Natural Blues: Structure meets function in plant natural products

Alan Houghton

A thesis submitted for the degree of Doctor of Philosophy

University of East Anglia, Norwich, UK

School of Biological Sciences

May 2023

Abstract

Colour plays a crucial role in food manufacturing, providing product consistency, addressing off-colouring issues, and enhancing consumer appeal. However, the availability of blue food colorants is severely limited, with only a few options in the US (Brilliant blue FCF and indigo carmine) and an additional choice in Europe. Given the rising demand for naturally derived additives, confectionery companies have struggled to find a suitable naturally derived substitute. Anthocyanins, plant-based pigments that exhibit a range of colours from red to blue, have been explored as potential food colorants. However, their inherent instability at the pH required for generating blue colours has presented a significant challenge. The complex mixture of anthocyanins derived from butterfly pea consists of 15 delphinidin-3(6"-malonyl), 3',5'-O-triglucoside derivatives acylated with p-coumarate which are highly stable and produce a blue colour at mildly acid pH. In this study, a new flash chromatography method for the rapid fractionation of ternatins was developed, allowing for efficient and food safe separation based on the terminal sidechain residues. This method offers a high capacity and quick separation technique for food applications. Through this research, it was demonstrated empirically that acyl residues stabilised the anthocyanin against degradation by stacking over the delphinidin C-ring. Glycosylation weakens these sidechain interactions, thereby decreasing stability. In both cases, acylation and glycosylation resulted in a bathochromic shift. The mechanisms by which acylation and glycosylation cause bathochromic shifts were explored using a combination of molecular dynamics and quantum mechanical simulations. Quantum mechanics simulations of delphinidin-3-O-glucoside rotamers have demonstrated that UV-VIS absorbance λ_{max} increases with an increasing dihedral angle. Molecular dynamics simulations of ternatins showed decreased interaction between the acyl residue and chromophore upon glycosylation, with an associated increase in dihedral angle. This thesis establishes ternatins as potential 'natural' replacements for synthetic blue food colourants.

Acknowledgements

Firstly, I'd like to express my heartfelt gratitude to my supervisor Prof Cathie Martin, for giving me the opportunity to work on this project. I am truly grateful for their support, throughout the course of this undertaking. Their guidance and encouragement pushed me to always strive for excellence and have been instrumental in my growth as a researcher. I would also like to thank the John Innes Centre and the UKRI-BBSRC Norwich Research Park Biosciences Doctoral Training Partnership for their services and financial support which has allowed this work to happen.

I'd also like to thank Dr Martin Rejzek, Dr Lionel Hill and Dr Paul Brett for providing the necessary resources and facilities for my project, as well as for your advice when things inevitably went wrong.

I've been privileged to meet and work with so many amazing people during my studies, but I'd like to give special thanks to Matt, Ruoxi, Brenda and Yang for making our lab a great environment to work in. Thanks to Tom, Jonny, Ocean, Connor and Matthew for being so supportive and providing relief when things were tough.

To Lauren, thank you for putting up with me. Your kindness, love and support throughout the years have been a constant source of encouragement.

Author's Declaration

The research described in this thesis was conducted at the John Innes Centre between October 2018 and May 2023 under the supervision of Prof Cathie Martin and Dr Martin Rejzek from the John Innes Centre. No part of this thesis has previously been submitted for a degree at this or any other academic institution. All data described here are original and were obtained by the author unless except in the cases outlined below and attributed in the text.

Sections from the general introduction in **Chapter 1** were previously published in:

Natural Blues: Structure Meets Function in Anthocyanins, Houghton A, Appelhagen I, Martin C., Plants. 2021; 10(4):726.

List of Tables

List of Figures

List of Abbreviations

Å	-	Angstrom
ACN	-	Acetonitrile
AM1	-	Austin model 1
B3LYP	-	Becke-3-parameter Lee-Yang-Parr
BCC	-	Bond charge correction
CD	-	Circular dichroism
CPCM	-	Conductor-like polarisable continuum model
CT	-	Charge transfer
fs	-	Femtosecond
GAFF	-	General atomic force field
GPU	-	Graphics processing unit
HF	-	Hartree-fock
HOMO	-	Highest occupied molecular orbital
HREX	-	Hamiltonian replica exchange
K	-	Kelvin
MD	-	Molecular dynamics
MeOH	-	Methanol
mM	-	Millimolar
nm	-	Nanometre
NMR	-	Nuclear magnetic resonance
NOE	-	Nuclear overhauser effect
NPT	-	Constant pressure and temperature
ns	-	Nanosecond

NVT	-	Constant number of particles, volume and temperature
ps	-	Picosecond
PVPP	-	Polyvinylpolypyrrolidone
RESP	-	Restrained electrostatic potential
TD-DFT	-	Time-dependent density functional theory
TIP3P	-	Transferable intermolecular potential with 3 points
UUF	-	Universal force field
μL	-	Microlitre
μM	-	Micromolar

Table of Contents

Chapter 1

General Introduction

Chapter 1: General Introduction

1.1 Introduction

Throughout history, our culture has been intrinsically linked to the colours of pigments and dyes accessible to us (Barnett, Miller and Pearce, 2006). Colour is one of the primary visual indicators of food spoilage, so it comes as no surprise that we have evolved with innate psychological responses to abnormally coloured food; effects so strong that food not matching colour expectations can taste less appealing than otherwise identical, "colour-correct" foods (Spence, 2015).

To aid in product uniformity, and to correct for colour-loss upon storage of food items, synthetic colourants have been used historically in industrial applications to colour or colour-correct foods . Of the synthetic blue pigments approved for use in food products, only three exist, of which all are derived from petrochemical products: brilliant blue FCF, indigo carmine and patent blue V (**Figure 1.1**). The safety of these colourants has been questioned, starting with the Food and Drug Act of 1906 in the USA, and more recently promoting a shift towards natural sources of pigment.

With recent trends towards use of natural pigments in food, manufacturers have invested heavily in developing replacements for synthetic colourants. The replacement of all colours with natural alternatives has proved challenging however, and natural blue pigments have proved particularly elusive (Newsome, Culver and Van Breemen, 2014), due to their increased sensitivity to light, temperature, and oxygen, as well as their low water solubility compared to synthetic dyes.

To date, only one natural pigment is readily available for use as a blue food colourant—phycocyanin (**Figure 1.1**), derived from "Spirulina", which is composed of a mixture of three species of cyanobacteria, namely*, Arthrospira platensis*, *Arthrospira fusiformis*, and *Arthrospira maxima*.

Brilliant blue

Patent V

Indigo Carmine
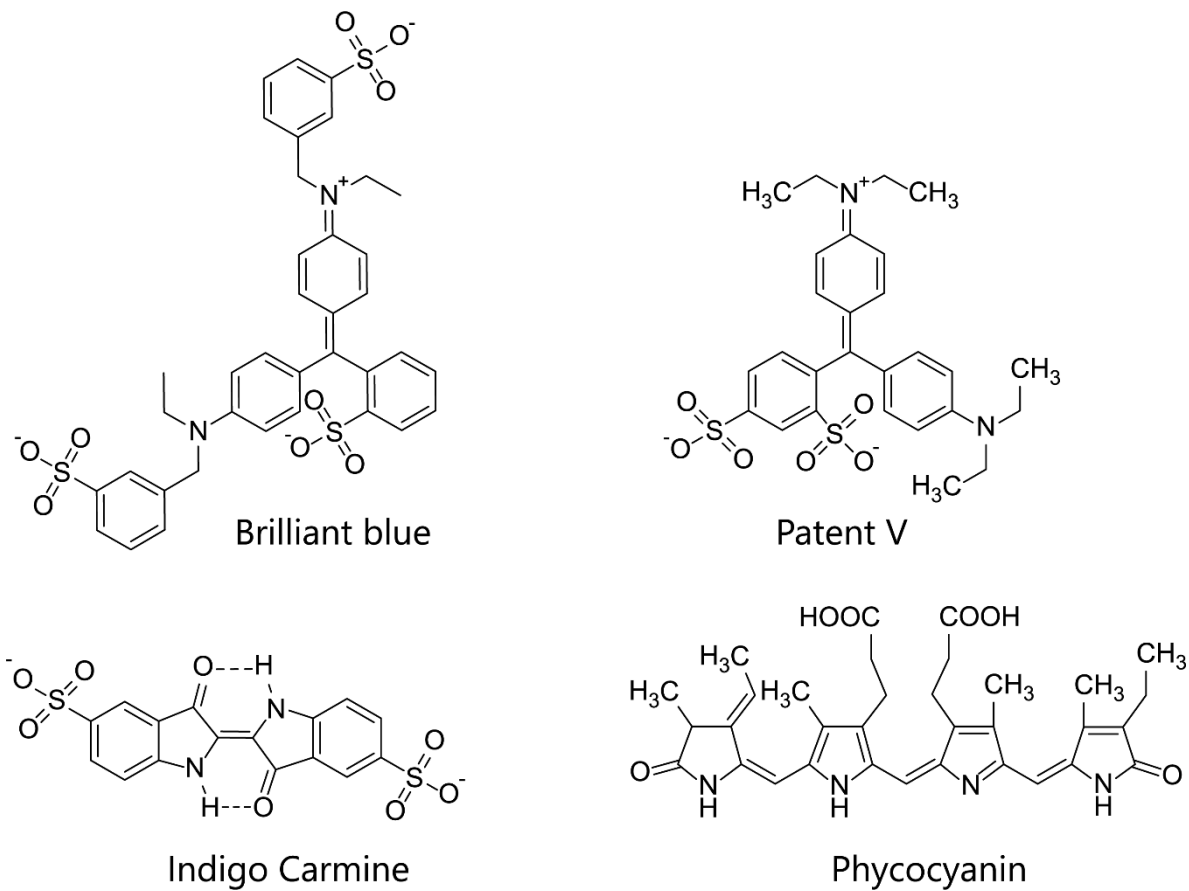
Phycocyanin

Figure 1.1 Chemical structures of European Food Safety Authority approved blue food colourants; brilliant blue, patent blue V, indigo carmine and phycocyanin

1.2 Availability of natural pigments

Plants produce an array of natural colourants, which can be broadly separated into four major groups (**Figure 1.2**): chlorophyll (green), carotenoids (yellow – red), betalains (yellow and red), and flavonoids (yellow – blue). Plants utilise these pigments to perform essential functions such as photosynthesis, radical scavenging, and attracting pollinators. Many fruits and vegetables are coloured with red and orange pigments, which have the greatest contrast against green foliage. This increases the likelihood they will be seen and eaten, allowing their seeds to be dispersed widely.

This seed dispersal strategy is so effective that three unique red pigments have evolved: lycopene (a carotenoid found in fruits such as tomatoes), betacyanins (betalains found in some of the *Caryophyllales*) (Chandran *et al.*, 2014), and anthocyanins (such as the major pigment in strawberries, pelargonidin 3-O-glucoside) (Lin-Wang *et al.*, 2014).

In contrast, natural blue pigments are rare in nature. Despite the appearance of blue colouration on many animals, only the blue of one species of butterfly, *Nessaea obrinus*, is the result of pigmentation. Instead, many apparently blue organisms create multi-layered nanostructures to reflect shorter wavelength light selectively, giving them their iridescent blue hues (Moyroud *et al.*, 2017). Unlike pigments however, structural colour is iridescent and will change in appearance, depending on the viewing or illumination angle.

Despite the wide variety of pigments synthesised by plants, few blossoms or fruits are blue. In fact, some estimates suggest only 10% of angiosperm species include blue varieties (Lee, 2007). A major difficulty in producing blue pigments arises, in part, due to how pigments are coloured in the first place.

Chlorophyll a
(Chlorophyll)

Betacyanin
(Betalain)

Cyanidin
(Anthocyanin)

Lycopene
(Carotenoid)
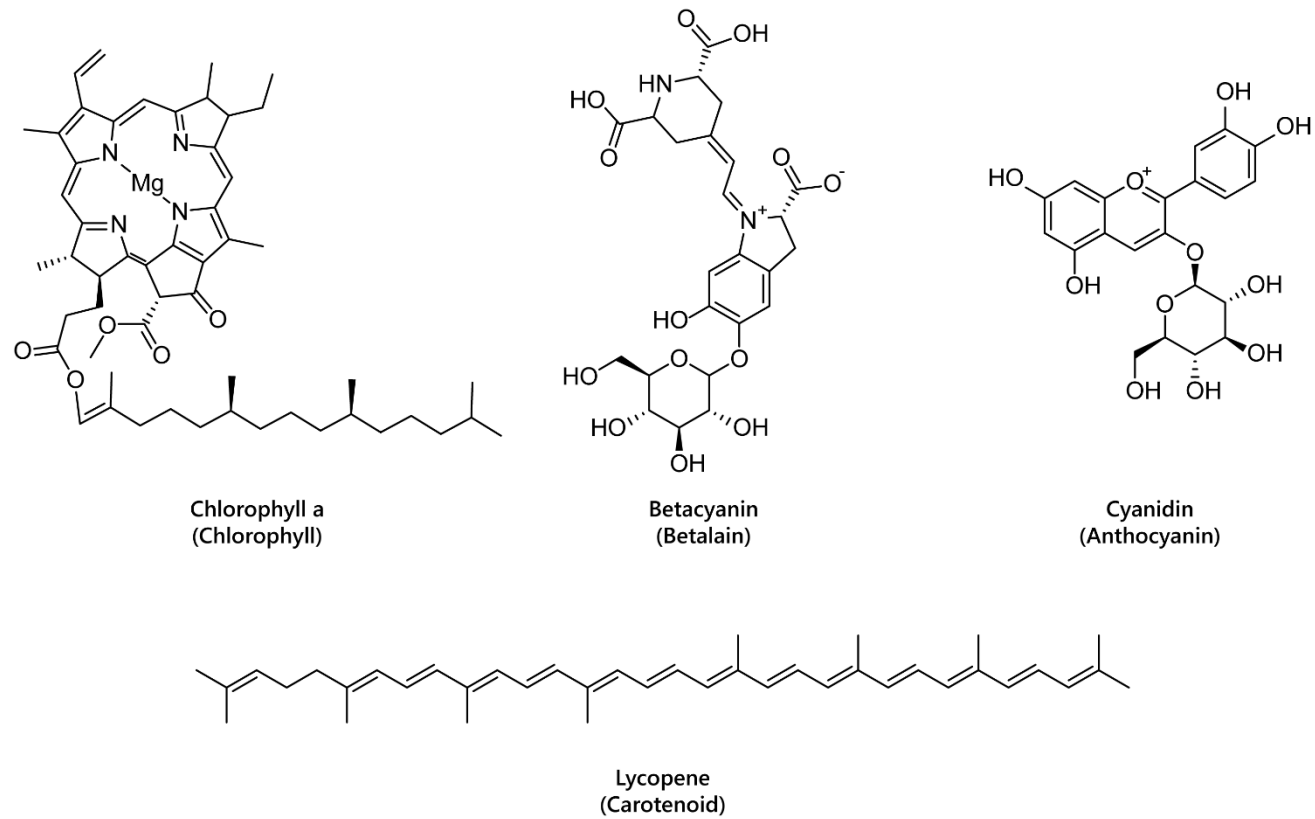
Figure 1.2 Chemical structures of plant derived pigments; chlorophyll a, betacyanin, lycopene and cyanidin

1.3 The difficulty of producing blue pigments

Electrons in the highest occupied molecular orbital of chromophores (HOMO) can absorb a discrete amount of energy (corresponding to a specific wavelength of light) which promotes an electron to the lowest unoccupied molecular orbital (LUMO) (Englman and Jortner, 1970). The light reflected is a mixture of all unabsorbed wavelengths, which creates the colours we see. Alternating π bonds increase the delocalisation of electrons, reducing the energy an electron requires to move from the ground state to the excited state (Meier, Stalmach and Kolshorn, 1997). As the HOMO-LUMO energy gap shrinks, longer wavelength (lower energy) light is absorbed, causing a bluing in colour. To produce true blues, the absorbed wavelength of light is relatively long at between 580–620 nm.

Compounds with the conjugated systems necessary to absorb light in this range are typically large and complex, requiring multiple biosynthetic steps and heavy energy investments in their biosynthesis. Compounding this problem, blue colouration provides only a relatively niche evolutionary advantage in attracting specific pollinators (Real, 1981). Combined, these are the principal reasons that blue pigments are relatively uncommon in nature. However, blue-flowered varieties have been heavily selected by breeders and horticulturalists, increasing the range of blue flowers available to enthusiastic gardeners.

Because of this, a diverse array of natural blue shades is available in floral systems, based on pigmentation by anthocyanins. However, anthocyanins are typically unstable in solution, especially under conditions where blue colouration could potentially be observed. This is due to the rapid formation of colourless forms, and the eventual isomerisation into yellow forms, leading to browning of products. To stabilise the anthocyanin quinonoid base and generate blue colours, anthocyanins form complexes through intramolecular copigmentation, intermolecular copigmentation, and metal chelation. Many of these strategies have been adopted in nature by plants that produce blue flowers or fruit.

1.4 General Features of Anthocyanins

Anthocyanins are a subclass of flavonoids; polyphenolic derivatives consisting of a fused benzopyrylium core (C- and A-rings) with an additional phenyl group attached at C2 (B-ring) formed by condensing a C6-C3 unit (*p*-coumaroyl-CoA) with three molecules of malonyl CoA.

Twenty-seven naturally occurring anthocyanidins have been identified to date, however, six account for approximately 92% of all anthocyanins reported (Andersen and Jordheim, 2013). These six are pelargonidin, cyanidin, delphinidin, peonidin, petunidin, and malvidin (**Table 1.1**). The distributions of these six anthocyanidin classes in fruits and vegetables are cyanidin 50%, delphinidin 12%, pelargonidin 12%, peonidin 12%, petunidin 7%, and malvidin 7% (Castañeda-Ovando *et al.*, 2009).

Anthocyanidins are really only stable at very acid pH values (around pH 1); at physiological pH, they are quickly hydrated to form colourless forms (hemiketals). They are glycosylated at C3 in order to stabilise natural anthocyanidins, and these compounds are termed anthocyanins. Further decoration of the anthocyanin with carbohydrate, aromatic, and aliphatic moieties produces the wide array of anthocyanins seen in nature. It is these compounds which are responsible for almost all the orange/blue colours of floral systems.

Substitution of the B-ring with either hydroxyl or methoxy groups influences the absorbance maxima of the anthocyanidin, and hence alters their colour from orange (pelargonidin) to bluish-purple (malvidin). Blue flowers commonly contain delphinidin derivatives, although there are a few exceptions where cyanidin-based anthocyanins confer blue (Sasaki and Nakayama, 2015).

The addition of hydroxyl groups on the B-ring tends to increase the maximum absorption of the anthocyanins in the visible range (λ_{max} in nm) under specified conditions. Thus, for pelargonidin 3-glucose, cyanidin 3-glucoside, and delphinidin 3-glucoside, the respective visible λ_{max} values are 510, 530, and 543 nm in 1% HCl in methanol (Santiago *et al.*, 2014).

7

Table 1.1 General structures of common anthocyanins (Houghton, Appelhagen and Martin, 2021).
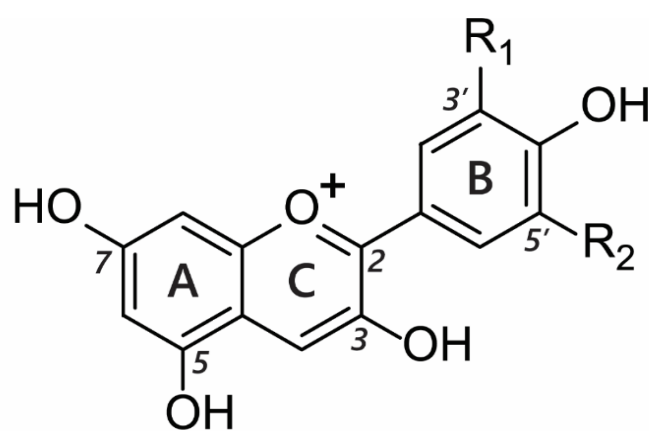
Anthocyanidin	R1	R2
Pelargonidin	H	H
Cyanidin	H	OH
Delphinidin	OH	OH
Peonidin	OCH_3	H
Petunidin	OH	OCH_3
Malvidin	OCH_3	OCH_3

Due to the electronegativity of the oxygen atom in the B-ring (**Table 1.1**), substituent groups delocalise electron density away from the aromatic centre. The additional substituent groups in the ortho and para positions of the B-ring have a lone pair of electrons present on the oxygen atom and are adjacent to the sp2 orbital of the B-ring. Electrons are donated through resonance with the C2-C1' bond, which becomes more electron-rich in the presence of *ortho* and *para* hydroxyl groups on the B-ring (Mao, Head-Gordon and Shao, 2018).

This increased electron density increases the energy of both the ground state (HOMO) and excited states (LUMO). As these interactions occur through the non-bonding orbital (which is energetically closer to the HOMO), the ground state experiences a greater magnitude of effect, effectively shrinking the transition energy gap and promoting absorption of longer wave-length light and a bathochromic shift.

1.5 Biosynthesis of anthocyanins

The flavonoid biosynthetic pathway has been well characterised using model species such as petunia (*Petunia hybrida),* snapdragon *(Antirrhinum majus)*, arabidopsis *(Arabidopsis thaliana)* and maize (*Zea mays)* (Sun, Li and Huang, 2012; Shi and Xie, 2014). The flavonoid biosynthetic pathway is a continuation from the general phenylpropanoid biosynthetic pathway and is well conserved between plants species (**Figure 1.3**).

After the formation of *p*-coumaroyl CoA, there is a stepwise condensation reaction with three molecules of malonyl CoA (arising from fatty acid metabolism). The condensation reaction produces naringenin chalcone and is facilitated by chalcone synthase, the first committed step of flavonoid biosynthesis (Zhang, Butelli and Martin, 2014). Naringenin chalcone then undergoes isomerisation mediated by chalcone isomerase, forming naringenin which is the earliest intermediate with a flavonoid skeletal core (Jiang *et al.*, 2015).

The biosynthetic pathway of flavonoids is divergent from naringenin, either undergoing a dehydration reaction catalysed by flavone synthase to produce flavones or hydroxylated by flavanone-3-hydroxylase (F3H) to from dihydrokaempferol.
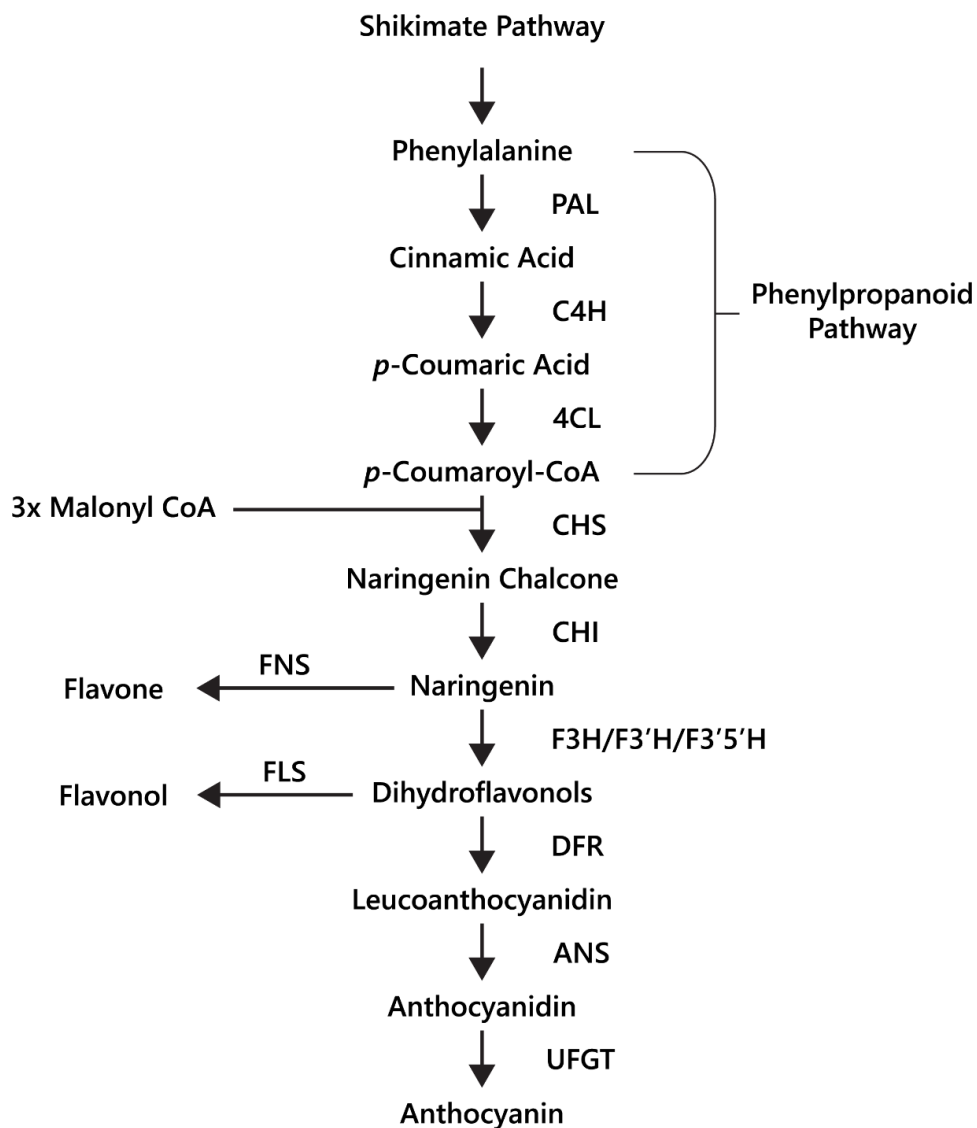
Figure 1.3 The core pathway for anthocyanin biosynthesis with related pathways for flavone and flavonol biosynthesis. The general phenylpropanoid pathway is catalysed by phenylalanine ammonia lyase (PAL), cinnamate 4-hydroxylase (C4H) and 4-coumaryol CoA ligase (4CL). The enzymes involved in flavonoid biosynthesis are chalcone synthase (CHS), chalcone isomerase (CHI), flavanone 3-hydroxylase (F3H), flavanone 3'-hydroxylase (F3'H) and flavanone 3',5'-hydroxylase (F3'5'H). Anthocyanins are synthesised by dihydroflavonol reductase (DFR), anthocyanin synthase (ANS) and one or more glycosylations initially catalysed by a UDP-flavonoid glycosyl transferase (UFGT) on the C3 hydroxyl group (Zhang, Butelli and Martin, 2014).

Dihydrokaempferol can be further hydroxylated by flavonoid-3'-hydroxylase (F3'H) forming dihydroquercetin, or by flavonoid-3',5'-hydroxylase (F3'5'H) leading to dihydromyricetin. The dihydroflavonols are then oxidised by the oxidoreductase flavonol synthase leading to the formation of flavan-4-ols (kaempferol, quercetin and myricetin), or undergo dehydration by dihydroflavonol-4-reductase (DFR) to form leucoanthocyanidins.

The final reaction is a reduction of the leucoanthocyanidin C-ring, catalysed by anthocyanidin synthase (ANS) and forming the fully oxidised heterocyclic ring system of anthocyanidins. The biosynthesis of anthocyanidins may include further decoration; mainly by glycosylation, methylation and acylation, which is particularly important for their stabilisation.

1.5.1 Enzyme localisation

Successive enzymes involved in the phenylpropanoid and flavonoid pathways have been proposed to be grouped together into weakly bound multienzyme complexes, termed metabolons, which coordinate the activity of enzymes and metabolites in shared pathways (Nakayama, Takahashi and Waki, 2019). This allows for the concentration of intermediates at the site of synthesis, preventing loss to the bulk phase of the cell cytoplasm (Winkel, 2004).

It is suggested that the flavonoid biosynthetic pathway is organised into a linear array of enzymes (**Figure 1.4**) associated with the cytoplasmic face of the rough endoplasmic reticulum (RER) (Burbulis and Winkel-Shirley, 1999). Membrane-bound P450 monooxygenases, such as cinnamate 4 hydroxylase (C4H) and flavonoid-3'-hydroxylase (F3'H) function as nucleation sites which serve to recruit other soluble enzymes into membrane associated complexes (Saslowsky and Winkel-Shirley, 2001).
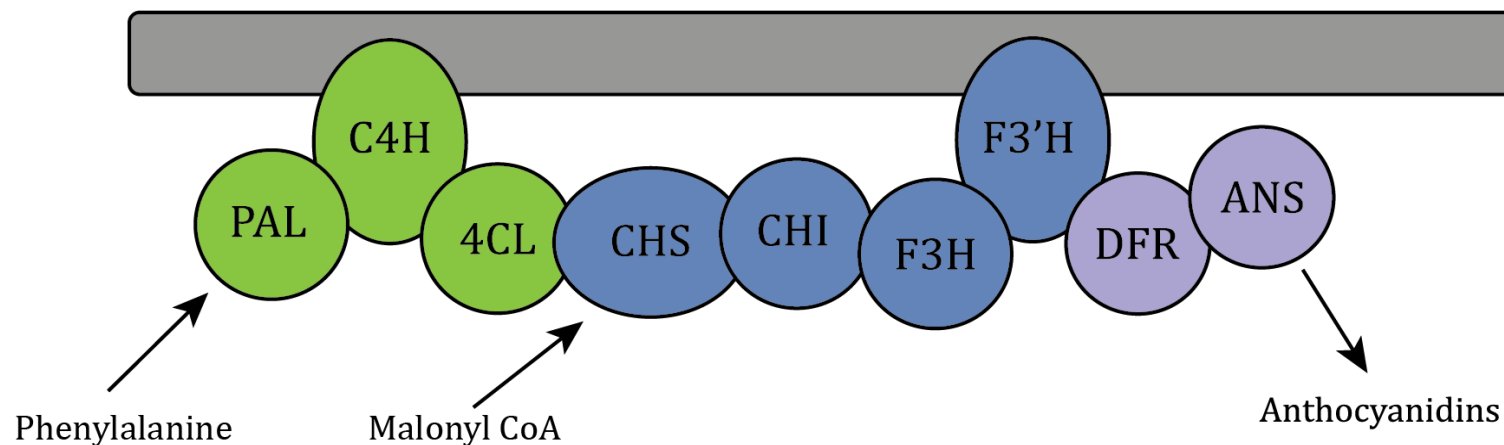
Figure 1.4 Schematic of the branch pathway of anthocyanin metabolism in *Arabidopsis*, organised as an individual metabolon. Schematic shows the enzymes required for production of 3',4'-dihydroflavonoids. F3'H is not required for 4'-hydroflavonoids, and F3'5'H (not shown in diagram) is required for 3',4',5'-trihydroflavonoids. Enzymes are; phenylalanine ammonia lyase (PAL), cinnamate 4-hydroxylase (C4H), chalcone synthase (CHS), chalcone isomerase (CHI), flavonoid 3-hydroxylase (F3H), flavonoid 3'-hydroxylase (F3'H), dihydroflavonol reductase (DFR), Anthocyanin synthase (ANS). Enzymes are coloured green for general phenylpropanoid pathway, blue for early biosynthetic enzymes and violet for late biosynthetic enzymes. C4H and F3'H are shown as membrane anchored proteins (Winkel, 2004).

Chalcone synthase and chalcone isomerase, the enzymes responsible for the first committed steps of flavonoid biosynthesis, have been localised to the RER (Hrazdina, Zobel and Hoch, 1987), which is consistent with the multienzyme complex hypothesis. Although the evidence for the localisation of other key enzymes is sparse, the multienzyme hypothesis is considered a favourable model for biosynthesis of anthocyanins (Shi and Xie, 2014).

1.5.2 Glycosylation of anthocyanins

Glycosylation is important in increasing the hydrophilicity and stability of anthocyanidins. It is arguably the most important post-biosynthetic step, as anthocyanidins are inherently unstable under physiological conditions. Mutants of red grapes missing the genes responsible for the initial glycosylation of C3-OH of anthocyanidins accumulate no anthocyanidins in their berries, despite having an intact anthocyanin biosynthetic pathway (Boss, Davies and Robinson, 1996; He *et al.*, 2010).

In plants, glycosylation is carried out by UDP-glucose : flavonoid glycosyltransferases (UF3GT), which catalyse the O-glycosylation, recognising either anthocyanins or anthocyanidins as sugar acceptors and UDP-glucose as sugar donors (Sparvoli *et al.*, 1994). A variety of other sugar donors exist, such as UDP-galactose, UDP-rhamnose, UDP-xylose and UDP-arabinose (Springob *et al.*, 2003).

UFGTs belong to family 1 glycosyltransferases, characterised by the presence of a conserved C-terminal domain termed the plant secondary product glycosyltransferase (PSPG) box (Cheng *et al.*, 2014), which is the domain responsible for binding the UDP moiety of the sugar donor. Genes encoding many UFGTs have been identified in a range of plant species, and have been divided into two groups, based on their specificity for sugar acceptors (Cheng *et al.*, 2014). Group one is responsible for the glycosylation of anthocyanidins, such as anthocyanidin 3-O-glycosyltransferase. The second group are involved in the further glycosylation of anthocyanins, such as the UDP-glucose : anthocyanin 3',5'-O-glucosyltransferase of *C. ternatea*, which glucosylates the 3' and 5' hydroxyl group of 3-O-(6"-O-malonyl)-β -glucoside) (Kogawa *et al.*, 2007).

1.5.3 Anthocyanin localisation

Although synthesis of anthocyanins has been proposed to occur at the rough endoplasmic reticulum (RER), anthocyanins need to be stored under acidic conditions within the vacuole to gain stability and function as coloured pigments. Little is known about the transport of these molecules through the tonoplast. There are currently two models proposed, the first is membrane localised transporters, and the second is vesicle mediated transport through the endomembrane system.

1.5.3.1 Transporter-mediated

Transporters can be divided into two major groups, Primary transport mediated by ATP-binding cassette (ABC) transporters, and secondary transporters which depend on a proton gradient. ATP-binding cassette (ABC) transporters couple hydrolysis of ATP to transport their substrates across the tonoplast. To date, only multidrug resistance-associated protein (MRP)-type ABC transporters have been implicated in flavonoid transport.

MRP-type ABC transporters can transport glutathione conjugates to the vacuole (Zhao and Dixon, 2010). These seemingly rely on the glutathione S-transferase (GST), which conjugates the flavonoid with glutathione (Petrussa *et al.*, 2013). Transport of flavonoids in their glutathionated form (GS-X) is inhibited by vanadate, an inhibitor of ABC-mediated transport, but not by inhibitors that dissipate the proton gradient (Kang *et al.*, 2011). However, evidence is increasing against the involvement of GST conjugation in anthocyanin localisation, as no anthocyanin-glutathione conjugates have ever been observed in plant cells. Instead, the GST protein itself may serve as a chaperone for anthocyanins, increasing their solubility and may play a role in protecting anthocyanins from oxidative degradation while escorting them to the tonoplast (Sun, Li and Huang, 2012).

Secondary transporters rely on V-type H^+-ATPase or H^+-pyrophosphate (V-PPase) proton pumps to maintain proton gradients across the tonoplast (Zhao and Dixon, 2010). Multidrug and toxic efflux (MATE) transporters then use the proton gradient to import substrate across the tonoplast. *TT12,* which encodes a MATE antiporter in *Arabidopsis thaliana* has been demonstrated as necessary for vacuolar uptake of

proanthocyanidin precursors and cyanidin-3-O-glucoside *in vitro* (Marinova *et al.*, 2007). MATE transporters of anthocyanins have also been identified in *Vitis vinifera* (grapevine) (Gomez *et al.*, 2009, 2011).

The genetic control of the vacuolar proton gradient in petals has been investigated in morning glory, where a structural gene was found which encodes a Na^+/H^+ antiporter, that is highly expressed before petal opening (Yamaguchi *et al.*, 2001). This mechanism is important in controlling the colour of Japanese morning glory, which presents reddish brown flower buds that change to a brilliant blue as the petals unfold. This colour change accompanies an increase in pH of the vacuoles from 6.6 in buds to an unusually high value of 7.7 in mature petals. The increase in vacuolar pH is conferred by a tonoplast-located Na^+/H^+ exchanger responsible for increasing the vacuolar pH by exchanging H^+ for Na^+ in vacuoles of petal cells during flower opening (Eiro and Heinonen, 2002; Yoshida, Mori and Kondo, 2009). This increased vacuolar pH has been shown to be specific to anthocyanin accumulating epidermal cells (Yoshida *et al.*, 1995).

1.5.3.2 Vesicle-mediated

The idea of vesicle-mediated transport was based on the discovery of anthocyanins and other flavonoids within discrete membrane bound structures (termed anthocyanoplasts) (Pecket and Small, 1980; Falcone Ferreyra, Rius and Casati, 2012), which may be imported into the vacuole through an autophagic mechanism (Pourcel *et al.*, 2010). Vesicles were found to be involved in the transport of flavonoids in maize cells which were induced to accumulate anthocyanins (Grotewold *et al.*, 1998). Similar transporters to those found in the tonoplast could be responsible for the loading of pigment into the ER compartment and vesicles (Petrussa *et al.*, 2013). If transporters in the vesicles and vacuole occur at a similar density per membrane unit, small vesicles would be energised more efficiently due to their large surface-to-volume-ratio (Martinoia *et al.*, 2012).

Microscopic studies have suggested that there might be multiple vesicle trafficking pathways (Lin, Irani and Grotewold, 2003). One is a *trans* Golgi network (TGN)-independent pathway which transports substrate to the cell wall, which may be

suited to the long-distance transport of flavonoids (Buer, Muday and Djordjevic, 2007).

MATE2 in the leaves and flowers of *Medicago truncatula* has been localised using fluorescence microscopy to vesicle structures. This MATE transporter shows higher transport capacity for anthocyanins in contrast to other flavonoid glycosides (Zhao *et al.*, 2011). Decoration of the anthocyanin may play an important role in targeting to the vacuole. Decoration of the anthocyanin with a malonic acid residue increases the affinity and transport efficiency into MATE2-containing vesicles (Petrussa *et al.*, 2013). A similar selectivity for cargo was reported for grape MATE transporters (Gomez *et al.*, 2009).

Direct genetic, biochemical or molecular evidence for vesicle mediated transport is still lacking. Particularly the cargo proteins and *N*-ethylmaleimide sensitive factor attachment protein receptors (SNAREs) required for the budding, trafficking, docking and fusion of vesicles associated with anthocyanin transport have yet to be identified (Zhao and Dixon, 2010).

1.6 Anthocyanin Multistate Equilibrium

All features of anthocyanin pigmentation need to be explained in the context of the multistate equilibrium which exists for anthocyanins in solution. Anthocyanins function as weak diacids due to their chromophore containing a strongly electron-withdrawing pyrilium ring, giving the phenolic hydroxyl groups at C4', C5, and C7 fairly acidic properties. Of these hydroxyl groups, the C7-OH is the most acidic (pKa1 ≈ 4), followed by the C4' hydroxyl (pKa2 > 7) for simple anthocyanins (Dangles and Fenger, 2018). As weak diacids, anthocyanins become deprotonated as pH increases which, in turn, increases electron polarisation and increases the absorption maxima towards higher wavelengths.

Anthocyanin colour therefore is the result of the ratio of the three coloured species in equilibrium at any specific pH. At very acidic pH values (around pH < 2), the formation of the red flavylium ion (AH+) is favoured (**Figure 1.5**). This species is fully protonated and has a delocalised positive charge across the chromophore. As the pH increases above pKa1, the first deprotonation occurs, converting flavylium ions into

the neutral quinonoid base (A) and the colour changes from red to purple. At pH values above pKa2, the quinonoid base is deprotonated further, forming the anionic quinonoid base (A−) with a negative delocalised charge (Basílio and Pina, 2016) (**Figure 1.5**). This deprotonation sequence is accompanied by two shifts in λ_{max}, the first (AH+→ A) typically of 20–30 nm and the second (A→ A−) a further shift of 50–60 nm (Dangles and Fenger, 2018).

Due to the instability of A (and A−), for simple anthocyanins, the conversion of AH+ to A is generally less thermodynamically favourable than hydration at C2, giving the colourless hemiketal (B), and tautomerisation of B to give colourless isomeric chalcones (Ct and Cc). As the most stable products above pH 3, species B and chalcones deplete the available pool of the coloured species, and the red or purple colours begin to disappear as the pH is increased.
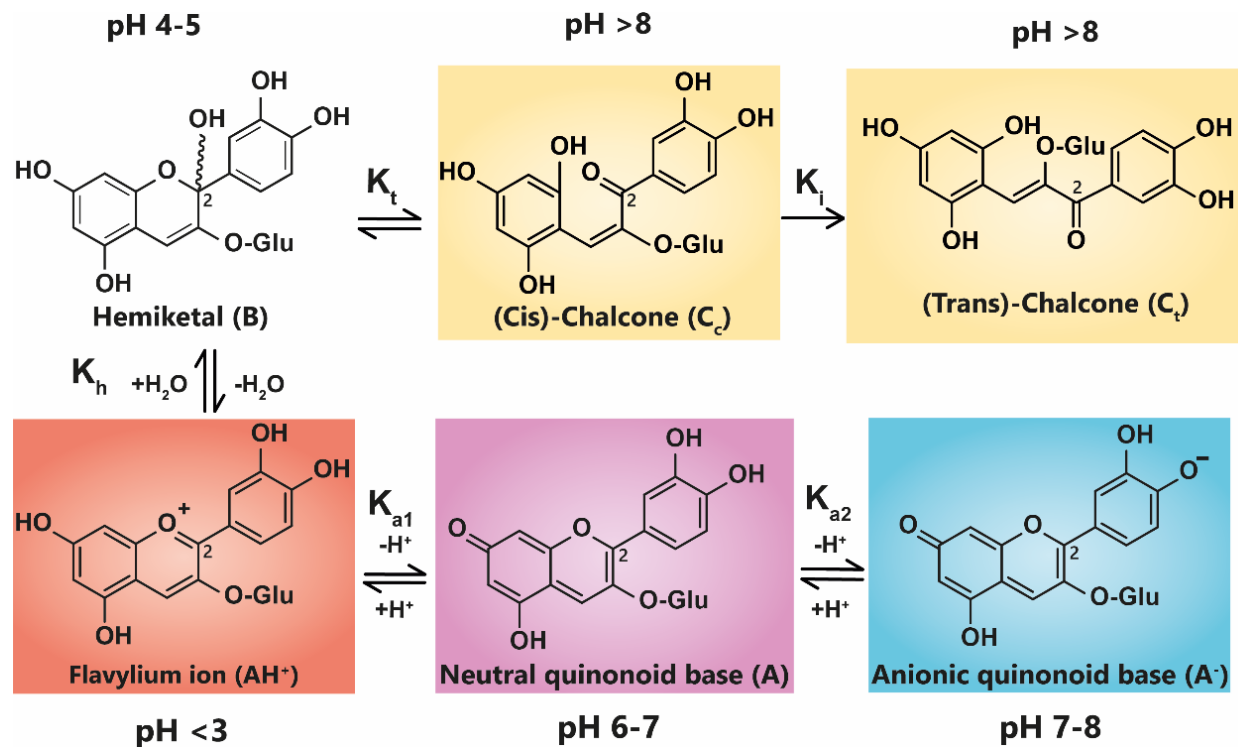
Figure 1.5 Structural transformation of cyanidin-3-O-glycoside as a function of pH. The three cyanidin species are given with their expected colours as outlined. Deprotonation from the C5-OH has been omitted for simplicity (Houghton, Appelhagen and Martin, 2021).

1.7 Copigmentation

The diversity of colours observed in floral systems cannot be explained by differences in chemical substituent groups and the pH equilibrium alone; without interactions that stabilise the chromophore at physiological pH within the vacuole (which is usually about pH 5.5), most flowers would be devoid of their attractive hues and would instead be colourless or pale yellow (as a result of hemiketal and chalcone formation, respectively; **Figure 1.5**). Anthocyanins are large, planar compounds which lend themselves to the formation of non-covalent interactions with molecules termed copigments (**Figure 1.6**). Generally, these interactions are favourable towards the stabilisation of the coloured species through blocking hydration of the flavylium cation to the colourless hemiketal by the addition of water at C2 (He *et al.*, 2019). These interactions occur with the highest affinity between the coloured forms of the anthocyanin species and the copigment, and less so with the hemiketal and chalcone forms due to decreased conjugation of the tricyclic core upon ring opening (Dangles and Fenger, 2018) (**Figure 1.6**). Observations from the anthocyanins found typically in red wines have demonstrated these copigmentation interactions produce the largest hyperchromic shifts at around pH 3.6, or when A is the dominant species (Brouillard *et al.*, 1991).
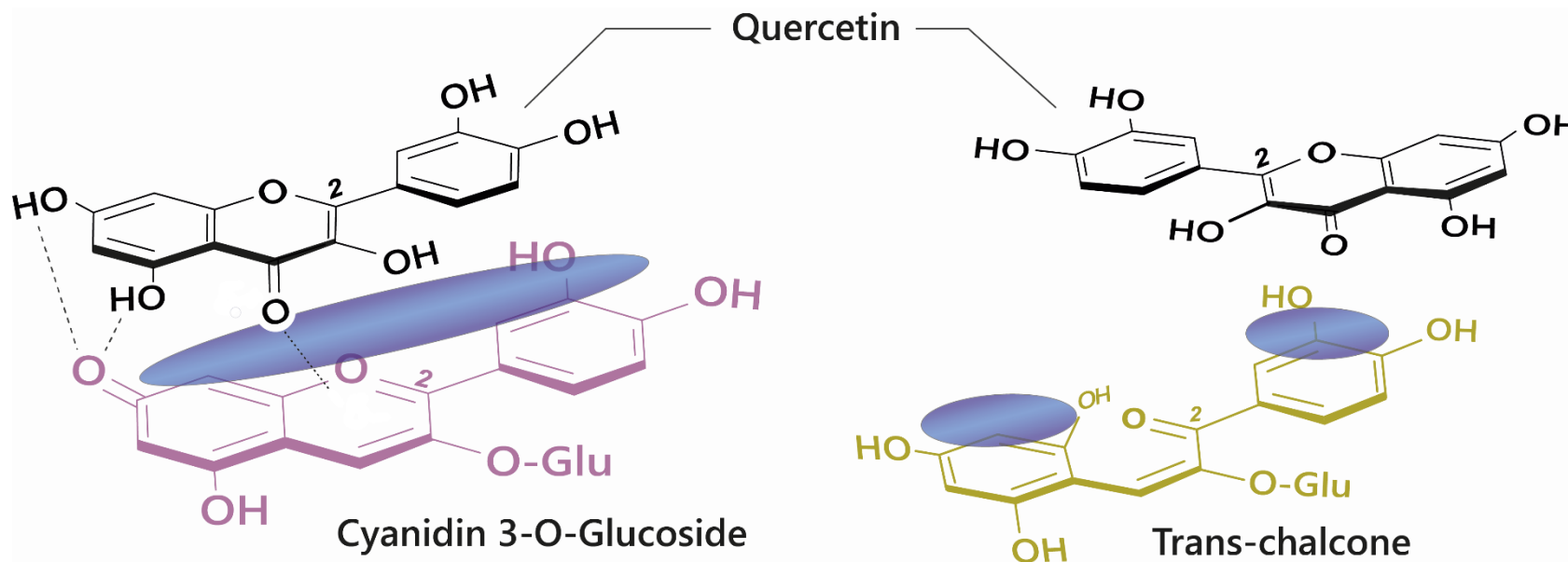
Figure 1.6 Schematic of an anthocyanin : copigment interaction (Houghton, Appelhagen and Martin, 2021). **Left**: Co-pigmentation interaction between cyanidin 3-O-glucoside and quercetin.lue clouds above the structures denote polarisable surface area. Direct electrostatic interactions form between substituent groups, stabilising the coloured species. **Right**: Ring opening breaks the conjugated system across the tricyclic ring, significantly reducing the planar surface for copigmentation.

1.7.1 Self-Association

Anthocyanins can also form self-association interactions with other anthocyanins. However, these complexes are typically less strongly associated than intermolecular copigmentation, requiring a relatively high concentration of anthocyanins before they are observed (around 1 mM) (González-Manzano *et al.*, 2008). Self-association interactions are strongest between neutral species and are destabilised by charge repulsion between pairs of positively charged flavylium cations and between negatively charged anionic bases (Dangles and Fenger, 2018) (**Figure 1.7**).

This self-associative behaviour is dependent on the substitutions present on the B-ring. It has been shown that the self-association constant decreases in the series delphinidin 3-O-glucoside > pelargonidin 3-O-glucoside > cyanidin 3-O-glucoside. The addition of a hydroxyl group to the C3' position introduces slight torsion into the B-ring: C-ring linkage, reducing the planarity and thus the proximity at which the cyanidin molecules can associate.

The opposite effect is observed with methylation where self-association increases from pelargonidin 3-O-glucoside < peonidin 3-O-glucoside < malvidin 3-O-glucoside. The steric hindrance of the methoxy groups on the B-ring may dictate the orientation of the complex, favouring associations with better alignment.

Decoration with sugars also influences the self-association stacking arrangement. Although not aromatic, the sugars are heteroatoms, with electron density polarised towards the oxygen. This property may induce further polarisation in the interacting molecule, strengthening the association, but the position and size of the sugar determines the overall alignment of the complex. Glycosylation at C3 and C5 of delphinidin increases the propensity for self-association, likely due to enhanced polarisation and the presence of additional hydrogen bonds. In contrast, this effect is reversed for malvidin 3,5-O-diglucoside, in which the addition of the C5 glucose sterically hinders self-association.
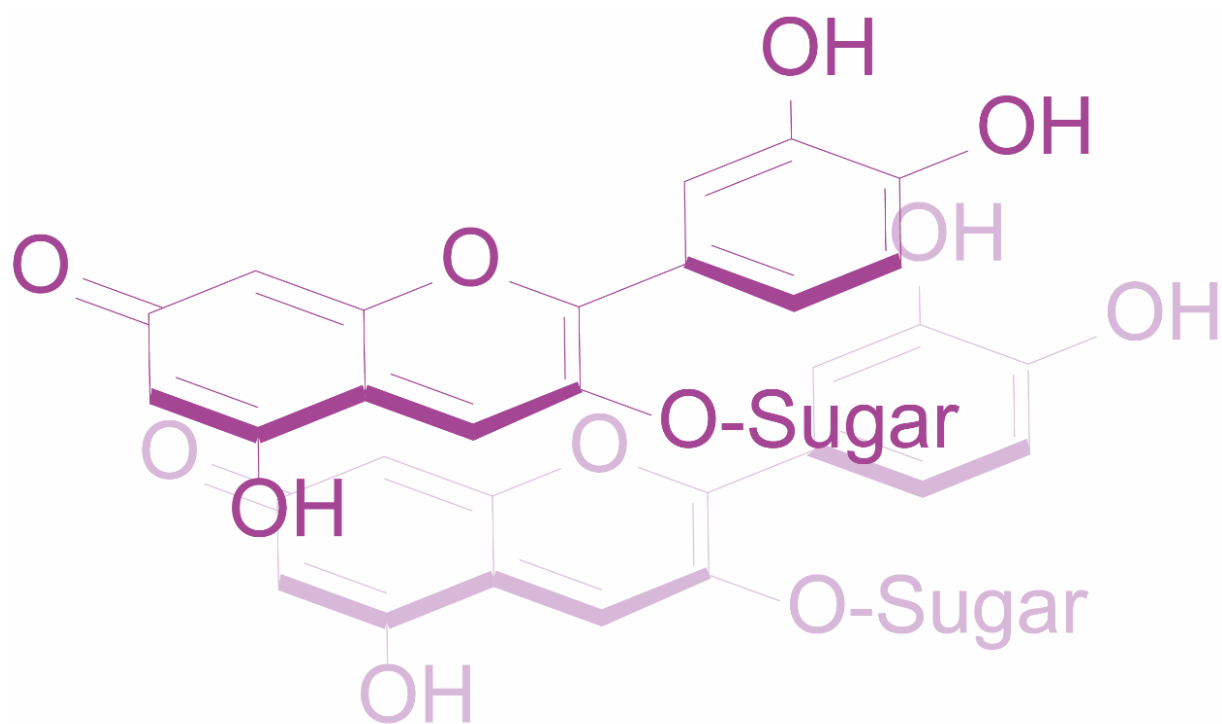
Figure 1.7 Self-association interaction between two cyanidin-3-O-glucoside quinonoid bases (Houghton, Appelhagen and Martin, 2021).

1.7.2 Effect of Copigmentation on Colour

The effects of copigmentation can be divided into three categories: the first two involve the selective stabilisation of the coloured species which manifest as an hyperchromic shift (**Figure 1.8, A**). The first effect is due to favourable interactions between the coloured anthocyanin species and the copigment, which depletes the available pool of free unbound anthocyanins and drives the equilibrium towards the formation of more coloured species (Lambert *et al.*, 2011). Secondly, copigmentation enhances the resistance of the chromophore to hydration at C2 by steric hinderance. This increases colour half-life and reducing colour fading by reducing reversible hydration, as well as the irreversible hydrolysis of the hydrated anthocyanins (Li *et al.*, 2018).

The third effect of copigmentation is the bathochromic shift (**Figure 1.8, B**), caused by distortions in the electron orbitals between closely interacting aromatics (Silva *et al.*, 2016). Light is an electromagnetic wave, and relative to the strength of the electromagnetic field, atomic nuclei are too dense to experience any substantial polarisation. However, electrons which are comparatively small experience a polarising effect when exposed to electromagnetic fields.

As electrons carry a negative charge and oscillate in electromagnetic fields (harmonic oscillation, they induce dynamic polarisation in the molecular orbitals, termed the dipole moment. Electrons which oscillate at the same frequency as an incident photon (termed degenerate) can absorb energy from those photons, promoting them from the highest occupied molecular orbital (HOMO) to the lowest unoccupied molecular orbital (LUMO). This additional energy is radiated to the surroundings within picoseconds through a number of mechanisms as the electron moves back into the ground state (Dangles and Fenger, 2018)
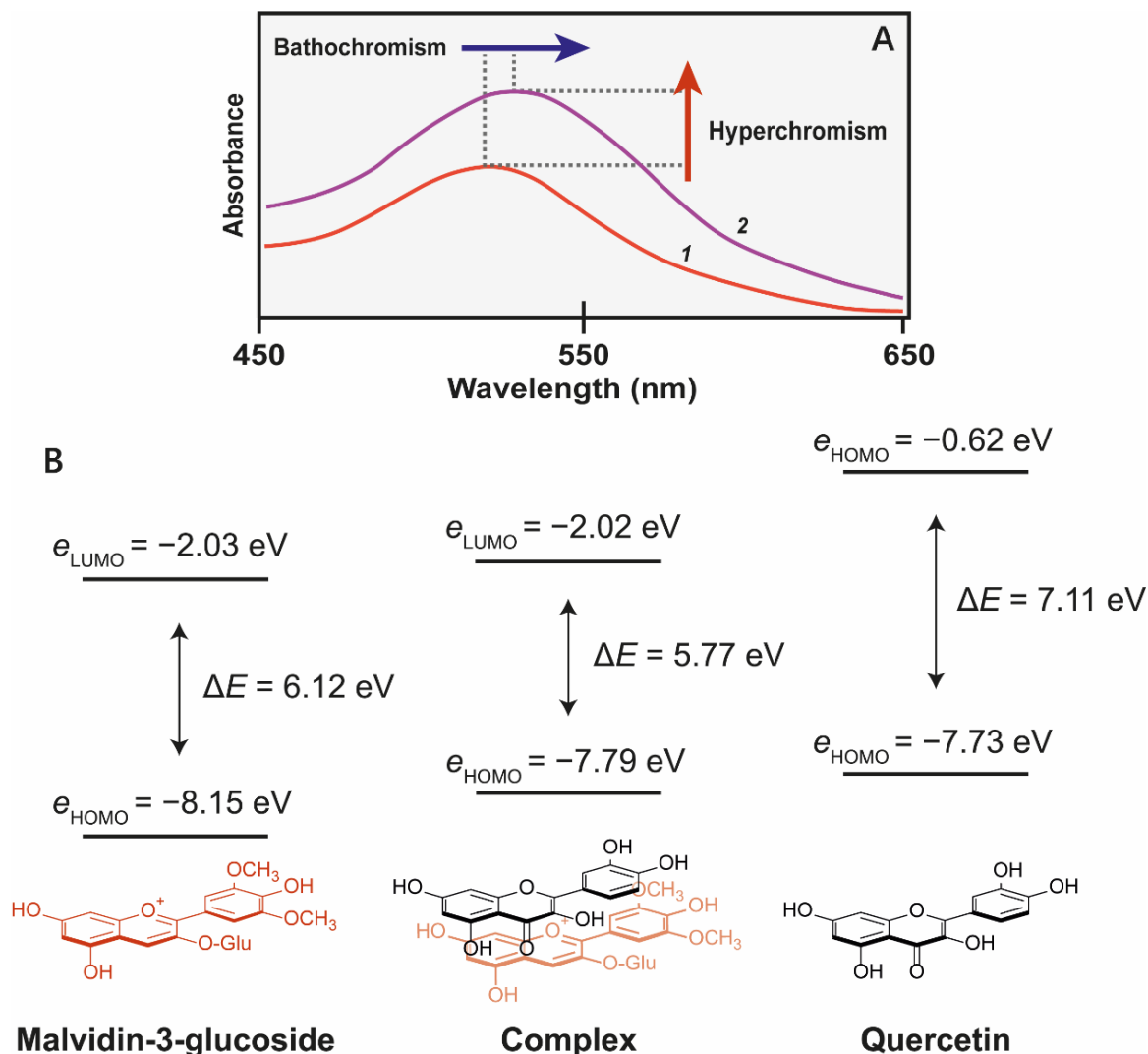
Figure 1.8 Effects of copigmentation interaction between malvidin 3-glucoside and quercetin (Houghton, Appelhagen and Martin, 2021). (**A**) Co-pigmentation of malvidin 3-glucoside with quercitin at pH 3.6. 1 = 500 μM malvidin 3-glucoside, 2 = 500 μM malvidin 3-glucoside + 1 mM quercitin (Lambert et al., 2011). (**B**) The molecular orbital correlation diagram of malvidin 3-O-glucoside and quercetin shows the energy changes as a result of co-pigmentation complex formation. Reduction of highest occupied molecular orbital of chromophores (HOMO)–lowest unoccupied molecular orbital (LUMO) energy gap (ΔE) results in a bathochromatic shift (Li et al., 2018).

The changes in electron density between the excited and ground state cause a change in the molecule's dipole moment, and therefore the energy absorbed corresponds to the frequency of oscillation between these two different dipole moments, termed the transition dipole moment (μ_{ge}, **Figure 1.9, A**). The transition dipole is proportional to the polarisability (α) of the compound (**Figure 1.9, C**), and therefore distortion of the chromophore electron orbitals by the copigment decreases the energy gap, resulting in a bathochromic shift diagrammed in figure 1.8.

In particularly strongly associated pigment : copigment complexes, interactions result in the formation of a through-space charge transfer excitation state. This occurs when there is partial mixing of the π orbitals of the two interacting molecules. After photoexcitation to a higher energy state, the molecule then rearranges the electron density to a lower energy state. This reduces the excitation energy gap between HOMO and LUMO, and increases the absorption wavelength maxima, shifting the colour towards blue (Li *et al.*, 2018). In most charge transfer interactions, orbital mixing is relatively small, and so the charge transfer band appears as a shoulder peak (or broadening of the main absorption peak as seen in **Figure 1.8**).
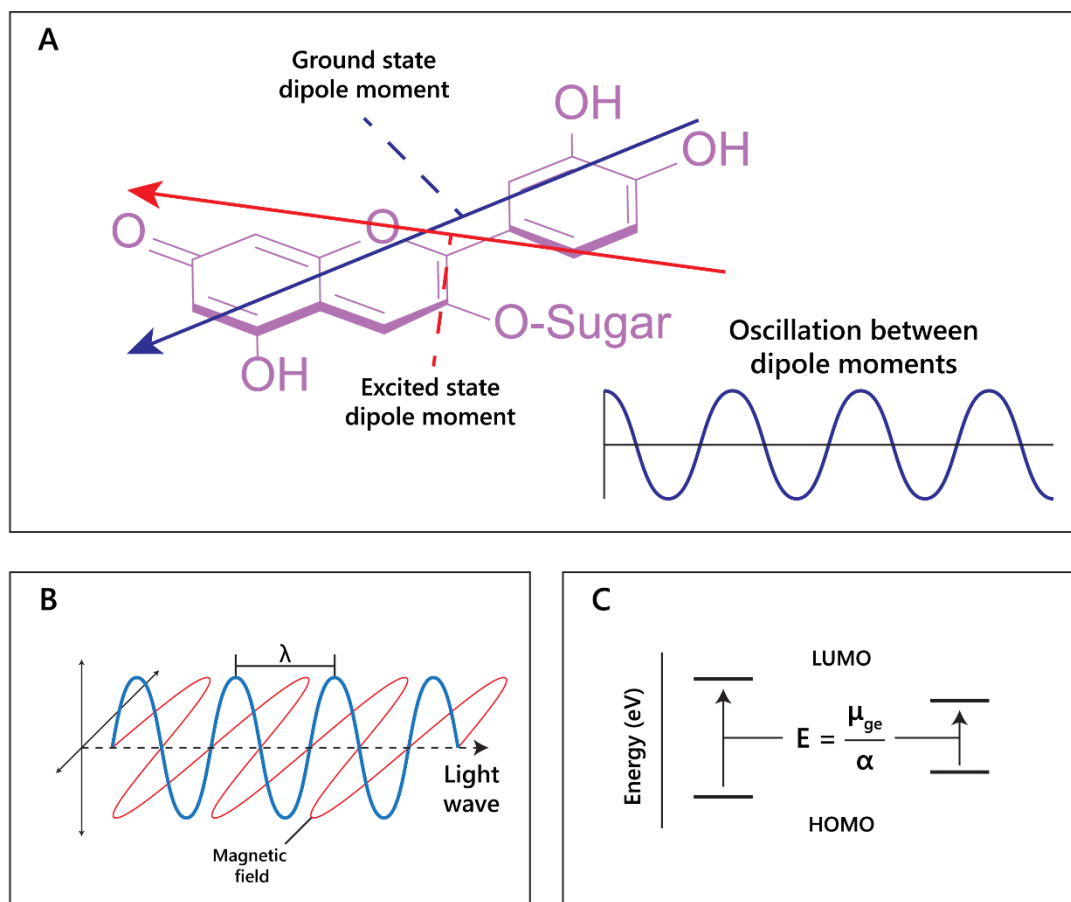
Figure 1.9 Simplified schematic of light absorption by cyanidin 3-O-glucoside. **(A)** When cyanidin absorbs electromagnetic radiation, the energy of the radiation can cause an electron to transition from a lower energy state (HOMO) to a higher energy state (LUMO). This transition can be accompanied by a change in distribution of electron density, and generation of internal magnetic forces due to moving electric charges. These transitions are the result of the magnetic component of electromagnetic waves **(B)**, with the frequency of oscillation between dipole moments corresponding to the frequency of the electromagnetic radiation. **(C)** The energy (E) of the interaction between electrons and electromagnetic radiation is dependent on the strength of the oscillating electric dipole (μ_{ge}) and the polarisation of the molecule — α; in highly polarised systems such as anthocyanins, the transition energy is reduced, allowing absorption of lower energy (longer wavelength) electromagnetic radiation.

1.7.3 Types of Copigment

Owing to their large and rigid, polarisable surface, anthocyanins can theoretically form vertical stacking assemblies with over ten-thousand compounds (Trouillas *et al.*, 2016). However, the major natural copigment compounds identified are the hydrolysable tannins, flavonoids (such as flavones, flavonols, and other anthocyanins), and phenolic acids (**Figure 1.10**).

The basic requirement for a copigment is the presence of a planar arene moiety, which favours interaction with the anthocyanin chromophore. Increasing the size of the aromatic system or the substituent groups can have profound effects on the degree of copigmentation (Escribano-Bailón, Rivas-Gonzalo and García-Estévez, 2018). Planar flavonols such as kaempferol, quercetin, and myricetin are conjugated over their A-, B-, and C-rings and thus have the highest affinity for copigmentation.

Hydroxycinnamic acid (HCA) derivatives such as *p*-coumaric, caffeic, ferulic, and sinapic acids are commonly observed as both inter- and intramolecular copigments, producing a similar degree of colour modulation to planar flavonoids (Dangles and Fenger, 2018). Increasing substitution of the arene group with hydroxyl and methoxy substituents increases the thermodynamic binding constant, and the affinity for copigmentation increases along the series: p-coumarate < caffeic acid < ferulic acid < sinapic acid (Marković, Petranović and Baranac, 2005; Galland *et al.*, 2007).

Despite the advances in understanding copigmentation interactions, it is still unclear whether this increased affinity for highly substituted groups is attributable to distortion in pigment electron density, or the formation of additional direct electrostatic interactions between the substituent groups and the anthocyanin chromophore, although it is highly likely that both mechanisms play a role.

Hydroxycinnamic acids

p-Coumarate

Caffeic acid

Ferulic acid

Sinapic acid

Benzoic acid derivatives

Vanillic acid

Gallic acid

Flavonols

Flavones

Flavan-3-ols
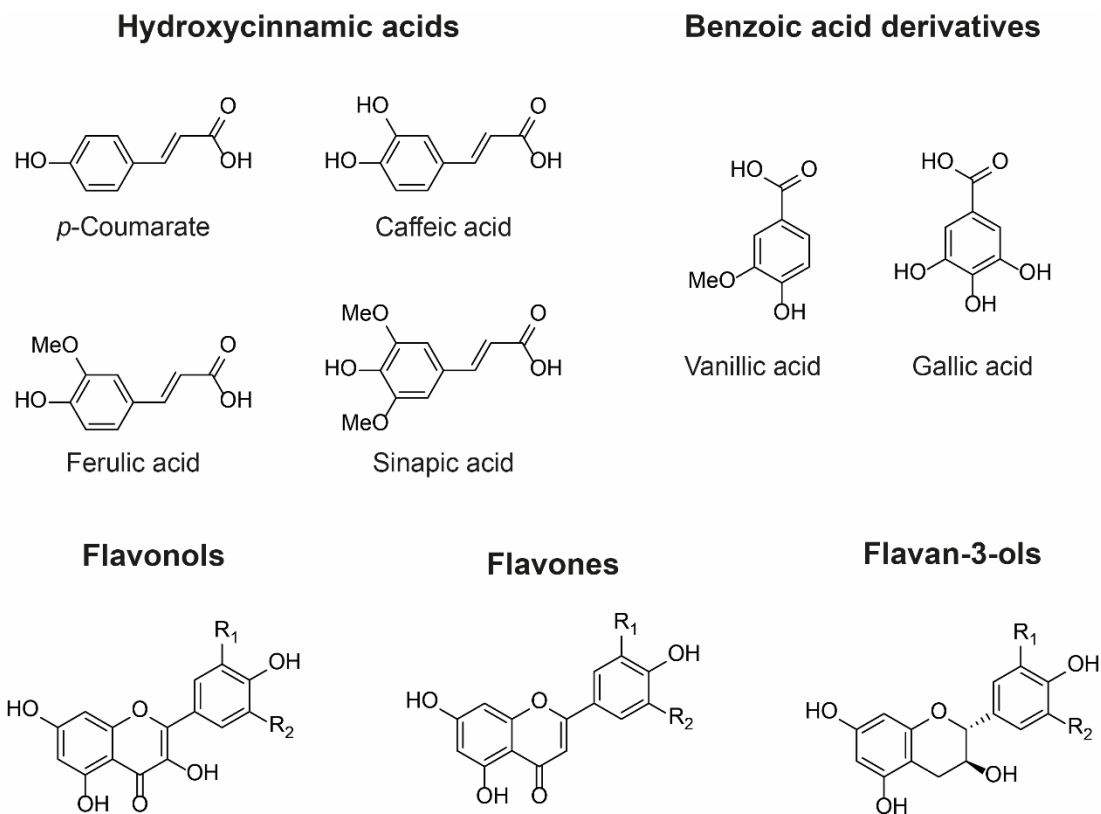
Figure 1.10 Examples of chemical structures of commonly observed copigments (Houghton, Appelhagen and Martin, 2021). In nature, copigmentation is dependent on co-localisation of anthocyanins and copigments within the vacuole. Consequently, flavone and flavonol copigments are glycosylated in order to facilitate their transport to the vacuole. Hydroxycinnamoyl quinic acids such as 3-O-caffeoyl and 3-O-*p*-coumaroyl quinic acids can be transported to the vacuole as aglycones and are found as copigments in hydrangea flowers, for example (Ito *et al.*, 2019).

28

1.7.4 Blues from Intermolecular Copigmentation

Blue colouration by intermolecular copigmentation alone is not common, and when it occurs, it is usually the result of copigmentation of anthocyanins with glycosylated flavonols and flavones. The blue-purple-coloured flowers of cranesbill, Geranium "Johnson's Blue", *Geranium pratense*, and *Geranium sanguineum* all contain the same anthocyanins: malvidin 3-O-glucoside-5-O-(6"-O-acetylglucoside) and a minor fraction of malvidin 3-O-(6"-O-acetyl glucoside) (Markham, Mitchell and Boase, 1997).

The variation in colour between these varieties (Figure 7) results from differences in the molar concentrations of flavonol copigments kaempferol and myricetin 3-O-glucoside and 3-O-sophoroside, with the blue-coloured petals of Geranium 'Johnson's Blue' and *G. pratense* containing significantly higher molar ratios of flavonols than *G. sanguineum* (Markham, Mitchell and Boase, 1997; Yoshida, Mori and Kondo, 2009) (**Figure 1.11**).

Pigment-copigment complexes are weak interactions and are attenuated by heating, dilution, and changes in pH. The formation of these complexes requires high molar ratios of copigment relative to the concentration of anthocyanins, the ratio depending on the affinity of the copigment for the anthocyanin. Observations of copigmentation with flavonols and flavones demonstrate the highest stabilising effect on anthocyanin/copigment complexes at ratios of 1 : 2 − 1 : 20, whereas hydroxycinnamic acids require concentrations between 1 : 10 and 1 : 100 for comparable hyperchromic and bathochromic shifts (Eiro and Heinonen, 2002; Lambert *et al.*, 2011).

Figure 1.11 The different colours of varieties of cranesbill are determined by intermolecular copigmentation between malvidin 3-O-glucoside-5-O-(6-O-acetylglucoside) and varying relative amounts of flavonol copigments kaempferol and myricetin 3-O-glucoside and 3-O-sophoroside [30]. From (Houghton, Appelhagen and Martin, 2021)

Using genetic engineering, Noda *et al*. (Noda *et al.*, 2017) were able to generate blue through the expression of the gene encoding UDP—glucose : anthocyanin 3',5'-O-glucosyltransferase from butterfly pea (*Clitoria ternatea*) and the flavonoid 3',5'-hydroxylase from Canterbury bells (*Campanula medium*) in chrysanthemum, which does not naturally have the capacity to make blue flowers (**Figure 1.12**), in part because it lacks a gene encoding flavonoid 3',5'-hydroxylase and consequently the ability to make delphinidin-based anthocyanins. However, expression of flavonoid 3',5'-hydroxylase alone in chrysanthemum produced only a purple/violet colour (Brugliera *et al.*, 2013; Noda *et al.*, 2013).

Expression of the gene encoding UDP—glucose : anthocyanin 3',5'-O-glucosyl transferase resulted in the modification of delphinidin 3-O-(6"-O-malonyl) glycoside in order to produce delphinidin 3-O-(6"-O-malonyl) glycoside-3'5'-di-O-glucoside (ternatin C5). In isolation, the glycosylation of 3' and 5' hydroxyl groups on the B-ring of the anthocyanin caused a reddening of the anthocyanin with the λ_{max} in the visible range dropping from 527 nm for delphinidin 3-O-(6"-O-malonyl) glycoside to 511 nm for delphinidin 3-O-(6"-O-malonyl) glycoside-3'5'-di-O-glucoside at pH 2.7.

In fact the bluing of the chrysanthemum flowers was due to an increased capacity for copigmentation with the flavones luteolin 7-O-(6"-O-malonyl) glucoside and tricetin 7-O-(6"-O-malonyl)glucoside, which are present naturally in chrysanthemum flowers, which turned from violet in the presence of delphinidin 3-O-(6"-O-malonyl) glycoside, to blue in the presence of delphinidin 3-O-(6"-O-malonyl) glycoside-3'5'-di-O-glucoside.
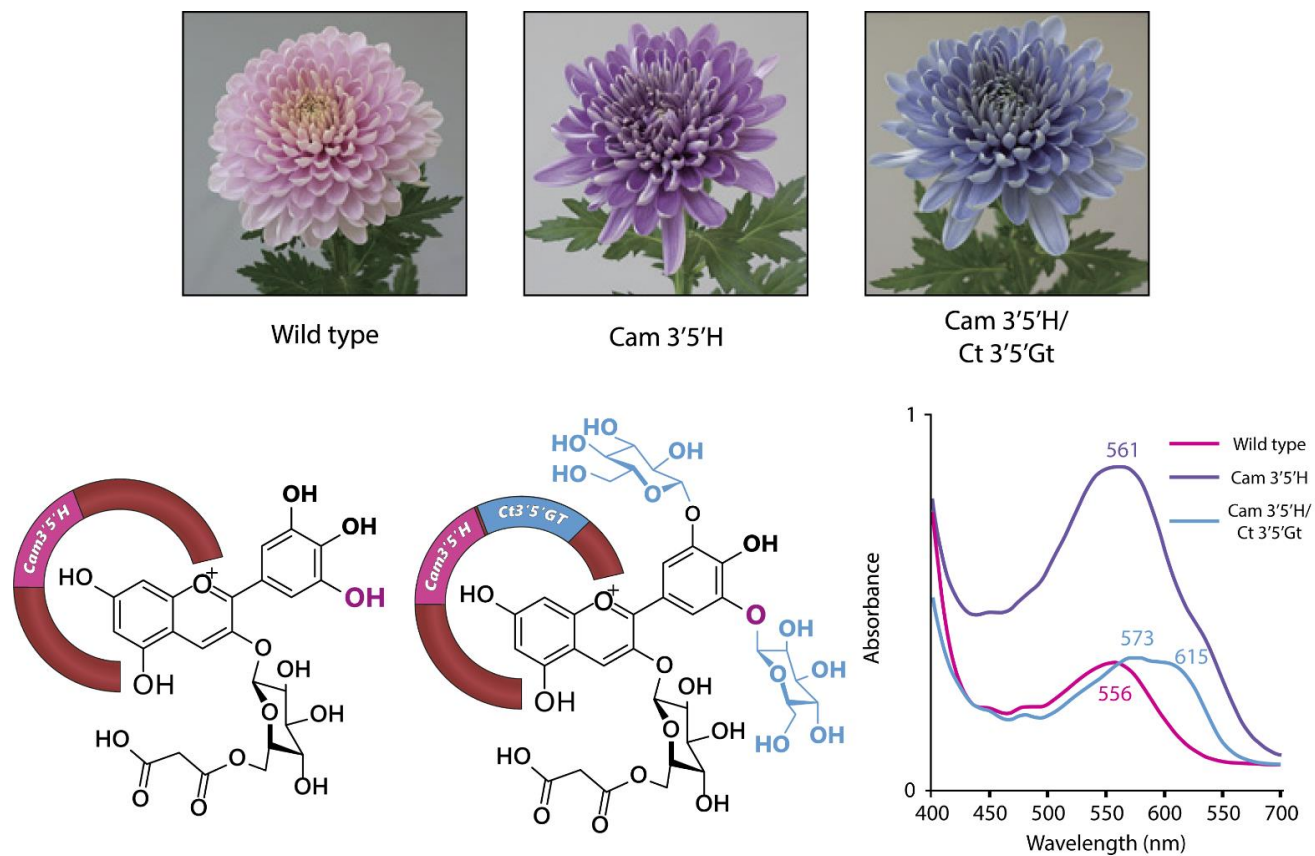
Figure 1.12 Noda *et al*. (Noda *et al.*, 2013) achieved blue colours in chrysanthemum through the expression of Cm3'5'H (flavonoid 3',5'-hydroxylase) and Ct3'5'GT (UDP—glucose : anthocyanin 3',5'-O-glucosyltransferase). Hydroxylation of cyanidin 3-O-(6"-O- malonyl) glucoside at 5' was not sufficient to generate blue colours, requiring the addition of two glucose groups to the B-ring, demonstrated by the shift in visible λ_{max} shown on the right (Houghton, Appelhagen and Martin, 2021).

The bathochromic shift of λ_{max} in the visible range was greater for the shift from cyanidin to delphinidin-based anthocyanins than for the di-glycosylation of the B-ring. However, delphinidin 3-O-(6"-O-malonyl)glycoside-3'5'-di-O-glucoside in combination with the flavonol copigments showed a strong shoulder of absorption at 590 nm, giving the bluing effect of copigmentation (Figure 8). For the delphinidin-based anthocyanins, copigmentation with flavonols increased the relative abundance of quinonoid base in the multistate equilibrium, whereas the abundance of anionic quinonoid base increased in the equilibrium of delphinidin 3-O-(6"-O-malonyl)glycoside-3'5'-di-O-glucoside with its flavonol copigments (Noda *et al.*, 2017).

In nature, intermolecular copigmentation is dependent upon the co-localisation of the copigment with the anthocyanins in the vacuole of petal cells. For this reason, flavonol and flavone copigments in flowers are always glucosylated since glycosylation provides the signal for their transport to the vacuole. For other copigments such as caffeoyl and *p*-coumaroyl quinic acids in blue hydrangeas, accumulation in vacuoles does not involve glycosylation signals (Ito *et al.*, 2019).

Most blue flowers produce anthocyanins that are stabilised by additional copigmentation interactions, which fall into two categories, namely, intramolecular association and metal cation chelation.

1.7.5 Intramolecular Association

Anthocyanins show a wide array of chemical diversity; of the nearly 700 species reported, close to 300 are acylated with one or more aromatic or aliphatic acyl groups, while approximately 100 of these contain both (Andersen and Jordheim, 2006). Although the acylated anthocyanins which produce the blue colours of flower petals are structurally diverse, the predominant anthocyanins are usually delphinidin derivatives.

Anthocyanins which are acylated with aromatic hydroxycinnamic (HCA) derivatives form non-covalent intramolecular associations with their covalently linked HCA moieties, which fold over the chromophore, protecting the chromophore from hydration (Malien-Aubert, Dangles and Amiot, 2001) (**Figure 1.13**).
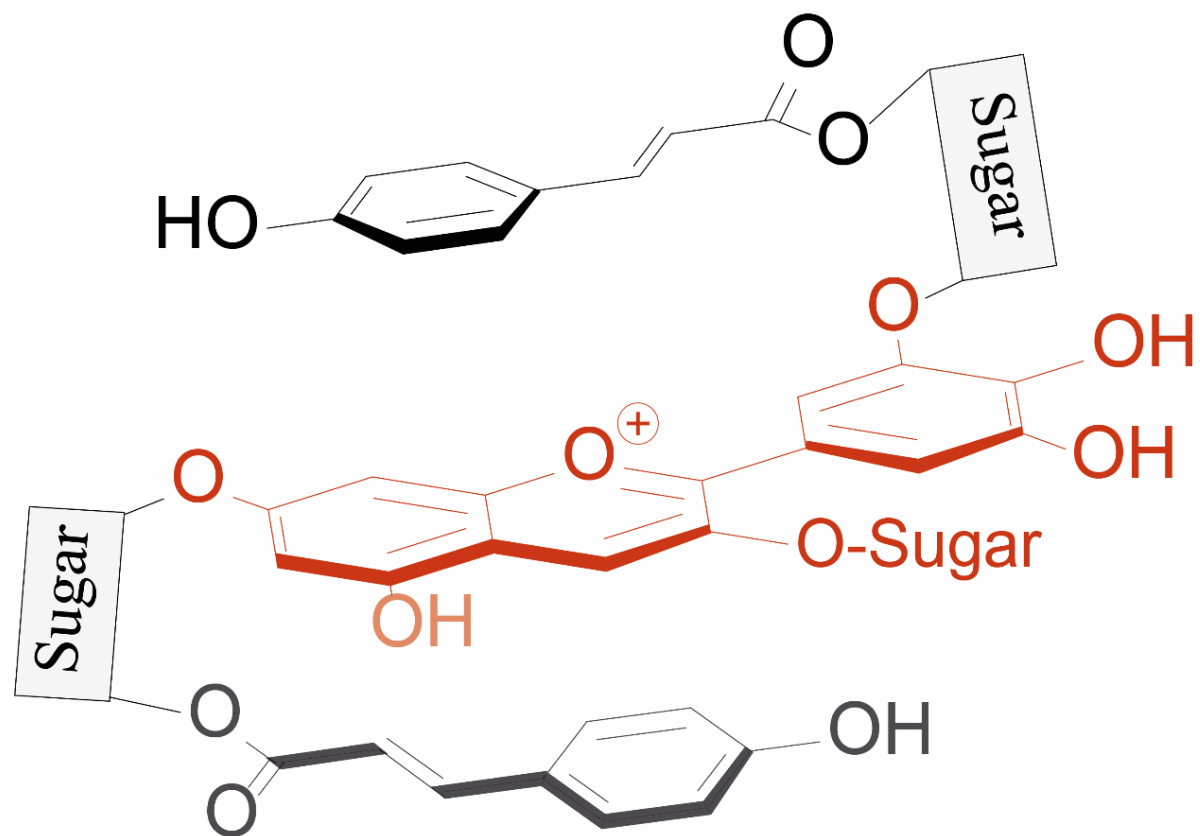
Figure 1.13 Typical structure of poly-acylated anthocyanins allowing non-covalent interaction between the anthocyanin chromophore and covalently linked acyl moieties (intramolecular copigmentation) (Houghton, Appelhagen and Martin, 2021).

These interactions can be loosely categorised into two types: type 1 where the anthocyanin forms intramolecular copigmentation interactions with HCAs both above and below the chromophore, and type 2 where the anthocyanin forms intramolecular interactions on only one plane, leaving one plane exposed for the formation of self-association interactions with other anthocyanins, or copigmentation with a non-anthocyanin copigment (Trouillas *et al.*, 2016).

As these copigments are covalently linked (intramolecular) their effects on colour and stability are not dependent on concentration. As an example, the polyacylated anthocyanin pelargonidin 3-O-(2"(caffeoylglucosyl)-6"-(caffeoylglucosylcaffeoyl) glucoside)-5-O-glucoside produces a bathochromic shift compared to the non-acylated anthocyanin equivalent to the addition of external caffeic acid copigment at a concentration of 0.6 M (Dangles, Saito and Brouillard, 1993).

These types of interactions are dependent on specific structural features of the anthocyanin which, as a consequence of the diversity of aromatically acylated anthocyanins and potential copigments, are incompletely defined. In general, for HCA moieties linked to either the B- or C-rings of the anthocyanidin by monosaccharides or disaccharides, intramolecular copigmentation (as observed in ternatins, heavenly blue anthocyanin (HBA) from Japanese morning glory, and gentiodelphin) is more common than self-association. Anthocyanins with extended acyl residue linkers or repeating sugar–HCA sequences linked to sugars attached to the A- or C-rings favour self-association (as observed with tecophilin) (Trouillas *et al.*, 2016).

The position of linkage of glycosyl–HCAs to the anthocyanin core imposes some geometric constraints, influencing how well the copigment can align with the anthocyanin chromophore, and hence the strength of the association (as measured by the degree of bathochromatic shift). The investigation of mono-acylated anthocyanins has led to the proposal of the following ranking: 3'~5' > 7 > 3 .
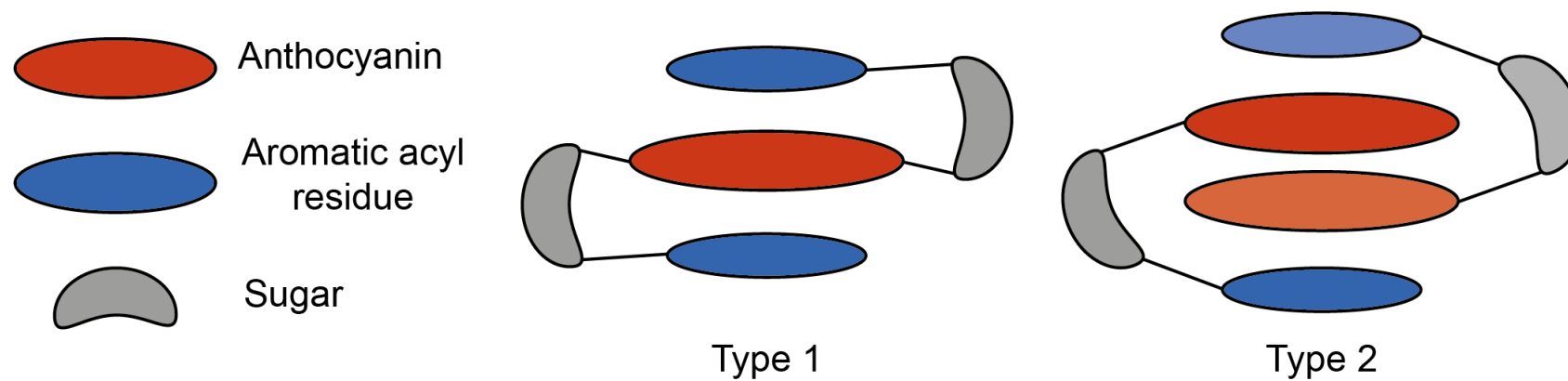
Figure 1.14 Simplified schematic of the two major types of intramolecular copigmentation. Type 1 complexes form intramolecular copigmentation interactions above and below the anthocyanin chromophore. Type 2 complexes form only one intramolecular interaction, and self-association interactions on the reverse plane.

Similarly, the position of acylation on the sugar can influence stacking interactions. Acyl groups are usually bound to the C6"-OH of the primary sugar, allowing enough bond flexibility for efficient stacking (Moloney *et al.*, 2018). Acyl groups have been observed less frequently at C2"-OH, C3"-OH and C4"-OH groups of the primary sugar than on the C6" OH group (Zhao *et al.*, 2017). In contrast to acyl groups linked to primary sugar OH groups, acylation of the secondary sugar on groups such as at C4"-OH appear to not interact with the chromophore at all (Nerdal and Andersen, 1992).

Studies of di-acylated anthocyanins from red cabbage have shown that addition of a single sinapoyl group to the C6" position of first sugar (G1) on C3 of cyanidin 3-O-sophorose-5-O-glucoside causes a bathochromic shift of 10 nm in λ_{max} in the visible range, while addition of a second sinapoyl group on C2" of G2 gives a bathochromic shift of 24 nm in λ_{max}. The equivalent bathochromic shifts that could be achieved with addition of free sinapic acid to cyanidin 3-O-sophorose-5-O-glucoside and cyanidin 3-O-(6"-sinapoyl) sophorose-5-O-glucoside were 4 and 11 nm, respectively (Moloney *et al.*, 2018), demonstrating the effectiveness of intramolecular copigmentation compared to intermolecular copigmentation.

Acylated anthocyanin complexes have increased colour stability and generally increased aromatic acylation increases blueness and anthocyanin stability. The addition of water to C2 of the flavylium ion is key to fading of anthocyanins because it causes the formation of colourless hemiketal epimers. By promoting the formation of quinonoid bases in the multistate equilibrium, HCA decoration increases anthocyanin stability because quinonoid bases cannot undergo hydration reactions. Aromatic acylation also increases the strength of π stacking interactions and slows irreversible colour loss by chemical degradation involving cleavage of the C-ring, although these latter processes are not well understood (Moloney *et al.*, 2018).

Consequently, aromatic acylation causes big increases in the stability of anthocyanin pigments, as illustrated in **Figure 1.15** for extracts of di-acylated anthocyanins from the desert bluebell (*Phacelia campanularia*) and quadri-acylated anthocyanins from butterfly pea (*Clitoria ternatea*) compared to the non-acylated cyanidin 3-O-rutinoside.
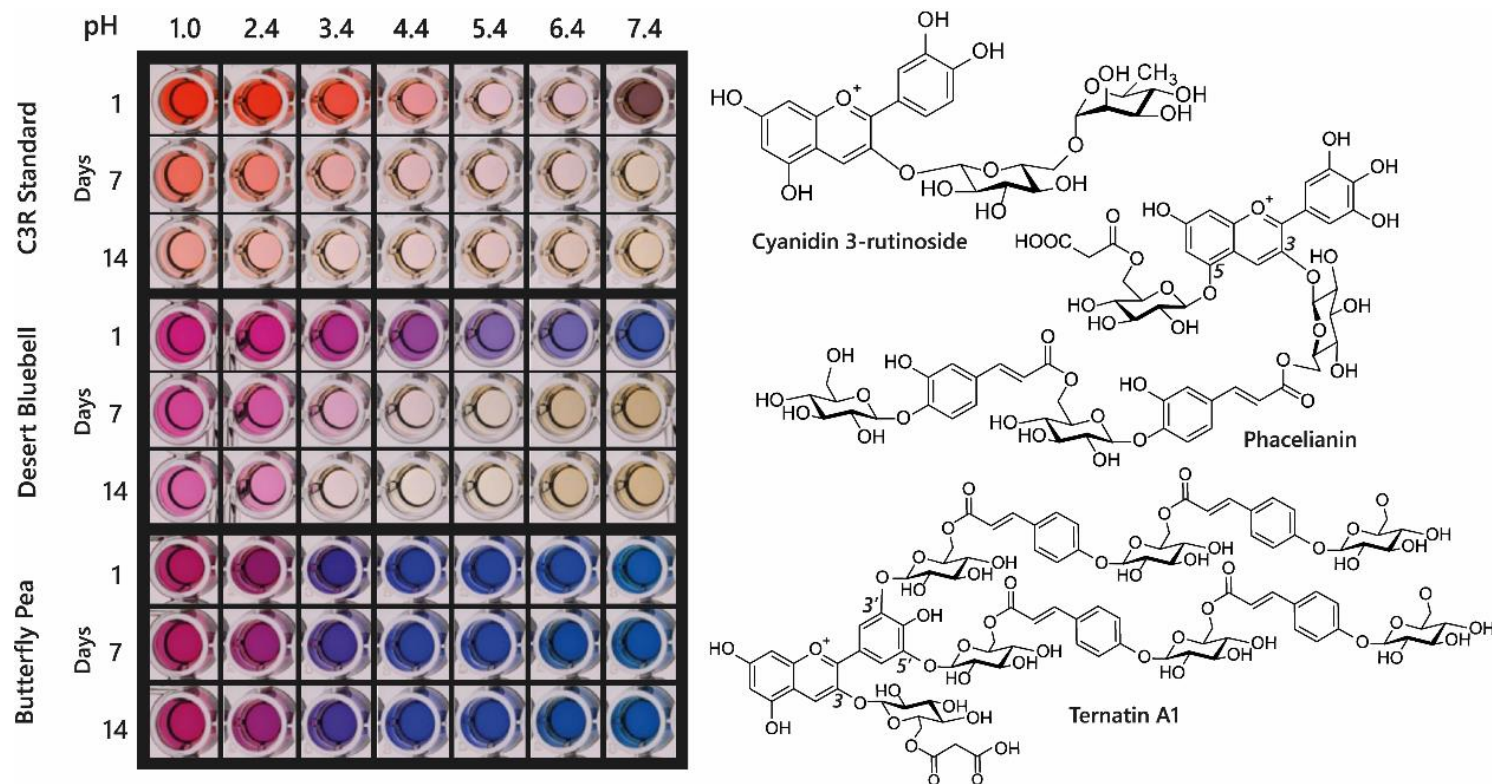
Figure 1.15 Demonstration of increasing blueness and increasing stability associated with multiple acylation of anthocyanins. Cyanidin 3-O-rutinoside was purchased commercially, and anthocyanins were extracted from desert bluebell (*Phacelia campanularia*) and butterfly pea (*Clitoria ternatea*). The most highly acylated anthocyanins in each extract are illustrated to the right. The plates show the colours of the anthocyanin extracts at different pH values and following storage at room temperature for 1, 7, and 14 days (Houghton, Appelhagen and Martin, 2021).

1.7.6 Aggregation of Aromatically Acylated Anthocyanins

Depending on the position of HCA attachment and the presence of additional stabilising glycosylation, some acylated anthocyanins aggregate, even at low concentration, at pH values greater than 2 and less than 8. Studies of aggregation undertaken at elevated temperatures have suggested that aggregation is the result of intermolecular rather than intramolecular associations (Mendoza *et al.*, 2020). Aggregation of this type can occur naturally in plant cells giving rise to the formation of anthocyanic vacuolar inclusions (AVIs) (Markham *et al.*, 2000; Kallam *et al.*, 2017). The production of aggregated acylated anthocyanins in AVIs is used in some flowers to create dark spots of colour, as seen in the central region of the sepals of lisianthus (*Eustoma grandiflora*). In the case of lisianthus, it is the absence of additional glycosylation of the C3 of delphinidin and cyanidin 3-O-galactoside-5-O-(*p*-coumaroyl) glucoside and 3-O-galactoside-5-O-(feruloyl)glucoside on the inner epidermis of the base of the sepals that is associated with AVI formation compared to the delphinidin and cyanidin-3-O-(rhamnosyl) galactoside-5-O-(*p*-coumaroyl) glucoside and 3-O-(rhamnosyl) galactoside-5-O-(feruloyl) glucoside.

These observations have been interpreted in terms of reduced glycosylation favouring AVI formation, which fits well with in vitro observations that 5-O-glycosylation reduces AVI formation in vivo (Kallam *et al.*, 2017). The role of AVI formation in patterning the flowers of lisianthus is illustrated in **Figure 1.16**.
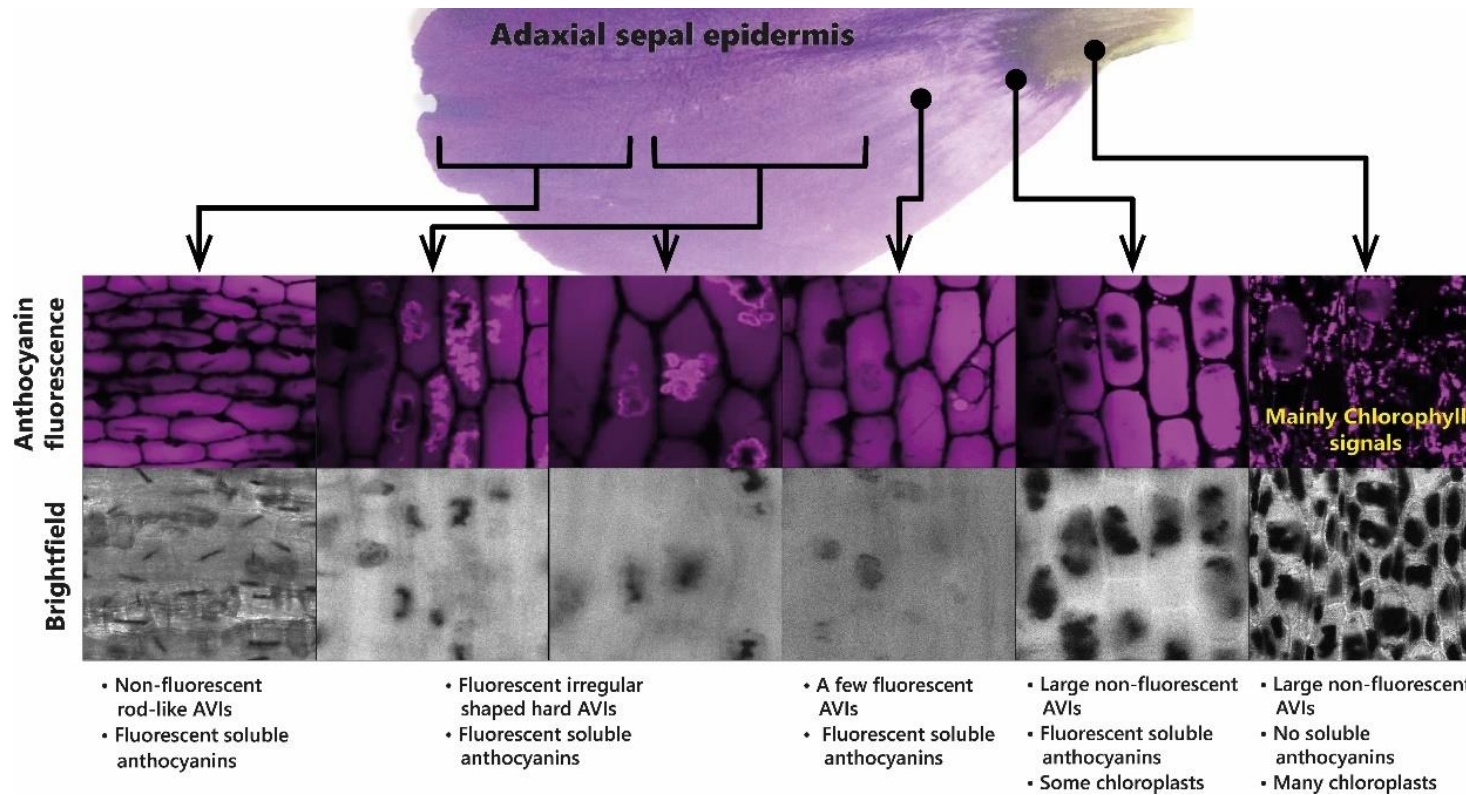
Figure 1.16 Presence of anthocyanin aggregates in the sepals of lisianthus is associated with the production of aggregates of acylated anthocyanins. The anthocyanic vacuolar inclusions (AVIs) that give the dark colour to the base of the sepals do not fluoresce under UV light. The soluble anthocyanins in lisianthus sepals fluoresce as do some additional anthocyanin aggregates from the outermost regions of the sepals (Houghton, Appelhagen and Martin, 2021).

Anthocyanins from blue flowers commonly also contain aliphatic acyl groups, such as malonic acid. Addition of malonyl groups does not impact the colour of anthocyanins directly (Luo *et al.*, 2007). The contribution of aliphatic acyl groups to stacking interactions (and hence indirectly to colouration) is unclear, although a few functions have been proposed. Acylation with a malonyl group was observed to lower the pKa of anthocyanins in *Matthiola incana*, possibly by stabilising the neutral species through the formation of hydrogen bonds between the malonyl group and the oxygen at C7 of the quinonoid base (Trouillas *et al.*, 2016).

Acylation with multiple aromatic groups reduces the solubility of anthocyanins in the aqueous cytoplasm. This may pose a problem during the biosynthesis of poly-acylated anthocyanins, such as the ternatins; the largest of which is acylated with four *p*-coumaroyl moieties. Therefore, acylation with aliphatic organic acids on glucose moieties at C3, C5, or C7 may help keep these anthocyanins soluble in the aqueous cytoplasm during biosynthesis (Luo *et al.*, 2007; Kallam *et al.*, 2017).

1.8 Anthocyanin metal complexes

Anthocyanins such as cyanidin, delphinidin, and petunidin, with two free hydroxyl groups on the B-ring, demonstrate a colour change upon chelating multivalent metal cations.

The metal cation competes with the hydrogen ions and the red flavylium cations are transformed to form the blue quinonoid base anions (Sigurdson *et al.*, 2016). The complexes are then stabilised through typical copigmentation interactions with external copigments, or through self-association with other anthocyanins, where both the copigment and anthocyanin are coordinated by the metal cation.

1.8.1 Fuzzy-metal complexes

These complexes still require high relative molar concentrations of copigment, and are blue only in aqueous solution, where their colour is similarly attenuated by dilution.

Further, their colour is lost upon attempts to crystallise or isolate these pigments. For this reason, they are referred to as "fuzzy metal complexes", due to the difficulties in clarifying the mechanisms of association (Yoshida, Mori and Kondo, 2009).

The best studied cases of fuzzy metal complexes are those present in *Hydrangea macrophylla*. These flowers range in colour from red to mauve to blue, and all contain delphinidin 3-O-glucoside, copigments (mainly 3-O-caffeoyl and 3-O-*p*-coumaroyl quinic acids), and Al^{3+} in differing concentrations (**Figure 1.17**).

Neither Al^{3+} or 3-O-caffeoylquinic acid alone could produce blue colour when mixed with delphinidin 3-O-glucoside, the anthocyanin present in both blue and red hydrangeas, in the molar ratios found in the flower sepals. The blue variety, however, had significantly higher concentrations of copigment and Al^{3+} compared to the red variety (Takeda, Kariuda and Itoi, 1985).

In a recent publication, it was shown that the anthocyanin accumulates in the sub-epidermal layer of sepals of both red and blue varieties. The concentration of caffeoyl/p-coumaroyl 3-O-quinic acids, which serve as copigments, are high in several cell layers of the sepals of hydrangea. However, it is the high levels of aluminium in the subepidermal layer that confer the specificity of blue, achieved through the association of anthocyanins, copigments, and Al^{3+} ions in the subepidermal layer of blue flowers (**Figure 1.17**). In the comparator red-flowered variety, levels of Al^{3+} were found to be much lower in the subepidermal cell layer of the sepals (Ito *et al.*, 2019).

Historically, this mechanism has been exploited to glean structural information about the anthocyanin under investigation, the addition of Al^{3+} producing a characteristic blueing of the anthocyanin indicating the presence of two free hydroxyl groups present on the B-ring. Gardeners have used hydrangea flower colour to demonstrate the effect of soil pH on flowers; plants grown on acid soils are bluer because low pH results in more bioavailable Al^{3+} to form the fuzzy metal complex. Cultivation of the same varieties on more alkaline soils results in red flowers due to the low availability of Al^{3+} under these conditions (Hayter, 1949; Chalker-Scott, 2018)

	Red Protoplast	Blue Protoplast
Cell Colour (λ_{max}, nm)	536	588
pH	3.1	3.6
Anthocyanin	24.6 mM	10.9 mM
3-O-Caffeoylquinic acid	6.7 mM	4.6 mM
5-O-Caffeoylquinic acid	0.3 mM	2.9 mM
Al^{3+}	0.3 mM	2.4 mM
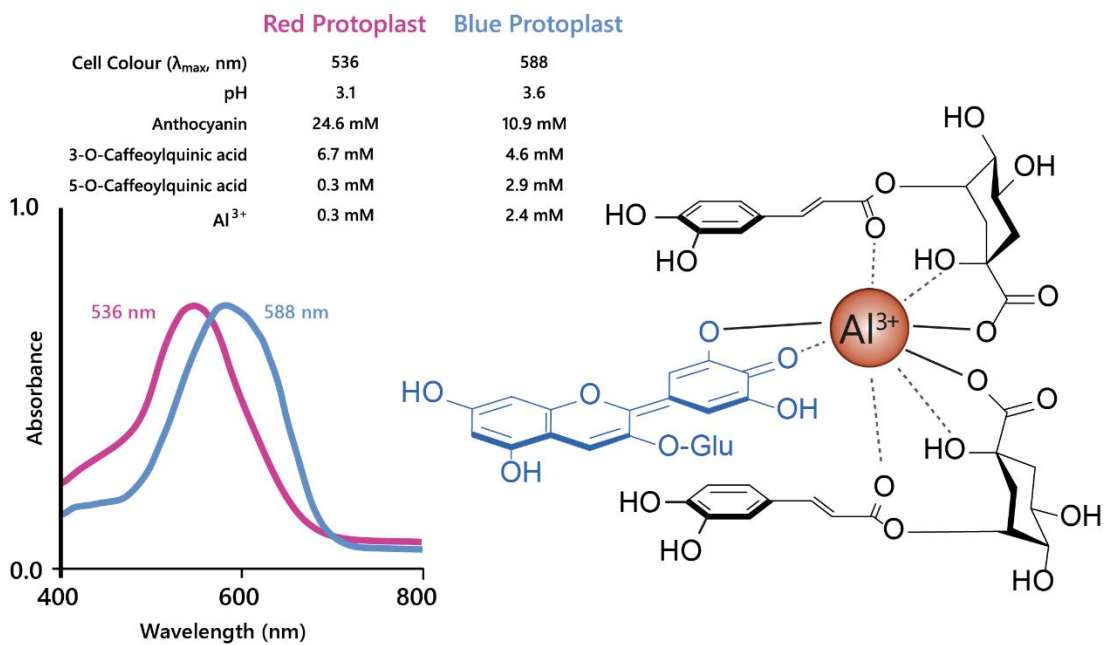
Figure 1.17 Hydrangeas produce blue flowers due to the formation of metal chelates between delphinidin, copigments, and Al^{3+}. This complex stabilises the anionic quinonoid base species, producing a bathochromic shift of 52 nm, as shown with the shift in λ_{max} for the visible absorption spectra shown on the left (Houghton, Appelhagen and Martin, 2021).

1.8.2 Metalloanthocyanins

A specific subset of metal-coordinated anthocyanins form large supramolecular structures termed metalloanthocyanins, which have been demonstrated to perform an important role in the development of blue colour in several floral systems. These complexes are composed of stoichiometric amounts of anthocyanins, flavones, and metal ions. For all metalloanthocyanins, these ratios are fixed at 6:6:2, respectively (Yoshida, Mori and Kondo, 2009).

The formation of these complexes requires similar structural properties to those seen in the more typical metal-coordination with anthocyanins—two unsubstituted neighbouring hydroxyl groups on the B-ring. Both the anthocyanins and flavones self-associate in pairs with distances between the aromatic rings of the respective pair of approximately 3.3 Å, indicative of hydrophobic interaction (Shiono, Matsugaki and Takeda, 2008) (**Figure 1.18**).

To date, only a few metalloanthocyanins have been identified with colours ranging from purple to blue. These known metalloanthocyanins consist of either delphinidin, cyanidin, or petunidin and Mg^{2+}, Al^{3+}, or Fe^{3+} as metal cations centralised within the complex (Yoshida, Mori and Kondo, 2009). Generally, in nature cyanidin-based anthocyanins tend to form complexes with Fe^{3+}, whereas delphinidin-based anthocyanins form complexes with Mg^{2+}, although there are exceptions to this "rule". Of course, in vitro these predominant metals may be capable of being substituted by others (unavailable in nature), which may impact the colour and stability of the complex (Ratanapoompinyo *et al.*, 2017).
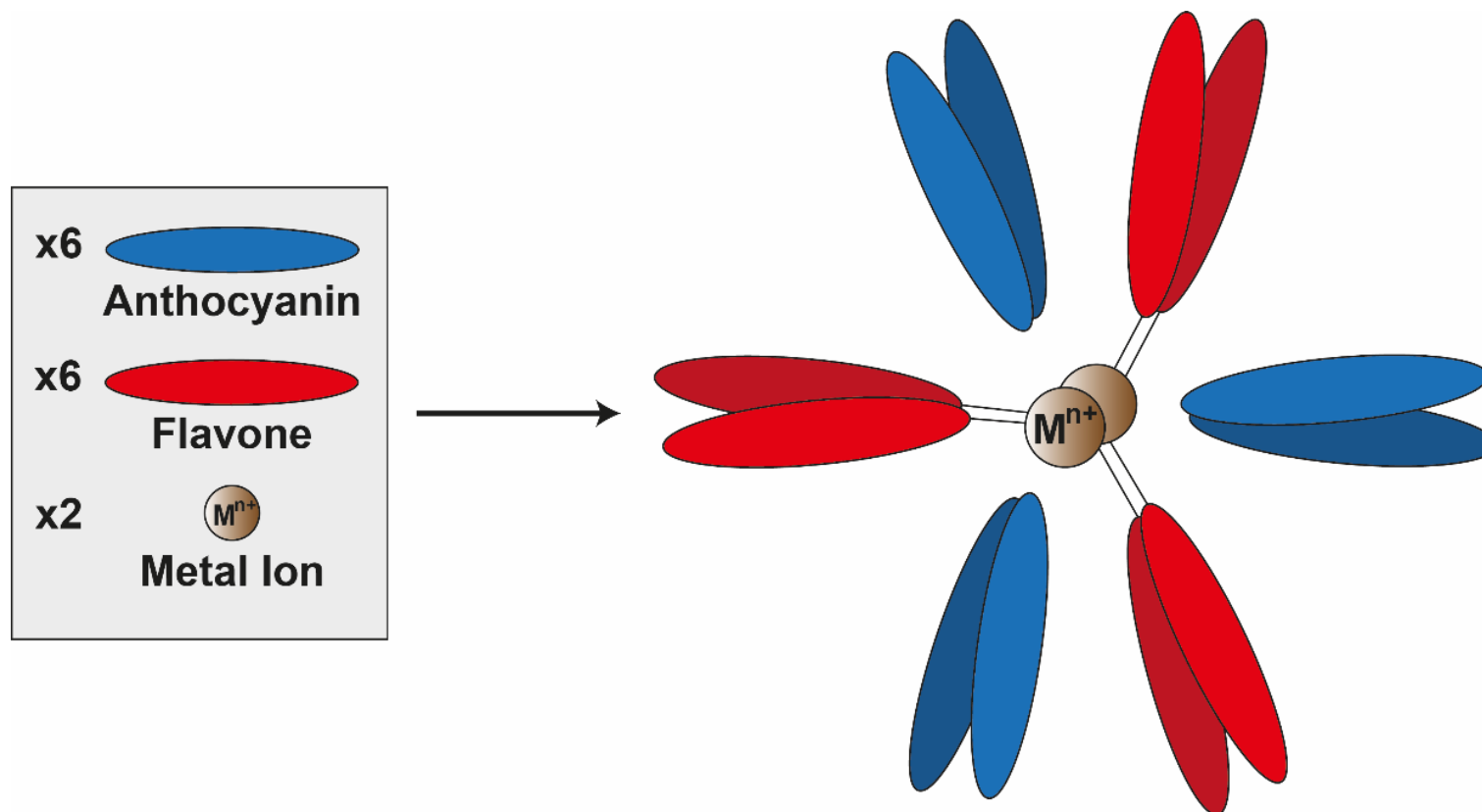
Figure 1.18 Assembly of metalloanthocyanin complex by self-assembly of six anthocyanin molecules, six flavone molecules, and two metal cations (Houghton, Appelhagen and Martin, 2021).

1.9 Forces Involved in Copigmentation

Hunter and Sanders (Hunter and Sanders, 1990) originally proposed that the π electron density of an unsubstituted aromatic ring results in a quadrupole moment, with a partial negative charge above both aromatic faces, and partial positive charge around the periphery. π-interaction is an attractive force between electrical quadrupoles, overpowering the repulsion of π-electron clouds. Where aromatic systems share a similar electron density, σ–π interactions are favourable and should therefore favour an edge-to-face or off-centre parallel arrangement, disfavouring face-centred arrangements (**Figure 1.19**).

The addition of electron-withdrawing substituents to an aromatic (arene) polarises π electron density away from the aromatic centre, reversing the direction of the quadrupole (Li *et al.*, 2018). Pairs of electron-rich and electron-deficient aromatics preferentially pair in a face-centred arrangement, referred to as "aromatic donor–acceptor interaction" (**Figure 1.19**).

Pigment–copigment interactions are often attributed overwhelmingly to π–π interactions. However, these fail to account for other molecular forces which can initiate and drive copigmentation. The magnitude of the interaction is the product of all forces acting between the two aromatic systems, maximising electrostatic forces and dispersive attraction.
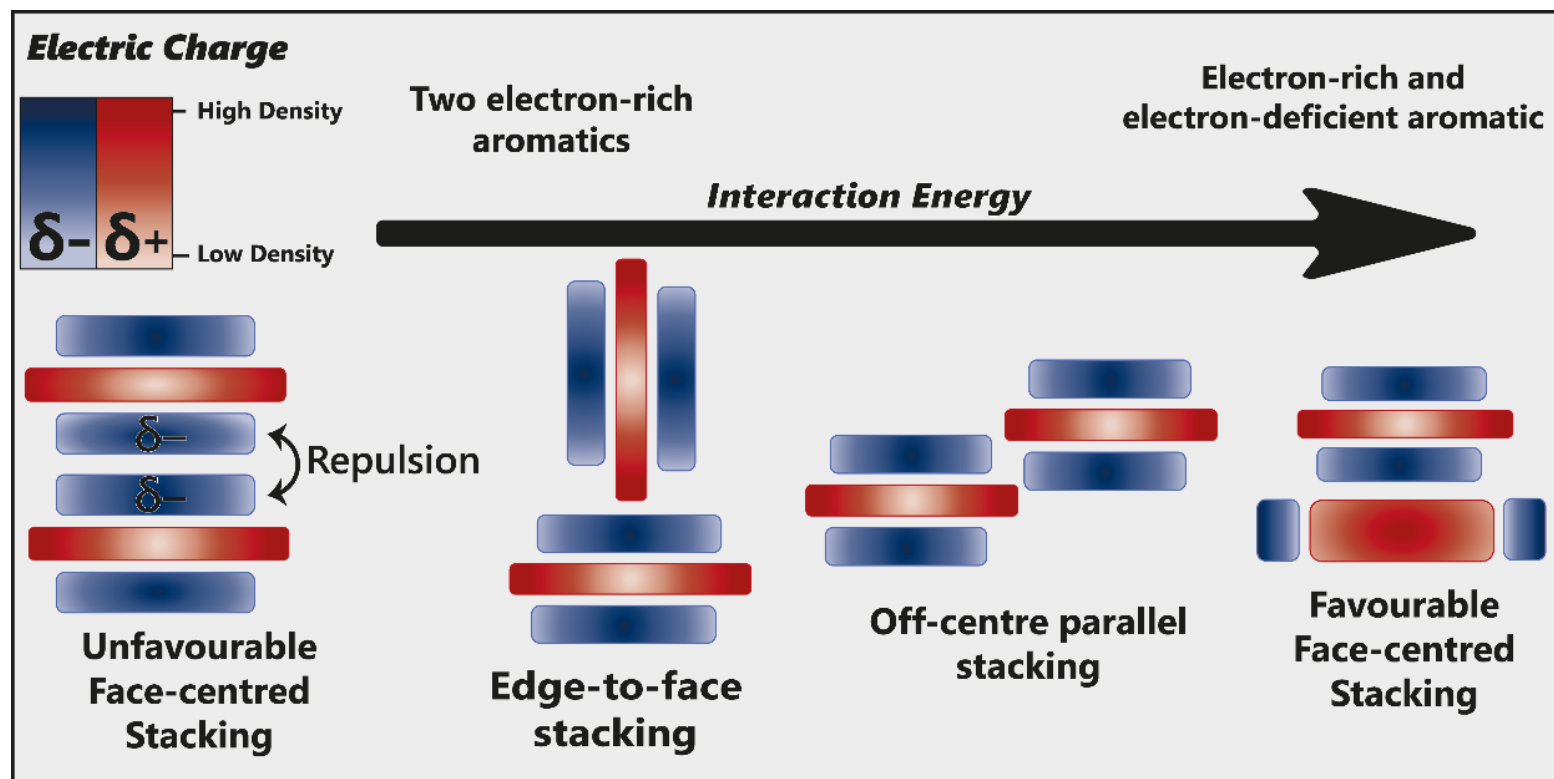
Figure 1.19 Alternative modes of stacking are shown, emphasising the locations of electrostatic attraction or repulsion. As the interaction energy increases, so too does complex stability (Houghton, Appelhagen and Martin, 2021).

π–π interactions are not limited to arene moieties; cyclical aliphatic molecules have also been demonstrated to take part in these interactions, suggesting that the requirement for electron delocalisation (as seen in aromatics) is not strict (Molcanov and Kojić-Prodić, 2019). This is particularly noteworthy within non-aromatic heterocycles, such as glucose, where the presence of an oxygen atom introduces asymmetry within the heterocycle, resulting in polarisation. Within any anthocyanin or copigment, there will be regions of high- and low-electron density. Therefore, the overall strength of intra- and intermolecular copigmentation interactions will depend on the alignment of these regions (Hunter and Sanders, 1990).

Such interactions may also explain why the addition a single glucose moiety at C3' and C5' (to produce ternatin C5) was sufficient to generate blue colouration in butterfly pea and also in genetically modified chrysanthemum (Marković, Petranović and Baranac, 2005; Noda *et al.*, 2017). These glucose groups may increase the dispersive attraction, and possibly direct electrostatic interactions, pulling the copigment closer to the chromophore and reducing the transition energy gap, thus producing a bluer colour (**Figure 1.12**).

1.9.1 The Hydrophobic effect

Anthocyanins form the strongest copigmentation interactions when dissolved in polar solvents and there is a decrease in magnitude of the copigmentation effect in non-aqueous solvents, where copigmentation has been attributed largely to the hydrophobic effect (Brouillard *et al.*, 1991).

These are both classical (entropic) and nonclassical (enthalpic). Water molecules solvating anthocyanins and their copigments form a highly organised hydrogen bond network around the aromatic intramolecular region (hydrophobic cavity) (Trouillas *et al.*, 2016). As the pigment and copigment come together, the hydrogen bonds between water molecules in the solvation shell are broken and released from the hydrophobic cavity, increasing the system entropy and enthalpy (**Figure 1.20**). The barrier to complex formation is further reinforced by the required decrease in the system entropy (requiring energy) as the acyl group rotates about the C6" axis to form a highly ordered structure.

Figure 1.20 Classical and non-classical hydrophobic forces involved in the formation of copigmentation complexes. As the two interacting aromatics approach each other, hydrogen bonds between the highly organised hydration shell are broken, increasing both enthalpy and entropy. New hydrogen bonds form between the displaced water molecules and the bulk water, decreasing enthalpy. A further decrease in enthalpy is observed due to the formation of attractive dispersive and electrostatic interactions. All observed copigmentation interactions are exothermic (−ΔH) and spontaneous (−ΔG0) at room temperature (23 °C) (Houghton, Appelhagen and Martin, 2021).

49

However, all observed copigmentation interactions are spontaneous at room temperature. To account for the energy deficit from the initial bond-breaking, energy is counterbalanced by a decrease in enthalpy by the formation of new hydrogen bonds between the liberated water molecules. In all cases investigated, pigment-copigment complexes have been exothermic (Dangles and Brouillard, 1992).

Generally, intermolecular copigmentation requires water as a solvent, indicating the importance of the hydrophobic effect in generating stable complexes (Alluis. and Dangles, 2001). However, in some intramolecular copigmentation complexes, the hydrophobic effect is not strictly necessary for interactions to occur. There have been many observations of anthocyanins forming pigment-copigment interactions in non-polar solvents, where hydrophobic forces would be minimal (Terahara *et al.*, 1998). Hydrophobic forces may drive the arene moieties close enough for dispersive and electrostatic attraction, which then dictate the final spatial arrangement and hence, the degree of charge transfer resulting in observed colour changes.

1.10 Project aims

Numerous studies have focused on the colour and stability altering effects of anthocyanic decoration. However, due to limited number of decorations typically observed, these studies themselves have been limited to the effects of only a few decorations. The extracts of dried butterfly pea flowers contain 15 delphinidin-3-(6"-malonate),3',5'-O-triglucoside derivatives (ternatins), decorated with alternating glucose and *p*-coumarate residues on the 3' and 5' hydroxyl groups. These can be broadly categorised into four groups: The A series, which feature glucose residues at each of the terminal positions of the 3' and 5' sidechain; the D series, which have *p*-coumarate residues at each of the terminal positions; the B series which have one glucose and one *p*-coumarate at the terminal positions; and the C series, where additional decorations are added to the 3' sidechain only. Therefore, this study focussed on providing insight into the stability-enhancing and colour-modulating effects of iterative decorations with sugar and organic acid residues on anthocyanin colour and stability.

To achieve this goal, high purity isolates of ternatin anthocyanins were generated and their molecular structures confirmed using a combination of high-resolution mass spectroscopy and nuclear magnetic resonance spectroscopy.

The colour and stability of each isolate was characterised using UV-VIS spectroscopy and related observed differences to the decorations present on each anthocyanin.

Finally, the conformational mechanisms responsible for the changes in stability and colour enhancement were demonstrated at atomic scale resolution. The overall aim of this thesis was to analyse comprehensively and to evaluate the suitability of ternatin anthocyanins as replacements for synthetic blue food colourants.

Chapter 2

Isolation and structural elucidation of ternatin anthocyanins from *Clitoria ternatea*

Chapter 2: Isolation and structural elucidation of ternatin anthocyanins from *Clitoria ternatea*

2.1 Introduction

2.1.1 Butterfly pea (*Clitoria ternatea*)

Butterfly pea is an herbaceous perennial legume, native to equatorial Asia, particularly Thailand, Malaysia and the Philippines; although it is now naturalised in all semi-arid tropics of Africa, Australia and Asia. As butterfly pea is a nitrogen fixing legume, the plant has found use as a cover crop in orchards, serving as a feed crop for pasture animals, and has an excellent capacity as a pioneer species for disturbed land (Duncan, 2017). More recently, butterfly pea garnered significant interest for the range of its valuable natural products (Oguis *et al.*, 2019).

One of these notable natural products are the anthocyanins which colour the flowers. Butterfly pea makes intensely blue pigments in its flowers (**Figure 2.1**), involving the synthesis of highly decorated delphinidin-based anthocyanins, which show extraordinary stability in solution when compared to anthocyanins from other blue flowers such as *Gentiana triflora* which produces gentiodelphin (Tasaki *et al.*, 2022). Nineteen different anthocyanins (**Table 2.1**) have been identified in the flower petals of butterfly pea, of which fifteen are classified as ternatins, with variations in their 3', 5' acyl-glycosyl sidechains (Kazuma, Noda and Suzuki, 2003).

Figure 2.1 Flowers of wild-type *Clitoria ternatea*.

Table 2.1 Current list of identified anthocyanins and flavonol glycosides present in the petals of *Clitoria ternatea*

Anthocyanins	
Delphinidin 3-(2"-rhamnosyl-6"-malonyl)-glucoside	Ternatin B4
Delphinidin 3-(6"-malonyl)-glucoside	Ternatin C1
Delphinidin 3-neohesperidoside	Ternatin C2
Delphinidin 3-glucoside	Ternatin C3
Ternatin A1	Ternatin C4
Ternatin A2	Ternatin C5
Ternatin A3	Ternatin D1
Ternatin B1	Ternatin D2
Ternatin B2	Ternatin D3
Ternatin B3	

Flavanol glycosides	
Kaempferol 3-(2G-rhamnosyl-rutinoside)	Quercetin 3-rutinoside
Kaempferol 3-neohesperidoside	Kaempferol 3-glucoside
Kaempferol 3-(2"-rhamnosyl-6"-malonyl)-glucoside	Myricetin 3-(2G-rhamnosyl-rutinoside)
Kaempferol 3-rutinoside	Myricetin 3-neohesperidoside
Kaempferol 3-glucoside	Myricetin 3-(2"-rhamnosyl-6"-malonyl)-glucoside
Quercetin 3-(2G-rhamnosyl-rutinoside)	Myricetin 3-rutinoside
Quercetin 3-neohesperidoside	Myricetin 3-glucoside
Quercetin 3-(2"-rhamnosyl-6"-malonyl)-glucoside	

2.1.2 Ternatins

Ternatins are a series of poly-acylated derivatives of delphinidin-3(6"-malonyl),3',5'–O-triglucosides present in the petals of butterfly pea. These anthocyanins feature repeating glycosylation and acylation with glucose and *p*-coumarate residues on both the 3' and 5' position hydroxyl groups. Each of the ternatins can be considered an intermediate in the synthesis of ternatin A1, the most decorated anthocyanin present. There are a large number of ternatin precursors (**Figure 2.2**), from the smallest, ternatin C5 which is the malonated delphinidin 3, 3', 5'-triglucoside, to the largest, ternatin A1, which has two 5-unit side chains (Glu-Cou-Glu-Cou-Glu) and a C3 position 6"-O-malonyl-glucopyranose (Terahara *et al.*, 1996, 1998).

Ternatins have been classified into four groups based on their decorations in the terminal position within the 3' and 5' sidechains. Group A terminate with two glucose molecules (G, G), group B with glucose and *p*-coumarate and series D with *p*-coumarate moieties on the 3' and 5' positions. Group C features the extension of one of the chains asymmetrically. Members of each series are numbered based on their elution order within their respective group. Variations in retention time are a result of differences in molecular weight within a series, and differences in hydrophobicity and sidechain symmetry between series. Series D (with two terminal acyl groups) are the most hydrophobic, with D1 eluting latest. Series B (with 1 terminal acyl group) is less hydrophobic than series D, and series A (with 0 terminal acyl groups) elutes earliest. Sidechain asymmetry determines elution order when terminal decorations and molecular masses are equal (as seen with B2 and B3). B2 has a more symmetrical sidechain compared to B3, and hence has better retention by the stationary phase of the reverse-phase column and elutes more slowly. Ternatin A1 (**Figure 2.3**) is not only the largest monomeric anthocyanin but has also been shown to be one of the most stable anthocyanins in neutral aqueous solutions, likely due to the high degree of acylation (Terahara *et al.*, 1990).
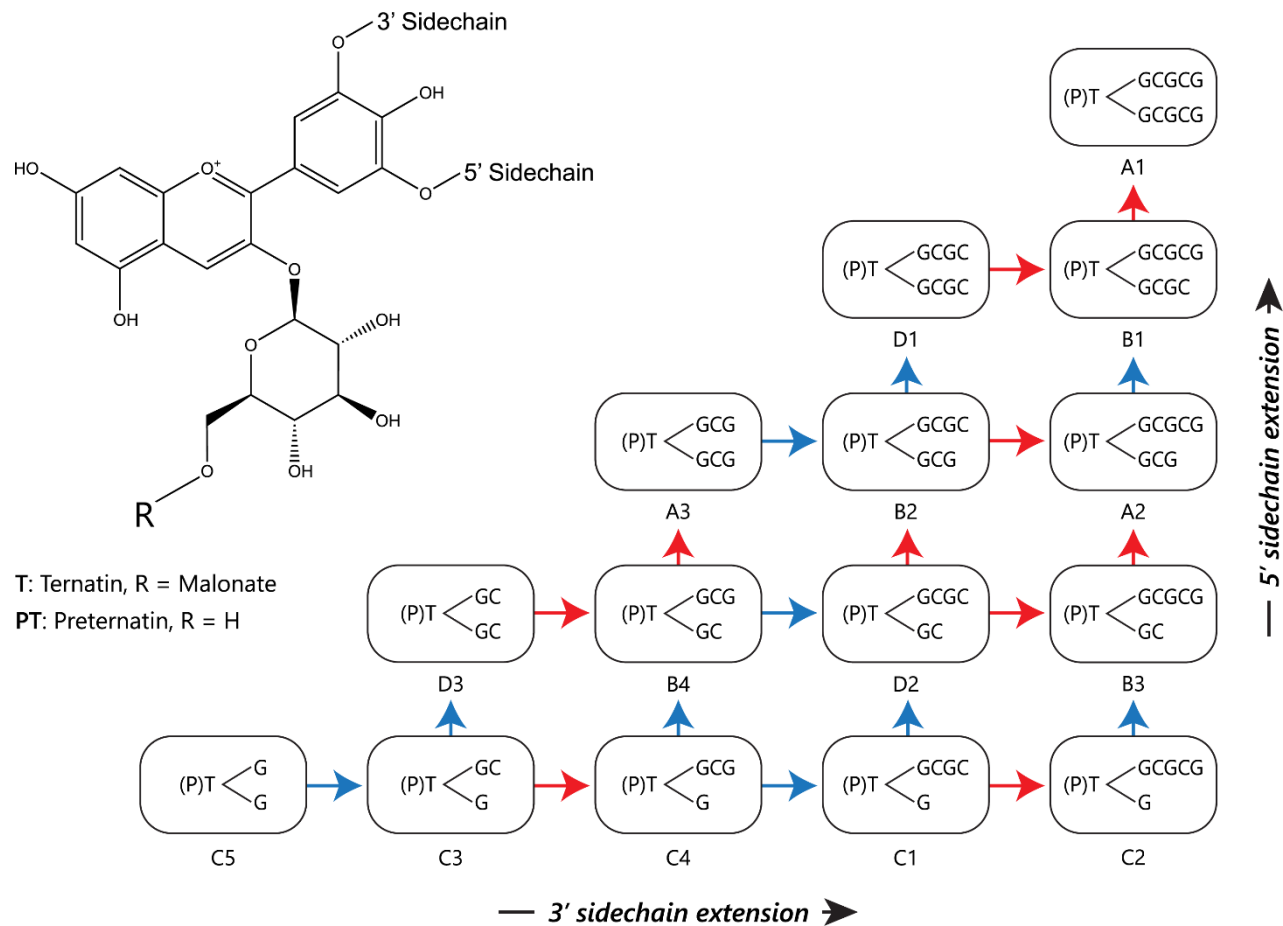
Figure 2.2 Possible biosynthetic pathways of ternatins. T: ternatin; PT: preternatin; G: glucose; C: *p*-coumarate. Symbolised structures show 3', 5' side chain construction of corresponding (pre)ternatins. Red arrows indicate glycosylation, blue arrows indicate acylation with *p*-coumarate.
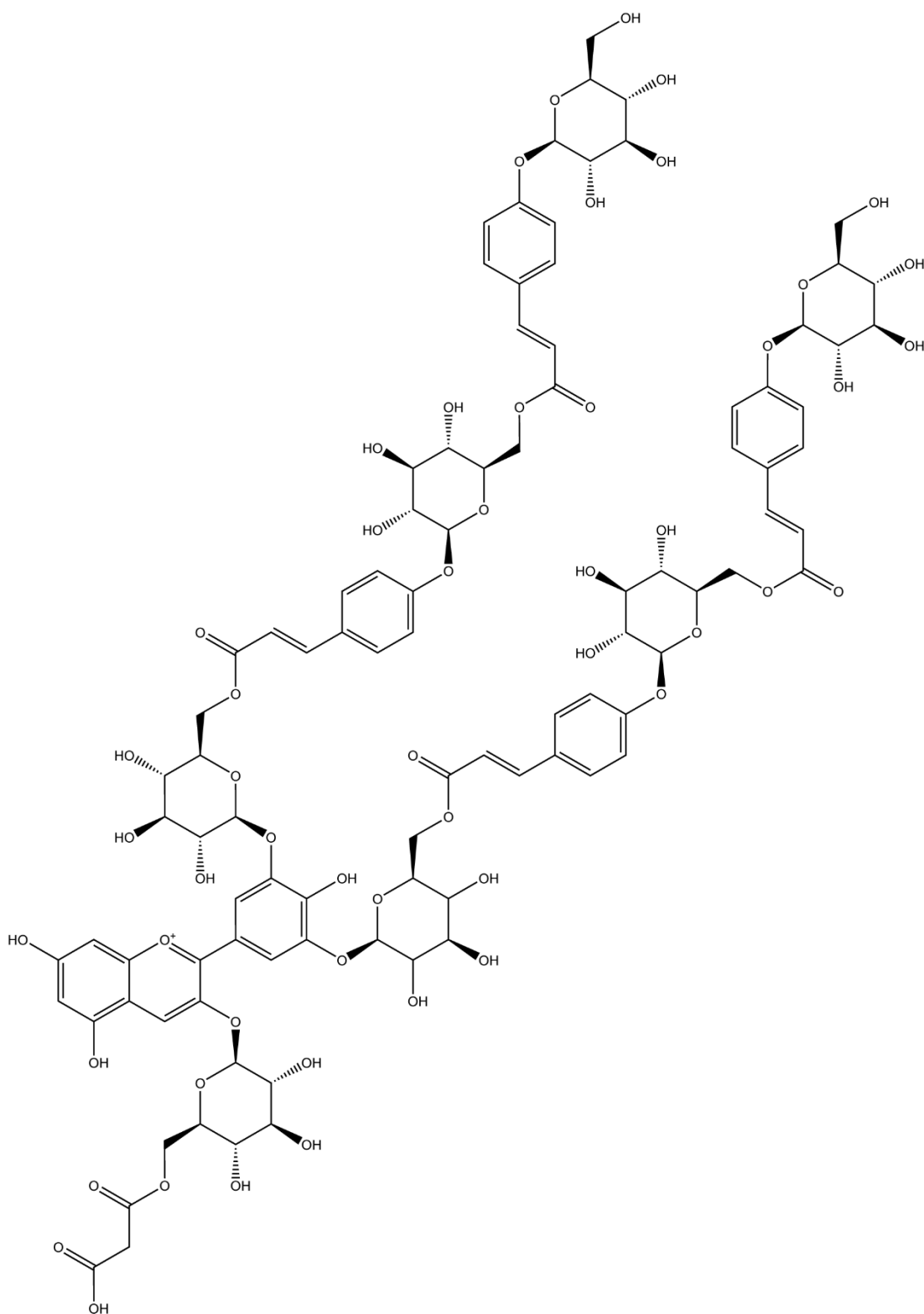
Figure 2.3 Structural formula of ternatin A1.

The anthocyanin profile of butterfly pea has also been demonstrated to change with flower age (Terahara *et al.*, 1998). Younger flowers show increased concentrations of pre-ternatins (ternatins lacking a 6"-O-malonyl group) and increased concentrations of smaller ternatin intermediates. In mature flowers, the most abundant ternatins are B2, B1 and D1, and these may be the most important pigments contributing to the overall blue colour.

2.1.3 Extraction of anthocyanins

Anthocyanin extraction typically follows similar methods, irrespective of the plant. Plant material due to be extracted is first freeze dried to inactivate enzymes present that would hydrolyse anthocyanic decorations. The first step of extraction involves the breakdown of plant material into smaller fragments or powder. This can involve sonication, bead milling or physical maceration using either a mortar and pestle, or some form of blender. Breakdown of the material causes the plant cells to rupture and allows better access for solvents, providing a better yield from the extraction.

Because of the polar nature of anthocyanins, they are highly soluble in water and alcohol and are insoluble in apolar organic solvents. They are generally extracted using aqueous alcohols; typically methanol, however ethanol can be used in place of methanol where the extract is used for food. Extraction solvents should also include a small amount of acid (0.1-1% HCl as standard) to convert the anthocyanin to their flavylium cation species and enhance their solubility in polar solvents.

It was once considered a rare occurrence for the anthocyanins to be acylated by an aliphatic acid; anthocyanins acylated with aromatic phenolic acids such as hydroxycinnamic acid or benzoic acid residues appeared to be more common (Bloom and Geissman, 1973). However, many malonylated anthocyanins have since been discovered by using alternative extraction techniques including weak acids. Extraction of these malonylated anthocyanins using standard procedures using methanolic HCl results in intermediate methyl ester formation at the free carboxyl group and the eventual loss of the acyl group (Harborne, 1988).

This degradation of the anthocyanin occurs primarily at the concentration and drying steps, where heat is applied to the solution to drive off solvent. With use of mineral

acids, anthocyanins are typically concentrated during the drying process, increasing the likelihood of esterification of the free carboxyl groups of acylated anthocyanins. Therefore, to prevent deacylation, it is often recommended to use weak acids such as acetic, tartaric or citric acids, or to use volatile acids such as trifluoroacetic acid. Working at low temperatures also increases anthocyanin stability and helps to prevent their degradation.

As the anthocyanins are insoluble in non-polar solvents, ethyl acetate, chloroform, hexane and diethyl ether can be used to semi-purify the plant extract. The organic phase of the partition contains the bulk of non-polar products such as plant sterols, carotenoids and chlorophylls, whereas the anthocyanins are localised in the aqueous phase. However, this can cause loss of some flavonoid aglycones, particularly the methylated flavones. The aqueous phase contains the polar flavonoids, which can then be analysed. Anthocyanins are generally reactive, and quickly degrade in solution. Therefore, they should be kept under a nitrogen atmosphere at low temperature and without exposure to light.

Growing interest in the antioxidant potential of anthocyanins, and their use in industry has led to a demand for more efficient extraction processes: reducing solvent consumption, reducing environmental impact and increasing extraction yield in shorter extraction times.

2.1.4 Chapter aims and objectives

This chapter presents the results from the isolation and structural elucidation of ternatins from butterfly pea. The main objectives for this chapter were to isolate each of the ternatins bearing decorations on both the 3' and 5' positions and to elucidate the structure of each of these isolates, in particular, to find key interactions between the delphinidin chromophore and the aromatic acyl residues.

Investigating the structural configuration of each ternatin may help to elucidate the structural characteristics which give rise to the enhanced stability of these anthocyanins in neutral conditions. Furthermore, structural elucidation should provide insight into the effects of iterative decorations on the resulting spatial geometry of these complexes.

2.2 Materials and methods

2.2.1 Analytical High-performance liquid chromatography – Mass Spectroscopy (HPLC-MS)

Analytical HPLC-MS was run using a QExactive HPLC system (ThermoFisher, *USA*), with a photodiode array (PDA) and quadrupole-orbitrap mass spectrometer as detectors. Samples were dissolved in 50 % acetonitrile : Milli-Q water, with 1% formic acid to a concentration of 5 mg mL^{-1}, and 1 µL (5 µg) was injected onto a Luna C18(2) column (4.6 x 100 mm, 3 µm). Products were monitored at 350 nm for UV-absorbing compounds, and 540 nm for anthocyanins.

The mass spectrometer was calibrating using Pierce calibration standards according to the manufacturer's instruction. Spectra were recorded using positive electrospray ionisation with a scan range of 200-2500 m/z, 70,000 resolution and data-dependent Top-3 MS^2. For MS^2, an isolation window of 4.0 m/z was used, and a normalised collision energy of 20%. HPLC was run with a flow rate of 0.6 mL min^{-1} and column temperature of 40 °C. The mobile phase was a gradient mixture of solvent A (1 % formic acid in Milli-Q water) and solvent B (1 % formic acid in acetonitrile) as follows: 5 % B (3 min), 5 % to 35 % (28.2 min), 35 % to 95 % (1.8 min), 95 % (2 min), 95 % to 5 % (0.8 min), 5 % (4.2 min).

2.2.2 Ternatin extraction

2.2.2.1 Crude extraction

100 g of dried flower petals of *Clitoria ternatea L.* were extracted with 1 L of 50 % methanol : Milli-Q water with 1 % formic acid for 24 hours at 4 °C. The extract mixture was then centrifuged at 7,200 x *g* , and the supernatant decanted and filtered through a nylon filter. The solids from the centrifugation and filtration steps were then reextracted a further two times, resulting in a collective extractant volume of 3 L.

The extraction solution was then dried *in vacuo* in a 1 L flask, using a rotary evaporator (Buchi, Switzerland) at 35 °C until the volume was reduced to ≈ 100 mL.

Due to the high degree of sugars and other matrix interferants, at this volume the extract formed a resin-like consistency. To liberate the extract for future purification steps, 900 mL of 1% formic acid : Milli-Q water was added, and the flask shaken on an orbital shaker until the resin at the bottom of the flask was no longer visible.

2.2.2.2 Extract sugar removal with XAD16N

The 1 L of extraction solution from section 2.2.1.1 was absorbed onto an open glass column (65 x 600 mm, i.d. x bed hight) packed with XAD16-N microporous resin at a flow rate of ≈ 5 mL min^{-1}. The column was washed with 2 L of 1 % formic acid : Milli-Q water at a flow rate of ≈ 10 mL min^{-1}, and the yellow eluent was discarded.

To elute the anthocyanins, the column was washed with 70 % ethanol with 1 % formic acid at a flow rate of ≈ 5 mL min^{-1}, and the eluent collected into a 2 L glass beaker. The column was washed until all the visible anthocyanins were removed from the column, and the eluent transferred to a 1 L rotary evaporator flask and dried *in vacuo* at 35 °C.

2.2.3 Polyvinylpolypyrrolidone (PVPP) fractionation

2.2.3.1 Preparing the adsorbent

Polyvinylpolypyrrolidone (PVPP) was first hydrated in 1 L of Milli-Q water in a 2 L beaker. The mixture was gently, but thoroughly, mixed to ensure all material had contact with the water, and the adsorbent allowed to hydrate over 24 hours.

After hydrating, the adsorbent was then resuspended by gentle mixing, and allowed to settle over a period of 2 hours. After the bulk of the resin had settled, the suspended fine particles were decanted and the resin : water mixture refilled to 1 L. The processes of mixing, settling and decanting fine particles were repeated until the supernatant was clear of all suspended particles.

2.2.3.2 Column chromatography

2.2.3.2.1 Preparing the column

After preparing the adsorbent as outlined in section 2.2.3.1, the volume of the aqueous portion of the adsorbent was reduced by decanting the water, until the mixture had a slurry-like consistency. A 65 mm i.d. glass column was half filled with Milli-Q water, and the adsorbent slurry was poured into the column. The column stopcock was opened halfway, and the column gently knocked along the sidewalls whilst the solvent drained to ensure even packing without the inclusion of air bubbles. The final bed height of the column was 120 mm.

2.2.3.2.2 Running the column

An extraction derived as described in section 2.2.3 was dissolved in 500 mL of 1 % trifluoracetic acid (TFA) in Milli-Q water, and 50 mL was loaded onto the column. The column was then washed with 400 mL of 1 % aqueous TFA, followed by 400 mL of 30 % ethanol with 1 % TFA, and finally with 400 mL of 60 % ethanol with 1 % TFA. As the separation on the column developed, three distinct bands were observed, which were collected as fractions and analysed by HPLC-MS (**Section 2.2.1**).

2.2.3.3 Flash chromatography

2.2.3.3.1 Preparing the flash cartridge

 PVPP (400 g) was dry sieved through a 60 mesh (250 μm) in batches of ≈ 50 g, and the material which passed through the sieve was discarded. After sieving, the stationary phase was then prepared in Milli-Q water. Like the method of packing the glass column in section 2.2.3.2.1, the absorbent underwent a cycle of mixing, settling and decanting of the fines until the aqueous portion was clear. An empty 490 g flash chromatography cartridge was then slurry packed with the adsorbent, and the flash cartridge was inserted into a Biotage Isolara flash chromatography system and flushed with 500 mL of Milli-Q water at a flow rate of 40 mL min^{-1}. The sintered glass frit was then fitted to the top of the stationary phase, with care taken to ensure no air entered the top of the cartridge.

2.2.3.3.2 Running flash chromatography

An extract derived by the method described in section 2.2.2 was dissolved in 100 mL of 20 % ethanol with 1 % formic acid. This was then loaded onto the flash cartridge by syringe, and flash chromatography run with a flow rate of 35 mL min^{-1}, using a solvent gradient of solvent A (1 % formic acid in Milli-Q water) and solvent B (1 % formic acid in ethanol) as follows: 5% B (200 mL), 5 % to 100 % (1900 mL), 100 % (400 mL).

After an initial void volume of 500 mL, fractions were collected every minute until the end of the run. Products were detected using a photodiode array at wavelengths of 540 nm and 350 nm, for detection of anthocyanins and flavonols respectively. At the end of the run, fractions corresponding to each of the peaks in the 540 nm absorbance trace were pooled together and dried in a rotatory evaporator. Fractions were then analysed using the method outlined in section 2.2.1.

2.2.4 Preparative HPLC

Samples were dissolved to a concentration of 400 mg mL^{-1} in 50 % acetonitrile with 1 % formic acid and 500 μL (200 mg) were injected onto a Gemini-NX C18 preparative column (150 x 21.2 mm, 5 μm). Preparative HPLC was run on a Dionex Ultimate 3000 (ThermoFisher, U.S.A.) at room temperature and with a flow rate of 20 mL min^{-1} using a variable wavelength detector set to 540 nm for detection of anthocyanins. For each of the PVPP fractions, the solvent gradient was as follows:

2.2.4.1 Series A ternatins (PVPP Fraction 1)

The mobile phase gradient was a gradient mixture of solvent A (1 % formic acid in Milli-Q water) and solvent B (1 % formic acid in acetonitrile) as follows: 5 % B (1 min), 5 % to 20 % (13.2 min), 20 % (2 min), 20 % to 95 % (1.3 min), 95 % (5 min), 95 % to 5 % (1 min), 5 % (5 min).

2.2.4.2 Series B ternatins (PVPP Fraction 2)

The mobile phase gradient was a gradient mixture of solvent A (1 % formic acid in Milli-Q water) and solvent B (1 % formic acid in methanol) as follows: 25 % to 29 % B (1 min), 29 % to 42 % (19 min), 42 % to 95 % (1 min), 95 % (5 min), 95 % to 25 % (1 min), 25 % (4 min).

2.2.4.3 Series D ternatins (PVPP Fraction 3)

The mobile phase gradient was a gradient mixture of solvent A (1 % formic acid in Milli-Q water) and solvent B (1 % formic acid in acetonitrile) as follows: 5 % B (1 min), 5 % to 10 % (1 min), 10 % to 30 % (24 min), 30 % to 95 % (1 min), 95 % (5 min), 95 % to 5 % (1 min), 5 % (4 min).

2.2.5 Semi-preparative HPLC

Fractions from the preparative HPLC step were dried in a rotary evaporator at 35 °C and pooled into a clean pre-weighed vial. The fractions were then redried, and their weights recorded. Each fraction was then dissolved to a concentration of 10 mg mL^{-1} in 50 % acetonitrile with 1 % formic acid.

Semi-preparative fractions were run on an Agilent 1290 (Agilent, U.S.A.) with a photodiode array set to 540 nm and 350 nm for detection of anthocyanins and flavonols respectively. For each preparative HPLC fraction, 500 µL (5 mg) were injected onto a Luna C18(2) (250 x 10 mm, 5 µm) column held at 40 °C. Samples were run at a flow rate of 3.5 mL min^{-1} with the following gradient mixtures of solvent A (1 % formic acid in Milli-Q water) and solvent B (1 % formic acid in acetonitrile).

2.2.5.1 Series A ternatin semi-prep

Semi-preparative HPLC was run on the fractions collected as described in section 2.2.3.4.1, using the parameters specified in section 2.2.3.5. For each of the fractions, the following solvent gradients were used:

2.2.5.1.1 Ternatin A1

For A1, the gradient conditions were as follows: 18 % B (1 min), 18 % to 24 % (10 min), 24 % (4 min) 24 % to 100 % (0.1 min), 100 % (4.9 min), 100 % to 18 % (0.1 min) and 18 % (4.9 min).

2.2.5.1.2 Ternatin A2

For A2, the gradient conditions were as follows: 18 % B (1 min), 18 % to 22 % (10 min), 22 % (4 min) 24 % to 100 % (0.1 min), 100 % (4.9 min), 100 % to 18 % (0.1 min) and 18 % (4.9 min).

2.2.5.1.3 Ternatin A3

For A3, the gradient conditions were as follows: 15 % B (1 min), 15 % to 23 % (10 min), 24 % (4 min) 24 % to 100 % (0.1 min), 100 % (4.9 min), 100 % to 18 % (0.1 min) and 15 % (4.9 min).

2.2.5.2 Series B ternatins semi-prep

All series B ternatin preparative HPLC fractions (section 2.2.3.4.2) were run using the following gradient: 5 % B (3.4 min), 5% to 10 % (0.1 min), 10 % to 23 % (31.9 min), 23 % (2 min), 23 % to 95 % (0.6 min), 95 % (2 min), 95 % to 5 % (0.1 min) and 5 % (5 min).

2.2.5.3 Series D ternatin semi-prep

Semi-preparative HPLC was run on the fractions collected as described in section 2.2.3.4.3, using the parameters specified in section 2.2.3.5. For each of the fractions, the following solvent gradients were used:

2.2.5.3.1 Ternatin D1

For D1, the gradient conditions were as follows: 15 % to 18 % B (1 min), 18 % to 25 % (15 min), 25 % (4 min), 25 % to 100 % (0.1 min), 100 % (4.9 min), 100 % to 15 % (0.1 min) and 15 % (4.9 min).

2.2.5.3.2 Ternatin D2

For D2, the gradient conditions were as follows: 15 % to 20 % B (1 min), 20 % to 30 % (10 min), 30 % to 100 % (0.1 min), 100 % (4.9 min), 100 % to 15 % (0.1 min) and 15 % (4.9 min).

2.2.6 Nuclear magnetic resonance (NMR) spectroscopy

^{1}H (600 MHz) and ^{13}C (150 MHz) spectra were acquired on a Bruker AVANCE NEO (Bruker Billerica, MA, USA) equipped with a Bruker TCI cryoprobe. Spectra were acquired at ambient temperature in deuterated dimethyl sulfoxide (DMSO-d_6, C_2D_6OS) with 10% deuterated trifluoracetic acid (TFA-d, $C_2DF_3O_2$). > 1 mg of each lyophilised anthocyanin isolate was dissolved in 200 µL of solvent and transferred to a 3 mm NMR tube. A total of 16 scans were recorded for ^{1}H spectra, and 4000 scans for ^{13}C spectra. Chemical shifts were recorded in parts per million (ppm) and were referenced to the DMSO-d_5 (C_2D_5HOS) residual solvent resonances (^{1}H δ = 2.5 ppm, ^{13}C δ = 39.52 ppm).

For each isolate, two-dimensional NMR experiments were performed to aid in assignment of ^{1}H and ^{13}C signals. These included COSY (correlation spectroscopy, number of scans = 4) for determining geminal and vicinal ^{1}H -^{1}H connectivity, and TOCSY (total correlation spectroscopy, number of scans = 4) to determine ^{1}H -^{1}H correlation within spin systems. In ^{13}C spectra, multiplicity of carbon atoms substituted with hydrogens were derived from HSQCed (multiplicity edited heteronuclear single quantum coherence spectroscopy, number of scans = 4), where methylene carbons (CH_2) are phased negative, and methine (CH) and methyl (CH_3) are phased positive.

Additional information about carbon atom multiplicity was acquired using DEPTQ (dimensionless enhancement by polarisation transfer, number of scans = 2000) which has the phasing of HSQCed but includes quaternary carbons (C) phased negatively. HSQC-TOCSY (heteronuclear single quantum coherence-total correlation, number of scans = 2) was used to solve overlapping cross-peaks by resolving signals into the ^{13}C dimension.

HMBC (homonuclear multiple bond coherence, number of scans = 8), which gives correlations between protons and carbons separated by two or three bonds, was used to assign carbon-carbon connectivity. ROESY (rotational Overhauser effect spectroscopy, number of scans = 24) was used to determine molecular conformation via through-space ^{1}H-^{1}H coupling.

2.3 Results

2.3.1 Polyvinylpolypyrrolidone (PVPP) fractionation

One of the major problems in isolating ternatins from butterfly pea were the large variety of anthocyanins present in the extractant. This presented the challenge that individual ternatin concentrations were low, compared to the total concentration of anthocyanins present in the analyte. Therefore, a separatory method to reduce the analyte complexity was required prior to separation using reverse-phase high-performance chromatography. In previous publications, separation of ternatins was achieved using polyvinylpolypyrrolidone (PVPP).

To demonstrate the separative properties of PVPP, an initial 1 g of crude extract (as produced using the method outlined in **Section 2.2.2**) dissolved in 20 % ethanol with 50 mM TFA was loaded onto a PVPP column with total bed volume of 400 cm^3. Separation was then performed using a stepwise gradient 1 bed volume of 20 % ethanol : 50 mM TFA, 40 % ethanol : 50 mM TFA and finally 60 % ethanol : 50 mM TFA with a flow rate of approximately 3 mL min^{-1}. Over the course of 8 hours, the analyte had developed into three distinct bands, as shown in Figure 2.4. Each of these bands were collected as individual fractions and, following drying *in vacuo* at 35 °C, were analysed using the analytical HPLC-MS method outlined in section 2.2.1.

As determined by the molecular ion mass, and retention time, analytical HPLC-MS (**Figure 2.5**) showed that each of the collected fractions contained primarily anthocyanins that had matching 3' and 5' terminal sidechain residues. Fraction 1 eluted earliest and contained ternatin A3 and A2, both of which have glucose residues as their terminal 3' and 5' sidechain residues, however ternatin A1 was missing. Fraction 2 contained ternatin B4, B3, B2 and B1 which all had a *p*-coumarate and glucose residue as their terminal residues, whereas fraction 3 which eluted last contained ternatin D3, D2 and D1, each of terminate with *p*-coumarate on both the 3' and 5' sidechains.
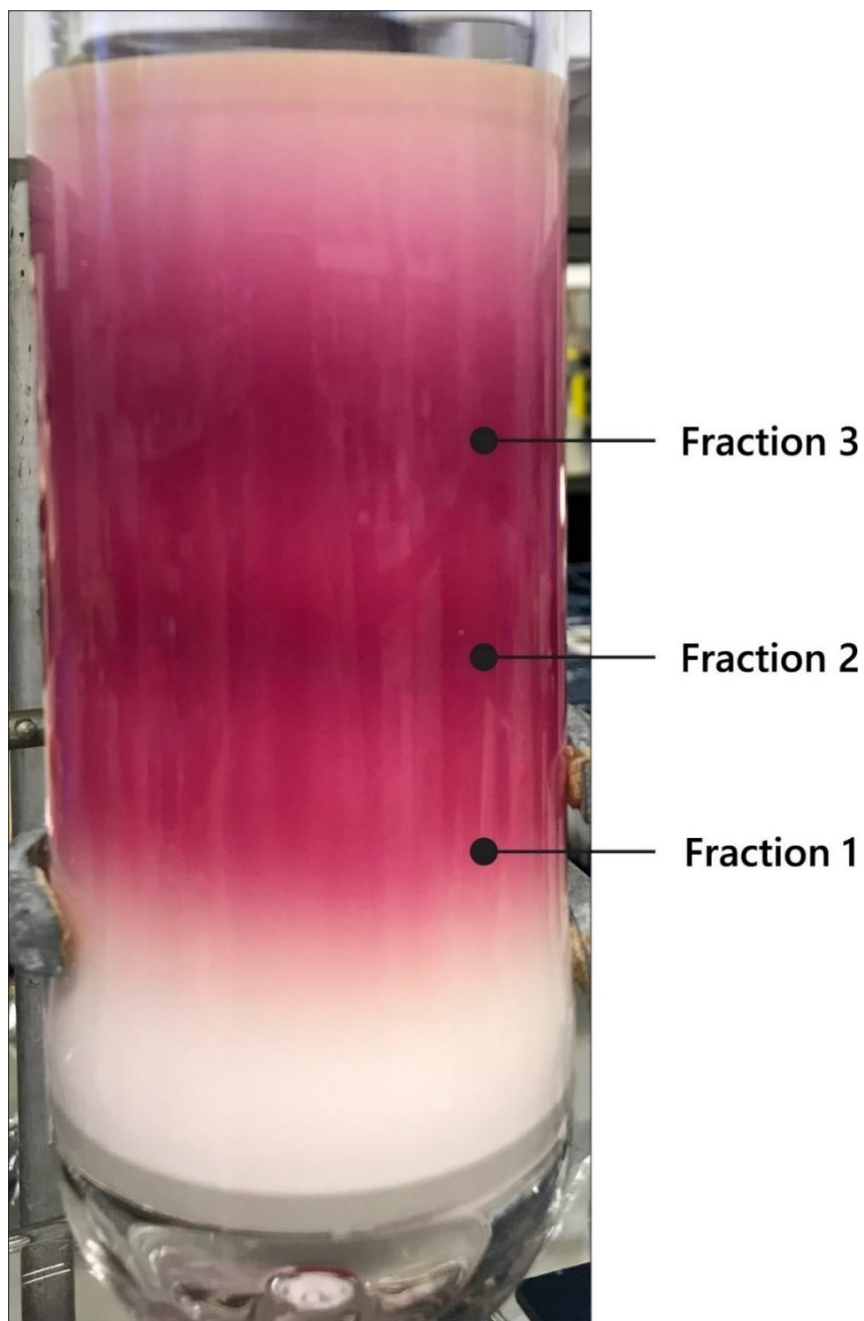
Figure 2.4 Fractionation of crude butterfly pea extract using polyvinylpolypyrrolidone chromatography. Elution was performed using a stepwise gradient of 20 %, 40 % and 60 % ethanol with 50 mM TFA. As the column developed, three distinct bands were observed. Fractions were collected as each band eluted from the column.
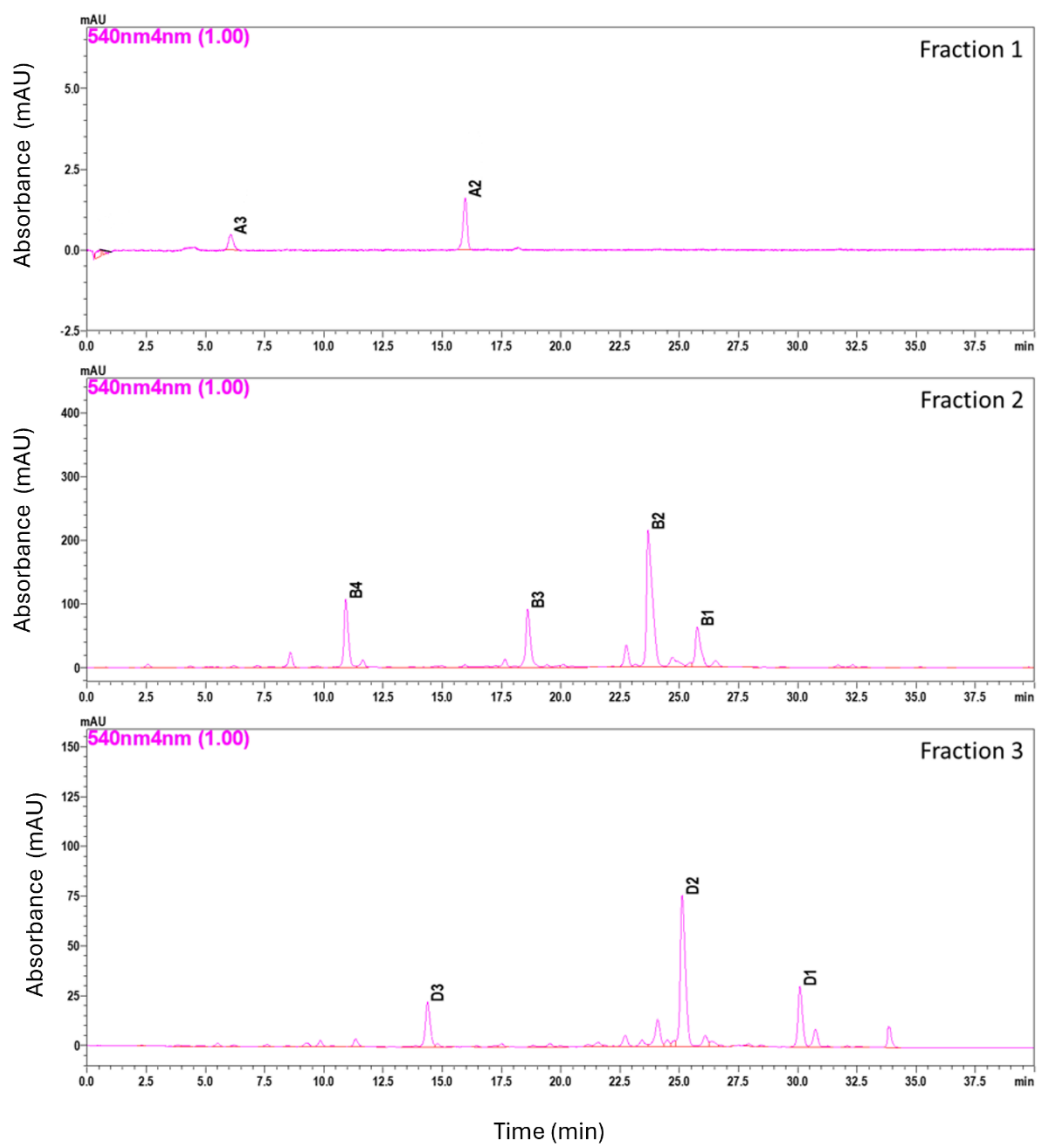
Figure 2.5 Analytical HPLC chromatograph traces of collected fractions from PVPP column chromatography. Chromatographs show absorbance at 540 nm and peak identity was confirmed by MS/MS. Fraction 1 contained only anthocyanins with two 3' and 5' end residues (series A). Fraction 2 contained primarily anthocyanins with different 3' and 5' residues (series B), while fraction 3 contained primarily anthocyanins with *p*-coumarate as their terminal residues (series D).

Across three independent runs, the PVPP stationary phase used to fractionate ternatin mixtures produced a few complications, namely the long development time in excess of six hours per experiment as well as the lack of reproducibility between experiments. Furthermore, as the separation was run on an open column, detection of bands and subsequent collection of fractions was done visually. This led to inconsistency in the contents of each fraction, such as the absence of ternatin A1 which likely eluted from the column prior to the fraction being collected.

Consequently, the next step was to determine whether the speed and reproducibility of this separation step could be increased. This was performed using a flash chromatography column packed with PVPP, and the separation performed using a Isolera Biotage. The total extractant mass of 6 g from 100 g of dried flower material was initially dissolved in 50 mL of 20 % ethanol : 1 % formic acid and loaded onto the flash cartridge. The flash cartridge was then run using a linear gradient of 5 % ethanol : 1 % formic acid for 200 mL, followed by an increasing organic concentration from 5 % ethanol : 1 % formic acid to 99 % ethanol : 1 % formic acid over 1900 mL at a flow rate of 35 mL min^{-1}. Absorbance at 540 nm was recorded as the separation proceeded to generate a chromatograph (**Figure 2.6**). After an initial void volume of 500 mL, 35 mL fractions were then collected each minute. Peaks observed in the chromatograph were pooled together, dried *in vacuo* yielding three fractions with masses 1.29 g, 1.98 g and 1.25 g for fractions 1, 2 and 3 respectively. Each fraction was then analysed using the analytical HPLC method described in section 2.2.1.
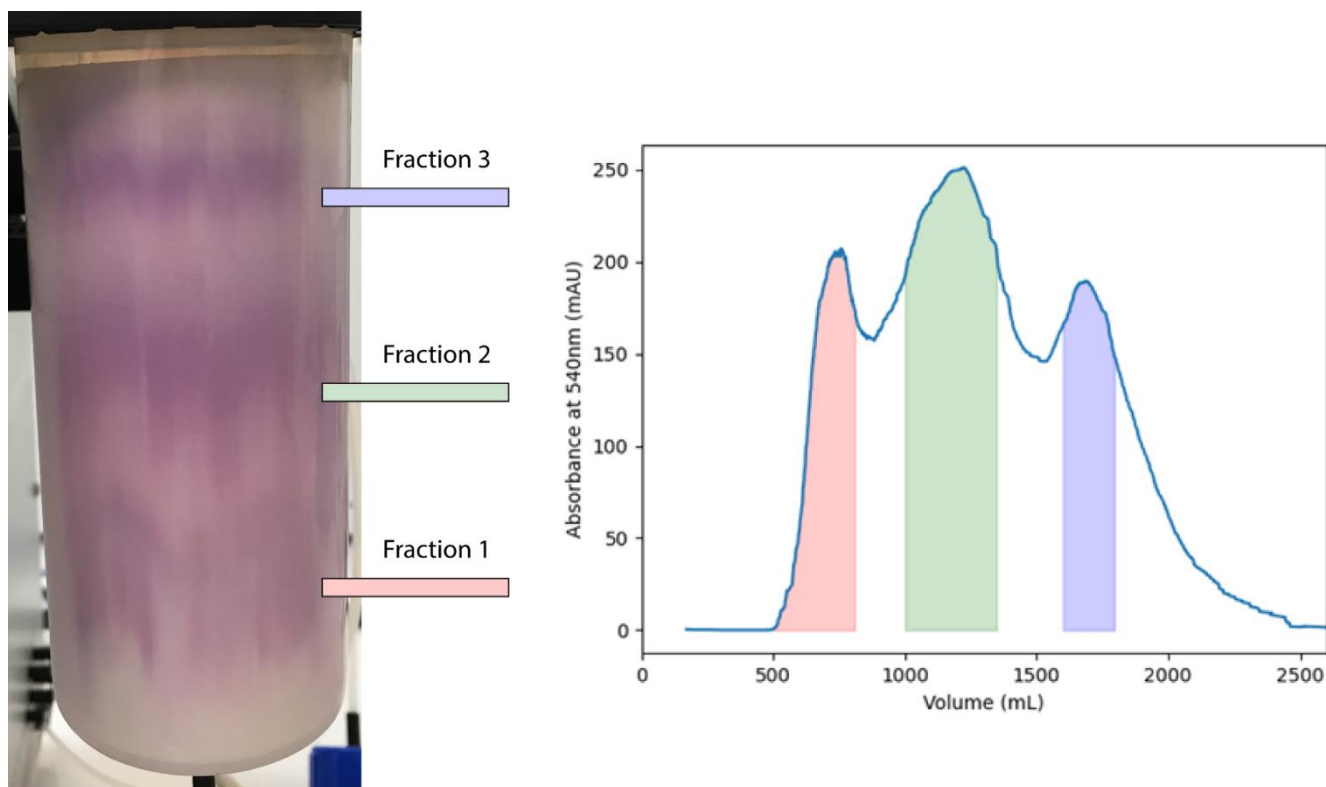
Figure 2.6 Flash chromatography of butterfly pea extract using polyvinylpolypyrrolidone. Chromatography was run using a gradient solvent system of ethanol : 1 % formic acid and water : 1 % formic acid at a flow rate of 30 mL min^{-1}. Absorbance at 540 nm was recorded to generate the chromatogram. The absorbance trace shows the development of ternatins into three distinct bands. The regions highlighted were collected and pooled for further analysis.

Despite running at a significantly increased flow rate compared to the previous column chromatography method, separation of the ternatin mixture was unaffected. Analytical HPLC confirmed that each of the fractions contained only the anthocyanins with matching terminal sidechain residues. The first fraction contained primarily group A ternatins A3 ($[M]^+$ = 1491.38 m/z), A2 ($[M]^+$ = 1799.47 m/z) and A1 ($[M]^+$ = 2107.56 m/z) (**Figure 2.7**). The next fraction to elute contained primarily anthocyanins from group B, ternatins B4 ($[M]^+$ = 1329.33 m/z), B3 ($[M]^+$ = 1637.42 m/z), B2 ($[M]^+$ = 1637.42 m/z), and B1 ($[M]^+$ = 1945.51 m/z) (**Figure 2.8**). The last fraction to elute contained primarily group D ternatins D2 ($[M]^+$ = 1475.37) and D1 ($[M]^+$ = 1783.46) (**Figure 2.9**), however ternatin D3 was missing from the fraction.
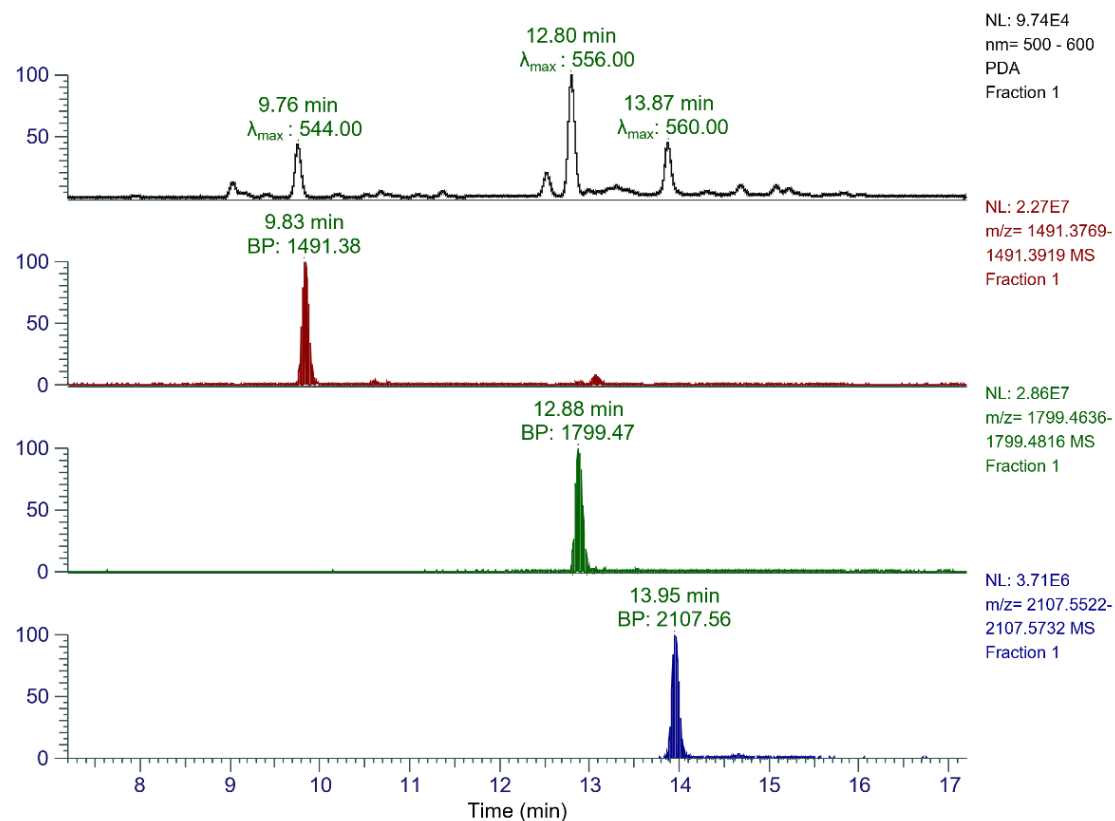
Figure 2.7 Analytical HPLC-MS trace of the first fraction collected from PVPP flash chromatography. The top chromatograph shows UV-VIS absorbance between 500 – 600 nm, while subsequent traces show total ion count. HPLC yielded three peaks with retention time: 9.76 min, 12.80 min and 13.87 min. These corresponded to ternatin A3 (1491.38 m/z), ternatin A2 (1799.47 m/z) and ternatin A1 (2107.56 m/z) respectively. Additional peaks corresponded to demalonyl-ternatins and minor contaminants from coelution of proceeding fractions.

Figure 2.8 Analytical HPLC-MS trace of the second fraction collected from PVPP flash chromatography. The top chromatograph shows UV-VIS absorbance, while subsequent traces show total ion count. HPLC yielded four peaks with retention time: 11.47 min, 13.70 min, 15.37 min and 15.81 min. These corresponded to ternatin B4 (1329.33 m/z), ternatin B3 (1637.42 m/z), ternatin B2 (1637.42 m/z) and ternatin B1 (1945.51 m/z) respectively. Additional peaks corresponded to demalonyl-ternatins and minor contaminants from coelution of preceding and proceeding fractions.
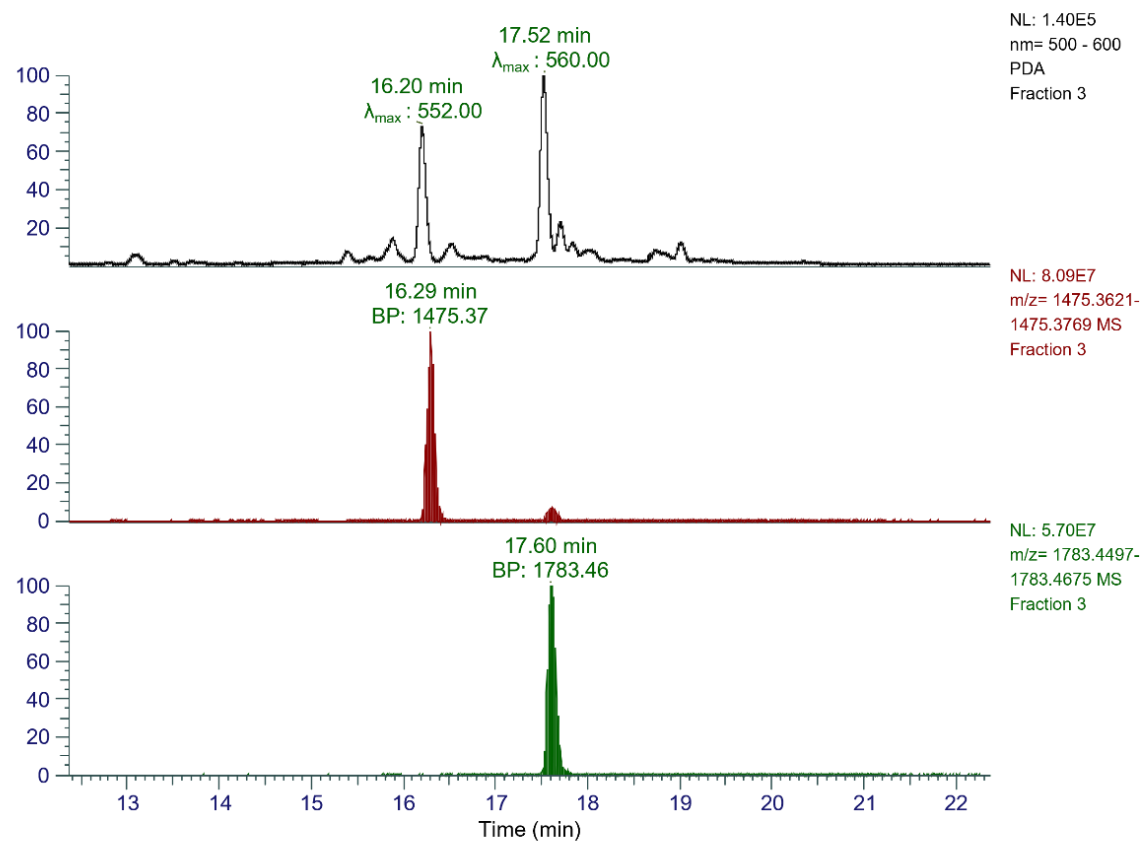
Figure 2.9 Analytical HPLC-MS trace of the third fraction collected from PVPP flash chromatography. The top chromatograph shows UV-VIS absorbance, while subsequent traces show total ion count. HPLC yielded two peaks with retention time: 16.20 min and 17.52 min. These corresponded to ternatin D2 (1475.37 m/z) and ternatin D1 (1783.46 m/z) respectively. Additional peaks corresponded to demalonyl-ternatins and minor contaminants from coelution of preceding fractions.

2.3.2 Ternatin Isolation

Following separation, each of the fractions from the PVPP separation were then separated using a preparative HPLC system as described in the methods in section 2.2.4. 200 mg of each fraction were dissolved in 500 µL of 50 % acetonitrile. Preparative HPLC of fraction 1 (using method outlined in **Section 2.2.4.1**) yielded three major peaks at retention times of 11.0 minutes, 13.6 minutes and 14.8 minutes (**Figure 2.10, A**). These peaks corresponded to ternatin A3, A2 and A1 with recovered masses of 33.2 mg, 82.8 mg and 43.5 mg respectively. Preparative HPLC of fraction 2 (using method outlined in **Section 2.2.4.2**) yielded four major peaks with retention times of 6.7 minutes, 9.5 minutes, 13.0 minutes and 15.0 minutes (**Figure 2.10, B**). Collection and analysis by analytical HPLC showed these peaks corresponded to ternatins B4, B3, B2 and B1 with yields of 25.9 mg, 36.1 mg, 142.8 mg and 76.5 mg respectively. Preparative HPLC of fraction 3 (using method outline in **Section 2.2.4.3**) resulted in two major peaks at retention times of 16.3 minutes and 18.1 minutes (**Figure 2.10, C**), corresponding to ternatin D2 and D1, of which 434.2 mg and 117.7 mg was recovered.

After preparative-HPLC, collected fractions were then re-isolated using semi-preparative HPLC as outlined in section 2.2.5. Semi-preparative HPLC yielded highly pure isolates of ternatins A1 (3.94 mg) (**Figure 2.11**), A2 (3.47 mg), A3 (5.39 mg) D1 (1.75 mg) and D2 (6.36 mg) Despite multiple attempts at isolating B1, B2, B3 and B4, contaminants remained. The masses recovered for these ternatins were 0.47 mg, 0.76 mg, 1.77 mg and 0.21 mg respectively.
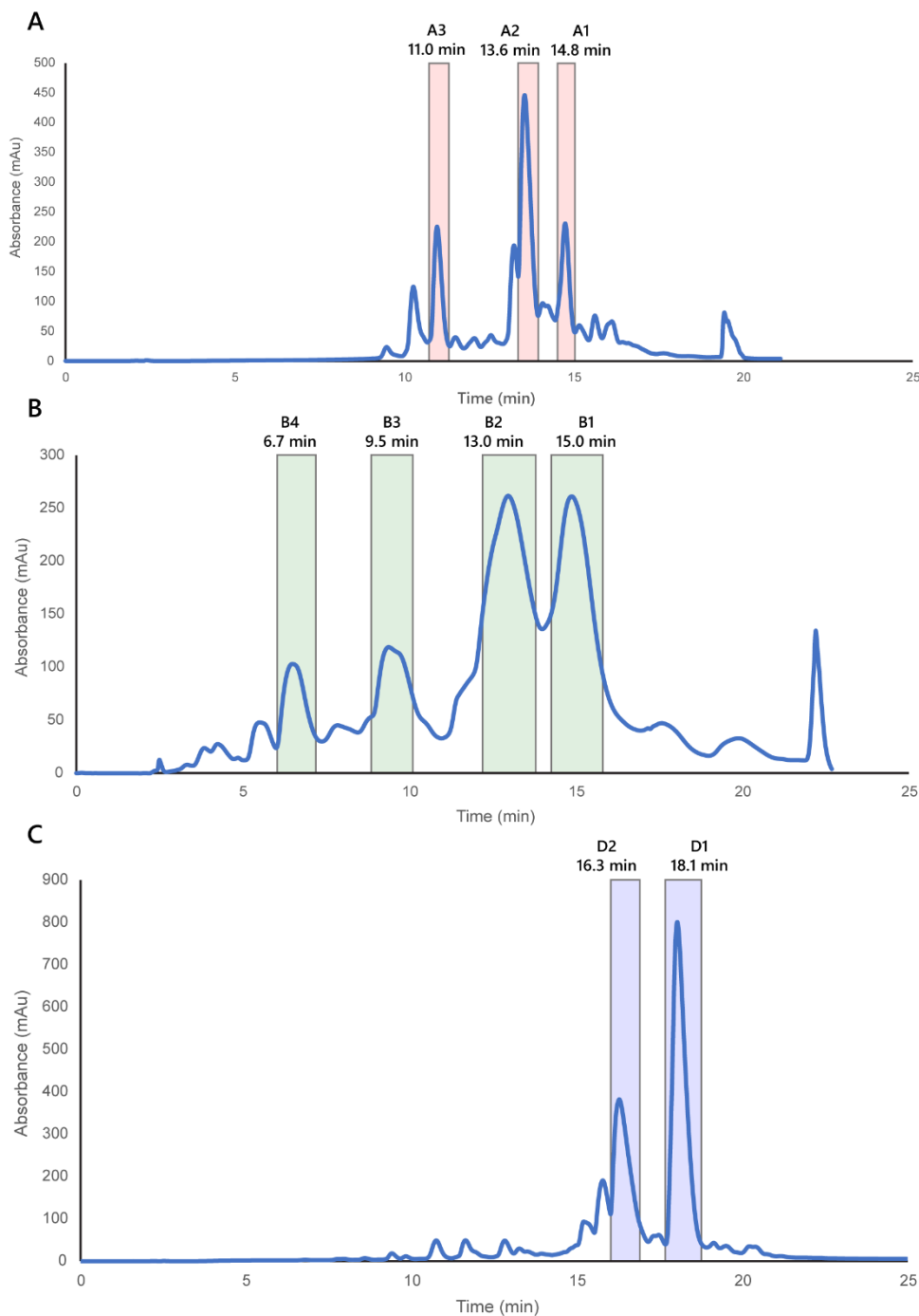
Figure 2.10 Preparative-HPLC chromatography of PVPP flash fractions .(**A**) preparative-HPLC of fraction 1 yielded three peaks at retention times 11.0 min (A3), 13.6 min (A2) and 14.8 min (A1). (**B**) fraction 2 yielded four peaks at retention times of 6.7 min (B4), 9.5 min (B3), 13.0 min (B2) and 15.0 min (B1). (**C**) fraction 3 yielded two peaks at retention times of 16.3 min (D2) and 18.1 min (D1). Peak identity was confirmed by HPLC-MS of the collected fractions.
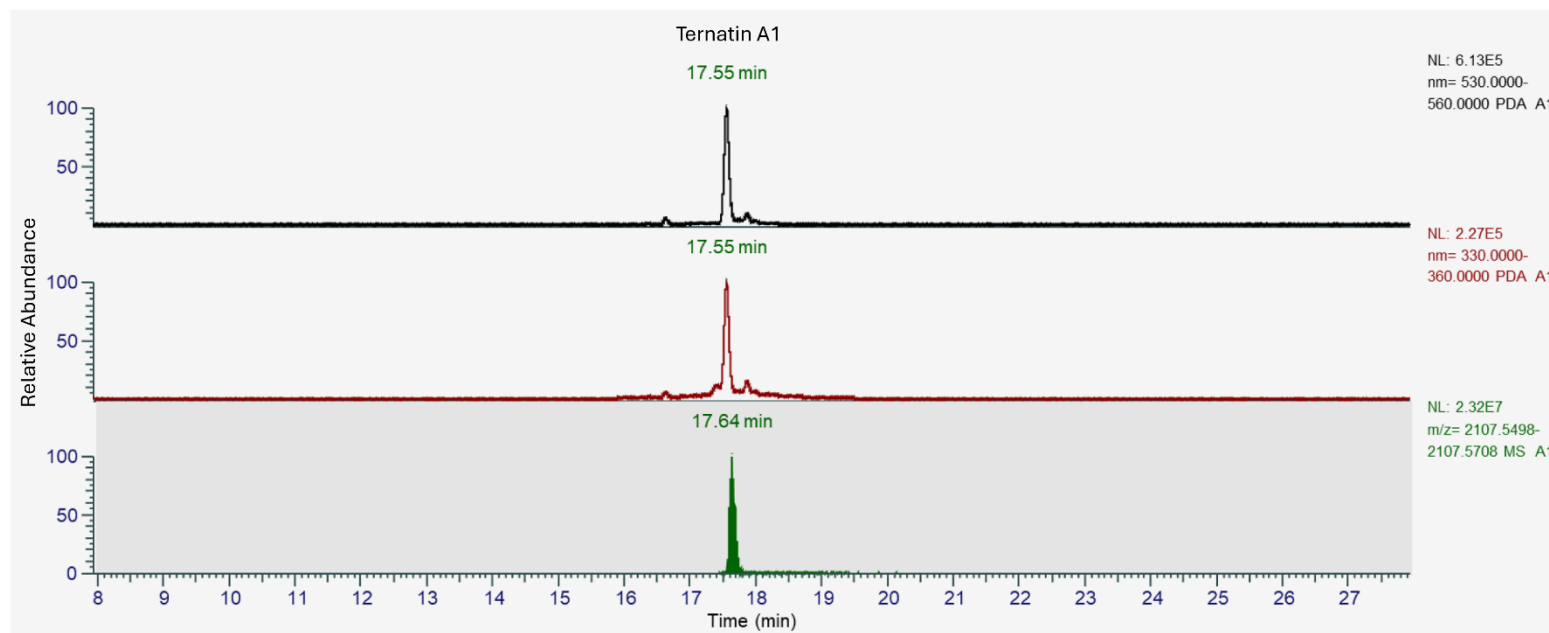
Figure 2.11 Analytical HPLC-MS trace of the isolated ternatin A1. The top chromatograph shows maximum UV-VIS absorbance between 530 – 560 nm for detection of anthocyanin contaminants, while the middle chromatograph shows maximum absorbance between 330 – 360 nm for detection of flavonol contaminants. Bottom trace shows extracted ion count within the mass range corresponding to ternatin A1. Analysis shows one major peak with retention time of 17.55 min and corresponding mass of 2107.56 m/z (ternatin A1), with very minor contamination.

2.3.3 Nuclear magnetic resonance (NMR) spectroscopy

Isolated ternatins were dissolved in 200 µL of DMSO-d_6 with 10 % TFA-d (9 : 1) and analysed using a combination of NMR experiments. Due to the presence of contaminants in isolates B2, B3 and B4, assignment of the NMR spectra was unattainable. The detailed structures of each ternatin isolate, A1, A2, A3, B1, D1 and D2, were established through ^{1}H- and ^{13}C − NMR spectroscopy, with bond carbon atom multiplicity established using HSQCed and DEPTQ, and connectivity identified through the use of COSY, TOCSY, HSQC-TOCSY, HMBC and ROESY (**Section 2.2.6**).NMR assignments were made *de novo*, and compared to assignments previously published in literature (Terahara *et al.,* 1990; Terahara *et al.,* 1996; Terahara *et al.,* 1998).

2.3.3.1 Delphinidin signals

Proton NMR (DMSO-d_6 600 MHz) spectra showed that all the ternatin anthocyanins have distinctive singlets of the delphinidin ring system found in the low magnetic field region of the spectra. The observed chemical shifts (δ_H) of these protons ranged between 8.55 – 6.85 ppm (**Table2.2**). The Dp 4 proton was observed to be the most deshielded, with resonances between 8.55 – 8.44 ppm, with the most decorated ternatins observed to be shifted further downfield compared to the less decorated ternatins. The Dp 2' and Dp 6' signals of the B-ring protons of ternatins A1, D1 and A3 (**Figure 2.12**) were observed to be fused as a broad singlet. This is likely the result of the symmetrical sidechains present on both the 3' and 5' positions of the anthocyanin B-ring, leading to similar chemical environments observed for both the Dp 2' and Dp 6' protons. In the ternatins with asymmetric sidechains, ternatins D2 and A2, this broad singlet is observed to split into two broad singlets, with the proton adjacent to the longer sidechain shifted slightly further downfield. Due to contaminants present in the ternatin B2 isolate, it was difficult to ascertain whether the signals of Dp 2' and Dp 6' were fused, or separated. However, previous NMR assignments showed these proton signals as a fused singlet observed at a chemical shift of 8.05 ppm (Terahara *et al.*, 1996). Both ternatin B1 and B2 have asymmetrical 3' and 5' sidechains, but both contain the first three glucose-acyl-glucose resides on the B-ring 3' and 5' positions. Therefore, decorations past the first three residues appear not to result in splitting of this singlet.

Table 2.2 ^{1}H NMR spectral data of isolated ternatins delphinidin ring systems (600 MHz, in DMSO-d_6-TFA-d = 9:1, δ ppm from residual DMSO signal = 2.5 ppm)

[1]H	T-A3	T-D2	T-A2	T-D1	T-A1
Dp-4	8.44	8.46	8.47	8.55	8.53
Dp-2'	8.00	7.99	8.00	8.02	8.00
Dp-6'	8.00	7.95	7.99	8.02	8.00
Dp-8	6.89	6.85	6.89	6.90	6.90
Dp-6	6.57	6.63	6.61	6.63	6.65

Table 2.3 ^{1}H NMR spectral data of isolated ternatins acyl decorations (600 MHz, in DMSO-d_6 : TFA-d = 9:1, δ ppm from residual DMSO signal = 2.5 ppm)

[1]H	T-A3	T-D2	T-A2	T-D1	T-A1
Ⅰ-α	6.11	6.03	6.12	6.08	6.1
Ⅱ-α	-	5.98	6.07	6.08	6.1
Ⅲ-α	-	6.30	6.41	6.28	6.41
Ⅳ-α	-	-	-	6.28	6.41
Ⅰ-β	7.26	7.22	7.27	7.25	7.26
Ⅱ-β	-	7.14	7.24	7.25	7.26
Ⅲ-β	-	7.41	7.49	7.42	7.50
Ⅳ-β	-	-	-	7.42	7.50
Ⅰ-2&6	7.11	7.07	7.11	7.12	7.12
Ⅱ-2&6	-	6.96	7.09	7.12	7.12
Ⅲ-2&6	-	7.37	7.52	7.39	7.53
Ⅳ-2&6	-	-	-	7.39	7.53
Ⅰ-3&5	6.81	6.83	6.8	6.83	6.83
Ⅱ-3&5	-	6.52	6.83	6.83	6.83
Ⅲ-3&5	-	6.71	6.98	6.73	7.00
Ⅳ-3&5	-	-	-	6.73	7.00

Figure 2.12 ^{1}H NMR spectrum of ternatin A3. The spectrum was recorded using a 600 MHz NMR spectrometer at ambient temperature (25 °C), with the sample dissolved in DMSO-d_6 with 10 % TFA-*d*. The structural formula of ternatin A3 is shown on the left.

2.3.3.2 Organic acid signals

The signals of the *p*-coumarate residues showed a high degree of symmetry within the proton spectra of ternatins A1, D1 and A3, further demonstrating the symmetry of the anthocyanin sidechains(**Appendix 2.1 – 2.5, (A)**). Within all *p*-coumarate residues, the α and β protons had large coupling constants of ≈ 16 Hz, indicating they are all in the trans (*E*) configuration. In the ^{1}H spectra of free *p*-coumarate (Oh *et al.,* 2022), the observed chemical shifts of the 2&6 and 3&5 proton signals are present at 7.56-7.46 ppm (overlapping the β signal) and 6.79 ppm respectively. Compared to each of the *p*-coumarate decorations of the ternatins, the signals of the 2&6 and 3&5 protons are shifted upfield (**Table 2.3**).

The change in chemical shift was greater for *p*-coumarate residues Ⅰ and Ⅱ, compared to residues Ⅲ and Ⅳ. This indicates that these shifts were likely the result of interactions between the magnetic fields of the *p*-coumarate, and the delphinidin ring systems, such as would be found within a stacked conformation. The stacking of acyl residues over the chromophore were further reinforced by the observation of long-range NOEs between the Dp-4 and Ⅰ & Ⅱ 2&6 protons (**Figure 2.13**).
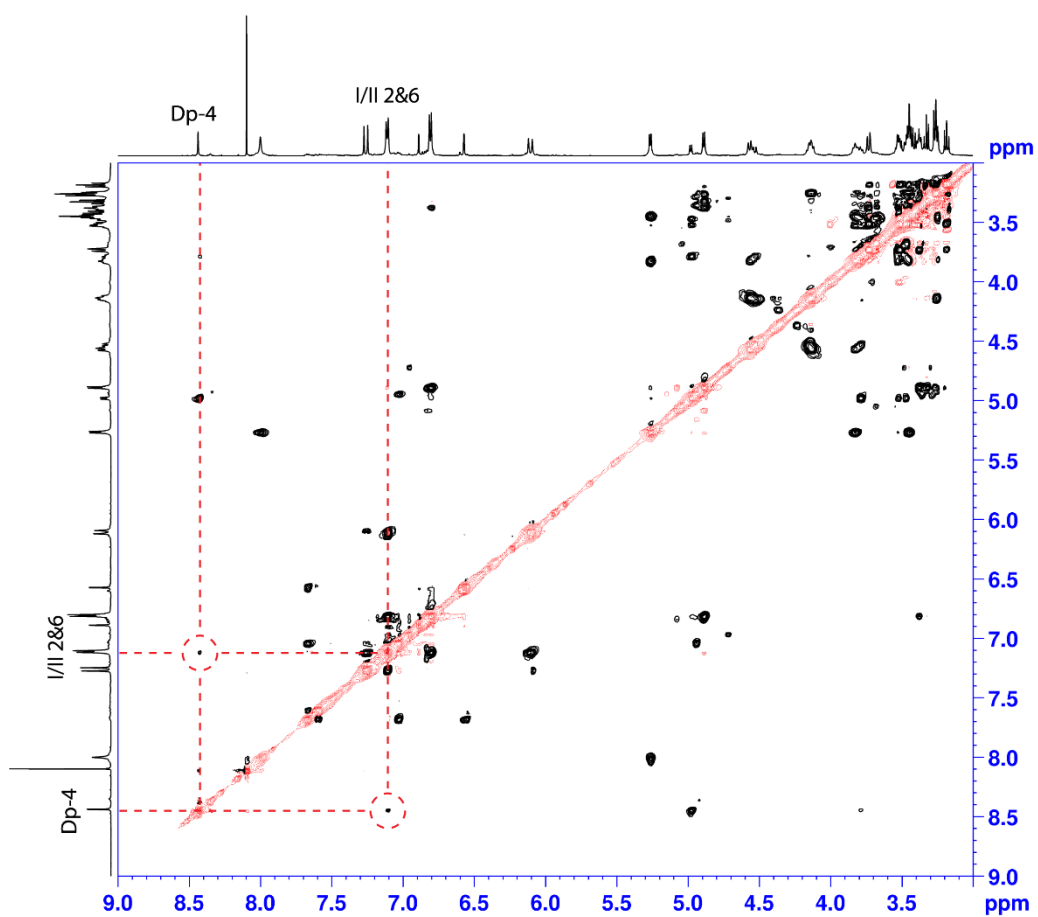
Figure 2.13 ROESY NMR spectrum of ternatin A3. The highlighted cross peak between Dp-4 and C_I and C_{II} 2&6 indicate close proximity between the delphinidin core and the acyl residues.

2.4 Discussion

2.4.1 Fractionation of crude butterfly pea extract using polyvinylpolypyrrolidone (PVPP)

Extracts of butterfly pea flowers contain 15 delphinidin-3-O-glucoside(6"-malonate)-3',5'-diglucoside derivatives, as well as a further 15 delphinidin-3,3',5'-O-triglucosides and several decorated flavonol glycosides. This presents a major challenge in the isolation of individual ternatins, especially as many of the products coelute on C18 stationary phase. For this reason, a method to fractionate the extract prior to isolation on HPLC systems was required.

Polyvinylpolypyrrolidone has been used previously in the initial fractionation of ternatins (Terahara *et al.*, 1996, 1998). Fractionation with PVPP column chromatography yielded three fractions (**Figure 2.4; Figure 2.5**) which contained only anthocyanins with matching terminal sidechain residues. The first to elute were primarily the A series ternatins, with two terminal glucose residues. B series ternatins with one glucose and one *p*-coumarate eluted next, followed by the final fraction, which contained only D series ternatins which two terminal *p*-coumarate residues. Collection of the fractions was carried out manually, by observation of the pigment bands as they eluted from the column. Consequently, ternatin A1 was missing from fraction 1, suggesting it eluted prior to collection of the fraction. Manual operation of the column presented difficulties for separation, namely in the form of reproducibility due to poor control over flowrate, mobile phase composition and detection of analyte. Coupled with long development times in excess of > 6 hours (Wrolstad and Struthers, 1971), PVPP is often overlooked as a stationary phase for the separation of anthocyanins, despite offering an orthogonal method of separation (Hrazdina, 1970) and having a larger binding capacity compared to macroporous and ion exchange stationary phases (Wen *et al.*, 2016).

Therefore, a flash chromatography method was developed, utilising PVPP and allowing automation of fraction collection, wavelength detection, and fine control over gradient and flowrate. Using this method, total separation time was reduced from 8 hours, to 1 hour, while maintaining similar fractionation of the extract (**Figure 2.6**). However, ternatin D3 was not present in the final fraction, nor in any of the

fractions collected from the region surrounding the corresponding absorbance peak. Additionally, fractions suffered from significant coelution, requiring re-chromatography to fully separate the ternatin groups. Further optimisation of the chromatography parameters such as flow rate, mobile phase system and gradient, and a more thorough analysis of the resulting fractions might localise D3 for collection, while also improving resolution and thus, the purity of fractions for downstream isolation.

2.4.2 Isolation of ternatins

The isolates for B2, B3 and B4 contained contaminants as observed in the analytical HPLC chromatograph at a wavelength of 280 nm. Despite re-isolation with methanol, these contaminants proved difficult to remove. The fact that after using orthogonal stationary phases, these contaminants were still present indicated that they share many chemical properties with the anthocyanins themselves. Therefore, further separation would have required re-isolation using a stationary phase with a different retention mechanism. Separation with pentafluorophenyl (PFP) stationary phases were attempted, where the retention mechanism relied on aromatic interactions between the analyte and stationary phase, but significant coelution remained. Further attempts could be attempted taking advantage of the characteristics of anthocyanins such as ion exchange chromatography because anthocyanins are weak diacids. However, use of such methods may prove problematic as the stability of anthocyanins at near neutral pH may result in degradation of the analyte. These refinements were not pursued further due to time constraints.

2.4.3 Ternatin Structural Conformation

The structures of isolated ternatins were confirmed by a combination of one dimensional (^{1}H, ^{13}C, DEPTQ), and two-dimensional experiments (COSY, TOCSY, HSQC, HMBC and ROESY) (**Appendix 2.1 – 2.5**). The NMR spectra of ternatins are defined by two distinct regions; the delphinidin and *p*-coumarate signals located in the high magnetic region (> 6 ppm), and the glucose and malonate signals in the low magnetic region (< 6 ppm). Chemical shifts of the *p*-coumarate protons were shifted upfield with respect to the reported signals of free *p*-coumarate (Oh *et al.*, 2022), an effect which was noticeably stronger for *p*-coumarate Ⅰ & Ⅱ (**Table 2.3**). Similarly, Dp

proton signals were shifted upfield with increasing decoration (**Table 2.2**), indicating interaction between the magnetic fields of the acyl residue, and the delphinidin scaffold. With the addition of NOE signals detected between Dp 4 and Ⅰ&Ⅱ 2&6 (**Figure 2.13**), NMR spectra show ternatins adopt a sandwich-type stacking arrangement with acyl residues localised near the anthocyanin C-ring, and closer proximity of these acyl residues with increasing number of decorations.

2.4.4 Conclusions

Since the initial isolation and structural characterisation of ternatins by Terahara *et al.* (1990; 1996; 1998), a large amount of work carried out on butterfly pea anthocyanins has been conducted on crude extractions, rather than isolated anthocyanins (Lakshan *et al, 2020;* Fu *et al.,* 2021*)*. This is due in part to complex mixture of ternatins, increasing the difficult of isolating ternatins. This chapter has demonstrated a high-capacity separation technique using polyvinylpolypyrrolidone, which can quickly fractionate ternatins depending on their terminal residues. Through using this fractionation technique, future research into the antioxidant and radical scavenging properties, as well as natural colourant evaluation, can be better correlated with the ternatin terminal decorations. Additionally, it is unlikely that individual isolates would be cost effective as replacement food colourants. From an industrial perspective, fractionation of ternatins into groups containing the most desirable properties, such as colour and stability could be achieved using this low-cost flash chromatography method. Using these isolates, the structure-colour relationship is explored in subsequent chapters.

Chapter 3

Effects of anthocyanin decoration on colour and stability

Chapter 3: Effects of anthocyanin decoration on colour and stability

3.1 Introduction

3.1.1 Anthocyanin structure – stability relationship

Anthocyanins are often suggested as replacement food colourants, due to their ability as weak diacids, to produce red and purple. These anthocyanins, such as those from purple-fleshed sweet potato, already see applications in beverages, confectionery and dairy products (Luo *et al.*, 2022). For the generation of blue colours, anthocyanins must be present in either the neutral or anionic base deprotonation states, and this generates a problem for their application; these anthocyanins protonation states are generally unstable. Copigmentation interactions, either intermolecular or intramolecular, can improve both stability and generate bathochromic shifts, improving the molar ratio of the neutral and anionic base and generating a bluer colour.

Butterfly pea produces ternatins, a group of polyacylated anthocyanins which include ternatin A1, the most decorated anthocyanin discovered to date. These anthocyanins have previously been isolated, and their structures characterised (Terahara *et al.*, 1996, 1998), however stability and colour characteristics have not yet been assessed. This is required to determine their suitability for potential food colourant applications.

3.1.2 Converting UV-VIS absorption spectra to colour measurements

It is important to quantify colour accurately when analysing pigments for their potential uses in manufacture. When looking at UV-VIS data, it can often be difficult to quantify the colour characteristics, particularly with spectra containing multiple peaks, where small changes in absorbance can have profound effects on the resulting colour.

Colour perception is the result of all wavelengths *not* absorbed by the compound. Therefore, to measure colour, the absorbance spectra must first be converted to transmission by taking the antilog of the absorbance value (*eq*. 3.1).

$$T = 10^{-Abs} \tag{3.1}$$

The transmission spectra are then weighted with an illuminant spectral power distribution curve. The standard illuminant used by the International Commission on Illumination (CIE) is termed D65K (**Figure 3.1**), which refers to the spectral emission of blackbody radiation at a temperature of 6,500 Kelvin and is equivalent to white noon-daylight in the Northern hemisphere.

The weighted spectra are then integrated with the three CIE colour match curves (**Figure 3.2**), which correspond to the sensitivity and distribution of cone cells present in the human eye. As $\bar{y}$ is the photopic luminosity function, these values are then scaled by calculating the photopic response of the illuminant (*eq.* 3.2 − 3.5).

$$X = \frac{100}{N} \int_\lambda I\,(\lambda)\,\bar{x}(\lambda)\,d\lambda \tag{3.2}$$

$$Y = \frac{100}{N} \int_\lambda I\,(\lambda)\,\bar{y}(\lambda)\,d\lambda \tag{3.3}$$

$$Z = \frac{100}{N} \int_\lambda I\,(\lambda)\,\bar{z}(\lambda)\,d\lambda \tag{3.4}$$

where N is the photopic response of the illuminant

$$N = \int_\lambda I(\lambda)\,\bar{y}(\lambda)\,d\lambda \tag{3.5}$$

$\text{I} \;=\; \text{transmittance spectra, and}$

$\bar{x}, \bar{y} \text{ and } \bar{z} \;=\; \text{the CIE colour match curves between 360 nm − 760 nm}$
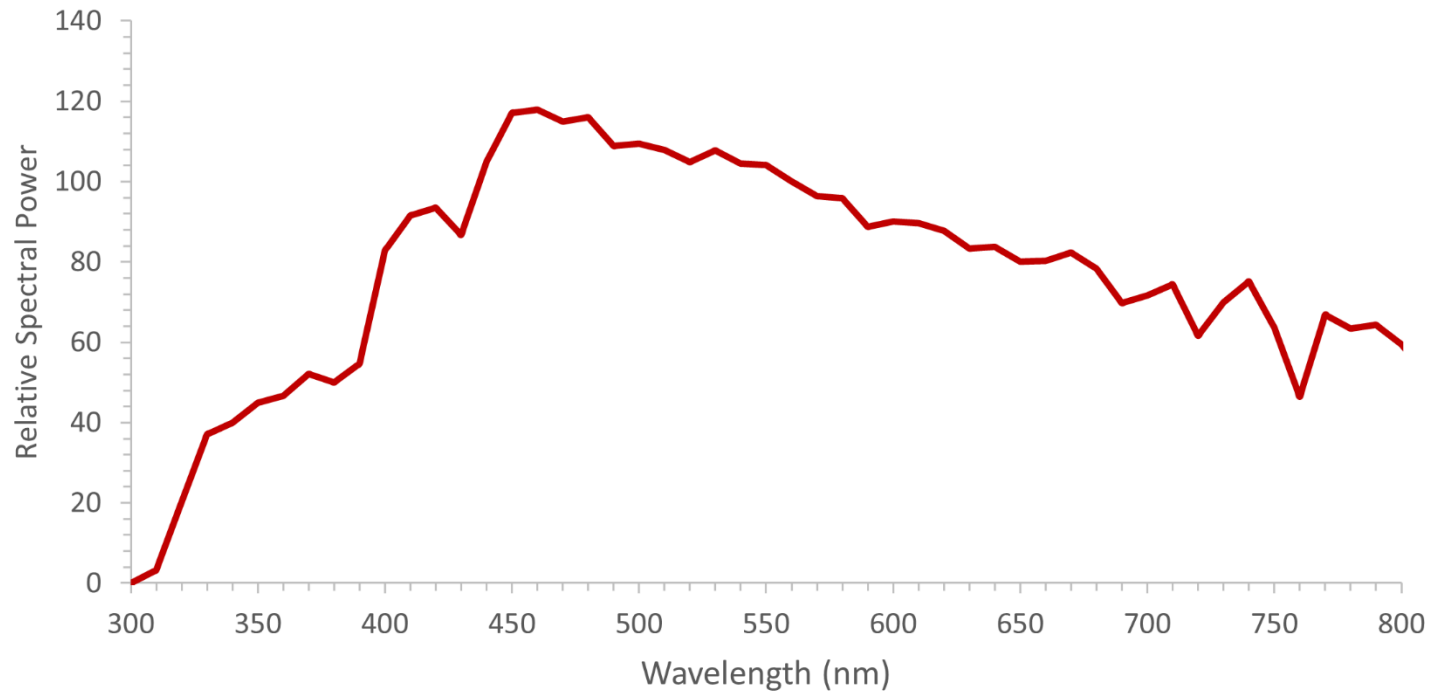
Figure 3.1 The relative spectral power of the white daylight (D65K) illuminant as a function of wavelength. Colour perception is dependent on the light that is returned from an object, and therefore reliant on the relative spectral of the illuminant
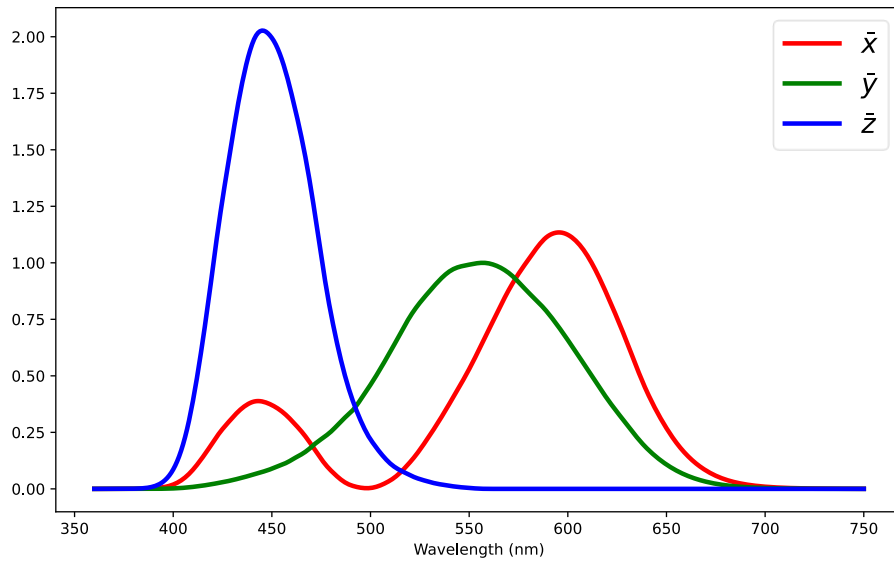
Figure 3.2 The three colour match functions produced by the Commission Internationale de l'Eclairage (CIE) of the 1964 CIE XYZ colour space, using a 10° observer angle.

3.1.3 Converting from CIE XYZ to CIE L*a*b* and CIE LHC coordinates

Although CIE XYZ coordinates meet the criteria of ascribing a numerical colour value to the UV-VIS spectra, the transformation lacks clarity with regards to how changes in the X or Z values affect the resulting colours. A further transformation with equations 3.6 – 3.12 can be utilised to convert X and Z to a* and b* respectively, where a* represents green-red (-a* = green; +a* = red) and b* represents blue-yellow (-b* = blue; +b* = yellow). The Y value is converted to L*, or lightness, where a value of 100 at the origin represents white, and 0 represents black.

The lightness value L*

$$L^* = 166\, f\left(\frac{Y}{Y_n}\right) - 16 \tag{3.6}$$

The red – green axis value a*

$$a^* = 500\left(f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right) \tag{3.7}$$

And the blue – yellow axis b*

$$b^* = 200\left(f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right) \tag{3.8}$$

*Where X, Y and Z have the meaning given in (**eq. 3.3**), (**eq. 3.4**) and (**eq. 3.5**) respectively, and X_n, Y_n and Z_n describe a white achromatic reference point*

$$X_n = 95.0489, \qquad Y_n = 100, \qquad Z_n = 108.8840 \tag{3.9}$$

The domain of the function f is divided into two parts, where t is given the meaning

$$t = \frac{X}{X_n}, \frac{Y}{Y_n} \; or \; \frac{Z}{Z_n} \tag{3.10}$$

The functional $f(t)$ is corrected as

$$f(t) = \begin{cases} \sqrt[3]{t} & if\ t > \delta^3 \\ \dfrac{t}{3\delta^2} + \dfrac{4}{29} & Otherwise \end{cases} \tag{3.11}$$

And δ is given the meaning

$$\delta = \frac{6}{29} \tag{3.12}$$

These Cartesian coordinates can then be transformed to polar coordinates with equations 3.13 – 3.14, where the colour is represented as the angle between the origin and data point, and chromaticity the distance from the origin (**Figure 3.3**). This ascribes a single numerical value to the colour and its intensity, which allows for easy comparison between different pigments.

The hue angle h°, where a* and b* have the meaning given in (**eq. 3.7**) and (**eq. 3.8**) respectively

$$h° = tan^{-1}\left(\frac{b*}{a*}\right) \tag{3.13}$$

And the chromaticity as

$$C^* = \sqrt{a^{*2} + b^{*2}} \tag{3.14}$$
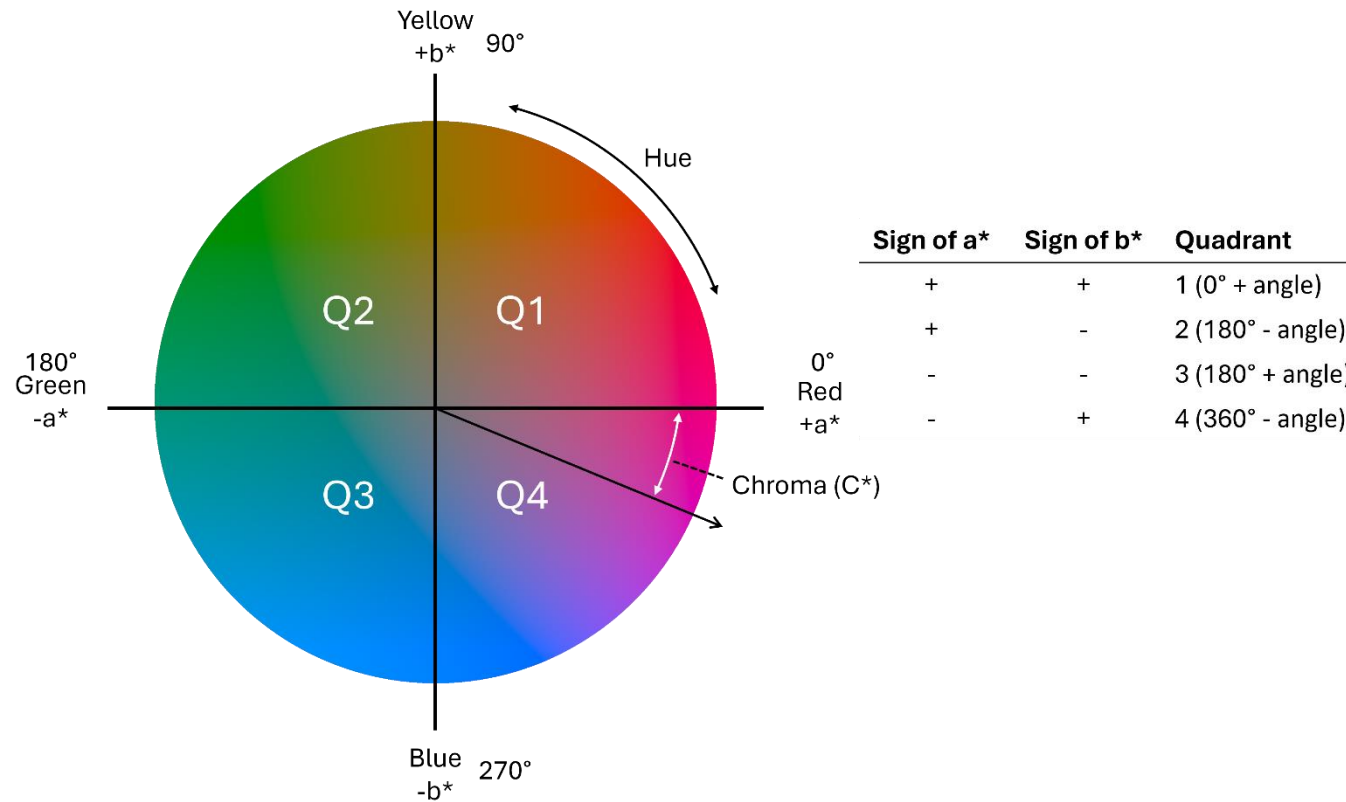
Figure 3.3 CIE LHC is a polar representation of L*a*b* Cartesian coordinates. Hue represents the arc tangent between the origin and data point, where 0° is red, 90° yellow, 180° green and 270° blue. Chroma is the distance from the origin and corresponds to the intensity of the colour. To calculate h°, first the quadrant is identified based on the sign of the a* and b* value. h° is the sum of the quadrant plus the angle calculated from arctan of $\frac{b*}{a*}$.

3.1.4 Aims and objectives

There were two main aims for this chapter. The first was to analyse the observed differences between the different ternatins to understand the effects of side chain decoration (glycosylation and acylation) on the colour and stability of the anthocyanins under different conditions. The extraction and purification of the ternatins provided a unique resource to examine sequential additions of aromatic acyl and glucose residues on the properties of the pigments, which could potentially illuminate the molecular mechanisms determining these properties

The second aim was to compare the colour and stability of individual ternatin species to the complex mixture extracted from flowers that can be used as a food colourant, to evaluate whether changes in ternatin composition might improve the functional properties with respect to their use as food colourants.

3.2 Materials and methods

3.2.1 Reagents and standards

Anthocyanin standard delphinidin 3-O-glucoside chloride (>95%) was purchased from Extrasynthèse (Genay, France). Analytical standards of *p*-coumarate (>95%) were purchased from Merck (Darmstadt, Germany). Ultrapure water was produced from a Milli-Q Direct water purification system (Millipore, Burlington, Mass., U.S.A.).

3.2.2 Buffer preparation

3.2.2.1 Potassium chloride : Hydrochloric acid (pH 1 – 2)

A stock solution of potassium chloride (KCl) was prepared by dissolving 14.91 g of KCl in one litre of Milli-Q water, giving a final concentration of 0.2 M. A second stock solution of hydrochloric acid (HCl) was prepared by mixing 16.5 mL of 37 % (w/w) and 983.5 mL of Milli-Q water, giving a final concentration of 0.2 M.

To prepare pH 1, 50 mL of KCl stock solution was added to 134 mL of HCl stock solution in a ratio of 0.37 : 1. For pH 2, 50 mL of KCl was added to 13 mL of HCl with a ratio of 3.85 : 1. The pH of these buffers was then measured with a pH meter and adjusted to their required pH values by addition of 1 M HCl and 1 M NaOH.

3.2.2.2 Citric acid :Disodium phosphate buffer (pH 3 – 8)

One litre of 0.2 M stock solution of disodium phosphate (Na_2HPO_4) was prepared by dissolving 28.38 g of Na_2HPO_4 in Milli-Q water. A second stock solution of citric acid ($C_6H_8O_7$) was prepared by dissolving 19.21 g of citric acid in one litre in Milli-Q water, giving a final concentration of 0.1 M. Buffers were then prepared as described in Table 1. The pH of each buffer was measured with a pH meter and adjusted with 1 M HCl and 1 M NaOH.

Table 3.1 Ratios of disodium phosphate and citric acid required for specific pH (per 100 mL)

pH	0.2 M Na_2HPO_4 (mL)	0.1 M $C_6H_8O_7$ (mL)
3	20.55	79.45
4	38.55	61.45
5	51.5	48.5
5.2	53.6	46.4
5.4	55.75	44.25
5.6	58	42
5.8	60.45	39.55
6	63.15	36.85
7	82.35	17.65
8	97.25	2.75

3.2.3 UV-VIS absorbance spectra

3.2.3.1 Single samples

1 mL of sample was loaded into a quartz cuvette with a fill volume of 1.5 mL and path length of 1 cm. The UV-VIS absorbance spectrum was then recorded between 340 nm and 750 nm, with a spectral resolution of 1 nm using a BioSpectrometer (Eppendorf, Germany). Solutions were kept at room temperature and protected from light. Data were collected in a comma-separated values file for further analysis.

3.2.3.2 Plate reader

UV-VIS absorbance spectra were recorded between 340 nm and 750 nm with a resolution of 1 nm using a CLARIOstar plate reader (BMG, Germany). Samples (200 µL) were loaded into the wells of a 96 well plate. Spectra were recorded at room temperature and protected from light. Path length correction was done autonomously by the instrument. Data were exported as comma-separated values files for further analysis.

3.2.4 Crude extract spectra

In triplicate, 5 mg of a crude extract of anthocyanins (extracted using the method outlined in **Section 2.2.2.1**) were dissolved in 1 mL of each of the pH buffers specified in sections 3.2.2.1 and 3.2.2.2. Aliquots (200 µL) of each pH buffered extract were then loaded onto a 96 well plate.

Samples were allowed to equilibrate at room temperature in the dark for 30 minutes prior to analysis. UV-VIS absorbance spectra were obtained using the method outlined in section 3.2.3.2. Spectra were recorded each day for a total of 14 days. In between readings, samples were transferred to microcentrifuge tubes to prevent evaporation in the plate. Samples were stored at room temperature, and with exposure to natural light.

This procedure was repeated for extracts after the sugar removal step using XAD16N (**Section 2.2.2.2**), and for each of the PVPP fractions (**Section 2.2.3.3**).

3.2.5 Ternatin isolate absorbance

Each isolate (from **Section 2.2.3.5**) was dissolved in 50 mM aqueous HCl to a concentration of 1 mM. These isolates were then diluted to a final concentration of 50 μM by addition of each buffer in sections 3.2.2.1 and 3.2.2.2. UV-VIS spectra were then recorded in triplicate using the method outlined in section 3.2.3.2.

3.2.6 Functional Colour: A tool for high-throughput colorimetry

In order to calculate CIE L*a*b* colour space coordinates for a large quantity of samples, a computational tool was developed using the python programming language (python, v.3.10). This algorithm automates the processing of UV-VIS absorbance spectra collected between 340 nm and 750 nm, producing CIE XYZ, CIE L*a*b*, sRGB and HSV colour space coordinates. As this algorithm is a post-processing technique, it provides a rapid assay method when used in conjunction with a plate reader, particularly when compared to the single cuvette capacity of traditional colorimeters. The program also features a graphical user interface, equipped with a file browser and batching system, allowing users to select multiple files to run within a single instance. Functional Colour can be accessed via the GitHub repository: https://github.com/AlanHoughton/Functional-Colour

3.2.6.1 Processing Inputs

Spectra generated from sections 3.2.6 and 3.2.7 were initially formatted in Microsoft Excel (Version 2110 Build 16.0.14527.20234), moving the "wavelength" column to column A, and sample spectra to all subsequent columns. The data were then loaded into the program as dataframe type objects using the pandas module (Mckinney, 2010). This dataframe was then converted into an array using the NumPy module (Harris *et al.*, 2020) and compared against the selected illuminant data to ensure that the spectral range and resolution were identical.

3.2.6.2 Converting spectra to illuminant weighted reflectance spectra

The absorbance spectra were converted to transmission spectra by taking the antilog of the absorbance value for each value in the array. The transmission spectra were then weighted with an illuminant spectrum by pointwise multiplication.

3.2.6.3 Calculating CIE XYZ from weighted transmission spectra

The weighted transmission spectra were converted to CIE XYZ colour space coordinates by taking the integral of the transmission spectra and each of the three tristimulus curves; $\bar{x}$, $\bar{y}$ and $\bar{z}$.

3.2.6.4 Converting CIE XYZ to CIE L*a*b*

A further transformation with equations 3.6 – 3.12were utilised to convert X and Z to a* and b* respectively, where a* represents green-red (-a* = green; +a* = red) and b* represents blue-yellow (-b* = blue; +b* = yellow). The Y value was converted to L*, or lightness, where a value of 100 at the origin represents white, and 0 represents black.

3.2.6.5 Converting Cartesian coordinates to a polar representation

The a* and b* values of the Cartesian CIE L*a*b* coordinates were further transformed to polar coordinates using equations 3.13 and 3.14.

3.2.6.6 Verifying calculated colours

To ensure the accuracy of the colour matching program, CIE L*a*b* colour measurements were made using the same solutions described in section 3.2.5. 1 mL of each solution was transferred to a cuvette with path length of 1 cm. Lab parameters were then recorded using a HunterLab UltraScan VIS spectrophotometer set to total transmission mode (TTRAN), with D65K illuminant and 10° observer angle. These measured values were then compared against values calculated using the python algorithm.

3.2.7 Thermal degradation assays

Thermal degradation of ternatin isolates was performed at pH 7 at 55 °C in an oven (Fenger *et al.*, 2019). Isolates were diluted to 50 µM in disodium phosphate : citric acid buffer preheated to 55 °C, and 500 µL aliquots immediately transferred to a microcentrifuge tube and plunged into dry ice until frozen. Each hour, a further aliquot 500 µL was removed and cooled in dry ice for a period of 8 hours, and a final sample at 24 hours.

After all samples had been collected, frozen samples were thawed on ice and loaded onto a 96 well plate. UV-VIS absorbance spectra between 340 nm and 750 nm were then collected using a CLARIOstar plate reader. The rate of colour loss was quantified as: $\% \ Remaining \ Coloured \ Species = 100 \times \frac{\text{Abs}_{\lambda\text{max}}(t)}{Abs_{\lambda\text{max}}(t_0)}$ and plotted as a function of time. The half-life of ternatins at pH 7 were compared to the values at room temperature.

3.3 Results

3.3.1 Stability of butterfly pea anthocyanins

Copigmentation interactions are particularly important for plants to generate and enhance colours to attract pollinators. However, copigmentation with non-covalently bound molecules, termed intermolecular copigmentation, is a comparatively weaker interaction when compared to intramolecular copigmentation. This is due to the requirements for a relatively higher concentration of copigments for the efficient formation of stacking type arrangements, which place a larger metabolic burden on the plant. In contrast to this, ternatins feature covalently bound acyl residues, which can form these stacking type arrangements with their attached copigment, termed intramolecular copigmentation. This is often more stable, due to the closer proximity of the covalently bound copigment, thereby increasing the probability for these interactions to occur.

Figure 3.4 shows the UV-VIS absorbance spectra of the crude butterfly pea extract in citric acid : sodium phosphate buffer at pH 2.5 to 7.5. At pH 2.5, the predominant protonation state of the anthocyanins is the cationic species, with a λ_{max} of 550 nm. As the pH increases to > 3 the neutral base was present, with a corresponding λ_{max} of 575 nm. Additionally, a new absorbance peak was present at a wavelength of 620 nm, a wavelength where the anionic base is typically observed. For true blue colours, the wavelength of absorbed light is within 580 – 640 nm. Extracts of butterfly pea flowers at pH 4.5 – 7.5 had a high degree of absorptivity within this region, and a resulting liquid colour of blue.

With anthocyanins which feature two or more proximal hydroxyl groups on the B – ring, the neutral species can chelate metal cations such as Al^{3+}, causing deprotonation and the subsequent formation of the anionic base. However, ternatins are decorated with glucosyl – acyl sidechains on the 3' and 5' hydroxyl sites, which occlude the interaction with metal cations. Therefore, this observed peak is likely the result of copigmentation interactions.

106
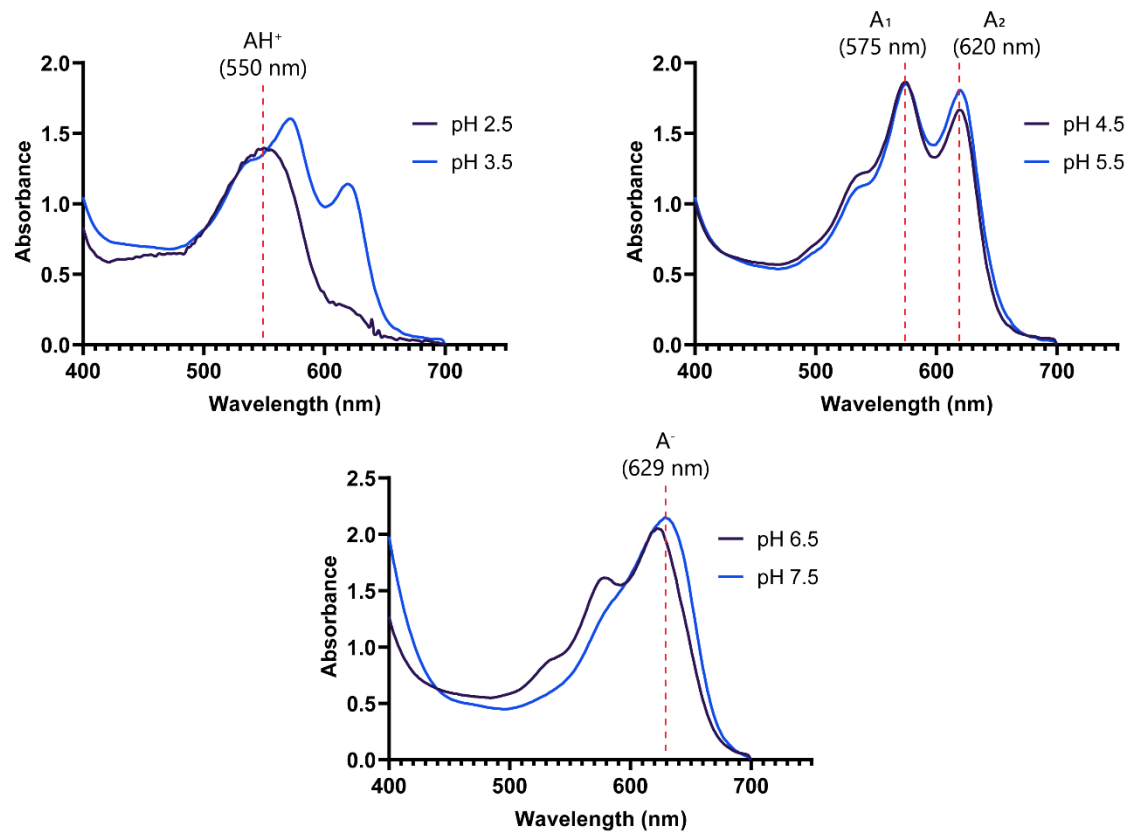
Figure 3.4 UV-VIS absorption spectra of 1 mg mL^{-1} butterfly pea crude extract between pH 2.5 and 7.5. The λ_{max} of the three major species was: 550 nm for the flavylium cation (AH$^+$), 575 nm for the neutral base (A$_1$), and 629 nm for the anionic base (A$^-$). An additional absorbance peak was present at a wavelength of 620 nm (A$_2$) between pH 2.5 and 5.5, before overlapping with the anionic base at pH 6.5 – 7.5.

To demonstrate the stability of anthocyanins extracted from butterfly pea, the absorbance spectra of an extract (as outlined in section 2.2.3.1) were recorded between 400 nm and 700 nm over a period of 79 days over a pH range of 2.5 to 7.5. Figure 3.5 shows the remaining coloured species, calculated as $\frac{A}{A_0} \times 100$ for each sampling point. As expected, the anthocyanins were most stable at an acid pH of 2.5, with a calculated half-life of 98.7 days. At this pH, the vast majority of anthocyanin species exist in the cationic protonation state, where the primary mechanism for colour loss is through the hydration of the C2 carbon, leading to the formation of the colourless hemiketal. As was shown in the ROESY NMR spectra of ternatin A3 (**Figure 2.13** and **Appendix 2.1 − 2.5, (I)**), ternatins formed sandwich-type stacking interactions around the B − and C − rings, which sterically hindered water from the C2 position, thereby stabilising the anthocyanin against hydration. Interestingly, the anthocyanins were less stable as the pH increased to 3.5 with a half-life of 41.2 days, before their stability increased at pH 4.5 and 5.5, with half-lives of 49.5 and 71.4 days.

As pH increased above pH 2, degradation occurred through poorly characterised oxidative and hydrolytic processes. These reactions take place principally at positions along the anthocyanin C − ring, namely the C2-C3, C3-C4 and C2-C1' bonds. At pH 5.5 the predominant species present was the neutral form, and the observed increase in stability suggested a preference for stacking type interactions occurring around the B − and C − rings. At pH values closer to neutral, anthocyanin stability was observed to decrease, with half-lives of 53.5 and 26.0 days for pH 6.5 and 7.5 respectively. This coincides with the formation of the anionic base, lending credence to the enhanced stability of the neutral species against hydrolytic and oxidative degradation at elevated pH.

Although crude extracts of ternatins have a high degree of stability and produce blue colours between pH 4 and 7, they represent a complex mixture of ternatin anthocyanins which collectively define their overall colour and stability.

A

B

	pH 2.5	pH 3.5	pH 4.5	pH 5.5	pH 6.5	pH 7.5
Half-life (days)	98.7	41.2	49.5	71.4	53.5	26.0
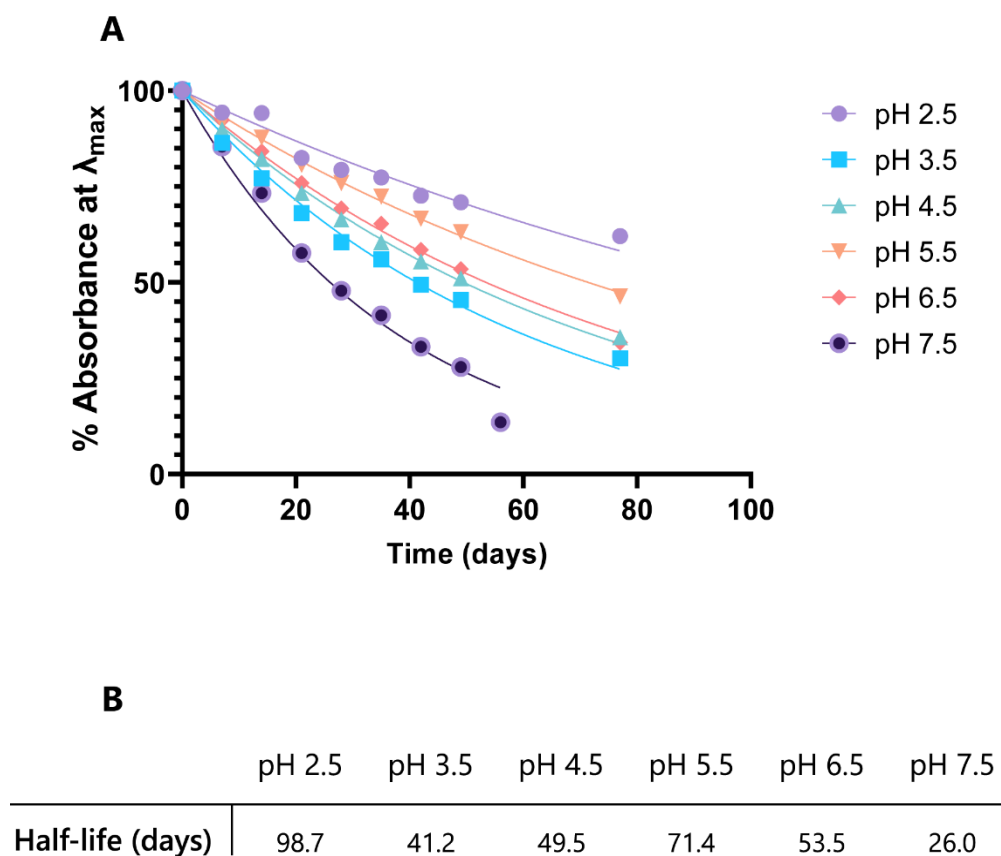
Figure 3.5 Stability of crude butterfly pea extract different pH values. (**A**) shows the remaining absorbance at λ_{max} for a 1 mg mL^{-1} sample of crude anthocyanin extract in citric acid : sodium phosphate buffer between pH 2.5 – 7.5, calculated as $\frac{A}{A_0} \times 100$; where A_0 was the initial absorbance value, and A the measured absorbance value. (**B**) shows the calculated half-life for absorbance at each pH. Half-life was calculated as $\ln\frac{2}{K}$, where K is the rate constant of the exponential decay equation used to fit the data. All R^2 values were greater than 0.95.

3.3.2 UV-VIS absorbance spectra of ternatin isolates

Because a crude extract of butterfly pea flowers represents a complex mixture of anthocyanins and other potential planar copigments, its colour is dependent on each of the components present as well as any potential intermolecular interactions. Therefore, to understand the contribution of each iterative decoration of ternatins on colour characteristics one needs to assay the overall colour and stability of each.

The individual ternatins (A1, A3, B1-B4 and D1) represent anthocyanins with progressive additions of glucose and aromatic acyl (*p*-coumarate) residues on the 3' and 5' carbon residues of the B-ring. By examining the colour and stability of each ternatin the contributions of the individual decorations to the properties of the pigments were defined, allowing the generation of hypotheses about the molecular mechanisms underpinning these properties.

3.3.2.1 Decoration of ternatins

Ternatin A1, the most decorated of the ternatins, has the largest bathochromic shifts of all the ternatins. The UV-VIS absorbance spectra of Ternatin A1 (**Figure 3.6**) shows four distinct peaks which evolved as the solvent pH increased. At acidic pH (pH < 2), the flavylium cation (AH^+) was the predominant species, with a λ_{max} of 554 nm. As the pH increased to 3, a shoulder peak emerged at a wavelength of 577 nm, corresponding to the neutral species (A). An additional absorbance peak was observed from pH 1 at a wavelength of 624 nm, which increased in amplitude as the pH increased, becoming a substantially distinct peak at pH 3. At these pH values the environment was too acidic for the formation of an appreciable amount of anionic base (A^-), which would typically be the species responsible for the peaks within this wavelength region. At pH values of 7 and 8, where the anionic base (A^-) begins to form, a peak with λ_{max} of 637 nm was observed.
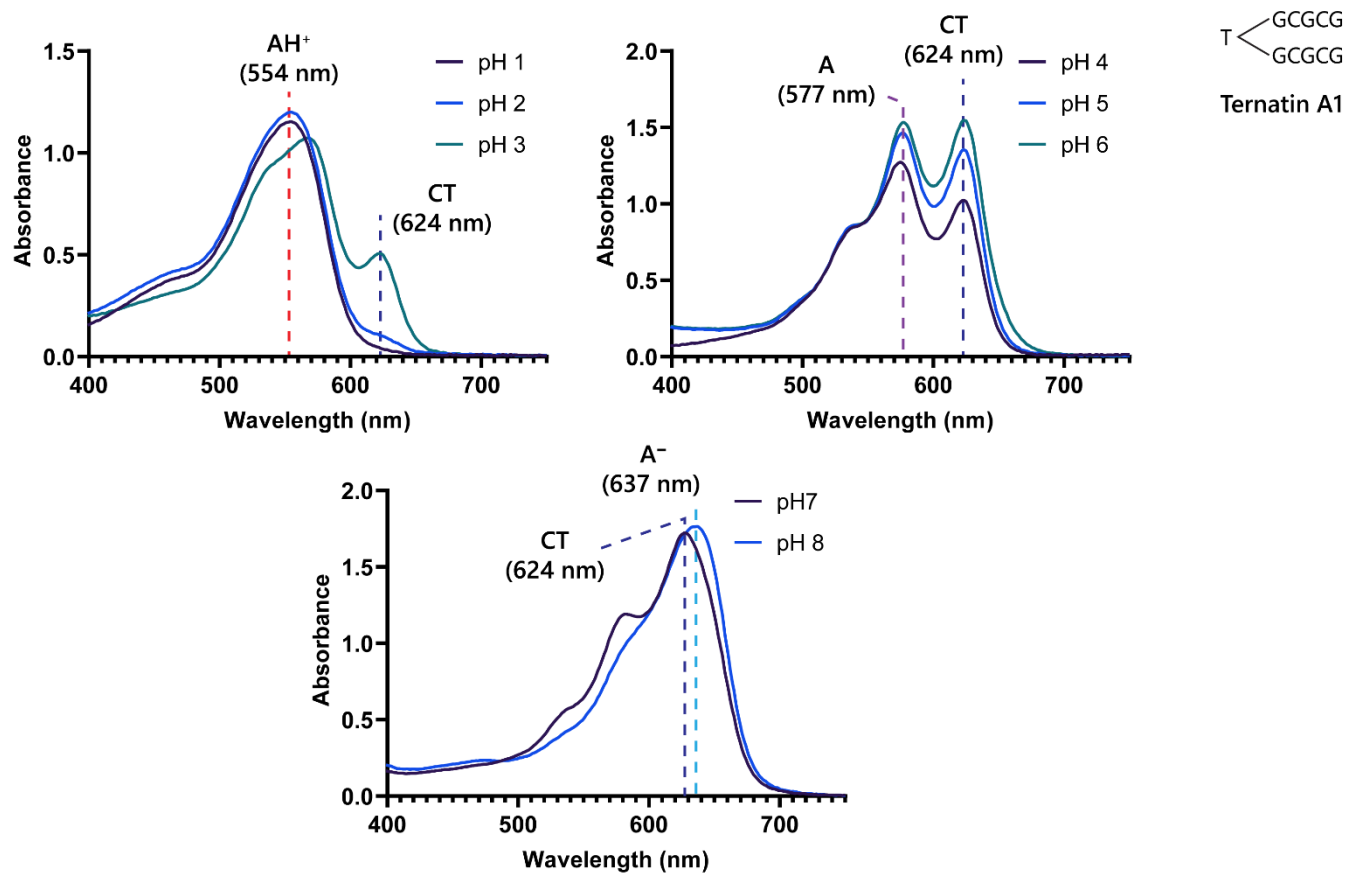
Figure 3.6 UV-VIS absorbance spectra of Ternatin A1 (60 µM) at pH 1 to 8. The three major species are: AH^+ (flavylium cation) with a λ_{max} of 554 nm, A (neutral species) at 577 nm and A^- (anionic base) at 637 nm. An additional peak appears from pH 3 onwards corresponding to the formation of a charge transfer state (CT) with λ_{max} of 637 nm. This absorbance peak is the primary contributor to the development of blue liquid colouration at slightly acid pH.

The additional peak observed is can likely be attributed to the interaction between the delphinidin chromophore and the acyl decorations, forming a charge transfer (CT) interaction. In such interactions, the overlapping delocalised π molecular orbitals of the chromophore and the acyl decorations and chromophore mix , with the concomitant rearrangement of electron density resulting in an orbital with lower energy. As a result, there is a decrease in the energy gap between the ground state and excited state, leading to the observed CT absorbance peak. The overall effect of this interaction is an increase in the wavelength of light absorbed, and thus a bluer colour.

The source of the blue colour exhibited by ternatins within the pH range of 4 to 7 is attributed to the charge transfer (CT) peak. Prior to the formation of the anionic base, each isolate demonstrates a comparable CT band in their UV-VIS absorbance spectra (see **Appendix 3.1 – 3.8**). However, the specific wavelength of this band is influenced by the number of acyl decorations present on the 3' and 5' sidechains (**Table 3.2**). In almost all cases, increasing the number of aromatic acyl decorations led to an increase in the λ_{max} of the CT absorbance peak.

3.3.2.2 Effect of acylation on UV-VIS spectra

The most significant bathochromic shifts of the charge transfer absorbance peak occurred upon acylation of the 3' sidechain, such as in the instance of B4 → D2 and A3 → B3, with an increase of 9 nm and 10 nm respectively (**Table 3.2**). In contrast, acylating the 5' sidechain of ternatin B3 to form D1 resulted in a mere 1 nm increase in λ_{max} of the CT peak. The magnitude of bathochromic shift for each protonation state displayed a consistent pattern: a 12 nm increase for B4 → D2, 10 nm increase for A3 → B3 for the cation, while the anionic base exhibited increases of 13 nm and 6 nm. Similarly, for the neutral base, there was a 5 nm increase for B4 → D2, 7 nm increase for A3 → B3, and 7 nm increase for the anionic base. Acylation of the 5' sidechain of B3 (to form D1) led to a significantly smaller increase for the λ_{max} of each protonation state, 1 nm for the anionic base and no change for the neutral base. A larger increase however was observed for the cationic species, with an associated bathochromic shift of 5 nm.

Table 3.2 λ_{max} of each protonation state of the ternatin isolates. AH$^+$: flavylium cation, A: neutral quinoidal base, CT: Charge transfer peak, A$^-$: anionic quinoidal base. Diagrammatic representation of each ternatin structure is given in the leftmost column, where: T = Delphinidin-3-O-glucoside(6"-malonate), G = glucose and C = p-coumarate.

	AH$^+$	A	CT	A$^-$	Structure
A1	554	577	624	637	T ⟨ GCGCG / GCGCG
B1	551	576	622	632	T ⟨ GCGCG / GCGC
D1	554	574	621	632	T ⟨ GCGC / GCGC
A2	550	574	620	632	T ⟨ GCGCG / GCG
B3	549	574	620	631	T ⟨ GCGC / GCG
A3	539	567	610	615	T ⟨ GCG / GCG
B2	547	572	617	625	T ⟨ GCGCG / GC
D2	550	572	618	626	T ⟨ GCGC / GC
B4	538	567	609	613	T ⟨ GCG / GC

B – ring rotation has been proposed to influence the wavelength of light absorbed, with a larger C3 - C2 - C1' - C2' dihedral angle resulting in a larger bathochromic shift. Copigmentation appears to be primarily driven by interaction between the aromatic moiety of the copigment through $\pi - \pi$ interaction, with sugar units contributing little to the interaction. This may explain the larger bathochromic shifts associated with acylation of the 3' side chain compared to acylation on the 5' side chain of the B ring. By acylating the 3' sidechain, the longer sidechain bearing more acyl residues experiences a greater magnitude of force acting on one side of the B – ring. This would induce a stronger "pull" on one side of the B – ring, thus increasing the dihedral angle and inducing a larger bathochromic shift.

3.3.2.3 Effect of glycosylation on UV-VIS spectra

Contrary to acylation, glycosylation of the ternatins led to significantly smaller effects on the λ_{max} of each of the protonation states, and of the CT complex. Glycosylation of 5' sidechain of B4, to form A3, led to an increase of 1 nm for the cationic species and CT complex, 2 nm for the anionic base and no change for the neutral base. Ternatin D2 can undergo glycosylation forming regioisomers B2 or B3, with the glucose added to either the 3' or 5' sidechain respectively. Upon glycosylation of the 3' sidechain forming B2, a decrease in λ_{max} of 3 nm was observed for the cationic species, and 1 nm for the CT complex and anionic base. Conversely, glycosylation of the 5' sidechain forming ternatin B3 caused a bathochromic shift of 2 nm for the neutral base and CT complex, and 5 nm for the anionic base.

Decoration with sugar units increases the hydrophilicity of the sidechain, and hence the interaction strength with the bulk solvent, effectively overcoming some of the interaction energy between the acyl residues and chromophore rings. This is partially offset by the potential for formation of hydrogen bonds between the π – orbitals of the chromophore and the O – atom of the sugar hydroxyl.

However, as a sugar unit added to the 3' sidechain is the fifth member of the sidechain (B3 -> A2), the distance between the sugar and chromophore was too great any for potential interactions. Instead, this interaction appeared to weaken the interaction between the sidechain and chromophore and may have resulted in a reduction of the unilateral forces that influenced the dihedral angle. Upon glycosylation of the 5' sidechain (B4-> A3; B1-> A1) however, the potential interactions between the sidechain and chromophore were weakened, thereby increasing the asymmetric forces which may act to rotate the B – ring. However, the impact of this on colour as defined by the bathochromic shift of the CT was limited to 1 nm (B4->A3) or 2 nm (B1-> A1).

3.3.3 Sidechain decorations influence on stability:

As reported in section 3.3.1, colour loss of anthocyanins occurs in two main ways: the first involving water addition at the C2 of the pyrylium ring and the formation of the colourless hemiketal. This is then followed by the opening of the pyrylium ring and eventual isomerisation, forming the cis-chalcone and trans-chalcone respectively. This degradation route occurs only with the flavylium cation, requiring an electrophilic region localised to the C2 position. The second route of degradation is facilitated by oxidative and hydrolytic reactions which occur at bonds positioned within the anthocyanin C – ring, with autoxidation presumed to occur by the spontaneous generation of H_2O_2 by electron or hydrogen atom transfer from the neutral or anionic base to O_2 (Dangles and Fenger, 2018).

As ternatins produce a blue colour at near neutral pH, to determine the stability of each ternatin isolate stored in an aqueous solution the UV-VIS absorbance spectra were sampled over a period of 37 days at pH 7 at room temperature. The absorbance at λ_{max} were recorded from each sampling point, and the residual coloured species were then determined as the percentage of remaining absorbance at λ_{max}, calculated as $\frac{A}{A_0} \times 100$ (**Figure 3.7**). The half-life was calculated from the rate of decay of the peak at λ_{max}, and represents the time taken for the absorbance to equal half of the initial starting value.
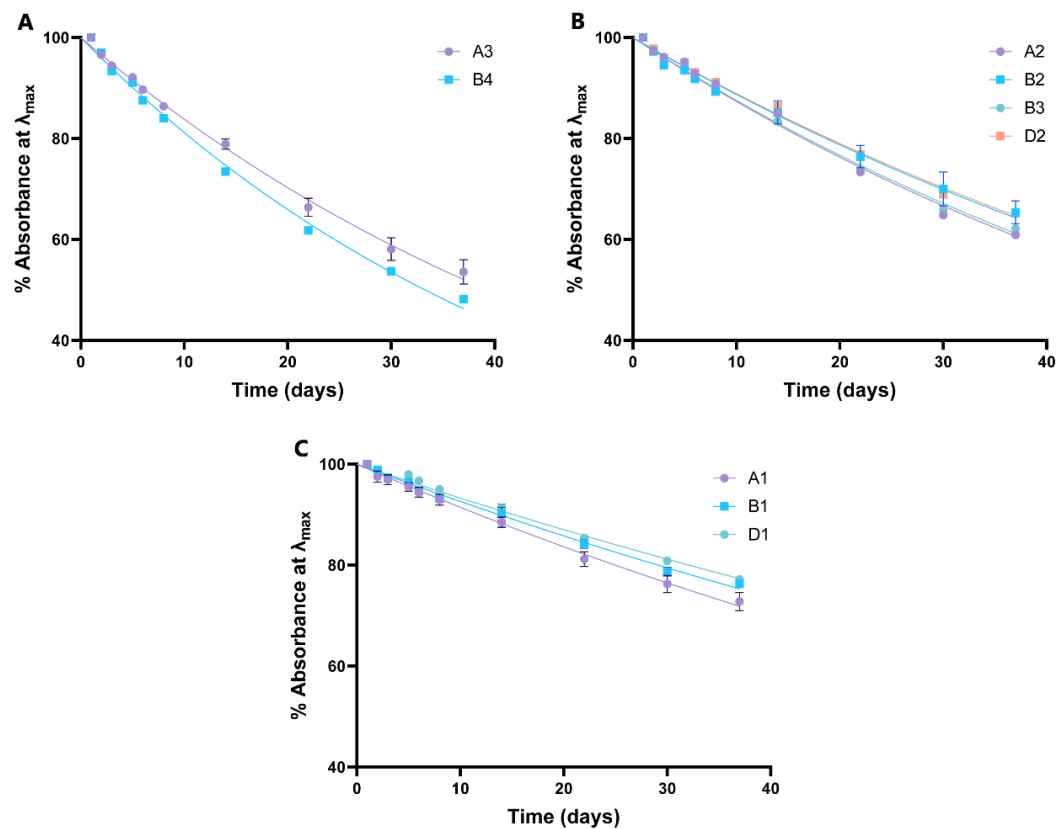
Figure 3.7 Stability of ternatin isolates (60 µM) over time at pH 7 and temperature of 23 °C. Percentage of residual coloured species were calculated as $\frac{A}{A_0} \times 100$, where A_0 was the initial absorbance value, and curves were fitted using the one phase exponential decay equation. Error bars represent standard deviation, and each data point represents the mean of three independent experiments (n=3). (**A**) ternatins decorated with two acyl residues; (**B**) ternatins decorated with three acyl residues; (**C**) ternatins decorated with four acyl residues.

Ternatin B4, the least decorated anthocyanin tested, had an absorbance corresponding to 48.2 % of the initial value (**Figure 3.7, A**), equivalent to a half-life of 31.8 days (**Table 3.3**). Despite being the least stable of the ternatins examined, the stability of this ternatin was comparable to that of the highly stable anthocyanins present in red cabbage.

Of the most decorated ternatins, D1 showed the highest stability, with 77.2 % remaining after 37 days (**Figure 3.7, C**) and with a calculated half-life of 99.7 days at room temperature (**Table 3.3**). Glycosylation of the 3' sidechain of ternatin D1, produced ternatin B1, which showed a slight decrease in stability of the anthocyanin, with 76.3 % remaining after 37 days and a decrease in the calculated half-life of 9.3 days.

Further glycosylation of ternatin B1 to A1, the most decorated anthocyanin, led to a larger decrease in stability with 72.8 % of the original absorbance remaining after 37 days (**Figure 3.7, C**), and a corresponding decrease in half-life of 12.8 days.

Without exception, the half-life of the anthocyanin increased with increasing number of acyl decorations. The largest change was upon acylation of B3 to D1, with a corresponding increase of 47.5 days (**Figure 3.8**). Similarly, acylation of A2 which produced B1, led to an increase in half-life of 39.1 days. Acylation of A3 to B3, and B4 to D2 led to increases in half-life of 12.9 and 25.6 days respectively.

Whilst acylation increased anthocyanin stability, in almost all cases glycosylation resulted in an overall decrease in half-life (**Figure 3.8**). The only exception was the glycosylation of ternatin B4 to A3, which lead to an increase in half-life of 6 days. Because the main route of degradation involves interaction with the carbons present on the anthocyanin C − ring, sterically hindering access to these positions may improve stability. Addition of glucose units potentially increased the interactions with the aqueous solvent, thus reducing the net forces acting to keep the stacked configuration closed. As the glycosylation of B4 to A3 is added in the third position of the 5' side chain, the distances between the glucose and the chromophore or the opposing sidechain may be short enough to facilitate interactions. This would counteract the increased hydrophilic forces acting to pull the sidechain away from the chromophore.

117

Table 3.3 Calculated half-life of ternatin isolates (60 µM) over time at pH 7 at 23 °C. Half-life was calculated as: $\ln(\frac{2}{K})$, where K is the rate of decay. Diagrammatic representation of each ternatin structure is given in the leftmost column, where: T = Delphinidin-3-O-glucoside(6″-malonate), G = glucose and C = p-coumarate.

	Half-life (days)	Structure
A1	77.6	T〈 GCGCG / GCGCG
B1	90.4	T〈 GCGCG / GCGC
D1	99.7	T〈 GCGC / GCGC
A2	51.3	T〈 GCGCG / GCG
B3	52.2	T〈 GCGC / GCG
A3	39.3	T〈 GCG / GCG
B2	58.0	T〈 GCGCG / GC
D2	58.9	T〈 GCGC / GC
B4	33.3	T〈 GCG / GC

Figure 3.8 Schematic showing the relationship between structural changes, half-life and λ_{max} of ternatin isolates at pH 7. T = Delphinidin-3-O-glucoside(6"-malonate), G = glucose and C = p-coumarate. Red arrows indicate glycosylation, while blue arrows indicate acylation.

Ternatins B3 and B2 are regioisomers, differing in which of the sidechains the terminal glucose residue is attached. In the case of ternatin B3, the terminal glucose resides on the 5' sidechain, whereas the terminal glucose of B2 is found on the 3' sidechain (**Figure 3.9**). Despite having the same total number of glucose and acyl units, the half-life of ternatin B2 was greater than that of B3 by 5.8 days (**Figure 3.8**). Both ternatins featured two acyl residues on the 3' sidechain, and only one on the 5' sidechain.

By adding the glucose to the 5' sidechain, the interactive forces may have been weakened asymmetrically; strengthening the interaction with the 3' sidechain but weakening the 5' sidechain interaction. In this case, too large a force acting on the 3' sidechain may have increased the B – ring dihedral, thus increasing the distance between the chromophore and the 5' sidechain. However, in the case of ternatin B2 the 3' sidechain stacking interactions are stabilised by association between the chromophore and two acyl residues.

Adding the glucose to the terminal position of the 3' sidechain must overcome the interaction forces between both acyl residues and chromophore. Therefore, upon glycosylation of the 3' sidechain the hydrophilic forces acting to cause destabilisation of the stacking interaction are reduced, contributing to the increased stability of B2 compared to B3.
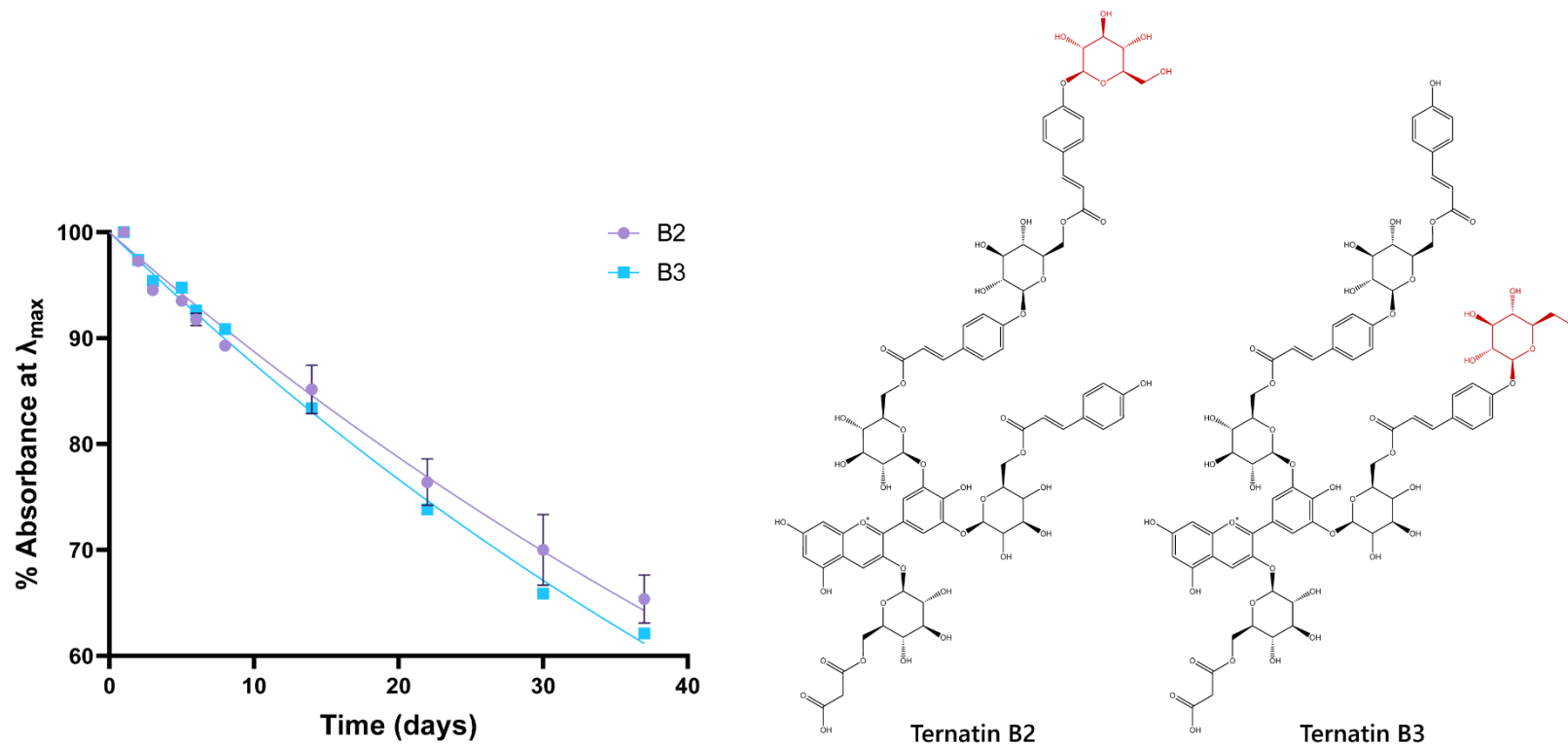
Figure 3.9 Decay of the UV-VIS absorption spectra of isolates (60 µM) ternatin B2 compared to B3 at pH 7. Within ternatin B2, the terminal glucose residue is present on the 3' sidechain, while the terminal glucose residue of B3 is present on the 5' sidechain. This glucose residue is highlighted red in the structural formula.

3.3.4 Thermal stability of ternatins

As plant foods are thermally processed, either as part of manufacture or prior to use in order to inhibit microbial growth, heating will impact anthocyanin content, and thus their ability to function as colourants. Curves of colour loss were generated by sampling the UV-VIS absorbance of samples stored at 55 °C each hour for a period of 8 hours, followed by two additional sampling points at 24 and 48 hours.

Ternatins obtain their stability due to the formation of sandwich-type stacking of the *p*-coumarate residues over the chromophore. However, at elevated temperatures energy provided to the system is likely capable of overcoming the interactive forces which stabilise the complex. This decreased the probability of the anthocyanin adopting a closed conformation, and therefore increased the likelihood for interaction between nucleophiles and the anthocyanin ring system.

Unlike at room temperature, the A series ternatins (those with two terminal glucose units) were the most resistant to degradation at elevated temperatures, with 88 % of the residual absorbance remaining after 48 hours for ternatin A1 and A3, and 87 % for ternatin A2 (**Figure 3.10**). The stability of the B series ternatins, with one *p*-coumarate and one glucose decoration terminating the sidechains, increased upon decoration. The lowest thermal stability observed for all ternatins was ternatin B4, with 46 % remaining after 48 hours (**Figure 3.10, C**). Stability of the B series increased with increasing decoration, from 56 % of the original absorbance remaining for B3, 65 % for B2 and 85 % for B1.

Unexpectedly, the D series ternatins which were observed to be the most stable at room temperature (**Figure 3.7**), were less stable compared to the A series ternatins After 48 hours at 55 °C, only 62 % and 66 % of the absorbance remained for D1 and D2 respectively (**Figure 3.10**).

The aromatic rings of the *p*-coumarate decorations themselves are electrophilic, and likely subject to nucleophilic addition. In this case, it may be possible that such reactions occurred leading to the degradation of the acyl residues themselves, and thus prevented copigmentation interactions, increasing accessibility of the anthocyanin ring system to further degradation by nucleophilic addition.
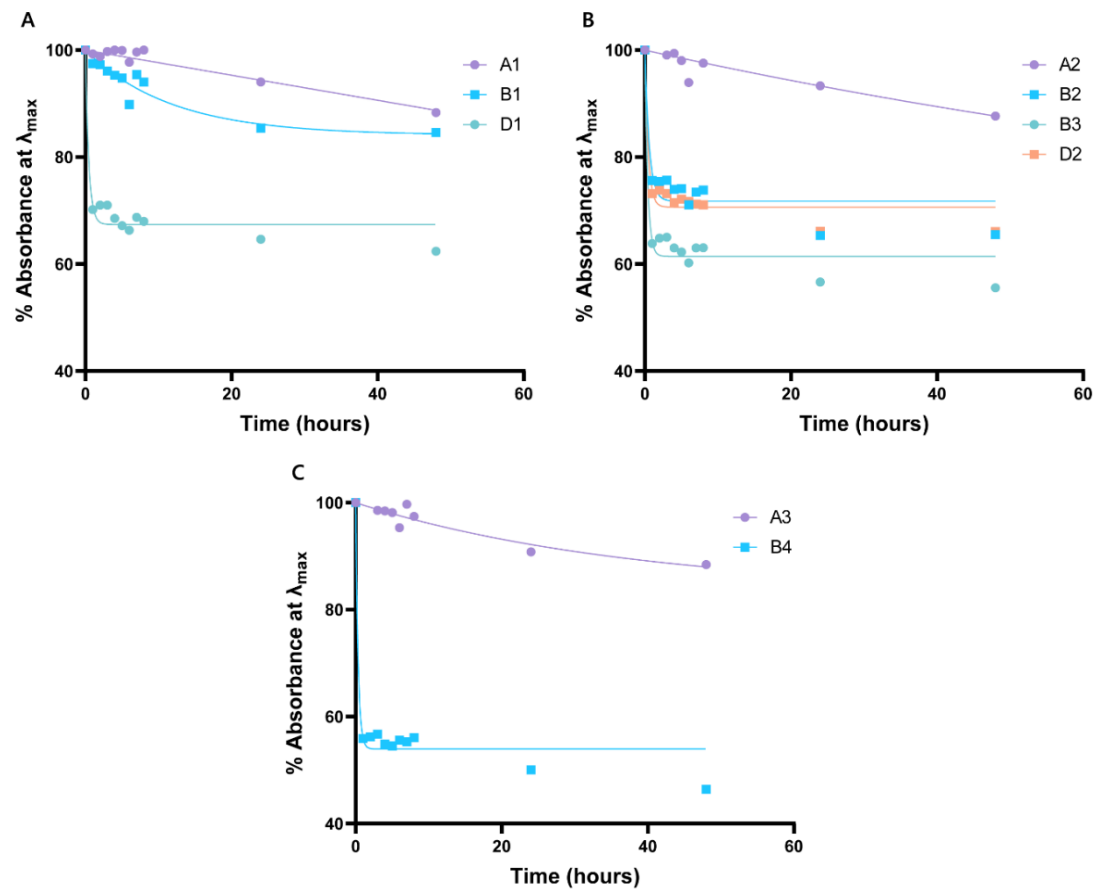
Figure 3.11 Stability of ternatin isolates (60 µM) over time at 55°C, at pH 7. (**A**) ternatins decorated with two acyl residues; (**B**) ternatins decorated with three acyl residues; (**C**) ternatins decorated with four acyl residues. Ternatins were incubated at 55°C in and oven, and each hour for the first seven hours, a 1 mL aliquot was taken. Two additional samples were taken at 24 and 48 hours. Aliquots were immediately cooled on ice and UV-VIS spectra recorded. Curves were fit using a one-phase decay function.

3.3.5 Measuring CIE colour space

In order to effectively compare the colour characteristics of the ternatin isolates to synthetic blue pigments, it was necessary to convert the UV-VIS absorbance spectra into CIE L*a*b* colour space coordinates. Typically, this would be carried out using a colorimeter. However, as these instruments only have capacity for single samples, measuring a large number of samples would be cumbersome. Therefore, a post-processing application was created to calculate automatically the CIE colour space coordinates from UV-VIS absorption spectra, measured using a plate reader. This enabled rapid acquisition of L*a*b* colour space measurements using a high throughput acquisition technique.

3.3.5.1 Verifying colour matching application

To verify the accuracy of the colour matching application, the CIE-L*a*b* values of crude butterfly pea extract at pH 1, 5 and 8 were measured using a colorimeter, and compared to the values calculated from UV-VIS absorbance spectra acquired using a spectrophotometer (**Figure 3.11**).

The values measured using the colorimeter showed close alignment to those calculated using the colour-match application, with colour differences (ΔE) of 2.73, 3.08 and 3.24 for pH 1, 5 and 8 respectively. These differences in colour fell slightly above the "just noticeable difference" cut-off value of 2.3, which is the minimum value difference required for a colour difference to be perceptible.

However, when these CIELAB values are converted into RGB values for display and compared against the original samples (**Figure 3.11**), the colorimeter measurements appeared to underestimate the saturation of the colours compared to the python script (as outlined in **Section 3.2.6**) values. The script values appear slightly darker, and more closely resemble the colour of the solutions in the cuvettes, although the difference is marginal.
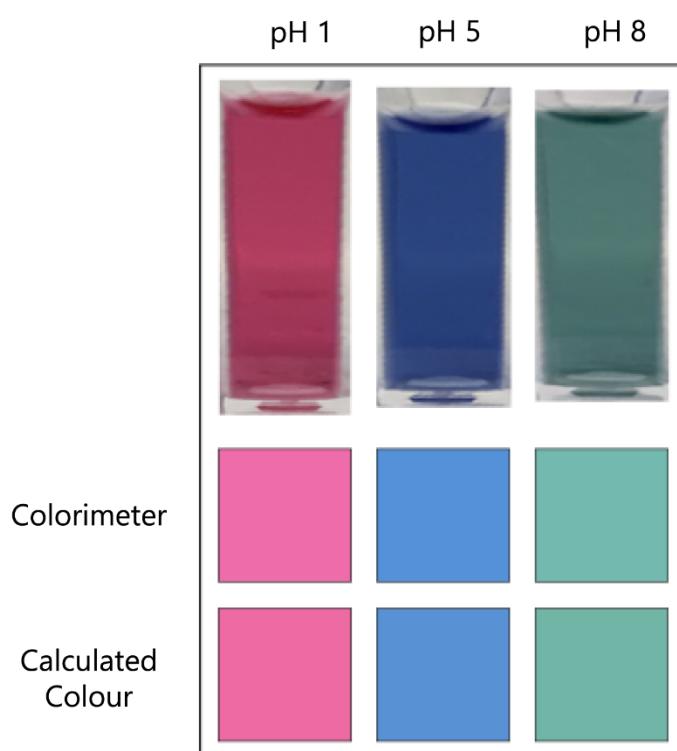
	Colorimeter			Calculated Colour			
	L*	a*	b*	L*	a*	b*	ΔE
pH 1	63.1 ± 0.5	56.0 ± 0.7	-8.0 ± 0.2	62.8 ± 0.4	56.6 ± 0.8	-5.4 ± 0.1	2.73
pH 5	59.0 ± 0.3	1.6 ± 0.2	-42.4 ± 0.3	59.4 ± 0.3	0.5 ± 0.1	-39.6 ± 0.	3.08
pH 8	70.5 ± 0.5	-24.4 ± 0.3	-1.8 ± 0.4	69.0 ± 0.6	-25.2 ± 0.4	0.9 ± 0.4	3.24

Figure 3.12 CIE-L*a*b* values of three solutions of butterfly pea crude extract (1 mg mL^{-1}) in citric acid : sodium phosphate buffer at pH 1, 5 and 8. L*a*b* values ± standard error (n = 5) were recorded using an HunterLab UltraScan VIS (Colorimeter), and calculated using the python application from UV-VIS absorption spectra recorded using an Eppendorf BioSpectrometer. Colour difference (ΔE) was calculated as $\Delta E = \sqrt{(L_2^* - L_1^*)^2 + (a_2^* - a_1^*)^2 + (b_2^* - b_1^*)^2}$. Samples were photographed in natural daylight. The coloured squares are visual representations of the measured and calculated L*a*b* values, converted to sRGB values for display.

3.4 Discussion

I have used the individual ternatins from the flowers of butterfly pea to investigate the impact of sequential glycosylation and acylation on the colour and stability of the anthocyanins working through intramolecular co-pigmentation. My empirical data show that glycosylation generally increases the λ_{max} of the anthocyanin slightly (involving bathochromic shifts of between 1 and 2 nm), but generally also leads to a decreased stability, as evidenced by systematic decreases in half-life.

The only example where glycosylation increased stability of the anthocyanin was ternatin B4 to A3. The delocalised π – orbitals above the anthocyanin rings are negatively charged, while the H – atoms of the sugar hydroxyl groups are positively charged due to the inductive effect of the oxygen atom. As the added glucose residue is in the third position on the sidechain, the distance between the glucose and chromophore may have been short enough to allow for π – H interactions to form, which would have increased the forces acting to produce stacked configurations. At greater distances however, glycosylation acts to decrease the interaction force between the sidechain and delphinidin rings. This is likely due to increased hydrophilic interaction with the aqueous solvent, which acts to pull the sidechain away from the anthocyanin and therefore decreases stability.

The small bathochromic shift as a result of glycosylation is therefore unlikely to be the result of the electronic effects of stacking interactions directly. It has been indicated that B – C – ring dihedral angle has a larger effect on the λ_{max} than the direct interaction between sidechain and anthocyanin (Denish *et al.*, 2021). This is particularly evident in stability and λ_{max} of the regioisomers ternatin B2 and B3 (**Figure 3.9**). Ternatin B2 has a longer half-life of 5.8 days days compared to B3, while ternatin B3 has a slight bathochromic shift of 3 nm. The glycosylation of B2 occurs on the 3' sidechain, which is decorated with two acyl residues. Therefore, the increased hydrophilic forces induced by the additional glucose are smaller, relative to the attractive forces of the acyl residues, and so the impact on sidechain interaction with the anthocyanin are minimal. Conversely, Decoration of the 5' sidechain, as in the case of ternatin B3, only has to overcome the attractive forces of one acyl residue, and so glycosylation has a larger destabilising effect. As the

126

interactive forces are asymmetrical, with two acyl residues on the 3' sidechain and only a single residue on the 5' sidechain, further weakening of the 5' sidechain attractive forces to induce additional distortion on the B – ring dihedral angle.

3.4.1 Charge transfer states enhance blue colour of anthocyanins

The empirical data presented in figure 3.4 (and **Appendix 3.1 – 3.8**)show a distinctive UV-VIS absorption peak for ternatins present at pH 3 and above. At this pH the equilibrium favours the flavylium cation, with a minor presence of the neutral species. In the case of ternatin A1, this absorption peak is present at a wavelength of 624 nm, shifted much further towards blue than the flavylium cation peak at 554 nm and the neutral species at 575 nm.

This absorption peak is therefore likely the result of a charge transfer interaction, involving the rearrangement of electron density within the molecule to a lower energy state. Upon absorption of incident light, the electrons have a lower energy pathway between the highest occupied molecular orbital and the new energy state, thus decreasing the energy gap and hence the energy of the light required for excitation.

Chapter 4

Using molecular dynamics as a tool to understand the conformational landscape of anthocyanins

Chapter 4: Using molecular dynamics as a tool to understand the conformational landscape of anthocyanins

4.1 Introduction

The interactions that anthocyanins undertake with planar copigment molecules has been extensively studied using a combination of spectroscopic techniques. Since copigmentation induces often large changes in the visible spectrum, UV-VIS spectroscopy is most commonly used to characterise copigmentation (Trouillas *et al.*, 2016). This provides information on the tendency of copigmentation interactions to occur, and the magnitude of these interactions, but cannot provide accurate information on the atomic geometry of the complex by which the changes in visible spectra occur.

Two experimental techniques exist which are used to infer the structural conformation of anthocyanin complexes: circular dichroism (CD) and NMR spectroscopy. As anthocyanin sugar residues are chiral themselves, self-association can result in chiral stacking of complexes which are particularly sensitive to CD spectroscopy (Mendoza *et al.*, 2018).

NMR is generally considered the gold standard of spectroscopic techniques for characterising anthocyanin : copigment complexes. Using the Nuclear Overhauser Effect (NOE), the interactions through space between protons of the anthocyanin and copigment can show how the interactions impact colour and stability (Mateus *et al.*, 2002; Fernandes *et al.*, 2015). However, NMR is incapable of distinguishing these interactions beyond "strong" (~ 4 Å) and "weak" (> 4 Å). In addition, due to the anthocyanin chromophore lacking vicinal protons on opposing sides of the B-ring, NMR is incapable of characterising B- ring rotation.

For the desired atomic scale resolution, X-ray crystallography is the only experimental technique available. However, pigment : copigment complexes are labial, with only metalloanthocyanin having been successfully crystallised (Shiono,

Matsugaki and Takeda, 2005), making X-ray crystallography unfeasible. A more recent computational approach using molecular dynamics simulations and quantum mechanical calculations are being applied to anthocyanins, allowing the study of copigmentation and anthocyanin : macromolecule interaction in atomic scale resolution.

4.1.1 Molecular Dynamics

Molecular dynamics simulations involve calculating the forces acting upon each particle within a system and updating their positions and velocities in small time steps. This relies on two fundamental principles, firstly that the continuous state of a system can be solved by integrating Newton's equation of motion (Xing, Chen and Zhang, 2023), and secondly the Born-Oppenheimer approximation which states that as atomic nuclei are much heavier than electrons, only the position and velocity of atoms are required to describe the state of a molecular system (Singh *et al.*, 2018). These simulations can capture a wide range of molecular interactions of macromolecules, revealing the position and trajectory of all atoms at femtosecond resolution (Hollingsworth and Dror, 2018). This approach has been used previously to study interactions between anthocyanins with benzoic acid and hydroxycinnamic acid copigments (Qian *et al.*, 2017), anthocyanin : polysaccharide polymer interaction (Fernandes *et al.*, 2014), as well as protein : anthocyanin binding (Xie *et al.*, 2020).

Molecular dynamics simulations can be broken up into distinctive stages. The first being the application of parameters which describe how forces act on the molecule. These force fields are applied to each atom in the system, describing the interatomic potentials (González, 2011). Secondly, the molecular system is solvated and equilibrated to a given temperature (or velocity) for each particle. This is achieved through statistical mechanics, or thermodynamic ensembles, which dictate how velocity (temperature) and particle density (pressure) are scaled. Finally, simulations are run with the position, velocity and energy of each atom recorded at each timestep. Data can then be extracted and analysed to yield information of the conformation of the system, as well as energy of interactions (Xie *et al.*, 2020).

4.1.2 Replica-exchange Molecular Dynamics

Sampling with classical molecular dynamics is limited to specific low energy regions of the conformational space. This can be particularly problematic when multiple energy minima exist within a molecular conformation, separated by free-energy barriers often larger than the available thermal energy of the system (Huber, Torda and van Gunsteren, 1994). The result of these local energy minima, termed saddle points, cause the molecular trajectory to become "stuck" in a local minimum, preventing sampling the full conformational space.

Multiple methods to sample the wider conformational space have been developed, with the most widely used method being parallel tempering or replica exchange molecular dynamics (Hansmann, 1997). In this method, several replicas of the system are simulated independently, in parallel and under different conditions. At a predefined interval, neighbouring pairs of replicas are exchanged with a specific transition probability (Qi *et al.*, 2018). In most parallel tempering experiments, temperature is used as the parameter that changes among the replicas, allowing replicas to escape the local minima by exchanging coordinates with simulations run at a higher temperature.

However, the number of replicas required to span a preassigned temperature range grows with the size of the system, requiring more computational power to run the simulation. In addition, the increase in temperature as replicas are exchanged can interfere with calculation of the free energy landscape. In contrast, Hamiltonian replica exchange works by scaling the force fields between replicas, such that replicas at "higher temperature" interact less strongly (Bussi, 2014). This has the benefit of allowing escape from local energy minima, without the addition of energy to the system. This technique has only recently been applied to the computation study of anthocyanin spectral properties (Rusishvili *et al.*, 2019) and structural conformation (Denish *et al.*, 2021).

4.1.3 Aims and objectives

This chapter presents simulations from molecular dynamic modelling of intramolecular copigmentation of the anthocyanins from butterfly pea (*Clitoria ternatea*) There were two goals for this chapter; the first was to determine the structural characteristics of the ternatins, specifically the intermolecular distance between the acyl decorations and the chromophore, their propensity to form these stacking type interactions and the average dihedral angle of the chromophore B – ring. The second aim was to determine the effects of B – ring rotation has on transition energy between frontier molecular orbitals, and the resulting absorbance and colour characteristics of the anthocyanins.

Determining the effects that iterative decoration has on the macromolecular geometry may help to clarify the mechanism by which more heavily decorated ternatins are stabilised more than their less decorated precursors. In addition, understanding how structural characteristics of the chromophore influence the light absorption of the molecule, could reveal how they alter the colour of ternatins.

4.2 Materials and methods

4.2.1 Generating molecular models

For both the cation and neutral species, the ternatin molecular structures were initially drawn in ChemDraw (v22.0) and opened into the 3D modelling software Avogadro (v 1.2) (Hanwell *et al.*, 2012). Each of the glucose and *p*-coumaric acid units were carefully assessed to ensure the correct stereochemistry was maintained (β – glucose and trans(Z)-*p*-coumaric acid) and resonance structure of the delphinidin chromophore was corrected. The molecular geometry was partially optimised with the energy minimisation function plugin in Avogadro, using the steepest descent with the universal force field (UFF). Optimisation was run until the energy change was below 0.01 kJ mol^{-1} between steps. These unfolded structures were then saved as .mol2 files.

The models were then manipulated around the 3'/5' hydroxyl – sugar linkages and 6" – 1 sugar – organic acid ester linkages until the organic acid moieties were positioned in an offset-parallel (stacked) arrangement with the anthocyanin chromophore. These folded structures were then also saved in .mol2 format.

4.2.2 Hamiltonian replica exchange molecular dynamics

4.2.2.1 Assignment of general atomic force field (GAFF)

In order to run molecular dynamics simulations which were representative of natural systems, the forces which act on the atoms within that system needed to be described, thoroughly. To accomplish this, first the interactions between atoms needed to be characterised with a force field, which included terms that described the bond strength, angles and torsions between atoms. The second requirement was the calculation of a series of partial charges for each atom, which described the Coulombic interactions within the molecular system. To fulfil this objective, the general atomic force field (GAFF) was assigned for each bond type, and partial charges for each atom in the molecule were calculated with the Amber Antechamber package, using the antechamber python parser interface (Acpype, v 2022.7.21) (Sousa Da Silva and Vranken, 2012).

For all molecules simulated with molecular dynamics, force fields were first assigned for all bond types, then the partial charges were calculated using the AM1-BCC method. Using this method, electron distribution was first calculated by the semi-empirical quantum chemical method AM1, and then bond charge corrections were applied (BCC), yielding charges comparable in quality to those calculated by HF/6-31G* RESP (Jakalian *et al.*, 2000). Parameters for the command-line input for the Acpype force field calculation are given below:

```
acpype -i input.mol2 -o gmx -n [charge] -a gaff -c bcc -s
[time (s)]
```

The input .mol2 file for which the calculation was performed was specified by the `-i` flag. The output format was specified by the `-o` flag, and automatically formatted the output into a topology file (.top), a coordinate file (.gro), a position restraint file and the input molecule parameters (.itp). The `-a` and `-c` flags specified the force field and charge calculation method respectively (in this case GAFF and AM1-BCC). Finally, the net charge of the molecule was given by flag `-n`, and the time limit for the calculation by `-s`.

4.2.2.2 Modifying Acpype output files for simulations

Although Acpype outputs Gromacs compatible files, the structure of the topology file is compressed with the parameters for the molecule specified in a separate file. In order to run the simulations, all sections of the input molecule parameters (.itp) file had to be concatenated into the topology file (.top). In addition, the force field and partial charge descriptions for both the solvent and any potential counter ions (which were required for charged systems) had to be added to the topology. In the case of the ternatin cations, this required both the inclusion of water and chloride ions.

To process the large volume of output files and add in the required force field elements, a python script was created called Topology_Processor.py (**Appendix 4.1**), which took the following command-line input:

```
python3 ./Topology_Processor.py -i input.itp -t input.top -w -
s cl
```

In this example, the input molecular properties (`-i`) were transferred to the topology file (`-t`) in the correct location. By providing the `-w` flag, the parameters for the transferable intermolecular potential with 3 points (TIP3P) water model were included in the relevant sections within the topology file. Last, the -s flag denoted which of the counter ions to include, from a selection of sodium (na), potassium (k), chloride (cl) or bromide (br). Parameters for the TIP3P water model were taken from (Mark and Nilsson, 2001), and counter ions from (Loche *et al.*, 2021).

4.2.2.3 Equilibration of simulation system

Prior to simulations, the molecular model simulation cell had to be established, solvent and counter ions added, and the system had to be equilibrated thoroughly. The Gromacs package (Gromacs, v2022.6.1) patched with Plumed (Plumed, v2.7) was used to carry out the following steps.

4.2.2.3.1 Simulation cell, solvation and counter ions

For each system a cubic cell was generated centred on the molecule with a minimum distance of 1 nm from the boundary cell wall. Molecules were then fully solvated in TiP3P water with between 2,150 and 9,677 water molecules for the smallest (ternatin C5) and largest (ternatin A1) systems respectively.

For molecules with a net charge not equal to zero, solvent molecules were replaced with a counter ion bearing the opposite charge. In the case of ternatin cations, one solvent molecule was replaced by a chloride ion bringing the overall system charge to neutral.

To ensure appropriate geometry and prevention of steric clashes, each of the solvated, electroneutral structures were then relaxed through a short energy minimisation simulation using the steepest decent algorithm. The energy minimisation was run for a maximum of 500 steps, with coordinates written every 10 steps.

4.2.2.3.2 Temperature equilibration (NVT)

To bring the system to the required experimental temperature of 300 K (26.85 °C), an ensemble with a constant number of particles, volume, and temperature (NVT) was generated by simulating the system for a short period, while applying harmonic forces to the solute to restrict its movement. In effect, only the solvent water molecules were simulated.

For each of the ternatins, a simulation of 100 ps with a timestep of 2 fs was run with coordinates output every 500 steps. To ensure equilibration to 300 K, velocity rescaling was active, which was a modified version of the Berendsen thermostat as implemented in Gromacs. Post simulation, the temperature progression was plotted to confirm each system reached thermal equilibrium.

4.2.2.3.3 Pressure equilibration (NPT)

A final equilibration using a constant pressure and temperature (NPT) ensemble was run as a continuation of the NVT ensemble by providing the checkpoint file (.cpt) from the previous simulation. Simulations were again run for 100 ps with a 2 fs timestep and coordinates output every 500 steps.

Pressure coupling was controlled using the Berendsen barostat, with a reference pressure of 1 bar. Simulation cell pressure was controlled by isotropic scaling of the box vectors. Similar to the temperature equilibration, the density progression was plotted as a 10 ps running average to confirm the system had equilibrated at the required density.

4.2.2.4 Partial Tempering of molecular system

4.2.2.4.1 Processing the topologies

For Hamiltonian replica exchange molecular dynamics (HREX-MD) using Gromacs patched with Plumed, atoms for which the force field scaling should be applied, must be specified. For Plumed to recognise this selection, in each of the topology files an underscore had to be appended to the end of each atom type in the selection (e.g. ca → ca_).

To process the large number of topologies, a second python script was created called Plumed_HREX_TopoProc.py (**Appendix 4.2**), which parsed the topology file, identified all molecules present and allowed the user to select where to apply the Plumed processing flag. In the case of the ternatins simulated, only the atoms with connectivity to the anthocyanin were selected for force field scaling.

4.2.2.4.2 Scaling the force fields

After processing the topology, a geometric progression was assembled between the two reference temperatures: 300 K and 1100 K. For ten replicates per system, a temperature gradient was established using the following equation:

$$f(t) = \prod_{i=0}^{n-1} t_{min} \times e^{i \times \left(\frac{\ln\left(\frac{t_{max}}{t_{min}}\right)}{(n-1)} \right)} \tag{4.1}$$

where t_{min} and t_{max} are the minimum and maximum reference temperatures respectively, while *n* is the number of replicates. This was then converted into a series of values by which to scale the force field parameters by dividing the minimum reference temperature by each calculated temperature in the gradient.

Using both the output from Plumed_HREX_TopoProc.py and the generated lambda value list, a series of topology files was generated with scaled force fields using the Plumed partial_tempering command. Each of these were then converted into Gromacs portal binary run input files.

4.2.2.5 Running the simulations

For both the cation and neutral species of the folded and unfolded geometry, using the processed NPT ensemble generated in the previous sections, ten replicas were simulated in parallel. Simulations were run on a high-performance computing cluster, with calculations shared between 2 GPUs and 40 processor cores.

The simulation was run for 25 ns per replica with a 2 fs timestep and coordinate
exchanges were attempted every 1000 timesteps, with an average acceptance rate
of 20% between adjacent replicas, for a combined total of 250 ns. After
demultiplexing the output trajectories, only the replicate trajectory at 300 K was
retained for analysis.

From the trajectory, the free energy for key interatomic distances between the acyl
decorations and each of the chromophore rings, torsion angle and inter-acyl distance
were extracted using the Plumed driver tool.

4.2.3 Classical molecular dynamics simulations

Having screened each ternatin for the lowest energy conformer using Hamiltonian
replica exchange, a molecular model of that structure was prepared for classical
molecular dynamics simulation using the method outlined in section 4.2.2. Ten
replicas for each ternatin were simulated for 50 ns with a 2 fs timestep.

To quantify the relative ratio of open and folded configurations, the minimum
distance between atoms of the chromophore and those of the acyl decorations were
tracked. For ternatins bearing three or four *p*-coumarate molecules, the inter-acyl
distances were also tracked.

To determine the relationship between acyl decorations and steric hindrance of
water molecules at the chromophore C2 position, the radial distribution function
(RDF) was used to measure the probabilistic density between the oxygen atom of
the water solvent, and the C2 carbon. Mean percentage differences between RDFs
were calculated using the following equation:

$$(4.2) \quad \frac{|RDF_1 - RDF_2|}{\left(\frac{RDF_1 + RDF_2}{2}\right)} \times 100$$

4.2.4 Examination of delphinidin-3-glucoside rotamers using quantum mechanical methods

To isolate the effects that rotation of the chromophore B-ring has on the transition energy gap and hence colour, quantum dynamics calculations were run on delphinidin-3-glucoside using the ORCA package (ORCA, v5.0.3) (Neese, 2012).

4.2.4.1 Generating an optimised structure

Using the method outlined in section 4.2.1, a molecular model of the cation delphinidin-3-glucoside was generated. The molecular model first underwent a geometry optimisation step in the gas phase using the functional B3LYP (Becke, 3-parameter, Lee–Yang–Parr) with basis set 6-311**G+. Convergence was reached when movement of any atom in the system resulted in an increase of the total system energy. After optimisation, the vibrational frequencies were calculated and analysed to ensure that all were positive, since negative frequencies indicate that the geometry converged to an energy saddle point. In cases where a negative frequency was present, the bond was displaced along the vibration vector and the optimisation was rerun on this displaced geometry. This was repeated until all vibrational frequencies were positive.

After optimisation in the gas phase, the geometry was then optimised with the same functional and basis set but including a conductor-like polarisable continuum model (CPCM), applying water as the solvent. Vibrational frequencies were then checked to ensure all frequencies were positive.

The optimised geometry was then loaded into Discovery studio (Dassault, Discovery studio visualizer 2021). Twelve additional rotamers were generated by setting the torsion angle of the C3, C2, C1' and C2' dihedral between 0° and 90° with a step size of 10°. This resulted in eleven rotamers, including the initial optimised configuration.

4.2.4.2 Identifying the optimum functional for absorbance calculations

Absorbance spectra of molecules were calculated using time-dependant density functional theory (TD-DFT). Taking the optimised geometry generated with the functional B3LYP and basis set 6-311**G+, ground-state hessians were generated using the following functionals: B3LYP, B3P86, M06, M062X and PBE0. Using the ground-state hessian, from the vibrational frequency calculation, the first five excited states were then calculated using an "adiabatic hessian after step" method, as implemented in the ORCA package.

Each of the calculated transition lines (derived from the oscillator strength and transition dipole moments) were then broadened with a Gaussian function. The calculated spectra were then compared to UV-VIS absorbance spectra collected from delphinidin-3-glucoside at a concentration of 60 μM in aqueous 0.1 % HCl.

4.2.4.3 Calculating absorbance spectra

After identifying the most appropriate functional to simulate absorbance spectra, the molecular geometry of each of the 12 rotamers was relaxed through a geometry optimisation until converged with B3LYP, and basis set 6-311**G+ with the torsion angle constrained at the given angle. Using these relaxed structures, the ground-state hessian was calculated and fed forward to the TD-DFT calculation method as above. The absorbance spectra were then convoluted by Gaussian broadening of the transition band spectra, and the colour calculated using the method described in section 3.2.6.

4.3 Results

4.3.1 Replica exchange molecular dynamics

4.3.1.1 Equilibration and potential energy gradient

Initially, molecular models of each of the ternatin cation and neutral species were simulated to determine the spatial conformation and to calculate the free energy of stacking type arrangements. Each structure first underwent a series of equilibration steps to ensure the model system was at the specified experimental conditions. Figure 4.1 shows the energy minimisation, and the temperature, pressure and density equilibration of ternatin A1. Each model was checked to ensure temperature was at 300 K and an average density of ≈ 1000 kg m^{-3}. After equilibration, a series of replicas were generated with scaled force field parameters. Figure 4.2 shows the potential energy overlaps between the replicas. For replica exchange, replicas must have sufficient overlap between these potential energy distributions, to ensure coordinates can be exchanged successfully. At the same time, to ensure that the conformational space had been explored thoroughly (and that the system had not been trapped within a local energy minimum), scaling of the interaction energies must have had a wide coverage. As can be seen in figure 4.2, by using a geometric progression, a larger proportion of the replicas have reduced interaction potential, without sacrificing energy overlap.
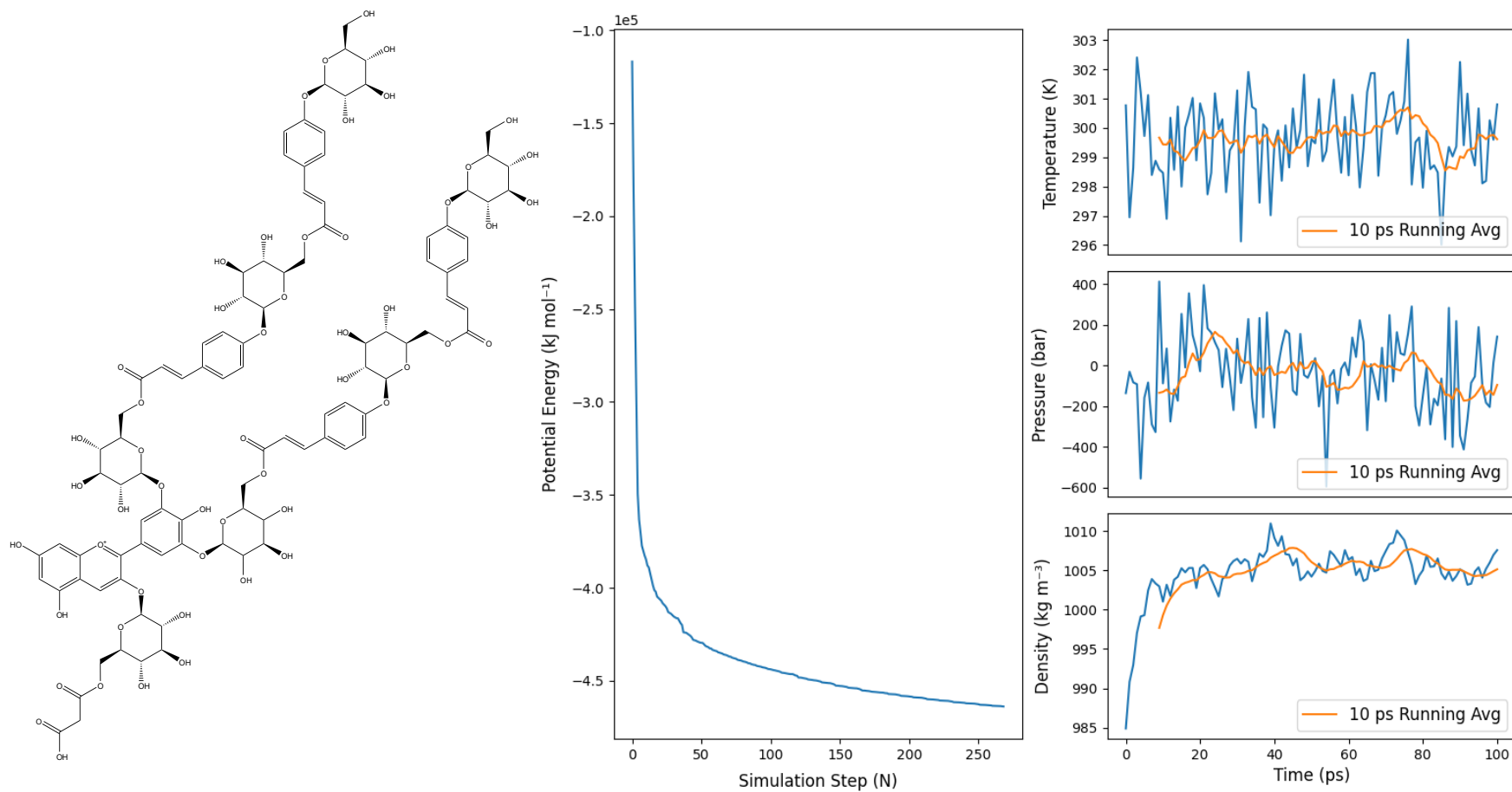
Figure 4.1 Energy minimisation, temperature, pressure and density during equilibration of ternatin A1 model. Left: Structural formula of ternatin A1. Right: Potential energy convergence of ternatin A1 model during geometry relaxation, temperature convergence (NVT) to 300 K, pressure and density (NPT) convergence to 0 bar

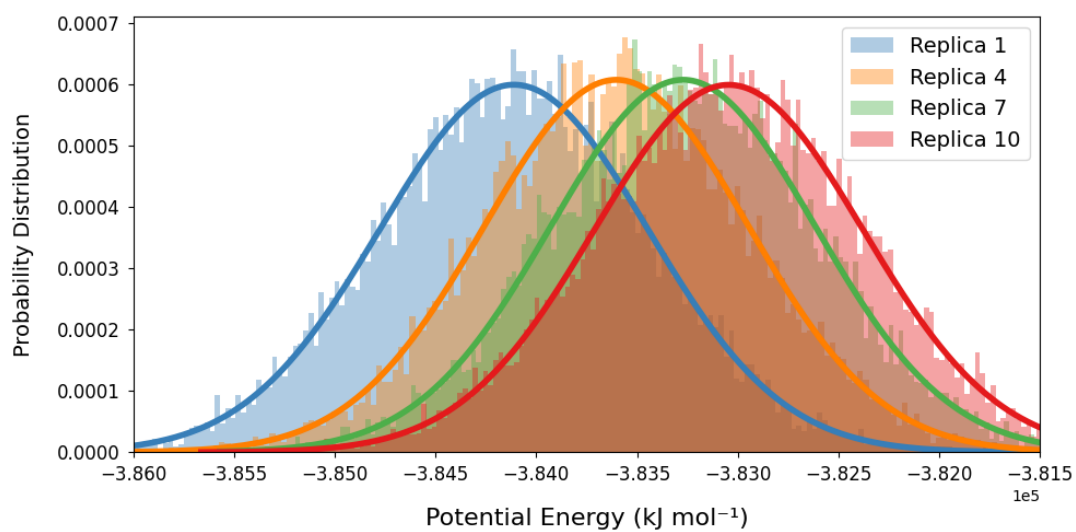**Figure 4.2** Geometric progression scaling of interaction energies between ten replicas of ternatin A1 simulated for 1 ns. Only the interaction energy of replicas 1, 4, 7 and 10 are shown for clarity. Overlap of potential energy allows for exchange of coordinates between replicas. Geometric progression ensures that more replicas are simulated with lower interaction energy to improve conformational space sampling.

4.3.1.2 Free energy of association between inner 3' and 5' acyl decorations

Replica exchange simulations with ten replicas were performed on each of the ternatin models, with a simulation period of 25 ns. The trajectories were then demultiplexed so that a contiguous simulation was achieved, with no interruptions in the trajectory, and the first trajectory simulated with full interaction potentials was used in this analysis. Using the Plumed plugin, free energy of association between atoms of the chromophore and of the acyl decorations were calculated. Figures 4.3 – 4.6 show the free energy distribution as a function of acyl decoration distance.

Ternatin D3 (**Figure 4.3**) shows two distinct energy minima; the first when both acyl residues are within 4 Å, indicating a strong stacking arrangement. The second energy minimum is more diffuse and spread over a larger range of distances, from 4.5 – 6 Å for the 3 ' acyl residue and 5- 6 Å. In addition, ternatin D3 has an energy pathway for dissociation of the 5' acyl residue as seen by the region of low energy leading away from the diffuse energy minimum. As the acyl residue linkages are flexible ester bonds, this lower energy pathway indicates the 5' acyl residue can drift away from the chromophore, likely by the formation of electrostatic interactions with the glucose residues present at the 3', 5' and 3 hydroxyl groups, displacing the acyl residue away from the chromophore.

Introducing asymmetry through glycosylation of ternatin D3 to produce ternatin B4 weakens the close association of the chromophore with the acyl decorations (**Figure 4.4**). For this intermediate the lowest free energy basin is diffused across a larger area, ranging from 5 – 6 Å for the 3' acyl residue, and 4 – 5.5 Å for the 5' acyl residue. The increased distance of the 3' acyl residue points towards weaker association between the decoration and the chromophore, where this effect may be due in part to an overall increase in hydrophilic interaction forces with the water solvent. The counteracting force may effectively reduce some of the electrostatic interaction which would otherwise pull the aromatic rings into closer association. Despite the energy minimum residing at a greater distance between the acyl residues and the chromophore, ternatin B4 has a more discrete energy basin with higher energy required to escape the stacked arrangement when compared to ternatin D3, as indicated by the region of relatively high energy surrounding the minima.

Ternatin A3 (**Figure 4.5**) has a strong energy minimum 3.5 – 4.5 Å for both the 3' and 5' acyl residues. Similar to ternatin D3 (although to a lesser extent), ternatin A3 has an energy pathway leading to dissociation of the 5' acyl residue, indicating that the 3' residue experiences higher attractive forces with the chromophore. Interestingly, this energy channel requires a decrease in the 3' acyl residue, hinting that the opposite acyl residues harbour some attractive forces for each other. As the sugar residues overhang the chromophore, electrostatic interactions between the terminal 3' and 5' sugars may provide some of this predicted attractive force.
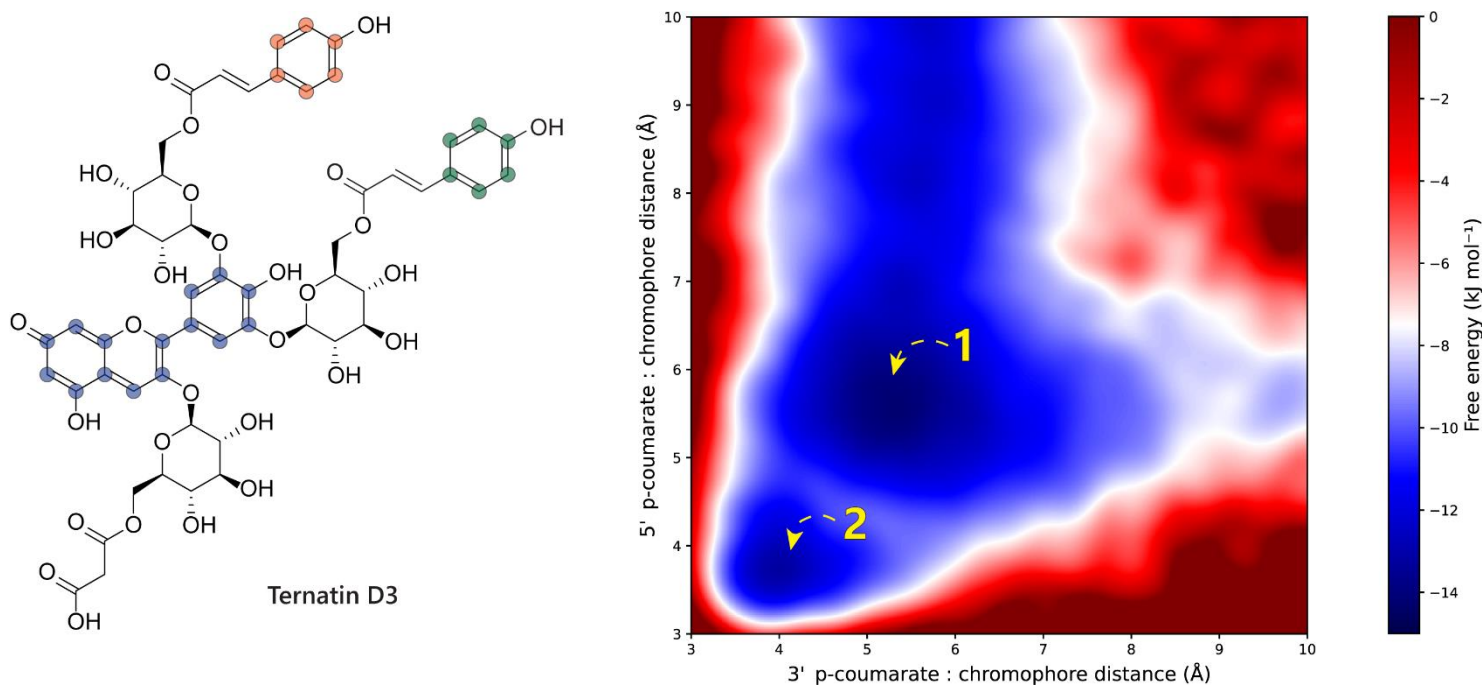
Figure 4.3 Free energy surface for distances acyl groups and ternatin D3 chromophore. Left: Structural formula of ternatin D3. Free energy of association was measured between the 3' acyl residue atoms (orange) and 5' acyl residue atoms (green) as a function of distance from the chromophore atoms (blue). Interactions between the chromophore and acyl residues are at the lowest energy when acyl residues are at a distance of between 5-6 Å (1). A second energy minima is observed when the 3' residue is at 4 Å and the 5' residue is at 3.8 Å from the chromophore (2).

Figure 4.4 Free energy surface for distances acyl groups and ternatin B4 chromophore. Left: Structural formula of ternatin B4. Free energy of association was measured between the 3' acyl residue atoms (orange) and 5' acyl residue atoms (green) as a function of distance from the chromophore atoms (blue). Interactions between the chromophore and acyl residues are at the lowest energy when the 3'acyl residue is at 5-6 Å and the 5' acyl residue is between 4.5-5.8 Å (1). A second faint energy minima is observed both the 3' and 5' residues are at a distance of ≈ 4 Å (2).
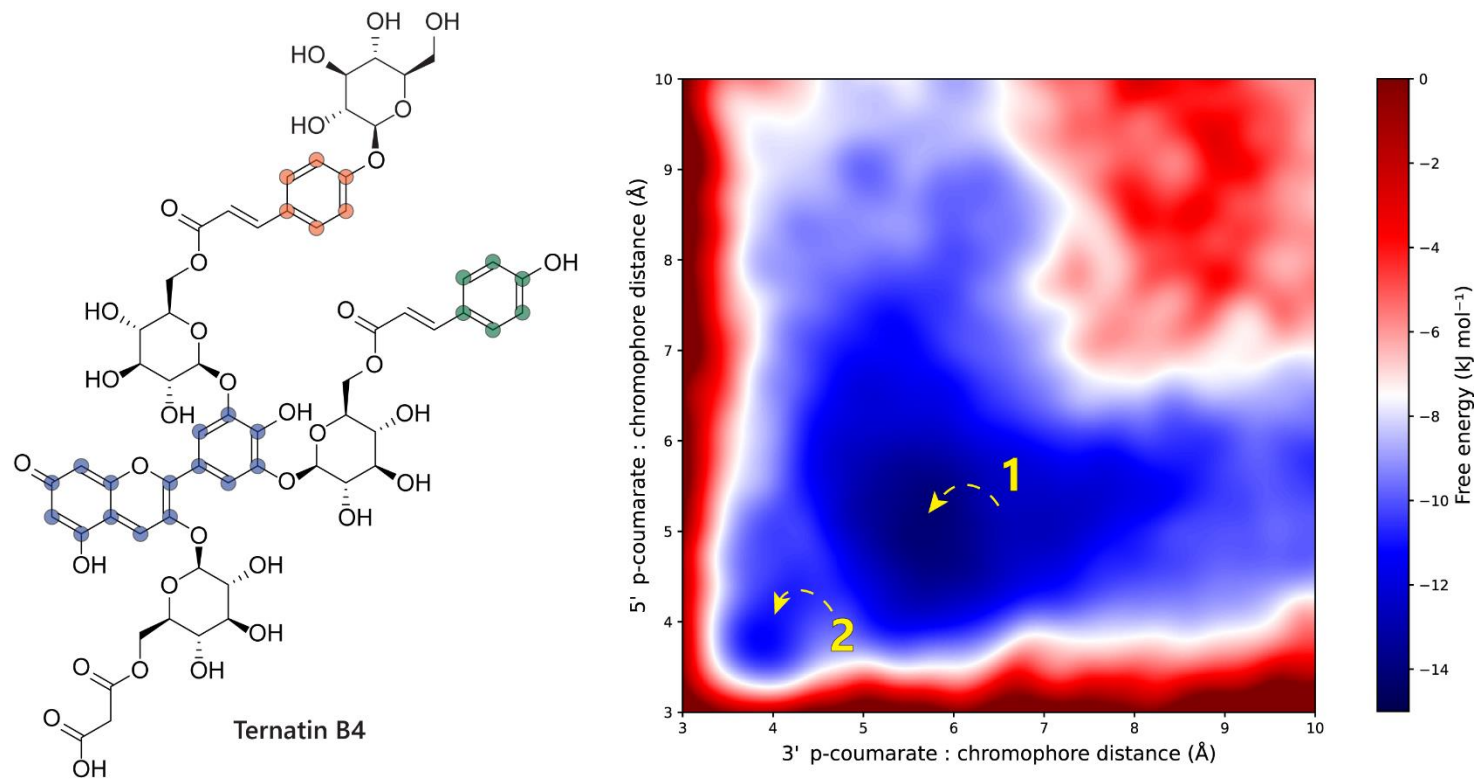
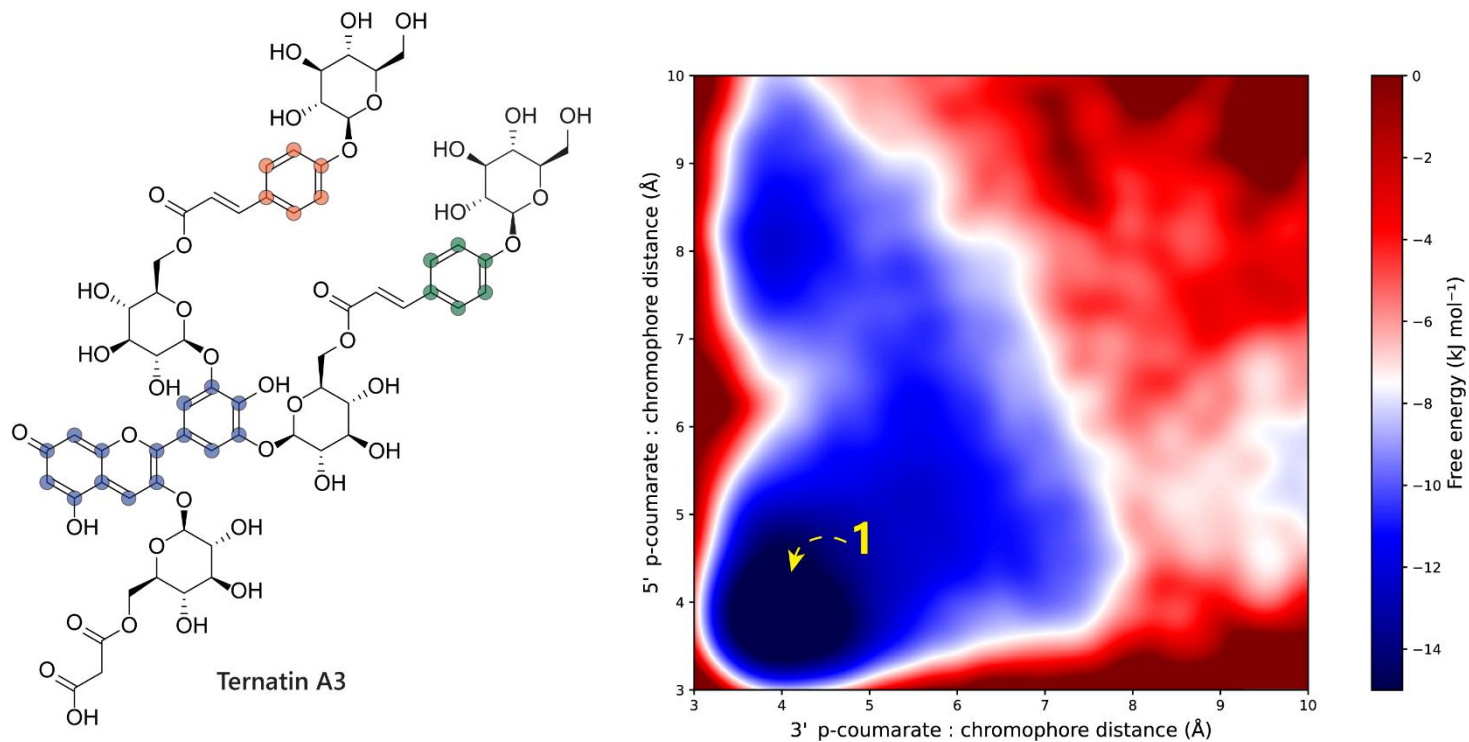**Figure 4.5** Free energy surface for distances acyl groups and ternatin A3 chromophore. Left: Structural formula of ternatin A3. Free energy of association was measured between the 3' acyl residue atoms (orange) and 5' acyl residue atoms (green) as a function of distance from the chromophore atoms (blue). Free energy surface shows only one strong energy minimum when both acyl residues are between 3.5-4.8 Å from the chromophore (1).

4.3.1.3 Localisation of acyl residues to chromophore rings

The free energy of association between the acyl residues and each of the chromophore ring systems was calculated to obtain more information regarding the spatial geometry of the intramolecularly stacked structures. As shown in figure 4.6 (**A**), the single 3' acyl residue of ternatin C3 formed the most energetically favourable interaction with the C – ring of the delphinidin chromophore, with a energy minimum located at 3.5 Å. Furthermore, this lowest energy state occurred when the acyl residue was within close proximity to the B – ring, with a distance of 4.5 Å. A second low energy conformation was also present when the acyl residue was equidistance from each of the chromophore rings with a distance of 5.5 – 6 Å.

Ternatin D3 also forms a strong interaction between the 3' acyl residue and the region of space between the C – ring and B – ring, with an energy minimum at a distance of 3.8 Å and 4.5 Å respectively (**Figure 4.6, B**). The interaction between the 5' acyl residue was less defined however, with three energy basins at distances of 3.8 Å, 4.8 Å and 5.5 Å. Instead, this acyl residue appears to form a closer association with the A – ring at a distance of 4 – 5 Å . Similar to figure 4.3, the 5' acyl residue has a relatively flat energy basin between the A and C rings, indicating a propensity for the residue to drift out of interaction range with the chromophore, possibly through the formation of interactions with surrounding molecules, resulting in a decrease in the stability of the stacked conformation.
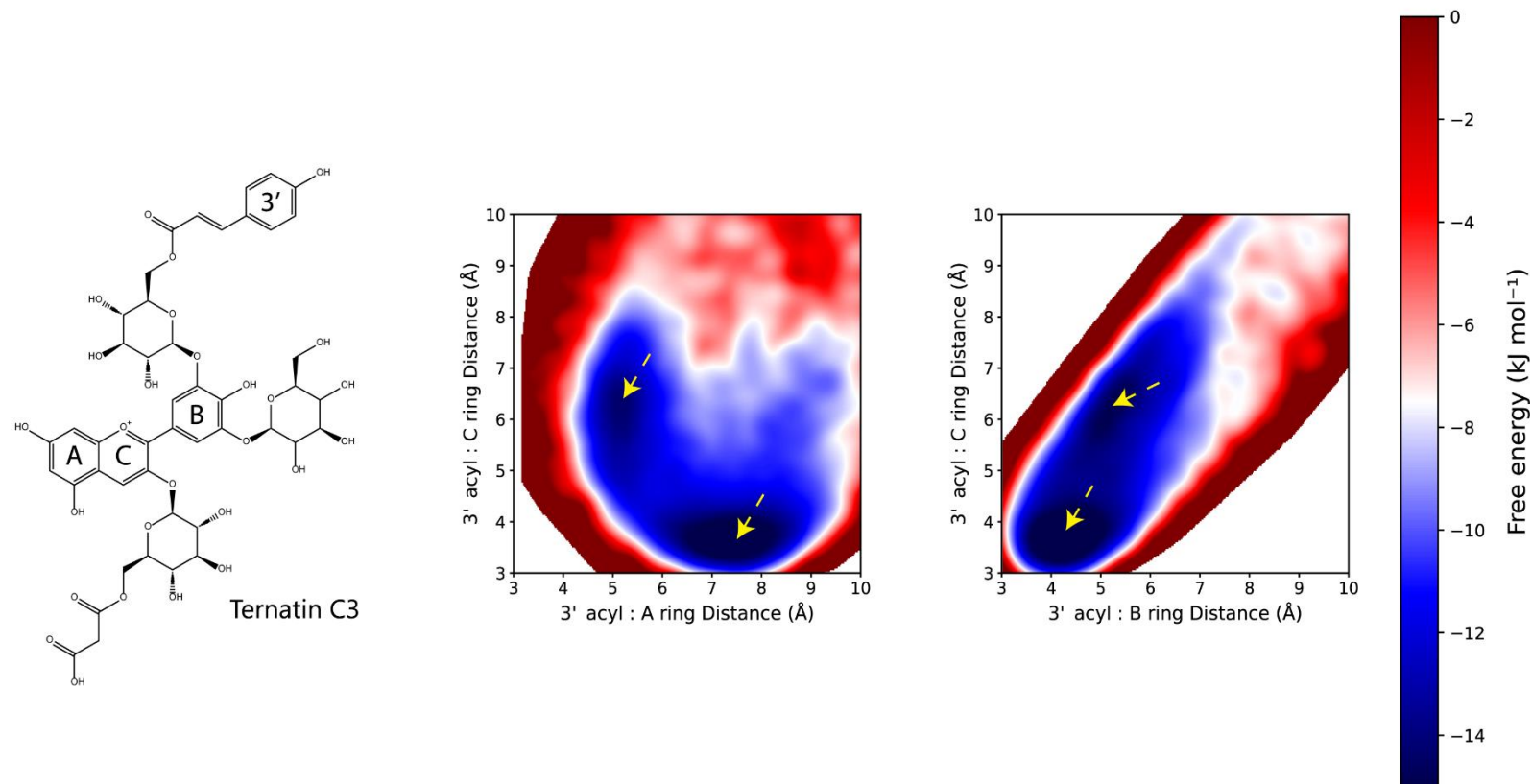
(A)

Figure 4.6 (A) Free energy of association between the 3' (top) acyl residue and anthocyanin ring systems. The structural formula of ternatin C3 is shown on the left; Free energy of interaction between the A and C chromophore rings (middle); C and B rings (right). Energy minima are highlighted with a yellow arrow.

(B)
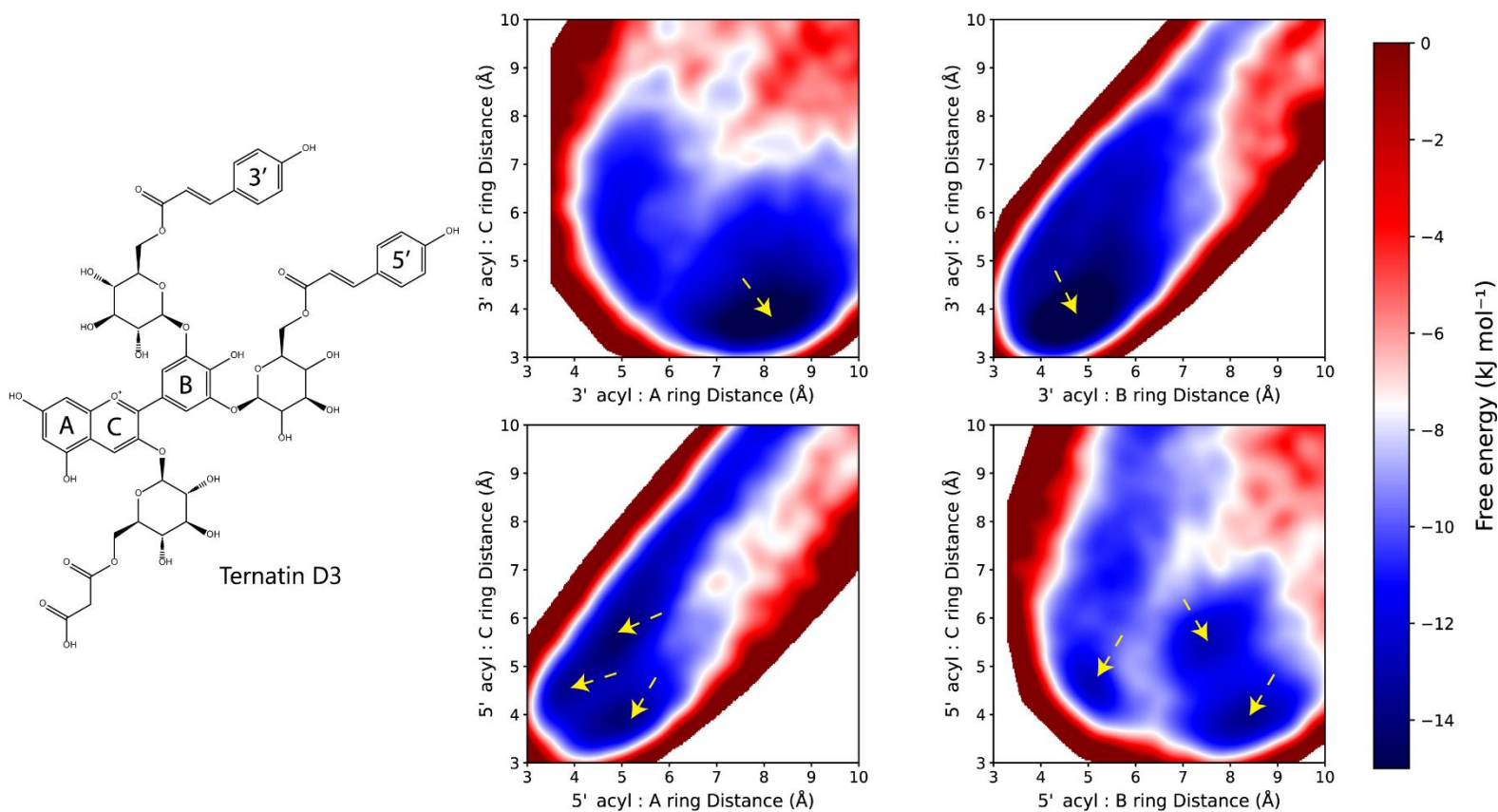
Figure 4.6 (B) Free energy of association between the 3' (top) and the 5' (bottom) acyl residue and anthocyanin ring systems. The structural formula of ternatin D3 is shown on the left; Free energy of interaction between the A and C chromophore rings (middle); C and B rings (right). Energy minima are highlighted with a yellow arrow.
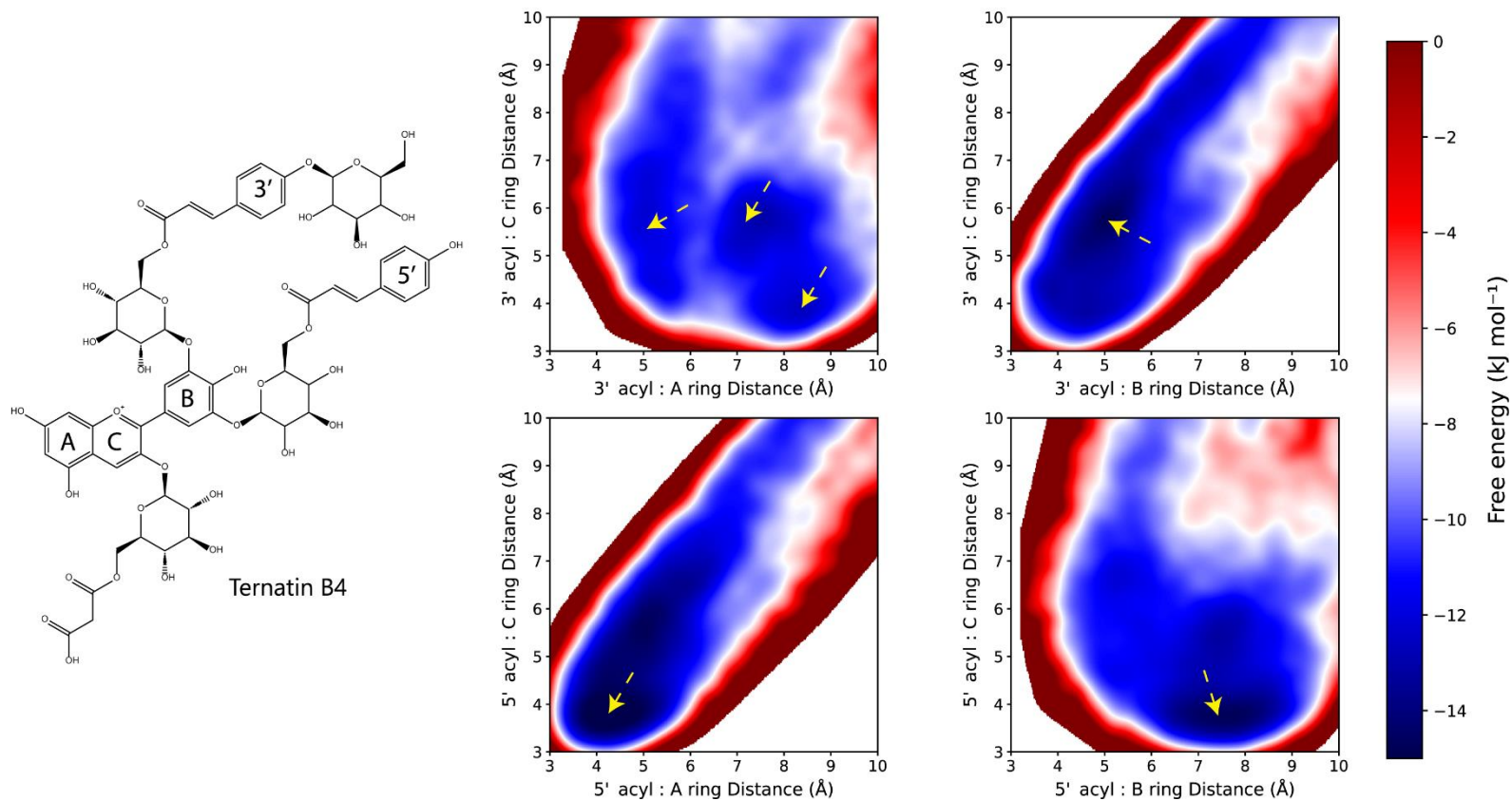
Figure 4.6 (C) Free energy of association between the 3' (top) and the 5' (bottom) acyl residue and anthocyanin ring systems. The structural formula of ternatin B4 is shown on the left; Free energy of interaction between the A and C chromophore rings (middle); C and B rings (right). Energy minima are highlighted with a yellow arrow.

(B)

Figure 4.6 (D) Free energy of association between the 3' (top) and the 5' (bottom) acyl residue and anthocyanin ring systems. The structural formula of ternatin A3 is shown on the left; Free energy of interaction between the A and C chromophore rings (middle); C and B rings (right). Energy minima are highlighted with a yellow arrow.
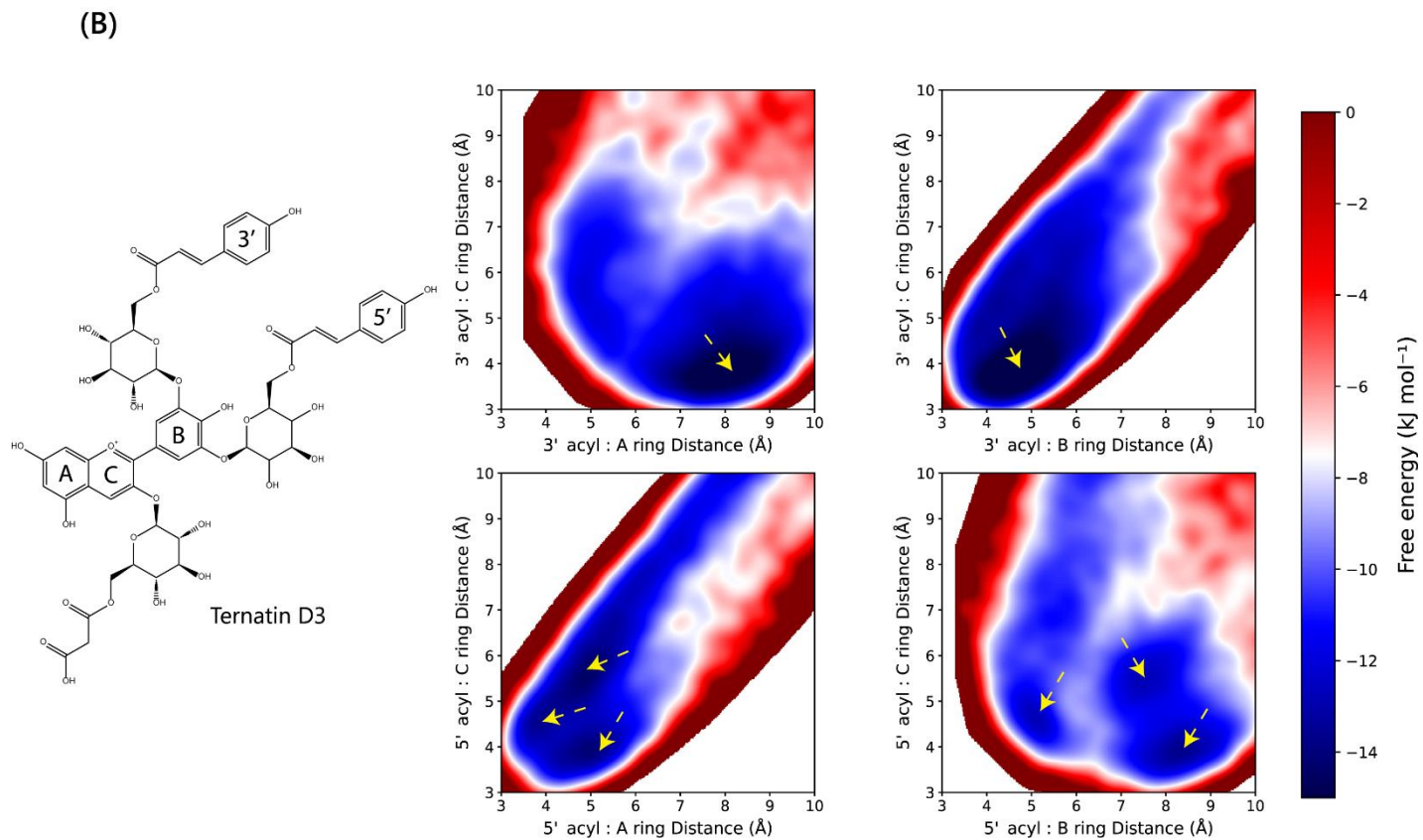
Glycosylation of the 3' sidechain from ternatin D3 to ternatin B4 (**Figure 4.6, C**) decreases the distance between the 5' acyl residue and both the A and C rings, with a distinct energy minimum at distances of 4.5 Å and 3.8 Å respectively. The interactions between the 3' acyl residue and the chromophore are considerably weaker, with association distances of 5.2 Å from the B – ring and 5.8 Å from the C – ring.

Within the free energy surface of ternatin A3 (**Figure 4.6, D**), energy minima for both the 3' and 5' acyl residues are distinct compared with the surrounding free energy landscape. The relatively lower free energy surrounding the minima is indicative of larger barriers to disassociation and hence, strong intramolecular forces stabilising the interaction between the chromophore and the acyl residues. Despite the strong interaction, the minima were diffuse, ranging from 3.8 Å – 6 Å between the 3' acyl residue and the chromophore C – ring, 3.8 – 4.5 Å from the B – ring and 5 Å from the A – ring. The 3' acyl residue displayed a preference for stacking over the B – C ring interface, however compared to the 3' acyl group of ternatin C3, this interaction is biased towards the B – ring, and encompasses a larger range of distances. The addition of glucose residues appears to weaken tight association of the residues, allowing the acyl residue to shift over a larger area of the chromophore. Unlike the case of ternatin B4 however, the high energy barriers to disassociation reduce the probability of the anthocyanin unfolding into an open conformation.

4.3.1.4 Effect of inner 3' and 5' sidechain on B – ring dihedral angle free energy

Figure 4.6 shows the free energy as a function of the B – ring dihedral angle. Ternatin C5, with only a single glucose residue linked through the 3' and 5' hydroxyl groups had a relatively symmetrical free energy landscape, with a slight bias towards a positive dihedral angle. In this structure, there were four observable energy basins at dihedral angles of -117.1°, -60.0°, 63.4° and 119.4°. However, when the reference angle for each dihedral angle was calculated, these dihedral angles were equivalent to 62.9°, 60°, 63.4° and 60.6° respectively, equating to a largely similar dihedral angle despite the orientation of the B – ring. Acylation of the 3' sidechain induces a significant bias towards -50.9° (reference angle of 50.9°), suggesting that negative rotation of the B – ring allowed for increased spatial alignment of the acyl residue with the chromophore. Both ternatin D3 and B4 had similar torsion angle free energy landscapes, with minima located in the positive direction of rotation.

Comparing the dihedral angle of ternatin D3, ternatin B4 and ternatin A3 with the free energy acyl : chromophore association (**Figure 4.7**), both ternatin D3 and ternatin B4 featured two local energy minima at approximately the same distances between the acyl residue and the chromophore, as well as two observable minima in the dihedral angle free energy surface. In contrast, ternatin A3 has only one minimum in both the acyl residue : chromophore distance and the dihedral angle free energies . This was indicative that a dihedral angle of around ≈ 138° was required for the close association of acyl residues with the chromophore, whereas ≈ 60° leads to weaker interactions.
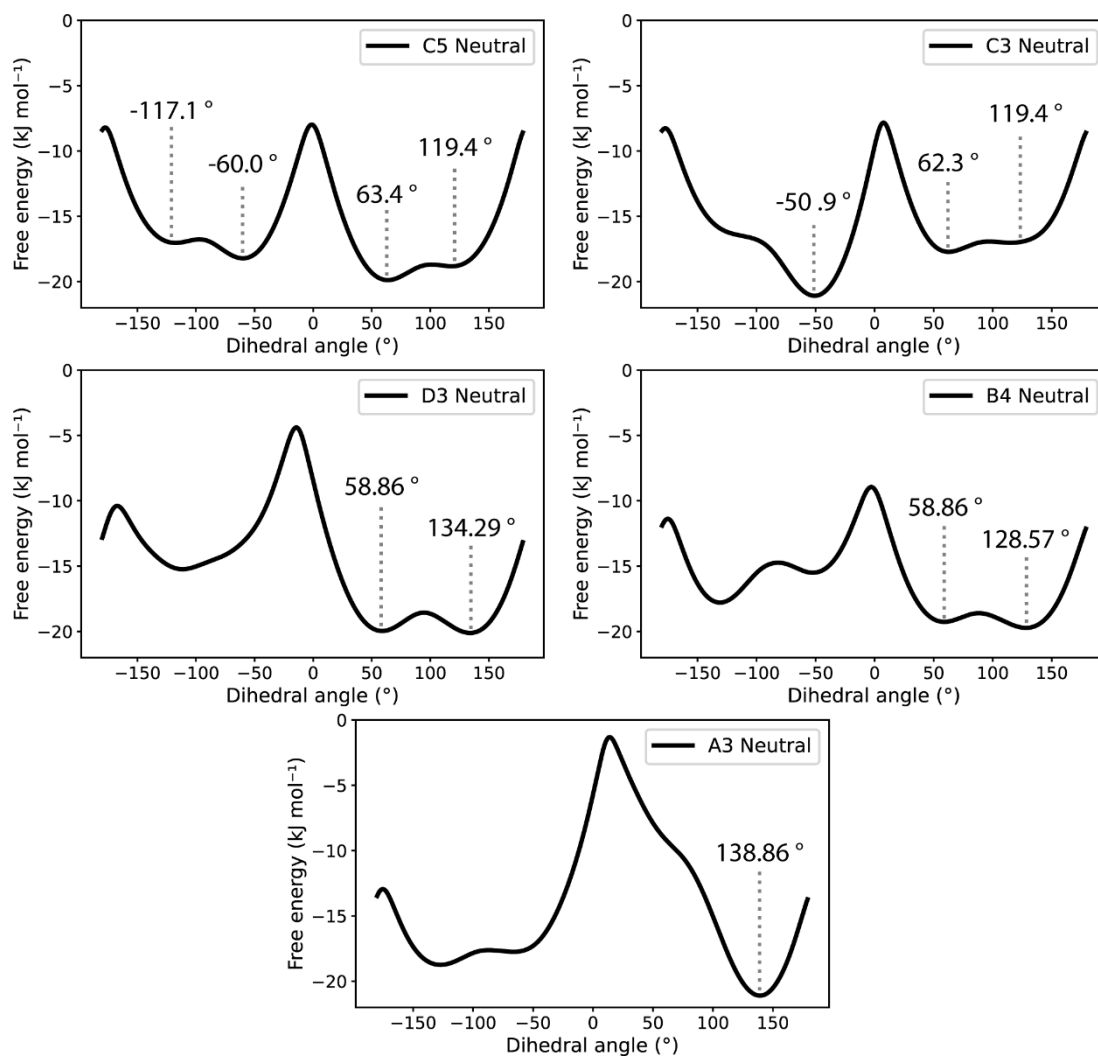
Figure 4.7 Free energy of the C3, C2, C1", C2" dihedral angles of ternatins C5, C3, D3, B4 and A3.

4.3.2 Classical molecular dynamics

Irreversible degradation of anthocyanins occurs first by nucleophilic attack of the C2 atom by water, followed by decomposition through a poorly characterised network of side reactions. Therefore, by modelling each of the anthocyanins in a full explicit water solvent, the radial distribution function (RDF) between the water molecule and C2 can indicate the probability of initiation of this degradation pathway. By comparing the RDF between each of the ternatins, the influence of iterative decorations on the stability of the anthocyanin could be inferred.

4.3.2.1 First decorations – ternatin C5 → ternatin A3

Figure 4.8 shows the radial distribution functions of the first five neutral ternatin intermediates. The first decorations of delphinidin which form ternatin C5, (via glycosylation of the 3' and 5' hydroxyl groups) result in an absolute mean percentage difference of 8.6 % (7.7 % with a cut-off distance of 5 Å) when compared to the simulation of delphinidin-3-glucoside. Addition of acyl groups considerably reduce the probability of water molecules in close proximity to the C2 atom. These decreases are largely symmetrical between acylation of the 3' sidechain and the 5' sidechain, with absolute percentage differences of 8.2 % from ternatin C5 to C3, and 7.4 % from ternatin C3 to D3. These effects are far more pronounced however between $0 - 5$ Å, with mean decreases of 40.3 % and 42 % for C5 → C3 and C3 → D3 respectively. With these first decorations, the acyl residues fold over the chromophore and likely reduce the density of water by steric hindrance. Further glycosylation of the side chain decreases the density of water around the C2 atom, with absolute mean decreases of 4.1 % for D3 → B4 and 5.0 % for B4 → A3. Again, the largest magnitude of decrease occurs within the interaction range of $0 - 5$ Å, with RDF differences of 27.9 % and 21.6 % for D3 → B4 and B4 → A3 respectively.
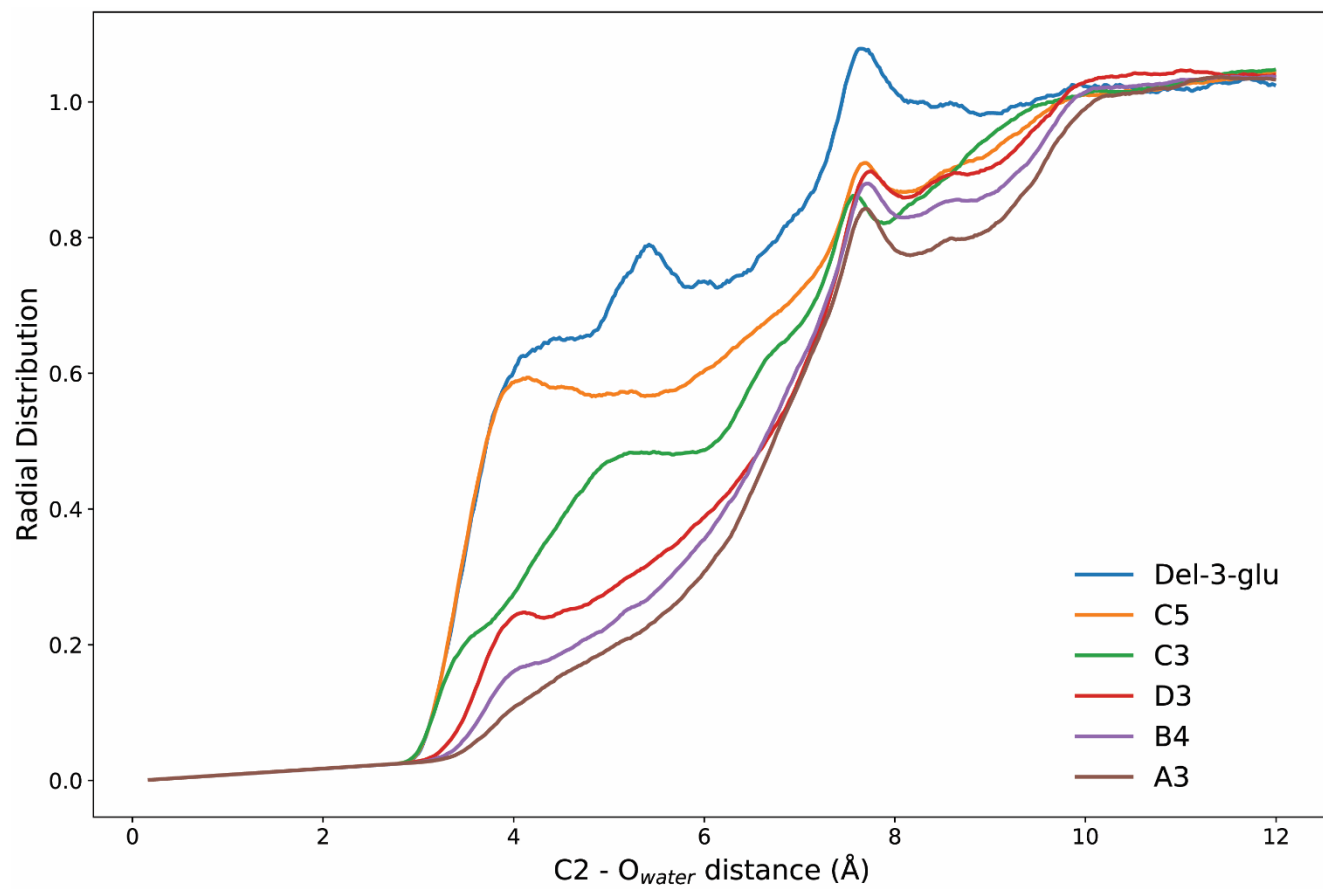
Figure 4.8 Radial distribution function for ternatins with 0, 1 and 2 acyl residues. Radial distribution shows the probability density between C2 and oxygen atoms of water as a function of distance for delphinidin-3-glucoside, ternatin C5, ternatin C3, ternatin D3, ternatin B4 and ternatin A3

4.3.2.2 Outer decorations – ternatin A3 → A1

Unlike in the inner most decorations of ternatins, further acylation of ternatin A3 sidechains with *p*-coumaric acid results in an increased density of water at the C2 atom for both acyl residues. Figure 4.9 shows that the initial decoration from A3 → B2 causes an absolute decrease in water density of 3.1 %, however most of this decrease is outside the typical range of interaction. When looking at the density within 5 Å, ternatin B2 has an increased density of water of 21. 3 %. Similarly, B2 → D1 shows a slight reduction in absolute density of 0.6 %, however has an increased probable density of 25.2 % within interaction range. Although water density within this 5 Å interaction range was increased, the overall probability that water initiates its trajectory towards the C2 atom decreased. This may be the result of additional steric hinderance by the acyl residues, however after passing through this obstacle the water molecules are then trapped in a closer association with C2, likely by hydrophobic forces from the same acyl residues. Decoration with a glucose residue on the 3' sidechain from D1 → B1 restores some of this decreased water density, with an absolute difference of 7.1 %. The largest magnitude of effect occurs within the interaction range between C2 and water, with a 37.6 % reduction in water density. Decoration of the 5' sidechain from B1 → A1 had a lesser effect, with an absolute difference of 3.0 %, however had a small increase in density within 5 Å of 2.8 %. In both cases of further decoration of the 3' and 5' sidechains, addition of the glucose residue results in an absolute decrease in water density from the C2 atom. This effect is likely the result of further steric hinderance, blocking water molecules which would otherwise have a trajectory towards close association with the C2 atom. Compared to the RDF of ternatin D1, ternatin B1 has a lower density of water within 5 Å.
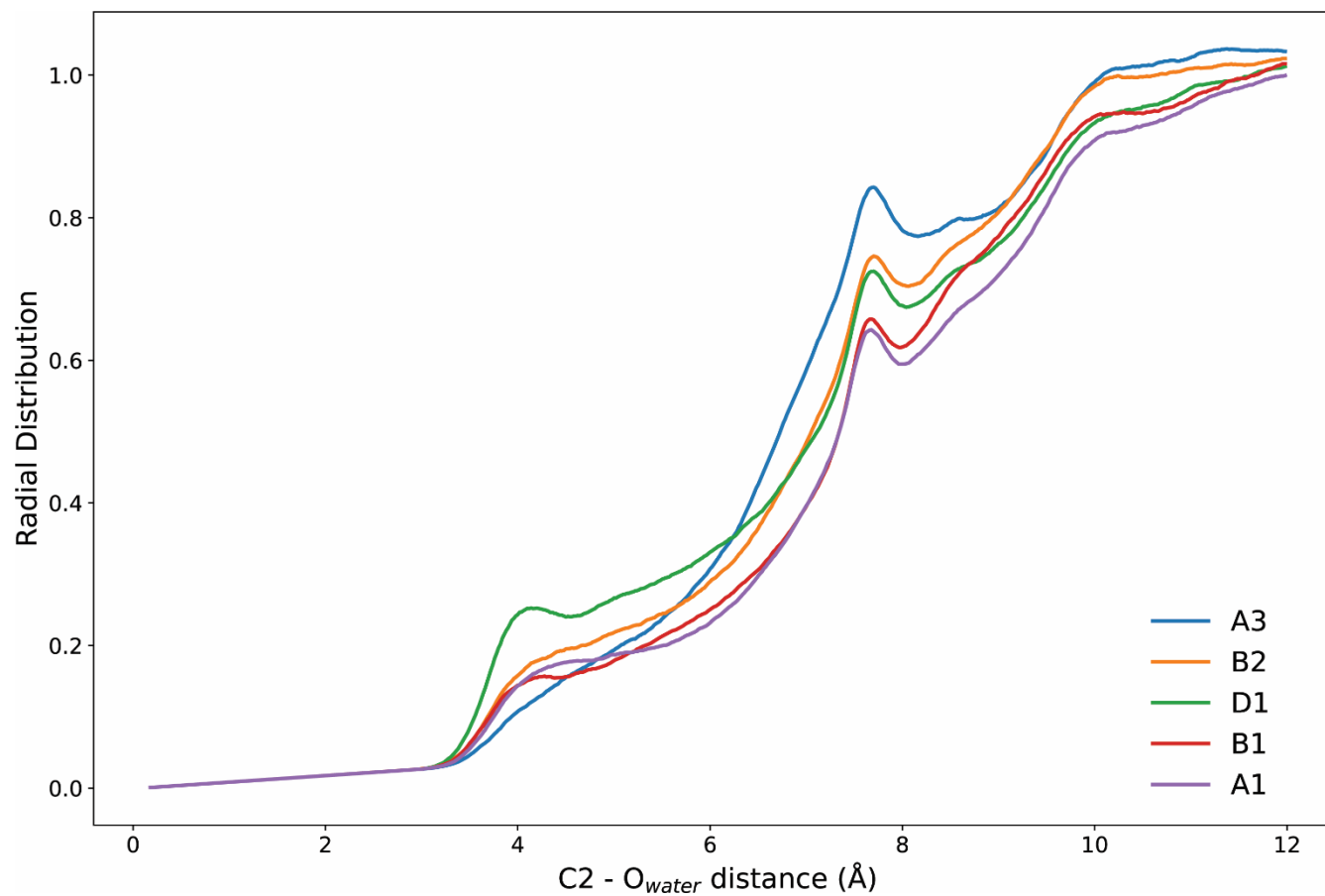
Figure 4.9 Radial distribution function for ternatins with 3 and 4 acyl residues. Radial distribution shows the probability density between C2 and oxygen atoms of water as a function of distance for ternatin A3, ternatin B2, ternatin D1, ternatin B1 and ternatin A1

4.3.3 Quantum mechanical simulations of delphinidin-3-O-glucoside rotamers

4.3.3.1 Selecting an appropriate functional

Since the advent of computational techniques for calculating absorbance from a molecular structure, a wide variety of functionals have been developed for use in conjunction with TD-DFT. Even within the narrower area of spectral calculations of anthocyanins, multiple functionals have been used for these calculations. Given that the functional component of the calculation is integral to accuracy of the results, simulations with five of the most common functionals (B3LYP, B3P86, M06, M062X and PBE0) were conducted on a geometry optimised delphinidin-3-O-glucoside model, and the first five transitions calculated were compared to experimentally acquired spectra.

Results from these calculations showed that the functionals B3LYP (**Figure 4.10, A**), B3P86 (**Figure 4.10, B**) and M06 (**Figure 4.10, D**) provided the closest approximation of the experimental results, with absorbance λ_{max} values of 504.4 nm, 503.6 nm, and 501.6 nm respectively, compared to the experimental λ_{max} of 516 nm. Although PBE0 (**Figure 4.10, E**) and M062X (**Figure 4.10, D**) have been previously reported to be highly accurate for spectral calculations of anthocyanins, with this implementation the absorption transition energy was significantly underestimated with a λ_{max} of 490.7 nm and 452.6 nm respectively.

Of the functionals B3LYP, B3P86 and M06, only B3P86 failed to predict the transition at 343 nm, while all functionals underestimated the oscillator strength of the $\approx$ 432 nm transition observed in the experimental spectra .
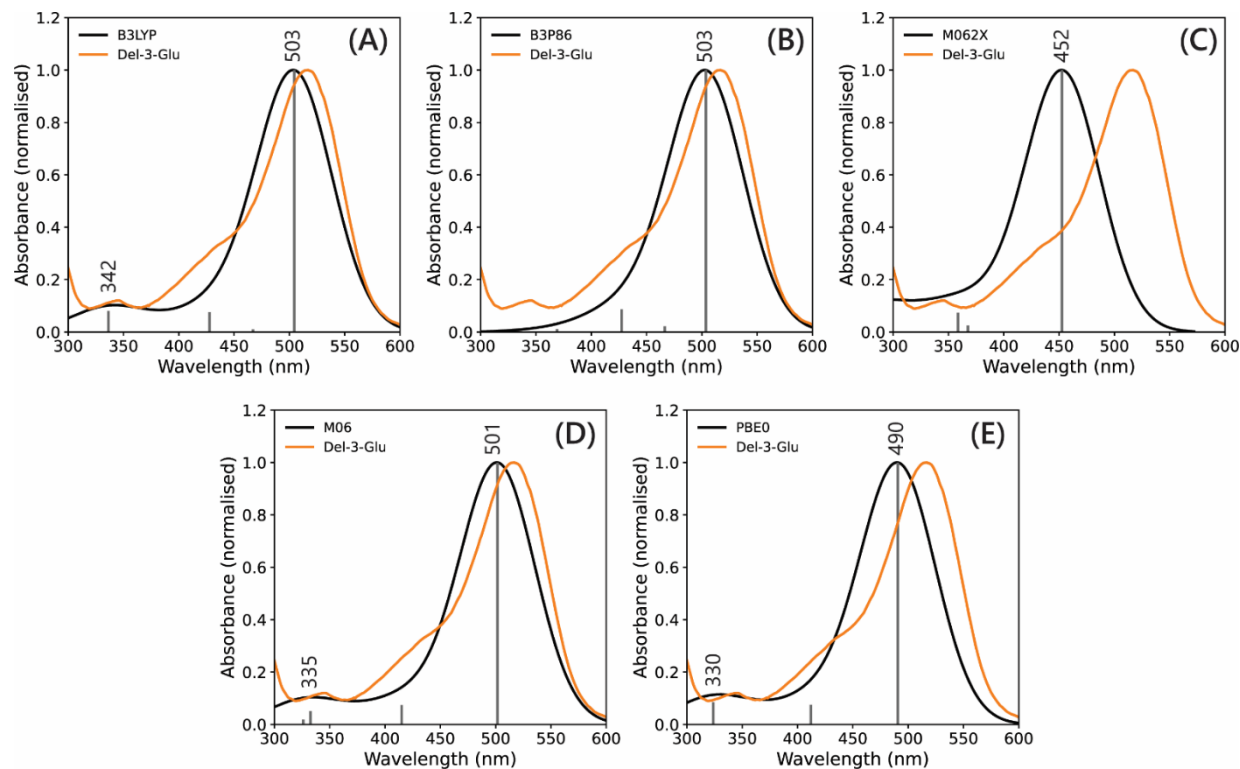
Figure 4.10 Simulated UV-VIS absorbance spectra of delphindin-3-O-glucoside. The black line shows the calculated UV-VIS absorbance, while the orange line shows experimentally acquired spectra of delphinidin-3-O-glucoside at 60 μM in aqueous 1 % HCl. Calculations were run using a basis set of 6-311+G(d,p) with a CPCM water model. The following functionals were used in the calculations: (A) B3LYP; (B) B3P86; (C) M062X; (D) M06; and (E) PBE0

4.3.3.2 Calculated colour of simulated spectra

As the colour characteristics of a pigment are the result of the collective absorbance between 350 nm and 750 nm, morphing of the simulated spectra into colour space coordinates can help more clearly evaluate the accuracy of the calculated results. Therefore, in order to validate the functionals tested, each of the simulated UV-VIS spectra were converted into CIE L*a*b* colour space coordinates using the method outlined in section 3.2.6.

Observations of the output colours calculated using the simulated UV-VIS spectra in figure 4.10 show clearly that B3LYP, B3P86 and M06 all closely match the experimental colour of delphinidin-3-O-glucoside (**Figure 4.11, A**). However, between the calculated colours, there are noticeable differences between the accuracy of the simulated spectra.

B3LYP, despite underestimating the experimentally obtained major absorbance peak at λ_{max} of 514 nm by 11.6 nm, as the bulk of the simulated absorbance overlaps the experimental absorbance peaks, the colour of the simulated spectra was very similar to that of delphinidin-3-O-glucoside (**Figure 4.10, A**). Calculating the Euclidian between the colour output of B3LYP and experimental colour values gave a colour space difference of 2.18, slightly below the "just noticeable difference" value of $\Delta E^*ab \approx 2.3$. Most of this difference was the result of the underestimation of the oscillator strength of the 432 nm transition band gap, which in turn corresponded to the $\bar{y}$ (L*) component (400 nm – 500 nm) of the tristimulus curves and was evident as a slight increase in the L* value, or lightening of the reported colour (figure 4.10, B).

Similarly, both B3P86 and M06 overestimated the L* value, but also more significantly underestimated the a* value (less red) and overestimated the b* values (more yellow) (**Figure 4.11, B**), leading to their increased colour differences of 3.02 and 5.75 respectively.

With this implementation of the calculation, M062X and PBE0 performed poorly in accurately predicting the absorbance spectra of delphinidin. As the main transition energy was predicted at -63.4 nm and -25.3 nm for M062X and PBE0, the absorbance

peak was shifted outside of the x̄ component (500 nm – 650 nm) and into the ȳ region of the tristimulus curves, leading to large differences in all L*a*b* values (**Figure 4.11, B**), and colour differences of 89.33 and 28.52 respectively.
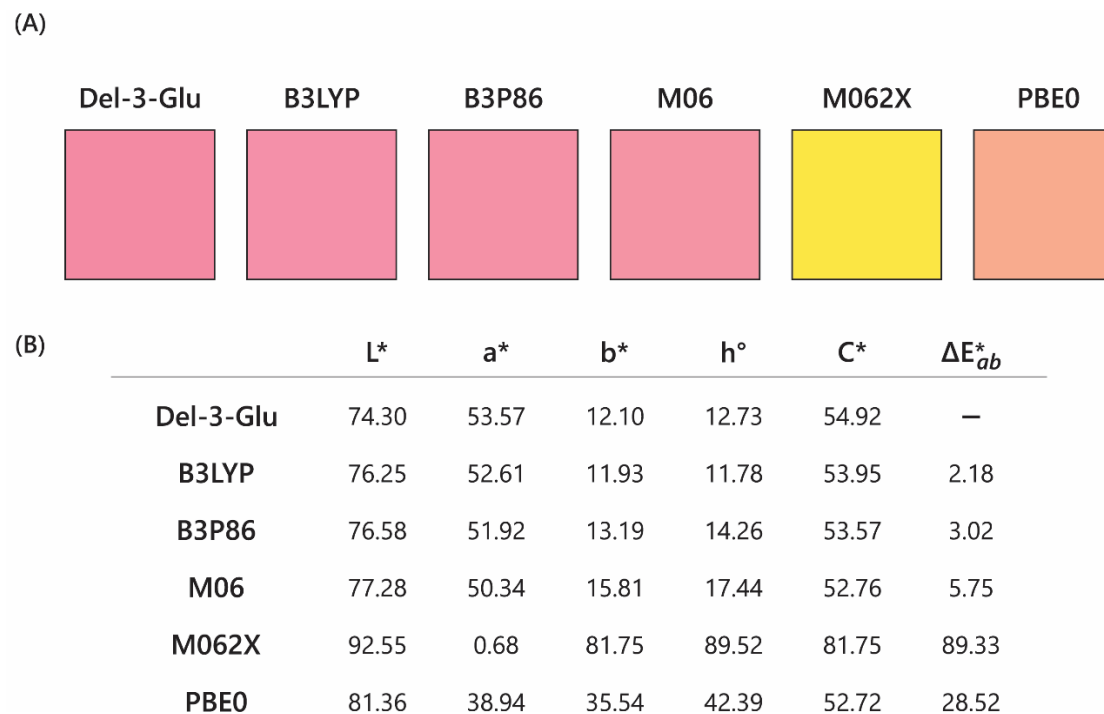
(A)

| Del-3-Glu | B3LYP | B3P86 | M06 | M062X | PBE0 |

(B)

	L*	a*	b*	h°	C*	ΔE^*_{ab}
Del-3-Glu	74.30	53.57	12.10	12.73	54.92	—
B3LYP	76.25	52.61	11.93	11.78	53.95	2.18
B3P86	76.58	51.92	13.19	14.26	53.57	3.02
M06	77.28	50.34	15.81	17.44	52.76	5.75
M062X	92.55	0.68	81.75	89.52	81.75	89.33
PBE0	81.36	38.94	35.54	42.39	52.72	28.52

Figure 4.11 (**A**) Representative colour of simulated UV-VIS absorbance spectra; (**B**) CIE L*a*b* h° C* colour space parameters for each simulated spectrum.

ΔE*ab was calculated using the following formula: $\Delta E^*_{ab} = \sqrt{(L^*_1 - L^*_2)^2 + (a^*_1 - a^*_2)^2 + (b^*_1 - b^*_2)^2}$

4.3.4 Influence of B – ring rotation on colour

4.3.4.1 UV-VIS absorbance

As the functional B3LYP was the most accurate in predicting the UV-VIS absorbance spectrum, this functional was selected for calculation of the excitation energies of the delphinidin rotamers. Using the same geometry optimised molecular model from the functional screening experiment, a series of conformations were generated, differing only in the dihedral angle of the B – ring.

Each of rotamers were then briefly relaxed with a short geometry optimisation and then the absorbance spectra were calculated. As can be seen in figure 4.12, increasing the dihedral angle of the B – ring induced a bathochromic shift. Between dihedral angles of 0 ° and 60 °, the bathochromic shift is relatively small, with a shift in λ_{max} of 1.1 nm, 3.7 nm, 4 nm, 6.6 nm, 10.6 nm and 16.7 nm for each 10 ° increment when compared to the 0 ° conformation.

However, past a dihedral angle of 60 °, this observed bathochromic shift became much larger, with shifts of 30 nm, 44.4 nm and 55.1 nm for 70 °, 80 ° and 90 ° angles. The highest occupied molecular orbital (HOMO) – lowest unoccupied molecular orbital (LUMO) transition was normalised to 1, and the spectra converted into CIE L*a*b* colour space values. From this plot, rotation of the B – ring clearly induces a colour shift from red to purple as the dihedral angle increased (**Figure 4.13**). The dihedral angle also influenced the probability of light absorption, as seen by a decrease in oscillator strength as the angle increases (**Table 4.1**).

Table 4.1 Wavelength maxima (nm), oscillator strength (f) and transition energy between the HOMO and LUMO (eV) for each delphindin-3-glucoside rotamer

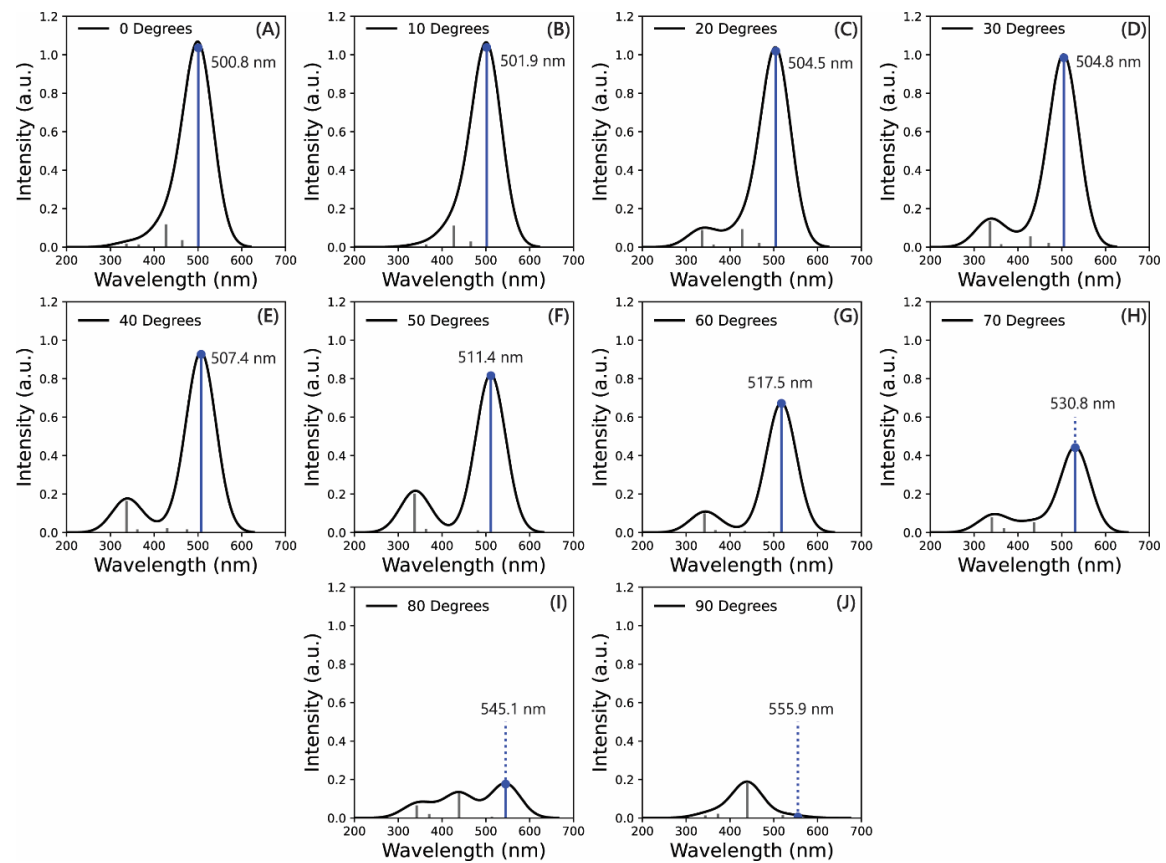| Dihedral Angle (°)	Wavelength (nm)	Oscillator Strength (f)	E (eV)
0	500.8	1.0363	2.476
10	501.9	1.0382	2.470
20	504.5	1.0193	2.458
30	504.9	0.9845	2.456
40	507.4	0.9268	2.444
50	511.4	0.8158	2.424
60	517.4	0.6717	2.396
70	530.8	0.4416	2.336
80	545.1	0.1763	2.274
90	555.0	0.0054	2.234

Figure 4.12 Simulated UV-VIS absorption spectra of B – ring rotamers. The HOMO – LUMO transition is indicated in blue
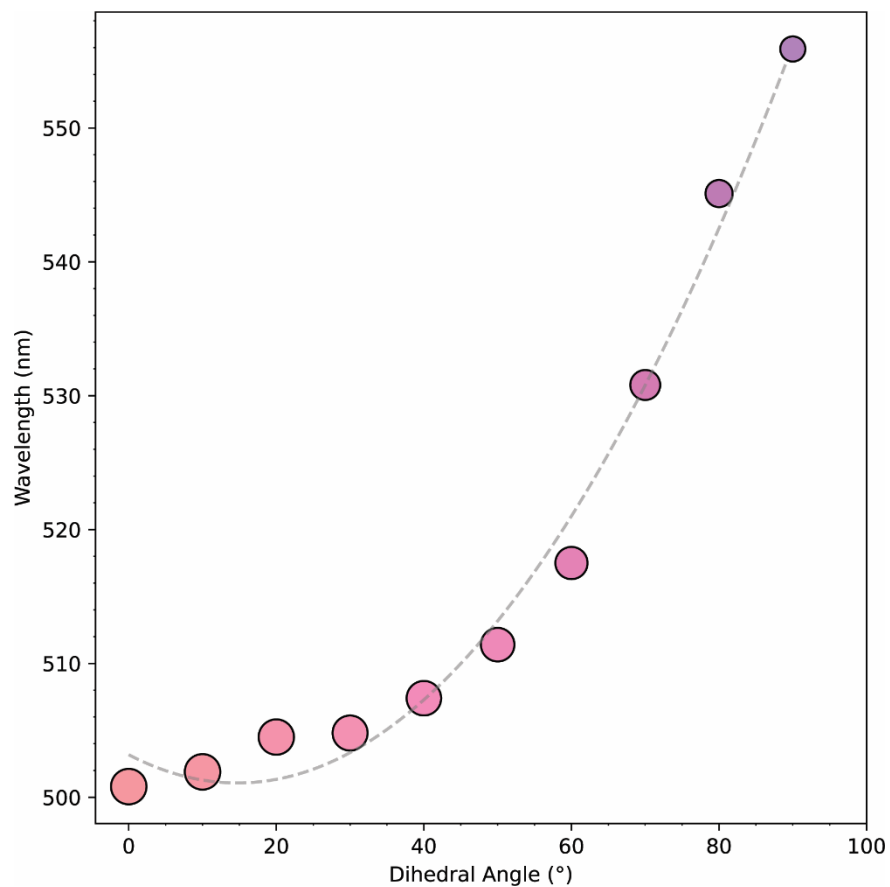
Figure 4.13 Calculated effect of λ_{max} of the HOMO-LUMO transition as a function of dihedral angles. The colour of each circle is the expected colour, marker size corresponds to the oscillator strength of the transition.

4.3.4.2 Effect of dihedral angle on frontier molecular orbitals

Molecular orbitals are regions of space around atomic nuclei where there is an increased probability of finding an electron, with frontier molecular orbitals referring specifically to the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO). Plotting of the frontier molecular orbitals (**Figure 4.14**) for the rotamers reveal a delocalisation of the HOMO and LUMO orbitals as B – ring dihedral angle increased.

At a dihedral angle of 0°, both the HOMO and LUMO were localised across the whole of the chromophore system. As the dihedral angle increased from 0° to 60° the spatial density of the HOMO and LUMO begins to segregate, with an increased LUMO density localised more around the A and C rings, and the HOMO localised around the B – ring. As the dihedral angle increased further to 90°, the HOMO and LUMO density were nearly completely separate, with almost all of the LUMO electronic density found over the A and C rings, and the HOMO density found almost exclusively around the B – ring.

As seen in figure 4.12, the largest colour change is observed as the dihedral angle increases from 60° to 70°, with smaller increases as the dihedral approaches 90°. This correlates with the observed separation between the HOMO and LUMO orbitals between 60° and 90° observed in figure 4.14, indicating that the separation between frontier molecular orbitals decreases the transition energy gap, resulting in a bathochromic shift.
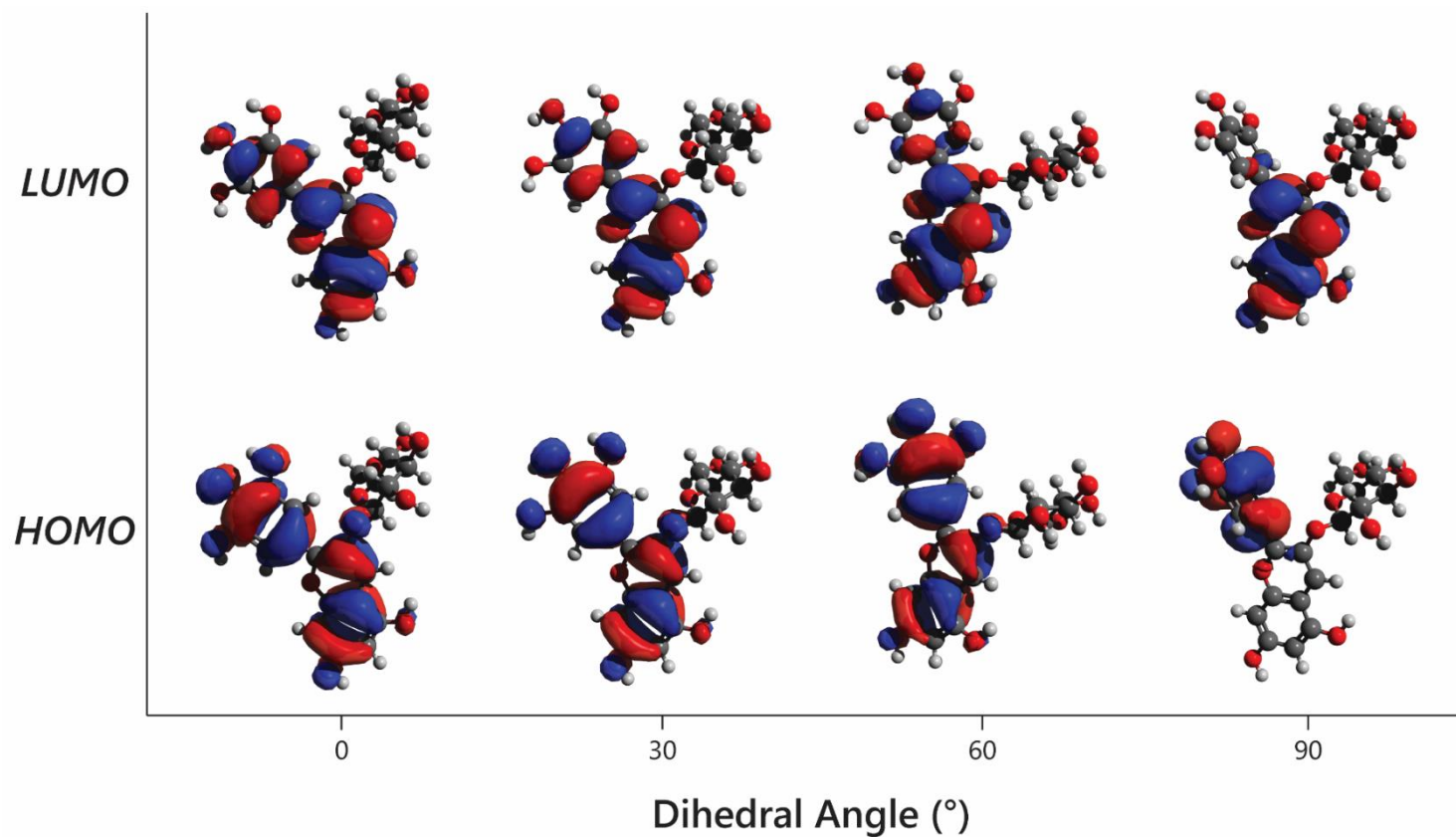
Figure 4.14 Frontier molecular orbitals (MOs) involved in the first electronic transition of each delphinidin-3-O-glucoside rotamer. As the dihedral angle increased, both the HOMO and LUMO orbitals began to segregate into two distinct regions. This was most evident at a dihedral angle of 90°, where the HOMO orbital was localised entirely on the delphinidin B-ring, while the LUMO orbital was present only on the A- and C-rings

4.4 Discussion

This work provided the basis for a better understanding the influence anthocyanin decorations have on the spatial conformation of stacking type interactions. However, to optimise an anthocyanin for use as a replacement blue colourant, a more comprehensive understanding of anthocyanin decoration at atomic resolution will be required. This will likely be dependent on the molecular simulations of a broad range of anthocyanins which have been shown to produce blue colours experimentally, to better realise the stability enhancing effects attained by the wide variety of covalently bound decorations, as well as through intramolecular copigmentation.

4.4.1 How decoration influences stacking configurations in intramolecular copigmentation

Anthocyanin copigmentation influences colour in two major ways; by stabilising the coloured species against hydration of the C2 atom and by charge transfer between the copigment structure and the anthocyanin chromophore.

Stability enhancement can be accomplished by reducing the accessibility of nucleophiles to the anthocyanin ring system. This is achieved by stacking of the acyl residues over the chromophore, sterically hindering water molecules. Radial distribution simulations show a decreasing density of water as the number of decorations increase between ternatin C5 and A3 (**Figure 4.8**). However, these trends were reversed with ternatins bearing additional decorations past A3 (**Figure 4.9**) and would suggest that increasing number of decorations does not necessarily equate to increased stability by steric hinderance.

The geometry of these intermolecular stacks favoured the 3' acyl residue in close association with the B – and C – rings, whereas the 5' acyl residue stacked over the A – and C – rings (**Figure 4.3 – 4.5**).

This is likely as a result of the proximal location of the two sidechains and the dihedral angle of the B – ring, where rotation of the B – ring is required to facilitate overlap of the acyl residues and the chromophore.

172

The close association of acyl residues with the chromophore facilitated molecular orbital mixing (Trouillas *et al.*, 2016), stabilising the transition state and effectively reducing the transition energy required to excite electrons from the HOMO to the LUMO. By reducing the transition energy gap to a lower energy, lower energy light (higher wavelength) were required for excitation, thereby inducing a bathochromic shift and resulting blue-shifting of the colour of the anthocyanins colour.

Additional decorations are at a greater distance from the chromophore, and so experience less attractive force. Moreover, they exert a pull force as they drift around the molecule, acting to pull on the whole sidechain and may displace the acyl residues from the chromophore, increasing nucleophile accessibility.

As X-ray crystallography remains unviable for atomic resolution investigation of anthocyanins, accurate molecular modelling was an essential tool for understanding the structural characteristics which give rise to enhanced stability and colour characteristics. Although the influence of decorations on the association between acyl decorations has been characterised, the molecular dynamic simulations of the ternatin intermediates was of insufficient duration to allow for ring opening events to occur.

In order to fully understand the comparative strength of association between each of the ternatins, longer duration simulations would likely be required to gather sufficient disassociation events to calculate the probability of these open conformations. Furthermore, to understand the changes in transition energy (including the charge transfer state) upon binding of copigments, the transition energy for each ternatin needs to be calculated using a quantum mechanical method, such as time-dependent density functional theory

4.4.2 Simulation UV-VIS absorption

One of the challenges of characterising the effects of anthocyanin decoration on the UV-VIS spectra is to obtain isolates free from contaminants which may act as copigments. Work documented in this thesis has demonstrated a method for the accurate *ab initio* calculation of UV-VIS spectral data and CIE L*a*b* colour space coordinates with a Euclidian colour difference of 2.18, generated from a simple

molecular structure. This process may be further applied to anthocyanins either found in nature or used as a way to screen the colour characteristics potential for "designer" anthocyanins generated through genetic manipulation, prior to undertaking potentially time consuming and labour-intensive experiments.

4.4.3 Why dihedral angle influences absorbance

There are several factors which influence the λ_{max} of UV-VIS absorption in anthocyanins. The results presented in figure 4.12 demonstrate one such mechanism, by which increasing the B – ring dihedral angle produced a bathochromic shift. This effect was evident in anthocyanins which lacked any decorations which would otherwise interact with the chromophore, and where there was no other copigment present. One such example of this effect was with ternatin C5. Terahara et al. (Terahara *et al.*, 1998) reported a λ_{max} of 523 nm for the cationic species, which when compared with the cation of delphinidin-3-O-glucoside (λ_{max} = 516 nm) represented an increase of 7 nm. Measured at low concentration, the only influence present on the transition energy is that exerted by the 3' and 5' glucose residues, which induce a torsion in the B – ring dihedral angle of 63.4° (**Figure 4.11**).

Noda et al.(Noda *et al.*, 2017) were able to generate blue coloured chrysanthemums by expression of the UDP-glucose : anthocyanin 3',5'-O-glucosyltransferase gene. This blue colouration was achieved via copigmentation with luteolin 7-O-(6"-malonyl)glucoside and tricetin 7-O-(6"-O-malonyl)glucoside natively biosynthesised. In contrast, when delphinidin-3-O-glucoside only was produced by expression of a flavonoid 3',5' hydroxylase, the resulting flower colour was violet. This effect may be explained in part by the increased dihedral angle of ternatin C5 compared to delphinidin, especially if the copigment induces additional rotation of the B – ring.

The effects of dihedral angle on the λ_{max} are due to a delocalisation of the HOMO and LUMO orbitals to regions of space with little overlap. This effect has been previously characterised in the simplest of the organic blue pigments, azulene (Michl and Thulstrup, 1976). Azulene is a bicyclic organic compound which is composed of a five-membered ring fused to a seven-membered ring. In such an arrangement, electronic charge is polarised, with the five-membered ring becoming electronegative, while the seven-member ring is electropositive. Azulene produces

a blue colour due to the spatial segregation of the HOMO and LUMO. The transition energy required to promote an electron to the excited state is proportional to the square of the dipole moment of the overlap density between the HOMO and LUMO, with smaller overlap resulting in a reduction in transition energy. This effect is similarly observed in figure 4.13., where a larger B – ring dihedral angle effectively segregates electronic density between the anthocyanin B – ring and the A- and C-ring systems.

Chapter 5

General Discussion

Chapter 5: General Discussion

5.1 Overview

This thesis aimed to provide an in-depth analysis and characterisation of the anthocyanins present in butterfly pea, assess the effects that iterative 3' and 5' glucosyl-acyl decoration had on stability and colour properties of the pigments, and provide insight into the molecular mechanisms responsible for these effects. Numerous objectives have been achieved, including: A) Development of a high capacity stationary phase for fractionation of anthocyanins using flash chromatography; B) Validation of the ternatin structures and generation of data on intramolecular copigmentation interactions; C) Development of a python application for high-throughput CIE-L*a*b* colour space measurements from UV-VIS spectral data; D) Characterisation of the relationship between ternatin decoration, stability and colour properties; E) Modelling of the intramolecular interactions using molecular dynamics and quantum mechanics, to understand how anthocyanin decorations influence stability and colour.

5.2 PVPP flash chromatography: A novel strategy for anthocyanin isolation

The extracts of dried butterfly pea flowers contain 15 delphinidin-3-(6"-malonate),3',5'-O-triglucoside derivatives (ternatins), decorated with alternating glucose and *p*-coumarate residues on the 3' and 5' hydroxyl groups. These can be broadly categorised into four groups: The A series, which feature glucose residues at each of the terminal positions of the 3' and 5' sidechain; the D series, which have *p*-coumarate residues at each of the terminal positions; the B series which have one glucose and one *p*-coumarate at the terminal positions; and the C series, where additional decorations are added to the 3' sidechain only.

As well as the main anthocyanins, crude extracts contain a further 15 delphinidin-3,3',5'-O-triglucoside derivatives (pre-ternatins), and multiple decorated flavonols. The presence of these additional products make isolation of ternatins particularly challenging, especially when considering the flavonols are present at ≈ x4 the

concentration relative to the anthocyanins (Kazuma, Noda and Suzuki, 2003). Due to the low relative concentration of anthocyanins, numerous separations using HPLC were required before an appreciable amount of purified anthocyanin could be obtained, made more difficult co-elution of products on C18 stationary phases. Therefore, to generate ternatin isolates, a method was required to reduce the complexity of the sample prior to preparative-HPLC.

Using column chromatographic separation with polyvinylpolypyrrolidone (PVPP) previously described in the literature , crude anthocyanic extracts were fractionated into fractions containing anthocyanins with similar terminal sidechain residues (**Figure 2.2**). Fractionation of butterfly pea extract resulted in three fractions: series A ternatins with two terminal glucose residues were the first to elute, followed by series B ternatins with one terminal glucose and one terminal *p*-coumarate, and finally series D ternatins with two terminal acyl residues.

PVPP is a clarifying agent, typically used in winemaking to adjust the colour profile of rosé wine and to reduce the overall polyphenol content which contributes to bitterness. PVPP has been shown to have a larger binding capacity for anthocyanins and better recovery rates than macroporous and cation exchange resins. In the past, PVPP had gained interest as a potential stationary phase for the isolation of anthocyanins, with the ability to retain anthocyanins via hydrogen bonding and π-π interaction. However, using open column chromatography resulted in low reproducibly between separations, coupled with long development times for separation. It is likely that these reasons caused PVPP to fall out of favour as a stationary phase.

Consequently, a flash chromatography method was developed using PVPP, which allowed for a higher degree of control over the separation, including solvent flow rate and composition. This method yielded similar results to the column chromatography; but the separation time was reduced from six hours to one hour, and reproducibility was greatly improved. Co-elution of fractions still occurred using the flash chromatography system, and therefore re-running the fractions where overlap occurred was necessary. However, this could be improved further by optimisation of chromatographic parameters, including flow rate, solvent type, and gradient profile. In addition to providing an automated, orthogonal method for

separating anthocyanins, the major benefits of this method were the scalability and suitability of the system for industrial application.

PVPP is relatively inexpensive and has a well-established use in beverage manufacture. Furthermore, flash chromatography systems are capable of kilogram scale separations, and the mobile phase used in these separations is food safe and therefore requires no adaptation. Flash chromatography with PVPP could be used to decrease the concentration of other products present in extracts and so increase the relative concentration of specific anthocyanins, making the extract a more effective food colourant.

5.3 How decoration influences colour and stability of ternatins

Blue colouration of anthocyanins is typically conferred by the neutral and anionic base of anthocyanins although these are generally less energetically favourable compared to the colourless hemiketal. It is well established that copigmentation interactions are the primary interactions which stabilise these coloured forms by preventing nucleophilic addition. This is driven predominantly through steric hinderance, occluding water from the electrophilic C2 position. Aromatically acylated anthocyanins, such as ternatins form intramolecular interactions between the aromatic rings of the anthocyanin, and the decorating aromatic acyl group. This stacking contributes to the overall stability of the neutral and anionic bases.

Due to low purity of ternatins B4, B3 and B2, only the study of six isolates using NMR was possible. However, of the isolates studied long range NOE signals (**Figure 2.13**) between the delphinidin C4 proton, and the 2 & 6 protons present on the innermost 3' and 5' aromatic acyl decorations, C_I and C_{II} were observed. The presence of these signals indicated that the protons were in close proximity to each other, suggesting that the ternatins adopted a sandwich-stacking configuration. This stacking over the chromophore would stabilise the coloured forms, shifting the equilibrium away from the formation of colourless hemiketals. This was further evident when the chemical shifts of the delphinidin C4-H in the core of each ternatin were observed to increase with increasing decoration and were all more shielded than those of nonacylated delphinidin-3-O-glucoside (**Table 2.2**).

5.4 Stability

Due to anthocyanins being weak diacids, their use as colourants faces challenges due to their propensity to change colour in response to pH. For anthocyanins where the neutral and anionic bases are stabilised (such as those involved in copigmentation interactions), these colours are generally red at acidic pH, red/purple at mildly acidic pH, and purple/blue at near neutral pH. Ternatins are unusual, in that they are capable of producing a blue liquid colour at mildly acidic pH.

The aromatic groups of acyl residues induces a diamagnetic anisotropy effect through space, where the ring currents of one aromatic group generate a magnetic field which influences the shielding of protons in close proximity (Pauling, 1936; Bertini, Luchinat and Parigi, 2002). Aromatic residues result in an increase in shielding (shift downfield) when the proton in question is located near the aromatic decoration, increasing in magnitude the as the proton becomes more centred under the decoration (Martin *et al.*, 2006). [1]H spectra of ternatin isolates showed no chemical shifts related to the C8-H, C2'-H or C6'-H. However, the chemical shift of C4-H (located within the anthocyanin C-ring) increased with increasing number of decorations from 8.44 ppm for A3, to 8.55 ppm for D1. Similar results were observed from the molecular dynamics simulations of the ternatins, with the 3' and 5' acyl sidechains producing the lowest energy configurations when the sidechains were closest to the C − ring. The glycosylation of D1 to form B1, and B1 to form A1 led to a decrease in chemical shift for this proton, with chemical shifts of 8.54 ppm and 8.55 ppm respectively. This indicated a weakening of the interaction.

Ternatin D1 was shown to be the most stable anthocyanin at room temperature and neutral pH, with a half-life of 99.7 days. This stability decreased upon glycosylation to 90.4 days for ternatin B1 and 77.6 days for ternatin A1. Interestingly, molecular dynamics simulations showed an increased radial distribution of water for D1 compared to B1 or A1, which suggested that steric hindrance of water may not be the only factor influencing stability. The stacked configurations exist in equilibrium between the open and closed forms. Therefore, it may be the case that ternatin D1 has a high probability of being in the ring closed configuration, thus protecting its electrophilic centres. Unfortunately, molecular dynamics simulations did not capture any of these ring opening events, so preventing the calculation of the probability of

these configurations. Further simulations with longer durations should be carried out to determine whether the probability of complex unfolding, and how this contributes to stability.

The increasing chemical shift of C4-H could be due to tighter association between the acyl residue and the ternatin C – ring, which would serve to increase the stability of the anthocyanin.

At elevated temperatures of 55 °C however, these stability trends reverse, with the A series ternatins having greater stability than the D series. At these temperatures, sufficient energy is provided to overcome the π-π the stacked configuration, increasing the probability of the ring-open configuration. At higher temperatures, the neutral and anionic bases have greater susceptibility to auto-oxidation. My results also suggest that the free hydroxyl group of *p*-coumarate may be involved in promoting the degradation of the anthocyanin, although the mechanism is unclear.

5.5 Importance of B – ring dihedral angle for colour

There are two types of decoration, sugars and aromatic acyl groups, on the side chains of the ternatins decorated on the 3' and 5' hydroxyl groups of the B-ring. Generally, glucose units decrease the overall stability and acyl units increase stability. UV-VIS spectroscopy of isolated ternatins showed an increase in stability and bathochromic shifts as the number of decorations increase.

Anthocyanins decorated with aromatic acyl groups can form intramolecular copigmentation interactions, which stabilise the coloured forms. Sequential acylation of the 3' and 5' side chains increased the bathochromic shift and the λ_{max} of the anthocyanin (**Table 3.2**) at neutral pH. The empirical data presented in figure 3.6 show a distinctive UV-VIS absorption peak present at pH 3. At this pH, the equilibrium favours the flavylium cation, with a minor presence of the neutral species. In the case of ternatin A1, this absorption peak is present at a wavelength of 622 nm, shifted much further towards blue than the flavylium cation peak at 553 nm and the neutral species at 575 nm. Similar CT peaks in the blue range were seen in the absorbance spectra for other ternatins (**Appendix 3.1 – 3.8**).

This absorption peak was therefore likely the result of charge transfer interactions, involving the rearrangement of electron density within the molecule to a lower energy state. Upon absorption of incident light, the electrons have a lower energy pathway between the highest occupied molecular orbital and the new energy state, thus decreasing the energy gap and hence the energy of the light required for excitation.

Anthocyanins which feature stacking type interactions both above and below the chromophore (sandwich-type stacking) display a well-defined charge transfer peak, present at pH values which would otherwise correspond to the neutral anthocyanin species.

Generally, glycosylation increased the λ_{max} of the anthocyanin, but leads to a decreased stability, as evidenced by systematic decreases in half-life. As glucose residues have multiple hydroxyl groups, the observed increases in λ_{max} may have been the result of the formation of additional hydrogen bonds between the sidechains and the anthocyanin chromophore. This would allow for a tighter association between the chromophore and the intermolecular copigment, thus more strongly influencing the electron polarisation of the molecule. Conversely, due to the presence of these hydroxyl groups, the hydrophilic forces would increase and stability decrease. Molecular dynamics simulations showed that the dihedral angle is dependent on the decorations of the ternatin. Although colour modulation from anthocyanin – copigment interactions is primarily considered in the context of π-π interactions (and thus overlapping molecular orbitals), this is not a comprehensive description of the relationship between the molecular geometry and colour properties. This was primarily evident in the double glycosylation of D1, forming B1 and subsequently A1, with an accompanying bathochromic shift of 1 nm and 2 nm respectively at pH7 (**Figure 3.8**). Although these glycosylations increased the λ_{max} of the anthocyanin, the stability at pH 7 was decreased by 9.3 and 12.8 days respectively (**Table 3.3**).

Rotation of the B – ring influenced the distance at which acyl residues stack over the chromophore. When one sidechain was very long, and the other was capped with a *p*-coumaroyl group there was a lower λ_{max} due to weakening of the sidechain interactions asymmetrically. When glucose is added to this *p*-coumaroyl group,

there is a small increase in λ_{max}, but this is generally accompanied by a decrease in stability as illustrated by comparing B2 with A2 (figure 3.6)

Similarly, when the *p*-coumarate residue on the 5' sidechain of ternatin D2 was glycosylated, there was an accompanying 2 nm increase in λ_{max}, as the additional hydrophilic forces acted to weaken the interactions with the 5' sidechain. This likely allowed more favourable interaction with the two acyl residues present on the 3' sidechain, thus inducing a larger torsion force on the B – ring dihedral angle.

Where one sidechain was decorated asymmetrically, such as in the case of the regioisomers B3 and B2, glycosylation of the shorter side chain appeared to increase the λ_{max} but decrease the stability of the ternatin (compare D2 to B3 and B2 to A2 in figure 3.6). This suggested that the interaction forces which hold the complex together were weakened, likely due to the additional glucose residues providing additional hydrophilic interactions with the aqueous solvent, working to pull the sidechains away from the chromophore.

In recent literature, it was shown that the B – ring dihedral angle of acylated cyanidin anthocyanins contributed more to the bathochromic shift than the electronic interaction between the chromophore and intramolecular copigment (Denish *et al.*, 2021). By calculating the electronic transition energy for delphinidin-3-O-glucoside rotamers, the relationship between B – ring dihedral angle, and λ_{max} (**Figure 4.10**) has been demonstrated. Replica exchange molecular dynamics showed an increased B – ring torsion angle of B3 compared to B2, which likely corresponded to the observed bathochromic shifts (B3 = 620 at pH7, B2 = 618 nm at pH7; figure 3.6). Tracking the dihedral angle of all ternatins showed no global correlation between dihedral angle and λ_{max} but my replica exchange simulations may not have been long enough for full exploration of the conformational space, and therefore future work should include further simulations for longer durations, to ensure this relationship is explored, fully.

5.6 Concluding remarks

Anthocyanins have huge potential to be used as replacements for the synthetic blue pigments currently used for colouring food products, but use of anthocyanins as

replacement pigments for synthetic blue No 1 is still limited by the lack of stability, sensitivity to pH change and temperature.

By understanding how these features change with increasing numbers of decorations, and how these alter the structural properties of anthocyanins, such as B – ring dihedral angle and copigment orientation, the understanding of the underlying mechanisms responsible for the generation of stable blue colours has improved. Highly decorated ternatins show remarkable stability at neutral pH as well as blue colours at pH 4 - 8, similar to that of indigo carmine.

The results in this thesis also offer insight into the mechanisms responsible for stabilising and modulating the colour of ternatins, which could be used for beneficial applications, for example, by increasing the concentration of D ternatins by flash chromatography with PVPP to give a stronger blues with greater stability than the unfractionated extract from flowers.

The findings presented in this thesis provide valuable insights into the mechanisms that contribute to the stabilisation and modulation of colour in ternatins, which could promote their replacement of synthetic blue colourants.

References

Alluis, B. and Dangles, O., (2001) 'Quercetin (= 2-(3, 4-dihydroxyphenyl)-3, 5, 7-trihydroxy-4H-1-benzopyran-4-one) glycosides and sulfates: chemical synthesis, complexation, and antioxidant properties'. *Helvetica Chimica Acta*, 84(5), pp.1133-1156.

Andersen, Ø.M. and Jordheim, M. (2006) *The anthocyanins*. 2nd Edition. Edited by Oyvind.M. Andersen and K.R. Markham. Boca Raton, FL: CRC Press.

Andersen, Ø.M. and Jordheim, M. (2013) 'Anthocyanins in Health and Disease - Basic Anthocyanin Chemistry and Dietary Sources', in T.C. Wallace and M.M. Giusti (eds) *Anthocyanins in Health and Disease - Basic Anthocyanin Chemistry and Dietary Sources*. CRC Press, pp. 30–107.

Barnett, J.R., Miller, S. and Pearce, E. (2006) 'Colour and art: A brief history of pigments', *Optics & Laser Technology*, 38(4–6), pp. 445–453.

Basílio, N. and Pina, F. (2016) 'Chemistry and photochemistry of anthocyanins and related compounds: A thermodynamic and kinetic approach', *Molecules*, 21(11), pp. 1502.

Bertini, I., Luchinat, C. and Parigi, G. (2002) 'Magnetic susceptibility in paramagnetic NMR', *Progress in Nuclear Magnetic Resonance Spectroscopy*, 40(3), pp. 249–273.

Bloom, M. and Geissman, T.A. (1973) 'Malonic acid: The acyl moiety of the Mimulus luteus anthocyanin', *Phytochemistry*, 12(8), pp. 2005–2006

Boss, P.K., Davies, C. and Robinson, S.P. (1996) 'Expression of anthocyanin biosynthesis pathway genes in red and white grapes', *Plant molecular biology*, 32(3), pp. 565–569.

Brouillard, R., Wigand, M.C., Dangles, O. and Cheminat, A., (1991) 'PH and solvent effects on the copigmentation reaction of malvin with polyphenols, purine and pyrimidine derivatives', *Journal of the Chemical Society, Perkin Transactions 2*, (8), pp. 1235–1241.

Brugliera, F., Tao, G.Q., Tems, U., Kalc, G., Mouradova, E., Price, K., Stevenson, K., Nakamura, N., Stacey, I., Katsumoto, Y. and Tanaka, Y., (2013) 'Violet/Blue Chrysanthemums—Metabolic Engineering of the Anthocyanin Biosynthetic Pathway Results in Novel Petal Colors', *Plant and Cell Physiology*, 54(10), pp. 1696–1710.

Buer, C.S., Muday, G.K. and Djordjevic, M.A. (2007) 'Flavonoids Are Differentially Taken Up and Transported Long Distances in Arabidopsis', *Plant Physiology*, 145(2), pp. 478–490.

Burbulis, I.E. and Winkel-Shirley, B. (1999) 'Interactions among enzymes of the Arabidopsis flavonoid biosynthetic pathway', *Proceedings of the National Academy of Sciences of the United States of America*, 96(22), pp. 12929–12934.

Bussi, G. (2014) 'Hamiltonian replica exchange in GROMACS: a flexible implementation', *Molecular Physics*, 112(3–4), pp. 379–384.

Castañeda-Ovando, A., de Lourdes Pacheco-Hernández, M., Páez-Hernández, M.E., Rodríguez, J.A. and Galán-Vidal, C.A., (2009) 'Chemical studies of anthocyanins: A review', *Food Chemistry*. Elsevier, pp. 859–871.

Chalker-Scott, L. (2018) 'How are Hydrangea flower colours determined'.

Chandran, J., Nisha, P., Singhal, R.S. and Pandit, A.B., (2014) 'Degradation of colour in beetroot (Beta vulgaris L.): A kinetics study', *Journal of Food Science and Technology*, 51(10), pp. 2678–2684.

Cheng, J., Wei, G., Zhou, H., Gu, C., Vimolmangkang, S., Liao, L. and Han, Y., (2014) 'Unraveling the mechanism underlying the glycosylation and methylation of anthocyanins in peach', *Plant physiology*, 166(2), pp. 1044–1058.

Dangles, O. and Brouillard, R. (1992) 'Polyphenol interactions. The copigmentation case: thermodynamic data from temperature variation and relaxation kinetics. Medium effect', *Canadian Journal of Chemistry*, 70(8), pp. 2174–2189.

Dangles, O. and Fenger, J.-A. (2018) 'The Chemical Reactivity of Anthocyanins and Its Consequences in Food Science and Nutrition', *Molecules*, 23(8), p. 1970.

Dangles, O., Saito, N. and Brouillard, R. (1993) 'Anthocyanin intramolecular copigment effect', *Phytochemistry*, 34(1), pp. 119–124.

Denish, P.R., Fenger, J.A., Powers, R., Sigurdson, G.T., Grisanti, L., Guggenheim, K.G., Laporte, S., Li, J., Kondo, T., Magistrato, A. and Moloney, M.P., (2021) 'Discovery of a natural cyan blue: A unique food-sourced anthocyanin could replace synthetic brilliant blue', *Science Advances*, 7(15).

Duncan, J. (2017) 'Cover Crop Options for Hot and Humid Areas', *National Center for Appropriate Technology – ATTRA*.

Eiro, M.J. and Heinonen, M. (2002) 'Anthocyanin color behavior and stability during storage: Effect of intermolecular copigmentation', *Journal of Agricultural and Food Chemistry*, 50(25), pp. 7461–7466.

Englman, R. and Jortner, J. (1970) 'The energy gap law for radiationless transitions in large molecules', *Molecular Physics*, 18(2), pp. 285–287.

Escribano-Bailón, M.T., Rivas-Gonzalo, J.C. and García-Estévez, I. (2018) 'Wine Color Evolution and Stability', in *Red Wine Technology*. Elsevier, pp. 195–205.

Falcone Ferreyra, M.L., Rius, S.P. and Casati, P. (2012) 'Flavonoids: biosynthesis, biological functions, and biotechnological applications', *Frontiers in Plant Science*, 3(SEP).

Fenger, J.A., Moloney, M., Robbins, R.J., Collins, T.M. and Dangles, O., (2019) 'The influence of acylation, metal binding and natural antioxidants on the thermal stability of red cabbage anthocyanins in neutral solution', *Food & Function*, 10(10), pp. 6740–6751.

Fernandes, A., Brás, N.F., Mateus, N. and de Freitas, V., (2014) 'Understanding the molecular mechanism of anthocyanin binding to pectin', *Langmuir*, 30(28), pp. 8516–8527.

Fernandes, A., Brás, N.F., Mateus, N. and de Freitas, V., (2015) 'A study of anthocyanin self-association by NMR spectroscopy', *New Journal of Chemistry*, 39(4), pp. 2602–2611.

Fu, X., Wu, Q., Wang, J., Chen, Y., Zhu, G. and Zhu, Z., (2021) 'Spectral characteristic, storage stability and antioxidant properties of anthocyanin extracts from flowers of butterfly pea (Clitoria ternatea l.)', *Molecules*, 26(22), p.7000.

Galland, S., Mora, N., Abert-Vian, M., Rakotomanomana, N. and Dangles, O., (2007) 'Chemical synthesis of hydroxycinnamic acid glucosides and evaluation of their ability to stabilize natural colors via anthocyanin copigmentation', *Journal of Agricultural and Food Chemistry*, 55(18), pp. 7573–7579.

Gomez, C., Terrier, N., Torregrosa, L., Vialet, S., Fournier-Level, A., Verries, C., Souquet, J.M., Mazauric, J.P., Klein, M., Cheynier, V. and Ageorges, A., (2009) 'Grapevine MATE-Type Proteins Act as Vacuolar H+-Dependent Acylated Anthocyanin Transporters', *Plant Physiology*, 150(1), pp. 402–415.

Gomez, C., Conejero, G., Torregrosa, L., Cheynier, V., Terrier, N. and Ageorges, A., (2011) 'In vivo grapevine anthocyanin transport involves vesicle-mediated trafficking and the contribution of anthoMATE transporters and GST', *Plant Journal*, 67(6), pp. 960–970.

González, M.A. (2011) 'Force fields and molecular dynamics simulations', *Collection SFN*, 12, pp. 169–200.

González-Manzano, S., Santos-Buelga, C., Dueñas, M., Rivas-Gonzalo, J.C. and Escribano-Bailón, T., (2008) 'Colour implications of self-association processes of wine anthocyanins', *European Food Research and Technology*, 226(3), pp. 483–490.

Grotewold, E., Chamberlin, M., Snook, M., Siame, B., Butler, L., Swenson, J., Maddock, S., St. Clair, G. and Bowen, B., (1998) 'Engineering Secondary Metabolism in Maize Cells by Ectopic Expression of Transcription Factors', *The Plant Cell*, 10(5), pp. 721–740.

Hansmann, U.H.E. (1997) 'Parallel tempering algorithm for conformational studies of biological molecules', *Chemical Physics Letters*, 281(1–3), pp. 140–150.

Hanwell, M.D., Curtis, D.E., Lonie, D.C., Vandermeersch, T., Zurek, E. and Hutchison, G.R., (2012) 'Avogadro: An advanced semantic chemical editor, visualization, and analysis platform', *Journal of Cheminformatics*, 4(8), pp. 1–17.

Harborne, J.B. (Jeffrey B.) (1988) 'The Flavonoids : advances in research since 1980', p. 621.

Harris, C.R., Millman, K.J., Van Der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J. and Kern, R., (2020) 'Array programming with NumPy', *Nature 2020 585:7825*, 585(7825), pp. 357–362.

Hayter, C.N. (1949) *Growing Blue Hydrangeas and Other Hydrangea Hints.*

He, F., Mu, L., Yan, G.L., Liang, N.N., Pan, Q.H., Wang, J., Reeves, M.J. and Duan, C.Q., (2010) 'Biosynthesis of anthocyanins and their regulation in colored grapes', *Molecules (Basel, Switzerland)*, 15(12), pp. 9057–9091.

He, J., Li, X., Silva, G.T., Quina, F.H. and Aquino, A.J., (2019) 'Quantum Chemical Investigation of the Intramolecular Copigmentation Complex of an Acylated Anthocyanin', *Article J. Braz. Chem. Soc*, 30(3), pp. 492–498.

Hollingsworth, S.A. and Dror, R.O. (2018) 'Molecular dynamics simulation for all', *Neuron*, 99(6), p. 1129.

Houghton, A., Appelhagen, I. and Martin, C. (2021) 'Natural blues: Structure meets function in anthocyanins', *Plants*, 10(4).

Hrazdina, G. (1970) 'Column Chromatographic Isolation of the Anthocyanidin-3,5-Diglucosides from Grapes', *Journal of Agricultural and Food Chemistry*, 18(2), pp. 243–245. Available at:

Hrazdina, G., Zobel, A.M. and Hoch, H.C. (1987) 'Biochemical, immunological, and immunocytochemical evidence for the association of chalcone synthase with endoplasmic reticulum membranes', *Proceedings of the National Academy of Sciences of the United States of America*, 84(24), pp. 8966–8970.

Huber, T., Torda, A.E. and van Gunsteren, W.F. (1994) 'Local elevation: A method for improving the searching properties of molecular dynamics simulation', *Journal of Computer-Aided Molecular Design*, 8(6), pp. 695–708.

Hunter, C.A. and Sanders, J.K.M. (1990) 'The Nature of π-π Interactions', *Journal of the American Chemical Society*, 112(14), pp. 5525–5534.

Ito, T., Aoki, D., Fukushima, K. and Yoshida, K., (2019) 'Direct mapping of hydrangea blue-complex in sepal tissues of Hydrangea macrophylla', *Scientific Reports*, 9(1), pp. 1–9.

Jakalian, A., Bush, B.L., Jack, D.B. and Bayly, C.I., (2000) 'Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method', *Journal of Computational Chemistry*, 21, pp. 132–146.

Jiang, W., Yin, Q., Wu, R., Zheng, G., Liu, J., Dixon, R.A. and Pang, Y., (2015) 'Role of a chalcone isomerase-like protein in flavonoid biosynthesis in Arabidopsis thaliana', *Journal of experimental botany*, 66(22), pp. 7165–7179.

Kallam, K., Appelhagen, I., Luo, J., Albert, N., Zhang, H., Deroles, S., Hill, L., Findlay, K., Andersen, Ø.M., Davies, K. and Martin, C., (2017) 'Aromatic Decoration Determines the Formation of Anthocyanic Vacuolar Inclusions', *Current Biology*, 27(7), pp. 945–957.

Kang, J., Park, J., Choi, H., Burla, B., Kretzschmar, T., Lee, Y. and Martinoia, E., (2011) 'Plant ABC Transporters', *The Arabidopsis Book / American Society of Plant Biologists*, 9, p. e0153.

Kazuma, K., Noda, N. and Suzuki, M. (2003) 'Flavonoid composition related to petal color in different lines of Clitoria ternatea', *Phytochemistry*, 64(6), pp. 1133–1139.

Kogawa, K., Kato, N., Kazuma, K., Noda, N. and Suzuki, M., (2007) 'Purification and characterization of UDP-glucose: anthocyanin 3',5'-O-glucosyltransferase from Clitoria ternatea', *Planta*, 226(6), pp. 1501–1509.

Lakshan, S.A.T., Pathirana, C.K., Jayanath, N.Y., Abeysekara, W.P.K.M. and Abeysekara, W.K.S.M., (2020) 'Antioxidant and selected chemical properties of the flowers of three different varieties of Butterfly Pea (Clitoria ternatea L.)', *Ceylon Journal of Science*, 49(2), pp.195-201.

Lambert, S.G., Asenstorfer, R.E., Williamson, N.M., Iland, P.G. and Jones, G.P., (2011) 'Copigmentation between malvidin-3-glucoside and some wine constituents and its importance to colour expression in red wine', *Food Chemistry*, 125(1), pp. 106–115.

Lee, D.W. (2007) 'Nature's palette : the science of plant color', p. 409.

Li, Y., Prejanò, M., Toscano, M. and Russo, N., (2018) 'Oenin and Quercetin Copigmentation: Highlights From Density Functional Theory', *Frontiers in Chemistry*, 6(JUN), p. 245.

Lin, Y., Irani, N.G. and Grotewold, E. (2003) 'Sub-cellular trafficking of phytochemicals explored using auto-fluorescent compounds in maize cells', *BMC Plant Biology*, 3(1), pp. 1–12.

Lin-Wang, K., McGhie, T.K., Wang, M., Liu, Y., Warren, B., Storey, R., Espley, R.V. and Allan, A.C., (2014) 'Engineering the anthocyanin regulatory complex of strawberry (Fragaria vesca)', *Frontiers in Plant Science*, 5(NOV), p. 651.

Loche, P., Steinbrunner, P., Friedowitz, S., Netz, R.R. and Bonthuis, D.J., (2021) 'Transferable Ion Force Fields in Water from a Simultaneous Optimization of Ion Solvation and Ion-Ion Interaction', *Journal of Physical Chemistry B*, 125(30), pp. 8581–8587..

Luo, J., Nishiyama, Y., Fuell, C., Taguchi, G., Elliott, K., Hill, L., Tanaka, Y., Kitayama, M., Yamazaki, M., Bailey, P. and Parr, A., (2007) 'Convergent evolution in the BAHD family of acyl transferases: Identification and characterization of anthocyanin acyl transferases from Arabidopsis thaliana', *Plant Journal*, 50(4), pp. 678–695.

Luo, X.E., Wang, R., Wang, J., Li, Y., Luo, H., Chen, S., Zeng, X.A. and Han, Z., (2022) 'Acylation of Anthocyanins and Their Applications in the Food Industry: Mechanisms and Recent Research Advances', *Foods 2022, Vol. 11, Page 2166*, 11(14), p. 2166.

Malien-Aubert, C., Dangles, O. and Amiot, M.J. (2001) 'Color stability of commercial anthocyanin-based extracts in relation to the phenolic composition. Protective effects by intra- and intermolecular copigmentation', *Journal of Agricultural and Food Chemistry*, 49(1), pp. 170–176.

Mao, Y., Head-Gordon, M. and Shao, Y. (2018) 'Unraveling substituent effects on frontier orbitals of conjugated molecules using an absolutely localized molecular orbital based analysis', *Chemical Science*, 9(45), pp. 8598–8607.

Marinova, K., Pourcel, L., Weder, B., Schwarz, M., Barron, D., Routaboul, J.M., Debeaujon, I. and Klein, M., (2007) 'The Arabidopsis MATE Transporter TT12 Acts as a Vacuolar Flavonoid/H+-Antiporter Active in Proanthocyanidin-Accumulating Cells of the Seed Coat', *The Plant Cell*, 19(6), pp. 2023–2038.

Mark, P. and Nilsson, L. (2001) 'Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K', *Journal of Physical Chemistry A*, 105(43), pp. 9954–9960.

Markham, K.R., Gould, K.S., Winefield, C.S., Mitchell, K.A., Bloor, S.J. and Boase, M.R., (2000) 'Anthocyanic vacuolar inclusions - Their nature and significance in flower colouration', *Phytochemistry*, 55(4), pp. 327–336.

Markham, K.R., Mitchell, K.A. and Boase, M.R. (1997) 'Malvidin-3-O-glucoside-5-O-(6-acetylglucoside) and its colour manifestation in "Johnson's Blue" and other "blue" geraniums', *Phytochemistry (United Kingdom)*.

Marković, J.M.D., Petranović, N.A. and Baranac, J.M. (2005) 'The copigmentation effect of sinapic acid on malvin: a spectroscopic investigation on colour enhancement', *Journal of Photochemistry and Photobiology B: Biology*, 78(3), pp. 223–228.

Martin, N.H., Loveless, D.M., Main, K.L. and Wade, D.C., (2006) 'Computation of through-space NMR shielding effects by small-ring aromatic and antiaromatic hydrocarbons', *Journal of molecular graphics & modelling*, 25(4), pp. 389–395.

Martinoia, E., Meyer, S., De Angeli, A. and Nagy, R., (2012) 'Vacuolar transporters in their physiological context', *Annual review of plant biology*, 63, pp. 183–213.

Mateus, N., Silva, A.M., Santos-Buelga, C., Rivas-Gonzalo, J.C. and de Freitas, V., (2002) 'Identification of Anthocyanin-Flavanol Pigments in Red Wines by NMR and Mass Spectrometry', *Journal of Agricultural and Food Chemistry*, 50(7), pp. 2110–2116.

Mckinney, W. (2010) 'Data Structures for Statistical Computing in Python', *SciPy*, 445(1), pp. 51-56.

Meier, H., Stalmach, U. and Kolshorn, H. (1997) 'Effective conjugation length and UV/vis spectra of oligomers', *Acta Polymerica*, 48(9), pp. 379–384.

Mendoza, J., Basílio, N., Pina, F., Kondo, T. and Yoshida, K., (2018) 'Rationalizing the Color in Heavenly Blue Anthocyanin: A Complete Kinetic and Thermodynamic Study', *Journal of Physical Chemistry B*, 122(19), pp. 4982–4992.

Mendoza, J., Oliveira, J., Araújo, P., Basílio, N., Teixeira, N., Brás, N.F., Pina, F., Yoshida, K. and de Freitas, V., (2020) 'The peculiarity of malvidin 3-O-(6-O-p-coumaroyl) glucoside aggregation. Intra and intermolecular interactions', *Dyes and Pigments*, 180, p. 108382.

Michl, J. and Thulstrup, E.W. (1976) 'Why is azulene blue and anthracene white? a simple mo picture', *Tetrahedron*, 32(2), pp. 205–209.

Molcanov, K. and Kojić-Prodić, B. (2019) 'Towards understanding π-stacking interactions between non-Aromatic rings', *IUCrJ*. International Union of Crystallography, pp. 156–166.

Moloney, M., Robbins, R.J., Collins, T.M., Kondo, T., Yoshida, K. and Dangles, O., (2018) 'Red cabbage anthocyanins: The influence of D-glucose acylation by hydroxycinnamic acids on their structural transformations in acidic to mildly alkaline conditions and on the resulting color', *Dyes and Pigments*, 158, pp. 342–352.

Moyroud, E., Wenzel, T., Middleton, R., Rudall, P.J., Banks, H., Reed, A., Mellers, G., Killoran, P., Westwood, M.M., Steiner, U. and Vignolini, S., (2017) 'Disorder in convergent floral nanostructures enhances signalling to bees', *Nature*, 550(7677), pp. 469–474.

Nakayama, T., Takahashi, S. and Waki, T. (2019) 'Formation of Flavonoid Metabolons: Functional Significance of Protein-Protein Interactions and Impact on Flavonoid Chemodiversity', *Frontiers in Plant Science*, 10, p. 821.

Neese, F. (2012) 'The ORCA program system', *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2(1), pp. 73–78.

Nerdal, W. and Andersen, Ø.M. (1992) 'Intermolecular aromatic acid association of an anthocyanin (petanin) evidenced by two-dimensional nuclear overhauser enhancement nuclear magnetic resonance experiments and distance geometry calculations', *Phytochemical Analysis*, 3(4), pp. 182–189.

Newsome, A.G., Culver, C.A. and Van Breemen, R.B. (2014) 'Nature's palette: The search for natural blue colorants', *Journal of Agricultural and Food Chemistry*, 62(28), pp. 6498–6511.

Noda, N., Aida, R., Kishimoto, S., Ishiguro, K., Fukuchi-Mizutani, M., Tanaka, Y. and Ohmiya, A., (2013) 'Genetic engineering of novel bluer-colored chrysanthemums produced by accumulation of delphinidin-based anthocyanins', *Plant and Cell Physiology*, 54(10), pp. 1684–1695.

Noda, N., Yoshioka, S., Kishimoto, S., Nakayama, M., Douzono, M., Tanaka, Y. and Aida, R., (2017) 'Generation of blue chrysanthemums by anthocyanin B-ring hydroxylation and glucosylation and its coloration mechanism', *Science Advances*, 3(7), p. e1602785.

Oguis, G.K., Gilding, E.K., Jackson, M.A. and Craik, D.J., (2019) 'Butterfly pea (Clitoria ternatea), a cyclotide-bearing plant with applications in agriculture and medicine', *Frontiers in Plant Science*, 10, p. 645.

Oh, C.M., Choi, J.Y., Bae, I.A., Kim, H.T., Hong, S.S., Noah, J.K. and Boo, Y.C., (2022) 'Identification of p-coumaric acid and ethyl p-coumarate as the main phenolic components of hemp (Cannabis sativa L.) roots' *Molecules*, 27(9), p.2781.

Pauling, L. (1936) 'The Diamagnetic Anisotropy of Aromatic Molecules', *The Journal of Chemical Physics*, 4(10), pp. 673–677.

Pecket, R.C. and Small, C.J. (1980) 'Occurrence, location and development of anthocyanoplasts', *Phytochemistry*, 19(12), pp. 2571–2576.

Petrussa, E., Braidot, E., Zancani, M., Peresson, C., Bertolini, A., Patui, S. and Vianello, A., (2013) 'Plant Flavonoids—Biosynthesis, Transport and Involvement in Stress Responses', *International Journal of Molecular Sciences*, 14(7), p. 14950.

Pourcel, L., Irani, N.G., Lu, Y., Riedl, K., Schwartz, S. and Grotewold, E., (2010) 'The formation of Anthocyanic Vacuolar Inclusions in Arabidopsis thaliana and implications for the sequestration of anthocyanin pigments', *Molecular plant*, 3(1), pp. 78–90.

Qi, R., Wei, G., Ma, B. and Nussinov, R., (2018) 'Replica Exchange Molecular Dynamics: A Practical Application Protocol with Solutions to Common Problems and a Peptide Aggregation and Self-Assembly Example', *Methods in molecular biology (Clifton, N.J.)*, 1777, p. 101.

Qian, B.J., Liu, J.H., Zhao, S.J., Cai, J.X. and Jing, P., (2017) 'The effects of gallic/ferulic/caffeic acids on colour intensification and anthocyanin stability', *Food chemistry*, 228, pp. 526–532.

Ratanapoompinyo, J., Nguyen, L.T., Devkota, L. and Shrestha, P., (2017) 'The effects of selected metal ions on the stability of red cabbage anthocyanins and total phenolic compounds subjected to encapsulation process', *Journal of Food Processing and Preservation*, 41(6), p. e13234.

Real, L.A. (1981) 'Uncertainty and Pollinator-Plant Interactions: The Foraging Behavior of Bees and Wasps on Artificial Flowers', *Ecology*, 62(1), pp. 20–26.

Rusishvili, M., Grisanti, L., Laporte, S., Micciarelli, M., Rosa, M., Robbins, R.J., Collins, T., Magistrato, A. and Baroni, S., (2019) 'Unraveling the molecular mechanisms of color expression in anthocyanins', *Physical Chemistry Chemical Physics*, 21(17), pp. 8757–8766.

Santiago, M.C.P.D.A., Gouvêa, A.C.M.S., Godoy, R.L.D.O., Borguini, R.G., Pacheco, S., Nogueira, R.I., Nascimento, L.D.S.D.M.D. and Freitas, S.P., (2014) 'Analytical standards production for the analysis of pomegranate anthocyanins by HPLC', *Brazilian Journal of Food Technology*, 17(1), pp. 51–57.

Sasaki, N. and Nakayama, T. (2015) 'Achievements and Perspectives in Biochemistry Concerning Anthocyanin Modification for Blue Flower Coloration', *Plant and Cell Physiology*, 56(1), pp. 28–40.

Saslowsky, D. and Winkel-Shirley, B. (2001) 'Localization of flavonoid enzymes in Arabidopsis roots', *The Plant journal : for cell and molecular biology*, 27(1), pp. 37–48.

Shi, M.-Z. and Xie, D.-Y. (2014) 'Biosynthesis and metabolic engineering of anthocyanins in Arabidopsis thaliana', *Recent patents on biotechnology*, 8(1), pp. 47–60.

Shiono, M., Matsugaki, N. and Takeda, K. (2005) 'Phytochemistry: structure of the blue cornflower pigment', *Nature*, 436(7052), p. 791.

Shiono, M., Matsugaki, N. and Takeda, K. (2008) 'Structure of commelinin, a blue complex pigment from the blue flowers of Commelina communis', *Proceedings of the Japan Academy Series B: Physical and Biological Sciences*, 84(10), pp. 452–456.

Sigurdson, G.T., Robbins, R.J., Collins, T.M. and Giusti, M.M., (2016) 'Evaluating the role of metal ions in the bathochromic and hyperchromic responses of cyanidin derivatives in acidic and alkaline pH', *Food Chemistry*, 208, pp. 26–34.

Silva, V.O., Freitas, A.A., Maçanita, A.L. and Quina, F.H., (2016) 'Chemistry and photochemistry of natural plant pigments: the anthocyanins', *Journal of Physical Organic Chemistry*, 29(11), pp. 594–599.

Singh, A., Vanga, S.K., Orsat, V. and Raghavan, V., (2018) 'Application of molecular dynamic simulation to study food proteins: A review', *Critical reviews in food science and nutrition*, 58(16), pp. 2779–2789.

Sousa Da Silva, A.W. and Vranken, W.F. (2012) 'ACPYPE - AnteChamber PYthon Parser interfacE', *BMC Research Notes*, 5(1), pp. 1–8.

Sparvoli, F., Martin, C., Scienza, A., Gavazzi, G. and Tonelli, C., (1994) 'Cloning and molecular analysis of structural genes involved in flavonoid and stilbene biosynthesis in grape (Vitis vinifera L.)', *Plant Molecular Biology*, 24(5), pp. 743–755.

Spence, C. (2015) 'On the psychological impact of food colour', *Flavour*, 4(1), p. 21.

Springob, K., Nakajima, J.I., Yamazaki, M. and Saito, K., (2003) 'Recent advances in the biosynthesis and accumulation of anthocyanins', *Natural product reports*, 20(3), pp. 288–303.

Sun, Y., Li, H. and Huang, J.R. (2012) 'Arabidopsis TT19 Functions as a Carrier to Transport Anthocyanin from the Cytosol to Tonoplasts', *Molecular Plant*, 5(2), pp. 387–400.

Takeda, K., Kariuda, M. and Itoi, H. (1985) 'Blueing of sepal colour of Hydrangea macrophylla', *Phytochemistry*, 24(10), pp. 2251–2254.

Tasaki, K., Watanabe, A., Nemoto, K., Takahashi, S., Goto, F., Sasaki, N., Hikage, T. and Nishihara, M., (2022) 'Identification of Candidate Genes Responsible for Flower Colour Intensity in Gentiana triflora', *Frontiers in Plant Science*, 13, p. 1989.

Terahara, N., Saito, N., Honda, T., Toki, K. and Osajima, Y., (1990) 'Structure of ternatin A1, the largest ternatin in the major blue anthocyanins from clitoria ternatea flowers', *Tetrahedron Letters*, 31(20), pp. 2921–2924.

Terahara, N., Oda, M., Matsui, T., Osajima, Y., Saito, N., Toki, K. and Honda, T., (1996) 'Five new anthocyanins, ternatins A3, B4, B3, B2, and D2, from Clitoria ternatea flowers', *Journal of natural products*, 59(2), pp. 139–144.

Terahara, N., Toki, K., Saito, N., Honda, T., Matsui, T. and Osajima, Y., (1998) 'Eight new anthocyanins, ternatins C1-C5 and D3 and preternatins A3 and C4 from young Clitoria ternatea flowers', *Journal of Natural Products*, 61(11), pp. 1361–1367.

Trouillas, P., Sancho-García, J.C., De Freitas, V., Gierschner, J., Otyepka, M. and Dangles, O., (2016) 'Stabilizing and Modulating Color by Copigmentation: Insights from Theory and Experiment', *Chemical Reviews*, 116(9), pp. 4937–4982.

Wen, Y., Chen, H., Zhou, X., Deng, Q., Zhao, C. and Gong, X., (2016) 'A polyamide resin based method for adsorption of anthocyanins from blackberries', *New Journal of Chemistry*, 40(4), pp. 3773–3780..

Winkel, B.S.J. (2004) 'Metabolic channeling in plants', *Annual review of plant biology*, 55, pp. 85–107.

Wrolstad, R.E. and Struthers, B.J. (1971) 'Polyvinylpyrrolidone column chromatography of strawberry, rhubarb, and raspberry anthocyanins', *Journal of Chromatography A*, 55(2), pp. 405–408.

Xie, L., Xie, J., Xu, Y. and Chen, W., (2020) 'Discovery of anthocyanins from cranberry extract as pancreatic lipase inhibitors using a combined approach of ultrafiltration, molecular simulation and spectroscopy', *Food & Function*, 11(10), pp. 8527–8536.

Xing, C., Chen, P. and Zhang, L. (2023) 'Computational insight into stability-enhanced systems of anthocyanin with protein/peptide', *Food Chemistry: Molecular Sciences*, 6, p. 100168.

Yamaguchi, T., Fukada-Tanaka, S., Inagaki, Y., Saito, N., Yonekura-Sakakibara, K., Tanaka, Y., Kusumi, T. and Iida, S., (2001) 'Genes encoding the vacuolar Na+/H+ exchanger and flower coloration', *Plant & cell physiology*, 42(5), pp. 451–461.

Yoshida, K., Kondo, T., Okazaki, Y. and Katou, K., (1995) 'Cause of blue petal colour', *Nature*. Nature Publishing Group, p. 291.

Yoshida, K., Mori, M. and Kondo, T. (2009) 'Blue flower color development by anthocyanins: From chemical structure to cell physiology', *Natural Product Reports*. Nat Prod Rep, pp. 884–915.

Zhang, Y., Butelli, E. and Martin, C. (2014) 'Engineering anthocyanin biosynthesis in plants', *Current Opinion in Plant Biology*, 19, pp. 81–90.

Zhao, C.L., Yu, Y.Q., Chen, Z.J., Wen, G.S., Wei, F.G., Zheng, Q., Wang, C.D. and Xiao, X.L., (2017) 'Stability-increasing effects of anthocyanin glycosyl acylation', *Food Chemistry*. Elsevier Ltd, pp. 119–128.

Zhao, J., Huhman, D., Shadle, G., He, X.Z., Sumner, L.W., Tang, Y. and Dixon, R.A., (2011) 'MATE2 Mediates Vacuolar Sequestration of Flavonoid Glycosides and Glycoside Malonates in Medicago truncatula', *The Plant Cell*, 23(4), pp. 1536–1555.

Zhao, J. and Dixon, R.A. (2010) 'The "ins" and "outs" of flavonoid transport', *Trends in plant science*, 15(2), pp. 72–80.
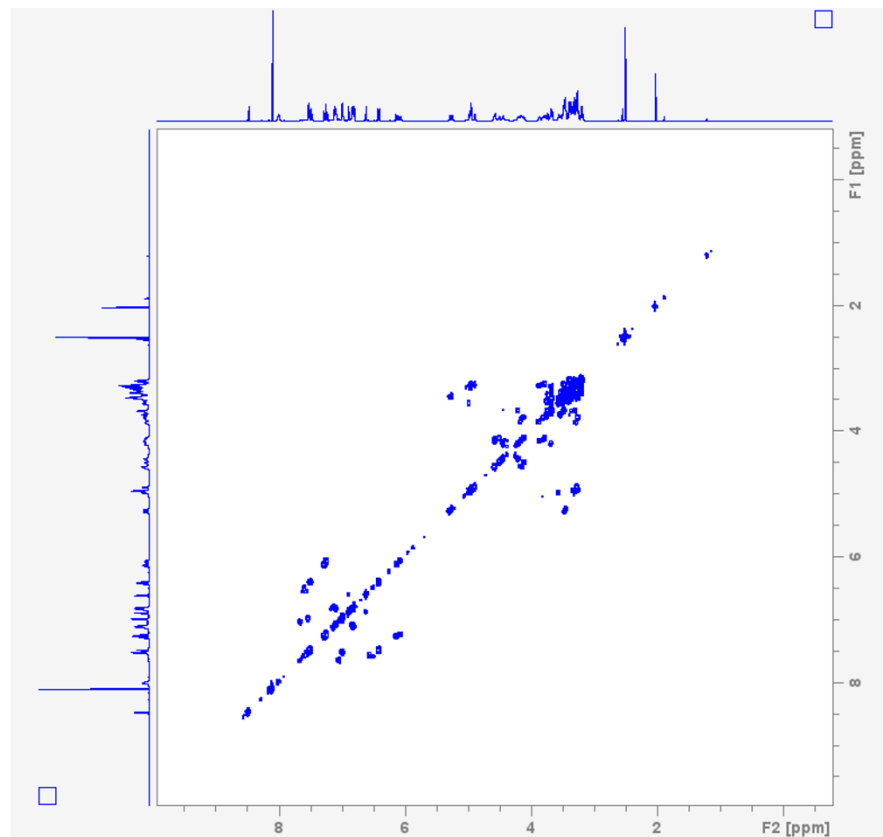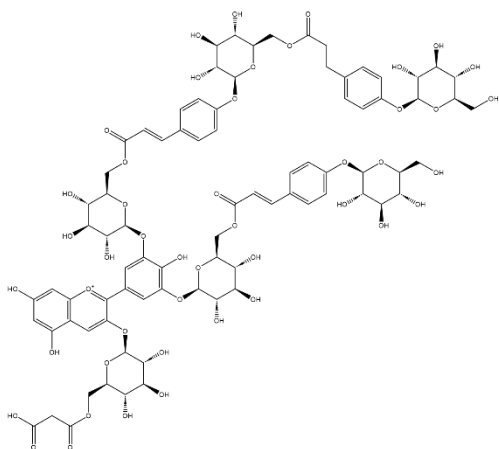
Appendix
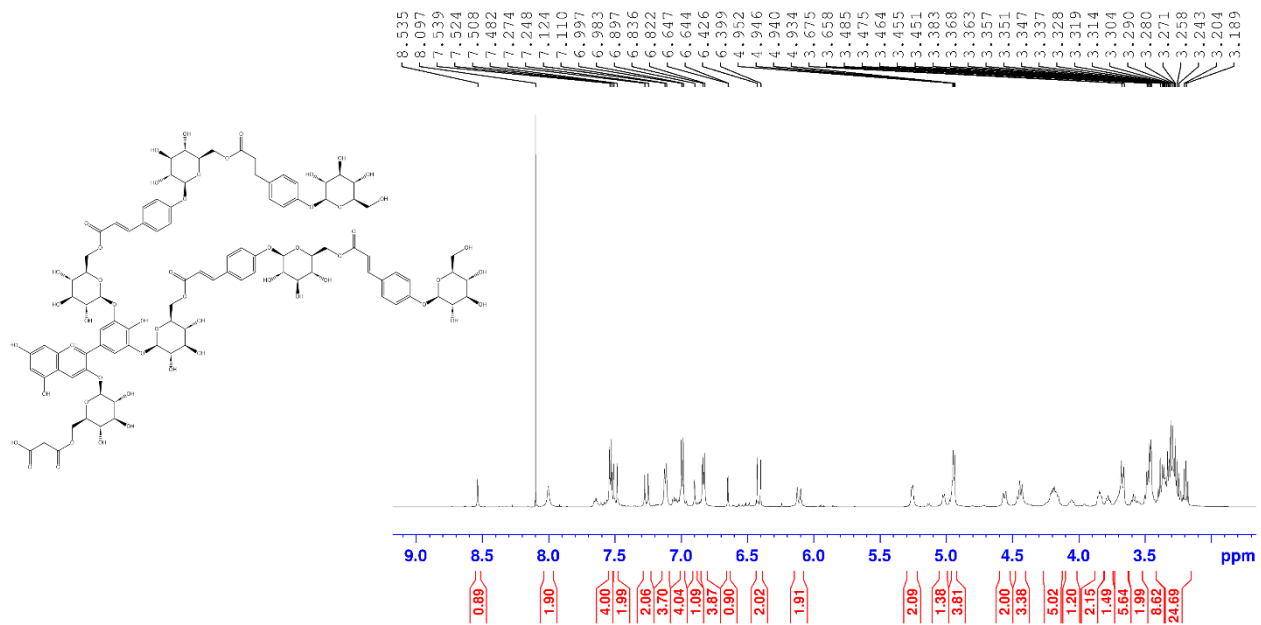
Appendix 2.1 (A) ^{1}H NMR spectra of ternatin A1 in DMSO-d_6/TFA-d, 9/1 (600 MHz). Number of Scans = 16, Bruker pulse program = zg30
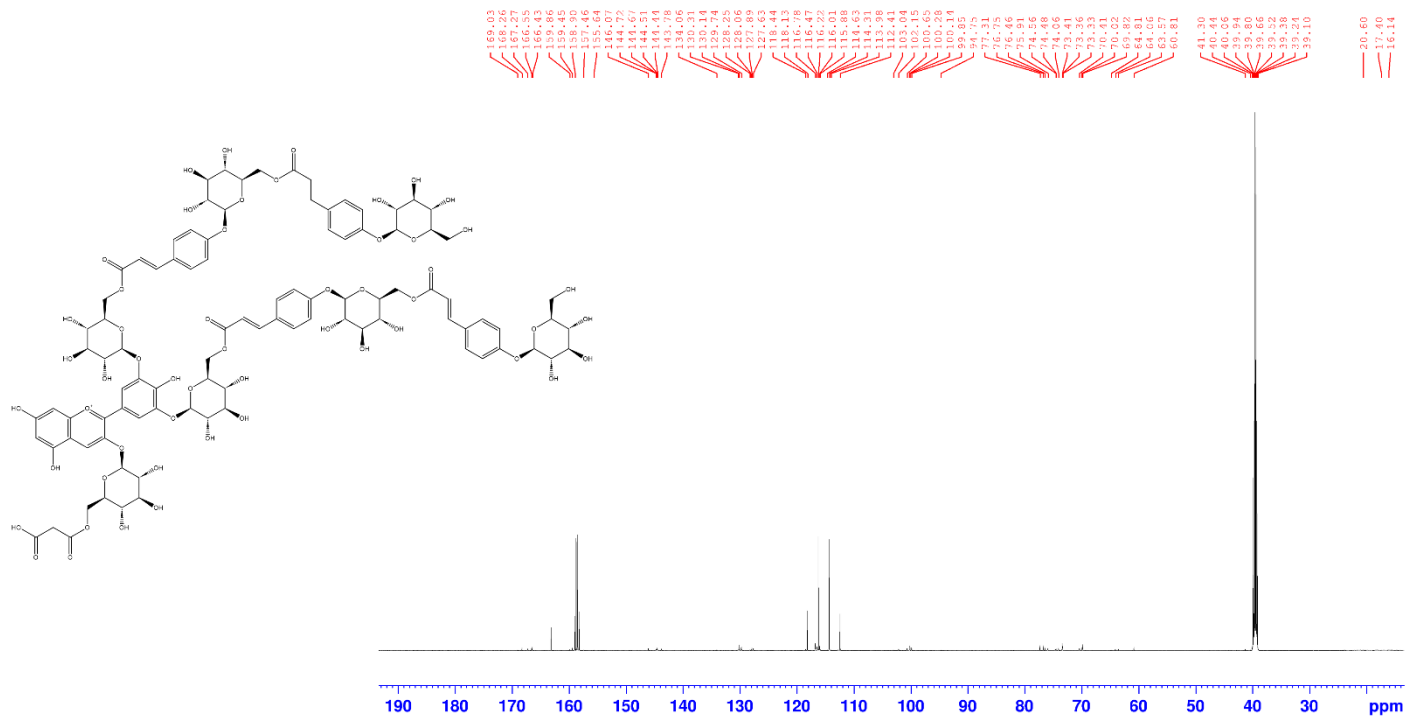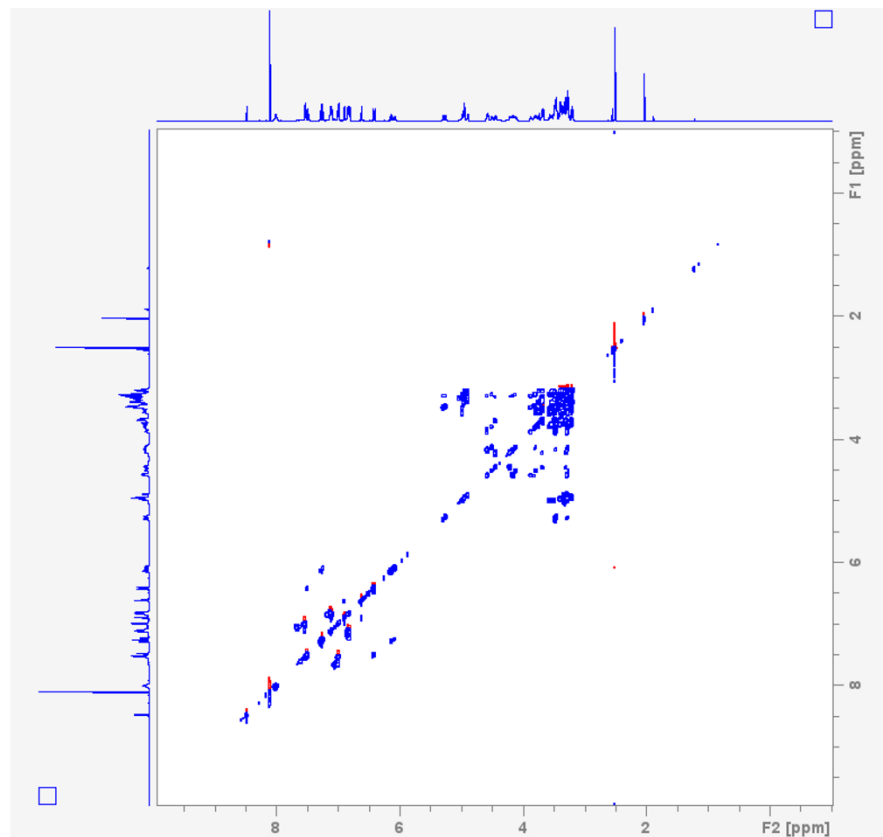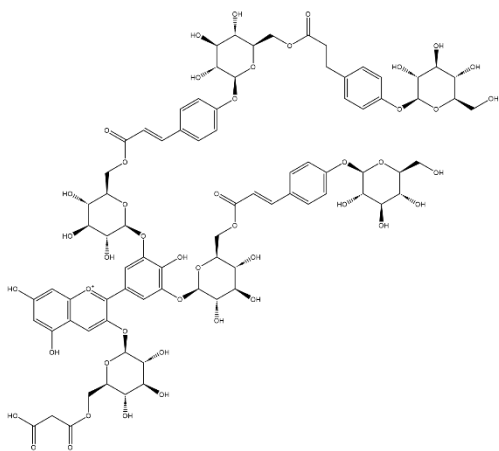
Appendix 2.1 (B) ^{13}C NMR spectra of ternatin A1 in DMSO-d_6/TFA-d, 9/1 (150 MHz). Number of Scans = 4,000, Bruker pulse program = zgpg30
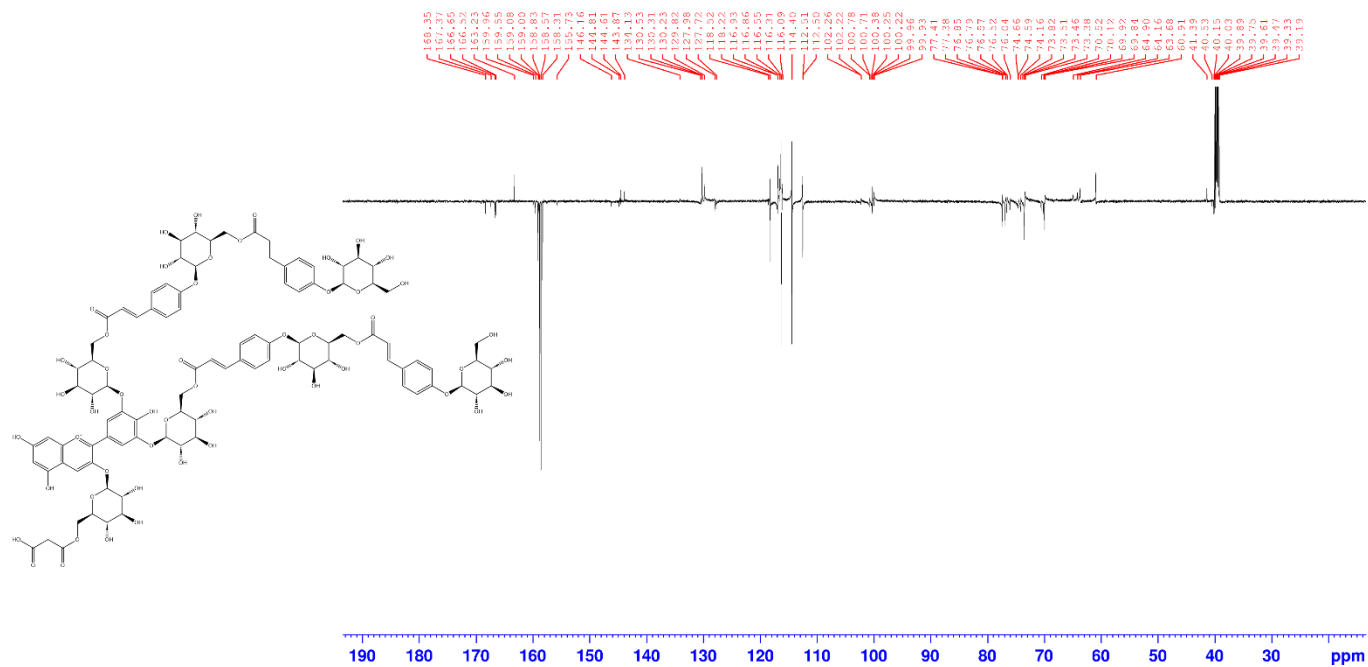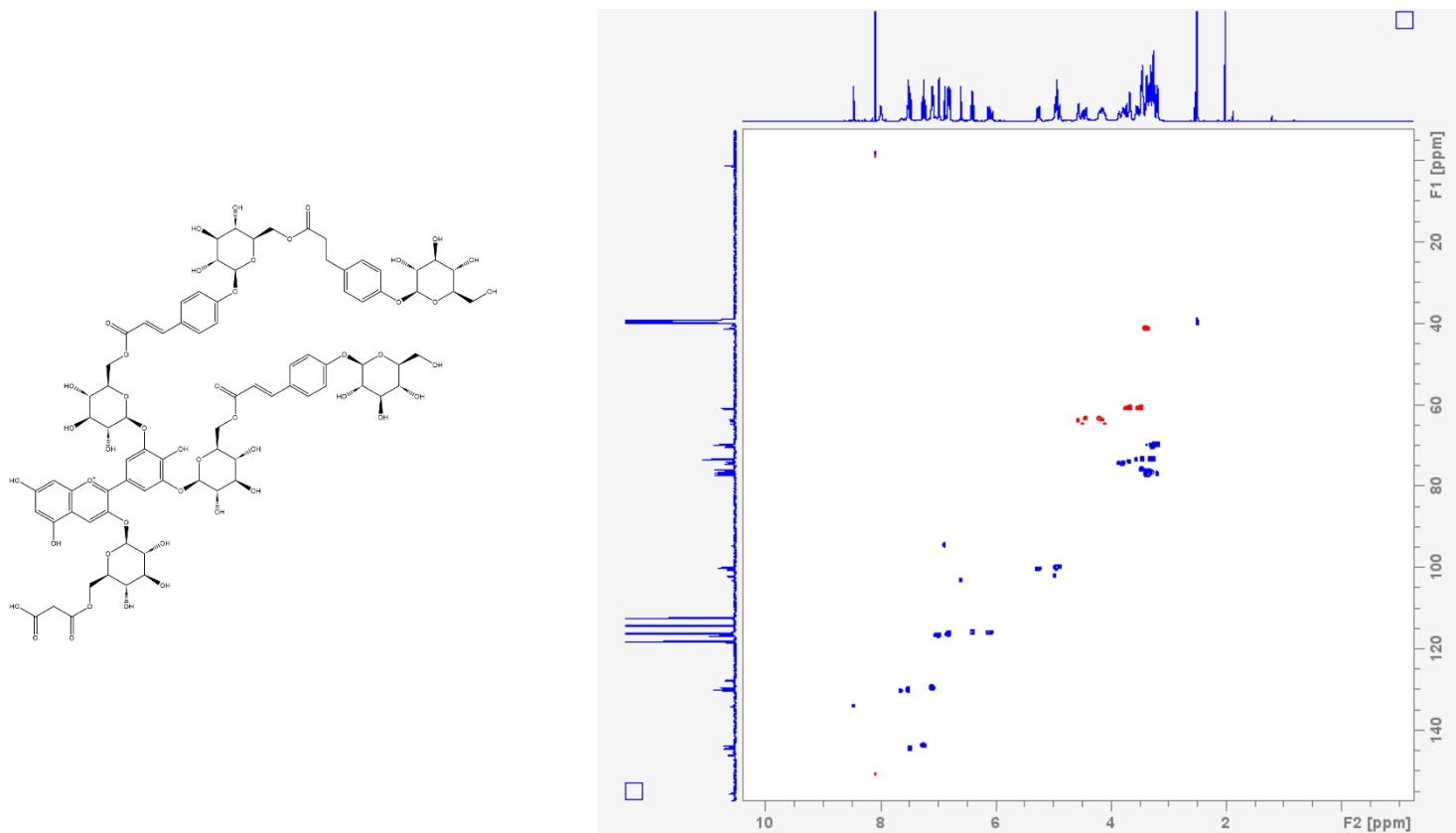
Appendix 2.1 (C) DEPTQ NMR spectra of ternatin A1 in DMSO-d_6/TFA-d, 9/1 (600 MHz). Number of Scans = 2,000, Bruker pulse program = deptqgpsp. Quaternary carbons (C) and methylene carbons (CH$_2$) are phased negative, while methine (CH) and methyl (CH$_3$) are phased positive.
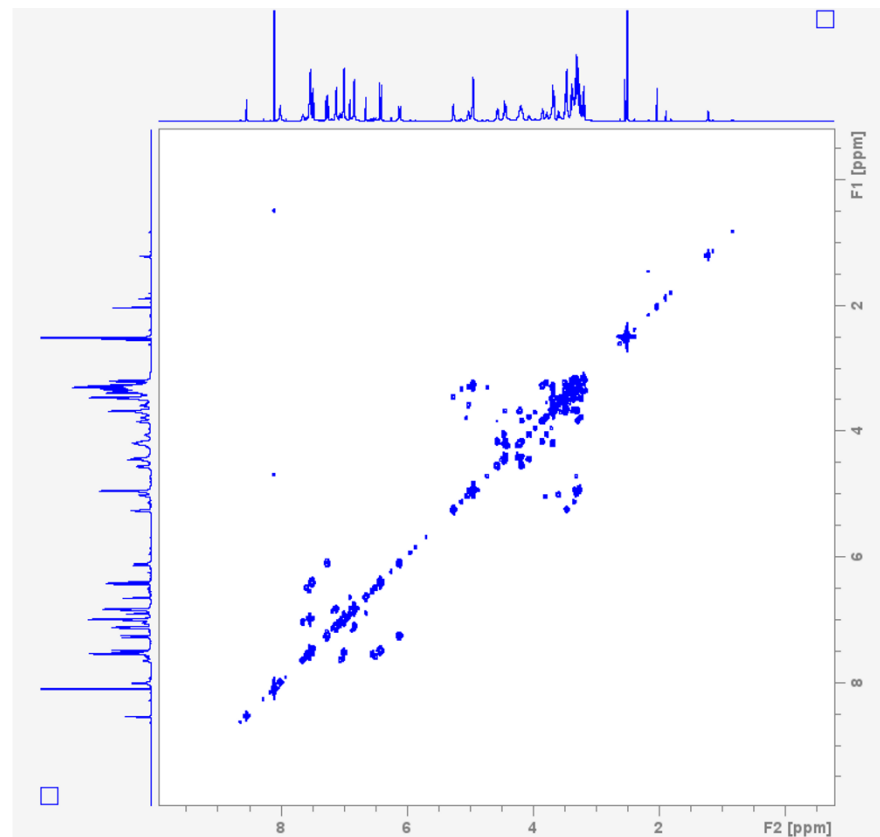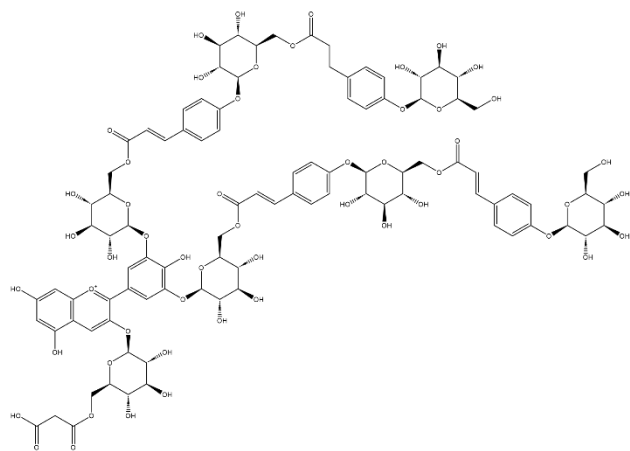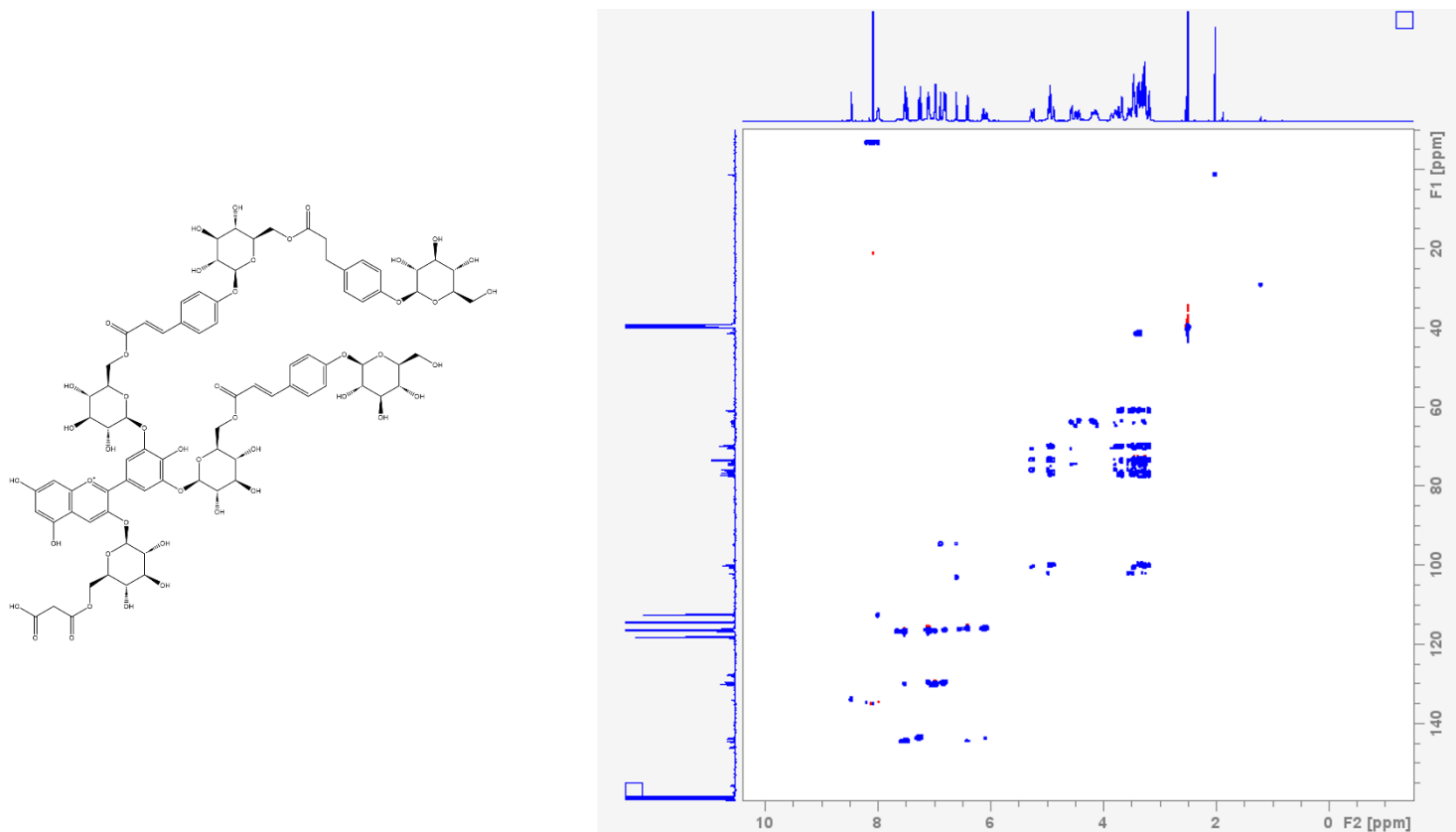
Appendix 2.1 (D) COSY (H-H correlation) NMR spectra of ternatin A1 in DMSO-d_6/TFA-d, 9/1 (600 MHz). Number of Scans = 4 , Bruker pulse program = cosygpppqf.ptype.
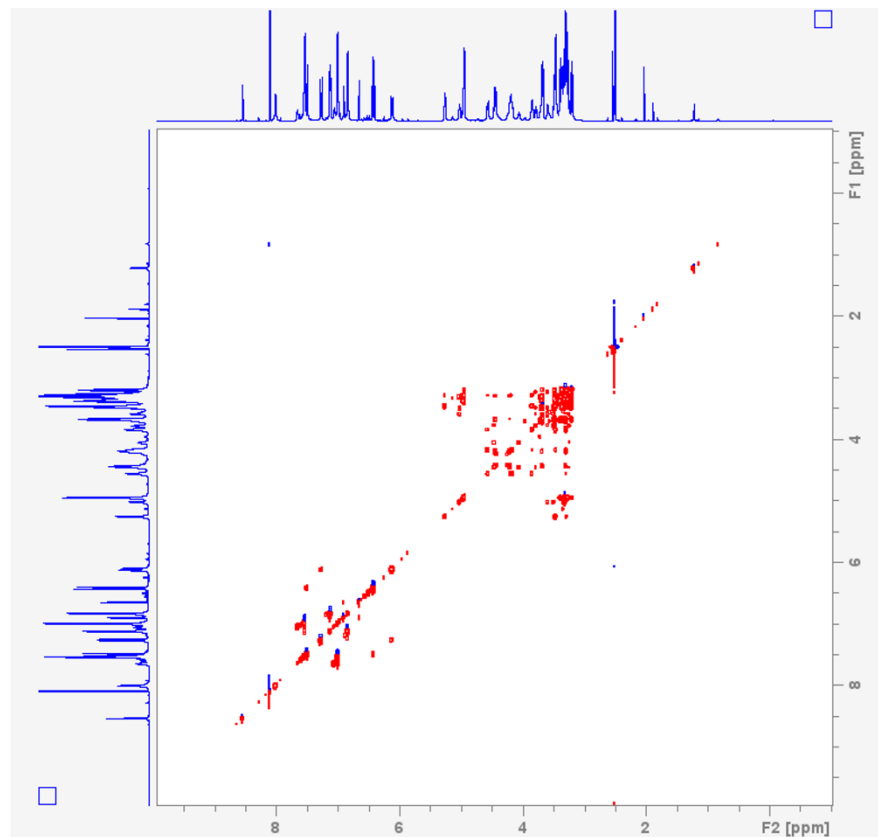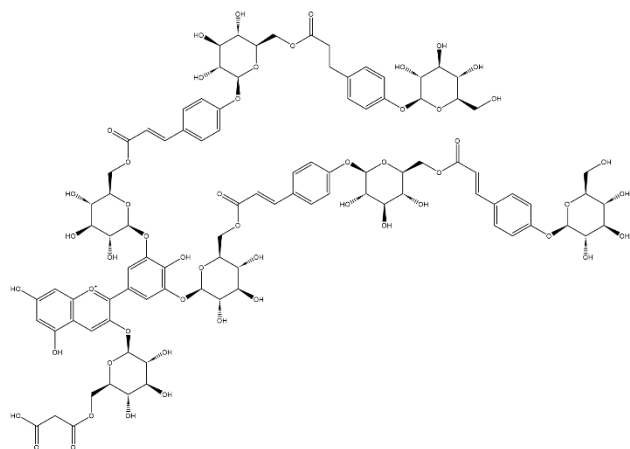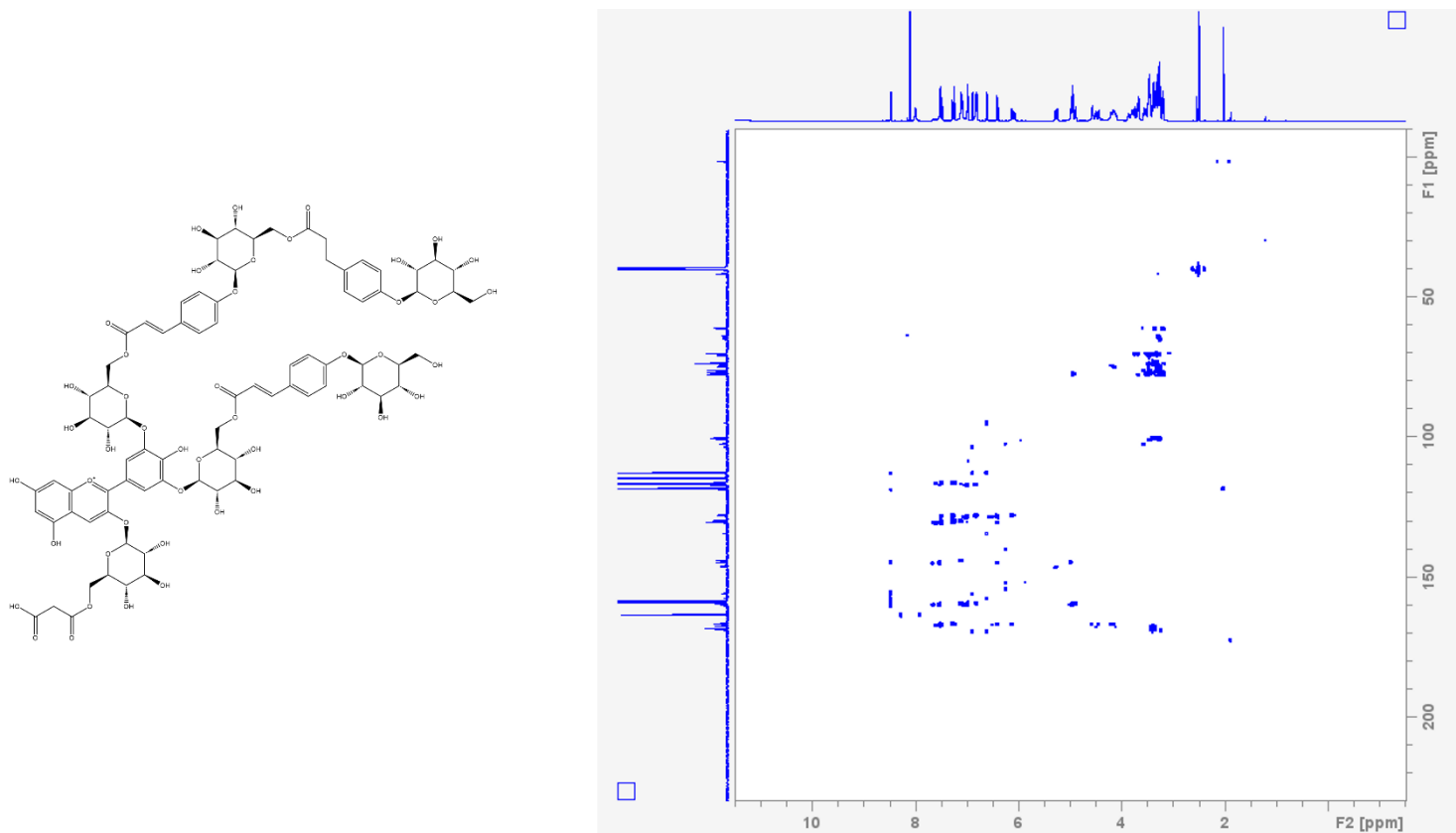
Appendix 2.1 (E) TOCSY (long range H-H correlation) NMR spectra of ternatin A1 in DMSO-d_6/TFA-d, 9/1 (600 MHz). Number of Scans = 4 , Bruker pulse program = dipsi2gpphzs
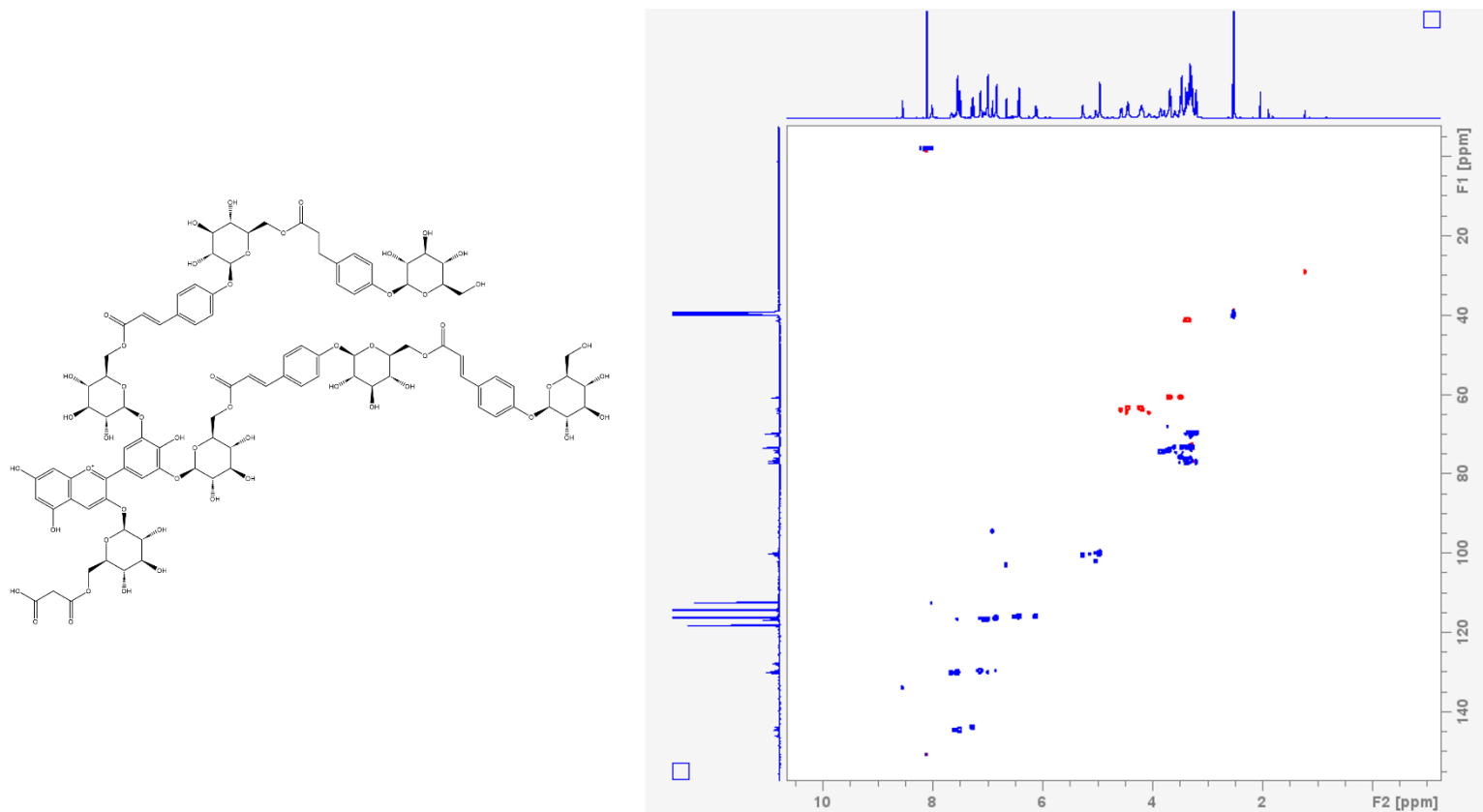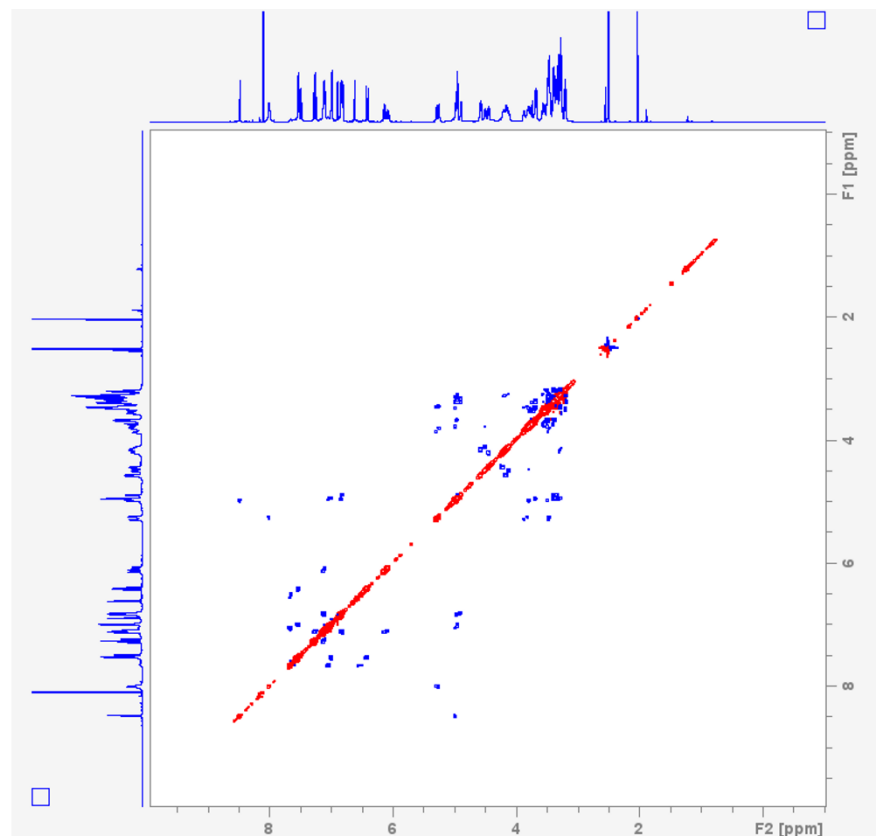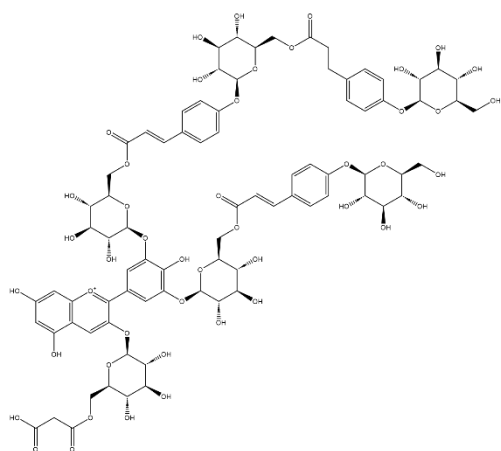
Appendix 2.1 (F) HSQC (H-C correlation) NMR spectra of ternatin A1 in DMSO-d_6/TFA-d (600 MHz for F2 frequency axis, and 150 MHz for F1 frequency axis). Number of Scans = 32 , Bruker pulse program = hsqcedetgpsisp2.3. Crosspeaks are phased such that methylene protons (CH$_2$) are negative (red), while methine (CH) and methyl (CH$_3$) are phased positive (blue).
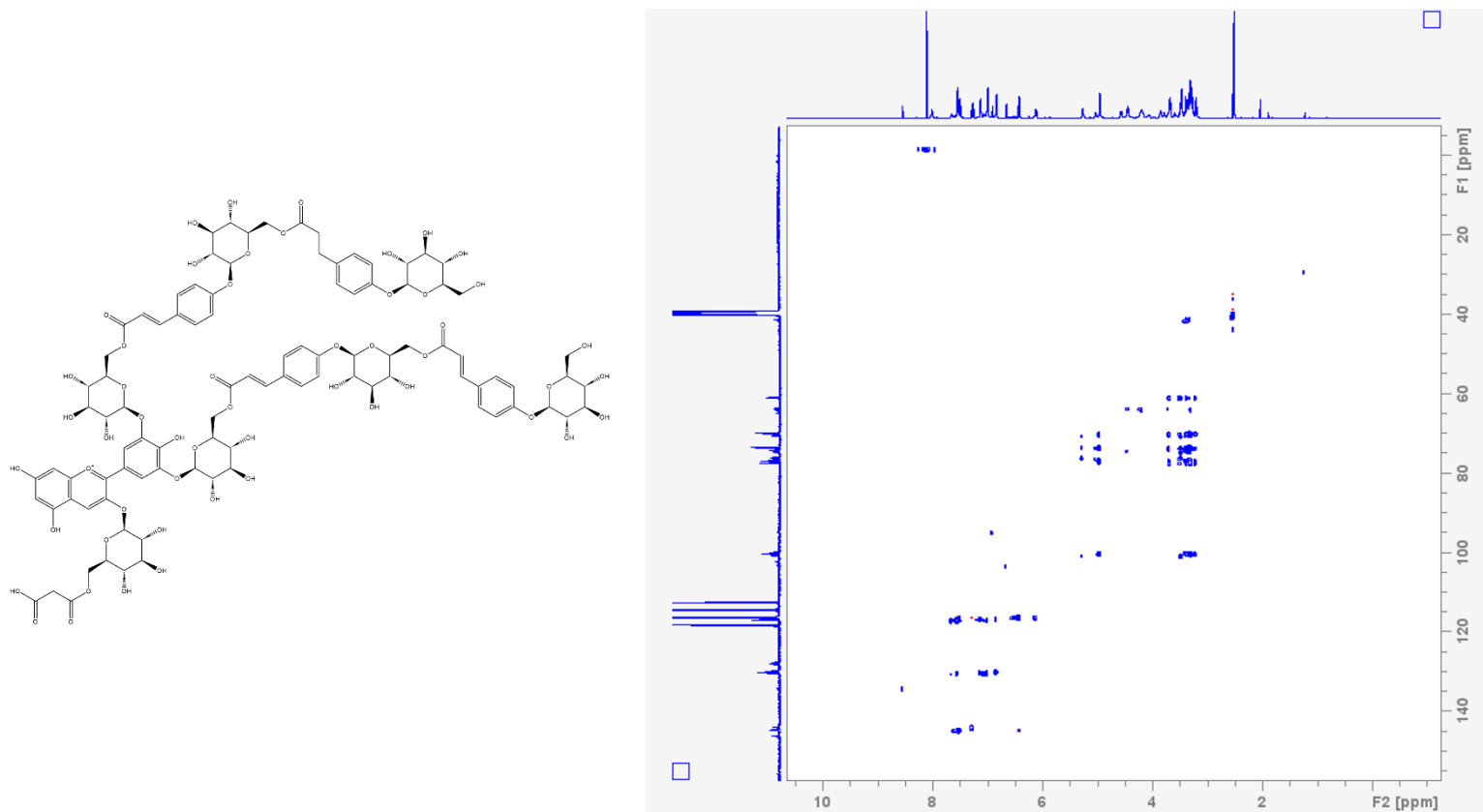
Appendix 2.1 (G) HSQC-TOCSY (H-C correlation, with long range H-H correlations) NMR spectra of ternatin A1 in DMSO-d_6/TFA-d, 9/1 (600 MHz for F2 frequency axis, and 150 MHz for F1 frequency axis). Number of Scans = 2 , Bruker pulse program = hsqcdietgpsisp2
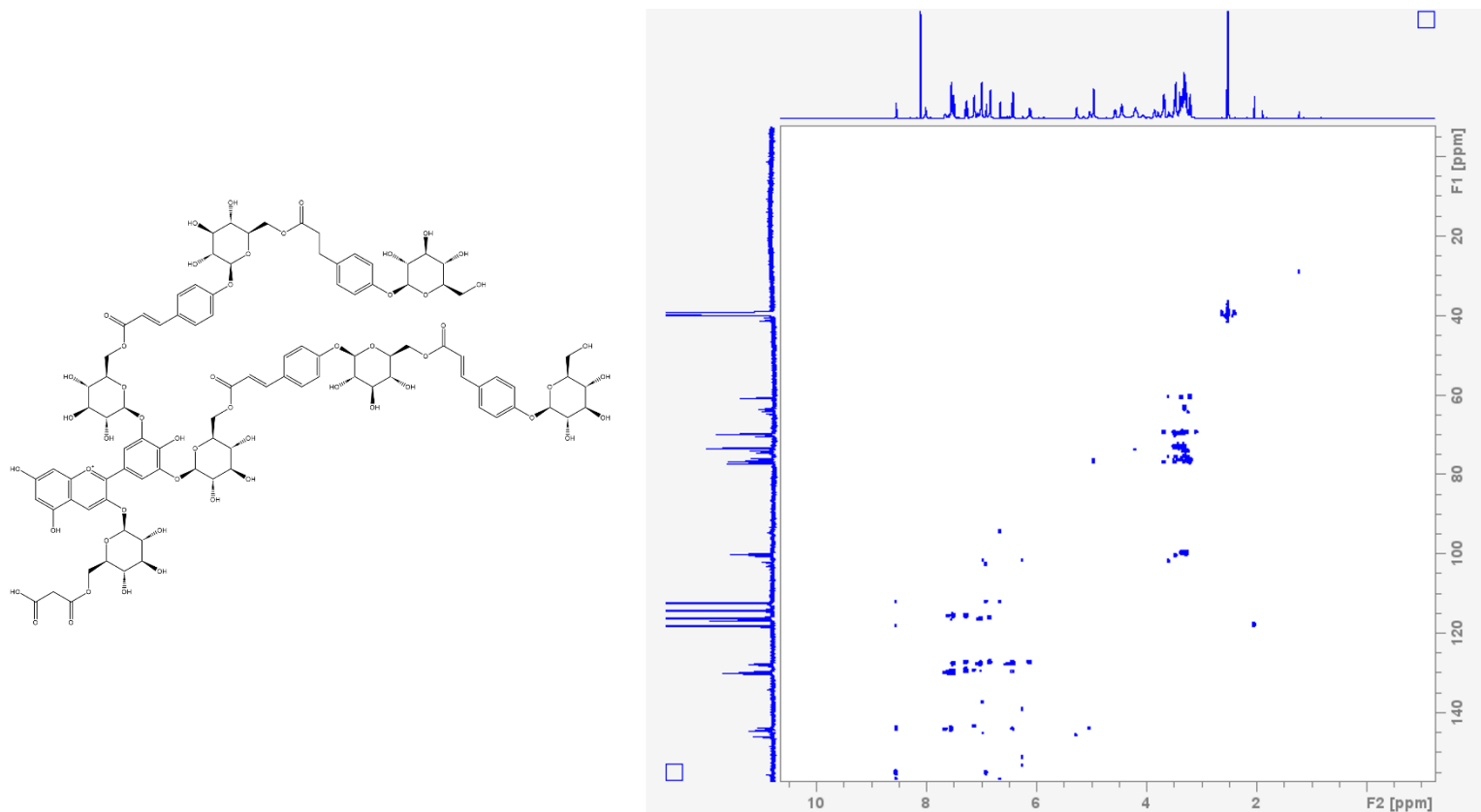
Appendix 2.1 (H) HMBC (Multiple bond H-C correlation) NMR spectra of ternatin A1 in DMSO-d_6/TFA-d, 9/1 (600 MHz for F2 frequency axis, and 150 MHz for F1 frequency axis). Number of Scans = 8, Bruker pulse program = hmbcetgpl2nd.2.
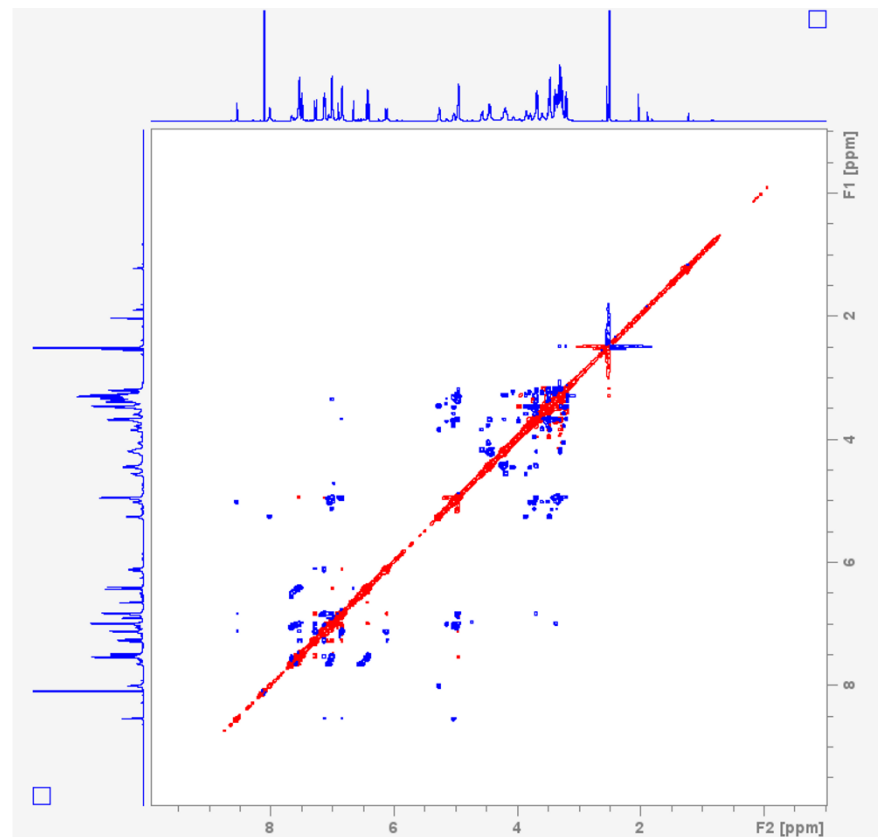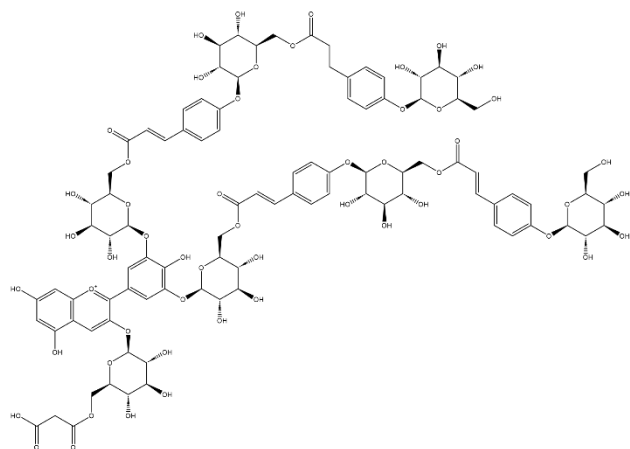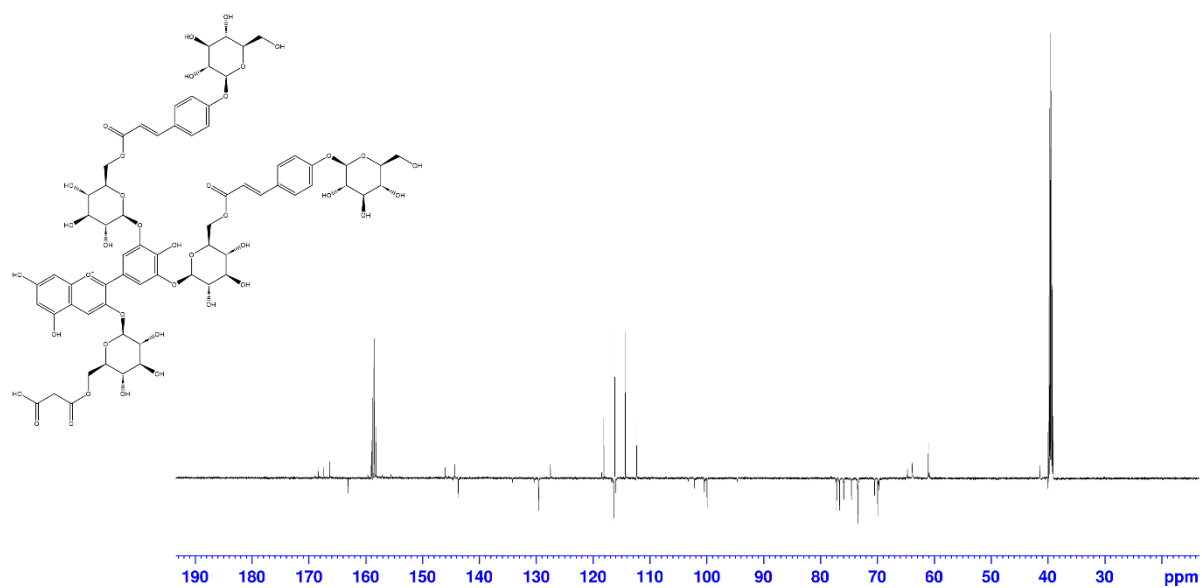
Appendix 2.1 (I) ROESY (Through space H-H correlation) NMR spectra of ternatin A1 in DMSO-d_6/TFA-d, 9/1 (600 MHz). Number of Scans = 24, Bruker pulse program = roesyadjsphpp.ptg.
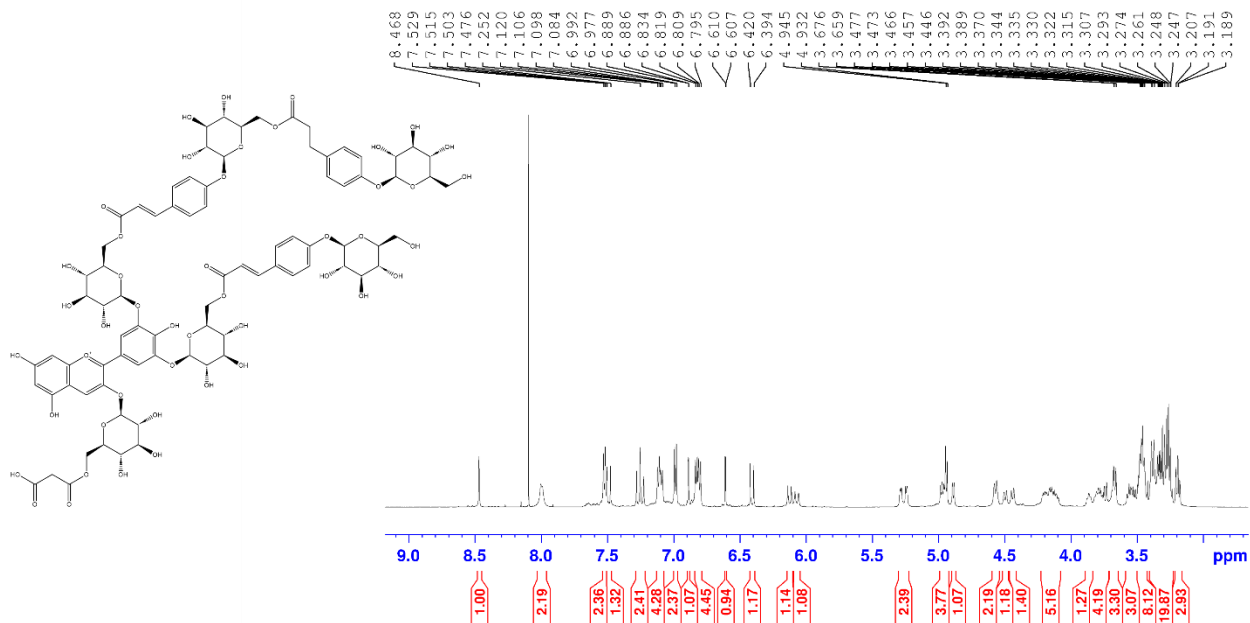
Appendix 2.2 (A) ^{1}H NMR spectra of ternatin A2 in DMSO-d_6/TFA-d, 9/1 (600 MHz). Number of Scans = 16, Bruker pulse program = zg30
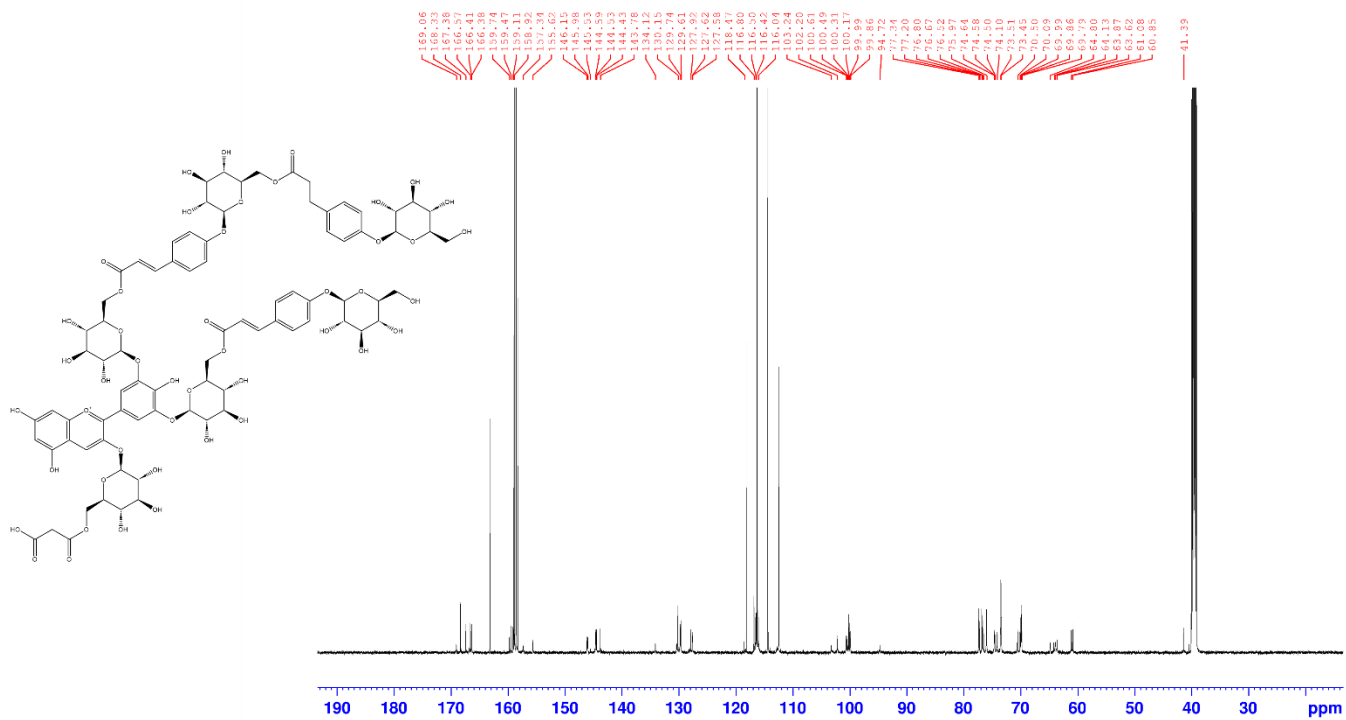
Appendix 2.2 (B) ^{13}C NMR spectra of ternatin A2 in DMSO-d_6/TFA-d, 9/1 (150 MHz). Number of Scans = 4,000, Bruker pulse program = zgpg30
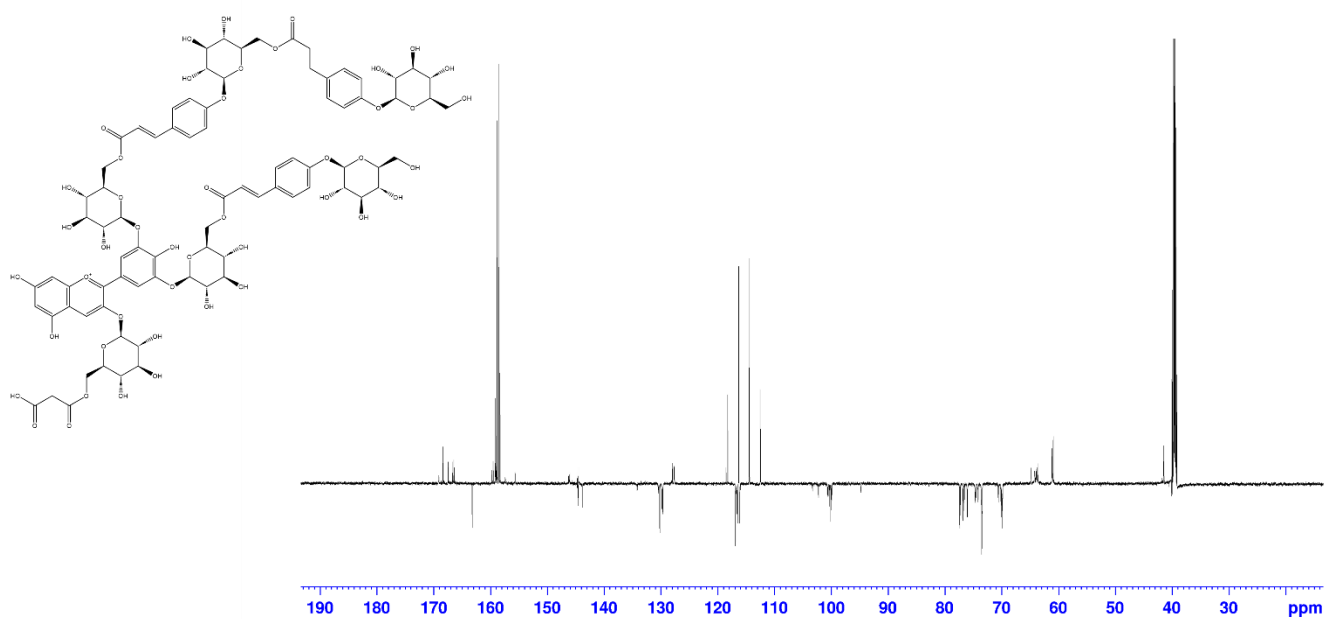
Appendix 2.2 (C) DEPTQ NMR spectra of ternatin A2 in DMSO-d_6/TFA-d, 9/1 (600 MHz). Number of Scans = 2,000, Bruker pulse program = deptqgpsp. Quaternary carbons (C) and methylene carbons (CH$_2$) are phased negative, while methine (CH) and methyl (CH$_3$) are phased positive.

Appendix 2.2 (D) COSY (H-H correlation) NMR spectra of ternatin A2 in DMSO-d_6/TFA-d, 9/1 (600 MHz). Number of Scans = 4 , Bruker pulse program = cosygpppqf.ptype.
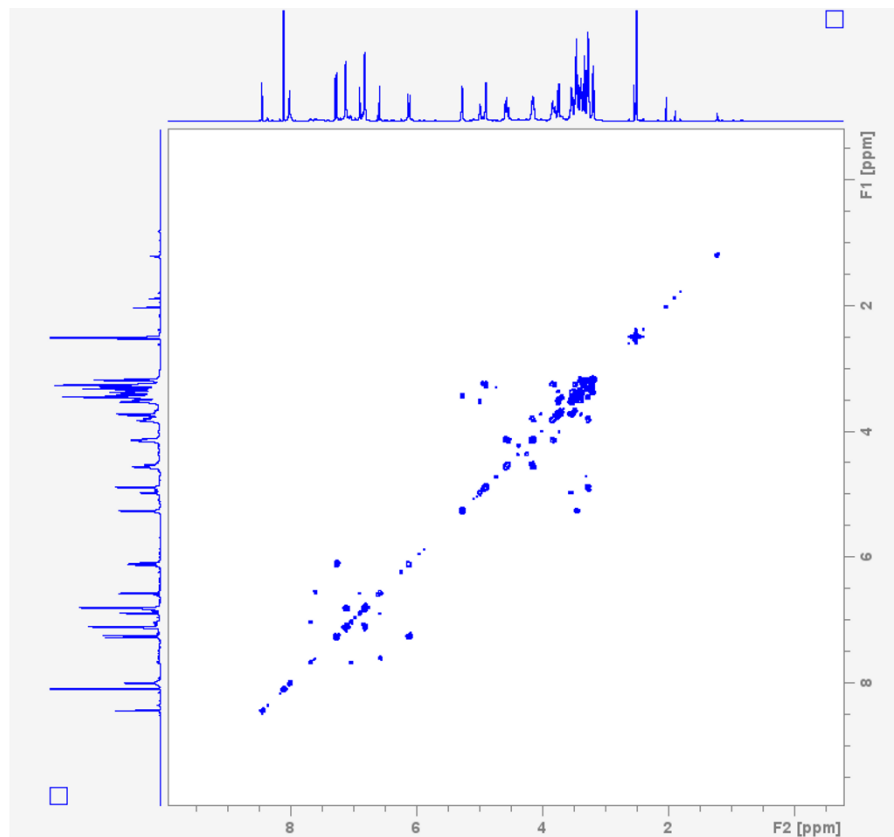
Appendix 2.2 (E) TOCSY (long range H-H correlation) NMR spectra of ternatin A2 in DMSO-d_6/TFA-d, 9/1 (600 MHz). Number of Scans = 4 , Bruker pulse program = dipsi2gpphzs

Appendix 2.2 (F) HSQC (H-C correlation) NMR spectra of ternatin A2 in DMSO-d_6/TFA-d, 9/1 (600 MHz for F2 frequency axis, and 150 MHz for F1 frequency axi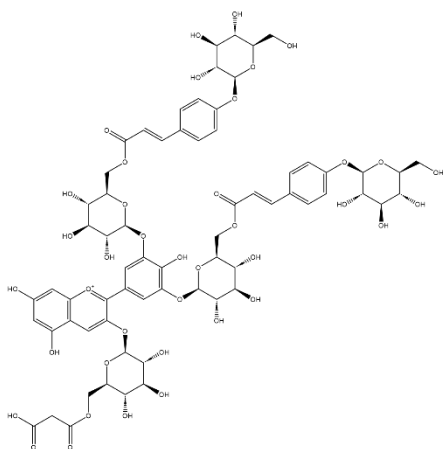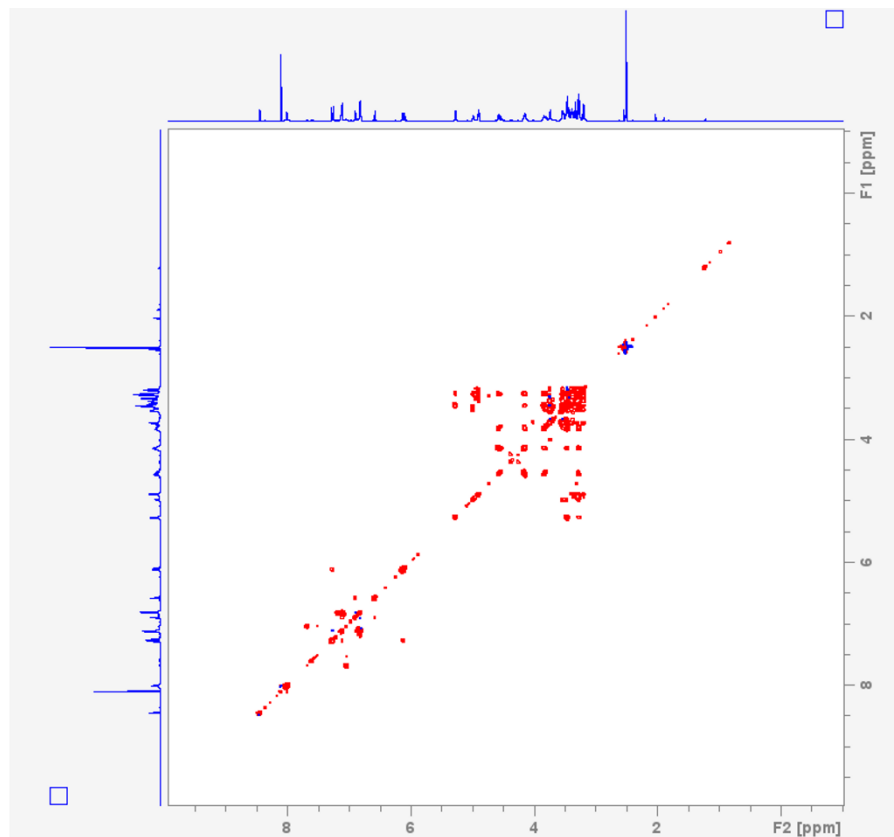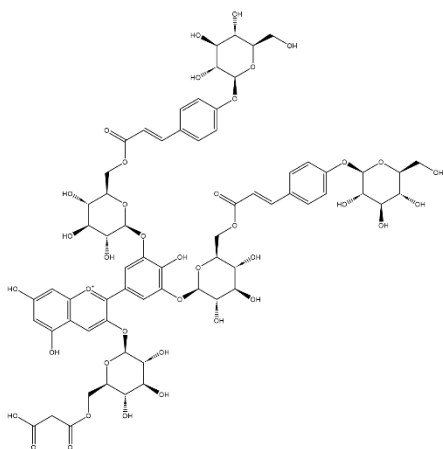s). Number of Scans = 32 , Bruker pulse program = hsqcedetgpsisp2.3. Crosspeaks are phased such that methylene protons (CH$_2$) are negative (red), while methine (CH) and methyl (CH$_3$) are phased positive (blue).

Appendix 2.2 (G) HSQC-TOCSY (H-C correlation, with long range H-H correlations) NMR spectra of ternatin A2 in DMSO-d_6/TFA-d, 9/1 (600 MHz for F2 frequency axis, and 150 MHz for F1 frequency axis). Number of Scans = 2 , Bruker pulse program = hsqcdietgpsisp2

Appendix 2.2 (H) HMBC (Multiple bond H-C correlation) NMR spectra of ternatin A2 in DMSO-d_6/TFA-d, 9/1 (600 MHz for F2 frequency axis, and 150 MHz for F1 frequency axis). Number of Scans = 8, Bruker pulse program = hmbcetgpl2nd.2.
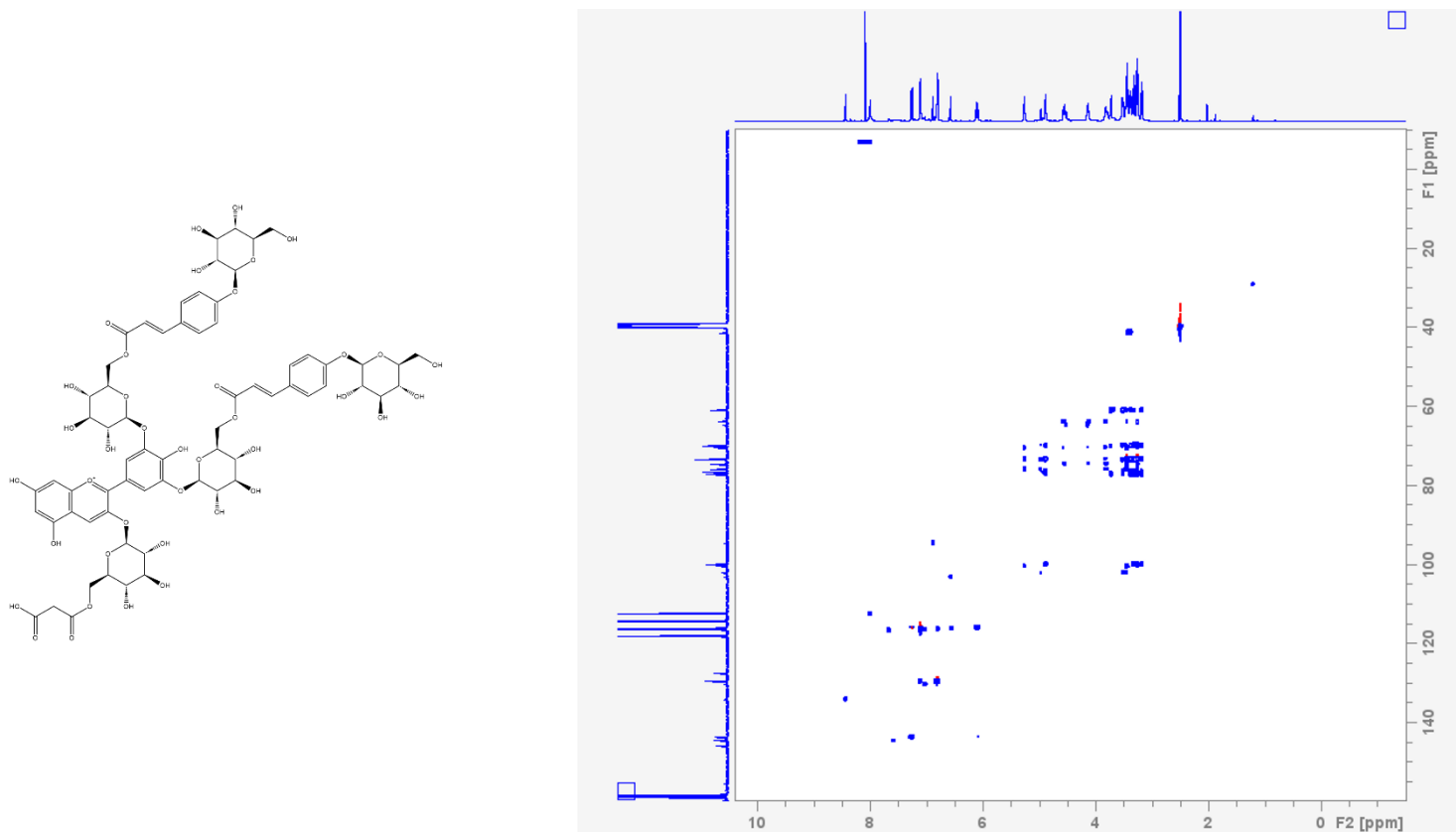
Appendix 2.2 (I) ROESY (Through space H-H correlation) NMR spectra of ternatin A2 in DMSO-d_6/TFA-d, 9/1 (600 MHz). Number of Scans = 24, Bruker pulse program = roesyadjsphpp.ptg.
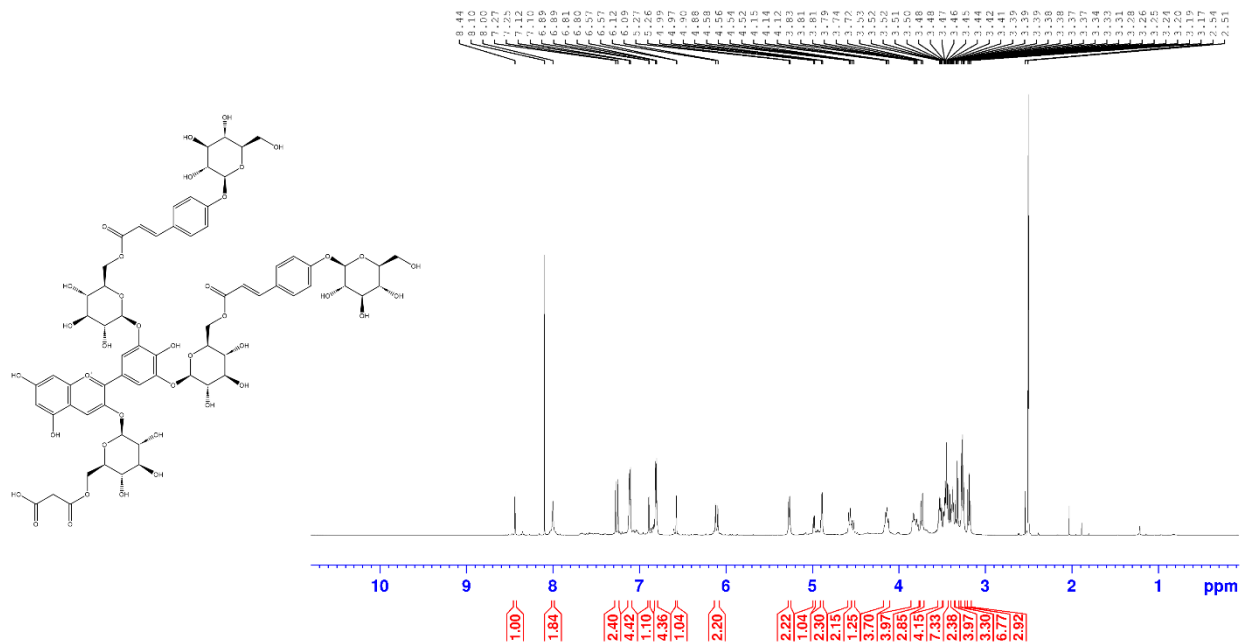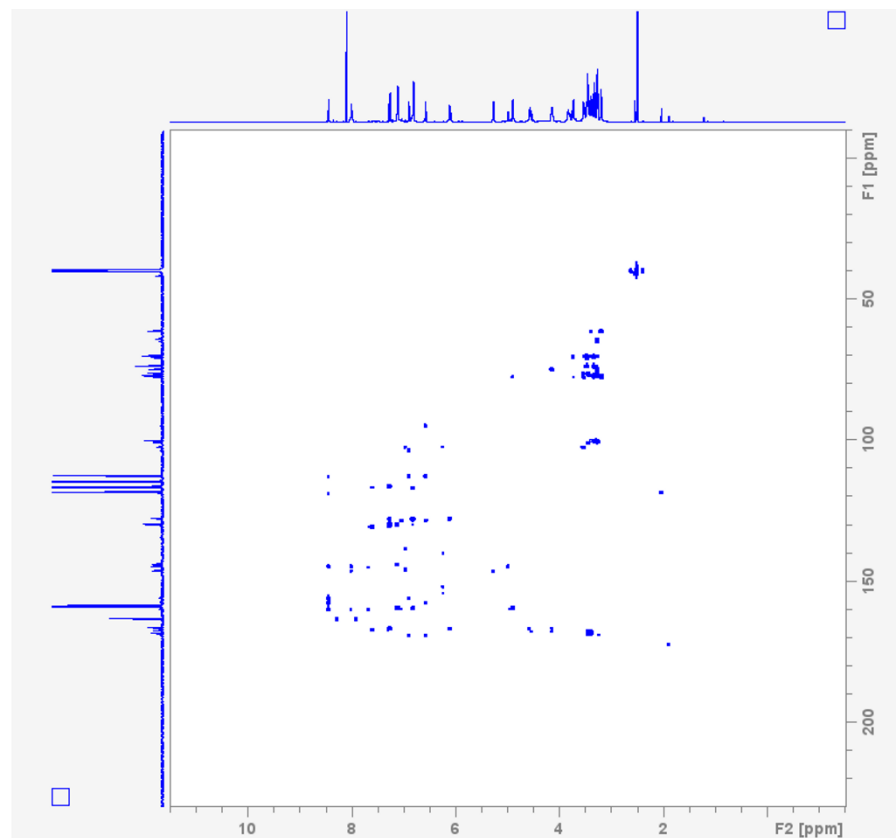
Appendix 2.3 (A) ^{1}H NMR spectra of ternatin A3 in DMSO-d_6/TFA-d, 9/1 (600 MHz). Number of Scans = 16, Bruker pulse program = zg30

Appendix 2.3 (B) ^{13}C NMR spectra of ternatin A3 in DMSO-d_6/TFA-d, 9/1 (150 MHz). Number of Scans = 4,000, Bruker pulse program = zgpg30

Appendix 2.3 (C) DEPTQ NMR spectra of ternatin A3 in DMSO-d_6/TFA-d, 9/1 (600 MHz). Number of Scans = 2,000, Bruker pulse program = deptqgpsp. Quaternary carbons (C) and methylene carbons (CH$_2$) are phased negative, while methine (CH) and methyl (CH$_3$) are phased positive.

Appendix 2.3 (D) COSY (H-H correlation) NMR spectra of ternatin A3 in DMSO-d_6/TFA-d, 9/1 (600 MHz). Number of Scans = 4 , Bruker pulse program = cosygpppqf.ptype.
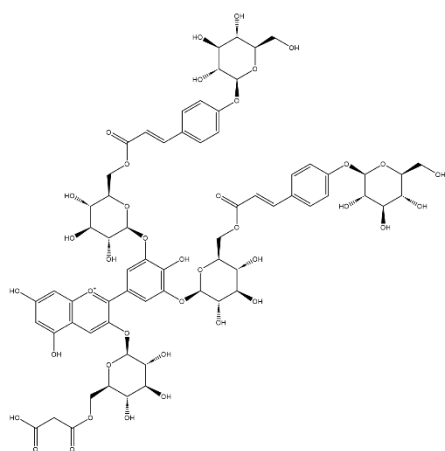
Appendix 2.3 (E) TOCSY (long range H-H correlation) NMR spectra of ternatin A3 in DMSO-d_6/TFA-d, 9/1 (600 MHz). Number of Scans = 4 , Bruker pulse program = dipsi2gpphzs
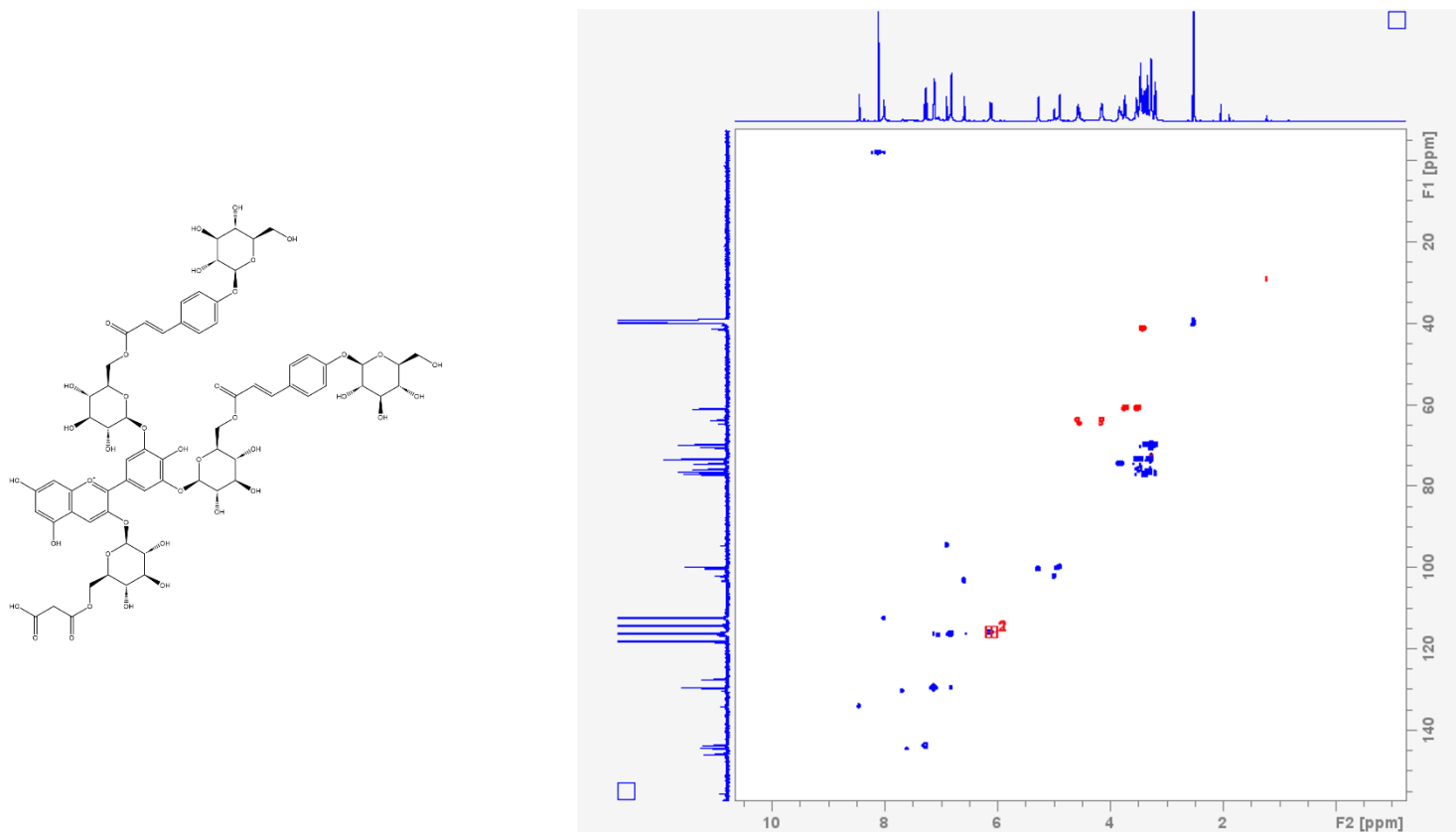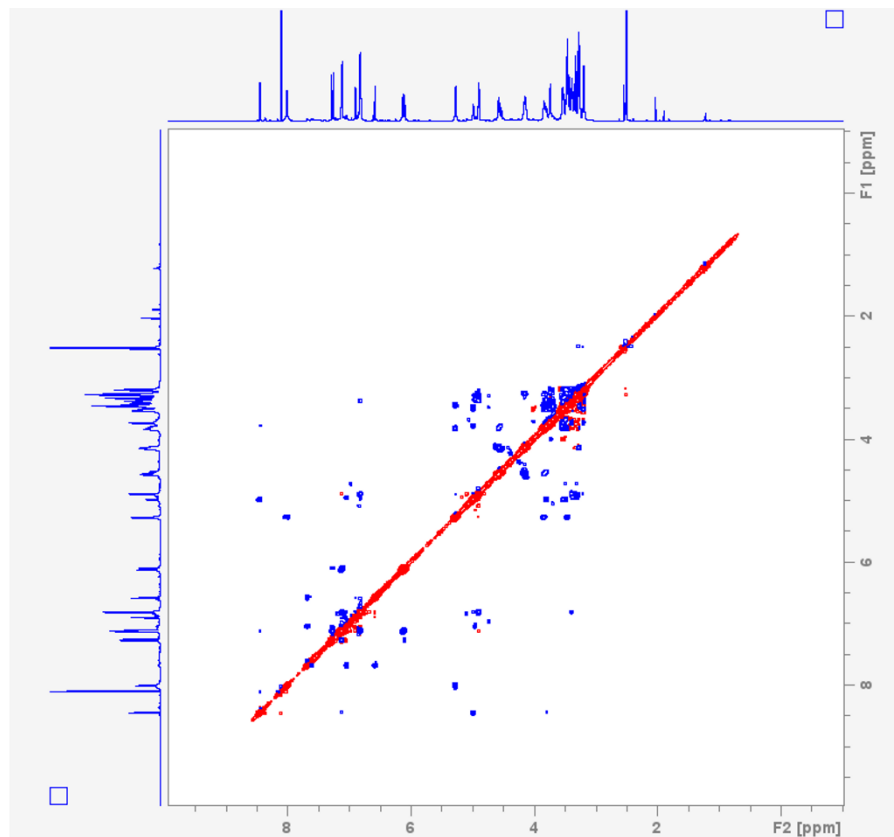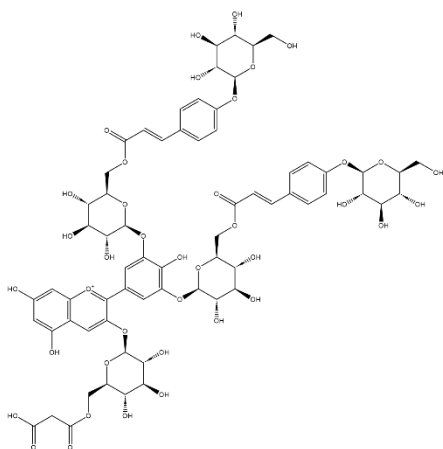
Appendix 2.3 (F) HSQC (H-C correlation) NMR spectra of ternatin A3 in DMSO-d_6/TFA-d, 9/1 (600 MHz for F2 frequency axis, and 150 MHz for F1 frequency axis). Number of Scans = 32 , Bruker pulse program = hsqcedetgpsisp2.3. Crosspeaks are phased such that methylene protons (CH$_2$) are negative (red), while methine (CH) and methyl (CH$_3$) are phased positive (blue).
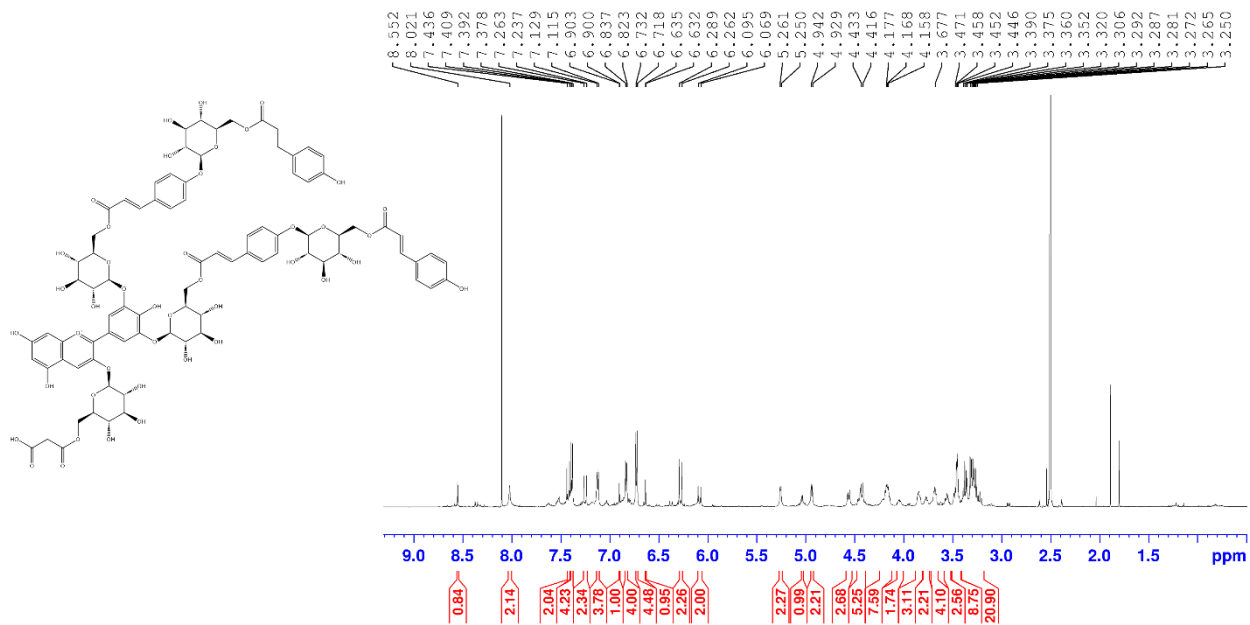
Appendix 2.3 (G) HSQC-TOCSY (H-C correlation, with long range H-H correlations) NMR spectra of ternatin A3 in DMSO-d_6/TFA-d, 9/1 (600 MHz for F2 frequency axis, and 150 MHz for F1 frequency axis). Number of Scans = 2 , Bruker pulse program = hsqcdietgpsisp2
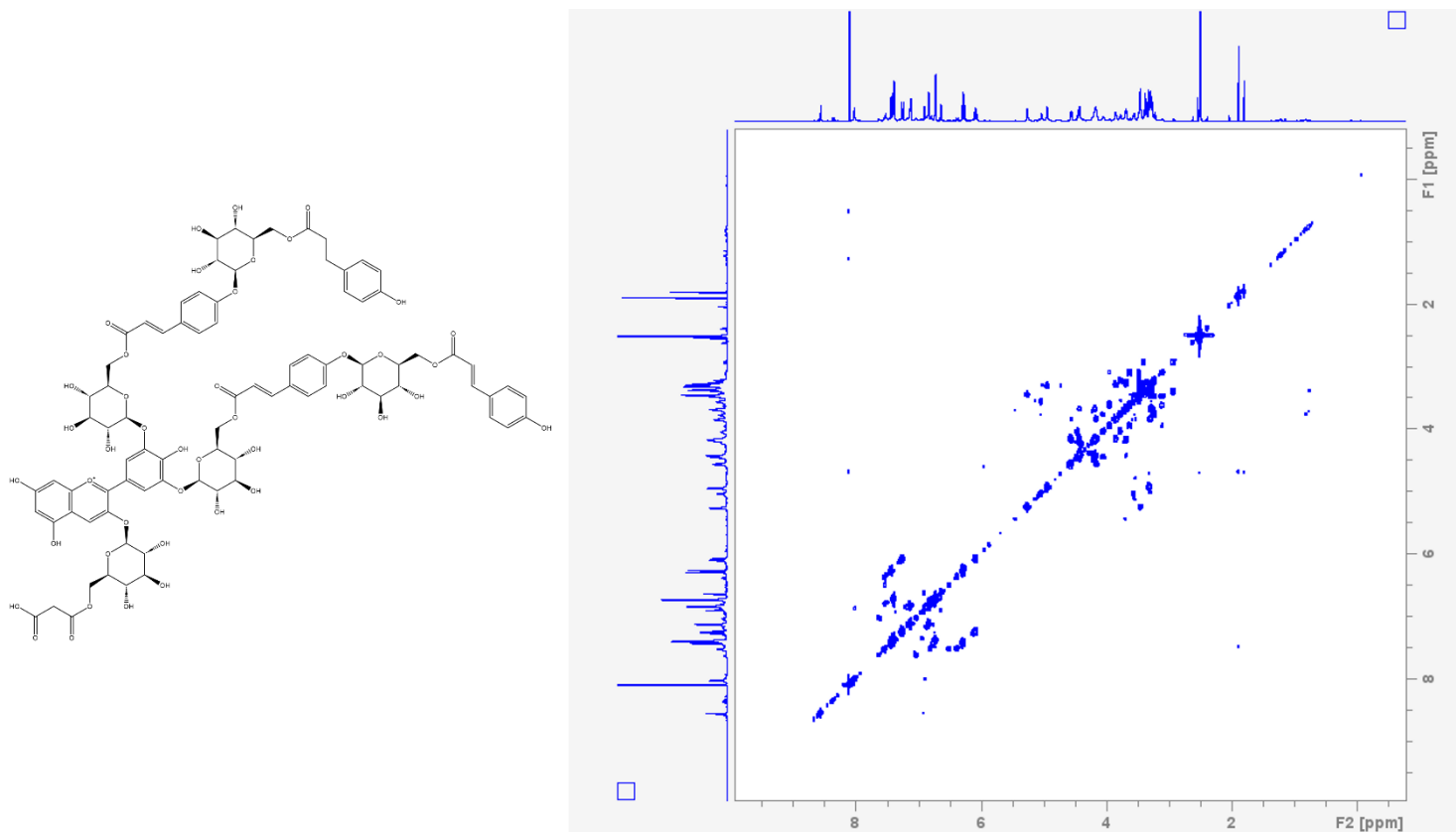
Appendix 2.3 (H) HMBC (Multiple bond H-C correlation) NMR spectra of ternatin A3 in DMSO-d_6/TFA-d, 9/1 (600 MHz for F2 frequency axis, and 150 MHz for F1 frequency axis). Number of Scans = 8, Bruker pulse program = hmbcetgpl2nd.2.
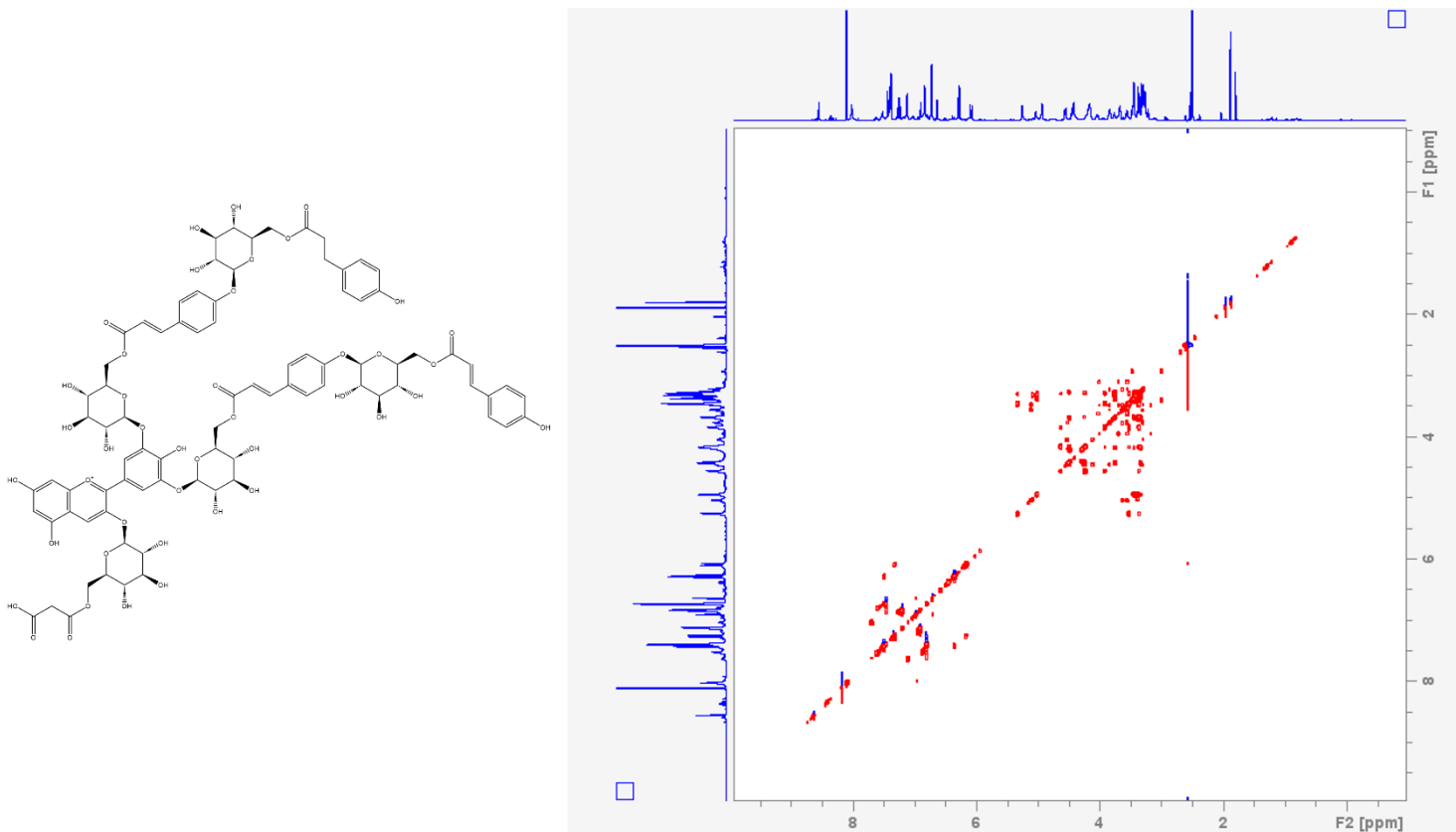
Appendix 2.3 (I) ROESY (Through space H-H correlation) NMR spectra of ternatin A3 in DMSO-d_6/TFA-d, 9/1 (600 MHz). Number of Scans = 24, Bruker pulse program = roesyadjsphpp.ptg.
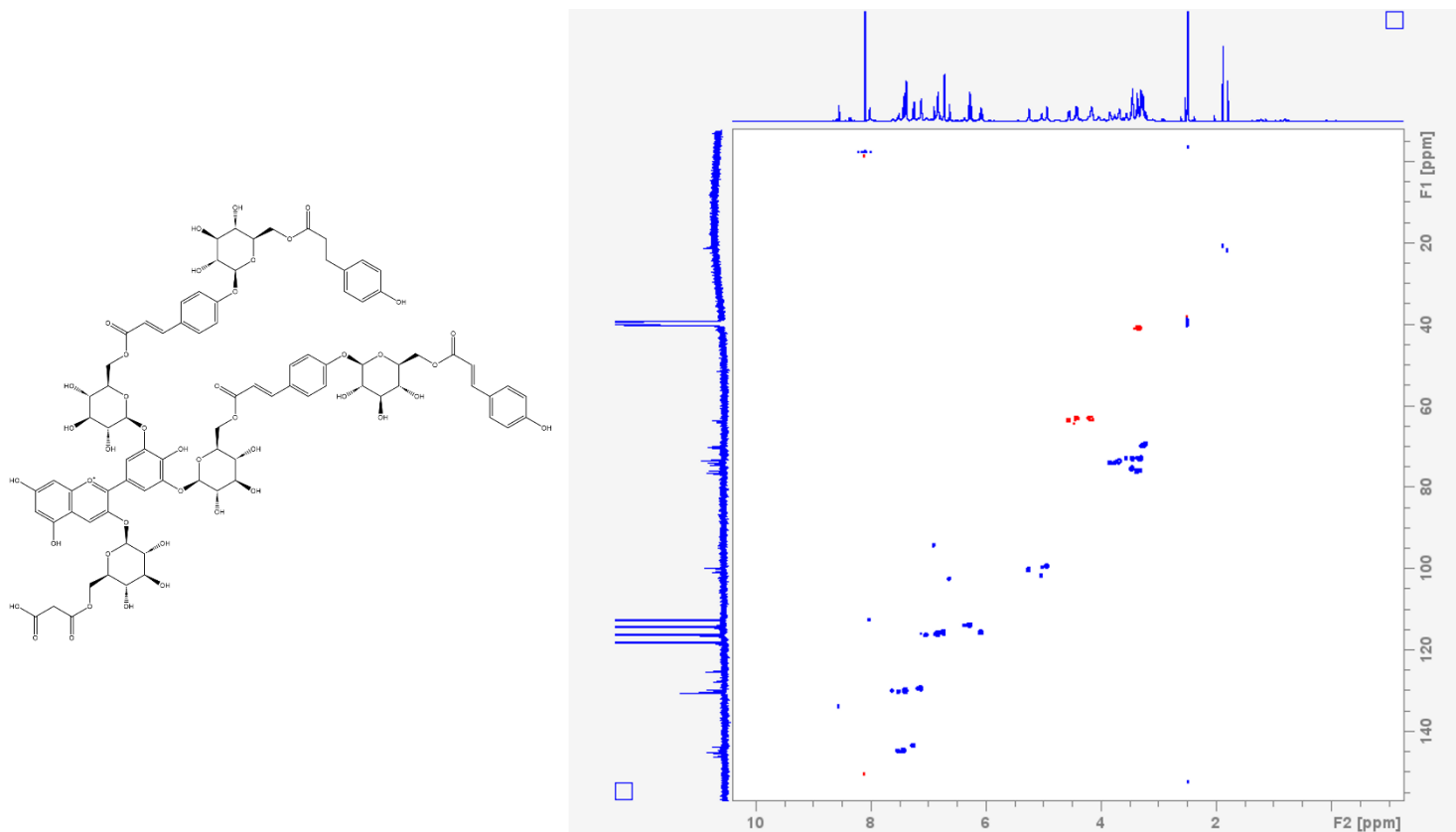
Appendix 2.4 (A) ^{1}H NMR spectra of ternatin D1 in DMSO-d_6/TFA-d, 9/1 (600 MHz). Number of Scans = 16, Bruker pulse program = zg30

Appendix 2.4 (B) ^{13}C NMR spectra of ternatin D1 in DMSO-d_6/TFA-d, 9/1 (150 MHz). Number of Scans = 4,000, Bruker pulse program = zgpg30
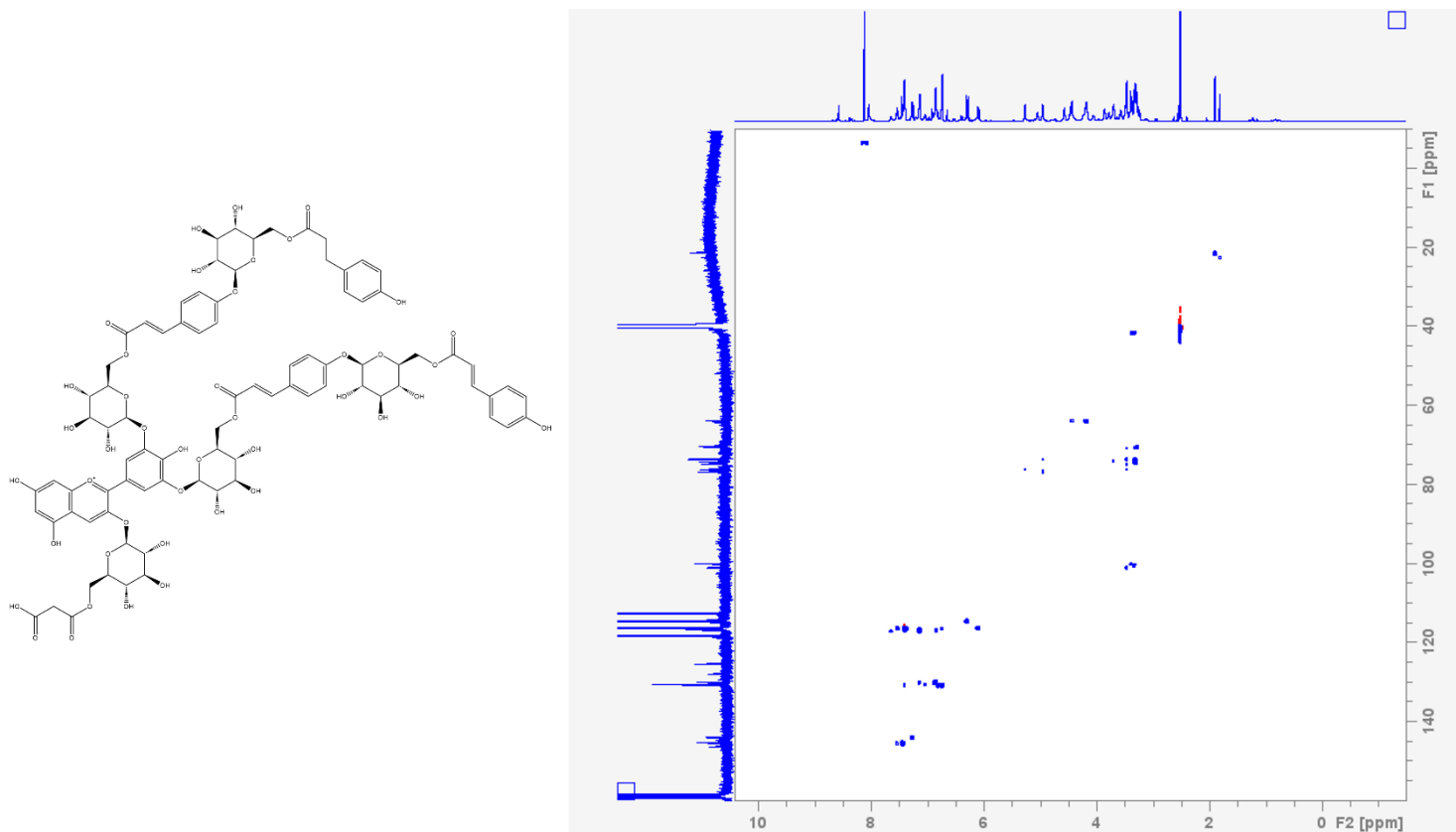
Appendix 2.4 (C) COSY (H-H correlation) NMR spectra of ternatin D1 in DMSO-d_6/TFA-d, 9/1 (600 MHz). Number of Scans = 4 , Bruker pulse program = cosygpppqf.ptype.
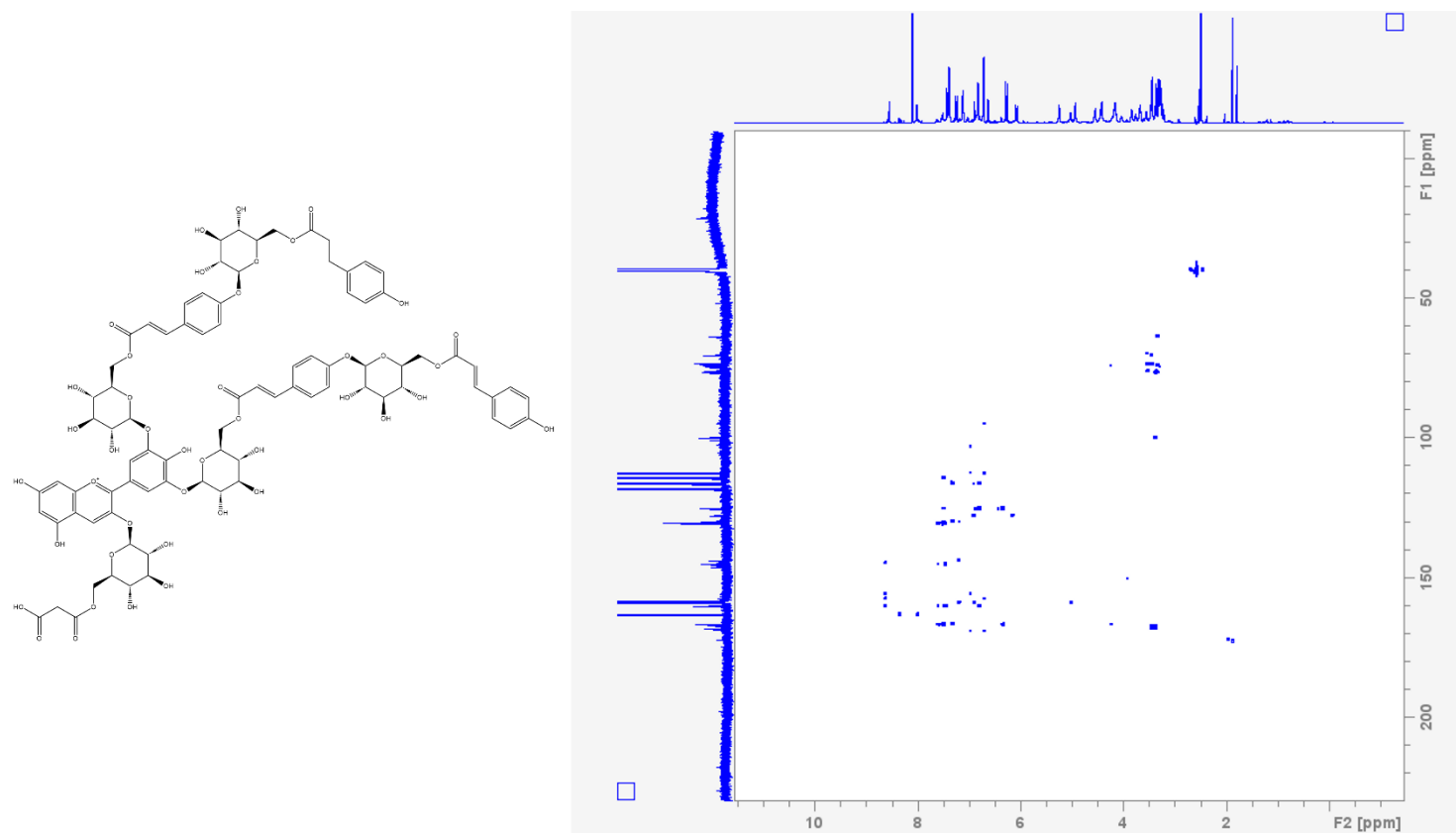
Appendix 2.4 (D) TOCSY (long range H-H correlation) NMR spectra of ternatin D1 in DMSO-d_6/TFA-d, 9/1 (600 MHz). Number of Scans = 4 , Bruker pulse program = dipsi2gpphzs
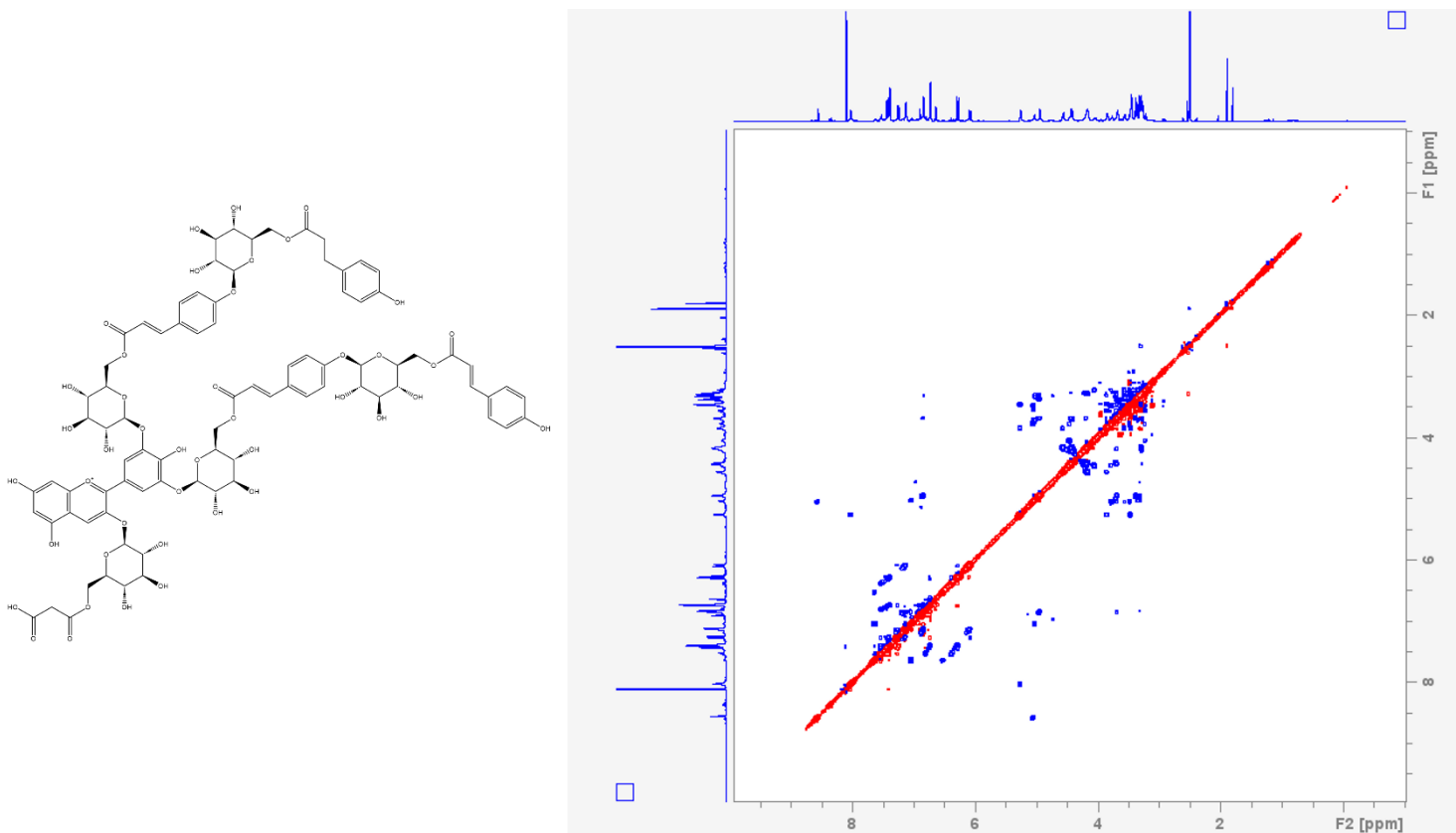
Appendix 2.4 (E) HSQC (H-C correlation) NMR spectra of ternatin D1 in DMSO-d_6/TFA-d, 9/1 (600 MHz for F2 frequency axis, and 150 MHz for F1 frequency axis). Number of Scans = 32 , Bruker pulse program = hsqcedetgpsisp2.3. Crosspeaks are phased such that methylene protons (CH$_2$) are negative (red), while methine (CH) and methyl (CH$_3$) are phased positive (blue).

Appendix 2.4 (F) HSQC-TOCSY (H-C correlation, with long range H-H correlations) NMR spectra of ternatin D1 in DMSO-d_6/TFA-d, 9/1 (600 MHz for F2 frequency axis, and 150 MHz for F1 frequency axis). Number of Scans = 2 , Bruker pulse program = hsqcdietgpsisp2

Appendix 2.4 (G) HMBC (Multiple bond H-C correlation) NMR spectra of ternatin D1 in DMSO-d_6/TFA-d, 9/1 (600 MHz for F2 frequency axis, and 150 MHz for F1 frequency axis). Number of Scans = 8, Bruker pulse program = hmbcetgpl2nd.2.

Appendix 2.4 (H) ROESY (Through space H-H correlation) NMR spectra of ternatin D1 in DMSO-d_6/TFA-d, 9/1 (600 MHz). Number of Scans = 24, Bruker pulse program = roesyadjsphpp.ptg.
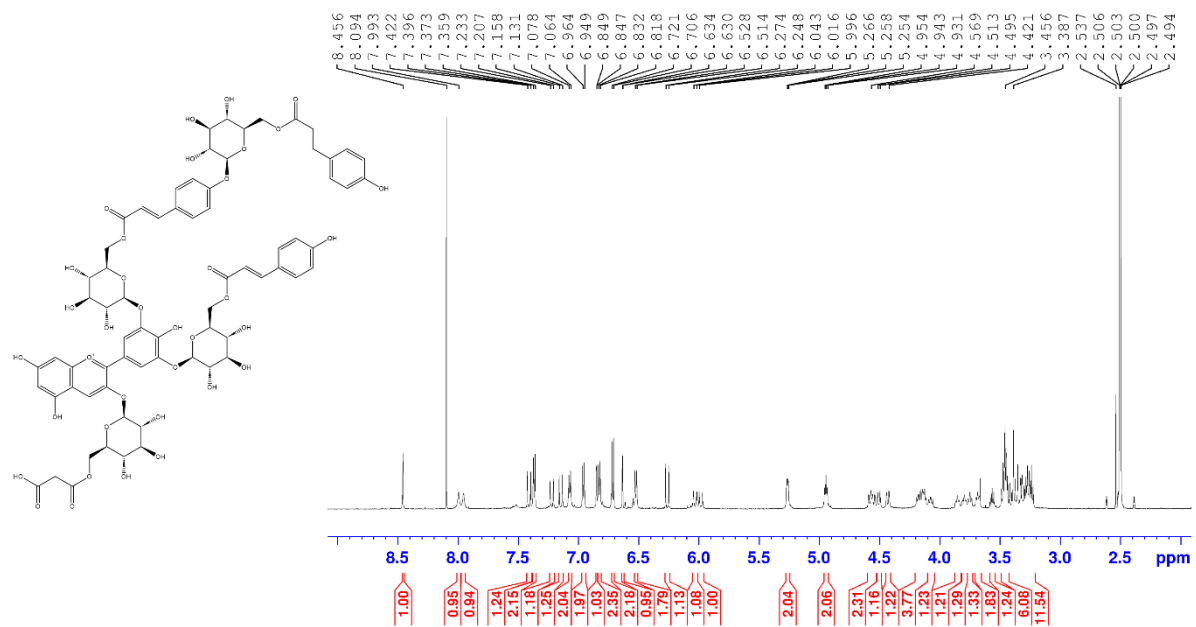
Appendix 2.5 (A) ^{1}H NMR spectra of ternatin D2 in DMSO-d_6/TFA-d, 9/1 (600 MHz). Number of Scans = 16, Bruker pulse program = zg30
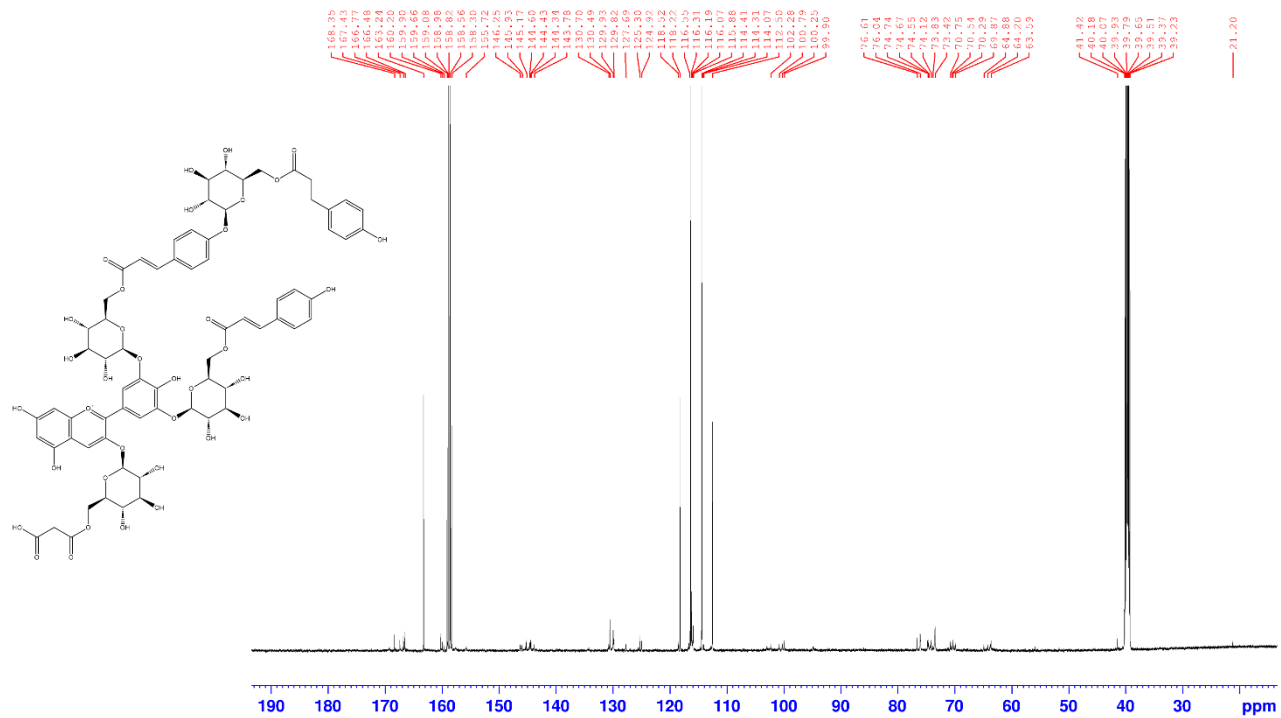
Appendix 2.5 (B) ^{13}C NMR spectra of ternatin D2 in DMSO-d_6/TFA-d, 9/1 (150 MHz). Number of Scans = 4,000, Bruker pulse program = zgpg30

Appendix 2.5 (C) DEPTQ NMR spectra of ternatin D2 in DMSO-d_6/TFA-d, 9/1 (600 MHz). Number of Scans = 2,000, Bruker pulse program = deptqgpsp. Quaternary carbons (C) and methylene carbons (CH_2) are phased negative, while methine (CH) and methyl (CH_3) are phased positive.

Appendix 2.5 (D) COSY (H-H correlation) NMR spectra of ternatin D2 in DMSO-d_6/TFA-d, 9/1 (600 MHz). Number of Scans = 4 , Bruker pulse program = cosygpppqf.ptype.

Appendix 2.5 (E) TOCSY (long range H-H correlation) NMR spectra of ternatin D2 in DMSO-d_6/TFA-d, 9/1 (600 MHz). Number of Scans = 4 , Bruker pulse program = dipsi2gpphzs
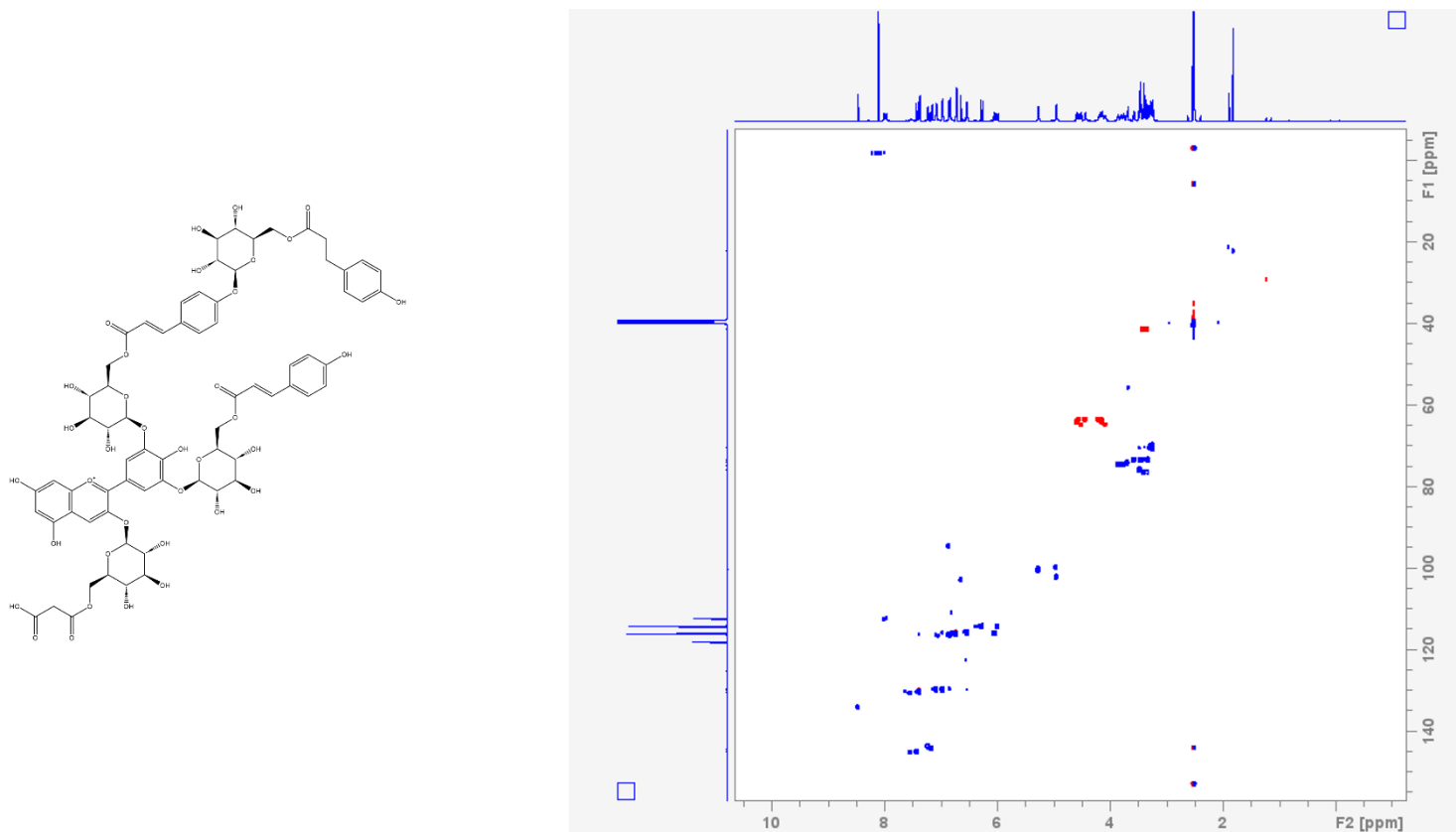
Appendix 2.5 (F) HSQC (H-C correlation) NMR spectra of ternatin D2 in DMSO-d_6/TFA-d, 9/1 (600 MHz for F2 frequency axis, and 150 MHz for F1 frequency axi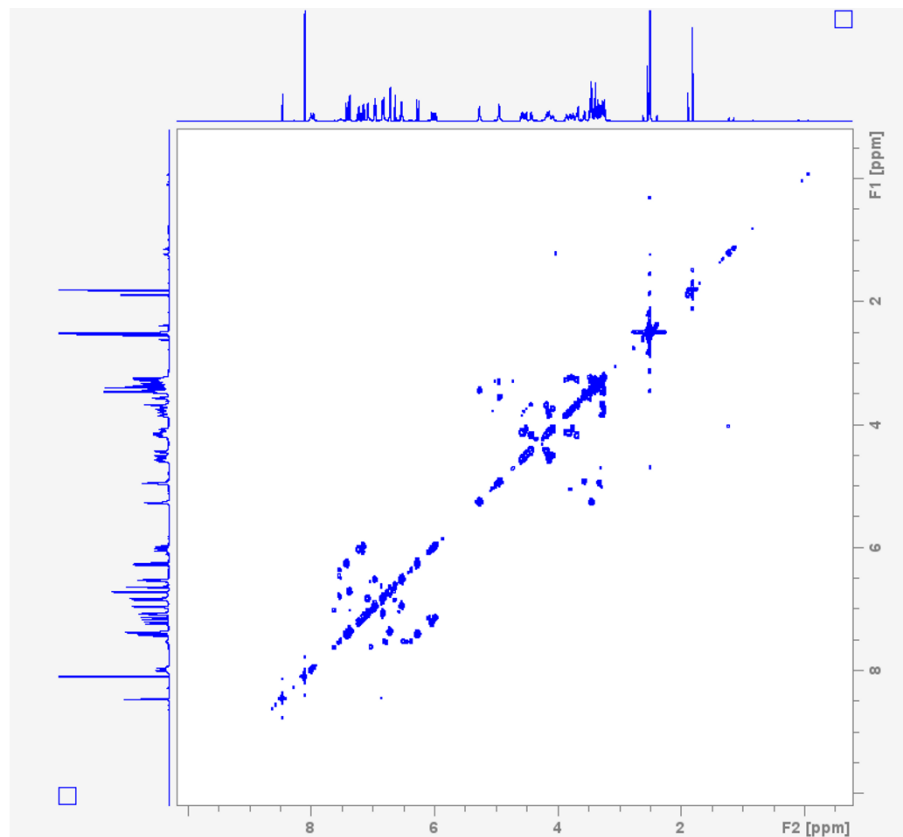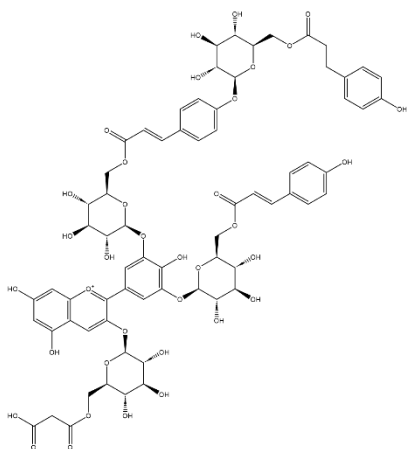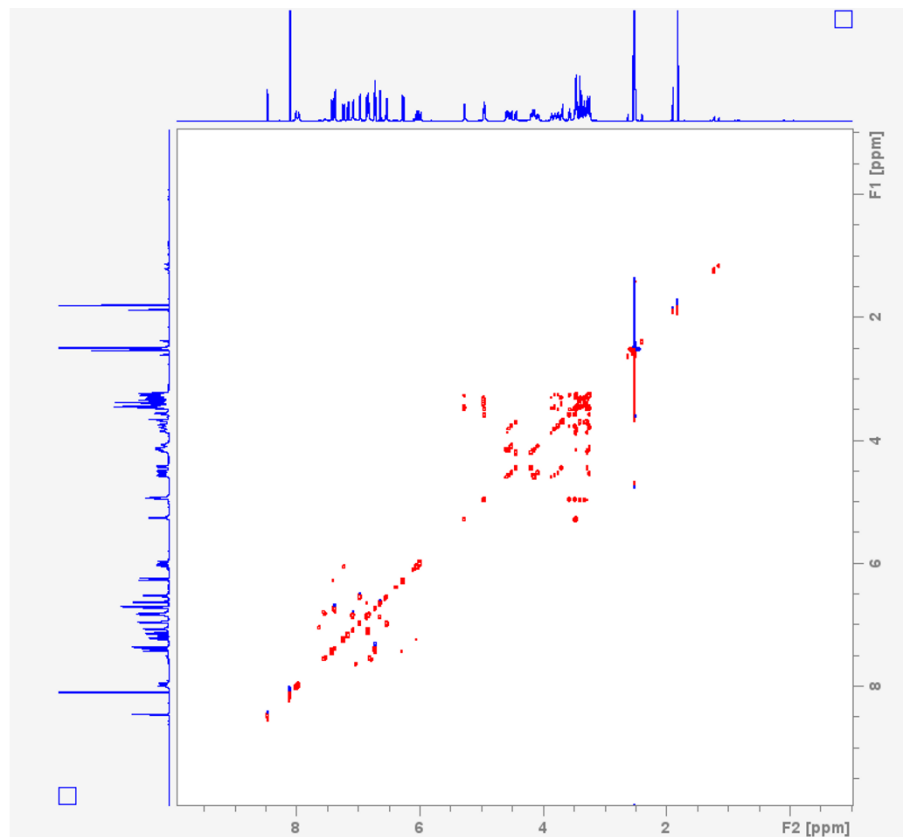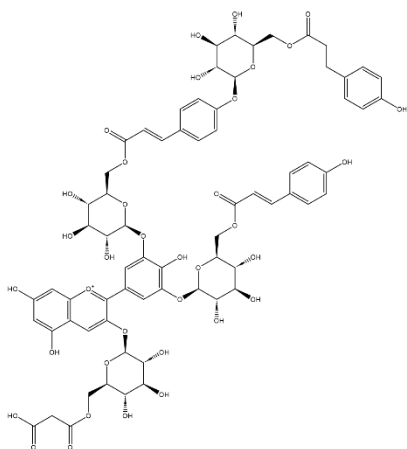s). Number of Scans = 32 , Bruker pulse program = hsqcedetgpsisp2.3. Crosspeaks are phased such that methylene protons (CH$_2$) are negative (red), while methine (CH) and methyl (CH$_3$) are phased positive (blue).

Appendix 2.5 (G) HSQC-TOCSY (H-C correlation, with long range H-H correlations) NMR spectra of ternatin D2 in DMSO-d_6/TFA-d, 9/1 (600 MHz for F2 frequency axis, and 150 MHz for F1 frequency axis). Number of Scans = 2 , Bruker pulse program = hsqcdietgpsisp2

Appendix 2.5 (H) HMBC (Multiple bond H-C correlation) NMR spectra of ternatin D2 in DMSO-d_6/TFA-d, 9/1 (600 MHz for F2 frequency axis, and 150 MHz for F1 frequency axis). Number of Scans = 8, Bruker pulse program = hmbcetgpl2nd.2.
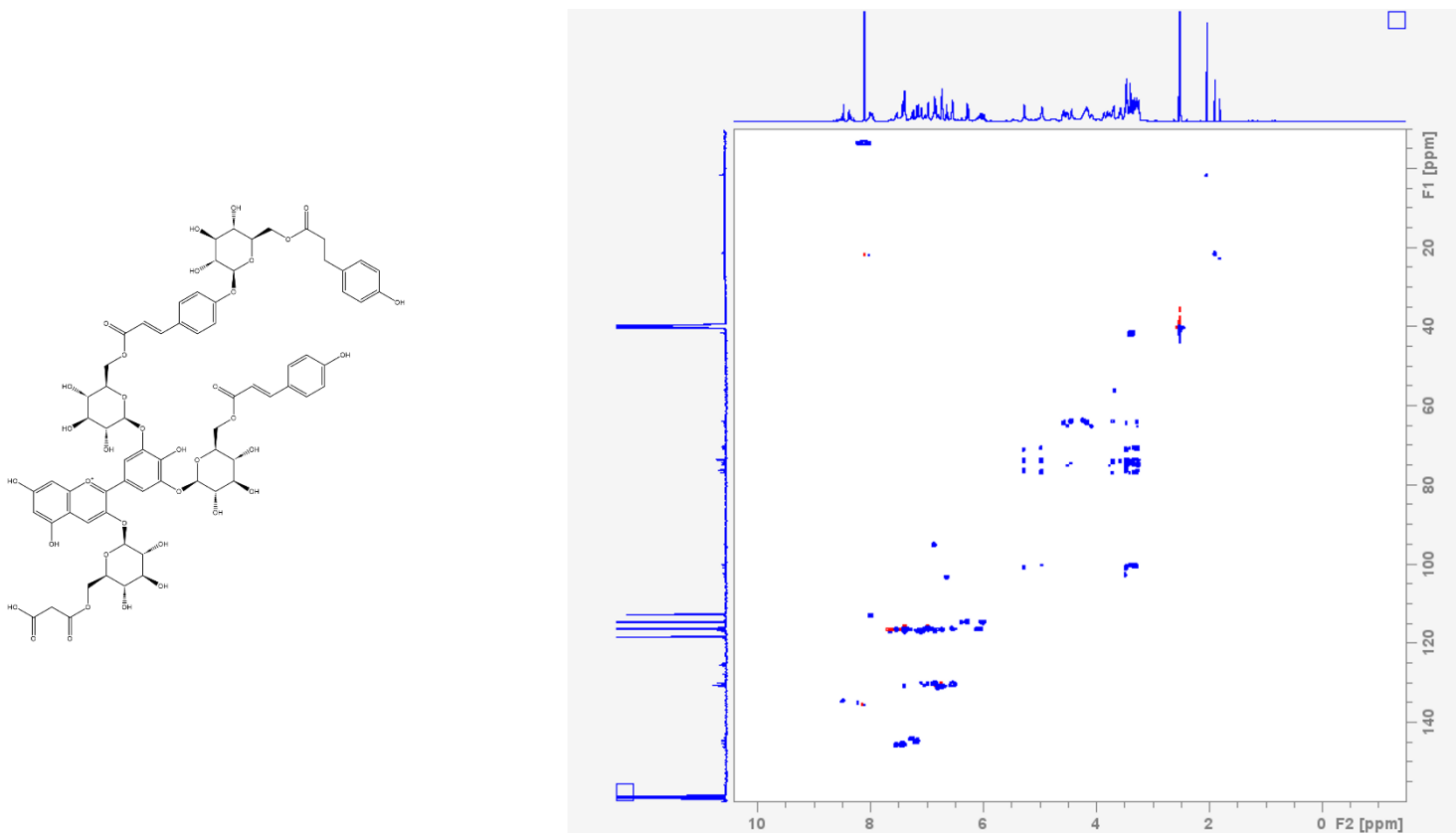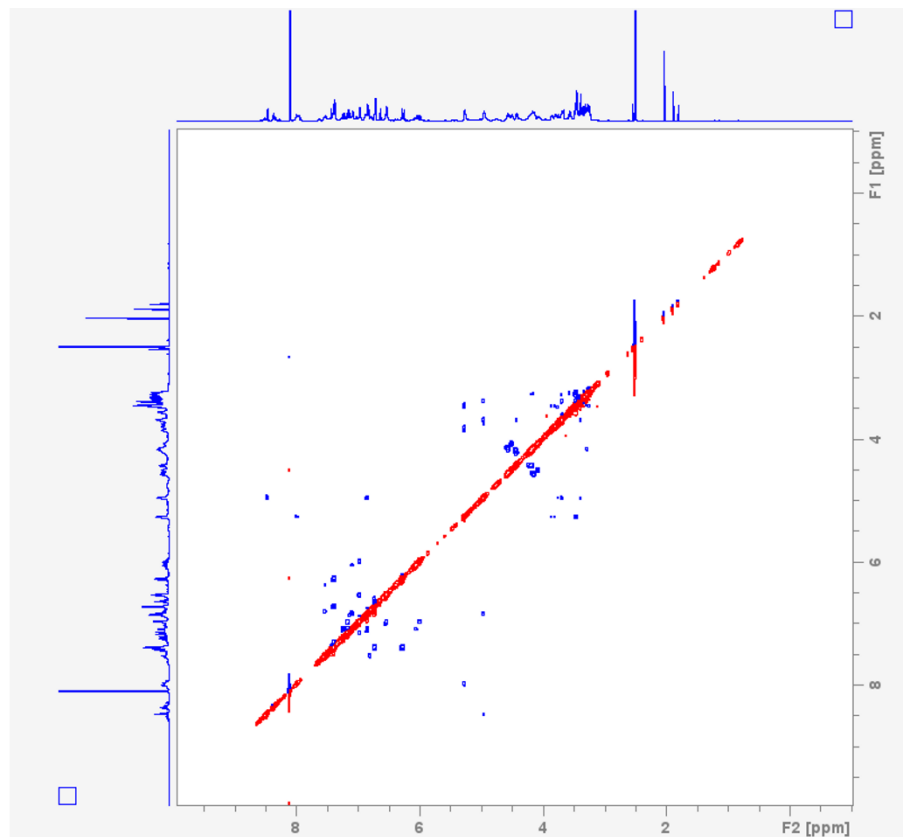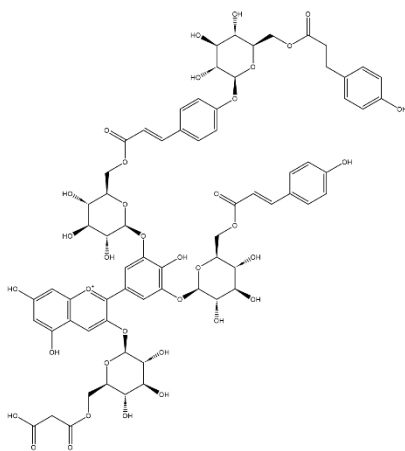
Appendix 2.5 (I) ROESY (Through space H-H correlation) NMR spectra of ternatin D2 in DMSO-d_6/TFA-d, 9/1 (600 MHz). Number of Scans = 24, Bruker pulse program = roesyadjsphpp.ptg.
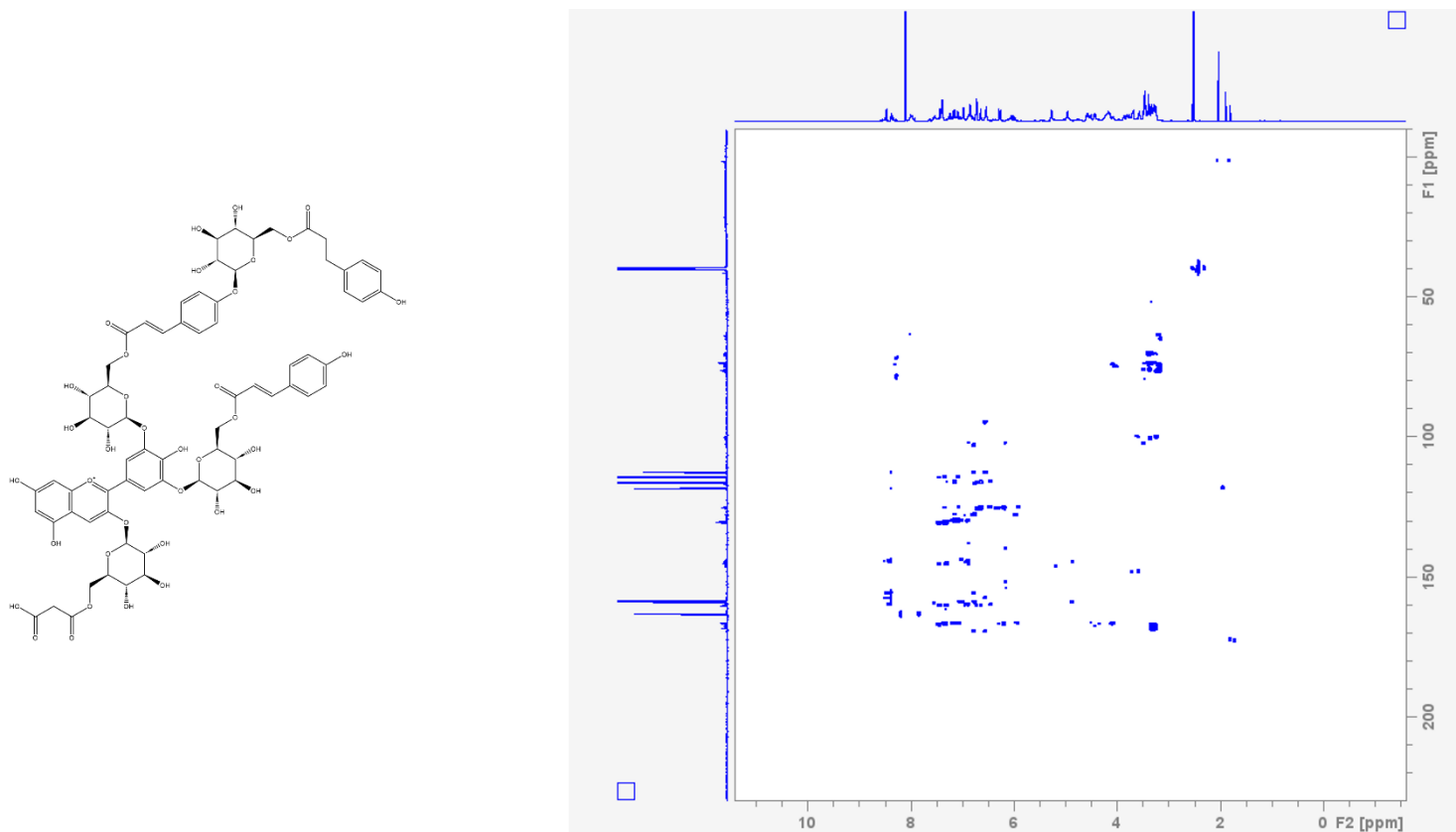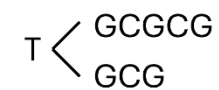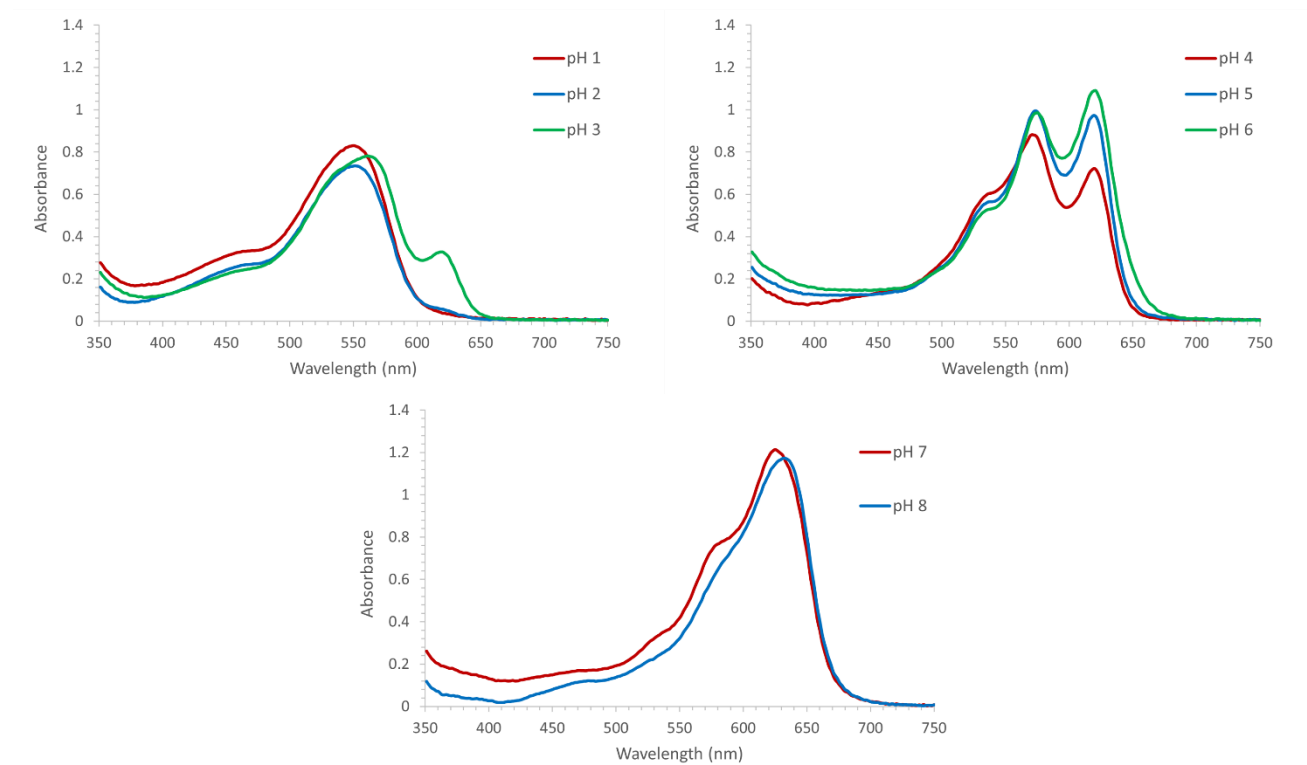
Appendix 3.1 UV-VIS absorbance spectra of ternatin A2 (60 µM) between pH 1 and 8.
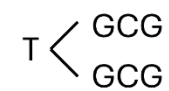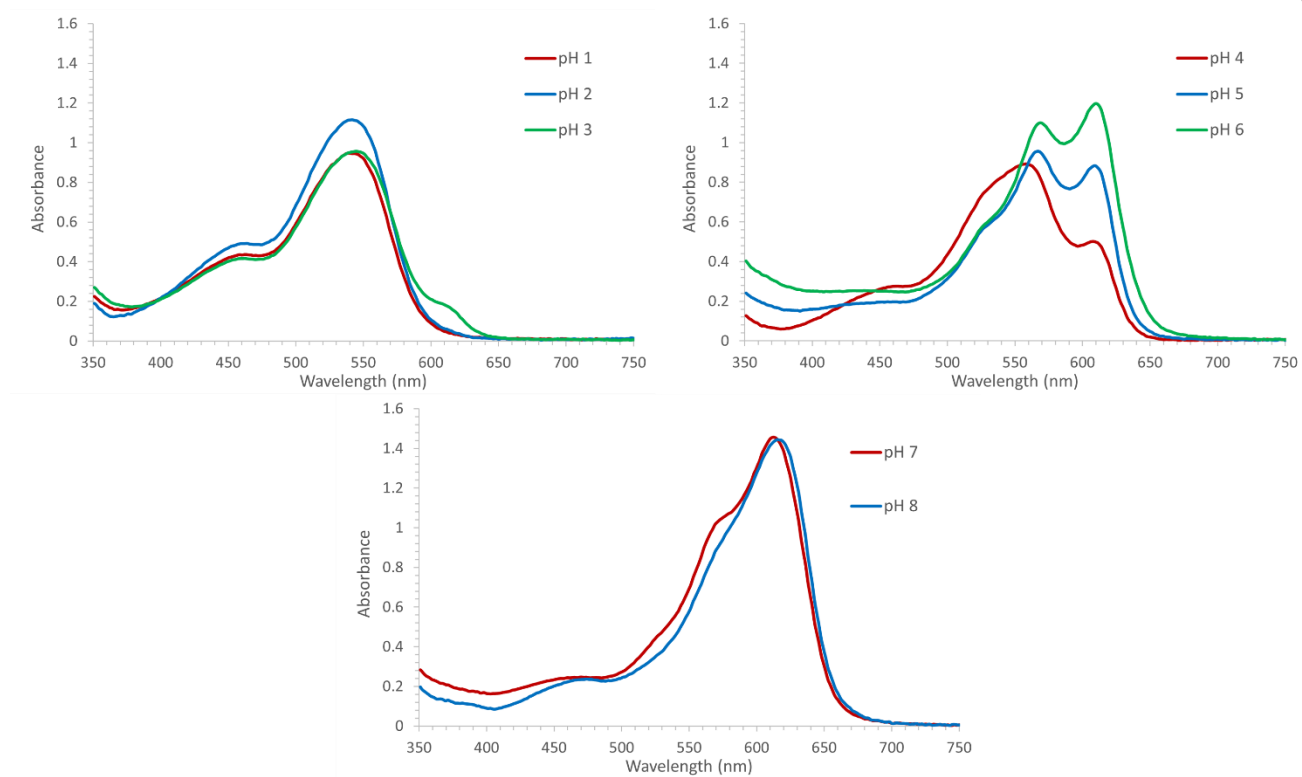
Appendix 3.2 UV-VIS absorbance spectra of ternatin A3 (60 µM) between pH 1 and 8.
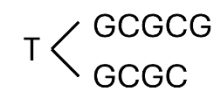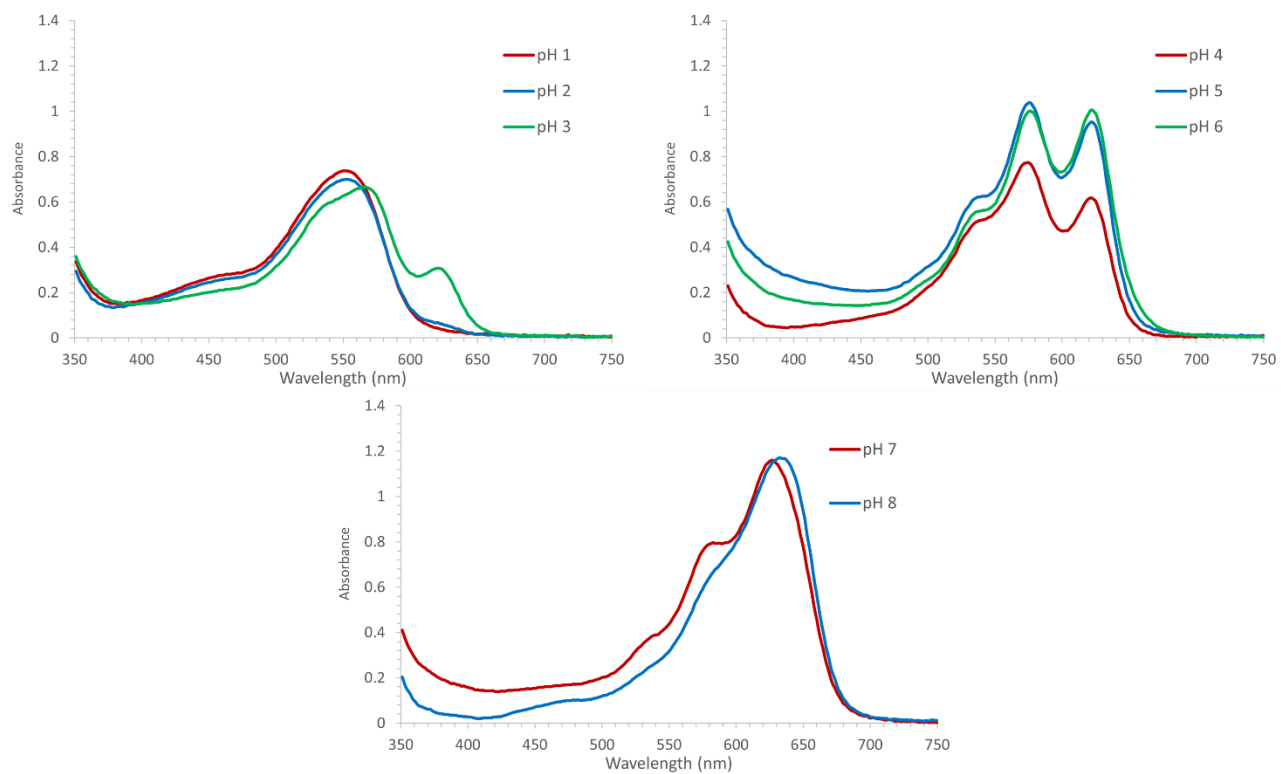
Appendix 3.3 UV-VIS absorbance spectra of ternatin B1 (60 μM) between pH 1 and 8.
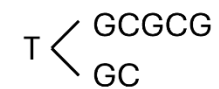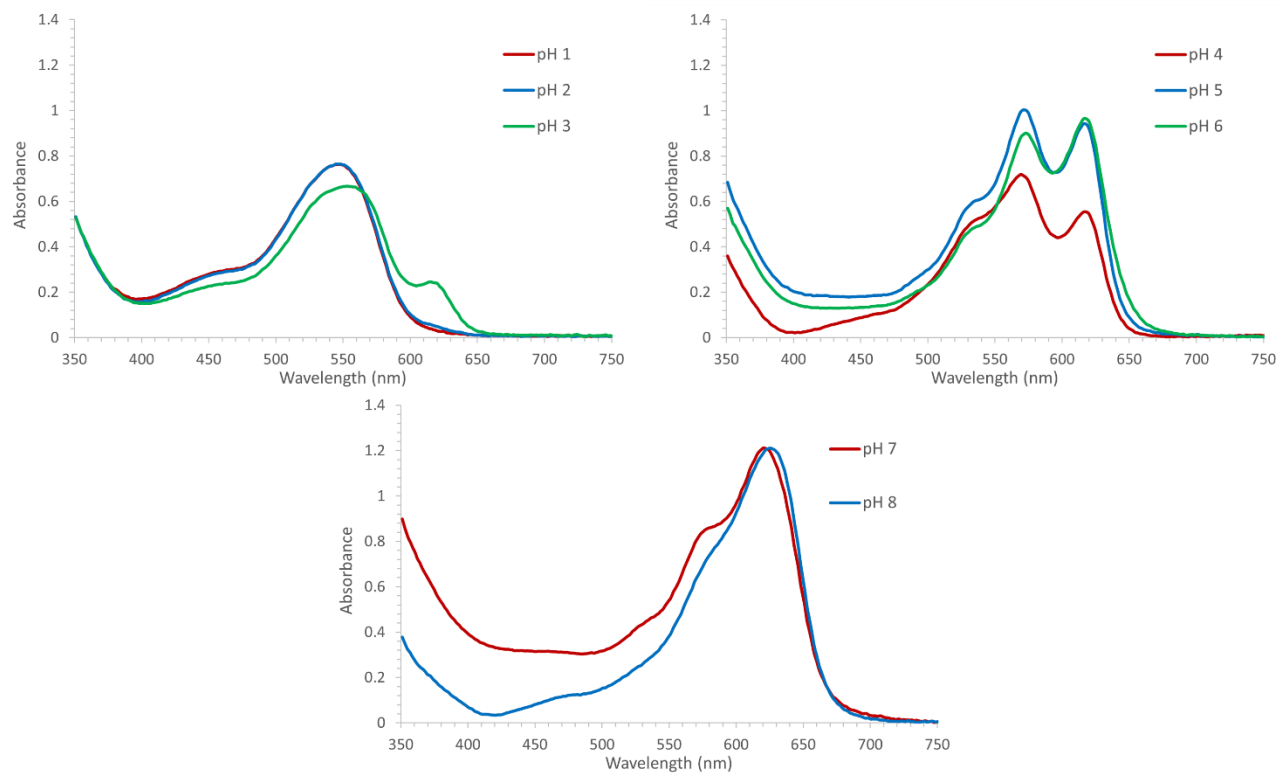
Appendix 3.4 UV-VIS absorbance spectra of ternatin B2 (60 µM) between pH 1 and 8.
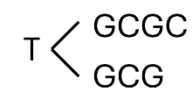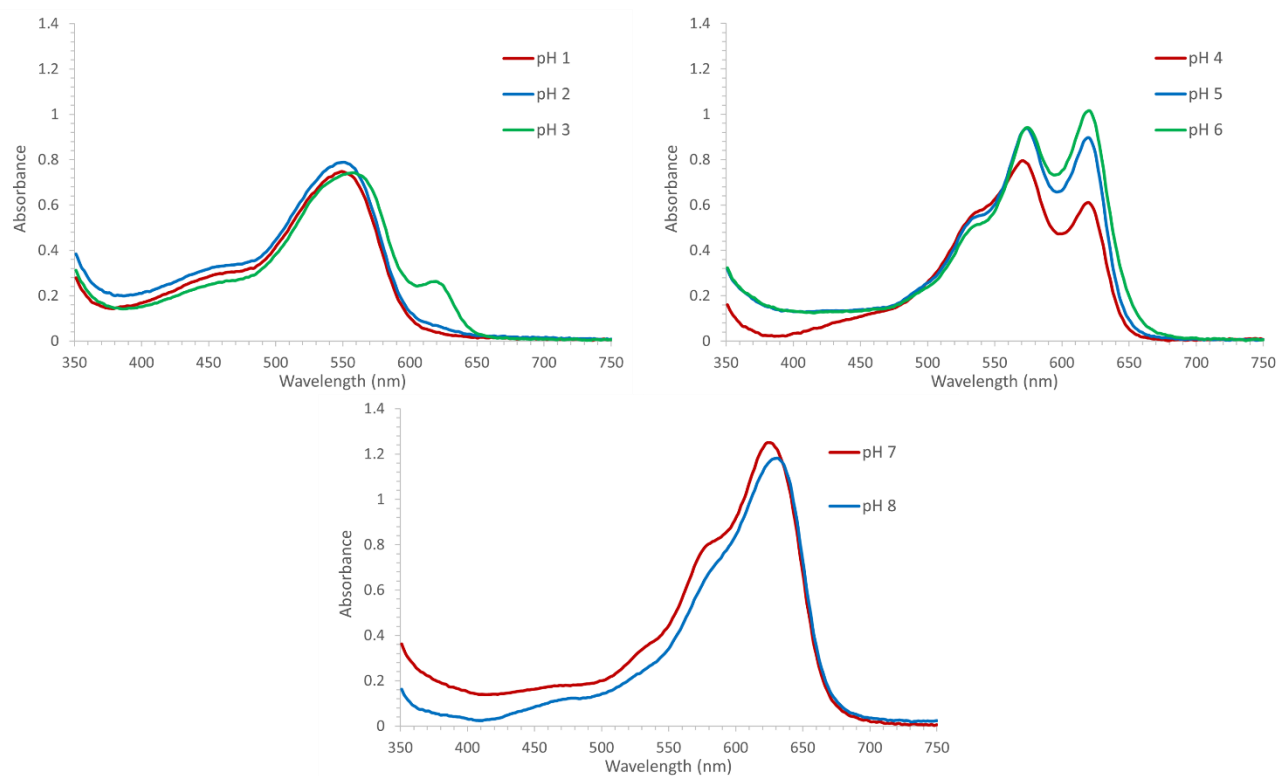
Appendix 3.5 UV-VIS absorbance spectra of ternatinB3 (60 μM) between pH 1 and 8.
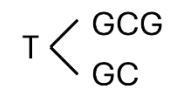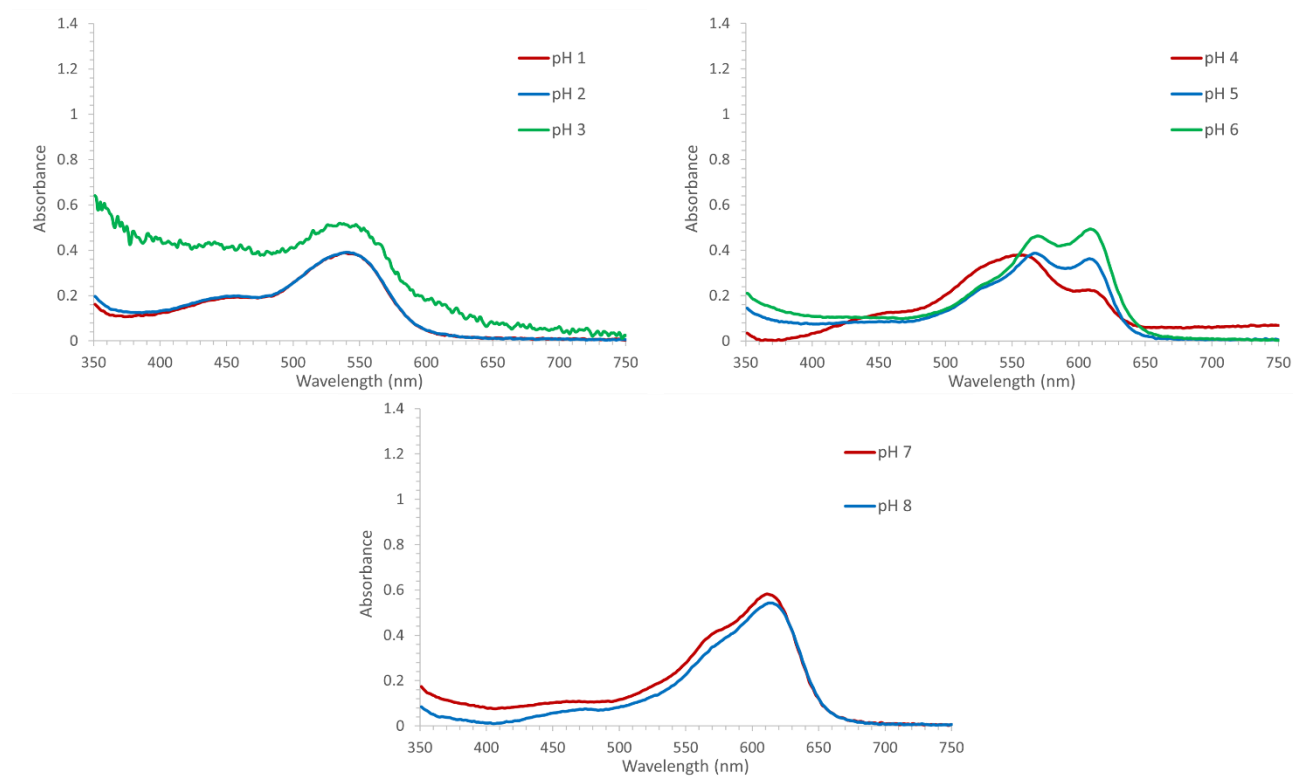
Appendix 3.6 UV-VIS absorbance spectra of ternatin B4 (60 µM) between pH 1 and 8.
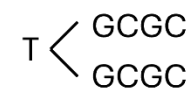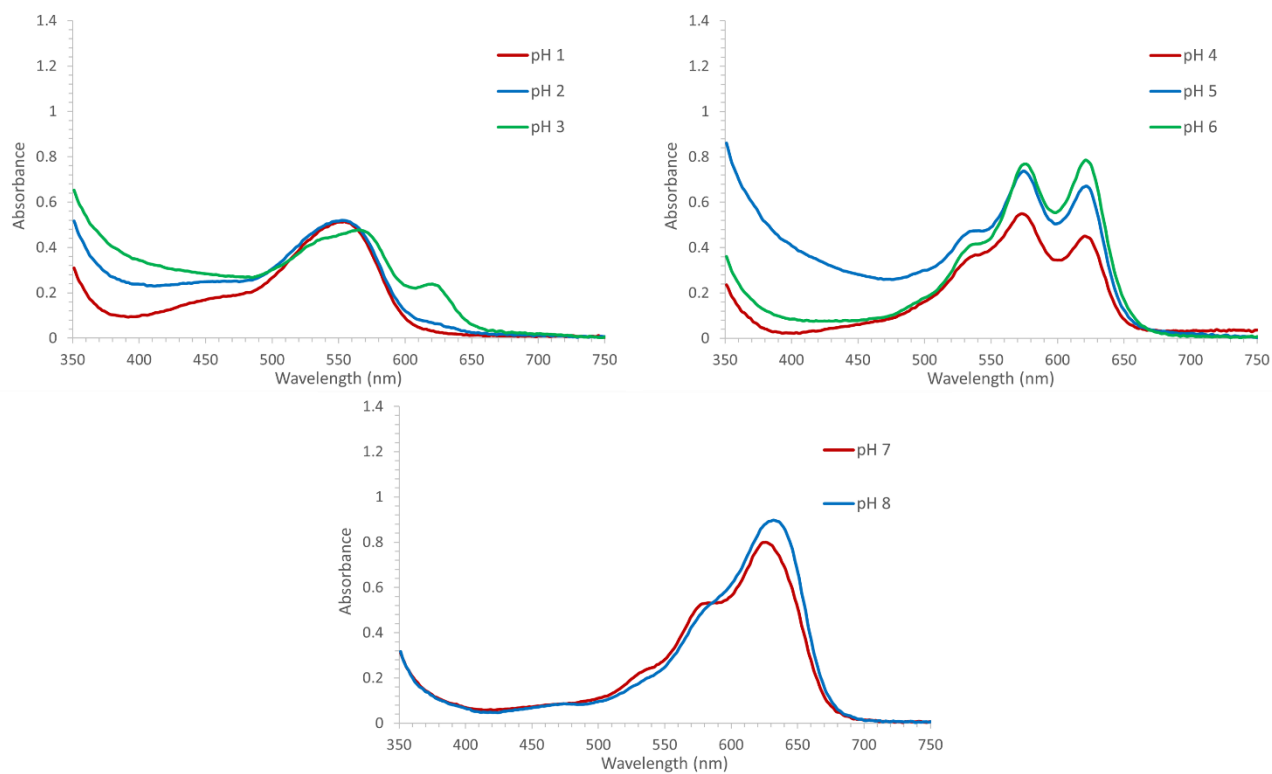
Appendix 3.7 UV-VIS absorbance spectra of ternatin D1 (60 µM) between pH 1 and 8.
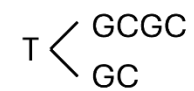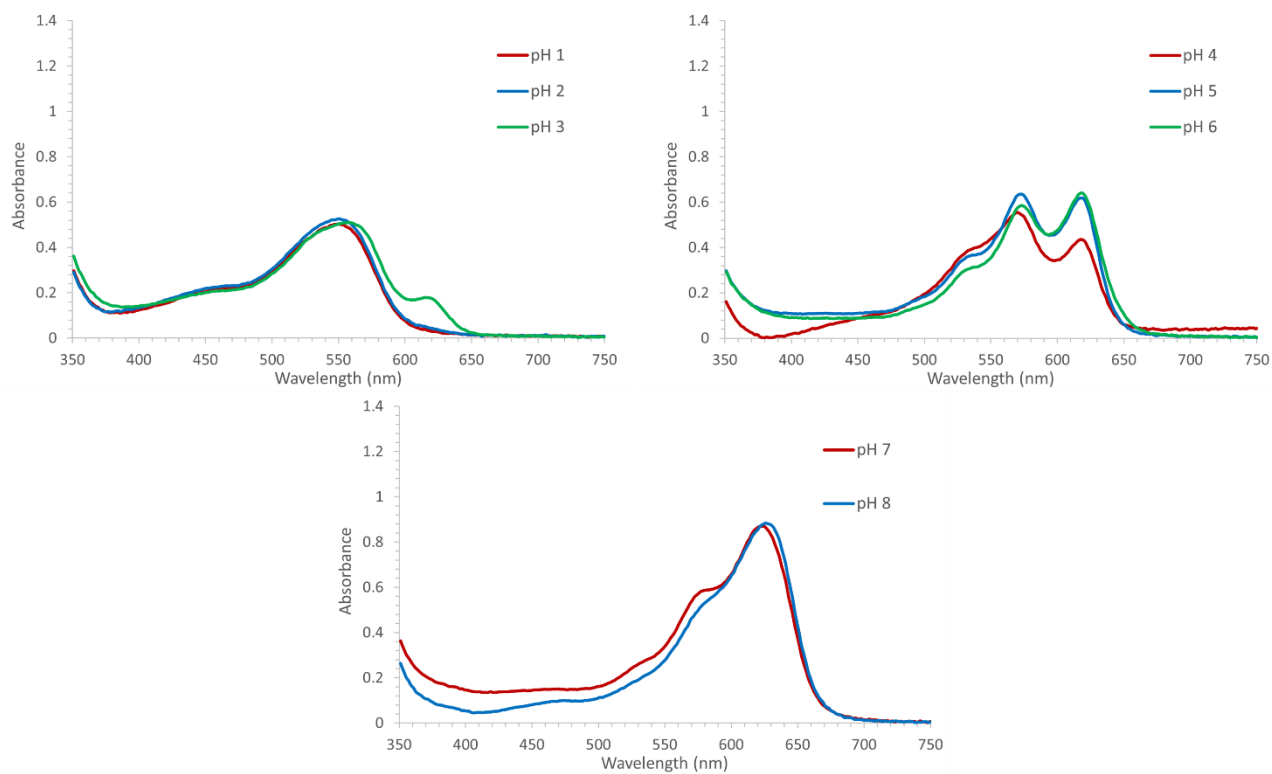
Appendix 3.8 UV-VIS absorbance spectra of ternatin D2 (60 µM) between pH 1 and 8.

Appendix 4.1 Topology_Processor.py

```python
1.  """
2.  Python -v 3.10 UTF-8
3.  Author: Alan Houghton
4.  Date Created: 26/11/22
5.
6.  Acpype TIP3P Topology formatter
7.
8.  - Reads in the *GMX.itp and *GMX.top files generated by acpype and formats
    them into a single *.top file
9.  - with TIP3P water model atomtypes and moleculetypes data.
10.
11. """
12.
13. import argparse
14. import sys
15. import os
16.
17.
18. # Counter Ions:
19.
20. chloride = [
21.     [
22.         " CL        CL            0.00000  0.00000   A     4.30000e-
    01  4.20000e-01"
23.     ],
24.     [
25.         "[ moleculetype ]",
26.         "; molname        nrexcl ; Loche et al. 2021 Transferable Ion Force
    Fields in Water...",
27.         "   Cl              1",
28.         "",
29.         "[ atoms ]",
30.         ";   nr   type  resnr residue  atom    cgnr    charge    mass",
31.         "    1     CL     1      Cl    CL      1       -1       35.453"
32.     ]
33. ]
34.
35. bromide = [
36.     [
37.         " BR        BR            0.00000  0.00000   A     4.430000e-
    01  7.50000e-01"
38.     ],
39.     [
40.         "[ moleculetype ]",
41.         "; molname        nrexcl ; Loche et al. 2021 Transferable Ion Force
    Fields in Water...",
42.         "   Br              1",
43.         "",
44.         "[ atoms ]",
45.         ";   nr   type  resnr residue  atom    cgnr    charge    mass",
46.         "    1     BR     1      Br    BR      1       -1       22.9898"
47.     ]
48. ]
49.
50. sodium = [
51.     [
52.         " NA        NA            0.00000  0.00000   A     2.31000e-
    01  4.50000e-01"
53.     ],
54.     [
55.         "[ moleculetype ]",
```

```
56.          "; molname        nrexcl ; Loche et al. 2021 Transferable Ion Force
     Fields in Water...",
57.          "  Na             1",
58.          "",
59.          "[ atoms ]",
60.          ";   nr    type   resnr residue   atom     cgnr     charge      mass",
61.          "    1     NA      1     Na    NA      1      1       22.9898"
62.      ]
63. ]
64.
65. potassium = [
66.      [
67.          " NA       NA          0.00000  0.00000   A     2.31000e-
     01   4.50000e-01"
68.      ],
69.      [
70.          "[ moleculetype ]",
71.          "; molname        nrexcl ; Loche et al. 2021 Transferable Ion Force
     Fields in Water...",
72.          "  Na             1",
73.          "",
74.          "[ atoms ]",
75.          ";   nr    type   resnr residue   atom     cgnr     charge      mass",
76.          "    1     NA      1     Na    NA      1      1       22.9898"
77.      ]
78. ]
79.
80.
81. # TIP3P Water
82.
83. water = [
84.      [
85.          " OW       OW          0.00000  0.00000   A     3.15075e-
     01   6.35968e-01",
86.          "
     HW       HW          0.00000  0.00000   A     0.00000e+00   0.00000e+00"
87.      ],
88.      [
89.          "[ moleculetype ]",
90.          "; molname       nrexcl ; TIP3P model",
91.          "  SOL            2",
92.          "",
93.          "[ atoms ]",
94.          ";   nr     type    resnr   residue atom    cgnr     charge      mass",
95.          "    1     OW      1     SOL    OW     1      -0.834   16.00000",
96.          "    2     HW      1     SOL    HW1    1       0.417    1.00800",
97.          "    3     HW      1     SOL    HW2    1       0.417    1.00800",
98.          "",
99.          "[ bonds ]",
100.            "; i       j        funct   length  force_constant",
101.            "1       2       1       0.09572
     502416.0   0.09572      502416.0",
102.            "1       3       1       0.09572
     502416.0   0.09572      502416.0",
103.            "",
104.            "[ angles ]",
105.            "; i       j       k       funct    angle   force_constant",
106.            "2       1       3       1       104.52 628.02       104.52
     628.02"
107.        ]
108.    ]
109.
110.
111.    def cmd_lineparser(*args: list):
112.
113.        parser = argparse.ArgumentParser(prog="AcpypeTIP3P",
```

250

```
114.                                          add_help=False,
115.                                          formatter_class=argparse.RawTex
     tHelpFormatter
116.                                      )
117.          group_inputs = parser.add_argument_group("Inputs")
118.
119.          group_inputs.add_argument("-i", "--itp_file",
120.                                    type=str,
121.                                    action="store",
122.                                    help="Path to *_GMX.itp file cotaining
     molecule data generated by acpype",
123.                                    metavar="\b",
124.                                    default=None
125.                                    )
126.
127.          group_inputs.add_argument("-t", "--top_file",
128.                                    type=str,
129.                                    action="store",
130.                                    help="Path to *_GMX.top file cotaining
     molecule data generated by acpype",
131.                                    metavar="\b",
132.                                    default=None
133.                                    )
134.
135.          group_optional = parser.add_argument_group("Optional")
136.
137.          group_optional.add_argument("-w", "--add_TIP3P",
138.                                      action="store_true",
139.                                      help="Add the TIP3P water to the
     force field, and generate molecule section.\n"
140.                                      "Warning: This script can
     detect whether the section already exisits\n"
141.                                      "Defaults to False")
142.
143.          group_optional.add_argument("-s", "--ion_types",
144.                                      type=str,
145.                                      nargs="*",
146.                                      action="store",
147.                                      default=None,
148.                                      choices=["na", "k", "cl", "br"],
149.                                      help="Add counter ions into the
     force field, and generate a molecule type section.\n"
150.                                      "You can pick from Potassium
     (k), Sodium (na), Chloride (cl) or Bromide (br) "
151.                                      "(or a combination of these
     (e.g. [ -s na k cl br ] ) "
152.                                      )
153.
154.          args = parser.parse_args(*args)
155.
156.          return args
157.
158.      def insert_atomtypes_topology(topology_lines: list, lines_to_add:
     list):
159.
160.          # Reverse order of the lines_to_add to preserve index
161.          lines_to_add.reverse()
162.
163.          # Enumerate through the itp file until [ atomtypes ] section is
     found
164.          atom_type_flag = False
165.          for i, line in enumerate(topology_lines):
166.
167.              if line == "[ atomtypes ]\n":
168.                  atom_type_flag = True
169.
```

```
170.              if atom_type_flag == True and line == "\n":
171.
172.                  # Add items, making sure that the string ends with \n
173.                  for item in lines_to_add:
174.                      if item.endswith("\n"):
175.                          topology_lines.insert(i, item)
176.                      else:
177.                          topology_lines.insert(i, f"{item}\n")
178.
179.                  return topology_lines
180.
181.          raise ValueError(f"Couldn't find the [ atomtypes ] section in
    {topology_lines}")
182.
183.      def insert_moleculetype_topology(topology_lines: list, list_to_add:
    list):
184.
185.          list_to_add.reverse()
186.
187.          system_flag = False
188.
189.          for i, line in enumerate(topology_lines):
190.
191.              # Locate the system header
192.              if "system" in line:
193.                  system_flag = True
194.                  system_index = i
195.
196.              if system_flag:
197.                  for spacer in ["\n", "\n"]:
198.                      topology_lines.insert(system_index, spacer)
199.
200.                  for add_line in list_to_add:
201.                      if add_line.endswith("\n"):
202.                          topology_lines.insert(system_index, add_line)
203.                      else:
204.                          topology_lines.insert(system_index,
    f"{add_line}\n")
205.
206.                  return topology_lines
207.
208.          raise ValueError(f"Couldn't locate the [ system ] header in
    {topology_lines}")
209.
210.
211.
212.      def main(top=None, itp=None, ion_type=None, add_TIP3P=False):
213.
214.          sys_path = sys.path[0]
215.          data_path = f"{sys_path}{os.sep}Data"
216.          output_path = f"{sys_path}{os.sep}Output"
217.
218.          # ================ #
219.
220.          # Check user inputs
221.
222.          if top is None:
223.              raise FileNotFoundError(f"Missing *.top file: {top}")
224.
225.          if ion_type is not None:
226.              for ion in ion_type:
227.                  if ion not in ["na", "k", "cl", "br"]:
228.                      raise ValueError(f"Ion type {ion_type} is not
    supported. [ -s ] must be na, k, cl or br "
229.                                       f"(or a list of these values)")
230.
```

```
231.          if not os.path.isdir(output_path):
232.              os.mkdir(output_path)
233.
234.          # ================ #
235.
236.          # Read user submitted files
237.
238.          # Read Molecule ITP file
239.
240.          if itp:
241.              with open(f"{itp}") as f:
242.                  itp = f.readlines()
243.                  f.close()
244.
245.          # Read Molecule Topology File
246.          with open(f"{top}") as f:
247.              topology = f.readlines()
248.              f.close()
249.
250.          # ================ #
251.
252.          # Find counter ion data (if required)
253.
254.          ions = None
255.
256.          if ion_type is None:
257.              pass
258.          else:
259.              search_ions = {
260.                  "na": sodium,
261.                  "k": potassium,
262.                  "cl": chloride,
263.                  "br": bromide
264.              }
265.
266.              ions = []
267.
268.              for value in ion_type:
269.                  if value in search_ions.keys():
270.                      ions.append(search_ions[value])
271.
272.          ########################
273.
274.          # Start modifying topology
275.
276.          # If replacing #include *.itp statements
277.          if itp is not None:
278.
279.              if add_TIP3P:
280.                  itp = insert_atomtypes_topology(itp, water[0])
281.
282.              if ions:
283.                  for ion in ions:
284.                      itp = insert_atomtypes_topology(itp, ion[0])
285.
286.          # Otherwise we're altering the topology file
287.          else:
288.
289.              if add_TIP3P:
290.                  topology = insert_atomtypes_topology(topology, water[0])
291.
292.              if ions:
293.                  for ion in ions:
294.                      topology = insert_atomtypes_topology(topology,
        ion[0])
295.
```

```
296.            # Position to insert itp data
297.            system_flag = False
298.
299.            # Replace the #include topology section of the topology file
     with the whole itp file
300.            if itp is not None:
301.
302.                # Reverse itp to preserve indexing
303.                itp.reverse()
304.
305.                for i, line in enumerate(topology):
306.
307.                    # Insert molecule data into topol file
308.                    if "include" and "topology" in line:
309.                        _ = topology.pop(i)
310.                        _ = topology.pop(i)
311.                        for item in itp:
312.                            topology.insert(i, item)
313.                        break
314.
315.            if add_TIP3P:
316.                topology = insert_moleculetype_topology(topology, water[1])
317.
318.            if ions:
319.                for ion in ions:
320.                    topology = insert_moleculetype_topology(topology,
     ion[1])
321.
322.
323.            output = "".join(topology)
324.
325.            if os.path.isdir(output_path):
326.                out_name = f"{output_path}{os.sep}{top.split(os.sep)[-1]}"
327.            else:
328.                out_name = output_path
329.
330.            with open(f"{out_name}", "w") as f:
331.                f.write(output)
332.
333.            return
334.
335.        if __name__ == "__main__":
336.            args = cmd_lineparser()
337.
338.            top = args.top_file
339.            itp = args.itp_file
340.            add_water = args.add_TIP3P
341.            add_ions = args.ion_types
342.
343.            main(top, itp, add_ions, add_water)
```

Appendix 4.2 Plumed_HREX_TopoProc.py

```python
1.  """
2.  Python=3.11 (UTF-8)
3.
4.  Author: Alan Houghton
5.  Date Created: 22/12/22
6.  Date Last Modified: 27/12/22
7.
8.  Version 1.0
9.
10. Usage:
11.
12. python3 ./Plumed_HREX_TopoProc.py -i [path to input *.top file] -o [path of
    output *.top file]
13.
14.     for use in bash script (assuming that the molecule to process is the
    first to appear in *.top file):
15.
16. printf 0 | python3 ./Plumed_HREX_TopoProc.py -i [path to input *.top file]
    -o [path of output *.top file]
17.
18.
19. About:
20.
21. Plumed_HREX_TopoProc is a python script for processing Gromacs topology
    files (*.top) for Hamiltonian replica exchange
22. simulations (HREX simulation) prior to the plumed partial_tempering step
    (where the documentation says to use
23. gmx grompp -pp .
24.
25. The script locates all [ moleculetypes ] within the *.top file, and allows
    the user to select a molecule group by index
26.  to flag for partial tempering by appending an underscore to the atom type
    in the [ atoms ] section ( e.g. ca -> ca_ ).
27.  The index follows the order of the [ moleculetypes ] in the *.top file, in
    that index 0 is the first listed molecule
28.  in the *.top file.
29.
30. """
31.
32. import argparse
33. import pandas as pd
34. import os
35.
36.
37. def read_file(file_path):
38.     """
39.     Read the contents of a file into a list of lines.
40.
41.     :param file_path: The path of the file to read
42.     :return lines: A list of lines from the file
43.     """
44.     with open(file_path) as f:
45.         lines = f.readlines()
46.     return lines
47.
48.
49. # Extract a section based on the section header ( e.g. [ atoms ] )
50. # Returns a list of dictionaries for each molecule found
51. def extractSection(lines, section_name):
52.     """
53.      Extract a section from a list of lines based on the section name.
54.
55.      :param lines: A list of lines from a Gromacs topology file
```

```
56.        :param section_name: The name of the section to extract (e.g. "atoms",
    "bonds", etc.)
57.        :return output: A list of dictionaries containing the extracted
    sections
58.        """
59.
60.     # List of accepted section names
61.     accepted_sections = ["moleculetype", "atoms", "atomtypes",
62.                          "bonds", "paris", "angles",
63.                          "dihedrals", "system", "molecules"]
64.
65.     # Check if the provided section name is valid
66.     if section_name not in accepted_sections:
67.         raise TypeError(f"extractSection: section_name must be in the
    accepted_sections list:\n{accepted_sections}\n"
68.                        f"User input: {section_name}")
69.
70.     # Initialize a list to store the extracted sections
71.     output = []
72.
73.     # Iterate through the lines
74.     for i, line in enumerate(lines):
75.
76.         # Check if the current line is the start of the desired section
77.         if f"[ {section_name} ]" in line:
78.             # Determine the start and end indices of the section
79.             start_index = i
80.             end_index = None
81.             idx = start_index + 1
82.             while end_index is None:
83.                 if lines[idx].startswith("\n") or lines[idx].startswith("[
    "):
84.                     end_index = idx
85.                 else:
86.                     idx += 1
87.
88.                 # If the end of the section could not be found, raise an
    error
89.                 if idx > (len(lines) - 1):
90.                     raise IndexError(f"Couldn't find the end of
    {section_name} in the provided list...\n"
91.                                      f"Started search at:
    {lines[start_index]} : (0) index {start_index}")
92.
93.             # Extract the name of the section
94.             name = __getMoleculeName(lines, start_index, section_name)
95.
96.             # Extract the section from the list of lines
97.             section = lines[start_index: end_index]
98.
99.             # Combine the name, start and end indices, and lines into a
    dictionary
100.                combined_section = {
101.                    "name": name,
102.                    "start_index": start_index,
103.                    "end_index": end_index,
104.                    "section": section
105.                }
106.
107.                # Append the dictionary to the list of extracted
    sections
108.                output.append(combined_section)
109.
110.        return output
111.
112.
```

```python
113.        # Returns name found in the moleculetype section
114.        def __getMoleculeName(lines, index, section_name):
115.            """
116.            Extract the name of a molecule from the [ moleculetype ] section
    of a Gromacs topology file.
117.            This function starts at the index position of the section_name
    in extractSection, and searches
118.            upstream (reverse) until it finds the [ moleculetype ] section.
    If the section_name is [ atomtypes ],
119.            then search direction is reversed.
120.
121.            :param lines: A list of lines from a Gromacs topology file
122.            :param index: The index of the line from which to start
    searching for the
123.            :param atomtypes: If true, will reverse the search direction (as
    moleculetype section is downstream from atomtypes)
124.            """
125.
126.            # Determine the direction to search for the section header
127.            if f"[ {section_name} ]" == "[ atomtypes ]":
128.                search_direction = 1
129.            else:
130.                search_direction = -1
131.
132.            start_index = None
133.            end_index = None
134.            name = None
135.
136.            # Search for the [ moleculetype ] header
137.            while index > 0:
138.                # Determine the start and end indices of the section
139.                if "[ moleculetype ]" in lines[index]:
140.                    start_index = index
141.                    idx = start_index + 1
142.                    while end_index is None:
143.                        # Determine the start and end indices of the section
144.                        if lines[idx].startswith("\n") or
    lines[idx].startswith("[ "):
145.                            end_index = idx
146.                            break
147.                        else:
148.                            idx += 1
149.                    break
150.                else:
151.                    index += search_direction
152.
153.            # Extract the section from the list of lines
154.            section = lines[start_index: end_index]
155.            for item in section:
156.                if item.startswith("[ ") or item.startswith(";") or
    item.startswith("\n"):
157.                    pass
158.                else:
159.                    name = item.split(" ")
160.
161.            # Iterate through the list of words to find the first non-empty
    word
162.            for item in name:
163.                if item == "":
164.                    pass
165.                else:
166.                    name = item
167.                    break
168.
169.            return name
170.
```

```
171.
172.        # Remove all matching values form list
173.        def remove_values_from_list(the_list, val):
174.            return [value for value in the_list if value != val]
175.
176.
177.        # Convert lines from topology section to a pandas dataframe
178.        def sectionToDataframe(section):
179.            """
180.                Convert a section from a Gromacs topology file into a Pandas
       dataframe.
181.
182.                :param section: A list of lines from a section of a Gromacs
       topology file
183.                                (optional) If the dictionary returned from
       function extractSection is provided, the dictionary
184.                                will be updated with the dataframe and a
       "dataframe" key will be added to the dictionary to
185.                                indicate that the data has been converted to
       a dataframe
186.                :return dataframe: A Pandas dataframe containing the data
       from the section
187.            """
188.
189.            if isinstance(section, dict):
190.                section_dict = section
191.                section = section_dict["section"]
192.
193.            else:
194.                section_dict = None
195.
196.            # Check that input is a list
197.            if not isinstance(section, list):
198.                raise TypeError(f"sectionToDataframe {type(section)} is not
     a list:\n"
199.                                f"\tSection must be a raw text list from
     extractSection")
200.
201.            # remove the section header ( [ head ] ) - not used
202.            _ = section.pop(0)
203.
204.            # Initialize an empty list to store the values from the section
205.            output_list = []
206.
207.            # Extract the values from the section
208.            for item in section:
209.                lines = item.split("\n")
210.                for line in lines:
211.                    # Split the line into a list of values and remove empty
       values
212.                    val = remove_values_from_list(line.split(" "), "")
213.                    # Add the list of values to the output list if it is not
       empty
214.                    if val:
215.                        output_list.append(val)
216.
217.            # Extract the headers from the first row of the output list
218.            headers = output_list.pop(0)
219.            # Check if the first header starts with a semicolon
220.            if headers[0] == ";":
221.                # Remove the semicolon from the first header
222.                _ = headers.pop(0)
223.                # Add the semicolon back to the first header
224.                headers[0] = f"; {headers[0]}"
225.
```

```
226.            # Pad the headers list with empty strings if it is shorter than
     the first row of the output list
227.            if len(headers) < len(output_list[0]):
228.                while len(headers) < len(output_list[0]):
229.                    headers.append("")
230.
231.            # Create a dataframe from the output list using the headers as
     column names
232.            dataframe = pd.DataFrame(output_list, columns=headers)
233.
234.            if section_dict:
235.                section_dict["section"] = dataframe
236.                section_dict["dataframe"] = True
237.                # Return the updated dictionary
238.                return section_dict
239.            else:
240.                # Return the dataframe if the update_dict argument was not
     provided
241.                return dataframe
242.
243.
244.        def processTopology(atoms_dataframe):
245.            """
246.            Modify the "type" column of an atoms dataframe from a Gromacs
     topology file.
247.
248.            :param atoms_dataframe: A Pandas dataframe containing the
     "atoms" section of a Gromacs topology file. If the
249.                                    dataframe is wrapped in a dictionary
     with a "dataframe" key set to True, the dictionary
250.                                    will be updated with the modified
     dataframe
251.            :return atoms_dataframe: The modified atoms dataframe or a
     dictionary containing the modified atoms dataframe
252.                                    with a "section" key
253.            """
254.
255.            # Check if atoms_dataframe is a dictionary with a "dataframe"
     key
256.            if isinstance(atoms_dataframe, dict) and
     atoms_dataframe["dataframe"] == True:
257.                output_dictionary = atoms_dataframe
258.                atoms_dataframe = atoms_dataframe['section']
259.
260.            else:
261.                output_dictionary = None
262.
263.            # Modify the "type" column of the atoms dataframe
264.            atom_type = atoms_dataframe['type'].to_numpy()
265.            for i, value in enumerate(atom_type):
266.                atom_type[i] = f"{value}_"
267.            atoms_dataframe['type'] = atom_type
268.
269.            # Return the modified dataframe as a dictionary if it was
     originally a dictionary, otherwise return the dataframe
270.            if output_dictionary:
271.                output_dictionary['section'] = atoms_dataframe
272.                return output_dictionary
273.            else:
274.                return atoms_dataframe
275.
276.
277.        def dataframeToLines(dataframe):
278.            """
279.            Convert a Pandas dataframe into a list of lines, including the
     column headers.
```

```python
280.
281.            :param dataframe: A Pandas dataframe
282.            :return lines: A list of strings, each string representing a
        line in the dataframe
283.            """
284.            # Convert the dataframe to a string
285.            dataframe_string = dataframe.to_string(header=False,
        index=False)
286.            # Split the string by line breaks
287.            lines = dataframe_string.split("\n")
288.            # Remove empty lines
289.            lines = [f"{line}\n" for line in lines if line]
290.            # Insert the column headers at the beginning of the list
291.            columns = "   ".join(dataframe.columns) + "\n"
292.            lines.insert(0, columns)
293.            return lines
294.
295.
296.        def dataframeToString(dataframe):
297.            """
298.            Convert a Pandas dataframe into a string.
299.
300.            :param dataframe: A Pandas dataframe
301.            :return dataframe_string: A string containing the data from the
        dataframe
302.            """
303.            # Convert the dataframe to a list of lines
304.            lines = dataframeToLines(dataframe)
305.            # Join the lines into a single string
306.            return "\n".join(lines)
307.
308.
309.        def replaceLines(lines, replacement_lines, start_index):
310.            """
311.            Replace a range of lines in a list with a new list of lines.
312.
313.            :param lines: The original list of lines
314.            :param replacement_lines: The list of lines to insert
315.            :param start_index: The index at which to start replacing lines
        in the original list
316.            :return lines: The modified list of lines
317.            """
318.
319.            # Reverse the order of the replacement lines
320.            replacement_lines.reverse()
321.
322.            # Calculate the number of lines to remove from the original list
323.            num_lines_to_remove = len(replacement_lines)
324.
325.            # Remove the lines to be replaced from the original list
326.            for i in range(num_lines_to_remove):
327.                lines.pop(start_index)
328.
329.            # Insert the replacement lines into the original list
330.            for line in replacement_lines:
331.                lines.insert(start_index, line)
332.            return lines
333.
334.
335.        def writeLinesToFile(lines, file_path):
336.            """
337.            Write a list of lines to a file.
338.
339.            :param lines: The list of lines to write to the file
340.            :param file_path: The path to the file
341.            """
```

```
342.          # Open the file in write mode
343.          with open(file_path, "w", newline="\n") as f:
344.              # Write the lines to the file
345.              f.writelines(lines)
346.
347.
348.      def getName(molecule_dict):
349.          return molecule_dict['name']
350.
351.
352.      def getDataframe(molecule_dict):
353.          if 'dataframe' in molecule_dict.keys() and
   molecule_dict['dataframe'] == True:
354.              return molecule_dict['dataframe']
355.          else:
356.              raise AttributeError(f"{molecule_dict} has no dataframe
   assigned. Use ")
357.
358.
359.      def cmd_lineparser(*args: list):
360.          parser = argparse.ArgumentParser(prog="Plumed_HREX_Processor",
361.                                           add_help=False,
362.                                           formatter_class=argparse.RawTex
   tHelpFormatter
363.                                           )
364.
365.          group_inputs = parser.add_argument_group("Inputs")
366.
367.          group_inputs.add_argument("-i", "--top_file",
368.                                    required=True,
369.                                    type=str,
370.                                    action="store",
371.                                    help="Path to *.top file to process",
372.                                    metavar="\b",
373.                                    default=None
374.                                    )
375.          group_outputs = parser.add_argument_group("Output")
376.
377.          group_outputs.add_argument("-o", "--output",
378.                                     required=True,
379.                                     type=str,
380.                                     action="store",
381.                                     help="Name of the output file",
382.                                     metavar="\b",
383.                                     default=None
384.                                     )
385.
386.          return parser.parse_args(*args)
387.
388.
389.      def main():
390.          args = cmd_lineparser()
391.          input_path = args.top_file
392.          output_path = args.output
393.
394.          print("Input path:", input_path)
395.          print("Output path:", output_path)
396.
397.          raw_input = None
398.
399.          if os.path.isfile(input_path):
400.              raw_input = read_file(input_path)
401.
402.          unformatted_sections = extractSection(raw_input, "atoms")
403.
404.          if len(unformatted_sections) > 1:
```

261

```
405.                print("Multiple molecule types found... "
406.                    "\nPlease select which groups to apply processing tag
       to:")
407.                print("\nIndex : Molecule Name")
408.
409.                for i, item in enumerate(unformatted_sections):
410.                    print(f"{i} : {unformatted_sections[i]['name']}")
411.
412.                # todo custom groupings? e.g. mol and sol
413.
414.                selection = input()
415.
416.                if not isinstance(selection, str) and selection not in
       ["0123456789"]:
417.                    raise IndexError
418.
419.                selected_section = unformatted_sections[int(selection)]
420.
421.                print(f"\nSelected: {selection} :
       {selected_section['name']}")
422.
423.            else:
424.                selected_section = unformatted_sections[0]
425.
426.            if isinstance(selected_section, list):
427.                raise NotImplementedError("Can't yet process two groups at
       once, run this twice for each group")
428.
429.            else:
430.                selected_molecules = sectionToDataframe(selected_section)
431.
432.            processed_molecule = processTopology(selected_molecules)
433.
434.            new_lines = dataframeToLines(processed_molecule['section'])
435.
436.            formatted_lines = replaceLines(raw_input, new_lines,
       processed_molecule["start_index"] + 1)
437.
438.            # Write the formatted string to the output file
439.            writeLinesToFile(formatted_lines, output_path)
440.
441.
442.        if __name__ == "__main__":
443.            main()
```