Identifying cellular targets of 3',4',5',5,7-pentamethoxyflavone in colorectal cancer

By
Emily Friar

Division of Natural Sciences, University of Kent

A thesis submitted for the degree of Doctor of Philosophy
February 2024

Page count: 398

Abstract

Despite increasing positive prognosis for those living in high income countries (HICs), the incidence of colorectal cancer (CRC) in low- and middle-income countries (LMICs) such as Malaysia, is rising and increasingly becoming a burden for people in these regions. Natural products with promising bioactive properties have been widely investigated for clinical usage and often show positive properties such as reduced toxicity. Flavonoids are polyphenol phytochemicals that often possess bioactive properties. Flavones, a sub-class of the flavonoid family, have shown promise as potential drug compounds in areas such as anti-inflammatory, anti-obesity, antioxidant, and anticancer medicine. 3',4',5',5,7-Pentamethoxyflavone (PMF) has been shown in previous research to possess anticancer and chemopreventive properties in CRC, this research aims to identify potential cellular targets of PMF to give insight into the mechanisms of action of the compound which are currently unknown. A structure-activity relationship study identified the 3'-position as a suitable modification point on the compound structure that could be changed without compromising biological activity. After synthesis of a photoaffinity labelling probe, protein cellular targets were then investigated using pull down experiments and mass spectrometry proteomics. RNA sequencing was also used to investigate changes in gene expression in CRC cells after treatment with the active compound to further aid in identifying a mechanism of action. Members of the RAB subfamily of small GTPases were isolated using the proteomics analysis. Pathway analysis showed that the key interacting proteins were involved in the cellular response to stress and protein folding. This suggests that the flavone could be involved in the unfolded protein response (UPR). RNA sequencing showed that exposure of the cells to the flavone compound caused changes in expression of genes associated with the cell cycle and the UPR. Future studies of PMF should aim to validate the interaction between the protein targets identified and the compound. Synthetic modification of PMF can improve both the binding properties to the target proteins and aqueous solubility of the compound and enable progression as a potential anticancer treatment.

Declaration

I declare that all the work presented in this thesis submission was my own work and written in my own words. It has not been submitted to any other institution for a qualification or for any other degree.

Emily Friar

February 2024

COVID Mitigation

Work for this thesis commenced in September 2019 and the COVID-19 pandemic caused disruption to the work from March 2020. Lockdown prevented access to the research laboratories entirely for four months and access was reduced until April 2021 implemented by a rota system. This resulted in the total loss of approximately seven months of laboratory access. Opportunities to travel to interact and research with collaborators at UNIMAS Malaysia were lost due to the pandemic, and the resulting reduction of consumables budget caused by the government budget cuts during the pandemic. There was also disruption to equipment and consumables access that hindered efficient progress during and after the times indicated above. Collectively these problems significantly hindered the progress of the research and funding extensions were not able to be obtained.

Dr Christopher Serpell Prof Michelle Garrett

Acknowledgements

The work presented in this thesis would not have been possible without, firstly, funding from the Global Challenges Doctoral Centre, University of Kent and support from our collaborators Taylor's University and Dr Isabel Lim Fong. Secondly, without my two supervisors, Professor Michelle Garrett and Dr Christopher Serpell. They provided me with guidance and support throughout the years and offered me compassion and understanding when I needed it the most.

A thank you is given to the members of both the Garrett and Serpell research groups, with a special mention for Edith Blackburn, without whom my practical biology skills would still amount to zero, and to Geraud Sansom, who *almost* never doubted my ability to cross even some of the toughest hurdles I faced. Most importantly however, a thank you to my companion throughout the total rollercoaster that was this journey, Bini Claringbold, there is no doubt that without your support (and kind post-it notes) I would not have reached this point.

I would also like to give thanks to my friends in both my PGR community and beyond, importantly my fellow women in STEM, for consistently being a place to celebrate even the smallest wins, and to work through some of the lows.

My greatest thanks though, goes to my family, for never wavering in their unparalleled support for me, for all the love and encouragement they have given me, without which I could have never achieved this. I am so incredibly grateful.

Table of Contents

Abbreviations

[%w/w]	Percent weight by weight
°C	Degrees Celsius
µ	Micro
^{13}C	Carbon-13
^{1}H	Hydrogen-1
4-HNE	4-hydroxy-2-nonenal
ABC	ATP-binding cassette
Abs	Absorbance
ACD	Advanced Chemistry Development (NMR)
AML	Acute myeloid leukemia
Apc	Adenomatous polyposis coli
ASPSCR1	UBX domain containing tether for SLC2A4
ASPSCR1	UBX domain containing tether for SLC2A4
ATP	Adenosine triphosphate
AU	Absorbance units
Bn	Benzyl
BSA	Bovine serum albumin
C/EBP-α	CCAAT enhancer binding protein alpha
C/EBP-β	CCAAT enhancer binding protein beta
c/w	Cells per well
Calcd.	Calculated
CAND1	Cullin associated and neddylation dissociated 1
CCCK-8 assay	Cell counting kit 8 assay
Cck	Cell counting kit
CD	Cyclodextrin
$CDCl_3$	Deuterated chloroform
CDDP	Cisplatin
CDK	Cyclin-dependent kinases
cDNA	Complementary DNA
CDS	Chromatography Data System
Conc	Concentration
COSY	Correlated Spectroscopy (NMR)
COX	Cyclooxygenase
CRC	Colorectal cancer
CuAAC	Cu(I)-catalyzed azide–alkyne cycloaddition
d	Doublet
DCK	Deoxycytidine kinase
DCM	Dichloromethane
dd	Doublet of doublets
DEGs	Differentially expressed genes
DMAP	4-Dimethylaminopyridine
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DMSO-d_6	Deuterated DMSO
DNA	Deoxyribonucleic acid
dRNA-seq	Direct RNA sequencing
DTT	Dithiothreitol
ECL	Enhanced Chemiluminescence
EDC.HCl	N-Ethyl-N′-(3-dimethylaminopropyl)carbodiimide hydrochloride
EDTA	Ethylenediaminetetraacetic acid
eq	Equivalents
ER	Endoplasmic reticulum
ERGIC	ER-Golgi intermediate compartment
ESI-HRMS	Electrospray ionisation high resolution mass spectrometry
ESI-MS	Electrospray ionisation mass spectrometry
EtOAc	Ethyl Acetate
FAP	familial adenomatous polyposis
FBS	Fetal bovine serum
FC	Fold change
FITC	Fluorescein isothiocyanate
FLAM	Fludarabine, cytarabine, and mitoxantrone

g	Grams
GCLC	Glutamate-Cysteine Ligase Catalytic Subunit
GI_{50}	Concentration that reduced the cell viability by 50% when compared to untreated controls
GPCR	G-protein coupled receptor
GTP	Nucleotide guanosine triphosphate
h	Hours
H	hydrogens
HCl	Hydrochloric acid
HDI	Human development index
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)
Hex	Hexane
HIC	High income countries
HMBC	Heteronuclear Multiple Bond Correlation (NMR)
HSQC	Hetereonuclear Single Quantum Coherence (NMR)
HO-1	Heme Oxygenase 1
HOBt	Hydroxybenzotriazole
HPLC	High-performance liquid chromatography
HRMS	High resolution mass spectrometry
HRP	Horseradish peroxidase
Hsp40	Heat shock protein 40
HSQC	Heteronuclear single quantum coherence spectroscopy (NMR)
ICAT	Isotope-coded affinity tags
ICR	Institute of Cancer Research, London, UK
IEDDA	Inverse electron demand Diels-Alder
IgG	Immunoglobulin G
IMDM	Iscove's Modified Dulbecco's Medium
IMS	Industrial methylated spirits
J	Coupling constant
K_d	Dissociation constant
kDa	Kilodaltons
KITLG	KIT proto-oncogene ligand
KRAS	Kirsten rat sarcoma viral oncogene homolog
LASP1	LIM and SH3 protein 1
LCMS	Liquid chromatography mass spectrometry
LMICs	Low and middle income countries
LogFC	Log_2(Fold Change)
LRRFIP1	Leucine-rich repeat in Flightless-1 interaction protein 1
m	multiplet
M	molar
m/z	Mass over charge
MAP	MUTYH-associated polyposis
MCM7	Minichromosome maintenance complex component 7
MeCN	Acetonitrile
MeOH	Methanol
MHz	Megahertz
mol	moles
MOPS	3-(N-morpholino)propanesulfonic acid
mRNA	Messenger RNA
MRP	Multidrug resistance-associated protein
MTT assay	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl-2H-tetrazolium bromide assay
MYCBP2	MYC binding protein 2
NaOH	Sodium hydroxide
NAP	Nucleosome assembly protein
nm	Nanometres
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser Effect Spectroscopy (NMR)
NP40	Nonidet-P40 substitute
NQO1	NAD(P)H Quinone Dehydrogenase 1
Nrf2	Nuclear factor erythroid 2-related factor 2
NSAIDs	Non-steroidal anti-inflammatory drugs
OBn	Benzyl
OH	Hydroxyl
PAGE	Polyacrylamide gel electrophoresis
PAINS	Pan Assay Interference Compounds
PAL	Photoaffinity labelling
pBa	*p*-benzoyl-L-phenylalanine
PBS	Phosphate buffered saline

PC1	Principle component 1
PCA	Principle component analysis
PCDH7	Transmembrane receptor protocadherin 7
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PGE-2	Prostaglandin E-2
Pgp	P-glycoprotein
Pgp	P-glycoprotein
PGs	Prostaglandins
ph	Phenyl
pH	Potential of hydrogen
PL	Pancreatic lipase
PMF	5,7,3',4',5'-pentamethoxyflavone
Poly-A	Polyadenylated
PPAR-γ	Peroxisome proliferator-activated receptor gamma
PPIs	Protein-protein interactions
PVDF	polyvinylidene difluoride
RHEB	RAS homolog enriched in brain
RHO	RAS Homolog Family Member
RNA	Ribonucleic acid
RNA-seq	Ribonucleic acid sequencing
ROS	Reactive oxygen species
rpm	Revolutions per minute
rRNA	Ribosomal RNA
rt	Room temperature
RT-qPCR	Reverse-transcription quantitative polymerase chain reaction
s	Singlet
SAR	Structure-activity relationship
SDS	Sodium dodecyl-sulfate
SDS-PAGE	Sodium dodecyl-sulfate polyacrylamide gel electrophoresis
SEMG1	Semenogelin 1
siRNA	Small interfering RNA
SRB	Sulforhodamine B
SREBF1	Sterol regulatory element-binding protein 1
TBAI	tetrabutyl ammonium iodide
TBATB	*n*-tetrabutylammonium tribromide
TBS-T	Tris-buffered saline with 0.1% Tween
TCA	Trichloroacetic acid
TCF	T cell factor
TEAB	Triethylammonium bicarbonate
TFA	Trifluoroacetic acid
TGN	Trans-golgi network
THF	Tetrahydrofuran
THPTA	Tris(benzyltriazolylmethyl)amine
TLC	Thin-layer chromatography
TLR	Toll-like receptor
TMF	Trimethoxyflavone
TMF	Trimethoxyflavone
TMT	Tandem mass tags
TPD	trifluoromethyl phenyl diazirine
Tris	tris(hydroxymethyl)aminomethane
UCL	University College London
UPR	Unfolded protein response
UV	Ultraviolet
V	volts
xg	times gravity
XO	Xanthine Oxidase
Δ	Chemical shift

Standard abbreviations for masses and volumes are used throughout e.g. g, L, mg, µg, as are mathematical operators such as Log. Element symbols have been abbreviated throughout.

1 Introduction

1.1 Flavonoids

1.1.1 Flavonoids and their sources

Flavonoids are polyphenolic phytochemicals that are produced in a wide number of plants. They are often responsible for the colour and aroma of flowers. The flavones apigenin and luteolin have been found in the pink, purplish red and reddish orange flowers in *Paeonia* cultivars, commonly known as peonies.[1] Flavonoids are also consumed widely in the diet in the form of seeds, citrus fruits and tea, all of which contain large numbers of these compounds, with over 8000 unique chemical structures identified in total.[2,3] Flavonoids can occur naturally in aglycone and glycoside forms, with the glycosidic linkage usually found at the C5 or C7 position. Natural flavonoid compounds possess a common phenylchromanone (C_6-C_3-C_6) structure, with two benzene rings that may have substitutions often referred to as the A and B rings, connected by a heterocyclic pyran or pyrone ring, referred to as the C ring (Figure 1.1).[4]

Flavonoids are further divided into seven major sub-categories: flavonols, flavanols, flavones, isoflavonoids, flavanones, dihydroflavonols and anthocyanidins, depending on the substitution pattern of the central ring (Figure 1.1).[5] Typical substitutions on the A and B rings include hydroxyl (-OH) and methoxyl (-OMe) groups.[6] The exact structure of the flavonoid can affect its biological properties and behaviour.[7,8]

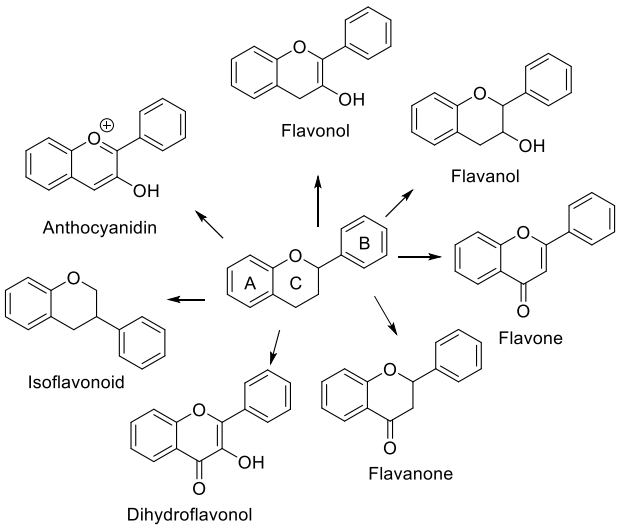


Figure 1.1: Structures of the seven major sub-categories of flavonoids: flavonols, flavanols, flavones, isoflavonoids, flavanones, dihydroflavonols and anthocyanidins.[5] The central phenylchromanone structure is given commonly used A, B and C ring labels.

1.1.2 Flavonoids as biologically active compounds

Historically, flavonoids have been shown to exhibit bioactive properties including antibacterial, antioxidant and anticancer properties, that can reduce the risk of disease.[9]

The bioactive properties identified have ranged from studies of the antidipsotropic properties of flavonoids found in *Radix puerariae* to the effect of genistein exposure on sexual differentiation in rats.[10] Plants containing these bioactive compounds have been utilised in Eastern traditional medicine for thousands of years but, are used sparingly in Western medicine despite large amounts of research conducted. The number of research papers focusing on the use of flavonoids in health has risen drastically over the last 25 years, a PubMed search for the term flavonoids reports 1,522 results for the year 1998 which can be compared to 12,304 results for the year 2022.[11]

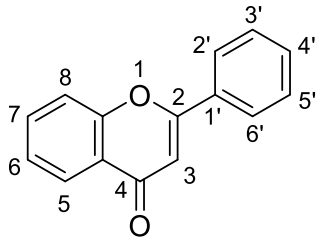
The study of the bioactive properties of flavonoids is however often restricted to the benefits accrued from consuming flavonoids within the diet. Research has primarily focused on the data from epidemiological studies, which can be difficult to compare accurately as the dietary intake of flavonoids varies significantly across the world.[12] Studies often involve monitoring the administration of a high-

flavonoid content food type, such as parsley, and monitoring changes in health using a variety of biomarkers. Plant sources also contain a mixture of polyphenolic compounds, therefore isolating individual compounds, if any, with enhanced potential is complex, there is also the consideration that it is the unique mixture of plant compounds that can give certain food types their beneficial properties.

The complex nature of the research has precluded valuable insight into key mechanisms of action of these compounds, which is vital when aiming to develop a compound with bioactive properties into a drug.

1.1.3 Flavones

The flavone subclass of flavonoids has been shown to contain compounds with a wide range of promising bioactive properties, which could offer great therapeutic potential if properties and mechanisms of actions can be harnessed and optimised. Flavones have the key structure of a pyrone C ring, with a double bond between C2 and C3 in the skeleton, oxidised at the C4 position and no substitution at the C3 position (Figure 1.2).[13] Commonly investigated flavones include apigenin, tricin and luteolin.



Flavone

Figure 1.2: Chemical structure of the flavone subclass of flavonoids including numbered locations on the flavonoid core structure. Substitutions are often found on the two phenyl rings; these can result in significant changes in properties.

1.1.4 Biological activities of flavones

Significant amounts of research have been carried out into the biological properties of flavonoids and specifically flavones as a subclass. This has primarily been focussed on the antioxidant and anticancer

properties of these compounds, including their capacity to re-sensitise cancer cells to chemotherapy and as chemopreventive agents. Due to the variety of properties they have been suggested to have, the uses of flavones in areas such as inflammation and obesity have also been investigated.

A key issue faced, however, when considering the biological activities of flavones is their oral bioavailability. Defined as the fraction of orally administered drug that reaches systemic circulation, poor oral bioavailability is a barrier to development of these drugs *in vivo*. Cellular uptake of most flavonoids studied has been shown to be substantial, as the compounds can easily penetrate cultured cells in their aglycone form. However, when administered orally low levels are seen in circulation. When consumed as part of the diet, the form in which the flavonoid is consumed can alter its bioavailability, *in vivo* behaviour and metabolism.[14] In contrast however, these difficulties are lessened when examining the behaviour of methoxylated flavones. They have been shown to be resistant to various metabolic processes including glucuronidation and sulfonation, which are key to the rapid removal of flavone molecules from circulation in the blood, and therefore are much more bioavailable.[13,15]

However, alongside the promise of the various biological activities of flavones as potential medicines, caution is also needed. Some biologically active natural products, like flavonoids, have been shown to possess activity in a wide range of areas. While this can be potentially useful for developing new drug compounds, sometimes the compounds are instead known to be 'promiscuous'. Pan Assay Interference Compounds (PAINs) are compounds that have been identified as causing behaviours in cells that can result in assay interference and appear to be the result of a desired biological interaction between compound and target, or result in a desired outcome, but do not possess any true activity that can be developed further. PAINs compounds interfere with biochemical assays in various ways, such as by metal chelation, which can affect protein behaviour; by redox mechanisms; causing aggregation or affecting readout by exhibiting fluorescence.[16]

Flavones have not been named as PAINs compounds in the past, however, it is important to note that other natural compounds that display a wide range of activities have often been classified as PAINs,

such as curcumin. Awareness of this potential downfall of natural product drug discovery indicates the need for the identification of interacting binding partners of the compound being studied, leading to an eventual full understanding of the mechanism of action of these compounds for their required roles.[17,16]

1.1.4.1 Antioxidant properties

Oxidative processes in cells result in the generation of reactive oxygen species (ROS) and cellular levels of these are generally controlled *via* normal metabolic processes. The body possesses endogenous defence systems to control levels of ROS, however various factors, including those in the environment such as cigarette smoke, can contribute to an imbalance in the reactive oxygen species (ROS) produced, resulting in a state known as 'oxidative stress'. An excess of ROS can cause damage to cellular structures such as lipids and DNA. Levels of oxidative stress have been shown to contribute to various metabolic diseases such as cardiovascular disease and also to many cancers and neurological disorders.[3,18] Many flavone structures have been shown to demonstrate antioxidant properties, and the consumption of these in the diet may offer a level of protection from ROS related conditions.[19]

Antioxidant activity has been shown to occur via two broad mechanisms: the prevention of initiation, and chain breaking mechanisms that act by scavenging radicals, mostly hydroxyl radicals.[20] Studies of the structure-activity relationships of the antioxidant activity of flavones have been widespread (Figure 1.3).

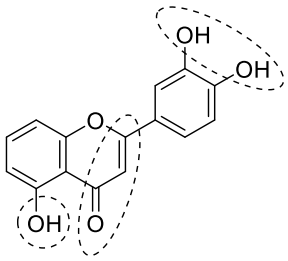


Figure 1.3: Structure of 5,3',4'-trihydroxyflavone. The dashed circles indicate key areas of the flavone structure that have been shown to be involved in antioxidant activity.[3,6,21,22]

A general consensus finds that a higher number of hydroxyl substituents on the structure increases antioxidant activity.[20,21] This is also thought to be the reason flavonols often demonstrate higher levels of antioxidant activity compared to flavones, due to the presence of an extra hydroxyl group at the 3 carbon in the C ring. The presence of a 2,3-double bond alongside the 4-keto group also confers a level of activity due to the improved electron delocalisation abilities, offering flavones superior activity to other flavonoids without this property, such as flavanones and flavanols.[3,6,21] The presence of two hydroxyl groups on the B ring has also shown to be beneficial, the structure can allow the formation of an intramolecular hydrogen bond between the groups, which can stabilise phenoxyl radicals.[22] A C5 hydroxyl substitution has also been shown to be valuable, as well as a catechol moiety on the B ring. These proximal groups are thought to improve the metal chelating abilities of these molecules, improving overall antioxidant capacity.[6,21]

Flavones have also been shown to interfere with the activity of some enzymes involved in the generation of ROS, resulting in indirect antioxidant activity. The activity of the free radical-producing enzyme xanthine oxidase (XO), an enzyme involved in the metabolism of purines, is often inhibited in the presence of flavone molecules.[6] The flavone scaffold has recently been utilised to develop lead compounds for the inhibition of XO, mitigating the effects of excess levels of ROS and other XO related conditions, such as gout.[6,23]

Flavones have also been shown to influence expression of antioxidant genes by modulating redox sensitive transcription factors e.g. Nuclear factors κB (NFκB) and NfE2 related factor-2 (Nrf2).[24] The result of this is increased expression of genes encoding detoxifying and antioxidant proteins.[24] Synergistic relationships between flavones and other endogenous antioxidants such as ascorbate and tocopherol have also been demonstrated, enhancing the capabilities of the flavones.[6]

1.1.4.2 Anticancer activity of flavones

One of the largest areas of research concerning flavonoids, and in particular the flavone subclass, is their anticancer properties. Natural products have long been a key source for anticancer agents, with

almost 50% of small molecule drugs approved from the 1940s to 2014 being natural products or their derivatives.[25] Paclitaxel (Taxol®) is one of the most successful natural product anticancer agents and is used worldwide to treat a number of cancer types including ovarian, breast, lung and pancreatic cancer.[26] Paclitaxel was originally isolated from the bark of the pacific yew tree in 1971 and was approved for medical use in the US in 1992. Paclitaxel is still commonly used as a stand-alone therapy and in combination with other potent anticancer drugs such as cisplatin.[27] Like flavonoids, paclitaxel has been shown to exhibit many other bioactivities and has shown promise in areas such as skin disorders and inflammation.[28]

A mixture of flavone structures have been tested for their anticancer abilities, including: apigenin, chrysin, luteolin, baicalein, tangertin and acacetin (Figure 1.4A).[9] The flavones appear to exert their anticancer abilities *via* different mechanisms depending on the structure and cancer cell type. The overall mechanisms by which these compounds cause anticancer effects is also often unknown.

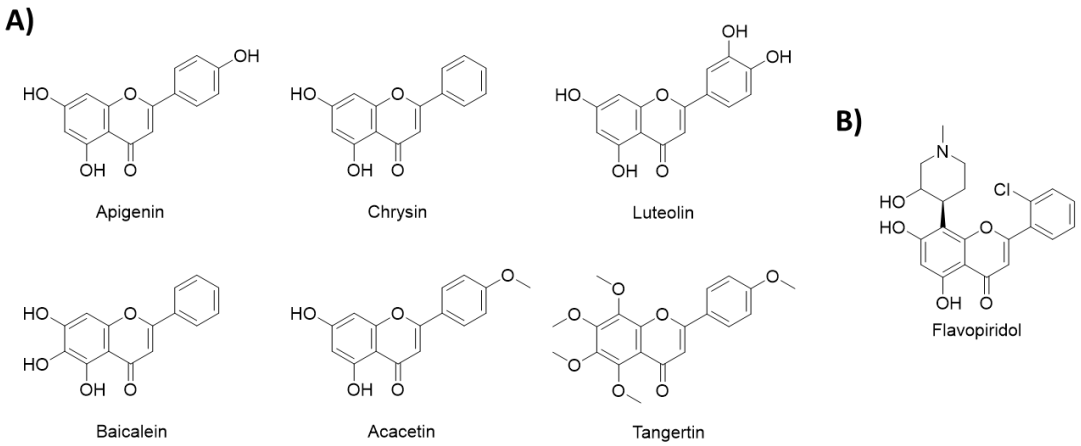


Figure 1.4: A) Structures of apigenin, chrysin, luteolin, baicalein, acacetin and tangertin, all tested for their anticancer abilities. B) Structure of Flavopiridol (alvocidib).

The compound Flavopiridol, also known as Alvocidib, is a synthetic flavone compound with a substituted piperidine ring conjugated at C8 (Figure 1.4B). This compound was investigated in phase II clinical trials, combined with cytarabine and mitoxantrone (FLAM) therapy, for the treatment of newly diagnosed high-risk acute myeloid leukaemia and was recommended for further trials. Alvocidib has

been shown to be a potent inhibitor of serine-threonine cyclin-dependent kinases (CDKs) with most activity towards CDK9 and CDK7 and its main mechanism of action is driven by its effects on transcriptional regulation by inhibition of the kinases.[29,30]

1.1.4.3 Chemotherapeutic sensitisation of drug resistant cancer cells

Many flavones have also been shown to re-sensitize drug-resistant cancer cell lines to established cancer drugs, causing them to regain efficacy. For example, apigenin has been shown to have a chemo-sensitization effect on hepatocellular carcinoma cells that are doxorubicin resistant (BEL-7402/ADM) and inhibited growth of human hepatoceullar carcinoma xenografts in mice. Luteolin has been shown to sensitize cancer cells to the anticancer effects of cisplatin, a clinically important drug used for cancer chemotherapy. These effects were shown to be dependent on an increase in p53 protein level, resulting in enhanced levels of cisplatin-induced apoptosis.[31]

Many flavone compounds have also been shown to modulate multidrug resistance. Multiple flavones have been shown to interfere with the role that ATP-binding cassette (ABC) transporters play in the cross resistance of cancer cells to multiple chemotherapeutics. Regulation of oxidative stress is also known to be important in the development of multidrug resistance, as with the anti-oxidant effects of flavones, the ability to scavenge ROS to control oxidative stress levels, could be a contributing factor in the reversal of resistance. Many other potential mechanisms of action of flavones in the area of resistance reversal have also been identified, including the regulation of hypoxia, cancer stem cell growth suppression and effects on cell autophagy mechanisms. Currently, it is assumed that the multi-functional anti-resistance properties of flavones is not attributed to a single target, but rather a broad spectrum of target modification.[32]

1.1.4.4 Chemopreventive properties

Alongside their potential as anticancer treatments, significant research has been conducted into the chemopreventive properties of consuming flavonoids as part of the diet.[8,10] Epidemiological studies have analysed the link between dietary intake of flavonoids and the risk of developing cancer,

examining positive correlations between higher dietary flavonoid intake and lower risk of developing certain cancers.[7,33] However, the current data is somewhat inconclusive, as the dietary and lifestyle habits of people varies significantly, making it hard for a direct comparison.

Chrysin has been identified as a key candidate in the field of chemoprevention, however the key mechanism of action of this flavonoid is yet to be determined. It has been shown to reduce cell proliferation by inducing the activity of antioxidant and detoxifying enzymes, reducing activity of cytochrome P450 enzymes and inducing apoptosis.[34]

Acacetin, has been shown to exhibit chemopreventive properties alongside a number of other biological properties. It has been shown to inhibit many clinically-relevant enzymes including cytochrome P450 enzymes CYP1A1 and CYP1B1.[35] It was also shown to block phosphorylation of AKT and inhibit PI3-K activity, suggesting associations with this kinase-based signalling pathway.[35]

Walle and Walle studied the chemopreventive properties of multiple flavone compounds with different levels of methylation on oral cancer cells, showing that some fully methylated flavones have great potential as cancer chemotherapeutic or chemopreventive agents. Analysis of the effect flavones, differing in levels of methylation, had on cell proliferation levels suggested that methylated flavones are more potent.[36] The outcomes were seen to be different depending on cell type however, which could suggest that the trend is not seen for all types of cancer cell.

Application of the chemopreventive capabilities of flavones to reduce the risk of cancer relies on the consistency of administration of the compounds and the ability of them to reach the cells in the desired form. The low levels of compound bioavailability for flavones when administered as part of the diet can result in the compounds not having the desired effects. This can be due to low levels of compound uptake, or metabolism. It was also suggested that increased methylation of flavones can improve oral bioavailability and tissue accumulation which can have large impacts on the efficacy of an anticancer drug and can lead to potential for chemopreventive agents as dietary supplements. This is currently thought to be due to lower susceptibility of the compounds to metabolism.[37]

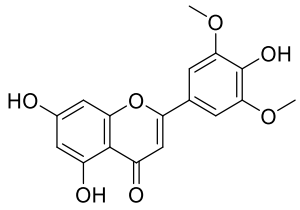
1.1.4.5 Anti-inflammatory properties

Flavones have also been investigated in the area of anti-inflammation, which plays a substantial role in a number of conditions including asthma, multiple sclerosis and rheumatoid arthritis. Steroidal and non-steroidal anti-inflammatories are ineffective as treatments for chronic inflammatory disorders and therefore alternatives are desired. Cyclooxygenase (COX) enzymes are widely known to play a vital role in the synthesis of prostaglandins (PGs), which leads to inflammation. The COX protein has 3 isoforms, COX-1, COX-2 and more recently discovered COX-3.[38,39] COX-2 is an inducible enzyme that responds to stimuli in regions of inflammation in the body to produce prostaglandins.[38,39] Many flavone structures have been shown to play a role in the regulation of COX-2 expression and therefore play an inhibitory role in the production of prostaglandin E_2 (PGE_2), reducing inflammation.[40] Ligand docking studies have also shown that many flavonoid compounds, particularly those with a double bond between C2 and C3 show binding within the COX-2 active site, which can lead to enzyme inhibition.[41] The anti-inflammatory properties of flavones may be connected to their activity as anti-cancer agents as COX-2 has been shown to play a role in carcinogenesis and resistance to treatments and is over-expressed in most tumour types.[42]

1.1.4.6 Anti-obesity properties

Global levels of obesity have risen rapidly and link to many other potential health risks, including cardiovascular diseases, hypertension, cancer and diabetes.[43] Natural products such as flavones pose a promising treatment for conditions such as obesity due to their generally low levels of toxicity.[44,45]

Studies comparing the association between obesity and higher dietary intake of flavonoids as a whole, have suggested that the increased consumption could be linked to lower levels of body fat, although results are not conclusive.[46] High oral dose supplementation of the flavone tricin (Figure 1.5) has been shown to reduce body fat mass and body weight in *in vivo* mouse models of obesity over a 12-week period. The compound was shown to exert these anti-obesogenic properties by causing a decrease in lipogenic marker and transcription factor expression, responsible for fat synthesis in the liver and adipocyte differentiation respectively.[47]



Tricin

Figure 1.5: Structure of tricin, a flavone that has shown potential anti-obesity properties in vivo.[47]

It has been shown that treatment with flavones can prevent preadipocyte differentiation and inhibit the activity of pancreatic lipase (PL). PL is responsible for the hydrolysis of triglycerides and majority of dietary fat processing. Inhibition of this enzyme can result in the prevention of dietary fat absorption.[48] Structure-activity relationship studies of flavone docking in PL shows that the C2-C3 double bond and the keto-group at C4 are responsible for the superior activity of flavones at PL inhibition compared to other flavonoids.[49]

There is also potential for the use of flavones as method of regulating type 2 diabetes. Apigenin and other flavonoids have been reported to exhibit anti-hyperglycemic effects and improve glucose tolerance *in vivo*.[50] Luteolin (Figure 1.4) can also increase insulin sensitivity through a reduction in inflammation-related insulin resistance.[51]

1.2 Cancer

Cancer ranks as the first or second leading cause of death before the age of 70 in 112 countries across the world, with an estimated almost 10 million cancer-related deaths in 2020.[52] The burden of cancer incidence and death is rapidly increasing due to an overall ageing population and changes in predominance of known cancer risk factors.[52]

The development of cancer involves the conversion of normal healthy cells into malignant cells. Six changes in cell physiology are classed as essential in the development of cancer: self-sufficiency in growth signals, insensitivity to growth inhibitory signals, evasion of apoptosis (programmed cell death), limitless replicative potential, sustained blood vessel formation, tissue invasion and metastasis; these are known as the hallmarks of cancer.[53]

1.2.1 Colorectal cancer

Colorectal cancer (CRC) is responsible for approximately 10% of cancer related deaths and is currently the second most commonly diagnosed cancer in women and the third in men worldwide.[52,54] The development of CRC usually starts from an aberrant crypt which can develop in to a polyp, which is a small growth in the lining of the intestines.[55] The crypts are currently thought to originate from stem cell or stem-cell-like cells, which can accumulate genetic mutations over time that are responsible for the initiation and maintenance of tumours. These polyps can progress to CRC over a 10-15 year period.[54] There are two distinct pathways of polyp development, 70-90% of CRCs follow the adenoma-carcinoma pathway and 10-20% originate from the serrated neoplasia pathway (Figure 1.6). These pathways are well studied and the understanding of the adenoma-carcinoma sequence has allowed the development of reliable biomarkers that can monitor progression over time.[56]

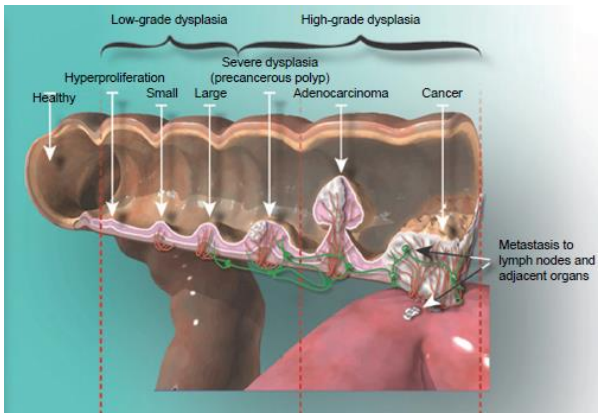


Figure 1.6: Histological changes in the development of CRC following the adenoma-carcinoma sequence.[55]

Up to 30% of CRC patients report a family history of the disease, with 3%-5% of cases resulting from clearly defined inherited syndromes, such as Lynch syndrome, familial adenomatous polyposis (FAP) and MUTYH-associated polyposis (MAP). The genetic causes of many of these inheritable conditions have been determined, and clinical genetic testing for many is available.[57,58,59]

Specific mutations have been shown to be instrumental in the development of both FAP type cancers and in many sporadic cancers, which trigger activation of the Wnt signalling pathway. Many of these

mutations are found in the adenomatous polyposis coli (*APC)* tumour suppressor gene, that codes for the APC protein. Inactivation of this gene can lead to the development of large numbers of adenomatous polyps in the colon and rectum.[60,61] Loss of the APC protein triggers accumulation of β-catenin and can cause activation of T cell factor (TCF)-responsive genes that encode transcription factors.[62] The tumour protein p53 (TP53) and Kirsten rat sarcoma viral oncogene homolog (KRAS) genes are also commonly mutated in cancer and can lead to the development and progression of CRC.[63]

Some cancers can be treated effectively using surgery, this can become more complex as the cancer progresses and so is preferred when the cancer is caught early due to early detection, for example *via* screening programmes.[54] Chemotherapy and radiotherapy are also regularly used treatment options, particularly if the cancer has metastasised. Metastatic CRC is commonly treated using 5-fluorouracil, a drug that inhibits DNA synthesis, oxaliplatin or irinotecan, the latter of which, was originally developed from the natural product camptothecin.[64,65]

1.2.1.1 Colorectal cancer in Malaysia

The understanding of the stages of CRC development has allowed for increased screening, which has resulted in downward trends in mortality in some countries such as the United States and the United Kingdom, however these trends are not moving in this positive direction in other parts of the world.[66,67] CRC is becoming an increasing burden in low and middle income countries (LMICs). Analysis of cancer survival rates in 279 countries from 1995-2009 show that 5-year survival rates of colon and rectum cancer over this time improved for most developed countries, those with the highest human development index (HDI) rankings e.g. Norway, Switzerland and Australia. This trend however is reversed when examining LMICs, as survival rates remain stable or even fall.[68] Many LMICs in Southeast Asia show similar profiles for colorectal incidence and survival rates, a key area that is increasingly experiencing this burden is Malaysia.[69] The changes in distribution of colorectal cancer incidence is a clear example of the cancer transition, in which countries with increasing socio-economic growth are showing rapid rises in the occurrence of cancers that are already prevalent in high-income countries.[67]

The increase in incidence of CRCs in Malaysia may be attributed to a number of factors, an increase in affluence has led to many adopting a more westernised lifestyle, such as a change in diet leading to increased obesity[70] and high levels of smoking.[71] Malaysia is also experiencing an ageing of its population and with the majority of CRCs diagnosed in those over the age of 40, this can result in elevated levels of this disease.[69]

1.2.1.2 Chemoprevention of CRC

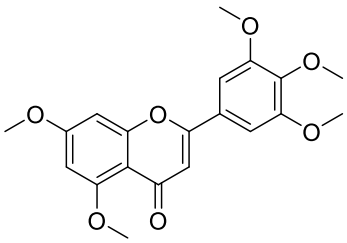
Cancer chemoprevention involves the pharmacological intervention to halt or reverse the process of carcinogenesis. This can mean the prevention of the development of pre-cancerous cells into cancerous cells or to prevent the development of new cancers after treatment.[56] The long progression of the CRC development sequence, and the ability to detect the presence of adenomas before they progress along the sequence, makes colorectal cancers promising targets for chemopreventive strategies.[72]

Some strategies of CRC chemoprevention have been investigated. Research has shown that long-term intake of non-steroidal anti-inflammatory drugs (NSAIDs) can reduce the risk of developing CRC. This has resulted in the identification of cyclooxygenase (COX) enzymes, specifically COX-2, as a molecular target for potential CRC chemopreventives.[73] Metformin is a drug that was developed for the control of hyperglycaemia for patients with type-2 diabetes. It has also shown promising anticancer properties, and epidemiological studies have shown that it can lead to reduced cancer incidence. Commonly used vitamin supplements such as calcium, vitamin D and omega-3 polyunsaturated fatty acids have also been suggested to reduce CRC incidence, however studies of this are often unreliable and the mechanisms by which these agents reduce risks are currently unknown. The ideal properties for an effective chemopreventive agent include: 1) low-toxicity levels to healthy cells, as the patient would likely be required to take the drug for extended periods of time; 2) ability to be consumed orally with good bioavailability; 3) a known and reliable mechanism of action.[74]

Flavones, alongside other phytochemicals, have shown promise as potential chemopreventive as well as chemotherapeutic agents.[36,74] As they are regularly consumed in the diet it is estimated that the level of toxicity caused by these compounds is low, but the oral bioavailability of flavones in their natural product form is often found to be low. As previously discussed, research by Walle *et al.* has shown that an increase in methylation of flavones can result in superior bioavailability compared to hydroxyl-substituted and non-substituted structures.[36]

1.3 3’,4’,5’,5,7-Pentamethoxyflavone (PMF)

A polymethoxyflavone compound that has shown particularly promising biological properties is 3’,4’,5’,5,7-pentamethoxyflavone (PMF) (Figure 1.7). PMF was isolated from the leaves and root bark of the Malaysian plant, *Murraya paniculate,* commonly known as orange jasmine.[75,76] The compound has been shown to be both a promising anticancer and chemopreventive agent. The multiple methoxy groups offer beneficial properties, such as improved tissue accumulation, as they prevent the compound from rapid metabolism *in vivo* and therefore improve the potential of the compound as a drug.[36.37]



3',4',5',5,7-pentamethoxyflavone (PMF)

Figure 1.7: Structure of 3’,4’,5’,5,7-pentamethoxyflavone (PMF).

Recently, studies have shown the antiproliferative activity of PMF in cancer cells. Vongdeth *et al.* synthesized and evaluated the *in vitro* antiproliferative activity of a selection of polymethoxyflavones and polymethoxychalcones, including PMF. The study showed that a higher number of methoxy group substituents on the structure resulted in higher antiproliferative activity overall.[77]

PMF has also been studied on drug resistant cancer cells, it has been shown to sensitize the cisplatin resistant A549 (A549/CDDP) non-small cell lung cancer line, to cisplatin and therefore reduce combined IC_{50} values. PMF was thought to interfere with the expression of nuclear factor erythroid 2-related factor 2 (Nrf2), which is thought to play a large role in cisplatin resistance. This behaviour has been seen for other flavone molecules, such as apigenin and luteolin, and could indicate a potential mechanism of action of flavones in cancer cells.[78,79,80,81]

PMF has shown particular promise when investigating its effects on preventing and treating CRC in the preclinical setting. The number of methoxy groups present around the flavone structure has also been shown to be important in determining the activity of a compound. Cai *et al.* studied the chemopreventive properties of PMF against other natural flavones, tricin and apigenin. The results suggested that PMF was superior to tricin and apigenin in efficacy. This study also tested a potential mechanism of action of the flavones. Specifically, their chemopreventive behaviour has been associated with their anti-inflammatory properties, so the interference with prostaglandin generation was also monitored. The results showed that the expression of COX-2 was not affected by incubation of the APC10.1 cell line with PMF, however it did result in inhibition of PGE-2 production. It was also shown that PMF can dock into the active site of the COX enzyme *in silico,* suggesting that this could be a potential mechanism of action of the flavones.[82] This study is discussed further in Chapter 3.

The exact mechanism of action of PMF that results in its anticancer and chemopreventive properties, however, is still currently unknown. Potential interacting proteins and changes in gene expression caused by the compound have previously been investigated, however, there have been to date, no conclusions made.

1.4 Strategies for identification of targets of biologically active compounds

Early-stage drug discovery relies on the identification of biologically active small molecules. Many of these are natural products that exhibit bioactive properties, often used as traditional medicines.[83] It is

of high importance to utilise these properties in medicine for the treatment of a variety of ailments, many of which currently have poor treatments.

Natural products derived from traditional medicines are sometimes viewed with suspicion, for a number of reasons. Firstly, their touted potential therapeutic uses are often wide ranging and unspecific, meaning that their efficacy as targeted medicines is often questioned. Secondly, there is generally very little information available about biological mechanism of action. This is frequently due to a lack of conclusive evidence of the molecular targets and protein interactions.[84] The identification of molecular targets, such as interacting proteins, for a biologically active small molecule can therefore be a vital step towards the elucidation of the mechanism of action of the compound, and thus significantly improve its chances for development as a drug.[84,85]

Methods that aim to identify targets of small molecules can usually be placed in two categories: those that probe physical interactions between active molecule and target, and those that use changes in cell phenotype to link to cellular targets.[86] Typically these methods are used in tandem e.g. phenotypic analysis is often used as a method of target validation after isolation of potential targets.[87,88] Studying the spectrum of interactions of a bioactive molecule in a cell can also offer advantages by predicting off-target effects which can be useful in turn for predicting toxicity and side effects of potential drug compounds.[89]

As target identification of small molecules is a crucial stage in the development of new medicines, technologies have been developed that can allow for a non-biased approach to detecting proteins that interact with the compound in cells.

1.4.1 Affinity methods and other protein target identification methods

A classical approach for affinity-based target identification is the use of an affinity matrix. A bioactive compound can be immobilised on a support or conjugated to a suitable isolatable group such as biotin, and on interaction with cell lysate, binds with the target molecule(s). After washing to elute unbound proteins, the compound and target moiety can be cleaved from the support allowing for isolation and

identification of binding partners (Figure 1.8).[90,91] This is a simple yet effective method of identifying and isolating key binding partners and has been proven useful in many instances, such as to confirm the covalent binding mechanism of pironetin to α-tubulin. The specific binding site of pironetin was also established *via* isolation of the covalently modified target after biotinylation of the compound.[85,92]

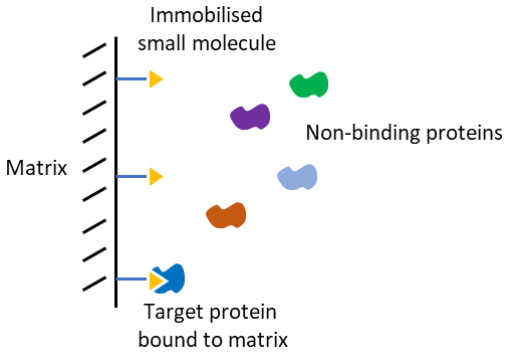


Figure 1.8: Traditional affinity matrix technology, involving the immobilisation of a bioactive compound on a support matrix. A mixture of target and non-target proteins can then be exposed to the support and the binding protein will form a linkage with the solid support. The non-binding proteins can then be eluted, and the target protein identified.

Although this strategy has proved successful for numerous small molecule-target identifications, these methods can encounter several problems. Firstly, the attachment of the compound to an affinity probe, such as an agarose bead, requires an understanding of the structure-activity relationships (SAR) of the compound of interest. Modification of the compound must be completed at a site that does not hinder the biological activity of the compound, this often requires large amounts of synthetic effort and analysis to identify the correct modification point.[85,90,93] The use of a simple affinity matrix also relies on the strength of binding interactions, which must be sufficient to withstand the multiple washing steps required to remove proteins with moderate or low binding affinity that may be non-specifically associated with the matrix scaffold. Some proteins can also suffer loss of function after detergent washing steps, which can result in loss of binding.[93]

Methods that do not focus on probing physical interactions have also been used successfully but present their own difficulties. Cell-based readouts that assess the phenotypic behaviour downstream as a form of target validation can be deceptive as there is no evidence of genuine interactions between the small molecule and target, resulting in the reporting of large numbers of targets and activities that are not all valid, such as for the promiscuous natural product curcumin.[94] Phenotypic studies are more commonly used for the study of natural products since the methods applied are more widely known and challenging syntheses are avoided.

Another method utilised for target identification is thermal proteome profiling. This monitors the changes in thermal stability of a protein due to the binding of a compound. A compound (ligand) can offer a degree of thermal stabilization to its target protein upon binding, and this can be analysed by monitoring denaturing temperatures to isolate binding partners. However, this requires the systematic evaluation of large numbers of candidate protein targets or can encounter bias.[95,96]

1.4.2 Photoaffinity labelling

The disadvantages of the affinity matrices, phenotypic studies, and thermal proteome profiling have led to the development of an alternative technology known as photoaffinity labelling (PAL) which is now commonly used as a tool for identifying the protein target(s) of a specific compound and also for isolating the binding site(s).[97] As opposed to the reversible interactions upon which affinity matrices rely, PAL uses a crosslinking technique to form a permanent covalent link between the small molecule and the target protein, upon activation with light.[98] To achieve this, the small molecules are modified with a photoreactive probe, which remains inert until activation with specific wavelengths of light. Upon photochemical activation, a highly reactive species is formed, which rapidly forms a covalent link with proximate binding partners (Figure 1.9). This covalent attachment can then allow for the isolation and analysis of the protein target(s). It can allow for even transient and weak interactions between the ligand and biomolecule to be detected, this is particularly useful when probing for unknown target molecules.[99]

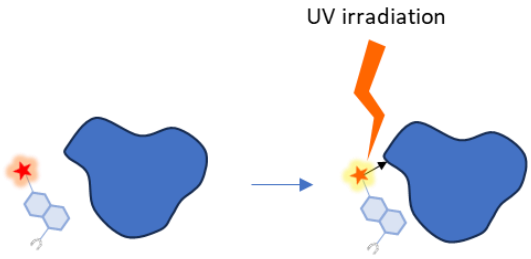


Figure 1.9: Scheme of photoaffinity labelling (PAL) showing the generation of a covalent bond between small molecule and target protein on activation with light.

Generally, a photoaffinity probe consists of three components: 1) *The affinity molecule*. This refers to the small molecule of interest. Attachment of a small molecule into a photoaffinity probe involves the identification of a modification site. This requires analysis of structure activity relationships (SAR) to identify a potential pharmacophore, upon which the photoaffinity probe should be affixed in a way that prevents loss of activity of the bioactive molecule. Acquisition of SAR can require large amounts of synthetic effort to generate tagged analogues for comparison.[96] 2) *A photo-crosslinking agent*. A variety of crosslinking agents are available, each are suited to different tasks.[89] The ideal crosslinking moiety will remain inert in the dark and have a high level of chemical stability, to conditions such as variable pH. Upon activation, the agent will react rapidly with neighbouring molecules to form a stable covalent link that can withstand isolation and analysis procedures.[97] 3) *A reporter tag*. The choice of the reporter tag allows for isolation or visualisation of the covalently modified ligand-target adduct for analysis. The tag should not interfere with the binding mechanism of the small molecule with the target protein and should efficiently isolate those proteins that have been covalently modified.[100]

An example of this is the photoaffinity probe designed by Sun and co-workers (Figure 1.10) to identify unknown targets of oleanolic acid. This consists of three components, the affinity molecule (a modified version of the parent molecule, oleanolic acid) the photocrosslinking group based on a benzophenone moiety, and a biotin-based reporter tag.[101]

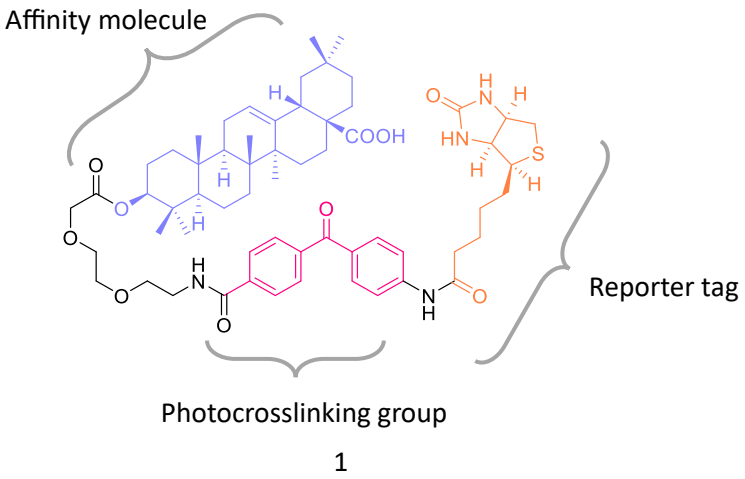


Figure 1.10: Example of a photoaffinity probe used to identify the target proteins of oleanolic acid.[101]

1.4.2.1 Photocrosslinking agents

Many potential photocrosslinking agents have been developed in efforts to provide the ideal properties for the specific detection of ligand-protein interactions. A few key groups popular for use as crosslinking agents are phenyl azides, benzophenones, and diazirines, which can be further categorised into aliphatic diazirines and aryl diazirines (Figure 1.11A).[89,97] Each of these groups forms a different reactive species upon activation (Table 1.1**Error! Reference source not found.**), which results in the formation of a covalent link with neighbouring biomolecules.

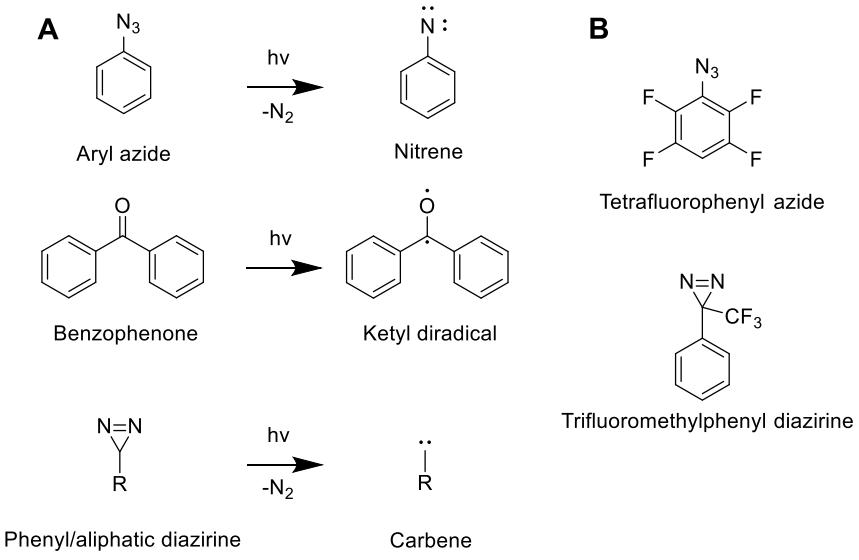


Figure 1.11: A) Structures of key photoactivatable groups used as crosslinking agents and the reactive species formed upon activation. B) Structures of modified PAL agents with improved stability.

Table 1.1: Summary of photoactivatable species.[102]

Photoactivatable group	Reactive species formed	Wavelength of activation
Phenyl azide	Nitrene	254 - 400 nm
Tetrafluorophenyl azide	Nitrene	320 - 350 nm
Benzophenone	Ketyl diradical	350 - 360 nm
Aliphatic diazirine	Carbene	350 nm
Phenyl diazirine	Carbene	340 - 380 nm
Trifluoromethyl phenyl diazirine	Carbene	350 - 355 nm

1.4.2.1.1 Aryl azides

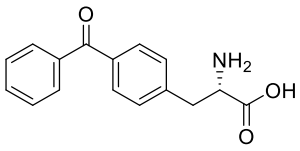
Phenyl or aryl azides rely on the functionality of the azide component. First presented in 1969, aryl azides are now easily synthesized from the corresponding amine *via* their corresponding diazonium salts and are available commercially, making them a very popular choice for photoaffinity labelling.[103] Alkyl azides are a less popular choice due to poor stability, they can be reduced to amines by thiols, which are commonly encountered in biological systems.[97,104,105] Upon irradiation, the azide component yields a molecule of N_2 to form a highly reactive singlet nitrene intermediate with a lifetime of approximately 0.1 ms.[106]

For a photoactivatable group to be useful as a probe for ligand-protein interactions, the activation wavelengths must be in a region that does not cause damage to the biomolecules in the sample, typically this is wavelengths of greater than 300 nm. The maximum absorption for aryl azides falls below 300 nm, making it potentially damaging to proteins.[107] A minor peak however, exists at 460 nm which is compatible with damage-free activation.[89] Modifications of the substituents on the phenyl ring can improve the properties of the azide. For example, tetrafluorophenyl azides exhibit red shifting of the excitation wavelengths to 320-350 nm and prevent unwanted rearrangements.[108] Due to the disadvantages associated with aryl azide probes, other types of crosslinking agent are now more popular for use in target identification.[109]

1.4.2.1.2 Benzophenones

Benzophenones are popular photocrosslinking agents that were first developed for this use in the 1970s. On irradiation of the species, the ground state ketone is excited to form a ketyl diradical (Figure

1.11A).[110] Benzophenones offer a number of advantages which account for their popularity. Many building blocks of the probe are commercially available and the final structure tolerates many solvents and synthetic steps, making incorporation of these structures into probes more straightforward than their competitors.[111] Benzophenones also exhibit reduced sensitivity to ambient light, easing synthesis.[112] The generation of a *p*-benzoyl-L-phenylalanine (*p*Ba) group (Figure 1.12) has allowed the incorporation of the photoreactive group into peptides in a site-specific manner, which can offer opportunities to study protein-protein interactions (PPIs).[113]



p-benzoyl-L-phenylalanine

Figure 1.12: Structure of *p*-benzoyl-L-phenylalanine (*p*Ba), used as an amino acid substitute to incorporate a photoreactive moiety into a peptide structure in a site specific manner.[113]

A further key advantage of this group is the irradiation wavelength, irradiation with light of 350-360 nm is required for the generation of a triplet diradical species, which can abstract hydrogen from neighbouring unreactive C-H bonds resulting in a covalent crosslink. The ability to react with any C-H bond means the reactive species formed can crosslink to all amino acids, though a preference for methionine has been demonstrated.[110] The excited state has been shown to last 120 µs, longer than the reactive species lifetimes for nitrenes and carbenes. As the ketyl radical is unreactive towards water, without a proximate X-H bond it will relax back to ground state and can then be excited again. This property increases the efficiency of crosslinking, as more interactions can be detected, but can also result in large amounts of nonspecific labelling.[107,112]

Benzophenones also show some shortcomings. The size of the group can be problematic as it can cause steric interference with the binding between small molecule and target. This can obscure the true targets of the small molecule, as some binding may be weak/transient and therefore undetected if a bulky group is blocking interactions, though this can be mitigated by the use of a linker group to create

distance.[96,114] The reversible nature of the reactive species formation can also lead to the use of longer irradiation times than necessary for other crosslinking agents, which can cause nonspecific labelling.[97]

1.4.2.1.3 Diazirines

Diazirine photocrosslinking agents irreversibly generate highly reactive carbenes upon irradiation. These species are highly electrophilic and can almost immediately react with proximal biomolecules to form stable covalent crosslinks.[115] Upon irradiation with favourable wavelengths in the region of 350 – 380 nm, which is widely clear of the potentially damaging threshold for biomolecules, singlet carbene species are formed by the extrusion of a molecule of N_2.[110] These carbenes are able to insert into many bond types, including: C-H, X-H, C-X single bonds, where X represents a heteroatom, and C-C double and triple bonds to form the corresponding three membered ring.[116] The singlet carbene can also be converted to a triplet state *via* intersystem crossing. These react through a radical mechanism but form similar crosslinks to the singlet form (Figure 1.13).[99,107] The carbon-based crosslinks formed by carbenes are more stable to a variety of conditions than the nitrogen-based adducts formed *via* a nitrene crosslink.[98] Short nanosecond half-lives of the carbene species results in the short reaction times needed for a photolabeling experiment, which is advantageous.[117]

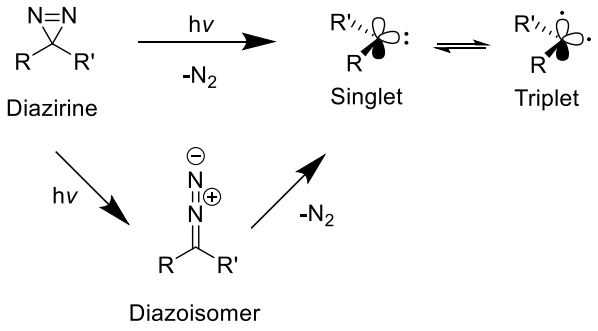


Figure 1.13: Formation of singlet and triplet carbenes from a diazirine *via* the formation of a diazoisomer.[118]

Due to the propensity of carbene to react with a wide variety of bond types, photocrosslinking yields can be seen to be much lower than with other agents, such as benzophenones, as the active species is often rapidly quenched with water or solvent. This, however, is favourable in many situations as it

can result in high levels of selectivity for the probe, as photoadducts are only formed with species in the immediate space. High levels of selectivity are particularly beneficial when there is a relatively low affinity between the probe and the target, reducing large amounts of nonspecific labelling. The low efficiency of the probe can be compensated with the use of higher concentrations.[97,117]

A key benefit of the use of diazirines is their inherently small size, this can prevent interference between the small molecule and binding site, allowing incorporation of the reactive species in close proximity to the small molecule, and increase the probability of forming an adduct in the binding region of the target.[111] Other photocrosslinking agents, such as benzophenones, can cause interference with the molecule binding to the target and are typically required to be installed at a greater distance from the targeting molecule due to their relative larger size.

One drawback of diazirine crosslinking agents is that during activation, a proportion (> 30%) of the diazirine is converted to a diazoisomer, which is not readily reactive. This isomer slowly converts to the reactive carbene species, which can allow for diffusion from target sites resulting in nonselective binding of the probe. The effect has been mitigated by the development of modified diazirine agents.[107,115]

The main disadvantages of diazirine usage as crosslinking agents is their difficult synthesis, which can require many complicated steps, however widespread usage offering a more reliable synthetic method, alongside increased commercial sources for these agents is reducing the impact of this problem.[98]

1.4.2.1.4 Aryl diazirines

Aryl diazirines were first developed for use as photolabeling agents in 1973 by Knowles and co-workers, and are now the most popular form of diazirine used for crosslinking.[119] Phenyl diazirines are chemically stable in a wide variety of conditions and maintain a useful activation wavelength of 340-380 nm.[89,107] Synthesis of probes *via* an alkoxyl-phenyl diazirine intermediate can allow for the simplification of synthesis as many analogues can be generated.[115]

Though many advantages are offered by the development of phenyl diazirine agents, it has been shown that the diazoisomer is formed as a major side product during activation, drastically reducing crosslinking yields.[89] The diazo form has the ability to abstract a proton from a nearby source, generating a diazonium, which leads to a carbocation formation that readily reacts with nucleophiles. This can cause high levels of ligand independent reactivity.[99]

The development of the trifluoromethylphenyl diazirine (TPD) photolabeling group has circumvented some of the problems encountered when using a phenyl diazirine (Figure 1.11). The introduction of the highly electron withdrawing trifluoromethyl group results in the stabilisation of the diazoisomer and reduces levels of nonspecific labelling, with the highest selectivity of all photocrosslinking agents.[107,115,116] The structure retains favourable properties with an activation wavelength range of 350-355 nm and greatly improved chemical and ambient light stability. It has also been noted that irradiation with a second wavelength (302 nm) can offer a 'boost' to yields by converting the unreactive diazoisomer to a carbene.[99,120]

Mizui and co-workers developed a phenyl diazirine based photoaffinity probe (Figure 1.14) to identify the binding targets of the antitumour natural product pladienolide, finding that splicing factor SF3b is an important protein target of the natural product, suggesting its use as a therapeutically important antitumour drug target.[121]

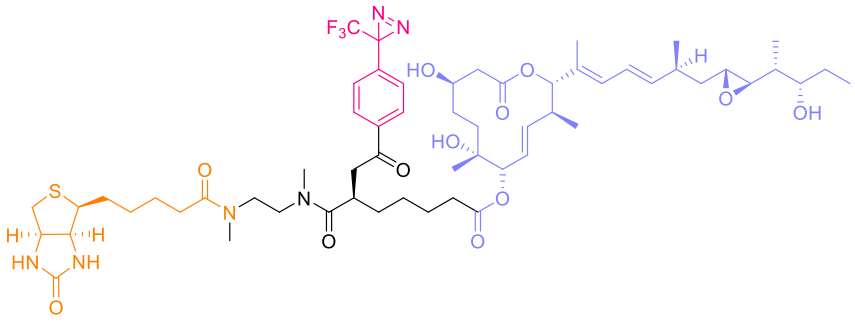


Figure 1.14: An example photoaffinity probe used to identify targets of potential antitumour drug candidate pladienolide.[121]

1.4.2.1.5 Aliphatic diazirines

Although aryl diazirines are most popular due to the properties that the aromatic ring offers, aliphatic diazirine crosslinking agents are also widely used as they offer the unique benefit of size. The very small size of aliphatic diazirines enable them to be incorporated very close or within the small molecules of interest without compromising binding efficiency, this can enable the highly accurate detection of binding sites within target biomolecules.[110] The use of an alkyl diazirine can be necessary when the small molecule of interest does not tolerate modification to the structure well, or changes in activity are very sensitive.[89] An example of this is seen by the identification of kinase cellular targets of Dasatinib, a chemotherapeutic agent. This work was carried out by Shi *et al.* by the incorporation of an aliphatic diazirine attachment to the compounds structure, allowing the development of a cell permeable probe (Figure 1.15).[122]

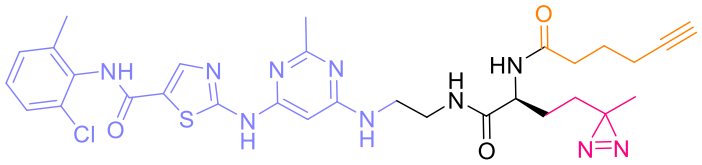


Figure 1.15: Structure of an aliphatic diazirine probe developed from Dasatinib, used to identify unknown kinase targets in vitro.[122]

1.4.2.2 Reporter tags

The third function within a photoaffinity probe involves the installation of a reporter or purification tag. Three common reporter tags are used in most situations; radiolabels, fluorophores and biotin.[123] Historically radiolabels have been a popular choice. Commonly used isotopes include I^{125} or H^{3}, as they enable highly sensitive detection. Their small size allows them to be incorporated within, or extremely close, to the small molecule of interest, resulting in minimal perturbations to the structure. Radioisotope labels are also very chemically stable and can allow the visualisation of the small molecule-target interaction *in vivo*. However, they present a few disadvantages. The synthesis of radioisotope-labelled probed can be complicated as the radioactive elements can be hazardous and

therefore require specific handling measures, and there are a limited set of starting materials.[97,108] Their short half-lives make the probes prone to degradation and therefore complicates analysis. The radioisotope identification method also does not allow for any means of enrichment of the bound compounds and isolation of the proteins using chromatography can be complex, restricting the analysis to visualisation of labelled proteins using protein gels.[97,109,114]

Fluorescent and biotin reporter tags are more widely used and offer advantages over the use of radiolabels, such as ease of handling and visualisation. Biotin reporter tags also give the ability of isolation and purification of tagged proteins which can significantly increase analysis opportunities, such as with proteomics analysis. Fluorescent and biotin reporter tags will be further discussed in Chapter 5.

1.4.2.2.1 Linker length and cleavable probes

The properties of a photoaffinity probe are also dependent on the nature of the linker group between reactive moieties. It is generally regarded that a 'minimalist' approach to the attachment of different reactive parts is favoured. The use of a conjugated handle that contains both the photoreactive crosslinking group and the reporter tag, affixed to the small molecule at a single modification point can reduce compromises in biological activity due to multiple modification sites.[89,123]

The most common linker groups used for photoaffinity labelling are polyethylene glycol (PEG) chains, these offer the advantageous properties of hydrophilicity, which can improve solubility, and chemical stability.[90,96,115] Aliphatic alkyl and peptide chains are also popular choices as linker groups but can suffer with problems of hydrophobicity and degradation respectively and may cause nonspecific labelling.[96]

Appropriate linker length and flexibility are key to developing a suitable photoaffinity probe as they can lead to a variety of outcomes.[97] An ideal linker group is balanced to enable effective crosslinking of the target biomolecule without causing steric hinderance. Short linker groups have been seen to reduce the bioactivity of a small molecule due to crowding, interrupting the binding mechanisms with

target molecules due to an impeded approach. They have also shown the ability to form crosslinks with the probe itself, reducing efficiency of the reaction.[97] High proximity between the photoreactive species and the target molecule however, can improve labelling efficiency by increasing the likelihood of a crosslink forming with the target protein.[124] Longer linker lengths can combat the issue of steric hinderance and preserve bioactivity however, can have compromised labelling efficiencies and can lead to increased levels of nonspecific labelling.[90,125] The length of the linker is often determined empirically for each target-probe pair as the ideal length will vary.[96,126,127]

Different types of linker group can also introduce a level of flexibility or rigidity into the probe. The probability of adopting a geometry that enables the covalent modification of the target molecule is improved by the use of a more flexible linker group, improving crosslinking efficiency, however, a more rigid structure can provide more information into the binding site of the small molecule.[128]

The shape of the probe has also been shown to be important in the development of an ideal photoaffinity probe, linear probes have been demonstrated to produce higher levels of nonspecific labelling than branched probes. This is thought to be due to higher levels of conformational flexibility imposed by linear linker groups, allowing for more interactions with proteins to be generated.[129]

The use of a rigid polyproline rod-like linker (Figure 1.16) has been demonstrated to create a distance between a biotin reporter tag and small molecule, ensuring the small molecule can successfully interact with a target biomolecule without interference from the biotin-avidin complex used for isolation of the targets.[130] Polyproline linkers have also shown success in isolating low abundance target proteins.[96]

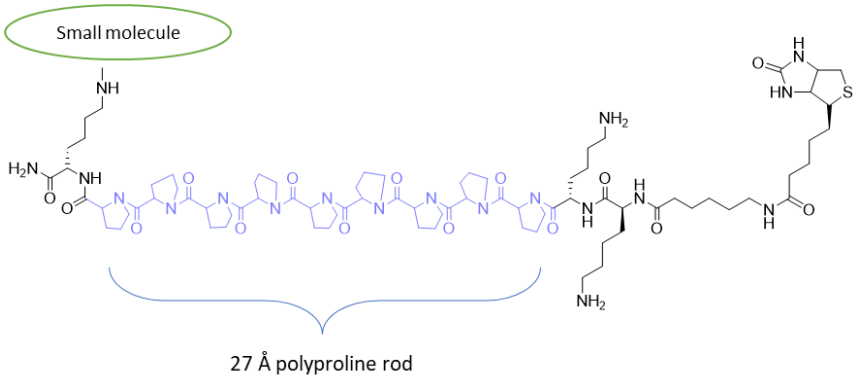


Figure 1.16: The structure of an affinity probe possessing a rigid polyproline rod used to create distance between a biotin reporter tag and affinity molecule, allowing for targets to be identified.[130]

The linker group can also be utilised to add another functionality to the probe. Problems frequently arise when biotin is chosen as the reporter tag, the extremely high affinity between biotin and avidin/streptavidin, which gives the tag its favourable properties for purification (using avidin/streptavidin coated beads), can cause difficulties when trying to isolate and analyse any target biomolecules, as elution from the complex can require harsh conditions.[96] The use of monomeric avidin is sometimes employed as it has a lower affinity, however large amounts and high concentrations of this protein are sometimes needed. Chemical cleavage is an alternative, however this can cause damage to the biomolecules.[131]

These difficulties have led to the development of cleavable linker groups that allow controllable release of the target molecules from a purification method, such as an agarose bead or biotin-avidin complex.[99] The most common form of cleavable linker is the incorporation of a disulphide bond, which, is stable in many biological conditions and, on treatment with mild reducing agents, is easily broken to release the probe from its isolation handle (Figure 1.17).[132] However, disulphide bonds can undergo a process known as disulphide exchange and have been shown to be prematurely cleaved in some cellular systems and reducing buffers.[133]

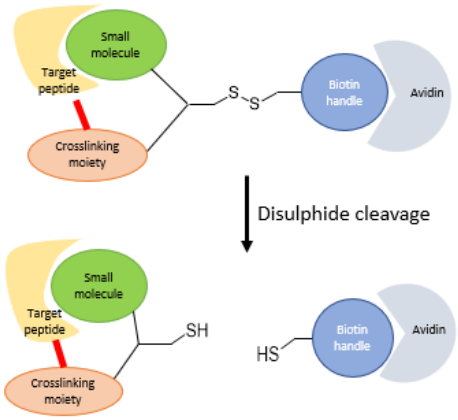


Figure 1.17: A cleavable disulphide linker can offer a method of cleaving any isolated target molecules from a purification method such as a biotin-avidin complex.

Alternative cleavable linker groups have also been developed. A linker group that can be cleaved *via* an enzyme has been developed by Hatanaka and co-workers that utilises a glutamic acid moiety that is proteolyzed *via* a V8 protease to isolate the target molecules from a biotin-avidin complex. The use of this linker group however is only appropriate when working with biomolecules that are stable to V8 protease digestion.[131] Acyl sulphonamide linkers have also been employed to sever a biomolecule from a purification complex. Acyl sulphonamides are chemically stable in acid and base conditions and, upon *N*-alkylation of the moiety, can be cleaved with nucleophiles under mild conditions.[99,134]

Bogyo *et al.* developed a diazobenzene based linker group that can be cleaved in mild reducing conditions using sodium dithionite, the cleavage is compatible with biochemical systems.[133]

A ligand transfer strategy is also occasionally employed, this can allow for the simplification of analysis of the target proteins after isolation, an example of this involves the use of a tosylate ester linkage. Upon attack by nucleophilic residues within the biomolecule, the covalent link is transferred to the target and the small molecule is released (Figure 1.18).[135]

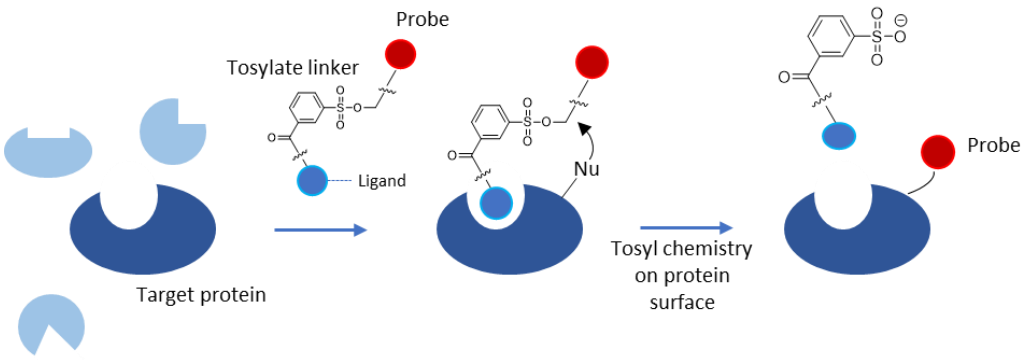


Figure 1.18: Schematic of ligand directed tosyl chemistry used to transfer an isolated target protein to a probe to simplify analysis.[135]

1.4.2.3 Proteomics analysis and validation of tagged proteins

After cross linking of the bioactive compound to protein targets in the cell (or cell lysate), these targets must be identified and validated." This typically involves the enrichment of the target and the identification *via* proteomics. Proteomics is the study of functions, interactions, structure and composition of proteins and mixed protein samples.[136]

Mass spectrometry analysis is a preferred method of complex protein sample analysis to identify a compound's target proteins. The application of mass spectrometry-based proteomics is possible due to gene and genome databases and the development of protein ionisation methods.[137] The technique uses quantitative proteomics to analyse the abundance of proteins in different samples and compare these, known as relative quantification. The use of isotope-coded affinity tags (ICAT) for chemical tag labelling of proteins is a commonly employed strategy for the comparison of samples from different samples.[110] ICAT often uses heavy and light isotopes of the same element e.g. $^{13}C_6/^{12}C_6$ to label proteins, the samples can then be combined for analysis. The isotope labels create a known mass shift which is detectable with mass spectrometry, allowing the relative abundance of proteins in different samples to be compared.[107]

Tandem mass spectrometry, known as MS/MS, combined with liquid chromatography (LC), is a commonly used method that enables the identification of proteins in a sample. The technique involves

more than one stage of mass spectrometry analysis in sequence. Initially, the proteins are ionised and separated by their mass to charge ratio (m/z), the ions are then subjected to fragmentation and the masses of the fragments calculated (Figure 1.19).[137,138] The use of this technique in proteomics allows small amino acid sequences to be identified and these sequences can be used to identify the protein in question. LC-MS/MS can also be used to identify binding sites of the small molecules with the protein of interest.[124]

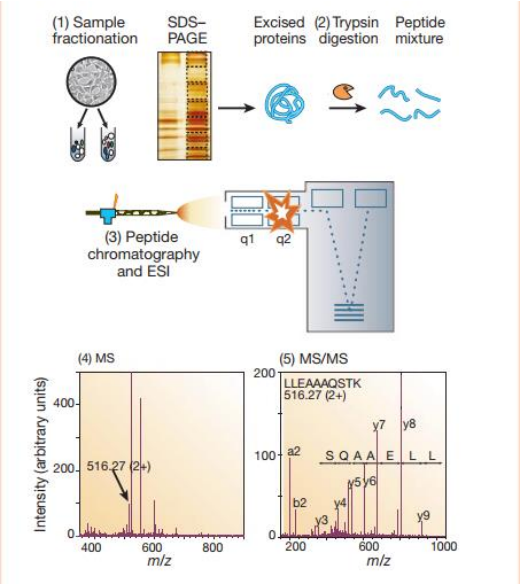


Figure 1.19: Generic mass spectrometry proteomics experiment by Aebersold *et al.* The steps involve excision or isolation of proteins, digestion and chromatography followed by tandem mass spectrometry analysis of the protein sequence. ESI = Electrospray Ionisation. [137]

After identification of the potential target proteins, these must be validated, using one or more methods. Immunobloting, the use of a specific antibody for a candidate target, can validate an assignment if the antibody indicates the protein at the correct size of the crosslinked protein in a gel-based analysis.

Competition assays are another commonly used validation strategy. Excess unmodified small molecule is exposed to the lysate prior to administering the probe, and if the target-probe interaction is specific

to the small molecule, then binding sites are saturated and the probe should not be able to bind, validating the target interaction.[96]

Phenotypic studies can also aid target validation, such as siRNA knockdown, to determine if knockdown/loss of the target in the cell can result in the same phenotype as treatment with the small molecule.[96]

1.4.3 RNA sequencing and analysis

Treatment of cancer cells with anticancer compounds can result in cellular changes including altering gene expression and protein levels in a cell. Changes in these factors can give insight into the mechanism of action of compounds with unknown behaviours. Methods of genetic target identification that monitor changes in DNA or RNA can offer the advantage that the compound does not need to be synthetically modified for use in the experiment. This advantage can mean that the targets of bioactive compounds that are particularly sensitive to modifications can still be investigated.[139] Traditional methods of gene expression analysis involve reverse-transcription quantitative polymerase chain reaction (RT-qPCR) techniques to measure the levels of mRNA in a sample, however these methods are low-throughput, candidate-driven and can be costly and time consuming.[139,140] Higher throughput methods such as gene expression microarray and next-generation sequencing can allow large scale unbiased analysis of changes in gene expression simultaneously.[140] RNA sequencing offers some advantages over microarray techniques including superior resolution that allows specific changes in gene expression to be identified.[141]

1.4.4 Summary

A variety of technologies have been developed to determine the interactions of biologically active compound in the cell. Affinity methods offer an advantage of giving insight into the exact biomolecules that a compound is interacting with, leading to the possible elucidation of a key target molecule of the compound. This can support the development of a biologically active compound with improved properties, as binding properties can be optimised when a target is known. The affinity methods

discussed, however, cannot effectively determine the effect the compound is causing on interaction with the biomolecules, be that inhibition, activation or otherwise.

Analysis of changes in gene and protein expression in a cell, such as the analysis of differential gene expression with RNA sequencing, can give further insight into the consequences of treatment of cells with active compounds. These methods can give information about the specific impacts exposure to a compound has, leading to a deeper understanding of the mechanisms in which the compounds cause the desired outcomes, e.g. cell death.

Overall, a combination of technologies must be employed to effectively understand the mechanisms of action of a biologically active natural product, in order to allow development into an effective drug treatment.

1.5 Project aims

The aim of this work is to identify potential cellular targets of the bioactive natural product 3',4',5',5,7-pentamethoxyflavone (PMF) so that in the future its anticancer properties can be developed into medicines. Protein cellular targets will be identified from pull down experiments using a photoaffinity probe conjugated to the active compound. To develop this probe, a structure-activity relationship study will be carried out that firstly requires the synthesis of modified analogues of PMF and then *in vitro* cell assays to evaluate the growth inhibition properties of the analogues and compare them to the parent compound. The activity of the analogues will be used to identify a modification point on the structure that can be used to attach the photoaffinity probe while retaining the anticancer properties of the structure. From this modification point, the photoaffinity probe will be synthesised and used to isolate protein binding interactions which will be analysed using mass spectroscopy proteomics to identify key binding proteins of the compound. RNA sequencing of CRC cells will also be performed after treatment with the active compound to assess changes in gene expression to further examine the behaviour and mechanisms of the compound *in vitro*.

2 Design and synthesis of 3',4',5',5,7-pentamethoxyflavone and modified flavone analogues for the investigation of the structure-activity relationship.

2.1 Introduction

2.1.1 Structure activity relationship studies

The first step in the identification of the target proteins of PMF using photoaffinity labelling methods, is to identify a suitable site for modification of the parent structure. The modification position can be used to affix a photoaffinity probe to the biologically active compound to isolate interacting proteins in the cell. To identify a suitable modification point on the compound, a structure-activity relationship (SAR) must be established, this can identify a location on the parent compound that is suitable for modification and is able to be altered without the loss of vital biological activity. An SAR can be used to identify the part of a medicinally active compound that is responsible for its primary biological activity, such as binding to its target protein, this part of a molecule is known as its pharmacophore.[142] SAR studies are often used to facilitate the improvement of compound characteristics when developing lead compounds into drug candidates.[143] To collect this data on the viable modification points of the parent compound, PMF, a series of analogues were designed to be synthesised and their biological activities tested and compared to determine a point of attachment for the photoaffinity probe.

2.1.2 Synthesis of the parent compound 3',4',5',5,7-pentamethoxyflavone (PMF).

Synthesis of the parent natural compound, PMF, is the primary objective when designing a structure activity relationship study as it provides a reference point for analysis.[144] PMF has been synthesised many times previously.[77,145] Typically PMF, and as such other flavones, have been synthesised *via* two routes, either by a Baker-Venkataraman rearrangement from a substituted ketone[144,145,146] or by the formation of a chalcone intermediate from a base catalysed aldol condensation and then a cyclisation

reaction catalysed by iodine, to form the flavone.[77,147,148,149,150,151,152] Both routes are detailed in Figure 2.1. Bose *et al.* demonstrated the synthesis of a flavone following an alternative route, using a 3-step process from a 2'-acetoxychalcone intermediate by bromination using *n*-tetrabutylammonium tribromide (TBATB) then dehydrobromination before a cyclisation in sodium methoxide solution to form the desired flavone product [153]

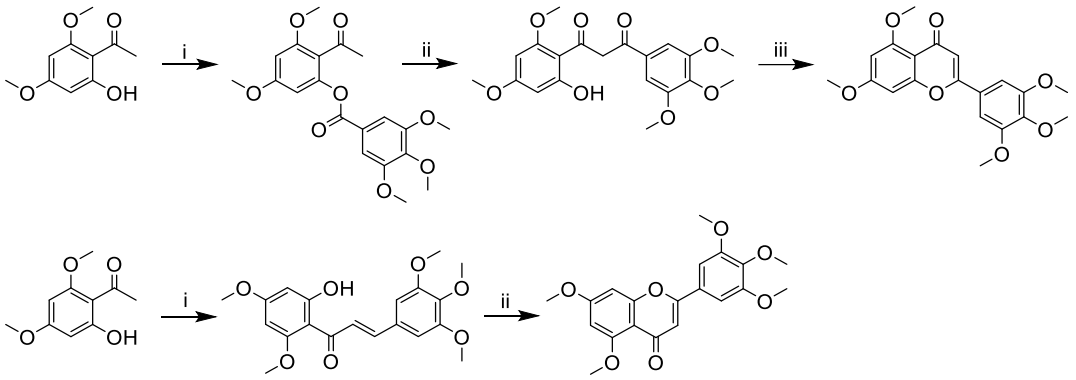


Figure 2.1: A: Synthesis of PMF using a Baker-Venkataraman rearrangement method used by Kawaii *et al*.[146] i) Dicyclohexylcarbodiimide, N,N-dimethyl-4-aminopyridine, 3,4,5-trimethoxybenzoic acid; ii) KOH, pyridine, 100˚C, 10 min; iii) 20% H_2SO_4/acetic acid, 100˚C, 10 min. B: Synthesis of PMF by formation of a chalcone intermediate by an aldol condensation used by Vongdeth *et. al.*[77] i) 3,4,5-trimethoxybenzaldehyde, 50% NaOH (aq), ethanol, rt; ii) I_2, DMSO, H_2SO_4, reflux.

All routes described give efficient synthesis of the desired flavone structure, alongside other polymethoxy and hydroxyflavones synthesised. For this work, the route selected was a modified version of route B (Figure 2.1) *via* a chalcone intermediate structure as the route only involved 2 steps and due to ready availability of most of the required substituted ketone and aldehyde starting materials.

2.1.3 Selection of analogues

The four distinct chemical environments of the methoxy substituents on PMF, offer a very useful range of modification points for the development of polymethoxyflavone analogues (Figure 2.2). These

analogues can be compared in their biological activities to identify a position that can be changed synthetically without compromise.

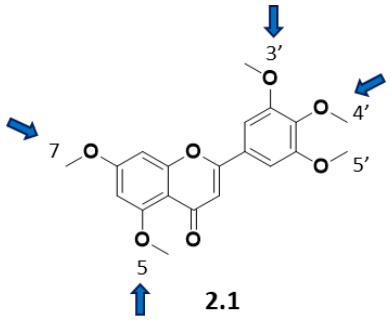


Figure 2.2: Structure of PMF, **2.1** with arrows pointing towards the four modification points identified for analogue synthesis.

Flavone analogues were designed using these four modification points, by replacement of one of the methoxy group substituents in the 5- ,7- , 3'- or 4'-position. The 5'-methoxy group substituent was not replaced to produce an analogue as rotation about the bond connecting the C and B rings of the flavone structure creates a symmetrical ring and places the 3'- and 5'-methoxy substituents in identical chemical environments. This symmetry is expected to be lost at low temperatures. The single environment can be seen in the ^{1}H NMR spectrum (Figure 2.7) as the CH_3 peak corresponding to the 3'- and 5'-methoxy substituents is a singlet with an integral value of six hydrogens.

The four modified analogues were synthesised with a benzyl group in place of the original methoxy substituent. The benzyl group was chosen for a few reasons: the benzyl group is known to be simple to install and can be removed in an orthogonal manner, resulting in its frequent usage as a protecting group.[154] The properties of the benzyl group also serve its purpose as an SAR analogue modification, the size of the phenyl group is such that it can mimic the modification of the structure with a photoaffinity label and could potentially disrupt any binding of the compound to a specific protein binding pocket in the same way (Figure 2.3). The aromatic properties of the benzyl group also create a different environment to the methoxy substituent so could disrupt any bonds formed that are

dependent on the chemical properties of the methoxy group between the molecule and the target protein if installed in the pharmacophore region of the molecule.

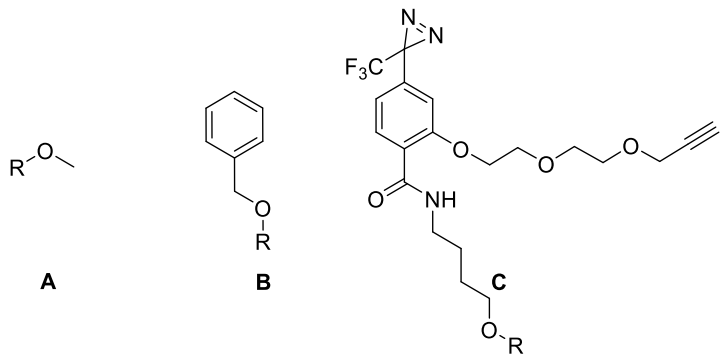


Figure 2.3: Structures of: A - methoxy group substituent; B - benzyl group substituent; C – phenyl diazirine photoaffinity probe. The structures are shown for comparison of size and shape of the benzyl group modification used to mimic the structure of the photoaffinity probe used for the pull down study. R = the remainder of the PMF structure.

2.1.4 Collaborative work investigating anti-obesity properties of polymethoxyflavones.

During the process of synthesising the PMF analogues necessary for this study, two additional compounds were isolated and investigated. As previously discussed, flavones, particularly polymethoxyflavones have been shown to have a wide selection of biological activities, many of these could be potentially beneficial in a medicinal context, such as for anti-obesity.[155] Two hydroxyflavones were synthesised and their anti-obesity properties tested in collaboration with researchers in the Wong research group at Taylor's University, Malaysia. This work supports recent results published by the group showing anti-pancreatic lipase and anti-adipogenic properties of PMF.[156]

2.1.5 Chapter aims

This Chapter aims to design, synthesise and characterise PMF and four modified flavone analogues selected for biological testing to develop a structure-activity relationship. This will facilitate the selection of a suitable modification point on the parent structure that can be chemically altered

without compromising the desired biological activity of the compound, to allow the attachment of a chemical probe structure for the purpose of target identification.

2.2 Materials and methods

2.2.1 Materials

1-(2-Hydroxy-4,6-dimethoxyphenyl)ethenone and 3,4-dimethoxy-5-hydroxybenzaldehyde were purchased from Biosynth International, Inc. 3,4-Dimethoxy-5-hydroxybenzaldehyde, dry chlorobenzene, dry DCM, iodine, 3,4,5-trimethoxybenzaldehyde and sodium hydroxide were purchased from Acros Organics. 4-Hydroxy-3,5-dimethoxybenzaldehyde and pentamethylbenzene were purchased from Fluorochem. Aluminium trichloride and benzyl bromide were purchased from Alfa Aesar. Boron trichloride was purchased from Sigma Aldrich. Deuterated solvents were purchased from Cambridge Isotope Laboratories Inc. Chloroform, DCM, diethyl ether, DMF, DMSO, ethanol, ethyl acetate, formic acid, hydrochloric acid, hexane, methanol and magnesium sulphate were purchased from Fisher Scientific. Sodium thiosulphate pentahydrate was purchased from Glentham Life Sciences. Potassium carbonate was purchased from Thermo Fisher Scientific.

2.2.2 Instrumentation

2.2.2.1 NMR

A Bruker AVII 400 MHz spectrometer was used to record NMR spectra and each spectrum was calibrated to the known chemical shift of the residual solvent peak of the deuterated solvent used. Chemical shifts were reported in part per million (ppm) and J coupling values were reported in Hz. Spectral data was processed using ACD labs software. Proton NMR spectra were obtained at 400 MHz and carbon-13 spectra were obtained at 100 MHz.

2.2.2.2 Biotage

Automated column purification was carried out using a Biotage Isolera One system using various columns and solvent systems, detailed further in the experimental methods section. Columns for purification were purchased from Biotage.

2.2.2.3 *Liquid chromatography mass spectrometry (LCMS)*

Electrospray mass spectrometry data was obtained using a Thermo MSQPlus instrument fitted with a Zorbax SB-C18 5 µm 3.0 x 150 mm column using H_2O + 0.1% formic acid and MeOH + 0.1% formic acid or H_2O + 0.1% TFA and MeCN + 0.1% TFA mobile phases. Data was analysed using Chromeleon™ Chromatography Data System (CDS) Software.

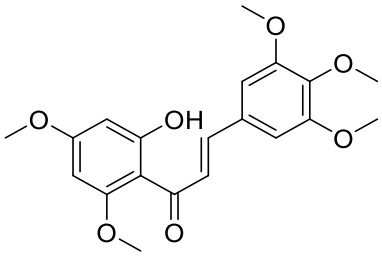
2.2.2.4 *High resolution mass spectrometry (HRMS)*

A Bruker micrOTOF-Q LCMS system was used to obtain high-resolution mass spectrometry data, samples were dissolved in HPLC-grade methanol and injected using direct injection mode with a mobile phase system of 50:50 MeOH and H_2O. Data was processed using Bruker Compass Hystar software.

2.2.3 Experimental methods

2.2.3.1 *Synthesis of PMF*

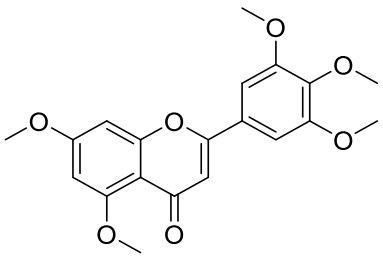
2.2.3.1.1 Preparation of (E)-1-(2-hydroxy-4,6-dimethoxyphenyl)-3-(3,4,5-trimethoxyphenyl)prop-2-en-1-one, **2.2**.



1-(2-Hydroxy-4,6-dimethoxyphenyl)ethanone (1.00 g, 5.10 mmol, 1 eq) was dissolved in methanol (10 mL) with 3,4,5-trimethoxybenzaldehyde (1.00 g, 5.10 mmol, 1 eq) and sodium hydroxide (0.820 g, 20.4 mmol, 4 eq). The solution was refluxed for 20 h at 60 °C. Addition of a few drops of 6M HCl caused formation of a bright yellow precipitate. This was collected with filtration and washed with water. The crude product was recrystallised from ethanol to give the desired compound (1.29 g, 3.45 mmol, 67.7%). ^{1}H NMR (400 MHz, DMSO-d_6) δ_H 13.12 (1H, s, OH), 7.67-7.53 (2H, m, H - α, H - β), 7.04 (2H, s, H – 2', H – 6'), 6.14 (2H, dd, J = 13.2, 2.3, H – 3, H - 5), 3.87 (3H, s, OCH_3 - 6), 3.84 (6H, s, OCH_3 - 3',OCH_3 - 5'), 3.82 (3H, s, OCH_3 - 4), 3.70 (3H, s, OCH_3 - 4').[77] (in line with published ^{1}H data). ^{13}C NMR (100 MHz, DMSO-d_6) δ_c 193.7 (CO), 165.4 (C - 4), 164.9 (C - 6), 161.8 (C-OH), 153.4 (C-3',C-5'), 143.0 (C-β), 139.7 (C-4'), 130.6 (C-α), 127.3 (C-1'),

106.9 (C-1), 106.1 (C-2'), 94.1 (C-3), 91.3 (C-5), 60.4 (4' -OCH_3), 56.4 (OCH_3 - 4) , 56.2 (3' – OCH_3, 5' – OCH_3), 55.9 (OCH_3 - 6). ESI-MS: m/z $[M+H]^+$ Calcd. 375 found 375, $[M+Na]^+$ Calcd. 397, found 397.

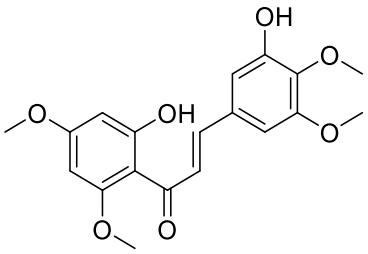
2.2.3.1.2 Synthesis of 5,7-dimethoxy-2-(3,4,5-trimethoxyphenyl)-4H-chromen-4-one, **2.1**.



Iodine (0.070 g, 0.270 mmol, 0.2 eq) was added to (E)-1-(2-hydroxy-4,6-dimethoxyphenyl)-3-(3,4,5-trimethoxyphenyl)prop-2-en-1-one (0.500 g, 1.34 mmols, 1 eq) in DMSO (10 mL). The reaction mixture was refluxed at 130 °C until complete consumption of starting material. Reaction progress was monitored *via* TLC (95:5 DCM:MeOH). The reaction mixture was cooled and poured into 10% [%w/w] sodium thiosulphate pentahydrate solution (100 ml) resulting in an off-white precipitate. This was filtered and washed with water. The crude product was purified using reverse phase Biotage Isolera purification (eluent: H_2O + 0.1% formic acid: MeOH + 0.1% formic acid 5-100%, cartridge: Biotage SNAP Ultra C18 12g) to give the desired product. (0.140 g, 0.369 mmol, 27.6%). ^{1}H NMR (400 MHz, DMSO-d_6) δ_H 7.32 (2H, s, H - 2', H – 6'), 6.93 (1H, d, J = 2.2, H - 8), 6.88 (1H, s, H-α), 6.52 (1H, d, J = 2.2, H - 6), 3.92 (3H, s, OCH_3 - 5), 3.91 (6H, s, OCH_3 - 3', OCH_3 - 5'), 3.84 (3H, s, OCH_3 - 7), 3.74 (3H, s, OCH_3 - 4').[146] (in line with published ^{1}H data). ^{13}C NMR (100 MHz, DMSO-d_6) δ_c 175.8 (CO), 163.7 ($COCH_3$ – 7), 160.2 (C – β), 159.4 ($COCH_3$ – 5), 159.2 (C – 8a), 153.3 ($COCH_3$ – 3', $COCH_3$ – 5'), 140.2 ($COCH_3$ – 4'), 126.2 (C – 1'), 108.3 (C – 4a), 108.2 (C – α), 103.6 (C – 2', C – 6'), 96.3 (C – 6), 93.5 (C – 8), 60.2 (OCH_3 – 4'), 56.3 (OCH_3 – 3', OCH_3 – 5'), 56.1 (OCH_3 – 5, OCH_3 – 7). ESI-MS: m/z $[M+H]^+$ Calcd. 373, found 373, $[M+Na]^+$ Calcd. 395, found 395.

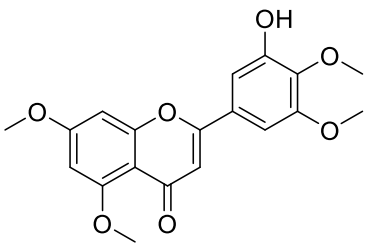
2.2.3.2 *Synthesis of the **3'-OBn** analogue*

2.2.3.2.1 Synthesis of (E)-3-(3-hydroxy-4,5-dimethoxyphenyl)-1-(2-hydroxy-4,6-dimethoxyphenyl)prop-2-en-1-one, **2.4**.



1-(2-Hydroxy-4,6-dimethoxyphenyl)ethanone (0.500 g, 2.55 mmol, 1 eq) was dissolved in methanol (15 mL) with 3,4-dimethoxy-5-hydroxybenzaldehyde (0.460 g, 2.55 mmol, 1 eq) and sodium hydroxide (0.16 g, 10.2 mmol, 4 eq). The solution was refluxed for 24 h at 80 °C then cooled to 5 °C and addition of a few drops of 6M HCl caused a bright yellow precipitate to form. This was collected with filtration and washed with ice-cold methanol. The crude compound was purified using Biotage Isolera purification (eluent: Hexane:Ethyl Acetate, 5-100%, cartridge: SNAP KP-Sil 50g) to give a yellow product (0.440 g, 1.20 mmol, 47.0%). ^{1}H NMR (400 MHz, DMSO-d_6) δ_H 13.43 (1H, s, OH - 2), 9.43 (1H, s, OH – 3'), 7.63 – 7.49 (2H, m, H - 3, H - 5), 6.84 (2H, dd, J= 8.8, 1.9, H - 2', H - 6'), 6.14 (2H, dd, J = 13.0, 2.3, H - α, H - β), 3.90 (3H, s, OCH_3 - 5) 3.82 (6H, s, OCH_3 - 4, OCH_3 - 5') 3.71 (3H, s, OCH_3 - 4'). ^{13}C NMR (100 MHz, DMSO-d_6) δ_c 192.2 (CO), 165.4 ($COCH_3$ – 4), 165.4 ($COCH_3$ – 6), 161.8 (CHO - 2), 153.4 ($COCH_3$ – 5'), 150.8 (CHO – 3'), 142.8 (C - β), 138.6 ($COCH_3$ – 4'), 130.2 (C - α), 126.4 (C – 2'), 109.2 (C – 2'), 106.3 (COC), 104.4 (C – 6'), 93.9 (C – 3), 91.1 (C - 5), 59.6 (OCH_3 – 4'), 59.9 (OCH_3 – 4'), 55.8, 55.7 (OCH_3). ESI-MS: m/z $[M+H]^+$ Calcd. 361, found 361, $[M+Na]^+$ Calcd. 383, found 383.

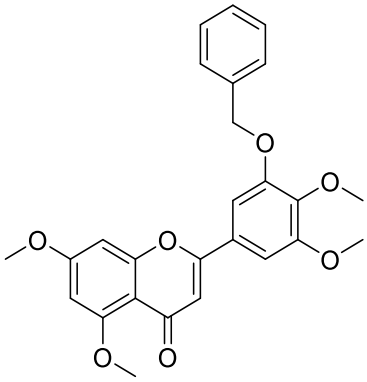
2.2.3.2.2 Synthesis of 2-(3-hydroxy-4,5-dimethoxyphenyl)-5,7-dimethoxy-4H-chromen-4-one, **2.5**.



Iodine (70.4 mg, 0.28 mmols, 0.2 eq) was added to (E)-3-(3-hydroxy-4,5-dimethoxyphenyl)-1-(2-hydroxy-4,6-dimethoxyphenyl)prop-2-en-1-one (0.500 g, 1.39 mmols, 1 eq) in DMSO (10 mL) which was refluxed at 130 °C for 2 h. The reaction mixture was then cooled and poured into 10% [%w/w] sodium thiosulphate pentahydrate solution (50 mL) to give a cloudy orange solution with brown solid precipitate which was filtered to collect precipitate. The filtrate was dried under reduced pressure and combined with the solid precipitate to give the crude product as brown

solid of 90% purity (0.370 g). This was purified in batches using reverse phase Biotage Isolera purification (eluent: H_2O + 0.1% formic acid: MeOH + 0.1% formic acid 5-100%, cartridge: Biotage SNAP Ultra C18 12g) (0.180 g, 0.500 mmol, 48.0%). ^{1}H NMR (400 MHz, DMSO-d_6) δ_H 9.54 (s, 1H, OH), 7.13 (dd, J = 11.6 Hz, 1.9 Hz, 2H, H-2',H-6'), 6.81 (d, J = 2.2 Hz, 1H, H-6), 6.69 (s, 1H, H-α), 6.51 (d, J = 2.3 Hz, 1H, H-8) 3.90 (s, 3H, OCH_3-5) 3.88 (s, 3H, OCH_3-5') 3.83 (s, 3H, OCH_3-7) 3.74 (s, 3H, OCH_3-4'). ^{13}C NMR (100 MHz, DMSO-d_6) δ_c 176.1 (CO), 164.1 ($COCH_3$ – 7), 160.7 (C - 2), 160.0 ($COCH_3$ – 5), 159.6 (C – 8a), 153.9 ($COCH_3$ – 5'), 151.3 (CHO), 139.5 ($COCH_3$ – 4'), 126.5 (C – 1'), 108.7 (C – 4a), 108.2 (C – 2'), 107.7 (C – α), 102.0 (C – 6'), 96.7 (C – 6), 93.7 (C – 8), 60.5 (OCH_3 – 4'), 56.5 (OCH_3 – 5'), 56.5 (OCH_3 – 7), 56.4 (OCH_3 – 5). ESI-HRMS: m/z $[M+H]^+$ Calcd. 359.1125, found 359.1167. Found m/z value is 0.0042 greater than calculated.

2.2.3.2.3 Synthesis of 2-(3-(benzyloxy)-4,5-dimethoxyphenyl)-5,7-dimethoxy-4H-chromen-4-one, 2.6.

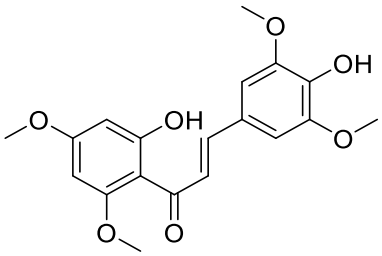


Benzyl bromide (79.7 µL, 0.670 mmol, 1.2 eq) was added to a solution of 2-(3-hydroxy-4,5-dimethoxyphenyl)-5,7-dimethoxy-4H-chromen-4-one (0.200 g, 0.560 mmol, 1 eq) and potassium carbonate (0.150 g, 1.12 mmol, 2 eq) in DMF (10 mL). The solution was stirred at rt for 72 h, forming a pale yellow solution. This was diluted with water (20 mL) causing a milky white solution. The solution was filtered to give white solid. Remaining product was extracted from the filtrate with DCM and dried under reduced pressure to form a white solid. These were combined to give the crude product which was purified using Biotage Isolera purification (eluent: DCM: MeOH 0-10%, cartridge: 50g KP-SIL snap.) to give the desired product. (0.160 g, 0.350 mmol, 62.3 %). ^{1}H NMR (400 MHz, DMSO-d_6) δ_H 7.53-7.31 (7H, m, Bn - H, H – 2', H – 5'), 6.88 (1H, d, J = 2.3, H – 8), 6.86 (1H, s, H – α), 6.51 (1H, d, J = 2.3, H – 6), 5.26 (2H, s, CH_2), 3.92 (3H, s, OCH_3 – 5), 3.90 (3H, s, OCH_3 – 5'), 3.83 (3H, s, OCH_3 - 7), 3.76 (3H, s, OCH_3 – 4'). ^{13}C NMR (100 MHz, DMSO-d_6) δ_c 176.2 (CO), 164.2 ($COCH_3$ – 7), 160.7 (C -2), 159.8 ($COCH_3$ – 5), 159.6 (C – 8a), 153.9 ($COCH_3$ – 5'), 152.7 (COBn), 141.1 ($COCH_3$ – 4'), 137.5 (Bn –

C1), 129.0 (Bn – C3, Bn – C5), 128.4 (Bn – C4), 128.2 (Bn – C2, Bn – C5), 126.6 (C – 1'), 108.8 (C – 4a), 108.6 (C – α), 105.7, 104.3 (C – 2'/C – 6'), 96.7 (C – 6), 93.9 (C – 8), 70.9 (CH_2), 60.7 (OCH_3 – 4'), 56.8 (OCH_3 – 5'), 56.5 (OCH_3 – 7), 56.5 (OCH_3 – 5). ESI-HRMS: m/z $[M+H]^+$ Calcd. 449.1595, found 449.1619.

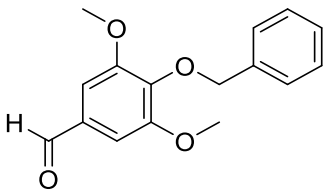
2.2.3.3 *Synthesis of **4'-OBn** analogue*

2.2.3.3.1 Synthesis of (E)-3-(4-hydroxy-3,5-dimethoxyphenyl)-1-(2-hydroxy-4,6-dimethoxyphenyl)prop-2-en-1-one.



1-(2-Hydroxy-4,6-dimethoxyphenyl)ethanone (0.500 g, 2.55 mmol, 1 eq) was dissolved in methanol (10 mL) with 4-hydroxy-3,5-dimethoxybenzaldehyde (0.460 g, 2.55 mmol, 1 eq) and sodium hydroxide (0.440 g, 10.2 mmol, 4 eq). The solution was refluxed for 20 h at 80 °C. Addition of a few drops of 6M HCl caused a bright yellow precipitate to form. This was collected with filtration and washed with water. The crude product was recrystallised from ethanol to give the desired compound (0.130 g, 0.350 mmol, 13.6%). ^{1}H NMR (400 MHz, DMSO-d_6) δ_H 13.34 (1H, s, OH - 2), 9.07 (1H, s, OH – 4'), 7.6-7.56 (2H, m, H - 3, H - 5), 7.02 (2H, s, H - 2'), 6.15 (2H, dd, J = 14.7, 2.3, H-α, H-β), 3.89 (3H, s, OCH_3 - 4), 3.84 (6H, s, OCH_3 - 3',OCH_3 - 5'), 3.83 (3H, s, OCH_3 - 6).[77] (in line with published ^{1}H data). ^{13}C NMR (100 MHz, DMSO-d_6) δ_c 192.8 (CO), 165.5 (C-4/C-6), 165.3 (C-4/C-6), 162.0 (OH - 2), 148.6 (C - 3', C - 5'), 144.4 (C - β), 139.1 (OH – 4'), 125.6 (C - α), 125.1 (C - 1'), 107.1 (C - 1), 106.8 (C - 2', C - 6'), 94.4 (C - 3), 91.5 (C - 5), 56.6, 56.5, 56.1 (OCH_3). ESI-MS: m/z Calcd. $[M+H]^+$ 361, found 361, Calcd. $[M+Na]^+$ 383, found 383.

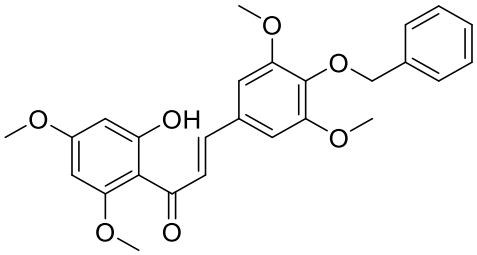
2.2.3.3.2 Synthesis of 4-(benzyloxy)-3,5-dimethoxybenzaldehyde, **2.7**.



Benzyl bromide (0.780 mL, 6.59 mmol, 1.2 eq) was added to a solution of 4-hydroxy-3,5-dimethoxybenzaldehyde (1.00 g, 5.49 mmol, 1 eq) and potassium carbonate (1.52 g, 11.0 mmol, 2 eq) in DMF (10 mL). The solution was stirred at rt for 24 h. The white emulsion formed was diluted with water (20 mL) causing product to precipitate which was isolated with filtration. Remaining product was extracted with diethyl ether and dried under reduced pressure to form an oil. These were combined to give pure

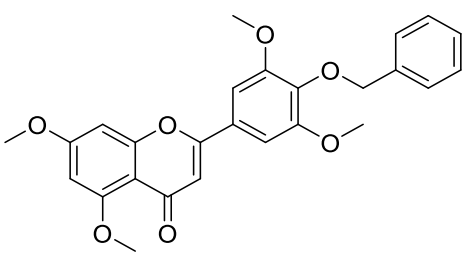
product as a white solid (1.22 g, 4.48 mmol, 81.5%). ^{1}H NMR (400 MHz, DMSO-d_6) δ_H 9.94 (1H, s, HCO), 7.51-7.34 (5H, m, ph-H), 7.31 (2H, s, H-2,H-6), 5.09 (2H, s, CH_2), 3.91 (6H, s, OCH_3). ^{13}C NMR (100 MHz, DMSO-d_6) δ_c 192.3 (CHO), 154.0 ($COCH_3$), 142.1 (COBn), 137.8 (CCH_2), 132.3 (CCHO), 128.6, 128.5, 128.4 (C-Bn), 107.2 (CHCCHO), 74.5 (CH_2), 56.6 (CH_3). ESI-HRMS: m/z $[M+Na]^+$ Calcd. 295.0946, found 295.0953.

2.2.3.3.3 Synthesis of (E)-3-(4-(benzyloxy)-3,5-dimethoxyphenyl)-1-(2-hydroxy-4,6-dimethoxyphenyl)prop-2-en-1-one, 2.8.



1-(2-Hydroxy-4,6-dimethoxyphenyl)ethanone (0.310 g, 1.59 mmol, 1 eq) was dissolved in methanol (20 mL) with 4-(benzyloxy)-3,5-dimethoxybenzaldehyde (0.430 g, 1.59 mmol, 1 eq) and sodium hydroxide (0.250 g, 6.34 mmol, 4 eq). The solution was refluxed for 24 h at 80 °C. Addition of a few drops of 6M HCl caused a bright yellow precipitate to form. This was collected with filtration and washed with methanol to give product as a yellow solid (0.560 g, 1.24 mmol, 78.0%). ^{1}H NMR (400 MHz, DMSO-d_6) δ_H 13.15 (1H, s, OH), 7.68-7.29 (7H, m, Bn-H, H-3, H-5), 7.05 (2H, s, H-2', H-4'), 6.15 (2H, dd, J = 9.8, 2, H-α, H-β), 4.96 (2H, CH_2), 3.88 (3H, s, OCH_3-4), 3.84 (6H, s, OCH_3-3', OCH_3-5'), 3.81 (3H, s, OCH_3-6). ^{13}C NMR (100 MHz, DMSO-d_6) δ_c 192.2 (CO), 165.0 ($COCH_3$ – 4), 164.5 ($COCH_3$ – 6), 161.4, (CHO), 153.1 ($COCH_3$ – 3', $COCH_3$ – 5'), 142.5 (COCHCH), 138.2 (C-Bn), 137.4 (CCH_2OBn), 130.3, 127.9, 127.8, 127.6 (Ph), 126.9 (C – 1'), 106.5 (C – 1), 105.7 (CH – 2', CH – 6'), 93.7 (CH – 3), 90.8 (CH – 5), 73.8 (CH_2), 55.9, 55.9, 55.4 (OCH_3). ESI-HRMS: m/z $[M+H]^+$ Calcd. 451.1751, found 451.1783.

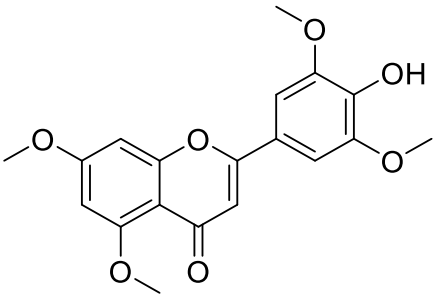
2.2.3.3.4 Synthesis of 2-(4-(benzyloxy)-3,5-dimethoxyphenyl)-5,7-dimethoxy-4H-chromen-4-one, 2.9.



Iodine (0.0600 g, 0.250 mmols, 0.2 eq) was added to (E)-3-(4-(benzyloxy)-3,5-dimethoxyphenyl)-1-(2-hydroxy-4,6-dimethoxyphenyl)prop-2-en-1-one (0.550 g, 1.23 mmols, 1 eq) in DMSO (10 mL) then refluxed at 130 °C for 1 h. The

reaction mixture was then cooled and poured into 10% [%w/w] sodium thiosulphate pentahydrate solution (150 mL) to give a white precipitate. This precipitate was filtered and washed with water to give a white solid of 90% purity (0.560 g). The crude product was purified in batches using reverse phase Biotage Isolera purification (eluent: H_2O + 0.1% formic acid: MeOH + 0.1% formic acid 5-100%, cartridge: Biotage SNAP Ultra C18 12g). (0.230 g, 0.507 mmol, 41.5%). ^{1}H NMR (400 MHz, DMSO-d_6) δ_H 7.47-7.33 (5H, m, Bn-H), 7.30 (2H, s, H-2', H-6'), 6.90 (1H, d, J = 2.3, H-8), 6.87 (1H, s, H-α), 6.50 (1H, d, J = 2.3, H-6), 5.00 (2H, s, CH_2), 3.91 (3H, s, OCH_3-5), 3.90 (6H, s, H-3', H-5'), 3.83 (3H, s, OCH_3-7). ^{13}C NMR (100 MHz, DMSO-d_6) δ_c 175.8 (CO), 163.7 ($COCH_3$ – 7), 160.2 (C – β), 159.4 ($COCH_3$ – 5), 159.2 (C – 8a), 153.4 ($COCH_3$ – 3', $COCH_3$ – 5'), 138.9 (Bn – 1), 137.6 (COBn), 128.2 (Bn – 3, Bn – 5), 128.1 (Bn – 2, Bn – 6), 127.9 (Bn – 4), 126.3 (C – 1'), 108.3 (C – 4a), 108.2 (C – α), 103.5 (C – 2', C – 6'), 96.3 (C – 6), 93.5 (C – 8), 74.0 (CH_2), 56.4 (OCH_3 – 3', OCH_3 – 5'), 56.1 (OCH_3 – 5, OCH_3 – 7). ESI-HRMS: m/z $[M+H]^+$ Calcd. 449.1595, found 449.1662.

2.2.3.3.5 Synthesis of 2-(4-hydroxy-3,5-dimethoxyphenyl)-5,7-dimethoxy-4H-chromen-4-one, **2.16**.

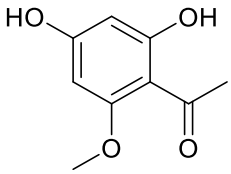


BCl_3 (0.540 mL, 1.00 M in CH_2Cl_2, 2 eq) was added to a stirred solution of 2-(4-(benzyloxy)-3,5-dimethoxyphenyl)-5,7-dimethoxy-4H-chromen-4-one (0.120 g, 0.270 mmol, 1 eq) pentamethylbenzene (0.120 g, 0.800 mmol, 3 eq) in dry DCM (4 mL) via syringe at -84 °C dropwise over 10 mins. After 15 mins, due to complete consumption of starting material, the reaction was quenched with $CHCl_3$:MeOH (10:1, 5 mL) and warmed to rt. A resulting orange solid was removed *via* filtration. The filtrate was concentrated under reduced pressure and purified using reverse phase Biotage Isolera purification (eluent: H_2O + 0.1% formic acid: MeOH + 0.1% formic acid 5-100%, cartridge: Biotage SNAP Ultra C18 12g) to give the desired product (0.070 g, 0.200 mmol, 25.0%). ^{1}H NMR (400 MHz, DMSO-d_6) δ_H 9.20 (1H, s, OH), 7.28 (2H, s, H-2', H-6'), 6.90 (1H, d, J = 2.3, H-8), 6.76 (1H, s, H-α), 6.49 (1H, d, J = 2.3, H-6), 3.91 (3H, s, OCH_3-5), 3.88 (6H, s, H-3', H-5'), 3.83 (3H, s, OCH_3-7). ^{13}C NMR (100 MHz, DMSO-d_6) δ_c 175.9 (CO), 163.7 (C – 7), 160.3 (C – 1), 160.1 (C – 5), 159.2 (C – 9), 148.3 (C – 3', C – 5'), 139.2 (CHO),

120.7 (C – 1'), 108.3 (C – 4), 106.9 (C – 2), 103.9 (C – 2', C – 6'), 96.3 (C – 6), 93.5 (C – 8), 79.3 79.0, 78.7 ($CHCl_3$), 56.4 (OCH_3 – 3', OCH_3 – 5'), 56.1 (OCH_3 – 5, OCH_3 – 7). ESI-HRMS: m/z $[M+H]^+$ Calcd. 359.1125, found 359.1143. Found m/z value is 0.0032 greater than calculated.

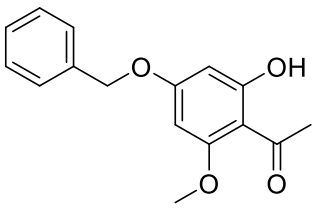
2.2.3.4 Synthesis of 7-OBn analogue

2.2.3.4.1 Synthesis of 1-(2,4-dihydroxy-6-methoxyphenyl)ethan-1-one, **2.10**.



$AlCl_3$ (0.500 g, 3.75 mmol, 1.47 eq) was added to 1-(2-hydroxy-4,6-dimethoxyphenyl)ethan-1-one (0.500 g, 2.55 mmol, 1 eq) and placed in inert atmosphere, dry chlorobenzene (10 mL) was added and the reaction mixture refluxed at 130 °C for 6 h. After cooling, the resulting mixture was concentrated under reduced pressure and ice cold H_2O:HCl (1:1, 20 mL) was added and the mixture sonicated to give a white precipitate which was isolated by filtration. The crude product was purified using Biotage Isolera purification (eluent: Hex:EtOAc 10-40%, cartridge: Biotage SNAP Ultra 10g) to give the desired product as a white solid. (0.110 g, 0.600 mmol, 23.4%). ^{1}H NMR (400 MHz, DMSO-d_6) δ_H 13.80 (1H, s, OH - 2), 10.69 (1H, s, OH - 4), 5.96 (1 H, d, J = 2.2, H – 5), 5.86 (1H, d, J = 2.2, H – 3), 3.81 (3H, s, OCH_3), 2.51 (3H, s, $COCH_3$).[157] (in line with published ^{1}H data). ^{13}C NMR (100 MHz, DMSO-d_6) δ_c 202.4 (CO), 166.3 (C - 2), 165.2 (C – 4), 163.4 (C – 6), 104.6 (C – 1), 95.6 (C – 3), 91.4 (C – 5), 55.9 (OCH_3), 32.6 (CH_3). ESI-MS: m/z $[M+K]^+$ Calcd. 221, found 221.

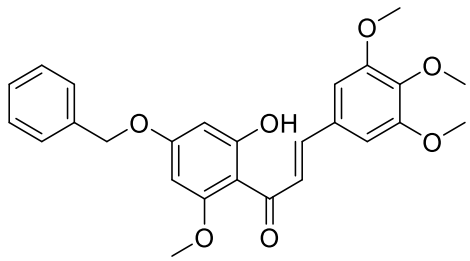
2.2.3.4.2 Synthesis of 1-(4-(benzyloxy)-2-hydroxy-6-methoxyphenyl)ethan-1-one, **2.11**.



Benzyl bromide (71.0 µL, 0.600 mmol, 1 eq) was added to a solution of 1-(2,4-dihydroxy-6-methoxyphenyl)ethan-1-one (0.110 g, 0.600 mmol, 1 eq) and potassium carbonate (0.080 g, 0.600 mmol, 1 eq) in DMF (3 mL). The solution was stirred at rt for 24 h. This was diluted with water (10 mL) causing a milky white solution. The solution was filtered to give white solid. Resulting crude product was purified using with column chromatography (Hex:EtOAc, 10-30%) to give a white solid (0.070 g, 0.260 mmol, 43.2%). ^{1}H NMR (400 MHz, $CDCl_3$) δ_H 13.88 (1H, s, OH), 7.29 – 7.18 (5H, m, Bn – H), 6.00 (1H, d, J = 2.4, H – 3), 5.86 (1H, d, J = 2.4, H - 5), 4.92 (2H, s, CH_2), 3.70 (3H, s, OCH_3), 2.47 (3H,

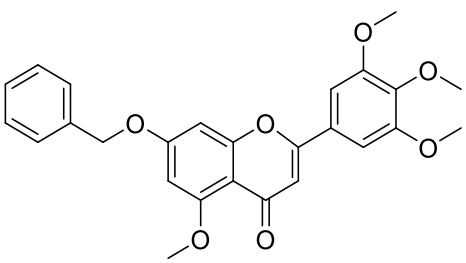
s, $COCH_3$).[158] (in line with published ^1H data). ^{13}C NMR (100 MHz, $CDCl_3$) δ_c 203.1 (CO), 167.5 (C – 2), 165.1 (C – 4), 162.9 ($COCH_3$), 135.8 (Bn – 1), 128.7, 128.3, 127.6, 106.1 (C – 1), 94.4 (C – 3), 91.3 (C – 5), 70.2 (CH_2), 55.5 (OCH_3), 32.9 (CH_3). ESI-HRMS: m/z $[M+H]^+$ Calcd. 273.1121, found 273.1135.

2.2.3.4.3 Synthesis of (E)-1-(4-(benzyloxy)-2-hydroxy-6-methoxyphenyl)-3-(3,4,5-trimethoxyphenyl)prop-2-en-1-one, 2.12.



1-(4-(Benzyloxy)-2-hydroxy-6-methoxyphenyl)ethan-1-one (0.070 g, 0.260 mmol, 1 eq) and 3,4,5-trimethoxybenzaldehyde (0.050 g, 0.260 mmol, 1 eq) were dissolved in methanol (5 mL) with sodium hydroxide (0.040 g, 1.04 mmol, 4 eq) and refluxed at 80 °C for 24 h. The reaction was cooled and a few drops of 6M HCl were added to cause a bright yellow precipitate to form. The precipitate was filtered and collected as a pure product. (0.050 g, 0.110 mmol, 43.5%). ^1H NMR (400 MHz, DMSO-d_6) δ_H 13.07 (1H, s, OH), 7.67 – 7.53 (2H, m, H – α, H – β), 7.48 – 7.34 (5H, m, Bn – H), 7.04 (2H, s, H – 2', H – 6'), 6.24 (2H, dd, J = 15.5, 2.3, H – 6, H – 8), 5.18 (2H, s, CH_2), 3.87 (3H, s, OCH_3 - 5), 3.84 (6H, s, OCH_3 – 3', OCH_3 – 5'), 3.71 (3H, s, OCH_3 – 4'). ^{13}C NMR (100 MHz, DMSO-d_6) δ_c 192.4 (CO), 164.5 ($COCH_3$ – 5), 164.1 (COBn), 161.5 (CHO), 153.1 ($COCH_3$ – 5'), 142.8 (H – β), 139.6 ($COCH_3$ – 4'), 136.4 (BnH - 1), 130.3 (BnH - 3, BnH – 5), 128.5 (BnH – 2, BnH - 6), 128.1 (BnH - 4), 127.9 (C – 1'), 127.1 (H - α), 106.8 (C – 4a), 105.9 (C – 2', C – 6'), 94.7 (C – 8), 91.7 (C – 6), 69.7 (CH_2), 60.1 (OCH_3 – 4'), 56.2 (OCH_3 – 5), 56.0 (OCH_3 – 3', OCH_3 – 5'). ESI-MS: m/z $[M+H]^+$ Calcd. 451.1751 found 451.1757.

2.2.3.4.4 Synthesis of 7-(benzyloxy)-5-methoxy-2-(3,4,5-trimethoxyphenyl)-4H-chromen-4-one, 2.13.

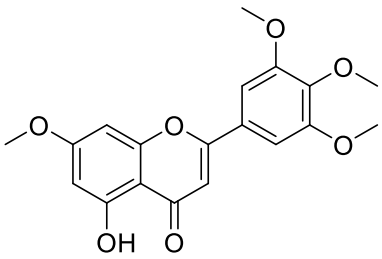


Iodine (5.76 mg, 0.020 mmol, 0.2 eq) was added to of (E)-1-(4-(benzyloxy)-2-hydroxy-6-methoxyphenyl)-3-(3,4,5-trimethoxyphenyl)prop-2-en-1-one (51.1 mg, 0.110 mmols, 1 eq) in DMSO (5 mL) and refluxed at 130 °C for 2.5 h. The

reaction mixture was cooled and poured into 10% [%w/w] sodium thiosulphate pentahydrate solution (50 mL) to give a cloudy orange solution with brown solid precipitate which was collected by filtration. The crude product was purified via column chromatography (DCM:MeOH, 0-10%) and then using reverse phase Biotage Isolera purification (eluent: H_2O + 0.1% formic acid: MeOH + 0.1% formic acid 5-100%, cartridge: Biotage SNAP Ultra C18 17g) to give the desired product. (10.7 mg, 0.020 mmol, 21.7%). ^{1}H NMR (400 MHz, DMSO-d_6) δ_H 7.52 – 7.37 (5H, m, Bn – H), 7.29 (2H, s, H – 2', H – 6'), 7.04 (1H, d, J = 2.2, H – 6/H – 8), 6.86 (1H, s, H – α), 6.59 (1H, d, J = 2.2, H – 6/H – 8), 5.24 (2H, s, CH_2), 3.90 (6H, s, OCH_3 – 3', OCH_3 – 5'), 3.82 (3H, s, OCH_3 – 5), 3.74 (3H, s, OCH_3 – 4'). ^{13}C NMR (100 MHz, DMSO-d_6) δ_c 175.9 (CO), 162.9 (COBn) , 160.3 (C – 1) , 159.5 ($COCH_3$ – 5) , 159.2 (C – 9), 153.3 ($COCH_3$ – 3', $COCH_3$ – 5'), 140.3 ($COCH_3$ – 4') , 136.2 (Bn – 1), 128.6 (Bn – 3, Bn – 5) , 128.3 (C – 1'), 128.2 (Bn – 2, Bn – 6), 126.3 (Bn - 4), 108.5 (C – 4), 108.2 (C – 2), 103.7 (C – 2', C – 6'), 96.8 (C – 6), 94.5 (C – 8), 70.2 (CH_2), 60.3 (OCH_3 – 4'), 56.4 (OCH_3 – 3', OCH_3 – 5'), 56.2 (OCH_3 – 5) ESI-HRMS: m/z $[M+H]^+$ Calcd.449.1595, found 449.1618.

2.2.3.5 Synthesis of 5-OBn analogue

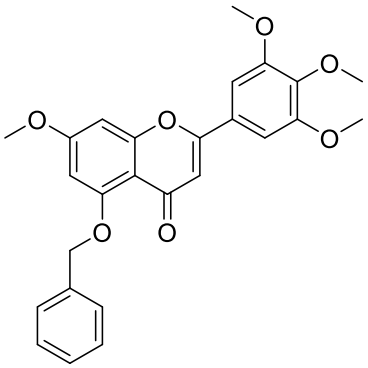
2.2.3.5.1 Synthesis of 5-hydroxy-7-methoxy-2-(3,4,5-trimethoxyphenyl)-4H-chromen-4-one, 2.14.



$AlCl_3$ (0.260 g, 1.97 mmol, 1.47 eq) was added to 5,7-dimethoxy-2-(3,4,5-trimethoxyphenyl)-4H-chromen-4-one (0.500 g, 1.34 mmol, 1 eq) and placed in inert atmosphere, dry chlorobenzene (20 mL) was added and the reaction mixture refluxed at 180 °C for 24 h. After cooling, the resulting mixture was concentrated under reduced pressure, ice cold H_2O: HCl (1:1, 40 mL) was added and the mixture sonicated to give an orange precipitate which was isolated by filtration. The crude product was purified using column chromatography (Hex: EtOAc, 0-50%) to give the desired product as a white solid. (0.130 g, 0.360 mmol, 27.1%). ^{1}H NMR (400 MHz, DMSO-d_6) δ_H 12.85 (1H, s, OH), 7.37 (2H, s, H – 2', H – 6'), 7.14 (1H, s, H – α), 6.86 (1H, d, J = 2.2, H – 8), 6.39 (1H, d, J = 2.2, H – 6), 3.91 (6H, s, OCH_3 – 3', OCH_3 – 5'), 3.89 (3H, s, OCH_3 – 7), 3.76 (3H, s, OCH_3 – 4').[159] (in line with published ^{1}H data). ^{13}C NMR (100 MHz, DMSO-

d_6) δ_c 182.3 (CO), 165.5 ($COCH_3$ – 7), 163.5 (C -1), 161.3 (CHO), 157.6 (C – 8a), 153.5 ($COCH_3$ – 3', $COCH_3$ – 5'), 141.1 ($COCH_3$ – 4'), 126.1 (C – 1'), 105.4 (C – α), 105.1 (C - 4a), 104.5 (C – 2', C – 6'), 98.4 (C – 6), 93.1 (C – 8), 60.5 (OCH_3 – 4'), 56.6 (OCH_3 – 3', OCH_3 – 5'), 56.4 (OCH_3 – 7). ESI-MS: m/z $[M+H]^+$ Calcd. 359, found 359, $[M+Na]^+$ Calcd. 381, found 381.

2.2.3.5.2 Synthesis of 5-(benzyloxy)-7-methoxy-2-(3,4,5-trimethoxyphenyl)-4H-chromen-4-one, 2.15.

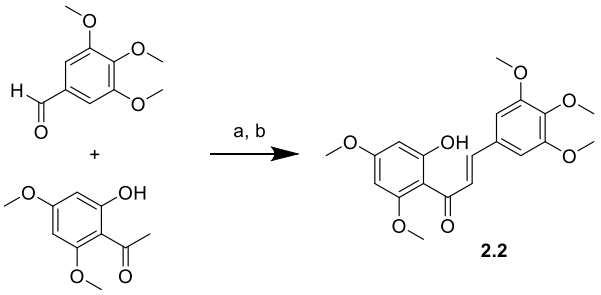


Benzyl bromide (0.120 mL, 1.01 mmol, 2 eq) was added to a solution of 5-hydroxy-7-methoxy-2-(3,4,5-trimethoxyphenyl)-4H-chromen-4-one (0.130 g, 0.350 mmol, 1 eq) and potassium carbonate (0.140 g, 1.00 mmol, 3 eq) in DMF (10 mL). The solution was stirred at rt for 96 h. This was diluted with water (20 mL) causing a milky white solution. The solution was filtered to give white solid. Resulting crude product was purified using reverse phase Biotage Isolera purification (eluent: H_2O + 0.1% formic acid: MeOH + 0.1% formic acid 5-100%, cartridge: Biotage SNAP Ultra C18 12g) to give the desired product. (0.070 g, 0.160 mmol, 45%). ^{1}H NMR (400 MHz, DMSO-d_6) δ_H 7.61-7.31 (5H, m, Bn – H), 7.29 (2H, s, H – 2', H – 6'), 6.91 (1H, d, J = 2.2, H – 8), 6.85 (1H, s, H – α), 6.59 (1H, d, J = 2.2, H – 6), 5.22 (2H, s, CH_2), 3.90 (6H, s, OCH_3 – 3', OCH_3 – 5'), 3.88 (3H, s, OCH_3 - 7), 3.73 (3H, s, OCH_3 – 4'). ^{13}C NMR (100 MHz, DMSO-d_6) δ_c 175.8 (CO), 163.6 ($COCH_3$ – 7), 159.5 (C – 1), 159.2 ($COCH_2Bn$), 158.9 (C – 8a), 153.3 ($COCH_3$ – 3', $COCH_3$ – 5'), 140.2 ($COCH_3$ – 4'), 136.9 (Bn – 1), 128.4 (Bn – 3, Bn – 5), 127.6 (Bn – 4), 126.9 (Bn – 2, Bn – 6), 126.3 (C – 1'), 108.7 (C – 4a), 108.2 (C – α), 103.6 (C – 2', C – 6'), 97.8 (C – 6), 93.8 (C – 8), 69.9 (CH_2), 60.2 (OCH_3 – 4'), 56.3 (OCH_3 – 3', OCH_3 – 5'), 56.1 (OCH_3 – 7). ESI-HRMS: m/z $[M+H]^+$ Calcd. 449.1595, found 449.1683. Found m/z value is 0.0088 greater than calculated.

2.3 Results and Discussion

2.3.1 Synthesis of PMF

Unmodified PMF (**2.1**) was synthesised in two steps, firstly the formation of a chalcone intermediate (**2.2**) and then oxidation to the flavone structure from an adapted flavone synthesis method. The synthetic route followed is a commonly adopted route for the synthesis of the flavone structure, mimicking natural biosynthetic flavone formation pathways in plants, and has been followed in the literature many times.[160,161] The chalcone intermediate was prepared using a crossed aldol condensation between an aldehyde and ketone component in the presence of base under reflux in methanol (Scheme 2.1).



Scheme 2.1: Synthesis of intermediate chalcone, 2.2 for synthesis of 5,7,3',4',5'-pentamethoxyflavone. Reagents and conditions: a) NaOH, MeOH, 60 °C, 20 h; b) 6M HCl.

The desired compound precipitated out of solution as a bright yellow solid after neutralisation by dropwise addition of 6M HCl and the solid purified by recrystallisation from ethanol with a yield of 68%. The structure was characterised with ^{1}H, ^{13}C NMR alongside 2D NMR spectra (HSQC, HMBC and COSY) and ESI mass spectrometry (Appendix 8.1).

After purification and characterisation of the pentamethoxychalcone structure, catalytic iodine was used to form the 6-membered flavone ring, facilitated by the use of DMSO as a solvent.[162] DMSO encourages regeneration of the iodine catalyst, allowing its use in small quantities. Flavone formation from 2'-hydroxychalcones is proposed to occur via an oxidative cyclisation mechanism this, and the mechanism for I_2 regeneration via DMSO, are shown in Figure 2.4.[152] The cyclisation is thought to

occur via an *oxo*-Michael addition that is trapped by the addition of an X-Y reagent, commonly I_2, or by isomerising to a flavanone structure and then oxidising to the flavone (**2.1**)[152]

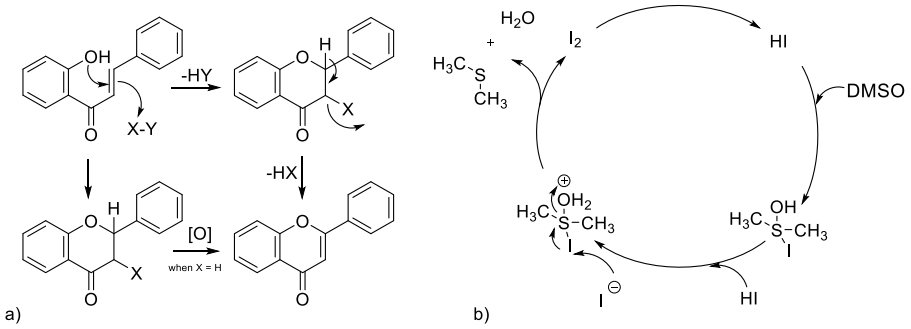
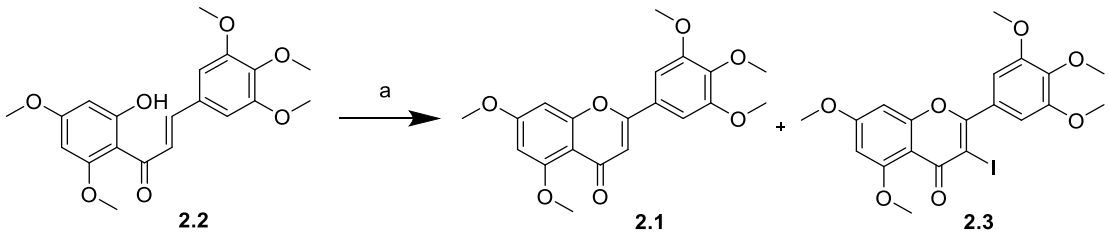


Figure 2.4: a) Two proposed mechanisms for the formation of flavones via oxidative cyclisation. b) Proposed mechanism of catalytic I_2 regeneration using DMSO.[152]

Refluxing the intermediate 2-hydroxychalcone structure **2.2**, in DMSO at 130 °C until complete consumption of starting material allowed the formation of the desired pentamethoxyflavone structure in good yields. Addition of the cooled reaction mixture to a solution of 10% [%w/w] sodium thiosulphate which reduces the excess iodine in solution caused the precipitation of an off-white solid which could be isolated *via* filtration (Scheme 2.2). The crude product isolated with filtration was found to be > 90 % pure, with a single small impurity formed as a by-product during the reaction. The side product was identified as an iodoflavone (**2.3**) and was characterised after purification using ^{1}H NMR and ESI-MS (Appendix 8.2).

The quantity of iodo side product produced was found to be time dependent, so reaction times were shortened by the careful monitoring of the consumption of starting material *via* TLC to optimise desired product formation.



Scheme 2.2: Synthesis of 5,7,3',4',5'-pentamethoxyflavone, **2.1**, using catalytic iodine and undesired iodoflavone side product, **2.3**. Reagents and conditions: a) I_2, DMSO, 130 °C.

Due to a lack of visible separation between the desired flavone and iodoflavone side product **2.3**, using normal phase TLC, alternative methods of purification were investigated. Visualisation of the crude sample using LC-MS with a MeOH/H_2O + 0.1 % formic acid mobile phase showed resolved peaks in a UV chromatogram (Figure 2.5), indicating that the impurity could be separated by reverse-phase chromatography with a C18 column.

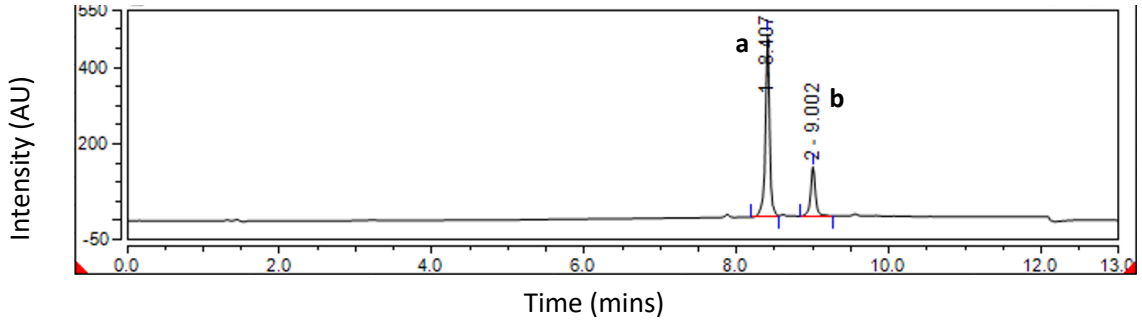
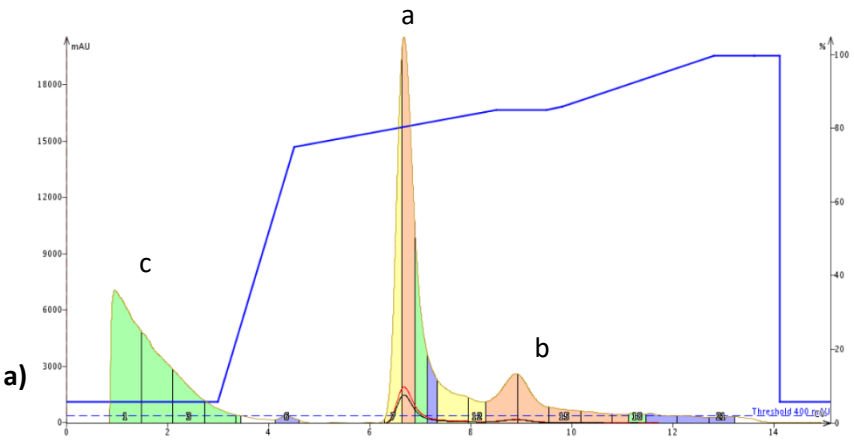


Figure 2.5: UV chromatogram of crude product from flavone reaction taken from LCMS analysis, showing resolved peaks of desired product 2.1, a, at T = 8.4 min and iodoflavone impurity 2.3, b, at T = 9.0 min. LCMS gradient: 5 – 100% MeOH/H_2O + 0.1% formic acid.

Reverse-phase purification was carried out using a Biotage® Isolera™ One system, using a C18 SNAP column as a stationary phase and methanol and distilled 18.2 MΩ ultrapure water with 0.1 % formic acid, as mobile phases. The crude mixture was dissolved in DMF and loaded to the column in liquid phase, a gradient of 5-100 % MeOH allowed separation of the impurity from the desired product (Figure 2.6).

a)



b)

Column volume	% MeOH
0 – 3	5
3 – 4.5	5 – 75
4.5 – 8.5	75 – 85
8.5 – 9.5	85
9.5 – 12.5	85 – 100
12.5 - 14	100
14 - 15	5

Figure 2.6: a) Graph from Biotage® Isolera™ One system purification of 5,7,3',4',5'-pentamethoxyflavone, showing isolation of the desired product peak, a, from the impurity, b, the graph also shows the DMF solvent front leaving the column, c. b) Table showing methanol gradient for purification.

This method of purification resulted in the desired product in varying yields, due to fluctuating efficiency of separation. The product was characterised by ^{1}H, ^{13}C NMR alongside 2D NMR spectra (HSQC, HMBC and COSY) and ESI mass spectrometry (Figure 2.7, Appendix 8.3). The purity of the samples carried forward for biological testing was analysed using the UV chromatogram data from LCMS (Appendix 8.3.7).

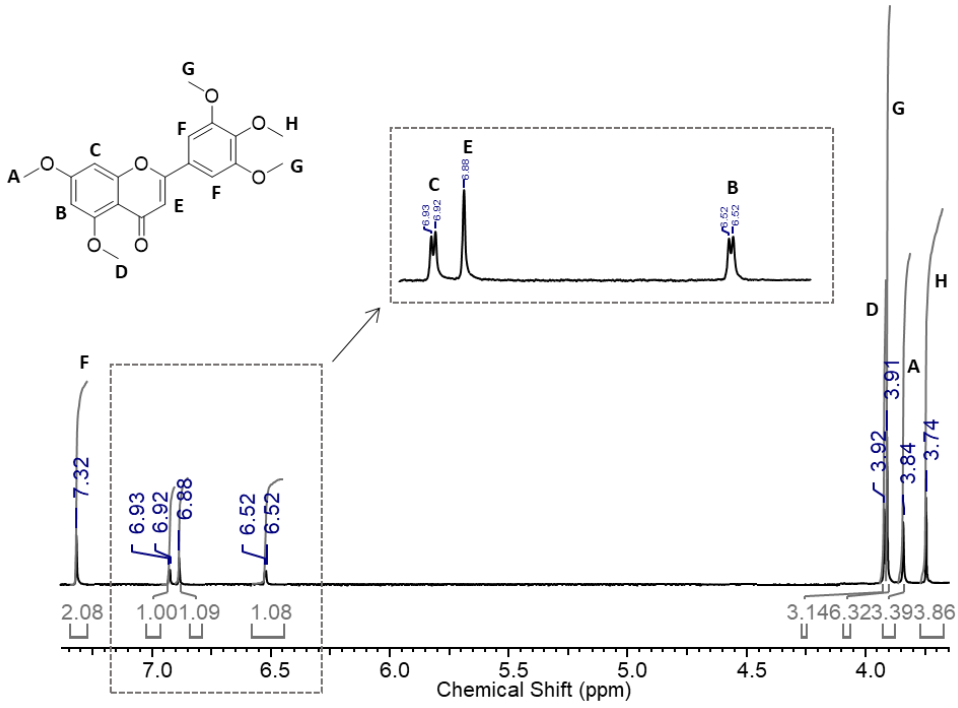
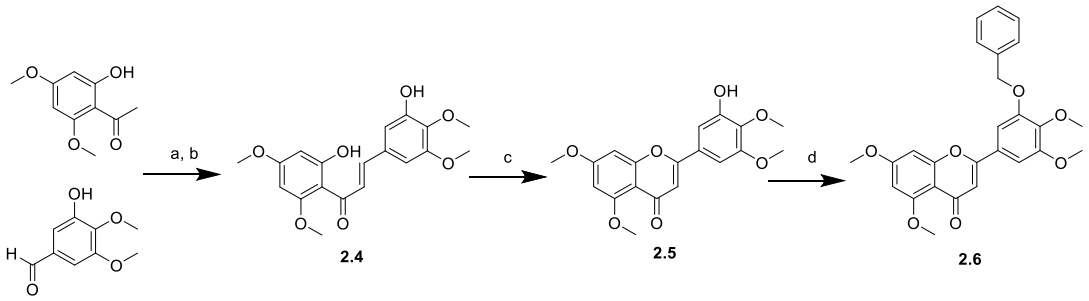


Figure 2.7: Assignments of ^{1}H NMR for 5,7,3',4',5'-pentamethoxyflavone showing magnification of multiplet structures of coupling phenyl hydrogens.

2.3.2 Synthesis of the 3'-OBn analogue

Synthesis of each modified flavone required a slightly different route, though followed the same key steps as the synthesis of PMF. The synthesis of the flavone modified with a benzyl group in the 3'-position is depicted in the scheme below (Scheme 2.3).



Scheme 2.3: Synthetic route of 3'-benzyl-5,7,4',5'-tetramethoxyflavone. Reagents and conditions: a) NaOH, MeOH, 80 °C, 24 h; b) 6M HCl; c) I_2, DMSO, 130 °C, 2h; d) BnBr, K_2CO_3, rt, 72 h.

The first step of the synthesis followed the same aldol reaction as used for PMF, resulting in the formation of a bright yellow precipitate on addition of 6M HCl, this was purified using a Biotage Isolera system in a hexane:ethyl acetate normal phase system. Catalytic iodine was then added to cyclise the chalcone **2.4**, to form the flavone **2.5**. The formation of the flavone structure with a hydroxyl substituent in the 3' – position was shown to be successful, without the use of a protecting group in the early stages of synthesis; this was not seen for analogues **4'-OBn** and **7-OBn**, which will be discussed further in sections 2.3.3 and 2.3.4 respectively. Characterisation data for both the chalcone and flavone structures can be found in Appendix 8.4 and 8.5 respectively. The final step to produce the benzyl-substituted analogue was achieved by stirring with benzyl bromide and potassium carbonate at rt *via* an S_N2 substitution mechanism aided by the phenyl ring. The final product **2.6**, was characterised by ^{1}H and ^{13}C NMR, ESI-MS and purity assessed with LCMS. (Figure 2.8, Appendix 8.6).

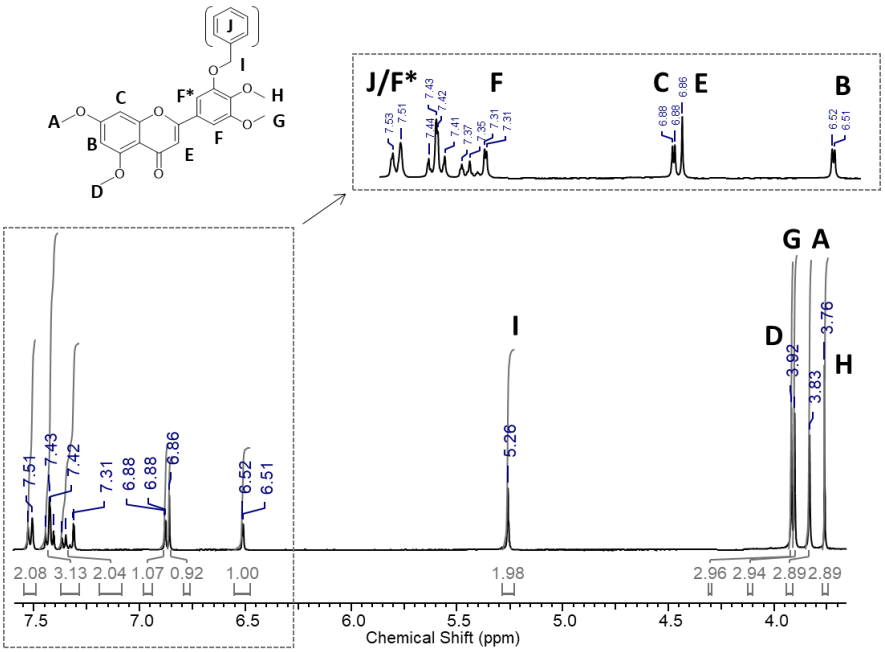
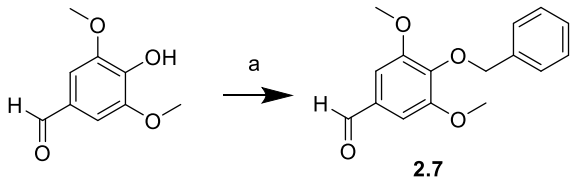


Figure 2.8: Assignments of ^{1}H NMR for 3'-benzyl-5,7,4',5'-tetramethoxyflavone showing magnification of multiplet structures of coupling phenyl hydrogens.

2.3.3 Synthesis of the 4'-OBn analogue

Synthesis of the flavone structure modified in the 4'-position **2.9**, was initially tested following the same initial step of the protocol used for modification in the 3'-position, detailed in experimental

methods 2.2.3.2.1. A chalcone was formed by reacting 1-(2-hydroxy-4,6-dimethoxyphenyl)ethanone and 4-hydroxy-3,5-dimethoxybenzaldehyde (Appendix 8.7), however when carrying this forward to the cyclisation step product formation was unsuccessful. It was suggested by Masesane that the iodine-catalysed flavone formation is not tolerant to the presence of hydroxyl groups in both the 4'- and 7-position on the flavone ring.[152] To circumvent this issue, the benzyl substituent, which is used as the ring modification for the SAR study, was installed as an initial step in the synthesis to act as a protecting group for the free hydroxyl and allow the formation of the flavone structure **2.7** during the cyclisation step.



Scheme 2.4: Formation of benzyl ether, **2.7**, as a protecting group for the hydroxyl group in the 4'-position of 4-hydroxy-3,5-dimethoxybenzaldehyde. Reagents and conditions: a) BnBr, K_2CO_3, DMF, rt, 24 h.

The benzyl protecting group was introduced prior to the aldol step of the flavone synthesis, formation of the benzyl ether was achieved in the same conditions as the final step of the synthesis of 3'-benzyl-5,7,4',5'-tetramethoxyflavone as depicted in Scheme 2.4. The desired product formed after stirring at rt for 24 h and was isolated initially *via* filtration to obtain the precipitate formed upon addition of water and then *via* extraction of the filtrate with diethyl ether and concentration to an oil under vacuum. Aldehyde **2.7** was analysed using ^{1}H and ^{13}C NMR and mass spectrometry, proton NMR assignments can been seen in

Figure **2.9** and the remaining characterisation data in the Appendix 8.8.

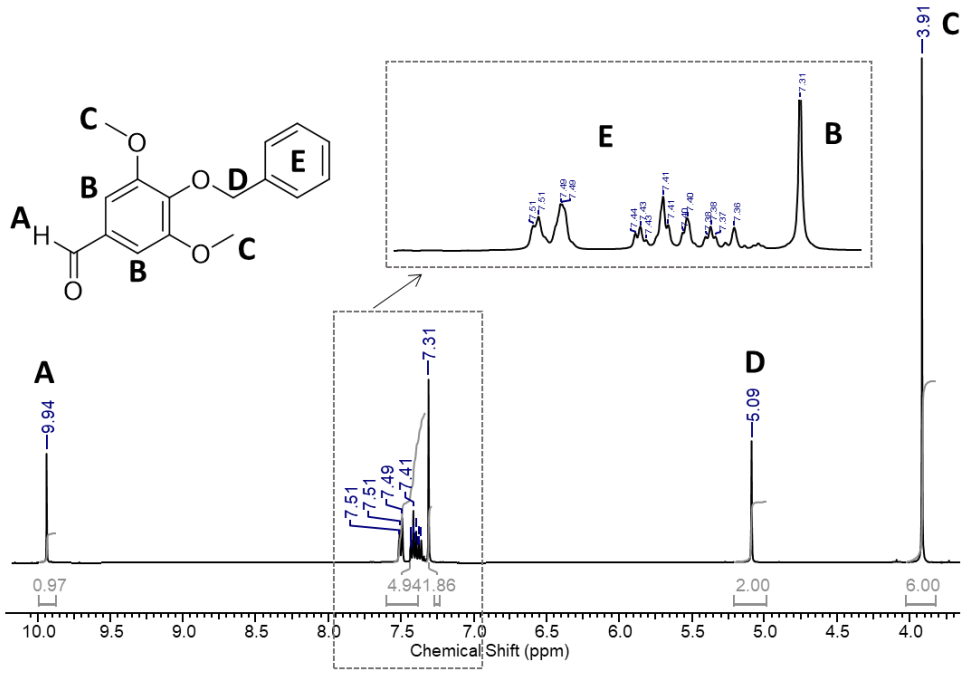
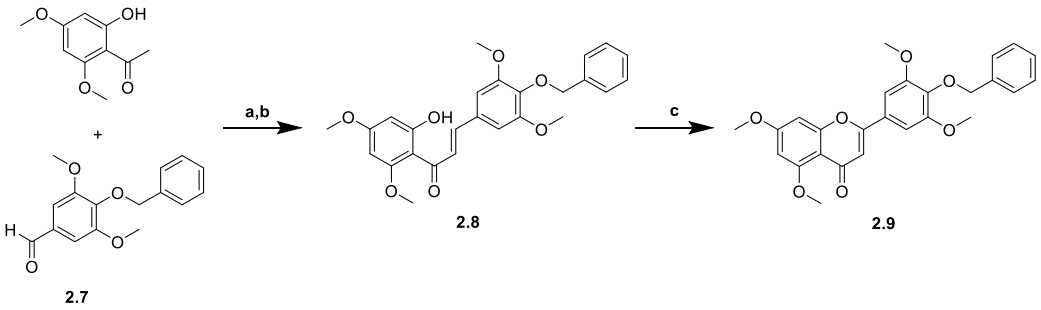


Figure 2.9: ^{1}H NMR assignments for **2.7**, 4-(benzyloxy)-3,5-dimethoxybenzaldehyde showing magnification of multiplet structures of coupling phenyl hydrogens.

The benzyl protected benzaldehyde was then used to form 4'-benzyl-5,7,3',5'-tetramethoxyflavone, 2.9, using the conditions for the aldol and cyclisation steps used for the previously discussed flavone structures (Scheme 2.5).

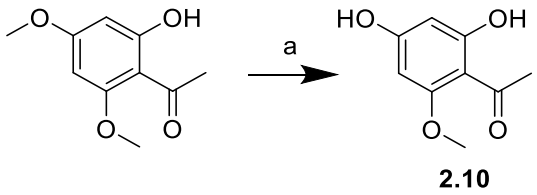


Scheme 2.5: Synthesis of 4'-benzyl-5,7,3',5'-tetramethoxyflavone, **2.9**, from 4-(benzyloxy)-3,5-dimethoxybenzaldehyde. Reagents and conditions: a) NaOH, MeOH, 80 °C, 24 h; b) 6M HCl; c) I_2, DMSO, 130 °C, 1 h.

Intermediate **2.8** and final product **2.9** were characterised using proton, carbon and 2D NMR alongside high-resolution mass spectrometry, due to the novel nature of intermediate compound **2.8**. Purity values were again assessed by LCMS analysis. (Appendix 8.9-8.10).

2.3.4 Synthesis of the 7-OBn analogue

Synthesis of the analogue modified by a benzyl ether in the 7-position of the structure required a preliminary synthetic step to obtain the required starting material **2.10** (Scheme 2.6). The commercially available starting material, 1-(2-hydroxy-4,6-dimethoxyphenyl)ethan-1-one, was subjected to a demethylation reaction using $AlCl_3$ adapted from the method suggested by Wei *et al.*[163]



2.10

Scheme 2.6: Synthesis of 1-(2,4-dihydroxy-6-methoxyphenyl)ethan-1-one, **2.10**, from 1-(2-hydroxy-4,6-dimethoxyphenyl)ethan-1-one via a single demethylation. Reagents and conditions: a) $AlCl_3$, dry chlorobenzene, 130 °C, 6 h.

The reaction results in a demethylation preferentially in the *para* position, however extended reaction times can result in total demethylation to give the undesirable trihydroxyketone product, so monitoring *via* TLC was necessary. Purification *via* column chromatography using a Biotage Isolera system allowed for the isolation of the required species, however did not afford very high yields, at 23%. Further reaction optimization of this synthetic route could result in higher yields by shortening reaction times to produce more product. Compound **2.10** was characterised using ^{1}H, ^{13}C, 2D NMR and ESI-MS (

Figure **2.10**, Appendix 8.11).

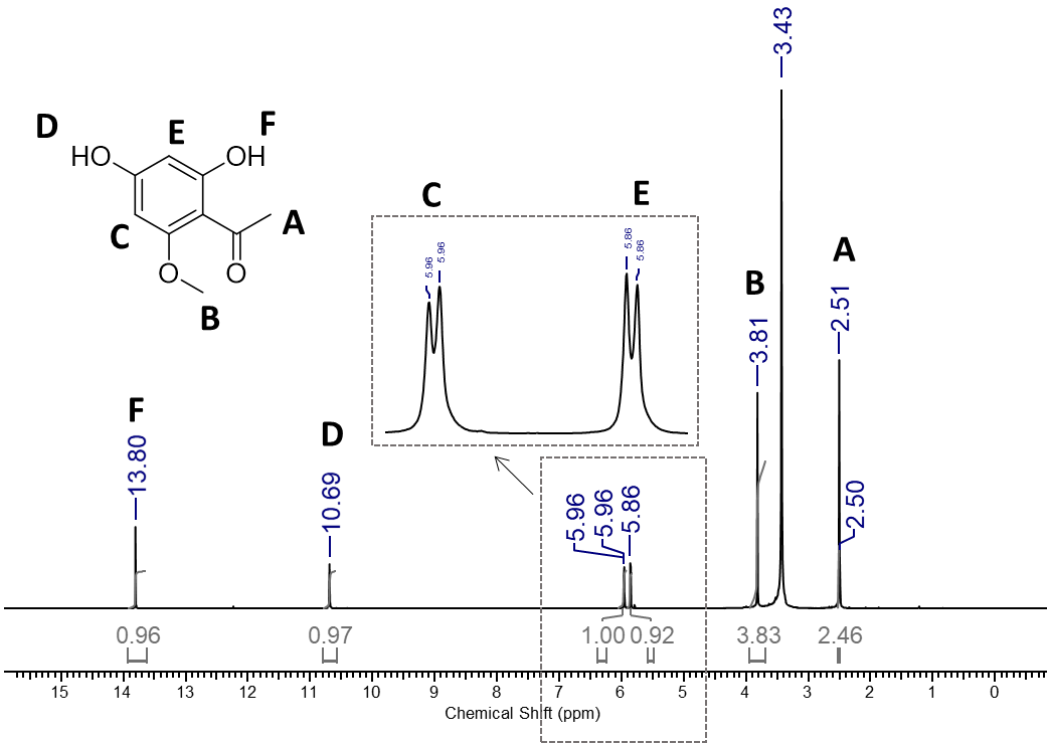


Figure 2.10: ^{1}H NMR assignments of 1-(2,4-dihydroxy-6-methoxyphenyl)ethan-1-one, **2.10**.

At this point in the synthesis the benzyl ether was formed, this was to prevent interference of the second phenol group with the chalcone and flavone synthetic steps, as these could hinder flavone formation similarly to the synthesis of the **4'-OBn** analogue. The formation of the desired benzyl ether was achieved using the same reaction conditions as for the other analogues, however, only one equivalent of benzyl bromide was used to prevent the formation of a dibenzyl compound. The desired isomer was formed as the major product, similarly to the synthesis by Schreier *et al.* who obtained the same *para*-substituted product[158] - thought to be due to the preferable sterics associated with ether formation in the *para*-position to the ketone - and the product was isolated using column chromatography. Ketone **2.11** was characterised by ^{1}H, ^{13}C alongside 2D NMR analysis including COSY, HSQC and HMBC experiments, and ESI-MS. Confirmation of the location of the benzyl ether on the ring was accomplished using Nuclear Overhauser Effect Spectroscopy (NOESY) NMR analysis (Figure 2.11) which allowed visualisation of through space interactions between the benzyl CH_2 and both aryl hydrogens, as depicted by the letter a on the spectrum, and the absence of interactions between this

CH_2 and the ketone methyl group, b, validating the formation of the correct isomer. (Figure 2.11, Appendix 8.12).

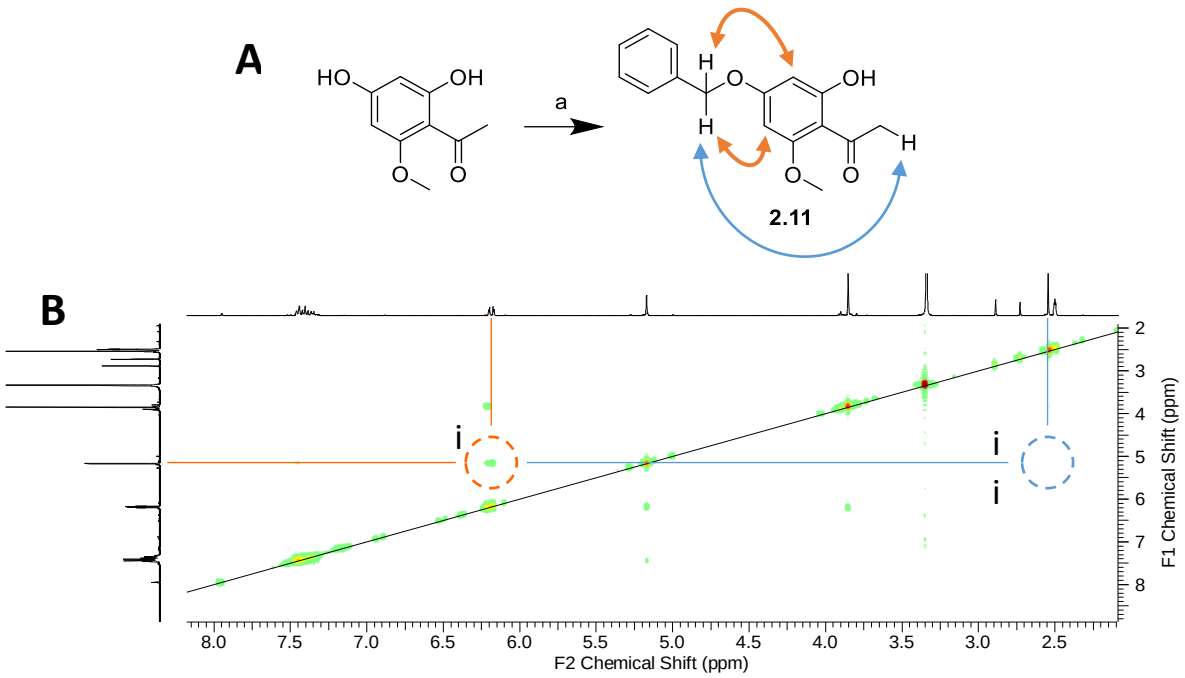
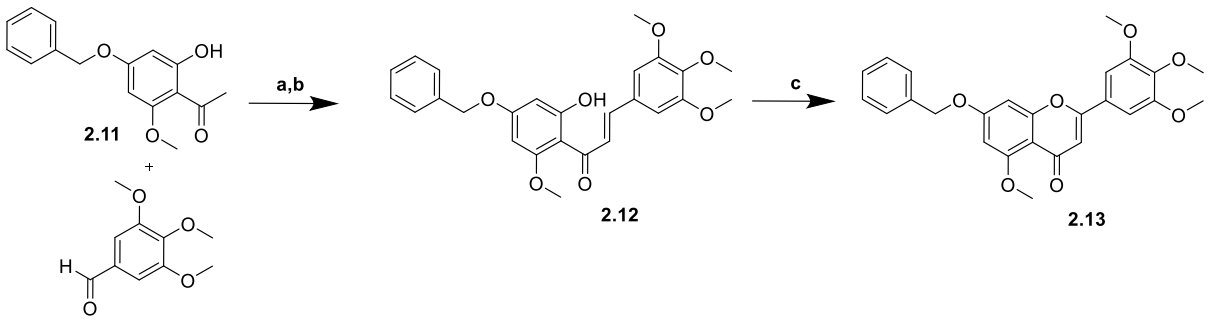


Figure 2.11: A) Synthetic scheme for the formation of 1-(4-(benzyloxy)-2-hydroxy-6-methoxyphenyl)ethan-1-one, **2.11**. Reagents and conditions: a) 1 eq. BnBr, K_2CO_3, DMF, 24 h, rt. B) NOESY spectrum with circles depicting key interactions: i) interaction of the benzyl CH_2 and both aryl hydrogens; ii) the absence of interactions between the CH_2 and the ketone methyl group.

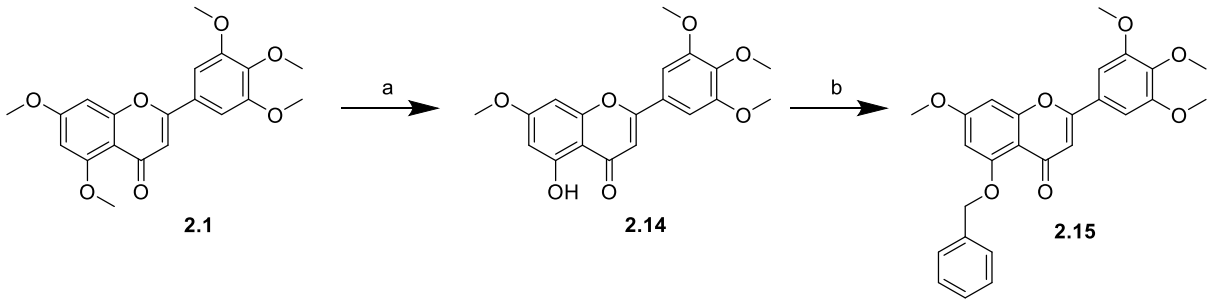
The benzyl protected species was then carried forward in the synthesis of the flavone analogue, using the same reaction conditions as used for the previously discussed structures (Scheme 2.7). Firstly, the formation of chalcone **2.12**, using an aldol synthesis and then cyclisation using catalytic iodine to produce the final flavone structure. Both compounds, **2.12** and **2.13**, were characterised using ^{1}H and ^{13}C NMR and HRMS (Appendix 8.13-8.14).



Scheme 2.7: Synthesis of 7'-benzyl-5,3',4',5'-tetramethoxyflavone **2.13,** from 1-(4-(benzyloxy)-2-hydroxy-6-methoxyphenyl)ethan-1-one. Reagents and conditions: a) NaOH, MeOH, 80 °C, 24 h; b) 6M HCl; c) I_2, DMSO, 130 °C, 2.5 h.

2.3.5 Synthesis of 5-OBn analogue

The route taken to isolate the final flavone analogue, 5-benzyl-7,3',4',5'-tetramethoxyflavone, was different to that of the other modified analogues since the required starting material, 1-(2,6-dihydroxy-4-methoxyphenyl)ethan-1-one, which would be necessary to facilitate this by a mono-benzyl group protection and then the using same method followed for the other flavone formations, was not commercially available. A route using a single demethylation of 1-(2-hydroxy-4,6-dimethoxyphenyl)ethan-1-one, was also unavailable, as the major product formed is the undesired structural isomer, used for the formation of 7-benzyl-5,3',4',5'-tetramethoxyflavone.



Scheme 2.8: Synthesis of 5-benzyl-7,3',4',5'-tetramethoxyflavone **2.15**, from 5,7,3',4',5'-pentamethoxyflavone **2.1**. Reagents and conditions: a) $AlCl_3$, dry chlorobenzene, 130 °C, 24 h. b) BnBr, K_2CO_3, DMF, rt, 96 h.

An alternative method, therefore, was used, this method utilises a selective demethylation of the parent compound 5,7,3',4',5'-pentamethoxyflavone **2.1**, using aluminium trichloride in dry chlorobenzene to remove the methyl group in the 5 position and leave a hydroxyl in its place, as shown in **Scheme 2.8**, this method builds on the technique used for demethylation to prepare the dihydroxyketone required for the synthesis of 7-benzyl-5,3',4',5'-tetramethoxyflavone. The selectivity of the demethylation in the 5-postion of the pentamethoxyflavone is thought to occur due to the chelation of $AlCl_3$ to the neighbouring carbonyl. The location of an electron-withdrawing group placed *ortho* to the methoxy group facilitates the removal of a methyl in this location.[164,165] The product obtained, **2.14**, was characterised by ^{1}H, ^{13}C and 2D NMR and ESI-MS (Figure 2.12, Appendix 8.15). The proton NMR of the product showed the appearance of a signal at 12.96 ppm, when compared to the PMF parent compound, indicating the presence of a hydroxyl group. The signal at 3.92 ppm seen in the PMF spectra that corresponds to the protons in the methoxy group in the 5-position of PMF, is lost in the product NMR and a shift downfield of the signal seen at 3.84 ppm that supports the deshielding of the 7-methoxy substituent after removal of the methyl group.

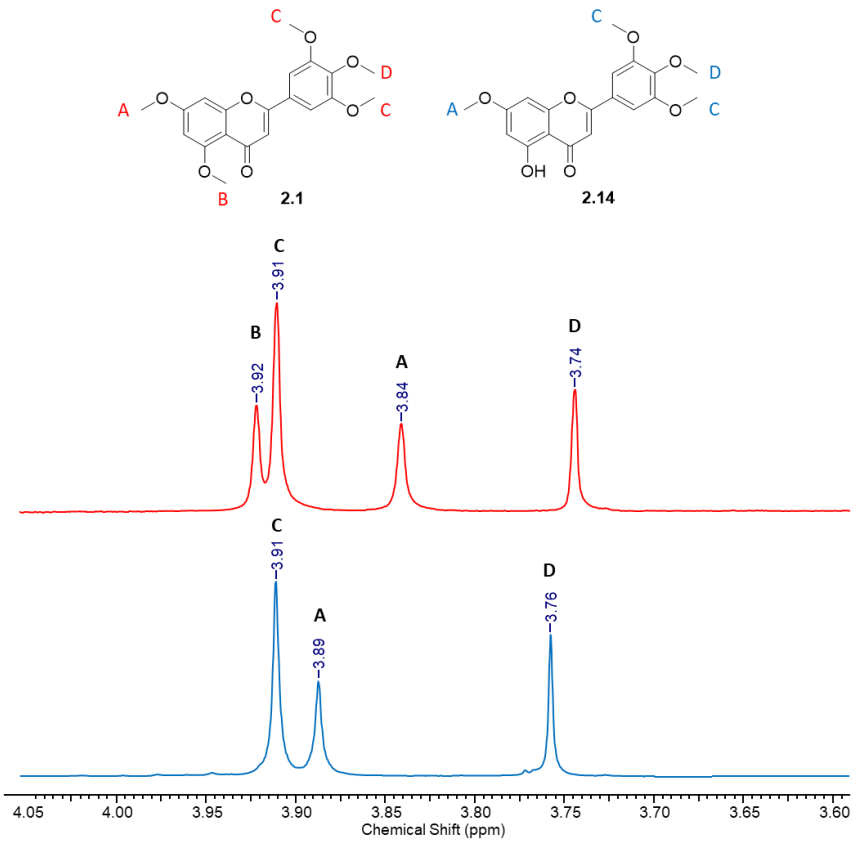


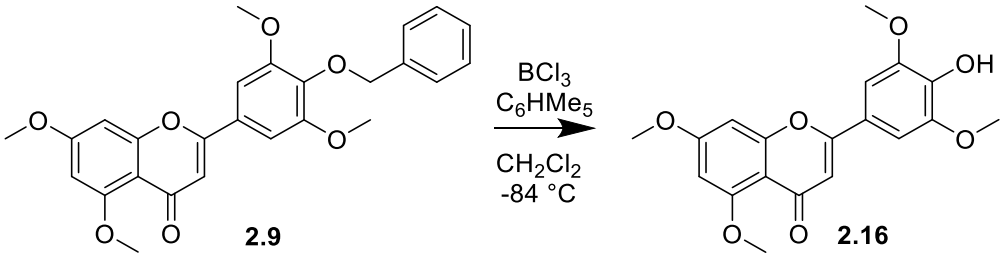
Figure 2.12: Comparison of ^{1}H NMR spectra of ^{1}H NMR showing the peaks corresponding to methoxy substituents on both 5,7,3',4',5'-pentamethoxyflavone **2.1**, and 5-hydroxy-7,3',4',5'-tetramethoxyflavone **2.14**. Letters assigning the methoxy groups to the peaks in the spectra show a loss of the 5 position methoxy group hydrogens.

The benzyl group formation in the 5-position was then carried out in the same reaction conditions as previously used to give **2.15**, this was characterised using ^{1}H, ^{13}C and 2D NMR, HRMS and LCMS analysis. (Appendix 8.16).

2.3.6 Synthesis of hydroxyflavone analogues for study of anti-obesity properties.

Synthesis of two hydroxyflavones, **3'-OH** and **4'-OH**, was carried out to facilitate the research into the potential anti-obesity properties of the flavones. Synthesis of the **3'-OH** flavone, was simple due to its presence as an intermediate compound (**2.5**), in the synthesis of the **3'-OBn** analogue **2.6**, so was easily isolated for analysis.

As previously described, synthesis of the **4'-OBn** analogue **2.9**, required the introduction of a benzyl ether protecting group to ensure the success of the flavone cyclisation step. To generate the **4'-OH** flavone **2.16**, a debenzylation method was needed. This was initially trialled by reacting **2.9** with a Pd/C catalyst in the presence of hydrogen gas. Unfortunately, the debenzylation was not successful and an alternative route was sought. Tokuyama *et al.* published a route for debenzylation using BCl_3 in the presence of pentamethylbenzene, Scheme 2.9, which proved successful in this instance.[166] The hydroxyflavone formed was purified using the Biotage Isolera reverse phase methods developed for other analogues and characterised by ^{1}H and ^{13}C NMR alongside HRMS to confirm its structure (Appendix 8.17).



Scheme 2.9: Synthesis of 2-(4-hydroxy-3,5-dimethoxyphenyl)-5,7-dimethoxy-4H-chromen-4-one **2.16**, from 2-(4-(benzyloxy)-3,5-dimethoxyphenyl)-5,7-dimethoxy-4H-chromen-4-one **2.9**. Reagents and conditions: a) BCl_3, C_6HMe_5, CH_2Cl_2, -84 °C.

The anti-obesity properties of the **4'-OH** and **3'-OH** flavones were studied by collaboration with B. Ahmad and E. H. Wong at Taylor's University, Malaysia. Results indicate that both hydroxyflavones showed promising anti-obesity activity such as a competitive reduction of the activity of pancreatic lipase, suppression of 3T3-L1 mouse embryonic 4 fibroblast cell differentiation during the early stages of adipogenesis and a reduction in the expression of key adipogenic and lipogenic marker genes. Results were consistently favourable for the **4'-OH** flavone suggesting that a hydroxy group in this position may have been beneficial to the structure-activity relationship of the analogues when examining anti-obesity activities.[167]

2.4 Conclusion

Synthesis of the parent compound PMF and four benzyl modified analogues was completed successfully and the final structures, as well as intermediate compounds, were purified and characterised using a selection of NMR and mass spectrometry methods. The final benzyl substituted analogues were then able to be used in cell culture assays to assess and compare biological activity.

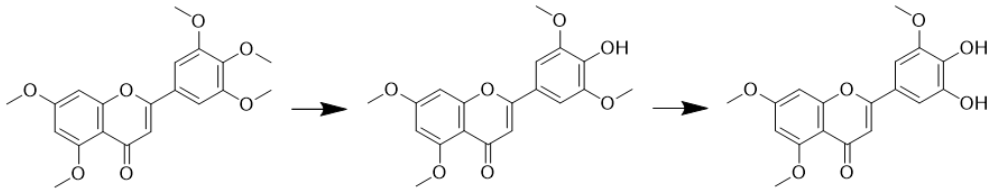
Two mono-hydroxy flavone structures were also synthesised, purified, and characterised for collaborative projects with Ahmed *et al.* that investigated the anti-obesogenic properties of polymethoxy and hydroxyflavones.

3 Evaluation of antiproliferative activity of 3',4',5',5,7-pentamethoxyflavone and modified flavone analogues to establish a suitable modification position.

3.1 Introduction

3.1.1 Antiproliferative activity and current structure-activity relationship understanding of 3',4',5',5,7-pentamethoxyflavone (PMF).

As for many other flavones, 3',4',5',5,7-pentamethoxyflavone (PMF) has been shown in previous research to exhibit antiproliferative activity in numerous cancerous cell lines, as discussed in Chapter 1. Vongdeth *et al.* recently synthesized and evaluated the *in vitro* antiproliferative activity of a selection of polymethoxyflavones and polymethoxychalcones, including PMF. This compound was screened against HeLa (cervical carcinoma), HCC1954 (breast cancer) and SK-OV-3 (ovarian cancer) cell lines and showed moderate antiproliferative activity in the micromolar range in these cell lines with GI_{50} values of 40.82, 53.84 and 30.17 µM respectively, assessed in this work using a CCK-8 assay. The positive control used for comparison, cisplatin, was found to give GI_{50} values of 13.30, 29.32 and 18.66 µM for the same cell lines. Comparison of the activity of the different polymethoxyflavones in the study allowed a brief insight into structure-activity relationships (SAR) within these compounds. Replacement of the methoxy substituent with a hydroxyl group in the 4'-position of the structure resulted in loss of antiproliferative activity in the assay used, but this activity was regained with subsequent replacement of the methoxy group in the 5'-position with a second hydroxy substituent (Figure 3.1).[77]



PMF ii iii

Cell line	GI_{50}		
	PMF	ii	iii
HeLa	40.82	>100	4.83
HCC1954	53.84	>100	8.58
SK-OV-3	30.17	>100	10.64

Figure 3.1: Representation of GI_{50} values of different polymethoxyflavones in 3 cancer cell lines tested by Vongdeth *et al.* showing the SAR of the hydroxyl groups on the B ring of the flavone backbone. Activity of the compounds was assessed using a CCK-8 assay. Structures refer to synthesised flavones 9c (PMF), 9b (ii) and 9a (iii).[77]

Sousa *et al.* evaluated the cytotoxic effect of treatment of 3 cancer cell lines, T24 (bladder carcinoma), TOV-21-G (ovarian adenocarcinoma) and HepG2 (liver carcinoma), with five polymethoxyflavone compounds, including PMF. The compound showed cytotoxic activity when treating the T24 and TOV-21-G cell lines, with GI_{50} values of 22.91 and 45.46 µM respectively. However, treatment of HepG2 cells resulted in GI_{50} values of > 100 µM showing inactivity of the flavone.[168]

The anticancer activity of PMF has been studied more extensively for the potential treatment of colorectal cancer (CRC). As introduced in Chapter 1, Cai *et al.* studied the chemopreventive properties of PMF for colorectal cancer by measuring its ability to inhibit the growth of preneoplastic APC10.1 adenoma cells; a model cell line developed from Apc^{Min} mice for *in vitro* studies of chemopreventive agents for familial adenomatous polyposis and colorectal carcinomas.[169,170] PMF was compared to two structurally related natural flavones, tricin and apigenin (Figure 3.2), to evaluate the significance of the O-methylation on efficacy. The results showed that PMF exhibited superior anticancer properties to tricin and apigenin in all areas studied and was found to be twice as potent as tricin and three times as potent as apigenin in inhibition of APC10.1 cell growth, with GI_{50} values of 6, 13 and 18 µM

respectively. This study also showed that PMF interferes with the cyclooxygenase (COX) pathway and inhibits the generation of prostaglandin E-2 (PGE-2) by HCA-7 colon adenocarcinoma cells at lower concentrations than apigenin and tricin. Increases in PGE-2 production by cancer cells have been linked to several processes involved in carcinogenesis. PMF was also found to bind preferentially to the active site of COX-1 and COX-2 enzymes *in silico* when compared to the binding affinity of apigenin and tricin.[170,171]

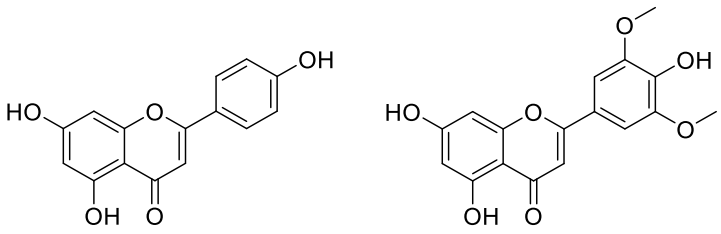


Figure 3.2: Structures of apigenin and tricin, structurally related natural flavones used for comparison of antiproliferative activity with PMF.

Research submitted for the thesis of Isabel L. Fong, a collaborator for this work, studied the antiproliferative properties of PMF, apigenin and tricin on APC10.1 cells, and showed that PMF exhibited superior antiproliferative activity to that of apigenin and tricin, with GI_{50} values of 5.26 µM after 3 days of treatment and 4.99 µM after 6 days of treatment. PMF was also found in this study to possess better abilities than the other flavones to cause cell cycle arrest and trigger apoptosis. In a cDNA microarray analysis PMF was shown to modulate several signalling pathways including Wnt, PI3K/AKT/GSK3β and JAK/STAT.[172]

3.1.2 Seeding density assay background

To successfully evaluate the antiproliferative activity of a potential anticancer compound, as the first step, a cell viability assay is usually conducted. This assay can assess the ability of the compound to inhibit cell growth or promote cell death. When grown in culture, cells exhibit cell growth kinetics, consisting of four phases: lag phase, logarithmic (log) phase, stationary phase and decline phase

(Figure 3.3).[173] To ensure the reliability of the antiproliferative assay the cells must remain in a constant rate of growth for the full time span of the assay being carried out, known as log phase growth. To assess the correct number of cells that can be plated to ensure that they can remain in log phase growth for the time period of the assay, a seeding density assay is conducted.

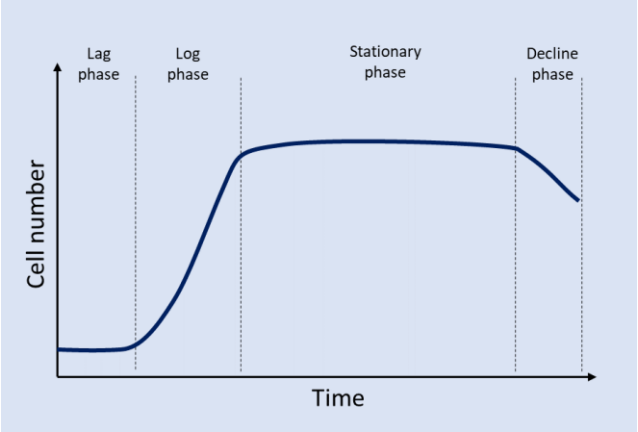


Figure 3.3: Cell growth kinetics in cell culture.[173]

3.1.3 Sulforhodamine cell viability assay

The Sulforhodamine B (SRB) cell viability assay was developed by Skehan *et al.* in 1990 for the purpose of screening drug candidates for potential cytotoxic and antiproliferative properties.[174] The assay is based on staining of proteins with SRB dye; this dye can bind to proteins in a pH dependent manner by electrostatically associating with basic amino acid residues. After treatment with the compound being evaluated, the cells are fixed to the base of the well in a 96-well plate using trichloroacetic acid (TCA). After fixation, the cells are treated with the dye dissolved in 1% acetic acid. The SRB dye has an anionic aminoxanthene structure with two sulfonic groups, and a bright pink colour, which binds to the cellular proteins. After washing away the excess, the dye can then be resolubilised with a weak base such as Tris and the optical density measured by recording absorbance at 490 nm using a plate reader. Results of the SRB assay are linear with respect to the cell number and cellular protein content, giving an accurate value of cell proliferation during the time of the assay.[175,176]

3.1.4 Improvement of aqueous solubility problems of flavonoids by encapsulation

Many flavonoids, like other polyphenols, have been widely investigated for their biological activities, however a key hurdle to the development of these compounds into approved drugs is their lack of aqueous solubility. Poor solubility in water has been widely reported for flavones, which can not only hinder *in vitro* studies and assessments, but can also create problems when considering the bioavailability and metabolism of the compounds *in vivo*.[177,178] Numerous strategies have been considered to improve the aqueous solubility of these compounds, such as increased levels of methylation and disruption of the planarity of the molecule by modification in the 3 position with a halogen.[178,179] When encountering problems with solubility of PMF in this work, a method of increasing the solubility of the compound without modification to the core flavone structure was sought. Complexation of flavones within a hydrophilic molecule such as cyclodextrin is a concept that has been investigated before.[180] Cyclodextrins (CDs) are cyclic oligosaccharides with hydrophobic cavities that increase in size depending on the number of glucose units in the ring. This cavity can be useful for shielding the hydrophobic molecule from the aqueous solution to enhance solubility of the compound and prevent precipitation.[181] This method was investigated in collaboration with Dr Emerald Taylor in the Serpell group, as a potential strategy to combat some of the solubility problems faced with PMF and its analogues.

3.1.5 Chapter aims

The aims of this chapter are to assess the biological activity of PMF in a CRC cell line and then to compare this to the activity of the previously synthesised modified polymethoxyflavone analogues to develop an SAR for the compound. From this, we aim to identify a suitable site on the molecule that can be synthetically modified without compromise of biological activity. This site modification can then be carried forward to allow the synthesis of a chemical probe which can be used for the isolation and identification of target proteins using a pull down study.

3.2 Materials and methods

3.2.1 Materials

Iscove Modified Dulbecco Medium (IMDM), Fetal bovine serum (FBS) and phosphate buffered saline (PBS) were purchased from Gibco. Serological pipettes, sterile 96-well plates and T75 cell culture flasks were purchased from Sarstedt. Trypan blue, Sulforhodamine B (SRB), trichloroacetic acid (TCA) and sterile DMSO were purchased from Sigma-Aldrich. Trypsin-EDTA was purchased from ThermoFisher Scientific. CRC cell lines HT29, SW620 and HCT116 were originally from ATCC.

3.2.2 Instrumentation

3.2.2.1 Plate reader

Absorbances measured for SRB assays were taken using 490 nm in a Perkin-Elmer Victor X4 plate reader.

3.2.2.2 Incubator

Cells were grown in a Thermo-scientific Forma Series II Water Jacket CO_2 incubator at 37 °C with 5% CO_2.

3.2.2.3 Microscope images

Microscope images were taken using GXCapture-T software visualising observations from an OLYMPUS CKX53 light microscope at 40x magnification setting.

3.2.3 Cell culture methods and equipment

Cell lines were defrosted from cryopreserved stocks and placed into IMDM supplemented with 10% FBS (*v/v*). Cells were grown as adherent cultures in T75 cell culture flasks with biweekly passaging of cells and addition of fresh media. Passaging of cells involved the removal of media, a wash with 1.5 mL PBS and then addition of 1.5 mL 0.25% white trypsin (both of which were warmed in a Grant water bath to 37 °C before use after storage at 4 °C). A small portion of the trypsinised cells were placed into a new T75 cell culture flask and 15 mL fresh supplemented media added. Cells were then placed in the

incubator. Cell cultures were grown at 37 °C in a 5% CO_2 environment, see section 3.2.2.2 . All cell handling was carried out in a sterile laminar-flow fume hood and equipment followed standard sterilisation procedure including autoclaving, 70% IMS spray and aseptic technique.[182]

3.2.3.1 Determination of cell number

After trypsinisation of the cells in a T75 flask and placement into fresh IMDM + 10% FBS during passaging, 50 µL of this suspension was removed and mixed with 50 µL trypan blue (0.4%) and cells were counted using a Brand™ Blaubrand™ Neubauer haemocytometer (0.0025 mm^2, 0.1 mm depth). Average cell count was converted to cells/mL values using the equation: $(Average\ cell\ count) \times 2 \times 10^4 = cells\ per\ mL.$

3.2.3.2 Seeding density assays

Determination of correct seeding densities for each cell line was carried out using a seeding density assay. This assay involved the plating of cells in a 96-well plate at varying numbers of cells/well and then fixation and staining of one plate each day over n + 1 days (where n = the number of days of an intended antiproliferative assay). The fixation and staining of plates followed the same protocols as those used for the SRB antiproliferative assay, detailed in 3.2.3.3.

Cells were removed from the flask, using the trypsinisation procedure followed for passaging, and placed into fresh media to create a suspension. Cells were counted using methods detailed in 3.2.3.1, and stocks were diluted to the required cell concentrations using IMDM + 10% FBS. Cells were plated in to sterile 96-well plates at the required concentrations in 200 µL of IMDM + 10% FBS per well. One row and column on the top, bottom and left of the plate were filled instead with sterile water to prevent evaporation, the unfilled columns on the right side were instead filled with only IMDM + 10% FBS with no cells to act as a media control. Growth curves were plotted using Excel and Graphpad Prism 9.

3.2.3.3 Sulforhodamine B (SRB) cell viability assay

Cells were plated at an equal concentration determined by the seeding density assay in sterile 96-well plates in 160 µL per well. One column and row around the plate was not filled with cells, and instead the top and bottom rows and left-hand column were filled with sterile water to prevent evaporation of media, and the left-hand column was filled with IMDM + 10% FBS to act as a media control. Untreated plates were placed in the incubator for 48 h to allow cells to adhere to the surface of the well plate. After 48 h, compounds were dissolved in IMDM + 10% FBS and DMSO vehicle and added to wells as 40 µL at 5 x the final required concentrations. Dilution of the 40 µL compound solution in the 160 µL media already in the plate resulted in the desired final concentrations and DMSO vehicle percentage. Plates were then incubated for the required number of days as determined by preliminary assays and then all plates were fixed. Fixation procedure involved the removal of media from the plates and addition of 70 µL 10% TCA solution to each well. After 30 minutes at room temperature, the TCA solution was removed with distilled water and 70 µL of 0.4% SRB dye in 1% acetic acid per well was added to stain cells, this was left for another 30 minutes at room temperature. Plates were then washed with 1% acetic acid solution and dried at 37 °C overnight. Plates were then treated with 100 µL 10 mM Tris buffer and shaken on a plate shaker at 200 rpm for 10 minutes. The absorbance of the solution in each well was then read at 490 nm in Victor X4 plate reader. Data obtained from the plate reader was converted to GI_{50} values using Excel and Graphpad Prism 9. Normalisation of absorbance data was achieved by subtracting the average absorbance media only control well measurements and an unconstrained sigmoidal curve was fitted using GraphPad Prism.

(https://www.graphpad.com/guides/prism/latest/curve-fitting/reg_dr_inhibit_normalized_variable.htm)

3.2.3.4 Formulation of PMF/cyclodextrin complexes

Formation of the PMF/cyclodextrin complexes was carried out by Dr Emerald Taylor as part of collaborative research within the group. Biological testing and analysis of results was carried out within this work.

Complexes of PMF and γ-cyclodextrin were prepared by dissolution of the γ-cyclodextrin (27.87 mg, 0.02 mmol, 4 eq) component in minimum quantities of distilled 18.2 MΩ ultrapure water (4.7 mL) and added to a solution of PMF (2.00 mg, 5.37 µmol, 1 eq) in acetone (3.15 mL) and the mixture stirred at rt for 5 days. The acetone was then removed by heating the solution to 35 °C for 2 h, resulting in the formation of precipitate. Additional distilled 18.2 MΩ ultrapure water was then added (5.3 mL) to give 10 mL of aqueous suspension in total. Biological testing of these complexes was carried out following methods discussed previously in section 3.2.3.3.

3.3 Results

3.3.1 Determination of required seeding densities of CRC cell lines.

Seeding density assay data was plotted as growth curves of the cells over time.

3.3.1.1 HT29 cell line

Growth curves were carried out in triplicate to ensure biological reproducibility. The data shown below is a representative example of each experiment (Figure 3.4). The seeding density assays for HT29 cells were initially carried out for cell concentrations between 400 cells/well and 12,800 cells/well. Complete data for these experiments can be found in Appendix 8.18.

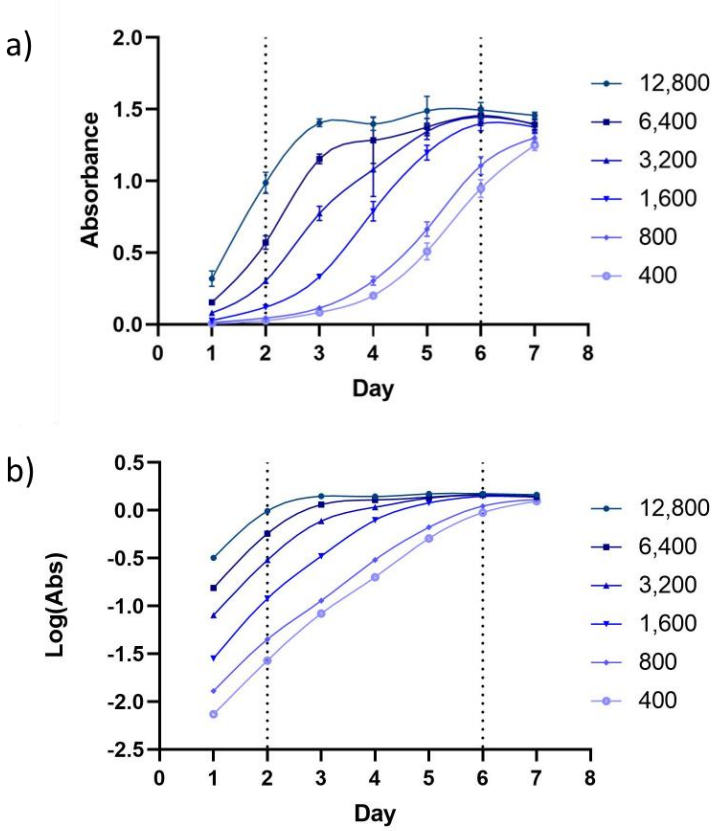


Figure 3.4: Representative example growth curves for HT29 cells a) Shows the absorbance measured at 490 nm, corresponding to the number of cells in each well, with each data point taken as an average of 6 data points from each plate. Standard deviation bars of the mean are also shown for each point. b) Log graphs of mean absorbance values for each cell concentration over 7 days. Dotted vertical lines represent the time interval of compound exposure that would be used for an SRB antiproliferative assay.

Following this, a further seeding density assay was carried out to ensure that cell growth remained in log phase for the extended length of time used for the HT29 cell line antiproliferative assay after modification of conditions, discussed in section 3.3.2.1 (Figure 3.5).

The assay was carried out using cell numbers from 100 cells/well to 3,200 cells/well and was completed as a single experiment as it confirmed earlier data.

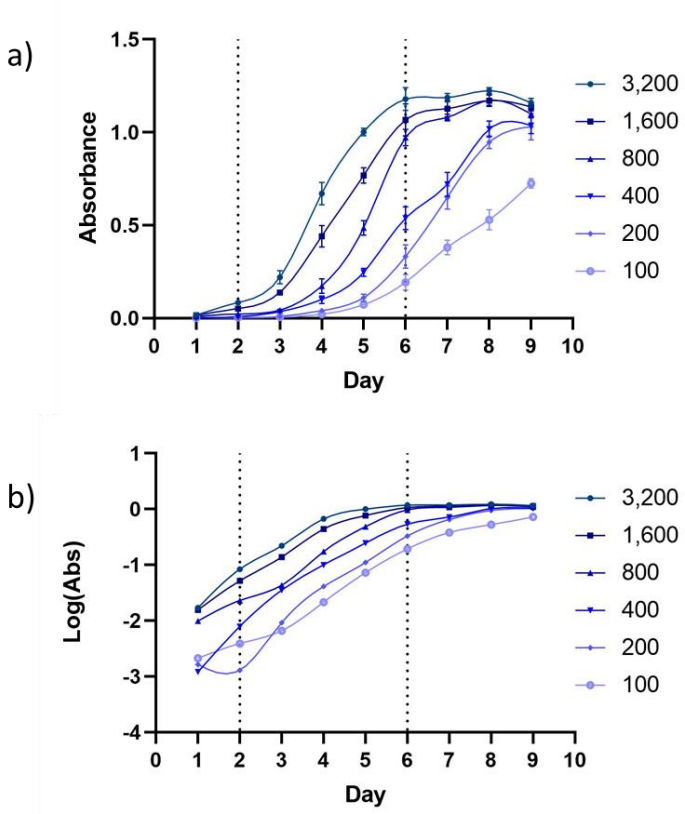


Figure 3.5: a) Shows the absorbance measured at 490 nm, corresponding to the number of cells in each well, with each data point taken as an average of 6 data points from each plate. Standard deviation bars of the mean are also shown for each point. b) Log graphs of mean absorbance values for each cell concentration over 9 days. Dotted vertical lines represent the time interval of compound exposure that would be used for an SRB antiproliferative assay.

Both seeding density assay experiments for HT29 cells showed that plating cells at 400 cells/well resulted in the cells maintaining log phase growth between days 2 and 6 or days 2 and 8 for the extended time assay.

3.3.1.2 HCT116 cell line

The seeding density assays for HCT116 cells were carried out for cell concentrations between 100 cells/well and 3,200 cells/well (Figure 3.6). Growth curves were carried out in triplicate to ensure biological reproducibility, data shown below is a representative example of each experiment, supplementary data can be found in Appendix 8.19.1-8.19.2.

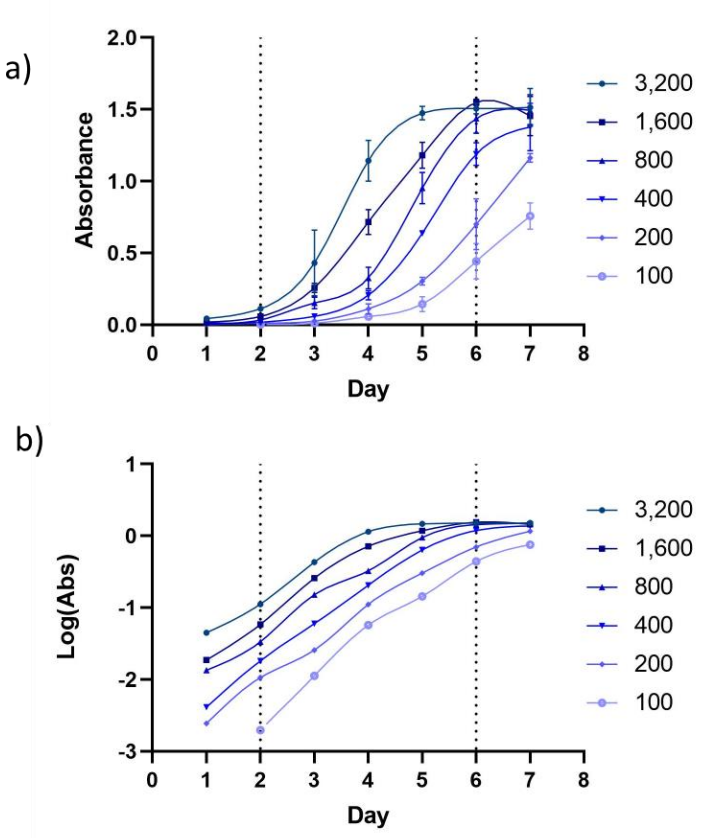


Figure 3.6: a) Shows the absorbance measured at 490 nm, corresponding to the number of cells in each well, with each data point taken as an average of 6 data points from each plate. Standard deviation bars of the mean are also shown for each point. b) Log graphs of mean absorbance values for each cell concentration over 7 days. Dotted vertical lines represent the time interval of compound exposure that would be used for an SRB antiproliferative assay.

Data points have been omitted for day 1 for 100 cells/well as the mean absorbance data was below 0 and returned an undefined number which could not be plotted.

Growth curves for HCT116 cells also showed that when plating cells in a 96-well plate that 400 cells/well results in log phase growth between day 2 and day 6 as necessary for an SRB antiproliferative assay.

3.3.2 Determination of antiproliferative activity of compounds using SRB assays.

3.3.2.1 Optimisation of conditions for SRB antiproliferative assays in HT29 cell line.

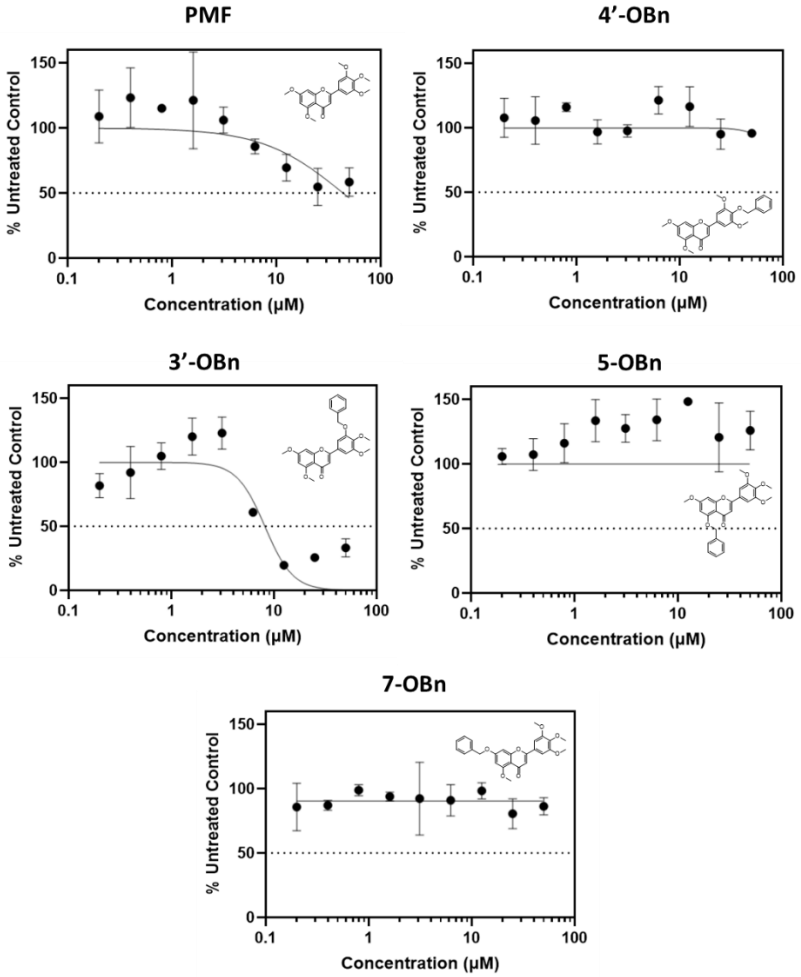


Figure 3.7: Preliminary antiproliferative data for PMF and benzyl flavone analogues tested on HT29 cells. Conditions: Concentration range: 0.2 µM – 50 µM; Vehicle: 0.25% DMSO in 1:1 dilution, Duration of compound treatment: 96 h. Data represents n = 1 biological repeats. Untreated control cells were exposed to media only.

Preliminary data obtained from an SRB assay (Figure 3.7), in the conditions stated above showed a reduction in cell viability when treating with parent compound PMF. At the highest concentrations, cell growth was found to be an average of 58% of the growth of the untreated control. The **4'-OBn**, **5-OBn** and **7-OBn** analogues showed no biological activity in this assay, with the **5-OBn** analogue showing potential growth promoting behaviour. Treatment with the **3'-OBn** analogue resulted in a

loss of cell viability, giving a GI_{50} value of 8 ± 5 µM for this single experiment, this suggested a higher potency for this compound than for PMF which gave a calculated GI_{50} value of 44 ± 68 µM. GI_{50} values for the other flavone analogues were unable to be calculated or lay outside of the concentration range used.

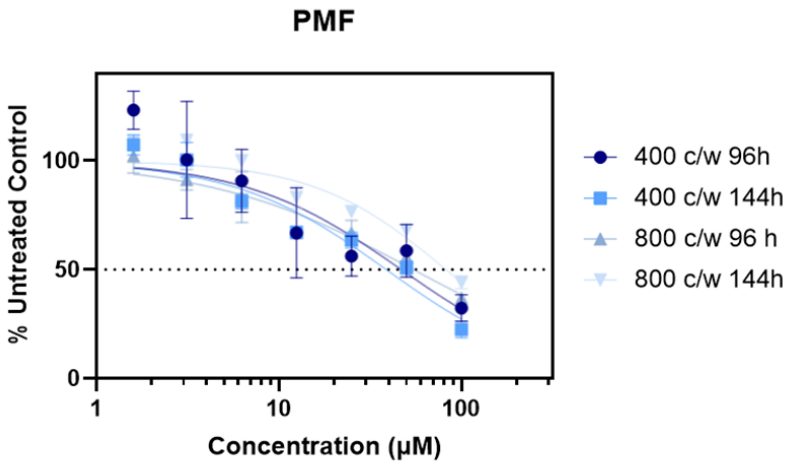


Figure 3.8: Matrix assay data showing antiproliferative data for PMF in varied assay conditions. Variable conditions: Cell number/well and compound exposure time. Consistent conditions: Concentration range: 1.6 µM – 100 µM; Vehicle: 0.5% DMSO in 1:1 dilution. Data represents n = 1 biological repeats. Untreated control cells were exposed to media only.

SRB assays were carried out under four distinct conditions, depicted in the table below (Table 3.1), to determine more optimum conditions to assess the antiproliferative activity of PMF. Cells were plated at either 400 or 800 cells/well, allowed to adhere for 48 h and were then exposed to PMF. After either 96 or 144 h of compound exposure, plates were fixed and stained to give absorbance data that could be shown as antiproliferative graphs. Assay number 2, 400 cells/well and 144 h compound exposure time, produced data which displayed the most potent activity of PMF, with a GI_{50} value of 39 ± 13 µM. These conditions were selected for further HT29 SRB assays to best display the activity of PMF. Due to compound stock concentrations, max concentrations of the compounds were administered in a 0.5% DMSO vehicle in the SRB assays.

Table 3.1: Table showing conditions used for SRB matrix assay.

Assay number	*Cell number (cells/well)*	*Compound exposure duration (h)*	*PMF GI_{50} (µM)*
1	400	96	47 ± 41
2	400	144	39 ± 13
3	800	96	55 ± 22
4	800	144	82 ± 28

During the process of the matrix assay previously discussed, it became apparent that at higher concentrations (> 12.5 µM), PMF was seen to precipitate out of the media solution as crystalline needles (Figure 3.9), this was understood to be due to a lack of aqueous solubility.

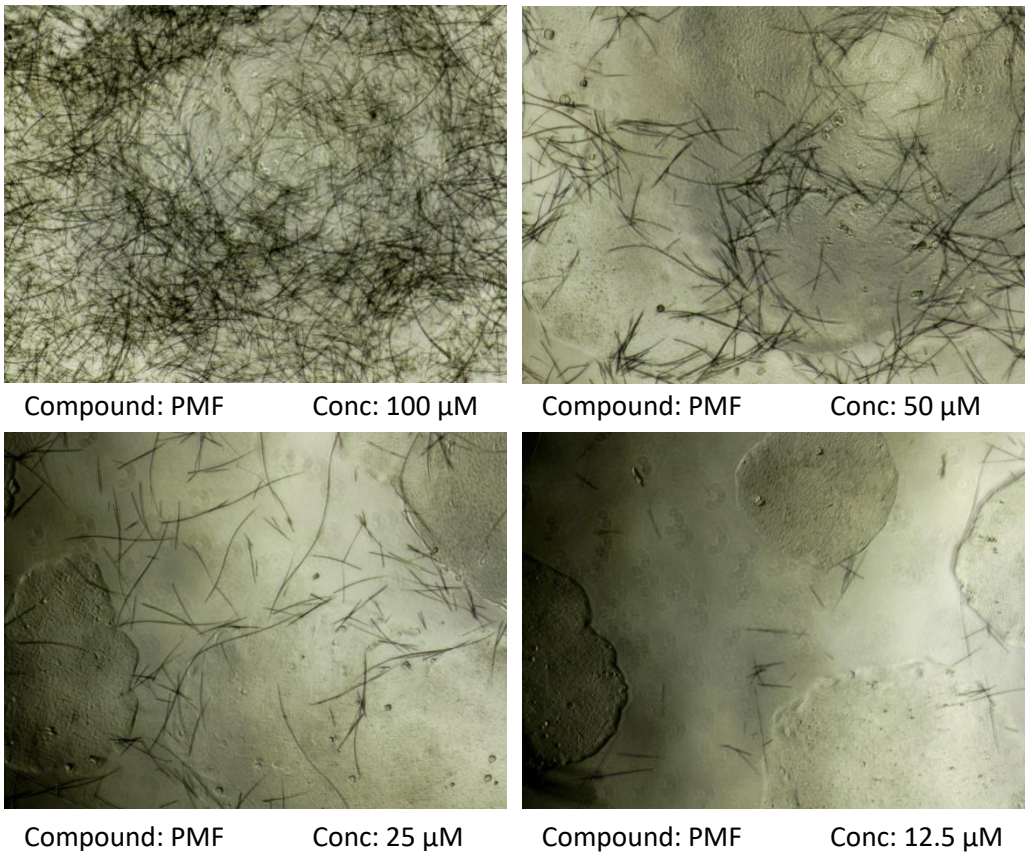


Figure 3.9: Images taken at 40x magnification of wells in a 96 well-plate 144 h after compound exposure. Images show visible precipitation of PMF, seen as needle-like structures, which increases as concentration of PMF increases. Images for concentrations < 12.5 µM have been omitted as no precipitation was observed. Scale bars not available.

It was suggested that this could be hindering the full effects of PMF from being seen in the assays as the compound was reaching saturation levels in the well, it is expected that only the dissolved PMF would be able to exert biological activity on the cells. Use of a higher percentage of DMSO as a vehicle could prevent this issue and increase the solubility of the hydrophobic compound. A small study was carried out, in which a 1% DMSO vehicle was used to deliver compounds to the cells at the maximum concentrations, followed by a 1:1 dilution for each concentration below. This assay resulted in a large loss of cell viability for the full range of concentrations tested. To supplement this test a 1% DMSO vehicle control test was also included, this was administered to the cells without the addition of the drug candidate to evaluate the effect of the vehicle on cell viability (Figure 3.10).

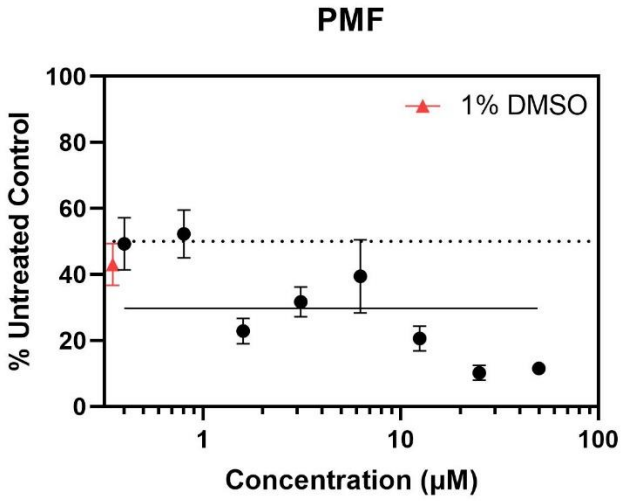


Figure 3.10: Antiproliferative data for PMF tested on HT29 cells using a 1% DMSO vehicle. The red marker indicates the cell viability for cells not treated with compound but only with the 1% DMSO control. Conditions: Concentration range: 0.2 µM – 50 µM; Vehicle: 1% DMSO in 1:1 dilution; Duration of compound treatment: 144 h. Data represents n = 1 biological repeat. Untreated control cells were exposed to media only.

It was proposed that the DMSO vehicle may be exhibiting toxicity to the cells and interfering with the results of the assay by compounding with the activity of PMF to produce artificially low cell viability numbers.

To investigate the effect a higher percentage of DMSO vehicle would have on cell viability, and to identify a useable DMSO vehicle percentage, a dose response curve for DMSO on HT29 cells was determined (Figure 3.11).

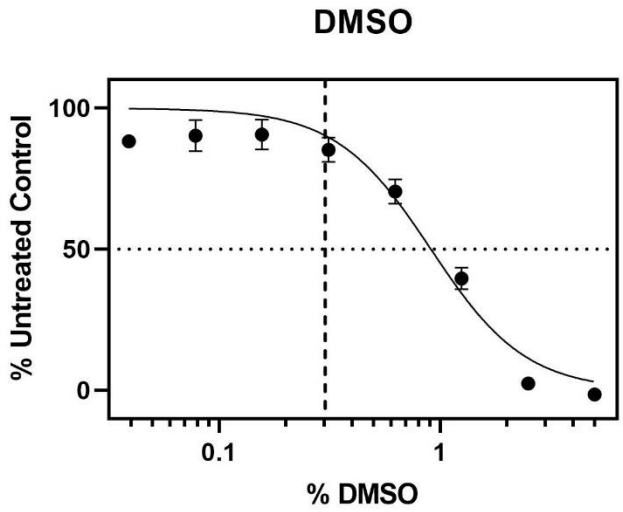


Figure 3.11: Dosage curve showing cell viability of HT29 cells after treatment with an increasing percentage of DMSO. Concentration range: 0.01% – 5%; compound exposure time: 144 h; cell density: 400 cells/well. The dashed vertical line indicates the location of 0.3% DMSO. This graph is representative of n > 3 biological repeats. Untreated control cells were exposed to media only.

The dose response curve shown for DMSO (Figure 3.11) indicated that the highest percentage DMSO that could be administered to HT29 cells without a loss in cell viability was 0.3%, as shown by the vertical dashed line. This was then taken forward for use as the maximum percentage of DMSO that could be used as a compound vehicle.

To maximise the solubility of PMF, and therefore the other flavone analogues, a 0.3% DMSO vehicle was used for delivery of all compound concentrations without subsequent 1:1 dilution as the concentration of flavone decreased, which is typical when conducting a cell viability assay. Instead, 0.3% DMSO was used as the vehicle for all concentrations of compound in the range used, to maximise

solubility and prevent precipitation. Stock solutions were also adjusted to allow a wider range of concentrations to be tested while maintaining the 0.3% upper limit.

3.3.2.2 Antiproliferative activity of PMF and modified flavone analogues on HT29 cells.

Conditions optimised using the earlier described methods were then carried forward to obtain final data for the antiproliferative activity of the flavones in the HT29 cell line (Figure 3.12).

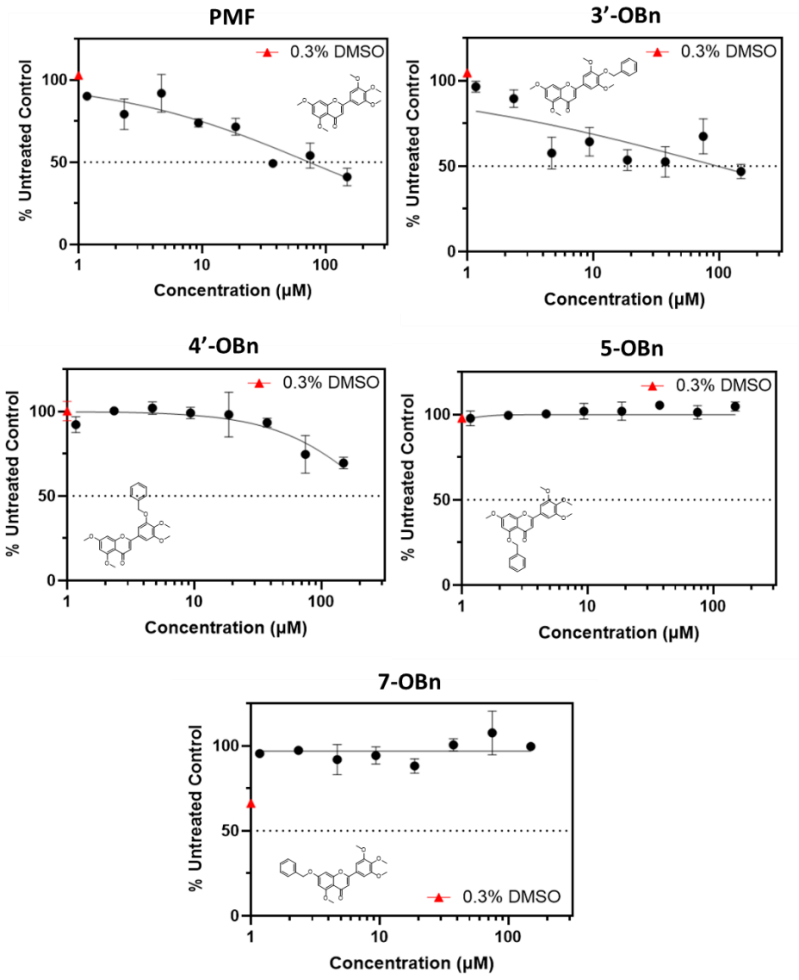


Figure 3.12: Antiproliferative activity of parent compound PMF and modified analogues in HT29 cells. Each data point represents the mean of at least 3 data points and error bars of 1 standard deviation above and below the mean are included. Concentration range: 1.17 μM – 150 μM; compound exposure time: 144 h; cell density: 400 cells/well. The red diamond indicates the mean cell viability of cells treated with only 0.3% DMSO vehicle control. Untreated control cells were exposed to media only.

Figure 3.12 shows the growth of cells as a percentage of the untreated controls when exposed to an increasing concentration of compound for 144 h. As the concentration of compound administered was increased for the parent compound, PMF, the cell viability was decreased, as manifested by the decreasing percentage of cells alive at the end point of the assay. The growth inhibitory power of the compound, or GI_{50} value, calculated from this data was found to be 74 ± 47 μM. Modification of the flavone in the 3'-position with the addition of the benzyl group, as seen in preliminary assays, was shown to give a similar pattern of data to that of the PMF experiment, suggesting that changes in this position were not sufficient to compromise the biological activity. The GI_{50} value calculated for the **3'-OBn** analogue for the assay shown was 97 ± 290 μM. Some growth inhibition was seen when treating with the **4'-OBn** analogue, however even at the highest concentrations of compound administered the cell growth did not fall below 50% of that of the untreated control cells, suggesting a compromise of biological activity. A GI_{50} value was calculated from extrapolated data and given as 279 μM, however as this value falls outside the concentration range used, it may not be taken as reliable.

Treatment with both the **5-OBn** and **7-OBn** analogues resulted in low levels of growth inhibition; despite an increase in concentration of compound administered, the growth of cells did not fluctuate greatly from that of the untreated control cells, suggesting that modification in these positions greatly reduced biological activity.

A problem encountered, however, when conducting this experiment for three biological repeats, was variability in results. Although PMF consistently caused a reduction in cell viability after treatment, the extent to which this was seen fluctuated to a large extent between individual experiments (Figure 3.13, Table 3.2).

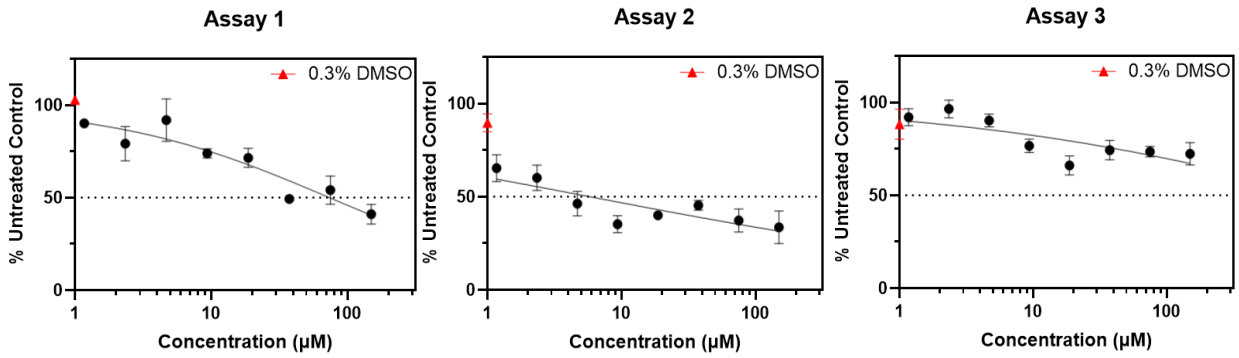


Figure 3.13: Three individual assay repeats showing treatment of HT29 cells with PMF at increasing concentrations. The graphs show variable data despite controlled conditions. Concentration range: 1.17 µM – 150 µM; compound exposure time: 144 h; cell density: 400 cells/well. The red diamond indicates the mean cell viability of cells treated with only 0.3% DMSO vehicle control. Untreated control cells were exposed to media only.

Table 3.2: GI_{50} values of PMF and the modified analogues for each biological repeat and the average value of the three experiments, when treating HT29 cells. *GI_{50} values for this experiment were unable to be calculated.

Assay number	*GI_{50} (µM)*				
	PMF	**3'-OBn**	**4'-OBn**	**5-OBn**	**7-OBn**
1	74 ± 47	97 ± 290	> 150	> 150	> 150
2	6 ± 6	*	> 150	> 150	> 150
3	> 150	> 150	*	*	*

It was supposed that the variability in the activity shown by PMF could be in part due to the precipitation of the compound in solution discussed earlier, which was still being observed despite the use of the 0.3% DMSO vehicle. Inconsistency was also seen for the treatment with the **3'-OBn** analogue, which was also exhibiting poor solubility. The data produced for treatment with the **4'-OBn**, **7-OBn**, and **5-OBn** analogues however was shown to be consistent, thought to be due to their relative lower activity levels, the full set of triplicate data is detailed in Appendix 8.20.

3.3.2.3 Determination of the sensitivity of CRC cell lines to PMF.

The anticancer activity of flavones for colorectal cancer has been reported for many cell lines, including those previously discussed, HT29 cells, as well as more uncommon cancer cell types such as APC10.1 mouse adenoma cells. [169,183] During the process of this investigation, a study by Fernández *et al.* was carried out to evaluate the effects of various flavonoids in colorectal cancer cell lines, HT29 (colorectal adenocarcinoma), HCT116 (colorectal carcinoma) and T84 (colon carcinoma metastasis). It found that the HCT116 cells were consistently more sensitive to the anticancer properties of the flavonoids, notably the two flavones studied, apigenin and luteolin.[184]

This development led to a preliminary investigation to evaluate the sensitivity of other cell lines to PMF and compare with data previously obtained for HT29 cells. HCT116 cells were given focus, with a less in-depth study with SW620 cells, as they were not evaluated in the previously discussed study.

It was supposed that a higher sensitivity to the compound's anticancer effects could give more reproducible data than that obtained for the HT29 cell line discussed in section 3.3.2.2, as the impact that poor aqueous solubility of PMF had on the HT29 data obtained would be minimised if a lower concentration of compound was required to produce the desired antiproliferative effects.

3.3.2.4 SW620 cell line PMF sensitivity

As an initial assessment for cell line sensitivity, cells were seeded in to 96-well plates at 400 or 800 cells/well and exposed to PMF in the concentration range 1.17 µM – 150 µM for either 144 or 96 h respectively, alongside a DMSO control curve in the range of 5% - 0.04%. After exposure, the cells were fixed and stained with SRB dye to assess remaining protein content and give a value of cell viability when compared to the untreated control, the results of these assays are shown in Figure 3.14.

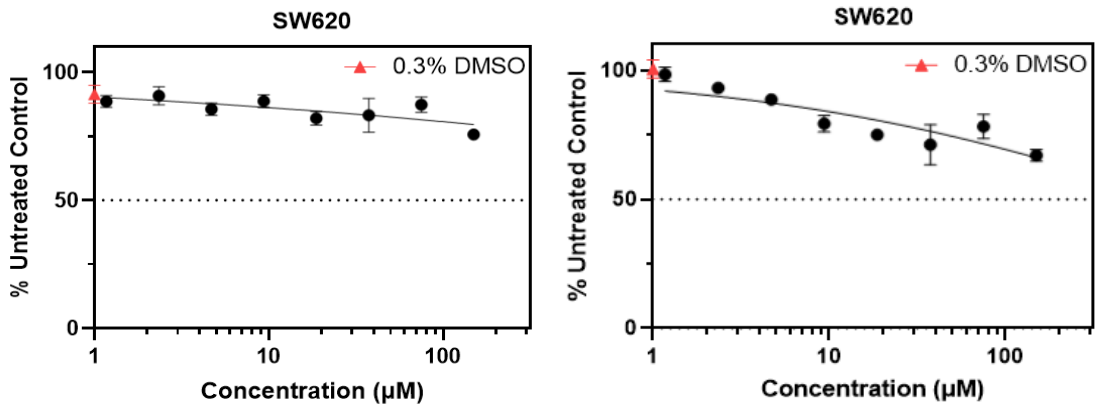


Figure 3.14: Antiproliferative data for SW620 cell line when treated with PMF. a) Concentration range: 1.17 µM – 150 µM; compound exposure time: 144 h; cell density: 400 cells/well. b) Concentration range: 150 µM – 1.17 µM; compound exposure time: 96 h; cell density: 800 cells/well. The red diamond indicates the mean cell viability of cells treated with only 0.3% DMSO vehicle control. Untreated control cells were exposed to media only.

A GI_{50} value for this data could not be accurately calculated as the number of viable cells did not fall below 50% of the untreated control, even at the highest concentration of 150 µM. This data gave early insight that this cell line was not more sensitive to the effects of PMF than the HT29 cells, so further studies were not carried out.

3.3.2.5 HCT116 cell line PMF sensitivity

To evaluate the sensitivity of the HCT116 cell line to treatment with PMF, assays were carried out in four conditions. Cells were seeded in 96-well plates at 400 and 800 cells/well and exposed to PMF in the concentration range 150 µM – 1.17 µM for both 144 and 96 h respectively, alongside a DMSO control curve in the range of 5% - 0.04%. The results obtained for this assay can be seen in Figure 3.15.

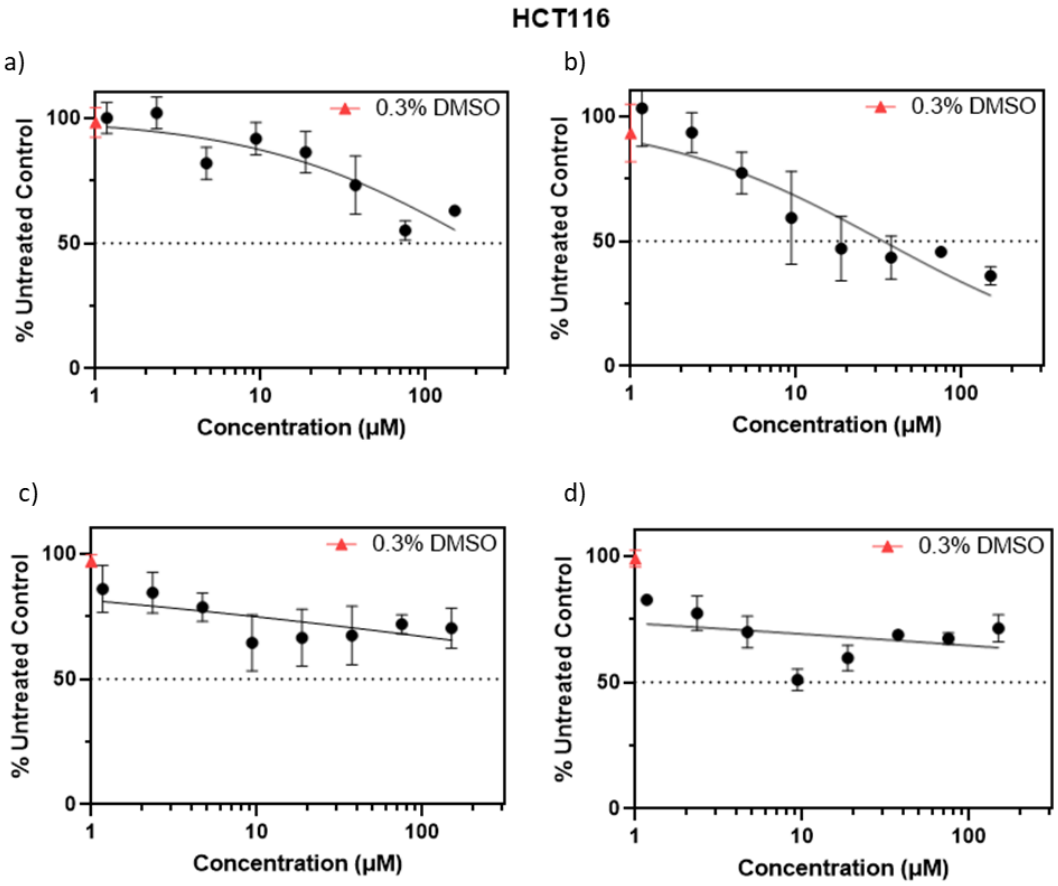


Figure 3.15: Antiproliferative data for HCT116 cell line when treated with PMF. a) Concentration range: 150 µM – 1.17 µM; compound exposure time: 96 h; cell density: 800 cells/well. b) Concentration range: 150 µM – 1.17 µM; compound exposure time: 96 h; cell density: 400 cells/well. c) Concentration range: 150 µM – 1.17 µM; compound exposure time: 144 h; cell density: 800 cells/well. d) Concentration range: 150 µM – 1.17 µM; compound exposure time: 144 h; cell density: 400 cells/well. The red diamond indicates the mean cell viability of cells treated with only 0.3% DMSO vehicle control. Untreated control cells were exposed to media only.

The data presented (Figure 3.15) shows some activity of PMF in the HCT116 cell line for all conditions tested, however, condition b, showed the greatest loss in cell viability when compared to the untreated control. In comparison with the biological activity of PMF in HT29 cells, which gave an average calculated GI_{50} value of 40 µM, excluding the much higher calculated GI_{50} value for assay 3 (> 500 µM). The GI_{50} value calculated for the treatment of HCT116 cells in this condition was found to be

34 µM. Due to the potential increased sensitivity of the new cell line, a more in depth study was carried out under these conditions, to obtain a triplicate set of data for PMF and assess the activity of the modified analogues in HCT116 cells.

3.3.3 Biological activity of PMF and modified flavone analogues on HCT116 cells.

Biological activity of PMF and the four modified flavone analogues was evaluated in HCT116 cells using a treatment concentration range of 1.17 µM – 150 µM with a compound exposure time of 96 h. Triplicate data representing three biological repeats for this experiment can be found in Appendix 8.21. The data presented below (Figure 3.16), shows the biological activity for PMF in HCT116 cells, the mean GI_{50} calculated from the three biological repeats of this experiment was found to be 25 µM, shown in Table 3.3. Higher activity levels were seen for all four modified analogues when treating HCT116 cells than for HT29 cells, in keeping with the results observed for treatment with PMF, discussed in section 3.3.2.5. An average GI_{50} value could also be calculated for the **3'-OBn** flavone analogue and was found to be 7 µM.

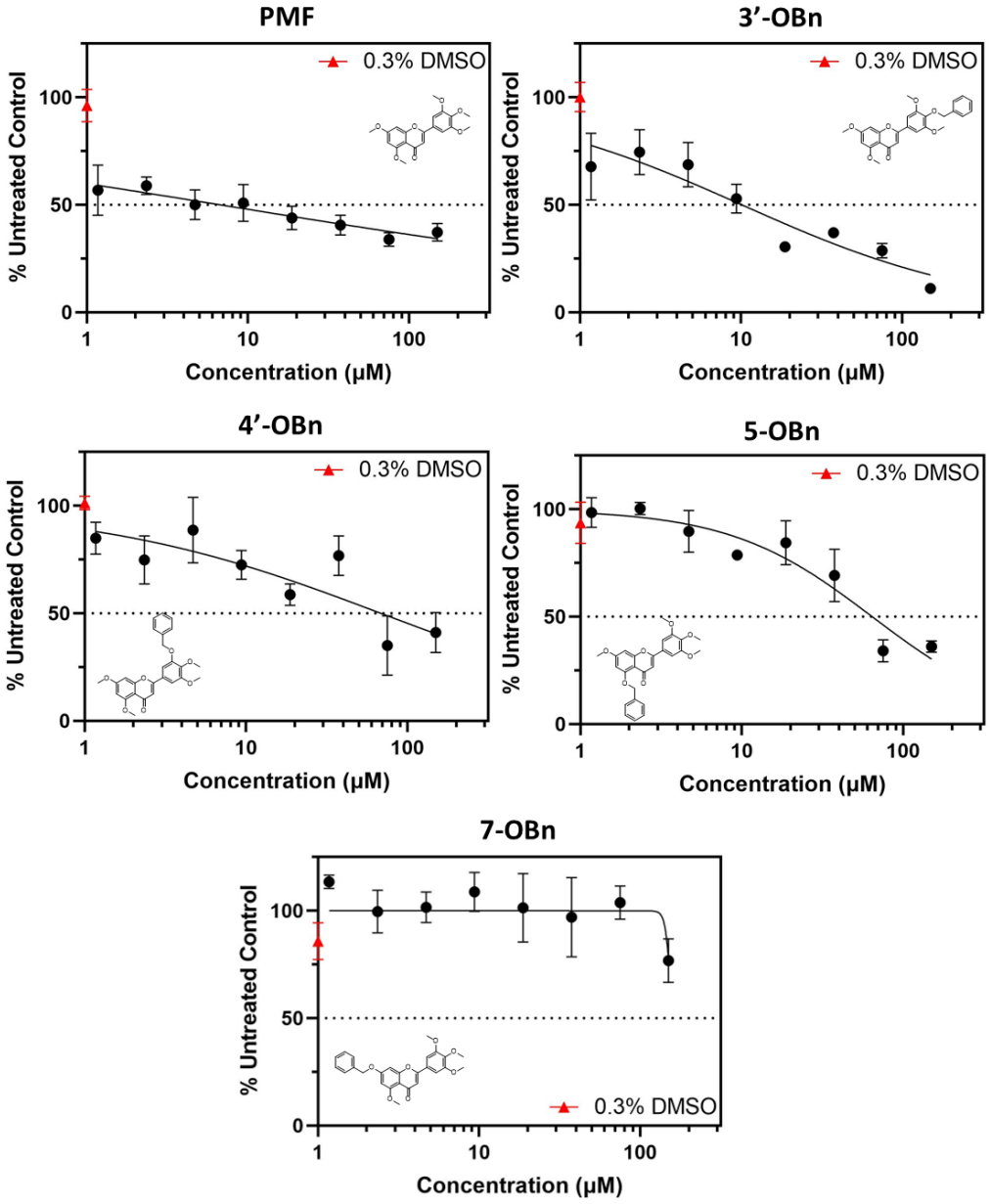


Figure 3.16: Antiproliferative activity of parent compound PMF and modified analogues in HCT116 cells. Each data point represents the mean of at least 3 data points and error bars of 1 standard deviation above and below the mean are included. Concentration range: 1.17 μM – 150 μM; compound exposure time: 96 h; cell density: 400 cells/well. The red diamond indicates the mean cell viability of cells treated with only 0.3% DMSO vehicle control. Data presented here represents n = 3 biological repeats. Untreated control cells were exposed to media only.

Treatment with the other three analogues resulted in at least one biological repeat in which the calculated GI_{50} value fell outside of the treatment range (quoted as > 150 μM in Table 3.3) and

therefore was excluded from the calculation of the mean value. As all three GI_{50} values calculated for treatment with the **7-OBn** analogue were > 150 µM a mean could not be effectively determined, this is due to the observed inactivity of the modified flavone in the cells, shown in the graphs as a straight line instead of the expected 'S' curve observed when treating with a biologically active compound.

Table 3.3: GI_{50} values of PMF and the modified analogues for each biological repeat and the average value of the three experiments when treating HCT116 cells. Some values have been omitted from the calculation of the mean as the GI_{50} value for that experiment fell outside of the treatment concentration range and is not reliable. *Mean has been calculated from n < 3 data sets.

Assay number	*GI_{50} (µM)*				
	PMF	**3'-OBn**	**4'-OBn**	**5-OBn**	**7-OBn**
1	10 ± 8	10 ± 5	69 ± 91	64 ± 25	> 150
2	59 ± 33	3 ± 2	> 150	> 150	> 150
3	7 ± 6	7 ± 2	> 150	134 ± 102	> 150
Mean	25	7	69*	99*	N/A

3.3.4 Treatment of HCT116 cells with PMF/γ-cyclodextrin complexes to assess improvement of compound solubility and increase in anti-cancer activity.

Biological testing of PMF/*γ*-cyclodextrin formulation, prepared in a collaborative project within the research group, was carried out to investigate the potential improvement of aqueous solubility of PMF when complexed with *γ*-cyclodextrin. The formulation was prepared following the method detailed in section 3.2.3.4.

An initial control assay was carried out to evaluate the effect of *γ*-cyclodextrin alone on HCT116 cells, (Figure 3.17). This experiment confirmed that *γ*-cyclodextrin was not toxic to cells at the concentrations used and could be administered to cells at much higher concentrations without compromise of cell viability.

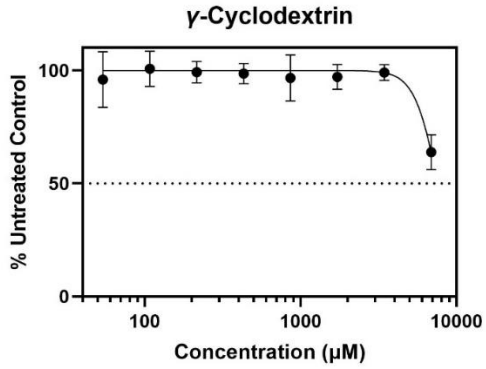


Figure 3.17: Assay data from the treatment of HCT116 cells with *γ*-cyclodextrin to assess loss of cell viability. Concentration range: 53.75 µM – 6.88 mM; Vehicle: Water; compound exposure time: 96 h; cell density: 400 cells/well. Data represents n = 1 biological repeats. Untreated control cells were exposed to media only.

The cells were then treated with the PMF/*γ*-cyclodextrin formulation to assess an improvement in antiproliferative activity in comparison to treatment with pure PMF. Cells were treated with the formulation using the maximum concentration obtained from preparation, which was then diluted 5-fold when placed in the IMDM media already administered to cells in a 96-well plate after 48 h adherence time, resulting in a concentration range of 107.43 µM – 0.84 µM (concentrations quoted for these assays refer to the estimated concentration of PMF in the solution). It is anticipated that concentration of PMF in solution may differ from the estimated concentration as the formulations had residual precipitate. Images of the solution immediately after addition of the compounds were taken to assess relative solubility/crystal formation of PMF and the PMF/*γ*-cyclodextrin formulation (Figure 3.18). The images show that when compared to the precipitation of PMF at similar concentrations, the visible crystal formation of the formulation was lower.

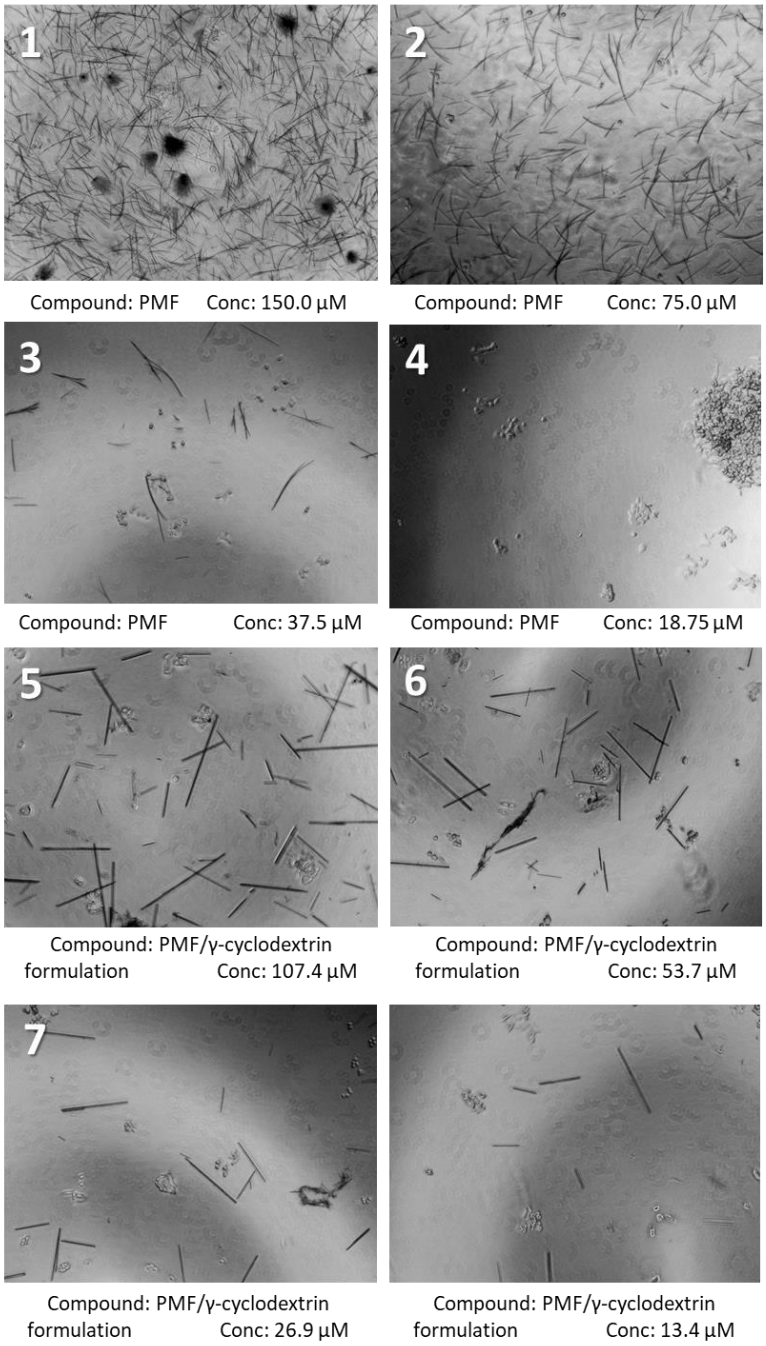


Figure 3.18: Images taken using a microscope at 40x magnification, showing precipitation formed after administering compound to cells in media. Images 1-4 taken after treatment with PMF at concentrations 150 µM, 75 µM, 37.5 µM and 18.75 µM. Images 5-8 taken after treatment with PMF/γ-cyclodextrin formulation at concentrations 107.4 µM 53.7 µM, 26.9 µM and 13.4 µM. Scale bars not available.

The results from the antiproliferative assay experiments can be seen in the combined graphs (Figure 3.19). Each experiment was completed in triplicate.

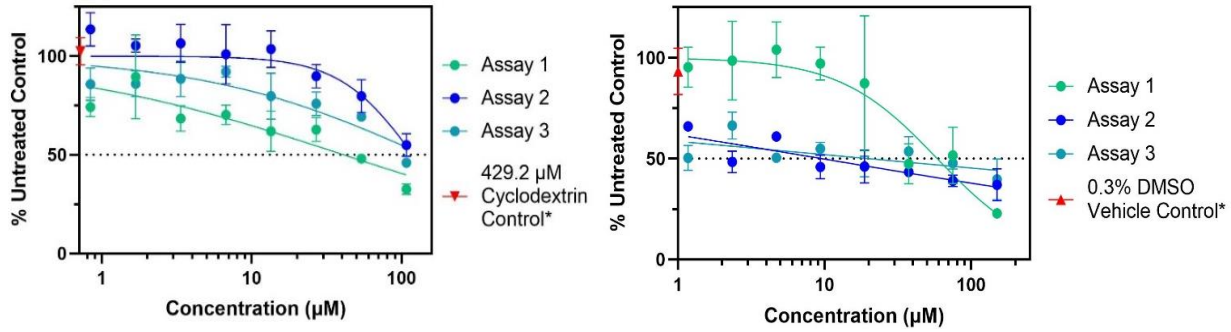


Figure 3.19: a) Combined graph showing triplicate data obtained from treatment of HCT116 cells with a PMF/*γ*-cyclodextrin formulation. Concentration range: 0.84 µM – 107.43 µM; Vehicle: Water; compound exposure time: 96 h; cell density: 400 cells/well. The red diamond indicates the mean cell viability of cells treated with 429.2 µM *γ*-cyclodextrin, calculated as an average from two assays. b) Combined graph showing triplicate data obtained from treatment of HCT116 cells with PMF only for comparison. Concentration range: 150 µM – 1.2 µM; Vehicle: 0.3% DMSO; compound exposure time: 96 h; cell density: 400 cells/well. The red diamond indicates the mean cell viability of cells treated with only 0.3% DMSO vehicle control, this is shown from assay 1 and is representative of all data. Untreated control cells were exposed to media only.

The GI_{50} values calculated for the treatment with PMF in these assays were found to be 59 ± 33 µM, 10 ± 8 µM and 20 ± 97 µM for assays 1, 2 and 3 respectively. When compared to the treatment of the cells with the PMF/*γ*-cyclodextrin formulation, the activity of PMF alone was found to be higher. A GI_{50} value for treatment with the PMF/*γ*-cyclodextrin formulation could only be calculated effectively for assay 1 and was found to be 41 ± 39 µM, the values calculated for assays 2 and 3 were found to be > 107.43 µM which was the maximum concentration used and therefore the data is not reliable.

3.4 Discussion

3.4.1 Discussion of sensitivity of various colorectal cancer cell lines to treatment with PMF and modified flavone analogues.

The HT29 colorectal cancer cell line was initially treated with PMF and the modified analogues to evaluate their biological activity, however the results obtained did not closely match those that have been reported previously for treatment of CRC cells with PMF and other flavones, and higher GI_{50} values were found for PMF than anticipated. Variation in GI_{50} values reported for specific compounds by different researchers can be expected, due to differences in the assays used and the conditions in which the assays are carried out, however the antiproliferative activity of PMF seen for the assays conducted in this work was much lower than that reported and therefore alternative cell lines were investigated.

It was suspected, based on data published by Fernández *et al.,* that other cell lines taken from different colorectal cancers could be more sensitive to the treatment with the flavones and result in lower GI_{50} values than found when treating HT29 cells.[184] Both SW620 and HCT116 cell lines were then treated with PMF in preliminary tests to assess their sensitivity to the flavone. SW620 cells were found not to be sensitive to treatment with PMF and gave GI_{50} values of > 150 µM. Despite structurally similar flavones 5,6-dihydroxy-7,3',4'-trimethoxyflavone and 5,6,4'-trihydroxy-7,3'-dimethoxyflavone showing low GI_{50} values of 9.5 µM and 5.0 µM respectively when treating this cell line *in vitro*, suggesting a link between the multiple hydroxyl groups in this compounds and their activity in SW620 cells.[185] Preliminary tests with the HCT116 cell line showed lower GI_{50} values than found for treatment of the HT29 cells, suggesting that it was more sensitive to the exposure to PMF than the HT29 cells. Many flavones and other flavonoids have been shown to exert anticancer activity on HCT116 cells, so these results are in agreement with wider scientific findings.

Zuo *et al.* showed that luteolin, a tetrahydroxyflavone, supressed cell proliferation of HCT116 cells in a dose-dependent manner.[186] Zhang *et al.* also evaluated the effect of two polymethoxyflavones,

nobiletin and 5-demethylnobiletin, on both HT29 and HCT116 cells, and concluded that HCT116 cells were more sensitive to treatment with both flavones than the HT29 cells by evaluating a loss in signal for lipid, protein and nucleic acid biomolecules in the cells seen by Raman microscopy.[187]

3.4.2 Analysis of structure-activity relationship of PMF and modified flavone analogues when treating HCT116 CRC cells.

The flavone analogues synthesised in Chapter 2 were developed from single point modifications to the methoxy groups around the rings, at positions 5, 7, 3' and 4'. Comparison of the biological activity of these modified analogues allows for the potential determination of the active part of the molecule, the pharmacophore, as modification in the region will likely result in the loss of biological activity being assessed. The results presented in section 3.3.3 show the antiproliferative activity of PMF when used to treat HCT116 CRC cells, as discussed previously, this cell line was more sensitive to the treatment with PMF than HT29 and SW620 cells, resulting in an average GI_{50} value of 25.02 µM. The analogues that have retained desired biological activity, and therefore have not been modified in the active region should retain a GI_{50} value similar, or less than that of PMF. The results of treatment with the modified analogues showed that biological activity was retained to some degree for the analogues modified in the 3', 4' and 5-positions of modification, with a total loss of activity seen for the modification in the 7-position. The **3'-OBn** analogue, however, resulted in the best retention of biological activity, and even showed a lower GI_{50} value than that recorded for the parent compound PMF. This result, alongside the lower GI_{50} value of 68.54 µM calculated for treatment with the **4'-OBn** analogue, suggests that modification to the methoxy groups in the B ring of the flavone structure results in lower impact to biological activity compared to modifications to the A ring (Figure 3.20). This suggests that the A ring without modification could be important to the activity of the compound in the cell, potentially *via* binding to the target protein through specific bonds or producing a molecular shape that allows the compound to fit in to the active site of the target protein.

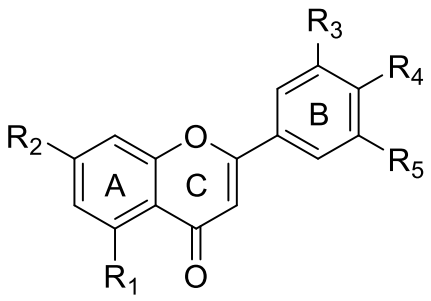


Figure 3.20: Flavone structure showing A, B and C rings with modifiable groups included as R groups 1-5.

From this data, it can be concluded that modification to the 3'-position of the molecule resulted in the least impact to biological activity and therefore offers a potential site for further synthetic modifications. The aim of this study was to identify this potential modification point to allow the synthetic attachment of a chemical probe to PMF, this can then be used to carry out a pull down study to identify potential protein binding partners of the compound in cells. Changes to the 3'-methoxy group of the molecule with a chemical probe should ensure complete retention of protein binding interactions and allow the true target proteins of the compound to be identified with the pull down study.

3.4.3 Discussion of change in activity of PMF when encapsulated in a *γ*-cyclodextrin formulation.

When studying the biological activity of PMF in CRC cells using the SRB assay method, it became clear that the compound had low levels of aqueous solubility, resulting in the precipitation occuring in the form of crystals when conducting the assay. The precipitation interfered with the effectiveness of the assay as the concentration of the compound in solution was limited, and the concentrations administered to the cells when dissolved in DMSO were not maintained after dilution in media and crystal formation occurred. An attempt to rectify this issue was tested by trying to encapsulate the hydrophobic flavone molecule in the carbohydrate structure of *γ*-cyclodextrin to improve its aqueous solubility properties. Images taken using a microscope showed a lower amount of visible crystal

formation when compared to treatment with PMF alone, however the results from the antiproliferative assays did not result in lower GI_{50} values for the formulation. Further study is required to investigate the potential benefits of complexation of PMF with *γ*-cyclodextrin, or other potential encapsulation methods, like that shown for the encapsulation of apigenin in liposomes.[188] Nanoization strategies could also be an alternative effective method for improving the aqueous properties of flavones, Zhao *et al.* showed a method of nanoemulsion formation of Baicalin, a biologically active flavone glucuronide, that improved the oral availability of the compound in rats.[189,190]

3.5 Conclusion

This chapter aimed to evaluate the anticancer activity of PMF in various CRC cell lines using an antiproliferative SRB assay technique, and then to evaluate the activity of the previously synthesised flavone analogues that had been modified with a benzyl group at four modification points around the structure. The activity of these modified analogues to inhibit cell growth could then be compared with that of PMF to develop a structure activity relationship between the methoxy groups on the structure and the behaviour of the compounds *in vitro*. The results showed that the HCT116 cell line was the most sensitive to treatment with PMF and resulted in the highest levels of growth inhibition compared to the treatment of the HT29 and SW620 cell lines. The results of testing the antiproliferative activity of the modified analogues showed that modification of the structure at the 7-position resulted in loss of desired biological activity in HCT116 cells, making it an unsuitable site for modification of the parent structure with a chemical probe. Some biological activity was maintained when treating with the three other analogues tested, **3'-OBn**, **4'-OBn** and **5-OBn**, though the highest levels of biological activity seen were found for treatment with the **3'-OBn** analogue, as this gave lower average GI_{50} values than the other analogues. From this data, it was determined that modification to the B ring of the flavone backbone resulted in the least compromise of biological activity, and therefore could be used as a suitable modification point on the structure for attachment of a chemical probe moiety needed to carry out a pull down study.

A brief investigation was also carried out to assess the benefits of encapsulating PMF in a *γ*-cyclodextrin carbohydrate structure to try to improve aqueous solubility properties. Microscope images taken from treatment of cells with the developed PMF/*γ*-cyclodextrin formulation showed decreased levels of precipitation when compared to PMF alone, however the results of SRB antiproliferative assays did not show an improvement of activity for the formulation. Further work is needed to assess the benefits of this encapsulation.

It is clear from the investigation carried out that the solubility properties of PMF and the modified flavone analogues is undesirable for the development of a potential drug molecule. Throughout the process of drug discovery and development, particularly when working with natural compounds, it is often the case that the exact structure of the molecule that initially shows promising biological activity is not suitable for use as a drug molecule, due to its properties. The compound often requires a series of modifications and improvements to increase the drug-like properties of the molecule and make it a more suitable candidate for use in the clinic. Flavones, and PMF specifically, are a clear example of this process and the aqueous solubility properties of the compound can be improved upon at later stages of development.

4 Synthesis of probe compounds and biological activity validation

4.1 Introduction

4.1.1 Photoaffinity probe design

A pull down study using photoaffinity labelling (PAL) requires the development of a photoaffinity probe, as discussed in Chapter 1. This probe molecule should consist of three key components (Figure 4.1): the photo-crosslinking group, the targeting group and the reporter tag. For this study, the probe design incorporated a trifluoromethylphenyl diazirine (TPD) crosslinking agent, attached to the flavone targeting group by an alkyl linker. This type of crosslinking group has been shown to be less likely to result in non-specific labelling than other crosslinking species[191]. The third component of the probe molecule is the reporter tag - incorporation of the reporter tag into the structure of this probe is facilitated by the attachment of an alkyne-PEG chain bonded to the phenyl ring of the TPD group. Incorporation of the alkyne group attached to a PEG-chain as a reporter handle offers many benefits to the probe structure including: (1) enabling the probe to be attached to various reporter tag species using click chemistry techniques, such as biotin or a fluorophore, to allow for different visualisation/identification techniques to be used; (2) the clickable handle results in a smaller sized probe molecule than incorporating a larger reporter tag, such as those mentioned, into the probe from the offset, this can prevent potential disruption of binding interactions with proteins if the reporter tag is too bulky.[129] The small size of the alkyne handle also improves cell penetration if live cell labelling experiments are being conducted.[192]

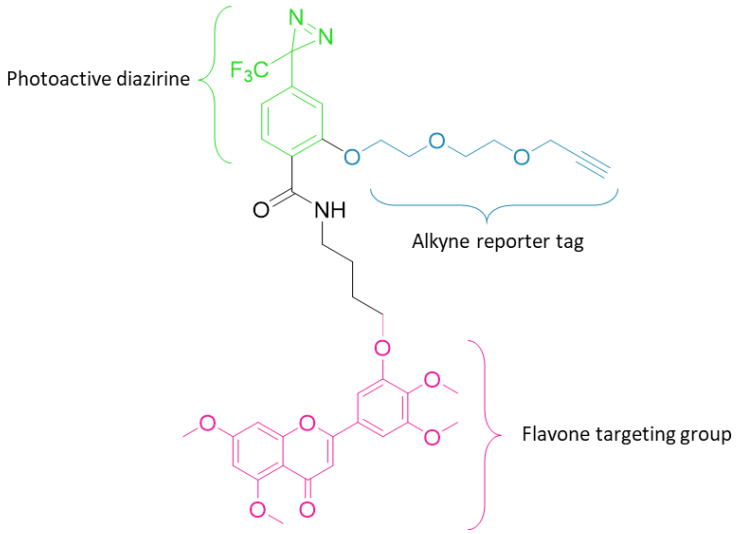


Figure 4.1: Design of photoaffinity probe for use in pull down study of interacting binding proteins of pentamethoxyflavone (PMF).

4.1.2 Synthesis of a trifluoromethylphenyl diazirine photoactive group

The photo-crosslinking agent selected for this study was a trifluoromethylphenyl diazirine (TPD), which is one of the most commonly used diazirines, the synthesis of this agent has been widely studied and typically follows the same 4 key steps for the production of the photoactive diazirine group (Figure 4.2).[193,194] The synthetic route starts from a phenyl ketone, which for the TPD species here is 2,2,2-trifluoroacetophenone. This is then converted to an oxime intermediate by condensation with hydroxylamine. Following the formation of the oxime, a sulfonyl ketoxime intermediate is synthesised, after tosylation or mesylation.[195] The formation of the three-membered diaziridine ring is typically achieved by reacting the activated oxime with liquid ammonia at low temperatures.[196] Oxidation of the diaziridine intermediate to the reactive diazirine can then be carried out with a range of oxidising reagents such as iodine in the presence of triethylamine, chromium trioxide in sulfuric acid, or silver (I) oxide.[196,197]

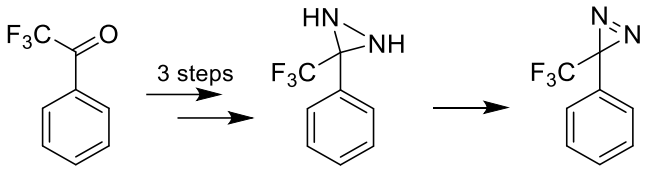


Figure 4.2: Key intermediates for the synthesis of a TPD photoreactive species.

Alternative methods to achieve the synthesis of the desired TPD compound have been suggested. Wang *et al.* developed a method that enabled synthesis of the final diazirine product from the tosyl intermediate, without isolation of the diaziridine compound.[198] The routes proposed however, require the use of liquid ammonia, which is not a desirable reagent due to its safety hazards as a volatile, toxic, and corrosive substance, as well as the expense and difficulty of maintaining a supply.[199]

An ammonia-free synthetic method was proposed by Kumar *et al.* by synthesis of a ketimine intermediate, followed by a reaction with O-mesitylenesulfonyl hydroxylamine to form the diaziridine and could be a promising alternative to the standard synthetic route.[199]

4.1.3 UV irradiation induced reactivity of photolabeling diazirines

Irradiation of the diazirine group with the required wavelength of UV light causes the excitation of the group; for phenyl diazirines this is 350 – 380 nm. The strained three-membered ring containing bonded nitrogens is converted to a singlet carbene by the extrusion of a molecule of N_2, and a rearrangement of electrons on the carbon results in the formation of the carbene (Figure 4.3). Both the phenyl and electron-withdrawing trifluoromethyl substituents on a TPD crosslinking group give stability to prevent the formation of an alternative diazoisomer, reducing non-specific labelling and promoting reactivity from only the singlet state.[99,120] The stability of the carbene is dependent on the solvent used but can range from < 1 ns to microseconds, with shorter lifetimes occuring in polar solvents such as water.[200]

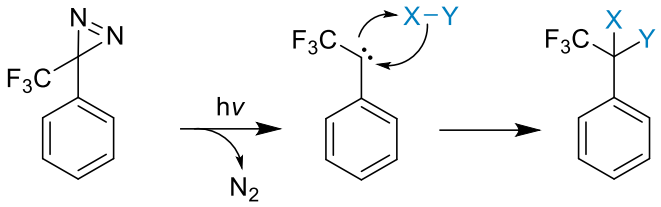


Figure 4.3: Scheme showing formation of a singlet carbene after irradiation of a TPD with UV light, then insertion into an X-Y bond (typically where X = hydrogen).[99]

4.1.4 Chapter aims

This chapter aims to complete the synthesis of the active and inactive probe molecules required to complete the pull down study, including the TPD photolabeling group, the alkyne reporter handle and the flavone targeting group for the active probe, modified in the 3'-position. This includes the synthesis of a trifluoromethylphenyl diazirine group. The chapter also aims to validate the anti-proliferative activity of both the intermediate flavone group after modification with a Boc-protected butyl amine linker group, and the full active probe molecule. The timescale of decomposition of the photoactive diazirine after exposure of UV light will also be discussed.

4.2 Materials and methods

4.2.1 Materials

N-(4-Bromobutyl)phthalimide, *tert*-butyl (4-bromobutyl)carbamate, n-butyl lithium, ammonium chloride, sulfamic acid, sodium chlorite and 3-(2-(2-bromoethoxy)ethoxy)prop-1-yne were purchased from Sigma-Aldrich. 3-Hydroxy-4,5-dimethoxybenzaldehyde was purchased from Biosynth International, Inc. Potassium carbonate was purchased from Thermo Fisher Scientific. DMF, ethyl acetate, hexane, methanol, diethyl ether, magnesium sulphate, sodium hydrogen carbonate and hydrochloride acid were purchased from Fisher Scientific. Deuterated solvents were purchased from Cambridge Isotope Laboratories Inc. HCl (4N in dioxane), 3-bromoanisole, methyl trifluoroacetate, hydroxylamine hydrochloride, DMAP, triethylamine, dichloromethyl methyl ether, EDC.HCl and silver oxide were purchased from Fluorochem. Sodium hydroxide, dry pyridine, dry DCM, dry THF, dry

ethanol, 7N NH_3 in methanol, titanium tetrachloride, tetrabutylammonium iodide and n-butyl amine were purchased from Acros Organics. Tosyl chloride and boron tribromide (1M in CH_2Cl_2) were purchased from Alfa Aesar.

4.2.2 Instrumentation

4.2.2.1 Synthetic instrumentation

Instrumentation and methods required for purification and characterisation of synthesised compounds including NMR, Biotage, LCMS and HRMS, have been discussed previously and can be found in section 2.2.

4.2.2.2 UV-Vis

UV-Vis spectra were recorded using a NanoDrop™ One/OneC Microvolume UV-Vis Spectrophotometer. Data was processed using Excel and Graphpad Prism 9.

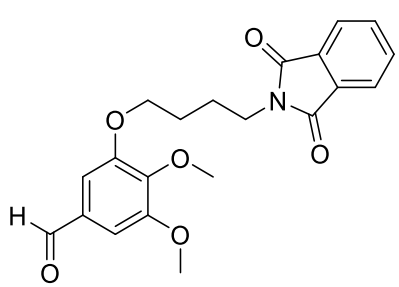
4.2.2.3 UV lamp

UV irradiation was carried out using an Alonefire SV18 12W 365nm UV Torch.

4.2.3 Experimental methods

4.2.3.1 Synthesis of 2-(3-(4-aminobutoxy)-4,5-dimethoxyphenyl)-5,7-dimethoxy-4H-chromen-4-one.

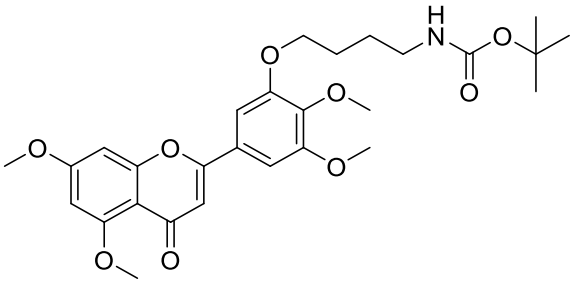
4.2.3.1.1 Synthesis of 3-(4-(1,3-dioxoisoindolin-2-yl)butoxy)-4,5-dimethoxybenzaldehyde, **4.1**.



N-(4-Bromobutyl)phthalimide (0.186 g, 0.660 mmol, 1.2 eq) was added to a solution of 3-hydroxy-4,5-dimethoxybenzaldehyde (0.100 g, 0.550 mmol, 1 eq) and potassium carbonate (0.152 g, 1.10 mmol, 2 eq) in DMF (15 mL). The solution was stirred vigorously at rt for 48 h. The reaction mixture was then diluted with water (20 mL) and extracted with ethyl acetate. The resulting crude product was purified using Biotage Isolera purification (eluent: Hexane: EtOAc 5-100%, cartridge: Biotage SNAP Ultra 10g) to give the desired product as a pale yellow oil (0.095 g, 0.250 mmol, 45.2%). ^{1}H NMR (400 MHz, $CDCl_3$-d_6) δ_H 9.85 (1H, s, CHO), 7.86 (2H, dd, J = 5.5 Hz, 3 Hz, NCOCC*H*), 7.73 (2H, dd, J = 5.5 Hz, 3 Hz, NCOCCHC*H*),

7.12 - 7.11 (2H, m, H – 2, H – 6), 4.13 – 4.10 (2H, m, COCH_2CH$_2$CH2$_2$CH$_2$N), 3.93 (3H, s, OCH$_3$ - 5), 3.92 (3H, s, OCH$_3$ – 4), 3.81 – 3.78 (2H, m, COCH$_2$CH$_2$CH$_2$CH_2N), 1.95 – 1.90 (4H, m, COCH$_2$CH_2CH_2CH$_2$N). ^{13}C NMR (100 MHz, CDCl$_3$-d_6) δ_c 191.1 (CHO), 168.4 (NCO), 153.7 (*C*OCH$_2$CH$_2$CH$_2$CH$_2$N), 152.9 (COCH$_3$ – 5), 143.9 (COCH$_3$ – 4), 134.0 (NCOCCH*C*H), 132.0 (NCO*C*CHCH), 131.6 (HOC*C*), 123.2 (NCOC*C*HCH), 108.4 (C – 2), 106.2 (C – 6), 68.5 (CO*C*H$_2$CH$_2$CH$_2$CH$_2$N), 61.0 (OCH$_3$ – 4), 56.3 (OCH$_3$ – 5), 37.5 (COCH$_2$CH$_2$CH-$_2$*C*H$_2$N), 26.5 (COCH$_2$*C*H$_2$CH$_2$CH$_2$N), 25.3 (COCH$_2$CH$_2$*C*H$_2$CH$_2$N). ESI-HRMS: m/z [M+H]$^+$ Calcd. 384.1442, found 384.1470, [M+Na]$^+$ Calcd. 406.1267, found 406.1286.

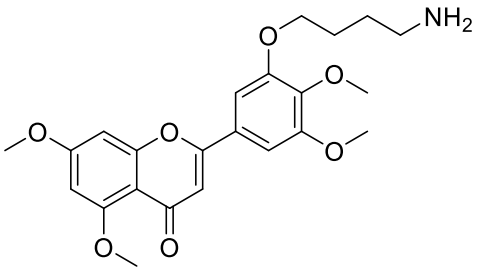
4.2.3.1.2 Synthesis of tert-butyl (4-(5-(5,7-dimethoxy-4-oxo-4H-chromen-2-yl)-2,3-dimethoxyphenoxy)butyl)carbamate, **4.2**.



tert-Butyl (4-bromobutyl)carbamate (98.8 mg, 0.390 mmol, 1.2 eq) was added to a solution of 2-(3-hydroxy-4,5-dimethoxyphenyl)-5,7-dimethoxy-4H-chromen-4-one (117 mg, 0.390 mmol, 1 eq) and potassium carbonate (90.2 mg, 0.650 mmol, 2 eq) in DMF (10 mL). The solution was stirred vigorously at rt for 48 h. This was diluted with water (20 mL) and extracted with chloroform. The resulting crude product was purified using column chromatography (eluent: 0-20% Ethyl acetate:methanol) to give the desired product as a pale yellow oil (97.0 mg, 0.180 mmol, 56.1%). ^{1}H NMR (400 MHz, CDCl$_3$-d_6) δ_H 7.01 (2H, s, H – 2', H – 6'), 6.64 (1H, s, COCH), 6.55 (1H, d, J = 2.2 Hz, H – 8), 6.33 (1H, d, J = 2.2 Hz, H – 6), 4.82 (1H, s, NH), 4.09 - 4.05 (2H, m, OCH$_2$), 3.91 (6H, s, OCH$_3$ – 5, OCH$_3$ – 7), 3.90 (3H, s, OCH$_3$ – 5'), 3.88 (3H, s, OCH$_3$ – 4'), 3.20 (2H, t, J = 3.2 Hz, CH$_2$NH), 1.91 – 1.85 (2H, m, OCH$_2$CH$_2$), 1.74 – 1.67 (2H, m, CH$_2$CH$_2$NH), 1.41 (9H, s, CCH$_3$CH$_3$CH$_3$). ^{13}C NMR (100 MHz, CDCl$_3$-d_6) δ_c 177.4 (CO), 164.2 (COCH$_3$ – 5'), 160.7 (COCH$_3$ – 7/COCH$_3$ – 5), 160.7 (C -1), 159.7 (C- 9), 156.0 (NHCOO), 153.5 (COCH$_3$ – 7/COCH$_3$ – 5), 152.7 (COCH$_2$), 141.2 (COCH$_3$ – 4'), 126.3 (C – 1'), 108.7 (C - 4), 108.2 (C – 2), 104. 6, 103.3 (C – 2', C – 6'), 96.2 (C – 6), 92.8 (C – 8), 79.0 (C(CH$_3$)$_3$), 68.9 (OCH$_2$), 60.9 (OCH$_3$ – 4'), 56.3 (OCH$_3$ – 5/ OCH$_3$ – 7), 56.2 (OCH$_3$ – 5/ OCH$_3$

– 7), 55.8 (OCH_3 – 5'), 40.1 (CH_2NH), 28.3 (CH_3), 26.75 (OCH_2CH_2), 26.4 (CH_2CH_2NH). ESI-HRMS: m/z $[M+H]^+$ Calcd. 530.2385, found 530.2386.

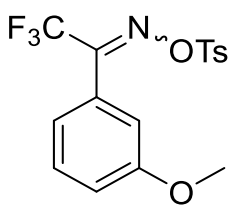
4.2.3.1.3 Synthesis of 2-(3-(4-aminobutoxy)-4,5-dimethoxyphenyl)-5,7-dimethoxy-4H-chromen-4-one, **4.3**.



N-(4-(5-(5,7-Dimethoxy-4-oxo-4H-chromen-2-yl)-2,3-dimethoxyphenoxy)butyl)pivalamide (44.0 mg, 0.083 mmol) was added to a stirred solution of HCl (5 mL, 4N in dioxane) at 0 °C, the reaction was then warmed to room temperature and stirred for 2 h. Ethyl acetate was added (30 mL) and the solution washed with aqueous 4M NaOH (3 x 30 mL) until the solution turned colourless. The aqueous layer was then extracted with ethyl acetate again and the combined organic layers were then dried under reduced pressure to give the desired product. (28.0 mg, 0.065 mmol, 78.5%).^{1}H NMR (400 MHz, DMSO-d_6) δ_H 7.29 (2H, d, J = 2.4 Hz, H – 2', H – 6'), 6.91 (1H, d, J = 2.3, H - 8), 6.88 (1H, s, COCH), 6.51 (1H, d, J = 2.3, H – 6), 4.11 (2H, t, J = 6.3, OCH_2), 4.02 (Ethyl acetate), 3.91 (3H, s, OCH_3 – 5), 3.89 (3H, s, OCH_3 – 5'), 3.83 (3H, s, OCH_3 – 7), 3.75 (3H, s, OCH_3 – 4'), 3.56 (Dioxane), 2.67 (2H, s (br), CH_2NH), 1.82 – 1.76 (2H, m, OCH_2CH_2), 1.61 – 1.54 (2H, m, CH_2CH_2NH). ^{13}C NMR (100 MHz, DMSO-d_6) δ_c 175.8 (CO), 170.4 (Ethyl acetate), 163.8 ($COCH_3$ – 5), 160.2 ($COCH_3$ – 7), 159.5 (C – 1), 159.2 (C – 9), 153.3 ($COCH_3$ – 5'), 152.7 ($COCH_2$), 140.4 ($COCH_3$ – 4'), 126.2 (C – 1'), 108.3 (C – 4), 108.2 (C – 2), 104.5, 103.5 (C – 2', C – 6'), 96.3 (C – 6), 93.5 (C – 8), 68.7 ($COCH_2$), 66.4 (Dioxane), 60.2 (OCH_3 – 4'), 59.8 (Ethyl acetate), 56.3 (OCH_3 – 5', OCH_3 – 5), 56.1 (OCH_3 – 7), 26.3 ($COCH_2CH_2$, $CH_2CH_2NH_2$) 20.8 (Ethyl acetate). ESI-HRMS: m/z [M+H]+ Calcd. 430.1860, found 430.1864.

4.2.3.2 *Synthesis of N-butyl-2-(2-(2-(prop-2-yn-1-yloxy)ethoxy)ethoxy)-4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzamide.*

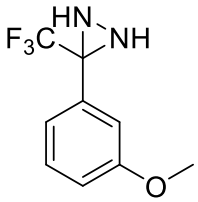
4.2.3.2.1 Synthesis of 2,2,2-trifluoro-1-(3-methoxyphenyl)ethan-1-one *O*-tosyl oxime, **4.6**.



n-Butyl lithium (9.48 mL, 23.7 mmol, 1.5 eq) was added slowly to a solution of 3-bromoanisole (2.00 mL, 15.8 mmol, 1 eq) in dry THF (32 mL) at – 84 °C and stirred for 2 h under N_2. Methyl trifluoroacetate (2.07 mL, 20.5 mmol, 1.3 eq) in dry THF (13 mL), cooled to – 84 °C, was then added initial solution over 30 minutes and stirred for 30 minutes more. The solution was then quenched with sat. ammonium chloride solution (12 mL) then allowed to warm to rt gradually. The solution was then extracted with Et_2O (25 mL), washed with water (3 x 15 mL) and brine (15 mL), then dried with magnesium sulphate and filtered. The organic layers were then concentrated to give a crude mixture of 2,2,2-trifluoro-1-(3-methoxyphenyl)ethan-1-one which was used without purification (2.50 g). ^{1}H NMR (400 MHz, $CDCl_3$-d_6) δ_H 7.67 (1H, dq, J = 7.7, 1.3 Hz, Ph – 6), 7.58 (1H, s, Ph – 2), 7.47 (1H, t, J = 8 Hz, Ph – 5), 7.28 – 7.25 (1H, m, Ph – 4), 3.89 (3H, s, OCH_3).[201] (in line with published ^{1}H NMR data). ^{13}C NMR (100 MHz, $CDCl_3$-d_6) δ_C 180.4 (q, CO), 159.9 ($COCH_3$), 131.0 (Ph – 5) , 130.1 (Ph – 2), 122.7 (Ph – 6), 122.3 (Ph - 4), 166.6 (q, CF_3), 113.9 (Ph - 1), 55.5 (CH_3). Hydroxylamine hydrochloride (1.15 g, 16.53 mmol, 1.25 eq) was dissolved in dry ethanol (9 mL) and dry pyridine (12 mL) and added to the crude mixture dissolved in dry ethanol (9 mL) and dry pyridine (12 mL) under N_2 and the reaction was refluxed at 85 °C for 19 h. The solution was then concentrated under vacuum and extracted with diethyl ether (25 mL) and washed with water (25 ml) and then 1M HCl (3 x 25 mL) and brine (25 mL). Organic layers were dried with magnesium sulphate, filtered and concentrated under vacuum to give the crude mixture (2.8 g). ^{1}H NMR (400 MHz, $CDCl_3$-d_6) δ_H 9.08 (1H, s, OH), 7.41 (1H, td, J = 7.8 Hz, Ph – 5), 7.10 – 7.08 (1H, m, Ph – 6), 7.06 – 7.03 (2H, m, Ph – 2, Ph – 4), 3.85 (3H, s, CH_3).[202] ^{13}C NMR (100 MHz, $CDCl_3$-d_6) δ_C 159.11 ($COCH_3$), 147.2 (q, CN), 129.4 (Ph – 5), 126.8 (Ph – 1), 120.5 (Ph – 6), 120.2 (q, CF_3), 115.9 (Ph – 4), 113.9 (Ph – 2), 55.04 (CH_3). ESI-MS $[M+H]^+$ Calcd. 242, found 242. The crude mixture was dissolved in

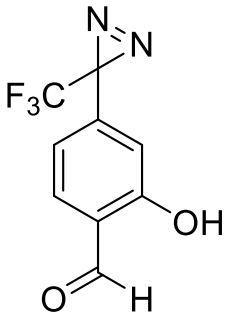
dry DCM (30 mL) and added to a solution of DMAP (0.16 g, 1.31 mmol, 0.11 eq) and tosyl chloride (2.91 g, 15.26 mmol, 1.29 eq) and triethylamine (2.66 mL, 19.10 mmol, 1.5 eq) which had been stirred at 0 °C for 30 minutes. The reaction mixture was stirred at rt for 20 h and then washed with 1M HCl (3 x 50 mL), dried with magnesium sulphate, filtered, and reduced under pressure. The desired product was obtained after recrystallization from ethanol (2.59 g, 0.16 mmol, 58.5 %). ^{1}H NMR (400 MHz, $CDCl_3$-d_6) δ_H 7.91 – 7.88 (2H, m, CH_3CCHCH), 7.41 – 7.37 (4H, CH_3CCH, Ph - 5), 7.06 (1H, ddd, J = 8.4, 2.6, 0.9 Hz, Ph - 5), 6.96 – 6.94 (1H, m, Ph – 6), 6.89 (1H, m, Ph – 4), 3.83 (3H, s, OCH_3), 2.49 (3H, s, CH_3).[202] ^{13}C NMR (100 MHz, $CDCl_3$-d_6) δ_C 159.9 ($COCH_3$), 146.4 (CH_3C), 130.3 (Ph – 5), 130.2 (CH_3CC), 129.6 (CH_3CCHC), 128.7 (CF_3), 120.8 (Ph – 4), 117.5 (Ph – 6), 114.2 (Ph – 2), 55.7 (OCH_3), 22.1 (CH_3). ESI-MS: [M+H$^+$] calcd. 374 found 373.

4.2.3.2.2 Synthesis of 3-(3-methoxyphenyl)-3-(trifluoromethyl)diaziridine, **4.7**.



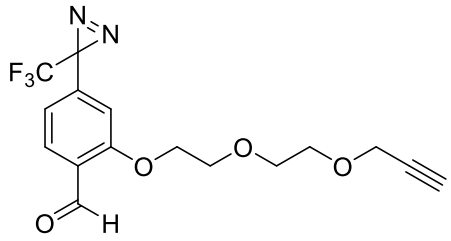
2,2,2-trifluoro-1-(3-methoxyphenyl)ethan-1-one *O*-tosyl oxime (1.63 g, 4.35 mmol) was dissolved in 7N NH_3 in methanol (18 mL) at 0 °C under N_2 and then allowed to warm to rt and stirred for 72 h. After venting with N_2, the reaction was quenched with sat. $NaHCO_3$ (3 mL) and filtered to remove the precipitate. The filtrate was then diluted with DCM (25 mL), washed with water (3 x 25 mL), dried with magnesium sulphate, filtered, and dried under vacuum to give a colourless crude oil. Desired product was obtained after column chromatography (eluent: 90:10 Hex:Et_2O) (0.260 g, 1.20 mmol, 27.6%). ^{1}H NMR (400 MHz, $CDCl_3$-d_6) δ_H 7.29 – 7.25 (1H, m, Ph – 5), 7.16 – 6.91 (3H, m, Ph – 2, Ph – 4, Ph – 6), 3.75 (3H, s, CH_3), 2.81, 2.32 (2H, s, NH).[202] (in line with published ^{1}H NMR data.) ^{13}C NMR (100 MHz, $CDCl_3$-d_6) δ_C 159.2 (CHO), 132.6 (Ph – 1), 129.4 (Ph – 5), 123.1 (CF_3), 119.8, 115.3, 113.2 (Ph – 2, Ph – 4, Ph – 6), 54.7 (CHO).

4.2.3.2.3 Synthesis of 2-hydroxy-4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzaldehyde, **4.10**.



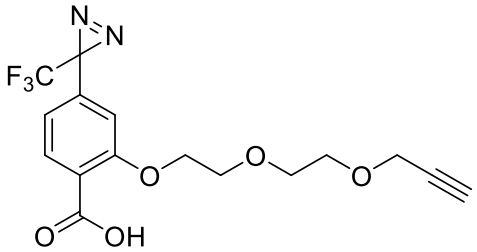
Synthesis of 3-(3-methoxyphenyl)-3-(trifluoromethyl)-3H-diazirine: 3-(3-methoxyphenyl)-3-(trifluoromethyl)diaziridine (0.260 g, 1.20 mmol, 1 eq) was dissolved in dry Et_2O (8 mL) and added dropwise to a solution of silver oxide (2.78 g, 12.0 mmol, 10 eq) in dry Et_2O (8 mL). The solution was protected from the light and stirred at rt for 24 h. The solution was then filtered and dried with magnesium sulphate and dried under reduced pressure. Completion of the reaction was monitored using TLC (Eluent: 80:20, Et_2O:Hex). The crude compound was then used directly in the next step of the synthesis. Synthesis of 2-methoxy-4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzaldehyde: the crude compound was dissolved in dry DCM (10 mL), placed in an inert atmosphere and cooled to 0 °C. Titanium tetrachloride (197.5 µL, 1.80 mmol, 1.5 eq) was added to the solution, followed by dichloromethyl methyl ether (162.9 µL, 1.80 mmol, 1.5 eq). After 10 mins, the solution was allowed to warm to rt and stirred for 2 h. The reaction was then quenched with water (2 mL), diluted with DCM (25 mL), washed with water (2 x 15 mL) then sat. $NaHCO_3$ solution (2 x 15 mL). The solution was then dried with magnesium sulphate, filtered and then dried under reduced pressure. Synthesis of 2-hydroxy-4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzaldehyde: the crude compound was then used directly in the next step of the synthesis, it was dissolved in dry DCM (5.5 mL) and boron tribromide (1M in CH_2Cl_2, 1.32 mL, 1.32 mmol , 1.1 eq) was added dropwise. The reaction was stirred at rt for 2 h, and product formation monitored by TLC (Eluent: 80:20, Et_2O:Hex). The solution was then quenched with water (2 mL) and diluted with DCM (15 mL) and washed with sat. $NaHCO_3$ (2 x 20 mL). The solution was then dried with magnesium sulphate and under reduced pressure following filtration. The crude compound was then purified with column chromatography (eluent: 10:90, Et_2O:Hex) to give the desired product (24.0 mg, 0.100 mmol, 8.68%). ^{1}H NMR (400 MHz, $CDCl_3$-d_6) δ_H 11.06 (1H, s, OH), 9.93 (1H, s, CHO), 7.61 (1H, m, Ph – 6), 6.81 (2H, m, Ph – 3, Ph – 5).[202] (in line with published ^{1}H NMR data.) ^{13}C NMR (100 MHz, $CDCl_3$-d_6) δ_C 196.2 (CO), 161.7 (CHO), 138.3 (CHO), 134.3 (Ph – 6), 122.0 (q, CF_3), 121.1 (Ph – 4), 117.6, 116.2 (Ph – 3, Ph – 5).

4.2.3.2.4 Synthesis of 2-(2-(2-(prop-2-yn-1-yloxy)ethoxy)ethoxy)-4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzaldehyde, **4.11**.



2-Hydroxy-4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzaldehyde (100 mg, 0.430 mmol, 1 eq) was dissolved in dry DMF (10 mL) with 3-(2-(2-bromoethoxy)ethoxy)prop-1-yne (99.0 mg, 0.480 mmol, 1.1 eq), potassium carbonate (96.1 mg, 0.700 mmol, 1.6 eq) and tetrabutylammonium iodide (16.1 mg, 0.043 mmol, 0.1 eq) and refluxed at 65 °C for 24 h in an inert atmosphere. The mixture was then diluted with ethyl acetate (20 mL), washed with water (3 x 20 mL) and brine (20 mL). Organic layers were then combined and dried to give a crude mixture which was purified using column chromatography (eluent: 1:1 Et_2O:Hex) to give the purified product. (85.0 mg, 0.240 mmol, 54.9%). ^{1}H NMR (400 MHz, $CDCl_3$-d_6) δ_H 10.49 (1H, s, CHO), 7.85 (1H, d, J = 8.2 Hz, Ar-6), 6.85 (1H, dt, J = 8.2 Hz, 0.7, Ar-3, Ar-5), 6.74 (1H, s, CHCOCH_2), 4.28 - 4.26 (2H, m, OCH_2CH$_2$OCH$_2$CH$_2$OCH$_2$), 4.21 (2H, d, J = 2.4 Hz, OCH$_2$CH$_2$OCH$_2$CH$_2$OCH_2), 3.95 -3.92 (2H, m, OCH$_2$CH_2OCH$_2$CH$_2$OCH$_2$), 3.78 - 3.71 (4H, m, OCH$_2$CH$_2$OCH_2CH_2OCH$_2$), 2.45 (1H, t, J = 2.4 Hz, CH$_2$CCH). ^{13}C NMR (100 MHz, $CDCl_3$-d_6) δ_c 188.8 (*C*HO), 160.9 (*C*OCH$_2$), 136.6 (Ar-*C*CN$_2$CF$_3$), 128.8 (Ar-*C*HCCHO), 125.8, 123.1, 120.4, 117.6 (q, J_{CF} = 273.7 Hz, *C*F$_3$), 125.5 (Ar-*C*CHO), 118.8 (Ar-*C*HCCN$_2$), 110.8 (Ar-*C*HCOCH$_2$), 79.4 (*C*≡CH), 74.7 (C≡*C*H), 70.7 (OCH$_2$CH$_2$O*C*H$_2$CH$_2$OCH$_2$), 69.4 (O*C*H$_2$CH$_2$OCH$_2$CH$_2$OCH$_2$), 69.0 (OCH$_2$CH$_2$OCH$_2$*C*H$_2$OCH$_2$), 68.5 (OCH$_2$*C*H$_2$OCH$_2$CH$_2$OCH$_2$), 58.4 (OCH$_2$CH$_2$OCH$_2$CH$_2$O*C*H$_2$), 29.1, 28.7, 28.3, 27.9 (q, J_{CF} = 40.9 Hz, *C*N$_2$). ESI-MS: m/z [M+K]$^+$ Calcd. 395, found 395.

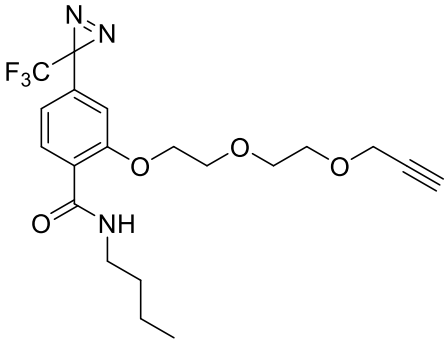
4.2.3.2.5 Synthesis of 2-(2-(2-(prop-2-yn-1-yloxy)ethoxy)ethoxy)-4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzoic acid, **4.12**.



2-(2-(2-(Prop-2-yn-1-yloxy)ethoxy)ethoxy)-4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzaldehyde (82.0 mg, 0.230 mmol, 1 eq) was dissolved in THF (2.4 mL) under N_2 and sulfamic acid (41.3 mg, 0.430 mmol, 1.85 eq), dissolved in water (1.6 mL) was added.

After stirring for 5 minutes, sodium chlorite (31.2 µL, 25% w/v, 3.32 M solution, 1.80 eq) was injected. Solution was stirred at rt for 4 hours, quenched with water and extracted with DCM (3 x 10 mL) then concentrated to give the desired compound that required no purification. (77.0 mg, 0.210 mmol, 89.9 %). ^{1}H NMR (400 MHz, $CDCl_3$-d_6) δ_H 8.18 (1H, d, J = 8.3, Ar-CHCCOOH) 6.95 (1H, dd, J = 8.3, 0.7, CHCHCCOOH), 6.78 (1H, s, $CHCOCH_2$), 4.40 – 4.38 (2H, m, $OCH_2CH_2OCH_2CH_2OCH_2$), 4.19 (2H, d, J = 2.4, $OCH_2CH_2OCH_2CH_2OCH_2$), 3.96 – 3.94 (2H, m, $OCH_2CH_2OCH_2CH_2OCH_2$), 3.78 – 3.72 (4H, m, $OCH_2CH_2OCH_2CH_2OCH_2$), 2.44 (1H, t, J = 2.4 Hz, C≡C*H*). ^{13}C NMR (100 MHz, $CDCl_3$-d_6) δ_c 164.5 (COOH), 157.4 ($COCH_2$), 135.8 (CCN_2), 134.2 (Ar-CHCCOOH), 123.0 (CF_3), 120.3 (Ar-$CHCCN_2$), 119.7 (CCOOH), 111.5 ($CHCOCH_2$), 79.3 (*C*≡CH), 74.7 (C≡*C*H), 70.6 ($OCH_2CH_2OCH_2CH_2OCH_2$), 69.5 ($OCH_2CH_2OCH_2CH_2OCH_2$), 69.1 ($OCH_2CH_2OCH_2CH_2OCH_2$), 68.5 ($OCH_2CH_2OCH_2CH_2OCH_2$), 58.4 ($OCH_2CH_2OCH_2CH_2OCH_2$), 28.5 – 28.1 (CN_2). ESI-MS: m/z $[M+Na]^+$ Calcd. 395, found 395.

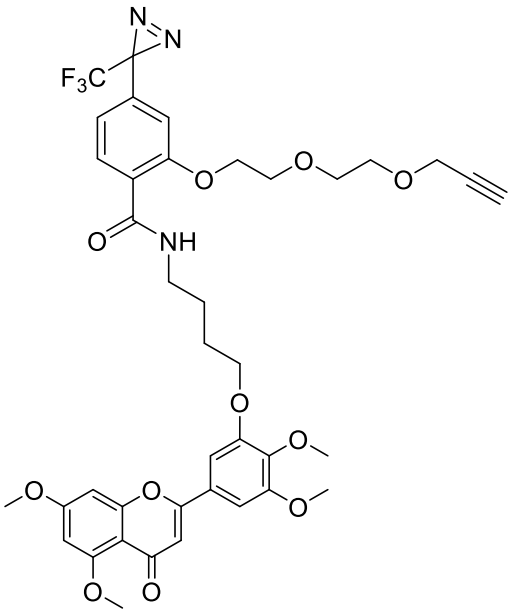
4.2.3.2.6 Synthesis of N-butyl-2-(2-(2-(prop-2-yn-1-yloxy)ethoxy)ethoxy)-4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzamide, **4.13**.



EDC.HCl (30.9 mg, 0.160 mmol, 2 eq), HOBt (21.8 mg, 0.160 mmol, 2 eq) and 2-(2-(2-(prop-2-yn-1-yloxy)ethoxy)ethoxy)-4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzoic acid (30.0 mg, 0.080 mmol, 1 eq) were dissolved in DMF (6 mL) with n-butyl amine (120 µL, 0.120 mmol, 1.5 eq) and stirred at 0 °C for 5 minutes. Triethyl amine (22.5 µL, 0.160 mmol, 2 eq) was added and the solution stirred at rt for 24 h. The solution diluted with 1M HCl (10 mL) and extracted with ethyl acetate (3 x 15 mL) then concentrated to give the crude compound. The crude mixture was purified initially using column chromatography (eluent: 40:60 EtOAc:Hex) then a second silica plug (eluent: Hex then 40:60 EtOAc:Hex then EtOAc) to remove grease. This gave the desired compound as a pale yellow oil (13.0 mg, 0.030 mmol, 38.2%). ^{1}H NMR (400 MHz, $CDCl_3$-d_6) δ_H 8.22 (1H, d, J = 8.2 Hz, CHCCONH), 8.03 (1H, s, NH), 6.89 (1H, dd, J = 8.2, 0.8 Hz, $CHCCN_2$),

6.69 (1H, s, $CHCOCH_2$), 4.28 – 4.26 (2H, m, $OCH_2CH_2OCH_2CH_2OCH_2$), 4.18 (2H, d, J = 2.3 Hz, $OCH_2CH_2OCH_2CH_2OCH_2$), 3.93 – 3.91 (2H, m, $OCH_2CH_2OCH_2CH_2OCH_2$), 3.76 – 3.71 (4H, m, $OCH_2CH_2OCH_2CH_2OCH_2$), 3.45 (2H, td, J = 7.1, 5.7 Hz, $CONHCH_2$), 2.43 (1H, t, J = 2.4 Hz, C≡CH), 1.65 – 1.56 (m, $CONHCH_2CH_2$), 1.46 – 1.37 (m, $CONHCH_2CH_2CH_2$), 0.96 (t, J = 7.3 Hz, CH_3). ^{13}C NMR (400 MHz, $CDCl_3$-d_6) δ_C 164.1 (CONH), 156.6 (CO), 133.2 (Ph – 4), 133.0 (Ph – 6), 123.6 (Ph – 1), 122.6 (CF_3), 119.5 (Ph – 5), 110.5 (Ph – 3), 79.3 (*C*=CH), 74.8 (C=*C*H), 70.5 ($OCH_2CH_2OCH_2CH_2OCH_2CH$), 69.1 ($OCH_2CH_2OCH_2CH_2OCH_2CH$, $OCH_2CH_2OCH_2CH_2OCH_2CH$), 68.2 ($OCH_2CH_2OCH_2CH_2OCH_2CH$), 58.5 ($OCH_2CH_2OCH_2CH_2OCH_2CH$), 39.7 ($NHCH_2$), 31.9 ($NHCH_2CH_2$), 20.3 ($NHCH_2CH_2CH_2$), 14.1 ($NHCH_2CH_2CH_2CH_2$). ESI-MS: m/z $[M+H]^+$ Calcd. 428.2, found 428.4, m/z $[M+Na]^+$ Calcd. 450.2, found 450.5. ESI-HRMS: m/z $[M+Na]^+$ Calcd. 428.1792, found 428.1799.

4.2.3.3 Synthesis of N-(4-(5-(5,7-dimethoxy-4-oxo-4H-chromen-2-yl)-2,3-dimethoxyphenoxy)butyl)-2-(2-(2-(prop-2-yn-1-yloxy)ethoxy)ethoxy)-4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzamide, 4.14.



EDC.HCl (25.8 mg, 0.140 mmol, 2 eq), HOBt (18.2 mg, 0.140 mmol, 2 eq) and 2-(2-(2-(prop-2-yn-1-yloxy)ethoxy)ethoxy)-4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzoic acid (27.6 mg, 0.070 mmol, 1.1 eq) were dissolved in DMF (6 mL) with 2-(3-(3-aminopropoxy)-4,5-dimethoxyphenyl)-5,7-dimethoxy-4H-chromen-4-one (28.0 mg, 0.070 mmol, 1 eq) and stirred at 0 °C for 5 minutes. Triethylamine (18.8 µL, 0.140 mmol, 2 eq) was added and the solution stirred at rt for 24 h. The solution was then diluted with 1M HCl (10 mL) and extracted with ethyl acetate (3 x 15 mL) then concentrated to give the crude compound. The crude mixture was purified using column chromatography (eluent: EtOAc to 90:10 EtOAc:MeOH) to give the desired compound as a pale yellow oil (12.6 mg, 0.020 mmol, 23.9%). ^{1}H NMR (400 MHz, $CDCl_3$-d_6) δ_H δ_H 8.22 (1H, br t, J = 5 Hz, NH), 8.12 (1H, d, J = 8.2 Hz, Ar-

CHCCONH), 7.28 (2H, dd, J = 14.1, 1.6 Hz, H – 2', H – 6'), 7.02 (1H, d, J = 8.2 Hz, $CHCCN_2$), 6.84 (1H, s, $CHCOCH_2$), 6.72 (1H, d, J = 2.3 Hz, H – 8), 6.58 (1H, s, H – 2), 6.45 (1H, s, J = 2.1 Hz, H – 6), 4.38 – 4.35 (2H, m, $OCH_2CH_2OCH_2CH_2OCH_2$), 4.23 (2H, t, J = 6.2 Hz, $CONHCH_2CH_2CH_2CH_2$), 4.15 (2H, d, J = 2.3 Hz, $OCH_2CH_2OCH_2CH_2OCH_2$), 3.94 – 3.91 (8H, m, OCH_3 – 5, OCH_3 – 5', $OCH_2CH_2OCH_2CH_2OCH_2$), 3.88 (3H, s, OCH_3 – 7), 3.83 (3H, s, 4' – OCH_3), 3.74 – 3.51 (4H, m, $OCH_2CH_2OCH_2CH_2OCH_2$), 3.53 (2H, q, J = 6.5 Hz, $CONHCH_2$), 2.91 – 2.88 (m, C≡CH), 1.98 – 1.93 (2H, m, $CONHCH_2CH_2CH_2$), 1.90 – 1.84 (2H, m, $CONHCH_2CH_2CH_2CH_2$). ^{13}C NMR (400 MHz, $CDCl_3$-d_6) δ_C 175.7 (CO), 164.1 (CONH),163.4, 153.8 ($COCH_3$ – 5, $COCH_3$ – 5'), 160.9 ($COCH_3$ – 7), 159.9 (C – 4), 159.7 (C – 2), 157.0 (Ph – 2), 153.1 ($COCH_2$ – 3'), 141.2 ($COCH_3$ – 4'), 132.5 (Ph – 6), 132.4 (Ph – 1), 126.7 (C – 1'), 124.4 (CCN_2), 123.5 (q, CF_3), 119.1 (Ph – 5), 111.1 (Ph – 3), 108.9 (C – 9), 108.2 (C – 2), 104.8, 103.5 (C – 2', C – 6'), 95.9 (C – 6), 93.1 (C – 8), 79.8 (OCH_2CCH), 75.1 (OCH_2CCH), 69.0 ($OCH_2CH_2OCH_2CH_2OCH_2$), 68.9, 55.9, 55.5 (OCH_3 – 5', OCH_3 – 5, $OCH_2CH_2OCH_2CH_2OCH_2$), 68.6 ($OCH_2CH_2OCH_2CH_2OCH_2$), 60.0 (OCH_3 – 4'), 57.8 ($OCH_2CH_2OCH_2CH_2OCH_2$), 39.1 ($NHCH_2$), 26.8 ($NHCH_2CH_2CH_2$), 26.2 ($NHCH_2CH_2$). ESI-HRMS: m/z $[M+H]^+$ Calcd. 784.2688, found 784.2728. Found m/z value is 0.004 greater than calculated.

4.2.3.4 Evaluation of biological activity of modified PMF and active and inactive probe compounds

Evaluation of the biological activity of PMF after modification with the linker group, as well as the final active and inactive probe compounds, was carried out using methods discussed in section 3.2.

4.2.3.5 UV-Vis analysis of diazirine reactivity timescale

The timescale in which the UV reactive probe molecule would degrade was assessed using UV-Vis spectroscopic techniques. This was carried out by recording the UV-Vis absorbance of the biologically inactive probe compound without any irradiation, then identifying the maximum wavelength in the 330 – 420 nm section of the spectrum, as this corresponded to the absorbance of the UV reactive diazirine moiety in the probe. The samples were then irradiated with UV light at 365 nm and UV-Vis spectra recorded at a series of timepoints. The absorbance value at this maximum decreased over time and was plotted to assess the length of irradiation time required to achieve complete loss of the

diazirine moiety. This experiment was carried out in triplicate to ensure reliability of the results obtained.

4.3 Results and Discussion

4.3.1 Attachment of a linker group to PMF in a suitable modification point

Results from the SAR study of the modified flavone analogues, discussed in Chapter 2, showed that modification of the PMF structure in the 3'-position resulted in the least change of desired biological activity with respect to PMF, and compared to modification in the other positions assessed. From these results, it was anticipated that the probe molecule required for the pull down study could be synthesised at this position without loss of biological activity. To synthesise this probe molecule, a linker group must first be affixed to the flavone structure to allow attachment of the UV active diazirine probe. To facilitate the attachment of the desired TPD moiety to the flavone structure, a carboxylic acid group is typically synthesised in the *para* position of the molecule (Figure 4.4), this can then be attached to an amine group using widely tested amide coupling chemistry methods.[203]

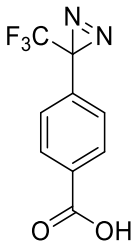
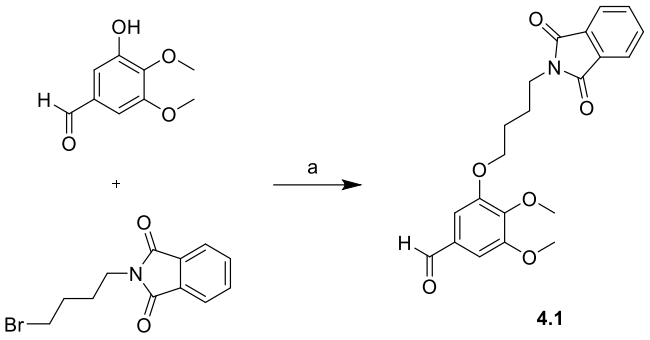


Figure 4.4: Structure of trifluoromethylphenyl diazirine photoreactive molecule with a carboxylic acid group installed in the *para* position to allow for simple attachment to targeting and reporter groups by amide coupling chemistry methods.

To use the amide coupling technique to attach the targeting flavone structure to the photoactive group, first an amine linker group must be attached to the flavone.

4.3.1.1 Synthetic attachment of an amine linker group to the flavone structure using a phthalimide protecting group.

The initial method trialled for adding the amine attachment to the flavone compound, involved the addition of an alkyl phthalimide. Firstly, the phthalimide group was introduced using a substitution of the hydroxyl group in the 3'-position of the aldehyde precursor to the final flavone structure, 3-hydroxy-4,5-dimethoxybenzaldehyde, with *N*-(4-bromobutyl)phthalimide (Scheme 4.1). This synthetic route was followed in a similar manner to the introduction of the benzyl group modification carried out for the synthesis of the **4'-OBn** analogue discussed in section 2.3.3, however the conditions used were altered to extend reaction times from 24 h to 48 h after monitoring the progress of the reaction with TLC.



Scheme 4.1: Attachment of alkyl phthalimide to 3-hydroxy-4,5-dimethoxybenzaldehyde to generate the starting material required to install an alkyl amine linker in to the polymethoxyflavone structure in the required 3'-position. Reagents and conditions: a) K_2CO_3, DMF, rt, 48 h.

This compound, **4.1**, was synthesised successfully with a yield of 45.2%, and characterised by ^{1}H and ^{13}C NMR alongside 2D NMR techniques and HRMS (Figure 4.5, Appendix 8.22).

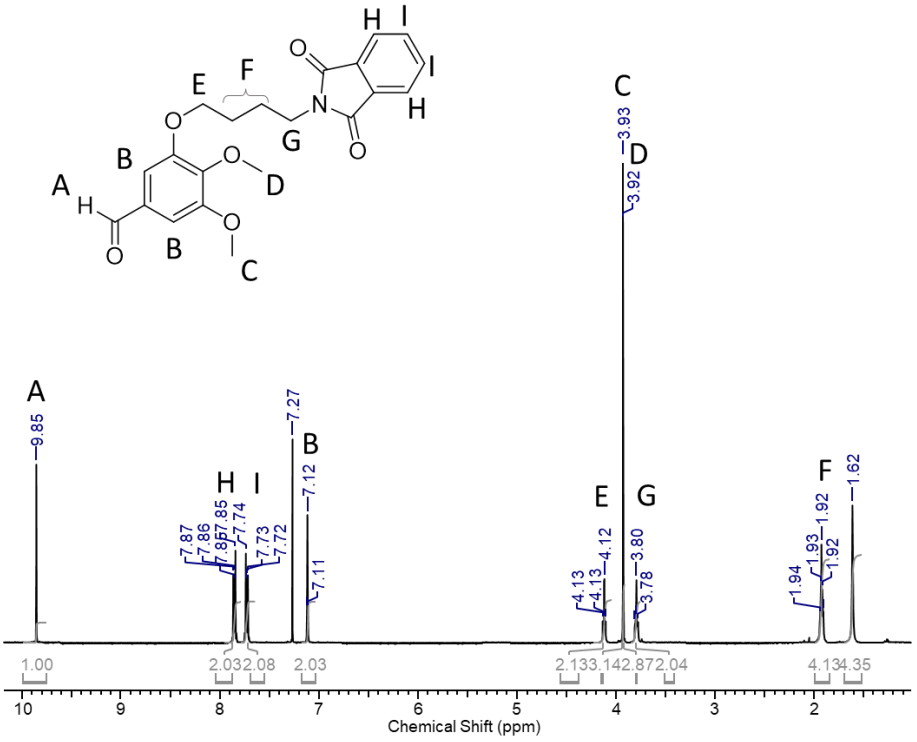


Figure 4.5: Assigned ^{1}H NMR spectrum for 3-(4-(1,3-dioxoisoindolin-2-yl)butoxy)-4,5-dimethoxybenzaldehyde, **4.1**.

Synthesis of the substituted chalcone using an aldol reaction as the next step of the route to the desired flavone however, was unsuccessful, though formation of the desired flavone structure was visible by NMR analysis, it appeared as though the phthalimide group had decomposed during the reaction. The strong basic conditions used for the chalcone synthesis step, described in Chapter 2, proved to be too harsh for the phthalimide protecting group, as the group can be base labile,[204] and resulted in the formation of a complex mixture containing an alkyl amine, this mixture was difficult to purify and therefore an alternative route was sought.

The next route tested followed a similar method to that used for the synthesis of the **3'-OBn** flavone analogue described in section 2.3.2. The hydroxy flavone intermediate, **2.5**, was synthesised and then attachment of the phthalimide protected amine linker was tested as a final step of the synthesis. However, when carrying out the substitution of the phenol group with the bromoalkyl phthalimide shown in Scheme 4.1, in the presence of potassium carbonate in DMF, the reaction proved

unsuccessful. LCMS analysis of the crude product formed from the reaction showed peaks corresponding to the desired product, but also ions of the product in various points of phthalimide decomposition (Figure 4.6).

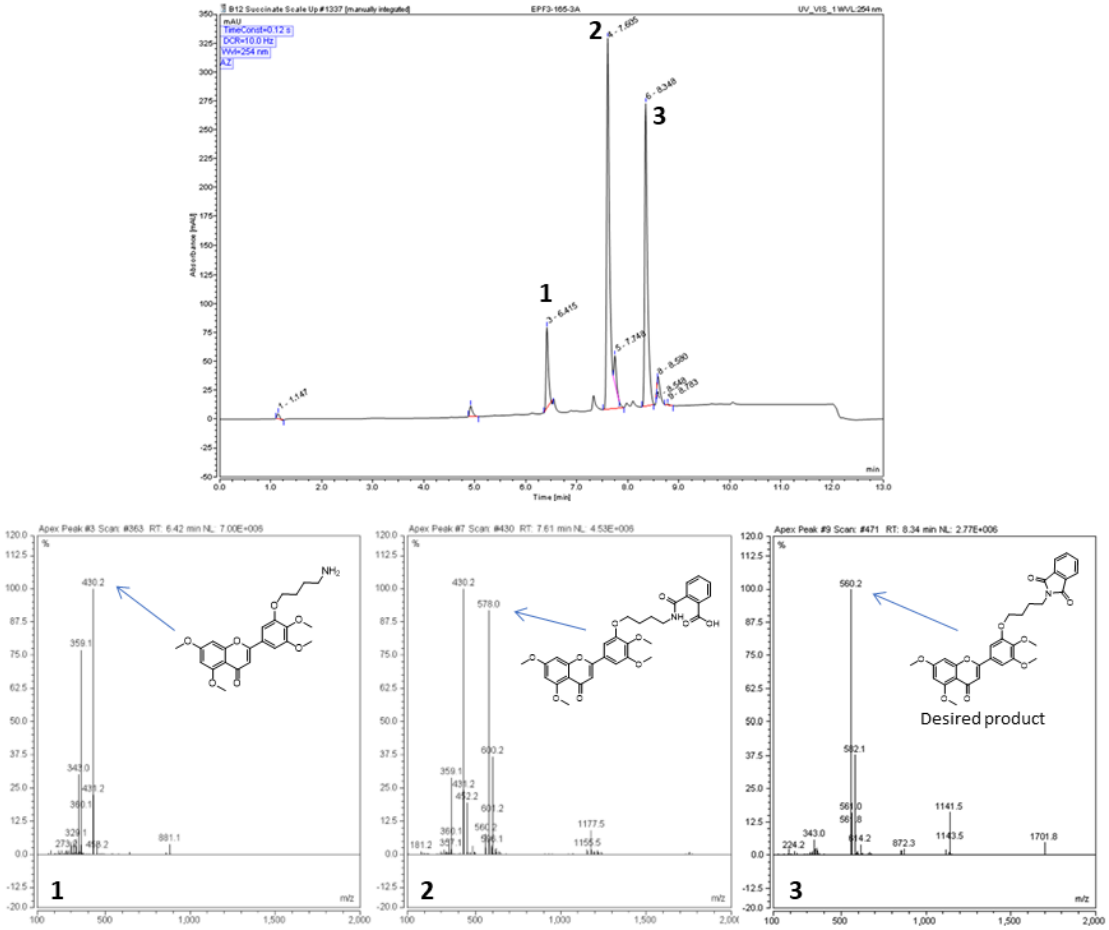


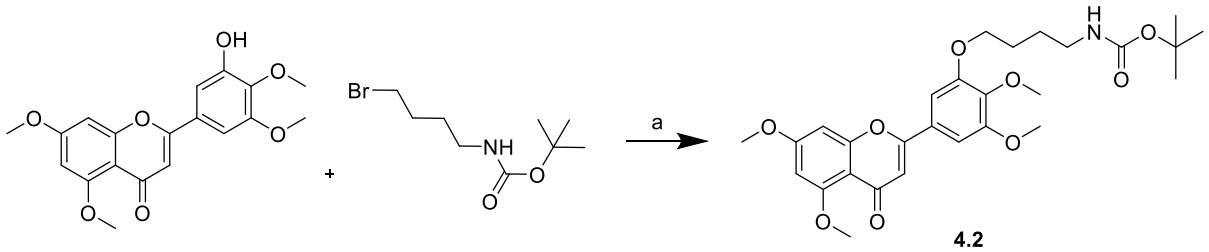
Figure 4.6: LCMS analysis of crude mixture of products after testing a synthetic route to 2-(4-(5-(5,7-dimethoxy-4-oxo-4H-chromen-2-yl)-2,3-dimethoxyphenoxy)butyl)isoindoline-1,3-dione. The first image shows the UV chromatogram from the LCMS with the peaks 1,2 and 3 indicated. Images 2-4 show the ESI-MS spectra of these peaks respectively with the uncharged compounds the peaks refer to indicated with arrows.

It could be the case that the presence of the potassium carbonate as a base in the solution could encourage the undesired decomposition of the phthalimide protecting group. To avoid a lengthy purification process that may come from trying to separate the crude mixture to isolate the desired

product at likely low yields, an alternative strategy was investigated. This strategy aimed to determine a method that could be scaled effectively for subsequent synthesis for the pull down probe.

4.3.1.2 Synthetic attachment of an amine linker group to the flavone structure using a Boc protecting group.

A Boc protected amine linker was trialled next. The 3'-hydroxy flavone, **2.5**, was reacted with 4-(Boc-amino)butyl bromide in the presence of potassium carbonate in DMF, Scheme 4.2, to give the desired Boc protected product after column purification.



Scheme 4.2: Synthetic route for N-(4-(5-(5,7-dimethoxy-4-oxo-4H-chromen-2-yl)-2,3-dimethoxyphenoxy)butyl)pivalamide, **4.2**. Reagents and conditions: a) 4-(Boc-amino)butyl bromide, K_2CO_3, DMF, rt, 48 h.

The reaction proceeded smoothly, without the formation of the free amine, and was obtained with a 56.1% yield after purification. The product was characterised by ^{1}H, ^{13}C and 2D NMR as well as mass spectrometry (Figure 4.7, Appendix 8.23).

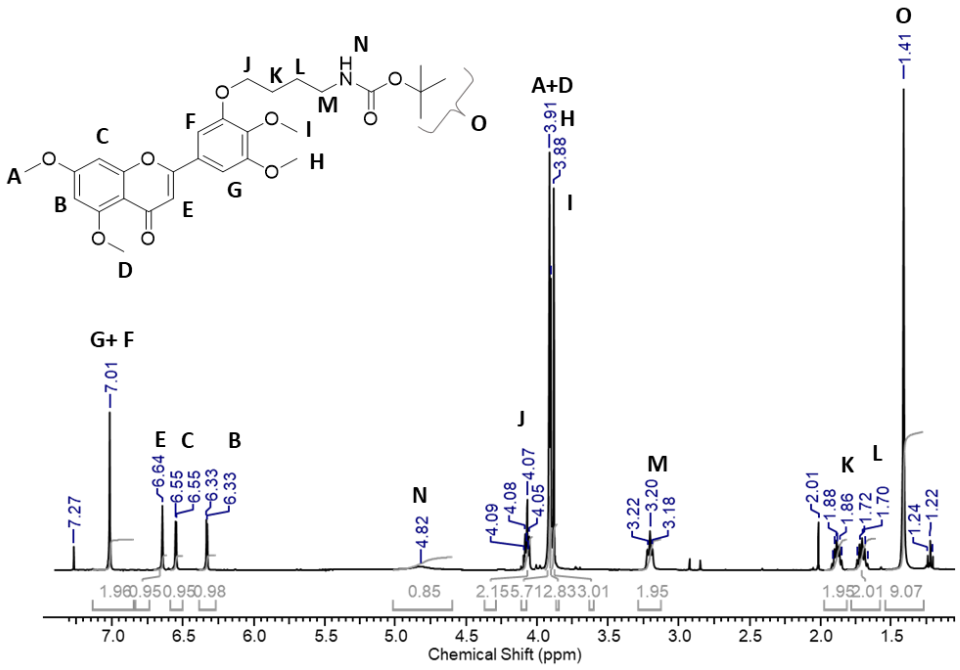
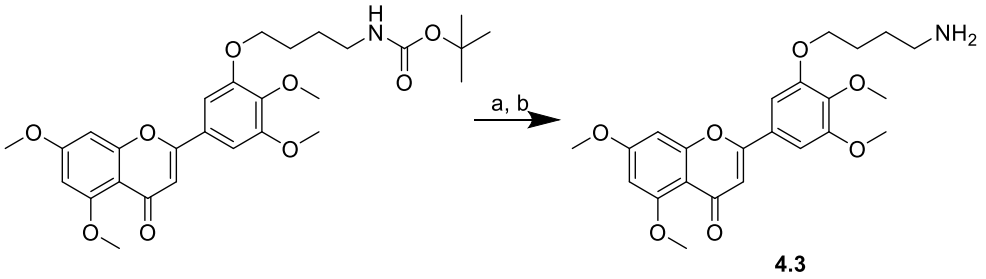


Figure 4.7: Assignments of ^{1}H NMR spectrum for *N*-(4-(5-(5,7-dimethoxy-4-oxo-4H-chromen-2-yl)-2,3-dimethoxyphenoxy)butyl)pivalamide, **4.2**.

From the compound **4.2**, the required deprotected amine, **4.3**, could be obtained (Scheme 4.3). The Boc protecting group was removed by stirring the solution in 4M HCl in dioxane, resulting in formation of the positively charged ammonium ion, and subsequent washing of the solution with aqueous NaOH solution allowed the formation of the neutral alkyl amine group. This compound was characterised by ^{1}H, ^{13}C and 2D NMR as well as HRMS (Appendix 8.24). The crude mixture of this step was used directly in further steps of the synthesis of the probe compound, as the polar nature of the amine group made purification of the product difficult.



Scheme 4.3: Synthetic scheme for deprotected alkyl amine, **4.3**. Reagents and conditions: a) 4M HCl in dioxane, 0°C → rt, 2h. b) NaOH (aq.).

4.3.2 Validation of anti-proliferative activity of PMF with a Boc protected amine linker

Results of the SAR study, discussed in Chapter 2, indicated that modification of PMF in the 3'-position resulted in the best retention of biological activity, compared to modifications of the 4', 5 and 7-methoxy groups, therefore this position was selected as the attachment point for the photoreactive probe. To ensure that further modification of the structure in this location with a larger group retained activity, the anti-proliferative activity of **4.2** was evaluated. This compound also showed improved aqueous solubility characteristics when compared to the previously tested flavones, potentially due to a disruption of the planarity of the molecule when introducing the linker group, this could disrupt crystallisation and result in increased solubility. Lewin *et al.* reported improved aqueous solubility levels when modifying flavone glycosides by bromination and chlorination in the 3-position.[179]

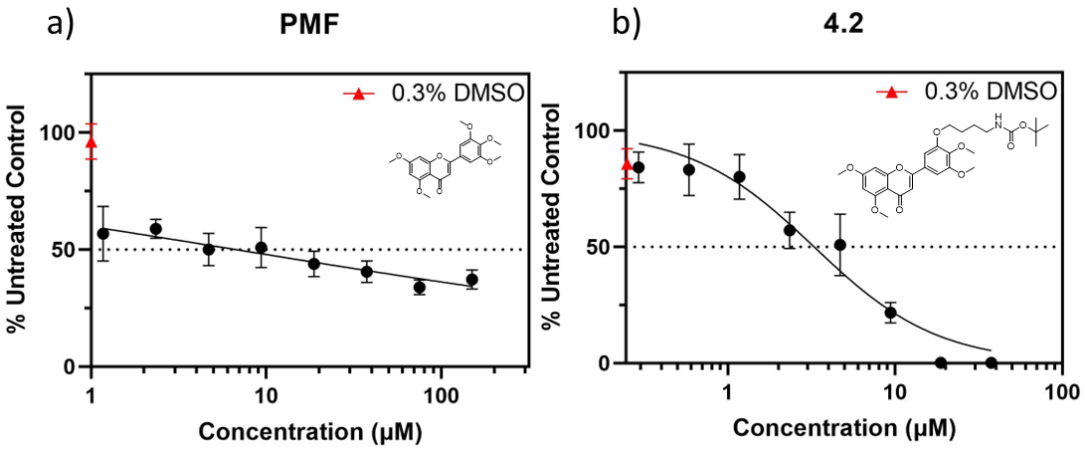


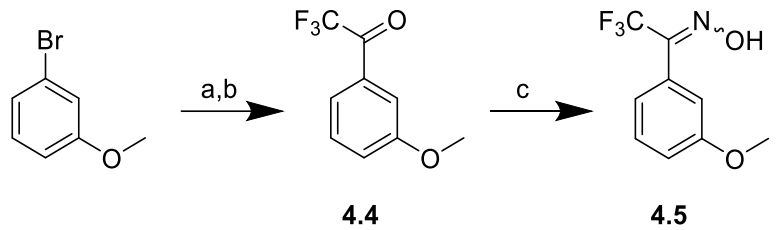
Figure 4.8: Representative SRB graphs showing anti-proliferative activity of the modified flavone, **4.2**, (b) compared to a representative graph of PMF activity (a). Each data point is produced from the mean of at least 3 data points and error bars of 1 standard deviation above and below the mean are included. Concentration range: a) 150 μM – 1.17 μM; b) 37.5 μM – 0.29 μM; compound exposure time: 96 h; cell density: 400 cells/well. The red diamond indicates the mean cell viability of cells treated with only 0.3% DMSO vehicle control. Data presented here is representative of n =3 biological repeats.

The anti-proliferative activity of the modified flavone was retained, validating the SAR study. The newly synthesised compound, **4.2**, was shown to be more potent than both PMF and the **3'-OBn** analogue, possibly due to the enhanced aqueous solubility characteristics previously discussed. The new compound resulted in an average GI_{50} value of 4 μM, compared to average GI_{50} values of 25 and 7 μM for PMF and the **3'-OBn** analogue respectively (Figure 4.8). Complete data for this experiment can be found in Appendix 8.25. The decision was made not to conduct an anti-proliferative study of the free amine compound, **4.3**, due to the difficulty in purification of the crude mixture which would hinder the reliability of the assay. It was also suspected that there could be interference from the free amine, which would not be present in the final probe molecule, with the biological activity of the compound as it could potentially become charged in solution and cause a change in *in vitro* activity compared to the uncharged parent compound. From this study it was confirmed that attachment of

the probe to the flavone in this position would be favourable to retain biological activity and the synthesis of the probe was continued at this location.

4.3.3 Synthesis of TPD photoaffinity species and attachment of alkyne reporter group

Synthesis of the TPD component of the probe followed the typically used route of conversion of a phenyl trifluoromethyl ketone to an oxime intermediate, followed then by a tosylation. From the tosyl intermediate, a diaziridine is then formed, which can then be converted to the desired diazirine. The synthetic route followed in this work was carried out mostly in known conditions. The first step involved the conversion of 3-bromoanisole to a trifluoromethylphenyl ketone and then the synthesis of an oxime from this (Scheme 4.4).



4.4 **4.5**

Scheme 4.4: Synthesis of a trifluoromethyl phenyl oxime intermediate from 3-bromoanisole. Reagents and conditions: a) *n*-BuLi, THF, -84 °C, 2h. b) Methyl trifluoroacetate, THF, -84 °C, 1h. c) Hydroxylamine hydrochloride, pyridine, EtOH, 85 °C, 19 h.

The successful synthesis of both intermediates, the trifluoromethylphenyl ketone **4.4** and oxime **4.5** was verified and the compounds characterised with ^{1}H and ^{13}C as well as 2D NMR and the oxime was also verified with ESI-MS (Appendix 8.26.1-8.26.1.2). Both compounds were carried forward to the next step of synthesis without purification to preserve yield. The oxime was then tosylated with tosyl chloride, DMAP, and triethylamine in DCM, to give **4.6**. This compound was purified from the crude mixture by recrystallisation from ethanol and characterised by ^{1}H, ^{13}C and 2D NMR as well as ESI-MS (Figure 4.9, Appendix 8.26.3).

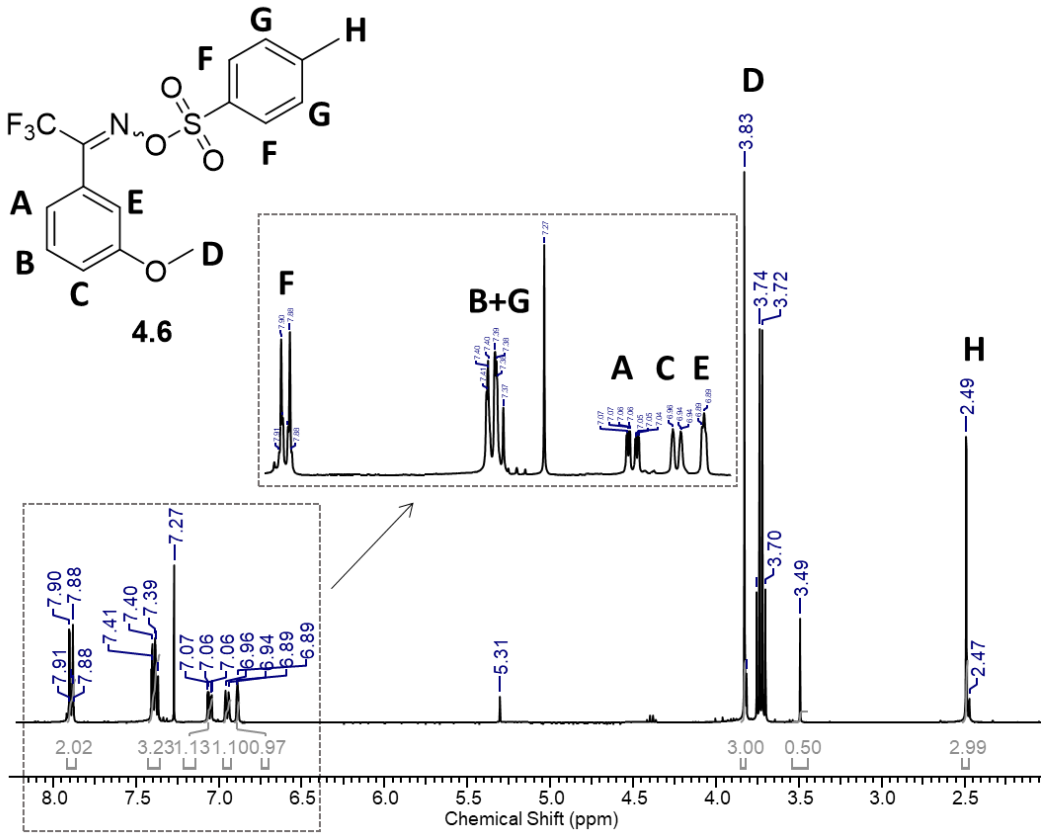
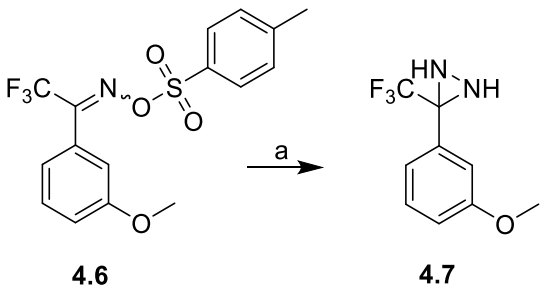


Figure 4.9: Assignments of ^{1}H NMR spectrum of 2,2,2-trifluoro-1-(3-methoxyphenyl)ethan-1-one O-tosyl oxime, **4.6**.

The next step of the route required the conversion of the tosyl species to a diaziridine with the formation of a key N-N bond. This step is typically carried out using anhydrous liquid ammonia in diethyl ether at -78 °C, which affords the desired product in very high yields, for example Chandrachud *et al*. isolated a trifluoromethyl phenyl diaziridine with a yield of 95%.[205] However, use of liquid ammonia as a reagent in this synthesis carries some difficulties due to the safety hazards previously mentioned. Work within the Serpell research group carried out by Dr Geraud Sansom, resulted in an improved method for this step (Scheme 4.5).

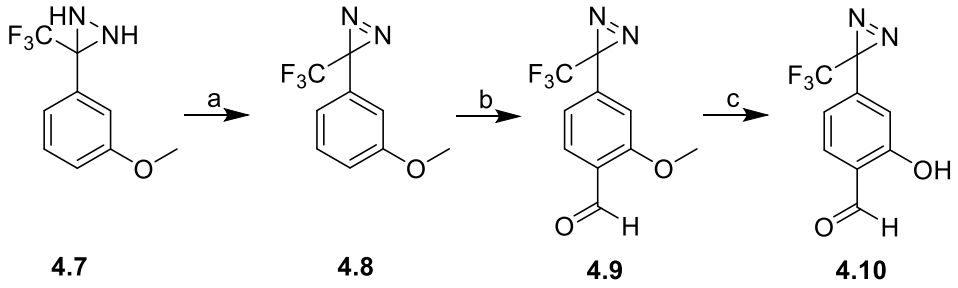


4.6 **4.7**

Scheme 4.5: Synthesis of 3-(3-methoxyphenyl)-3-(trifluoromethyl)diaziridine, **4.7**. Reagents and conditions: a) 7N NH_3 in methanol, 0 °C → rt, 72 h.

The conditions involved dissolving the pure tosylate starting material, **4.6**, in excess methanolic ammonia at 0 °C and allowing the mixture to warm to rt and stir for 72 h. Monitoring of the reaction progress with TLC showed formation of a product spot which was isolated after column chromatography purification of the crude mixture. Verification of the desired synthetic product was achieved using ^{1}H, ^{13}C and 2D NMR, which showed consistent results with those in literature reports of this intermediate (Appendix 8.27).[206] This method resulted in much lower yields of 27%, however the relative simplicity of the method made it a favourable choice for this step.

Synthesis of the diazirine was then achieved using silver oxide in diethyl ether and the crude mixture of this reaction was used directly in the next steps required to allow attachment of the TPD species into the final probe structure. The *meta* methoxy group of the structure was demethylated to allow attachment of a PEG-alkyne reporter handle and an aldehyde was installed in the *para* position to allow attachment to the targeting flavone structure.



Scheme 4.6: Synthetic route to obtain 2-hydroxy-4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzaldehyde. Reagents and conditions: a) Silver oxide, dry Et_2O, rt, 24 h. b) Titanium chloride, dichloromethyl methyl ether, dry DCM, 0 °C → rt, 2 h. c) Boron tribromide, dry DCM, rt, 2 h.

The final product **4.10**, was purified by column chromatography and characterised by ^{1}H and ^{13}C as well as 2D NMR (Figure 4.10, Appendix 8.28). The ^{1}H NMR spectrum of **4.10** matches the assignments for this structure shown by Mayer *et al.* validating that the desired compound had been synthesised successfully.[202] The synthesis following the route detailed above to obtain **4.10** was successful, however the yields obtained for the final compound were much lower than required to proceed with synthesis of the final probe compound, at only 9% over the final 3 steps. In the interest of time and the necessity to move the product onwards, the decision was made to source **4.10** commercially for the next steps of the synthesis.

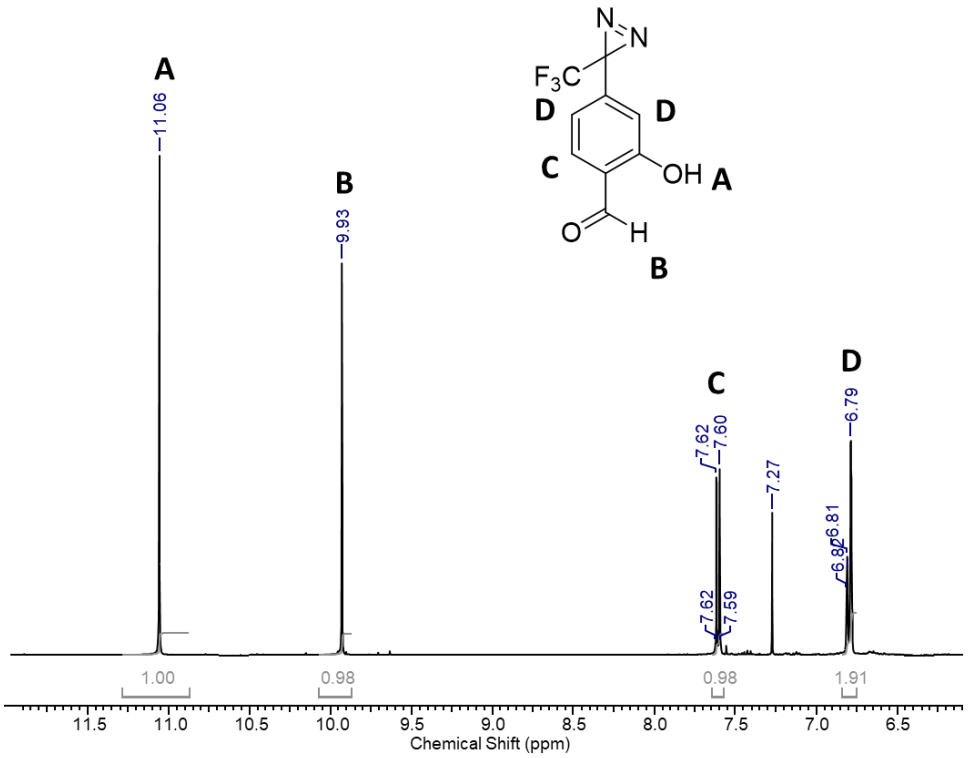
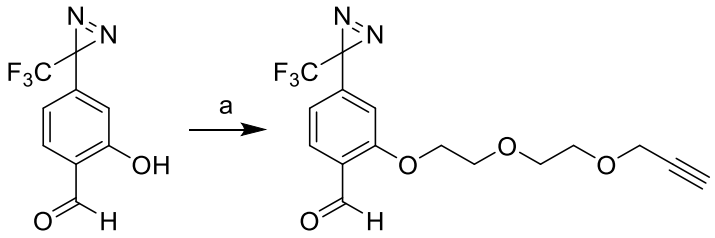


Figure 4.10: Assignments of ^{1}H NMR spectrum of 2-hydroxy-4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzaldehyde.

A PEG-alkyne reporter handle was then attached to the TPD species using an alkylation step, this reporter handle allows the probe to be bound to a reporter group such as biotin using click chemistry methods. Propargyl-PEG bromide obtained commercially was incorporated into the probe structure by refluxing with **4.10** in the presence of catalytic tetrabutyl ammonium iodide (TBAI) and potassium carbonate (Scheme 4.7). The use of TBAI, typically used for N-alkylation, activates the alkylating agent by converting the bromine to a more reactive iodine group, this encourages the formation of the ether.



Scheme 4.7: Synthesis of 2-(2-(2-(prop-2-yn-1-yloxy)ethoxy)ethoxy)-4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzaldehyde. Reagents and conditions: a) 3-(2-(2-bromoethoxy)ethoxy)prop-1-yne, TBAI (00.1 eq), K_2CO_3, DMF, 65 °C.

The product was purified from the crude mixture using column chromatography and characterised by ^{1}H, ^{13}C and 2D NMR and ESI-MS (Appendix 8.29).

4.3.4 Synthesis of active and inactive probe compounds

The next step in the synthesis of the probe compound required oxidation of the aldehyde to a carboxylic acid; to allow the formation of an amide bond with the amine linker attached to the targeting flavone group. This was carried out in the using sodium chlorite in acidic conditions, a reaction also known as the Pinnick oxidation.[207] Compound **4.11** was treated with sodium chlorite and sulfamic acid as a scavenger in a THF/H_2O solvent mixture and stirred at rt for 4 h. The crude mixture produced from this reaction was used without purification as the ^{1}H NMR analysis showed good levels of conversion with a ratio of 0.07:1 of remaining aldehyde in solution to desired product (Figure 4.11). The product was also characterised with ^{13}C, 2D NMR and ESI-MS (Appendix 8.30).

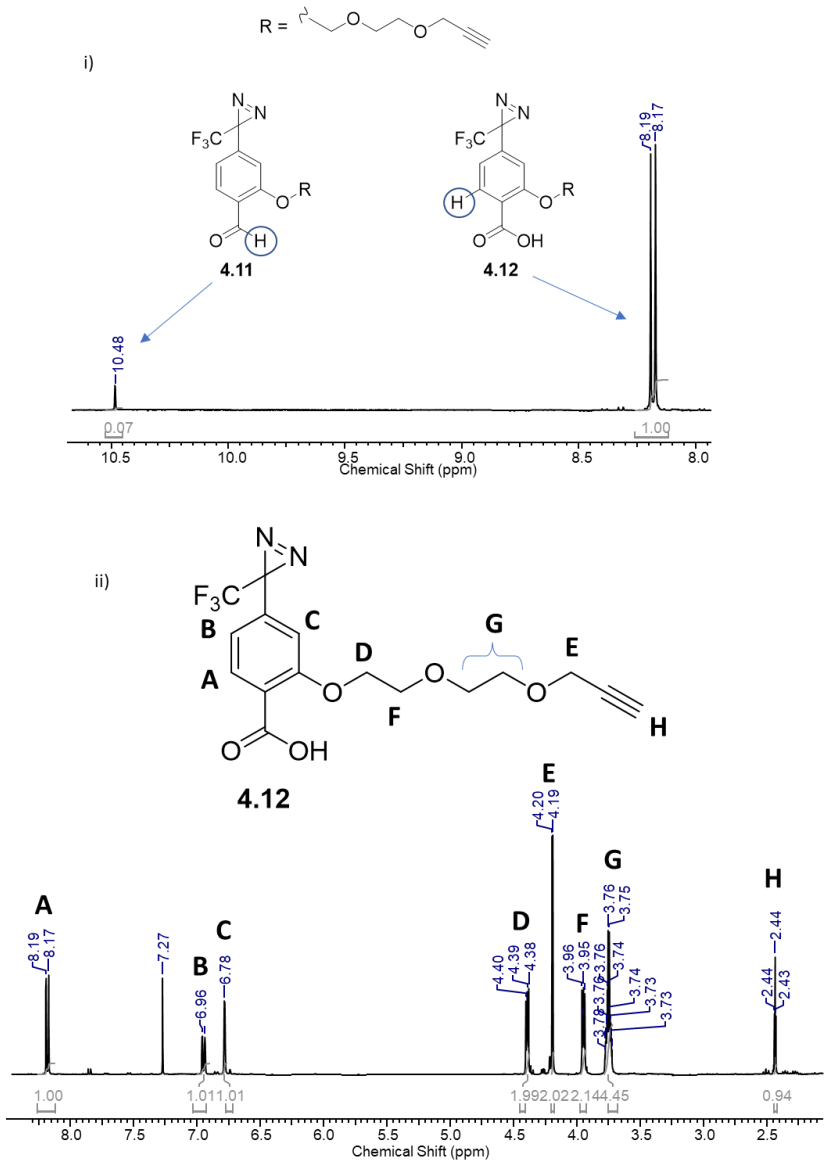


Figure 4.11: i) Comparison of the integrals of key hydrogen peaks in the aldehyde, **4.11,** and carboxylic acid, **4.12,** and show the ratio of 0.07:1. ii) Assignments of ^{1}H NMR spectrum of 2-(2-(2-(prop-2-yn-1-yloxy)ethoxy)ethoxy)-4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzoic acid, **4.12**.

Alongside the complete probe compound needed for the pull down study, an inactive version of the probe was also synthesised to aid in identifying non-specific labelling that could result from interactions of proteins with the TPD species. The inactive probe compound, **4.13**, was designed with a butyl chain in place of the targeting group, to mimic the alkyl linker group used in the active probe compound, **4.14** (Figure 4.12).

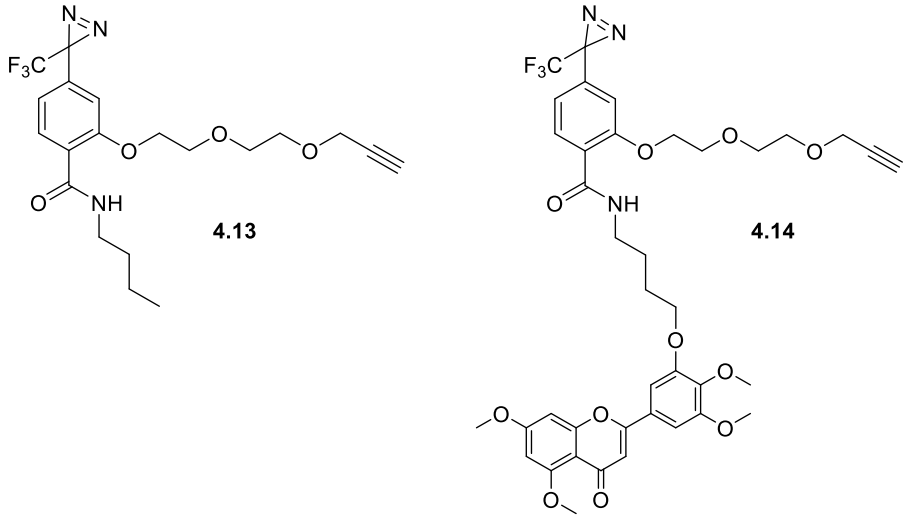
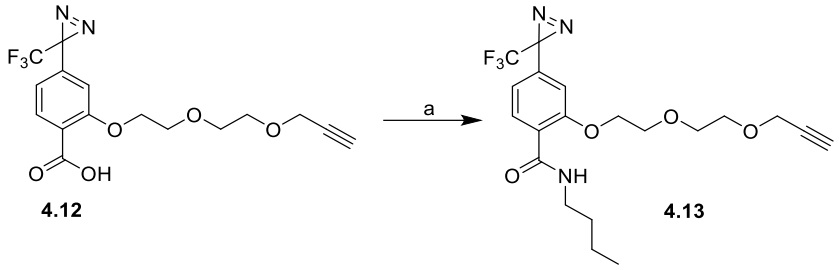


Figure 4.12: Structures of inactive probe compound, **4.13**, and active probe compound, **4.14**.

The butyl chain was attached to the TPD structure using amide coupling methods, *N*-butyl amine was added to a solution of the carboxylic acid, **4.12**, with EDC.HCl, HOBt and Et_3N in DMF and stirred at rt for 24 h (Scheme 4.8). The carbodiimide (EDC.HCl) activates the carboxylic acid by producing a reactive *O*-acylisourea which reacts with the HOBt to form an ester intermediate that is more susceptible to attack from the amine to produce the amide product.[208] The desired product was purified by two rounds of column chromatography and isolated with a 38% yield, the compound was characterised by ^1H, ^{13}C, 2D NMR and ESI-MS (Appendix 8.31).



Scheme 4.8: Synthesis of N-butyl-2-(2-(2-(prop-2-yn-1-yloxy)ethoxy)ethoxy)-4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzamide, **4.13**. Reagents and conditions: a) N-butylamine, EDC.HCl, HOBt, Et_3N, DMF, 0 °C → rt, 24 h.

The active probe compound, **4.14**, was synthesised in the same conditions as the inactive probe, however lower equivalents of the amine flavone, **4.3**, were used due to lower availability of this

reagent. The probe compound was purified from the crude mixture using column chromatography with a 24% yield. The compound was characterised by ^{1}H, ^{13}C, 2D NMR and HRMS due to the novel nature of the compound (Figure 4.13, Appendix 8.32).

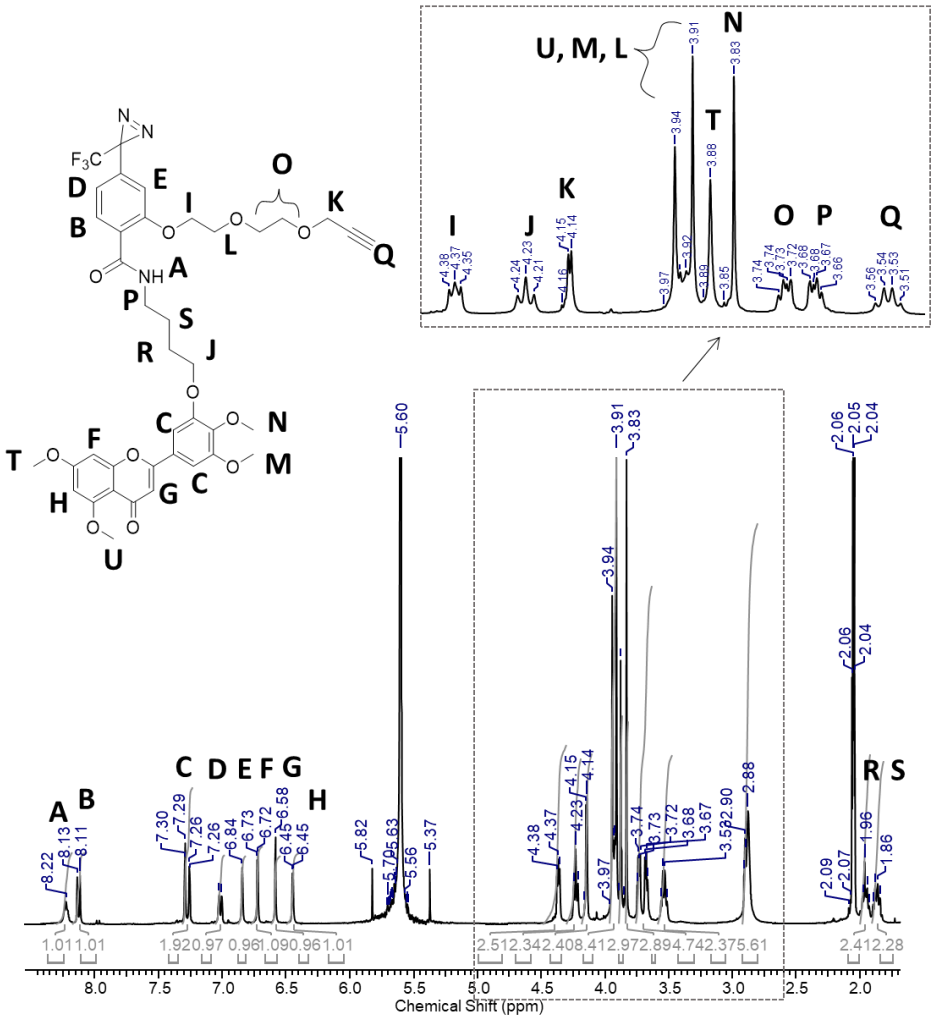


Figure 4.13: Assignments of ^{1}H NMR spectrum for N-(4-(5-(5,7-dimethoxy-4-oxo-4H-chromen-2-yl)-2,3-dimethoxyphenoxy)butyl)-2-(2-(2-(prop-2-yn-1-yloxy)ethoxy)ethoxy)-4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzamide, **4.14**.

4.3.5 Evaluation of anti-proliferative activity of active and inactive probe compounds

To verify that the polymethoxyflavone targeting group compound retained its anti-proliferative activity after modification with the TPD and PEG reporter handle, SRB assays were conducted. The

assays were also used to ensure that the inactive probe molecule did not have any inherent anti-proliferative activity (Figure 4.14).

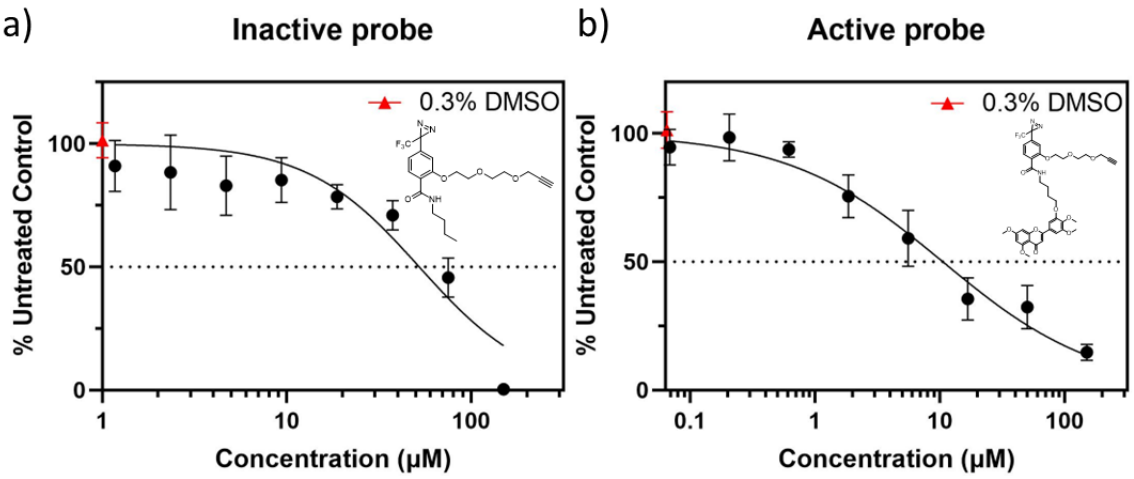


Figure 4.14: a) Representative SRB graph showing anti-proliferative activity of the inactive probe, **4.13**. b) Representative SRB graph showing anti-proliferative activity of the active probe, **4.14**. Each data point is produced from the mean of at least 3 data points and error bars of 1 standard deviation above and below the mean are included. Concentration range: a) 150 μM – 1.17 μM; b) 150 μM – 0.07 μM; compound exposure time: 96 h; cell density: 400 cells/well. The red diamond indicates the mean cell viability of cells treated with only 0.3% DMSO vehicle control. Data presented here is representative of n =3 biological repeats. Untreated control cells were exposed to media only.

The assays validated the activity of the active probe after modification and showed an average GI_{50} value of 8 μM. The inactive probe showed some anti-proliferative activity, likely due to the photoactive diazirine component which could potentially react with cellular components, however the average GI_{50} value was found to be almost an order of magnitude higher than the active probe at 62 μM. Complete data for these assays can be found in Appendix 8.33.

4.3.6 Determination of UV irradiation timescale required for diazirine decomposition

The decomposition of the TPD species after irradiation with 365 nm UV light was monitored using UV-Vis spectroscopy. The wavelength region between 330 – 420 nm, contains the absorbance peak

corresponding to diazirine, and on irradiation this peak is lost (Figure 4.15). The data presented for the 0 second timepoint was not used as the maximum absorbance, as it fell below that of the 2 and 4 second time points, it is suspected that there may have been a change in structure after initial exposure to the UV light as this was seen for all three repeat experiments.

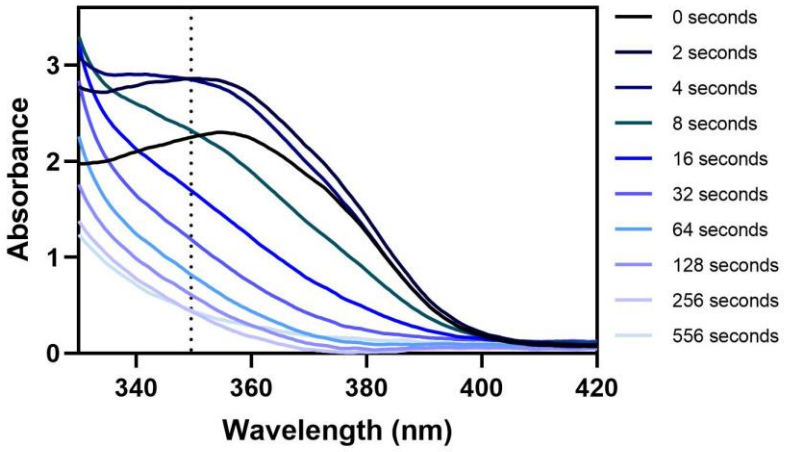


Figure 4.15: Average UV-Vis absorbance of the inactive probe compound, **4.13**, in the region of 330 – 420 nm after irradiation with 365 nm UV light for extended periods of time in seconds.

The maximum of the peak in this region was determined to be 349.5 nm, after each irradiation timepoint the absorbance at this wavelength was recorded and averaged for the three repeats. This absorbance was plotted against time to show the exponential decay of the diazirine over time. From these values, percentage conversion as a fraction of the observed absorbance at the two second timepoint was calculated. The percentage conversion against time shows rapid conversion over the initial 64 seconds of UV exposure and then it begins to plateau as the decomposition slows (Figure 4.16).

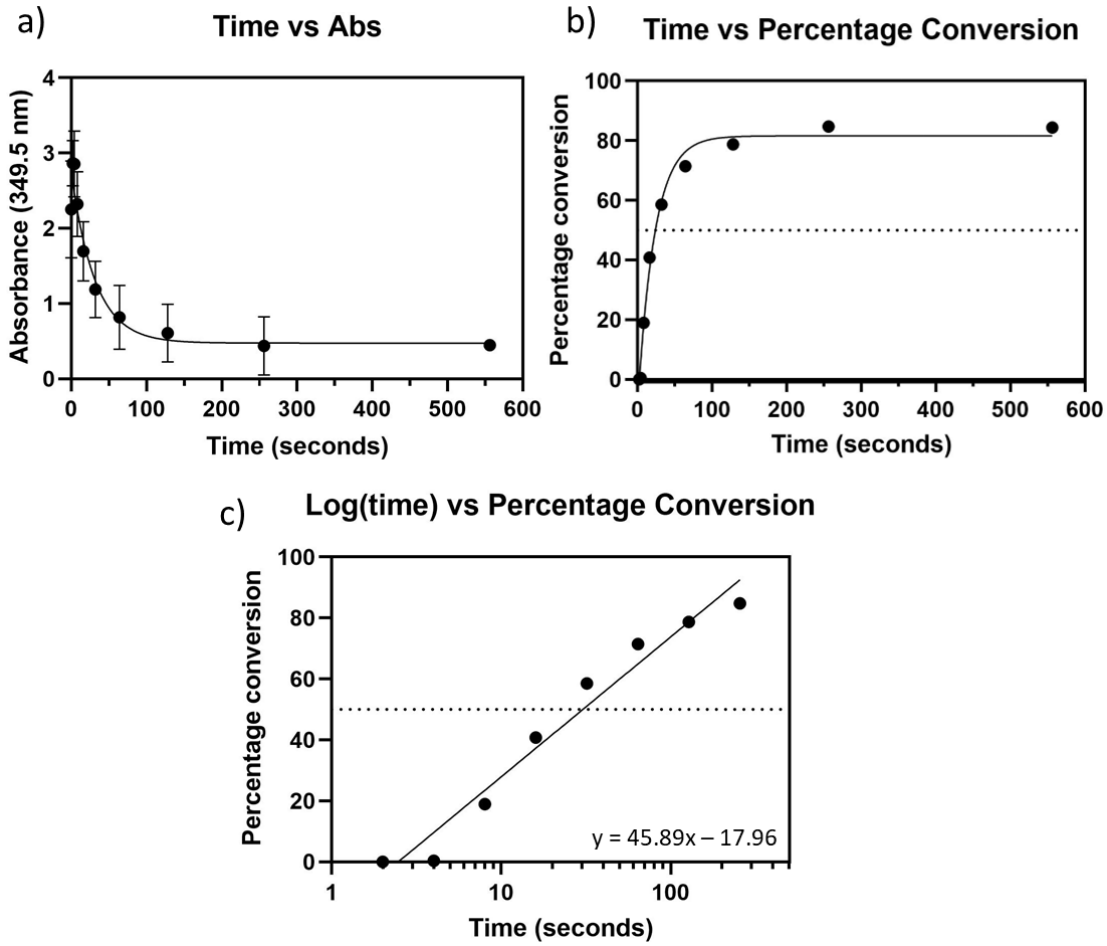


Figure 4.16: a) Graph of time vs Abs at 349.5 nm. b) Time vs percentage conversion of the diazirine species, monitored by loss of the absorbance peak at 349.5 nm, calculated as a fraction of the two second timepoint. c) Graph of log(time) vs percentage conversion of the diazirine species with a linear regression line. The absorbance value for the 556 second timepoint was removed from this plot as it was the same as the value for 256 seconds indicating a complete plateau/complete conversion.

The data in Figure 4.16c can be used to find the timepoint at which 50% conversion occurs which was calculated to be 30 seconds. Figure 4.16b shows however, that complete conversion can be classed as having occurred before the 256 second timepoint as the absorbance value did not decrease after this. From this an irradiation time of 5 minutes (300 seconds) was taken as sufficient to ensure complete conversion of the diazirine.

4.4 Conclusion

This Chapter presented the successful attachment of a butyl amine linker to the 3'-position of the polymethoxyflavone targeting group. After modification of the flavone with a Boc protected version of this amine, the anti-proliferative activity was evaluated and found to be maintained, therefore synthesis progressed as the 3'-position was validated as a suitable modification point. The photoactive TPD component of the probe was then successfully synthesised, however the yield obtained was lower than required to progress to the probe synthesis and subsequent pull down study, so further supplies of the desired compound were sought commercially. The active probe compound was then synthesised and its anti-proliferative activity evaluated: it was shown to have maintained its activity after modification. An inactive probe molecule was also synthesised, that consisted of the TPD species and the PEG-alkyl reporter handle, but in place of the flavone targeting group a butyl chain was attached; this showed much lower anti-proliferative activity and therefore could be used as a control probe in the pull down study.

The timescale of UV irradiation required to fully activate the diazirine component of the tag was also determined and conversion was found to plateau after the 256 seconds timepoint, suggesting that irradiation of the sample for 300 seconds would be sufficient to ensure total conversion.

5 Analysis of proteins interacting with the targeting compound by western blot and LCMS proteomics

5.1 Introduction

5.1.1 Potential cellular targets of PMF responsible for anticancer activity

The mechanism of action in which PMF exerts anticancer and chemopreventive properties has been previously investigated, but to date results have been inconclusive. Cai *et al.* examined the effect of three polymethoxyflavones, including PMF, on the production of prostaglandin E-2 (PGE-2), a product of the cyclooxygenase (COX) pathway.[170] Enzymatic inhibition of the enzymes in this pathway, specifically COX-2, have been shown to be a potential method of CRC prevention and treatment.[73] The work showed that PMF had superior ability to structurally similar flavones tricin and apigenin, in compromising PGE-2 generation, at a similar level of potency to the compound resveratrol, which has been previously investigated as a CRC chemopreventative.[170,209] This suggested that the mechanism of action of the flavone could be associated with the COX pathway.

Hou *et al.* presented that treatment of cisplatin resistant A549 lung cancer cells (A549/CDDP) with PMF, resulted in re-sensitisation of the cell line to cisplatin, reversing resistance. It was shown that this could be due to a reduction in the expression of Nuclear factor erythroid 2-related factor 2 (Nrf2), leading to a reduction of downstream genes, including HO-1 (Heme Oxygenase 1), NQO1 (NAD(P)H Quinone Dehydrogenase 1) and GCLC (Glutamate-Cysteine Ligase Catalytic Subunit) in a dose-dependent manner. This study suggests key interactions between PMF and Nrf2, which plays important roles in the regulation of many cytoprotective genes.[79]

PMF has also been shown to chemosensitize acute myeloid leukemia (AML) cell lines to treatment with vincristine when cell lines are over-expressing P-glycoprotein (Pgp), at similar levels of activity to verapamil, a known substrate for Pgp. Pgp is an adenosine triphosphate (ATP) dependent efflux pump that is a known membrane drug transporter and is responsible for high levels of multi-drug

resistance.[210] The study found however that PMF was not a competitive inhibitor of Pgp and therefore was exerting sensitising effects in a different manner. The compound did not exhibit the same sensitisation activity when treating AML-2/DX100 cells, which over-express the multidrug resistance-associated protein (MRP), suggesting the interactions were specifically associated with Pgp.[211]

When investigating the anti-obesogenic effects of PMF and 6,2′,4′-trimethoxyflavone (TMF), Ahmed *et al.* showed that PMF inhibited the activity of pancreatic lipase (PL) in a dose-dependent manner with a competitive inhibition mechanism. The compound also reduced mRNA expression of key adipogenic genes PPAR-γ, C/EBP-α, C/EBP-β, and SREBF1.[156]

Work on the colorectal cancer chemopreventive mechanisms of PMF by I. L. Fong, investigated the changes in protein expression of mechanistically important proteins in the adenomas and mucosa of $Apc^{Min/+}$ mice treated with PMF. Expression of minichromosome maintenance complex component 7 (MCM7) and β-catenin were significantly reduced in the adenomas studied, compared to control mice. MCM7 has been shown to play a critical role in the initiation of carcinogenesis and β-catenin is involved in regulation and coordination of cell–cell adhesion and gene transcription. The study also showed some increase in expression of protein p21 which is linked with cell cycle arrest.[172]

The studies presented above give insight into the phenotypic outcomes of some cell lines after treatment with PMF and indicates indirect interactions with a number of proteins involved in cell signalling and the cell cycle. The work in this chapter aims to identify key proteins and protein families that are physically interacting with PMF in the setting of colorectal cancer, and to investigate the pathways associated with these interacting proteins.

5.1.2 Pull down studies

Chapter 1 introduced a method of detecting protein targets of a biologically active compound using photoaffinity labelling (PAL) and pull down methods. This method utilises a three-component probe molecule consisting of a photo affinity label, a targeting compound, and a reporter tag. This probe can be exposed to cellular proteins and can then be activated with the required wavelength of light to form

covalent linkages between the probe compound and interacting proteins in the lysate. For this study a trifluoromethylphenyl diazirine (TPD) photoactive group was used to generate covalent bonds, targeted by the modified flavone structure identified in Chapter 3 using an SAR study.

5.1.2.1 PAL reporter tags

The key to visualisation and purification of proteins from a pull down study is through the choice of a reporter tag. Fluorophore reporter tags are a popular choice for photoaffinity labelling. The use of fluorescence can allow for visualisation of the small molecule and interacting proteins, this is typically used with sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis or can be employed in *in vitro* experiments to allow subcellular visualisation of the bound probe molecule to the protein of interest. The use of *in vitro* analysis can validate identified interacting proteins and give insight in to potential mechanisms of action.[96,97] There are many different types of fluorophore available for use, each can offer specific advantages, usually fluorophores are hydrophobic and have good levels of cell permeability, although some can cause problems due to photobleaching.[97,109] Fluorescent tags can offer high sensitivity as they possess a high quantum emission yield.[212] An example of the use of various reporter tags was shown by Ziang and co-workers; they identified potential target molecules of BAY R3401, a hypoglycaemic agent for the treatment of type 2 diabetes. A variety of photoaffinity probes were tested, and the crosslinked species were isolated using a biotin purification handle. This aided the research towards a potential mechanism of action for this drug candidate, which previously had not been established.[125]

A fluorescein isothiocyanate (FITC) fluorescent reporter tag (Figure 5.1) was used for the work described in this chapter. The FITC tag is one of the most widely used fluorescent probes for this type of study.[212] FITC has excitation and emission spectrum peak wavelengths of 494 nm and 518 nm respectively, allowing the labelled proteins to be visible using a gel scanner. There are some drawbacks to the use of a FITC reporter tag, such as high levels of unwanted background fluorescence which can occur due to residual unreacted FITC in the sample.[213]

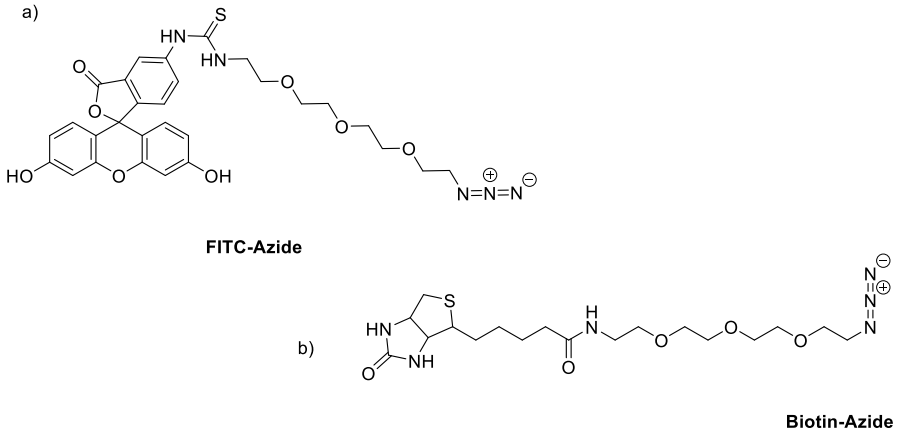


Figure 5.1: Structures of azide reporter tags used in this work. a) Fluorescein isothiocyanate (FITC) azide fluorescent reporter tag. b) Biotin-azide isolation reporter tag.

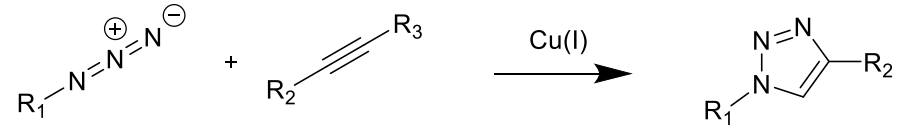
Fluorophore handles however, do not offer means of purification; a popular isolation handle with this ability is biotin. The chemical conjugation of a biotin reporter tag to a photoaffinity probe allows the isolation and enrichment of captured targets, this is possible due to the extremely high affinity (K_d = 10^{-15} M) of biotin for tetrameric protein avidin, the biotin handle can form a strong but noncovalent interaction that enables purification. The complex can withstand numerous washing steps to isolate only biotinylated samples, isolated products can then be digested for analysis via LCMS methods.[110,115] Streptavidin, a structurally similar tetrameric protein isolated from *Streptomyces avidinii,* possesses similarly high biotin-binding properties to the protein avidin. Streptavidin is often used for biotin-containing assays as it does not contain the glycoprotein portion of the protein resulting in reduced non-specific binding. Biotin also offers dual functionality as it can be used for both isolation and detection, the conjugation of a reporter molecule to the streptavidin, such as the enzyme horseradish peroxidase (HRP) can give highly accurate detection using chemiluminescence.[115]

Biotin also poses some drawbacks, since attachment of a biotin handle within a photoaffinity probe at early stages of the experimental process can result in undesired chemical properties. Incorporation of a biotin handle in a probe molecule can result in poor cell permeability and the large size can negatively affect activity of the small molecule.[114] The strong interaction between biotin and avidin can also cause difficulties in removing the isolated species from the complex for analysis.[97]

5.1.2.2 Clickable probes

Incorporation of an affinity tag such as biotin into a probe can create many complications for the optimisation of a photoaffinity labelling experiment, including problems with cell permeability, solubility and potential steric blocking. Assays that evaluate the activity of a small molecule can be unsuccessful as the reporter tags can significantly reduce bioactivity due to steric hinderance. Without accurate evaluation of the biological activity of a probe, it is impossible to assess its effectiveness for identifying target molecules.[96]

Due to this problem, a strategy of incorporating an identification/purification handle after the covalent modification of the target biomolecule can circumvent these problems. These methods often involve bio-orthogonal reactions known as 'click' reactions, one of the most used of these is an in-cell or in-lysate Cu(I) catalysed alkyne-azide cycloaddition (CuAAC), which enables the conjugation of the reporter tag at a later stage. The copper(I) catalysed reaction is favoured due to its bio-orthogonality and stability in biological conditions. Typically in these protocols, an alkyne moiety is installed into the photoaffinity probe which forms a covalent crosslink with the interacting protein, an azide moiety conjugated to the desired reporter tag is then added to the mixture with a Cu(I) catalyst which induces a controlled reaction between the alkyne and azide to form a stable triazole product (Scheme 5.1). The alkyne-azide components can be exchanged, however this has been shown to increase levels of background protein labelling.[93]



Azide component Alkyne component 1,4 cycloaddition product

Scheme 5.1: Scheme showing the reaction between azide and alkyne components of a Cu(I) catalysed alkyne-azide cycloaddition (CuAAC), also known as 'click' chemistry reaction.

Fluorescent and biotin reporter tags are commonly incorporated *via* a click reaction, mainly due to their large size and often hydrophobic nature. The use of both reporter tags, conjugated via a 'clickable'

handle, enabled the identification of potential protein-protein interactions of γ-secretase in live cells.[214]

A traditional copper(I) click reaction can be problematic when carrying out live cell analysis, as the copper can be toxic to cells. Alternatives to the copper-based reaction have therefore been developed. Strain promoted click reactions that involve azide and cyclooctyne moieties and inverse electron demand Diels-Alder (IEDDA) reactions between tetrazine and cyclopropene/*trans*-cyclooctene are favourable in these situations.[123]

5.1.3 Bradford protein assay

Prior to the analysis of binding proteins of the biologically active compound, the total protein quantification of the samples is required to equalise sample protein content to allow direct comparison of protein band intensity in visualisation techniques such as SDS-PAGE, or western blotting. The Bradford protein assay is a method of protein concentration determination in a solution by a colorimetric assay. First developed by M Bradford in 1976, the assay uses Bradford reagent (Figure 5.2), which, upon binding to proteins, undergoes a colour change from its red form to its blue form which shifts the absorption maximum of the dye from 465 nm to 595 nm[215]. The anionic form of the Coomassie brilliant blue dye binds to arginine, histidine, phenylalanine, tryptophan and tyrosine residues through Van der Waals forces and electrostatic interactions to form a protein-dye complex that stabilises the dye and allows the absorption to be measured using a 96-well plate reader.[216,217] The quantity of the anionic blue form of the dye measured in the sample can be used to calculate quantitative protein concentration. Readings of this absorption can be compared to a standard curve of known concentrations of a particular protein such as bovine serum albumin (BSA) or immunoglobulin G (IgG) to calculate the protein concentration of the unknown sample.

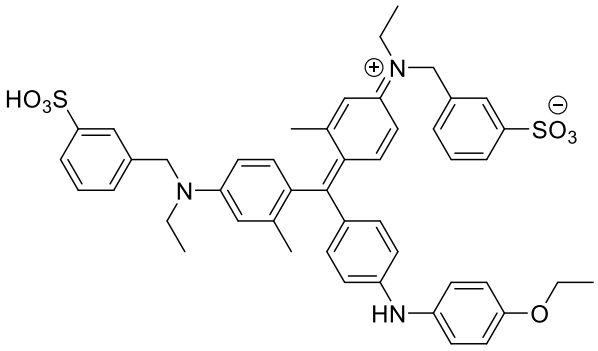


Figure 5.2: Structure of Coomassie brilliant blue G-250 protein binding dye, used for assessing protein concentration in solution.

The calculated protein concentration can then be used to dilute the protein lysate to the required protein concentration for the experiment. The method is widely used due to its sensitivity to proteins and minimal interference from solution contaminants.[215]

5.1.4 Gel electrophoresis and western blotting

SDS-PAGE is a technique used for separating proteins in a mixture according to the size of the proteins. When compared to a protein ladder of known lengths, the molecular mass of the proteins of interest can be determined and the composition of the protein mixture investigated. A protein ladder is a mixture of proteins of defined molecular weights that can be used as a reference when run alongside proteins of undetermined molecular weights. Denatured proteins, usually in the form of protein lysate, are administered to the 'top' of a polyacrylamide gel and a current is applied across the gel, resulting in the negatively charged protein molecules being attracted to the anode at the 'bottom' of the gel. Proteins with smaller molecular masses move more quickly through the gel due to less interactions of the protein with the gel matrix, resulting in separation by size. Proteins can be visualised after electrophoresis using several methods, such as staining of the gel with Coomassie blue protein stain which allows the detection of all proteins on the gel. The Coomassie stain results in the appearance of blue protein bands on the gel due to interactions described for the protein concentration determination Bradford assay.[218]

A widely used method for identification of the proteins separated using SDS-PAGE is the technique of western blotting. This technique involves the electrophoretic transfer of the proteins from the gel to a membrane, typically nitrocellulose or polyvinylidene difluoride (PVDF).[219] Specific proteins can then be detected on the membrane using many techniques. A commonly used technique involves the use of an antibody that is specific to the target protein, which has been conjugated to a detectable label, such as a fluorophore, allowing direct visualisation of the target protein, this is known as an immunoblot.[220] Biotinylated proteins, or antibodies in a secondary antibody step, can be detected using a labelled streptavidin or avidin binding partner, such as streptavidin-HRP.[221]

5.1.5 Tandem mass tags (TMT) for LCMS proteomics analysis

Quantitative liquid chromatography mass spectrometry (LCMS) proteomics for the work described in this chapter, was carried out using tandem mass tags (TMT). These tags allow labelling of samples from different treatment groups with isobaric sample labels. The samples can then be combined and analysed simultaneously with the treatment group identified after analysis.[222] Label-free approaches for proteomics experiments are possible, however these methods can be susceptible to small environmental changes which can result in small changes in quantitation being missed and require significant instrument usage. Analysis of all treatment samples at the same time is therefore generally found to be more reproducible.[222]

The tags used for TMT labelling consist of three components including a mass reporter, mass normaliser and an amine reactive group (Figure 5.3). To create the isobaric structures needed to label multiple treatment groups, each distinct tag is created with a different selection of heavy isotopes, ^{13}C and ^{15}N, in various regions of the structure.[223]

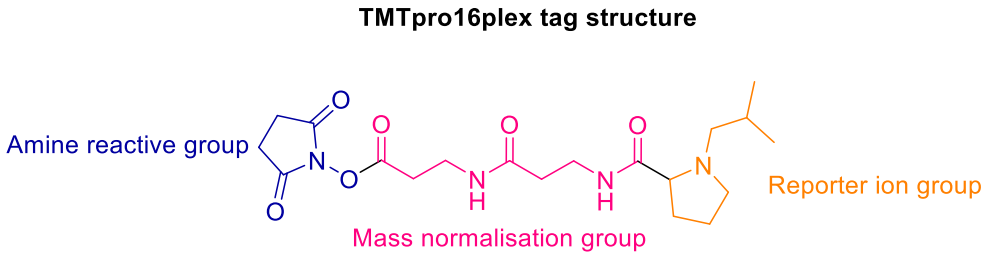


Figure 5.3: Structure of a TMTpro16plex tag used for quantitative proteomics analysis. The structure consists of three components: a mass reporter, mass normaliser and an amine reactive group.[223]

The amine reactive group allows labelling of proteins on the N-termini and on the lysine side chains. During analysis, the reporter group is cleaved from the label and therefore identification of the treatment sample in which the protein was found is possible.[223]

5.1.6 Chapter aims

Work in this Chapter aims to use pull down methods and LCMS proteomics analysis to identify potential binding proteins of 3',4',5',5,7-pentamethoxyflavone. From the structure activity relationship study carried out and detailed in Chapter 2, a probe molecule was developed that allows the formation of covalent bonds between the probe compound, **4.14**, and neighbouring biomolecules upon irradiation with UV light. This probe molecule can be mixed with cell lysate, and interacting proteins can be covalently linked to the probe molecule for visualisation and isolation. Visualisation and isolation of the proteins that are interacting with the target molecule can be achieved with western blot and pull down techniques to give insight into a potential mechanism of action of the parent molecule.

5.2 Materials and methods

Materials, instrumentation and experimental methods for cell culture and irradiation with UV light were discussed in section 3.2.

5.2.1 Materials

HEPES, free acid and DTT (dithiothreitol) were purchased from Melford biolaboratories Ltd. SDS micropellets were purchased from Fisher BioReagents. Ethylenediaminetetraacetic acid (EDTA), Streptavidin HP SpinTrap columns, sodium chloride, copper (III) sulphate pentahydrate, tris(hydroxymethyl)aminomethane (Tris), sucrose, TrisHCl, magnesium chloride, methanol, acetic acid, chloroform and glycine were purchased from Fisher Scientific. Triton™ X-100, sodium fluoride, sodium orthovanadate, bromophenol blue, bovine serum albumin (BSA), sodium ascorbate, β-glycerophosphate, Coomassie brilliant blue R stain and TWEEN® 20 were purchased from Sigma-Aldrich. Dried skimmed milk powder was from Marvel (Premier Foods group). Complete™ proteinase inhibitor cocktail tablets, PVDF western blotting membranes and Nonidet-P40 substitute (NP40) were obtained from Roche. Phosphate buffered saline (PBS) was purchased from Gibco. Clarity Western ECL substrate reagents, streptavidin-HRP and Bio-Rad protein assay dye reagent concentrate were purchased from Bio-Rad. 1-Step™ Transfer buffer, PageRuler™ prestained protein ladder, Pierce™ streptavidin magnetic beads, triethylammonium bicarbonate buffer (TEAB), Pierce™ silver stain kit and extra thick western blotting filter papers were purchased from ThermoScientific. Tris(benzyltriazolylmethyl)amine (THPTA) was purchased from Combi-blocks. NuPAGE™ 4-12% Bis-Tris gels 1.0 x 12 well precast gels and NuPAGE™ MOPS SDS running buffer (20x) Invitrogen™ and BenchMark™ Fluorescent Protein Standard were purchased from Invitrogen. FITC-azide was synthesised by colleagues in the Serpell research group. Azido-PEG3-biotin conjugate was purchased from Alfa Aesar.

5.2.2 Instrumentation

5.2.2.1 Centrifuges

Falcon tubes (≤ 15 mL) were centrifuged using a MSE Sanyo Centaur 2 MSE Centrifuge. Falcon tubes (50 mL) were centrifuged using a ThermoScientific Heraeus Megafuge 16R centrifuge. Eppendorf tubes (≤ 2 mL) were centrifuged using a Sigma 1-14K centrifuge.

5.2.2.2 SDS-PAGE and western blot equipment

SDS-PAGE was run using a XCELL Surelock gel box and Bio-Rad PowerPac 200. Samples were heated using a Stuart Scientific Block heater. Gels were imaged using a GeneSys GBOX Chemi XX6 gel imager with GeneSys V1.5.2.0 software. A Stuart gyro-rocker STR9 was used to agitate gels and western blots. Proteins were transferred to western blot membranes using a Bio-Rad Trans-Blot® Turbo™ Transfer system.

5.2.2.3 Determination of solution pH

Determination of solution pH was achieved using an Orion Star A111 pH meter.

5.2.3 Experimental methods

5.2.3.1 Preparation of cell pellets

Cells were seeded into 3 x T75 cell culture flasks and grown to 80% confluency. The cells were then trypsinised and counted using methods discussed in section 3.2.3.1. Stock solutions of cells in IMDM + 10% FBS were then divided into falcon tubes to give 20,000,000 cells per tube. The tubes were centrifuged at 106 xg for 5 minutes. The media was then carefully removed, and the cell pellet resuspended in 1 mL ice-cold PBS and transferred to pre-cooled (4 °C) 1.5 mL Eppendorf tubes. The tubes were centrifuged at 1,000 xg for 5 minutes at 4 °C. The PBS was then removed, and the pellets resuspended in a further 1 mL ice-cold PBS as a wash step. The tubes were centrifuged at 3,671 rpm for 5 minutes. The wash step was repeated for a second time. The PBS was then carefully removed to leave the cell pellet which was snap frozen with dry ice. Pellets were then stored at -80 °C until use.

5.2.3.2 Preparation of cell lysates and determination of protein concentration

Cell pellets were removed from the freezer and allowed to defrost at rt, 100 µL of ice-cold lysis buffer (Appendix 8.34) was added to each pellet and the tube placed on ice for 30 minutes. Eppendorf tubes were centrifuged at 14,000 rpm for 10 minutes at 4 °C. The lysate supernatant (now called cell lysate) was collected, and the pellet discarded. A sample of this cell lysate was taken from the stock and used

for protein concentration determination using a Bradford Assay. The cell lysate was diluted to 3 mg/mL using lysis buffer and stored at – 80 °C until use.

5.2.3.2.1 Bradford protein assay

A protein concentration standard curve was generated using BSA at concentrations of 0.05, 0.1, 0.2, 0.3, 0.4 and 0.5 mg/mL diluted in distilled water from a 1 mg/mL stock solution. These standard solutions (10 µL), as well as 10 µL of distilled water as a blank, were each added to wells of a flat bottom 96-well plate in triplicate with 200 µL Bio-Rad protein assay dye, which had been diluted 1:4 with distilled water. Dilutions of the cell lysate in distilled water (1/10, 1/20, 1/40, 1/80, 1/160 etc.) were also added to the 96-well plate 10 µL/well in triplicate along with 200 µL Bio-Rad protein assay dye (1:4 dilution in distilled water) per well. The 96-well plate was placed on to a plate shaker for 10 minutes. The absorbance of the protein standards and diluted cell lysate samples were recorded at 595 nm using a plate reader. The standard curve of BSA concentrations was generated in Excel using the mean absorbance values of each standard, minus the blank, and a linear regression line fitted. The equation of the regression line was used to calculate x-axis values for the cell lysate dilutions, which were added to the standard curve graph. Diluted lysate readings that fell within the points of the standard curve were then averaged and the undiluted protein concentration calculated. A new standard curve was generated for each lysate protein concentration evaluated.

5.2.3.3 Preclearing of cell lysate to remove biotinylated proteins

Prior to these steps, the centrifuge was cooled to 4°C. Cell lysate was precleared before use with streptavidin columns to remove endogenously biotinylated proteins. Initially the Streptavidin HP SpinTrap columns were equilibrated with lysis buffer, by centrifuging each column with 400 µL lysis buffer for 1 minute at 13,000 × g, this was repeated twice. Cell lysate (500 µL per column) was then added and the columns placed on a plate rocker for 1 h on ice. The bottom of the column was then removed and the column placed into an Eppendorf tube. With the caps loosened the tubes were centrifuged for 1 minute at 13,000 x g to collect the precleared lysate. Protein concentrations were determined with a Bradford assay before the preclearing step.

5.2.3.4 *Dilution of probe and competition compound stocks*

Stocks of inactive and active probe compounds (**4.13** and **4.14**) as well as compound **4.2**, which was used for competition studies, were prepared in DMSO as stock solutions at the concentrations shown below (Table 5.1).

Table 5.1: Concentrations of compounds required for pull down study diluted in DMSO.

Compound	Concentration
Active probe (**4.14**)	2 mM
Inactive probe (**4.13**)	2 mM
Competition compound (**4.2**)	50 mM

5.2.3.5 *Methods for the pull down study*

5.2.3.5.1 Incubation of competition compound and active and inactive probe compounds

Lysate (3 mg/mL) was defrosted and aliquoted into the required volumes in 1.5 mL Eppendorf tubes on ice. For samples which required pre-incubation with the competition compound, **4.2**, this was added to the lysates at the required volume from the stock solution shown in Table 5.1, and the samples were placed on ice for 1h. The inactive or active probe compound was then added in the correct volume required in DMSO from the stock solution shown in Table 5.1; the samples were then placed on ice in the dark for 1 h.

5.2.3.5.2 Irradiation of the samples with UV light

After incubation, the samples were placed in an ice water bath and irradiated with UV light (365 nm) from a torch placed 4 cm from the closed Eppendorf tube. Each tube was irradiated for 5 minutes for optimal crosslinking, determined by evaluation of crosslinking time requirements discussed in Chapter 4.

5.2.3.5.3 Methods for carrying out the click chemistry reaction with the alkyne reporter handle

Click reactions were carried out on two scales, depending on the volume of cell lysate used for the experiment (Table 5.2). After the crosslinking step, PBS was added to the crosslinked lysate with the

required azide reporter tag (Biotin-N_3 or FITC-N_3) and the tube briefly vortexed. Tris(benzyltriazolylmethyl)amine (THPTA) (100 mM in distilled water) and $CuSO_4$ (20 mM in distilled water) were added and the solution vortexed briefly. Sodium ascorbate (300 mM in distilled water) was then added to initiate the click reaction and the solution vortexed again before incubating at rt for 3 h while protected from ambient light.

Table 5.2: Volumes of reagents used for different scales of click reaction.

Reagent	**100 µL lysate**	**500 µL lysate**
PBS	200 µL	2000 µL
Azide (reporter tag)	20 µL	100 µL
THPTA	20 µL	100 µL
$CuSO_4$	20 µL	100 µL
Sodium ascorbate	20 µL	100 µL

After the 3 h incubation, proteins were extracted using a protein precipitation method. To each sample methanol, chloroform and distilled 18.2 MΩ ultrapure water was added at the required volume (Table 5.3), and the solution vortexed briefly after each addition. The samples were then centrifuged at 5,580 × g for 5 minutes at rt and the upper aqueous layer carefully removed. More methanol was then added and the solution vortexed briefly. The samples were then centrifuged at 5,580 × g for 5 minutes at rt to pellet the protein and the supernatant removed and discarded. Methanol was then added for a wash step and the samples were centrifuged and the supernatant removed and discarded. More methanol was then added, and the wash step repeated. The sample tube lids were then opened, and the protein pellets allowed to air dry for a minimum of 15 minutes to remove residual methanol. Protein pellets were then used directly for the following step or stored at -80 °C until use. Volumes of solvents were varied depending on the scale of reaction being carried out (Table 5.3).

Table 5.3: Volumes of reagents used for different scales of protein precipitation.

Reagent	**100 µL lysate**	**500 µL lysate**
Methanol (1st step)	600 µL	3,000 µL
Chloroform	150 µL	750 µL
18.2 MΩ ultrapure water	400 µL	2,000 µL
Methanol (2nd step)	450 µL	2,250 µL
Methanol (wash step)	450 µL	2,250 µL

5.2.3.5.4 Isolation of biotinylated proteins using streptavidin beads

Proteins covalently bound to the active or inactive probe and then exposed to the biotin azide click chemistry step were isolated from the proteins pellets as described here. Pellets were dissolved in HEPES ((4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) buffer (Appendix 8.35) containing 1% SDS (sodium dodecyl sulfate). Pierce™ streptavidin magnetic beads were vortexed and added to an Eppendorf tube. A magnet was used to collect the beads to the side of the tube and the storage solution was removed and replaced with TBS-T (Tris-buffered saline with 0.1% Tween) (Appendix 8.36) washing buffer and the beads re-dispersed. The beads were then collected using the magnet and the TBS-T washing buffer removed and the protein pellet dissolved in HEPES buffer added to the beads. The samples were placed on a plate shaker for 1 h at rt. The magnet was then used to collect the beads to the side of the tube and the supernatant removed, discarded, and replaced with TBS-T washing buffer. The beads were washed with TBS-T washing buffer twice more. Samples which were used for LCMS proteomics analysis were then washed further with TEAB (Triethylammonium bicarbonate, 1.0 M, pH 8.5±0.1) buffer 3 times. At this step a sample from each condition was removed for analysis with SDS-PAGE and western blot, remaining sample was stored at -20 °C for further use or LCMS proteomics. Volumes of reagents were varied depending on the scale of reaction being carried out (Table 5.4).

Table 5.4: Volumes of reagents used for different scales of streptavidin bead pull down

Reagent	**100 µL lysate**	**500 µL lysate**
HEPES buffer	150 µL	1,500 µL
Pierce™ streptavidin beads	25 µL	125 µL
TBST (1st step)	500 µL	1,500 µL
TBS-T (washing steps)	150 µL	750 µL
TEAB buffer	-	750 µL

5.2.3.6 SDS-PAGE analysis

5.2.3.6.1 FITC-N_3 labelled protein analysis

Samples that had been bound to a FITC-azide reporter tag were used directly for the SDS-PAGE analysis and were dissolved in 60 µL 1 x sample buffer (Appendix 8.37) and heated in a heating block for 10

minutes at 85 °C. Samples were then sonicated to break up residual DNA and then aliquoted into gel wells. Samples that were not used directly were stored at -20 °C.

5.2.3.6.2 Biotin-N_3 labelled protein analysis

Samples bound to a biotin-azide reporter tag and isolated using streptavidin beads were then dissolved in 50 µL 1 x sample buffer (Appendix 8.37) and heated in a heating block for 10 minutes at 85 °C. A magnet was used to collect the beads to the side of the tube and the supernatant removed for analysis, the beads were then discarded. Samples that were not used directly were stored at -20 °C.

5.2.3.6.3 SDS-PAGE procedure

Pre-cast sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) gel cassettes were removed from packaging after storage at 4 °C, the wells were washed with 1 x MOPS running buffer (Diluted from 20x MOPS running buffer) three times and then filled with running buffer. Cassettes were fitted into the buffer core of the gel box and the gel tension wedge locked. Running buffer (200 mL) was added to the upper buffer chamber of the gel tank. Frozen samples were heated with a heating block at 90°C for 1 minute to denature proteins. Samples dissolved in 1 x sample buffer were added to the wells, alongside the required protein ladder in well 1 (20 µL). Wells that did not have sample were filled with 20 µL 1 x sample buffer. After addition of the samples to the wells of the gel cassette, 600 mL of 1 x MOPS running buffer was added to the lower buffer chamber and the lid securely fixed. Electrophoresis was carried out with a Bio-rad PowerPac 200 at constant 200 V for 50 minutes. Gels were then removed from the cassettes and imaged, stained or carried forward for western blotting.

5.2.3.6.4 Fluorescent imaging of FITC bound proteins

Samples that had been bound to a fluorescent reporter tag (FITC-N_3) were imaged after SDS-PAGE using fluorescent imaging. A fluorescent ladder was used when running these gels to aid visualisation of the protein molecular weights. After removal from the cassette, completed gels were placed into distilled water and imaged using the gel imager using a FITC setting.

5.2.3.6.5 Coomassie staining

For Coomassie staining, gels were rinsed with distilled water and then submerged in Coomassie stain (Appendix 8.38) for 1h on a gel rocker. The gel was then placed into de-stain solution for 1 h. The de-stain solution was then replaced, and the gel rocked for 1 h further. The de-stain was then replaced again, and the gel photographed.

5.2.3.6.6 Silver staining

Gels were stained using a Pierce™ silver stain kit. Gels were rinsed with 18.2 MΩ ultrapure water for 5 minutes, and then the water replaced for another 5 minutes. The gel was then fixed with 30% ethanol/10% acetic acid/60% 18.2 MΩ ultrapure water solution for 15 minutes, the solution was then replaced, and the gel fixed for another 15 minutes. The gel was then washed twice for 5 minutes in 10% ethanol/90% 18.2 MΩ ultrapure water, followed by a wash in 18.2 MΩ ultrapure water for 5 minutes. The gel was then incubated in sensitizer working solution for 1 minute then rinsed twice with deionised water. The gel was then placed in silver stain working solution for 20 minutes then washed twice with deionised water for 20 seconds. Developer working solution was then added until bands reached the desired intensity and then replaced and washed briefly with 5% acetic acid stop solution. The stop solution was then replaced and left for 10 minutes at which point the gel was photographed (Appendix 8.39).

5.2.3.7 Western blotting procedure

Samples that had been bound to a biotin reporter tag (Biotin-N_3) required western blotting for visualisation of the biotinylated proteins. After SDS-PAGE, proteins were transferred to 0.2 µm pore Immobilon-P polyvinylidene fluoride (PVDF) membrane using a Bio-Rad Trans-Blot® Turbo™ transfer system. The gel was removed from the cassette and placed in to 1-Step™ transfer buffer. Three thick western blotting filter papers were soaked in transfer buffer and the PVDF membrane was activated in methanol. Two filter papers were placed below the membrane and the gel was placed on top and another filter paper was placed above, and the 'sandwich' placed into the transfer machine. The transfer was carried out using the Bio-Rad setting for 30 minutes and then the membrane was placed

in to 20 mL of blocking buffer (5% milk powder in TBS-T). Gels were stained with Coomassie staining after transfer to ensure all proteins had transferred. The membrane was rocked on a gel rocker in blocking buffer for 1 h and then placed in to 20 mL blocking buffer with 20 µL of streptavidin-HRP conjugate and rocked at 4 °C for 12 h. The membrane was then washed 5 x for 15 minutes with TBS-T. The membrane was then placed in to ECL chemiluminescence reagents (1:1, 7 mL) for 5 minutes and then imaged using the gel imager using an ECL setting.

5.2.3.8 Proteomics LCMS analysis conducted at the Institute of Cancer Research, London

Samples for proteomics analysis were produced in triplicate under three conditions. Treatment of protein lysate was carried out using the same methods used for the preparation of western blot samples previously detailed, but at a larger scale (500 µL). Protein lysate was incubated with DMSO, the active probe compound ('active tag') or pre-incubated with competition compound, **4.2**, and then exposed to the active probe compound ('active tag + competition'). The experimental protocol was then completed to isolate biotinylated proteins covalently linked to the active probe compound bound to magnetic streptavidin beads ('active tag' and 'active tag + competition' samples) or to isolate proteins bound non-specifically to the streptavidin beads ('DMSO' samples). The washed beads were then frozen at -20 °C and transported on dry ice. These samples were then processed and analysed by Dr Fernando Jr Sialana at the Institute of Cancer Research, London, UK (ICR) for proteomics analysis.

5.2.3.8.1 Methods of sample preparation and analysis for proteomics used by ICR.

Complete methods for the proteomics analysis carried out by Dr Fernando J. Sialana at the ICR can be found in Appendix 8.40.[224]

5.2.3.8.2 Methods of data analysis used for proteomics data obtained from ICR.

Data received from the ICR proteomics analysis was processed and analysed using Excel, Phantasus v1.19.3[225] and BioVenn.[226] Pathway analysis of the interacting proteins identified was carried out using Metascape.[227]

5.3 Results

Investigations were conducted, presented in this Chapter, to use pull down techniques to identify proteins that interact with the flavone compound in cell lysate. The probe molecule synthesised in Chapter 4 was used to covalently link interacting proteins and visualise them, using a fluorescent and biotin reporter tag, *via* SDS-PAGE and proteomics analysis.

5.3.1 Protein concentration determination using Bradford protein assay

Prior to incubation with either the active or inactive probe, standardisation of protein lysate samples was required to allow direct comparison of protein bands using SDS-PAGE and western blotting. Determination of protein concentration of cell lysate, calculated from dilutions of concentrated lysate (20,000,000 cells, 100 µL), was achieved using a Bradford protein assay. The absorbance of dilutions of the cell lysate was measured after adding Bradford protein assay dye. This was compared to the absorbance of standard solutions of BSA, which were used to generate a standard curve (Figure 5.4).

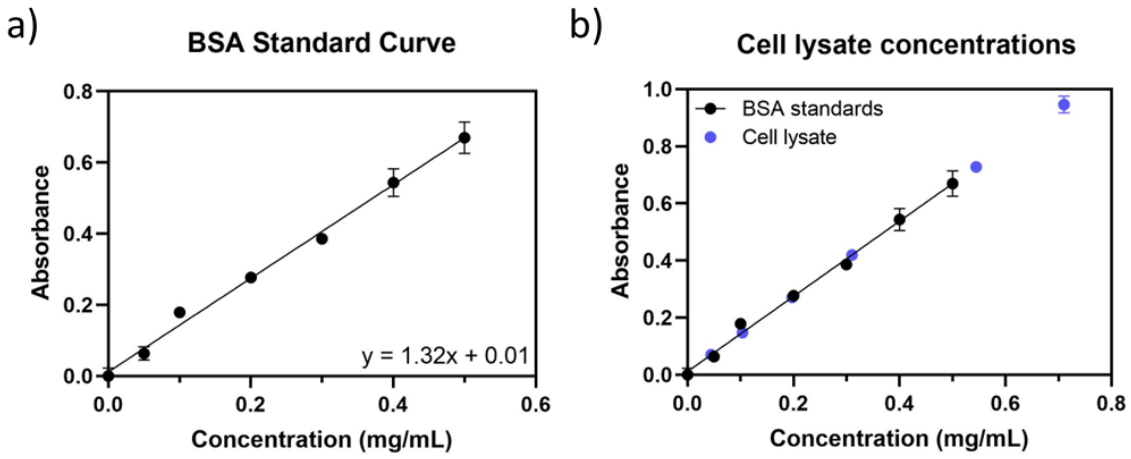


Figure 5.4: a) Standard curve of absorbance of BSA protein standards fitted with a linear regression line with the equation y = 1.32x + 0.01, r^2 = 0.99. b) BSA standard curve with additional points of diluted cell lysate, with concentrations calculated from average absorbance values using the regression line equation.

The equation from the linear regression line was used to convent the absorbance values of the diluted lysate samples to protein concentrations, then these were used to calculate the undiluted concentrations using the dilution factor. Points which fell within the linear regression line were averaged to give the concentration of the lysate (Table 5.5). In the example presented here, the lysate concentration was found to be 14.79 mg/mL.

Table 5.5: Data used to calculate concentration of cell lysate from a Bradford protein assay. Values shown in red were used to calculate the average concentration of the cell lysate.

Lysate dilution factor	10	20	40	80	160	320	640
Mean absorbance (AU)	0.946	0.728	0.420	0.272	0.148	0.070	0.000
Diluted concentration from equation (mg/mL)	0.710	0.545	0.310	0.198	0.104	0.045	-0.008
Undiluted concentration (mg/mL)	7.105	10.892	*12.416*	*15.818*	*16.639*	*14.295*	-5.358

5.3.2 Detection of interacting proteins using a fluorescent reporter tag

Small-scale pull down experiments were carried out using a FITC-azide reporter tag for visualisation of the proteins that had been covalently linked to the probe molecule, indicating a bonding interaction between the protein and the flavone targeting group. During initial pull down condition optimisation, some samples were attached to the FITC-azide reporter tag using click chemistry. The conditions tested are detailed in **Table 5.6** for two experiments using the FITC-azide reporter.

Table 5.6: Conditions of two experiments using the FITC-azide reporter tag.

Condition	**Experiment 1**	**Experiment 2**
Lysate volume used (3 mg/mL)	100 µL (0.3 mg)	100 µL (0.3 mg)
Probe volumes used (2 mM)	0.5 - 50 µL	0.5 – 25 µL
FITC-azide volume used (1 mM)	8 µL	8 µL
Length of time of click reaction incubation	30 minutes	1 h
Max protein loaded per well	15 µL (0.075 mg)/ 25 µL (0.125 mg)	15 µL (0.075 mg)

Fluorescent imaging of the gels after electrophoresis is shown in Figure 5.5, with the conditions of each sample displayed above. The gels show low levels of fluorescence in most lanes, with a higher intensity seen at the bottom of the gel for each lane, shown by a blue box. The higher fluorescence at the base of the gel may be due to residual unreacted FITC-azide remaining in the samples from the click reaction step, or probe that had reacted with the FITC-azide but had not formed a covalent link with a protein in the lysate and remained in solution as a free molecule. The FITC-azide has a molecular weight of 607.6 g/mol, and the probe product of the click reaction has a molecular weight of 1391.4 g/mol, so these compounds would be expected to travel to the bottom of the gel during electrophoresis. The background fluorescence seen on both gels is higher for the samples in which the active probe compound, **4.14**, was administered, instead of the inactive control probe, **4.13**, this suggests a higher level of association of the active probe compound with the proteins than the inactive probe, which is to be expected from the experiment as the inactive control probe does not have the targeting group.

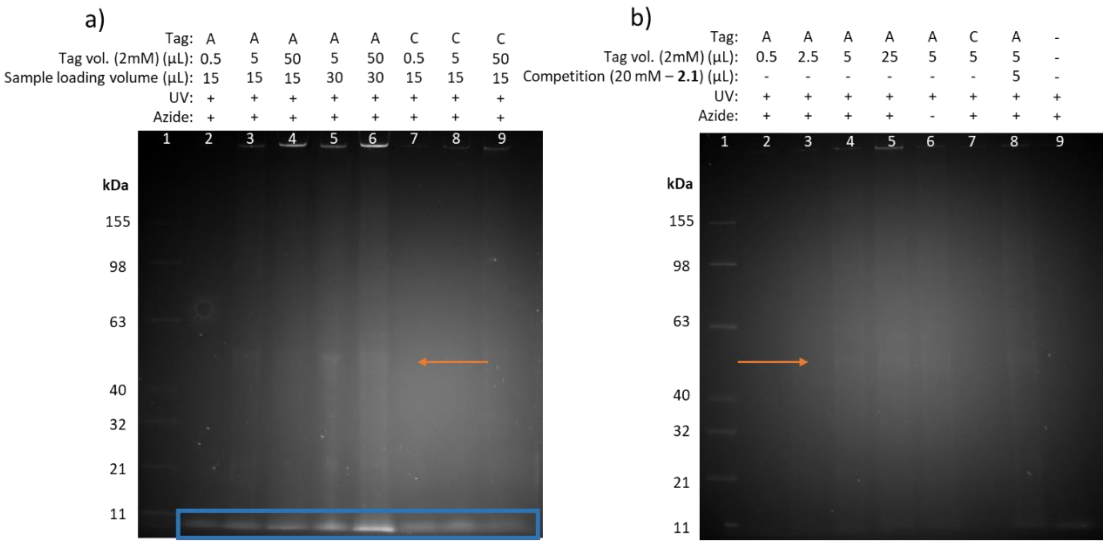


Figure 5.5: Fluorescent gel images of pull down experiments using FITC-azide reporter tag. The ladder (lane 1) is shown on the left hand side with protein molecular weight in kDa shown. A = Active probe compound, **4.14,** C = Control probe compound, **4.13**. a) Experiment 1 (Table 5.6) b) Experiment 2 (Table 5.6), this experiment included a condition in which compound **2.1** was used as a competition

compound. The blue box indictes higher levels of fluorescence seen at the base of the gel. Orange arrows indicate areas of discussion.

The gels from both experiment 1 and 2 however, do not show any prominent bands that suggest a key binding protein. Some more intense regions were seen between 63 and 40 kDa, indicated with the orange arrows, but this was also present in lane 6 of experiment 2 which was used as a no azide control and so did not have the reporter tag added to the sample, suggesting that it is background fluorescence and not a key interaction. It was considered that there may have been errors made while carrying out the SDS-PAGE protocol for these experiments and that was the reason for the inconclusive results obtained, therefore after fluorescent imaging of the gel in experiment 2, the gel was stained using Coomassie staining to ensure that the gel ran successfully (Figure 5.6).

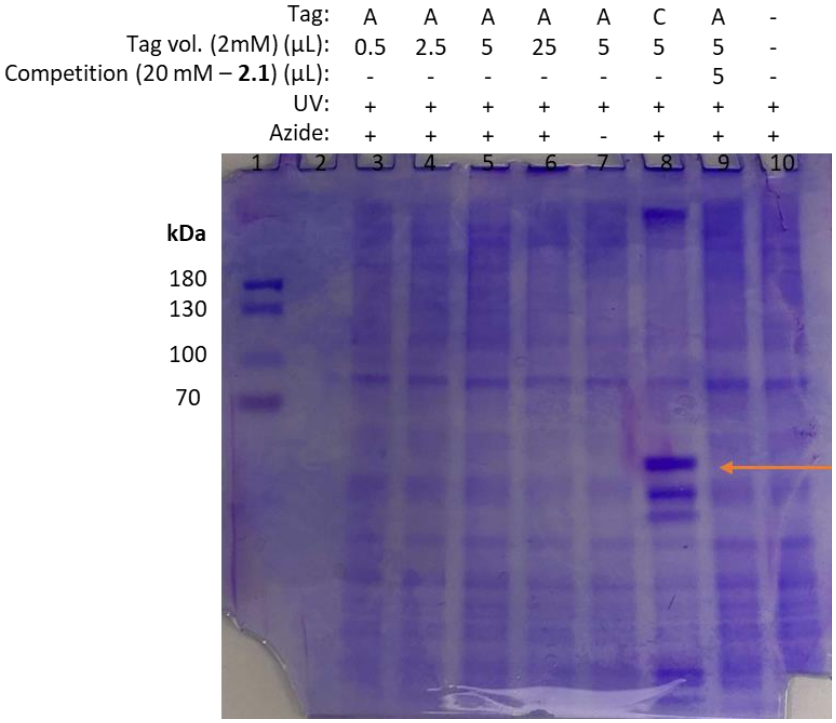


Figure 5.6: Coomassie stained SDS-PAGE gel for experiment 2 (Table 5.6). The ladder (lane 1) is shown on the left hand side with protein molecular weight in kDa shown. A = Active probe compound, **4.14,** C = Control probe compound, **4.13**. This experiment included a condition in which compound **2.1** was used as a competition compound. The orange arrow indicates prominent bands discussed in the text.

The Coomassie stained gel showed consistent protein distribution across all lanes, with the exception of a few prominent bands seen in lane 8, shown by the orange arrow. This sample contained the inactive probe molecule, **4.13**, so the protein precipitation of this sample may have been different to those with the active probe, **4.14**. However, the results from this stained gel showed that the SDS-PAGE was successful, and the reason for the inconclusive results seen in the fluorescent gel were still unknown. There were many potential explanations for the inconclusive results seen for the FITC-azide reporter tag experiments, discussed further in section 5.4.1, however due to time constraints of the project, it was decided that the research would move to focusing on results from the biotin-azide reporter tag. Use of the biotin-azide reporter tag would result in the final samples required for the proteomics analysis of the binding proteins and therefore western blot analysis of the proteins isolated with this method was prioritised.

5.3.3 Optimisation of conditions for detection of interacting proteins using a biotin reporter tag

Initial experiments utilising the biotin-azide reporter tag involved optimisation of the experimental conditions. Conditions were altered to try to optimise binding of the target group on the probe molecule to the protein and ensure successful click reactions between the azide and alkyne, while reducing non-specific labelling, which can increase background signal on the western blot and proteomics analysis. The sample loading volume, volume of azide, probe volume and click reaction incubation time were all potential experimental conditions that could be altered for optimisation.

Table 5.7: Experimental conditions of experiment 3, using a biotin-azide reporter tag.

Condition	Experiment 3
Lysate volume used (3 mg/mL)	100 µL (0.3 mg)
Probe volumes used (2 mM)	0.5 µL
Biotin-azide volume used (1 mM)	8 µL
Length of time of click reaction incubation	30 minutes
Max protein loaded per well	15 µL (0.09 mg)

Seven individual lysate samples were analysed, after incubation with either the active probe compound, **4.14**, or the inactive control compound, **4.13**. UV light was then used to form a covalent linkage between the interacting proteins and the photoactive diazirine. Click chemistry techniques were used to attach a biotin reporter tag to the probe-protein complex and then methanol based protein precipitation methods were used to isolate the proteins. The samples were analysed using SDS-PAGE and then transferred to a PVDF membrane. The membrane was incubated with streptavidin-HRP to detect the biotinylated proteins. Initial experimental conditions (Table 5.7) resulted in the western blot images shown in Figure 5.7. Biotinylated proteins are visible on the blot as ECL reagents were used to detect the HRP from the streptavidin-HRP reaction. The blot showed distinct bands of high intensity between 70 and 100 kDa as well as a band at 130 kDa (indicated by orange boxes), which were visible in all samples regardless of the probe. These bands are representative of endogenously biotinylated proteins, and therefore are likely not bound to the probe, but instead associate directly with the streptavidin-HRP. Mammalian cells possess four endogenous biotinylated proteins, pyruvate carboxylase (130 kDa), 3-methylcrotonyl coA carboxylase (75 kDa), propionyl coA carboxylase (72 kDa) and acetyl coA carboxylase (220 kDa), three of which are visible on the blot shown[228]. A lower exposure image of this blot (Appendix 8.40) clearly shows two distinct bands in the region of 70 – 100 kDa, verifying the two biotinylated proteins discussed above.

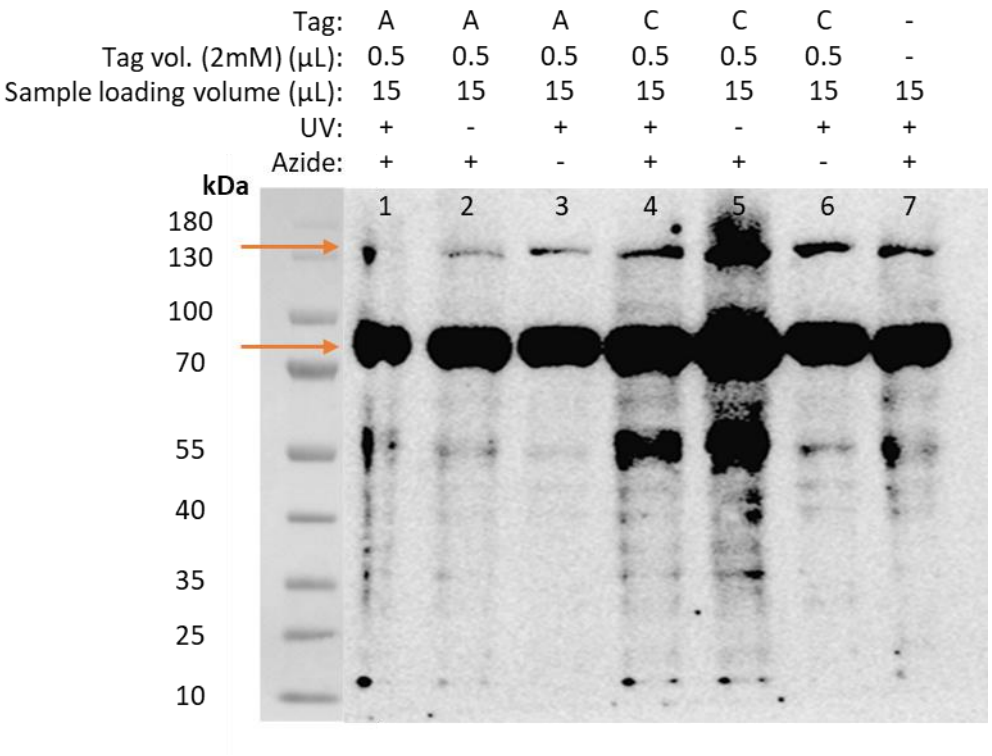


Figure 5.7: Western blot images of pull down experiment 3 using biotin-azide reporter tag, then detection of biotinylated protiens using streptavidin-HRP. The ladder is shown on the left hand side with protein molecular weights in kDa shown. A = Active probe compound, **4.14,** C = Control probe compound, **4.13**. Endogenously biotinylated proteins are indicated with orange arrows.

Samples treated with the inactive control probe showed a higher number and a higher intensity of bands than those treated with the active control probe, suggesting that the targeting group was increasing accumulation of the probe with interacting target proteins, therefore increasing selectivity. However, a control sample that was not treated with the either probe compound (lane 7) showed many of the bands present in the other sample lanes, indicating that the bands seen were not genuine interactions of the probe with proteins in the lysate.

Experiment 4 focused on increasing the concentration of the active probe and inactive control probes, and increasing loading volume of sample, to aid in visualisation of genuine protein-probe complex bands (Table 5.8).

Table 5.8: Experimental conditions of experiment 4, using a biotin-azide reporter tag.

Condition	Experiment 4
Lysate volume used (3 mg/mL)	100 µL (0.3 mg)
Probe volumes used (2 mM)	0.5 µL – 50 µL
Biotin-azide volume used (1 mM)	8 µL
Length of time of click reaction incubation	30 minutes
Max protein loaded per well	15 µL (0.075 mg)/ 25 µL (0.125 mg)

Probe compounds were added to the samples at 0.5, 5 and 50 µL volume, and for the active probe compound, **4.14**, the loading volume for the gel was tested at 15 and 25 µL (Figure 5.8). The key finding gained from this experiment was that the addition of 50 µL of probe compound resulted in high intensity band smearing, due to an oversaturation of probe molecule in the lysate mixture. A lower exposure image (Appendix 8.41) shows that the addition of 50 µL of active probe resulted in complete smearing of the lane (lanes 3 and 5), but some clear bands could be seen at 50 µL of the inactive control probe (lane 8). Some bands that were not visible at 0.5 µL probe volume became visible after addition of 5 µL of probe in the active probe lanes (lanes 1-5), these are indicated by the orange arrows. These bands are not seen in the inactive control compound lanes (lanes 6-8) suggesting a specific interaction between the targeting group and the interacting protein molecule.

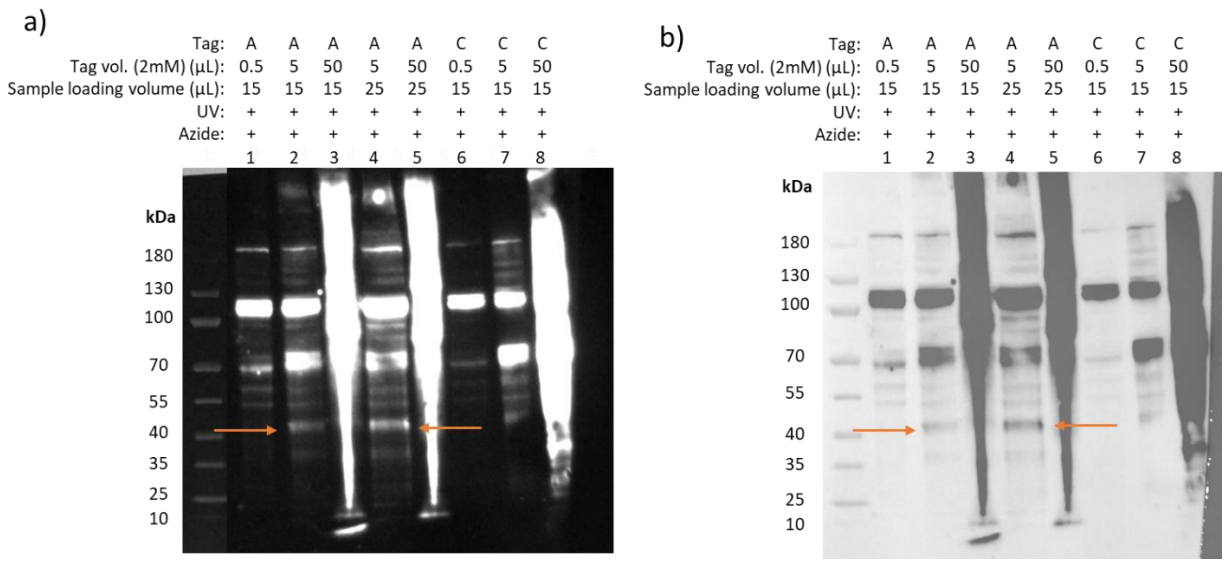


Figure 5.8: Western blot images of pull down experiment 4 using biotin-azide reporter tag, then detection of biotinylated protiens using streptavidin-HRP. The ladder is shown on the left hand side with protein lengths in kDa shown. A = Active probe compound, **4.14,** C = Control inactive probe compound, **4.13**. a) and b) show the same blot in difference contrast settings. Orange arrows indicate bands which have become present in the higher probe volume samples.

Experiment 5 introduced a competition sample, this sample is pre-incubated with the parent compound, **2.1**, to saturate binding sites of interacting proteins and prevent association of the probe. Specific interactions of the targeting compound and the protein should therefore be removed from the western blot image of a competition sample, as the probe has not bound to saturated proteins. Condition optimisation was continued for this experiment, with probe volumes used in a range of 0.5 µL – 25 µL. The results of this experiment (Figure 5.9) shows some bands increasing in intensity as the volume of probe compound used increased, shown by the blue arrow, others were classified as non-specific interactions due to their presence in the inactive control probe sample. On addition of the competition compound, **2.1**, to the lysate sample, some precipitation was observed due to the lack of aqueous solubility of the compound, discussed in Chapter 3. The western blot result for this sample showed high levels of intensity and some smearing at high exposure, which may be due to the precipitation of the compound. An image taken at a lower exposure level (Appendix 8.42) shows the

bands of this sample (lane 7) more clearly, however no bands were seen to have been lost, compared to those without competition compound. The length of time the sample was incubated with the click reagents before protein precipitation was also increased for this experiment from 30 minutes to 1 h, to increase the likelihood of the reporter tag successfully binding to the probe-protein complex.

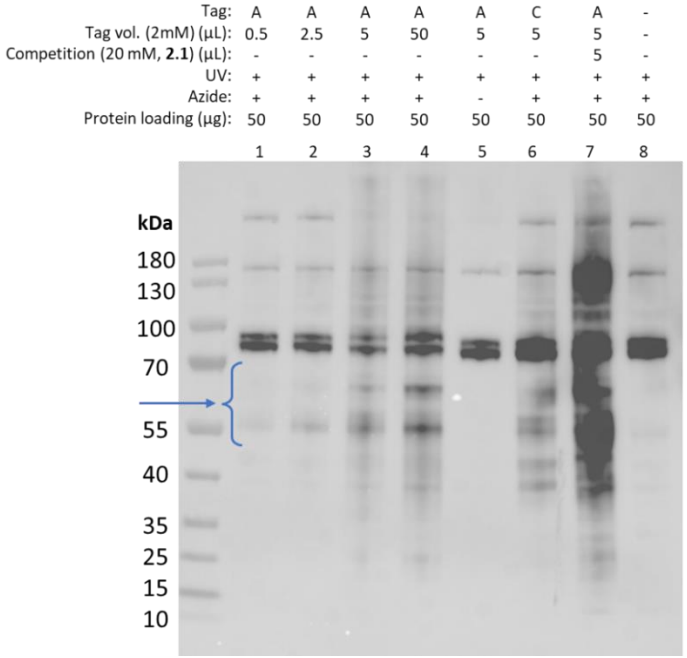


Figure 5.9: Western blot image of pull down experiment 5 using biotin-azide reporter tag, then detection of biotinylated protiens using streptavidin-HRP. The protein ladder is shown on the left hand side with protein molecular weights in kDa shown. A = Active probe compound, **4.14,** C = Control probe compound, **4.13**. Competiton compound used: **2.**1. Blue arrow indicates bands which have become present in the higher probe volume samples.

Building on the improvements of band intensity seen for the active probe samples in experiment 5, click reaction incubation times were extended to 3 h for some samples in following experiments, and compared to those incubated for 1 h as previously shown. It was also considered that the intense bands seen for the endogenously biotinylated proteins could be masking specific probe-protein interactions, so a 'preclearing' step was introduced before the lysate was exposed to the probe compound using streptavidin HP SpinTrap columns. A 'pull down' step was also introduced for all samples after the click reaction in which the protein was dissolved and incubated with magnetic

streptavidin beads, which allowed isolation of only biotinylated proteins from the solution. These proteins were removed from the beads for SDS-PAGE analysis. Experimental conditions for experiment 6 are shown in Table 5.9.

Table 5.9: Experimental conditions of experiment 6, introducing extended click reaction incubation times and biotinylated protein preclearing step. All samples were treated with 8 µL of biotin-azide reporter tag and a maximum protein loading concentration of 0.12 mg was used for each sample. A = Active probe compound, **4.14,** C = Control probe compound, **4.13**.

Sample	Probe compound (0.5 µL)	Preclearing step	Click reaction incubation time
1	-	-	1 h
2	A	-	1 h
3	-	✓	1 h
4	A	✓	1 h
5	C	✓	1 h
6	-	-	3 h
7	A	-	3 h
8	-	✓	3 h
9	A	✓	3 h
10	C	-	3 h
11	C	✓	3 h

Pre-incubation of the cell lysate with the streptavidin successfully removed the endogenously biotinylated proteins from the western blot, however there were also no specific interactions visible in these lanes. The increase of click reaction incubation time resulted in higher levels of band intensity compared to the 1 h incubation time, even at a lower probe volume of 0.5 µL (lane 2 vs lane 7, Figure 5.10).

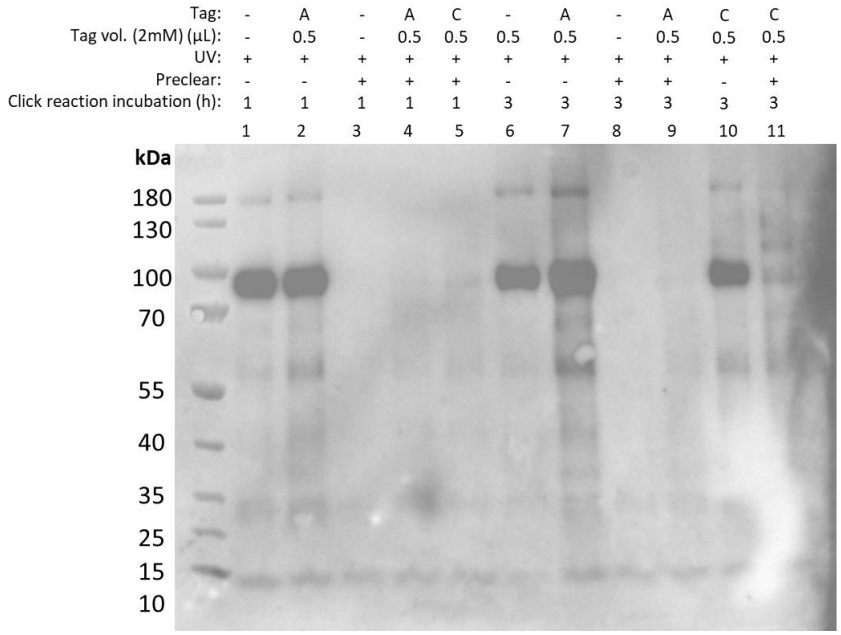


Figure 5.10: Western blot image of pull down experiment 6 using biotin-azide reporter tag, then detection of biotinylated protiens using streptavidin-HRP. The protein ladder is shown on the left hand side with protein molecular weights in kDa shown. A = Active probe compound, **4.14,** C = Control probe compound, **4.13**. Some lysates were precleared with streptavidin before incubation with probe compounds, as indicated.

The extended click reaction incubation time of 3 h was maintained for subsequent experiments as the intensity of bands observed was greater than for the 1 h or 30 minute incubation times. Experiments 7a and 7b focused on evaluating if increasing the concentration of biotin-azide used in the click reaction step resulted in higher levels of specific band visibility, while also determining if preclearing the lysate before probe incubation was advantageous. Experiment 7a used lysate that had been precleared with streptavidin before probe exposure and experiment 7b used lysate that retained the endogenous biotinylated proteins, a full SDS-PAGE and western blot experiment was carried out for both and the results compared. Probe compounds were added at volumes of 0.5 – 2.5 µL and some samples were preincubated with a competition compound before probe exposure, this compound was selected as compound **4.2**, which had shown a similar *in vitro* activity profile to that of the parent

compound PMF but had improved aqueous solubility characteristics. Both experiments involved a 'pull down' step with magnetic streptavidin beads after precipitation as discussed for experiment 6.

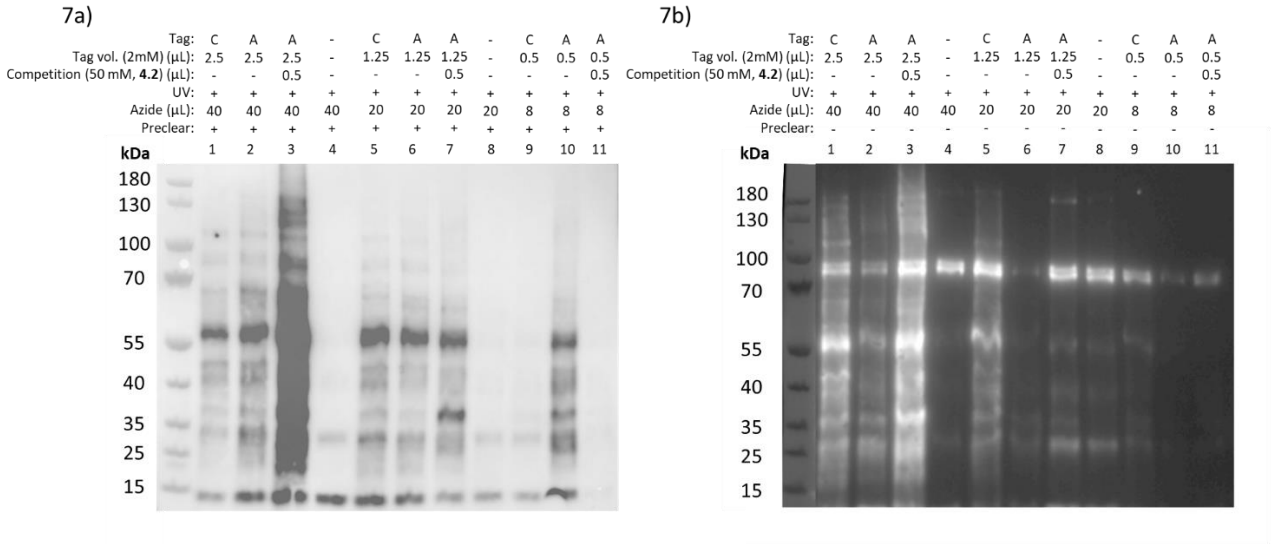


Figure 5.11: Western blot images of pull down experiments 7a and 7b using biotin-azide reporter tag, then detection of biotinylated protiens using streptavidin-HRP. The protein ladder is shown on the left hand side with protein molecular weights in kDa shown. A = Active probe compound, **4.14,** C = Control probe compound, **4.13**. Competition compound: **4.2**. 7a) All lysate was precleared with streptavidin before incubation with probe compounds. 7b) Lysate with all biotinylated proteins remaining. Images shown are in different contrast settings.

Addition of 1.25 µL of active control probe compound, **4.13**, was selected for further experiments as it resulted in higher intensity and number of bands compared to 0.5 µL of probe and 2.5 µL appeared to result in higher levels of non-specific binding, especially seen for the competition lane (lane 3, experiment 7a) which had smearing across all protein sizes. Scaling the quantity of azide used alongside the volume of probe did not result in any negative outcomes therefore was maintained. These experiments did not provide sufficient data to support the use of the preclearing step so further studies were carried out to determine if this was beneficial to the protocol.

Experiment 8 involved the analysis of four sample conditions: a no tag control; a 1.25 µL treatment of active probe, **4.14**; treatment of the sample with the same volume of active probe but after pre-

incubation with competition compound **4.2** and treatment with 1.25 µL of inactive control probe, **4.13**. Each of these sample conditions was tested in cell lysate that had been precleared with streptavidin and lysate that had not been precleared (Figure 5.12).

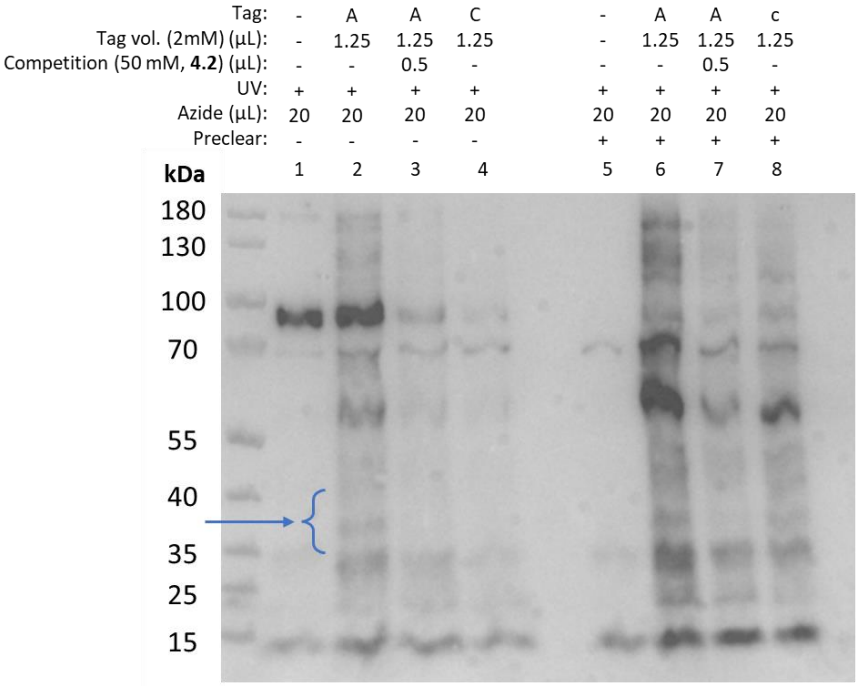


Figure 5.12: Western blot images of pull down experiment 8 using biotin-azide reporter tag, then detection of biotinylated proteins using streptavidin-HRP. The protein ladder is shown on the left hand side with protein molecular weights in kDa shown. A = Active probe compound, **4.14,** C = Control probe compound, **4.13**. Compeition compound: **4.2**. Blue arrow indicates bands which decreased in intensity in the sample treated with the competition compound before incubation with the active probe. Precleared samples are as indicated.

Preclearing of the lysate effectively removed the endogenous biotinylated proteins as these are not visible in the western blot image for those samples (lanes 5-8), however, it was decided that there could be a potential loss of important binding proteins during this preclearing step that was not detectable with the western blot analysis, with low levels of discernible benefit to the clarity of the results therefore was not carried forward for further experiments and LCMS analysis.

Some bands were seen in the sample treated with only the active probe, that were not visible in the sample that has been pre-incubated with the competition compound, this suggests a specific interaction between the probe and the protein, shown by the blue arrow in Figure 5.12. The optimised conditions tested in experiment 8 were selected for scale up of the experiment, to isolate tagged proteins for proteomics analysis.

5.3.4 Final results of western blot analysis of interacting proteins using a biotin reporter tag

Samples produced for LCMS proteomics analysis were required to be at a larger scale than those used for western blot analysis as the samples required downstream processing before analysis. Samples were scaled to 500 µL of cell lysate from 100 µL used for previous experiments (Table 5.10). A wider range of concentrations of competition compound, **4.2**, were also studied, at a one in three dilution from 10 eq of competition compound compared to active probe compound: 10 eq, 3.33 eq, 1.11 eq and 0.37 eq. This was compared to treatment with only the active probe compound, **4.14**, a DMSO/no probe control and an inactive probe compound, **4.13**, control. This experiment was performed in triplicate to ensure reproducibility of the results obtained, and an example set of results for this experiment are presented in Figure 5.13, supplementary data can be found in Appendix 8.43.

Table 5.10: Experimental conditions of experiments 9 - 12 for proteomics analysis.

Condition	Experiments 9 - 12
Lysate volume used (3 mg/mL)	500 µL (1.5 mg)
Probe volume used (2 mM)	6.25 µL
Biotin-azide volume used (1 mM)	100 µL
Length of time of click reaction incubation	3 h
Preclearing of lysate with streptavidin columns (before treatment with probe)	No
Isolation of tagged proteins using streptavidin beads (after treatment with probe)	Yes
Competition compound concentrations used (eq to probe)	10 eq, 3.33 eq, 1.11 eq, 0.37 eq

A portion of each sample was taken from the larger scale experiment volume for analysis *via* SDS-PAGE and western blot and the remainder of the sample was retained for proteomics analysis. A prominent band can be seen between 35 and 40 kDa, which decreases in intensity as the equivalents of competition compound (added prior to incubation with the probe compound) increases, this suggests a specific interaction between the protein and probe compound. This result is seen consistently for the triplicate results of the experiment in these conditions.

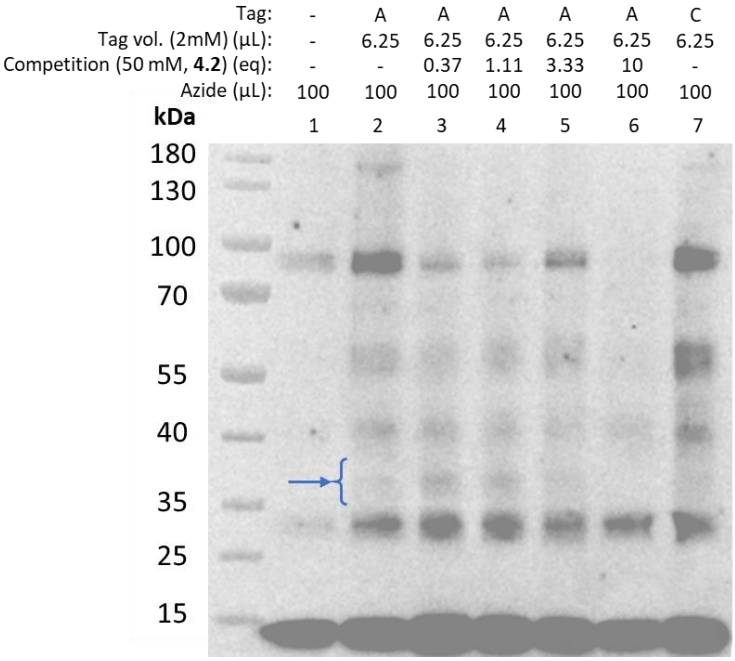


Figure 5.13: Western blot images of pull down experiment 9 using biotin-azide reporter tag, then detection of biotinylated protiens using streptavidin-HRP. The protein ladder is shown on the left hand side with protein molecular weights in kDa shown. A = Active probe compound, **4.14,** C = Control probe compound, **4.13**. Competition compound: **4.2**. Blue arrow indicates bands which decreased in intensity as the concentration of competition compound added before incubation with the active probe increased.

A second set of SDS-PAGE gels were also produced for experiments 9 to 11, using the same samples, and were stained with silver staining instead of transferring the proteins for western blot analysis. This enabled the visual determination of complete protein content across the gel which is beneficial for proteomics analysis, as all proteins are detected using LCMS, not only those detectable with

streptavidin-HRP due to biotinylation (Appendix 8.44). The silver-stained gels confirmed consistent protein loading across the wells and effective distribution of the proteins vertically across the gel, confirming the SDS-PAGE ran correctly. It is worth of note, however, that the silver-stained gels did not show a protein band in the 35-40 kDa range that has previously been discussed.

5.3.5 Proteomics analysis of pull down study samples

After samples were labelled and combined for proteomics analysis by collaborators at the ICR, the samples were analysed with methods discussed in section 5.2.3.8.2 and the data was processed and returned. Data was received with intensity values for each protein detected in each sample. A total of 2069 distinct proteins were identified using the proteomics methods across all nine samples (Appendix 8.45) The data obtained from the proteomics experiment allowed for detection and quantification of the proteins in the samples and therefore the proteins that were found in each treatment group could be compared using differential expression calculations. Table 5.11 describes the samples submitted for analysis.

Table 5.11: Sample names assigned to the nine samples submitted for proteomics analysis. Sample names are assigned as EF = researcher initals, 1,2,3 = Sample condition, A,B,C = Biological repeat. Lane reference refers to the lanes assigned in final western blot experiment (Figure 5.13).

Sample	Treatment condition	Lane ref.	Active Probe Vol (2 mM)	Competition compound eq
EF1A	DMSO	1	0	0
EF1B	DMSO	1	0	0
EF1C	DMSO	1	0	0
EF2A	Active tag	2	6.25	0
EF2B	Active tag	2	6.25	0
EF2C	Active tag	2	6.25	0
EF3A	Active tag + competition	6	6.25	10
EF3B	Active tag + competition	6	6.25	10
EF3C	Active tag + competition	6	6.25	10

This data required normalisation across the different samples before the results of the pull down study could be investigated. Normalisation methods for data of this type vary depending on the quality of

the data obtained and the degree to which samples within replicates differ. Biological variation can account for some degree of difference in samples from different replicates, however contamination and human error can create larger differences between samples of the same treatment group. Normalisation of the data can improve these differences and allow data to be compared more effectively.[229]

5.3.5.1 Processing of proteomics data using multiple normalisation methods

A principal component analysis (PCA) was carried out, using Phantasus, to analyse the data before normalisation to indicate variation between the triplicate samples under the same treatment conditions (Figure 5.14). The initial PCA plot shows that samples of different replicates are not sufficiently clustered together, indicating high levels of variation between samples of the same treatment group. Normalisation of the data was used to reduce the variation between the samples and the PCA plots allowed the effectiveness of normalisation methods to be evaluated.

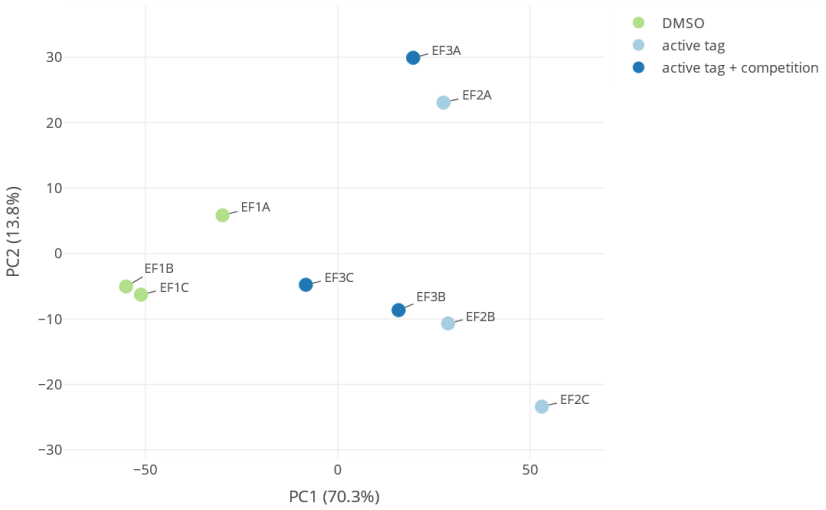


Figure 5.14: PCA plot showing variation of samples using raw data (without normalisation methods). Samples under identical treatment conditions are shown in the same colour and number, biological replicates are distinguished by letters: A, B and C.

Five different methods of normalising the data were trialled to reduce variation of samples between replicates (Table 5.12). All five methods resulted in increased clustering of biological repeat samples

(Repeat 1, 2 or 3) while maintaining the difference between sample treatment conditions (Samples A, B and C). This is shown in the PCA plots (Table 5.12), seen by increased clustering of the coloured points on the x-axis. Methods 3, 4 and 5 resulted in the best improvement of biological repeat sample clustering.

Table 5.12: Methods of normalisation of raw proteomics data. x = raw data point of each protein for that sample.

Method number	Formula
1	$(\frac{x}{sum\ of\ values\ for\ that\ sample} \times maximum\ intensity\ of\ all\ samples)$
2	$(\frac{x}{mean\ of\ values\ of\ that\ sample} \times average\ of\ means\ for\ that\ treatment)$
3	$(\frac{x}{median\ of\ values\ for\ that\ sample} \times median\ of\ medians\ of\ that\ treatment)$
4	$(\frac{x}{sum\ of\ values\ for\ that\ sample} \times mean\ of\ all\ data\ points\ in\ that\ treatment)$
5	$(\frac{x}{sum\ of\ values\ for\ that\ sample} \times median\ of\ all\ data\ points\ in\ that\ treatment)$

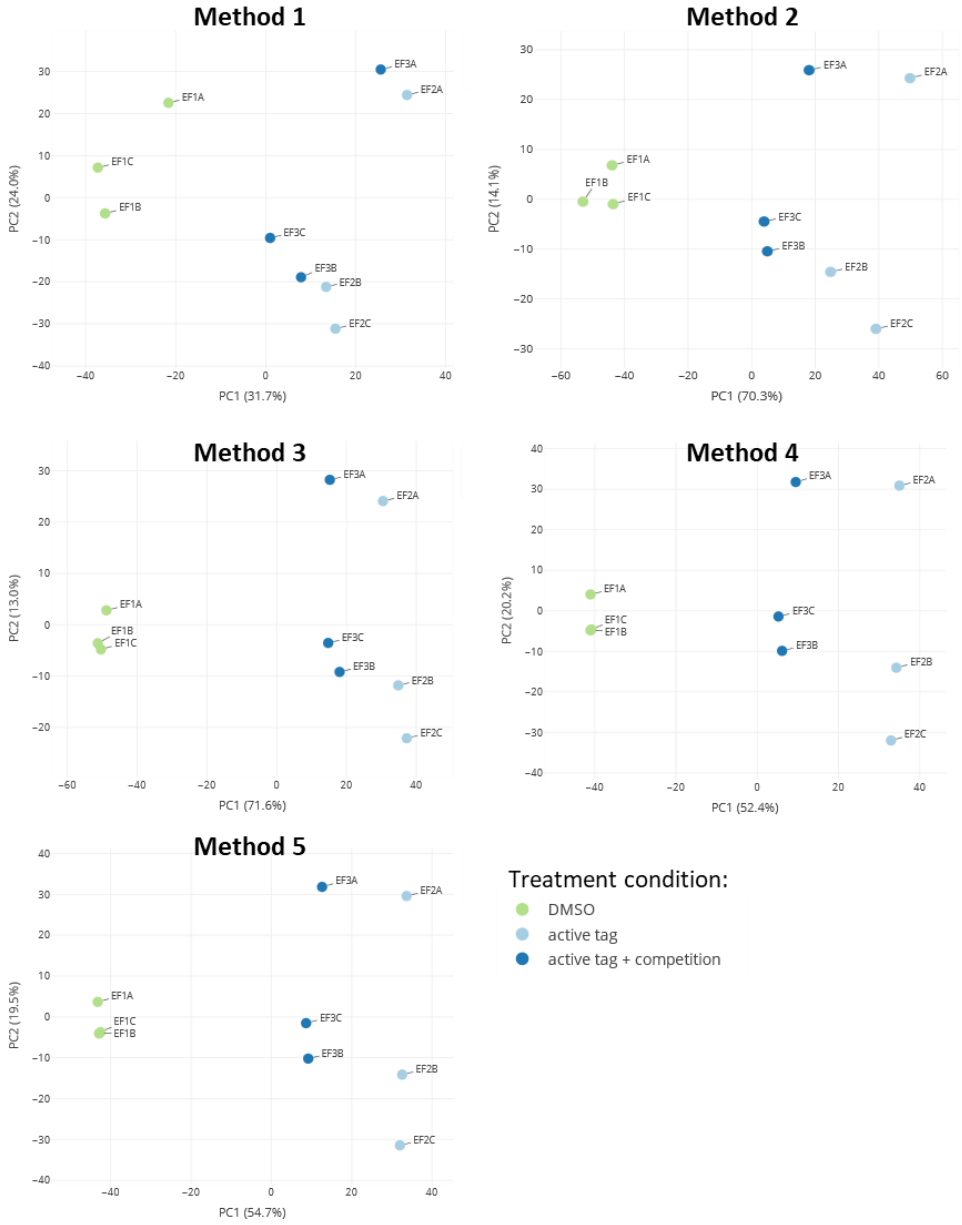


Figure 5.15: PCA plots showing clustering of samples after normalisation with five different methods. Methods detailed in Table 5.12. Samples under identical treatment conditions are shown in the same colour and number, biological replicates are distinguished by letters: A, B and C.

It was decided that some data analysis would be carried out using all normalisation methods, to extract potential binding proteins that were consistently found as hits for each method. Phantasus software was used to adjust the data by $\log_2$ and calculated the differential levels of the proteins in each treatment group using differential expression calculations using Limma. To identify proteins that were found to have increased in intensity (enriched) when treated with the active probe compound, **4.14**,

compared to the control samples that were treated with only DMSO a Limma comparison for 'DMSO vs active tag' was carried out. The proteins that decreased in intensity (depleted) when the competition compound, **4.2**, was pre-incubated with the lysate before addition of the active probe compound, **4.14**, were identified using a Limma comparison of 'tag vs competition'.

Log(Fold change) (Log_2(FC)) and p.value values were calculated and fold change and limits were determined (Table 5.13). To show proteins highly enriched by the addition of the active probe compound, a fold change limit of 1.5 was applied for 'DMSO vs tag' and to detect proteins that were lost from the sample after treatment with the competition compound a fold change limit of 0.8 for 'tag vs competition' was applied. A p.value limit of less than 0.05 was also applied to all data.

Table 5.13: Fold change and Log_2(FC) limits determined for the differential expression analysis of the treatment comparisons.

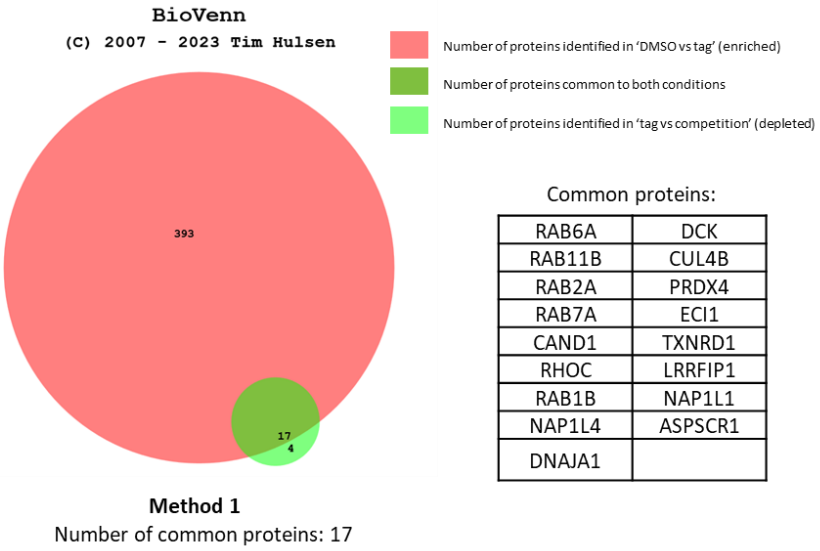
Comparison	Fold change limit	Log_2(FC) limit
DMSO vs Tag	1.5	0.585
Tag vs Competition	0.8	-0.3

For each method of normalisation, proteins that met the criteria detailed above were identified and this data was exported to Excel. Table 5.14 shows the number of proteins identified for each method of normalisation, for both the 'DMSO vs tag' and the 'tag vs competition' comparison after applying the Log_2(FC) and p.value limits.

Table 5.14: Number of proteins found for both comparisons using each normalisation method after applying fold change and p.value limits.

	Number of proteins identified	
Normalisation Method	**DMSO vs Tag**	**Tag vs Competition**
1	410	21
2	1848	389
3	1893	124
4	1494	221
5	1589	176

The lists of identified proteins from the two comparison conditions were then compared with each other, for each normalisation method, using BioVenn (Figure 5.16). The proteins symbols were placed in to BioVenn and the number of proteins that were common to both lists, were extracted using BioVenn. For normalisation method 1, there were 17 proteins found to be common to both lists.



Common proteins:

RAB6A	DCK
RAB11B	CUL4B
RAB2A	PRDX4
RAB7A	ECI1
CAND1	TXNRD1
RHOC	LRRFIP1
RAB1B	NAP1L1
NAP1L4	ASPSCR1
DNAJA1	

Figure 5.16: Example Venn diagram produced by comparing proteins meeting fold change limits for two comparison conditions 'DMSO vs tag' and 'tag vs competition', using normalisation method 1. Proteins common to both lists are shown in a table. Complete data for each normalisation method can be found in Appendix 8.47.

The lists of overlapping proteins for the two comparisons were extracted using all methods of normalisation (Appendix 8.48) and the proteins that appeared frequently were identified. Specifically, seven proteins were found in all five methods of normalisation (Table 5.15). Proteins common to only four methods of normalisation were also identified (Appendix 8.49).

Table 5.15: Proteins common to all lists after application of the fold change and p.value limits, using five different methods of normalisation. Protein symbols and names were generated from databases using Phantasus from UniProt codes.

UniProt	Symbol	Protein name
Q9BZE9	ASPSCR1	ASPSCR1, UBX domain containing tether for SLC2A4
Q86VP6	CAND1	cullin associated and neddylation dissociated 1
P27707	DCK	deoxycytidine kinase
Q32MZ4	LRRFIP1	LRR binding FLII interacting protein 1
P55209	NAP1L1	nucleosome assembly protein 1 like 1
Q99733	NAP1L4	nucleosome assembly protein 1 like 4
Q9H0U4	RAB1B	RAB1B, member RAS oncogene family

Protein families that are seen to occur regularly include the nucleosome assembly protein (NAP) family, DNAJ heat shock protein family and the RAB protein family which form part of the RAS superfamily of small GTPases.

5.3.5.2 Proteins identified as significant after application of fold change limits after data normalisation with one selected method.

Further analysis of the proteomics results was also carried out by focusing on one normalisation method to identify the proteins with high fold changes. Normalisation method 3 was chosen as the best method tested for reducing the variation of samples between biological repeats, as seen in the PCA plot (Figure 5.15) which shows the samples aligned significantly on the x-axis (PC1). The first principal component accounts for 71.6% of the plot for this method, so it was selected for further investigation. Using Phantasus, application of the p.value and log(FC) cut off values discussed previously were applied to identify significant proteins, and it was found that 1893 proteins were increased upon addition of the active probe compared to the DMSO control and 124 proteins were decreased in intensity when treating the sample with the competition compound, **4.2**, before addition of the active probe (Figure 5.17). Complete lists of the proteins identified can be found in Appendix 8.46.

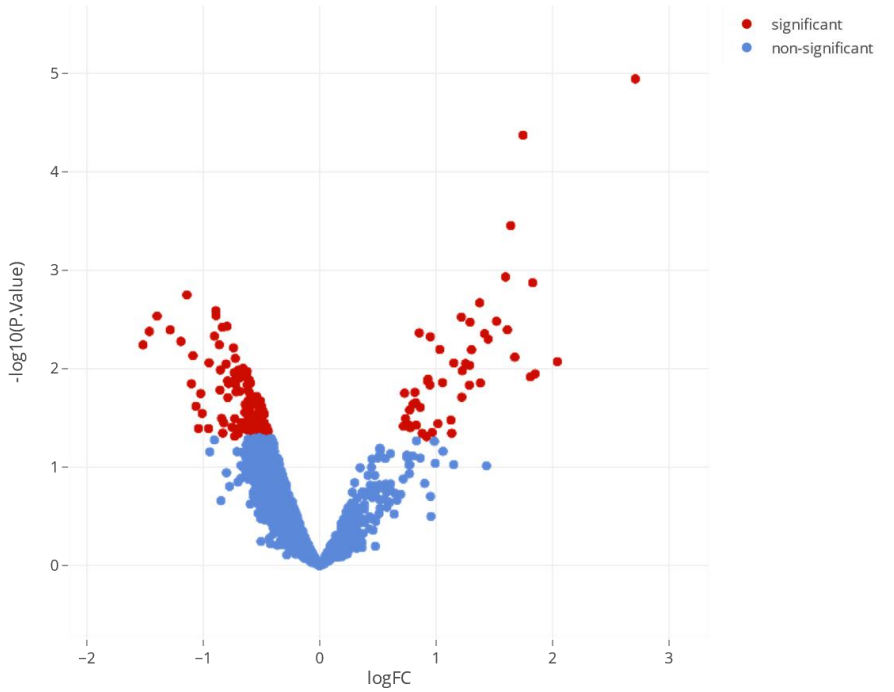


Figure 5.17: Volcano plot showing proteins detected for samples comparing active probe (tag) vs competition compound. Red dots indicate proteins meeting the significance criteria of $p \leq 0.05$. Significant negative log(FC) values indicate a decrease in the intensity (FC = 0.8) of the protein when the sample was treated with the competition compound compared to the samples without.

As seen in the Venn diagram (Appendix 8.47), 123 proteins were common to both comparison conditions. These can be seen in the labelled volcano plot (Appendix 8.51). Metascape was used to analyse the pathway involvement of these 123 proteins (Figure 5.18).[227] The overlapping proteins were analysed as a gene list, and pathways associated with the proteins in the list were identified and ranked by significance. The pathway showing the highest level of involvement with this list, based on the number of genes from the list found in that pathway group, was cellular responses to stress. RAB geranylgeranylation was also identified as a key pathway association, due to the high level of RAB proteins found in the list. Signalling by Rho GTPases, micro GTPases and RHOBTB3 is also shown as a key enriched geneset, due to the association of the probe with RhoC.

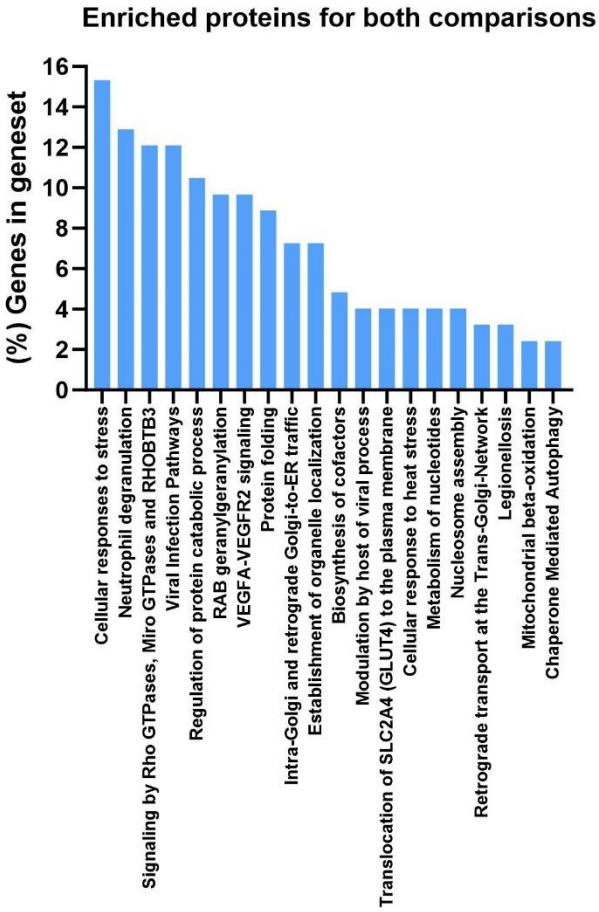


Figure 5.18: Top 20 enriched genesets identified using Metascape analysis of the overlapping 123 proteins pulled down and identified using proteomics.

To identify the proteins that were reduced in intensity the most by treatment with the competition compound, indicating true interactions between the parent compound and the proteins, the proteins with the largest negative fold change difference between the active probe samples and the competition samples were identified. Initially, this was carried out by ranking the log(FC) values for the 'tag vs competition' comparison from smallest to largest and then the top 20 proteins were selected. Semenogelin 1 (SEMG1) was found to fall in this selection of the top 20 proteins with the largest drop in intensity, however this protein had a log(FC) value of 0.24 for the 'DMSO vs tag' comparison – corresponding to a fold change of 1.18 – which did not meet the selected criteria of a minimum 1.5 fold change from the DMSO control to the active probe samples, so was removed from the top 20 selected and replaced with the next highest log(FC) value. These protein symbols were added to the

labelled volcano plot previously discussed (Figure 5.19). Nine proteins from the identified top 20 hits were labelled as members of the RAB family, suggesting key interactions of the target compound with this protein group.

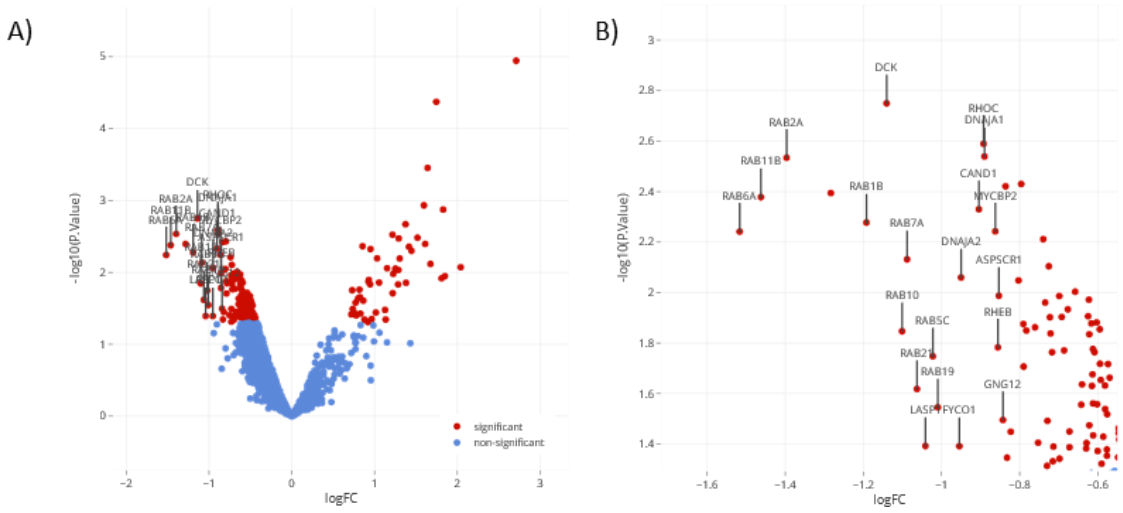
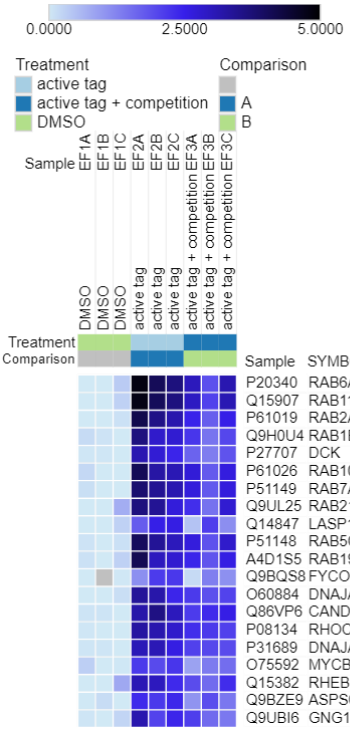


Figure 5.19: A) Volcano plot showing differentially expressed proteins for the tag vs competition comparison, labelled proteins fall as the top 20 largest fold changes for this comparison. B) Magnified image of the volcano plot.

The list of top 20 hit proteins also contained four of the seven proteins identified as hits in all five normalisation methods: UBX domain containing tether for SLC2A4 (ASPSCR1), cullin associated and neddylation dissociated 1 (CAND1), deoxycytidine kinase (DCK) and member RAS oncogene family (RAB1B). A further eight of the proteins seen in the top 20 hits were found in four out of the five normalisation methods, highlighted in Appendix 8.49.

A heatmap was generated from these top 20 hit proteins to visualise the variation in intensity for each protein in the different samples and compare these with the replicates to identify any anomalies (Figure 5.20). To generate the heatmap, the mean of the intensity of the DMSO samples for these proteins was calculated and this was subtracted from all values to centre the values around 0. Dark colours show that some proteins had much higher intensities when treated with the active probe than others.



Sample	SYMBOL	GENENAME	logFC	AveExpr	t	P.Value	Mean
P20340	RAB6A	RAB6A, member RAS oncogene family	-1.5166	9.2804	-3.7334	0.0057382	5.8545
Q15907	RAB11B	RAB11B, member RAS oncogene family	-1.4623	10.466	-3.9553	0.0041869	7.0375
P61019	RAB2A	RAB2A, member RAS oncogene family	-1.3964	9.3333	-4.2143	0.0029247	6.4020
Q9H0U4	RAB1B	RAB1B, member RAS oncogene family	-1.1921	6.8063	-3.7912	0.0052825	4.3202
P27707	DCK	deoxycytidine kinase	-1.1404	7.0118	-4.5854	0.0017796	4.8244
P61026	RAB10	RAB10, member RAS oncogene family	-1.1009	8.1850	-3.1182	0.014229	5.2652
P51149	RAB7A	RAB7A, member RAS oncogene family	-1.0882	10.397	-3.5591	0.0073854	7.4896
Q9UL25	RAB21	RAB21, member RAS oncogene family	-1.0629	4.1123	-2.7736	0.024114	1.3950
Q14847	LASP1	LIM and SH3 protein 1	-1.0410	3.5681	-2.4395	0.040533	1.8081
P51148	RAB5C	RAB5C, member RAS oncogene family	-1.0221	9.5431	-2.9675	0.017898	6.5820
A4D1S5	RAB19	RAB19, member RAS oncogene family	-1.0095	4.7810	-2.6655	0.028508	1.9976
Q9BQS8	FYCO1	FYVE and coiled-coil domain containing 1	-0.95426	3.1321	-2.4385	0.040598	1.7871
O60884	DNAJA2	DnaJ heat shock protein family (Hsp40) member A2	-0.94981	7.6618	-3.4462	0.0087161	5.1458
Q86VP6	CAND1	cullin associated and neddylation dissociated 1	-0.90425	10.910	-3.8767	0.0046777	8.1001
P08134	RHOC	ras homolog family member C	-0.89228	8.6068	-4.3080	0.0025751	5.9681
P31689	DNAJA1	DnaJ heat shock protein family (Hsp40) member A1	-0.88984	9.7314	-4.2229	0.0028906	7.2788
O75592	MYCBP2	MYC binding protein 2, E3 ubiquitin protein ligase	-0.86228	3.2929	-3.7350	0.0057248	1.6058
Q15382	RHEB	Ras homolog, mTORC1 binding	-0.85562	6.2106	-3.0211	0.016493	3.8370
Q9BZE9	ASPSCR1	ASPSCR1, UBX domain containing tether for SLC2A4	-0.85313	3.6978	-3.3335	0.010299	1.8076
Q9UBI6	GNG12	G protein subunit gamma 12	-0.84272	3.9698	-2.5915	0.031985	1.8531

Figure 5.20: Heatmap showing the 20 proteins with the largest log(FC) for the tag vs competition comparison. Light blue colours indicate values close to zero and dark colours into black indicate higher numbers. The average expression for these proteins is also shown in the table, generated by Phantasus.

The heatmap clearly shows that of the top 20 hit proteins, the nine RAB proteins identified are found to have some of the largest negative log(FC) changes, making up nine out of the top eleven proteins in the list.

The log(FC) values of these top 20 hit proteins can also be compared to the average expression of these proteins in cells (Figure 5.21). The average expression of these proteins was determined using Phantasus and is not specific to the cell line. The blue bars show the average expression given for each of the proteins displayed and the orange points correspond to the log(FC) value for each of these proteins. The presence of proteins with lower levels of average expression but retaining significant fold change values indicates more specific interaction of this protein with the target compound, as protein abundance in the lysate can influence the likelihood of the active probe associating. Shown in this analysis, LIM and SH3 protein 1 (LASP1), RAB5C and DNAJ heat shock protein family (Hsp40) member

A2 (DNAJA2) are worthy of note due to their relative low expression and low log(FC) values suggesting key interactions.

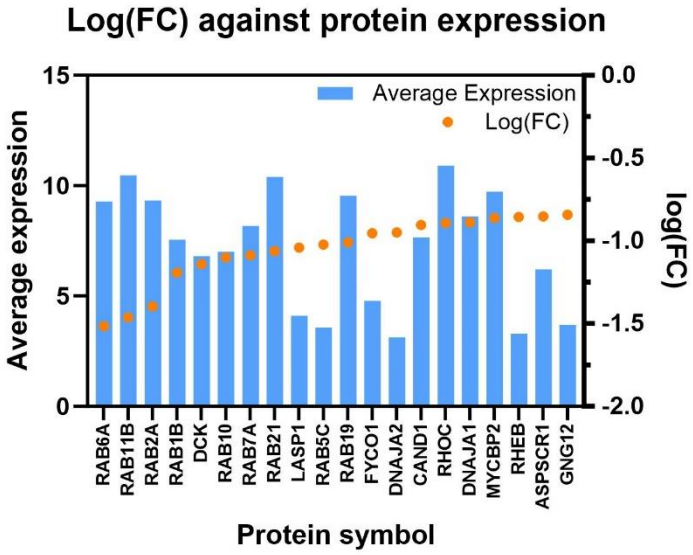


Figure 5.21: Comparison of the log(FC) values for each of the top 20 hit proteins previously discussed, and their average expression, determined by Phantasus.

5.4 Discussion

5.4.1 Detection of interacting proteins using a fluorescent reporter tag

Results of the analysis of proteins interacting with the TDP probe using a fluorescent reporter tag were inconclusive. The results from the detection of fluorescent proteins using SDS-PAGE showed high levels of background fluorescence, including in control lanes for samples that were not treated with the azide reporter tag. The gels showed no distinct fluorescent bands suggesting key protein binding interactions. Throughout condition optimisation for the western blot analysis of interacting proteins visualised using a biotin reporter tag, the experimental conditions were changed to improve detection of the proteins. This involved increasing the length of incubation time allowed for the reporter tag to react with the probe molecule during the click reaction step, for experiments 1 and 2 this was 30 mins and 1 h respectively but for final western blot experiments this was increased to 3 h. The compound used for the competition samples was also changed from compound **2.1** to compound **4.2** due to low

levels of aqueous solubility, which could have improved the effectiveness of the comparison between the competition and no competition lanes of the experiment.

The high levels of background fluorescence seen in Figure 5.5 hindered the visualisation of prominent bands on the gels, this could be due to remaining unreacted FITC-azide in the sample mixture, this is known to be problematic in this type of experiment and methods are being developed to overcome this issue.[213] Size-exclusion chromatography has been previously used to remove unreacted FITC molecules, however this can result in the loss of proteins. Chaganti *et al.* developed a method of purifying the products from the FITC tag using tandem affinity purification of fluorescently labelled proteins, which could be adapted to these experimental conditions to improve outcomes.[213]

The flavones molecules studied in this work also possess intrinsic fluorescence, with emission wavelengths between 400 and 600 nm after excitation at 365 nm (Appendix 8.52), that could have interfered with the detection of the FITC labelled proteins in the gel, this could be further investigated and choice of a fluorescent reporter tag that does not emit in the same region as the flavone molecules when irradiated could be chosen to reduce background.

5.4.2 Detection of interacting proteins using a biotin reporter tag

Results from western blot analysis of interacting proteins with the biotin reporter tag showed interactions with a wide range of proteins of different molecular weights, though a prominent band was seen in the molecular weight region of 35 – 40 kDa. The band in this molecular weight region decreased in intensity as the concentration of competition compound increased and then was lost completely at the highest level of competition compound. A second band that was lost on addition of the highest concentration of competition compound with a molecular weight slightly higher than the 55 kDa weight marker, though this did not decrease gradually as competition concentration increased. As discussed in the introduction of this chapter, there has been previous work that suggested that PMF suggests interaction with the COX enzymes, and result in the inhibition of PGE-2 production in CRC cells, however the molecular weight of the COX monomeric units are approximately 71 kDa, indicating

that these are not the proteins that are being detected in the western blots as interacting with the probe molecule in this experiment.[230]

Other suspected target proteins, Nrf2 and Pgp, both do not fall in the range of the key bands in the western blot with molecular weights of 95 – 110 kDa and 170 KDa respectively.[231,232] However, the molecular weight of Nrf2 when visualising with SDS-PAGE had previously been debated to be seen as a band seen in the molecular weight region of 55 – 65 kDa, which would match a band seen in the western blot analysis; the higher molecular weight quoted for Nrf2 band lies in the region of the western blot occupied by endogenously biotinylated proteins seen as a result of detecting with streptavidin-HRP and therefore may be hidden by these proteins.[231]

MCM7 and β-catenin that were associated with the biological activity of PMF in work by I. L. Fong, have molecular weights of 80 and 92 kDa respectively, also not matching the key bands seen in the western blot analysis. Identification of the proteins isolated by the pull down studies and their comparative intensity in the competition samples is therefore imperative to gaining insight into the target proteins of the compound and therefore it's behaviour in cells.

5.4.3 Identification of interacting proteins using LCMS proteomics analysis

5.4.3.1 Discussion of quality of data obtained from proteomics analysis

The PCA of the samples without normalisation showed that there was a large degree of variation between samples under the same biological conditions (same treatment), this could be accounted for to some degree by biological variation as the lysates used for the samples were generated from different cell pellets, however it may also indicate potential error in the production of the samples as the intensities recorded for each sample varied significantly. However, there was no trend seen in the intensity of proteins detected in each sample compared to the biological repeat they corresponded to, indicating that there was not compromised lysate extraction or protein precipitation in any one replicate specifically (Figure 5.22).

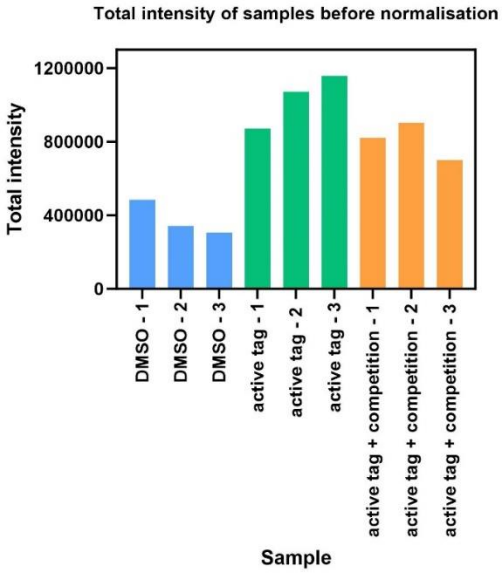


Figure 5.22: Total intensities of protein samples detected by proteomics analysis, before normalisation methods.

5.4.3.2 Pathway analysis of proteins increased in intensity by addition of the probe compound and reduced by the competition compound.

Metascape analysis of the overlapping proteins identified using normalisation method 3 gave insight into the ways in which the interactions between the probe compound and the proteins identified could influence molecular mechanisms occuring in the cells.[227] A large number of the proteins identified were found to fall into the geneset of cellular response to stress; this is a wide category that encompasses a range of biological processes that all indicate a response of the cell to disruptive stimuli. Binding to proteins in this category could indicate protective mechanisms of the flavone compound which could be responsible for the anti-cancer effects of the compound *in vitro*.

An important result of cells undergoing stress is the unfolded protein response (UPR). This is activated to rebalance protein folding mechanisms in the endoplasmic reticulum (ER) of the cells after a disturbance. Disruption of these vital processes can lead to apoptosis or cancer.[233] The pathway analysis of the proteins pulled down suggests involvement not only in cellular response to stress but also protein folding, cellular response to heat stress and chaperone mediated autophagy. Autophagy

is the orderly- degradation and recycling of cellular components through a lysosome-dependent mechanism, involved in cell survival and death that is often activated during cell stress by signalling caused by the UPR.[234] Flavones have previously been studied in reference to the UPR, such as a study carried out by Wu *et al.,* investigating the effects of luteolin and apigenin on the lipid peroxidation product 4-hydroxy-2-nonenal (4-HNE), which has roles in activating the UPR and can trigger the activation of other related UPR pathways.[235]

The pathway analysis also supports the suggestion that the flavone compound influences the behaviour of protein pathways regulated by members of the RAS superfamily of small GTPases. The compound could be potentially disrupting the activity of many GTPases, which could consequently impact the protein trafficking processes in the cells. RAB geranylgeranylation was identified as a pathway involvement of the pulled down proteins. Geranylation occurs as a post-translational modification of RAB proteins during their synthesis. Inhibition of this process, by inhibition of a key enzyme, geranylgeranyl diphosphate synthase, has previously been shown to be a potential mechanism of cancer treatment for cancers including osteosarcoma, Ewing sarcoma and pancreatic ductal adenocarcinoma.[236,237]

It is important to note that although the results from the pull down analysis identified proteins that are involved in these pathways it is not possible from this study to elucidate the exact mechanism in which the interaction of the targeting flavone compound could be having an effect on these proteins *in vitro* and therefore how the pathways involved may be altered, and further studies are needed to evaluate how the compound is impacting these pathways in order to give more insight into the true mechanisms of action of the compound.

5.4.3.3 Discussion of key protein interactions of the probe compound identified and relevance to CRC and flavonoid treatment.

Proteins that were detected *via* multiple normalisation methods and those identified in the top 20 hits of method 3 are to be taken as proteins of interest in the identification of the target proteins of PMF.

Those that occur frequently are most worthy of note. The data indicates a high level of specific interactions between the targeting group and the family of RAB proteins, suggesting important interactions with this group. The RAB family of proteins form the largest group in the superfamily of proteins called the RAS family. Mutations in these proteins are regularly found in colorectal cancers (CRC), with more than 30% of CRC driven by mutations in the RAS family.[238] Targeting of RAB GTPases to affect membrane trafficking of cargo has previously been studied as a potential strategy for cancer treatment, including CRC.[239] RAB proteins have previously been associated with flavones in a study by Ötzkan *et al.* showing a non-significant reduction in expression levels of RAB3A in brain tissue after administration of 7,8-dihydroxyflavone.[240] However, there is limited knowledge to date on how flavones may interact with RAB proteins in a cancer setting. The top 20 hit proteins also included RAS Homolog Family Member C (RHOC) a member of the Rho sub-group of the RAS superfamily and also acts as a GTPase, like many of the RAB family of proteins. Previous work from Ullmannova *et al.* has reported restoration of the expression levels of DLC1, a gene encoding for a Rho GTP-ase activating protein (GAP), in HT29 CRC cells after treatment with a dietary flavone. This change in gene expression after treatment resulted in suppression of cell proliferation *in vitro* and reduction of tumorigenicity *in vivo*.[241] The RHEB (RAS homolog enriched in brain) protein is a member of the RAS superfamily and also possesses GTPase activity.[242] The high number of RAS related proteins identified with the pull down proteomics experiment described in this chapter suggests that a high likelihood that the activity of PMF and the related flavones studied are involved with the RAS family and gives insight into the potential mechanisms of action of these compounds.

The other proteins identified as interacting with the pull down probe could also suggest mechanisms of action of the compound, a selection of these have previously been studied for their associations with cancer, specifically CRC, and previously investigated interactions with flavonoids.

Deoxycytidine kinase (DCK) regulation plays an important role in cancer and chemotherapy resistance, specifically its involvement in the activation of deoxycytidine analogue gemcitabine. Deficiency of DCK has been associated with gemcitabine resistance.[243] No links have been established between

treatment with flavonoids and changes in DCK expression, however there could be interactions occurring in colorectal cancer cells, as suggested by the data from the proteomics analysis.

NAP1L1 and NAP1L4 are members of the nucleosome assembly protein (NAP) family, these proteins bind to histones which are involved in the assembly of chromatin.[244] NAP1L1 has also been identified as a biomarker for the progression of CRC as mRNA expression is increased. It has roles in mediating nucleosome formation alongside cell cycle progression and proliferation.[245]

LIM and SH3 protein 1 (LASP1) promotes colon cancer cell proliferation and contributes to invasiveness in late-stage CRC tissues.[246] A potential interaction between the flavones tested and this protein could result in halted progression of CRC, if this protein was found to be involved in the mechanism of action of PMF.

Both DNAJA1 and DNAJA2, also known as heat shock protein family (hsp40) members A1 and A2, were identified in the top 20 hit proteins. These proteins are part of the family of heat shock proteins that are produced by cells in response to stress and have roles in protein assembly and trafficking.[247] DNAJA1 is involved in cancer metastasis and proliferation in colorectal and other cancers, while links between DNAJA2 and cancer are less understood.[247,248] Interactions between flavonoids and heat shock proteins have been investigated in the past, Hosokawa *et al.* examined the effect of quercetin and other flavonoids, including flavone and luteolin, on the synthesis of a selection of heat shock proteins, in HeLa and COLO320 DM (colorectal adenocarcinoma cancer) cells. Hsp40 members were not specifically investigated but the flavonoids showed the ability to supress the synthesis of heat shock proteins in these cells.[249] Luteolin also affected the expression of some heat shock proteins when studied for its antiadipogenic and pro-osteogenic effects, increasing links between the anticancer effects of flavone and potential anti-obesity effects previously mentioned.[250] The detection of these proteins as potential targets supports the findings that the flavone targeting compound could be interfering with the cells' response to stress, including effecting the UPR, as these proteins are

molecular chaperone proteins and also have involvement in protein folding and refolding when the UPR is activated.[234]

The LRRFIP1 (leucine-rich repeat in Flightless-1 interaction protein 1) protein has been shown to be involved in the promotion of metastasis and pathogenesis of CRC.[251,252] However, no links have yet been established between the impact of flavonoids on the behaviour of LRRFIP1.

Inhibition of MYC binding protein 2 (MYCBP2) by manipulation of expression levels of tumour suppressor miR-1247 resulted in methylator colon cancers when studied by Liang *et al.* suggesting that interactions with this protein could be beneficial for CRC treatment.[253]

5.5 Conclusion

This chapter aimed to isolate and identify proteins in cell lysate prepared from HCT116 cells that interact with the modified flavone analogue by using western blot and LCMS proteomics analysis of a pull down study, facilitated by the attachment of the targeting compound to a photoactive probe molecule that could form covalent links with the proximate proteins in the lysate. Proteins that were interacting in a significant way with the flavone targeting compound were identified using a competition study. This involved pre-treatment of the cell lysate with a non-tagged flavone compound to compete for key binding sites of the proteins and reduce the interaction of the probe compound.

Western blot analysis indicated that proteins in the molecular weight region of 35-40 kDa were absent from the samples pre-treated with the competition compound, indicating that proteins of this molecular weight could be potential targets for the flavone. The identification of the RAB proteins as potential key interacting proteins does not however match the molecular weight region identified here, as the molecular weight of RAB proteins is 20-30 kDa.[254]

Proteomics analysis with LCMS was carried out and identified a selection of proteins that were found to have high fold changes in samples that were treated with the probe compound when compared to a DMSO control. Significant proteins were identified by examining the fold change decreases of the interacting proteins when the sample was pre-treated with the competition compound. Proteins that

were identified regularly using multiple normalisation methods and those that were identified as significant when data was analysed using the most effective normalisation method were considered as the most important protein interactions. Due to time constraints the data was analysed in the specific way detailed in this chapter, however it is important to consider that fold change ratios are not the only indicator of significant interactions between the flavone compound and proteins in the lysate, and therefore some potential key binding proteins could be missed, and further analysis of the proteomics data could give more insight.

The nature of the pull down study carried out also has drawbacks to be considered when evaluating the outcomes, the identification of an interacting protein with this method does not give insight into the way in which the flavone compound could be influencing these proteins *in vitro* and therefore it is impossible to determine the exact way in which these compounds may be affecting the cellular processes from this data, and how the pathways discussed may be effected.

Short term future work, required to validate the findings from this study, would be to probe the pull down western blot for the proteins identified using the LCMS proteomics analysis to evaluate if these proteins decrease in intensity accordingly with the increase in competition compound concentration. Docking studies of PMF and the other modified flavone compounds in the identified proteins could also give insight into the mode in which the compounds are interacting with the proteins.

More in depth studies of the proteins pulled down in this experiment could also be carried out to potentially locate binding sites of the flavone on specific proteins, such as for the commonly occurring RAB proteins, to identify a motif or type of interaction occuring. As the pull down probe was designed with a photoactive moiety that caused formation of a covalent linkage between the proximal proteins and the probe the data in its current form does not give information as to the nature of the interactions, e.g. covalent or non-covalent linkages.

Additional insight into the activity of the flavone compound in cells, and validation of the results presented here, could be given by investigation into the effect of changes in expression or inhibition

of the interacting proteins identified. This could be carried out by examining the cellular response to changes in protein levels, such as by using DepMap or carrying out siRNA knockdown experiments.

6 Investigation into the mechanism of action of PMF by changes in RNA expression

6.1 Introduction

6.1.1 Investigating time-response of CRC to treatment with flavones to identify mechanisms of action

Widely used cytotoxicity assays typically use one measure of cellular proliferation, such as the SRB assay which measures cellular protein content. Other assays, like the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl-2H-tetrazolium bromide (MTT) assay, measure metabolic activity. These commonly used assays often only calculate cell viability at a pre-set time point, this means that the behaviour of the cells in response to compound exposure during the time interval of the assay is unknown. Measuring cell viability at different time-points after compound exposure can give more insight into the behaviour and therefore potential mechanism of action of the compound. Rapid reductions in cell viability after compound exposure can indicate that the compound is having direct cell killing results, however reduction in the number of viable cells that requires multiple doubling times can indicate that the compound is preventing growth over time.

6.1.2 Investigating changes in gene expression to identify mechanisms of action

Changes in gene expression in cells after exposure to a biologically active compound can give important insight into the behaviour of compound *in vitro* and aid in the identification of a potential mechanism of action of the compound. A common method of transcriptome analysis for drug discovery is expression microarray.[141] Investigation into the anticancer effects of PMF using expression microarray techniques was carried out by Dr Isabel L. Fong, a collaborator for this work. Changes in gene expression were studied after treatment of APC10.1 cells, a cell type developed as model for inherited CRC caused by mutations in the *Apc* gene, with PMF.[169] The results were analysed to identify pathways associated with the genes that were altered in expression. Seven of the pathways identified were

closely related to CRC: T-cell receptor signalling, Hedgehog signalling, Wnt signalling, B-cell receptor signalling, cytokine-cytokine receptor interaction, apoptosis, and Toll-like receptor (TLR) signalling. These results suggested some pathways that could be involved in the mechanism of action of the flavones, mainly PMF.[172] Microarray analysis, however, poses some drawbacks when identifying differentially expressed genes and is now viewed as an inferior technique to RNA sequencing for some projects.[141]

RNA sequencing has been established as an important tool for investigating changes in RNA in cells. This can be to identify biomarkers, or to investigate mechanisms of resistance to a compound. The technique is most often used to analyse differential gene expression of RNA in cells in different conditions.[255,256]

The most popular method for RNA sequencing uses the Illumina short-read technology. The usual workflow for this technique, specifically the workflow followed in this work, involves the initial extraction of the RNA from the cells. Selection of messenger RNA (mRNA) from the total extracted RNA follows. Ribosomal RNA (rRNA) often makes up a high percentage of the total RNA in the sample, so to effectively identify differentially expressed genes (DEGs) in the cells the mRNA needs to be examined primarily. The mRNA can be isolated from the total RNA by selecting polyadenylated (poly-A) RNAs. Poly-A tails are found on mRNA molecules at the 3' end and can be isolated using poly-T oligos attached to a substrate.[257] The RNA strands are then fragmented and complementary DNA (cDNA) is synthesised from the RNA strands, using random complementary oligonucleotides, which is then synthesised into double-stranded cDNA. The strands are then 'A-tailed' and unique identifying barcodes are added so the samples can be analysed simultaneously but identified after analysis.[257] The library is then amplified using the polymerase chain reaction (PCR) and the strands can be sequenced. The multiple steps of short read RNA sequencing can, however, lead to the introduction of imperfections and biases.[255]

Alternative methods of sequencing have also been developed and are becoming more common. Long read sequencing and direct RNA sequencing (dRNA-seq) technologies are enabling more complex problems to be addressed. These methods combat some of the problems encountered with short read sequencing, such as removing the need for mapping the short sequence reads or reducing biases creating by the use of PCR or reverse transcription.[255]

6.1.3 Chapter aims

This chapter aims to further gain insight into the potential mechanisms of action of the flavone compounds synthesised and studied in previous chapters, to further learn about the behaviour of PMF in CRC. The chapter will assess the time-response of CRC cells after treatment with a biologically active flavone to identify if the compound is likely having a cell killing or a cell growth inhibition mechanism depending on the rate at which cell viability is lost over time.

Genes that are affected by treatment of the cells with different concentrations of compound will also be assessed using an RNA sequencing experiment. The changes in gene expression can be investigated to find potential targets of the compound, to identify a potential mechanism of action. The pathways associated with these changes in gene expression can also give further insight into the effects the compound is having on the cells.

6.1.4 Materials and methods

Materials, instrumentation and experimental methods for cell culture were discussed in section 3.2.

6.1.5 Materials

Cell culture and SRB assay materials were detailed in section 3.2. RNA extraction materials were obtained from the Monarch® total RNA miniprep kit.

6.1.6 Instrumentation

Instrumentation used for recording absorbance measurements for SRB assays and taking microscope images is detailed in section 3.2. RNA purity measurements were recorded on a NanoDrop™ Spectrophotometer, detailed in section 4.2. Cell pelleting instrumentation is detailed in section 5.2.

6.1.6.1 RNA extraction kit

RNA extraction was performed using Monarch® total RNA miniprep kit according to the use manual.

6.1.7 Experimental methods

6.1.7.1 Assays evaluating the biological activity of flavones over time

To conduct research into the rate of cell growth/death after treatment with PMF and compound **4.2**, cells were identically plated on to 4 96-well plates for each compound, at standard seeding densities used for SRB assays discussed in chapters 3 and 4 (400 cells/well). Cells were allowed to adhere to the surface of the 96-well plates for 48 h and then compounds were administered at concentration ranges determined from previous SRB assays. PMF: 1.17 µM - 150 µM ; Compound **4.2**: 0.29 µM - 37.5 µM. The cells on one plate for each compound was then fixed and stained at 24 h, 48 h, 72 h and 96 h. The absorbance of SRB dye was then determined for each plate, and SRB graphs generated for each timepoint and GI_{50} values calculated. Three biological repeats of each experiment was carried out.

6.1.7.2 Determination of required seeding density for analysis of changes in RNA

To determine the required seeding density of cells in a T25 cell culture flask for the extraction of RNA for analysis, cells were trypsinised from larger flasks and placed in to IMDM stock. Cells were counted using a haemocytometer and then diluted to the required concentrations using more IMDM. Cells were seeded in to T25 cell culture flasks at: 1.25 x 10^3 cells/ml, 2.5 x 10^3 cells/ml, 5 x 10^3 cells/ml, 1 x 10^4 cells/ml, 2 x 10^4 cells/ml and 4x 10^4 cells/ml in 5 mL IMDM for each flask. The cells were allowed to adhere to the flask for 48 h and then a further 24 h to mimic the required 72 h of growth required for the RNA analysis experiment. After 72 h, the cells were trypsinised and counted again using a haemocytometer and images taken showing cell confluency for each seeding density.

6.1.7.3 Evaluation of the effect of a modified flavone analogue on RNA expression in CRC cells

6.1.7.3.1 Seeding of cell culture flasks at the required seeding density and treatment of cells with required concentrations of compound

After determination of the required seeding density for the RNA analysis experiment, the correct number of cells was counted and seeded in to T25 flasks in 5 mL of IMDM. Cells were allowed to adhere to the surface of the cell culture flasks for 48 h and then compounds were administered according to Table 6.1. After 24 h exposure to the compound, the flasks were trypsinised and the cells collected. Cells were also collected from the original media alongside those that were adhered to the surface of the flask.

Table 6.1: Treatment conditions of cells for analysis of the RNA expression after treatment with modified flavone analogue, **4.2**.

Flask	Compound	Concentration		Vehicle control
A	-	0 µM		0.3 % DMSO
B	**4.2**	1 x GI_{50}	3.68 µM	0.3 % DMSO
C	**4.2**	3 x GI_{50}	11.07 µM	0.3 % DMSO

6.1.7.3.2 Production of cell pellets for the extraction and analysis of RNA expression

Cells collected in IMDM after treatment were then pelleted. The pelleting protocol followed was described in section 5.2.3.1 and the same for this experiment.

6.1.7.3.3 Extraction of RNA from cell pellets after treatment

RNA was extracted from cell pellets using the Monarch® total RNA miniprep kit. Cell pellets were thawed and resuspended in RNA lysis buffer (300 µL). Samples were placed in to gDNA removal columns and centrifuged at 16,000 xg for 30 seconds. The flow through was collected and the column discarded. An equal volume of ethanol was added to the mixture, and it was transferred to an RNA purification column and centrifuged at 16,000 xg for 30 seconds. The flow through was discarded. RNA wash buffer (500 µL) was added to the column and the column centrifuged at 16,000 xg for 30 seconds. DNase I in reaction buffer (80 µL) (made according to kit instructions) was added to the column and

incubated for 15 minutes at room temperature. RNA priming buffer (500 µL) was added to the column and centrifuged at 16,000 xg for 30 seconds and the flow through was discarded. RNA wash buffer (500 µL) was added to the column and centrifuged at 16,000 xg for 30 seconds, the flow through was discarded. RNA wash buffer (500 µL) was added for a second time and the columns centrifuged at 16,000 xg for 2 minutes, the flow through was discarded. Nuclease-free water (100 µL) was added to the centre of the column matrix and the column centrifuged at 16,000 xg for 30 seconds. The flow through was collected. Samples were stored at -80 °C until use.

6.1.7.3.4 Purity analysis of RNA after extraction from cell pellets

Absorbance ratios for each RNA sample were recorded using a NanoDrop™ Spectrophotometer, ratios of absorbance at 280 nm, 260 nm and 230 nm were recorded by recording a water blank and then measuring 2 µL of each sample.

6.1.7.3.5 Data analysis of RNA expression sequencing data processed by UCL

6.1.7.3.5.1 Sequencing of RNA samples carried out by UCL

Detailed sample preparation and sequencing methods used by UCL can be found in Appendix 8.53. RNA from the samples was quantified using an Aligent 4200 Tapestation system. The RNA for each sample (500 ng) was processed using a Kapa mRNA HyperPrep kit. Sequencing samples were quantified using the Qubit High Sensitivity DNA assay and normalised to 10 nM. Equal volumes of each library were pooled and re-quantified. Samples were sequenced using NextSeq 2000 instrument. After sequencing samples were demultiplexed and the quality of the RNA sequencing was evaluated using numerous methods and complied into a MultiQC report (Appendix 8.54).[258]

6.1.7.3.5.2 Analysis of RNA sequencing results provided by UCL

Investigation into the differentially expressed genes identified with by the RNA sequencing experiment was provided by UCL in the form of a statistical report generated using DESeq2 and SARTools R (Appendix 0).[259]

6.1.7.3.5.3 Further analysis of RNA sequencing results

Differentially expressed features identified from the RNA sequencing experiment were analysed using Excel, iVenn[260], Metascape[227] and STRING[261] software tools.

6.2 Results

6.2.1 Time dependent activity of flavones on HCT116 cells

CRC cells were treated with PMF and the boc-modified flavone analogue, **4.2**, to assess the rate at which the compounds caused cell death or growth inhibition. Investigation into the growth response of the cells over time can give insight to the mechanism of action of the compounds and indicate if the compound is slowing cell growth or causing cell death. Compound **4.2** was selected as a comparison compound for PMF due to the improved solubility characteristics and lower GI_{50} values than the unmodified PMF and other flavone analogues, discussed in Chapter 4. SRB assays were used to evaluate cell growth, as described in Chapter 3, at time intervals of 24 h to 96 h, at the required concentration range, and GI_{50} were values calculated for each interval. The overlay of the dose response curves shows the change in response of the cells, measured by protein content with SRB assay, to the different concentrations over time (Figure 6.1). Assays for each compound were completed in triplicate (Appendix 8.56).

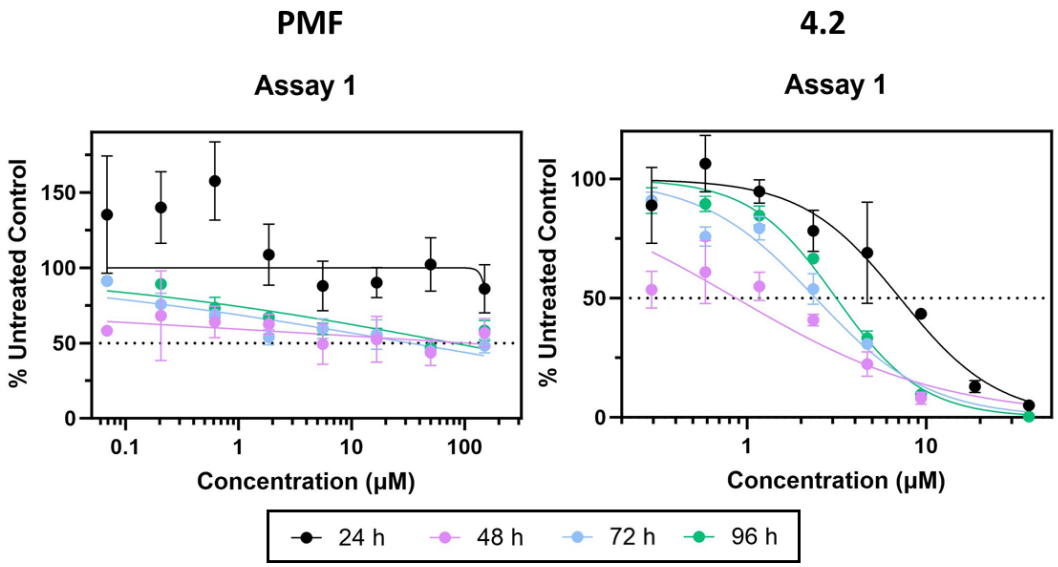


Figure 6.1: Overlay SRB graphs of time interval assay 1 for PMF and compound **4.2** showing changes in response to compounds over time. Concentration range: PMF 1.17 μM – 150 μM; Compound **4.2** 0.29 μM – 37.5 μM. Cell density: 400 cells/well. Each data point is produced from the mean of at least 3 data points and error bars of 1 standard deviation above and below the mean are included. Graphs for other biological repeats of this experiments can be found in Appendix 8.56.

Treatment with compound **4.2** again resulted in much lower GI_{50} values, similar to those seen before. Table 6.2 shows the calculated GI_{50} values for both PMF and compound **4.2** at each time point. The graphs, however, show inconsistent results across the three assays. SRB results using the flavones have been shown to be inconsistent in previous studies described in Chapter 3.

Table 6.2: GI_{50} values calculated for both PMF and compound **4.2** at each time point.

	GI_{50} (μM)											
Compound	**Assay 1**				**Assay 2**				**Assay 3**			
	24 h	**48 h**	**72 h**	**96 h**	**24 h**	**48 h**	**72 h**	**96 h**	**24 h**	**48 h**	**72 h**	**96 h**
PMF	171.5	103.9	32.9	79.5	2.9	227.6	8.5	17.9	155.8	6.6	80.4	104.9
4.2	7.2	0.9	2.4	3.2	3.4	0.7	2.7	2.5	0.8	2.5	1.7	2.9

It was expected that a certain timepoint would result in the lowest GI_{50} value, indicating the highest level of activity of the compound occurring before this point. Due to the variability, averages were calculated of the three assays to gain a better understanding of the behaviour of the compounds over

time (Figure 6.2). The averaged results for PMF show that the lowest GI_{50} value is seen for the 72 h time point, this indicates that the compound is not causing immediate cell killing effects on exposure. The cells appear to require multiple doubling times for the compound to reduce cell growth to this level. For compound **4.2**, the lowest GI_{50} value was shown to be at 48 h. This indicates that the compound was exerting a higher level of activity during an earlier time period than PMF, which is in line with the lower overall GI_{50} values seen for this compound, compared to the other modified flavones and PMF.

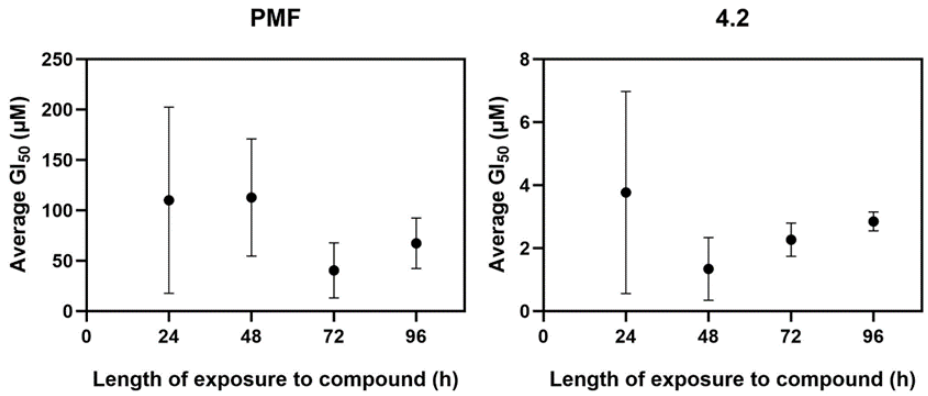


Figure 6.2: Average GI_{50} values of each time point for PMF and compound **4.2** showing error bars of one standard deviation for each time point.

6.2.2 Analysis of differential gene expression after treatment using RNA sequencing

6.2.2.1 Determination of correct seeding density for RNA extraction experiment

To determine an appropriate seeding density of cells for the extraction of RNA, an experiment was carried out to compare seeding densities. In order to extract the required quantity of RNA to allow for sequencing, the cells were grown and treated in tissue culture (T25) flasks instead of wells in a 96-well plate. Six seeding densities were evaluated after 72 h to assess the confluency and cell count of the cells in the T25 flaks and determine a suitable seeding level (Table 6.3).

Table 6.3: Seeding densities of HCT116 cells in T25 flasks trialled to determine the appropriate density for RNA extraction.

Flask	Cells seeded	Cells after 72 h (counted)
1	1.25 x 10^3 cells/mL *6,250 cells*	2.8 x 10^4 cells/mL *154,000 cells*
2	2.5 x 10^3 cells/mL *12,500 cells*	2.8 x 10^4 cells/mL *154,000 cells*
3	5 x 10^3 cells/mL *25,000 cells*	10.6 x 10^4 cells/mL *583,000 cells*
4	1 x 10^4 cells/mL *50,000 cells*	N/A
5	2 x 10^4 cells/mL *100,000 cells*	4.42 x 10^5 cells/mL *2,431,000 cells*
6	4 x 10^4 cells/mL *200,00 cells*	8.56 x 10^5 cells/mL *4,708,000 cells*

Cells were seeded at densities ranging from 1.25 x 10^3 cells/mL in 5 mL to 4 x 10^4 cells/mL in 5 mL of IMDM and incubated for growth for 72 h, after this the cells were trypsinised and counted. Images were taken showing the confluency of the cells after the 72 h to determine which seeding density resulted in the most optimum confluency of the cells after this time (Figure 6.3). Cells should be approximately 80% confluent and not overly in contact as this can inhibit growth.[262]

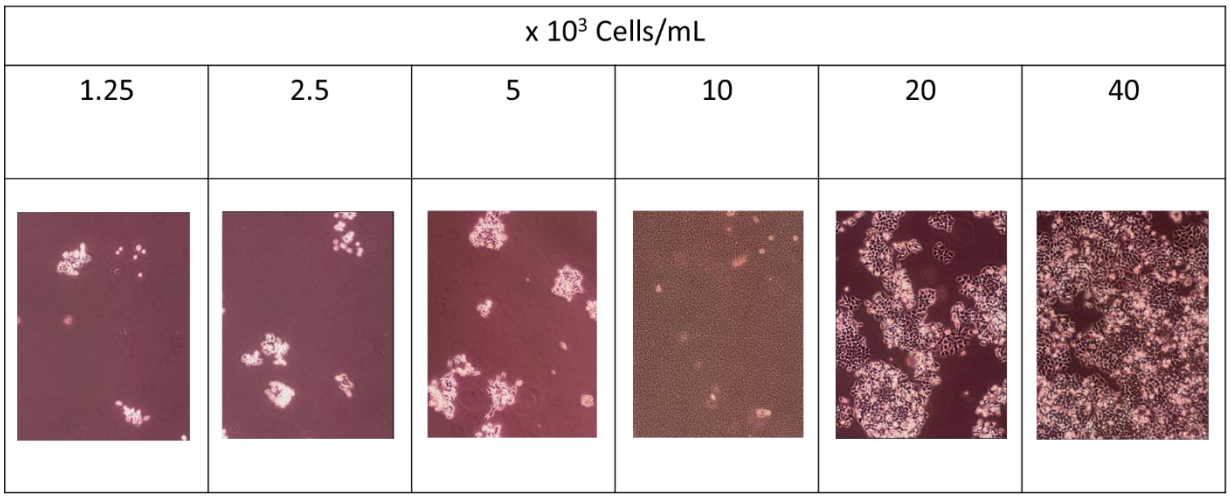


Figure 6.3: Microscope images of cell confluency after 72 h of growth at different seeding density levels measured at 40x magnification. Scale bars not available.

In the experiment carried out, flask 4, seeded at 1 x 10^4 cells/mL, was found to be contaminated at the 72 h timepoint, therefore was discounted from the experimental results. Flask 5, seeded at 2 x 10^4 cells/mL, was found to have the desired level of confluency after 72 h and was selected as the optimum seeding density for the RNA extraction experiment.

6.2.2.2 Evaluation of RNA quality after extraction from cell pellets

6.2.2.2.1 Assessment of extracted RNA quality using a NanoDrop™ Spectrophotometer

Cells were seeded and treated with the required compound or DMSO control and then pelleted as described in section 6.1.7. The RNA was then extracted from the pelleted cells using the Monarch® total RNA miniprep kit. After extraction, the quality of the RNA can be assessed using a NanoDrop™ spectrophotometer. The $A_{260/280}$ and $A_{260/230}$ ratios recorded using the spectrophotometer indicate the quality of the extracted RNA.[263] These ratios are calculated from the absorbance maximum at 260 nm divided by the absorbance maximum at 280 nm and 260 nm divided by the absorbance at 230 nm. The $A_{260/280}$ ratio indicates contamination from phenol, guanidine or other extraction reagents in the sample and the ratio is found in the range of 1.9-2.1 if the sample is pure. The $A_{260/230}$ ratio indicates the presence of unwanted chemicals in the sample such as phenol, salt, peptides or carbohydrates and has the desired range of 2.0-2.2 for pure samples. The nine samples produced for the sequencing experiment were assessed using the spectrophotometer and the results of the ratios calculated (Table 6.4). The values shown in red for the $A_{260/230}$ ratio were found to be outside the desired range but deemed permissible as they were not above 2.3 in value.

Table 6.4: RNA purity ratios calculated using a spectrophotometer for the samples prepared for RNA sequencing. Values in red are found to be outside of the desired range.

Sample	$A_{260/280}$	$A_{260/230}$
1a	2.082	*2.220*
1b	2.062	2.194
1c	2.078	2.111
2a	2.058	2.126
2b	2.061	2.155
2c	2.058	2.022
3a	2.066	2.073
3b	2.075	2.180
3c	2.078	*2.217*

6.2.2.2.2 MultiQC analysis of sequenced samples

After sequencing of the provided samples was performed, analysis of the quality of the data was performed by UCL. This analysis was compiled to produce an mRNA-Seq Quality control report (MultiQC) report (Appendix 8.54). The report provided general statistics of sample quality (Table 6.5). The results show low rRNA contamination for all samples and a value of approximately 30% duplication rate for all samples, the number of reads for the samples was also sufficiently high. These statistics indicate that the sequencing of the samples was a good quality and the data could be carried forward for analysis and interpretation.

Table 6.5: General statistics of RNA sequenced sample quality.

Sample Name	M Reads Mapped	% rRNA	dupInt	% Dups	5'-3' bias	M Aligned	Error rate	M Non-Primary	M Reads Mapped	% Mapped	% Proper Pairs	M Total seqs	% Aligned	% GC
EF1	62.5	0.00%	0.11%	30.3%	1.15	28.0	0.29%	6.5	56.0	100.0%	100.0%	56.0	89.4%	26.0
EF2	67.8	0.00%	0.10%	31.7%	1.15	30.3	0.30%	7.2	60.7	100.0%	100.0%	60.7	89.2%	28.1
EF3	66.5	0.00%	0.10%	33.0%	1.20	29.7	0.27%	7.1	59.4	100.0%	100.0%	59.4	89.1%	27.5
EF4	48.3	0.00%	0.11%	28.8%	1.20	21.7	0.24%	4.8	43.5	100.0%	100.0%	43.5	90.1%	20.2
EF5	52.7	0.00%	0.10%	28.3%	1.22	23.6	0.25%	5.5	47.1	100.0%	100.0%	47.1	89.9%	21.9
EF6	56.0	0.01%	0.10%	30.9%	1.19	25.1	0.25%	5.8	50.1	100.0%	100.0%	50.1	89.8%	23.3
EF7	67.4	0.00%	0.11%	30.2%	1.15	30.2	0.23%	7.0	60.4	100.0%	100.0%	60.4	90.4%	28.0
EF8	62.0	0.00%	0.09%	29.1%	1.18	27.8	0.24%	6.4	55.6	100.0%	100.0%	55.6	90.3%	25.8
EF9	66.0	0.00%	0.11%	33.9%	1.18	29.5	0.26%	7.0	59.1	100.0%	100.0%	59.1	89.5%	27.4

6.2.2.3 RNA expression sequencing data after treatment of CRC cells with compound

Differential expression analysis of the genes was performed using DESeq2 and a report was produced using SARTools R package.[259] Each sequence identified was assigned a unique identifier/feature ID and features that were differentially expressed between sample conditions were analysed. 57773 unique features were identified in this analysis. Firstly, the data was normalised then the differential expression of genes between the DMSO, 1 x GI_{50} and 3 x GI_{50} sample conditions was analysed. The samples were assigned sample numbers for each of the three biological repeats of the three conditions (Table 6.6).

Table 6.6: Sample names assigned to the nine samples submitted for RNA sequencing by UCL.

Sample	Treatment condition	Sample name
1a	DMSO	EF1
1b	1 x GI_{50}	EF2
1c	3 x GI_{50}	EF3
2a	DMSO	EF4
2b	1 x GI_{50}	EF5
2c	3 x GI_{50}	EF6
3a	DMSO	EF7
3b	1 x GI_{50}	EF8
3c	3 x GI_{50}	EF9

6.2.2.4 Raw data visualisation and normalisation

The total number of mapped and counted reads for each sample is shown in Figure 6.4, samples EF4, EF5 and EF6 showed lower overall total read counts than the repeat samples in the same treatment condition. Some variation of total read count for raw data is to be expected between replicates.

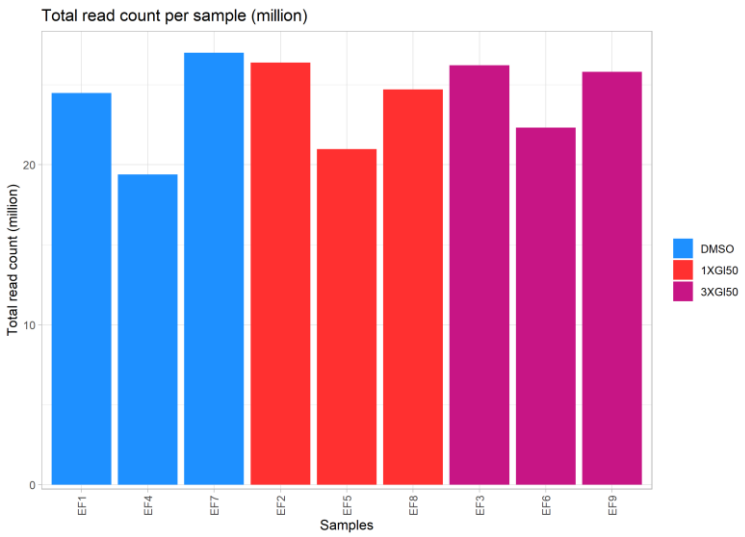


Figure 6.4: Total read count of each sample indicating each unique feature identified by RNA sequencing.

Some identified features will have no read count in other samples, this is expected to be similar for samples in the same treatment condition. Features with no read count in all nine samples were not included in the DESeq2 analysis, for this data that value was found to be 30,752 features (53.23%) of the total count (Appendix 8.57). In some cases, a single or a few features can account for a high proportion of the reads (up to or above 20%) (Figure 6.5). The results show that for all samples the most expressed sequence is the same: ENSG00000210082_MT-RNR2. In all samples this accounts for a low percentage of the total read count (< 2.2%).

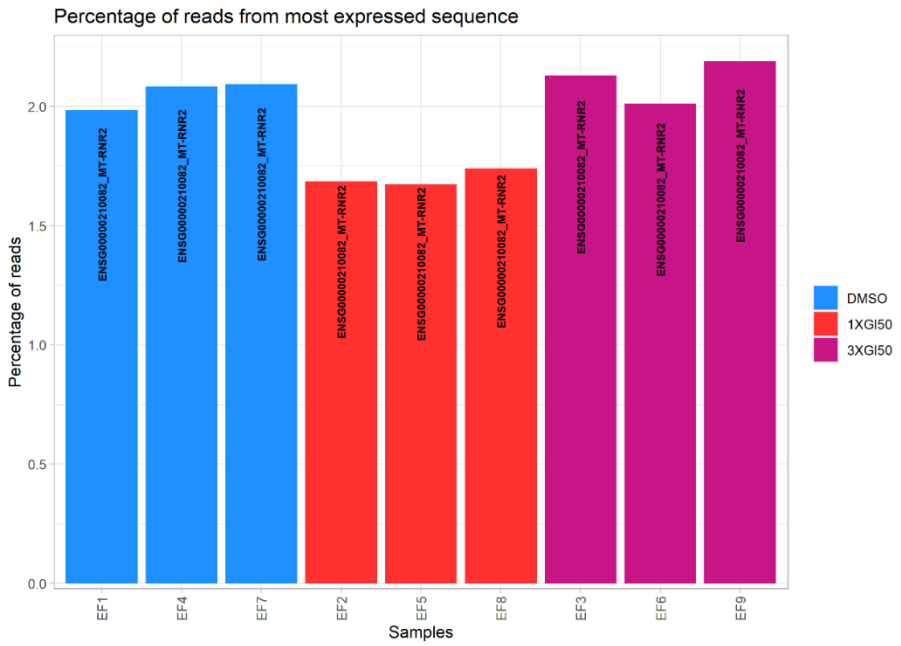


Figure 6.5: Percentage of reads associated with the sequence having the highest count (provided in each box on the graph) for each sample.

A pairwise scatterplot was produced to assess the similarities between samples, both from replicates of the same experiment and between sample treatment conditions (Figure 6.6). The similarity index values were calculated as a SERE statistic.[264] Values of 0 indicate that samples are identical, similarity index values of 1 indicate technical repeats and higher values indicate biological variability which is expected to be higher between samples of different treatment conditions.

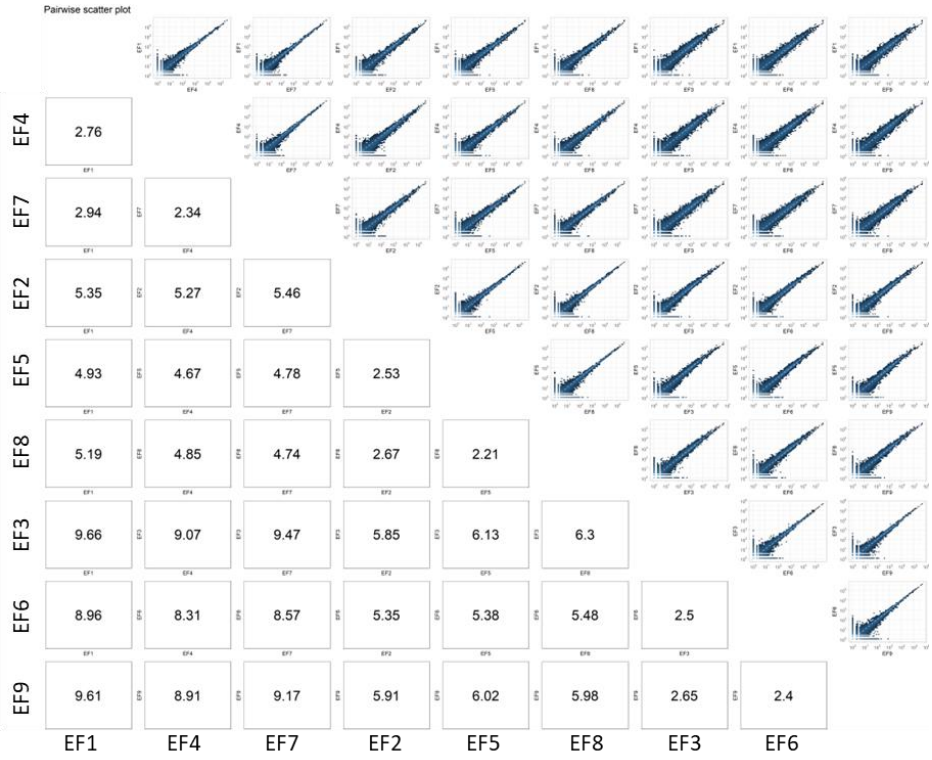


Figure 6.6: Pairwise scatter plot showing similarity index values calculated as SERE statistics between samples.

The scatter plot shows, as expected, that there is a higher level of variation between samples of different treatment conditions, and lower level of variation for samples that are replicates of the same treatment.

Sample variability was also assessed using hierarchical clustering and a principal component analysis (Appendix 8.58). The hierarchical clustering of the samples shows that the samples are clustered based on treatment condition as expected, it also shows that the samples in treatment conditions 1 x GI_{50} and 3 x GI_{50} are the most similar, suggesting that the treatment with the compound at 1 x GI_{50} resulted in similar outcomes to the samples treated at 3 x GI_{50} and differ more significantly than those treated only with DMSO. The principle component analysis shows that the first principle component separated samples by treatment condition, indicated by the clustering of the samples on the x-axis, which is representative of 62.48% of the analysis.

The data was then normalised to allow the samples to be directly compared. Scaling factors were calculated during the DESeq2 analysis using the function locfunc='median', and feature counts were normalised by dividing each value by the scaling factor for that sample (Table 6.7).

Table 6.7: Scaling factors calculated using DESeq2 analysis using the function locfunc='median', raw values are divided by the scaling factor to normalise the samples.

	EF1	EF4	EF7	EF2	EF5	EF8	EF3	EF6	EF9
Size factor	1.09	0.82	1.15	1.13	0.89	1.05	1.05	0.9	1.05

The normalisation of the samples was validated by a comparison of the count distribution box plots (Figure 6.7). The plots show that after the application of the normalisation the count distributions are much more stable and the averages of the samples are in line, this validates the normalisation process and the scaling factors used.

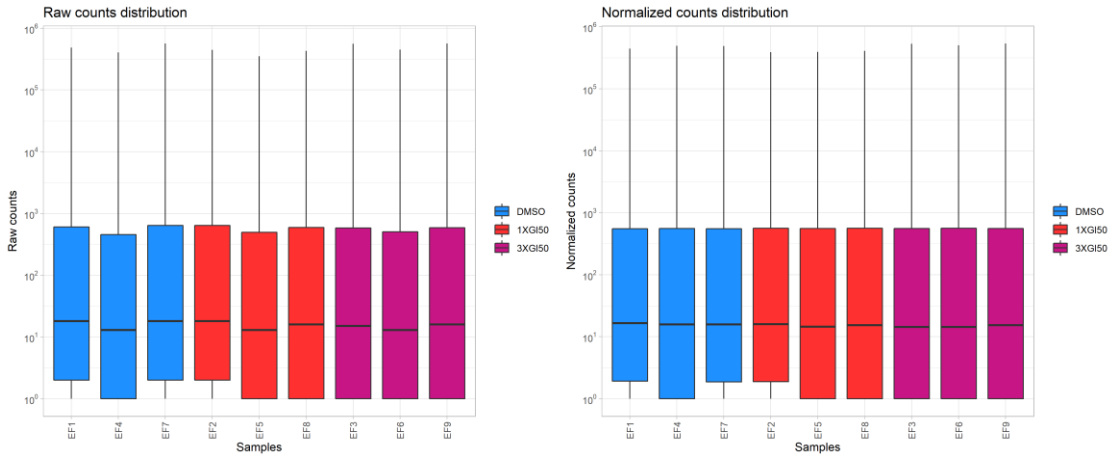


Figure 6.7: Box plots showing the count distribution of each sample before and after normalisation. The plots show that the distributions are more stable after the normalisation process.

After the normalisation process the differential expression of the features could be examined. The results underwent a process of independent filtering, to improve the detection power of the differential expression analysis. The filtering process defines a threshold on the mean of the

normalised counts and identifies features that have very low counts to discard from the data (Table 6.8).

Table 6.8: Number of discarded features using independent filtering from the base mean threshold of each treatment condition.

Test vs Ref	BaseMean Threshold	# discarded
1 x GI_{50} vs DMSO	7.28	43064
3 x GI_{50} vs DMSO	3.38	41587
3 x GI_{50} vs 1 x GI_{50}	17.3	44542

The number of differentially expressed features for each condition comparison was calculated, these were also divided into genes that were increased in counts and those that were decreased in counts (Table 6.9).

Table 6.9: Number of up-, down- and total number of differentially expressed features for each comparison.

	Number of differentially expressed features		
Comparison condition	**Down**	**Up**	**Total**
1 x GI_{50} vs DMSO	2017	2151	4168
3 x GI_{50} vs DMSO	3790	3794	7584
3 x GI_{50} vs 1 x GI_{50}	2045	1954	3999

MA-plots were generated to visualise the DEGs for each comparison condition (Figure 6.8). The plots show log fold-changes as a function of the mean intensity for each feature. Larger log fold-changes were seen for the comparison of the DMSO vs 3 x GI_{50} condition indicating a larger change in counts for the features assessed.

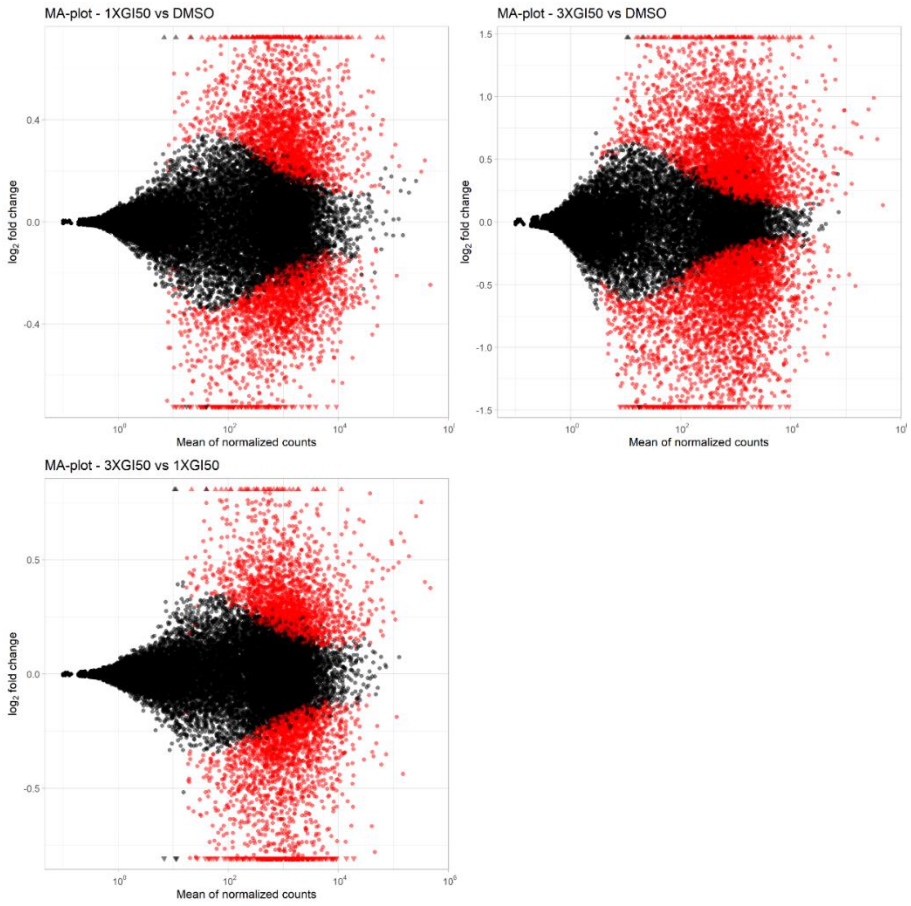


Figure 6.8: MA-plots showing differentially expressed genes for each comparison condition. Differentially expressed features are shown in red and triangles represent to features having log_2(FC) values outside of the axes.

6.2.2.5 Analysis and interpretation of RNA sequencing data.

To identify the differentially expressed genes that are the most significantly affected by the treatment with the compound, 2-fold change cut-offs for the changes in expression were applied to narrow the results. Genes that were significantly upregulated were identified using a fold change cut-off of 2 (log_2(FC) ≥ 1) and genes that were significantly downregulated were identified using a cut-off of 0.5 (log_2(FC) ≤ -1) (Table 6.10). A p-value cut-off of 0.05 was also applied to the results.

Table 6.10: Fold change and Log_2(FC) limits determined for the differential expression analysis of RNA sequencing results to identify significantly affected genes.

Change in gene regulation	Fold change limit	Log_2(FC) limit
Upregulated genes	2	1
Downregulated genes	0.5	-1

After application of these fold change cut-offs, the number of significantly up- and downregulated genes was determined (Table 6.11). More genes were significantly differentially expressed for the 3 x GI_{50} treatment condition, this is expected as the higher concentration of compound will likely have a larger impact on the RNA in the cells. The lists of differentially expressed features for each condition can be found in Appendix 8.59.

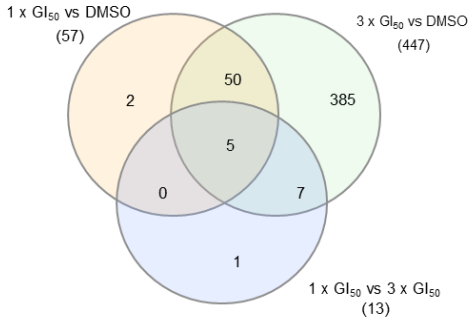
Table 6.11: Number of significantly differentially expressed genes identified after application of fold change limits for each comparison condition.

	Number of differentially expressed features		
Comparison condition	**Down**	**Up**	**Total**
1 x GI_{50} vs DMSO	48	57	105
3 x GI_{50} vs DMSO	603	447	1050
3 x GI_{50} vs 1 x GI_{50}	74	13	87

To identify up- and downregulated genes of interest amongst the data, the lists of significantly differentially expressed features were compared using iVenn software.[260] The features that were common to the three comparison condition lists for both up and downregulated genes were determined to be the most important for consideration (Figure 6.9). It should be noted that of the six commonly decreased genes one is identified as KRT23 which codes for a keratin protein. Identification of keratins in the experimental results is likely to be due to contamination in sample preparation. Features that were only common to two comparison conditions were also determined (Appendix 8.60-8.61).

A)

Features increased compared to reference:

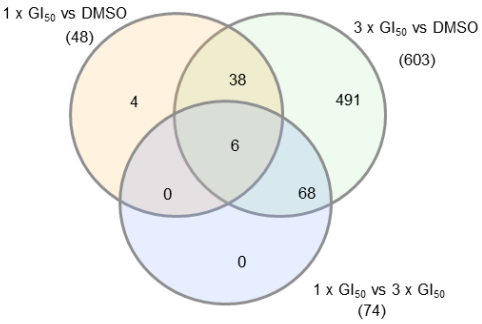


5 common features to all conditions:

Feature	Fold change values		
	1 x GI_{50} vs DMSO	3 x GI_{50} vs DMSO	1 x GI_{50} vs 3 x GI_{50}
VGF	3.40	7.375	2.12
SLFN5	2.595	6.749	2.512
HIST1H2BD	2.49	7.297	2.618
HIST1H2AC	2.35	7.933	2.975
DNAJB9	2.116	5.376	2.357

B)

Features decreased compared to reference:



6 common features to all conditions:

Feature	Fold change values		
	1 x GI_{50} vs DMSO	3 x GI_{50} vs DMSO	1 x GI_{50} vs 3 x GI_{50}
KITLG	0.338	0.161	0.483
KRT23	0.373	0.116	0.41
GPR110	0.375	0.127	0.363
PCDH7	0.393	0.132	0.397
CYP24A1	0.414	0.157	0.399

Figure 6.9: Comparison of significantly differentially expressed features for all treatment conditions. A) Venn diagram shows the number of features increased when compared to the reference treatment condition and the number of common features. The table shows the feature name determined by the sequencing experiment and the fold change values for each comparison. B) Venn diagram shows the number of features decreased when compared to the reference treatment condition and the number of common features. The table shows the feature name determined by the sequencing experiment and the fold change values for each comparison.

The Metascape tool was used to analyse the genes identified as up- and downregulated by treatment with compound **4.2**.[227] The tool was used to explore the biological pathways associated with the genes identified. The genes altered by treatment of the cells with the 3 x GI_{50} concentration of the compound were studied with the Metascape tools, as the difference in the genes between the cells treated with the DMSO control and the cells treated with compound **4.2** is greatest for this comparison. Analysis was carried out for the upregulated genes, downregulated genes and the complete set of differentially

expressed genes to find which pathways are the most significantly affected by the treatment (Figure 6.10, Appendix 8.62). For the increased differentially expressed genes, the pathway analysis showed that the genes in the set were involved in neutrophil degranulation, regulation of catabolic processes and response to nutrient levels as the most significant pathways associated with the genes. Also worthy of note are the pathways that show similarities to the pathways identified for the proteomics experiments discussed in Chapter 5, such as autophagy, unfolded protein response (UPR) and cellular response to hypoxia.

The pathway analysis for the decreased DEGs showed a more skewed result, in that the percentage of genes in the geneset for the top hits in this analysis was much higher than for the increased DEGs. Also, many of the top identified pathways are all very similar: cell cycle, mitotic cell cycle and positive regulation of cell cycle processes. Appendix 8.62 shows the pathways associated with the both the increased and decreased DEGs as some parts of the pathway can be affected in different ways by treatment with the compound.

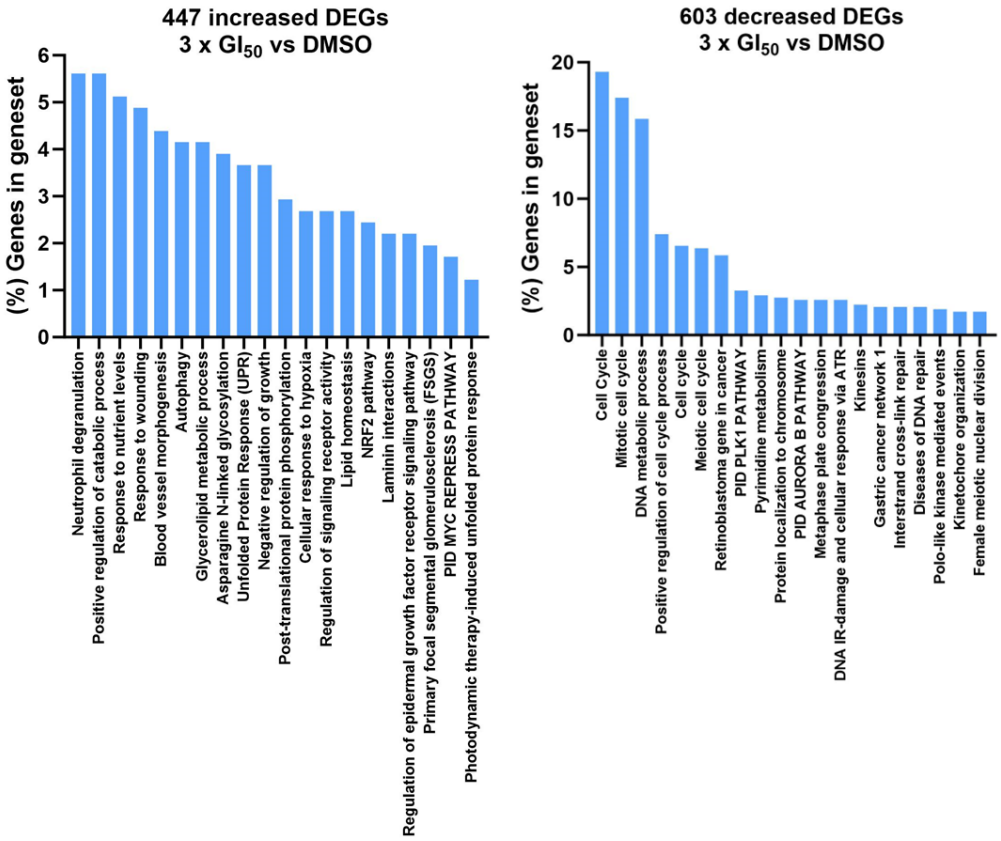


Figure 6.10: Top 20 enriched genesets, identified using Metascape analysis, for the up- and downregulated genes identified as significant for the 3 x GI_{50} vs DMSO comparison.

The STRING: protein-protein interaction networks functional enrichment analysis tool was also used to examine the way that the identified DEGs interact with each other and the biological processes that they are involved in.[261] The lists of increased and decreased DEGs identified for the 3 x GI_{50} vs DMSO comparison were analysed by the STRING tool to examine the networks. The interaction networks for the DEGs can be found in Appendix 8.62.1-8.63. The images show that there are a significant number of edges (lines depicting the interaction of two nodes/genes) linking in the genes in the networks. For the decreased DEGs this was found to be 625 compared to 358 that would be expected for a network of the same size of random proteins, and 8992 compared to an expected 1717 for the decreased DEGs. This suggests that there are some pathways or biological processes that are being specifically targeted by treatment with the compound.

Table 6.12 shows the biological processes identified with the STRING analysis that were associated with the increased DEGs. The table shows that for the increased DEGs the three biological processes identified include response to unfolded protein and response to endoplasmic reticulum stress. The false discovery rate is calculated as a p-value indicating how significant the enrichment is for the process. These processes are in line with the results seen for the proteomics analysis of the pull down study discussed in Chapter 5.

Table 6.12: Biological processes identified as functional enrichments when examining the upregulated features identified in the 3 x GI_{50} vs DMSO comparison. Detailed descriptions of the table headings can be found in Appendix 8.64 and were obtained from data analysis using STRING.[261]

Term description	Observed gene count	Background gene count	Strength	False discovery rate
Response to unfolded protein	14	122	0.76	0.0073
Catabolic process	71	1976	0.26	0.0094
Response to endoplasmic reticulum stress	18	223	0.61	0.0094

For the decreased DEGs the number of related biological processes identified was much higher, the top 10 are shown in Table 6.13, sorted by false discovery rate. Appendix 8.64 shows the other processes identified. These results are similar to the pathways identified using the Metascape pathway analysis, showing that the DEGs are strongly involved in the cell cycle.

Table 6.13: Top 10 biological processes identified as functional enrichments when examining the downregulated features identified in the 3 x GI_{50} vs DMSO comparison. Detailed descriptions of the table headings can be found in Appendix 8.64 and were obtained from data analysis using STRING.[261]

Term description	Observed gene count	Background gene count	Strength	False discovery rate
Cell cycle	166	1246	0.67	9.06E-57
Cell cycle process	122	835	0.71	1.51E-43
Mitotic cell cycle	103	631	0.76	6.94E-40
Mitotic cell cycle process	93	537	0.78	1.55E-37
Cell division	84	527	0.75	3.06E-31
Chromosome segregation	64	286	0.89	1.84E-30
Regulation of cell cycle process	95	716	0.67	3.33E-30
Nuclear chromosome segregation	57	229	0.94	5.06E-29
Nuclear division	64	323	0.84	5.63E-28
Organelle fission	65	349	0.81	3.92E-27

6.3 Discussion

6.3.1 Time dependent activity of flavones on HCT116 cells

The results of the averaged GI_{50} values calculated after treatment of the compounds with PMF and compound **4.2** show that the lowest GI_{50} value timepoints were 72 h and 48 h respectively. Both compounds showed that greater than 24 h was required to result in low GI_{50} values. HCT116 cells have a doubling time of approximately 18 h, indicating that the cells required exposure to the compounds for multiple doubling times to result in a reduction in viable cells able to be detected with the assay.[265] If the compounds were directly triggering cell death, in a manner similar to traditional anticancer drug cisplatin, the cells would require shorter timespans of compound exposure, suggesting that the compounds may only be inhibiting cell growth or proliferation.[266] This indicates that the mechanism of action of the flavones is involved in preventing cell growth. Treatment with compound **4.2** resulted in the lowest GI_{50} value on average at the 48 h timepoint, which is earlier than the 72 h timepoint for treatment with PMF. GI_{50} values overall were found to be lower for treatment with compound **4.2**, this has been discussed as potentially due to the increased aqueous solubility characteristics of the compound, however when considering that the cell proliferation inhibition activity of this compound appears to be superior to that of PMF, it may indicate that the structural changes of the compound

may have also improved its *in vitro* properties and resulted in a stronger proliferation inhibition mechanism.

6.3.2 Discussion of RNA expression sequencing data

6.3.2.1 Discussion of RNA sample quality assessment and normalisation

RNA was extracted from the HCT116 cells after treatment with DMSO, 1 x GI_{50} or 3 x GI_{50} concentrations of compound **4.2**. The RNA quality was assessed using $A_{260/280}$ and $A_{260/230}$ ratios calculated using spectrophotometer measurements. The measurements showed that the extracted RNA was of good quality and the assessment carried out by UCL and shown in the MultiQC report supports this by showing low levels of contamination of the samples. Normalisation methods were effective in ensuring the samples were comparable, so the number of differentially expressed samples could be effectively analysed.

6.3.2.2 Identification of the differentially expressed genes when comparing different RNA samples

Sequencing of the samples resulted in the identification of many unique features, a significant proportion (53.23%) of these were found to exhibit no read count for all samples and therefore were removed from the analysis of differentially expressed genes. However, despite the removal of these, the number of features identified remained high, ensuring that the sequencing could provide sufficient insight into the effects of treatment with the compound.

Genes that were altered by the treatment of the cells with the two doses of the compound were identified by comparing the features in different treatment conditions. To allow effective analysis of the data, which showed high numbers of differentially expressed genes when the treatment conditions were compared, only the genes with significant fold changes from DMSO treatment to the compound treatment were considered. Genes that were more significantly altered by the treatment with the compound are likely to exhibit the largest fold changes, compared to the DMSO control. It is possible, however, that some genes that are important to the mechanism of action of the compound and the other flavones, including PMF, may be directly impacted however the fold changes detected may fall

outside of the cut-off range. More thorough analysis of the data obtained could aid in identifying any of these potential target genes.

6.3.2.3 Identification of significant DEGs

Using the Venn diagrams, genes that were common to all the comparison conditions were identified, these were classed as the genes that were most worthy of note. For each of the features common to all three comparison conditions the fold change values were in line with expectations. For the selected increased features, the fold change values increased when comparing the 1 x GI_{50} vs DMSO condition to the 3 x GI_{50} vs DMSO condition, showing a higher level of this feature was detected in the 3 x GI_{50} samples, indicating a dose-dependent response of the genes to the treatment with the compound. For the selected decreased features, the fold change values were lower for the 3 x GI_{50} vs DMSO condition compared to the 1 x GI_{50} vs DMSO condition, showing that the amount of this gene detected was lower when treating the cells with a higher concentration of compound, again supporting a dose-dependent response. The fold change values seen for these common features are also of a particularly large size, as the upregulated features have fold change values ranging 5.376 - 7.933 for the 3 x GI_{50} vs DMSO comparison. Compared to the cut-offs applied that determined a fold change of ≥ 2 to be a significant change these numbers are of note. For the downregulated features the fold change values range from 0.116 – 0.161 for the 3 x GI_{50} vs DMSO comparison, when the level classed as significant is ≤ 0.5. These values suggest that the compound is resulting in substantial impacts on the behaviour of these genes in the cell.

The five common upregulated genes identified/ features, that were increased in comparison with the reference condition, included two HIST1H2 features. The HIST1H2AC gene encodes for the histone H2A type 1-C protein and HIST1H2BD encodes for the histone H2B type 1-D protein, both fall in families of core histones and are involved in forming the structure of the nucleosome and are involved in the folding of DNA.[267,268] Histones are regulated by the cell cycle and vary greatly in number during different parts of the cell cycle, identifying that the compound caused a significant increase in the detection of

this gene is in support of the Metascape analysis of the decreased genes that showed a strong association with the cell cycle pathways.[267]

DNAJ homolog subfamily B member 9, encoded for by the DNAJB9 gene, is notable due to the identification of two DNAJ homolog subfamily members using the proteomics analysis experiments discussed in Chapter 5. The DNAJ protein family members act as co-chaperones to heat shock protein 70s. DNAJB9 is localised to the endoplasmic reticulum (ER) and is thought to be upregulated in situations of ER stress and therefore likely involved in protein folding and degradation responses.[269]

KIT proto-oncogene ligand (KITLG) was identified as a commonly decreased feature in all the comparisons. This ligand is a target of miR-34c which is known to be silenced in CRC, overexpression of miR-34c has shown to induce apoptosis and inhibit cell proliferation by silencing the KITLG factor.[270] Decrease in expression of the KITLG factor by compound **4.2**, shown by these results, could indicate a mechanism of action of the compounds anti-colorectal cancer effects.[270,271]

The GPR110 gene encodes for the G-protein coupled receptor (GPCR) 110 protein as a part of the adhesion-GPCR family. The GPCR proteins are responsible for a range of functions, involving activating Rho GTPases. This strengthens the associations seen in the pull down experiments between the activity of the flavone compounds and the GTPases.[272]

The transmembrane receptor protocadherin 7, coded for by the PCDH7 gene, has been shown to be involved in cell apoptosis, proliferation and invasion. Results from the experiments shown in this chapter indicate that PCDH7 was decreased dose-dependently in the HCT116 cells after treatment with compound **4.2**. Previous studies by Liu *et al.* have shown that PCDH7 is involved in the chemosensitivity of colon cancer cells. Knockdown of the gene, in line with the decreased expression seen for the experiments carried out here, resulted in increased sensitivity of colon cancer cells to chemotherapy agents. This suggests that the reduction in expression of PCDH7 in the HCT116 cell line could result in the positive impacts of the cell growth inhibition seen after treatment with the flavone compounds.[273]

The CYP24A1 gene encodes for the 24-hydroxylase enzyme, which is responsible for the catabolism of 1,25-dihydroxyvitamin D_3.[274] The enzyme is associated with proliferation and metastasis of cancer cells. Overexpression of CYP24A1 in tumours has also been associated with poorer prognosis of patients for cancers other than CRC.[274] Studies have also shown that expression levels can be used as biomarkers for CRC tumorigenesis and that CYP24A1 overexpression can have prooncogenic effects.[274,275] Reduction in CYP24A1 levels could therefore indicate improvement in the progression of the treated cells or could be an insight into the mechanism of action of the compound in cells, specifically for the chemopreventive effects of the PMF previously discussed.

The role of the Wnt10b gene in CRC is not yet fully understood, however due to its structural similarity with Wnt10a some roles of the gene can be inferred. Higher expression levels of Wnt10a have been associated with advanced tumour stage, and knockdown of Wnt10a can inactivate Wnt signalling, prevent proliferation and supress the invasiveness of CRC cells.[276] However, the role of Wnt10b in CRC is still unknown.

6.3.2.4 Discussion of the pathway analysis of the DEGs using Metascape and STRING

The analysis of the pathways associated with the differentially expressed genes showed consistency with the proteomics experiments. The increased genes showed pathway associations for autophagy and the unfolded protein response (UPR). There are also associations between the increased genes and the Nrf2 pathway, this has been previously been shown to be an interaction of PMF in cells and was discussed in Chapter 5.[277] The Metascape analysis also shows post-translational phosphorylation as a related geneset, meaning that compound **4.2** could have some interactions with the kinases in this pathway. Deoxycytidine kinase (DCK) was identified as a key interacting protein using the proteomics experiment analysis, indicating that the flavone compounds may be involved in phosphorylation processes.[278]

The pathways associated with the decreased DEGs are heavily related to cell cycle processes. This association can give an indication into the mechanism of action of the compound in cells. Further

studies of this association, such as a cell cycle analysis of the CRC cells after treatment with the compounds would be able to give more insight into the ways in which the compound is interfering with the cell cycle specifically.

The analysis of the complete set of DEGs showed some pathways that are also related to the results seen for the proteomics. The identification of signalling by Rho GTPases as a relevant pathway is consistent with the identification of a Rho GTPase as one of the top binding partners for the proteomics experiment.

The STRING analysis of the DEG networks gave some interesting insights. The same selection of cellular processes was identified as for the Metascape analysis, including the UPR and the cell cycle. Endoplasmic reticulum (ER) stress and the UPR are strongly linked as the UPR is evoked to restore ER proteostasis.[279] The network for the decreased DEGs showed a very high number of interactions between the genes, further analysis into the connection between these genes and their exact roles in cells could be beneficial to identify the behaviour of these compounds and their impact. There has previously been strong links identified between various RAB proteins and the ER, particularly in response to ER stress. Mateus *et al.* demonstrated that RAB7a controls ER homeostasis and responds to ER stress resulting in controlling the ER morphology. RAB7a was identified as a key interacting protein in the proteomics experiments.[280] The links between ER stress, the UPR and the RAB proteins could indicate a potential mechanism of action of the flavone compounds.

6.4 Conclusion

This chapter aimed to initially study the rate of cell growth inhibition observed for CRC cells when treated with PMF and compound **4.2**. The studies showed that the cells required exposure to the compounds for multiple doubling times to achieve the cell growth inhibition outcome. Treatment with compound **4.2** resulted in the lowest GI_{50} values after 48 h of exposure, indicating that the compound was causing more potent effects than treatment with PMF, which required 72 h of exposure. These results indicated a cell growth inhibition mechanism instead of a cell killing mechanism.

The chapter also presented the results for an experimental analysis of differentially expressed RNA after treatment of HCT116 cells with varying concentrations of compound **4.2**. DEGs were identified for the different treatments of the cells, and divided into genes that were increased compared to the DMSO control and those that were increased. The differentially expressed genes identified were analysed using pathway analysis tools Metascape and STRING to identify their cellular process involvement. The results showed that the DEGs were strongly involved in the UPR, ER stress response and the cell cycle. These cellular processes align with the interacting proteins and pathway analysis seen in Chapter 5 from the proteomics experiments.

Further analysis of the DEGs identified using the RNA sequencing results could give more insight into the specific changes occurring within the cell after treatment with the compound. Some changes in gene expression that may be important indicators of the mechanism of action of the compounds may have been missed with this analysis as they may not have reached the fold change cut-offs that were applied to the data but may still be significant.

7 Discussion

7.1 Introduction

Natural products have historically been a source for treatments for cancer. Almost 65% of the anticancer drugs developed since 1981 are either natural products or have been derived from the compounds.[281] Phytochemicals such as flavonoids are often reported to possess a wide range of bioactive properties, however, understanding of their direct behaviours and mechanisms of action to exert these properties is limited. The anticancer capabilities of flavones has been widely reported and studied, identifying for example that increased methylation of the structure can result in improved oral bioavailability.[37] PMF has been shown to possess promising anticancer and chemopreventive properties in CRC, but to date, a specific mechanism of action of the compound *in vitro* has not been identified.[282] The work reported in this thesis aimed to investigate the cellular targets and interactions of PMF in CRC to aid in the identification of potential targets or a mechanism of action of the compound. A pull down study was used to identify interacting protein targets of the compound. To carry out this study, a photoaffinity probe was synthesised. Synthesis of the photoaffinity probe required initial modification of the active compound. To ensure that the active compound was not modified in a manner that compromised biological activity, first, a series of modified flavone analogues were synthesised. The *in vitro* biological activity of the compound was then compared to the activity of the parent compound PMF to identify a suitable modification point. The complete probe molecule was then synthesised and the *in vitro* activity of the probe compound validated. Pull down studies were then carried out by exposing lysate of CRC cells to the probe molecule and initiating the formation of covalent linkages between the probe and the proximal proteins in the lysate. These proteins were visualised using SDS-PAGE and Western blot analysis, then the interacting proteins were identified using LCMS analysis. Changes in RNA expression caused by treatment of CRC cells with different concentrations of active flavone compound were also studied and analysed using RNA sequencing techniques to further investigate potential cellular interactions of the compound *in vitro* and potentially lead to the identification of a mechanism of action of the compound.

7.2 Synthesis and evaluation of biological activity of the parent compound and modified flavone analogues

The parent compound, PMF, was successfully synthesised and isolated, alongside four flavone analogues modified in different positions on the structure. The route followed in this work to synthesise the flavones utilised an iodine catalyst to form the cyclised flavone structure, this followed an adapted route used by Vongdeth *et. al.*[77] Difficulties were encountered during this step of the synthesis as purification of the desired compound became difficult due to contamination from an iodine-substituted cyclisation side product. The purification of this side product from the target compound is thought to be novel as the structure has not previously been published. Although the desired quantities of the compound were able to be isolated for the study, development of an alternative synthetic route that did not result in this side product could be beneficial for future uses of these compounds. Alternatively, the iodine-substituted side product could also have been utilised as an additional modification point on the structure for the attachment of the probe, had the modification positions tested all resulted in loss of biological activity.

PMF and the modified analogues were tested for their biological activity in two CRC cell lines, HT29 and HCT116, showing higher levels of activity in the HCT116 cell line overall. This reflected the sensitivity of these two cell lines seen by Fernández *et al*.[184] Modification of the flavone structure at the 4', 5 or 7-positions resulted in loss of biological activity in this cell line, evaluated using SRB assays. Loss of activity after modification of the flavone structure at the 4' position had previously been indicated by Vongdeth *et al.* in cancer cell lines of different tumour origins (HeLa, HCC1954 and SK-OV-3).[77] Modification of the structure in the 3'-position resulted in retention of biological activity and some increase in potency compared to PMF. Therefore, this point of modification was selected as the attachment point of the active compound to the pull down probe. Assessment of the *in vitro* biological activity of PMF and the analogues was complicated due to unfavourable low levels of aqueous solubility of the compounds. Precipitation of the compound as needle-like crystals was observed during the assays used to evaluate the biological activity of the compounds. Low levels of aqueous

solubility and poor oral bioavailability have been reported for phytochemicals as a whole, and flavones specifically, however reports have shown that increased methylation improves bioavailability properties.[37] Despite the successful identification of a suitable modification point carried out in this work, for the development of polymethoxyflavones into clinical drugs, significant improvements in the aqueous solubility of the compounds would be required. Alternatively, a method of encapsulation and delivery of the compound could also be investigated. Encapsulation of PMF in *γ*-cyclodextrin rings was tested by Dr Emerald Taylor within the Serpell group, however this did not improve the *in vitro* characteristics of PMF. Future work in this area would be required to improve solubility of PMF to enable its development into a potential drug.

7.3 Synthesis and evaluation of biological activity of the photoaffinity probe for the pull down study

A photoaffinity probe was designed using a TPD moiety and an alkyne click reporter handle, attached to the active flavone structure in the 3'-position. The diazirine was synthesised using previously developed methods for the most part.[193] One step, that typically required the use of liquid ammonia, was changed to allow the use of methanolic ammonia instead, a technique developed within the Serpell group by Dr Geraud Sansom. This simplified the synthesis and reduced the cost and safety risks associated with the step significantly and the final desired diazirine compound was successfully synthesised. However, the synthetic yield of the photoaffinity moiety obtained was too low for the amount of the probe required for the pull down study. Due to time constraints on the project, intensified by lab access restrictions due to the COVID-19 pandemic, optimisation of the reaction conditions to improve yield or increase the scale of the reactions for photoaffinity moiety was not feasible. Commercial sources of the photoaffinity species were sought to allow movement of the project into the next stages. If more time was available, the conditions for these reactions could have been optimised to allow complete synthesis of the probe compound.

Attachment of the linker group to the flavone for the synthesis of the probe, in the 3'-position, resulted in an increase in *in vitro* potency of the compound when its biological activity was assessed, thought to be due to increasing aqueous solubility characteristics seen for the structure. This could give insight into potential methods of structure modification that could improve the characteristics of PMF to make it more suitable as a clinical drug, since it has now been shown that modification at the 3'-position does not result in the loss of biological activity but can improve physical properties. The biological activity of the active and inactive probes was also assessed using antiproliferative assays with the active probe retaining biological activity as desired. As photoaffinity probes can be used to carry out live cell labelling, this could be possible using the probe synthesised in this work. Confirmation of activity of the active probe indicated that the probe was able to permeate the cell membrane of HCT116 cells. Cell based assay assessment of the proteins interacting with the probe could give more interesting insight into the mechanism of action of the compound.[283]

7.4 Identification and analysis of interacting proteins and changes in gene expression caused by treatment with the biologically active flavone

Using the photoaffinity labelling probe, proteins that interacted with the active compound could be isolated for identification. Initially, proteins that were associating with the probe were crosslinked to the probe and a fluorophore reporter tag was conjugated to the probe-protein complex using click chemistry. Results from the SDS-PAGE analysis of the fluorescently tagged probe were inconclusive, as proteins were not able to be visualised using the gel imaging equipment. Optimisation of this experiment could be very beneficial to identifying or validating key binding proteins of the compound. The conditions of the pull down experiment were optimised when using the biotin reporter tag, including extending the length of incubation used for the click reaction and changing the volume of the reporter tag used. These condition optimisations could improve the quality of data obtained with the FITC reporter tag if tested again. A different fluorophore could also be advantageous and could also give better results.

The results obtained from the SDS-PAGE experiments using the biotin reporter tag showed that there was a reduction in binding of proteins in the 35-40 kDa molecular weight region, due to the competition compound saturating the binding sites. However, useful results from these experiments were difficult to identify. This is mainly due to the large number of proteins being visualised in the gels, as the samples were complete cell lysates. The large number of proteins made gels crowded and therefore depletion of the bands when carrying out the competition study was difficult to see. The SDS-PAGE analysis of the lysates could further be used for validation of the protein targets identified using the proteomics, using a secondary antibody probe for the interacting proteins.

The interacting proteins were also isolated using streptavidin beads that bound to the biotin reporter tag attached to the probe in solution, using click chemistry. Mass spectrometry proteomics analysis was used to identify the isolated proteins and compare the quantities of the proteins in samples exposed to different treatment conditions. Control samples that were not exposed to the probe compound were compared to those that were treated with the probe, these were further compared to samples that had been treated with a competition compound prior to exposure to the probe.

Using multiple methods of analysis, members of the RAB GTPase family of proteins were regularly seen as top hits. As a result of this and the observation that associated pathways such as RAB geranylgeranylation were identified in the Metascape pathway analysis, this family of proteins, and in particular the members identified by this study, can be concluded as important interacting proteins of the flavone compound.

Assessment of the RAB proteins that were identified as key binders, Table 7.1, was carried out to determine if there is a consistent cellular localisation or function that this group is associated with that could give more insight into the mechanisms behind the association of the probe with these proteins, Table 7.1.

Table 7.1: Subcellular localisation and function of top hit RAB proteins

RAB protein	Subcellular localisation	Function
6a	Golgi apparatus[284]	Regulates membrane traffic from the Golgi apparatus to the endoplasmic reticulum (ER)[285]
11b	Vesicular compartments of recycling endosome and the trans-Golgi network (TGN)[286]	Controls mitotic spindle function and cell division[286]
2a	ER-Golgi intermediate compartment (ERGIC)[287]	Regulates COPI-dependent vesicular transport from the ERGIC[287]
7a	Late endosomes[288]	Regulates transport to endocytic degrative compatments[288,289]
1b	ER-golgi compartments[290]	Regulates anterograde traffic from the ER towards the Golgi[290]
19	Golgi apparatus[291]	Associated with cell cycle regulation[238]
10	Endoplasmic reticulum[292]	Regulates ER structure and dynamics[292]
21	Early endosomes[293]	Regulates integrin internalisation, endosomal sorting of cargo, autophagy and sorting at the Golgi[293]
5c	Early endosomes[294]	Early endosome fusion and endocytosis[294]

As shown in Table 7.1, there is no clear link between the localisation or function of the RAB proteins identified. This suggests that probe associated with RAB proteins with a wide variety of functions across the cell, suggesting that the association is more likely to be due to the structure of the protein than the specific setting of the proteins that is causing association. RAB proteins are now known to have 30-40% sequence identity and possess a highly conserved motif. Thorough analysis to find a consistent motif in the top hit proteins identified as hits could indicate why the proteins were associating with the probe molecule and should be investigated in future work.

To study the link between the proteins interacting with PMF and its potential mechanism of action, it is important to assess the role that the top hits play in cancer and in CRC specifically, although this is less understood. Jiang *et al*. showed that, of the 62 RAB proteins analysed, RAB10 and RAB19 were upregulated in CRC and these upregulated RAB proteins were shown to be related to cell cycle pathways. In this work, cell cycle pathways were shown to be significantly impacted by the exposure of the HCT116 cell line when RNA expression analysis was carried out, showing links between the

differentially expressed genes analysed in Chapter 6 and the identification of RAB proteins as top hits.[238]

Other hit pathways identified during the RNA expression analysis were associated with the UPR and autophagy. The UPR is activated when large amounts of misfolded or unfolded proteins accumulate in the ER. Autophagy processes regulate cellular metabolism by digesting and recycling proteins and cell organelles. Both of these mechanisms have been strongly associated with each other and with colorectal cancer.[295] ER stress caused by inflammation has been linked to both processes, suggesting that the anti-inflammatory properties of flavones and in particular PMF could be associated with their effect on these processes. Flavones have also been shown to influence expression of antioxidant genes by modulating redox sensitive transcription factors eg. Nuclear factors κB (NFκB), which is involved in the UPR and inflammatory cascade, and NfE2 related factor-2 (Nrf2).[234,24] The result of this is increased expression of genes encoding detoxifying and antioxidant proteins.[24]

Autophagy and the UPR involvement in cancer have also been liked with obesity, suggesting that biological effects that PMF has been shown to possess for both anti-obesity and anti-cancer uses could be due to a common mechanism of action associated with these cellular processes.[295.234]

Future work in this area should be focused on identifying the results of the influence of PMF on the processes of the UPR and autophagy; the specific effects of PMF on the cell cycle of CRC cells; and the influence of PMF on the behaviour of RAB proteins to elucidate how these interactions result in the favourable biological effects.

In summary, this thesis work has made the novel discovery that the RAB family of small GTPases interact with a tagged analogue of PMF in lysates prepared from CRC cells. Pathway analysis of the proteomics data also indicates links to the UPR, and this is supported by gene expression studies with PMF in CRC cells. These discoveries provide candidate cellular targets of PMF and a relationship to the UPR, which could be an important step in the pipeline of development of these natural compounds into potential drug treatments for CRC.

Bibliography

1 T. Iwashina, *Nat. Prod. Commun.*, 2015, **10**, 529–544.

2 E. Middleton, C. Kandaswami and T. C. Theoharides, *Pharmacol. Rev.*, 2000, **52**, 673–751.

3 P. G. Pietta, *J. Nat. Prod.*, 2000, **63**, 1035–1042.

4 A. Hazafa, K.-U.- Rehman, N. Jahan and Z. Jabeen, *Nutr. Cancer*, 2019, **0**, 1–12.

5 C. A. Rice-Evans, N. J. Miller and G. Paganga, *Free Radic. Biol. Med.*, 1996, **20**, 933–956.

6 M. Catarino, J. Alves-Silva, O. Pereira and S. Cardoso, *Curr. Top. Med. Chem.*, 2014, **15**, 105–119.

7 O. Prakash, A. Kumar, P. Kumar and A. Ajeet, *Am. J. Pharmacol. Sci.*, 2013, **1**, 104–115.

8 H. Chang, L. Lei, Y. Zhou, F. Ye and G. Zhao, *Nutrients*, 2018, **10**, 950.

9 M. Abotaleb, S. Samuel, E. Varghese, S. Varghese, P. Kubatka, A. Liskova and D. Büsselberg, *Cancers (Basel).*, 2018, **11**, 28.

10 J. R. Levy, K. A. Faber, L. Ayyash and C. L. Hughes, *Proc. Soc. Exp. Biol. Med.*, 1995, **208**, 60–66.

11 F. Perez-Vizcaino and C. G. Fraga, *Arch. Biochem. Biophys.*, 2018, **646**, 107–112.

12 B. A. Graf, P. E. Milbury and J. B. Blumberg, *J. Med. Food*, 2005, 8, 281–290.

13 G. L. Hostetler, R. A. Ralston and S. J. Schwartz, *Adv. Nutr. An Int. Rev. J.*, 2017, **8**, 423–435.

14 S. Kumar and A. K. Pandey, *Sci. World J.*, 2013, **2013**, 1–16.

15 D. F. Birt, S. Hendrich and W. Wang, *Pharmacol. Ther.*, 2001, **90**, 157–177.

16 J. Baell and M. A. Walters, *Nature*, 2014, **513**, 481–483.

17 J. B. Baell, *J. Nat. Prod.*, 2016, **79**, 616–628.

18 E. Birben, U. M. Sahiner, C. Sackesen, S. Erzurum and O. Kalayci, *World Allergy Organ. J.*, 2012, 5, 9–19.

19 P. Knekt, J. Kumpulainen, R. Järvinen, H. Rissanen, M. Heliövaara, A. Reunanen, T. Hakulinen and A. Aromaa, *Am. J. Clin. Nutr.*, 2002, **76**, 560–568.

20 M. Singh, M. Kaur and O. Silakari, *Eur. J. Med. Chem.*, 2014, **84**, 206–239.

21 J. Hyun, Y. Woo, D. S. Hwang, G. Jo, S. Eom, Y. Lee, J. C. Park and Y. Lim, *Bioorganic Med. Chem. Lett.*, 2010, **20**, 5510–5513.

22 M. Majewska, M. Skrzycki, M. Podsiad and H. Czeczot, *Acta Pol. Pharm. - Drug Res.*, 2011, **68**, 611–615.

23 R. Dhiman, S. Sharma, G. Singh, K. Nepali and P. M. Singh Bedi, *Arch. Pharm. (Weinheim).*, 2013, **346**, 7–16.

24 F. Dajas, A.-C. Juan Andres, A. Florencia, E. Carolina and R.-M. Felicia, *Cent. Nerv. Syst. Agents Med. Chem.*, 2013, **13**, 30–35.

25 D. J. Newman and G. M. Cragg, *J. Nat. Prod.*, 2016, **79**, 629–661.

26 A. L. Demain and P. Vaishnav, *Microb. Biotechnol.*, 2011, **4**, 687–699.

27 M. A. Jordan and L. Wilson, *Nat. Rev. Cancer*, 2004, **4**, 253–265.

28 D. Zhang, R. Yang, S. Wang and Z. Dong, *Drug Des. Devel. Ther.*, 2014, **8**, 279–284.

29 J. F. Zeidner and J. E. Karp, *Leuk. Res.*, 2015, **39**, 1312–1318.

30 J. F. Zeidner, M. C. Foster, A. L. Blackford, M. R. Litzow, L. E. Morris, S. A. Strickland, J. E. Lancet, P. Bose, M. Y. Levy, R. Tibes, I. Gojo, C. D. Gocke, G. L. Rosner, R. F. Little, J. J. Wright, L. A. Doyle, B. D. Smith and J. E. Karp, *Leuk. Res.*, 2018, **72**, 92–95.

31 R. Shi, Q. Huang, X. Zhu, Y. B. Ong, B. Zhao, J. Lu, C. N. Ong and H. M. Shen, *Mol. Cancer Ther.*, 2007, **6**, 1338–1347.

32 Q. Ye, K. Liu, Q. Shen, Q. Li, J. Hao, F. Han and R. W. Jiang, *Front. Oncol.*, 2019, **9**, 1–16.

33 P. Batra and A. K. Sharma, *3 Biotech*, 2013, **3**, 439–459.

34 E. R. Kasala, L. N. Bodduluru, R. M. Madana, K. V. Athira, R. Gogoi and C. C. Barua, *Toxicol. Lett.*, 2015, **233**, 214–225.

35 R. B. Semwal, D. K. Semwal, S. Combrinck, J. Trill, S. Gibbons and A. Viljoen, *Phytochem. Lett.*, 2019, **32**, 56–65.

36 T. Walle, N. Ta, T. Kawamori, X. Wen, P. A. Tsuji and U. K. Walle, *Biochem. Pharmacol.*, 2007, **73**, 1288–1296.

37 T. Walle, *Semin. Cancer Biol.*, 2007, **17**, 354–362.

38 X. de Leval, J. Delarge, F. Somers, P. de Tullio, Y. Henrotin, B. Pirotte and J.-M. Dogne, *Curr. Med. Chem.*, 2012, **7**, 1041–1062.

39 J. M. Schwab, H. J. Schluesener, R. Meyermann and C. N. Serhan, *Prostaglandins Leukot. Essent. Fat. Acids*, 2003, **69**, 339–343.

40 T. T. Dao, Y. S. Chi, J. Kim, H. P. Kim, S. Kim and H. Park, *Bioorganic Med. Chem. Lett.*, 2004, **14**, 1165–1167.

41 M. K. Gadhwal, U. J. Joshi and P. P. Shetgiri, .

42 N. Hashemi Goradel, M. Najafi, E. Salehi, B. Farhood and K. Mortezaee, *J. Cell. Physiol.*, 2019, **234**, 5683–5699.

43 X. Pi-Sunyer, *Postgrad. Med.*, 2009, **121**, 21–33.

44 G. Galati and P. J. O'Brien, *Free Radic. Biol. Med.*, 2004, **37**, 287–303.

45 M. Kawser Hossain, A. Abdal Dayem, J. Han, Y. Yin, K. Kim, S. Kumar Saha, G.-M. Yang, H. Choi and S.-G. Cho, *Int. J. Mol. Sci.*, 2016, **17**, 569.

46 M. Marranzano, S. Ray, J. Godos and F. Galvano, *Int. J. Food Sci. Nutr.*, 2018, **69**, 1020–1029.

47 D. Lee and J. Y. Imm, *J. Agric. Food Chem.*, 2018, **66**, 9989–9994.

48 B. Ahmad, E. P. Friar, M. S. Vohra, M. D. Garrett, C. J. Serpell, I. L. Fong and E. H. Wong, *Phytochemistry*, 2020, 180, 112513.

49 M. M. Li, Y. T. Chen, J. C. Ruan, W. J. Wang, J. G. Chen and Q. F. Zhang, *Curr. Res. Food Sci.*, 2023, **6**, 100424.

50 A. P. Rauter, A. Martins, C. Borges, H. Mota-Filipe, R. Pinto, B. Sepodes and J. Justino, *Phyther. Res.*, 2010, **24**, 133–138.

51 L. Ding, D. Jin and X. Chen, *J. Nutr. Biochem.*, 2010, **21**, 941–947.

52 H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal and F. Bray, *CA. Cancer J. Clin.*, 2021, **71**, 209–249.

53 D. Hanahan and R. A. Weinberg, *Cell*, 2000, **100**, 57–70.

54 E. Dekker, P. J. Tanis, J. L. A. Vleugels, P. M. Kasi and M. B. Wallace, *Lancet*, 2019, **394**, 1467–1480.

55 K. Simon, *Clin. Interv. Aging*, 2016, 11, 967–976.

56 S. Umezawa, T. Higurashi, Y. Komiya, J. Arimoto, N. Horita, T. Kaneko, M. Iwasaki, H. Nakagama and A. Nakajima, *Cancer Sci.*, 2019, **110**, 3018–3026.

57 R. M. Barnett, E. Borras, N. Jewel Samadder and E. Vilar, *Hered. Color. Cancer*, 2018, 349–365.

58 K. W. Jasperson, T. M. Tuohy, D. W. Neklason and R. W. Burt, *Gastroenterology*, 2010, **138**, 2044–2058.

59 W. M. Grady, *Gastroenterology*, 2003, **124**, 1574–1594.

60 R. Fodde, *Eur. J. Cancer*, 2002, **38**, 867–871.

61 R. Najdi, R. Holcombe and M. Waterman, *J. Carcinog.*, 2011, **10**, 5.

62 D. Huang, W. Sun, Y. Zhou, P. Li, F. Chen, H. Chen, D. Xia, E. Xu, M. Lai, Y. Wu and H. Zhang, *Cancer Metastasis Rev.*, 2018, **37**, 173–187.

63 H. Lei and K. Tao, *J. Cell. Mol. Med.*, 2020, **24**, 11828–11836.

64 M. R. Islam, S. Akash, M. M. Rahman, F. T. Nowrin, T. Akter, S. Shohag, A. Rauf, A. S. M. Aljohani and J. Simal-Gandara, *Chem. Biol. Interact.*, 2022, **368**, 110170.

65 C. Bailly, *Pharmacol. Res.*, 2019, **148**, 104398.

66 B. Mangla and K. Fink, *Am. Fam. Physician*, 2003, **67**, 129–130.

67 M. Arnold, M. S. Sierra, M. Laversanne, I. Soerjomataram, A. Jemal and F. Bray, *Gut*, 2017, **66**, 683–691.

68 C. Allemani, H. K. Weir, H. Carreira and et al., *Lancet*, 2015, **385**, 977–1010.

69 S. K. Veettil, K. G. Lim, N. Chaiyakunapruk, S. M. Ching and M. R. Abu Hassan, *Asian J. Surg.*, 2017, **40**, 481–489.

70 Y. Y. Lee and W. A. M. W. Muda, *Nutr. Res. Pract.*, 2019, **13**, 159–168.

71 H. K. Lim, S. M. Ghazali, C. C. Kee, K. K. Lim, Y. Y. Chan, H. C. Teh, A. F. M. Yusoff, G. Kaur, Z. M. Zain, M. H. N. Mohamad and S. Salleh, *BMC Public Health*, 2013, **13**, 1.

72 A. Leslie, F. A. Carey, N. R. Pratt and R. J. C. Steele, *Br. J. Surg.*, 2002, **89**, 845–860.

73 R. A. Gupta and R. N. DuBois, *Nat. Rev. Cancer*, 2001, **1**, 11–21.

74 S. Rajamanickam and R. Agarwal, *Drug Dev. Res.*, 2008, **69**, 460–471.

75 T. Kinoshita and K. Firman, *Phytochemistry*, 1997, **45**, 179–181.

76 F. IMAI, K. ITOH, N. KISHIBUCHI, T. KINOSHITA and U. SANKAWA, *Chem. Pharm. Bull.*, 1989, **37**, 119–123.

77 K. Vongdeth, P. Han, W. Li and Q. A. Wang, *Chem. Nat. Compd.*, 2019, **55**, 11–17.

78 X. Tang, H. Wang, L. Fan, X. Wu, A. Xin, H. Ren and X. J. Wang, *Free Radic. Biol. Med.*, 2011, **50**, 1599–1609.

79 X. Hou, X. Bai, X. Gou, H. Zeng, C. Xia, W. Zhuang, X. Chen, Z. Zhao, M. Huang and J. Jin, *Mol. Cells*, 2015, **38**, 396–401.

80 A. M. Gao, Z. P. Ke, J. N. Wang, J. Y. Yang, S. Y. Chen and H. Chen, *Carcinogenesis*, 2013, **34**, 1806–1814.

81 X. H. V Dolsodwlq, R. I. Wkh, K. L. Q. D. Q. L, L. X. Dqj and L. X. Hq, *Asian pacific J. Cancer Prev.*, 2014, **5**, 5–10.

82 H. Cai, S. Sale, R. Schmid, R. G. Britton, K. Brown, W. P. Steward and A. J. Gescher, *Cancer Prev. Res.*, 2009, **2**, 743–750.

83 F. E. Koehn and G. T. Carter, *Nat. Rev. Drug Discov.*, 2005, **4**, 206–220.

84 T. W. Corson and C. M. Crews, *Cell*, 2007, **130**, 769–774.

85 N. Kanoh, *Nat. Prod. Rep.*, 2016, **33**, 709–718.

86 C. P. Hart, *Drug Discov. Today*, 2005, **10**, 513–519.

87 E. Dominguez, A. Galmozzi, J. W. Chang, K. L. Hsu, J. Pawlak, W. Li, C. Godio, J. Thomas, D. Partida, S. Niessen, P. E. O'Brien, A. P. Russell, M. J. Watt, D. K. Nomura, B. F. Cravatt and E. Saez, *Nat. Chem. Biol.*, 2014, **10**, 113–121.

88 C. Harrison, *Nat. Rev. Drug Discov.*, 2014, **13**, 102–103.

89 S. Pan, H. Zhang, C. Wang, S. C. L. Yao and S. Q. Yao, *Nat. Prod. Rep.*, 2016, **33**, 612–620.

90 M. Kawatani and H. Osada, *Medchemcomm*, 2014, **5**, 277–287.

91 S. J. Dilly, M. J. Bell, A. J. Clark, A. Marsh, R. M. Napier, M. J. Sergeant, A. J. Thompson and P. C. Taylor, *Chem. Commun.*, 2007, 2808–2810.

92 T. Usui, H. Watanabe, H. Nakayama, Y. Tada, N. Kanoh, M. Kondoh, T. Asao, K. Takio, H. Watanabe, K. Nishikawa, T. Kitahara and H. Osada, *Chem. Biol.*, 2004, **11**, 799–806.

93 A. L. MacKinnon and J. Taunton, *Curr. Protoc. Chem. Biol.*, 2009, **1**, 55–73.

94 K. M. Nelson, J. L. Dahlin, J. Bisson, J. Graham, G. F. Pauli and M. A. Walters, *J. Med. Chem.*, 2017, **60**, 1620–1637.

95 H. Franken, T. Mathieson, D. Childs, G. M. A. Sweetman, T. Werner, I. Tögel, C. Doce, S. Gade, M. Bantscheff, G. Drewes, F. B. M. Reinhard, W. Huber and M. M. Savitski, *Nat. Protoc.*, 2015, **10**, 1567–1593.

96 S. Y. Seo and T. W. Corson, *Small molecule target identification using photo-affinity chromatography*, Elsevier Inc., 1st edn., 2019, vol. 622.

97 E. Smith and I. Collins, *Future Med. Chem.*, 2015, **7**, 159–183.

98 Y. Sadakane and Y. Hatanaka, *Anal. Sci.*, 2006, **22**, 209–218.

99 G. W. Preston and A. J. Wilson, *Chem. Soc. Rev.*, 2013, **42**, 3289–3301.

100 H. A. Flaxman and C. M. Woo, *Biochemistry*, 2018, **57**, 186–193.

101 L. Zhang, Y. Zhang, J. Dong, J. Liu, L. Zhang and H. Sun, *Bioorganic Med. Chem. Lett.*, 2012, **22**, 1036–1039.

102 J. Michalak, H. Bin Zhai and M. S. Platz, *J. Phys. Chem.*, 1996, **100**, 14028–14036.

103 K. Barral, A. D. Moorhouse and J. E. Moses, *Org. Lett.*, 2007, **9**, 1809–1811.

104 G. W. J. Fleet, R. R. Porter and J. R. Knowles, *Nature*, 1969, **224**, 511–512.

105 Y. 1 Y. Xia and L. Peng, Chem. Rev., 2013, 113, 7880–7929.Xia and L. Peng, *Chem. Rev.*, 2013, **113**, 7880–7929.

106 P. P. Geurink, L. M. Prely, G. A. Van Der Marel, R. Bischoff and H. S. Overkleeft, *Top. Curr. Chem.*, 2012, 324, 85–113.

107 J. Cheng and T. J. Deming, *Pept. Mater.*, 2011, **310**, 1–26.

108 S. A. Fleming, *Tetrahedron*, 1995, **51**, 12479–12520.

109 N. Yarravarapu, L. Geffert, C. K. Surratt, M. Cascio and D. J. Lapinsky, *Bioorganic Med. Chem. Lett.*, 2018, **28**, 3431–3435.

110 D. P. Murale, S. C. Hong, M. M. Haque and J. S. Lee, *Proteome Sci.*, 2017, **15**, 1–34.

111 J. Sumranjit and S. J. Chung, *Molecules*, 2013, **18**, 10425–10451.

112 Y. Xia and L. Peng, *Chem. Rev.*, 2013, **113**, 7880–7929.

113 I. S. Farell, R. Toroney, J. L. Hazen, R. A. Mehl and J. W. Chin, *Nat. Methods*, 2005, **2**, 377–384.

114 D. J. Lapinsky, *Bioorganic Med. Chem.*, 2012, **20**, 6237–6247.

115 T. Tomohiro, M. Hashimoto and Y. Hatanaka, *Chem. Rec.*, 2005, **5**, 385–395.

116 N. Kanoh, K. Honda, S. Simizu, M. Muroi and H. Osada, *Angew. Chemie*, 2005, **117**, 3625–3628.

117 K. Sakurai, S. Ozawa, R. Yamada, T. Yasui and S. Mizuno, *ChemBioChem*, 2014, **15**, 1399–1403.

118 S. F. Musolino, Z. Pei, L. Bi, G. A. DiLabio and J. E. Wulff, *Chem. Sci.*, 2021, **12**, 12138–12148.

119 R. A. G. Smith and J. R. Knowles, *J. Am. Chem. Soc.*, 1973, **95**, 5072–5073.

120 A. B. Kumar, J. D. Tipton and R. Manetsch, *Chem. Commun.*, 2016, **52**, 2729–2732.

121 Y. Kotake, K. Sagane, T. Owa, Y. Mimori-Kiyosue, H. Shimizu, M. Uesugi, Y. Ishihama, M. Iwata and Y. Mizui, *Nat. Chem. Biol.*, 2007, **3**, 570–575.

122 H. Shi, C.-J. Zhang, G. Y. J. Chen and S. Q. Yao, *J. Am. Chem. Soc.*, 2012, **134**, 3001–3014.

123 H. Guo and Z. Li, *Medchemcomm*, 2017, **8**, 1585–1591.

124 A. Herner, J. Marjanovic, T. M. Lewandowski, V. Marin, M. Patterson, L. Miesbauer, D. Ready, J. Williams, A. Vasudevan and Q. Lin, *J. Am. Chem. Soc.*, 2016, **138**, 14609–14615.

125 L. Zhang, Z. Yan, Y. Wang, C. Song and G. Miao, *Molecules*, 2019, **24**, 2394.

126 J. Das, S. Roy, S. Halnor, A. K. Das and A. Basak, *Org. Biomol. Chem.*, 2017, **15**, 1122–1129.

127 M. M. Rowland, H. E. Bostic, D. Gong, A. E. Speers, N. Lucas, W. Cho, B. F. Cravatt and M. D. Best, *Biochemistry*, 2011, **50**, 11143–11161.

128 G. Dormán and G. D. Prestwich, *Biochemistry*, 1994, **33**, 5661–5673.

129 H. Park, J. Y. Koo, Y. V. V. Srikanth, S. Oh, J. Lee, J. Park and S. B. Park, *Chem. Commun.*, 2016, **52**, 5828–5831.

130 S. I. Sato, Y. Kwon, S. Kamisuki, N. Srivastava, Q. Mao, Y. Kawazoe and M. Uesugi, *J. Am. Chem. Soc.*, 2007, **129**, 873–880.

131 M. Hashimoto, S. Okamoto, K. Nabeta and Y. Hatanaka, *Bioorganic Med. Chem. Lett.*, 2004, **14**, 2447–2450.

132 N. Kanoh, H. Takayama, K. Honda, T. Moriya, T. Teruya, S. Simizu, H. Osada and Y. Iwabuchi, *Bioconjug. Chem.*, 2010, **21**, 182–186.

133 S. H. L. Verhelst, M. Fonović and M. Bogyo, *Angew. Chemie - Int. Ed.*, 2007, **46**, 1284–1286.

134 J. J. Park, Y. Sadakane, K. Masuda, T. Tomohiro, T. Nakano and Y. Hatanaka, *ChemBioChem*, 2005, **6**, 814–818.

135 T. Tamura, S. Tsukiji and I. Hamachi, *J. Am. Chem. Soc.*, 2012, **134**, 2216–2226.

136 S. Al-Amrani, Z. Al-Jabri, A. Al-Zaabi, J. Alshekaili and M. Al-Khabori, *World J. Biol. Chem.*, 2021, **12**, 57–69.

137 R. Aebersold and M. Mann, *Nature*, 2003, **422**, 198–207.

138 F. W. McLafferty, *Science (80-.).*, 1981, **214**, 280–287.

139 M. Schenone, V. Dančík, B. K. Wagner and P. A. Clemons, *Nat. Chem. Biol.*, 2013, **9**, 232–240.

140 H. D. VanGuilder, K. E. Vrana and W. M. Freeman, *Biotechniques*, 2008, **44**, 619–626.

141 Z. Khatoon, B. Figler, H. Zhang and F. Cheng, *Drug Dev. Res.*, 2014, **75**, 324–330.

142 G. W. Milne, M. C. Nicklaus and S. Wang, *SAR QSAR Environ. Res.*, 1998, **9**, 23–38.

143 R. Guha, *Methods Mol. Biol.*, 2013, **993**, 81–94.

144 C. J. Mills, N. N. Mateeva and K. K. Redda, *J. Heterocycl. Chem.*, 2006, **43**, 59–64.

145 N. N. Mateeva, R. N. Kode and K. K. Redda, *J. Heterocycl. Chem.*, 2002, **39**, 1251–1258.

146 S. Kawaii, T. Ikuina, T. Hikima, T. Tokiwano and Y. Yoshizawa, *Anticancer Res.*, 2012, **32**, 5239–5244.

147 D. S. Lim, D. H. Lim, J. H. Lee, E. T. Oh and Y. S. Keum, *J. Agric. Food Chem.*, 2017, **65**, 3056–3064.

148 M. Łużny, E. Kozłowska, E. Kostrzewa-Susłow and T. Janeczko, *Catalysts*, 2020, **10**, 1–13.

149 A. Pawlak, M. Henklewska, B. Hernández Suárez, M. Łużny, E. Kozłowska, B. Obmińska-Mrukowicz and T. Janeczko, *Molecules*, 2020, **25**, 4362.

150 M. Łużny, T. Tronina, E. Kozłowska, E. Kostrzewa-Susłow and T. Janeczko, *J. Agric. Food Chem.*, 2021, **69**, 3879–3886.

151 W. Lan, F. Lu, M. Regner, Y. Zhu, J. Rencoret, S. A. Ralph, U. I. Zakai, K. Morreel, W. Boerjan and J. Ralph, *Plant Physiol.*, 2015, **167**, 1284–1295.

152 I. B. Masesane, *Int. J. Chem. Stud.*, 2015, **3**, 53–59.

153 G. Bose, E. Mondal, A. T. Khan and M. J. Bordoloi, *Tetrahedron Lett.*, 2001, **42**, 8907–8909.

154 J. Clayden author, N. Greeves author and S. G. Warren author, *Organic chemistry*, Oxford University Press, Oxford, 2nd editio., 2012.

155 B. Ahmad, E. P. Friar, E. Taylor, M. S. Vohra, C. J. Serpell, M. D. Garrett, J. S. E. Loo, I. L. Fong and E. H. Wong, *Eur. J. Pharmacol.*, 2023, **938**, 175445.

156 B. Ahmad, E. P. Friar, E. Taylor, M. S. Vohra, C. J. Serpell, M. D. Garrett, J. S. E. Loo, I. L. Fong and E. H. Wong, *Eur. J. Pharmacol.*, 2023, **938**, 175445.

157 L. Xia and Y. R. Lee, *Bull. Korean Chem. Soc.*, 2012, **33**, 2087–2090.

158 D. Kavvadias, P. Sand, K. A. Youdim, M. Z. Qaiser, C. Rice-Evans, R. Baur, E. Sigel, W. D. Rausch, P. Riederer and P. Schreier, *Br. J. Pharmacol.*, 2004, **142**, 811–820.

159 Z. Sipos and K. Kónya, *Synth.*, 2018, **50**, 1610–1620.

160 W. Lan, F. Lu, M. Regner, Y. Zhu, J. Rencoret, S. A. Ralph, U. I. Zakai, K. Morreel, W. Boerjan and J. Ralph, *Plant Physiol.*, 2015, **167**, 1284–1295.

161 W. Liu, Y. Feng, S. Yu, Z. Fan, X. Li, J. Li and H. Yin, *Int. J. Mol. Sci.*, 2021, **22**, 1–18.

162 K. V. Sashidhara, M. Kumar and A. Kumar, *Tetrahedron Lett.*, 2012, **53**, 2355–2359.

163 Y. Wei, J. Tang, X. Cong and X. Zeng, *Green Chem.*, 2013, **15**, 3165–3169.

164 J. C. Aponte, M. Verástegui, E. Málaga, M. Zimic, M. Quiliano, A. J. Vaisberg, R. H. Gilman and G. B. Hammond, *J. Med. Chem.*, 2008, **51**, 6230–6234.

165 Z. T. Du, J. Lu, H. R. Yu, Y. Xu and A. P. Li, *J. Chem. Res.*, 2010, 222–227.

166 K. Okano, K. I. Okuyama, T. Fukuyama and H. Tokuyama, *Synlett*, 2008, 1977–1980.

167 B. Ahmad, E. P. Friar, M. S. Vohra, N. Khan, C. J. Serpell, M. D. Garrett, J. S. E. Loo, I. L. Fong and E. H. Wong, *Chem. Biol. Interact.*, 2023, **379**, 110503.

168 J. A. C. Sousa, A. C. C. Reis, C. L. de B. Magalhães, B. de M. Silva and G. C. Brandão, *Nat. Prod. Res.*, 2022, **0**, 1–6.

169 S. Sale, I. L. Fong, C. de Giovanni, L. Landuzzi, K. Brown, W. P. Steward and A. J. Gescher, *Eur. J. Cancer*, 2009, **45**, 2731–2735.

170 H. Cai, S. Sale, R. Schmid, R. G. Britton, K. Brown, W. P. Steward and A. J. Gescher, *Cancer Prev. Res.*, 2009, **2**, 743–750.

171 M. Al-Fayez, H. Cai, R. Tunstall, W. P. Steward and A. J. Gescher, *Cancer Chemother. Pharmacol.*, 2006, **58**, 816–825.

172 I. L. Fong, University of Leicester, 2011.

173 A. L. Cornelia Kasper, Verena Charwat, *Cell Culture Technology*, Springer International Publishing, 2018.

174 P. Skehan, R. Storeng, D. Scudiero, A. Monks, J. Mcmahon, D. Vistica, J. T. Warren, H. Bokesch, S. Kenney and M. R. Boyd, *J. Natl. Cancer Inst.*, 1990, **82**, 1107–1112.

175 W. Voigt, in *Chemosensitivity*, Humana Press, New Jersey, pp. 039–048.

176 K. T. Papazisis, G. D. Geromichalos, K. A. Dimitriadis and A. H. Kortsaris, *J. Immunol. Methods*, 1997, 208, 151–158.

177 B. H. Havsteen, *The biochemistry and medical significance of the flavonoids*, 2002, vol. 96.

178 T. Walle, *Int. J. Mol. Sci.*, 2009, **10**, 5002–5019.

179 G. Lewin, A. MacIuk, A. Moncomble and J. P. Cornard, *J. Nat. Prod.*, 2013, **76**, 8–12.

180 H. Kim, H. W. Kim and S. Jung, *Bull. Korean Chem. Soc.*, 2008, **29**, 590–594.

181 M. Christoff, L. T. Okano and C. Bohne, *J. Photochem. Photobiol. A Chem.*, 2000, **134**, 169–176.

182 J. Picot, *Human Cell Culture Protocols*, 2004.

183 H. Cai, S. Sale, R. Schmid, R. G. Britton, K. Brown, W. P. Steward and A. J. Gescher, *Cancer Prev. Res.*, 2009, **2**, 743–750.

184 J. Fernández, B. Silván, R. Entrialgo-Cadierno, C. J. Villar, R. Capasso, J. A. Uranga, F. Lombó and R. Abalo, *Biomed. Pharmacother.*, 2021, **143**, 112241.

185 J. M. Seo, H. M. Kang, K. H. Son, J. H. Kim, C. W. Lee, H. M. Kim, S. I. Chang and B. M. Kwon, *Planta Med.*, 2003, **69**, 218–222.

186 Q. Zuo, R. Wu, X. Xiao, C. Yang, Y. Yang, C. Wang, L. Lin and A. N. Kong, *J. Cell. Biochem.*, 2018, **119**, 9573–9582.

187 H. Zhang, J. Zheng, A. Liu, H. Xiao and L. He, *J. Agric. Food Chem.*, 2016, **64**, 9708–9713.

188 M. Paini, S. R. Daly, B. Aliakbarian, A. Fathi, E. A. Tehrany, P. Perego, F. Dehghani and P. Valtchev, *Colloids Surfaces B Biointerfaces*, 2015, **136**, 1067–1072.

189 H. Chen, C. Khemtong, X. Yang, X. Chang and J. Gao, *Drug Discov. Today*, 2011, **16**, 354–360.

190 L. Zhao, Y. Wei, Y. Huang, B. He, Y. Zhou and J. Fu, *Int. J. Nanomedicine*, 2013, **8**, 3769–3779.

191 T. Tomohiro, M. Hashimoto and Y. Hatanaka, *Chem. Rec.*, 2005, **5**, 385–395.

192 E. Smith and I. Collins, *Future Med. Chem.*, 2015, **7**, 159–183.

193 J. R. Hill and A. A. B. Robertson, *J. Med. Chem.*, 2018, **61**, 6945–6963.

194 L. Dubinsky, B. P. Krom and M. M. Meijler, *Bioorganic Med. Chem.*, 2012, **20**, 554–570.

195 I. Protasova, B. Bulat, N. Jung and S. Bräse, *Org. Lett.*, 2017, **19**, 34–37.

196 Z. P. Tachrim, L. Wang, Y. Murai and M. Hashimoto, *Molecules*, 2021, **26**, 4496.

197 J. R. Hill and A. A. B. Robertson, *J. Med. Chem.*, 2018, **61**, 6945–6963.

198 L. Wang, Y. Murai, T. Yoshida, A. Ishida, K. Masuda, Y. Sakihama, Y. Hashidoko, Y. Hatanaka and M. Hashimoto, *Org. Lett.*, 2015, **17**, 616–619.

199 A. B. Kumar and R. Manetsch, *Synth. Commun.*, 2018, **48**, 626–631.

200 T. Zhang and A. E. Ondrus, *Synlett*, 2021, **32**, 1053–1059.

201 V. A. Pistritto, J. M. Paolillo, K. A. Bisset and N. E. Leadbeater, *Org. Biomol. Chem.*, 2018, **16**, 4715–4719.

202 T. Mayer and M. E. Maier, *European J. Org. Chem.*, 2007, 4711–4720.

203 H. Deng, Q. Lei, N. Yang, S. Dai, H. Peng, K. Yang, Z. Xiao, D. Wang, Z. Yu, N. Li and W. Li, *Anal. Chem.*, 2022, **94**, 10118–10126.

204 M. N. Khan and A. A. Khan, *J. Chem. Soc. Perkin Trans. 2*, 1979, 796–798.

205 P. P. Chandrachud, L. Wojtas and J. M. Lopchuk, *J. Am. Chem. Soc.*, 2020, **142**, 21743–21750.

206 E. J. Lawrence, G. G. Wildgoose, L. Aldous, Y. A. Wu, J. H. Warner, R. G. Compton and P. D. McNaughter, *Chem. Mater.*, 2011, **23**, 3740–3751.

207 A. A. Hussein, A. A. M. Al-Hadedi, A. J. Mahrath, G. A. I. Moustafa, F. A. Almalki, A. Alqahtani, S. Shityakov and M. E. Algazally, *R. Soc. Open Sci.*, 2020, **7**, 191568.

208 G. T. Hermanson, in *Bioconjugate Techniques*, Elsevier, 2013, pp. 229–258.

209 S. Sale, R. G. Tunstall, K. C. Ruparelia, G. A. Potter, W. P. Steward and A. J. Gescher, *Int. J. Cancer*, 2005, **115**, 194–201.

210 I. I. Ahmed Juvale, A. A. Abdul Hamid, K. B. Abd Halim and A. T. Che Has, *Heliyon*, 2022, **8**, e09777.

211 C. H. Choi, J. H. Kim and S. H. Kim, *Biochem. Biophys. Res. Commun.*, 2004, **320**, 672–679.

212 N. Barbero, C. Barolo and G. Viscardi, *World J. Chem. Educ. Vol. 4, 2016, Pages 80-85*, 2016, **4**, 80–85.

213 L. K. Chaganti, N. Venkatakrishnan and K. Bose, *Biosci. Rep.*, 2018, **38**, 1–8.

214 T. Ballard, H. E. Murrey, K. F. Geoghegan, C. W. Am Ende and D. S. Johnson, *Medchemcomm*, 2014, **5**, 321–327.

215 M. M. Bradford, *Anal. Biochem.*, 1976, **72**, 248–254.

216 J. E. Noble and M. J. A. Bailey, *Chapter 8 Quantitation of Protein*, Elsevier Inc., 1st edn., 2009, vol. 463.

217 N. J. Kruger, *Methods Mol. Biol.*, 1994, **32**, 9–15.

218 K. Weber and M. Osborn, *J. Biol. Chem.*, 1969, **244**, 4406–4412.

219 W. N. Burnette, *Anal. Biochem.*, 1981, **112**, 195–203.

220 A. Sormunen, E. Koivulehto, K. Alitalo, K. Saksela, N. Laham-Karam and S. Ylä-Herttuala, *Methods Protoc.*, 2023, **6**, 43.

221 M. Mishra, S. Tiwari, A. Gunaseelan, D. Li, B. D. Hammock and A. V. Gomes, *Electrophoresis*, 2019, **40**, 1731–1739.

222 M. K. Sivanich, T. Gu, D. N. Tabang and L. Li, *Proteomics*, 2022, **22**, 2100256.

223 J. Li, J. G. Van Vranken, L. Pontano Vaites, D. K. Schweppe, E. L. Huttlin, C. Etienne, P. Nandhikonda, R. Viner, A. M. Robitaille, A. H. Thompson, K. Kuhn, I. Pike, R. D. Bomgarden, J. C. Rogers, S. P. Gygi and J. A. Paulo, *Nat. Methods*, 2020, **17**, 399–404.

224 F. J. Sialana, T. I. Roumeliotis, H. Bouguenina, L. Chan Wah Hak, H. Wang, J. Caldwell, I. Collins, R. Chopra and J. S. Choudhary, *J. Proteome Res.*, 2022, **21**, 1842–1856.

225 M. Kleverov, D. Zenkova, V. Kamenev, M. Sablina, M. N. Artyomov and A. A. Sergushichev, *bioRxiv*, , DOI:10.1101/2022.12.10.519861.

226 T. Hulsen, J. de Vlieg and W. Alkema, *BMC Genomics*, 2008, **9**, 1–6.

227 Y. Zhou, B. Zhou, L. Pache, M. Chang, A. H. Khodabakhshi, O. Tanaseichuk, C. Benner and S. K. Chanda, *Nat. Commun.*, 2019, **10**, 1523.

228 J. M. Niers, J. W. Chen, R. Weissleder and B. A. Tannous, *Anal. Chem.*, 2011, **83**, 994–999.

229 L. Ting, M. J. Cowley, S. L. Hoon, M. Guilhaus, M. J. Raftery and R. Cavicchioli, *Mol. Cell. Proteomics*, 2009, **8**, 2227–2242.

230 P. Brooks, P. Emery, J. F. Evans, H. Fenner, C. J. Hawkey, C. Patrono, J. Smolen, F. Breedveld, R. Day, M. Dougados, E. W. Ehrich, J. Gijon-Baños, T. K. Kvien, M. H. Van Rijswijk, T. Warner and H. Zeidler, *Rheumatology*, 1999, **38**, 779–788.

231 A. Lau, W. Tian, S. A. Whitman and D. D. Zhang, *Antioxidants Redox Signal.*, 2013, **18**, 91–93.

232 A. Seelig, *Front. Oncol.*, 2020, **10**, 1–16.

233 A. Read and M. Schröder, *Biology (Basel).*, 2021, **10**, 1–10.

234 E. A. V. Beilankouhi, M. A. Sajadi, I. Alipourfard, P. Hassani, M. Valilo and R. Safaralizadeh, *Pathol. Res. Pract.*, 2023, **248**, 154706.

235 P. S. Wu, J. H. Yen, M. C. Kou and M. J. Wu, *PLoS One*, 2015, **10**, 1–23.

236 S. L. Haney, M. L. Varney, Y. S. Chhonker, S. Shin, K. Mehla, A. J. Crawford, H. J. Smith, L. M. Smith, D. J. Murry, M. A. Hollingsworth and S. A. Holstein, *Oncogene*, 2019, **38**, 5308–5320.

237 S. L. Haney, D. Feng, Y. S. Chhonker, M. L. Varney, J. T. Williams, L. M. Smith, J. B. Ford, D. J. Murry and S. A. Holstein, *Drug Dev. Res.*, 2023, **84**, 62–74.

238 X. Jiang, L. Yang, Q. Gao, Y. Liu, X. Feng, S. Ye and Z. Yang, *Front. Genet.*, 2022, **13**, 1–15.

239 H. Jin, Y. Tang, L. Yang, X. Peng, B. Li, Q. Fan, S. Wei, S. Yang, X. Li, B. Wu, M. Huang, S. Tang, J. Liu and H. Li, *Front. Cell Dev. Biol.*, 2021, **9**, 1–13.

240 S. Ötzkan, W. E. Muller, W. Gibson Wood and G. P. Eckert, *NeuroMolecular Med.*, 2021, **23**, 130–139.

241 V. Ullmannova and N. C. Popescu, *Cancer Detect. Prev.*, 2007, **31**, 110–118.

242 N. Parmar and F. Tamanoi, *Enzymes*, 2010, **27**, 39–56.

243 C. L. Van der Wilt, J. R. Kroep, A. M. Bergman, W. J. P. Loves, E. Alvarez, I. Talianidis, S. Eriksson, C. J. Van Groeningen, H. M. Pinedo and G. J. Peters, *Adv. Exp. Med. Biol.*, 2000, **486**, 287–290.

244 P. Rodriguez, D. Munroe, D. Prawitt, L. L. Chu, E. Bric, J. Kim, L. H. Reid, C. Davies, H. Nakagama, R. Loebbert, A. Winterpacht, M. J. Petruzzi, M. J. Higgins, N. Nowak, G. Evans, T. Shows, B. E. Weissman, B. Zabel, D. E. Housman and J. Pelletier, *Genomics*, 1997, **44**, 253–265.

245 C. J. S. Queiroz, F. Song, K. R. Reed, N. Al-Khafaji, A. R. Clarke, D. Vimalachandran, F. Miyajima, D. M. Pritchard and J. R. Jenkins, *Front. Oncol.*, 2020, **10**, 1–11.

246 N. Chen, X. Han, X. Bai, B. Yin and Y. Wang, *Am. J. Transl. Res.*, 2020, **12**, 6490–6500.

247 S. Yang, X. Ren, Y. Liang, Y. Yan, Y. Zhou, J. Hu, Z. Wang, F. Song, F. Wang, W. Liao, W. Liao, Y. Ding and L. Liang, *Oncogene*, 2020, **39**, 249–261.

248 A. Kaida, S. Yamamoto, A. Parrales, E. D. Young, A. Ranjan, M. A. Alalem, K. ichi Morita, Y. Oikawa, H. Harada, T. Ikeda, S. M. Thomas, F. j. Diaz and T. Iwakuma, *Oncogene*, 2021, **40**, 5013–5025.

249 N. Hosokawa, K. Hirayoshi, A. Nakai, Y. Hosokawa, N. Marui, M. Yoshida, T. Sakai, H. Nishino, A. Aoike, K. Kawai and K. Nagata, *Cell Struct. Funct.*, 1990, **15**, 393–401.

250 S. M. Kwon, S. Kim, N. J. Song, S. H. Chang, Y. J. Hwang, D. K. Yang, J. W. Hong, W. J. Park and K. W. Park, *J. Nutr. Biochem.*, 2016, **30**, 24–32.

X. Fan, L. Liu, Y. Shi, F. Guo, H. Wang, X. Zhao, D. Zhong and G. Li, *World J. Surg. Oncol.*, 2020, **18**, 1–14.

K. Ariake, H. Ohtsuka, F. Motoi, D. Douchi, M. Oikawa, T. Rikiyama, K. Fukase, Y. Katayose, S. Egawa and M. Unno, *Cancer Lett.*, 2012, **325**, 99–107.

J. Liang, W. Zhou, N. Sakre, J. DeVecchio, S. Ferrandon, A. H. Ting, S. Bao, I. Bissett, J. Church and M. F. Kalady, *Br. J. Cancer*, 2018, **119**, 1267–1277.

G. Li, *Curr. Drug Targets*, 2011, **12**, 1188–1193.

R. Stark, M. Grzelak and J. Hadfield, *Nat. Rev. Genet.*, 2019, **20**, 631–656.

M. Hong, S. Tao, L. Zhang, L. T. Diao, X. Huang, S. Huang, S. J. Xie, Z. D. Xiao and H. Zhang, *J. Hematol. Oncol.*, 2020, **13**, 1–16.

L. J. Cseke and L. Larka, *Handb. Mol. Cell. Methods Biol. Med. Third Ed.*, 2016, **2015**, 141–164.

P. Ewels, M. Magnusson, S. Lundin and M. Käller, *Bioinformatics*, 2016, **32**, 3047–3048.

H. Varet, L. Brillet-Guéguen, J. Y. Coppée and M. A. Dillies, *PLoS One*, 2016, **11**, 1–8.

H. Heberle, V. G. Meirelles, F. R. da Silva, G. P. Telles and R. Minghim, *BMC Bioinformatics*, 2015, **16**, 1–7.

D. Szklarczyk, R. Kirsch, M. Koutrouli, K. Nastou, F. Mehryary, R. Hachilif, A. L. Gable, T. Fang, N. T. Doncheva, S. Pyysalo, P. Bork, L. J. Jensen and C. Von Mering, *Nucleic Acids Res.*, 2023, **51**, D638–D646.

S. K. Schnyder, J. J. Molina and R. Yamamoto, *Sci. Rep.*, 2020, **10**, 1–14.

W. W. Wilfinger, K. Mackey and P. Chomczynski, *Biotechniques*, 1997, **22**, 474–481.

S. K. Schulze, R. Kanwar, M. Gölzenleuchter, T. M. Therneau and A. S. Beutler, *BMC Genomics*, 2012, **13**, 7–9.

C. Gongora, L. Candeil, N. Vezzio, V. Copois, V. Denis, C. Breil, F. Molina, C. Fraslon, E. Conseiller, B. Pau, P. Martineau and M. Del Rio, *Cancer Biol. Ther.*, 2008, **7**, 822–832.

H. Alborzinia, S. Can, P. Holenya, C. Scholl, E. Lederer, I. Kitanovic and S. Wölfl, *PLoS One*, 2011, **6**, e19714.

W. F. Marzluff, P. Gongidi, K. R. Woods, J. Jin and L. J. Maltais, *Genomics*, 2002, **80**, 487–498.

W. Albig, P. Kioschis, A. Poustka, K. Meergans and D. Doenecke, *Genomics*, 1997, **40**, 314–322.

Y. Shen, L. Meunier and L. M. Hendershot, *J. Biol. Chem.*, 2002, **277**, 15947–15956.

S. Yang, W. Li, H. Sun, B. Wu, F. Ji, T. Sun, H. Chang, P. Shen, Y. Wang and D. Zhou, *BMC Cancer*, 2015, **15**, 1–11.

Z. Yang, S. Liu, Y. Wang, Y. Chen, P. Zhang, Y. Liu, H. Zhang, P. Zhang, Z. Tao and K. Xiong, *Thorac. Cancer*, 2020, **11**, 1944–1954.

P. K. Chaudhary and S. Kim, *An insight into gpcr and g-proteins as cancer drivers*, 2021, vol. 10.

Z. Liu, Y. Xu, X. Liu and B. Wang, *Biochem. Cell Biol.*, 2022, **100**, 445–457.

H. Sun, C. Wang, M. Hao, R. Sun, Y. Wang, T. Liu, X. Cong and Y. Liu, *Hum. Pathol.*, 2016, **50**, 101–108.

H. C. Horváth, P. Lakatos, J. P. Kósa, K. Bácsi, K. Borka, G. Bises, T. Nittke, P. A. Hershberger, G. Speer and E. Kállay, *J. Histochem. Cytochem.*, 2010, **58**, 277–285.

X. Nie, H. Liu, L. Liu, Y. D. Wang and W. D. Chen, *Front. Oncol.*, 2020, **10**, 1–13.

X. Hou, X. Bai, X. Gou, H. Zeng, C. Xia, W. Zhuang, X. Chen, Z. Zhao, M. Huang and J. Jin, *Mol. Cells*, 2015, **38**, 396–401.

S. Hazra, A. Szewczak, S. Ort, M. Konrad and A. Lavie, *Biochemistry*, 2011, **50**, 2870–2880.

Y. Lin, M. Jiang, W. Chen, T. Zhao and Y. Wei, *Biomed. Pharmacother.*, 2019, **118**, 109249.

D. Mateus, E. S. Marini, C. Progida and O. Bakke, *BBA - Mol. Cell Res.*, 2018, **1865**, 781–793.

D. J. Newman and G. M. Cragg, *J. Nat. Prod.*, 2020, **83**, 770–803.

H. Cai, S. Sale, R. Schmid, R. G. Britton, K. Brown, W. P. Steward and A. J. Gescher, *Cancer Prev. Res.*, 2009, **2**, 743–750.

S. A. Head and J. O. Liu, *J. Vis. Exp.*, , DOI:10.3791/54529.

A. Iwaki, K. Moriwaki, T. Sobajima, M. Taniguchi, S. ichiro Yoshimura, M. Kunii, S. Kanda, Y. Kamada, E. Miyoshi and A. Harada, *FASEB J.*, 2020, **34**, 9450–9465.

T. Matanis, A. Akhmanova, P. Wulf, E. Del Nery, T. Weide, T. Stepanova, N. Galjart, F. Grosveld, B. Goud, C. I. De Zeeuw, A. Barnekow and C. C. Hoogenraad, *Nat. Cell Biol.*, 2002, **4**, 986–992.

I. Joseph, J. Flores, V. Farrell, J. Davis, J. Bianchi-Smak, Q. Feng, S. Goswami, X. Lin, Z. Wei, K. Tong, Z. Feng, M. P. Verzi, E. M.

Bonder, J. R. Goldenring and N. Gao, *EMBO Rep.*, , DOI:10.15252/embr.202256240.

287 T. Sugawara, F. Kano and M. Murata, *Sci. Rep.*, 2014, **4**, 1–14.

288 A. Margiotta, C. Progida, O. Bakke and C. Bucci, *Biochim. Biophys. Acta - Mol. Cell Res.*, 2017, **1864**, 367–381.

289 K. W. Cheng, J. P. Lahad, J. W. Gray and G. B. Mills, *Cancer Res.*, 2005, **65**, 2516–2519.

290 T. Weide, M. Bayer, M. Köster, J. Siebrasse, R. Peters and A. Barnekow, *EMBO Rep.*, 2001, **2**, 336–341.

291 B. Goud, S. Liu and B. Storrie, *Small GTPases*, 2018, **9**, 66–75.

292 A. R. English and G. K. Voeltz, *Nat. Cell Biol.*, 2013, **15**, 169–178.

293 T. Del Olmo, C. Lacarrière-Keïta, C. Normandin, D. Jean, F.-M. Boisvert and S. Jean, *Biol. Open*.

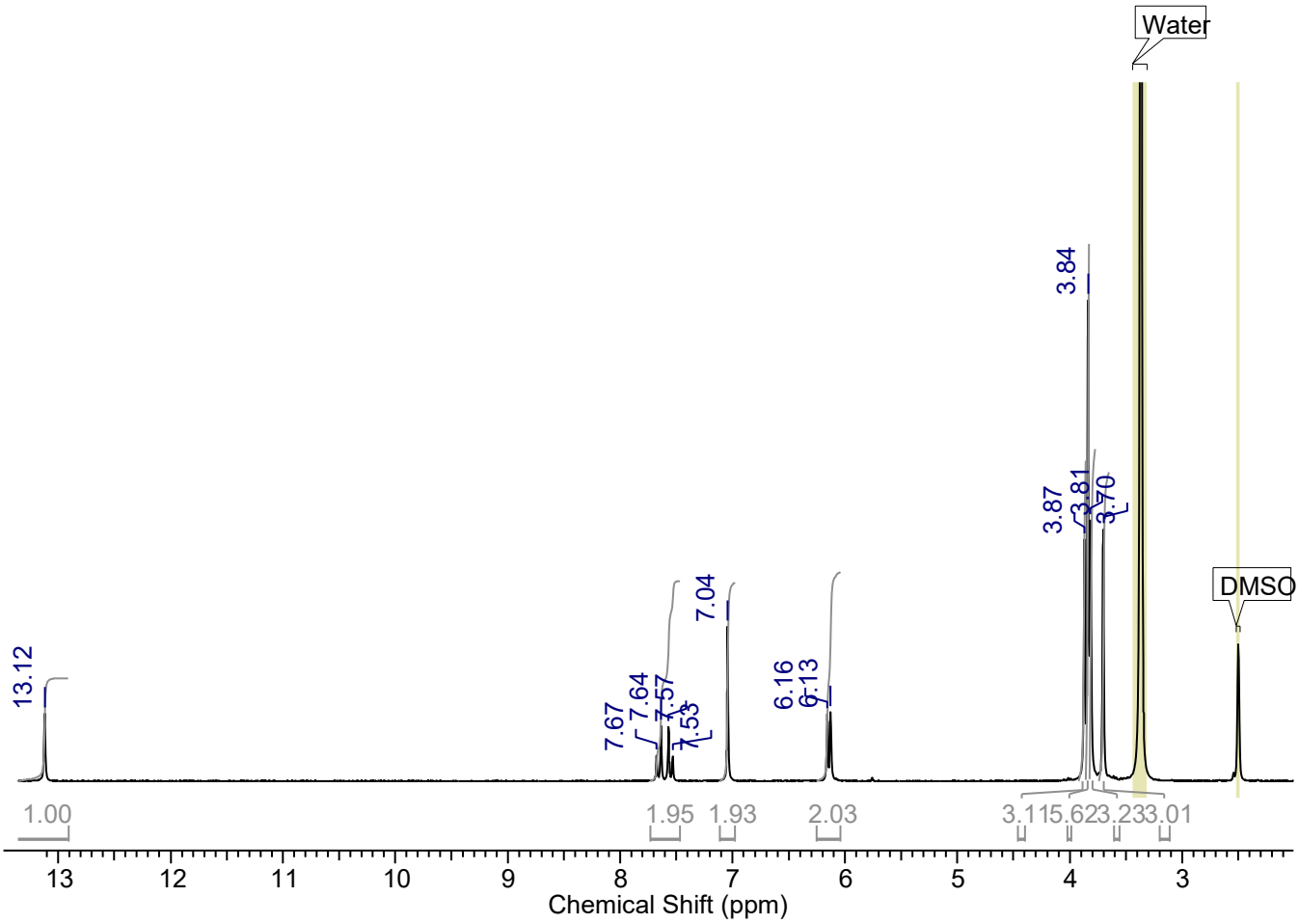
294 C. Bucci, A. Lütcke, O. Steele-Mortimer, V. M. Olkkonen, P. Dupree, M. Chiariello, C. B. Bruni, K. Simons and M. Zerial, *FEBS Lett.*, 1995, **366**, 65–71.

295 P. Mokarram, M. Albokashy, M. Zarghooni, M. A. Moosavi, Z. Sepehri, Q. M. Chen, A. Hudecki, A. Sargazi, J. Alizadeh, A. R. Moghadam, M. Hashemi, H. Movassagh, T. Klonisch, A. A. Owji, M. J. Łos and S. Ghavami, *Autophagy*, 2017, **13**, 781–819.

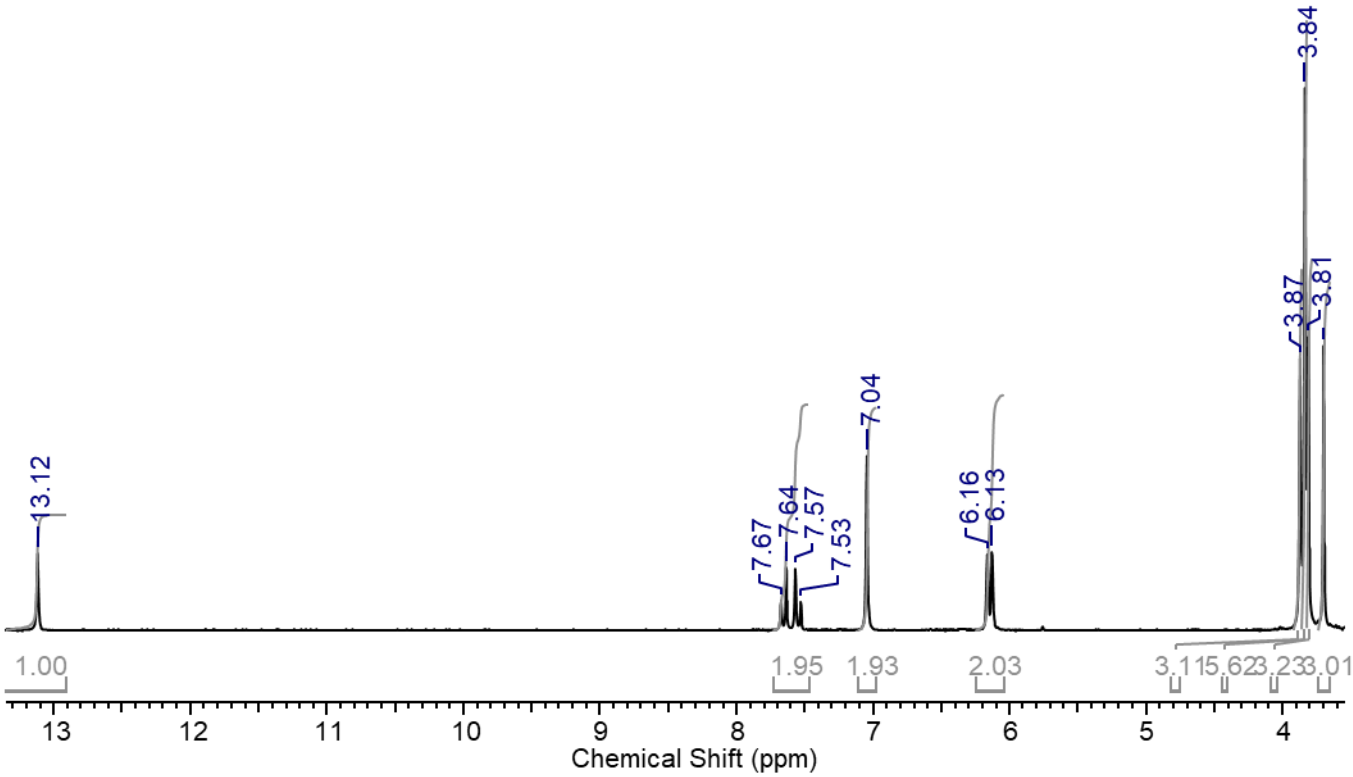
8 Appendix

8.1 Characterisation data for the synthesis of (E)-1-(2-hydroxy-4,6-dimethoxyphenyl)-3-(3,4,5-trimethoxyphenyl)prop-2-en-1-one, compound 2.2.

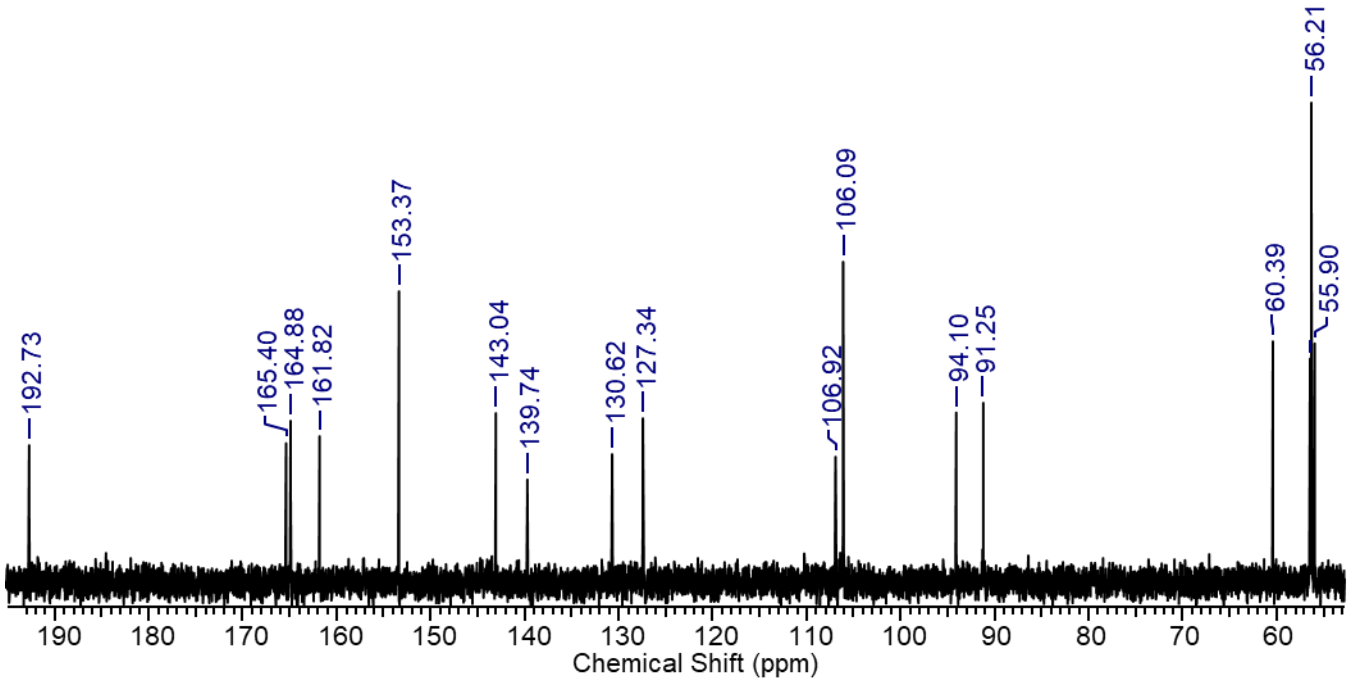
8.1.1 Extended ^{1}H NMR spectrum



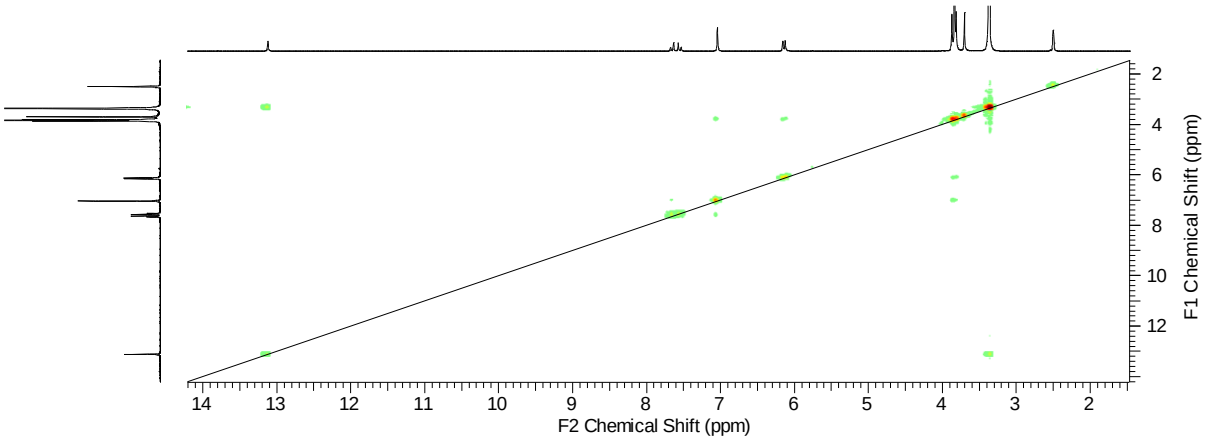
8.1.2 Magnified ^{1}H NMR spectrum



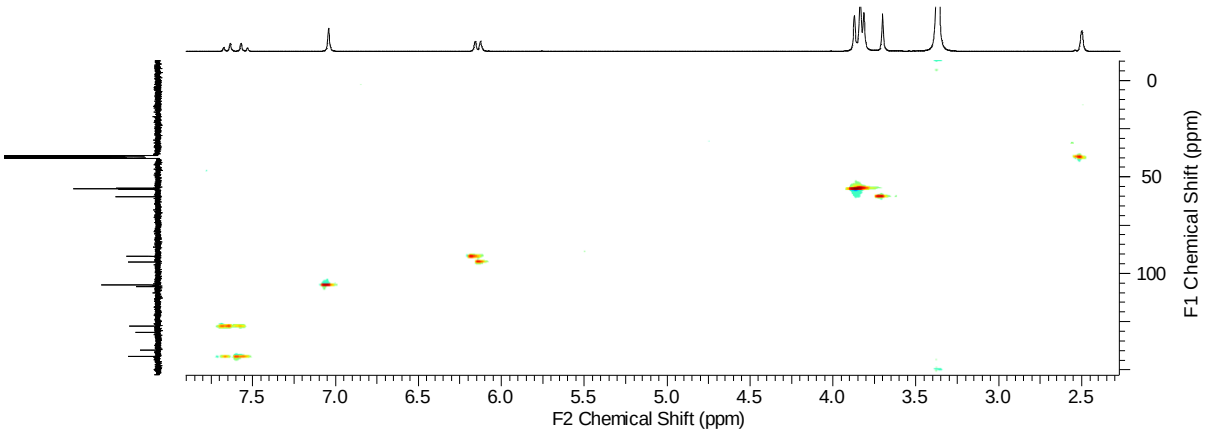
8.1.3 Magnified ^{13}C NMR spectrum



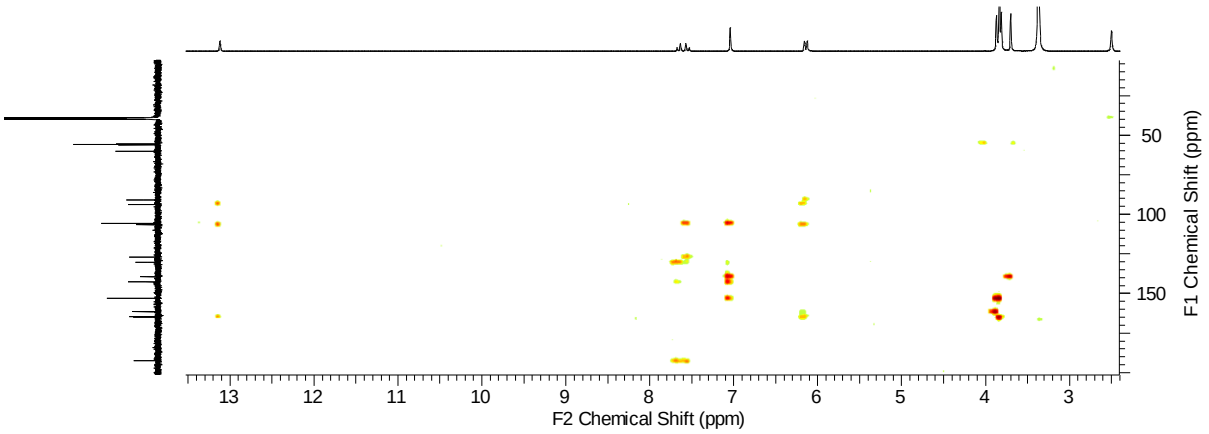
8.1.4 COSY spectrum



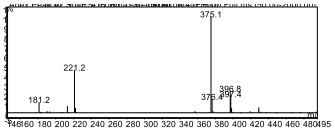
8.1.5 HSQC spectrum



8.1.6 HMBC spectrum

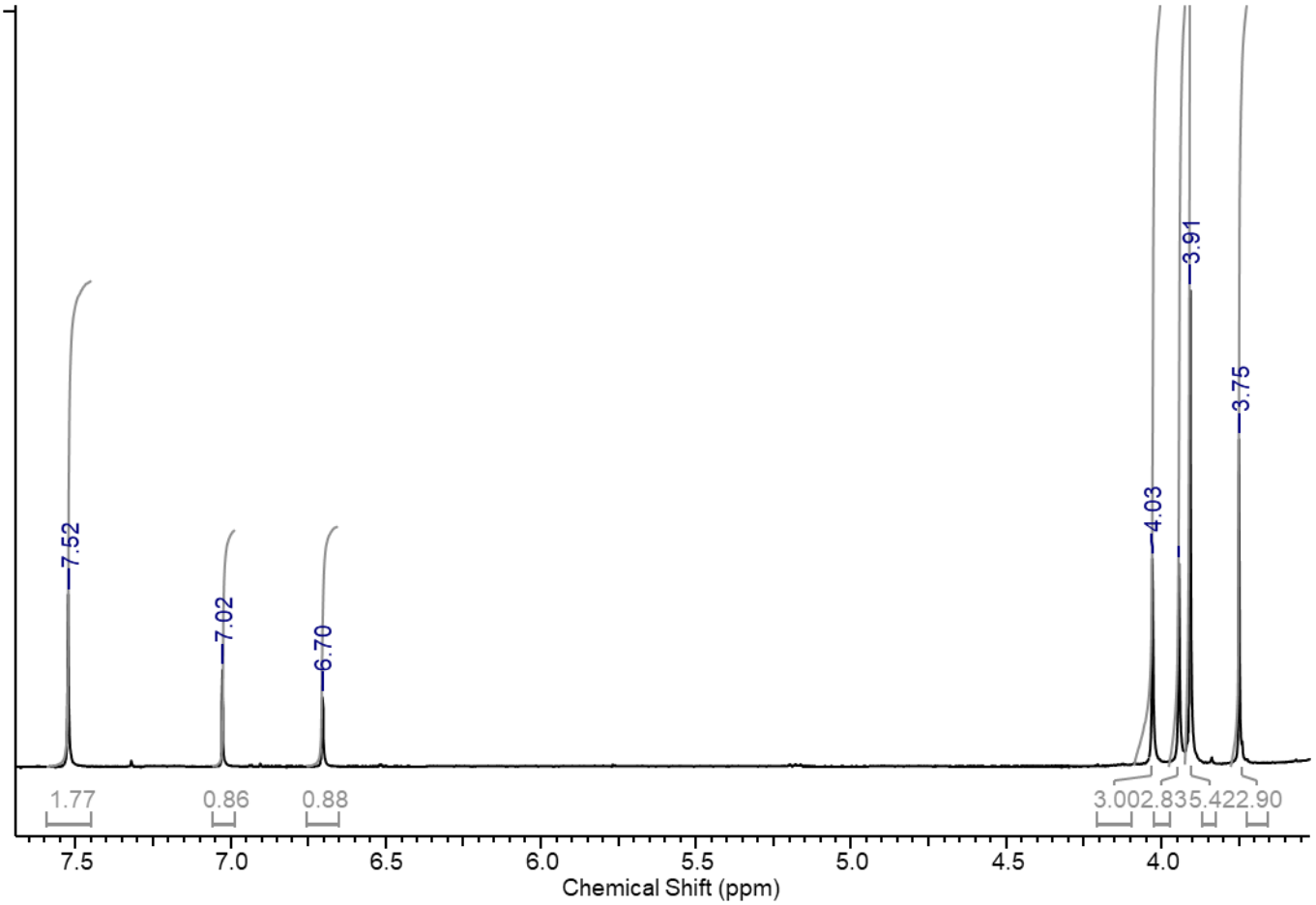


8.1.7 ESI-MS spectrum

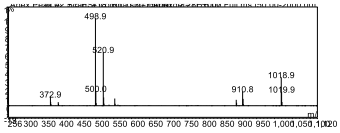


8.2 Characterisation data for the synthesis of ((3-iodo-5-methoxy-4-oxo-2-(3,4,5-trimethoxyphenyl)-4H-chromen-7-yl)oxy)methylium, compound 2.3

8.2.1 Magnified ^{1}H NMR spectrum

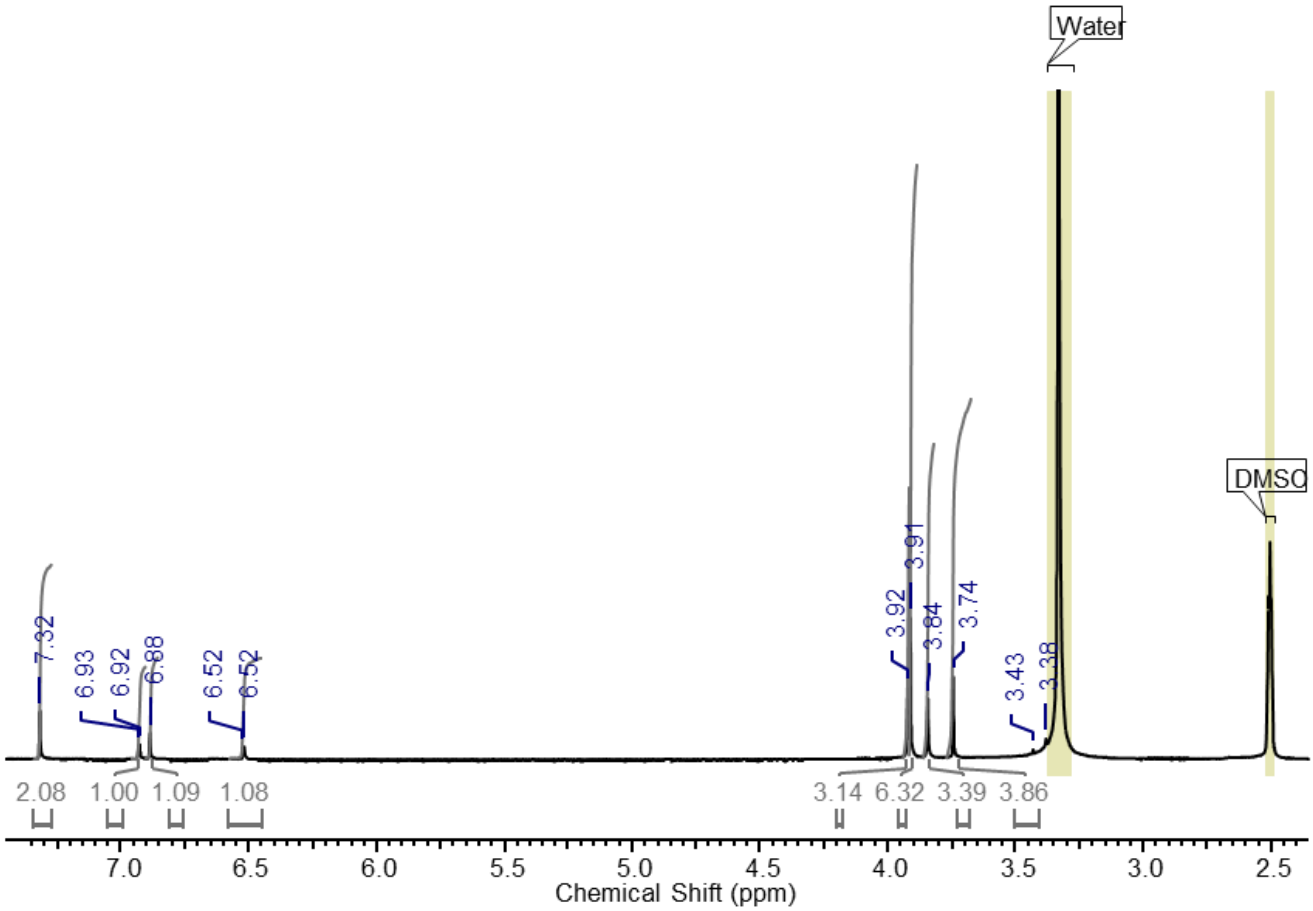


8.2.2 ESI-MS spectrum

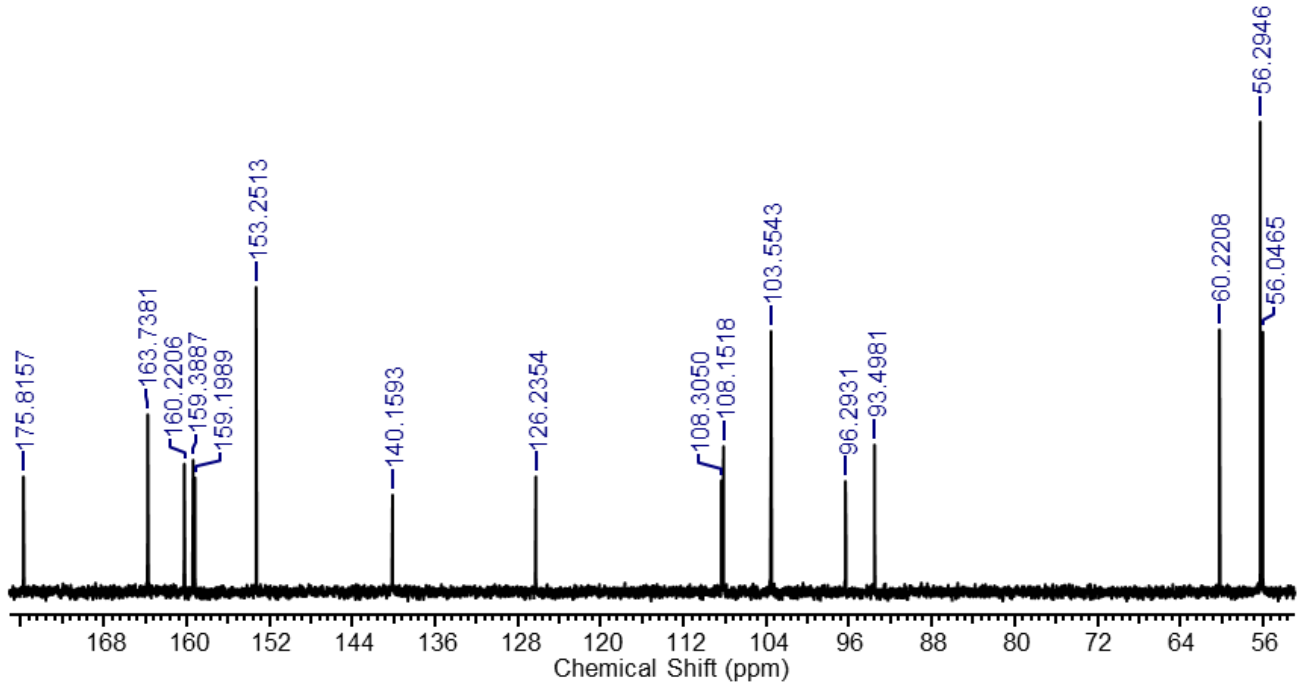


8.3 Characterisation data for the synthesis of 5,7-dimethoxy-2-(3,4,5-trimethoxyphenyl)-4H-chromen-4-one, compound 2.1.

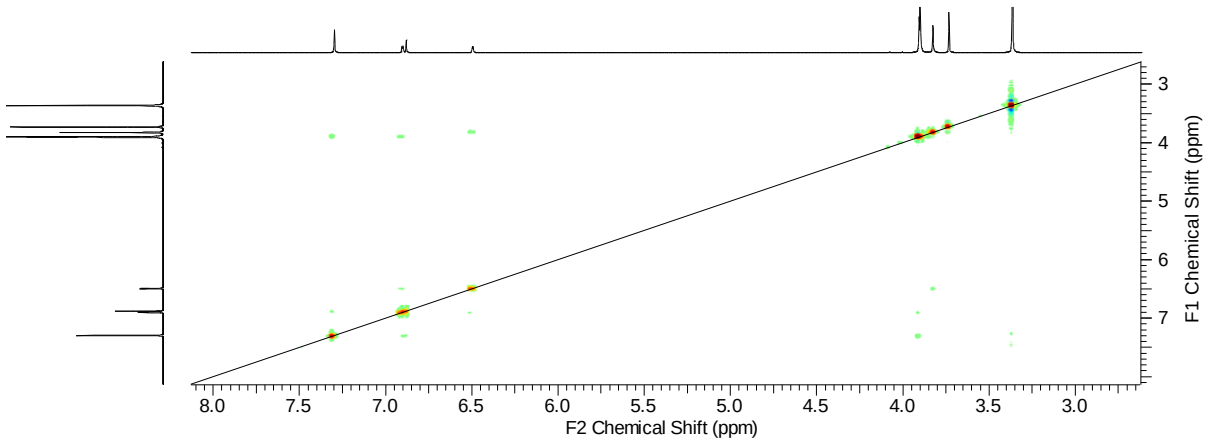
8.3.1 Expanded ^{1}H NMR spectrum



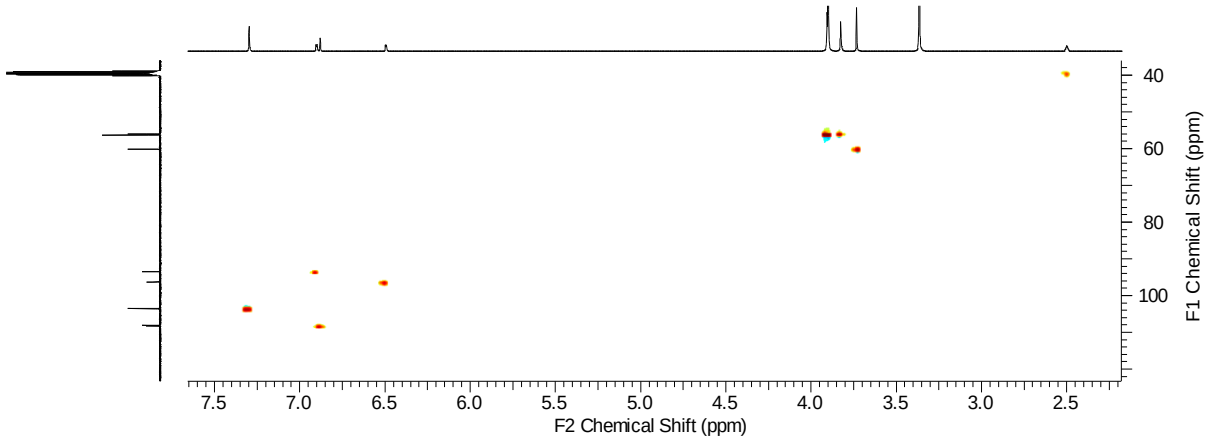
8.3.2 Magnified ^{13}C NMR spectrum



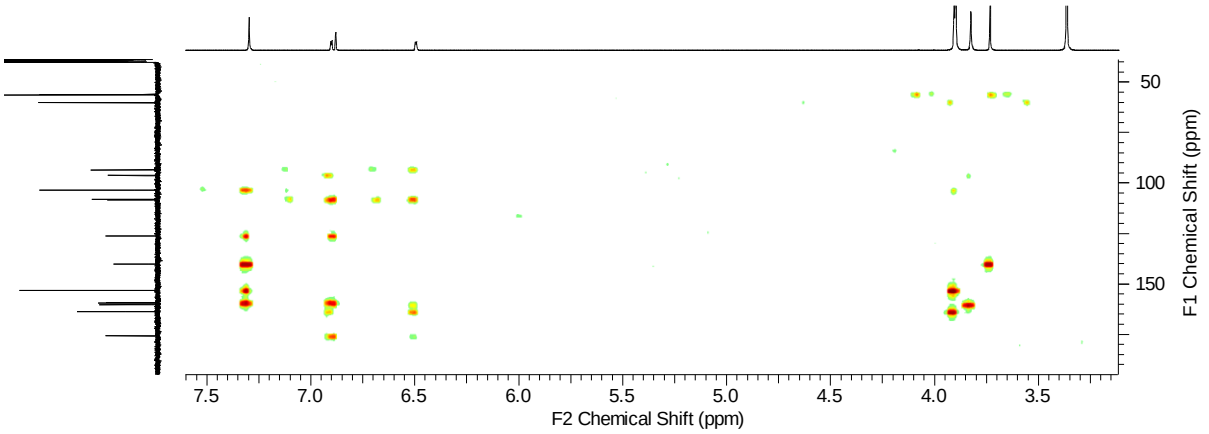
8.3.3 COSY spectrum



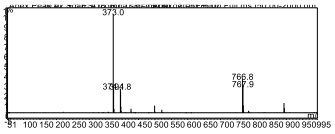
8.3.4 HSQC spectrum



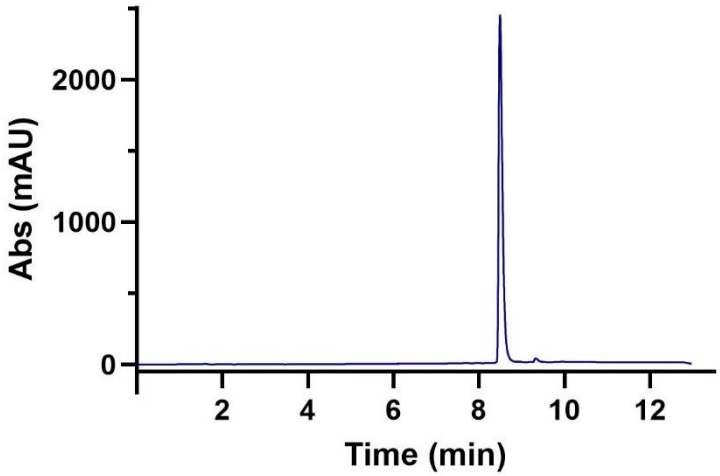
8.3.5 HMBC spectrum



8.3.6 ESI-MS spectrum

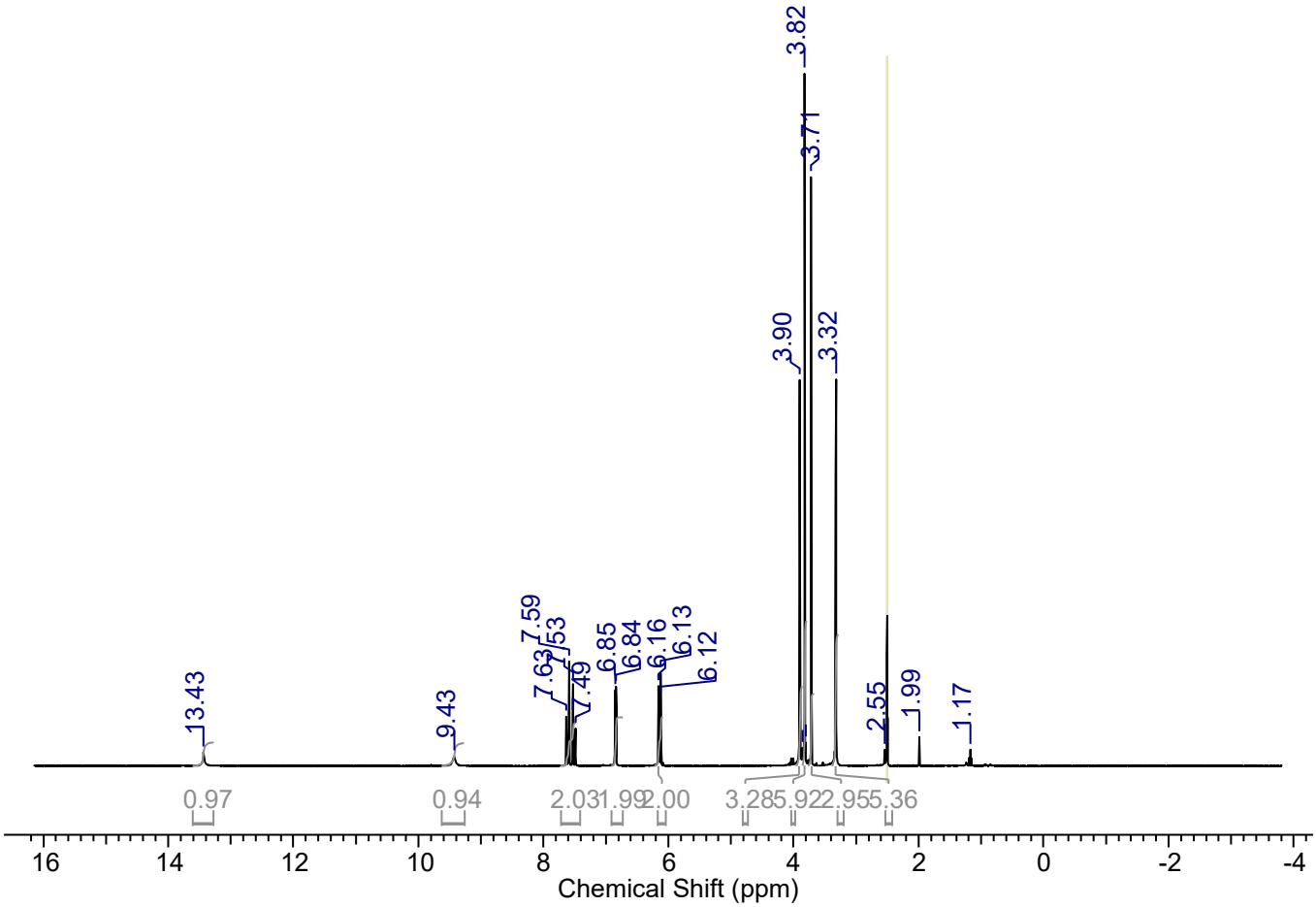


8.3.7 UV chromatogram of purified compound showing absorbance at 254 nm.

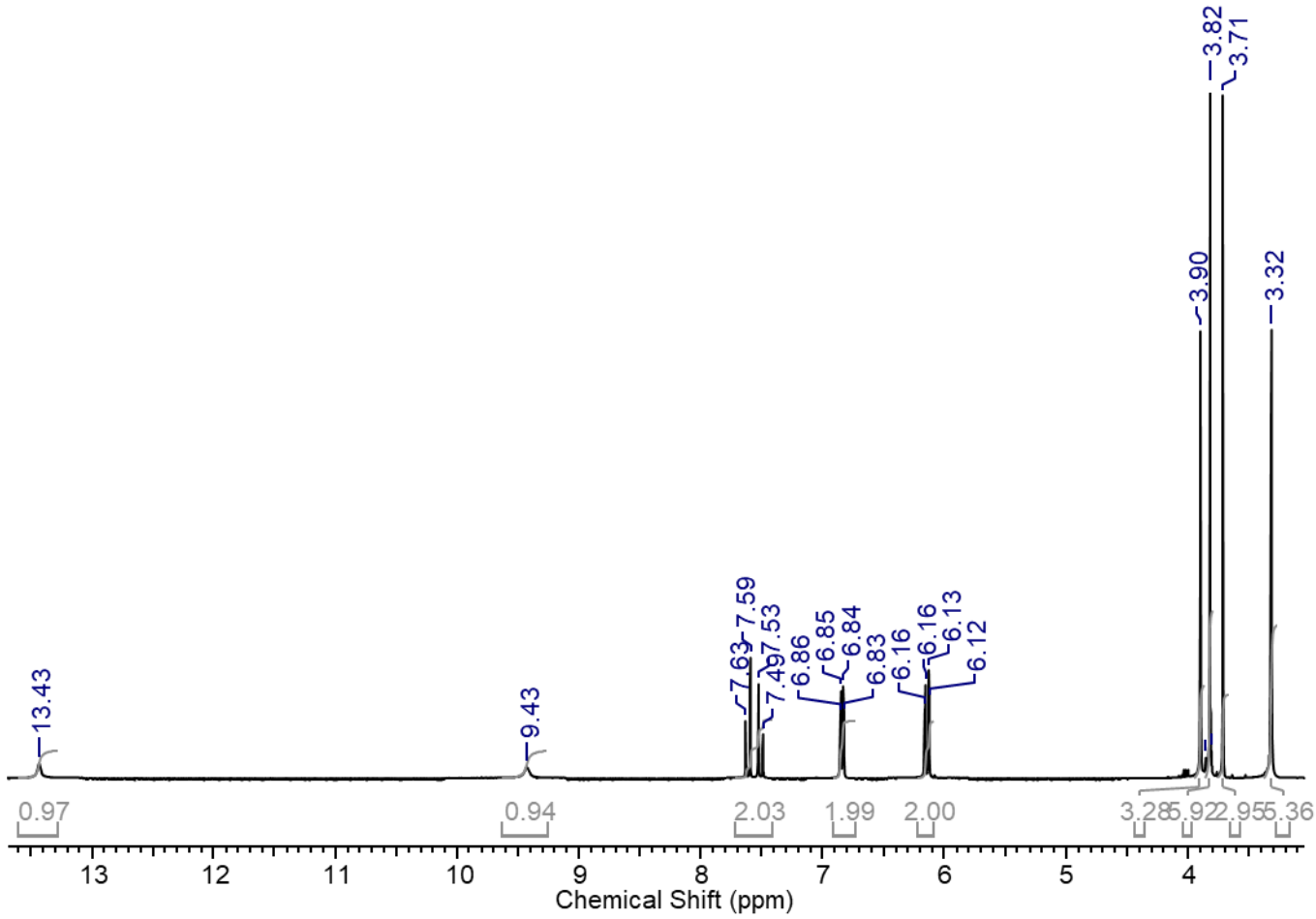


8.4 Characterisation data for the synthesis of (E)-3-(3-hydroxy-4,5-dimethoxyphenyl)-1-(2-hydroxy-4,6-dimethoxyphenyl)prop-2-en-1-one, compound 2.4.

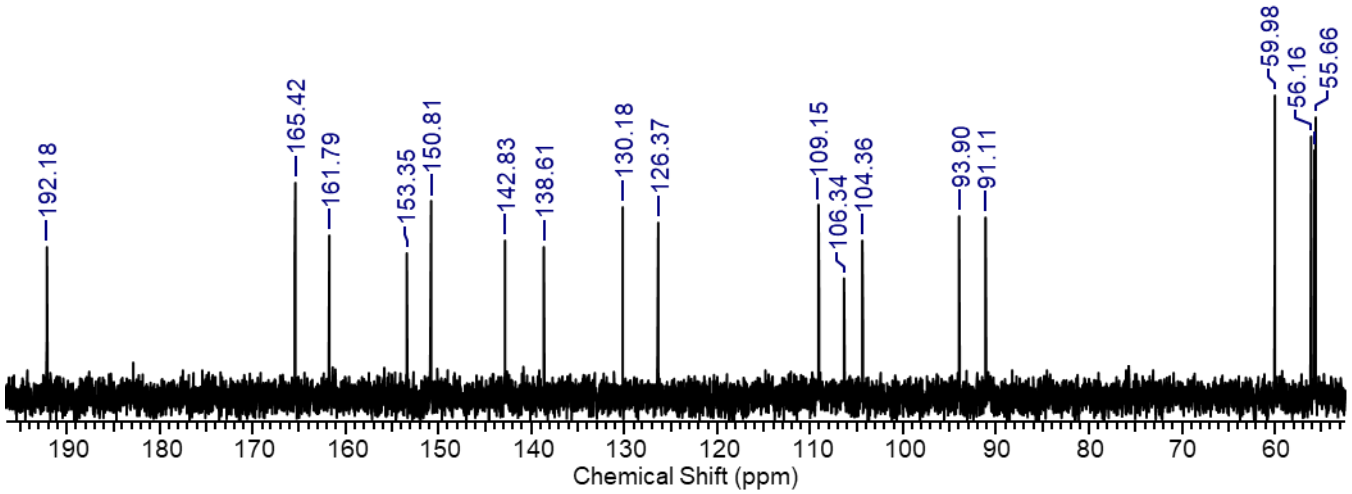
8.4.1 Extended ^{1}H NMR spectrum



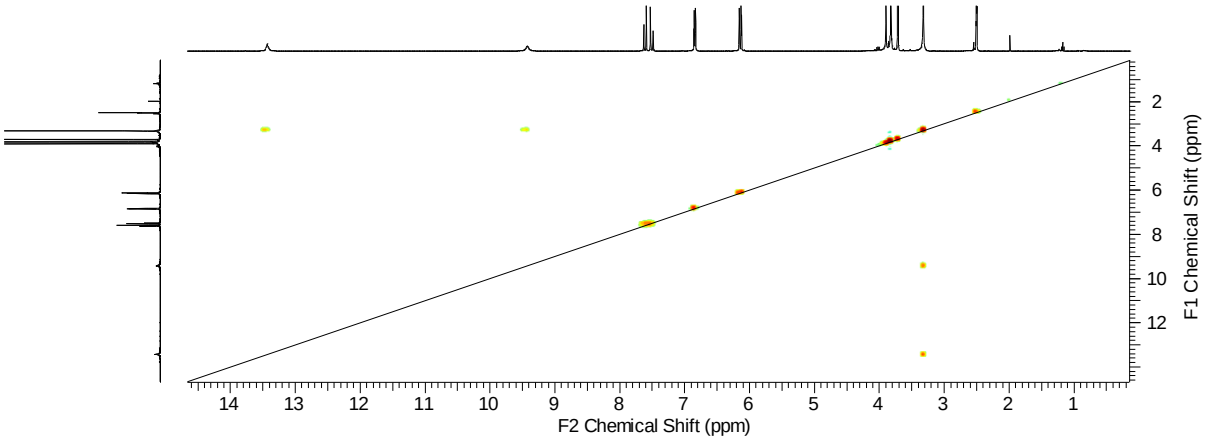
8.4.2 Magnified ^{1}H NMR spectrum



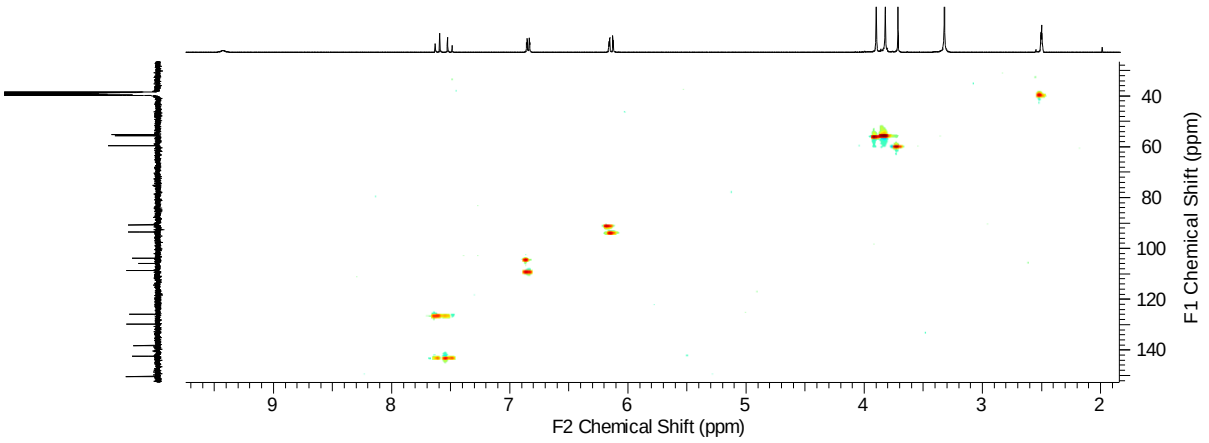
8.4.3 Magnified ^{13}C NMR spectrum



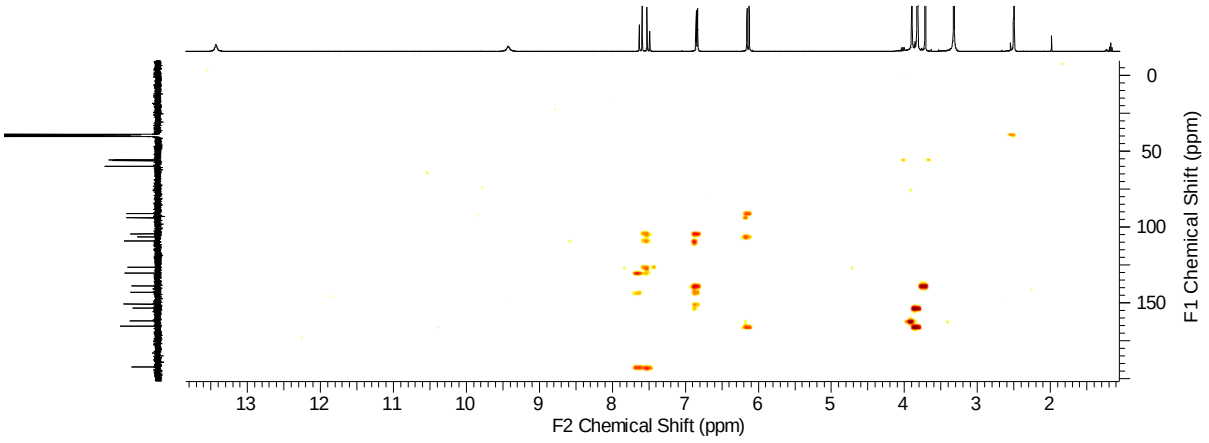
8.4.4 COSY spectrum



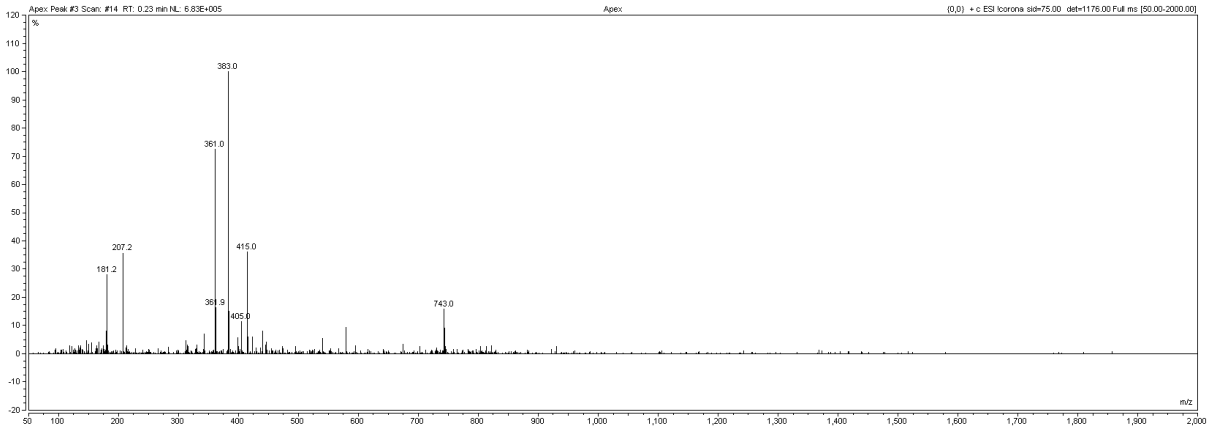
8.4.5 HSQC spectrum



8.4.6 HMBC spectrum

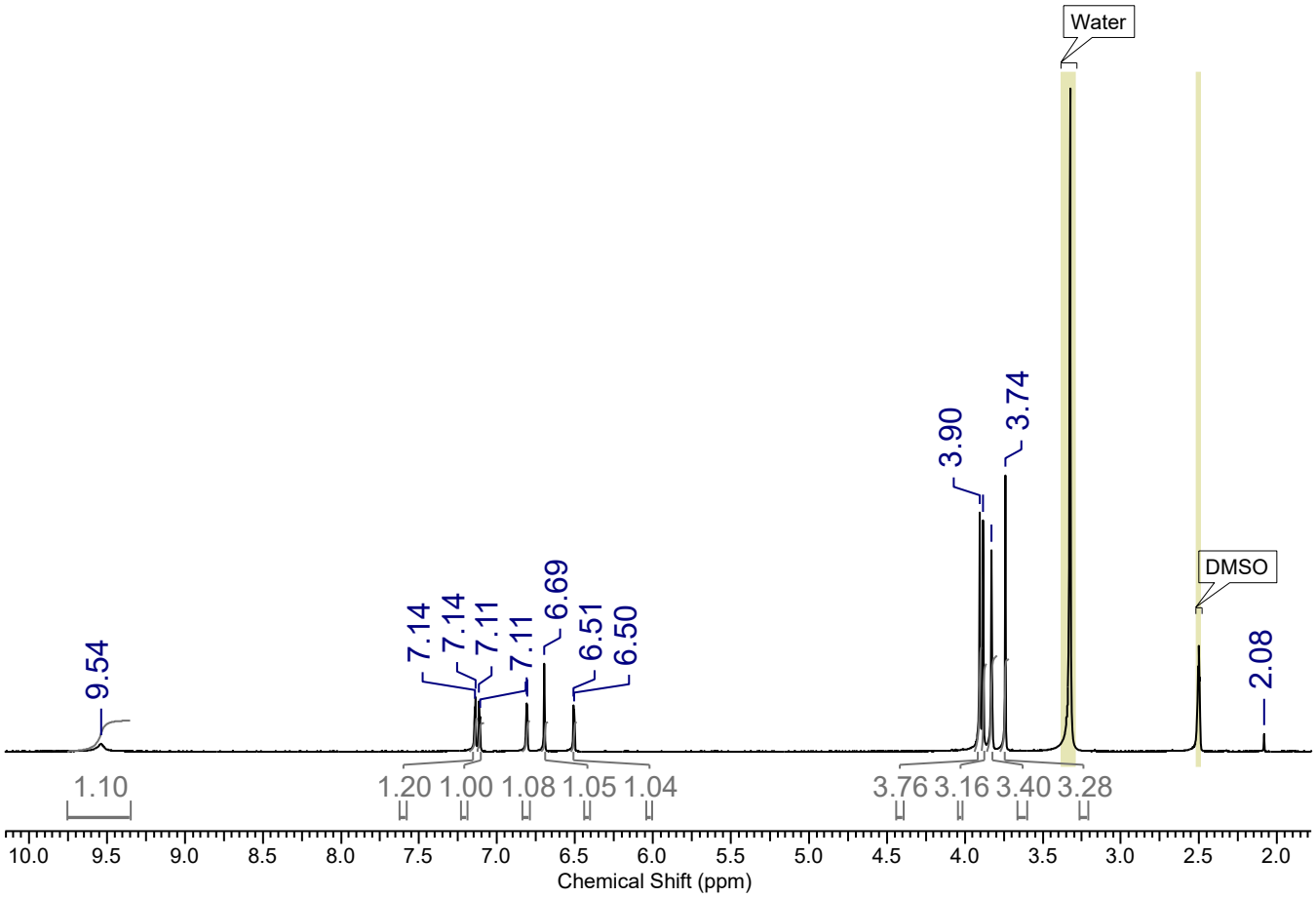


8.4.7 ESI-MS spectrum

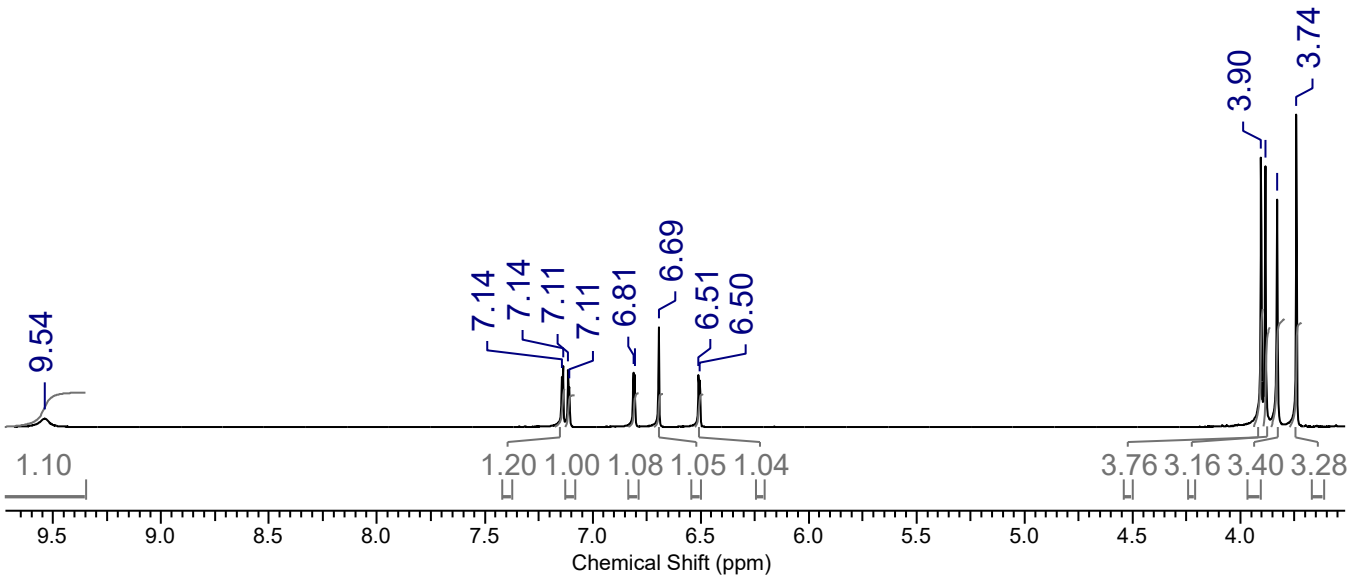


8.5 Characterisation data for the synthesis of 2-(3-hydroxy-4,5-dimethoxyphenyl)-5,7-dimethoxy-4H-chromen-4-one, compound 2.5.

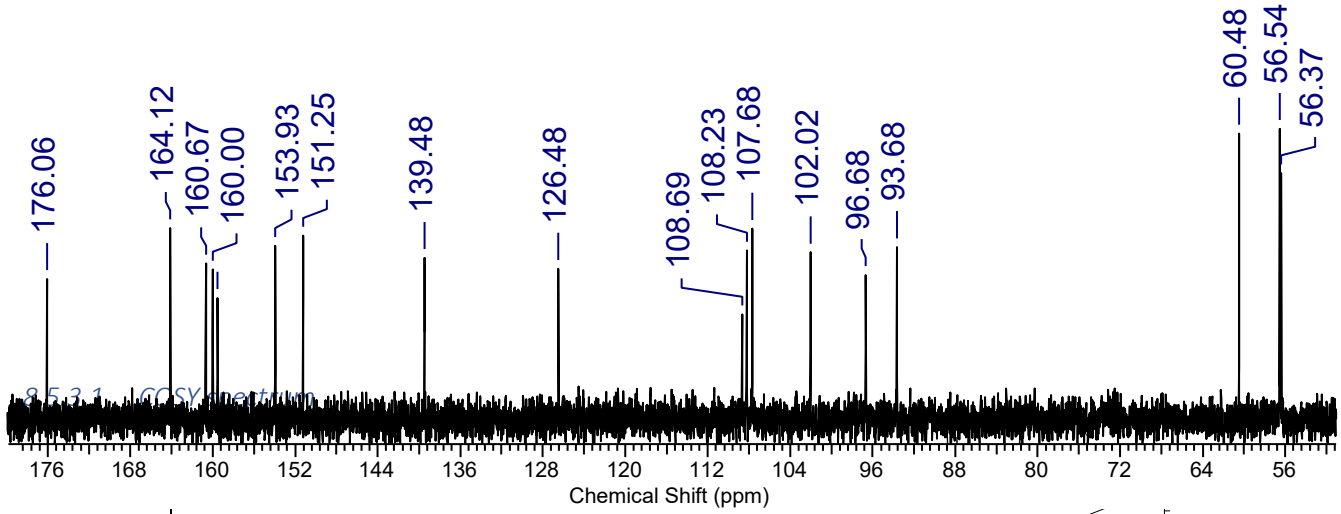
8.5.1 Extended ^{1}H NMR spectrum



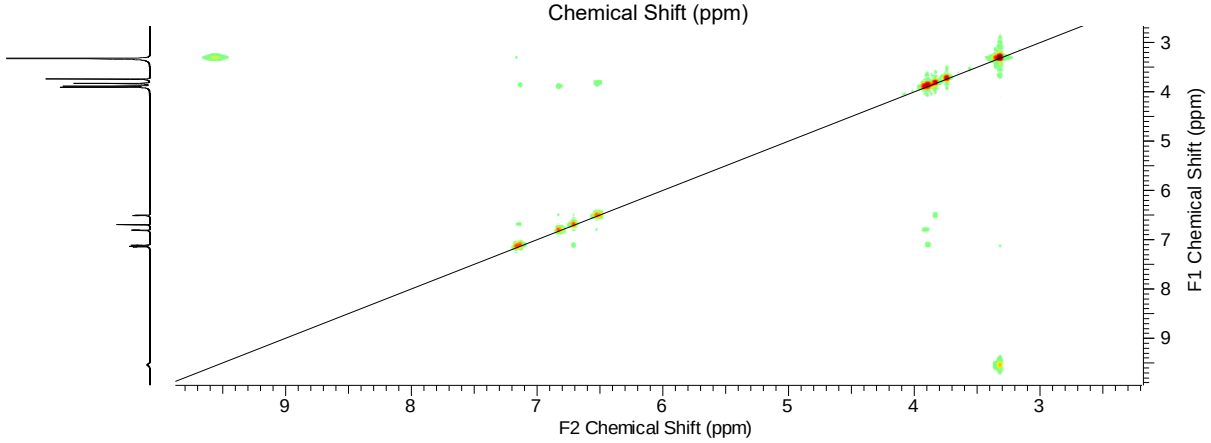
8.5.2 Magnified ^{1}NMR spectrum



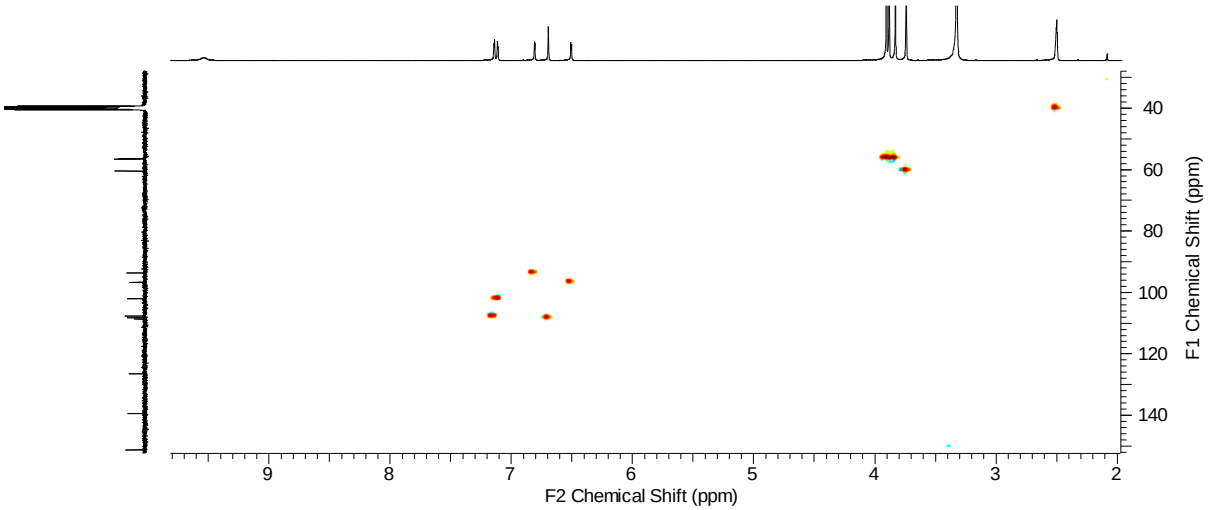
8.5.3 Magnified ^{13}C NMR spectrum



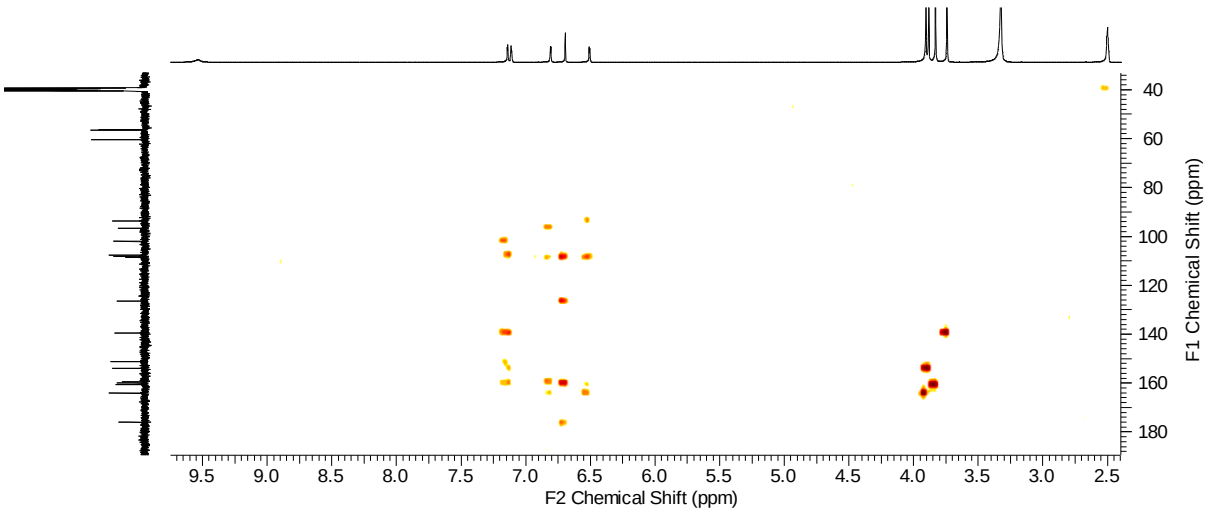
8.5.3.1 COSY spectrum



8.5.4 HSQC spectrum

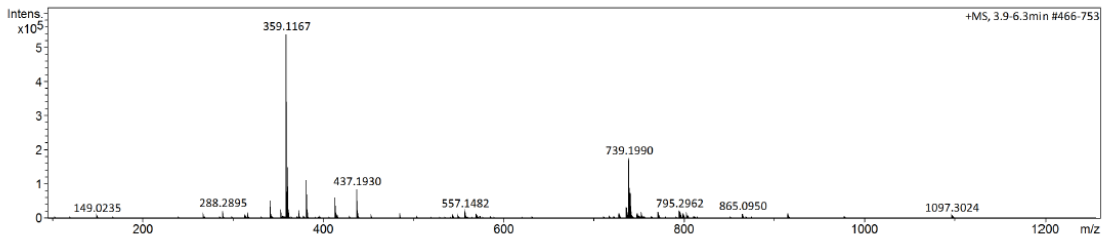


8.5.5 HMBC spectrum

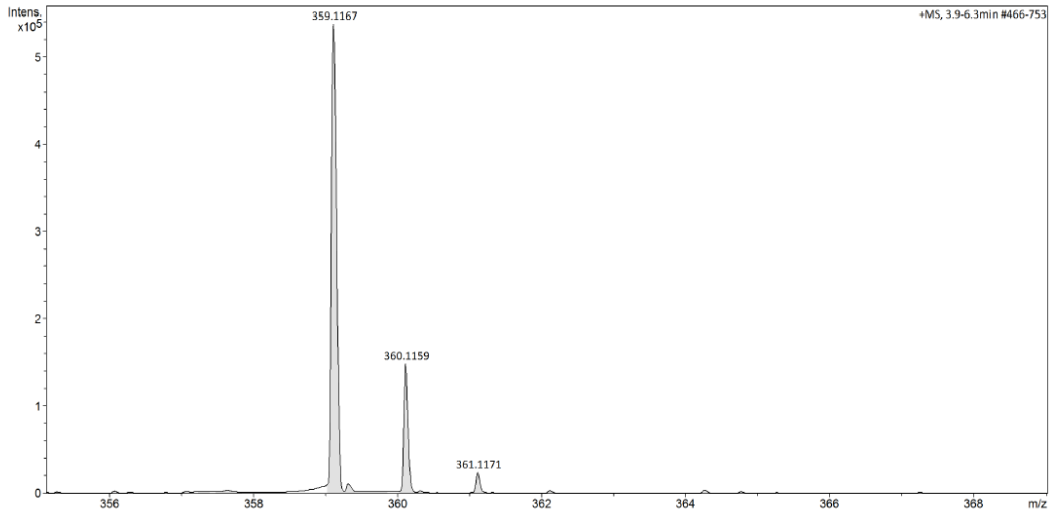


8.5.6 HRMS spectra

a)

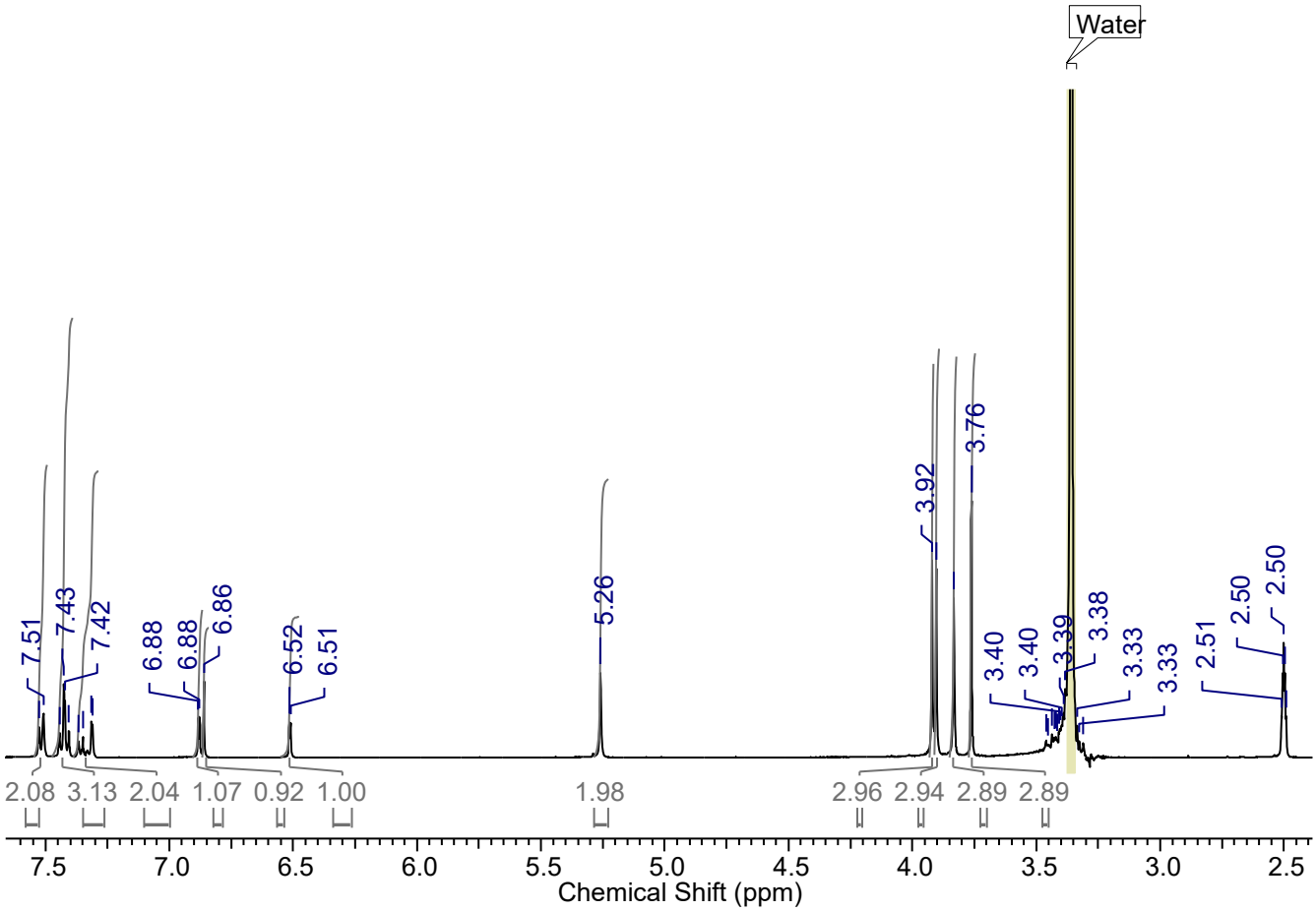


b)

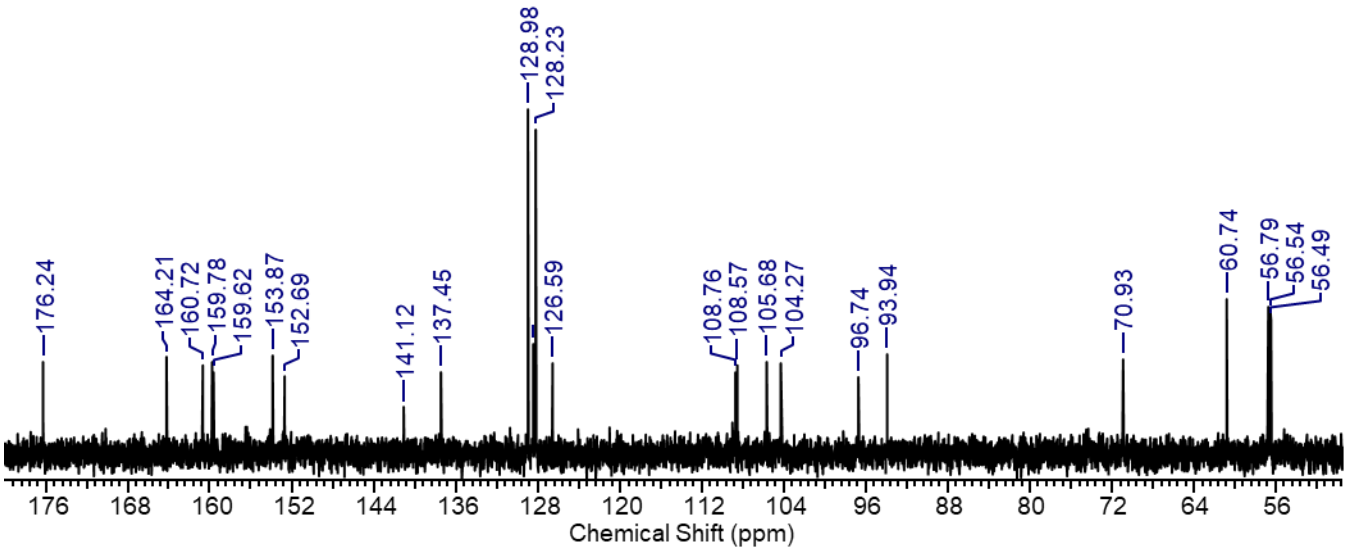


8.6 Characterisation data for the synthesis of 2-(3-(benzyloxy)-4,5-dimethoxyphenyl)-5,7-dimethoxy-4H-chromen-4-one, compound 2.6.

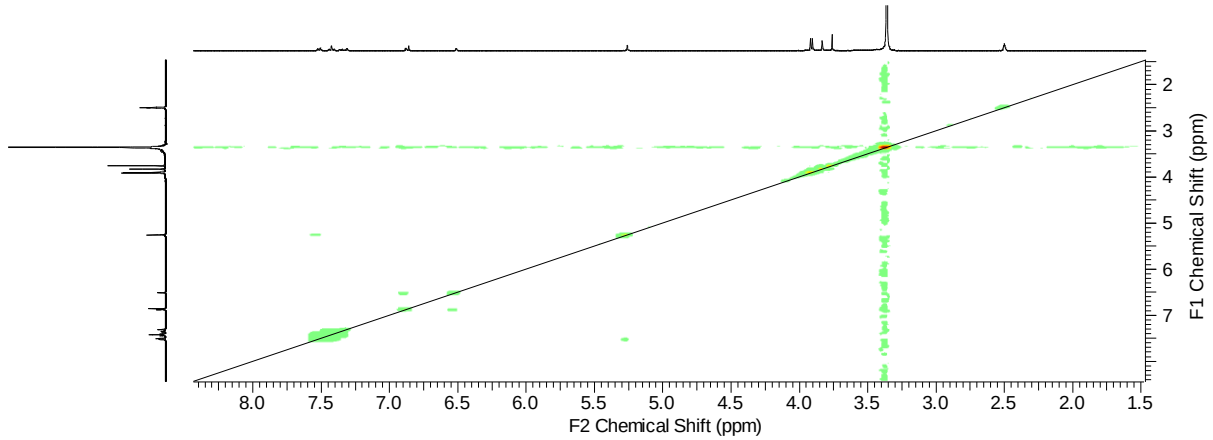
8.6.1 Extended ^{1}H NMR spectrum



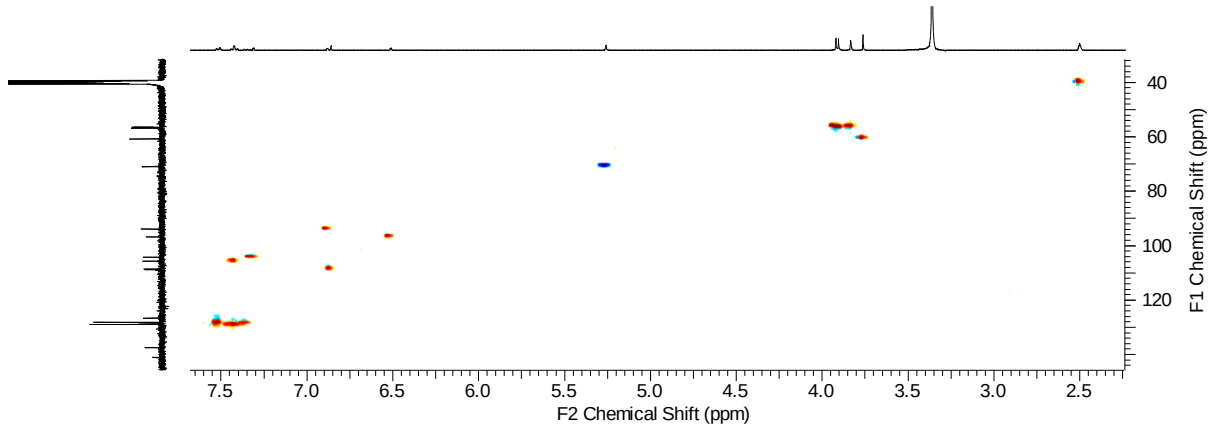
8.6.2 Magnified ^{13}C NMR spectrum



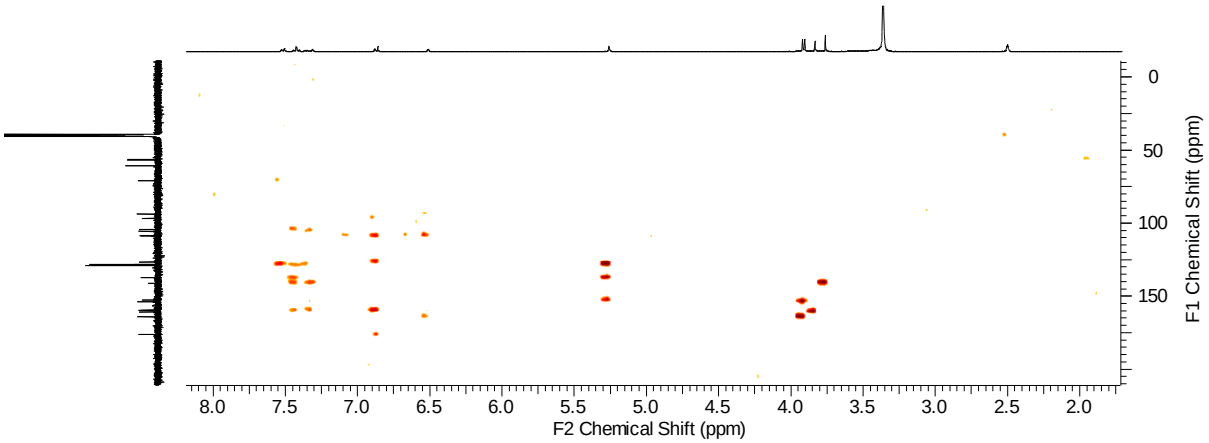
8.6.3 COSY spectrum



8.6.4 HSQC spectrum

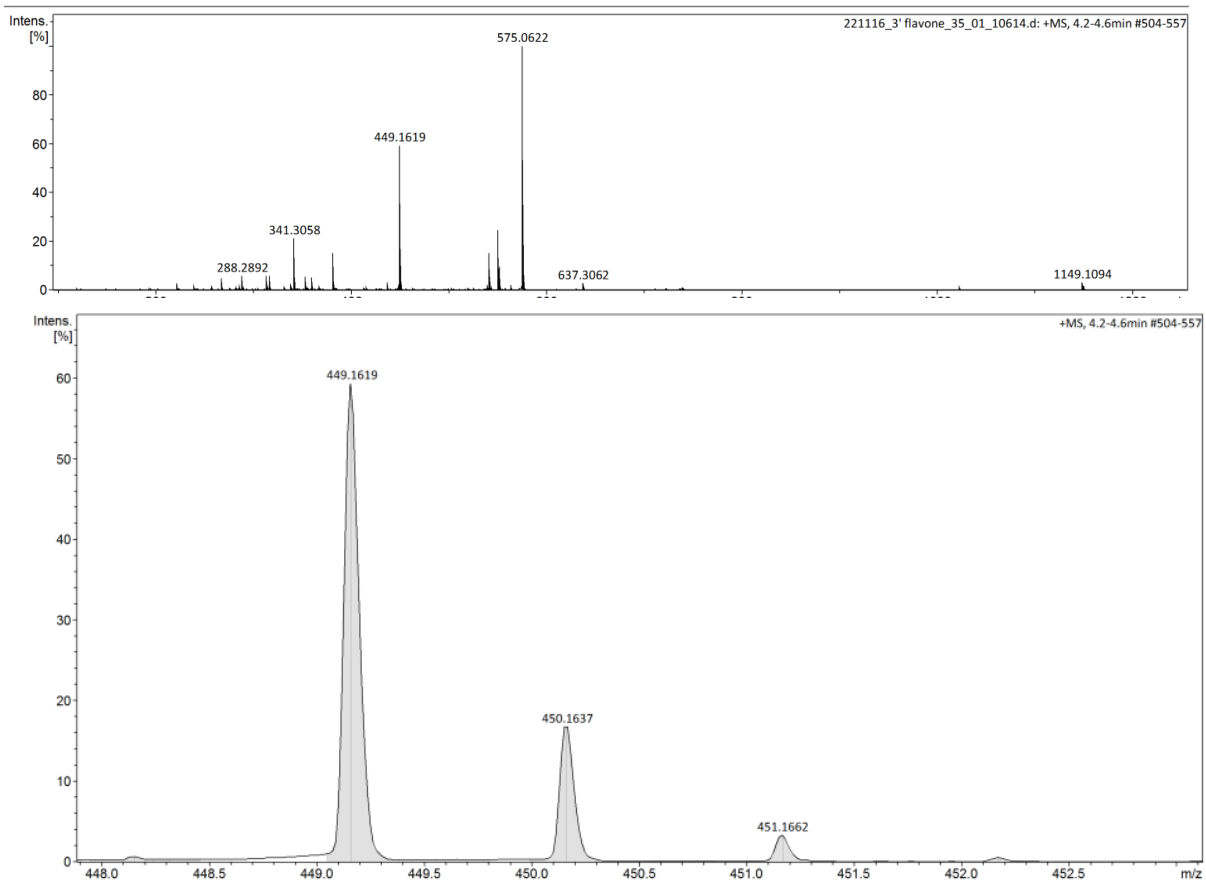


8.6.5 HMBC spectrum

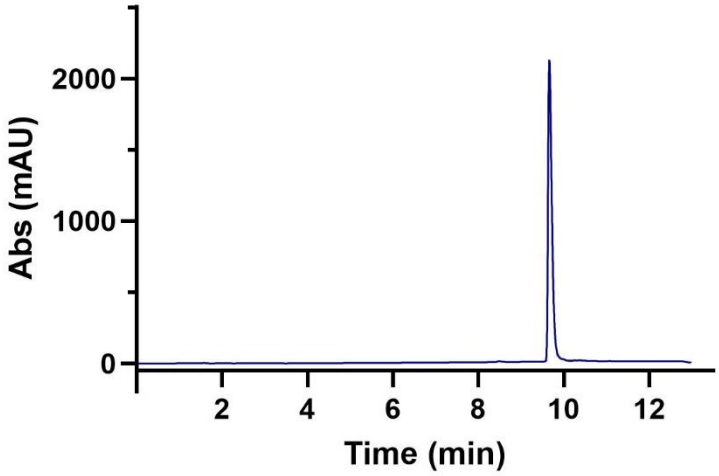


8.6.6 HRMS spectra

a)

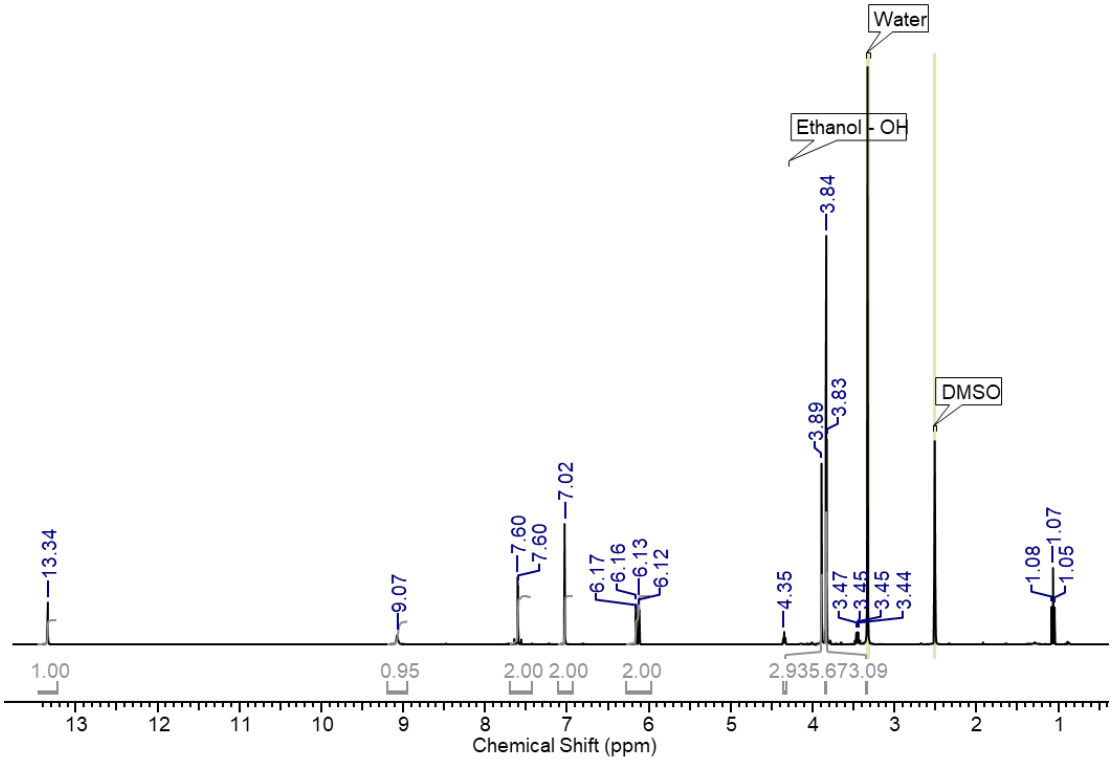


8.6.7 UV chromatogram of purified compound showing absorbance at 254 nm.

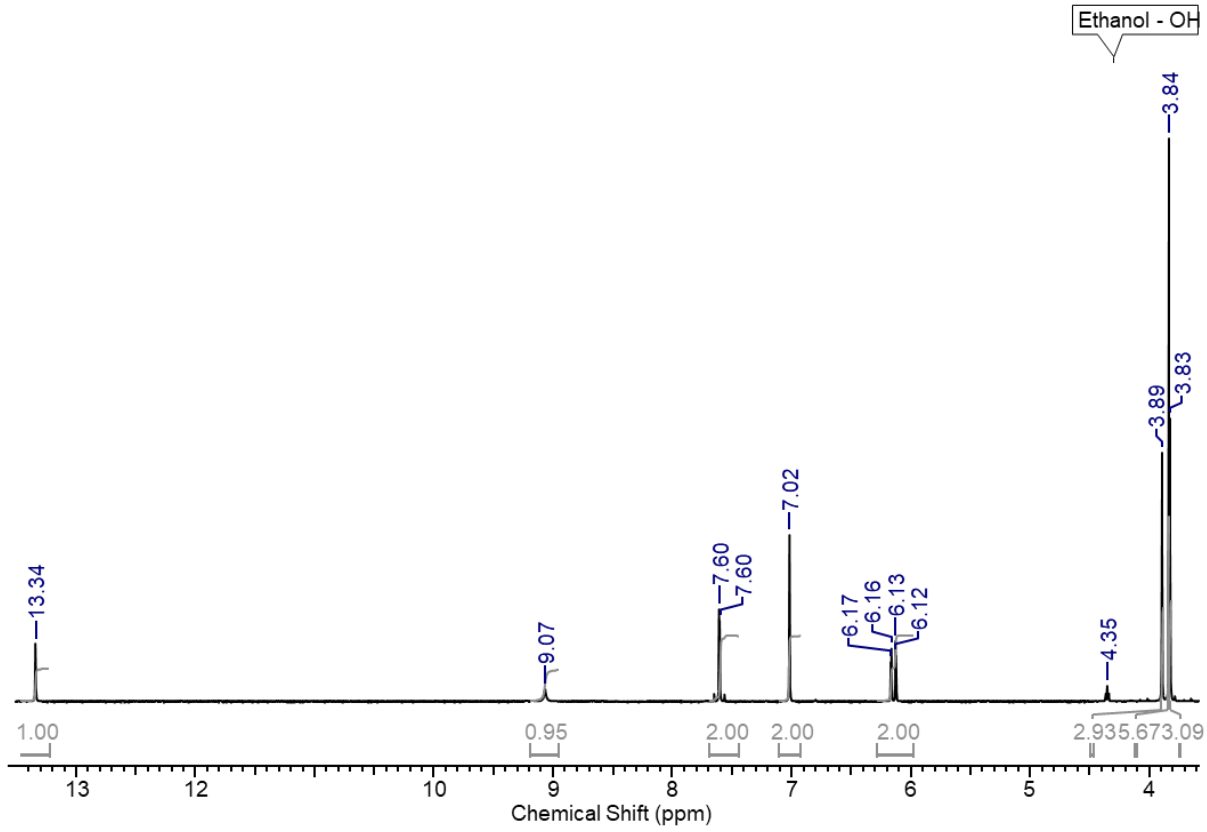


8.7 Characterisation data for the synthesis of (E)-3-(4-hydroxy-3,5-dimethoxyphenyl)-1-(2-hydroxy-4,6-dimethoxyphenyl)prop-2-en-1-one.

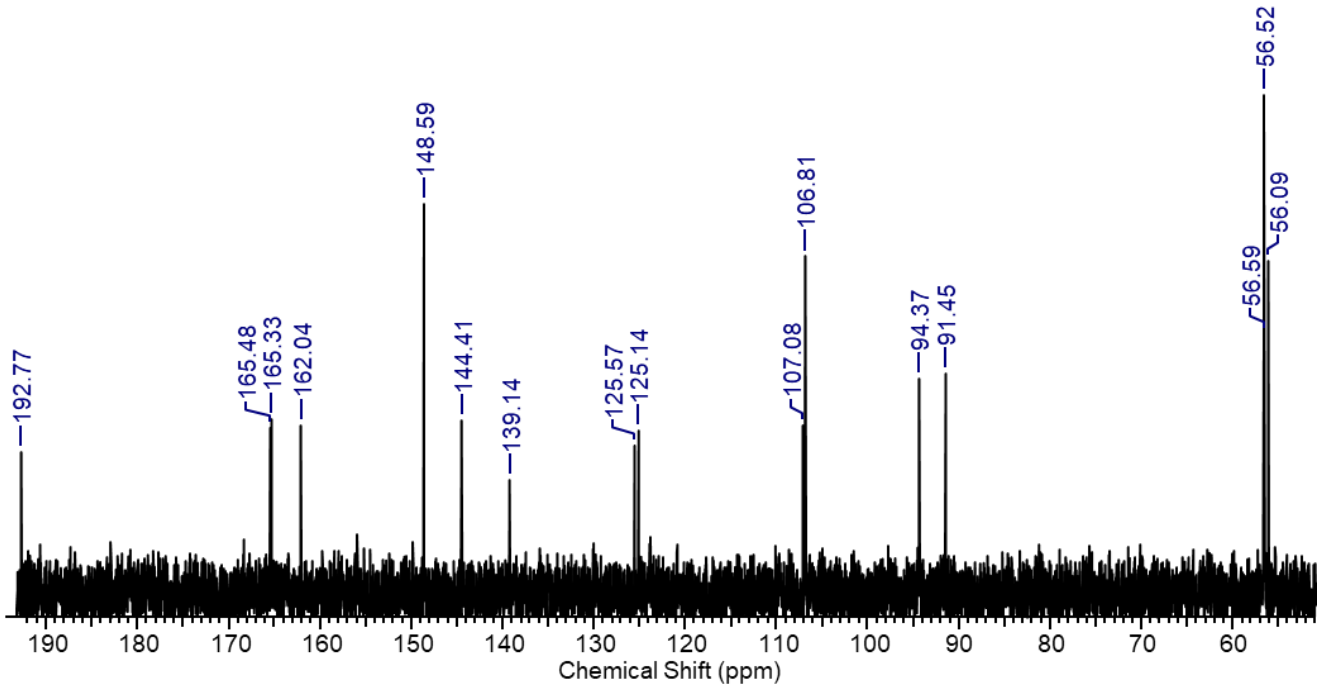
8.7.1 Extended ^{1}H NMR spectrum



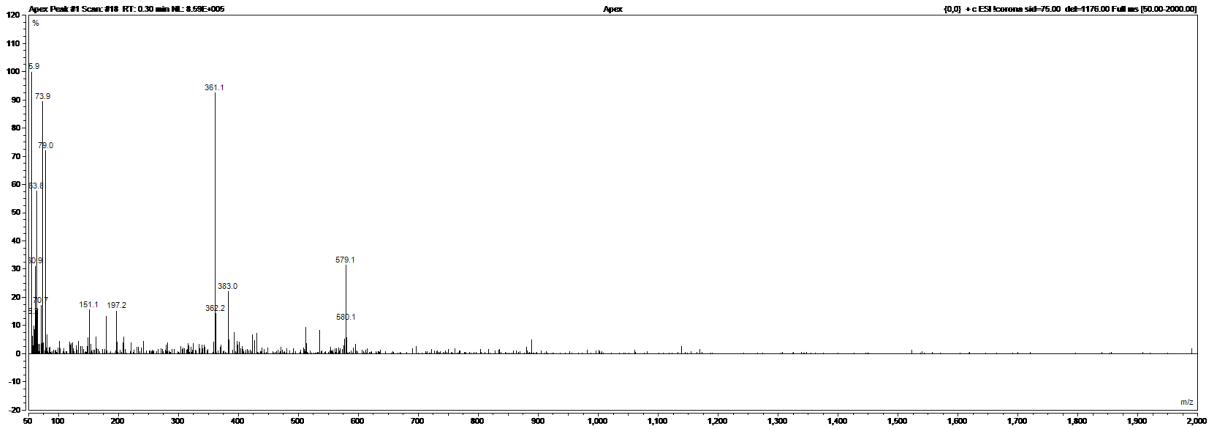
8.7.2 Magnified ^{1}H NMR spectrum



8.7.3 Magnified ^{13}C NMR spectrum

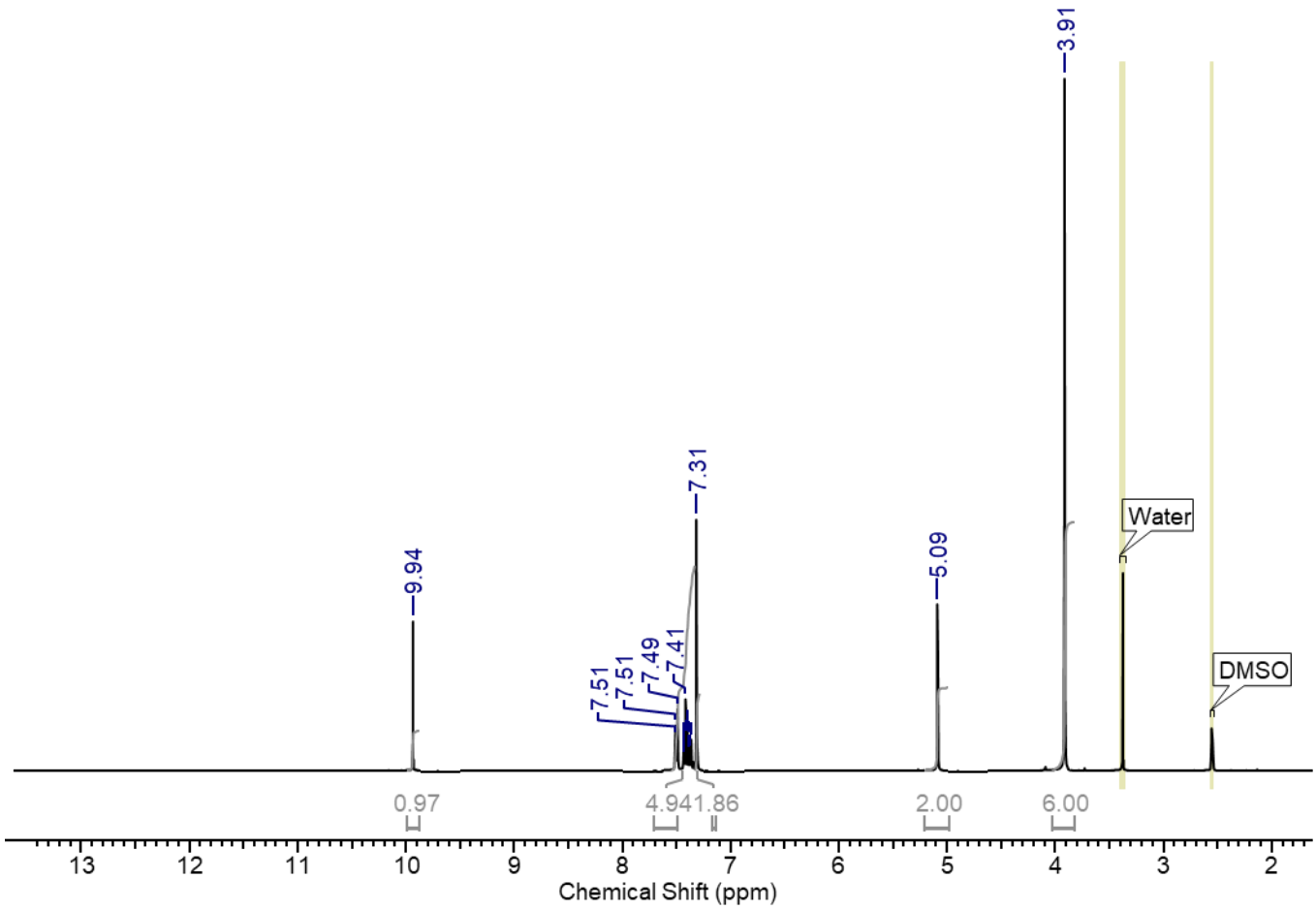


8.7.4 ESI-MS spectrum

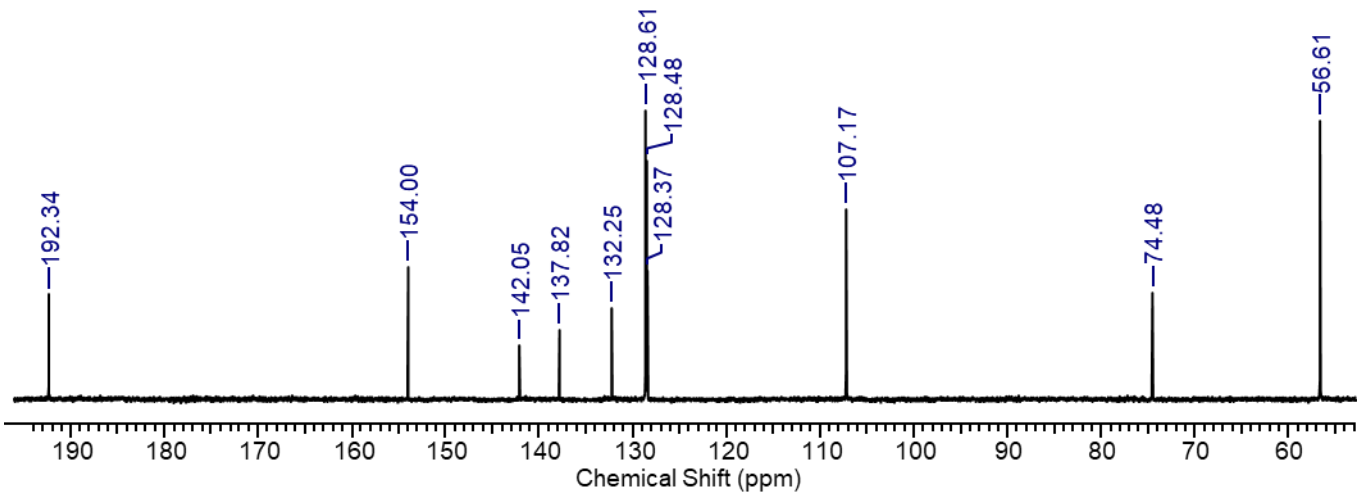


8.8 Characterisation data for the synthesis of 4-(benzyloxy)-3,5-dimethoxybenzaldehyde, compound 2.7.

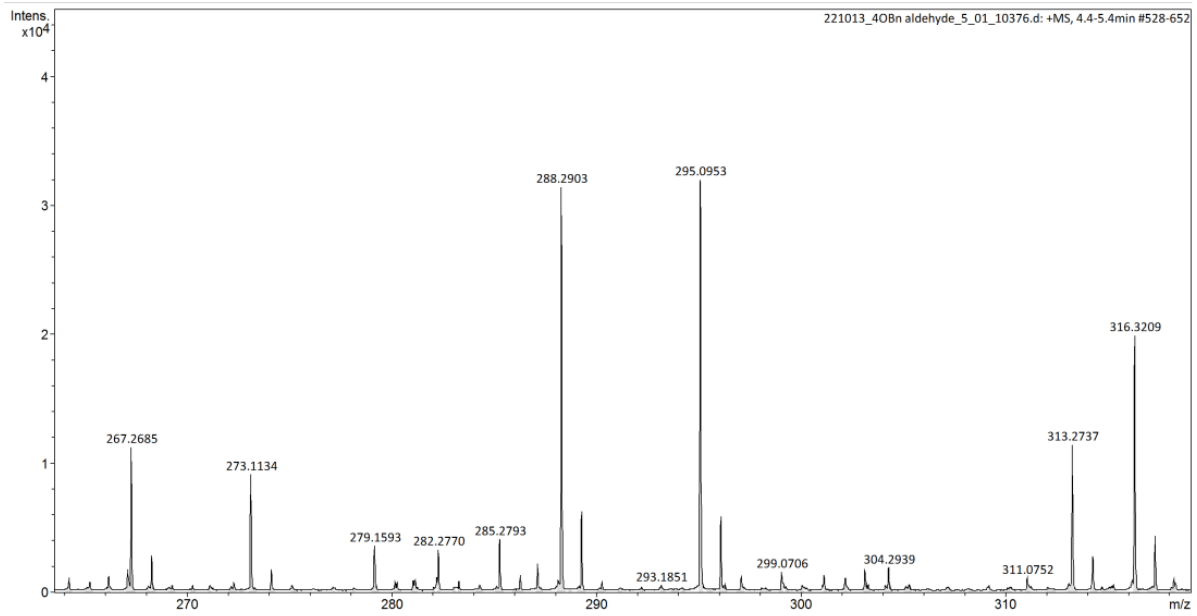
8.8.1 Extended ^{1}H NMR spectrum



8.8.2 Magnified ^{13}C NMR spectrum

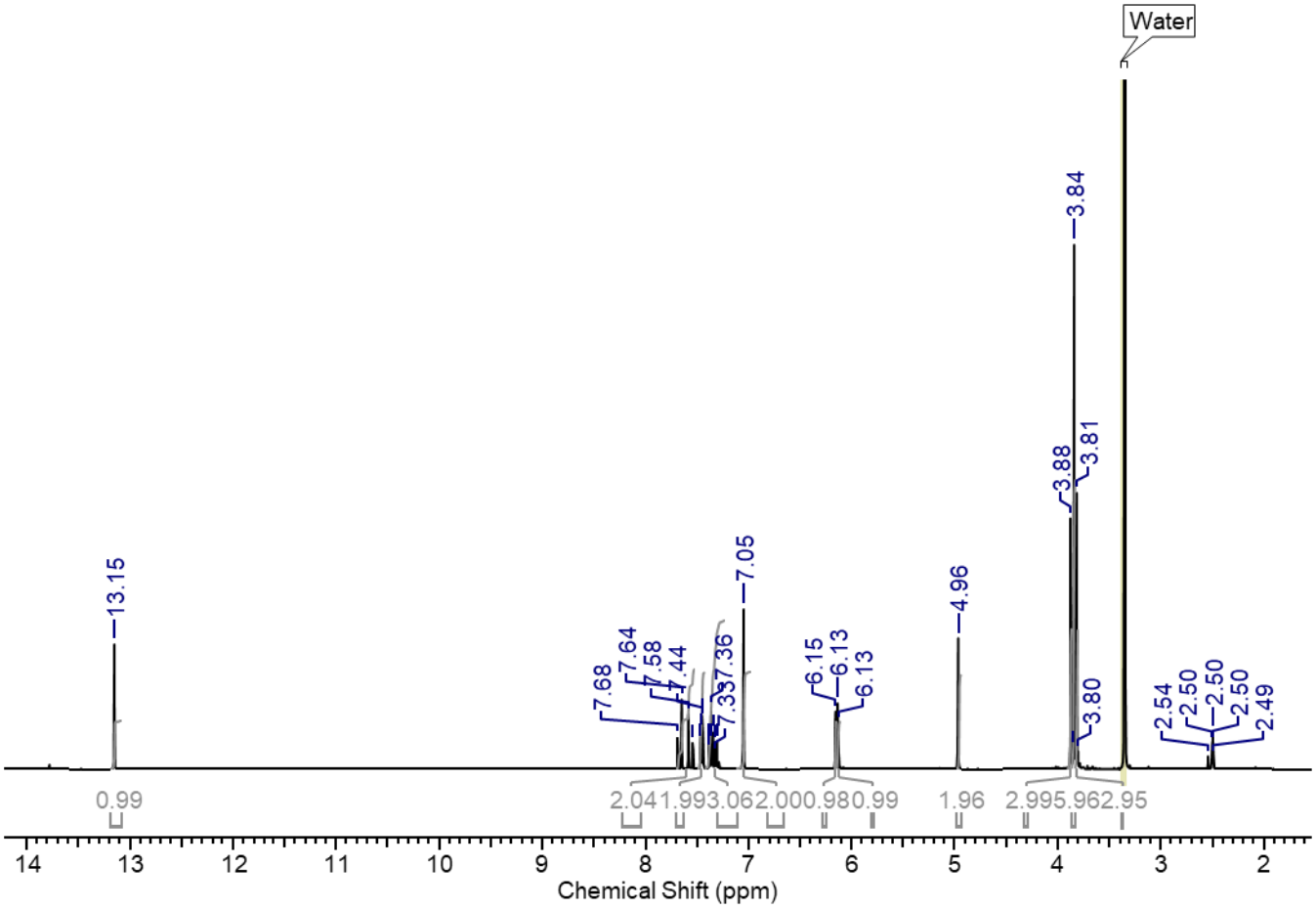


8.8.3 HRMS spectrum

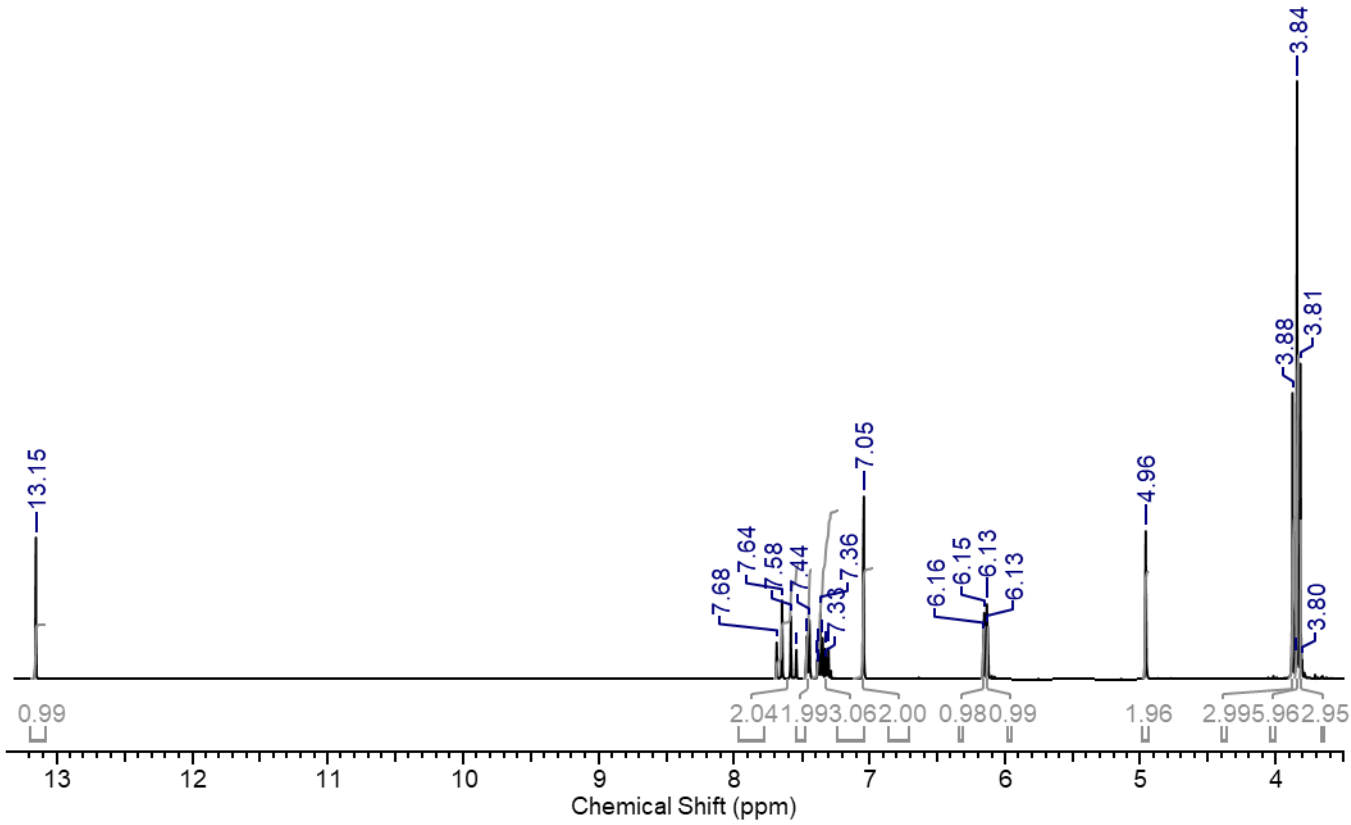


8.9 Characterisation data for the synthesis of (E)-3-(4-(benzyloxy)-3,5-dimethoxyphenyl)-1-(2-hydroxy-4,6-dimethoxyphenyl)prop-2-en-1-one, compound 2.8.

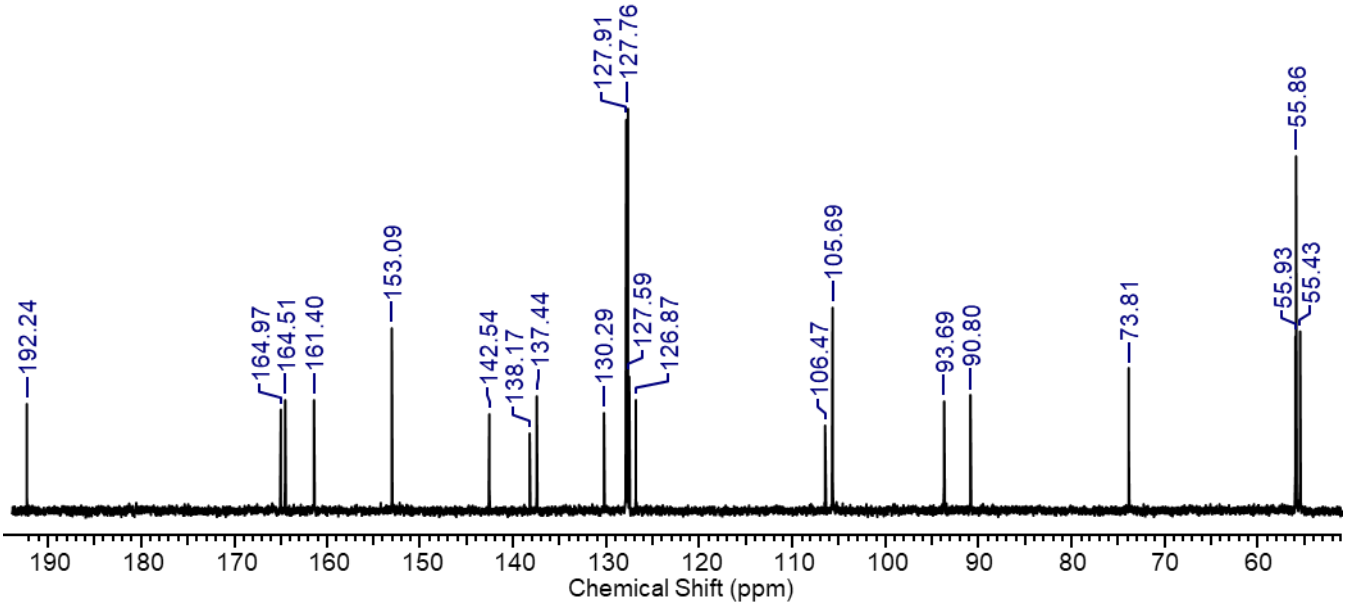
8.9.1 Extended ^{1}H NMR spectrum



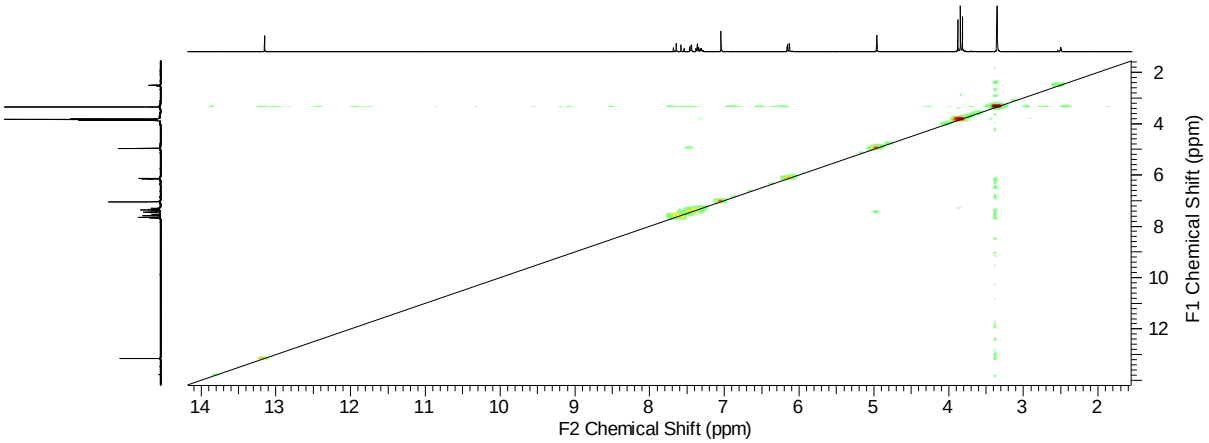
8.9.2 Magnified ^{1}H NMR spectrum



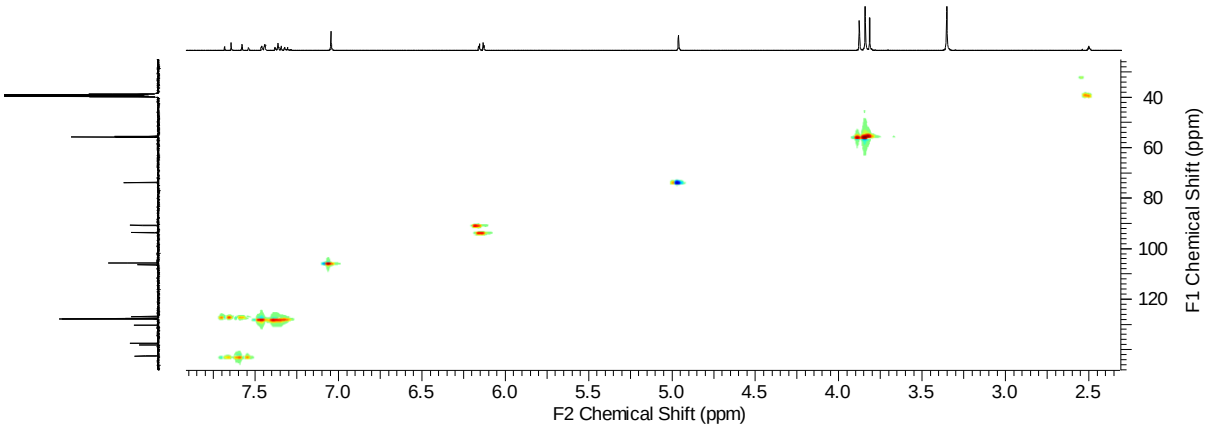
8.9.3 Magnified ^{13}C NMR spectrum



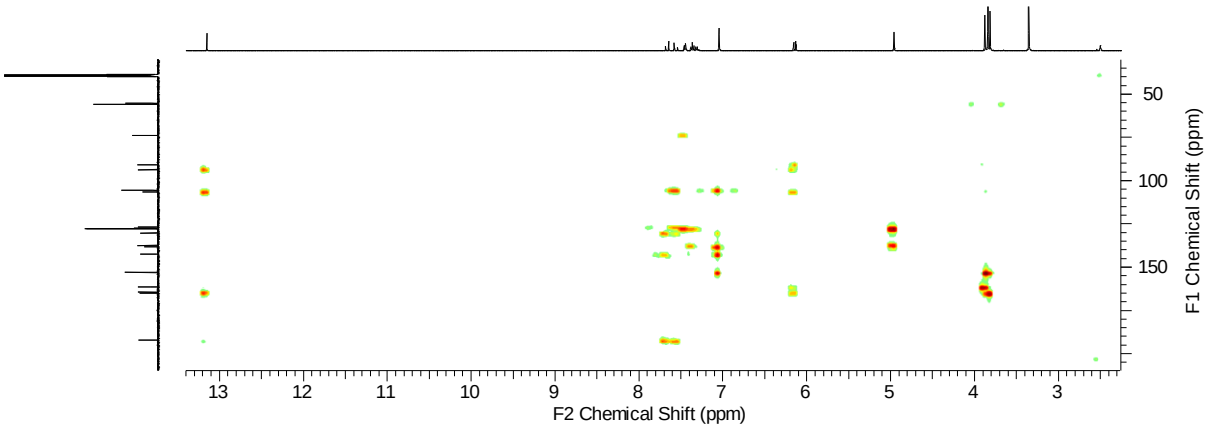
8.9.4 COSY spectrum



8.9.5 HSQC spectrum

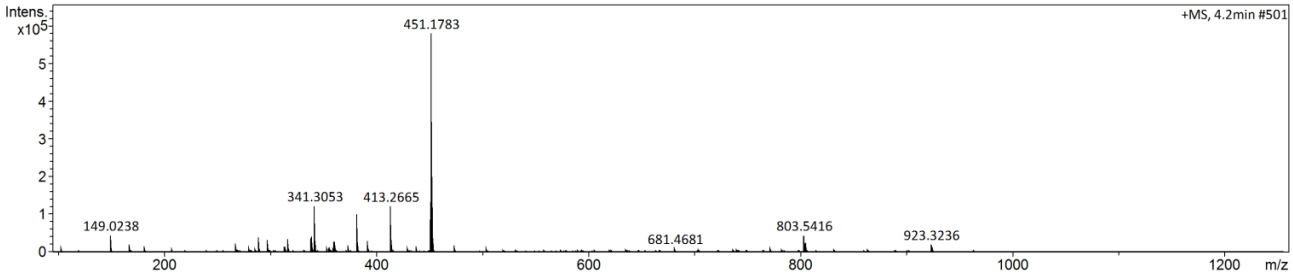


8.9.6 HMBC spectrum

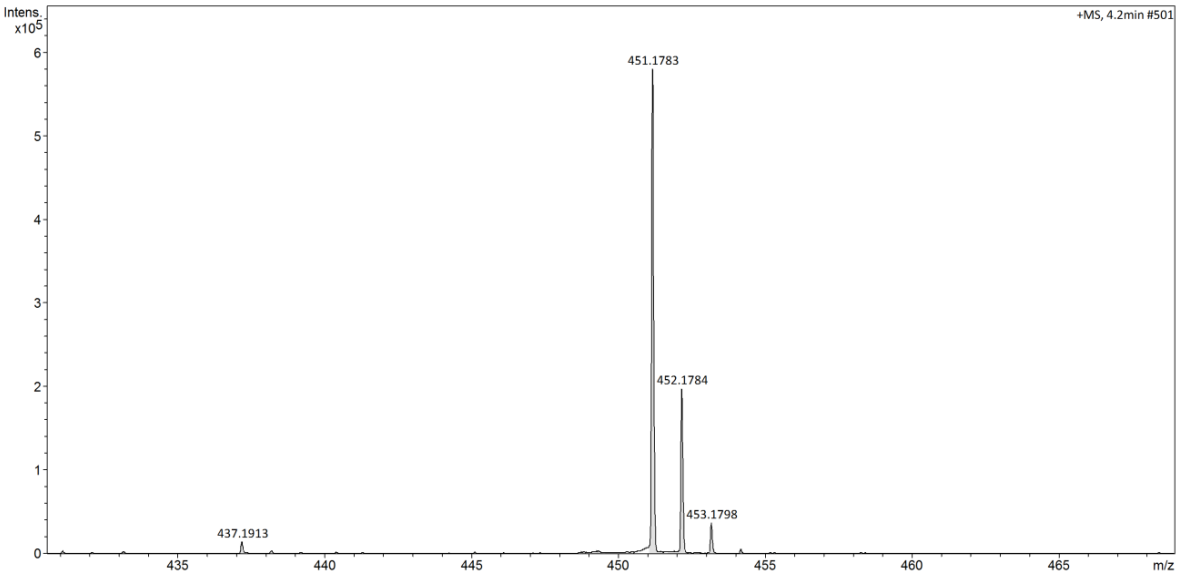


8.9.7 HRMS spectra

a)

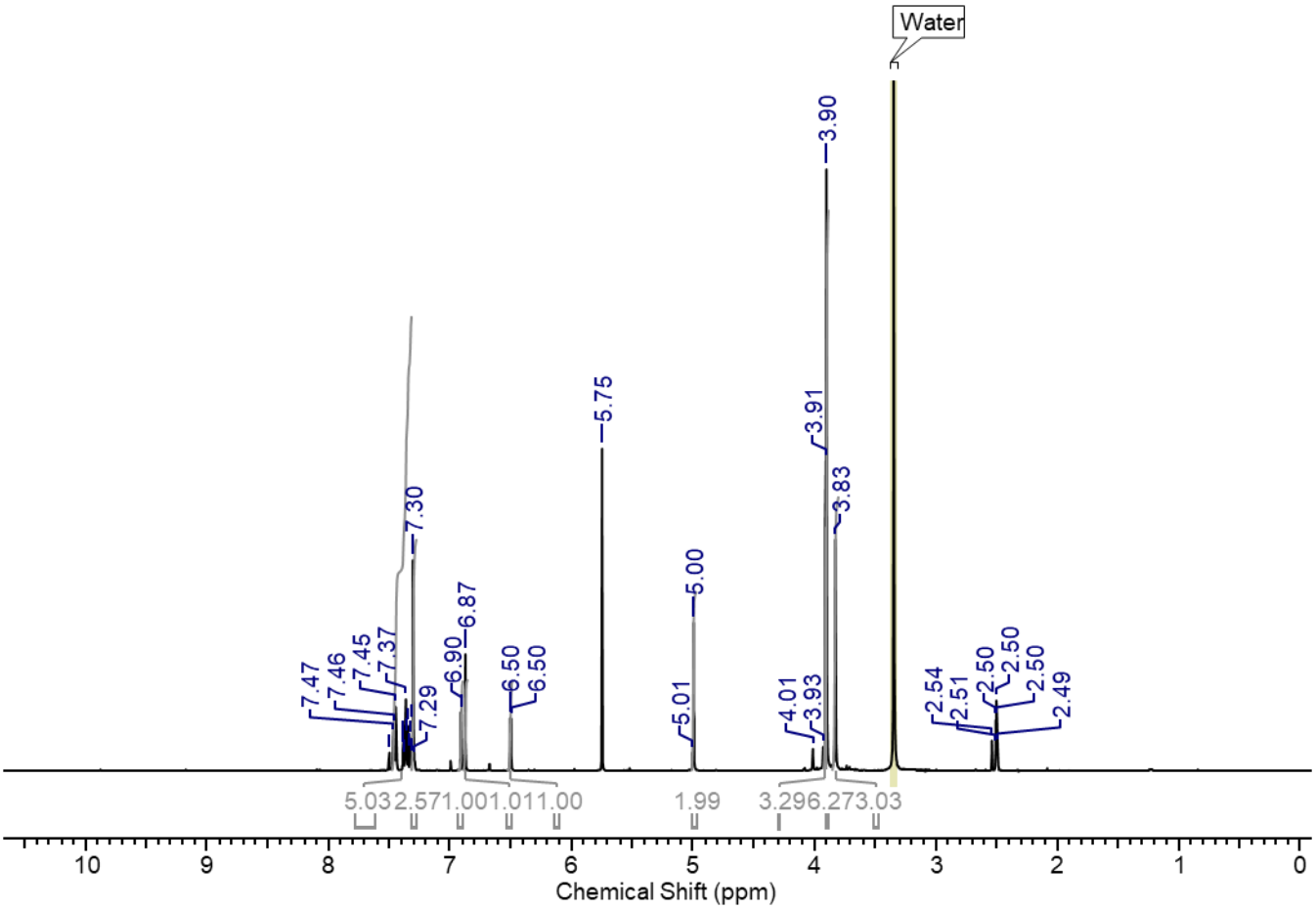

Intens.
x10^5
+MS, 4.2min #501
451.1783
341.3053
413.2665
149.0238
681.4681
803.5416
923.3236
m/z

b)

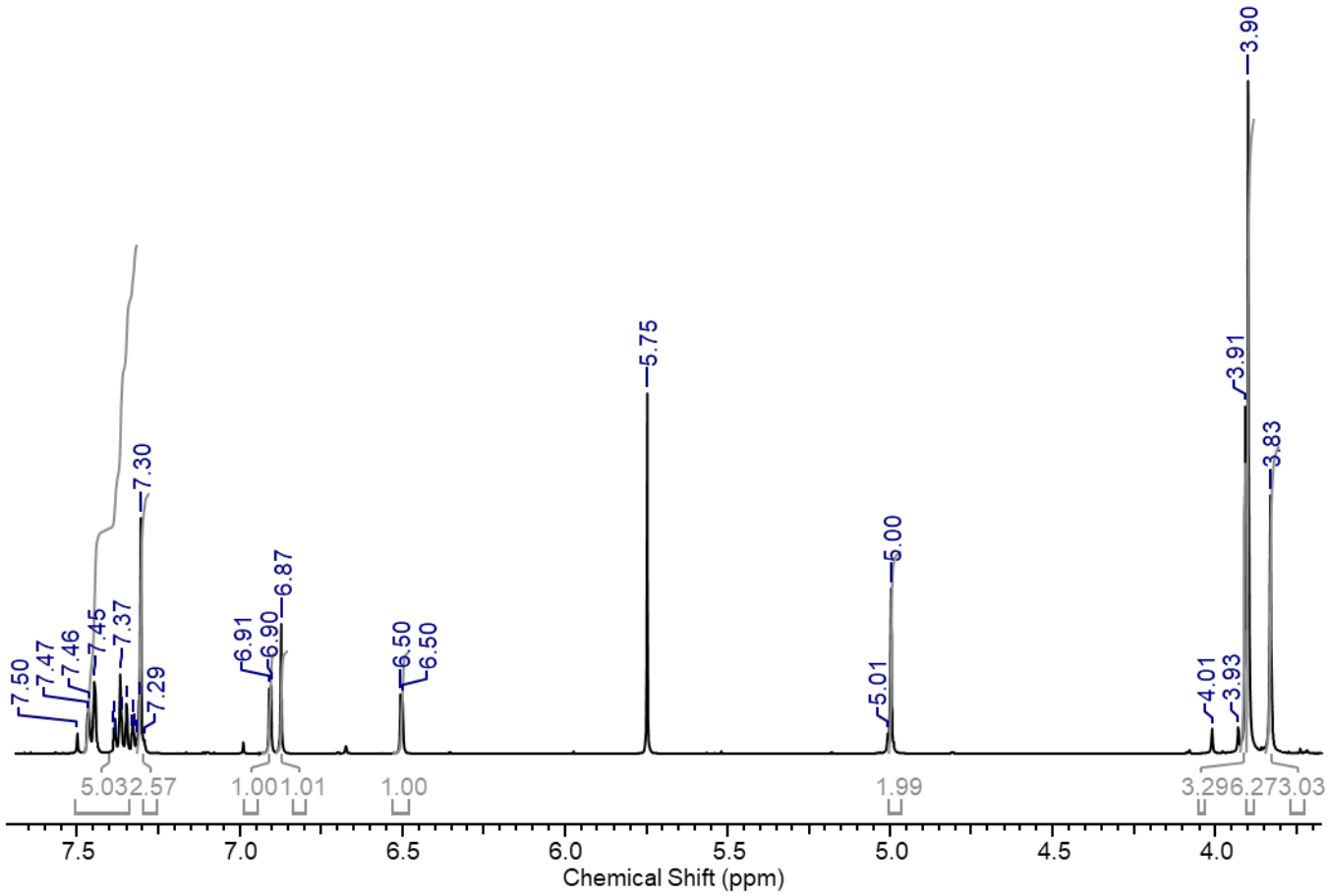

Intens.
x10^5
+MS, 4.2min #501
451.1783
452.1784
453.1798
437.1913
m/z

8.10 Characterisation data for the synthesis of 2-(4-(benzyloxy)-3,5-dimethoxyphenyl)-5,7-dimethoxy-4H-chromen-4-one, compound 2.9.

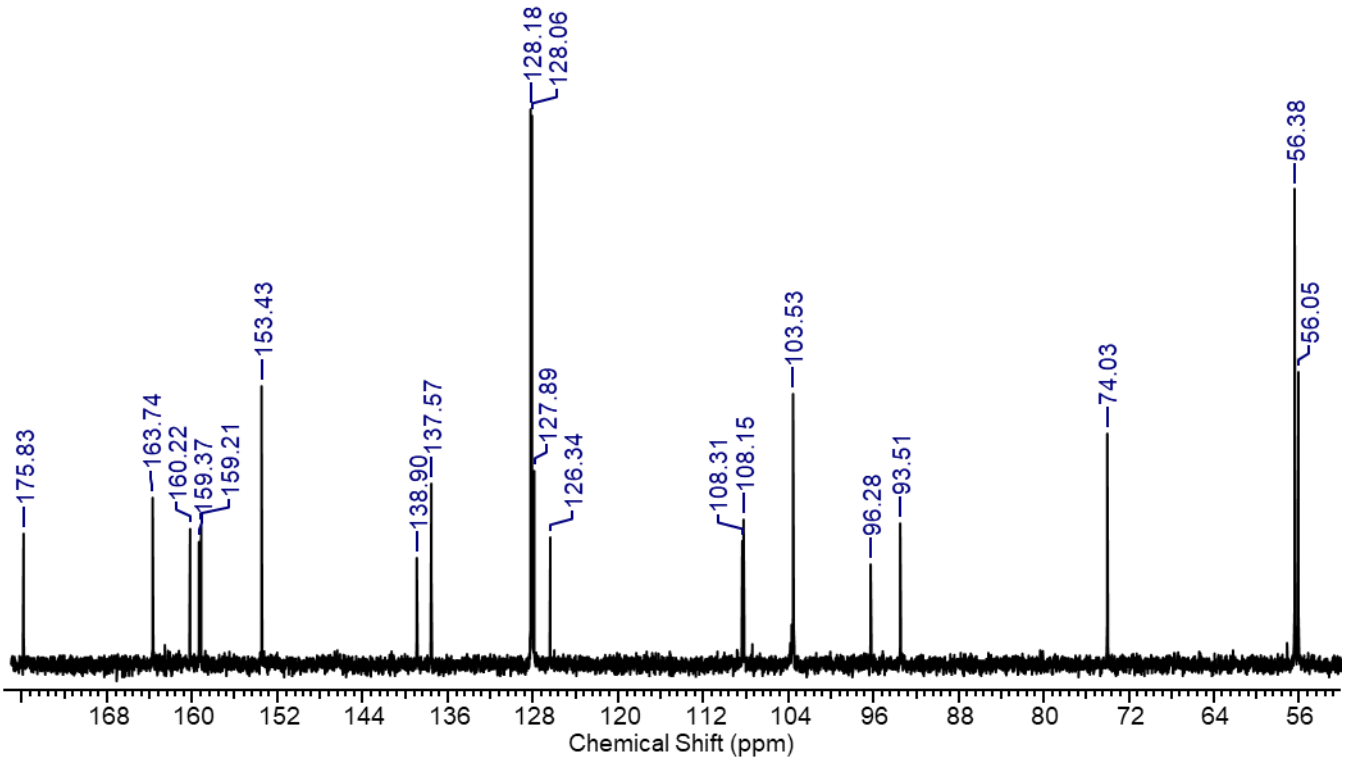
8.10.1 Extended ^{1}H NMR spectrum



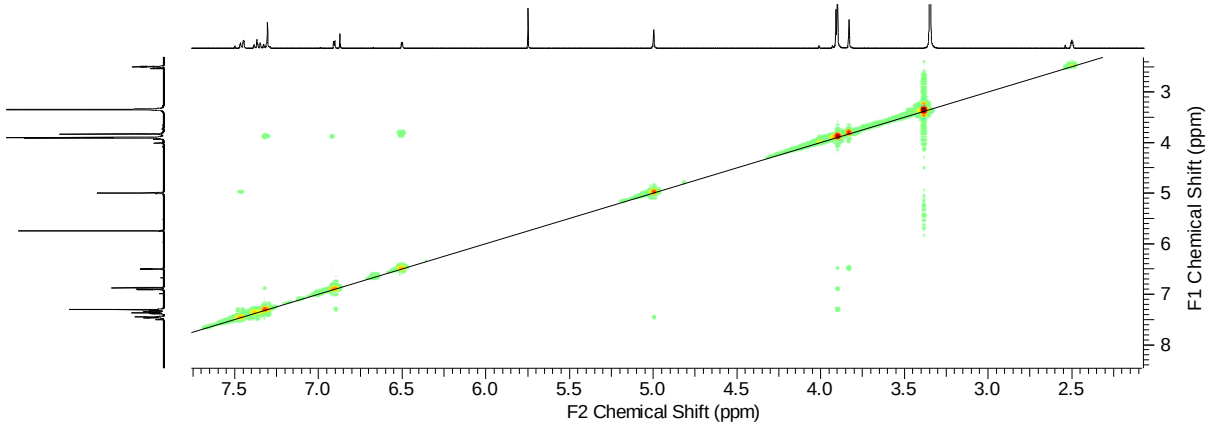
8.10.2 Magnified ^{1}H NMR spectrum



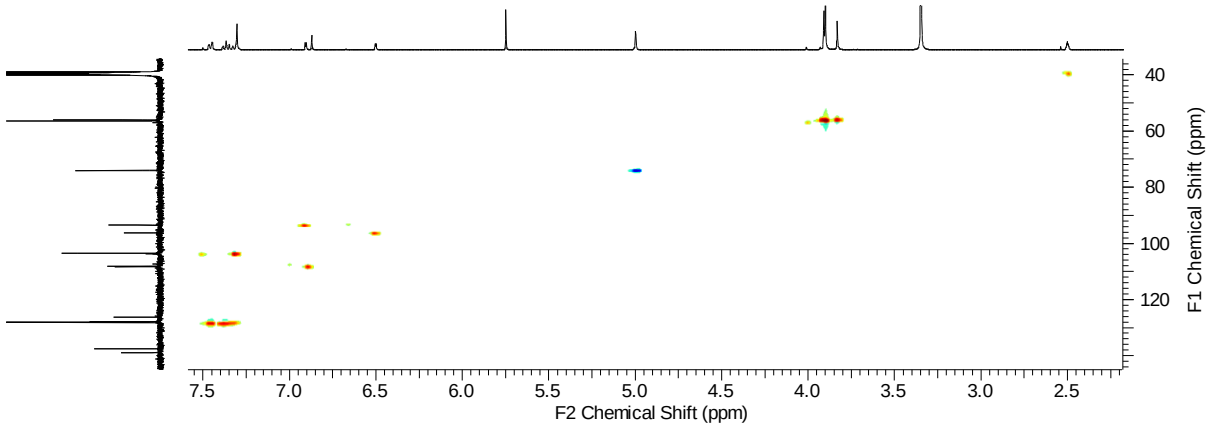
8.10.3 Magnified ^{13}C NMR spectrum



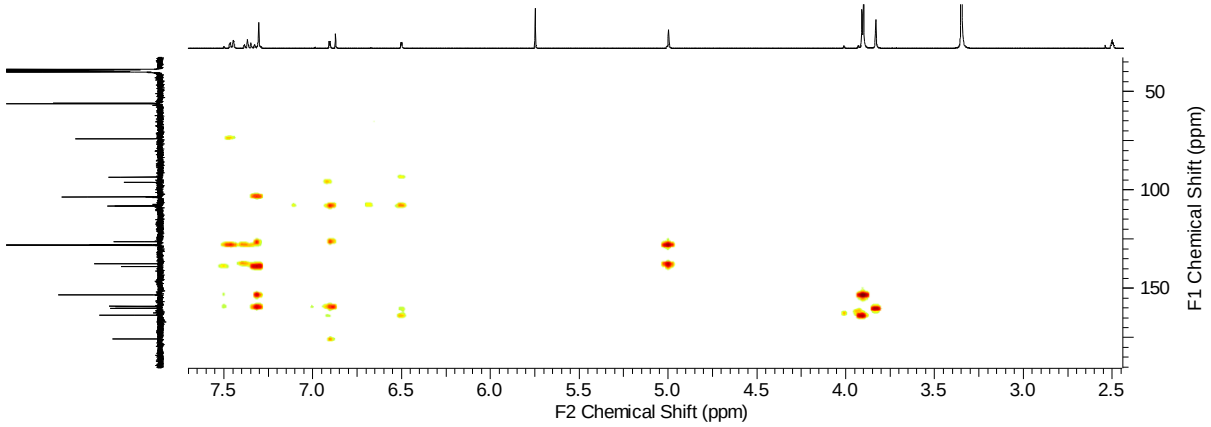
8.10.4 COSY spectrum



8.10.5 HSQC spectrum

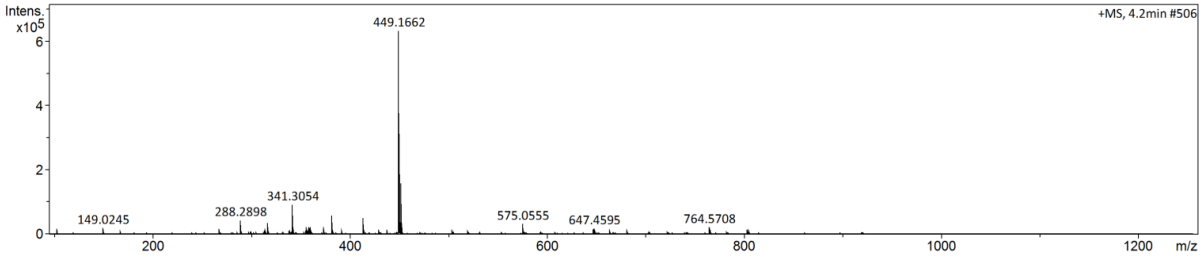


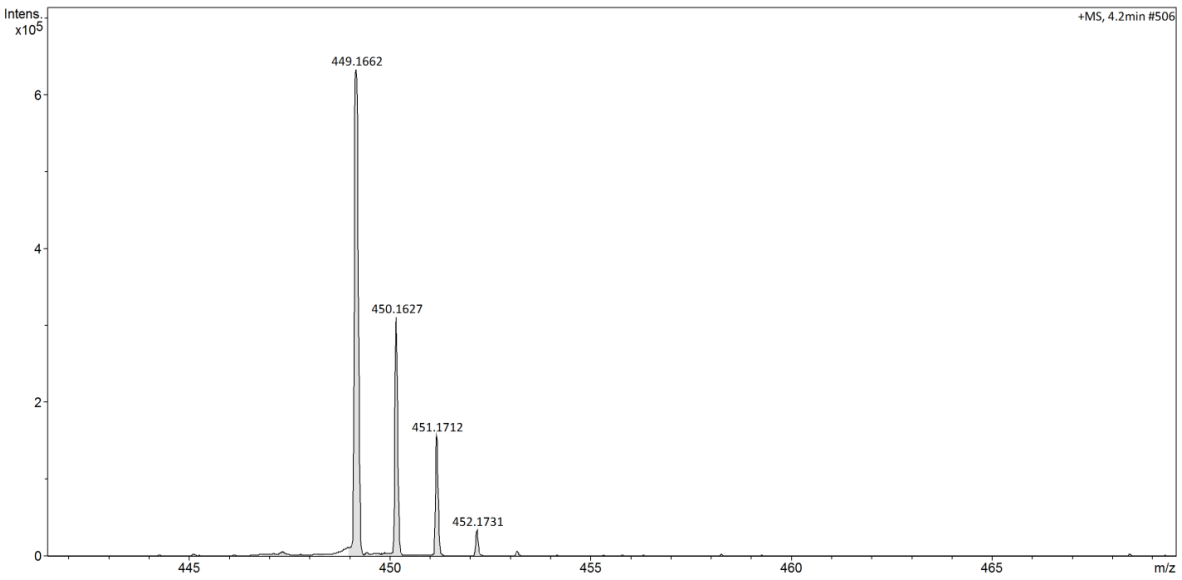
8.10.6 HMBC spectrum



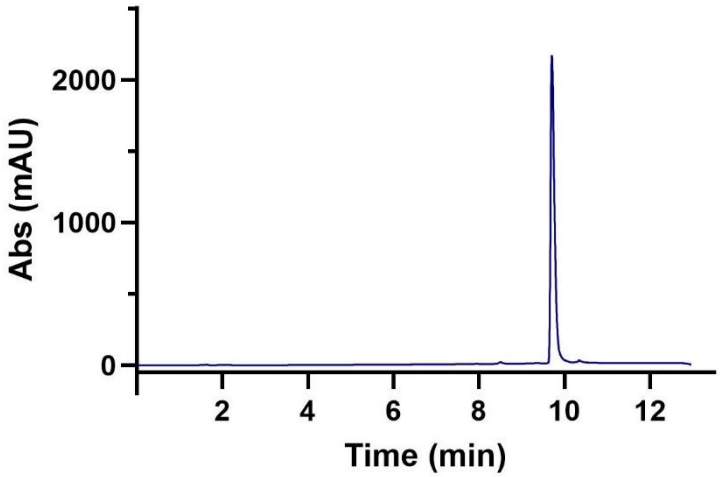
8.10.7 HRMS spectra

a)



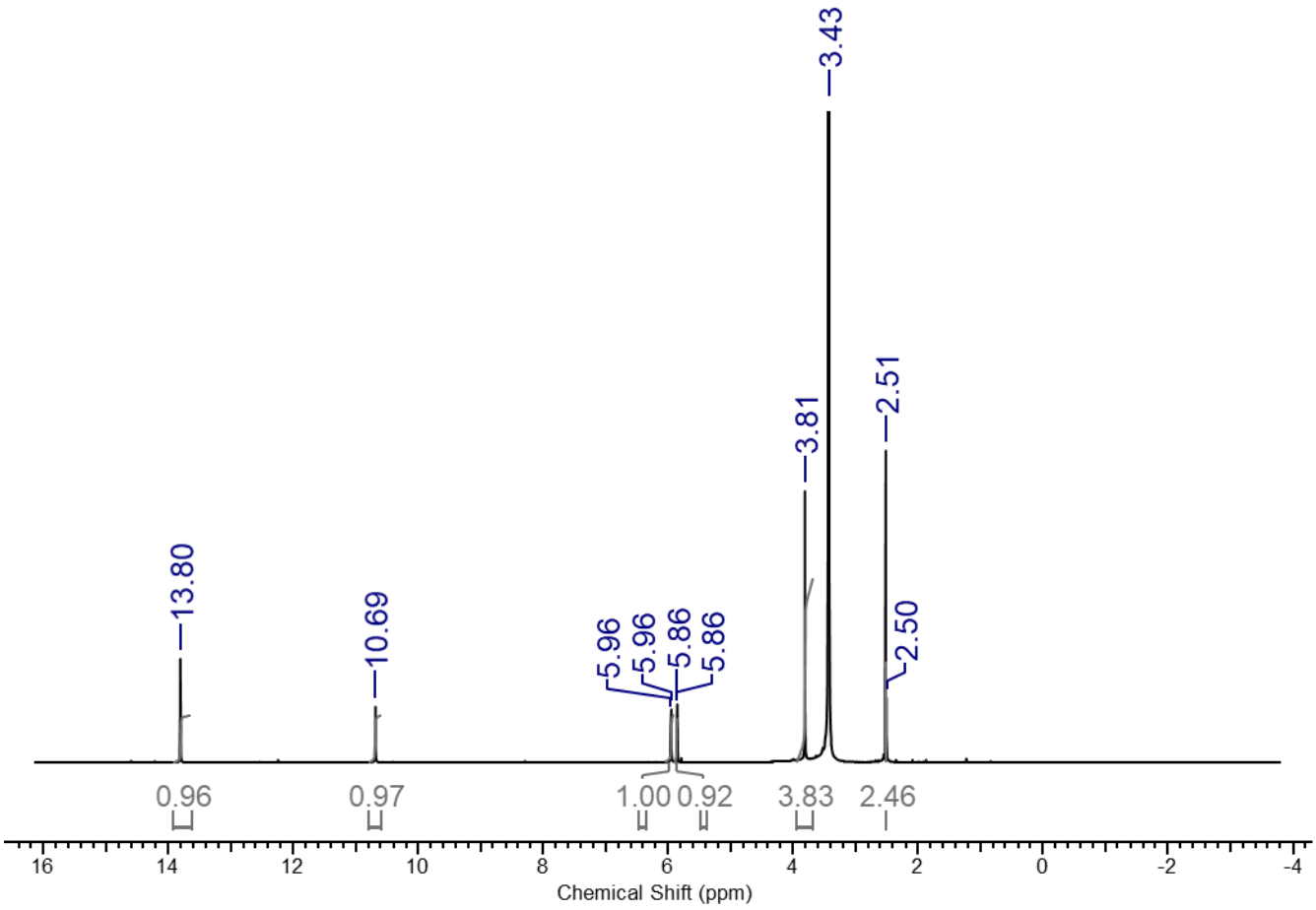


8.10.8 UV chromatogram of purified compound showing absorbance at 254 nm.

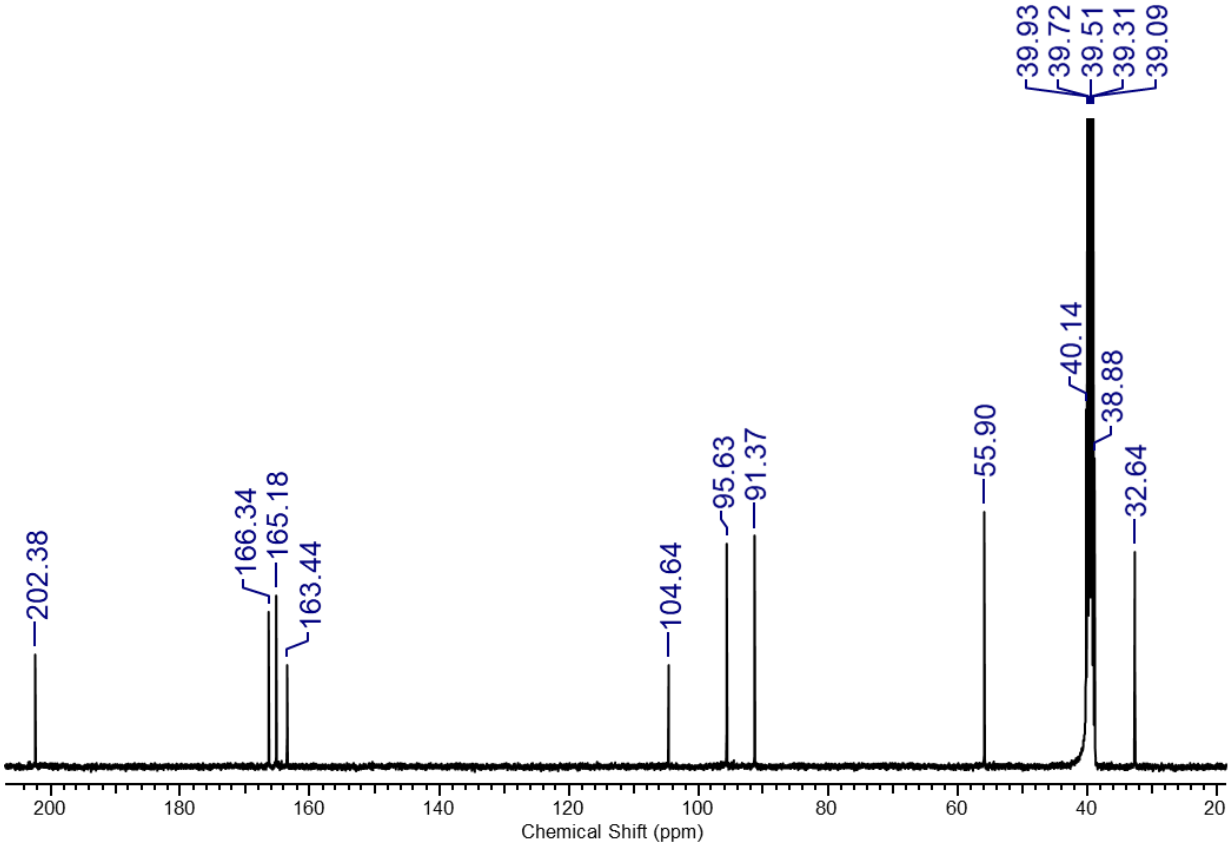


8.11 Characterisation data for the synthesis of 1-(2,4-dihydroxy-6-methoxyphenyl)ethan-1-one, compound 2.10.

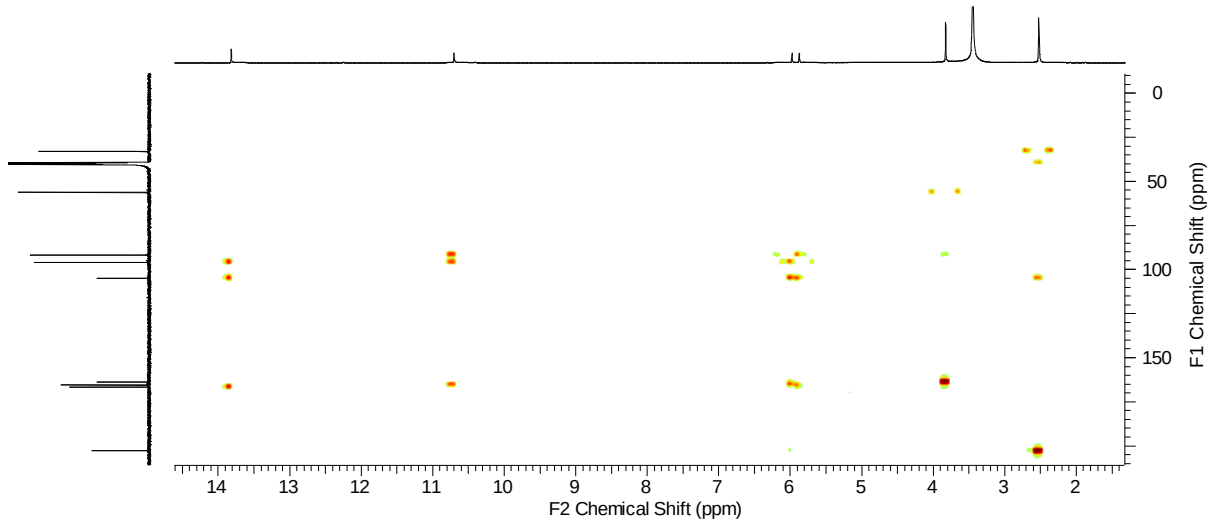
8.11.1 Extended ^{1}H NMR spectrum



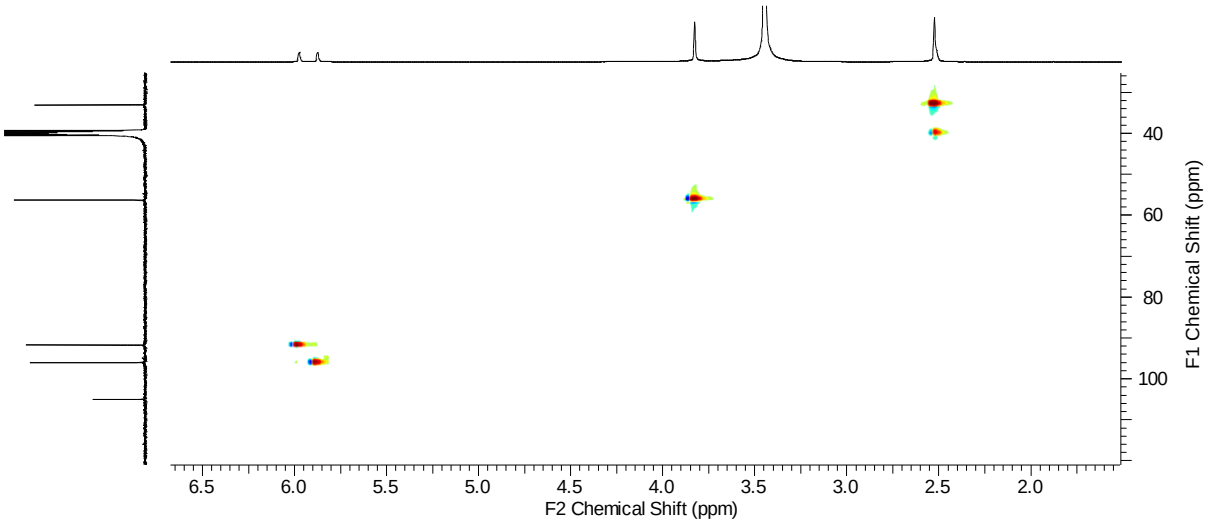
8.11.2 Magnified ^{13}C NMR spectrum



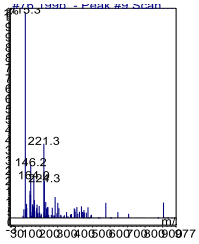
8.11.3 HMBC spectrum



8.11.4 HQSC spectrum

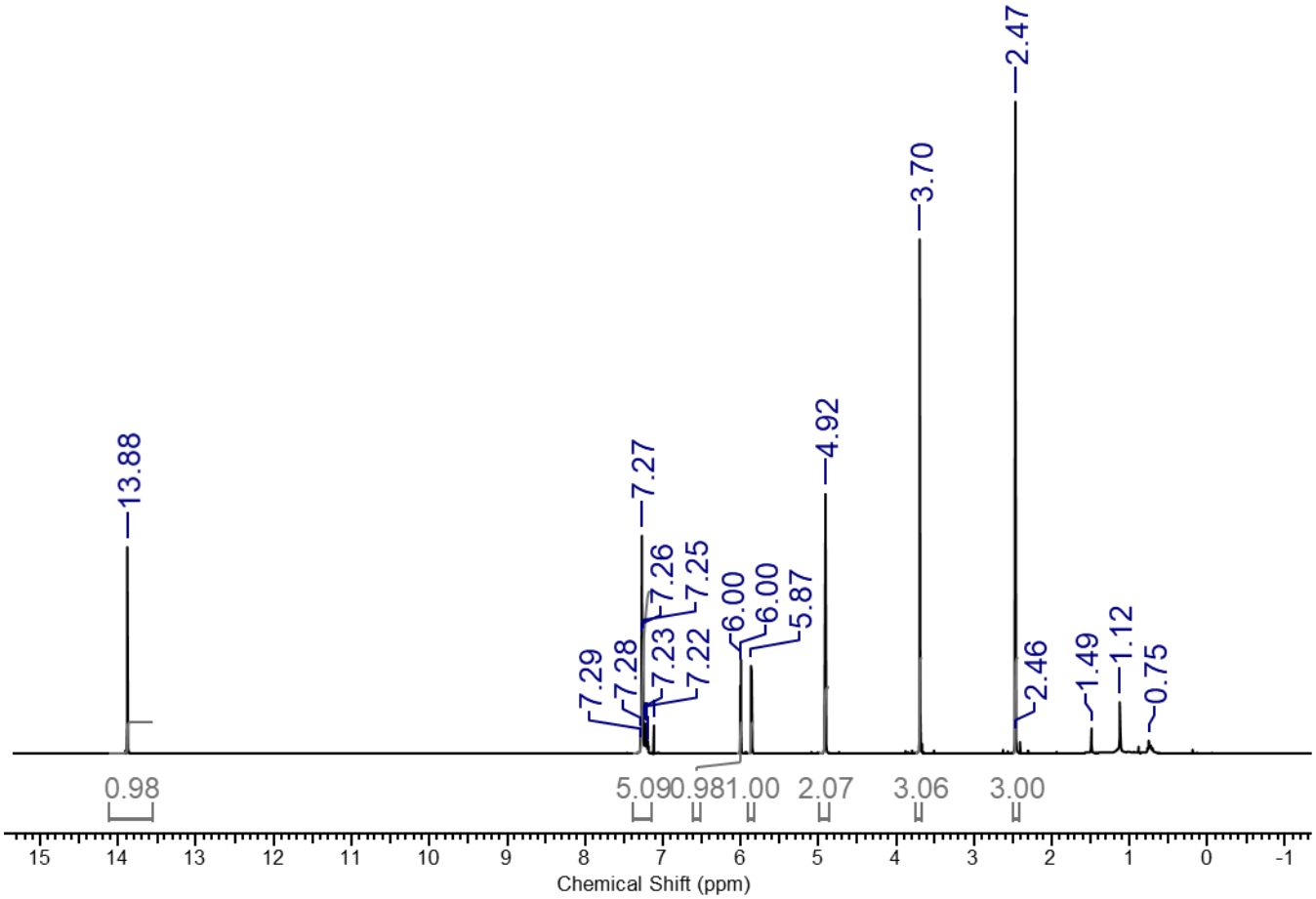


8.11.5 MS spectrum

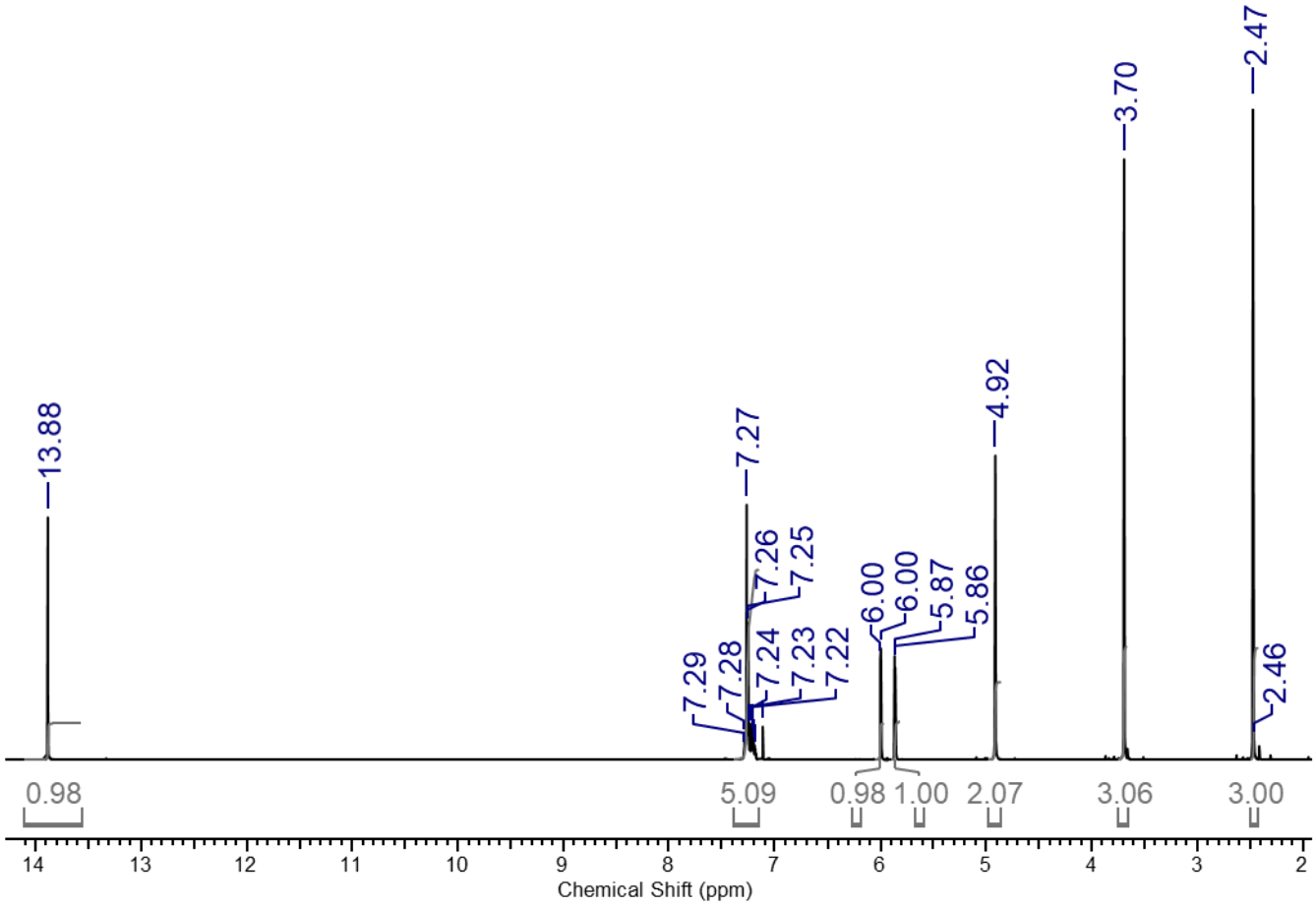


8.12 Characterisation data for the synthesis of 1-(4-(benzyloxy)-2-hydroxy-6-methoxyphenyl)ethan-1-one, compound 2.11.

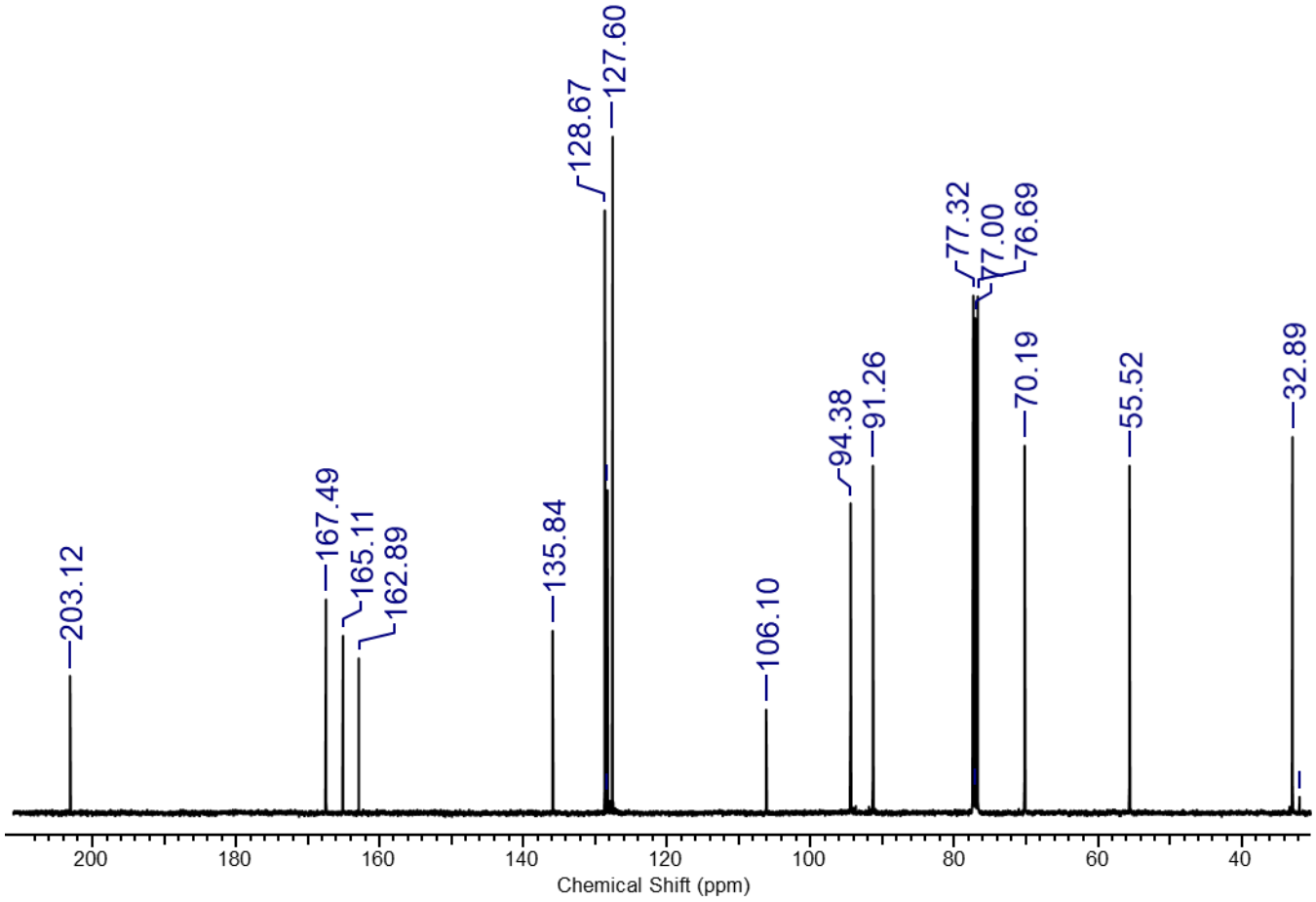
8.12.1 Extended ^{1}H NMR spectrum



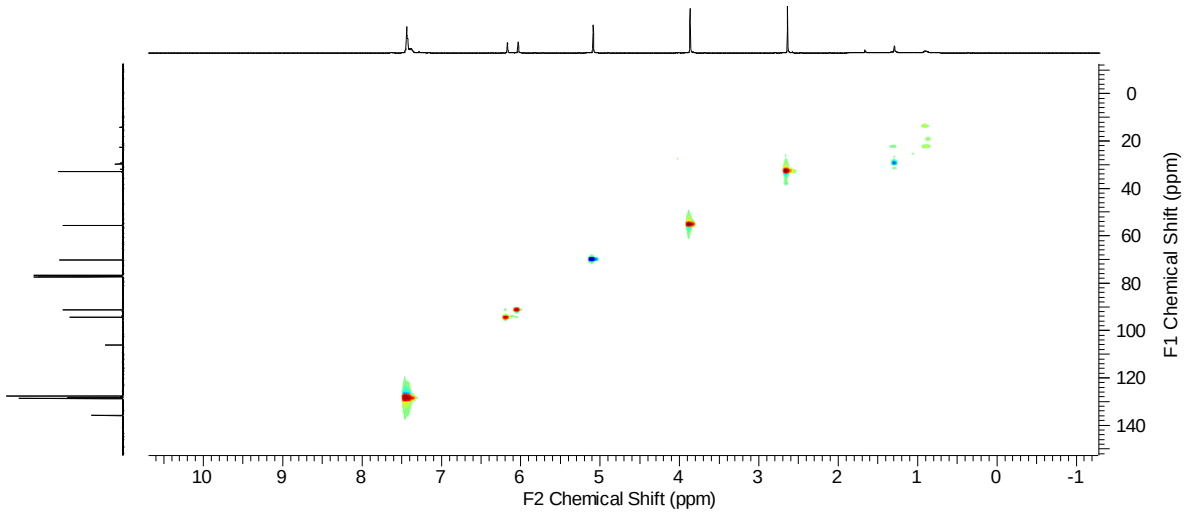
8.12.2 Magnified ^{1}H NMR spectrum



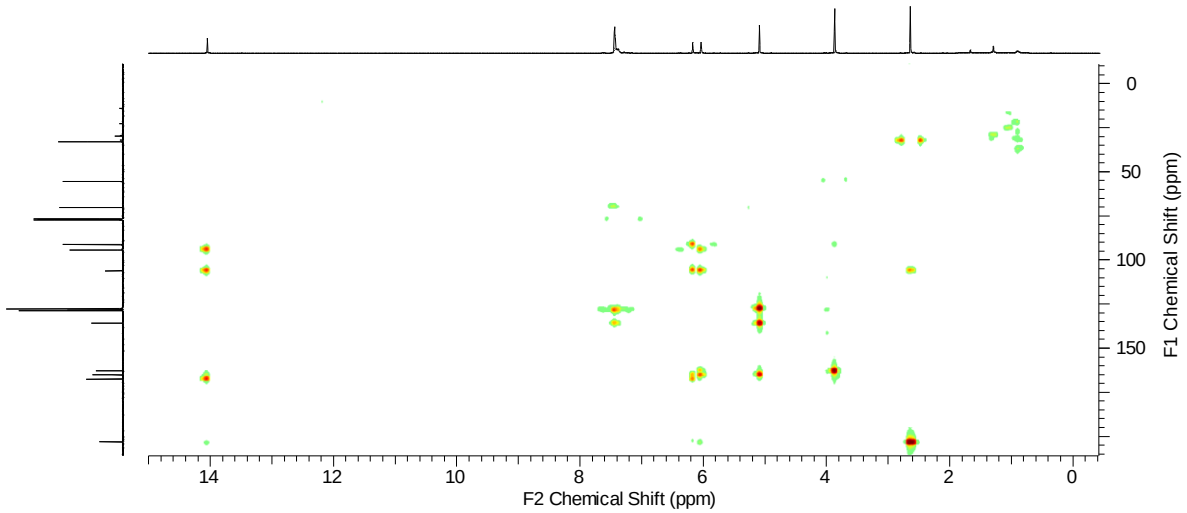
8.12.3 Magnified ^{13}C NMR spectrum



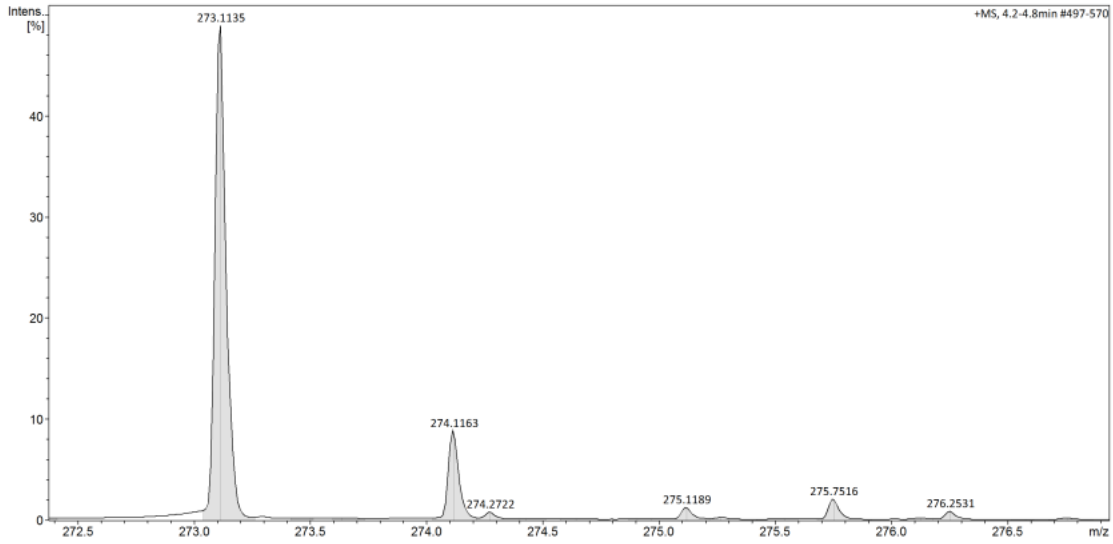
8.12.4 HQSC spectrum



8.12.5 HMBC spectrum

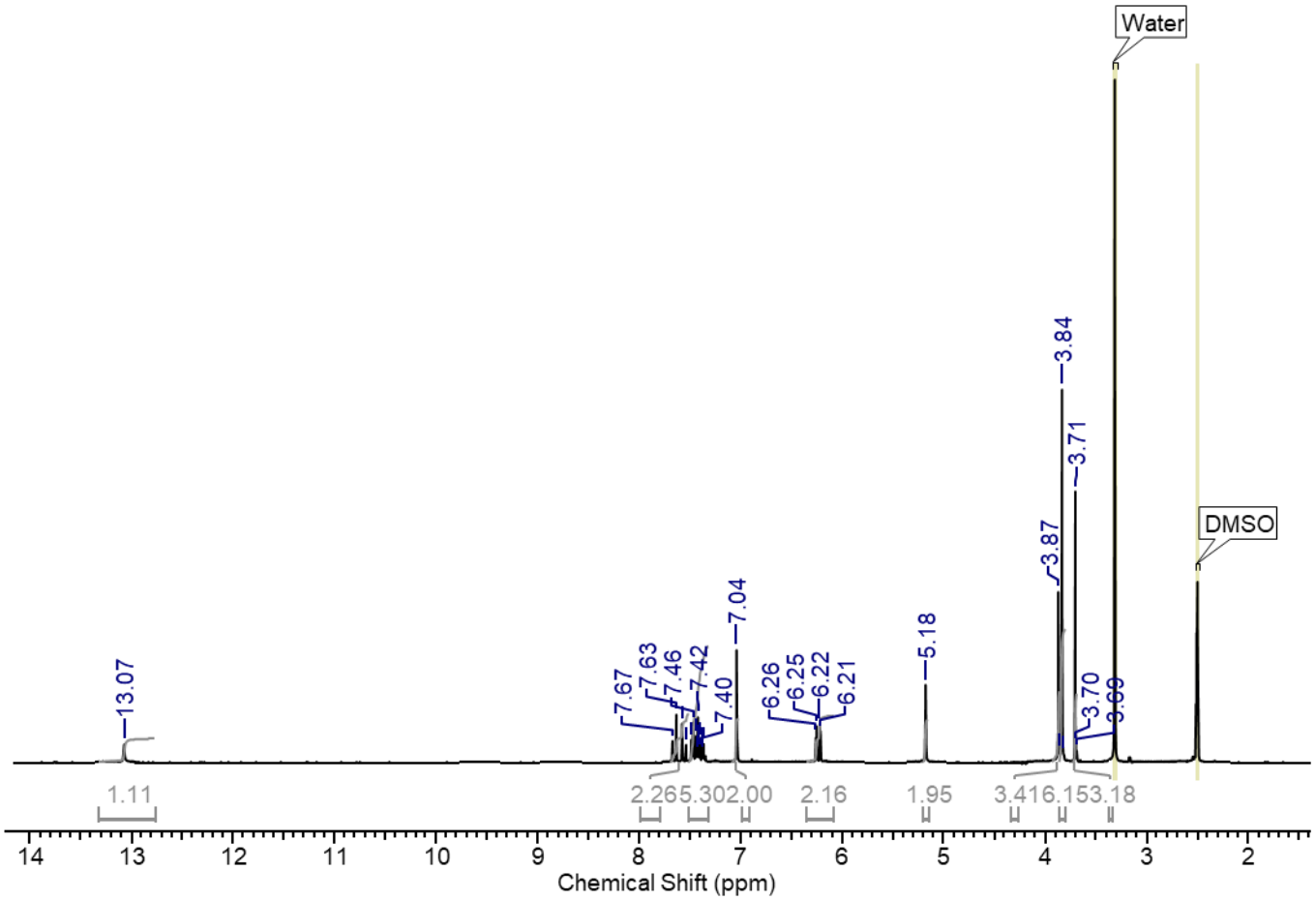


8.12.6 HRMS spectrum

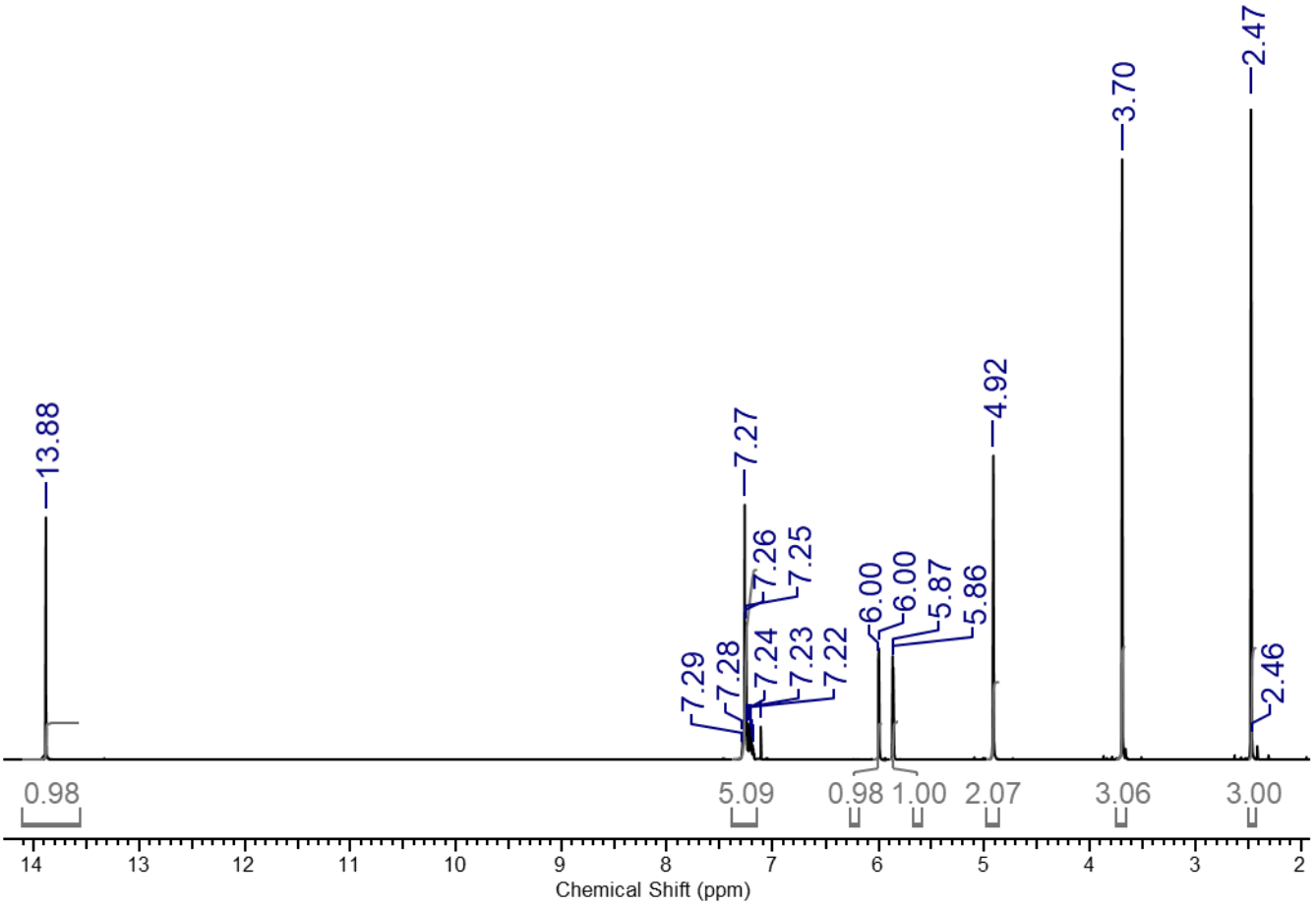


8.13 Characterisation data for the synthesis of (E)-1-(4-(benzyloxy)-2-hydroxy-6-methoxyphenyl)-3-(3,4,5-trimethoxyphenyl)prop-2-en-1-one, compound 2.12.

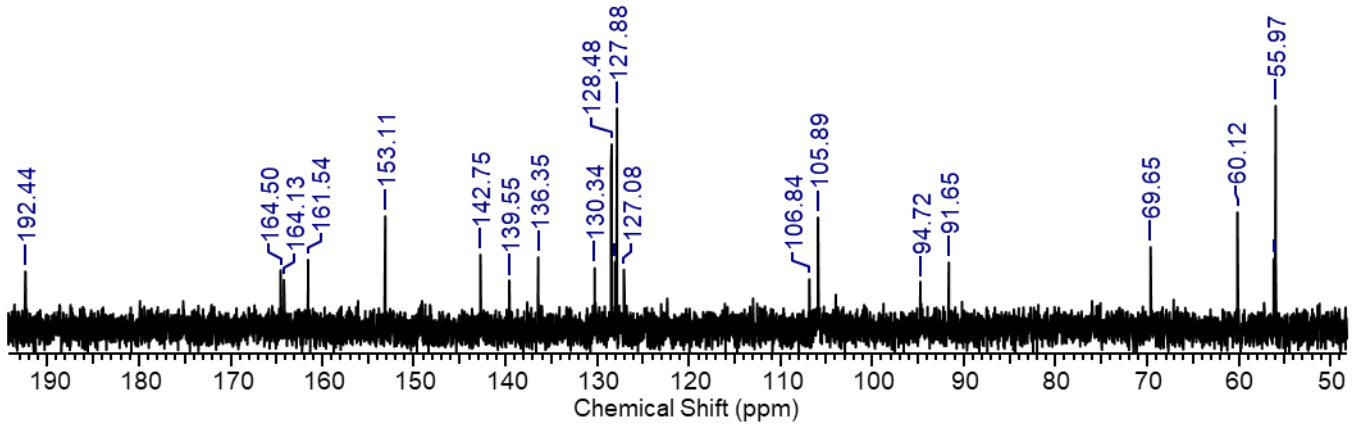
8.13.1 Extended ^{1}H NMR spectrum



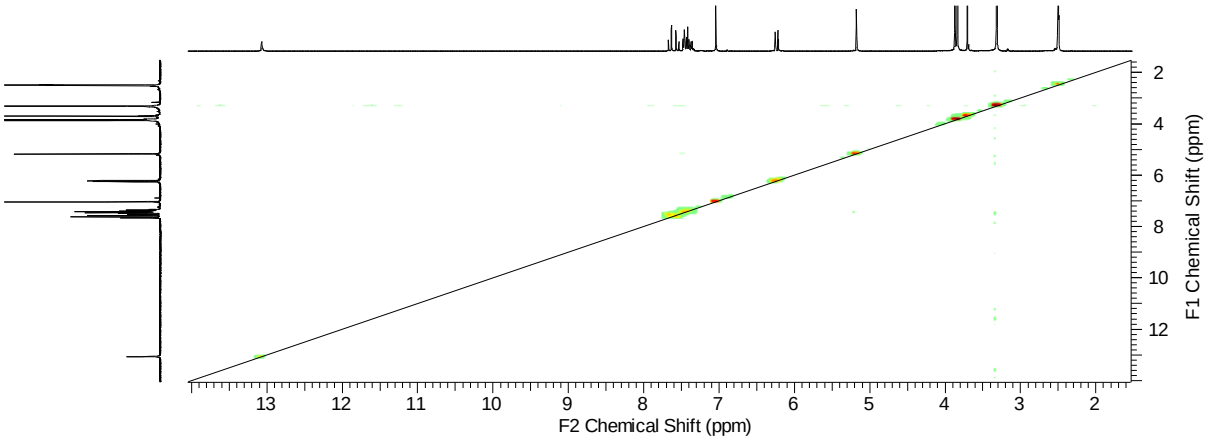
8.13.2 Magnified ^{1}H NMR spectrum



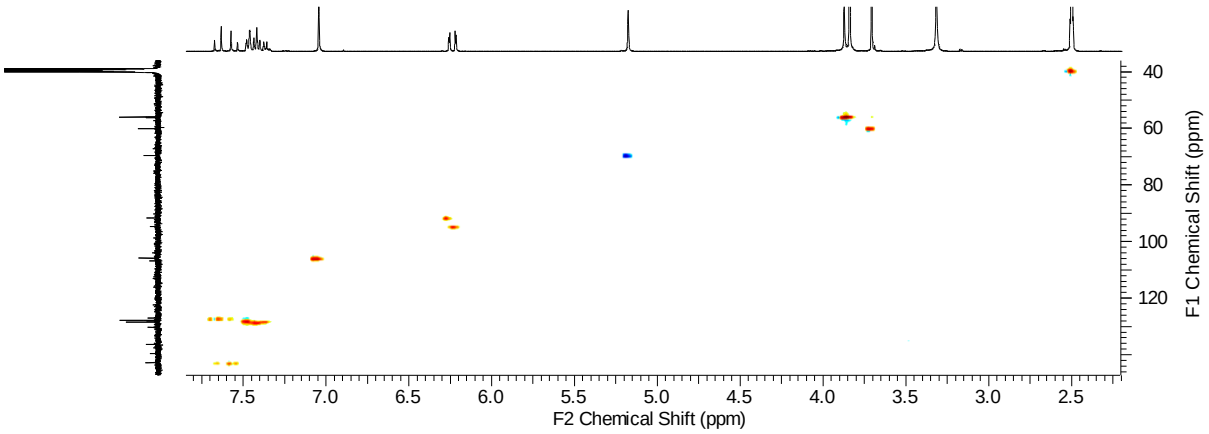
8.13.3 Magnified ^{13}C NMR spectrum



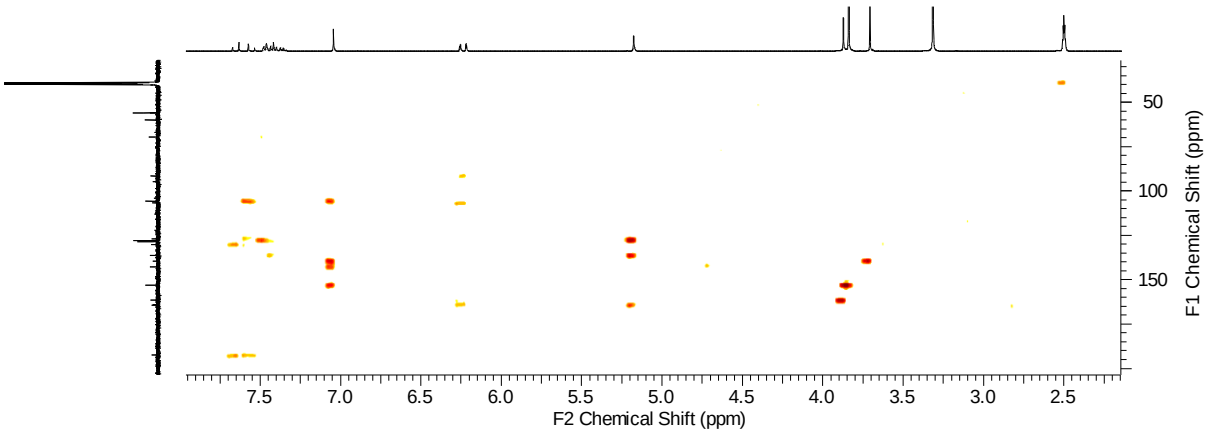
8.13.4 COSY spectrum



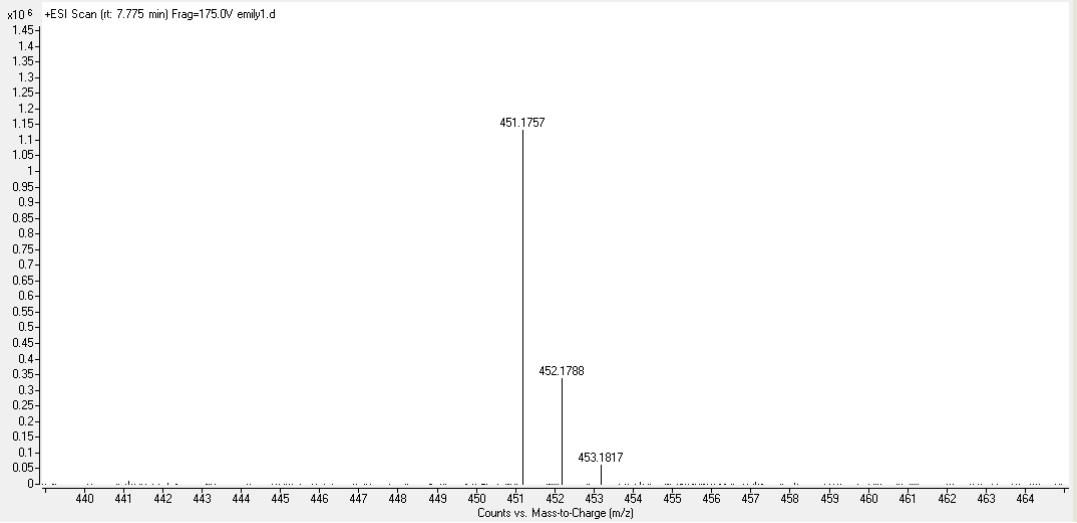
8.13.5 HQSC spectrum



8.13.6 HMBC spectrum

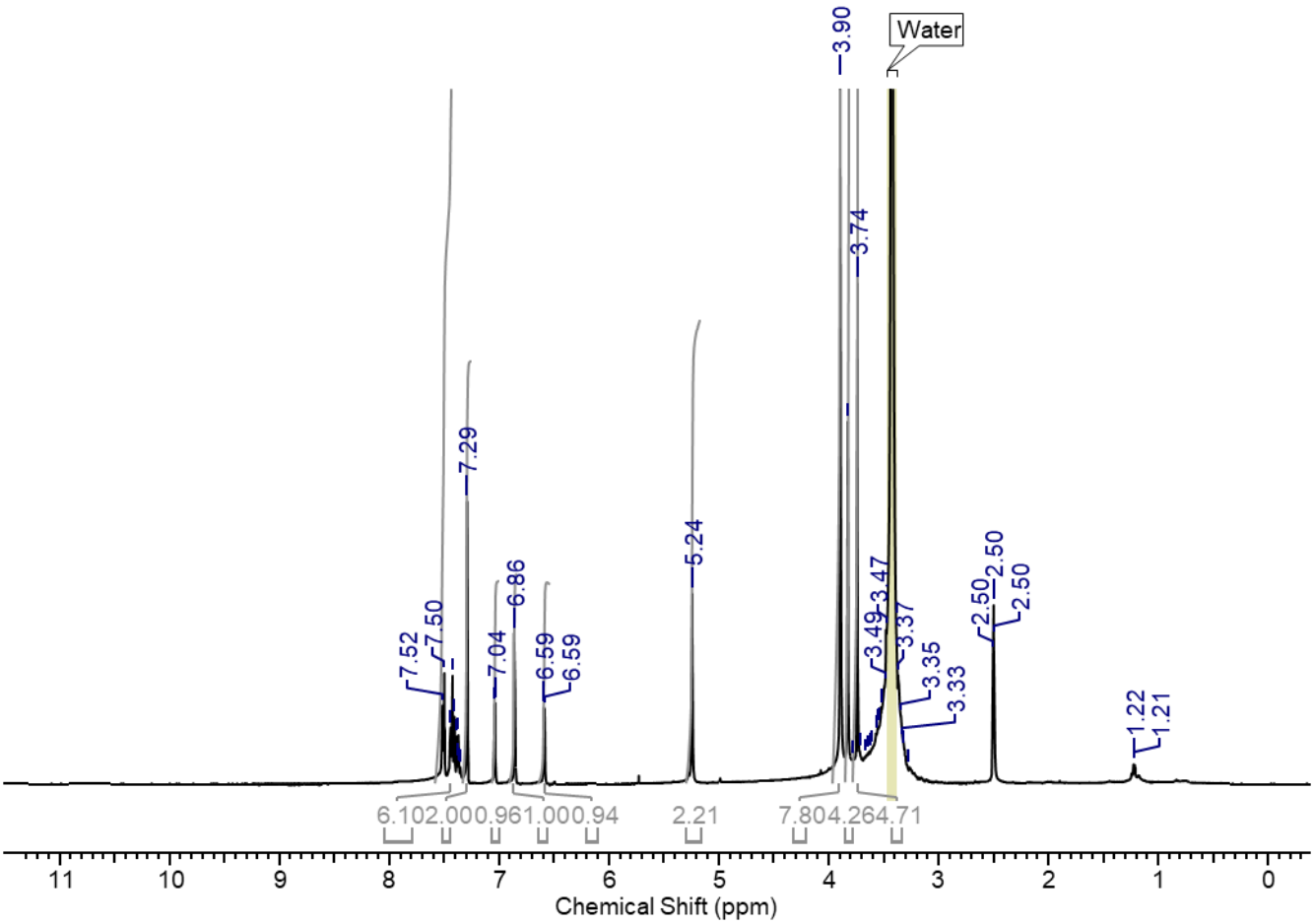


8.13.7 HRMS spectrum

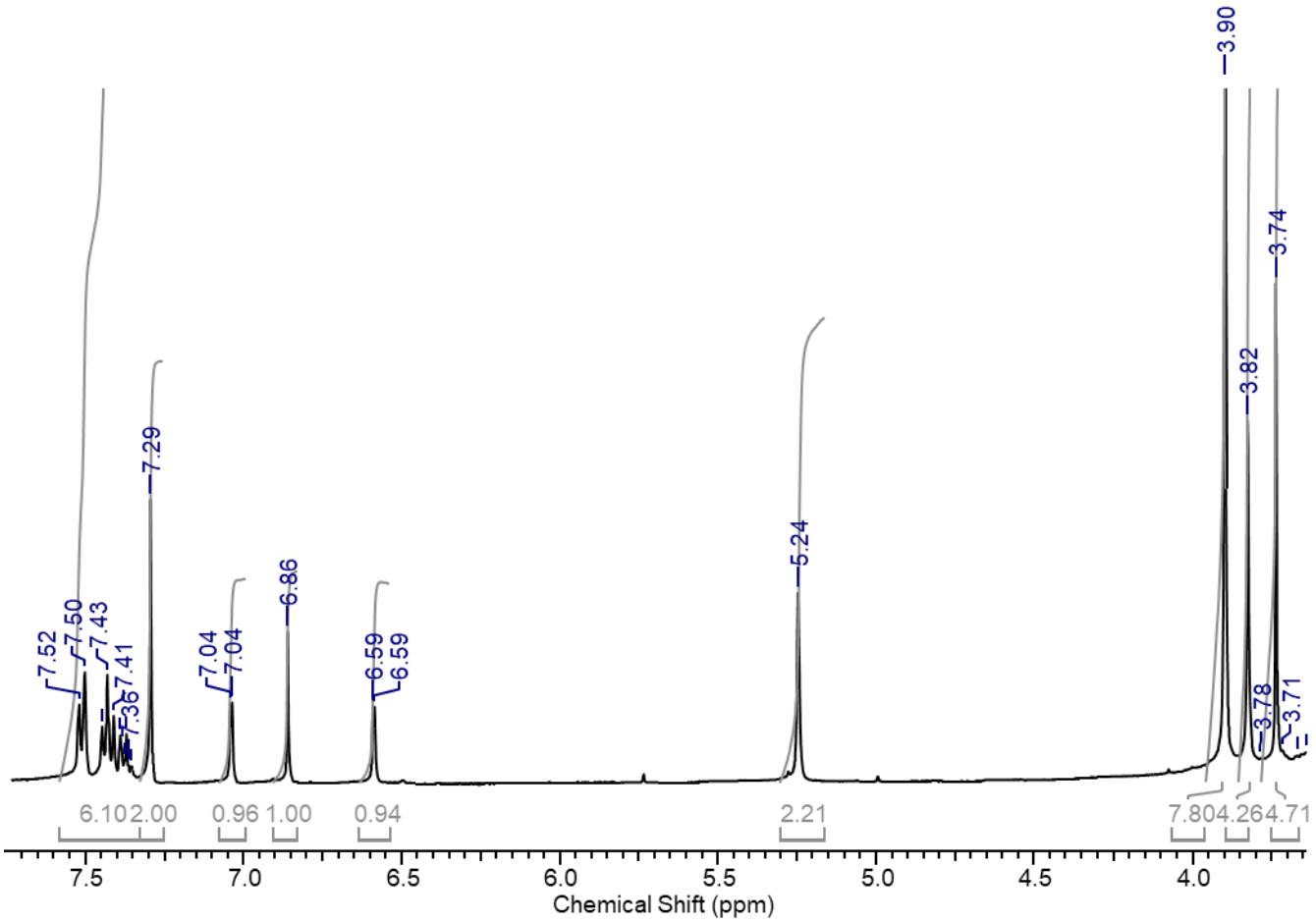


8.14 Characterisation data for the synthesis of 7-(benzyloxy)-5-methoxy-2-(3,4,5-trimethoxyphenyl)-4H-chromen-4-one, compound 2.13.

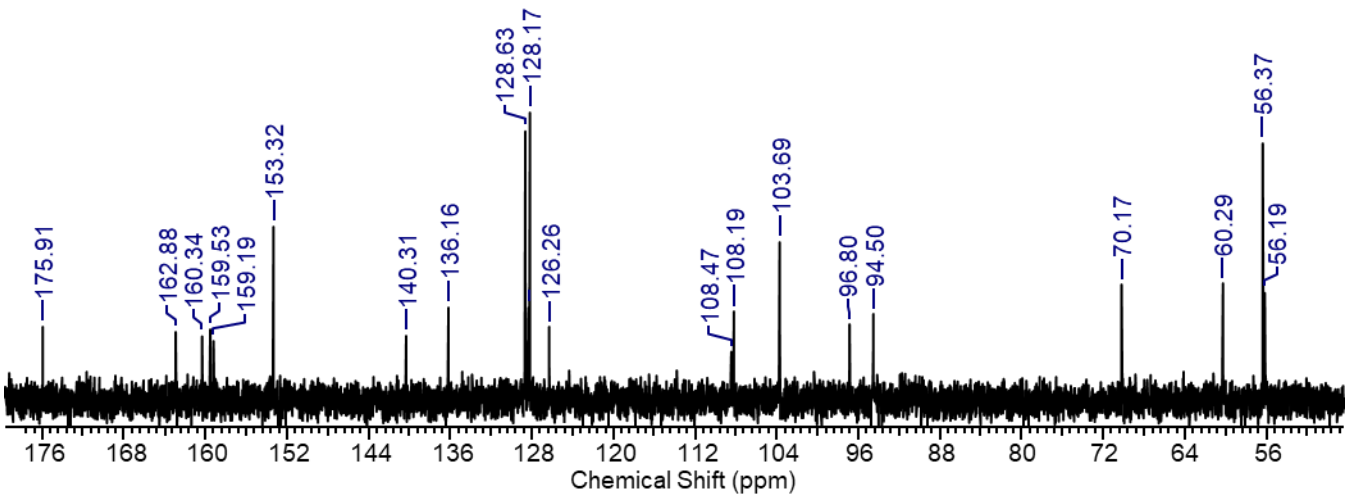
8.14.1 Extended ^{1}H NMR spectrum



8.14.2 Magnified ^{1}H NMR spectrum

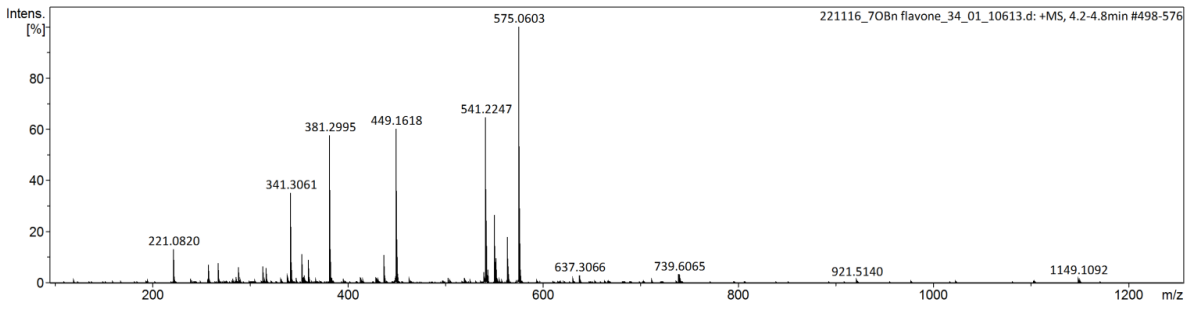


8.14.3 Magnified ^{13}C NMR spectrum

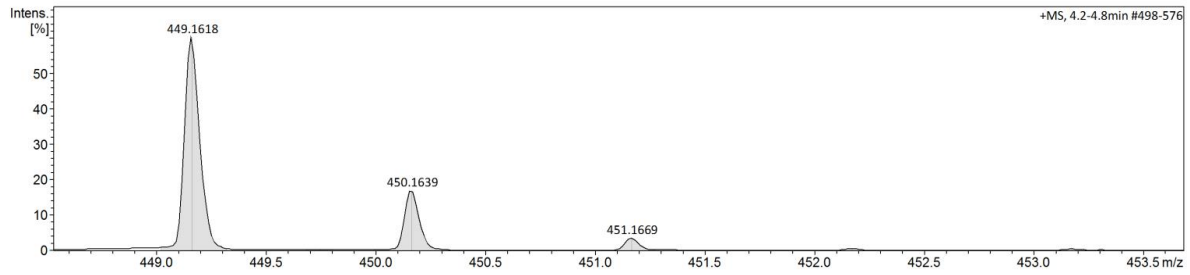


8.14.4 HRMS spectrum

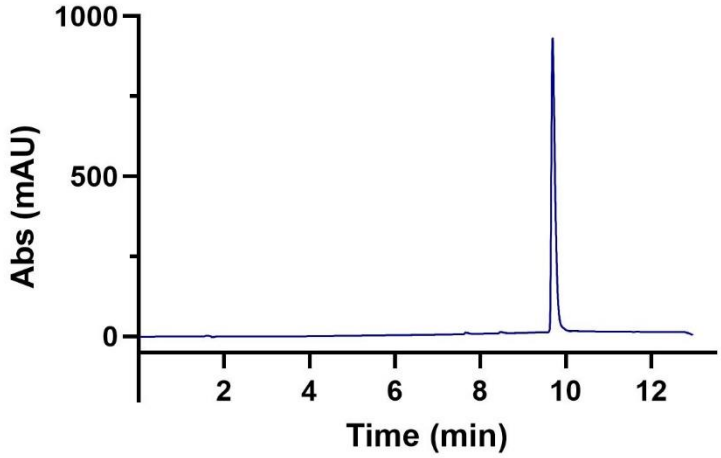
a)


Intens.
[%]
221116_7OBn flavone_34_01_10613.d: +MS, 4.2-4.8min #498-576
575.0603
541.2247
449.1618
381.2995
341.3061
221.0820
637.3066
739.6065
921.5140
1149.1092
m/z

b)

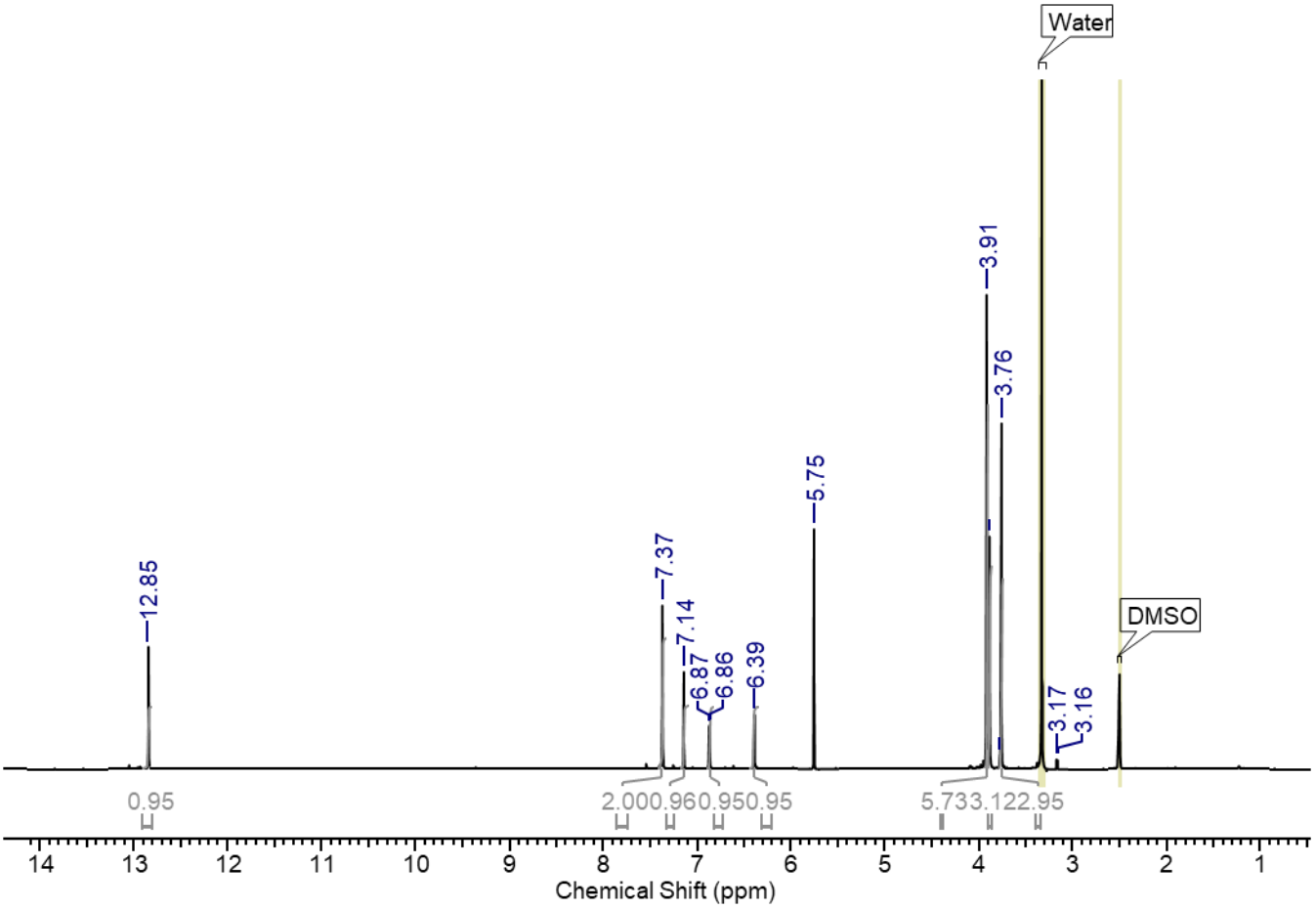

Intens.
[%]
+MS, 4.2-4.8min #498-576
449.1618
450.1639
451.1669
m/z

8.14.5 UV chromatogram of purified compound showing absorbance at 254 nm.

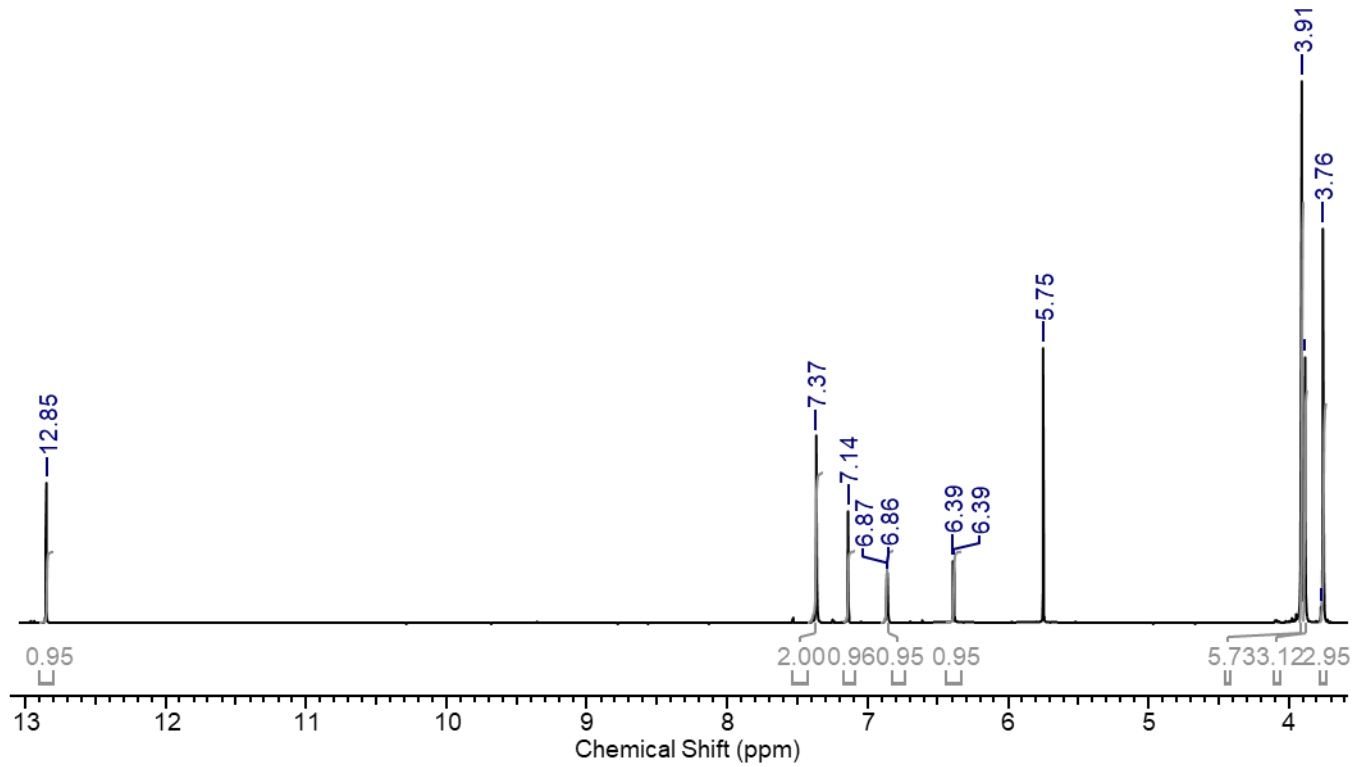


8.15 Characterisation data for the synthesis of 5-hydroxy-7-methoxy-2-(3,4,5-trimethoxyphenyl)-4H-chromen-4-one, compound 2.14.

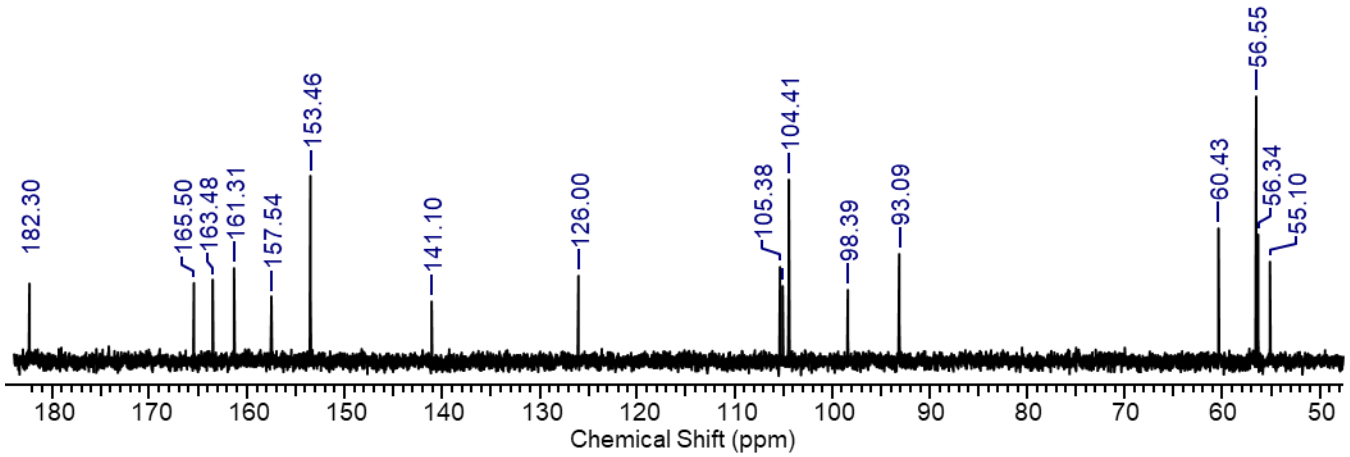
8.15.1 Extended ^{1}H NMR spectrum



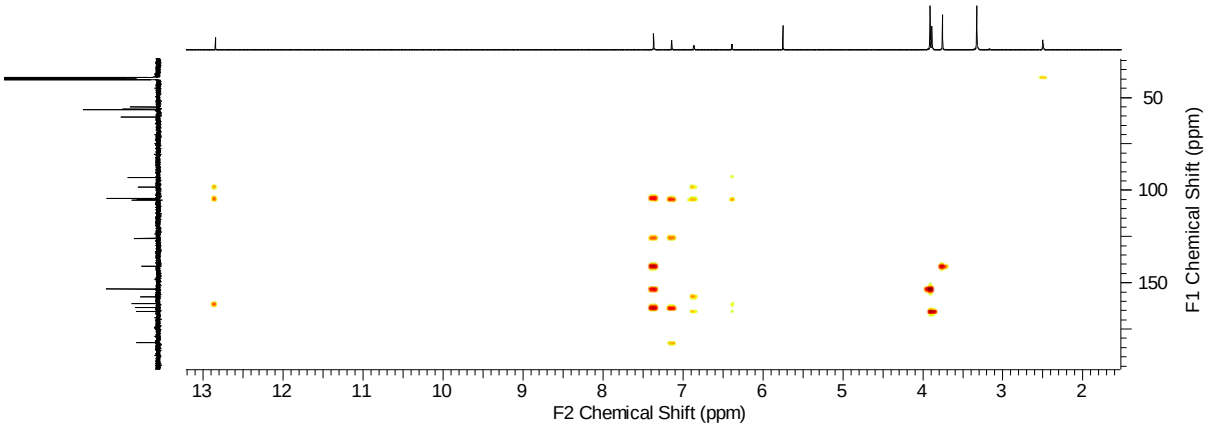
8.15.2 Magnified ^{1}H NMR spectrum



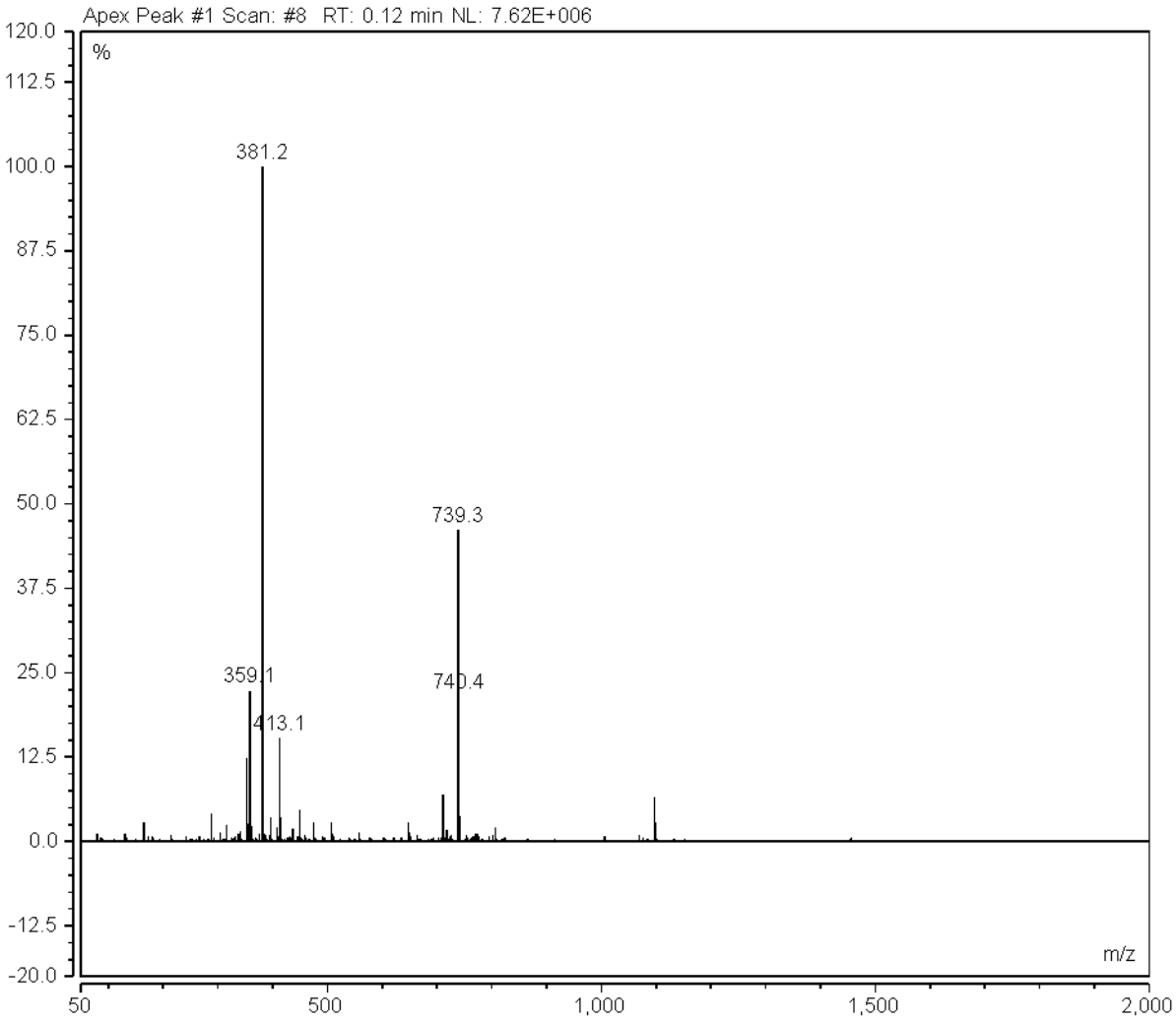
8.15.3 Magnified ^{13}C NMR spectrum



8.15.4 HMBC spectrum

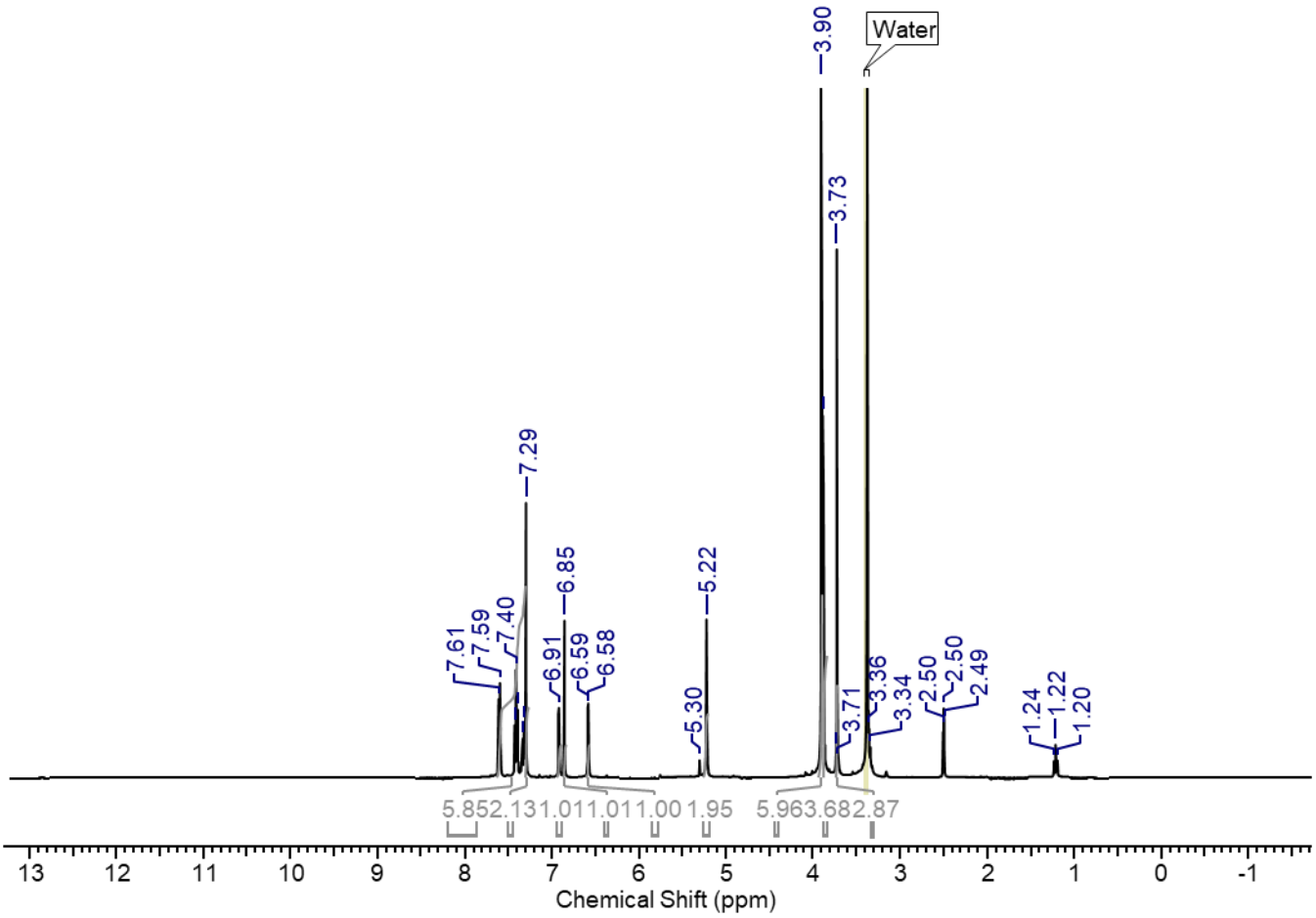


8.15.5 ESI-MS spectrum

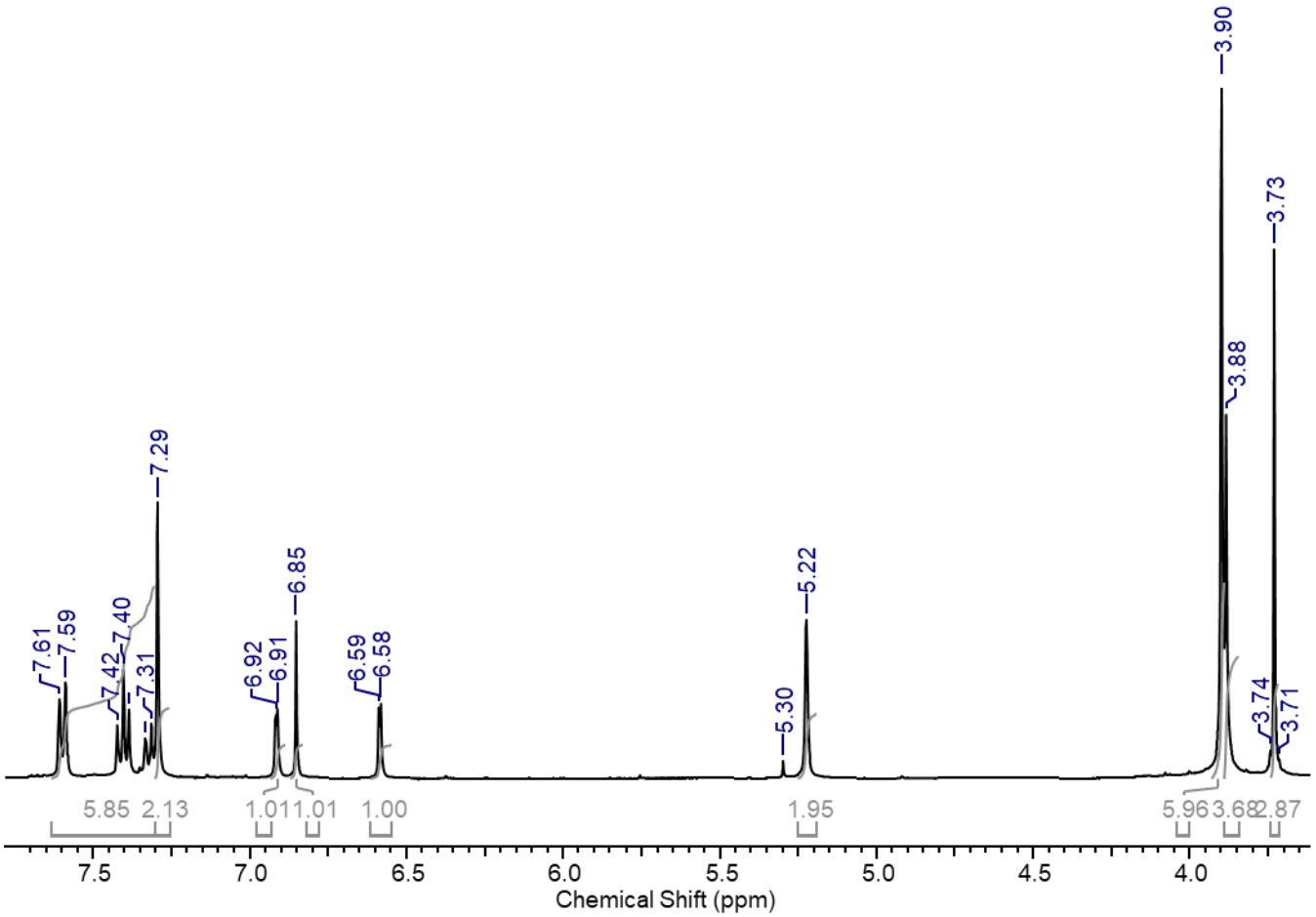


8.16 Characterisation data for the synthesis of 5-(benzyloxy)-7-methoxy-2-(3,4,5-trimethoxyphenyl)-4H-chromen-4-one, compound 2.15.

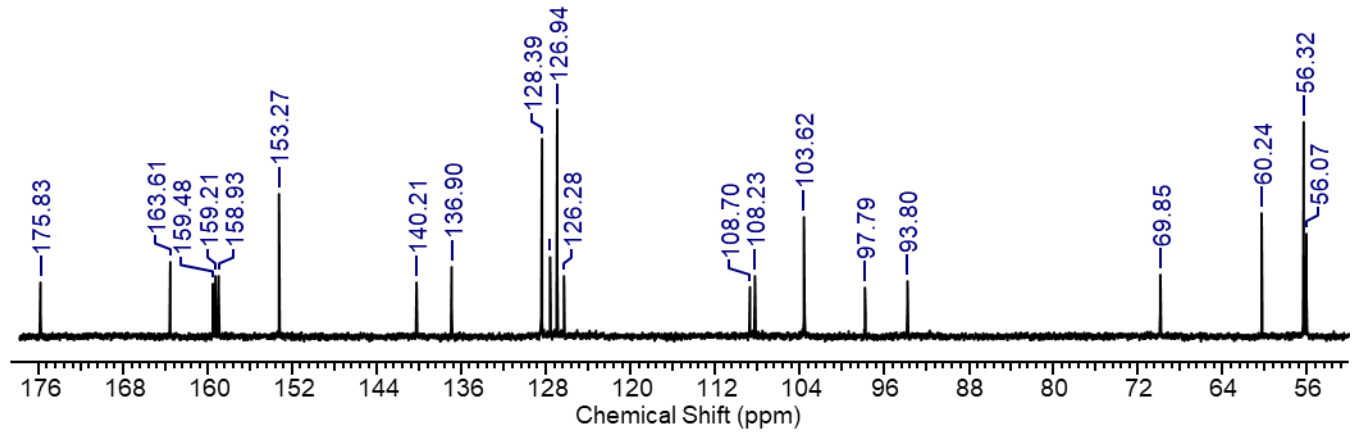
8.16.1 Extended ^{1}H NMR spectrum



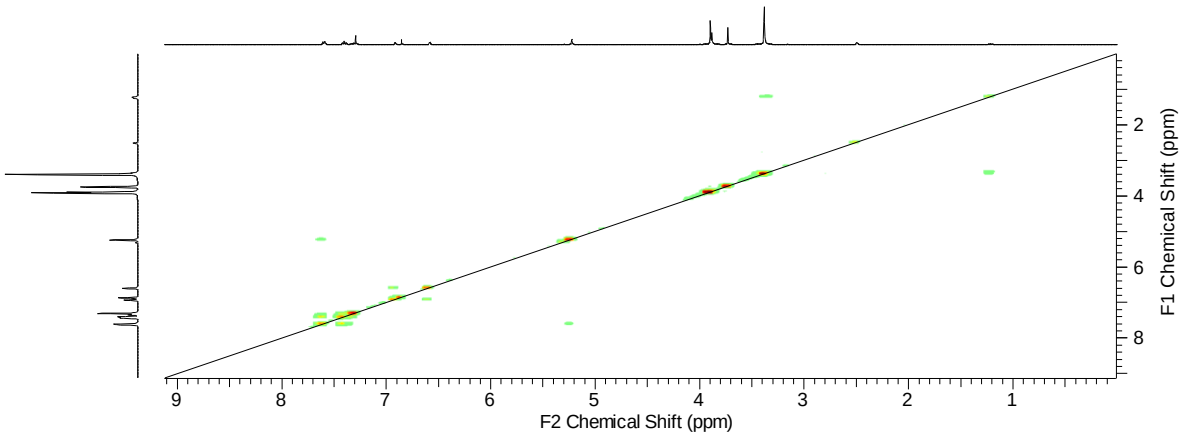
8.16.2 Magnified ^{1}H NMR spectrum



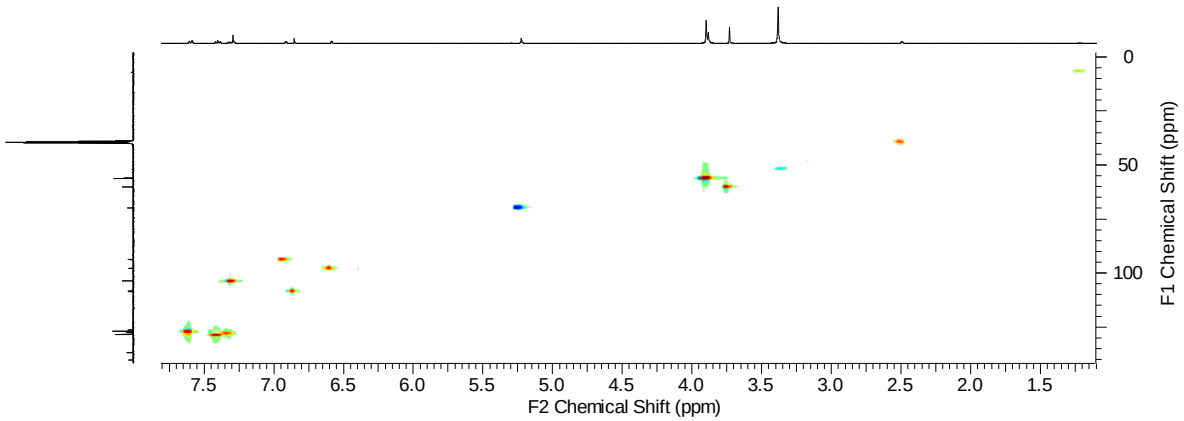
8.16.3 Magnified ^{13}C NMR spectrum



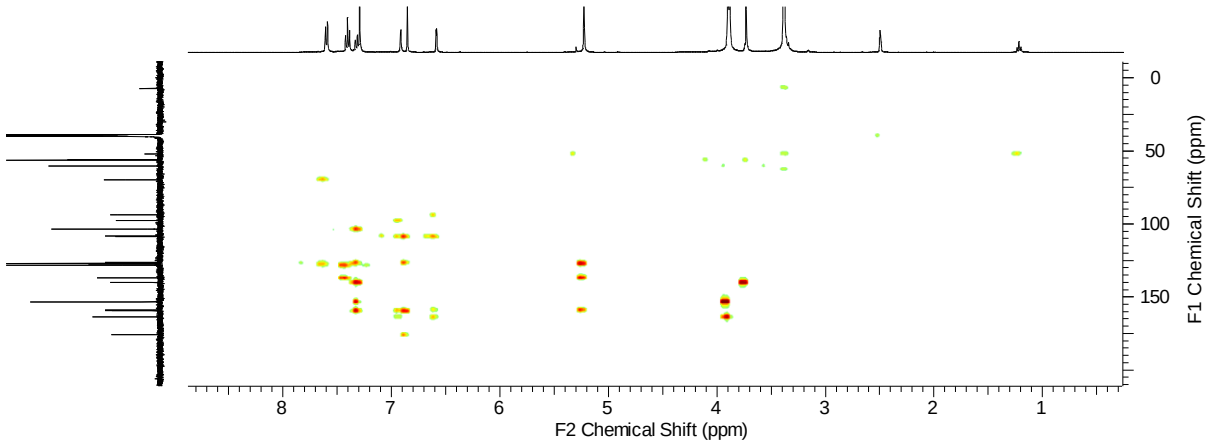
8.16.4 COSY spectrum



8.16.5 HQSC spectrum

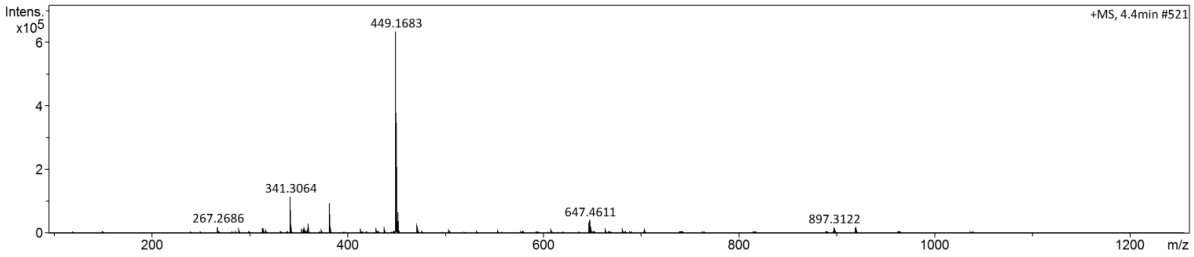


8.16.6 HMBC spectrum

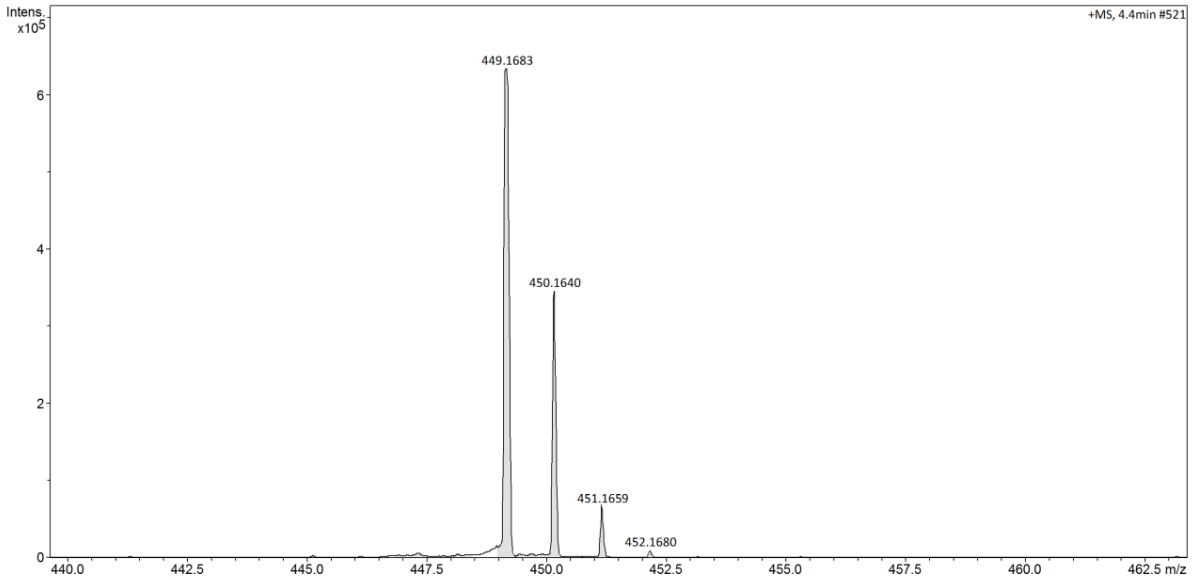


8.16.7 HRMS spectrum

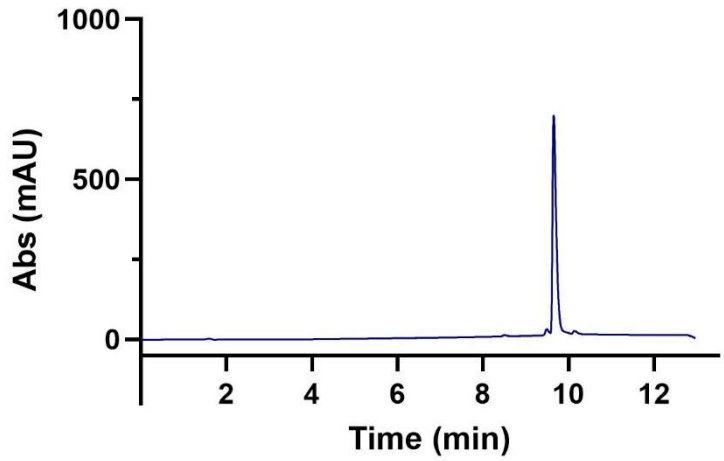
a)


Intens.
x10^5
+MS, 4.4min #521
449.1683
341.3064
267.2686
647.4611
897.3122
m/z

b)

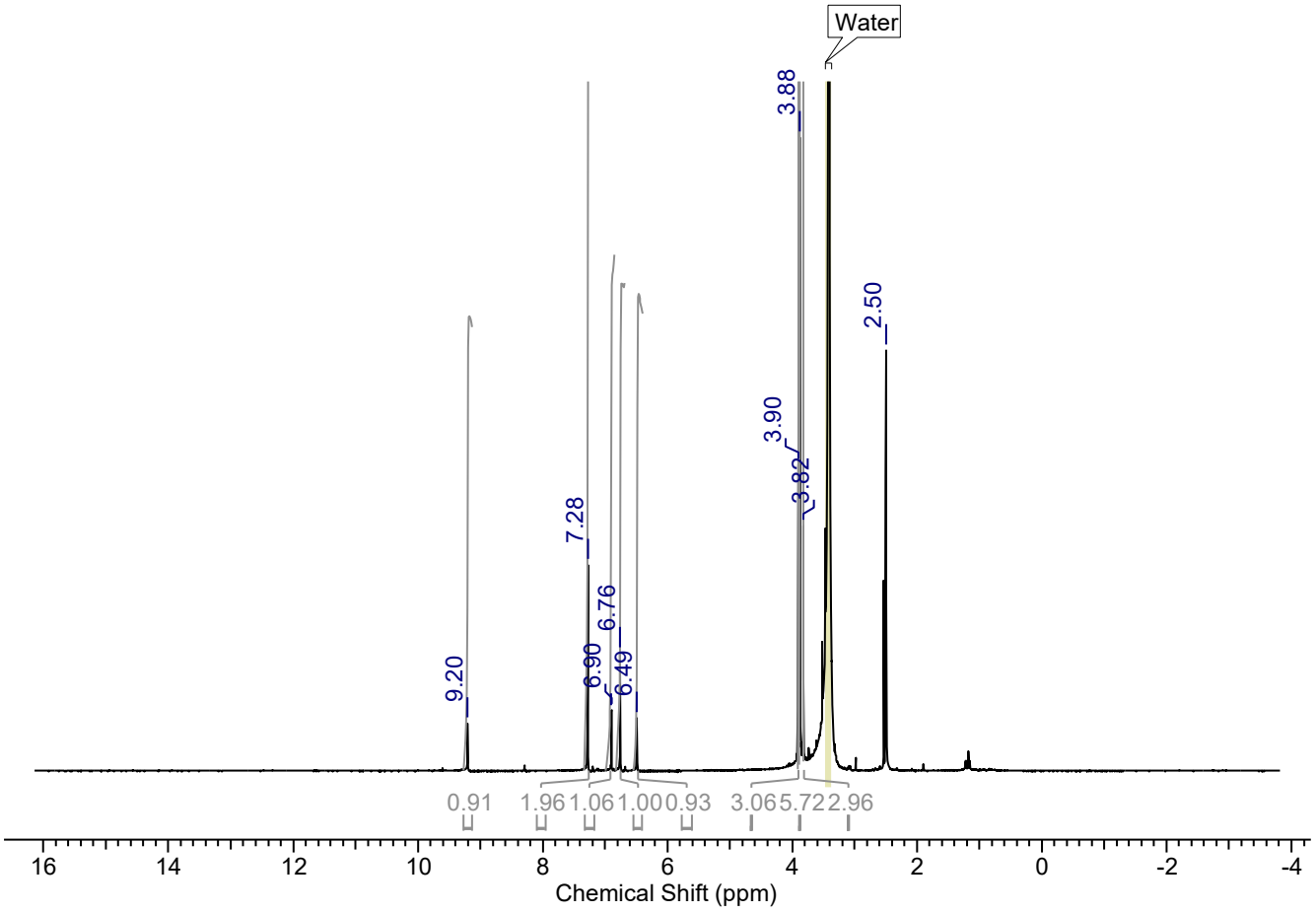

Intens.
x10^5
+MS, 4.4min #521
449.1683
450.1640
451.1659
452.1680
m/z

8.16.8 UV chromatogram of purified compound showing absorbance at 254 nm.

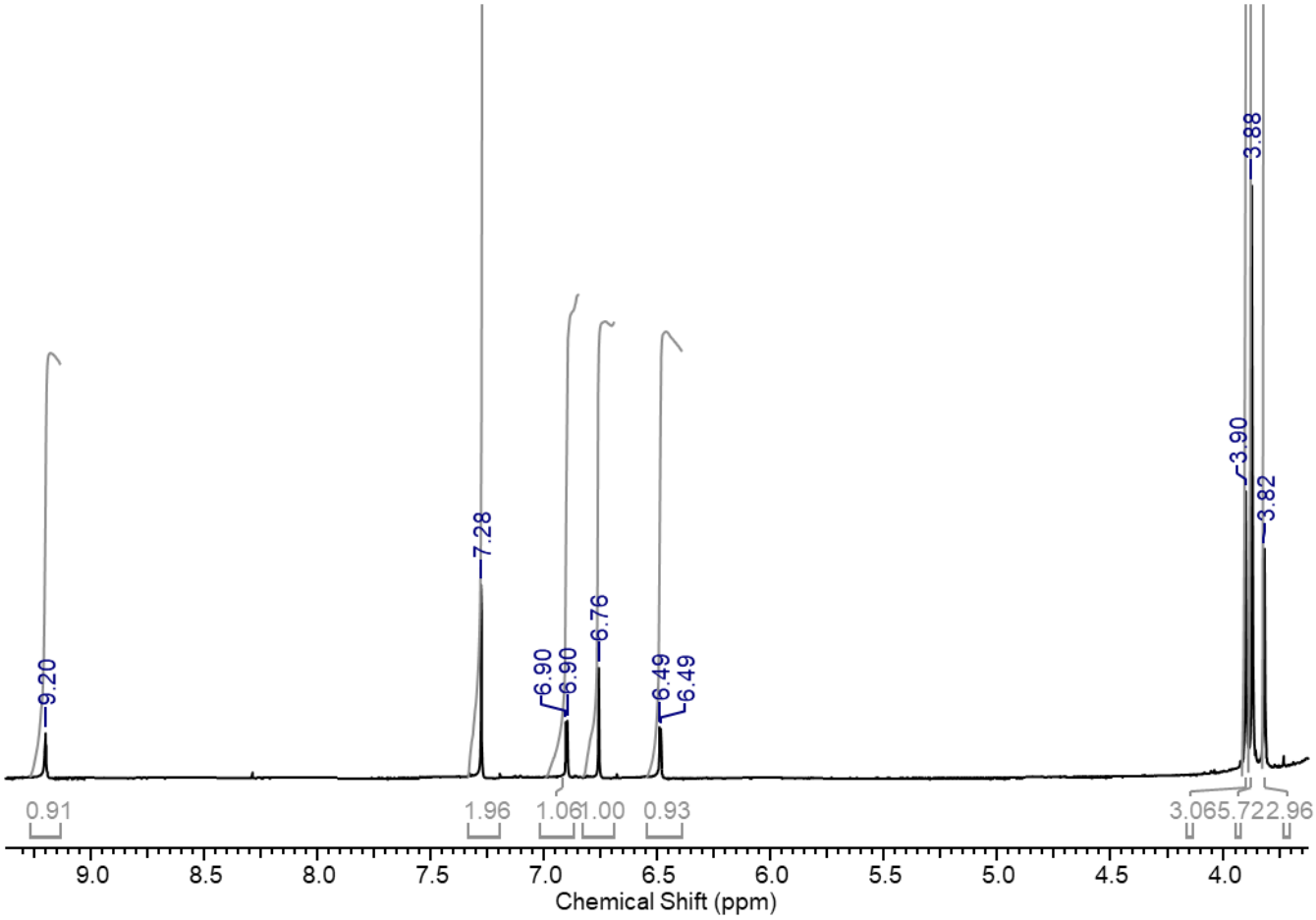


8.17 Characterisation data for the synthesis of 2-(4-hydroxy-3,5-dimethoxyphenyl)-5,7-dimethoxy-4H-chromen-4-one, compound 2,16.

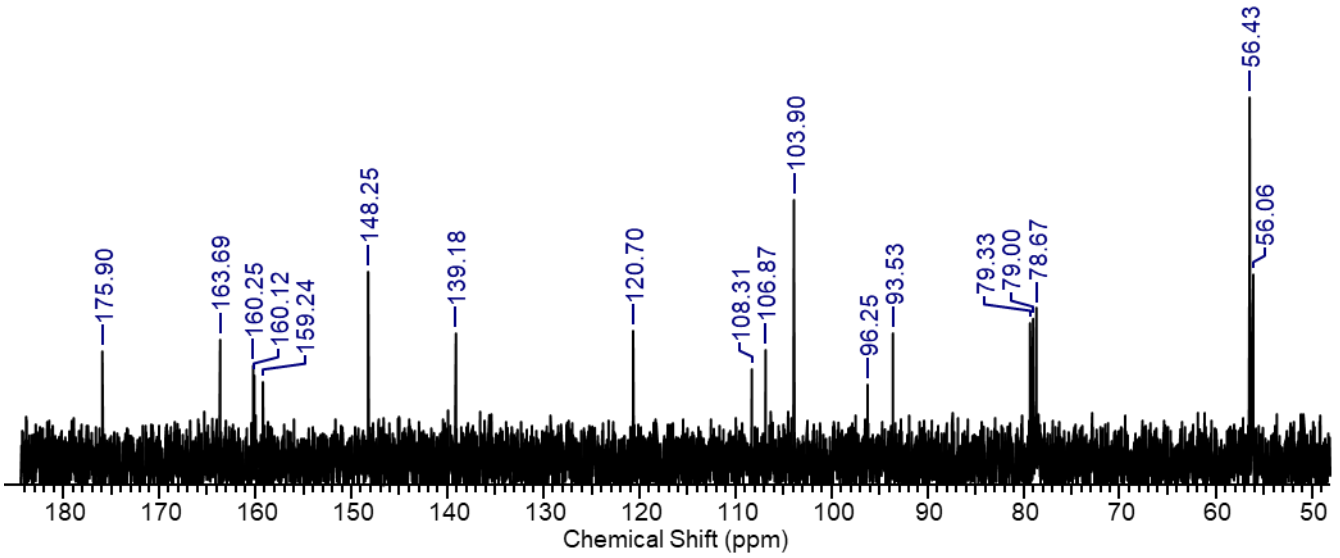
8.17.1 Extended ^{1}H NMR spectrum



8.17.2 Magnified ^{1}H NMR spectrum

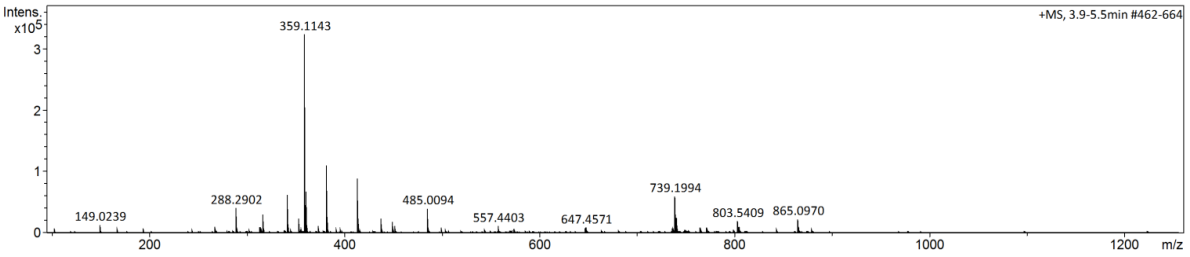


8.17.3 Magnified ^{13}C NMR spectrum

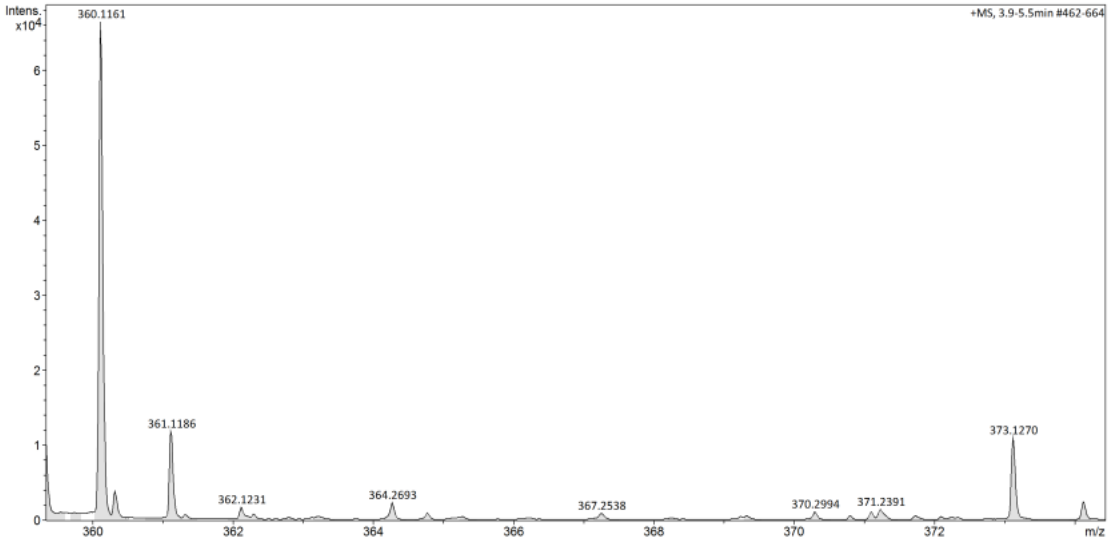


8.17.4 HRMS spectra

a)



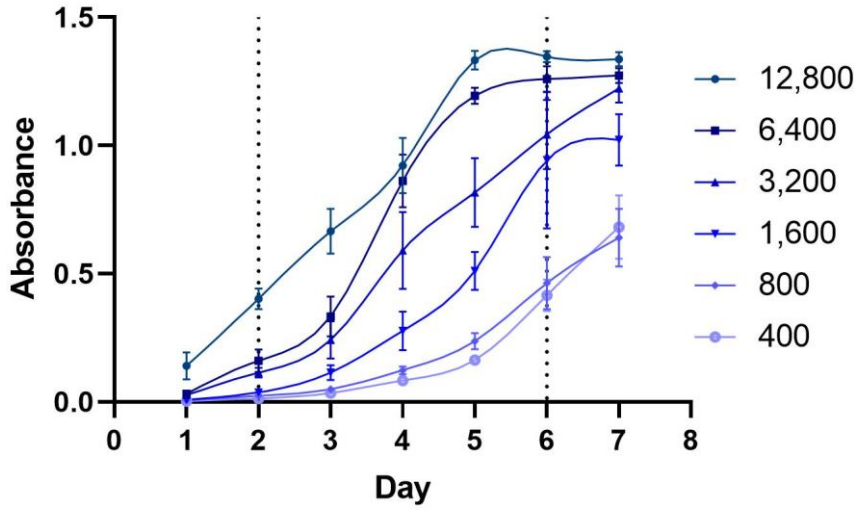
b)



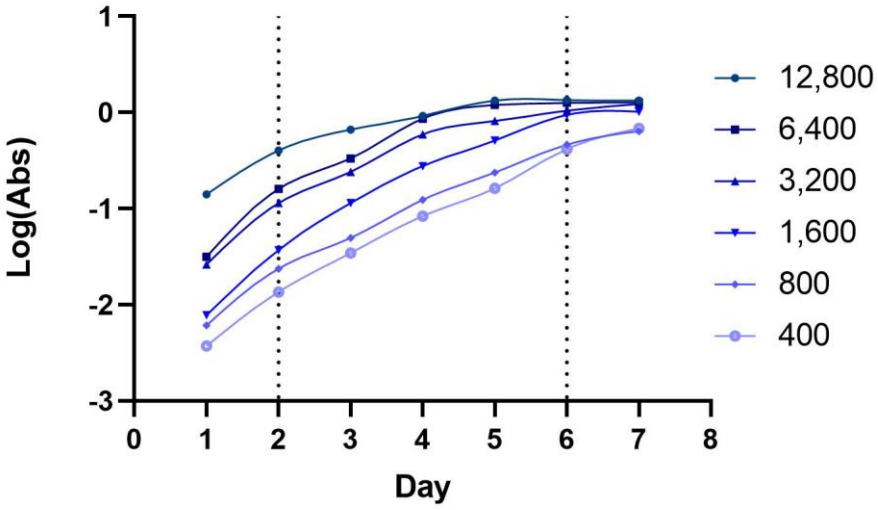
8.18 HT29 Seeding Density Assays

8.18.1 Assay 2

a) Growth curve 2 of HT29 cells created from absorbance data of each well which is proportional to the number of cells in each well.

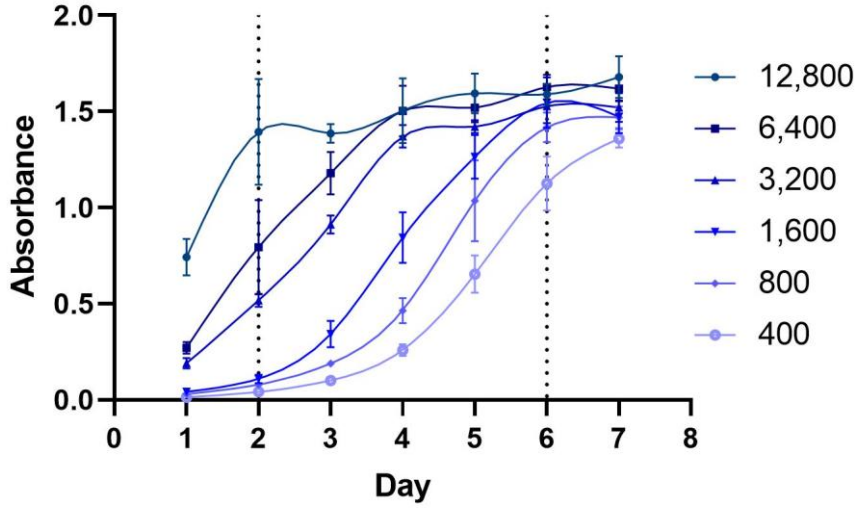


b) Growth curve 2 of HT29 cells created from log of absorbance data of each well which is proportional to the number of cells in each well.

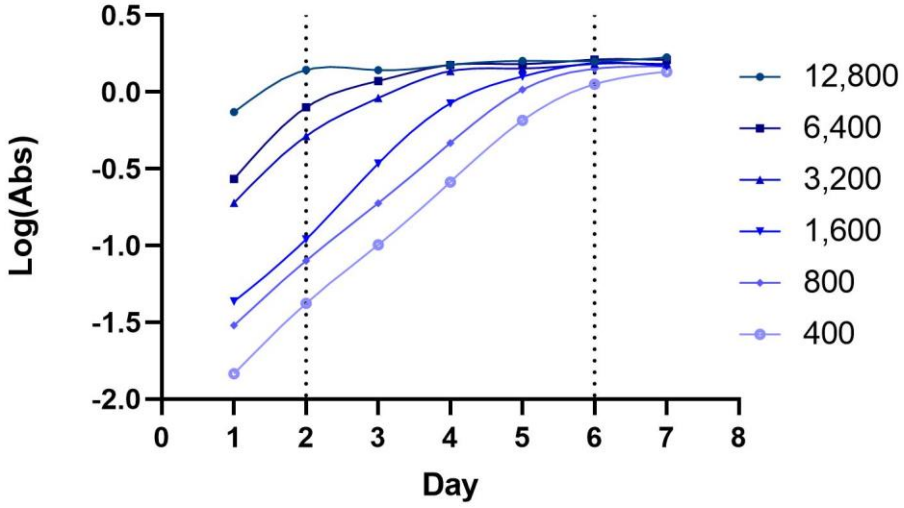


8.18.2 Assay 3

a) Growth curve 3 of HT29 cells created from absorbance data of each well which is proportional to the number of cells in each well.



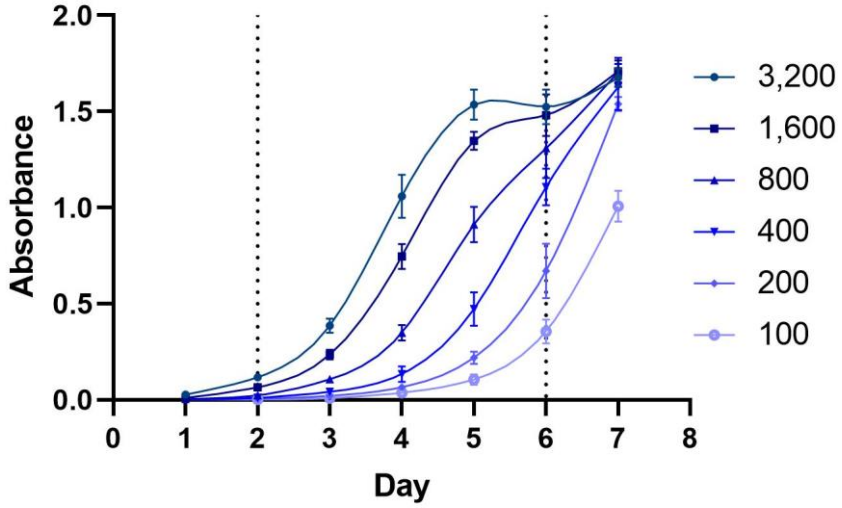
b) Growth curve 3 of HT29 cells created from log of absorbance data of each well which is proportional to the number of cells in each well.



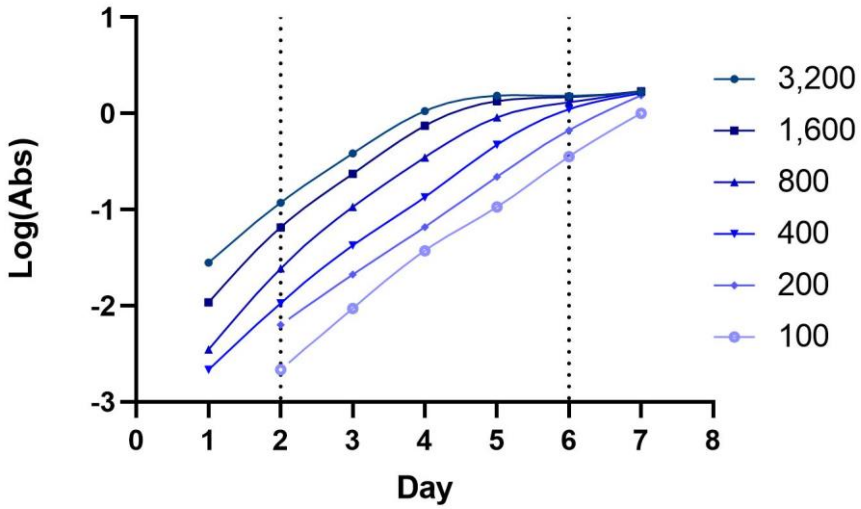
8.19 HCT116 Seeding Density Assays

8.19.1 Assay 2

a) Growth curve 2 of HCT116 cells created from absorbance data of each well which is proportional to the number of cells in each well.



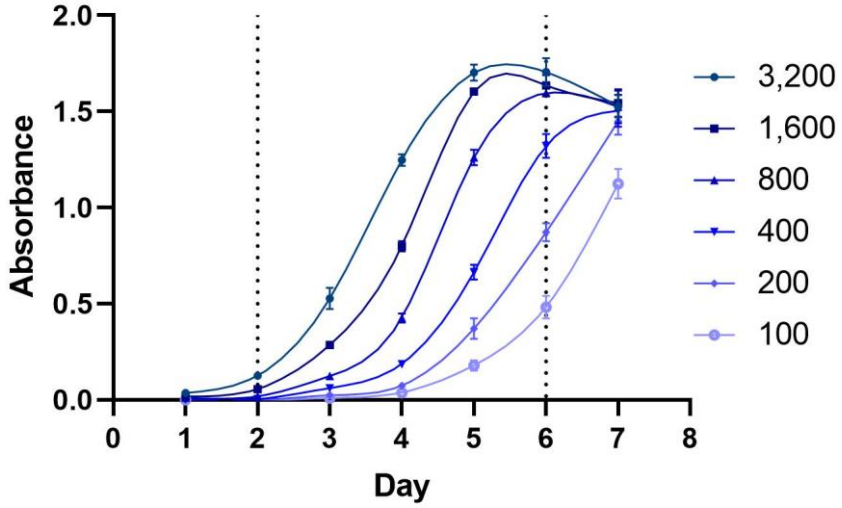
b) Growth curve 2 of HCT116 cells created from log of absorbance data of each well which is proportional to the number of cells in each well.*



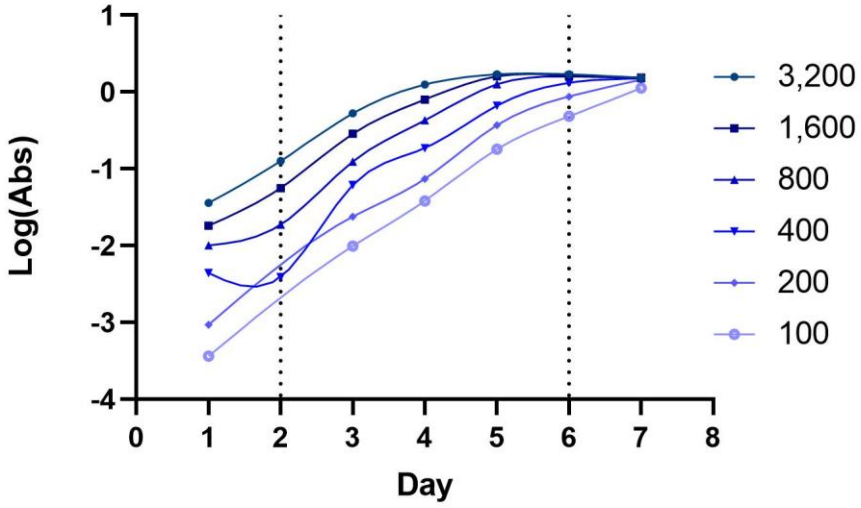
* Data points have been omitted for day 1 for 100 and 200 cells/well as the mean absorbance data was below 0 and returned an undefined number which could not be plotted.

8.19.2 Assay 3

a) Growth curve 3 of HCT116 cells created from absorbance data of each well which is proportional to the number of cells in each well.

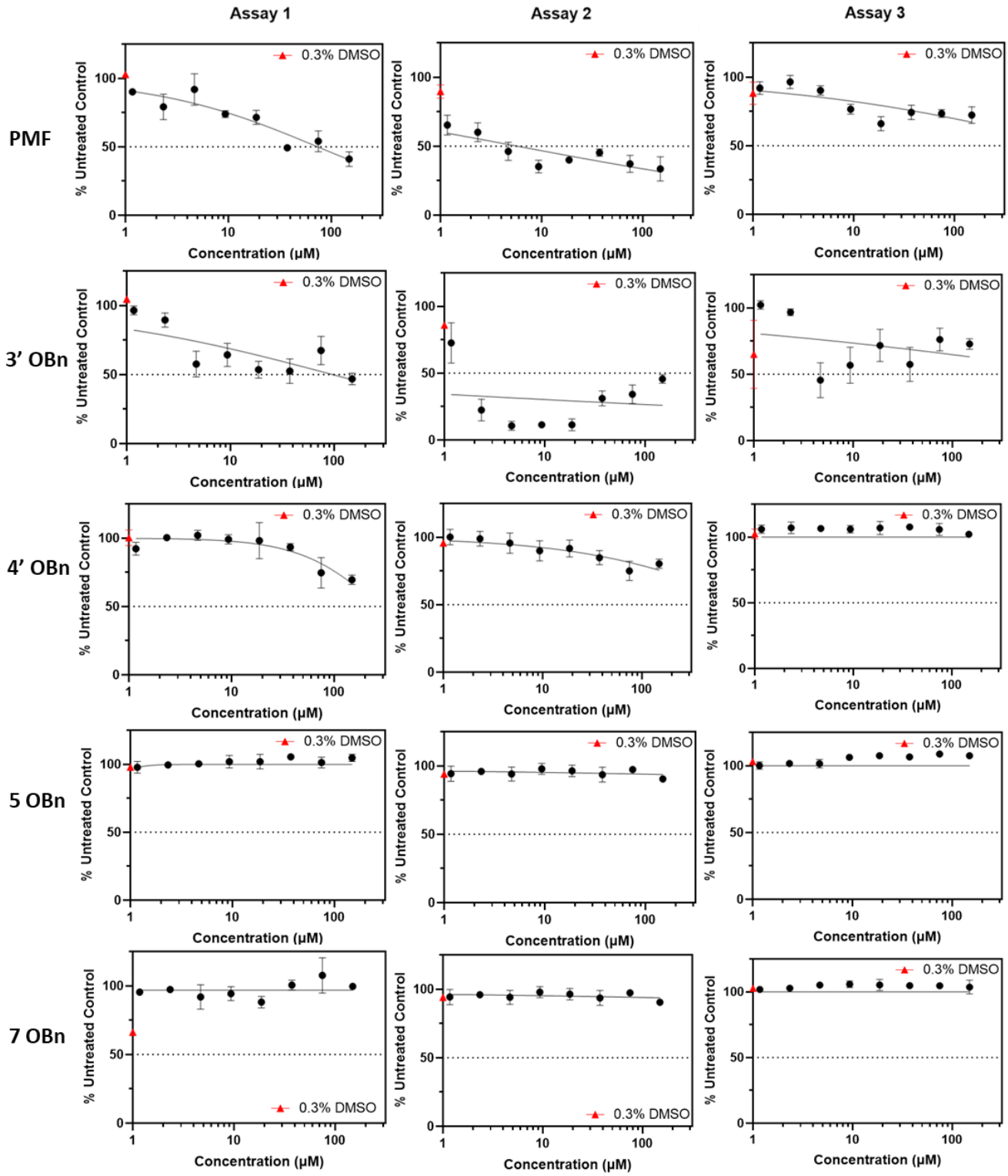


b) Growth curve 3 of HCT116 cells created from log of absorbance data of each well which is proportional to the number of cells in each well.*

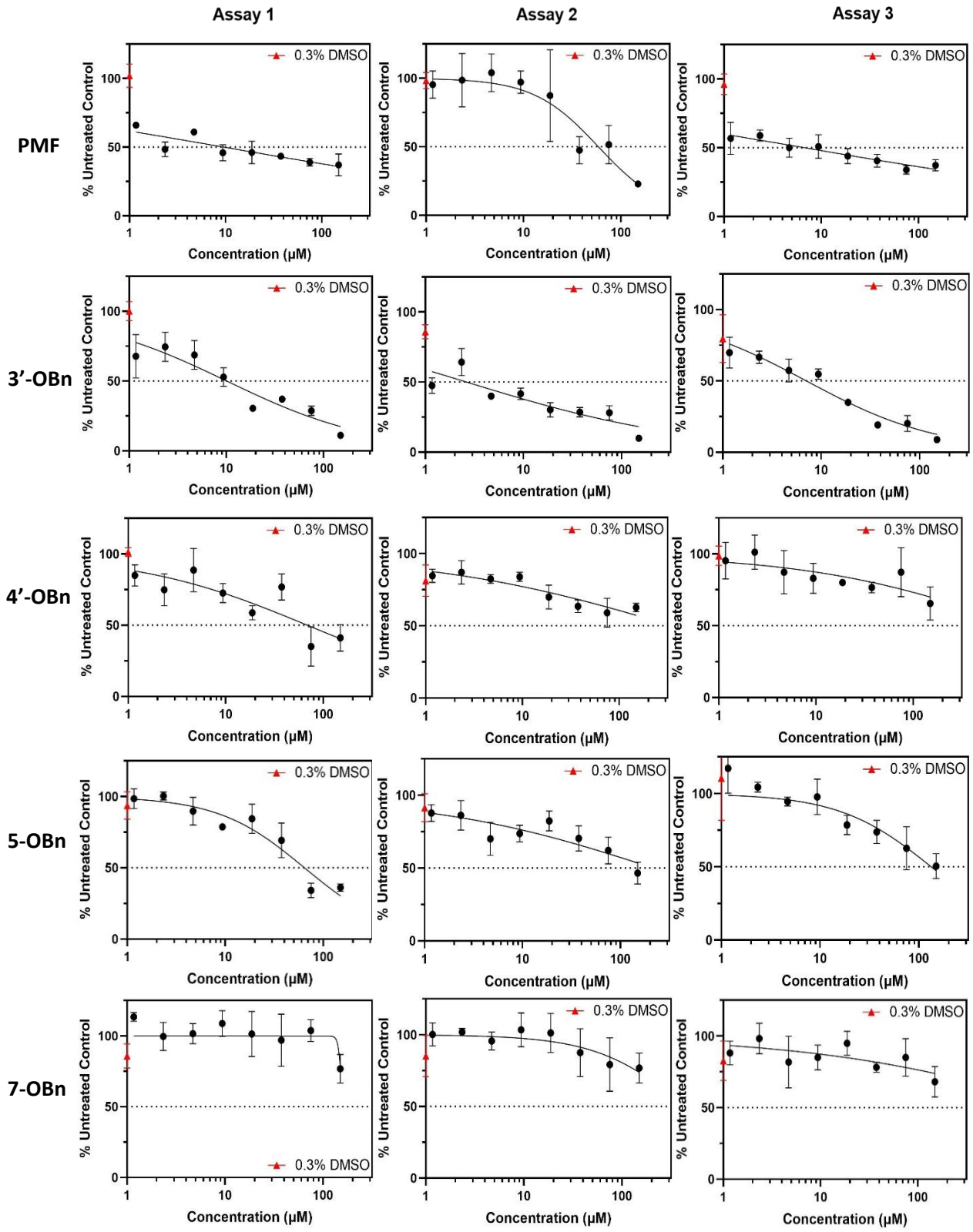


* Data points have been omitted for day 1 for 100 and 200 cells/well as the mean absorbance data was below 0 and returned an undefined number which could not be plotted.

8.20 Triplicate data for PMF and other analogues when treating HT29 cells.

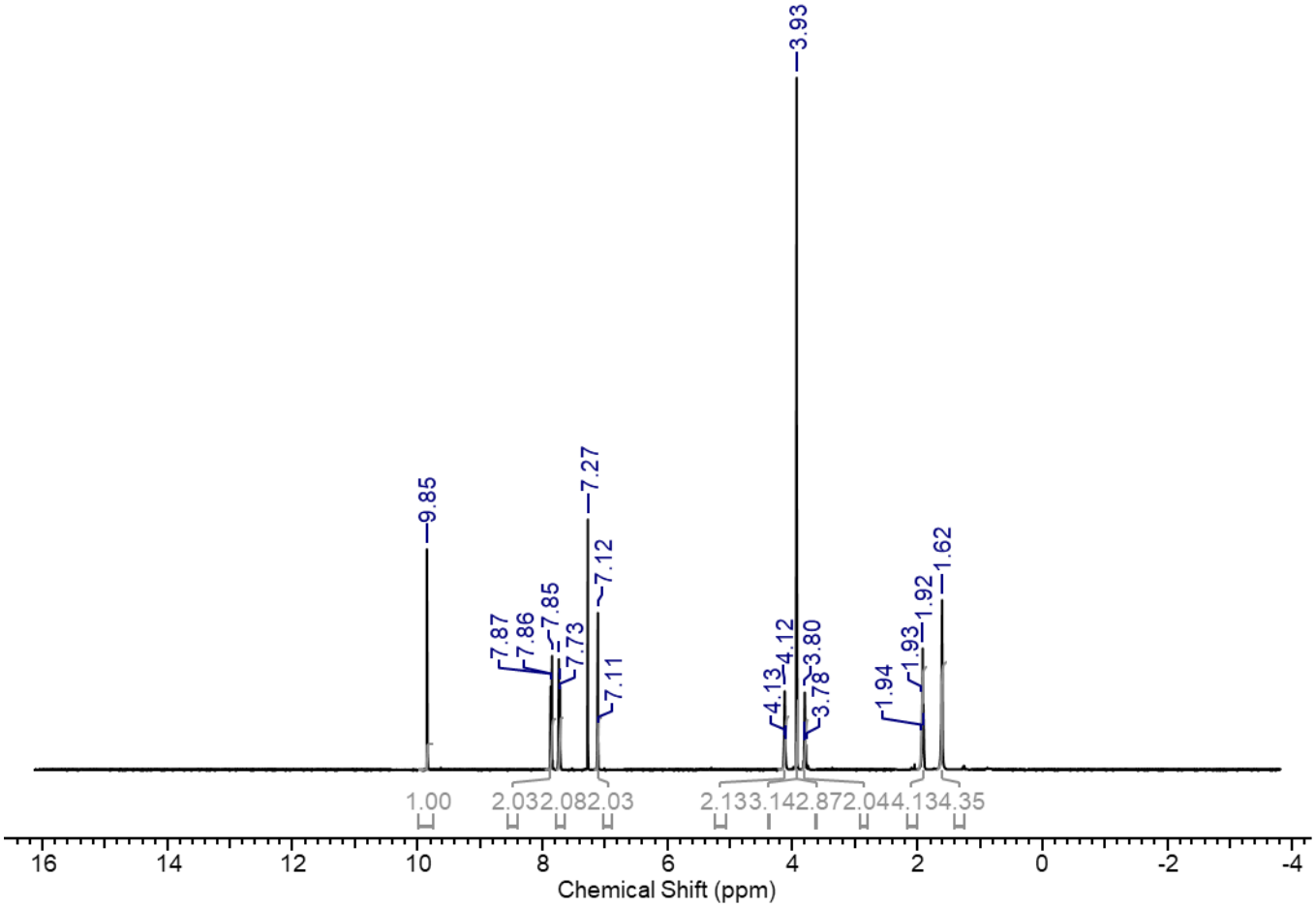


8.21 Triplicate data for PMF and other analogues when treating HCT116 cells.

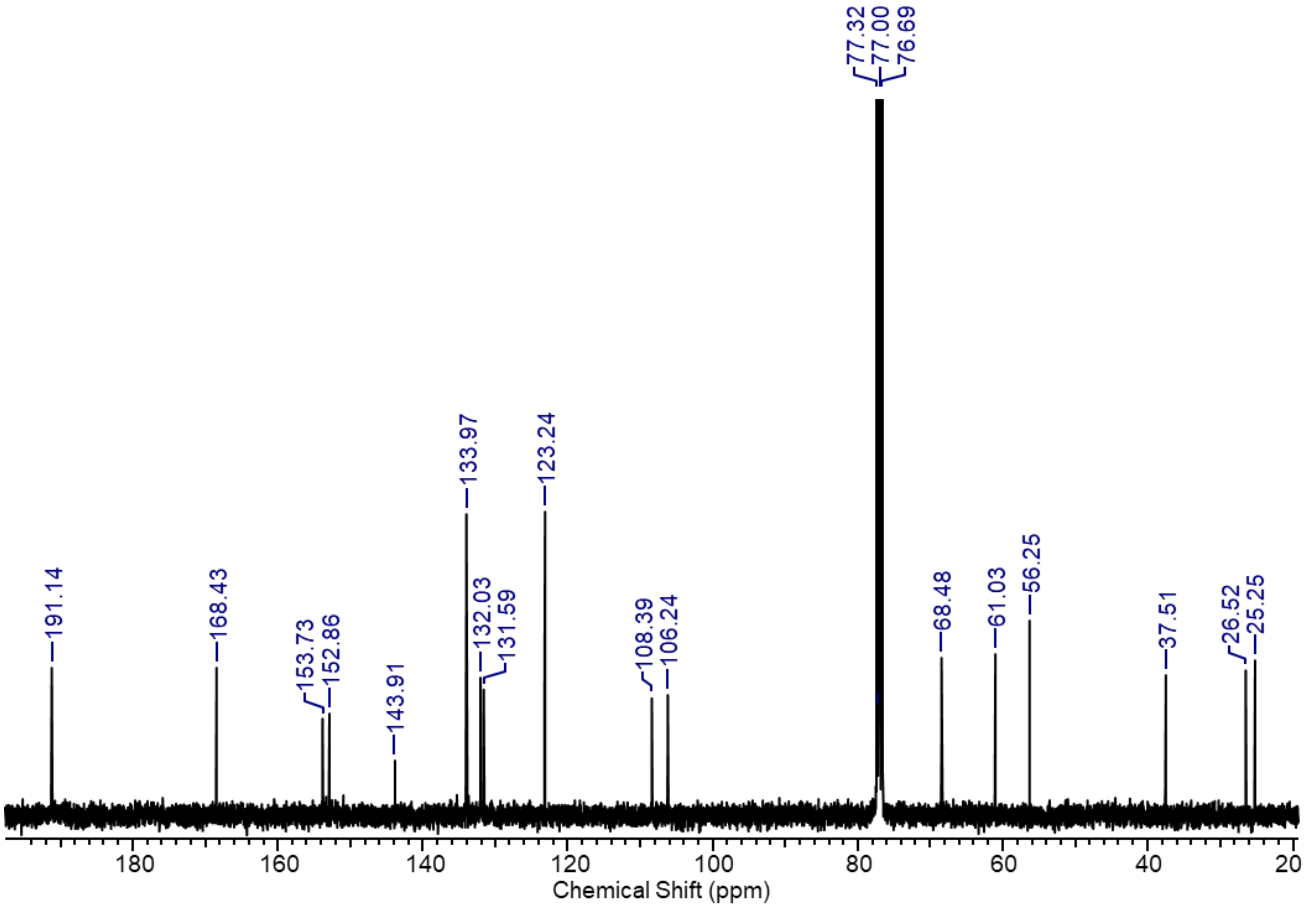


8.22 Characterisation data for the synthesis of 3-(4-(1,3-dioxoisoindolin-2-yl)butoxy)-4,5-dimethoxybenzaldehyde, **4.1**.

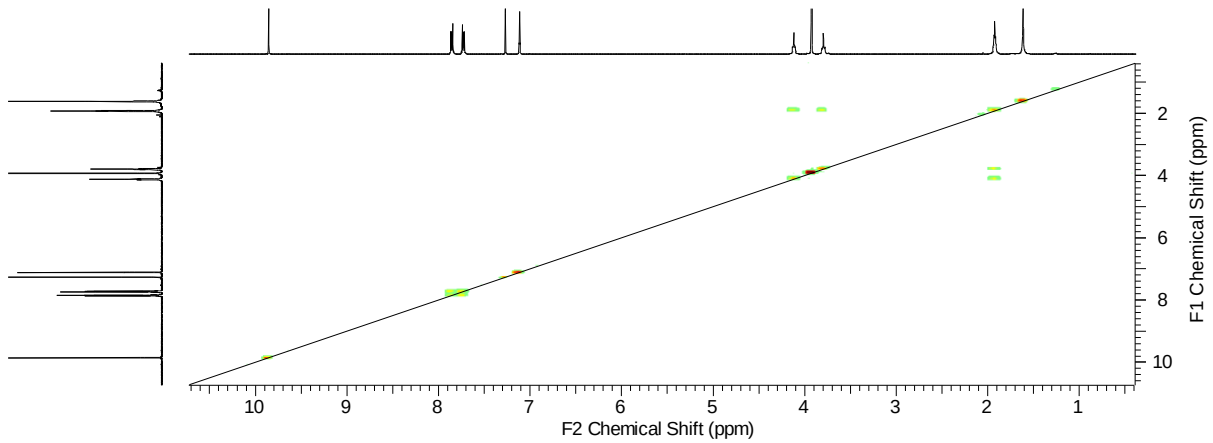
8.22.1 Extended ^{1}H NMR spectrum



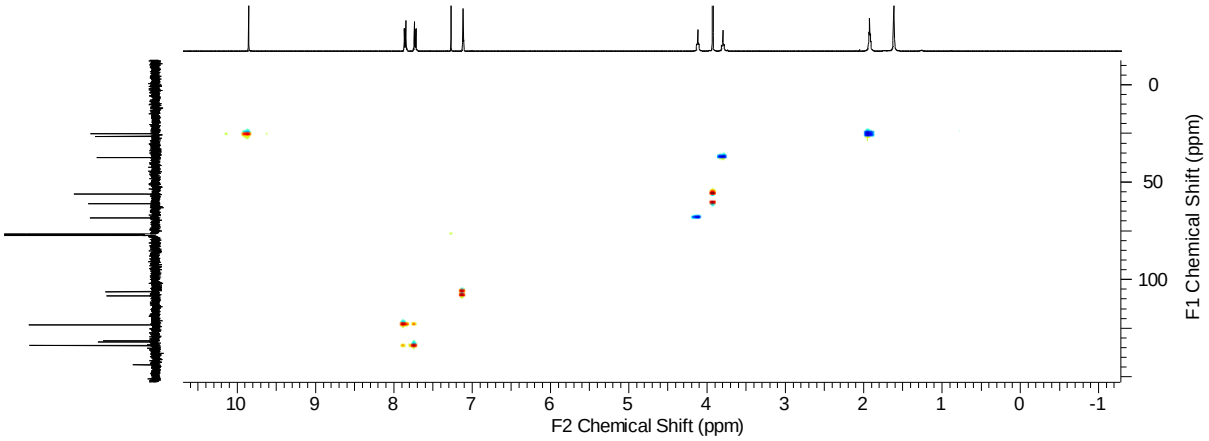
8.22.2 Magnified ^{13}C NMR spectrum



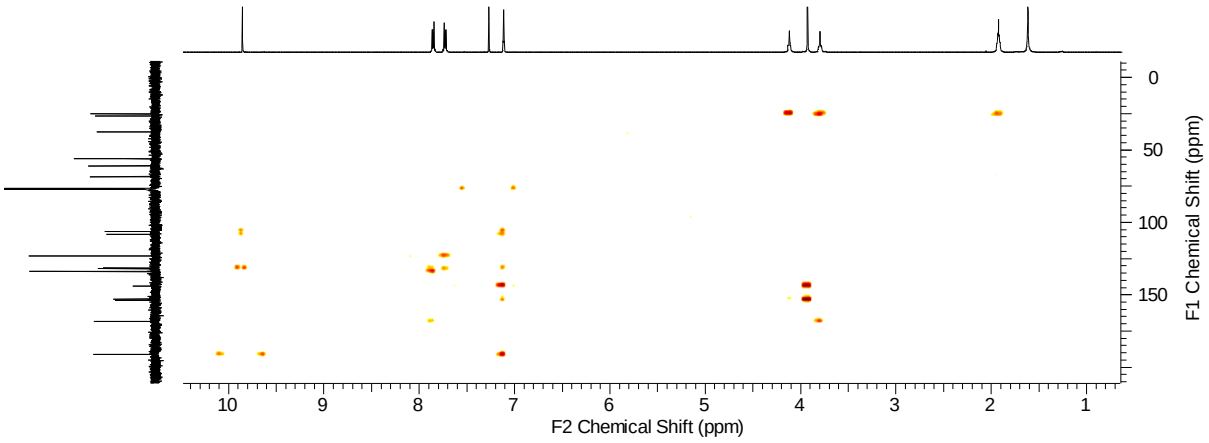
8.22.3 COSY NMR spectrum



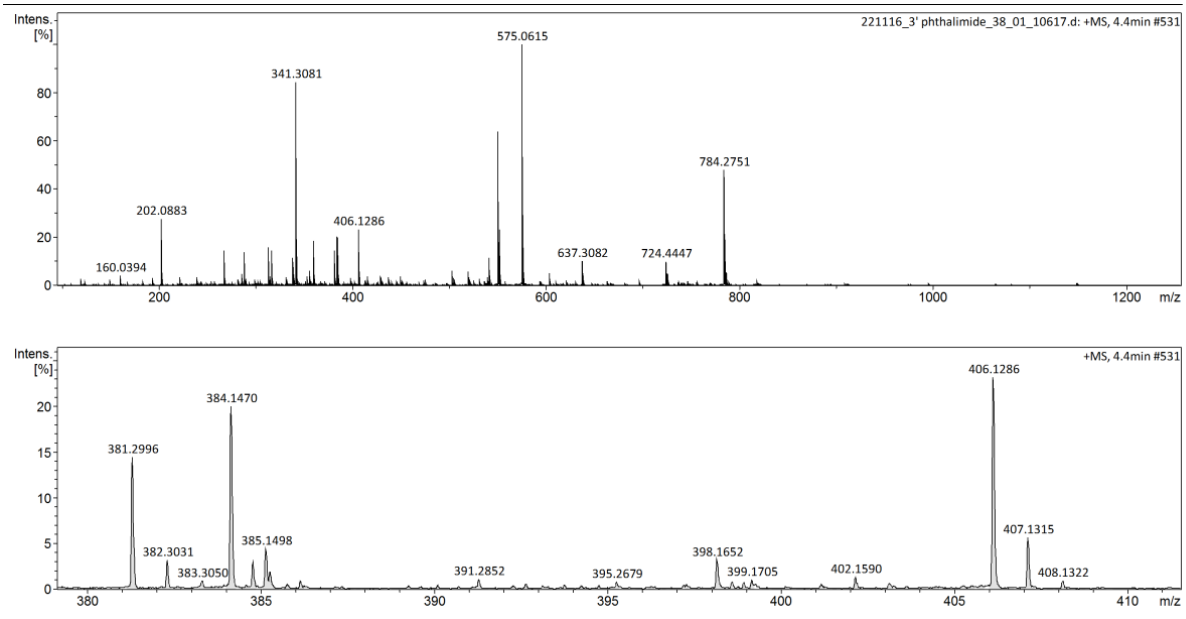
8.22.4 HSQC NMR spectrum



8.22.5 HMBC NMR spectrum

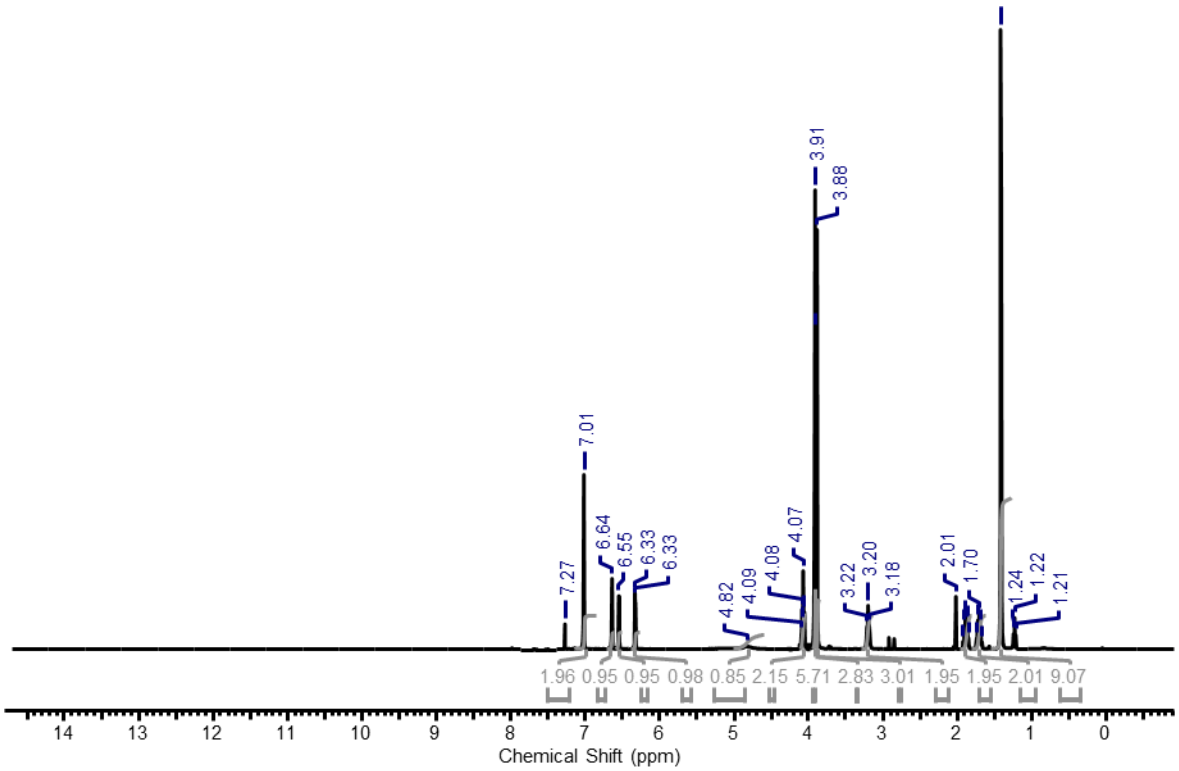


8.22.6 HR-MS spectrum

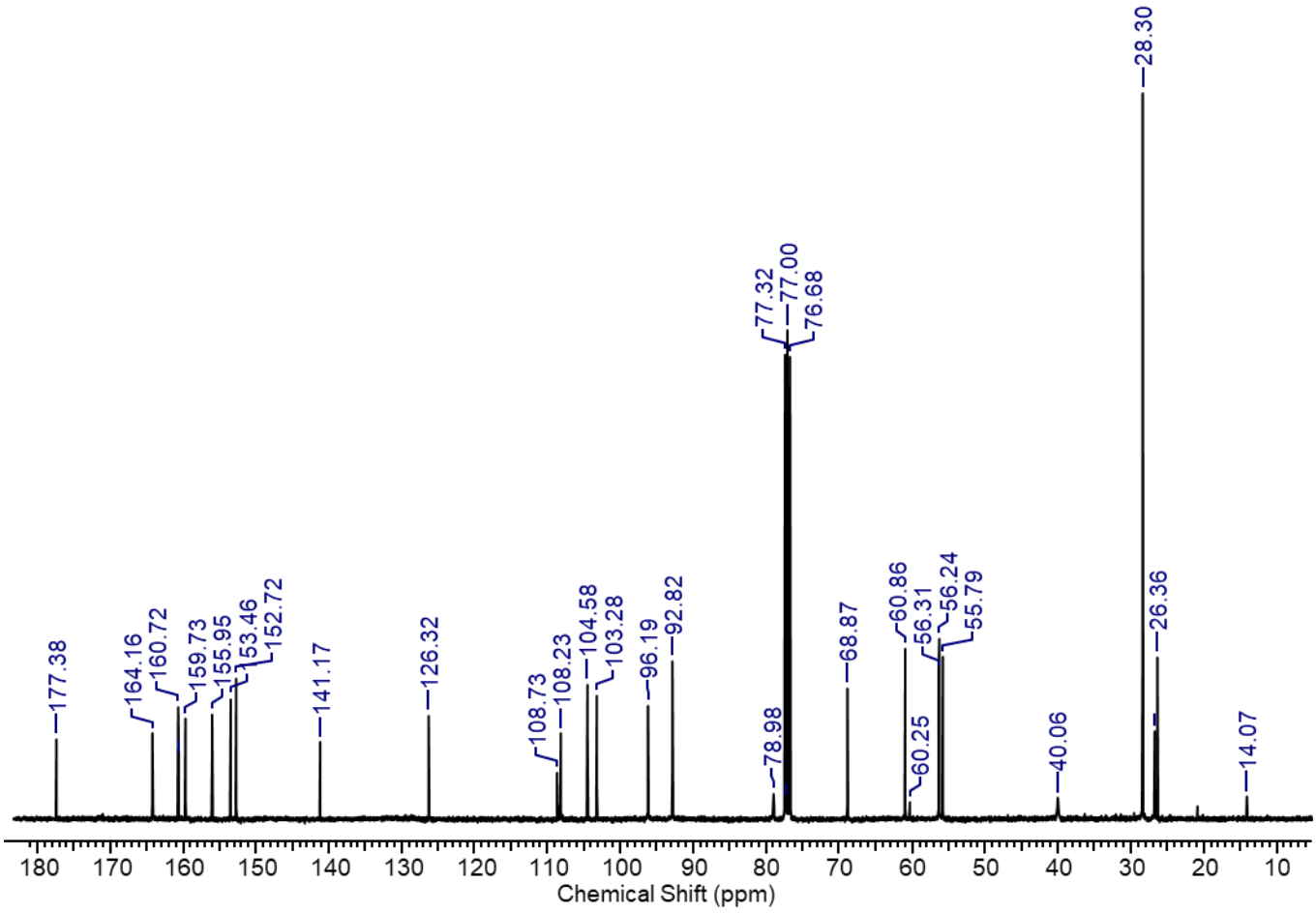


8.23 Characterisation data for synthesis of tert-butyl (4-(5-(5,7-dimethoxy-4-oxo-4H-chromen-2-yl)-2,3-dimethoxyphenoxy)butyl)carbamate, **4.2**.

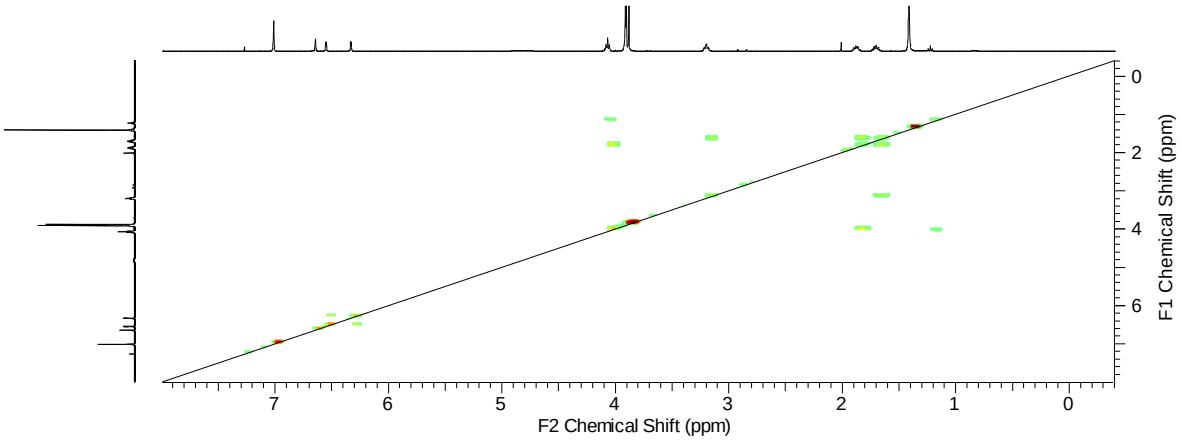
8.23.1 Extended ^{1}H NMR spectrum



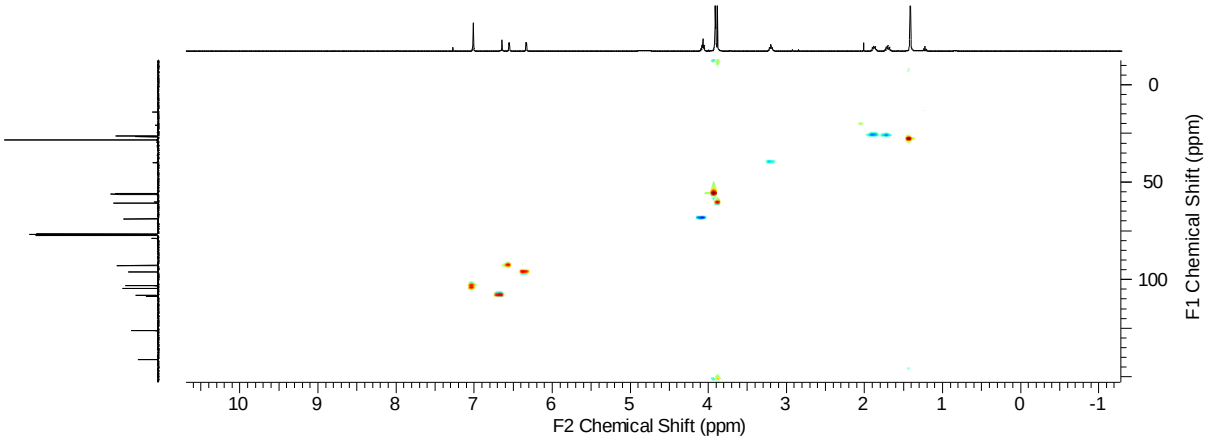
8.23.2 Magnified ^{13}C NMR spectrum



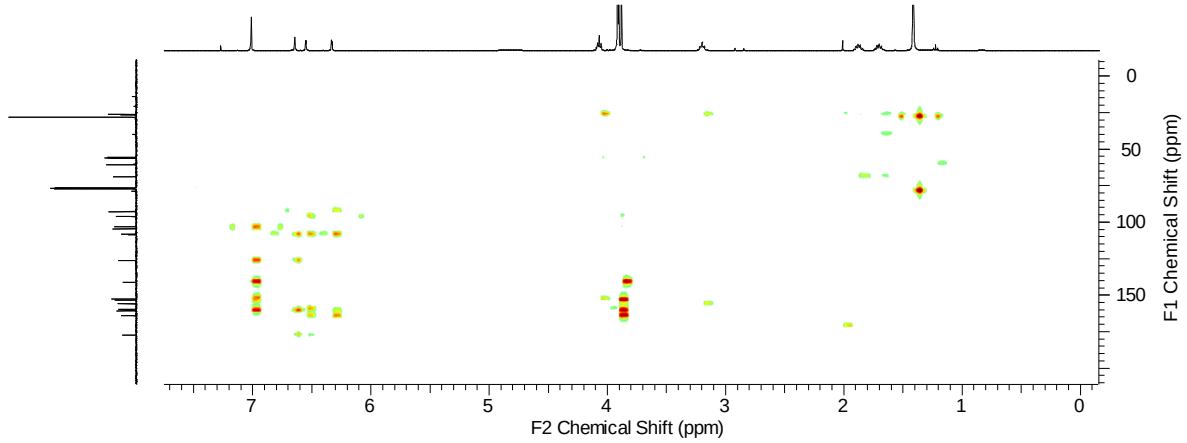
8.23.3 COSY NMR spectrum



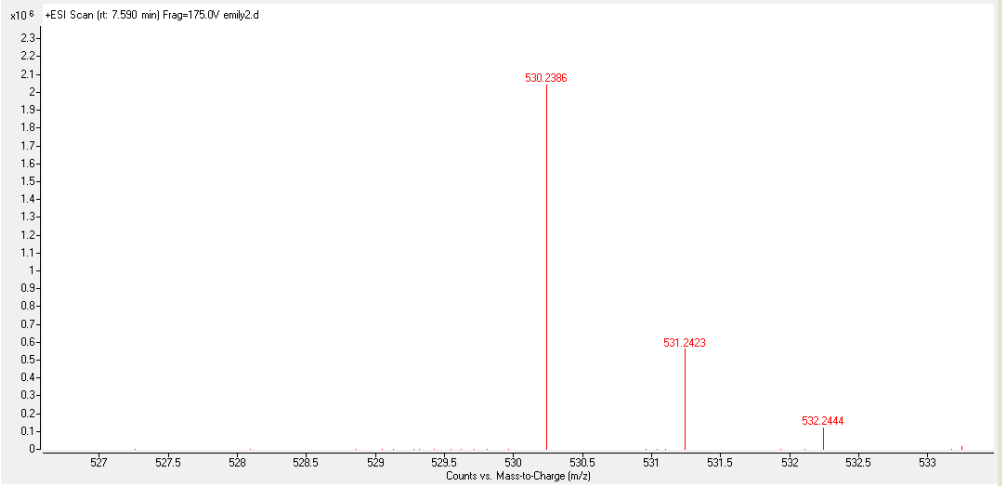
8.23.4 HSQC NMR spectrum



8.23.5 HMBC NMR spectrum

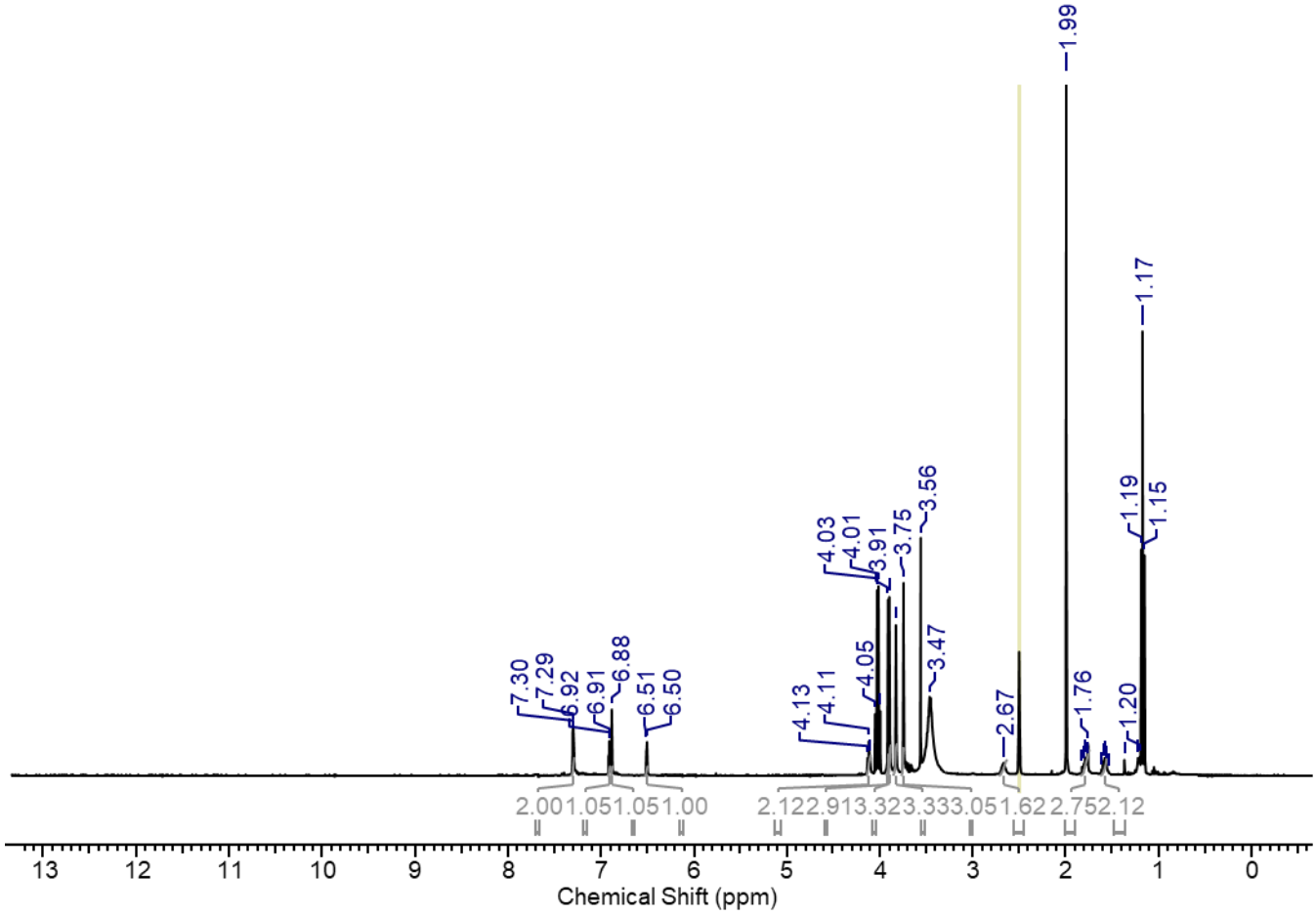


8.23.6 High resolution ESI-MS

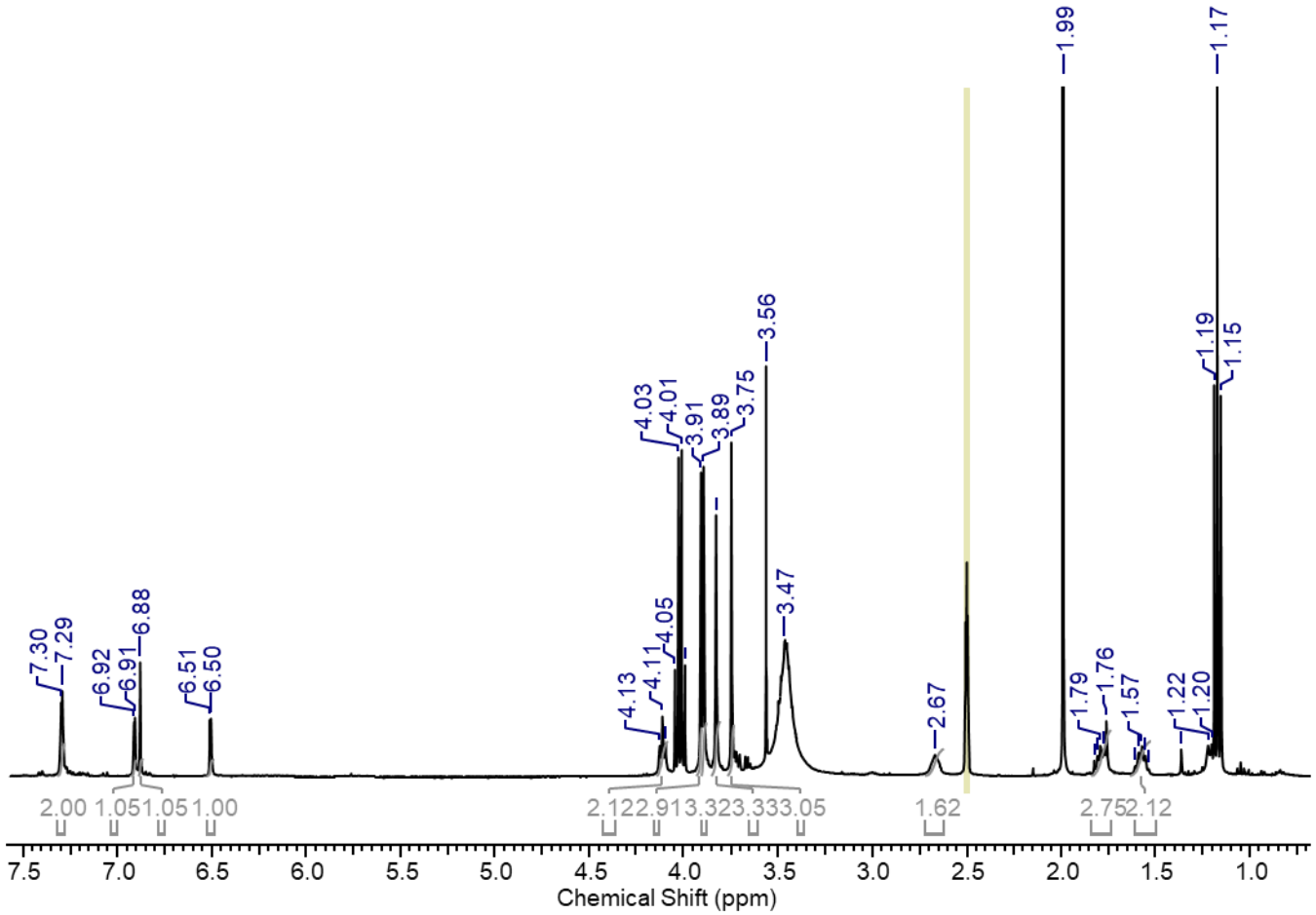


8.24 Characterisation data for synthesis of 2-(3-(4-aminobutoxy)-4,5-dimethoxyphenyl)-5,7-dimethoxy-4H-chromen-4-one, **4.3**.

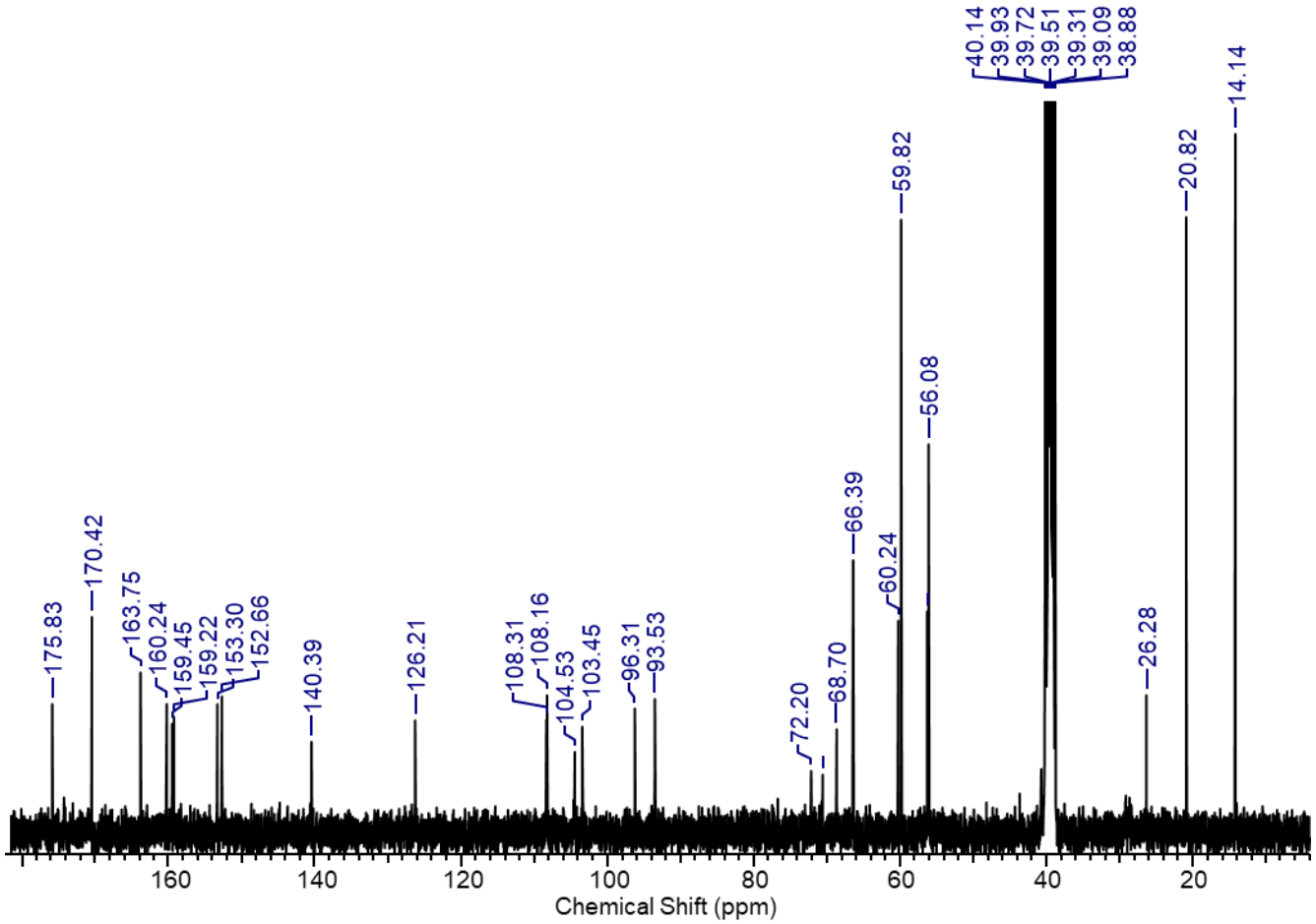
8.24.1 Extended ^{1}H NMR spectrum



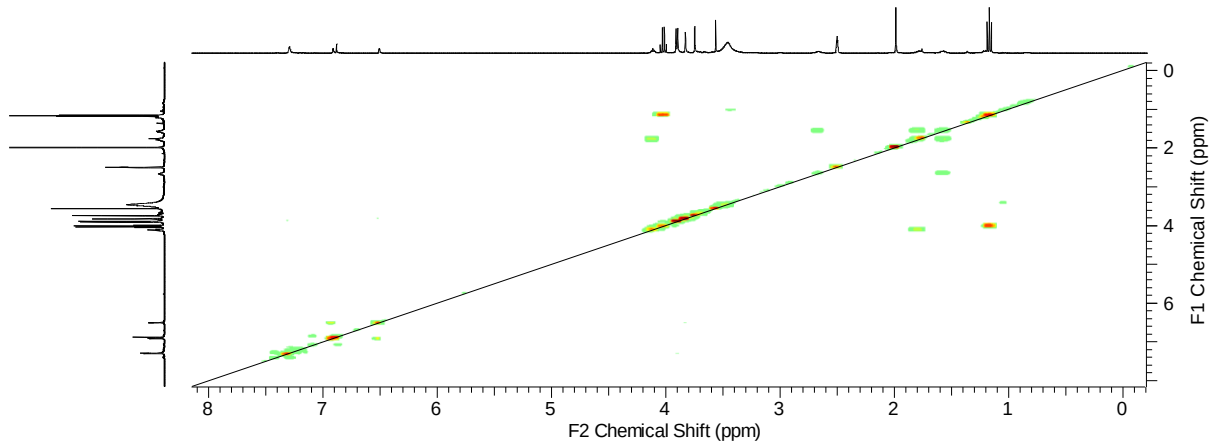
8.24.2 Magnified ^{1}H NMR spectrum.



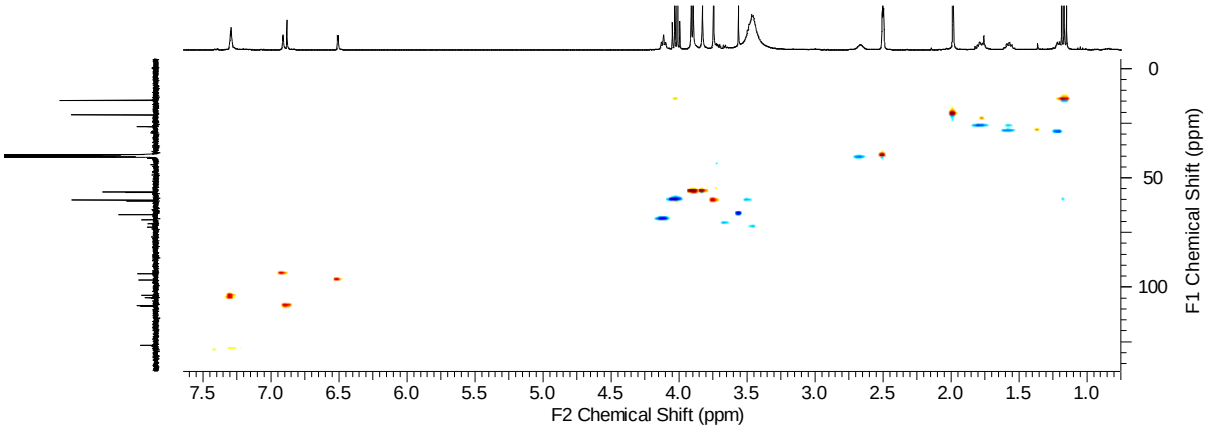
8.24.3 Magnified ^{13}C NMR spectrum



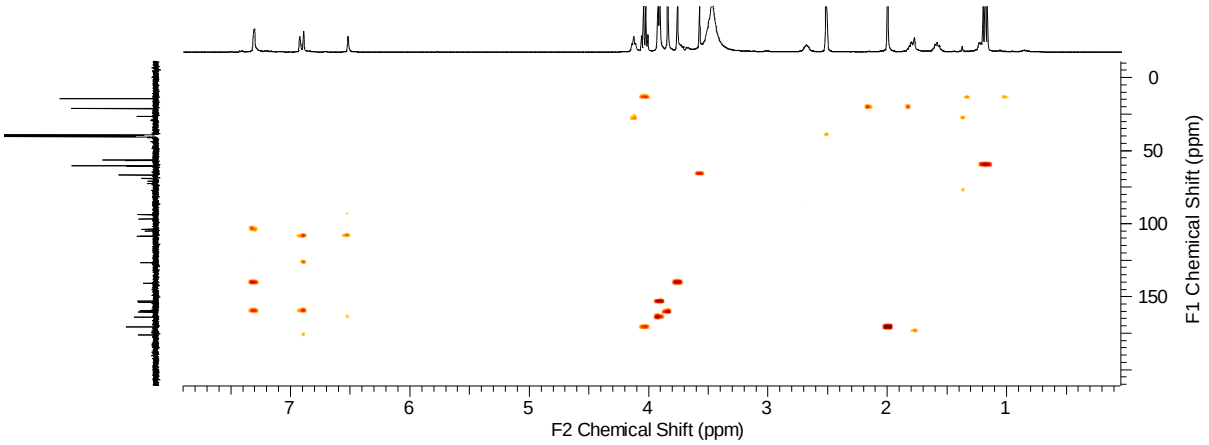
8.24.4 COSY NMR spectrum



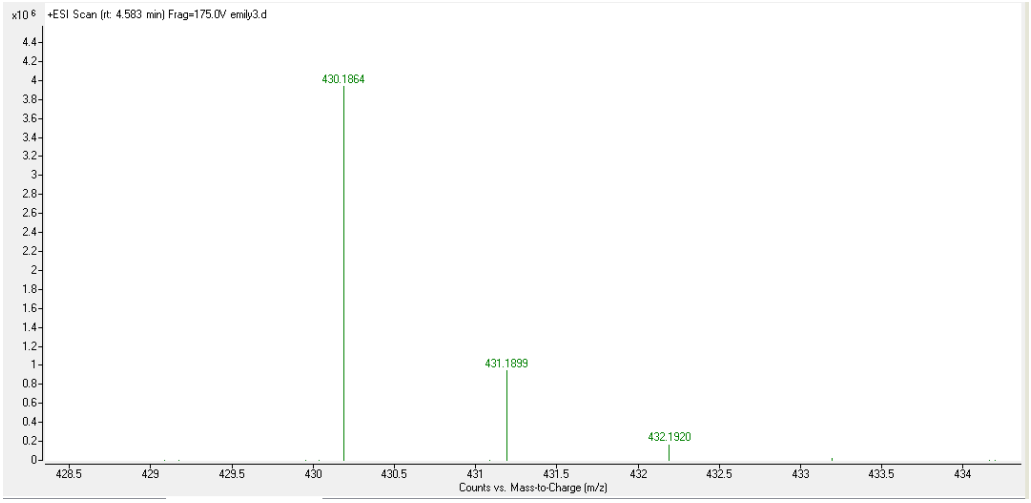
8.24.5 HSQC NMR spectrum



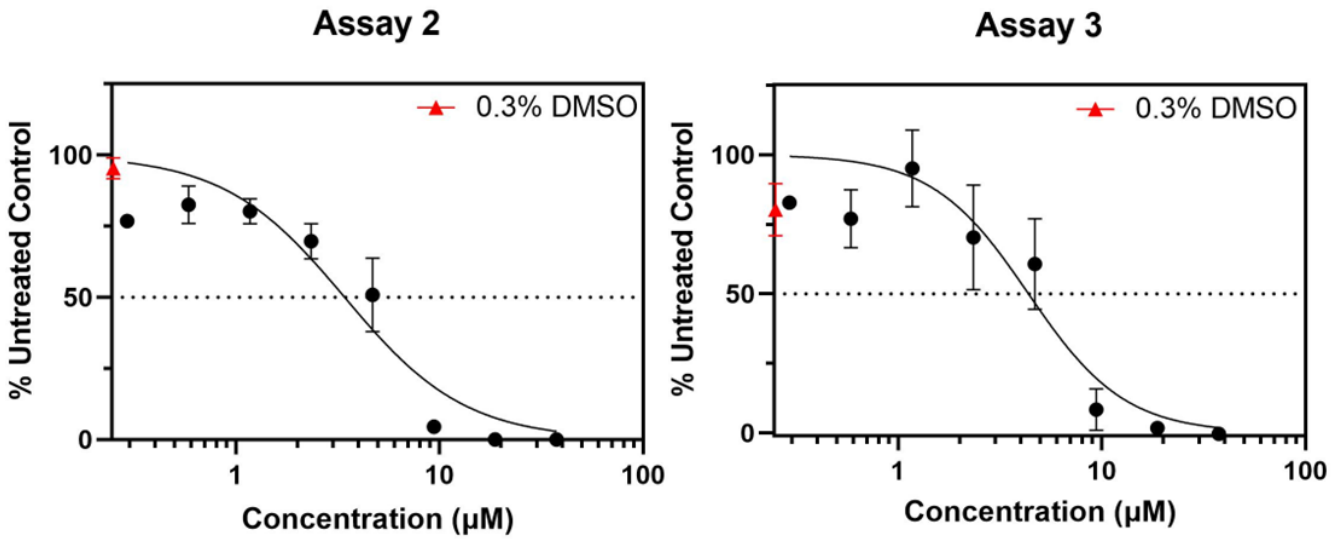
8.24.6 HMBC NMR spectrum



8.24.7 High resolution ESI-MS



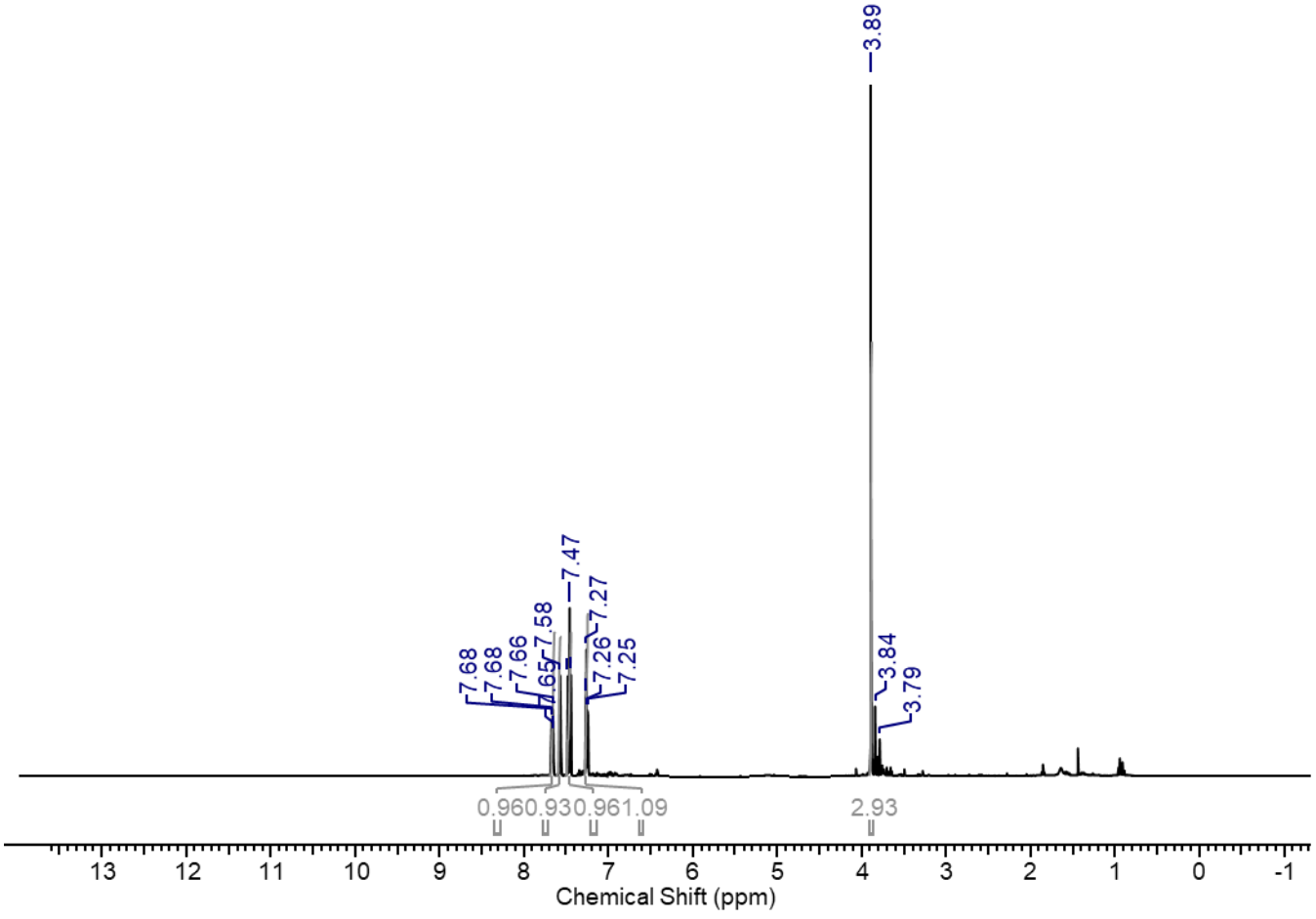
8.25 Triplicate data for anti-proliferative activity of tert-butyl (4-(5-(5,7-dimethoxy-4-oxo-4H-chromen-2-yl)-2,3-dimethoxyphenoxy)butyl)carbamate, 4.2.



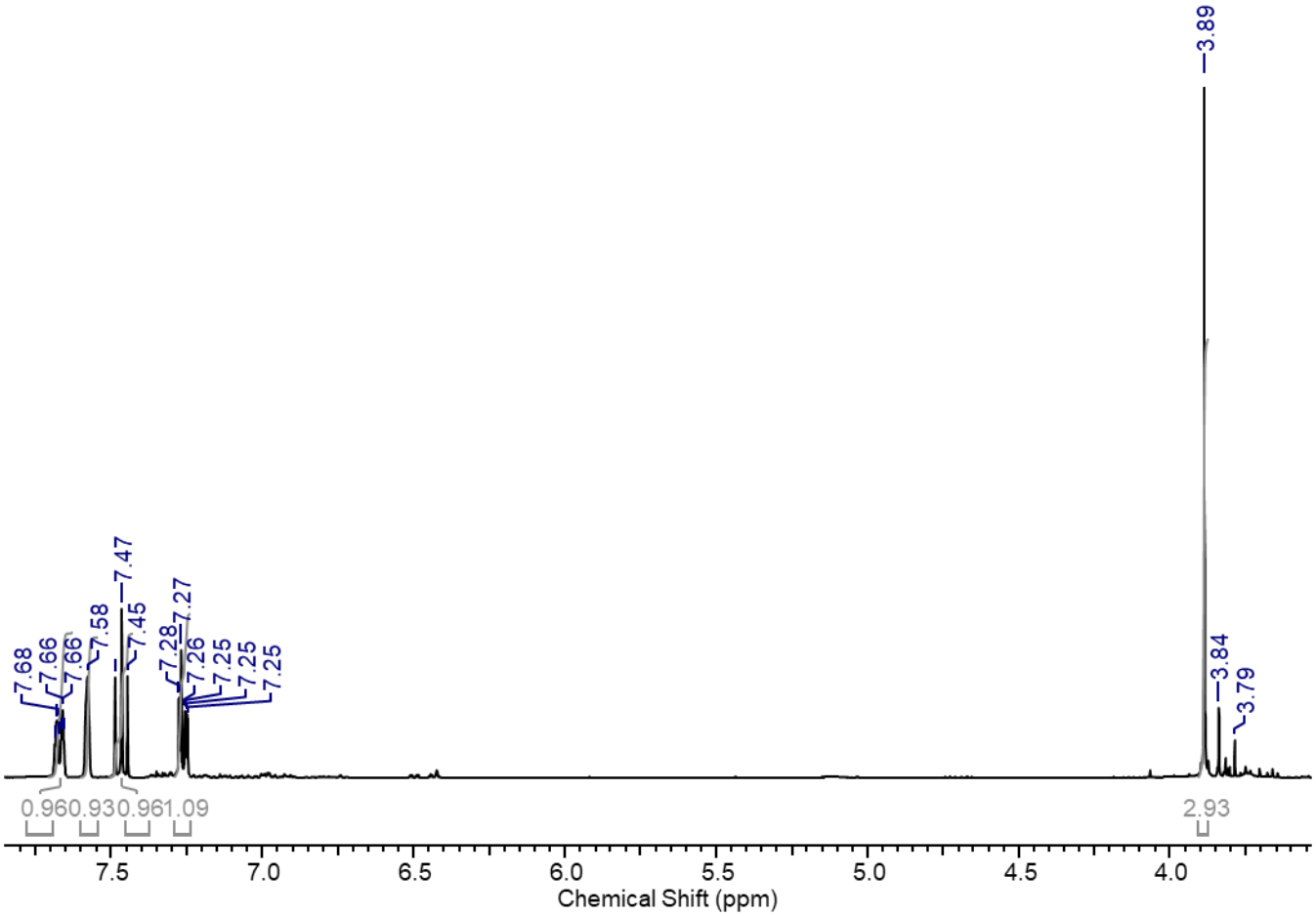
8.26 Characterisation data for synthesis of 2,2,2-trifluoro-1-(3-methoxyphenyl)ethan-1-one O-tosyl oxime.

8.26.1 Synthesis of 2,2,2-trifluoro-1-(3-methoxyphenyl)ethan-1-one

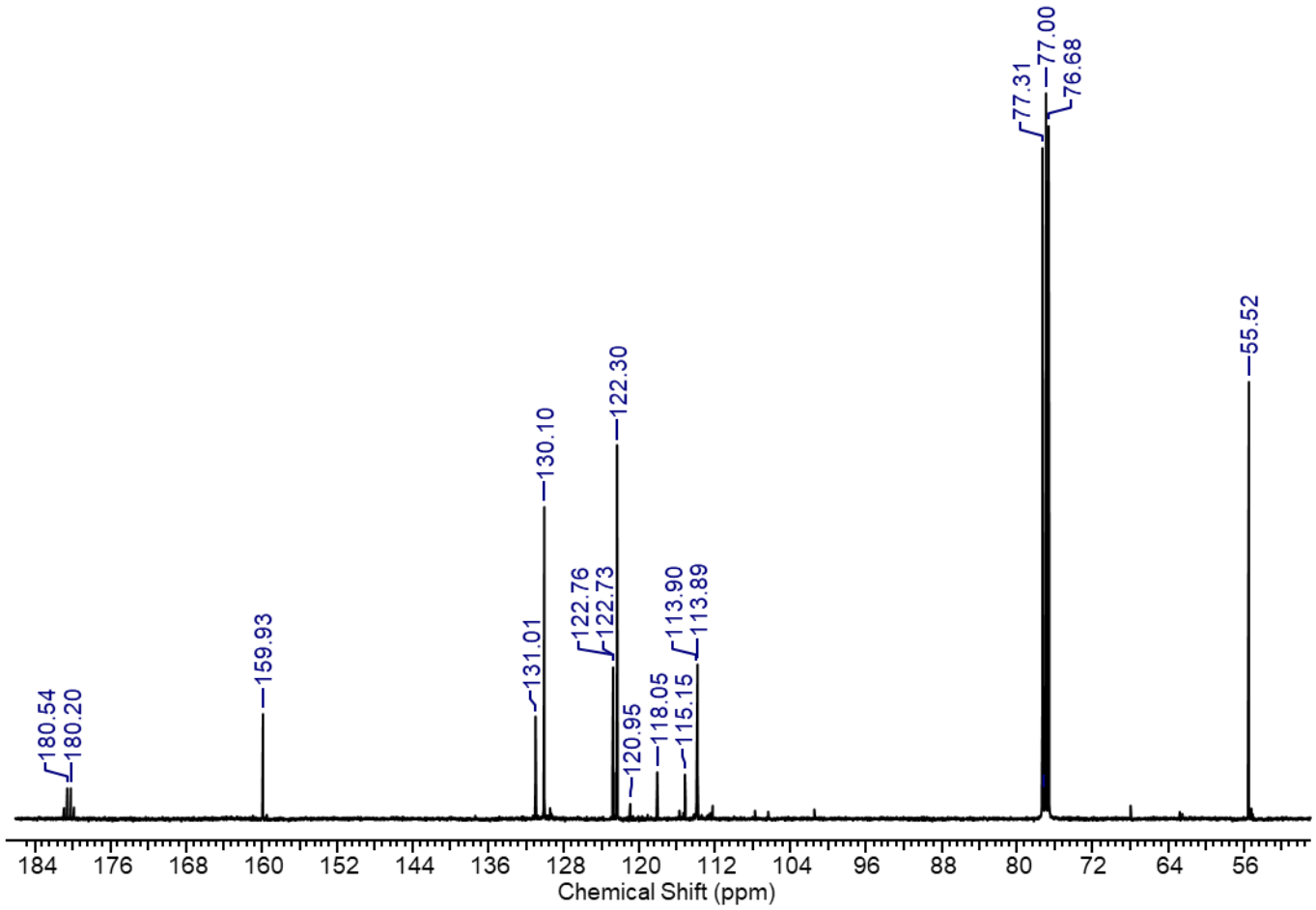
8.26.1.1 Extended ^{1}H NMR spectrum



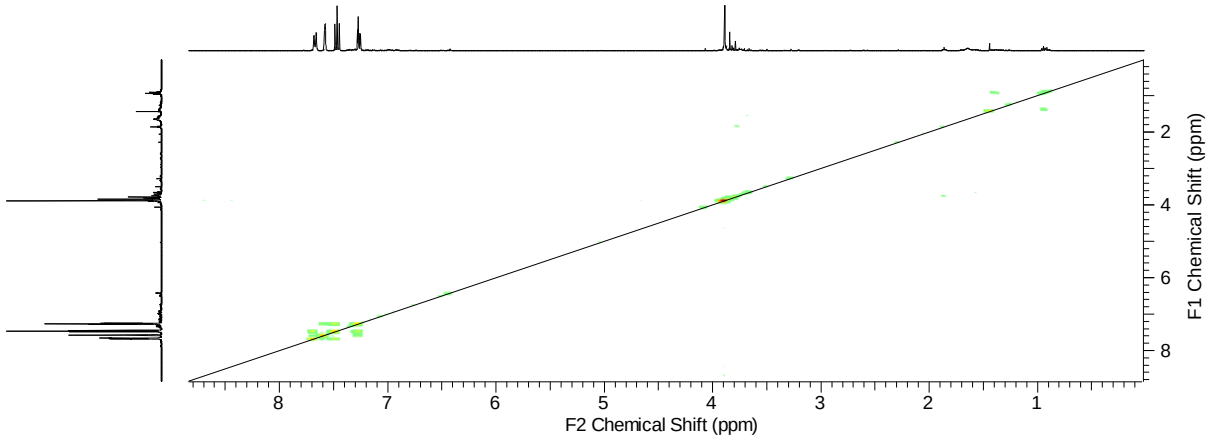
8.26.1.2 Magnified ^{1}H NMR spectrum



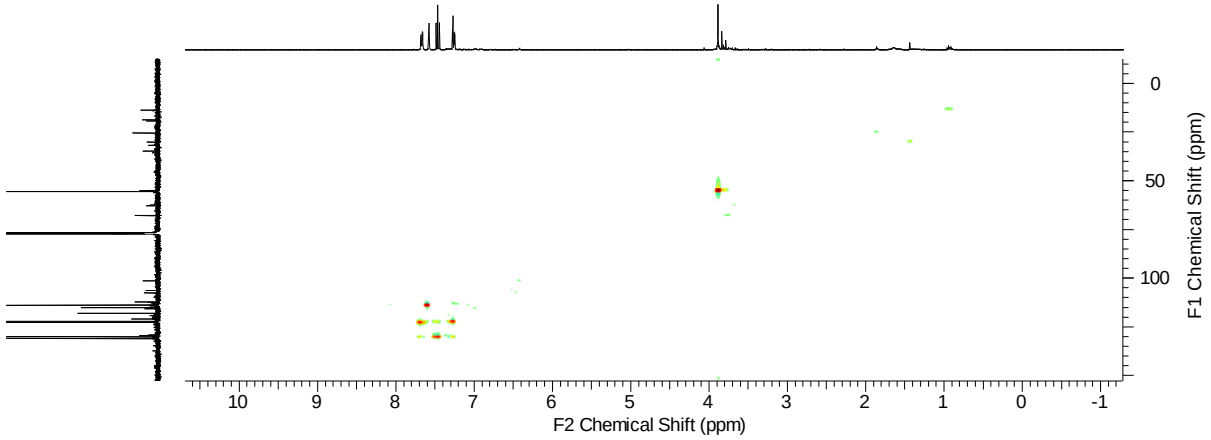
8.26.2 Magnified ^{13}C NMR spectrum



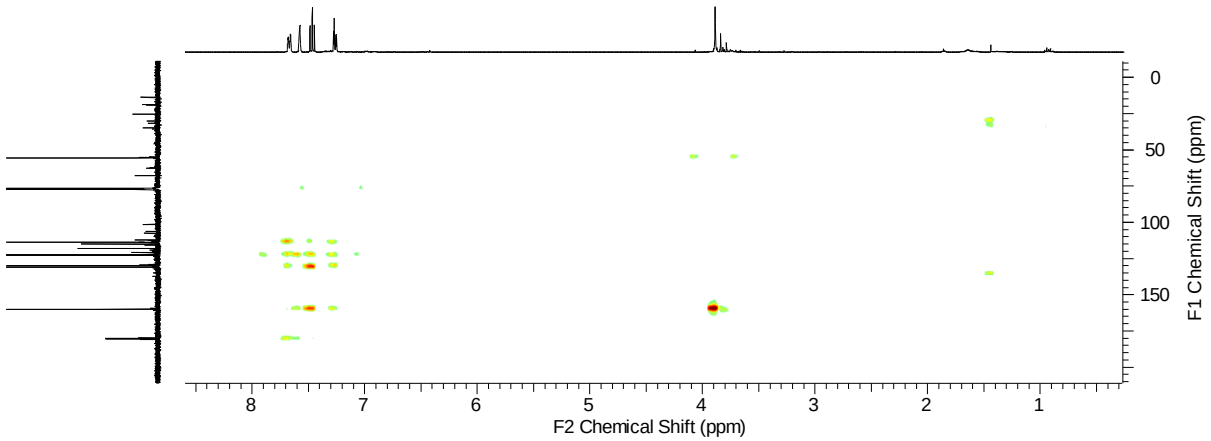
8.26.2.1 COSY NMR spectrum



8.26.2.2 HSQC NMR spectrum

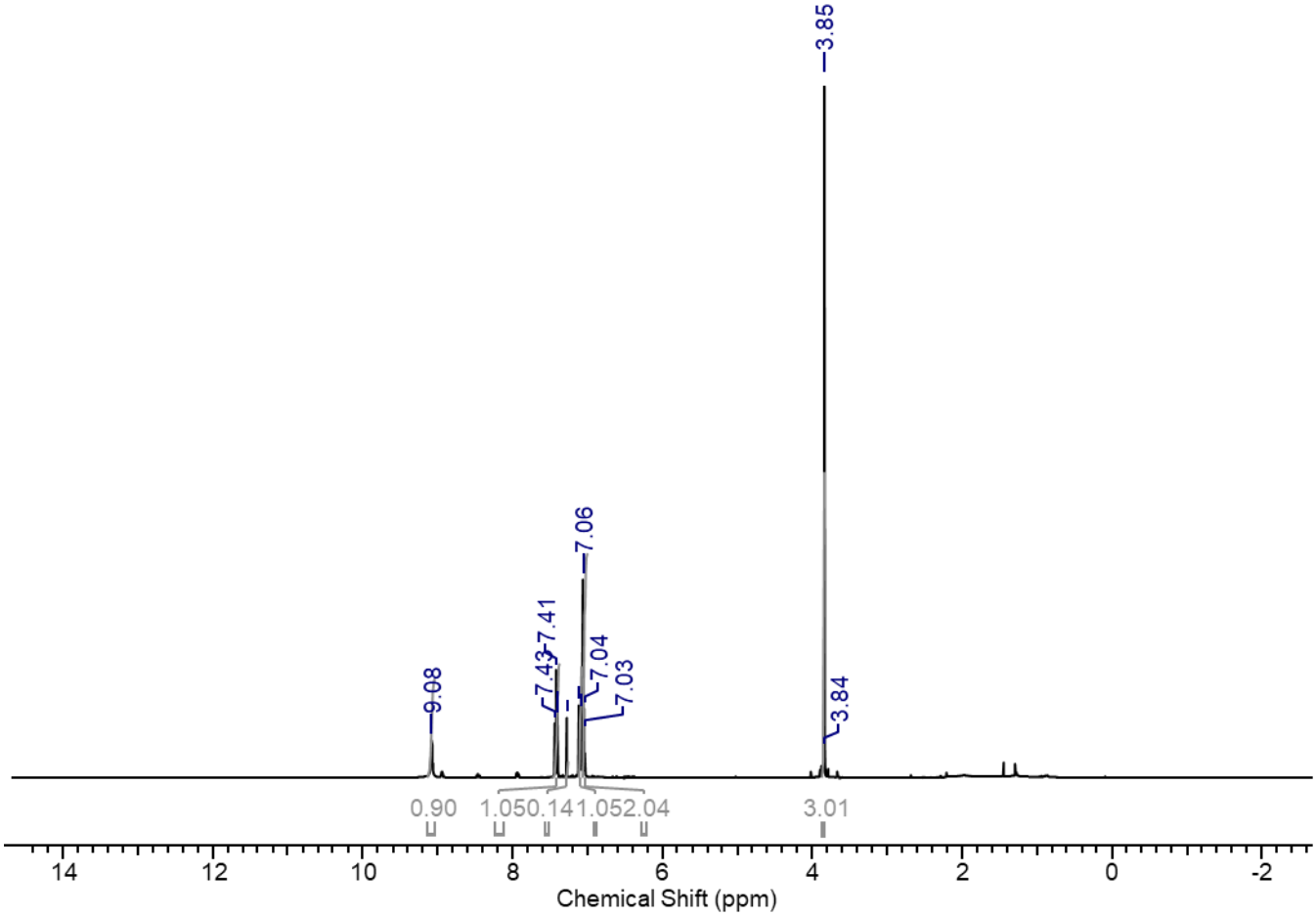


8.26.2.3 HMBC NMR spectrum

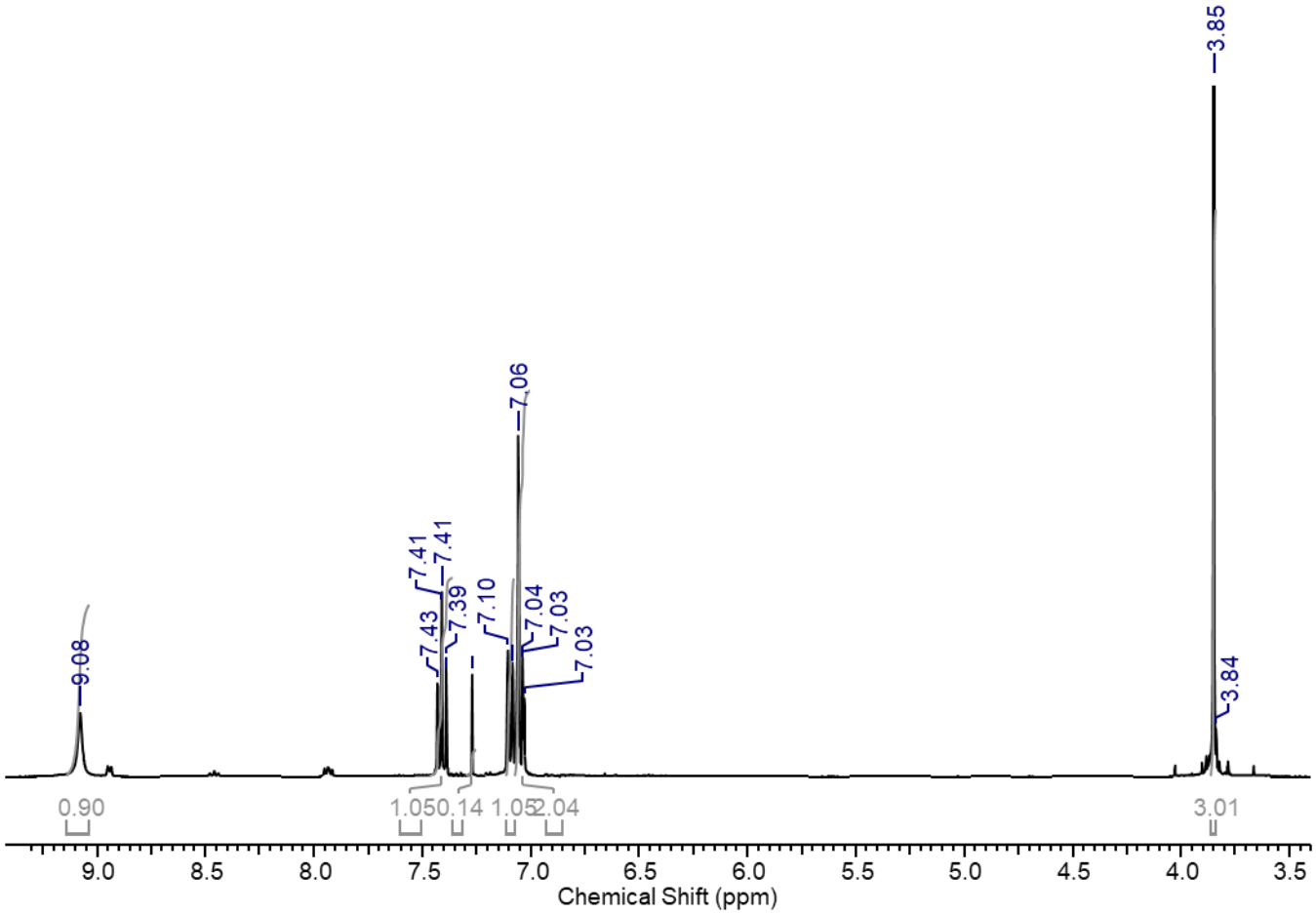


8.26.3 Synthesis of 2,2,2-trifluoro-1-(3-methoxyphenyl)ethan-1-one oxime.

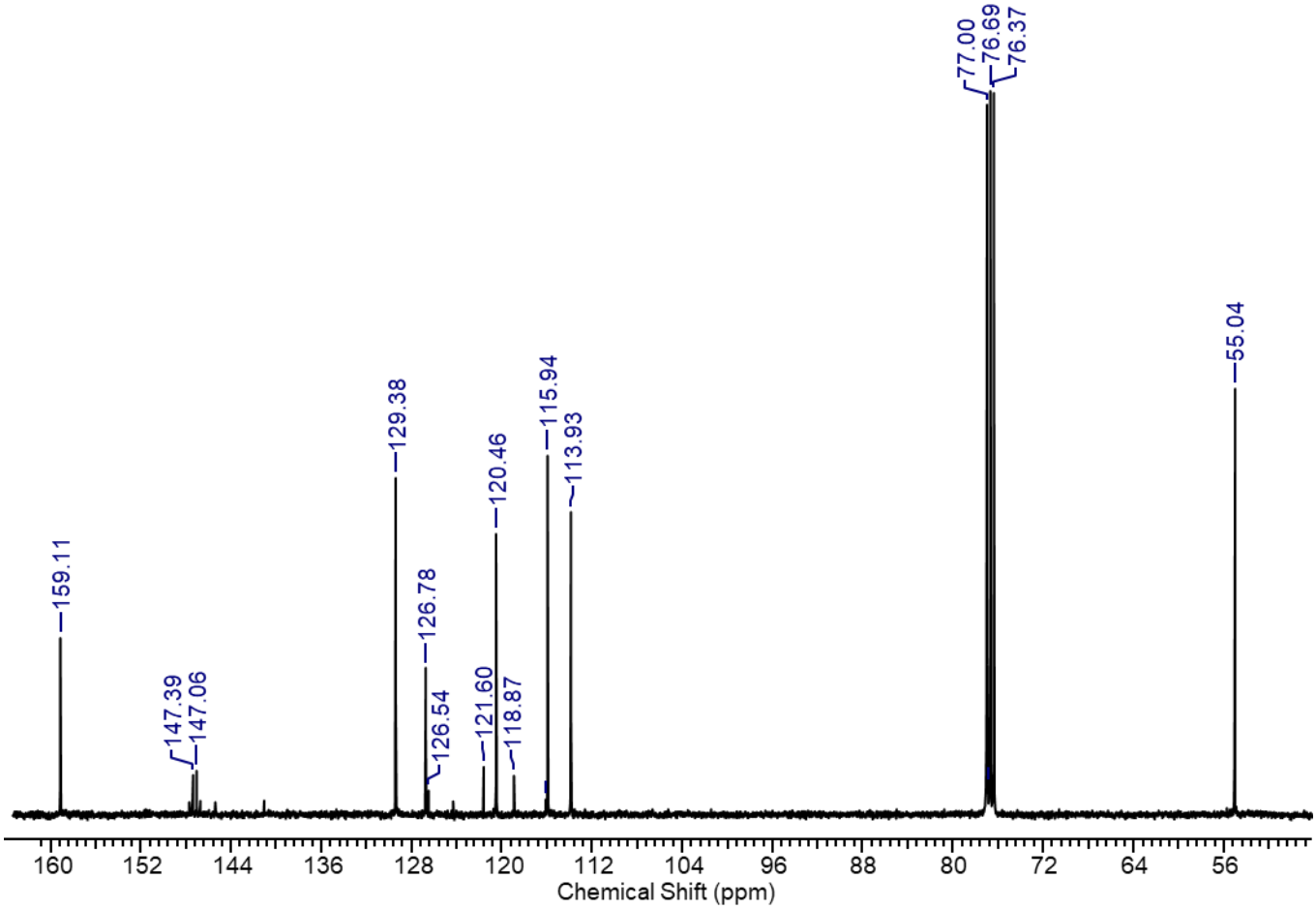
8.26.3.1 Extended ^{1}H NMR spectrum



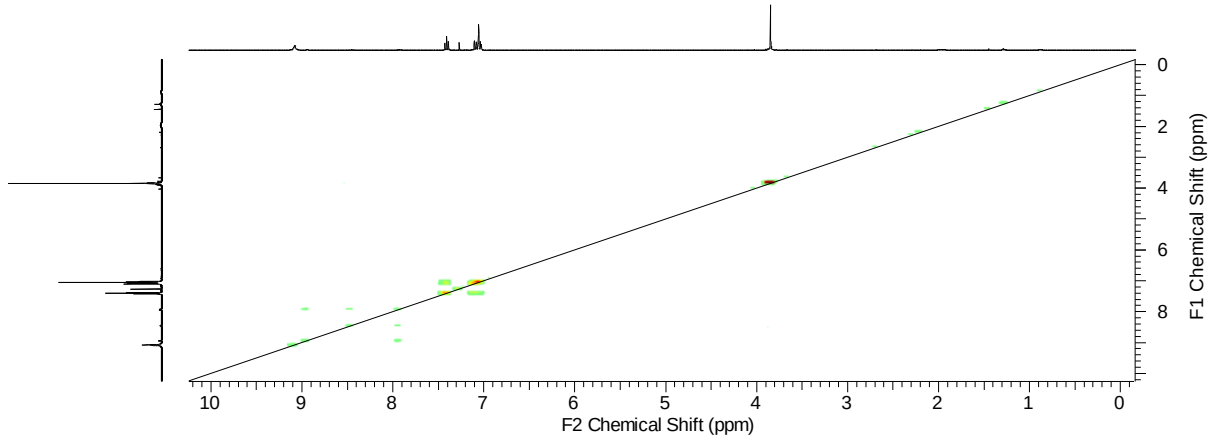
8.26.3.2 Magnified ^{1}H NMR spectrum



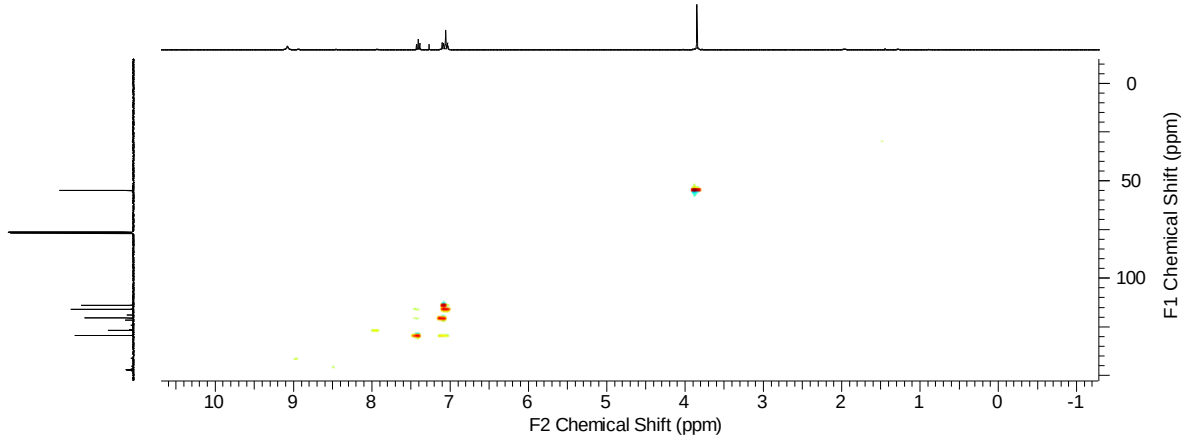
8.26.3.3 Magnified ^{13}C NMR spectrum



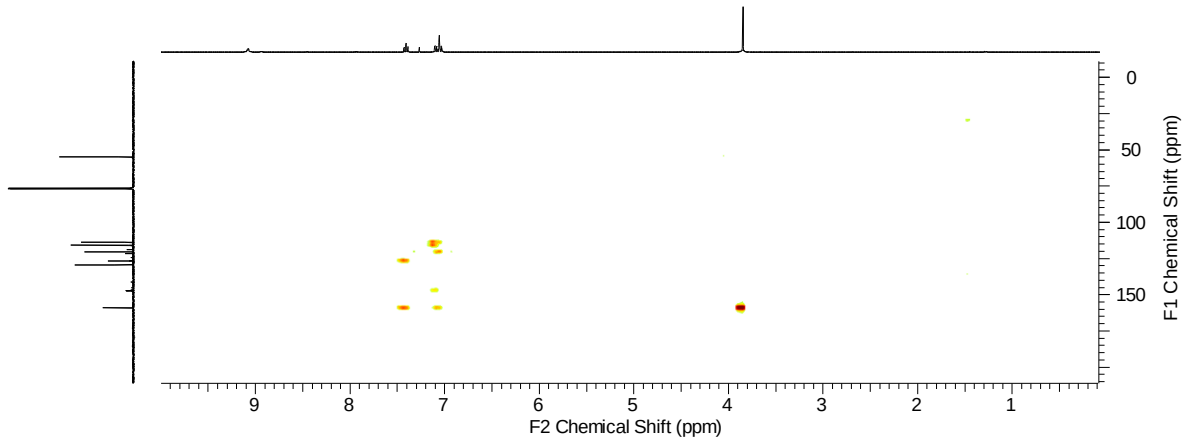
8.26.3.4 COSY NMR spectrum



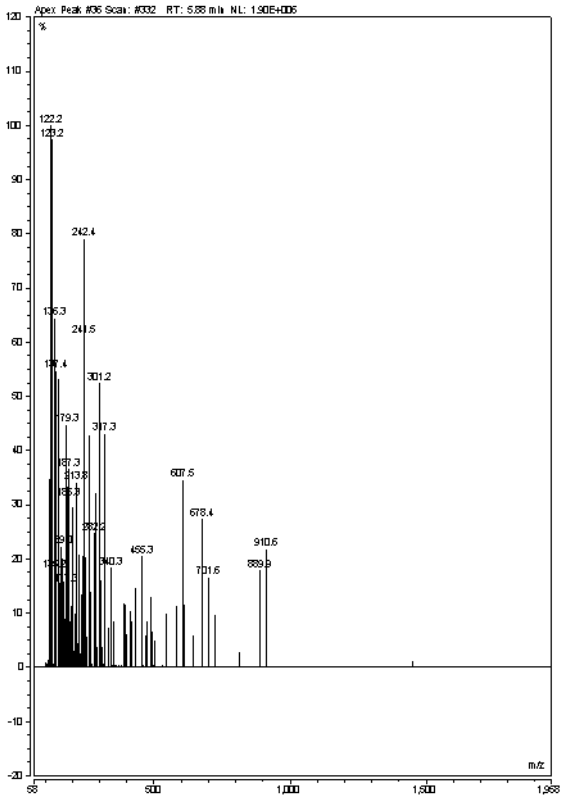
8.26.3.5 HSQC NMR spectrum



8.26.3.6 HMBC NMR spectrum

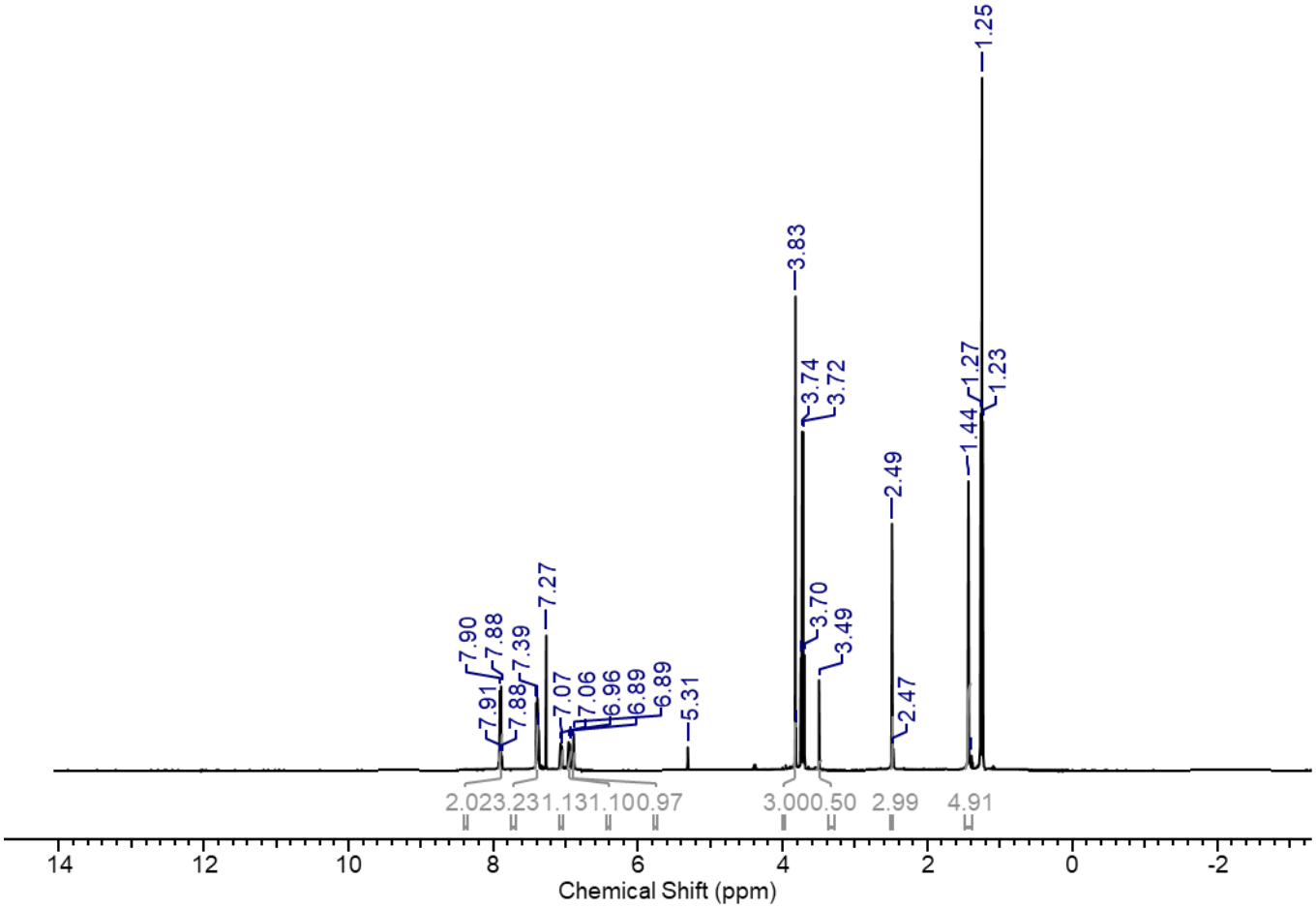


8.26.3.7 ESI-MS

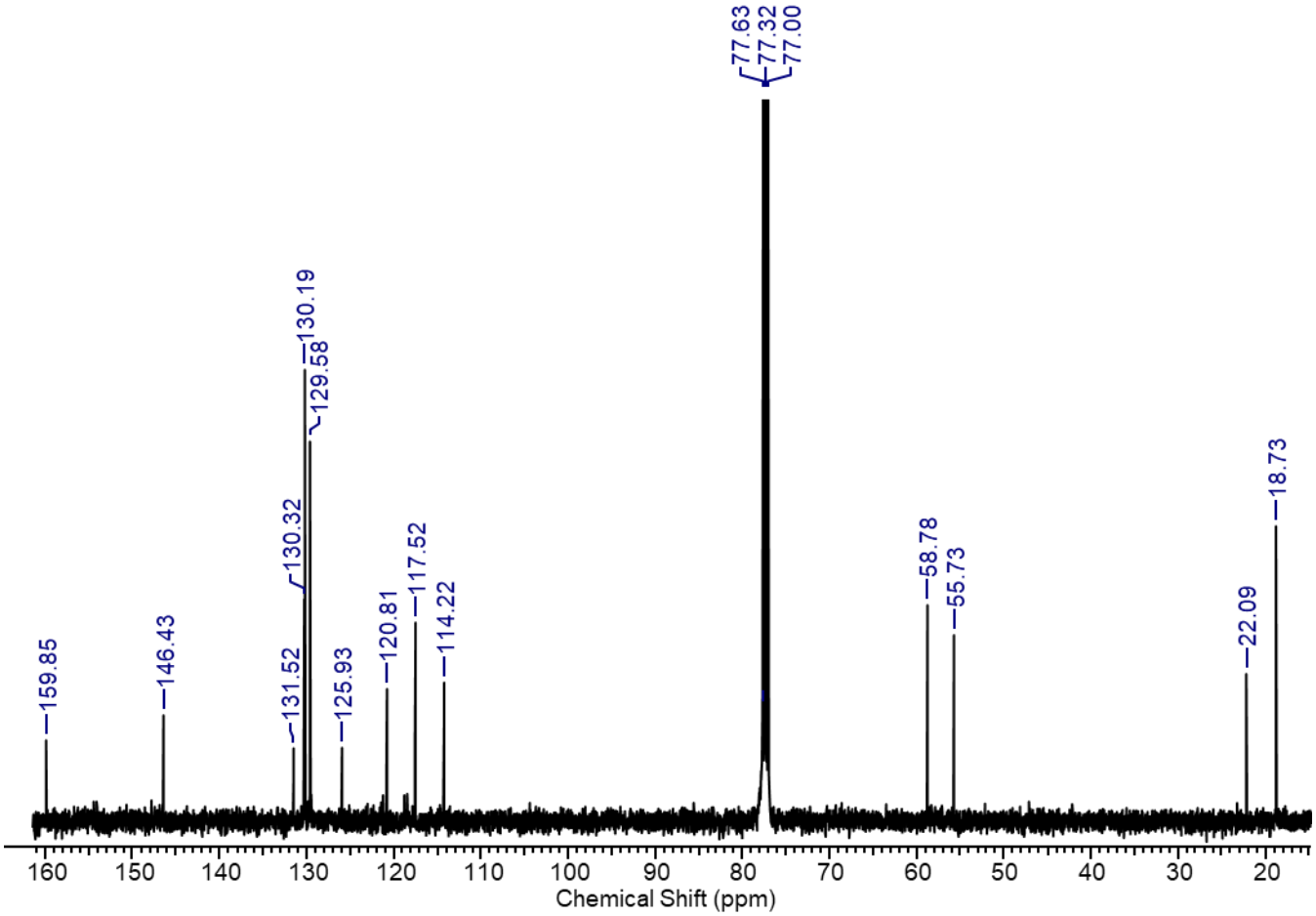


8.26.4 Characterisation data for synthesis of 2,2,2-trifluoro-1-(3-methoxyphenyl)ethan-1-one O-tosyl oxime.

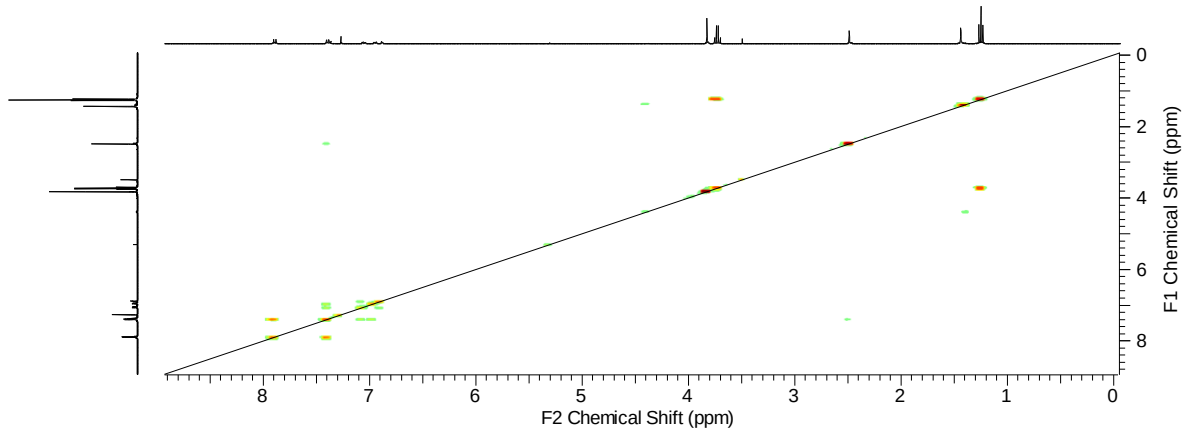
8.26.4.1 Extended ^{1}H NMR spectrum



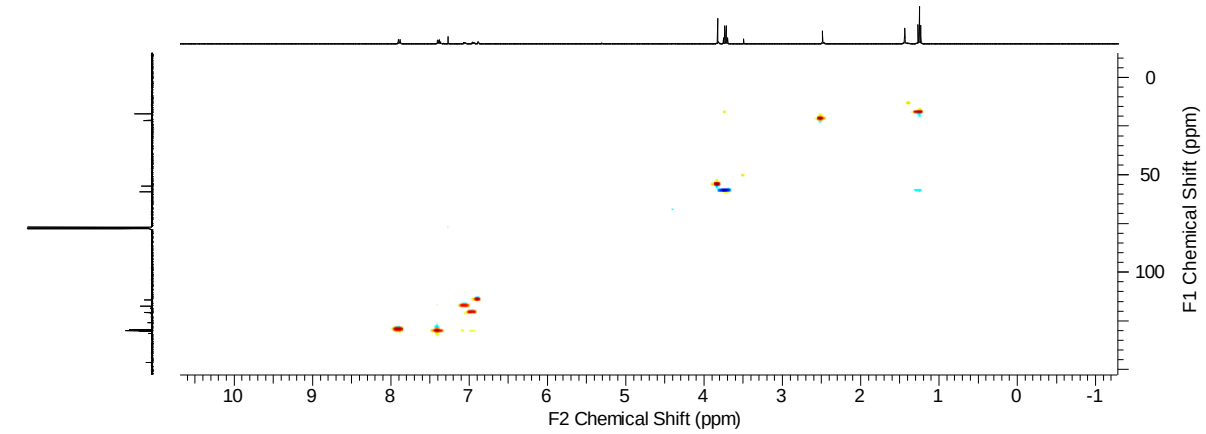
8.26.4.2 Magnified ^{13}C NMR spectrum



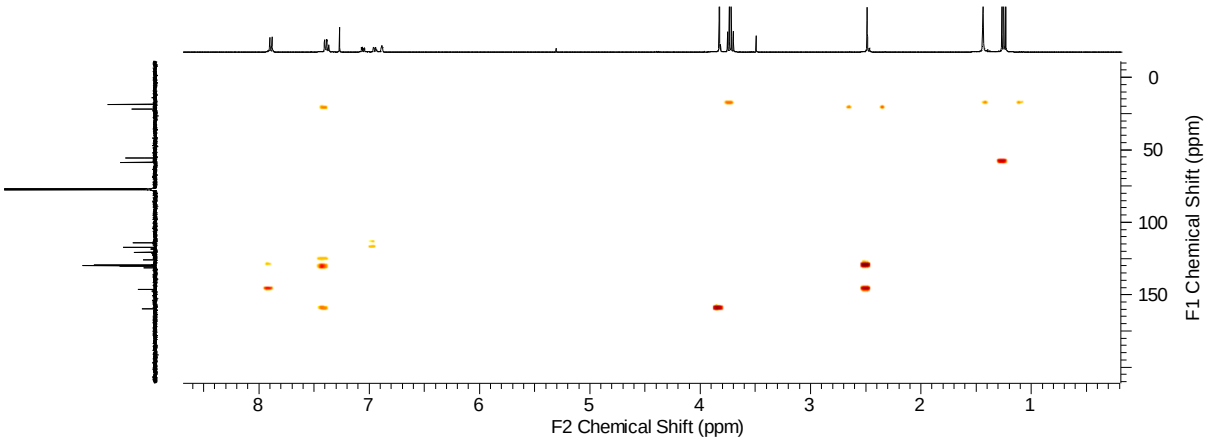
8.26.4.3 COSY NMR spectrum



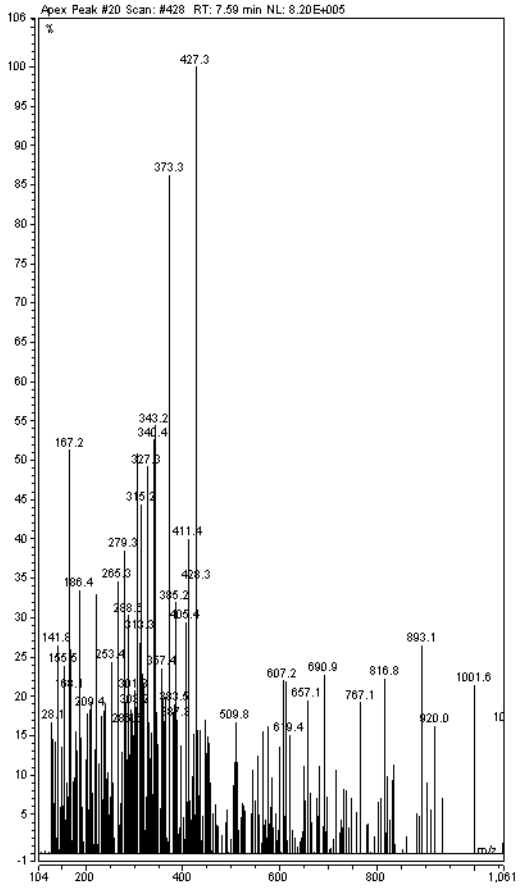
8.26.4.4 HSQC NMR spectrum



8.26.4.5 HMBC NMR spectrum

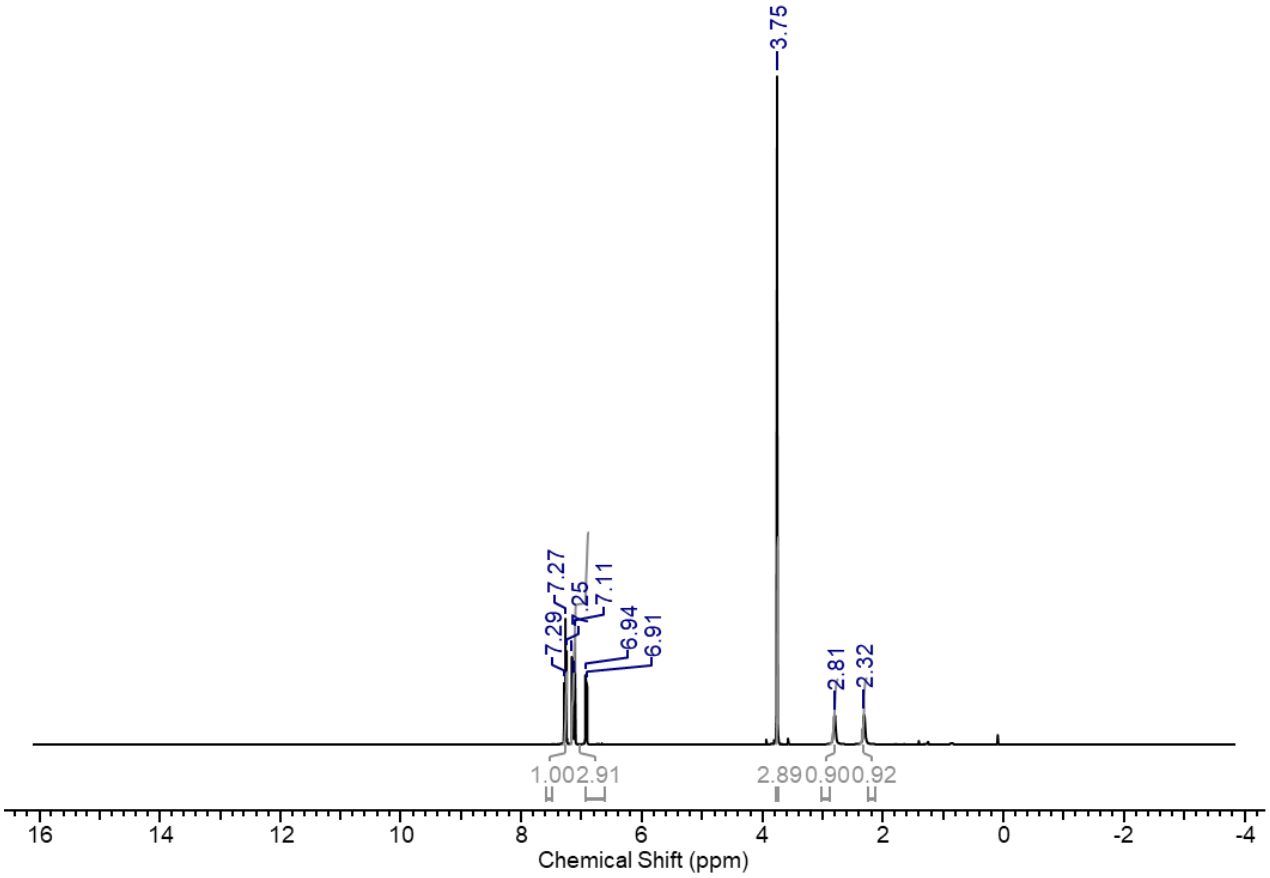


8.26.4.6 ESI-MS spectrum

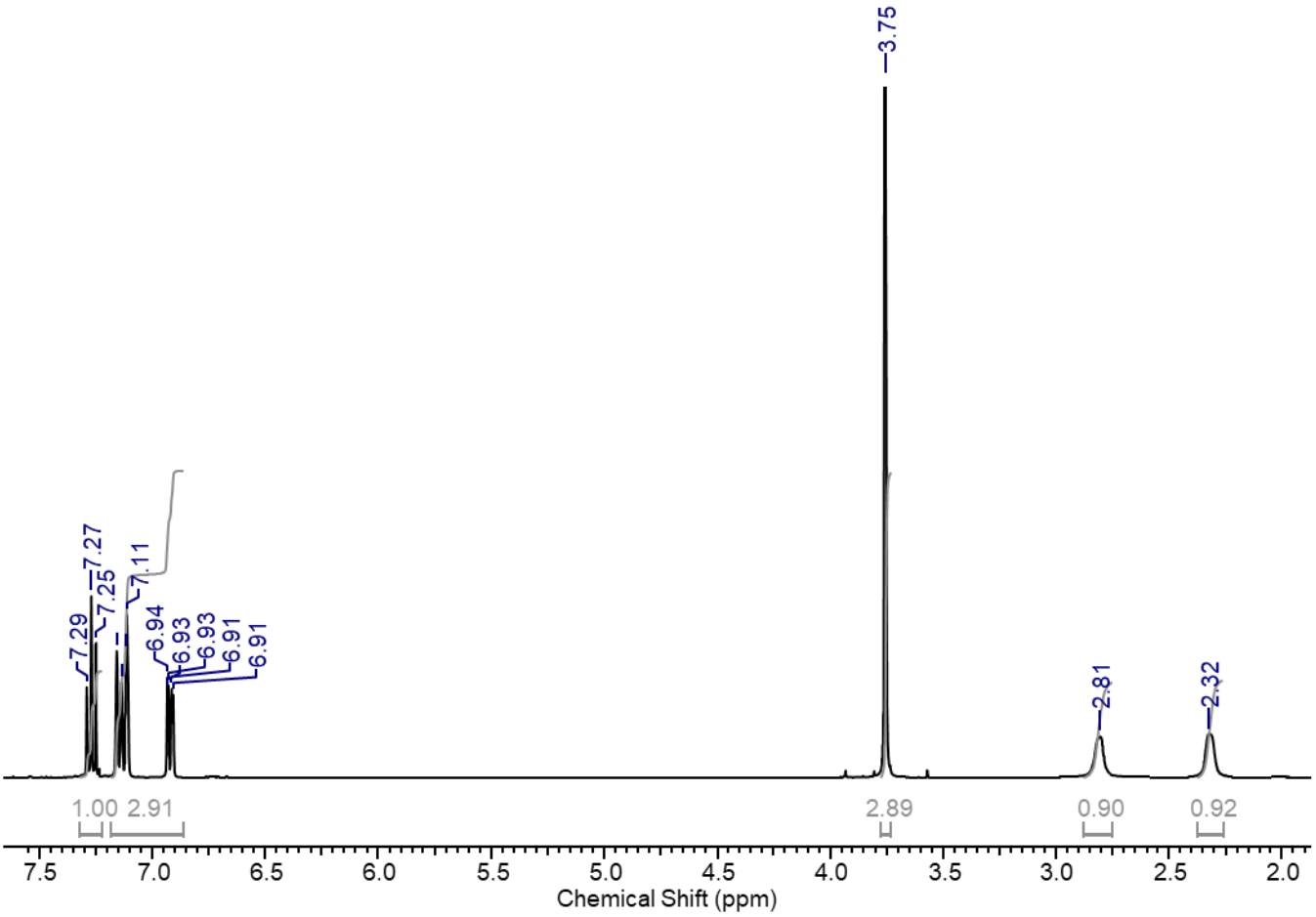


8.27 Characterisation data for synthesis of 3-(3-methoxyphenyl)-3-(trifluoromethyl)diaziridine.

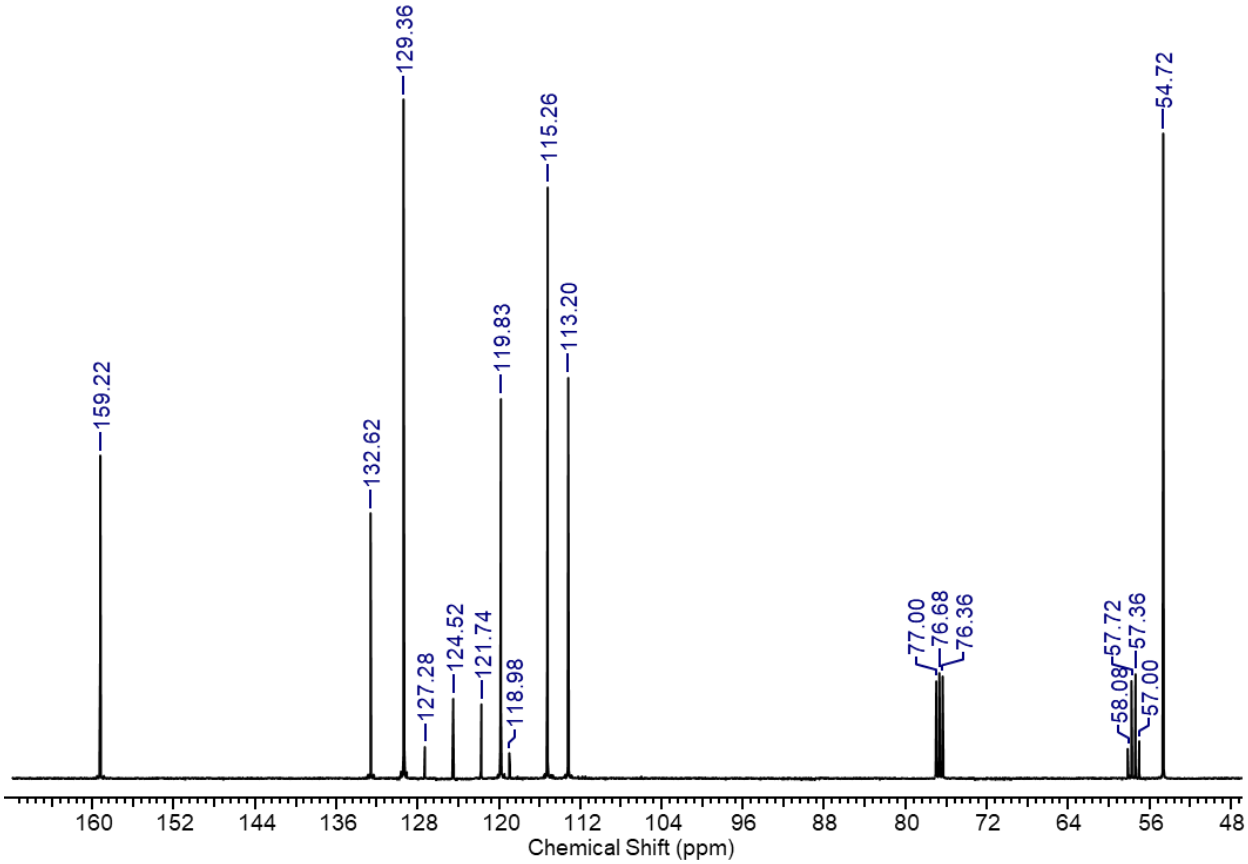
8.27.1 Extended ^{1}H NMR spectrum



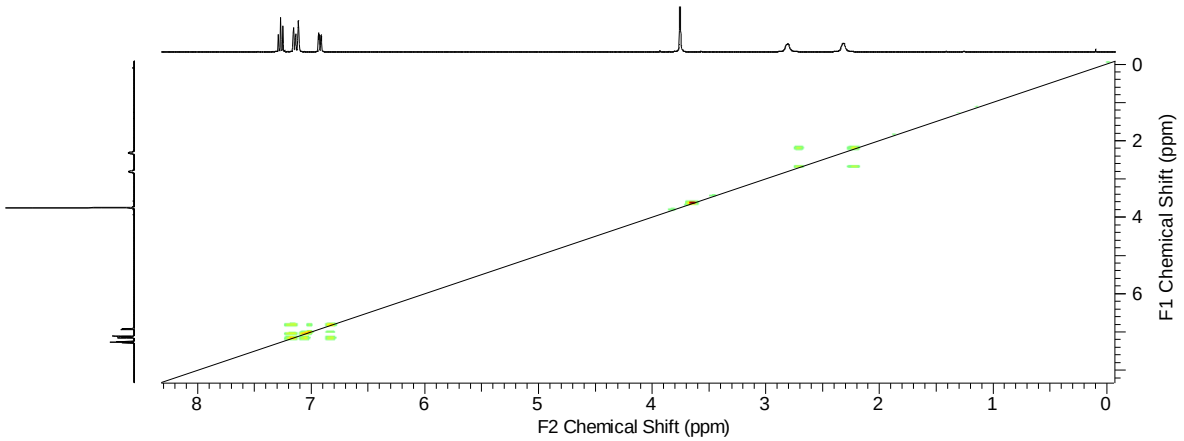
8.27.2 Magnified ^{1}H NMR spectrum



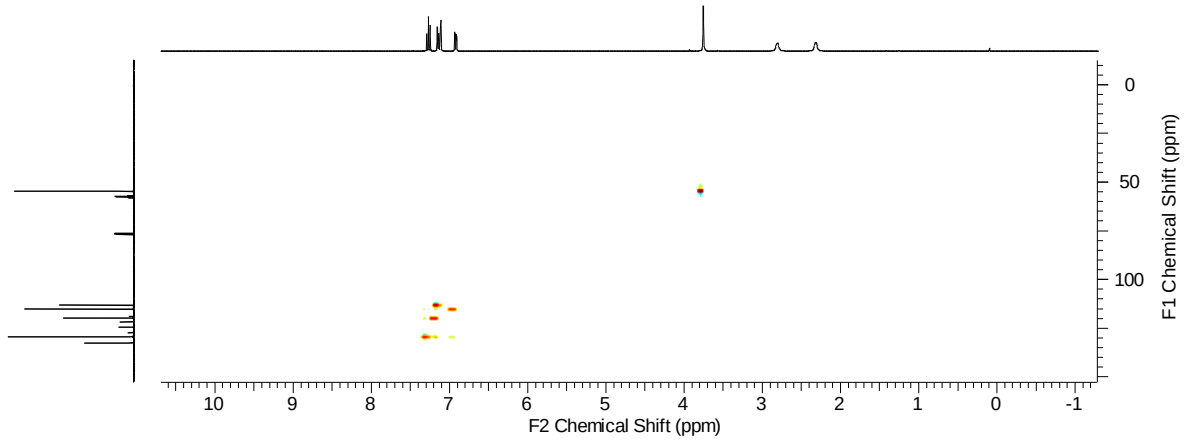
8.27.3 Magnified ^{13}C NMR spectrum



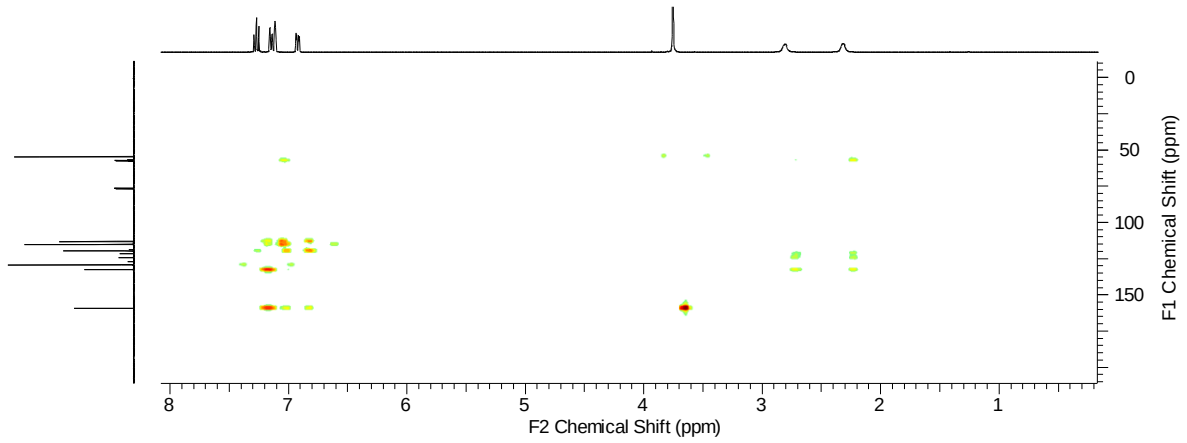
8.27.4 COSY NMR spectrum



8.27.5 HSQC NMR spectrum

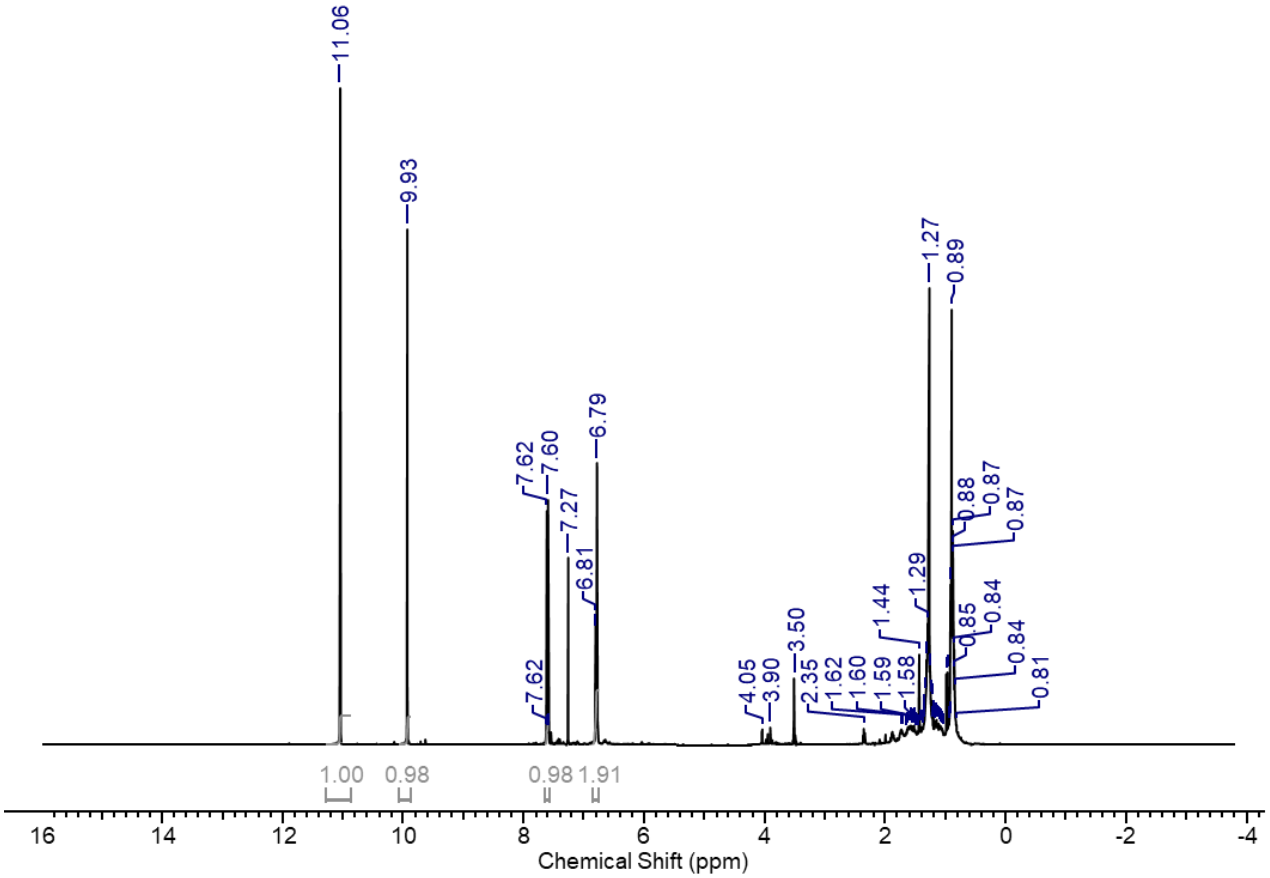


8.27.6 HMBC NMR spectrum

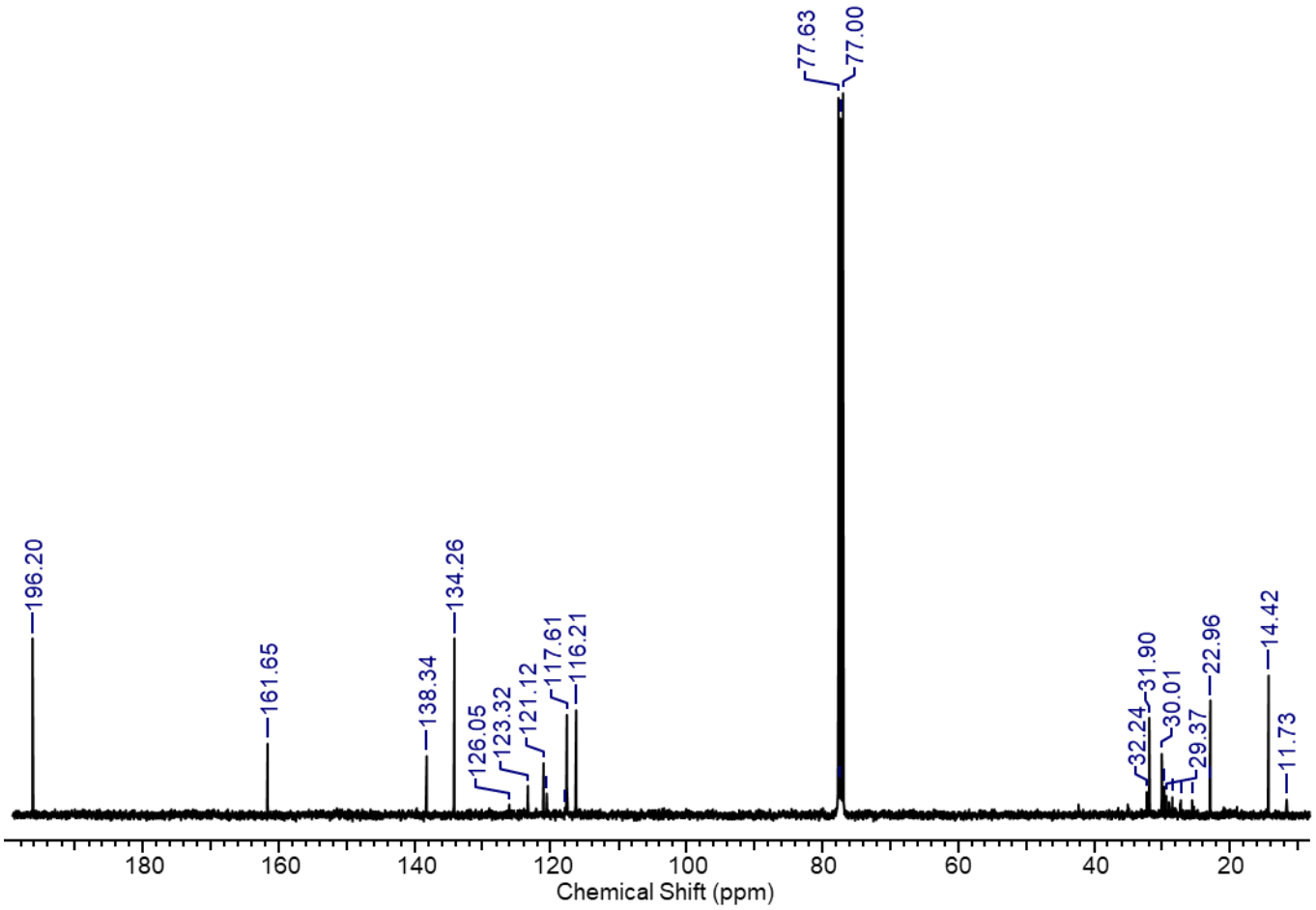


8.28 Characterisation data for synthesis of 2-hydroxy-4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzaldehyde.

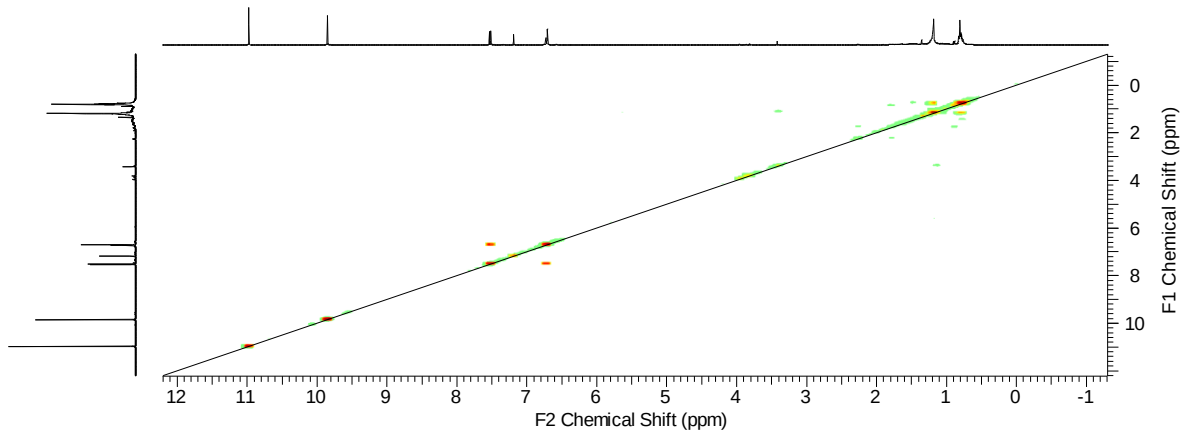
8.28.1 Extended ^{1}H NMR spectrum



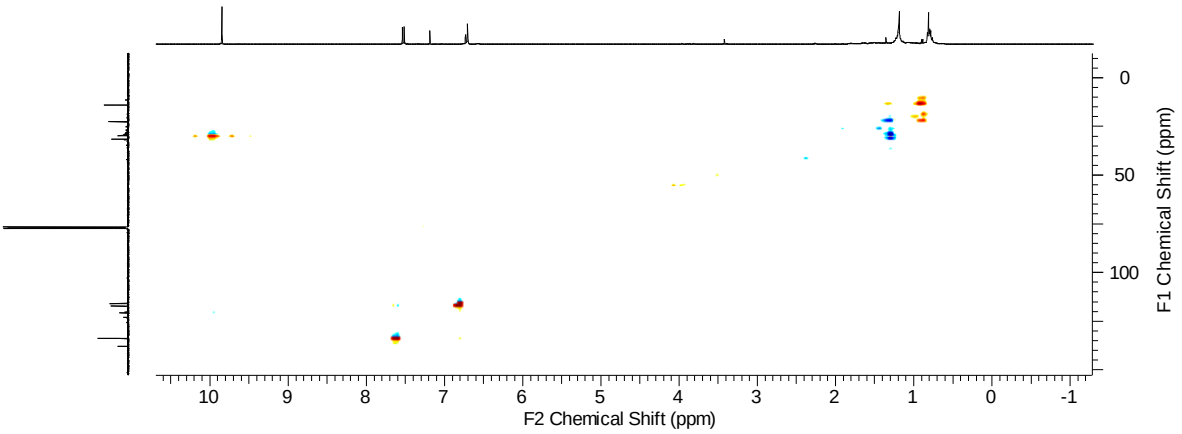
8.28.2 Magnified ^{13}C NMR spectrum



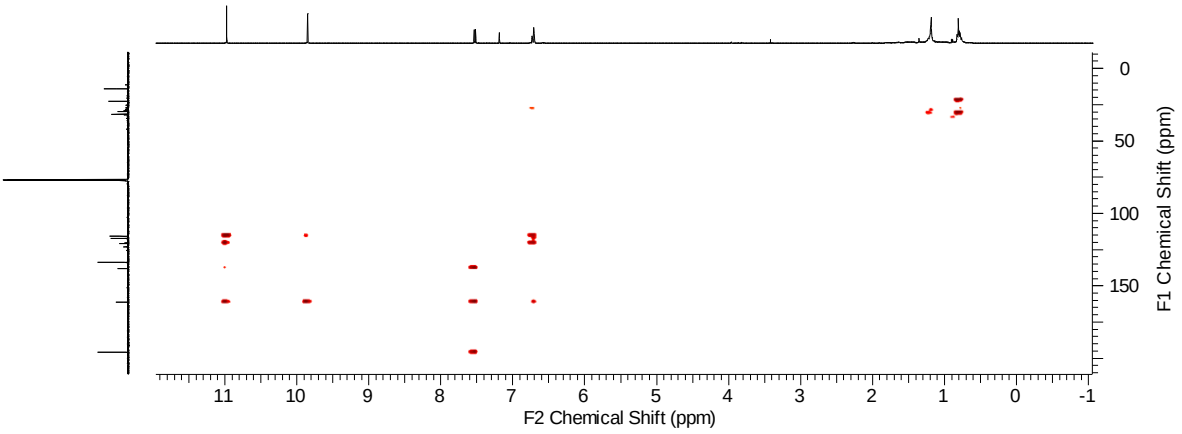
8.28.3 COSY NMR spectrum



8.28.4 HSQC NMR spectrum

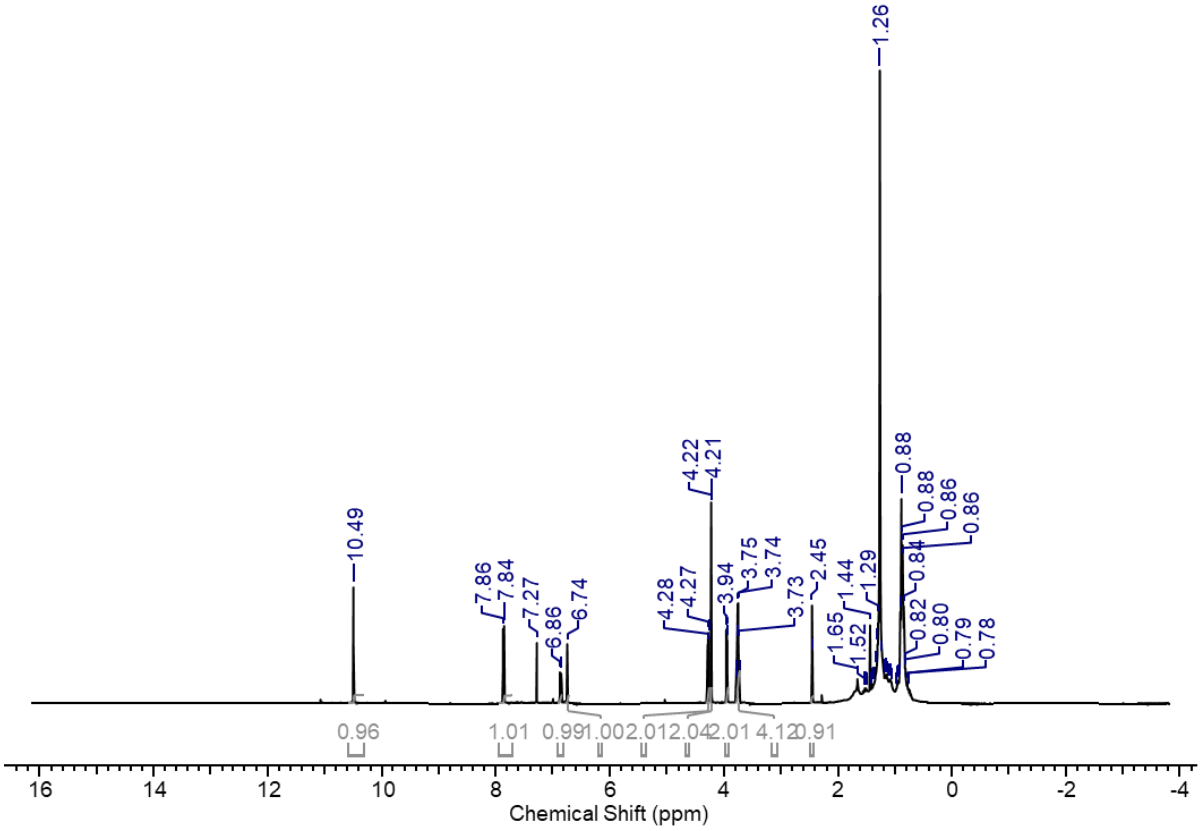


8.28.5 HMBC NMR spectrum

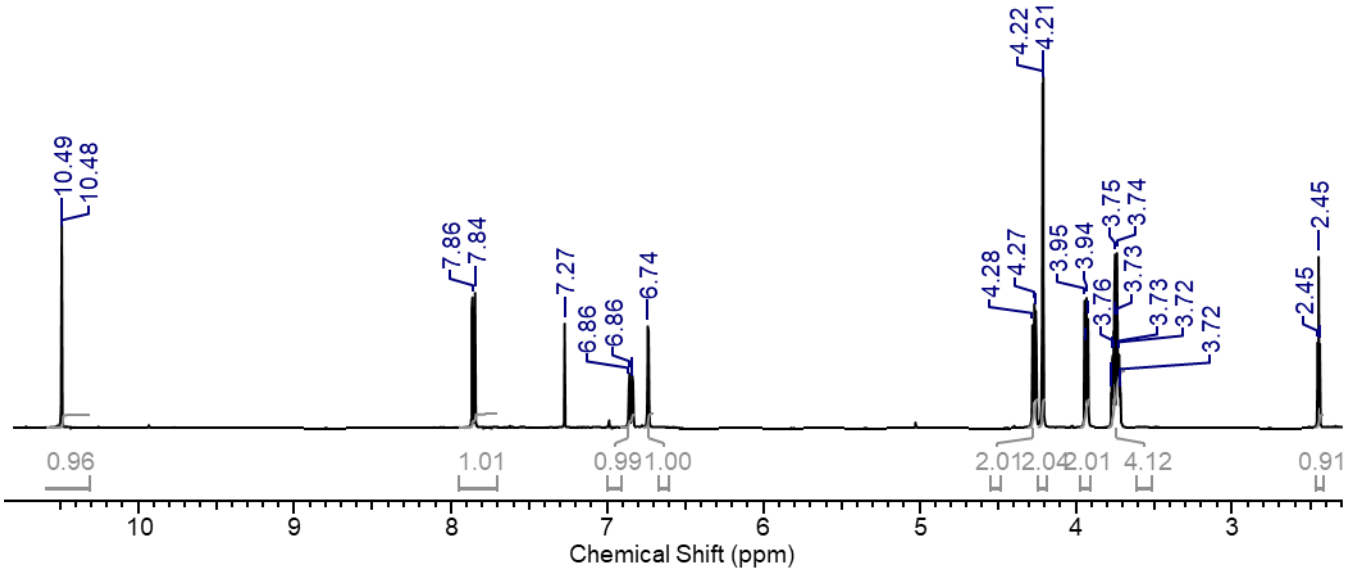


8.29 Characterisation data for the synthesis of 2-(2-(2-(prop-2-yn-1-yloxy)ethoxy)ethoxy)-4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzaldehyde.

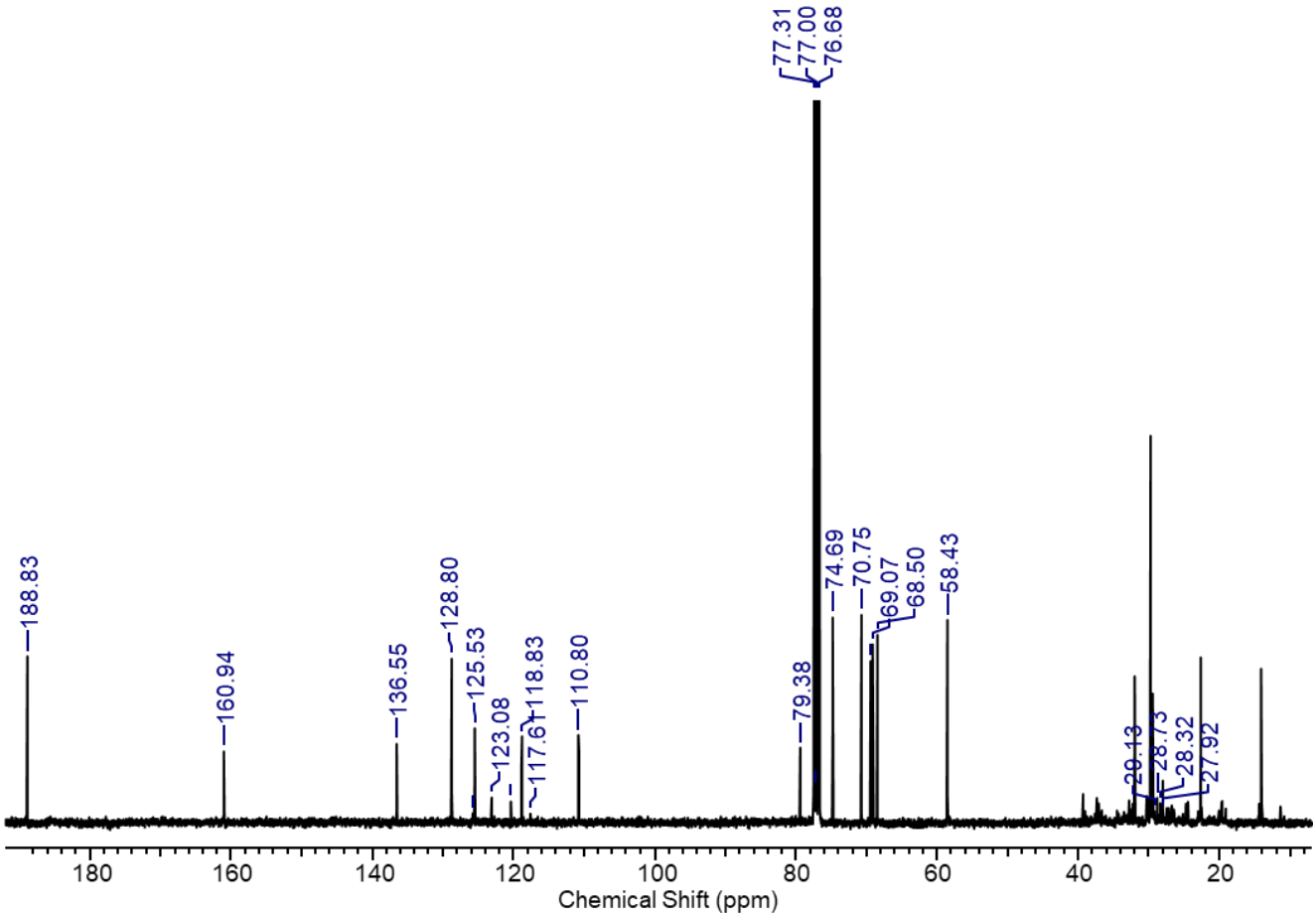
8.29.1 Extended ^{1}H NMR spectrum



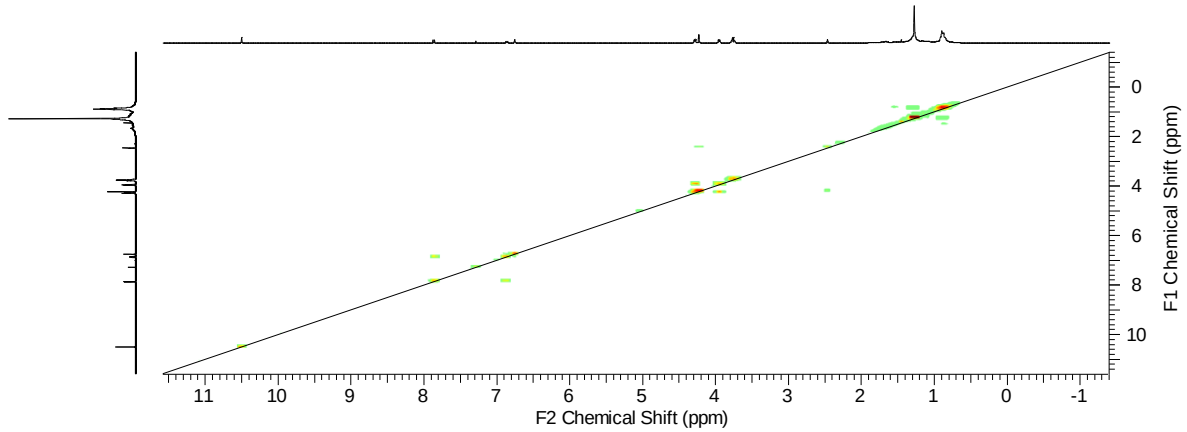
8.29.2 Magnified ^{1}H NMR spectrum



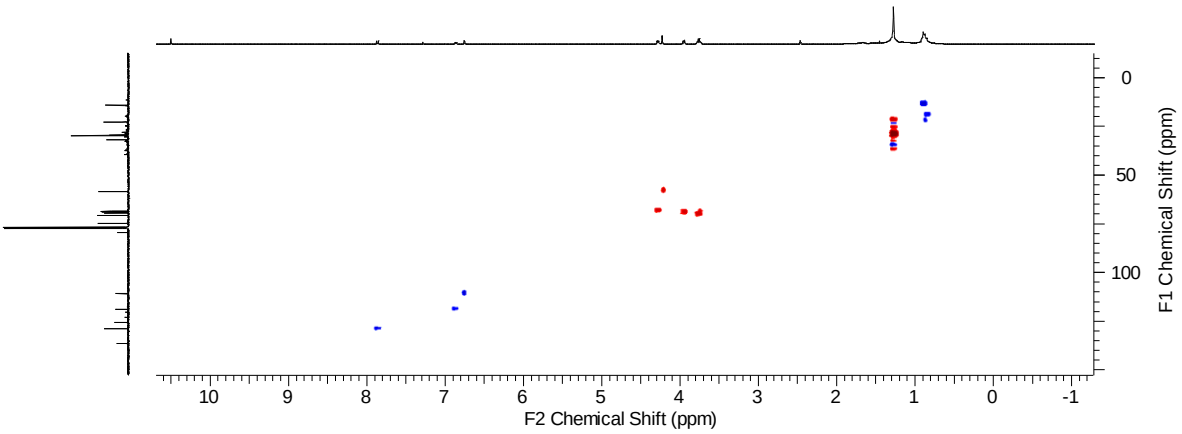
8.29.3 Magnified ^{13}C NMR spectrum



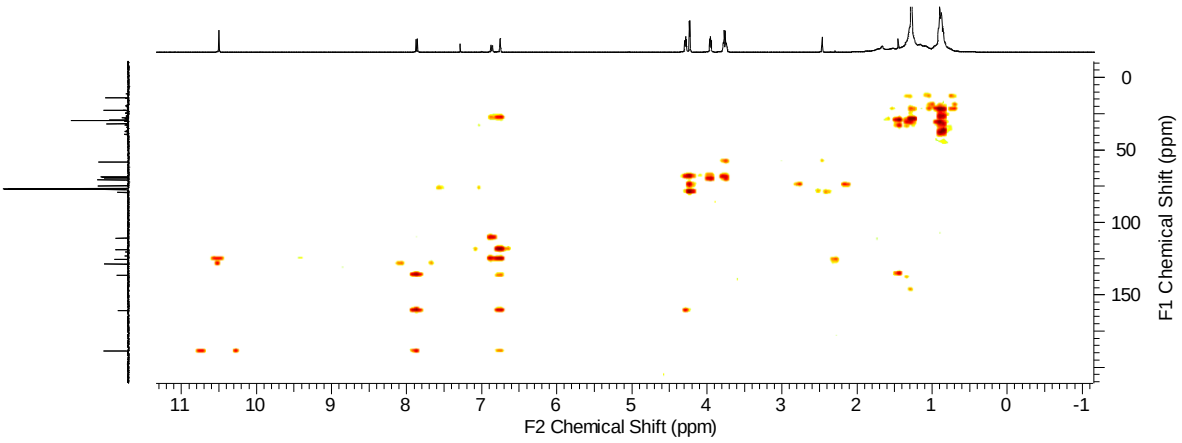
8.29.4 COSY NMR spectrum



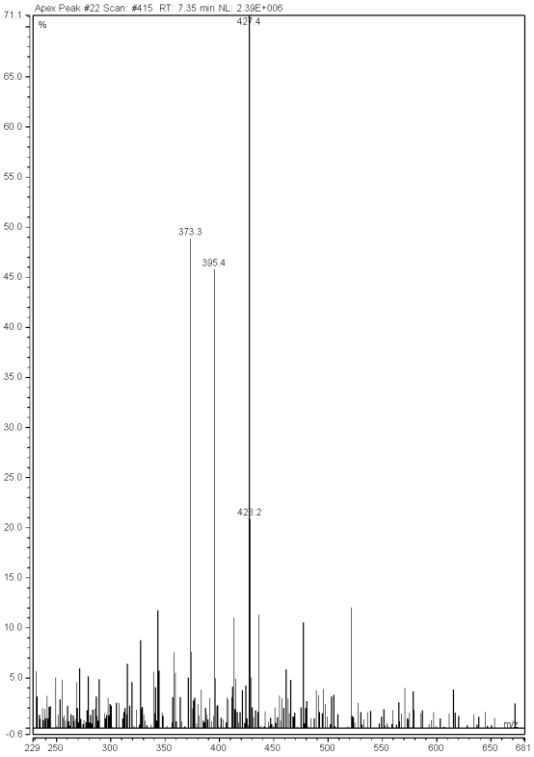
8.29.5 HSQC NMR spectrum



8.29.6 HMBC NMR spectrum

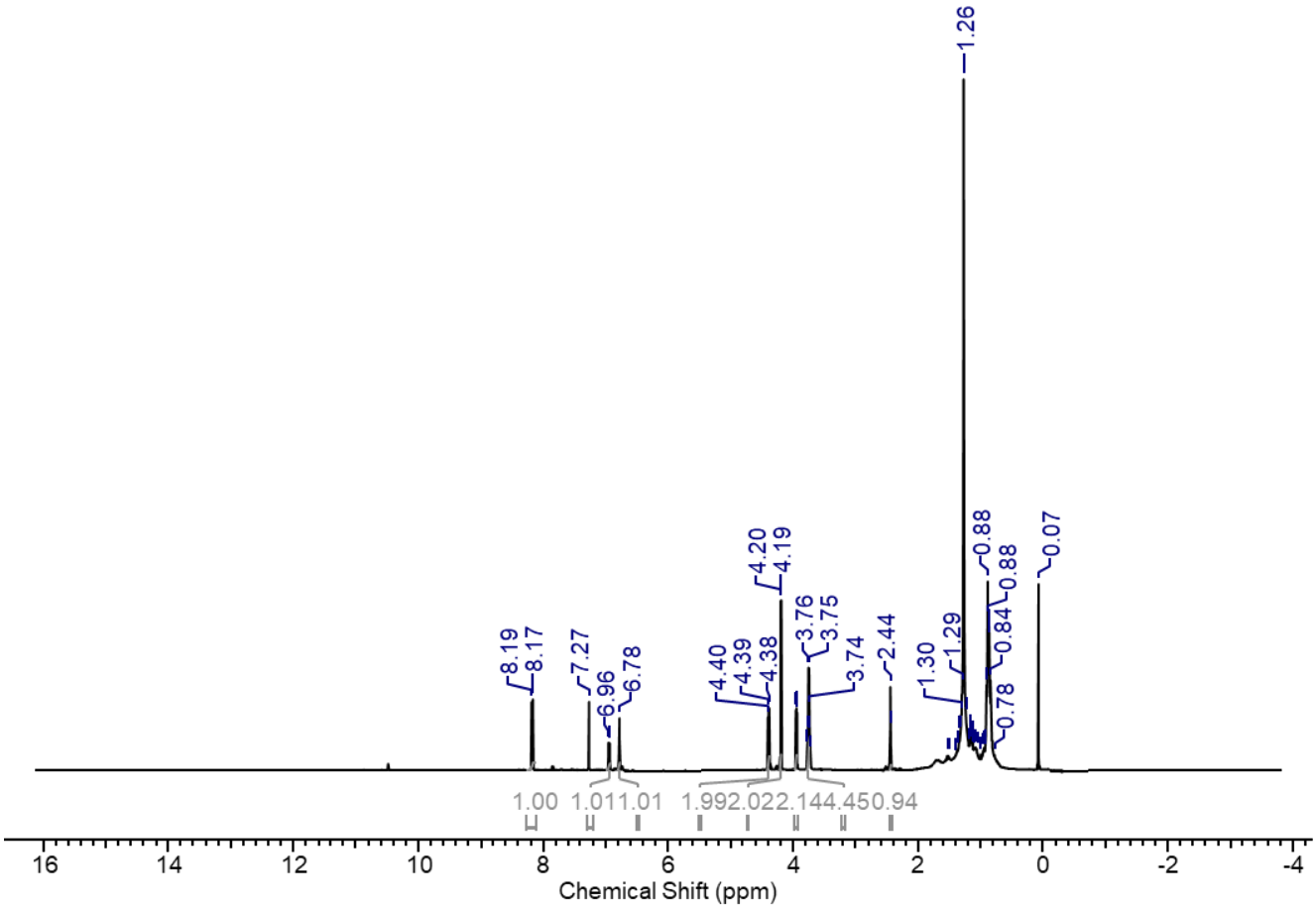


8.29.7 ESI-MS spectrum

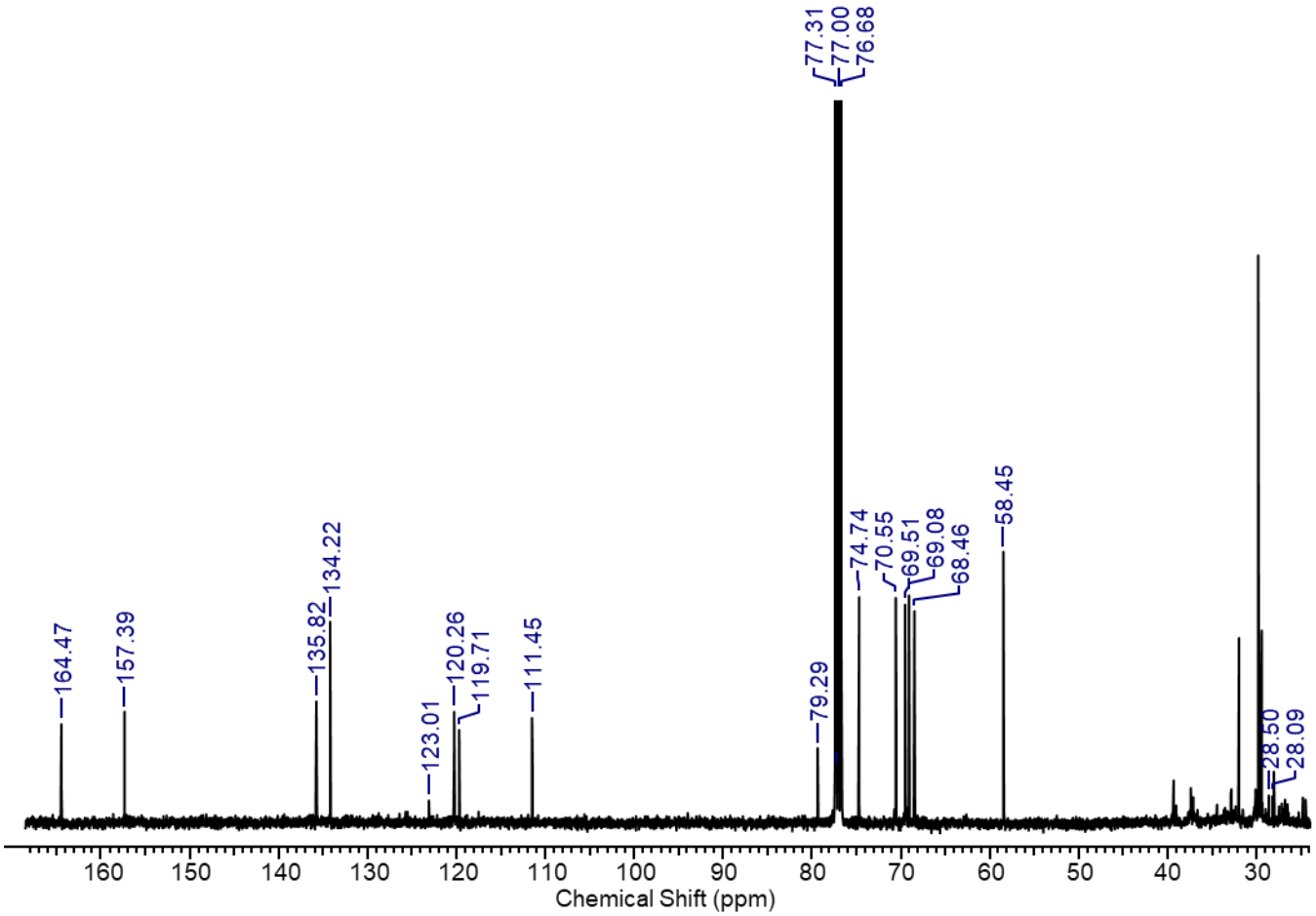


8.30 Characterisation data for synthesis of 2-(2-(2-(prop-2-yn-1-yloxy)ethoxy)ethoxy)-4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzoic acid.

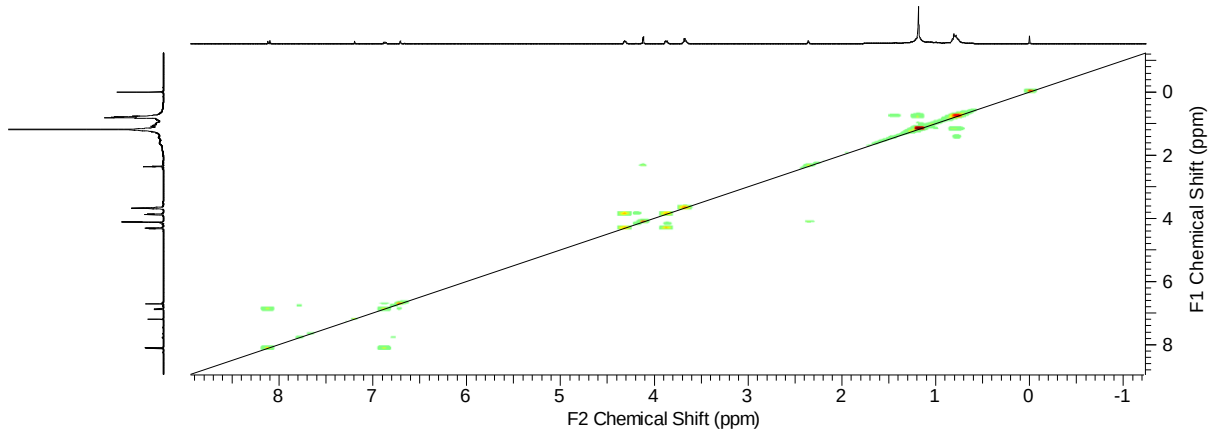
8.30.1 Extended ^{1}H NMR spectrum



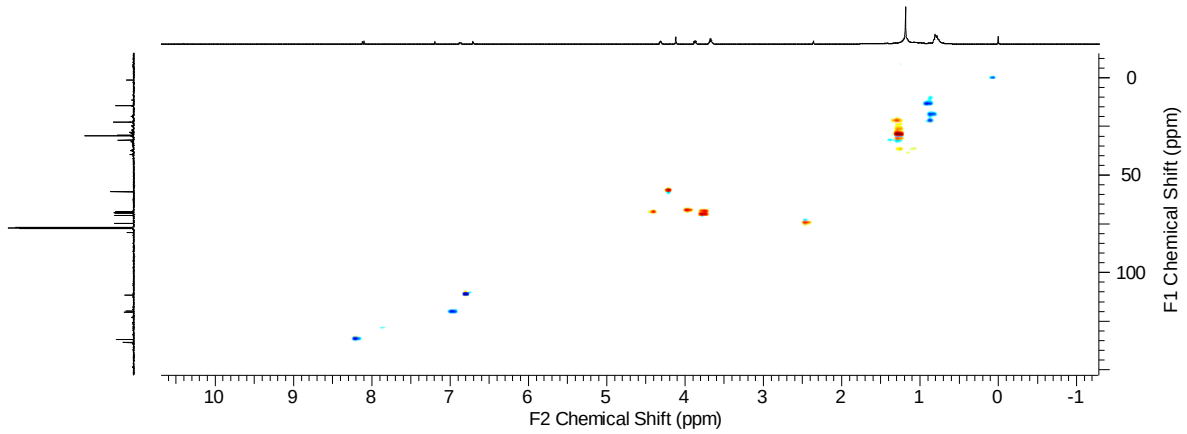
8.30.2 Magnified ^{13}C NMR spectrum



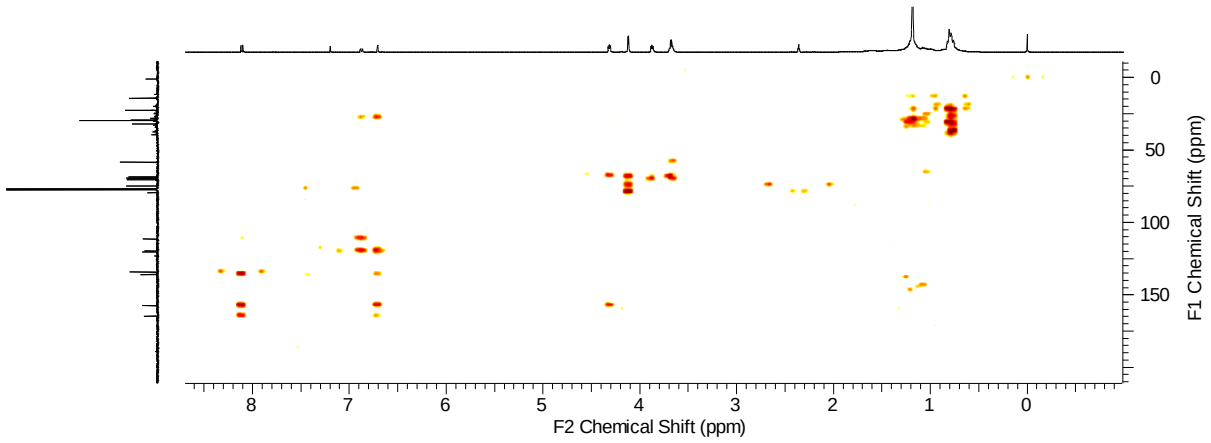
8.30.3 COSY NMR spectrum



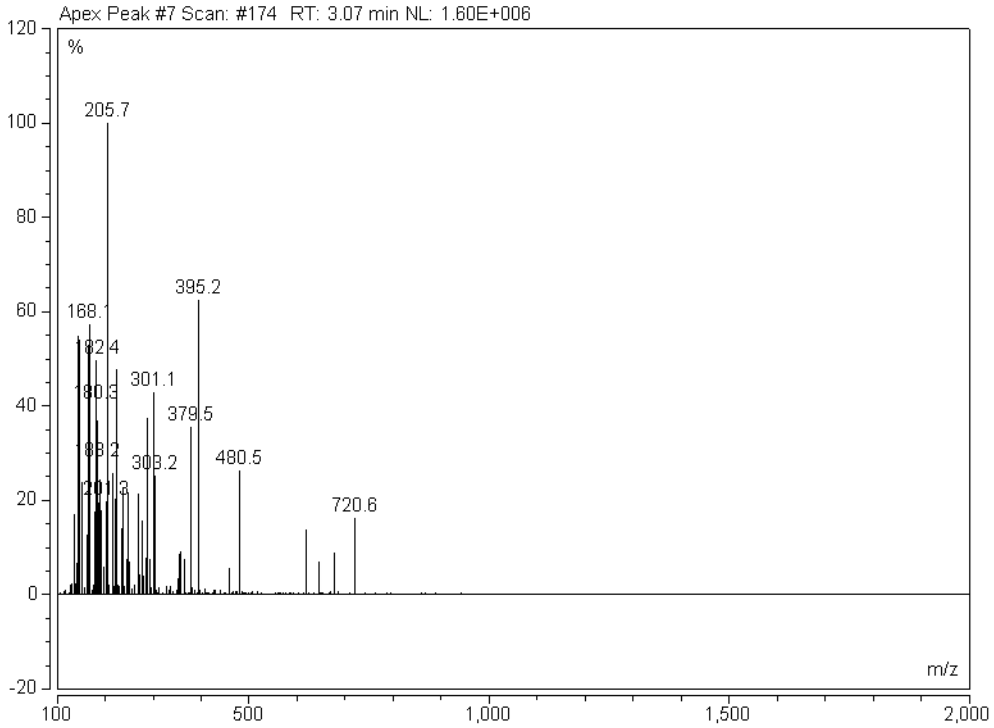
8.30.4 HSQC NMR spectrum



8.30.5 HMBC NMR spectrum

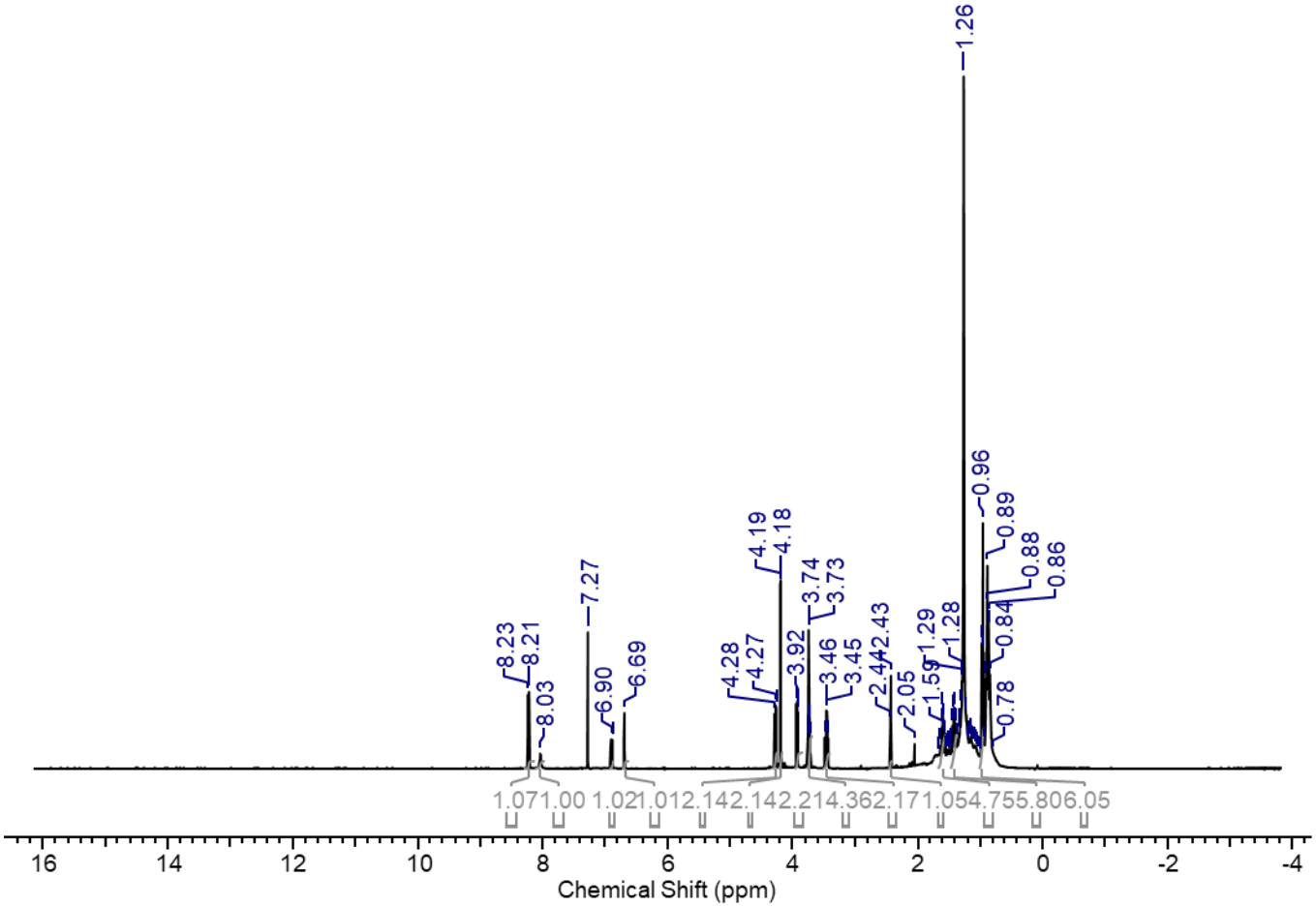


8.30.6 ESI-MS spectrum

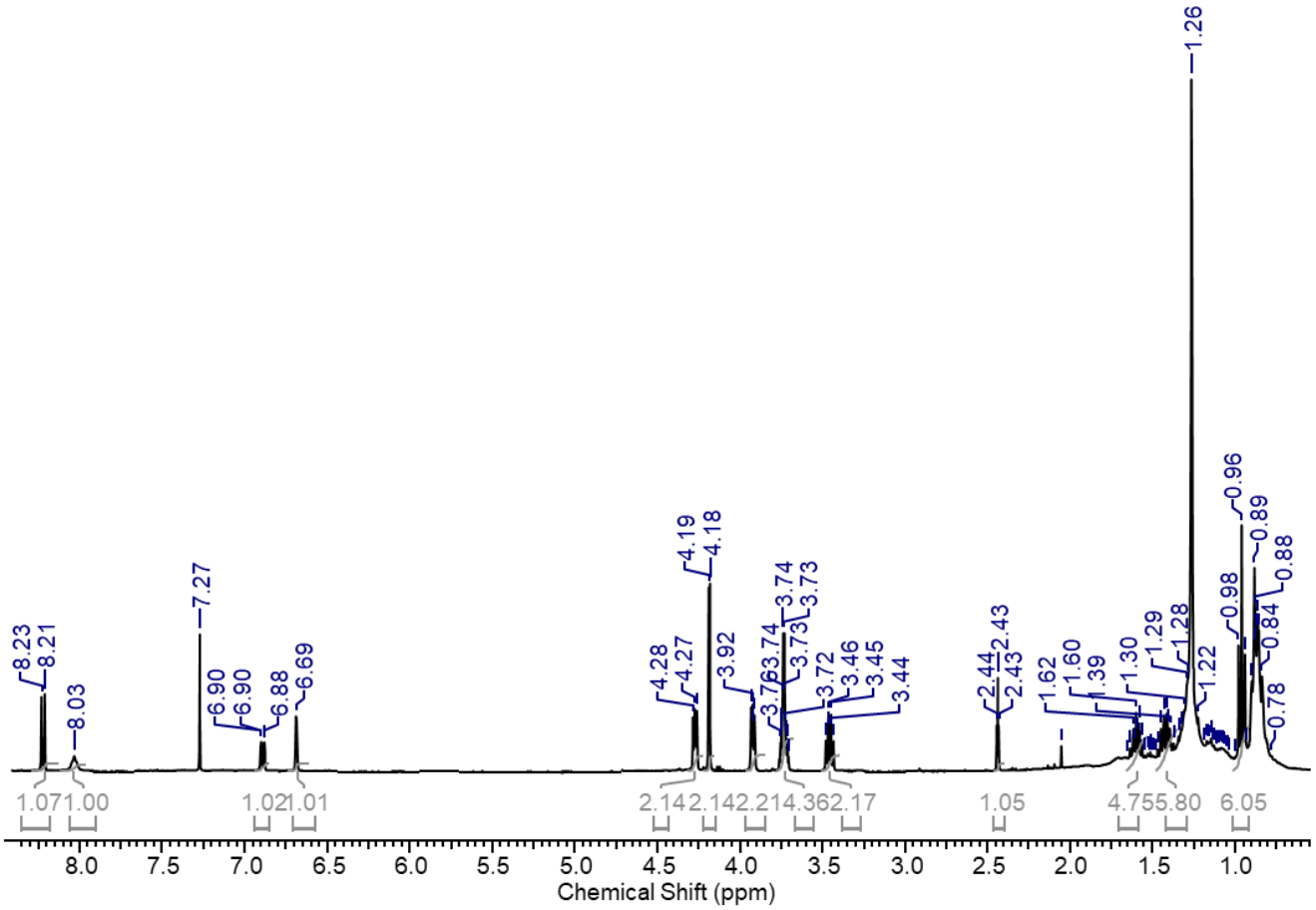


8.31 Characterisation data for the synthesis of N-butyl-2-(2-(2-(prop-2-yn-1-yloxy)ethoxy)ethoxy)-4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzamide.

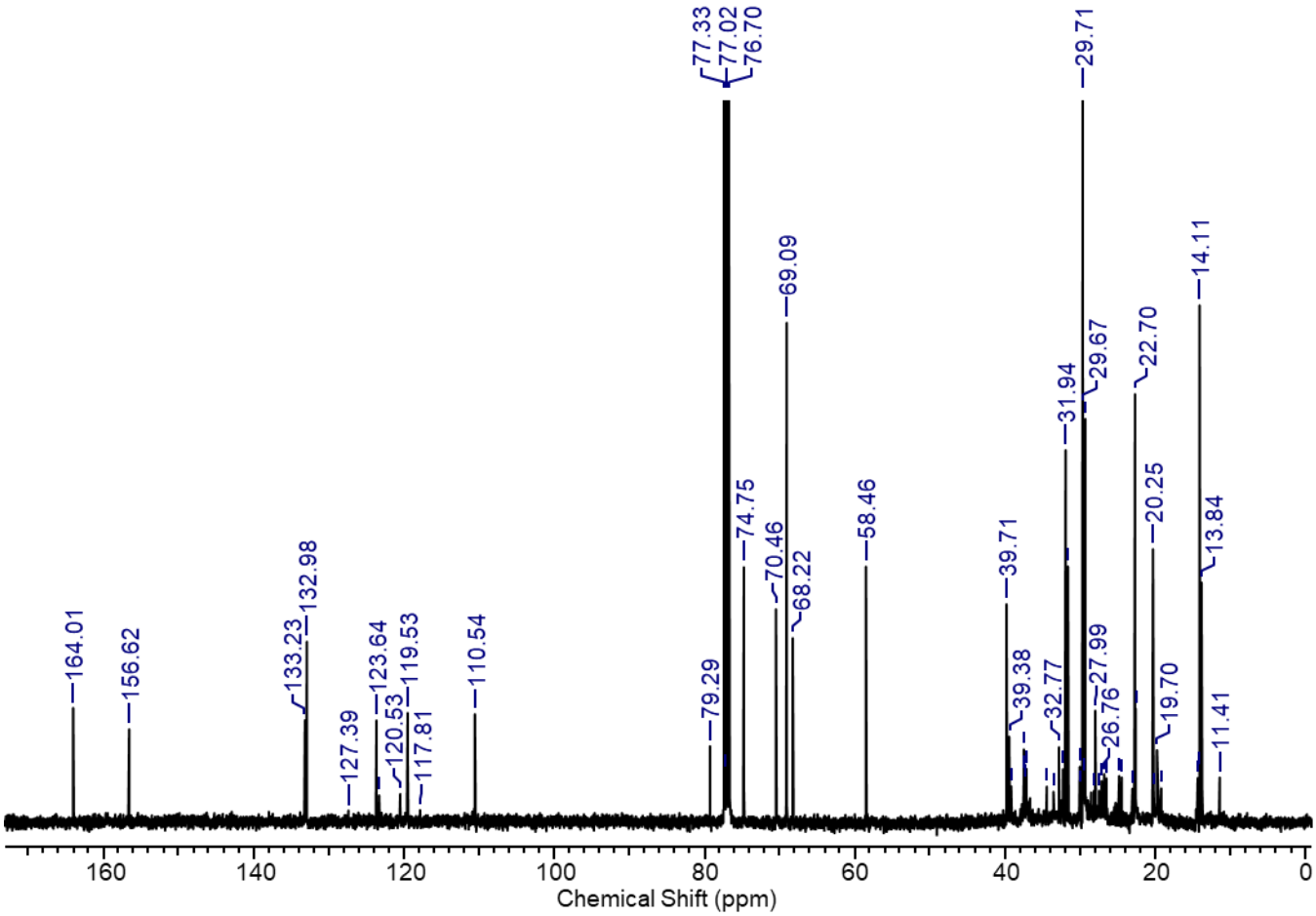
8.31.1 Extended ^{1}H NMR spectrum



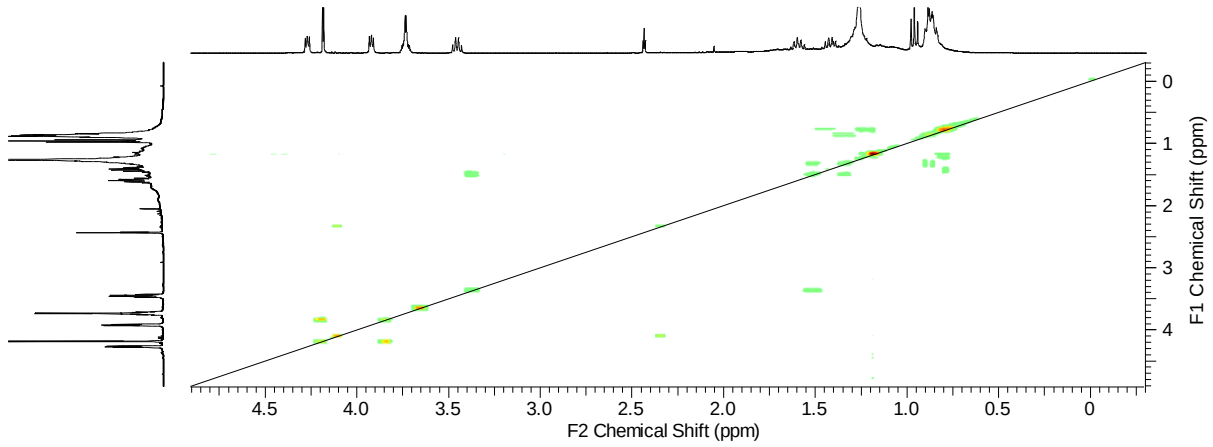
8.31.2 Magnified ^{1}H NMR spectrum



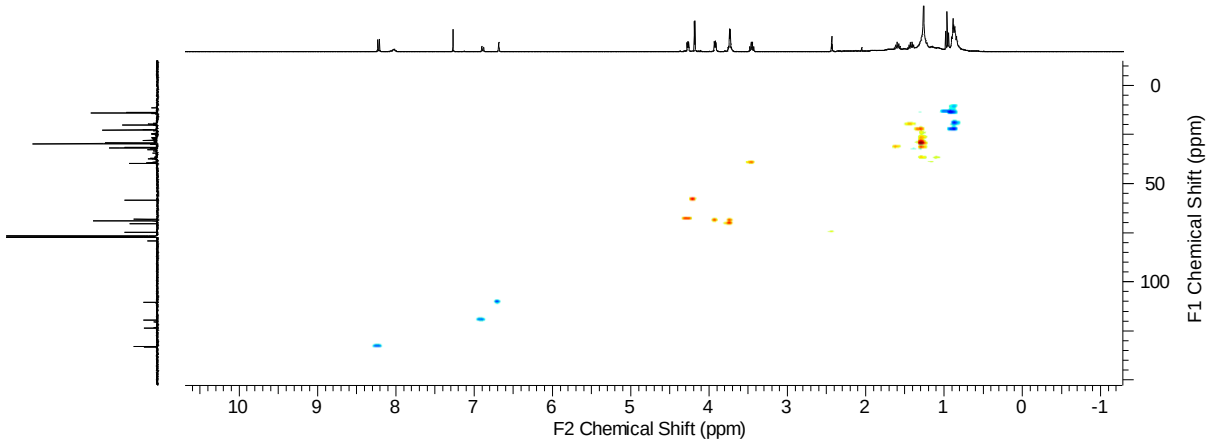
8.31.3 Magnified ^{13}C NMR spectrum



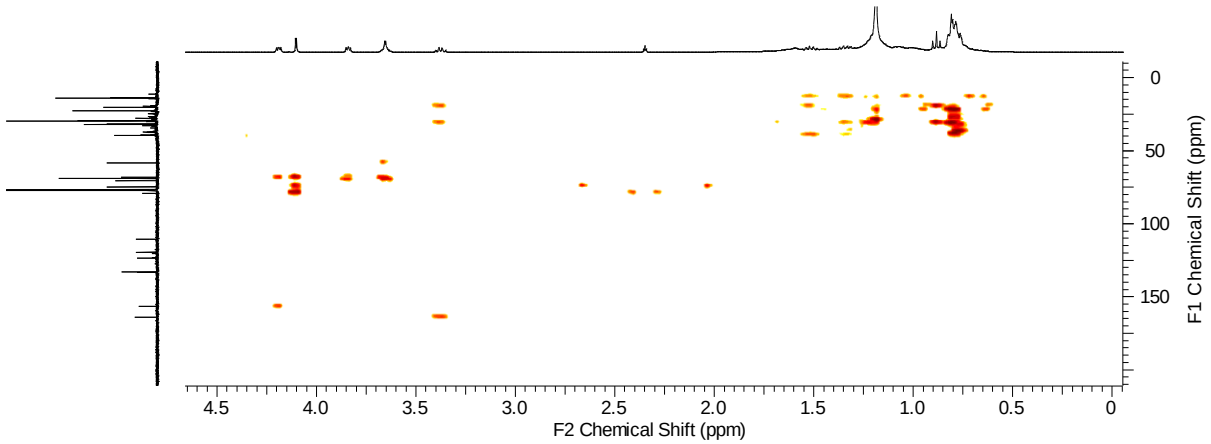
8.31.4 COSY NMR spectrum



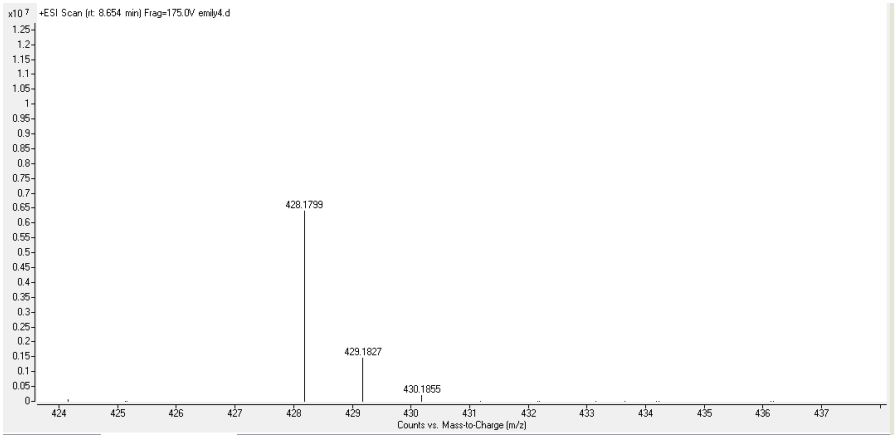
8.31.5 HSQC NMR spectrum



8.31.6 HMBC NMR spectrum

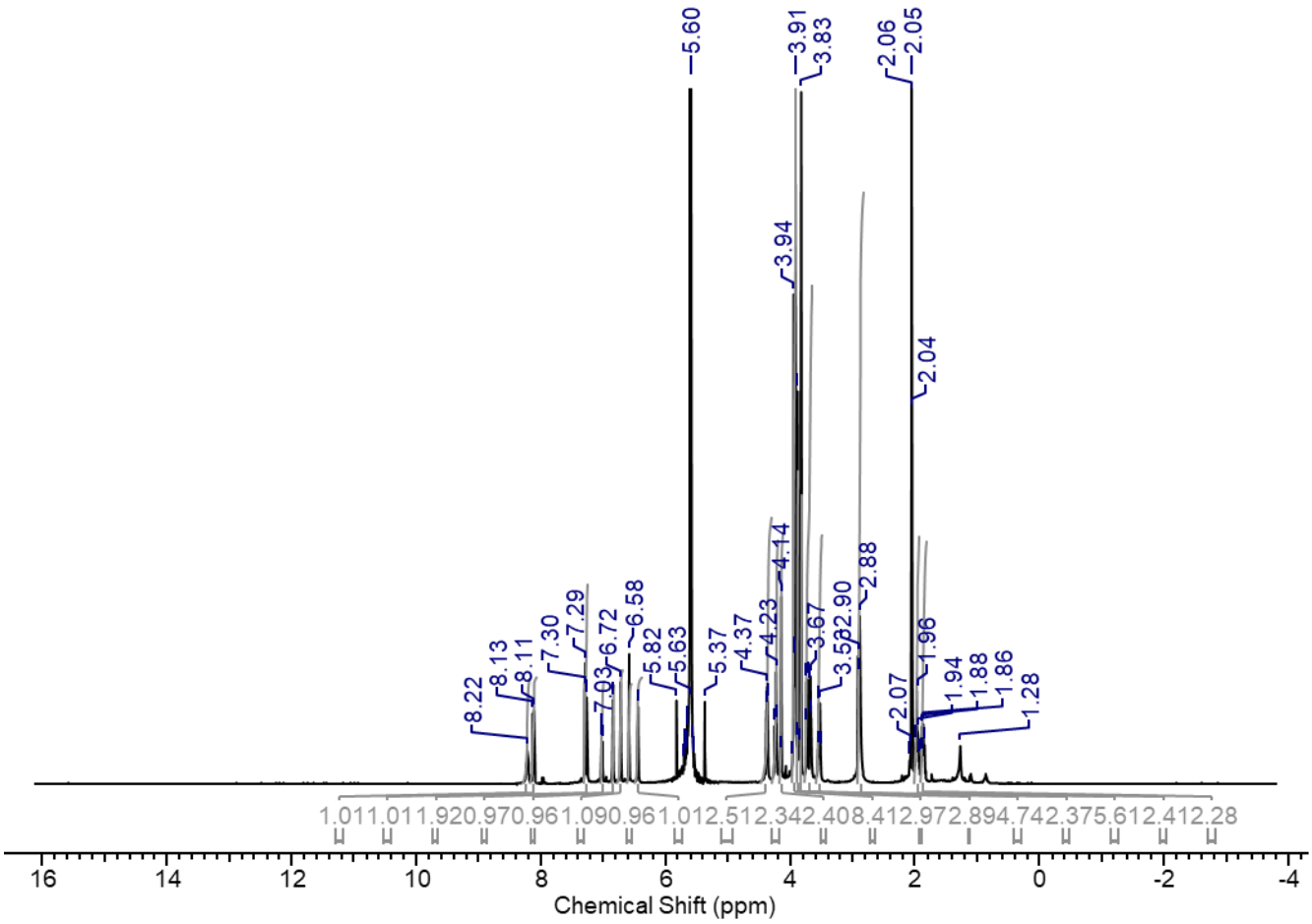


8.31.7 ESI-HRMS spectrum

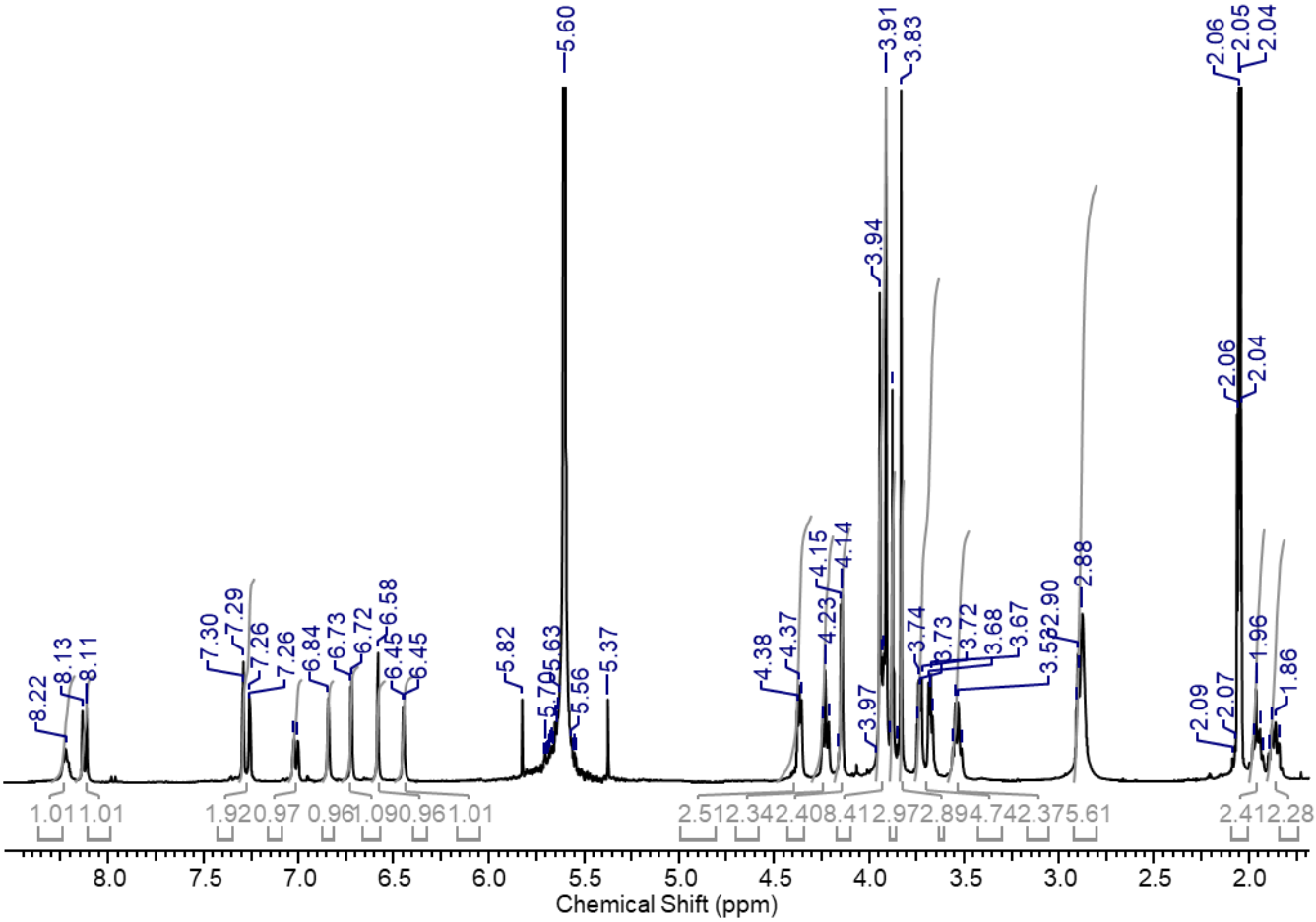


8.32 Characterisation data for the synthesis of N-(4-(5-(5,7-dimethoxy-4-oxo-4H-chromen-2-yl)-2,3-dimethoxyphenoxy)butyl)-2-(2-(2-(prop-2-yn-1-yloxy)ethoxy)ethoxy)-4-(3-(trifluoromethyl)-3H-diazirin-3-yl)benzamide.

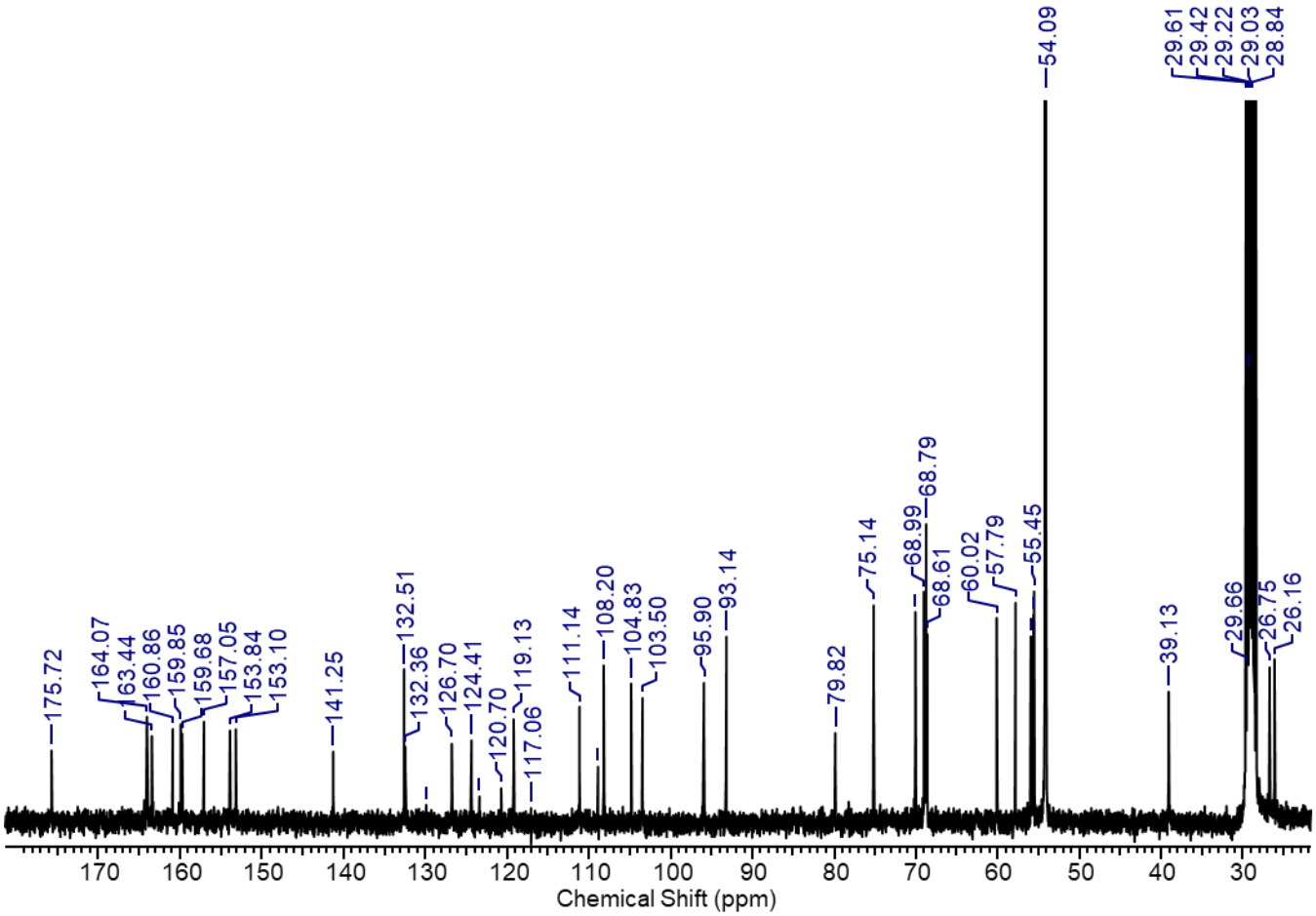
8.32.1 Extended ^{1}H NMR spectrum



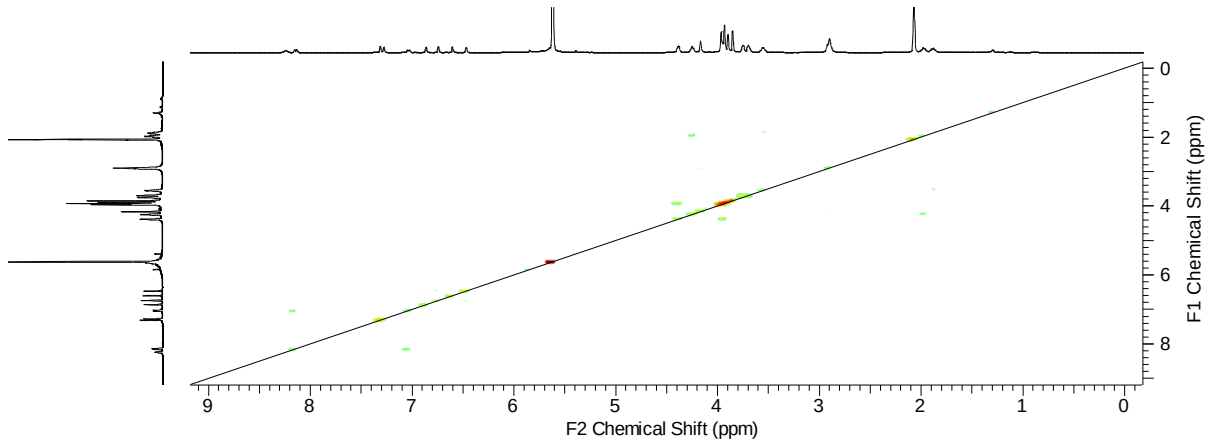
8.32.2 Magnified ^{1}H NMR spectrum



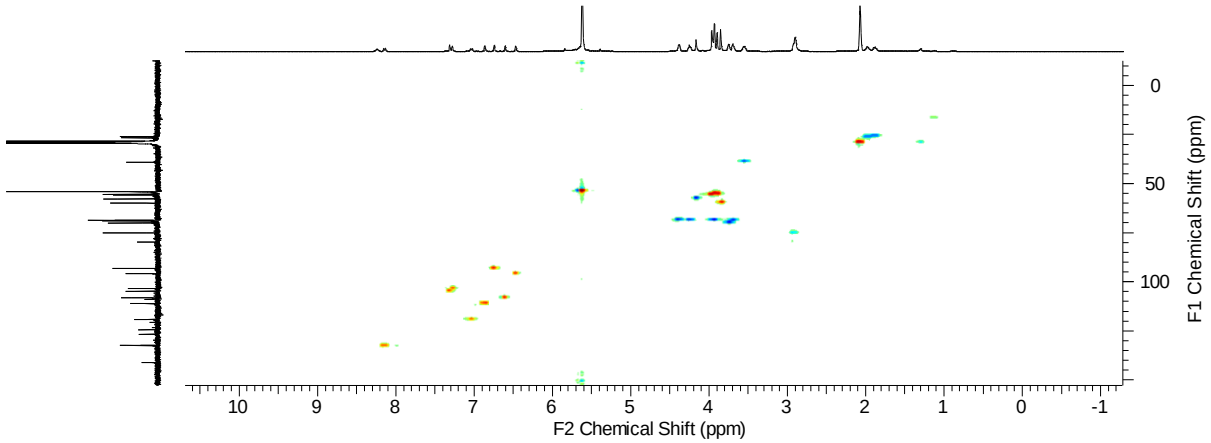
8.32.3 Magnified ^{13}C NMR spectrum



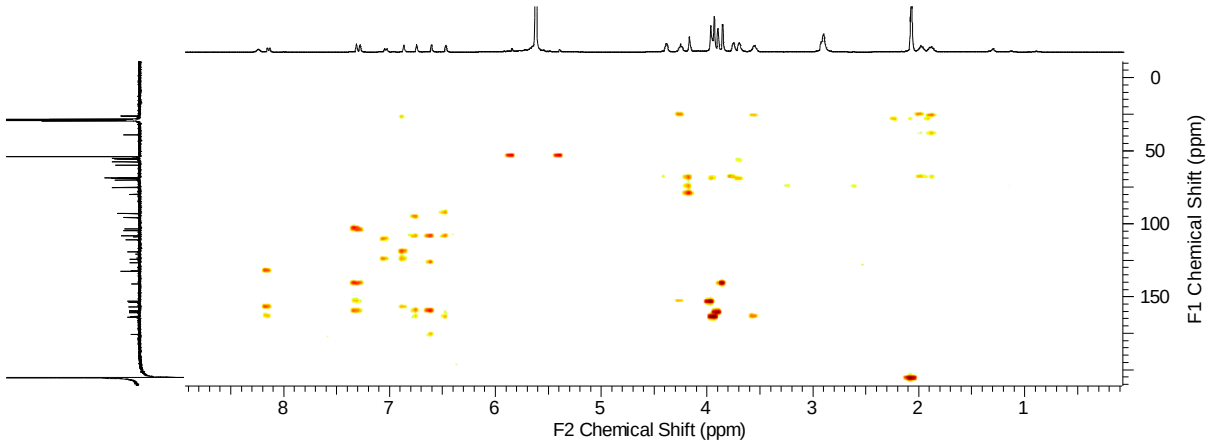
8.32.4 COSY NMR spectrum



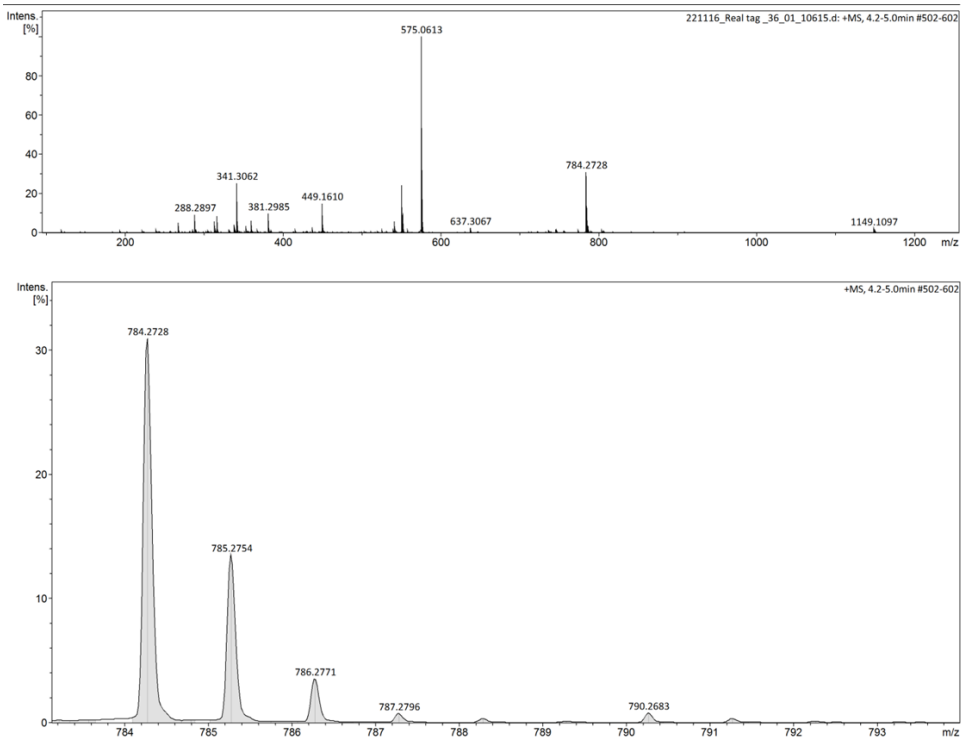
8.32.5 HSQC NMR spectrum



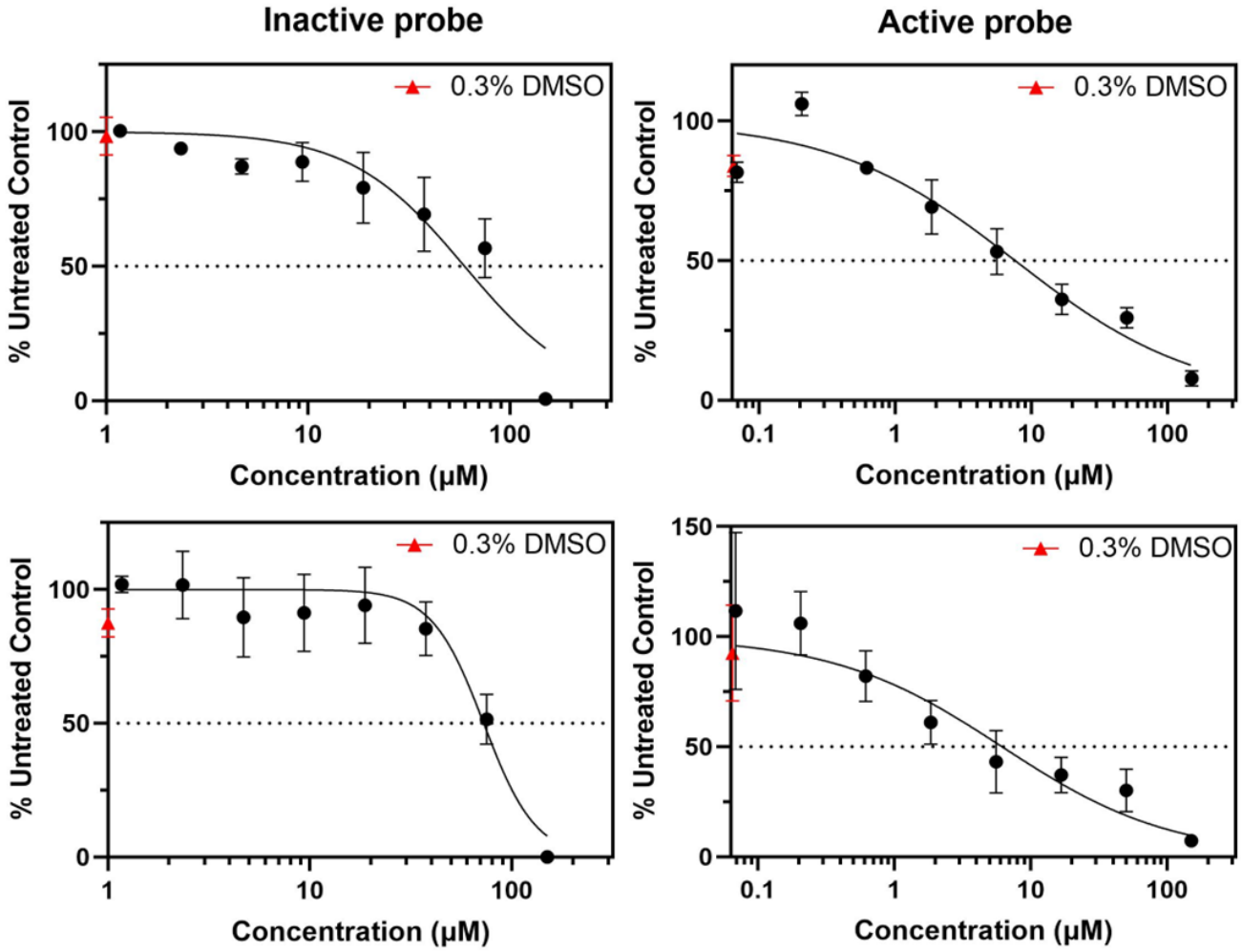
8.32.6 HMBC NMR spectrum



8.32.7 HR-MS spectrum



8.33 Triplicate data for anti-proliferative activity of active and inactive probe molecules, 4.13 and 4.14.



8.34 Lysis buffer

50mM HEPES pH7.4

250mM NaCl

0.1%NP40

1mM Pefabloc/AEBSF*

10μg/ml Aprotinin*

20μM Leupeptin*

1mM DTT

1mM EDTA

1mM NAF

10mM β-Glycerophosphate

0.1mM Sodium Orthovanadate

* from Boehringer/Roche proteinase inhibitor cocktail tablet

8.35 HEPES buffer

25 mM HEPES

150 mM NaCl

2 mM $MgCl_2$

1% SDS

pH: 7.5

8.36 TBS-T buffer

1 x TBS-T buffer

50 mM Tris pH 8.0

150 mM NaCl

0.1% TWEEN® 20

Distilled water

8.37 SDS-PAGE sample buffer

4 x Sample buffer recipe:

40% Sucrose

6.82% Tris Base

6.66% TrisHCl

8.0% SDS

0.06% EDTA

Bromophenol Blue (trace)

0.2M DTT (0.308g/10ml)

Sample buffer was diluted to 1 x for use in SDS-PAGE.

8.38 Coomassie stain

Stain

0.1% Coomassie brilliant blue

40% Methanol

10% Acetic Acid

De- Stain

10% Methanol

10% Acetic Acid

8.39 Silver stain reagents

Pierce™ silver stain kit contained silver stain developer solution, silver stain solution, silver stain enhancer solution and silver stain sensitizer solution and was used according to manufacturer instructions.

8.40 Proteomics analysis methods provided by the ICR.

Phosphoproteomics Methods

Sample Preparation. Beads from chemical pulldown experiments were added were added with a buffer containing 100 µL 0.1 M triethylammonium bicarbonate (TEAB), 5 mM tris-2-carboxyethyl phosphine (TCEP), and 10 mM iodoacetamide, sonicated and incubated for 45mins at room temperature. Proteins are digested by the addition of 10uL trypsin stock solution (500 ng/µL in 0.1% formic acid, Pierce, #90059) were added to each sample and incubated for 2 h at 37 °C, then at RT overnight in a shaker at 600 rpm. Beads were magnetised and supernatant containing the peptides are dried by centrifugal vacuum concentrator. Dried peptides were resuspended in 25uL of 100mM TEAB and were TMT-labelled were labelled with a tandem mass tag (TMT) multiplex reagent vial (Thermo Scientific #A34808) according to the manufacturer instructions. according to manufacturer's instructions. All the TMT-labelled samples were combined in equal amounts to a single tube. The sample was then dried with a centrifugal vacuum concentrator.

High-pH reversed-phase phosphopeptide fractionation. Offline peptide fractionation was based on high pH Reverse Phase (RP) chromatography using the Waters XBridge C18 column (2.1 × 150 mm, 3.5 µm) on a Dionex Ultimate 3000 HPLC system at a 0.85% gradient with flow rate 0.2 mL/min. Mobile phase A was 0.1% ammonium hydroxide, and mobile phase B was 100% acetonitrile, 0.1% ammonium hydroxide. Retention time-based fractions are collected and pooled into thirty-two samples for LC-MS analysis.

Liquid chromatography-mass spectrometry analysis. LC–MS analysis LC-MS analysis was performed on the Dionex UltiMate 3000 UHPLC system coupled with Orbitrap Ascend Tribrid mass spectrometer (Thermo Scientific). Samples were analyzed with the nanoEase M/Z Peptide BEH C18 C18 column (75 µm × 25 cm, 1.7 µm) at 45 °C. Mobile phase A was 0.1% formic acid and mobile phase B was 80% acetonitrile, 0.1% formic acid. The gradient separation method was as follows: 80 min gradient up to 35% B, for 5 min up to 95% B, for 5 min isocratic at 95% B, re-equilibration to 5% B in 5 min, for 5 min isocratic at 5% B.

For MS analysis, precursors between 375 and 1,500 m/z were selected, with mass resolution of 120,000, automatic gain control (AGC) of 4×10^5, and IT (injection time) of 50 ms, with the top speed mode in 3 s, and the precursors were isolated for collision-induced dissociation (CID) fragmentation with a quadrupole isolation width of 0.7 Th (Thomson unit). Collision energy was set at 35%, with AGC at 1×10^4 and IT at 50 ms. MS^3 quantification was obtained with higher-energy collisional dissociation (HCD) fragmentation of the top 5 most abundant CID fragments isolated with synchronous precursor selection (SPS). Quadrupole isolation width was set at 0.7 Th, collision energy was applied at 65%, and the AGC setting was at 1×10^5 with IT at 105 ms. The HCD MS^3 spectra were acquired for the mass range 100 to 500 with a resolution of 50,000. Targeted precursors were dynamically excluded for further isolation and activation for 45 s with 7 ppm mass tolerance.

Database search and protein quantification. The SEQUEST-HT search engine was used to analyse the acquired mass spectra in Proteome Discoverer 2.4 (Thermo Scientific) for protein identification and

quantification. The precursor mass tolerance was set at 20 ppm, and the fragment ion mass tolerance was set at 0.02 Da. Spectra were searched for fully tryptic peptides with maximum 2 mis-cleavages. TMT on lysine residues and peptide N termini (+229.163 Da) and carbamidomethylation of cysteine residues (+57.0215 Da) were set as static modifications. Phosphorylation of serine/ threonine/ tyrosine (+79.966 Da), oxidation of methionine residues (+15.9949 Da) and deamidation of asparagine and glutamine (+0.9848 Da) were set as a variable modification. Peptide confidence was estimated with the Percolator. The peptide FDR was set at 0.01, and validation was based on the q value and a decoy database search. All spectra were searched against UniProt-SwissProt proteomes of *Homo sapiens* protein entries appended with contaminants. The reporter ion quantifier node included a TMT quantification method with an integration window tolerance of 15 ppm and integration method based on the most confident centroid peak at the MS^3 level. Only unique peptides were used for quantification, with protein groups considered for peptide uniqueness. Peptides with an average reporter signal-to-noise ratio of >3 were used for protein quantification. Correction for isotopic impurity of reporter quantification values is applied. Only spectra with at least 50% of the SPS masses matching to the identified peptides are used for quantification.

References:

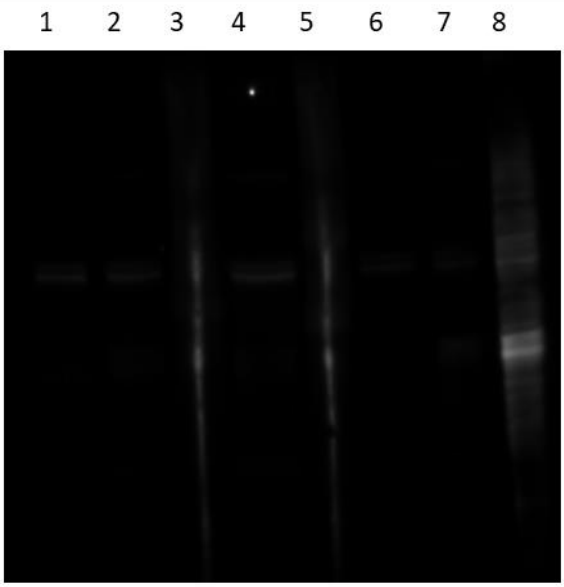
SimPLIT: Simplified Sample Preparation for Large-Scale Isobaric Tagging Proteomics

Fernando J. Sialana, Theodoros I. Roumeliotis, Habib Bouguenina, Laura Chan Wah Hak, Hannah Wang, John Caldwell, Ian Collins, Rajesh Chopra, and Jyoti S. Choudhary

Journal of Proteome Research **2022** *21* (8), 1842-1856

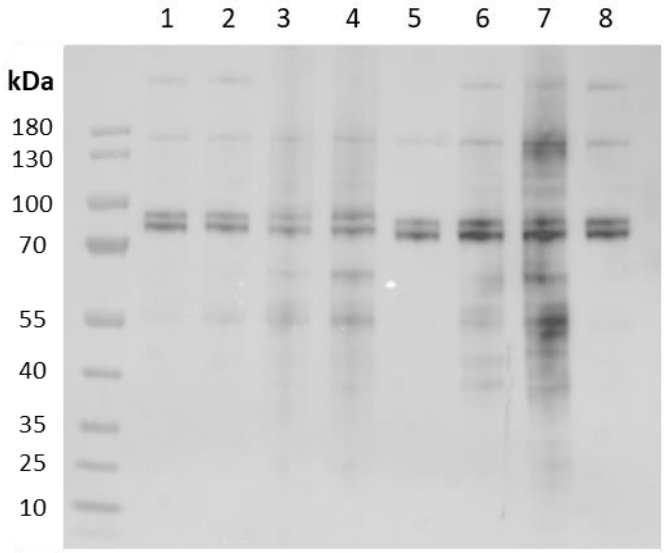
DOI: 10.1021/acs.jproteome.2c00092

8.41 Supplementary image of western blot for experiment 4.



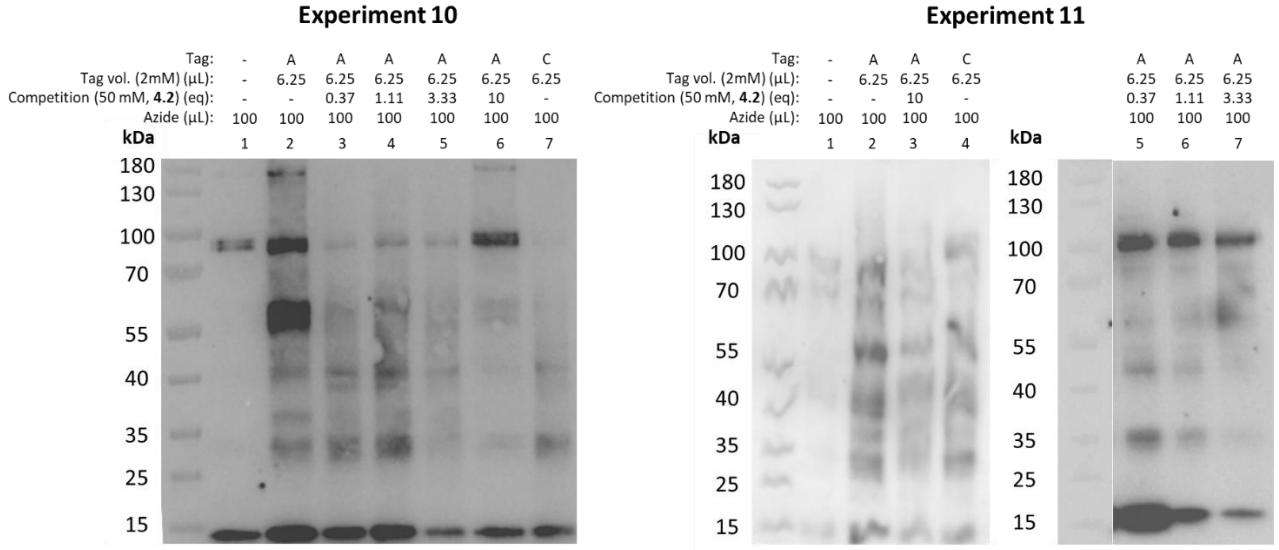
Experiment 4

8.42 Supplementary image of western blot for experiment 5.

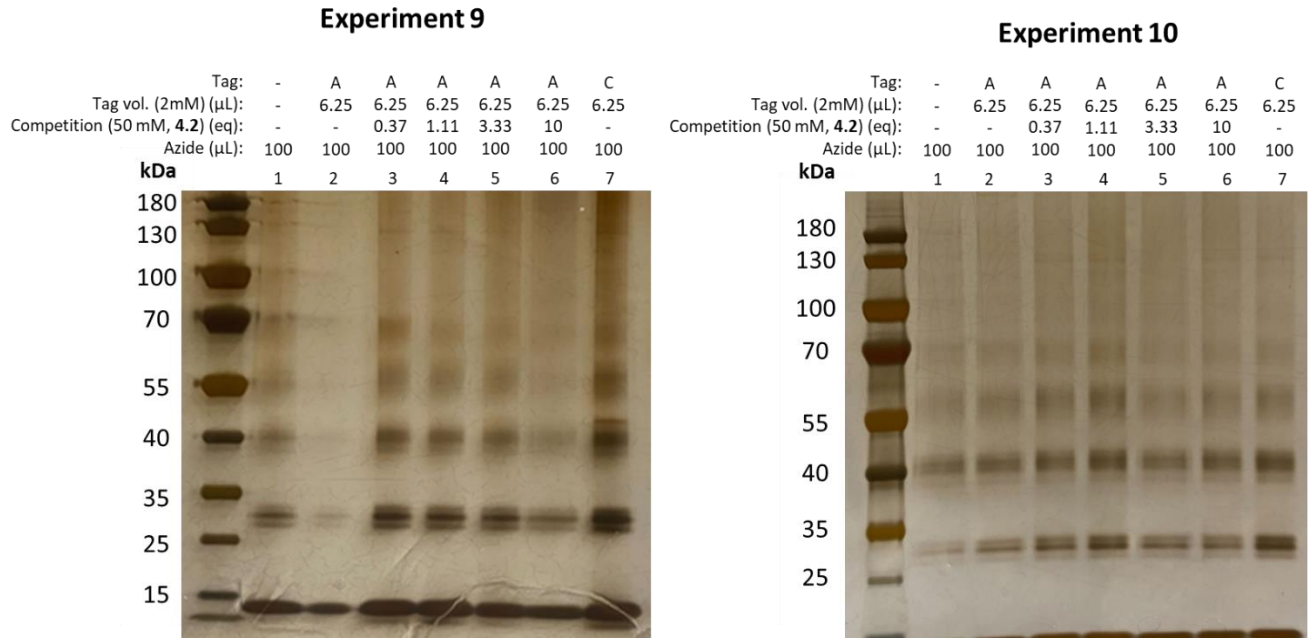


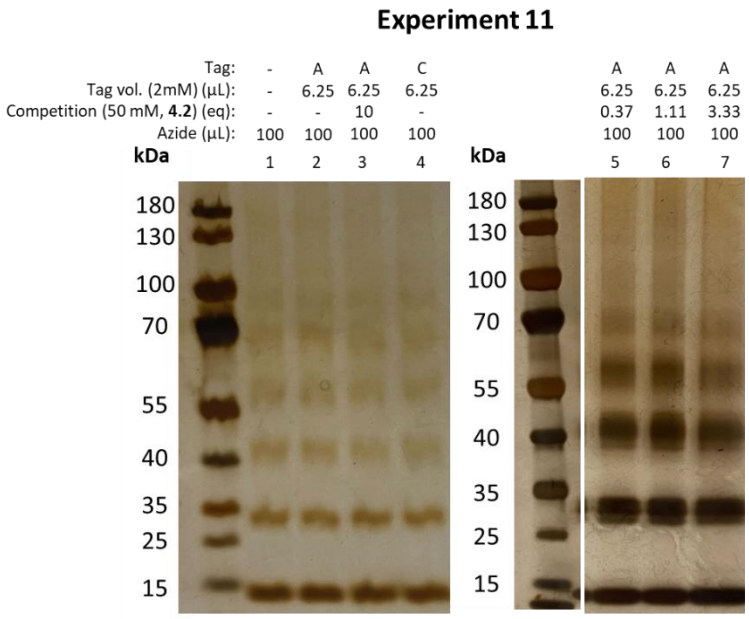
Experiment 5

8.43 Supplementary data from western blot experiments 10 and 11.

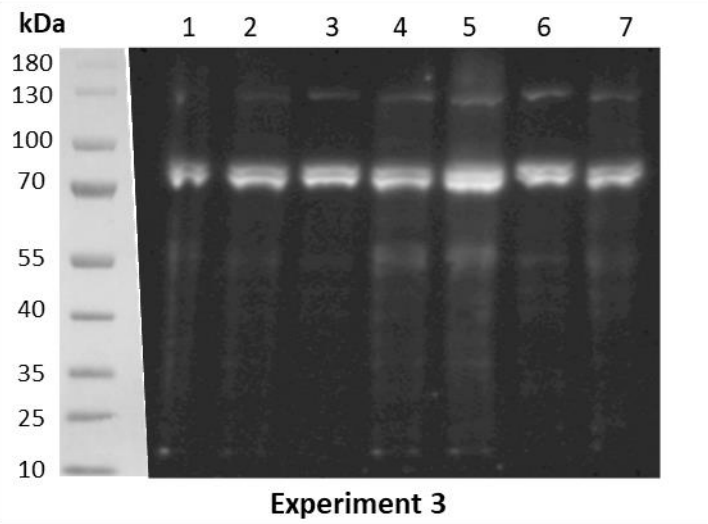


8.44 Silver stained SDS-PAGE gels for experiments 9-11.





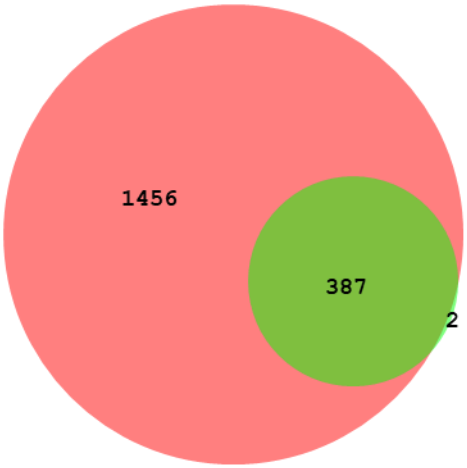
8.45 Supplementary image of western blot for experiment 3.



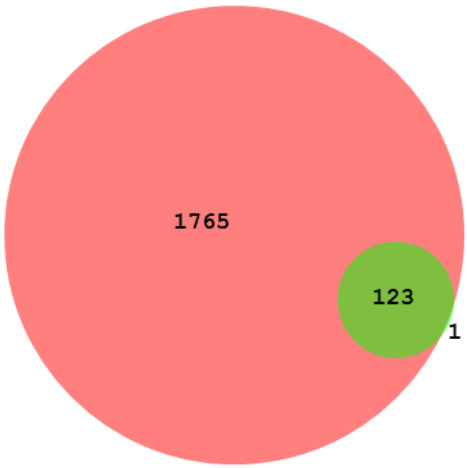
8.46 Raw proteomics data from ICR

See additional document

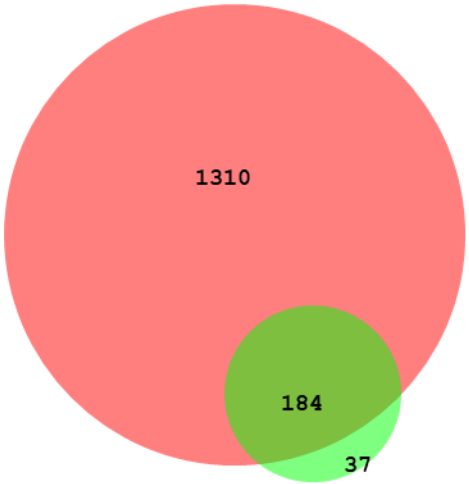
8.47 Venn diagrams for each normalisation condition and number of overlaps.



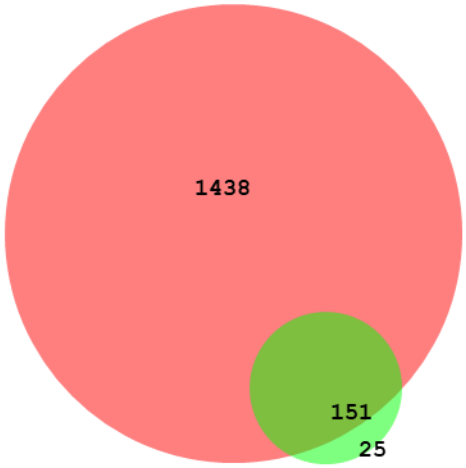
Method 2
Number of common proteins: 387



Method 3
Number of common proteins: 123



Method 4
Number of common proteins: 184



Method 5
Number of common proteins: 151

8.48 List of overlapping proteins for each normalisation method

See additional document

8.49 Proteins common to four methods of normalisation

UniProt	Symbol	Protein name
Q8NFV4	ABHD11	abhydrolase domain containing 11
P30520	ADSS	adenylosuccinate synthase
P35573	AGL	amylo-alpha-1, 6-glucosidase, 4-alpha-glucanotransferase
Q92688	ANP32B	acidic nuclear phosphoprotein 32 family member B
Q9UBB4	ATXN10	ataxin 10
Q8N163	CCAR2	cell cycle and apoptosis regulator 2
P60953	CDC42	cell division cycle 42

Q86XP3	DDX42	DEAD-box helicase 42
Q16698	DECR1	2,4-dienoyl-CoA reductase 1
P31689	DNAJA1	DNAJ heat shock protein family (Hsp40) member A1
O60884	DNAJA2	DNAJ heat shock protein family (Hsp40) member A2
Q8WXX5	DNAJC9	DNAJ heat shock protein family (Hsp40) member C9
Q86YB8	ERO1B	endoplasmic reticulum oxidoreductase 1 beta
Q9BQS8	FYCO1	FYVE and coiled-coil domain containing 1
P48506	GCLC	glutamate-cysteine ligase catalytic subunit
Q9UBI6	GNG12	G protein subunit gamma 12
O14929	HAT1	histone acetyltransferase 1
P15559	NQO1	NAD(P)H quinone dehydrogenase 1
P13667	PDIA4	protein disulfide isomerase family A member 4
P00558	PGK1	phosphoglycerate kinase 1
O60664	PLIN3	perilipin 3
P30041	PRDX6	peroxiredoxin 6
P10644	PRKAR1A	protein kinase cAMP-dependent type I regulatory subunit alpha
Q15907	RAB11B	RAB11B, member RAS oncogene family
P61106	RAB14	RAB14, member RAS oncogene family
A4D1S5	RAB19	RAB19, member RAS oncogene family
P61019	RAB2A	RAB2A, member RAS oncogene family
P20340	RAB6A	RAB6A, member RAS oncogene family
Q15382	RHEB	RAS homolog, mTORC1 binding
Q9Y6Y8	SEC23IP	SEC23 interacting protein
P09661	SNRPA1	small nuclear ribonucleoprotein polypeptide A'
Q8TC07	TBC1D15	TBC1 domain family member 15
Q96LD4	TRIM47	tripartite motif containing 47
Q9UBT2	UBA2	ubiquitin like modifier activating enzyme 2
Q08AM6	VAC14	Vac14, PIKFYVE complex component

Proteins highlighted in yellow are seen in the list of top 20 protein hits identified for normalisation method 3.

8.50 Differentially expressed proteins identified using normalisation method 3

See additional document

8.51 Volcano plot showing proteins significantly decreased in the competition samples that also had a greater than 1.5 fold change increase from DMSO to tag.

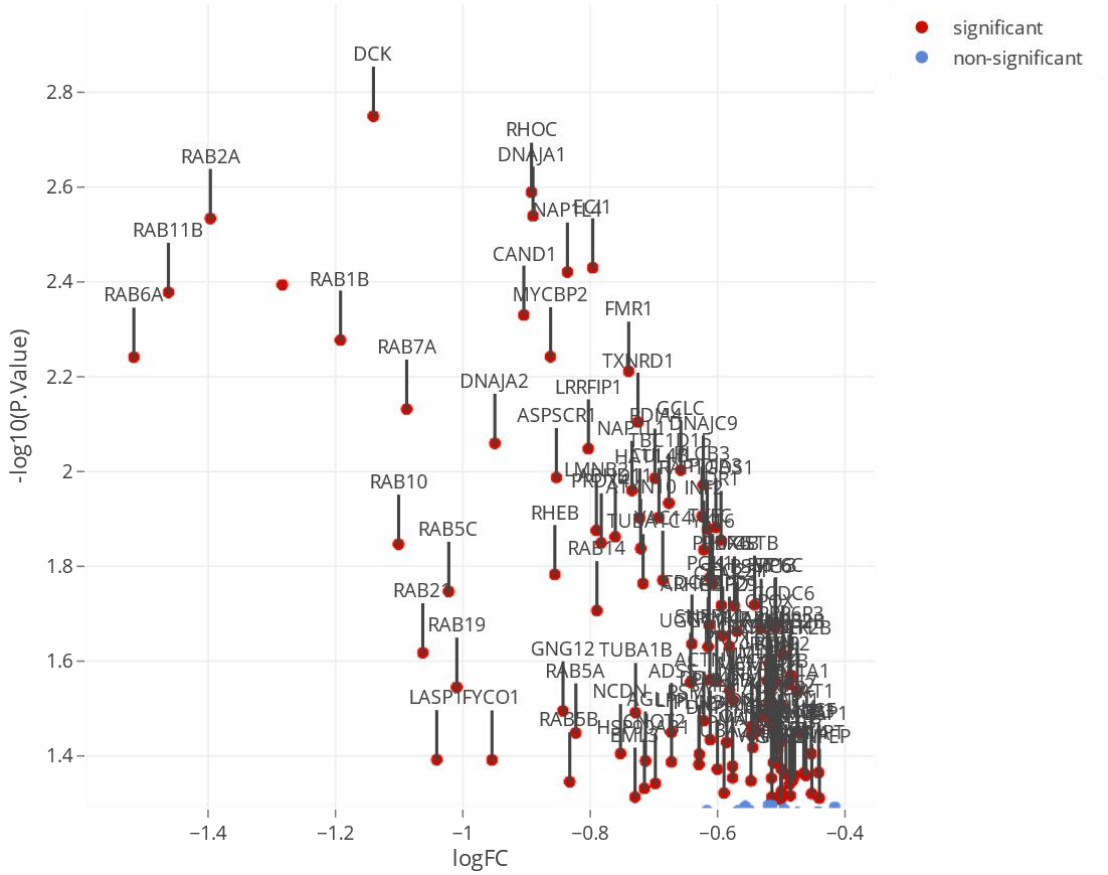
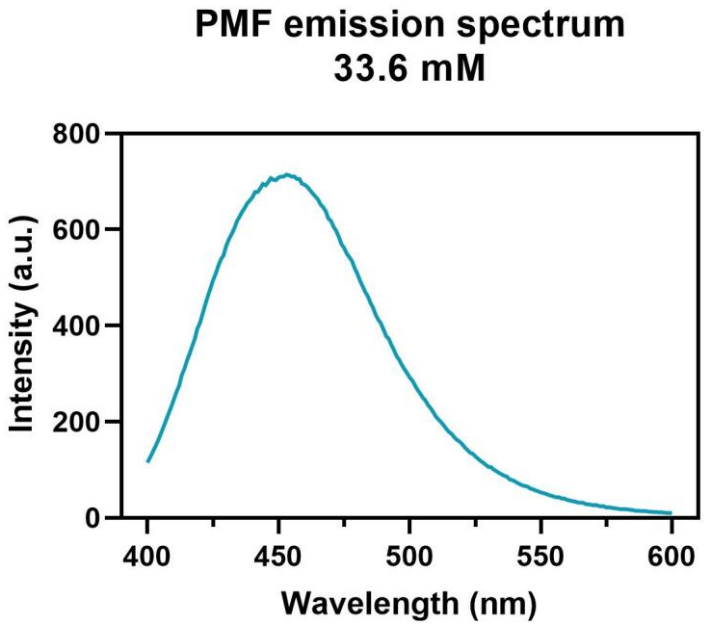


Image shows a magnified section of the volcano plot seen in Figure 5.17 with the addition of protein symbols for the 123 overlapping proteins.

8.52 Emission spectrum of PMF, **2.1**, after irradiation with 365 nm light.



8.53 UCL RNA sample preparation and sequencing methods

See additional document

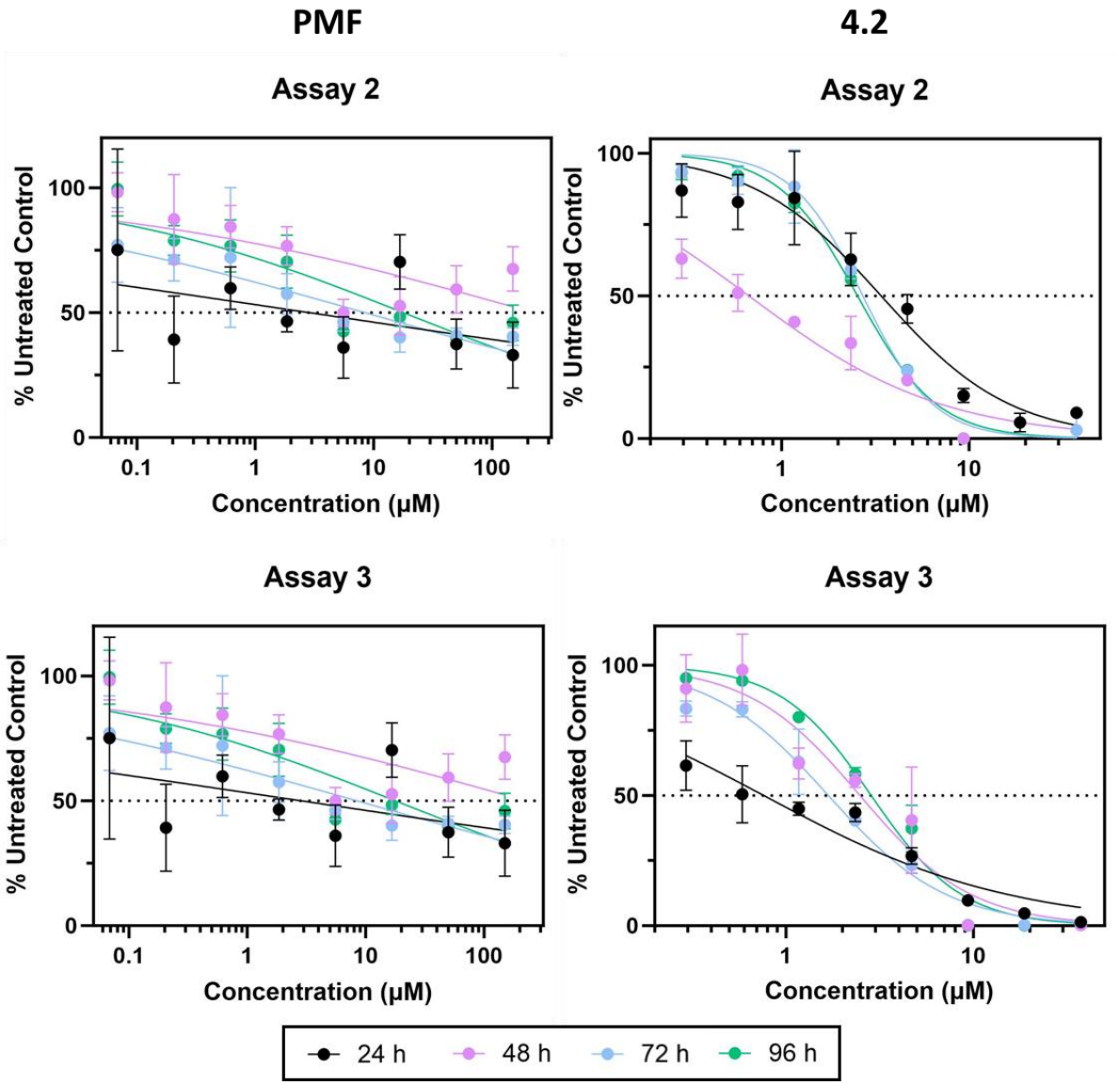
8.54 MultiQC report assessing the quality of the RNA sequenced data.

See additional document

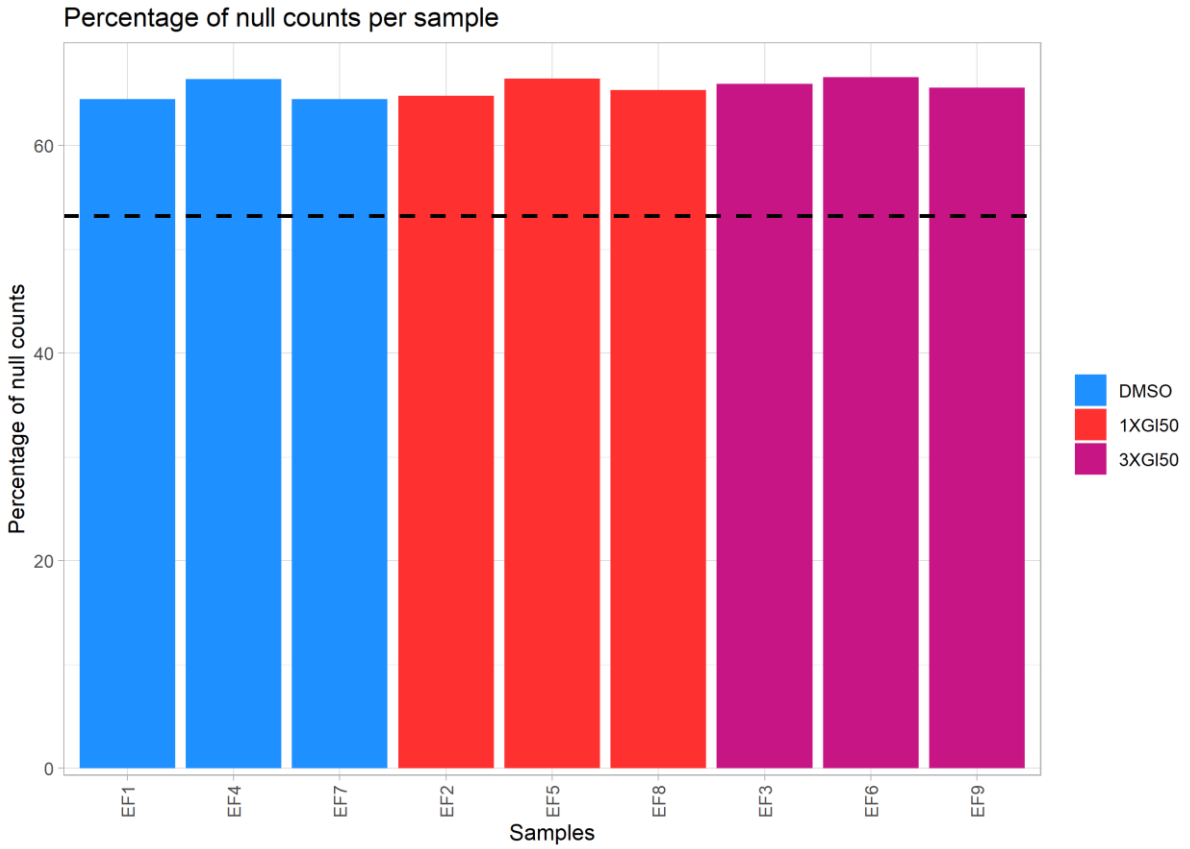
8.55 Statistical report of differentially expressed genes identified in the RNA sequencing experiment.

See additional document

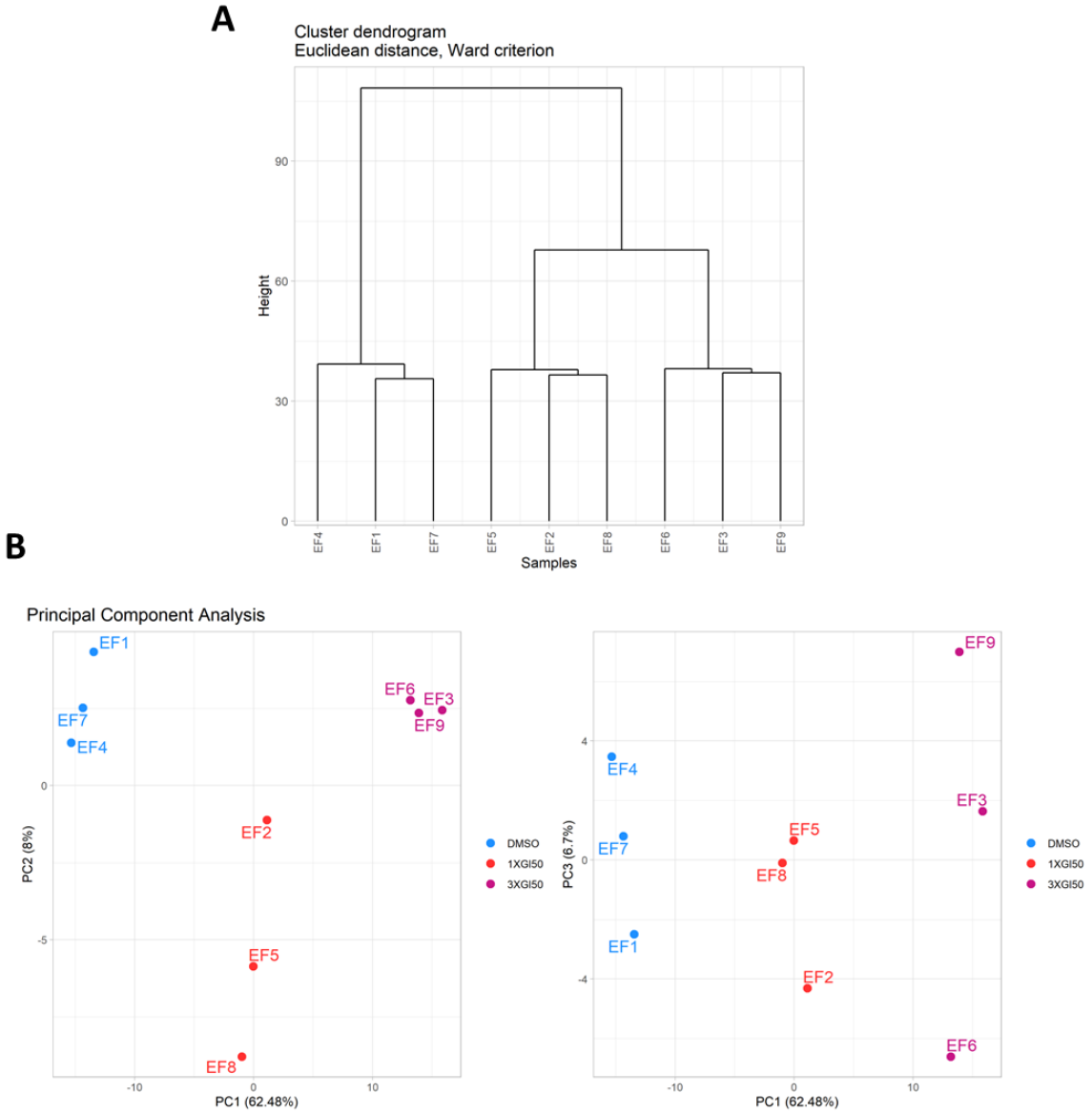
8.56 Supplementary graphs for time dependent SRB assays of PMF and compound 4.2.



8.57 Percentage of features with null read counts in each sample.



8.58 Analysis of sample variability



A) Hierarchical clustering of samples

B) First two components of a principle component analysis, showing percentages of variance assigned to each axis.

8.59 Significant differentially expressed features for each condition identified using RNA sequencing. Feature names correspond to gene names and were generated during RNA sequencing.

1 x GI_{50} vs DMSO	1 x GI_{50} vs DMSO	3 x GI_{50} vs DMSO	3 x GI_{50} vs DMSO	1 x GI_{50} vs 3 x GI_{50}	1 x GI_{50} vs 3 x GI_{50}
Upregulated	**Downregulated**	**Upregulated**	**Downregulated**	**Upregulated**	**Downregulated**
TM9SF1	RP11-603J24.9	TM9SF1	CTD-2287O16.3	RP11-603J24.9	GPR110
BOLA2	SYT13	BOLA2	ANXA10	HIST1H2AC	RAD51AP1
SHPK	KIAA1199	SHPK	NLRP11	HIST1H2BD	E2F2
CLIC3	GRIN2B	HIST1H2AC	KRT23	CHAC1	PCDH7
WNT9A	SCG2	VGF	GPR110	SLFN5	CDK1
MT2A	MRAP2	HIST1H2BD	PCDH7	DNAJB9	CYP24A1

VGF	DACT1	CLIC3	HNF4G	ASNS	MCM10
TNFRSF6B	ANXA10	SLFN5	FAM198B	PNRC1	E2F8
SMIM14	BDH2	MT2A	SYT13	FAM228B	KRT23
AHNAK2	KITLG	LA16c-380H5.5	CYP24A1	VGF	RAD54L
ABLIM2	TLR3	WNT9A	KIAA1199	FAM129A	MTFR2
SCN9A	RASSF2	SMIM14	KITLG	KDM5B	OAS3
TM7SF2	GAL	AHNAK2	CCAT1	ADM2	RIBC2
ITGB4	KRT23	DNAJB9	GRIN2B		GINS2
ITGA2	TNFSF18	SCN9A	LINC00326		RAD51
SLFN5	GPR110	ABCA3	PRIMA1		CENPW
AGPAT4	KRT80	MT1X	BDH2		POLQ
MAP2	AL162759.1	ASNS	RP11-551L14.1		DLEU2
HIST1H2BD	MKX	TNFRSF6B	TLR3		PLK1
MT1E	PCDH7	H6PD	SLC7A2		SPC24
DHRS2	SLC7A2	NDRG1	RASSF2		PBK
MT1X	RP11-551L14.1	ITGA2	KRT80		TMSB15A
HIST1H2AC	ACSL5	AP003068.18	WNT10B		MND1
H6PD	KRT7	AGPAT4	DHRS3		SPC25
THSD4	SPTBN2	FAM214A	TRANK1		RAD54B
ABCA3	ARL14EPL	ABLIM2	ARL14EPL		DHRS3
ITGA3	CYP24A1	TM7SF2	IL18R1		MYB
TUBA1A	PRIMA1	THSD4	RAD51AP1		XRCC3
SEC24D	IL18R1	MT1E	SCG2		KIAA0101
HID1	DPP4	ZFP41	ACSL5		CDCA5
NEU1	HS3ST1	SMPD1	GPD1		CCNA2
ADIRF	KLK6	PCLO	MOB3B		EXO1
NDRG1	NELL2	HELZ2	KRT20		GMNN
TLN2	LONRF2	ITGA3	C3orf80		TTK
C6orf1	FAM198B	CCDC92	BMP6		FAM64A
CCDC92	NLRP11	OSGIN1	GPR87		UBE2C
FAM110C	SLC38A5	CLDN9	TFR2		CDCA3
GLRX	RAB7L1	COL4A3	E2F8		RRM2
DNAJB9	RBM24	MXD4	MRAP2		MAD2L1
H1F0	DGKE	CHAC1	RBM24		CENPQ
L1CAM	BANK1	FICD	PLA2G4A		PSRC1
LOXL2	IQGAP2	SEC24D	ZBED2		PRIM1
ABCC3	ITGB8	RHOF	BATF3		CDC45
SFN	HNF4G	TMEM132B	DMBX1		DSCC1
FICD	KLF11	EFR3B	DGKE		BUB1B
RHOF	TRANK1	FAM129A	IFIT1		ANLN
NPTXR	ZNF618	RP11-50E11.3	OAS3		ZBED2
SMPD1	WNT10B	PNRC1	SLC38A5		PKMYT1
EPPK1		EVI5L	ILDR2		ESCO2
EVI5L		HIST1H2AG	SEMA3A		CDC25A
ITFG3		ITGB4	FAM81A		CENPK
TRNP1		PDE2A	TESC		GINS1
LA16c-380H5.5		MIDN	DUSP2		NDC80

TMEM158		HIST2H2BE	PIK3R3		NMU
ZFYVE28		HID1	TNS3		SGOL1
PINK1		MAP2	ZNF618		KITLG
LGALS1		LURAP1L	CENPQ		ZWINT
		C12orf76	RAB7L1		NUF2
		SERPINB1	DLEU2		CENPA
		RP11-284F21.9	RHOBTB1		WNT10B
		YPEL5	C9orf84		AURKB
		GLRX	MTFR2		ESPL1
		NEU1	FHAD1		HMMR
		RPL32P29	IQGAP2		ARHGEF39
		ANGPTL4	CDK1		MIS18BP1
		L1CAM	GAL		CDCA8
		HERPUD1	TNFSF18		DLGAP5
		HIST2H2AA4	MKX		KIFC1
		PRAF2	KLHL13		PIF1
		ABCC3	NRG4		RP11-424C20.2
		FAM110C	LONRF2		HMGB2
		ORAI3	ITGB3BP		C11orf82
		LDHD	ABCC2		NCAPH
		RP11-509E16.1	CA12		SLC43A3
		SLC33A1	SLC43A3		
		CREB3L2	DPP4		
		BAIAP3	VASH2		
		ICA1	DACT1		
		IFITM10	E2F2		
		PLCD1	ARHGEF39		
		LYRM9	RP11-230G5.2		
		BTG1	MAD2L1		
		LOXL2	SPC25		
		EPPK1	GPRC5B		
		TRNP1	LINC00973		
		CAPS	LRRCC1		
		HGSNAT	APOBEC3B		
		MAP1LC3B	SAMD13		
		PSMG3-AS1	SMAD9		
		PPIC	ERVMER34-1		
		SYVN1	MYT1		
		RP11-284F21.10	MAT1A		
		ARSD	TRIML2		
		TPPP	FAM222A-AS1		
		LAMB2	AL162759.1		
		IRF7	RGCC		
		PLEC	PLCL2		
		RAP2C-AS1	TMSB15A		
		NPTXR	CYP1A1		
		CHPF2	ABCA12		

		MXD1	CENPW		
		ACSS2	RP11-497H16.9		
		RP11-196G18.22	MND1		
		PINK1	KLRC2		
		NUCB2	TUBBP5		
		LTK	KLK6		
		LAMC2	DDN		
		SFN	ALDH1A3		
		ZER1	LONRF3		
		C6orf226	FAM184A		
		ADIRF	CTSC		
		C6orf1	FAM72C		
		TVP23C-CDRT4	TGFB3		
		OPLAH	KCNK5		
		SLC7A11	CAV1		
		LDLR	KIAA0101		
		PAG1	PSRC1		
		DOPEY2	HOXC13		
		EPHB3	HS3ST1		
		KDM5B	CCNA2		
		OPTN	LIMCH1		
		PAQR8	HSPE1		
		DAP	CDC25A		
		PCK2	ADORA1		
		MIR31HG	GNG4		
		RP11-196G18.24	MCM10		
		ST6GALNAC2	TRIM6		
		GNG13	ZNF488		
		LRRC24	DLGAP1		
		CDCP1	S100A3		
		AHNAK	MCM2		
		KIAA0513	DSCC1		
		HIST1H2BC	AADAT		
		BX322557.10	PRIM1		
		LIPH	MRVI1		
		RP1-74M1.3	ZNF532		
		ISG20	SYTL3		
		CORO2B	C1R		
		AP003068.23	PLK1		
		OGDH	NMU		
		ZFYVE1	TNNC1		
		NBEA	BUB1B		
		ZNF554	FAM216A		
		MAML3	DCLRE1A		
		HSPG2	ANKLE1		
		SNX29	MNS1		
		WIPI1	CFHR1		

		STX1A	CDC45		
		ABCD1	FGFBP1		
		MUC12	HS6ST2		
		RP11-284F21.7	SGOL1		
		AP003419.11	37865		
		FAM89B	AC074212.3		
		KCTD21	PBK		
		GRK5	LETM2		
		FGF19	S100A2		
		LGALS1	PRF1		
		SLC23A3	OLFML2B		
		GOLGB1	POLE2		
		ARMCX3	CHRFAM7A		
		PTPRH	CENPK		
		BTBD19	CTC-479C5.12		
		ANO8	EBF2		
		CD68	GINS2		
		SOWAHB	HMMR		
		SERPINB8	ESCO2		
		WDR5B	MTBP		
		CTD-2651B20.3	CDCA3		
		SAT1	PRRT4		
		LINC00205	LCP1		
		TOM1L2	TRAF5		
		SNPH	MYB		
		SQSTM1	RAET1E		
		AQPEP	PHLDB2		
		SLC22A18	KRT222		
		CDK2AP2	NDC80		
		MANF	RFC3		
		ZBED6	KIF14		
		HMGA1	RP11-160O5.1		
		MFGE8	CRABP2		
		ATP10D	FAM64A		
		RGAG4	SEMA3D		
		S100A16	CC2D2A		
		ZSWIM8	FLVCR2		
		WWC3	GINS1		
		FAM114A1	FJX1		
		NOP14-AS1	FAM72D		
		TUBA1A	BARD1		
		KDM7A	ZNF365		
		PPL	HMGA2		
		RP11-96C23.5	AURKA		
		DENND6B	TMEM19		
		FAM228B	TTK		
		PLD1	ANK3		

		ONECUT3	HMGB2		
		CTNNBIP1	HNRNPDL		
		ADM2	SCARA3		
		PLXNA3	ARMC4		
		NLRP1	FDXACB1		
		ITFG3	LIPE		
		ARFGAP3	RAD54L		
		PGM3	UCA1		
		DCLK1	ESPL1		
		FAM127A	FAM72A		
		RHBDF1	ZNF704		
		RAB5B	HAUS1		
		ZNF185	CCDC34		
		CRISPLD2	PRC1		
		PIP5KL1	KIF18B		
		ABHD4	ITGA9		
		IDS	FAM13A		
		TIMP1	MBNL3		
		FKBP14	RAC3		
		PPP2R5B	HIRIP3		
		KLF2	TMPO		
		WFS1	MCM7		
		RP11-53I6.2	NUF2		
		TNXB	RAD54B		
		CAPN5	FANCI		
		LAMB3	P2RX5		
		GPR173	SGPP2		
		VWA5B2	DLGAP5		
		NEAT1	GMNN		
		LINC00657	CDCA5		
		ULK1	TMEM97		
		TMEM158	JADE2		
		PTK6	SMC2		
		PNPLA8	RAB3IL1		
		MAFA	BORA		
		C17orf103	BRIP1		
		ZFYVE28	XRCC3		
		COL4A4	C1orf21		
		SIRT2	FAM83D		
		ABTB1	PCOLCE2		
		DUSP4	IL17RB		
		SEZ6L2	DNAJC12		
		CCDC120	KIF20B		
		FAM46C	POLA1		
		RABIF	NEIL3		
		DNASE1L1	GNAL		
		RPS6KA2	MFSD4		

		PGM1	PKMYT1		
		CALCOCO1	CHRD		
		ARSA	RTKN2		
		MIR24-2	EXO1		
		RGS6	P2RY1		
		SELM	MYPN		
		CRELD1	SERF1B		
		AP000692.10	C11orf82		
		TSPAN31	LOXL3		
		FAM102A	PHF19		
		RAC2	TTPA		
		FAM222A	FANCD2		
		RP11-1275H24.1	FAM198A		
		TP53INP2	FBXO5		
		JUND	SYK		
		GATSL2	SPP1		
		ORMDL2	SKA3		
		HBE1	PIF1		
		REPS2	ARHGDIB		
		HYAL3	STEAP3		
		HSPA5	LYPD5		
		DUSP18	WDHD1		
		NTSR1	CDCA7		
		TRAPPC9	LIN9		
		MIR22HG	KIF22		
		HAP1	ORC6		
		FGF18	MMACHC		
		NANS	PLD6		
		RP11-137L10.6	MCM4		
		TMEM150A	RAD51		
		ALPK3	HJURP		
		LRP11	EREG		
		PNPLA2	LMNB1		
		GRAMD4	KIAA1524		
		FTH1P8	SLC16A6		
		BTBD11	EMP2		
		SRXN1	LAMA2		
		STXBP1	CCNB1		
		RP11-324I22.4	GINS3		
		TBC1D8	RAB15		
		RNF223	BMP4		
		FBXO16	DBNDD2		
		ABCC10	KRT7		
		PPP1R26-AS1	MCM3		
		NHS	ZWINT		
		CRB1	SPC24		
		S100A11	ANLN		

		PPAPDC1B	BDNF		
		MYO1E	NEK2		
		C3	POLQ		
		MSLN	SLC19A1		
		CHST15	NAV3		
		MCF2L	MT-TS2		
		STX5	EXOSC9		
		BNIP3L	FAM132B		
		SESN2	MELK		
		LMF1	MPP7		
		DISP2	TRIP13		
		SLC41A2	C4orf21		
		FTH1	BIRC5		
		GABARAPL1	KNSTRN		
		IRGQ	NCAPH		
		CYP4F11	SFXN4		
		CDH1	CCDC74A		
		AGPAT9	EDNRA		
		PPAP2B	KRR1		
		AC005336.4	TYMS		
		ZSWIM6	KIF2C		
		P4HB	MYBL2		
		MVP	SLC39A8		
		MALAT1	RAI14		
		SQRDL	KIF18A		
		ULBP2	SGOL2		
		PITPNM1	MDFIC		
		TP53I3	RP11-424C20.2		
		IRF6	RACGAP1		
		EXD3	COCH		
		PDLIM7	RP11-114H23.1		
		EIF2AK3	TRPC3		
		PCDH1	CDCA8		
		CTIF	TROAP		
		UPP1	ELFN2		
		KCTD11	ZNF367		
		RP1-151F17.2	SNTB1		
		ZFP36L1	NCAPG		
		TAOK3	FAM72B		
		MID2	KLF11		
		PYGB	SLC25A48		
		EFNB2	UBE2T		
		CRAT	TOP2A		
		CBR3-AS1	DEPDC1		
		BCL6	CCDC58		
		C14orf79	CARD10		
		ALDH2	RASGRP1		

		PTPRM	KIF23		
		NMRK1	MACC1		
		MYO7A	ANO4		
		BACE1	DEPDC1B		
		ERLEC1	SKP2		
		ATP2A3	MIS18BP1		
		PHLDA2	CENPU		
		CNFN	AURKB		
		TMEM125	DUT		
		SH3BGRL3	RIBC2		
		UGDH-AS1	PRPS2		
		RWDD2A	NUSAP1		
		SYNC	FANCB		
		BSDC1	FRK		
		KRCC1	PAICS		
		TCP11L2	SLC9B2		
		ZNF425	CCDC138		
		RP4-694A7.2	CEP152		
		CTD-2521M24.9	SNRPG		
		SERINC2	CDC25C		
		CREB3	KIF15		
		ZCCHC24	HAUS7		
		ITPRIPL2	MCM5		
		SYTL1	RP11-705C15.2		
		IDUA	SRSF7		
		EHD2	MRTO4		
		EPHA2	WDR4		
		ANXA5	EFNA2		
		MYO1D	PRKCQ		
		FNDC3B	EOMES		
		PSAP	ZNF695		
		CYP4F3	HES1		
		PGPEP1	BANK1		
		TBC1D16	METTL7A		
		KLHL24	CLCN1		
		RP11-163N6.2	NUP107		
		TMTC2	ORC1		
		GATS	TMEM98		
		RP11-268J15.5	KIF11		
		FTL	XRCC4		
		MFSD11	DNA2		
		CTB-25B13.12	EPHB6		
		ST3GAL3	C1D		
		DLGAP4	36951		
		PANX2	MRM1		
		CCDC30	RP11-253E3.3		
		LMLN	GPM6A		

		GFPT1	WNT16		
		EPAS1	BRIX1		
		EXOC6B	WDR76		
		GDPGP1	CNTRL		
		OS9	PXN-AS1		
		TTC3	KIAA1009		
		TGFA	PRKCQ-AS1		
		PLAUR	RANBP1		
		SEMA4B	TCOF1		
		DAAM1	OBSL1		
		HIST1H1C	RP11-22B23.1		
		RP11-379H18.1	RGS3		
		SLC25A53	LINC00669		
		AC004466.1	LAD1		
		ABCA5	PFDN4		
		KHNYN	RP11-386G11.10		
		ZSWIM5	METAP1D		
		RP11-319G6.1	CXXC4		
		GAS6-AS2	MS4A15		
		EPHA10	FSD1		
		TEX14	NDC1		
		PTK2B	HNRNPD		
		CHMP1B	MMS22L		
		PIAS3	ASB9		
		GTF2IRD2B	KIF20A		
		G6PD	TPX2		
		PLD3	ARHGAP11A		
		HBEGF	DDX11		
		TPT1-AS1	ADORA2A		
		FAM115C	CNTNAP2		
		SFXN3	RP11-872D17.8		
		LINC00883	UHRF1		
		RP5-1148A21.3	PRKACB		
		SAMD5	CNTLN		
		GPRC5A	WDR90		
		RP11-67L2.2	LIMD2		
		GJB4	DGAT2		
		TRIM8	HAS3		
		ELK3	PARPBP		
		DNAJC3	ANKRD29		
		ENTPD7	ATAD2		
		RRBP1	BLM		
		MOB3A	ZNF267		
		INSIG1	FAR2		
		POLD4	C2orf88		
		HERC1	NME1		
		KSR1	UBE2C		

		ZNF70	EEF1E1		
		SGSM1	ZNF595		
		C16orf45	ERCC6L		
		KRBA2	DHODH		
		PLK3	CDCA2		
		EPG5	TIPIN		
		JUP	LRIG3		
		TUSC3	VSNL1		
		ZSWIM4	KIFC1		
		TNKS1BP1	HMGB1P5		
		CETN2	TNFRSF18		
		TLN2	KNTC1		
		SLC22A17	TYRO3		
		HYOU1	NETO2		
		AC093673.5	IFIT3		
		RP11-166P13.3	SLIRP		
		TSPYL4	HTR7		
		RP11-61L23.2	CENPL		
			ICA1L		
			HOXC5		
			NLRP4		
			NUDCD1		
			PNPT1		
			MKI67		
			ZBTB2		
			DUSP19		
			DMC1		
			RP11-586D19.1		
			GTF3C6		
			SASS6		
			ZNF239		
			CENPF		
			ARHGAP19		
			ADCY3		
			SCNN1A		
			CENPP		
			FAM131C		
			AC005329.7		
			ECT2		
			ASPM		
			ANKRD32		
			RAD51D		
			RADIL		
			RP11-540A21.2		
			ZNF888		
			DTL		
			VRK1		

			FAM96A		
			GTSE1		
			WEE1		
			BTN3A2		
			GIN1		
			CGNL1		
			SLC1A3		
			KCTD12		
			RPF2		
			FANCG		
			LRR1		
			DCK		
			ZNF124		
			KCNQ4		
			FBN1		
			NUP35		
			MGME1		
			SHCBP1		
			NFYB		
			FAM213A		
			FGD3		
			CLMP		
			CDKN3		
			GPR160		
			CPED1		
			MSH2		
			TBC1D31		
			CCDC86		
			DCTPP1		
			ICAM3		
			SH3TC2		
			SLC37A2		
			SETMAR		
			CHN1		
			OR51I1		
			GCSH		
			ANP32E		
			FOXM1		
			RRM2		
			CENPA		
			DDX12P		
			CENPE		
			CDC7		
			SCN5A		
			GATA3		
			PIK3AP1		
			PRR22		

			DUSP23		
			RP11-480I12.5		
			TUBA1B		
			MCM8		
			SNHG17		
			ADAM21		
			PLK4		
			CTD-3148I10.9		
			SSFA2		
			PTTG1		
			TICRR		
			KIF24		
			ZWILCH		
			STARD8		
			AMER1		
			HAT1		
			TPRKB		
			SUV39H1		
			MIS18A		
			POLR3K		
			ATAD5		
			MSI1		
			TAF1A		
			C19orf40		
			CMPK2		
			RBMX		
			GSG2		
			POLR3G		
			CHAF1A		
			RP11-818F20.5		
			CEP55		
			TCERG1		
			CCNE2		
			C1orf112		
			ZNF273		
			RCC1		
			RRM1		
			COQ3		
			LRRC49		
			TRIP6		
			EED		
			MAGOHB		
			CD83		
			CENPM		
			SLC27A5		
			PPIH		
			TOMM5		

			SLC16A14		
			NR5A2		
			C2orf70		
			HDGF		
			POLR2K		
			ATAD3B		
			NTHL1		
			SUSD2		
			COA7		
			PHF5A		
			PSMC3IP		
			PDCD2L		
			RP11-874J12.4		
			ZNF850		
			HLX		
			TRABD2A		
			FRMPD3		
			CCNB2		
			NELFCD		
			MCM6		
			CACYBP		
			CCNF		
			HMGB1		
			TRIM59		
			DTYMK		
			TRAF2		
			GAREML		
			HSPA14		
			GEMIN6		
			PLEKHG2		
			NRG2		
			SNRPEP4		
			GPR19		

8.60 Upregulated features common to multiple comparison conditions, determined using iVenn.

1 x GI_{50} vs DMSO and 3 x GI_{50} vs DMSO	1 x GI_{50} vs DMSO and 1 x GI_{50} vs 3 x GI_{50}	3 x GI_{50} vs DMSO and 1 x GI_{50} vs 3 x GI_{50}
Total: 50	Total: 0	Total: 7
TM9SF1		ASNS
BOLA2		CHAC1
SHPK		FAM129A
CLIC3		PNRC1
WNT9A		KDM5B
MT2A		FAM228B
TNFRSF6B		ADM2

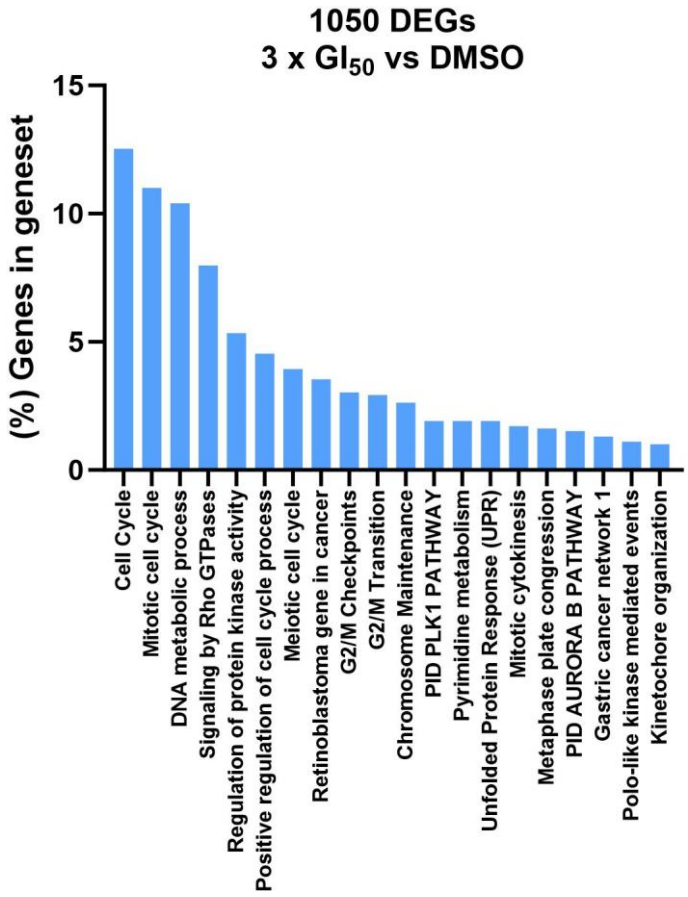
SMIM14		
AHNAK2		
ABLIM2		
SCN9A		
TM7SF2		
ITGB4		
ITGA2		
AGPAT4		
MAP2		
MT1E		
MT1X		
H6PD		
THSD4		
ABCA3		
ITGA3		
TUBA1A		
SEC24D		
HID1		
NEU1		
ADIRF		
NDRG1		
TLN2		
C6orf1		
CCDC92		
FAM110C		
GLRX		
L1CAM		
LOXL2		
ABCC3		
SFN		
FICD		
RHOF		
NPTXR		
SMPD1		
EPPK1		
EVI5L		
ITFG3		
TRNP1		
LA16c-380H5.5		
TMEM158		
ZFYVE28		
PINK1		
LGALS1		

8.61 Downregulated features common to multiple comparison conditions, determined using iVenn.

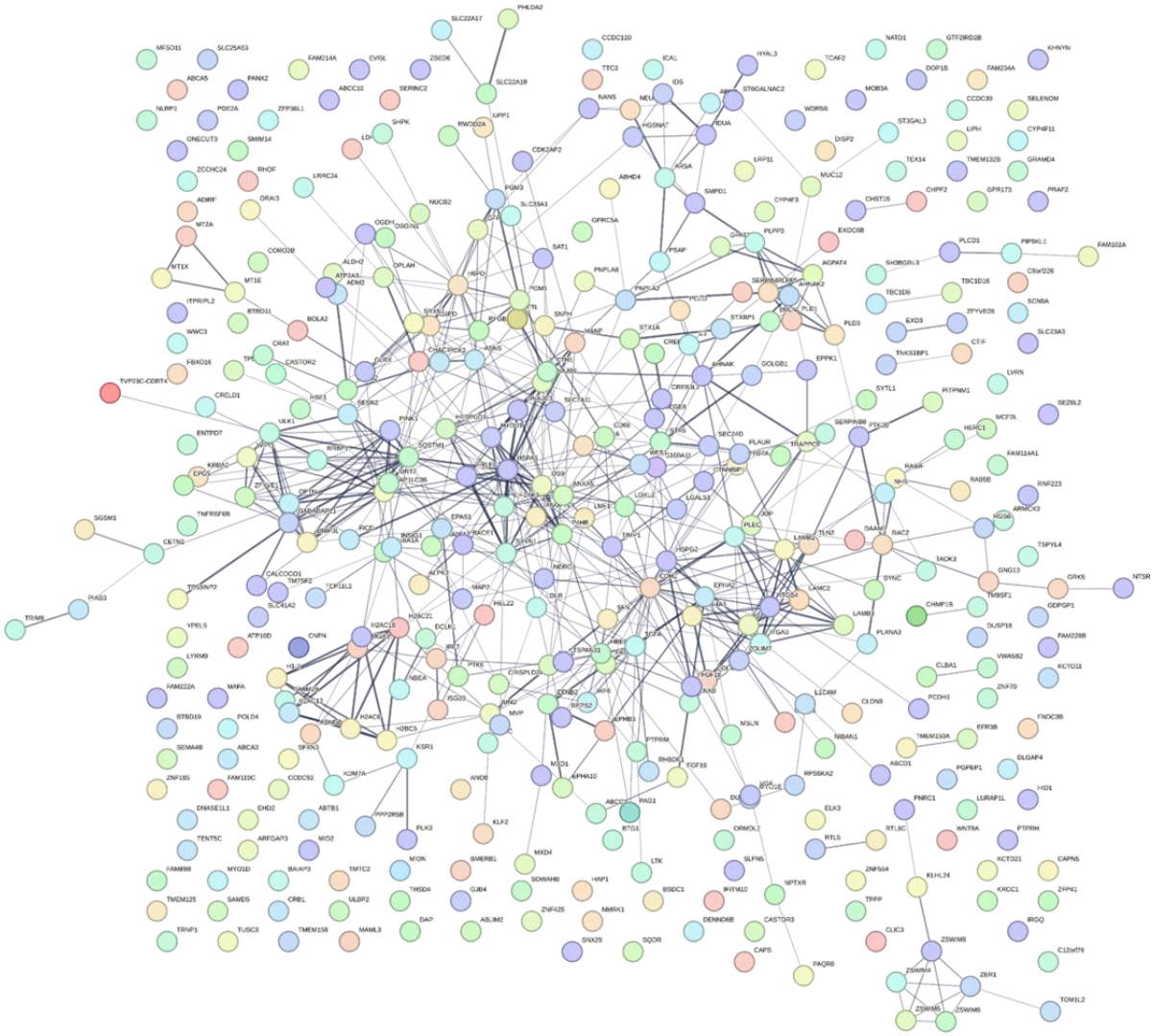
1 x GI_{50} vs DMSO and 3 x GI_{50} vs DMSO	1 x GI_{50} vs DMSO and 1 x GI_{50} vs 3 x GI_{50}	3 x GI_{50} vs DMSO and 1 x GI_{50} vs 3 x GI_{50}
Total: 38	Total: 0	Total: 68
SYT13		DHRS3
KIAA1199		RAD51AP1
GRIN2B		E2F8
SCG2		ZBED2
MRAP2		OAS3
DACT1		CENPQ
ANXA10		DLEU2
BDH2		MTFR2
TLR3		CDK1
RASSF2		SLC43A3
GAL		E2F2
TNFSF18		ARHGEF39
KRT80		MAD2L1
AL162759.1		SPC25
MKX		TMSB15A
SLC7A2		CENPW
RP11-551L14.1		MND1
ACSL5		KIAA0101
KRT7		PSRC1
ARL14EPL		CCNA2
PRIMA1		CDC25A
IL18R1		MCM10
DPP4		DSCC1
HS3ST1		PRIM1
KLK6		PLK1
LONRF2		NMU
FAM198B		BUB1B
NLRP11		CDC45
SLC38A5		SGOL1
RAB7L1		PBK
RBM24		CENPK
DGKE		GINS2
BANK1		HMMR
IQGAP2		ESCO2
HNF4G		CDCA3
KLF11		MYB
TRANK1		NDC80
ZNF618		FAM64A
		GINS1
		TTK
		HMGB2

		RAD54L
		ESPL1
		NUF2
		RAD54B
		DLGAP5
		GMNN
		CDCA5
		XRCC3
		PKMYT1
		EXO1
		C11orf82
		PIF1
		RAD51
		ZWINT
		SPC24
		ANLN
		POLQ
		NCAPH
		RP11-424C20.2
		CDCA8
		MIS18BP1
		AURKB
		RIBC2
		UBE2C
		KIFC1
		RRM2
		CENPA

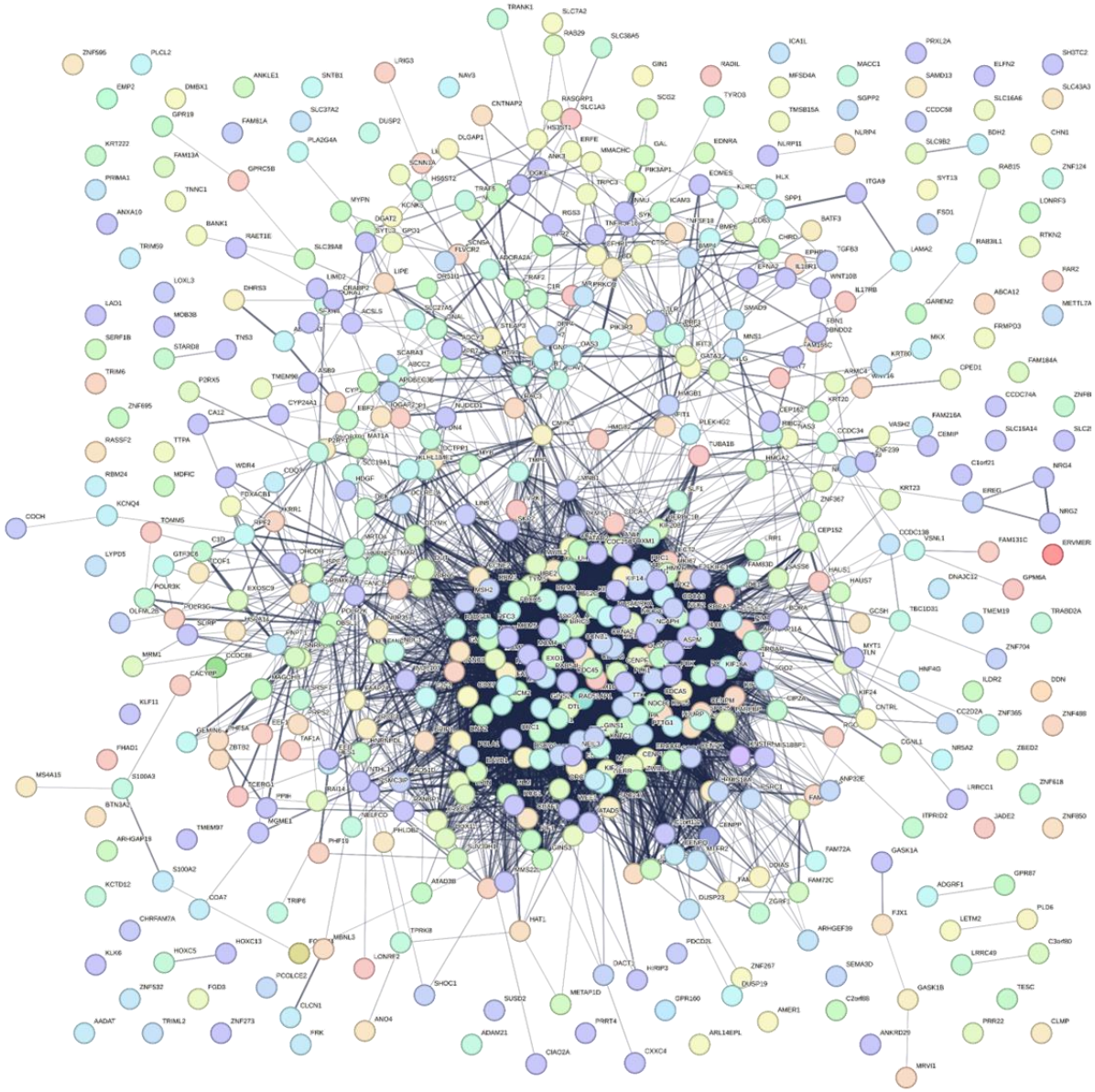
8.62 Top 20 enriched genesets identified using Metascape analysis for the genes identified as significant for the 3 x GI50 vs DMSO comparison, complete analysis of both up and downregulated genes.



8.62.1 STRING analysis network for the upregulated genes identified in the 3 x GI_{50} vs DMSO comparison.



8.63 STRING analysis network for the downregulated genes identified in the 3 x GI_{50} vs DMSO comparison.



8.64 Biological processes identified as functional enrichments when examining the downregulated features identified in the 3 x GI50 vs DMSO comparison.

Term description	Observed gene count	Background gene count	Strength	False discovery rate
Cell cycle	166	1246	0.67	9.06e-57
Cell cycle process	122	835	0.71	1.51e-43
Mitotic cell cycle	103	631	0.76	6.94e-40
Mitotic cell cycle process	93	537	0.78	1.55e-37
Cell division	84	527	0.75	3.06e-31
Chromosome segregation	64	286	0.89	1.84e-30
Regulation of cell cycle process	95	716	0.67	3.33e-30

Nuclear chromosome segregation	57	229	0.94	5.06e-29
Nuclear division	64	323	0.84	5.63e-28
Organelle fission	65	349	0.81	3.92e-27
Regulation of cell cycle	112	1108	0.55	8.36e-27
Dna metabolic process	93	785	0.62	2.71e-26
Chromosome organization	102	968	0.57	1.94e-25
Sister chromatid segregation	40	144	0.99	2.09e-21
Regulation of mitotic nuclear division	37	118	1.04	3.24e-21
Mitotic sister chromatid segregation	37	122	1.03	7.99e-21
Regulation of mitotic cell cycle	66	493	0.67	1.47e-20
Mitotic nuclear division	42	175	0.92	1.47e-20
Regulation of nuclear division	38	145	0.96	1.29e-19
Dna replication	43	203	0.87	2.48e-19
Regulation of cell cycle phase transition	60	431	0.69	2.48e-19
Dna repair	63	497	0.65	1.63e-18
Positive regulation of cell cycle process	44	251	0.79	4.57e-17
Dna-templated dna replication	34	137	0.94	7.74e-17
Cellular response to dna damage stimulus	74	744	0.54	1.44e-16
Regulation of mitotic cell cycle phase transition	48	332	0.7	9.24e-16
Regulation of chromosome segregation	31	128	0.93	5.25e-15
Negative regulation of cell cycle process	42	272	0.73	1.60e-14
Meiotic cell cycle	40	250	0.75	3.06e-14
Regulation of chromosome separation	28	109	0.95	5.10e-14
Double-strand break repair	36	204	0.79	7.20e-14
Positive regulation of cell cycle	46	349	0.66	9.75e-14
Dna recombination	38	235	0.75	1.22e-13
Meiotic nuclear division	33	171	0.83	1.26e-13
Meiotic cell cycle process	34	188	0.8	2.38e-13
Regulation of mitotic sister chromatid segregation	21	55	1.13	4.57e-13
Negative regulation of mitotic nuclear division	19	41	1.21	6.23e-13
Regulation of chromosome organization	38	253	0.72	8.31e-13
Negative regulation of cell cycle	45	359	0.64	8.31e-13

Dna conformation change	26	104	0.94	8.31e-13
Negative regulation of mitotic sister chromatid separation	18	35	1.25	8.31e-13
Dna geometric change	25	95	0.96	9.74e-13
Regulation of mitotic sister chromatid separation	25	96	0.96	1.17e-12
Nucleic acid metabolic process	129	2203	0.31	2.52e-12
Cell cycle checkpoint signaling	30	157	0.82	2.56e-12
Regulation of mitotic metaphase/anaphase transition	24	91	0.96	2.98e-12
Negative regulation of cell cycle phase transition	35	225	0.74	3.45e-12
Negative regulation of mitotic metaphase/anaphase transition	17	33	1.26	3.61e-12
Negative regulation of chromosome organization	23	84	0.98	5.07e-12
Mitotic cell cycle checkpoint signaling	27	127	0.87	5.33e-12
Regulation of sister chromatid segregation	25	105	0.92	5.35e-12
Microtubule cytoskeleton organization	54	542	0.54	6.13e-12
Dna duplex unwinding	23	89	0.96	1.30e-11
Spindle organization	29	160	0.8	1.83e-11
Mitotic spindle assembly checkpoint signaling	16	31	1.26	1.83e-11
Negative regulation of mitotic cell cycle	33	214	0.73	1.97e-11
Organelle organization	171	3470	0.24	7.31e-11
Meiosis i	25	126	0.84	1.58e-10
Microtubule cytoskeleton organization involved in mitosis	25	129	0.83	2.43e-10
Regulation of dna-templated dna replication	18	57	1.04	3.72e-10
Nucleobase-containing compound metabolic process	141	2722	0.26	6.23e-10
Regulation of dna replication	25	136	0.81	6.49e-10
Microtubule-based process	63	803	0.44	6.49e-10
Dna replication initiation	14	29	1.23	1.11e-09
Negative regulation of mitotic cell cycle phase transition	27	167	0.75	1.20e-09

Heterocycle metabolic process	146	2891	0.25	1.31e-09
Nuclear dna replication	13	24	1.28	2.04e-09
Mitotic spindle organization	21	97	0.88	2.30e-09
Regulation of organelle organization	79	1190	0.37	2.43e-09
Meiotic chromosome segregation	21	98	0.87	2.65e-09
Cellular aromatic compound metabolic process	146	2936	0.24	3.69e-09
Double-strand break repair via homologous recombination	22	114	0.83	5.01e-09
Regulation of dna metabolic process	48	541	0.49	5.56e-09
Dna unwinding involved in dna replication	12	22	1.28	1.09e-08
Non-membrane-bounded organelle assembly	35	314	0.59	1.12e-08
Organic cyclic compound metabolic process	152	3181	0.22	2.01e-08
Positive regulation of cell cycle phase transition	21	115	0.8	3.25e-08
Positive regulation of mitotic cell cycle phase transition	19	92	0.86	3.86e-08
Cell cycle g2/m phase transition	16	62	0.95	6.25e-08
Spindle assembly	19	96	0.84	7.08e-08
Cell cycle phase transition	25	183	0.68	1.43e-07
Cellular component organization	223	5436	0.16	1.75e-07
Cellular response to stress	89	1572	0.3	2.23e-07
Regulation of cell division	25	188	0.67	2.28e-07
Double-strand break repair via break-induced replication	9	12	1.42	3.29e-07
Mitotic dna replication	9	12	1.42	3.29e-07
Cellular component organization or biogenesis	228	5639	0.15	3.48e-07
Positive regulation of mitotic cell cycle	20	121	0.76	3.65e-07
Chromosome localization	17	84	0.85	4.01e-07
Regulation of cell cycle g2/m phase transition	19	111	0.78	5.40e-07
Regulation of g2/m transition of mitotic cell cycle	18	99	0.8	5.98e-07
Metaphase plate congression	15	65	0.91	7.25e-07
Regulation of cellular component organization	117	2365	0.24	7.39e-07

Establishment of chromosome localization	16	77	0.86	8.09e-07
Mitotic cell cycle phase transition	23	173	0.67	9.38e-07
Kinetochore organization	10	21	1.22	9.72e-07
Centromere complex assembly	11	29	1.12	1.12e-06
G2/m transition of mitotic cell cycle	14	58	0.93	1.37e-06
Cytoskeleton organization	72	1229	0.31	2.45e-06
Interstrand cross-link repair	12	41	1.01	2.52e-06
Homologous recombination	14	63	0.89	3.29e-06
Negative regulation of organelle organization	32	351	0.5	4.96e-06
Reciprocal meiotic recombination	13	56	0.91	6.58e-06
Regulation of cytokinesis	16	94	0.77	8.69e-06
Cellular response to stimulus	243	6357	0.13	1.02e-05
Mitotic cytokinesis	15	84	0.79	1.26e-05
Cellular nitrogen compound metabolic process	150	3463	0.18	1.33e-05
Regulation of dna-templated dna replication initiation	8	15	1.27	1.49e-05
Positive regulation of cell cycle g2/m phase transition	10	31	1.05	1.69e-05
Homologous chromosome segregation	13	63	0.86	2.03e-05
Chromosome organization involved in meiotic cell cycle	14	76	0.81	2.32e-05
Mitotic metaphase plate congression	12	53	0.9	2.51e-05
Female meiotic nuclear division	10	34	1.01	3.37e-05
Protein localization to chromosome	15	92	0.76	3.38e-05
Negative regulation of cellular component organization	46	691	0.37	3.71e-05
Cellular macromolecule metabolic process	115	2512	0.2	4.64e-05
Regulation of transferase activity	55	912	0.32	5.01e-05
Regulation of attachment of spindle microtubules to kinetochore	8	19	1.17	5.67e-05
Positive regulation of g2/m transition of mitotic cell cycle	9	28	1.05	6.88e-05
Spindle elongation	7	13	1.27	8.21e-05

Regulation of cell cycle checkpoint	11	49	0.89	8.21e-05
Regulation of biological process	390	11655	0.07	9.62e-05
Centrosome cycle	14	88	0.74	0.0001
Mitotic spindle assembly	11	51	0.88	0.00011
Regulation of cellular process	372	11025	0.07	0.00013
Dna strand elongation involved in dna replication	7	15	1.21	0.00016
Sister chromatid cohesion	10	42	0.92	0.00016
2-deoxyribonucleotide biosynthetic process	7	15	1.21	0.00016
Kinetochore assembly	7	16	1.18	0.00022
Positive regulation of cellular process	211	5584	0.12	0.0003
Attachment of spindle microtubules to kinetochore	8	26	1.03	0.00035
Mitotic dna replication initiation	5	5	1.54	0.00039
Mitotic spindle midzone assembly	6	11	1.28	0.00044
Positive regulation of cell cycle checkpoint	7	19	1.11	0.00052
Positive regulation of chromosome separation	8	28	1	0.00052
Positive regulation of dna metabolic process	25	304	0.46	0.00062
Positive regulation of mitotic cell cycle spindle assembly checkpoint	6	12	1.24	0.00062
Pyrimidine deoxyribonucleoside triphosphate metabolic process	5	6	1.46	0.00064
Mitotic recombination	7	20	1.09	0.00066
Reproduction	74	1514	0.23	0.00077
Response to stimulus	276	7835	0.09	0.00081
Protein localization to chromosome, centromeric region	9	41	0.88	0.00081
Regulation of mitotic cell cycle spindle assembly checkpoint	7	21	1.07	0.00083
Chromosome condensation	9	42	0.87	0.00093
Dna-templated dna replication maintenance of fidelity	10	54	0.81	0.00093
Cenp-a containing chromatin assembly	5	7	1.4	0.001

2-deoxyribonucleotide metabolic process	9	43	0.86	0.0011
Reproductive process	73	1504	0.23	0.0011
Dna integrity checkpoint signaling	14	113	0.64	0.0011
Cellular process	467	14826	0.04	0.0012
Positive regulation of transferase activity	37	586	0.34	0.0012
Intracellular signal transduction	73	1518	0.23	0.0014
Biological regulation	402	12385	0.05	0.0018
Meiotic spindle organization	6	16	1.12	0.002
Protein localization to kinetochore	6	16	1.12	0.002
Response to stress	135	3358	0.15	0.0022
Meiotic spindle assembly	5	9	1.29	0.0022
Positive regulation of chromosome segregation	7	26	0.97	0.0024
Mitotic chromosome condensation	6	17	1.09	0.0025
Positive regulation of biological process	224	6207	0.1	0.0025
Negative regulation of cell division	6	17	1.09	0.0025
Positive regulation of cell division	12	92	0.66	0.0026
Spindle assembly involved in female meiosis i	4	4	1.54	0.0031
Regulation of cyclin-dependent protein serine/threonine kinase activity	13	111	0.61	0.0034
Positive regulation of protein phosphorylation	42	747	0.29	0.0039
Cellular component assembly	104	2467	0.17	0.0039
Pyrimidine nucleotide metabolic process	9	53	0.77	0.004
Regulation of exit from mitosis	6	19	1.04	0.004
Positive regulation of phosphorylation	45	826	0.28	0.0042
Regulation of molecular function	124	3085	0.15	0.0048
Positive regulation of molecular function	73	1587	0.21	0.0049
Strand invasion	4	5	1.45	0.0051
Positive regulation of protein modification process	52	1018	0.25	0.0053

Mitotic dna integrity checkpoint signaling	11	85	0.66	0.0056
Regulation of protein kinase activity	38	663	0.3	0.0058
Regulation of nitrogen compound metabolic process	207	5734	0.1	0.0058
Female gamete generation	15	154	0.53	0.0059
Regulation of protein phosphorylation	55	1108	0.24	0.0065
Pyrimidine nucleotide biosynthetic process	7	32	0.88	0.0065
Replication fork processing	8	45	0.79	0.0074
Positive regulation of cytokinesis	8	45	0.79	0.0074
Regulation of phosphorylation	60	1251	0.22	0.0075
Female meiosis i	5	13	1.13	0.0076
Positive regulation of dna-directed dna polymerase activity	5	13	1.13	0.0076
Positive regulation of binding	16	181	0.49	0.0095
Dna biosynthetic process	11	92	0.62	0.0099
Pyrimidine nucleoside triphosphate metabolic process	6	24	0.94	0.0107
Pyrimidine deoxyribonucleotide metabolic process	6	24	0.94	0.0107
Regulation of binding	25	375	0.37	0.0109
Pyrimidine deoxyribonucleotide biosynthetic process	4	7	1.3	0.0117
Cellular component biogenesis	109	2702	0.15	0.0117
Positive regulation of mitotic cytokinesis	4	7	1.3	0.0117
Regulation of protein modification process	70	1560	0.2	0.0125
Organelle assembly	42	799	0.26	0.0133
Protein-dna complex assembly	18	229	0.44	0.0139
Telomere maintenance	11	97	0.6	0.0143
Regulation of kinase activity	41	778	0.26	0.0149
Positive regulation of phosphate metabolic process	46	912	0.25	0.0159

Pyrimidine-containing compound metabolic process	10	82	0.63	0.0159
Homologous chromosome pairing at meiosis	8	52	0.73	0.0162
Double-strand break repair via synthesis-dependent strand annealing	4	8	1.24	0.0162
Regulation of macromolecule metabolic process	219	6249	0.09	0.0168
Dna replication checkpoint signaling	5	17	1.01	0.019
Protein kinase a signaling	5	17	1.01	0.019
Regulation of primary metabolic process	208	5899	0.09	0.0191
Histone exchange	6	28	0.87	0.0201
Negative regulation of cellular process	172	4736	0.1	0.0216
Mitotic sister chromatid cohesion	5	18	0.99	0.0232
Immune effector process	24	375	0.35	0.0234
Chromosome separation	6	29	0.86	0.0234
Regulation of dna biosynthetic process	12	124	0.53	0.0268
Double-strand break repair via nonhomologous end joining	7	43	0.75	0.0276
Establishment of organelle localization	24	380	0.34	0.0276
Positive regulation of mitotic sister chromatid separation	5	19	0.96	0.028
Positive regulation of dna biosynthetic process	9	73	0.63	0.0281
Protein-dna complex subunit organization	19	268	0.39	0.0287
Lymphocyte activation	27	457	0.31	0.0324
Regulation of mitotic spindle organization	7	45	0.74	0.0344
Cell population proliferation	37	712	0.26	0.0348
Telomere organization	12	129	0.51	0.0357
Regulation of catalytic activity	95	2370	0.15	0.0357
Lymphocyte activation involved in immune response	12	130	0.51	0.0377
Adaptive immune response based on somatic recombination of immune receptors built from	14	169	0.46	0.0377

immunoglobulin superfamily domains				
Positive regulation of attachment of spindle microtubules to kinetochore	4	11	1.1	0.0379
Mesenchymal cell differentiation	14	170	0.46	0.0394
T cell mediated immunity	6	33	0.8	0.04
Dna replication-independent chromatin assembly	6	33	0.8	0.04
Histone phosphorylation	6	33	0.8	0.04
Dttp biosynthetic process	3	4	1.42	0.0404
Regulation of nucleobase-containing compound metabolic process	149	4074	0.11	0.0436
Negative regulation of g2/m transition of mitotic cell cycle	8	63	0.65	0.0449
Oocyte maturation	5	22	0.9	0.0456
Meiotic chromosome separation	5	22	0.9	0.0456
T-circle formation	4	12	1.07	0.047
Signal transduction in response to dna damage	12	135	0.49	0.0481
Macromolecule metabolic process	201	5781	0.08	0.0483
Mitotic dna damage checkpoint signaling	9	81	0.59	0.0499

Table heading descriptions:

Observed Gene count: Number indicates how many proteins in your network are annotated with a particular term.

Background Gene count: Number indicates how many proteins in total (in your network and in the background) have this term assigned.

Strength: Log10(observed / expected). This measure describes how large the enrichment effect is. It's the ratio between i) the number of proteins in your network that are annotated with a term and ii) the number of proteins that we expect to be annotated with this term in a random network of the same size.

False Discovery Rate: This measure describes how significant the enrichment is. Shown are p-values corrected for multiple testing within each category using the Benjamini–Hochberg procedure.