derivative of the wild type *A. vinelandii* DJ and has the ability of nitrogen fixation. Briefly, nifL, a nitrogenase anti-activator, was deleted from the *A. vinelandii* chromosome, resulting in a stable *A. vinelandii* mutant strain that can express nitrogenase constitutively and excrete ammonium to the surrounding medium (Ortiz-Marquez et al., 2012).

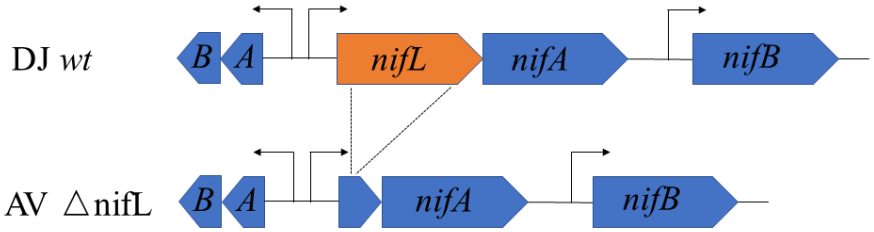


Figure 2-2 Illustration of *A. vinelandii* ΔnifL strain by deletion of the *nifL* gene in *A. vinelandii* DJ. Adapted from Ref (Ortiz-Marquez et al., 2012).

The strain was kindly gifted by Dr Leonardo Curatti on an agar slant. First, one colony was inoculated into nitrogen-free Burk's medium and cultivated for two days. Then one colony was selected to streak on the new Burk's agar plate to purify the culture. Stocks were maintained at -80 ℃ with 20% sterile glycerol (v/v). Unless stated otherwise, *A. vinelandii* ΔnifL was cultivated in Burk's medium (Appendix A) at 30 ℃ with agitation at 150 rpm. All works were performed in a laminar flow hood under sterile conditions Growth was monitored using spectrophotometer at an absorbance wavelength of 600 nm or colony-forming units (CFUs) on Burk's agar plates.

2.2.2 Ammonium quantification

Ammonia concentration was determined using the modified Nessler method according to Jeong *et al.* (2013) (Jeong et al., 2013). Firstly, 1 mL cell culture was centrifuged at 13000 rpm for 5 mins. 500 µL of the supernatant was diluted with 500 µL of ammonia-free water (VWR Chemicals) and then 20 µL of mineral stabilizer solution (HACH) was added to the samples. Next, 20 µL of 0.135% (w/v) PVA reagent (Sigma-Aldrich) and 40 µL of Nessler reagent (Sigma-Aldrich) were added to the samples followed by mixing well with vortex. The samples were incubated for 10 min at room temperature. Finally, the absorbance of 425 nm was measured, and ammonium concentrations were calculated from the ammonium standard curve generated with a series of ammonia standard solutions (Sigma-Aldrich) of 1, 2, 3, 4, 5 mg/L with triplicates. New PVA reagents and standard curves were prepared every time.

2.3 Protein analysis

2.3.1 Protein extraction

Protein extraction from the culture was performed using bead-based homogenization method . Firstly, 25 mL of cell cultures were taken and harvested by centrifuge at 5000 rpm for 10 min. Then the pellets were washed by resuspending in 10 mL of PBS buffer and repeating the spin. Next, 500 µL of lysis buffer (2% sodium dodecylsulfate [SDS; w/v], 40 mM Tris base pH 8.5, 60 mM dithiothreitol [DTT]) was added to resuspend the pellet. The cells were frozen at -80 °C overnight and quickly thawed at 37 °C to allow partial cell breakage. For the remainder of the procedure, 7 µL of 100× Halt™ protease inhibitor cocktail (Thermo Scientific) was added to protect proteins from degradation by endogenous proteases released during protein extraction and the samples were kept on ice. About 500 µL of 429-600nm acid-washed glass beads (Sigma-Aldrich) were added to the samples to leave 2 to 3 mm of cell suspension above the level of settled beads. The cells were broken by vigorous vortex mixing of the tube for 20 times in cycles of mixing for 30 s and cooling on ice for 30 s. Lysed cells were centrifuged at 13000 rpm, 4 °C for 10 mins and then left on ice until the foam subsided. The resulting green supernatant fraction was transferred to new pre-cold 1.5 mL LoBind Eppendorf tubes and stored at -20 °C or used for the following protein analysis.

2.3.2 Protein purification

Crude protein samples were purified using 2D Clean-Up Kit (GE Healthcare) to remove excess salts, buffers and other contaminants. The protein purification steps were following the manufacturer's instructions. Firstly, 100 µL of protein samples were transferred into 1.5 mL LoBind Eppendorf tubes. Then 300 µL precipitant was added to the sample, followed by mixing well with vortex and then incubated on ice for 15 mins. 300 µL co-precipitant was added and samples were mixed by vortex briefly then centrifuged at 8000 ×g for 10min. Remove as much of the supernatant as possible rapidly. Sufficient co-precipitant was added to the samples (about 3-4× times the size of pellet) followed by centrifugation at 8000×g for 5min. Discard the supernatant. Enough distilled water was added to disperse the pellets. 1mL pre-chilled wash buffer and 5ul wash additive were added to wash the samples by incubating at -20 ℃ for at least 30 mins with vortex for 20-30 s once every 10 mins. After that, the samples were centrifuged at 8000 ×g for 10 mins. Discard the supernatant and allow the pellet to air dry. Finally, 100 µL lysis buffer was added to fully dissolve the cleaned protein. Centrifuge at

8000 ×g for 10 mins to remove any insoluble material and to reduce any foam. The supernatant can be loaded to SDS-PAGE or stored at -80 ℃.

2.3.3 Protein concentration measurement

Protein concentrations were measured using BradfordUltra reagent (Expedeon), which is based on the reaction of coomassie with protein in an acidic medium, and the colour change from brown to blue. This coloured substance has specific absorption at 595 nm. Briefly, samples, standards and a blank were mixed with BradfordUltra reagent at a ratio of 1:15. All the samples were mixed well and incubated for 5 min at room temperature. Finally, the absorbance of 595 nm was measured, and protein concentrations were calculated from the BSA standard curve generated with a series of BSA standard solutions of 0.2, 0.4, 0.6, 0.8 and 1 mg/mL with triplicates. 2% (w/v) SDS, the same concentration as the protein samples, was used to dissolve BSA to eliminate the influence of detergent on the BradfordUltra assay. BSA standard curve was prepared every time.

2.3.4 Protein SDS-PAGE

To analyse proteins with the size range of 10-170 kDa, protein SDS-PAGE was run using NuPAGE™ 12% Bis-Tris Gel (Invitrogen). Briefly, 10 µL protein sample, 5 µL 4× LDS sample buffer, 2 µL 10× reducing agent and 3 µL dH_2O were mixed well and then incubated at 70 ℃ water bath for 10 min. Running buffer was prepared by adding 50 mL 20× MOPS SDS Running Buffer (Expedeon) to 950mL dH_2O to 1× SDS Running Buffer. 0.5mL antioxidant (NuPAGE) was added to 200 mL 1× Running Buffer for reduced samples. NuPAGE™ 12% pre-cast Gel was placed in the mini gel tank. Then 600 mL of 1× running buffer was added to the lower chamber, and 200 mL 1× Running Buffer with antioxidant was added to the upper chamber. 20 µL of the reacted protein samples were added into the gel holes and 5 µL of EZ-Run™ Pre-Strained *Rec* Protein Ladder (Fisher BioReagents™) was added to each gel. After the SDS-PAGE, gels were washed by distilled water and stained using ReadyBlue™ Protein Gel Stain (Sigma-Aldrich) overnight.

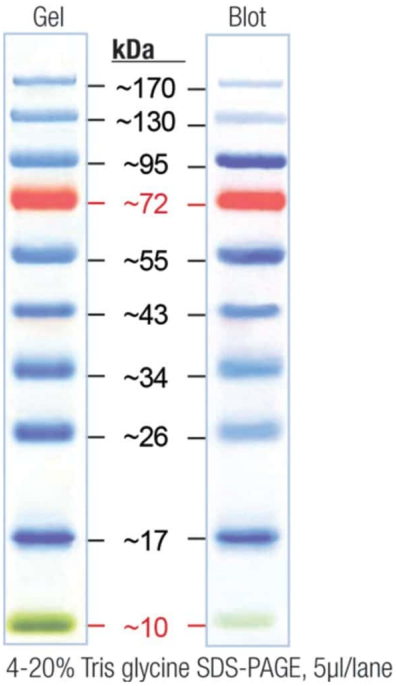


Figure 2-3 Protein ladder used for protein SDS-PAGE experiment. Image taken from Fisher scientific website (https://www.fishersci.com/).

2.4 Mass spectrometry

2.4.1 Protein reduction, alkylation and tryptic digestion

Protein lysates were digested using the methods reported by Hitchcock et al. (2016)(Hitchcock et al., 2016) and Wan Razali et al. (2022) (Hitchcock et al., 2016; Wan Razali et al., 2022) with modifications. Briefly, cleaned-up protein pellets were dissolved in 30 µL urea buffer (8 M urea/ 100 mM Tris-HCl pH 8.5/ 5 mM DTT) followed by water bath sonication to fully suspend. Protein concentrations were determined using a NanoDrop™ 2000 spectrophotometer (Thermo Scientific, Delaware, USA) with urea buffer as a blank. 50 µg of protein samples (previously mixed based on protein or cell number) were diluted to 10 µL with urea buffer and incubated at 37 °C for 30 min to reduce the protein. 1.5 µL of 100 mM iodoacetamide was added to the protein solutions and incubated in the dark at room temperature for 30 min. 10 µL MS grade trypsin (Promega, Wisconsin, USA) was added in a 1:50 (w/w) protease:protein ratio to the protein solutions and the solutions were diluted with 58.5 µL 50 mM Tris-HCl (pH 8.5)/ 10 mM $CaCl_2$ to a final urea concentration of 1 M. The protein solutions were incubated

overnight in a 37° C water bath. Trypsin digestion was terminated by adding formic acid to a final concentration of 1%. Digested peptides were desalted using Bond Elut OMIX C18 tips (Agilent Technologies) following the manufacturer's instructions and dried using a SpeedVac.

2.4.2 LC-MS/MS for proteomics

Liquid chromatography-tandem mass spectrometry (LC-MS/MS) proteomic analysis was performed following the methods previously reported (Hanson et al., 2016) with modification. Dried peptide pellets were dissolved in 50 µL loading buffer, consisting of 3% acetonitrile and 0.1% trifluoroacetic acid in water, and sonicated in a water bath for 3 min for full suspension, after which they were cleared by centrifugation at 13,000 rpm for 2 min. LC-MS/MS was performed and analyzed by nano-flow liquid chromatography (U3000 RSLCnano, Thermo Fisher Scientific, United Kingdom) coupled to a hybrid quadrupole-orbitrap mass spectrometer (Q Exactive HF, Thermo Fisher Scientific, United Kingdom). Peptides were separated on an Easy-Spray C18 column (75 µm × 50 cm) using a 2-step gradient from 3% solvent A (0.1% formic acid in water) to 10% B over 5 min and then to 50% solvent B (0.1% formic acid in 80% acetonitrile) at 300 nl min-1, 40°C. The mass spectrometer was programmed for data-dependent acquisition with 10 product ion scans (resolution 30,000, automatic gain control 1e5, maximum injection time 60 ms, isolation window 1.2 Th, normalized collision energy 27, and intensity threshold 3.3e4) per full MS scan (resolution 120,000, automatic gain control 1e6, maximum injection time 60 ms) with a 20-s exclusion time.

2.5 Proteomics data analysis

2.5.1 Database setup

For protein identification of the monoculture and co-culture samples, a reference FASTA protein database was created using all protein sequences of *S. elongatus* PCC 7942 and *A. vinelandii* DJ appended with CscB from *E. coli* W and SPS from *Synechocystis* sp. PCC 6803 from Uniprot (https://www.uniprot.org/, Feb 2020). All reviewed and unreviewed lists were selected and merged manually. The final database contained 7889 protein sequences.

2.5.2 Database searching - MaxQuant

Database search and protein identification were performed with MaxQuant software using the built-in Andromeda search engine. MaxLFQ algorithm was used for label-free quantification. Detailed parameters were set as listed in Table 2-1.

Table 2-1 Parameter settings of MaxQuant software used for protein identification and quantification.

Grouping	Parameters	Settings
Label-free quantification	Variable modification	Oxidation (M), Acetyl (Protein N-term)
	Fixed modification	Carbamidomethyl (C)
	LFQ minimum ratio count	2
	FastLFQ	on
Digestion	Digestion Enzyme	Trypsin/P
	Maximum missed cleavages	2
Instrument	Instrument type	Orbitrap
	First search peptide tolerance	20 ppm
Sequences	Minimum peptide length for unspecific search	8
	Maximum peptide length for unspecific search	25
Quantification	Peptides for quantification	Unique and razor
MS/MS analyser	FTMS MS/MS match tolerance	20 ppm
	ITMS MS/MS match tolerance	0.5 Da
Identification	PSM FDR	0.01

Protein FDR	0.01
Minimum peptides	1
Minimum razor and unique peptides	1
minimum score for modified peptides	40

LFQ, Label-free quantification; FTMS, Fourier Transform Mass Spectrometry; ITMS, Ion trap mass spectrometry; PSM, peptide spectrum matches; FDR, False Discovery Rate. Unlisted parameters were set as default.

2.5.3 Protein identification and quantification

For protein identification of the monoculture and co-culture samples, a reference database was created using all protein sequences of *S. elongatus* PCC 7942 (2874 sequences) and *A. vinelandii* DJ (5013 sequences) appended with CscB from *E. coli* and SPS from *Synechocystis* sp. PCC6803 from Uniprot (https://www.uniprot.org/, Feb 2020), resulting in a final database of 7889 protein sequences. Raw MS data files were processed using MaxQuant (2.0.3.0) and its built-in Andromeda search engine for peptide identification and protein inference (Cox et al., 2014). Default settings were used with search parameters set to include the following modifications: Oxidation (M) and Acetyl (Protein N-term) (variable); Carbamidomethyl (C) (fixed). Peptide-spectrum matches and protein identifications were filtered using a target-decoy approach at a false discovery rate (FDR) of 1%. Label free quantification (LFQ) and intensity-based absolute quantification (iBAQ) options were selected (Schwanhäusser et al., 2011).

Proteins were quantified using 6 metrics based on spectral counts (PSMs, unique peptides, and NSAF) and spectral intensity (intensity, iBAQ intensity, and LFQ intensity) at both the protein level and cell level. PSMs, unique peptides, intensity, iBAQ intensity, and LFQ intensity values were obtained from MaxQuant analysis. All the values were obtained from MaxQuant output except NSAF, which is equal to PSMs count divided by protein length.

2.5.4 Downstream bioinformatics analysis

2.5.4.1 Statistical analysis

Statistical analysis of protein identification was performed using the LFQ-Analyst website (https://bioinformatics.erc.monash.edu/apps/LFQ-Analyst/) based on the ProteinGroup.txt file. First, contaminant proteins, reverse sequences and proteins identified "only by site" were filtered out. In addition, proteins that have been only identified by a single peptide and proteins not identified/quantified consistently in the same condition have been removed as well. The LFQ data was converted to log2 scale, samples were grouped by conditions and missing values were imputed using the 'Missing not At Random' (MNAR) method, which uses random draws from a left-shifted Gaussian distribution of 1.8 StDev (standard deviation) apart with a width of 0.3. Protein-wise linear models combined with empirical Bayes statistics were used for the differential expression analyses. The limma package from R Bioconductor was used to generate a list of differentially expressed proteins for each pair-wise comparison. A cutoff of the adjusted p-value of 0.05 (Benjamini-Hochberg method) along with a |log2 fold change| of 1 has been applied to determine significantly regulated proteins in each pairwise comparison.

2.5.4.2 Functional annotation and pathway assignment

Differentially expressed proteins were mapped to each strain using the search and colour function in KEGG (https://www.genome.jp/kegg/mapper/color.htmL). Gene Ontology (GO) terms associated with differentially expressed proteins were identified using GO enrichment tool in OmicSolution. Clusters of orthologous groups (COGs) were analysed using EggNOG (http://eggnog-mapper.embl.de/) with default settings. Similarities of uncharacterised protein sequence were searched using BLASTP (National Center for Biotechnology Information; NCBI) with default settings (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

2.6 Flow cytometry

Flow cytometric analysis was carried out on an A60-Micro PLUS flow cytometer (Apogee Flow Systems, Hemel Hempstead, UK) equipped with 405 nm, 488 nm and 561 nm diode lasers. Three photomultiplier tubes (PMTs) were installed to collect small-angle light scatter (SALS), medium-angle light scatter (MALS) and large-angle light scatter (LALS) signals. Before sample analysis, the flow cytometer was calibrated using a reference silica beads mix

with diameters ranging from 110 to 1300 nm (ApogeeMix, #1493). Cells were measured using 405-MALS (325 V) and 561 nm orange (500 V) lasers. Data were acquired at a flow rate of 1.5 μL/min with a sample volume of 130 μL under the sheath fluid pressure of 150 mbar and recorded in the Histogram software (Apogee Flow Systems, Hemel Hempstead, UK). All data from the Apogee flow cytometer wee analysed using FlowJo.

Chapter 3 Construction of a carbon and nitrogen fixing synthetic microbial consortium for PHB production

3.1 Abstract

Microbial consortia have attracted much attention as a novel and effective bioproduction platform owing to their unique advantages compared with axenic culture. In this chapter, a cross-feeding bi-culture system was established using two engineered strains: a) *Synechococcus elongatus* cscB/SPS, a photosynthetic cyanobacterium that can synthesize and export sucrose to the medium, and b) *Azotobacter vinelandii* ΔnifL, a nitrogen-fixing bacterium that can secrete ammonium, which were grown under optimized conditions for at least two weeks. The results show that *S. elongatus* cscB/SPS could not grow in the absence of a nitrogen fixing *A. vinelandii* after the depletion of the initial nitrogen source. And the co-culture appears to bring inhibition of the heterotroph, but *A. vinelandii* ΔnifL can still remain alive. This artificial microbial consortium has also been confirmed having the capacity of PHB production.

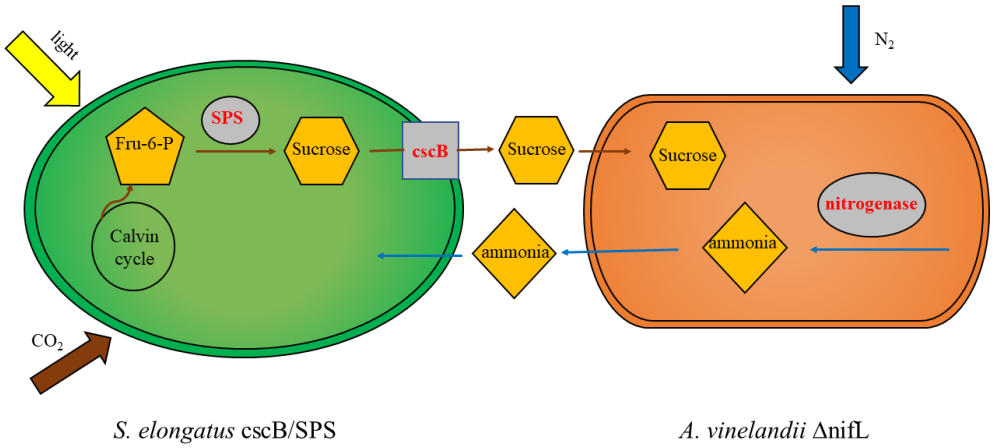


Figure 3-1 Illustration of the synthetic cross-feeding microbial co-culture of *S. elongatus* cscB/SPS and *A. vinelandii* ΔnifL. CO_2 is fixed in *S. elongatus* cscB/SPS through the Calvin cycle to produce sucrose, which is secreted to the medium by the

activity of heterologous CscB. *A. vinelandii* ΔnifL uptakes the sucrose in the medium for growth and production. N_2 is fixed in *A. vinelandii* ΔnifL and converted to ammonium via nitrogenase and secreted to the medium, which is absorbed by *S. elongatus* cscB/SPS. CO_2, carbon dioxide; N_2, nitrogen; SPS, sucrose phosphate synthase; CscB, sucrose permease. Fru-6-P, Fructose-6-phosphate.

3.2 Introduction

Microbial communities consisting of two or more microbial strains or species are not only essential for various Earth's ecosystem, but also have potential for the industrial bioproduction, yet examples of synthetic communities with mixed metabolism are limited in the laboratory (Hays et al., 2017). Because of their distinct characteristics, including undertaking more complex tasks and enduring more changing environments, microbial consortia have been shown to be capable of performing more complicated tasks and enduring more changing environments than individual populations (Brenner et al., 2008). However, their applications are restricted by the high cost owing to large demand of sugar and nitrate as nutrients and low productivity.

Recent research has explored the creation and application of microbial communities as a solution to mitigate the energy burden and environmental pollution associated with bioproduction substrates. 'Autotrophic-heterotrophic' co-culture systems have gained increasing investigation owing to their advantages such as high stability, robustness (Hays et al., 2017), and higher substrate utilisation efficiency (Argun et al., 2009). For example, Weiss *et al*. introduced sucrose permease (cscB) from *Escherichia coli* W to *S. elongatus* PCC7942 to promote the sucrose production with addition of NaCl, and then co-cultured with *Halomonas boliviensis*. This artificial co-culture has the ability of produce polyhydroxybutyrate with a production rate up to 28.3 mg L^{-1} d^{-1}, which can last for 5 months (Weiss et al., 2017).

Carbon and nitrogen are the most expensive components in bioproduction. Cyanobacterial carbohydrates have been explored as a developing renewable feedstock in industrial biotechnology for the manufacture of fuels and chemicals, with prospective production rates when compared to crop-based feedstock (Löwe et al., 2017a). *S. elongatus* is a model freshwater cyanobacterium with the capacity to produce biotechnologically relevant chemicals and biofuels, including acetone (Chwa et al., 2016), glycerol (Hirokawa et al., 2017), fatty acids (Santos-Merino et al., 2018), etc.

The strain used in this study was engineered to secrete sucrose, using carbon sourced from CO_2 fixation, and hence has generated significant interest in the microbial biotechnology community (Abramson et al., 2016; Hays et al., 2017). *A. vinelandii* is an obligate aerobic nitrogen-fixing bacterium that has been used as a model to study nitrogen fixation and siderophore production (Baars et al., 2016). The strain used here can secrete ammonium. In addition, it also has the ability to produce PHB, the alternative to fossil fuel-based plastics, up to 80–90% of its biomass (El-shanshoury et al., 2013; Smith and Francis, 2016a). Considering carbon and nitrogen are the two most costly nutrients during microbial cultivation and bioproduction (Yen et al., 2014; Tan et al., 2016), these two strains were combined for synthetic microbial co-culture, in which the atmospheric carbon (CO_2) and nitrogen (N_2) are fixed into organic forms to enable cross-feeding (as sucrose and ammonium, respectively).

In this chapter, a synthetic microbial community was established, based on the photosynthetic cyanobacterium *S. elongatus* cscB/SPS reported in Abramson et al., 2016, which is a derivate of *S. elongatus* PCC 7942 by integrating sucrose permease (cscB) from *E. coli* and overexpressing sucrose phosphate synthase (SPS) from *Synechocystis* sp. PCC 6803 (Abramson et al., 2016), and the nitrogen fixation bacterium *A. vinelandii* ΔnifL which is a derivate of *A. vinelandii* DJ by knockout of the *nifL* gene, a repressor of the nitrogenase (Ortiz-Marquez et al., 2012). In this system, sucrose produced by *S. elongatus* cscB/SPS is the carbon source for *A. vinelandii* ΔnifL and ammonia produced by *A. vinelandii* ΔnifL is the nitrogen source for *S. elongatus* cscB/SPS, allowing the microbial community to grow without addition of carbon and nitrogen nutrients (Figure 3-1). This designed microbial community has been verified to have the potential function for PHB production.

3.3 Methods

3.3.1 Strains background and monocultures

3.3.1.1 *S. elongatus* cscB/SPS monoculture growth and sucrose production

S. elongatus cscB/SPS used in this project is a derivative of the wild-type (wt) *S. elongatus* PCC 7942 (Chapter 2, Figure 2-1). Two enzymes, sucrose permease (*cscB*) and sucrose phosphate synthase (*SPS*), were introduced into *S. elongatus* PCC 7942 to promote sucrose production. More details regarding strain modification are available in Ducat et al. and Abramson et al.'s papers (Ducat et al., 2012; Abramson et al., 2016). *S. elongatus* cscB/SPS was cultivated in BG-11 medium (Appendix A) at 30 °C with agitation at 120 rpm and light intensity at 120 $\mu E \cdot m^{-2} \cdot s^{-1}$. Antibiotics were added to achieve the following final concentrations: kanamycin (Km), 50 μg/ml and chloramphenicol (Cm), 25 μg/mL. All works were performed in a laminar flow hood under sterile conditions. Growth was monitored using a spectrophotometer (Thermo Scientific) at an absorbance wavelength (λ) of 750 nanometres (nm) or a flow cytometer (details shown in section 2.6). Growth rate and doubling time were calculated using equation 3-1 and 3-2, respectively.

$$\mu = \frac{ln_{OD2} - ln_{OD1}}{t_2 - t_1} \text{ (Equation } 3-1)$$

$$t_d = \frac{ln2}{\mu} \text{ (Equation } 3-2)$$

Sucrose production in *S. elongatus* cscB/SPS was triggered by the addition of 0.5 mM IPTG to induce the *cscB* and *SPS* expression. Sucrose concentration in the culture was determined using the Sucrose/ D-Glucose Assay Kit (Megazyme, Bray, Ireland). The method depends on the production of a red-coloured quinoneimine dye compound by the reaction of glucose oxidase and peroxidase at pH 7.4 using *p*-hydroxybenzoic acid and 4-aminoantipyrine, before which sucrose is hydrolysed by the enzyme β-

fructosidase to D-glucose and D-fructose. The colour intensity is directly proportional to the sucrose concentration. 1 mg/mL D-Glucose standard solution was measured every time with three replicates as the control for calculation of the factor to convert from absorbance to D-Glucose amount.

3.3.1.2 *A. vinelandii* ΔnifL monoculture growth and ammonium production

A. vinelandii ΔnifL used in this project is a derivative of the wild type *A. vinelandii* DJ (Chapter 2, Figure 2-2). NifL, a nitrogenase anti-activator, was deleted from the *A. vinelandii* chromosome, resulting in a stable *A. vinelandii* mutant strain that can express nitrogenase constitutively and excrete ammonium to the surrounding medium (Ortiz-Marquez et al., 2012). More details regarding strain modification are available in Ortiz-Marquez et al. *A. vinelandii* ΔnifL was cultivated in Burk's medium (Appendix A) at 30 degrees with agitation at 150 rpm. All works were performed in a laminar flow hood under sterile conditions Growth was monitored using spectrophotometer at an absorbance wavelength of 600 nm or colony-forming units (CFUs) on Burk's agar plates.

Ammonia concentration was determined using the modified Nessler method according to Jeong *et al*. (2013) (Jeong et al., 2013). Firstly, 1 mL cell culture was centrifuged at 13000 rpm for 5 min. 500 µL of the supernatant was diluted with 500 µL of ammonia-free water and then 20 µL of mineral stabilizer was added to the samples. Next, 20 µL of 0.135% (w/v) polyvinyl alcohol (PVA) reagent and 40 µL of Nessler reagent were added to the samples followed by mixing well with vortex. The samples were incubated for 10 min at room temperature. Finally, the absorbance of 425 nm was measured, and ammonium concentrations were calculated from the ammonium standard curve generated with a series of ammonia chloride (NH_4Cl) standard solutions of 20, 40, 60, 80, and 100 µM with three replicates. New PVA reagents and standard curves were prepared every time.

3.3.2 Microbial consortium construction

3.3.2.1 Formulation of co-culture medium

A co-culture medium was created by adjustment of Burk's medium for *A. vinelandii* and BG-11 medium for *S. elongatus*. Detailed experiments were carried out in order to optimise the co-culture medium with limited amounts of nitrogen and carbon sources. Different media were designed for co-culture by varying concentrations of carbon and nitrogen source, salt and buffer (Table 3-1, SAC1-SAC7). Cells from *S. elongatus* cscB/SPS and *A. vinelandii* ΔnifL growing to the logarithmic phase were collected for co-culture.

Table 3-1 Co-culture media compositions adjusted from standard Burk's medium and BG-11 medium. Components are shown in grams per liter with the exception of BG-11 trace element solution listed in Table 3-2.

Media composition (g/L)		SAC 1	SAC 2	SAC 3	SAC 4	SAC 5	SAC 6	SAC 7
Carbon source	Sucrose				5	5	5	5
Nitrogen source	$NaNO_3$			0.06	0.06	0.06	0.06	0.4
Burk's phosphate buffer	KH_2PO_4	0.2	0.2	0.2	0.2	0.2	0.2	0.2
	K_2HPO_4	0.8	0.8	0.8	0.8	0.8	0.8	0.8
Burk's salts	$CaCl_2•2H_2O$	0.61	0.61	0.61	0.61	0.61	0.61	0.61
	$MgSO_4•7H_2O$	0.81	0.81	0.81	0.81	0.81	0.81	0.81
	$Na_2MoO_4•2H_2O$	0.0108	0.0108	0.0108	0.0108	0.0108	0.0108	0.0108
	$FeSO_4•7H_2O$	0.005	0.005	0.005	0.005	0.005	0.005	0.015

	$K_2HPO_4•3H_2O$				0.04	0.04	0.2	0.2
	$MgSO_4•7H_2O$	0.075	0.075	0.075	0.075	0.075	0.375	0.375
	$CaCl_2•2H_2O$	0.036	0.036	0.036	0.036	0.036	0.036	0.036
BG-11 stock solutions	Citric acid	0.006	0.006	0.006	0.006	0.006	0.006	0.006
	$EDTANa_2•2H_2O$	0.011	0.011	0.011	0.011	0.011	0.011	0.011
	Na_2CO_3	0.02	0.02	0.02	0.02	0.02	0.02	0.02
	Ammonium ferric citrate green					0.006	0.006	0.006
BG-11 trace element		1 mL	1 mL	1 mL	1 mL	1 mL	1 mL	1 mL
pH buffer	$NaHCO_3$		2	2				
	HEPES					2.383	2.383	2.383

Table 3-2 Composition of BG-11 trace element solution.

Composition	Concentration (g/L)
H_3BO_3	2.86
$MnCl_2•4H_2O$	1.81
$ZnSO_4•7H_2O$	0.22
$Na_2MoO_4•2H_2O$	0.39
$CuSO_4•5H_2O$	0.08

$Co(NO_3)_2•6H_2O$	0.05

3.3.2.2 Starting ratio of synthetic co-culture

To assess the influence of starting ratio on the growth of synthetic co-culture, *S. elongatus* cscB/SPS and *A. vinelandii* ΔnifL cells were grown to the logarithmic phase and then mixed at different cell number ratios of 50:50, 60:40, 70:30, 80:20, 90:10, and 100:0. The growth of co-culture was measured using a flow cytometer.

3.3.2.3 Cultivation and characterization of synthetic co-culture

To start the synthetic microbial co-culture, *S. elongatus* cscB/SPS axenic culture was firstly grown in BG-11 medium to an exponential phase and then inoculated into SAC7 medium after washing with PBS solution to remove residential nutrients. Two days after the induction of *S. elongatus* cscB/SPS by 0.5 mM IPTG, *A. vinelandii* ΔnifL was inoculated at a cell number ratio of 80:20 of *S. elongatus* to *A. vinelandii* to ensure that sucrose production and export process was ready for feeding it (Smith and Francis, 2016b; Löwe et al., 2017b). Co-culture members were grown in 500 mL volume flasks and incubated at 30 degrees with agitation of 120 rpm and continuous illumination of 120 $\mu E \cdot m^{-2} \cdot s^{-1}$. Five g/L sucrose and 0.4 g/L $NaNO_3$ were used to trigger the growth of two members and their cross-feeding.

To have the approximate growth status, the total optical density of 1 mL culture was measured using a spectrophotometer at a wavelength of 600 nm for *A. vinelandii* ΔnifL and 750 nm for *S. elongatus* cscB/SPS, respectively. In terms of *S. elongatus* is easily detectable using chlorophyll a, flow cytometer (FCM) was chosen for monitoring the growth of *S. elongatus* cscB/SPS (details shown in section 2.6). The colony forming units (CFUs) were measured for the cell counting of *A. vinelandii* ΔnifL with serial dilution and plating on Burk's-BB agar. The cell number of *A. vinelandii* ΔnifL was calculated using the CFUs results and the standard curve between CFUs and cell number shown in Appendix B1. All experiments were conducted in three biological

replicates.

3.3.3 Model of growth under substrate-limited condition

The microbial cells growing in a batch culture generally follows Monod kinetics (Gupta and Gupta, 2021). It describes a system that is limited by the nutrient concentration and follow Michaelis-Menten kinetics. The equation for Monod kinetics can be written as follows:

$$\mu = \frac{\mu_{max} * S}{K_s + S} \text{ (Equation } 3-3)$$

Where μ represents the specific growth rate; μ_{max} represents the maximum specific growth rate; *S* represents the concentration of growth-limiting substrate; K_s is saturation constant, defined as the substrate concentration at which the specific growth rate reaches half of the maximum growth rate.

In the synthetic microbial co-culture, nutrient concentration are necessarily low due to the productivity of the microorganisms, resulting in relatively low concentrations of sucrose and ammonium. Therefore, Monod kinetics are well suited to describe the synthetic co-culture system.

3.3.4 PHB determination

PHB determination was conducted by gas chromatography (GC) technique using the modified procedures reported by Sayyed et al. 2020 (Sayyed et al., 2020). Briefly, about 10-20 mg freeze-dried co-culture cells were mixed with 2 mL chloroform and 2 mL methanolic sulphuric acid. For depolymerization and methanolysis of PHB, the mixture was heated at 100˚C for 4 h with occasional shaking followed by subsequently cooling at room temperature and adding 2 mL of distilled water. The organic layer at the bottom part was collected by a glass Pasteur pipette and subjected to spectral analysis (Sayyed et al., 2020).

Samples were analysed by GC (Perkin Elmer AutoSystem XL Autosampler) equipped with Zebron ZB-5plus Capillary GC column. Half μL of the sample was injected with

a sampling rate of 6.25 pts/s. Hydrogen was used as the carrier gas. The oven temperature of the column was programmed from 60 ˚C for 3 min, then ramped at a rate of 10 ˚C /min to 230 ˚C and held at this temperature for 2 min (Sayyed et al., 2020). Methyl benzoate was used as the internal standard. This value was used for PHB yield calculation in dry cells using equations 3-4 to 3-6.

HB standard equation:

$$y = 2.991 * x \text{ (Equation } 3-4)$$

Concentration of sample:

$$\text{HB concentration (g/L)} = \frac{\text{Peak area of HB}}{2.991} * 10\text{^}(-3) \text{ (Equation } 3-5)$$

PHB yield:

$$\text{PHB (\%)} = \frac{\text{HB weight (g/L)}}{\text{total dry cell weight (g/L)}} * 100 \text{ (Equation } 3-6)$$

3.4 Results and discussion

3.4.1 Growth and production in monocultures

3.4.1.1 Growth and sucrose production of *S. elongatus* cscB/SPS

S. elongatus cscB/SPS cells were cultivated in BG-11 medium as described above. The strain can reach stationary phase after 7 days (Figure 3-2 A) with specific growth rate (Equation 3-1) of 0.214 ± 0.079 day^{-1} and doubling time (Equation 3-2) of 3.655 ± 1.707 days using OD_{750} taken at day 5 and day 7 time points.

Sucrose production was induced by the addition of IPTG. IPTG concentration of 1 mM is usually the standard concentration employed in the induction of protein expression in molecular biology (Larentis et al., 2014). However, considering the negative effect of IPTG on the growth of bacteria (Gomes et al., 2020) and IPTG concentration for the

induction of CscB and SPS expression, 0.5 mM IPTG was chosen for the induction of sucrose. Sucrose production at different IPTG was also examined, and the results showed that sucrose production ability is similar under 0.5 mM and 1 mM IPTG (Figure 3-2 B). The sucrose production rate under 0.5 mM IPTG were 0.109 ± 0.046 g/(L·day) with a maximum sucrose concentration of 0.780 ± 0.327 g/L on day 7 within the measurement.

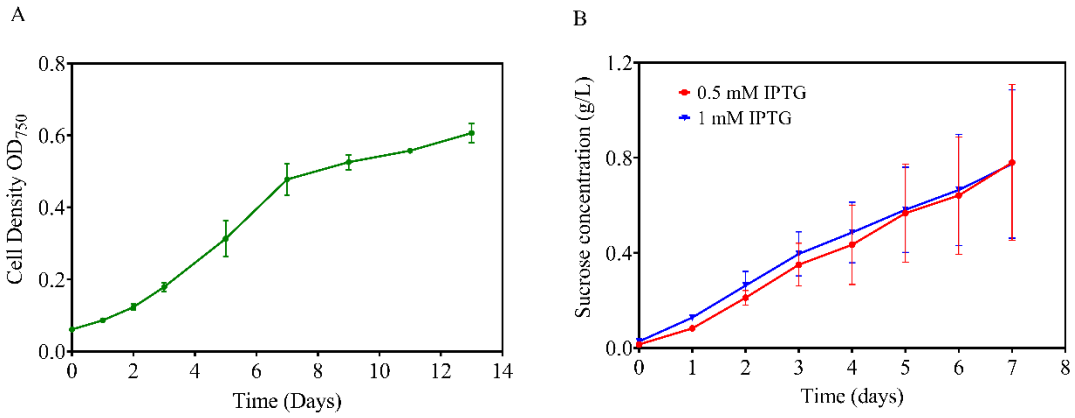


Figure 3-2 Growth and production of *S. elongatus* cscB/SPS strain. (A) Growth kinetics of *S. elongatus* cscB/SPS cultured under standard conditions. Growth was measured by optical density at OD750. (B) Sucrose production of *S. elongatus* cscB/SPS under different IPTG concentration. All data shown were repeated in three biological replicates. Error bars represent standard deviation.

3.4.1.2 Growth and ammonium production of *A. vinelandii* ΔnifL

A. vinelandii ΔnifL cells were cultivated in Burk's medium as described in section 3.3.1.2. The strain can reach stationary phase after 5 days (Figure 3-3 A) with a specific growth rate (Equation 3-1) of 1.803 ± 0.026 day^{-1} and doubling time (Equation 3-2) of 0.384 ± 0.006 days using OD_{600} taken at day 2 and day 3 time points.

Ammonia concentration was determined using the modified Nessler method (Jeong et al., 2013). The standard ammonium curve was conducted using NH_4Cl, result shown in

Appendix B2. The highest ammonium productivity was observed on day 5 with a concentration of 1.095 ± 0.032 mM (Figure 3-3 B), which is equivalent to 0.059 ± 0.002 g/L NH_4Cl.

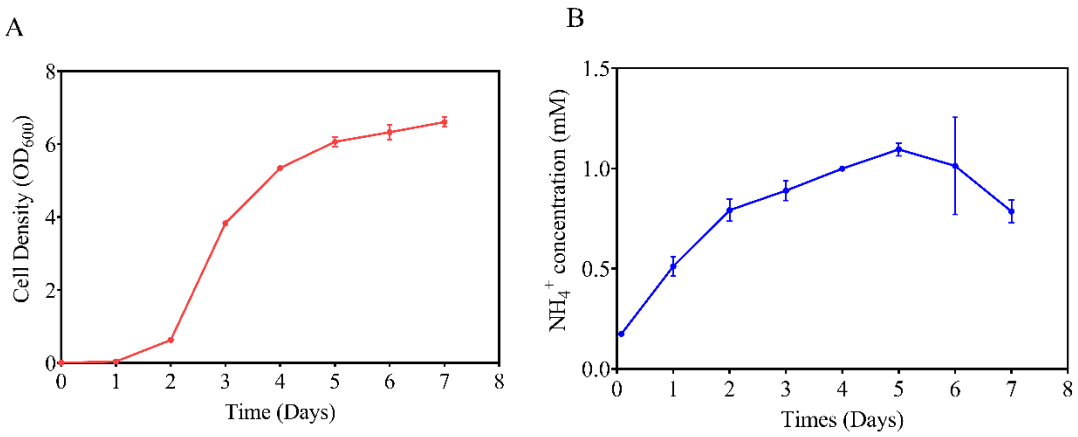


Figure 3-3 Growth and production of *A. vinelandii* ΔnifL strain. (A) Growth kinetics of *A. vinelandii* ΔnifL cultured under standard conditions. (B) Ammonium production of *A. vinelandii* ΔnifL under standard conditions. All data shown were repeated in three biological replicates. Error bars represent standard deviation.

3.4.1.3 Substrate limitation

To test the influence of sucrose limitation on the growth of *A. vinelandii* ΔnifL, different sucrose concentrations were used in Burk's medium, i.e., 0 mM, 1 mM, 10 mM and 50 mM (Figure 3-4). Maximum growth rate was calculated between day 2 and day 3 using equation 3-3 (Perni et al., 2005), and results are shown in table 3-2. The results demonstrated that *A. vinelandii* ΔnifL can grow under various sucrose concentrations, but rarely grow below 1 mM (342.3 mg/L) sucrose. Moreover, the higher the sucrose concentration in the medium, the higher the specific growth rate it had (Figure 3-4 B).

$$\mu = \frac{OD_2 - OD_1}{t_2 - t_1} \quad \text{(Equation } 3-3\text{)}$$

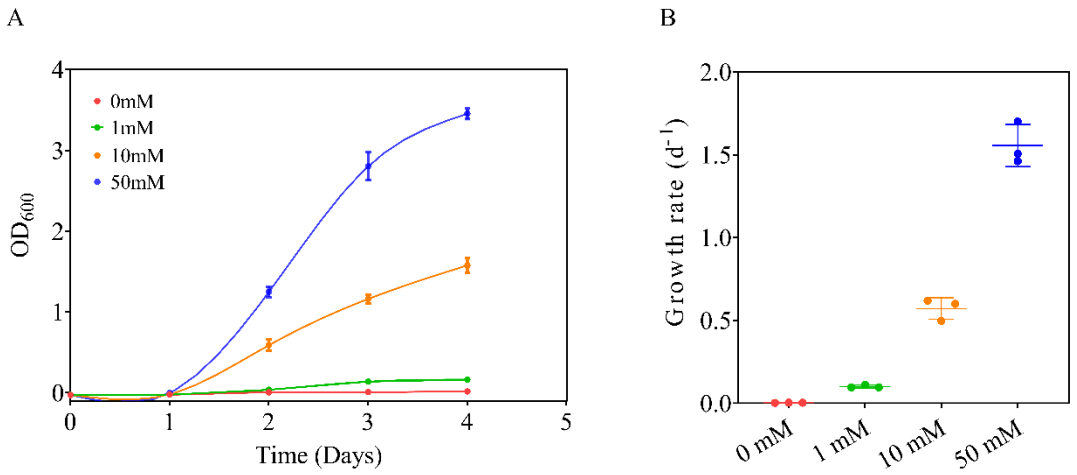


Figure 3-4 Growth kinetics of *A. vinelandii* ΔnifL grown under different sucrose concentration. (A) Growth curve of *A. vinelandii* ΔnifL cultured at different sucrose concentrations in Burk's medium, varying from 0 mM to 50 mM. (B) Specific growth rate of *A. vinelandii* ΔnifL at different sucrose concentration from day 2 to day 3. Each experiment was repeated in three biological replicates. Error bars represent standard deviation.

Table 3-3 Growth kinetics of *A. vinelandii* ΔnifL grown under different sucrose concentration

Sucrose concentration (mM)	SGR (d^{-1})	Doubling Time (d)
0	0.003 ± 0.001	231.049
1	0.101 ± 0.0087	6.863
10	0.572 ± 0.0656	1.212
50	1.556 ± 0.1272	0.445

SGR: specific growth rate.

3.4.2 Optimisation of co-culture medium

Previous studies have revealed an increase in sucrose production in engineered *S. elongatus* cscB/SPS (Ducat et al., 2012; Abramson et al., 2016) and ammonium production in engineered *A. vinelandii* △nifL (Ortiz-Marquez et al., 2012) compared to their wild types under IPTG induced. To make the synthetic microbial co-culture grow, different co-culture media were tested with varying compositions of carbon source, nitrogen source, salt, and buffer, based on BG-11 medium for *S. elongatus* and Burk's medium for *A. vinelandii*. (Figure 3-5).The purpose of this chapter was to construct a self-sufficient system, therefore, carbon and nitrogen sources were omitted from these two media initially (Table 3-1, SAC1). However, no growth evidence was shown on either *S. elongatus* cscB/SPS or *A. vinelandii* ΔnifL when cultivated in the SAC1 medium (Figure 3-5 A). This means initial carbon and nitrogen sources are needed to trigger the growth of both strains and the generation of cross-feeding.

According to Tu et al. (2018), sodium bicarbonate ($NaHCO_3$) can be used as an external carbon source to cultivate microalga growth (Tu et al., 2018). Therefore, 10% $NaHCO_3$ (Smith and Francis, 2016a) was then added into the co-culture medium (Table 3-1, SAC2). However, neither *S. elongatus* cscB/SPS nor *A. vinelandii* ΔnifL showed growth when cultivated in SAC2 medium (Figure 3-5 B).

Then different media compositions were tested, mainly differing in carbon source, nitrogen source, phosphate, sulfate, and iron. To trigger the growth of *S. elongatus* cscB/SPS, 0.06 g/L sodium nitrate ($NaNO_3$) and 0.006 g/L ammonium ferric citrate green were added as initial nitrogen (Table 3-1, SAC3 and SAC5). To promote the growth of *A. vinelandii* ΔnifL, 5 g/L sucrose was added into the co-culture medium (Table 3-1, SAC4). Additionally, the concentrations of $K_2HPO_4 \cdot 3H_2O$ and $MgSO_4 \cdot 7H_2O$ components were increased 5 times to exclude possible limitations when adding an additional microorganism (Table 3-1, SAC6). 10 mM HEPES was added as a pH stabilizer. Figure 3-5 C-F demonstrated that *S. elongatus* cscB/SPS and *A. vinelandii* ΔnifL co-culture grown in media SAC3 – SAC6 only showed a small amount

of growth on the first or second day but could not maintain growth for a long time.

Finally, a co-culture medium was constructed by adding 0.4 g/L $NaNO_3$ to provide more initial nitrogen sources for *S. elongatus* cscB/SPS (Table 3-1, SAC7). Figure 3-5 G showed that *S. elongatus* cscB/SPS and *A. vinelandii* ΔnifL co-culture grew well in the SAC7 medium with the highest OD_{600} of 4.915 and OD_{750} of 4.035 on the 4^{th} and 5^{th} day, respectively. The results suggested that the initial carbon and nitrogen sources are compulsory for the start of co-culture.

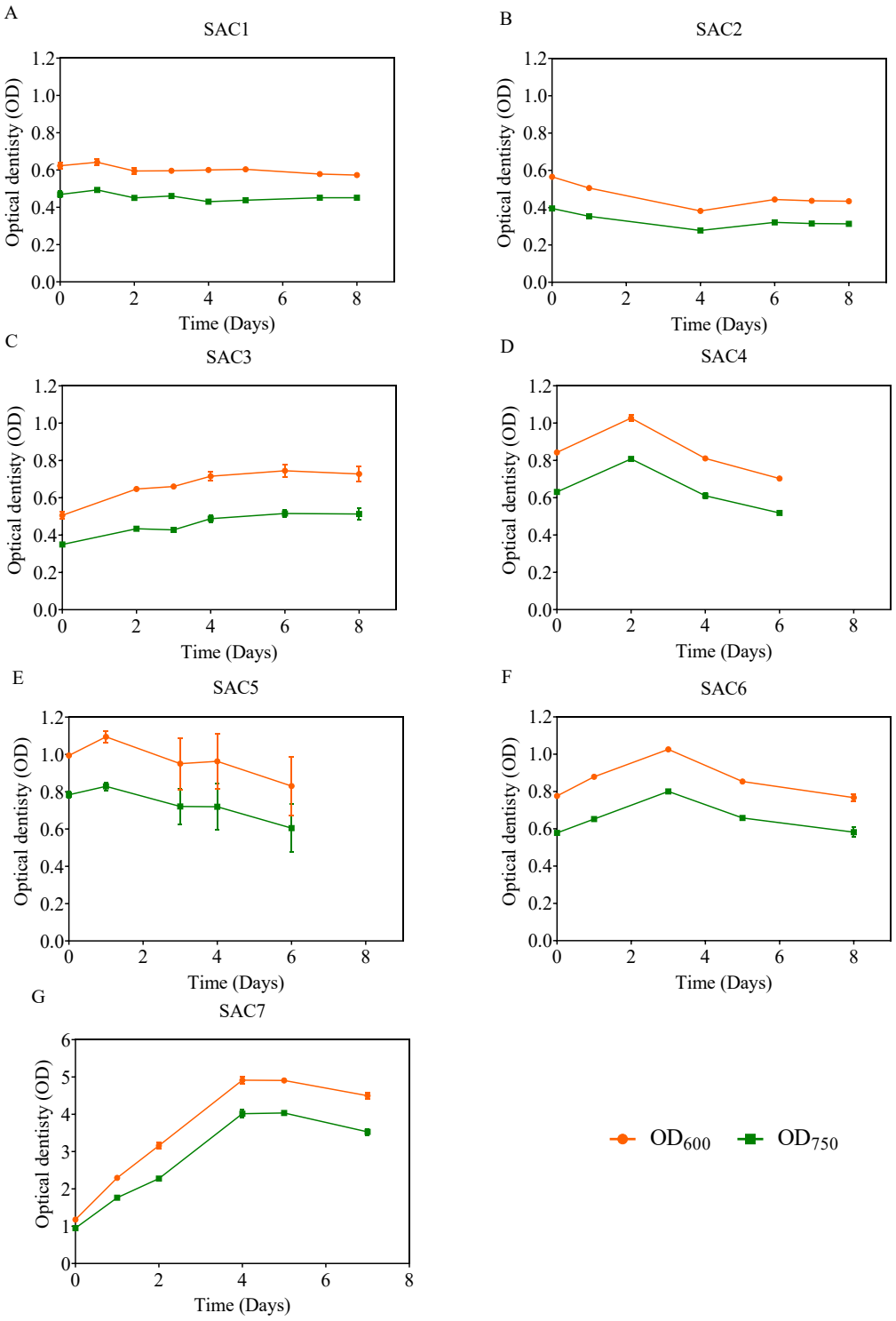


Figure 3-5 Growth curve of co-culture under different media compositions. (A) SAC1 medium, (B) SAC2 medium, (C) SAC3 medium, (D) SAC4 medium, (E) SAC5 medium, (F) SAC6 medium, (G) SAC7 medium. Orange lines represent the growth curves of *A. vinelandii* ΔnifL measured at OD_{600}. Green lines represent growth curves

of *S. elongatus* cscB/SPS measured at OD_{750}. Each experiment was repeated in three biological replicates. Error bars represent standard deviation.

3.4.3 Initial inoculation ratio influences the synthetic co-culture

The initial inoculation ratio is another key experimental parameter for the growth of synthetic co-culture, as it can not only affect the final ratio of the co-culture but also regulate the metabolic capacity of the coculture (Gao et al., 2021). Considering the growth rate of *S. elongatus* cscB/SPS is much slower than that of *A. vinelandii* △nifL (Figure 3-2 and 3-3), more cells of *S. elongatus* cscB/SPS were inoculated to the co-culture at the beginning with different ratios of 50%, 60%, 70%, 80%, 90%, and 100%. The growth of *S. elongatus* cscB/SPS in co-culture was tracked using a flow cytometer. The growth curves of *S. elongatus* cscB/SPS at different inoculation ratios exhibited that apart from *S. elongatus* cscB/SPS monoculture (100% *S. elongatus*), *S. elongatus* cscB/SPS initial ratio of 80% showed a higher growth rate (Figure 3-6 and Table 3-4). Therefore, we used the initial inoculation ratio of 80: 20 of *S. elongatus* cscB/SPS and *A. vinelandii* △nifL for the following experiments.

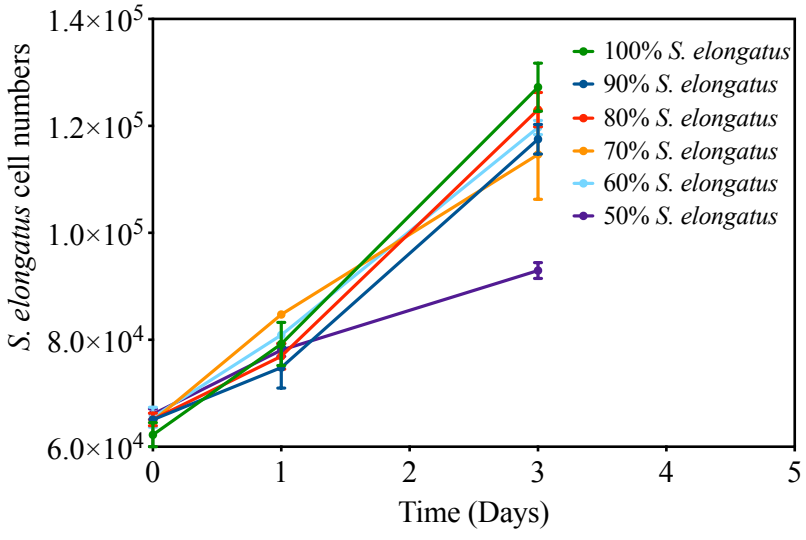


Figure 3-6 Specific growth of *S. elongatus* cscB/SPS at different inoculation ratios. Ratios were reported as a percentage of *S. elongatus* cscB/SPS cell number, e.g., 60% *S. elongatus* indicates a ratio of 60: 40 of *S. elongatus* cscB/SPS: *A. vinelandii* △nifL.

Table 3-4 Specific growth rate of *S. elongatus* cscB/SPS at different starting ratios in co-culture.

Starting ratio	100% *S. elongatus*	90% *S. elongatus*	80% *S. elongatus*	70% *S. elongatus*	60% *S. elongatus*	50% *S. elongatus*
SGR (day^{-1})	0.238	0.197	0.212	0.190	0.201	0.114

SGR: Specific growth rate.

3.4.4 Kinetic and production analysis of co-culture

S. elongtatus cscB/SPS and *A. vinelandii* ΔnifL co-culture were grown under carbon and nitrogen limited conditions (5 g/L sucrose and 0.4 g/L $NaNO_3$, respectively) in SAC7 medium as described above at pH 7.3. Growth curves of each strain grown in co-culture and monoculture control are shown in Figure 3-7. It was observed that co-culture has different influence on the autotroph and heterotroph. The *A. vinelandii* ΔnifL strain had a higher specific growth rate of 1.926 ± 0.052 day^{-1} at the beginning, compared with that in monoculture of 1.403 ± 0.084 day^{-1} using the data taken on day 0 and 2 (Figure 3-7 A). While the *S. elongatus* cscB/SPS strain had a lower specific growth rate of 0.423 ± 0.117 day^{-1} at the beginning, compared with that in monoculture of 0.620 ± 0.005 day^{-1} using the data taken on day 0 and 2 (Figure 3-7 B). This may be because the heterotroph dominated at the start of the competition for nutrients, and the higher sucrose concentration in the co-culture medium that produced by *S. elongatus* cscB/SPS.

After 4 days, the axenic *S. elongatus* cscB/SPS culture began to collapse (Figure 3-7 B) and exhibited the phenomenon of chlorosis in the response to nitrogen starvation (Forchhammer and Schwarz, 2019). Whereas it can maintain at a fluctuating and

relatively stable level in co-culture, which is due to the benefit of the ammonium feeding from the *A. vinelandii* ΔnifL. The number of *A. vinelandii* ΔnifL cells in co-culture tended to be stable in co-culture after two-day fast growing, while it still exhibited growth trend in the axenic culture (Figure 3-7 A). This may be owing to the nutrients limitation, like metals, caused by the competition of *S. elongatus* cscB/SPS, except for sucrose.

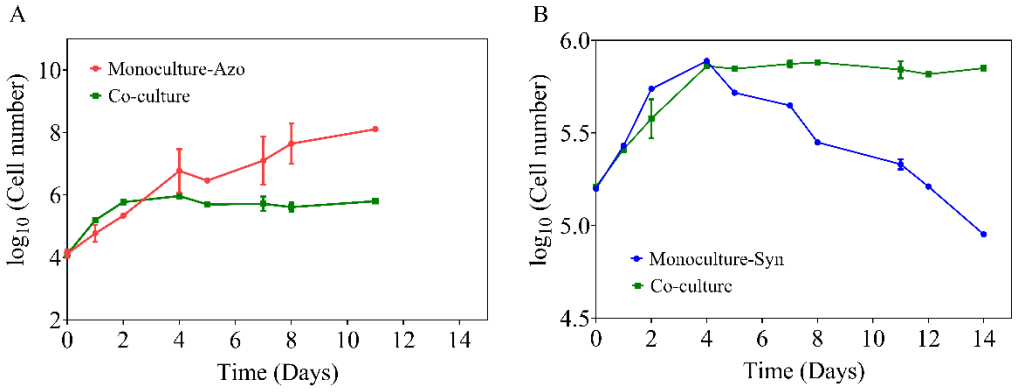


Figure 3-7 Growth curves of each strain in co-culture and monoculture control. (A) Growth curve of *A. vinelandii* ΔnifL in co-culture (green) and monoculture (red). (B) Growth curve of *S. elongatus* cscB/SPS in co-culture (green) and monoculture (blue). Each experiment was repeated in three biological replicates. Error bars represent standard deviation.

The sucrose and ammonium production in co-culture are depicted in Figure 3-8. Since the initial nitrogen source provided, the highest ammonium concentration accumulated was 7.84 mg/L (434.64 μM) after two days of co-culturing in limited medium (Figure 3-8), followed by a rapid decline owing to the need of ammonium from *S. elongatus* cscB/SPS, which then maintained at about 0.5 mg/L after 10 days. Sucrose concentration in the co-culture medium dropped quickly at the first five days because of the rapid growth of *A. vinelandii* ΔnifL, which could maintain at about 1 g/L after 12 days. These observations prove the hypothesis that the sucrose produced from *S.*

elongatus cscB/SPS and ammonium produced from *A. vinelandii* ΔnifL can support each other to growth to form a self-sufficient microbial system.

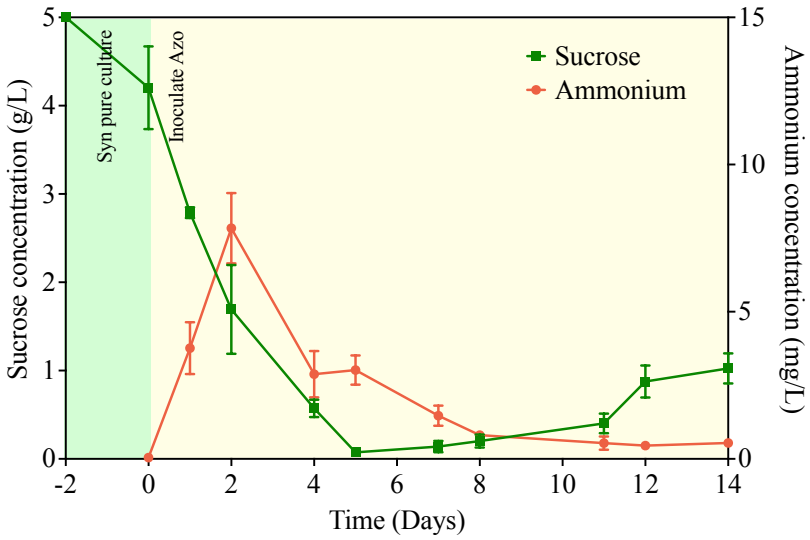


Figure 3-8 Sucrose and ammonium production in the co-culture. *A. vinelandii* ΔnifL was inoculated after *S. elongatus* cscB/SPS under IPTG induction for two days. Sucrose produced from *S. elongatus* cscB/SPS was measured using the Sucrose/ D-Glucose Assay Kit (green line). Ammonium produced from *A. vinelandii* ΔnifL was measured using the modified Nessler method (red line). Each experiment was repeated in three biological replicates. Error bars represent standard deviation.

3.4.5 The potential of PHB production using air CO_2 and N_2

A. vinelandii is a common gram-negative, PHA-producing soil bacterium that can fix atmospheric nitrogen and produce different products, such as PHA (Mok et al., 2021; Yoshida et al., 2022). The application of biopolymers produced from different kinds of microorganisms is usually limited by the high cost, especially carbon source, which accounts for approximately 50% of the entire production process of PHAs (Jiang et al., 2016). Cyanobacterial carbohydrates have been suggested as emerging renewable feedstocks for the production of fuels and chemicals in industrial biotechnology (Löwe

et al., 2017a), therefore, *A. vinelandii* △nifL was expected to produce PHA using the carbon source secreted from *S. elongatus* cscB/SPS in this synthetic microbial co-culture. The nature of PHB accumulated in the co-culture on the 9th day was evaluated by GC analysis. The retention time of the peak of 3.90 was identical to the methyl-hydroxybutyrate, which is the characteristic feature of PHB, similar results were observed by Nurbas and Kutsal (2004) (Nurbaş and Kutsal, 2004). The area under the peak denotes the PHB content (Figure 3-9), which was determined to be 0.47 % of dry cell weight. This result demonstrates the potential application of our synthetic microbial co-culture for biopolymer production using air CO_2 and N_2.

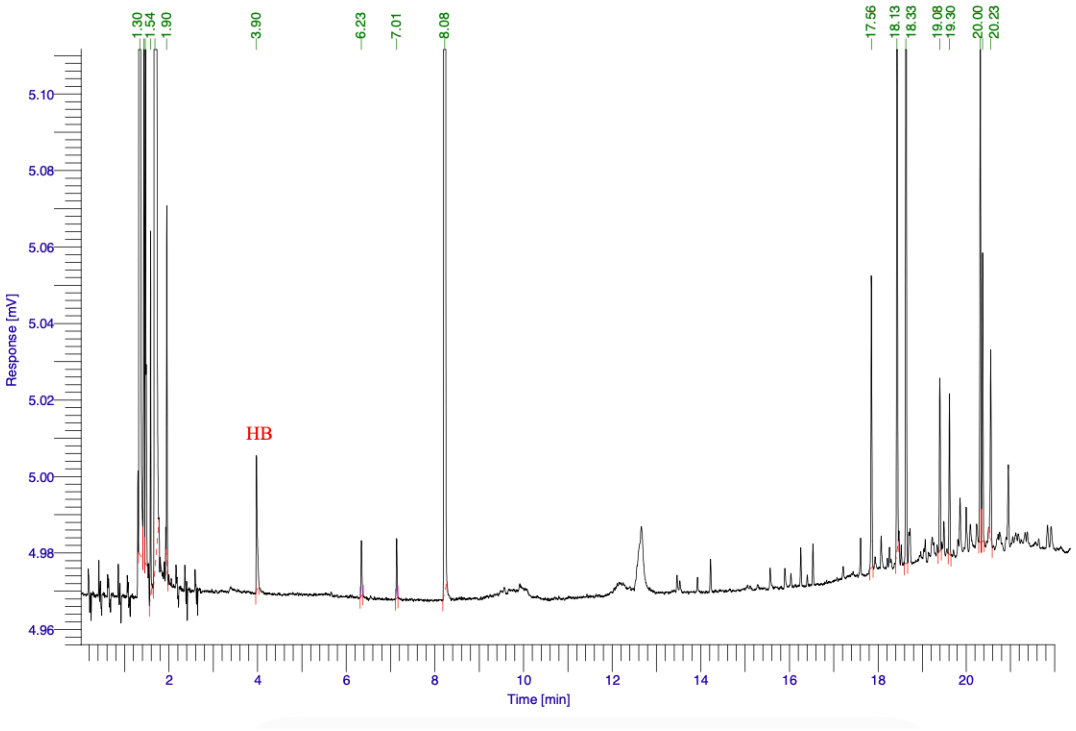


Figure 3-9 Gas chromatography (GC) analysis of polymer synthesised in co-culture. 0.5 µL Sample was analysed by GC equipped with Zebron ZB-5plus Capillary GC column. The spectra obtained were recorded with methyl benzoate as the internal standard.

3.5 Conclusion

In this work, a mixed culture of engineered *S. elongatus* cscB/SPS and *A. vinelandii* ΔnifL was established on lab scale, which can be co-cultured together for at least two weeks. The sucrose produced from *S. elongatus* cscB/SPS and the ammonium produced from *A. vinelandii* ΔnifL can feed each other, achieving a carbon and nitrogen self-sufficient system, which decreases the cost of growth and production greatly. The strategy of combining photosynthetic organisms with heterotrophic bacteria to produce industrially relevant compounds is successful. Although with low PHB productivity in this co-culture system, the biopolymer yield can be improved by other approaches such as engineering targeted synthesis pathways. This chapter provides a perspective on the bioproduction of a wide range of biological products from air CO_2 and N_2 in synthetic microbial co-culture.

Chapter 4 Establishing a label-free proteomics workflow for microbial co-cultures

4.1 Abstract

The value of synthetic microbial communities in biotechnology is gaining traction due to their ability to undertake more complex metabolic tasks than monocultures. However, a thorough understanding of strain interactions, productivity and stability is often required to optimize growth and scale up cultivation. Quantitative proteomics can provide valuable insights into how microbial strains adapt to changing conditions in biomanufacturing or bioremediation scenarios. However, current workflows and methodologies are not suitable for simple artificial co-culture systems where strain ratios are dynamic. In this chapter, a standard workflow for co-culture proteomics was established using an exemplar system containing two key members, *Azotobacter vinelandii* and *Synechococcus elongatus*. The state of co-culture label-free proteomics was assessed, with particular attention to the variability in protein abundance estimates across co-culture samples in this highly peptide-centric technique. Firstly, the performance of popular proteomics analysis software, MaxQuant, was tested for pure and mixed culture proteins. Factors affecting the quantitative accuracy of co-culture proteomics were investigated, including peptide physicochemical characteristics such as molecular weight, isoelectric point, hydrophobicity, and dynamic range, as well as factors relating to protein identification such as varying proteome size and shared peptides between species. Different quantification methods based on spectral counts and intensity were evaluated, demonstrating good correlations between protein amount and the six quantification methods at the protein level. A new normalization method was proposed, named "LFQRatio", to reflect the relative contributions of the two distinct cell types emerging from the cell ratio changes during co-cultivation. LFQRatio can be applied to real co-culture proteomics experiments, providing accurate insights into quantitative proteome changes in each strain.

4.2 Introduction

There is growing interest in a variety of fields in creating synthetic microbial consortia. This can be for the fundamental understanding of how microbes interact (Christie-Oleza et al., 2017), microbial evolution studies (Good et al., 2017; Cairns et al., 2018) or for biotechnology purposes (Minty et al., 2013; Pandhal and Noirel, 2014; Cairns et al., 2018). Quantitative proteomics provides a powerful tool to interrogate how these microbes interact and their relative metabolic status in different conditions.

Although proteomics workflows to quantify protein abundance changes in different conditions (e.g., nutrient limitation, light and dark cycles, etc.) are widely available for axenic (pure) microbial cultures (Wegener et al., 2010; Guerreiro et al., 2014), applying these common workflows to synthetic co-cultures, i.e., when several strain types are cultivated together, presents certain challenges. This is particularly relevant when comparing conditions with highly variable cell-type ratios, as this will cause large differences in protein abundance between samples. Advanced software algorithms can deal with some systematic biases among samples, e.g., samples processed on different days or with different MS performances (Cappadona et al., 2012; Cox et al., 2014), as well as for differences in protein extraction efficiency among co-culture cell types; however, they produce unreliable quantification data when analysing samples with large cell number differences (Kleiner et al., 2017), such as comparing mono-cultures and co-cultures, or different co-culture time points. In metaproteomics experiments, species abundances are often quantified and considered when interpreting the findings of the data. Quantifying cell numbers can be undertaken using well-established methods such as 16S rRNA gene amplicon sequencing and fluorescence *in situ* hybridization. Biomass abundances can even be calculated using mass spectra, and methodologies have been demonstrated to quantify organisms in saliva from multiple individuals and microbial mats from two alkaline soda lakes (Kleiner et al., 2017; Kleiner, 2019). However, a method to account for the two distinct protein populations present in samples for quantitative proteomics analysis remains to be established.

Key considerations for such an approach are the physicochemical characteristics of consortium member proteomes, such as the range of isoelectric points (*pI*) (Kozlowski, 2017), molecular weight (M_w) (Angel et al., 2012), and hydrophobicity (Warwood et al., 2013), as well as the dynamic range distribution of protein abundances within the proteome (Zubarev, 2013), must be considered when analysing mixed-strain proteomic samples, as they will affect protein

extraction efficiency, peptide ionization in the mass spectrometer and overall proteome identification, among other factors. There are also protein database-related challenges, for bioinformatics analysis including variable proteome sizes (Kumar et al., 2017) and shared peptides between the strains that must be considered in terms of protein assignment and relative quantification (Helliwell et al., 2018).

Previous proteomic studies of synthetic co-cultures, often compare the bi-culture to monoculture, and hence, physicochemical and bioinformatics factors that affect quantification do not need to be considered (Helliwell et al., 2018). If the aim is to infer metabolic changes in each strain in the co-culture, there are cultivation strategies that can overcome technical challenges. One option is spatial separation of the strains, for example in growth chambers separated by a semi-permeable membrane that allows metabolite exchange but not cell mixing (Benomar et al., 2015; Thøgersen et al., 2018). Standard proteomics workflows can then be applied to each sample separately. Furthermore, if strains are sufficiently different in size or fluorescence, they can be separated by flow cytometry and cell sorting (Rüger et al., 2012; Engel et al., 2019) or differential centrifugation (Pajarillo et al., 2017), although this process has its inherent limitations: a) overlap in the size or fluorescence distributions of the strains; and b) changing cell cultivation volumes (Schlembach et al., 2021), which could affect metabolism and therefore the proteome (Llufrio et al., 2018).

Quantification in proteomics has largely moved to label-free methods due to the rapid improvement in the sensitivity of liquid chromatography (LC), mass spectrometry hardware and the accuracy of proteomics data analysis tools (Al Shweiki et al., 2017). Label-free quantification (LFQ) has the benefit of cost savings, less stringent chemistry requirements in extraction buffers and no limitation of sample numbers, compared with label-based methods (Välikangas et al., 2018). LFQ is widely used to analyse global proteome changes in different biological conditions, allowing the quantification of thousands of proteins to be determined using the “total protein approach” (Wiśniewski et al., 2014; Wiśniewski and Mann, 2016). There are two main categories of relative quantification methods for label-free proteomics. The first is based on spectral counting data generated during protein identification, such as peptide counts, and peptide-to-spectrum matches (PSMs). The second is based on spectral intensity, such as peptide peak heights and peak areas (Zhu et al., 2010; Blein-Nicolas and Zivy, 2016). Spectral counting-based methods have gained widespread use due to their ease of implementation (Lundgren et al., 2010); however, they are significantly affected by the

dynamic exclusion settings of the mass spectrometer, which can obscure the relationship between the detected number of counts and protein abundance, especially for lower abundance proteins. Intensity-based methods can be applied as an alternative, although MS2 methods are not always accurate, as peptide fragmentation sometimes does not occur at the maximum of the elution peak (Goeminne et al., 2018). In addition, some methods combining both approaches have been developed, such as ProPCA (Dicker et al., 2010).

To develop a quantitative proteomics analytical workflow for mixed microbial cultures, the proteome physiochemical characteristics, proteome sizes, and shared peptides between *S. elongatus* and *A. vinelandii* were compared. Then a series of mixed samples containing known ratios of 1) each cell type or 2) protein extracts from each strain were generated and the variability in estimated protein abundance for the sample series was analyzed. Estimation was assessed using 6 metrics based on spectral counts (PSMs, unique peptides, and NSAF) and spectral intensity (intensity, iBAQ intensity, and LFQ intensity) at both the protein level and cell level. Guided by this data, a novel normalization method was developed, named "LFQRatio", which, unlike the analytical methods currently available, can accommodate large differences in cell number ratios observed between co-culture conditions. LFQRatio factors in the LFQ intensity ratio of each protein and the total protein intensity to generate accurate label-free proteome quantification data for both strains within the microbial co-culture.

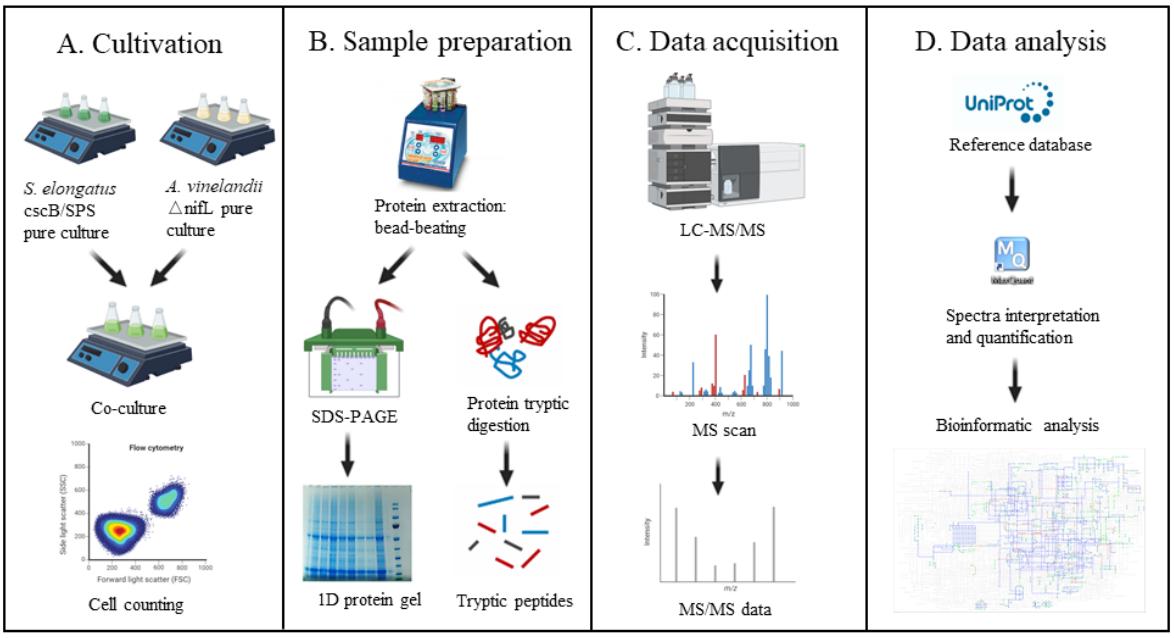


Figure 4-1 Workflow of performing label-free proteomics for co-culture. (Icons obtained from

BioRender)

4.3 Methods

4.3.1 Cell cultivation

S. elongatus culture was grown in BG11 medium (Rippka et al., 1979), pH 8.0, at 30 °C with agitation at 120 rpm and light intensity at 120 $\mu E \cdot m^{-2} \cdot s^{-1}$ in air environment. *A. vinelandii* was cultivated in Burk's medium(Dos Santos, 2011) at 30 °C with agitation at 150 rpm.

4.3.2 Cell counting

Flow cytometric analysis was carried out on an A60-Micro PLUS flow cytometer (Apogee Flow Systems, Hemel Hempstead, UK) equipped with 405 nm, 488 nm and 561 nm diode lasers. Three photomultiplier tubes were installed to collect small-angle light scatter, medium-angle light scatter and large-angle light scatter signals. Before sample analysis, the flow cytometer was calibrated using reference silica beads mix with diameters ranging from 110 to 1300 nm (ApogeeMix, #1493). Cells were measured using 405-MALS (325 V) and 561 nm orange (500 V) lasers. Data were acquired at a flow rate of 1.5 µL/min with a sample volume of 130 µL under the sheath fluid pressure of 150 mbar and recorded in the Histogram software (Apogee Flow Systems, Hemel Hempstead, UK).

4.3.3 Protein extraction, purification, quantification and SDS-PAGE

Cell cultures were grown to exponential phase and harvested by centrifuge at 5000 rpm, 4 °C for 10 min. The pellets were washed by resuspending in 10 mL of PBS buffer and repeating the spin. 500 µL of lysis buffer (2% sodium dodecylsulfate [SDS; w/v], 40 mM Tris base, pH 8.5, 60 mM dithiothreitol [DTT]) was added to resuspend the pellet. The cells were frozen at -80 °C overnight and quickly thawed at 37 °C to allow partial cell breakage. 7 µL of 100× Halt™ protease inhibitor cocktail (Thermo Scientific, Illinois, USA) was added to protect proteins from degradation by endogenous proteases released during protein extraction and the samples were kept on ice. Complete cell breakage was achieved by vigorous vortex mixing of the samples with 500 µL of 425-600 nm acid-washed glass beads (Sigma-Aldrich, Missouri, USA) 20 times in cycles of mixing for 30 s and cooling on ice for 30 s. Lysates were collected by centrifugation at 13000 rpm, 4 °C for 10 min.

Crude protein samples were purified using 2D Clean-Up Kit (GE Health, Buckinghamshire, UK) to remove excess salts, buffers and other contaminants following the manufacturer's

instructions. Protein concentrations were measured using BradfordUltra reagent (Expedeon, Cambridgeshire, UK) following the manufacturer's instructions using bovine serum albumin as the protein standard. 100 µg of purified proteins were loaded into NuPAGE™ 12% Bis-Tris Gel (Thermo Scientific, California, USA) running at 200 V for 55 min for protein separation. After the SDS-PAGE, gels were washed with distilled water and stained using ReadyBlue™ Protein Gel Stain (Sigma-Aldrich, Darmstadt, Germany) overnight.

4.3.4 Sample preparation

Two types of synthetic mixes were prepared, i.e., protein level mixes and cell level mixes, to assess different quantification methods of co-culture. Protein mixes were made by mixing the extracted proteins of *S. elongatus* and *A. vinelandii* at ratios of 100:0, 95:5, 90:10, 75:25, 50:50, 25:75, 10:90, 5:95, 0:100, named pSA1, pSA2, pSA3, pSA4, pSA5, pSA6, pSA7, pSA8, and pSA9, respectively. Cell mixes were prepared by mixing *S. elongatus* and *A. vinelandii* cells at ratios of 100:0, 95:5, 90:10, 75:25, 50:50, 25:75, 10:90, 5:95, 0:100, named cSA1, cSA2, cSA3, cSA4, cSA5, cSA6, cSA7, cSA8, and cSA9, respectively. The proteins of mixed cells were extracted as the same method shown above.

4.3.5 In-solution trypsin digestion and peptide purification

Protein lysates were digested using the methods reported by Hitchcock et al. (2016)(Hitchcock et al., 2016) and Wan Razali et al. (2022) (Hitchcock et al., 2016; Wan Razali et al., 2022) with modifications. Briefly, cleaned-up protein pellets were dissolved in 30 µL urea buffer (8 M urea/ 100 mM Tris-HCl pH 8.5/ 5 mM DTT) followed by water bath sonication to fully suspend. Protein concentrations were determined using a NanoDrop™ 2000 spectrophotometer (Thermo Scientific, Delaware, USA) with urea buffer as a blank. 50 µg of protein samples (previously mixed based on protein or cell number) were diluted to 10 µL with urea buffer and incubated at 37 °C for 30 min to reduce the protein. 1.5 µL of 100 mM iodoacetamide was added to the protein solutions and incubated in the dark at room temperature for 30 min. 10 µL MS grade trypsin (Promega, Wisconsin, USA) was added in a 1:50 (w/w) protease:protein ratio to the protein solutions and the solutions were diluted with 58.5 µL 50 mM Tris-HCl (pH 8.5)/ 10 mM $CaCl_2$ to a final urea concentration of 1 M. The protein solutions were incubated overnight in a 37° C water bath. Trypsin digestion was terminated by adding formic acid to a final concentration of 1%. Digested peptides were desalted using Bond Elut OMIX C18 tips (Agilent Technologies) following the manufacturer's instructions and dried using a SpeedVac.

4.3.6 Shotgun LC-MS/MS analysis

Liquid chromatography-tandem mass spectrometry (LC-MS/MS) proteomic analysis was performed following the methods previously reported (Hanson et al., 2016) with modification. Dried peptide pellets were dissolved in 50 µL loading buffer, consisting of 3% acetonitrile and 0.1% trifluoroacetic acid in water, and sonicated in a water bath for 3 min for full suspension, after which they were cleared by centrifugation at 13,000 rpm for 2 min. LC-MS/MS was performed and analyzed by nano-flow liquid chromatography (U3000 RSLCnano, Thermo Fisher Scientific, United Kingdom) coupled to a hybrid quadrupole-orbitrap mass spectrometer (Q Exactive HF, Thermo Fisher Scientific, United Kingdom). Peptides were separated on an Easy-Spray C18 column (75 µm × 50 cm) using a 2-step gradient from 3% solvent A (0.1% formic acid in water) to 10% B over 5 min and then to 50% solvent B (0.1% formic acid in 80% acetonitrile) at 300 nl min^{-1}, 40°C. The mass spectrometer was programmed for data-dependent acquisition with 10 product ion scans (resolution 30,000, automatic gain control 1e5, maximum injection time 60 ms, isolation window 1.2 Th, normalized collision energy 27, and intensity threshold 3.3e4) per full MS scan (resolution 120,000, automatic gain control 1e6, maximum injection time 60 ms) with a 20-s exclusion time.

4.3.7 Protein identification and quantification

For protein identification of the monoculture and co-culture samples, a reference database was created using all protein sequences of *S. elongatus* PCC 7942 (2874 sequences) and *A. vinelandii* DJ (5013 sequences) appended with CscB from *E. coli* and SPS from *Synechocystis* sp. PCC6803 from Uniprot (https://www.uniprot.org/, Feb 2020), resulting in a final database of 7889 protein sequences. Raw MS data files were processed using MaxQuant (2.0.3.0) and its built-in Andromeda search engine for peptide identification and protein inference (Cox et al., 2014). Default settings were used with search parameters set to include the following modifications: Oxidation (M) and Acetyl (Protein N-term) (variable); Carbamidomethyl (C) (fixed). Peptide-spectrum matches and protein identifications were filtered using a target-decoy approach at a false discovery rate (FDR) of 1%. Label free quantification (LFQ) and intensity-based absolute quantification (iBAQ) options were selected (Schwanhäusser et al., 2011).

Proteins were quantified using 6 metrics based on spectral counts (PSMs, unique peptides, and NSAF) and spectral intensity (intensity, iBAQ intensity, and LFQ intensity) at both the protein level and cell level. PSMs, unique peptides, intensity, iBAQ intensity, and LFQ intensity values

were obtained from MaxQuant analysis. All the values were obtained from MaxQuant output except NSAF, which is equal to PSMs count divided by protein length.

4.3.8 Assessment of physicochemical characteristics and shared peptides

The theoretical *pI* and *Mw* of all *S. elongatus* and *A. vinelandii* proteins were assessed by R scripts (Appendix file C3 and C4) using the proteome sequences obtained from Uniprot (Feb 2020). The hydrophobicity of *S. elongatus* and *A. vinelandii* proteomes was calculated using an R script (Appendix file C5) by GRAVY scores. Theoretically and actually shared peptides between *S. elongatus* and *A. vinelandii* were compared. A theoretical tryptic digest was performed based on the protein sequences of *S. elongatus* and *A. vinelandii* retrieved from Uniprot (Feb 2020). An R script (Appendix file C6) was prepared to read the protein sequences and theoretically digest them into tryptic peptides by cleaving them after arginine and lysine. Peptides ranging in length from 6 to 25 amino acids were identified and compared. Measured shared peptides were analyzed by comparison of the resulting peptide sequences of each strain using Hiplot (https://hiplot.com.cn/cloud-tool/drawing-tool/detail/113). The influence of database size was evaluated by searching the *S. elongatus* and *A. vinelandii* co-culture mass spectrometry raw data against individual and merged databases.

4.4 Results and discussion

4.4.1 Preliminary global proteomics analyses

4.4.1.1 Global proteome identification of *S. elongatus* cscB/SPS and *A. vinelandii* ΔnifL pure protein and mixed protein using MaxQuant software

To verify whether it was feasible to use label-free proteomics for global analysis of the proteome of *S. elongatus* cscB/SPS and *A. vinelandii* △nifL and the co-culture, a series of preliminary experiments and analysis was carried out. Firstly, 10 μg proteins of *S. elongatus* cscB/SPS, *A. vinelandii* △nifL and their 50:50 mixture were digested with trypsin. All raw data files generated from MS were processed using MaxQuant (version 2.0.3.0) against the merged database of *S. elongatus* PCC 7942, *A. vinelandii* DJ, CscB of *E. coli* W and SPS of *Synechocystis* sp. PCC 6803 downloaded from Uniprot. Detailed parameters setting were shown in chapter 2, section 2.5. All the proteins represented by two or more peptides were used

for the following analysis.

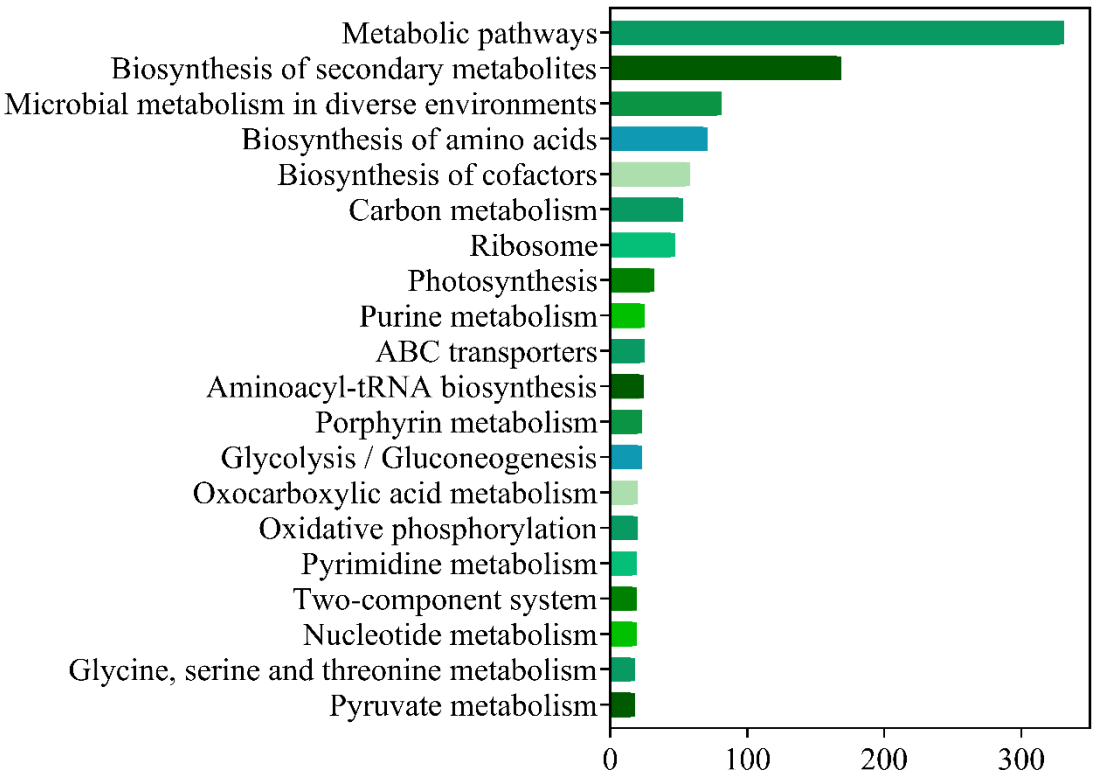
Table 4-1 Peptide and protein identification of pure protein and mixed protein of *S. elongatus* cscB/SPS and *A. vinelandii* △nifL using MaxQuant

Proteins	Protein numbers	Peptide numbers	Identification rate (%)
Syn	1015	8471	35.3
Azo	1266	13711	25.3
SA-H	867 for Syn, 1008 for Azo	7096 for Syn, 9553 for Azo	

Syn: *S. elongatus* cscB/SPS; Azo: *A. vinelandii* △nifL; SA-H: 50:50 mixture of *S. elongatus* cscB/SPS and *A. vinelandii* △nifL.

After removing contaminants and reverse sequences, a total of 13711 unique tryptic peptides for *A. vinelandii* △nifL and 8471 for *S. elongatus* cscB/SPS with 1% FDR were identified by shotgun proteomics approach, respectively (Table 4-1). These detected peptides corresponded to 1266 proteins of the *A. vinelandii* △nifL and 1015 proteins of the *S. elongatus* cscB/SPS with unique peptides equal to or more than 2, accounting for 37% and 67% of the *A. vinelandii* and *S. elongatus* proteome, respectively. The identified proteins mapped to 103 KEGG pathways for *S. elongatus* cscB/SPS and 116 KEGG pathways for *A. vinelandii* △nifL. Top 20 KEGG pathways were shown in Figure 4-2. More pathway information is shown in Appendix Table C1 and C2.

A


Metabolic pathways
Biosynthesis of secondary metabolites
Microbial metabolism in diverse environments
Biosynthesis of amino acids
Biosynthesis of cofactors
Carbon metabolism
Ribosome
Photosynthesis
Purine metabolism
ABC transporters
Aminoacyl-tRNA biosynthesis
Porphyrin metabolism
Glycolysis / Gluconeogenesis
Oxocarboxylic acid metabolism
Oxidative phosphorylation
Pyrimidine metabolism
Two-component system
Nucleotide metabolism
Glycine, serine and threonine metabolism
Pyruvate metabolism
0
100
200
300

B

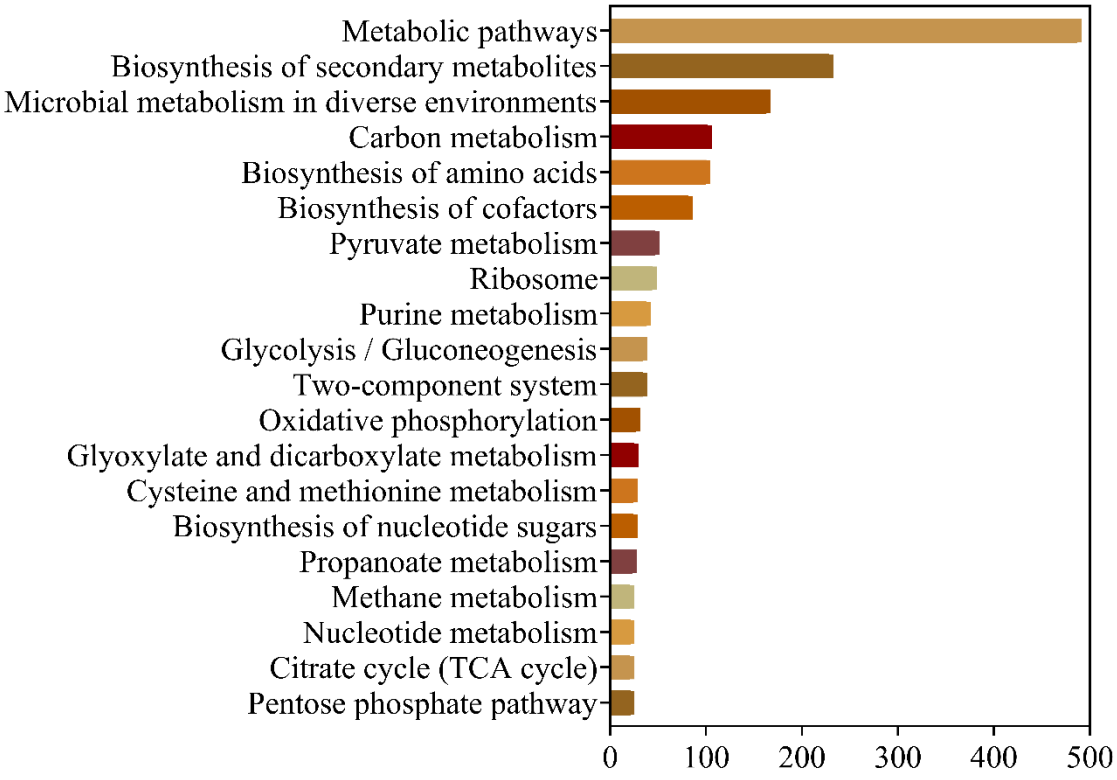

Metabolic pathways
Biosynthesis of secondary metabolites
Microbial metabolism in diverse environments
Carbon metabolism
Biosynthesis of amino acids
Biosynthesis of cofactors
Pyruvate metabolism
Ribosome
Purine metabolism
Glycolysis / Gluconeogenesis
Two-component system
Oxidative phosphorylation
Glyoxylate and dicarboxylate metabolism
Cysteine and methionine metabolism
Biosynthesis of nucleotide sugars
Propanoate metabolism
Methane metabolism
Nucleotide metabolism
Citrate cycle (TCA cycle)
Pentose phosphate pathway
0
100
200
300
400
500

Figure 4-2 Top 20 mapped KEGG pathways by the identified proteins of *S. elongatus* cscB/SPS (A) and *A. vinelandii* △nifL (B).

4.4.1.2 Preliminary physicochemical characterisation

The physicochemical characteristics of proteins differ between species and can affect protein extraction efficiency and peptide ionization in the mass spectrometer among other factors, thus influencing protein identification and quantification. Protein *pI* is the pH value at which the surface of a molecule carries no net electric charge. Proteins are at their least soluble when the buffer pH is equal to the *pI* value, therefore this parameter must be considered when optimizing a buffer system to maximize the solubility of proteins. The theoretical *pI* ranges of *S. elongatus* and *A. vinelandii* were assessed using R scripts (Appendix file C3), which exhibited similar *pI* ranges of 3.20–13.03 and 3.28–13.16, respectively. The frequency plots (Figure 4-3 A and B) revealed that more than 50% of proteins from both species have *pI* values of 5–7. The M_w of the proteins in the *S. elongatus* and *A. vinelandii* databases were also assessed using R scripts (Appendix file C4). The M_w of proteins in the *S. elongatus* database ranged from 1.05 to 204.08 kDa (Figure 4-3 C), of which proteins with M_w of 1-100 kDa accounted for 98.75% of all proteins. The M_w range of *A. vinelandii* was from 1.04 to 579.04 kDa (Figure 4-3 D), of which protein M_w of 1-100 kDa accounted for 98.64%. Our results showed that the proteomes of *S. elongatus* and *A. vinelandii* have similar M_w, although *A. vinelandii* contains three large protein that were listed in Table 4-2.

Table 4-2 List of three largest protein of *A. vinelandii* △nifL

Protein IDs	Protein names	Gene names	*Mw* (kDa)
C1DIQ6	Non-ribosomal peptide synthase	*Avin_25570*	579.04
C1DM35	Type I fatty acid synthase ArsA	*arsA*	267.77
C1DDX7	Glutamate	*Avin_18860*	182.23

	dehydrogenase

The hydrophobicity of proteins in the *S. elongatus* and *A. vinelandii* databases were assessed using R scripts (Appendix file C5) and expressed as the grand average of hydropathy (GRAVY) scores. Negative GRAVY values indicate that the proteins are non-polar, whereas positive values indicate that the proteins are polar. The frequency plots in Figure 4-3 E and F showed that most proteins have GRAVY scores of -0.5 to 0.5, accounting for 93.25% and 90.19 % of *S. elongatus* and *A. vinelandii* protein databases, respectively. Therefore, the difference in physicochemical characteristics between the two species was considered to be minimal based on physicochemical comparisons.

Dynamic range is the range of MS1 peak intensities that enables the detection of peptides.(Zubarev, 2013; Yimer et al., 2017) A broad dynamic range, or a very highly abundant protein in one strain, could affect quantification in co-culture. Therefore, it is crucial to examine the protein dynamic range of each strain in a microbial community before further analysis. The abundance of detected *S. elongatus* and *A. vinelandii* proteins was quantified by absolute quantification (iBAQ). The scatter plots illustrating the dynamic range with log10 iBAQ intensity show that the dynamic range of both strains covered 5 orders of magnitude (Figure 4-3 G and H), suggesting that the influence of dynamic range on MS1 detection would be minimal between the two strains. The 10 most abundant proteins belong to photosynthesis, sucrose biosynthetic process, carbon fixation, regulation of nitrogen utilization and photorespiration for *S. elongatus* cscB/SPS and nitrogen fixation, protein folding for *A. vinelandii* △nifL. The ten most abundant proteins of *S. elongatus* cscB/SPS and *A. vinelandii* △nifL based on iBAQ intensity were listed in Table 4-3. There are several more highly abundant proteins in *S. elongatus* cscB/SPS, which means lower abundance proteins can be masked by this. Therefore, a dynamic exclusion setting needed to consider to minims this effect during the mass spectrometry procedure.

Table 4-3 List of the ten most abundant proteins of *S. elongatus* cscB/SPS and *A. vinelandii* △ nifL based on iBAQ intensity.

Protein IDs	Protein names	Gene names	Log10 (iBAQ)
S. elongatus cscB/SPS			
P13530	C-phycocyanin alpha subunit	*cpcA1; cpcA2*	9.67
P06539	C-phycocyanin beta subunit	*cpcB1; cpcB2*	9.64
Q31RG1	Allophycocyanin, beta subunit	*Synpcc7942_0326*	9.42
Q31RG0	Allophycocyanin alpha chain	*Synpcc7942_0327*	9.41
Q31NL7	Photosystem I reaction center subunit IV	*psaE*	9.31
Q55440	Sucrose-phosphate synthase	*sps*	9.16
Q03511	Carboxysome shell protein	*ccmK2*	9.04
P0A3F4	Nitrogen regulatory protein P-II	*glnB*	9.04
Q31NB2	Ribulose bisphosphate carboxylase small subunit	*cbbS*	8.98
Q31NT9	Photosystem I reaction center subunit III (PSI-F)	*Synpcc7942_1250*	8.91
A. vinelandii △nifL			
C1DIS6	Heat shock Hsp20 protein	*Avin_25770*	8.49
C1DH27	Flavodoxin	*nifF*	8.33
C1DHG0	Phasin protein	*phbP*	8.09
C1DKJ8	Elongation factor Tu (EF-Tu)	*tuf*	8.04
P84253	Molybdenum storage protein subunit beta	*mosB*	8.03
C1DPG7	Cold shock domain family protein	*Avin_11700*	7.97
C1DGZ6	Nitrogenase iron protein	*nifH*	7.95
C1DRH0	Acyl carrier protein (ACP)	*acpP*	7.93

P84308	Molybdenum storage protein subunit alpha	*mosA*	7.93
C1DQC1	Co-chaperonin GroES	*groES*	7.84

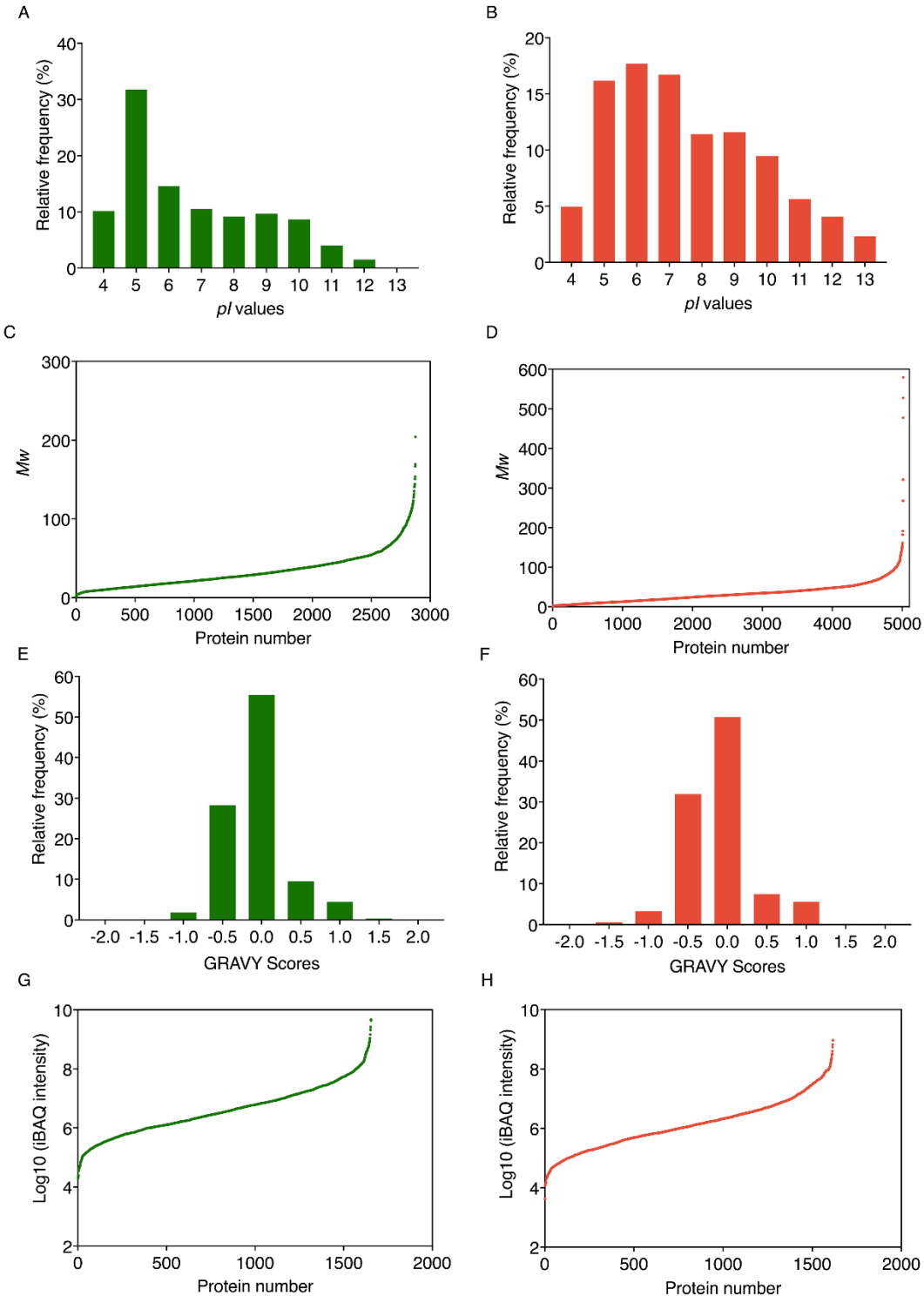


Figure 4-3 Theoretical *pI*, M_w, hydrophobicity values and protein dynamic range estimation of the proteins in *S. elongatus* (A, C, E and G) and *A. vinelandii* (B, D, F and H) databases. (A) Frequency plot of *pI* of *S. elongatus* database. (B) Frequency plot of *pI* of *A. vinelandii* database. (C) M_w range of all proteins in *S. elongatus* database. (D) M_w range of all proteins in *A. vinelandii* database. (E) Frequency plot of GRAVY scores of proteins in *S. elongatus* database. (F) Frequency plot of GRAVY scores of proteins in *A. vinelandii* database. (G) Protein dynamic range estimation of *S. elongatus*. (H) Protein dynamic range estimation of *A. vinelandii*. The iBAQ intensity values for the detected proteins were plotted with log10 iBAQ intensity on the y-axis, and proteins were ranked by iBAQ intensity on the x-axis. Theoretical data were calculated using the R script. *pI*: isoelectric point; M_w: molecular weight; GRAVY: grand average of hydropathy.

4.4.1.3 Proteome size and shared peptides analysis

Database searching is the preferred method for protein identification from digital spectra of the mass-to-charge ratio (m/z) of protein samples detected by a mass spectrometer (Kumar et al., 2017). The quality of the database is one of the main influencing factors in the discovery of proteins present in the sample, including the database size. The size of the database determines the computational power required for analysis and the number of peptides identified from the search, and therefore the biological conclusions drawn. To test the effects of database size on protein identification, the spectra of *S. elongatus* and *A. vinelandii* were searched against the individual databases and a larger, merged database. The Venn diagrams illustrate the identified protein numbers of *S. elongatus* (Figure 4-4 A) and *A. vinelandii* (Figure 4-4 B) searched against individual databases and the merged database, using proteins with two or more unique peptides. Compared with the individual databases, searching against the merged database reduced the number of identified proteins due to the increased complexity of the merged database content affecting the FDR. This effect was greater on the small proteome database (*S. elongatus*, 2876 sequences) compared to the larger proteome database (*A. vinelandii*, 5013 sequences). However, considering less than 2% of proteins were not identified when searching against the larger, merged database, the merged database was used in our analyses. However, this should be considered in co-culture proteomics workflows, particularly where proteome sizes of co-culture members vary more widely.

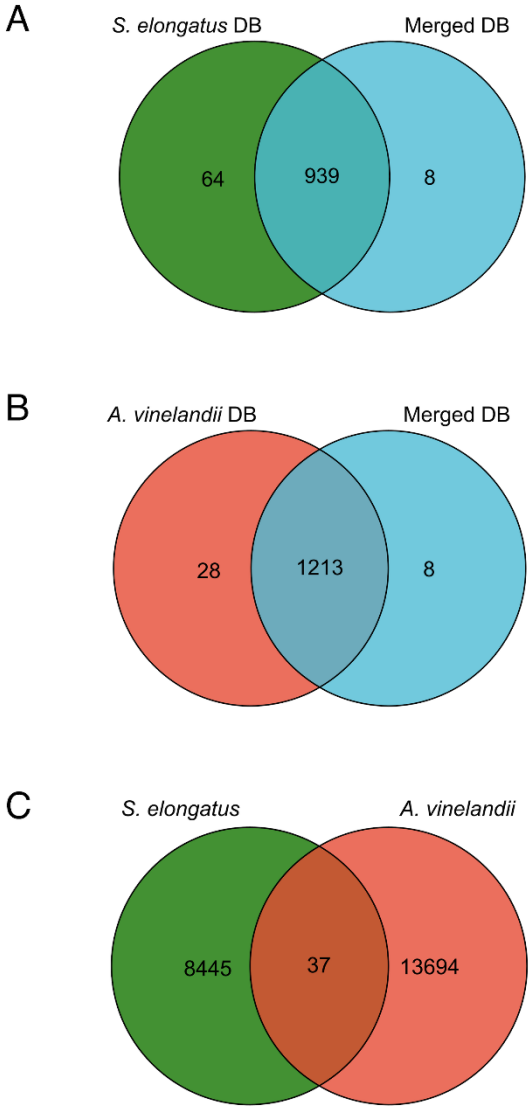


Figure 4-4 Venn diagrams showing identified protein numbers of *S. elongatus* (A), *A. vinelandii* (B) and shared peptides between *S. elongatus* and *A. vinelandii* (C). (A) Identified protein numbers of *S. elongatus* searching against the individual *S. elongatus* database (green) and merged *S. elongatus* and *A. vinelandii* database (blue). (B) Identified protein numbers of *A. vinelandii* searching against the individual *A. vinelandii* database (red) and merged *S. elongatus* and *A. vinelandii* database (blue). Venn diagrams were generated using Hiplot. (C) Shared and unique peptides detected between *S. elongatus* and *A. vinelandii*. DB: database.

The LFQ experiment was aimed at identifying proteins in *S. elongatus* and *A. vinelandii* co-

cultures. Therefore, it is crucial to consider the peptides that are shared between the two organisms. For example, peptides from *S. elongatus* might be matched to similar *A. vinelandii* proteins, and thus influence protein quantification (Helliwell et al., 2018). Theoretical shared peptides were calculated using R scripts (Appendix file C6) and 0.19 % of tryptic peptides of the specified size (8 to 25 amino acids) were shared between *S. elongatus* PCC 7942 and *A. vinelandii* DJ. Measured shared peptides were also analyzed by comparison of the resulting peptide sequences of each strain, and 0.17 % of shared peptides were obtained (Figure 4-4 C), which was similar to the theoretical value. Since this is deemed to be a low number, the assumption was made unlikely that shared peptides will significantly interfere with the protein identification and quantification of each organism.

4.4.2 Assessment of different quantification methods at protein level

To generate an accurate proteome quantification, it is crucial to test the relationships between different quantification methods and actual protein abundance. Nine synthetic protein mixes were made by mixing the extracted proteins of *S. elongatus* and *A. vinelandii* at ratios of 100:0, 95:5, 90:10, 75:25, 50:50, 25:75, 10:90, 5:95, and 0:100 to total amounts of 10 µg (i.e. a 95:5 ratio mix is 9.5 µg *S. elongatus* protein with 0.5 µg *A. vinelandii* protein) (Figure 4-5 A). Different methods for quantifying biomass contribution were validated by HPLC-MS/MS.

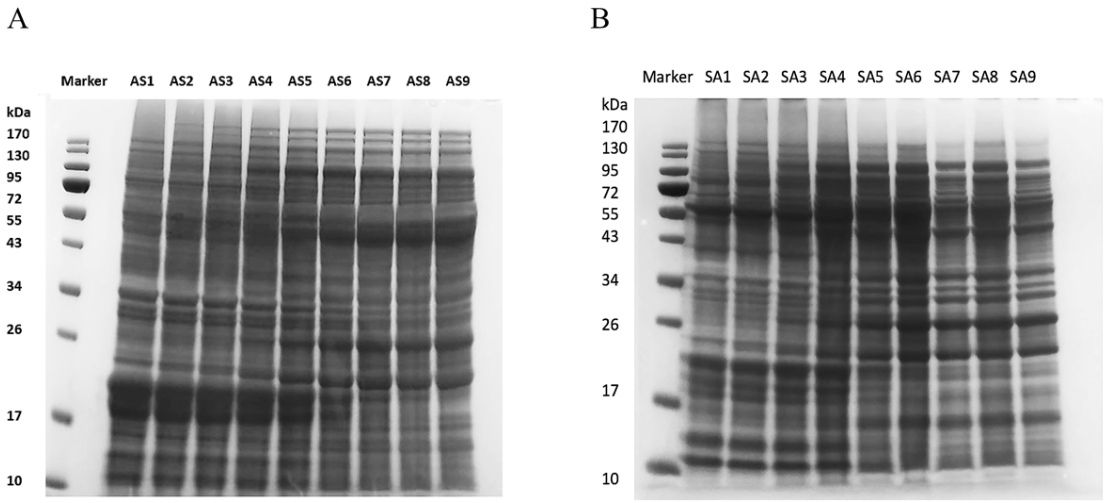


Figure 4-5 SDS-PAGE gel shows 2D cleaned-up proteins of protein mixes (A) and cell mixes (B)

Here, three approaches based on the spectral data (i.e., PSMs, unique peptide, and NSAF) and three approaches based on intensity (i.e., total intensity, iBAQ intensity, and LFQ intensity) were assessed for the proteome quantification. This is based on the assumption that the abundance of a protein is reflected by the proportion of its counts or intensity to the total counts and intensities as shown in equations 4-1 to 4-6.

$$\frac{\text{Protein mass}}{\text{Total protein mass}} \approx \frac{\text{Protein PSMs count}}{\text{Total PSMs count}} \quad (\text{Equation } 4-1)$$

$$\frac{\text{Protein mass}}{\text{Total protein mass}} \approx \frac{\text{Protein unique peptide count}}{\text{Total unique peptide count}} \quad (\text{Equation } 4-2)$$

$$\frac{\text{Protein mass}}{\text{Total protein mass}} \approx \frac{\text{Protein NSAF count}}{\text{Total NSAF count}} \quad (\text{Equation } 4-3)$$

$$\frac{\text{Protein mass}}{\text{Total protein mass}} \approx \frac{\text{Protein intensity}}{\text{Total intensity}} \quad (\text{Equation } 4-4)$$

$$\frac{\text{Protein mass}}{\text{Total protein mass}} \approx \frac{\text{Protein iBAQ intensity}}{\text{Total iBAQ intensity}} \quad (\text{Equation } 4-5)$$

$$\frac{\text{Protein mass}}{\text{Total protein mass}} \approx \frac{\text{Protein LFQ intensity}}{\text{Total LFQ intensity}} \quad (\text{Equation } 4-6)$$

4.4.2.1 Quantification based on spectral counting

The protein abundance contributions of *S. elongatus* cscB/SPS and *A. vinelandii* △nifL protein mixtures were estimated using three spectral counts-based quantification methods, i.e., PSMs count, unique peptide count, and NSAF. The theoretical protein ratios and quantified protein ratios by different counts-based parameters of *S. elongatus* cscB/SPS and *A. vinelandii* △nifL were shown in Figure 4-6. The trends between the theoretical protein ratios and measured protein ratios using different parameters were the same, which is reasonable. However, there is a bias between the theoretical values and measured values. Therefore, to visualise the accuracy of protein quantification by different parameters, the relative errors between the theoretical protein ratios and the measured protein ratios were calculated based on Equation 4-7. It can be seen that protein ratios quantified by unique peptides show very large relative errors, whereas the relative errors of PSMs and NSAF-based quantification were small (Figure 4-7). This is consistent with the relationship between absolute protein quantification by different

parameters and protein amounts shown in Figure 4-8, where there is a higher R-squared (R^2) for PSMs and NSAF quantification. Assumptions of equations 1, 2, and 3 were also be verified by Figure 4-8, in which absolute protein quantification by all different count-based parameters showed high level of correlation with protein amounts with all R^2 values of greater than 0.89.

$$\text{Relative error} = \frac{\text{Measured value} - \text{real value}}{\text{real value}} \quad (\text{Equation } 4-7)$$

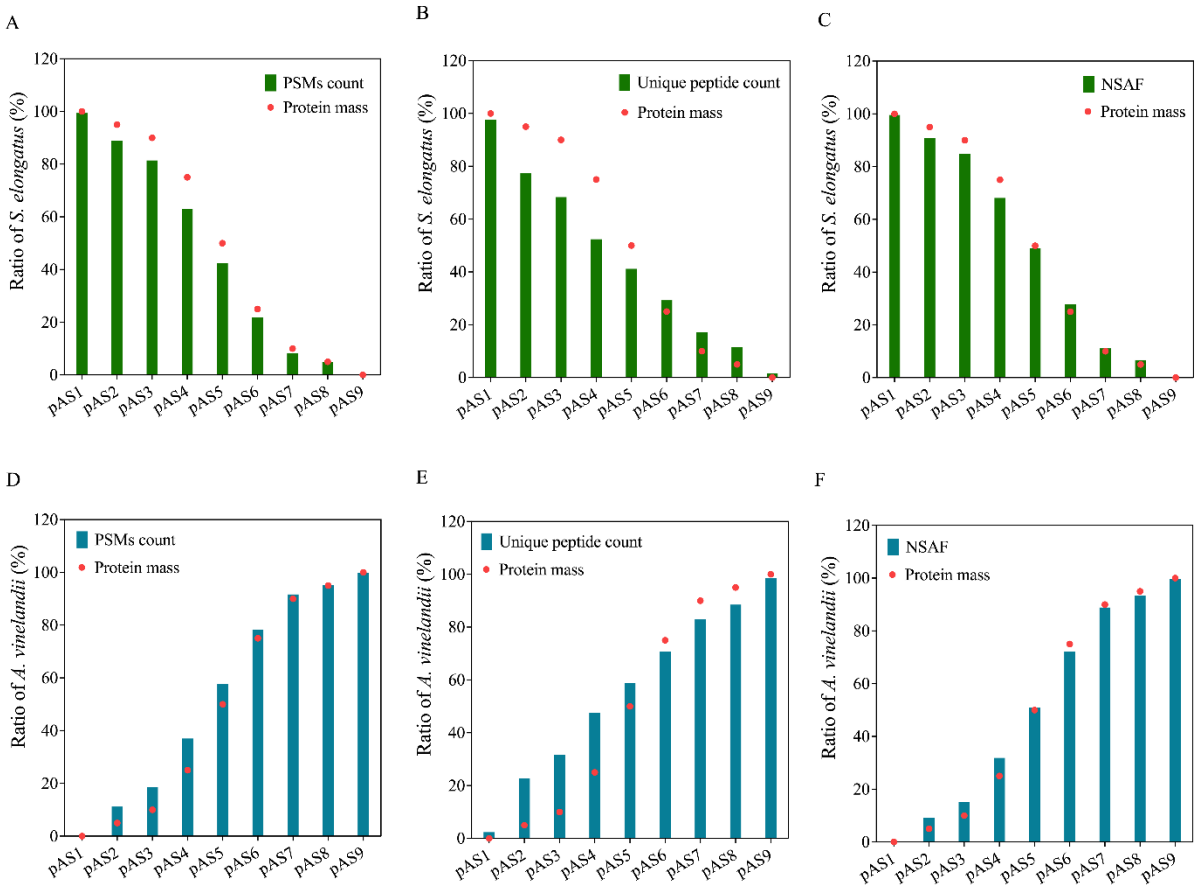


Figure 4-6 Theoretical protein ratios (red dots) and quantified protein ratios by different count-based methods (PSMs count, unique peptide count, and NSAF; columns) of *S. elongatus* cscB/SPS (A, B, and C) and *A. vinelandii* △nifL (D, E, and F) at protein level. PSMs and unique peptide count were from MaxQuant output.

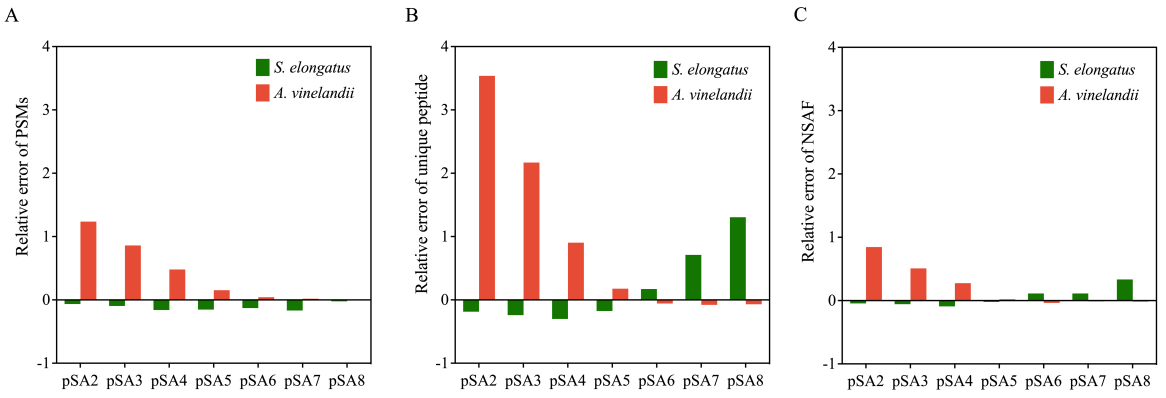


Figure 4-7 Relative errors of different count-based quantification methods of *S. elongatus* cscB/SPS (red) and *A. vinelandii* △nifL (green) at protein level. Value of 1 in y axis means that measured value is 2-fold more than the theoretical value. Relative error is equal to the measured protein ratio minus theoretical protein ratio and then divided by theoretical protein ratio.

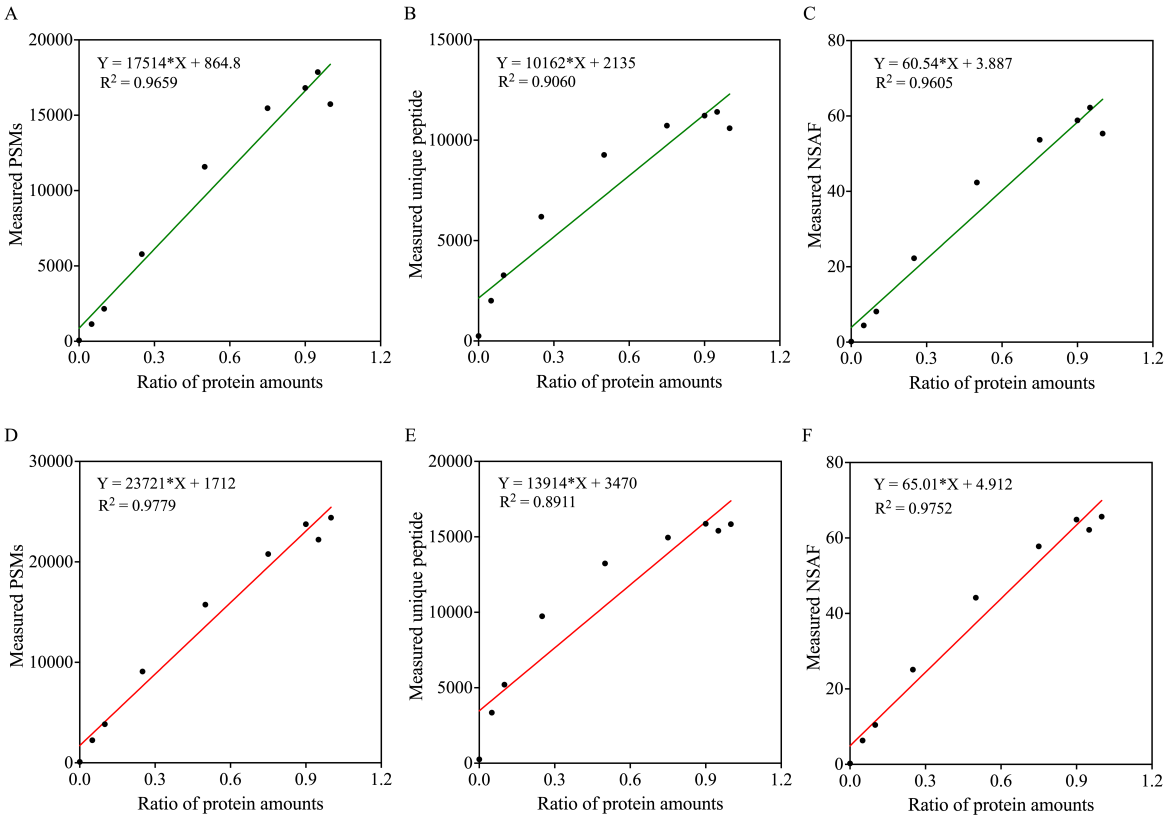


Figure 4-8 Relationship between absolute protein quantifications calculated using different count-based methods, and protein amounts in co-cultures of *S. elongatus* (top, green) and *A.*

vinelandii (bottom, red) at the protein level. i.e. 0.2 refers to 20% of the protein mix being sourced from the specified organism. Protein quantification of *S. elongatus* samples was achieved using three count-based methods: (A) PSMs, (B) unique peptide, and (C) NSAF. Protein quantification of *A. vinelandii* samples was achieved using three count-based methods: (D) PSMs, (E) unique peptide, and (F) NSAF. Linear fittings were analysed using GraphPad Prism.

4.4.2.2 Quantification based on intensity

Three intensity-based quantification methods (total intensity, iBAQ intensity, LFQ intensity) were assessed for the protein abundance contribution of *S. elongatus* cscB/SPS and *A. vinelandii* △nifL at protein level. IBAQ intensity is the total intensity divided by the number of peptides of one protein, while LFQ intensity is similar to the iBAQ intensity, but the intensities are normalised by the built-in algorithm in MaxQuant to exclude some outliers (Cox et al., 2014).

The theoretical protein ratios and quantified protein ratios by different intensity-based parameters of *S. elongatus* cscB/SPS and *A. vinelandii* △nifL were shown in Figure 4-9. The trends between the theoretical protein ratios and measured protein ratios using different parameters were the same. Figure 4-10 showed that the relative errors between the theoretical protein ratios and the measured protein ratios were small, suggesting that the intensity-based quantification is closer to actual protein ratio than the spectral counts-based methods. The relationship between absolute protein quantification by different intensity parameters and protein amounts is shown in Figure 4-11. All the total values of different measured intensities have good linear relationship with the absolute protein amount, which verified assumption equations 4-4, 4-5, and 4-6.

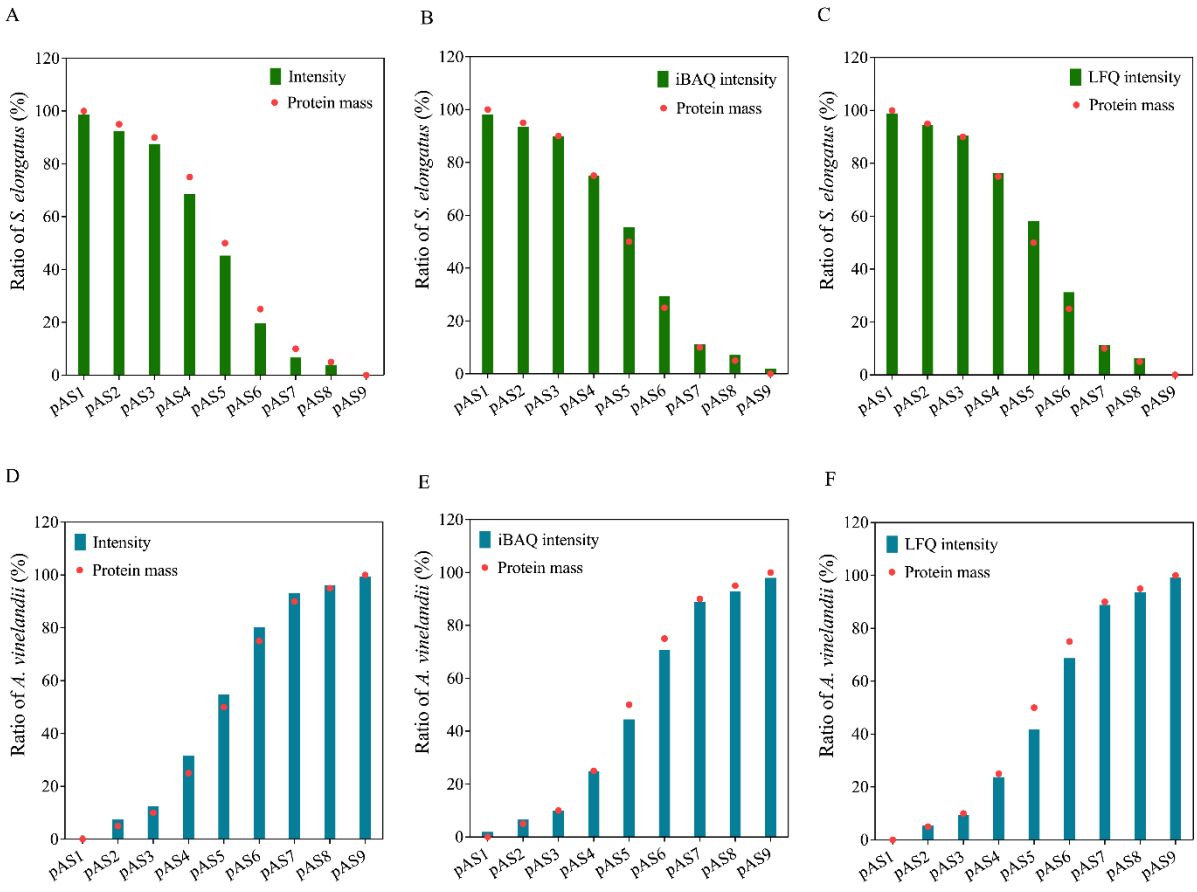


Figure 4-9 Theoretical protein ratios (red dots) and quantified protein ratios by different count-based methods (total intensity, iBAQ intensity, and LFQ intensity; columns) of *S. elongatus* cscB/SPS (A, B, and C) and *A. vinelandii* △nifL (D, E, and F) at protein level. Total intensity, iBAQ intensity, and LFQ intensity values were from MaxQuant outputs.

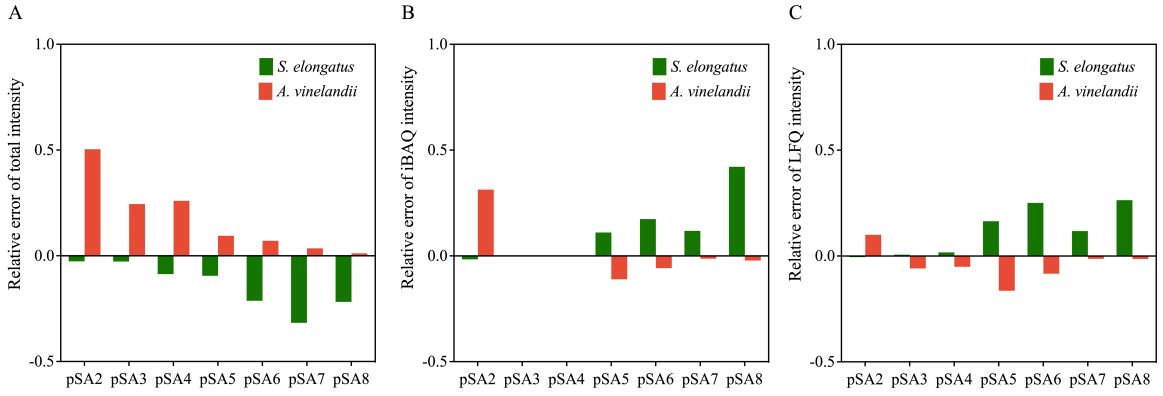


Figure 4-10 Relative errors of different intensity-based quantification methods of *S. elongatus* cscB/SPS (red) and *A. vinelandii* △nifL (green) at protein level. Relative error is equal to the

measured protein ratio minus theoretical protein ratio and then divided by theoretical protein ratio.

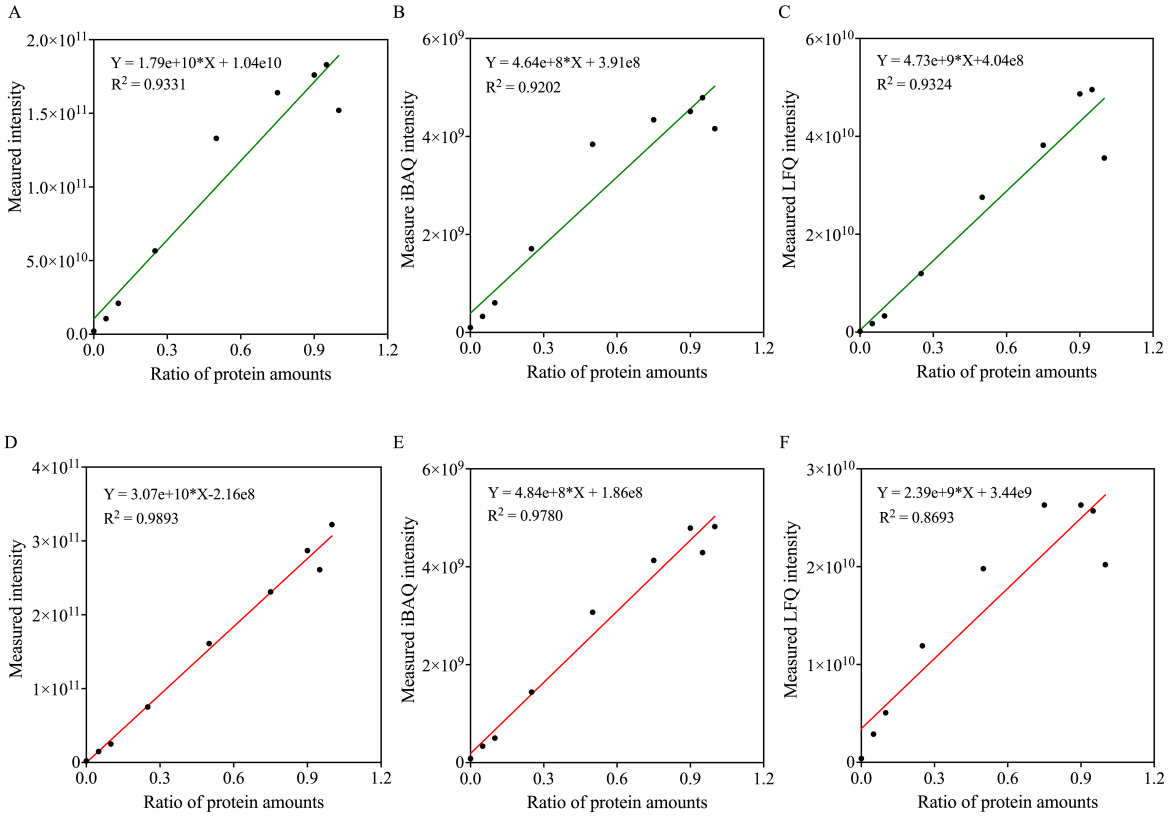


Figure 4-11 Relationship between absolute protein quantification by different intensity-based methods and protein amount ratios of *S. elongatus* (top, green) and *A. vinelandii* (bottom, red) at the protein level. Protein quantification of *S. elongatus* samples was achieved using three intensity-based methods: (A) intensity, (B) iBAQ intensity, and (C) LFQ intensity. Protein quantification of *A. vinelandii* samples was achieved using three intensity-based methods: (A) intensity, (B) iBAQ intensity, and (C) LFQ intensity. Linear fittings were analysed using GraphPad Prism.

Although the results demonstrate good linear relationships between each of the parameters tested and ratio of protein amounts at different protein mixtures (Figure 4-8 and Figure 4-11), the relative errors of quantification using the intensity-based methods were much smaller than those using count-based methods (Figure 4-7 and Figure 4-10). This is because spectral counts, which represent the number of MS2 spectra assigned to each protein, include all redundancies

of peptide identification, such as charge states, missed cleavages, modifications, and multiple detections of the same peptide resulting from the expired dynamic exclusion (Blein-Nicolas and Zivy, 2016). These redundancies may obscure the relationship between the spectral counts and protein abundance, especially when the machine settings are changed (Goeminne et al., 2018). Therefore, relative label-free quantification calculated using ion intensities enables more accurate proteome quantification in this synthetic co-culture system.

4.4.3 Assessment of different quantification method at cell level

Although mixing proteins extracted from *S. elongatus* and *A. vinelandii* cells in pre-defined ratios provides insight into protein quantification, showing good linear correlation to spectral counts and intensity, in actual proteomics experiments, cell numbers can vary in ratio in different conditions or over time. Therefore, to investigate the impact of this phenomenon, *S. elongatus* and *A. vinelandii* cells were mixed at nine pre-determined cell number ratios of 100:0, 95:5, 90:10, 75:25, 50:50, 25:75, 10:90, 5:95, and 0:100 to assess different quantification methods at the cell level (Figure 4-5 B).

Three approaches based on the spectral data (i.e., PSMs, unique peptide, and NSAF) and three approaches based on intensity (i.e., total intensity, iBAQ intensity, and LFQ intensity) were assessed. The premise of using these approaches is to assume that the cell abundance is reflected by the proportion of its counts or intensity to the total counts and intensities as shown in equations 4-8 to 4-13.

$$\frac{\text{Cell number of one strain}}{\text{Total cell number}} \approx \frac{\text{PSMs count of one strain}}{\text{Total PSMs count}} \quad (\text{Equation } 4-8)$$

$$\frac{\text{Cell number of one strain}}{\text{Total cell number}} \approx \frac{\text{Unique peptide count of one strain}}{\text{Total unique peptide count}} \quad (\text{Equation } 4-9)$$

$$\frac{\text{Cell number of one strain}}{\text{Total cell number}} \approx \frac{\text{NSAF count of one strain}}{\text{Total NSAF count}} \quad (\text{Equation } 4-10)$$

$$\frac{\text{Cell number of one strain}}{\text{Total cell number}} \approx \frac{\text{Intensity of one strain}}{\text{Total intensity}} \quad (\text{Equation } 4-11)$$

$$\frac{\text{Cell number of one strain}}{\text{Total cell number}} \approx \frac{\text{IBAQ intensity of one strain}}{\text{Total iBAQ intensity}} \quad (\text{Equation } 4-12)$$

$$\frac{\text{Cell number of one strain}}{\text{Total cell number}} \approx \frac{\text{LFQ intensity of one strain}}{\text{Total LFQ intensity}} \quad (\text{Equation } 4-13)$$

4.4.3.1 Quantification based on spectral counts

The protein contributions of *S. elongatus* cscB/SPS and *A. vinelandii* △nifL cell mixtures were estimated using three spectral counts-based quantification methods, i.e., PSMs count, unique peptide count, and NSAF. The theoretical cell ratios and quantified cell ratios by different counts-based parameters of *S. elongatus* cscB/SPS and *A. vinelandii* △nifL were shown in Figure 4-12. The trends between the theoretical cell ratios and measured cell ratios using different parameters were the same, but the relative errors between them were very large (Figure 4-13). Fitting plots for *S. elongatus* exhibited good linear relationships, with R^2 values > 0.9 between absolute protein quantification by different count-based methods and cell numbers (Figure 4-14 A, B, and C), whereas poor linear relationships were observed in *A. vinelandii* (Figure 4-14 D, E, and F), revealing weak correlation between cell number and spectral counts.

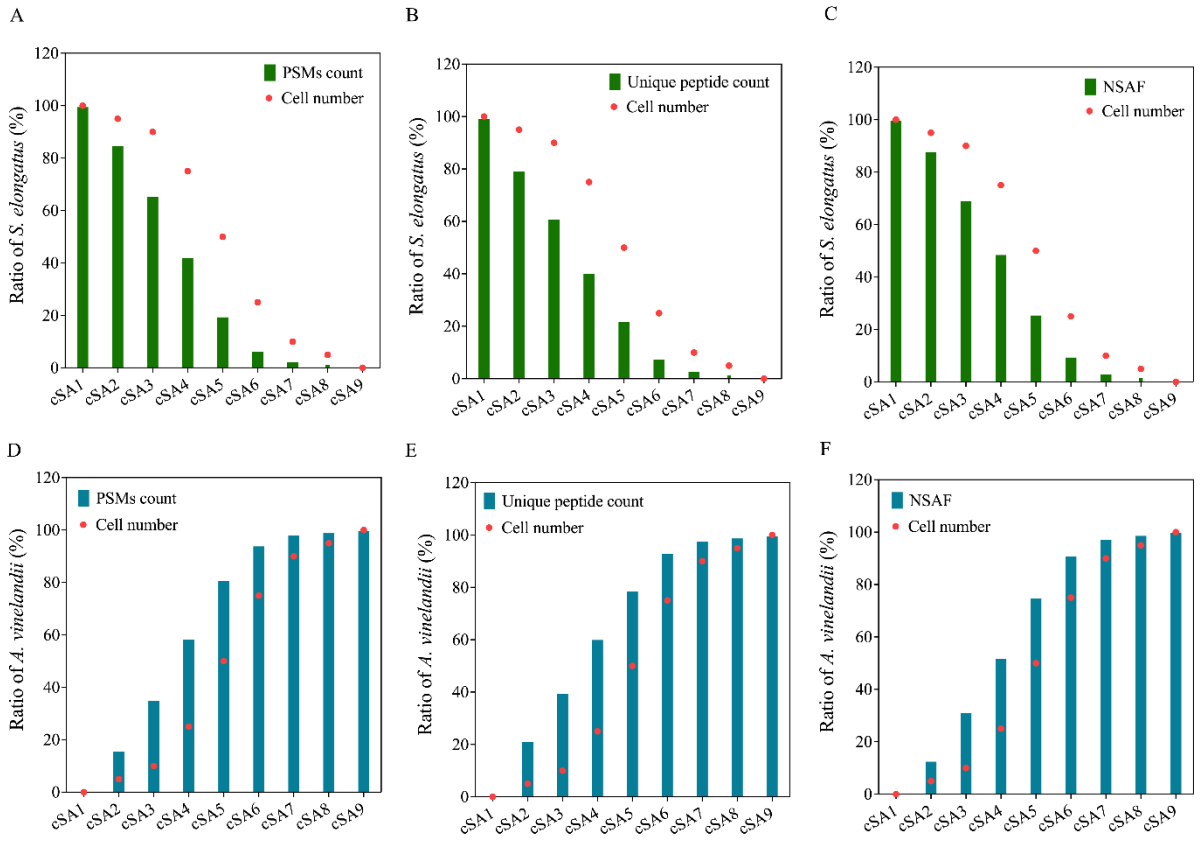


Figure 4-12 Theoretical protein ratios (red dots) and quantified protein ratios by different count-based methods (PSMs count, unique peptide count, and NSAF; columns) of *S. elongatus* cscB/SPS (A, B, and C) and *A. vinelandii* △nifL (D, E, and F) at cell level. PSMs and unique

peptide count were from MaxQuant output. NSAF is equal to PSMs count divided by protein length.

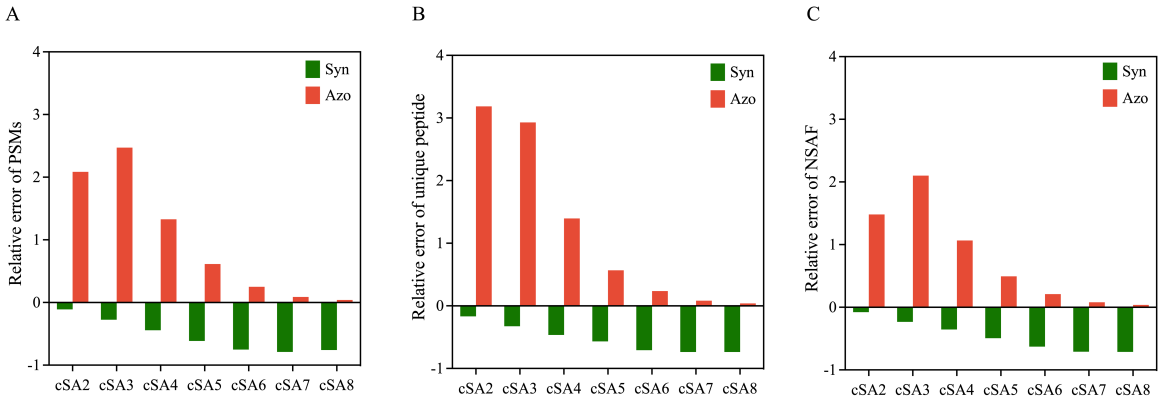


Figure 4-13 Relative errors of different count-based quantification methods of *S. elongatus* cscB/SPS (red) and *A. vinelandii* △nifL (green) at cell level. Relative error is equal to the measured cell ratio minus theoretical cell ratio and then divided by theoretical cell ratio.

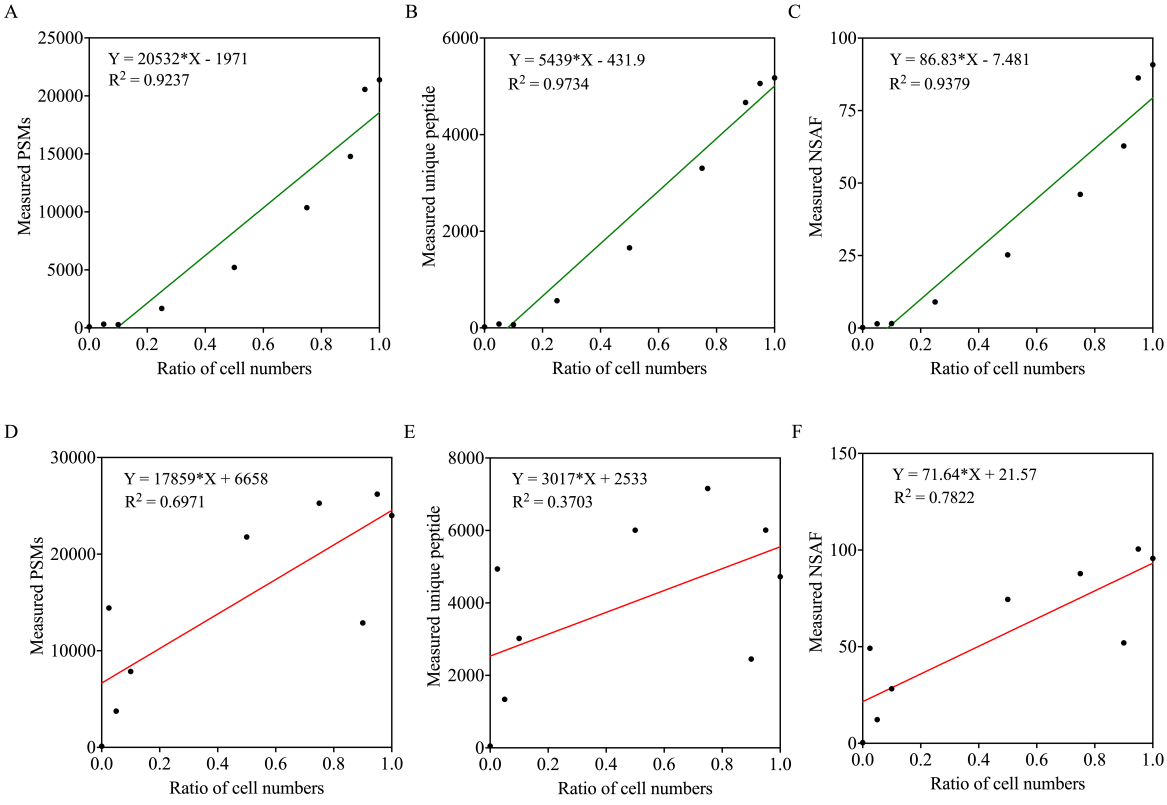


Figure 4-14 Relationship between absolute protein quantification by different intensity-based

methods of *S. elongatus* (top, green) and *A. vinelandii* (bottom, red) mixture at the cell level. Protein quantification of 9 different cell ratios of *S. elongatus* samples was achieved using three count-based methods: (A) PSMs, (B) unique peptide, and (C) NSAF. Protein quantification of 9 different cell ratios of *A. vinelandii* samples was achieved using three count-based methods: (D) PSMs, (E) unique peptide, and (F) NSAF. Linear fittings were analysed using GraphPad Prism.

4.4.3.2 Quantification based on intensity

To assess the protein abundance contribution of *S. elongatus* cscB/SPS and *A. vinelandii* △nifL at cell level, three intensity-based quantification methods (total intensity, iBAQ intensity, LFQ intensity) were performed.

The theoretical cell ratios and quantified protein ratios by different intensity-based parameters of *S. elongatus* cscB/SPS and *A. vinelandii* △nifL were shown in Figure 4-15. The trends between the theoretical protein ratios and measured protein ratios using different parameters were the same but with some distinctions, especially in cSA3 to cSA7. Figure 4-16 shows that the relative errors between the theoretical cell ratios and the measured protein ratios of all these three parameters (intensity, iBAQ intensity and LFQ intensity) are large, indicating that the intensity-based quantification method is not suitable for measuring the actual protein abundance in cell mixture samples. Linear regressions were also generated between absolute protein quantification calculated using different intensity-based methods and the cell numbers of *S. elongatus* and *A. vinelandii* samples mixed at pre-determined cell number ratios (Figure 4-17). Very poor correlation between cell number and spectral intensity was observed for *A. vinelandii*, which overturned the assumption equations 4-4, 4-5, and 4-6.

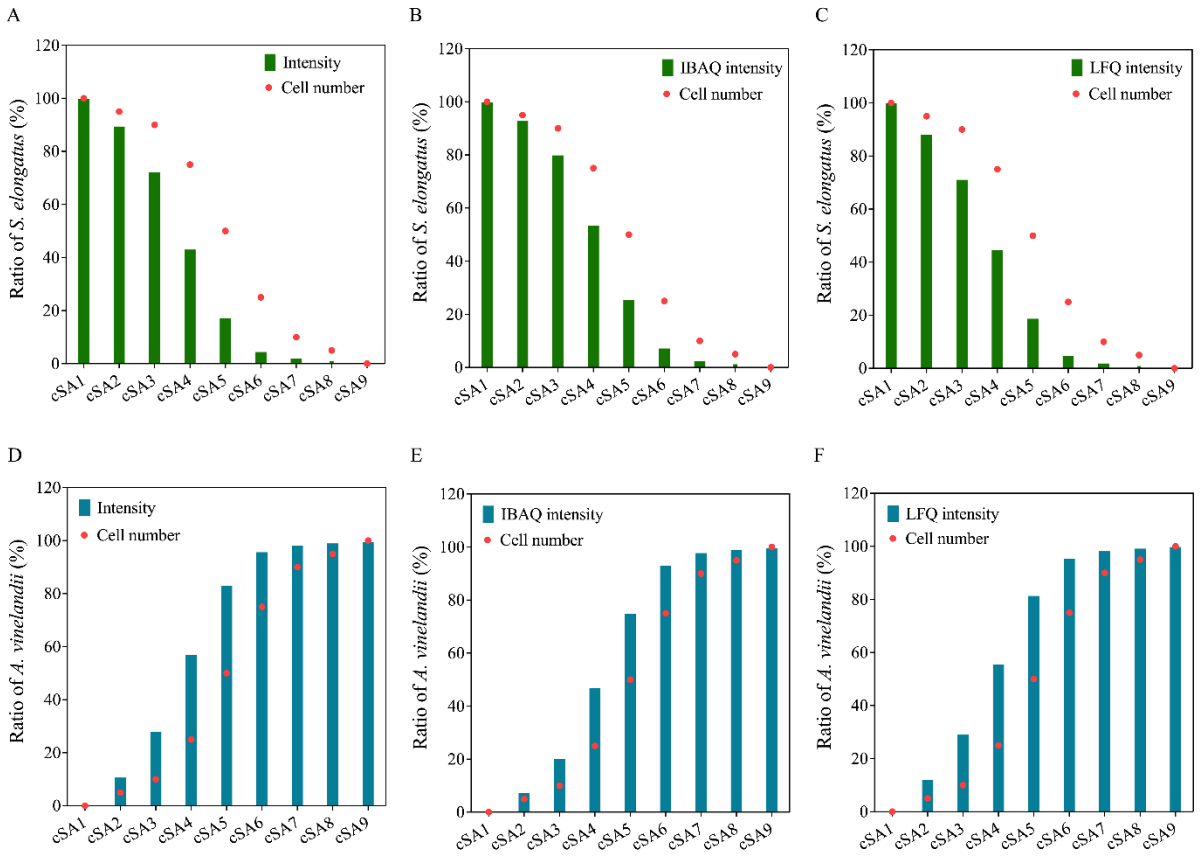


Figure 4-15 Theoretical protein ratios (red dots) and quantified protein ratios by different count-based methods (total intensity, iBAQ intensity, and LFQ intensity; columns) of *S. elongatus* cscB/SPS (A, B, and C) and *A. vinelandii* △nifL (D, E, and F) at cell level. Total intensity, iBAQ intensity, and LFQ intensity values were from MaxQuant outputs.

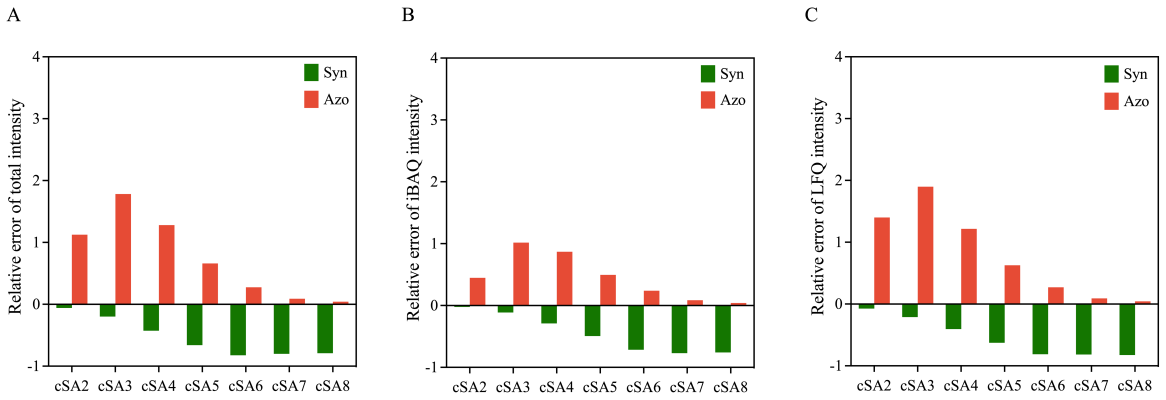


Figure 4-16 Relative errors of different intensity-based quantification methods of *S. elongatus* cscB/SPS (red) and *A. vinelandii* △nifL (green) at cell level. Relative error is equal to the measured protein ratio minus theoretical protein ratio and then divided by theoretical protein

ratio.

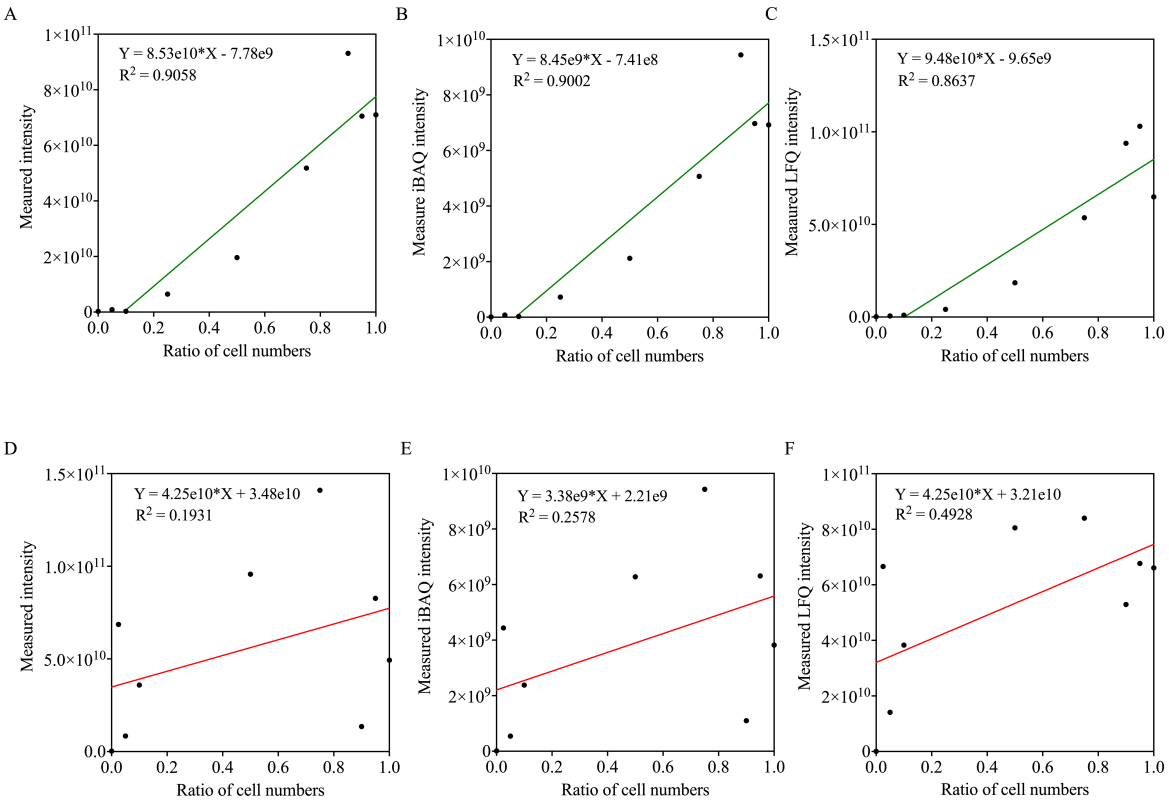


Figure 4-17 Relationship between absolute protein quantification by different intensity-based methods and cell number ratios of *S. elongatus* (top, green) and *A. vinelandii* (bottom, red) at the cell level. Protein quantification of *S. elongatus* samples of 9 different ratios was achieved using three intensity-based methods: (A) intensity, (B) iBAQ intensity, and (C) LFQ intensity. Protein quantification of *A. vinelandii* samples of 9 different ratios was achieved using three intensity-based methods: (A) intensity, (B) iBAQ intensity, and (C) LFQ intensity. Linear fittings were analysed using GraphPad Prism.

This section validated that all the proteomic quantification methods showed a good linear relationship between the quantification parameters and protein amounts when mixes were undertaken at protein level, but not in the cell level mixes. This is because when loading same amount of proteins into mass spectrometer, it is easy to do for pure culture, but it is impossible to figure out the protein amount for each strain in co-culture. This means there is a big bias in protein quantification using different proteome quantification proxy in co-culture. Therefore,

a normalisation should be performed in the cell mixture samples to eliminate the influence at cell level.

4.4.4 Normalisation of co-culture

4.4.4.1 Normalising protein quantification data by cell number alone is unsuitable for quantifying proteins in mixed cultures

Previous studies showed that weak correlations were found between cell number and spectral counts or spectral intensities in *A. vinelandii*, revealing that the protein quantification is not accurate at the cell level. Therefore, a method was developed to adjust protein quantification values, accounting for cell numbers. LFQ intensity protein quantification data were used in the calculation, as the LFQ intensity is a good proxy for protein amount with low relative error (Figure 4-10). The normalisation method was achieved by the calculation that, for each individual protein detected, its LFQ intensity was divided by the number of cells for its respective strain (Equation 4-14).The values obtained are theoretically the same, as the total protein amount per cell in different cell mixtures of *S. elongatus* or *A. vinelandii* is the same.

$$\text{Normalised LFQ intensity} = \frac{\text{LFQ intensity of one strain}}{\text{cell number of the strain}} \quad (\text{Equation } 4-14)$$

To validate our method for normalization by cell number, protein amounts per cell of *S. elongatus* (Figure 4-18 A) and *A. vinelandii* (Figure 4-18 B) were calculated from normalized LFQ intensity using the linear relationship equations between LFQ intensity and protein amount shown in Figure 4-11 C and F. The inferred protein amount per cell varied widely in different cell mixtures, with relative standard deviation (RSD) of 83% and 68% for *S. elongatus* and *A. vinelandii*, respectively (Figure 4-18 C), which was also far from the actual protein amount per cell values of 5.49×10^{-7} μg for *S. elongatus* and 8.36×10^{-7} μg for *A. vinelandii* (Figure 4-18 D). Differentially expressed proteins (DEPs) were also analyzed using in-browser LFQ-Analyst software (https://analyst-suite.monash-proteomics.cloud.edu.au/apps/lfq-analyst//) (Shah et al., 2020); about 66.19% of all proteins were identified as DEPs among all pairwise comparisons. However, to generate the artificial cell mixes, we mixed cells harvested from the same *S. elongatus* and *A. vinelandii* cultures to minimize variation; therefore, we did not expect to see differential expression between proteins in each cell mixture. This verified that the normalisation method using LFQ intensity divided by cell number does not apply to

cell mixtures.

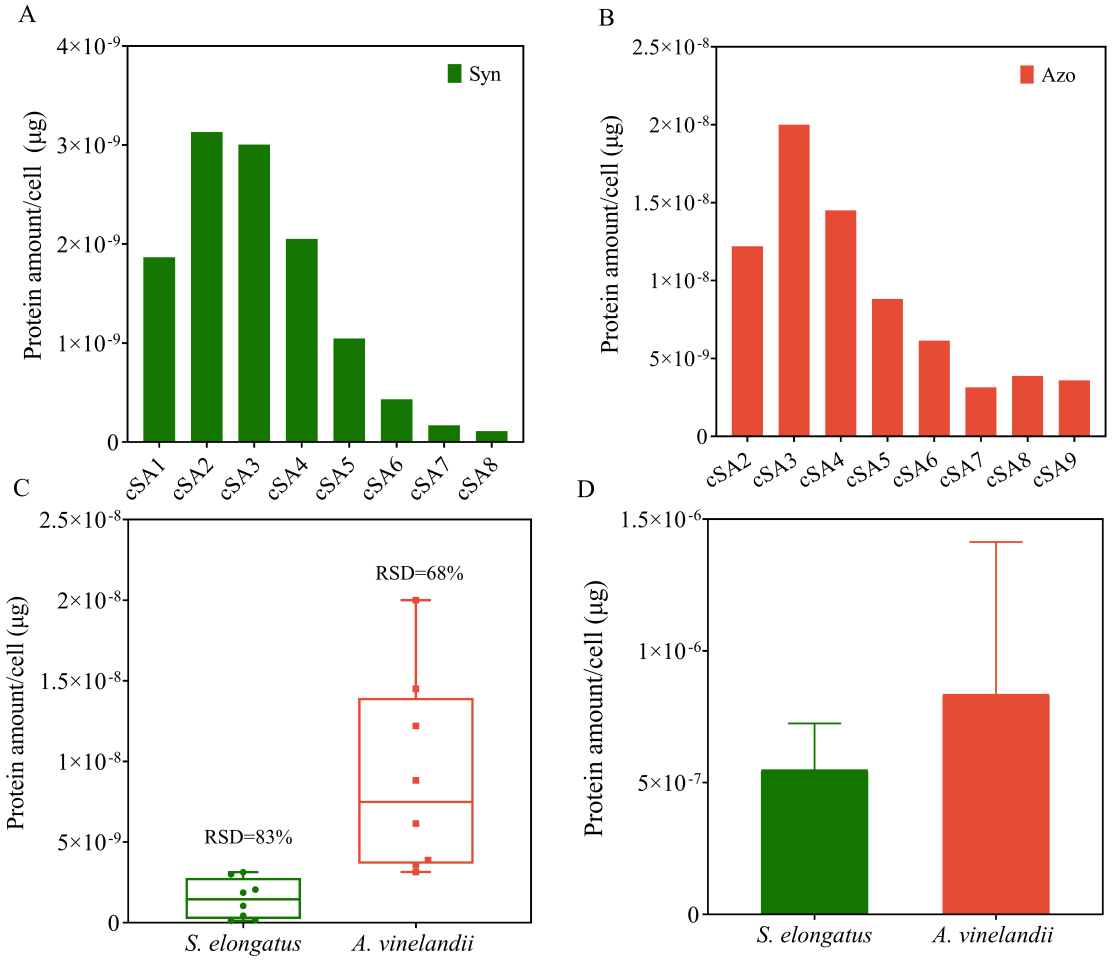


Figure 4-18 Normalisation analysis of proteomic data in cell mixes samples based on cell number. Normalised LFQ intensities were converted to protein amount per cell using the linear relationship equations between LFQ intensity and protein amount. A. Protein amount per cell of *S. elongatus* cscB/SPS in cSA1 to cSA8 after normalisation. B. Protein amount per cell of *A. vinelandii* △nifL in cSA2 to cSA9 after normalisation. C. Distribution of protein amount per cell with RSD across different cell mixtures of *S. elongatus* cscB/SPS (green) and *A. vinelandii* △nifL (red) calculated by normalised LFQ intensity. RSD% = SD / Mean. RSD: Relative standard deviation. SD: Standard Deviation. D. Actual protein amount per cell of *S. elongatus* cscB/SPS (green) and *A. vinelandii* △nifL (red) with five different cell numbers and two replicates of each.

4.4.4.2 Normalisation by LFQ intensity ratio

To minimize the impact of cell number changes, a normalization method should be developed for co-culture proteomics. Considering the biological states of the *S. elongatus* and *A. vinelandii* cells were the same across the different cell mix samples, we assumed that each protein amount per cell should be the same across different cell mix samples. This means there are no differentially expressed proteins among sample mixtures. Based on this, the protein abundances for each cell mix ratio were normalized using a different method: for each individual protein detected, its LFQ intensity was divided by the sum of all protein LFQ intensities for its respective strain (Equation 4-15), named LFQRatio normalization.

$$\text{Normalised LFQ intensity} = \frac{\text{LFQ intensity of one protein}}{\text{total LFQ intensities of the strain}} \quad (\text{Equation } 4-15)$$

To validate the LFQRatio normalization method, the pairwise correlation of LFQRatio normalized results for *S. elongatus* and *A. vinelandii* ratios of 10:90, 20:80, 30:70, 40:60, 50:50, 60:40, 70:30, 80:20, and 90:10 were analyzed (Figure 4-19), which exhibited good pairwise relationships for the cell mixtures ($R^2 > 0.90$). Differential protein expression was analyzed in the LFQRatio normalized cell mixes, revealing that only 0.05% of proteins were significantly differentially expressed among all pairwise group comparisons. Therefore, the LFQRatio normalisation approach was considered suitable for normalizing co-culture quantitative proteomics data via minimising the influence of cell number changes on protein quantification.

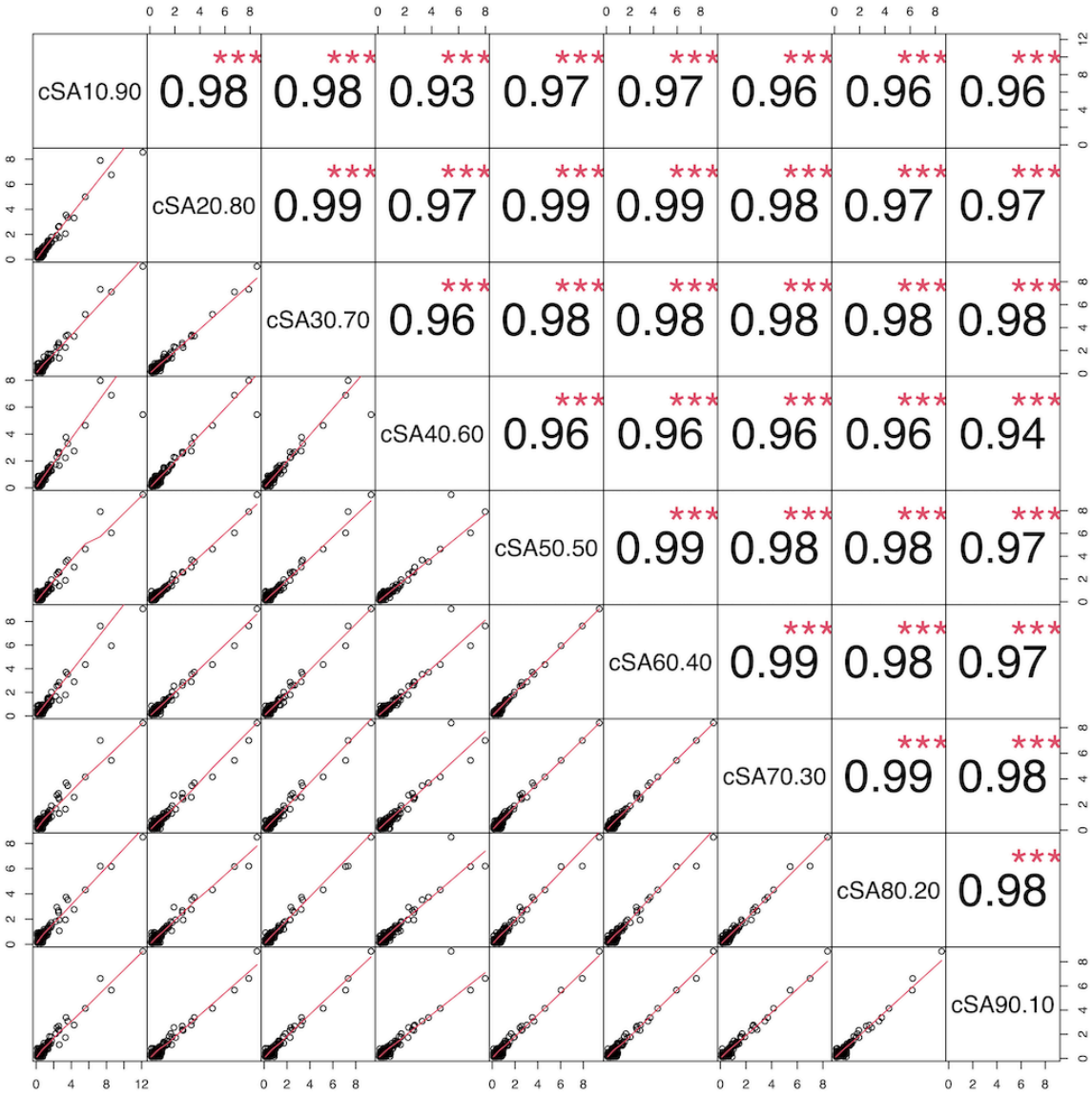


Figure 4-19 Pairwise correlation plots of LFQRatio normalised cell mixes. The Scatterplot matrix represents the entire dataset of *S. elongatus* and *A. vinelandii* ratios of 10:90, 20:80, 30:70, 40:60, 50:50, 60:40, 70:30, 80:20, and 90:10 without missing values. The red lines represent the pairwise linear regression slope. Regression equations and R-squared are shown in each plot. Correlation coefficients were calculated using the Pearson method.

4.5 Conclusion

Multiple factors affect the quantification of peptides from individual strains in co-culture proteomics, including physicochemical and bioinformatics aspects. Protein quantification was

assessed using six different quantification methods, verifying a good linear relationship between the protein amount and the six selected parameters at the protein level. Different ratios of cell mixes were constructed to mimic the co-culture system, which revealed that the correction between cell numbers and quantification parameters can be poor. Therefore, a new normalization method, “LFQRatio”, was proposed, which minimizes the influence of multiple factors such as protein extraction efficiency, different cell numbers and cultivation conditions on proteome quantification. A standard workflow was present to determine the individual proteome responses of two different strains cultivated in the same vessel. This will enable researchers to gain new insights into multi-strain interactions and their mutual impact on metabolic processes, which were previously unattainable.

Chapter 5 Proteomics analysis of a synthetic microbial community – *S. elongatus* cscB/SPS and *A. vinelandii* ΔnifL

5.1 Abstract

Microbial communities play a key role in a wide range of ecosystems and have the potential to enhance the bioproduction of desired chemicals. Understanding the interactions among members is the key to the rational design of microbial communities. In this chapter, label-free shotgun proteomics workflow, a high-throughput technique, established in Chapter 4, was used to analyse the proteome changes between the *A. vinelandii* △nifL and *S. elongatus* cscB/SPS monocultures and the synthetic co-culture that was established in Chapter 3. Characterisation of functional categories and changes was achieved in monocultures and co-culture. Experiments showed a higher relative abundance of proteins in nitrogen fixation in *A. vinelandii* △nifL and some proteins in the photosynthesis pathway in *S. elongatus* cscB/SPS after co-culture, indicating a cross-feeding between these microorganisms. Metal limitations were observed from the proteomic results, which provide a basis for the rational design of the co-culture medium. Furthermore, the proteome changes of the synthetic co-culture over time were investigated, revealing potential targets for optimisation to keep the synthetic microbial co-culture healthier and longer-lived.

5.2 Introduction

Synthetic microbial consortia have gained increased attention for industrial bioproduction of a variety of products. Recent advances have improved the ability to control temporal, spatial, and community compositional organization (Grandel et al., 2021). However, the mechanisms of how the members interact with each other, via processes such as predation, competition, mutualism or symbiosis, remain poorly understood (Christie-Oleza et al., 2017; La Sarre et al., 2017).

There have been attempts at mathematical models and numerical simulation to uncover how the stability and productivity of the synthetic consortia are affected by the system structure and parametrisation (Johns et al., 2016; Di and Yang, 2019). Integrated omics techniques including genomics and transcriptomics are universal approaches for species analysis, which have been used in environmental communities (Boaro et al., 2014; Koo et al., 2017) and human microbiome (Ames et al., 2017; Huang et al., 2020). However, their application is limited by the fact that mRNA abundance usually provides little information about protein activity and is not a substitute for detailed functional analysis of genes (Feder and Walser, 2005), thus little is known about the mechanisms of translational control in dynamic and interacting microbial communities. For these reasons, solutions that improve the entire proteomics workflow are highly valued. With the development of mass spectrometry technology and advances in bioinformatics tools, mass spectrometry-based proteomics has become widely used for the identification, characterisation, and quantification of proteins.

Although different techniques have been used to analyse microbial communities, a quantitative understanding of the changes induced by co-culture of different species is still limited. Quantitative proteomics can provide functional insight into co-cultures, with information on both species, where unique peptides are present, a field referred to as metaproteomics. This approach has been used in many studies using samples from the environment (Sieber et al., 2015; Sedlacek et al., 2016), gut bacteria (Aakko et al., 2020), and other pathological research (Mayneris-Perxachs et al., 2022). However, proteomics is rarely used in the interaction analysis among artificial co-cultures, which could provide information on rational synthetic microbial consortia design.

In this chapter, a label-free shotgun proteomics approach with high-throughput capability was applied to the synthetic microbial community of *S. elongatus* cscB/SPS and *A. vinelandii* ΔnifL

that was constructed in Chapter 3, and a detailed proteomics analysis workflow and comprehensive proteomics analysis were established (Figure 5-1). Although these strains were successfully combined and shown to survive without organic forms of carbon and nitrogen in the medium (Smith 2016), they exhibited signs of physiological stress and ratios of each cell type varied drastically over time. A quantitative proteomics analysis could be used to characterize metabolic constraints from the perspective of both strains, making them an ideal model co-culture for this study. The protein changes of *S. elongatus* cscB/SPS and *A. vinelandii* ΔnifL were compared between the monocultures and co-culture, resulting in significant protein changes in nitrogen fixation and photosynthesis pathways. PHB synthesis-related proteins exhibited more abundance in *A. vinelandii* ΔnifL after co-culture, validating the promotion of PHB production in co-culture. In addition, proteome changes in synthetic co-culture over time were examined, which provides prospective objects for the rational optimisation of artificial microbial communities.

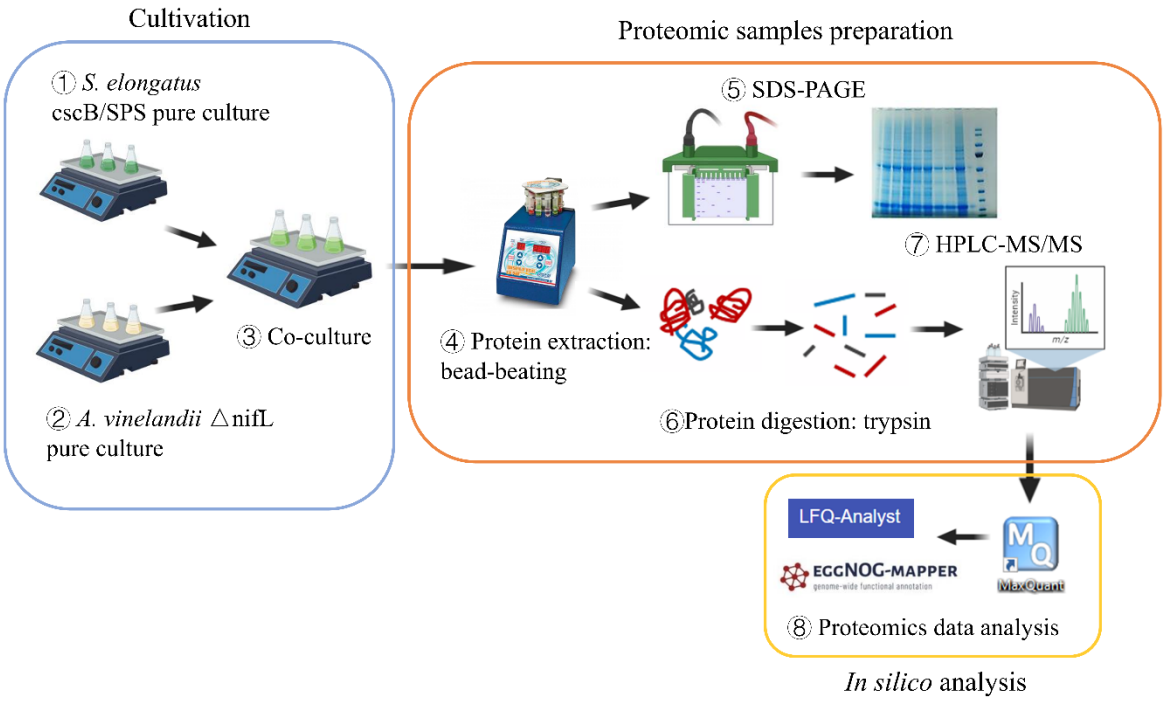


Figure 5-1 Schematic flow diagram of proteomics analysis for co-culture, including cultivation, proteomic samples preparation and *in silico* analysis

5.3 Methods

5.3.1 Co-culture setup

S. elongatus cscB/SPS was cultured in BG11 medium with 50 μg/mL kanamycin and 25 μg/mL chloramphenicol. *A. vinelandii* ΔnifL was cultured in Burk's medium. After growing to the early mid-log phase for 7 days, *S. elongatus* cscB/SPS was washed with PBS buffer and inoculated into SAC7 medium with 0.5 mM IPTG to induce sucrose production. *A. vinelandii* ΔnifL was grown to an early mid-exponential phase, washed with PBS buffer and then inoculated into IPTG-induced *S. elongatus* cscB/SPS culture with an initial ratio of 20:80 of *A. vinelandii* ΔnifL to *S. elongatus* cscB/SPS. The *S. elongatus* cscB/SPS and *A. vinelandii* ΔnifL co-culture were cultured in 500 mL flasks with 250 mL SAC7 medium as the composition shown in Chapter 3. The experiment was kept at 30 degrees under constant light intensity at 120 $\mu E \cdot m^{-2} \cdot s^{-1}$ with shaking at 120 rpm. *S. elongatus* cscB/SPS and *A. vinelandii* ΔnifL monocultures were cultivated at the same conditions as controls. Each experiment was repeated in three biological replicates. 25 mL samples were taken out from monocultures on the 4^{th} day, and 25 mL samples were taken out from co-culture on days 0, 4, 8, and 16 for proteomics analysis.

5.3.2 Proteomics sample preparation and mass spectrometry

Cell samples from each biological replicate were collected for proteomics analysis by centrifuge at 5000 rpm, 4 °C for 10 min. Protein extraction from the culture was performed using a bead-based homogenization method. Crude protein samples were purified using 2D Clean-Up Kit (GE Health) to remove excess salts, buffers and other contaminants. Protein concentration was measured using UltraBradford reagent according to the manual. In-solution digestion was conducted according to the methods reported by Hitchcock et al. (2016) to get the tryptic peptides (Hitchcock et al., 2016). Liquid chromatography-tandem mass spectrometry (LC-MS/MS) proteomic analysis was performed following the methods previously reported (Hanson et al., 2016) with modification using reverse-phase LC on a Dionex Ultimate 3000 RSLCnano system coupled online to a Q Exactive HF mass spectrometer (Thermo Scientific). Details are shown in Chapter 2, sections 2.3 and 2.4.

5.3.3 Protein identification and quantification

For protein identification of the monoculture and co-culture samples, a reference database was

created using all protein sequences of S. elongatus PCC 7942 (2874 sequences) and A. vinelandii DJ (5013 sequences) appended with CscB from E. coli and SPS from Synechocystis sp. PCC6803 from Uniprot (https://www.uniprot.org/, Feb 2020), resulting in a final database of 7889 protein sequences. Raw MS data files were processed using MaxQuant (2.0.3.0) and its built-in Andromeda search engine for peptide identification and protein inference (Cox et al., 2014). Default settings were used with search parameters set to include the following modifications: Oxidation (M) and Acetyl (Protein N-term) (variable); Carbamidomethyl (C) (fixed). Peptide-spectrum matches and protein identifications were filtered using a target-decoy approach at a false discovery rate (FDR) of 1%. Label-free quantification (LFQ) and intensity-based absolute quantification (iBAQ) options were selected (Schwanhäusser et al., 2011). Detailed parameter settings are shown in Chapter 2 section 2.5. To get confident proteins, identified proteins with unique peptides with 2 or more were used for further analysis.

5.3.4 Normalisation

To minimize the impact of cell number changes on differential expression analysis, normalization was carried out on co-culture and monocultures based on the LFQRatio method (Equation 5-1) that was developed in Chapter 4. Briefly, for each individual protein detected, its LFQ intensity was divided by the sum of all protein LFQ intensities for its respective strain.

$$\text{Normalised LFQ intensity} = \frac{\text{LFQ intensity of one protein}}{\text{total LFQ intensities of the strain}} \quad (\text{Equation } 5-1)$$

5.3.5 Differential expression analysis

Quantitative results containing protein LFQ intensities as obtained from the experiments and normalization described above, were used for the differentially expressed protein analysis using LFQ-Analyst (https://bioinformatics.erc.monash.edu/apps/LFQ-Analyst/). The normalized 'proteinGroups' table from MaxQuant and an experimental design matrix were uploaded into LFQ-Analyst for analysis. After pre-filtering to remove proteins that have been only identified by site and proteins with a high percentage of missing values, all LFQ intensities were converted to $\log_2$ scales and replicates were conditionally grouped according to the information provided in the experimental design table. Missing values were imputed using the Missing not At Random (MNAR) method in Perseus-type. Finally, a linear model of protein

combined with empirical Bayesian statistics was used for differential expression analysis. Significant proteins were filtered with an adjusted p-value cutoff of 0.05 and $\log_2$ fold change cutoff of 1. Benjamini Hochberg (BH) method was used for FDR correction.

5.3.6 Functional categories and pathway assignment

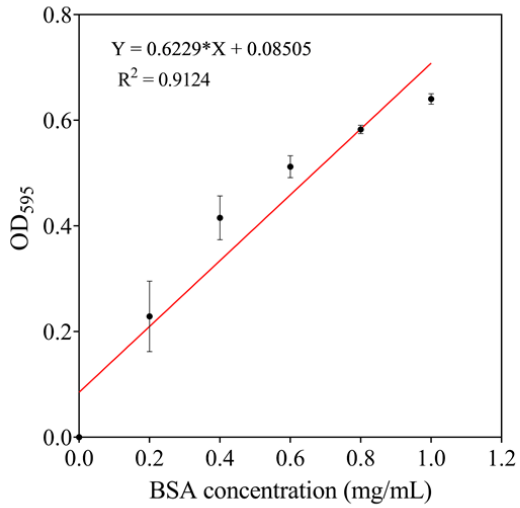
The functional categories distribution of identified proteins was analysed using Uniprot (https://www.uniprot.org/) and EggNOG-Mapper (http://eggnog-mapper.embl.de/) with default settings. Gene ontology (GO) of the differentially expressed proteins were analysed using Omicsolution (https://www.omicsolution.org/wkomics/main/) based on the GO background downloaded from Uniprot. KEGG mapper (https://www.kegg.jp/kegg/mapper/search.html) was used to assign proteins to corresponding KEGG pathways.

5.4 Results and discussion

5.4.1 Protein concentration measurement and SDS-PAGE

Protein concentrations were measured using BradfordUltra reagent (Expedeon), which is based on the reaction of coomassie with protein in an acidic medium, and the colour change from brown to blue. This coloured substance has specific absorption at 595 nm. BSA standard curve was generated with a series of BSA standard solutions of 0.2, 0.4, 0.6, 0.8 and 1 mg/mL with triplicates (Figure 5-2 A). Extracted protein concentration was calculated based on the BSA standard curve. About 30 µg of 2D cleaned-up proteins were loaded into 1D SDS-PAGE gel (Figure 5-2 B) to check protein samples.

A



B

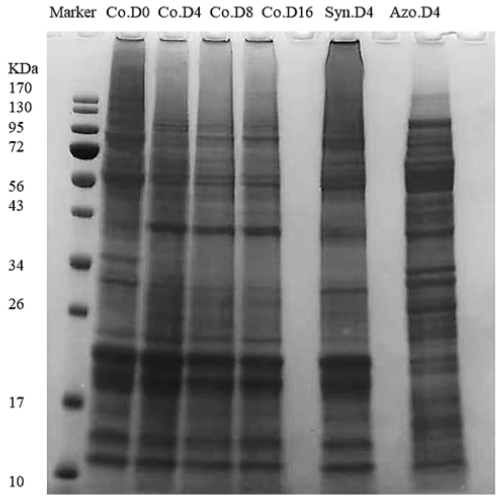


Figure 5-2 Preliminary analysis of protein samples. (A) Standard curve of BSA standard solutions. Three replicates of each point. (B) SDS-PAGE gel showing about 30 μg of 2D cleaned-up protein extracted from co-culture on day 0, 4, 8, and 16, and *S. elongatus* cscB/SPS and *A. vinelandii* △nifL monoculture control on day 4. Samples were cultured in SA7 medium. Samples were labelled as follows: 'Co.D0' for the co-culture sample on day 0, 'Co.D4' for the co-culture sample on day 4, 'Co.D8' for the co-culture sample on day 8, 'Co.D16' for co-culture sample on day 16, 'Syn.D4' for *S. elongatus* cscB/SPS monoculture control on day 4, 'Azo.D4' for *A. vinelandii* △nifL monoculture control on day 4.

5.4.2 Proteomics analysis of monoculture and co-culture on Day 4

Proteomes of the monocultures and co-culture were compared on day 4 when the cells in co-culture were in the exponential to stationary phase, and the growth curves are shown in Chapter 3 (Figure 3-6 B). 50 μg proteins of each sample were digested with trypsin followed by C18 Zip Tips cleaning and the resulting peptides were loaded into Orbitrap Q-Exactive Mass Spectrometer (Thermo Fisher Scientific). All raw data files generated from MS were processed using MaxQuant (version 2.0.3.0) against the merged database of *S. elongatus* PCC 7942, *A. vinelandii* DJ, CscB from *E. coli*, and SPS from *Synechocystis* sp. PCC 6803 downloaded from UniProt. Detailed parameters setting are shown in Chapter 2, section 2.5.

5.4.2.1 Proteomics results summary

After removing contaminants and reverse sequences, a total of 13552 tryptic peptides from *A. vinelandii* △nifL and 15003 peptides from *S. elongatus* cscB/SPS were identified across all biological and technical replicates with 1% FDR. These detected peptides corresponded to 1629 proteins of *A. vinelandii* △nifL and 1768 proteins of *S. elongatus* cscB/SPS with two or more unique peptides, covering 32.5% of the *A. vinelandii* △nifL proteome (5013 proteins) and 61.5% of the *S. elongatus* cscB/SPS proteome (2876 proteins), which is comparable to other proteomics analyses of the microorganisms (Wegener et al., 2010; Guerreiro et al., 2014; Chowdhury-Paul et al., 2018).

In addition, the consistency of identified proteins between the monocultures and co-culture was investigated. In the samples containing equal amounts of total peptides, 872 of the 1547 (56.4%) proteins identified in *A. vinelandii* △nifL were also found in coculture. Similarly, 967 of the 1767 (54.7%) proteins identified in *S. elongatus* cscB/SPS were also found in coculture (Figure 5-3). Since the monocultures and co-culture contain equal amounts of peptides, the peptides from each strain in co-culture should be roughly half that of the monocultures, therefore, it is reasonable that co-culture holds about 50% identification rate of each monoculture.

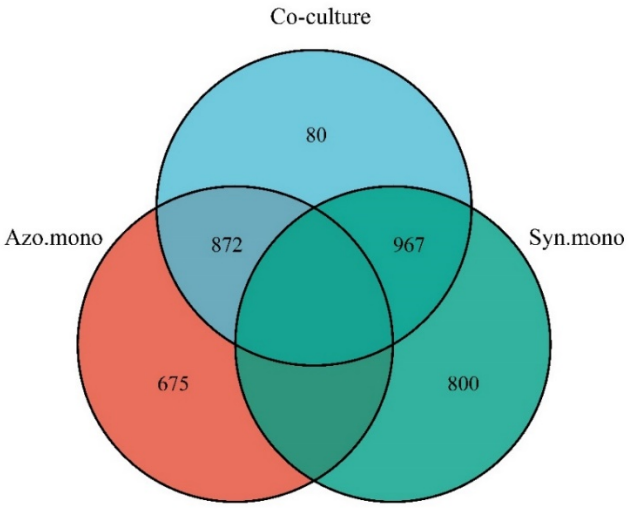


Figure 5-3 Venn figure of identified protein numbers of *A. vinelandii* △nifL and *S. elongatus* cscB/SPS in co-culture and monoculture. Samples containing a mixture of about 50 µg proteins from *A. vinelandii* △nifL and *S. elongatus* cscB/SPS (blue cycle) and samples containing about 50 µg proteins from either *A. vinelandii* △nifL alone (red cycle) or *S. elongatus* cscB/SPS

alone (green cycle) were loaded into LC-MS/MS with three replicates of each. Protein identification numbers were from MaxQuant processing.

Surprisingly, 80 proteins were identified only in co-culture (Figure 5-3). After removing the proteins with the effect of ignoring the proteins with unique peptide of only 1, 55 proteins (6 from *S. elongatus* cscB/SPS and 49 from *A. vinelandii* △nifL) mainly related to energy production, transcription, translation, coenzyme metabolism, and cell cycle control, were found only in co-culture, which were shown in Table 5-1. Although equipment detection sensitivity is a bias for the identification of these 55 proteins, they can be still regarded as higher abundant proteins after co-culture. Among of them, proteins mainly show the up-regulation of organic molecule catabolism, such as sugar and lipid, revealing an increased need of these nutrients during co-culture. The expression of nitrogenase molybdenum-iron protein (*nifD*) show the up-regulation of nitrogen fixation in *A. vinelandii* △nifL, which may be promoted by the need of its partner. The expression of molybdopterin biosynthesis protein (*moeB2* and *Avin_02710*) indicated the competition for micronutrient molybdenum in co-culture. A BLASTP search (NCBI) of uncharacterised protein Q8GJK9 showed 72% similarity to an OsmC (osmotically inducible protein C) family protein (E-value 6e-22), suggesting an oxidative stress because of the co-culture (Lesniak et al., 2003). A BLASTP search (NCBI) of uncharacterised protein C1DM30 showed 93% similarity to an FimV family protein (E-value 0), which contains a peptidoglycan binding domain and may be related to the bacterial secretion system (Wehbi et al., 2011).

Table 5-1 List of 55 proteins only identified in co-culture compared with monocultures on day 4

Grouping	Protein IDs	Description	Gene names	KEGG pathways
A. vinelandii				
Energy production	C1DGZ7	Nitrogenase molybdenum-iron protein alpha chain	*nifD*	Nitrogen metabolism; Chloroalkane and chloroalkene

				degradation
	C1DEJ8	ATP synthase subunit alpha	*atpA*	Oxidative phosphorylation
	C1DIH8	Nitroreductase	*Avin_24750*	Nitrotoluene degradation
	C1DNN9	Di-heme cytochrome c peroxidase, CCP_MauG family	*Avin_10230*	N/A
	C1DFD7	FAD linked oxidoreductase	*Avin_21430*	N/A
	C1DJB9	FAD-dependent pyridine nucleotide-disulphide oxidoreductase	*Avin_04480*	N/A
	C1DR03	Na(+)-translocating NADH-quinone reductase subunit B	*nqrB*	N/A
	C1DRV1	Thioredoxin	*Avin_36790*	N/A
Amino acid transport and metabolism	C1DRR1	Histidinol-phosphate aminotransferase	*hisC*	Tyrosine metabolism; Histidine metabolism; Novobiocin biosynthesis; Biosynthesis of amino acids; Phenylalanine, tyrosine and tryptophan biosynthesis; Phenylalanine metabolism

	C1DIC5	Shikimate dehydrogenase (NADP(+))	*aroE*	N/A
	C1DSV7	Cysteine desulfurase, sufS	*Avin_39100*	N/A
Nucleotide metabolism and transport	C1DJ86	Glucokinase	*glk-2*	Galactose metabolism; Carbon metabolism; Glycolysis/gluconeogenesis; Streptomycin biosynthesis; Amino sugar and nucleotide sugar metabolism; Biosynthesis of nucleotide sugars; Starch and sucrose metabolism
Lipid metabolism	C1DF21	2-hydroxymuconic semialdehyde hydrolase	*Avin_42230*	Degradation of aromatic compounds; Ethylbenzene degradation; Degradation of aromatic compounds
	C1DSA9	Short-chain dehydrogenase/reductase SDR	*Avin_16490*	N/A
	C1DIR4	Non-ribosomal peptide synthase: Amino acid adenylation	*Avin_25650*	N/A
Coenzyme metabolism	C1DI34	Molybdopterin biosynthesis protein, MoeB	*moeB2*	Sulfur relay system

	C1DI35	Molybdopterin biosynthesis protein	*Avin_02710*	N/A
	C1DM32	Type I fatty acid synthase ArsD	*arsD*	N/A
	C1DEL4	Oxygen-independent coproporphyrinogen III oxidas	*hemN*	Biosynthesis of cofactors; Porphyrin metabolism
	C1DJH1	Glutamate--cysteine ligase	*gshA*	Biosynthesis of cofactors; Cysteine and methionine metabolism; Glutathione metabolism
Transcription	C1DJI0	Two-component response regulator OmpR	*ompR*	Two-component system
	C1DQF1	Sigma54-dependent activator protein, FleQ	*fleQ*	Two-component system; Flagellar assembly
	C1DDM2	Response regulator, transcriptional regulatory protein (Two-component), ColR	*colR*	N/A
	C1DHP9	Transcriptional regulator, LysR family	*Avin_45170*	N/A
	C1DSY5	Bacterial regulatory protein, TetR family	*Avin_39380*	N/A

	C1DDK3	Sigma54-dependent activator protein	*Avin_17610*	N/A
Translation	C1DLS8	tRNA/tmRNA (uracil-C(5))-methyltransferase	*trmA*	N/A
	C1DSQ5	CinA-related protein	*Avin_38570*	Nicotinate and nicotinamide metabolism
	C1DRL4	ATP-dependent RNA helicase DeaD	*deaD*	RNA degradation
Cell cycle control	C1DQ76	Cell division protein ZapE	*zapE*	N/A
	P77817	Cell division protein FtsZ	*ftsZ*	N/A
	C1DQA4	Cell division protein FtsZ	*ftsZ*	N/A
	C1DQM4	Probable septum site-determining protein MinC	*minC*	N/A
Cell wall/membrane/envelop biogenesis	C1DQ93	Peptidoglycan D,D-transpeptidase FtsI	*ftsI*	Beta-Lactam resistance; Peptidoglycan biosynthesis
	C1DGI6	ErfK/YbiS/YcfS/YnhG family protein	*Avin_23090*	N/A
	C1DIW7	Chaperone SurA	*surA*	N/A
	C1DM30	Uncharacterized protein	*Avin_29500*	N/A
Cell motility	C1DQ	Uncharacterized	*Avin_34230*	N/A

	D6	protein		
Replication and repair	C1DPL6	ATP-dependent DNA helicase RecQ	*recQ*	RNA degradation
	C1DEM9	ComEA-related protein	*Avin_20070*	N/A
	C1DMI6	Superfamily I DNA and RNA helicases and helicase subunits-related protein	*Avin_51880*	N/A
Inorganic ion transport and metabolism	C1DKE0	3(2),5-bisphosphate nucleotidase CysQ	*cysQ*	Sulfur metabolism
Post-translational modification, protein turnover, chaperone functions	C1DJN9	Alkyl hydroperoxide reductase subunit F	*Avin_26320*	N/A
	C1DGR3	Aspartyl/Asparaginyl beta-hydroxylase family protein	*Avin_44410*	N/A
	C1DH25	Peptidylprolyl isomerase	*nifM*	N/A
Others	C1DSZ1	YaeQ family protein	*yaeQ*	N/A
	C1DQI5	Uncharacterized protein	*Avin_34720*	N/A
	C1DDR5	Uncharacterized protein	*Avin_18230*	N/A
	C1DQZ1	Uncharacterized protein	*Avin_14480*	N/A
S. elongatus				

Post-translational modification, protein turnover, chaperone functions	Q8GMT3	Peptidase, metallopeptidases	*sek0003*	N/A
Secondary Structure	Q31P61	Uncharacterized protein	*Synpcc7942_1128*	N/A
Others	Q31LR7	TPR repeat	*Synpcc7942_1972*	N/A
	Q8GJK9	Uncharacterized protein	*Synpcc7942_1818*	N/A
	Q31M71	Uncharacterized protein	*Synpcc7942_1818*	N/A
	Q31NF4	Uncharacterized protein	*Synpcc7942_1385*	N/A

Grouping was determined using functional categories through EGGNOG-Mapper.

The functional categories distribution of identified proteins was analysed using Uniprot (https://www.uniprot.org/) and EggNOG-Mapper (http://eggnog-mapper.embl.de/), resulting in 19 distinct functional categories of *S. elongatus* cscB/SPS and 18 distinct functional categories of *A. vinelandii* △nifL. The grouping of identified proteins provides an overview of protein functional distribution of the two strains in co-culture and monocultures (Figure 5-4). The top three abundant functional categories in *A. vinelandii* △nifL monoculture were assigned to energy production and conversion (11.4%), amino acid transport and metabolism (9.3%), and translation (7.8%), while the top three abundant in *S. elongatus* cscB/SPS monoculture were energy production and conversion (9.5%), translation (8.1%), and coenzyme transport and metabolism (6.7%). The predominant protein groupings in co-culture were the same as that in monoculture, apart from that more proteins assigned to translation (11%) than amino acid transport and metabolism (9.3%) in *A. vinelandii* △nifL in co-culture.

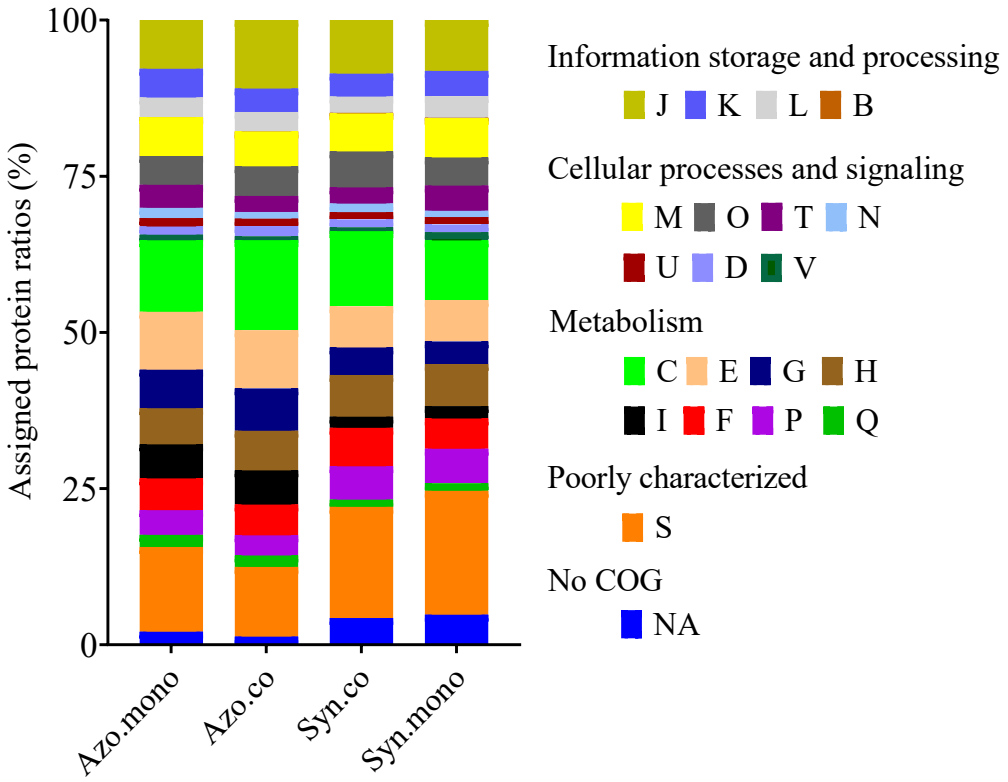


Figure 5-4 Comparison of the distributions of the identified proteins across samples by their functional categories in Clusters of Orthologous Genes (COGs). COG categories are as follows: J (Translation, ribosomal structure and biogenesis), K (Transcription), L (Replication, recombination and repair), B (Chromatin structure and dynamics), M (Cell wall/membrane/envelope biogenesis), O (Posttranslational modification, protein turnover, chaperones), T (Signal transduction mechanisms), N (Cell motility), U (Intracellular trafficking, secretion, and vesicular transport), D (Cell cycle control, cell division, chromosome partitioning), V (Defense mechanisms), C (Energy production and conversion), E (Amino acid transport and metabolism), G (Carbohydrate transport and metabolism), H (Coenzyme transport and metabolism), I (Lipid transport and metabolism), F (Nucleotide transport and metabolism), P (Inorganic ion transport and metabolism), Q (Secondary metabolites biosynthesis, transport and catabolism), S (Function unknow), and NA (No COG assigned). Portions in columns were ranked by first class of COGs (i.e., Information storage and processing, Cellular processes and signaling, metabolism, and Poorly characterized) and by percentage from high to low of *A. vinelandii* △nifL monoculture. Azo.mono: *A. vinelandii* △nifL monoculture; Azo.co: *A. vinelandii* △nifL in co-culture; Syn.co: *S. elongatus* cscB/SPS in co-culture; Syn.mono: *S. elongatus* cscB/SPS monoculture.

5.4.2.2 Comparison of *A. vinelandii* ΔnifL proteome in monoculture and co-culture conditions

The proteome of *A. vinelandii* △nifL in monoculture and in co-culture with *S. elongatus* cscB/SPS was compared on day 4 when the cells in co-culture were in exponential to stationary phase, and the growth curves are shown in Chapter 3 (Figure 3-6 B). Extracted proteins from *A. vinelandii* △nifL were identified and quantified, and totals of 1547 and 934 proteins were detected in monoculture control and co-culture, respectively (Figure 5-5 A). Among them, 54.2% of the proteins (872) were identified in both conditions, while 675 and 62 proteins were uniquely present in control and co-culture, respectively (Figure 5-5 A).

Differentially expressed proteins (DEPs) were analysed using LFQ-Analyst (https://bioinformatics.erc.monash.edu/apps/LFQ-Analyst/). A comparison of the global proteome profiles between co-culture and monoculture control based on the LFQ intensity from MaxQuant revealed an obvious separation between the two groups (Figure 5-5 B). The heatmap delivered an overview of all DEPs across all samples, suggesting that co-culture with *S. elongatus* cscB/SPS changed the expression of some proteins associated with the biological activities and cellular process of *A. vinelandii* △nifL (Figure 5-5 C). Among the identified proteins, 213 proteins (13.08 %) were significantly up-regulated with a $\log_2$ fold change of 1 (p.adj < 0.05), while 321 proteins (19.71 %) were significantly down-regulated with a $\log_2$ fold change of 1 (p.adj < 0.05) after co-culture with *S. elongatus* cscB/SPS (Figure 5-5 D). These significant proteins were assigned to 78 up-regulated KEGG pathways and 86 down-regulated KEGG pathways in *A. vinelandii*.

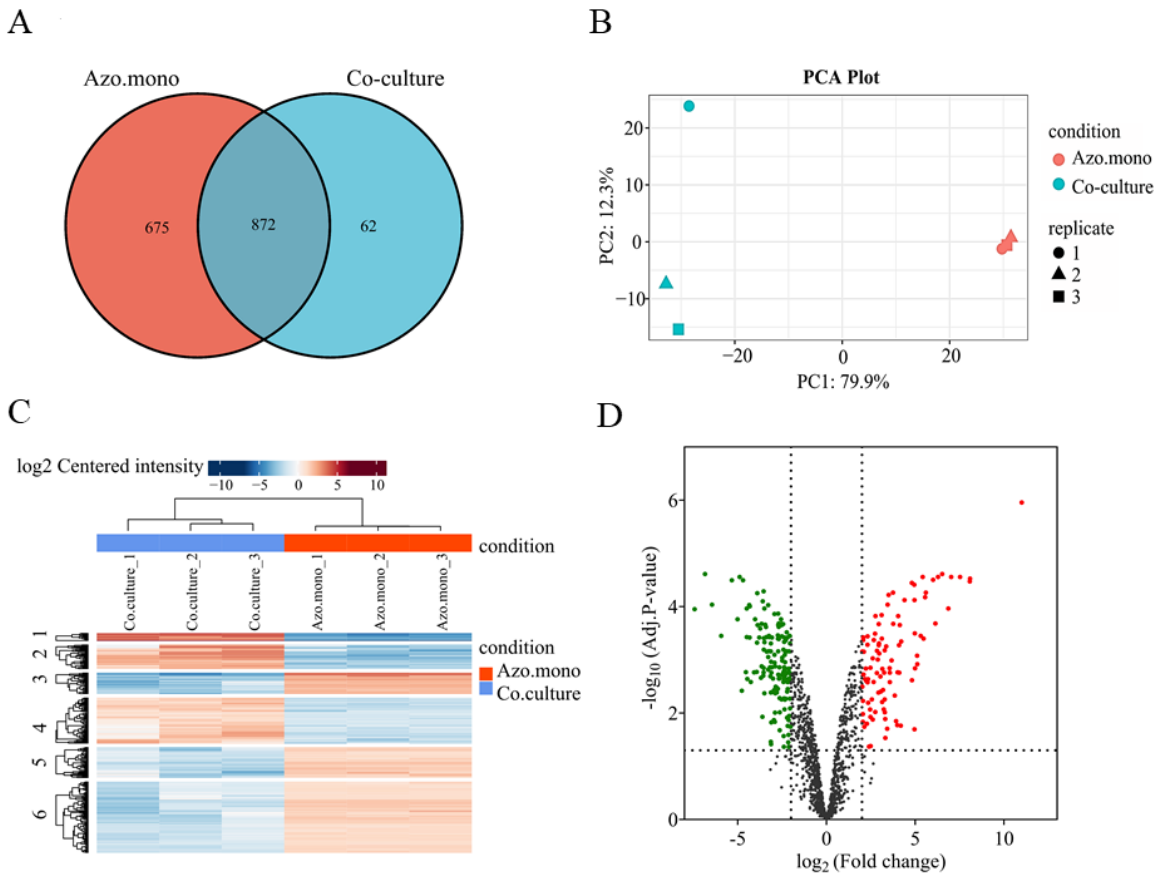


Figure 5-5 Comparison of *A. vinelandii* △nifL proteome in monoculture and co-culture conditions. (A) Venn diagram of unique and overlapped proteins in monoculture (red) and co-culture (blue). (B) Principal Component Analysis (PCA) plot of individual samples showed distinct clustering of each condition. Each sample was in three replicates. (C) Heatmap provided an overview of all DEPs (rows) across all samples (Columns). (D) Volcano plot of significantly up-regulated (red, 99) and down-regulated (green, 155) proteins after co-culture. The dotted lines represented $\log_2$ fold change cutoff of 1 and adjusted *p-value* cutoff of 0.05.

DEPs analysis showed that the greatest up-regulated protein observed in *A. vinelandii* △nifL after co-culture was for nitrogenase protein alpha chain (*nifD*; Table 5-2), which had a $\log_2$ fold change of 11 in co-culture relative to monoculture. Similarly, Nitrogenase molybdenum-iron protein beta chain (*nifK*) and nitrogenase iron protein (*nifH*) exhibited co-culture/monoculture $\log_2$ fold change of 7.53 and 4.13, respectively. The nitrogenase protein is the key protein in *A. vinelandii* for nitrogen fixation, which contains two components, component I (Fe-Mo protein) encoded by *nifD* and *nifK* gene and component II (Fe protein)

encoded by *nifH* gene (Hamilton et al., 2011). Correspondingly, Flavodoxin 1, an electron donor to nitrogenase (Gangeswaran and Eady, 1996), also showed higher abundance with log_2 fold change of 4.09 after co-culture. The higher abundance of all these three proteins indicates the increase of ammonia synthesis after co-culture with *S. elongatus* cscB/SPS (KEGG pathway shown in Appendix Figure D1). This is because, in co-culture, *S. elongatus* cscB/SPS consumes ammonia and causes less ammonia concentration in the medium compared to monoculture, which promotes ammonia production from *A. vinelandii* △nifL. A protein related to metabolizing sucrose to a sole carbon source, sucrose-6-phosphate hydrolase (log_2 FC 3.75), was also up-regulated (KEGG pathway shown in Appendix Figure D1), which verified the increased demand for carbon metabolism in *A. vinelandii* △nifL. Interestingly, three cell division-related proteins, cell division protein ZapE (log_2 FC 5.56), cell division protein FtsZ (log_2 FC 4.81), and cell division topological specificity factor (log_2 FC 3.57) showed higher abundance after co-culture, suggesting that co-culture promoted *A. vinelandii* △nifL cell division. However, it did not show higher growth rate compared with the monoculture control (1.651 ± 0.8548 d^{-1} in monoculture and 0.222 ± 0.0826 d^{-1} in co-culture), reasons behind this requires further study.

Table 5-2 Top 50 significantly higher relative abundance proteins of *A. vinelandii* △nifL in co-culture compared to monoculture.

Grouping	Protein IDs	Description	log_2 FC	Adj. *p*-value
Nitrogen fixation	C1DGZ7	Nitrogenase protein alpha chain	11	1.10E-06
	C1DGZ8	Nitrogenase molybdenum-iron protein beta chain	7.53	2.76E-05
	C1DGZ6	Nitrogenase iron protein	4.13	2.21E-04
Energy production and	C1DIH8	Nitroreductase	6.52	2.43E-05
	C1DNN9	Di-heme cytochrome c peroxidase, CCP_MauG family	6.29	2.76E-05

conversion	C1DS49	NADH:flavin xenobiotic reductase	4.98	3.17E-04
	P52964	Flavodoxin 1	4.09	5.28E-04
	C1DGT4	FAD dependent oxidoreductase	3.31	1.29E-02
Cell cycle control and mitosis	C1DQ76	Cell division protein ZapE	5.56	6.63E-05
	P77817	Cell division protein FtsZ	4.81	3.57E-05
	C1DQM6	Cell division topological specificity factor	3.57	1.74E-03
Amino acid transport and metabolism	C1DEE5	Lactoylglutathione lyase	5.31	3.54E-04
	C1DI85	Flavin-containing mono amine oxidase	4.01	1.71E-03
Nucleotide metabolism and transport	C1DRU5	Adenylyl-sulfate kinase	4.82	2.41E-03
	C1DJ86	Glucokinase	4.4	7.50E-05
	C1DI76	Pyrimidine biosynthesis regulatory protein, PyrR	4.1	1.74E-03
Carbohydrate metabolism and transport	C1DMH9	Sucrose-6-phosphate hydrolase	3.75	5.45E-05
	C1DKP7	KDPG aldolase	3.5	6.01E-05
Coenzyme metabolism	C1DM32	Type I fatty acid synthase ArsD	6.87	1.08E-04
	C1DI34	Molybdopterin biosynthesis protein, MoeB	4.95	3.87E-05
	C1DI35	Molybdopterin biosynthesis protein	3.85	4.67E-03
	C1DJH1	Glutamate--cysteine ligase	3.29	9.67E-03
	C1DSA9	Short-chain dehydrogenase/reductase SDR	4.96	2.01E-02
Transcription	C1DRL4	ATP-dependent RNA helicase DeaD	5.12	1.20E-03
	C1DEL3	Fnr-like negative transcriptional	4.99	1.44E-03

		regulator of CydAB		
	C1DKZ9	Cold shock protein, CspD	4.21	4.51E-04
	C1DSY5	Bacterial regulatory protein, TetR family	4.04	1.50E-04
	C1DJI0	Two-component response regulator OmpR	3.67	1.02E-03
Replication and repair	C1DEM9	ComEA-related protein	6.14	2.07E-04
Cell wall/membrane/envelop biogenesis	C1DGI6	ErfK/YbiS/YcfS/YnhG family protein	3.92	1.41E-02
	C1DRS2	dTDP-4-dehydrorhamnose 3,5-epimerase	3.57	2.61E-03
Post-translational modification, protein turnover, chaperone functions	C1DGR3	Aspartyl/Asparaginyl beta-hydroxylase family protein	5.6	5.45E-05
	C1DJN9	FAD-dependent pyridine nucleotide-disulphide oxidoreductase	5.42	2.76E-05
	C1DH25	Peptidylprolyl isomerase	3.32	2.93E-02
Inorganic ion transport and metabolism	C1DJC0	Carbon monoxide dehydrogenase	8.08	3.35E-05
	C1DPF4	Integral membrane protein, TerC family	3.76	2.10E-04
Secondary Structure	C1DM34	Type III polyketide synthase	5.11	8.33E-04
	C1DM35	Type I fatty acid synthase ArsA	3.37	5.56E-04
Other	C1DQI5	Uncharacterized protein	8.09	2.96E-05
	C1DH01	Uncharacterized protein	5.47	3.98E-04
	C1DQZ1	Uncharacterized protein	3.42	1.96E-02

C1DDQ1	Uncharacterized protein	7.03	2.76E-05
C1DQD6	Uncharacterized protein	3.56	1.43E-03
C1DHG0	Phasin protein	6.01	3.13E-05
C1DGT6	Uncharacterized protein	4.19	1.72E-02
C1DRS6	Hydrolase (HAD superfamily) protein	3.99	1.68E-02
C1DIF2	Esterase, poly(3-hydroxybutyrate) depolymerase	3.45	9.16E-05

Many proteins related to energy production and conversion show lower abundance in *A. vinelandii* △nifL after co-culture (Table 5-3). Key proteins in oxidative phosphorylation, including NADH-ubiquinone oxidoreductase subunit E (log_2 FC -3.93), NADH-quinone oxidoreductase subunit A (log_2 FC -3.28), succinate dehydrogenase hydrophobic membrane anchor subunit (log_2 FC -4.49), cytochrome b (log_2 FC -3.52), cytochrome c oxidase (log_2 FC -3.28), and ATP synthase subunit A (log_2 FC -3.79), corresponding to the components of protein complexes I, II, III, VI and V (Matlin, 2016; Kalnenieks et al., 2019; Moosavi et al., 2019), exhibited lower abundance after co-culture, revealing a down-regulation of oxidative phosphorylation pathway (Appendix Figure D1).

Apart from the top 50 proteins listed in the table, many proteins related to ribosome function showed lower abundance (KEGG pathway shown in Appendix D1). One cold shock protein, which has been shown to contribute to pH, osmotic, starvation, oxidative, and ethanol stress tolerance as well as to host cell invasion (Keto-Timonen et al., 2016), exhibited higher abundance (log_2 FC 4.21) after co-culture. This indicates the stress to *A. vinelandii* △nifL resulting from the co-culture.

Table 5-3 Top 50 significantly lower relative abundance proteins of *A. vinelandii* △nifL in co-culture compared to monoculture.

Grouping	Protein IDs	Description	log_2 FC	Adj. *p*-value
Energy production and conversion	C1DL12	Isocitrate lyase	-5.93	3.54E-04
	C1DIE4	Formate dehydrogenase, alpha subunit	-4.89	2.76E-05
	C1DEX4	Acetoin:2,6-dichlorophenolindophenol oxidoreductase alpha subunit, AcoA	-4.49	1.08E-04
	C1DM59	Succinate dehydrogenase hydrophobic membrane anchor subunit; sdhD	-4.49	3.73E-04
	C1DEX7	2,3-butanediol dehydrogenase	-4.33	3.80E-04
	C1DL17	NADH-ubiquinone oxidoreductase subunit E; nuoE	-3.93	1.68E-03
	C1DNU2	Pyruvate dehydrogenase E1 component subunit alpha	-3.91	4.74E-04
	C1DP78	Dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase complex	-3.88	2.21E-04
	C1DND9	ATP synthase subunit a; atpB	-3.79	3.79E-04
	C1DML5	Aldehyde dehydrogenase (NAD+)	-3.65	1.32E-03
	C1DNT7	Methylmalonate-semialdehyde dehydrogenase	-3.61	1.17E-02
	C1DQ81	Cytochrome b; petB	-3.52	3.90E-04

	C1DKF8	Cytochrome c Catalase; cccA	-3.5	4.23E-03
	C1DEX5	Acetoin:2,6-dichlorophenolindophenol oxidoreductase beta subunit AcoB	-3.49	4.66E-04
	C1DNU3	TPP-dependent dehydrogenase, E1 component beta subunit	-3.41	2.41E-04
	C1DEM3	Cytochrome c oxidase; ccoN	-3.28	8.81E-04
	C1DL14	NADH-quinone oxidoreductase subunit A; nuoA	-3.28	9.33E-05
Amino acid transport and metabolism	C1DQU2	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase	-4.54	1.70E-03
	C1DE68	Cysteine desulfurase IscS	-4.35	9.27E-05
	C1DII7	Zinc-containing alcohol dehydrogenase superfamily	-3.73	2.57E-04
	C1DNK4	Urease subunit alpha	-3.69	2.17E-03
	C1DJ12	Glycine dehydrogenase (decarboxylating)	-3.54	5.13E-05
	C1DDX7	NAD-dependent glutamate dehydrogenase	-3.24	3.54E-04
Nucleotide metabolism and transport	C1DGA8	Xanthine dehydrogenase, molybdopterin binding subunit	-4.02	1.72E-04
Carbohydrate metabolism and transport	C1DLL4	Aquaporin Z	-5.33	3.18E-05
	C1DRP5	Phosphogluconate dehydratase	-3.65	2.10E-04
	C1DHU4	Alpha-1,4 glucan phosphorylase	-3.59	7.92E-05

Coenzyme metabolism	C1DHY3	3-demethoxyubiquinol 3-hydroxylase	-3.9	5.58E-05
	C1DM80	Mo processing, homeostasis	-3.22	3.72E-04
Lipid metabolism	C1DFK2	Acetyl-coenzyme A synthetase	-4.08	1.71E-03
	C1DNT6	Enoyl-CoA hydratase/isomerase	-3.54	1.41E-03
	C1DEX2	Short-chain dehydrogenase/reductase SDR	-3.54	1.71E-03
Translation	C1DQ52	Sigma 54 modulation protein/ribosomal protein S30EA	-3.18	1.38E-03
Transcription	C1DI45	DNA-directed RNA polymerase subunit omega	-3.28	1.52E-03
Cell wall/membrane/envelop biogenesis	C1DLG2	OmpW family outer membrane porin	-3.38	3.73E-04
Post-translational modification, protein turnover, chaperone functions	C1DIS7	60 kDa chaperonin	-7.43	1.11E-04
	C1DEN7	Heat shock protein Hsp20	-5.01	1.72E-04
	C1DSQ8	Heat shock protein Hsp20	-4.77	3.79E-03
	C1DSQ9	Heat shock protein Hsp20	-3.25	6.86E-04
Inorganic ion transport and metabolism	C1DN36	Cytochrome c-type biogenesis protein	-4.26	2.60E-03
Secondary Structure	C1DF30	Isochorismatase hydrolase	-3.45	2.18E-04
Signal Transduction	C1DFW5	DUF883_C domain-containing protein	-3.53	6.43E-04
Intracellular	C1DHQ7	Bacterial secretion system protein	-4.32	9.83E-05

trafficing and secretion				
Others	C1DI92	Uncharacterized protein	-6.45	9.16E-05
	C1DGE8	Uncharacterized protein	-4.7	3.18E-05
	C1DJ09	PepSY domain-containing protein	-6.84	2.43E-05
	C1DML1	Zinc-containing alcohol dehydrogenase superfamily	-4.46	2.27E-03
	C1DJP2	Aldo/keto reductase protein	-3.5	1.08E-04

Functional analysis and enrichment were carried out using Omicsolution tool (https://www.omicsolution.org/wkomics/main/), which searched for Gene ontology (GO) terms within the DEPs based on the GO background downloaded from the UniProt. Figure 5-6 exhibits the most enriched GO terms in proteins with significantly higher or lower relative abundance ($p < 0.05$) of *A. vinelandii* in co-culture compared to the *A. vinelandii* monoculture. Many proteins involved in metabolic processes of organic and inorganic compounds showed higher relative abundance in *A. vinelandii* after co-cultured with *S. elongatus*, including phosphorus metabolic process, phosphate-containing compound metabolic process, and glycolytic process, indicating more demand for these nutrients, such as carbon, phosphate, etc (Figure 5-6 A). Proteins related to organonitrogen compound biosynthetic process showed higher abundance after co-culture, suggesting the increase of nitrogen assimilation ability of A. vinelandii as expected (Figure 5-6 A). Interestingly, the proteins in the GO term 'poly-hydroxybutyrate biosynthetic process' were enriched, revealing a promotion of PHB production in A. vinelandii when co-cultured with S. elongatus (Appendix Table D3). Whereas biological process GO terms enriched within the lower abundance proteins in A. vinelandii after co-culture included cellular respiration, aerobic respiration, and tricarboxylic acid cycle, indicating possible oxygen limitation in co-culture, which may be due to a reduction in oxygen solubilization because of the higher cell density (Figure 5-6 B).

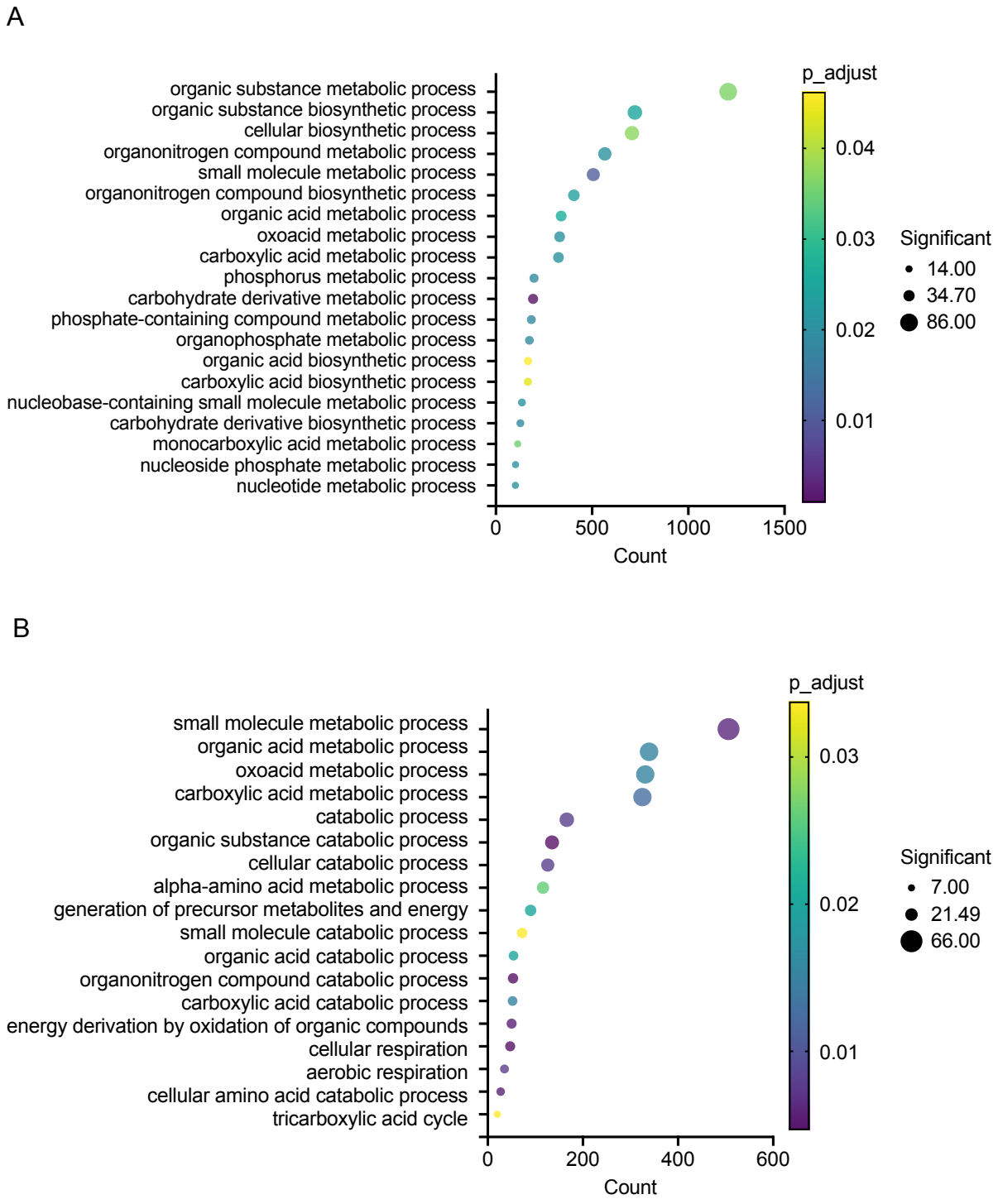


Figure 5-6 Enriched biological process GO terms of differentially expressed proteins in *A. vinelandii* △nifL in co-culture compared to monoculture. (A) Top 20 GO terms in proteins with significantly higher relative abundance ($p < 0.05$) in *A. vinelandii* △nifL after coculture. (B) GO terms in proteins with significantly lower relative abundance ($p < 0.05$) in *A. vinelandii* △nifL after coculture. Significantly DEPs with a $\log_2$ fold change cutoff of 1. Adjusted p-value

for GO enrichment is 0.05. Figures were created using GraphPad Prism.

5.4.2.3 Comparison of *S. elongatus* cscB/SPS proteome in monoculture and co-culture conditions

The difference between the proteome of *S. elongatus* cscB/SPS in monoculture and in co-culture with *A. vinelandii* △nifL was analysed. Extracted proteins from *S. elongatus* cscB/SPS were identified and quantified, and totals of 1767 and 985 proteins were detected in monoculture control and co-culture, respectively (Figure 5-7 A). Among them, 54.2% of the proteins (967) were identified in both conditions, while 800 and 18 proteins were uniquely present in control and co-culture, respectively (Figure 5-7 A).

A comparison of the global proteome profiles between co-culture and monoculture control based on the LFQ intensity from MaxQuant showed an obvious separation between the two groups (Figure 5-7 B). The heatmap delivered an overview of all DEPs across all samples, suggesting that co-culture with *A. vinelandii* △nifL affected the cellular process and biological activities of *S. elongatus* cscB/SPS (Figure 5-7 C). Among the identified proteins, 66 proteins (3.73 %) were significantly up-regulated with a $\log_2$ fold change of 1 (p.adj < 0.05), while 451 proteins (25.51 %) were significantly down-regulated with a $\log_2$ fold change of 1 (p.adj < 0.05) after co-culture with *A. vinelandii* △nifL (Figure 5-7 D). These significant proteins were assigned to 15 up-regulated KEGG pathways and 88 down-regulated KEGG pathways in *S. elongatus*.

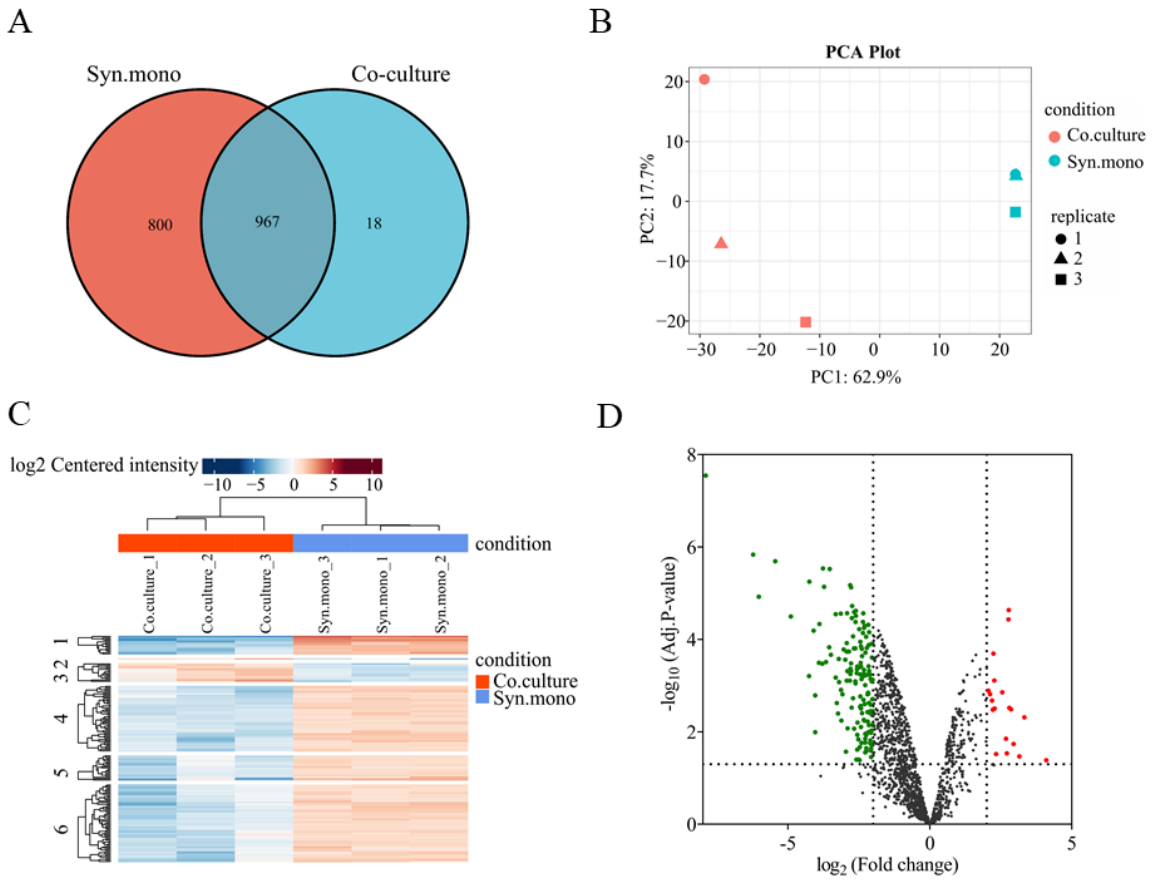


Figure 5-7 Comparison of *S. elongatus* cscB/SPS proteome in monoculture and co-culture conditions. (A) Venn diagram of unique and overlapped proteins in monoculture (red) and co-culture (blue). (B) Principal Component Analysis (PCA) plot of individual samples showed distinct clustering of each condition. Each sample was in three replicates. (C) Heatmap provided an overview of all DEPs (rows) across all samples (Columns). (D) volcano plot of significantly up-regulated (red, 19) and down-regulated (green, 151) proteins after co-culture. The dotted lines represented a $\log_2$ fold change cutoff of 1 and an adjusted *p-value* cutoff of 0.05.

DEPs analysis of *S. elongatus* cscB/SPS in co-culture compared to monoculture showed many proteins related to photosynthesis in higher abundance, for example, photosystem components proteins, including photosystem I iron-sulfur center ($\log_2$ FC 1.16), photosystem I reaction center subunit IV ($\log_2$ FC 1.14), photosystem II reaction center Psb28 protein ($\log_2$ FC 1.18), and phycobilisome maturation protein ($\log_2$ FC 1.49), as well as electron carrier proteins participating in electron transfer between the cytochrome b6-f complex and P700 PSI,

including plastocyanin (log_2 FC 2.77) and cytochrome c6 (log_2 FC 2.12) (Table 5-4). Whereas the component protein of cytochrome b6-f complex, cytochrome b6-f complex subunit 4 (log_2 FC -4.04), and proteins related to the reaction center of PSII, cytochrome b559 subunit beta (log_2FC -2.8), were down-regulated in *S. elongatus* cscB/SPS after co-culture (Table 5-5) (KEGG pathway shown in Appendix D2). This may result from the harmful reactive oxygen species (ROS) (Pospíšil, 2016; Shimakawa and Miyake, 2018). And some proteins related to carbon fixation and export, e.g., sucrose phosphate synthase (log_2 FC -1.5), and sucrose permease (log_2 FC -1.04) also showed lower abundant (KEGG pathway shown in Appendix D2), which is not desired. This may be because the ammonium produced by its partner is not sufficient for the growth and production of *S. elongatus* cscB/SPS.

Table 5-4 Top 50 significantly higher relative abundance proteins of *S. elongatus* cscB/SPS in co-culture compared to monoculture.

Grouping	Protein IDs	Description	log_2 FC	Adj. *p*-value
Energy production and conversion	P55020	Plastocyanin; petE	2.77	3.69E-05
	P25935	Cytochrome c6; petJ	2.12	1.53E-03
	Q31MF5	Ferredoxin-thioredoxin reductase, catalytic chain; ftrC	1.98	0.0428
	Q31RP8	Cytochrome c6; petJ.1	1.97	0.0113
	Q31KK2	Cytochrome c domain-containing protein	1.81	0.0218
	Q9Z3G5	PBS lyase HEAT-like repeat	1.78	0.018
	Q31M	Probable amidase	1.64	0.0197

	Z1			
	P18655	Superoxide dismutase [Fe]; sodB	1.5	0.00507
	Q44116	Phycobilisome maturation protein; cpcF	1.49	0.0215
	Q31Q35	Allophycocyanin alpha chain-like	1.35	0.0224
	Q31QV2	Photosystem I iron-sulfur center; psaC	1.16	0.0163
Amino acid transport and metabolism	Q31QR5	VOC domain-containing protein	2.8	0.0148
	Q31ND8	VOC domain-containing protein	1.31	0.00623
Nucleotide metabolism and transport	Q31PQ0	Adenylyl-sulfate kinase	1.2	0.0453
	Q31QM9	Acireductone dioxygenase	1.2	0.00697
Coenzyme metabolism	Q31PM2	Porphobilinogen deaminase	1.38	0.0246
	Q31SA7	ATP-dependent dethiobiotin synthetase BioD	1.13	0.0468
Translation	Q31KR7	Aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit C	1.36	0.0315
	Q31KN7	Ribosome hibernation promoting factor	1.35	0.0339
Cell wall/membrane/ envelop	Q31MT3	Beta-Ig-H3/fasciclin	2.55	0.00936
	Q8KU	ANL06	1.48	0.0389

biogenesis	W7			
Post-translational modification	Q31KR6	Glutaredoxin	1.39	0.0386
Inorganic ion transport and metabolism	Q8KUU6	ANL35	2.27	0.00686
	Q31MC5	Uncharacterized protein	2	0.00552
	Q31QM2	Rhodanese-like	1.18	0.0209
Signal Transduction	Q31RU2	Uncharacterized protein	2.19	0.0113
	Q79PF9	Response regulator receiver domain protein (CheY-like)	1.77	0.0056
Others	Q31P39	DUF305 domain-containing protein	3.33	0.0196
	Q31NF4	Uncharacterized protein	2.95	0.0461
	Q31KY3	DUF1400 domain-containing protein	2.86	0.0151
	Q31KM3	Nif11 domain-containing protein	2.28	0.0148
	Q31QW3	Uncharacterized protein	2.06	0.00888
	Q31R69	Uncharacterized protein	1.89	0.00969
	Q31QW8	Uncharacterized protein	1.75	0.00623

	Q31MM8	Uncharacterized protein	1.62	0.0045
	Q31MU1	Phenazine biosynthesis PhzC/PhzF protein	1.6	0.0181
	Q31LZ0	Cupin_5 domain-containing protein	1.53	0.0337
	Q31L92	PINc domain-containing protein	1.43	0.0415
	Q31KK9	Uncharacterized protein	1.18	0.00887
	Q31ML0	Photosystem II reaction center Psb28 protein	1.18	0.00686
	Q31NL7	Photosystem I reaction center subunit IV	1.14	0.0296

Table 5-5 Top 50 significantly lower relative abundance proteins of *S. elongatus* cscB/SPS in co-culture compared to monoculture.

Grouping	Protein IDs	Description	log_2 FC	Adj. *p*-value
Energy production and conversion	P15347	Iron stress-induced chlorophyll-binding protein; isiA	-6.23	1.46E-06
	Q54710	Cytochrome b6-f complex subunit 4; petD	-4.04	1.02E-02
	Q31RT9	Type 2 NADH dehydrogenase	-2.93	1.20E-04
	Q8VPV7	CO_2 hydration protein; chpY	-2.81	6.69E-06
	Q8KPP2	Cytochrome b559 subunit beta; psbF	-2.8	4.29E-04
	Q31QL4	Thioredoxin reductase	-2.76	3.00E-03

	Q31LK5	Glyco_transf_28 domain-containing protein	-2.9	8.43E-03
	Q31N12	Nitrogen assimilation regulatory protein	-3.73	7.27E-06
	Q8VPV9	NAD(P)H dehydrogenase, subunit NdhF3 family	-2.96	2.70E-02
Amino acid transport and metabolism	Q31NC9	2-isopropylmalate synthase	-3.28	2.38E-03
	Q8KPV4	Dehydrogenase subunit-like protein	-2.78	3.30E-04
	Q31KY1	Aminotransferase	-2.75	3.46E-05
Nucleotide metabolism and transport	Q31Q19	Phosphoribosylformylglycinamidine synthase subunit PurS	-5.45	2.04E-06
	Q8GAA0	1-deoxy-D-xylulose-5-phosphate synthase	-3.77	2.91E-06
Coenzyme metabolism	Q31RS6	Cobyric acid synthase	-6.03	1.19E-05
	Q31PF2	Biosynthetic arginine decarboxylase	-3.14	2.77E-05
	Q31PY6	LL-diaminopimelate aminotransferase	-3.08	1.54E-03
	Q31QL9	Adenosylhomocysteinase	-2.98	2.75E-04
Lipid metabolism	Q31K57	AB hydrolase-1 domain-containing protein	-3.79	3.37E-04
Translation	Q31N05	Piwi domain-containing protein	-4.9	3.19E-05
	P63200	30S ribosomal protein S12	-4.25	5.62E-06
	Q935Y0	Bacterial translation initiation factor 1 (BIF-1)	-3.91	3.21E-04
	Q31LW9	Leucine--tRNA ligase	-3.32	2.84E-05
	Q31M43	Ribonuclease J	-3.32	6.70E-04

	Q31KD2	Glycine--tRNA ligase alpha subunit	-3.2	8.39E-04
	Q31PL3	CBS	-2.82	7.81E-04
	Q31KP9	GTPase Der	-2.92	2.71E-05
	Q31RX7	GTPase Era	-2.71	8.60E-05
Transcription	Q31LN5	Sigma-24 (FecI-like)	-2.93	1.84E-04
Replication and repair	Q31NS6	N6-adenine-specific DNA methylase-like	-2.74	1.89E-05
Cell wall/membrane/envelop biogenesis	Q31N76	Probable porin	-4.04	1.63E-03
	Q31NW5	ABC-transporter membrane fusion protein	-2.85	1.83E-04
	Q31MT1	Mannose-1-phosphate guanylyltransferase (GDP)	-2.78	7.48E-06
Post-translational modification, protein turnover, chaperone functions	P53533	Chaperone protein ClpB 1	-4.26	6.22E-04
	Q93AK0	Cell division protein Ftn2	-3.12	5.72E-03
	Q31PJ1	ATP-dependent zinc metalloprotease FtsH	-2.91	4.86E-04
Signal Transduction	Q31M30	Diguanylate cyclase/phosphodiesterase with PAS/PAC sensor(S)	-4.1	6.48E-05
Intracellular trafficing and secretion	Q31KG0	Biopolymer transport ExbB like protein	-3.21	2.65E-04
Defense mechanisms	Q31KM0	Hydrophobe/amphiphile efflux-1 HAE1	-3.04	1.65E-04
Others	Q31MV9	Uncharacterized protein	-3.89	4.65E-05

Q31QT9	Uncharacterized protein	-3.66	3.09E-04
Q31RP4	Possible high light inducible polypeptide HliC	-7.9	2.85E-08
Q31N64	Sodium-dependent bicarbonate transporter	-3.55	1.48E-04
Q55237	ORF4 (Fragment)	-3.49	2.15E-04
Q31PW6	DUF4062 domain-containing protein	-3.24	4.01E-03
Q31MA6	UPF0182 protein Synpcc7942_1783	-3.13	1.24E-03
Q31NQ0	Putative modulator of DNA gyrase	-2.93	7.83E-04

Functional analysis and enrichment were carried out using Omicsolution tool (https://www.omicsolution.org/wkomics/main/), which searched for Gene ontology (GO) terms within the DEPs based on the GO background of *S. elongatus* PCC 7942, cscB from *E. coli*, and SPS from *Synechocystis* sp. PCC 6803 downloaded from the UniProt. Figure 5-8 exhibits the most enriched GO terms in proteins with significantly lower relative abundance ($p < 0.05$) of *S. elongatus* in co-culture compared to the *S. elongatus* monoculture. Many proteins involved in the biosynthetic processes showed lower relative abundance in *S. elongatus* after co-cultured with A. vinelandii, which may relate to the higher growth rate.

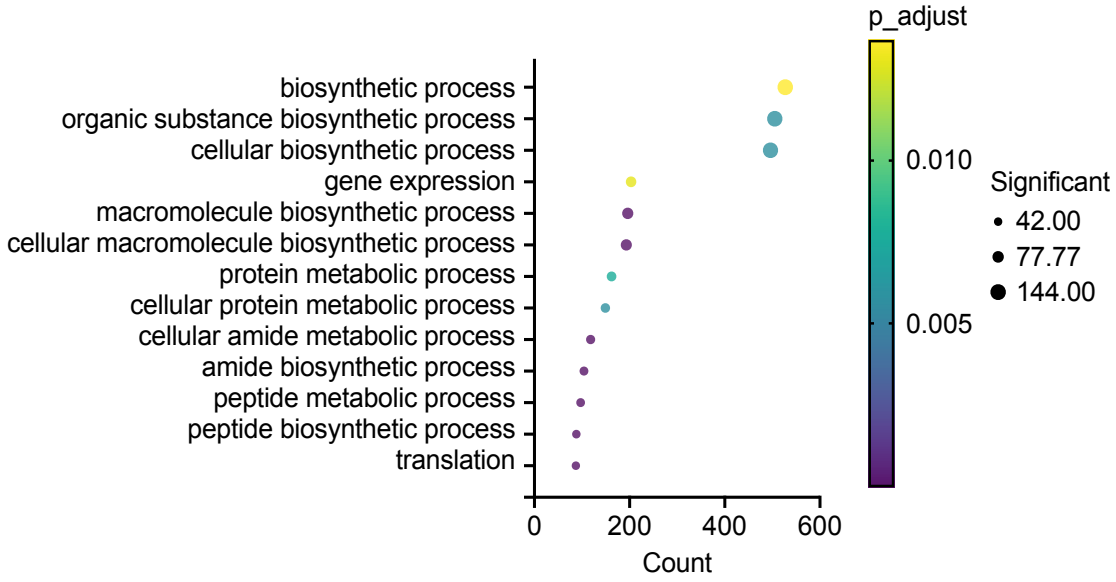


Figure 5-8 Enriched biological process GO terms of significantly lower relative abundance ($p < 0.05$) proteins in *S. elongatus* cscB/SPS after co-culture. Significantly DEPs with a $\log_2$ fold change cut-off of 1. Adjusted p-value for GO enrichment is 0.05. Figures were created using GraphPad Prism.

5.4.3 Characterisation of microbial community using quantitative shotgun proteomics at different time points

5.4.3.1 Proteomics results summary

Following *S. elongatus* cscB/SPS and *A. vinelandii* △nifL co-culturing in triplicate, proteins were extracted from days 0, 4, and 8 (Chapter 3, Figure 3-6) corresponding to the initial phase, exponential phase, and stationary phase of *S. elongatus* cscB/SPS, and subjected to LC-MS/MS analysis. Label-free shotgun proteomics approach was used to compare co-culture proteome on different time points, followed by MaxQuant analysis for peptide sequences generation and spectral matching (Cox et al., 2014). After removing contaminants and reverse sequences, identified proteins with two or more unique peptides in each sample were shown in Figure 5-9 A. The number of *S. elongatus* cscB/SPS had a severe decrease on day 4 compared to day 0 (Figure 5-9 B), which may be due to nitrogen starvation in the medium, as nitrate reduction plays an important role in electron sink and therefore affects the biological activity (Klotz et al., 2015). After that protein numbers keep relatively stable in both strains (Figure 5-9 B).

After normalization of the LFQ intensities of each protein using LFQRatio method established in Chapter 3, statistical analysis of quantified proteins was carried out using LFQ-Analyst. PCA analysis showed distinct clustering of each time point, although biological replicate 1 (showed in technical replicates 1 and 2) clustered away from biological 2 and 3 (showed in technical replicates 3, 4, 5, and 6) on the 8th day (Figure 5-9 C). The heatmap provided an overview of all DEPs across all samples, indicating that proteome in co-culture fluctuated with time (Figure 5-9 D).

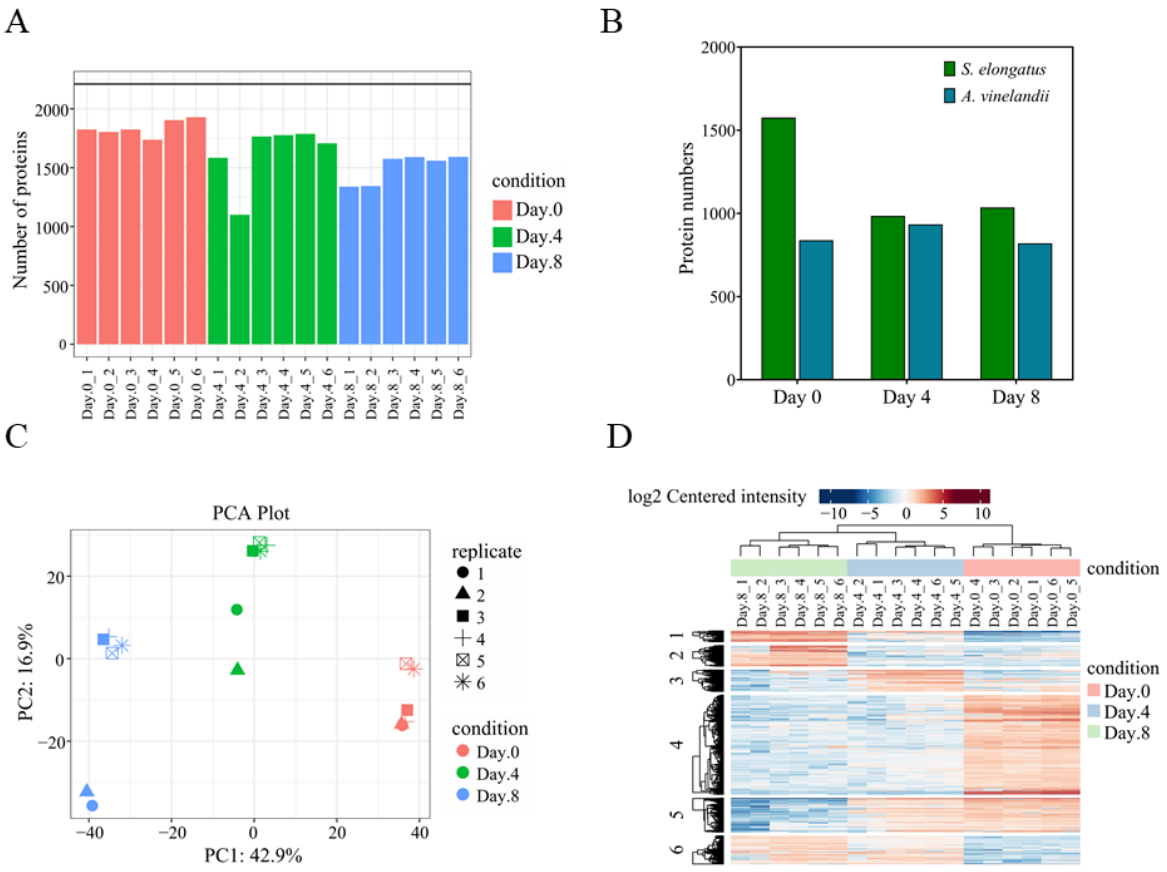


Figure 5-9 Proteomics data analysis of *S. elongatus* cscB/SPS and *A. vinelandii* △nifL co-culture at different time points. (A) Histogram of total protein numbers of each sample in three biological replicates and two technical replicates. (B) Identified proteins of *S. elongatus* cscB/SPS (green) and *A. vinelandii* △nifL (blue) in co-culture at different time points. (C) PCA plot of individual samples showed distinct clustering of each condition. (D) Heatmap provided an overview of all DEPs (rows) across all samples (columns) in 6 clusters.

LFQ-Analyst results showed 1293 proteins (58.48%) differently expressed (*p.adj* < 0.05) of 2211 proteins among samples with $\log_2$ fold change cutoff of 1. Figure 5-10 exhibited higher and lower relative abundance proteins of *S. elongatus* cscB/SPS and *A. vinelandii* △nifL in co-culture on day 4 and day 8, respectively. It can be observed that lots of proteins showed higher (100) or lower (520) relative abundance in *S. elongatus* cscB/SPS on day 4 compared to day 0, while many proteins showed higher (94) or lower (188) relative abundance in *A. vinelandii* △nifL on day 8 compared to day 4. This indicates the cyanobacterium, *S. elongatus* cscB/SPS, tends to adapt to the co-culture environment at the beginning (from day 0 to day 4),

while the heterotroph, *A. vinelandii* △nifL, adapts to the co-culture environment a little late (from day 4 to day 8).

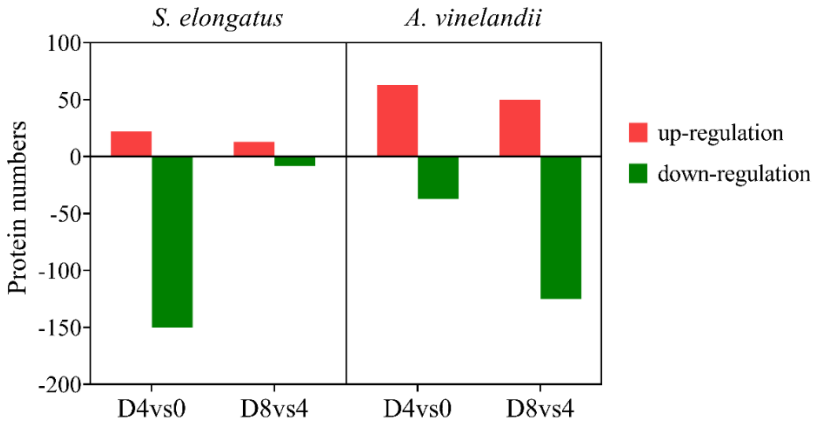


Figure 5-10 Higher (red) and lower (green) relative abundance of proteins in *S. elongatus* cscB/SPS (left panel) and *A. vinelandii* △nifL (right panel) on day 4 and day 8 in co-culture compared to their monocultures.

5.4.3.2 Significant pathways analysis on day 4 compared to day 0

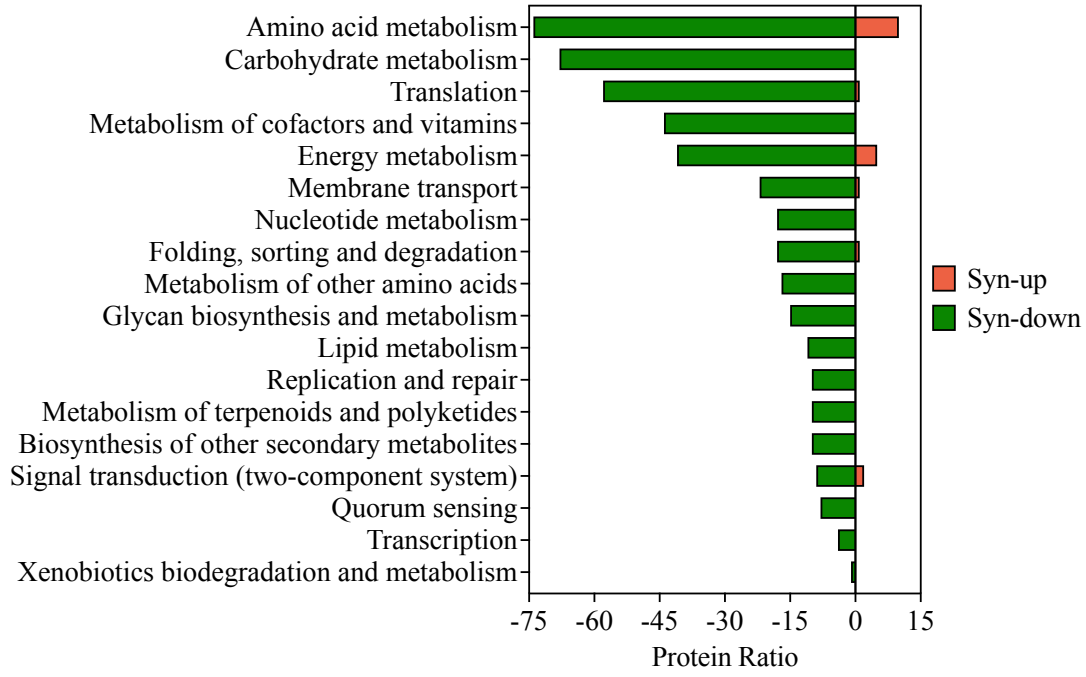
To figure out the roles of the differentially expressed proteins during co-culture, the DEPs were mapped to their corresponding terms in the KEGG database (Figure 5-11). Most KEGG pathways in S. *elongatus* cscB/SPS exhibited relatively lower abundance in co-culture on day 4 (exponential phase) compared to day 0 (initial phase), such as amino acid metabolism, carbohydrate metabolism, energy metabolism, translation and metabolism of cofactors and vitamins (Figure 5-11 A), suggesting a decreased demand for these related nutrients, which was consistent with slower growth rate in co-culture than in monoculture from day 0 to day 4 (Figure 3-6 B, Chapter 3). It is worth mentioning that nitrate transport permease and nitrate transport ATP-binding subunits C and D (with log2 FC of -6.05, -4.5, and -9.1, respectively), involved in membrane transport of nitrate uptake, exhibited significant down-regulation on day 4 compared to day 0. This is due to the depletion of the initial nitrate in the medium.

Whereas there were still some pathways that exhibited relatively higher abundance, involving amino acid metabolism, energy metabolism, and signal transduction. An increase in photosynthesis was achieved by the upregulation of plastocyanin, a copper-containing protein that mediates electron transfer (Geerts et al., 1994). Signal transduction was realised through

two-component regulatory systems (TCSs) (via CheY protein), a stimulus-response mechanism that senses environmental changes (Kera et al., 2020). This information indicated protein changes of S. *elongatus* cscB/SPS impacted by *A. vinelandii* △nifL.

More up-regulated KEGG pathways were detected in *A. vinelandii* △nifL during co-culture from day 0 to day 4. The top three higher relative abundance pathways were carbohydrate metabolism, amino acid metabolism, and cell motility, while the top three lower relative abundance pathways were assigned to amino acid metabolism, metabolism of cofactors and vitamins, and carbohydrate metabolism (Figure 5-11 B). Among the higher abundance proteins, two of them were mapped to sucrose metabolism, i.e., sucrose-6-phosphate hydrolase and glucokinase, which revealed increased demand for carbon sources for *A. vinelandii* △nifL. Interestingly, glycan biosynthesis and metabolism, metabolism of terpenoid and polyketides were up-regulated in *A. vinelandii* △nifL, indicating an metabolic changes to adapt to co-culture with S. *elongatus* cscB/SPS.

A



B

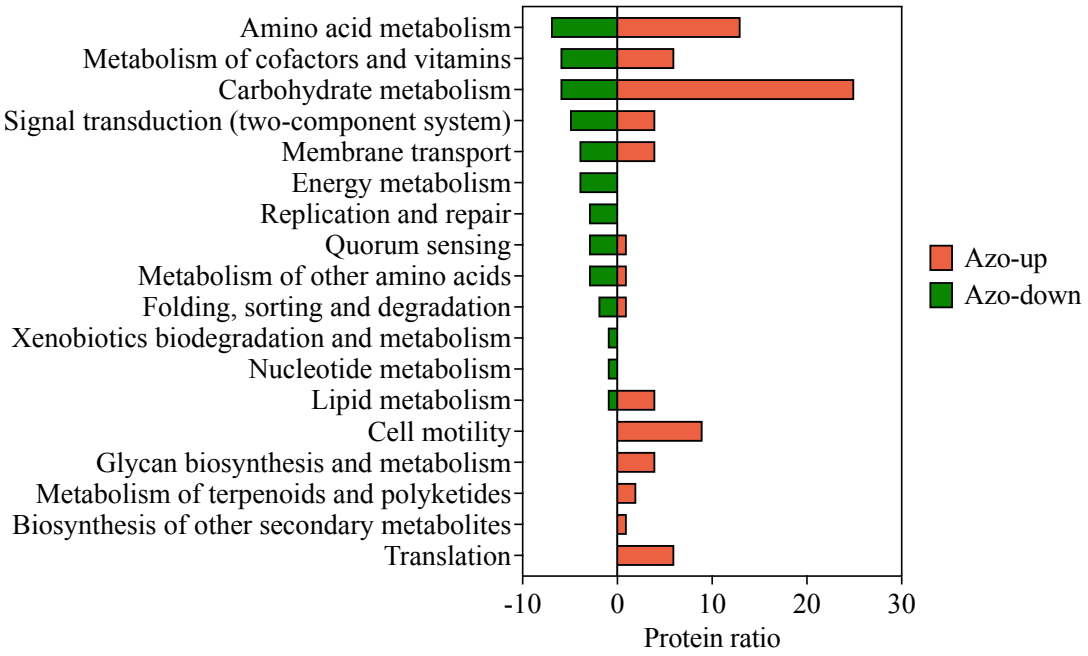


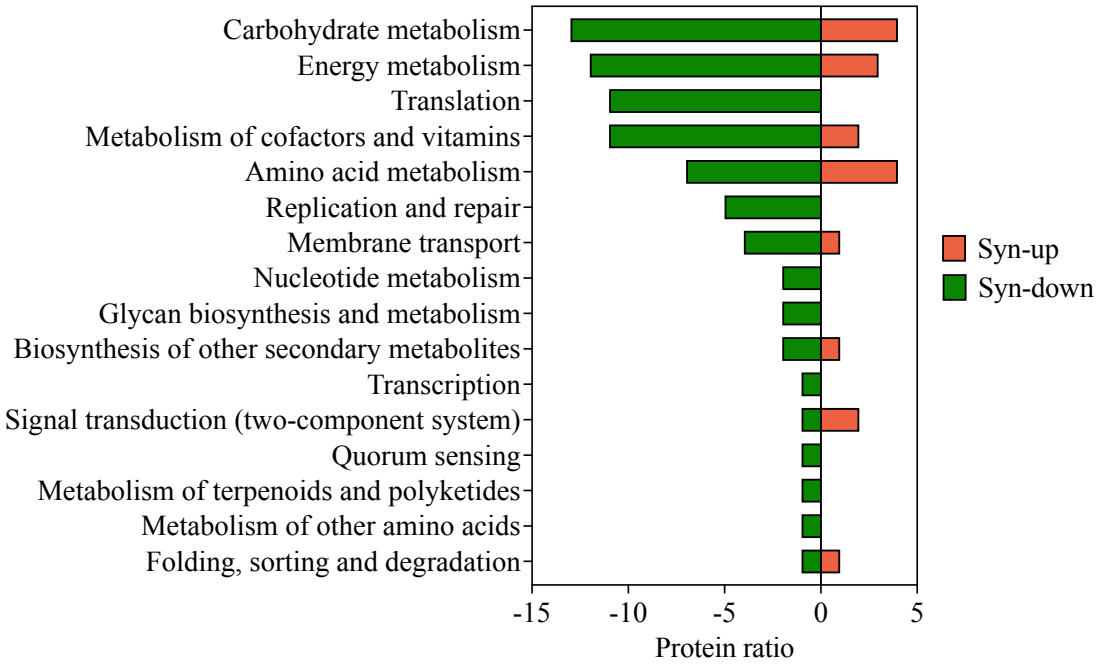
Figure 5-11 KEGG pathways of significant proteins in *S. elongatus* cscB/SPS and *A. vinelandii* △nifL on day 4 compared to day 0. (A) KEGG pathways within higher (red) and lower (green) abundance proteins in *S. elongatus* cscB/SPS on day 4 compared to day 0. (B) KEGG pathways within higher (red) and lower (green) abundance proteins in *A. vinelandii* △nifL on day 4

compared to day 0. Level II KEGG pathway terms used.

5.4.3.2 Significant pathways analysis on day 8 compared to day 4

From day 4 to day 8, the major down-regulated KEGG pathways in *A. vinelandii* △nifL strain during co-culture occurred in carbohydrate metabolism, amino acid metabolism, and energy metabolism (Figure 5-12 B. Whereas, a greatly up-regulation was observed in two-component system with 7 proteins, including OmpR family (osmolarity-sensing histidine protein kinase EnvZ, flagellin FliC, and tripartite tricarboxylate transporter periplasmic receptor protein TctC), CitB family (TRAP dicarboxylate transporter-periplasmic solute receptor subunit DctP), LytTR family (encystment and alginate biosynthesis response regulator; AlgR), NtrC family (RNA polymerase sigma-54 factor RpoN), and other family (polysaccharide export protein), indicating the increase of signal transduction in co-culture between the two species (Appendix D3). It is worth mentioning that three proteins related to flagellar assembly were up-regulated while two proteins related to bacterial chemotaxis were down-regulated in the cell motility pathway. Chemotaxis is the process by which bacteria respond to environmental changes and move towards more favourable conditions, which is achieved by modulating the direction of flagellar (Haneline et al., 1991; Bren and Eisenbach, 2000). The down-regulation of bacterial chemotaxis suggested co-culture environment tend to balance from day 4 to day 8, however the reasons for the up-regulation of flagellar assembly requires further investigation. KEGG pathways in *S. elongatus* cscB/SPS did not change too much from day 4 to day 8, a relatively remarkable down-regulation occurred in metabolism of cofactors and vitamins, manipulated by thiamine biosynthesis protein ThiC, protoporphyrin IX magnesium-chelatase, and quinolinate synthetase A.

A



B

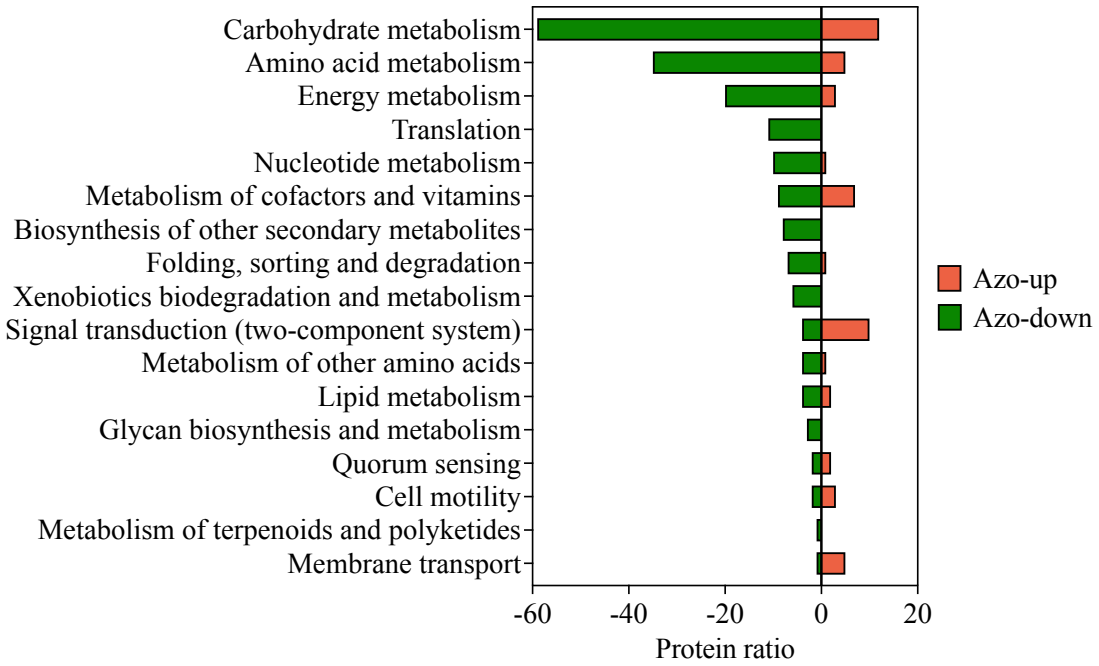


Figure 5-12 KEGG pathways of significant proteins in *S. elongatus* cscB/SPS and *A. vinelandii* △nifL on day 8 compared to day 4. (A) KEGG pathways within higher (red) and lower (green) abundance proteins in *S. elongatus* cscB/SPS on day 8 compared to day 4. (B) KEGG pathways

within higher (red) and lower (green) abundance proteins in *A. vinelandii* △nifL on day 8 compared to day 4. Level II KEGG pathway terms used.

Apart from the significantly changed proteins, the engineered proteins related to ammonium and sucrose production were also evaluated. NifL, a repressor protein for nitrogenase, was knocked out from *A. vinelandii* and it was not detected in the proteomics results either. Exogenous sucrose phosphate synthase (SPS) from *Synechocystis* sp. PCC 6083 and sucrose permease (cscB) from *E. coli* were both identified during the co-culture, with slight down regulation from day 0 to day 4 (log_2 fold change of -0.467 and -0.544, respectively), which is due to the addition of initial sucrose. From day 4 to day 8, SPS (log_2 FC 1.17) and cscB (log_2 FC 0.258) were up-regulated, indicating the increase of sucrose synthesis and output from the *S. elongatus* cscB/SPS.

The engineered genes aimed at the increase of the ammonium production from *A. vinelandii* △nifL and sucrose production from *S. elongatus* cscB/SPS in monoculture, and this also enhanced the cross-feeding of the nitrogen and carbon sources from each other, which was verified by the up-regulation of related ammonium-produced proteins (NifD, NifK, and NifH) and sucrose-produced proteins (CscB and SPS). After proteomics analysis, the potential metabolic targets are engineered sucrose uptake in *A. vinelandii* △nifL and ammonium transporter in *S. elongatus* cscB/SPS. Sucrose permease did not show a significantly up-regulation, and it can be improved by increase the salt pressure to promote the sucrose output (Ducat et al., 2012).

5.5 Conclusions

In this chapter, a comprehensive proteomics analysis was performed on the synthetic co-culture of a nitrogen fixation bacterium *A. vinelandii* △nifL and photosynthetic cyanobacterium *S. elongatus* cscB/SPS, identifying proteins for 32.5% of the *A. vinelandii* △nifL proteome and 61.5% of the *S. elongatus* cscB/SPS proteome with two or more unique peptides. Proteome changes between the co-culture and monocultures control were investigated, showing the up-regulation of nitrogen fixation and photosynthesis pathways, which is evidence of the cross-feeding between the two members. In addition, co-culture proteomics changes over time were

studied, demonstrating some interactions between co-cultures, including general carbon metabolism, two-component system, bacterial chemotaxis, etc. These proteomics results provide detailed information on the protein changes between the synthetic microbial community and reveal potential targets for optimisation to keep the synthetic microbial co-culture healthier and longer-lived.

Chapter 6 Conclusion and future work

The overall aims of this thesis were to establish a self-sufficient cross-feeding synthetic microbial co-culture, which can utilise atmospheric CO_2 and N_2 as the carbon and nitrogen sources for PHB production (Chapter 3). To deepen the understanding of interactions in co-culture, a proteomics pipeline was developed to analyse the protein and metabolic changes in the co-culture, with the consideration of many factors affecting the quantitative accuracy of co-culture proteomics (Chapter 4). This proteomics pipeline was applied to the synthetic microbial co-culture, revealing the nutrient exchange and growth stress in the co-culture (Chapter 5). In this chapter, the main findings of this thesis will be summarised, and future research directions will be proposed.

6.1 Construction of a carbon and nitrogen fixing synthetic microbial consortium for PHB production

Synthetic microbial co-culture is not easy to research as many interactions can occur among microbial communities, through either positive effect such as mutualism, protocooperation, and commensalism, or negative effects such as amensalism, predation, and competition (Kong et al., 2018). In Chapter 3, a synthetic co-culture of engineered *S. elongatus* cscB/SPS and *A. vinelandii* ΔnifL was established based on a designed mutualism interaction on a laboratory scale and could be co-cultured together for at least two weeks. The co-culture medium is crucial for co-culture growth, different media experiments suggested that initial carbon and nitrogen sources are compulsory for the start of co-culture (Figure 3-5). The initial inoculation ratio is another key experimental parameter for the growth of synthetic co-cultures as it not only affects the final ratio of the co-cultures but also regulates the metabolic capacity of the co-cultures (Gao et al., 2021). The experiment showed that the co-culture grew better when the starting ratio of *S. elongatus* cscB/SPS and *A. vinelandii* △nifL is 80:20. Long-term growth experiments illustrate metabolic coupling between *S. elongatus* cscB/SPS and *A. vinelandii* △nifL, demonstrating robust nutrient exchange between the phototroph and heterotroph.

The synthetic co-culture composed of *S. elongatus* cscB/SPS and *A. vinelandii* △nifL has the ability to convert atmospheric CO_2 and N_2 into PHB, which has been confirmed by GC analysis (Figure 3-8). Sucrose produced by *S. elongatus* cscB/SPS and ammonium produced by *A. vinelandii* △nifL can be used as mutualistic nutrients for carbon and nitrogen self-sufficiency,

greatly reducing growth and production costs. This proves that the strategy of combining heterotrophic bacteria with photosynthetic organisms to produce industrially relevant compounds is successful. This will be one of the most sustainable manufacturing technologies and will enable manufacturing even in extreme environments. Although the yield of PHB from this co-culture system is low, the yield of biopolymers can be increased by other methods such as engineering-directed synthetic pathways, which will be discussed in the future study directions.

6.2 Establishing a label-free proteomics workflow for microbial co-cultures

Label-free quantitative (LFQ) proteomics, which is used to analyse overall proteomic changes under different biological conditions, is widely applicable to microbial monocultures (Wegener et al., 2010; Guerreiro et al., 2014), however, applying these common workflows to synthetic co-cultures, i.e., when several strain types are cultured together, poses certain challenges. In Chapter 4, a comprehensive study was conducted about how to establish a label-free proteomics workflow for microbial co-cultures, including analysis of factors affecting the quantitative accuracy, different quantification methods, and appropriate normalisation approach. Before the co-culture proteomics experiments, many factors affecting the quantitative accuracy were summarised and analysed including the physicochemical characteristics of consortium member proteomes, such as the range of isoelectric points (*pI*) (Kozlowski, 2017), molecular weight (M_w) (Angel et al., 2012), and hydrophobicity (Warwood et al., 2013), as well as the dynamic range distribution of protein abundances within the proteome (Zubarev, 2013). Although these factors have been verified as a minimal difference in my co-culture members *S. elongatus* and *A. vinelandii*, these should be considered in other co-culture proteomics workflows, particularly where proteome sizes of co-culture members vary more widely.

A comprehensive analysis of 6 different label-free quantification methods was performed at the protein level and cell level, demonstrating that all the tested parameters had good linear relationships with protein amount. Modelling verified LFQ intensities do not correlate to cell ratios in cell mixes. Therefore, a normalisation approach 'LFQRatio' was created to minimise the influence of cell number changes on protein quantification when analysing the proteome during co-culture. This normalisation method also reveals insight into how to normalise the proteomic data when loading an unknown amount of protein from different strains in bi-culture or even multiple cultures.

6.3 Proteomics analysis of a synthetic microbial community – *S. elongatus* cscB/SPS and *A. vinelandii* ΔnifL

In Chapter 5, using the established shotgun proteomics approach, proteome changes between the co-culture and monoculture control were investigated, revealing a higher relative abundance of proteins in nitrogen fixation in *A. vinelandii* △nifL and some proteins in the photosynthesis pathway in *S. elongatus* cscB/SPS after co-culture, indicating cross-feeding between the two members. Metal limitations were observed from the proteomic results, which provide a basis for the rational design of the co-culture medium. Furthermore, the evolution of co-culture proteomics was investigated, revealing several connections between co-cultures such as general carbon metabolism, two-component systems, bacterial chemotaxis, and so on. These proteomics findings give precise information on protein changes in the synthetic microbial community and show nutrient stress resulting from the co-culture, which could be a good target for engineering the key enzymes to enhance the symposium and bioproduction.

In addition, since applying a quantitative proteomics approach to a synthetic microbial co-culture is a novel concept, there are still many aspects worthy of investigation, as discussed below.

6.4 Future study directions

6.4.1 Trace element adjustment in the co-culture medium

The proteomics results in Chapter 5 revealed many protein changes in the two-component system. Two-component signal transduction system (TCS) enables bacteria to sense, respond, and adapt to changes in the environment or intracellular state (Zschiedrich et al., 2016). Among them, the tonB-dependent ferrisiderophore receptor protein (pfeA, log2FC of 1.75), which is involved in the transport of siderophores into the periplasm, showed a higher relative abundance in *A. vinelandii* ΔnifL, indicating iron limitation after co-culture (Ferguson and Deisenhofer, 2002). Therefore, the co-culture medium can be adjusted by increasing the iron concentration in order to make the synthetic microbial co-culture longer-lived.

6.4.2 Enhanced PHB production in the co-culture

The synthetic microbial co-culture system has been verified a potential of PHB production using atmospheric CO_2 and N_2 in chapter 3, however, the PHB yield is low. An efficient way to enhance the PHB production in the co-culture is to genetically engineer the producer *A.*

vinelandii. For example, overexpressing the genes of key enzymes or introducing exogenous highly active enzymes in the PHB synthesis pathways. Class I PHA synthases preferentially utilize the short-chain length (SCL) monomers (C3-C5), and have been used to enhance PHA production in *E. coli* (Shi et al., 2022).

Another approach to enhance the PHB production is by adjusting the C/N ratio. Nitrogen limitation has been poved favourable for PHB production (Zhou et al., 2022), and the microbial co-culture system in this study fulfills this condition. Johnson et al. have reported that microbial reaction rates varied with the limiting substrate: high acetate uptake was found in carbon-limiting condition (medium C/N ratio 6-13.2 Cmol/Nmol), whereas in nitrogen-limiting condition (medium C/N ratios 6-13.2 Cmol/Nmol) C/N ratios 15-24 Cmol/Nmol) high ammonia uptake rates was achieved (Johnson et al., 2010). In the case of strong nitrogen limitation, biomass has a higher baseline PHA content, but carbon limitation usually results in biomass with a higher maximum PHA storage capacity (Johnson et al., 2010). Thus, C/N ratio could be monitored and adjusted during the growth of synthetic co-culture to achieve high PHB production.

6.4.3 Membrane proteomics to analyse interactions between the two members

In Chapter 4 and Chapter 5, a label-free proteomics workflow for analysing synthetic microbial co-culture was established. However, membrane proteins are not collected by specific method and therefore some important information may be lost. It is expected that 30% of natural proteins are embedded in biological membranes (Tan et al., 2008), therefore, membrane proteins are good targets as they can reflect some physical contact and chemical signals during the co-culture. An efficient membrane protein extraction and characterised method for microbial co-culture should be explored based on Han et al.'s research (Han et al., 2008).

6.4.4 Introduce a third member to utilise sucrose and ammonium for bioproduction

S. elongatus cscB/SPS and *A. vinelandii* △nifL have been verified to be co-cultured together, *S. elongatus* cscB/SPS can utilise air CO_2 as the carbon source to produce sucrose, and *A. vinelandii* △nifL can use N_2 as the nitrogen source to produce ammonium. Therefore, we can introduce a third member, such as *E. coli*, which can use sucrose and ammonium for bioproduction. According to this idea, carbon dioxide and nitrogen in the atmosphere can be used to achieve bioproduction and synthesis of a variety of high-value biological products by introducing an appropriately engineered third member.

References

Aakko, J., Pietilä, S., Suomi, T., Mahmoudian, M., Toivonen, R., Kouvonen, P., et al. (2020). Data-Independent Acquisition Mass Spectrometry in Metaproteomics of Gut Microbiota - Implementation and Computational Analysis. *J. Proteome Res.* 19, 432–436. doi:10.1021/acs.jproteome.9b00606.

Aasfar, A., Bargaz, A., Yaakoubi, K., Hilali, A., Bennis, I., Zeroual, Y., et al. (2021). Nitrogen Fixing Azotobacter Species as Potential Soil Biological Enhancers for Crop Nutrition and Yield Stability. *Front. Microbiol.* 12, 1–19. doi:10.3389/fmicb.2021.628379.

Abramson, B. W., Kachel, B., Kramer, D. M., and Ducat, D. C. (2016). Increased photochemical efficiency in cyanobacteria via an engineered sucrose sink. *Plant Cell Physiol.* 57, 2451–2460. doi:10.1093/pcp/pcw169.

Aebersold, R., and Mann, M. (2003). Mass spectrometry-based proteomics. *Nature* 422, 198–207. doi:doi: 10.1038/nature01511.

Al Shweiki, M. R., Mönchgesang, S., Majovsky, P., Thieme, D., Trutschel, D., and Hoehenwarter, W. (2017). Assessment of Label-Free Quantification in Discovery Proteomics and Impact of Technological Factors and Natural Variability of Protein Abundance. *J. Proteome Res.* 16, 1410–1424. doi:10.1021/acs.jproteome.6b00645.

Alam, M. A., Vandamme, D., Chun, W., Zhao, X., Foubert, I., Wang, Z., et al. (2016). Bioflocculation as an innovative harvesting strategy for microalgae. *Rev. Environ. Sci. Bio/Technology* 15, 573–583. doi:10.1007/s11157-016-9408-8.

Alleman, A. B., Mus, F., and Peters, J. W. (2021). Metabolic Model of the Nitrogen-Fixing Obligate Aerobe Azotobacter vinelandii Predicts Its Adaptation to Oxygen Concentration and Metal Availability. *MBio* 12, 1–15. doi:10.1128/mBio.02593-21.

Allison, S. D., and Martiny, J. B. H. (2009). "Resistance, resilience, and redundancy in microbial communities," in *In the Light of Evolution* (National Academies Press), 149–166. doi:10.17226/12501.

Altelaar, A. F. M., Munoz, J., and Heck, A. J. R. (2013). Next-generation proteomics : towards an integrative view of proteome dynamics. *Nat. Rev. Genet.* 14, 35–48. doi:10.1038/nrg3356.

Ames, N., Ranucci, A., Moriyama, B., and Wallen, G. (2017). The Human Microbiome and Understanding the 16S rRNA Gene in Translational Nursing Science. *Nurs Res* 66, 184–197. doi:10.1097/NNR.0000000000000212.

Amin, S. A., Hmelo, L. R., Van Tol, H. M., Durham, B. P., Carlson, L. T., Heal, K. R., et al. (2015). Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* 522, 98–101. doi:10.1038/nature14488.

Anand, S., Samuel, M., Ang, C.-S., Keerthikumar, S., and Mathivanan, S. (2017). "Label-Based and Label-Free Strategies for Protein Quantification," in *Proteome Bioinformatics* (Humana Press, New York, NY), 31–43. doi:doi: 10.1007/978-1-4939-6740-7_4.

Andersson, I. (2008). Catalysis and regulation in Rubisco. *J. Exp. Bot.* 59, 1555–1568. doi:10.1093/jxb/ern091.

Ang, K. S., Lakshmanan, M., Lee, N.-R., and Lee, D.-Y. (2018). Metabolic Modeling of Microbial Community Interactions for Health, Environmental and Biotechnological Applications. *Curr. Genomics* 19, 712–722. doi:10.2174/1389202919666180911144055.

Angel, T. E., Aryal, U. K., Hengel, S. M., Baker, E. S., Kelly, R. T., Robinson, E. W., et al. (2012). Mass spectrometry-based proteomics: Existing capabilities and future directions. *Chem. Soc. Rev.* 41, 3912–3928. doi:10.1039/c2cs15331a.

Arachea, B. T., Sun, Z., Potente, N., Malik, R., Isailovic, D., and Viola, R. E. (2012). Detergent selection for enhanced extraction of membrane proteins. *Protein Expr. Purif.* 86, 12–20. doi:10.1016/j.pep.2012.08.016.

Argun, H., Kargi, F., and Kapdan, I. K. (2009). Effects of the substrate and cell concentration on bio-hydrogen production from ground wheat by combined dark and photo-fermentation. *Int. J. Hydrogen Energy* 34, 6181–6188.

doi:10.1016/j.ijhydene.2009.05.130.

Arike, L., Valgepea, K., Peil, L., Nahku, R., Adamberg, K., and Vilu, R. (2012). Comparison and applications of label-free absolute proteome quantification methods on Escherichia coli. *J. Proteomics* 75, 5437–5448. doi:10.1016/j.jprot.2012.06.020.

Aschenbrenner, I. A., Cernava, T., Berg, G., and Grube, M. (2016). Understanding microbial multi-species symbioses. *Front. Microbiol.* 7, 1–9. doi:10.3389/fmicb.2016.00180.

Asfahl, K. L., and Schuster, M. (2017). Social interactions in bacterial cell-cell signaling. *FEMS Microbiol. Rev.* 41, 92–107. doi:10.1093/femsre/fuw038.

Baars, O., Zhang, X., Morel, F. M. M., and Seyedsayamdost, M. R. (2016). The siderophore metabolome of Azotobacter vinelandii. *Appl. Environ. Microbiol.* 82, 27–39. doi:10.1128/AEM.03160-15.

Bali, A., Blanco, G., Hill, S., and Kennedy, C. (1992). Excretion of ammonium by a nifL mutant of Azotobacter vinelandii fixing nitrogen. *Appl. Environ. Microbiol.* 58, 1711–1718. doi:doi: 10.1128/aem.58.5.1711-1718.1992.

Banerjee, S., and Mazumdar, S. (2012). Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte. *Int. J. Anal. Chem.* 2012, 1–40. doi:10.1155/2012/282574.

Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., and Kuster, B. (2007). Quantitative mass spectrometry in proteomics: A critical review. *Anal. Bioanal. Chem.* 389, 1017–1031. doi:10.1007/s00216-007-1486-6.

Barney, B. M., Eberhart, L. J., Ohlert, J. M., Knutson, C. M., and Plunkett, M. H. (2015). Gene deletions resulting in increased nitrogen release by Azotobacter vinelandii: Application of a novel nitrogen biosensor. *Appl. Environ. Microbiol.* 81, 4316–4328. doi:10.1128/AEM.00554-15.

Bellenger, J. P., Arnaud-Neu, F., Asfari, Z., Myneni, S. C. B., Stiefel, E. I., and Kraepiel, A. M. L. (2007). Complexation of oxoanions and cationic metals by the biscatecholate siderophore azotochelin. *J. Biol. Inorg. Chem.* 12, 367–376.

doi:10.1007/s00775-006-0194-6.

Benomar, S., Ranava, D., Cárdenas, M. L., Trably, E., Rafrafi, Y., Ducret, A., et al. (2015). Nutritional stress induces exchange of cell material and energetic coupling between bacterial species. *Nat. Commun.* 6, 1–10. doi:10.1038/ncomms7283.

Blein-Nicolas, M., and Zivy, M. (2016). Thousand and one ways to quantify and compare protein abundances in label-free bottom-up proteomics. *Biochim. Biophys. Acta - Proteins Proteomics* 1864, 883–895. doi:10.1016/j.bbapap.2016.02.019.

Boaro, A. A., Kim, Y. M., Konopka, A. E., Callister, S. J., and Ahring, B. K. (2014). Integrated 'omics analysis for studying the microbial community response to a pH perturbation of a cellulose-degrading bioreactor culture. *FEMS Microbiol. Ecol.* 90, 802–815. doi:10.1111/1574-6941.12435.

Bren, A., and Eisenbach, M. (2000). How Signals Are Heard during Bacterial Chemotaxis : Protein-Protein Interactions in Sensory Signal Propagation. *J. Bacteriol.* 182, 6865–6873.

Brenner, K., and Arnold, F. H. (2011). Self-organization, layered structure, and aggregation enhance persistence of a synthetic biofilm consortium. *PLoS One* 6, 1–7. doi:10.1371/journal.pone.0016791.

Brenner, K., You, L., and Arnold, F. H. (2008). Engineering microbial consortia: a new frontier in synthetic biology. *Trends Biotechnol.* 26, 483–489. doi:10.1016/j.tibtech.2008.05.004.

Cai, Z., Liu, G., Zhang, J., and Li, Y. (2014). Development of an activity-directed selection system enabled signi fi cant improvement of the carboxylation ef fi ciency of Rubisco. 5, 552–562. doi:10.1007/s13238-014-0072-x.

Cairns, J., Jokela, R., Hultman, J., Tamminen, M., Virta, M., and Hiltunen, T. (2018). Construction and characterization of synthetic bacterial community for experimental ecology and evolution. *Front. Genet.* 9. doi:10.3389/fgene.2018.00312.

Cappadona, S., Baker, P. R., Cutillas, P. R., Heck, A. J. R., and Van Breukelen, B. (2012). Current challenges in software solutions for mass spectrometry-based quantitative proteomics. *Amino Acids* 43, 1087–1108. doi:10.1007/s00726-012-1289-8.

Carmo-Silva, E., Scales, J. C., Madgwick, P. J., and Parry, M. A. J. (2015). Optimizing Rubisco and its regulation for greater resource use efficiency. *Plant Cell Environ.* 38, 1817–1832. doi:10.1111/pce.12425.

Chen, Y., and Liu, L. (2019). "Targeted Proteomics," in *Functional Proteomics Methods and Protocols*, eds. X. Wang and M. Kuruc (Humana Press), 265–277.

Chowdhury-Paul, S., Pando-Robles, V., Jiménez-Jacinto, V., Segura, D., Espín, G., and Núñez, C. (2018). Proteomic analysis revealed proteins induced upon Azotobacter vinelandii encystment. *J. Proteomics* 181, 47–59. doi:10.1016/j.jprot.2018.03.031.

Christie-Oleza, J. A., Sousoni, D., Lloyd, M., Armengaud, J., and Scanlan, D. J. (2017). Nutrient recycling facilitates long-term stability of marine microbial phototroph-heterotroph interactions. *Nat. Microbiol.* 2, 1–10. doi:10.1038/nmicrobiol.2017.100.

Chwa, J. W., Kim, W. J., Sim, S. J., Um, Y., and Woo, H. M. (2016). Engineering of a modular and synthetic phosphoketolase pathway for photosynthetic production of acetone from CO2 in Synechococcus elongatus PCC 7942 under light and aerobic condition. *Plant Biotechnol. J.* 14, 1768–1776. doi:10.1111/pbi.12536.

Clarridge, J. E. (2004). Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clin. Microbiol. Rev.* 17, 840–862. doi:10.1128/CMR.17.4.840-862.2004.

Cornelis, P., Matthijs, S., and Van Oeffelen, L. (2009). Iron uptake regulation in Pseudomonas aeruginosa. in *BioMetals*, 15–22. doi:10.1007/s10534-008-9193-0.

Cox, J., Hein, M. Y., Luber, C. A., Paron, I., Nagaraj, N., and Mann, M. (2014). Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. *Mol. Cell. Proteomics*

13, 2513–2526. doi:10.1074/mcp.

Croft, M. T., Lawrence, A. D., Raux-Deery, E., Warren, M. J., and Smith, A. G. (2005). Algae acquire vitamin B12 through a symbiotic relationship with bacteria. *Nature* 438, 90–93. doi:10.1038/nature04056.

Cydzik-Kwiatkowska, A., and Zielińska, M. (2016). Bacterial communities in full-scale wastewater treatment systems. *World J. Microbiol. Biotechnol.* 32, 1–8. doi:10.1007/s11274-016-2012-9.

Demoulin, C. F., Lara, Y. J., Cornet, L., François, C., Baurain, D., Wilmotte, A., et al. (2019). Cyanobacteria evolution: Insight from the fossil record. *Free Radic. Biol. Med.* 140, 206–223. doi:10.1016/j.freeradbiomed.2019.05.007.

Dexter, J., and Fu, P. (2009). Metabolic engineering of cyanobacteria for ethanol production. *Energy Environ. Sci.* 2, 857–864. doi:10.1039/b811937f.

Di, S., and Yang, A. (2019). Analysis of productivity and stability of synthetic microbial communities. *J. R. Soc. Interface* 16, 20180859. doi:10.1098/rsif.2018.0859.

Dicker, L., Lin, X., and Ivanov, A. R. (2010). Increased power for the analysis of label-free LC-MS/MS proteomics data by combining spectral counts and peptide peak attributes. *Mol. Cell. Proteomics* 9, 2704–2718. doi:10.1074/mcp.M110.002774.

Donohoe, D. R., Garge, N., Zhang, X., Sun, W., O'Connell, T. M., Bunger, M. K., et al. (2011). The microbiome and butyrate regulate energy metabolism and autophagy in the mammalian colon. *Cell Metab.* 13, 517–526. doi:10.1016/j.cmet.2011.02.018.

Dos Santos, P. C. (2011). "Molecular Biology and Genetic Engineering in Nitrogen Fixation," in *Nitrogen Fixation: Methods and Protocols*, ed. M. W. Ribbe (Totowa, NJ: Humana Press), 81–92. doi:10.1007/978-1-61779-194-9_6.

Dreisewerd, K. (2003). The desorption process in MALDI. *Chem. Rev.* 103, 395–425. doi:10.1021/cr010375i.

Duan, K., Sibley, C. D., Davidson, C. J., and Surette, M. G. (2009). “Chemical interactions between organisms in microbial communities.,” in *Contributions to microbiology*, 1–17. doi:10.1159/000219369.

Duan, Y., Luo, Q., Liang, F., and Lu, X. (2016). Sucrose secreted by the engineered cyanobacterium and its fermentability. *J. Ocean Univ. China* 15. doi:10.1007/s11802-016-3007-8.

Ducat, D. C., Avelar-Rivas, J. A., Way, J. C., and Silvera, P. A. (2012). Rerouting carbon flux to enhance photosynthetic productivity. *Appl. Environ. Microbiol.* 78, 2660–2668. doi:10.1128/AEM.07901-11.

Eiteman, M. A., Lee, S. A., and Altman, E. (2008). A co-fermentation strategy to consume sugar mixtures effectively. *J. Biol. Eng.* 2, 1–8. doi:10.1186/1754-1611-2-3.

El-shanshoury, A. E. R., Kenawy, E., and Amara, A. a (2013). Optimization of Polyhydroxybutyrate (PHB) production by Azotobacter vinelandii using experimental design. *Int. J. Curr. Microbiol. Appl. Sci.* 2, 227–241.

Eng, A., and Borenstein, E. (2019). Microbial community design: methods, applications, and opportunities. *Curr. Opin. Biotechnol.* 58, 117–128. doi:10.1016/j.copbio.2019.03.002.

Engel, C., Schattenberg, F., Dohnt, K., Schröder, U., Müller, S., and Krull, R. (2019). Long-term behavior of defined mixed cultures of geobacter sulfurreducens and Shewanella oneidensis in bioelectrochemical systems. *Front. Bioeng. Biotechnol.* 7. doi:10.3389/fbioe.2019.00060.

Esteves-Ferreira, A. A., Inaba, M., Fort, A., Araújo, W. L., and Sulpice, R. (2018). Nitrogen metabolism in cyanobacteria: metabolic and molecular control, growth consequences and biotechnological applications. *Crit. Rev. Microbiol.* 44, 541–560. doi:10.1080/1040841X.2018.1446902.

Feder, M. E., and Walser, J. C. (2005). The biological limitations of transcriptomics in elucidating stress and stress responses. *J. Evol. Biol.* 18, 901–910. doi:10.1111/j.1420-9101.2005.00921.x.

Ferguson, A. D., and Deisenhofer, J. (2002). TonB-dependent receptors-structural perspectives. *Biochim. Biophys. Acta* 1565, 318–332. Available at: www.bba-direct.com.

Flores, E., and Herrero, A. (2005). Nitrogen assimilation and nitrogen control in cyanobacteria. *Biochem. Soc. Trans.* 33, 164–167. doi:DOI: 10.1042/BST0330164.

Forchhammer, K. (2004). Global carbon/nitrogen control by PII signal transduction in cyanobacteria: From signals to targets. *FEMS Microbiol. Rev.* 28, 319–333. doi:10.1016/j.femsre.2003.11.001.

Forchhammer, K., and Schwarz, R. (2019). Nitrogen chlorosis in unicellular cyanobacteria – a developmental program for surviving nitrogen deprivation. *Environ. Microbiol.* 21, 1173–1184. doi:10.1111/1462-2920.14447.

Fukami, K., Nishijima, T., and Ishida, Y. (1997). Protective and inhibitory effects of brain extracts on the growth of bacteria. *Live Food Mar. Larvic.* 358, 185–191. doi:10.1007/BF02766185.

Gangeswaran, R., and Eady, R. R. (1996). Flavodoxin 1 of Azotobacter vinelandii: Characterization and role in electron donation to purified assimilatory nitrate reductase. *Biochem. J.* 317, 103–108. doi:10.1042/bj3170103.

Gao, C. H., Cao, H., Cai, P., and Sørensen, S. J. (2021). The initial inoculation ratio regulates bacterial coculture interactions and metabolic capacity. *ISME J.* 15, 29–40. doi:10.1038/s41396-020-00751-7.

Geerts, D., Schubert, H., De Vrieze, G., Borrias, M., Matthijs, H. C. P., and Weisbeek, P. J. (1994). Expression of Anabaena PCC 7937 plastocyanin in Synechococcus PCC 7942 enhances photosynthetic electron transfer and alters the electron distribution between photosystem I and cytochrome-c oxidase. *J. Biol. Chem.* 269, 28068–28075. doi:10.1016/s0021-9258(18)46896-2.

Ghosh, S., Chowdhury, R., and Bhattacharya, P. (2016). Mixed consortia in bioprocesses : role of microbial interactions. *Appl. Microbiol. Biotechnol.* 100, 4283–4295. doi:10.1007/s00253-016-7448-1.

Goeminne, L. J. E., Gevaert, K., and Clement, L. (2018). Experimental design and data- analysis in label-free quantitative LC/MS proteomics: A tutorial with MSqRob. *J. Proteomics* 171, 23–26. doi:10.1016/j.jprot.2017.04.004.

Gomes, L., Monteiro, G., and Mergulhão, F. (2020). The impact of IPTG induction on plasmid stability and heterologous protein expression by escherichia coli biofilms. *Int. J. Mol. Sci.* 21. doi:10.3390/ijms21020576.

Good, B. H., McDonald, M. J., Barrick, J. E., Lenski, R. E., and Desai, M. M. (2017). The dynamics of molecular evolution over 60,000 generations. *Nature* 551, 45–50. doi:10.1038/nature24287.

Grandel, N. E., Reyes Gamas, K., and Bennett, M. R. (2021). Control of synthetic microbial consortia in time, space, and composition. *Trends Microbiol.* 29, 1095–1105. doi:10.1016/j.tim.2021.04.001.

Grant, M. A. A., Kazamia, E., Cicuta, P., and Smith, A. G. (2014). Direct exchange of vitamin B 12 is demonstrated by modelling the growth dynamics of algal-bacterial cocultures. *ISME J.* 8, 1418–1427. doi:10.1038/ismej.2014.9.

Greenblum, S., Turnbaugh, P. J., and Borenstein, E. (2012). Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. *Proc. Natl. Acad. Sci. U. S. A.* 109, 594–599. doi:10.1073/pnas.1116053109.

Guerreiro, A. C. L., Benevento, M., Lehmann, R., Van Breukelen, B., Post, H., Giansanti, P., et al. (2014). Daily rhythms in the cyanobacterium Synechococcus elongatus probed by high-resolution mass spectrometry-based proteomics reveals a small defined set of cyclic proteins. *Mol. Cell. Proteomics* 13, 2042–2055. doi:10.1074/mcp.M113.035840.

Gundry, R. L., White, M. Y., Murray, C. I., Kane, L. A., Fu, Q., Stanley, B. A., et al. (2010). Preparation of Peptides for Mass Spectrometry Analysis in Bottom-Up Proteomics Workflows. *Curr. Protoc. Mol. Biol.*, 10.25.1-10.25.23. doi:10.1002/cpz1.85.

Guo, L. (2015). From Discovery-Based to Targeted Proteomics.

Gupta, R., and Gupta, N. (2021). *Fundamentals of Bacterial Physiology and Metabolism*. Singapore: Springer.

Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* 17, 994–999. doi:doi.org/10.1038/13690.

Hamilton, J. J., Calixto Contreras, M., and Reed, J. L. (2015). Thermodynamics and H2 Transfer in a Methanogenic, Syntrophic Community. *PLoS Comput. Biol.* 11, 1–20. doi:10.1371/journal.pcbi.1004364.

Hamilton, T. L., Ludwig, M., Dixon, R., Boyd, E. S., Santos, P. C. Dos, Setubal, C., et al. (2011). Transcriptional Profiling of Nitrogen Fixation in Azotobacter vinelandii □ †. 193, 4477–4486. doi:10.1128/JB.05099-11.

Han, C. L., Chien, C. W., Chen, W. C., Chen, Y. R., Wu, C. P., Li, H., et al. (2008). A multiplexed quantitative strategy for membrane proteomics: Opportunities for mining therapeutic targets for autosomal dominant polycystic kidney disease. *Mol. Cell. Proteomics* 7, 1983–1997. doi:10.1074/mcp.M800068-MCP200.

Han, X., Jin, M., Breuker, K., and McLafferty, F. W. (2006). Extending Top-Down Mass Spectrometry to Proteins with Masses Greater Than 200 Kilodaltons. *Science (80-.).* 314, 109–113. doi:DOI: 10.1126/science.1128868.

Haneline, S., Connelly, C. J., and Melton, T. (1991). Chemotactic Behavior of Azotobacter vinelandii. *Appl. Environ. Microbiol.* 57, 825–829. doi:DOI: 10.1128/aem.57.3.825-829.1991.

Hanemaaijer, M., Olivier, B. G., Röling, W. F. M., Bruggeman, F. J., and Teusink, B. (2017). Model-based quantification of metabolic interactions from dynamic microbial-community data. *PLoS One* 12, 1–19. doi:10.1371/journal.pone.0173183.

Hanson, A. J., Guho, N. M., Paszczynski, A. J., and Coats, E. R. (2016). Community proteomics provides functional insight into polyhydroxyalkanoate production by a mixed microbial culture cultivated on fermented dairy manure. *Appl. Microbiol. Biotechnol.* 100, 7957–7976. doi:10.1007/s00253-016-7576-7.

Hays, S. G., and Ducat, D. C. (2015). Engineering cyanobacteria as photosynthetic feedstock factories. *Photosynth. Res.* 123, 285–295. doi:10.1007/s11120-014-9980-0.

Hays, S. G., Patrick, W. G., Ziesack, M., Oxman, N., and Silver, P. A. (2015). Better together: Engineering and application of microbial symbioses. *Curr. Opin. Biotechnol.* 36. doi:10.1016/j.copbio.2015.08.008.

Hays, S. G., Yan, L. L. W., Silver, P. A., and Ducat, D. C. (2017). Synthetic photosynthetic consortia define interactions leading to robustness and photoproduction. *J. Biol. Eng.* 11. doi:10.1186/s13036-017-0048-5.

Helliwell, K. E., Pandhal, J., Cooper, M. B., Longworth, J., Kudahl, U. J., Russo, D. A., et al. (2018). Quantitative proteomics of a B12-dependent alga grown in coculture with bacteria reveals metabolic tradeoffs required for mutualism. *New Phytol.* 217, 599–612. doi:10.1111/nph.14832.

Herrero, A., Muro-Pastor, A. M., and Flores, E. (2001). Nitrogen control in cyanobacteria. *J. Bacteriol.* 183, 411–425. doi:10.1128/JB.183.2.411-425.2001.

Heyer, R., Schallert, K., Büdel, A., Zoun, R., Dorl, S., Behne, A., et al. (2019). A Robust and Universal Metaproteomics Workflow for Research Studies and Routine Diagnostics Within 24 h Using Phenol Extraction, FASP Digest, and the MetaProteomeAnalyzer. *Front. Microbiol.* 10, 1883. doi:10.3389/fmicb.2019.01883.

Hickman, J. W., Kotovic, K. M., Miller, C., Warrener, P., Kaiser, B., Jurista, T., et al. (2013). Glycogen synthesis is a required component of the nitrogen stress response in Synechococcus elongatus PCC 7942. *Algal Res.* 2, 98–106. doi:10.1016/j.algal.2013.01.008.

Hirokawa, Y., Matsuo, S., Hamada, H., Matsuda, F., and Hanai, T. (2017). Metabolic engineering of Synechococcus elongatus PCC 7942 for improvement of 1,3-propanediol and glycerol production based on in silico simulation of metabolic flux distribution. *Microb. Cell Fact.* 16, 1–12. doi:10.1186/s12934-017-0824-4.

Hitchcock, A., Jackson, P. J., Chidgey, J. W., Dickman, M. J., Hunter, C. N., and

Canniffe, D. P. (2016). Biosynthesis of Chlorophyll a in a Purple Bacterial Phototroph and Assembly into a Plant Chlorophyll-Protein Complex. *ACS Synth. Biol.* 5, 948–954. doi:10.1021/acssynbio.6b00069.

Huang, H., Ren, Z., Gao, X., Hu, X., Zhou, Y., Jiang, J., et al. (2020). Integrated analysis of microbiome and host transcriptome reveals correlations between gut microbiota and clinical outcomes in HBV-related hepatocellular carcinoma. *Genome Med.* 12, 1–14. doi:10.1186/s13073-020-00796-5.

Jagtap, P., Bandhakavi, S., Higgins, L., Mcgowan, T., Sa, R., Stone, M. D., et al. (2012). Workflow for analysis of high mass accuracy salivary data set using MaxQuant and ProteinPilot search algorithm. *Proteomics* 12, 1726–1730. doi:10.1002/pmic.201100097.Workflow.

Jagtap, P. D., Blakely, A., Murray, K., Stewart, S., Kooren, J., Johnson, J. E., et al. (2015). Metaproteomic analysis using the Galaxy framework. *Proteomics* 15, 3553–3565. doi:10.1002/pmic.201500074.

Jeong, H., Park, J., and Kim, H. (2013). Determination of NH+ in environmental water with interfering substances using the modified nessler method. *J. Chem.* 2013. doi:10.1155/2013/359217.

Jiang, G., Hill, D. J., Kowalczuk, M., Johnston, B., Adamus, G., Irorere, V., et al. (2016). Carbon sources for polyhydroxyalkanoates and an integrated biorefinery. *Int. J. Mol. Sci.* 17. doi:10.3390/ijms17071157.

Jiang, L. L., Zhou, J. J., Quan, C. S., and Xiu, Z. L. (2017). Advances in industrial microbiome based on microbial consortium for biorefinery. *Bioresour. Bioprocess.* 4, 11. doi:10.1186/s40643-017-0141-0.

Johns, N. I., Blazejewski, T., Gomes, A. L. C., and Wang, H. H. (2016). Principles for designing synthetic microbial communities. *Curr. Opin. Microbiol.* 31, 146–153. doi:10.1016/j.mib.2016.03.010.

Johnson, J. S., Spakowicz, D. J., Hong, B. Y., Petersen, L. M., Demkowicz, P., Chen, L., et al. (2019). Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat. Commun.* 10, 1–11. doi:10.1038/s41467-019-

13036-1.

Johnson, K., Kleerebezem, R., and van Loosdrecht, M. C. M. (2010). Influence of the C/N ratio on the performance of polyhydroxybutyrate (PHB) producing sequencing batch reactors at short SRTs. *Water Res.* 44, 2141–2152. doi:10.1016/j.watres.2009.12.031.

Kabat, A. M., Srinivasan, N., and Maloy, K. J. (2014). Modulation of immune development and function by intestinal microbiota. *Trends Immunol.* 35, 507–517. doi:10.1016/j.it.2014.07.010.

Kalnenieks, U., Balodite, E., and Rutkis, R. (2019). Metabolic Engineering of Bacterial Respiration : High vs . Low P/O and the Case of Zymomonas mobilis. *Front. Bioeng. Biotechnol.* 7, 1–11. doi:10.3389/fbioe.2019.00327.

Kazamia, E., Smith, A. G., Nguyen, T. T. Van, Hodson, S. J., Sasso, S., Warren, M. J., et al. (2012). Mutualistic interactions between vitamin B12-dependent algae and heterotrophic bacteria exhibit regulation. *Environ. Microbiol.* 14, 1466–1476. doi:10.1111/j.1462-2920.2012.02733.x.

Kearney, S. M., Thomas, E., Coe, A., and Chisholm, S. W. (2021). Microbial diversity of co-occurring heterotrophs in cultures of marine picocyanobacteria. *Environ. Microbiomes* 16, 1–15. doi:10.1186/s40793-020-00370-x.

Kenny, D. J., and Balskus, E. P. (2018). Engineering chemical interactions in microbial communities. *Chem. Soc. Rev.* 47, 1705–1729. doi:10.1039/c7cs00664k.

Kera, K., Yoshizawa, Y., Shigehara, T., Nagayama, T., Tsujii, M., Tochigi, S., et al. (2020). Hik36–Hik43 and Rre6 act as a two-component regulatory system to control cell aggregation in Synechocystis sp. PCC6803. *Sci. Rep.* 10, 1–9. doi:10.1038/s41598-020-76264-2.

Keto-Timonen, R., Hietala, N., Palonen, E., Hakakorpi, A., Lindström, M., and Korkeala, H. (2016). Cold Shock Proteins: A Minireview with Special Emphasis on Csp-family of Enteropathogenic Yersinia. *Front. Microbiol.* 7, 1–7. doi:10.3389/fmicb.2016.01151.

Khasheii, B., Mahmoodi, P., and Mohammadzadeh, A. (2021). Siderophores: Importance in bacterial pathogenesis and applications in medicine and industry. *Microbiol. Res.* 250. doi:10.1016/j.micres.2021.126790.

Kim, H. J., Du, W., and Ismagilov, R. F. (2011). Complex function by design using spatially pre-structured synthetic microbial communities: Degradation of pentachlorophenol in the presence of Hg(ii). *Integr. Biol.* 3, 126–133. doi:10.1039/c0ib00019a.

Kleiner, M. (2019). Metaproteomics: Much More than Measuring Gene Expression in Microbial Communities. *mSystems* 4, e00115-19. doi:10.1128/msystems.00115-19.

Kleiner, M., Thorson, E., Sharp, C. E., Dong, X., Liu, D., Li, C., et al. (2017). Assessing species biomass contributions in microbial communities via metaproteomics. *Nat. Commun.* 8, 1558. doi:10.1038/s41467-017-01544-x.

Klotz, A., Reinhold, E., Doello, S., and Forchhammer, K. (2015). Nitrogen starvation acclimation in Synechococcus elongatus: Redox-control and the role of nitrate reduction as an electron sink. *Life* 5, 888–904. doi:10.3390/life5010888.

Kong, W., Meldgin, D. R., Collins, J. J., and Lu, T. (2018). Designing microbial consortia with defined social interactions. *Nat. Chem. Biol.* 14, 821–829. doi:10.1038/s41589-018-0091-7.

Koo, H., Mojib, N., Hakim, J. A., Hawes, I., Tanabe, Y., Andersen, D. T., et al. (2017). Microbial communities and their predicted metabolic functions in growth laminae of a unique large conical mat from Lake Untersee, East Antarctica. *Front. Microbiol.* 8, 1–15. doi:10.3389/fmicb.2017.01347.

Kouzuma, A., and Watanabe, K. (2014). Microbial ecology pushes frontiers in biotechnology. *Microbes Environ.* 29, 1–3. doi:10.1264/jsme2.me2901rh.

Kozlowski, L. P. (2017). Proteome-pI: Proteome isoelectric point database. *Nucleic Acids Res.* 45, D1112–D1116. doi:10.1093/nar/gkw978.

Kraepiel, A. M. L., Bellenger, J. P., Wichard, T., and Morel, F. M. M. (2009). Multiple roles of siderophores in free-living nitrogen-fixing bacteria. *BioMetals*

22, 573–581. doi:10.1007/s10534-009-9222-7.

Kumar, D., Yadav, A. K., and Dash, D. (2017). “Choosing an Optimal Database for Protein Identification from Tandem Mass Spectrometry Data,” in *Proteome Bioinformatics*, eds. S. Keerthikumar and S. Mathivanan (New York: Humana Press, New York, NY), 17–29. doi:10.1007/978-1-4939-6740-7_3.

Kusakabe, T., Tatsuke, T., Tsuruno, K., Hirokawa, Y., Atsumi, S., Liao, J. C., et al. (2013). Engineering a synthetic pathway in cyanobacteria for isopropanol production directly from carbon dioxide and light. *Metab. Eng.* 20, 101–108. doi:10.1016/j.ymben.2013.09.007.

Kuypers, M. M. M., Marchant, H. K., and Kartal, B. (2018). The microbial nitrogen-cycling network. *Nat. Rev. Microbiol.* 16, 263–276. doi:10.1038/nrmicro.2018.9.

La Sarre, B., McCully, A. L., Lennon, J. T., and McKinlay, J. B. (2017). Microbial mutualism dynamics governed by dose-dependent toxicity of cross-fed nutrients. *ISME J.* 11, 337–348. doi:10.1038/ismej.2016.141.

Lan, E. I., and Liao, J. C. (2011). Metabolic engineering of cyanobacteria for 1-butanol production from carbon dioxide. *Metab. Eng.* 13, 353–363. doi:10.1016/j.ymben.2011.04.004.

Landels, A., Evans, C., Noirel, J., and Wright, P. C. (2015). Advances in proteomics for production strain analysis. *Curr. Opin. Biotechnol.* 35, 111–117. doi:10.1016/j.copbio.2015.05.001.

Larentis, A. L., Nicolau, J. F. M. Q., Esteves, G. D. S., Vareschini, D. T., De Almeida, F. V. R., Dos Reis, M. G., et al. (2014). Evaluation of pre-induction temperature, cell growth at induction and IPTG concentration on the expression of a leptospiral protein in E. coli using shaking flasks and microbioreactor. *BMC Res. Notes* 7, 1–13. doi:10.1186/1756-0500-7-671.

Lesniak, J., Barton, W. A., and Nikolov, D. B. (2003). Structural and functional features of the Escherichia coli F1-ATPase. *Protein Sci.* 12, 2838–2843. doi:10.1023/A:1005519801891.

Li, Z., Adams, R. M., Chourey, K., Hurst, G. B., Hettich, R. L., and Pan, C. (2012).

Systematic comparison of label-free, metabolic labeling, and isobaric chemical labeling for quantitative proteomics on LTQ orbitrap velos. *J. Proteome Res.* 11, 1582–1590. doi:10.1021/pr200748h.

Liang, F., and Lindblad, P. (2017). Synechocystis PCC 6803 overexpressing RuBisCO grow faster with increased photosynthesis. *Metab. Eng. Commun.* 4, 29–36. doi:10.1016/j.meteno.2017.02.002.

Lindemann, S. R., Bernstein, H. C., Song, H. S., Fredrickson, J. K., Fields, M. W., Shou, W., et al. (2016). Engineering microbial consortia for controllable outputs. *ISME J.* 10, 2077–2084. doi:10.1038/ismej.2016.26.

Llácer, J. L., Espinosa, J., Castells, M. A., Contreras, A., Forchhammer, K., and Rubio, V. (2010). Structural basis for the regulation of NtcA-dependent transcription by proteins PipX and PII. *Proc. Natl. Acad. Sci. U. S. A.* 107, 15397–15402. doi:10.1073/pnas.1007015107.

Llufrio, E. M., Wang, L., Naser, F. J., and Patti, G. J. (2018). Sorting cells alters their redox state and cellular metabolome. *Redox Biol.* 16, 381–387. doi:10.1016/j.redox.2018.03.004.

Löwe, H., Hobmeier, K., Moos, M., Kremling, A., and Pflüger-Grau, K. (2017a). Photoautotrophic production of polyhydroxyalkanoates in a synthetic mixed culture of Synechococcus elongatus cscB and Pseudomonas putida cscAB. *Biotechnol. Biofuels* 10, 1–11. doi:10.1186/s13068-017-0875-0.

Löwe, H., Schmauder, L., Hobmeier, K., Kremling, A., and Pflüger-Grau, K. (2017b). Metabolic engineering to expand the substrate spectrum of Pseudomonas putida toward sucrose. *Microbiologyopen* 6, 1–9. doi:10.1002/mbo3.473.

Lundgren, D. H., Hwang, S. Il, Wu, L., and Han, D. K. (2010). Role of spectral counting in quantitative proteomics. *Expert Rev. Proteomics* 7, 39–53. doi:10.1586/epr.09.69.

Mærk, M., Jakobsen, Ø. M., Sletta, H., Klinkenberg, G., Tøndervik, A., Ellingsen, T. E., et al. (2020). Identification of Regulatory Genes and Metabolic Processes Important for Alginate Biosynthesis in Azotobacter vinelandii by Screening of a

Transposon Insertion Mutant Library. *Front. Bioeng. Biotechnol.* 7. doi:10.3389/fbioe.2019.00475.

Marzluf, G. A. (1997). Genetic regulation of nitrogen metabolism in the fungi. *Microbiol. Mol. Biol. Rev.* 61, 17–32. doi:10.1128/mmbr.61.1.17-32.1997.

Matlin, K. S. (2016). The Heuristic of Form : Mitochondrial Morphology and the Explanation of Oxidative Phosphorylation. *J. Hist. Biol.* 49, 37–94. doi:10.1007/s10739-015-9418-3.

Maynard, R. H., Premakumar, R., and Bishop, P. E. (1994). Mo-independent nitrogenase 3 is advantageous for diazotrophic growth of Azotobacter vinelandii on solid medium containing molybdenum. *J. Bacteriol.* 176, 5583–5586. doi:10.1128/jb.176.17.5583-5586.1994.

Mayneris-Perxachs, J., Castells-Nobau, A., Arnoriaga-Rodríguez, M., Garre-Olmo, J., Puig, J., Ramos, R., et al. (2022). Caudovirales bacteriophages are associated with improved executive function and memory in flies, mice, and humans. *Cell Host Microbe* 30, 340-356.e8. doi:10.1016/j.chom.2022.01.013.

McRose, D. L., Baars, O., Morel, F. M. M., and Kraepiel, A. M. L. (2017). Siderophore production in Azotobacter vinelandii in response to Fe-, Mo- and V-limitation. *Environ. Microbiol.* 19, 3595–3605. doi:10.1111/1462-2920.13857.

McRose, D. L., Seyedsayamdost, M. R., and Morel, F. M. M. (2018). Multiple siderophores: bug or feature? *J. Biol. Inorg. Chem.* 23, 983–993. doi:10.1007/s00775-018-1617-x.

Megger, D. A., Pott, L. L., Ahrens, M., Padden, J., Bracht, T., Kuhlmann, K., et al. (2014). Comparison of label-free and label-based strategies for proteome analysis of hepatoma cell lines. *Biochim. Biophys. Acta - Proteins Proteomics* 1844, 967–976. doi:10.1016/j.bbapap.2013.07.017.

Menhart, N., Thariath, A., and Viswanatha, T. (1991). Characterization of the pyoverdines of Azotobacter vinelandii ATCC 12837 with regard to heterogeneity. *Biol Met.* 4, 223–232. doi:DOI: 10.1007/BF01141185.

Merrick, M. J., and Edwards, R. A. (1995). Nitrogen Control in Bacteria M.

Microbiol. Rev. 59, 1–19. Available at: papers2://publication/uuid/4A82FEB2-4DDF-40CC-9027-0A895C5FDE85.

Minty, J. J., Singer, M. E., Scholz, S. A., Bae, C.-H., Ahn, J.-H., Foster, C. E., et al. (2013). Design and characterization of synthetic fungal-bacterial consortia for direct production of isobutanol from cellulosic biomass. *Proc. Natl. Acad. Sci.* 110, 14592–14597. doi:10.1073/pnas.1218447110.

Mok, P. S., Chuah, J. A., Najimudin, N., Liew, P. W. Y., Jong, B. C., and Sudesh, K. (2021). In vivo characterization and application of the pha synthase from azotobacter vinelandii for the biosynthesis of polyhydroxyalkanoate containing 4-hydroxybutyrate. *Polymers (Basel).* 13. doi:10.3390/polym13101576.

Moosavi, B., Berry, E. A., Lei, X., Wen, Z., Yang, C., and Fu, G. (2019). The assembly of succinate dehydrogenase : a key enzyme in bioenergetics. *Cell. Mol. Life Sci.* 76, 4023–4042. doi:10.1007/s00018-019-03200-7.

Nadler, W. M., Waidelich, D., Kerner, A., Hanke, S., Berg, R., Trumpp, A., et al. (2017). MALDI versus ESI: The Impact of the Ion Source on Peptide Identification. *J. Proteome Res.* 16, 1207–1215. doi:10.1021/acs.jproteome.6b00805.

Noar, J. D., and Bruno-Bárcena, J. M. (2018). Azotobacter vinelandii: The source of 100 years of discoveries and many more to come. *Microbiol. (United Kingdom)* 164, 421–436. doi:10.1099/mic.0.000643.

Noar, J., Loveless, T., Navarro-Herrero, J. L., Olson, J. W., and Bruno-Bárcena, J. M. (2015). Aerobic hydrogen production via nitrogenase in azotobacter vinelandii CA6. *Appl. Environ. Microbiol.* 81, 4507–4516. doi:10.1128/AEM.00679-15.

Nurbaş, M., and Kutsal, T. (2004). Production of PHB and P(HB-co-HV) biopolymers by using Alcaligenes Eutrophus. *Iran. Polym. J.* 13, 45–51. Available at: https://www.researchgate.net/publication/242744012.

Oda, Y., Huang, K., Cross, F. R., Cowburn, D., and Chait, B. T. (1999). Accurate quantitation of protein expression and site-specific phosphorylation. *Proc. Natl. Acad. Sci. U. S. A.* 96, 6591–6596. doi:10.1073/pnas.96.12.6591.

Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., et al. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* 1, 376–386. doi:10.1074/mcp.M200025-MCP200.

Ortiz-Marquez, J. C. F., Nascimento, M. Do, Dublan, M. de los A., and Curatti, L. (2012). Association with an ammonium-excreting bacterium allows diazotrophic culture of oil-rich eukaryotic microalgae. *Appl. Environ. Microbiol.* 78, 2345–2352. doi:10.1128/AEM.06260-11.

Pajarillo, E. A. B., Kim, S. H., Valeriano, V. D., Lee, J. Y., and Kang, D. K. (2017). Proteomic view of the crosstalk between Lactobacillus mucosae and intestinal epithelial cells in co-culture revealed by Q exactive-based quantitative proteomics. *Front. Microbiol.* 8. doi:10.3389/fmicb.2017.02459.

Palatsi, J., Illa, J., Prenafeta-Boldú, F. X., Laureni, M., Fernandez, B., Angelidaki, I., et al. (2010). Long-chain fatty acids inhibition and adaptation process in anaerobic thermophilic digestion: Batch tests, microbial community structure and mathematical modelling. *Bioresour. Technol.* 101, 2243–2251. doi:10.1016/j.biortech.2009.11.069.

Palmblad, M., Bindschedler, L. V, and Cramer, R. (2007). Quantitative proteomics using uniform 15 N-labeling , MASCOT , and the trans-proteomic pipeline. *Proteomics* 7, 3462–3469. doi:10.1002/pmic.200700180.

Pandeswari, P. B., and Sabareesh, V. (2019). Middle-down approach : a choice to sequence and characterize proteins / proteomes by mass spectrometry. *RSC Adv.* 9, 313–344. doi:10.1039/c8ra07200k.

Pandhal, J., and Noirel, J. (2014). Synthetic microbial ecosystems for biotechnology. *Biotechnol. Lett.* 36, 1141–1151. doi:10.1007/s10529-014-1480-y.

Park, Y., Je, K. W., Lee, K., Jung, S. E., and Choi, T. J. (2008). Growth promotion of Chlorella ellipsoidea by co-inoculation with Brevundimonas sp. isolated from the microalga. *Hydrobiologia* 598, 219–228. doi:10.1007/s10750-007-9152-8.

Patel, R. S., Roy, M., and Dutta, G. K. (2012). Mass spectrometry- a review. *Vet.*

World 5, 185–192. doi:10.5455/vetworld.2012.185-192.

Perni, S., Andrew, P. W., and Shama, G. (2005). Estimating the maximum growth rate from microbial growth curves: Definition is everything. *Food Microbiol.* 22, 491–495. doi:10.1016/j.fm.2004.11.014.

Pospíšil, P. (2016). Production of reactive oxygen species by photosystem II as a response to light and temperature stress. *Front. Plant Sci.* 7, 1–12. doi:10.3389/fpls.2016.01950.

Ramos-Fernández, A., López-Ferrer, D., and Vázquez, J. (2007). Improved Method for Differential Expression Proteomics Using Trypsin-catalyzed 18O Labeling with a Correction for Labeling Efficiency. *Mol. Cell. Proteomics 6.7* 6, 1274–86.

Riedlinger, J., Schrey, S. D., Tarkka, M. T., Hampp, R., Kapur, M., and Fiedler, H. P. (2006). Auxofuran, a novel metabolite that stimulates the growth of fly agaric, is produced by the mycorrhiza helper bacterium Streptomyces strain AcH 505. *Appl. Environ. Microbiol.* 72, 3550–3557. doi:10.1128/AEM.72.5.3550-3557.2006.

Rippka, R., Deruelles, J., Waterbury, J. B., and Stanier, R. Y. (1979). Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *J. Gen. Microbiol.* 111, 1–61. doi:10.1099/00221287-111-1-1.

Roell, G. W., Zha, J., Carr, R. R., Koffas, M. A., Fong, S. S., and Tang, Y. J. (2019). Engineering microbial consortia by division of labor. *Microb. Cell Fact.* 18, 35. doi:10.1186/s12934-019-1083-3.

Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., et al. (2004). Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* 3, 1154–1169. doi:10.1074/mcp.M400129-MCP200.

Rüger, M., Bensch, G., Tüngler, R., and Reichl, U. (2012). A flow cytometric method for viability assessment of Staphylococcus aureus and Burkholderia cepacia in mixed culture. *Cytom. Part A* 81 A, 1055–1066. doi:10.1002/cyto.a.22219.

Saliba, A. E., Westermann, A. J., Gorski, S. A., and Vogel, J. (2014). Single-cell

RNA-seq: Advances and future challenges. *Nucleic Acids Res.* 42, 8845–8860. doi:10.1093/nar/gku555.

Santos-Merino, M., Garcillán-Barcia, M. P., and De La Cruz, F. (2018). Engineering the fatty acid synthesis pathway in Synechococcus elongatus PCC 7942 improves omega-3 fatty acid production. *Biotechnol. Biofuels* 11, 1–13. doi:10.1186/s13068-018-1243-4.

Santos-Merino, M., Singh, A. K., and Ducat, D. C. (2019). New applications of synthetic biology tools for cyanobacterial metabolic engineering. *Front. Bioeng. Biotechnol.* 7. doi:10.3389/fbioe.2019.00033.

Sardiu, M. E., and Washburn, M. P. (2011). Building Protein-Protein Interaction Networks with Proteomics and Informatics Tools. *J. Biol. Chem.* 286, 23645–23651. doi:10.1074/jbc.R110.174052.

Sayyed, R., Wani, S., Alarfaj, A., Syed, A., and El-Enshasy, H. (2020). Production, purification and evaluation of biodegradation potential of PHB depolymerase of Stenotrophomonas sp. RZS7. *PLoS One* 15, e0220095.

Schlembach, I., Grünberger, A., Rosenbaum, M. A., and Regestein, L. (2021). Measurement Techniques to Resolve and Control Population Dynamics of Mixed-Culture Processes. *Trends Biotechnol.* 39, 1093–1109. doi:10.1016/j.tibtech.2021.01.006.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., et al. (2011). Global quantification of mammalian gene expression control. *Nature* 473, 337–342. doi:10.1038/nature10098.

Sedlacek, C. J., Nielsen, S., Greis, K. D., Haffey, W. D., Revsbech, N. P., Ticak, T., et al. (2016). Effects of bacterial community members on the proteome of the ammonia-oxidizing bacterium Nitrosomonas sp. strain Is79. *Appl. Environ. Microbiol.* 82, 4776–4788. doi:10.1128/AEM.01171-16.

Setubal, J. C., Dos Santos, P., Goldman, B. S., Ertesvåg, H., Espin, G., Rubio, L. M., et al. (2009). Genome sequence of Azotobacter vinelandii, an obligate aerobe specialized to support diverse anaerobic metabolic processes. *J. Bacteriol.* 191,

4534–4545. doi:10.1128/JB.00504-09.

Shah, A. D., Goode, R. J. A., Huang, C., Powell, D. R., and Schittenhelm, R. B. (2020). Lfq-Analyst: An easy-To-use interactive web platform to analyze and visualize label-free proteomics data preprocessed with maxquant. *J. Proteome Res.* 19, 204–211. doi:10.1021/acs.jproteome.9b00496.

Shestakov, S. V, and Ti, A. I. Y. E. (1970). Evidence for Genetic Transformation in Blue-Green Alga Anacystis nidulans. 375, 372–375. doi:DOI: 10.1007/BF00441199.

Shi, M., Li, M., Yang, A., Miao, X., Yang, L., Pandhal, J., et al. (2022). Class I Polyhydroxyalkanoate (PHA) Synthase Increased Polylactic Acid Production in Engineered Escherichia Coli. *Front. Bioeng. Biotechnol.* 10, 1–7. doi:10.3389/fbioe.2022.919969.

Shimakawa, G., and Miyake, C. (2018). Oxidation of P700 Ensures Robust Photosynthesis. *Front. Plant Sci.* 9, 1–15. doi:10.3389/fpls.2018.01617.

Sieber, J. R., Crable, B. R., Sheik, C. S., Hurst, G. B., Rohlin, L., Gunsalus, R. P., et al. (2015). Proteomic analysis reveals metabolic and regulatory systems involved in the syntrophic and axenic lifestyle of Syntrophomonas wolfei. *Front. Microbiol.* 6, 1–9. doi:10.3389/fmicb.2015.00115.

Siuti, N., and Kelleher, N. L. (2008). Decoding protein modifications using top-down mass spectrometry. *Nat. Methods* 4, 817–821. doi:doi: 10.1038/nmeth1097.

Sivasakthi, S., Saranraj, P., and Sivasakthivelan, P. (2017). Biological Nitrogen Fixation by Azotobacter sp.-A Review. *Indo-Asian J. Multidiscip. Res.* 3, 1274–1284. doi:10.22192/iajmr.2017.3.5.6.

Smith, M. J., and Francis, M. B. (2016a). A Designed A. vinelandii-S. elongatus Coculture for Chemical Photoproduction from Air, Water, Phosphate, and Trace Metals. *ACS Synth. Biol.* 5, 955–961. doi:10.1021/acssynbio.6b00107.

Smith, M. J., and Francis, M. B. (2016b). A Designed A. vinelandii-S. elongatus Coculture for Chemical Photoproduction from Air, Water, Phosphate, and Trace Metals (supplements). *ACS Synth. Biol.* 5, 955–961.

doi:10.1021/acssynbio.6b00107.

So, A. K. C., and Espie, G. S. (1998). Cloning, characterization and expression of carbonic anhydrase from the cyanobacterium Synechocystis PCC6803. *Plant Mol. Biol.* 37, 205–215. doi:10.1023/A:1005959200390.

Sun, R., Sun, P., Zhang, J., Esquivel-Elizondo, S., and Wu, Y. (2018). Microorganisms-based methods for harmful algal blooms control: A review. *Bioresour. Technol.* 248, 12–20. doi:10.1016/j.biortech.2017.07.175.

Tan, J. P., Jahim, J. M., Wu, T. Y., Harun, S., and Mumtaz, T. (2016). Use of corn steep liquor as an economical nitrogen source for biosuccinic acid production by Actinobacillus succinogenes. in *IOP Conference Series: Earth and Environmental Science* (Institute of Physics Publishing), 012058. doi:10.1088/1755-1315/36/1/012058.

Tan, S., Hwee, T. T., and Chung, M. C. M. (2008). Membrane proteins and membrane proteomics. *Proteomics* 8, 3924–3932. doi:10.1002/pmic.200800597.

Therien, J. B., Zadvornyy, O. A., Posewitz, M. C., Bryant, D. A., and Peters, J. W. (2014). Growth of Chlamydomonas reinhardtii in acetate-free medium when co-cultured with. *Biotechnol. Biofuels* 7, 154.

Thøgersen, M. S., Melchiorsen, J., Ingham, C., and Gram, L. (2018). A Novel Microbial Culture Chamber Co-cultivation System to Study Algal-Bacteria Interactions Using Emiliania huxleyi and Phaeobacter inhibens as Model Organisms. *Front. Microbiol.* 9. doi:10.3389/fmicb.2018.01705.

Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., et al. (2003). Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* 75, 1895–1904. doi:10.1021/ac0262560.

Timp, W., and Timp, G. (2020). Beyond mass spectrometry , the next step in proteomics. *Sci. Adv.* 6, eaax8978. doi:DOI: 10.1126/sciadv.aax8978.

Ting, C. S., Rocap, G., King, J., and Chisholm, S. W. (2002). Cyanobacterial photosynthesis in the oceans: The origins and significance of divergent light-

harvesting strategies. *Trends Microbiol.* 10, 134–142. doi:10.1016/S0966-842X(02)02319-3.

Trabelsi, D., Ben Ammar, H., Mengoni, A., and Mhamdi, R. (2012). Appraisal of the crop-rotation effect of rhizobial inoculation on potato cropping systems in relation to soil bacterial communities. *Soil Biol. Biochem.* 54, 1–6. doi:10.1016/j.soilbio.2012.05.013.

Tshikantwa, T. S., Ullah, M. W., He, F., and Yang, G. (2018). Current trends and potential applications of microbial interactions for human welfare. *Front. Microbiol.* 9. doi:10.3389/fmicb.2018.01156.

Tu, Z., Liu, L., Lin, W., Xie, Z., and Luo, J. (2018). Potential of using sodium bicarbonate as external carbon source to cultivate microalga in non-sterile condition. *Bioresour. Technol.* 266, 109–115. doi:10.1016/j.biortech.2018.06.076.

Tyanova, S., Temu, T., and Cox, J. (2016). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* 11, 2301–2319. doi:10.1038/nprot.2016.136.

Unnithan, V. V., Unc, A., and Smith, G. B. (2014). Mini-review: A priori considerations for bacteria-algae interactions in algal biofuel systems receiving municipal wastewaters. *Algal Res.* 4, 35–40. doi:10.1016/j.algal.2013.11.009.

Välikangas, T., Suomi, T., and Elo, L. L. (2018). A systematic evaluation of normalization methods in quantitative label-free proteomics. *Brief. Bioinform.* 19, 1–11. doi:10.1093/bib/bbw095.

Valladares, A., Montesinos, M. L., Herrero, A., and Flores, E. (2002). An ABC-type, high-affinity urea permease identified in cyanobacteria. *Mol. Microbiol.* 43, 703–715. doi:10.1046/j.1365-2958.2002.02778.x.

Vázquez-Bermüdez, M. F., Paz-Yepes, J., Herrero, A., and Flores, E. (2002). The NtcA-activated amt1 gene encodes a permease required for uptake of low concentrations of ammonium in the cyanobacterium Synechococcus sp. PCC 7942. *Microbiology* 148, 861–869. doi:10.1099/00221287-148-3-861.

Verberkmoes, N. C., Russell, A. L., Shah, M., Godzik, A., Rosenquist, M., Halfvarson, J., et al. (2009). Shotgun metaproteomics of the human distal gut microbiota. *ISME J.* 3, 179–189. doi:10.1038/ismej.2008.108.

Villa, J. A., Ray, E. E., and Barney, B. M. (2014). Azotobacter vinelandii siderophore can provide nitrogen to support the culture of the green algae neochloris oleoabundans and scenedesmus sp. BA032. *FEMS Microbiol. Lett.* 351, 70–77. doi:10.1111/1574-6968.12347.

Vincent, W. F. (2009). "Cyanobacteria," in *Encyclopedia of Inland Waters*, ed. G. E. Likens (Academic Press), 226–232.

Völkel, P., Faou, P. Le, and Angrand, P. (2010). Interaction proteomics : characterization of protein complexes using tandem affinity purification – mass spectrometry. *Biochem. Soc. Trans.* 38, 883–887. doi:10.1042/BST0380883.

Waghmode, T. R., Kurade, M. B., Khandare, R. V, and Govindwar, S. P. (2011). International Biodeterioration & Biodegradation A sequential aerobic / microaerophilic decolorization of sulfonated mono azo dye Golden Yellow HER by microbial consortium GG-BL. *Int. Biodeterior. Biodegradation* 65, 1024–1034. doi:10.1016/j.ibiod.2011.08.002.

Wan, C., Zhao, X., Guo, S., Alam, A., and Bai, F. (2013). Bioresource Technology Bioflocculant production from Solibacillus silvestris W01 and its application in cost-effective harvest of marine microalga Nannochloropsis oceanica by flocculation. *Bioresour. Technol.* 135, 207–212. doi:10.1016/j.biortech.2012.10.004.

Wan Razali, W. A., Evans, C. A., and Pandhal, J. (2022). Comparative Proteomics Reveals Evidence of Enhanced EPA Trafficking in a Mutant Strain of Nannochloropsis oculata. *Front. Bioeng. Biotechnol.* 10. doi:10.3389/fbioe.2022.838445.

Warwood, S., Byron, A., Humphries, M. J., and Knight, D. (2013). The effect of peptide adsorption on signal linearity and a simple approach to improve reliability of quantification. *J. Proteomics* 85, 160–164. doi:10.1016/j.jprot.2013.04.034.

Wegener, K. M., Singh, A. K., Jacobs, J. M., Elvitigala, T., Welsh, E. A., Keren, N., et al. (2010). Global proteomics reveal an atypical strategy for carbon/nitrogen assimilation by a cyanobacterium under diverse environmental perturbations. *Mol. Cell. Proteomics* 9, 2678–2689. doi:10.1074/mcp.M110.000109.

Wehbi, H., Portillo, E., Harvey, H., Shimkoff, A. E., Scheurwater, E. M., Howell, P. L., et al. (2011). The peptidoglycan-binding protein fimv promotes assembly of the pseudomonas aeruginosa type IV pilus secretin. *J. Bacteriol.* 193, 540–550. doi:10.1128/JB.01048-10.

Wei, S., Bian, Y., Zhao, Q., Chen, S., Mao, J., Song, C., et al. (2017). Salinity-Induced Palmella Formation Mechanism in Halotolerant Algae Dunaliella salina Revealed by Quantitative Proteomics and Phosphoproteomics. *Front. Plant Sci.* 8, 810. doi:10.3389/fpls.2017.00810.

Weiss, T. L., Young, E. J., and Ducat, D. C. (2017). A synthetic, light-driven consortium of cyanobacteria and heterotrophic bacteria enables stable polyhydroxybutyrate production. *Metab. Eng.* 44, 236–245. doi:10.1016/j.ymben.2017.10.009.

Wichard, T., Bellenger, J.-P., Morel, F. M. M., and Kraepiel, A. M. L. (2009). Role of the siderophore azotobactin in the bacterial acquisition of nitrogenase metal cofactors. *Environ. Sci. Technol.* 43, 7218–7224. doi:10.1021/es8037214.

Wilkins, M. (2009). Proteomics data mining. *Expert Rev. Proteomics* 6, 599–603.

Wisén, S., Bergman, B., and Mannervik, B. (2004). Mutagenesis of the cysteine residues in the transcription factor NtcA from Anabaena PCC 7120 and its effects on DNA binding in vitro. *Biochim. Biophys. Acta - Gene Struct. Expr.* 1679, 156–163. doi:10.1016/j.bbaexp.2004.06.003.

Wiśniewski, J. R., Hein, M. Y., Cox, J., and Mann, M. (2014). A “proteomic ruler” for protein copy number and concentration estimation without spike-in standards. *Mol. Cell. Proteomics* 13, 3497–3506. doi:10.1074/mcp.M113.037309.

Wiśniewski, J. R., and Mann, M. (2016). A proteomics approach to the protein normalization problem: Selection of unvarying proteins for MS-based

proteomics and western blotting. *J. Proteome Res.* 15, 2321–2326. doi:10.1021/acs.jproteome.6b00403.

Wu, C., Tran, J. C., Zamdborg, L., Durbin, K. R., Li, M., Dorothy, R., et al. (2013). A Protease for Middle Down Proteomics Cong. *Nat. Methods* 9, 822–824. doi:10.1038/nmeth.2074.A.

Xu, L., Cheng, X., and Wang, Q. (2018). Enhanced Lipid Production in Chlamydomonas reinhardtii by Co-culturing With Azotobacter chroococcum. *Front. Plant Sci.* 9, 741. doi:10.3389/fpls.2018.00741.

Yan, W., and Chen, S. S. (2005). Mass spectrometry-based quantitative proteomic profiling. *Briefings Funct. Genomics Proteomics* 4, 27–38.

Yan, Y., Yang, J., Dou, Y., Chen, M., Ping, S., Peng, J., et al. (2008). Nitrogen fixation island and rhizosphere competence traits in the genome of root-associated Pseudomonas stutzeri A1501. *PNAS* 105, 7564–7569. doi:doi.org/10.1073/pnas.0801093105.

Ye, C., Zou, W., Xu, N., and Liu, L. (2014). Metabolic model reconstruction and analysis of an artificial microbial ecosystem for vitamin C production. *J. Biotechnol.* 182–183, 61–67. doi:10.1016/j.jbiotec.2014.04.027.

Yen, H.-W., Hu, I.-C., Chen, C.-Y., and Chang, J.-S. (2014). "Design of Photobioreactors for Algal Cultivation," in *Biofuels from Algae*, eds. A. Pandey, D.-J. Lee, Y. Chisti, and C. R. Soccol (Amsterdam: Elsevier), 23–45. doi:https://doi.org/10.1016/B978-0-444-59558-4.00002-4.

Yimer, S. A., Birhanu, A. G., Kalayou, S., Riaz, T., Zegeye, E. D., Beyene, G. T., et al. (2017). Comparative proteomic analysis of Mycobacterium tuberculosis lineage 7 and lineage 4 strains reveals differentially abundant proteins linked to slow growth and virulence. *Front. Microbiol.* 8, 1–14. doi:10.3389/fmicb.2017.00795.

Yoneyama, F., Yamamoto, M., Hashimoto, W., and Murata, K. (2011). Azotobacter vinelandii gene clusters for two types of peptidic and catechol siderophores produced in response to molybdenum. *J. Appl. Microbiol.* 111, 932–938.

doi:10.1111/j.1365-2672.2011.05109.x.

Yoneyama, F., Yamamoto, M., Hashimoto, W., and Murata, K. (2015). Production of polyhydroxybutyrate and alginate from glycerol by azotobacter vinelandii under nitrogen-free conditions. *Bioengineered* 6, 209–217. doi:10.1080/21655979.2015.1040209.

Yoshida, N., Takase, R., Sugahara, Y., Nambu, Y., and Hashimoto, W. (2022). Direct production of polyhydroxybutyrate and alginate from crude glycerol by Azotobacter vinelandii using atmospheric nitrogen. *Sci. Rep.* 12, 8032. doi:10.1038/s41598-022-11728-1.

Zehr, J. P., Jenkins, B. D., Short, S. M., and Steward, G. F. (2003). Nitrogenase gene diversity and microbial community structure: a cross-system comparison. *Environ. Microbiol.* 5, 539–554. Available at: http://pfam.wustl.edu.

Zhang, A., Carroll, A. L., and Atsumi, S. (2017). Carbon recycling by cyanobacteria: Improving CO2 fixation through chemical production. *FEMS Microbiol. Lett.* 364, 1–7. doi:10.1093/femsle/fnx165.

Zhang, Y., Fonslow, B. R., Shan, B., Baek, M., and Yates, J. R. (2013). Protein Analysis by Shotgun / Bottom-up Proteomics. *Chem. Rev.* 113, 2343−2394.

Zhou, K., Qiao, K., Edgar, S., and Stephanopoulos, G. (2015). Distributing a metabolic pathway among a microbial consortium enhances production of natural products. *Nat. Biotechnol.* 33, 377–383. doi:10.1038/nbt.3095.

Zhou, W., Colpa, D. I., Geurkink, B., Euverink, G. J. W., and Krooneman, J. (2022). The impact of carbon to nitrogen ratios and pH on the microbial prevalence and polyhydroxybutyrate production levels using a mixed microbial starter culture. *Sci. Total Environ.* 811, 152341. doi:10.1016/j.scitotenv.2021.152341.

Zhu, W., Smith, J. W., and Huang, C. M. (2010). Mass spectrometry-based label-free quantitative proteomics. *J. Biomed. Biotechnol.* 2010. doi:10.1155/2010/840518.

Zschiedrich, C. P., Keidel, V., and Szurmant, H. (2016). Molecular Mechanisms of Two-Component Signal Transduction. *J. Mol. Biol.* 428, 3752–3775. doi:10.1016/j.jmb.2016.08.003.

Zubarev, R. A. (2013). The challenge of the proteome dynamic range and its implications for in-depth proteomics. *Proteomics* 13, 723–726. doi:10.1002/pmic.201200451.

Zuroff, T. R., Xiques, S. B., and Curtis, W. R. (2013). Consortia-mediated bioprocessing of cellulose to ethanol with a symbiotic Clostridium phytofermentans/yeast co-culture. *Biotechnol. Biofuels* 6, 1–12. doi:10.1186/1754-6834-6-59.

Appendix A: Media recipes and supplements for Chapter 2

Appendix Table A1: BG11 medium

Components	Chemical formula	g/200ml dH_2O	Volume (ml)
Sodium nitrate	$NaNO_3$	5	60
Dipotassium phosphate trihydrate	$K_2HPO_4 \cdot 3H_2O$	0.8	10
Magnesium sulfate heptahydrate	$MgSO_4 \cdot 7H_2O$	1.5	10
Calcium chloride dihydrate	$CaCl_2 \cdot 2H_2O$	0.72	10
Citric acid		0.12	10
Ammonium ferric citrate green		0.12	10
Disodium ethylenediaminetetraacetate dihydrate	$EDTANa_2 \cdot 2H_2O$	0.22	10
Sodium carbonate	Na_2CO_3	0.4	10
Trace element solution			1

Make up to 1L solution with distilled water. Autoclave.

Appendix Table A2: BG11 trace element solution

Components	Chemical formula	g/200ml dH_2O
Boric acid	H_3BO_3	0.572
Manganese chloride tetrahydrate	$MnCl_2 \cdot 4H_2O$	0.362
Zinc sulfate heptahydrate	$ZnSO_4 \cdot 7H_2O$	0.044
Sodium molybdate dihydrate	$Na_2MoO_4 \cdot 2H_2O$	0.078

Copper (II) sulfate pentahydrate	$CuSO_4 \cdot 5H_2O$	0.016
Cobalt (II) nitrate hexahydrate	$Co(NO_3)_2 \cdot 6H_2O$	0.01

Make up to 200 mL solution with distilled water. Sterile with 0.22 µ filter.

Appendix Table A3: Burk's medium

Components	Volume (mL)
Phosphate buffer	900
10× salts solution	100

Make up to 1 L solution. Autoclave and add 1 mL sterile 5g/L $FeSO_4$ solution.

Appendix Table A4: Burk's phosphate buffer:

Components	Chemical formula	Amount (g)
Monobasic potassium phosphate	KH_2PO_4	0.2
Dibasic potassium phosphate	K_2HPO_4	0.8

Make up to 900 mL solution with distilled water.

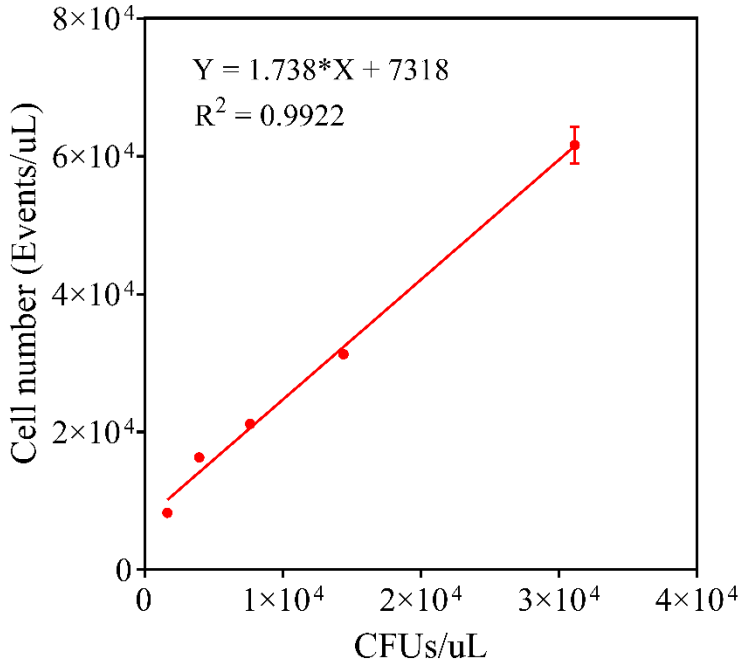
Appendix Table A5: Burk's 10× salts solution:

Components	Chemical formula	Amount (g)
sucrose		200
Magnesium sulfate heptahydrate	$MgSO_4 \cdot 7H_2O$	2
Calcium chloride dihydrate	$CaCl_2 \cdot 2H_2O$	0.9

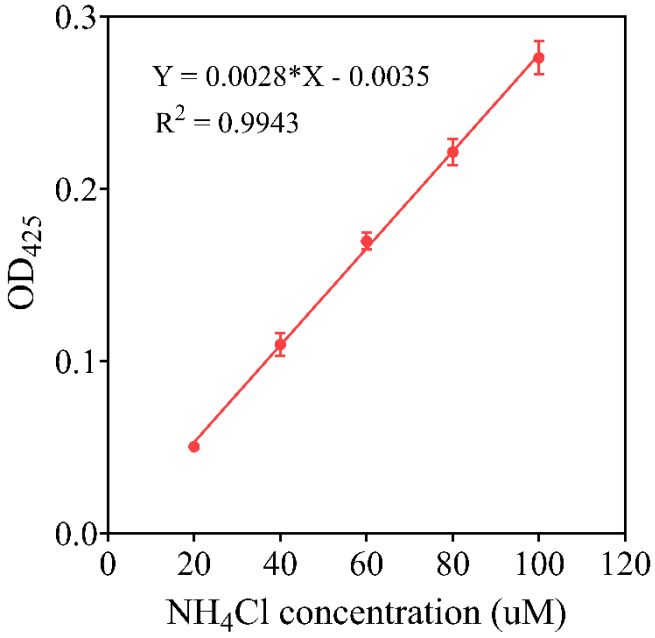
Sodium molybdate dihydrate	$Na_2MoO_4 \cdot 2H_2O$	0.0026

Make up to 1 L solution with distilled water.

Appendix B: Supplementary material for Chapter 3



Appendix Figure B1: Standard curve of relationship between cell number and CFUs in *A. vinelandii* ΔnifL. Cell number was counted by flow cytometer. CFUs was counted on Burk's agar plate. All data shown were repeated in three replicates. Error bars shown in SD. $P < 0.0001$



Appendix Figure B2: Ammonium standard curve using NH_4Cl as standard. All data shown was repeated in three replicates. Error bars shown in SD. $P < 0.0001$.

Appendix C: Supplementary material for Chapter 4

Appendix Table C1: Identified KEGG pathways of *S. elongatus* cscB/SPS

KEGG IDs	KEGG pathways	Matched proteins
syf01100	Metabolic pathways	423
syf01110	Biosynthesis of secondary metabolites	208
syf01120	Microbial metabolism in diverse environments	94
syf01240	Biosynthesis of cofactors	85
syf01230	Biosynthesis of amino acids	84
syf01200	Carbon metabolism	56
syf03010	Ribosome	49
syf02010	ABC transporters	45
syf00195	Photosynthesis	38
syf02020	Two-component system	35
syf00230	Purine metabolism	34
syf00190	Oxidative phosphorylation	28
syf00860	Porphyrin and chlorophyll metabolism	27
syf00970	Aminoacyl-tRNA biosynthesis	25
syf00270	Cysteine and methionine metabolism	24
syf00010	Glycolysis / Gluconeogenesis	24
syf02024	Quorum sensing	22
syf00620	Pyruvate metabolism	21
syf00520	Amino sugar and nucleotide sugar metabolism	21

syf00240	Pyrimidine metabolism	21
syf01210	2-Oxocarboxylic acid metabolism	20
syf00250	Alanine, aspartate and glutamate metabolism	18
syf00400	Phenylalanine, tyrosine and tryptophan biosynthesis	18
syf00630	Glyoxylate and dicarboxylate metabolism	18
syf00260	Glycine, serine and threonine metabolism	18
syf00920	Sulfur metabolism	17
syf00196	Photosynthesis - antenna proteins	16
syf00030	Pentose phosphate pathway	16
syf00480	Glutathione metabolism	15
syf00710	Carbon fixation in photosynthetic organisms	15
syf00910	Nitrogen metabolism	15
syf00680	Methane metabolism	15
syf00500	Starch and sucrose metabolism	15
syf00330	Arginine and proline metabolism	13
syf00550	Peptidoglycan biosynthesis	13
syf03430	Mismatch repair	13
syf00220	Arginine biosynthesis	12
syf00900	Terpenoid backbone biosynthesis	12
syf03440	Homologous recombination	12
syf03018	RNA degradation	12
syf00790	Folate biosynthesis	11
syf00300	Lysine biosynthesis	11

syf00760	Nicotinate and nicotinamide metabolism	11
syf00290	Valine, leucine and isoleucine biosynthesis	11
syf00770	Pantothenate and CoA biosynthesis	11
syf00670	One carbon pool by folate	10
syf03070	Bacterial secretion system	10
syf01212	Fatty acid metabolism	10
syf00051	Fructose and mannose metabolism	10
syf00061	Fatty acid biosynthesis	10
syf00020	Citrate cycle (TCA cycle)	10
syf00540	Lipopolysaccharide biosynthesis	10
syf00340	Histidine metabolism	10
syf03030	DNA replication	10
syf00541	O-Antigen nucleotide sugar biosynthesis	10
syf03410	Base excision repair	9
syf03060	Protein export	9
syf00521	Streptomycin biosynthesis	9
syf00640	Propanoate metabolism	8
syf00130	Ubiquinone and other terpenoid-quinone biosynthesis	8
syf04122	Sulfur relay system	7
syf03420	Nucleotide excision repair	7
syf00730	Thiamine metabolism	7
syf00780	Biotin metabolism	7

syf00460	Cyanoamino acid metabolism	5
syf00380	Tryptophan metabolism	5
syf00450	Selenocompound metabolism	5
syf03020	RNA polymerase	5
syf00561	Glycerolipid metabolism	5
syf00660	C5-Branched dibasic acid metabolism	5
syf00261	Monobactam biosynthesis	5
syf00906	Carotenoid biosynthesis	5
syf01503	Cationic antimicrobial peptide (CAMP) resistance	5
syf01502	Vancomycin resistance	4
syf00523	Polyketide sugar unit biosynthesis	4
syf00280	Valine, leucine and isoleucine degradation	4
syf00052	Galactose metabolism	4
syf00650	Butanoate metabolism	4
syf00740	Riboflavin metabolism	4
syf00360	Phenylalanine metabolism	4
syf00350	Tyrosine metabolism	4
syf00750	Vitamin B6 metabolism	3
syf00562	Inositol phosphate metabolism	3
syf00332	Carbapenem biosynthesis	3
syf00471	D-Glutamine and D-glutamate metabolism	3
syf00071	Fatty acid degradation	3
syf01220	Degradation of aromatic compounds	2

syf00623	Toluene degradation	2
syf00627	Aminobenzoate degradation	2
syf00310	Lysine degradation	2
syf01501	beta-Lactam resistance	2
syf00430	Taurine and hypotaurine metabolism	2
syf00401	Novobiocin biosynthesis	2
syf00625	Chloroalkane and chloroalkene degradation	2
syf00511	Other glycan degradation	2
syf00410	beta-Alanine metabolism	2
syf00361	Chlorocyclohexane and chlorobenzene degradation	2
syf00364	Fluorobenzoate degradation	2
syf00525	Acarbose and validamycin biosynthesis	2
syf00564	Glycerophospholipid metabolism	1
syf00590	Arachidonic acid metabolism	1
syf00040	Pentose and glucuronate interconversions	1
syf00633	Nitrotoluene degradation	1
syf00473	D-Alanine metabolism	1
syf00053	Ascorbate and aldarate metabolism	1
syf00626	Naphthalene degradation	1

Appendix Table C2: Identified KEGG pathways of *A. vinelandii* ΔnifL

KEGG IDs	KEGG pathways	Matched proteins
avn01100	Metabolic pathways	567
avn01110	Biosynthesis of secondary metabolites	261
avn01120	Microbial metabolism in diverse environments	186
avn01200	Carbon metabolism	116
avn01230	Biosynthesis of amino acids	109
avn01240	Biosynthesis of cofactors	107
avn02020	Two-component system	66
avn00230	Purine metabolism	55
avn00620	Pyruvate metabolism	53
avn03010	Ribosome	49
avn02010	ABC transporters	44
avn00010	Glycolysis / Gluconeogenesis	39
avn00630	Glyoxylate and dicarboxylate metabolism	36
avn00190	Oxidative phosphorylation	35
avn00640	Propanoate metabolism	35
avn00680	Methane metabolism	30
avn00270	Cysteine and methionine metabolism	29
avn00650	Butanoate metabolism	26
avn00030	Pentose phosphate pathway	25
avn00020	Citrate cycle (TCA cycle)	25

avn02024	Quorum sensing	25
avn00250	Alanine, aspartate and glutamate metabolism	24
avn01212	Fatty acid metabolism	24
avn00970	Aminoacyl-tRNA biosynthesis	24
avn00240	Pyrimidine metabolism	23
avn00260	Glycine, serine and threonine metabolism	23
avn01210	2-Oxocarboxylic acid metabolism	22
avn00280	Valine, leucine and isoleucine degradation	22
avn00500	Starch and sucrose metabolism	21
avn03018	RNA degradation	19
avn00330	Arginine and proline metabolism	19
avn00520	Amino sugar and nucleotide sugar metabolism	18
avn03070	Bacterial secretion system	18
avn00220	Arginine biosynthesis	18
avn00540	Lipopolysaccharide biosynthesis	17
avn00061	Fatty acid biosynthesis	16
avn00860	Porphyrin and chlorophyll metabolism	16
avn02040	Flagellar assembly	15
avn00920	Sulfur metabolism	15
avn00760	Nicotinate and nicotinamide metabolism	14
avn00350	Tyrosine metabolism	14
avn00051	Fructose and mannose metabolism	14
avn00071	Fatty acid degradation	14

avn00400	Phenylalanine, tyrosine and tryptophan biosynthesis	14
avn02030	Bacterial chemotaxis	14
avn00770	Pantothenate and CoA biosynthesis	14
avn00362	Benzoate degradation	14
avn00340	Histidine metabolism	14
avn00052	Galactose metabolism	13
avn00290	Valine, leucine and isoleucine biosynthesis	13
avn00790	Folate biosynthesis	13
avn00910	Nitrogen metabolism	12
avn03060	Protein export	12
avn00900	Terpenoid backbone biosynthesis	12
avn04122	Sulfur relay system	12
avn00380	Tryptophan metabolism	12
avn00480	Glutathione metabolism	12
avn00780	Biotin metabolism	12
avn00550	Peptidoglycan biosynthesis	11
avn00300	Lysine biosynthesis	11
avn00541	O-Antigen nucleotide sugar biosynthesis	11
avn00360	Phenylalanine metabolism	11
avn00310	Lysine degradation	11
avn00730	Thiamine metabolism	11
avn03430	Mismatch repair	10
avn01220	Degradation of aromatic compounds	10

avn00450	Selenocompound metabolism	10
avn00430	Taurine and hypotaurine metabolism	10
avn00521	Streptomycin biosynthesis	10
avn00130	Ubiquinone and other terpenoid-quinone biosynthesis	10
avn00670	One carbon pool by folate	9
avn00562	Inositol phosphate metabolism	8
avn01501	beta-Lactam resistance	8
avn01503	Cationic antimicrobial peptide (CAMP) resistance	8
avn03440	Homologous recombination	7
avn03030	DNA replication	7
avn00261	Monobactam biosynthesis	7
avn00625	Chloroalkane and chloroalkene degradation	7
avn03420	Nucleotide excision repair	7
avn00660	C5-Branched dibasic acid metabolism	7
avn00072	Synthesis and degradation of ketone bodies	7
avn00561	Glycerolipid metabolism	7
avn00740	Riboflavin metabolism	6
avn00410	beta-Alanine metabolism	6
avn00750	Vitamin B6 metabolism	6
avn00564	Glycerophospholipid metabolism	6
avn03410	Base excision repair	5
avn02060	Phosphotransferase system (PTS)	5

avn00523	Polyketide sugar unit biosynthesis	5
avn01502	Vancomycin resistance	4
avn00626	Naphthalene degradation	4
avn03020	RNA polymerase	4
avn00040	Pentose and glucuronate interconversions	4
avn00471	D-Glutamine and D-glutamate metabolism	3
avn00401	Novobiocin biosynthesis	3
avn00281	Geraniol degradation	3
avn00785	Lipoic acid metabolism	2
avn00525	Acarbose and validamycin biosynthesis	2
avn00633	Nitrotoluene degradation	2
avn00643	Styrene degradation	2
avn00053	Ascorbate and aldarate metabolism	2
avn00903	Limonene and pinene degradation	2
avn00622	Xylene degradation	2
avn00332	Carbapenem biosynthesis	2
avn01055	Biosynthesis of vancomycin group antibiotics	2
avn00930	Caprolactam degradation	2
avn00627	Aminobenzoate degradation	2
avn00600	Sphingolipid metabolism	2
avn00473	D-Alanine metabolism	2
avn00590	Arachidonic acid metabolism	1
avn00909	Sesquiterpenoid and triterpenoid biosynthesis	1

avn00460	Cyanoamino acid metabolism	1
avn00642	Ethylbenzene degradation	1
avn00998	Biosynthesis of various secondary metabolites	1
avn00592	alpha-Linolenic acid metabolism	1
avn00100	Steroid biosynthesis	1
avn00621	Dioxin degradation	1
avn00364	Fluorobenzoate degradation	1
avn00361	Chlorocyclohexane and chlorobenzene degradation	1
avn00624	Polycyclic aromatic hydrocarbon degradation	1
avn00623	Toluene degradation	1
avn01040	Biosynthesis of unsaturated fatty acids	1

Appendix file C3. R script for analysing *pI* of *S. elongatus* or *A. vinelandii* proteome

```
library(seqinr)
library(Peptides)
# Load the proteome data
seq = read.fasta("sequence.fasta", seqtype="AA")
# Calculate the pI for the proteome
pI_scores <- pI(syn, pKscale = "EMBOSS")
pI_score <- mean(pI_scores)
print(pI_score)
```

Appendix file C4. R script for analysing molecular weight (M_w) of *S. elongatus* or *A. vinelandii* proteome

```
library(seqinr)
library(Peptides)
# Load the proteome data
seq = read.fasta("sequence.fasta", seqtype="AA")
# Calculate the Molecular weight of the proteome
mw_scores <- mw( syn,
   monoisotopic = FALSE,
   avgScale = "expasy",
   label = "none",
   aaShift = NULL
)
mw_score <- mean(mw_scores)
print(mw_score)
```

Appendix file C5. R script for analysing hydrophobicity of *S. elongatus* or *A. vinelandii* proteome

```
library(seqinr)
library(Peptides)
# Load the proteome data
seq = read.fasta("sequence.fasta", seqtype="AA")
# Calculate the hydrophobicity score for the proteome
hydro_scores <- hydrophobicity(seq, scale = "KyteDoolittle")
```

```
gravy_score <- mean(hydro_scores)
print(gravy_score)
```

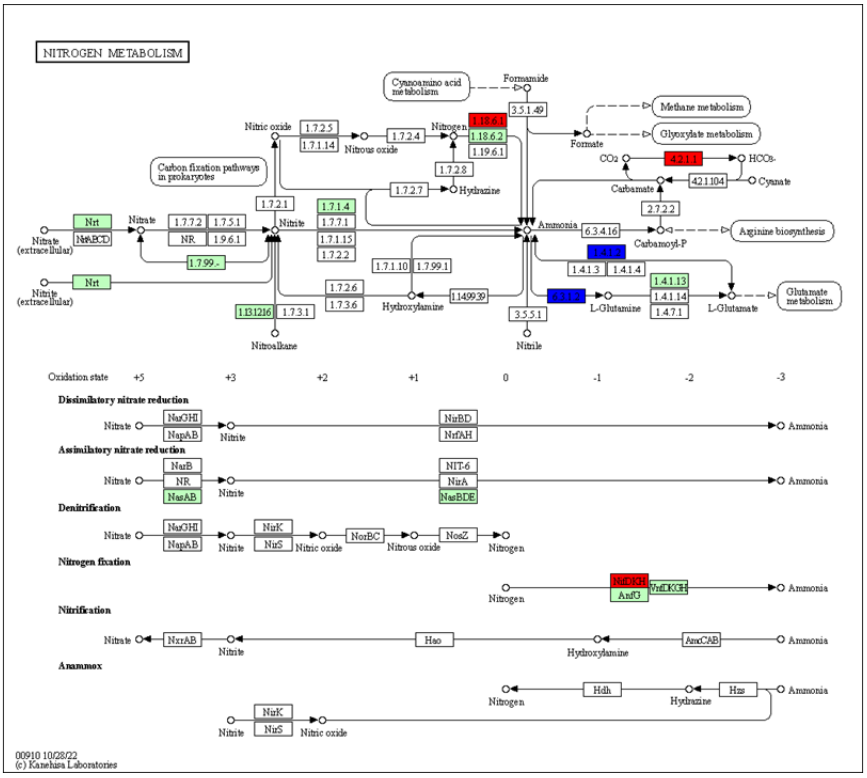
Appendix file C6. R script for analysing shared peptides between *S. elongatus* and *A. vinelandii*

```
library(cleaver)
library(seqinr)
# Load the proteome data
avin = read.fasta("Avin_00010.fasta", seqtype="AA")
syn = read.fasta("Synpcc7942_0001.fasta", seqtype="AA")
avin = sapply(avin, FUN=function (s) paste(s, collapse=""))
syn  = sapply(syn,  FUN=function (s) paste(s, collapse=""))
# Trypsin digestion
avin_pept = unlist(cleave(avin, enzym="trypsin"))
names(avin_pept) = NULL
avin_len = nchar(avin_pept)
avin_pept = avin_pept[8 <= avin_len & avin_len <= 25]
avin_pept = unique(avin_pept)
syn_pept = unlist(cleave(syn, enzym="trypsin"))
names(syn_pept) = NULL
syn_len  = nchar(syn_pept)
syn_pept = syn_pept[8 <= syn_len & syn_len <= 25]
syn_pept = unique(syn_pept)
# Calculate shared peptides
```

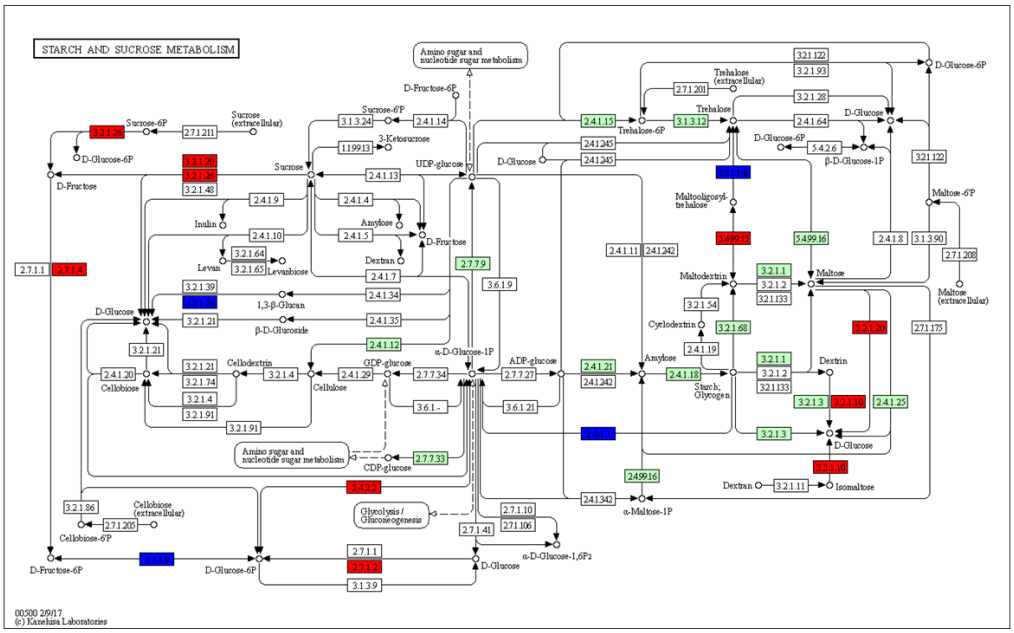
length(intersect(avin_pept, syn_pept))

Appendix D: Supplementary material for Chapter 5

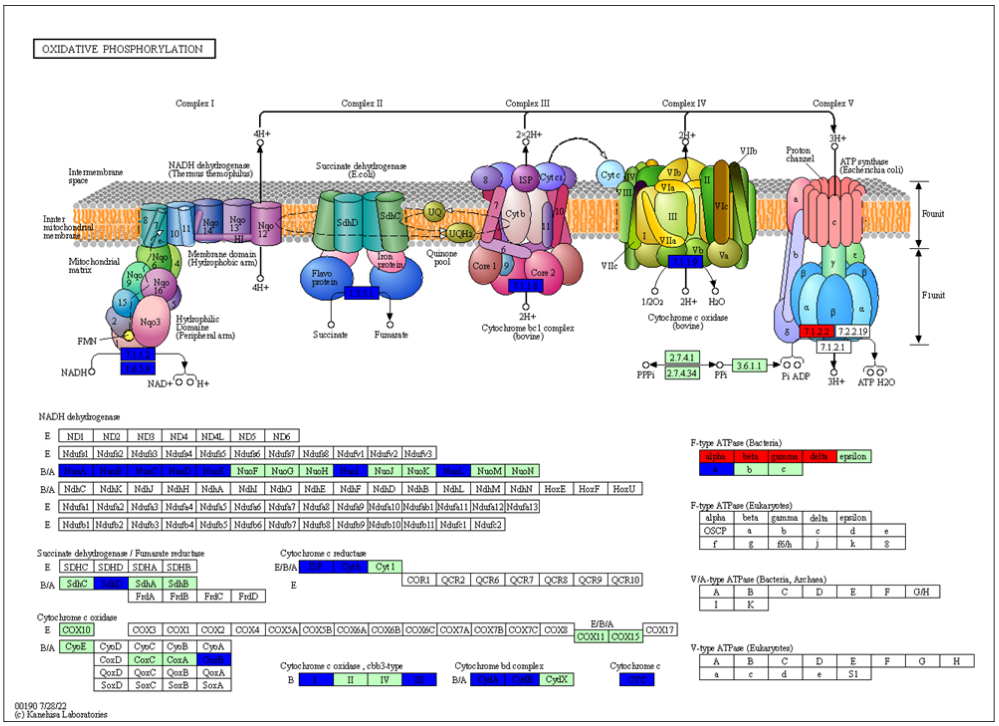
A



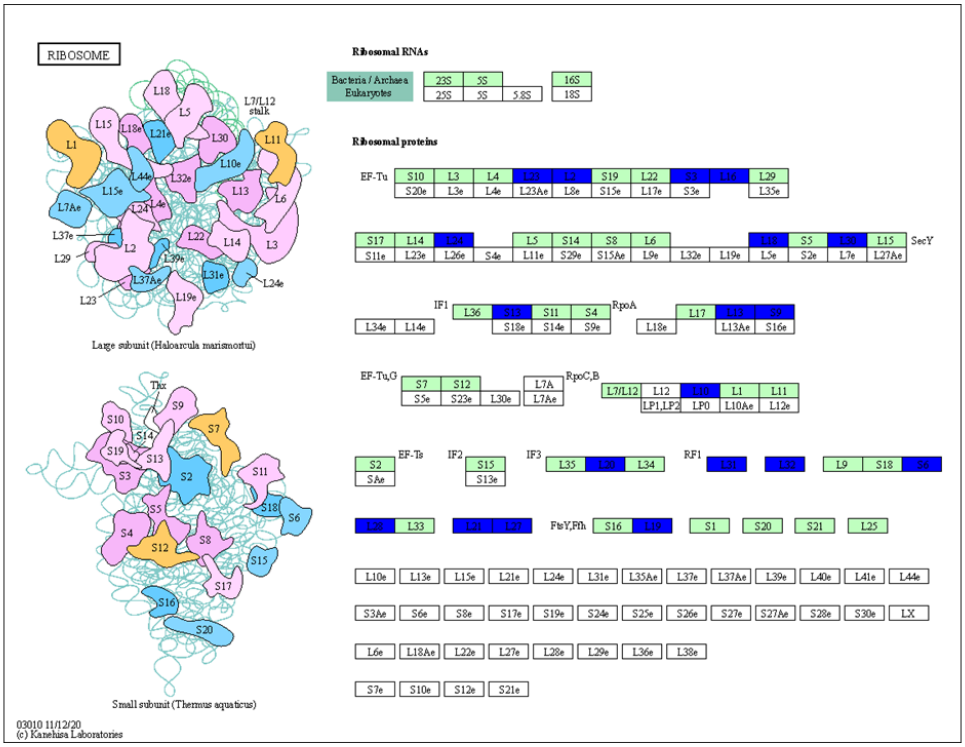
B



C

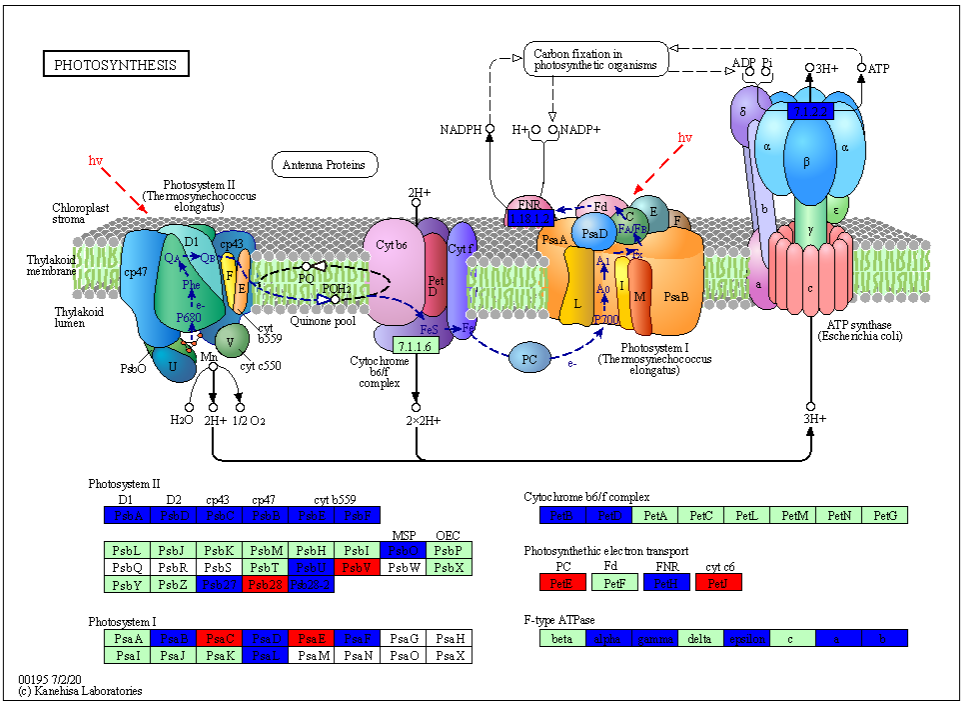

OXIDATIVE PHOSPHORYLATION
Complex I
Complex II
Complex III
Complex IV
Complex V
NADH dehydrogenase (Thermus thermophilus)
Succinate dehydrogenase (E.coli)
Cytochrome bc1 complex (bovine)
Cytochrome c oxidase (bovine)
Proton channel
ATP synthase (Escherichia coli)
Intermembrane space
Inner mitochondrial membrane
Mitochondrial matrix
Membrane domain (Hydrophobic arm)
Hydrophilic Domain (Peripheral arm)
Succinate
Fumarate
Quinone pool
NADH
NAD+
FMN
Pi ADP
ATP H2O
NADH dehydrogenase
F-type ATPase (Bacteria)
F-type ATPase (Eukaryotes)
V/A-type ATPase (Bacteria, Archaea)
V-type ATPase (Eukaryotes)
Succinate dehydrogenase / Fumarate reductase
Cytochrome c reductase
Cytochrome c oxidase
Cytochrome c oxidase , cbb3-type
Cytochrome bd complex
Cytochrome c
00190 7/28/22
(c) Kanehisa Laboratories

D

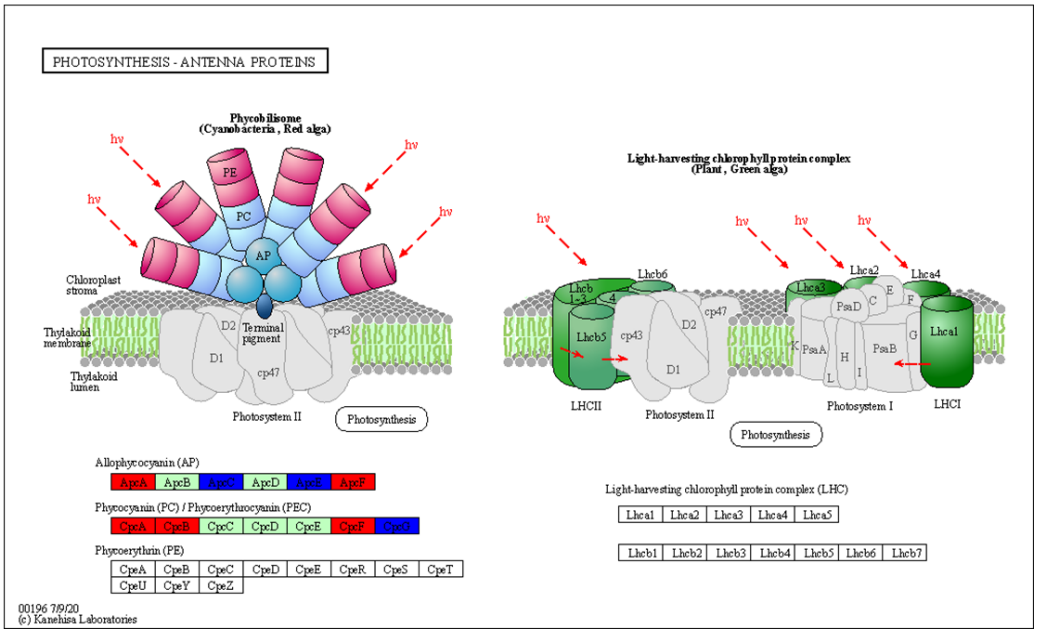

RIBOSOME
Large subunit (Haloarcula marismortui)
Small subunit (Thermus aquaticus)
Ribosomal RNAs
Bacteria / Archaea
Eukaryotes
Ribosomal proteins
EF-Tu
EF-Ts
EF-Tu,G
IF1
IF2
IF3
RF1
RpoA
RpoC,B
SecY
FtsY,Ffh
03010 11/12/20
(c) Kanehisa Laboratories

Appendix Figure D1: Interested KEGG pathways of *A. vinelandii* △nifL showing higher abundant proteins (red) or lower abundant proteins (blue) in co-culture compared with monoculture. (A) Nitrogen metabolism, (B) Starch and sucrose metabolism, (C) Oxidative phosphorylation, (D) Ribosome.

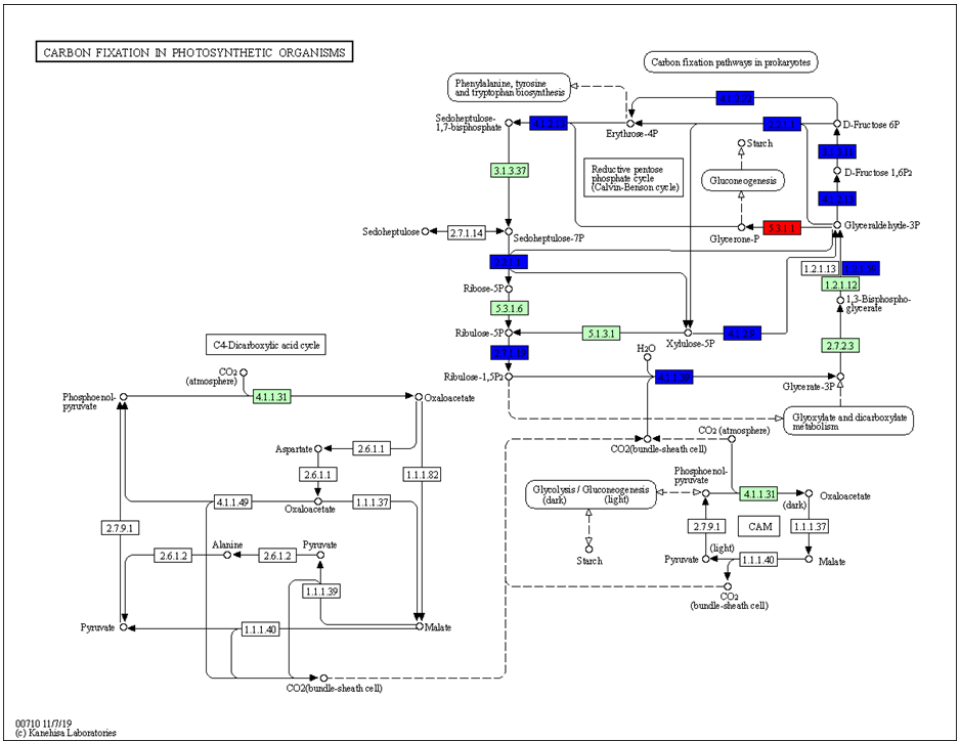
A



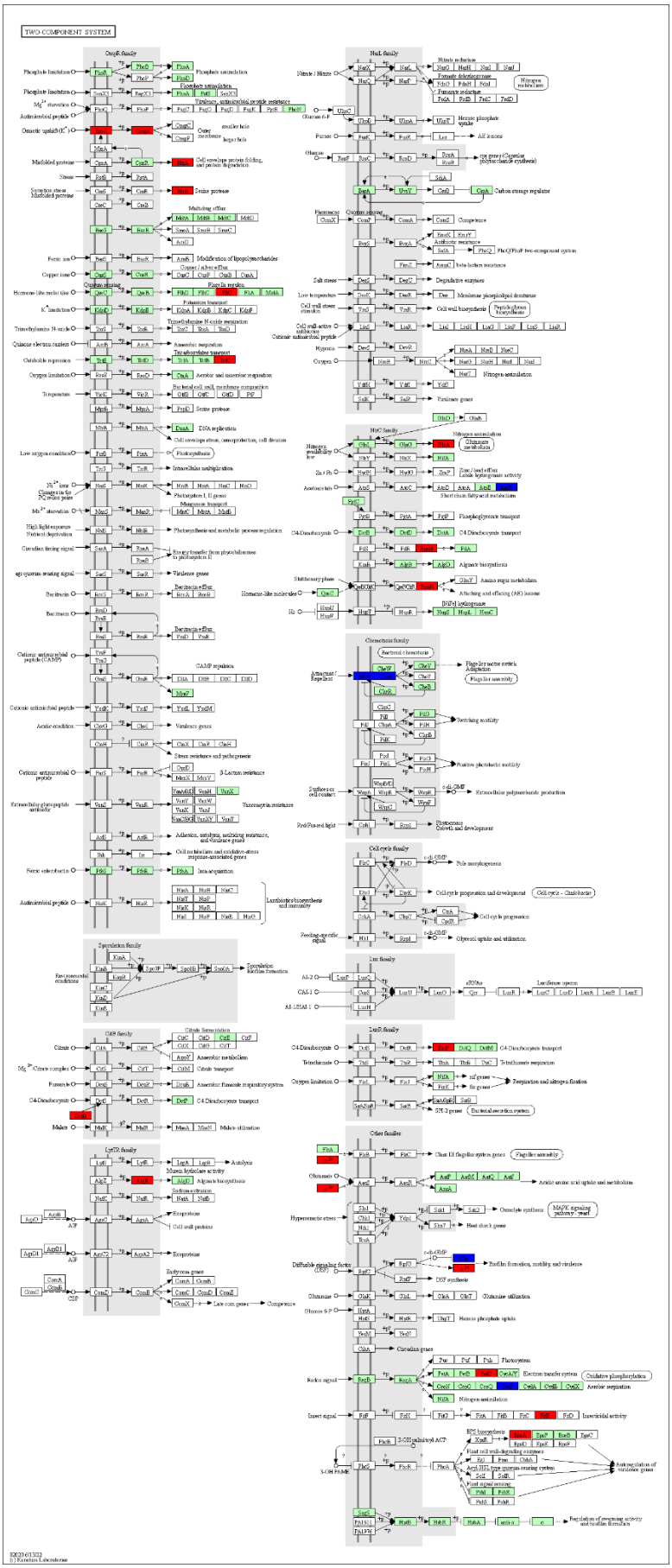
B



C



Appendix Figure D2: Interested KEGG pathways of *S. elongatus* cscB/SPS showing higher abundant proteins (red) or lower abundant proteins (blue) in co-culture compared with monoculture. (A) Photosynthesis, (B) Photosynthesis – Antenna proteins, (C) Carbon fixation in photosynthesis organisms.


TWO-COMPONENT SYSTEM

Appendix Figure D3: Interested KEGG pathway of *A. vinelandii* △nifL in co-culture on day 8 compared to day 4.