Investigation and Exploitation of Mycosporine-like amino acids from Rhodophytes and Bacteria

by

Bibi Nazia Auckloo

Thesis submitted for the degree of
Doctor of Philosophy in Life Sciences

University of Warwick, School of Life Sciences
June 2022

Table of Contents

Acknowledgement

First of all, I thank God for everything that I have in my life.

I thank the University of Warwick for awarding me the Chancellor's International Scholarship to pursue my PhD in School of Life Sciences.

A massive thanks to my supervisor Christophe Corre for believing in me (especially when starting a PhD in life sciences and had no clue about anything related to molecular biology). Thank you for taking the time to reply to my email 4 years ago despite the fact that you knew nothing about me and I'm grateful for accepting me in your group and it has been an honour to work with you for the last 3.5 years. Furthermore, a massive thanks to my co-supervisor Lijiang Song who had always been of great help in the analytical chemistry side of my work throughout my PhD. After all those long meeting hours, endless discussions, and hard work, I hope to have reached both of my supervisor's expectation as their student.

A big thank you to both of my advisory panel members, two powerful women scientists, Professor Elizabeth Wellington, and Dr Manuela Tosin for their valuable support and guidance during my studies. Big thanks also go to the MAA team including the Vasilios Stavros lab, Martin Wills and Emma MacPherson lab and to all my lab mates in the Corre group, especially Rohini, Laurence, Pan, Patrick, Matt, and Jack Ho as well as Dr Matthew Jenner, Dr Fabrizio Alberti, and the Alberti group. Spending countless hours in the lab with those people has been remarkable in every possible way. A huge thank you to some of my closest friends, Will, Angela, Abbie, Linda, Sunday, Anisa, Catriona, and Severine who had always been there for me.

Some very important people that I definitely need to thank: My mentor, Dr Mitrasen Bhikajee who has always encouraged me to persevere in life; My surgeon, Dr Maurice who saved my life in 2014; and Dr Pillai who has always supported me.

I owe my PhD to my lovely mum who has never given up on me and who has always believed in me; my dad who did his best for me and my brother to prosper in life; my adorable brother (my partner in crime) who is the best in every possible way; my dog Pipo who is in heaven; and my adopted Mummy Helen, her husband Louis and my adopted Grandma Doreen who made the last 3.5 years of my life extraordinary in England. My biological and my adopted family mean the world to me. Sometimes, you need someone to trust you and believe in you to be able to strive for success.

Declaration

This PhD thesis is submitted to the University of Warwick for the degree of Doctor of Philosophy in Life Sciences.
The thesis has not been submitted for a degree at another university.

The work presented was carried out by myself except the sections mentioned below:

Chapter 3, Section 3.5: Transient Electronic Absorption Spectroscopy (TEAS) was done by Abbie Whittock, a PhD student under the supervision of Professor Vasilios Stavros, Chemistry department, University of Warwick.

Chapter 4: Synthesis of AC23 (substrate used during enzymatic assay) was done by Adam Cowden under the supervision of Professor Martin Wills, Chemistry department, University of Warwick.

Appendix [2]: Proteomics samples were run at the Warwick Proteomics Facility (University of Warwick).

Published Work

Parts of this thesis (Chapter 3) have been published as shown below:

1) Whittock, A. L.; Woolley, J. M.; **Auckloo, N**.; Corre, C.; Stavros, V. G. "Investigating the ultrafast dynamics and long-term photostability of an isomer pair, usujirene and palythene from the mycosporine-like amino acid family". *Molecules*. **2022**, 27, 2272. DOI: https://doi.org/10.3390/molecules27072272.

2) Whittock, A. L.; **Auckloo, N**.; Cowden, A. M.; Turner, M. A. P.; Woolley, J. M.; Wills, M.; Corre, C.; Stavros, V. G. "Exploring the blueprint of photoprotection in mycosporine-like amino acids". *J. Phys. Chem. Lett.* **2021**, 12, 3641-3646. DOI: 10.1021/acs.jpclett.1c00728.

Additional work carried out in collaboration with Stavros group from the Chemistry department, University of Warwick (not included in this thesis) has also been published as shown below:

3) Abiola, T. T.; **Auckloo, N.**; Woolley, J. M.; Corre, C.; Poigny, S.; Stavros, V. G. "Unravelling the photoprotection properties of garden cress sprout extract". *Molecules*. **2021**, 26, 7631. DOI: https://doi.org/10.3390/molecules26247631.

Abstract

Pharmaceutical industries, especially skin-care related industries are focused on the development of new eco-friendly, affordable and photostable sunscreen formulations which possess potent photoprotective effects against the increasing levels of ultraviolet radiation exposure. The target in this study was 'nature-inspired' molecules from rhodophytes as well as microbial natural products structurally related to mycosporine-like amino acids (MAAs).

The MAAs shinorine and porphyra-334 were purified from the commercially available Helioguard^{TM} 365 whereas the cis-trans UVA absorbers namely usujirene/palythene were extracted from the edible rhodophyte *Palmaria palmata*, characterised using UHPLC-HRMS and ^{1}H NMR and their photoprotective mechanism was determined with the collaboration of the Chemistry department, University of Warwick.

A novel chemo-enzymatic procedure for accessing MAA-inspired metabolites was developed where the focal point of this research was the fourth enzyme, being a NRPS-like enzyme. Based on bioinformatic analyses, homologous enzymes containing NRPS-like domains with distinct substrate specificity were identified in the genome of diverse cyanobacteria where a new technique was proposed involving "A-domain" swapping in NRPS-like enzymes leading to novel recombinant enzymes that was projected to direct the biosynthesis of novel MAAs using a synthetic substrate, AC23, synthesized in the Chemistry department, University of Warwick. Furthermore, a plausible mechanism of the TE domain of the NRPS-like enzymes in this study was also proposed. Further research should be carried out, using the native substrate, namely mycosporine-glycine to confirm the functionality of the NRPS-like enzymes involved in this study.

As such, NRPS-like enzymes engineering can be further explored and exploited to generate nature-inspired photoprotective natural products with desired properties for pharmaceutical industries as well as for benefiting human in counteracting the harmful effects of UV radiation.

List of Abbreviation

AMP	Adenosine Monophosphate
ATP	Adenosine Triphosphate
A-domain	Adenylation domain
BPC	Base Peak Chromatogram
CL	Crude Lysate
CoA	Co-enzyme A
C-domain	Condensation domain
DHQS	Dehydroquinate Synthase
DNA	Deoxyribonucleic acid
EIC	Extracted Ion Chromatogram
ESI	Electrospray Ionisation
HPLC	High Performance Liquid Chromatography
IPTG	Isopropylthio-ß-D-galactoside
IR	Infrared Radiation
kb	Kilobase pairs
kDa	Kilo Dalton
KLD	Kinase-Ligase-DpnI
LB	Luria-Bertani
MAA	Mycosporine-like Amino Acid
Ni-NTA	Nickel-Nitriloacetic acid
NMR	Nuclear Magnetic Resonance
NRPS	Non-Ribosomal Peptide Synthetase

OD	Optical Density
O-MT	O-methyltransferase
PAR	Photosynthetically Active Radiation
PCP	Peptidyl Carrier Protein
PCR	Polymerase Chain Reaction
ROS	Reactive Oxygen Species
RP-HPLC	Reverse-phase High Performance Liquid Chromatography
SDS-PAGE	Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis
TE	Thioesterase
TEAS	Transient Electronic Absorption Spectroscopy
UHPLC-HRMS	Ultra-High Performance Liquid Chromatography- High Resolution Mass Spectrometry
UVR	Ultra-Violet Radiation
WHO	World Health Organisation

List of Illustrations and Tables

Chapter 1: General Introduction

Planet Earth accounts for 70% of ocean with 30% being the land mass. The biosphere is considered as an untapped source of scaffolds with priceless and unique compounds holding an assortment of biological properties[1,2]. Moreover, it is well-known that both terrestrial and marine organisms are able to withstand competitive and extreme conditions like salinity, pressure, temperature, dissolved oxygen, drought and low amount of nutrients to name a few by their capacity of generating remarkable chemicals with particular configurations and bioactivities[3,4]. Various natural products with a medley of chemical structures and biochemical uniqueness originate directly from both terrestrial and marine flora and fauna which make them potent candidates for the development of new and effective drugs for pharmaceutical and medical purposes[5–8].

1.1 A specific problem: Harmful UV radiation

The planet has always been exposed to infrared radiation (IR) absorbing at >800 nm, photosynthetically available radiation (PAR) absorbing at 400-750 nm and ultra-violet radiation (UVR). However, due to the problematic concern about climate change where the stratospheric ozone layer has been depleted recently, an increase in UV intensity attaining the earth has been seen[9]. UVA are known to absorb at 315-400 nm whereas the more energetic UVB absorb at 280-315 nm where both can be detrimental to human beings as well as marine life. UVC is known to absorb at 100-280 nm and is blocked by the Earth's atmosphere[10–12] as shown in Figure 1.1 below.

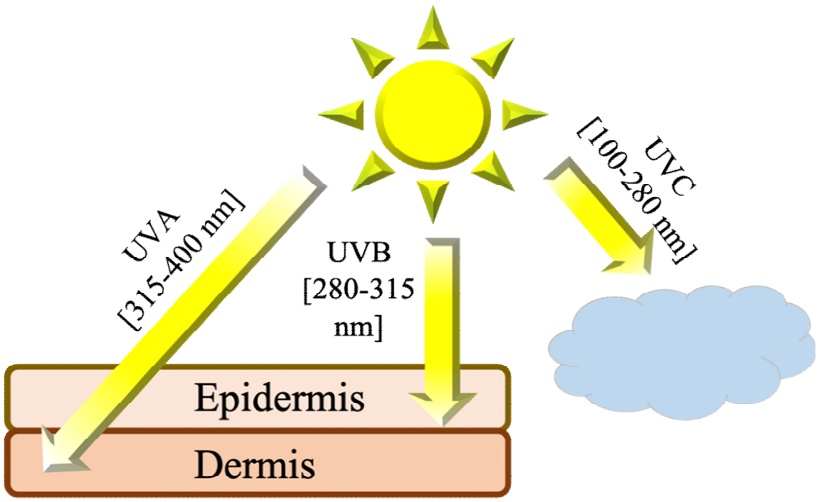


Figure 1.1: UVA absorbs in the range of 315 – 400 nm; UVB absorbs in the range of 280 – 315 nm and UVC absorbs from 100 – 280 nm.

UV radiation levels are influenced by various factors including the position of the sun where the higher the sun in the sky, the higher is the UV radiation level and it is also known that UV radiation is higher in regions closer to the equator. Altitude plays another role in the level of UV rays where every 1000 metres increase in altitude leads directly to an increase of 10% - 12% of UV radiation level. Not to be neglected is the varying ozone level over the year as well as reflection or scattering of UV radiation where for instance, 80% of UV radiation can be reflected by fresh snow, and up to 25% by the beach sand (Figure 1.2)[13].

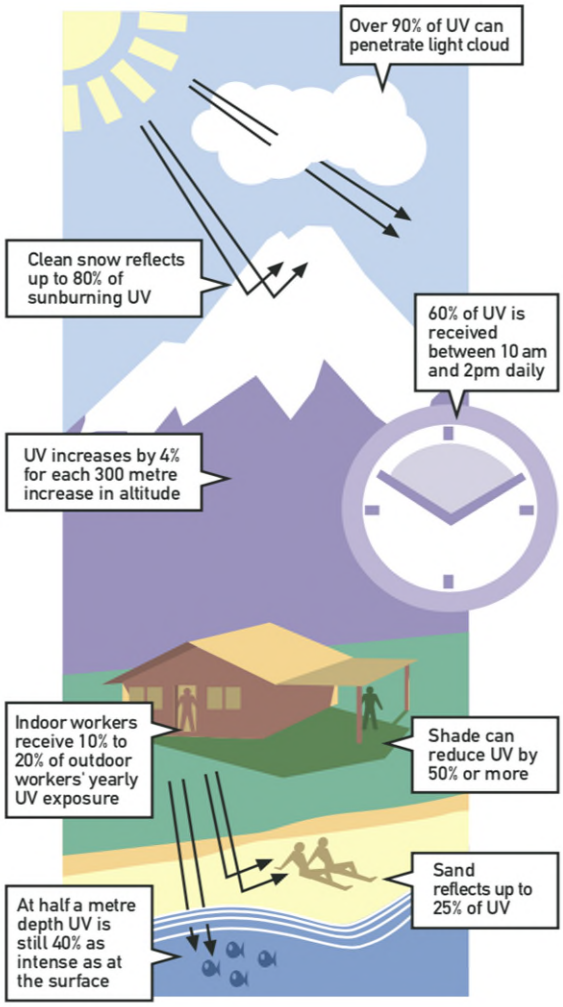


Figure 1.2: UV display in different environment as depicted by the World Health Organisation[13].

According to the World Health Organisation (WHO), overexposure to UV rays have been seen to have various consequences on human health including skin cancers, eye problem such as photokeratitis, photo-ageing (premature skin ageing) or sunburns among others[14].

Two to three million non-melanoma skin cancers and about 132000 diagnosed melanoma skin cancers occur every year around the globe[13]. As such, cutaneous malignant melanoma (a type of skin cancer appearing as a mole) was seen to affect mostly fair-skinned males from higher latitudes who were sporadically exposed to increased UV radiation during holidays[15]. Another disease related to overexposure to sunlight was seen to be squamous cell carcinoma which was more prominent at lower latitude[16]. Squamous cell carcinoma, another type of skin cancer appearing either as

flat reddish or brownish crusty patches have the tendency to be prevalently noticed in areas of chronic inflammation and scarring which might be attributed to the fact that those areas lack pigmentation in dark skinned individuals. Based on a previous study, squamous cell carcinoma was revealed to be more deadly in deeply pigmented people, in developed countries than the most common, less aggressive basal cell carcinoma, (a flesh-coloured mole or a bump) which is more flagrant in white populations[17,18]. Another impact of excess exposure to UV radiation on human skin relates to photo-induced ageing which includes sun-damaged alterations to the skin such as solar keratoses as well as modifications of intrinsic or chronologic ageing[19]. Incidences from sunburn have been seen since decades and the latter keep increasing in many parts of the world now. Substantial sunburns incidents among youngsters continue to increase where 72% between the age of 11-18 years old were seen in the United States during at least one summer whereas 48% of parents have declared at least one sunburn from their kids in the United Kingdom[20,21]. As a whole, the effects of harmful UV radiation on the skin correlate to the generation of reactive oxygen species (ROS) as well as DNA damage and protein deactivation as shown in Figure 1.3.

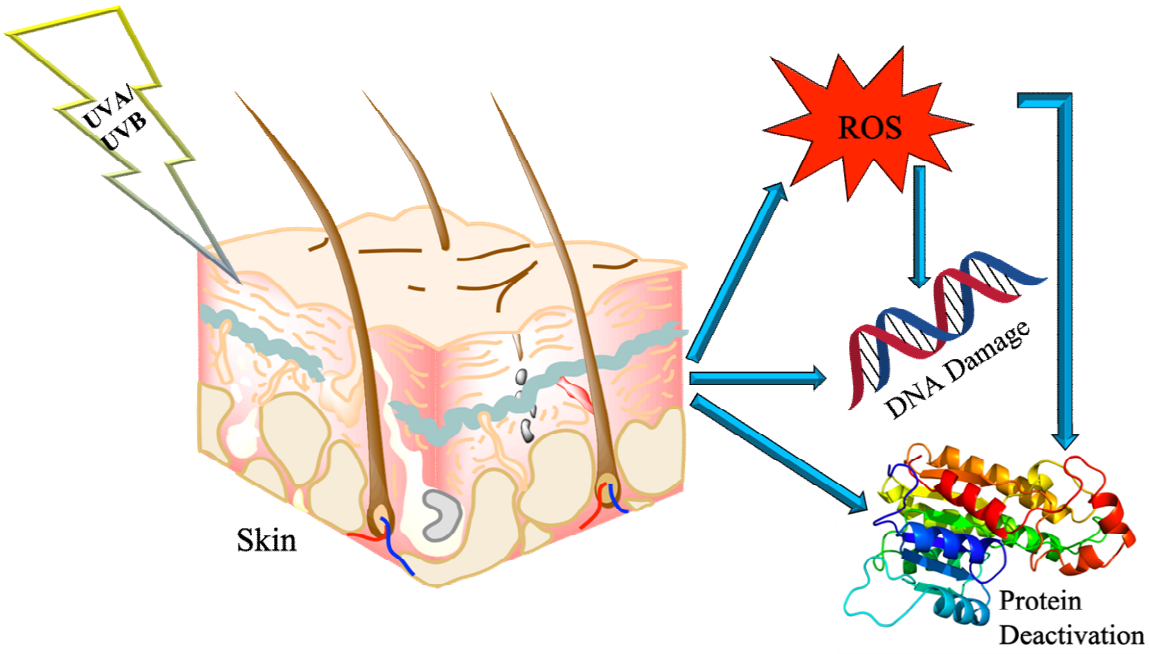


Figure 1.3: The effect of UVA and UVB on the skin.

An important concept which has long been hard to be understood is the UV Index (UVI) which refers to the measurement of the intensity of UV radiation on the earth's surface that has an impact on human skin where diverse UVI range are attributes to the level of UV radiation exposure as seen in Figure 1.4[13].

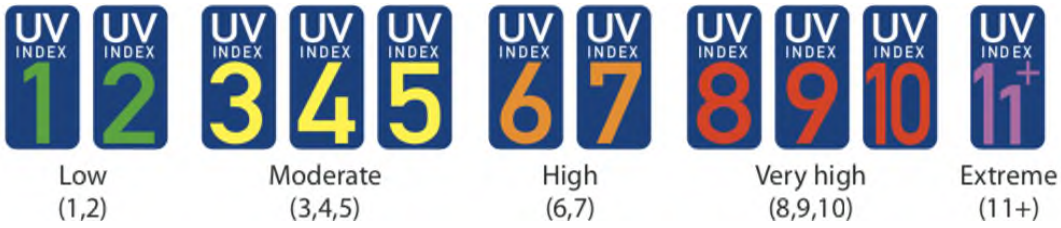


Figure 1.4: The International colour codes representing UV Index as depicted by the World Health Organisation[13].

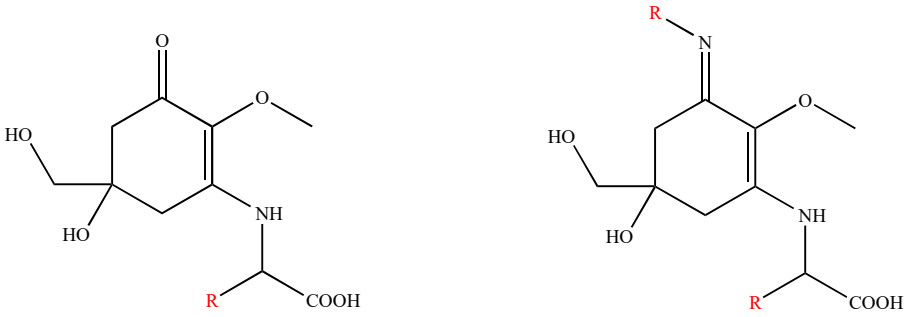
On a different note, prolonged sun exposure may be the reason for up to 20% of cataracts cases of blindness in various countries such as India, Pakistan and regions near the equator. In addition to the current situation, computational models have predicted that an increase of 300000 non-melanoma and 4500 melanoma skin cancers as well as approximately 1.7 million more cataracts cases annually worldwide for a 10% decline in the stratospheric ozone[13].

1.2 Solution: Mycosporine-like amino acids (MAAs) as photoprotective natural products

The marine ecosystem is not to be spared from the harmful effects of UV rays. The epipelagic zone (the sunlight zone: <200 m) as well as the intertidal zone are both exposed to the highest levels of UVR. UV exposure can trigger cellular impairment, stimulating DNA damages leading to the generation of reactive oxygen species (ROS) like superoxide radical or singlet oxygen which are known to be harmful to living organisms[22].

Therefore, marine organisms came up with one of the best strategies to cope with the high UV radiation by developing protection mechanisms involving the regulation and expression of particular enzymes which in turn lead to the production of UV absorbing molecules known as the mycosporine-like amino acids (MAAs).

MAAs are small, colourless, water-soluble compounds with a low-molecular weight of < 400 Da (with the exception of glycosylated MAAs which can go up to around 1000 Da) that absorb UV radiation in the range of 310 to 360 nm. They are present as intracellular secondary metabolites in various organisms[23,24] and should not be considered as pigments as they are transparent to visible light with high molar absorptivity (ε = 28100–50000 M^{-1} cm^{-1}) for UVA and UVB. Those compounds are composed of a cyclohexenone or cyclohexenimine chromophore core where amino acid substituents or imino alcohol substituents are attached through imine linkages resulting in a blend of resonance tautomers aiding in UV absorption[25,26]. One of the pivotal roles of MAAs is therefore their effective photoprotective capacities which depends on a number of factors such as their concentrations, the levels of UVR exposure, depth and salinity. Another aspect that showed variations in the MAAs' UV range absorption is the attached side groups and nitrogen substituents[27].



Cyclohexenone or cyclohexenimine chromophore core
R = amino acid or imino alcohol substituents

Figure 1.5: The cyclohexenone or cyclohexenimine chromophore core where R = amino acid or imino alcohol substituent

Significant levels of these UV-absorbing molecules were seen in a variety of phytoplankton species where they absorbed almost 70% of the harmful UV radiation prior to damaging molecular targets in the cytoplasm of the diatom *Phaeocystis antarctica*[28]. Another study demonstrated that a significant amount of MAAs were present in the surface layer of the siphonal mantle of the giant clam *Tridacna crocea* which had the capacity of absorbing about 90% of 320 nm radiation which in fact revealed a protective effect on the zooxanthellae *Symbiodinium* sp., in the animal tissue[29].

In addition to bearing photoprotective effects, some MAAs were seen to act as antioxidants in order to quench ROS and hamper cellular damage. The Dead Sea halotolerant cyanobacterium *Aphanothece halophytica* was seen to produce mycosporine-2-glycine as the predominant MAA. Mycosporine-2-glycine was revealed to significantly inhibit free radical scavenging activities with SC50 value of 22 ± 1.4 μM, thus displaying its potent antioxidant activity. Moreover, pre-incubation of the human melanoma A375 and NHSF cells with 0.31 μM mycosporine-2-glycine showed a shielding effect upon exposure to 10 μM hydrogen peroxide. During oxidative stress, specific transcription factors and activator proteins are triggered among which is the transcription factor-nuclear factor-κF (NF-κB). However, upon treatment with mycosporine-2-glycine, this expression of this particular protein declined when NHSF cells were exposed to hydrogen peroxide, resulting in some kind of barrier of NF-κB activation by mycosporine-2-glycine. It was also shown that this particular MAA was not genotoxic, meaning that it had the ability to protect against DNA damage making mycosporine-2-glycine an effective antioxidant factor against oxidative stress-mediated cell death[30].

Two glycosylated novel MAAs were isolated from the cyanobacterium *Nostoc commune* as shown in Figure 1.6. The first one was identified as a pentose-bound porphyra-334 derivative with molecular mass of 478 Da with an absorption peak at 335 nm and a molar absorption coefficient of 33173 M^{-1} cm^{-1}. The second one showed a molecular mass of 1050 Da with an absorption maxima at 312 nm along with a shoulder at 340 nm where the UV absorption spectrum revealed the presence of two different chromophores namely 3-aminocyclohexen-1-one and 1,3-diaminocyclohexen, which corresponded to absorption peaks of 312 and 340 nm respectively. The calculated molar absorption coefficient was deduced to be 58800 $M^{-1}cm^{-1}$ at 312 nm. Furthermore, due to the specific absorption peak at 3399 cm^{-1} in the IR spectrum, it was suggested that there was presence of high number of hydroxyl groups, here referring to pentose and hexose sugars. Both novel compounds exhibited potent antioxidant activities by showing ABTS radical scavenging ability whereas the 1050 Da MAA was the only one among the two which had the capacity of scavenging DPPH radicals *in vitro*[31].

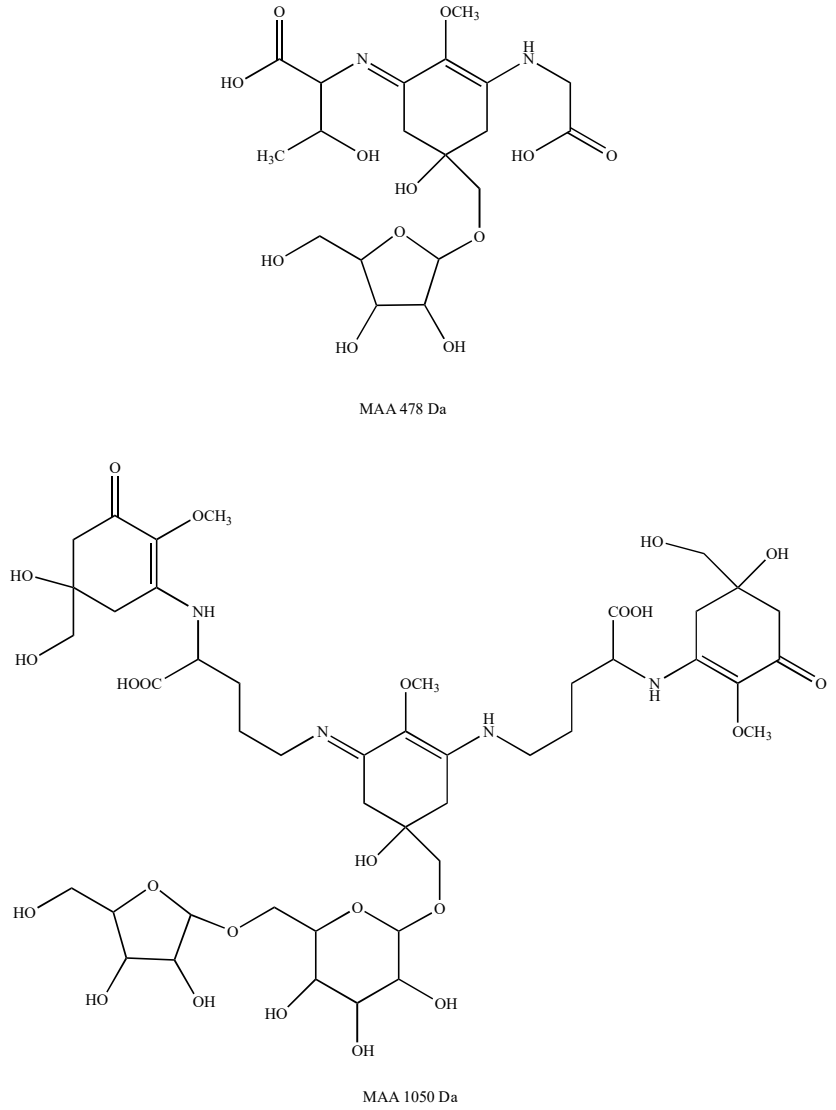


Figure 1.6: Predicted structures of the 478 Da MAA and 1050 Da MAA[31].

These glycosylated MAAs could be interpreted as both being UV-absorber molecules as well as antioxidant compounds related to anhydrobiosis (life without water) in the cyanobacterium *N. commune*[31].

1.2.1 Brief History of MAAs through time

The very first marine organism to display UV absorbing capacities (λ_{max} 305 nm) was seen in the gas gland of the epipelagic Portuguese man-of-war by Wittenburg in the 60s[32] which was later isolated by Price (1969) where the latter found the constituent to absorb at 310 nm. Tsujino (1961), discovered UV absorbing substances in Rhodophytes followed by another research group that revealed the presence of UV

absorbing matters termed as S-320 with a broad maximum absorbance at 320 nm in aqueous extracts of various corals in the *Acropora* genus and *Pocillopora* genus as well as a cyanobacterium isolated from the Great barrier Reef. After several years, S-320 was seen to be triggered by UV exposure as proof of its decline in the tissues of *Pocillopora damicornis* on long-term eradication of UVR. Since then, research had been ongoing where mycosporine-glycine was isolated from a tropical zoanthid, *Palythoa tuberculosa* with maximum absorption at 310 nm which possessed the same basic structure as a compound (P-310) isolated from the terrestrial fungus *Stereum hirsutum* during sporulation. In 1979, a diverse array of compounds with the mycosporine backbone was isolated namely palythinol, palythine and palythene from *Palythoa tuberculosa*. Porphyra-334, another compound with the mycosporine backbone with UV absorption capacities at 334 nm was isolated from the red algae *Porphyra tenera* followed by the isolation and characterization of shinorine (λ_{max} 334 nm) from the red algae *Chondrus yendoi* and another UV absorbing molecule at 357 nm namely usujirene isolated from the rhodophyte *Palmaria palmata*. Another compound named asterina-330, isolated from the starfish *Asterina pectinifera*, with UV absorption of 330 nm was later characterized. The above has been discussed in literature[33–35]. As a whole, an assortment of those UV absorbing compounds have since been identified in taxonomically diverse organisms ranging from marine prokaryotes, cyanobacterium[36,37], microalgae[34], macroalgae as well as in both symbiotic invertebrates-microbes and in non-symbiotic invertebrates[38–40]. Euhalothece-362 was isolated from the unicellular cyanobacterium *Euhalothece* sp. strain LK-1 located from the upper layer of a gypsum crust on the bottom of a hypersaline pond in Israel. The latter was revealed to produce mycosporine-2-glycine which absorb at 331 nm as well as Euhalothece-362 with absorption maxima of 362 nm, which is the only discovered mycosporine-like amino acid that absorb at that wavelength till now. Euhalothece-362 was seen to have molecular formula $C_{14}H_{23}N_2O_7$ with $[M+H^+]$ being 331.1495 based on HPLC-MS analysis. In-depth HPLC separation of hydrolysate of Euhalothece-362 displayed an α-alanine peak which was equivalent to $C_3H_7NO_2$ as seen in Figure 1.7. The red shift in absorbance maxima of this particular compound was attributed to the presence of an additional conjugated double bond, which is also found in both palythene and its cis isomer usujirene. Nevertheless, the exact structural formula was hard to be put forward as the

four hydroxyl groups as well as the carboxyl group of the alanine moiety might have acted as an iron chelator and thus hamper NMR analysis due to magnetic interferences[41]. A proposed chemical structure of euhalothece-362 was henceforth proposed as shown below.

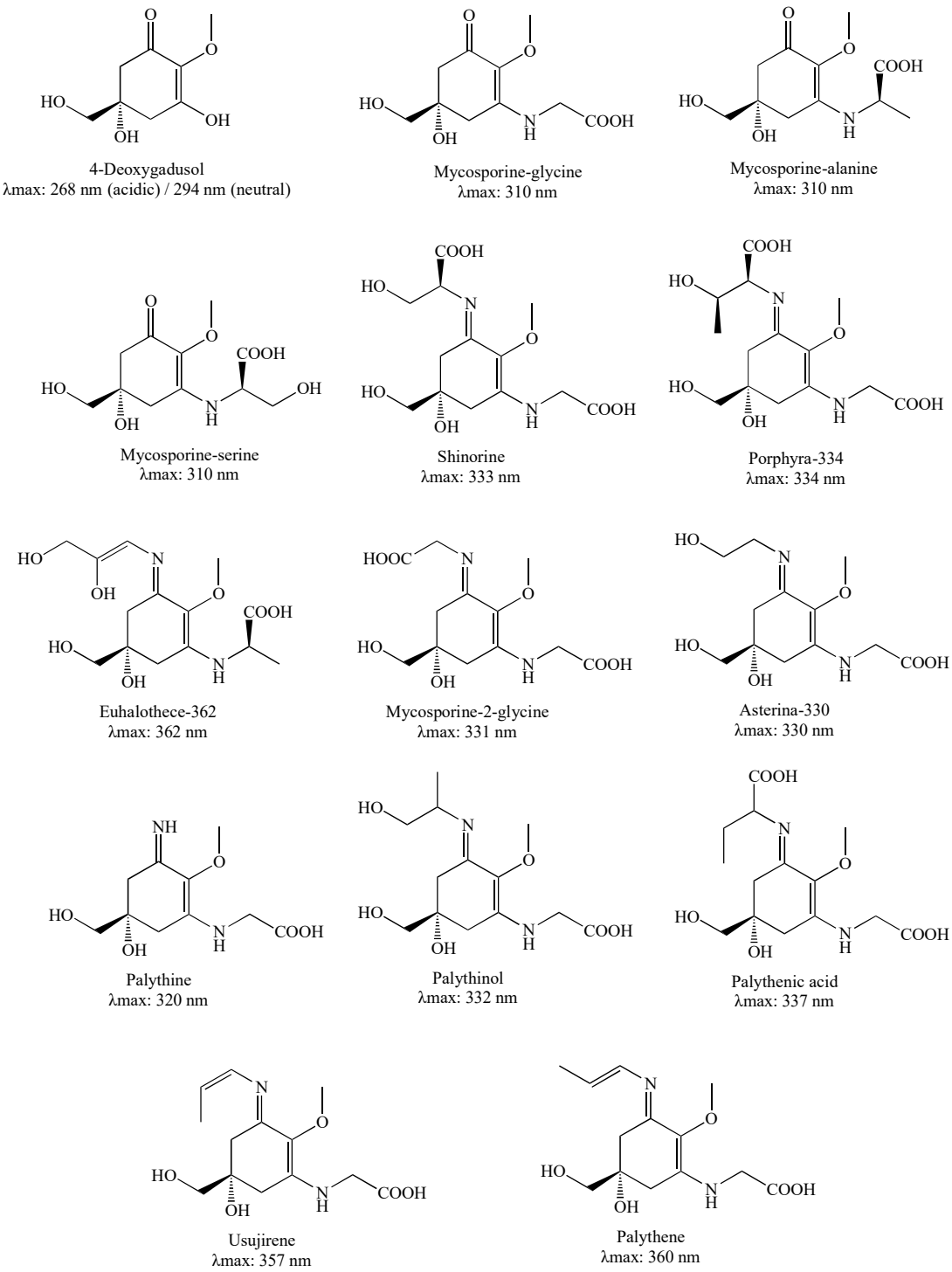


Figure 1.7: Chemical structures and wavelength of a range of cyclohexenone and cyclohexenimine MAAs.

1.2.2 Extraction and characterisation of known MAAs from literature

There have been a range of techniques for the extraction of mycosporines and mycosporine-like amino acids based on literature. Different extraction solvents were used including ethanol, ranging from 80% ethanol for 24 hours at 4ºC for the extraction of UV absorbing metabolites from the eleven species of Antarctic diatoms[28] to absolute ethanol where an unidentified substance namely P_{310} was detected spectrophotometrically in various fungal extracts exposed to six hours of near-ultraviolet radiation[42]. In addition, aqueous extracts from the red tide dinoflagellate *Alexandrium excavatum* yielded various mycosporine-like amino acids with absorbance maxima from 310–360 nm in the UV spectrum upon exposure to high light intensity (200 μE m^{-2} s^{-1})[25]. Another popular solvent used for the extraction of those photoprotective metabolites is methanol, ranging from 20% methanolic extraction where thirteen distinct MAAs were seen from thirteen cyanobacterial strains[36], 25% methanol extraction at 45ºC yielded higher total MAA concentration such as shinorine, palythine, asterina-330 or porphyra-334 from freshwater phytoplankton, the copepod *Cyclops abyssorum tatricus*, and the rhodophyte *Porphyra* sp.[43] up to absolute methanol for the extraction of MAAs in tropical marine fish eye tissues from the Great Barrier Reef[44]. Analytical techniques such as the use of reversed phase high performance liquid chromatography (HPLC) utilizing C_{18}-phase has been among the most common method of characterizing MAAs when particular solvent concentration is used as mobile phase. For instance, in early days, Nakamura and co-workers applied the concept of reversed phase HPLC using prepacked columns like MCI Hypersil ODS HY-5U (5 μm, Mitsubishi), ALTEX Ultrasphere ODS (5 μm) and Develosil ODS-3 (3 μm, Nomura), 25 cm, 4.6 mm with dilute acetic acid as mobile phase being eluted isocratically where the retention times of the different MAAs were all dependant on each column's interaction with the various compounds which was afterwards attributed to the presence of specific functional groups like carboxyl groups or the hydrophobic properties[45].

Popular methods for the identification and characterization of mycosporines and mycosporine-like amino acids is the combination of HPLC and mass spectrometry (LC-MS). Two UV-B absorbing molecules namely mycosporine-glutaminol-glucoside (*m/z* 465) and mycosporine-glutamicol-glucoside (*m/z* 466) with absorption

maxima at 310 nm were successfully identified from various microcolonial fungi[46]. Extraction efficiency is an important factor to consider and that was one of the main focus of Dunlap et al., (1986), in order to identify and quantify the chemical composition of S-320 (mixture of water-soluble molecules) from the Great Barrier Reef coral *Acropora Formosa* where three MAAs namely mycosporine-glycine, palythine and palythinol were successfully extracted by using 20% tetrahydrofuran (THF) and 80% methanol followed by quantification of the isolated natural products by isocratic HPLC employing diverse mobile phase composition including 10% methanol and 90% water when using Whatman, 10 μ Partisil, ODS-3 (22.1 mm ID x 50cm) column; 0.1% acetic acid, 10% methanol and 89.9% water when using Brownlee, Spheri-5, RP-8 (4.6 mm ID x 25 cm), with RP-8 guard (4.6 mm ID x 5cm); and 5mM aqueous phosphate buffer when using Brownlee, Spheri-5, Amino (4.6 mm ID x 25 cm)[47]. Two novel MAAs named mycosporine-taurine and mycosporine-2 glycine were successfully extracted from the temperate sea anemone *Anthopleura elegantissima* by using 80% methanol followed by fractionation of the crude extract using HPLC with methanol, water, acetic acid as mobile phase and ultimate purification was done on a 25-cm Brownlee amino column using a mobile phase comprising of 40 mM ammonium acetate and 17.5 mM acetic acid in 80% methanol[48]. Another popular solvent used is acetonitrile while employing gradient elution methods of natural products in HPLC. As such, oligosaccharide-mycosporine amino acids were extracted from fresh colonies of *Nostoc commune* collected from Germany using 25% methanolic extraction for 1.5 hours at 40ºC. After getting rid of proteins, the oligosaccharide-mycosporine amino acids were afterwards purified by reversed-phase HPLC (5 μm RP-18 column, 125 x 4 mm) starting from 100% water to 5% acetonitrile in water for 30 minutes[49]. In 2005, Carreto and co-workers displayed the ability of separating over 20 MAAs from four algal cultures, three phytoplankton samples, four rhodophytes and three symbiotic organisms by coupling a C_{18} column system with an optimized aqueous mobile phase comprising trifluoroacetic acid (TFA) and ammonium as ion-suppression/ion-pairing agents with two columns connected in series at a temperature of 35ºC[50]. Another method applied for the characterization of mycosporines and mycosporines-like amino acids is the solid–liquid extraction technique where lyophilized fungal, cyanobacterial and actinomycetes strains samples were pulverized with liquid nitrogen and extracted with 0.2% aqueous acetic acid with

0.5% methanol (v/v) at 4ºC for 12 hours on a shaker, followed by a reversed phase liquid chromatography/mass spectrometry identifying each mycosporines and MAAs by their retention time, absorption maxima as well as their fragmentation patterns[46].

1.3 Brief Introduction on Macroalgae

There are three main categories of macroalgae namely Rhodophyta which refers to red algae, Chlorophyta which refers to green algae and Pheophyta which refers to brown algae. Rhodophytes are possibly one of the most ancient groups of eukaryotic algae. Different types of algae thrive under various conditions ranging from fresh to brackish to marine environment in tropical, temperate and arctic regions[51]. Red seaweeds are exceptional at tolerating a wider range of light levels in comparison to other groups of photosynthetic lives due to the fact that they possess chlorophyll a along with other accessory pigments such as phycobiliproteins like R-phycoerythrin and R-phycocyanin which contribute to the possibility of the latter to live in greater depth[52]. Among the three types of seaweeds, red algae are well-thought-out as the wealthiest one in terms of secondary metabolites assortment and abundance comprising more than 1500 diverse compounds belonging to various classes of natural products[53].
Moreover, it is recognized worldwide that the consumption of seaweeds possessing nutritional properties boosts up human benefits to some extent. Among the 10,000 marine macroalgal species present, a small portion of only about 145 species are currently used as food source[54]. As such, macroalgae contain about fifty-six different minerals such as calcium, iron or iodine to name a few as well as some trace elements[55], plus a quite high level of natural antioxidants in the form of polyunsaturated fatty acids like omega 3 and omega 6 which displayed certain responsibilities in the prevention of diabetes and cardiovascular diseases[56]. Eicosanoids, which are oxygenated derivatives of fatty acids, are normally present in rhodophytes which plays significant roles in immune responses and homeostasis[57]. Vitamins present in the form of water-soluble or fat-soluble vitamins are present in a variety of macroalgae. For instance, vitamins A, B1, B12, C, D and E are found in various quantities in corresponding types of seaweeds which have mitigation effects of ageing process as well as anaemia. Another vital component of seaweeds is polysaccharides which can reach up to 60% that can behave as a good dietary fibre source for the digestive system[58].

Since ages, macroalgae top the list of organisms possessing an assortment of molecules bearing medicinal properties. However, their commercial exploitation by pharmaceutical industries were not common until recent years where more and more scientists had put their focus on those sea plants for the hunt of bioactive natural products[59]. Seaweeds have been demonstrated to exhibit various health benefits which can act against heart diseases, inflammation, breast cancer or certain bacterial pathogens causing dreadful infections[55]. In 2011, a particular research had disseminated macroalgae as an antiviral agent which displayed the ability to hinder the transmission of the human immune deficiency virus along with other sexually transmitted viruses such as herpes simplex virus making seaweeds one of the candidates to focus on for advanced scientific research. Another component that is known to be present in seaweeds is mineral which varies based on the life cycle of the latter and it had been suggested to have a positive effect in the treatment of thyroid goiter[60].

1.3.1 Rhodophyte *Palmaria palmata* [Characteristics and Life cycle]

Palmaria palmata (Linnaeus) (Weber & Mohr 1805), previously known as *Rhodymenia palmata*, as well as commonly called Dulse or Dillisk in English and Dilleasc or Creathnach in Irish is a rhodophyte (red alga) that are found on the northern coasts of the Atlantic Ocean, commonly distributed throughout England and Ireland as well as ranging from Portugal to the Baltic coasts as well as the Icelandic coasts. *P. palmata* have also been seen on the shores of Arctic Russia, Arctic and Atlantic Canada, Alaska, Korea and Japan[61–63].

Table 1.1: Taxonomy of *Palmaria palmata*

Phylum	**Rhodophyta**
Class	Florideophyceae
Order	Palmariales
Family	Palmariaceae
Genus	*Palmaria*
Species	*P. palmata*
Authority	(Linnaeus) Weber & Mohr, 1805

P. palmata have usually a leathery-membranous texture with flattened fronds of approximately 50-300 mm in length which can sometimes go up to 1000 mm and 30-80 mm in width which ranged from a rusty purple to brownish colour. This type of seaweed arises from a discoid holdfast with a small inconspicuous stipe around 5 mm long intensifying progressively to form simple or dichotomously or palmately divided fronds with variable blades shape. These rhodophytes are widely distributed in the intertidal and subtidal zones and can reside to a depth of up to 20 m in exceptionally clear waters. They may be found on rocky shores, mussels or on upper sections of the phaeophyte (brown alga) *Laminaria hyperborea* stipes[62].

Life cycle of *Palmaria palmata*

P. palmata has an unusual life cycle involving a diplohaplontic (one in which full-grown haploid and diploid forms alternate) and a heteromorphic life cycle (different forms occur at different stages in the cycle) consisting of an asexual tetrasporophyte stage alternating with a sexual gametophyte stage and with an extreme sexual dimorphism. There's a similarity between the morphology of the diploid tetrasporophytes and the male gametophytes until about a year where irregular, dark red tetrasporangial sori are formed on their fronds starting around November till March mostly around Ireland coasts. These tetraspores settle when released and developed into female gametophytes and male gametophytes[61,64–66]. The carpogonia, which is the female sex organ of the red algae, are then fertilized by the males' gametes released from the year before. As such, a zygote is formed by fertilization and a new tetrasporophyte develops from the female gametophyte which overgrows the latter. It is therefore suggested that *P. palmata* is a pseudoperennial seaweed[67] based on the fact that it partly loses its thallus during summer season[68].

1.3.2 *Palmaria palmata* as a MAAs producer

P. palmata exhibit various biological activities such as antioxidant properties where the latter could be of benefits to prevent oxidative stress that could otherwise lead to neurodegenerative diseases as well as this macroalgae could also be of interest to hinder or treat gastric ulcers as well as cancers[69,70] as it possesses a specific type of algal protein family called phycobiliprotein[71].

Apart from the above-mentioned properties, the edible *P. palmata* have been seen to produce a diversity of MAAs according to literature. Eight different MAAs namely mycosporine-glycine, shinorine, porphyra-334, palythine, asterina-330, palythinol, palythene and usujirene were produced from this red seaweed collected in the Arctic Kongsfjord, Spitsbergen, Norway[72]. The quantity of various MAAs produced were affected by the effects of specific conditions including solar without UVA+B, solar without UVB as well as the full solar spectrum at different depth. It was noted that the transplantation of dulse from 3 metres depth to 0.2 metres depth followed by one-week exposure to diverse radiation treatment led to a significant increase in MAAs production, with each MAA responding differently under each specific condition. In addition, the highest level of those UV-absorbing compounds was revealed under the full solar spectrum radiation after the transplantation of *P. palmata* from 7.5 metres depth to 1 metre depth. As a whole, it could be said that there is a correlation between MAAs production and depth as well as upon exposure to various levels of UV radiation[72]. It can also be deducted that macroalgae experience an array of doses of photosynthetically active radiation (PAR) and UV depending on the characteristics of the water column such as the optical properties as well as the season, cloud cover or even sun angle and intensity. Furthermore, an important aspect not to be neglected is the break-up or melting of sea ice, especially in spring, which might contribute to greater depth as well as significant incursion of suspended particles in certain areas where *P. palmata* are present which thus paly a consequence on the yield of the UV-absorbing molecules[73]. As such, the production of MAAs in red algae might be more substantial in tropical regions which are exposed to stronger solar radiation in comparison to temperate areas[74]. Macroalgae consist of different sections as explained above and it had been proven that there was a microscale variation in the UV-absorbing molecules' concentrations along the thalli of *P. palmata* where young apical tips of the latter were seen to display greater MAAs yield that the older basal areas of the seaweeds[72] which was in accordance to other studies done in the rhodophyta *Eucheuma striatum*[75].

MAAs with total ion masses 245.00, 333.10, 289.10, 303.10, 347.10 and 285.10 were isolated from the red seaweed *P. palmata* where the various MAAs were incorporated into cow milk to counteract the degradation effect of natamycin, a natural antimicrobial preservative used in food, upon exposure to light[76]. The concentration of natamycin in milk decreased to almost 83% after 8 days at 4ºC due to photo-

degradation. However, the addition of MAAs into the milk boosted up the stability of natamycin where the latter was seen to be retained at 82.2 ± 0.9% and 92.2 ± 0.9% when low and high MAA concentrations were respectively incorporated.

1.4 Linking Biosynthetic genes and MAAs production

1.4.1 MAA biosynthesis in freshwater Cyanobacteria

[1] *Anabaena variabilis* ATCC 29413

Anabaena variabilis ATCC 29413 is a filamentous freshwater cyanobacterium that is known to produce the MAA shinorine[77]. Four genes, [*Ava_3858-Ava_3855*] where *Ava_3858* represents a DHQS homolog, *Ava_3857* being a O-methyltransferase, *Ava_3856* being a ATP-grasp and *Ava_3855* corresponding to a NRPS-like enzyme, are involved in the biosynthesis of shinorine in this particular species where the whole 6.5 kb biosynthetic gene cluster as well as two truncated constructs with *Ava_3858-Ava_3856* and *Ava_3858-Ava_3857* were cloned and expressed in the heterologous host *E. coli*. Strains containing the truncated constructs *Ava_3858-Ava_3856* produced 4-deoxygadusol (pH = 7: λ_{max} = 294 nm and pH = 2: λ_{max} = 268 nm) and mycosporine-glycine (λ_{max} = 310 nm) whereas strains lacking both *Ava_3856* and *Ava_3855* generated 4-deoxygadusol only. The strains containing the entire cluster were seen to produce 4-deoxygadusol and shinorine (λ_{max} = 333 nm) where all the compounds were characterized based on their wavelength, standard (from Helioguard 365) as well as analytical spectroscopy like ^{1}H NMR and high-resolution mass spectroscopy.

The biosynthetic pathway of mycosporines and mycosporine-like amino acids has long been assumed to be via the shikimate pathway. However, an important discovery made during this study unravelled that the biosynthetic pathway of mycosporines and mycosporine-like amino acids might in fact be the pentose phosphate pathway (which is a crucial primary metabolic pathway) where the enzyme sedoheptulose 7-phosphate cyclase (SH7PC) catalyse the cyclization of C_7-sugar phosphates to cyclic compounds by using sedoheptulose 7-phosphate (SH7P) as substrate[77].

There are three known sedoheptulose 7-phosphate cyclases till now and the one involved in MAAs production is desmethyl-4-deoxygadusol synthase (DDGS) which has the capacity to convert sedoheptulose 7-phosphate to desmethyl-4-deoxygadusol (DDG)[77,78]. As a whole, the sugar phosphate cyclases could be thought as a bridge

between primary and secondary metabolism by converting pentose phosphate pathway's intermediates to secondary metabolites.

In 2007, Xu and co-workers were the first to report the presence of a distinct group of sugar phosphate cyclases (DDGS) in diverse cyanobacteria[79]. As a consequence, DDGS together with another enzyme namely *O*-methyltransferase converted SH7P into 4-deoxygadusol (4-DG) which in turn was converted to mycosporine-glycine by an ATP-grasp enzyme in the presence of glycine and finally to the known MAA shinorine by a NRPS-like protein. This NRPS-like enzyme contains a typical adenylation, thiolation and thioesterase domains involving an ATP-dependent condensation reaction where L-serine gets activated as aminoacyl-AMP by the adenylation domain and transferred to the thiolation domain (peptidyl carrier protein domain), thus releasing AMP. Successively, the L-serine reacted with the ketone group of the mycosporine-glycine precursor forming a peptide bond and finally the compound produce, here referring to shinorine is released by the thioesterase domain[77] [Figure 1.8].

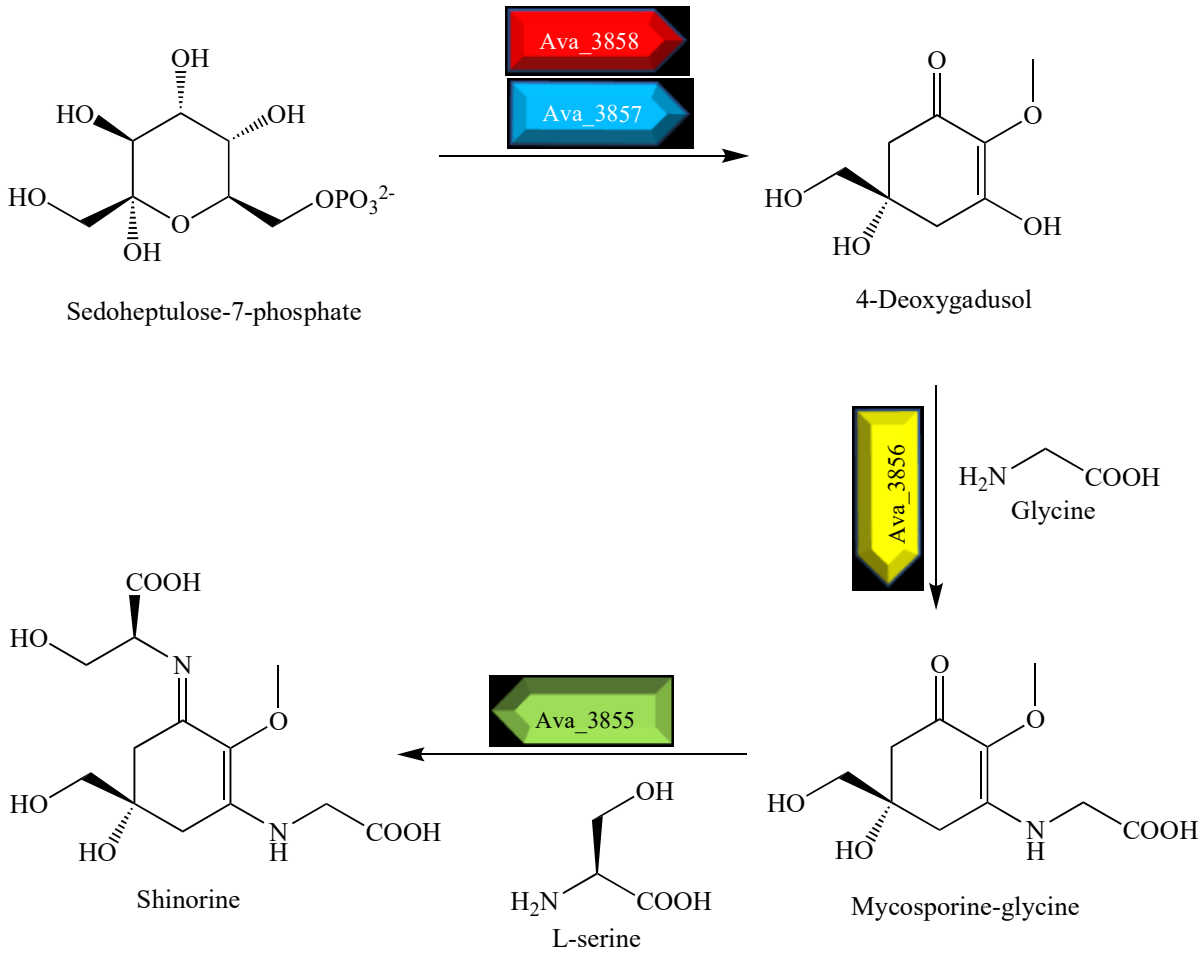


Figure 1.8: Shinorine biosynthetic pathway from the cyanobacteria *Anabaena variabilis*

[2] *Nostoc punctiforme* ATCC 29133

Nostoc punctiforme ATCC 29133 is also a filamentous freshwater cyanobacteria which contains four genes involved in the biosynthesis of oxomycosporine and iminomycosporines. The four genes involved in the biosynthesis of MAA in *Nostoc punctiforme* ATCC 29133 were seen to be *NpR5600*, *NpR5599*, *NpR5598* and *NpF5597* (also termed as *mysA*, *mysB*, *mysC* and *mysD* respectively) where the genes *NpR5600* and *NpR5599* were perceived to be homologs of *Ava_3858* and *Ava_3857* respectively, where *NpR5600* in *N. punctiforme* codes for demethyl 4-deoxygadusol synthase using sedoheptulose 7-phosphate as a substrate and *NpR5599* encodes O-methyltransferases that catalyze the methylation of demethyl 4-deoxygadusol synthase, thus leading to the production of 4-deoxygadusol[77,80]. In order to understand the proper biosynthetic role of the two genes *NpR5598* and *NpF5597*, the three genes

biosynthetic gene cluster (*NpR5600* to *NpR5598*) as well as the four genes cluster (*NpR5600* to *NpF5597*) were cloned and heterologously expressed in *E. coli*. The three genes, *NpR5600*, *NpR5599* and *NpR5598* expressed in *E. coli* led to the synthesis of mycosporine-glycine where *NpR5598* could be confirmed as a homolog of *Ava_3856* by catalysing the condensation of glycine onto 4-deoxygadusol. However, the fourth gene namely *NpF5597* was noted to be different from the fourth gene namely *Ava_3855* in *A. variabilis*. Due to the fact that *NpF5597* enzyme was insoluble for enzymatic activity testing *in vitro*, the experiment was done *in vivo* where the host was made to contain both plasmids bearing genes *NpR5600* to *NpR5598* for the production of significant amount of mycosporine-glycine to be used as substrate as well as the one containing genes *NpR5600* to *NpF5597* where various iminomycosporines namely the most abundant shinorine (m/z = 333.17), mycosporine-2-glycine (m/z = 303.16) as well as porphyra-334 (m/z = 347.15) were produced and confirmed by UV-Vis and LC-MS. The enzyme NpF5597 was seen to differ from the NRPS-like enzyme AV_3855 in *A. variabilis* and was thus attributed to another enzyme family termed as ATP- grasp ligases (D-ala D-ala ligase) based on sequence similarity where a plausible explanation would be that the latter activated mycosporine-glycine by phosphorylation to allow the addition of L-serine to the activated precursor, leading to the formation of shinorine[81].

1.4.2 MAA biosynthesis in Halotolerant Cyanobacteria

Aphanothece halophytica

The halotolerant cyanobacterium *Aphanothece halophytica* is an extremophile which can survive under specific conditions like high salinity and alkaline pH. The MAA mycosporine-2-glycine (λ_{max} = 331 nm) with m/z 303.02 was revealed to be produced from *Aphanothece halophytica* where the yield was seen to be higher under more saline conditions. Four genes were found to be involved in the biosynthesis of mycosporine-2-glycine based on shotgun sequence analysis of *Aphanothece halophytica* as well as a homology search to the MAA clusters of *Anabaena variabilis* (*Ava_3858* to *Ava_3855*) and *Nostoc punctiforme* (*NpR5600* to *NpR5598* and *NpF5597*). Three genes namely *Ap3857*, *Ap3856* and *Ap3855* were found to be homologous to *Ava_3857/NpR5599*, *Ava_3856/NpR5598* and *NpF5597* respectively where *Ap3857* was thus considered to encode a O-Methytransferase (Ap-OMT),

Ap3856 encoding a C-N ligase (Ap-CNligase) and *Ap3855* encoding a D-Ala-D-Ala ligase (Ap-AAligase). Those three genes were seen to be located close to each other, in comparison to *Ap3858* (Ap-DDG), which was situated far from the upper area of the *Ap3857* gene. In other words, there was a gap between *Ap3858* and *Ap3857* which might be attributed to an unknown N-terminal domain (154 amino acid residues). The four MAA genes including *Ap3858* to *Ap3855* were cloned and heterologously expressed in *E. coli* as well as in a freshwater cyanobacterium namely *Synechococcus elongatus* PCC7942. Mycosporine-2-glycine was seen to be significantly produced when *E. coli* as well as in *Synechococcus elongatus* cells were grown under an increased saline condition, leading to the fact that the four putative MAA genes might be induced by salt upshock. In simpler terms, Ap-OMT might have catalysed the reaction of demethyl 4-deoxygadusol (DDG) to 4-deoxygadusol which was afterwards catalysed by Ap-CNligase in the presence of glycine to form mycosporine-glycine. Based on substrate specificity, glycine was utilized by Ap-AAligase for the last step involving the condensation of glycine onto mycosporine-glycine resulting into the MAA product mycosporine-2-glycine which apart from a UV protector, might as well be an osmoprotectant defending against salt stress[82].

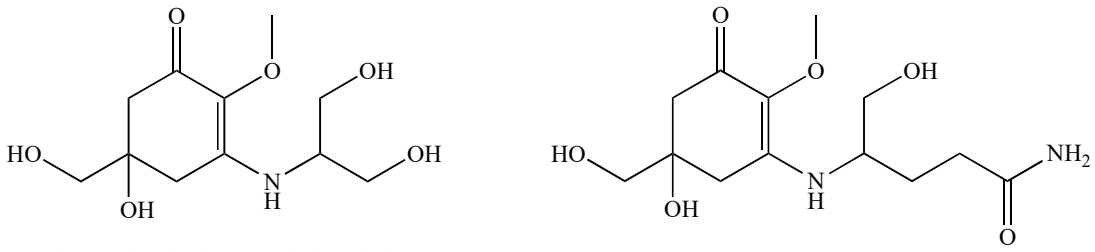
1.4.3 MAA biosynthesis in terrestrial Gram-positive bacteria

Actinosynnema mirum DSM 43827 and *Pseudonocardia* sp. strain P1 are two Gram-positive bacteria from the order Actinomycetales that were seen to possess a mycosporine-like amino acid gene cluster homologous to the biosynthetic gene clusters found in cyanobacteria. The biosynthetic gene clusters in *A. mirum* [(*amir_4256-4259*) (*mysA-mysD*)] and in *Pseudonocardia* sp. P1 [(*pseP1_010100031425-010100031440*) (*mysA-mysD*)] consisted of four genes encoding putative dimethyl 4-deoxygadusol (DDG) synthase, O-methyltransferase (O-MT), ATP-grasp family protein, and D-Ala D-Ala ligase homolog. The first three genes of *A. mirum* DSM 43827 (*Amir_4259–Amir_4257*) and *Pseudonocardia* sp. P1 (*PseP1_010100031440–PseP1_010100031430*) showed similarity to *Ava_3858–Ava_3856* of *A. variabilis* ATCC 29413 whereas the fourth gene of *amir_4256* and *pseP1_ 010100031425* were closer to D-Ala D-Ala ligase-like protein of *Nostoc punctiforme* ATCC 29133. The entire gene cluster for MAA biosynthesis from *A. mirum* DSM 43827 (6.3 kb) and *Pseudonocardia* sp. P1 (8.2 kb), were cloned in

Streptomyces avermitilis SUKA22 under a constitutively expressed promoter by an *in vivo* cloning strategy mediated by λ-RED recombination. After culturing *A. mirum* DSM 4382, *Pseudonocardia* sp. P1, and *S. avermitilis* SUKA22 transformants in various liquid media under diverse conditions, shinorine was detected in *Pseudonocardia* sp. P1 grown in Trypticase soy broth medium whereas no MAA-like compound was detected from *A. mirum* DSM 43827, indicating that the gene cluster might be in a cryptic state. Three compounds absorbing at 333 nm (*m/z* = 333.1288); 333 nm (*m/z* = 317.1334) and 334 nm (*m/z* = 347.1454) were attributed to shinorine, mycosporine-glycine-alanine (first time discovered) and porphyra-334 respectively in the transformants carrying *mysA-mysD* genes of *A. mirum* DSM 43827 placed under the control of the promoter along with the presence of the biosynthetic intermediates namely 4-deoxygadusol and mycosporine-glycine. On the contrary, only shinorine was produced from the *S. avermitilis* SUKA22 transformants carrying the entire *mys* gene cluster of *Pseudonocardia* sp. P1. One of the plausible reasons for the production of a low yield of porphyra-334 and mycosporine-glycine-alanine from the transformant encoding the *mysD* gene from *A. mirum* DSM 4382 might be due to the substrate chosen by the D-Ala D-Ala ligase homolog which was seen to mainly catalyse the condensation of serine to mycosporine-glycine generating the dominant MAA shinorine[83].

1.4.4 MAA biosynthesis in fungi

The percentage of MAAs found in MAAs producing organisms does not exceed 1% of their dry weight[36]. Cells greater than 100 µm have a definite advantage over those with sizes ranging from 10 to 100 µm regarding the accumulation of those photoprotective molecules. Cells which are lower than 10 µm in size exhibit little protection from the only MAAs dissolved in the cytoplasm[84,85]. Fungi require UV light for their reproductive organs formation and these UV lights are detected by light-induced pigments absorbing at 240 and 310 nm. The primary compound namely 2-methoxy-3-bis(hydroxymethyl)methylamino-5-hydroxymethyl-2-cyclohexen-1-one[86] absorbing at 310 nm was isolated from the sporophores of the basidiomycete *Stereum hirsutum*[87].



2-methoxy-3-bis(hydroxymethyl)methylamino-5-hydroxymethyl-2-cyclohexen-1-one

Mycosporine glutaminol

Figure 1.9: Chemical structures of 2-methoxy-3-bis(hydroxymethyl)methylamino-5-hydroxymethyl-2-cyclohexen-1-one and mycosporine glutaminol isolated from fungi.

Basidiomycetes, *Zygomycetes*, *Ascomycetes* and *Deuteromycetes* are common fungal classes which contains mycosporines whereas fungal classes like *Sporidiobolales* and *Wallemiales* lack those UV absorbing molecules[88,89]. Both terrestrial as well as aquatic fungi are known to produce mycosporine where some might be saprophytic or phytopathogenic or others opportunistic. They are scattered around temperate regions as well as in harsh environments such as hypersaline conditions, deserts, or the polar areas. *Pyronema omphalodes* cultures gave rise to various mycosporines such as nor-mycosporine-glutamine in the primordia along with the more stable mycosporine-glutaminol-glucoside in the mature ascoma[90]. The phytopathogenic anamorphic fungus *Ascochyta fabae* produced mycosporine during the reproduction process which were seen to be accumulated in spores[91] which could thus show that these UV-absorbing compounds are generally sensed in sporulating mycelia[34,87,91] depending on the irradiation wavelength, length of time of irradiation as well as the media composition of cultures. In addition, mycosporines have also been detected in the extracellular matrix and the outer cell wall layers of extremophilic microcolonial fungi[92].

Not much is known about the biosynthesis of MAAs in fungi except from the fact that they are important for sporulation in some of them. The deuteromycete *Trichothecium roseum* was revealed to generate a high amount of mycosporine glutaminol during sporulation. Incorporation of labelled compounds namely [1-^{14}C] acetic acid, [1-^{14}C] glyceric acid and [U-^{14}C] 3-DHQ in the cultures of *Trichothecium roseum* showed that 3-DHQ originated from the first section of the shikimate pathway and acted as a precursor in mycosporine production[93].

1.4.5 MAA biosynthesis in symbiotic dinoflagellates

Reef-building corals are predominantly found in shallow, transparent, oligotrophic water providing habitats for more than 25% of all known marine organisms. Corals of the order Scleractinian coral species share a symbiotic relationship with unicellular photosynthetic dinoflagellates of the genus Symbiodinium and those symbiotic dinoflagellates were reported to produce mycosporine-glycine, mycosporine-2-glycine, shinorine, porphyra-334 and palythine[94,95]. The Symbiodinium expressed sequence tag database was used to identify sequences corresponding to MAA biosynthetic genes where their accession numbers were found using BLAST search. Homologs of two O-methyltransferases genes [Genebank accession number: FE864321 and FE865550] were detected where the transfer of methyl group by O-methyltransferases catalysis are proposed to be involved in the MAA biosynthesis pathway[77,80] as well as gene sequence encoding homolog of ATP-grasp were also seen in *Symbiodinium* sp. [Genebank accession number: EH058154] with conserved domain similar to the cyanobacterium *A. variabilis*[77] which are known to catalyse ATP-dependent synthesis. Furthermore, homolog of a NRPS-like enzyme was discovered in the symbiotic dinoflagellate *Symbiodimium* sp. [Genebank accession number: EH058166, EH037827] where the latter are anticipated to be involved in the adenylation of amino acid by catalysing the formation of substrate-coenzyme A [96]. The similarity in gene sequences involved in MAA biosynthesis between *Symbiodinium* sp. and other species like cyanobacteria or others[77] showed that symbiotic dinoflagellates demonstrated a monophyletic clustering with their bacterial and algal homologs thus proving their microbial origin based on phylogenetic analyses of their putative MAA biosynthetic genes[96].

1.5 Brief Introduction to Non-Ribosomal Peptide Synthetases (NRPSs)

A diverse array of secondary metabolites, especially peptides are synthesized by non-ribosomal multi-modular peptide synthetases. The enzymes are built up with modules which leads to the determination of the primary product structure. Each module consists of various individual catalytic domains with each having its specific enzymatic reaction which contribute to the entire process of selecting and covalently tether a building block involving a condensation reaction with the nascent chain, thus producing specific substructures. In general, NRPSs are involved into specific amino

acids activation by converting them into their respective aminoacyl thioesters leading to the succeeding peptide bond formation between the activated amino acids. During this process, both proteinogenic as well as nonproteinogenic amino acids can be activated and incorporated into natural products by NRPSs[97].

1.5.1 General mechanism of NRPSs system

Each module consists of an adenylation (A) domain which activates the corresponding substrate from a vast variety of substrates and tethers them as thioesters to the peptidyl carrier protein (PCP) domain next to it. The domains are more prone in picking up L-amino acids and can be very specific in activating the proteinogenic amino acids and a diverse array of nonproteinogenic amino acids along with various other carboxylic acids like aryl acids for the assembly of the metabolic natural product. The epimerization domains might also be present in some modules as well as the condensation domain which are greatly selective toward the alpha-carbon epimer of the aminoacyl thioester present in natural products[97]. The NRPS system also comprises of a stand-alone thioesterase (TE) domain which catalyses the release of the assembled product from the carrier protein domain of the last module. There are two ways in which the TE domain is known to carry out this chain release reaction. Firstly, is the hydrolysis or intermolecular condensation with a soluble amine which give rise to linear metabolic products with a carboxyl terminus while the second possibility is the formation of an ester bond by an intramolecular amide resulting into cyclic natural products[98,99].

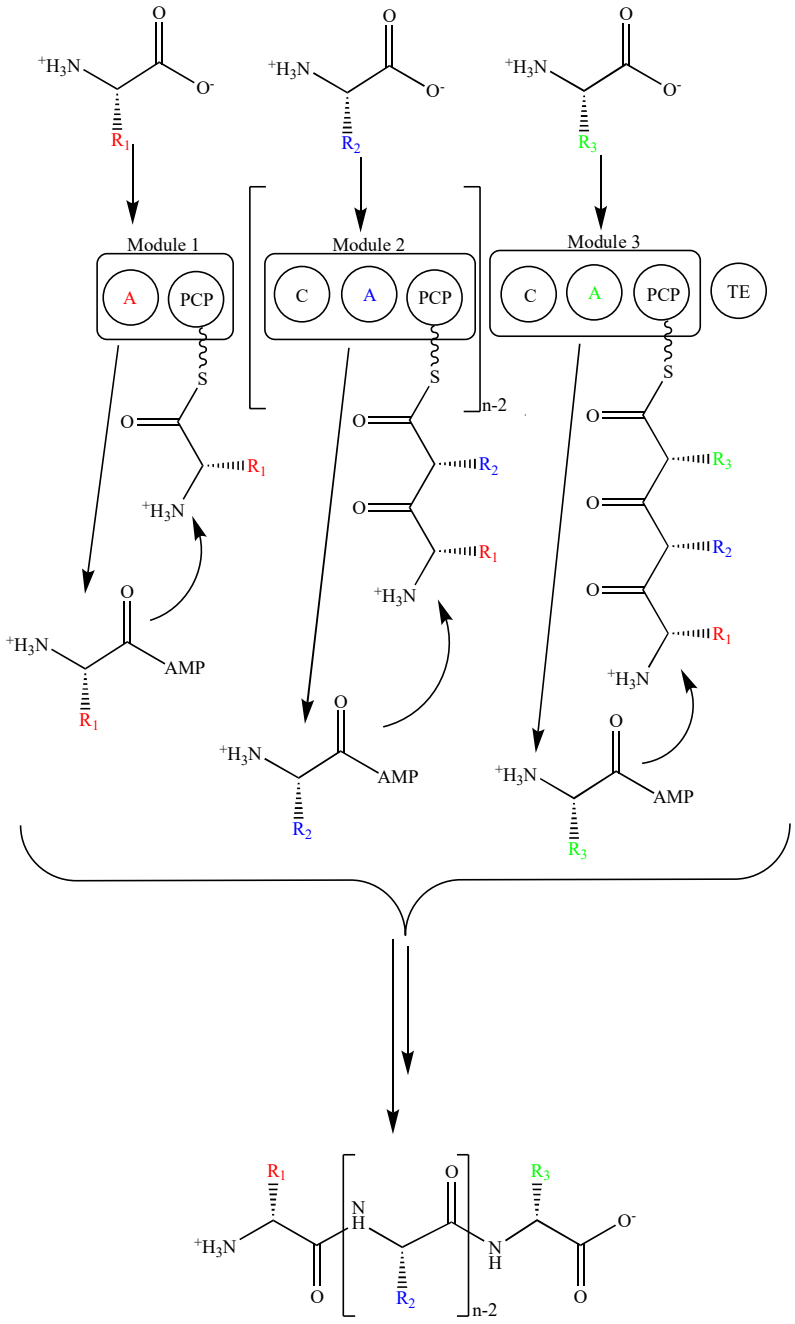


Figure 1.10: NRPSs mechanism: The adenylation (A) domain specifically chooses its substrates and tethers them to the peptidyl carrier protein (PCP) by the formation of a thioester bond. The A and PCP domains are always present in NRPS modules whereas the condensation (C) is normally present in chain extension modules. The epimerization (E) and methyltransferase (MT) domains are optional where the E domain is usually involved into a change in configuration of one chiral centre and the MT is involved into the addition of a methyl group. The thioesterase (TE) domain is ultimately involved into catalysing the release of the peptide product from the NRPS.

1.5.2 Specificity of Adenylation domains in NRPSs system

The adenylation domains are regarded as the primordial commander of substrate specificity in NRPSs[100]. The A domain is present in each module of the NRPSs with the explicit duty of recognizing amino acids from a cellular pool followed by their chemical activation through adenylation. The adenylated substrate is then transferred to the adjacent peptidyl carrier protein forming a thioester bond with the phosphopantetheine thiol by a condensation reaction. Based on literature, more precisely, the crystal structure of the A domain PheA activating L-Phe in gramicidin S synthetase GrsA, it was revealed that 10 amino acids residues were responsible for the activation of L-Phe in PheA[101–103]. An assortment of techniques such as sequence analysis, biochemical characterisation or online software like pks/nrps analysis (http://nrps.igs.umaryland.edu) or even antiSMASH bacterial version (https://antismash.secondarymetabolites.org/#!/start) are useful tools to depict substrate specificity in order to help predict structures of novel non-ribosomal peptides.

1.5.3 Plausible mechanisms of NRPS-like enzyme for the production of shinorine in *Anabaena variabilis*

This study focuses on the fourth enzyme namely the NRPS-like enzyme, here referring to Ava_3855 which consist of an adenylation (A), peptidyl carrier protein (PCP) and thioesterase (TE) domains[104].

Figure 1.11: Domains of Ava_3855 in the terrestrial cyanobacterium *Anabaena variabilis*.

There are 2 ways in which the substrate serine might be attached to the precursor mycosporine-glycine in order to produce shinorine. The first plausible mechanism might involve the TE domain in the imine formation where an enol ester intermediate arises due to rearrangement of O to N from a 1,4 addition of the serine nitrogen to the

activated cyclohexenimine core. In this aspect, the precursor mycosporine-glycine is acting as a strong nucleophile. The second plausible mechanism is mycosporine-glycine acting as a weak nucleophile where a direct condensation reaction occurs followed by hydrolysis of the chain reaction by the TE domain[77].

Plausible Mechanism 1:

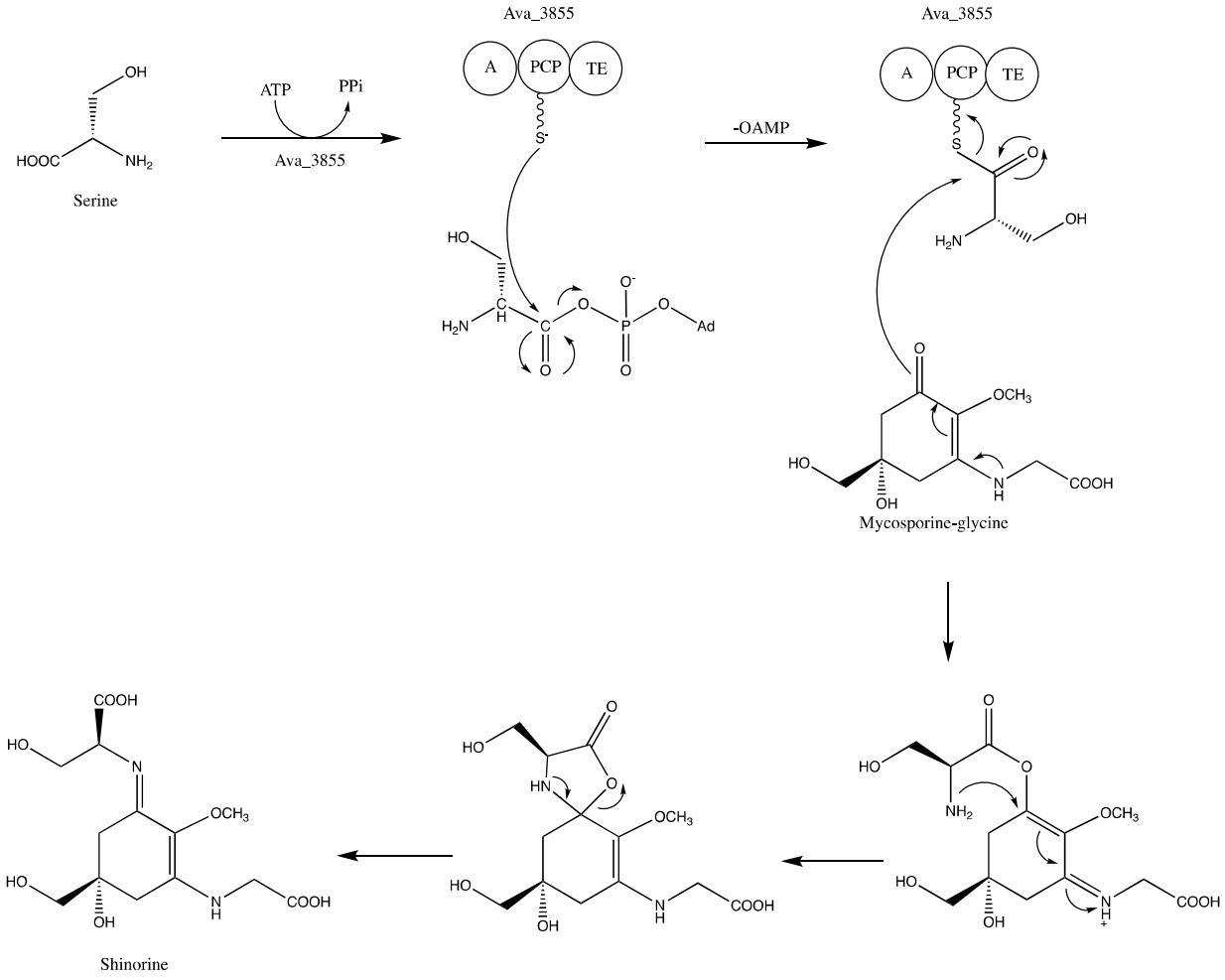


Figure 1.12: The precursor mycosporine-glycine acting as a strong nucleophile. (A = adenylation domain; PCP = peptidyl carrier protein; TE = thioesterase domain).

Plausible Mechanism 2:

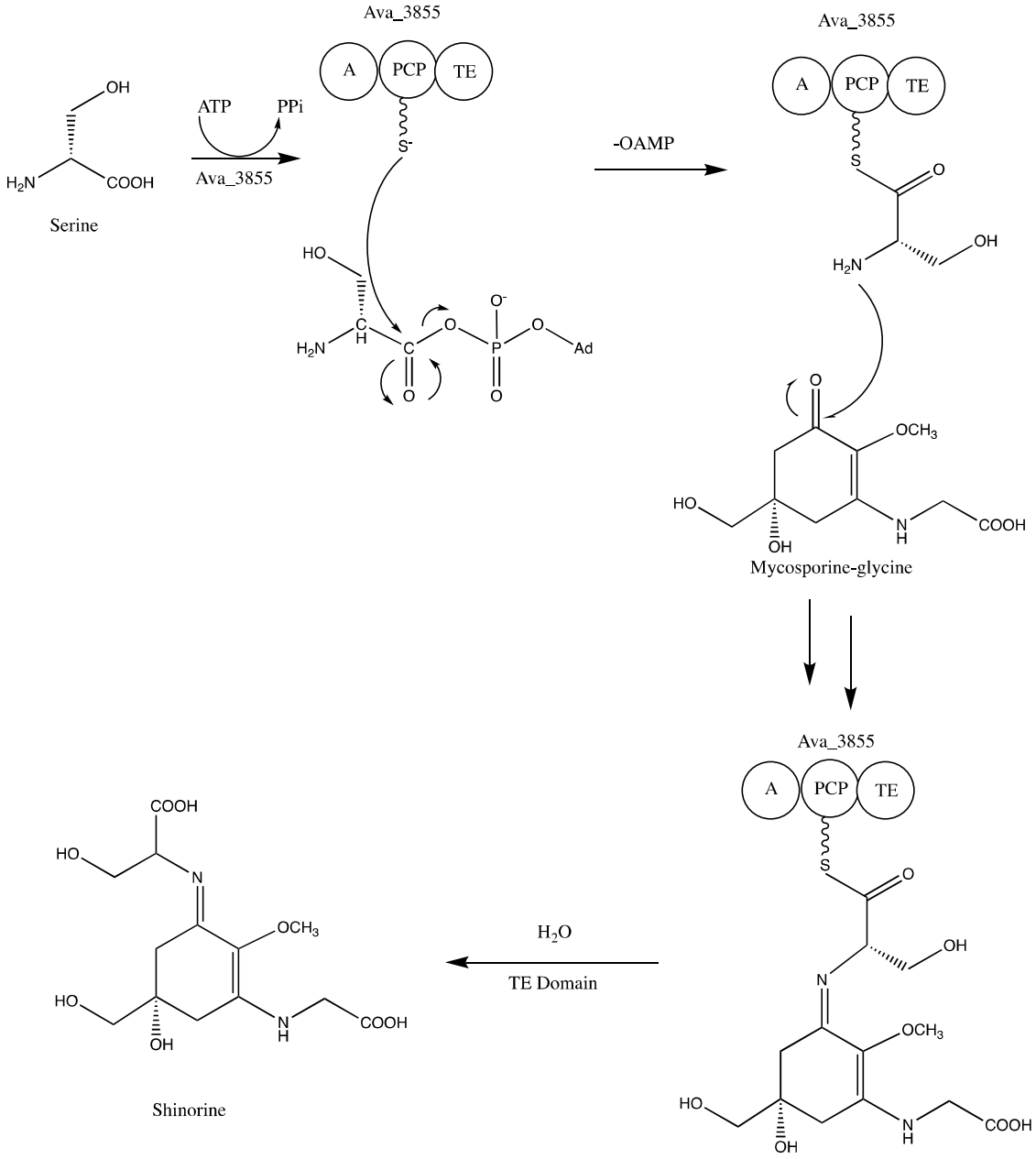


Figure 1.13: The precursor mycosporine-glycine acting as a weak nucleophile. (A = adenylation domain; PCP = peptidyl carrier protein; TE = thioesterase domain).

1.6 Industrial importance of photoprotective molecules

Sunlight as described thoroughly above has its pros and cons. Consumers are being more and more cautious about their skin needs that, cosmetic industries are revolutionising quite abruptly. Mechanical stress like sunburns, premature skin ageing, melanoma and non-melanoma skin cancers are some examples of the detrimental

effect of the harmful UV rays on the skin. Every year, in November, the Australasian College of Dermatologists together with the Cancer Council organize the "National Skin Cancer Action week" to promote awareness of skin cancers caused by the sun as well as the necessity for their early detection[105]. Moisturizers, skin creams, makeup products, body lotions, and other care products are being scrutinized by consumers nowadays about the ingredients integrated as well as the presence of SPF which have become trendy on the market. The sunscreen market is a fast growing one where the global sun care value is expected to go beyond 24 billion US dollars by 2029[106]. Some commercially available sunscreens contain ingredients like avobenzone, which is known to be photo-unstable[107]. Furthermore, upon photoexcitation of avobenzone, a high proportion of the photoexcited population was not seen to return to its original ground state along with formation of numerous photoproducts were detected[108–111]. In addition, the commercially available oxybenzone has been associated with coral bleaching and hence banned in various countries[112]. To remediate to this, eco-friendly, affordable, photostable, nature inspired photoprotective molecules are needed.

1.7 Aims and objectives of Research

This study focuses on the purification of MAAs from Helioguard 365, the exploration and exploitation of naturally occurring MAAs from the Rhodophyte *Palmaria palmata* in order to elucidate their photoprotective mechanism by tracking their energy flow and dynamics after photoexcitation as well as the development of NRPS-like biocatalysts for MAAs production using *E. coli* as a heterologous host.

Objectives of Research

The objectives of this research are:

[1] The purification and characterisation of the MAAs shinorine and porphyra-334 from the commercially available Helioguard[365].

[2] The extraction, purification, and characterisation of known MAAs, especially the cis-trans isomers, usujirene/palythene from the rhodophyte *Palmaria palmata*.

The is the first time that the photoprotective mechanism of these purified MAAs has been elucidated using TEAS by tracking their energy flow and dynamics after photoexcitation.

[3] The exploration of the NRPS-like genes involved in MAAs production in cyanobacteria using Bioinformatics

[4] The cloning of 5 selected NRPS-like genes from *Anabaena variabilis*, *Chlorogloeopsis fritschii*, *Chryseobacterium indologenes*, *Moorea producens* and *Scytonema hofmanni* involved in MAA production in pET151 using Directional TOPO cloning for expression and purification.

[5] The expression and purification of the NRPS-like enzyme Ava_3855 followed by confirmation and functionality of the protein using intact protein mass spectrometry

[6] The chemo-enzymatic assay using Ava_3855 and a synthetic substrate namely AC23 to monitor the ATP-dependent condensation reaction.

[7] The development of a novel biocatalyst by an A-domain swap from the NRPS-like gene Ava_3855 from *Anabaena variabilis* to the A-domain of the NRPS-like gene from *Moorea producens*, assessing its prediction of selecting and activating L-proline as substrate.

[8] The cloning of the recombinant NRPS-like gene in pET28MBP vector followed by the expression and purification of the new engineered enzyme MBPNEE

[9] The chemo-enzymatic assay using MBPNEE and a synthetic substrate namely AC23 to monitor the ATP-dependent condensation reaction involving the predicted L-proline activation.

[10] A proposed role of the TE domain in Ava_3855 and MBPNEE in comparison to other NRPS-like enzymes.

Chapter 2: Materials and Methods

2.1 Bacterial strains and plasmids

Table 2.1: *Escherichia coli* strains used

E. coli	**Reference**
One Shot™ TOP10 Chemically Competent *E. coli*	ThermoFisher Scientific
BL21 Star™ (DE3) One Shot™ Chemically Competent *E. coli* (From Champion™ pET151 Directional TOPO™ Expression Kit)	ThermoFisher Scientific
NEB® Turbo Competent *E. coli*	New England Biolabs
NEB 5-alpha Competent *E. coli* (From Q5® Site-Directed Mutagenesis Kit)	New England Biolabs
NEB® 5-alpha Competent *E. coli* (From NEBuilder® HiFi DNA Assembly Cloning Kit)	New England Biolabs
BAP1 – *E. coli* strain	B.A. Pfeifer, S.J. Admiraal, H. Gramajo, D.E. Cane, C. Khosla. Biosynthesis of complex polyketides in a metabolically engineered strain of E. coli. Science 291, 1790-2 (2001)

Table 2.2: Inserts used

Species Accession ID	Species	NRPS-like Gene ID
CP000117.1	*Anabaena variabilis* ATCC 29413	ABA23460.1 WP_011320550.1
NZ_AJLN01000152.1	*Chlorogloeopsis fritschii* PCC 6912 contig00157	WP_016876762.1
NZ_CP018786.1	*Chryseobacterium indologenes* strain AA5	WP_079242079.1
NZ_CP017599.1	*Moorea producens* PAL-8-15-08-1	WP_202972154.1 WP_070393404.1
NZ_KK073768.1	*Scytonema hofmanni* UTEX 2349 Tol9009DRAFT_TPD.8	WP_038297685.1

Table 2.3: Recombinant plasmids

Vector	Insert (*E. coli* codon-optimised)	Insert size	Recombinant plasmid	Recombinant plasmid size
pET151	Ava_3855	2637 bp	pAna	8400 bp
pET151	UYC_RS0133560	4047 bp	pChloro	9810 bp
pET151	BUE84_RS13530	3885 bp	pChryseo	9648 bp
pET151	BJP34_RS17220	4014 bp	pMoo	9777 bp
pET151	TOL9009_RS40645	3411 bp	pScy	9174 bp
pET151	NEE	2649 bp	pNEE	8412 bp
pET28MBP	NEE	2649 bp	pMBPNEE	9087 bp

2.2 Culture media

2.2.1 Solid media

Luria-Bertani (LB) Agar (Lennox)

Difco Peptone	10 g

Difco Yeast extract	5 g
Sodium Chloride	5 g
Agar	12 g
Distilled water	up to 1000 ml

2.2.2 Liquid media

Luria-Bertani (LB) broth

Difco Peptone	10 g
Difco Yeast extract	5 g
Sodium Chloride	5 g
Distilled water	up to 1000 ml

2.3 Antibiotics

The table below shows the typical antibiotic concentrations in media used for appropriate *E. coli* selection.

Table 2.4: Antibiotics used during this study

Antibiotic	**Stock Solution (mg/ml)**	***E. coli* culture in LB (µg/ml)**
Kanamycin	50	50
Ampicillin	100	100

2.4 Buffers and solutions

2.4.1 Protein lysis buffer (Buffer A) at pH = 8 [For Ava_3855]

Component	**Concentration**
Tris-HCl	20 mM
NaCl	500 mM
$MgCl_2$	10 mM

2.4.2 Protein lysis buffer (Buffer B) at pH = 8.5 [For MBPNEE]

Component	**Concentration**
Tris-HCl	20 mM
NaCl	500 mM
$MgCl_2$	10 mM

2.4.3 Protein elution Buffer C at pH = 8 [For Ava_3855]

Component	Concentration
Tris-HCl	20 mM
NaCl	500 mM
$MgCl_2$	10 mM
Imidazole	5 mM

Protein elution buffer (Buffer C with different imidazole concentration)
The elution of protein from Ni-NTA was done using a stepwise imidazole gradient in buffer C starting from 25 mM, 50 mM, 75 mM, 100 mM, 125 mM, 150 mM and 200 mM imidazole concentration.

2.4.4 Protein elution Buffer D at pH = 8.5 [For MBPNEE]

Component	Concentration
Tris-HCl	20 mM
NaCl	500 mM
$MgCl_2$	10 mM
Imidazole	5 mM

Protein elution buffer (Buffer D with different imidazole concentration)
The elution of protein from Ni-NTA was done using a stepwise imidazole gradient in buffer D starting from 25 mM, 50 mM, 75 mM, 100 mM, 125 mM, 150 mM and 200 mM imidazole concentration.

2.4.5 Protein storage buffer E at pH = 8 [For Ava_3855]

Component	Concentration
Tris-HCl	20 mM
NaCl	50 mM
Glycerol	5% v/v

2.4.6 Protein storage buffer F at pH = 8.5 [For MBPNEE]

Component	Concentration
Tris-HCl	20 mM
NaCl	50 mM
Glycerol	5% v/v

2.5 Usage of *Escherichia coli* as host

2.5.1 Growth conditions of *E. coli*

E. coli was grown at 37ºC overnight either on LB agar plates or LB broth on the shaker with the appropriate antibiotics.

2.5.2 Storage of *E. coli* strains as glycerol stock

Stock of *E. coli* strains were prepared with 500 µl of the overnight LB cultures with the corresponding antibiotics mixed with 500 µl of 50% v/v glycerol and stored at -80ºC.

2.5.3 Preparation of electrocompetent *E. coli* cells

The desired *E. coli* strain to be used was streaked on LB plates and incubated at 37ºC overnight. One colony was picked up to inoculate 10 ml LB and incubated at 37ºC on a shaker overnight. 500 µl of the overnight culture was inoculated into 50 ml LB and left to grow at 37ºC on a shaking incubator at 180 rpm until an OD_{600} of about 0.5 was

reached. The culture was then transferred to an ice-cold 50 ml falcon tube and stored on ice for 10 minutes. The cell pellets were recovered after centrifugation at 4000 rpm for 10 minutes at 4ºC. The supernatant was discarded, and the pellet was gently resuspended in ice-cold water and centrifuged at 4000 rpm for 10 minutes at 4ºC. The supernatant was discarded, and the pellet was gently resuspended in ice-cold 20% glycerol followed by centrifugation at 4000 rpm for 10 minutes at 4ºC. The supernatant was discarded, and the pellet was gently resuspended in ice-cold 20% glycerol and was then aliquoted (50 µl) into separate 1.5 ml microcentrifuge tubes, flash frozen in liquid nitrogen and finally stored at -80ºC.

2.5.4 Preparation of chemically competent *E. coli*

The *E. coli* strain to be used was streaked on LB plates and incubated at 37ºC overnight. The following day a single colony was picked up to inoculate 10 ml LB and placed on a shaking incubator at 37ºC overnight. 500 µl of the overnight culture was inoculated into 50 ml LB and left to grow at 37ºC on a shaking incubator at 180 rpm until an OD_{600} of about 0.5 was reached. The culture was then transferred to an ice-cold 50 ml falcon tube and stored on ice for 10 minutes. The cell pellets were recovered after centrifugation at 4000 rpm for 10 minutes at 4ºC. The supernatant was discarded, and the cell pellets were resuspended in 10 ml ice-cold 30 mM $CaCl_2$ and stored on ice. The cell pellets were recovered after centrifugation at 4000 rpm for 10 minutes at 4ºC. The supernatant was discarded, and the pellet was resuspended in ice-cold $CaCl_2$ and aliquoted (50 µl) into separate 1.5 ml microcentrifuge tubes, flash frozen in liquid nitrogen and finally stored at -80ºC.

2.5.5 Transformation of electrocompetent *E. coli* cells

The electrocompetent cell was transferred on ice and allowed to thaw. Meanwhile, 3-5 µl of the desired DNA was added to an ice-cold electro-cuvette followed by 50 µl of the thawed electrocompetent cells. The cuvette was placed in the electroporator, and electroporation was carried out using the following conditions: 25 µF capacitor, 200 ohms resistance, 2 kV voltage (for 1 mm cuvette). After the electroporation process, 950 µl of LB was immediately added and transferred to a 1.5 ml Eppendorf tube. The tube was then incubated on a shaking incubator at 37ºC for 1 hour. The

culture (50 µl to 200 µl) were plated on each LB agar plates containing the appropriate antibiotics and the latter were incubated in the 37ºC incubator overnight.

2.5.6 Transformation of chemically competent *E. coli* cells

The chemically competent cell was transferred on ice and allowed to thaw. DNA (3-5 µl) was added to the thawed competent cell under sterile conditions and mixed by gently flicking the tube. The tube was stored on ice for 30 minutes and was gently flicked at half time. The tube was then transferred to a water bath preheated to 42ºC and incubated for 45 seconds. The tube was transferred back on ice for 5 minutes and 950 µl of LB was afterwards added and mixed thoroughly. The mix was incubated on a shaking incubator at 37ºC for 1 hour. The culture (50 µl to 200 µl) were plated on each LB agar plates containing the appropriate antibiotics and the latter were incubated in the 37ºC incubator overnight.

2.6 Plasmid isolation from *E. coli*

Plasmid isolation (Miniprep) was done using the ThermoFisher Scientific GeneJET Plasmid Miniprep kit. The cell pellet from an overnight 5 ml *E. coli* culture in LB with the appropriate antibiotics was centrifuged for purification of high-copy plasmids. The pelleted cells were resuspended completely by pipetting up and down in 250 µl resuspension buffer solution. The cell suspension was then transferred to an Eppendorf tube. A volume of 250 µl lysis buffer was added and mixed thoroughly by inverting the tube 6 times until the solution became viscous followed by the addition of 350 µl of the neutralization solution and the latter was thoroughly mixed immediately by inverting the tube 6 times. The tube was spined for 5 mins to pellet the cell debris and DNA. The supernatant was then transferred to the GeneJET spin column, making sure to avoid disturbing the white precipitate. The tube GeneJET spin column was centrifuged for 1 min; the flow-through was discarded and the column was placed back into the same collection tube. 500 µl of the wash solution was added to the GeneJET spin column and centrifuged for 1 min. The column was placed back into the same collection tube after the flow-through was discarded. The wash process was carried out twice. The GeneJET spin column was then transferred to a fresh 1.5 ml Eppendorf tube where 50 µl of the elution buffer was added to the centre of the column

membrane. The latter was incubated for 2 mins at room temperature and centrifuged for 2 mins for elution of the plasmid DNA.

2.7 DNA analysis, *In vitro* manipulation and Quantification of DNA

2.7.1 Agarose gel electrophoresis

Agarose gels were prepared and run in 1X TAE (Tris-acetate-EDTA) buffer. 1% agarose gels were usually used for electrophoresis. The gels were mainly stained using GelRed nucleic acid gel stain which is ultra-sensitive and environmentally friendly. The NEB 1 kb DNA ladder ranging from 500 bp to 10 kb in size was used for its suitability for both small and large DNA fractions.

2.7.2 DNA fragments recovery from agarose gel

The Monarch® DNA gel extraction kit from NEB was used for DNA fragments recovery from agarose gel. The DNA fragment was excised from the agarose gel using a clean razor blade and transferred to a 1.5 ml Eppendorf tube. The mass of the gel slice was calculated, and 4 volumes of the monarch gel dissolving buffer was added to the gel slice (400 µl buffer per 100 mg agarose). The sample was incubated at 50ºC, inverting periodically until the gel slice was completely dissolved. The dissolved sample was then loaded onto the column and spined for 1 min. The flow-through was discarded and 200 µl of DNA wash buffer was added to the column and spined for 1 min. The above washing process was repeated twice. The column was afterwards transferred to a fresh 1.5 ml microcentrifuge tube and 10 µl of elution buffer was added to the centre of the column membrane. The latter was incubated for 1 min at room temperature and spined for 1 min for elution of the corresponding DNA.

2.7.3 DNA digestion using restriction enzymes

Plasmid DNA digestion was carried out based on the corresponding restriction enzymes used along with their protocol including the corresponding buffers; the enzyme's optimum incubation temperature; the enzyme's digestion period and the enzyme's heat inactivation period if applicable. The total volume of 10 µl was used for a typical restriction digest for DNA check [1 µl plasmid DNA; 1 µl buffer; 0.5 µl restriction endonuclease and 7.5 µl water].

2.7.4 Quantification of DNA and protein

DNA was quantified using the NanoDrop™ 2000 Spectrophotometer which is the only microvolume UV-Vis spectrophotometer scanning the full spectrum (190-840 nm) with sample volume as low as 0.5 µl. The sample is usually pipetted directly onto an optical measurement surface and the measurement of DNA sample at A260 while the measurement of protein sample at A280 are both made within 5 seconds.

2.8 PCR protocols

2.8.1 General PCR

PCR from plasmid DNA was typically carried out with 1-25 ng/ µl as template.

Table 2.5: General PCR

Components	25 µl Reaction volume	Final concentration
Q5® High-Fidelity 2X Master Mix	12.5	1X
10 µM Forward primer	1.25	0.5 µM
10 µM Reverse primer	1.25	0.5 µM
Template DNA	1	1-25 ng
Nuclease-free water	9.0	

Table 2.6: General PCR cycle conditions

Step	Temperature	Time	
Initial Denaturation	98ºC	30 seconds	25 cycles
Denaturation	98ºC	10 seconds	
Annealing	50-72ºC*	30 seconds	
Extension	72ºC	30 seconds/kb	
Final extension	72ºC	2 minutes	
Hold	4ºC		

*Depending on the primers

2.8.2 Gradient PCR

Gradient PCR is a technique which permit the empirical determination of an optimal annealing temperature using the least number of steps. The thermocycler provides a gradient function allowing 12 different annealing temperatures in one single run which result in higher chance of proper annealing temperatures, meaning higher primers binding affinity to the template DNA.

2.8.3 Colony PCR

Colony PCR was used mainly as a quick screening of recombinant vectors after cloning. A general PCR was usually carried out with the exception of using a bacterial colony instead of template DNA with an extended initial denaturation period.

2.8.4 PCR product clean-up

The Monarch® PCR & DNA Clean up Kit was used for PCR clean-up. The PCR sample was diluted with DNA clean up binding buffer by pipetting up and down properly. The sample was loaded onto the insert column and spined for 1 min. The flow-through was discarded and 200 µl of DNA wash buffer was added and spined for 1 minute. The flow-through was discarded and the above washing step was repeated. The column was afterwards transferred to a fresh 1.5 ml Eppendorf tube and 10 µl of elution buffer was added to the centre of the column membrane. The latter was incubated for 1 min at room temperature and spined for 1 min for elution of the corresponding DNA.

2.9 Sanger Sequencing

All cloned plasmids; engineered plasmids or PCR products were sent for sanger sequencing for confirmation of correct constructs or sequence. 5 µl of plasmid DNA with 5 µl of the corresponding primer were sent for next day sanger sequencing.
The sequencing data were then uploaded and analysed on Benchling by applying sequence alignment between the template and the corresponding sequencing data.

2.10 Directional TOPO Cloning in pET151

The Champion™ pET151 Directional TOPO™ Expression Kit with BL21 Star™ (DE3) One Shot™ Chemically Competent *E. coli* was used. Five of the eight synthetic genes were cloned in respective pET151/D-TOPO® vector. This kit utilizes a highly efficient, 5-minute cloning strategy ("TOPO® Cloning") to directionally clone a blunt-end PCR product into the vector for high-level, T7-regulated expression in *E. coli*. This type of cloning makes use of the enzyme topoisomerase I from *Vaccinia* virus which binds after 5'-CCCTT in one strand of the vector and cleaves the phosphodiester backbone forming a covalent bond between the 3' phosphate of the cleaved strand and a tyrosyl residue (Tyr-274) of topoisomerase I.

Forward primers were designed by addition of 4 bases (CACC) while a stop codon was added in all reverse primers followed by amplification of the synthetic DNA sequences by PCR.

2.10.1 Primer design for synthetic genes and PCR

Table 2.7: Primer design for Directional TOPO cloning

Name	Forward Primer (5' to 3')	Reverse Primer (5' to 3')
Anabaena variabilis	CACCATGGACATTAACACC GAAACC	TCAACTGTTGTTTTCC AGACAATCTTG
Chlorogloeopsis fritschii	CACCATGGACATTAACACC GAGAC	TCACAGGAAAACGCT ATCGCTC
Chryseobacterium indologenes	CACCATGGATATCAATACG GAGACG	TCACGGGTTTTTCTGG CTATTAAAGC
Moorea producens	CACCATGGATATTAACACG GAGACG	TCACACGCGGCTGCTC AGACGTTC
Scytonema hofmanni	CACCATGGATATCAACACG GAGAC	TCAAAAGCCTTCGAG ACAGCTTTTC

A 25 µl PCR reaction mix was prepared as described above in section 2.8.1. PCR was carried out with corresponding annealing temperature as follows.

Table 2.8: PCR protocol of insert preparation for Directional TOPO cloning

Step	Temperature	Time	
Initial Denaturation	98ºC	30 seconds	25 cycles
Denaturation	98ºC	10 seconds	
Annealing	Ava_3855 (66ºC)	30 seconds	
	UYC_RS0133560 (68ºC)		
	BUE84_RS13530 (68ºC)		
	BJP34_RS17220 (68ºC)		
	TOL9009_RS40645 (68ºC)		
Extension	72ºC	30 seconds/kb	
Final extension	72ºC	2 minutes	
Hold	4ºC		

The PCR products were cleaned-up using the monarch kit as explained above in section 2.8.4 and the latter were directionally cloned after the overhang of the cloning vector (GTGG) occupies the 5' end of the PCR product followed by annealing and stabilization of the PCR product in the correct orientation.

2.10.2 Directional TOPO Cloning and transformation in One shot TOP10 competent cells

The cloning strategy was done according to the kit where a total of 6 µl of reaction mix was prepared. The TOPO cloning reaction set up is shown below.

Table 2.9: Directional TOPO cloning protocol

Reagents	Volume
PCR product (> 25 ng/µl)	0.5 µl
Salt solution	1 µl
Sterile water	3.5 µl
TOPO vector	1 µl
Total volume	6 µl

The above reaction was gently mix and incubated for 5 minutes at room temperature. The reaction mix was afterwards placed on ice followed by transformation in One shot TOP10 competent cells.

2.10.3 PCR check using T7 primers

PCR was used to check the cloning result using two designated primers (T7 FWD and T7 REV).

Table 2.10: Primers' design used for confirmation of successful cloning

Primer	Sequence (5' to 3')
T7	TAATACGACTCACTATAGGGG
T7 Reverse	TAGTTATTGCTCAGCGGTGG

A 25 µl PCR reaction mix was prepared as described above in section 2.8.1. PCR was carried out with corresponding annealing temperature as follows.

Table 2.11: PCR cycle conditions when using the T7 primers

Step	Temperature	Time	
Initial Denaturation	98ºC	30 seconds	25 cycles
Denaturation	98ºC	10 seconds	
Annealing	60ºC	30 seconds	
Extension	72ºC	30 seconds/kb	
Final extension	72ºC	2 minutes	
Hold	4ºC		

2.10.4 Restriction digest using PstI-HF

A restriction digest using PstI-HF® was carried out to check for the correct recombinant plasmids.

Table 2.12: Restriction digest protocol using pstI

Component	Volume
Plasmid DNA	1 µl
Buffer (CutSmart)	1 µl
Restriction enzyme (PstI-HF)	0.5 µl
Water	7.5 µl

The above reaction mix was incubated at 37ºC for 1 hour and was then run on a 1% agarose gel.

2.10.5 Primers design and Sanger sequencing

The cloned plasmids were prepared and sent for Sanger sequencing. Primers were designed inside the inserts in order to capture every bit of DNA during the sequencing process and to eradicate every single possibility of missing any point mutations. 5 µl of plasmid DNA and 5 µl of 10 µM primer were sent for sequencing in 1.5 ml Eppendorf tubes.

Primers used for Sanger sequencing:

(1) Recombinant plasmid containing DNA sequences from *Anabaena variabilis*

Table 2.13: Sequencing primers' design for confirmation of correct pAna

Sequencing primers	Sequence (5' to 3')
T7	TAATACGACTCACTATAGGGG
Ana 1322 REV	CACAACCGCTTCACGAACTTG
Ana 1342 FWD	CAAGTTCGTGAAGCGGTTGTG
T7 Reverse	TAGTTATTGCTCAGCGGTGG

(2) Recombinant plasmid containing DNA sequences from *Chlorogloeopsis fritschii*

Table 2.14: Sequencing primers' design for confirmation of correct pChloro

Sequencing primers	Sequence (5' to 3')
T7	TAATACGACTCACTATAGGGG
Chloro 1007 REV	CAGTGGGAACGGATAGGCTTG
Chloro 1027 FWD	CAAGCCTATCCGTTCCCACTG
Chloro 2064 REV	CACGTGGTTTCGCGATCACCAGAC
Chloro 2087 FWD	GTCTGGTGATCGCGAAACCACGTG
Chloro 3101 REV	CAGCGGGAGTTCTTTGCCGAATTG
Chloro 3124 FWD	CAATTCGGCAAAGAACTCCCGCTG
T7 Reverse	TAGTTATTGCTCAGCGGTGG

(3) Recombinant plasmid containing DNA sequences from *Chryseobacterium indologenes*

Table 2.15: Sequencing primers' design for confirmation of correct pChryseo

Sequencing primers	Sequence (5' to 3')
T7	TAATACGACTCACTATAGGGG
Chry 1000 REV	CTGTCTTTAACCAGAATGTGGCTG
Chry 1023 FWD	CAGCCACATTCTGGTTAAAGACAG

Chry 2011 REV	CTTATAAACTGGTTCGATGGCGCTCTTC
Chry 2038 FWD	GAAGAGCGCCATCGAACCAGTTTATAAG
Chry 3025 REV	CACGGGCCTCGAACAGAATC
Chry 3044 FWD	GATTCTGTTCGAGGCCCGTG
T7 Reverse	TAGTTATTGCTCAGCGGTGG

(4) Recombinant plasmid containing DNA sequences from *Moorea producens*

Table 2.16: Sequencing primers' design for confirmation of correct pMoorea

Sequencing primers	Sequence (5' to 3')
T7	TAATACGACTCACTATAGGGG
Moo 1004 REV	GAGCGTGTAGCTCTTGAGTTCTTG
Moo 1027 FWD	CAAGAACTCAAGAGCTACACGCTC
Moo 2023 REV	CACGGCGCACTTCTTGGCTAATC
Moo 2045 FWD	GATTAGCCAAGAAGTGCGCCGTG
Moo 3046 REV	GCCCAGCTCAAAGAAGTTGTCC
Moo 3067 FWD	GGACAACTTCTTTGAGCTGGGC
T7 Reverse	TAGTTATTGCTCAGCGGTGG

(5) Recombinant plasmid containing DNA sequences from *Scytonema hofmanii*

Table 2.17: Sequencing primers' design for confirmation of correct pScy

Sequencing primers	Sequence (5' to 3')
T7	TAATACGACTCACTATAGGGG
Scy 894 REV	CGAAACTAACCGCGATGGCATTAC
Scy 917 FWD	GTAATGCCATCGCGGTTAGTTTCG
Scy 1896 REV	GTTCTGGGCGGTTCAGATAACCGC
Scy 1919 FWD	GCGGTTATCTGAACCGCCCAGAAC
Scy 2878 REV	CAGTTGCTGGGCGATTTCGTAC
Scy 2899 FWD	GTACGAAATCGCCCAGCAACTG
T7 Reverse	TAGTTATTGCTCAGCGGTGG

The successfully cloned plasmids were transformed in both *E. coli* BL21 Star™ (DE3) and *E. coli* BAP1 as explained in section 2.5.6 and 2.5.5 respectively and stored at -80ºC as glycerol stocks.

2.11 Q5® Site-Directed Mutagenesis

The Q5® Site-Directed Mutagenesis Kit was used to make a site-specific mutagenesis in one of the synthetic DNA sequences. The kit was chosen for its quick and simple procedure where the custom non-overlapping mutagenic primers were designed using the NEBaseChanger™ where the corresponding annealing temperature was also provided which thus allowed effective exponential amplification resulting in a higher yield of the desired mutated plasmid. A PCR master mix containing Q5 Hot Start High-Fidelity DNA Polymerase and a unique multi-enzyme Kinase-Ligase-DpnI (KLD) mix is provided in the kit which reduces error rate and also enables the reaction to happen at room temperature.

The purpose of using site-directed mutagenesis was for a base insertion in the *Moorea producens* synthetic DNA sequence in pUC57.

2.11.1 Primer design and Exponential amplification

Table 2.18: Primers' design for site-directed mutagenesis

Name	Forward Primer (5' to 3')	Reverse Primer (5' to 3')
DNA sequence from *Moorea producens* in pUC57	TACAGTGGTGGCATCCTC GTGCTG	CCACGTACTGAAGATTT CTTGGAAGC

Exponential Amplification:

Table 2.19: General PCR protocol for exponential amplification for site-directed mutagenesis

Components	Volume	Final concentration
Q5® Hot Start High-Fidelity 2X Master Mix	12.5 µl	1X
10 µM Forward primer	1.25 µl	0.5 µM
10 µM Reverse primer	1.25 µl	0.5 µM
Template DNA	1 µl	1-25 ng

Nuclease-free water	9.0 µl	

Cycle conditions:

Table 2.20: Cycle conditions for PCR for site-directed mutagenesis

Step	Temperature	Time	
Initial Denaturation	98ºC	30 seconds	25 cycles
Denaturation	98ºC	10 seconds	
Annealing	61ºC	30 seconds	
Extension	72ºC	30 seconds/kb	
Final extension	72ºC	2 minutes	
Hold	4ºC		

After PCR, the reaction mix was added to the KLD enzyme mix for 5 minutes at room temperature for circularization and template eradication by the DpnI enzyme.

2.11.2 Kinase-Ligase-DpnI Reaction and Transformation in chemically competent cells

Table 2.21: Site-directed mutagenesis protocol

Components	Volume	Final Concentration
PCR product	1 µl	
2X KLD reaction buffer	5 µl	1X
10X KLD enzyme mix	1 µl	1X
Nuclease-free water	3 µl	

After 5 minutes incubation at room temperature, the reaction mix was transformed in chemically competent *E. coli* cells. A total volume of 5 µl of the KLD mix was added to 50 µl of chemically competent cells and the transformation was carried out as described in section 2.5.6 above.

2.11.3 Primers design and Sanger sequencing

The desired mutated plasmid (DNA sequence from *Moorea producens* in pUC57) was prepared and sent for sanger sequencing. Primers were designed inside the insert in order to capture every bit of DNA, especially the mutated area, during the sequencing process. 5 µl of plasmid DNA and 5 µl of 10 µM primer were sent for sequencing in 1.5 ml Eppendorf tubes.

(1) pUC57 containing DNA sequences from *Moorea producens*

Table 2.22: Sequencing primers' design for confirmation of successful site-directed mutagenesis

Sequencing primers	Sequence (5' to 3')
Moo FWD 1718 pUC57	GAATACCAAGGCCAGATGGTGTGTC
Moo REV 2519 pUC57	GGCTCTCACCTTGTTCGCCAAAC

After obtaining the desired mutated plasmid (DNA sequence from *Moorea producens* in pUC57), the DNA sequence from *Moorea producens* was captured and cloned in pET151 by directional TOPO cloning as described in the above section 2.11. The successfully cloned plasmids were transformed in both *E. coli* BL21 Star™ (DE3) and *E. coli* BAP1 as explained in section 2.5.6 and 2.5.5 respectively and stored at -80ºC as glycerol stocks.

2.12 NEBuilder® HiFi DNA Assembly Cloning [Gibson assembly] for A-domain swap

The NEBuilder® HiFi DNA Assembly Master Mix was used to swap the A-domain from *Anabaena variabilis* to the A-domain of *Moorea producens*. This technique has the ability to assemble different sizes of DNA fragments with varied overlaps of usually 30 base pairs with ease and simplicity.

2.12.1 Primer design and Exponential amplification

Table 2.23: Primers' design for Gibson assembly

Name	Forward Primer (5' to 3')	Reverse Primer (5' to 3')
***Anabaena variabilis* in pET151**	CAAGTTAAGATTCGCGGCTACCGC	GCGTTCAACTTGGGCAACGATGAGTTC
***Moorea producens* in pET151**	CATGAACTCATCGTTGCCCAAGTTGAACGCACCCCGGATGCGGTGGCG	CTCAATGCGGTAGCCGCGAATCTTAACTTGATTGTCGATGCGACCCAG

Exponential Amplification:

Table 2.24: General PCR protocol for exponential amplification for Gibson assembly

Components	Volume	Final concentration
Q5® High-Fidelity 2X Master Mix	12.5 µl	1X
10 µM Forward primer	1.25 µl	0.5 µM
10 µM Reverse primer	1.25 µl	0.5 µM
Template DNA	1 µl	1-25 ng
Nuclease-free water	9.0 µl	

Cycle conditions:

Table 2.25: Cycle conditions for PCR for Gibson assembly

Step	Temperature	Time	
Initial Denaturation	98ºC	30 seconds	25 cycles
Denaturation	98ºC	10 seconds	
Annealing	72ºC	30 seconds	
Extension	72ºC	30 seconds/kb	
Final extension	72ºC	2 minutes	
Hold	4ºC		

2.12.2 PCR product gel extraction; DpnI digestion of PCR product and PCR Cleanup

The respective PCR products were extracted from the gel using the Monarch® DNA gel extraction kit from NEB as described above. DpnI is known to cleave methylated DNA and is a good way to get rid of the template DNA after PCR.

Table 2.26: Restriction digest protocol using DpnI

Component	Volume
Plasmid DNA	15 µl
Buffer (CutSmart)	5 µl
Restriction enzyme (DpnI-HF)	1 µl
Water	29 µl

The above reaction mix was incubated at 37°C for 1 hour and was then heat inactivated at 80°C for 20 minutes.

The PCR product was cleaned-up using the monarch kit as explained above in section 2.8.4. The cleanup PCR product was quantified using the NanoDrop™ 2000 spectrophotometer and calculations were done for the Gibson assembly cloning process.

2.12.3 Gibson assembly and Transformation

The Gibson assembly master mix contains an exonuclease enzyme chews back 5' ends to form single stranded 3' overhangs easing fragments annealing at the overlap region while the polymerase enzyme fills in gaps within each annealed fragment followed by the ligation of the assembled DNA by DNA ligase. The ratio of vector to insert used was 1:5.

The reaction mix was incubated at 50°C for 1 hour in a thermocycler followed by incubation on ice. The reaction mix was then transformed in 50 µl NEB 5-alpha Competent *E. coli* as described in section 2.5.6 above.

2.12.4 PCR check using designed primers

The next day, 20 colonies were picked, and overnight cultures of 5 ml were set up followed by Miniprep the next day. PCR was carried out using two designated primers to check for the correct construct as shown below. The 'T7' forward primer was chosen as it was found on the plasmid backbone and the 'Moo 2023 REV' reverse primer was found in the insert which would give a good confirmation of whether the cloning worked or not.

Table 2.27: Primers' design for confirmation of successful Gibson assembly by PCR check

Primer	Sequence (5' to 3')
T7	TAATACGACTCACTATAGGGG
Moo 2023 REV	CACGGCGCACTTCTTGGCTAATC

2.12.5 Restriction digest using HpaI

A restriction digest was also set up using the restriction endonuclease HpaI as shown below to check the engineered plasmid.

Table 2.28: Restriction digest protocol using HpaI

Component	Volume
Plasmid DNA	1 µl
Buffer (CutSmart)	1 µl
Restriction enzyme (HpaI)	0.5 µl
Water	7.5 µl

2.12.6 Primers design and Sanger sequencing

The engineered plasmid was prepared and sent for sanger sequencing. Primers were designed inside the inserts in order to capture every bit of DNA during the sequencing process and to eradicate every single possibility of missing any mutations. 5 µl of plasmid DNA and 5 µl of 10 µM primer were sent for sequencing in 1.5 ml Eppendorf tubes.

Primers used for sanger sequencing:

(1) Engineered plasmid involving the A-domain swap

Table 2.29: Sequencing primers' design for confirmation of successful A-domain swap using Gibson assembly

Sequencing primers	Sequence (5' to 3')
T7	TAATACGACTCACTATAGGGG
GA Ana pREV	GCGTTCAACTTGGGCAACGATGAGTTC
GA Moo FWD	CATGAACTCATCGTTGCCCAAGTTGAACGCACCCCGGATG CGGT GGCG
Moo 2023 REV	CACGGCGCACTTCTTGGCTAATC
Moo 2045 FWD	GATTAGCCAAGAAGTGCGCCGTG
Ana 1322 REV	CACAACCGCTTCACGAACTTG
Ana 1342 FWD	CAAGTTCGTGAAGCGGTTGTG
T7 Reverse	TAGTTATTGCTCAGCGGTGG

The successfully engineered plasmid was transformed in both *E. coli* BL21 Star™ (DE3) and *E. coli* BAP1 as explained in section 2.5.6 and stored at -80^{o}C as glycerol stocks.

2.13 NEBuilder® HiFi DNA Assembly Cloning [Gibson assembly] of engineered enzyme (NEE) in pET28MBP

The engineered enzyme (previously in pET151) was cloned in pET28MBP.

2.13.1 Primer design and Exponential amplification

Table 2.30: Primers' design for Gibson assembly

Name	Forward Primer (5' to 3')	Reverse Primer (5' to 3')
pET28MBP	CACCACCACCACCACCA CTGAGATCCGGC	ACTGAGGGATCCTTGGA AGTATAGATTTTC
NEE in pET151	GAAAATCTATACTTCCA AGGATCCCTCAGTATGG ACATTAACACCG	GCCGGATCTCAGTGGTG GTGGTGGTGGTGACTGT TGTTTTCCAGAC

Exponential Amplification:

Table 2.31: General PCR protocol for exponential amplification for Gibson assembly

Components	Volume	Final concentration
Q5® High-Fidelity 2X Master Mix	12.5 µl	1X
10 µM Forward primer	1.25 µl	0.5 µM
10 µM Reverse primer	1.25 µl	0.5 µM
Template DNA	1 µl	1-25 ng
Nuclease-free water	9.0 µl	

Cycle conditions:

Table 2.32: Cycle conditions for PCR for Gibson assembly

Step	Temperature	Time	
Initial Denaturation	98ºC	30 seconds	25 cycles
Denaturation	98ºC	10 seconds	
Annealing	72ºC	30 seconds	
Extension	72ºC	30 seconds/kb	
Final extension	72ºC	2 minutes	
Hold	4ºC		

2.13.2 PCR product gel extraction; DpnI digestion of PCR product and PCR Cleanup

The respective PCR products were extracted from the gel using the Monarch® DNA gel extraction kit from NEB as described above. DpnI is known to cleave methylated DNA and is a good way to get rid of the template DNA after PCR.

Table 2.33: Restriction digest protocol using DpnI

Component	Volume
Plasmid DNA	15 µl
Buffer (CutSmart)	5 µl
Restriction enzyme (DpnI-HF)	1 µl
Water	29 µl

The above reaction mix was incubated at 37°C for 1 hour and was then heat inactivated at 80°C for 20 minutes.

The PCR product was cleaned-up using the monarch kit as explained above in section 2.8.4. The cleanup PCR product was quantified using the NanoDrop™ 2000 spectrophotometer and calculations were done for the Gibson assembly cloning process.

2.13.3 Gibson assembly and Transformation

The Gibson assembly was carried out with the ratio of vector to insert being 1:5.

The reaction mix was incubated at 50°C for 1 hour in a thermocycler followed by incubation on ice. The reaction mix was then transformed in 50 µl NEB 5-alpha Competent *E. coli* as described in section 2.5.6 above.

2.13.4 PCR check using designed primers

The next day, 10 colonies were picked, and overnight cultures of 5 ml were set up followed by Miniprep the next day. PCR was carried out using two designated primers to check for the correct construct as shown below. The ‘T7’ forward primer and the ‘T7 REV’ primers were chosen as both are found on the plasmid backbone which would give a good confirmation of whether the cloning worked or not.

Table 2.34: Primers' design for confirmation of successful Gibson assembly by PCR check

Primer	Sequence (5' to 3')
T7	TAATACGACTCACTATAGGGG
T7 Reverse	TAGTTATTGCTCAGCGGTGG

2.13.5 Primers design and Sanger sequencing

The plasmid was prepared and sent for sanger sequencing. Primers were designed inside the inserts in order to capture every bit of DNA during the sequencing process and to eradicate every single possibility of missing any mutations. 5 µl of plasmid DNA and 5 µl of 10 µM primer were sent for sequencing in 1.5 ml Eppendorf tubes.

Primers used for Sanger sequencing:

Table 2.35: Sequencing primers' design for confirmation of successful cloning of NEE in pET28MBP using Gibson assembly

Sequencing primers	Sequence (5' to 3')
T7	TAATACGACTCACTATAGGGG
Pmbp REV	ACTGAGGGATCCTTGGAAGTATAGATTTTC
Imbp FWD	GAAAATCTATACTTCCAAGGATCCCTCAGTATGGACATTAACACCG
Moo 2023 REV	CACGGCGCACTTCTTGGCTAATC
Moo 2045 FWD	GATTAGCCAAGAAGTGCGCCGTG
Ana 1322 REV	CACAACCGCTTCACGAACTTG
Ana 1342 FWD	CAAGTTCGTGAAGCGGTTGTG

T7 Reverse	TAGTTATTGCTCAGCGGTGG

The successfully engineered plasmid was transformed in both *E. coli* BL21 Star™ (DE3) as explained in section 2.5.6 and stored at -80ºC as glycerol stocks.

2.14 Production of mycosporine-glycine (Precursor)

pET29b-Ava3858-3856 was a gift from Emily Balskus & Christopher T Walsh (Addgene plasmid # 45745 ; http://n2t.net/addgene:45745 ; RRID:Addgene_45745). [The genetic and molecular basis for sunscreen biosynthesis in cyanobacteria. Balskus EP, Walsh CT. *Science. 2010 Sep 24;329(5999):1653-6. doi: 10.1126/science.1193637. Epub 2010 Sep 2.* 10.1126/science.1193637 PubMed 20813918].

The plasmid pET29b Ava3858-3856 consists of a DHQS homolog, a O-methyltransferase and an ATP-grasp enzyme cloned in the pET29b vector which led to the production of mycosporine-glycine.

2.14.1 Isolation of pET29b Ava3858-3856

The plasmid was obtained as agar stab in *E. coli* Top10 which was streaked on LB plate with the appropriate antibiotic, here referring to 50 µg/ml kanamycin and left to grow overnight at 37ºC in the incubator. A colony was picked up the next day and inoculated in 5 ml LB with 5 µl kanamycin and left to grow overnight in a shaking incubator at 37ºC. The following day, the plasmid was isolated by miniprep as described in section 2.6.1 and quantified on the NanoDrop™ 2000 Spectrophotometer.

2.14.2 Transformation of pET29b Ava3858-3856 in *E. coli*

A volume of 3 µl of the plasmid was transformed in both *E. coli* BL21 Star™ (DE3) and *E. coli* BAP1 as explained in section 2.5.6 and stored at -80ºC as glycerol stocks.

2.14.3 Heterologous expression of pET29b Ava3858-3856 in *E. coli*

Overnight LB culture containing the corresponding amount of kanamycin was prepared and used to inoculate a total of 7 L of fresh LB cultures supplemented with 50 µg/ml kanamycin. The cultures were incubated at 37ºC on a shaking incubator until

the OD_{600} reached about 0.5, cooled to 15ºC and induced with 0.5 mM IPTG followed by overnight incubation at 15ºC.

The following day, the cells were pelleted by centrifugation at 6000 rpm for 20 minutes and the latter was resuspended in methanol, lysed on ice by sonicating with the following conditions: 8 X 30 seconds pulses with 1 minute recovery in between each pulse. The lysate was spined down at 4000 rpm for 45 minutes and the supernatant was dried using the rotary evaporator.

2.14.4 Purification and Characterisation of mycosporine-glycine

The crude extract was resuspended in HPLC-grade water, filtered using 0.22-micron nylon filter and prepared for High Performance Liquid Chromatography (HPLC) and Ultra High-Performance Liquid Chromatography - High Resolution Mass Spectrometry (UHPLC-HRMS).

2.14.5 High Performance Liquid Chromatography (HPLC)

The purification of mycosporine-glycine was performed on the Agilent 1260 Infinity Series system which consisted of a binary pump; a Diode Array Detector (DAD); a standard autosampler; a sample degasser; a column compartment and a fraction collector. A preparative reverse-phase C18 column of 21.2 X 100 mm with particle size of 5 micron from Agilent was used for the purification of mycosporine-glycine. The following conditions were used: 5% solvent B for 15 minutes; increasing to 15% solvent B for 15 minutes; increasing to 100% solvent B for 10 minutes (solvent A = water + 0.1% formic acid; solvent B = methanol + 0.1% formic acid) with a flow rate of 5 ml/min. Unfortunately, the yield of mycosporine-glycine was too low to be used in this research.

2.15 Extraction of MAAs from Rhodophytes

2.15.1 Purification and characterisation of MAAs from Helioguard^{TM} 365

All reagents and solvents used were of analytical grade unless otherwise stated.

The Helioguard^{TM} 365 samples (Mibelle Biochemistry) were received as an aqueous preparation containing 0.1 % mycosporine-like amino acids (MAAs) according to the manufacturer's report. The ethanol and most of the water was removed under vacuum on a rotary evaporator and the residue was resuspended in HPLC-grade water and filtered through 0.25 μm filters in preparation for reverse-phase High Performance Liquid Chromatography (RP-HPLC). The purification was performed using a PrepHT XDB-C18 column (21.2 x 100 mm, particle size 5 μm) connected to an Agilent 1260 HPLC equipped with a binary pump and diode-array detection detector. Injection volumes of 200 μL were used with the following conditions: 5 % B for 15 min; 10 % B for 10 min; then increasing to 100 % B over 5 min; and 100 % B for 10 min (solvent A = 0.1 % formic acid in water; solvent B = 0.1 % formic acid in methanol) with a flow rate of 5 mL/min. Fractions were collected based on absorbance at 333 nm using an automated fraction collector. The retention time of shinorine was 4.35 min while the retention time of porphyra-334 was 6.15 min. The fractions containing each compound were combined respectively and the solvent was evaporated using a centrifugal evaporator.

Each compound was confirmed by Ultra High- Performance Liquid Chromatography - High Resolution Mass Spectrometry (UHPLC-HRMS) analysis. A sample of porphyra-334 was dissolved in D_2O, and a ^1H NMR was recorded on a Bruker Avance III 400 MHz spectrometer.

2.15.2 Purification and characterisation of MAAs from *Palmaria palmata*

Palmaria palmata (Dulse) was bought online from different companies from Amazon. The samples were cut into small pieces, except for the flakes ones and sequentially extracted in 25% HPLC-grade methanol using 2 different methods including 45ºC for 2 hours and 4ºC overnight followed by sonication the next day respectively. After extraction, the solvent was pooled together, centrifuged at 4000 rpm for 30 minutes to get rid of debris and the clear solvent evaporated using the centrifugal evaporator (SP Genevac EZ-2 Series). The crude extract was then resuspended in HPLC-grade water, filtered using a syringe-driven filter membrane, size 0,22 μm for further analysis.

Reverse-phase HPLC was performed using a PrepHT XDB-C18 column (21.2 x 100 mm, particle size 5-micron) connected to an Agilent 1260 HPLC equipped with a

binary pump and DAD detector for the purification of MAAs. The first purification step was the injection of 500 µL filtered sample under the following conditions: 95 : 5 solvent A/ solvent B for 15 minutes, increasing to 90 : 10 solvent A/ solvent B for 10 minutes, increasing to 100% solvent B over 5 minutes and finally 100% solvent B for 10 minutes (solvent A = 0.1% formic acid in water; solvent B = 0.1% formic acid in methanol) with a flow rate of 5 ml/min. Fractions were collected every minute based on absorbance at 310 nm, 333 nm and 360 nm using an automated fraction collector. Fractions were combined into batches 1 to 4 where batches 1 and 2 were further purified using 95 : 5 solvent A/ solvent B for 15 minutes with varying collection period except for batches 3 and 4 which were purified under the following conditions: 95 : 5 solvent A/ solvent B for 10 minutes, increasing to 90 : 10 solvent A/ solvent B for 10 minutes. After each purification step, the fractions of each compound were pooled respectively, and the solvent was evaporated using the centrifugal evaporator. No formic acid was used in both solvent A and solvent B during purification of usujirene/palythene.

Each purified compound was resuspended in HPLC-grade water and filtered using the centrifugal filter membrane, size 0,22 µm for UHPLC-HRMS analysis. A sample of usujirene/palythene mix was dissolved in D_2O, and a ^1H NMR was recorded on a Bruker Avance III 400 MHz spectrometer.

Ultra-High-performance Liquid Chromatography- High Resolution Mass Spectrometry (UHPLC-HRMS)

Ultra-High-performance Liquid Chromatography- High Resolution Mass Spectrometry was performed with a reverse phase Zorbax Eclipse Plus C18 column, size 2.1 x 100 mm, particle size 1.8 µm) connected to a Dionex 3000RS UHPLC coupled to Bruker Ultra High Resolution (UHR) Q-TOP MS MaXis II mass spectrometer using ESI (Electrospray Ionization in positive mode) with 10 mM sodium formate through loop injection. A *m/z* range of 50-2500 was used. Column elution was done with the following gradient: 0% B to 25% B for 30 mins (solvent A = water + 0.1% formic acid and solvent B = acetonitrile + 0.1% formic acid).

2.16 Over-expression of NRPS-like enzyme by IPTG induction [50 ml culture]

Each expression construct from *E. coli BL21* and *E. coli* BAP1 was streaked on respective antibiotics LB plates. One colony was picked up to prepare overnight cultures of each expression constructs in LB supplemented with corresponding antibiotics. Overnight cultures were used to inoculate a total of 50 ml of fresh LB cultures supplemented with appropriate antibiotics. The cultures were incubated at 37ºC on a shaking incubator until the OD_{600} reached about 0.5, cooled to 15ºC and induced with corresponding concentration of IPTG as shown below followed by overnight incubation at 15ºC.

Table 2.36: IPTG concentration used for each construct during induction

Recombinant plasmid in *E. coli* BAP1	**IPTG concentration**
pAna	0.5 mM
pChloro	0.2 mM
pChryseo	0.2 mM
pMoorea	0.2 mM
pScy	0.2 mM
pNEE	0.2 mM
pET28MBPNEE	0.2 mM

2.17 Benchtop purification of NRPS-like enzyme from 50 ml culture

The following day, the cells were pelleted by centrifugation at 4000 rpm for 20 minutes and were resuspended in 3 ml of the corresponding buffer as shown below.

Table 2.37: Lysis buffer and washing buffer used for each construct

Construct	**Lysis Buffer**	**Washing/ Elution Buffer**

pAna	A	C
pChloro	B	D
pChryseo	B	D
pMoorea	B	D
pScy	B	D
pNEE	B	D
pET28MBPNEE	B	D

The cells were lysed by sonication on ice with the following conditions: 5 X 30 seconds pulses with 1 minute recovery in between each pulse. The lysate was clarified by centrifugation at 4000 rpm for 45 minutes at 4^{o}C. Meanwhile, 200 µl of Ni Sepharose was added to 1.5 ml washing buffer in an Eppendorf tube, centrifuged for 3 minutes at 13000 rpm at 4^{o}C and the supernatant was discarded. Half of the lysed soluble fraction (1.5 ml) was added to the Ni Sepharose and incubated for 5 minutes on ice, occasionally inverting the tubes. The latter were spined down and the supernatant discarded. The rest of the lysed soluble fraction (1.5 ml) was added to the Ni Sepharose and incubated for 5 minutes on ice, occasionally inverting the tubes. The latter were spined down and the supernatant discarded. 1.5 ml of the corresponding washing buffer was added, incubated for 5 minutes on ice, occasionally inverting the tubes and centrifuged for 3 minutes. 0.5 ml of the corresponding elution buffer [200 mM imidazole in the corresponding buffer as described above] was used to elute the protein of interest. Analysis of the protein was done by 12% SDS-PAGE protein gel.

2.17.1 SDS-PAGE (12%) protein gel protocol

Loading: Total volume = 12 ml

Component	**Volume**
Distilled water	3.9 ml
30% Acrylamide	4.8 ml
1.5M Tris (pH= 8.8)	3 ml
10% SDS	120 µl
10% APS	120 µl
TEMED	12 µl

Stacking 6%: Total volume = 5 ml

Component	**Volume**
Distilled water	2.6 ml
30% Acrylamide	1 ml
0.5M Tris (pH= 8.8)	1.25 ml
10% SDS	50 µl
10% APS	50 µl
TEMED	5 µl

The loading was poured into plates and left to set for about 20 minutes followed by the making of the stacking which was afterwards top up. The comb was gently added and allowed to set for another 20 minutes.

The SDS running buffer was prepared as follows.

For 1X:

Component	**Mass**
Tris	3.03 g
Glycine	14.4 g
SDS	1 g
Distilled water	Up to 1000 ml

The samples were prepared by adding SDS loading dye and heated at 95ºC for 10 minutes.

The gel was run for about 10 minutes at 160V without the samples in order to equilibrate the latter. The samples were then loaded (about 10 µl) and 4 µl color prestained protein standard, broad range (10-250 kDa) NEB ladder was used, and the gel was run for 1 hour at 150V.

After the run, the gel was transferred into a clean plastic container, coated with Instant Blue dye and left on the rocker for a minimum of 15 minutes to overnight. The Instant Blue was afterwards discarded, and water was added to the gel and viewed.

2.18 Large-scale over-expression and purification of NRPS-like enzyme using HisPur™ Ni-NTA Resin

Each expression construct from *E. coli BL21* and *E. coli* BAP1 was streaked respectively on appropriate antibiotics LB plates. One colony was picked up to prepare overnight cultures of each expression constructs in LB supplemented with corresponding antibiotics. Overnight cultures were used to inoculate a total of 2 L of fresh LB cultures supplemented with corresponding antibiotics. The cultures were incubated at 37ºC on a shaking incubator until the OD_{600} reached about 0.5, cooled to 15ºC and induced with corresponding concentration of IPTG as shown below followed by overnight incubation at 15ºC.

The following day, the cells were pelleted by centrifugation at 6000 rpm for 20 minutes and were resuspended in 30 ml of the corresponding buffer (based on the 50 ml culture). Lysozyme as well as one tablet of the protease inhibitor cocktail were added to the buffers before the cells were lysed by sonication on ice with the following conditions: 8 X 30 seconds pulses with 1 minute recovery in between each pulse. The lysate was clarified by centrifugation at 4000 rpm for 45 minutes at 4ºC. The supernatant was then incubated with 3 ml of Ni-NTA resin for 2 hours at 4ºC.

The mixture was spined down at 3000 rpm for 5 minutes and the unbound fraction was discarded. The Ni-NTA was resuspended in the corresponding wash buffer for each and loaded into a column and washed with 10 ml of the corresponding wash buffer.

The NRPS-like enzymes were eluted from the Ni-NTA using a stepwise imidazole gradient in the corresponding buffer (25 mM; 50 mM; 75 mM, 100 mM, 125 mM 150 mM and 200 mM) and 3 ml fractions were collected for each.

Analysis of the protein was done by SDS-PAGE using precast 4-15% Tris-HCl protein gel. The fractions containing the targeted protein were combined and concentrated to

about 8 ml using an Amicon Ultra 50K molecular weight cutoff spin filter. The NRPS-like enzyme from *Anabaena variabilis* was dialyzed thrice against 2 L of storage buffer E while the engineered NRPS-like enzyme was dialyzed thrice against 2 L of storage buffer F.

The protein was quantified using the NanoDrop™ 2000 Spectrophotometer as described in section 2.7.4. The protein solutions were then aliquoted (50 µL), flash frozen in liquid nitrogen and stored at -80^{o}C until further use.

2.19 UHPLC-ESI-Q-TOF-MS analysis of intact Ava_3855 protein

Purified Ava_3855 was converted from "apo" to "holo" form by incubation of 270 µM enzyme with 5 mM $MgCl_2$, 0.1 mM CoA and 500 µM sfp enzyme in a total volume of 50 µL for 1 hour at room temperature. The loading of the substrate L-serine was initiated by adding 5 mM ATP and 1 mM L-serine in a final volume of 50 µL where the reaction mix was incubated for 1 hour at room temperature. The synthetic precursor (cyclohexenone backbone) AC23 (2 mM) was then added to the reaction mix and incubated for 1 hour at room temperature.

Intact protein analysis of both the "apo" and "holo" form, followed by the loading and unloading of the substrate L-serine on the T domain were analysed on a Bruker MaXis II ESI-Q-TOF-MS connected to a Dionex 3000 RS UHPLC fitted with an ACE C4-300 RP column (100 x 2.1 mm, 5 µm). A gradient elution starting from 5-100% acetonitrile containing 0.1% formic acid for 30 mins was used. The mass spectrometer was operated in positive ion mode with a scan range of 200-3000 m/z. Source conditions were: end plate offset at -500 V; capillary at -4500 V; nitrogen as nebulizer gas at 1.8 bar and dry gas at 9.0 L min^{-1}; dry temperature at 200^{o}C. Ion transfer conditions were: ion funnel RF at 400 Vpp; multiple RF at 200 Vpp; quadrupole low mass at 200 m/z; collision energy at 8.0 eV; collision RF at 2000 Vpp; transfer time at 110.0 µs; pre-pulse storage time at 10.0 µs.

2.20 Chemo-enzymatic assay

A chemo-enzymatic assay was set up using the synthetic precursor AC23 (The synthetic substrate was synthesized by Adam Cowden, PhD student under the supervision of Martin Wills in the Chemistry department). A final volume of 100 µL solution containing 5 mM $MgCl_2$, 0.1 mM CoA, 500 µM Sfp, 1 mM L- serine, 2 mM

AC23, 2 µM Ava_3855 and 5 mM ATP were used for the assay which was carried out at room temperature overnight.

The reaction mix was quenched with ice-cold acetonitrile, vortexed and dried using the centrifugal evaporator. The residue were resuspended in 50:50 HPLC-grade water : HPLC-grade methanol followed by filtration using the centrifugal filter membrane, size 0,22 µm.

Ultra-High-performance Liquid Chromatography- High Resolution Mass Spectrometry (UHPLC-HRMS)

Ultra-High-performance Liquid Chromatography- High Resolution Mass Spectrometry was performed with a reverse phase Zorbax Eclipse Plus C18 column, size 2.1 x 100 mm, particle size 1.8 µm) connected to a Dionex 3000RS UHPLC coupled to Bruker Ultra High Resolution (UHR) Q-TOP MS MaXis II mass spectrometer using ESI (Electrospray Ionization in positive mode) with 10 mM sodium formate through loop injection. A *m/z* range of 50-2500 was used. Column elution was done with the following gradient: 0% B to 100% B for 30 mins (solvent A = water + 0.1% formic acid and solvent B = acetonitrile + 0.1% formic acid)

Chapter 3: Extraction, purification, and characterisation of naturally occurring Mycosporine-like amino acids from Rhodophytes and their photoprotective mechanisms

3.1 Brief Introduction

Rhodophytes have long been known as producers of MAAs[113]. The MAAs, among other functions, have potential of protecting the DNA of these organisms from the damaging effects of ultraviolet radiation, a scientific term known as, "Photoprotection". It was shown that MAAs such as shinorine and porphyra-334 had the capacity of releasing approximately 97% of the absorbed UV radiation into heat to its surroundings based on a photoacoustic calorimetry test[114]. The need to track the UV to heat energy transfer process in real time is thus a vital detail to consider for the development of these photoprotective molecules in pharmaceutical industries for future purposes.

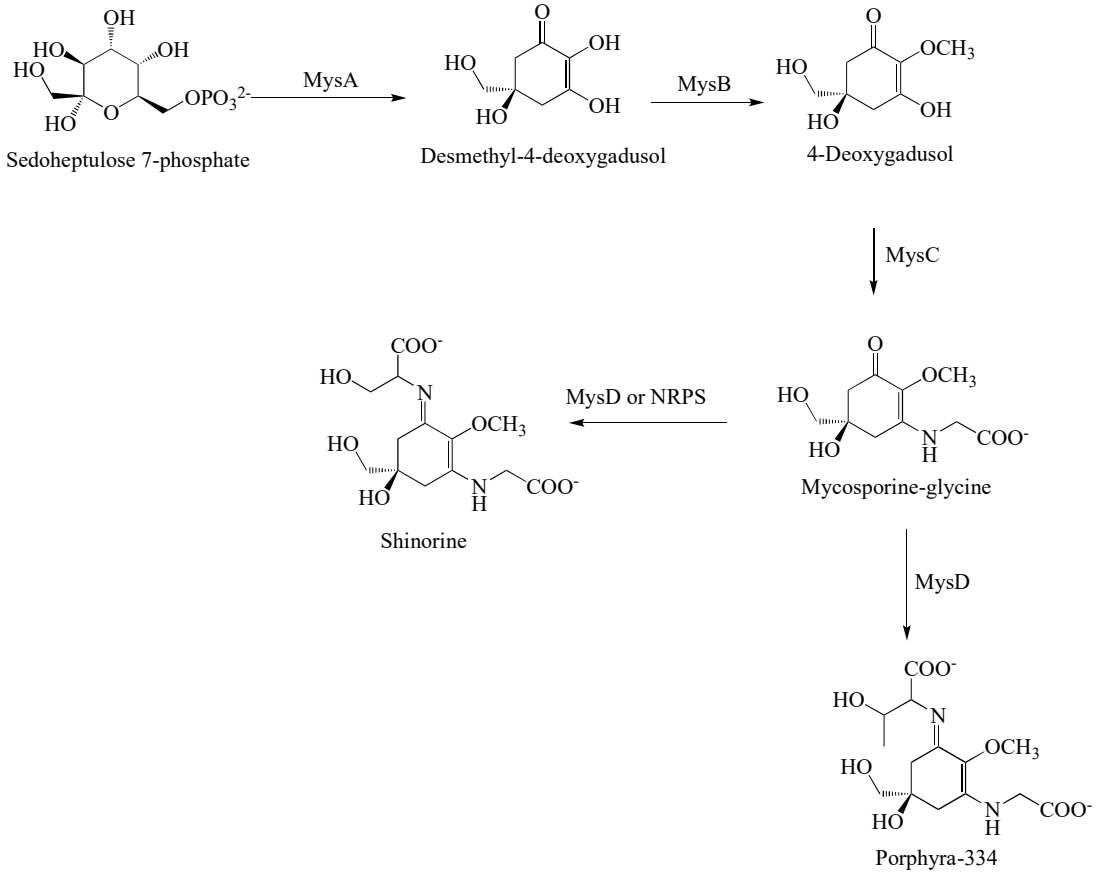
The goal of this experiment was to purify and characterise shinorine and porphyra-334 from Helioguard 365 and to extract, purify and characterise the cis-trans isomers usujirene/ palythene from *Palmaria palmata* where the photoprotective mechanism including the tracking of the energy flow using Transient Electronic Absorption Spectroscopy (TEAS) was carried out for the first time on these specific molecules.

3.2 Purification and characterisation of shinorine and porphyra334 from Helioguard™ 365

Porphyra umbilicalis, commonly known as purple laver, belongs to the Bangiophyceae group. They are about 20 cm in length with tough irregular-shaped frond found in harsh environments like the intertidal zones with temperature variations, severe osmotic stress as well as elevated irradiation levels[113,115]. 1% of the dry weight of *Porphyra* is usually composed of MAAs, the most common one being porphyra-334. The biosynthetic genes encoding the production of MAAs in *Porphyra* are a *mysa-mysb* and *mysc-mysd* fusion proteins which are found next to each other but transcribed on opposite DNA strands with adjacent 5' ends. These two protein-fusion

arrangements suggest the effectiveness along with the boosted advantage of MAAs biosynthesis upon increased level of UV irradiation[113].

A.



B.

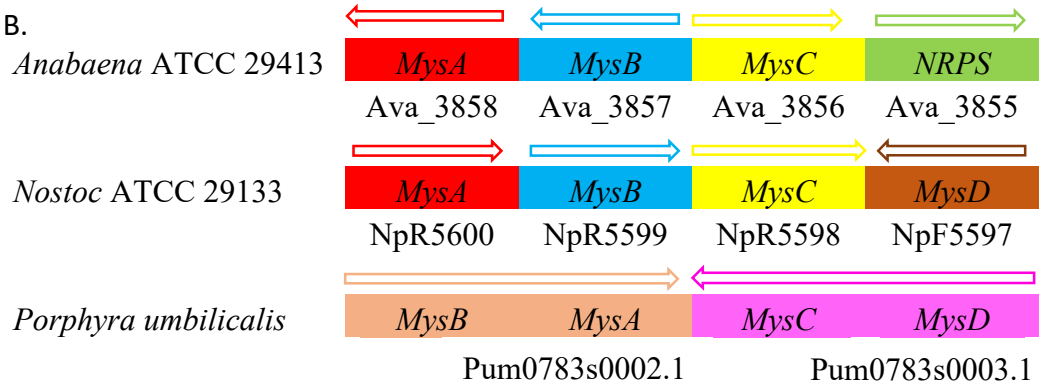


Figure 3.1: A: Biosynthetic pathway of shinorine and porphyra-334 from sedoheptulose 7-phosphate in cyanobacteria and rhodophyte. B: Gene fusion in the red alga *Porphyra umbilicalis* compared to individual gene cluster in the cyanobacteria *Anabaena* ATCC 29413 and *Nostoc* ATCC 29133.

Helioguard^{TM} 365, a commercially available product provided by Mibelle Group Biochemistry consist of photoprotective metabolites namely shinorine and porphyra-334 from the Rhodophyte *Porphyra umbilicalis.* The sample as shown in Figure 2 was first dried and resuspended in HPLC-grade water which was prepared for purification of the UVA- absorbing metabolites by HPLC.

Figure 3.2: The sample Helioguard 365 from Mibelle Group Biochemistry. The product is a liposomal preparation of mycosporine-like amino acids from the rhodophyte *Porphyra umbilicalis* comprising of 80% *Porphyra umbilicalis* extract with 0.1% mycosporine-like amino acids, 3% phospholipids, 15% absolute ethanol, 0.16% sodium lactate and deionized water.

As described in Chapter 2, the Helioguard 365 sample was purified giving rise to shinorine at retention time 4.35 min while the retention time of porphyra-334 was 6.15 min.

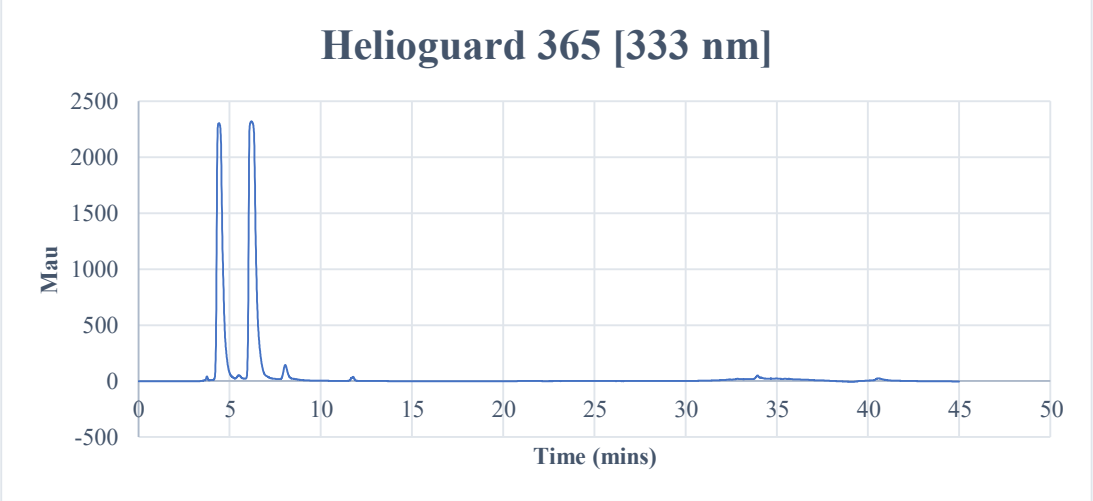


Figure 3.3: The HPLC chromatogram at 333 nm of shinorine and porphyra-334 purified from Helioguard 365. The Helioguard 365 extract was purified resulting in shinorine at 4.35 min and porphyra-334 at 6.15 min.

Shinorine and porphyra-334 were successfully purified based on the wavelength, also taking into consideration their respective retention time. Each individual fraction was then prepared for LC-HRMS for confirmation.
The ultraviolet (UV) chromatogram at 333 nm for each fraction showed a single peak with the retention time correlating to the correct mass spectra.

Shinorine from Helioguard^{TM} 365 (λ_{max} = 333 nm)

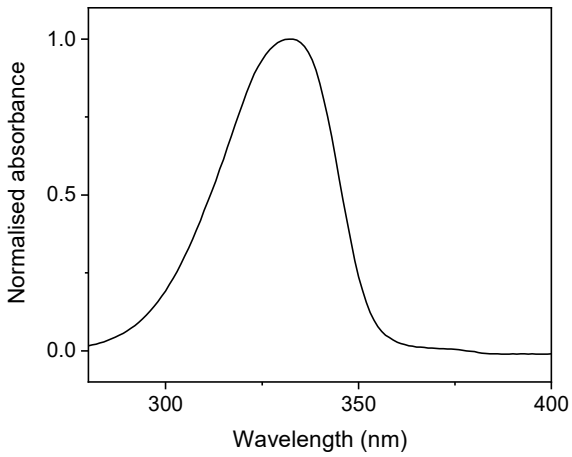


Figure 3.4: UV-Vis of purified shinorine from Helioguard^{TM} 365.

3.2.1 UHPLC-HRMS analysis

Shinorine: HRMS (ESI-TOF) *m/z*: $[M+H]^+$ Calculated for $C_{13}H_{21}N_2O_8^+$ 333.1292; Found 333.1291; error 0.3 ppm

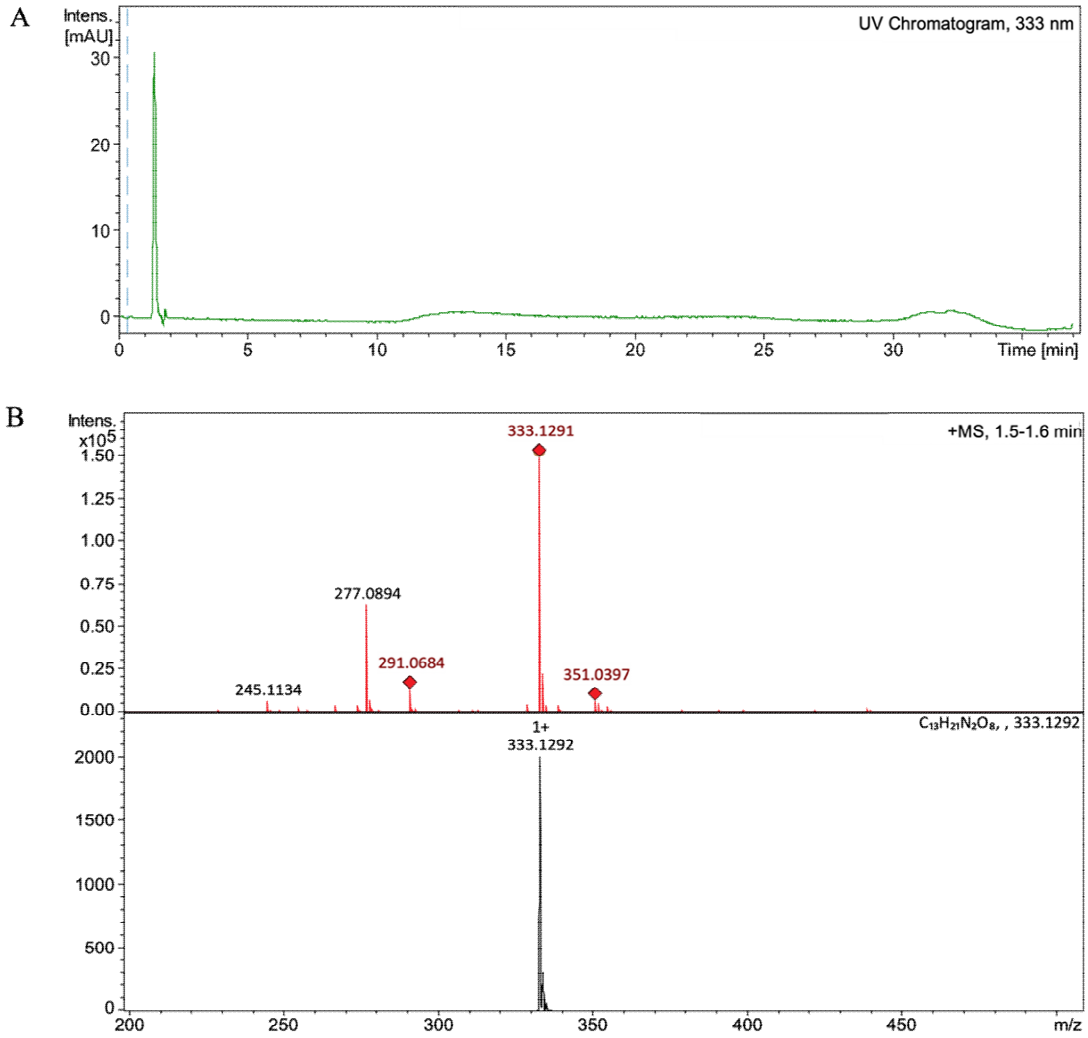


Figure 3.5: The chromatogram of purified shinorine from Helioguard 365. A: UV chromatogram at 333 nm for shinorine (Green). B: Comparison of the measured (Red) and simulated (Black) $[M+H]^+$ mass spectra for shinorine after UHPLC-HRMS analysis.

Porphyra-334 from Helioguard^{TM} 365 (λ_{max} = 333 nm)

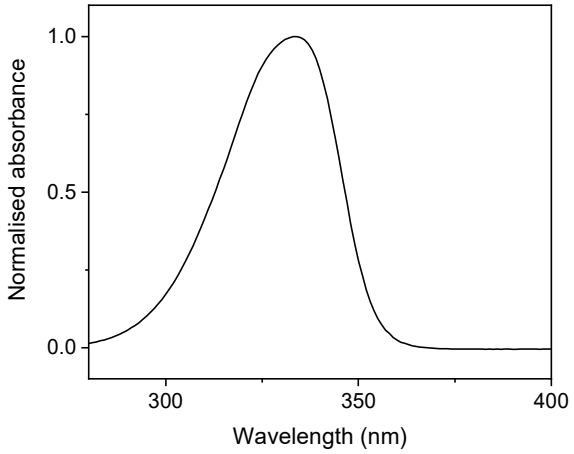


Figure 3.6: UV-Vis of purified porphyra-334 from Helioguard^{TM} 365.

Porphyra-334: HRMS (ESI-TOF) *m/z*: $[M+H]^+$ Calculated for $C_{14}H_{23}N_2O_8^+$ 347.1449; Found 347.1448 ; error 0.3 ppm

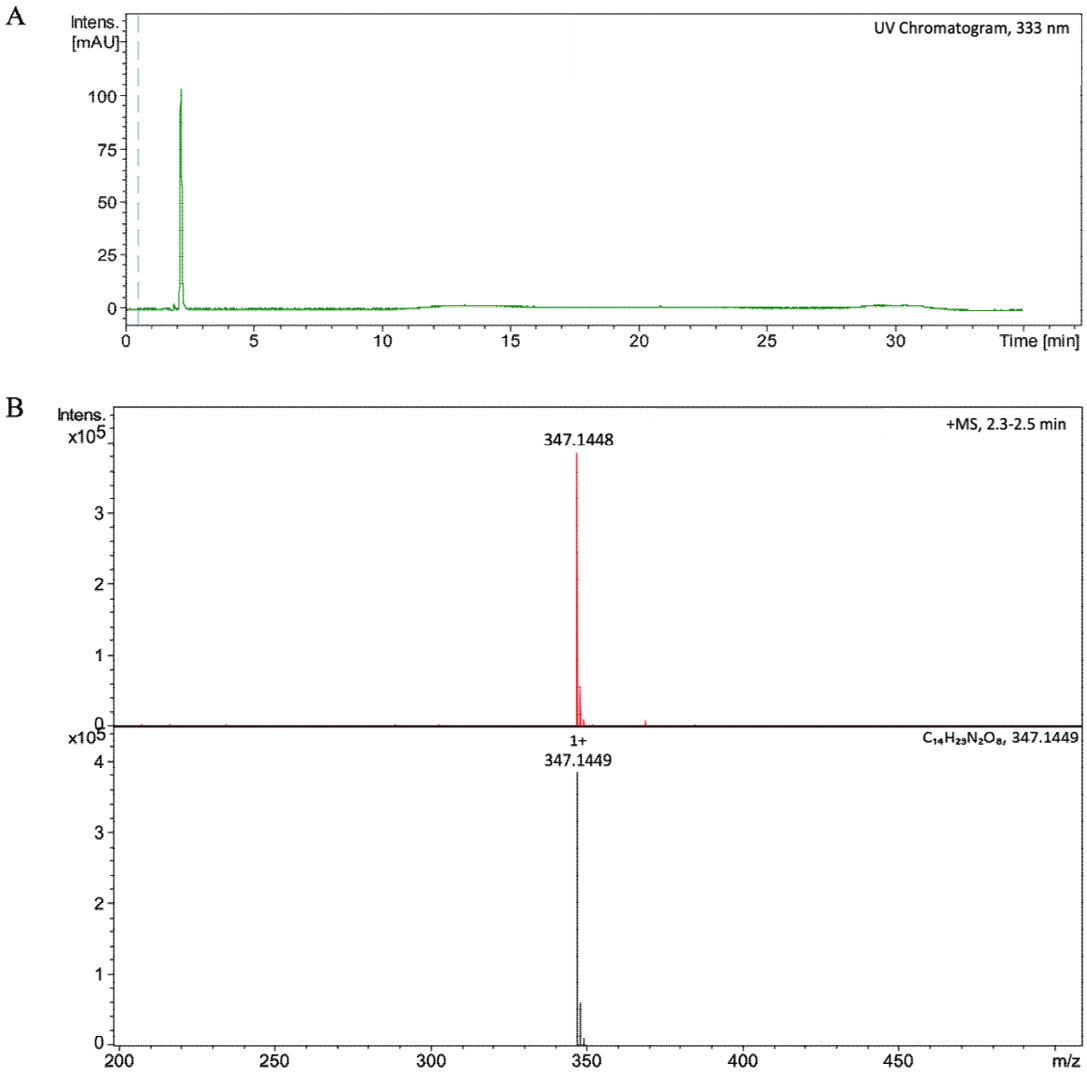
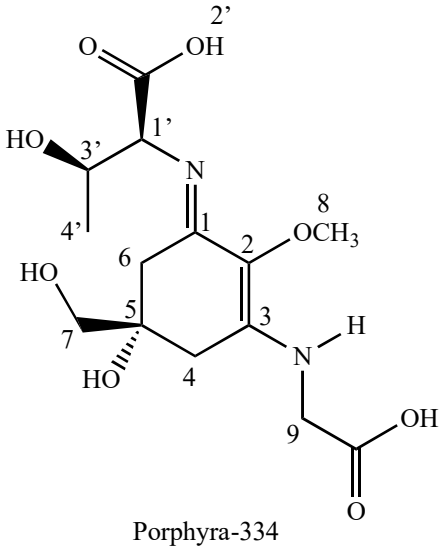


Figure 3.7: The chromatogram of purified porphyra-334 from Helioguard 365. A: UV chromatogram at 333 nm for porphyra-334 (Green). B: Comparison of the measured (Red) and simulated (Black) $[M+H]^+$ mass spectra for porphyra-334 after UHPLC-HRMS analysis.

3.2.2 ^{1}H NMR analysis of porphyra-334

Porphyra-334, being a UVA absorber has a high molar extinction coefficient, ε = 42,300 M^{-1} cm^{-1} at λ_{max} = 334 nm along with its significant photostability[114,116]. The ^{1}H NMR data of porphyra-334 obtained in this study was in agreement with a previous study carried out by Klisch *et al.*,[117]. The ^{1}H NMR spectra of porphyra-334 obtained in deuteriumm oxide is reported below.



Porphyra-334

Table 3.1: ^{1}H NMR data (400 MHz, δ in ppm, *J* in Hz) of porphyra-334 in D_2O

Position	**δ_H, in ppm (*J* in Hz)**
4	2.70 (1H, d, *J* = 17.8 Hz); 2.80 (1H, d, *J* = 17.8 Hz)
6	2.68 (1H, d, *J* = 17.6 Hz); 2.83 (1H, d, *J* = 17.6 Hz)
7	3.49 (2H, s)
8	3.61 (3H, s)
9	4.11 (2H, s)
1'	4.19 (1H, d, *J* = 4.0 Hz)
4'	1.18 (3H, d, *J* = 6.4 Hz)

OH, NH, COOH protons exchange for deuterium forming OD, ND, COOD, thus cannot be detected from the ^{1}H NMR spectrum.

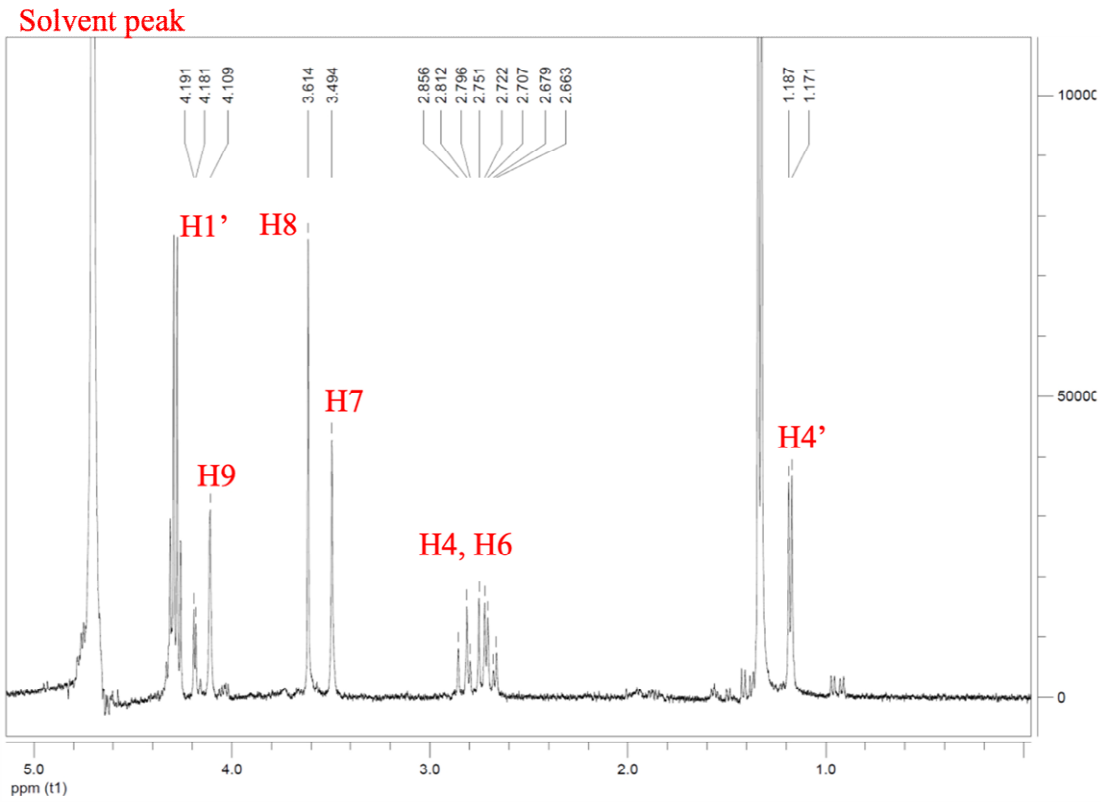


Figure 3.8: ^{1}H NMR spectrum (400 MHz) of porphyra-334 in D_2O with each peak highlighted.

3.3 Extraction, Purification and Characterisation of MAAs from *Palmaria palmata*

Palmaria palmata has long been known to produce MAAs. *Palmaria palmata* (Dulse) were bought from various companies including the Atlantic kitchen dulse seaweed, Icelandic söl dulse seaweed, the Cornish seaweed dulse and dulse seaweed seaspoon on Amazon as shown below. Dulse is known to be an edible seaweed and is even considered as a delicacy in Ireland.

Figure 3.9: Different sources of *Palmaria palmata* (Dulse) used in this study. All the samples were bought from Amazon.

Several MAAs have successfully been identified from the edible rhodophyte *P. palmata* by 25% methanolic extraction as shown below.

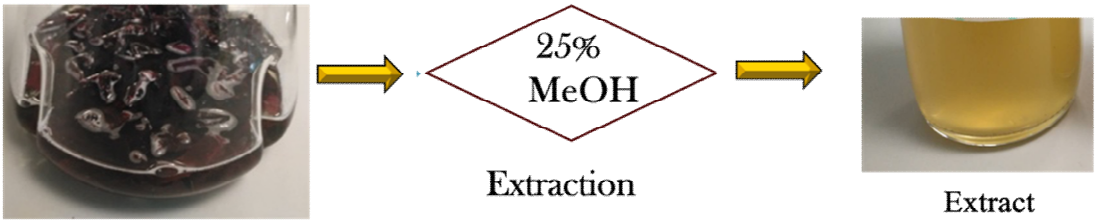


Figure 3.10: Extraction of MAAs from *Palmaria palmata* using 25% methanol.

Dulse from four different sources were extracted with 25% methanol. However, the sample that was focused on was the Cornish seaweed company where seaweed flakes (20 g) were extracted and prepared for analysis. The HPLC profile below showed one HPLC run from 5% to 25% methanol for 30 mins (as explained in detail in Chapter 2) at two different wavelengths to have a preliminary overview of everything produced under this first pre-purification step.

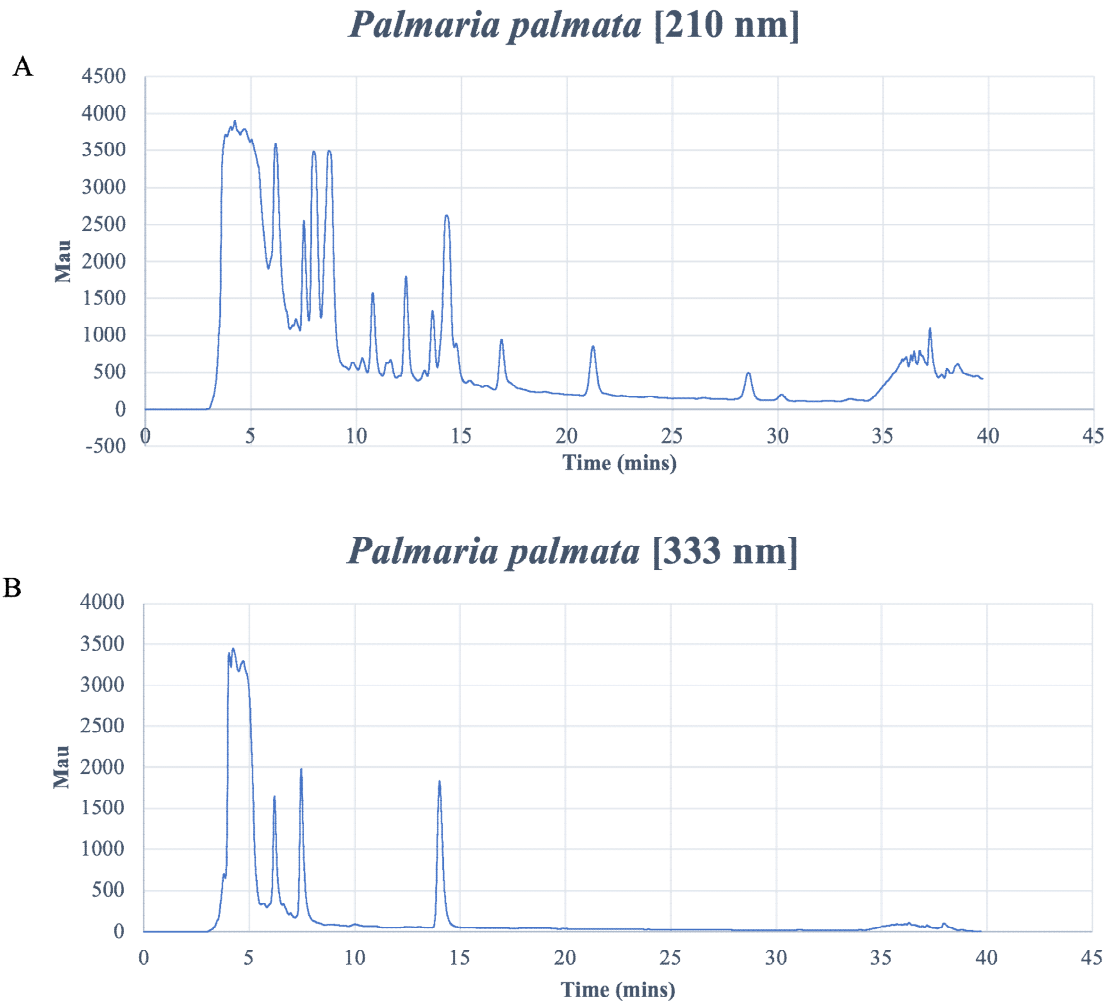


Figure 3.11: HPLC profile of *P. palmata* after 25% methanolic extraction. A: The chromatogram shows the wavelength at 210 nm. B: The chromatogram shows the wavelength at 333 nm.

Further purification (as described in chapter 2) gave rise to two pure MAAs namely shinorine [10 mg] and prophyra-334 [20 mg]; palythine and asterina-330 were two other MAAs that were detected in particular fractions (not pure) where the latter were prepared for UHPLC-HRMS for confirmation.

3.3.1 UHPLC-HRMS analysis

[1] Shinorine (λ_{max} = 333 nm)

Shinorine: HRMS (ESI-TOF) *m/z*: $[M+H]^+$ Calculated for $C_{13}H_{21}N_2O_8^+$ 333.1292; Found 333.1291; error 0.3 ppm

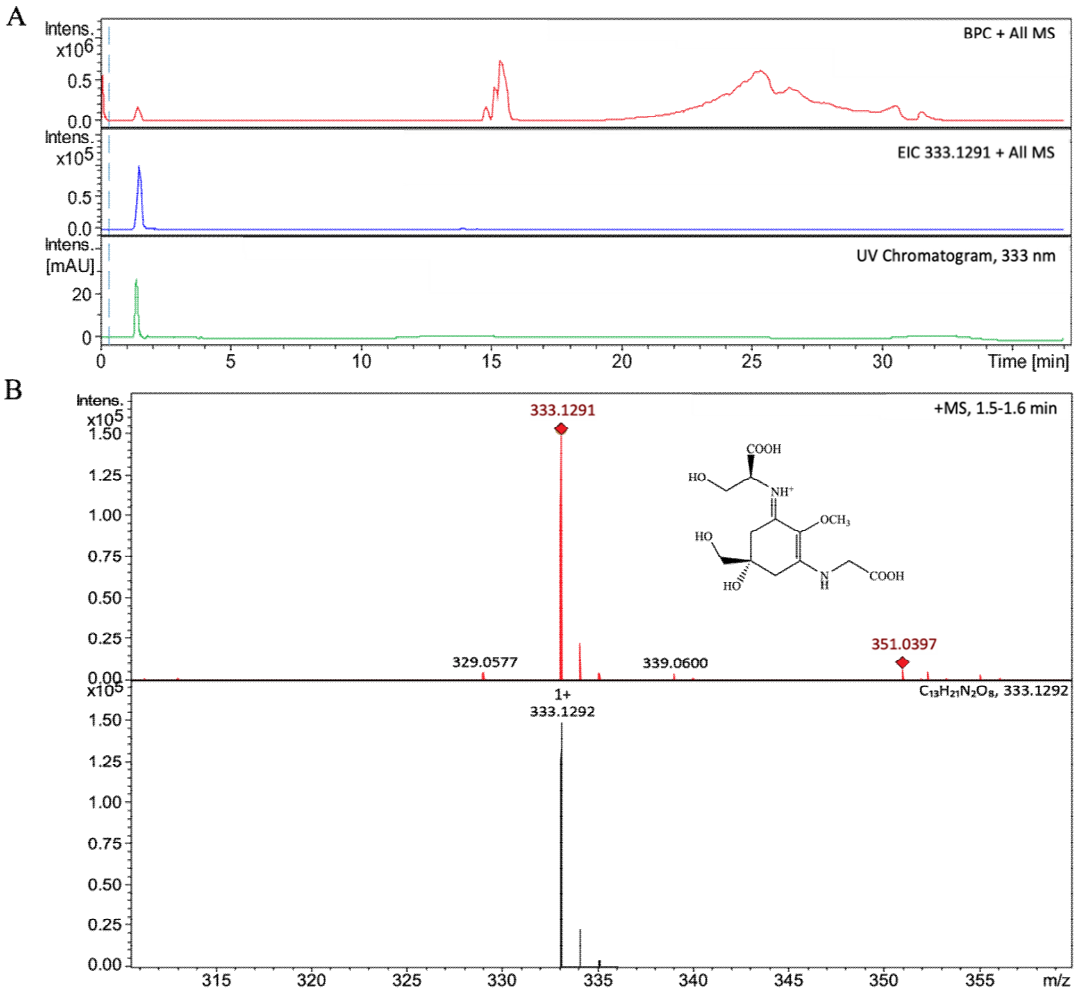


Figure 3.12: The chromatogram of purified shinorine from *P. palmata*. A: Base peak chromatogram (BPC) (Red); Extracted ion chromatogram (EIC) for *m/z* 333.1291 corresponding to shinorine (Blue); UV chromatogram at 333 nm for shinorine (Green). B: Comparison of the measured (Red) and simulated (Black) $[M+H]^+$ mass spectra for shinorine after UHPLC-HRMS analysis.

[2] Porphyra-334 (λ_{max} = 333 nm)

Porphyra-334: HRMS (ESI-TOF) *m/z*: $[M+H]^+$ Calculated for $C_{14}H_{23}N_2O_8^+$ 347.1449; Found 347.1448 ; error 0.3 ppm

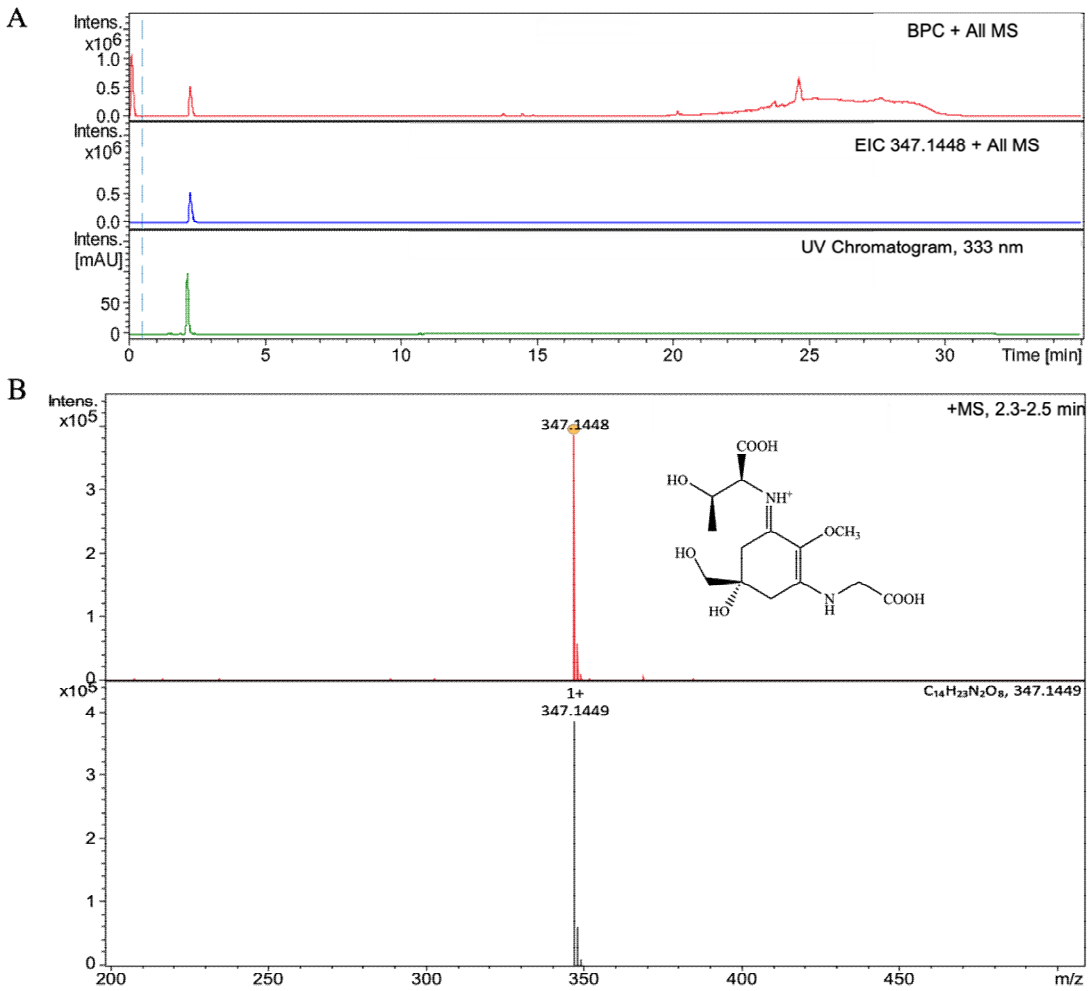


Figure 3.13: The chromatogram of purified porphyra-334 from *P. palmata*. A: Base peak chromatogram (BPC) (Red); Extracted ion chromatogram (EIC) for *m/z* 347.1448 corresponding to porphyra-334 (Blue); UV chromatogram at 333 nm for porphyra-334 (Green). B: Comparison of the measured (Red) and simulated (Black) $[M+H]^+$ mass spectra for porphyra-334 after UHPLC-HRMS analysis.

Detection of other MAAs

[1] Palythine (λ_{max} = 320 nm)

Palythine: HRMS (ESI-TOF) *m/z*: $[M+H]^+$ Calculated for $C_{10}H_{17}N_2O_5^+$ 245.1132; Found 245.1131 ; error 0.4 ppm

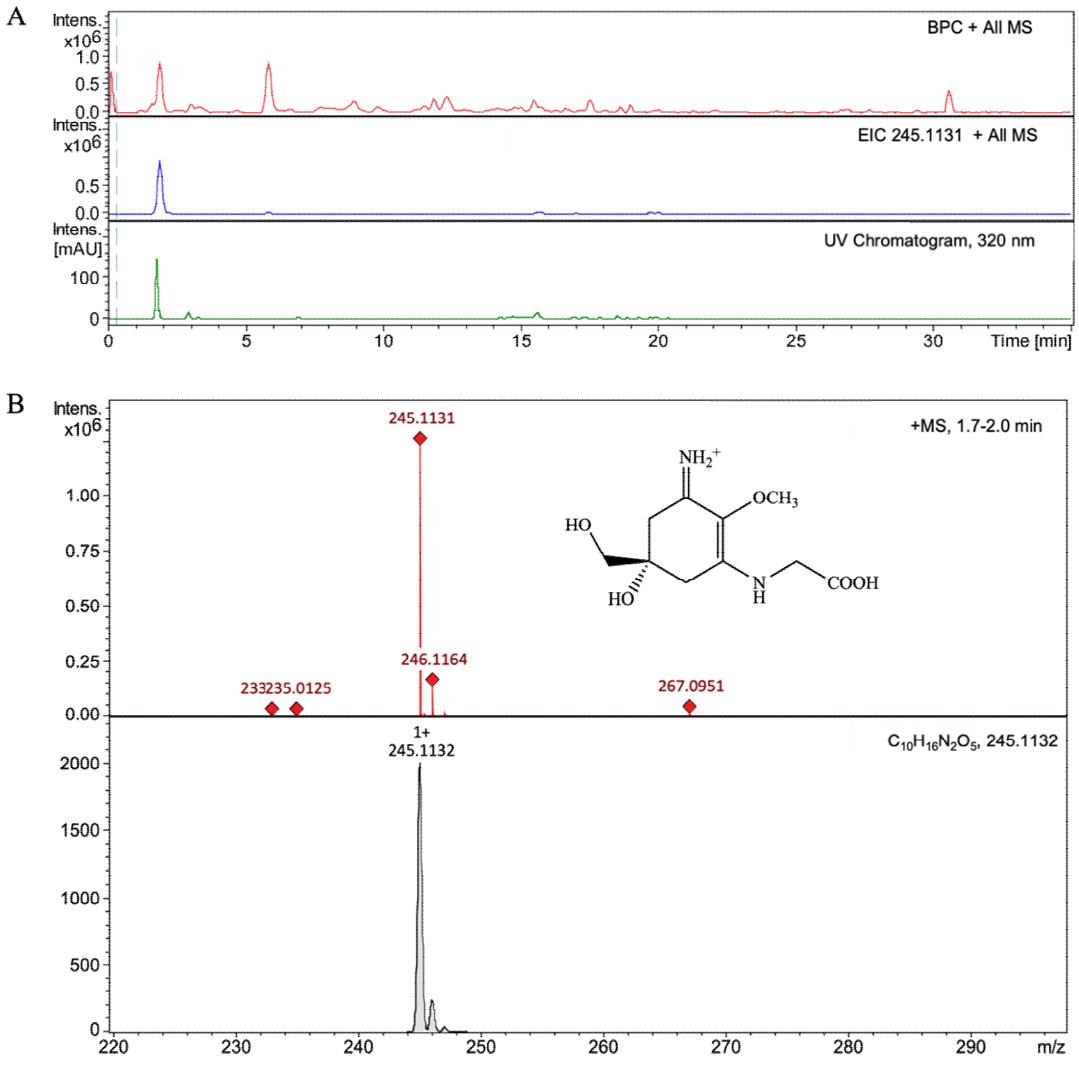


Figure 3.14: The chromatogram of palythine from *P. palmata*. A: Base peak chromatogram (BPC) (Red); Extracted ion chromatogram (EIC) for *m/z* 245.1131 corresponding to palythine (Blue); UV chromatogram at 320 nm for palythine (Green). B: Comparison of the measured (Red) and simulated (Black) $[M+H]^+$ mass spectra for palythine after UHPLC-HRMS analysis.

[2] Asterina-330 (λ_{max} = 330 nm)

Asterina-330: HRMS (ESI-TOF) *m/z*: $[M+H]^+$ Calculated for $C_{12}H_{21}N_2O_6^+$ 289.1394; Found 289.1396 ; error 0.6 ppm

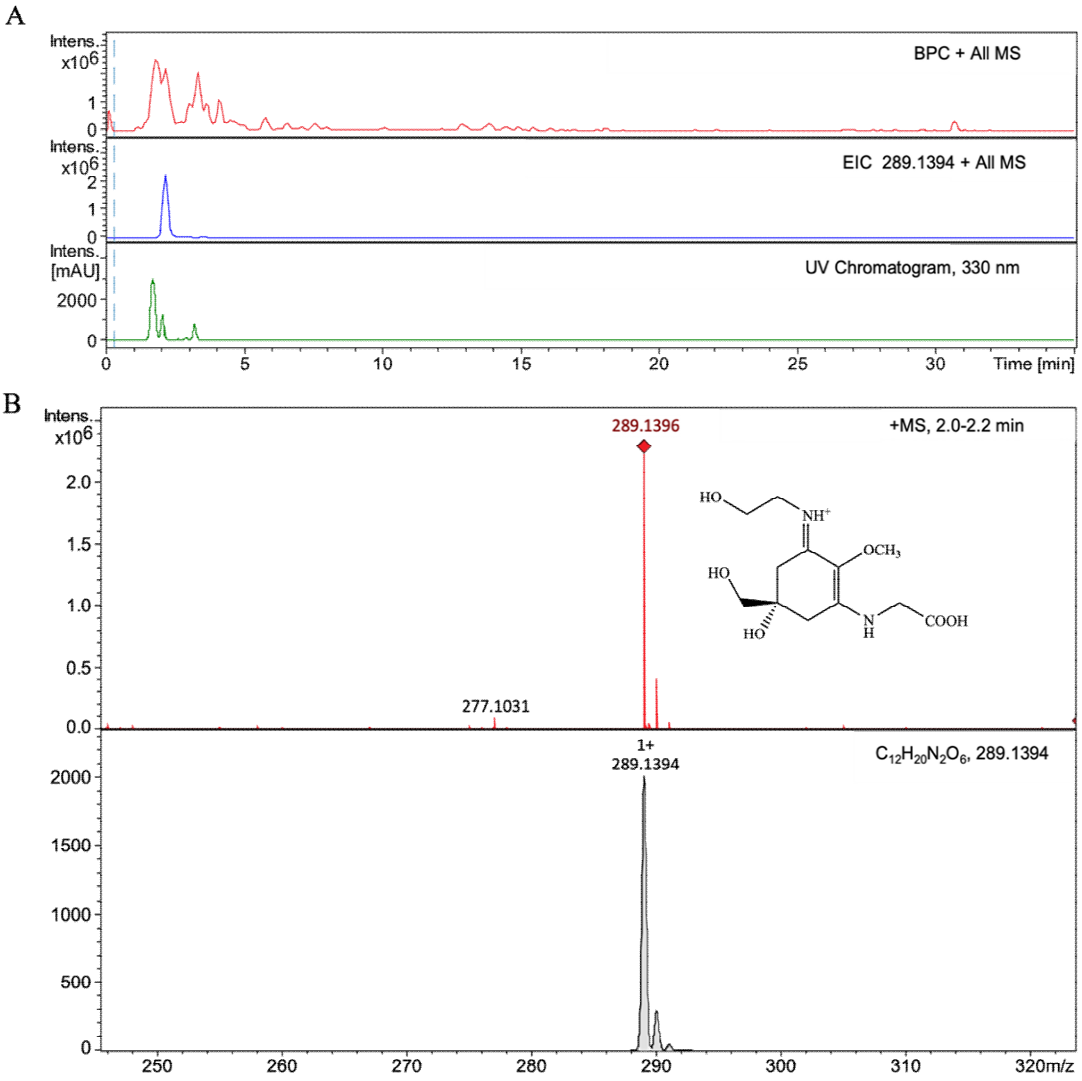


Figure 3.15: The chromatogram of asterina-330 from *P. palmata*. A: Base peak chromatogram (BPC) (Red); Extracted ion chromatogram (EIC) (Blue) for *m/z* 289.1394 corresponding to asterina-330; UV chromatogram (Green) at 320 nm for asterina-330. B: Comparison of the measured (Red) and simulated (Black) $[M+H]^+$ mass spectra for asterina-330 after UHPLC-HRMS analysis.

3.4 Focus on the high UVA-range absorbers namely usujirene/palythene

One of the main focus of this study was the purification of the known MAAs usujirene/palythene from *Palmaria palmata* which absorb at 357 nm and 360 nm respectively. So far, these cis-trans isomers are among the highest UVA-range absorbers after the MAA euhalothece-362, isolated from a halophilic cyanobacterium *Euhalothece* sp. on a gypsum crust on the bottom of a hypersaline pond in Israel,

which has an absorption maximum of 362 nm[41]. Traces of usujirene/palythene was detected in one of the dulse samples bought on Amazon, which was the Cornish seaweed sample. However, due to its extremely low yield, it was challenging to get enough during purification. Consequently, more samples were bought in the hope of being able to purify enough for further analysis.

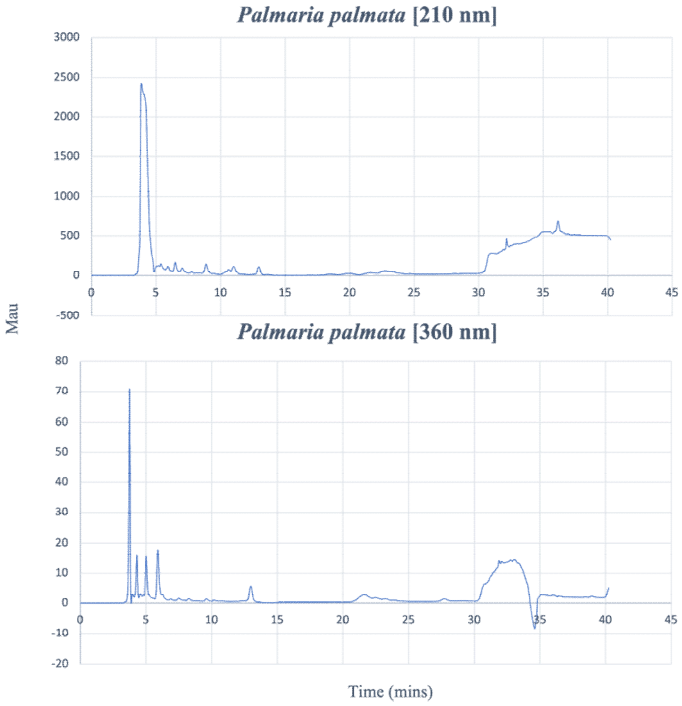


Figure 3.16: HPLC profile of *P. palmata* (Cornish Seaweed Sample) after 25% methanolic extraction. The chromatograms show the wavelength at 210 nm and 360 nm. The small peak at around 13 mins showed the presence of usujirene/palythene mix.

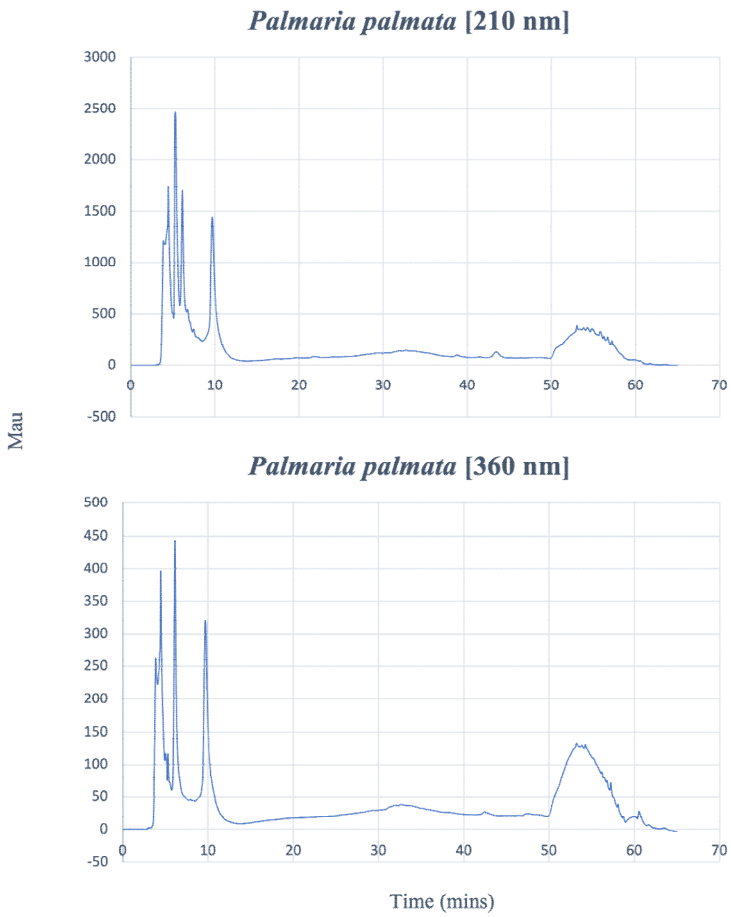


Figure 3.17: HPLC profile of *P. palmata* (Sol Dulse Sample) after 25% methanolic extraction. The chromatograms show the wavelength at 210 nm and 360 nm. No detection of usujirene/palythene.

It was crystal clear to see that no matter how many samples were bought, the yield of usujirene/palythene was simply indirectly proportional to the seaweed sample. Some samples which came from the same supplier did not have any of these cis-trans UV absorbers whereas other batches from the same supplier revealed traces (was so little to even get an appropriate amount) and finally one of the batches that we received from the supplier contained a high amount of usujirene/palythene as shown below.

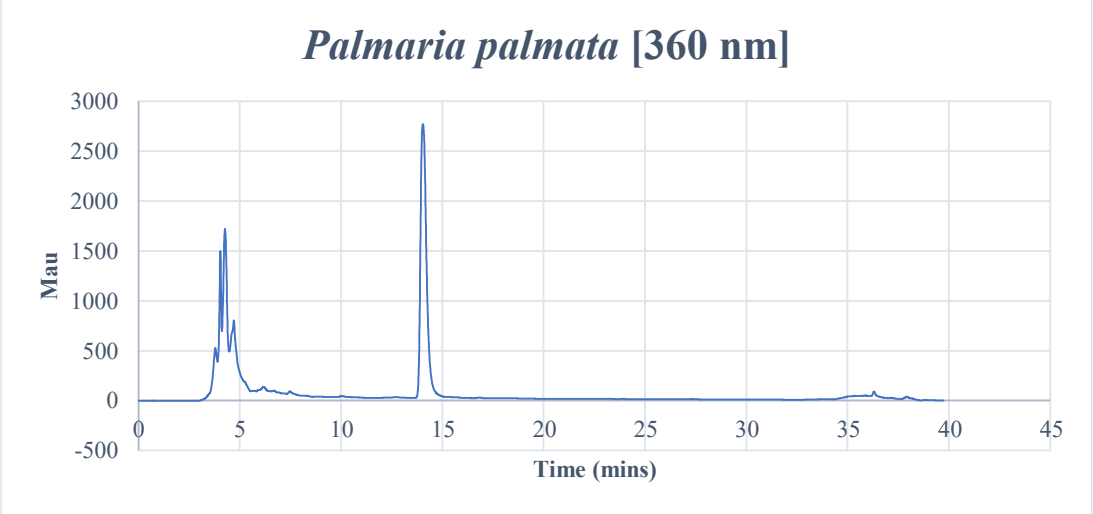


Figure 3.18: HPLC profile of *P. palmata* after 25% methanolic extraction. The chromatogram shows the wavelength at 360 nm resulting in purification of usujirene/palythene mixture at about 14 min.

During the hunt of these molecules, it became obvious to us that the yield of usujirene/palythene might be dependent on the area of sample collection, the weather conditions including cloud cover and the amount of UV radiation received in that particular area, the chemo-physical parameters of the water such as salinity, dissolved oxygen as well as the depth in which the seaweeds were collected. One interesting study carried out by Karsten *et al*., clearly showed that the yield of MAAs was directly proportional to the depth at which they were collected, where the greatest yield of the total MAAs was seen to be 2.74 mg g^{-1} dry weight at 1.5 m compared to 0.68 mg g^{-1} dry weight at 3 m in depth. In simpler terms, the shallower the collection point, the higher is the amount of MAAs produced. And to prove that concept, they did a transplantation study to show the variation in yield of MAAs from 3 m deep to 0.2 m deep. To complement this study, they also tested the impact of the full solar radiation along with and without UVA and UVB respectively to monitor the level of MAAs in *Palmaria palmata* where it was seen that the most significant total MAAs yield was during the full solar spectrum[72]. As such, it is reliable enough to say that the yield of MAAs, here especially referring to usujirene/palythene, is to a great extent dependent on their environment.

After extraction from various *Palmaria palmata* batches from the Cornish seaweed supplier, I finally managed to obtain a significant amount (70 mg) of usujirene/palythene for further analysis.

Usujirene/Palythene (λ_{max} = 357/360 nm)

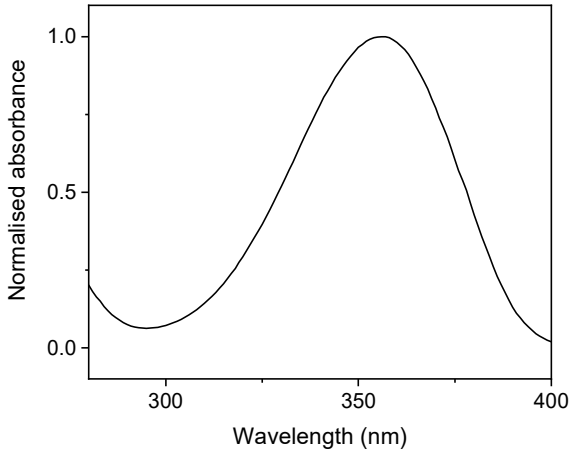


Figure 3.19: UV-Vis of purified usujirene/ palythene from *Palmaria palmata*.

3.4.1 UHPLC-HRMS analysis

Usujirene/palythene: HRMS (ESI-TOF) *m/z*: $[M+H]^+$ Calculated for $C_{13}H_{21}N_2O_5^+$ 285.1445; Found 285.1449; error 1.4 ppm

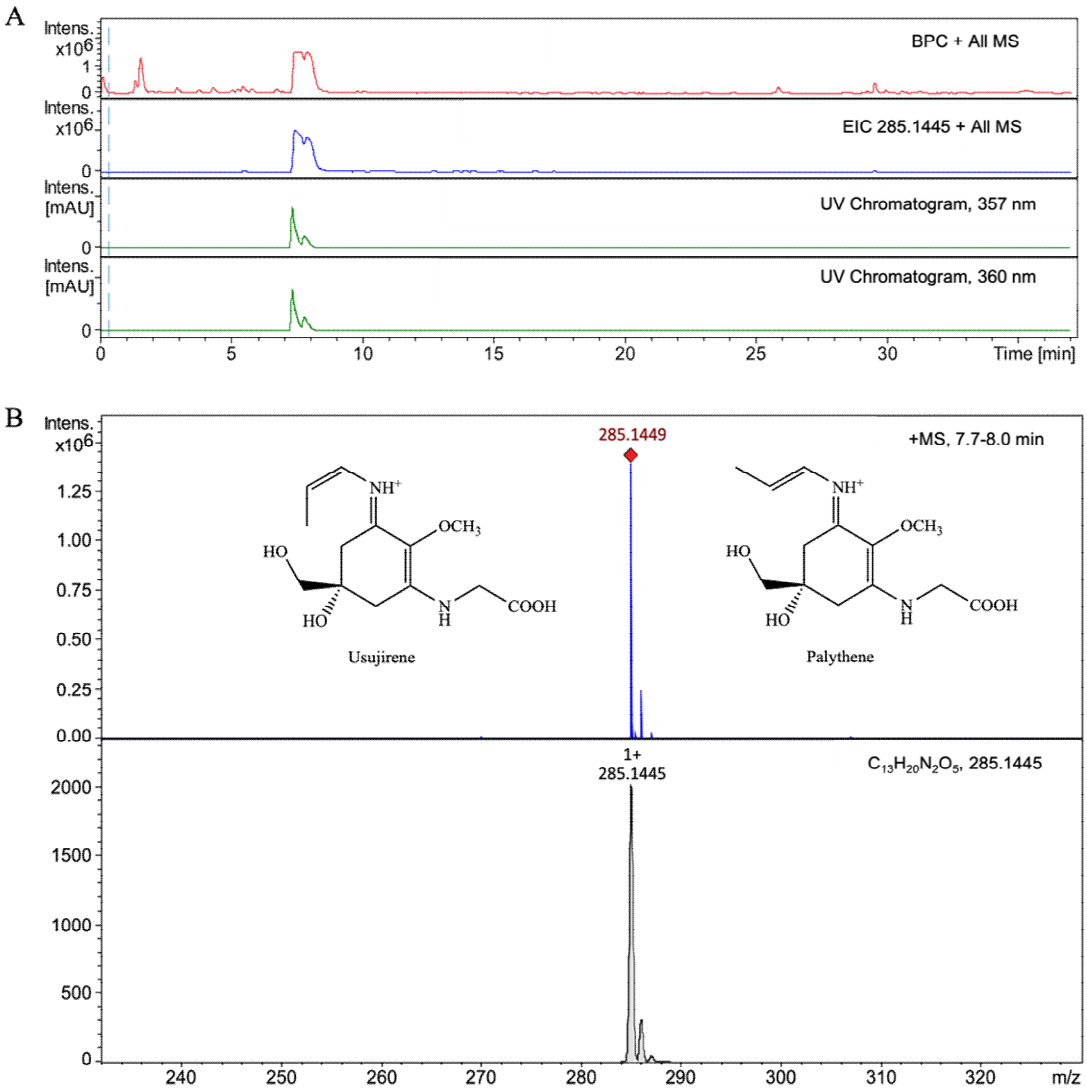


Figure 3.20: Chromatograms after UHPLC-HRMS analysis of the usujirene and palythene collected fractions. A: Base peak chromatogram (BPC) (Red). Extracted ion chromatogram (EIC) for m/z 285.1445 corresponding to usujirene and palythene (Blue). UV chromatogram at 357 nm and 360 nm (Green). B: Comparison of the measured (Red) and simulated (Black) $[M+H]^+$ mass spectra for usujirene and palythene.

Palythine: HRMS (ESI-TOF) *m/z*: $[M+H]^+$ Calculated for $C_{10}H_{17}N_2O_5$+ 245.1132; Found 245.1131; error 0.4 ppm

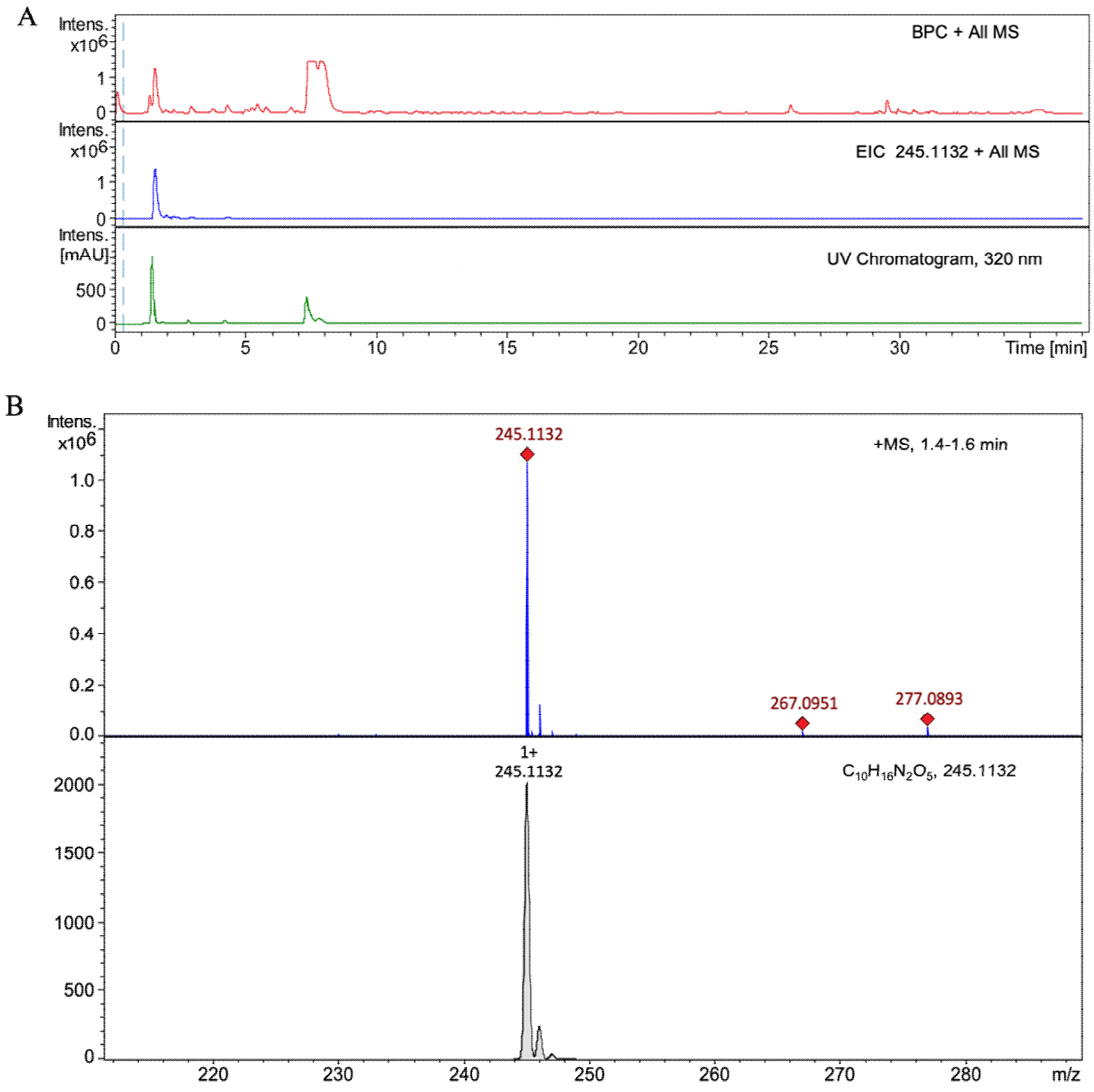


Figure 3.21: Chromatograms after UHPLC-HRMS analysis of the usujirene and palythene collected fractions containing palythine. A: Base peak chromatogram (BPC) (Red). Extracted ion chromatogram (EIC) for m/z 245.1132 corresponding to palythine (Blue). UV chromatogram at 320 nm (Green). B: Comparison of the measured (Red) and simulated (Black) $[M+H]^+$ mass spectra for palythine.

An important factor to consider while purifying usujirene/palythene was the elimination of formic acid in both solvent A (water) and solvent B (methanol) due to the fact that these cis-trans isomers were seen to be prone to palythine conversion under acidic conditions. Therefore, acids were avoided during the whole purification process to lengthen stability of these UVA absorbers.

3.4.2 Proposed mechanism for the conversion of usujirene/ palythene to palythine under acidic conditions

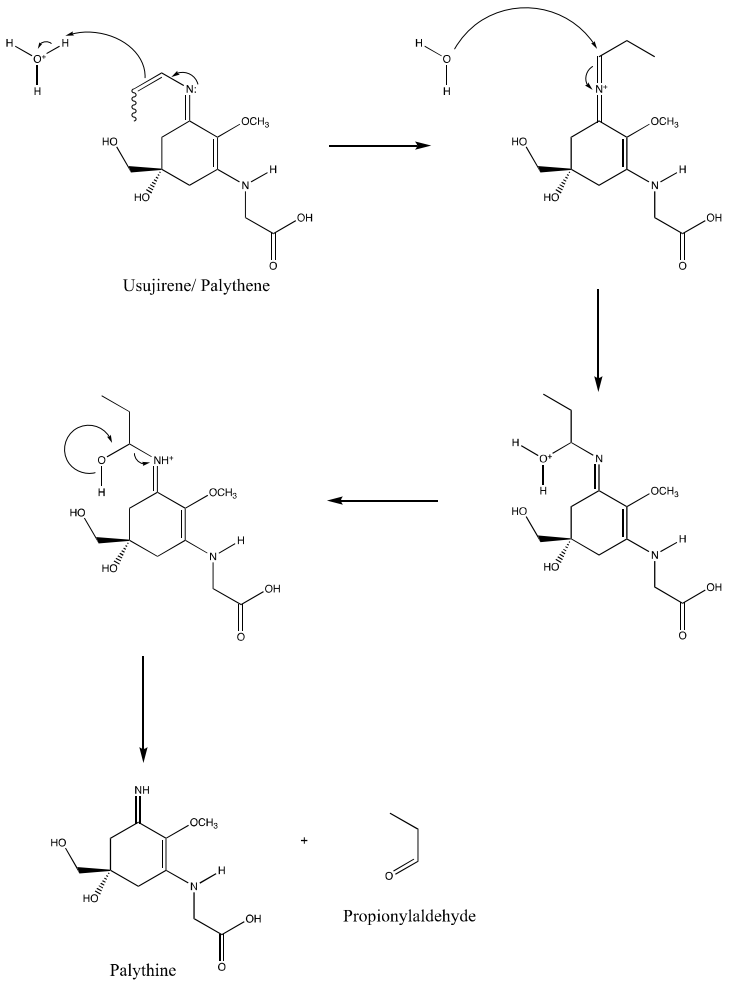
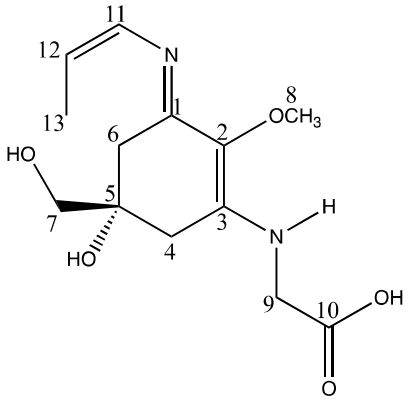


Figure 3.22: A proposed mechanism involving the hydrolysis of usujirene/ palythene to form palythine under acidic conditions.

The lone pair of electrons on the nitrogen atom of the mycosporine backbone get transferred to the adjacent C-C bond until the C=C attack the electron deficient H_3O^+ forming a C-OH attached to the mycosporine backbone leaving the nitrogen atom electron deficient. The O-H bond connected to the C-CH_2CH_3 get released, thus giving back the required electron to the nitrogen atom, resulting in the formation of palythine and propionylaldehyde as end products.

3.4.3 NMR analysis of usujirene/palythene based on literature

The ^{13}C NMR and ^{1}H NMR of usujirene based a paper entitled "Antioxidant Effect of the Constituents of Susabinori (*Porphyra yezoensis*)" [118].

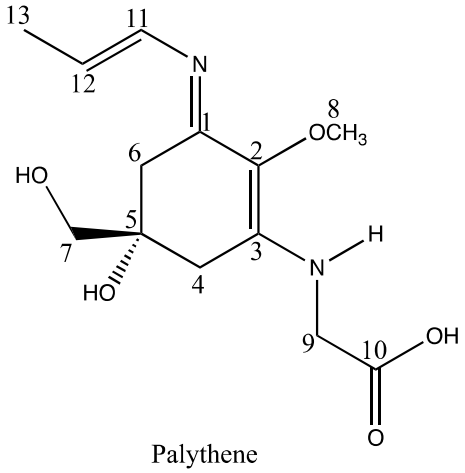


Usujirene

Table 3.2: ^{1}H NMR data (δ in ppm, *J* in Hz) of usujirene in D_2O[118]

Position	δ_C in ppm; D_2O	δ_H, in ppm, D_2O (*J* in Hz)
1	155.6	
2	127.9	
3	164.1	
4	34.9	2.74 (1H, *d*, *J* = 17.6), 2.92 (1H, *d*, *J* = 17.6 Hz)
5	72.2	
6	34.6	2.74 (1H, *d*, *J* = 17.6), 2.91 (1H, *d*, *J* = 17.6 Hz)
7	69.4	3.47 (2H, *s*)
8	59.9	3.68 (3H, *s*)
9	48.2	3.99 (2H, *s*)
10	173.4	
11	123.6	6.41 (1H, br. *d*, *J* = 7.6 Hz)
12	116.5	5.29 (1H, *dq*, *J* = 7.6, 7.1 Hz)
13	11.4	1.78 (3H, *dd*, *J* = 1.7, 7.1 Hz)

The ^{13}C NMR and ^{1}H NMR of palythene based a paper entitled "Isolation and structure of two new amino acids, palythinol and palythene from the zoanthid *Palythoa tuberculosa*" [119].



Palythene

Table 3.3: ^{1}H NMR data (δ in ppm, *J* in Hz) of palythene in D_2O[119]

Position	δ_C in ppm; D_2O	δ_H, in ppm, D_2O (*J* in Hz)
1	161.5	
2	126.4	
3	154.2	
4	33.8	2.88 (2H, ABq, *J* = 17 Hz)
5	71.8	
6	33.8	2.96 (2H, ABq, *J* = 17 Hz)
7	68.4	3.66 (2H, *s*)
8	60.3	3.71 (3H, *s*)
9	47.6	4.11 (2H, *s*)
10	175.4	
11	124.5	6.58 (1H, br. *d*, *J* = 13 Hz)
12	117.9	5.75 (1H, *dq*, *J* = 6, 13 Hz)
13	15.2	1.88 (3H, *dd*, *J* = 2, 6 Hz)

3.4.4 NMR analysis of purified usujirene/palythene from *P. palmata*

The ^{1}H NMR analysis was carried out in D_2O using a 400 MHz NMR spectrometer (Avance III, Bruker). The ^{1}H NMR spectrum is shown below noting the presence of usujirene/ palythene with some impurities in the sample.

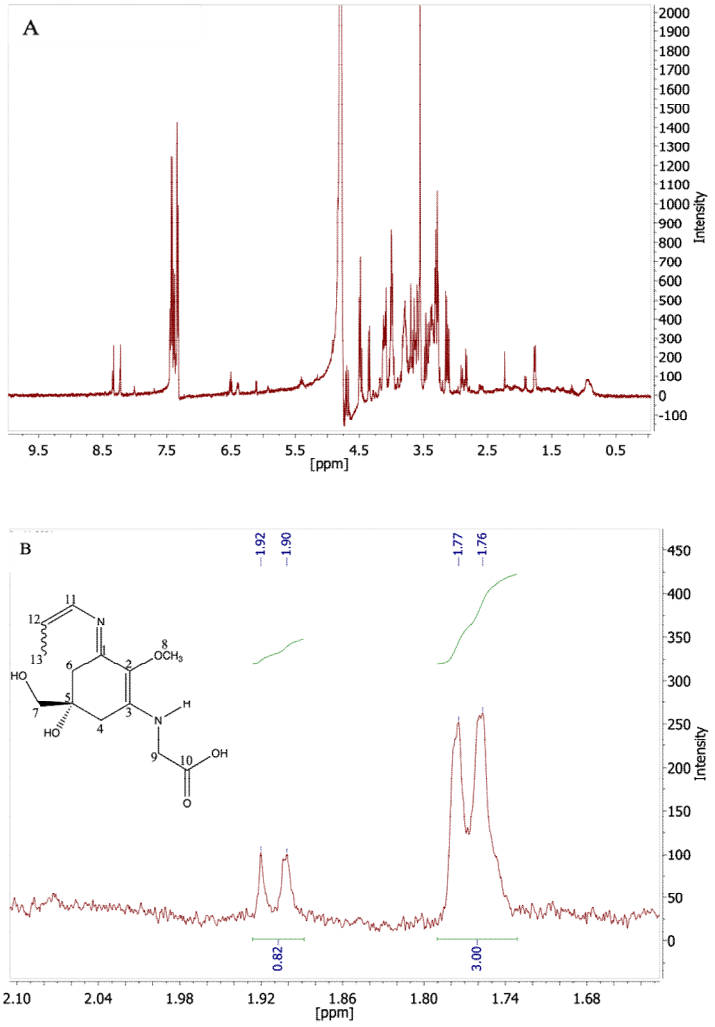


Figure 3.23: A: ^{1}H NMR spectrum of usujirene/ palythene in D_2O. B: ^{1}H NMR spectrum of usujirene/ palythene in D_2O around 1.7-2.1 ppm to focus on the position of the methyl group on C-13.

The signal corresponding to the methyl group on C-13 was focused on based on previous assignments[118,119]. The ratio of usujirene/ palythene was calculated by integrating these particular signals which led to 3.7:1 respectively. As such, it was clear to see that the ratio was consistent with the UV chromatograms shown above.

3.5 Transient Electronic Absorption Spectroscopy (TEAS)

The purpose of this research was to elucidate the photoprotective mechanism of the various MAAs purified and characterised by using Transient Electronic Absorption Spectroscopy (TEAS). The concept of TEAS relates to the ability to track the energy

flow and dynamics of the molecule after photoexcitation, thus resulting in understanding its photoprotective mechanism.

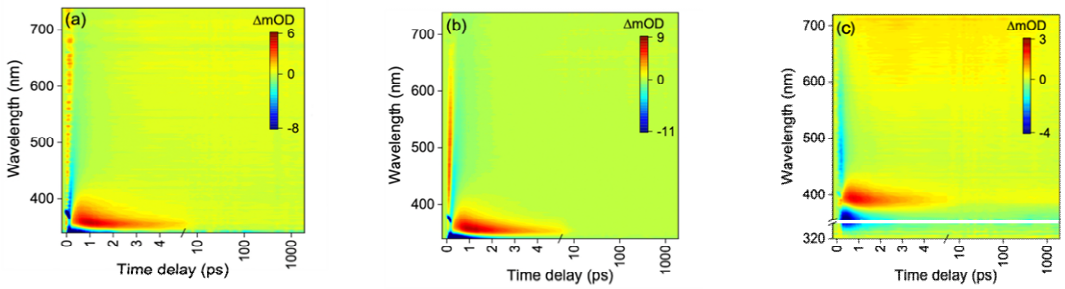


Figure 3.24: Transient absorption spectra (TAS) presented as false colour heat maps for (a) shinorine; (b) porphyra-334; (c) usujirene/palythene photoexcited at their respective λ_{max}. Time delays are plotted linearly until 5 ps and then as a logarithmic scale from 5 to 1900 ps.

The transient absorption spectra (TAS) in Figure 3.24 had similar spectral signatures for each studied MAA. Immediately after time-zero, there was a ground state bleach (negative feature) that appeared in the same spectral region as each MAAs UV-visible spectrum and represented a depletion in electronic ground state (S_0) population upon photoexcitation. Moreover, from time-zero was stimulated emission (negative feature) which covered a broad spectral region between ~350-700 nm. Stimulated emission is when an incoming photon interacted with a molecule in the excited state resulting in it dropping to a lower energy level and emitting a second photon in the process. The presence of this feature implied that the molecule was in its electronic excited state and given that spectral red shifting was observed with time suggested that the MAAs were evolving along their electronic excited state (S_1) potential energy surface towards the S_1/S_0 conical intersection (a point at which two energy levels are degenerate). Shifted from time-zero was an excited state absorption (positive feature) and indicated an absorption from a state that was not the original ground state of the molecule. It was believed that this excited state absorption was due to vibrational cooling of the MAAs from a high lying vibrational level in the electronic ground state to the ground state vibrational level in the electronic ground state after passing through the S_1/S_0

conical intersection. At this point, the MAAs had fully relaxed and were able to reabsorb another photon.

A very quick conversion of the absorbed photon energy to harmless heat within ~1 ps demonstrated the efficiency of these MAAs as photoprotective molecules as shown on the false colour heat maps in Figure 3.24. Fast relaxation back to the molecules original ground state is ideal as it reduces the probability for harmful side reactions leading to photoproducts or the reemission of a photon.

3.6 Discussion

The MAAs shinorine and porphyra-334 were successfully purified from Helioguard 365 while the cis-trans isomers usujirene/palythene were extracted from the rhodophyte *Palmaria palmata* using 25% methanol and characterised using UHPLC-HRMS as well as ^{1}H NMR. One issue during this experiment was the detection of usujirene/palythene in the various samples utilized. It was seen that the presence of usujirene/palythene was not consistent in all the different batches bought from either the same or different suppliers compared to the presence of shinorine and porphyra-334 no matter what supplier/ sources they came from. The reason for this might have been the physico-chemical environment such as salinity, temperature, cloud cover, UV intensity, or nutrients available in that area to name a few[120–122]. Depth at which the seaweed was collected might be another explanation for the inconsistency in the presence/ quantity of usujirene/palythene in the various samples[123]. A study involving the transplantation of *Palmaria palmata* from 3 m deep to 0.2 m deep along with the full solar radiation revealed the significantly higher yield of MAAs[72] indicating that the yield of MAAs is greatly dependent on their surroundings. Furthermore, the cis-trans isomers were noted to be prone to degradation under acidic conditions which was seen by its hydrolysis to the other MAA palythine. Therefore, the purification technique had to be optimized by eradicating formic acid usage in the solvents used to be able to acquire enough usujirene/palythene for further characterisation. Porphyra-334 was successfully characterised using ^{1}H NMR and the ratio of usujirene/palythene mix was also determined to be 3.7:1 respectively based on the integration of particular signals on the ^{1}H NMR spectrum.

It was seen that shinorine, porphyra-334 and the cis-trans isomers usujirene/palythene mix managed a quick conversion of the absorbed energy to harmless heat within

approximately 1 picosecond compared to the common UVA filter, avobenzone, which has been found to cause coral bleaching, hence banned in numerous countries[112]. Upon photoexcitation of avobenzone, a high proportion of the photoexcited population was not seen to return to its original ground state[108,109], as well as resulted in several photoproducts[110,111]. Furthermore, previous research focused on photostability of commercial sunscreens where a photo-unstable formulation is one with a AUCI (Area Under Curve Index: calculated by the area under the curve after irradiation divided by the area under the curve before irradiation) of less than 0.8 after 120 minutes of UVA and UVB irradiation[124,125]. Thus, photostability measurements on the MAAs showed an AUCI of greater than 0.8 which would automatically qualify them as being photostable. In addition, having such a fast relaxation mechanism back to its ground state, MAAs restrict the formation of harmful photoproducts which was also assessed by irradiation of the sample for several hours. The outcome of this constant irradiation revealed the decline of the peak absorbance by approximately 1% after 5 hours. In conclusion, the purified MAAs namely shinorine, porphyra-334 and the usujirene/palythene mix from both Helioguard 365 as well as *Palmaria p*almata respectively exhibited excellent photoprotective properties which can contribute into development of novel, eco-friendly, cost-effective UVA filters for use in sunscreen.

Chapter 4: Exploitation of the NRPS-like enzymes for the conversion of oxo-type to imino-type mycosporine-like amino acids

4.1 Brief Introduction

Anabaena variabilis is a terrestrial filamentous, heterocyst-forming cyanobacterium playing a major role in nitrogen fixation, thus, contributing significantly as a natural biofertilizer. Due to the high level of UV radiation impinging on natural habitats, *A. variabilis* along with other terrestrial or aquatic cyanobacteria like *Nostoc* sp., *Scytonema* sp. and *Moorea* sp., to name a few have developed mechanisms into coping with these incoming harmful rays which can lead to protein alteration, DNA damage[126,127,128,129,130], slow growth, nitrogen fixation issues or membrane permeability problems[126,131,132,133]. During their evolutionary history, cyanobacteria have come up with a diverse list of counterattacks such as DNA repair by photoreactivation or excision repair[134,135], carotenoids or specific antioxidant accumulation[136,137]. To add to that list is a very important asset developed by those organisms is the production of a natural UV shield known as mycosporine-like amino acids (MAAs)[138]. MAAs are small molecules with a basic cyclohexanone or cyclohexenimine core structure which absorb UV radiation in the range of 310 to 362 nm[41,139]. Their roles in those organisms are related to various biological processes such as photoprotection from harmful UV rays; antioxidant capacities; sporulation involvement[91,140,141] as well as their contribution in osmotic regulation[33,142].

The entire 6.5 kb biosynthetic gene cluster of the MAA shinorine was successfully identified in the cyanobacterium *Anabaena variabilis* by Balskus *et al*.,[77]. There are four biosynthetic genes in the cluster where a dehydroquinate synthase (DHQS) homolog [Ava_3858] adjacent to an O-methyltransferase (O-MT) [Ava_3867] gave rise to the cyclohexanone 4-deoxygadusol. The third enzyme namely an ATP-grasp homolog (Ava_3856) was revealed to convert 4-deoxygadusol into mycosporine-glycine in the presence of glycine, ATP and Mg^{2+} cofactors by phosphorylating the mycosporine backbone 4-deoxygadusol. The fourth enzyme involved is a NRPS-like enzyme which comprised of three domains including an adenylation (A), peptidyl

carrier protein (PCP) and thioesterase (TE) domains[104]. The binding pocket of the A domain of Ava_3855 was predicted to activate L-serine based on the active site residues[102] which was also confirmed by the ATP-^{32}PPi exchange assay[77], thus resulting in the generation of the MAA shinorine.

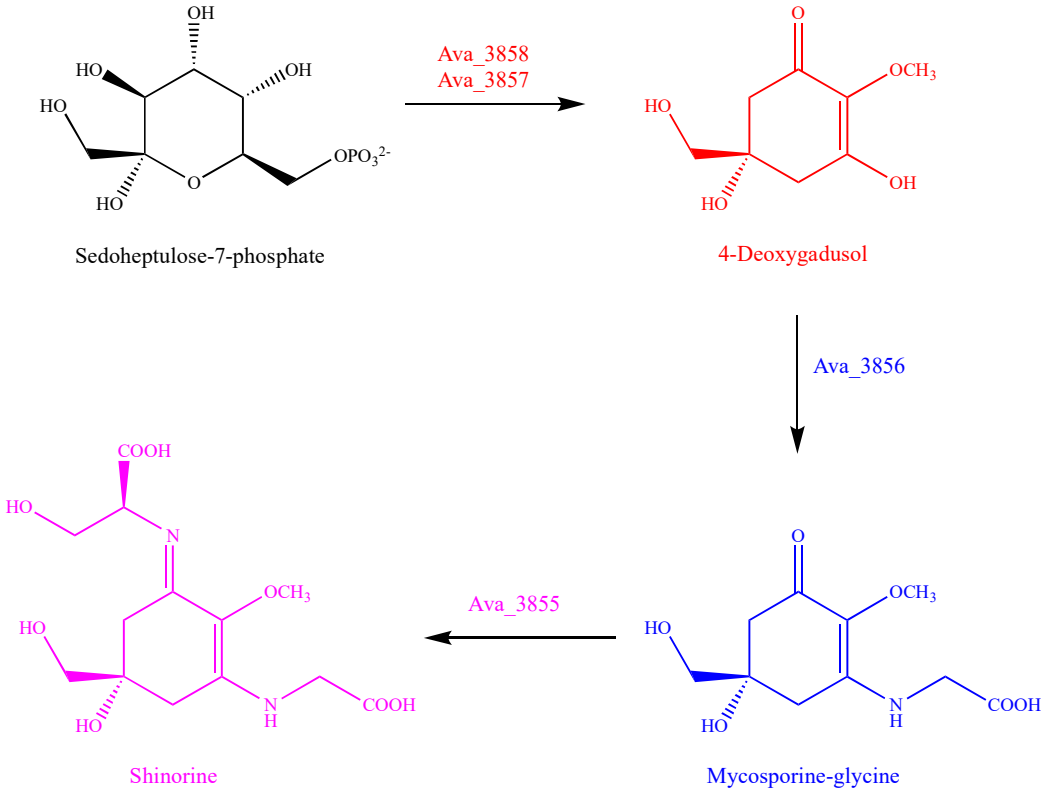


Figure 4.1: The biosynthetic pathway of shinorine from the terrestrial cyanobacterium *Anabaena variabilis*

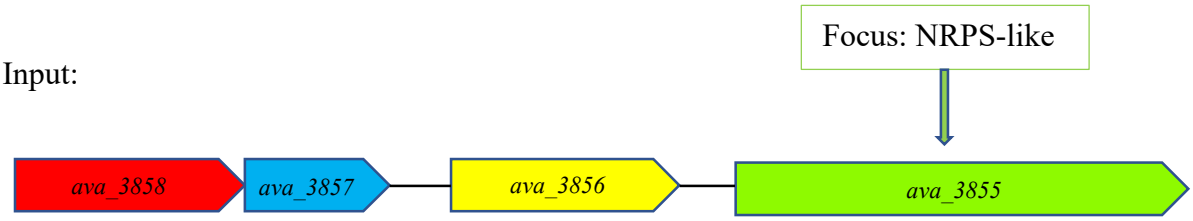
The goal of this experiment was to investigate the NRPS-like enzymes involved in MAAs production from various cyanobacteria as well as to confirm and demonstrate the functionality of the protein of interest into converting the oxo-type to imino-type MAAs.

4.2 Exploration and exploitation of NRPS-like enzyme involved in the generation of MAAs using Bioinformatics

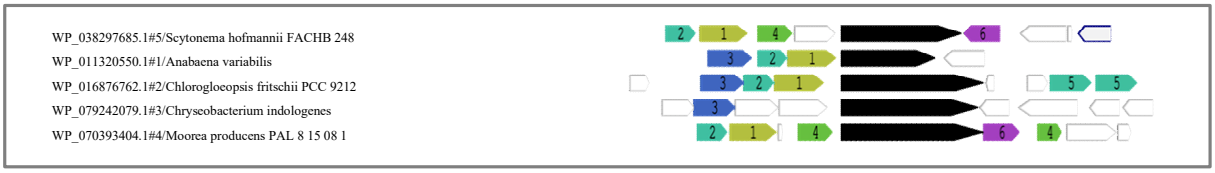
The amino acid sequence of the NRPS-like enzyme Ava_3855 (see Appendix), involved in the generation of the MAA shinorine in the cyanobacterium *Anabaena variabilis* was used to retrieve NRPS-like enzyme homologues related to the production of MAAs from cyanobacteria. A BLAST search of the amino acid

sequence Ava_3855 was run against RefSeq proteins on NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) and several NRPS-like enzymes of interest were selected. The accession numbers of the selected proteins were pasted into the WebFlaGs website (http://www.webflags.se/)[143,144] generating the output as shown below.

Input:



Output:



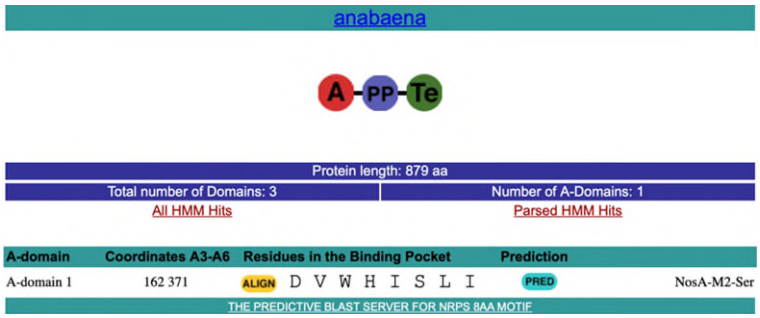
Number	Enzyme
1	ATP-grasp enzyme
2	Class I SAM-dependent methyltransferase
3	Sedoheptulose 7-phosphate cyclase/ 3-dehydroquinate synthase
4	DMT family transporter
5	Glutathione-dependent formaldehyde dehydrogenase
6	Methyltransferase domain-containing protein

Figure 4.2: Genome mining of several NRPS-like genes using the amino acid sequence of Ava_3855 as template.

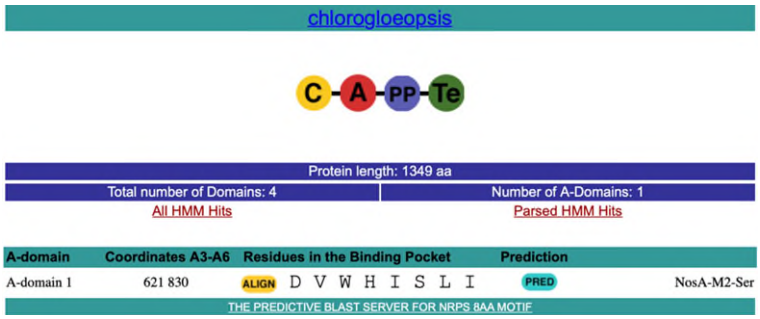
The five selected NRPS-like amino acid sequences were used and input in the PKS/NRPS Analysis Website (http://nrps.igs.umaryland.edu/) to confirm the A-domain specificity of each one of them. Below is the output of each of the amino acid sequences corresponding to its specific substrate. The A-domain specificity can be predicted based on specific amino acid residues, normally eight to ten residues (lining the binding pocket), which are the key aspect in the selection of the cognate substrate activating it as its aminoacyl adenylate using ATP. In simple terms, it can be said that signature sequences and the substrate specificities are connected thus leading to

specific codes for specific amino acids activation in non-ribosomal peptide synthetases[102].

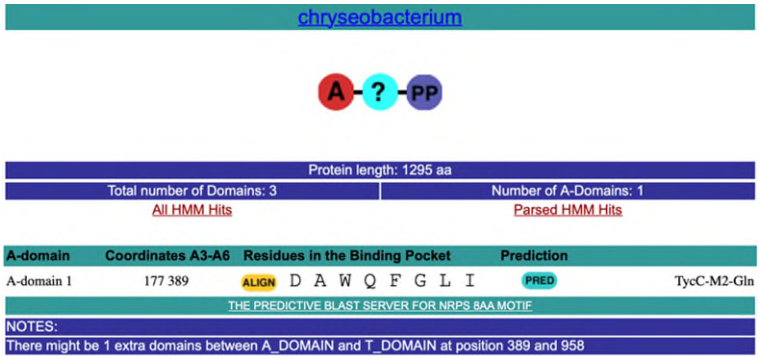
Species	NRPS-like Enzyme	A-domain specificity
Anabaena variabilis	Ava_3855	Serine



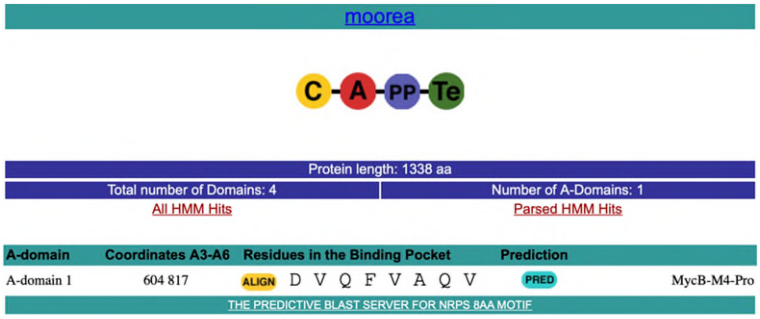
Species	NRPS-like Enzyme	A-domain specificity
Chlorogloeopsis fritschii	UYC_RS0133560	Serine



Species	NRPS-like Enzyme	A-domain specificity
Chryseobacterium indologenes	BUE84_RS13530	Glycine



Species	NRPS-like Enzyme	A-domain specificity
Moorea producens	BJP34_RS17220	Proline



Species	NRPS-like Enzyme	A-domain specificity
Scytonema hofmannii	TOL9009_RS40645	No hit

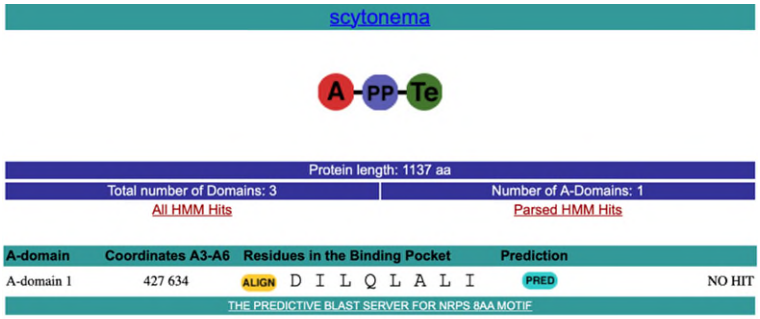


Figure 4.3: Output from PKS/NRPS Analysis website for each NRPS-like amino acid sequence query depicting their corresponding A-domain specificity.

Homologues of the NRPS-like enzyme Ava_3855 were found in various cyanobacteria (both terrestrial and marine). Based on the A-domain prediction (especially after using the PKS/NRPS Analysis tool), it was noted that some of the homologues had different substrate affinity showing that those MAAs might be novel, especially the one predicted to activate L-proline as substrate as no MAAs with proline moiety has been discovered so far.

Five *E. coli* codon optimized synthetic NRPS-like genes (See Appendix) were cloned into pUC57-Kan and delivered as plasmid DNA sample as well as glycerol stock from Genewiz. Extra amino acids (MDINTETE) were added in front of each NRPS-like sequence to have the same start codons for all of them as well as it is known that the selected start codons enhance the translation rate, thus boosting up protein production. Furthermore, based on a study carried out by Nissley *et al*., the nascent proteome behaviour can be governed by codon translation. It is to be noted that there exists a series of cotranslational processes including but not limited to chaperone binding, glycosylation or cotranslational folding[145]. Therefore, carefully choosing the start codons for a particular protein of interest is important to avoid or delay proper translation rate. In this case, the starting codon do not really affect the rate of translation since translation already starts at the beginning of the His-tag (due to the fact that cloning was done in pET151) which is several codons before the N-terminus of the start of the protein of interest.

4.3 Recombinant plasmid generation using Directional TOPO Cloning

The strains containing the corresponding *E. coli* codon optimized NRPS-like genes from *Anabaena variabilis*, *Chlorogloeopsis fritschii*, *Chryseobacterium indologenes*, *Moorea producens* and *Scytonema hofmanni* cloned in pUC57-Kan plasmids were streaked followed by plasmid DNA isolation as described in Chapter 2. PCR was used to amplify the DNA sequences using the forward primer starting with CACC in order to allow Directional TOPO cloning. This type of cloning makes use of the enzyme topoisomerase I from *Vaccinia* virus which binds after 5'-CCCTT in one strand of the vector and cleaves the phosphodiester backbone forming a covalent bond between the 3' phosphate of the cleaved strand and a tyrosyl residue (Tyr-274) of topoisomerase I.

The reverse primers were designed to integrate a stop codon at the end of each NRPS-like gene sequence to halt translation (Each primer is shown in Chapter 2).

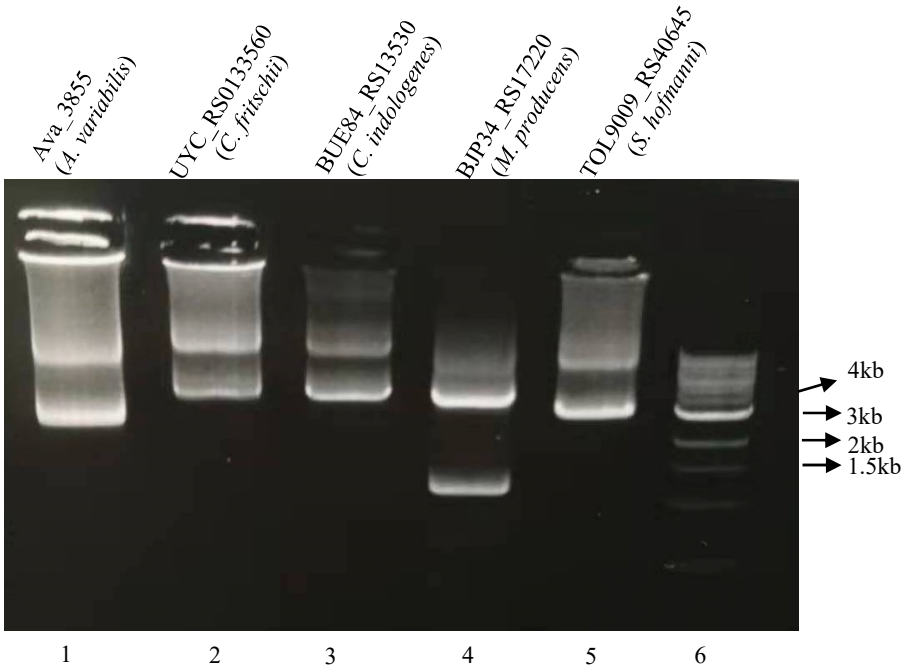


Figure 4.4: PCR products from pUC57 with the expected size analysed on 1% agarose gel and visualized under UV with lane 6 being the NEB 1 kb DNA ladder.

Table 4.1: Summary of PCR results

PCR product from pUC57 encoding each NRPS-like gene from	**Expected band size on gel**	**Lane**
Anabaena variabilis	2644 bp	1
Chlorogloeopsis fritschii	4054 bp	2
Chryseobacterium indologenes	3892 bp	3
Moorea producens	4021 bp	4
Scytonema hofmanni	3418 bp	5

After obtaining the PCR products, the latter were purified using the Monarch® DNA Gel Extraction Kit and prepared for Directional TOPO cloning.

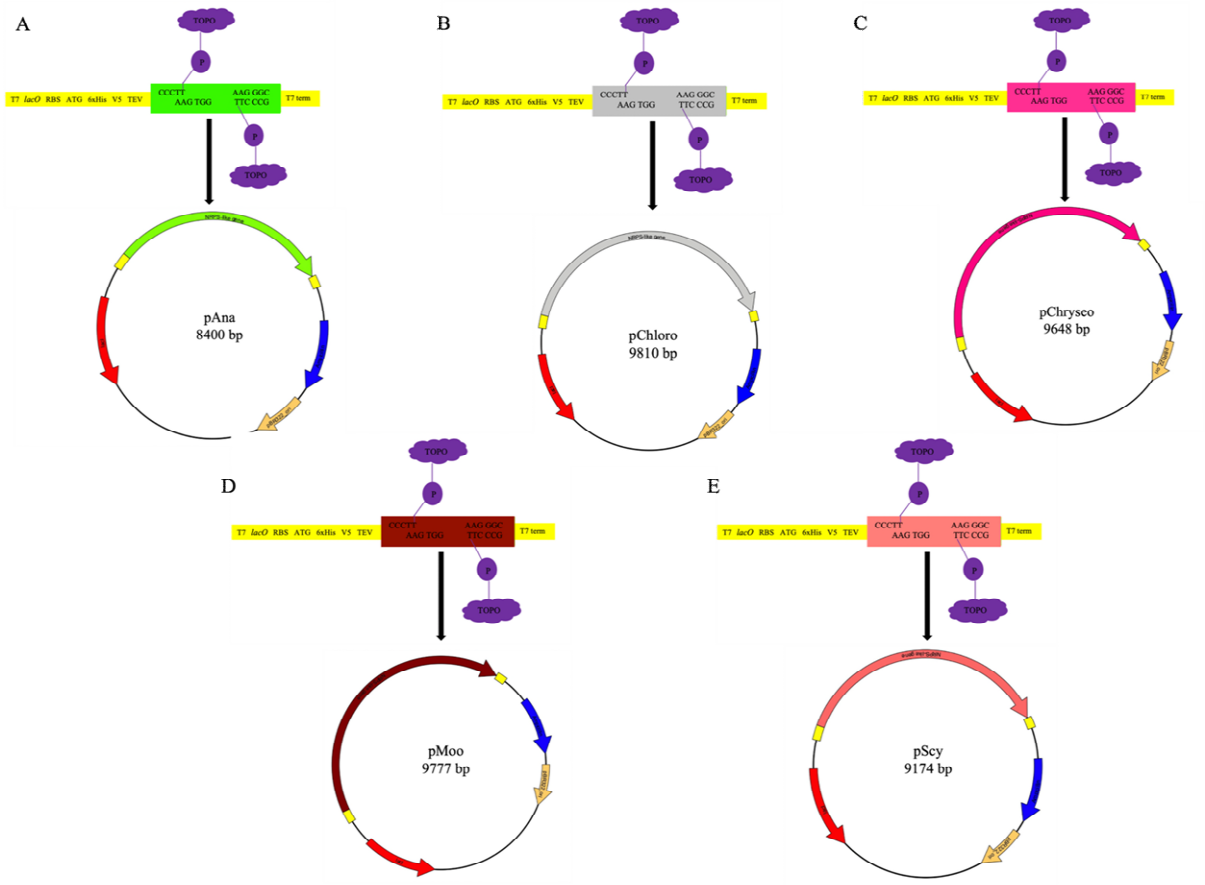


Figure 4.5: Directional TOPO cloning of each NRPS-like gene. Recombinant plasmid map of: A: *ava_3855* in pET151 to generate pAna; B: *uyc_rs0133560* in pET151 to generate pChloro; C: *bue84_rs13530* to generate pChryseo; D: *bjp34_rs17220* to generate pMoo; E: *tol9009_rs40645* to generate pScy.

The plasmid pET151 contains a T7 promoter, lacO which allow induction by IPTG. Six Histidine tags followed by V5 epitope tag, and a TEV protease recognition site are found after the start codon methionine where the His tags are the primary reason for choosing this vector based on its ability to bind to nickel beads during purification. The plasmid also contains an ampicillin resistance marker for selection as well as pBR322 as origin of replication with a copy number of around 20.

After the directional TOPO cloning of the NRPS-like DNA sequences in pET151, the constructs were checked by PCR with designed primers as shown with the expected band size obtained for each based on the agarose gel as shown below with the GeneRuler 1 kb Plus ladder being used in the last lane.

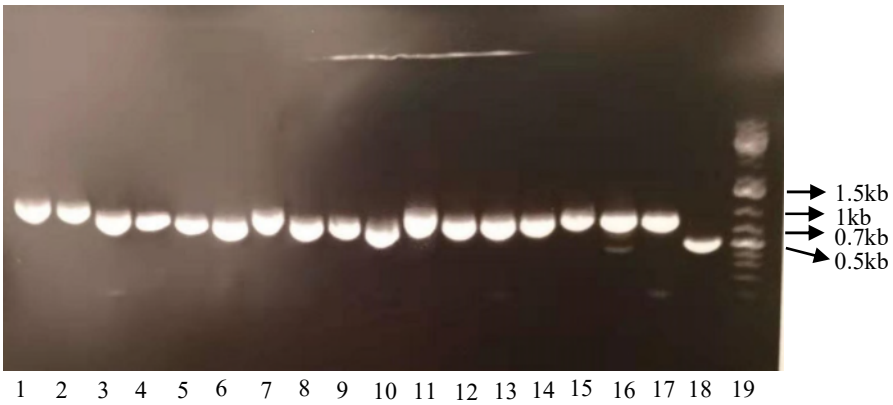


Figure 4.6: PCR check of each construct using corresponding primers as described in the table below. Analysis was done on 1% agarose gel and visualized under UV with lane 19 being the Generuler 1 kb Plus DNA ladder.

Table 4.2: Summary of construct with the corresponding primers.

Construct	**Primers**	**Expected band size**	**Lane on agarose gel**
pAna	T7 to Ana 1322 REV	1429 bp	1
	Ana 1342 FWD to T7 Reverse	1403 bp	2
pChloro	T7 to Chloro 1007 REV	1114 bp	3
	Chloro 1027 to Chloro 2064 REV	1081 bp	4
	Chloro 2087 to Chloro 3101 REV	1061 bp	5
	Chloro 3124 to T7 Reverse	1035 bp	6
pChryseo	T7 to Chry 1000 REV	1110 bp	7
	Chry 1023 to Chry 2011 REV	1039 bp	8
	Chry 2038 to Chry 3025 REV	1034 bp	9
	Chry 3044 to T7 Reverse	949 bp	10
pMoo	T7 to Moo 1004 REV	1114 bp	11

	Moo 1027 to Moo 2023 REV	1042 bp	12
	Moo 2045 to Moo 3046 REV	1045 bp	13
	Moo 3067 to T7 Reverse	1057 bp	14
pScy	T7 to Scy 894 REV	1004 bp	15
	Scy 917 to Scy 1896 REV	1026 bp	16
	Scy 1919 to Scy 2878 REV	1004 bp	17
	Scy 2899 to T7 Reverse	622 bp	18

A restriction digest using PstI as restriction enzymes was chosen as the latter was seen to cut once in the vector and several times to none on each recombinant plasmid based on a virtual digest carried out on Benchling. The gel below shows the results of the restriction digest when carried out on the various constructs with the GeneRuler 1 kb Plus DNA ladder being used in the first and last lane.

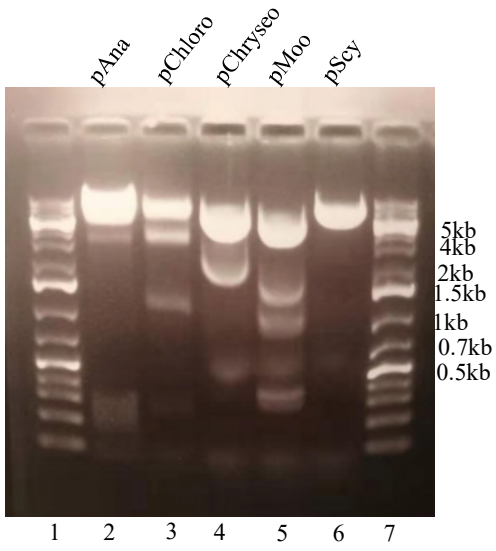


Figure 4.7: Restriction digest of each construct using PstI. Analysis was done on 1% agarose gel and visualized under UV with lanes 1 and 7 being the Generuler 1 kb Plus DNA ladder.

Table 4.3: Summary of bands size after PstI digestion.

Construct	**Size of digested band (bp)**	**Lane**
pAna	5279; 2891; 183; 47	2

pChloro	5702; 2725; 1125; 258	3
pChryseo	7133; 2515	4
pMoo	5828; 1891; 1281; 438; 339	5
pScy	9174	6

Based on the band size of each PCR product from each construct as well as the restriction digest performed using PstI, it could be determined that the directional TOPO cloning worked. Moreover, in order to confirm whether the construct does not contain any point mutation, the recombinant plasmid was sent with corresponding sequencing primers (See chapter 2) for Sanger sequencing.

The gene sequences of recombinant plasmids pAna; pChloro; pChryseo and pScy were found to be correct after sequencing. However, the NRPS-like gene sequence from *Moorea producens* was found to have a point mutation, in terms of missing a base 'G' after the recombinant plasmid was aligned to the sequencing data on Benchling as shown below.

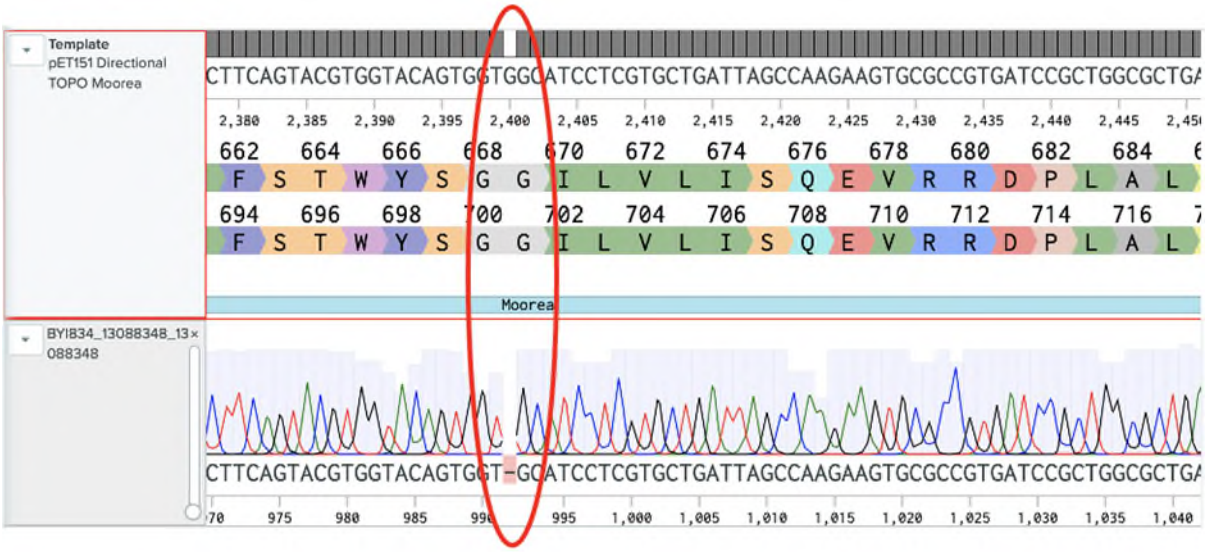


Figure: 4.8: Missing "G" in the NRPS-like gene sequence from *Moorea producens* in pET151

The solution opted to remediate to this problem was site-directed mutagenesis as described in the next section.

4.4 Site-directed mutagenesis of NRPS-like gene from *Moorea producens*

The Q5® Site-Directed Mutagenesis Kit was used to insert a base 'G' in the NRPS-like genes in pUC57 due to its smaller size [~ 6 kb] compared to pMoo [~ 9 kb] which would increase the chance of the mutation/ base addition to be successful. The nucleotide (G) was integrated into the 5' end of the forward primer while the reverse primer was designed to anneal back-to-back to the corresponding forward primer.

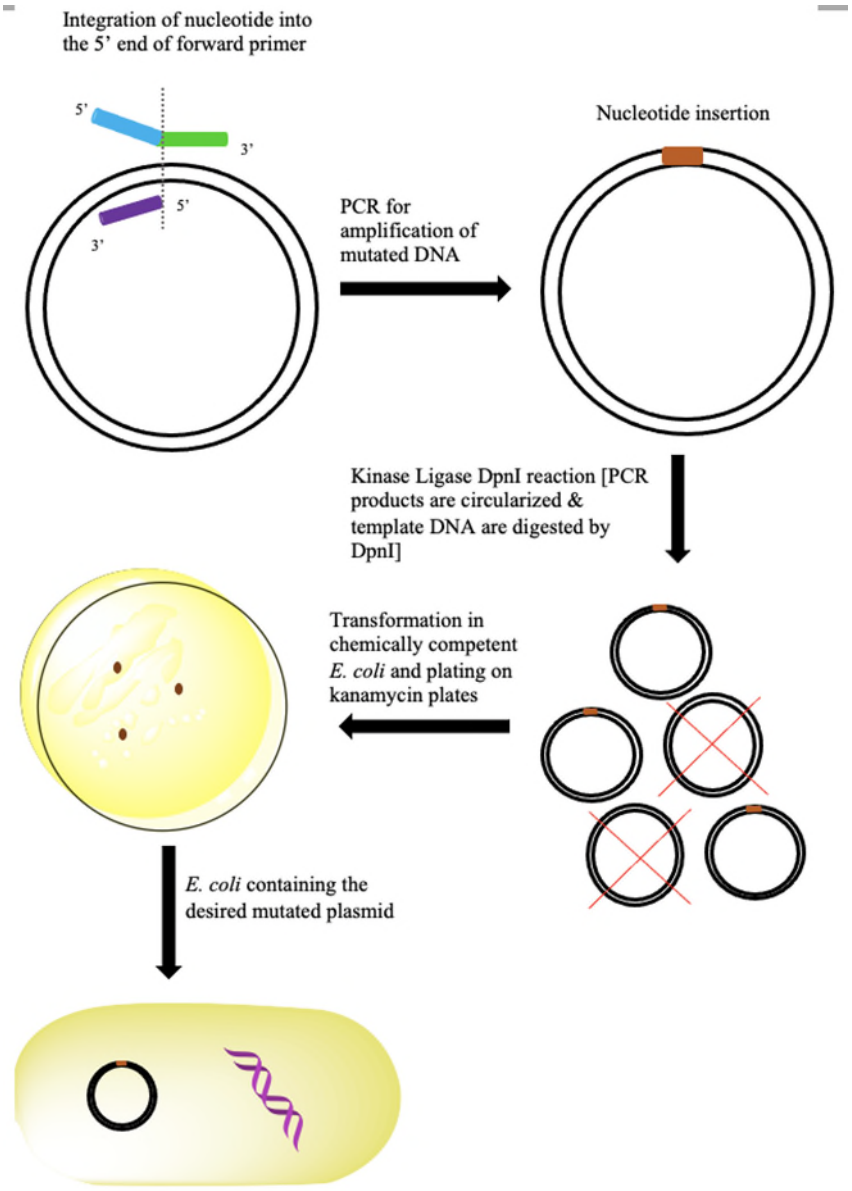


Figure 4.9: Schematic diagram of site-directed mutagenesis.

The mutated plasmid containing the NRPS-like gene from *Moorea producens* was sent for sequencing and the nucleotide 'G' was seen to be successfully inserted by the site-directed mutagenesis technique.

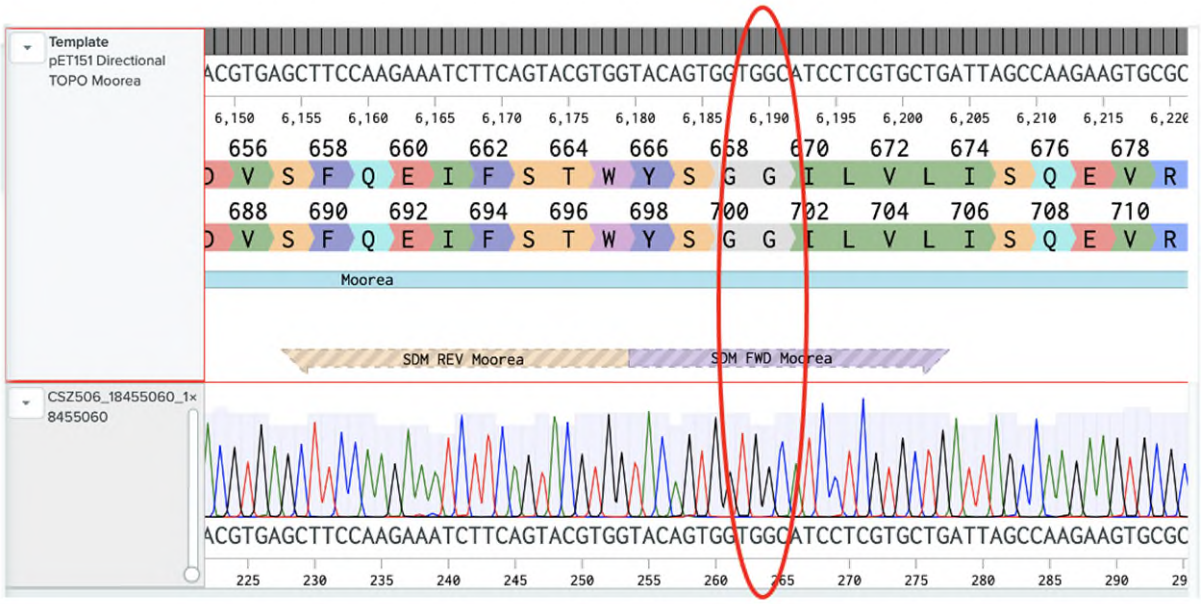


Figure 4.10: The corrected NRPS-like gene sequence from *Moorea producens*

All plasmids including the corrected one (shown above) were then transformed in *E. coli* protein expression strains namely *E. coli* BL21 and *E. coli* BAP1 for future expression experiment. Both strains contain a chromosomal copy of the phage T7 RNA polymerase gene which upon addition of the inducer IPTG, expression followed by transcription and translation of the protein of interest occur.

4.5 Expression and purification of Ava_3855 using HisPur™ Ni-NTA Resin

The *E. coli* BL21 and *E. coli* BAP1 strains each containing the *ava_3855* gene successfully cloned in pET151 were cultured in LB supplemented with 100 µg/ml ampicillin, induced with 0.5 mM IPTG at OD_{600} = 0.5 and incubated overnight at 15°C as described in materials and methods section in chapter 2.

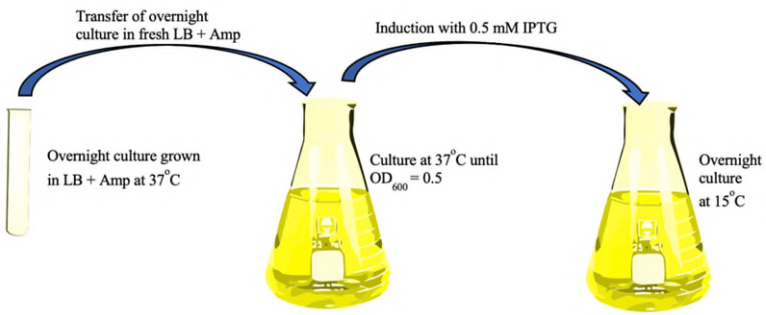


Figure 4.11: Schematic diagram for culture and expression of NRPS-like enzyme.

The following day, the lysis buffer A supplemented with lyzozyme and protease cocktail inhibitor as detailed in materials and methods were used for the expression construct, followed by sonication and centrifugation. The NRPS-like enzyme Ava_3855 was successfully expressed and purified using HisPur™ Ni-NTA Resin as explained in Chapter 2 with step wise imidazole elution. The process is shown below.

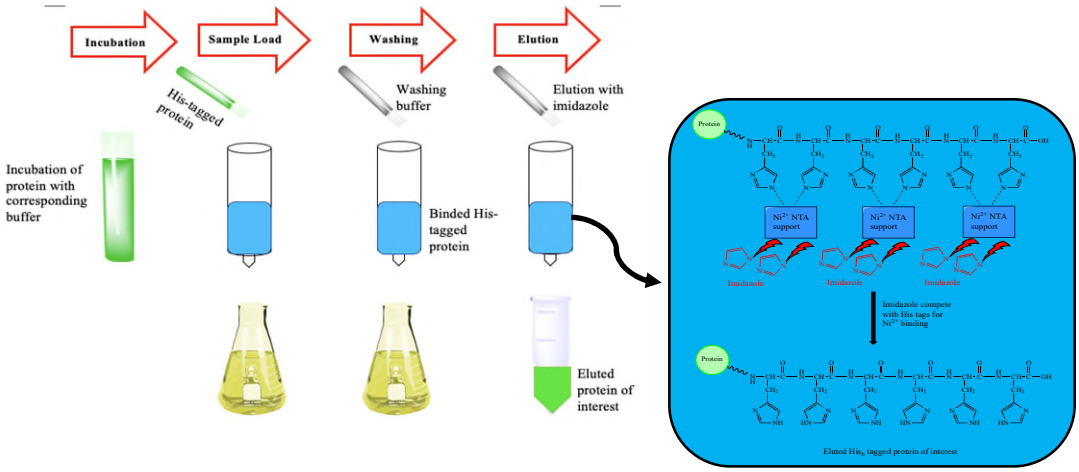


Figure 4.12: Schematic diagram of His-tagged protein purification

The SDS-PAGE gel result of the successfully purified NRPS-like enzyme Ava_3855 is shown below (Figure 4.13). The expected protein size (103.18 kDa) was seen and the latter was eluted with 150 – 200 mM imidazole.

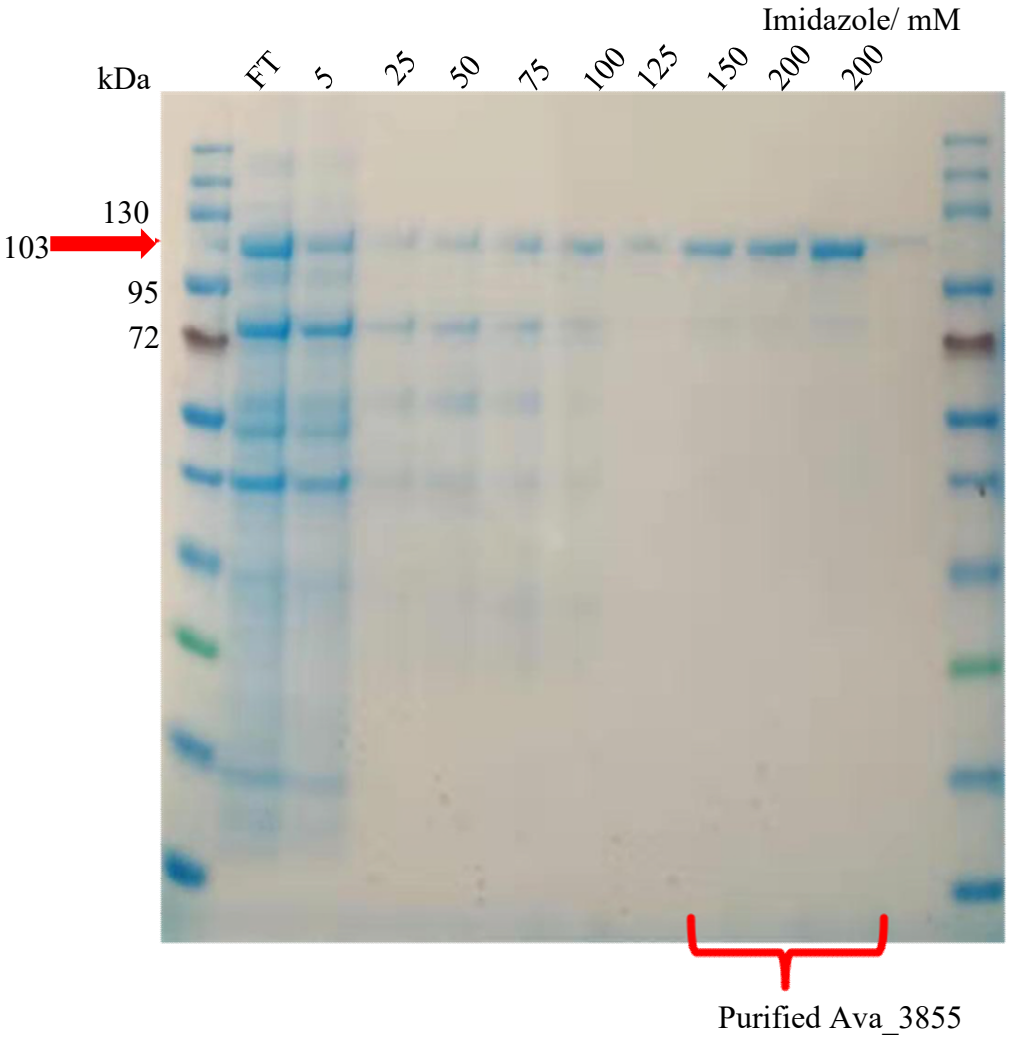


Figure 4.13: SDS-PAGE analysis of purified Ava_3855. The strain used is *E. coli* BL21.

The fractions containing the purified protein of interest were combined and concentrated using an Amicon Ultra 50K molecular weight cutoff spin filter. The NRPS-like enzyme from *Anabaena variabilis* was dialyzed thrice against 2 L of storage buffer E and the protein was aliquoted and stored at -80ºC for further experiment.

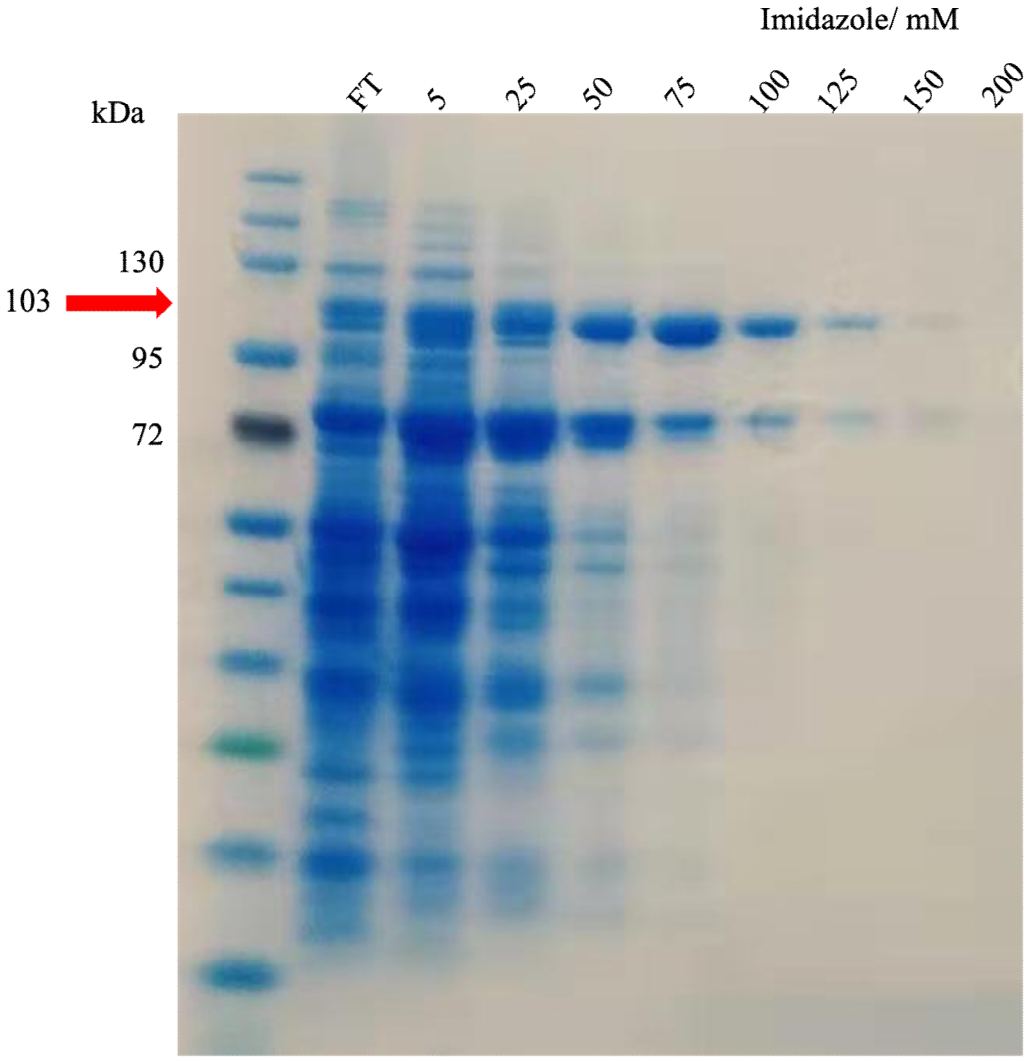


Figure 4.14: SDS-PAGE analysis of purified Ava_3855. The strain used is *E. coli* BAP1.

Both gels were compared, and it was clear that the expression of Ava_3855 was clearer and purer in *E. coli* BL21. The yield of protein expressed, however, seemed better in *E. coli* BAP1 (Figure 4.14), but due to the band at about 72 kDa, it was not convenient to use the latter as a host for future experiment. The band at 103 kDa as well as the one at 72 kDa was sent for proteomics to confirm the protein of interest and to also figure out what is the band at 72 kDa. Based on the proteomics results, the protein of interest was confirmed while the band at around 72 kDa was confirmed as being a protein directly produced by the *E. coli* BAP1 host namely propionate--CoA ligase (69 kDa) encoding the *prpE* gene. The difference between *E. coli* BAP1 and *E. coli* BL21 is the deletion of the *prpB*, *prpC* and *prpD* genes from *E. coli* BL21 and

integration of a T7 promoter in front of a *sfp* gene as well as a T7 promoter in front of the *prpE* gene, thus giving rise to *E. coli* BAP1[146] as shown in Figure 4.15.

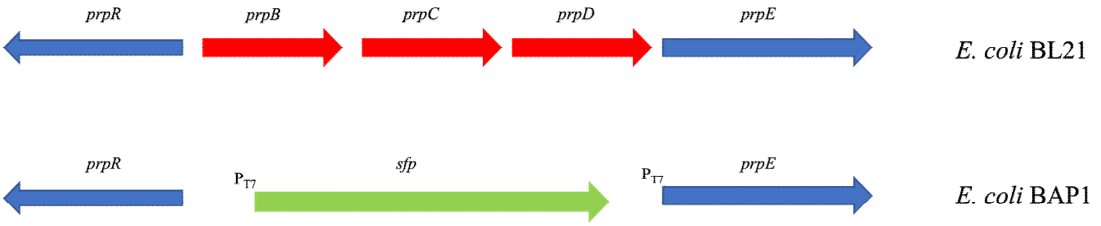


Figure 4.15: *E. coli* BL21 v/s *E. coli* BAP1. Modifications include the deletion of the *prpB*, *prpC* and *prpD* genes to integrate the *sfp* gene.

The reason to initially consider the use of *E. coli* BAP1 as host was the presence of the *sfp* gene which code for a phosphopantetheinyl transferase after induction which was dismissed to avoid interference with propionate--CoA ligase which seemed to be produced in quite high yield during protein purification, which might be competing with the protein of interest for the Nickel beads, thus lowering the yield of the enzyme of interest. Henceforth, *E. coli* BL21 was the host used for the expression of Ava_3855.

Protein 4'-phosphopantetheinylation

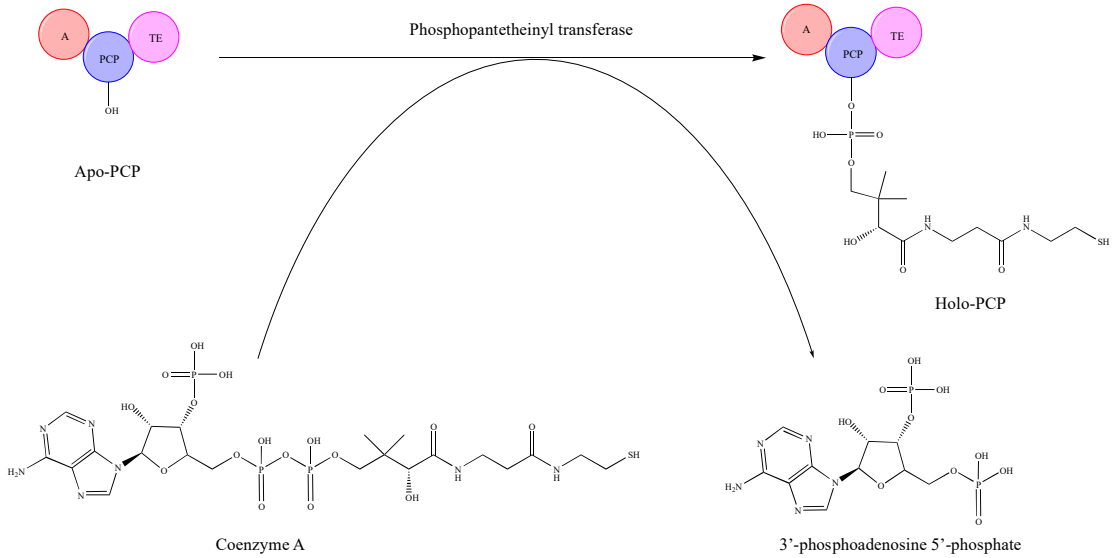


Figure 4.16: Protein 4'-phosphopantetheinylation. The conversion of the protein from "*apo*" to "*holo*" form by the addition of the phosphopantetheinyl arm prosthetic group.

4.6 Intact protein mass spectrometry of Ava_3855

The reverse phase C4 column was used for HPLC-MS of the intact protein with a linear gradient elution of 5-100% acetonitrile containing 0.1% formic acid for 30 mins. The stationary phase of this type of column has typically an ultra-inert silica surface with a butyl ligand (low hydrophobicity). The shorter alkyl chain such as that in the C4 column are appropriate for hydrophobic proteins due to the quick kinetic desorption between the stationary phase and the proteins. Thus, the hydrophobicity of the protein is the main parameter to consider during liquid chromatography, the more non-polar the protein, the shorter the retention time[147].

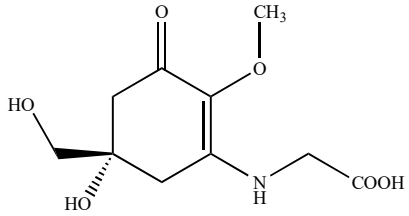
Deconvolution, being a prevailing tool during MS spectra analysis, was utilised for the de-isotoping and de-charging of detected ions allowing the resolution of overlapping chromatograms. The soft technique namely electrospray ionisation (ESI) was used for intact protein mass spectrometry which has the ability to generate gas-phase ions from high molecular weight molecules, thus deconvolution is a must to convert the peaks in the series of diverse charge states into a single peak on the molecular mass scale[148].

The purified Ava_3855 enzyme from *E. coli* BL21 was concentrated to a final concentration of 270 µM followed by the addition of 10 mM $MgCl_2$, 0.1 mM CoA and 500 µM Sfp to make up to a final volume of 50 µL. The reaction mix was incubated for one hour at room temperature to allow the conversion of the "apo" form of the enzyme to the phosphopantetheylated "holo" form of Ava-3855. The addition of 340 Da confirmed the phosphorylation of the PCP domain. 5 mM ATP along with 1 mM L-serine were added to the reaction mix, followed by incubation for one hour at room temperature. The addition of 87 Da showed the loading of serine on the PCP domain, which confirmed the functionality of Ava_3855 to be able to activate the amino acid L-serine (Figure 4.18).

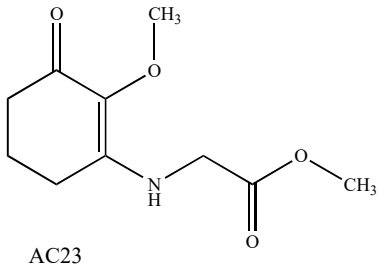
The natural precursor to be used for the formation of shinorine is mycosporine-glycine. However, due to the very low yield of mycosporine-glycine produced from the *E. coli* BL21 host containing the plasmid pET28b *ava_3858* to *ava_3856*, a different option was chosen.

The synthetic precursor AC23 (Synthesized by Adam Cowden, PhD student under the supervision of Martin Wills from Chemistry department, Warwick University) was

used instead. A final concentration of 2 mM of AC23 was added and incubated for another hour followed by mass spectrometry.



Mycosporine-glycine



AC23

Figure 4.17: Chemical structure of the natural precursor mycosporine-glycine and the synthetic molecule AC23.

The loss of 87 Da displayed in the Figure below revealed that L-serine was being unloaded from the enzyme, most probably by AC23 acting as a strong nucleophile.

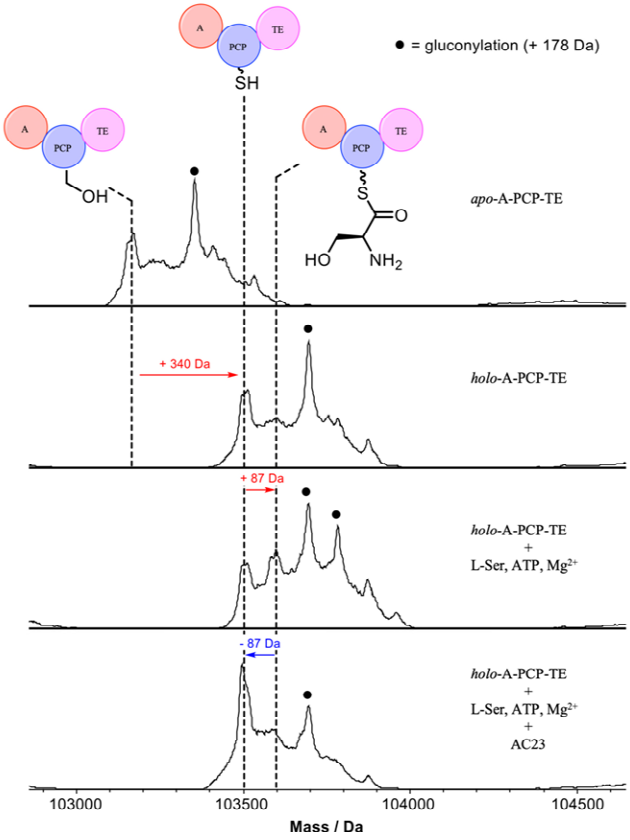


Figure 4.18: Deconvoluted intact protein mass spectra of Ava_3855. Mass shifts corresponding to conversion of "*apo*" to "*holo*" form and loading of L-serine on the PCP domain as well as unloading of L-serine from the PCP domain. Peaks with dots indicate N-terminal gluconoylation.

Gluconoylation is a known post-translational modification of recombinant heterologous proteins in *E. coli*. Proteins containing N-terminal His tags were seen to give rise to products with mass 258 and 178 Da greater than the predicted value. The mass of +178 Da is usually attributed to dephosphorylation of the acylation with a 6-phosphohexonic acid (+258 Da). An intracellular acylating agent namely 6-phosphoglucono-1,5-lactone (+258 Da) is normally formed by the enzyme glucose-6-phosphate dehydrogenase acting on glucose 6-phosphate during synthesis of nucleic acids. The eradication of the phospho group by the cell phosphatase, thus reduce the mass from +258 Da to +178 Da. The modification depend on various factors including the concentration of 6-phosphoglucono-1,5-lactone in the cell, the expression level of the recombinant protein as well as the period the His tag is exposed to an environment

containing the lactone. In simpler terms, this post-translational modification happened due to the lactone being reactive to the N-terminal of the His tag protein which was confirmed by Geoghegan *et al*., based on NMR analysis[149].

4.6.1 Proposed Reaction based on Intact Protein Mass Spectrometry

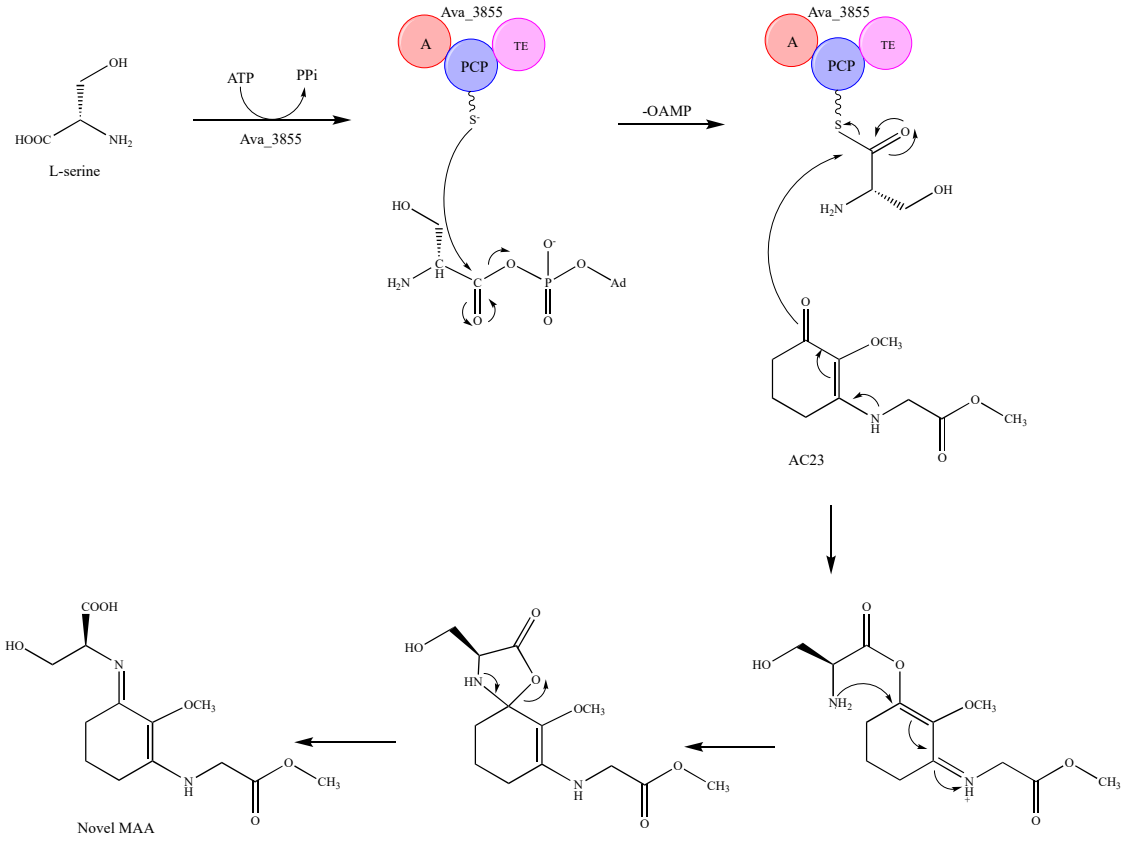


Figure 4.19: Proposed mechanism based on intact mass spectrometry results. AC23 might be acting as a strong nucleophile, thus releasing the activated L-serine substrate from the enzyme Ava_3855.

Based on the intact protein mass spectrometry results, a plausible mechanism was proposed. After the conversion of the "*apo*" to "*holo*" form by the addition of the phosphopantetheinyl arm prosthetic group, ATP was used to load the L-serine substrate on the PCP by a phosphodiester bond with the release of OAMP. The synthetic precursor, here referring to AC23, might be acting as a strong nucleophile, thus promoting the release of the L-serine from the enzyme, leading to the formation

of an enol ester intermediate, followed by rearrangement of O to N forming a cyclohexeneimine core, thus giving rise to a novel type of MAA.

4.7 Chemo-enzymatic assay using Ava_3855 and AC23

A chemo-enzymatic assay was set up in order to see whether any product could be detected using the synthetic precursor AC23. A final volume of 100 µL solution containing 5 mM $MgCl_2$, 0.1 mM CoA, 500 µM Sfp, 1 mM L- serine, 2 mM AC23, 2 µM Ava_3855 and 5 mM ATP were used for the assay which was carried out at room temperature overnight. The UHPLC-HRMS results are displayed below.

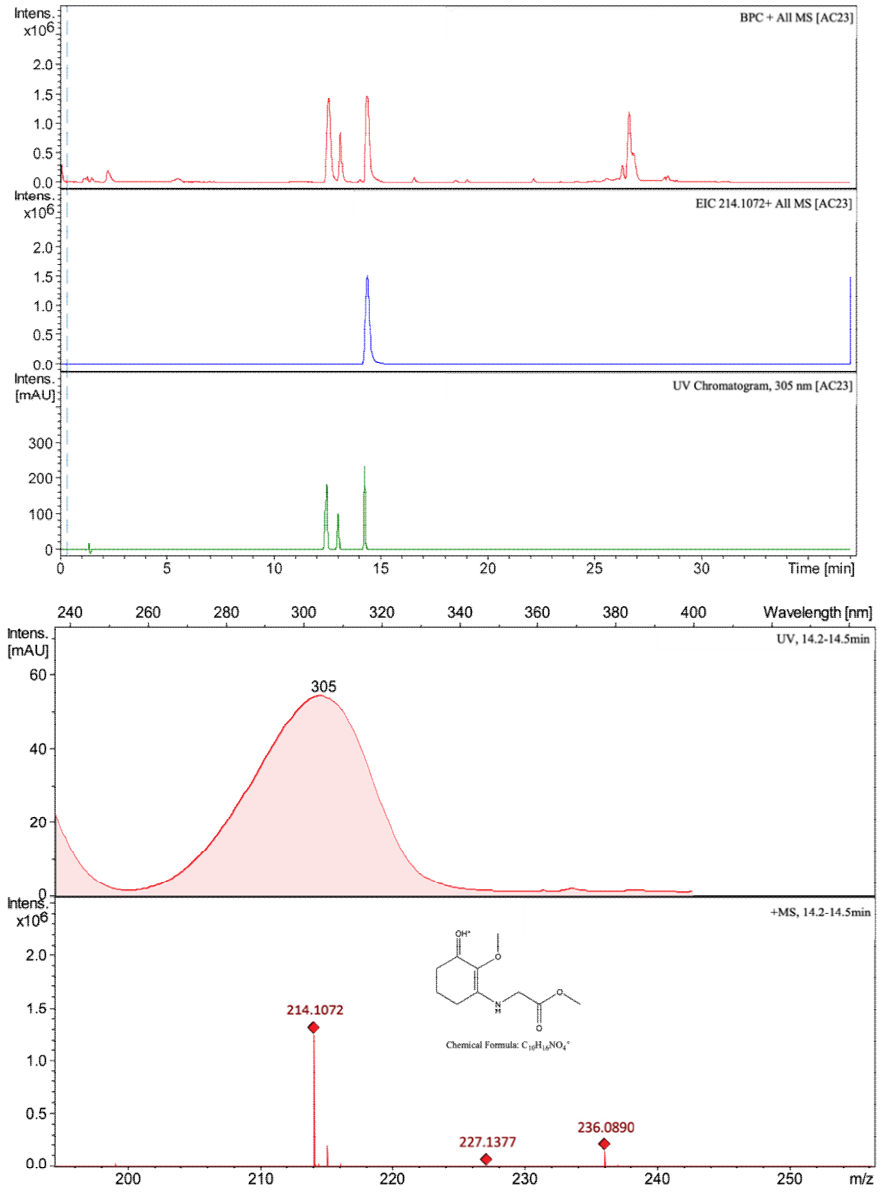


Figure 4.20: The chromatogram of AC23. Top: Base peak chromatogram (BPC) (Red); Extracted ion chromatogram (EIC) for *m/z* 214.1072 (Blue); UV chromatogram at 305 nm (Green). Bottom: UV and mass spectra of AC23 after UHPLC-HRMS analysis.

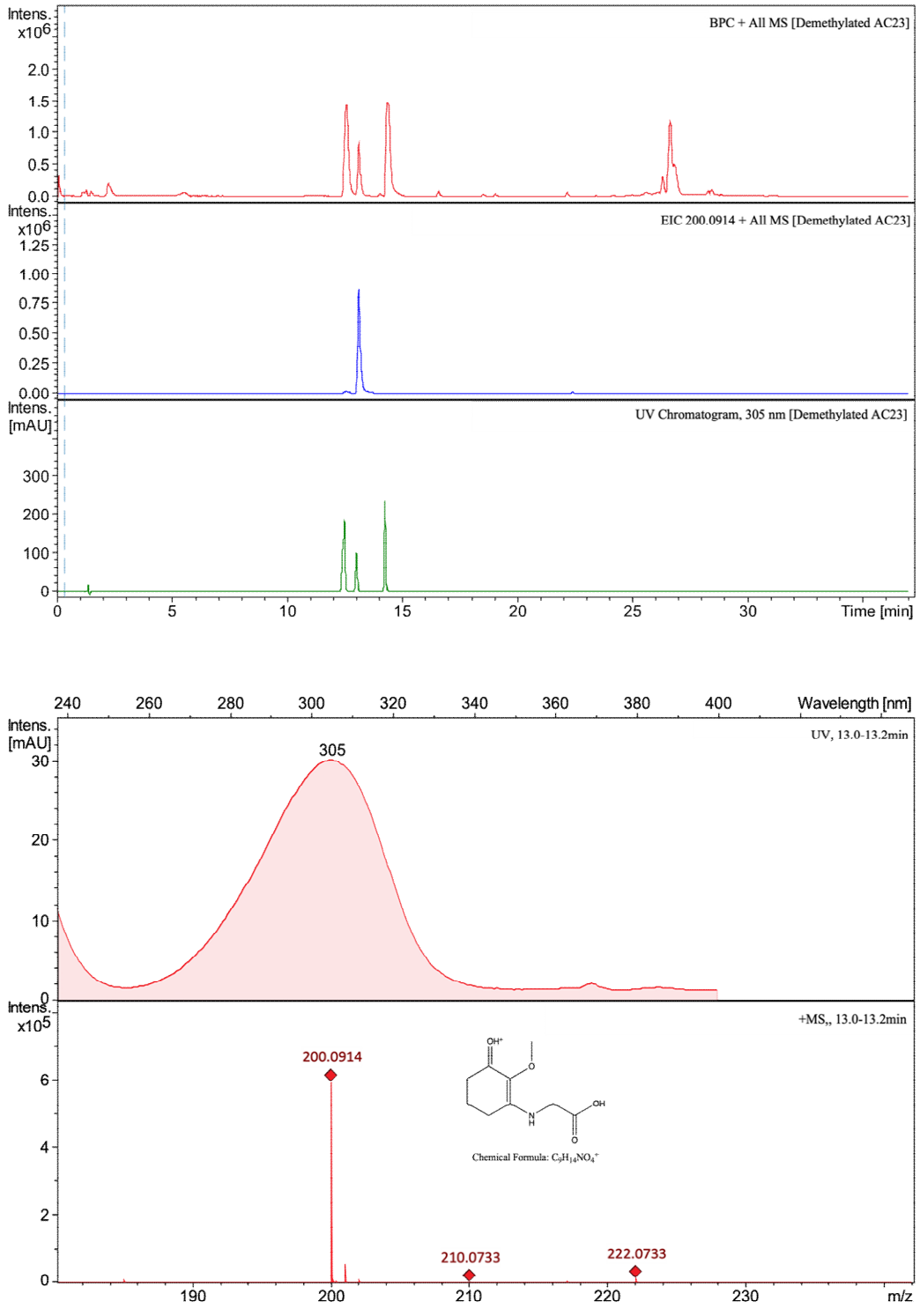


Figure 4.21: The chromatogram of demethylated AC23. Top: Base peak chromatogram (BPC) (Red); Extracted ion chromatogram (EIC) for *m/z* 200.0914 (Blue); UV chromatogram at 305 nm (Green). Bottom: UV and mass spectra of demethylated AC23 after UHPLC-HRMS analysis.

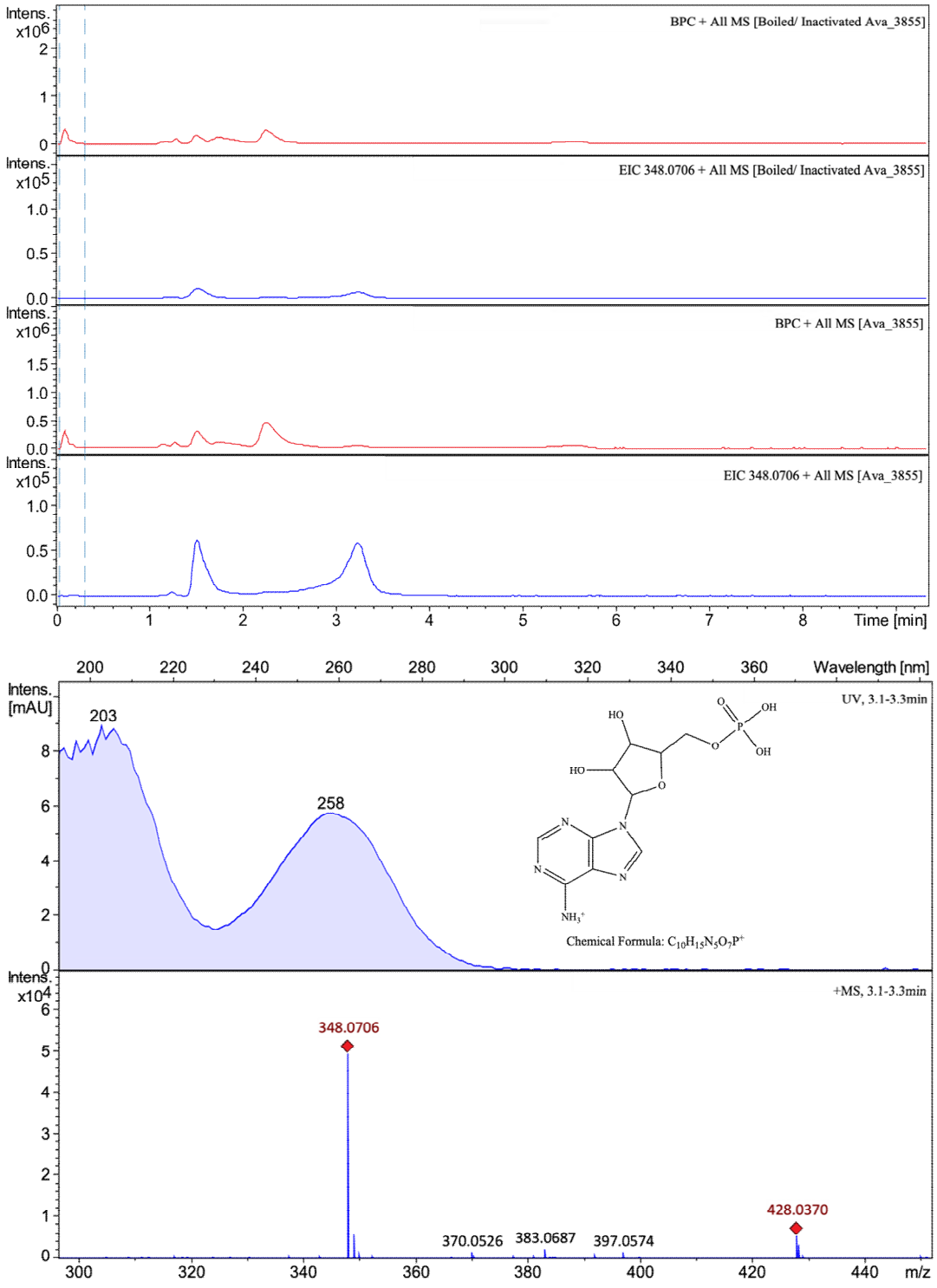


Figure 4.22: The chromatogram monitoring AMP production with boiled/inactivated Ava_3855 and with Ava_3855. Top: Base peak chromatogram (BPC) (Red); Extracted ion chromatogram (EIC) for *m/z* 348.0706 (Blue). Bottom: UV chromatogram at 258 nm and mass spectra of AMP after UHPLC-HRMS analysis.

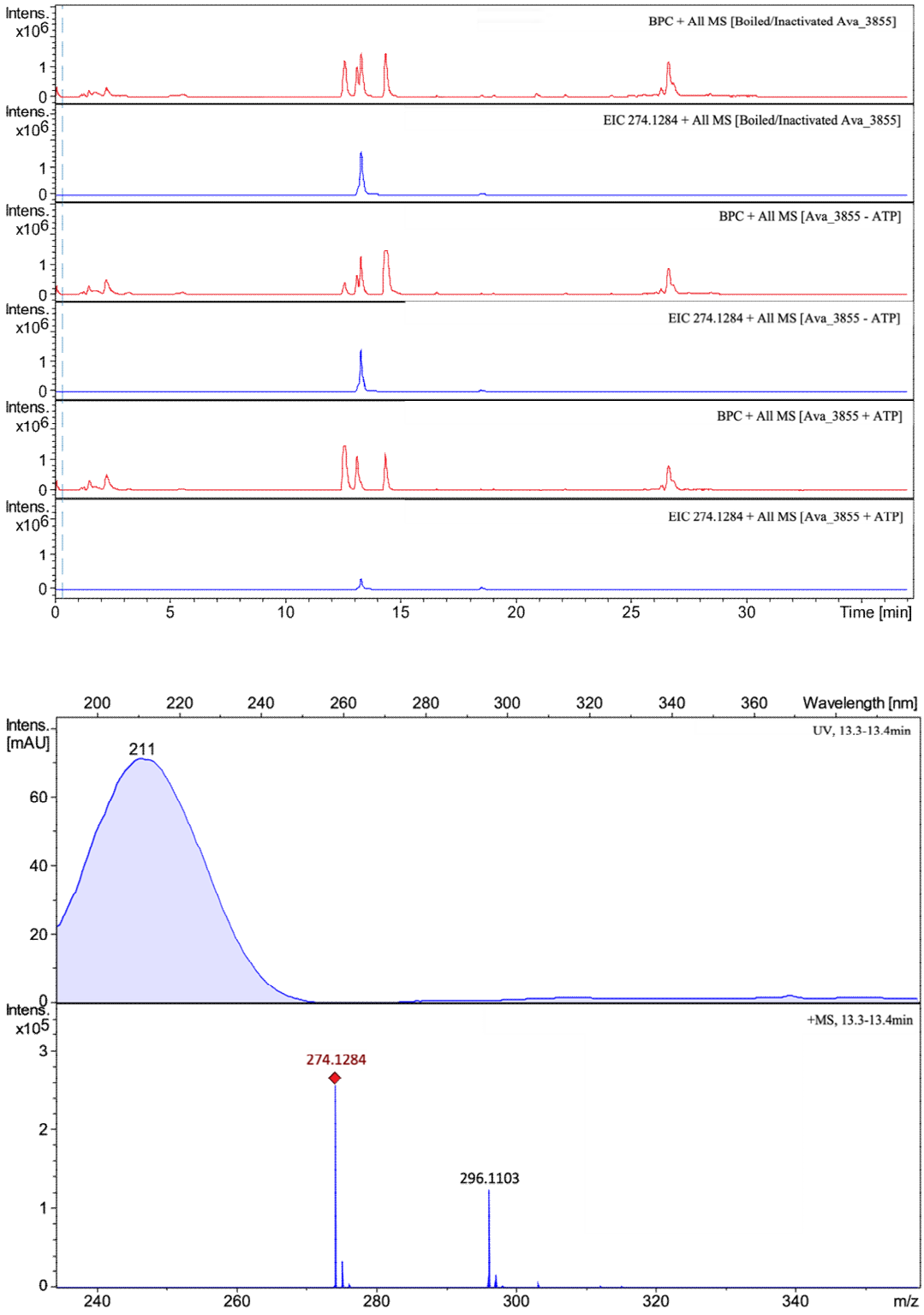


Figure 4.23: The chromatogram of boiled/inactivated Ava_3855; Ava_3855 without ATP and Ava_3855 with ATP. Top: Base peak chromatogram (BPC) (Red); Extracted ion chromatogram (EIC) for *m/z* 274.1284 (Blue). Bottom: UV chromatogram at 211 nm and mass spectra of novel compound after UHPLC-HRMS analysis.

Among the 5 expression constructs, only Ava_3855 could be purified. The other four NRPS-like proteins could not be expressed under a series of conditions as described in Chapter 2. The SDS-PAGE gel of the other four NRPS-like clearly showed no protein expression.

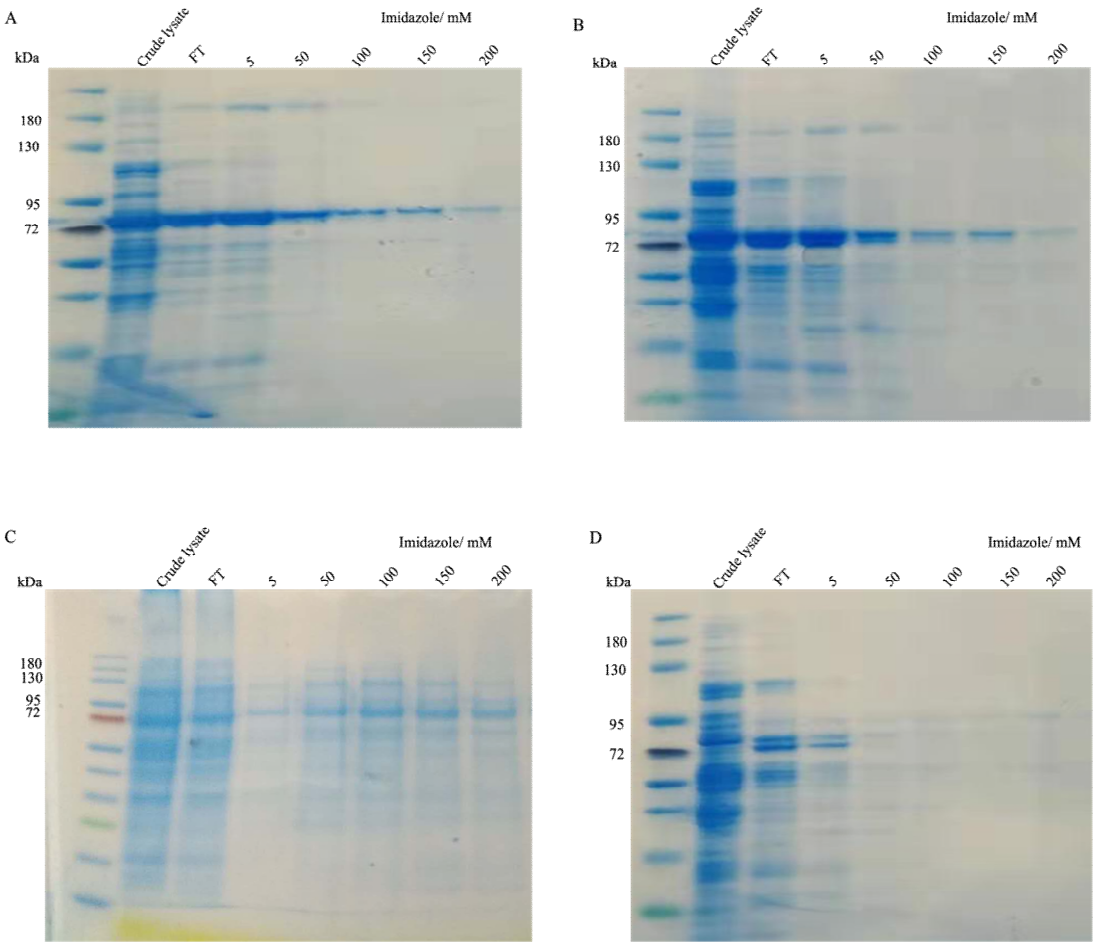


Figure 4.24: SDS-PAGE gel of A: UYC_RS0133560; B: BUE84_RS13530; C: BJP34_RS17220; D: TOL9009_RS40645 after expression and purification from the host *E. coli* BAP1 with Lane 1 being the NEB colour prestained protein standard, broad range (10-250 kDa).

After failure to express the other four enzymes, different parameters were taken into consideration to remediate to this problem. Diverse IPTG concentrations during induction, lysis buffers as well as pH were used to favour expression (Chapter 2). The

SDS-PAGE gels below show 0.2 mM IPTG used during induction and lysis buffer at pH 8.5.

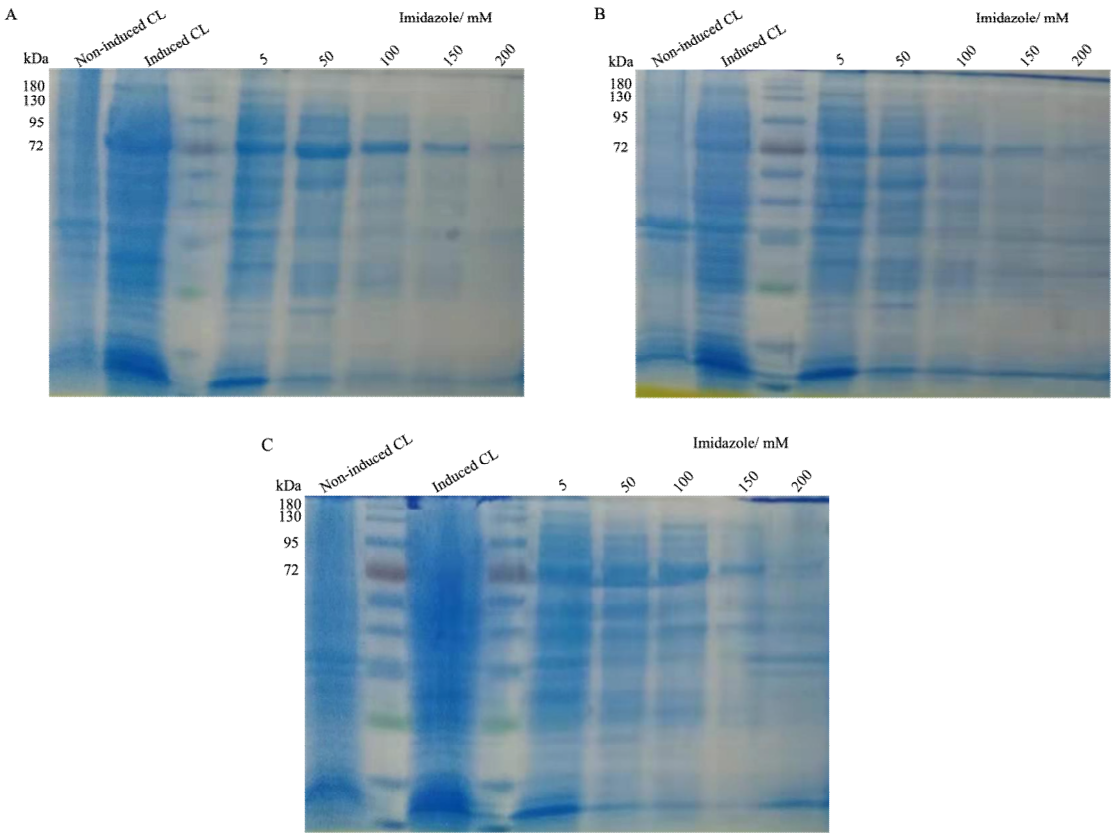


Figure 4.25: SDS-PAGE gel of A: UYC_RS0133560; B: BJP34_RS17220; C: TOL9009_RS40645 after expression and purification from the host *E. coli* BAP1 with Lane 3 in A and B and Lane 2 and 4 in C being the NEB colour prestained protein standard, broad range (10-250 kDa). [CL: Crude lysate].

Moreover, aliquots of the various cultures for BJP34_RS17220 were taken after different time points (0.5 hr, 1 hr, 2 hr, 5 hr, 19 hr) after induction (0.2 mM, 0.5 mM, 1 mM) to check for any expression at a distinct period as well as to check whether the protein is toxic or degrades after expression due to instability (See below).

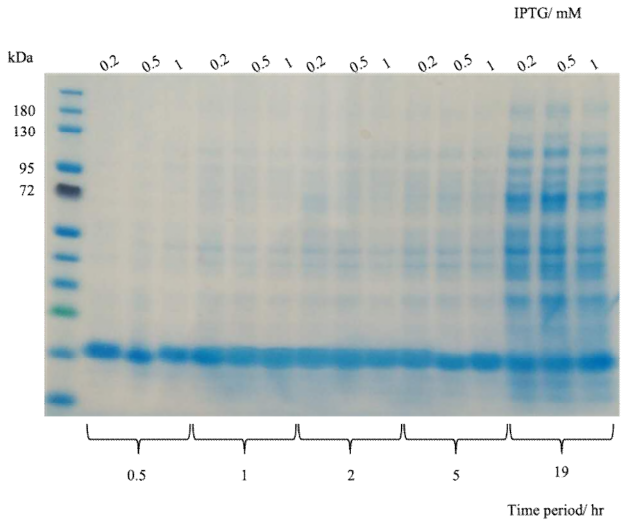


Figure 4.26: SDS-PAGE gel of BJP34_RS17220 after expression at different time point and IPTG concentration from the host *E. coli* BAP1 with Lane 1 being the NEB colour prestained protein standard, broad range (10-250 kDa).

4.8 Discussion

Four enzymes are known to be involved in the biosynthesis of shinorine namely Ava_3858 to Ava_3855 in the cyanobacterium *Anabaena variabilis*[77]. The NRPS-like enzyme Ava_3855 was successfully cloned in pET151, expressed in *E. coli* BL21 and purified using Nickel Sepharose beads. The host *E. coli* BL21 was chosen compared to *E. coli* BAP1 because of the presence of propionate--CoA ligase at 69 kDa which was confirmed by the proteomics results (See Appendix). The latter was seen to be expressed in high yield which could compete with the Nickel Sepharose beads during purification, therefore the choice for *E. coli* BL21. The enzyme Ava_3855 was both confirmed by the proteomics results (See Appendix) as well as through intact protein mass spectrometry. The results from the intact protein mass spectrometry clearly demonstrated the conversion of the "apo" to "holo" form of the enzyme followed by activation and loading of the substrate L-serine, thus displaying its functionality. Furthermore, upon addition of the synthetic precursor AC23, it was seen that the substrate L-serine got offloaded which might suggest its release through a nucleophilic attack from AC23.

However, after analysing the LC-MS results, it was seen that demethylated AC23 (Figure 4.19) might possibly react chemically or enzymatically with glycerol (which comes from the storage buffer of the enzyme), thus forming the novel compound with

chemical formula $C_{12}H_{19}NO_6$, *m/z* $[M+H]^+$ 274.1284 (Figure 4.23). This particular compound was found in the control (boiled Ava_3855); whole reaction without ATP as well as in the complete reaction mix. Two hypotheses can be put forward in this case. One being the fact that demethylated AC23 reacted with glycerol (from the protein storage buffer) chemically, without the involvement of Ava_3855, which might be confirmed by the spectrum above (Figure 4.23). Another plausible explanation is the consumption of ATP by Ava_3855, releasing AMP which is confirmed in Figure 4.22, where L- serine might be activated and loaded on the PCP, which is afterwards released by the nucleophilic attack of demethylated AC23. However, due to the presence of glycerol in the reaction mixture, the latter might contribute into the displacement of L- serine, once it has been added to the MAA backbone forming the enol ester intermediate as explained above prior to the ring rearrangement to form the hexeneimine core as shown below.

Another explanation might be the fact that AC23 or its demethylated version are not very suitable as substrates to be used by Ava_3855 due to the lack of the hydroxyl and hydroxymethyl groups. The hydroxyl and hydroxymethyl groups might play a role in positioning and stabilising the nucleophile close enough to the enzyme for the nucleophilic attack to occur. Therefore, the use of the natural substrate namely mycosporine-glycine might be the best one to use to increase the chance of the nucleophilic attack to proceed, thus leading to the formation of shinorine.

What was also noticed while doing this enzymatic assay was the precipitation of the enzyme Ava_3855 after a couple of hours which might relate to the enzyme having crashed out, suggesting that the yield of the MAA produced (if any) might be too low to detect via mass spectrometry.

Unfortunately, the other four NRPS-like enzymes namely UYC_RS0133560; BUE84_RS13530; BJP34_RS17220; and TOL9009_RS40645 could not be expressed despite the effort of trying to optimize the proper conditions for protein expression.

The following experiment was designed to swap the A-domain of the NRPS-like enzyme Ava_3855 and replace it with the A-domain of the NRPS-like enzyme BJP34_RS17220. This study is further discussed in the next Chapter.

Chapter 5: A-Domain swap in NRPS-like enzyme for production of novel MAAs

5.1 Brief Introduction

Non-ribosomal peptide synthetases (NRPSs) consist of multi-modular enzymes evolved from bacteria and fungi, which have the ability to generate an assortment of pharmacologically potent natural products based on the integration of a variety of proteinogenic and non-proteinogenic amino acids as building blocks. Each module is comprised of different catalytic domains consisting of an adenylation (A) domain whose role is to specifically selects and activate the appropriate amino acid based on the binding pocket of the A-domain; a 4'-phosphopantetheinylated thiolation (PCP) domain which acts as the selected amino acid carrier; a condensation (C) domain leading to peptide bond formation; and a thioesterase (TE) domain which is normally involved in the hydrolysis or cyclization of the mature product via intramolecular nucleophilic attack[150,151].

The focus of this section is mainly on the adenylation domain of NRPS. NRPS enzymes use ATP to catalyse the adenylation of a carboxylate substrate, here referring to an amino acid substrate, leading to the formation of an acyl-AMP intermediate.

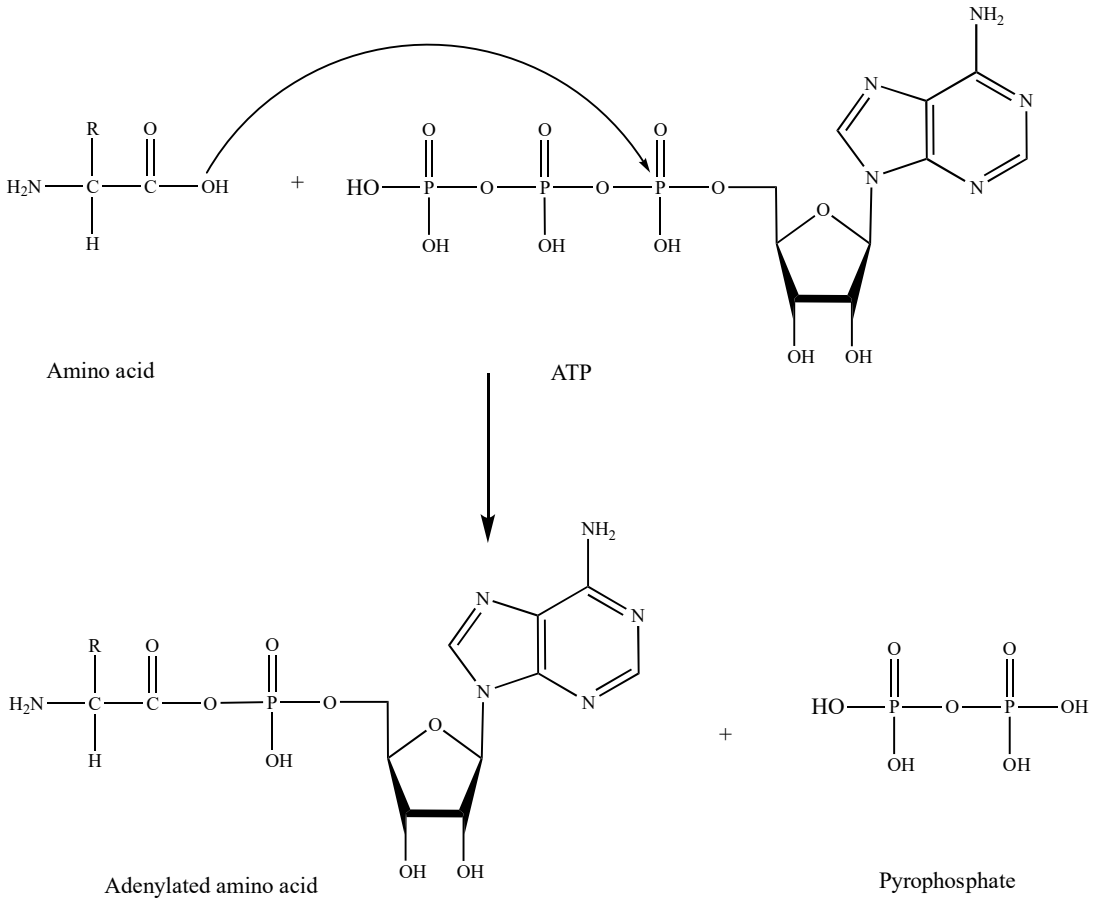


Figure 5.1: Formation of an adenylated amino acid by using ATP.

The selection and activation performed by the A-domain can be regarded as a security guard to allow entry or not. Eight to ten specific amino acids stretching about 100 amino acids apart are present in the binding pocket, thus allowing substrate recognition as well as stabilization of the amino acid substrate[152].

The enzyme binds to CoA, forming a phosphopantetheinyl prosthetic arm which attack the activated amino acid (acyl-AMP), thus releasing AMP[153]. Manipulation of domains using genetic engineering approaches has become quite prevalent recently, especially for the production of optimized natural products, taking into consideration that some disadvantages might arise in terms of non-functionality of the enzymes[154,155].

The goal of this chapter was to swap the A-domain of Ava_3855 to the A-domain of BJP34_RS17220 to generate an engineered NRPS-like enzyme which was predicted to select and activate the substrate L-proline, emphasizing on the fact that no MAA with the L-proline moiety has so far been discovered.

5.2 Sequence alignment for determination of A-domain swap positioning

Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) was used to carry out sequence alignment of the different NRPS-like amino acid sequences and antiSMASH bacterial version (https://antismash.secondarymetabolites.org/#!/start) were used to delineate the specific domains in each of them (See Appendix).

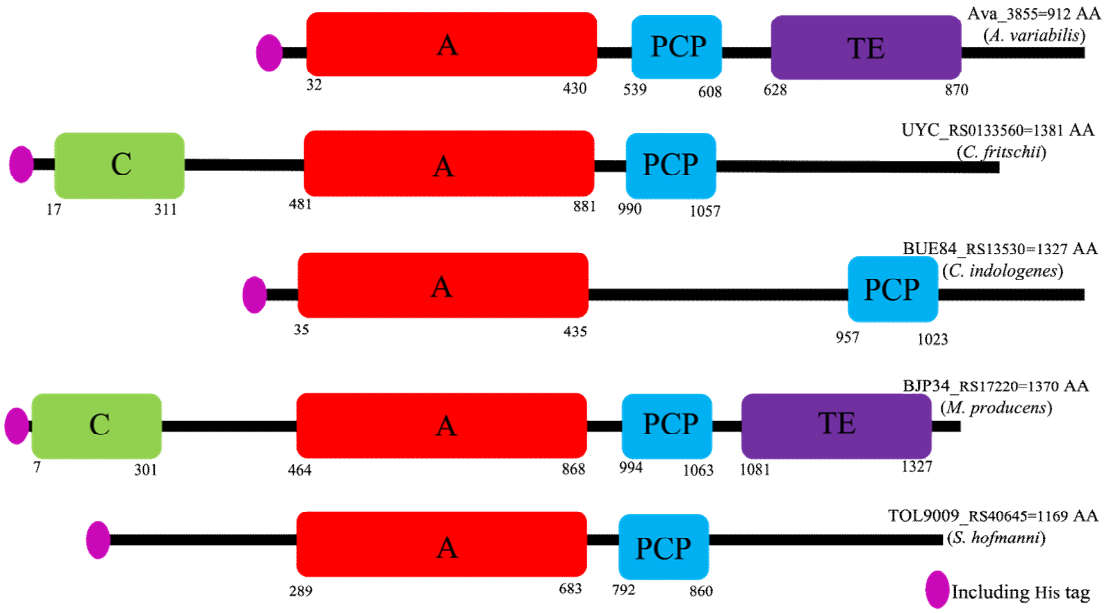


Figure 5.2: Comparison of domain organisation of NRPS-like proteins that direct MAA biosynthesis from various cyanobactria.

The A-domain that was focused on in this study was from BJP34_RS17220, due to its predicted ability to activate L- proline as substrate. It is to be noted that, no MAA with the proline moiety has been discovered so far. Based on the amino acid sequence alignment of Ava_3855 and BJP34_RS17220, a specific region was selected to do the swapping, keeping the conserved start amino acids (T for Threonine and P for Proline) of both sequences.

5.3 Application of Gibson Assembly technique for A-domain swap in NRPS-like enzyme

Gibson assembly technique was used to swap the A-domain of Ava_3855 to the A-domain of BJP34_RS17220 as shown below.

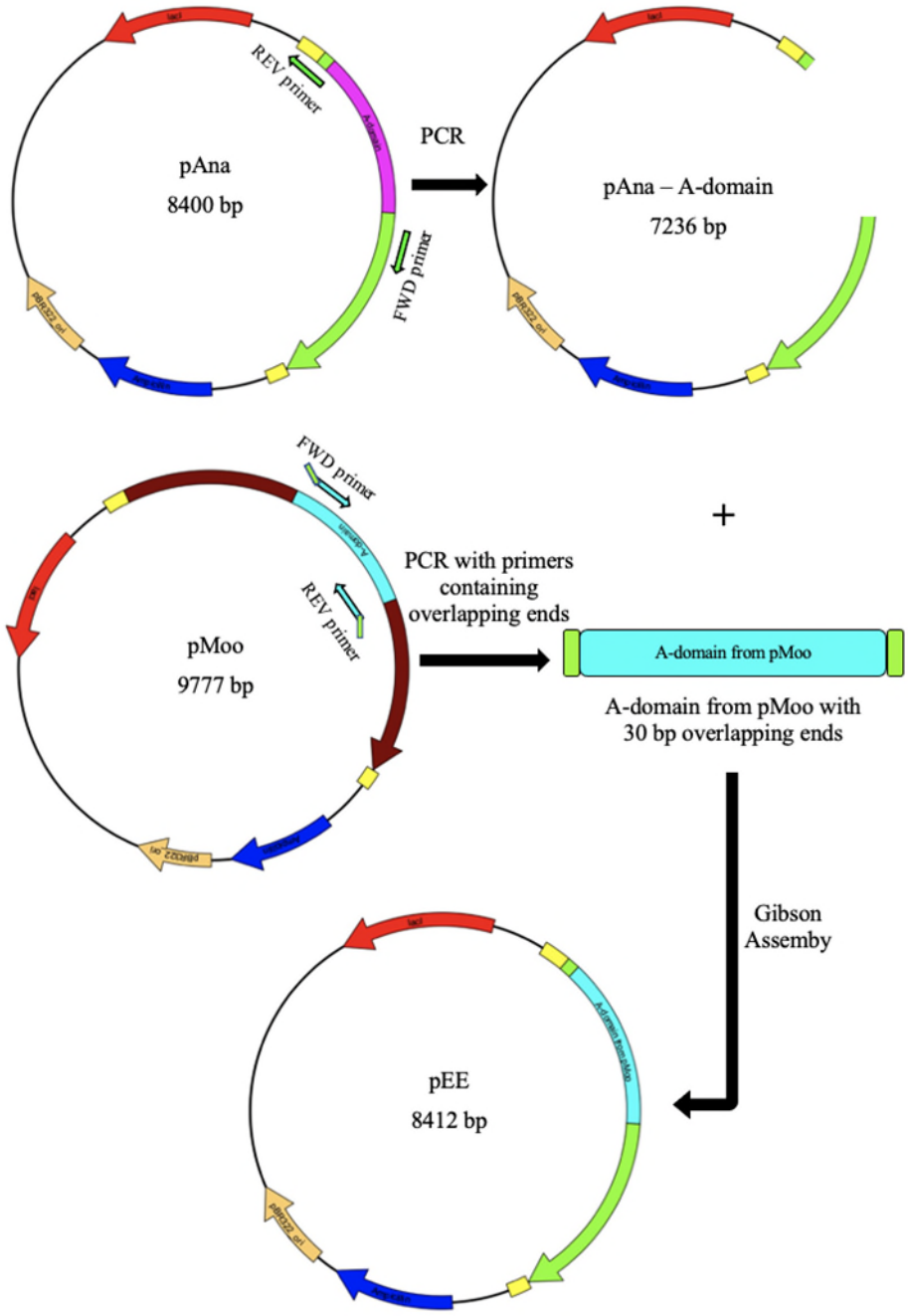


Figure 5.3: Schematic process of A-domain swapping for generation of pEE.

PCR was carried out to amplify the specific DNA sequences from both pAna and pMoo using two corresponding primers (as shown in Chapter 2).

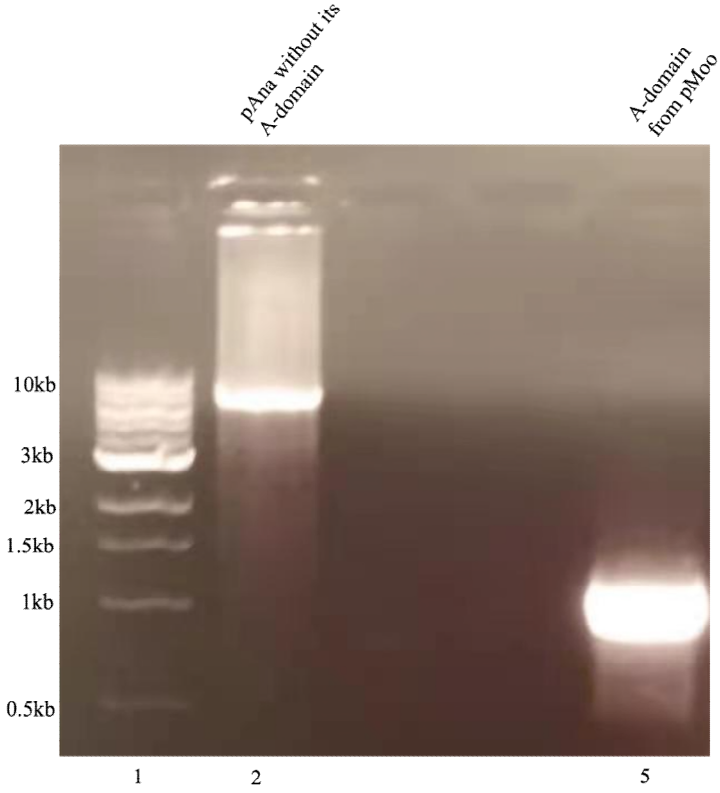


Figure 5.4: PCR products for Gibson assembly. The PCR products were analysed on 1% agarose gel and visualized under UV with the NEB 1 kb ladder in lane 1.

Table 5.1: Summary of PCR results

PCR product	**Expected band size**	**Lane on agarose gel**
Vector (pAna – A-domain)	7236 bp	2
Insert (A-domain from pMoo)	1236 bp	5

The PCR products were extracted from the gel followed by digestion with DpnI restriction endonuclease to make sure that all template DNA were eradicated by its ability to cleave methylated DNA, thus increasing the chance of successful cloning, and decreasing colonies of false positives after transformation. The DpnI digested PCR products were then cleaned using the Monarch® PCR & DNA Clean up Kit and quantified using the NanoDrop™ 2000 Spectrophotometer and the following calculations (Table 5.2) was done for the Gibson assembly cloning process. The NEBuilder® HiFi DNA Assembly Master Mix was used to assemble the 1236 bp insert constituting 30 bp overlapping ends with the 7236 bp vector to create the recombinant

plasmid pEE with a total length of 8412 bp. A ratio of 1:5 vector to insert was used for the cloning process.

Table 5.2: HiFi DNA Assembly calculation

Component	**Length (bp)**	**Concentration (ng/ µl)**	**Amount needed (pmols)**	**Amount needed (ng)**	**Volume (µl) [amt/conc]**
Vector	7236	52.0	0.01	48	0.92
Insert	1236	125.1	0.05	41	0.3
GA master mix					5
Water					3.78
Total volume					10.0

After carrying out Gibson assembly followed by transformation as explained in materials and methods (chapter 2), nine colonies were picked from the plates followed by an overnight culture and plasmid isolation by miniprep the next day. PCR were carried out using the 'T7' forward primer which anneal on the plasmid backbone and the 'Moo 2023 REV' reverse primer which anneal in the insert, thus revealing if the cloning worked or not.

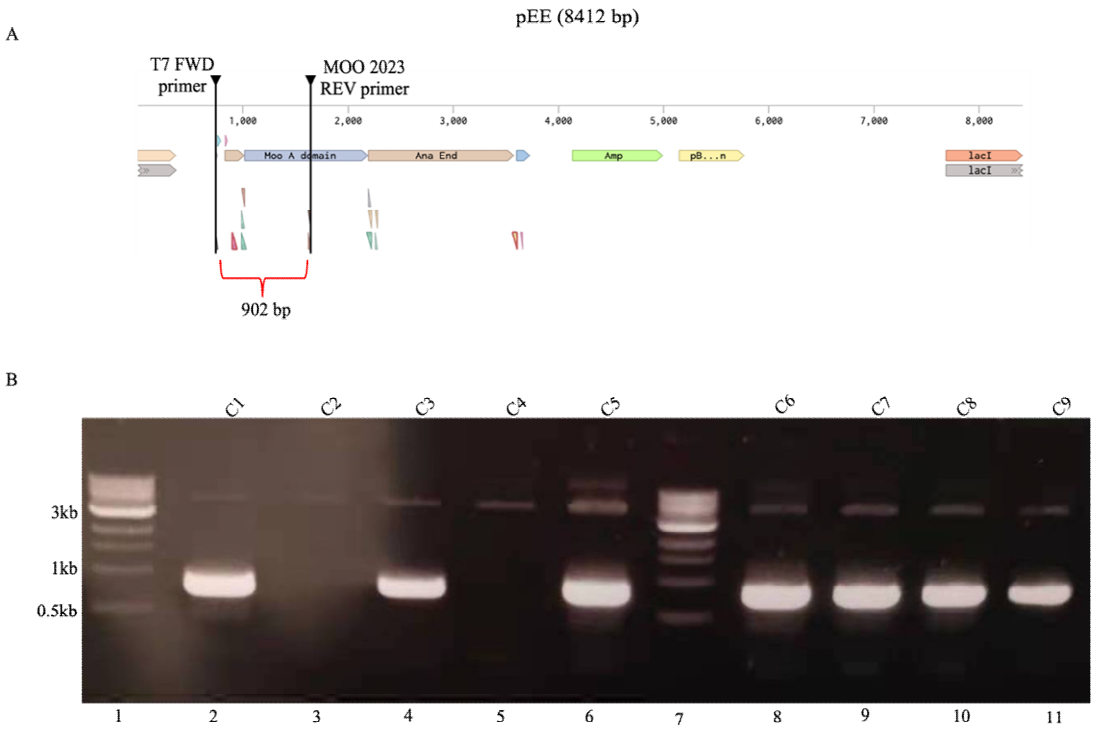


Figure 5.5: A: Primers annealing position in pEE. B: PCR screening of purified plasmids from nine colonies. The gel shows the PCR products from selected purified plasmids with the NEB 1 kb ladder loaded in the first and seventh lanes. C1 to C9 = colony number.

Table 5.3: Summary of PCR product

PCR product	Expected band size	Lane on agarose gel
T7 to Moo 2023 REV	902 bp	2; 4; 6; 8; 9; 10; 11

Among the nine selected plasmids, seven were shown to have the correct PCR product size including C1, C3, C5, C6-C9.

A restriction digest using HpaI as restriction enzyme was chosen as the latter was seen to cut in the A-domain of the NRPS-like gene *bjp34_rs17220* compared to none in the A-domain of the NRPS-like gene of *ava_3855* based on a virtual digest carried out on Benchling.

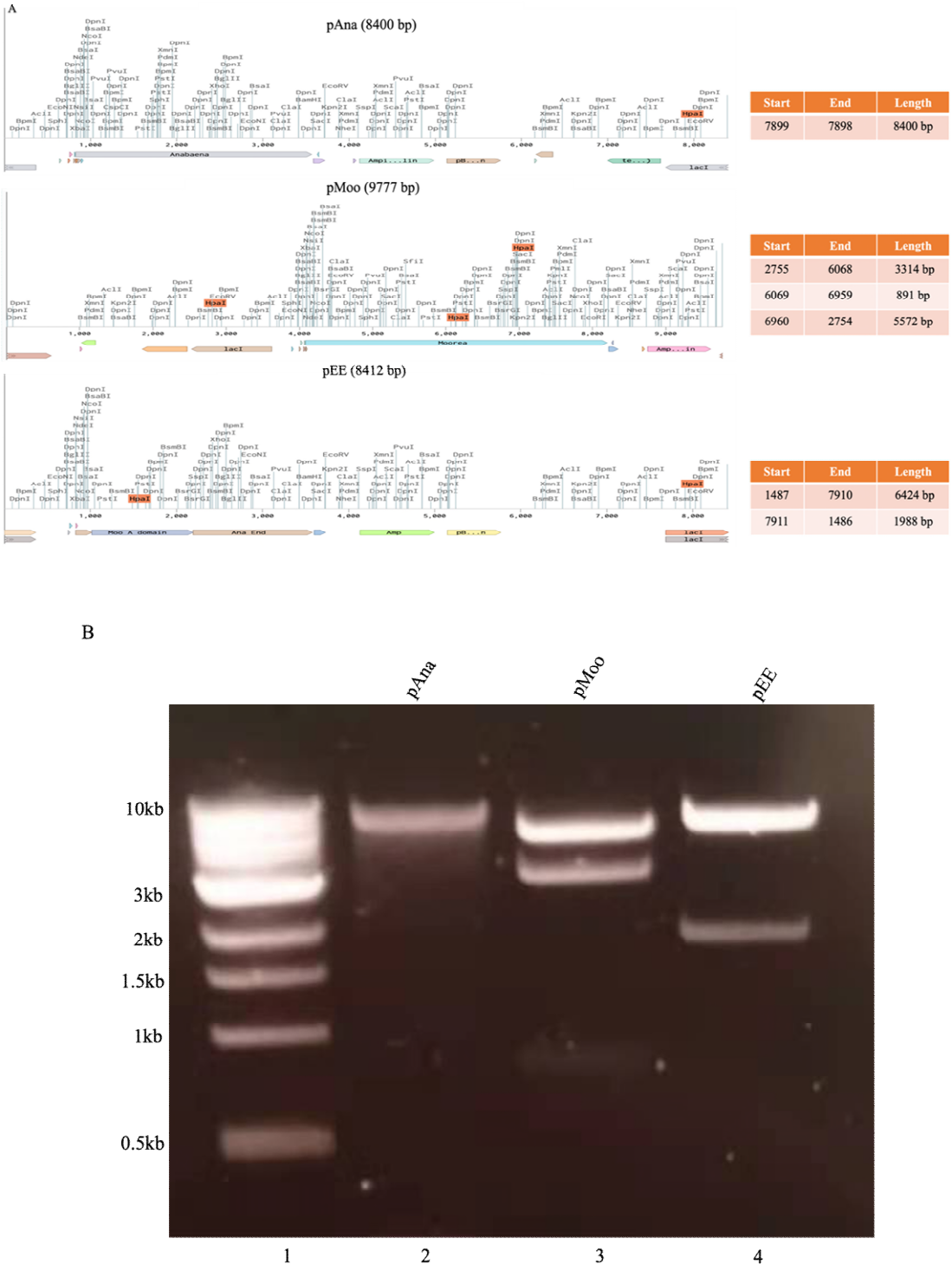


Figure 5.6: A: Position of HpaI in the three different plasmids and the corresponding fragments after digestion. B: Restriction digest of pAna; pMoo and pEE using HpaI.

The gel shows the results of the restriction digest when carried out on the various constructs with the NEB 1 kb DNA ladder being used in the first lane.

Table 5.4: Summary of bands size after HpaI digestion

Construct	Size of digested band (bp)	Lane
pAna	8400	2
pMoo	5572; 3314; 891	3
pEE	6424; 1988	4

Based on the band size of the PCR product as well as the restriction digest performed using HpaI, it could be determined that the Gibson assembly worked. The recombinant plasmid (pEE) was sent with corresponding sequencing primers (Chapter 2) for Sanger sequencing and the sequencing data were clearly positive and matched the required sequence. The engineered plasmid was then transformed in *E. coli* BL21 for future experiment.

5.4 Expression and purification of the recombinant protein NEE using HisPur™ Ni-NTA Resin

The *E. coli* strain containing the successfully engineered plasmid was cultured and the expression of the gene was firstly induced using 0.5 mM IPTG and incubated overnight at 15ºC. The following day, lysis buffer A (see chapter 2) supplemented with lyzozyme and protease inhibitor were used, followed by sonication (8 X 30 seconds pulses with 1 minute recovery in between each pulse) and centrifugation. HisPur™ Ni-NTA Resin was used to allow binding of the NEE with its His tag and eluted with different imidazole concentrations as shown below.

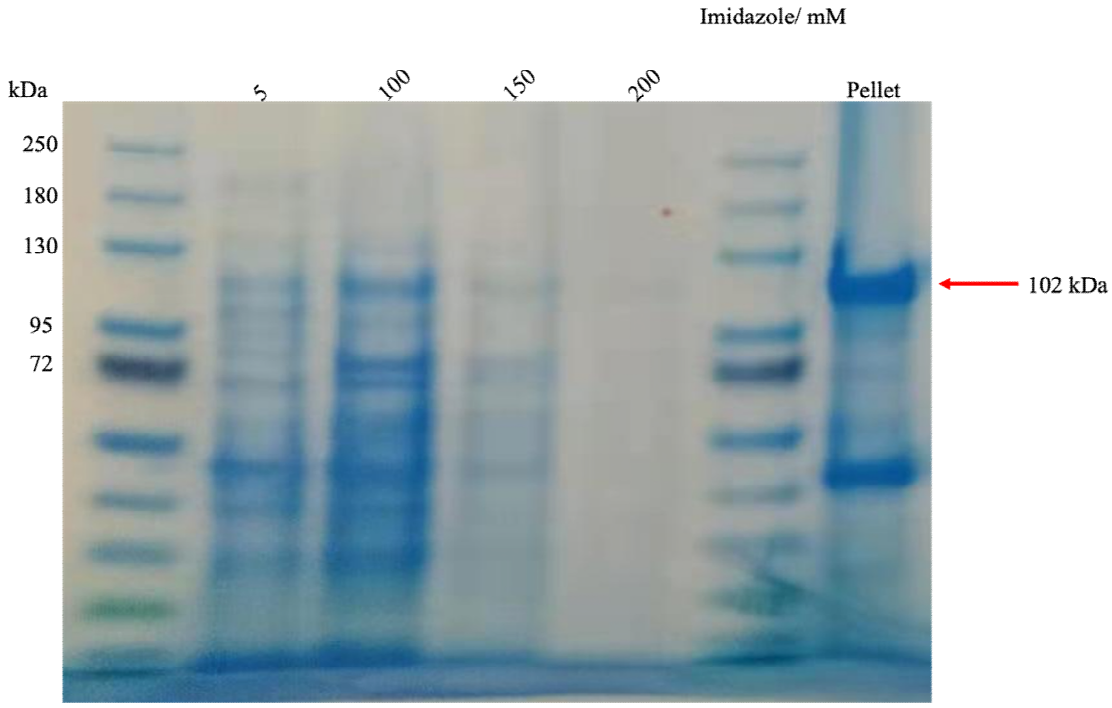


Figure 5.7: SDS-PAGE gel of NEE after expression and purification from the host *E. coli* BL21 with Lane 1 and 6 being the NEB colour prestained protein standard, broad range (10-250 kDa).

The gel above clearly revealed that the enzyme NEE was present in the insoluble fraction (Lane 8) with the molecular weight of 102 kDa. To remediate to this, a lower concentration of IPTG was used for induction (0.2 mM) and different lysis buffers with varying pH (8, 8.5, 9), 5% glycerol and 5% DMSO were employed to increase the chance of the enzyme's solubility.

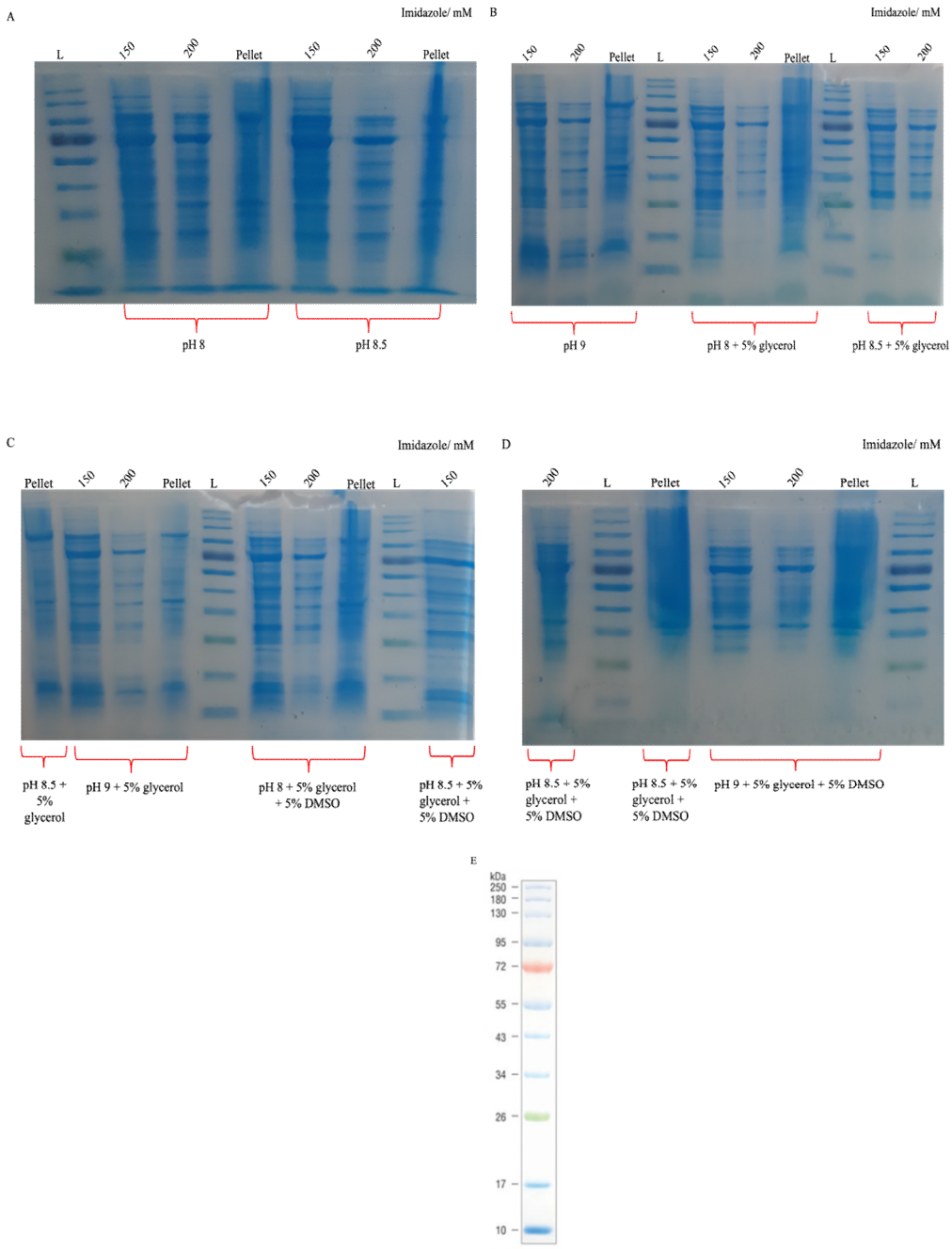


Figure 5.8: A-D: SDS-PAGE gel of NEE (0.2 mM IPTG for induction) using different lysis buffers after expression and purification from the host *E. coli* with L being the NEB colour prestained protein standard, broad range (10-250 kDa). E: NEB colour prestained protein standard, broad range (10-250 kDa) for reference.

The expression and purification of NEE was achieved using 0.2 mM IPTG for induction which reduced the formation of inclusion bodies significantly and allowed presence of the protein in the soluble fraction followed by usage of lysis buffer B with pH of 8.5 (refer to Chapter 2).

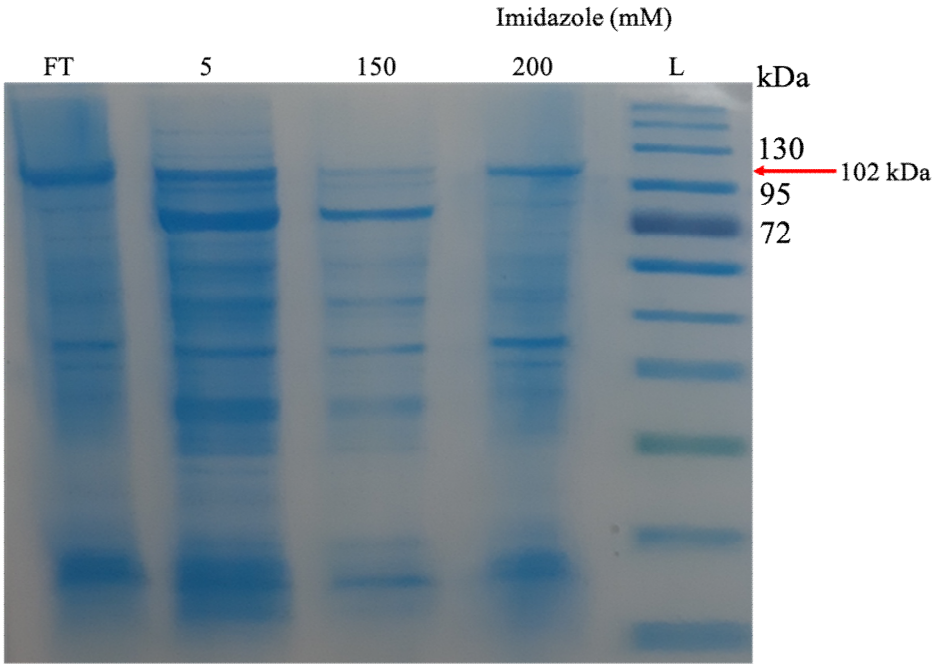


Figure 5.9: SDS-PAGE gel of the recombinant protein NEE with L being the NEB colour prestained protein standard, broad range (10-250 kDa). [FT = Flow Through]

However, the expression level of NEE was quite low after purification. Furthermore, despite large cultures (2 L) were done by repeating the above chosen conditions, it was seen that the protein expression was seen as just a fade band (as shown below in Figure 5.10) on the SDS-PAGE gel, which made it an unreliable way of purifying this engineered enzyme (yield = 0.210 mg/ ml).

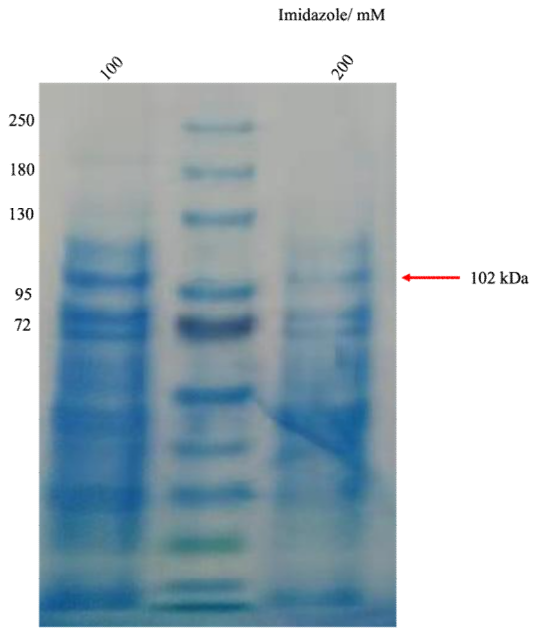


Figure 5.10: SDS-PAGE gel of the recombinant protein NEE with Lane 2 being the NEB colour prestained protein standard, broad range (10-250 kDa) after purification of NEE from 2 L culture.

The reason for this might have been due to the folding pattern of the new engineered protein after the A-domain swap where the enzyme might be more prone to be insoluble (as shown in Figures 5.7 and 5.8) or the His-tag might have been in a position which made it difficult to bind to the Nickel beads, thus reducing the level of binding affinity.

5.5 Recombinant plasmid generation of pET28MBPNEE using Gibson assembly

To remedy to the low expression level of NEE, a different plasmid was used namely pET28a-MBP-TEV which contained the maltose binding protein tag as well as two His-tags, one at the N-terminus and the other one at the C-terminus. This plasmid was chosen due to the presence of the MBP tag which was supposed to aid in solubilizing the protein of interest. Furthermore, the two His-tags were ideal in the sense that they could increase the binding affinity of the engineered enzyme to the Nickel beads, avoiding loss of the latter in the washing process during protein purification.

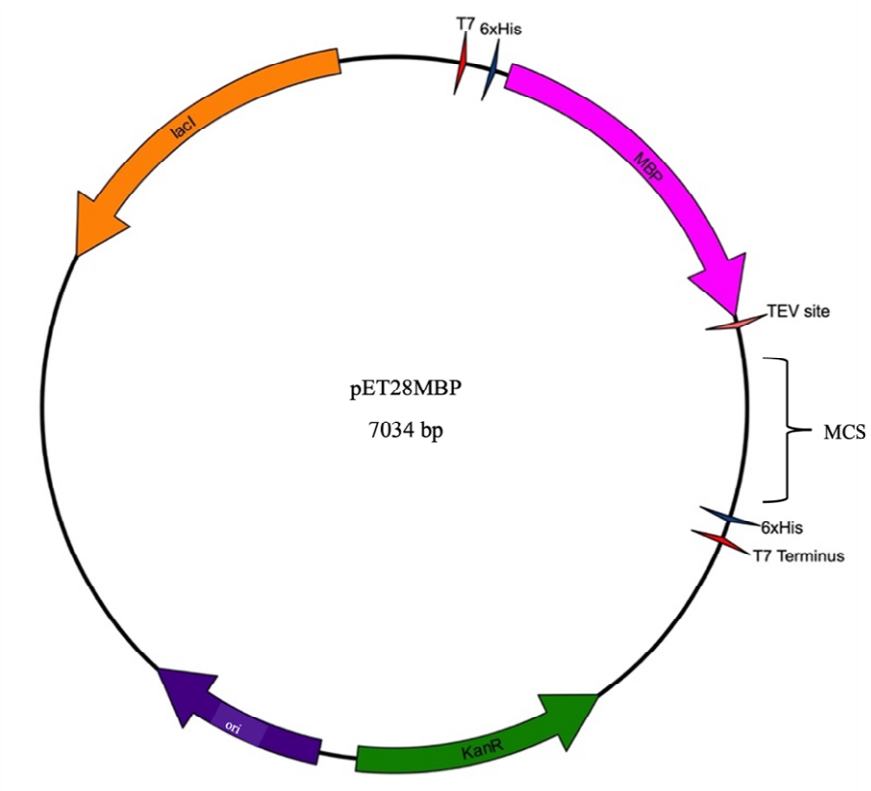


Figure 5.11: pET28MBP plasmid containing two His tags, one at the N-terminus and one at the C-terminus along with a maltose binding protein (MBP) tag to help with solubilization of the protein of interest.

Specific primers as given in Chapter 2 were designed to generate the PCR products capturing the whole engineered NRPS-like gene sequence with the swapped A-domain from pEE to be cloned in pET28MBP vector.

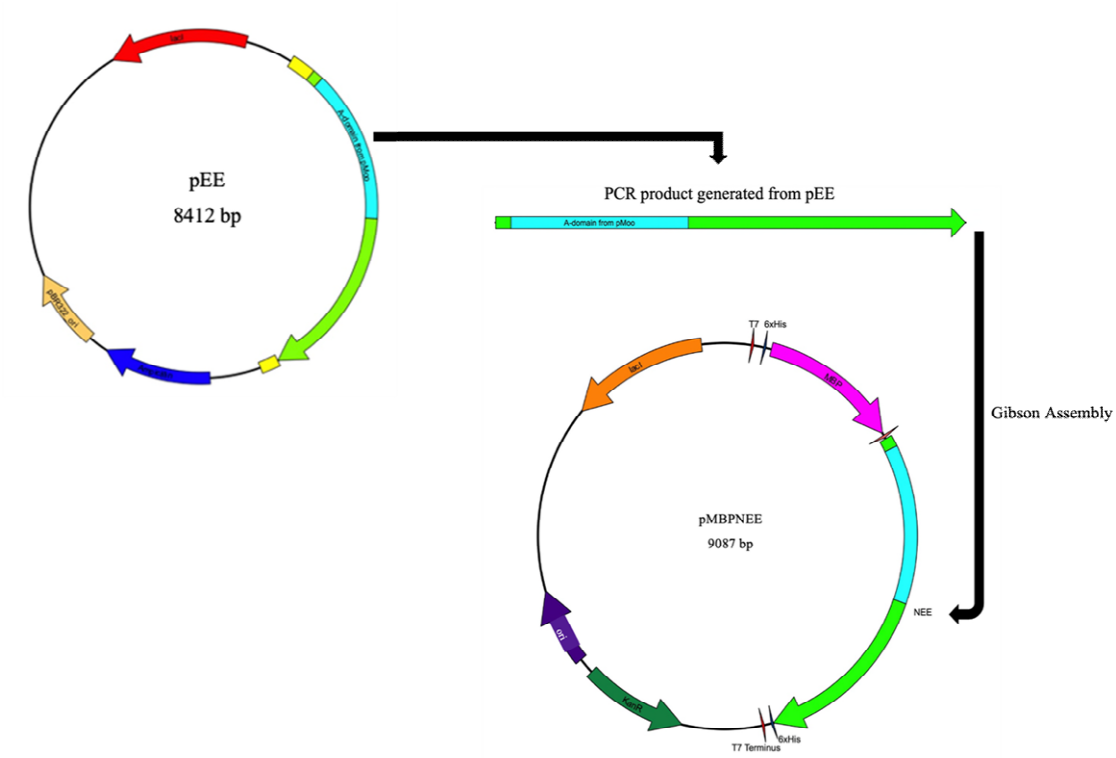


Figure 5.12: The recombinant NRPS-like enzyme NEE PCR product was cloned in pET28MBP to generate pMBPNEE.

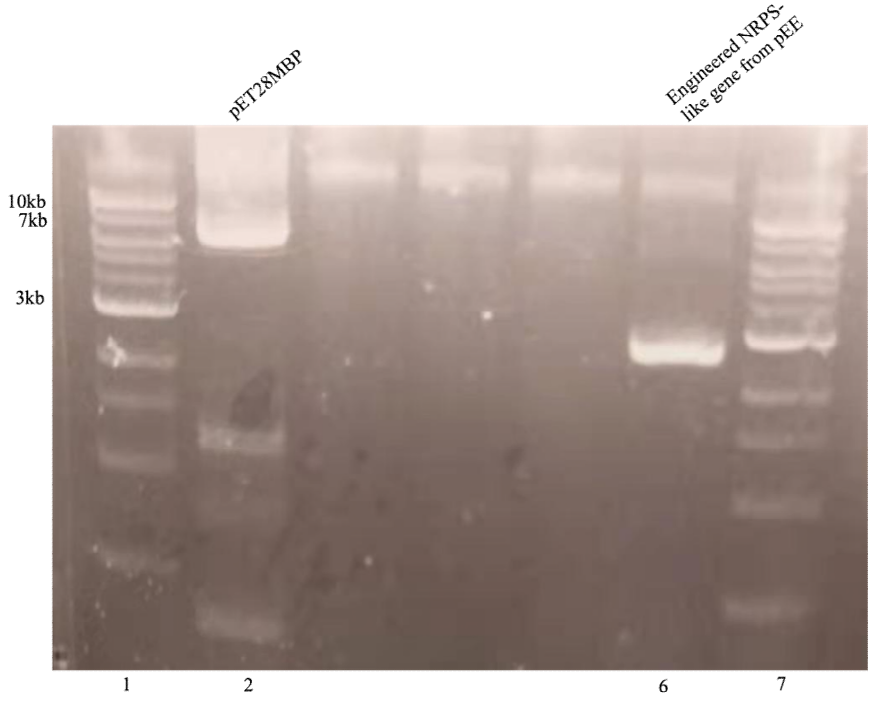


Figure 5.13: Corresponding PCR products with the expected size. PCR products were analysed on 1% agarose gel using the NEB 1 kb DNA ladder in the first and seventh lanes.

Table 5.5: Summary of PCR results

PCR product	Expected band size	Lane on agarose gel
Vector (pET28MBP)	6438 bp	2
Insert (Engineered NRPS-like gene)	2649 bp	6

After obtaining the PCR products, the corresponding bands were extracted from the gel followed by digestion with DpnI restriction endonuclease, cleaned using the Monarch® PCR & DNA Clean up Kit and quantified using the NanoDrop™ 2000 Spectrophotometer.

The NEBuilder® HiFi DNA Assembly Master Mix was used to assemble the 2649 bp insert constituting 30 bp overlapping ends with the 6438 bp vector to create the recombinant plasmid pMBPNEE with a total length of 9087 bp. The following calculations (as shown in Table 5.6) was done for the Gibson assembly cloning process.

Table 5.6: HiFi DNA Assembly calculation

Component	Length (bp)	Concentration (ng/ µl)	Amount needed (pmols)	Amount needed (ng)	Volume (µl) [amt/conc]
Vector	6438	44.0	0.01	42	0.95
Insert	2649	70.0	0.05	87	1.24
GA master mix					5
Water					2.81
Total volume					10.0

After carrying out Gibson assembly followed by transformation, PCR from ten minipreped colonies were carried out using the 'T7' forward primer and the 'T7 Reverse' primers (See chapter 2) to check whether the insert has been successfully cloned in the vector.

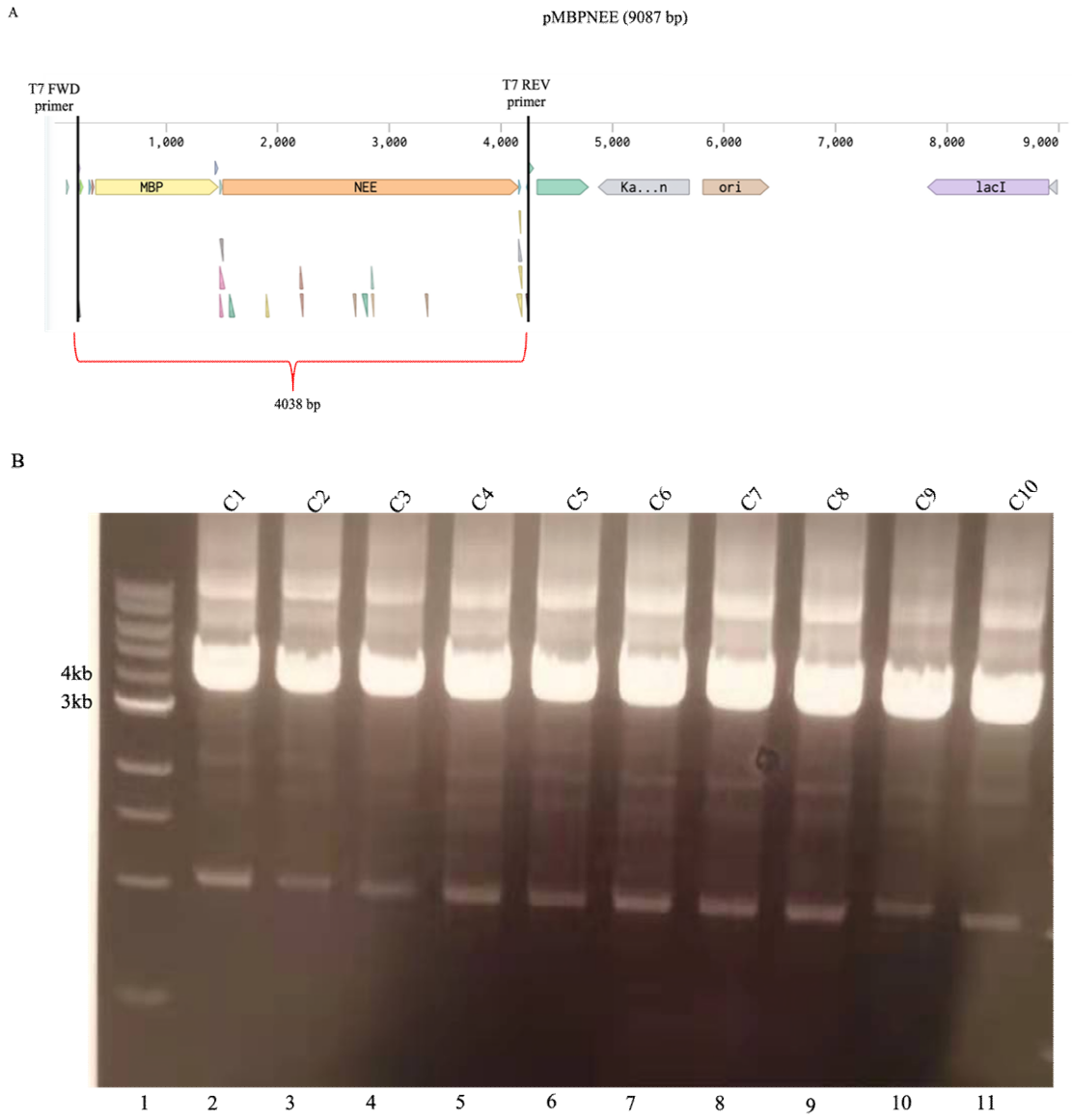


Figure 5.14: A: Primers annealing position in pEE. B: PCR screening of purified plasmids from ten colonies. The gel shows the PCR products from selected purified plasmids with the NEB 1 kb ladder loaded in the first Lane. C1 to C10 = colony number.

Table 5.7: Summary of PCR results

PCR product	Expected band size	Lane on agarose gel
From T7 Forward primer to T7 Reverse primer	4038 bp	2-11

Based on the band size of the PCR products as well as the confirmation of the Sanger sequencing results which clearly matched the expected sequence, it could be determined that the cloning worked leading to the generation of pMBPNEE. The recombinant plasmid was then transformed in *E. coli* BL21 for future experiment.

5.6 Expression and purification of pMBPNEE using HisPur™ Ni-NTA Resin

The *E. coli* BL21 strain containing the successful recombinant plasmid pMBPNEE was cultured in LB supplemented with 50 µg/ml kanamycin and induced with 0.2 mM IPTG at OD_{600} about 0.5 and incubated overnight at 15°C. The following day, the lysis buffer B supplemented with lyzozyme and protease inhibitor was used, followed by sonication (8 X 30 seconds pulses with 1 minute recovery in between each pulse) and centrifugation. The NRPS-like enzyme was purified using HisPur™ Ni-NTA Resin as explained in Chapter 2 with different imidazole gradient elution. The size of the engineered enzyme NEE fused to the protein expression MBP tag (43 kDa) led to a total predicted protein size of around 143 kDa as confirmed on the SDS-PAGE gel below.

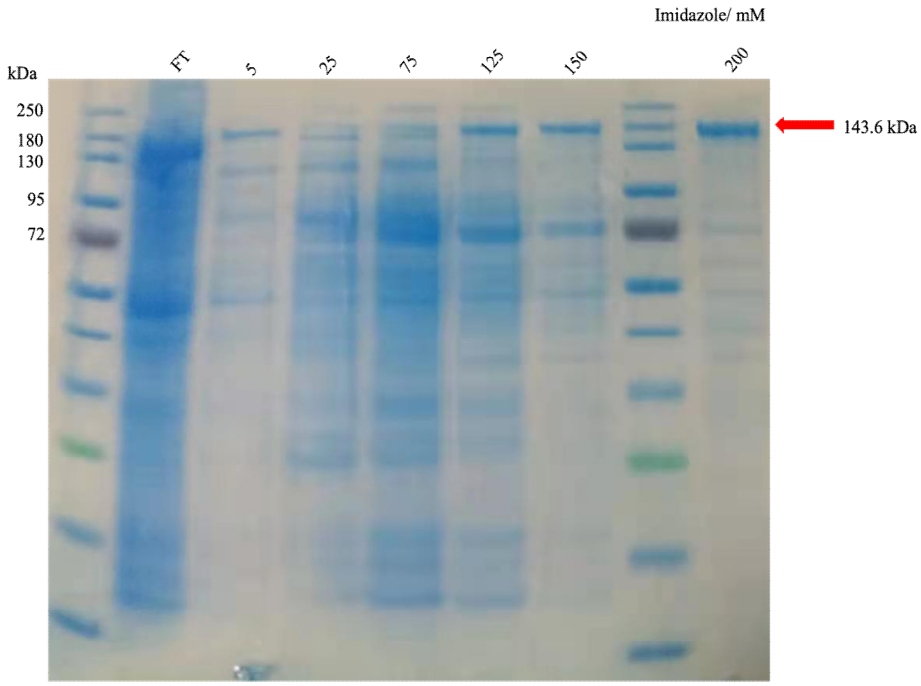


Figure 5.15: SDS-PAGE analysis of purified MBPNEE. The expected protein size of 143.6 kDa is seen eluting with 200 mM imidazole.

The fractions containing the purified protein of interest were combined and concentrated using an Amicon Ultra 50 kDa molecular weight cutoff spin filter. A buffer exchange with storage buffer F using the PD10 column was carried out and the protein was aliquoted and stored at -80ºC for further experiment.

5.7 Chemo-enzymatic assay using MBPNEE and the synthetic precursor AC23

A proposed mechanism involving the catalysis of the synthetic precursor AC23 with activated L-proline by MBPNEE is shown below where AC23 might be acting as a strong nucleophile.

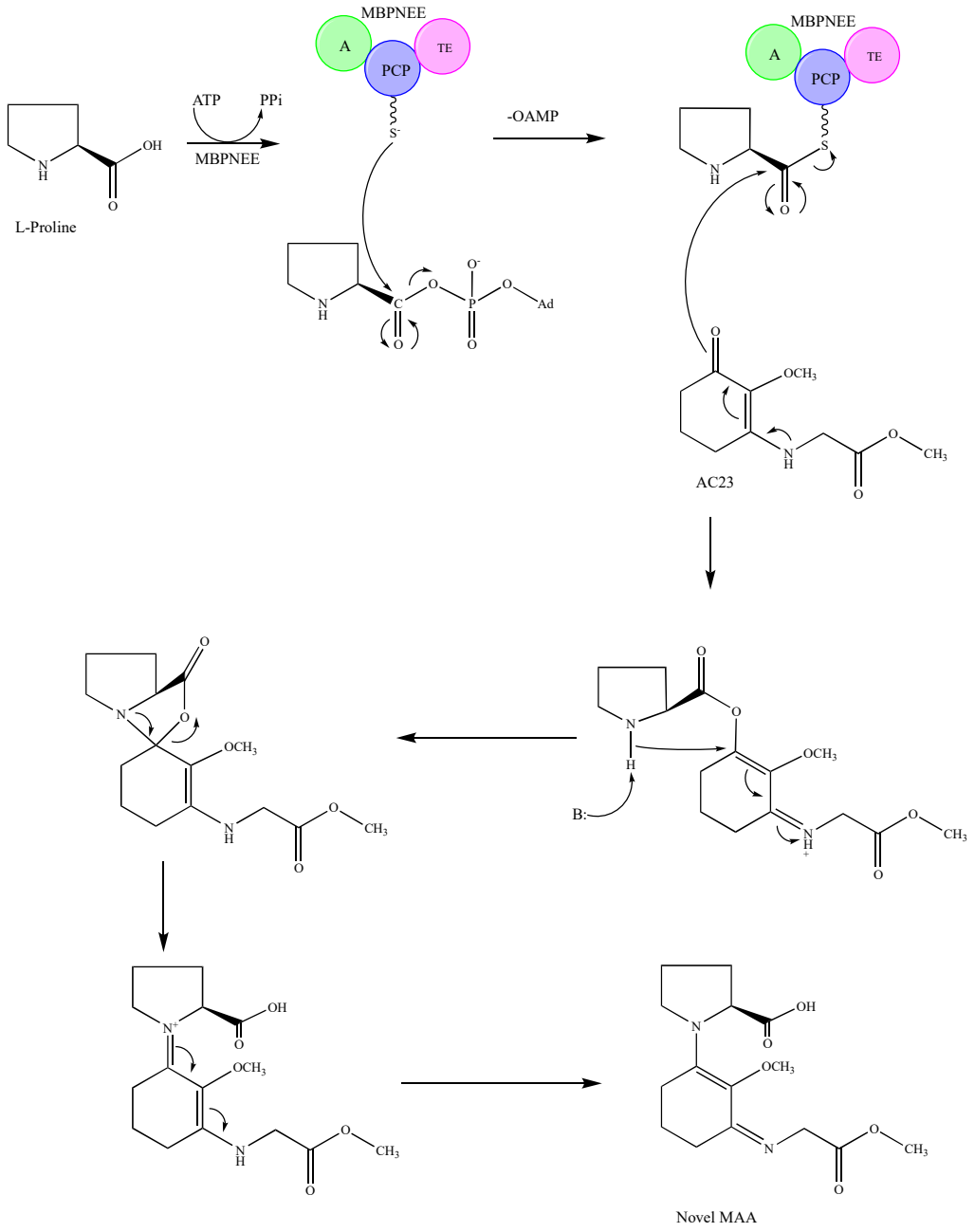


Figure 5.16: Proposed mechanism of AC23 with activated L-proline catalysed by MBPNEE. AC23 might be acting as a strong nucleophile, thus releasing the activated L-proline substrate from the recombinant enzyme MBPNEE.

After the conversion of the "apo" to "holo" form by the addition of the phosphopantetheinyl arm prosthetic group, ATP was used to load the L-proline substrate on the PCP by a phosphodiester bond with the release of OAMP. AC23 might be acting as a strong nucleophile, leading to the release of the L-proline from the engineered enzyme MBPNEE, leading to the formation of an enol ester

intermediate, followed by rearrangement of O to N forming a cyclohexeneimine core, thus giving rise to a novel type of MAA.

A chemo-enzymatic assay was set up to check for any product generated using the synthetic precursor AC23. A final volume of 100 µL solution containing 5 mM $MgCl_2$, 0.1 mM CoA, 500 µM Sfp, 1 mM L- proline, 2 mM AC23, 2 µM MBPNEE and 5 mM ATP were used for the assay which was carried out at room temperature overnight. The UHPLC-HRMS results are displayed below.

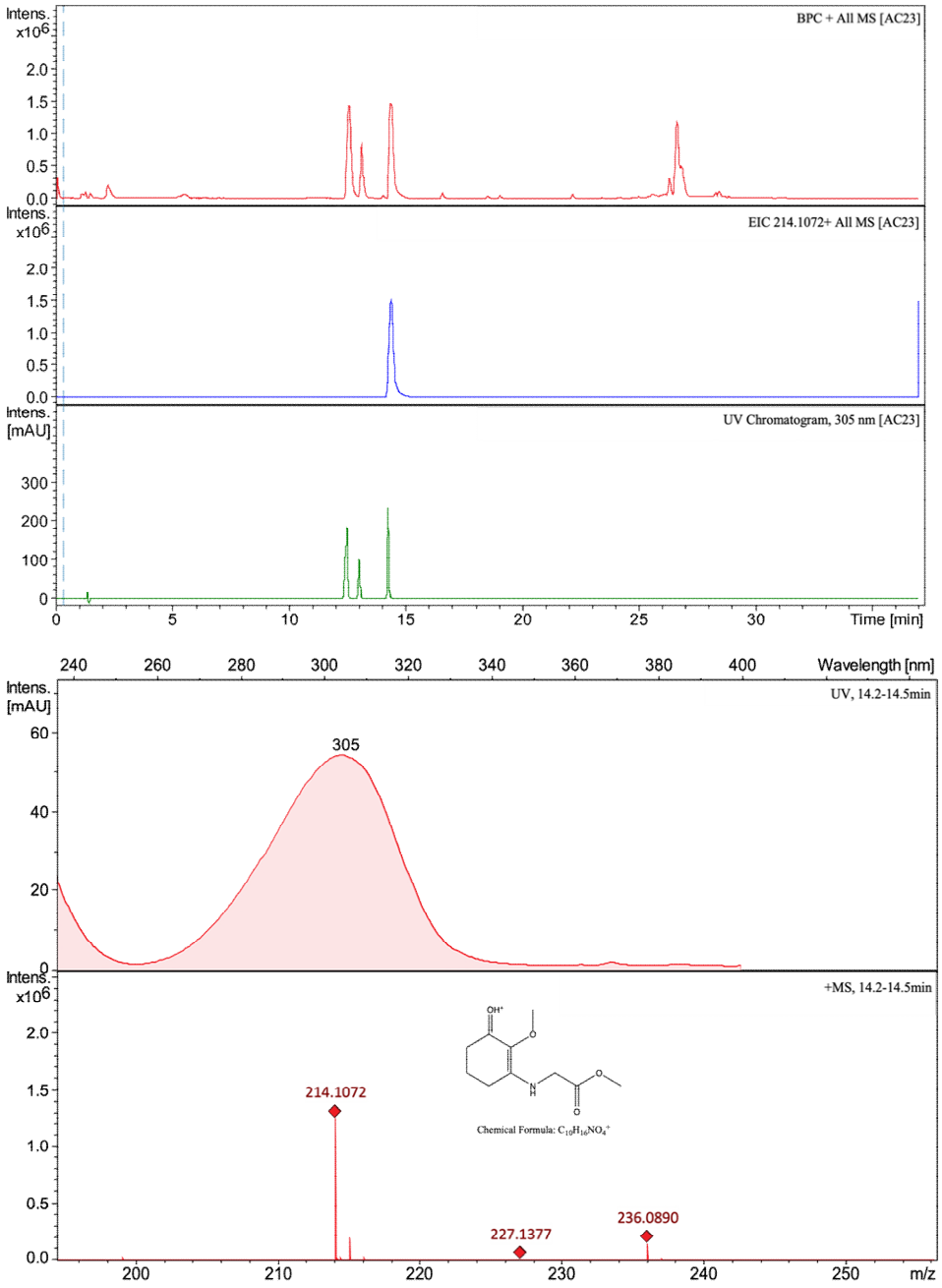


Figure 5.17: The chromatogram of AC23. Top: Base peak chromatogram (BPC) (Red); Extracted ion chromatogram (EIC) for *m/z* 214.1072 (Blue); UV chromatogram at 305 nm (Green). Bottom: UV and mass spectra of AC23 after UHPLC-HRMS analysis.

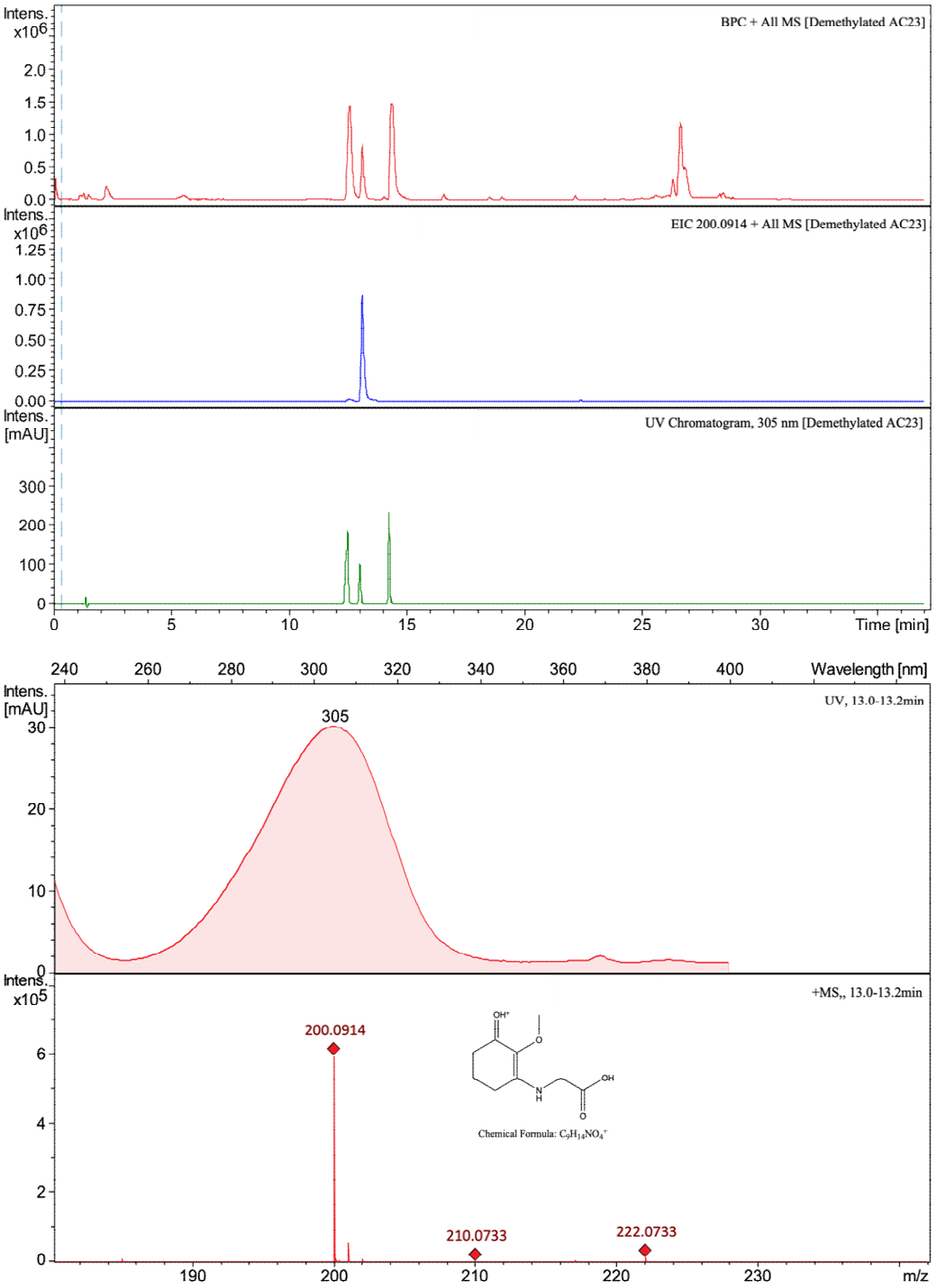


Figure 5.18: The chromatogram of demethylated AC23. Top: Base peak chromatogram (BPC) (Red); Extracted ion chromatogram (EIC) for *m/z* 200.0914 (Blue); UV chromatogram at 305 nm (Green). Bottom: UV and mass spectra of demethylated AC23 after UHPLC-HRMS analysis.

The consumption of ATP by the engineered enzyme MBPNEE followed by the release of AMP (Figure 5.19), revealed that L- proline might be activated and loaded on the PCP, which might afterwards be released by the nucleophilic attack of AC23.

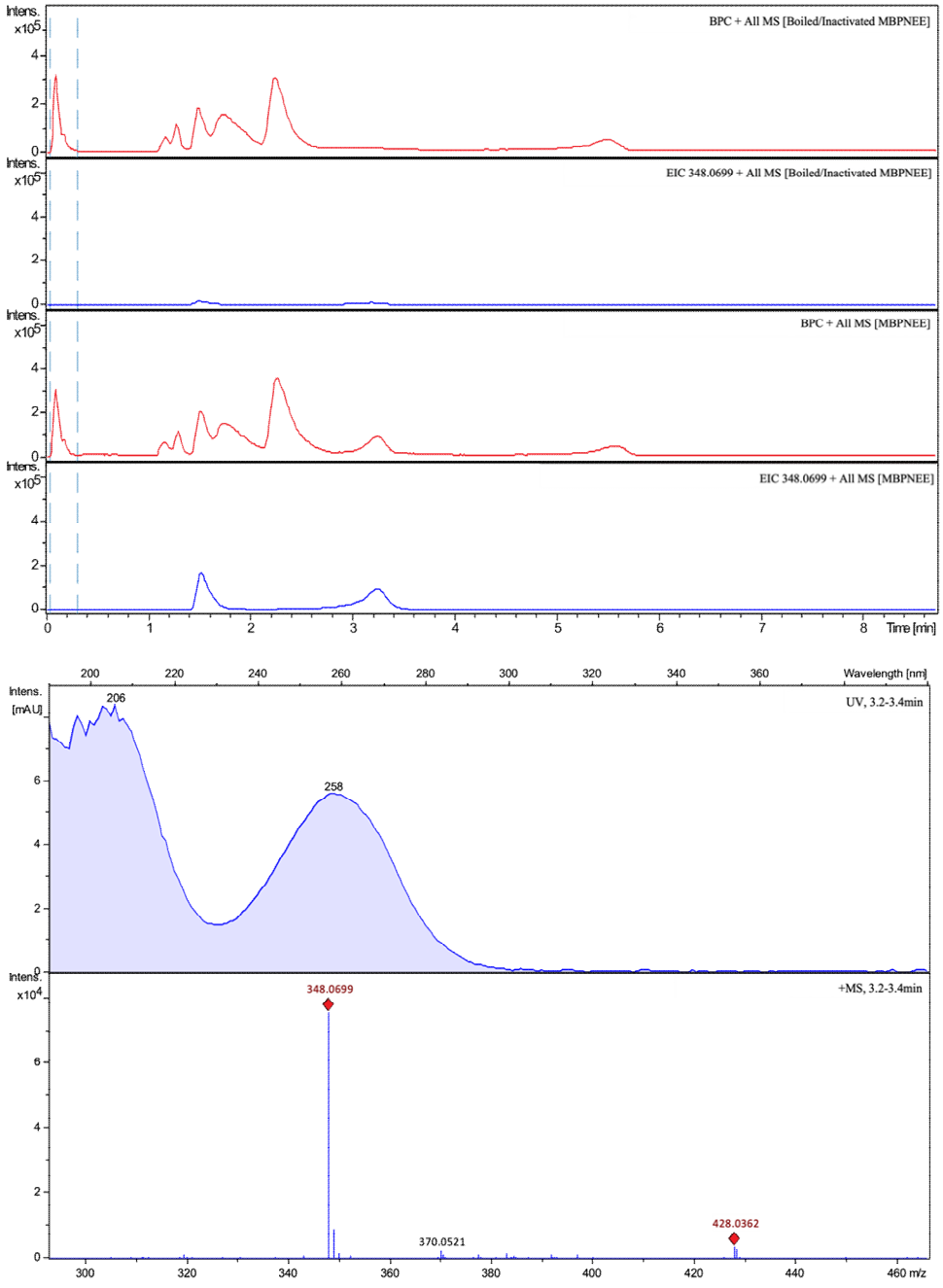


Figure 5.19: The chromatogram monitoring AMP production with boiled/inactivated MBPNEE and with MBPNEE. Top: Base peak chromatogram (BPC) (Red); Extracted ion chromatogram (EIC) for *m/z* 348.0706 (Blue). Bottom: UV chromatogram at 258 nm and mass spectra of AMP after UHPLC-HRMS analysis.

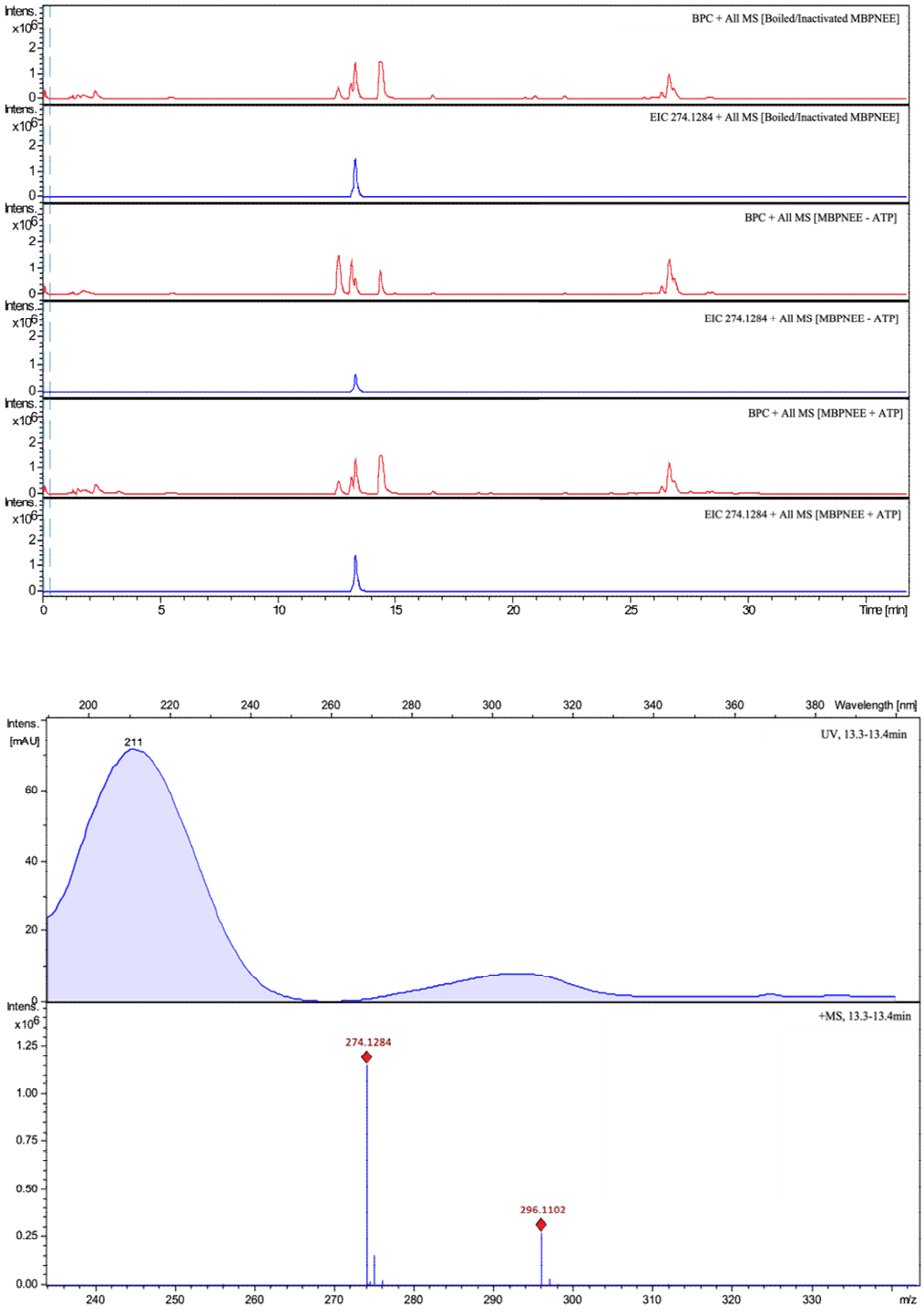


Figure 5.20: The chromatogram of boiled/inactivated MBPNEE; MBPNEE without ATP and MBPNEE with ATP. Top: Base peak chromatogram (BPC) (Red); Extracted ion chromatogram (EIC) for *m/z* 274.1284 (Blue). Bottom: UV chromatogram at 211 nm and mass spectra of novel compound after UHPLC-HRMS analysis.

5.8 Discussion

The A-domain of Ava_3855 was successfully swapped to the A-domain of BJP34_RS17220 which was predicted to activate L-proline as substrate. The engineered NRPS-like enzyme generated namely NEE was obtained in the insoluble fraction, however after optimization of the conditions during protein purification, negligible yield was obtained. This could be due to various factors including formation of inclusion bodies during purification or the whole folding pattern of the protein might have changed due to the A-domain swap, thus making the binding of the His-tags with the Nickel beads difficult.

Thus, to remediate to that, the recombinant NRPS-like enzyme NEE was cloned in pET28MBP which is known to help in protein solubilization due to the maltose binding protein tag as well as the presence of two His-tags at both the N-terminus and C-terminus of the protein of interest which was successfully purified from *E. coli* BL21 leading to the recombinant enzyme MBPNEE.

During the chemo-enzymatic assay, the novel compound with chemical formula $C_{12}H_{19}NO_6$ having *m/z* $[M+H]^+$ 274.1284 was detected in the control containing the boiled/ inactivated MBPNEE, the reaction mix without ATP as well as the whole reaction mix. As explained in Chapter 4, two possibilities could arise where demethylated AC23 reacted chemically with glycerol or the displacement of L-proline by glycerol after the formation of an enol ester intermediate with demethylated AC23. However, no predicted product (as shown in Figure 5.16) was detected after UHPLC-MS analysis which might be because of the low yield, or another possibility might be the instability of the enol ester intermediate which quickly get substituted by glycerol, thus forming the novel compound with chemical formula $C_{12}H_{19}NO_6$.

Another hypothesis might be the fact that the synthetic substrate AC23 or its demethylated version could not stay in proximity with the enzyme to unload the substrate proline due to the lack of the hydroxyl group which might be important for stability and closeness while the nucleophilic attack is happening. Therefore, the ideal substrate would be mycosporine-glycine to rule out every possible issue for the enzymatic reaction to happen.

Chapter 6: Proposed role of the TE domain in Ava_3855 and MBPNEE

6.1 Brief Introduction

Bioactive peptides are known to be produced by megaenzymes called nonribosomal peptide synthetases (NRPSs) without the need of ribosome, thus widening an assortment of around 500 non-proteinogenic amino acids,[156] including D-configured or N-methylated ones to be selected by these enzymes. NRPSs are organized into modules where each module is somehow a self-regulating functional block resulting in a cycle of peptide elongation. One module is in turn divided into specific domains namely the adenylation domain (A-domain), peptidyl carrier protein (PCP), condensation domain (C-domain) and thioesterase domain (TE domain). The A-domain is responsible for the selection of the appropriate amino acid which greatly depends on the amino acid residues in the binding pocket of the domain which is then activated as an aminoacyl adenylate through ATP hydrolysis. The PCP allows the selected amino acid to be tethered as a thioester via the phosphopantetheine arm resulting from coenzyme A, aminoacyl-S-PCP. The role of the C-domain is the peptide bond formation[157] resulting from a nucleophilic attack from the nitrogen of the amino group of the cognate aminoacyl-S-PCP which acts as a nucleophile with the peptidyl-S-PCP from the upstream module. Other domains have also been revealed like the epimerization domain (E-domain) which has the ability to change the stereochemistry of the C-terminal amino acid residues from L to D form[158].

The last step in the biosynthesis of full-length bioactive peptides is the release of the latter from the NRPSs, which is usually accomplished by the TE domain (approximately 28 kDa) which is normally situated at the C-terminus of the NRPS assembly lines[159].

The goal of this chapter was to determine a proposed role of the TE domain of Ava_3855 and MBPNEE by comparing them with other NRPS-like enzyme, thus trying to understand the product release mechanism.

TE domains belong to the alpha/beta hydrolase superfamily including esterases, lipases and proteases which was seen to play a significant role in chemoenzymatic synthesis of novel analogues, especially related to the regioselective and stereoselective peptide cyclization[160–162]. Moreover, TE domains have been revealed to share a conserved catalytic triad comprising of serine, aspartic acid and histidine residues based on recent structural characterization of NRPS modules[163–165]. The release mechanism of the full-length product assembled by the NRPS requires various steps.

The final product is primarily transferred from the last PCP to the hydroxyl group of the conserved serine residue in the TE domain active site[166] forming a peptidyl ester. The conserved aspartic acid residue stabilizes the histidine residue which in turn accepts the proton from the hydroxyl of the serine residue. On the other hand, hydroxyl oxygen of the serine residue acts as a nucleophile and attacks the carbonyl carbon of the peptidyl thioester tethered to the PCP giving rise to a peptidyl-O-Serine oxoester on the TE domain. An oxyanion hole possibly coming from the TE domain backbone might contribute to the stability of the negative charge on the thioester from the nucleophilic attack. There are two possibilities of the final release of the TE-bound peptidyl-O-Serine intermediate.

The first possibility being hydrolysis by water resulting in a linear peptide whereas the second option is the intramolecular attack of an amine (arising from the N-terminus or side chain of an amino acid) or a hydroxyl (arising from the side chain of an amino acid) leading to the production of a cyclic peptide[167].

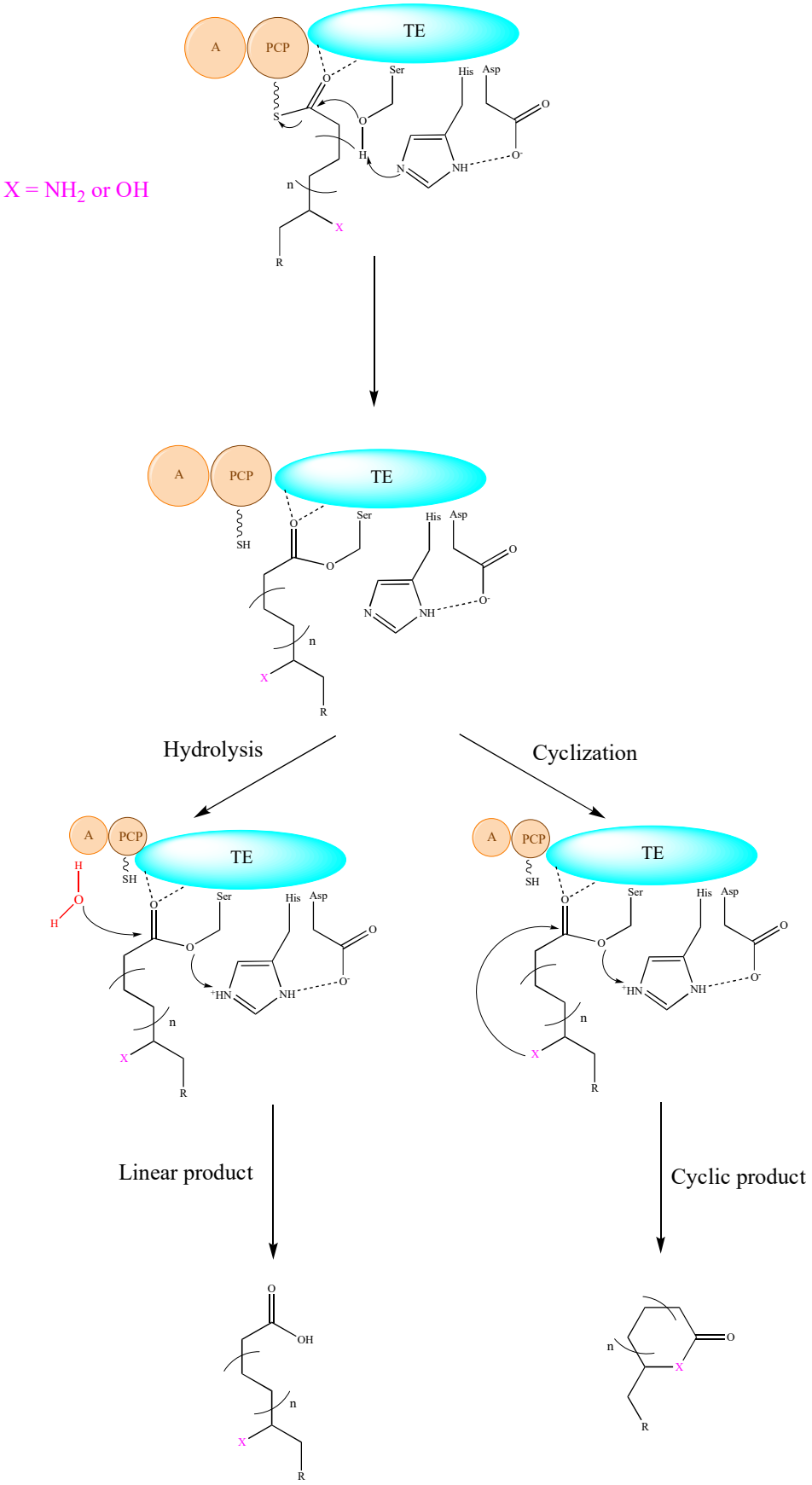
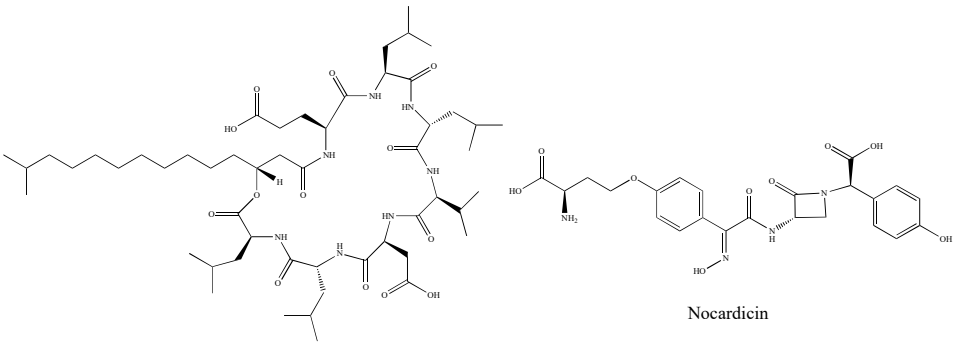


Figure 6.1: General mechanism of TE-catalysed NRP release. The oxyanion hole stabilizes the carbonyl oxygen of the acyl-S-PCP. Two possibilities of product release

namely hydrolysis or cyclization can occur giving rise to a linear product and a cyclic product respectively.

Based on literature, TE domains have also been unveiled to specific type of activities namely formation of cyclic peptides by dimerization or trimerization of peptide monomers[168,169], intramolecular transesterification[170] as well as cyclization along with macrothiolactonization[171], suggesting that they might contribute to supplementary structural diversity within the final peptide. As such, the TE domain of the NRPS-like enzymes Ava_3855 and BJP34_RS17220 were aligned against other characterized TE domains from surfactin (Srf), antibiotic fengycin (Fen) and nocardicin NRPS biosynthetic clusters.



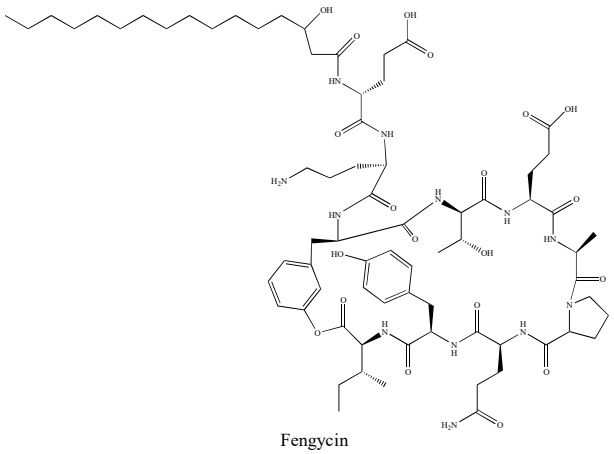


Figure 6.2: Structures of Surfactin, Nocardicin and Fengycin.

Based on previous studies, the TE domain of surfactin, from *Bacillus subtilis*, was shown to catalyze the cyclization and release of the lipoheptapeptide whereas the catalysis of the regioselective and stereoselective macrocyclization was performed by the TE domain of fengycin, from *Bacillus subtilis* F29-3[172,173].

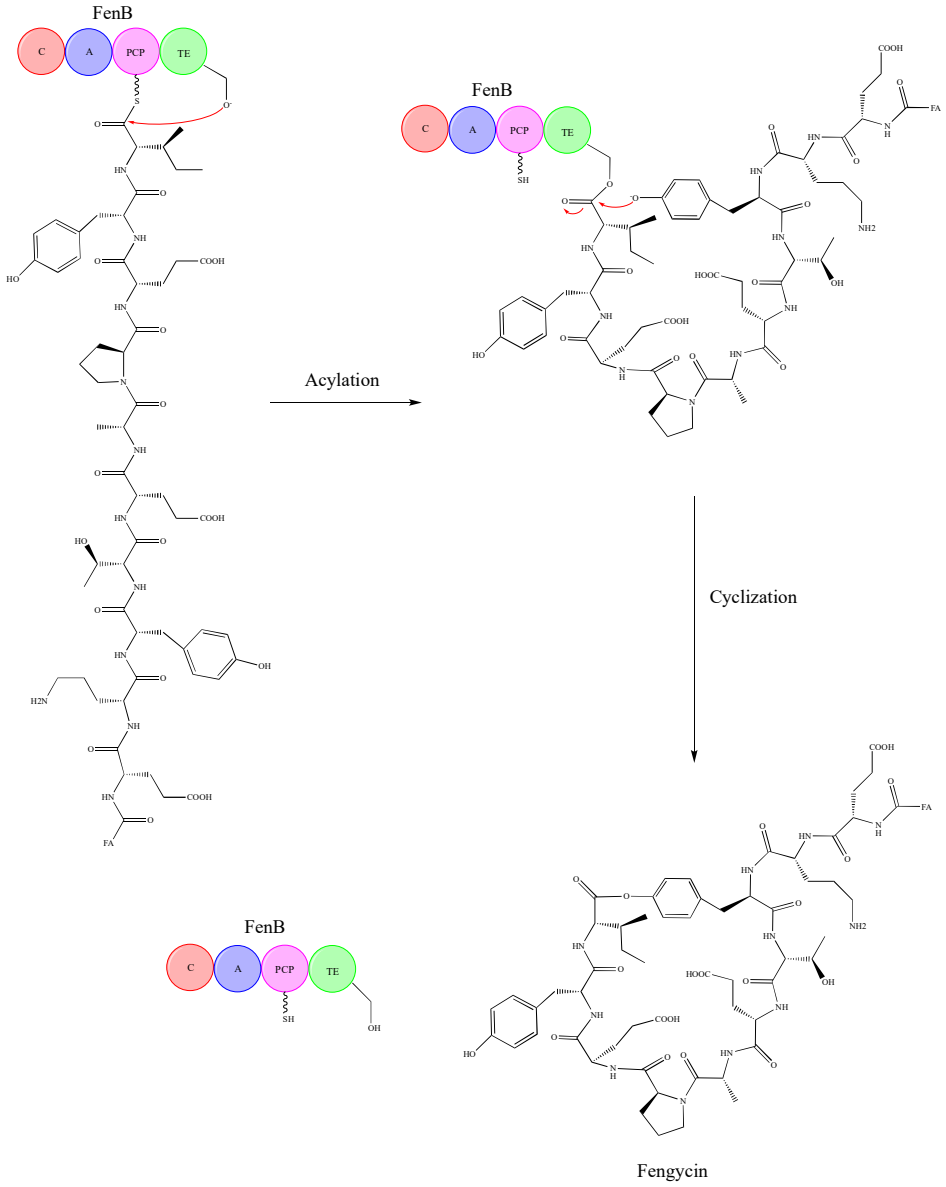


Figure 6.3: Fengycin release from FenB module. The active serine residue on the TE domain forms an oxo-ester bond with the L-Ile11 which in turn gets attacked by the hydroxylate group of L-Try4 which act as a nucleophile, thus leading to cyclization and release of Fengycin. [FA = Fatty Acid] [Adapted from Samel *et al.*,[173]].

Five fengycin synthetases (FenA-E) are arranged in a particular order namely FenC (287 kDa), FenD (291 kDa), FenE (287 kDa), FenA (407 kDa) and FenB (144 kDa) with a total of ten modules, each activating and integrating one specific amino acid. The last module, FenB, consists of the full length fengycin peptide tethered to the PCP which get transferred to the active serine residue of the TE domain forming an oxo-

ester bond between L-Ile11 and the serine residue active site. The hydroxylate group of the L-Tyr4 acts as a nucleophile by attacking the oxo-ester bond followed by a lactone formation, ultimately releasing the final product fengycin[173].

It is proposed that the electrophilicity and nucleophilicity of amino acid residues of the peptidyl substrate plays an important role in the determination of the substrate specificity of the NRPS TE domains. Structural information also suggests that the PCP and TE are usually at an appropriate distance to allow domain interactions and flow of communications in NRPS[174,175]. The less common type II TE domains, also known as *trans*-acting TE, are predominantly renowned to have an editing function in biosynthesis where atypical intermediates jammed on NRPS due to biosynthetic mistakes are eradicated. For instance, type II TE possess an accessible active site which delivers an essential rectification in the non-ribosomal peptide biosynthesis[176,177]. Furthermore, type II TE domains are specific to the hydrolysis mechanism giving rise to linear product which are typically tailored after release from the assembly lines[178].

6.2 TE domain comparison with other NRPS-like biosynthetic amino acid sequence

A TE domain alignment of Ava_3855 TE domain and BJP34_RS17220 TE domain against other characterized TE domains from surfactin, fengycin and nocardicin NRPS biosynthetic clusters was done as shown in the Figure below.

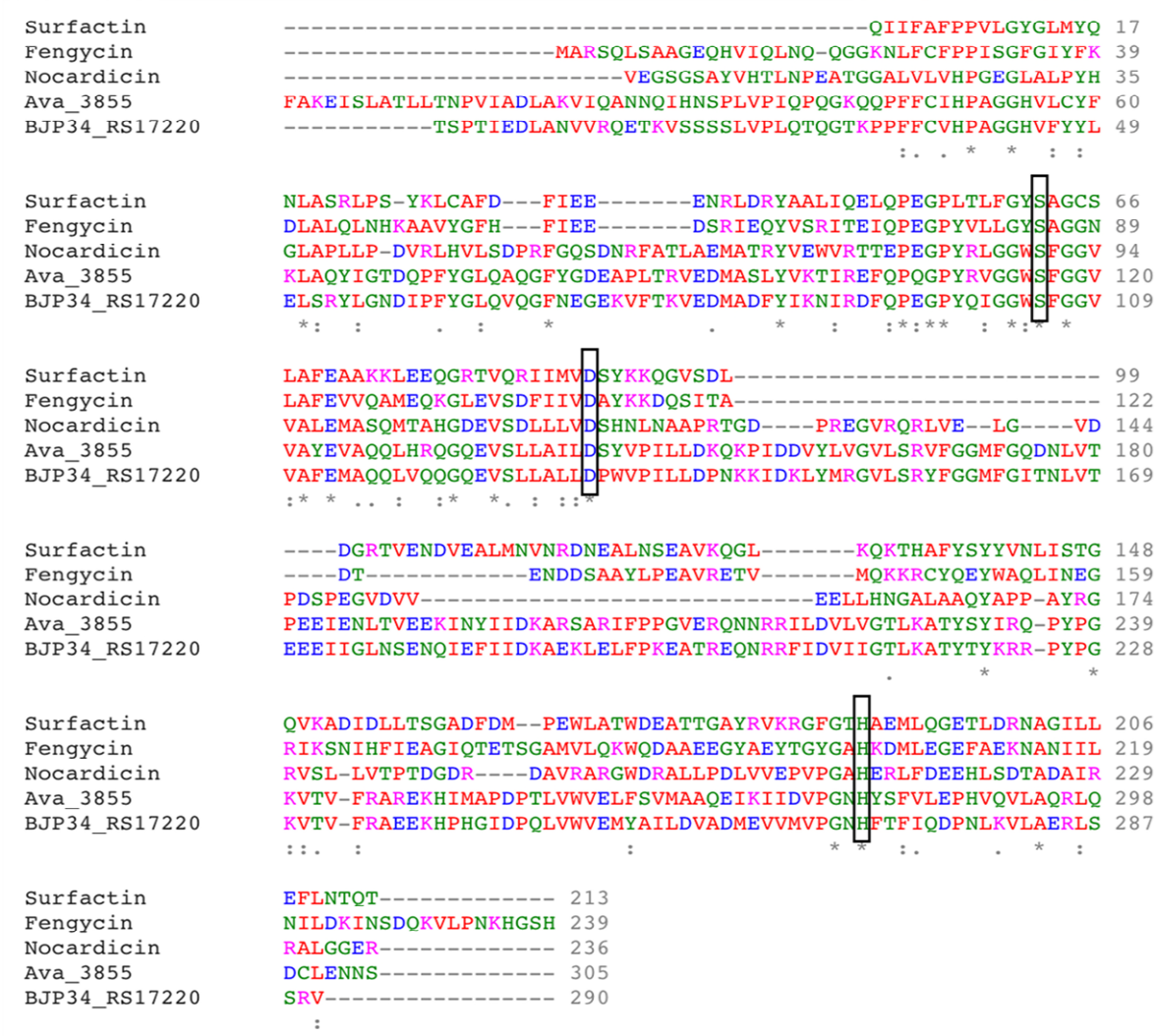
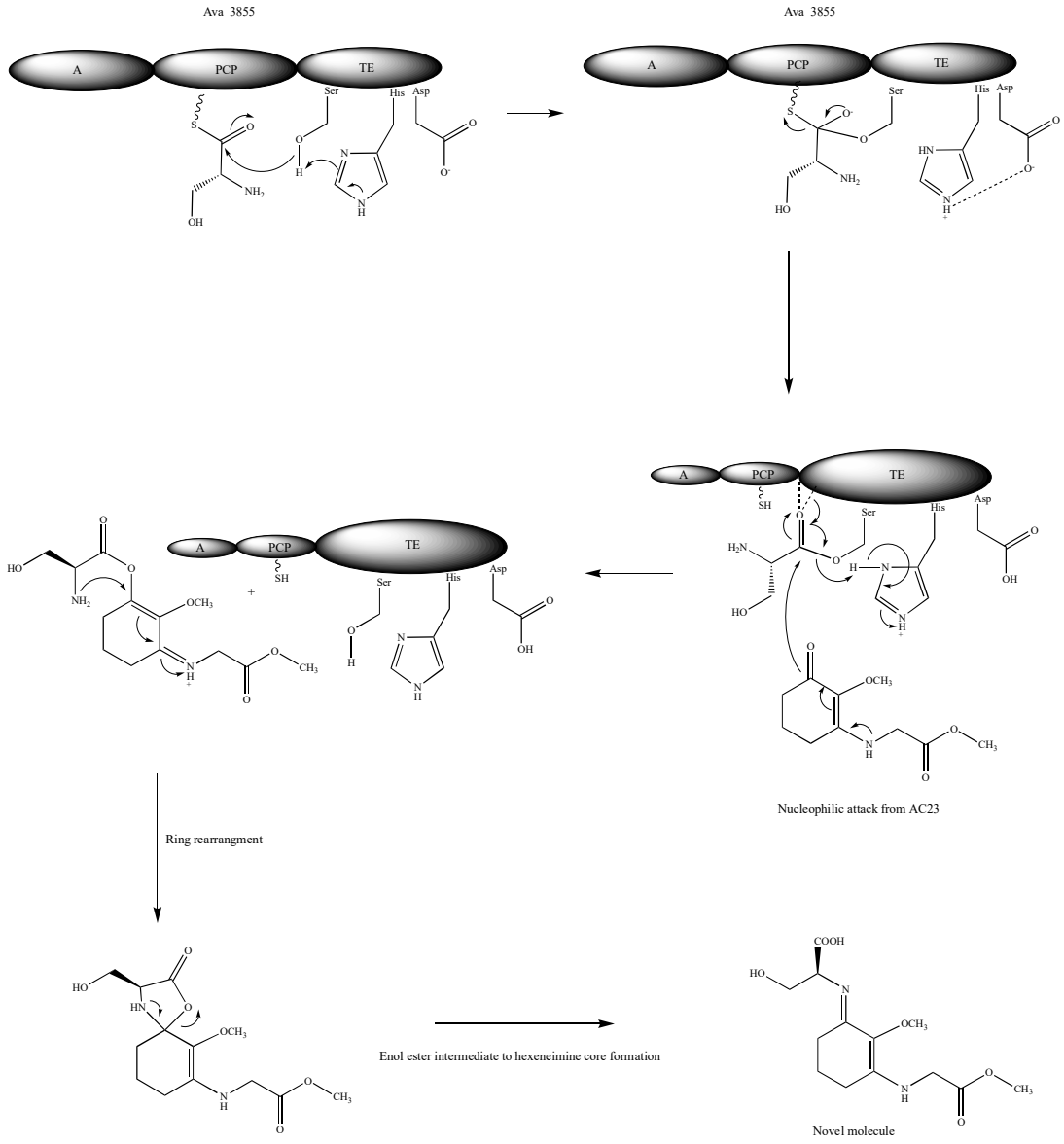


Figure 6.4: TE domain alignment of Ava_3855 TE domain and BJP34_RS17220 TE domain against other characterized TE domains from surfactin, fengycin and nocardicin NRPS biosynthetic clusters. The residues of the catalytic triad namely Serine, Aspartic acid and Histidine are conserved among the various TE domains as shown in the black box.

The presence of the conserved catalytic triad namely serine, aspartic acid and histidine was revealed in the TE domain of both Ava_3855 and BJP34_RS17220, suggesting that the TE domains of the latter might play an important role for the generation of the final product as discussed in the next section.

6.3 Proposed role of [A] Ava_3855 TE domain and [B] MBPNEE TE domain during novel compound (with AC23) final assembly



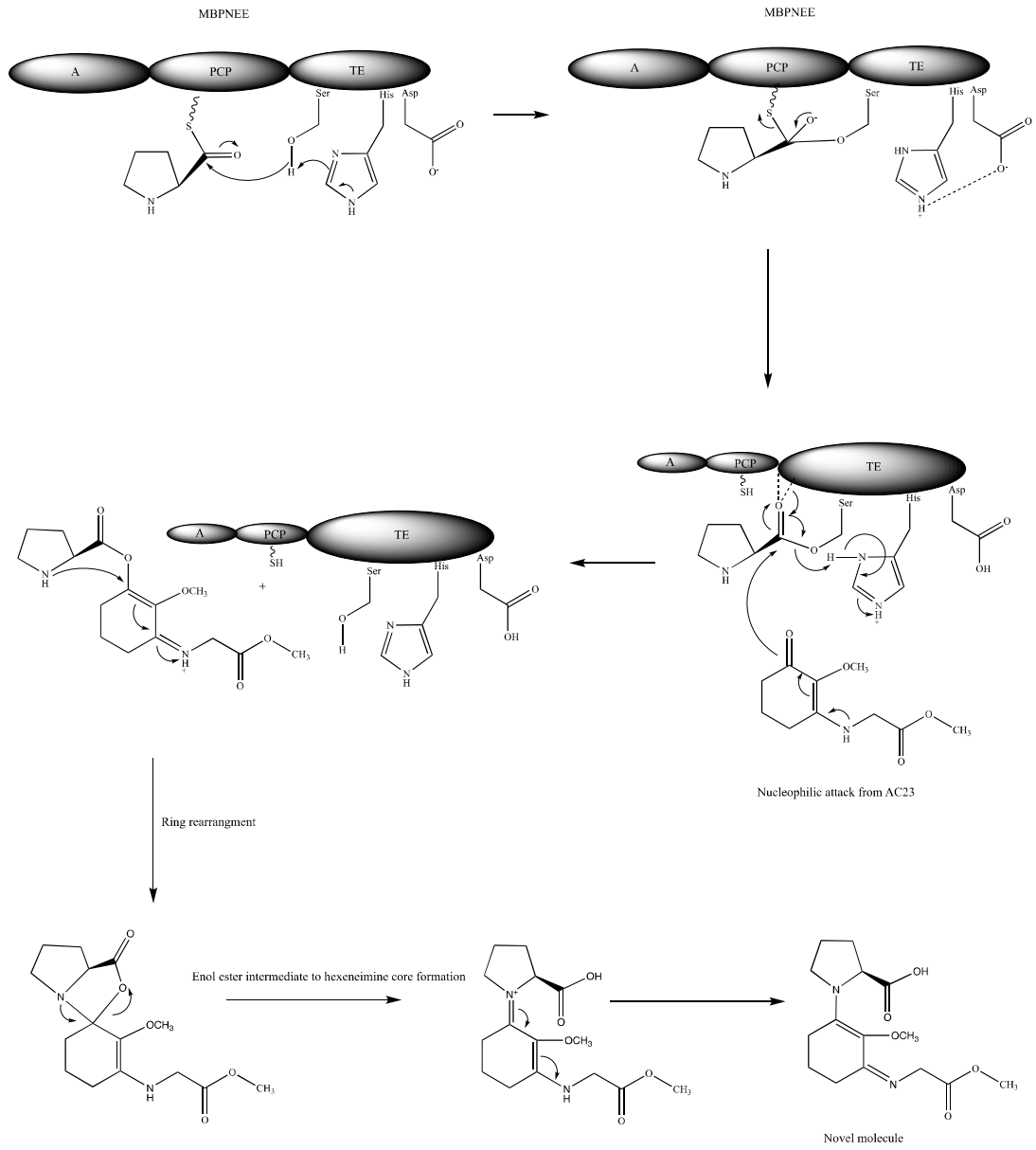


Figure 6.5: A: Proposed release mechanism involving Ava_3855 TE domain. The synthetic precursor AC23 plays the role of a strong nucleophile attacking the substrate serine giving rise to a novel MAA after ring rearrangement. B: Plausible release mechanism involving NEE TE domain. The synthetic precursor AC23 plays the role of a strong nucleophile attacking the substrate proline giving rise to a novel MAA after ring rearrangement.

6.4 Discussion

The presence of the TE domain in Ava_3855 and MBPNEE might play an important role in the release of the product. The catalytic triad namely serine, histidine and

aspartic acid are present in both enzymes' TE domains suggesting that it might be involved in the whole mechanism as shown above.

The lone pair of electrons on the nitrogen atom of the histidine molecule attacks the -OH group of the serine molecule in the TE domain which in turn attacks the carbonyl group of the activated serine substrate in Ava_3855 and proline substrate in NEE linked via a thioester bond on the PCP respectively. The thioester bond is broken leading to the transfer of the substrate serine and the substrate proline in Ava_3855 and NEE respectively to form a bond with the amino acid serine from their respective domain where the formation of an oxyanion hole coming from the TE domain arise, resulting in the stabilisation of the chosen substrate. The lone pair of electrons on the nitrogen atom of the synthetic molecule AC23 moves towards the carbonyl which in turn attacks the -COO of the loaded substrate (serine for Ava_3855 and proline for NEE) tethered to the serine of the TE domain. The molecule AC23 thus forms a bond with the corresponding carbonyl group of the respective substrate (serine for Ava_3855 and proline for NEE), leading to its release from the TE domain and the formation of an enol ester intermediate. The amine group from the corresponding substrate forms a bond with the carbon atom of the cyclohexanone backbone, subsequently rearranging the position of the substrate, leading to the generation of a cyclohexeneimine backbone having serine and proline respectively tethered, hence forming a novel molecule.

As seen in this chapter, the TE domain is an essential part in the NRPS system for product release either by hydrolysis or macrocyclization. A previous study successfully revealed that shifting of the TE domain forward in a NRPS system can generate new truncated cyclic or linear peptides based on the predicted sequence. For instance, the antifungal lipopeptide, plipastatin, is assembled by five NRPS genes namely *ppsA*, *ppsB*, *ppsC*, *ppsD* and *ppsE* respectively. The TE domain of plipastatin synthetase was moved forward and the truncated/ novel peptides were evaluated in order to have an insight of the catalytic activity and selectivity of the TE domain in the engineered enzymes. A key result during this study was the significance of the PCP-TE linker which became obsolete when different linkers were used to connect the TE domain with the PCP domain. Thus, the native linker in the plipastatin NRSP system were kept in all the hybrid enzymes which showed its importance through generation of the predicted cyclic or linear lipopeptides via LC-MS. It could also be

assumed that apart from tethering the PCP and TE domain, the inter-domain linker might also play a role in ensuring specificity as well as communication between domains. The TE domain in the engineered enzymes gave rise to linear products by using water as a strong nucleophile for product release. The plipastatin TE domain, on the contrary, could both catalyse the peptide hydrolysis as well as catalyse the product cyclization revealing the fact that the TE domain had the ability to distinguish the last amino acid to allow the cyclization process to occur[179]. Henceforth, shifting the TE domain along NRPS systems could be an alternative to generate an assortment of new analogues with more potent bioactivity.

Chapter 7: General Conclusion and Perspective

The sun has an assortment of benefits including its contribution for photosynthesis to occur where solar energy is converted into chemical energy by photosynthetic organisms or the metabolism of vitamin D regulation in humans. However, it can also be stated that solar radiation can be deleterious on various organisms causing DNA damage, sunburn, premature ageing, or diverse types of skin cancers to name a few[180]. Throughout evolution, numerous photoprotective mechanisms have been developed in nature to cope with the disastrous effects of those harmful UV radiation, namely UVA which absorbs between 320-400 nm and UVB which absorbs between 280-320 nm. Even though less than 7% of UVA and less than 1% of. UVB reach the Earth surface[181], the consequence can be devastating as said above. As such, organisms exposed to harmful UV radiation have come up with specific strategies to counteract the detrimental effects by developing defensive mechanisms, one among which is the production of photoprotective natural products which have the ability to act as a sunshield against damage as well as preventing formation of reactive oxygen species after UV irradiation[182]. Humans contain melanin, a class of compound found in the epidermis, which can be categorized into two chemically distinctive types namely the brown-black eumelanins and the red-yellow phaeomelanins which was seen to contribute to pigmentation as well as photoprotection[183]. On the other hand, higher plants were seen to produce phenylpropanoids as well as flavonoids which could aid in photoprotection.

Some autotrophs like cyanobacteria[184,185] as well as macro-algae[186] and heterotrophs like fungi[88], sea urchins, crustaceans[187] and fishes[188] have developed coping mechanisms against chronic UVR exposure by synthesizing small, colourless, polar UV-screening secondary metabolites known as MAAs[139]. Based on previous literature, it was also found that animals could acquire MAAs through their food or through their symbiotic relationship with their hosts, hence accumulating them[189,190].

MAAs are known to have high molar extinction coefficients (up to 50 000 M^{-1} cm^{-1}), thus making them strong candidates as UV absorbers which in turn lead to photoprotection[191].

As shown in Chapter 3, the MAAs content and composition varied in the different samples used. It was noted that the MAAs shinorine, porphyra-334, palythine were all present in the various *Palmaria palmata* samples, no matter what supplier/ sources

they came from. However, the cis-trans isomers usujirene/palythene were not as consistent as the other MAAs. A plausible explanation could be environmental factors like temperature, UVR exposure, cloud cover, salinity, dissolved oxygen, or nutrients affecting the production of those MAAs. Based on a previous study, a higher MAA content was found in a slightly alkaline environment, with significant UVR exposure and nutrients like nitrates and phosphates along with the salinity of the water[120]. The spectral variability and intensity[121,122]; spatial and temporal variability as well as seasonal fluctuations and depth[123] are all factors that could affect the level of MAAs produced in seaweeds. The extraction procedure of MAAs from seaweeds is another important factor to consider in order to optimize the yield and to also select a cheap and eco-friendly extraction protocol. The chosen extraction solvent during this study was 25% methanol in water which seemed to allow successful MAAs extraction followed by purification and identification by High-Performance Liquid Chromatography; Ultra High-Performance Liquid Chromatography – High Resolution Mass Spectrometry and ^{1}H Nuclear Magnetic Resonance spectroscopy respectively.

Moreover, the energy flow and dynamics of shinorine, porphyra-334 and usujirene/palythene were followed after photoexcitation to be able to understand their photoprotective mechanism for the first time. The overall deduction was the fact that MAAs have a quick relaxation time, meaning that, the absorbed energy was seen to be converted into harmless heat within picoseconds compared to the common UVA filter, avobenzone. Ultimately, those MAAs were unveiled as photostable after several hours of irradiation which make them ideal candidates to be further developed in pharmaceutical/ cosmetic industries.

The biosynthesis of MAAs have been a debate for quite some time now implying that MAAs arise from the pentose phosphate pathway[77,192,193] while other studies suggested that the shikimate pathway was the source[192,194]. In both pathways, it was seen that 4-deoxygadusol was the core structure of MAAs which was followed by the addition of glycine producing the mono-substituted cyclohexanone mycosporine-glycine intermediate. A second amino acid integration, such as serine gave rise to the cyclohexenimine shinorine while the addition of threonine generated porphyra-334[77,195].

There are four enzymes involved in shinorine biosynthesis in the terrestrial cyanobacteria *Anabaena variabilis* with the last enzyme being the NRPS-like enzyme Ava_3855 involved in attaching serine to the mycosporine-glycine intermediate by utilising ATP[77]. The NRPS-like enzyme was the focus of chapter 4 where Ava_3855 was successfully cloned in pET151 and expressed in *E. coli* BL21 followed by purification using Nickel Sepharose beads. The purified Ava_3855 was successfully confirmed by both proteomics and intact protein mass spectrometry which also confirmed the activation and loading of the substrate L-serine demonstrating its functionality. The NRPS-like enzyme Ava_3855 was subsequently engineered by swapping its A-domain by the A-domain of BJP34_RS17220 which was predicted to activate L-proline as substrate. The recombinant gene was successfully cloned in pET28MBP and expressed in *E. coli* BL21 followed by purification using Nickel Sepharose beads leading to the engineered enzyme MBPNEE.

During this study, the intermediate mycosporine-glycine could not be obtained in adequate amount for further *in vitro* experiments. Therefore, a synthetic analogue namely AC23 (prepared by Adam Cowden, PhD student under the supervision of Professor Martin Wills in the Chemistry department, Warwick University) was used during the chemo-enzymatic assay with the wild type (Ava_3855) as well as engineered enzyme (MBPNEE). Unfortunately, the predicted product could not be detected by LC-MS due to different possibilities related to the synthetic analogue might not be an appropriate intermediate to be used in this context or it might be the fact that the expected product was unstable and degraded as soon as it was formed. Another possibility was the lack of the hydroxyl and hydroxymethyl groups which were needed to allow proper positioning and proximity of the enzyme and substrate for the nucleophilic attack to occur.

However, a novel compound with chemical formula $C_{12}H_{19}NO_6$, *m/z* $[M+H]^+$ 274.1284 was detected in the control (boiled Ava_3855); whole reaction without ATP as well as in the complete reaction mix. It could be suggested that demethylated AC23 reacted with glycerol (from the protein storage buffer) chemically, without the involvement of the respective enzyme. The second possibility was the displacement of L- serine (when using Ava_3855) and L-proline (when using MBPNEE) after the respective substrate were activated, loaded on the PCP and nucleophilically attacked by demethylated AC23 to form an enol ester intermediate by glycerol.

Further studies should definitely be carried out involving the *in vitro* assay with the native intermediate namely mycosporine-glycine. Moreover, intact protein mass spectrometry of the engineered enzyme MBPNEE could be done to confirm the loading of L-proline on the PCP. Alpha fold can be used to track the specific differences in the A-domain of Ava_3855 and MBPNEE, thus analysing each respective A-domain binding pocket with the various amino acid residues involved in the substrate selection.

In addition, the di-domain (A-domain and PCP) only could be expressed and purified to exactly understand the role/ importance of the TE domain in both Ava_3855 and MBPNEE. By cutting the TE domain off, and carry out the full enzymatic assay with the native intermediate mycosporine-glycine, clearer understanding of the individual domain as well as the release mechanism could be achieved. Per se, biosynthetic pathways engineering, more precisely, mega synthetases like the NRPS-like enzymes engineering can lead to generation of novel, eco-friendly, affordable, photostable, nature inspired photoprotective natural products with desired properties for pharmaceutical industries as well as for benefiting human in counteracting the harmful effects of UV radiation.

References

1. Ammerman, J., Fuhrman, J., Hagström, Å. & Azam, F. Bacterioplankton growth in seawater: I. Growth kinetics and cellular characteristics in seawater cultures. *Mar. Ecol. Prog. Ser.* **18**, 31–39 (1984).
2. Kang, H. K., Seo, C. H. & Park, Y. Marine peptides and their anti-infective activities. *Mar. Drugs* **13**, 618–654 (2015).
3. Bilal, M. *et al.* Marine-Derived Biologically Active Compounds for the Potential Treatment of Rheumatoid Arthritis. *Mar. Drugs* **19**, (2020).
4. Vo, T. S. & Kim, S. K. Potential anti-HIV agents from marine resources: An overview. *Mar. Drugs* **8**, 2871–2892 (2010).
5. Aneiros, A. & Garateix, A. Bioactive peptides from marine sources: Pharmacological properties and isolation procedures. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **803**, 41–53 (2004).
6. Martin, L. P. *et al.* Phase II study of weekly PM00104 (ZALYPSIS®) in patients with pretreated advanced/metastatic endometrial or cervical cancer. *Med. Oncol.* **30**, (2013).
7. Tsoukalas, N. *et al.* Complete remission of a reccurrent mesenteric liposarcoma with rare histological features following the administration of trabectedin. *Oncol. Lett.* **7**, 47–49 (2014).
8. Bhatnagar, I. & Kim, S. K. Immense essence of excellence: Marine microbial bioactive compounds. *Mar. Drugs* **8**, 2673–2701 (2010).
9. Häder, D. P. *et al.* Effects of UV radiation on aquatic ecosystems and interactions with other environmental factors. *Photochem. Photobiol. Sci.* **14**, 108–126 (2015).
10. Banaszak, A. T. & Lesser, M. P. Effects of solar ultraviolet radiation on coral reef organisms. *Photochem. Photobiol. Sci.* **8**, 1276–1294 (2009).
11. Lesser, M. P. The effects of UV radiation in the marine environment. *Limnol. Oceanogr.* **46**, 211–211 (2001).
12. Panel, A. EPA Public Access. (2019) doi:10.1039/c7pp90043k.Environmental.
13. WHO. World Health Organization. Global Solar UV Index: A Pratical Guide. 28 (2002).
14. Roby, L., Tony, M., Wayne, S. & Bruce, A. Solar Ultraviolet Radiation:

Global burden of disease from solar ultraviolet radiation. *World Health* **55**, 987–999 (2006).

15. English, D. & Armstrong, B. Case-control study of cutaneous malignant melanoma. *Br. Med. J. (Clin. Res. Ed).* **296**, 1799 (1988).
16. Stockfleth, E., Rosen, T. & Shumack, S. Managing skin cancer. *Manag. Ski. Cancer* 1–226 (2010) doi:10.1007/978-3-540-79347-2.
17. Wehner, M. R. *et al.* HHS Public Access. **78**, 663–672 (2019).
18. Bang, K. M., Halder, R. M., White, J. E., Sampson, C. C. & Wilson, J. Skin cancer in black Americans: A review of 126 cases. *J. Natl. Med. Assoc.* **79**, 51–58 (1987).
19. Griffiths, C. E. M. Dowling Oration delivered at the Royal College of Physicians, London, Friday 5 June 1998. Retinoids: Renaissance and reformation. *Clin. Exp. Dermatol.* **24**, 329–335 (1999).
20. Merten, J. W., King, J. L., Walsh-Childers, K., Vilaro, M. J. & Pomeranz, J. L. Skin cancer risk and other health risk behaviors: A scoping review. *Am. J. Lifestyle Med.* **11**, 182–196 (2017).
21. Bourke, J. F. & Graham-Brown, R. A. C. Protection of children against sunburn: A survey of parental practice in Leicester. *Br. J. Dermatol.* **133**, 264–266 (1995).
22. Vile, G. F. & Tyrrell, R. M. Uva radiation-induced oxidative damage to lipids and proteins in vitro and in human skin fibroblasts is dependent on iron and singlet oxygen. *Free Radic. Biol. Med.* **18**, 721–730 (1995).
23. Sinha, R. P., Singh, S. P. & Häder, D. P. Database on mycosporines and mycosporine-like amino acids (MAAs) in fungi, cyanobacteria, macroalgae, phytoplankton and animals. *J. Photochem. Photobiol. B Biol.* **89**, 29–35 (2007).
24. Klisch, M. Mycosporine-Like Amino Acids and Marine Toxins - The Common and the Different. *Mar. Drugs* **6**, 147–163 (2008).
25. Carreto, J. I., Carignan, M. O., Daleo, G. & Marco, S. G. D. Occurrence of mycosporine-like amino acids in the red-tide dinoflagellate Alexandrium excavatum: UV-photoprotective compounds? *J. Plankton Res.* **12**, 909–921 (1990).
26. Karentz, McEuen, Land & Dunlap, W. C. o • N. **166**, 157–166 (1991).
27. Siezen, R. J. Microbial sunscreens. *Microb. Biotechnol.* **4**, 1–7 (2011).

28. Riegger, L. & Robinson, D. Photoinduction of UV-absorbing compounds in Antarctic diatoms and Phaeocystis antarctica. *Mar. Ecol. Prog. Ser.* **160**, 13–25 (1997).
29. Carreto, J. I. & Carignan, M. O. Mycosporine-like amino acids: Relevant secondary metabolites. chemical and ecological aspects. *Mar. Drugs* **9**, 387–446 (2011).
30. Cheewinthamrongrod, V., Kageyama, H., Palaga, T., Takabe, T. & Waditee-Sirisattha, R. DNA damage protecting and free radical scavenging properties of mycosporine-2-glycine from the Dead Sea cyanobacterium in A375 human melanoma cell lines. *J. Photochem. Photobiol. B Biol.* **164**, 289–295 (2016).
31. Matsui, K. *et al.* Novel glycosylated mycosporine-like amino acids with radical scavenging activity from the cyanobacterium Nostoc commune. *J. Photochem. Photobiol. B Biol.* **105**, 81–89 (2011).
32. Wittenberg, B. Y. J. B. The Source of Carbon Monoxide in the Float of the Portuguese Man-of-War, Physalia Physalis L. *J. Exp. Biol.* **37**, 698–705 (1960).
33. Dunlap, W. C. & Shick, J. M. Ultraviolet radiation-absorbing mycosporine-like amino acids in coral reef organisms: A biochemical and environmental perspective. *J. Phycol.* **34**, 418–430 (1998).
34. Řezanka, T., Temina, M., Tolstikov, A. G. & Dembitsky, V. M. Natural microbial UV radiation filters — Mycosporine-like amino acids. *Folia Microbiol. (Praha).* **49**, 339–352 (2004).
35. Bhatia, S. *et al.* Broad-spectrum sun-protective action of Porphyra-334 derived from Porphyra vietnamensis. *Pharmacognosy Res.* **2**, 45–49 (2010).
36. Garcia-Pichel, F. & Castenholz, R. W. Occurrence of UV-absorbing, mycosporine-like compounds among cyanobacterial isolates and an estimate of their screening capacity. *Appl. Environ. Microbiol.* **59**, 163–169 (1993).
37. Karsten, U. & Garcia-Pichel, F. Carotenoids and mycosporine-like amino acid compounds in members of the Genus microcoleus (Cyanobacteria): A chemosystematic study. *Syst. Appl. Microbiol.* **19**, 285–294 (1996).
38. Dunlap, W. C. Sunscreens, oxidative stress and antioxidant functions in marine organisms of the Great Barrier Reef. *Redox Rep.* **4**, 304–306 (1999).
39. Karsten, U., Sawall, T., West, J. & Wiencke, C. Ultraviolet sunscreen compounds in epiphytic red algae from mangroves. *Hydrobiologia* **432**, 159–

171 (2000).

40. García, P. E., Diéguez, M. C., Ferraro, M. A., Zagarese, H. E. & Pérez, A. P. Mycosporine-like amino acids in freshwater copepods: Potential sources and some factors that affect their bioaccumulation. *Photochem. Photobiol.* **86**, 353–359 (2010).
41. Volkmann, M., Gorbushina, A. A., Kedar, L. & Oren, A. Structure of euhalothece-362, a novel red-shifted mycosporine-like amino acid, from a halophilic cyanobacterium (Euhalothecesp.). *FEMS Microbiol. Lett.* **258**, 50–54 (2006).
42. Tan, K. R. & Epton, H. A. S. Ultraviolet-absorbing compounds associated with sporulation in Botrytis cinerea. *Trans. Br. Mycol. Soc.* **63**, 157–167 (1974).
43. Tartarotti, B. & Sommaruga, R. The effect of different methanol concentrations and temperatures on the extraction of mycosporine-like amino acids (MAAs) in algae and zooplankton. *Archiv fur Hydrobiologie* vol. 154 691–703 (2002).
44. Dunlap, W. C., Williams, D. M. B., Chalker, B. E. & Banaszak, A. T. Biochemical photoadaptation in vision: U.V.-absorbing pigments in fish eye tissues. *Comp. Biochem. Physiol. -- Part B Biochem.* **93**, 601–607 (1989).
45. Hideshi, N., Junichi, K. & Yoshimasa, H. ‘,* ?+yH u&$&‘~. *SubStance* **334**, 3341.
46. Volkmann, M. & Gorbushina, A. A. A broadly applicable method for extraction and characterization of mycosporines and mycosporine-like amino acids of terrestrial, marine and freshwater origin. *FEMS Microbiol. Lett.* **255**, 286–295 (2006).
47. Dunlap, W. C. & Chalker, B. E. Identification and quantitation of near-UV absorbing compounds (S-320) in a hermatypic scleractinian. *Coral Reefs* **5**, 155–159 (1986).
48. Stochaj, W. R., Dunlap, W. C. & Shick, J. M. Marine ,, = : = Biology. **156**, 149–156 (1994).
49. Bohm, G. A., Pfleiderer, W., Boger, P. & Scherer, S. Structure of a novel oligosaccharide-mycosporine-amino acid ultraviolet A/B sunscreen pigment from the terrestrial cyanobacterium Nostoc commune. *Journal of Biological Chemistry* vol. 270 8536–8539 (1995).

50. Carreto, J. I., Carignan, M. O. & Montoya, N. G. A high-resolution reverse-phase liquid chromatography method for the analysis of mycosporine-like amino acids (MAAs) in marine organisms. *Mar. Biol.* **146**, 237–252 (2005).
51. Cole, K. M. & G, S. R. Book Reviews: Book Reviews. *Blood* **41**, 738 (1973).
52. Baweja, P., Kumar, S., Sahoo, D. & Levine, I. *Biology of Seaweeds*. *Seaweed in Health and Disease Prevention* (Elsevier Inc., 2016). doi:10.1016/B978-0-12-802772-1.00003-8.
53. Pereira, R. & Vasconcelos, M. Chemical defense in the red seaweed Plocamium brasiliense: spatial variability and differential action on herbivores. *Brazilian J. Biol.* **74**, 545–552 (2014).
54. Delaney, A., Frangoudes, K. & Ii, S. A. *Society and Seaweed: Understanding the Past and Present*. *Seaweed in Health and Disease Prevention* vol. 2 (2016).
55. Philpott, J. & Bradford, M. Seaweed: Nature's Secret for a Long and Healthy Life? *Nutr. Pract. Winter* 1–21 (2006).
56. Devi, G. K., Manivannan, K., Thirumaran, G., Rajathi, F. A. A. & Anantharaman, P. In vitro antioxidant activities of selected seaweeds from Southeast coast of India. *Asian Pac. J. Trop. Med.* **4**, 205–211 (2011).
57. Susanto, E., Fahmi, A. S., Abe, M., Hosokawa, M. & Miyashita, K. Lipids, Fatty Acids, and Fucoxanthin Content from Temperate and Tropical Brown Seaweeds. *Aquat. Procedia* **7**, 66–75 (2016).
58. Mohamed, S., Hashim, S. N. & Rahman, H. A. Seaweeds: A sustainable functional food for complementary and alternative therapy. *Trends Food Sci. Technol.* **23**, 83–96 (2012).
59. Albertus, S. Medicinal and pharmaceutical uses of seaweed natural products: A review. *Probl. Peredachi Informatsii* **40**, 50–62 (2004).
60. Pal, A., Kamthania, M. C. & Kumar, A. Bioactive Compounds and Properties of Seaweeds—A Review. *OALib* **01**, 1–17 (2014).
61. Grote, B. Recent developments in aquaculture of Palmaria palmata (Linnaeus) (Weber & Mohr 1805): cultivation and uses. *Rev. Aquac.* **11**, 25–41 (2019).
62. Provan, J., Wattier, R. A. & Maggs, C. A. Phylogeographic analysis of the red seaweed Palmaria palmata reveals a Pleistocene marine glacial refugium in the English Channel. *Mol. Ecol.* **14**, 793–803 (2005).
63. Mouritsen, O. G. *et al.* On the human consumption of the red seaweed dulse

(Palmaria palmata (L.) Weber & Mohr). *J. Appl. Phycol.* **25**, 1777–1791 (2013).

64. Werner, A. & Dring, M. Aquaculture Explained Principal authors Aquaculture Explained Cultivating Palmaria palmata. *Irish Sea Fish. Board* 85 (2011).
65. Edwards, M. D. & Dring, M. J. Open-sea cultivation trial of the red alga, Palmaria palmata from seeded tetraspores in Strangford Lough, Northern Ireland. *Aquaculture* **317**, 203–209 (2011).
66. Faes, V. A. & Viejo, R. M. Structure and dynamics of a population of Palmaria palmata (Rhodophyta) in northern Spain. *J. Phycol.* **39**, 1038–1049 (2003).
67. Pang, S. J. & Lüning, K. Tank cultivation of the red alga Palmaria palmata: Year-round induction of tetrasporangia, tetraspore release in darkness and mass cultivation of vegetative thalli. *Aquaculture* **252**, 20–30 (2006).
68. J.R, S. & R.T, W. Sublittoral , Benthic Marine Algae of Southern Cape Cod and Adjacent Island : Seasonal Periodicity , Associations , Diversity , and Floristic Composition Author (s): James R . Sears and Robert T . Wilce Reviewed work (s): Published by : Ecological Soc. *America (NY).* **45**, 337–365 (2012).
69. González, R. *et al.* Anti-inflammatory activity of phycocyanin extract in acetic acid-induced colitis in rats. *Pharmacol. Res.* **39**, 55–59 (1999).
70. Remirez, D. *et al.* Effect of phycocyanin in zymosan-induced arthritis in mice - Phycocyanin as an antiarthritic compound. *Drug Dev. Res.* **48**, 70–75 (1999).
71. Wang, L. *et al.* Isolation, Purification and Properties of an R-Phycocyanin from the Phycobilisomes of a Marine Red Macroalga Polysiphonia urceolata. *PLoS One* **9**, e87833 (2014).
72. Karsten, U. & Wiencke, C. Factors controlling the formation of UV-absorbing mycosporine-like amino acids in the marine red alga Palmaria palmata from Spitsbergen (Norway). *J. Plant Physiol.* **155**, 407–415 (1999).
73. Bischof, K. *et al.* Acclimation of brown algal photosynthesis to ultraviolet radiation in Arctic coastal waters (Spitsbergen, Norway). *Polar Biol.* **20**, 388–395 (1998).
74. Karsten, U., Sawall, T. & Wiencke, C. A survey of the distribution of UV-absorbing substances in tropical macroalgae. *Phycol. Res.* **46**, 271–279

(1998).

75. Wood, W. F. Photoadaptive responses of the tropical red alga Eucheuma striatum Schmitz (Gigartinales) to ultra-violet radiation. *Aquat. Bot.* **33**, 41–51 (1989).
76. Teixeira, G. H. A. & O'Keefe, S. F. Short communication: Mycosporine-like amino acids protect natamycin against photodegradation in milk exposed to fluorescent or light-emitting diode light. *J. Dairy Sci.* **102**, 4972–4977 (2019).
77. Balskus, E. P. & Walsh, C. T. The genetic and molecular basis for sunscreen biosynthesis in cyanobacteria. *Science (80-.).* **329**, 1653–1656 (2010).
78. Shumpei, A., Pengfei, X., Corey, B., Patricia, F. & Taifo, M. Evolutionary divergence of Sedoheptulose 7- phosphate cyclases leads to several distinct cyclic products. *J Am Chem Soc.* **23**, 1–7 (2012).
79. Xiumei, W. *et al.* A Comparative Analysis of the Sugar Phosphate Cyclase Superfamily Involved in Primary and Secondary Metabolism. *Chembiochem.* **23**, 1–7 (2007).
80. Singh, S. P., Klisch, M., Sinha, R. P. & Häder, D. P. Genome mining of mycosporine-like amino acid (MAA) synthesizing and non-synthesizing cyanobacteria: A bioinformatics study. *Genomics* **95**, 120–128 (2010).
81. Gao, Q. & Garcia-Pichel, F. An ATP-Grasp ligase involved in the last biosynthetic step of the iminomycosporine shinorine in nostoc punctiforme ATCC 29133. *J. Bacteriol.* **193**, 5923–5928 (2011).
82. Waditee-Sirisattha, R., Kageyama, H., Sopun, W., Tanaka, Y. & Takabe, T. Identification and upregulation of biosynthetic genes required for accumulation of mycosporine-2-glycine under salt stress conditions in the halotolerant cyanobacterium Aphanothece halophytica. *Appl. Environ. Microbiol.* **80**, 1763–1769 (2014).
83. Miyamoto, K. T., Komatsu, M. & Ikeda, H. Discovery of gene cluster for mycosporine-like amino acid biosynthesis from Actinomycetales microorganisms and production of a novel mycosporine-like amino acid by heterologous expression. *Appl. Environ. Microbiol.* **80**, 5028–5036 (2014).
84. Garcia-Pichel, F., Wingard, C. E. & Castenholz, R. W. Evidence regarding the UV sunscreen role of a mycosporine-like compound in the cyanobacterium Gloeocapsa sp. *Appl. Environ. Microbiol.* **59**, 170–176 (1993).
85. Garcia-Pichel, F. A model for internal self-shading in planktonic organisms

and its implications for the usefulness of ultraviolet sunscreens. *Limnol. Oceanogr.* **39**, 1704–1717 (1994).

86. Favre-Bonvin J, Arpin N, B. C. Structure de la mycosporine (P310). *Can. J. Chem.* **54**, 1105–1113 (1976).
87. CM, L. Ultraviolet-absorbing substances associated with light-induced sporulation in fungi. *Can. J. Bot.* **43**, 185–200 (1965).
88. Libkind, D., Sommaruga, R., Zagarese, H. & Van Broock, M. Mycosporines in carotenogenic yeasts. *Syst. Appl. Microbiol.* **28**, 749–754 (2005).
89. Zalar, P., Sybren de Hoog, G., Schroers, H. J., Frank, J. M. & Gunde-Cimerman, N. Taxonomy and phylogeny of the xerophilic genus Wallemia (Wallemiomycetes and Wallemiales, cl. et ord. nov.). *Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol.* **87**, 311–328 (2005).
90. Bernillon J, Bouillant ML, Pittet JL, Favre-Bonvin J, A. N. Mycosporine glutamine and related mycosporines in the fungud Pyronema omphalodes. *Phytochemistry* **23**, 1083–1087 (1984).
91. Bandaranayake, W. M. Mycosporines : are they nature ' s sunscreens ? 159–172 (1997).
92. Gorbushina, A. A. *et al.* Black fungal colonies as units of survival: Hyphal mycosporines synthesized by rock-dwelling microcolonial fungi. *Can. J. Bot.* **81**, 131–138 (2003).
93. Jean Favre-Bonvin, Jacques Bernillon, Nadia Salin, N. A. Biosynthesis of mycosporines: Mycosporine glutaminol in Trichothecium roseum. *Phytochemistry* **26**, 2509–2514 (1987).
94. T. Banaszak, A., LaJeunesse, T. C. & Trench, R. K. The synthesis of mycosporine-like amino acids (MAAS) by cultured, symbiotic dinoflagellates. *J. Exp. Mar. Bio. Ecol.* **249**, 219–233 (2000).
95. Banaszak, A. T., Barba Santos, M. G., LaJeunesse, T. C. & Lesser, M. P. The distribution of mycosporine-like amino acids (MAAs) and the phylogenetic identity of symbiotic dinoflagellates in cnidarian hosts from the Mexican Caribbean. *J. Exp. Mar. Bio. Ecol.* **337**, 131–146 (2006).
96. Rosic, N. N. Phylogenetic analysis of genes involved in mycosporine-like amino acid biosynthesis in symbiotic dinoflagellates. *Appl. Microbiol. Biotechnol.* **94**, 29–37 (2012).
97. Christophe Corre, G. C. *Exploiting Genomics for New Natural Product*

Discovery in Prokaryotes. (Elsevier: Oxford, 2010).

98. Banskota, A. H. *et al.* Isolation and identification of three new 5-alkenyl-3,3(2H)-furanones from two Streptomyces species using a genomic screening approach. *J. Antibiot. (Tokyo).* **59**, 168–176 (2006).
99. Banskota, A. H. *et al.* Genomic analyses lead to novel secondary metabolites: Part 3 ECO-0501, a novel antibacterial of a new class. *J. Antibiot. (Tokyo).* **59**, 533–542 (2006).
100. Lautru, S. & Challis, G. L. Substrate recognition by nonribosomal peptide synthetase multi-enzymes. *Microbiology* **150**, 1629–1636 (2004).
101. Conti, E., Stachelhaus, T., Marahiel, M. A. & Brick, P. Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J.* **16**, 4174–4183 (1997).
102. Stachelhaus, T., Mootz, H. D. & Marahiel, M. A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **6**, 493–505 (1999).
103. Challis, G. L. & Ravel, J. Coelichelin, a new peptide siderophore encoded by the Streptomyces coelicolor genome: Structure prediction from the sequence of its non- ribosomal peptide synthetase. *FEMS Microbiol. Lett.* **187**, 111–114 (2000).
104. Fischbach, M. A. & Walsh, C. T. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: Logic machinery, and mechanisms. *Chem. Rev.* **106**, 3468–3496 (2006).
105. Insights, F. B. Market Research Report. (2020).
106. Department, S. R. Sun Care Industry - Statistics & Facts. (2022).
107. Afonso, S. *et al.* Photodegradation of avobenzone: Stabilization effect of antioxidants. *J. Photochem. Photobiol. B Biol.* **140**, 36–40 (2014).
108. Holt, E. L., Rodrigues, N. d. N., Cebrián, J. & Stavros, V. G. Determining the photostability of avobenzone in sunscreen formulation models using ultrafast spectroscopy. *Phys. Chem. Chem. Phys.* **23**, 24439–24448 (2021).
109. Dunkelberger, A. D., Kieda, R. D., Marsh, B. M. & Crim, F. F. Picosecond Dynamics of Avobenzone in Solution. *J. Phys. Chem. A* **119**, 6155–6161 (2015).
110. Schwack, W. & Rudolph, T. Photochemistry of dibenzoyl methane UVA filters Part 1. *J. Photochem. Photobiol. B Biol.* **28**, 229–234 (1995).

111. Lebedev, A. T. *et al.* Identification of avobenzone by-products formed by various disinfectants in different types of swimming pool waters. *Environ. Int.* **137**, 105495 (2020).
112. Downs, C. A. *et al.* Toxicopathological Effects of the Sunscreen UV Filter, Oxybenzone (Benzophenone-3), on Coral Planulae and Cultured Primary Cells and Its Environmental Contamination in Hawaii and the U.S. Virgin Islands. *Arch. Environ. Contam. Toxicol.* **70**, 265–288 (2016).
113. Brawley, S. H. *et al.* Insights into the red algae and eukaryotic evolution from the genome of Porphyra umbilicalis (Bangiophyceae, Rhodophyta). *Proc. Natl. Acad. Sci. U. S. A.* **114**, E6361–E6370 (2017).
114. Federico R Conde, M Sandra Churio, C. M. P. The deactivation pathways of the excited-states of the mycosporine-like amino acids shinorine and porphyra-334 in aqueous solution. *Photochem. Photobiol. Sci.* 960–967 (2004).
115. Pizzolla, P. . Porphyra umbilicalis Purple laver. in *In Tyler-Walters H. and Hiscock K. (eds) Marine Life Information Network: Biology and Sensitivity Key Information Reviews, [on-line]. Plymouth: Marine Biological Association of the United Kingdom.* (2008).
116. Whitehead, K. & Hedges, J. I. Analysis of mycosporine-like amino acids in plankton by liquid chromatography electrospray ionization mass spectrometry. *Mar. Chem.* **80**, 27–39 (2002).
117. Klisch, M., Richter, P., Puchta, R., Häder, D. P. & Bauer, W. The stereostructure of porphyra-334: An experimental and calculational NMR investigation. Evidence for an efficient 'proton sponge'. *Helv. Chim. Acta* **90**, 488–511 (2007).
118. Nakayama, R., Tamura, Y., Kikuzaki, H. & Nakatani, N. Antioxidant effect of the constituents of Susabinori (Porphyra yezoensis). *JAOCS, J. Am. Oil Chem. Soc.* **76**, 649–653 (1999).
119. Takano, S., Uemura, D. & Hirata, Y. Isolation and structure of two new amino acids, palythinol and palythene, from the zoanthid palythoa tubercolosa. *Tetrahedron Lett.* **19**, 4909–4912 (1978).
120. Vega, J. *et al.* Mycosporine-like amino acids from red macroalgae: Uv-photoprotectors with potential cosmeceutical applications. *Appl. Sci.* **11**, (2021).

121. Raj, S. *et al.* Microalgae as a source of mycosporine-like amino acids (Maas); advances and future prospects. *Int. J. Environ. Res. Public Health* **18**, (2021).
122. Rosic, N. N. Mycosporine-like amino acids: Making the foundation for organic personalised sunscreens. *Mar. Drugs* **17**, 1–17 (2019).
123. Jofre, J., Celis-Plá, P. S. M., Figueroa, F. L. & Navarro, N. P. Seasonal variation of mycosporine-like amino acids in three subantarctic red seaweeds. *Mar. Drugs* **18**, 1–17 (2020).
124. Hojerová, J., Medovcíková, A. & Mikula, M. Photoprotective efficacy and photostability of fifteen sunscreen products having the same label SPF subjected to natural sunlight. *Int. J. Pharm.* **408**, 27–38 (2011).
125. Gonzalez, H. *et al.* Photostability of commercial sunscreens upon sun exposure and irradiation by ultraviolet lamps. *BMC Dermatol.* **7**, 1–9 (2007).
126. Sinha RP, H. D. Photobiology and ecophysiology of rice field cyanobacteria. *Photochem. phytobiology* **64**, 887–896 (1996).
127. Gröniger, A., Sinha, R. P., Klisch, M. & Häder, D. P. Photoprotective compounds in cyanobacteria, phytoplankton and macroalgae - A database. *J. Photochem. Photobiol. B Biol.* **58**, 115–122 (2000).
128. Bothwell ML, Sherbot DM, P. C. Ecosystem response to solar ultraviolet-B radiation: influence of trophic-level interactions. *Science (80-.).* 97–100 (1994).
129. Häder, D. P., Kumar, H. D., Smith, R. C. & Worrest, R. C. Effects of solar UV radiation on aquatic ecosystems and interactions with climate change. *Photochem. Photobiol. Sci.* **6**, 267–285 (2007).
130. Vincent WF, R. S. Solar ultraviolet-B radiation and aquatic primary production : damage , protection , and recovery. *Environ. Rev.* **1**, 1–12 (1993).
131. Donkor V, H. D. Effects of solar and ultraviolet radiation on motility, photomovement and pigmentation in filamentous, gliding cyanobacteria. *FEMS Microbiol. Ecol.* **86**, 159–168 (1991).
132. Rastogi, R. P. *et al.* Ultraviolet radiation and cyanobacteria. *J. Photochem. Photobiol. B Biol.* **141**, 154–169 (2014).
133. Sinha, R. P., Richter, P., Faddoul, J., Braun, M. & Häder, D. P. Effects of UV and visible light on cyanobacteria at the cellular level. *Photochem. Photobiol. Sci.* **1**, 553–559 (2002).
134. Britt, A. B. Repair of DNA damage induced by ultraviolet radiation. *Plant*

Physiol. **108**, 891–896 (1995).

135. Kim ST, S. A. Photorepair of nonadjacent pyrimidine dimers by DNA photolyase. *Photochem. Photobiol.* **61**, 171–174 (1995).
136. Middleton, E. M. & Teramura, A. H. The role of flavonol glycosides and carotenoids in protecting soybean from ultraviolet-B damage. *Plant Physiol.* **103**, 741–752 (1993).
137. Mittler R, T. E. Oxidative stress responses in the unicellular cyanobacterium Synechococus PCC 7942. *Free Radic. Res. Commun.* **13**, 845–850 (1991).
138. Singh, S. P., Kumari, S., Rastogi, R. P., Singh, K. L. & Sinha, R. P. Mycosporine-like amino acids (MAAs): Chemical structure, biosynthesis and significance as UV-absorbing/screening compounds. *Indian J. Exp. Biol.* **46**, 7–17 (2008).
139. Cockell CS, K. Ultraviolet radiation screening compounds. *Biol. Rev.* **74**, 311–345 (1999).
140. Shick J.M, D. W. . Mycosporine-like amino acids and related gadusols: Biosynthesis, accumulation and UV-protective functions in aquatic organisms. *Annu. Rev. Physiol.* **64**, 223–262 (2002).
141. Oren, A. & Gunde-Cimerman, N. Mycosporines and mycosporine-like amino acids: UV protectants or multipurpose secondary metabolites? *FEMS Microbiol. Lett.* **269**, 1–10 (2007).
142. Oren, A. Mycosporine-like amino acids as osmotic solutes in a community of halophilic cyanobacteria. *Geomicrobiol. J.* **14**, 231–240 (1997).
143. Saha CK, Pires RS, Brolin H, Delannoy M, A. G. FlaGs and webFlaGs: discovering novel biology through the analysis of gene neighbourhood conservation. Bioinformatics. https://doi.org/10.1093/bioinformatics/btaa788 (2020).
144. Jimmy S, Saha CK, Kurata T, Stavropoulos C, Oliveira SRA, Koh A, Cepauskas A, Takada H, Rejman D, Tenson T, Strahl H, Garcia-Pina A, Hauryliuk V, A. G. A widespread toxin-antitoxin system exploiting growth control via alarmone signaling. *Proc. Natl. Acad. Sci. U. S. A.* (2020).
145. Nissley, D. A. & Obrien, E. P. Timing is everything: Unifying Codon translation rates and nascent proteome behavior. *J. Am. Chem. Soc.* **136**, 17892–17898 (2014).
146. Pfeifer, B. A., Admiraal, S. J., Gramajo, H., Cane, D. E. & Khosla, C.

Biosynthesis of complex polyketides in a metabolically engineered strain of E. coli. *Science (80-.).* **291**, 1790–1792 (2001).

147. Hamada, T., Tanaka, H., Izumine, H. & Ohira, M. Evaluation of an embedded polar C4 phase for hydrophobic protein analysis by reversed-phase liquid chromatography. *J. Chromatogr. A* **1043**, 27–32 (2004).
148. Xu, Guang; Stupak Jacek; Yang, Li; Hu, Luokai; Guo, Bo; Li, J. Deconvolution in mass spectrometry based proteomics. *Rapid Commun. Mass Spectrom.* **32**, 763–774 (2018).
149. Geoghegan, K. F. *et al.* Spontaneous α-N-6-phosphogluconoylation of a 'His tag' in Escherichia coli: The cause of extra mass of 258 or 178 Da in fusion proteins. *Anal. Biochem.* **267**, 169–184 (1999).
150. Kopp, F. & Marahiel, M. A. Macrocyclization strategies in polyketide and nonribosomal peptide biosynthesis. *Nat. Prod. Rep.* **24**, 735–749 (2007).
151. Bozhüyük, K. A. J. *et al.* De novo design and engineering of non-ribosomal peptide synthetases. *Nat. Chem.* **10**, 275–281 (2018).
152. Miyanaga, A., Cieślak, J., Shinohara, Y., Kudo, F. & Eguchi, T. The crystal structure of the adenylation enzyme VinN reveals a unique β-amino acid recognition mechanism. *J. Biol. Chem.* **289**, 31448–31457 (2014).
153. Reger, A. S., Wu, R., Dunaway-Mariano, D. & Gulick, A. M. Structural characterization of a 140° domain movement in the two-step reaction catalyzed by 4-chlorobenzoate:CoA ligase. *Biochemistry* **47**, 8016–8025 (2008).
154. Baltz, R. H. Model for Synthetic Biology To Accelerate the Evolution of Secondary Metabolite Biosynthetic Pathways. *ACS Synth. Biol.* **3**, 748–758 (2014).
155. Bozhüyük, K. A. J. *et al.* Modification and de novo design of non-ribosomal peptide synthetases using specific assembly points within condensation domains. *Nat. Chem.* **11**, 653–661 (2019).
156. Walsh, C.T, O'Brien, R.V, Khosla, C. Nonproteinogenic Amino Acid Building Blocks for Nonribosomal Peptide and Hybrid Polyketide Scaffolds. *Angew Chem Int Ed Engl.* **52**, 7098–7124 (2013).
157. Süssmuth, R. D. & Mainz, A. Nonribosomal Peptide Synthesis—Principles and Prospects. *Angew. Chemie - Int. Ed.* **56**, 3770–3821 (2017).
158. Payne, J. A. E., Schoppet, M., Hansen, M. H. & Cryle, M. J. Diversity of

nature's assembly lines-recent discoveries in non-ribosomal peptide synthesis. *Mol. Biosyst.* **13**, 9–22 (2017).

159. Grünewald, J. & Marahiel, M. A. Chemoenzymatic and Template-Directed Synthesis of Bioactive Macrocyclic Peptides. *Microbiol. Mol. Biol. Rev.* **70**, 121–146 (2006).
160. Kohli, R. M., Walsh, C. T. & Burkart, M. D. Biomimetic synthesis and optimization of cyclic peptide antibiotics. *Nature* **418**, 658–661 (2002).
161. Grünewald, J., Sieber, S. A., Mahlert, C., Linne, U. & Marahiel, M. A. Synthesis and derivatization of daptomycin: A chemoenzymatic route to acidic lipopeptide antibiotics. *J. Am. Chem. Soc.* **126**, 17025–17031 (2004).
162. Mahlert, C., Sieber, S. A., Grünewald, J. & Marahiel, M. A. Chemoenzymatic approach to enantiopure streptogramin B variants: Characterization of stereoselective pristinamycin I cyclase from Streptomyces pristinaespiralis. *J. Am. Chem. Soc.* **127**, 9571–9580 (2005).
163. Tanovic, A., Samel, S. A., Essen, L. O. & Marahiel, M. A. Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science (80-.).* **321**, 659–663 (2008).
164. Liu, Ye, Zheng, Tengfei, Bruner, S. . Structural basis for phosphopantetheinyl carrier domain interactions in the terminal module of nonribosomal peptide synthetases. *Chem Biol.* **18**, 1482–1488 (2011).
165. Drake, E. J. *et al.* Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature* **529**, 235–238 (2016).
166. Horsman, M. E., Hari, T. P. A. & Boddy, C. N. Polyketide synthase and non-ribosomal peptide synthetase thioesterase selectivity: Logic gate or a victim of fate? *Nat. Prod. Rep.* **33**, 183–202 (2016).
167. Trauger, J. W., Kohli, R. M., Mootz, H. D., Marahiel, M. A. & Walsh, C. T. Peptide cyclization catalysed by the thioesterase domain of tyrocidine synthetase. *Nature* **407**, 215–218 (2000).
168. Hoyer, K. M., Mahlert, C. & Marahiel, M. A. The Iterative Gramicidin S Thioesterase Catalyzes Peptide Ligation and Cyclization. *Chem. Biol.* **14**, 13–22 (2007).
169. Shaw-Reid, C.A, Kelleher, N.L, Losey, H.C, Gehring, A.M, Berg, C, Walsh, C. Assembly line enzymology by multimodular nonribosomal peptide synthetases: the thioesterase domain of E. coli EnfF catalyzes both elongation

and cyclolactonization. *Chem. Biol.* **6**, 385–400 (1999).

170. Ray, L, Yamanaka, K, Moore, B. . A Peptidyl Transesterifying Type I Thioesterase in Salinamide Biosynthesis. *Angew Chem Int Ed Engl.* **55**, 364–367 (2016).
171. Robbel, L., Hoyer, K. M. & Marahiel, M. A. TioS T-TE - A prototypical thioesterase responsible for cyclodimerization of the quinoline- and quinoxaline-type class of chromodepsipeptides. *FEBS J.* **276**, 1641–1653 (2009).
172. Bruner, S. D. *et al.* Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE. *Structure* **10**, 301–310 (2002).
173. Samel, S. A., Wagner, B., Marahiel, M. A. & Essen, L. O. The Thioesterase Domain of the Fengycin Biosynthesis Cluster: A Structural Base for the Macrocyclization of a Non-ribosomal Lipopeptide. *J. Mol. Biol.* **359**, 876–889 (2006).
174. Sieber, S. A., Tao, J., Walsh, C. T. & Marahiel, M. A. Peptidyl Thiophenols as Substrates for Nonribosomal Peptide Cyclases. *Angew. Chemie - Int. Ed.* **43**, 493–498 (2004).
175. Lin, H. & Walsh, C. T. A chemoenzymatic approach to glycopeptide antibiotics. *J. Am. Chem. Soc.* **126**, 13998–14003 (2004).
176. Schwarzer, D., Mootz, H. D., Linne, U. & Marahiel, M. A. Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 14083–14088 (2002).
177. Heathcote, M. L., Staunton, J. & Leadlay, P. F. Role of type II thioesterases: Evidence for removal of short acyl chains produced by aberrant decarboxylation of chain extender units. *Chem. Biol.* **8**, 207–220 (2001).
178. Harvey, B. M. *et al.* Insights into Polyether Biosynthesis from Analysis of the Nigericin Biosynthetic Gene Cluster in Streptomyces sp. DSM4137. *Chem. Biol.* **14**, 703–714 (2007).
179. Gao, L. *et al.* Translocation of the thioesterase domain for the redesign of plipastatin synthetase. *Sci. Rep.* **6**, 1–9 (2016).
180. Sampedro, D. Natural and artificial photoprotective agents. *Molecules* **26**, 11–12 (2021).
181. Arrieta, J. M., Weinbauer, M. G. & Herndl, G. J. Interspecific variability in

sensitivity to UV radiation and subsequent recovery in selected isolates of marine bacteria. *Appl. Environ. Microbiol.* **66**, 1468–1473 (2000).

182. Rastogi, R. P., Madamwar, D., Nakamoto, H. & Incharoensakdi, A. Resilience and self-regulation processes of microalgae under UV radiation stress. *J. Photochem. Photobiol. C Photochem. Rev.* **43**, 100322 (2020).
183. Oliveira, L. B. A., Fonseca, T. L. & Cabral, B. J. C. 15N NMR shifts of eumelanin building blocks in water: A combined quantum mechanics/statistical mechanics approach. *Molecules* **25**, (2020).
184. Geraldes, V., Jacinavicius, F. R., Genuário, D. B. & Pinto, E. Identification and distribution of mycosporine-like amino acids in Brazilian cyanobacteria using ultrahigh-performance liquid chromatography with diode array detection coupled to quadrupole time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **34**, (2020).
185. Jain, S. *et al.* Cyanobacteria as efficient producers of mycosporine-like amino acids. *J. Basic Microbiol.* **57**, 715–727 (2017).
186. Sun, Y. *et al.* Distribution, contents, and types of mycosporine-like amino acids (MAAs) in marine macroalgae and a database for Maas based on these characteristics. *Mar. Drugs* **18**, (2020).
187. Rosic, N. N. & Dove, S. Mycosporine-like amino acids from coral dinoflagellates. *Appl. Environ. Microbiol.* **77**, 8478–8486 (2011).
188. Osborn, A. R. *et al.* De novo synthesis of a sunscreen compound in vertebrates. *Elife* **4**, 1–15 (2015).
189. Newman, S. J., Dunlap, W. C., Nicol, S. & Ritz, D. Antarctic krill (Euphausia superba) acquire a UV-absorbing mycosporine-like amino acid from dietary algae. *J. Exp. Mar. Bio. Ecol.* **255**, 93–110 (2000).
190. Walter Helbling, E., Fernando Menchi, C. & Villafañe, V. E. Bioaccumulation and role of UV-absorbing compounds in two marine crustacean species from Patagonia, Argentina. *Photochem. Photobiol. Sci.* **1**, 820–825 (2002).
191. Geraldes, V. & Pinto, E. Mycosporine-like amino acids (Maas): Biology, chemistry and identification features. *Pharmaceuticals* **14**, 1–17 (2021).
192. Pope, M. A. *et al.* O-methyltransferase is shared between the pentose phosphate and shikimate pathways and is essential for mycosporine-like amino acid biosynthesis in Anabaena variabilis ATCC 29413. *ChemBioChem*

16, 320–327 (2015).

193. Katoch, M. *et al.* Heterologous production of cyanobacterial mycosporine-like amino acids mycosporine-ornithine and mycosporine-lysine in Escherichia coli. *Appl. Environ. Microbiol.* **82**, 6167–6173 (2016).
194. Spence, E., Dunlap, W. C., Shick, J. M. & Long, P. F. Redundant Pathways of Sunscreen Biosynthesis in a Cyanobacterium. *ChemBioChem* **13**, 531–533 (2012).
195. Geraldes, V. *et al.* Genetic and biochemical evidence for redundant pathways leading to mycosporine-like amino acid biosynthesis in the cyanobacterium sphaerospermopsis torques-reginae ITEP-024. *Algae* **35**, 177–187 (2020).

Appendix

[1] Amino Acid sequence and codon-optimized (for *E. coli*) sequences of each NRPS-like enzyme

[1] Original sequence of NRPS-like enzyme from *Anabaena variabilis*

MDINTETETHRDYDLSQSLHELIVAQVERTPEAIAVTFDKQQLTYQELNHKANQLGHYLQTLGVQPETLVGVCLE
RSLEMVICLLGILKAGGAYVPIDPEYPQERIAYMLEDSQVKVLLTQEKLLNQIPHHQAQTICVDREWEKISTQAN
TNPKSNIKTDNLAYVIYTSGSTGKPKGAMNTHKGICNRLLWMQEAYQIDSTDSILQKTPFSFDVSVWEFFWTLLT
GARLVIAKPGGHKDSAYLIDLITQEQITTLHFVPSMLQVFLQNRHVSKCSSLKRVICSGEALSIDLQNRFFQHLQ
CELHNLYGPTEAAIDVTFWQCRKDSNLKSVPIGRPIANTQIYILDADLQPVNIGVTGEIYIGGVGVARGYLNKEE
LTKEKFIINPFPNSEFKRLYKTGDLARYLPDGNIEYLGRTDYQVKIRGYRIEIGEIENVLSSHPQVREAVVIARD
DNAQEKQIIAYITYNSIKPQLDNLRDFLKARLPDFMIPAAFVMLEHLPLTPSGKVDRKALPKPDLFNYSEHNSYV
APRNEVEEKLVQIWSNILHLPKVGVTENFFAIGGNSLKALHLISQIEELFAKEISLATLLTNPVIADLAKVIQAN
NQIHNSPLVPIQPQGKQQPFFCIHPAGGHVLCYFKLAQYIGTDQPFYGLQAQGFYGDEAPLTRVEDMASLYVKTI
REFQPQGPYRVGGWSFGGVVAYEVAQQLHRQGQEVSLLAILDSYVPILLDKQKPIDDVYLVGVLSRVFGGMFGQD
NLVTPEEIENLTVEEKINYIIDKARSARIFPPGVERQNNRRILDVLVGTLKATYSYIRQPYPGKVTVFRAREKHI
MAPDPTLVWVELFSVMAAQEIKIIDVPGNHYSFVLEPHVQVLAQRLQDCLENNS

Codon optimized sequence of NRPS-like enzyme from *Anabaena variabilis* for *E. coli* (L:2637)

ATGGACATTAACACCGAAACCGAGACCCACCGCGATTACGATCTGAGTCAGAGTCTGCATGAACTCATCGTTGCC
CAAGTTGAACGCACGCCAGAGGCGATCGCGGTGACCTTCGATAAACAGCAGCTCACCTACCAAGAACTGAACCAC
AAGGCGAATCAGCTCGGCCACTATCTCCAAACGCTGGGCGTGCAGCCGGAGACGCTGGTTGGTGTTTGTCTGGAG
CGCAGTCTGGAGATGGTGATTTGTCTGCTGGGCATCCTCAAAGCGGGTGGTGCCTATGTTCCGATCGACCCAGAA
TATCCGCAAGAACGCATTGCCTACATGCTGGAGGATAGCCAAGTTAAAGTTCTGCTGACGCAAGAAAAGCTGCTG
AACCAGATTCCGCATCACCAAGCCCAAACCATCTGTGTGGATCGTGAGTGGGAGAAGATCAGTACGCAAGCCAAC
ACCAACCCGAAAAGCAATATCAAAACGGACAATCTGGCCTACGTTATCTACACGAGCGGTAGCACGGGCAAACCG
AAAGGCGCGATGAACACCCATAAGGGCATCTGTAACCGTCTGCTGTGGATGCAAGAGGCGTACCAAATCGACAGC
ACCGACAGCATTCTGCAGAAGACCCCGTTTAGCTTCGACGTGAGTGTTTGGGAGTTCTTCTGGACGCTCCTCACG
GGTGCGCGTCTCGTTATTGCCAAGCCGGGCGGCCATAAGGACAGTGCCTACCTCATTGATCTGATCACGCAAGAA
CAGATTACCACGCTGCATTTCGTGCCGAGCATGCTGCAAGTTTTTCTGCAGAACCGCCATGTGAGTAAGTGCAGT
AGTCTGAAGCGCGTGATCTGCAGTGGTGAAGCGCTGAGCATTGACCTCCAGAACCGTTTCTTCCAGCATCTCCAG
TGCGAGCTGCACAACCTCTATGGTCCGACGGAGGCGGCCATTGACGTGACGTTCTGGCAATGTCGCAAAGACAGC
AATCTGAAGAGCGTTCCGATTGGTCGCCCAATCGCGAATACCCAGATCTACATTCTGGACGCGGATCTCCAACCG
GTGAACATCGGCGTGACGGGTGAGATTTATATCGGCGGTGTGGGCGTGGCCCGTGGTTATCTGAACAAAGAGGAA
CTGACGAAAGAGAAATTCATTATCAACCCGTTCCCGAACAGCGAGTTCAAGCGTCTGTATAAAACGGGCGATCTG
GCGCGTTATCTGCCAGATGGCAACATTGAATATCTGGGCCGCACCGACTATCAAGTTAAGATTCGCGGCTACCGC
ATTGAGATCGGCGAAATTGAAAACGTGCTGAGCAGCCACCCGCAAGTTCGTGAAGCGGTTGTGATCGCCCGTGAC
GACAACGCGCAAGAAAAGCAGATCATTGCGTATATTACGTACAATAGCATCAAACCGCAGCTCGACAATCTGCGC
GACTTTCTGAAAGCCCGTCTCCCGGACTTCATGATCCCGGCGGCGTTCGTTATGCTCGAGCACCTCCCACTGACC
CCAAGCGGTAAAGTTGATCGCAAAGCGCTGCCGAAACCAGATCTGTTCAACTACAGTGAGCACAACAGCTATGTG
GCGCCGCGTAACGAAGTTGAAGAGAAGCTCGTTCAGATCTGGAGCAACATCCTCCACCTCCCGAAAGTTGGCGTT
ACGGAAAACTTTTTCGCCATCGGTGGCAACAGCCTCAAGGCCCTCCATCTGATCAGCCAGATTGAGGAACTCTTC
GCGAAGGAAATCAGTCTCGCCACGCTGCTGACCAACCCAGTTATTGCCGACCTCGCGAAGGTGATTCAAGCCAAC
AACCAGATCCATAACAGCCCACTCGTGCCGATCCAGCCACAAGGCAAGCAGCAACCATTCTTCTGTATTCATCCA
GCCGGTGGCCACGTTCTCTGCTATTTCAAGCTCGCGCAGTACATCGGCACGGATCAACCATTCTATGGTCTCCAA
GCGCAAGGCTTTTATGGCGATGAAGCCCCGCTGACGCGTGTTGAGGATATGGCCAGTCTGTACGTGAAGACGATT
CGCGAATTTCAACCGCAAGGCCCGTACCGCGTTGGCGGTTGGAGCTTCGGTGGCGTGGTTGCGTATGAAGTGGCC
CAGCAACTCCACCGTCAAGGCCAAGAAGTGAGTCTGCTGGCGATTCTGGATAGCTACGTGCCAATTCTGCTGGAC
AAACAGAAACCGATCGATGACGTGTATCTGGTGGGCGTTCTGAGCCGCGTTTTTGGCGGCATGTTCGGCCAAGAC
AATCTGGTGACCCCGGAAGAAATCGAAAATCTCACCGTGGAAGAAAAAATTAACTACATCATCGATAAAGCCCGC
AGCGCGCGTATCTTTCCGCCGGGTGTGGAACGCCAGAACAACCGCCGCATTCTGGATGTTCTGGTTGGCACGCTG
AAGGCGACCTACAGCTACATTCGCCAGCCATACCCGGGTAAAGTTACGGTGTTTCGTGCCCGCGAGAAGCATATT
ATGGCGCCGGATCCGACGCTGGTTTGGGTGGAACTGTTTAGCGTGATGGCGGCCCAAGAAATCAAGATCATTGAC
GTGCCGGGTAACCATTACAGCTTCGTGCTGGAACCGCACGTTCAAGTTCTCGCCCAGCGTCTGCAAGATTGTCTG
GAAAACAACAGT

[2] Original sequence of NRPS-like enzyme from *Nostocales cyanobacterium*

MDINTETEISNAEAITEFSLTHAQQVAWFLYKFAPQSYIDKLAFAVRFSQQLDYECLDRVFQSLVKRHPSLRTAY
IEKQGKVLQHVVDEASASVDFVSSTGWNEEKLQNQVLQSLQRPFDLASGSVVRLSVFSCTPTNHVILLAVHQIAC
DDRTLMLLVDELLKLYQARKNNIVIALPSVDSSYQEYVQRELNLLNSPEGEQLWYELKGRLDGELPTLNLPMSRP
RPPVRTYRGALHKFCISPNIASKLQQFAQTEDVELSTILLAAFQVILHRYTATEELLVGFKTEQAKSRNFNCKIG
NFNNLTVVRSSISSTLSFQELLSLTRSAVFDVIAHQDYPFPLLVRQLQLNSQLSHPPICQVGLTYQNLDKLETLS
TLFNQTNSSLSYFEIPEQRVEFDLNWEILEAGESLTCFLHYNRDLLDADAIVRMAGQLQILLAEIVTSPKQQVTQ
LPILTDSEQHQLLIEWNATHRDYNLSRCLHELFETQVDLTPEAIAVSFEQQQLTYRELNEKANQLAHHLQKKGVK
TEVLVGISVERSLQMVIGLLGILKAGGAYVPIDPESPQERIAYMLADSQVSLLLTQKKQVTQLPKSQAQIICLDA
DWEKISQEQTNNPASGVQPENLAYVIYTSGSTGKPKGAMNTHRGICNRLLWMQEAYQLTATDAVLQKTPFSFDVS
VWEFFWTLLTGARLVIAKPGGHRDRSYLVELISQEQITTLHFVPSMLQVFLESRDLDKCQSLRRVICSGEALTLD
LQEKFFQHLGCELHNLYGPTEAAIDVTFWRCQRQSHLRTVPIGRPIANTQIYILDPHLQPVPIGVTGELYIGGVG
VARGYLNREELTAECFITNPFRQSDRLYKTGDLARYLSDGNIEYIGRIDHQVKIRGFRVELGEIENALSQHPQVR
EAVVIVRQDKPGDKQIVAYIVSTENLPSPSLLREFLRQKLPDYMVPAAFILLETLPLTSNGKLDRRALPSPSLSN
FSESHNYIAPRNDIEQQLAEIWAQILDVHPVGVRDNFFELGGNSLLAIHLIAEIEEKFGKDLPLSELLTNPVIED
LAKILQTASNIFNNSPIVPIQPKGNKRPFFCVHPAGGHVFCYFNLARYLGREQPFYGLQAQGFNGEEEPLSRVED
MASLYVKAIQTVQPEGPYQIGGWSFGGVVAYEVAQQFYKQGHEVSLLAILDSYVPIVLDKQKEINNQYLVGVLSR
VFGGMFGQDNLVTQDEIKHLSVDEQIDYIIDKARKVKIFPPTVERRKNRRILDVLVGTLKATYSYVRQPYPGKAT
IFRAREKHIMAPDPTLVWVELFSIMAAKEIEIIDVPGSHYSFVLEPHVQVLAERLKTCLV

Codon optimized sequence of NRPS-like enzyme from *Nostocales cyanobactrium* for *E. coli* (L:4005)

ATGGATATCAACACCGAAACGGAGATCAGCAACGCCGAAGCGATTACCGAGTTCAGTCTCACGCATGCGCAGCAA
GTGGCGTGGTTTCTGTACAAATTCGCCCCGCAAAGCTACATTGACAAGCTGGCCTTTGCGGTGCGTTTTAGCCAG
CAGCTCGATTATGAGTGTCTGGATCGCGTTTTCCAAAGCCTCGTGAAACGTCACCCGAGTCTGCGCACGGCGTAT
ATCGAGAAGCAAGGCAAAGTGCTGCAACACGTTGTGGATGAAGCGAGCGCCAGCGTGGACTTTGTGAGCAGTACC
GGCTGGAATGAGGAGAAACTGCAAAACCAAGTTCTGCAGAGTCTCCAGCGTCCATTTGATCTGGCCAGTGGCAGT
GTTGTGCGCCTCAGCGTGTTCAGCTGCACCCCAACGAACCACGTTATTCTGCTGGCGGTGCATCAGATCGCGTGT
GATGACCGCACGCTGATGCTGCTGGTGGACGAACTGCTGAAACTCTACCAAGCCCGTAAGAACAACATTGTGATC
GCGCTGCCGAGCGTGGACAGCAGCTACCAAGAGTACGTGCAACGTGAGCTCAATCTGCTGAACAGTCCGGAGGGC
GAGCAACTGTGGTACGAACTGAAGGGCCGTCTGGATGGTGAACTCCCGACGCTGAATCTCCCGATGAGTCGTCCA
CGTCCGCCAGTTCGCACGTATCGTGGCGCGCTCCATAAATTCTGCATCAGCCCAAACATTGCGAGCAAGCTGCAA
CAGTTCGCGCAGACCGAGGATGTGGAACTGAGCACGATTCTGCTGGCGGCGTTCCAAGTTATTCTGCACCGCTAT
ACCGCGACCGAAGAACTGCTGGTGGGCTTCAAAACCGAGCAAGCCAAGAGCCGCAACTTTAATTGCAAAATTGGT
AATTTTAACAATCTGACCGTGGTTCGCAGCAGCATCAGCAGTACGCTGAGTTTTCAAGAACTGCTGAGTCTGACG
CGCAGTGCGGTGTTCGACGTGATCGCCCACCAAGATTACCCGTTCCCACTGCTGGTTCGCCAGCTCCAGCTGAAC
AGCCAGCTCAGTCACCCGCCGATTTGCCAAGTGGGTCTCACGTACCAGAACCTCGACAAACTCGAGACCCTCAGT
ACGCTGTTTAACCAGACCAACAGCAGTCTGAGCTACTTCGAAATCCCGGAGCAACGTGTGGAGTTTGATCTCAAC
TGGGAAATTCTGGAGGCCGGTGAGAGCCTCACGTGCTTTCTGCATTACAATCGCGATCTGCTGGACGCGGATGCG
ATCGTTCGTATGGCCGGTCAACTCCAGATTCTGCTGGCGGAAATCGTTACGAGCCCAAAACAGCAAGTTACGCAG
CTCCCAATTCTGACCGATAGTGAGCAGCACCAGCTGCTGATTGAGTGGAATGCGACGCACCGCGACTACAACCTC
AGTCGTTGTCTGCACGAGCTGTTTGAGACGCAAGTTGATCTGACCCCGGAGGCCATCGCCGTGAGCTTCGAGCAG
CAGCAGCTCACCTACCGCGAGCTCAATGAAAAGGCGAACCAACTCGCGCACCATCTGCAAAAGAAGGGCGTGAAA
ACGGAGGTGCTGGTGGGCATCAGCGTTGAACGCAGTCTCCAGATGGTGATCGGTCTGCTCGGTATTCTGAAAGCC
GGTGGCGCGTACGTTCCAATCGATCCGGAGAGTCCACAAGAGCGCATCGCGTATATGCTCGCGGATAGTCAAGTT
AGTCTGCTGCTGACCCAGAAGAAGCAAGTGACGCAGCTGCCAAAGAGCCAAGCCCAGATCATCTGTCTGGATGCC
GACTGGGAAAAAATCAGCCAAGAGCAAACCAATAACCCAGCCAGCGGTGTGCAGCCGGAGAATCTGGCCTACGTT
ATCTACACCAGTGGCAGTACCGGCAAACCGAAAGGTGCGATGAATACCCACCGTGGCATTTGCAACCGTCTGCTG
TGGATGCAAGAAGCCTATCAACTCACGGCCACCGATGCGGTTCTCCAAAAGACCCCGTTCAGCTTCGATGTGAGC
GTGTGGGAGTTCTTCTGGACGCTGCTGACGGGTGCCCGTCTGGTGATCGCCAAACCGGGCGGTCACCGCGACCGT
AGTTATCTGGTTGAACTCATCAGCCAAGAACAGATCACGACGCTGCATTTCGTGCCGAGCATGCTCCAAGTTTTC
CTCGAGAGTCGCGATCTCGACAAATGTCAGAGTCTGCGCCGCGTTATCTGCAGTGGTGAAGCGCTCACGCTCGAC
CTCCAAGAAAAATTCTTCCAGCATCTGGGCTGTGAGCTCCACAATCTGTACGGCCCGACGGAAGCGGCGATTGAT
GTGACGTTTTGGCGTTGCCAACGTCAGAGCCATCTGCGCACGGTTCCAATTGGCCGTCCGATCGCCAACACCCAA
ATCTACATTCTGGACCCGCATCTGCAACCGGTTCCGATCGGTGTGACCGGCGAACTGTATATTGGCGGTGTTGGT
GTTGCGCGCGGTTACCTCAACCGTGAAGAACTCACGGCCGAGTGCTTCATCACGAACCCGTTCCGTCAGAGCGAC
CGTCTGTACAAAACGGGTGATCTGGCCCGCTATCTGAGCGACGGCAATATCGAATACATTGGCCGCATCGACCAC
CAAGTGAAAATCCGCGGCTTCCGTGTGGAGCTCGGCGAGATCGAGAACGCGCTGAGCCAGCACCCACAAGTTCGC
GAAGCCGTGGTGATCGTGCGTCAAGATAAACCGGGCGACAAGCAGATCGTTGCCTACATCGTGAGCACCGAGAAC
CTCCCGAGCCCGAGTCTGCTGCGCGAATTCCTCCGCCAGAAACTGCCGGATTACATGGTGCCAGCGGCGTTCATT
CTGCTGGAGACGCTGCCGCTGACCAGCAATGGTAAGCTCGATCGTCGCGCGCTGCCAAGCCCAAGTCTCAGTAAC

```
TTTAGCGAAAGCCACAACTACATCGCCCCGCGTAATGACATTGAGCAACAGCTGGCGGAGATTTGGGCGCAGATT
CTGGATGTTCACCCAGTGGGCGTGCGCGACAATTTCTTTGAGCTGGGCGGCAACAGTCTGCTGGCCATCCATCTC
ATCGCGGAGATCGAGGAAAAATTTGGCAAGGATCTGCCGCTCAGCGAACTGCTGACGAACCCAGTTATCGAGGAT
CTGGCCAAGATTCTGCAGACCGCCAGCAATATCTTCAACAACAGCCCGATCGTTCCGATCCAGCCGAAGGGCAAC
AAGCGTCCGTTTTTCTGTGTTCACCCGGCGGGCGGCCATGTGTTCTGCTACTTCAACCTCGCGCGCTATCTGGGC
CGCGAACAGCCATTTTATGGTCTCCAAGCCCAAGGCTTCAATGGTGAAGAAGAACCACTGAGCCGCGTGGAAGAC
ATGGCGAGTCTGTATGTGAAGGCGATCCAAACCGTTCAGCCAGAAGGTCCATACCAAATCGGCGGTTGGAGTTTT
GGTGGCGTTGTTGCCTATGAGGTGGCCCAACAGTTCTACAAACAAGGTCATGAGGTGAGTCTGCTGGCGATTCTC
GATAGCTACGTTCCAATTGTGCTGGACAAACAAAAAGAGATTAACAACCAGTACCTCGTGGGCGTTCTGAGTCGC
GTGTTCGGCGGCATGTTTGGCCAAGATAACCTCGTTACGCAAGATGAAATCAAACATCTGAGCGTGGACGAGCAG
ATTGACTACATTATCGATAAGGCCCGCAAGGTGAAGATTTTCCCACCGACGGTTGAACGCCGCAAGAACCGTCGC
ATCCTCGATGTTCTCGTGGGCACCCTCAAAGCCACCTACAGCTATGTTCGCCAACCGTATCCGGGCAAAGCGACG
ATTTTCCGTGCCCGTGAAAAGCATATCATGGCGCCAGATCCAACGCTGGTTTGGGTGGAGCTCTTCAGTATCATG
GCCGCCAAGGAGATCGAGATCATCGATGTTCCGGGTAGCCACTACAGTTTTGTGCTGGAGCCACACGTTCAAGTT
CTGGCCGAGCGTCTGAAAACGTGTCTGGTT
```

[3] Original sequence of NRPS-like enzyme from *Chlorogloeopsis fritschii*

```
MDINTETEESSTESIIEFPLSETQRKAWFLSKLELQGCTDKIAFAIHWRQQVNVECLKQAFEIVIARHPSLHTAY
RDKHGELVQQVLATTTAEINFVDASNWSEDKLKEKILDSSQHPFDLAVGSVVRMSLFSRTPTTHVLLLVVHCIAS
DTWGLLLLLDELLTLYHQLKSNTAISLSSLTSSYQDYVKQEINLLNSPEGKQLGHYWQQYLAGELPTLELPSARS
RSPIRSYKGACYKFSIAPDITDKVRYLAITEDVDLSTVVLSAFQLMLYRYTATENILVGLLPQRYQSEFKNVVGN
FANSAVVKLAISGDVSFQRYLSQIQFAVAETIAHQAYPFPLLVRQLQLNSQLSHPPICQVGFAYHHLHELKIISK
LFDENFGELEYFEIPRQRTEFDLSLEILESQESLIGFLYYNSDLLDADTIARAAEHLQNLLVAIIANPQQLVARL
PLLSDREKHQLLVEWNSTSKDYDLSRCLHELFEAQAEKTPEAIALSFEEEKLTYRELNSRANQLAHYLQNLGVKP
EVLVGICVERSLEMVVGLLGILKAGGAYVPIDPEYPPERLAYMLADSQVSVLLTQQKLLARLPNHQAEIICLDRD
WEEISQEQNTNLTSGVKPDNLAYVIYTSGSTGKPKGAMNTHQGVCNRLLWMQQAYQLTSTDRVLQKTPFSFDVSV
WEFFWTLLNGARLVIAKPRGHRDSTYLVKLIIQEQITTVHFVPSMLQVFLDSRDVKQCSSLKRVICSGEALPLDL
QARFFQCLQCELHNLYGPTEAAIDVTAWQCQKQSHLKTIPIGRPVANTQIYILDSHLQPVPIGVVGELHIGGVQV
ARGYLNRPDLTQEKFIANPFGKAQESRLYKTGDLARYLSDGNIEYIGRTDYQVKIRGLRIELGEIENALSQHCQV
REAVVIVRCDRPGDKQLVAYITTEQEKPTPESLREFLKQKLPEYMVPVAFVILEALPLTPSGKLDRRALPKPNLF
SFSQSQGSVLPRNDTERELAKIWLDILTIQSVGVQDNFFEIGGTSLSAIYLIAAIEQQFGKELPLSVLLTNPTIE
ELAKVLHLSSEQTNNSPLIPIQPKGNKQPFFCVHPAGGHVLCYFNLARYLGTDQPFYGLQAQGFNDDEEPLTRVE
DMASTYVEAIRQFQPYGPYQIGGWSFGGIVAYEIAQQLHKQGQEVSLLAILDSYVPIILDPNKKIDNKYLVGVLS
RVFGGMFGQDNLVTPQEIQHLSVEERLDYIIEKARKAKIFPPGVERSKNRRILDVLVGTLKATYSYVRQPYPGKV
TVFRASEKHIMAPDPTLVWVELLSVMAVKDVKVVKIPGNHYTFILEPHVKVLAERLKSCLNSSSQNLESDSVFL
```

Codon optimized sequence of NRPS-like enzyme from *Chlorogloeopsis fritschii* for *E. coli* (L:4047)

```
ATGGACATTAACACCGAGACCGAGGAGAGCAGTACGGAGAGCATCATCGAATTCCCACTCAGCGAAACCCAGCGC
AAAGCGTGGTTTCTCAGTAAACTGGAGCTCCAAGGTTGCACGGACAAGATTGCCTTCGCCATCCACTGGCGTCAG
CAAGTTAATGTTGAGTGTCTCAAGCAAGCCTTCGAAATCGTGATTGCGCGCCATCCAAGTCTGCACACGGCGTAT
CGTGACAAACATGGCGAACTCGTGCAACAAGTTCTCGCCACGACCACCGCGGAGATCAACTTCGTTGACGCCAGT
AACTGGAGCGAGGATAAGCTCAAGGAAAAAATTCTGGACAGCAGCCAACACCCATTTGATCTGGCCGTTGGTAGT
GTGGTGCGTATGAGTCTGTTTAGTCGTACCCCAACCACGCACGTTCTGCTGCTCGTGGTGCACTGCATTGCGAGC
GACACGTGGGGTCTGCTGCTGCTGCTGGATGAACTGCTGACGCTCTACCACCAGCTCAAGAGCAACACCGCGATT
AGTCTCAGTAGTCTGACGAGTAGCTATCAAGATTATGTGAAGCAAGAAATCAATCTGCTCAACAGTCCGGAGGGT
AAACAGCTCGGTCACTACTGGCAACAGTACCTCGCGGGTGAGCTGCCAACGCTCGAACTGCCAAGTGCGCGCAGC
CGTAGTCCAATCCGCAGTTACAAGGGTGCGTGCTACAAGTTTAGTATTGCGCCGGACATTACCGATAAGGTGCGC
TACCTCGCGATTACGGAGGACGTGGATCTGAGCACCGTTGTTCTGAGTGCGTTTCAGCTGATGCTGTACCGCTAC
ACCGCGACCGAAAATATCCTCGTGGGTCTGCTGCCGCAACGCTATCAAAGCGAGTTCAAGAACGTGGTGGGCAAT
TTCGCGAACAGTGCGGTGGTTAAACTCGCGATCAGCGGCGATGTTAGCTTCCAGCGCTACCTCAGCCAGATCCAG
TTCGCGGTGGCGGAAACCATTGCGCATCAAGCCTATCCGTTCCCACTGCTGGTTCGCCAGCTGCAGCTGAACAGC
CAACTGAGCCATCCGCCGATTTGTCAAGTTGGCTTTGCGTACCACCATCTCCATGAGCTGAAGATCATCAGTAAG
CTCTTTGACGAGAATTTCGGCGAGCTCGAATATTTCGAAATCCCGCGCCAGCGCACCGAATTCGATCTGAGTCTG
GAAATTCTCGAAAGCCAAGAAAGTCTGATCGGTTTCCTCTACTACAATAGTGATCTGCTGGATGCGGACACGATT
GCCCGTGCCGCCGAGCATCTGCAGAATCTGCTGGTGGCGATTATTGCGAACCCGCAACAGCTCGTTGCGCGTCTG
CCGCTGCTGAGTGACCGCGAAAAACATCAGCTGCTGGTGGAGTGGAACAGTACGAGTAAGGACTACGATCTGAGC
CGCTGTCTGCATGAACTCTTTGAGGCGCAAGCCGAAAAAACCCCAGAAGCGATTGCGCTGAGTTTTGAAGAAGAA
AAACTCACGTACCGCGAACTGAATAGTCGCGCGAACCAGCTCGCCCATTACCTCCAGAATCTGGGCGTGAAACCA
GAAGTGCTGGTGGGCATCTGCGTGGAGCGCAGTCTGGAAATGGTGGTTGGTCTGCTGGGTATCCTCAAGGCCGGT
GGCGCGTATGTTCCAATCGATCCAGAGTATCCACCGGAACGCCTCGCCTACATGCTCGCCGATAGTCAAGTTAGC
GTTCTGCTGACCCAGCAAAAGCTCCTCGCCCGCCTCCCGAATCATCAAGCCGAGATTATCTGCCTCGATCGCGAC
```

```
TGGGAGGAAATCAGCCAAGAACAGAACACGAACCTCACGAGCGGCGTTAAACCGGATAACCTCGCGTATGTGATC
TACACCAGTGGCAGTACCGGCAAACCGAAAGGCGCCATGAACACGCATCAAGGTGTTTGCAACCGTCTGCTCTGG
ATGCAACAAGCCTACCAGCTGACGAGTACGGACCGCGTGCTCCAGAAGACGCCGTTCAGCTTCGACGTGAGTGTG
TGGGAGTTCTTCTGGACGCTGCTGAACGGTGCCCGTCTGGTGATCGCGAAACCACGTGGCCATCGCGACAGCACG
TACCTCGTTAAGCTGATCATTCAAGAACAGATCACGACCGTGCACTTCGTGCCGAGTATGCTCCAAGTGTTTCTG
GATAGCCGCGACGTGAAACAATGCAGCAGCCTCAAACGTGTTATCTGTAGCGGCGAAGCCCTCCCACTGGACCTC
CAAGCCCGCTTTTTTCAATGCCTCCAGTGCGAGCTGCACAATCTCTACGGCCCGACGGAAGCGGCCATCGACGTT
ACCGCGTGGCAGTGTCAGAAGCAGAGTCATCTGAAAACGATCCCAATCGGCCGCCCGGTTGCGAATACCCAGATC
TACATTCTGGATAGCCATCTGCAGCCGGTGCCGATTGGCGTTGTGGGTGAACTGCACATTGGCGGTGTGCAAGTG
GCGCGCGGTTATCTGAATCGCCCAGATCTGACCCAAGAAAAGTTCATCGCCAATCCGTTCGGTAAGGCCCAAGAA
AGCCGTCTGTACAAGACGGGTGATCTGGCCCGTTACCTCAGCGACGGCAACATCGAGTATATCGGCCGCACGGAT
TACCAAGTTAAAATCCGCGGTCTCCGCATTGAGCTCGGTGAGATCGAAAACGCCCTCAGCCAGCATTGCCAAGTT
CGCGAGGCGGTTGTGATTGTGCGCTGCGATCGTCCGGGCGATAAACAGCTCGTGGCCTACATCACCACCGAGCAA
GAAAAACCAACCCCAGAAAGTCTGCGCGAATTTCTGAAACAGAAACTCCCGGAATATATGGTGCCAGTGGCCTTT
GTTATCCTCGAGGCGCTCCCGCTGACCCCAAGTGGCAAGCTGGACCGTCGCGCGCTGCCGAAACCGAACCTCTTT
AGCTTCAGCCAGAGTCAAGGCAGTGTGCTCCCGCGCAATGACACCGAACGTGAACTGGCCAAAATCTGGCTGGAT
ATCCTCACGATTCAGAGCGTTGGCGTGCAAGATAACTTCTTCGAGATTGGCGGCACGAGTCTGAGTGCGATTTAC
CTCATCGCCGCGATTGAGCAGCAATTCGGCAAAGAACTCCCGCTGAGTGTGCTCCTCACCAATCCGACGATCGAG
GAACTCGCCAAGGTTCTGCATCTGAGCAGCGAGCAAACGAACAATAGTCCACTCATTCCGATCCAGCCGAAGGGT
AATAAGCAGCCATTCTTCTGCGTGCACCCAGCGGGCGGCCATGTGCTCTGTTATTTCAACCTCGCCCGCTATCTG
GGCACCGATCAACCGTTCTATGGTCTGCAAGCCCAAGGCTTTAATGACGATGAGGAACCGCTCACCCGCGTGGAG
GATATGGCCAGCACGTACGTGGAAGCCATTCGTCAGTTCCAGCCGTATGGCCCGTACCAGATTGGCGGCTGGAGC
TTTGGCGGCATTGTGGCGTACGAGATCGCCCAACAGCTCCATAAACAAGGCCAAGAAGTGAGTCTGCTCGCGATT
CTGGATAGTTACGTGCCGATCATCCTCGACCCGAACAAGAAAATTGATAATAAATATCTGGTGGGCGTTCTCAGC
CGCGTGTTTGGTGGCATGTTTGGCCAAGACAATCTGGTGACCCCACAAGAAATTCAGCATCTGAGTGTGGAAGAA
CGCCTCGACTACATCATCGAGAAGGCCCGCAAAGCGAAAATCTTCCCACCGGGTGTGGAGCGTAGTAAAAATCGC
CGTATCCTCGACGTGCTGGTTGGTACGCTGAAAGCCACCTACAGCTACGTTCGCCAGCCATACCCGGGCAAAGTT
ACCGTGTTCCGCGCGAGCGAAAAGCACATCATGGCCCCGGACCCGACGCTGGTTTGGGTTGAGCTGCTCAGCGTG
ATGGCGGTGAAGGATGTGAAAGTTGTTAAAATCCCGGGCAATCACTATACCTTCATCCTCGAACCGCATGTGAAA
GTTCTGGCCGAGCGCCTCAAGAGTTGTCTGAACAGCAGCAGCCAGAATCTCGAGAGCGATAGCGTTTTCCTG
```

[4] Original sequence of NRPS-like enzyme from *Chryseobacterium indologenes*

```
MDINTETENLLIKSYLDSNYFKKNYSKETLPYLLFSQQTNYGEKIYLRDYQKELTFNDVCIESLKISLYLRLLSL
SKEPVGIYMEPSVEMGLGIWGILMACKSYIPLSLDYPEERLQYMIDNSGINYILTSRECQKKLHVFNLKNTKIIC
LEDVKSFLSEYEITNTELDFWSSEEDLAYIIYTSGSTGNPKGVMIQHKNISHQMNWLESQHNIGFSTKILHKTPI
SFDAAQWEILSGCNGSEVVIGPVGIHKNPDQITGMIKQYGINTLQGVPTLLQALTENDFFRDCKTLKCIFSGGEA
LTRKLAASMFKQLPNAKITNLYGPTECTINISSHILVKDSMEDYKDTIPIGNIVNNHSFLLLNNKGNHCSNGETG
ELYISGIQVSNGYYKNTKQTLERFVTIDDQYYYKTGDLVVQDENLHLYFVGRLDNQVKLRGYRLELDEIKINIEK
HPWVRHAAVIISENTINSTQQLVAFIELNPNEAALMDQGAENAHHHASKSDQLQVKAQLSNSGCKNDKELTGKES
VKLPGYYSDSFQNEKIFARKSYRSFEGGDVHKDEVLNCLTKILTSKIDRFIPSSYKDMSFQDLGYIMRYFGAFYS
QERILPKYGYASPGALYATQIFLETANMKDLKNGLYYYHPIMHTLFLIYEFKECSGNEKSPITLHFIGKKSAIEP
VYKNNIKEVLEFEAGHIIGLFDNVLKQFQQYIGKGYHSEEIDCVKEKLFGCRKEDYYLGSFPIIRGDRKKEKDNI
AIEILIQFNKETLDNKGIYKYINGELMPLADEKIEKNDVIAINQRIYDNAEIGVCFLNHLETWDQYIELGRLLQT
FQMNDQMMGFVSSGYSSKSGNDLKAAKKIKKILKLNDKTGVYFALGGKISPDQKYSRGMKEDTVHMAGPAEIIKQ
DLASFLPDYMLPGRIVILQSLPRTSNGKIDLIEIQNSAQEIINKRITPFLLPENETQHRLLDLWKNLLKIDEIST
KDDFFELGGNSLSALLLINRINKELKCNLPVQILFEARDIVSLSDMIIQNTNTINSSRMISLNKAQGIQPVICWP
GLGGYPLNLKPLGKEISSTPVYGFQAYGLNNNEVPYTSLEEMALEDIKILKKKYPKGPYTLVGYSFGAKLAYEVA
FQLEKMNDTVDRILLIAPGSPNHRTMNERNKESISFDDPYFVTILYSVFTRSIDDTAIKNCLEKCKDRESFIHFI
NQENRDLGMDSIRRITHIVERTFYFVYNNLSNKKLNTEIIHIKAEGDQDSFIENNINFVNKVDFFESTFDHYDIL
KEEGVTEIMKMIRFNSQKNP
```

Codon optimized sequence of NRPS-like enzyme from *Chryseobacterium indologenes* for *E. coli* (L:3885)

```
ATGGATATCAATACGGAGACGGAGAATCTGCTGATCAAGAGCTATCTGGACAGTAACTACTTTAAAAAGAATTAC
AGCAAAGAAACGCTGCCATATCTGCTGTTCAGCCAACAGACCAACTACGGCGAAAAAATCTACCTCCGCGATTAC
CAAAAGGAACTGACGTTCAATGACGTTTGTATCGAGAGCCTCAAGATTAGCCTCTATCTCCGTCTGCTCAGTCTG
AGCAAAGAGCCAGTTGGCATCTATATGGAGCCGAGCGTTGAAATGGGCCTCGGTATCTGGGGTATTCTGATGGCG
TGCAAGAGCTACATCCCACTGAGCCTCGACTACCCAGAGGAGCGTCTGCAATACATGATCGACAACAGCGGCATT
AACTACATCCTCACGAGCCGCGAGTGCCAGAAAAAACTGCATGTTTTCAATCTGAAAAATACGAAGATTATCTGT
CTCGAAGACGTGAAGAGCTTTCTGAGCGAATACGAAATTACCAACACCGAGCTGGATTTCTGGAGCAGCGAAGAA
GATCTGGCGTATATCATCTACACCAGCGGCAGTACCGGCAACCCGAAAGGTGTTATGATTCAACATAAAAATATT
```

AGCCACCAAATGAACTGGCTGGAGAGCCAACACAATATTGGTTTTAGCACGAAAATTCTGCATAAGACCCCGATC
AGCTTCGATGCCGCGCAGTGGGAAATTCTGAGTGGCTGCAATGGTAGTGAAGTGGTTATCGGCCCGGTGGGCATC
CATAAAAACCCGGATCAGATCACGGGCATGATCAAGCAGTATGGCATTAACACGCTGCAAGGCGTGCCAACGCTG
CTCCAAGCGCTGACGGAAAACGATTTTTTCCGCGATTGTAAAACGCTGAAGTGCATCTTCAGCGGCGGCGAAGCG
CTGACGCGTAAACTGGCCGCCAGTATGTTCAAACAGCTGCCGAACGCGAAAATCACCAATCTGTATGGCCCGACG
GAGTGCACGATCAACATTAGCAGCCACATTCTGGTTAAAGACAGCATGGAAGACTACAAGGACACGATTCCGATC
GGCAACATCGTGAACAACCATAGCTTTCTGCTGCTGAACAACAAGGGCAACCATTGCAGCAATGGCGAGACGGGT
GAGCTGTACATTAGCGGCATCCAAGTTAGCAACGGTTACTATAAAAACACCAAACAGACGCTGGAACGCTTCGTG
ACCATTGATGACCAATACTATTACAAAACGGGTGATCTCGTGGTTCAAGATGAGAATCTCCATCTGTACTTCGTG
GGCCGTCTGGATAACCAAGTGAAGCTGCGCGGTTACCGCCTCGAGCTGGACGAAATTAAGATCAATATCGAGAAA
CATCCATGGGTGCGCCACGCGGCCGTGATCATTAGCGAAAACACCATTAATAGCACGCAGCAGCTGGTTGCGTTC
ATCGAACTGAACCCGAACGAAGCGGCCCTCATGGATCAAGGTGCGGAGAACGCCCATCATCATGCGAGTAAAAGC
GATCAGCTGCAAGTTAAAGCCCAGCTGAGCAACAGCGGCTGCAAGAATGACAAAGAGCTCACCGGCAAGGAAAGC
GTGAAGCTGCCGGGCTATTACAGTGACAGCTTTCAGAATGAGAAGATTTTTGCGCGCAAAAGCTATCGCAGCTTC
GAGGGCGGTGATGTTCACAAGGATGAAGTTCTGAACTGTCTGACGAAGATTCTGACGAGTAAGATCGACCGCTTC
ATCCCAAGTAGCTACAAGGACATGAGCTTCCAAGATCTGGGTTACATCATGCGCTACTTCGGTGCCTTTTATAGC
CAAGAGCGCATTCTGCCAAAGTATGGTTATGCCAGCCCGGGCGCGCTCTACGCGACCCAAATTTTCCTCGAGACC
GCGAATATGAAGGATCTGAAAAATGGTCTGTACTACTACCACCCGATCATGCACACGCTGTTCCTCATCTACGAG
TTCAAAGAGTGCAGCGGTAATGAGAAAAGTCCGATCACGCTGCACTTCATCGGTAAGAAGAGCGCCATCGAACCA
GTTTATAAGAATAATATCAAAGAAGTGCTGGAGTTCGAAGCCGGCCATATCATTGGTCTGTTCGATAATGTTCTC
AAACAATTTCAACAATATATTGGCAAAGGTTATCATAGCGAGGAGATCGATTGCGTTAAGGAGAAGCTGTTCGGT
TGCCGTAAGGAAGACTATTATCTCGGCAGCTTTCCGATTATCCGCGGTGATCGTAAGAAGGAAAAGGATAATATC
GCGATCGAGATCCTCATCCAGTTCAACAAGGAAACGCTGGATAACAAGGGCATCTACAAATATATCAATGGCGAA
CTCATGCCACTCGCCGATGAAAAAATCGAGAAGAATGATGTGATCGCCATCAACCAGCGTATCTACGACAACGCG
GAAATTGGTGTGTGCTTTCTCAACCATCTCGAAACGTGGGACCAGTACATCGAGCTGGGCCGTCTGCTGCAGACC
TTTCAGATGAATGACCAAATGATGGGCTTCGTGAGCAGTGGCTACAGTAGCAAGAGTGGCAATGATCTGAAAGCG
GCCAAAAAGATCAAAAAGATTCTCAAGCTGAATGATAAAACGGGCGTGTATTTCGCCCTCGGTGGCAAGATCAGT
CCGGACCAGAAATATAGCCGCGGTATGAAAGAGGACACGGTTCACATGGCGGGTCCAGCCGAGATCATCAAACAA
GATCTGGCGAGCTTCCTCCCGGACTACATGCTGCCGGGCCGTATCGTGATCCTCCAGAGTCTGCCACGCACGAGC
AATGGCAAGATTGATCTGATCGAGATCCAAAATAGTGCCCAAGAAATCATTAACAAGCGCATTACCCCATTTCTG
CTGCCGGAAAACGAAACCCAGCATCGTCTGCTGGACCTCTGGAAAAATCTGCTGAAAATCGACGAGATCAGCACG
AAGGACGACTTTTTCGAGCTCGGTGGTAACAGTCTCAGTGCGCTGCTGCTGATCAACCGCATCAATAAAGAGCTC
AAGTGCAACCTCCCGGTGCAGATTCTGTTCGAGGCCCGTGACATCGTTAGCCTCAGCGATATGATCATCCAGAAC
ACCAATACGATCAATAGCAGCCGCATGATCAGTCTGAACAAAGCCCAAGGCATCCAACCGGTTATCTGCTGGCCG
GGCCTCGGTGGCTACCCGCTCAACCTCAAACCGCTGGGTAAGGAGATCAGCAGCACCCCAGTTTACGGTTTTCAA
GCCTACGGCCTCAATAACAACGAAGTGCCATACACGAGTCTGGAAGAGATGGCCCTCGAAGACATTAAGATTCTG
AAAAAGAAGTATCCGAAGGGCCCATACACGCTGGTTGGCTACAGCTTTGGCGCCAAGCTGGCCTATGAGGTGGCC
TTCCAGCTCGAGAAGATGAACGACACCGTGGATCGTATTCTGCTGATTGCGCCGGGTAGCCCGAATCATCGCACG
ATGAATGAACGCAACAAAGAAAGCATCAGCTTCGACGACCCATACTTTGTGACGATTCTGTACAGCGTGTTCACG
CGTAGTATCGATGATACGGCCATCAAAAACTGCCTCGAGAAGTGCAAGGATCGTGAAAGCTTCATCCATTTTATC
AACCAAGAAAATCGTGATCTCGGTATGGATAGCATCCGCCGCATCACGCACATTGTTGAACGCACGTTTTACTTT
GTGTACAATAATCTGAGTAATAAAAAACTGAACACGGAGATCATCCACATCAAGGCCGAAGGCGATCAAGATAGC
TTCATCGAAAACAACATTAATTTCGTTAATAAAGTTGACTTTTTTGAGAGCACGTTCGACCACTACGACATTCTG
AAAGAAGAAGGCGTGACGGAGATCATGAAGATGATCCGCTTTAATAGCCAGAAAAACCCG

[5] Original sequence of NRPS-like enzyme from *Moorea producens*

MDINTETETSHIQQANWFLYKFQPTGLSDKLSIAVRIKSPVDIKTIETTLQALTERHAILRSIYYEEDGQIIQKI
RDTADIYLAKIDASSWSDEELNEQLSQRIKLPFNLENGSIFRACLFTRSATDYILVLTLHQIAADWESLLILVDD
LVGIYESTINCKPPNLLPINKSYRDYIHQEVEFINSVEGKQIGKYWRENLADDLPVLELPTTSPRPSMRTYDGAA
IKFTINPELTQQLKQLVKTQGVTIEEILLAVFKVLLYRYTGEQDILVALLQKRANRLLFEQVVGNFNNVTVARQA
ISATIKFTDLLNQVRQQLFELDRYQNYPFSLLIQELKSYTLSHPPICQVAFGYSQLDKLFNAQEIDIEYYEIPQQ
KVDFELSLEVTDLQSYSLGYFKYNTDILEAETVAKIAEHFQNLLTAFVANPDTPVGKLPMLSDGEQQQIVLAWNK
TQTDYPQNKSIHQLFESQVEKTPDAVAVVFEDQELTYSELNTKANQLAHYLQKLGIGPDVQVGICVERSLEMVVG
LLGILKAGGAYVPLDSSYPSERLAYMLSDAKVAVLLTQESLVTSLPEYQGQMVCLDSDWDAIAQFSEENLSRVVK
PKNLGYIIYTSGSTGKPKGVAMSQRALVNLIMWQQQEAIIGQGARTLQFSPISFDVSFQEIFSTWYSGGILVLIS
QEVRRDPLALMQFMAQKKIDKLFLPFVALQQLAAVAPQCQNLPQLREIVTAGEQLQVTPDLIELMNRLPYCRLQN
QYGPSESHVVSAYTLQGAATSWPALPPIGRPIANNQLYILNSELQPVPIGVAGELYIGGVGVANGYHNRPELTAE
KFIPDPFGEQGESRLYKTGDRARYLRDGNIEFLGRIDNQVKIRGFRIEIGEIEATLSQHPTVKETVVVVREDNPG
NKRLVAYIVPETETTSNSELSDTQVNTEIAQQIIPQLKQHLNQKLAEYMVPSAFVVLSKLPLTPNGKVNRRSLPA
PDLSSFSRSENFVAPRDSIEQKLAEIWSQLLNINPVGVKDNFFELGGHSLLAVSLMAKIQQHLDKQLPLSTLFTS
PTIEDLANVVRQETKVSSSSLVPLQTQGTKPPFFCVHPAGGHVFYYLELSRYLGNDIPFYGLQVQGFNEGEKVFT
KVEDMADFYIKNIRDFQPEGPYQIGGWSFGGVVAFEMAQQLVQQGQEVSLLALLDPWVPILLLDPNKKIDKLYMRG
VLSRYFGGMFGITNLVTEEEIIGLNSENQIEFIIDKAEKLELFPKEATREQNRRFIDVIIGTLKATYTYKRRPYP
GKVTVFRAEEKHPHGIDPQLVWVEMYAILDVADMEVVMVPGNHFTFIQDPNLKVLAERLSSRV

Codon optimized sequence of NRPS-like enzyme from *Moorea producens* for *E. coli* (L:4014)

```
ATGGATATTAACACGGAGACGGAGACGAGCCACATCCAGCAAGCGAACTGGTTTCTGTACAAATTCCAGCCGACC
GGTCTGAGCGACAAGCTGAGCATCGCGGTGCGCATTAAAAGCCCAGTGGACATCAAGACGATTGAGACCACGCTC
CAAGCGCTGACGGAGCGCCACGCGATTCTCCGCAGTATCTACTACGAAGAAGATGGCCAGATCATCCAAAAGATC
CGCGACACGGCGGATATCTATCTGGCCAAGATCGATGCGAGCAGTTGGAGCGACGAAGAACTCAATGAACAGCTC
AGCCAACGTATCAAGCTGCCGTTCAATCTGGAGAATGGCAGCATCTTCCGCGCGTGTCTGTTTACGCGTAGTGCC
ACGGACTACATTCTGGTTCTGACGCTGCACCAGATTGCCGCCGATTGGGAGAGTCTGCTCATTCTGGTGGATGAT
CTGGTTGGCATCTACGAGAGCACGATTAACTGCAAACCGCCAAATCTGCTCCCGATCAATAAGAGCTACCGCGAT
TACATCCACCAAGAAGTTGAGTTCATTAACAGCGTTGAGGGTAAGCAGATCGGCAAATACTGGCGCGAGAATCTG
GCGGACGACCTCCCAGTTCTCGAACTGCCAACCACCAGTCCGCGCCCAAGCATGCGTACGTATGATGGCGCGGCC
ATCAAATTTACGATCAACCCGGAGCTGACGCAACAGCTGAAGCAGCTCGTGAAAACCCAAGGCGTTACGATCGAG
GAGATTCTGCTGGCCGTGTTTAAAGTGCTCCTCTACCGTTATACCGGCGAACAAGATATTCTGGTGGCGCTGCTG
CAAAAACGTGCCAATCGTCTGCTGTTCGAGCAAGTTGTTGGCAACTTCAATAATGTGACCGTTGCCCGCCAAGCC
ATTAGTGCCACGATCAAATTCACGGATCTGCTGAATCAAGTTCGCCAGCAACTGTTTGAGCTCGACCGTTATCAA
AACTACCCGTTCAGTCTGCTGATCCAAGAACTCAAGAGCTACACGCTCAGCCATCCACCAATCTGCCAAGTTGCG
TTTGGTTATAGCCAGCTGGACAAGCTGTTCAATGCCCAAGAAATCGATATTGAATACTATGAAATCCCGCAGCAA
AAAGTTGACTTTGAGCTGAGTCTGGAAGTTACCGATCTGCAGAGTTACAGTCTGGGTTATTTTAAGTACAATACC
GACATTCTGGAAGCGGAGACCGTGGCGAAAATCGCCGAGCATTTTCAGAATCTGCTGACGGCCTTTGTGGCCAAC
CCAGATACCCCAGTTGGCAAACTCCCGATGCTGAGTGATGGCGAGCAGCAGCAAATCGTTCTGGCGTGGAACAAG
ACCCAGACGGATTACCCGCAGAACAAAAGCATCCACCAGCTGTTTGAAAGCCAAGTTGAAAAGACCCCGGATGCG
GTGGCGGTTGTGTTCGAGGATCAAGAACTCACCTACAGTGAGCTGAATACGAAGGCCAACCAGCTGGCCCACTAT
CTGCAGAAACTCGGTATTGGTCCAGATGTTCAAGTTGGCATCTGCGTGGAACGCAGTCTCGAAATGGTGGTGGGT
CTGCTGGGCATTCTGAAAGCGGGTGGCGCGTACGTTCCACTGGATAGCAGCTACCCGAGTGAGCGTCTCGCGTAC
ATGCTCAGCGACGCCAAAGTGGCGGTGCTGCTGACCCAAGAAAGTCTGGTGACGAGTCTGCCGGAATACCAAGGC
CAGATGGTGTGTCTCGACAGCGATTGGGATGCCATCGCCCAATTCAGTGAGGAAAATCTGAGCCGTGTGGTGAAA
CCAAAGAATCTGGGCTACATTATCTACACCAGCGGCAGTACGGGTAAGCCGAAAGGCGTTGCCATGAGTCAACGC
GCCCTCGTTAACCTCATTATGTGGCAGCAGCAAGAAGCCATTATCGGTCAAGGCGCGCGCACGCTGCAGTTTAGC
CCGATCAGTTTCGACGTGAGCTTCCAAGAAATCTTCAGTACGTGGTACAGTGGTGGCATCCTCGTGCTGATTAGC
CAAGAAGTGCGCCGTGATCCGCTGGCGCTGATGCAGTTTATGGCCCAGAAAAAGATCGACAAACTGTTTCTGCCG
TTCGTTGCGCTCCAGCAACTGGCGGCCGTTGCGCCACAGTGCCAGAATCTCCCGCAGCTCCGTGAGATTGTTACC
GCGGGTGAACAGCTGCAAGTTACGCCGGATCTGATCGAACTGATGAATCGTCTGCCGTACTGTCGTCTCCAAAAC
CAGTACGGCCCGAGCGAAAGCCATGTGGTTAGTGCGTACACCCTCCAAGGCGCCGCCACCAGTTGGCCAGCGCTC
CCACCGATTGGTCGTCCGATTGCCAACAACCAGCTCTACATCCTCAATAGTGAGCTGCAACCGGTTCCGATTGGT
GTTGCGGGCGAACTGTACATCGGTGGTGTTGGTGTGGCCAACGGCTATCATAACCGTCCAGAACTGACGGCCGAA
AAGTTCATCCCGGACCCGTTTGGCGAACAAGGTGAGAGCCGTCTGTACAAAACCGGTGATCGTGCGCGCTATCTG
CGCGACGGTAATATTGAGTTTCTGGGTCGCATCGACAATCAAGTTAAGATCCGTGGCTTTCGCATCGAGATCGGC
GAGATTGAGGCCACCCTCAGCCAGCATCCAACGGTGAAGGAGACGGTTGTGGTTGTGCGTGAAGATAACCCGGGC
AACAAGCGTCTGGTGGCCTATATCGTTCCGGAAACGGAGACGACCAGCAACAGTGAGCTCAGCGATACGCAAGTT
AACACCGAGATCGCCCAGCAGATCATTCCACAGCTGAAACAGCATCTGAACCAGAAACTGGCGGAGTACATGGTG
CCGAGCGCCTTTGTGGTGCTCAGCAAACTGCCGCTGACGCCGAATGGTAAAGTGAATCGCCGCAGTCTCCCAGCC
CCGGATCTGAGCAGCTTTAGTCGCAGTGAGAACTTCGTTGCGCCGCGCGATAGCATCGAACAGAAGCTGGCGGAA
ATCTGGAGCCAGCTGCTCAACATCAATCCGGTGGGTGTTAAGGACAACTTCTTTGAGCTGGGCGGTCATAGTCTG
CTCGCCGTGAGTCTGATGGCCAAGATTCAGCAGCATCTGGATAAACAGCTCCCGCTCAGCACGCTGTTCACGAGC
CCAACCATTGAAGATCTGGCGAACGTTGTGCGCCAAGAAACGAAAGTGAGCAGTAGCAGCCTCGTGCCACTGCAG
ACGCAAGGTACCAAACCGCCATTCTTCTGCGTTCATCCGGCGGGTGGCCACGTGTTTTACTATCTGGAGCTCAGC
CGTTACCTCGGCAACGACATCCCATTTTACGGTCTGCAAGTTCAAGGTTTCAATGAGGGTGAGAAGGTGTTCACC
AAAGTGGAGGACATGGCGGACTTCTATATCAAGAATATCCGCGATTTCCAACCGGAGGGTCCATACCAGATCGGT
GGTTGGAGTTTTGGTGGTGTGGTGGCGTTCGAGATGGCGCAACAGCTGGTTCAGCAAGGCCAAGAGGTTAGTCTC
CTCGCGCTGCTGGACCCATGGGTGCCGATTCTGCTGGACCCGAATAAAAAGATCGATAAACTCTACATGCGCGGC
GTTCTGAGCCGCTATTTTGGCGGCATGTTCGGTATTACCAATCTGGTGACGGAAGAAGAAATCATTGGCCTCAAC
AGCGAAAATCAGATTGAATTCATCATTGACAAGGCGGAAAAGCTCGAGCTGTTCCCAAAAGAAGCCACCCGTGAG
CAGAACCGTCGCTTCATCGACGTGATCATTGGCACGCTGAAAGCCACCTACACCTACAAGCGTCGCCCATACCCG
GGCAAGGTTACGGTTTTCCGTGCGGAAGAGAAGCACCCGCATGGCATTGATCCACAGCTCGTGTGGGTTGAGATG
TACGCCATCCTCGATGTGGCCGATATGGAAGTGGTTATGGTGCCGGGCAACCATTTCACCTTCATTCAAGATCCA
AATCTGAAAGTGCTGGCCGAACGTCTGAGCAGCCGCGTG
```

[6] Original sequence of NRPS-like enzyme from *Scytonema hofmanni*

```
MDINTETEQLKEYWEKQLTGELPLLNLPSTLPRSSIKTYDCSSHIFIIKNKLTGLLKDLANSESVSLENLILAAF
KVLLYRYTNETDILVALPLNQQYRPDWERFTSKVSNVILSRDFISESNSFKKFLHQVNHNVLEIKDYQDYPFALL
VKELQSKFNLSDSPICQASFTFHNGNQQENISNLFESLVKKQPENCEKIEVEFDLSLEVIQLSESLQSCFHYNSN
```

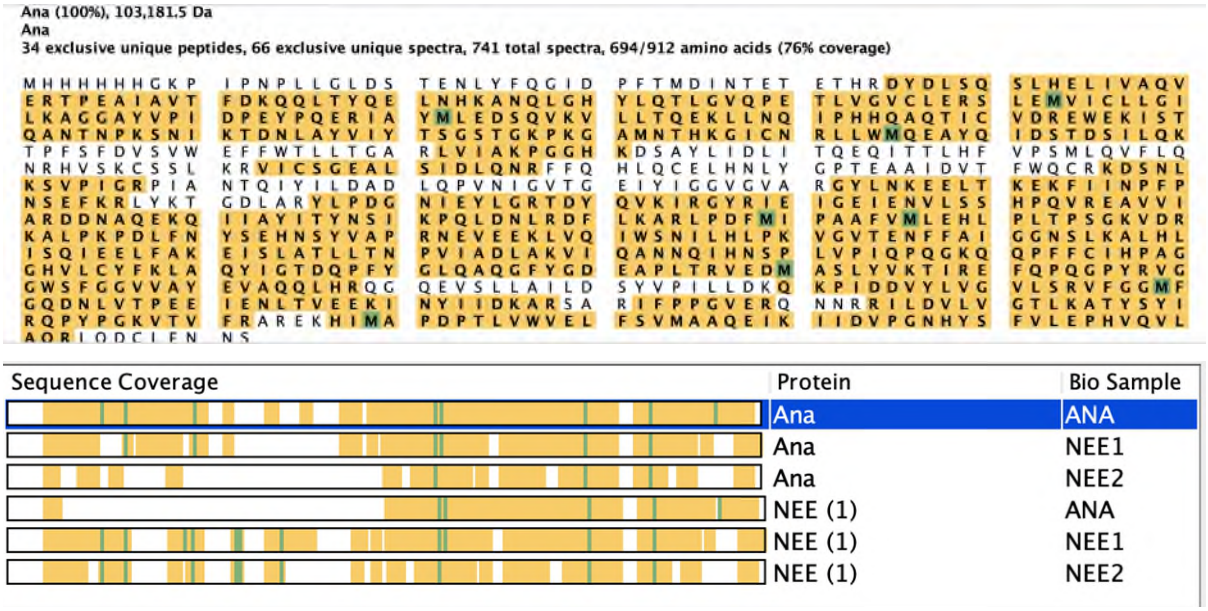
```
LLDETTVAEISRHFHNLLESIVSNPEQPVAQLPLLSESEEKKILVEWNATQTDYDLTMCLHQLIEAQVERNSNAI
AVSFAGEQLTYRELNQRANQLAHYLQTLGVKPEVLVGICIERSLEMLVGLLAILKAGAAYLPLDPRLPQERLELI
LSDSQVPVLLTDHQKLFADDEHRKVVCLRTNQENIATHSKDNPASRTTAKNLAYVIYTSGSTGKPKGVEIPHSAV
VNFLKSMQSEPGITESDVLLAVTTVSFDIAALELYLPLITGAQVVLATQEVASDGKLLKELIDTTGATIMQATPA
SWRMLLAAGWGGSPQLKILCGGEGLTSDLAQHLLEKSAAVWNLYGPTETTIWSTVSKVESAQLSHALVPIGRPIA
NTQIYILDSYLQPVPVGVIGELYIGGVGVARGYLNRPELTAERFVANHFSMGYCLYKTGDLARYLHDGNIEYVSR
IDHQVKIRGFRIELGEIENTLSAHPQVREAVVISRSDQSEEKQIVAYFTAQQQQPTPETLRDFLRQKLPDYMVPT
AFVILDALPLTPSGKVNRRALPKPTVSSFSQNNEFITPRNDTERRLEKIWSEILNIQPVGVKDNFFEIGGNSLSA
IHLIASIEQQFGKELPLSAILTNPTIEKFVNLLDTSSDTFDRSPLIPIQPKGNKQPFFCVHPAGGHVMSYFKLAQ
YLSADQPFYGLQAQGFHGEEEPLTRVEDMASLYIQAIQKFQPQGPYQIGGWSFGGVVAYEIAQQLHKQGHEVSLL
AILDSYVPILLDKNKKIDAPYLVGALSRYLGGMLGQDNLVTTDEIKHLSVDEQINYILDKAIKVKILPPSNQNRR
IVDVLVGTLKATYSYEKQPYPGKVTVFRAREKHIMAPDPTLVWVEFFSIMDAEDIKIVDVSGNHYTLILEPHLQV
LAECLKSCLEGF
```

Codon optimized sequence of NRPS-like enzyme from *Scytonema hofmanni* for *E. coli* (L:3411)

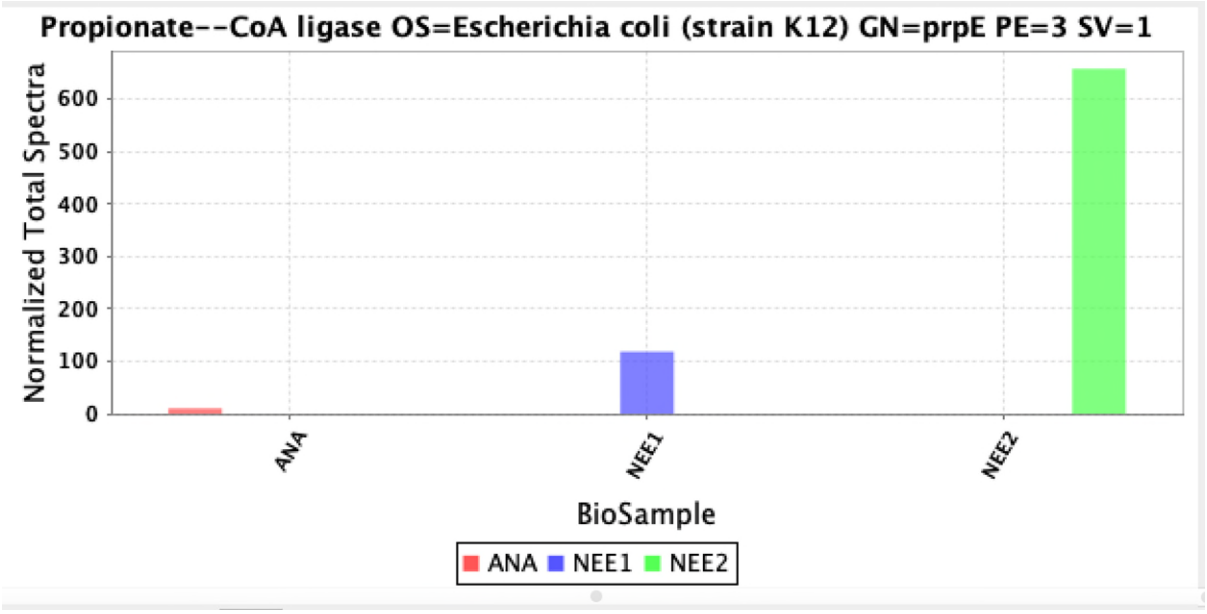
```
ATGGATATCAACACGGAGACGGAACAGCTGAAAGAGTACTGGGAGAAGCAGCTCACCGGCGAGCTGCCACTGCTC
AATCTGCCGAGCACGCTGCCGCGCAGTAGCATCAAGACCTACGATTGCAGCAGCCACATCTTTATCATCAAAAAT
AAACTGACGGGTCTGCTCAAGGATCTGGCGAACAGCGAGAGTGTGAGTCTGGAGAATCTCATTCTGGCGGCCTTC
AAGGTGCTGCTGTATCGTTACACGAACGAGACGGACATTCTGGTGGCGCTCCCACTGAATCAACAATACCGTCCG
GACTGGGAACGTTTCACCAGCAAGGTGAGCAACGTGATTCTGAGTCGCGACTTCATTAGCGAAAGCAATAGTTTT
AAAAAATTTCTGCACCAAGTGAACCACAACGTGCTGGAAATCAAGGACTACCAAGATTACCCGTTCGCGCTGCTG
GTTAAGGAGCTCCAGAGTAAATTCAATCTGAGCGACAGCCCGATCTGTCAAGCCAGCTTCACCTTTCATAATGGC
AACCAGCAAGAAAATATCAGTAACCTCTTCGAAAGCCTCGTTAAGAAACAACCGGAAAACTGCGAGAAAATTGAG
GTGGAGTTTGATCTGAGCCTCGAGGTTATCCAGCTCAGTGAGAGCCTCCAGAGCTGCTTCCACTACAACAGCAAT
CTGCTGGATGAAACGACCGTGGCGGAAATCAGCCGCCACTTCCACAATCTGCTGGAGAGTATCGTGAGCAATCCA
GAGCAGCCAGTGGCCCAACTCCCGCTGCTGAGTGAGAGCGAGGAGAAAAAGATCCTCGTTGAGTGGAATGCCACG
CAGACGGACTACGATCTCACGATGTGCCTCCATCAGCTGATCGAAGCGCAAGTTGAGCGCAACAGTAATGCCATC
GCGGTTAGTTTCGCGGGCGAGCAACTGACGTACCGCGAACTCAACCAACGTGCGAATCAGCTGGCCCACTATCTG
CAAACGCTGGGCGTTAAACCGGAGGTGCTCGTGGGCATCTGCATCGAGCGTAGTCTGGAAATGCTGGTGGGTCTG
CTCGCGATTCTGAAAGCGGGTGCCGCGTACCTCCCGCTGGATCCACGTCTCCCGCAAGAGCGTCTGGAGCTGATT
CTGAGTGACAGTCAAGTTCCGGTGCTGCTGACCGACCACCAAAAACTGTTTGCCGACGATGAGCACCGCAAGGTG
GTTTGTCTGCGTACCAACCAAGAAAACATCGCGACGCATAGCAAAGACAATCCAGCCAGCCGTACGACGGCCAAA
AATCTGGCGTACGTGATCTATACCAGCGGCAGCACGGGCAAACCGAAGGGCGTTGAAATTCCGCACAGTGCCGTG
GTGAACTTTCTGAAAAGCATGCAGAGCGAACCGGGTATCACGGAAAGCGATGTGCTGCTGGCGGTTACCACGGTT
AGCTTCGATATCGCCGCGCTGGAACTGTATCTGCCACTGATTACCGGTGCGCAAGTTGTGCTCGCCACCCAAGAA
GTTGCCAGCGACGGCAAGCTGCTCAAAGAACTGATCGATACCACCGGCGCGACGATTATGCAAGCGACGCCAGCC
AGTTGGCGTATGCTGCTCGCGGCGGGTTGGGGTGGTAGCCCGCAACTGAAAATTCTCTGCGGCGGTGAAGGCCTC
ACCAGTGATCTCGCGCAACATCTGCTGGAAAAGAGCGCGGCCGTGTGGAATCTGTACGGTCCGACGGAAACCACC
ATTTGGAGCACGGTTAGCAAGGTGGAGAGCGCCCAGCTGAGCCATGCGCTGGTGCCAATTGGCCGCCCGATCGCC
AACACCCAAATCTACATTCTGGACAGCTATCTGCAACCGGTTCCGGTTGGTGTGATTGGCGAACTGTACATCGGC
GGCGTGGGTGTTGCCCGCGGTTATCTGAACCGCCCAGAACTGACGGCCGAACGCTTCGTTGCGAATCACTTTAGT
ATGGGCTACTGTCTGTACAAAACCGGCGATCTCGCCCGCTATCTCCATGACGGCAACATTGAATACGTGAGCCGC
ATTGATCACCAAGTTAAGATCCGTGGCTTTCGTATCGAACTGGGCGAGATCGAAAATACGCTGAGTGCGCATCCG
CAAGTTCGTGAGGCGGTTGTTATCAGCCGCAGTGATCAGAGCGAAGAGAAGCAGATCGTTGCGTATTTCACCGCG
CAGCAGCAACAACCAACGCCGGAAACGCTGCGCGATTTTCTGCGTCAGAAACTGCCAGACTACATGGTGCCGACC
GCCTTCGTTATCCTCGATGCGCTGCCGCTGACGCCAAGTGGCAAGGTTAATCGCCGTGCGCTCCCAAAGCCGACG
GTGAGCAGCTTCAGCCAGAACAACGAGTTCATTACGCCGCGTAACGACACCGAGCGCCGTCTGGAAAAGATCTGG
AGCGAGATTCTGAACATCCAGCCGGTGGGCGTGAAAGACAACTTCTTCGAGATCGGTGGCAATAGTCTCAGTGCC
ATTCACCTCATCGCGAGCATCGAACAACAGTTCGGTAAAGAGCTGCCACTCAGCGCGATTCTGACCAACCCAACG
ATTGAGAAGTTCGTGAATCTGCTCGACACGAGCAGTGACACGTTTGACCGTAGCCCACTGATTCCGATCCAGCCG
AAAGGCAACAAGCAGCCATTCTTTTGTGTGCACCCGGCGGGCGGCCACGTTATGAGCTACTTCAAACTCGCCCAG
TATCTGAGTGCGGACCAACCGTTTTACGGTCTCCAAGCGCAAGGTTTCCACGGCGAGGAAGAACCACTCACGCGC
GTGGAAGACATGGCCAGCCTCTATATCCAAGCCATTCAGAAGTTTCAACCACAAGGCCCGTACCAGATTGGTGGC
TGGAGTTTCGGCGGCGTTGTTGCGTACGAAATCGCCCAGCAACTGCACAAACAAGGTCATGAGGTGAGTCTGCTG
GCCATTCTGGATAGCTACGTTCCGATTCTGCTCGACAAAAACAAGAAAATCGACGCCCCATATCTGGTGGGTGCC
CTCAGCCGTTACCTCGGTGGCATGCTGGGTCAAGATAATCTCGTTACCACGGACGAGATCAAGCATCTGAGCGTG
GATGAGCAAATCAACTACATTCTGGATAAAGCCATCAAAGTGAAAATTCTGCCGCCAAGCAACCAGAATCGCCGC
ATTGTTGACGTGCTGGTGGGCACCCTCAAGGCCACGTACAGCTATGAGAAACAGCCATATCCGGGCAAAGTTACG
GTTTTCCGTGCCCGCGAGAAACATATTATGGCGCCGGACCCAACCCTCGTGTGGGTGGAGTTCTTTAGCATTATG
GACGCGGAGGACATCAAAATTGTGGACGTGAGCGGTAATCATTACACGCTCATTCTGGAGCCACACCTCCAAGTT
CTGGCGGAATGTCTGAAAAGCTGTCTCGAAGGCTTT
```

[2] Proteomics Results

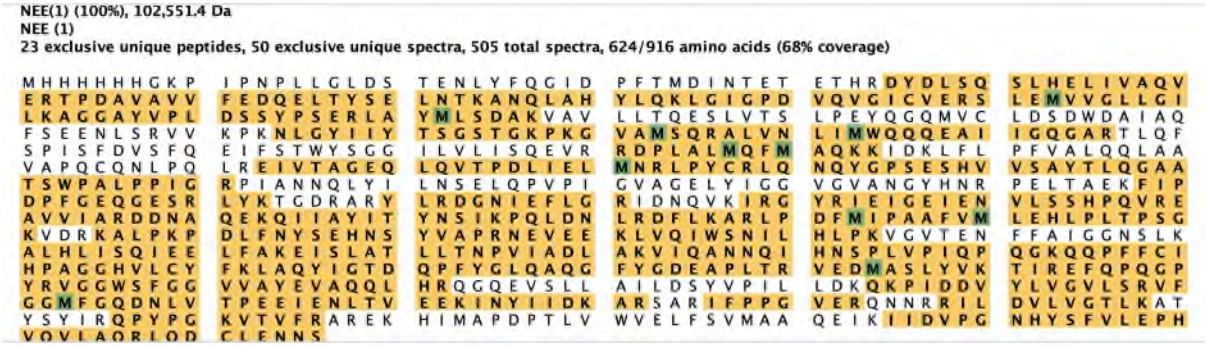
Confirmation of Ava_3855

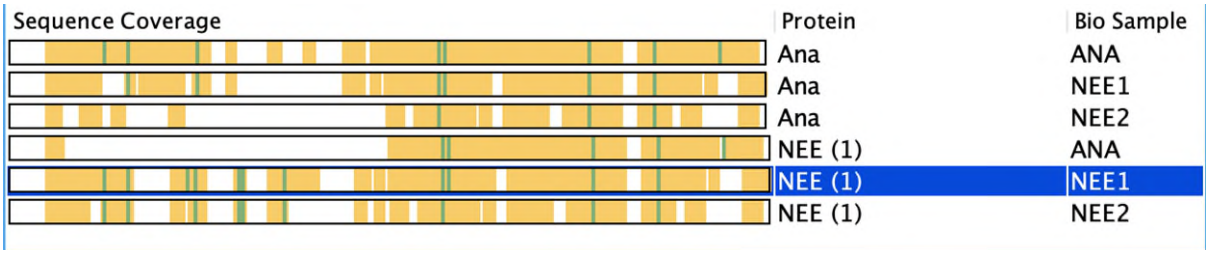


Confirmation of Propionate—CoA ligase

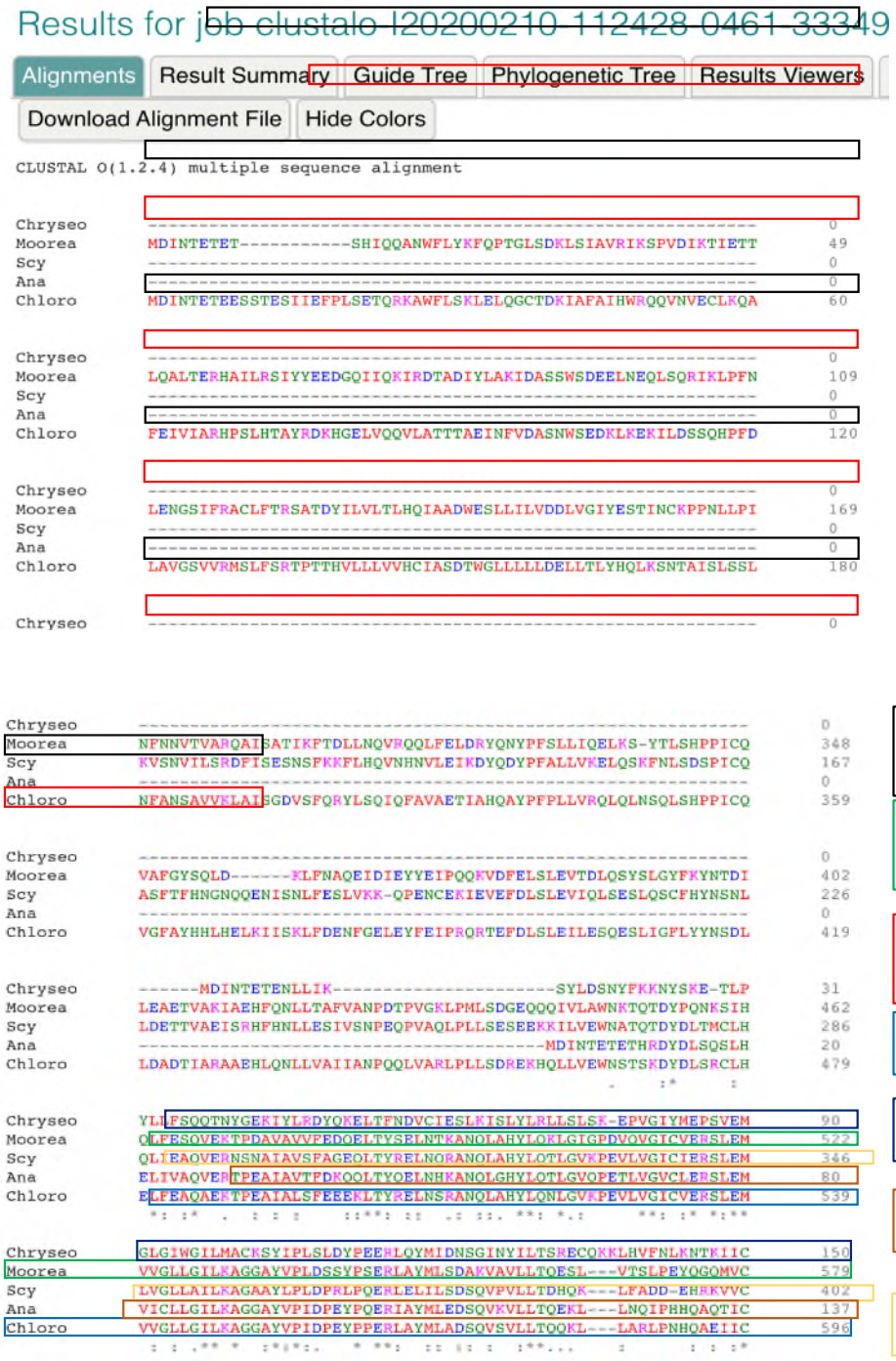


Confirmation of Engineered Enzyme

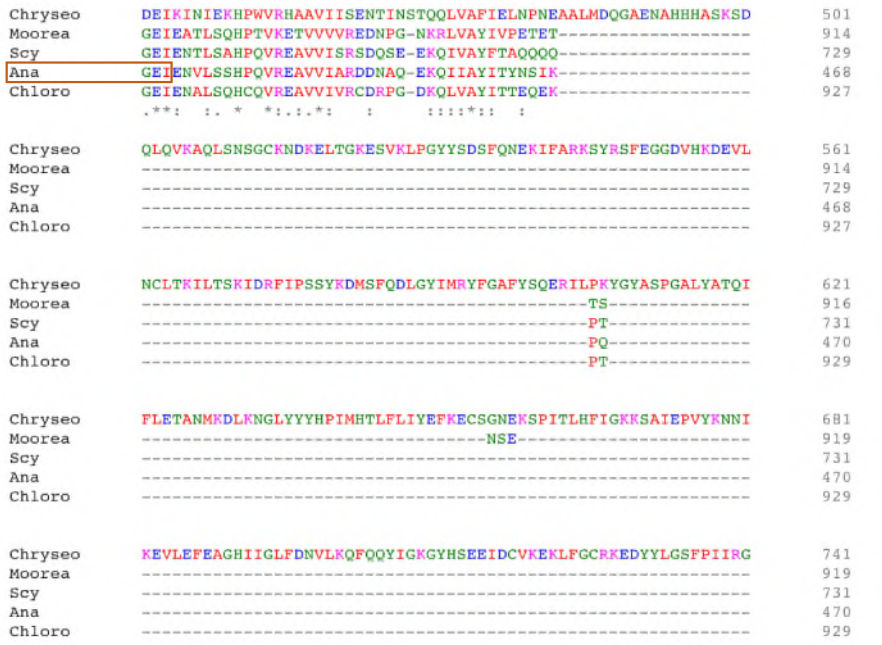




[3] Sequence Alignment using ClustalOmega and delineation of domains using antiSMASH bacterial version in NRPS-like enzyme




Chryseo
Moorea
Scy
Ana
Chloro
A- Domain for Moorea: 464-868 aa
A- Domain for Chloro: 481-881 aa
A- Domain for Chryseo: 35-435 aa
A- Domain for Ana: 33-431 aa
A- Domain for Scy: 289-683 aa


Chryseo
Moorea
Scy
Ana
Chloro
A- Domain for Ana: 33-431 aa


Chryseo
Moorea
Scy
Ana
Chloro
PCP for Moorea: 994-1063 aa
TE for Moorea: 1081-1327 aa
PCP for Chloro: 990-1057 aa
PCP for Chryseo: 957-1023 aa
PCP for Ana: 540-609 aa
PCP for Scy: 792-860 aa


Chryseo
Moorea
Scy
Ana
Chloro
TE for Moorea: 1081-1327 aa