Oxidative Heterodimerisation Of 4'-Hydroxycinnamate Esters With 4'-Hydroxycinnamic Acids As Potential HIV-1 Integrase Inhibitors And Identification Of Two Novel Homoisoflavonoids With Anti-cancer Potential.

Natural Product Drug Discovery, Medicinal and Computational Chemistry

A thesis submitted for the degree of Master of Philosophy

By

Cameron Garty

IBERS

Institute of Biology, Environmental and Rural Sciences, Aberystwyth University

PRIFYSGOL ABERYSTWYTH UNIVERSITY

August 2019

Declaration

I, Cameron Alexander Garty, declare that this thesis is my own work.

Signed_____ Date_____

Table of contents

List Figures

List of Schemes

List of tables

Abbreviations

1D NMR: One dimensional Nuclear Magnetic Resonance (Spectroscopy)

2D NMR: Two dimensional Nuclear Magnetic Resonance (Spectroscopy)

3'-P: Processing reaction

AIDs: Acquired immune deficiency syndrome

ART: Antiretroviral therapy

BIC: Bictegravir

CAB: Cabotegravir

cART: Combination antiretroviral therapy

CCD: Catalytic core domain

CH_3CN: Acetonitrile

cDNA: Complementary deoxyribonucleic acid

COSY: Correlation Spectroscopy

CTD: C-Terminal Domain

DFA: Dehydrodiferulates

DKAs: Diketo acids

DNA: Deoxyribonucleic acid

DOl: Dolutegravir

ELV: Elvitegravir

EtoAc: Ethyl Acetate

FDA: Food Drug administration

HAART: High active antiretroviral therapy

HIV: Human immunodeficiency virus

HMBC: Heteronuclear multiple bond correlation

HPLC: High performance liquid chromatography-Tandem mass spectrometry

HRESIMS: High-resolution electrospray ionisation mass spectrometry

HSQC: Heteronuclear single quantum coherence

IC_{50}: estimated half maximal inhibitory concentration

IN(I): integrase (Inhibitor)

INSTI: Integrase strand transfer inhibitor

LEDGF: Lens epithelium-derived growth factor

MD: Molecular dynamics

MeOH: Methanol

MIC: Minimum inhibitory concentration

MS: Mass spectrometry

MTT: Thiazolyl Blue Tetrazolium Blue

NMR: Nuclear magnetic resonance

NTD: N-terminal Domain

PDA: Photodiode Array

PFV: Primate Foamy Virus

PI: Protease inhibitor

PIC: Pre-Integration Complex

RAL: Raltegravir

R_f: Retention factor

RNA: Ribonucleic acid

RP: Reversed phase

SAR: Structure activity relationship

ST: Strand Transfer Reaction

TLC: Thin layer chromatography

TMS: Tetrmethylsilane

t_R: Retention Time

UV: Ultraviolet (Spectroscopy)

v/v: volume/volume

VEGF: Vascular endothelial growth factor

WHO: World Health Organisation

WT: Wild Type

λ_{max}: Wavelength of peak

Acknowledgments

Firstly, the author would like to say thanks to my supervisors Dr. Ifat Parveen Shah, Aberystwyth University and Prof. Michael Threadgill, University of Bath, whose supervision has taught me so much, and guidance made it possible to complete this thesis.

With special thanks to Dr. Tim Woodman, University of Bath for all the NMR spectra. Thanks to Dr. Barbara Hauck, Dr. Ana Winters and Dr. Helen Phillips, Aberystwyth University for assistance on HPLC mass spectrometry analysis. Thanks to Prof. Robert Nash and Mrs Hazel Sharp, Phytoquest Ltd, for the training on chromatography techniques and the use of the analytical equipment. Thanks to Prof. Andrea Brancale and Dr. Salvatore Ferla, Cardiff University School of Pharmacy and Pharmaceutical Science, for computational chemical analysis for chapter 3.

Thanks to Aberystwyth University and to the Joy Welch Foundation for providing the facilities and funding, in giving me this fantastic opportunity.

The Author would also like to thank my family, Sarah Garty, Thomas Garty, Njal Skinner, Ross Garty, Colin and Nita Porter, and finally to my partner Louise Long for their continuous help, love and support.

Chapter 1

General Introduction – Natural product drug discovery

1.0 *Introduction*

Drug discovery involves isolation and structural elucidation, to develop effective drug candidates to safely treat disease with little or no side effects, which is a huge challenge. Even though there has been substantial development of drugs for treating and managing diseases such as cancer, diabetes and HIV/AIDS, these diseases continue with significant mortalities. A current viable approach will be to revert back to nature for continued drug discovery. Over the past three decades, target based approaches has been the focus of commercial drug discovery (Zheng *et al.*, 2013). It works by adapting the performance of molecular targets that have a key role within the disease (Yoshida, 2019). On the other hand, phenotypic screening is used when there is no known knowledge of the target molecule, an assay based technique is used to represent a disease model, and it is still widely used (Zheng, Thorne and McKew, 2013; Yoshida, 2019). Both techniques reply on a large drug library, which may include natural products that are currently in clinical practise today. With 1,562 new chemical entities introduced between 1981-2014, 26% were isolated from natural products. A further 11% were synthetic, but with a natural product pharmacophore design (Newman and Cragg, 2012). Natural products equate for a quarter of all discovered drugs from 1981-2014 (Figure 1-1).

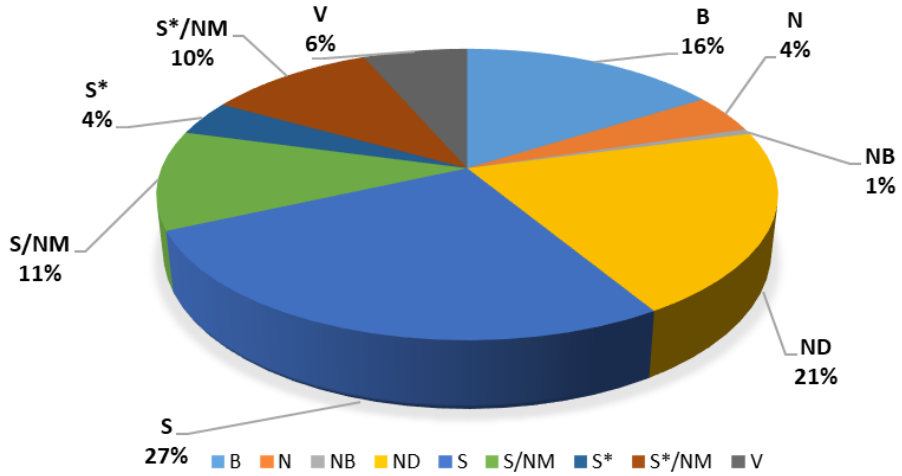


Figure 1-1: Chart displaying new chemical entities introduced between 1981-2014. S: synthetic. S*: synthetic – pharmacophore derived from a natural product. NM: mimic of natural product. V: vaccine. B: biological. N: unchanged natural product. NB: botanical drug. ND: natural product derivative. Information compiled from Newman and Cragg (2016)

1.1 *Natural product drug development*

Natural products offer a range of compounds with different chemical characteristics, which can be problematic to replicate synthetically. Natural products offer multiple potential advantages, including; stronger fused and bridging ring systems, a higher number of chiral centres and resulting in structural diversity (Diggers *et al.*, 2008; Drewy and Macarron, 2010). In addition, natural products tend to be more hydrophilic, facilitating uptake within cells and having higher affinity, requiring minimal modification (Ganesan, 2008; Drewy and Macarron, 2010). 50 % of natural products are isolated from plants, with an ever increasing proportion from marine and microorganisms (Havery, 2008) (Table 1-1). Plants are in every habitable environment, faced with stresses and challenges of evolution, therefore plants have developed vast libraries of molecules for protection biotic and abiotic factors (Weng, Philippe and Noel, 2012). It has been estimated that higher plants collectively synthesise 100,000 secondary metabolites (AGI, 2000), with only 10% of the higher plants having ever been screened for bioactivity (Gransalke, 2011).

Table 1-1: Biologically active drugs derived from natural products.

Natural product	**Drug name**	**Source**
Plant	Morphine	*Papaver sommiferum*
	Paclitaxel	*Taxcus brevifolia*
	Aspirin	*Salix alba* (willow)
	Pilocarpine	*Pilocarpus jaborandi* (Rutaceae)
	Withaferin	*Withania somnifera*
Bacteria	Mitomycin C	*Streptomyces caespitosus*
	Ixabepilone	*Sorangium cellulosum* (synthetic analogue)
Fungus	Cytochalasin B	*Helminthosporium dematioideu*
	Phalloidin	*Amanita phalloides* (death cap mushroom)
Marine	Dolastatin	*Dolabella auricularia* (marine mollusc)
	Jasplakinolide	*Jaspis johnstoni*
	Kabiramide	*Ircinia sp.*

When a natural extract has been reported biologically active, the conventional approach is bioassay-guided fractionation. This process takes place involving a five step process: (i) solvent extraction to obtain metabolites from the biomass, (ii) chromatography fractionation of the extracted metabolites, (iii) biological assay of each fraction, (iv) isolation of active molecules, (v) structural elucidation of the isolated compounds and evaluation of biological activity of the purified molecule (Colegate and Molyneux, 1993; Newman and Cragg, 2012; and Bucar, Wube and Schmid, 2013). This method was scientifically accepted in the 1950's (Schnider *et al*., 1952; Wekker, 2012). As a result, hundreds of enzymes were purified and identified during 1950s-1960s, allowing enzyme kinetic protocols to be developed for receptor pharmacology (Foreman, Johansen, 1996). Enzymes are an ideal target for drug discovery, due to the available drug interactions on the receptors (Lundstorm and Chiu, 2006; Leifert, 2009), demonstrating the significant importance of molecular receptors for drug development (Segel, 1975; Zheng *et al*., 2013).

Moreover, genome science introduced the target-based approach in the 1980s, where a new generation of assays were developed with the use of recombinant DNA. This method allowed rapid screening of large chemical libraries (Imming *et al*., 2006; Rask-Andersen *et al*., 2011). In conjunction with, assay miniaturisation and robotic automation and further recent advancements of lead identification, by x-ray crystallography, computational chemistry modelling and virtual screening (Swinney and Anthony, 2011), these methods have allowed fast high-throughput screens for activity of a range of compounds (Diller, 2008; Pereira and Williams, 2007; Zheng *et al*., 2013).

Drug discovery has always been a very expensive process. The average cost to develop a new drug over 10-12 year period is around $1 billion, when the losses of unsuccessful drugs are included it is estimated to be around $2.5 billion (Mullin, 2014; Sertkaya *et al*., 2016). The elevated cost in development is due to complex clinical trials, insurance policies and the greater focus of developing therapies for chronic diseases such as Alzheimer's, Parkinsons and Schizophrenia, where animal models are not comparative, and trials require a greater number of candidates to be followed for an extensive period of time. (DiMasi *et al*., 2003; Mullin, 2014; Zheng *et al*., 2016). The overall success rate from Phase I to food and drug administration (FDA) approval has previously estimated between 9.6 and 10.4% reported by Hay *et al.*, (2014) and

Thomas *et al*., (2016) respectively. However, a recent study by Massachusetts Institute of Technology reported that 13.8% of all drug development programs eventually lead to approval (Wong, Siah and Lo, 2019). Multiple unsuccessful results are not published, reasons for these failures are (1) bioactive compounds were degraded during the purification process, (2) bioactive compounds were at low concentrations for efficient purification, (3) bioactivity was due to the synergistic effect between multiple compounds (Nothias *et al*., 2018).

Meanwhile, the number of new chemical entities has fallen from an average of 51 per year (1981-2000), down to around 40 per year (2001-2010) (Newman and Cragg, 2012). A recent paper highlighted an increase in new chemical entities from 2011 – 2014 to around 50 per year (Newman and Cragg, 2016). The major increase in new chemical entities arose between 2013, where it doubled to 60 per year, 42% of these new compounds were characterised into the biological/vaccine classifications. The increase was due to approved vaccines for the avian influenza outbreak (Newman and Cragg, 2016).

1.2 *Traditional medicine systems*

Western medicine is inaccessible to many nations due to location and financial issues. In 2002, the world health organisation estimated that 60-80% of people in developing countries reply on traditional medicine for primary healthcare (WHO, 2002). Traditional medicine has a strong cultural belief, with over 2000 traditional medicine practitioners out numbering western medicine practitioners 100:1. The traditional healthcare system comprises of Ayurveda, Siddha, Yoga and Homeopathy, which is supported by 1.5 million practitioners, with Ayurveda being most widely used. There is a strong belief that natural therapies are safer than western medicine (WHO, 2007; Mukherjee and Wahile, 2006). With 80 % of India's population relying on traditional medicine for health care (WHO, 2002), it has resulted in over 25,000 herbal medicinal preparations on the Indian market. These herbal preparations are used for therapy of multiple diseases and conditions, however only 6% of species used have been extensively investigated (Mukherjee and Wahile, 2006). Additionally, over 40% of Chinese population from rural and urban areas prefer to use traditional Chinese medicine (Qi *et al*., 2011), resulting in the Chinese population spending around £56 billion on traditional medicine which is equivalent to 36% of the total national health care (LCB, 2014). Alternative therapies are also popular within developed nations,

with 40% of Americans having used traditional medicine in the last 12 months (Barnes *et al*., 2008).

Pioneering drug discovery starts by developing knowledge from traditional natural products for effective treatments of diseases. The significance of natural products in providing advanced drugs with unique characteristics cannot be undermined. The 2015 Nobel Prize for Physiology or Medicine went to Prof. Tu Youyou and Drs. Omura and Campbell for their discovery and development of artemisinin and avermectin respectively (Newman and Cragg, 2016).

Advances within technology have made it possible to understand the complexity of natural products. This has resulted in a substantial portfolio of isolated and synthesized compounds from natural products, to successful lead compounds that are currently clinically active. Natural product drug discovery has become a very successful approach for the advancement of novel therapeutic drugs. With the use of advancing scientific technology, new therapeutic moieties will be discovered. Consequently, natural product drug discovery stands to be a major provider to solving global health issues, such as; resistance and side effects, and achieving sustainable development for global health.

1.3 *Outline of the thesis*

Chapter 1: General introduction

Contains background for understanding the developments of natural product drug discovery.

Chapter 2: Oxidative heterodimerization of 4'-hydroxycinnamate esters with 4'-hydroxycinnamic acids, potential HIV-1 integrase inhibitors.

Contains background information on the following; HIV/AIDS, antiretroviral therapy/drugs, structure and function of HIV-1 integrase, structure activity relationships of HIV-1 integrase inhibitors, polyhydroxylated aromatic HIV-1 inhibitors. Materials and methods that apply to this chapter. Results of the isolated compounds and discussion of the purified and structurally elucidated compounds.

Chapter 3: Computational docking of potential hydroxycinnamate ester derivatives with the structure of HIV-1 integrase.

Contains background information on computational techniques and developments. Materials and methods that apply to this chapter. Results and discussion of the computational analysis and conclusion from the docking models.

Chapter 4: The identification of two novel homoisoflavonoids with anti-cancer properties from plant X.

Contains background information on the following; breast cancer, current treatments for breast cancer, drug discovery from natural products. Materials and methods that apply to this chapter. Results of analysis of the extracts and discussion of the impact of the results in this chapter in the context of the literature and the conclusion.

Chapter 5: General conclusion

Contains a brief overall conclusion of all results in this thesis.

Chapter 2

Oxidative heterodimerization of 4'-hydroxycinnamate esters with 4'-hydroxycinnamic acids as potential HIV-1 integrase inhibitors.

2.0 *Abstract*

Integrase strand-transfer inhibitors (INSTIs) have become essential in the treatment of HIV. They are the newest class of antiretroviral drugs. The three currently in clinical use are raltegravir, dolutegravir and elvitegravir. However, therapy is only successful as long as the virus is not resistant against the mode of action of the drug. The current treatments share a high degree of cross-resistance and have a low genetic barrier to resistance, which generated the development of the second-generation inhibitors, cabotegravir and bictegravir. Lithospermic acid, a caffeic trimer isolated from *Salvia miltiorrhiza*, has been shown to be a highly potent non-toxic HIV-1 integrase inhibitor. The synthesis of four different 4'-hydroxycinnamate esters coupled with 4'-hydroxycinnamate acids creates a portfolio of structurally similar compounds, potentially possessing similar pharmacophores to lithospermic acid. The homodimers and heterodimers were identified by high performance liquid chromatography - electrospray ionisation - tandem mass spectrometry (HPLC-UV-ESI-MS/MS) and their structures were elucidated by 1D and 2D NMR. Development of a purification method proved challenging, giving very low yields. The isolated compounds will be tested for inhibition of HIV-1 integrase and other biological activities.

2.1 *Introduction*

2.1.1 *HIV epidemiology, polymorphism*

To date, 78 million people have become infected with HIV and over 35 million deaths have resulted from HIV/AIDS and related diseases. Recent analysis from United Nations AIDS (UNAIDS) has shown that 36.9 million people are currently living with HIV globally, of whom 1.8 million are under the age of fifteen and a further 1.8 million [1.6 – 2.1 million] are newly infected individuals (Figure 2-1). UNAIDS also report an 18.4% rise in HIV cases since 2005, with a 47% drop in HIV-related mortality. The drop is due to the improvement of services allowing widespread access to antiretroviral therapy (ART). The World Health Organisation (WHO) has also revealed a supporting correlation, in that the number of people with consistent access to ART has also

multiplied ten-fold since 2005. To diminish the effects of HIV, it is vital to understand the fundamental nature of the virus.

In 1987, Gallo and co-workers reported the discovery of the first human retrovirus, Human T-cell leukaemia virus type 1 (HTLV-1). HTLV-1 was first of the four retroviruses that infect humans. The virus contains single-stranded RNA and causes T-cell leukaemia and T-cell lymphoma in adults. The group had also discovered the first cytokine, interleukin-2 (IL-2). IL-2 is an essential growth factor for T-lymphocytes; the cells proliferate significantly at higher levels of IL-2. This was one of the most important technological developments that helped the isolation of human immunodeficiency virus (HIV).

Gallo suggested that AIDS was possibly caused by a retrovirus similar to HTLV (Gallo *et al*., 1982). Gallo and his collaborators demonstrated that a virus called HTLV-III, believed to be related to the leukaemia viruses of his earlier work, was in fact the causative factor of AIDS (Gallo *et al*., 1984; Popovic *et al*; 1984). Furthermore, a research team based in France run by Montagnier isolated a virus called lymphadenopathy-associated virus (LAV). The sample was taken from a patient at risk of acquiring AIDS (Barré–Sinoussi *et al*., 1983). In 2008, Barré-Sinossi and Montagnier won the Nobel Prize for physiology or medicine for their discovery of the infectious agent. The infectious agent came to be known as HIV. Following the discovery of the HIV as the causative agents of AIDS, many advances were made in a short period. For example, between 1984-1985, all modes of transmission were characterised and the sequencing of the HIV-1 genome was completed (Sanches-Pescodar *et al*., 1985; Ratner *et al*., 1985 and Wain-Hobson *et al*., 1985). The spread of AIDs was rapid; soon after the first reported case in 1981, by the end of June 2017, 36.9 million people worldwide were estimated to be living with HIV (UNAIDS, 2017).

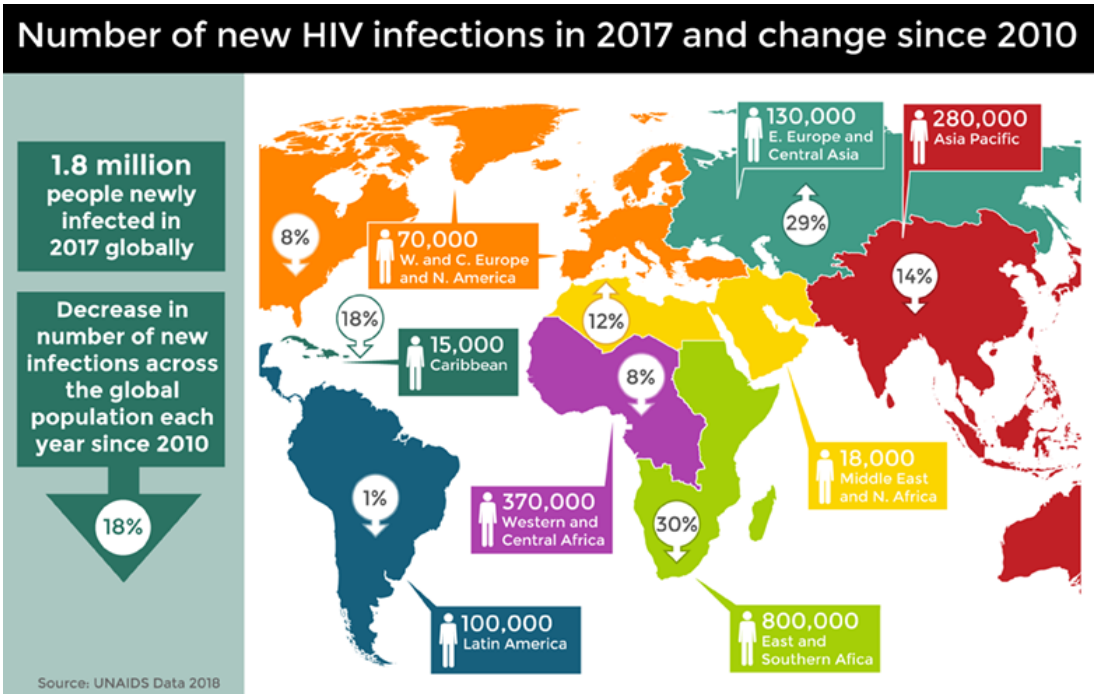


Figure 2-1: The change in global HIV prevalence 2007-2010, highlighting an increase of 29% of newly infected individuals in Eastern Europe and Central Asia. Middle East and North Africa having an increase of 12%. Nonetheless, there was an 18% decrease in number of new infections globally.

There is a geographical variation between and within countries and regions. In Eastern Europe and Central Asia, epidemics that were once characterised primarily by transmission among injecting drug users are now increasingly characterised by significant sexual transmission (UNAIDS, 2018). Additionally, Asian epidemics are becoming increasingly characterised by significant transmission among heterosexual couples. Africa still bears an inordinate share of the global HIV burden (Figure 2-1). The high prevalence in Africa can also be attributed to the fact that this region is believed to be the origin of HIV (Zhu *et al*., 1998). It is estimated that HIV entered the human population as early as 1931 through multiple infections from simian immunodeficiency virus (SIV)-infected non-human primates (Korber *et al*., 2000). Thus the virus went undetected for many years. During these years, the virus could spread unrecognised mainly because the late symptoms of AIDS coincide with symptoms of malnutrition and tuberculosis, which happened to be frequent problems in the infected population.

HIV can be divided into two main branches, known as HIV-1 and HIV-2. HIV-1 evolved from an SIV variant, which is present in chimpanzees, whereas the second type, HIV-2, is the result of a zoonotic infection from SIV found in another primate

species called sooty mangabeys (Heeney *et al*., 2006). Differences between the two variants lie in the organisation of the RNA. Both strains consist of long terminal repeats with *gag, pol* and *env* genes, as these are required for protein structure, whereas *tat, rev, nef, vif, vpr, and vpu* (*vpx* in HIV-2) are genes for infection and replication regulation. The altered genetic make-up of HIV-2 lowers its infectivity and virulence, compared to HIV-1 (Franales-Belasio *et al.,* 2010). HIV-1 is the prevalent and pathogenic strain and, as such, it is the priority for novel treatments to be developed.

Genetic variation is the key survival mechanism for HIV. For each strain and sub-type, the virions possesses slight differences in enzymatic processes and protein structures. This increasingly changes the baseline for drugs, making them impractical. Sub-type group M was the first to be discovered, it is the most common sub-type of HIV, with more than 90% of HIV/AIDS cases deriving from this infection (Thomson *et al*., 2002). Groups N and O total a further 10%. Group M is subdivided further to eight smaller subtypes (A, B, C, D, E, F, G, H, and J), based on clusters typically appearing in phylogenetic analyses of genetic sequences (Robertson *et al*., 2000). Each cluster is defined by a geographical location. HIV-1 sub-type B is most prevalent in the developed regions of the world and is therefore used for development of antiretroviral therapy (Kantor *et al*., 2005). Sub-type C, for example, has been reported with the highest global infections at 47.3% (Osmanov *et al*., 2002). In the modern day world, co-infections of different subtypes in a patient result in further recombination of the HIV genome. Therefore, creating circulating recombinant forms (CRFs), these sub-sub-subtypes are given a number such as group M A 01, 02 and B 01 (Buonaguro *et al*., 2007). The diversity of HIV generates one of the major challenges in HIV drug design (Walker and Burton, 2008). The less pathogenic HIV-2 shows a restricted worldwide distribution. HIV-2 has remained largely restricted to West Africa, with Guinea-Bissau and Senegal having the highest prevalence (de Silva *et al*., 2007).

However, overall prevalence rates are declining and, in most West African countries, HIV-2 is increasingly being replaced by HIV-1 (Van der loeff *et al*., 2006; Hamel *et al*., 2007). Most individuals that are infected with HIV-2 do not progress to AIDS. However, the minority who do pass onto overt disease cannot be distinguished clinically from HIV-1-infected patients (Rowland-Jones and White, 2007).

2.1.2 *Human Immunodeficiency virus type 1*

HIV-1 is a part of the retroviral family; it is a member of the genus of lentiviruses, which have a characteristic long incubation period. Electron microscopy of particles in infected cell cultures show spherical entities with a diameter of 100 -120 µm (Figure 2-2). Retroviruses store their genetic information in ribonucleic acid (RNA) and this requires a mechanism to translate RNA to deoxyribonucleic acid (DNA). Each viral particle contain two single-stranded RNAs that are tightly bound to viral nucleocapsid proteins and two viral enzymes, reverse transcriptase and integrase. These are essential for a successful infection of the host cell. The complex is protected by a cone-shaped capsid comprising approximately 2,000 copies of the capsid protein.

HIV has several major genes, coding for structural proteins. The *gag* (group specific antigen) gene is the precursor of four vial structural proteins; the nucleocapsid proteins (p6 and p7), the viral matrix (p17) and the viral capsid (p24) (Figure 2-3), therefore providing the physical infrastructure of the virus.

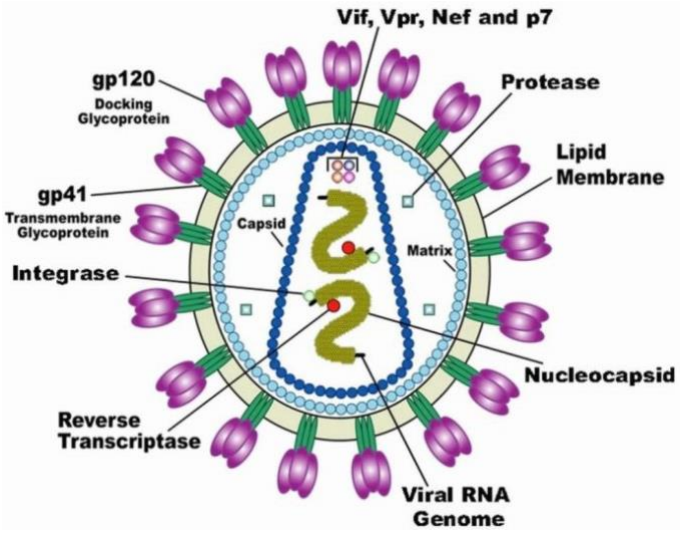


Figure 2-2: Schematic representation of a HIV virion. Indicating the locations of the nucleocapsid proteins. (Derived from US National Institute of Health).

HIV possesses a complex genome of 9.8 Kb containing accessory and regulatory genes (Figure 2-3). The polymerase gene (*Pol*) encodes for three proteins: the protease (p10), the reverse transcriptase (p66/p51) and the integrase (p32). The *Pol* gene provides the mechanism for the retroviral reproduction. The envelope gene (*env*) encodes for glycoproteins gp120 and gp41. The smaller gene encodes for the trans-activators; - *Tat* (Trans-Activator of Transcription), *Rev* (Regulator of virion expression, *vpr* (viral protein r). These activators enhance gene expression, and other regulator proteins – *vif*

(viral infectivity factor), *nef* (negative regulatory factor) and *vpu* (viral protein u) help to counter the defence mechanism of the host cell and keep it efficient in its reproduction. The protease enzyme functions as a tetramer with each monomer cleaved out by the protease from the C-terminal portion of the *gag-pop* polyprotein (Figure 2-3).

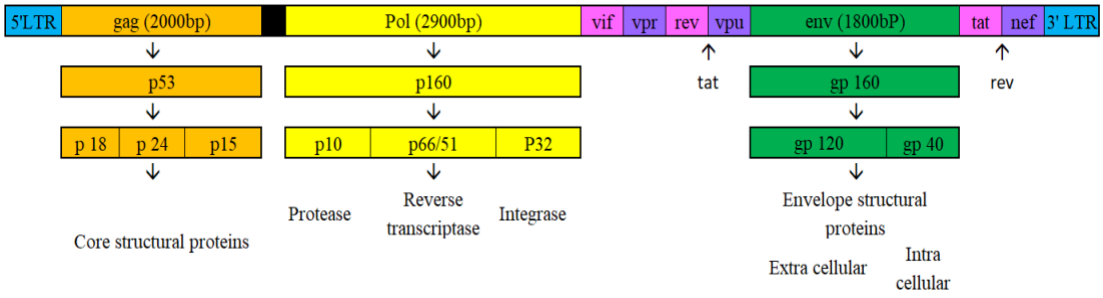


Figure 2-3: Representation of the 9.8 kb HIV-1 genome showing the organisation of genes. (Modified from Costin, 2007).

2.1.3 *HIV-1 replication cycle*

The replication cycle of HIV, illustrated by Figure 2-4, indicates all the principal steps of viral life cycle from fusion to the new matured infectious viral particles. The HIV replication cycle begins when the gp 120 protein on the viral surface binds to the CD4 receptor in the host cell surface by a mechanism involving conformational changes in both the CD4 receptor and the gp 120 protein. The role of the CD4 receptor in HIV cell entry was identified shortly after the isolation of HIV (Dalgleish *et al.*, 1984). The conformational change in the binding receptors leads to exposure of a further antigen molecule, which allows binding to the co-receptor, such as CCR5 (a chemokine receptor). Individuals harbouring a mutant of the CCR5 protein compromise the HIV-1 infectivity. This provides evidence of the co-receptor binding (Berger, Muphy & Farber, 1999). Following this, HIV is able to enter the cell by a process that involves fusion of the viral envelope with the plasma membrane of the host cell. Studies have demonstrated that HIV primarily enters the target cell by endocytosis, followed by fusion in the endosome and not by fusion directly at the plasma membrane (Uchil and Mothes, 2009).

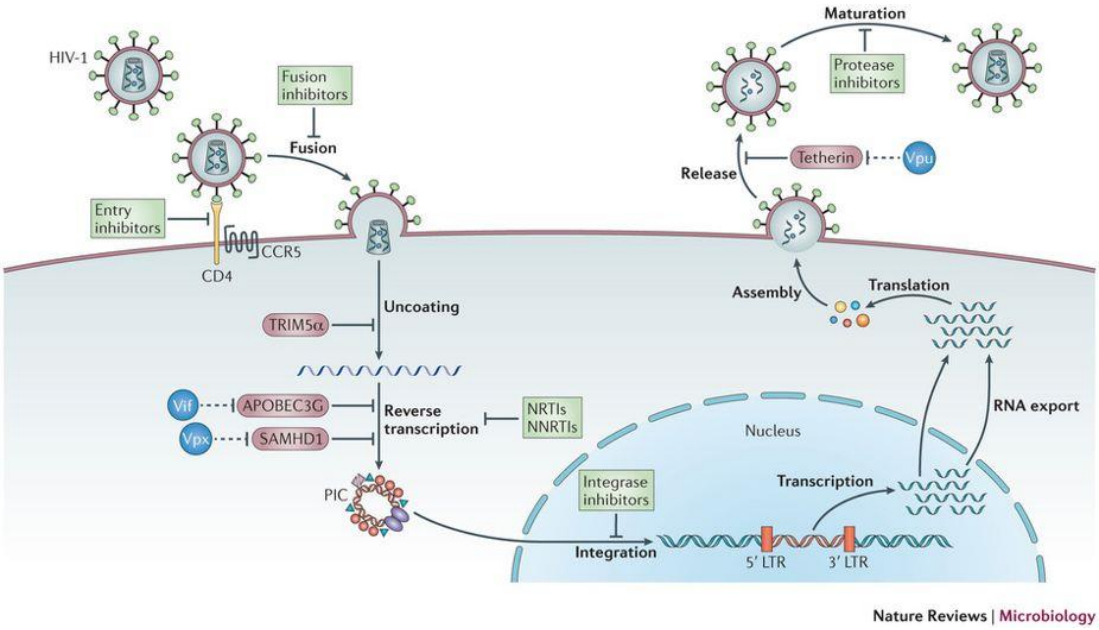


Figure 2-4: Highlights of the main stages with the life cycle of HIV-1 replication; binding to the CD4 receptor and co-receptors; fusion to the host cell membrane with the uncoating of the viral capsid; releasing the viral RNA (vRNA) into the cytoplasm; with the presents of reverse transcriptase vRNA to transformed to viral DNA (vDNA). Formation of the pre-integration complex (PIC) allowing movement into the nucleus. Integrase inserts the vDNA into the host genome; subsequently transcribed and translated to form new vRNA; with assembly at the cell surface membrane to form new immature viral forms. The illustration also provides the major categories of antiviral therapy in green boxes and the step of the life cycle they inhibit. (Diagram depicted from Barré-Sinoussi, Ross and Delfraissy, 2013).

The process of reverse transcription is catalysed by the viral enzyme reverse transcriptase (RT). The process involves the viral single-stranded RNA genome, which is copied into double stranded DNA (dsDNA). Together with viral host proteins, the dsDNA form the pro-integration complex (PIC) which is guided to the nuclear pore. This is an essential step of the process, as it prepares the viral genome for subsequent integration into the host chromosome. The viral enzyme integrase is the part of the PIC which catalyses the integration of viral DNA into the host chromosome. The virus is now referred to as the provirus, as integration is now complete. The next phase of viral replication involves transcription of the integrated DNA provirus into messenger RNA (mRNA), which is the spliced into smaller pieces. The spliced mRNA fragments are translated into regulatory proteins Tat and Rev (Figure 2-3), encouraging production of new virus, and then are exported to the cytoplasm from the nucleus. The poly-proteins env and *gag/gag-pol* are transported to the viral membrane, where formation of new viral particles takes place. Gp160 is transported to the Golgi complex after

going through the endoplasmic reticulum of the host cell. Gp160 is broken up into gp120 and gp41. The virion buds from the host cell containing gag and gag-pol and associated proteins with the plasma membrane along with the viral RNA. Maturation can occur either in the budding phase or in the immature virion after budding. After release, gag and gag-pol are cleaved into their mature forms of viral protease. After maturation, the HIV virion is able to further infect another cell.

2.1.4 *Antiretroviral therapy*

Recent advances in antiretroviral therapy have has led to improvements of quality of life with HIV-infected individuals. The significant advancement in the understanding of HIV replication and its pathogenesis has helped in the identification of pharmacological targets. Zidovudine (AZT) was the first anti-HIV agent to be licensed for clinical use, in 1987. Currently there are more than 30 drugs, belonging to seven different classifications of antiviral agents (Zeier *et al.*, 2011). The seven classes include nucleoside reverse transcriptase inhibitors (NRTIs); non-nucleotide reverse transcriptase inhibitors (NNRITs); protease inhibitors (PIs); integrase inhibitors (IN); chemokine receptor antagonist; fusion and entry inhibitors. These drugs are part of treatment known as highly active antiretroviral therapy (HAART). Each class targets distinct areas within the virus's replication cycle. To ensure that the aim of preventing the virus from developing resistance, a combination therapy is also used. This incorporates non-nucleoside reverse transcriptase with nucleoside analogue reverse-transcriptase inhibitors.

2.1.4.1 *Entry and fusion inhibitors*

The first class of inhibitors intercept the viral replication at the entry of the viral core into the cytosol of the host cell. Amidst all the anti-HIV inhibitors, entry and fusion inhibitors are the only class to activity target the host's proteins rather than a viral protein (Esté and Telenti, 2007). Entry inhibitors are classified according to the different stages they target; CD4 binding, co-receptor binding and fusion (Tilton and Doms, 2010). The FDA has approved maraviroc and enuvirtide for CCR5 and fusion inhibition, respectively. There are other strategies to inhibit other aspects of HIV entry; these currently under development (Tilton and Doms, 2010).

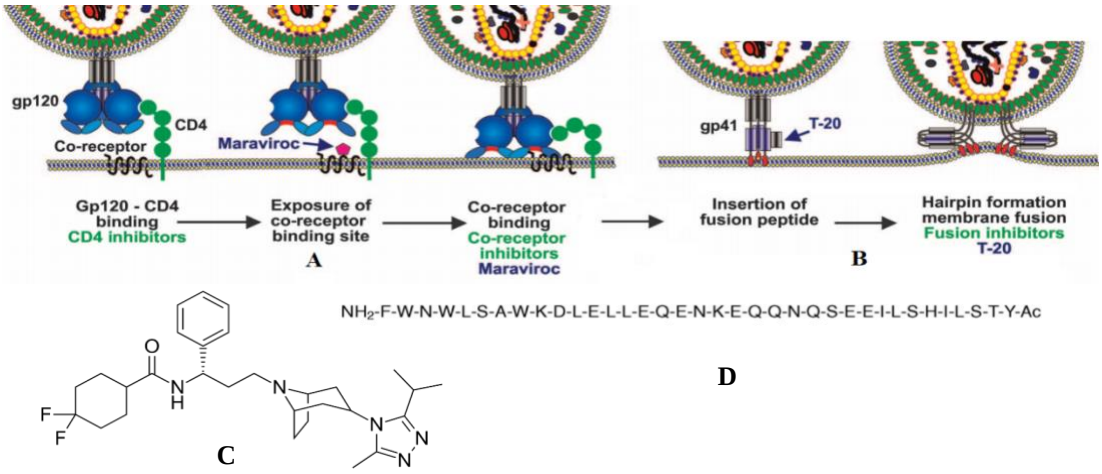


Figure 2-5: Steps in HIV-1 entry and the inhibitor effects of maraviroc and enfuvirtide. (**A**) The co-receptor maraviroc binding to CCR5 and inhibiting gp120 to fuse into the plasma membrane. (**B**) The fusion inhibitor ENF binds to the helical repeat gp41, therefore preventing the hairpin required for membrane fusion. (Picture modified from Forssmann *et al*., 2010). (**C**) Chemical structure of Maraviroc, the p120-CCR5 inhibitor. (**D**) The peptide drug that blocks gp 41 membrane fusion – enfuvirtide.

The co-receptor inhibitor maraviroc (Figure 2-5**C**) binds to the CCR5 and specifically inhibits infection by HIV-1 strains using the CCR5 and not those with CXCR4 co-receptor. Enfuvirtide (T-20 or ENF) prevents entry of HIV by inhibiting fusion. ENF binds to a subunit of gp41 and prevents the essential conformational change that facilitates the fusion of the two membranes (Figure 2-5**A**). Drug resistance mutations are usually located in the ENF-binding site on gp41 (Miller & Hazuda, 2004).

2.1.4.2 *Reverse transcriptase inhibitors* (*RTIs*)

The reverse transcriptase (RT) inhibitors interfere with the generation of a DNA copy of the viral genome. RT functions as a heterodimer to catalyse the conversion of the single-stranded genomic RNA into double-stranded DNA with duplicated long terminal repeats, which is integrated into cellular DNA by the viral integrase. RT is a heterodimer, comprising of p66 and p51 units (Kohstaedt *et al*., 1992). There are two classes of RT inhibitors, distinguished by their mode of action (Figure 2-6). Firstly there is the group called nucleotide RT inhibitors. They serve as chain-terminators of viral RT, consequently acting early in the viral replication cycle by inhibiting an essential step of proviral DNA synthesis (Cihalr and Ray, 2010).

The second group of RT inhibitors are non-nucleotide reverse transcriptase inhibitors (NNRTIs). These are nucleoside analogues; they carry out the inhibition of RT by binding to a hydrophobic pocket in the active site of the enzyme. This step impairs the flexibility of RT, hence preventing DNA synthesis. Mutations can cause resistance to NNRIs, which results in reduced affinity of the inhibitor and target. Usually, a single mutation that is selective by NNRTIs is sufficient to confer complete resistance to all compounds of the drug class (Clavel & Hance, 2004).

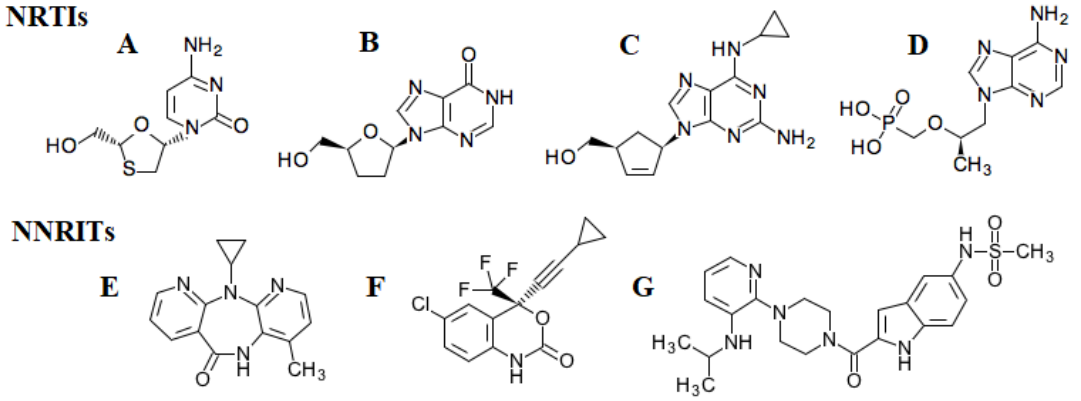


Figure 2-6: Selection of chemical structures of NRTIs and NNRTIs. NRTIs: (**A**) Lamivudine (**B**) Stavudine (**C**) Abacavir (**D**) Didanosine. NNRITs**:** (**E**) Nevirapine (**F**) Efavirenz (**G**) Delavirdine**.**

2.1.4.3 *Protease inhibitor (PIs)*

PIs interfere with the process of forming new mature infectious viral particles. Protease targets the gene amino-acid sequence in the *gag* (p55) and *gag-pol* (p160) polyproteins, which must be cleaved before budding viral particles can mature. This process produces three large proteins, p7, p17 and p24; they contribute to the structure of the virion, RNA packaging and transcriptase, integrase and protease (Huff, 1991). PIs are small molecules that bind to the active site at the centre of the homodimer and compete with its natural substrate. PIs prevent cleavage of gag-pol protein precursors in acutely and chronically infected cells (Flexner, 1998). Mutations in response to PIs have altered the shape of the active site and changed the sequence of the cleavage sites in the gag polyprotein, these developments together to reduce drug resistance while preserving the virus fitness (Ali *et al*., 2010). With the use of structural and biochemical data, ten FDA-approved inhibitors have been designed to mimic the substrate transition state. The hydroxy group of the inhibitor interacts with the

carboxyl group of the protease active site residues, Asp 25 and Asp 25', by hydrogen bonding (Lv, Chu and Wang, 2015) (Figure 2-7).

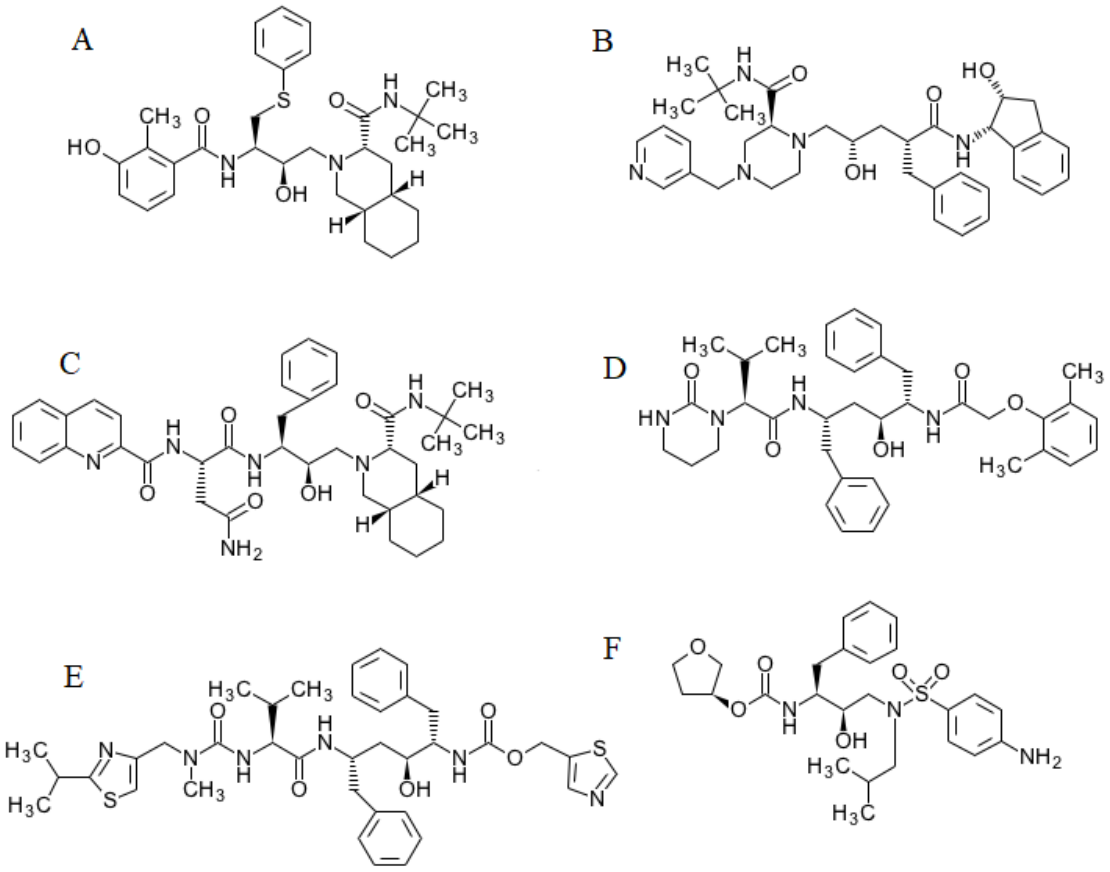


Figure 2-7: Six of the chemical structures of HIV protease inhibitors. (**A**) Nelfnavir (**B**) Indinavir (**C**) Sapuninavir (**D**) Lopinavir (**E**) Ritonavir (**F**) Amprenavir

2.1.5 *Integrase inhibitors (INIs)*

There are several ways to target integrase, inhibition of the LEDGF/p75 (Lens epithelium-derived growth factor/p75) – IN interaction and IN binding inhibitors (INBIs) but the most pursued is 3'-processing inhibitors (3'-INIs) and strand transfer inhibitors (INSTIs). Integrase inhibitors aim to prevent the enzyme activities of integrating the vDNA into the host chromosome. There are three domains (Figure 2-11); the N-terminal domain (NTD) contains a HH-CC zinc-finger motif that is partially responsible for multimerisation, optimal activity and protein stability. Within the integrase lies an active site residues D64, D116 and E152 (DDE motif) in the catalytic core domain (CCD) forms the catalytic triad. The C-terminal domain (CTD) binds non-specifically to DNA with high affinity.

2.1.5.1 *Inhibition of the LEDGF/p75 – integrase interaction.*

LEDGF/p75 is a cellular co-factor of HIV-1 integrase that promotes viral integration by securing the PIC to the chromatin. A discovery of a novel class of inhibitors, 2-(quinolin-3-yl) acetic acid derivatives. They work by blocking HIV replication at low micromolar concentrations through binding in the LEDGF/p75 binding pocket of HIV integrase, thereby blocking the interaction with LEDGF/p75 and interfering indirectly with the catalytic activity (Christ *et al*., 2012). These drugs are likely to be highly target-specific and less prone to the development of resistance due to the allosteric behaviours, displaying a dual mode of action; inhibition of IN-viral DNA assembly and inhibition of IN-LEDGF interaction (Tsiang *et al*., 2012).

2.1.5.2 *Integrase binding inhibitors*

A further class of INIs could be IN-binding inhibitors such as V-165 (Figure 2-8). V-165 inhibited HIV replication in MT-4 cells at an IC_{50} of 8.9 µM, this concentration is 14-fold below the cytotoxic concentration. Furthermore, V-165 inhibited both reverse transcriptase and integrase activities in enzymatic assays at micromolar concentrations. With further HAART success, in combination with zidovudine or nelfinavir inhibition increased (Pannecouque *et al*., 2002).

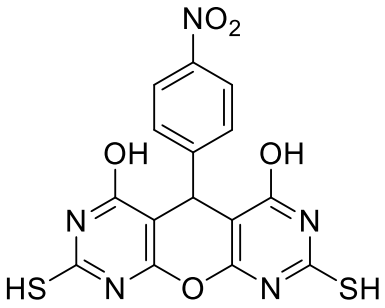


Figure 2-8: Structure of V-165

2.1.5.3 *Strand transfer (ST) and 3'-processing (3'-P) inhibitors*

The integration of the vDNA requires a three-step process (Figure 2-9). Firstly, 3'-processing step, occurring in the cytoplasm, IN constructs itself on the developed vDNA, cleaving bases, G-T from both 3' end of the DNA. Secondly, the PIC transports the vDNA into the nucleus, where IN acts as a catalyst to fuse the pre-processed 3'-end to the corresponding strand of the host DNA, creating an offset of four to six

nucleotides (Yoder and Bushman, 2000). Finally, the process of integration is complete with gap repair mechanisms from DNA repair enzymes (Miller *et al*., 1995).

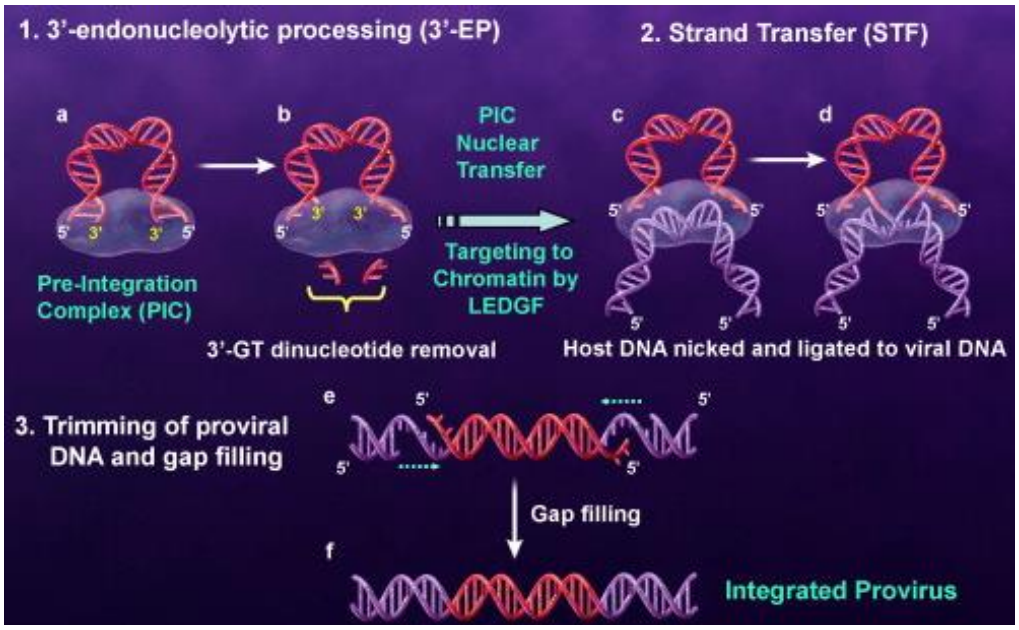


Figure 2-9: The integration of vDNA into the host DNA is essential to the HIV lifecycle. The process requires a twostep procedure facilitated by the integrase enzyme, (1) 3'-processing and (2) strand-transfer reaction. This reaction creates a reactive 3'-hydroxy at the end of the viral DNA. The PIC is transported across the nuclear membrane; with the help of LEDGF, it is targeted to the chromatinised host genomic DNA. The second reaction is the strand transfer reaction. Both 3'-hyroxyl ends of the vDNA are bonded to the host chromosomal DNA. The ST reaction results in a four to six nucleotide offset at the 5' ends. Cellular repair enzymes fill that gap, resulting in a mature integrate provirus from which the lifecycle can continue. Adapted from Pommier *et al.*, 2005 and depicted from McColl and Chen, 2010.

The main integrase inhibitors that are currently on the market are raltegravir, elvitegravir and dolutegravir (Figure 2-10). These compounds are split into two main subcategories. Raltegravir and elvitegravir share the mechanism of competing for the Mg^{2+}-binding sites (Bacchi *et al*., 2011). Elvitegravir has been determined to be more potent than raltegravir (Marinello *et al*., 2008) but has also been shown to demonstrate cytotoxic effects in non-infected cells (Sato *et al*., 2006). Raltegravir has the potential to inhibit both HIV-1 and HIV-2 replication and is more effective than other INSTIs with less adverse side-effects. Compared to dolutegravir, which focuses on interfering with catalytic integrase functions such as the PIC (McCormack, 2014).

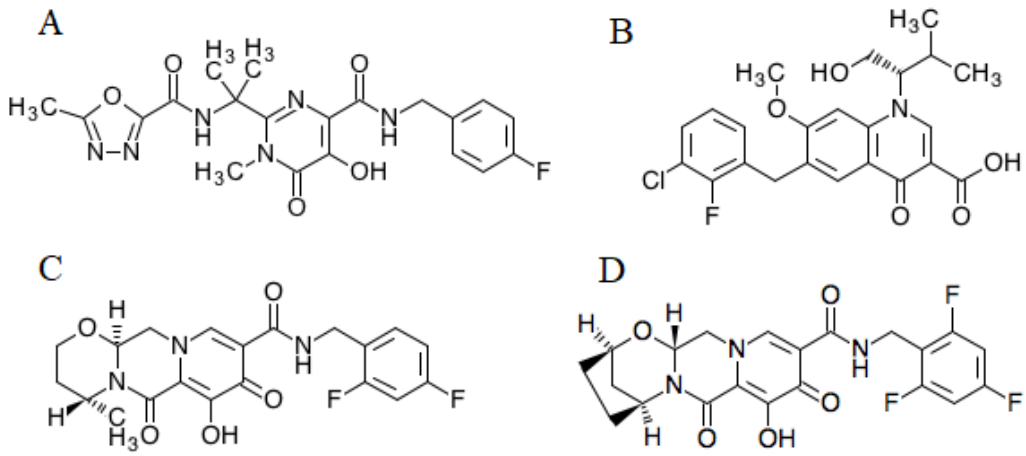


Figure 2-10: Structures of clinically active integrase strand transfer inhibitors. (A) Raltegravir, (B) elvitegravir (C) dolutegravir and (D) bictegravir.

However, resistance has already been shown during the therapy of raltegravir (Liao *et al*., 2010). Second-generation integrase inhibitors, such as MK-2084, were developed to over-come the resistance in raltegravir and are now currently under phase III clinical trials. It is also reported to show potency to inhibit HIV-1 enzymatic functions approximately four times longer than raltegravir (Hartman and Buckheit, 2012). Additionally, bictegravir (Figure 1-10: (D)) is a novel integrase inhibitor. The FDA has approved it in 2018 for combination therapy. Bictegravir shows synergistic effects with other inhibitors and cytotoxicity is absent in most non-infected cells, with higher antiviral effects than any other INSTIs (Tsiang *et al*., 2016).

2.1.6 *Highly active antiretroviral therapy (HAART)*

Combination therapy was developed, which involved combining several antiretroviral compounds into one therapy. This approach has the greatest benefit when combining drugs from NNRTIs and PI classes. Combination therapy can block the resistance effect more effectively for two reasons; (1) multiple drugs can inhibit replication more effectively than a single agent can, (2) multiple mechanisms are required for resistance to occur to all drugs. This marked the beginning of HAART. HAART combines a minimum of three drugs from at least two different drug classes targeting different and distinct proteins (Dybul *et al*., 2002; Clavel and Hance, 2004). This creates multiple barriers to HIV replication and subsequently lowers and reduces the possibility of a superior mutation. With increasing concerns regarding antiretroviral resistance, the death rate among HIV-positive individuals continues to decline (Crum *et al*., 2006).

The success of HAART in preventing the onset of AIDS is clear but HAART does carry limitations. Drug toxicity, non-adherence to therapy and viral resistance restrict treatment.

As this therapy relies on each class of compound on having a unique adverse effect, selecting a procedure with certain characteristics may be a challenge. For example, protease inhibitors produce adverse metabolic effects that may increase the risk of developing cardiovascular disease (Iloeje *et al*., 2005). Moreover, studies have shown the incidence of cancer is rising in HIV-1 infected patients, non-nucleotide reverse transcriptase inhibitors has been seen to cause radio-sensitivity in patients resulting in radiation-induced side-effects (Ulrike *et al*., 2015). Since each HAART compound has specific limitations, tailoring a therapy to an individual patient is of the utmost importance (Sension *et al*., 2007), ensuring effective inhibition is met with reduced side-effects.

2.1.7 *Integrase inhibitor resistance*

No cure exists for HIV and patients must remain on therapy for the entirety of their lives, therefore drug resistance is a major issue. The principles of integrase inhibitors is parallel to those of all inhibitor resistant patterns. Drug resistance has been documented for every currently available drug class (Wainberg, Zaharatos and Brenner, 2011). Virological failure is the failure to suppress and sustain a patient's load to less than 200 copies/mL (Parienti *et al*., 2004). The first generation of strand-transfer inhibitors, raltegravir and elvitegravir, are associated with development of resistance through three mutually exclusive pathways. Firstly INI resistance is caused by primary mutations that decrease the sensitivity of IN to the inhibitors (Cooper *et al*., 2008). In combination with secondary mutations that further reduce virus sensitivity, these mutations could also decrease the fitness of the virus (Loizidou *et al*., 2009). The second pathway is a genetic barrier to inhibitor resistance, defined by the number of mutations required for the loss of clinical inhibitor activity. The third pathway is extensive but incomplete cross-resistance among the inhibitors (Blanco *et al*., 2011).

2.1.8 *Structure and functions of HIV-1 integrase*

The HIV-1 integrase is a 288-amino-acid protein (32 KDa) that belongs to a family of polynucleotide transferases (Dyda *et al*., 1994; Bujacz *et al*., 1995). It encodes the C-terminal part of the *pol* gene of the HIV-1 genome. It has three structurally independent domains (Figure 2-11) (i) the N-terminal domain (NTD) (IN_{1-49}) with an non-conventional HHCC zinc finger motif, promoting protein multimerisation; (ii) the central core domain (CCD) ($IN_{-50-212}$) containing a canonical DDE motif and involved in DNA substrate recognition; (iii) the C-terminal domain (CTD) ($IN_{213-288}$), which binds DNA non-specially and helps to stabilise the IN-DNA complex (Asante-Appiah and Skalka, 1999). The structures of the isolated HIV-1 domains were characterised by X-ray crystallography and NMR analysis (Cai *et* al., 1997; Berman *et al*., 2000; Eijkeleboom *et al*., 1995). The NTD polypeptide has a structure consisting of four helices stabilised by a Zn^{2+} cation, in tetrahedral coordination with the HHCC motif formed by His12, His 16, Cys 40 and Cys 43 with the last 43-49 amino-acids disordered (Figure 1-11).

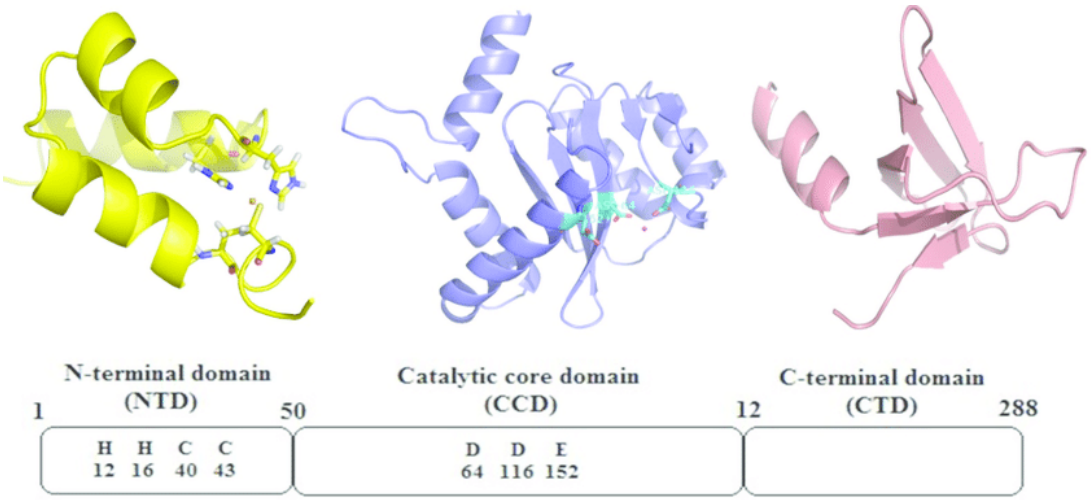
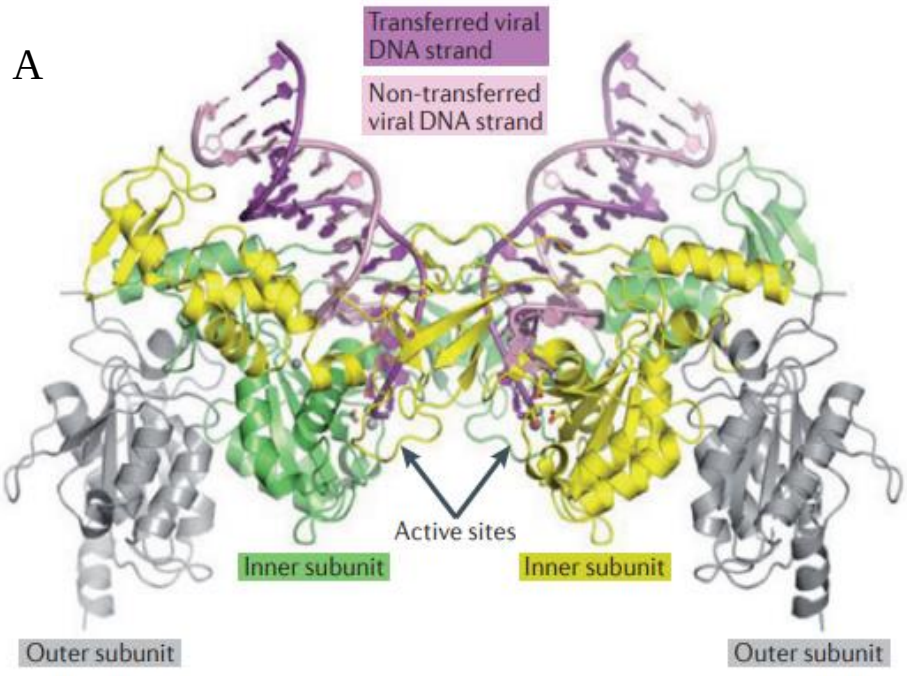


Figure 2-11: Structural and functional domains of HIV-1 integrase. IN is a three-domain multimeric protein that contains conserved amino-acid motifs in both the NTD and CCD with HHCC and DDE, respectively. NTD required for Zn^{2+} binding multimerisation, CCD Mg^{2+}/Mn^{2+} chelation catalysis and DNA binding and the CTD has of five antiparallel β-strands forming a β-barrel and adopting an SH3-fold. The CTD is required for non-specific DNA binding. All three domains are required for integration activity. (Depicted from Hajimahdi and Zarghi, 2016)

The core domain has a mixed α / β structure, with five β-sheets and six α-helices (Dyda *et al*., 1994). The active site residues D64, D116 and E125 are located in different structural elements: β-sheet, coil and helix, respectively. The CCD comprises a flexible residue loop from 140-149 that encompasses the catalytic site. Conformational

changes in this loop are required for 3'-P and ST reactions (Mouscadet *et al.*, 2010). The IN activities require the presence of the metallic co-factor(s), which bind to the catalytic residues of the DDE motif. The structures of Avian Sarcoma Virus (ASV) IN (Bujacz *et al.*, 1997) and the Tn5 transposase (Lovell *et al.*, 2002) have provided evidence of a two-metal active-site structure, which has been used to build metal-containing IN models (Karki *et al.*, 2004; Wang *et al.*, 2005).

The CCD and CTD were connected by a helix formed by residues 195-221. The local structure of each domain was similar to the structure of the isolated CCD, whereas the dimer C-terminal interface differed from that obtained by NMR. The complete structure of the HIV-1 homologue-primate foamy virus (PFV) integrase complexed with the substrate DNA has also been recently reported (Hare *et al.*, 2010). The available structural experimental data, together with the biochemical evidences, are useful for generation of the HIV-1 IN models (Engelman and Cherepanov, 2012) (Figure 2-12).



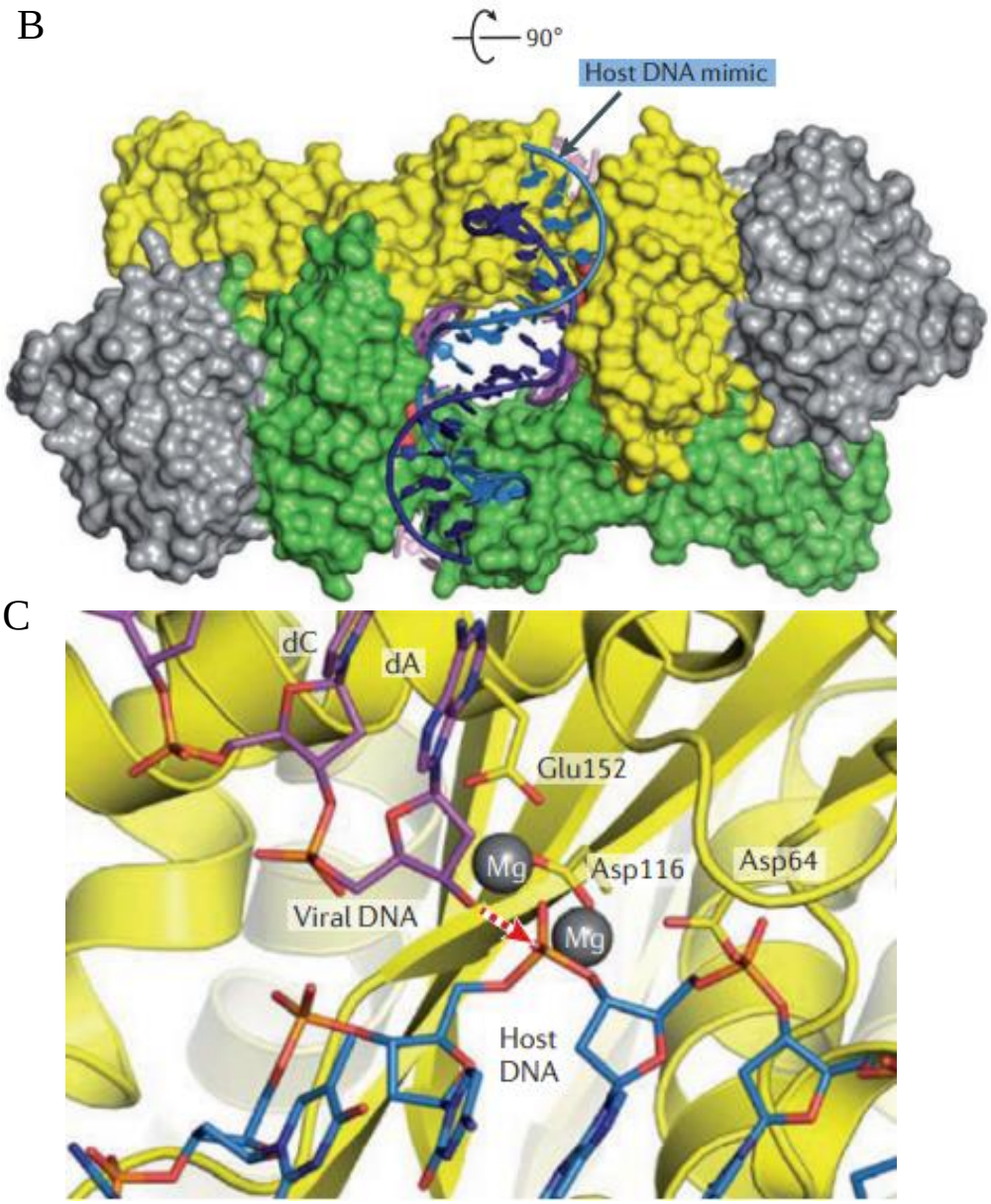


Figure 2-12: A) Illustration of the prototype foamy virus integrase structure (protein data bank (PDB) accession 3OY9). Viral IN forms a dimer of dimers structure in which the active subunits yellow and green, along with the chemically inactive grey regions. The transferred vDNA strand harbours the terminal dinucleotide and becomes joined to the chromosomal DNA by the action of the IN strand-transfer activity. Active-site triad is shown as red sticks and Mg^{2+} ions as grey spheres. **B)** The IN subunits are shown in fill-space mode. IN in a complex with chromosomal host DNA. **C)** Close-up of the DNA strand-transfer reaction, the Asp and Glu active-site residues of integrase are shown as yellow sticks. DNA is shown as magenta and blue sticks. Nucleophilic attack is shown in the direction of the dashed red arrow. The Figure has been depicted from Engelman and Cherepanov, 2012.

2.1.9 *Structure activity relationship of HIV-1 integrase inhibitors*

The HIV-1 IN is a key enzyme in the replication mechanism of a retrovirus, catalysing the covalent insertion of the reverse-transcribed DNA into the chromosomal DNA of the infected cells (Brown, 1990). Two structural components are necessary for integrase binding (i) a chelating traid, which coordinates the integrase activity and strictly requires the presence of metallic cationic cofactors (Sante-Appiah & Shalka, 1999). Within the metal binding region, the triad binds with two Mg^{2+}, therefore anchoring the inhibitor to the protein surface. (ii) The second structural unit of the pharmacophore model requires a hydrophobic benzyl moiety. This component will fit into the hydrophobic pocket near the active site (Wang *et al*., 2010). Furthermore, the model illustrates a flexible linker, thus allowing the closest fit into the hydrophobic pocket. The model also consists of a planar heteroatom-chelating motif (Figure 2-13).

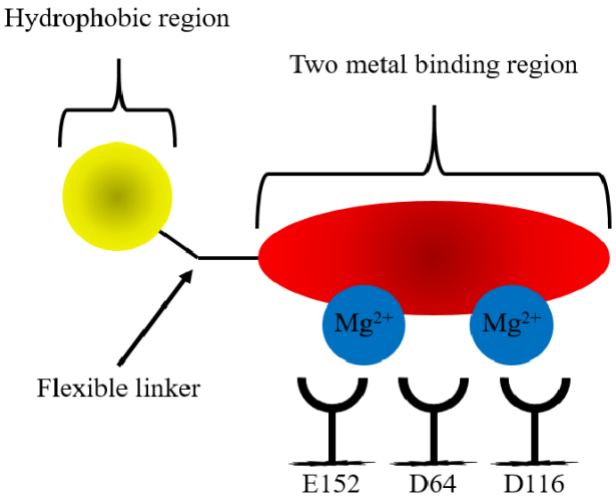


Figure 2-13: Two Mg^{2+} ions with the catalytic triad (E152, D64 and D116) binding pharmacophore of IN inhibitors.

Potent integrase inhibitors have a substituted benzyl component; this is critical for preserving 3'-end-inhibting potency. Research conducted by Chen *et al*., (2008) reported the removal of the benzyl group prevents inhibitory function. Moreover, lipophilic substituents are beneficial for the strand transfer inhibition, for example thiophen-2-yl, phenyl, and furanyl substitutions. Finally, heteroaromatic amines and amides also increase the 3'-processing inhibition function.

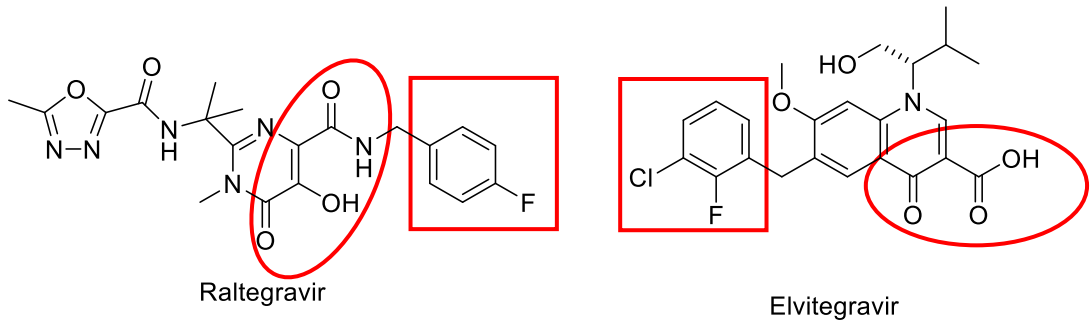


Figure 2-14: Structure activity relationship of elvitegravir and raltegravir. A benzyl group fits into the hydrophobic pocket (red square). A hydrophilic region (red oval) that interacts with the metallic cationic cofactors.

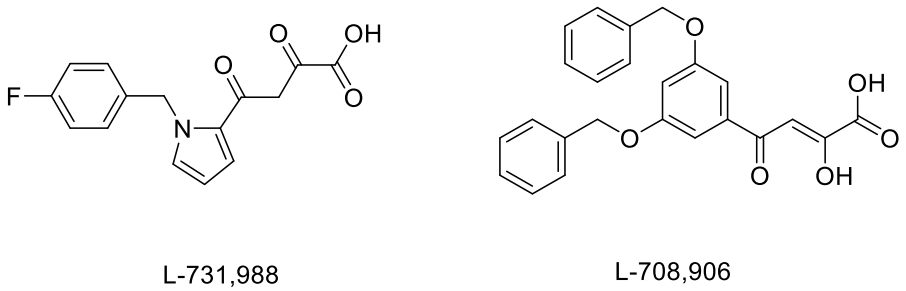
Catechol-based inhibitors that maintain a planar relationship with the bis-hydroxylated aryl ring increase their potency. Further optimisation could be included by a *meta*-chloro substituent; this enhances the interaction of the benzyl group with the adjacent hydrophobic pocket (Pendri *et al*., 2011) (Figure 2-14). A benzyl substituted hydroxy group improves metal-chelating capability while a methoxy group is much less potent due to the steric clash by the addition methyl group with the catalytic metals (Chen *et al*., 2008).

2.1.10 *Mechanisms of current integrase inhibitors*

Pharmacological development of integrase inhibitors has demonstrated that there are two principle strategies, which have been considered for the development of integrase inhibitors. (i) Non-catalytic inhibitors targeting IN/LEDGF interactions, and (ii) catalytic inhibitors targeting 3'-P or the strand-transfer reactions (Thierry *et al*., 2016). Since the early 1990s, a number of compounds either inhibiting just one or both of these reactions *in vivo* have been identified (Semenova, Marchad and Pommier, 2008).

However, the PIC with association of integrase with vDNA reconstituted *in vitro* or whether isolated from infected cells, is highly stable, therefore keeping the complex together after the 3'-processing reaction, thus ready for the subsequent ST reaction to occur (Smolov *et al*., 2006). It can be said that the complex has a slow catalytic activity and does not dissociate after 3'-processing. This weak catalytic activity is not detrimental in the host cells, due to a single integration event sufficient for overall function. Thus making is challenging to create competitive inhibitors for free integrase. Therefore, Hazuda *et al*. (2000) suggested that the PIC would be a suitable target for inhibitors. With the development of new assays, designed for screening

ligands of the integrase vDNA complex ultimately headed the structural elucidation and characterisation of the first strand-transfer inhibitors, L-731,988 and L-708,906 (Figure 2-15) (Hazuda *et al*., 2000).



L-731,988 L-708,906

Figure 2-15: The first strand-transfer inhibitors L-731,988 and L-708,906

L-731,988 and L-708,906 were the most active compounds identified (Hazuda *et al*., 2000). The diketo-acid functionality is an intrinsic feature of these compounds and they compete with the catalytic triad of the CCD, which opens following a change in the conformation, which is induced by the binding and inhibition of the 3'-processing of the vDNA (Hazuda *et al*., 2000). Furthermore, these compounds have a strong ability to inhibit the strand-transfer at nanomolar concentrations (Espeseth *et al*., 2000). This discovery of the compound class is a breakthrough, not only due to their potent antiviral activity but also because it was the first well-characterised selective targeting of integrase in HIV-infected cells. Lastly, Diketo acids (DKAs) exclusively inhibit integrase, as they show no inhibition of the other steps in the viral lifecycle.

The replacement of the carboxylate group by its tetrazole bioisotere led to the development of 5-CITEP and its analogue, S-1360 (Figure 2-16). This led to an increase in antiviral activity and to overcome pharmacological limits, such as serum protein binding (Zhuang *et al*., 2003). The first co-crystallised structure of CCD was with 5-CITEP, where it was found to be in close proximity to the DDE motif, confirming previous speculation that the CCD is actually the catalytic active site of the enzyme, making it possible to construct a model of the inhibitor indicating minimal interactions between the ligand and the resolved Mg^{2+} (Goldgur *et al*., 1999). S-1360 advanced into phase-II clinical trials but was withdrawn due to poor pharmacokinetic data. A Shionogi-GSK joint venture put RSC-1838 into pre-clinical development. This has now led to the novel integrase inhibitor GSK364735.

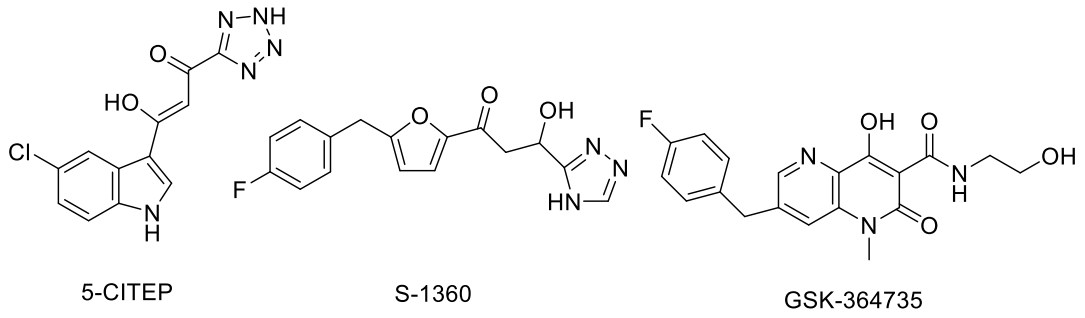


Figure 2-16: Structures of 5-CITEP, S-1360 and GSK-364735 integrase inhibitors.

Unfortunately, a problem with many of the DKAs developed is their toxicity. Further efforts on the DKAs motif have led to analogues (Figure 2-17), namely L-870,810 and L-870,812, with integrase inhibition in the nanomolar range. L-870,810 was very promising but, during phase II clinical trials, it was withdrawn due to their characteristic liver and kidney toxicity. L-870,812 was not developed further (Hazuda *et al*., 2004).

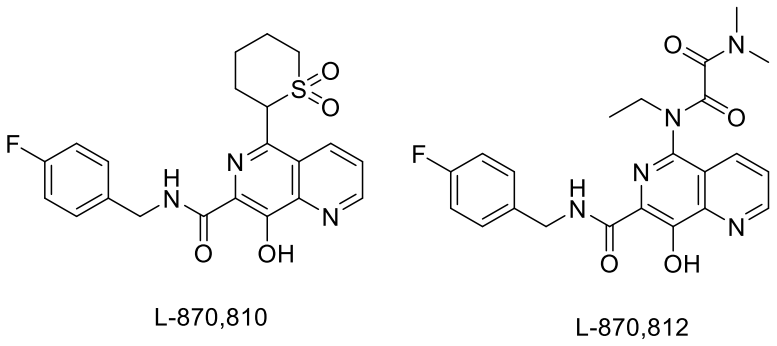


Figure 2-17: Chemical structures of L-870,810 and L-870,812. With the characteristic DKA motif, due to toxicity level both were withdrawn from clinical trials.

2.1.11 *Hydroxycinnamic acids*

The most widely-distributed phenolic acid in plants are termed, hydroxycinnamic acids (HCAs). Structurally they have a C_6-C_3 skeleton, demonstrated by *p*-coumaric acid, caffeic acid and ferulic acid. Phenolic compounds isolated from plants sources are secondary plant metabolites and are abundant in food and beverages (Herrmann, 1989; Adom and Liu, 2002). These structures play a vital role in survival, as they contribute to the defence against pathogen attacks and ultraviolet radiation (Manach *et al*., 2004; Treutter, 2010) and play important roles in growth and development (Rocha *et al*., 2012; El-Seedi *et al*., 2012). Hydroxycinnamic acids are abundant in large quantities in fruits and vegetables, where they can occur freely or as components of

plant polymers in the cell walls, cross-linking the sugars which provides strength and support (El-Seedi, 2007; Marques, 2009; Cartea *et al*., 2011).

2.1.12 *Alternative sources for HIV-1 inhibitors*

For millennia, plants have been essential for food production, cosmetics and building materials. Plants have played a dramatic role in medicine, due to the use of their natural products, these being defined as small-molecule secondary metabolites from living organisms. Traditional medicine uses more than 350,000 different plant species. In 2015, Professor Tu Youyou received the Nobel Prize in Physiology or Medicine for her discoveries concerning a novel therapy against malaria. Her knowledge of ancient texts of traditional Chinese herbs enabled her to isolate and characterise artemisinin (Nobel Foundation., 2015).

The lack of a pharmacological cure for HIV and the low cost and easy availability of herbal therapies have encouraged many patients to seek alternative therapy. Several reported studies of biochemical extracts have identified multiple natural polyphenols that inhibit HIV-1 integrase (Mazumder *et al*., 1997; Ahn *et al*., 2002; Cotelle, 2005; Akram *et al*., 2017). Hydroxylated aromatic derivatives represents one of the largest class of IN inhibitors (Maurin *et al*., 2003; Gordon *et al.,* 2007; Yang *et al*., 2010; Neamati *et al.*,2012)

Research has demonstrated that the most potent compounds for integrase inhibitors share two common characteristics, multiple aromatic rings and secondly a catechol moiety (Righi *et al*., 2017). Some natural compounds of this class belong to this class (Figure 2-18). Structure-activity analysis on this class demonstrated that the IN inhibitor activity is dependent on the catechol moiety (Bruke *et al*., 1995; Zhao *et al*., 1997). There are no definitive conclusions regarding their mechanism of actions. It is not clear if the hydroxyl groups chelate the Mg^{2+} ions or if they interact with the amino acids by hydrogen-bonding within the CCD. To date, none of the compounds are suitable for clinical development due to problems of cytotoxicity (Hajimahdi *et al*., 2016). Cytotoxic activity is believed to result from non-selective binding and, perhaps more importantly, from the formation of oxidised species, such as semiquinones or *ortho*-quinones, which form adducts with proteins or possibly DNA (Bolton and Dunlap, 2016).

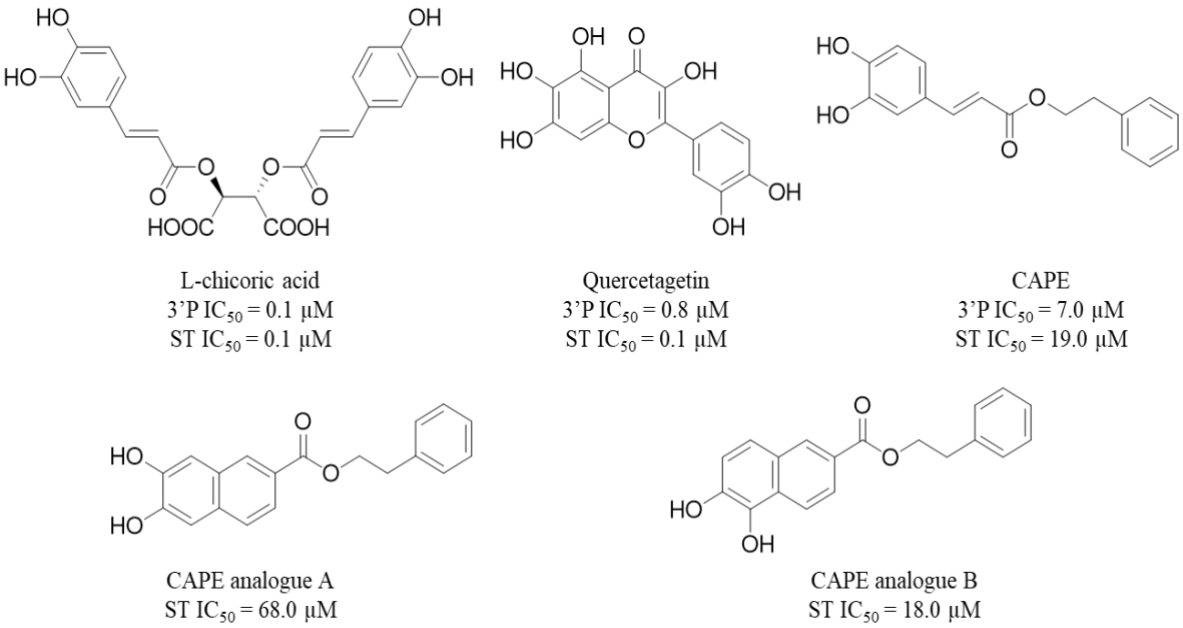


Figure 2-18: Structure and activity of polyhydroxylated aromatic derivatives. The structures of CAPE A and CAPE B were designed to mimic the two isomers of CAPE. As indicated the subtle conformational change with the catechol moiety has a significant effect on the inhibitor binding.

Further research has attempted to separate the mechanisms of cytotoxicity and IN inhibition. A number of mono-hydroxylated derivatives as well as methoxy and acetyl-protected catechol analogues were tested. However, none exhibited a significant inhibitory response. Furthermore, as illustrated in a paper by Bruke *et al*. (1995), two structural isomers of CAPE, with subtle conformational changes were found to have significantly different inhibitor results (Figure 2-18). This highlights the polyhydroxylated aromatics to have the strict requirement of a catechol moiety (Burke *et al*., 1995).

Additionally, nations such as China are rich in herbal resources and have a long history of use in traditional Chinese medicine (TCM), providing therapy for a variety of diseases. For example, *Emblica officinalis* has been reported for antioxidant, immunomodulatory and anticancer activities (Madhuri *et al*., 2011). *Momordica charantia* extracts were analysed for anti-HIV activity, this identified a protein called MAP 30 that inhibits HIV-1 integrase (Lee-Huang *et al*., 1995; Kong *et al*., 2003). *Lithospermum erythrorhizon and L. ruderale* are used for their antimicrobial and antitumor activity (Brigham *et al*., 1999; Fu *et al*., 2002). Root extracts of *L. ruderale* and *L. eryhrorhizon* produce a number of polyphenolic compounds, such as caffeic isomers including CAPE, rosmarinic acid and lithospermic acid A and lithospermic

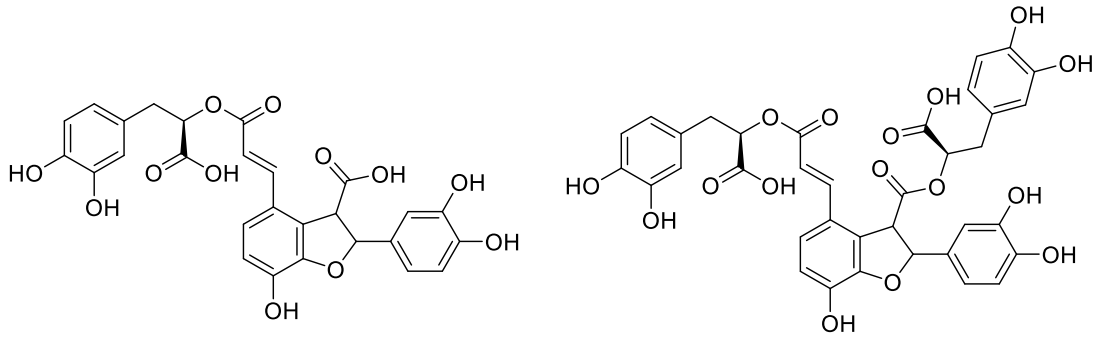
acid B (Figure 2-19) (Kelley *et al*., 1975; Findley and Jacobs, 1980; Yamamoto *et al*., 2000; Thuong *et al*., 2009).

2.1.13 *Lithospermic acid*

Lithospermic acid is an organic by-product; it was first discovered in the root extracts of *Lithospermum ruderale* by Johnson *et al.* (1963). The chemical constituents were used in tea as a contraceptive, reducing gonadotropic release (Train, Henrichs and Archer, 1941). Using a sub-species called *Lithospermum officinale*, lithospermic acid was isolated and structurally elucidated (Kelley *et al*., 1975). Related scientific reports were minimal until 2002. It has been reported that lithospermic acid has been isolated from *Salvia miltiorrhiza* (red sage). The dried roots of *S. miltiorrhiza* is one of the most popular herbs in TCM. There are publications supporting the effects of lithospermic acid as a potent antioxidant for ischemia, cardiovascular disease; a therapeutic aid for cerebrovascular and Parkinson's diseases, hepatoprotective and anti-HIV activity (Liu *et al*., 2007; Wang, Morris-Natshe and Lee, 2007; Chen *et al.,* 2009; Lin *et al*., 2015; Chan and Ho, 2015).

The overall discovery of lithospermic acid and its vast medicinal properties has demonstrated possible involvement of this compound to treat a variety of infections and diseases. An influential finding by Abd-Elazem *et al*. (2002) reported that lithospermic acids display potent inhibition of HIV-1 IN with no cytotoxicity, even at high concentrations – CC_{100} value > 297 µM in H9 cells (Abd-Elazem *et al*., 2002). The IC_{50} values for 3'-processing and the strand transfer reaction were found to be 0.83 µM and 0.48 µM respectively. Lithospermic acid is a promising lead on HIV-1 IN activity, furthermore, there is a weak to no activity on HIV-1 entry, reverse transcriptase and protease (Peng *et al*., 2008 Abd-Elazem *et al*., 2008). There are two known variations, lithospermic acid and lithospermic acid B (Figure 18). Alternatively, lithospermic acid B has the potential to be a promising integrase inhibitor, with weak to no activity on HIV-1 reverse transcriptase and protease (Peng *et al*., 2008). Chu and Liu (2011) also highlighted the potency of the two variants, while lithospermic acid B has a different mode of action to block p24 proteins in HIV-1 IN. The total synthesis of lithospermic acid was first completed by an asymmetric

intramolecular alkylation *via* rhodium-catalysed C-H activation in a ten-step reaction (O'Malley *et al*., 2005). The reaction concluded with a 5.9% overall yield.



Lithospermic acid A

Lithospermic acid B

Figure 2-19: Chemical structures of HIV-1 IN inhibitor lithospermic acid (A) and lithospermic acid (B). The carboxylic acid and esters present indicate the potential pharmacophore with the binding motif found in the CCD of integrase.

When comparing the structures of the two lithospermic acids, It can be noted they are constructed of caffeic acid monomers (Kelley *et al*., 1976). Lithospermic acid B contained four caffeic acid monomer and can be as a tetramer. Both variants share the classic structural motif of the other well-known IN inhibitors, the bis-catechol (Figure 2-18).

Caffeic acid is responsible for producing lignin in most plants (Boerjan *et al*., 2003) and has shown to possess many antioxidative, antihypertension, anti-tumour and anti-viral properties (Boerjan *et al*., 2003; Jiang *et al*., 2005 Wadwha *et al*., 2016). Ferulic acid is the most abundant aromatic acid in plant cell walls, and acts as a precursor for various aromatic compounds. It is naturally synthesised by plants from caffeic acid (Shahidi and Naczk, 2004). Free-radical-induced oxidative coupling of ferulates produce a range of dehydrodiferulates (Ralph *et al*., 1994). These dimeric compounds function in a similar way to the mono-counterparts by conferring structural reinforcement, by cross-linking polysaccharides to lignin in the cell wall (Fry and Miller, 1989; Yamaoto *et al*., 1989; Ralph *et al*., 1993; Santaigo and Malvar, 2010).

To investigate the oxidatively coupling of ethyl hydroxycinnamic acids with hydrocinnamates esters to synthetically construct homodimers and heterodimers of the constituents of lithospermic acid. This method will produce multiple analogues with a similar or identical pharmacophore presumably achieving similar antiretroviral effects. Ralph *et al.* (2004) have previously reported the homo-dimerization of ethyl ferulate (Figure 2-20), with mediocre yields, nonetheless done cheaply. These synthesised homodimers have been identified having providing important roles, including structural reinforcement by cross-linking polysaccharides to lignin in the cell wall (Fry and Miller, 1989; Yamaoto *et al*., 1989; Ralph *et al*., 1993; Santaigo and Malvar, 2010), extensibility during cell wall growth, digestibility and adhesion (Sanchez *et al*., 1996; Ng *et al*., 1998; Satiago *et al*., 2010). The ethyl ferulate homodimers have varying structures, these are dependent upon radical linkages (Lu *et al*., 2012), 5-5, 8-*O*-4, 8-8, and 8-5 diethyl diferulate.

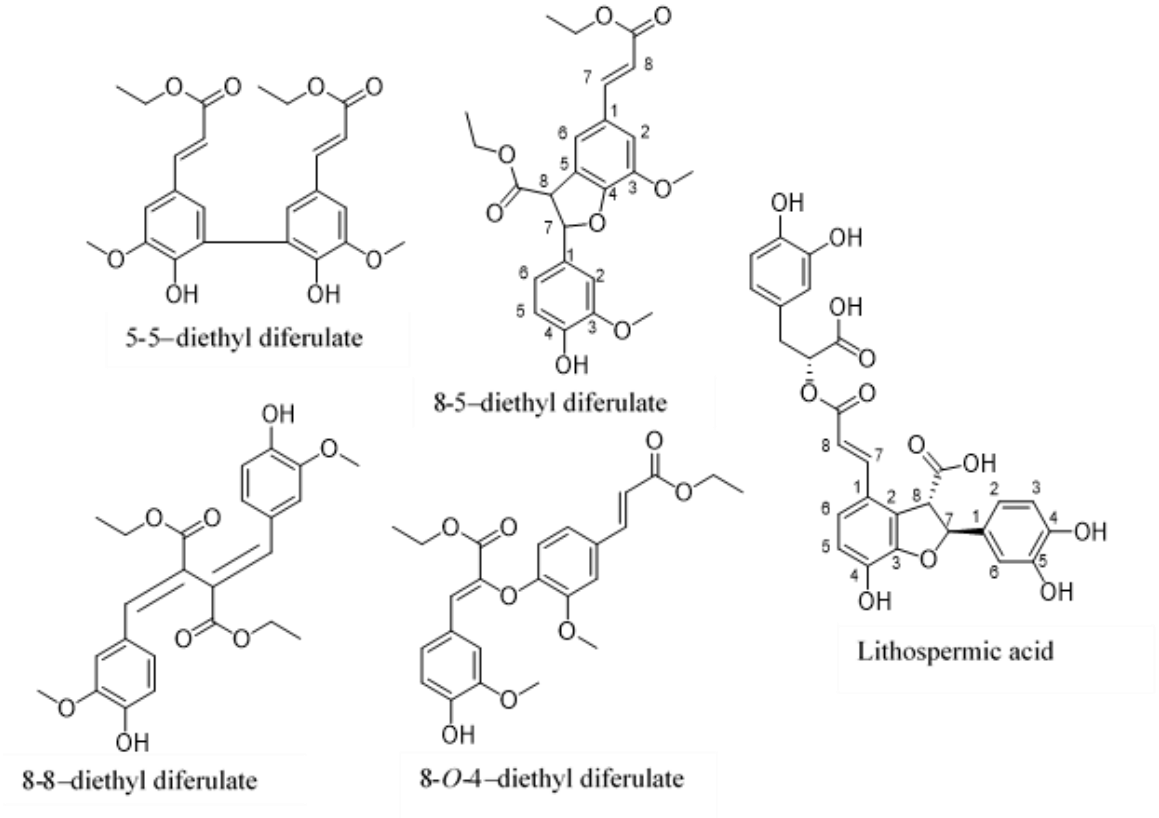


Figure 2-20: Analogues of diethyl diferulate and their similarity to a known HIV-1 IN inhibitor, lithospermic acid. The numbered carbons on 8-5-diethyl diferulate and lithospermic acid identify identical radical linkages. The numbered carbons also demonstrate the linkages, which form the different stereoisomers.

Noticeably, the four diethyl diferulate structures (Figure 2-20) share distinct similarities with that of lithospermic acid and, therefore, it could be suggested that the

similar pharmacophore could also exhibit similar biological activities. 8-5 diethyl diferulate and lithospermic acid both depend on a very distinct dihydrobenzofuran group within the centre of the molecules. The furan ring have been reported in literature to be very favourable for HIV-1 IN inhibition (Schultz *et al*., 2006). A major similarity between 8-5 diethyl diferulate and lithospermic acid is that both contain hydroxy groups at position C4 and a carboxyl group is attached to the furan group, all possible factors for the chelation within the catalytic triad.

2.2 *Specific aims*

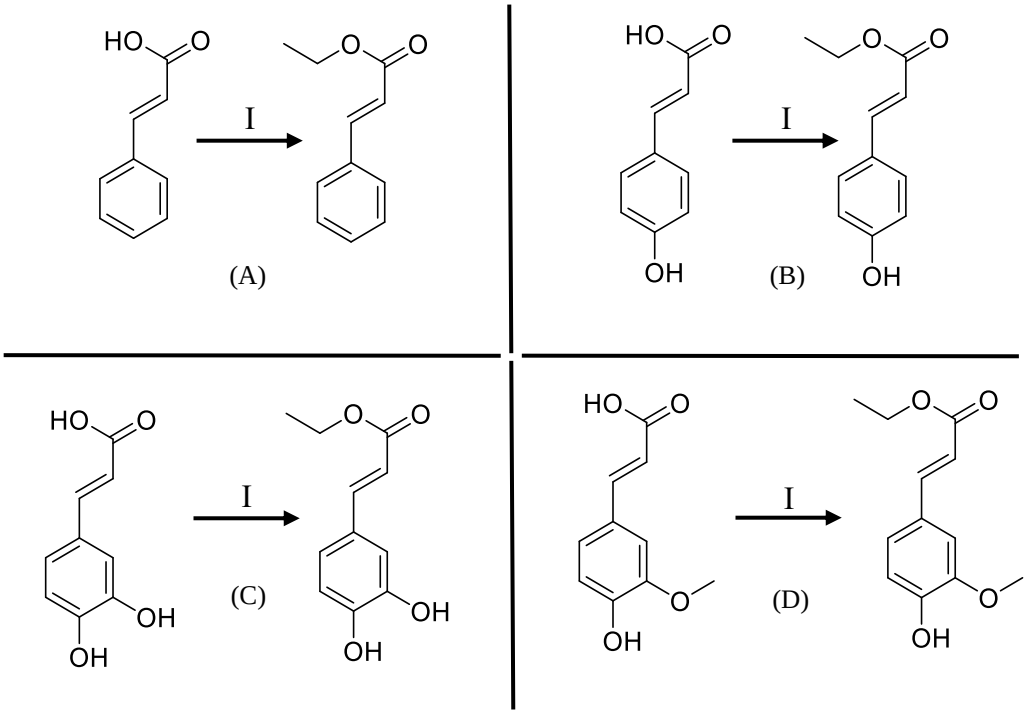
Natural product drug discovery has always been the foundation for new unique compounds to fight a broad range of diseases and infections, with multiple modes of action including; anti-inflammatory, anti-cancer and anti-viral. With the ever increasing demand for pharmacological agents to overcome toxicity and ever-growing resistance, new natural product drug discovery is vital. The aim of the study is to synthesise a portfolio of novel compounds structurally similar to lithospermic acid. With particular attention to the 8-5 radically-linked homo and heterodimer, which closely resemble the reported pharmacophore, therefore potentially possess similar anti-HIV-1 integrase activity.

2.3 *Objectives*

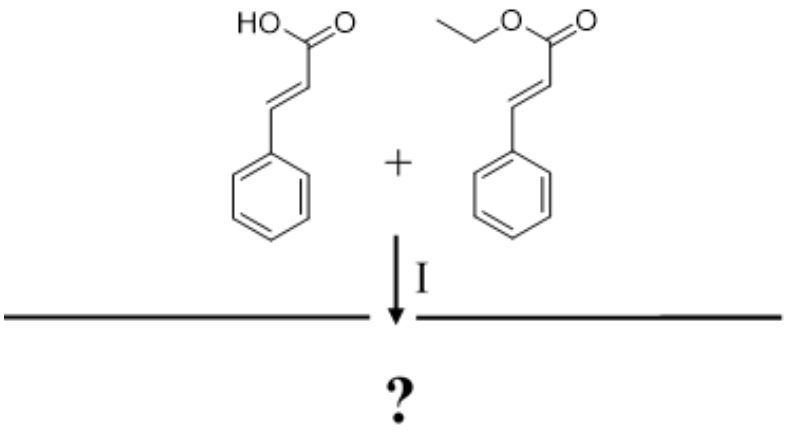
The objectives of the study are to synthesise ethyl cinnamate, ethyl coumarate, ethyl caffeate, and ethyl ferulate *via* an esterification reaction (Scheme 1). Furthermore, to heterodimerize oxidatively 4’-hydroxycinnamic acids with 4’-hydroxycinnamate esters *via* an reaction with functional pharmacophores in a methodical and cost-efficient route (Scheme 2- 5) (Ralph *et al*., 1998).

The acquired samples will be purified using standard laboratory techniques, such as C_{18} reversed-phase column chromatography with the use of ten grams solid-phase extraction columns. HPLC-UV-MS/MS, ultraviolet (UV) spectroscopy and photodiode-array (PDA) detectors will be used to identify the afforded samples for semi-preparative purification. The esters will be fully characterised and structurally elucidated with two-dimensional nuclear magnetic resonance spectroscopy (2D NMR) including ^{1}H, ^{13}C, and correlation spectroscopy (COSY), nuclear Overhauser effect spectroscopy (NOESY) HSQC and HMBC correlations. Melting point (mp) will also be completed for pure samples.

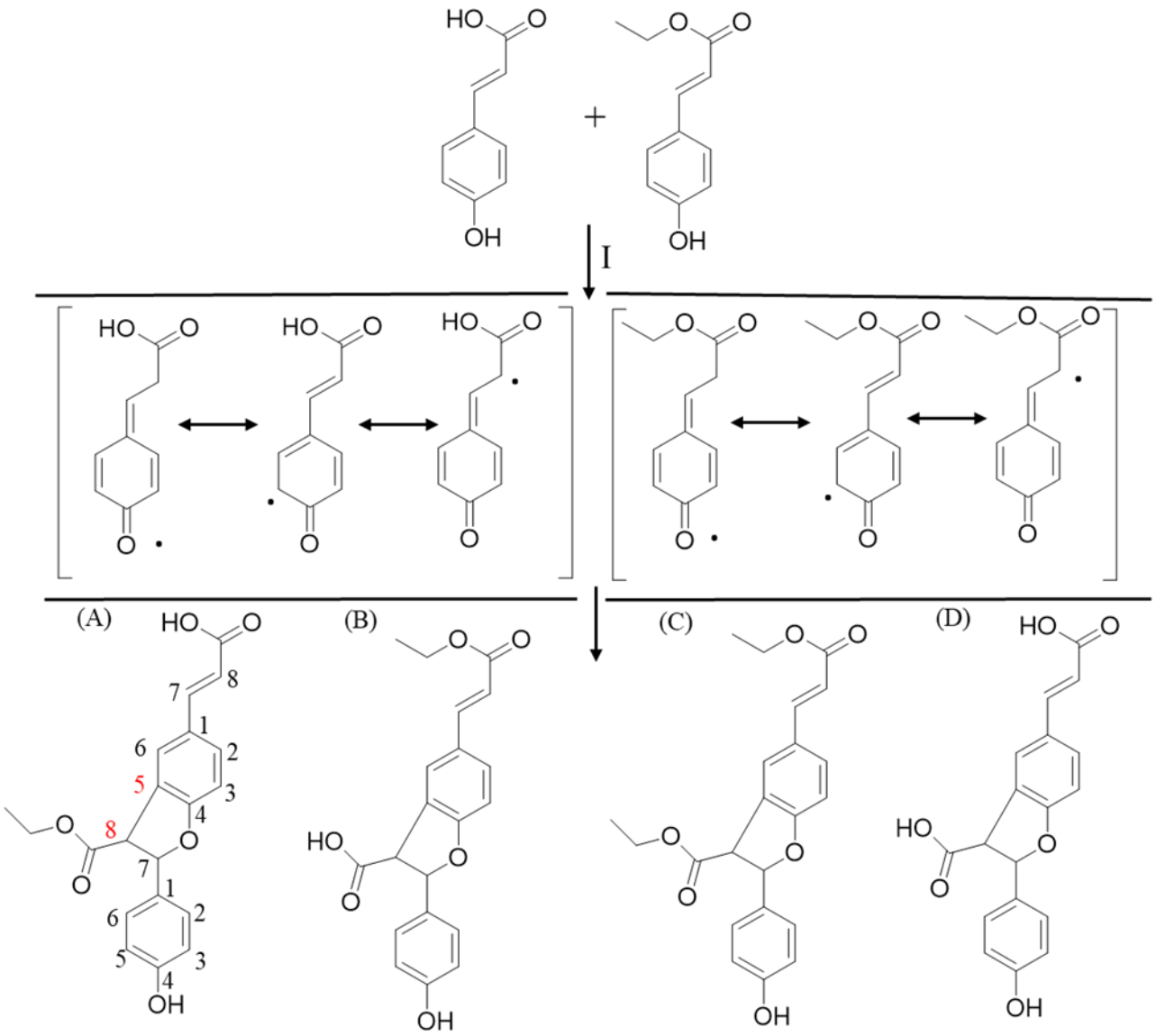
With the wide range of applications of lithospermic acid and its analogues, the purified compounds will be tested for anthelminthic (antiparasitic) and antimicrobial activities. If successful, potential analogues of lithospermic acid that have an anti-HIV-1 integrase effect would have been synthesised in a standard laboratory with minimal costs.



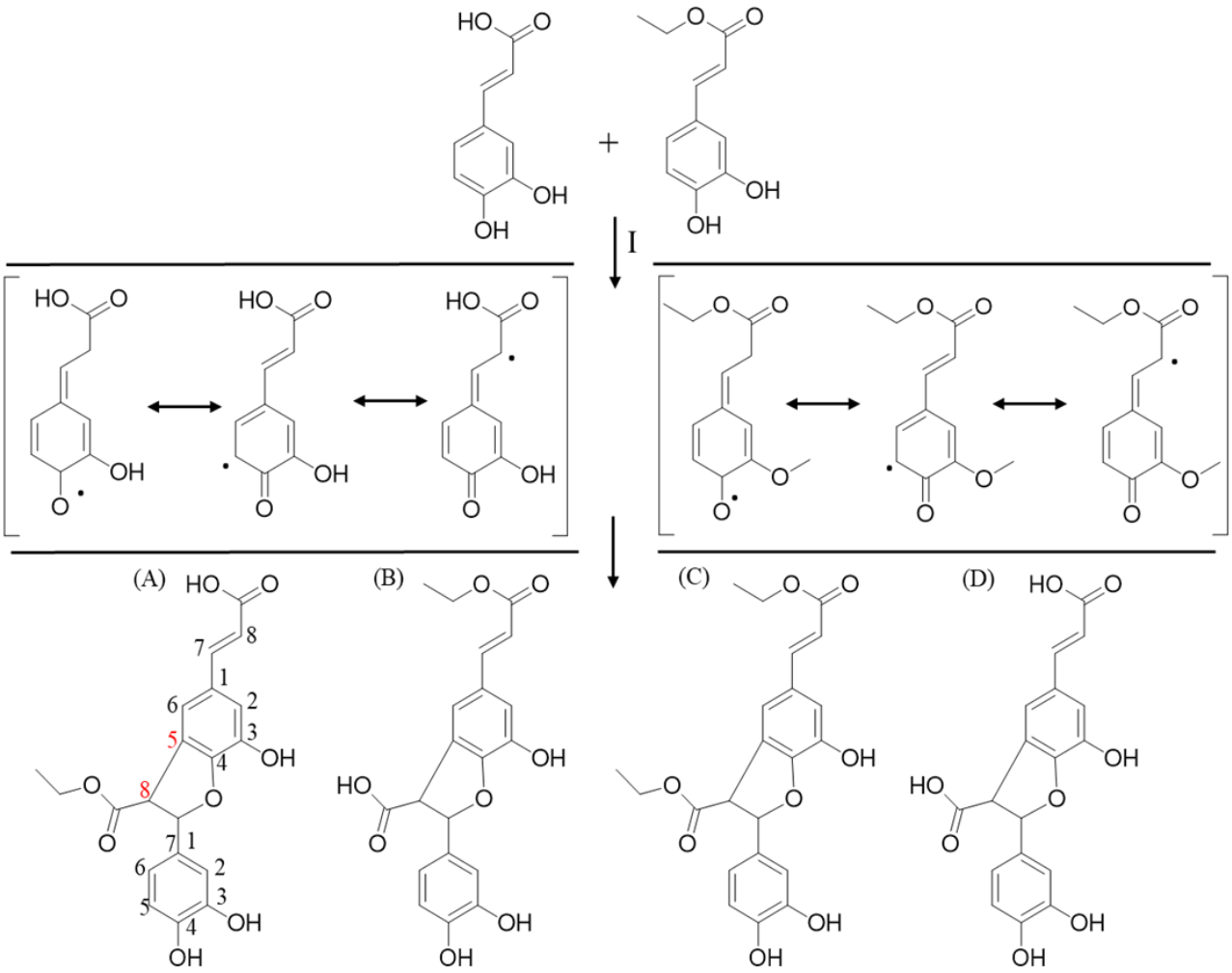
Scheme 2-1: Synthesis of hydroxycinnamate esters **(A)** Esterification of cinnamic acid to form ethyl hydrocinnamates. **(B)** Esterification of coumaric acid to form ethyl coumarate. **(C)** Esterification of caffeic acid to form ethyl caffeic. **(D)** Esterification of ferulic acid to form ethyl ferulate. .*Reagents and conditions:* **I** EtOH, acetyl chloride, ambient temperature for 24 h.



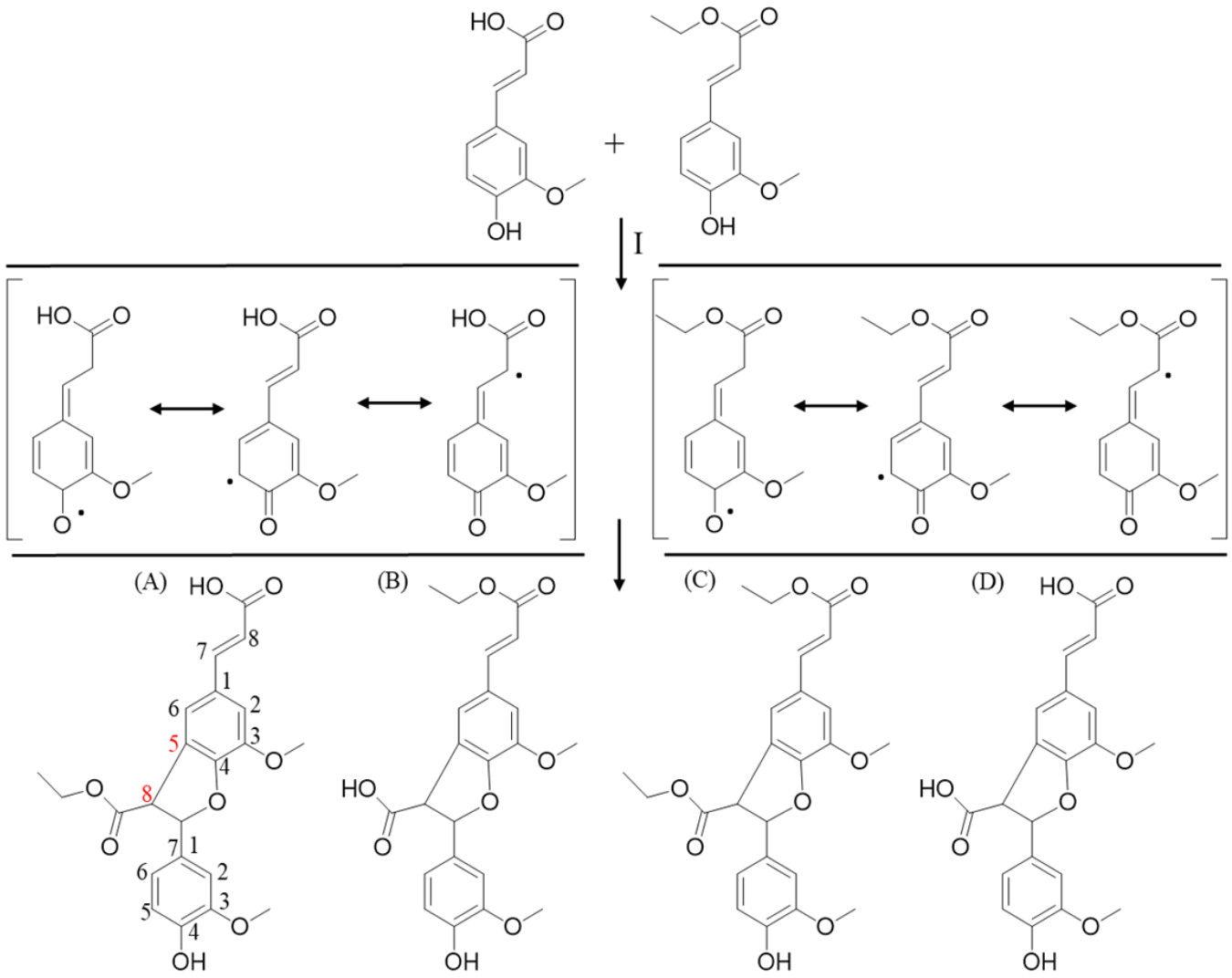
Scheme 2-2: Synthesis of constituent analogues of ethyl cinnamate and cinnamic acid; proposing the synthesis of two 8-5 ethyl cinnamate stereoisomers. To investigate whether the hydroxyl group is necessary for oxidative coupling. *Reagents and conditions:* (I) acetone, hydrogen peroxide urea tablets, pH4; horseradish peroxidase; at ambient temperature and atmospheric pressure.



Scheme 2-3: Synthesis of constituent analogues of ethyl *para*-coumarate and *para*-coumaric acid; proposing the synthesis of two 8-5 ethyl dicoumarate stereoisomers (A) and (B); 8-5 diethyl dicoumarate (C) and dicoumaric acid (D) *Reagents and conditions:* (I) acetone, hydrogen peroxide urea tablets, pH4; horseradish peroxidase; at ambient temperature and atmospheric pressure.



Scheme 2-4: Synthesis of constituent analogues of ethyl caffeate and caffeic acid; proposing the synthesis of two 8-5 ethyl dicaffeate stereoisomers (A) and (B); 8-5 diethyl dicaffeate (C) and dicaffeic acid (D). *Reagents and conditions:* (I) acetone, hydrogen peroxide urea tablets, pH4; horseradish peroxidase; at ambient temperature and atmospheric pressure.



Scheme 2-5: Synthesis of constituent analogues of ethyl ferulate and ferulic acid; proposing the synthesis of two 8-5 ethyl diferulate stereoisomers (A) and (B); 8-5 diethyl diferulate (C) and diferulic acid (D). *Reagents and conditions:* (I) acetone, hydrogen peroxide urea tablets, pH4; horseradish peroxidase; at ambient temperature and atmospheric pressure.

2.4 *Experimental and results*

2.4.1 *General experimental procedures*

Melting points were determined on a Mel-Temp melting point apparatus (Electrothermal Engineering Ltd Rochford, UK) with Fluke 51 II thermometer (Fluke Corporation, Everett, Washington, USA) in open capillaries and are uncorrected. ^{1}H, ^{13}C and 2-D NMR (COSY, HSQC and HMBC correlations) spectra were recorded on a Bruker Avance III instrument operating at 500 MHz for ^{1}H and 125 MHz for ^{13}C, using CD_3OD as the solvent and $SiMe_4$ as the internal standard. Chemical shifts (δ) are given in ppm relative to $SiMe_4$. Experiments were conducted at ambient temperature, unless otherwise noted. 1D NMR (^{1}H, ^{13}C) and 2D NMR (HSQC, HMBC and COSY) experiments used standard Bruker pulse programs, the instrument was run by Dr. Tim Woodman, University of Bath, Department of Pharmacy & Pharmacology.

HPLC-UV-ESI-MS/MS analysis was performed as previously described (Parveen *et al*., 2001) with a Thermo Finnigan LCMS System (Thermo Electron Corporation, Waltham, Massachusetts, USA). To a glass vial containing crude extract (1 mg) was added HPLC grade MeCN (1 mL). The tube was vortexed, centrifuged and 100 µL was transferred into a HPLC vial for analysis in both positive and negative ion mode. The system included a Finnigan Surveyor PDA and alongside a Finnigan LTQ (linear trap quadrupole) with an ESI source. Chromatography was performed on a Waters C_{18} reverse-phase (3.9 × 100 mm i.d., 4 µm) Nova-Pak column (Waters Corporation, Milford, Massachusetts, USA). The column temperature was constant at 50°C, the columns high pressure limit was 5000 psi and the temperature of the auto-sampler was maintained at 30 ºC. The sample injection volume was 10 µL, the detection wavelength was set at 240-400 nm with a flow rate of 1.0 mL min^{-1}. The mobile phase was composed of purified water with 0.1% formic acid (solvent A) and MeOH with 0.1% formic acid (solvent B). The process of column equilibration was performed with 95% A: 5% B. During the analyses, the percentage of B increased linearly to 100% over a period of 30 min. The parameters; capillary temperature 320 ºC, nitrogen sheath gas 30 arbitrary units and nitrogen auxiliary gas at 15 arbitrary units, were used to obtain mass spectra in positive and negative modes. In negative-ionisation mode, spray voltage was set to 4.8 KV, capillary voltage was 45 V and tube lens offset was set at 110 V.

Orbitrap Fusion Tribrid Mass spectrometer – LC-HRMS analysis was performed on an Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific) that was coupled to UltiMate 3000 liquid chromatography tower (Dionex, Thermo Scientific). A Hypersil Gold (Thermo Scientific) reverse phase C18 column (2.1 mm × 150 mm; particle size 5 µm) was used for chromatography, maintained at a temperature of 60°C. 10 µL of sample was injected. The mobile phases for gradient elution were ultra- pure water (18.2 MΩ) with 0.1% formic acid (mass spectrometry grade, Fluka), (solvent A) and MeOH (HPLC grade, Fisher Scientific) with 0.1% formic acid (solvent B). The flow rate was fixed at 0.6 mL min^{-1}. The initial condition was 100% A with a linear increase to 100% B over 7 min. 100% B was held for 3.5 min before equilibration at initial conditions for a further 2.5 min. Ions were generated in a HESI-II source with a source voltage of 3500/2500 V for positive/negative mode, sheath gas: 45, aux gas: 13, a vaporiser temperature of 358°C and an ion-transfer temperature of 342°C. Ions were detected in profile mode in the *m/z* 100-2000 range in the Orbitrap detector at a resolution of 240000 and an injection time of 100 ms in both positive and negative mode.

The reactions were monitored by thin layer chromatography (TLC) performed on a silica gel 60 F_{254} pre-coated on aluminium (Merck KGaA, Darmstadt, Germany) and spots were visualised under UV at 254 nm. Silica gel (70-230 mesh) (Sigma-Aldrich, St. Louis, Missouri, USA) was used for column chromatography. C18 silica gel 230-400 mesh was used as a binding agent for solid-phase extraction (SPE) (Supleco, analytical SUPERCLEAN LC-18 packing 60 mL, 10g column) and Biotage flash chromatography. Celite S was used as a filter aid (Merck KGaA, Darmstadt, Germany).

To identify the peak of interest and the corresponding UV patterns, a method development column is used. F14 Symmetry C8 (3.5 µm, 4.6 × 50 mm column) with a gradient of H_2O to ACN 100% over 13 min, containing 0.01% trifluoroacetic acid. All samples were analysed at a concentration of 10 mg mL^{-1} in MeOH. Once a peak of interest has been identified, individual method development can take place on a longer column to achieve separation. The column used for further method development was HICHROM ACE 5 C_{18} (3.5 µm, 4.6 × 250 mm); after development, it can be taken straight onto the semi-preparative system. Semi-preparative system: Water HPLC prep LC system controller 2000 coupled to a Water 2487 dual λ absorbance detector was

used for semi-preparative HPLC. The column used was a HICHROM ACE 10 C_{18} (250 × 21.2 mm). Samples were prepared at a concentration of 10 mg mL^{-1} by dissolving the crude product in the organic solvent used for mobile phase. All solvents used were of HPLC grade. Each purification had different starting conditions and mobile phase compositions, due to the previous analytical methods stated above. Before injection, samples were mixed with water to match starting conditions. The flow rate, wavelengths and injection volumes used were the same for all methods: 15 ml min^{-1} and 210 nm. The injection volume was 2.0 mL. All absorbance's were recorded on a Kipp and Zonen BD 12E Flatbed recorder BD12E.

All chemicals and solvents were obtained from commercial suppliers and were used without further purification: *trans*-4-hydroxy-3-methoxycinnamic acid, 3,4-dihydroxybenzeneacrylic acid; Horseradish peroxidase (Merck KGaA, Darmstadt, Germany); Acetyl chloride, +99%, ACROS Organics™ (Fisher Scientific, Waltham, Massachusetts, USA). Tetrahydrofuran (THF) and Na_2SO_4 (Sigma-Aldrich, St. Louis, Missouri, USA), HPLC-grade solvents: MeOH, EtOH, EtOAc, MeCN, hexane and petroleum ether (Fisher Scientific, Waltham, and Loughborough, UK).

2.5 *Synthesis of ethyl cinnamate*

Acetyl chloride (10.0 mL, 128 mmol) was added cautiously to cinnamic acid (5.00 g, 22.4 mmol) in EtOH (250 mL) and the mixture was stirred for 24 h. The progress was monitored by TLC (CH_2Cl_2). The EtOH was evaporated, the residue in EtOAc was washed thrice with saturated brine. Drying (Na_2SO_4) and evaporation gave ethyl cinnamate (4.85 g, 81%) as a colourless oil which formed white crystals at -18°C (lit. (Kikukawa *et al.*, 1981) oil).

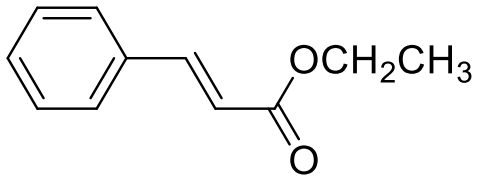


Figure 2-21: Structure of ethyl cinnamate

2.5.1 *Synthesis results of ethyl cinnamate*

Cinnamic acid was converted to ethyl cinnamate *via* an acid-catalysed esterification, with the use of acetyl chloride. The reaction mixture was monitored using TLC,

identifying a single spot that was less polar than the standard cinnamic acid. After the starting material cinnamic acid (R_f = 0.53) had been consumed and converted fully to ethyl cinnamate (R_f = 0.83), the solvent was evaporated.

Ethyl cinnamate gave a t_R of 21.05 min in the HPLC chromatogram Figure 2-22. In positive ion mode MS, gave a ion at *m/z* 177.0914 $[M+H]^+$ corresponded to the pseudomolecular ion of $C_{11}H_{12}O_2$ (calcd: 176.08373). Other ions were also evident, *m/z* 353 $[2M+H]^+$ and a sodium adducts *m/z* 199 $[M+Na]^+$. Secondary ionisation (MS_2) of the pseudomolecular ion 177 $[M+H]^+$ gave fragments *m/z* 99 $[(M+H)\text{-}C_6H_5]^+$, 130 $[(M+H)\text{-}HOCH_2CH_3]^+$, 135 $[(M+H)\text{-}CO_2]^+$, 145 $[(M+H)\text{-}2H_2O]$, 162 $[(M+H)\text{-}CH_3]^+$ and 159 $[(M+H)\text{-}H_2O]^+$ which are in accordance with the fragmentation patterns of 4'-hydroxycinnamate esters (Zhang *et al*., 2013). Relatively small amounts of by-products were visible in the chromatogram. The UV spectrum of ethyl cinnamate was characterised by two bands with 253 nm and 287 nm (Yabe *et al*., 1972). Band II was produced by the cinnamoyl group (Aksoz and Ertan, 2012). ^{1}H NMR gave a triplet at δ 1.34 integrating for 3 protons and a quartet integrating for two protons at δ 4.27 (*J* = 7.2 Hz). The further downfield chemical shifts are consistent with being further away from an electron-withdrawing function. The alkene protons at 7 and 8 were observed at δ 6.28 and δ 6.91 as doublets with a *trans* coupling constant of *J* = 16 Hz. ^{13}C NMR provided seven distinct chemical environments. The successful synthesis of ethyl cinnamate (Figure 2-23), the analysis of ^{1}H NMR and ^{13}C NMR of ethyl cinnamate is consistent of that found in literature (Wu *et al*., 1995; Kelly, 2011) ^{1}H NMR ($CDCl_3$) δ 1.34 (3 H, t, *J* = 7.1 Hz, CH_3), 4.27 (2H, q, *J* = 7.1 Hz, OCH_2), 6.44 (1 H, d, *J* = 16.2 Hz, $EtO_2CC\mathit{H}$=), 7.46 (1 H, tt, *J* = 7.2, 1.5 Hz, Ph 4-H), 7.43 (2 H, t, *J* = 8.0 Hz, Ph 3,5-H_2), 7.54 (2 H, d, *J* = 8.0 Hz, Ph 2,6-H_2), 7.73 (1 H, d, *J* = 16.2 Hz, Ph$C\mathit{H}$=); ^{13}C NMR ($CDCl_3$) δ 14.2 (CH_3), 61.4 (CH_2), 116.5 ($EtO_2C\mathit{C}H$=), 127.8 (Ph 4-C), 128.5 (Ph 2,6-C_2), 128.6 (Ph 3,5-C_2), 135.2 (Ph 1-C), 145.1 (Ph-$\mathit{C}H$=), 166.5 (CO_2); ESIMS$^+$ *m/z* 177 $[M + H]^+$.

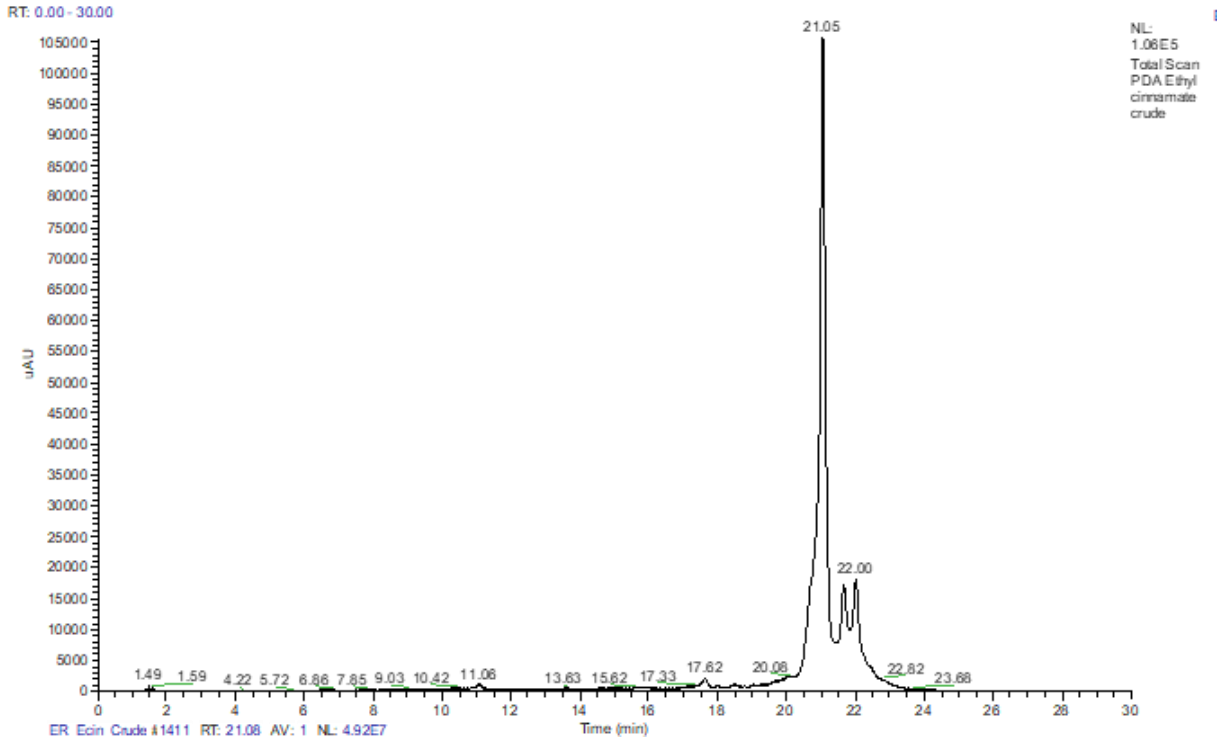


Figure 2-22: HPLC-UV-ESI-MS/MS of ethyl cinnamate

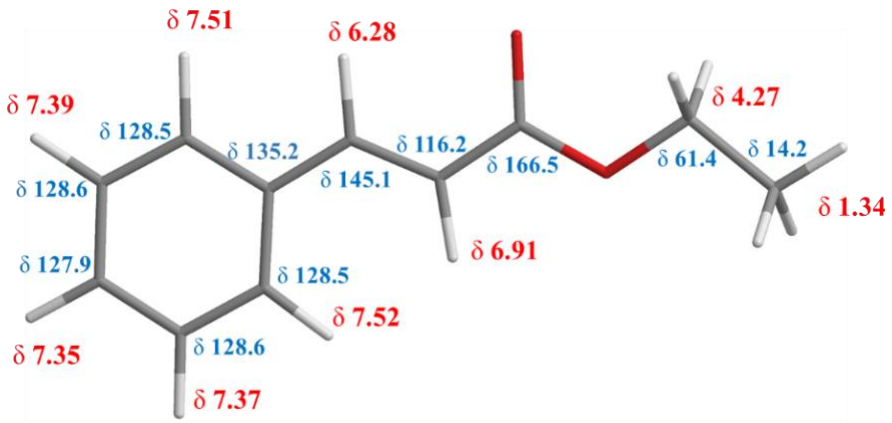


Figure 2- 23: Ethyl cinnamate structural formula with appropriate $_{1}$H NMR (Red) and $_{13}$C NMR (Blue) shift data.

2.6 *Oxidative heterodimerisation of cinnamic acid and ethyl cinnamate*

Ethyl coumarate (1.838 g, 9.57 mmol) and coumaric acid (1.570 g, 9.57 mmol) was added acetone (240 mL) gently heated until dissolved. Purified H_2O (440 mL) was added cautiously, followed by urea hydrogen peroxide (1.24 g, 13 mmol) in purified H_2O (30 mL). The pH was monitored and maintained at pH 4 (Schomberg, Salzmann, and Stephan, 1993), followed by the addition of horseradish peroxidase type 2 (10 mg) in H_2O (10 mL). The transparent mixture was stirred at ambient temperature for 1.5 h. Due to the lack of colour change, 2 mL of the reaction mixture was removed and added

ferulic acid, with an immediate colour change. Indicating the enzyme was still chemically active. Further addition of urea H_2O_2 (540 mg, 5.74 mmol) in H_2O (3 mL) was needed. The mixture was mixed for another 1h, after which the compound was obtained into EtOAc. The residue, in EtOAc was washed thrice with saturated brine. Drying (Na_2SO_4) and evaporation gave an off-white opaque crude oil (3.63 g).

2.6.1 *Results of heterodimerisation of cinnamic acid and ethyl cinnamate*

The attempted oxidative coupling reaction between ethyl cinnamate and cinnamic acid was unsuccessful. The reaction was monitored at pH 4 as the enzyme was added to the reaction mixture; there was no significant colour change. A sample of the reaction mixture was taken every hour and monitored by TLC (Silica gel, DCM 100%), for each TLC run, the two starting materials remained clearly visible on the TLC plate; R_f = 0.25 and 0.40 for cinnamic acid and ethyl cinnamate respectively. After 24 h at ambient temperature and pressure, the reaction was quenched. HPLC – ESI-UV-MS/MS (ESI negative mode) of the reaction mixture indicated two peaks at t_R 15.18 min, *m/z* 149.04 $[M+H]^+$ and 21.14 min, *m/z* 176.98 $[M+H]^+$, integrating for cinnamic acid and ethyl cinnamate pseudomolecular ions respectively.

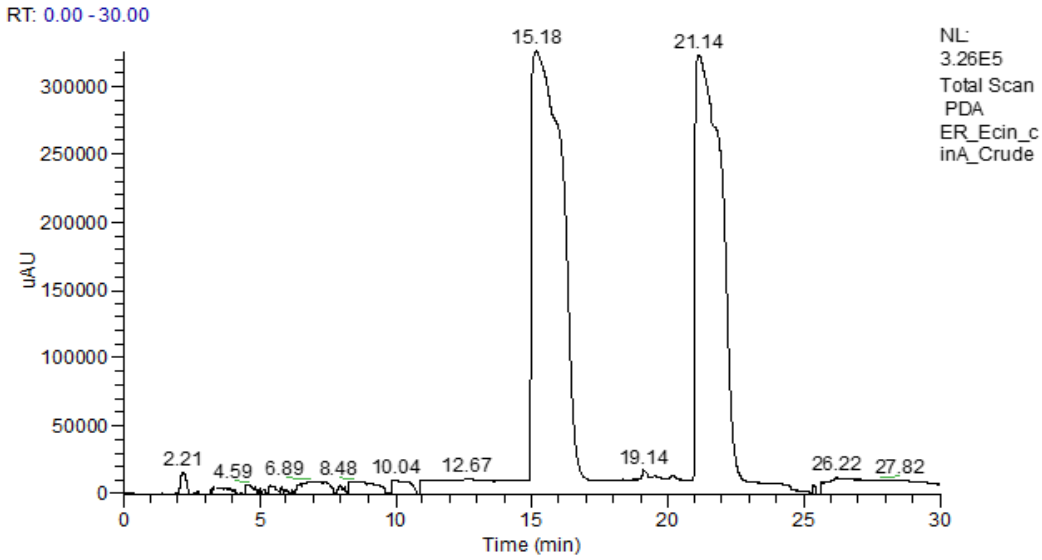


Figure 2-24: HPLC-UV of the attempted oxidative coupling of ethyl cinnamate t_R 21.14 min and cinnamic acid t_R 15.18 min respectively.

To ensure the enzyme was still chemically active and had not denatured in storage, 1.0 mL of the reaction mixture was removed during synthesis. To this mixture, ferulic acid was added, with the observation of an immediate colour change. The reaction mixture

was monitored via TLC. A gradual disappearance of ferulic acid and the presence of a more non-polar spot was observed, diferulic acid. Cinnamic acid and ethyl cinnamate were still visible with the consistent R_f values. The oxidative reaction had not taken place due to the lack of the hydroxy group on the phenol. This functional group is essential for redox reaction with H_2O_2 and the horseradish peroxidase enzyme.

2.7 *Synthesis of ethyl para-coumarate (ethyl 4-hydroxycinnamate)*

Acetyl chloride (36.0 mL, 460 mmol) was added cautiously to 4-hydroxycinnamic acid (5.00 g, 30.5 mmol) in EtOH (250 mL) and the mixture was stirred for 24 h. The progress was monitored by TLC (CH_2Cl_2 : EtOAc / 1:1). The EtOH was evaporated. The residue, in EtOAc, was washed thrice with saturated brine. Drying (Na_2SO_4) and evaporation gave 4-hydroxycinnmate (4.98 g, 85 %) as off-white crystals: m:p: 65-68 °C (lit. (Zhang *et al*., 2015) 74-75°C); R_f = 0.3 (CH_2Cl_2: EtOAc / 1:1).

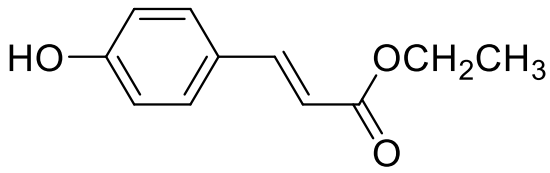


Figure 2-25: Structure of ethyl *para*-coumarate.

2.7.1 *Synthesis results of ethyl para-coumarate*

Para-coumaric acid was converted to ethyl *para*-coumarate *via* an acid-catalysed esterification reaction using acetyl chloride. The reaction mixtures was monitored using TLC. Ethyl *para*-coumarate is more non-polar than *para*-coumaric acid therefore the spot was higher on the TLC plate. Following absence of the starting material, ethyl *para*-coumarate gave an R_f = 0.3 (Silica gel, 100 % DCM). Yielding off white crystals in 85 % yield.

Ethyl *para*-coumarate was subsequently analysed using HPLC-ESI-UV-MS/MS (ESI negative mode) whereby the UV chromatogram indicated the presence of a single compound at t_R 17.57 min with a UV chromophore typical of the ester of 310 nm, this absorption is typical of ethyl *para*-coumarate (Sun *et al.*, 2007; Said *et al*., 2017). In negative ion mode the peak at t_R 17.57 min, *m/z* 191 $[M-H]^-$ corresponded to the pseudomolecular of $C_{11}H_{13}O_3$ (calcd: 192.07865). Secondary ionisation MS_2 of the

pseudomolecular ion *m/z* 191 [M-H]- yielded three fragments, a base peak at *m/z* 163 [M-H-ethene]-, typical of the loss of the ethyl chain by McLafferty rearrangement, further fragments include *m/z* 145 [M-HOCH$_2$CH$_3$]- and *m/z* 118 [M-CO$_2$-CH$_2$CH$_3$]-. In positive ion mode, a base peak was detected at *m/z* 193 [M+H]$^+$, whilst a less abundant ion at *m/z* 195 [(M+H)-ethene]$^+$ by McLafferty rearrangement. A formate adduct was also present as the base peak at *m/z* 236 [(M-H)+COOH]-, formed during the ionisation process. Further fragmentation of *m/z* 193 [M+H]$^+$ gave fragments in MS$_2$ ionisation, a base peak at *m/z* 147 [(M+H)-OHCH$_2$CH$_3$]$^+$, confirming the loss of an ethanol group, indicative of an ethyl ester moiety. An ion was observed at *m/z* 165 [(M+H)-CH$_2$CH$_3$]$^+$, indicating the loss of the ethyl fragment, also an ion at *m/z* 175 [(M+H)-H$_2$O]$^+$, the loss of water. The melting point of ethyl *para*-coumarate was 67-70ºC, this is consistent with literature (Plieninger *et al.*, 1965; Hixson *et al.*, 2012).

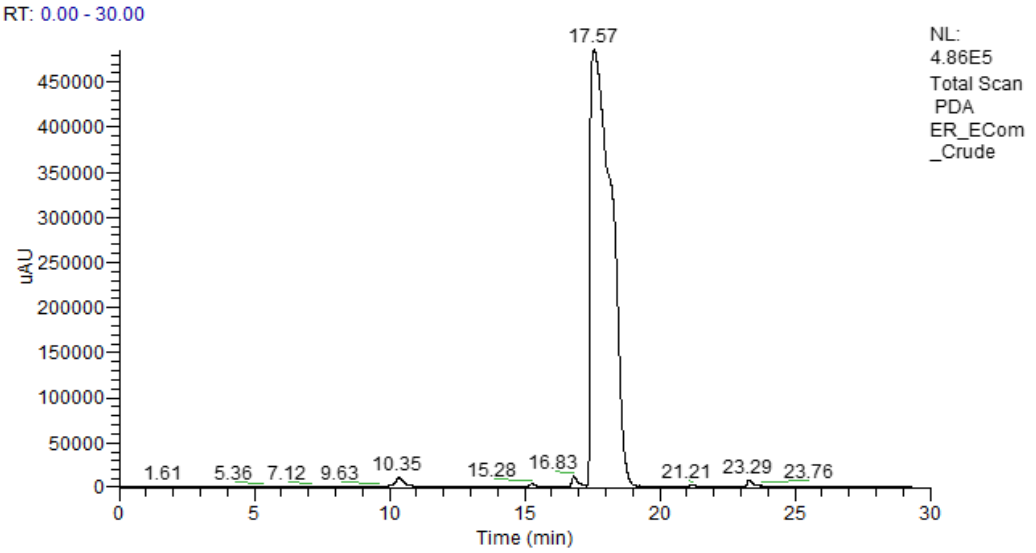


Figure 2-26: HPLC-UV of ethyl *para*-coumarate crude.

^{1}H NMR and ^{13}C NMR analysis (Figure 2-27) gave a triplet at δ 1.22 integrating for 3 protons. A quartet integrating for two protons at δ 4.19 (*J* = 7.2 Hz). The same chemical shift imply that the protons are in a symmetrical environment. The further downfield chemical shifts are consistent with being further away from an electron-withdrawing function. The alkene protons at 7 and 8 were observed at δ 6.31 and δ 7.48 as doublets with a trans coupling constant of *J* = 16 Hz. With six different proton environments and nine different carbon environments, the synthesised ethyl *para*-coumarate is consistent with that found in literature (Cernerud *et al.*, 1996; Lee *et al.*, 2013). ^{1}H NMR (CDCl$_3$) δ 1.22 (3 H, t, *J* = 7.1 Hz, CH$_3$), 4.19 (2 H, q, *J* = 7.1 Hz, CH$_2$), 6.31 (1

H, d, *J* = 16.2 Hz, =C*H*CO_2Et), 6.65 (2 H, d, *J* = 8.0 Hz, Ar 3,5-H_2), 7.48 (1 H d, *J* = 16.1 Hz, ArC*H*=), 7.57 (2 H, d, *J* = 8.0 Hz, Ar 2,6-H_2), 7.75 (1 H, d, *J* = 16.2 Hz, ArC*H*=); ^{13}C NMR ($CDCl_3$) δ 14.2 (CH_3), 61.3 (CH_2), 115.6 (=*C*HCO_2Et), 115.8 (Ar 3,5-C_2), 127.8 (Ar 1-C), 130.6 (Ar 2,6-C_2), 146.5 (Ar*C*H=), 157.7 (Ar 4-C), 167.1 (C=O).

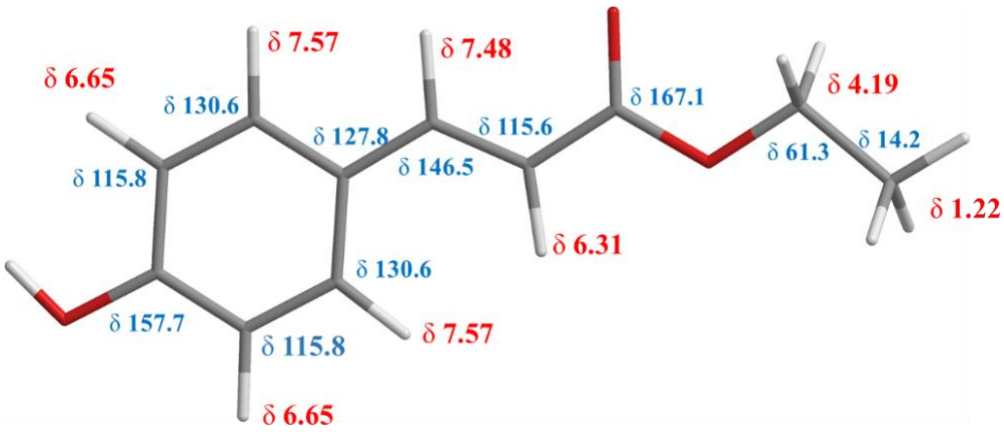


Figure 2-27: Ethyl *para*-coumarate structural formula with appropriate ^{1}H NMR (Red) and ^{13}C NMR (Blue) shift data.

2.7.2 *Oxidative heterodimerisation of ethyl para* coumarate and coumaric acid

Ethyl coumarate (1.838 g, 9.57 mmol) and coumaric acid (1.570 g, 9.57 mmol) was added acetone (240 mL) gently heated until dissolved. Purified H_2O (1440 mL) was added cautiously, followed by urea hydrogen peroxide (990 mg, 10.52 mmol) in purified H_2O (10 mL). The pH was monitored and maintained at pH 4 (Schomberg, Salzmann, and Stephan, 1993), followed by the addition of horseradish peroxidase type 2 (10 mg) in H_2O (10 mL). The dark yellow coloured mixture was stirred at ambient temperature for 1.5 h. Further addition of urea H_2O_2 (540 mg, 5.74 mmol) in H_2O (3 mL) was needed. The mixture was mixed for another 1h, after which the compound was obtained into EtOAc. The residue, in EtOAc was washed thrice with saturated brine. Drying (Na_2SO_4) and evaporation gave a crude orange/brown crystals (2.4815 g).

2.7.3 *Results of heterodimerization of ethyl para coumarate and coumaric acid.*

Ethyl *para*-coumarate and *para*-coumaric were oxidatively coupled *via* a biomimetic hydrogen peroxide horseradish peroxidase, one-step reaction. The starting materials were solubilised in warm acetone and the pH monitored and maintained at pH 4. 1 mL of the reaction mixture was removed and washed with the standard procedure. The TLC indicated starting materials were still present, a further 540 mg H_2O_2 urea tablets was added to the reaction mixture and stirred for a further 30 mins and monitored by TLC. After 30 mins the TLC results had a large streak from top to bottom, with three smaller visible bands. The crude product was extracted into EtOAc, washed with brine and dried Na_2SO_4. The solvent was remove and to produce 2.4815 g of orange/brown crystals, the sample was subsequently analysed by HPLC-ESI-UV-MS/MS (Figure 2-28).

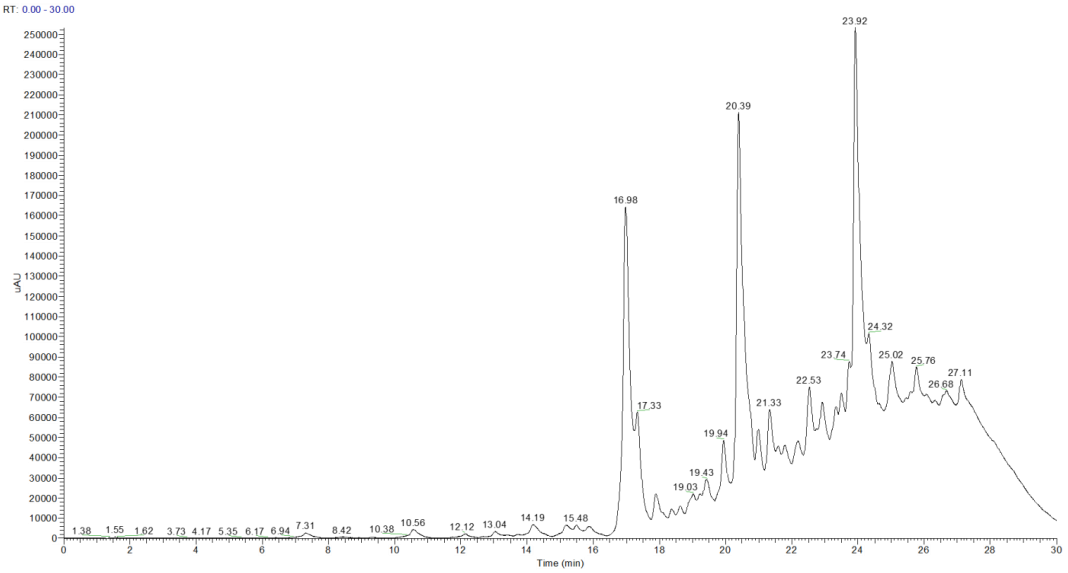


Figure 2-28: HPLC-UV-ESI chromatogram of compounds **1** at t_R 16.98 min, **2** at t_R 20.39 min and **3** at t_R 23.92 within the crude. After purification the t_R differ slightly these are recorded below.

2.7.3.1 *Isolera Biotage Flash Chromatography*

Using Biotage Flash Chromatography, the crude product was separated based on polarity, differences in UV absorbances and t_R. The crude extract (200 mg) was purified on an C-18 reverse phase cartridge using the mobile phase 65% water to 100 % MeOH over 70 min, at a flow rate of 25 mL/min. The sample was loaded onto the cartridge pre-absorbed onto C-18 silica. The run collected 14 fractions in total yield of 67 % mass recovery, (133.80 mg). These were determined based on the HPLC

chromatogram of the set wavelengths of 254 and 330 nm. (Biotage chromatogram available in supporting information S2). HPLC-ESI-UV-MS/MS analysis of the fractions highlighted the three fractions of interest, corresponding to the three sharp peaks in Figure 2-27. Compound **1** and **2** ionised better within the positive ion mode and therefore had more complex fragmentation patterns, compound **3** ionised better within the negative ion mode.

Table 2-1: Isolera Biotage Flash chromatographic fractions of oxidatively coupled ethyl para-coumarate and coumaric acid.

Fraction / Sample name GH ‘ ’	***Mass (mg)***	***Fraction / Sample name GH ‘ ’***	***Mass (mg)***
1	4.6	***8***	7.4
2	2.0	***9***	9.1
3	4.1	***10***	2.9
4	8.3	***11***	7.1
5	15.4	***12***	14.6
6	4.4	***13***	12.9
7	10.2	***14***	4.7

2.7.3.2 Ethyl para-coumarate and para-coumaric acid heterodimer

Compound **1** (*GH_2*): HPLC – ESI-UV-MS/MS (ESI negative mode) of **1** showed a peak at t_R 17.61 min, *m/z* 352.95 corresponding to $[M-H]^-$ for $C_{20}H_{18}O_6$ (calcd: 354.11034) the pseudomolecular ion of a heterodimer. λ_{max}, 321.00 nm. 2 mg 0.2 % yield. In negative ion mode at t_R 17.61 min, the base peak at 100 % intensity *m/z* $[2M-H]^-$ corresponds to cluster of the pseudomolecular ion. A second ion was observed at *m/z* 398 $[(M-H)+HOCH_2CH_3]^-$, the ethanol motif formed during the ionisation process. The corresponding dimer was also observed at *m/z* 752 $[2(M-H)+HOCH_2CH_3]^-$. Secondary MS_2 of the pseudomolecular ion *m/z* 353 $[M-H]^-$ yielded two fragments *m/z* 308 $[M-COOH]^-$, the loss of *m/z* 45 confirming the presence of a carboxylic acid group and *m/z* 262 $[M-HOCH_2CH_3.COOH]^-$, indicating the presence of one carboxylic acid and one ethyl ester. A base peak was detected at *m/z* 355 $[M+H]^+$, corresponding to the pseudomolecular ion. Further ions was observed at *m/z* 710 $[2M+H]^+$, integrating for the cluster of the pseudomolecular ion, *m/z* 337 $[(M+H)-H_2O]^+$, *m/z* 309 $[(M+H)-HOCH_2CH_3]^+$ representing the loss of the ethyl ester, *m/z* 193 [(M+H)- ethyl *para-*

coumarate]+, *m/z* 310 [(M+H)- COOH]+ . In MS_2 positive ion mode, the base peak was observed at *m/z* 322 [(M+H)-$CH_3.H_2O$]+. The sample contains another three smaller peaks therefore is unsuitable for melting point and NMR analysis, the sample is also unsuitable for further purification due to the low yield. Only 8 % of the total starting crude was loaded onto the C-18 Biotage column, further runs on the Biotage will yield the required amount for semi-preparative purification and subsequently NMR analysis.

Compound **2** (*GH_5*): HPLC – ESI-UV-MS/MS (ESI positive mode) of **2** showed a peak at t_R 21.05 min, *m/z* 353.1024 corresponding to [M+H] + for $C_{20}H_{18}O_6$ (calcd: 354.11034) the pseudomolecular ion a heterodimer. m.p. 107.0-109.4 ºC. λ_{max}, 314.00 nm. 15.40 mg 0.5 % yield. In negative ion mode the base peak ion was observed at *m/z* 706.00 [2M-H]-, corresponding to the cluster of the pseudomolecular ion. A second ion was observed at *m/z* 398 [(M-H)+$HOCH_2CH_3$]-. Negative MS_2 of the pseudomolecular ion *m/z* 352 [M-H]-, yielded three significant fragments, *m/z* 325 [M-CO]-, *m/z* 307 [M-$HOCH_2CH_3$]-, *m/z* 263 [M-$HOCH_2CH_3.CO_2$]-. In positive ion mode ion *m/z* 355 $[M+H]^+$ corresponds to the pseudomolecular ion. Further ions observed were *m/z* 337 [M-H_2O]+, *m/z* 369 [(M-H)+CH_3]+, in MS_2 positive ion mode, seven fragments were observed at *m/z* 309 [(M+H)-$HOCH_2CH_3$]+, equating to the loss of the ethyl ester, *m/z* 337 [(M+H)-H_2O]+, *m/z* 322 [(M+H)-$CH_3.H_2O$]+, *m/z* 291 [(M+H)-$HOCH_2CH_3.H_2O$]+, *m/z* 267 [(M+H)-$2CO_2$]+ and *m/z* 265 [(M-H)-$HOCH_2CH_3.CO_2$]+. ^{1}H NMR and ^{13}C NMR are shown below (Figure 2-29). ^{1}H NMR ($CDCl_3$) δ 1.31 (3 H, t, *J* = 7.1 Hz, CH_3), 4.26 (1 H, dq, *J* = 14.3, 7.2 Hz, CO_2CHH), 4.29 (1 H, dq, *J* = 14.3, 7.2 Hz, CO_2CHH) 4.37 (1 H, d, *J* = 7.5 Hz, benzofuran 3-H), 6.03 (1 H, d, *J* = 7.5 Hz, benzofuran 2 – H), 6.38 (1 H, d, *J* = 15.9 Hz, =$CHCO_2$), 6.87 (2 H, d, *J* = 8.7 Hz, Ph 3,5 – H_2), 6.92 (1 H, d, *J* = 8.4 Hz, benzofuran 7-H), 7.29 (2 H, d, *J* = 8.6 Hz, Ph 2,6-H_2), 7.61 (1 H, dd, *J* = 8.4, 1.8 Hz, benzofuran 6 – H), 7.65 (1 H, d, *J* = 16.0 Hz, benzofuran –C*H*=), 7.72 (1 H, brs, benzofuran 4- H).^{13}C NMR δ 14.46 (CH_3), 55.70 (benzofuran 3-C), 62.22 (CH_2), 87.72 (benzofuran 2-C), 110.76 (benzofuran 7-C), 116.26 (Ph 3,5-C_2), 116.47 (=*C*HCO_2H), 125.96 (benzofuran 4-C), 126.85 (benzofuran 3a-C), 128.49 (Ph 2,6-C_2), 128.75 (benzofuran 5-C), 131.53 (benzofuran 6-C), 159 (benzofuran 7a-C), 132 (Ph 1-C), 148.28 (benzofuran-CH=), 158 (Ph 4-C), 162.08 (CO_2H), 171.76 (*C*O_2Et) as seen in Figure 2-28.

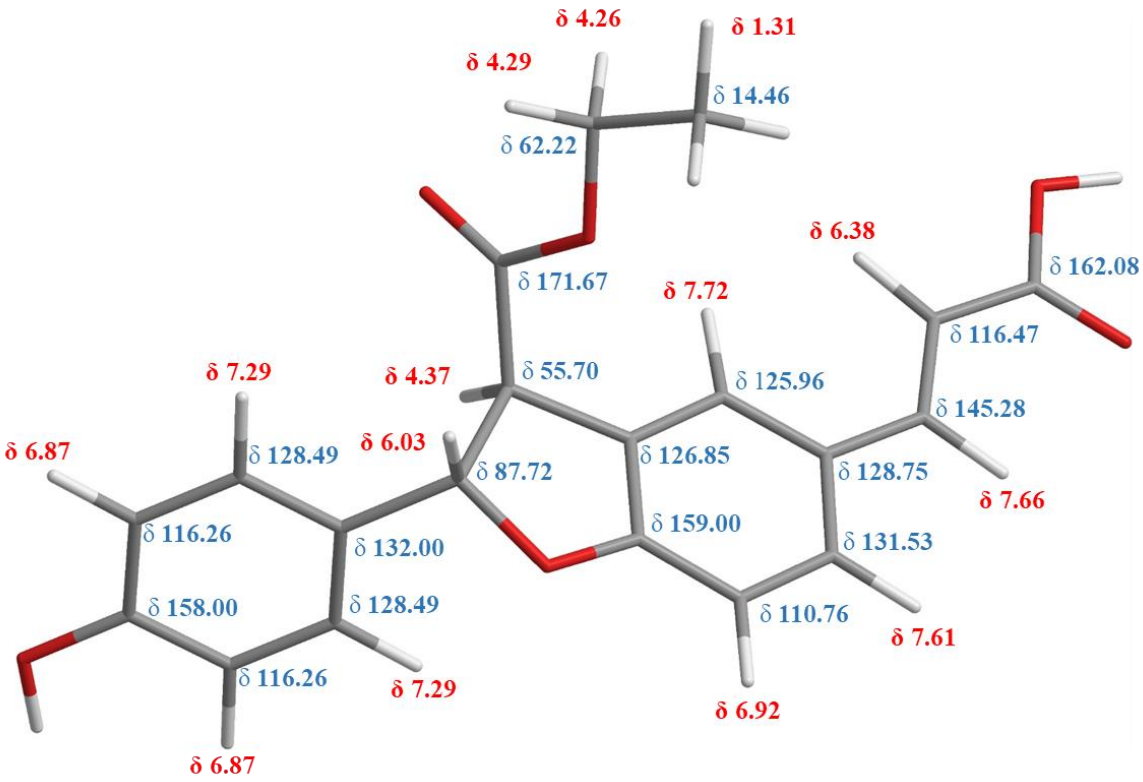


Figure 2-29: Compound **2** structural formula with appropriate $_{1}$H NMR (Red) and $_{13}$C NMR (Blue) shift data.

The $_{1}$H NMR spectrum of **2** showed that it was an 8-5 oxidative dimer of the conventional type, as there was a signal for the *trans*-disubstituted alkene. A combination of $_{1}$H, $_{13}$C, COSY, NOESY, HSQC and HMBC spectra was used to assign the structure.

A signal for the 4-hydroxyphenyl ring was seen in the 1D spectra. The *para*-disubstituted pattern was seen for the left ring in the above structure (Ar). Ar 3-H and 5-H both resonated as a doublet (J = 8.7 Hz) at δ 6.87, consistent with being *ortho* to an electron donating hydroxy group. The corresponding carbons (Ar 3-C and 5-C) were identified by an HSQC cross-peak at δ 116.26; this chemical shift is consistent with its environment. COSY cross-peak δ 7.29 identified Ar 2-H and 6-H. A three-bond HMBC correlation from Ar 1-C δ 132.00 to Ar 3-H δ 6.87 d, J = 8.7 Hz and a three-bond HMBC correlation from Ar 4-C δ 158.00 to Ar 2-H δ 7.29. Ar 2-H and 6-H both resonated as a doublet (J = 8.6 Hz) at δ 7.29, consistent with a phenyl ring. The corresponding carbons (Ar 2-C and Ar 6-C) was identified by an HSQC cross-peak at δ 128.49; this chemical shift is consistent with its environment. Ar 2-H have three-

bond HMBC cross-peak to Ar 4-C (δ 158.00) and to Ar 6-C (δ 128.49). The downfield chemical shift for Ar 4-C showed that it carried an electronegative atom (oxygen of hydroxy).

Benzofuran ring was identified with the benzene ring was trisubstituted with 3 H signals. Benzofuran 4-H resonated as a narrow mulitplet at δ 7.72, this chemical shift is consistent with being *ortho* to a weak electron-withdrawing group. HSQC gave correlation at δ 125.96 benzofuran 4-C. Benzofuran 6-H (δ 7.61) which was a double doublet (J = 8.4, 1.8 Hz), consistent with being *ortho* to a weak electron-withdrawing group. The corresponding carbon (6-C) was identified by an HSQC at cross-peak δ 131.53; this chemical shift is consistent with its environment. Benzofuran 7-H (δ 6.92) which was a doublet (J = 8.4 Hz). The corresponding carbon (7-C) was identified by HSQC at a cross-peak δ 110.76. The downfield chemical shift of 7-C showed that it is consistent with bring *ortho* to an electron donating group (oxygen). The COSY spectrum contained a cross-peak for the benzofuran 4-H (δ 7.72) with benzofuran 6-H (δ 7.61), and 6-H (δ 7.61) with benzofuran 7-H (δ 6.92).

Three-bond HMBC from the benzofuran 4-H to the benzofuran 6-H to C-H carbon signal at δ 145.28, showed that this was the benzofuran 5 substitute. HSQC linked this carbon to a doublet proton signal at δ 7.66 (benzofuran-CH=). The large coupling constant of J = 16.0 Hz indicated the *trans* disubstituted alkene. COSY linked δ 7.66 to δ 6.38 doublet, (J = 15.9 Hz). HSQC from benzofuran 7-H doublet identified C-7 the corresponding cross peak at δ 116.47. All these chemical shifts indicated that this alkene was attached at the distal end to a carbonyl group, and this unit therefore comprising an enone. HMBC from δ 7.66 (benzofuran-CH=) to δ 162.08, this chemical shift is indicative of ester or carboxylic acid. In the benzofuran three-bond HMBC from benzofuran 7-H identified benzofuran 3'-C at δ 126.85. HMBC from benzofuran 4-H identified benzofuran7'-C at δ 159.00.

Benzofuran 4-H was linked by HMBC to aliphatic carbon at δ 55.70 benzofuran 3-C. HSQC showed the benzofuran 3-H at δ 4.37 broad doublet (J = 7.5 Hz). The COSY spectrum contained a cross-peak from the benzofuran 2-H δ 6.03 doublet (J = 7.5 Hz). HSQC showed benzofuran 2-C at δ 87.72. HMBC three bond cross-peak from δ 6.03 to δ 159 gave a very weak signal. This signal is evidence of closure of the dihydrofuran ring. This cross-peak is weak because the (2-H)-(2-C)-(O)-(7a-C) dihedral angle is

close to 90°, as shown in an MM2-minimised model (Figure2-29). Furthermore three-bond HMBC δ 6.03 benzofuran 3-H (*J* = 7.5 Hz) to carbonyl at δ 171.67 (CO_2Et), carbon chemical shift is consistent with ester or carboxylic acid.

^1H spectrum showed an OCH_2CH_3 as δ 1.31 as a triplet (*J* = 7.1 Hz) and two double quartets (*J* = 14.3, 7.2 Hz) at δ 4.26 and δ 4.29. The 14.3 Hz is a geminal coupling. The two different chemical shifts imply that they are in an asymmetrical environment. HSQC identified the ethyl carbon signals at δ 14.46 and δ 62.22 (CH_2). HMBC cross-peaks from both OCH_2 protons to carbonyl signal at δ 171.67, showed that the ester was located at 3-C of the benzofuran ring, thus the carbonyl at the terminus of the alkene must be a carboxylic acid.

2.7.3.3 Diethyl para-coumarate homodimer

Compound **3** *(GH_10):* HPLC – ESI-UV-MS/MS (ESI negative mode) of **3** showed a peak at t_R 24.61 min, *m/z* 381.1138 corresponding to $[M-H]^-$ for $C_{22}H_{22}O_6$ (calcd 382.14164) the pseudomolecular ion of a diethyl dicaffeate compound. m:p: 98.3-100.1 ºC. λ_{max}, 314 nm. 29 mg 1.6 % yield. This sample contains small amounts of impurity therefore potentially making the melting point reading unreliable. In negative ion mode the base peak corresponds to the pseudomolecular ion *m/z* 381 $[M-H]^-$. The following ions were observed in full negative ESI mode *m/z* 762 $[2M-H]^-$, *m/z* 426 $[M-formate]^-$, *m/z* 545 [(2M-H)+coumaric acid]-. MS^2 of *m/z* 543 yielded three fragments *m/z* 524 $[543-H_2O]^-$, *m/z* 499 $[543-CO_2]^-$. *m/z* 451 $[543-2HOCH_2CH_3]^-$. MS^2 of the pseudomolecular ion *m/z* 381 $[M-H]^-$ yielded five fragments *m/z* 353 $[M-CO]^-$ *m/z* 335 $[M-HOCH_2CH_3]^-$, *m/z* 307 $[M-HOCH_2CH_3C=O]^-$, *m/z* 235 $[(M-H)-2HOCH_2CH_3C=O]^-$. In positive ion mode the base peak corresponded to *m/z* 383 $[M+H]^+$ the pseudomolecular ion. MS^2 fragmentation yielded *m/z* 766 $[2M+H]^+$, *m/z* 337 $[(M+H)-HOCH_2CH_3]^+$, *m/z* 291 $[(M+H)-2HOCH_2CH_3]^+$. ^1H NMR and ^{13}C NMR chemical shifts are shown below (Figure 2-30). ^1H NMR ($(CD_3)_2CO$) δ 1.28 (3 H, t, *J* = 6.9 Hz, $=CHCO_2CH_2CH_3$), 1.31 (3 H, t, *J* = 6.9 Hz, benzofuran-$CO_2CH_2CH_3$), 4.19 (2 H, q, *J* = 7.2 Hz, $=CHCO_2CH_2$), 4.26 (1 H, dq, *J* = 14.0, 7.1 Hz, benzofuran-CO_2CHH), 4.28 (1 H, dq, *J* = 14.0, 7.1 Hz, benzofuran-CO_2CHH), 4.37 (1 H, brd, *J* = 7.4 Hz, benzofuran 3-H), 6.03 (1 H, d, *J* = 7.5 Hz, benzofuran 2-H), 6.39 (1 H, d, *J* = 15.9 Hz, $=CHCO_2$), 6.91 (1 H, d, *J* = 8.3 Hz, benzofuran 7-H), 6.87 (2 H, d, *J* = 8.6 Hz, Ph 3,5-H_2), 7.29 (2 H, d, *J* = 8.6 Hz, Ph 2,6-H_2), 7.61 (1 H, dd, *J* = 8.2, 1.8 Hz,

benzofuran 6-H), 7.65 (1 H, d, *J* = 16.0 Hz, benzofuran-CH=), 7.72 (1 H, m, benzofuran 4-H). ^{13}C NMR ($(CD_3)_2CO$) d 14.45 (benzofuran-$CO_2CH_2CH_3$), 14.61 (=$CHCO_2CH_2CH_3$), 55.69 (benzofuran 3-C), 60.56 (=$CHCO_2CH_2$), 62.22 (benzofuran-CO_2CH_2), 87.73 (benzofuran 2-C), 110.76 (benzofuran 7-C), 116.29 (=$CHCO_2$), 116.44 (Ph 3,5-C_2), 125.98 (benzofuran 5-C), 126.85 (benzofuran 3a-C), 128.48 (Ph 1-C), 128.69 (benzofuran 4-C), 131.53 (benzofuran 6-C), 131.89 (Ph 2,6-C_2), 144.91 (CH=$CHCO_2$), 158.5 (Ph4-C), 162.11 (benzofuran 7a-C), 167.23 (=$CHCO_2$), 171.06 (benzofuran-CO_2) as seen in Figure 2-29. HRESIMS *m/z* 381.1337 [M - H]- ($C_{22}H_{21}O_6$ requires 381.1138).

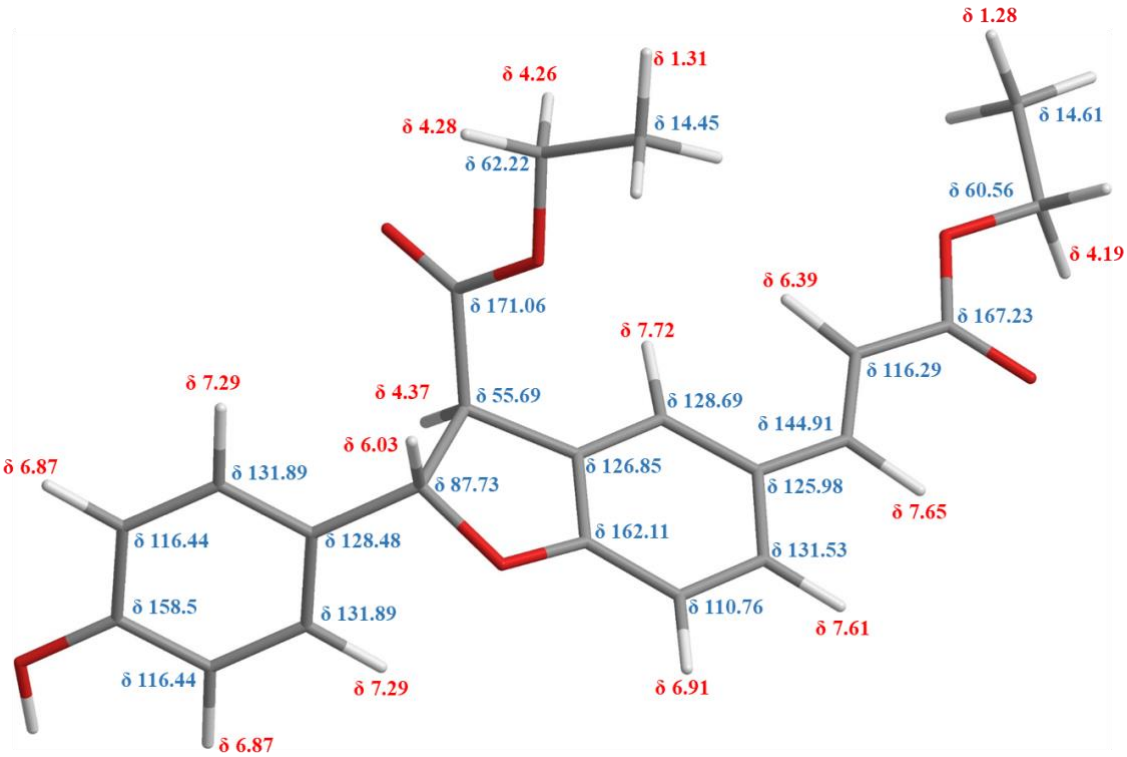


Figure 2-30: Compound **3** structural formula with appropriate ^{1}H NMR (RED) and ^{13}C NMR (Blue) shift data.

The ^{1}H NMR spectrum of **3** showed that it was an 8-5 oxidative dimer of the conventional type, as there was a signal for the *trans*-disubstituted alkene, with the additional ethyl ester replacing the carboxylic acid. A combination of ^{1}H, ^{13}C, COSY, NOESY, HSQC and HMBC spectra was used to assign the structure.

A signal for the 4-hydroxyphenyl ring was seen in the 1D spectra. The *para*-disubstituted pattern was seen for the left ring in the above structure (Ar). Ar 3-H and 5-H both resonated as a doublet (*J* = 8.7 Hz) at δ 6.87, consistent with being *ortho* to an electron donating hydroxy group. The corresponding carbons (Ar 3-C and 5-C)

were identified by an HSQC cross-peak at δ 116.44; this chemical shift is consistent with its environment. COSY cross-peak δ 7.29 identified Ar 2-H and 6-H. A three-bond HMBC correlation from Ar 1-C δ 128.49 to Ar 3-H δ 6.87 d, *J* = 8.7 Hz and a three-bond HMBC correlation from Ar 4-C δ 158.5 to Ar 2-H δ 7.29. Ar 2-H and 6-H both resonated as a doublet (*J* = 8.6 Hz) at δ 7.29, consistent with a phenyl ring. The corresponding carbons (Ar 2-C and Ar 6-C) was identified by an HSQC cross-peak at δ 132.00; this chemical shift is consistent with its environment.

Benzofuran ring was identified with the benzene ring was trisubstituted with 3 H signals. Benzofuran 4-H resonated as a narrow mulitplet at δ 7.72, this chemical shift is consistent with being *ortho* to a weak electron-withdrawing group. HSQC gave correlation at δ 128.96 benzofuran 4-C. Benzofuran 6-H (δ 7.61) which was a double doublet (*J* = 8.4, 1.8 Hz), consistent with being *ortho* to a weak electron-withdrawing group. The corresponding carbon (6-C) was identified by an HSQC at cross-peak δ 131.53; this chemical shift is consistent with its environment. Benzofuran 7-H (δ 6.91) which was a doublet (*J* = 8.4 Hz). The corresponding carbon (7-C) was identified by HSQC at a cross-peak δ 110.76. The downfield chemical shift of 7-C showed that it is consistent with bring *ortho* to an electron donating group (oxygen). The COSY spectrum contained a cross-peak for the benzofuran 4-H (δ 7.72) with benzofuran 6-H (δ 7.61), and 6-H (δ 7.61) with benzofuran 7-H (δ 6.91).

Three-bond HMBC from the benzofuran 4-H to the benzofuran 6-H to C-H carbon signal at δ 144.91, showed that this was the benzofuran 5 substitute. HSQC linked this carbon to a doublet proton signal at δ 7.65 (benzofuran-CH=). The large coupling constant of *J* = 16.0 Hz indicated the *trans* disubstituted alkene. COSY linked δ 7.65 to δ 6.39 doublet, (*J* = 15.9 Hz). HSQC from benzofuran 7-H (δ 6.91) doublet identified C-7 the corresponding cross peak at δ 116.29. All these chemical shifts indicated that this alkene was attached at the distal end to a carbonyl group, and this unit therefore comprising an enone. HMBC from δ 7.65 (benzofuran-CH=) to δ 167.23 (=CH*C*O_2), this chemical shift is indicative of ester or carboxylic acid. The $_1$H spectrum showed a triplet ($OCH_2C\mathit{H}_3$) at δ 1.28 (*J* = 6.9 Hz) and showed a quartet integrating for two protons at δ 4.19 (*J* = 7.2 Hz). The same chemical shift imply that the protons are in a symmetrical environment. The further downfield chemical shifts are consistent with being further away from an electron-withdrawing function. HSQC

identified the ethyl carbon signals at δ 14.61 (CH_3) and δ 62.22 (CH_2). HMBC cross-peak from the OCH_2 protons to the carbonyl signal at δ 167.23, thus the carbonyl at the terminus of the alkene must be the ethyl ester.

In the benzofuran three-bond HMBC from benzofuran 7-H identified benzofuran 3'-C at δ 126.85. HMBC from benzofuran 4-H identified benzofuran 7a-C at δ 162.11. Benzofuran 4-H was linked by HMBC to aliphatic carbon at δ 55.69 benzofuran 3-C. HSQC showed the benzofuran 3-H at δ 4.37 broad doublet (*J* = 7.5 Hz). The COSY spectrum contained a cross-peak from the benzofuran 2-H δ 6.03 doublet (*J* = 7.5 Hz). HSQC showed benzofuran 2-C at δ 87.73. HMBC three bond cross-peak from δ 6.03 to δ 162.11 gave a very weak signal. This signal is evidence of closure of the dihydrofuran ring. This cross-peak is weak because the (2-H)-(2-C)-(O)-(7a-C) dihedral angle is close to 90°, as shown in an MM2-minimised model (Figure2-30). Furthermore three-bond HMBC δ 6.03 benzofuran 3-H (*J* = 7.5 Hz) to carbonyl at δ 171.06 (benzofuran-CO_2), carbon chemical shift is consistent with ester.

^{1}H spectrum showed an OCH_2CH_3 as triplet at δ 1.31 (*J* = 7.1 Hz) and showed two double quartets (*J* = 14.3, 7.2 Hz) at δ 4.26 and δ 4.29. The 14.0 Hz is a geminal coupling. The two different chemical shifts imply that they are in an asymmetrical environment. HSQC identified the ethyl carbon signals at δ 14.45 and δ 62.22 (CH_2). HMBC cross-peaks from both OCH_2 protons to carbonyl signal at δ 171.06, showed that the ester was located at 3-C of the benzofuran ring. Thus completing the structural elucidation of the 8-5 diethyl dicoumarate.

2.8 *Synthesis of ethyl caffeate (ethyl 3,4-dihydroxycinnamate)*

Acetyl chloride (10.0 mL, 128 mmol) was added cautiously to caffeic acid (2.515 g, 15 mmol) in EtOH (250 mL) was slowly stirred for 24 h. The progress was monitored by TLC (CH_2Cl_2 / EtOAc, 1:1). The EtOH was evaporated. The residue, in EtOAc, was washed thrice with saturated brine. Drying (Na_2SO_4) and evaporation gave 3,4-dihydrocinnamte (2.214 g, 71%) as off-white crystals. mp 148.2-152.6 °C 9 (lit (Tawata *et al.*, 1996); R_f = 0.22 (CH_2Cl_2).

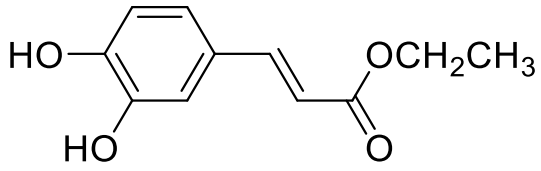


Figure 2-31: Structure of ethyl caffeate

2.8.1 *Synthesis results of ethyl caffeate*

Caffeic acid was converted to ethyl caffeate *via* an acid-catalysed esterification reaction using acetyl chloride. The HPLC-ESI-UV-MS/MS UV chromatogram of ethyl caffeate indicated the presence of a single compound at t_R 16.45 min with a UV chromophore typical of the ester of 324 nm, this absorption is typical of ethyl caffeate (Tan and Shahidi, 2012). In negative ion mode, the base peak at t_R 16.45 min with 100 % intensity, *m/z* 207.0658 $[M-H]^-$ corresponded to the $C_{11}H_{12}O_4$ (calcd: 208.07356) *pseudo*molecular. A second ion observed at *m/z* 414 $[2M-H]^-$ presence of the cluster Secondary MS_2 of the pseudomolecular ion *m/z* 207 $[M-H]^-$ yielded three fragments, the base peak at 100 % intensity *m/z* 178 $[M-HCH_2CH_3]^-$, *m/z* 162 $[M-CO_2]^-$ and *m/z* 134 $[M-CH_2CH_3O-C=O]^-$. A formate adduct was also present as the base peak at *m/z* 236 [(M-H)+COOH]-, formed during the ionisation process. Secondary ionisation MS_2 of the pseudomolecular ion *m/z* 191 $[M-H]^-$ yielded three fragments, a base peak at *m/z* 163 $[M-CH_2CH_3]^-$, typical of the loss of the ethyl chain , further fragments include *m/z* 144 $[M-HOCH_2CH_3]^-$ and *m/z* 118 [M-formate-CH_2CH_3]-. In positive ion mode, a base peak was detected at *m/z* 193 $[M+H]^+$, whilst a less abundant ion at *m/z* 195 $[(M+H)-CH_2CH_3]^+$, corresponding to the McLafferty rearrangement. Further fragmentation of *m/z* 193 $[M+H]^+$ gave fragments in MS_2 ionisation, a base peak at *m/z* 147 $[(M+H)-OHCH_2CH_3]^+$, confirming the loss of an ethanol group and indicative of an ethyl ester moiety. An ion at *m/z* 165 $[(M+H)-CH_2CH_3]^+$, corresponding to the McLafferty rearrangement, indicating the loss of the ethyl chain, also an ion at *m/z* 175 $[(M+H)-H_2O]$, the loss of water.

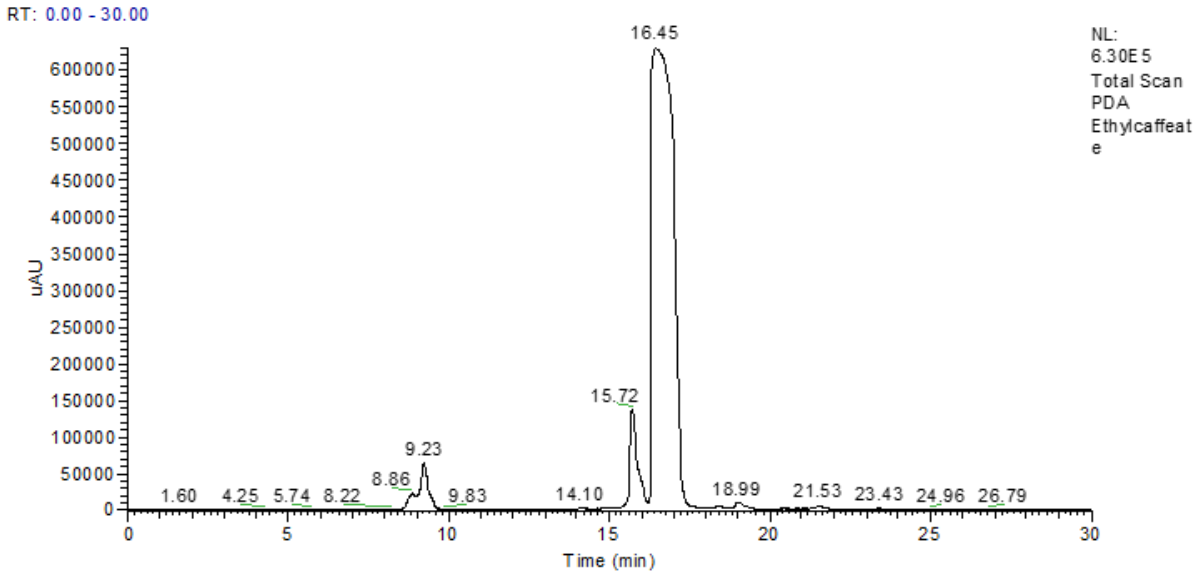


Figure 3-32: HPLC-UV chromatogram of ethyl caffeate t_R 16:45 min.

^{1}H NMR and ^{13}C NMR analysis (Figure 2-31) gave a triplet at δ 1.26 integrating for 3 protons, a quartet integrating for two protons at δ 4.25 (J = 7.2 Hz). The same chemical shift imply that the protons are in a symmetrical environment. The further downfield chemical shifts are consistent with being further away from an electron-withdrawing function. The two alkene protons at 7 and 8 were observed at δ 6.31 and δ 7.48 as doublets with a trans coupling constant of J = 16 Hz. With seven different proton environments and eleven different carbon environments, the synthesised ethyl caffeate is consistent with that found in literature (Dai *et al*., 2006 Zhao *et al*., 2006 He *et al*., 2009; Xiang *et al*., 2011). ^{1}H NMR ($CDCl_3$**)** δ 1.26 (3 H, t, J = 7.1 Hz CH_3,), 4.25 (2 H, q, J = 7.1 Hz, CH_2), 6.29 (1 H, d, J = 15.7 Hz, =C*H*CO_2Et), 6.88 (1 H, dd, J = 8.4, 0.5 Hz, Ar 5-H), 7.28 (1 H, dd, J = 1.9, 0.5 Hz, Ar 2-H), 7.69 (1 H, dd, J = 8.4, 1.9 Hz, Ar 6-H), 7.73 (1 H, d, J = 15.7 Hz, ArC*H*=); ^{13}C NMR ($CDCl_3$) δ 14.30 (CH_3), 60.31 (CH_2O), 109.21 (Ar 2-C), 115.50 (Ar 5-C), 115.63 (=*C*HCO_2), 122.69 (Ar 6-C), 127.51 (Ar 1-C), 144.50 (Ar*C*H=), 147.10 (Ar 3-C), 147.52 (Ar 4-C), 168.10 (C=O) as seen in Figure 2-33.

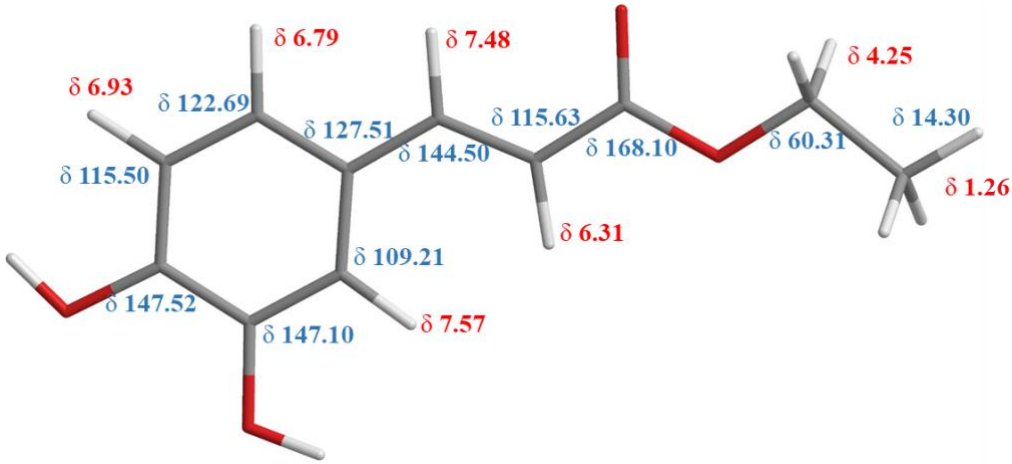


Figure 2-33: Ethyl caffeate structural formula with appropriate $_{1}$H NMR (Red) and $_{13}$C NMR (Blue) shift data.

2.9 *Oxidative heterodimerisation of caffeic acid and ethyl caffeate.*

Caffeic acid (1.22 g, 5.9 mmol) and ethyl caffeate (1.05 g, 5.9 mmol) was added acetone (100 mL) gently heated until dissolved. Purified H_2O (440 mL) was added cautiously, followed by urea hydrogen peroxide (617 mg g, 6.56 mmol) in purified H_2O (30 mL). The pH was monitored and maintained at pH 4 (Schomberg, Salzmann, and Stephan, 1993), followed by the addition of horseradish peroxidase type 2 (5 mg) in H_2O (10 mL). The pink coloured mixture was stirred at ambient temperature for 1.5 h. Indicating the enzyme was still chemically active. Further addition of urea H_2O_2 (316 mg, 3.28 mmol) in H_2O (3 mL) was needed. The mixture was mixed for another 1h, after which the compound was obtained into EtOAc. The residue, in EtOAc was washed thrice with saturated brine. Drying (Na_2SO_4) and evaporation gave a crude dark red crude oil (2.08 g). The HPLC-UV chromatogram of the crude synthesis is shown in Figure 2- 34.

2.9.1 Results of heterodimerisation of caffeic acid and ethyl caffeate.

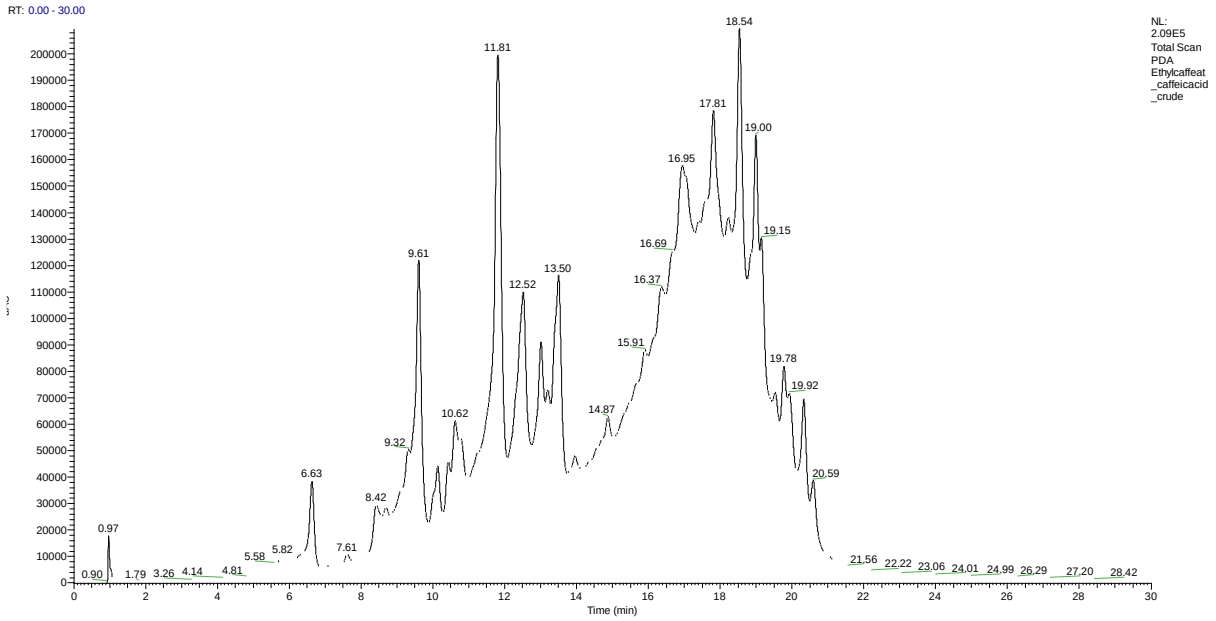


Figure 2-34: HPLC chromatogram of the ethyl caffeate and caffeic acid crude.

2.9.2 *Solid phase extraction and semi-preparative method development*

Solid-phase extraction was used to fractionate the crude sample. The crude material (1.0407 g) was bound on to reverse-phase silica and loaded onto a pre-equilibrated reverse phase cartridge. The fractions were then eluted with increasing gradient of MeOH (0 % - 60 % -100 %, 200 mL per wash). Fractionated peaks were evaporated, freeze-dried and weighed (Table 2-2). All samples where then subject to further HPLC-UV-ESI-MS/MS, to identify any further purification or NMR analysis if the yields allowed HPLC can be found within the supporting information (S9-II).

Table 2-2: Solid phase extraction of oxidatively coupled of ethyl caffeate and caffeic acid.

Fraction / Sample name DiC…	*MeOH %*	*Mass (mg)*
1	0	14.8
2	10	33
3	20	226.8
4	30	227.7
5	40	374.1
6	50	35
7	60	45.8
8	100	34.5

HPLC-UV-ESI-MS/MS identified two fractions where further separation was possible, as these fractions contained the compounds of interest with ideal chromatograms for semi-preparative purification. Method development began on DiC003 and DiC004 with the F14 C8 column, ACN 30-50 % and ACN 40-60% over 10 min, respectively, thus providing sufficient separation with the HICHROM ACE 5 C_{18} column. The method was loaded onto the semi-preparative system and eluted to give the following fractions (Table 2-3). Compound 4 – 9 supporting data can be found within the appendix S9-S15.

Table 2-3: Yields from semi-preparative purification of samples DiC003 and DiC004.

Sample Name DiC00 3…	*Mass (mg)*	*Sample name DiC00 4…*	*Mass (mg)*
A	5.7	***A***	10.7
B	4.3	***B***	44.1
C	22.7	***C***	9.1
D	14.7	***D***	28.0
E	2.9		
F	45.9		
G	12.9		
H	3.0		

2.9.3 Ethyl dicaffeate and caffeic acid heterodimers

*Compound **4** – (DiC003_D):* The HPLC-MS/MS (ESI negative mode) of **4** showed a peak at t_R 17.22 min, with an *m/z* 385.0923 corresponding to [M-H]- the pseudomolecular ion for $C_{20}H_{18}O_8$ (calcd 386.10017), 14.7 mg , 0.6 % yield. The UV spectrum λ_{max} 324.00 nm, this UV absorption if typical of hydroxycinnamic acid conjugates as described by Parveen *et al*., (2008). In negative mode MS_2 four fragment ions were observed; *m/z* 339 [M-$HOCH_2CH_3$]-, loss of EtOH, confirming the presence of an ethyl ester group; *m/z* 341 [M-CO_2]-, corresponding to the loss of CO_2; *m/z* 771 integrating for [2M-H]- *m/z* 367 [M-H_2O]- (loss of H_2O, indicating the presence of more than on hydroxy group). Furthermore, MS_2 fragment ions in positive mode were observed with *m/z* 341 [M-$HOCH_2CH_3$]- loss of EtOH, suggesting the presence of ethyl ester, *m/z* 206 [M-caffeic acid]- (loss of caffeic acid) *m/z* 250 [M-ethyl caffeate]- loss of ethyl caffeate, supporting that **4** is a heterodimer of ethyl caffeate and caffeic acid.

*Compound **5** – (DiC003_E):* The HPLC-MS/MS (ESI negative mode) of **5** showed a peal at t_R 14.97 min, with an *m/z* 385.0923 corresponding to [M-H]- the pseudomolecular ion for $C_{20}H_{18}O_8$ (calcd: 386.10017) (2.9 mg, 0.13 %). In full negative mode another ion was observed at *m/*z 771.1926, corresponding to [2M-H]- . In MS_2 fragmentation two ions were observed *m/z* 341 [M-COOH]- corresponding to the loss of carboxylic acid and *m/z* 295 [M-$HOCH_2CH_3$.COOH]-, indicating the loss the ethyl ester with the carboxylic acid. In positive ion mode the pseudomolecular ion does not ionise easily, a small peak at *m/z* 387.07 [M-H]- was observed, were as in MS_2 gave a molecular ion at *m/z* 342 [M-COOH]- corresponding to the loss of the formate ion. λ_{max} 324 nm this UV absorption is typical of hydroxycinnamic acid conjugates as described by Parveen *et al*. (2008). Only 2.9 mg was yielded, NMR analysis was not completed.

*Compound **6** - (DiC003_F):* The HPLC-MS/MS (ESI negative mode) of **6** showed a peak at t_R 15.03 min, with an *m/z* 385.0930, corresponding to [M-H]- the pseudomolecular ion for $C_{20}H_{18}O_8$ (calcd: 386.10017) (45.9 mg, yield 2 %). λ_{max} 343 nm this UV absorption is typical of hydroxycinnamic acid conjugates as described by Parveen *et al*. (2008). In negative ion mode the base peak ion was observed at *m/z* 768.00 [M-H]-, corresponding to the cluster of the pseudomolecular ion. Further ions

were observed at *m/z* 341 $[M\text{-}CO_2]^-$ and *m/z* 341 $[M\text{-}HOCH_2CH_3.CO_2]^-$. MS_2 of the pseudomolecular ion *m/z* 384 $[M\text{-}H]^-$, yielded a further significant fragment *m/z* 275 $[M\text{-}C_6H_3.2OH]^-$, corresponding to the benzene ring with two hydroxyl groups attached. *m/z* 307 $[M\text{-}HOCH_2CH_3]^-$, *m/z* 263 $[M\text{-}HOCH_2CH_3.CO_2]^-$. In positive ion mode ion a single peak at *m/z* 387 $[M+H]^+$ corresponds to the pseudomolecular ion. In positive ion mode MS_2, the following ions were observed, *m/z* 369 $[M\text{-}H_2O]^+$ consistent with the loss water, *m/z* 342 $[M\text{-}HOCH_2CH_3]^+$ conforming to the loss of EtOH from the ethyl ester, *m/z* 314 $[M+HCH_2CH_3.COOH]^+$ corresponding to the loss of EtOH and the carboxylic acid. Poor solubility precluded NMR analysis.

2.9.4 Diethyl dicaffeate homodimers

*Compound **7** - (DiC003_B):* The HPLC-MS/MS (ESI negative mode) of **7** showed a peak at t_R 17.06 min, with an *m/z* of 413.1241 corresponding to $[M\text{-}H]^-$ the pseudomolecular ion for $C_{22}H_{22}O_8$ diethyl dicaffeate (calcd: 414.13147) (5.7 mg, 0.2 % yield). The UV spectrum λ_{max} 341 nm this UV absorption is typical of hydroxycinnamic acid conjugates as described by Parveen *et al*. (2008). In negative ion mode the following ions were observed, a base peak at *m/z* 826 $[2M\text{-}H]^-$, corresponding to the pseudomolecular dimer, *m/z* 459 $[(M\text{-}H)+HOCH_2CH_3]^-$. In negative ion mode MS_2 yielded five significant fragments, a base peak at *m/z* 303 $[M\text{-}C_6H_3.2OH]^-$, *m/z* 367 $[M\text{-}HOCH_2CH_3]^-$ indicative of the ethyl ester, *m/z* 321 $[M\text{-}2HOCH_2CH_3]^-$, *m/z* 233 [M-caffeic acid]-, *m/z* 180 [M-ethyl caffeate]-. In positive ion mode, the pseudomolecular ion corresponds to *m/z* 415. In MS_2 of *m/z* 415 $[M+H]^+$ gave two fragments *m/z* 369 $[(M+H)\text{-}HOCH_2CH_3]^+$ and *m/z* 341 $[(M+H)\text{-}C_6H_3]^+$. Melting point was not possible due to another peak present in the PDA chromatogram t_R 19.34 min, *m/z* 411 $[M\text{-}H]^-$ pseudomolecular ion with a base peak *m/z* 822 $[2M\text{-}H]^-$ corresponding to the dimer of the pseudomolecular ion. The UV spectrum λ_{max} (nm) 266.00 and 311.00. Due to the low yield the sample is not suitable for further purification and thus not suitable for NMR analysis.

*Compound **8** - (DiC004_A):* The HPLC-MS/MS (ESI negative mode) of **8** showed a peak at t_R 17.72 min, *m/z* 385.0929 corresponding to $[M\text{-}H]^-$ the pseudomolecular ion for $C_{20}H_{18}O_8$ ethyl dicaffeate (calcd: 386.10017) (10.7 mg, 0.47 % yield). m.p.169.5 - 170.1°C. λ_{max} 321 nm this UV absorption is typical of hydroxycinnamic acid conjugates as described by Parveen *et al*. (2008). In negative ion mode the base peak

at *m/z* 770 [2M-H]- corresponds to the dimer of the pseudomolecular ion at *m/z* 385 [M-H]-, *m/z* 430 [(M-H)+COOH]-. Analysis of negative ion mode MS_2 yielded three fragments, *m/z* 341 [M-CO_2]-, *m/z* 294 [M-$HOCH_2CH_3.CO_2$]-, *m/z* 275 [M-H-C_6H_3.2OH]- corresponding to the loss of the ring A. In positive ion mode the pseudomolecular ion was observed at *m/z* 323 [(M+H)-$HOCH_2CH_3.H_2O$]+. Further fragmentation positive MS_2 only yielded one peak at 100 % intensity, *m/z* 341 [(M+H)-$HOCH_2CH_3$]+, integrating for the ethyl ester present. Poor solubility precluded NMR analysis

Compound ***9*** - *(DiC004_C):* The HPLC-MS/MS (ESI negative mode) of **9** showed two peaks at t_R 15.67 min and 17.73, both having an *m/z* 385.0929 corresponding to [M-H]- the pseudomolecular ion for $C_{20}H_{18}O_8$ ethyl dicaffeate (calcd: 386.10017) (44.1 mg, yield 2 %) . No melting point to due the sample containing two peaks. λ_{max} 338 nm for peak at t_R 15.67 min and λ_{max} 338 nm for peak at t_R 17.73 min, this peak corresponds to sample contamination between compounds **8** and **9.** These UV absorptions are typical of hydroxycinnamic acid conjugates as described by Parveen *et al*. (2008). HPLC-MS/MS analysis of **9** in negative mode, a base peak of *m/z* 770 [2M-H]-, corresponding to the dimer of the pseudomolecular ion. Further two ions were overserved at *m/z* 430 [(M-H)+COOH]-, the addition of a formation ion, and at *m/z* 1156 [(2M-H)+(2M-H)]-, corresponding to the another addition of the pseudomolecular ion to the dimer. Negative ion mode MS_2 yielded three fragments, *m/z* 339 [M-$HOCH_2CH_3$]-, *m/z* 275 [M-H-C_6H_3.2OH]-, with the base peak at *m/z* 229 [M-H-C_6H_3.2OH-COOH]-. In positive ion mode, the pseudomolecular ion was observed at *m/z* 387 $[M+H]^+$, with a base peak at *m/z* 341 [(M+H)-$HOCH_2CH_3$]+. Further purification is required for reliable NMR analysis.

2.10 *Synthesis of ethyl ferulate*

Acetyl chloride (10.0 mL, 128 mmol) was added cautiously to ferulic acid (5.030 g, 30 mmol) in EtOH (250 mL) was slowly stirred for 24 h. The progress was monitored by TLC (CH_2Cl_2 / EtOAc, 1:1). The EtOH was evaporated. The residue, in EtOAc, was washed thrice with saturated brine. Drying (Na_2SO_4) and evaporation To give ethyl 4-hydroxy-3-methoxycinnamate (4.317 g, 65%) as white crystals: mp 62-65ºC (lit. (Voisin-Chiret *et al*., 2007) mp 67°C); R_f = 0.21 (petroleum ether / EtOAc 9:1). HPLC t_R 18.99 min, λ_{max} 242 and 322 nm.

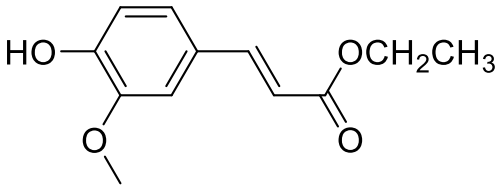


Figure 2-35: Structure of ethyl ferulate.

2.10.1 *Synthesis results of ethyl ferulate*

Ethyl ferulate showed a peak at t_R of 18.99 min, the λ_{max} 322 nm, this is characteristic of an esterified product; additionally it is constituent within literature (Ralph et al., 1998). In negative mode MS, a peak at *m/z* 221.0823 corresponding to $[M-H]^-$ for $C_{12}H_{14}O_4$ (calcd 221.08921) the pseudomolecular ion. A formate adduct was also present at *m/z* 267 $[(M-H)+HCOOH]^-$, forming during the ionisation process. MS^2 of the pseudomolecular ion *m/z* 221 $[M-H]^-$ gave two fragments, *m/z* 193 [Ferulate]-, corresponding to the ferulate unit and a base peak at *m/z* 206 $[M-H, CH_3]^-$ corresponding to the loss of a methyl moiety. Further fragmentation ion at *m/z* 223 gave an individual MS^2 ion at *m/z* 177 $[M-HOCH_2CH_3]^+$, indicating the loss of an ethanol group, indicative of an ethyl ester moiety. In positive mode, the base peak was identified at *m/z* 223 $[M+H]^+$, whilst a reduced abundant ion at *m/z* 445 $[2M+H]^+$ characterised as the dimer formation. The fragmentation patterns are in accordance with those previously published (Bunzel *et al*., 2008; Aliawish *et al*., 2014).

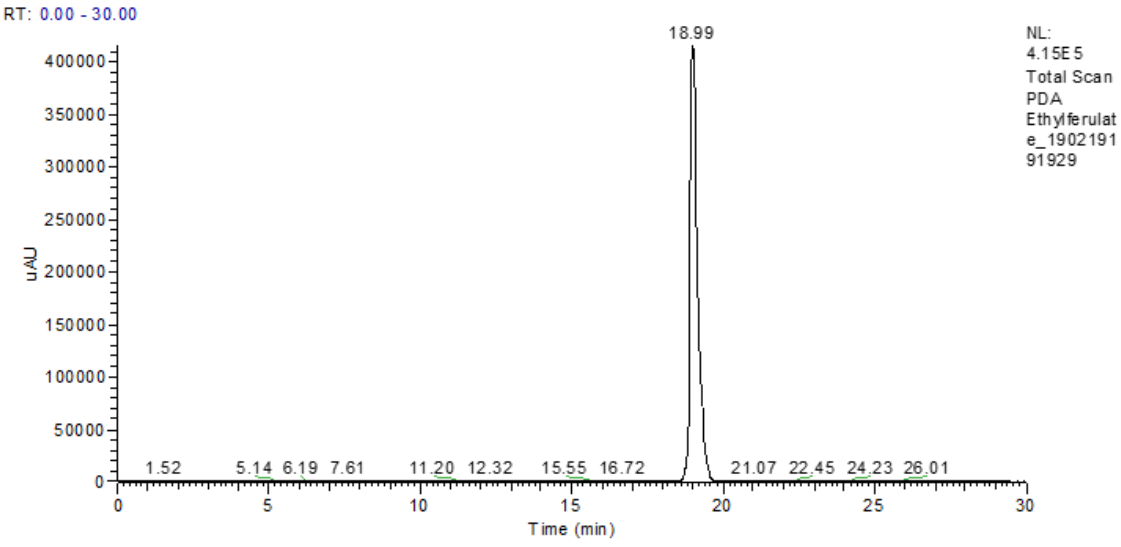


Figure 2-36: HPLC-UV of ethyl ferulate, t_R 18.99 min.

Analysis of ^{1}H NMR and ^{13}C NMR chemical shift date shown below (Figure 2-33) of ethyl ferulate is gave a triplet at δ 1.33 integrating for 3 protons, gave a doublet at δ 4.20 the alkene protons at 7 and 8 were observed at δ 6.32 and δ 7.48 as doublets with a trans coupling constant of J = 16 Hz. Furthermore a chemical shift was observed at δ 3.90 as a singlet integrating for 3 protons. With eight different proton environments and twelve different carbon environments, the synthesised ethyl ferulate is consistent with that found in literature (Li *et al*., 2009; Kumar, 2011; Voisin-Chiret *et al*., 2007).

^{1}H NMR ($CDCl_3$) δ 1.33 (3 H, t, J = 7.1 Hz, CH_2CH_3), 3.90 (3 H, s, OCH_3), 4.20 (2 H, q, J = 7.1 Hz, CH_2CH_3), 6.23 (1 H, d, J = 15.7 Hz, =$CHCO_2H$), 6.79 (1H, dd, J = 8.4, 0.4 Hz, Ar 5-H), 7.00 (1 H, dd, J = 1.9, 0.4 Hz, Ar 2-H), 4.16 (1 H, dd, J = 8.4, 1.9 Hz, Ar 6-H), 7.58 (1 H, d, J = 15.7 Hz, ArCH =); ^{13}C NMR ($CDCl_3$) δ 14.30 (OCH_2CH_3), 55.87 (OCH_3), 60.31 (CH_2), 109.21 (Ar 2-C), 114.63 (=$CHCO_2$Et),115.60 (Ar 5-C), 122.97 (Ar 6-C), 126.98 (Ar 1-C), 144.60 (ArCH=), 146.69 (Ar 3-C), 147.82 (Ar 4-C), 167.23 (C=O) data shown in Figure 2-37. HRESIMS m/z 221.0823 [M - H]- ($C_{12}H_{13}O_4$ requires 221.0814).

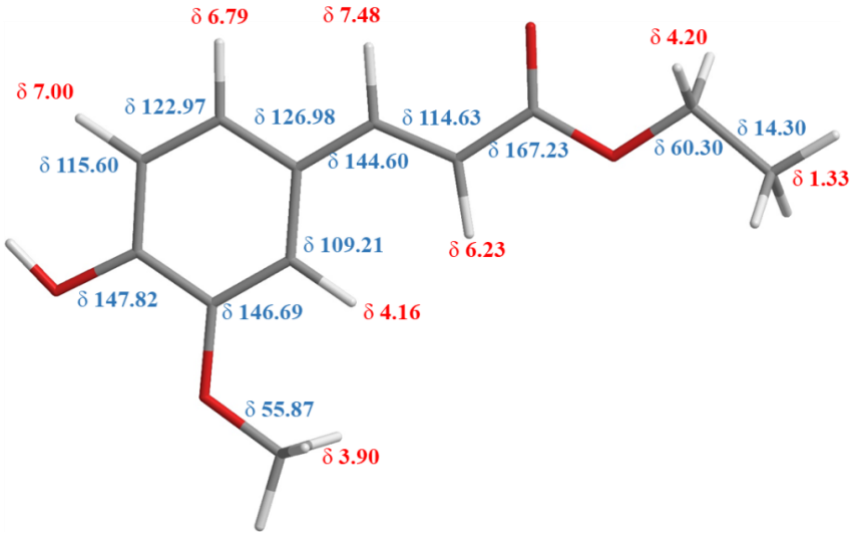


Figure 2-37: Ethyl ferulate structural formula with appropriate ^{1}H NMR (Red) and ^{13}C NMR (Blue) shift data.

2.11 *Oxidative heterodimerisation of ferulic acid and ethyl ferulate (I)*

Ferulic acid (1.21 g, 5.9 mmol) and ethyl ferulate (1.32 g, 5.9 mmol) was added acetone (100 mL) gently heated until dissolved. Purified H_2O (825 mL) was added cautiously, followed by urea hydrogen peroxide (617 mg g, 6.6 mmol) in purified H_2O

(15 mL). The pH was monitored and maintained at pH 4 (Schomberg, Salzmann, and Stephan, 1993), followed by the addition of horseradish peroxidase type 2 (5 mg) in H_2O (10 mL). The pink coloured mixture was stirred at ambient temperature for 1.5 h. Indicating the enzyme was still chemically active. Further addition of urea H_2O_2 (316 mg, 3.28 mmol) in H_2O (3 mL) was needed. The mixture was mixed for another 1h, after which the compound was obtained into EtOAc. The residue, in EtOAc was washed thrice with saturated brine. Drying (Na_2SO_4) and evaporation gave a crude dark orange/brown crude oil (2.08 g, 82 % recovery yield).

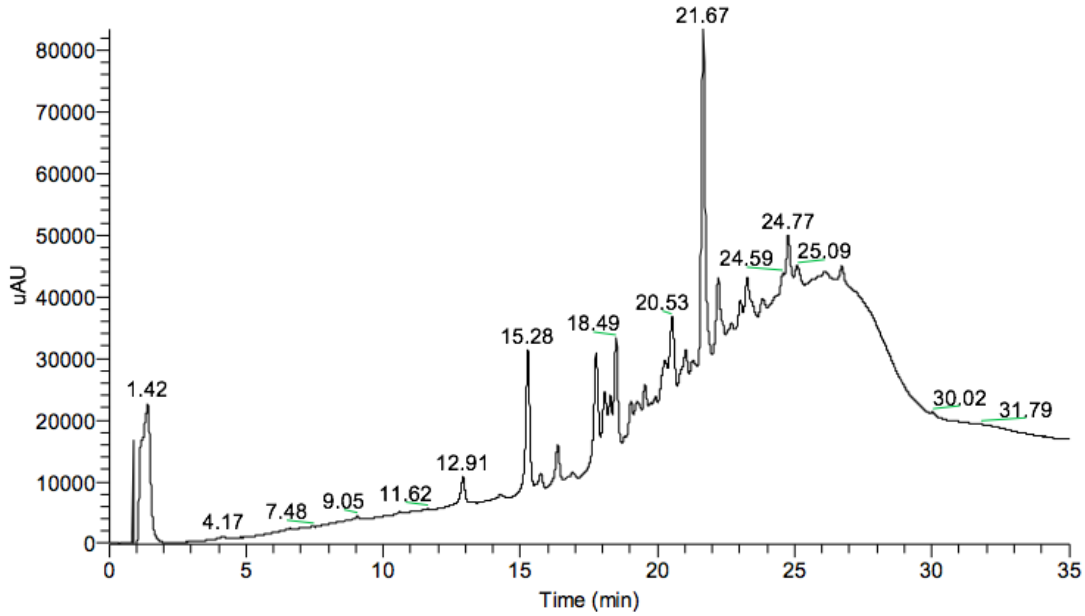


Figure 2-38: The PDA chromatogram of ethyl ferulate and ferulic acid crude (Reaction **I**).

Analysis of the crude by HPLC-ESI-UV-MS/MS (Figure 2-36) indicated there were multiple peaks which had the three desired molecular weights for the pseudomolecular ion at, *m/z* 384 corresponding to [M-H]- for $C_{20}H_{16}O_8$ for diferulate, observed at t_R 12.91. Ions *m/z* 413 corresponding to [M-H]- for $C_{22}H_{21}O_8$ for ethyl diferulate, at t_R 15.28 and 18.49 min. Ion ay *m/z* 442 corresponding to [M-H]- for $C_{24}H_{26}O_8$ for diethyl diferulate at t_R 21.67 min.

2.11.1 *Isolera Biotage Flash Chromatography and semi-preparative method development*

Isolera Biotage Flash Chromatography was performed using a pre-packed FLASH SNAP C_{18} 60 g reverse phase cartridges. The crude extract (0.5069 g) was bound onto C_{18} silica. The material was then loaded onto a pre washed and equilibrated cartridge (100% H_2O). Mobile phase consisted of 100 % H_2O to 100 % CH_3CN over 65 min, at

a flow rate of 40 mL min^{-1}. A photodiode array (PDA) monitored absorbance at 280 and 340 nm. Fractionated peaks were evaporated, freeze-dried and weighed, total of 374.4 mg, 75 % of material eluated, Table 2-4 below. All samples where then subject to further HPLC-UV-ESI-MS/MS, to identify any further purification or NMR analysis.

After HPLC analysis of the fractions Table 2-4. Fractions S2 and S5 where moved onto semi preparative method development using an F14 C3 column. Mobile phase initiated with 60 % aq. CH_3CN to 80 % aq. CH_3CN over 13 min. This method was then installed on the semi-preparative system using a HICHROM ACE 5 C_{18} column. Fraction S2 yielded two compounds (compounds **10** and **11**), fraction S5 yielded one compound (compound **12**).

Table 2-4: Isolera Biotage Flash Chromatography fractions of oxidatively coupled ethyl ferulate and ferulic acid.

Sample name: Biotage	***Mass (mg)***
S1	13.2
S2	81.1
S3	75.6
S4	40.5
S5	73.4
S6	39.1
S7	51.5

2.11.2 Ethyl diferulates

Compound **10** *(S2)***:** Ethyl (±)-4,5-*trans*-5-(4-hydroxy-4-methoxyphenyl)-3-(*E*-4-hydroxy-3-methoxyphenylmethylene)-2-oxotetrahydrofuran-4-carboxylate**.** The HPLC-MS/MS (ESI negative mode) of **10** showed a peak at t_R 15.42 min, *m/z* 413.1234 corresponding to $[M-H]^-$ for $C_{22}H_{22}O_8$ (calcd: 413.123645). This suggests a heterodimer. Yield 10.2 mg, (yield 0.4 %) λ_{max} 234, 338 nm, typical of hydrocinnamaic acid conjugates (Parveen *et al*., 2008) Orange oil. (10.2 mg, 0.4 % yield). 10.2 mg 0.4 % yield oil. Another ion was also evident, this was the formate adduct, *m/z* 458 $[(M-H)+COOH]^-$. MS_2 produced further fragments including *m/z* 353 $[M-H-COOH,CH_3]^-$, *m/z* 367 $[M-HOCH_2CH_3]^-$ and *m/z* 369 $[M-CO_2]^-$ at 100% intensity. In positive ion

mode MS, the peak at *m/z* 415 $[M+H]^+$ corresponded to the pseudomolecular ion. Other were also evident, these were a sodium adduct; *m/z* 437 $[M+Na]^+$, *m/z* 430 $[M+CH_3]^+$ and the dimer of the pseudomolecular ion *m/z* 830 $[2M+H]^+$. MS_2, gave three significant fragments; *m/z* 369 $[M-HOCH_2CH_3]^+$, *m/z* 397 $[M-H_2O]^+$ and *m/z* 324 $[M-HOCH_2CH_3.COOH]^+$.

The UV spectrum of **10** showed a strong absorption at 242.00 and 338.00 nm which reduced to a shoulder within the range of 360 - 400 nm. Compound **10** with appropriate ^1H NMR and ^{13}C NMR chemical shift data (Figure 2-35). ^1H NMR ($CDCl_3$) δ 1.14 (3 H, t, *J* = 6.8 Hz, $CH_2C\mathit{H}_3$), 3.65 (3 H, s, $ArOCH_3$), 3.76 (3 H, s, $Ar'OCH_3$), 4.10 (1 H, dd, *J* = 3.2, 2.0 Hz, furan 4-H), 4.14 (1 H, dq, *J* = 10.8, 7.1 Hz, $OC\mathit{H}HMe$), 4.16 (1 H, dq, *J* = 10.8, 7.1 Hz, $OCH\mathit{H}Me$), 5.88 (1 H, d, *J* = 3.2 Hz, furan 5-H), 6.75 (1 H, d, *J* = 0.8 Hz, Ar 2-H), 6.77 (1 H, dd, *J* = 8.0, 0.8 Hz, Ar 6-H), 6.86 (1 H, d, *J* = 8.0 Hz, Ar 5-H), 6.89 (1 H, d, *J* = 8.0 Hz, Ar' 5-H), 7.03 (1 H, d, *J* = 0.8 Hz, Ar' 2-H), 7.04 (1 H, dd, *J* = 8.0, 0.8 Hz, Ar' 6-H), 7.65 1 H, d, *J* = 2.0 Hz, =CH); ^{13}C NMR ($CDCl_3$) δ 13.94 ($CH_2\mathit{C}H_3$), 62.02 ($\mathit{C}H_2CH_3$), 53.78 (4-C), 56.00 ($ArO\mathit{C}H_3$, $Ar'O\mathit{C}H_3$), 80.25 (furan 5-C), 107 (furan 3-C), 107.54 (Ar 2-C), 112.09 (Ar' 2-C), 114.70 (Ar 5-C), 114.78 (Ar' 5-C), 118.27 (Ar 6-C), 125.49 (Ar' 6-C), 125.84 (Ar' 1-C), 131.27 (Ar 1-C), 140.87 (=CH), 146.64 (Ar 3-C), 146.82 (Ar' 3,4-C_2), 148.24 (Ar' 4-H), 170.24 ($\mathit{C}O_2Et$), 171.18 (furan 2-C). Data imaged below Figure 2-39. HRESIMS *m/z* 413.1234 $[M-H]^-$ ($C_{22}H_{21}O_8$ requires 413.1236).

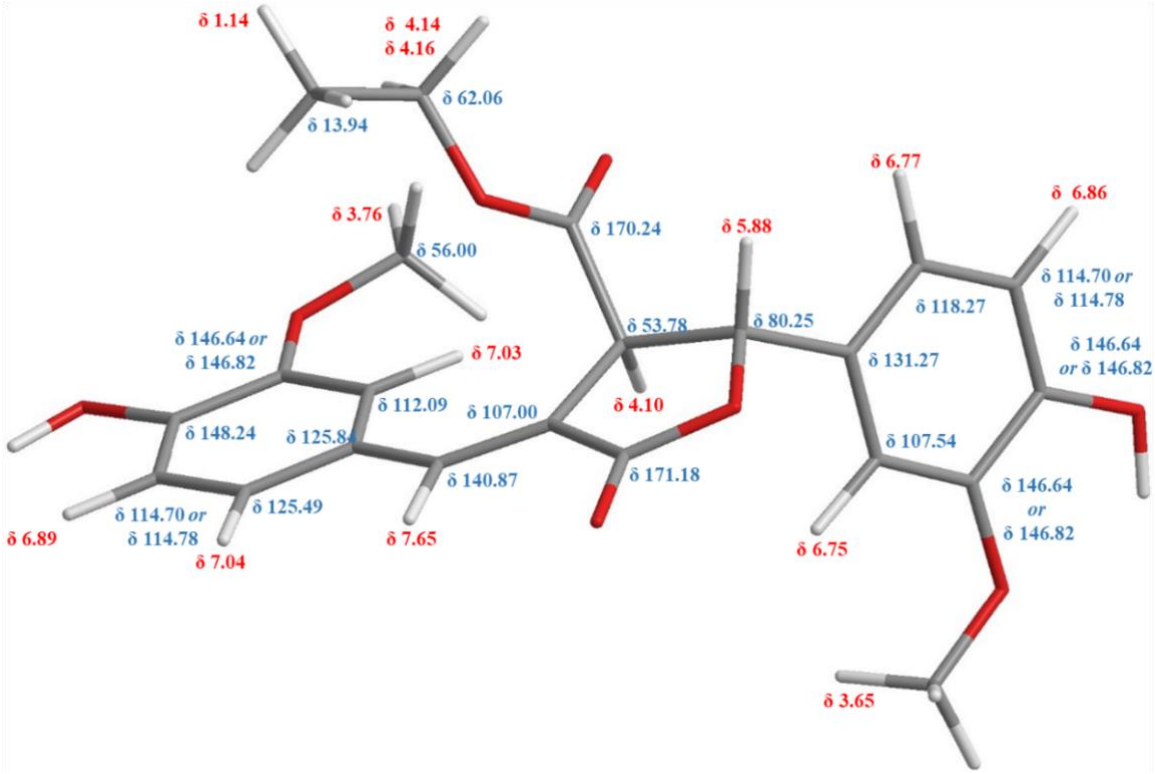


Figure 2-39: Compound **10** – Ethyl diferulate structural formula with appropriate $_{1}$H NMR (Red) and $_{13}$C NMR (Blue) shift data.

The $_{1}$H NMR spectrum of **10** showed that it was not a 8-5 oxidative dimer of the conventional type, in that there were n signals for a *trans*-disubstituted alkene. Due to the unique structure complete analysis using a combination of $_{1}$H, $_{13}$C, COSY, NOESY, HSQC and HMBC spectra in full detail was used to assign the structure.

Signals for two 4-hydroxy-3-methoxyphenyl rings were seen in the 1D spectrum. The 1,3,4-trisubstituted pattern was seen for the left ring in the above structure (Ar). Ar 2-H resonated as a narrow doublet (J = 0.8 Hz) at δ 7.03, consistent with being *ortho* to both an electron-donating group (hydroxy or methoxy) and a weakly electron-withdrawing function. The corresponding carbon (Ar 2-C) was identified by an HSQC cross-peak at δ 112.09; this chemical shift is consistent with its environment. Ar 2-H gave three-bond HMBC cross-peaks to Ar 4-C (δ 148.24) and to Ar 6-C (δ 125.49). The downfield chemical shift of Ar 4-C showed that it carried an electronegative atom (oxygen of hydroxy or methoxy). HSQC from Ar 6-C identified Ar 6-H (δ 7.04) which was a double doublet (J = 8.0, 0.8 Hz). The COSY spectrum contained a cross-peak from Ar 6-H to a doublet (J = 8.0 Hz) at d 6.89, which was therefore assigned to Ar H-5. The coupling constant between Ar 5-H and Ar 6-H is consistent with these two protons being *ortho* to each other ($_{3}J$) and the small coupling constant to Ar 2-H is

consistent with *meta* coupling (^4J). A three-bond HMBC correlation from Ar 5-H then identified Ar 1-C at δ 125.84 and Ar 3-C at δ 146.64 or δ 146.82. The downfield chemical shift of the latter showed attachment of an electronegative oxygen atom. A methoxy group (δ_H 3.76, δ_C 56.00) was shown to be attached at Ar 3-C by an HMBC cross-peak from this methoxy proton signal and Ar 3-C. Therefore, Ar 4-C carries a hydroxy group, consistent with the chemical shift.

The other aromatic ring showed similar NMR signals, suggesting a similar substitution pattern. Ar' 2-H resonated as a narrow doublet (J = 0.8 Hz) at δ 6.75, showing that it is ortho to an electron-donating group and to an electron-neutral group. It is noteworthy that this signal is markedly up field from the corresponding Ar 2-H signal (d 7.03) of the other benzene ring. Ar' 2-H is *meta*-coupled to Ar' 6-H (δ 6.77), which is also up field from the corresponding signal from Ar 6-H (δ 7.04). Thus this benzene ring Ar' carried an electron-neutral group at 1-C, in contrast to the weakly electron-withdrawing group attached at 1-C of the left benzene ring Ar. Ar' 2-C was identified at δ 107.54 and Ar' 6-C was identified at δ 118.25 by HSQC cross-peaks from Ar' 2-H and Ar' 6-H, respectively. A COSY cross-peak from Ar' 6-H revealed the Ar' 5-H signal as a doublet (J = 8.0 Hz) at δ 6.86. This peak showed an HSQC correlation with Ar' 5-C (δ 114.70 or δ 114.78). HMBC then identified the signals for the quaternary carbons (Ar' 1-C, δ 131.27; Ar' 3-C, δ 146.64 or δ 146.82; Ar' 4-C, δ 146.64 or δ 146.82). The methoxy group attached to 3-C of this ring resonated at δ_H 3.65 and δ_C 56.00.

The core of the structure was identified as follows. Ar 2-H and Ar 6-H both gave three-bond HMBC cross-peaks to a ^{13}C signal at δ 140.87. This chemical shift is consistent with the downfield shift for the β-carbon of an enone. HSQC linked this ^{13}C signal to a narrow doublet (J = 2.0 Hz) in the ^1H spectrum at δ 7.65, also consistent with being the β-H of an enone. The small coupling constant could have been benzylic coupling to Ar 2-H and Ar 6-H but the coupling constants and the multiplicities of the latter did not match and this was discounted. The signal at δ 7.65 is therefore ^4J allylically coupled, on the basis of the small coupling constant, and has no two-bond or three-bond couplings. A COSY cross-peak linked this signal to the double doublet (J = 3.2, 2.0 Hz) at δ 4.10. This proton was joined to the carbon at δ 53.78, as shown by HSQC. This ^1H signal also showed a COSY cross-peak to the signal at δ 5.58 (d, J = 3.2 Hz)

and HSQC identified the corresponding carbon signal at δ 80.25. Both the ^{1}H and ^{13}C chemical shifts were consistent with this position being both benzylic and attached to an acyloxy group. (*i.e.* a benzyl ester). HMBC cross-peaks from the ^{1}H signal at δ 5.58 to Ar' 2-C and to Ar' 6-C confirmed that Ar' was attached at this point. The ^{1}H NMR signals at δ 4.10 and δ 5.58 both gave HMBC cross-peaks to an ester carbonyl signal at δ 171.18. This confirms that these form a five-membered ring, the α-alkylidene-λ-lactone shown in the centre of the structure. The quaternary alkene carbon signal was identified at δ 107 by an HMBC correlation with the signal at δ_H 5.58. This up-field chemical shift is consistent with the α-carbon of an enone.

The ^{1}H signal at δ 5.58 showed a three-bond HMBC correlation to the other ester carbonyl signal at δ_C 170.24. This carbonyl signal was also linked by a two-bond HMBC to the signal at δ 4.10. Thus this ester carbonyl is attached as shown in the structure. The carbonyl resonance showed further HMBC cross-peaks to the OCH_2 signals for the ethyl group, confirming that this is a pendant ethyl ester. Interestingly, the OCH_2 protons are diastereotopic (and thus magnetically inequivalent) and resonate as two double quartets, with chemical shifts δ 4.14 and δ 4.16. The geminal ^{2}J coupling constant between these protons is 10.8 Hz, lower than usual and consistent with the proximity of the electronegative oxygen atom. A COSY cross-peak then linked these resonances to the $OCH_2C\mathit{H}_3$ signal at δ 1.14. The OCH_2 was evident at δ 62.16 and the $OCH_2\mathit{C}H_3$ at δ 13.94, according to HSQC. Thus the CO_2Et unit is attached as shown.

NOESY spectroscopy established the absolute configuration of the alkene and the relative configurations of the two chiral centres, as well as confirming the linkages deduced from HMBC, *etc*. Firstly, NOESY cross-peaks were observed between the methoxy proton signal at δ 3.76 and the Ar 2-H signal at δ 7.03 and between the methoxy proton signal at δ 3.65 and the Ar' 2-signal at δ 6.75, confirming the location of the methoxy groups. Secondly, Ar 2-H was shown to be close in space to the furanone signal at δ 4.10 by NOESY; thus these are on the same side of the alkene, which is proven to be of *E* configuration. Thirdly, a NOESY cross-peak between the furanone signal at δ 5.58 and the $OC\mathit{H}_2$ signals demonstrates that these are on the same face of the five-membered ring. Thus the structure is as shown in Figure 2-35, except that the material is racemic.

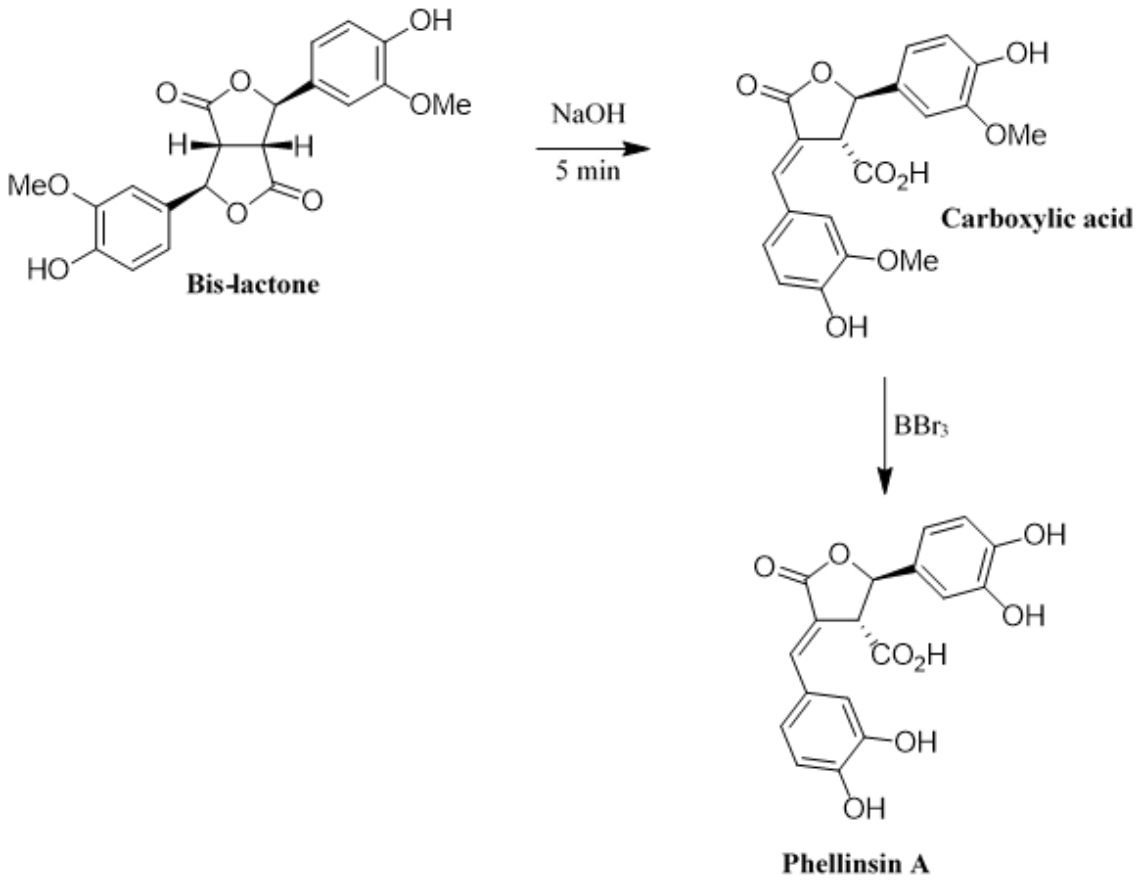


Figure 2-40: Synthesis of the corresponding carboxylic acid by base-catalysed β-elimination from the fused bis-lactone; Kim *et al*. (2005).

Compound **10** is formally novel. However, Kim *et al*. (2005) synthesised the corresponding carboxylic acid (Figure 2-36) by base-catalysed β-elimination from the fused bis-lactone (Figure 2-40). This elimination occurs once only under the mild condition, as the driving force is the relief of ring-strain which had been imposed by the 5:5 ring fusion in the bis-lactone. Demethylation then led to the natural product phellinsin A. Phellinsin A shows strong inhibition of chitin synthase (Hwang *et al*., 2000). Bae and Kim (2012) prepared pheninsin A by oxidative dimerisation of caffeic acid (3,4-dihydroxycinnamic acid). Pharbilignanol D, which is the O-demethylated analogue of **10**, has been reported as being a constituent of the seeds of *Pharbitis nil* (Morning Glory) (Kim et al., 2014). Furthermore, two similar compounds to compound **11** have been synthesised by Ralph *et al*. (1994), 12b the carboxylic acid and the synthesised methyl ester compound 38 (Ralph *et al*., 1994). The ^{1}H NMR that was reported to be very similar to the spectrum for compound **10**.

Compound **11** (*S2*): The HPLC-MS/MS (ESI negative mode) of **11** (Ethyl (±)-2,3-*trans*-5-(*E*-3-hydroxy-3-oxopropen-1-yl)-2-(4-hydroxy-3-methoxyphenyl)-7-methoxy-2,3-dihydrobenzo(b)furan-3-carboxylate) showed a peak at t_R 17.82 min, *m/z* 413.1240 corresponding to $[M-H]^-$ for $C_{22}H_{22}O_8$ (calcd: 413.123645). With a melting range of 160.8 – 162.0 °C. The formate adduct ion was also apparent *m/z* 458 $[M-HCOOH]^-$ as well as the pseudomolecular dimer and its formate adduct; *m/z* 826 $[2M-H]^-$ and *m/z* 371 $[2M-HCOOH]^-$ respectively. MS_2 produced additional fragments *m/z* 323 $[M-HOCH_2CH_3.COOH]$, *m/z* 367 $[M-HOCH_2CH_3]^-$, *m/z* 369 $[M-CO_2]^-$ *m/z* 395 $[M-H_2O]$. In positive ion mode MS, the peak at *m/z* 415 $[M+H]^+$, corresponded to the pseudomolecular ion. Other abundance ions include, *m/z* 351 $[(M+H)-H_2O.HOCH_2CH_3]^+$ and the dimer of the pseudomolecular ion *m/z* 830 $[2M+H]^+$. Positive mode MS_2 gave five significant fragments, *m/z* 397 $[M-H_2O]^+$, *m/z* 369 $[M-HOCH_2CH_3]^+$, *m/z* 383 $[M-HOCH_3]^+$, *m/z* 350 $[M-HOCH_2CH_3.H_2O]^+$, *m/z* 291 $[M-C_6H_3.OH.OCH_3]^+$. The UV spectrum of **11** showed a strong absorption at 242.00 and 323.00 nm which reduced to a shoulder within the range of 340 – 380 nm, which is typical of hydroxycinnamic acid conjugates (Parveen *et al*., 2008).. Compound **11** with appropriate ^1H NMR and ^{13}C NMR chemical shift data shown below (Figure 2-39). ^1H NMR ($CDCl_3$) δ 1.30 (3 H, t, *J* = 7.2 Hz, CH_2CH_3), 3.81 (3 H, s, $PhOCH_3$), 3.88 (3 H, s, benzofuran OCH_3), 4.20 (1 H, d, J = 8.0 Hz, benzofuran 3-H), 6.07 (1 H, d. *J* = 8.0 Hz, benzofuran 2-H), 4.22 (dd, *J* = 16.2, 7.0 Hz, $OCHHCH_3$), 4.25 (dd, *J* = 16.2, 7.0 Hz, $OCHHCH_3$), 6.27 (1 H, d, *J* = 15.6 Hz, propenyl 2-H), 6.85 (1 H, dd, *J* = 8.8, 0.8 Hz, Ph 6-H), 6.86 (1 H, m, Ph 2-H), 6.87 (1 H, dd, *J* = 8.8 Hz, Ph 5-H), 7.01 (1 H, m, benzofuran 6-H), 7.18 (1 H, m, benzofuran 4-H), 7.69 (1 H, d, *J* = 15.6 Hz, propenyl 1-H); ^{13}C NMR ($(CD_3)_2CO$) δ 13.0 (CH_2CH_3), 56.10 (benzofuran 3-C), 56.33 ($PhOCH_3$), 56.51 (benzofuran 7-OCH_3), 62.23 ($COCH_2CH_3$), 88.35 (benzofuran 2-C), 110.77 (Ph 2-C), 113.32 (benzofuran 6-C), 115.77 (Ph 5-C), 116.5 (propenyl 2-C), 118.82 (benzofuran 4-C), 120.23 (Ph 6-C), 130.1 (benzofuran 5-C), 132.00 (Ph 1-C), 145.78 (benzofuran 7a-C), 146.0 (benzofuran 7-C), 146.1 (propenyl 1-C), 148.1 (Ph 3-C), 148.2 (Ph 4-C), 150.0 (benzofuran 7a-C), 168.0(CO_2H), 171.14 (CO_2Et) Data shown n Figure 2-41.

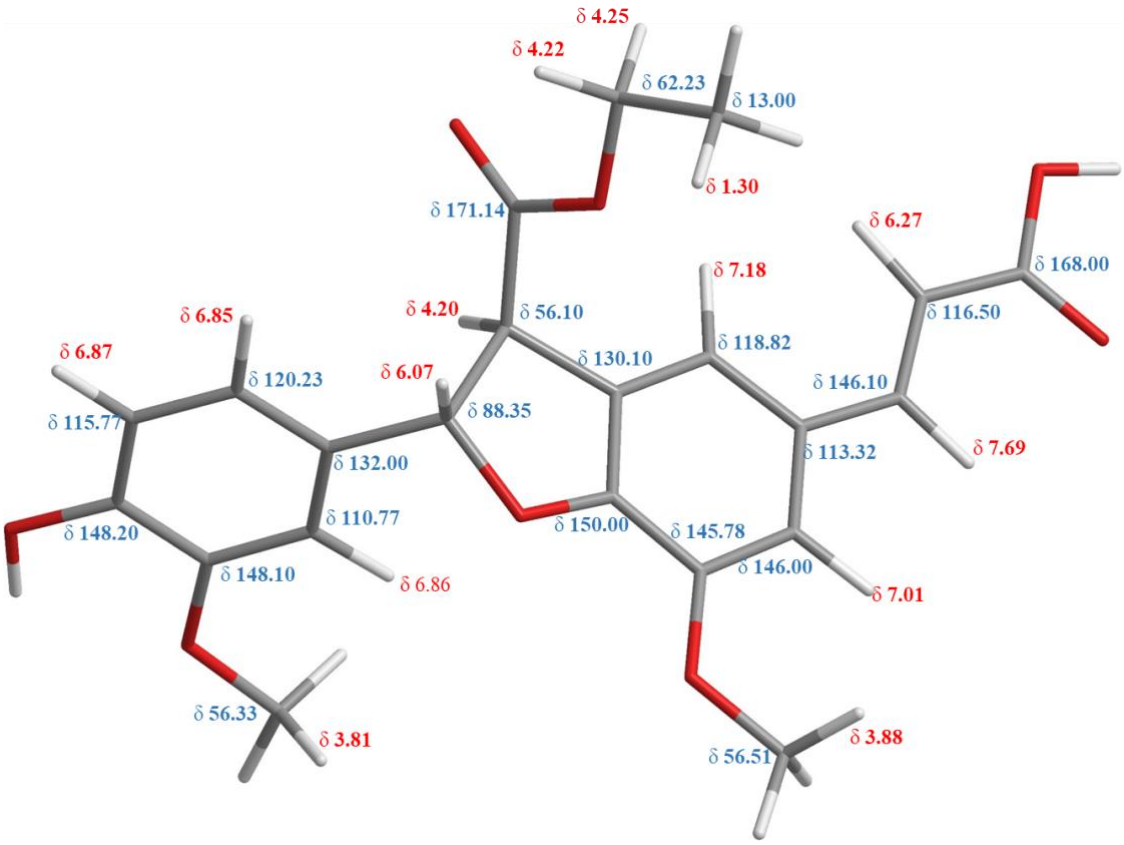


Figure 2-41: Compound **11,** 8-5 ethyl diferulate structural formula with appropriate $_{1}$H NMR (Red) and $_{13}$C NMR (Blue) shift data.

The $_{1}$H NMR spectrum of **11** showed that it was an 8-5 oxidative dimer of the conventional type, as there was a signal for the *trans*-disubstituted alkene. A combination of $_{1}$H, $_{13}$C, HSQC and HMBC spectra was used to assign the structure.

Signals for two 4-hydroxy-3-methoxyphenyl rings were seen in the 1D spectrum. The 1,3,4-trisubstituted pattern was seen for the left ring in the above structure (Ar). Ar 2-H resonated as a mulitplet at δ 6.86, consistent with being *ortho* to an electron donating group (methoxy). The corresponding carbon (Ar 2-C) was identified by an HSQC cross-peak at δ 110.77; this chemical shift is consistent with its environment. Ar 2-H gave three-bond HMBC cross-peaks to Ar 4-C (δ 148.20) and to Ar 6-C (δ 120.23). The downfield chemical shift of Ar 4-C showed that it carried an electronegative atom of oxygen (hydroxy or methoxy). HSQC from Ar 6-C identified Ar 6-H (δ 6.85) to be a double doublet (J = 8.8, 0.8 Hz) and a double doublet (J = 8.8 Hz) at Ar 5-H (δ 6.87). The coupling constant between Ar 5-H and Ar 6-H is consistent with these two protons being *ortho* to each other. A three-bond HMBC correlation from Ar 5-H then identified Ar 1-C at δ 132 and Ar 3-C at δ 148.10. The downfield chemical shift pattern

is due to the attachment of an electronegative oxygen atom. A methoxy group (δ_H 3.81, δ_C 56.33) was shown to be attached at Ar 3-C by HMBC cross-peak from this methoxy proton signal and Ar 3-C. Consequently, Ar 4-C carries a hydroxy group, consistent with the chemical shift, and the structure of ferulic acid.

The other aromatic ring showed similar NMR signals, suggesting a similar substitution pattern. The benzofuran ring was identified with the benzene ring was trisubstituted with 3 H signals. Benzofuran 4-H resonated as a narrow mulitplet at δ 7.18, this further up field chemical shift is consistent with being *ortho* to a weak electron-withdrawing group. HSQC gave correlation at δ 118.82 benzofuran 4-C. Benzofuran 6-H (δ 7.01) which was a mulitplet. The corresponding carbon (6-C) was identified by an HSQC at cross-peak δ 146.00; this chemical shift is consistent with its environment. The benzofuran OCH_3 resonated at δ 3.88 as a singlet, HMBC identified the corresponding δ_C 56.51 at benzofuran 7-OCH_3. The downfield chemical shift of 7-C showed that it is consistent with bring *ortho* to an electron donating group (oxygen). The COSY spectrum contained a cross-peak for the benzofuran 4-H (δ 7.18) with benzofuran 6-H (δ 7.01).

Three-bond HMBC from the benzofuran 4-H to the benzofuran 6-H to C-H carbon signal at δ 146.10, showed that this was the benzofuran 5 substitute. HSQC linked this carbon to a doublet proton signal at δ 7.69 (benzofuran-CH=). The large coupling constant of J = 15.6 Hz indicated the *trans* disubstituted alkene. These chemical shifts indicated that this alkene was attached at the distal end to a carbonyl group, and this unit therefore comprising an enone. HMBC from δ 7.69 (benzofuran-CH=) to δ 168.00 (=CHCO_2), this chemical shift is indicative of carboxylic acid or an ester.

In the benzofuran three-bond HMBC from benzofuran 7-H identified benzofuran 3'-C at δ 130.10. HMBC from benzofuran 4-H identified benzofuran 7a-C at δ 150.00. Benzofuran 4-H was linked by HMBC to aliphatic carbon at δ 56.10 benzofuran 3-C. HSQC showed the benzofuran 3-H at δ 4.20 broad doublet (J = 8.0 Hz). The COSY spectrum contained a cross-peak from the benzofuran 2-H δ 6.07 doublet (J = 8.0 Hz). HSQC showed benzofuran 2-C at δ 88.35. HMBC three bond cross-peak from δ 6.07 to δ 150.00, this signal is evidence of closure of the dihydrofuran ring. This cross-peak is weak because the (2-H)-(2-C)-(O)-(7a-C) dihedral angle is close to 90°, as shown in an MM2-minimised model (Figure2-37). Furthermore three-bond HMBC δ 6.07

benzofuran 3-H (J = 8.0 Hz) to carbonyl at δ 171.14 (benzofuran-CO_2), carbon chemical shift is consistent with ester.

^{1}H spectrum showed an OCH_2CH_3 as triplet at δ 1.30 (J = 7.2 Hz) and showed two double quartets (J = 16.2, 7.0 Hz) at δ 4.22 and δ 4.22. The 14.0 Hz is a indicates a *trans* disubstituted alkene. The two different chemical shifts imply that they are in an asymmetrical environment. HSQC identified the ethyl carbon signals at δ 13.00 (CH_3) and δ 62.23 (CH_2). HMBC cross-peaks from both OCH_2 protons to carbonyl signal at δ 171.14, showed that the ester was located at 3-C of the benzofuran ring.

2.11.3 Diethyl diferulate

Compound **12** *(S5):* Based on mass spectral, NMR and UV data it was confirmed that the synthesis of diethyl diferulate was successful with 73.4 mg, 14.4% yield, this is lower than the yield obtained by Ralph *et al*., (1994) at 28%. A large proportion of the mass will be by-product of the reaction. Further purification will be able to highlight an accurate yield. The HPLC-MS/MS (ESI negative mode) of **12** showed a signification peak at t_R 21.42 min in the PDA chromatogram, indicating the compound is much less non-polar and the ethyl diferulate heterodimers. The base peak *m/z* 441.1551 corresponds to $[M-H]^-$ $C_{24}H_{26}O_8$ (calcd: 442.16277). MS_2 fragmentation pattern; *m/z* 426 $[M-CH_3]^-$, *m/z* 397 $[M-CO_2]$, *m/z* 395 $[M-HOCH_2CH_3]^-$. *m/z* 380 $[M-HOCH_2CH_3-CH_3]$. In positive ion mode MS, the peak at *m/z* 443 $[M+H]^+$, corresponded to the pseudomolecular ion. Positive mode MS_2 provides four fractions, *m/z* 397 $[M-HOCH_2CH_3]^+$ at 100 % intensity with smaller less abundant ions at *m/z* 351 $[M-2HOCH_2CH_3]^+$, confirming the presence of two ethyl esters. Also *m/z* 884 $[2M+H]^+$ and *m/z* 907 $[2M+H+Na]^+$. The proposed structure of the diethyl diferulate was confirmed through ^{1}H NMR. Compound **12** melted at 149.9 – 154.8 ºC, which is consistent of the 8-5 diethyl diferulate reported by Ralph *et al*., (1994), the larger range is due to the impurities found within the sample. The ^{1}H NMR and ^{13}C NMR is consistent of 8-5 diethyl diferulate reported by Ralph *et al*., (1994;1998) (Figure 2-38). ^{1}H NMR ($(CD_3)_2CO$) d δ 1.28 (3 H, t, J = 6.9 Hz, $=CHCO_2CH_2C\mathit{H}_3$), 1.31 (3 H, t, J = 6.9 Hz, benzofuran-$CO_2CH_2C\mathit{H}_3$), 3.10 (3 H, s, $PhOC\mathit{H}_3$), 3.30 (3 H, s, benzofuran $OC\mathit{H}_3$), 4.19 (2 H, q, J = 7.2 Hz, $=CHCO_2C\mathit{H}_2$), 4.26 (1 H, dq, J = 14.0, 7.1 Hz, benzofuran-$CO_2C\mathit{H}$H), 4.28 (1 H, dq, J = 14.0, 7.1 Hz, benzofuran-$CO_2CH\mathit{H}$), 4.37(1 H, brd, J = 7.4 Hz, benzofuran 3-H), 6.03 (1 H, d, J = 7.5 Hz, benzofuran 2-H),

6.39 (1 H, d, *J* = 15.9 Hz, =CHCO$_2$), 6.81 (1 H, d, *J* = 8.3 Hz, benzofuran 7-H), 6.85 (1 H, d, *J* = 8.6 Hz, Ph 5-H),6.92 (1 H, d, *J* = 8.6 Hz, Ph 6-H), 7.10 (1 H, s, Ph 2-H), 6.81 (1 H, s, benzofuran 6-H), 7.67 (1 H, d, *J* = 16.0 Hz, benzofuran-CH=), 7.72 (1 H, m, benzofuran 4-H).^{13}C NMR ((CD$_3$)$_2$CO)): δ_c 14.23 (OCH$_2$*C*H$_3$), 14.34 (OCH$_2$*C*H$_3$), 56.00 (benzofuran 3-C), 56.22 (PhO*C*H$_3$), 56.40 (benzofuran 7-O*C*H$_3$), 167.29 (*C*OCH$_2$CH$_3$), 88.32 (benzofuran 2-C), 62.18 (CO*C*H$_2$CH$_3$), 60.54 (CO*C*H$_2$CH$_3$), 171.07 (*C*OCH$_2$CH$_3$), 110.65 (Ph 2-C), 113.10 (benzofuran 6-C), 115.76 (Ph 5-C), 116.60 (propenyl 2-C), 118.90 (benzofuran 4-C), 120.17 (Ph 6-C), 127.34 (benzofuran 5-C), 129.40 (benzofuran 1-C), 132.00 (Ph 1-C), 145.11 (propenyl 1-C), 145.78

(benzofuran 7-C), 147.91 (Ph 4-C), 147.91 (Ph 3-C), 150.92 (benzofuran 7a-C), 167.29 (*C*O$_2$Et) 171.07 (*C*O$_2$Et). HRESIMS *m/z* 411.1571 [M - H]-.

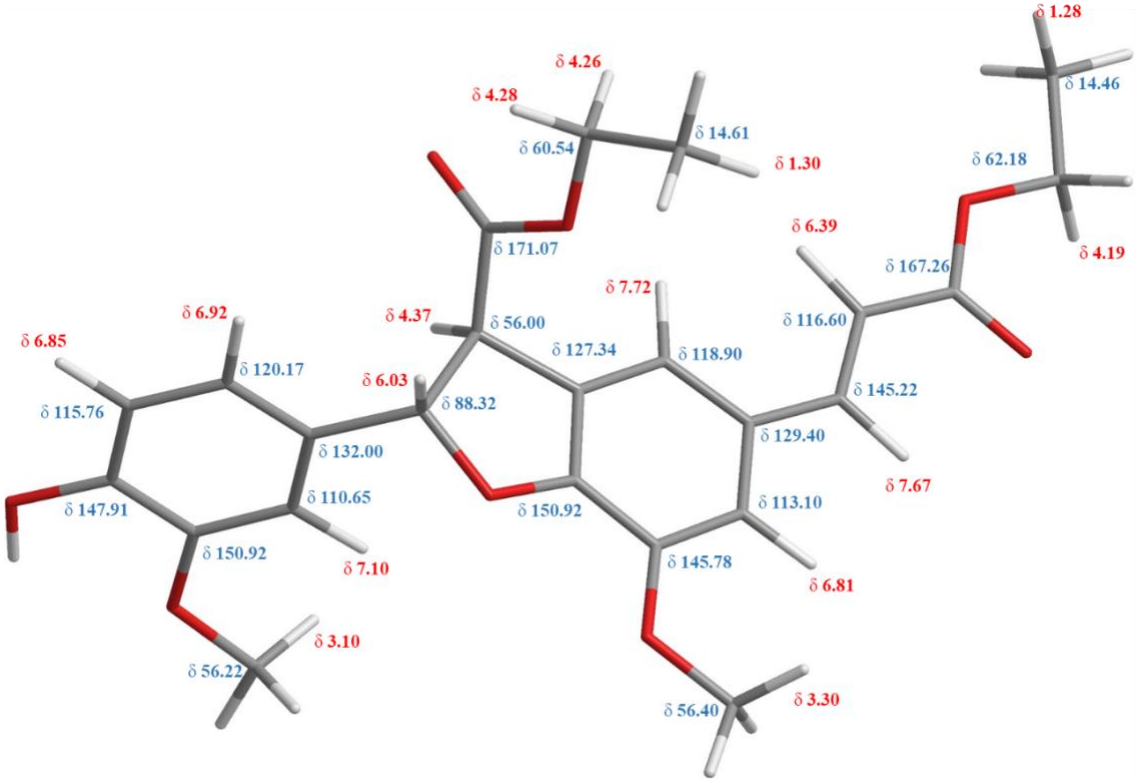


Figure 2-42: Compound **12**; 8-5 diethyl diferulate structural formula with appropriate ^{1}H NMR (Red) and ^{13}C NMR (Blue) shift data.

The ^{1}H NMR spectrum of **12** showed that it was an 8-5 oxidative dimer of the conventional type, as there was a signal for the *trans*-disubstituted alkene. A combination of ^{1}H, ^{13}C, HSQC and HMBC spectra was used to assign the structure. The ^{1}H spectrum showed an OCH$_2$CH$_3$ as triplet at δ 1.30 (*J* = 7.2 Hz) and showed two double quartets (*J* = 14.0, 7.1 Hz) at δ 4.26 and δ 4.28. The 14.0 Hz is a indicates

a *trans* disubstituted alkene. The two different chemical shifts imply that they are in an asymmetrical environment. HSQC identified the ethyl carbon signals at δ 14.61 (CH_3) and δ 60.54 (CH_2). HMBC cross-peaks from both OCH_2 protons to carbonyl signal at δ 171.07 ($COCH_2CH_3$), showed that the ester was located at 3-C of the benzofuran ring.

All chemical shifts are consistent with that of compound **11**. With the addition of the other ethyl ester attached to the terminus of the alkene. The ^1H spectrum showed a triplet (OCH_2CH_3) at δ 1.28 (J = 6.9 Hz) and showed a quartet integrating for two protons at δ 4.19 (J = 7.2 Hz). The same chemical shift imply that the protons are in a symmetrical environment. The further downfield chemical shifts are consistent with being further away from an electron-withdrawing function. HSQC identified the ethyl carbon signals at δ 14.61 (CH_3) and δ 62.18 (CH_2). HMBC cross-peak from the OCH_2 protons to the carbonyl signal at δ 167.26, thus the carbonyl at the terminus of the alkene must be the second ethyl ester.

Compound **12** is structurally similar to the compound pharbilignans C, which showed potent cytotoxicity against a variety of different human tumour cell lines (A549, SK-OV-3, SK-MEL-2 and HCT-15 with IC_{50} values of 1.42, 0.16, 0.20, 0.14 µM respectively (Kim *et al.*, 2014). The only structural different between compound **12** and pharbilignans C is that compound **12** contains two oxymethyl groups, whereas pharbilignans C is made up caffeic monomers, thus having hydroxy group on 3-C and benzofuran 7a-C instead. Further investigation into the cytotoxic activities of compound **12** may lead to significant drug leads.

2.12 Oxidative heterodimerisation of ferulic acid and ethyl ferulate (II)

Ferulic acid (1.21 g, 5.9 mmol) and ethyl ferulate (1.32 g, 5.9 mmol) was added acetone (100 mL) gently heated until dissolved. Purified H_2O (825 mL) was added cautiously, followed by urea hydrogen peroxide (617 mg g, 6.6 mmol) in purified H_2O (15 mL). The pH was monitored and maintained at pH 4 (Schomberg, Salzmann, and Stephan, 1993), followed by the addition of horseradish peroxidase type 2 (5 mg) in H_2O (10 mL). The pink coloured mixture was stirred at ambient temperature for 1.5 h. Indicating the enzyme was still chemically active. Further addition of urea H_2O_2 (316 mg, 3.28 mmol) in H_2O (3 mL) was needed. The mixture was mixed for another 1h, after which the compound was obtained into EtOAc. The residue, in EtOAc was

washed thrice with saturated brine. Drying (Na_2SO_4) and evaporation gave a crude dark orange/brown crude oil (2.51 g, 99 % recovery yield).

The oxidative heterodimerization of ferulic acid and ethyl ferulate was duplicated to increase the amount of yield and to provide more purified samples for NMR and biological assays. The replication of the oxidative coupling reaction provided a chance to improve the solubility of the starting materials and to ensure the reaction remains in solution. This was done by with the addition of aqueous acetone 1:1. This ensured the insoluble sticky mixture was once again solubilised. The crude was obtained into EtOAc, washed with brine and dried using anhydrous $NaSO_4$ to yield 2.511g, 98.77 % yield, this was an improvement of 16.78 % yield. The crude was subjected to HPLC-UV-ESI-MS/MS analysis (Figure 2-43). Chromatogram comparison of reaction I and II show signification similarities even with a column change. Both crude eluated at similar t_R 15.28 and 17.21 corresponding to the same *m/z* of 413.01 with corresponding *m/z* for the next five significant peaks.

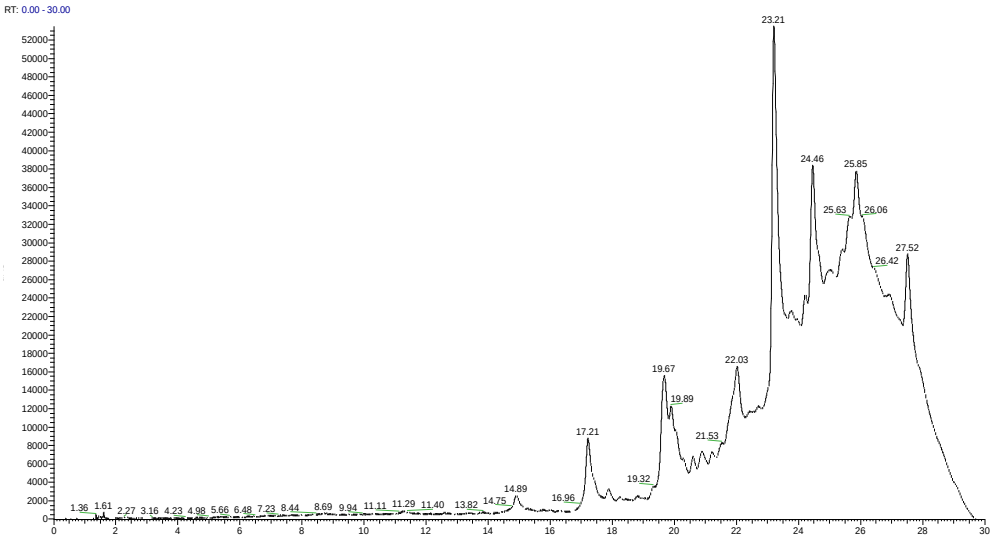


Figure 2-43: PDA chromatogram of ferulic acid and ethyl ferulate (reaction **II**)

2.12.1 Solid phase extraction

The crude substance was purified by SPE. Crude residue 0.600g was reabsorbed onto C_{18} reverse phase silica, loaded onto a Supleco, analytical SUPERCLEAN LC-18 packing 60 mL, 10g column. A solvent system divided into 10% increments from 0% – 70% MeOH. The column were washed with 200 mL of each solvent system under a water jet vacuum. Fractions were collected in a conical flask. The solvent was then

removed with the remaining residue of water removed via a freeze drier in pre-weighed vials (Table: 2-5). Total yield of the column separation 539.9 mg (90 % recovery yield).

Table 2-5: Solid phase extraction of oxidatively coupled ethyl ferulate and ferulic acid.

Sample Name	Mass (mg)
DiF001	0.8
DiF002	8.4
DiF003	21.0
DiF004	241.3
DiF005	38.4
DiF006	137.3
DiF007	92.7

2.12.2 Semi-preparative purification of DiF003 and DiF004

HPLC-UV-ESI-MS/MS enabled analytical characterisation of peaks of interest and official PDA-UV spectra to compare against with the water alliance and semi-preparative systems. Samples DiF003 (21.0 mg) and DiF004 (241.3 mg) demonstrated effective samples for suitable method development on the waters C_{18} semi-preparative system. The purification for DiF003 started with the equilibration of the column at 60 % H_2O and 30 % CH_3CN to 10 % H_2O and 80 % CH_3CN, over 42 min, with the UV photo diode array set at 205 nm. Method development of DiF004 concluded the same separation in only 11 min. The separated fractions were analysed again by HPLC-UV-ESI-MS/MS, and to identify those suitable for NMR analysis (Table 2-6).

Table 2-6: Weights of DiF003 and DiF004 after semi-preparative purification. Only the priority peaks were collected for sample DiF003.

Sample Name DiF003_...	*Mass (mg)*	*Sample Name DiF004_...*	*Mass (mg)*
7	*6.2*	*A*	*42.4*
9	*1.7*	*B*	*12.8*
17	*1.4*	*C*	*16.4*
20	*3.1*	*E*	*4.3*
-	-	*F*	*7.3*
-	-	*G*	*8.4*

*Compound **13** (DiF003_7):* The HPLC-MS/MS (ESI negative mode) of **13** presented a peak at t_R 13.42 min, HRESIMS *m/z* 385.0928 $[M-H]^-$ corresponding to a diferulic acid, $C_{20}H_{18}O_8$ (calcd: 386.10017) 6.2 mg 30 % yield. This suggests a diacid homodimer. Yield 6.2 mg, (0.27 % yield) as brown crystals. λ_{max} 324 nm, typical of hydroxycinnamic acid conjugates (Parveen *et al*., 2008). An ion was observed at *m/z* 341 $[M-H-CO_2]^-$, furthermore a dimer of the pseudomolecular ion was also observed at *m/z* 770 $[2M-H]^-$. MS_2 produced one additional fragment *m/z* 340 $[M-COOH]^-$. In positive ion mode, the pseudomolecular ion corresponded to *m/z* 387 $[M+H]^+$. Further ions observed included, *m/z* 342 $[(M+H)-COOH]^-$, *m/z* 297 $[(M+H)-2COOH]^+$, confirming the presence of the two carboxylic acids. MS_2 fragmentation yielded one other ion at *m/z* 369 $[M-H_2O]^-$. Further purification of the sample is needed for melting point and NMR analysis as the chromatogram contains three other smaller peaks.

*Compound **14** (DiF003_17):* The HPLC-MS/MS (ESI negative mode) of **14** showed a peak at t_R 17.97 min, HRESIMS *m/z* 413.1234 corresponding to $[M-H]^-$ for $C_{22}H_{22}O_8$ (414.13147) ethyl diferulate. This suggests a heterodimer (1.4 mg 0.06 % yield) as brown crystals. λ_{max} 337 nm, λ_{max} 324 nm, typical of hydroxycinnamic acid conjugates (Parveen *et al*., 2008). (1.4 mg 7 % yield). The formate adduct ion was also apparent *m/z* 458 $[M-HCOOH]^-$ as well as the pseudomolecular dimer and its formate adduct; *m/z* 826 $[2M-H]^-$ and *m/z* 371 $[2M-HCOOH]^-$ respectively. MS_2 produced additional fragments *m/z* 323 $[M-HOCH_2CH_3.COOH]$, *m/z* 367 $[M-HOCH_2CH_3]^-$, *m/z* 369 $[M-CO_2]^-$ *m/z* 395 $[M-H_2O]$. In positive ion mode, peak at *m/z* 415 $[M+H]^+$, corresponded to the pseudomolecular ion. Other abundance ions include, *m/z* 351 [(M+H)-

$H_2O.HOCH_2CH_3]^+$ and the dimer of the pseudomolecular ion *m/z* 830 $[2M+H]^+$. Positive mode MS_2 gave five significant fragments, *m/z* 397 $[M-H_2O]^+$, *m/z* 369 $[M-HOCH_2CH_3]^+$, *m/z* 383 $[M-HOCH_3]^+$, *m/z* 350 $[M-HOCH_2CH_3.H_2O]^+$, *m/z* 291 $[M-C_6H_3.OH.OCH_3]^+$. The UV spectrum of **14** showed a strong absorption at 324.00 nm. This compound has a very similar fragmentation pattern to that of compound **12**. The t_R of the two compounds differ by 1 min, when run as a mixed spot. Suggesting the presence of a structural isomer.

The HPLC-MS/MS (ESI negative mode) of the crude sample DiF004_B showed two peaks at t_R 16.71 min and 19.18 min, *m/z* 413.1234 corresponding to $[M - H]^-$ ($C_{22}H_{21}O_8$ requires 413.1236), corresponding to the molecular formula $C_{22}H_{22}O_8$. This suggests two heterodimers λ_{max} (nm) 337 nm, typical of hydroxycinnamic acid conjugates (Parveen *et al*., 2008). Fragmentation is consistent with compounds **11** and **12** respectively. Semi preparative purification of fraction DiF004_B yield two peaks, compound **11** *m/z* 413.1239 corresponding to $[M-H]^-$ for $C_{22}H_{22}O_8$ (calcd: 413.123645). Orange oil. 4.4 mg, (0.17 % yield) and compound **12** was overserved at t_R 19.12 min, *m/z* 413.1240 corresponding to $[M-H]^-$ for $C_{22}H_{22}O_8$ (calcd: 413.123645). m.p. 160.8 – 162.3. (2.3 mg, 1 % yield). 2.3 mg, (0.09 % yield).

The HPLC-MS/MS (ESI negative mode) of the crude fraction DiF004 C, showed a peak at t_R 22.79 min, *m/z* 441.1551 corresponding to $[M-H]^-$ the pseudomolecular ion for $C_{24}H_{26}O_8$ (calcd: 442.16277). The λ_{max} 324.00 nm (16.4 mg, 1.5 % yield). HPLC fragmentation, accurate mass, mix spot and melting point analysis confirmed the successful identification of the 8-5 diethyl diferulate previously as compound **13**.

With the aim of getting better separation solid phase extraction was preformed using a SUPERCLEAN LC-18 packed 60 mL column 10g. The crude extract (600 mg) was pre-absorbed onto C_{18} silica and loaded onto the Supleco column. The mobile phase consisted of aqueous MeOH in 10% increments from 0 – 70 % MeOH. The solvent was then removed, and then freeze dried to yield fractions table 1-6. The total yield of the column separation 539.9 mg, 89.9 % yield, this method of solid phase extraction increased the separation yield by 15% compared to Biotage flash chromatography. HPLC-UV-ESI-MS/MS identified two samples, both containing peaks of interest corresponding to the diacid, heterodimer and diethyl diferulate structures. DiF003 and DiF004 were both separated using the same purification gradient.

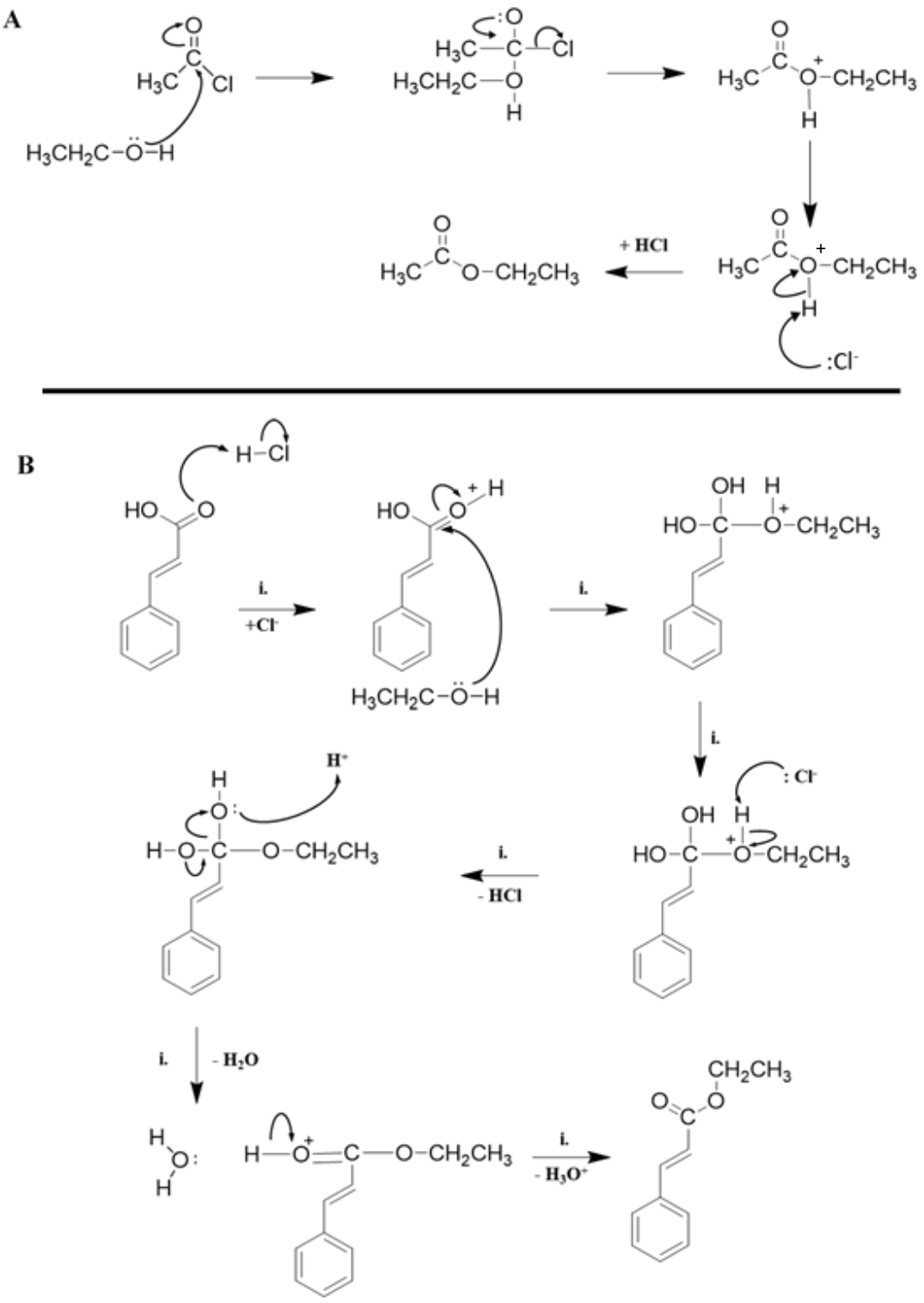
DiF003 provided a 60 % yield across four fractions after semi preparative purification. The HPLC-MS/MS of the obtained fraction DiF003_7 highlighted the presence of the proposed diferulic acid, *m/z* 385.0928 $[M-H]^-$ for $C_{20}H_{18}O_8$ (calcd: 386.10017). HPLC-ESI-MS/MS analysis yielded three ions, *m/z* 770 $[2M-H]^-$, *m/z* 341 $[M-CO_2]^-$ and *m/z* 208 [(ferulic acid-H) + CH_3]-. In MS_2 negative ion mode yielded a peak at 100 % intensity, *m/z* 341 $[M-H.CO_2]^-$, a further smaller ion included *m/z* 294 $[M-2COOH]^-$, supporting the evidence of the dimerisation of ferulic acid. The fragmentation pattern is also consistent with the research of Aljawish *et al*., (2014), further supporting the formation of 8-5 diferulic acid. Structural identification is not possible due to the low yield provided by this fraction. Further research has identified that naturally occurring phenolic hydroxycinnamic are mostly present in *trans* configuration. Clifford *et al*.,(2008) has shown that isomerisation into the *cis* form is prevalent when exposed to UV light.

2.13 *Discussion*

Since the beginning of the HIV epidemic, thirty-six million people are currently living with HIV/AIDs-related diseases. Unfortunately, 1.8 million are children below the age of fifteen (UNAIDS, 2016). This epidemic reduces the average lifespan by approximately fifteen years (Mathers *et al*., 2001). In an effort to reduce the child mortality rate, Bruzzese *et al.* (2018) have reported safe and effective HIV-1 therapy in infected children with the use of a combination dolutegravir mix. An increase in mortality of HIV-1 patients is significantly increased within the first six months of having a co-infection with hepatitis or tuberculosis, without HAART (Marcy *et al*., 2018). With drug resistance being recorded for every available drug class in patients with HIV, designing novel inhibitors based on structural similarities and building novel therapeutics with altered mechanisms of action has been made a top priority for scientists worldwide.

2.13.1 *Mechanism of esterification*

According to literature, there are many different mechanisms that can achieve the successful synthesis of the four hydroxycinnamic acids (Scheme 2-1), examples include; laccase-catalysed oxidation and celite-immobilised lipase conversions (Kuma and Kamwar, 2011; Aljawish *et al*., 2014). These biocatalysied reactions have specific requirements for pH and other environmental factors. In this study, a simple acid-catalysed esterification reaction was used for the synthesis of the three-hydroxycinnamic esters (Quideau and Ralph, 1992). The carbonyl oxygen on the hydroxycinnamic acid is protonated by HCl generated from the reaction between acetyl chloride and ethanol. The lone pair of electrons on the oxygen of ethanol functions as the nucleophile attacking the electrophilic carbonyl group, resulting in a tetrahedral intermediate. Protonation of the hydroxy group on the tetrahedral intermediate, results in the loss of water a molecule thus leading to deprotonation and resulting in the ester (Scheme 2-6).



Scheme 2-6: Proposed mechanism for the synthesis of hydroxycinnamate esters: ethyl cinnamate. (A) Formation of HCl with acetyl chloride and EtOH. (B) Esterification of ethyl cinnamate with a tetrahedral intermediate. *Reagents and Conditions: i Acetyl chloride and EtOH.*

2.13.2 *Mechanism of action of horseradish peroxidase*

Peroxidases are enzymes that are found in most aerobically respiring organisms, playing a vital role in preventing oxidative damage of cellular components caused by hydrogen peroxide (H_2O_2). By catalysing the heterolytic cleavage of the peroxide bond, peroxidase avoids the accumulation of reactive oxygen species (ROS) (Zamocky *et al*., 2008). Peroxidases, such as horseradish peroxidase, use H_2O_2 to oxidise organic substances, this is known as proxidatic reaction Figure 2-44.

The enzymatic cycle of peroxidases is initiated with H_2O_2-mediated oxidation of the resting state of the enzyme to a high-valent iron-oxo intermediate known as compound I, with the release of H_2O (Dunford, 1999; Derat *et al*., 2005), resulting in an oxoferryl porphyrin radical cation (Chuang and Vanwart, 1992; Ivancich *et al*., 1997; Kim *et al*., 2006). Peroxidases favourably catalyse compound I reduction by one-electron-donating substrates; typically phenols are used to recover the native ferric state, *via* another intermediate compound II. Compound II is described as an oxo/hydroxo ferryl haem intermediate (Derat and Shaik, 2006; Dunford and Dunford, 2010). All HPLC-UV-ESI-MS/MS, ^{1}H NMR and ^{13}C NMR data for the oxidative heterodimerisation of 4'-hydroxycinnamate esters with 4'-hydroxycinnamic acid can be found within the appendix supplementary information.

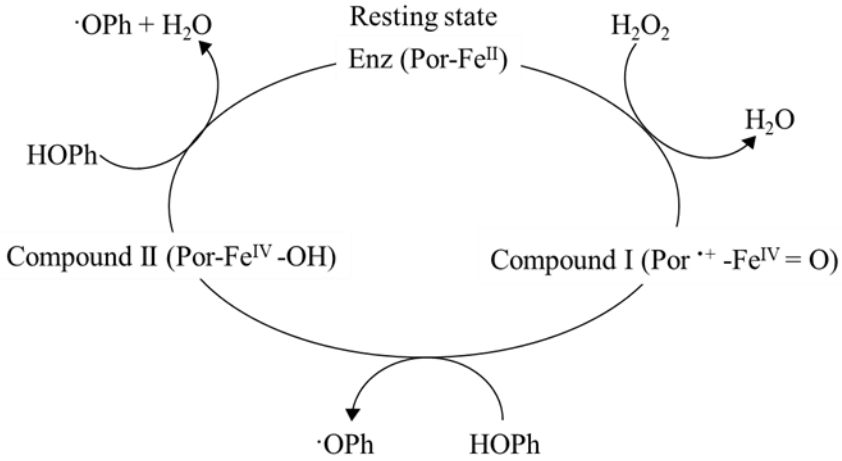
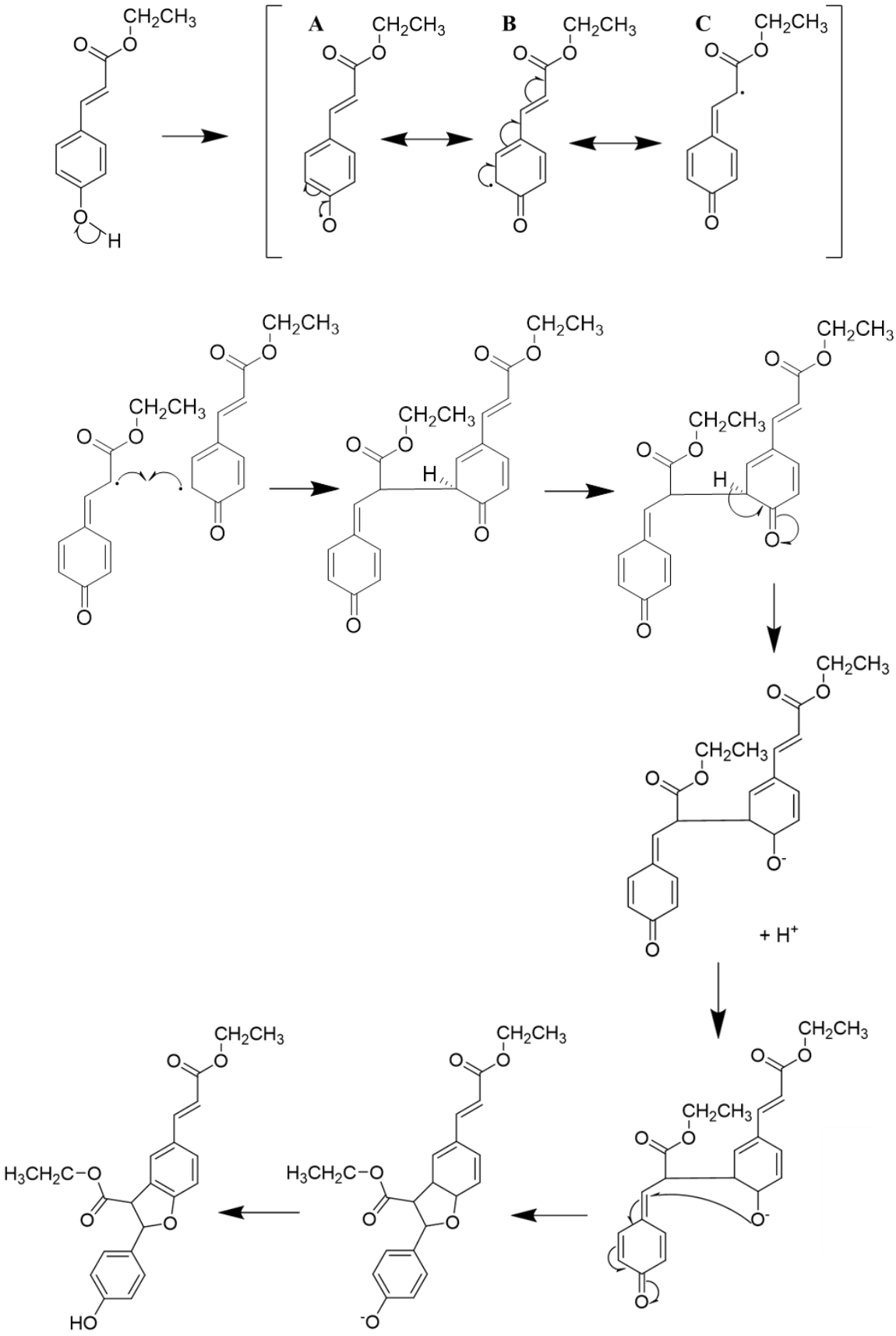


Figure 2-44: The peroxidatic reaction catalysed by peroxidase using hydrogen peroxide (H_2O_2). The cycle being with Compound I formation from the reduction of H_2O_2. Compound I reacts with a molecule of a one-electron donating substrate (e.g., phenol, PhOH), transforming to another intermediate termed compound II. Modified from Campomanes *et al*., (2015).

2.13.3 *Proposed reaction mechanism for the conventional 8-5 linkage and compound **10**.*

The oxidative cross-coupling reaction is initiated by the reduction of H_2O_2 by horseradish peroxidase, creating compound **I** (Figure 2-25). Compound **I** reacts with the hydroxyl group on the 4-C on the benzene ring, an electron donating substrate, creating another intermediate compound **II** and an oxygen radical. With further oxidation of a hydroxy phenol group, thus reduces ferric enzyme back to the resting native state. Resulting in two radical intermediates. Each radical intermediate yielded can tautomerize into three different tautomers (**A**, **B** and **C**) (Scheme 2-7). This mechanism has previously been reported by Ralph *et al*., (1994).



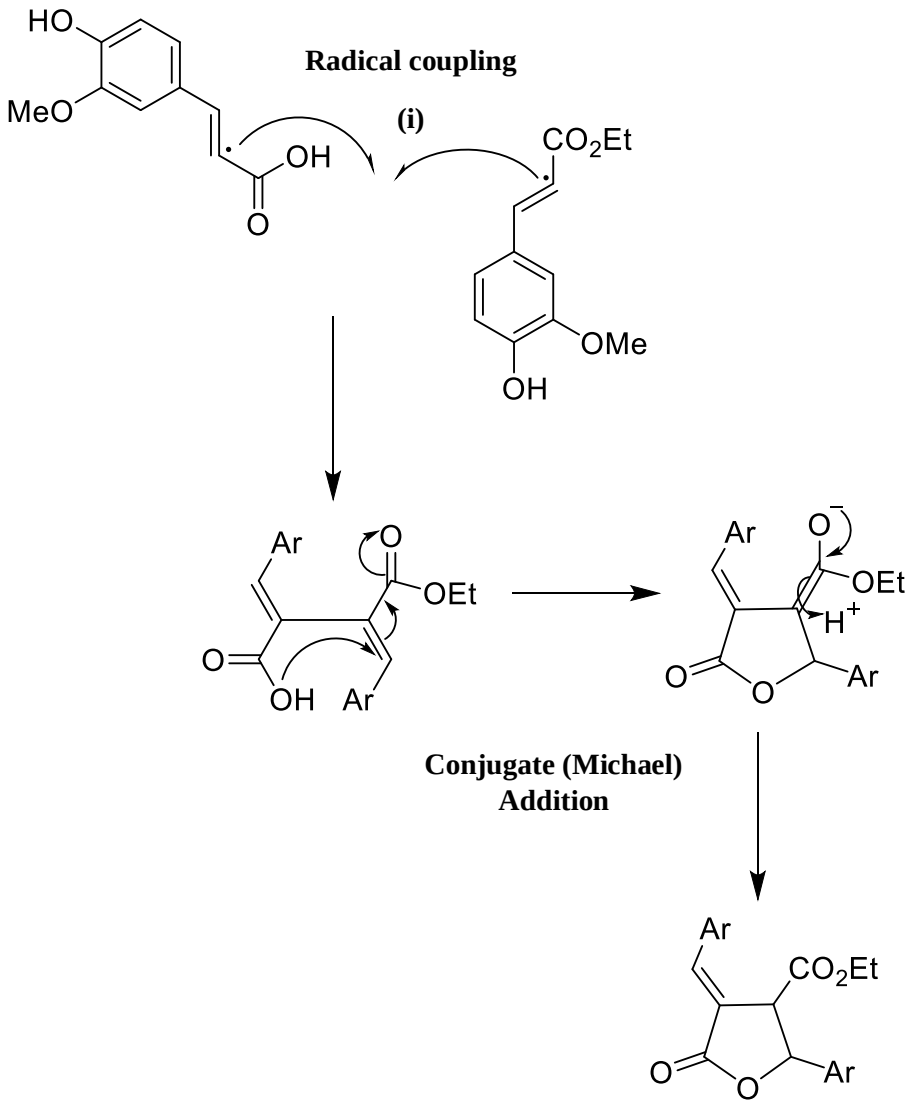
Scheme 2-7: Proposed mechanism for the formation of the 8-5 linkage. *Reagents and condition;* acetone, H_2O_2, peroxidase, pH4. Depicted and expanded from Ralph *et al*., (1994).

The crude chromatograms of each oxidative heterodimerisation highlights a significant challenge. One hydroxy group is essential for the oxidative reaction to take place (Campomanes *et al*., 2015) as identified by the successful synthesis ethyl cinnamate and cinnamic acid oxidative coupling reaction. Ethyl *para*-coumarate and *para*-coumaric acid oxidative coupling seemed to follow the proposed mechanism for the formation of the 8-5 linkage. No diacid heterodimers were visible with the HPLC-ESI-UV-MS/MS chromatogram. Isolation of two novel compounds; (**2**) 8-5 ethyl dicoumarate, where NMR analysis has identified the ethyl ester to be present off the furan ring, and (**3**) 8-5 diethyl dicoumarate. Further purification of compound **1**, may provide evidence of the ethyl ester at the top of the molecule, this may explain the significant difference in t_R (3.44 min).

The addition of any further hydroxy groups on the hydroxycinnamate acids, seem to complicate the reaction mixture, providing the radical with more chemical oxidative pathways, leading to a significant increase in stereoisomers produced. This issue is visible by the crude chromatograms of caffeic acid and ethyl caffeate oxidative coupling having the higher proportion of hydroxy groups. Thus consequently lowering the yield of the desired compounds with the 8-5 linkage, and creating significant challenges for purification, resulting in limited yields and poor solubility precluding NMR analysis. Solid Phase extraction was a purification technique used to break up and fragment the caffeic acid and ethyl caffeate crude, as the HPLC chromatogram highlighted it was unsuitable for crude Biotage Flash Chromatography. Followed by semi preparative purification for successful isolation, leading to poor solubility precluded NMR analysis.

By means of the addition of the less reactive methoxy group for the ethyl ferulate and ferulic acids, oxidative coupling seemed to reduce the complexity of stereoisomers formed. Furthermore, as shown by the ^{1}H NMR, the ethyl ester on the heterodimer is always located on the furan ring, so far they ethyl ester has not been observed on the top of the molecule, leaving the carboxylic acid group on the furan ring. A possible method of the free Michael addition, presence of the ester at the top makes the alkenic protons electrophilic, thus creating the ethyl ester on the furan ring. Alternatively, the ^{1}H NMR spectrum of **10** showed that it was not a 8-5 oxidative dimer of the conventional type, as no *trans*-disubstituted alkene was observed. The mechanism of how **10** is formed *via* a conjugate Michael addition, a nucleophilic addition of a

carbanion to an, β-unsaturated carbonyl compound, the mechanism has been illustrated below (Scheme 2-8).



Scheme 2-8: Proposed mechanism of the oxidative radical coupling of ethyl ferulate and ferulic acid to form compound **10**. *Regents and conditions: (i)* Acetone, hydrogen peroxide urea tablets, pH4; horseradish peroxidase; at ambient temperature and atmospheric pressure.

2.14 *Final remarks*

By using different hydrocinnamaic acids and their corresponding hydroxycinnamate esters this thesis has investigated the addition hydroxy and methoxy groups has on the synthesis of stereoisomers of hydroxycinnamate heterodimers, with various yields.

The overall aim of the study was to synthesis a portfolio of oxidatively coupled 4'-hydroxycinnmanic acids with the corresponding 4'-hydroxycinnmate esters, as

potential anti-HIV drug leads, with the 8-5 linkage based on the structural similarity to lithospermic acid.

To further refine the methodology of this project, multiple aspects must be investigated. The resulted chromatograms for each crude highlighted how differently each oxidative coupling reacted, thus purification techniques and yields varied significantly between each coupling. The project has yielded and obtained; diacid-homodimers, ethyl/acid-heterodimers and diethyl-homodimers. The less abundant diacid homodimers proved difficult for identification *via* the HPLC chromatograms, and thus purification and structurally elucidation was inadequate as yields where small. A possible solution is to synthesis a much higher proportion of a desired product is to use Le Chatelier's Principle. Thus to influence the formation of ethyl homodimer and diethyl heterodimers and therefore increasing the yields, by reducing diacid formation. To obtain diacids, simply only involve diacids within the reaction.

From the obtained MS data, starting materials was present in all most of the crudes. Suggesting that the coupling reaction was incomplete. Additional research into ensuring the reaction is complete and the compounds produced are stable needs to be explored. The thesis also used three purification techniques with varied results, Solid Phase extraction; Biotage Flash Chromatography and Semi preparative purification. Solid Phase extraction was used to fragment the complicated crudes of caffeic acid, ethyl caffeate and ferulic acid, ethyl ferulate. To improve this technique the use of a larger column would remove the issue of repeating smaller runs. Biotage Flash chromatography, was a great asset to have, by purifying the crude sample consistently and reliably. Furthermore, semi-preparative purification was required on caffeic and ferulic crude due to the complexity and how similar in polarity the compounds were.

Moreover, the presentation of some NMR results was successful and leading to the structural identification of multiple novel compounds (compounds **2, 3, 10** and **11**). The NMR results highlighted the ethyl homodimers only containing the ethyl ester present on the furan ring. Furthermore, the appearance of most NMR results were not successful due to most compounds yielding small amounts and containing multiple peaks; other samples experienced poor solubility issues with $CDCl_3$ especially the dicaffeate products, use of another deuterated solvent such as d.acetone may enable further structural elucidation of the more insoluble products. Finally, the isolated

compound can be screen in other biological assays, including but not limited to antimicrobial, anti-mutagenic and anthelminthic, within the institute of IBERS.

2.15 *Final conclusions*

The overall aim of the study was to synthesis a portfolio of hydroxycinnamic acids with their corresponding hydroxycinnamate ethyl esters, as potential anti-HIV candidates. These are structurally similar to a known potent non-toxic integrase inhibitor, lithospermic acid. First of all the synthesis of ethyl cinnamate, ethyl *para*-coumarate, ethyl caffeate and ethyl ferulate, with high yields of 81%, 85%, 71% and 65% respectively.

The oxidative coupling of cinnamic acid and ethyl cinnamate provided evidence towards the reaction mechanics. The *para*-hydroxy phenol is needed for the reaction to take place within the oxidative reduction relationship with the peroxidase enzyme. The coupling of ethyl *para*coumarate and *para*-coumaric acid yielded three significant peaks within the HPLC chromatogram, with effective Biotage purification of compounds **2** and **3** were structurally elucidated to be novel compounds, compound **1** had insufficient yields for NMR analysis. Further purification runs will allow structural elucidation of compound **1**. Oxidative coupling of ethyl caffeate and caffeic acid proved to be very challenging. The radical chemistry had the use of an extra hydroxyl group on C-3, thus creating multiple analogues, with further NMR solubility issues hampered structural elucidation analysis of this reaction. Furthermore, oxidative coupling of ferulic acid and ethyl ferulate, was not as challenging as caffeic acid. The use of two main purification techniques (Biotage chromatography and solid phase extraction), yielded similar results, the identification of the novel compound **10**, to which is similar to the known compound Phenllinsin A, and the novel compound **11**, 8-5 ethyl diferulate. With more time and laboratory training using the Biotage instrument, purification of the oxidative crudes may have increased the yields, possibly therefore not relying on the required semi-preparative system. Additional research is required to synthesise, identify and isolate the homodimer stereoisomers that possess the ethyl ester at the top of the compound, thus leaving the carboxylic acid off the furan ring. The accomplished synthetic pathway provided multiple novel compounds with fair yields.

Chapter 3

Computational docking of potential hydroxycinnamate ester derivatives with the structure of HIV-1 integrase.

3.0 *Abstract*

Competition of the Human Genome Project, triggered major developments within pharmacological industry. Virtual screening is used for hit identification and lead optimisation. A rational drug discovery approach with exceptional reliability. Structure-based drug development is the mutual method of predicting the orientation of the ligand when incorporated into the receptor. With significant biological evidence, computational interactions can be analysed by the calculation of docking scores, calculated using Van der Waals, conventional hydrogen bonding and π-cation interactions, without biological evidence, only visual comparisons can be analysed. This study focuses on the docking the stereoisomers mentioned in the above chapter (Chapter 2) within the Human Immunodeficiency Virus-Integrase (HIV-IN) protein. IN inhibitors can be divided into two regions; inhibitors of 3'processing and strand-transfer inhibitors. 3'-processing inhibitors bind at the donor vDNA binding site and strand-transfer inhibitors at the acceptor DNA-binding site. Current IN inhibitors belong to the IN strand-transfer group, as they display a β-hydroxy carbonyl, which is thought to interact with the two Mg^{2+} ions coordinating the three catalytic residues Asp64, Asp116 and Glu152. Cardiff University School of Pharmacy and Pharmaceutical science completed the molecular modelling using Molecular Operating Environment (MOE) 2018.10. From docking, it seems that the ester group on the tetrahydrofuran ring is forcing the compound far from the active site (8-5 ethyl diferulate) and this might result in a non-inhibition of integrase. In the docking simulation, the lateral acid group is co-ordinating the Mg^{2+} but research suggests that the acid group on the tetrahydrofuran should do the co-ordination, while the lateral chain acid should interact in a similar manner of the reported inhibitor L-731,988. When the ester is in the lateral chain and there is the free acid on the tetrahydrofuran ring, this co-ordination seems still possible.

3.1 *Introduction*

Traditionally, disease targets, such as receptor proteins, were identified *via* a complicated and lengthy process. With completion of the Human Genome Project in 2003, major developments have resulted in increasing number of new therapeutic targets for drug discovery. Scientific advancement has resulted in high-throughput protein purification, crystallography and NMR spectroscopy techniques. These have all contributed and established the structural details of proteins and protein-ligand complexes, therefore making molecular docking a part of modern-day drug discovery (Walters *et al*., 1998; Langer and Hoffmann, 2001; Bajorath, 2002; Jorgensen, 2004; Kitchen *et al*., 2004). More recently, lead discovery and identification is performed by *in-silico* techniques (Yamaotsu and Hirono, 2018). An example of this is called virtual screening (VS). VS is used for hit identification and lead optimisation (Gohlke and Klebe, 2002). VS is a rational drug discovery approach with great reliability (Bailey *et al*., 2001; Shoichet *et al*., 2002; Moitessier *et al*., 2008). This method can be split into two pathways; ligand-based and structure-based. When an active ligand is known but with limited structural information for target, ligand based methods such as quantitative structure activity relationship (QSAR) methods are employed. Structure-based drug development is the main common method of predicating the orientation of the ligand when incorporated into the receptor or enzyme. Using the known three-dimensional shape and the electrostatic exchanges to quantify the interactions such as, Van der Waals, conventional hydrogen bonds and π-cation interactions are used in the calculation of the docking scores (Yuriey *et al*., 2010). The scoring system can only be used and have a significance if the reaction between the target protein and the ligand has a known biological activity.

The lock and key theory was used on the earliest reported docking methods, therefore treating both the ligand and the receptor as rigid bodies. This results in their affinity being directly proportional to the geometrical fit between each other (Mezei, 2003). Koshland put forward the theory of the induced fit, suggesting that receptor and ligand should both be treated as flexible during docking (Hammes, 2002; Koshland, 1963). Subsequently, these flexible docking algorithms not only predict the binding mode of a molecule more accurately than rigid body algorithms, but also its binding affinity relative to other compounds (Verkhivker *et al*., 2000). Other the past two decades, more than sixty different docking tools and programs have been developed, such as

AutoDock (Osterberg *et al*., 2002) DOCK (Venkatachalam *et al*., 2003) and GOLD (Jones *et al*., 1997). The programs were tested for their abilities in producing the correct binding mode of a ligand to its biological target. AutoDock and GOLD had the most reliable sores, as GOLD was able to identify the correct ligand binding poses consistently with 90% accuracy (Wang *et al*., 2016).

3.1.1 *In-Silico docking of potential HIV-1 integrase inhibitors*

Human immunodeficiency virus type 1 (HIV-1) integrase (IN) enzyme and their inhibitors constitute a major advancement in AIDS research, highlighting potent antiretroviral effects in clinical practice (Pommier, Johnson and Marchand, 2005 and Savarino, 2006). As there is no known similar enzyme in humans, specific HIV-1 IN inhibitors can be designed to have minimal side-effects compared to other antiviral agents (Pommier, Johnson and Marchand, 2005; Rice and Baker, 2001). The family of enzymes have a similar 3D folding catalytic domain and a consistent catalytic triad of metal-coordinating carboxylates. HIV-1 IN is a composed of three separate regions, N-terminal (1-50), catalytic core (51-212) and C-terminal (213-288). The N-terminal domain is involved in enzyme multimerisation, while the C-terminal domain has a strong non-specific DNA-binding activity and is therefore called the DNA-binding domain. The catalytic core contains three residues: Asp_{64}, Asp_{116} and Glu_{152}, these coordinate with the metal cation, Mg^{2+}. IN is involved in two reactions 1) 3'-processing, and 2) strand-transfer. Reverse transcriptase transcribes viral RNA into DNA, IN-catalysed 3'-processing removes a 3'-terminal portion at both ends of the newly transcribed viral DNA (vDNA) (also referred to as donor DNA). Once 3'-processing has finished, IN multimers remain bound to both ends of the vDNA and these pre-integration complexes translocate into the nucleus (Stevenson, 2000). The second reaction catalysed by IN, strand-transfer, introduces both 3'-end of HIV vDNA into a host cell chromosome (also referred to an acceptor DNA). This process leaves a five base-pair, single-stranded gap at each junction between the integrated vDNA and the host acceptor DNA, in addition to the two-base flap at the 5'-end of the proviral DNA. With the use of cellular DNA repair enzymes, the newly formed DNA is filled in (Pommier *et al*., 2005).

IN inhibitors can be divided into two regions; inhibitors of 3'processing and strand-transfer inhibitors. Research shows that 3'-processing inhibitors bind at the donor

vDNA binding site, while the strand-transfer inhibitors occupy the position of the acceptor DNA-binding site, after IN forms a transitional complex with the donor DNA. This theory is supported by biochemical evidence (Espeseth *et al.*,2000; Pommier *et al.*,2006; Johnson *et al.*, 2006). IN inhibitors currently in clinical practice belong to the IN strand-transfer inhibitor group. These compounds chemically display a β-hydroxy carbonyl (Figure 3-1), this structure is thought to interact with the two metal ions coordinating the three catalytic residues, Asp_{64}, Asp_{116} and Glu_{152}.

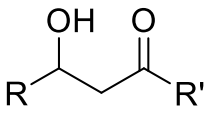


Figure 3-1: Chemical structure of the β-hydroxy carbonyl motif.

Using a two-Mg^{2+} metal model for HIV-1 IN catalytic core domain (CCD) in complex with a small molecule, 1-(5-chloroindol-3-yl)-3-(tetrazolyl)propane-1,3-dione (5-CITEP) (Figure 3-2) could be used as a surrogate for an IN/viral DNA complex, as it allows replication of contacts documented biochemically in viral DNA/IN complexes (Goldgur *et al.*, 1999). Research concluded that the docking showed efficiency and reproducibility, with the fitness of different compounds. The IN/5-CITEP complex correlated to $P<0.01$ with their 50% inhibitory concentration (IC_{50}s) in their strand-transfer assays *in vitro* (Savarino, 2007). The 3D model presented by Saverino (2007) supports the idea that IN strand-transfer inhibitors bind at the acceptor DNA-binding site in the IN/vDNA complex. The mechanism of enzyme inhibition, likely to be exploited by some natural products, releasing future approaches for inhibition of nucleic acid-influencing enzymes and their diseases.

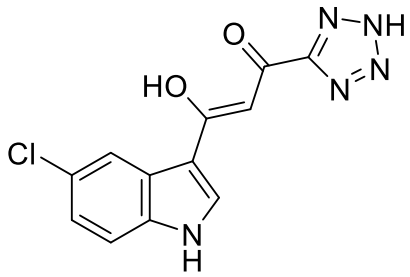


Figure 3-2: Structural representation of 5-CITEP, which occupies the donor, 3'-processing viral DNA binding site , therefore allowing the ST acceptor DNA binding site be too available for docking studies.

By using a 3D platform and exploiting crystallographic data on IN CCD in complex with 5-CITEP, which acts as a perfect surrogate for the terminal portion of the 3'-processing vDNA. This agent occupies the donor-binding site, allowing the acceptor-binding site to be available for docking studies (Savarino, 2007). Currently, it is not certain if the difference in the binding mechanism is correct, as there are several different publications on the molecular docking studies of known inhibitors, suggesting either the viral cDNA-binding site or the host DNA-binding site is as the active target (Savarino, 2007; DeAnda *et al*., 2013; Li *et al.*, 2015). Techniques applied to anti-HIV drug research can be achieved through two steps; 1) designing stereoisomers of a known active pharmacophore, which confer biological activity in the active site of the protein. 2) Scoring these conformations on enthalpy, to develop the closest fit (Meng *et al*., 2011). With these two perspectives, computational docking can be achieved and thus provide a prediction of the 3D ligand-receptor complexes.

3.1.2 *Specific aims and objectives*

The aim of this sub-project is to compare visually the binding poses of the synthesised hydroxycinnamic acids and dimers for further development. To be able to concept the different poses and the preferred position of the ethyl ester chains and hydroxy / phenol groups, or whether the ethyl chain is unfavourable in all positions. Furthermore, the presence of hydrogen-bonding, metal-ligand and π-π stacking interactions will be identified, as these play a major role in the stabilisation of the binding preference between protein and ligand. Cardiff University School of Pharmacy and Pharmaceutical Sciences have provided the 3D computational modelling PBD files of the ligand and protein interactions within the acceptors vDNA active site of HIV-1 integrase.

3.2 *Experimental*

3.2.1 *General procedures*

Molecular Modelling: All molecular docking studies were performed on a Viglen Genie Intel®CoreTM i7-3770 vPro CPU@ 3.40 GHz x 8 running Ubuntu 14.04. Molecular Operating Environment (MOE) 2018.10 (2) and Maestro (Schrödinger Release 2017-1) (1) were used as molecular modelling software. The HIV-1 integrase structure in complex with 5-CITEP (as surrogate of the terminal portion of 3'-processed viral DNA) and in complex with different INSTIs inhibitors reported by

Savarino (2007) was used. The protein was pre-processed using the Schrödinger Protein Preparation Wizard by assigning bond orders, adding hydrogens and performing a restrained energy minimisation of the added hydrogens using the OPLS_2005 force field instead of the addition of water molecules. Ligand structures were built with MOE and then prepared using the Maestro LigPrep tool by energy minimising the structures (OPLS_2005 force filed), generating possible ionisation states at pH 7±2, generating tautomers and low-energy ring conformers. In the preparation of the acceptor DNA-binding site, a 13 Å docking grid (inner-box 10 Å and outer-box 23 Å) was prepared using the docked inhibitor L-731,988 as centroid and including the two Mg^{2+} atoms and the 5CITEP as part of the binding site. In the donor DNA-binding site preparation, a 13 Å docking grid (inner-box 10 Å and outer-box 23 Å) was prepared using the co-crystallised 5CITEP as centroid and including the two Mg^{2+} atoms as part of the binding site. Using the default parameters in the VEGA program, force fields and charges were assigned according to AMBER and Gasteiger algorithms, respectively, and the molecules were energy-minimized by 50 cycles of conjugate gradients (CG) Molecular docking studies were performed using Glide SP Precision, keeping the default parameters and setting 5 as number of output poses per input ligand to include in the solution. The output data were saved as mol2 files. The docking results were visually inspected for their ability to bind the active site. The PDB (Protein Data Bank) files were visualised on Chimera 1.13.1 and Discovery Studio 2019 client for ligand interactions; both programmes are available as free downloads.

3.3 *Results*

Collaborators at Cardiff University School of Pharmacy and Pharmaceutical Sciences have docked eight compounds (Figure 3-3) in a model of the catalytic HIV-1 integrase, at the putative acceptor DNA-binding site. The 5-CITEP inhibitor occupying the donor-binding site. The results were visually inspected on Chimera 1.13.1, alongside the use of Discovery Studio 2019 2D diagram highlighting the interactions between ligand and receptor. The scoring system can only be used and have a significance if the reactions between the target protein and the ligand have a known biological activity.

Key:

- MG: 504 - Mg^{2+} A
- MG: 502 - Mg^{2+} B
- 1001004 - 5-CITEP

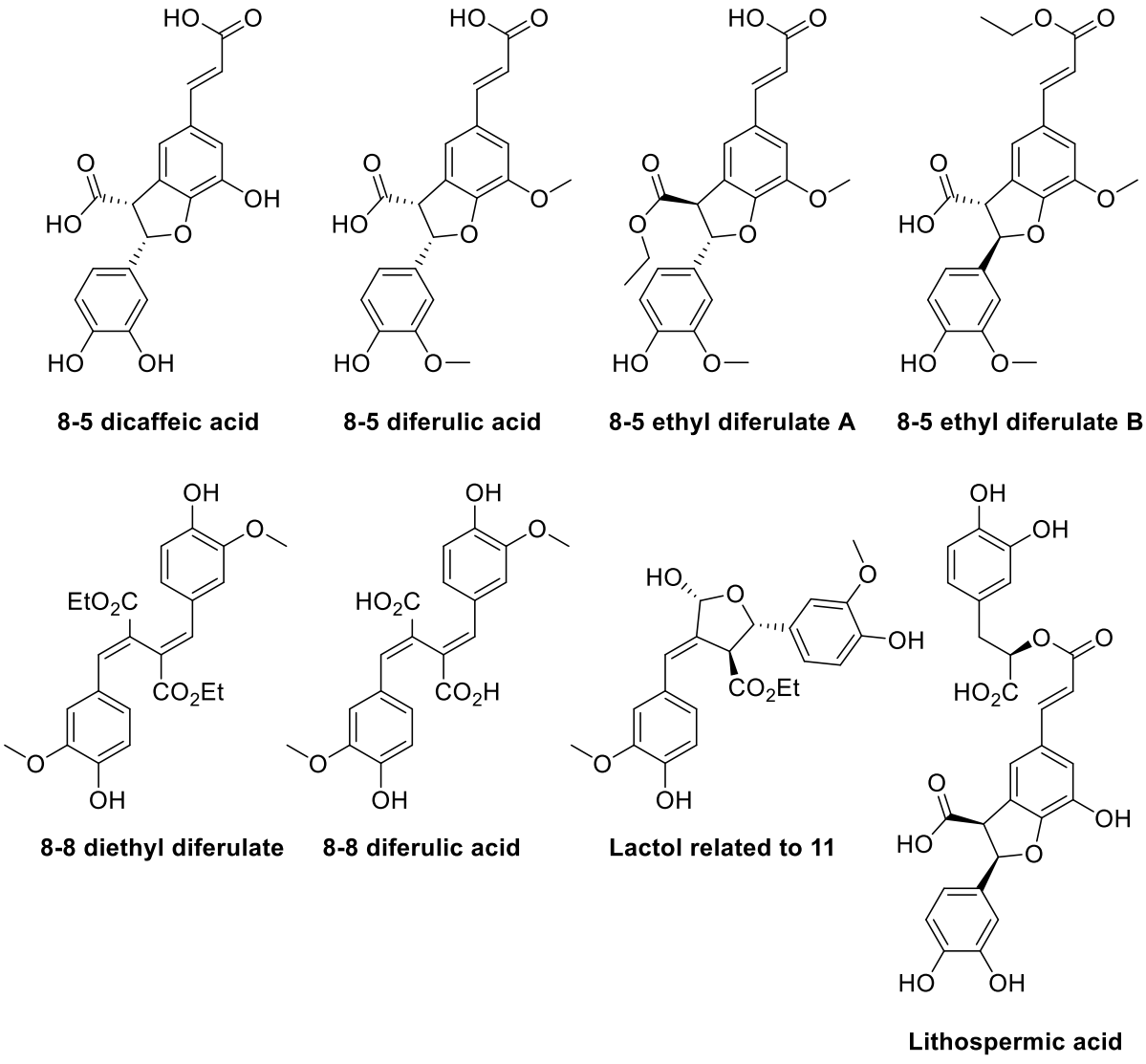


Figure 3-3: Chemical structures of potential HIV-1 integrase inhibitors. Lactol related to compound 11 from chapter 2.

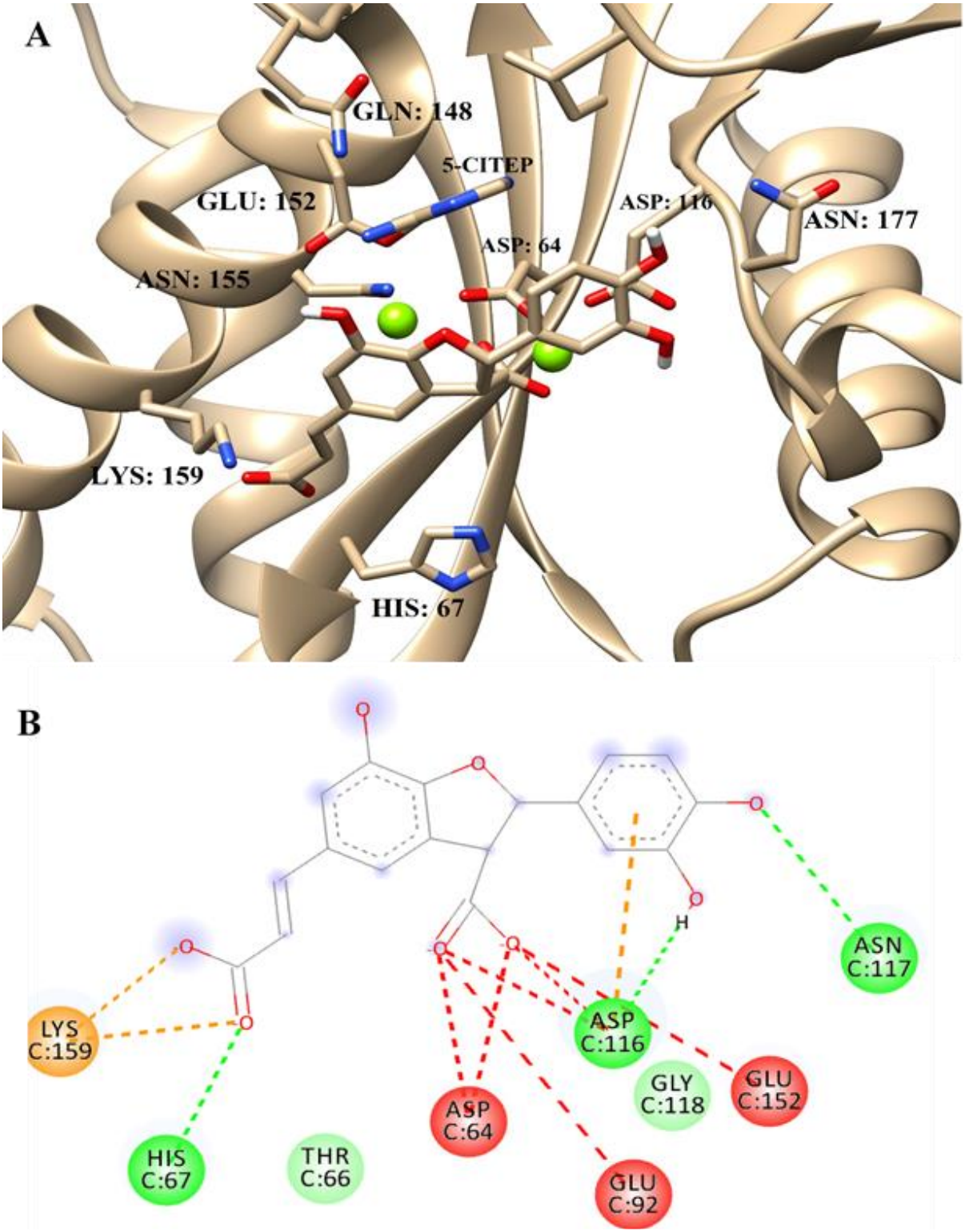


Figure 3-4: A: The docking results of HIV-1 IN binding with the ligand 8-5 dicaffeic acid is shown in gold with electronegative oxygen atoms in red. IN shown as a gold depth-cued rounded ribbon. All residues are labelled respectively. Mg^{2+} ions are represented as green spheres. **B**: 2-dimensional diagram highlighting the interactions between ligand and receptor.

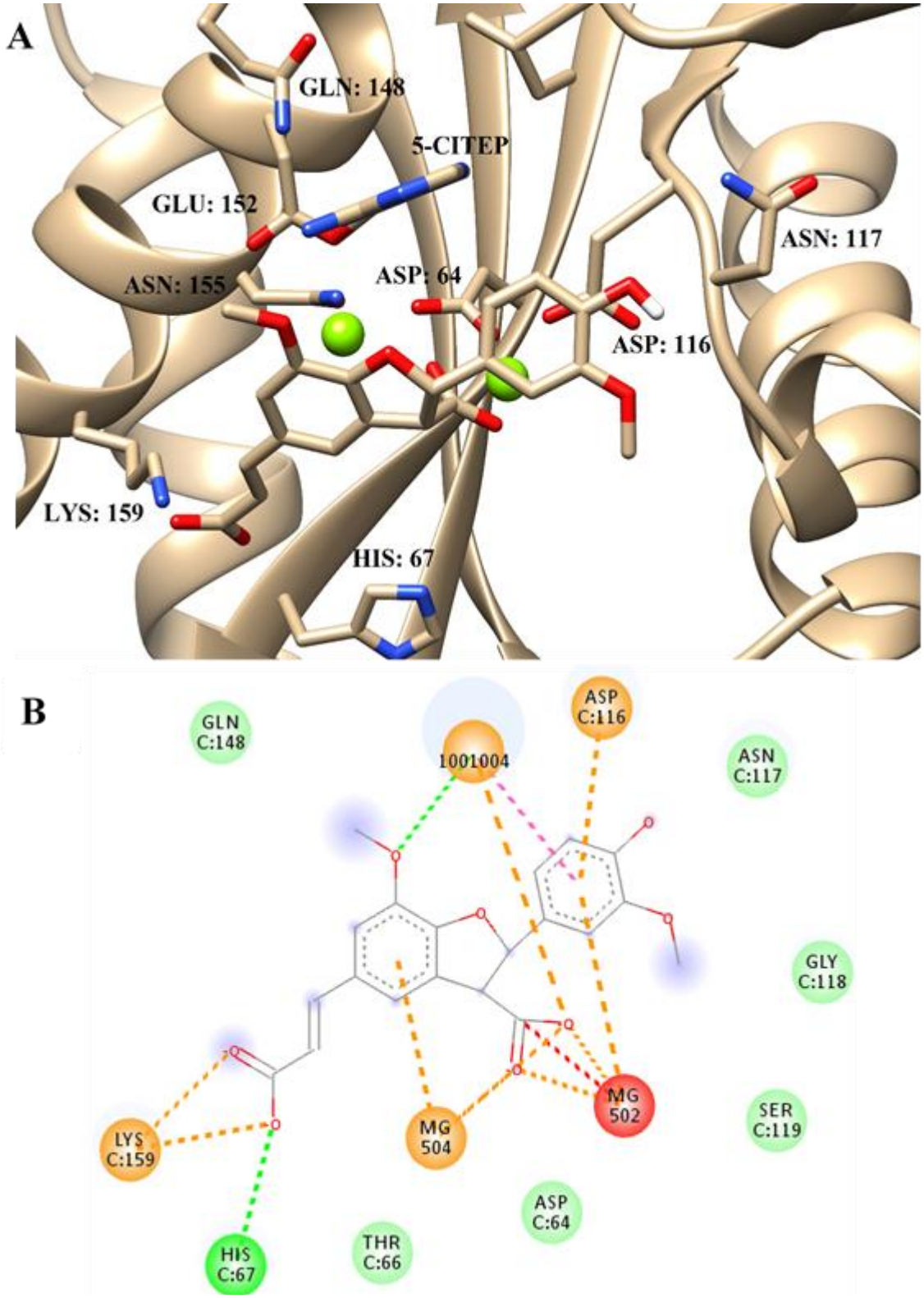


Figure 3-5: A: The docking results of HIV-1 IN binding with the ligand 8-5 diferulic acid is shown in gold with electronegative oxygen atoms in red. IN shown as a gold depth-cued rounded ribbon. All residues are labelled respectively. Mg^{2+} ions are represented as green spheres. **B**: 2-dimensional diagram highlighting the interactions between ligand and receptor.

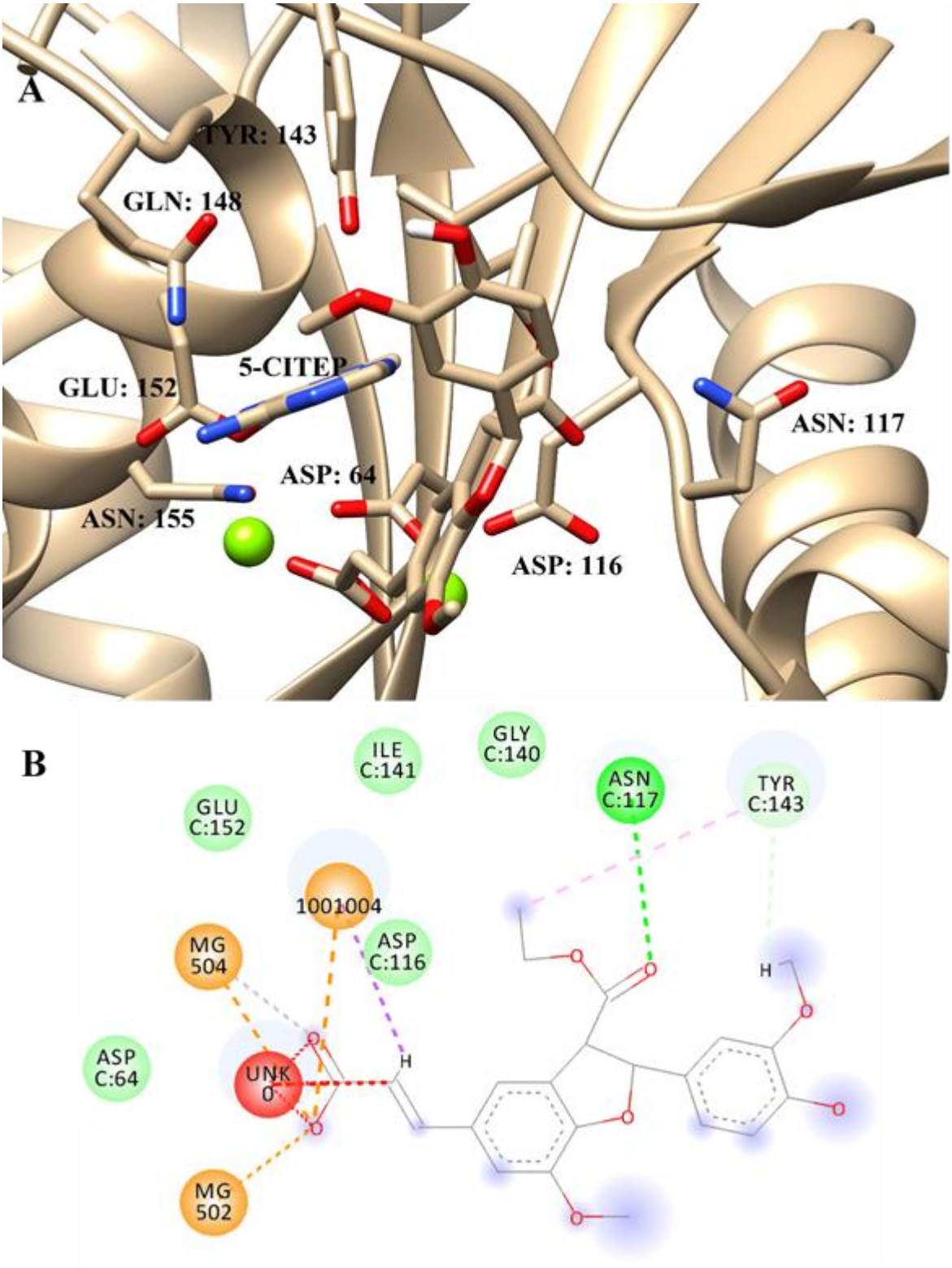


Figure 3-6: A: The docking results of HIV-1 IN binding with the ligand 8-5 ethyl diferulate A is shown in gold with electronegative oxygen atoms in red. IN shown as a gold depth-cued rounded ribbon. All residues are labelled respectively. Mg^{2+} ions are represented as green spheres. **B**: 2-dimensional diagram highlighting the interactions between ligand and receptor.

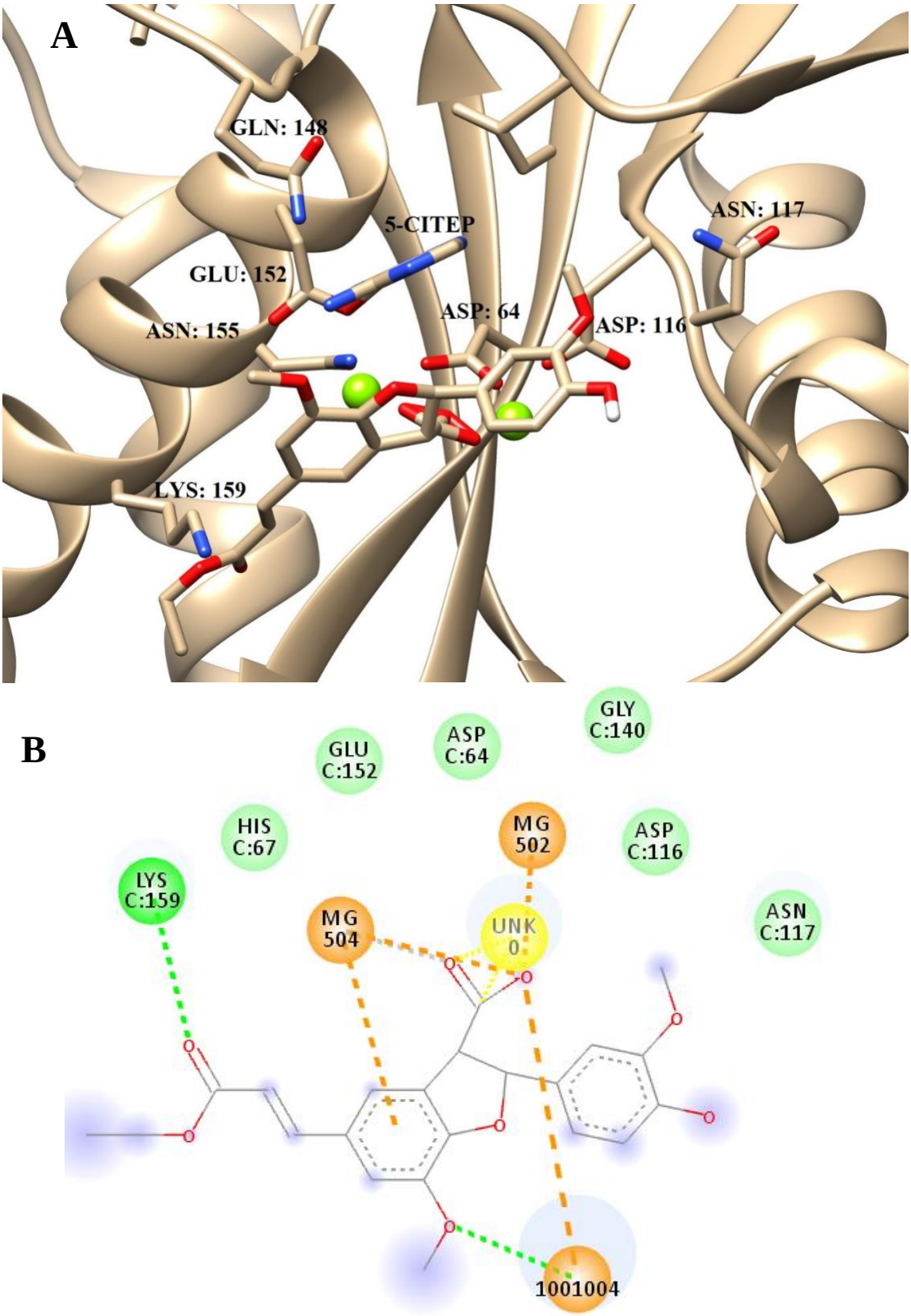


Figure 3-7: A: The docking results of HIV-1 IN binding with the ligand 8-5 ethyl diferulate B is shown in gold with electronegative oxygen atoms in red, with the ethyl ester forming at the top of compound. IN shown as a gold depth-cued rounded ribbon. All residues are labelled respectively. Mg^{2+} ions are represented at green spheres. **B:** 2-dimentional diagram highlighting the interactions between ligand and receptor.

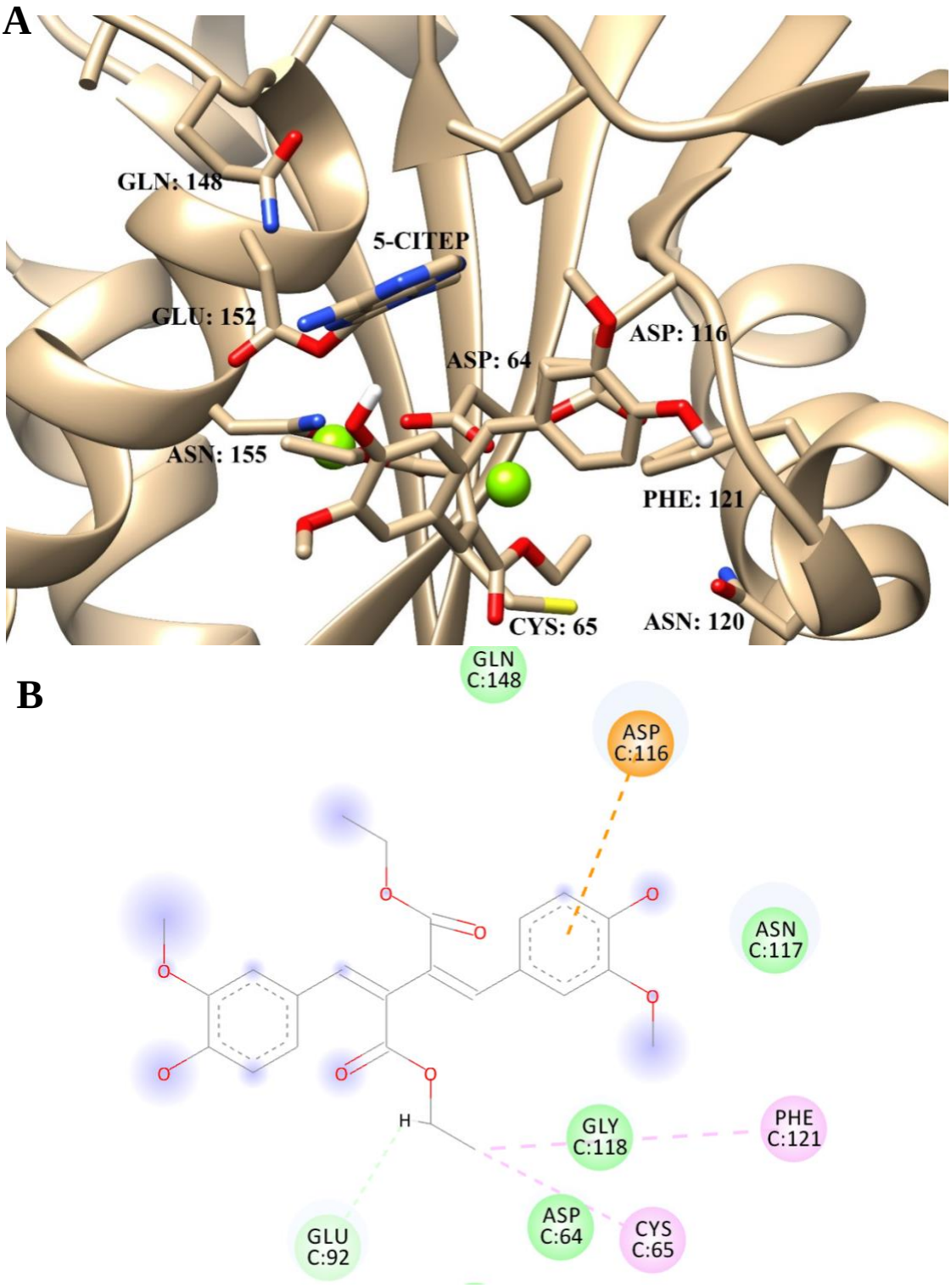


Figure 3-8: A: The docking results of HIV-1 IN binding with the ligand 8-8 diethyl diferulate, is shown in gold with electronegative oxygen atoms in red. IN shown as a gold depth-cued rounded ribbon. All residues are labelled respectively. Mg^{2+} ions are represented at green spheres. **B:** 2-dimentional diagram highlighting the interactions between ligand and receptor.

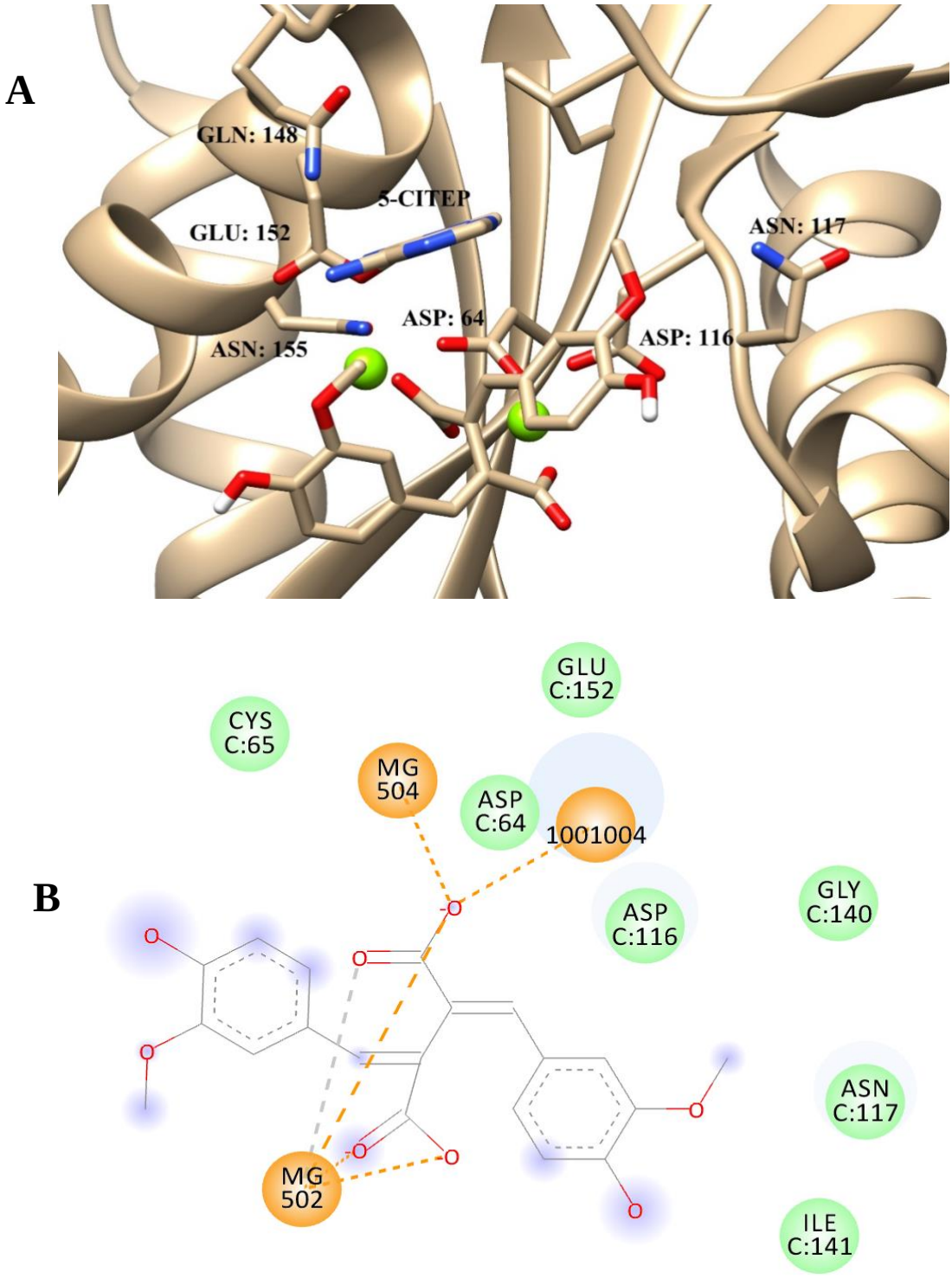


Figure 3-9: A: The docking results of HIV-1 IN binding with the ligand 8-8 diferulic acid, shown in gold with electronegative oxygen atoms in red. IN shown as a gold depth-cued rounded ribbon. All residues are labelled respectively. Mg^{2+} ions are represented at green spheres. **B:** 2-dimentional diagram highlighting the interactions between ligand and receptor.

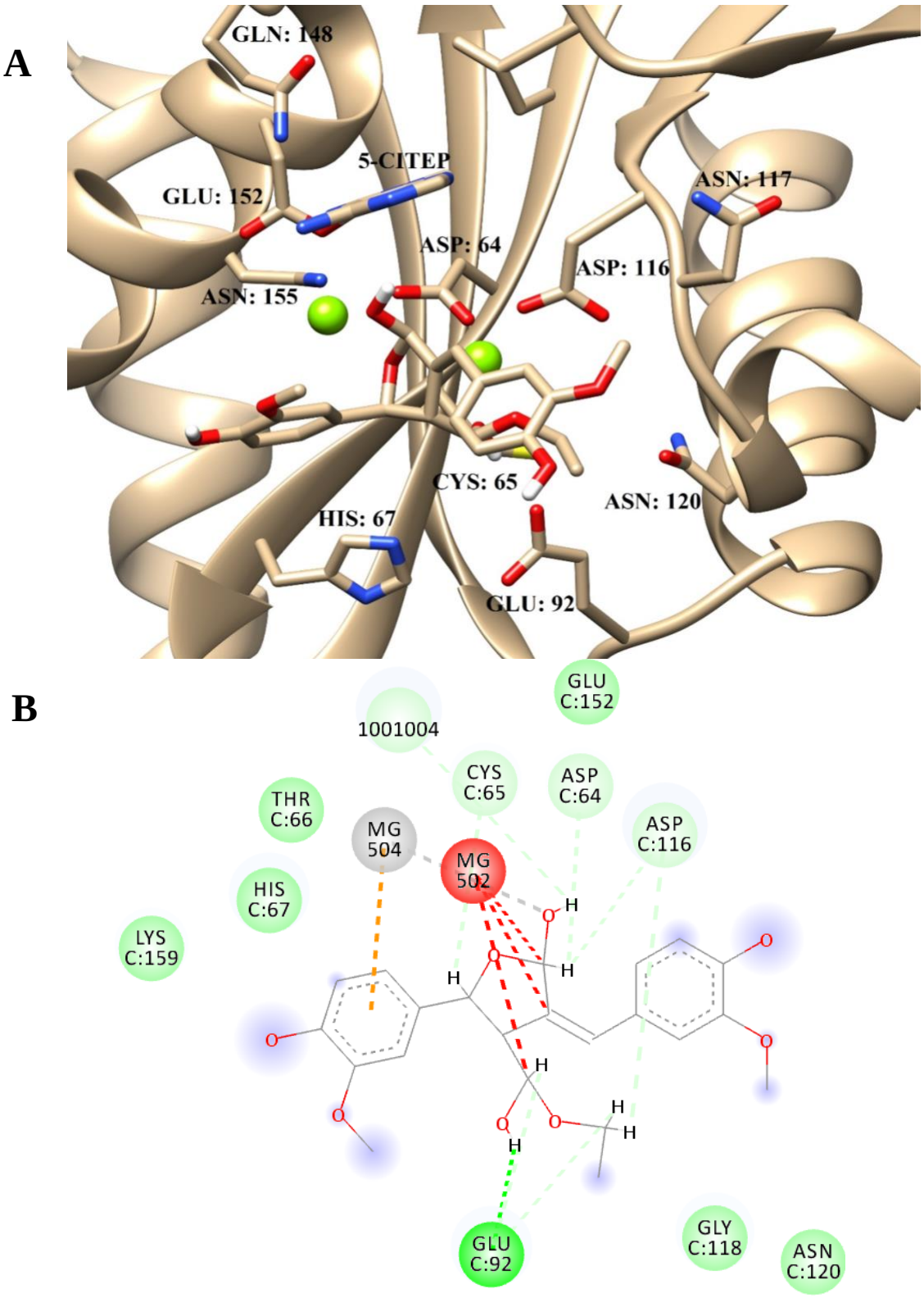


Figure 3-10: A: The docking results of HIV-1 IN binding lactol derived from compound **11** (Chapter 2), shown in gold with electronegative oxygen atoms shown in red. IN shown as a gold depth-cued rounded ribbon. All residues are labelled respectively. Mg^{2+} ions are represented at green spheres. **B:** 2-dimentional diagram highlighting the interactions between ligand and receptor.

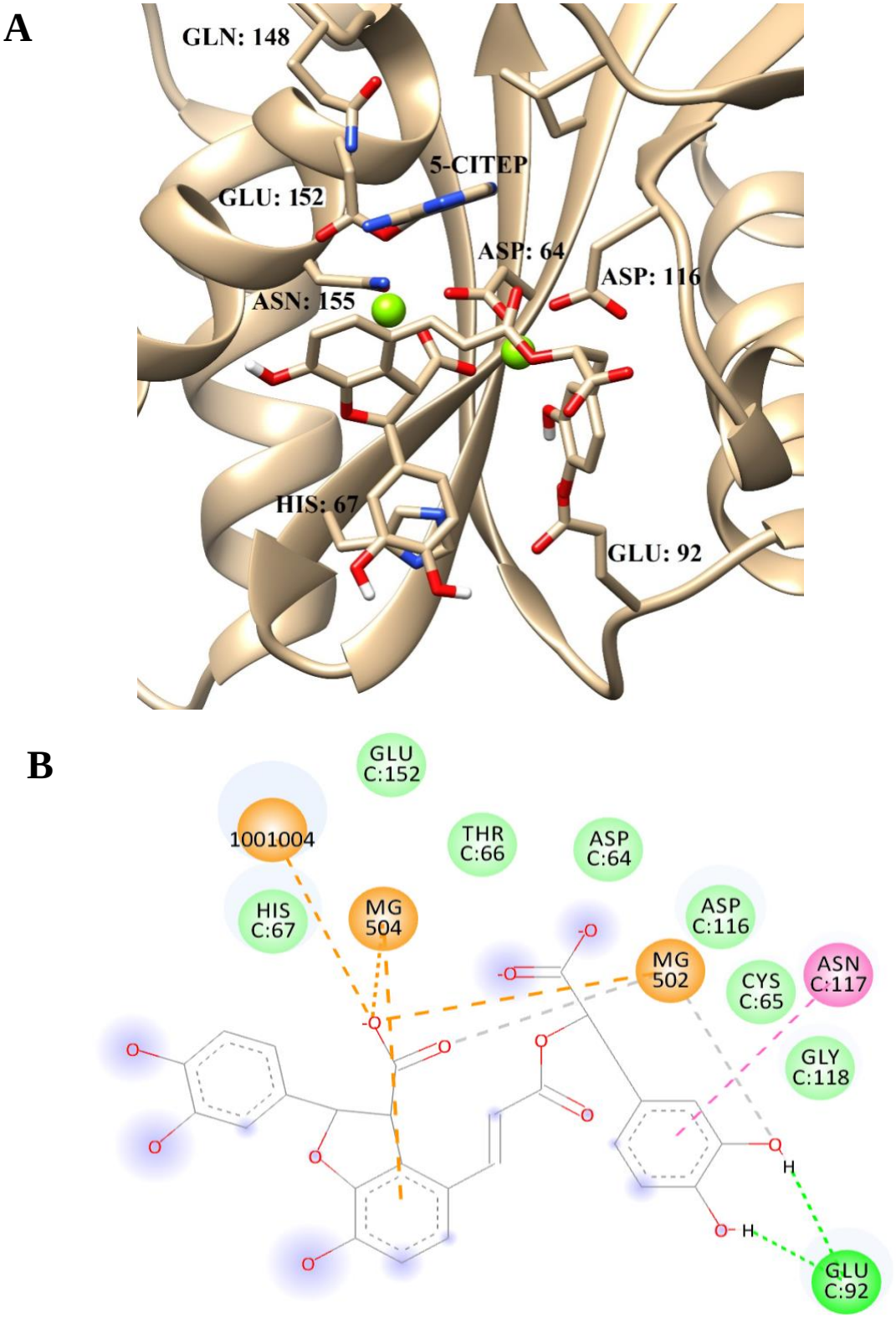


Figure 3-11: A: The docking results of HIV-1 IN binding with the ligand lithospermic acid, shown in gold with electronegative oxygen atoms shown in red. IN shown as a gold depth-cued rounded ribbon. All residues are labelled respectively. Mg^{2+} ions are represented at green spheres. **B:** 2-dimentional diagram highlighting the interactions between ligand and receptor.

3.4 *Discussion*

The docking results show eight different ligands and their interactions within the catalytic core of HIV-1 integrase putative acceptor DNA-binding site. The charges of the atoms on the structures are consistent with that of Savarino (2007), were the ligand structures were built with MOE and then prepared using the Maestro LigPrep tool by energy minimising the structures, generating possible ionisation states at pH 7±2, generating tautomers and low-energy ring conformers. They are all located near helix α-4 and are folded into the pocket created by residues Lys_{156}, Lys_{159}, His_{67}, Glu_{152}, Asp_{64}, Asp_{116}, Phe_{143} and Gln_{148}. Within the active core, there are three key amino-acid residues, Asp_{64} and Asp_{116} which bind to a Mg^{2+} cation, and Glu_{125} (Figure 3-12) (Drelich *et al.*,1992), the other remaining amino acid residues will form hydrogen bonds, metal-ligand and π-π stacking interactions (Nunthaboot *et al.*, 2013).

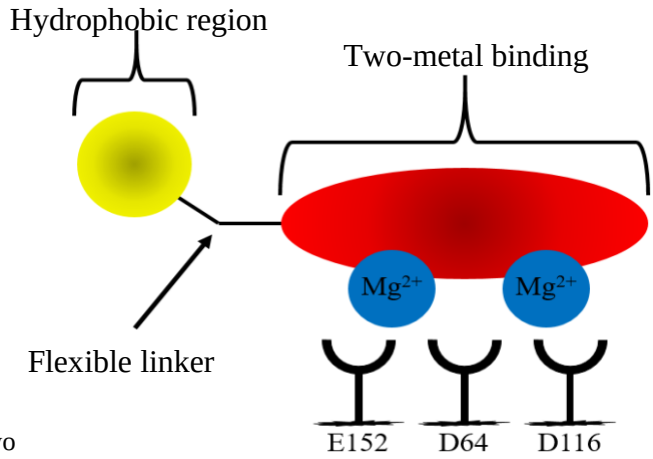


Figure 3-12: The two ıd Asp_{116}, illustrating the binding pharmacophore of IN inhibitors.

3.4.1 *Diacids*

Ligand 8-5 dicaffeic acid (Figure 3-4) showed electrostatic interaction from the terminal carboxylate anion to the positively charge Lys_{159}, with a conventional hydrogen bond from the side-chain N-H of Asn_{117} forming at the opposite side of the ligand. Furthermore, a π-cation and a hydrogen bond was observed between Asp_{116} and ring C. One of the three desired amino acid residues have favourable interactions. Exchanges between Asp_{64} and Glu_{152} have been observed as an unfavourable negative-negative repelling interactions with the carboxylic acid from the furan ring, this is illustrated by the red dotted lines (Figure 3-4: B). The interaction is forcing the ligand away from the active site, the potential reason why the two Mg^{2+} were not observed

with the Figure 3-4: B. Ligand 8-5 diferulic acid (Figure 3-5) binds in a very similar way to that of dicaffeic, with the electrostatic interaction from the terminal carboxylate anion to the positively charged Lys_{159} and hydrogen bonding between the carboxylate anion to His_{67}, and Asp_{116} again creates the π-cation on ring C. This docking simulation has recorded interactions between the two Mg^{2+} ions, with a π-cation interaction between the electron-rich ring A and Mg^{2+} A. The same Mg^{2+} is also ligated electrostatically to the carboxylate anions of Glu_{152} and Asp_{64}, two of the three required amino-acid residues. The central carboxylate of the ligand is bound electrostatically to Mg^{2+} B. The ion is also ligated to the carboxylate of Asp_{64} and Asp_{116}. Ligand 8-8 diferulic acid (Figure 3-9) binds in a very different pose. Dimerising two ferulic acid units on carbon-8 creates the two carboxylates within the centre of the ligand. Possible van der Waals interactions have been observed between the three essential amino acid residues. The two central carboxylates of the ligand have bound electrostatically to the Mg^{2+} ion A and B, thus using the similar β-hydroxy carbonyl motif for binding, this pose allows both carboxylic acids to interact with the two Mg^{2+} ions. Furthermore, an electrostatic interaction has been observed between the ligand and the 5-CITEP within the donor-binding site. In addition, no interactions were observed with the ligand and amino acids (Figure 3-9: B). Additionally the 8-8 diethyl diferulate was also docked; this compound also had a very different pose within the binding site. The observed π-anion interaction between Asp_{116} and the electron-rich ring A. Glu_{92} forms to a carbon hydrogen bond from the CH_2 of the ethyl ester. The methyl forms π-alkyl interactions with Cys_{65} and Phe_{121}. It seems the ethyl chains are inhibiting the electronegative ester away from Mg^{2+} interactions, as the Mg^{2+} are not present within the 2D model (Figure 3-9: B).

3.4.2 *Ethyl and diethyl dimers*

8-5 ethyl diferulate A (Figure 3-6) has a very different pose, whereby the compound only interacts with the core with the carboxylic anion on ring A, instead of the proposed binding pose of the Mg^{2+} interacting within the centre of the ligand. Furthermore, hydrogen bonds were observed from the side-chain OH of Tyr_{143} to the methoxy oxygen attached to ring C and from the side chain of N-H of Asn_{117} to the carboxylate attached to the furan ring. The carboxylate anion attached to the alkene is very close to the carboxylate of Asp_{64}. This may seem to be electrostatically repulsive but perhaps there is a bridging proton or water. The Asp_{116} carboxylate is also very close; this may also indicate the bridging proton of a bridging water molecule. Moreover, the ethyl of the ethyl ester has formed a π-alkyl interaction with Tyr_{143}, from the picture the distance looks significant, but the interactions has been labelled with the purple dotted line from the ethyl ester to Tyr_{143}.

8-5 ethyl diferulate B (Figure 3-7) has the ethyl ester present at the top of the molecule. This has changed the interaction with the catalytic site. A hydrogen bond is shown from the terminal ester to the side-chain $N^{+}H_3$ of Lys_{159}. Also there is an attractive electrostatic interaction between the central carboxylate and 5-CITEP within the donor-binding site and potential water-bridged hydrogen bonds from the carboxylate ion to Asp_{64} and Asp_{116}.

Lactol derived compound **11** identified within chapter 1 has also been docked to identify its binding pose (Figure 3-10). This compound coordinated very well with the two Mg^{2+} ions, a π-cation interaction between the electron-rich ring A and Mg^{2+} A has been observed, alongside the co-ordination between Mg^{2+} and the central ester carbonyl oxygen. Hydrogen bonds have been observed with the ethyl ester with the amino-acid residue Glu_{92}. Only one pose was sent returned to Aberystwyth University, the binding of lactol was very variable with different binding poses found, further modelling and structure activity relationship studies are required, (Personal communication, Dr. Salvatore Ferla). The lactol derived compound **11** has interactions with two (Asp_{64}, Asp_{116}) of the three amino acids residues for reliable binding.

The drug portfolio within the current work has been structurally based on the known HIV-1 inhibitor lithospermic acid and lithospermic acid B. These two polyphenolic structures inhibited with IC_{50} values in the range of 0.37-0.83 μM, strongly

suppressing serious HIV-1 infections in H9 cells (IC_{50} 6.9 µM). The compounds were also reported non-cytotoxic towards human cells at high concentrations (CC_{100} > 297 µM). In computational studies of lithospermic acid and lithospermic acid B against wild type and mutated strains, almost all major interactions between the amino-acids of IN and the ligand were preserved. This current study also simulated lithospermic acid for visual comparison on the docking pose (Figure 3-11). Hydrogen bonds were observed between the hydroxy-groups on ring D and Glu_{92}. Electrostatic interaction from the central carboxylate anion to the positively charged 5-CITEP, suggesting the ligand is well incorporated within the active site. Furthermore a π-cation interaction between the electron-rich ring C and Mg^{2+} A was observed. Besides, the negative metal acceptor interactions between Mg^{2+} and the furan lactol and the ethyl ester. Furthermore, an amide-π stacked interaction between Asn_{117} and the electro rich ring C. Hydrogen bonds between residues His_{67}, Asp_{64} and Asp_{116}, forming from the carboxylate anion and ester of ring C, are also consistent with those reported in the literature (Zhu *et al.*, 2004; Nunthabot *et al.*, 2013). Multiple van der Waals interactions have also been observed with the remaining amino-acid residues Phe_{121}, Gly_{118}, Cys_{65} and Thr_{66}.

3.5 *Conclusion*

In conclusion, all the ligands were docked into the acceptor DNA-binding site compounds; 8-5 dicaffeic acid (Figure 3-4), 8-5 diferulic acid (Figure 3-5), 8-5 ethyl diferulate B (Figure 3-7), 8-8 diethyl diferulate (Figure 3-8) and 8-8-diferulic acid (Figure 3-9) seem the most promising as potential integrase inhibitors. Lithospermic acid (Figure 3-11) binding poses was consistent with literature (Zhu *et al.*, 2004; Nunthabot *et al.*, 2013). Further work and modelling needs to be conducted to fully understand the binding mechanics of lactol pk1 (Figure 3-10), thus the rotatable bonds during the docking experiments need to investigated in more detail.

From docking, it seems that the ester group on the tetrahydrofuran ring is forcing the compound far from the active site (8-5 ethyl diferulate, Figure 3-6) and this might result in a non-inhibition of integrase. In the docking simulation, the lateral acid group is co-ordinating the Mg^{2+} but research suggests that the acid group on the tetrahydrofuran should do the co-ordination, while the lateral chain acid should interact in a similar manner of the reported inhibitor L-731,988 docked by Savarino *et al.*,

(2006) (Figure 3-13). Furthermore, when the ester is in the lateral chain and there is the free acid on the tetrahydrofuran ring, this co-ordination seems still possible. The particular stereoisomers were based on the structural similarity they possess with lithospermic acid, a known IN. The different orientation of the ethyl ester was investigated to understand which stereoisomer would be suitable for further investigation for an IN inhibitor. The lactol was modelled rather than the lactone compound **11** from chapter 2, due to the potential increased hydrogen bonding with Glu_{92}. Thus creating a similar structure to that of the β-hydroxy carbonyl motif (Drelich *et al*., 1992).

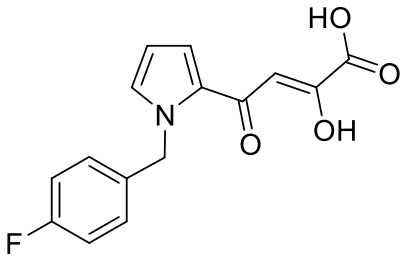


Figure 3-13: Chemical structure of L-731,988, reported inhibitor (Savarino *et al*., 2006). Compound highlighting the β-hydroxy carbonyl motif required for successful inhibition.

Docking of the compounds in the DNA donor-binding site was also performed. The other compounds above do not dock well in the site, as they occupy the binding site in a different manner of the co-crystallized 5-CITEP and therefore interactions with the Mg^{2+} ion are not always present. In conclusion, 8-5 diferulic acid, 8-5 dicaffeic acid, 8-8 diferulate acid and 8-8 diethyl diferulate look more promising.

Chapter 4

The identification of two novel homoisoflavonoids with anti-cancer properties from an undisclosed plant

4.0 *Abstract*

Breast cancer is the most common type of cancer affecting women worldwide, but also affects men. Annually there are approximately 55,000 new cases within women and 350 in men per year in the United Kingdom, with estimations of 1,884,213 new cases of breast cancer occurring in females worldwide by 2020. As the heterogeneous disease develops and evolves, new treatment systems are needed. Natural products, with particular emphasis on phenolic compounds, are an effective means of treating the disease. With limited research into the biological activities of homoisoflavonoids, this study aims to identify, purify and structurally elucidate two novel homoisoflavonoids with known potent cytostatic activities. HPLC-grade methanol was used to extract the polar compounds from the plant roots. Subsequently, the extract was analysed using HPLC-UV-ESI-MS/MS to identify two *pseudo*molecular ions at *m/z* 311 and *m/z* 341, with the characteristic fragmentation of homoisoflavonoids. Further biologically important compounds were identified within the leaf extractions, demonstrating the usefulness of natural products as essential drug discovery leads.

4.1 *Introduction*

Breast cancer affects 2.1 million women worldwide each year. A report by the World Health Organisation (WHO) highlighted that approximately 15% of all cancer-related deaths among women are caused by breast cancer, with an estimation of 627,000 women in 2015 (GLOBOCAN, 2015), with the highest incidence rate seen in economically developed countries (DeSantis *et al*., 2015). Breast cancer is the most common cancer in the United Kingdom, with an estimated 55,000 new cases within women and 350 in men per year, equating to 15% of all cancers diagnosed (CRUK, 2014).

Breast cancer has multiple different reported origins, with multiple different factors, such as hereditary (mutations) and hormonal changes; environmental stimuli including ionising radiation, lifestyle and cancerous agents (Coyle, 2004; Anand *et al*., 2008). Mutations in the DNA of a breast cell take place as a result to exposure to factors

detailed above. Small DNA errors, such as single bases and if repaired are harmless, if not repaired it will cause a pathway malfunction resulting in cell death. If a series of errors occur, thus providing a cell growth advantage over other cells, the dividing cells form a cluster which is termed as a tumour (Alberts *et al*., 2002). One of the most common inherited mutations of *BRCA1* or *BRCA2* have a 50-90% lifetime risk of cancer (Royal Marsden NHS Trust, 2013). Breast cancer has been classed as a heterogeneous disease, due to the varied tumour cells with distinct morphological and phenotypic profiles, these include; gene expression, metastatic potential, metabolism, proliferation and motility (Turashvili and Brogi, 2017). With various morphological and phenotypic profiles of breast cancer, consequently creating multiple traits for detection. Also each phenotypic profile may possess a unique clinical response to treatment (Weigelt, Geyer and Reis-Filho, 2010).

4.1.1 *Current treatments*

The current treatments available on the NHS include surgery, radiotherapy, chemotherapy, hormone therapy, individual biological target therapy, immunotherapy and palliative care (Dhankhar *et al*., 2010). Surgery is the leading approach to treatment of localised breast cancer (Matsen, 2013). Radiation therapy usually follows after surgery, to expose the remaining cancer cells to high levels of radiation (Bomford and Kunkler, 2002). The tumour shrinks with the added combination of chemotherapy (Lind, 2008). The treatments have been reported to cause serious side effects, further bacterial infection, and decreased sensation in the breast tissue and under the arm; skin problems within the treated area; alopecia, fatigue and nausea (Janssen *et al*., 2017; Werbel and Cohen, 2018). Even with the complications of side effects, the 5- year survival rate for female breast cancer is 90%. Even with the large incidence rate, it has become one of the highest survival rates, with prostate and melanoma of the skin with 99% and 92% respective survival rates if the cancer has been identified at a very early stage. Pancreatic, lung and liver cancer have an 8%, 18% and 18% survival rates, respectively (Siegel *et al*., 2018).

As the heterogeneous disease develops and evolves, new treatment systems are needed. Research has identified the current methods such as chemotherapy has its restrictions, such as the toxic effects on non-targeted cells, resulting in damaging human health (Ochwang *et al*., 2014), thus driving the demand for alternative

treatments with naturally derived anticancer actives, with plants being the ultimate source. Natural products have gained increasing attention in cancer chemotherapy as literature has reported, with the potential to be less toxic, thus having reduced side effects (Mishra and Tiwari, 2011). Many recent papers has identified natural product-derived anticancer drugs possess alternative modes of action, that achieve affective cell death (Gali-Muhtasib *et al*,. 2015; Khalid *et al*., 2016). With this knowledge researchers, such as Kotoku *et al*. (2016) and Bernardini *et al*. (2017), are investigating the clinical potential of these natural products from medicinal plants, to aid the pharmaceutical industry. Between 1940 and 2014, the use of natural products in the treatment of cancer has increased tremendously. Newman and Cragg, (2012) analysed that, 48.6% of the 175 small molecules isolated and derived from nature, and have been approved for cancer chemotherapy (Newman and Cragg, 2012).

4.1.2 *Secondary metabolites from plants as anticancer agents*

Plants have been a rich source of inexpensive natural compounds, secondary metabolites. The secondary metabolites may exhibit an extensive range of biological activities, with structural complexity where synthesis is difficult or not accomplished yet. They are mostly small organic molecules, which are produced by the organism, which are not essential to sustain life. These compounds include phenolics, such as polyphenols, flavonoids and coumarins, alkaloids and terpenoids. With new secondary metabolites being isolated and structurally elucidated every year, therefore creating an established and continuous source of new structures with the possibilities to analyse and control the development of cancer and many other diseases. Although some secondary metabolites have a novel anticancer mechanism, but are unable to be approved for clinical therapy, due to limited bioactivity or toxicity (Seca and Pinto, 2018). Secondary metabolites can be used for leads for drug development. Modifications to the chemical structure is a resourceful way to increase absorption, selectivity, metabolism and anticancer activity and decrease potential toxicity (Seca and Pinto, 2018).

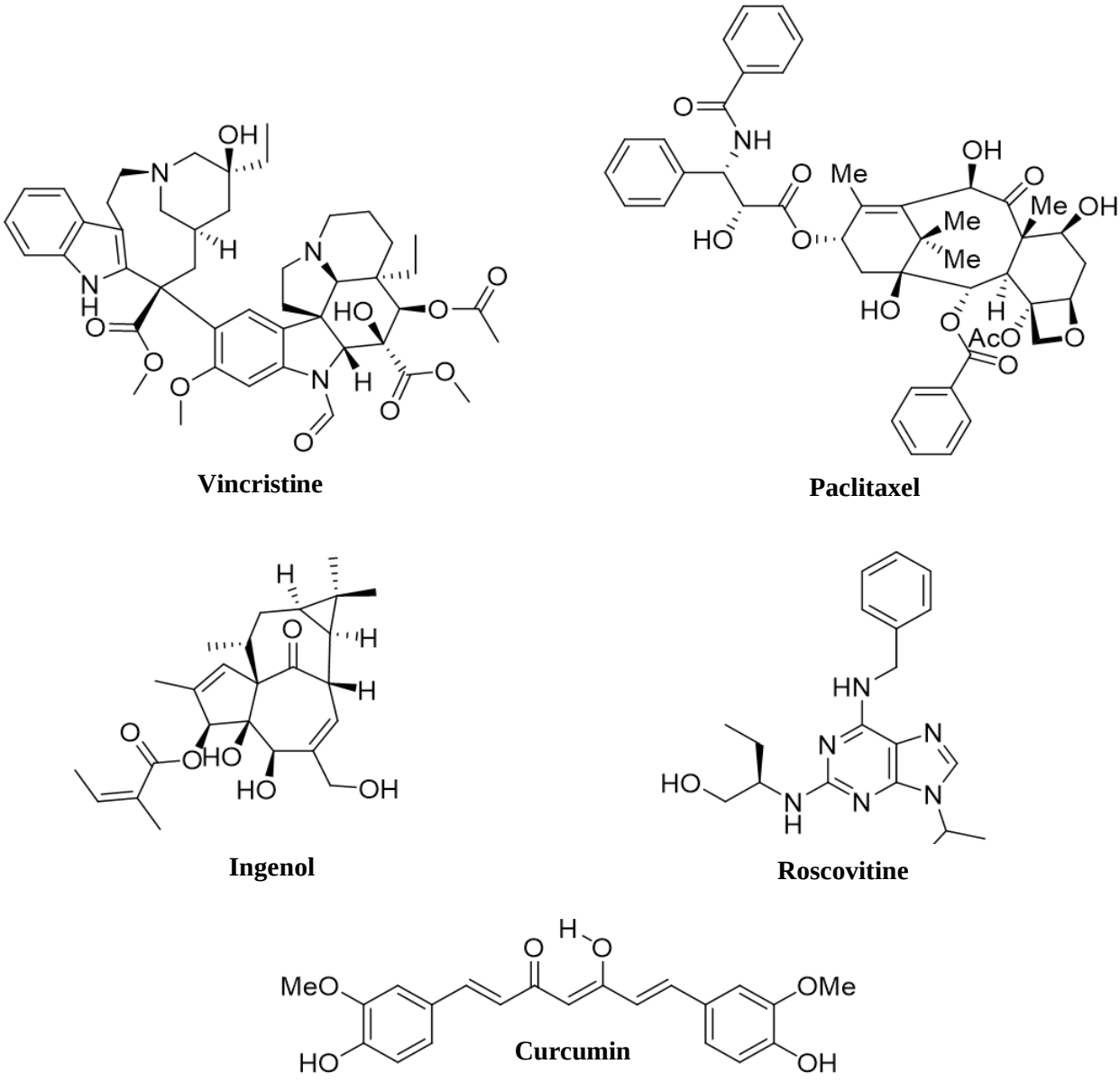


Figure 4-1: Vincristine; A natural product extracted from *Catharanthus roseus* with anticancer activity, which represses cell growth by altering the microtubular dynamics. Paclitaxel; a natural microtubule inhibitor extracted from *Taxus brevifolia*. Ingenol mebutate, extracted from *Euphorbia peplus* L. Roscovitine Curcumin.

Vincristine (Figure 4-1) is a naturally occurring alkaloid, extracted from the leaves from *Catharanthus roseus.* It was one of the first FDA-approved plant-derived anticancer agents (Newman and Cragg, 2016). It has been used in adult chemotherapy it has a more common use in paediatric oncology. When incorporated in a treatment system for acute leukaemia in children, the survival rate increases to 80% (Evans *et* al., 1963). Vincristine works in a concentration-dependent manner, interacting with the tubulin, inhibiting polymerisation of the mitotic spindle microtubules. It thereby disrupts the assembly of the mitotic spindle and halts the cell from dividing (Wang *et*

al., 2016). A recent report by Yoshihara *et al*., (2017) highlighted that combined chemotherapy with drugs demonstrating different mechanisms not only increased the activity of inhibition of the tumour but also decreased drug resistance and toxicity. Therefore, clinical trials are in progress with the combination of vincristine and dactinomycin (ClinicalTrials.gov Identifier: NCT02879643 and NCT01527149) (Yoshihara *et al*., 2017).

The discovery of paclitaxel (Figure 4-1) came with a new significant mode of action, of the tubulin-assembly. Paclitaxel was extracted in 1962 from the bark of *Taxus brevifolia* (Pacific Yew), as part of an initiative conducted the National Cancer Institute (NCI) to screen plants for anticancer therapy (Cragg, 1998). Issues with the supply and severe problems with formulating into a deliverable form causes a two decade gap before it was entered into clinical trials (Walsh and Goodman, 2002). Paclitaxel is the only clinically used drug that was discovered using plant-screening programmes between 1960 and 1981 (Walsh and Goodman, 1999) Since 1993, it has been sold under the name of Taxol® and it has become one of the most active cancer chemotherapy drugs (Weaver, 2014; Bernabeu *et al*., 2017). The complex structure of paclitaxel is a diterpenoid, that contains a complex 6,8,6-tricycle-fused skeleton, linked to a four member oxetane ring and it has amide, ketone, ester and alcohol functional groups (Figure 4-1). These functional groups are at unique positions that ensure that β-tubulin is inhibited specifically, thus preventing proper mitotic spindle assembly and chromosome segregations during cell division (Weaver, 2014). The success of paclitaxel is due to the effectiveness on different stages of tumours, alongside a broad spectrum of antitumour activities. This is down to the novel mechanism of action that targets cancer phenotype like cell proliferation (Weaver, 2014). Furthermore, paclitaxel is already a success within the pharmaceutical industry for cancer therapy, but its use in current in clinical trials for other treatments such as inhibiting botulinum neurotoxin (Dadgar, Ramjan, Floriano, 2013) and psoriasis (Ehrlich *et al*., 2004).

Ingenol mebutate (Figure 4-1) is a diterpene that had been extracted from *Euphorbia peplus* L. and was later identified as an active antitumour compound (Ramsay *et al.*, 2011). In a recent I/II clinical trial, sap from *E. peplus* was effective against human non-melanoma skin cancer (Ogbourne, and Parsons, 2014). The biological activity of ingenol is pH-dependent. Two structural characteristics are required for the anticancer

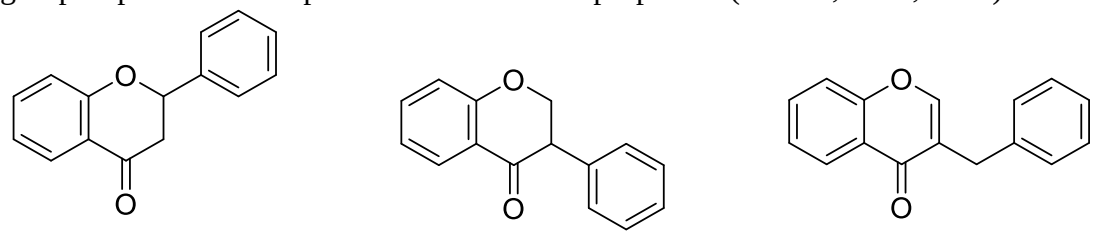
activity (1) the ester moiety (joining the ingene diterpene nucleus to the angelate unit) (2) the free hydroxy groups (Liang *et al*., 2013). Ingenol mebutate acts in a dose- and time- dependent manner against multiple cell lines (Benhadji *et al*., 2008: Serova *et al*., 2008). Reports have highlighted a list of adverse reactions, with moderate local skin responses, causing swelling, ulcers and flaking. Finally, it shows a safer current therapy and tolerability profile demonstrating a lack of systemic absorption (Collier *et al*., 2014; Tzogani 2014).

Roscovitine (Figure 4-1), extracted from *Raphanus sativus*, (Olomucine) is also known as CY-202 and seliciclib and is a 2,6,9-substituted purine analogue. The small molecule inhibits cyclin-dependent kinases (CDK1, CDK2, CDK5 and CDK7) (Cicenas *et al*., 2015), with direct competition at the ATP-binding sites. Roscovitine is currently being evaluated as a potential drug to treat cancers, inflammation, viral infections, neurodegenerative diseases and polycystic kidney disease (Cicenas *et al.*, 2015). Roscovitine is currently in phase II clinical trials for breast cancer (Foell *et al*., 2008; Slovackova *et al*., 2012).

Curcumin (Figure 4-1) is a polyphenolic compound that has been extracted from *Curcuma longa* L. (turmeric). *C. longa* has been used in Chinese and Indian traditional medicines for centenaries (Kocaadam and Sanlier, 2017), for a wide range of biological activities including anti-inflammatory, antioxidant, chemopreventive, chemotherapeutic and chemo-sensitising activity (Kocaadam and Sanlier, 2017). It has also been reported that curcumin is poorly absorbed from the gut of mice, with a chemopreventive potency indicates tentatively that the daily dose of 1.6 g of curcumin is required for efficacy in human (Perkins *et al*., 2002). It has been reported that curcumins anti-cancer properties work through the induction of cancer cell apoptosis and inhibiting the expression of fatty acid synthase (FAS) (Fan *et al*., 2016). Multiple reports have shown that curcumin can modulate a variety of cancer-related targets or pathways (Pavan *et al*., 2016; Imran *et al*., 2016; Kumar *et al*., 2016 and Kunnumakkara *et al.*, 2017) including the modulation of cytochromes P450 (CYP) enzymes. The mode of action increases the transcription of Nrf2 (nuclear factor erythroid 2-related factor 2) *via* the mitogen-activated protein kinase (MAPK) pathway (Schwerthei *et al.*, 2017), thus inhibiting tubulin polymerisation (Ramya *et al.*, 2017).

4.1.3 *Flavonoids*

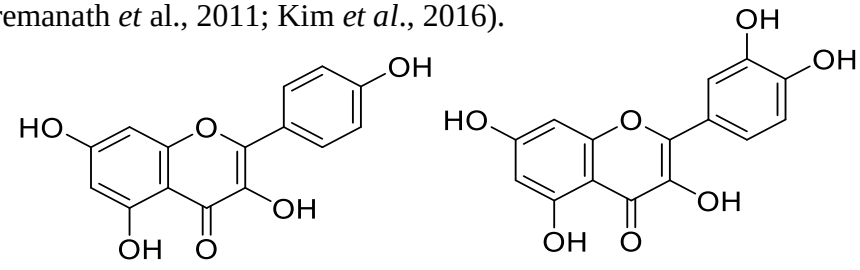
Anti-cancer properties of compounds found in the extract of natural products are often derived from phenols (Carocho and Ferreira, 2013). Phenols are an extensive class of chemical compounds that are present is all plant species; they can be classified into three major groups. (1) Cell wall phenolics; consist of hydroxycinnamic acids and lignin's (Ralph *et al*., 1994; Geissmann and Neukom, 1971). (2) Phenolic acids; consist of glycosides, amides and esters (Pereira *et al*., 2009). (3) Flavonoids; synthesised from the amino-acid phenylalanine (Havsteen, 2002). Flavonoids have been described as the most diverse amongst structures (Figure: 4-2), in providing the most important group of phenolics that possess anti-oxidative properties (Panche, *et al*., 2016).



Flavonoid **Isoflavonoid** **Homoisoflavone**

Figure 4-2: Basic structures of flavonoid, consisting of a C_{15} structure with a benzopyran ring bound to a phenyl ring on carbon-2. Isoflavonoid, consisting of a C_{15} structure with a benzopyran ring bound to a phenyl ring on carbon-3. Homoisoflavone, consisting of a C_{16} structure, with a chromone construction, C-3 of the benzopyran is bonded to a carbon-carbon linkage to a phenyl ring.

The biological function plays a fundamental role in inhibiting the growth of tumours and the proliferation of cancerous cell lines, such as MCF-7 in breast cancer (Roleira *et al*., 2015). Two natural flavonoids, kaempferol and quercetin (Figure: 4-3), isolated from *Ranunculaceae delphinium* and *Trigonella foenum-graecum*, have both been shown to have anti-proliferative activity against breast cancer cell lines (Lin *et al*., 2008; Premanath *et* al., 2011; Kim *et al*., 2016).

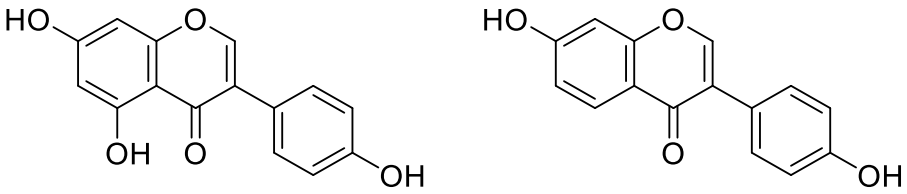


Kaempferol **Quercetin**

Figure 4-3: Chemical structures of two flavonoids kaempferol and quercetin, both possessing anti-proliferative activity against breast cancer cell lines.

4.1.4 *Isoflavonoids*

Isoflavonoids are flavonoids in where the B ring is linked to the heterocyclic ring at the C3 instead of the C2 position (Jaganath and Crozier, 2010; Jacob *et al.*, 2011). Important isoflavones are genistein and daidzein (Figure: 4-4). Both can be sourced from soy products, such as tofu and soy protein isolate. Geinistein and daidzein were observed to have a similar structure to an oestrogen steroid hormone (17β-oestradiol), therefore competing for the oestrogen receptor, contributing a favourable role in the treatment of hormone-related cancers (Banerjee *et al*., 2008). Other flavonoids have also been shown to reduce MCF-7 cell proliferation; these include 7-methoxyflavone, luteolin, apigenin, naringenin and chrysin (Le Bail *et al*., 1998).



Genistein **Daidzein**

Figure 4-4: Chemical structures of genistein and the related compound daidzein.

4.1.5 *Homoisoflavonoids*

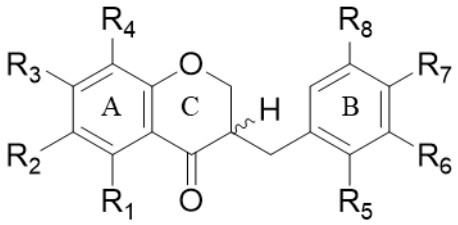
Homoisoflavonoids are a structurally similar class to flavonoids; they possess an extra carbon-carbon covalent bond, therefore made up of a 16 C-skeleton, with one heterocyclic ring, and two phenyl rings (Lin *et al* .,2014) (Figure 4-3). Currently, there are approximately 250 naturally occurring homoisoflavonoids that have been reported within literature (Lin *et al*., 2014; Zhou *et al*., 2015). An extensive range of biological actives has been reported in literature, including anti-oxidant, anti-mutagenic effects, immunomodulatory, anti-diabetic, anti-antigenic effects, anti-microbial and anti-inflammatory activates. These compounds were extracted from leaves, seeds, bark, roots and inter-bulb surfaces (Cunog *et al.*, 2012; Hu *et al*., 2017). The majority of homoisoflavonoids have predominately been found within six families: *Asparagacease, Fabaceae, Polygonaceae, Portulacaceae, Orchidaceae* and *Gentinanaceae*. The further species names have been continued in table 4-1 below. Homoisoflavonoids can be classified into five groups, based on structure. (**I**) protosappanins, (**II**) caesalpins, (**III**) brazilins, (**IV**) scillascillins and (**V**) sappanins.

Table 4-1 Examples of the family and genus of plants containing homoisoflavonoids.

Family	Genus
Asparagacease,	*Ophiopogon, Scilla, Muscari, Polygonatum, Liriope, Agave, Chlorophytum, Hyacintus, Pseudoprospero, Dramas, Ornithogalum, Ledebouria, Dracaena, Chionodoxa, Veltheimia, Bellevalia* and *Eucomis*
Fabaceae,	*Hoffmannseggia, Caesalpinia, Cassia, Pterocarpus* and *Haematoxylum*
Polygonaceae,	*Polygomum*
Portulacaceae,	*Portulaco*
Orchidaceae	*Cremastra*
Gentinanaceae	*Trachiadenus*

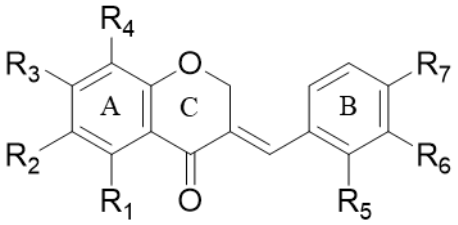
The sappanin category makes up 80% of all structurally elucidated homoisoflavonoids (Lin *et al*., 2014). Additionally, sappanins can be chemically cyclised and decyclised to form rearranged forms (Lin *et al*., 2008). Homoisoflavonoids of the sappanin-type can be split up into a further six sub-groups: : 3-benzylchroman type, 3-benzylchroman-3,4-diol type, 3-benzylchroman-4- one type, 3-benzylchroman-3-ol-4-one type , $\Delta_{2,3}$ 3-benzylchroman-4-one type , and $\Delta_{3,9}$ 3-benzylchroman-4-one type. The compounds that have demonstrated anti-cancer properties largely belong to 3-benzylchroman-4- one (Table 4-2) and $\Delta_{3,9}$ 3-benzylchroman-4-one (Table 4-3) class types (Lin, Liu and Ye, 2014; Simon *et al*., 2017; Demirayak *et al*., 2017). Both compounds contain four R groups on the chromone unit, while 3-benzylchroman-4-one also has further four R groups on the phenyl ring. On the other hand $_{3,\ 9}$,3 benzylchroman-4-one contains a double carbon bond between carbon 3 and 9, with only three R groups on the phenyl ring.

Table 4-2: Structures of 3-benzylchroman-4-one homoisoflavonoid, with their structures and sources.



No.	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	Source	Ref.
10	H	H	O CH_2O-	-	H	H	OCH_3	H	*Chlorophytum*	O'Donnell, Bucar and Gibbons, 2006
20	OH	H	OCH_3	H	H	H	OH	H	*Ledebouria*	Mutanyatta, Matapa and Shushu, 2003
21	OH	CH_3	OH	CH_3	H	H	OCH_3	H	*Ophiopogon*	Zhou *et al*., 2013
22	OH	OCH_3	OCH_3	H	H	H	OH	H	*Ledebouria*	Mutanyatta, Matapa and Shushu, 2003
23	OCH_3	H	OCH_3	OCH_3	H	H	OH	H	*Ledebouria*	Mutanyatta, Matapa and Shushu, 2003
24	OH	H	OH	H	OH	H	OH	H	*Liriope*	Tsai *et al*., 2013
25	OH	OCH_3	OH	H	H	OH	OCH_3	H	*Scilla*	Crouch, Bangani and Mulholland, 1999

Table 4-3: Structures of $\Delta_{3,9}$ 3-benzylchroman-4-one homoisoflavonoid, with their structures and sources.



No.	R_1	R_2	R_3	R_4	R_5	R_6	R_7	Source	Ref.
26	H	H	OH	H	H	H	OCH_3	*Caesalpinia*	Namikoshi Nakata and Saitoh, 1987; Roy *et al*., 2013
27	H	H	OH	OH	H	H	OCH_3	*Hoffmannseggia*	Wall *et al*., 1989
28	H	H	OCH_3	OH	H	H	OCH_3	*Hoffmannseggia*	Wall *et al*,. 1989
29	OH	H	OCH_3	OH	H	OH	OH	*Muscari*	Mašterova *et al*., 1991
30	OH	OCH_3	OH	H	H	OH	OH	*Scilla*	Silayo *et al.*, 1999
31	OCH_3	H	OH	H	H	OH	OH	*Muscari*	Mašterova *et al*., 1991

4.1.6 *Anti-microbial effects*

Natural products are an extensive source of compounds. O'Donnell (2006) isolated **19** from *Chlorophytum inorantum,* which exhibited activity against four strains of fast-growing myco-bacteria with MIC ranging from 16-256 µg mL^{-1}. With the aim to isolate an efflux pump inhibitor, Roy *et al*., (2013) isolated compound **26** from *Chlorophytum digyna* roots. Compound **26** had a reduction of the MIC of ethidium bromide against *Mycobacterium smegmatis*; it also exhibited significant efflux pump inhibition (Roy *et al*., 2013).

4.1.7 *Anti-mutagenic effects*

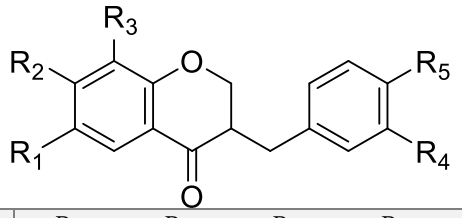
Mutagens have been classed as chemical or physical agents that change genetic material (Benigni and Bossa, 2011) Compounds **27** and **28**, intricatin and intricatinol, respectively, are of the $\Delta^{3,9}$ 3-benzylchroman-4-one category, and have demonstrated anti-mutagenic activities (Wall *et al*., 1989). Compounds **27** and **28** were both isolated from the roots of the flowering *Hoffmannseggia intrica*. They were subjected to three different mutagens, ethyl methanesulfonate (EMS), 2-aminoanthracane (2AN) and acetylaminofluorene (AAF) in the Gram-negative bacterium *Salmonella typhimurium*. This activity is possibly due to the presence of the two-hydroxy groups on phenyl A in the chromone structure of the compound. Furthermore, an extract of *Muscari racemosum* was shown to exhibit anti-mutagenic effects on *Saccharomyces cerevisiae* and four strains of *S. typhimurium* (Miadoková *et al*., 2002). In this experiment, there was a three-fold reduction in the mutagenic activity of 4-NQO (4-nitroquinoline 1-oxide), a mutagen and carcinogen used to test for anti-mutagenicity. Three homoisoflavonoids were extracted from the bulbs of *M. racemosum* these three compounds **29**, **30** and **31** fall in to the $\Delta^{3,9}$ 3-benzylchroman-4-one category (Table 2) (Miadoková *et al*., 2002).

4.1.8 *Cytotoxicity and anti-angiogenesis*

Reported homoisoflavonoids showed cytotoxicity against different cell lines from a wide range of tumours, such as breast cancer, colon cancer and lung cancer (Guo *et al*., 2013). One reported mechanism is to increase phosphorylation of Bcl-2, thus causing mitotic arrest in breast cell cancer cells. This pathway can be induced by compound **21**, with an increase in phosphorylation of Bcl-2, causing G2/M cell arrest and the up-regulation and expression of p21 and p53, thus causing decreased cell

viability, consequently causing apoptosis (Rafi *et al*., 2007). Isolation and structural characterisation were reported by Zhou *et al*., (2013). Compound **21** was extracted from *Polygonatum odoratum* (Vietnamese coriander). Mutanyatta *et al*., (2003) isolated and structurally elucidated five novel homoisoflavonoids **20**, **22** and **23** from the bulbs of *Ledebouria graminifolia*. They were then subsequently analysed at the US National Cancer Institute against sixty different human tumour cell lines for cytotoxic and anti-proliferative activities *in vitro*. Compound **22** demonstrated GI_{50} = 7 µg mL, 50 % growth inhibition in MCF-7 breast cancer cell lines (Boyd and Paul, 1995), while compounds **20** and **23** were inactive (Mutanyatta *et al*., 2003), indicating that the presence of the methoxy group is needed at C-6 phenyl A (R_2) for cytotoxicity activities. Nguyen *et al*., (2006) highlighted the extracts of roots *Disporopsis aspera* (*Asparagaceae)* to have cytotoxic activity against numerous human cancer cell lines, such as MCF-7, with IC_{50} ranging from 15 to 200 µM; the compound isolated has been termed disporopsin (**24**). This activity could possibly be due to the lack of methoxy groups on compound **24,** as it is the only compound reported above with only consisting of hydrogens and hydroxy groups are present. Furthermore, in a recent paper by El-Elimat *et al*., (2018), four new homoisoflavonoids were identified in a chloroform extract of the bulbs of *Bellevalia fexuosa*, along with thirteen known analogues. Cytotoxic analysis of the isolated compounds were evaluated using human cancer cell lines. Compound **32** and **33** (Table **4-4**) were found to be the most potent against MDA-MB-435 cancer cell line with IC_{50} 1.6 and 2.0 µM, respectively, furthermore OVCAR cancer cell line with IC_{50} 9.5 and 10.8 µM, respectively. **33** was most potent against MDA-MB-231 breast cancer cell line with IC_{50} 3.6 µM.

Table 4-4: Structures of the homoisoflavonoids **32** and **33**.



	R_1	R_2	R_3	R_4	R_5
32:	H	H	H	OH	CH_3
33:	H	CH_3	OCH_3	OH	CH_3

Angiogenesis is the formation of new blood vessels triggered by signals such as vascular endothelial growth factor (VEGF). Signals such as VEGF can promote growth and survival of solid cancer tumours, which are larger than a few millimetres. With new blood vessels providing oxygen and needed nutrients, metastases can be enhanced. Therefore, anti-angiogenesis agents such as **15** (Table 4-2) work by blocking the growth of blood vessels the support the tumour, thus inhibiting the growth and potentially the metastatic state of the cancer. Compound **15** is an homoisoflavonoid in the 3,9 3-benzylchroman-4-one category (Shim *et al.,* 2004); it was isolated from the bulb of *Cremastra appendiculata* (Orchidaceae). Compound **15** has been reported by Shim *et al*. for being a potent angiogenesis inhibitor, with a mechanism of inhibiting basic fibroblast growth factor (bFGF) *in vitro* and *in vivo* angiogenesis of the chorioallantoic membrane of chick embryos, with no toxicity (Shim *et al*., 2004).

As seen above, the structure diversity of homoisoflavonoids have received much attention with in literature. With over 250 naturally derived homoisoflavonoids classified into five categories, they have been become of great importance for research and development of natural bioactive compounds, with significant cytotoxic activities. Further isolation and characterisation of homoisoflavonoids from other medical plants are needed, with additional investigations on pharmacology, toxicological and chemical studies, for better validation of the therapeutic potential of homoisoflavonoids. Collaborators at the University of Bath and at Brunel University have grown the plant (confidential identity) and established cytotoxic activity on MDA-MB-231 breast cancer cell line (unpublished communication, Dr. Ifat Parveen Shah).

4.2 *Specific aim*

The overall aim of the study is to extract the two potentially novel homoisoflavonoids from roots of plant X, to investigate further whether the compounds of interest are also present within the leaf and stems and to develop an efficient purification procedure to provide material for structural elucidation.

4.3 *Objectives*

To identify the two cytotoxic compounds in the extracts of the roots of plant X. The acquired samples will be extracted, purified and structurally elucidated using standard

laboratory techniques. HPLC-UV-MS/MS, ultraviolet (UV) spectroscopy and photodiode-array (PDA) detectors will be used to identify the afforded samples for semi-preparative purification. The melting point (mp) of the pure compounds will also be recorded. The samples will be fully characterised and structurally elucidated with two-dimensional nuclear magnetic resonance spectroscopy (2D NMR) including $_{1}H$, $_{13}C$, and 3D NMR spectroscopy, correlation spectroscopy (COSY), nuclear Overhauser effect spectroscopy (NOESY).

4.4 *Experimental*

4.4.1 *General procedures*

Melting points were determined on a Mel-Temp melting point apparatus (Electrothermal Engineering Ltd Rochford, UK) with Fluke 51 II thermometer (Fluke Corporation, Everett, Washington, USA) in open capillaries, and are uncorrected. IR spectra were obtained using Bruker Equiox 55 (Bruker Corporation, Billerica, Massachusetts, USA) with Specac Golden Single Reflection Diamond Attenuated Total Reflectance accessory (Specac Ltd, Kent, UK).

HPLC-UV-ESI-MS/MS analysis was performed as previously described (Parveen *et al*., 2001) with a Thermo Finnigan LCMS System (Thermo Electron Corporation, Waltham, Massachusetts, USA). To a glass vial crude plant extract (1 mg) was added HPLC grade MeOH (1 mL). The tube was vortexed, centrifuged and a 100 µL aliquot was transferred into a HPLC vial for MS analysis in both positive and negative ion mode. The system included a Finnigan Surveyor PDA and alongside a Finnigan LTQ (linear trap quadrupole) with an ESI source. Chromatography was performed on a Waters C_{18} reverse-phase (3.9 × 100 mm, i.d. 4 µm) Nova-Pak column (Waters Corporation, Milford, Massachusetts, USA). The column temperature was constant at 5ºC and the temperature of the auto-sampler was maintained at 30ºC. The sample injection volume was 10 µL, the detection wavelength was set at 240-400 nm with a flow rate of 1 mL min^{-1}. The mobile phase was composed of purified water with 0.1% formic acid (solvent A) and MeOH with 0.1% formic acid (solvent B). The process of column equilibration was performed with 95% A: 5% B. During the analyses, the gradient of B increased linearly to 100% over a period of 30 min. The parameters; capillary temperature 320 ºC, nitrogen sheath gas 30 arbitrary units and nitrogen auxiliary gas at 15 arbitrary units, were used to obtain mass spectra in positive and negative modes. In negative ionisation mode, spray voltage was set to 4.8 KV,

capillary voltage was 45 V and tube lens offset was set at 110 V. Orbitrap Fusion Tribrid Mass spectrometer – LC-HRMS analysis was performed on an Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific) that was coupled to UltiMate 3000 liquid chromatography tower (Dionex, Thermo Scientific). A Hypersil Gold (Thermo Scientific) reverse phase C18 column (2.1 mm × 150 mm; particle size 5 µm) was used for chromatography, maintained at a temperature of 60 °C. 10 µL of sample was injected. The mobile phases for gradient elution were ultra-pure water (18.5 MΩ) with 0.1% formic acid (mass spectrometry grade, Fluka), (solvent A) and MeOH (HPLC grade, Fisher Scientific) with 0.1% formic acid (solvent B). The flow rate was fixed at 0.6 mL min^{-1}. The initial condition was 100% A with a linear increase to 100% B over 7 mins. 100% B was held for 3.5 minutes before equilibration at initial conditions for a further 2.5 min. Ions were generated in a HESI-II source with a source voltage of 3500/2500 V for positive/negative mode, sheath gas: 45, aux gas: 13, a vaporiser temperature of 358 °C and an ion transfer temperature of 342 °C. Ions were detected in profile mode in the 100-2000 *m/z* range in the Orbitrap detector at a resolution of 240000 and an injection time of 100 ms in both positive and negative mode. The reactions were monitored by thin layer chromatography (TLC) performed on a silica gel 60 F_{254} pre-coated on aluminium (Merck KGaA, Darmstadt, Germany) and visualised under UV at 254 nm. Silica gel (70-230 mesh) (Sigma-Aldrich, St. Louis, Missouri, USA) was used for column chromatography. Sample were at 10 mg/mL in MeOH. To identify the peak of interest and the corresponding UV patterns a method development column is used. F14 Symmetry C8 3.5 µm 4.6 × 50 mm column) with a gradient of Water to MeCN 100% over 13 min, containing 0.01 % TFA. Once peak of interest has been identified, individual method development can take place on a longer column to get separation. The column used for further method development was HICHROM ACE 5 C_{18} 3.5 µm 4.6 × 250 mm; after development it can be taken straight onto the semi-preparative system. Semi-preparative system; a Waters HPLC prep LC system controller 2000 coupled to a Waters 2487 dual λ absorbance detector was used for semi-preparative HPLC. The column used was a HICHROM ACE 10 C_{18} (250 × 21.2 mm). Samples were prepared at a concentration of 10 mg mL^{-1} by dissolving the crude material in the organic solvent used for the mobile phase. All solvents used were of HPLC grade. Before injection, samples were mixed with water to match the starting conditions. The flow rate, wavelengths and injection volumes

used were the same for all methods, 15 mL min^{-1}, 210 nm respectively. The injection volume was 2 mL. Data was recorded on a Kipp and Zonen BD 12E Flatbed recorder BD12E. All chemicals were obtained from commercial suppliers, and were used without further purification: Tetrahydrofuran (THF) and Na_2SO_4 (Sigma-Aldrich, St. Louis, Missouri, USA), HPLC-grade solvents: MeOH, EtOH, EtOAc, MeCN Hexane and petroleum ether (Fisher Scientific, Waltham, Massachusetts, USA).

4.4.2 *Plant X extractions*

Prior to extraction, plant X was stored at -20ºC. The plant X was then washed under a cold water tap by hand to remove any soil content, the root were then covered in foil and left to dry in dark ambient conditions for 72 h. The leaves and stem were then separated carefully from the roots. The two materials were then cut finely and transferred into a pestle and mortar, and powdered (7.1873 g). HPLC-grade MeOH (3 × 300 mL) was added to the roots and leaves and stem and the mixtures were stirred for 48 h at ambient temperature. The procedure was repeated three times to allow efficient extraction. All extracts from roots or stems and leaves were pooled together, then the MeOH was evaporated using a rotary evaporator, yielding 567 mg from the roots and 1511 mg from leaves and stem. 5 mL of MeOH from each extract remained; this was then centrifuged at 3000 rpm for 10 min. Semi-preparative purification used a HICHROM ACE 10 C_{18} (250 x 21.2 mm). Samples were loaded at 10 mg mL^{-1} in MeOH. The mobile phase consisted of 40 % MeCN 60 % H_2O – 60 % MeCN and 40 % H_2O over 15 mins.

*Compound **A:*** The HPLC – MS/MS of **A** showed a peak at t_R 14.04 min, (1.3 mg) as a white soild. λ_{max} 326 nm. HRESIMS *m/z* 311.0921 corresponding to $[M - H]^-$ for $C_{18}H_{15}O_5$ (calcd: 311.0912) (1.3 mg) for the *pseudo*molecular ion.

*Compound **B**:* The HPLC – MS/MS of **B** showed a peak at t_R 14.22 min, (2.6 mg) as a white solid. λ_{max} 326 nm. HRESIMS *m/z* 341.1025 corresponding to $[M - H]^-$ for $C_{19}H_{17}O_6$ (calcd: 341.0103) (2.6 mg) for the *pseudo*molecular ion.

*Compound **C**:* The HPLC-MS/MS of **C** showed a peak at t_R 14.39 min, *m/z* 739 corresponding to $[M - H]^-$ the *pseudo*molecular ion.

*Compound **D**:* The HPLC-MS/MS of **C** showed a peak at t_R 14.64 min, *m/z* 594 corresponding to $[M - H]^-$ the *pseudo*molecular ion.

4.5 *Results and Discussion*

4.5.1 *Plant X root analysis*

Plant X is a confidential species that was grown in a greenhouse in Batheaston, Bath, UK by Dr Susan Oldfield. It has been reported that plant X has significant cytotoxicic properties against the MDA-MB-231 breast cancer epithelia cell line (unpublished communication, Dr. Ifat Parveen Shah). With the use of 100% HPLC grade MeOH, the polar compounds within the roots were extracted. The solvent was evaporated and crude extract was centrifuged to obtain the supernatant. HPLC-UV-MS/MS of the roots identified two compounds at t_R of 16.32 and 16.47 min, with the UV absorption of 290 and 326 nm with *m/z* values of 313 and 343, respectively, (Figure: 4-5). Both peaks within the chromatogram had an UV spectra with λ_{max} in the range of 300-550 nm and λ_{max2} very close to the range of 240-285 nm which are the characteristic ranges of the flavonoid B and A rings respectively (Figure 4-8) (Sun *et al.*, 2011; Pinheiro and Justino, 2012; Chen *et al.*, 2018), this strongly suggests that the compounds are within the flavonoid family.

A semi-preparative purification method was developed and separation was completed with a mobile phase of 40-60 % MeCN over 15 min. Purification highlighted a third compound present in the middle of the two major peaks. Due to a small extraction yield, separation was unable to remove the smaller third peak from the samples. Instead, the unknown sample was included with compound **A**. Semi-preparative purification yielded 1.3 mg and 2.6 mg for compound **A** and compound **B** respectively. HPLC-UV-MS/MS analysis was conducted on the newly separated samples compound **A** (Figure **4-6**) and compound **B** (Figure **4-7**), the separated compounds had consistent UV absorbance patterns as within the crude mixture.

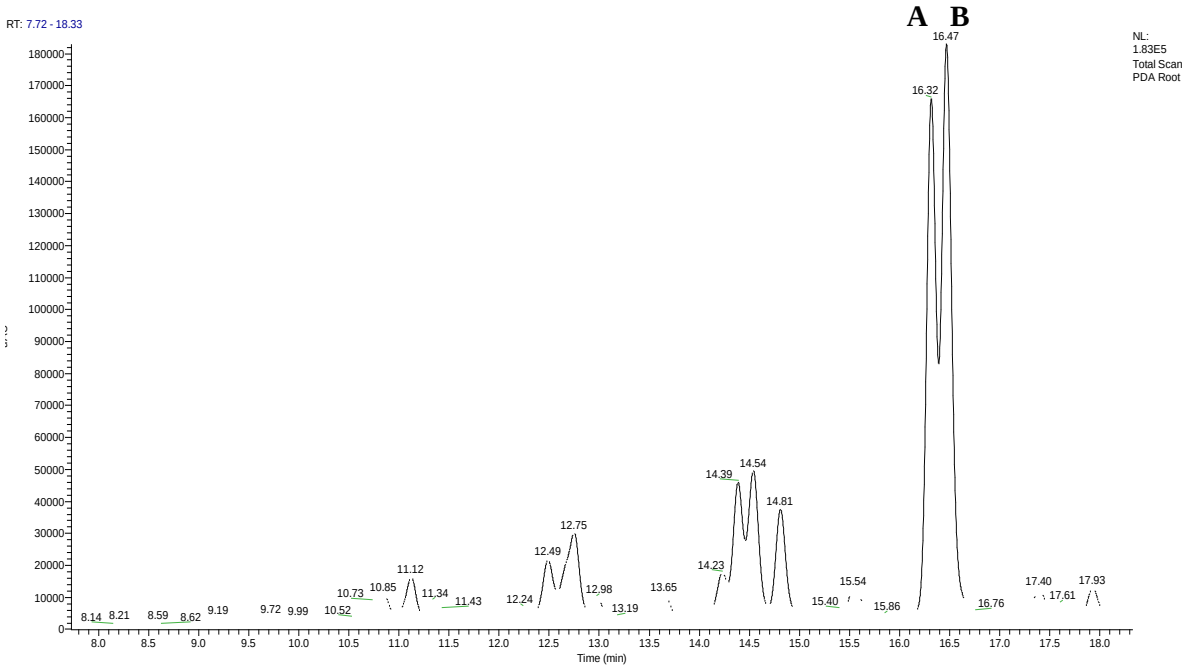


Figure 4-5: HPLC chromatogram of root extract of plant X, identifying compound **A** and **B** with t_R of 16.32 and 16.47 respectively.

*Compound **A**:* The HPLC – MS/MS of **A** showed a peak at t_R 14.04 min, HRMS *m/z* 311.0921 corresponding to $[M-H]^-$ for $C_{18}H_{16}O_5$ (calcd 312.09978) (1.3 mg) for the *pseudo*molecular ion (Figure: 4-6). In negative ion mode the base peak ion was observed at *m/z* 356 $[(M-H)+HCOOH]^-$, this formate adduct would have been formed during the ionisation process. A second ion was observed at *m/z* 623 $[2M-H]^-$ corresponding to the a [2 M – H]- cluster. Further ion were observed at *m/z* 341 $[(M-H)+2CH_3]^-$, corresponding to methyl groups, suggesting methoxy groups are present within the structure, *m/z* 367 $[(M-H)+2H_2O]^-$. A characteristic ion was also observed at *m/z* 249 $[M-H.2OCH_3]^-$, indicating the loss of two methoxy groups. MS^2 of the *pseudo*molecular ion *m/z* 311 $[M-H]^-$, yielded three fragments, *m/z* 205 $[M-H-C_{11}H_9O_4]$, and *m/z* 106 $[M-H-C_7H_7O]^-$ indicating fragmentation at C_{3-9} bond (Figure: 4-9). Further fragments at *m/z* 189 [M-H-112] and *m/z* 160 [M-H-151]-. In positive ion mode, the *pseudo*molecular ion was observed at *m/z* 313 $[M+H]^+$, *m/z* 335 $[M+Na]^+$, further ions observed were *m/z* 626 $[2M-H]^+$. MS^2 yielded four fragments at 295 $[(M+H)-H_2O]^+$, *m/z* 267 $[(M+H)-HCOOH]^+$ and *m/z* 253 $[(M+H)-2OCH_3.OH]^+$. The UV spectra are consistent with of flavonoids (Pinheiro and Justino, 2012). Analysis of

melting points are unsuitable due to the contamination of smaller peaks observed at t_R 14.22 min (Figure 4-6).

*Compound **B***: The HPLC – MS/MS of **B** showed a peak at t_R 14.22 min, HRMS *m/z* 341 corresponding to [M-H]- for $C_{19}H_{19}O_6$ (calcd 343.118165 (2.6 mg) for the *pseudo*molecular ion (Figure: 4-7). In negative mode, the base peak ion was observed at *m/z* 682 [2M-H]-. A further ion was observed at *m/z* 386 [(M-H)+ formic acid]-, formed during the ionisation process due to the presence of formic acid. MS_2 of the pseudomolecular ion *m/z* 341 [M-H]- yielded three fragments; *m/z* 325 [M-H.CH_3]-, *m/z* 297 [M-CO_2]- and *m/z* 282 [M-CO_2.CH_3]-. In positive ion mode, the *pseudo*molecular ion was observed at *m/z* 343 $[M+H]^+$, with ions observed at *m/z* 686 $[2M+H]^+$ and *m/z* 214 $[(M+H)-129]^+$. MS_2 fragmentation in positive ion mode yielded three fragments at *m/z* 205 and *m/z* 136, resulting from the fragmentation at $C_{3\text{-}9}$ (Figure: 4-9**)**. The base peak was observed at *m/z* 151, indicating a loss of 192 [(M+H)-205-CH]-.

Fragmentation analysis of compounds **A** and **B** in negative mode provide major fragments at the following; *m/z* 205 and 106 for compound **A** and *m/z* 206 and 136 for compounds **B**, with fragmentation at the $C_{3\text{-}9}$ bond. Based on the UV and the fragmentation data the two peaks were determined to be compounds **A** and **B** (Figure **4-11**).

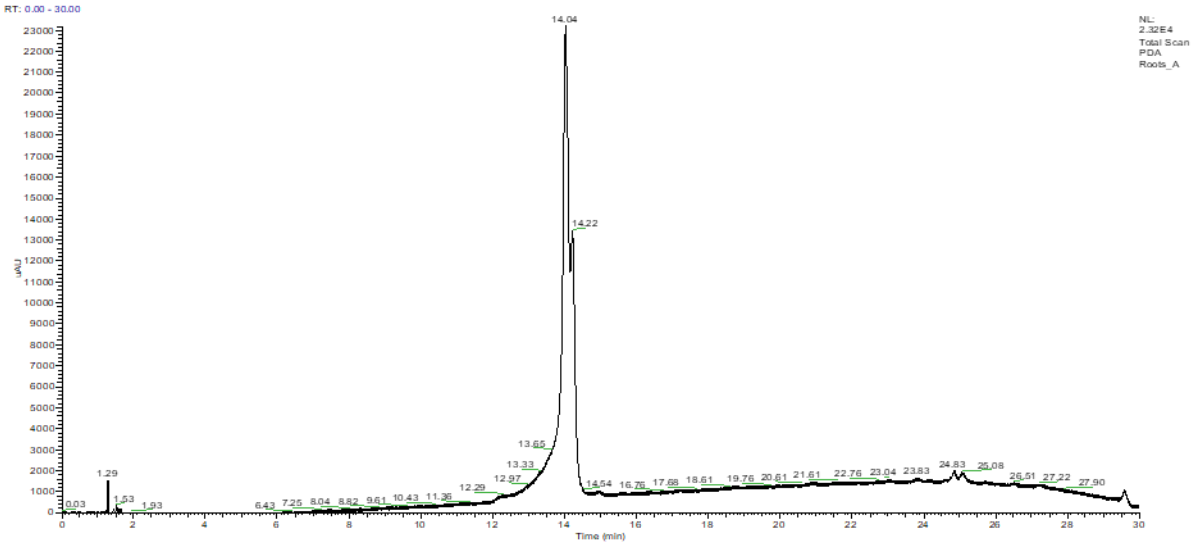


Figure 4-6: HPLC-UV-PDA chromatogram of compound **A**. After purification the t_R had changed to 14.04 min. Compound **A** still had the same corresponding 289 and 326 nm UV and fragmentation characteristic

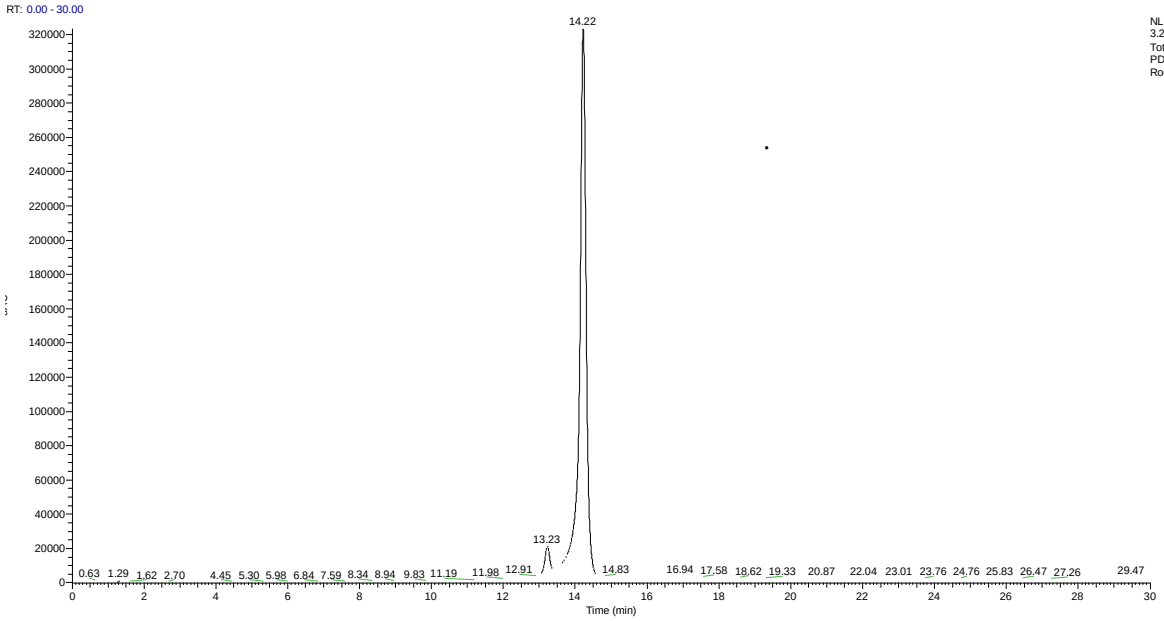


Figure 4-7: HPLC-UV-PDA chromatogram of compound **B**. After purification the t_R had changed to 14.22 min. Compound **B** still had the same corresponding 289 and 326 nm UV characteristic.

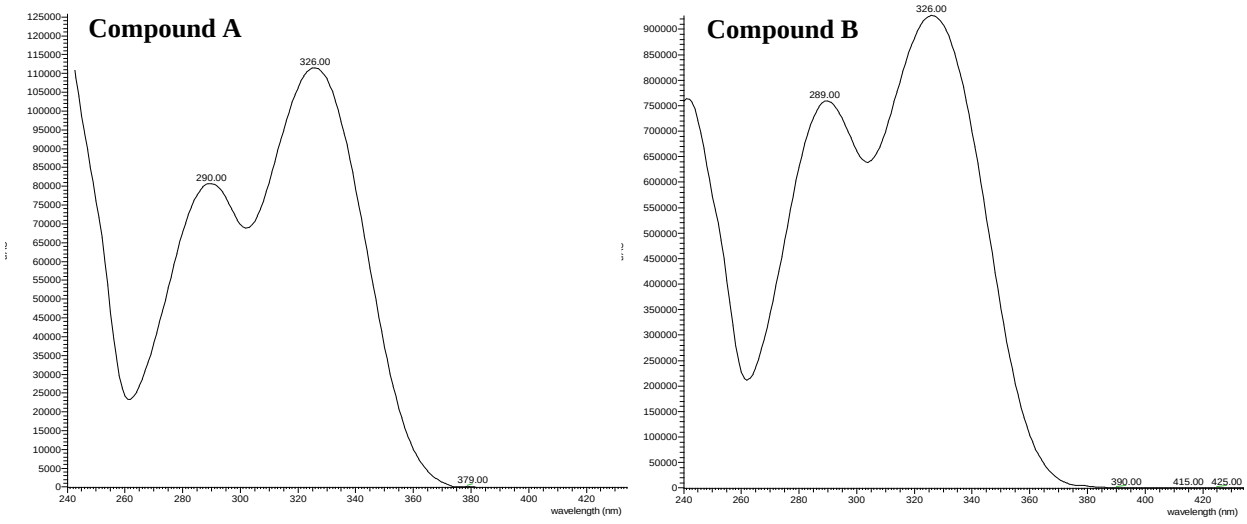


Figure 4-8: UV spectra of plant X compounds **A** and **B**. Both λ_{max1} and λ_{max2} in or very close to the ranges of flavonoids (Pinheiro and Justino, 2012).

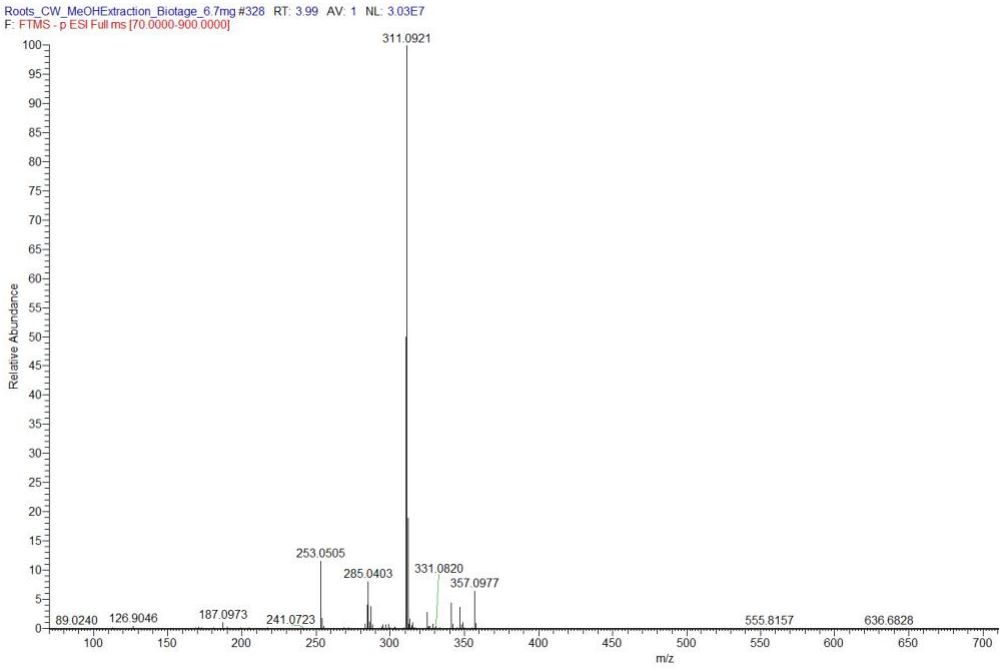


Figure 4-9: HRESIMS *m/z* of compound **A**.

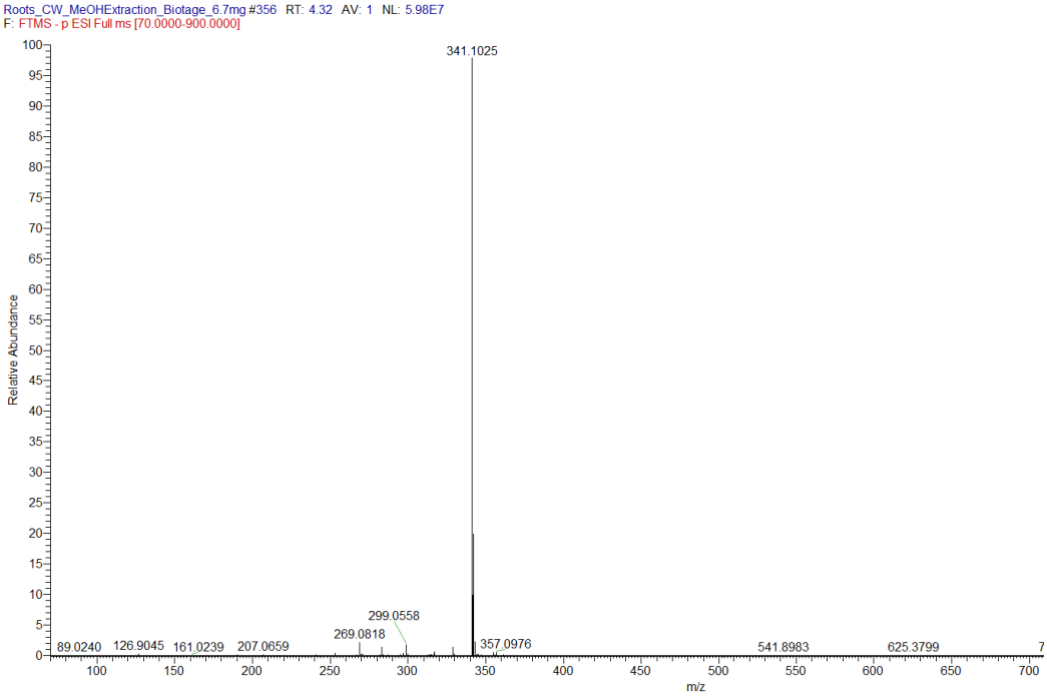


Figure 4-10: HRESIMS *m/z* of compound **B**.

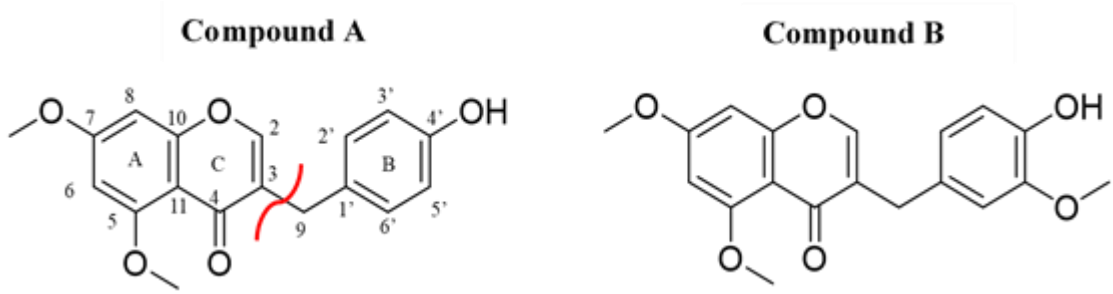


Figure 4-11: The suggested likely chemical structures of compounds **A** and **B**. Highlighting the $C_{3\text{-}9}$ bond of the major fragment by MS analysis. Carbon numbering from Escobar-Ramos *et al*., (2017).

Flavonoids roles within plant include; antibacterial, UV and free radical protection and regulation of plant growth (Havsteen, 2002 and Jones *et al*., 2007). These compounds are also present within nuts, vegetables and fruits, where they possess anti-oxidant and anti-cancer properties, alongside acting as anti-inflammatory agents (Carlo *et al*., 1998). Low toxicity makes flavonoids ideal potential drug leads. The toxicity depends on the groups attached, but a paper has highlighted flavonoids only demonstrate toxicity in very large amounts, such as 1 mg Kg^{-1} of daidzein caused female rate to have reduced mammary glands, with no effect on the offspring (Galati and O'Brien, 2004).

Flavonoids are within the class of polyphenols and are characterised by the C=O double bond on the C-ring and carbonyl group at C-4 (Havsteen, 2002). Isoflavones have the B-ring at position 3 on the C-ring (Fotsis *et al*., 1997). HPLC-ESI-UV-MS/MS was effective in characterising the structures for the bioactive compounds found within Plant X. Compounds A and B have been identified as homoisoflavonoids, due to the additional CH_2 group between the rings of B and C (Figure 4-11). Literature has highlighted that around 250 homoisoflavonoids are known. Compounds A and B have an unique molecular weight of homoisoflavonoids, within the comprehensive review by Lin *et al*., (2014). Suggesting that compounds **A** and **B** are generally synthesised by this species of plant X. Compounds **A** and **B** eluted at similar times suggesting a similar structure, differing only in the methoxy group in compound **B** on the B-ring and UV spectra, with similar high levels of conjugation. The fragmentation pattern of the homoisoflavonoid $C_{3\text{-}9}$ bond is consistent with that found in literature by Ye *et al*., (2005), who identified eighteen homoisoflavonoids in *Ophiopoogon japonicas*.

A bioactive homoisoflavanone from *Polygonatum odoratum* similar to alternative isomers of Compound **A** but without the C-ring double bond, has been described by Rafi and Vastano, (2007) as 8-methyl-DBP (Figure: 4-12). It was reported that 8-methyl-DBP inhibited growth in a clonogenic assay. The mechanism of action is identified as inducing G2/M cell cycle arrest and phosphorylation of Bcl-2 and increased expression of p53 (Rafi and Vastano, 2007). The cytotoxic activity of Compounds **A** and **B** could potentially by similar to compounds such as 8-methyl-DBP.

8-methyl-DBP

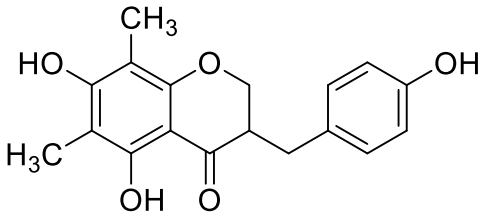


Figure 4-12: Structure of 8-methyl-DBP taken from Rafi and Vastano (2007).

4.5.2 *Plant X steam and leaf analysis*

The polar compounds of the stem and leaf of plant X were also extracted using 100% MeOH (3 ×300 mL) at ambient temperature and pressure. The solvent was removed to yield a brown residue (1511 mg). The crude sample was subjected to HPLC-ESI-UV-MS/MS analysis, where upon the HPLC chromatogram identified two compounds (Figure: 4-13). The UV spectrum of the compound **C** and **D** are consistent with that shown in figure 4-14. Compounds **C** and **D** show similar UV absorbance patterns which are also consistent with those reported UV characteristics of flavone-like compounds (Zhang *et al*., 2011).

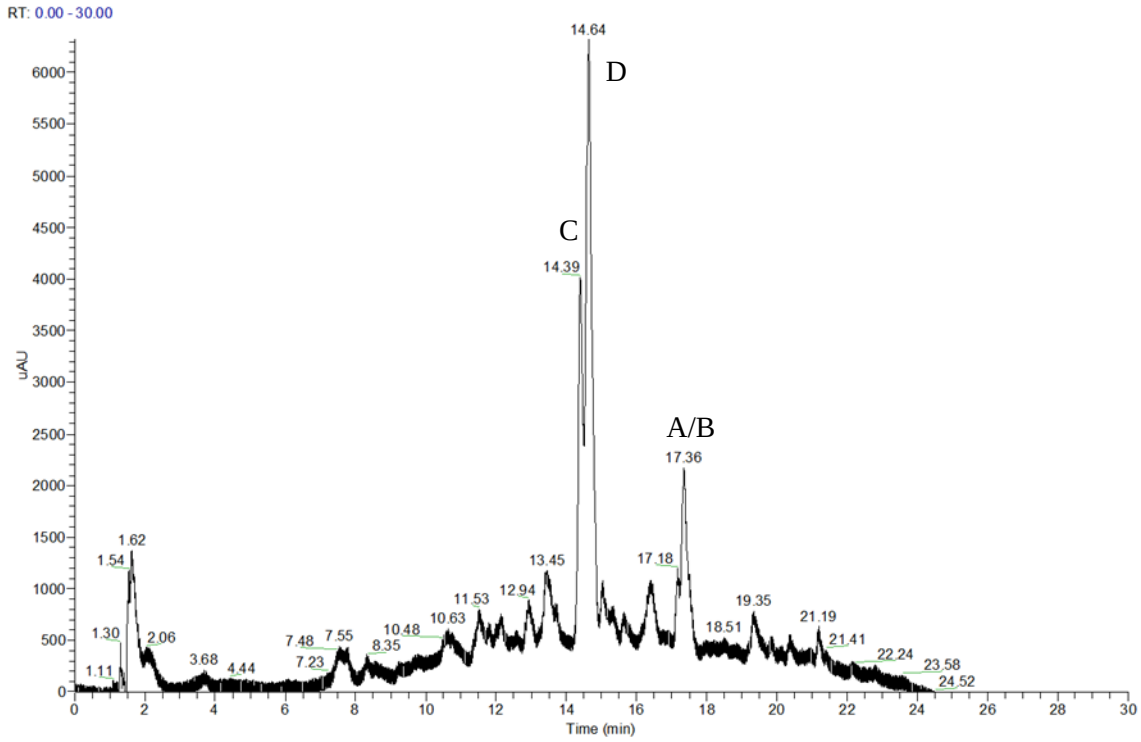


Figure 4-13: HPLC-UV-PDA chromatogram of the crude extract of stem and leaves. Compound **C** t_R 14.39 min and compound **D** t_R 14.64 min, (traces of compounds **A** and **B** are present at t_R 17.18 and 17.36 min).

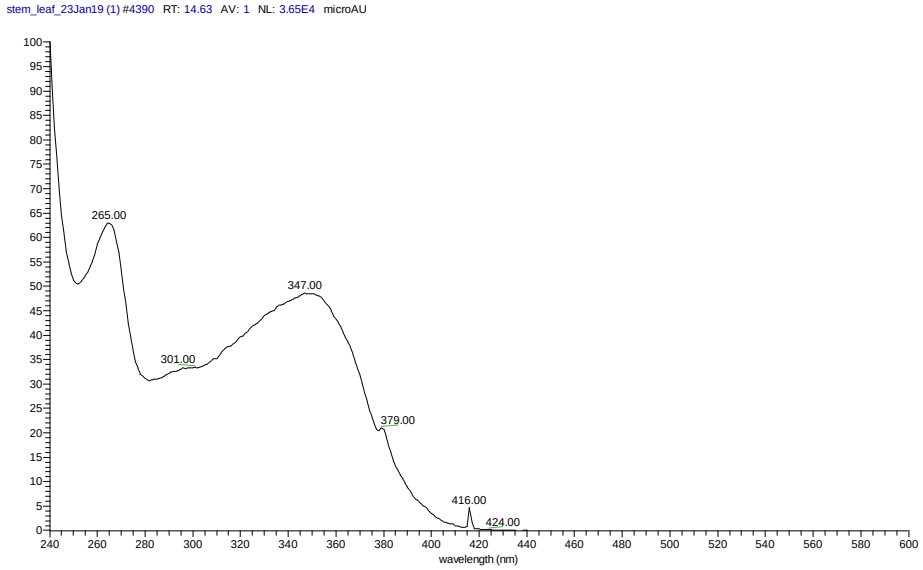


Figure 4-14: UV spectrum of compounds **C** and **D** from plant X stem and leaf material.

The HPLC–MS/MS (ESI negative mode) of **C** at t_R 14.39 min, *m/z* 739 $[M-H]^-$ corresponding to the pseudomolecular ion, further ions were observed at *m/z* 1479 $[2M-H]^-$ and *m/z* 785 $[(M-H)+HCOOH]^-$. MS_2 of the pseudomolecular ion *m/z* 739 $[M-H]^-$ yielded, five fragments *m/z* 721 $[(M-H)-H_2O]^-$, *m/z* 593 $[(M-H)-164]^-$, *m/z* 575 $[(M-H)-H_2O, 146]^-$, *m/z* 557 $[M-H, 162]^-$ and at *m/z* 285 $[aglycone-H]^-$. The ion at *m/z* 721, corresponding to the loss of a water molecule is consistent with the fragmentation

of homoisoflavonoids (Qi *et al*., 2010). The ions at *m/z* 575 and 593 correspond to the loss of a deoxyhexose unit, 146 and 164 Da units thus confirming the presence of an *O*-glycoside (Vukics *et al.,* 2008). The ion at *m/z* 285 indicates the presence of kaempferol. The UV spectrum (Figure: 3-12) and fragmentation patterns of compound **C** were typical of flavonoid *O*-glycosides, and are consistent of the flavone 3,5,7,4'-tetrahydroxyflavone-3-*O*-robinoside-7-*O*-rhamnoside (robinin) (Figure: 4-15) (March *et al*., 2004: Bammaw *et al*., 2016).

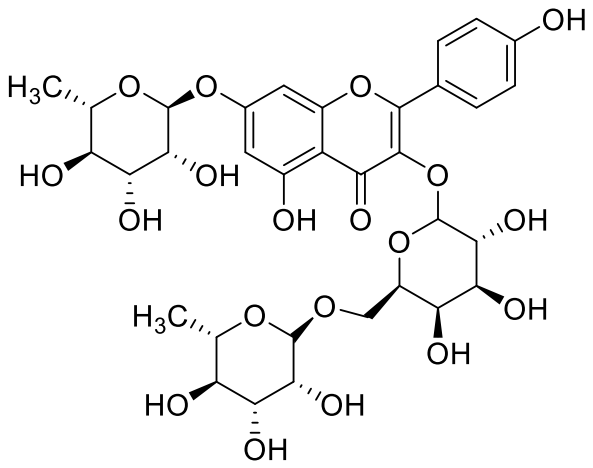


Figure 4-15: Structure of robinin (compound **C**).

Compound **C**, can be hydrolysed to form kaempferol (Bokkenheuser *et al*., 1987), which has many beneficial properties; these include; antidiabetic, anti-inflammatory, anti-tumour, neuroprotective and hepatoprotective (Wang *et al*., 2006; Arif *et al*., 2018; Wang *et al*., 2018). This compound has also demonstrated anti-proliferative effects on mouse colon cancer cell line (CT26), mouse melanoma cell line (B16F1) and human hepatoma cell line (HepG2) (Wang *et al*., 2018).

The HPLC-MS/MS (ESI negative mode) of **D** at t_R 14.64, *m/z* 593 [M-H]-, corresponding to the *pseudo*molecular ion; a formate adduct was present at *m/z* 639 [(M+H)-HCOOH]-. MS_2 yielded additional fragment ions, m/z 575 [M-H, H_2O]-, 503 [M-H, 90]-, 473 [M-H, 120]-, 447 [M-H, 146]-, 429 [M- H, 164]-. The ion at *m/z* 575 is indicative of the loss of a water molecule, indicative of the loss of a sugar, which is a consistent fragmentation patterns of homoisoflavonoids, (Qi *et al*., 2010). The ions observed at *m/z* 447 and *m/z* 426 correspond to the losses of 146 and 164 Da, thus indicating the presence of a flavonoid *O*-glycoside containing deoxyhexose units (Vukics and Guttmann, 2008). Further ions were observed at *m/z* 357 and *m/z* 327; these ions are suggestive of the presence of an aglycone and these are common

amongst *O*-glycosides (Geng *et al*., 2016). A ion at *m/z* 285 suggests that kaempferol could in fact be the aglycone present. Further purification will allow 2D and 3D NMR structural elucidation of this molecule.

The HPLC-MS/MS (ESI negative mode) also identified the compounds **A** and **B** with the in stem and leaf extraction (Figure: 4-13), although at significantly less abundant. This was expected due to the variety of biological activities flavonoids have within plants. Flavonoids act as UV-filters, function as signalling molecules, allelopathic and phytoalexin and antimicrobial compounds (Samanta, Das and Das, 2011). Therefore compounds **A** and **B** are most likely to be used as signalling and defensive activities within the plant X roots.

4.6 *Further work*

Further isolation and purification of compounds **A** and **B** will be essential for structural elucidation by NMR to confirm the presence of two novel homoisoflavonoids within plant X species. Further identification of the other potential structures and homoisoflavonoids within the crude extract may yield additional homoisoflavonoids for potential drug leads. Additionally, synthesis of compounds **A** and **B** from available precursors could be a potential route to secure enough material. The synthesised material would be tested against MDA-MB-231 breast cancer epithelial cell line, to confirm that these compounds are responsible for the effects reported (private communication, Dr. Ifat Parveen Shah). With the synthesised compounds, investigations into the nature of the cell death and thus the mechanism of action can be performed therefore confirming at what treatment dose the compounds are cytostatic or cytotoxic. To investigate further whether the effects are specific to breast cancer or have a common mode of action against other cancers. In-house biological investigations will continue to examine the antimicrobial and anthelminthic activities of the synthesised compounds **A** and **B**.

4.7 *Conclusion*

The aim of the study was to identify the two novel homoisoflavonoids with anti-breast cancer activities from plant X. The mass spectral data identified two homoisoflavonoid stereoisomers from the root material and the identification of two potentially useful compounds present in the stem and leaf material of plant X. There are multiple studies, which support the biological activity of homoisoflavonoids, including antimicrobial,

anthelmintic, anti-oxidant and anti-cancer. Further work is needed to extract additional material for structural elucidation, of the potentially novel compounds **A** and **B**, for prospective drug leads against MDA-MB-231 breast cancer of the epithelial cell line. Research suggests that homoisoflavonoids will make suitable drug candidates as they are tolerated by patients.

Limitations include insufficient quantities of the compounds from low yielding extractions of natural recourses. With the isolated compounds **A** and **B** sent for NMR, the data gained may provide a synthetic method development may provide affective route to obtain adequate supplies. Thus removing the limitation of insufficient yields from natural product extractions and thus reducing the drug development cost. In conclusion, the two potential novel homoisoflavonoids isolated from Plant X, compounds **A** and **B** may offer a low cost option for health care providers worldwide and will act as lead candidates to identify the molecular target for anticancer drug development programmes.

Chapter 5

General Conclusion

5.0 *General conclusion*

Successful drug discovery starts by developing knowledge from traditional medical systems and developing them into effective therapies for diseases. The significance of natural products in providing new chemical entities with unique biochemical properties cannot be undermined. Natural products and their derivatives have been reported for many years as a source of therapeutic agents and structural diversity. This thesis has incorporated natural product drug discovery with medicinal and computational chemistry over three different chapters.

Chapter 2 "Oxidative heterodimerisation of 4'-hydroxycinnamate esters with 4'-hydroxycinnamic acids, potential HIV-1 integrase inhibitors", gave an in-depth literature review into the epidemiology, structure and function of HIV-1, integrase and its current inhibitors. The aim of the study was to synthesis a large portfolio of hydroxycinnamic acids with their corresponding hydroxycinnamate ethyl esters. To create a potential lead as an anti-HIV IN candidate, with similar non-toxic but highly potent inhibitor, much like lithospermic acid.

The successful synthesis with good yields of the starting ethyl esters enabled the oxidative coupling to be completed. The first conclusion to be drawn is by the oxidative coupling of cinnamic and ethyl cinnamate, the *para*-hydroxyl group is essential for the radical coupling with the peroxidase enzyme. Thus providing evidence towards the mechanism of synthesis. The coupling of ethyl *para*-coumarate and *para*-coumaric acid proved to be the most controlled reaction, yielding three significant peaks, as shown in the HPLC chromatogram. Compounds **2** and **3** with successful structural elucidation of novel compounds, with further work identification of compound **1** is possible.

Furthermore, oxidative coupling of ethyl caffeate and caffeic acid proved to be very challenging. The radical chemistry had the use of an extra hydroxyl group on C3, thus creating multiple analogues, multiple purification techniques were employed to break up the crude, so it becomes suitable for HPLC semi-preparative purification, further NMR solubility issues hampered structural elucidation analysis of this reaction. The

radical coupling of ferulic acid and ethyl ferulate, seemed to be controlled more so than caffeic, this is due to the less reactive methoxy group on Ar 3-C. Oxidative coupling of ferulic acid and ethyl ferulate yielded the conventional 8-5 linkage in the form of a ethyl heterodimer (compound 12) and the diethyl homodimer (compound 13). Furthermore, it produced a unique linkage, were there was no *trans*-disubstituted alkene. The structure is very similar to that report by Ralph *et al.* (1994) (compound 39) a methyl ester and similar to the carboxylic acid synthesised by Kim *et al*. (2005) The proposed mechanism of compound **11** has also been discussed. ^{1}H NMR, the ethyl ester on the heterodimer is always located on the furan ring, so far they ethyl ester has not been observed on the top of the molecule, leaving the carboxylic acid group on the furan ring. A possible method of the free Michael addition, presence of the ester at the top makes the alkenic protons electrophilic, thus creating the ethyl ester on the furan ring. The isolated compound can be screen in other biological assays, including but not limited to antimicrobial, anti-mutagenic and anthelminthic.

Chapter 3 titled computational docking of potential hydroxycinnamate ester derivatives with the structure of HIV-1 integrase investigated a small portfolio of compounds from chapter 1 as potential IN inhibitors. All the ligands were docked into the acceptor DNA-binding site, from the docking studies, it appears that the ester group on the tetrahydrofuran ring of the 8-5 ethyl diferulate A is forcing the compound far from the active site; this would ultimately results in a non-inhibition of integrase. Furthermore the lateral acid group is co-ordinating with the Mg^{2+}, but literature suggests that the acid group on the tetrahydrofuran should be involved in the co-ordination Savarino, 2007; DeAnda *et al*., 2013; Li *et al.,* 2015). With an interaction similar to that of the reported inhibitor L-731,988 docked by Savarino *et al*., (2006) (Figure 3-13), which contains the desired β-hydroxy carbonyl motif (Pommier *et al*., 2005). The docking of the compounds within the DNA donor-binding site was also simulated. Resulting in intermittent interactions with the Mg^{2+} ions, as the compound occupied the DNA donor-binding site with a different pose of the co-crystallised 5-CITEP.

Chapter 4 titled the identification of two novel homoisoflavonoids with anti-cancer properties from plant X. The overall aim of the chapter was to identify two potentially novel homoisoflavonoids from plant X, with reported anti-breast cancer properties (unpublished communication, Dr. Ifat Parveen Shah). The HPLC-UV-ESI-MS/MS

identified two homoisoflavonoid stereoisomers from the root material with *m/z* 311.0921 and *m/z* 341.1025 with UV spectra consistent with that for flavone-like compounds. MS data also identified two potentially useful compounds present within the stem and leaf material, the data corresponded to robinin, which possess a range of favourable activities such as anti-proliferative, anti-inflammatory, anti-androgenic and activity anti-estrogenic. Traces of the two novel homoisoflavonoids were also identified within the stem and leaf material. Due to the continuous insufficient yields, synthesis from available precursors of the predicted structures provided by literature and MS data, should be investigated, with sequential analysis on MDA-MB-231 breast cancer cell lines. As demonstrated by this thesis natural product drug discovery has become a very successful approach for the advancement of novel potent therapeutic drugs. Compounds **A** and **B** may offer a low cost option for health care providers worldwide and will act as lead candidates to identify the molecular target for anticancer drug development programmes.

Recent advancement in analytical and computational techniques have opened multiple pathways to isolate and structurally elucidate complex natural products, using their structures to design and create innovative drugs. With computational molecular design such as predictive computational software has contributed to the discovery of molecular targets of natural products and their derivatives.

Currently this technology is predicting pharmacodynamics and pharmacokinetic data resulting in a reduced number of false positive leads in natural product drug discovery. This thesis has discussed plant-based natural product drug discovery and how computational models can influence medicinal chemistry and natural product drug discovery.

References:

Abd-Elazem, I., Chen, H., Bates, R., & Huang, R. (2002). Isolation of two highly potent and non-toxic inhibitors of human immunodeficiency virus type 1 (HIV-1) integrase from Salvia miltiorrhiza. *Antiviral Research*, 55, pp.91–106.

Aidsinfo.unaids.org. (2018). AIDSinfo | UNAIDS. [online] Available at: http://aidsinfo.unaids.org/ [Accessed 11 Jul. 2018].

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. (2002) Molecular Biology of the Cell. 4th ed. *New York: Garland Science* 91(3) pp.401

Ali, A., Bandaranayake, R., Cai, Y., King, N., Kolli, M., Mittal, S., Murzycki, J., Nalam, M., Nalivaika, E., Özen, A., Prabu-Jeyabalan, M., Thayer, K. and Schiffer, C. (2010). Molecular Basis for Drug Resistance in HIV-1 Protease. *Viruses*, 2(11), pp.2509-2535.

Aljawish, A., Chevalot, I., Jasniewski, J., Paris, C., Scher, J. and Muniglia, L. (2014). Laccase-catalysed oxidation of ferulic acid and ethyl ferulate in aqueous medium: A green procedure for the synthesis of new compounds. *Food Chemistry*, 145, pp.1046-1054.

Anand, P., Kunnumakara, B., Sundaram, C., Harikumar, B., Tharakan, T., Lai, S., Sung, B. and Aggarwal, B., (2008). Cancer is a Preventable Disease that Requires Major Lifestyle Changes. *Pharmaceutical Research*, 25(9), pp. 2097-2116.

Asante-Appiah, E. & Skalka, A. M. (1999). HIV-1 integrase: structural organization, conformational changes, and catalysis. Adv.Virus Res., 52, pp.351-369.

Bacchi, A., Carcelli, M., Compari, C., Fisicaro, E., Pala, N., Rispoli, G., Rogolino, D., Sanchez, T., Sechi, M. and Neamati, N. (2011). HIV-1 IN Strand Transfer Chelating Inhibitors: A Focus on Metal Binding. *Molecular Pharmaceutics*, 8(2), pp.507-519.

Bae, S.; Kim, H. (2012). Enzymatic transformation of caffeic acid with enhanced cyclo-oxygenase-2 inhibitory activity. *bioorganic & medicinal chemistry Lett*, 22, 793-796. *Compound 6 in this paper.*

Bailey D, Brown D. (2001). High-throughput chemistry and structure-based design: survival of the smartest. *Drug Discovery Today*; 6(2) pp.57–59.

Bajorath J. (2002). Integration of virtual and high-throughput screening. *Nature Reviews Drug Discovery*; 1(11): pp.882–894.

Bamawa, M., Ndjele, M. and Foma, M., (2016). Characterization of Leaf Phenolic Compounds of Sabicea johnstonii by HPLC-MSn. *Journal of Natural Products and Resources*. 2(2) pp.86–89

Banerjee S, Li Y, Wang Z, Sarkar H. (2008). Multi-targeted therapy of cancer by genistein. *Cancer Lett* 269: p.226–42.

Barré-Sinoussi, F., Chermann, C., Rey, F., Nugeyre, T., Chamaret, S., Gruest, J. (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science*, 220, pp. 868-871.

Ben Said, R., Hamed, A.I., Mahalel, U.A., Al-Ayed, A.S., Kowalczyk, M., Moldoch, J., Oleszek, W. and Stochmal, A., (2017). Tentative characterization of polyphenolic compounds in the male flowers of Phoenix dactylifera by liquid chromatography

coupled with mass spectrometry and DFT. *International journal of molecular sciences*, *18*(3), pp.512.

Benhadji, A.; Serova, M.; Ghoul, A.; Cvitkovic, E. Le Tourneau, C.; Ogbourne, M. Lokiec, F.; Calvo, F.; Hammel, P.; Faivre, S. (2008). Antiproliferative activity of PEP005, a novel ingenol angelate that modulates PKC functions, alone and in combination with cytotoxic agents in human colon cancer cells. *Cancer* , 99, pp.1808–1815.

Berger, A., Murphy, M., & Farber, M. (1999). Chemokine receptors as HIV-1 correctors: roles in viral entry, tropism, and disease. *Annual Review of Immunology*., 17, pp. 657-700.

Berman, M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T., Weissig, H (2000) the Protein Data Bank. *Nucleic Acids Research, 28,* pp.235-242.

Bernabeu, E., Cagel, M., Lagomarsino, E., Moretton, M., Chiappetta, A. (2017). Paclitaxel: What has been done and the challenges remain ahead. *International Journal of Pharmacology*., 526, pp.474–495.

Bernardini, S.; Tiezzi, A.; Laghezza Masci, V.; Ovidi, E. (2017). Natural products for human health: An historical overview of the drug discovery approaches. *Natural Product Res*earch, (16), pp.1926-1950

Blanco, J., Varghese, V., Rhee, S., Gatell, J. and Shafer, R. (2011). HIV-1 Integrase Inhibitor Resistance and Its Clinical Implications. *The Journal of Infectious Diseases*, 203(9), pp.1204-1214.

Bolton, J. and Dunlap, T. (2016). Formation and Biological Targets of Quinones: Cytotoxic versus Cytoprotective Effects. *Chemical Research in Toxicology*, 30(1), pp.13-37.

Bomford, C. & Kunkler, I., (2002). Walter and Millers Textbook of Radiotherapy: Radiation Physics, *Therapy and Oncology*. 6th ed. London: *Churchill Livingstone, Complete Book.*

Bray, F., Ren, J., Masuyer, E. & Ferlay, J., (2013). Estimates of global cancer prevalence for 27 sites in the adult population in 2008. *International Journal of Cancer,* 132(5), pp. 11331145.

Brigo, A., Lee, W., Fogolari, F., Mustata, L., & Briggs, M. (2005). Comparative molecular dynamics simulations of HIV-1 integrase and the T66I/M154I mutant: Binding modes and drug resistance to a diketo acid inhibitor. *Proteins-Structure Function and Bioinformatics*, 59, pp.723-741.

Brown, O. (1990). Integration of Retroviral DNA. *Current Topics in Microbiology and Immunology*, 157, pp.19-48.

Bujacz, G., Alexandratos, J., Wlodawer, A., Merkel, G., Andrake, M., Katz, A. (1997). Binding of different divalent cations to the active site of avian sarcoma virus integrase and their effects on enzymatic activity. *Journal of Biological Chemistry*., 272, pp. 18161-18168.

Bunzel, M., Heuermann, B., Kim, H. and Ralph J. (2008). Peroxidase-Catalysed Oligmerization of Ferulic Acid Esters *Journal of Agricultural and Food Chemistry* 56(21), pp. 10368-10375.

Buonaguro, L., Tornesello, M. and Buonaguro, F. (2007). Human Immunodeficiency Virus Type 1 Subtype Distribution in the Worldwide Epidemic: Pathogenetic and Therapeutic Implications. *Journal of Virology*, 81(19), pp.10209-10219.

Burke, R., Jr.; Fesen, M.; Mazumder, A.; Yung, J.; Wang, J.; Carothers, M.; Grunberger, D.; Driscoll, J.; Pommier, Y.; Kohn, K. (1995). Hydroxylated Aromatic Inhibitors of HIV-1 Integrase. J*ournal of Medicinal. Chem*istry 38, (21), pp.4171-4178.

Cai, M., Zheng, R., Caffrey, M., Craigie, R., Clore, M. and Gronenborn, A. (1997). Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nature Structural Biology*, 4(10), pp.839-840.

Carocho, M. and Ferreira, I., (2013). The role of phenolic compounds in the fight against cancer–a review. *Anti-Cancer Agents in Medicinal Chemistry (Formerly Current Medicinal Chemistry-Anti-Cancer Agents),* 13(8), pp.1236-1258.

Carroll, P., Utshudiema, J. & Rodrigues, J., (2017). The British breast cancer epidemic: Trends, patterns, risk factors, and forecasting. *Journal of American Physicians and Surgeons,* 22(1), pp. 8-16.

Centers for Disease Control (1981). Kaposi's sarcoma and Pneumocystis pneumonia among homosexual men--New York City and California. *Morbidity and Mortality Weekly Report* (MMWR)., 30, pp.305-308.

Centers for Disease Control (1981). Pneumocystis Pneumonia --- Los Angeles. *Morbidity and Mortality Weekly Report (MMWR)* 30, pp.1-3.

Centers for Disease Control (1982). Update on acquired immune deficiency syndrome (AIDS)-United States. *Morbidity and Mortality Weekly Report* 31.

Cernerud, M., Reina, J., Tegenfeldt, J. and Moberg, C. (1996). Chiral polymers via asymmetric epoxidation and asymmetric dihydroxylation. *Tetrahedron: Asymmetry*, 7(10), pp.2863-2870.

Chakrabarti, L., Guyader, M., Alizon, M., Daniel, M. D., Desrosiers, R. C., Tiollais, P, (1987). Sequence of simian immunodeficiency virus from macaque and its relationship to other human and simian retroviruses. *Nature*, 328, pp.543-547.

Chalifoux, V., Ringler, J., King, W., Sehgal, K., Desrosiers, C and Daniel, D. (1987). Lymphadenopathy in macaques experimentally infected with the simian immunodeficiency virus (SIV). *Am.J.Pathol.*, 128, pp.104-110.

Chaudhary, P., Tupe, S. and V Deshpande, M., (2013). Chitin synthase inhibitors as antifungal agents. *Mini reviews in medicinal chemistry*, *13*(2), pp.222-236.

Chemspider.com. (2019). *(E)-Ethyl caffeate | C11H12O4 | ChemSpider*. [online] Available at: http://www.chemspider.com/Chemical-Structure.4476132.html [Accessed 18 April. 2019].

Chen, H., Krucinski J, Miercke,W., Finer-Moore, S., Tang, H and Leavitt, D (2000). Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: A model for viral DNA binding. *Proceedings of the National Academy of Sciences of the United States of America*, 97,pp.8233-8238.

Chen, W.K., Wang, Y., Gao, X.T., Yang, X.H., He, F., Duan, C.Q. and Wang, J., 2018. Flavonoid and aromatic profiles of two Vitis vinifera L. teinturier grape cultivars. *Australian journal of grape and wine research*, 24(3), pp.379-389.

Chen, X., Tsiang, M., Yu, F., Hung, M., Jones, G. S and Zeynalzadegan, A. (2008). Modeling, analysis, and validation of a novel HIV integrase structure provide insights into the binding modes of potent integrase inhibitors. *Journal of Molecular Biology*, 380, pp.504-519.

Chiu, K. & Davies, R. (2004). Structure and function of HIV-1 integrase. *Current Topics in Medicinal Chemistry,* 4, pp.965-977.

Christ, F., Shaw, S., Demeulemeester, J., Desimmie, B., Marchand, A., Butler, S., Smets, W., Chaltin, P., Westby, M., Debyser, Z. and Pickford, C. (2012). Small-Molecule Inhibitors of the LEDGF/p75 Binding Site of Integrase Block HIV Replication and Modulate Integrase Multimerization. *Antimicrobial Agents and Chemotherapy*, 56(8), pp.4365-4374.

Chuang, J. and Van Wart, E., (1992). Resonance Raman spectra of horseradish peroxidase and bovine liver catalase compound I species. Evidence for predominant 2A2u pi-cation radical ground state configurations. *Journal of Biological Chemistry*, 267(19), pp.13293-13301.

Cicenas, J., Kalyan, K., Sorokinas, A., Stankunas, E., Levy, J., Meskinyte, I., Stankevicius, V., Kaupinis, A. and Valius, M., (2015). Roscovitine in cancer and other diseases. *Annals of translational medicine*, 3 pp.10.

Cihlar, T. and Ray, S. (2010). Nucleoside and nucleotide HIV reverse transcriptase inhibitors: 25 years after zidovudine. *Antiviral Re*s. 85(1), pp.39-58.

Clavel, F. & Hance, A. J. (2004). HIV drug resistance. *N.Engl.J.Med*., 350, pp.1023-1035.

Collier, J.; Ali, R.; Lear, T. Ingenol mebutate: A novel treatment for actinic keratosis. *Clin. Pract.* 2014, 11, pp.295–306.

Cooper, A., Steigbigel, T., Gatell, M., Rockstroh, K., Katlama, C and Yeni, P. (2008). Subgroup and resistance analyses of raltegravir for resistant HIV-1 infection. *New England Journal of Medicine,* 359, pp.355-365.

Costin, M. (2007).Cytopathic mechanisms of HIV-1. *Virol*.J. 4, pp.100.

Coyle, M., (2004). The effect of environment on breast cancer risk. *Breast cancer research and treatment,* 84(3), pp.273-288.

Cragg, M., (1998). Paclitaxel (Taxol): A Success Story With Valuable Lessons for Natural Product Drug Discovery and Development. *Medicinal Research Reviews*, 18(5), pp. 315-331.

Crouch NR, Bangani V, Mulholland DA. (1999). Homoisoflavanones from three South African Scilla species. *Phytochemistry* 51: pp.943–946.

CRUK(2014).CRUK Cancer incidence statistics. [online]. Available from:http://www.cancerresearchuk.org/cancer-info/cancerstats/incidence/ [Accessed January 10, 2019].

Crum, F., Riffenburgh, H., Wegner, S., Agan, K., Tasker, A and Spooner, M.(2006). Comparisons of causes of death and mortality rates among HIV-infected persons:

analysis of the pre-, early, and late HAART (highly active antiretroviral therapy) eras. *J.Acquir.Immune Defic.Syndr*., 41, pp.194-200.

Dadgar, S.; Ramjan, Z.; Floriano, W.B. (2013) Paclitaxel is an inhibitor and its boron dipyrromethene derivative is a fluorescent recognition agent for botulinum neurotoxin subtype A. *J. Med. Chem*. 56, pp.2791–2803.

Dai Z, Wang F, Wang GL, Lin RC (2006). Studies on chemical constituents of Balanophora spicata China. J. *China Mat. Med*., 21: pp.1798-800.

Dalgleish, G., Beverley, C., Clapham, R., Crawford, H., Greaves, F., and Weiss, A. (1984). The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature*, 312, pp.763-767.

De Silva, I., Cotten, M., and Rowland-Jones, L.(2008). HIV-2: the forgotten AIDS virus. Trends *Microbiol*. 16(12), pp.588-595.

De, L., Pedretti, A., Vistoli, G., Barreca, L., Villa, L and Monforte, P. (2003). Analysis of the full-length integrase-DNA complex by a modified approach for DNA docking. *Biochem.Biophys.Res.Commun*., 310, pp.1083-1088.

DeAnda, F., Hightower, K., Nolte, R., Hattori, K., Yoshinaga, T., Kawasuji, T. and Underwood, M. (2013). Dolutegravir Interactions with HIV-1 Integrase-DNA: Structural Rationale for Drug Resistance and Dissociation Kinetics. *PLoS ONE*, 8(10), pp. (e) 77448.

Delelis, O., Malet, I., Na, L., Tchertanov, L., Calvez, V, and Marcelin, G. (2009). The G140S mutation in HIV integrases from raltegravir-resistant patients rescues catalytic defect due to the resistance Q148H mutation. *Nucleic Acids Research*, 37, pp.1193-1201.

Demirayak, S., Yurttas, L., Gundogdu-Karaburun, N., Karaburun, A.C. and Kayagil, I., (2017). New chroman-4-one/thiochroman-4-one derivatives as potential anticancer agents. *Saudi Pharmaceutical Journal*, 25(7), pp.1063-1072.

Derat, E. and Shaik, S. (2006). Two-state reactivity, electromerism, tautomerism, and "surprise" isomers in the formation of Compound II of the enzyme horseradish peroxidase from the principal species, Compound I. *Journal of the American Chemical Society*, 128(25), pp.8185-8198.

Derat, E., Cohen, S., Shaik, S., Altun, A. and Thiel, W. (2005). Principal active species of horseradish peroxidase, compound I: a hybrid quantum mechanical/molecular mechanical study. *Journal of the American Chemical Society*, 127(39), pp.13611-13621.

DeSantis CE, Bray F, Ferlay J, Lortet-Tieulent J, Anderson BO, Jemal A (2015) International variation in female breast cancer incidence and mortality rates. *Cancer Epidemiol Biomarkers Prev* 24(10) pp.1495–1506.

Dhankhar R, Vyas SP, Jain AK, Arora S, Rath G and Goyal. (2010) AK Cells Blood *Substit Immobil Biotechnol.* 38(5) pp.230-49.

Doitsh, G., Cavrois, M., Lassen, K., Zepeda, O., Yang, Z., Santiago, M., Hebbeler, A. and Greene, W. (2010). Abortive HIV Infection Mediates CD4 T Cell Depletion and Inflammation in Human Lymphoid Tissue. *Cell*, 143(5), pp.789-801.

Dolan, J., Chen, A. P., Weber, I. T., Harrison, R. W., and Leis, J. (2009). Defining the DNA Substrate Binding Sites on HIV-1 Integrase. *Journal of Molecular Biology*, 385, pp.568-579.

Drelich, M., Wilhelm, R. and Mous, J., (1992). Identification of amino acid residues critical for endonuclease and integration activities of HIV-1 IN protein in vitro. *Virology,* 188(2), pp.459-468.

Dunford, H.B., 2010. Peroxidases and catalases: biochemistry, biophysics, biotechnology and physiology. *John Wiley & Sons.* pp. 210-250

Dybul, M., Fauci, S., Bartlett, G., Kaplan, E., & Pau, K. (2002). Guidelines for using antiretroviral agents among HIV-infected adults and adolescents. *Ann.Intern.Med.,* 137, pp. 381-433.

Dyda, F., Hickman, B., Jenkins, M., Engelman, A., Craigie, R., & Davies, R. (1994). Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science,* 266, pp.1981-1986.

Ehrlich, A.; Booher, S.; Becerra, Y.; Borris, L.; Figg, D.; Turner, L.; Blauvelt, A. (2004). Micellar paclitaxel improves severe psoriasis in a prospective phase II pilot study. J. Am. Acad. *Dermatol.* 2004, 50, pp.533–540.

Eijkelenboom, A., Puras Lutzke, R., Boelens, R., Plasterk, R., Kaptein, R. and Hård, K. (1995). The DNA-binding domain of HIV-1 integrase has an SH3-like fold. *Nature Structural & Molecular Biology*, 2(9), pp.807-810.

El-Elimat, T., Rivera-Chávez, J., Burdette, J., Czarnecki, A., Alhawarri, M., Al-Gharaibeh, M., Alali, F. and Oberlies, N. (2018). Cytotoxic homoisoflavonoids from the bulbs of Bellevalia flexuosa. *Fitoterapia*, 127, pp.201-206.

Ellis P., Schnitt J., Sastre-Garau X., Bussolati G., Tavassoli A., Eusebi V., Peterse L., Mukai K., Tabar L., Jacquemier J., Cornelisse J., Sasco J., Kaaks R., Pisani P., Goldgar E., Devilee P., Cleton-Jansen J., Borresen-Dale L., van't Veer L., Sapino A., (2003). Invasive breast carcinoma. In Tavassoli F.A., editor; , Devilee P., editor. (Eds.), *WHO* Classification of Tumours Pathology and Genetics of Tumours of the Breast and Female Genital Organs. Lyon Press.

Ellison, V., Gerton, J., Vincent, K. A., and Brown, O. (1995). An essential interaction between distinct domains of HIV-1 integrase mediates assembly of the active multimer. *J.Biol.Chem.,* 270, pp.3320-3326.

Engelman, A. and Cherepanov, P. (2012). The structural biology of HIV-1: mechanistic and therapeutic insights. *Nature Reviews Microbiology*, 10(4), pp.279-290.

Escobar-Ramos, A., Lobato-García, C., Zamilpa, A., Gómez-Rivera, A., Tortoriello, J. and González-Cortazar, M. (2017). Homoisoflavonoids and Chalcones Isolated from Haematoxylum campechianum L., with Spasmolytic Activity. *Molecules,* 22(9), p.1405.

Espeseth, A. S., Felock, P., Wolfe, A., Witmer, M., Grobler, J and Anthony, N. (2000). HIV-1 integrase inhibitors that compete with the target DNA substrate define a unique strand transfer conformation for integrase. *Proceedings of the National Academy of Sciences of the United States of America*, 97, pp.11244-11249.

Esposito, D. & Craigie, R. (1998). Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interaction. *EMBO J.*, 17, 5832-5843.

Estari, M., Venkanna, L., Sripriya, D., & Lalitha, R. (2012). Human immunodeficiency virus (HIV-1) reverse transcriptase inhibitory activity of Phyllanthus emblica plant extract. *Biologie et Médecine*, 4, pp.178–182.

Esté, J. and Telenti, A. (2007). HIV entry inhibitors. *The Lancet*, 370(9581), pp.81-88.

Evans, A.E.; Farber, S.; Brunet, S.; Mariano, P.J. (1963). Vincristine in the treatment of acute leukaemia in children. *Cancer*, 16, pp.1302–1306

Fan, H., Liang, Y., Jiang, B., Li, X., Xun, H., Sun, J., He, W., Lau, H.T. & Ma, X., (2016). Curcumin inhibits intracellular fatty acid synthase and induces apoptosis in human breast cancer MDA-MB-231 cells. *Oncology reports*, 35(5), pp. 2651-2656.

Fanales-Belasio, E., Raimondo, M., Suligoi, B. and Buttò, S. (2010). HIV virology and pathogenetic mechanisms of infection: a brief overview. *Annali dell'Istituto Superiore di Sanità*, 46(1). Pp 5-14.

Faure, A., Calmels, C., Desjobert, C., Castroviejo, M., Caumont-Sarcos, A., Tarrago-Litvak, L. et al. (2005). HIV-1 integrase crosslinked oligomers are active in vitro. *Nucleic Acids Res*., 33, pp.977-986.

Fesen, m., Pommier, Y., Leteurtre, F., Hiroguchi, S., Yung, J. and Kohn, K. (1994) Inhibition of HIV-1 integrase by flavones, caffeic acid phyenethyl ester (CAPE) and related compounds. *Biochemical Pharacology*, 48(3), pp.595-608

Flexner, C. (1998). HIV-protease inhibitors. *New England Journal of Medicine*. 338(18), pp.1281-1292.

Forssmann, W., The, Y., Stoll, M., Adermann, K., Albrecht, U., Tillmann, H., Barlos, K., Busmann, A., Canales-Mayordomo, A., Gimenez-Gallego, G., Hirsch, J., Jimenez-Barbero, J., Meyer-Olson, D., Munch, J., Perez-Castells, J., Standker, L., Kirchhoff, F. and Schmidt, R. (2010). Short-Term Monotherapy in HIV-Infected Patients with a Virus Entry Inhibitor Against the gp41 Fusion Peptide. *Science Translational Medicine*, 2(63) pp.360-369.

Fransen S, Gupta S, Danovich R, Hazuda D, Miller M, Witmer M, Petropoulos CJ, Huang W. (2009). Loss of raltegravir susceptibility by human immunodeficiency virus type 1 is conferred via multiple nonoverlapping genetic pathways. *J Virol* 83 pp11440 – 11446.

Fry, S.C. and Miller, J.G. (1989). Toward a working model of the growing plant cell wall: Phenolic cross-linking reactions in the primary cell walls of dicotyledons.

Galati, G. and O'Brien, P.J. (2004) Potential toxicity of flavonoids and other dietary phenolics: significance for their chemopreventive and anticancer properties. *Free Radical Biology and Medicine*. 37(3), pp.287–303.

Gali-Muhtasib, H.; Hmadi, R.; Kareh, M.; Tohme, R.; Darwiche, N. (2015) Cell death mechanisms of plant-derived anticancer drugs: Beyond apoptosis. *Apoptosis*. 20, pp.1531–1562.

Gallo, C., Salahuddin, Z., Popovic, M., Shearer, M., Kaplan, M and Haynes, F. (1984). Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science,* 224, pp.500-503.

Gallo, C., Sarin, S., Gelmann, P., Robert-Guroff, M., Richardson, E and Kalyanaraman, V. (1983). Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). *Science,* 220, pp.865-867.

Gallo, C, Blattner, A., Kalyanaraman, S., Robert-Guroff, M., Lister, A., Galton, A., Sarin, S., Crawford, H., Catovsky, D., and Greaves, M. (1982). The human type-C retrovirus, HTLV, in blacks from the Caribbean region, and relationship to adult T-cell leukemia/lymphoma. *International Journal of Cancer*, *30*(3), pp.257-264.

Geissmann, T. and Neukom, H. (1971). Vernetzung von Phenolcarbonsäureestern von Polysacchariden durch oxydative phenolische *Kupplung. Helvetica Chimica Acta,* 54(4), pp.1108-1112.

GLOBOCAN (2015) World Heath Organization International Agency for Research on Cancer GLOBOCAN 2008. [online]. Available from: http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx [Accessed January 10, 2019].

Goethals O, Clayton R, Van Ginderen M, Vereycken I, Wagemans E, Geluykens P, Dockx K, Strijbos R, Smits V, Vos A, Meersseman G, Jochmans D, Vermeire K, Schols D, Hallenberger S, Hertogs K. (2008). Resistance mutations in human immunodeficiency virus type 1 integrase selected with elvitegravir confer reduced susceptibility to a wide range of integrase inhibitors. *J Virol* 82 pp.10366 –10374.

Gohlke H, Klebe G. (2002) Approaches to the description and prediction of the binding affinity of smallmolecule ligands to macromolecular receptors. Angew Chem *Int Ed Engl*. 41(15) pp.2644– 2676.

Goldgur, Y., Craigie, R., Cohen, G., Fujiwara, T., Yoshinaga, T., Fujishita, T., Sugimoto, H., Endo, T., Murai, H. and Davies, D. (1999). Structure of the HIV-1 integrase catalytic domain complexed with an inhibitor: A platform for antiviral drug design. *Proceedings of the National Academy of Sciences*, 96(23), pp.13040-13043.

Goldgur, Y., Dyda, F., Hickman, B., Jenkins, M., Craigie, R., and Davies, R. (1998). Three new structures of the core domain of HIV-1 integrase: An active site that binds magnesium. *Proceedings of the National Academy of Sciences of the United States of America,* 95, pp. 9150-9154.

Gordon P, Griffith R and Keller A. (2007). Control of HIV through the inhibition of HIV-1 integrase: A medicinal chemistry perspective. *Journal of Medicinal Chemistry*. (3) pp.199- 220.

Greenwald, J., Le, V., Butler, L., Bushman, D., and Choe, S. (1999). The mobility of an HIV-1 integrase active site loop is correlated with catalytic activity. *Biochemistry*, 38, pp. 8892-8898.

Guiot, E., Carayon, K., Delelis, O., Simon, F., Tauc, P and Zubin, E (2006). Relationship between the oligomeric status of HIV-1 integrase on DNA and enzymatic activity. *Journal of Biological Chemistry*., 281, pp.22707-22719.

Guo, C., Eckler, M., McKenna, W., McKinsey, G., Rubenstein, J. and Chen, B. (2013). Fezf2 Expression Identifies a Multipotent Progenitor for Neocortical Projection Neurons, Astrocytes, and Oligodendrocytes. *Neuron*, 80(5), pp.1167-1174.

Hajimahdi, Z. and Zarghi, A. (2016). Progress in HIV-1 integrase inhibitors: A review of their chemical structure diversity. *Iranian journal of pharmaceutical research: IJPR*, *15*(4), pp.595.

Hamel, J., Sankale, L., Eisen, G., Meloni, S. T., Mullins, C., Gueye-Ndiaye, A., Mboup, S., and Kanki, P. J. (2007). Twenty years of prospective molecular epidemiology in Senegal: changes in HIV diversity. AIDS *Research of Human Retroviruses* 23(10), pp.1189-1196.

Hare, S., Gupta, S., Valkov, E., Engelman, A., & Cherepanov, P. (2010). Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature,* 464, pp.232-236.

Hare, S., Vos, A. M., Clayton, R. F., Thuring, J. W., Cummings, M. D., & Cherepanov, P. (2010). Molecular mechanisms of retroviral integrase inhibition and the evolution of viral resistance. *Proc.Natl.Acad.Sci.U.S.A,* 107, 20057-20062.

Haris, P.; Mary, V.; Aparna, P.; Dileep, K.V and Sudarsanakumar, C. (2017). A comprehensive approach to ascertain the binding mode of curcumin with DNA. Spectrochim. Acta A Mol. *Biomol. Spectrosc*., 175, pp.155–163.

Harper, E., Marselle, M., Gallo, C., and Wong-Staal, F. (1986). Detection of lymphocytes expressing human T-lymphotropic virus type III in lymph nodes and peripheral blood from infected individuals by in situ hybridization. *Proc.Natl.Acad.Sci.U.S.A,* 83, pp.772-776.

Hartman, T. and Buckheit, R. (2012). The Continuing Evolution of HIV-1 Therapy: Identification and Development of Novel Antiretroviral Agents Targeting Viral and Cellular Targets. *Molecular Biology International,* pp.1-17.

Havsteen, B. (2002). The biochemistry and medical significance of the flavonoids. *Pharmacology & Therapeutics*, 96(2-3), pp.67-202.

Hazuda J, Anthony J, Gomez P, Jolly M, Wai S, Zhuang L, Fisher E, Embrey M, Guare J, Egbertson S, Vacca P, Huff R, Felock J, Witmer V, Stillmock A, Danovich R, Grobler J, Miller D, Espeseth A, Jin L, Chen W, Lin H, Kassahun K, Ellis D, Wong K, Xu W, Pearson G, Schleif A, Cortese R, Emini E, Summa V, Holloway K, Young D: (2014) A naphthyridine carboxamide provides evidence for discordant resistance between mechanistically identical inhibitors of HIV-1 integrase. *Proc Natl Acad Sci USA* 31 pp.11233-11238

Hazuda J, Felock P, Witmer M, Wolfe A, Stillmock K, Grobler A, Espeseth A, Gabryelski L, Schleif W, Blau C, Miller D. (2000). Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. *Science* 287 pp.646 – 650.

Hazuda, J., Young, D., Guare, P., Anthony, J., Gomez, P andWai, S. (2004). Integrase inhibitors and cellular immunity suppress retroviral replication in rhesus macaques. *Science*, 305, pp.528-532

He Z, Yan F, Song J, Ye F, Liao X, Peng L, Ding S (2009). Chemical constituents from the aerial parts of Artemisia minor. *Journal of Natural Products*, (6) pp.198-1201.

Heeney, L., Dalgleish, G., and Weiss, A. (2006). Origins of HIV and the evolution of resistance to AIDS. *Science*, 313 pp. 462-466.

Hixson, L., Sleep, R., Capone, L., Elsey, M., Curtin, D., Sefton, A. and Taylor, K., (2012). Hydroxycinnamic acid ethyl esters as precursors to ethylphenols in wine. *Journal of agricultural and food chemistry*, *60*(9), pp.2293-2298.

Huff, J. (1991). HIV protease: a novel chemotherapeutic target for AIDS. *Journal of Medicinal Chemistry*, 34(8), pp.2305-2314.

Hwang, E.I., Yun, S., Kim, K., Kwon, B., Kim, H., Lee, H., Jeong, E. and Kim, S., (2000). Phellinsin A, a novel chitin synthases inhibitor produced by Phellinus sp. PL3. *The Journal of antibiotics*, *53*(9), pp.903-911.

Hwang, E., Yun, B., Kim, Y., Kwon, B., Kim, H., Lee, H., Jeong, W and Kim S.(2000) Phellinsin A, a novel chitin synthases inhibitor produced by *Phellinus sp.* PL3. J. *Antibiot.* 53, pp.903-911.

Ichimura, T., Otake, T., Mori, T. and Maruyama, S. (1999). HIV-1 Protease Inhibition and Anti-HIV Effect of Natural and Synthetic Water-soluble Lignin-like Substances. *Bioscience, Biotechnology, and Biochemistry*, 63(12), pp.2202-2204.

Iloeje, U., Yuan, Y., L'Italien, G., Mauskopf, J., Holmberg, S., Moorman, A., Wood, K. and Moore, R. (2005). Protease inhibitor exposure and increased risk of cardiovascular disease in HIV-infected patients. *HIV Medicine*, 6(1), pp.37-44.

Imran, M., Saeed, F.,Nadeem, M.; Arshad, U., Ullah, A.; Suleria, H. (2016) Cucurmin, anticancer and antitumor perspectives—A comprehensive review. Crit. Rev. *Food Sci. Nutr.*, 22, pp.1–23.

Ivancich, A., Jouve, H., Sartor, B. and Gaillard, J., (1997). EPR investigation of compound I in Proteus mirabilis and bovine liver catalases: formation of porphyrin and tyrosyl radical intermediates. *Biochemistry*, 36(31), pp.9356-9364

Jacob V, Hagai T and Soliman K. (2011). Structure-activity relationships of flavonoids. *Curr Org Chem*;15, pp.2641–57.

Jaganath IB, Crozier A. (2010) Dietary flavonoids and phenolic compounds. In: Fraga CG, editor Plant phenolics and human health: biochemistry, nutrition, and pharmacology . *Hoboken (NJ): Wiley*. p. 1–49.

Janssen, S., Käsmann, L., Fahlbusch, F., Rades, D. and Vordermark, D. (2017). Side effects of radiotherapy in breast cancer patients. *Strahlentherapie und Onkologie*, 194(2), pp.136-142.

Jiang, R., Lau, K., Hon, P., Mak, T., Woo, K. and Fung, K. (2005). Chemistry and Biological Activities of Caffeic Acid Derivatives from Salvia miltiorrhiza. *Current Medicinal Chemistry*, 12(2), pp.237-246.

Johnson, A., Marchand, C., Patil, S., Costi, R., Di Santo, R., Burke, T. and Pommier, Y. (2006). Probing HIV-1 Integrase Inhibitor Binding Sites with Position-Specific Integrase-DNA Cross-Linking Assays. *Molecular Pharmacology*, 71(3), pp.893-901.

Jorgensen, W. (2004). The many roles of computation in drug discovery. *Science*, 303(5665), pp.1813-1818.

Karamać M., Buciński A., Pegg R, Amarowicz R. (2005). Antioxidant and antiradical activity of ferulates. *Czech J. Food Sci*., 23: pp.64–68.

Karki, R., Tang, Y., Burke, T., and Nicklaus, M. (2004). Model of full-length HIV-1 integrase complexed with viral DNA as template for anti-HIV drug design. *Journal of Computer-Aided Molecular Design*, 18, pp.739-760.

Katz L and Baltz, R. (2016). Natural product discovery: Past, present and future. J. Ind. Microbiol. *Biotechnol.* 43, 155–176.

Khalid, E.B.; Ayman, E.E.; Rahman, H.; Abdelkarim, G.; Najda, A. Natural products against cancer angiogenesis. *Tumor Biol*. 2016, 37, 14513–14536.

Kikukawa, K., Nagira, K., Wada, F. and Matsuda, T., (1981). Reaction of diazonium salts with transition metals—III: Palladium (0)-catalyzed arylation of unsaturated compounds with arenediazoium salts. *Tetrahedron*, 37(1), pp.31-36.

Kim, E.; Lee, H, Hwang, E., Kim, S., Lee, W., Lee, S.; Jung, S. (2005) Stereochemistry of phellinsin A: A concise synthesis of α-arylidene-λ-lactones. *Synth. Commun*., 35, pp. 1231-1238.

Kim, K.,Woo, K., Moon, E., Choi, S., Kim, S., Choi, S and Lee, K. R. (2014). Identification of antitumor lignans from the seeds of Morning Glory (*Pharbitis nil*). *J. Agric. Food Chem.* 62, pp. 7746-7752. *Compound **4** in this paper.*

Kim, S, Hwang, K and Choi, K,. (2016). Treatment with kaempferol suppresses breast cancer cell growth caused by estrogen and triclosan in cellular and xenograft breast cancer models. *The Journal of nutritional biochemistry,* 28, pp.70-82.

Kim, S., Perera, R., Hager, L., Dawson, J. and Hoffman, B., (2006). Rapid freeze-quench ENDOR study of chloroperoxidase compound I: the site of the radical. *Journal of the American Chemical Society,* 128(17), pp.5598-5599.

Kitchen D, Decornez H, Furr J and Bajorath J. (2004). Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov.*; 3(11) pp.935–949.

Kocaadam, B.; ¸Sanlier, N. (2017) Curcumin, an active component of turmeric (Curcuma longa), and its effects on health. *Crit. Rev. Food Sci. Nutr.*, 57, pp.2889–2895.

Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A., and Steitz, T. A. (1992). Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* 256(5065), pp.1783-1790.

Korber, B., Muldoon, M., Theiler, J., Gao, F., Gupta, R., Lapedes, A. (2000). Timing the ancestor of the HIV-1 pandemic strains. *Science,* 288, pp.1789-1796.

Kotoku, N.; Arai, M.; Kobayashi (2016). Search for anti-angiogenic substances from natural sources. *Chem. Pharm. Bull*., 64, pp.128–134.

Krainer, F.W. and Glieder, A., (2015). An updated view on horseradish peroxidases: recombinant production and biotechnological applications. Applied microbiology and biotechnology, 99(4), pp.1611-1625.

Kumar, C. (2011). Parametric Optimization of Feruloyl Esterase Production from Aspergillus terreus Strain GA2 Isolated from Tropical Agro-Ecosystems Cultivating Sweet Sorghum. *Journal of Microbiology and Biotechnology,* 21(9), pp.947-953.

Kumar, G.; Mittal, S.; Sak, K.; Tuli, H. (2016) Molecular mechanisms underlying chemopreventive potential of curcumin: Current challenges and future perspectives. *Life Sci*., 148, pp.313–328.

Kunnumakkara, A.B.; Bordoloi, D.; Harsha, C.; Banik, K.; Gupta, S.(2017). Curcumin mediates anticancer effects by modulating multiple cell signaling pathways. *Clin. Sci*. 2017, 131, pp.1781–1799.

Lane, H. and Fauci, A. (1985). Immunologic abnormalities in the acquired immunodeficiency syndrome. *Annu.Rev.Immunol.,* 3, pp.477-500

Langer T, Hoffmann R. (2001) Virtual screening: an effective tool for lead structure discovery? *Curr Pharm Des*. 7(7): pp.509–527.

Le Bail, J., Laroche, T., Marre-Fournier, F. and Habrioux, G. (1998). Aromatase and 17β-hydroxysteroid dehydrogenase inhibition by flavonoids. *Cancer Letters*, 133(1), pp.101-106.

Lee, M., Deng, J., Briggs, J., & Duan, Y. (2005). Large-scale conformational dynamics of the HIV-1 integrase core domain and its catalytic loop mutants. *Biophysical Journal*, 88, pp.3133- 3146.

Lee, Y., Cho, J., Kim, C., Lee, S., Kim, W., Jeon, T., Park, K. and Moon, J. (2013). Coumaroyl quinic acid derivatives and flavonoids from immature pear (Pyrus pyrifolia nakai) fruit. *Food Science and Biotechnology*, 22(3), pp.803-810.

Lee-Huang, S., Zhang, H., Young-Tae, C., & Paul, L. (2003). Anti-HIV activity of olive leaf extracts (OLE) and modulation of host cell gene expression by HIV-1infection and OLE treatment. *Biochemical and Biophysical Research Communications,* 307, pp.1029–1037.

Li, N, Shi, Z., Tang, Y., Li, B. and Duan, J., (2009). Highly efficient esterification of ferulic acid under microwave irradiation. *Molecules*, *14*(6), pp.2118-2126.

Li, Y., Xuan, S., Feng, Y. and Yan, A. (2015). Targeting HIV-1 integrase with strand transfer inhibitors. *Drug Discovery Today*, 20(4), pp.435-449.

Liang, X.; Grue-Sørensen, G.; Månsson, K.; Vedsø, P.; Soor, A.; Stahlhut, M.; Bertelsen, M.; Engell, K.M.; Högberg, T. (2013) Syntheses, biological evaluation and SAR of ingenol mebutate analogues for treatment of actinic keratosis and non-melanoma skin cancer. *Bioorg. Med. Chem. Lett*., 23, pp.5624–5629.

Liao, C., Marchand, C., Burke Jr, T., Pommier, Y. and Nicklaus, M. (2010). Authentic HIV-1 integrase inhibitors. *Future Medicinal Chemistry*, 2(7), pp.1107-1122.

Lin, L., Liu, Q. & Ye, Y., 2014. Naturally occurring homoisoflavonoids and their pharmacological activities.. *Planta medica*, 80(13), pp. 1053-1066.

Lind, M., (2008). Principles of cytotoxic chemotherapy. *Medicine,* 36(1), pp. 19-23

Lins, R. D., Adesokan, A., Soares, T. A., & Briggs, J. M. (2000). Investigations on human immunodeficiency virus type 1 integrase/DNA binding interactions via molecular dynamics and electrostatics calculations. *Pharmacology & Therapeutics*, 85, pp.123-131.

Lins, R. D., Briggs, J. M., Straatsma, T. P., Carlson, H. A., Greenwald, J and Choe, S. (1999). Molecular dynamics studies on the HIV-1 integrase catalytic domain. *Biophysical Journal*, 76, pp.2999-3011.

Lins, R. D., Straatsma, T. P., and Briggs, J. M. (2000). Similarities in the HIV-1 and ASV integrase active sites upon metal cofactor binding. *Biopolymers*, 53, pp.308-315.

Liu, A., Guo, H., Ye, M., Lin, Y., Sun, J., Xu, M. and Guo, D., (2007). Detection, characterization and identification of phenolic acids in Danshen using high-performance liquid chromatography with diode array detection and electrospray ionization mass spectrometry. *Journal of Chromatography A*, *1161*(1-2), pp.170-182.

Locher, C., Witvrouw, M., Bethune, M., Burch, M., Mower, H., Davis, H. Clerc, E. (1996). Antiviral activity of Hawaiian medicinal plants against human immunodeficiency virus type-1 (HIV-1). *Phytomedicine*, 2, pp.259–264.

Loizidou, E. Z., Zeinalipour-Yazdi, C. D., Christofides, T., & Kostrikis, L. G. (2009). Analysis of binding parameters of HIV-1 integrase inhibitors: Correlates of drug inhibition and resistance. *Bioorganic & Medicinal Chemistry*, 17, pp.4806-4818.

Lovell, S., Goryshin, I., Reznikoff, W. and Rayment, I. (2002). Two-metal active site binding of a Tn5 transposase synaptic complex. *Nature Structural Biology*, 9(4), pp.278-281.

Maignan, S., Guilloteau, J. P., Zhou-Liu, Q., Clement-Mella, C., & Mikol, V. (1998). Crystal structures of the catalytic domain of HIV-1 integrase free and complexed with its metal cofactor: High level of similarity of the active site with other viral integrases. *Journal of Molecular Biology*, 282, pp.359-368.

Malet I, Delelis O, Valantin MA, Montes B, Soulie C, Wirden M, Tchertanov L, Peytavin G, Reynes J, Mouscadet JF, Katlama C, Calvez V, Marcelin AG. (2008). Mutations associated with failure of raltegravir treatment affect integrase sensitivity to the inhibitor in vitro. *Antimicrob Agents Chemother* 52: pp.1351–1358.

March, R.E., Miao, X.S. and Metcalfe, C.D., (2004). A fragmentation study of a flavone triglycoside, kaempferol-3-O-robinoside-7-O-rhamnoside. *Rapid communications in mass spectrometry*, 18(9), pp.931-934.

Margot NA, Hluhanich RM, Jones GS, Andreatta KN, Tsiang M, McColl DJ, White KL, Miller MD. (2012). In vitro resistance selections using elvitegravir, raltegravir, and two metabolites of elvitegravir M1 and M4. *Antiviral Res* 93:pp.288 –296.

Marinello, J., Marchand, C., Mott, B., Bain, A., Thomas, C. and Pommier, Y. (2008). Comparison of Raltegravir and Elvitegravir on HIV-1 Integrase Catalytic Reactions and on a Series of Drug-Resistant Integrase Mutants. *Biochemistry*, 47(36), pp.9345-9354.

Mašterov, I., Suchý, V., Uhrín, D., Ubik, K., Grančaiová, Z. and Bobovnický, B. (1991). Homoisoflavanones and other constituents from Muscari racemosum. *Phytochemistry*, 30(2), pp.713-714.

Matsen, C. and Neumayer, L., (2013). Breast cancer: a review for the general surgeon. *JAMA surgery*, *148*(10), pp.971-980.

Maurin C, Bailly F and Cotelle P. (2003) Structure-activity relationships of HIV-1 integrase inhibitors-enzymeligand interactions. *Curr. Med. Chem*. 10: pp.1795- 810.

McColl, D. and Chen, X. (2010). Strand transfer inhibitors of HIV-1 integrase: Bringing IN a new era of antiretroviral therapy. *Antiviral Research*, 85(1), pp.101-118.

McColl, D. , Fransen, S., Gupta, S., Parkin, N., Margot, N and Chuck, S. (2007). Resistance and cross-resistance to first generation integrase inhibitors: insights from a Phase II study of elvitegravir (GS-9137). *Antiviral Therapy*, 12, pp.11.

McCormack, P. (2014). Dolutegravir: A Review of Its Use in the Management of HIV-1 Infection in Adolescents and Adults. *Drugs*, 74(11), pp.1241-1252.

Metifiot, M., Maddali, K., Naumova, A., Zhang, X., Marchand, C., & Pommier, Y. (2010). Biochemical and pharmacological analyses of HIV-1 integrase flexible loop mutants resistant to raltegravir. *Biochemistry*, 49, pp.3715-3722.

Michel, F., Crucifix, C., Granger, F., Eiler, S., Mouscadet, J. F., Korolev, S. et al. (2009). Structural basis for HIV-1 DNA integration in the human genome, role of the LEDGF/P75 cofactor. *EMBO* J, 28, pp.980-991.

Miller, M. D. & Hazuda, D. J. (2004). HIV resistance to the fusion inhibitor enfuvirtide: mechanisms and clinical implications. *Drug Resist.Updat.*, 7, pp.89-95.

Miller, M., Bor, Y. and Bushman, F. (1995). Target DNA capture by HIV-1 integration complexes. *Current Biology*, 5(9), pp.1047-1056.

Mishra, B.B.; Tiwari, V.K. (2011). Natural products: An evolving role in future drug discovery. *Eur. J. Med. Chem*., 46, pp.4769–4807.

Moitessier N, Englebienne P, Lee D, Lawandi J, Corbeil CR. (2008) Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. *Br J Pharmacol*. 153(Suppl 1): pp.7–26.

Molteni, V., Greenwald, J., Rhodes, D., Hwang, Y., Kwiatkowski, W and Bushman, F. (2001). Identification of a small-molecule binding site at the dimer interface of the HIV integrase catalytic domain. *Acta Crystallographica Section D-Biological Crystallography*, 57, pp.536-544.

Morgan, D. A., Ruscetti, F. and Gallo, R. (1976). Selective in vitro growth of T lymphocytes from normal human bone marrows. *Science*, 193, pp.1007-1008.

Mouscadet, J. F. & Tchertanov, L. (2009). Raltegravir: molecular basis of its mechanism of action. Eur.J.Med.Res., 14 *Suppl* 3, pp.5-16.

Mouscadet, J. F., Arora, R., Andre, J., Lambry, J. C., Delelis, O., Malet, I. (2009). HIV-1 IN alternative molecular recognition of DNA induced by raltegravir resistance mutations. *J.Mol.Recognit.*, 22, pp.480-494.

Mouscadet, J., Delelis, O., Marcelin, A. and Tchertanov, L. (2010). Resistance to HIV-1 integrase inhibitors: A structural perspective. *Drug Resistance Updates*, 13(4-5), pp.139-150.

Mukhtar, M., Arshad, M., Mahmood, A., Pomerantz, R., & Parveen, Z. (2008). Antiviral potentials of medicinal plants. *Virus Research*, 131, pp.111–120.

Mutanyatta J, Matapa BG, Shushu DD, Abegaz BM. (2003). Homoisoflavonoids and xanthones from the tubers of wild and in vitro regenerated Ledebouria graminifolia and cytotoxic activities of some of the homoisoflavonoids. *Phytochemistry*, 62: pp.797–804

Namikoshi, M., Nakata, H. and Saitoh, T. (1987). Homoisoflavonoids from Caesalpinia sappan. *Phytochemistry,* 26(6), pp.1831-1833.

Neamati N. (2002) Patented small molecule inhibitors of HIV1 integrase: A ten-year saga. Expert Opin. *Ther. Pat*. 12: 709-24.

Newman, D.J.; Cragg, G.M. (2016) Natural products as sources of new drugs from 1981 to 2014. *J. Nat. Prod.*, 79, 629–661.

Ng, A., Harvey, A., Parker, M., Smith, A. and Waldron, K. (1998). Effect of axidative coupling in the temal stability of texture and cell wall chemistry of beet root *Beta vulgaris*. *Journal of Agricultural and Food Chemistry,* 46(8), pp.3365-3370.

Nguyen, A.T., Fontaine, J., Malonne, H. and Duez, P., (2006). Homoisoflavanones from Disporopsis aspera. *Phytochemistry,* 67(19), pp.2159-2163.

Nijhuis, M., van Maarseveen, N. M., Lastere, S., Schipper, P., Coakley, E and Glass, B. (2007). A novel substrate-based HIV-1 protease inhibitor drug resistance mechanism. *PLoS.Med*., 4, pp.36.

Nobel Foundation., (2015), https://www.nobelprize.org/uploads/2018/07/annual_report_15.pdf [visited 03/04/2018]

Notka, F., Meier, G., & Wagner, R. (2004). Concerted inhibitory activities of Phyllanthus amarus on HIV replication in vitro and ex vivo. *Antiviral Research*, 64, pp.93–102.

Nunthaboot, N., Lugsanangarm, K., Kokpol, S. and Abd-Elazem, I.S., (2013). Binding mode prediction of biologically active compounds from plant Salvia Miltiorrhiza as integrase inhibitor. *Bioinformation,* 9(8), pp.426.

Nutan, M., Charlene, S., Shweta, K., Ajay, K., Sharad, K., & Swadesh, M. (2013). Extracts from Acacia catechu suppress HIV-1 replication by inhibiting the activities of the viral protease and Tat. *Virology Journal,* 10, pp.309–311.

O'Donnell, G., Bucar, F. and Gibbons, S. (2006). Phytochemistry and antimycobacterial activity of Chlorophytum inornatum. *Phytochemistry*, 67, pp.178-182.

Ochwang'I DO, Kimwele CN, Oduma JA, Gathumbi PK, Mbaria JM, Kiama SG. (2014) Medicinal plants used in treatment and management of cancer in Kakamega County Kenya. *Journal of Ethnopharmacology;* 151 pp.1040–1055

Ogbourne, S.M.; Parsons, P.G. (2014) The value of nature's natural product library for the discovery of new chemical entities: The discovery of ingenol mebutate. *Fitoterapia* , 98, pp.36–44.

Onifade, A. and Jewell, A. (2012). 5-Month Herbal Therapy and Complete Sero-Reversion with Recovery in an Adult HIV/AIDS Patient. *Journal of Antivirals & Antiretrovirals*, 01(1), pp.56-69.

Osmanov, S., Pattou, C., Walker, N., Schwardländer, B. and Esparza, J. (2002). Estimated Global Distribution and Regional Spread of HIV-1 Genetic Subtypes in the Year 2000. JAIDS Journal of *Acquired Immune Deficiency Syndromes,* 29(2), pp.184-190.

Panche, A., Diwan, A. and Chandra, S. (2016). Flavonoids: an overview. *Journal of Nutritional Science,* 5(11),pp.1139-1144.

Pannecouque, C., Pluymers, W., Van Maele, B., Tetz, V., Cherepanov, P., De Clercq, E., Witvrouw, M. and Debyser, Z., (2002). New class of HIV integrase inhibitors that block viral replication in cell culture. *Current Biology*, 12(14), pp.1169-1177.

Parienti, J., Massari, V., Descamps, D., Vabret, A., Bouvet, E., Larouze, B. and Verdon, R. (2004). Predictors of Virologic Failure and Resistance in HIV-Infected Patients Treated with Nevirapine- or Efavirenz-Based Antiretroviral Therapy. *Clinical Infectious Diseases*, 38(9), pp.1311-1316.

Parveen, I., Threadgill, MD., Hauck, B., Donnison, I. and Winters, A. (2011). Isolation, identification and quantitation of hydroxycinnamic acid conjugates, potential platform chemicals, in the leaves and stems of Miscanthus×giganteus using LC–ESI-MSn. *Phytochemistry*, 72(18), pp.2376-2384.

Pascal, O., Chikwelu, L., Eunice, I., & Simon, L. (2004). In vitro activity of three selected South African medicinal plants against human immunodeficiency virus type 1 reverse transcriptase. *African Journal of Biotechnology*, 3, pp.555–559.

Pavan, A.R.; Silva, G.D.; Jornada, D.H.; Chiba, D.E.; Fernandes, G.F.; Man Chin, C.; Dos Santos, J.L. Unraveling the anticancer effect of curcumin and resveratrol. *Nutrients* 2016, 8, pp.628.

Pendri, A., Meanwell, N., Peese, K. and Walker, M. (2011). New first and second generation inhibitors of human immunodeficiency virus-1 integrase. *Expert Opinion on Therapeutic Patents*, 21(8), pp.1173-1189.

Pereira, D., Valentão, P., Pereira, J. and Andrade, P. (2009). Phenolics: *From Chemistry to Biology. Molecules*, 14(6), pp.2202-2211.

Perkins, S., Verschoyle, R.D., Hill, K., Parveen, I., Threadgill, M.D., Sharma, R.A., Williams, M.L., Steward, W.P. and Gescher, A.J., (2002). Chemopreventive efficacy and pharmacokinetics of curcumin in the min/+ mouse, a model of familial adenomatous polyposis. *Cancer Epidemiology and Prevention Biomarkers*, 11(6), pp.535-540.

Pinheiro, P. and Justino, G. (2012) Structural Analysis of Flavonoids and Related Compounds – A Review of Spectroscopic Applications. In V. Rao, ed. Phytochemicals - A Global Perspective of Their Role in Nutrition and Health. *InTech*, p. 538.

Plieninger, H.; Arnold, L.; Hoffmann, W. (1965). Preparation and properties of cyclohexadienone and cyclohexadienol derivatives II. *Chem. Ber.* 98, pp.1765−1773.

Pommier, Y., Johnson, A. and Marchand, C. (2005). Integrase inhibitors to treat HIV/Aids. *Nature Reviews Drug Discovery*, 4(3), pp.236-248.

Popovic, M., Sarngadharan, M. G., Read, E., & Gallo, R. C. (1984). Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and preAIDS. *Science*, 224, pp.497-500.

Premanath, R., Sudisha, J., Devi, N. and Aradhya, S. (2011). Antibacterial and Anti-oxidant Activities of Fenugreek (Trigonella foenum graecum L.) Leaves. *Research Journal of Medicinal Plant*, 5(6), pp.695-705.

Purushothaman KK, Kalyani K, Subramaniam K, Shanmughanathan SP. (1982). Structure of Bonducellin – a new homoisoflavone from Caesalpinia bonducella. *Indian J Chem* B 21: p383.

Quan, F. S., Young, J. L., Richard, W. C., & Sang, M. K. (2015). Salviae and Cinnamomi herbal medicines have antiviral activity against a broad range of human immunodeficiency virus. *International Journal Virology AIDS*, 2(1), p13.

Rafi MM, Vastano BC, Ho CT, Rosen RT. (2002). Novel compound isolated from Polygonum odoratum induces Bcl-2 phosphorylation and apoptosis in cancer cell tines. *Faseb J* 16: pp.A743–A743.

Ralph, J. (2009). Hydroxycinnamates in lignification. *Phytochemistry Reviews*, 9(1), pp.65-83.

Ralph, J., Ede, R., Robinson, N. and Main, L. (1987). Reactions of ß-ARYL Lignin Model Quinone Methides with Anthrahydroquinone and Anthranol. *Journal of Wood Chemistry and Technology*, 7(2), pp.133-160.

Ralph, J., Grabber, J. and Hatfield, R. (1995). Lignin-ferulate cross-links in grasses: active incorporation of ferulate polysaccharide esters into ryegrass lignins. *Carbohydrate Research*, 275(1), pp.167-178.

Ralph, J., Helm, R., Quideau, S. and Hatfield, R. (1992). Lignin–feruloyl ester cross-links in grasses. Part 1. Incorporation of feruloyl esters into coniferyl alcohol dehydrogenation polymers. *Journal of the Chemical Society*., Perkin Trans. 1, (21), pp.2961-2969.

Ralph, J., Quideau, S., Grabber, J. and Hatfield, R. (1994). Identification and synthesis of new ferulic acid dehydrodimers present in grass cell walls. *Journal of the Chemical Society*, Perkin Transactions 1, (23), p.3485.

Ramsay, J.R.; Suhrbier, A.; Aylward, J.H.; Ogbourne, S.; Cozzi, S.J.; Poulsen, M.G.; Baumann, K.C.; Welburn, P.; Redlich, G.L.; Parsons, P.G.(2011). The sap from Euphorbia peplus is effective against human nonmelanoma skin cancers. *Br. J. Dermatol*, 164, pp.633–636.

Ramya, P.V.; Angapelly, S.; Guntuku, L.; Singh Digwal, C.; Nagendra Babu, B.; Naidu, V.G.; Kamal, A. (2017) Synthesis and biological evaluation of curcumin inspired indole analogues as tubulin polymerization inhibitors. Eur. *J. Med. Chem*., 127, pp.100–114.

Ratner, L., Haseltine, W., Patarca, R., Livak, K. J., Starcich, B and Josephs, S. F.(1985). Complete nucleotide sequence of the AIDS virus, HTLV-III. *Nature,* 313,pp. 277-284.

Ren, G., Gao, K., Bushman, F. D., & Yeager, M. (2007). Single-particle image reconstruction of a tetramer of HIV integrase bound to DNA. *Journal of Molecular Biology*, 366, pp.286-294.

Robertson, D., Anderson, J., Bradac, J., Carr, J., Foley, B., Funkhouser, R. (2000). HIV-1 nomenclature proposal. *Science*, 288, pp.55-56.

Roleira, F., Tavares-da-Silva, E., Varela, C., Costa, S., Silva, T., Garrido, J. and Borges, F. (2015). Plant derived and dietary phenolic antioxidants: Anticancer properties. *Food Chemistry*, 183, pp.235-258.

Rowland-Jones, S. L. and Whittle, H. C. (2007). Out of Africa: what can we learn from HIV-2 about protective immunity to HIV-1? *Nat.Immunol*. 8(4), pp.329-331.

Roy, S.K., Kumari, N., Gupta, S., Pahwa, S., Nandanwar, H. and Jachak, S.M., (2013). 7-Hydroxy-(E)-3-phenylmethylene-chroman-4-one analogues as efflux pump inhibitors against Mycobacterium smegmatis mc2 155. *European journal of medicinal chemistry*, 66, pp.499-507.

Samanta, A., Das, G. and Das, S.K., (2011). Roles of flavonoids in plants. carbon, 100(6). Complete book.

Sanchez, M., Pena, M., Revilla, G. and Zarra, I. (1996). Changes in Dehydrodiferulic Acids and Perozidase Activity against Ferulic acid associated with Cell walls during growth *Pinus pinaster* Hypocotyl. *Plant phydiology*, 111(3), pp.941-946.

Sanchez-Pescador, R., Power, M. D., Barr, P. J., Steimer, K. S., Stempien, M. M., Brown-Shimer, S. L. (1985). Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2). *Science*, 227, pp.484-492.

Sante-Appiah. and Skalka, A. (1999). HIV-1 integrase: Structural organization, conformational changes, and catalysis. *Advances in Virus Research*, Vol 52, 52, pp.351-369.

Sato, M., Motomura, T., Aramaki, H., Matsuda, T., Yamashita, M., Ito, Y. (2006). Novel HIV-1 integrase inhibitors derived from quinolone antibiotics. *Journal of Medicinal Chemistry*, 49, pp.1506- 1508.

Savarino A. (2006) A historical sketch of the discovery and development of HIV-1 integrase inhibitors. *Expert Opin Investig Drug*, 12: pp.507-522.

Savarino, A., (2007). In-Silico docking of HIV-1 integrase inhibitors reveals a novel drug type acting on an enzyme/DNA reaction intermediate. *Retrovirology*, *4*(1), p.21.

Schomberg, D., Salzmann, M., and Stephan, D., (1993*) Enzyme Handbook* 7, EC 1.11.1.7: pp1-6

Schwertheim, S.; Wein, F.; Lennartz, K.; Worm, K.; Schmid, K.W.; Sheu-Grabellus, S.Y. (2017). Curcumin induces G2/M arrest, apoptosis, NF-κB inhibition, and expression of differentiation genes in thyroid carcinoma cells. J. Cancer Res. Clin. Oncol., 143, pp.1143–1154.

Seca, A. and Pinto, D. (2018). Plant secondary metabolites as anticancer agents: successes in clinical trials and therapeutic application. *International journal of molecular sciences*, 19(1), p.263.

Semenova, E. A., Marchand, C., & Pommier, Y. (2008). HIV-1 integrase inhibitors: update and perspectives. *Adv.Pharmacol*., 56, 199-228.

Sension, M. G. (2007). Long-Term suppression of HIV infection: benefits and limitations of current treatment options. *J.Assoc.Nurses AIDS Care* 18(1 Suppl), S2-10.

Serova, M.; Ghoul, A.; Benhadji, K.A.; Faivre, S.; Le Tourneau, C.; Cvitkovic, E.; Lokiec, F.; Lord, J.; Ogbourne, S.M.; Calvo, F. (2008). Effects of protein kinase C modulation by PEP005, a novel ingenol angelate, on mitogen-activated protein kinase and phosphatidylinositol 3-kinase signaling in cancer cells. *Mol. Cancer Ther*. 7, pp.915–922.

Shaw, G. M., Hahn, B. H., Arya, S. K., Groopman, J. E., Gallo, R. C., & Wong-Staal, F. (1984). Molecular characterization of human T-cell leukemia (lymphotropic) virus type III in the acquired immune deficiency syndrome. *Science*, 226, pp.1165-1171.

Shim JS, Kim JH, Lee JY, Kim SN, Kwon HJ. (2004). Anti-angiogenic activity of a homoisoflavanone from Cremastra appendiculata. *Planta Med* 70: pp171–173.

Shimura K, Kodama E, Sakagami Y, Matsuzaki Y, Watanabe W, Yamataka K, Watanabe Y, Ohata Y, Doi S, Sato M, Kano M, Ikeda S, Matsuoka M. (2008). Broad antiretroviral activity and resistance profile of the novel human immunodeficiency virus integrase inhibitor elvitegravir (JTK-303/ GS-9137). *J Virol* 82 pp.764 –774.

Shoichet, BK.; McGovern, SL.; Wei, B.; Irwin, JJ. (2002) Hits, leads and artifacts from virtual and high throughput screening. Molecular Informatics: *Confronting Complexity*. 65 pp.436-439.

Siegel, R., Miller, K. & Jemal, A., 2018. Cancer statistics, (2018). *CA: a cancer journal for clinicians*, 68(1), pp. 7-30.

Silayo A, Ngadjui BT, Abegaz BM.(1999) Homoisoflavonoids and stilbenes from the bulbs of Scilla nervosa subsp rigidifolia. *Phytochemistry* 52: pp947–955.

Simon, L., Salam, A.A.A., Kumar, S.M., Shilpa, T., Srinivasan, K.K. and Byrappa, K., (2017). Synthesis, anticancer, structural, and computational docking studies of 3-benzylchroman-4-one derivatives. *Bioorganic & medicinal chemistry letters*, 27(23), pp.5284-5290.

Slovackova, J., Smarda, J. and Smardova, J. (2012). Roscovitine-induced apoptosis of H1299 cells depends on functional status of p53. *Neoplasma*, 59(06), pp.606-612.

Smith SJ, Zhao XZ, Burke TR, Jr, Hughes SH. (2018). HIV-1 integrase inhibitors that are broadly effective against drug-resistant mutants. *Antimicrob Agents Chemother* 62: pp.5-18

Smolov, M., Gottikh, M., Tashlitskii, V., Korolev, S., Demidyuk, I., Brochon, J. C. et al. (2006). Kinetic study of the HIV-1 DNA 3 '-end processing - Single-turnover property of integrase. *Febs Journal*, 273, pp.1137-1151.

Stevenson M: (2000). HIV nuclear import: What's the flap? *Nat Med*, 6. pp.626-628

Sun, J., Liang, F., Bin, Y., Li, P. and Duan, C., (2007). Screening non-colored phenolics in red wines using liquid chromatography/ultraviolet and mass spectrometry/mass spectrometry libraries. *Molecules*, *12*(3), pp.679-693.

Sun, L., Zhang, J., Lu, X., Zhang, L. and Zhang, Y., (2011). Evaluation to the antioxidant activity of total flavonoids extract from persimmon (Diospyros kaki L.) leaves. *Food and chemical toxicology*, 49(10), pp.2689-2696.

Tan, Z. and Shahidi, F. (2012). A novel chemoenzymatic synthesis of phytosteryl caffeates and assessment of their antioxidant activity. *Food Chemistry*, 133(4), pp.1427-1434.

Tawata, S., Taira, S., Kobamoto, N., Zhu, J., Ishihara, M. and Toyama, S., (1996). Synthesis and antifungal activity of cinnamic acid esters. *Bioscience, biotechnology, and biochemistry*, 60(5), pp.909-910.

The Royal Marsden NHS Trust (2013) A Beginner's Guide to BRCA1 and BRCA2.

Thierry, E., Deprez, E. and Delelis, O. (2017). Different Pathways Leading to Integrase Inhibitors *Resistance. Frontiers in Microbiology*, 7.

Thomson, M., Pérez-Álvarez, L. and Nájera, R. (2002). Molecular epidemiology of HIV-1 genetic forms and its significance for vaccine development and therapy. *The Lancet Infectious Diseases*, 2(8), pp.461-471.

Tilton, J. C. and Doms, R. W. (2010) Entry inhibitors in the treatment of HIV-1 infection. *Antiviral Res*. 85(1), pp.91-100.

Tsai YC, Chiang SY, El-Shazly M, Wu CC, Beerhues L, Lai WC, Wu SF, Yen MH, Wu YC, Chang FR. (2013). The oestrogenic and anti-platelet activities of dihydrobenzofuroisocoumarins and homoisoflavonoids from Liriope platyphylla roots. *Food Chem pp*.140: pp.305–314.

Tsiang M, Jones GS, Hung M, Samuel D, Novikov N, Mukund S. (2012) Dithiothreitol causes HIV-1 integrase dimer dissociation while agents interacting with the integrase dimer interface promote dimer formation. *Biochemistry*.50 (10):pp.1567-1581.

Tsiang, M., Jones, G., Goldsmith, J., Mulato, A., Hansen, D., Kan, E., Tsai, L., Bam, R., Stepan, G., Stray, K., Niedziela-Majka, A., Yant, S., Yu, H., Kukolj, G., Cihlar, T., Lazerwith, S., White, K. and Jin, H. (2016). Antiviral Activity of Bictegravir (GS-9883), a Novel Potent HIV-1 Integrase Strand Transfer Inhibitor with an Improved Resistance Profile. *Antimicrobial Agents and Chemotherapy*, pp.1474-16.

Tsiang, M., Jones, G., Niedziela-Majka, A., Kan, E., Lansdon, E., Huang, W., Hung, M., Samuel, D., Novikov, N., Xu, Y., Mitchell, M., Guo, H., Babaoglu, K., Liu, X., Geleziunas, R. and Sakowicz, R. (2012). New Class of HIV-1 Integrase (IN) Inhibitors with a Dual Mode of Action. *Journal of Biological Chemistry*, 287(25), pp.21189-21203.

Turashvili, G. and Brogi, E., (2017). Tumor heterogeneity in breast cancer. *Frontiers in medicine*, 4, p.227.

Tzogani, K.; Nagercoil, N.; Hemmings, R.J.; Samir, B.; Gardette, J.; Demolis, P.; Salmonson, T.; Pignatti, F. (2014) The European Medicines Agency approval of ingenol mebutate (Picato) for the cutaneous treatment of non-hyperkeratotic, non-hypertrophic actinic keratosis in adults: Summary of the scientific assessment of the Committee for Medicinal Products for Human Use (CHMP). *Eur. J. Dermatol*., 24, pp.457–463.

Uchil, P. D. and Mothes, W. (2009). HIV Entry Revisited. *Cell*, 137, pp.402-404.

Unaids.org. (2017). Global AIDS Update 2017 | UNAIDS. [online] Available at: http://www.unaids.org/en/resources/documents/2016/Global-AIDS-update-2017 [Accessed 11 Jul 2018].

Vajpeyi, R., (2005). WHO classification of tumours: pathology and genetics of tumours of the breast and female genital organs. *J. Clin. Pathol.* (2005);58:pp.670–672

Van der Loeff, M. F., Awasana, A. A., Sarge-Njie, R., van der Sande, M., Jaye, A., Sabally, S., Corrah, T., McConkey, S. J., and Whittle, H. C. (2006). Sixteen years of HIV surveillance in a West African research clinic reveals divergent epidemic trends of HIV-1 and HIV-2. *Int.J.Epidemiol*. 35(5), pp.1322-1328.

Venkanna, L., and Estari, A. (2012). Inhibition of human immunodeficiency virus (HIV-1) reverse transcriptase by Cassia occidentalis (L) plant extract. *International Journal Science Engineering Research,* 3, pp.2229–5518.

Voisin-Chiret, A.S., Bazin, M.A., Lancelot, J.C. and Rault, S., (2007). Synthesis of new L-ascorbic ferulic acid hybrids. *Molecules,* 12(11), pp.2533-2545.

Vukics, V., Ringer, T., Kery, A., Bonn, G.K. and Guttman, A., (2008). Analysis of heartsease (Viola tricolor L.) flavonoid glycosides by micro-liquid chromatography coupled to multistage mass spectrometry. *Journal of Chromatography A,* 1206(1), pp.11-20.

Wainberg, M., Zaharatos, G. and Brenner, B. (2011). Development of Antiretroviral Drug Resistance. *New England Journal of Medicine,* 365(7), pp.637-646.

Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S., & Alizon, M. (1985). Nucleotide sequence of the AIDS virus, LAV. *Cell,* 40, pp.9-17.

Wall ME, Wani MC, Manikumar G, Taylor H, Mcgivney R. (1989). Plant antimutagens. 6. Intricatin and intricatinol, new antimutagenic homoisoflavonoids from Hoffmanosseggia intricata. *J Nat Prod* 52: pp774–778.

Wall, M., Wani, M., Manikumar, G., Taylor, H. and McGivney, R. (1989). Plant Antimutagens, 6. Intricatin and Intricatinol, New Antimutagenic Homoisoflavonoids from Hoffmanosseggia intricata. *Journal of Natural Products*, 52(4), pp.774-778.

Walsh, V. and Goodman, J. (1999). Cancer chemotherapy, biodiversity, public and private property: the case of the anti-cancer drug Taxol. *Social Science & Medicine* , 49(9), pp. 1215-1225.

Walsh, V. & Goodman, J. (2002). From taxol to Taxol: the changing identities and ownership of an anti-cancer drug. *Medical Anthropology*, 21(3-4), pp. 307-336.

Walters WP, Stahl MT, Murcko MA. (1998) Virtual screening - an overview. *Drug Discov Today*.; 3:pp.160–178.

Wang, J. Y., Ling, H., Yang, W., and Craigie, R. (2001). Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein. *EMBO J*., 20, pp.7333-7343.

Wang, L. D., Liu, C. L., Chen, W. Z., and Wang, C. X. (2005). Constructing HIV-1 integrase tetramer and exploring influences of metal ions on forming integrase-DNA complex. Biochemical and Biophysical *Research Communications*, 337, pp.313-319.

Wang, X., Morris-natshe, S., and Lee, K. (2007). New developments in the chemistry and biology of the bioactive constituents of Tanshen. *ChemInform*, 38 p.19

Wang, X.; Song, Y.; Su, Y.; Tian, Q.; Li, B.; Quan, J.; Deng, Y. (2016) Are PEGylated liposomes better than conventional liposomes? A special case for vincristine. *Drug Deliv*. 23, pp.1092–1100.

Wang, Y., Lv, Z. and Chu, Y. (2015). HIV protease inhibitors: a review of molecular selectivity and toxicity. *HIV/AIDS - Research and Palliative Care*, p.95.

Weaver, B.A. (2014). How Taxol/paclitaxel kills cancer cells. Mol. Biol. *Cell* , 25, pp. 2677–2681.

Weber, W., Demirdjian, H., Lins, R. D., Briggs, J. M., Ferreira, R., and McCammon, J. A. (1998). Brownian and essential dynamics studies of the HIV-1 integrase catalytic domain. *Journal of Biomolecular Structure & Dynamics,* 16, pp.733-745.

Weigelt, B., Geyer, F.C. and Reis-Filho, J.S., (2010). Histological types of breast cancer: how special are they?. *Molecular oncology*, 4(3).pp.192-208.

Werbel, T. and Cohen, P. (2018). Persistent Alopecia in a Breast Cancer Patient Following Taxane Chemotherapy and Adjuvant Endocrine Therapy: Case Report and Review of Post-treatment Hair Loss in Oncology Patients with Breast Cancer. *Cureus.* 25, pp.369-375.

Wijitkosoom, A., Tonmunphean, S., Truong, T. N., & Hannongbua, S. (2006). Structural and dynamical properties of a full-length HIV-1 integrase: Molecular dynamics simulations. *Journal of Biomolecular Structure & Dynamics,* 23, pp.613-624.

Willcox, M. (2004). Traditional herbal medicines for malaria. *BMJ*, 329(7475), pp.1156-1159.

Williams, S. L. and Essex, J. W. (2009). Study of the Conformational Dynamics of the Catalytic Loop of WT and G140A/G149A HIV-1 Integrase Core Domain Using Reversible Digitally Filtered Molecular Dynamics. *Journal of Chemical Theory and Computation*, 5, pp.411-421.

World Health Organization. (2015). Breast cancer. [online] Available at: https://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/ [Accessed 19 Jun. 2018].

Xiang, M., Su, H., Hu, J. and Yan, Y., (2011). Isolation, identification and determination of methyl caffeate, ethyl caffeate and other phenolic compounds from Polygonum amplexicaule var. sinense. *Journal of Medicinal Plants Research, 5*(9), pp.1685-1691.

Xu, C., Chen, G., Fu, C. and Huang, X. (1995). The Wittig Reaction of Stable Ylide with Aldehyde Under Microwave Irradiation: Synthesis of Ethyl Cinnamates. *Synthetic Communications*, 25(15), pp.2229-2233.

Xuan, Z., Liu-Meng, Y., Guang-Ming, L., Ya-Juan, L., Chang-Bo, Z., & Yong- Jun, L. (2012). Potent anti-HIV activities and mechanisms of action of a pine cone extracts from Pinus yunnanensis. *Journal Molecula*r, 17, pp.6916– 6929.

Yamaoto, G. and Bokelman, N. (1989). Plant Cell Wall Polymers. *Analytical Chemistry*, 61(19), pp.1102A-1102A.

Yamaotsu, N. and Hirono, S. (2018). In silico fragment-mapping method: a new tool for fragment-based/structure-based drug discovery. *Journal of Computer-Aided Molecular Design*, 32(11), pp.1229-1245.

Yang L, Xu X, Huang Y, Zhang B, Zeng C, He H, Wang C and Hu L. (2010). Synthesis of polyhydroxylated aromatics having amidation of piperazine nitrogen as HIV-1 integrase inhibitor. *Bioorg. Med. Chem. Lett.* 20: pp.5469-5471

Ye, M., Guo, D., Ye, G. and Huang, C. (2005) Analysis of homoisoflavonoids in Ophiopogon japonicus by HPLC-DAD-ESI-MSn. *Journal of the American Society for Mass Spectrometry*. 16(2), pp.234–243.

Yoder, K. E. & Bushman, F. D. (2000). Repair of gaps in retroviral DNA integration intermediates. *J.Virol.,* 74, pp.11191-11200.

Yoshihara, H.; Yoshimoto, Y.; Hosoya, Y.; Hasegawa, D.; Kawano, T.; Sakoda, A.; Okita, H.; Manabe, A. (2017) Infantile fibrosarcoma treated with postoperative vincristine and dactinomycin. *Pediatr. Int.* 59, pp.371–374.

Zamocky, M., Furtmüller, P.G. and Obinger, C. (2008). Evolution of catalases from bacteria to humans. *Antioxidants & redox signaling*, 10(9), pp.1527-1548.

Zeier, M Nachega, J.B., Mugavero, M.J.,., Vitória, M. and Gallant, J.E. (2011). Treatment simplification in HIV-infected adults as a strategy to prevent toxicity, improve adherence, quality of life and decrease healthcare costs. *Patient preference and adherence*, *5*, p.357.

Zhang, B., Lv, C., Li, W., Cui, Z., Chen, D., Cao, F., Miao, F. and Zhou, L. (2015). Ethyl cinnamate derivatives as promising high-efficient acaricides against Psoroptes cuniculi: synthesis, bioactivity and structure-activity relationship. *Chemical and Pharmaceutical Bulletin*, pp.c14-00765.

Zhang, Y., Wang, X., Li, L., Li, W., Zhang, F., Du, T. and Chu, X. (2013). Simultaneous determination of 23 flavor additives in tobacco products using gas chromatography–triple quadrupole mass spectrometry. *Journal of Chromatography A*, *1306*, pp.72-79.

Zhao D, Zhang W, Li MJ Liu X. (2006). Studies on chemical constituents of Acroptilon repens. China J. Mat. Med., 22: 1869-72. Zhu JJ, Zhang CF, Zhang M, Wang ZT (2006). Studies on chemical constituents in roots of Rumex dentatus. *China J. Mat. Med.,* 20: pp.1691-1693.

Zhao H, Neamati N, Mazumder A, Sunder S, Pommier Y and Burke Jr TR. (1997) Arylamide inhibitors of HIV-1 integrase. *J. Med. Chem*. 40: pp.1186-94.

Zhou C., Zou L, Mo J., Wang X., Yang B, He QJ, Gan L. (2013). Homoisoflavonoids from Ophiopogon japonicus. *Helv Chim Acta*, 96: pp.1397– 1405.

Zhu, T., Korber, B. , Nahmias, A. J., Hooper, E., Sharp, P., & Ho, D. (1998). An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. *Nature,* 391, pp.594-597.

Zhuang, L., Wai, J., Embrey, M., Fisher, T., Egbertson, M and Payne, L. (2003). Design and synthesis of 8-hydroxy-[1,6]naphthyridines as novel inhibitors of HIV-1 integrase in vitro and in infected cells. *Journal of Medicinal Chemistry*, 46, pp.453-456.

Appendix: Supplementary information

Oxidative Heterodimerization Of 4'-Hydroxycinnamate Esters With 4'-Hydroxycinnamic Acids As Potential HIV-1 Integrase Inhibitors And The Identification Of Two Novel Homoisoflavonoids With Anti-cancer Activity.

Cameron A Garty

Institute of Biological, Environmental & Rural Science (IBERS), Aberystwyth University, Aberystwyth SY23 3DA, United Kingdom.

Contents

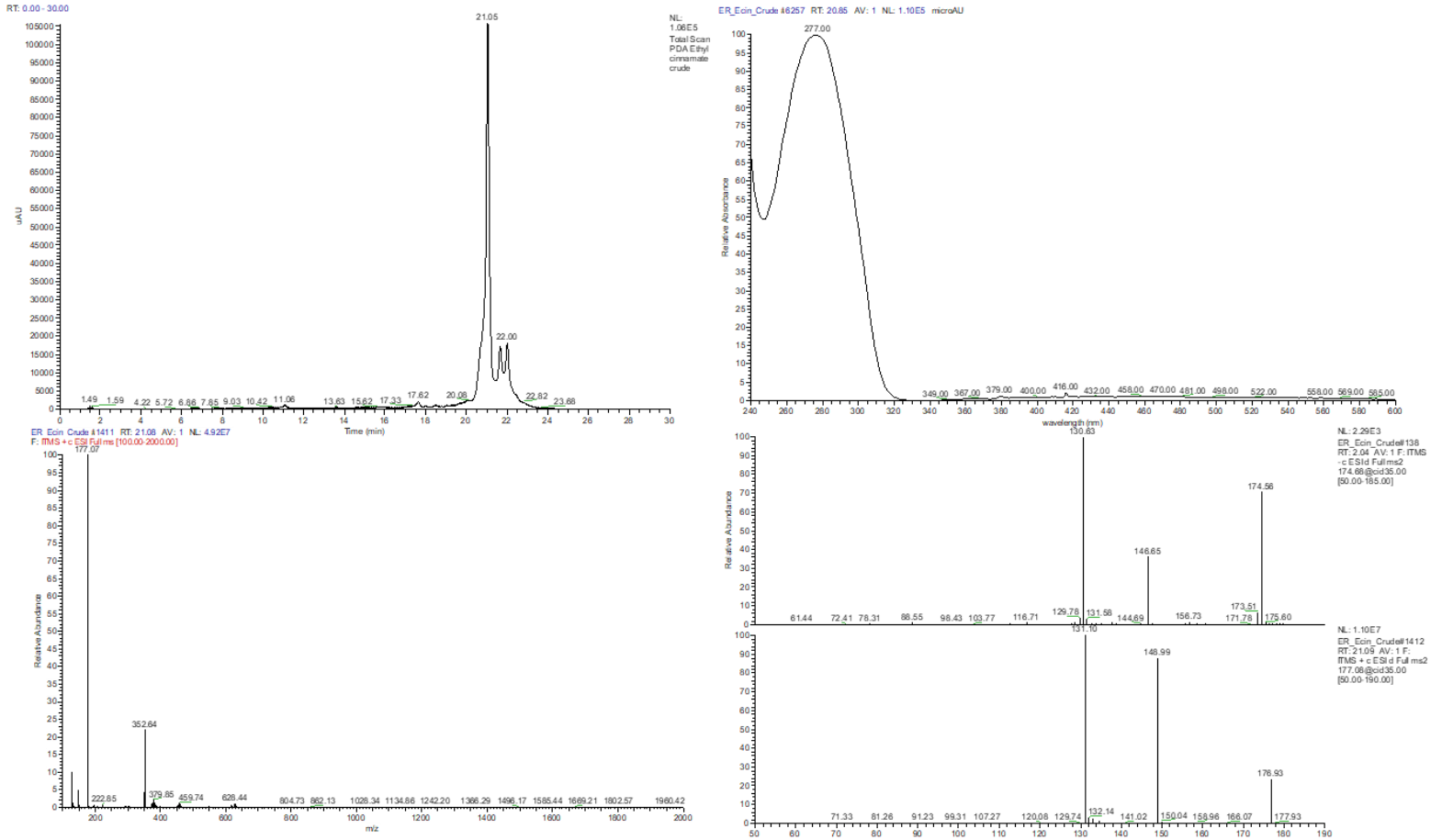
S16. Synthesis of ethyl ferulate.

S17. Oxidative coupling of ethyl ferulate and ferulic acid.

S18. Compound **10**

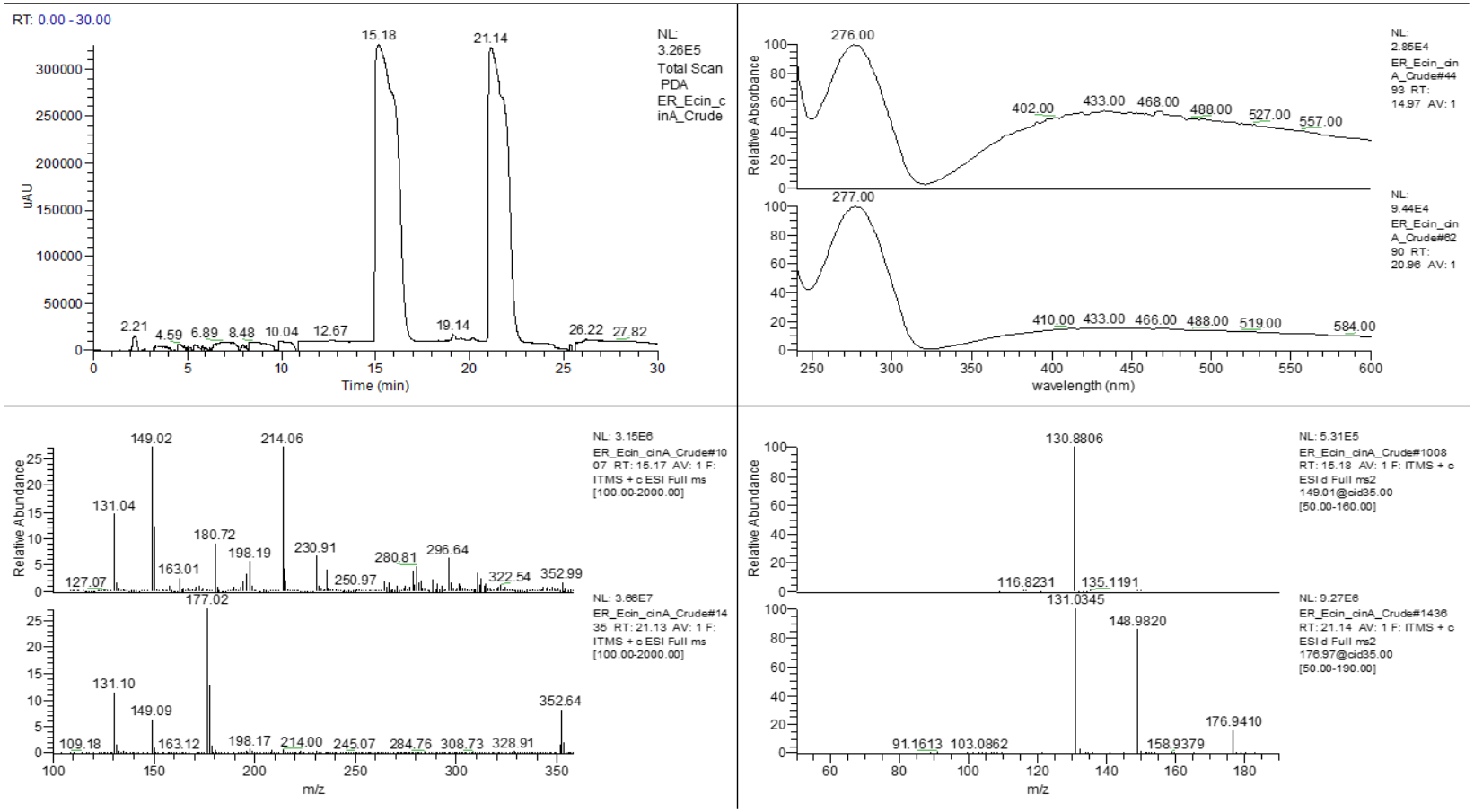
S19. Compound **11**

S1. Synthesis of ethyl cinnamate



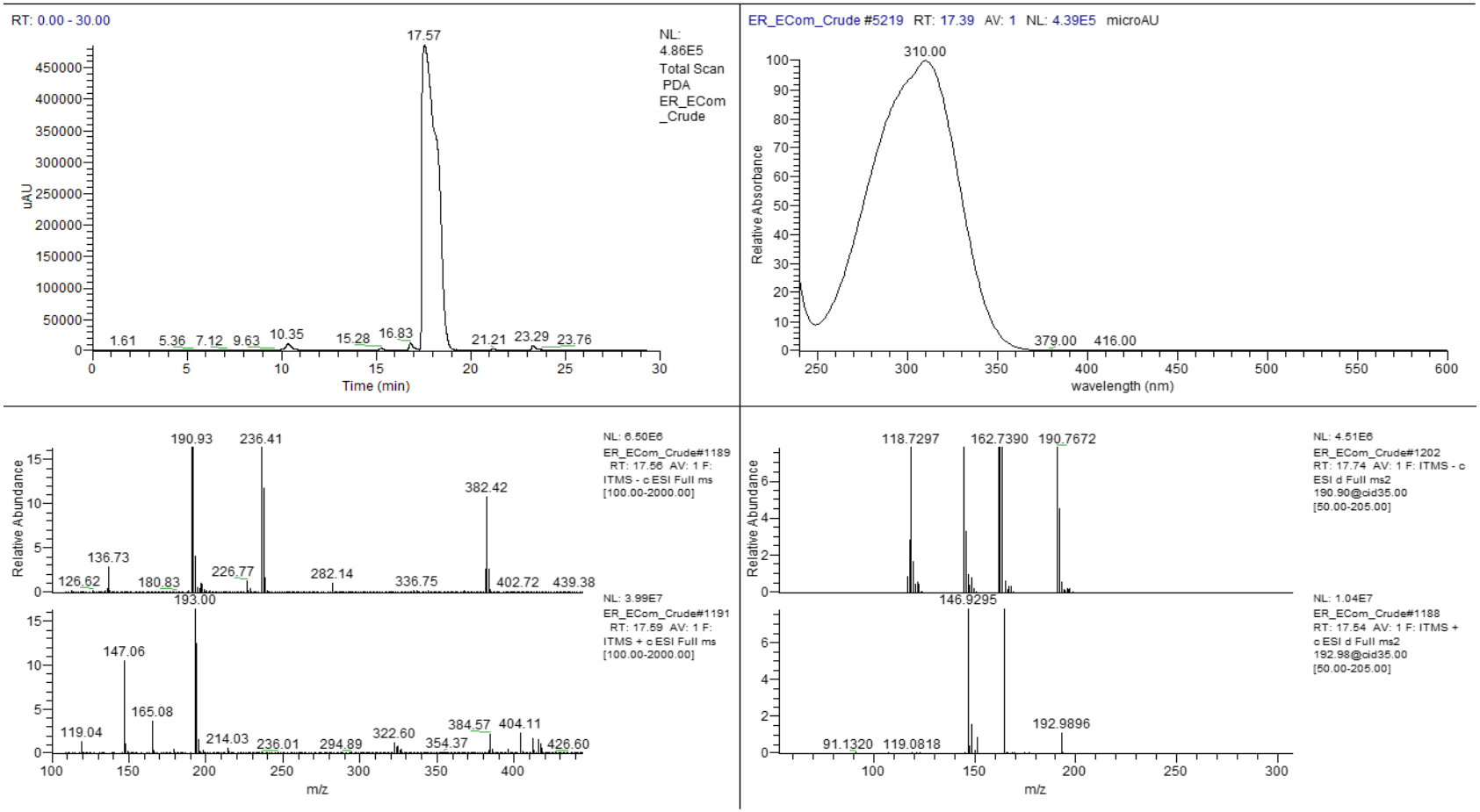
S1-I: HPLC-UV-ESI-MS/MS of ethyl cinnamate

S2. Oxidative coupling of ethyl cinnamate and cinnamic acid.



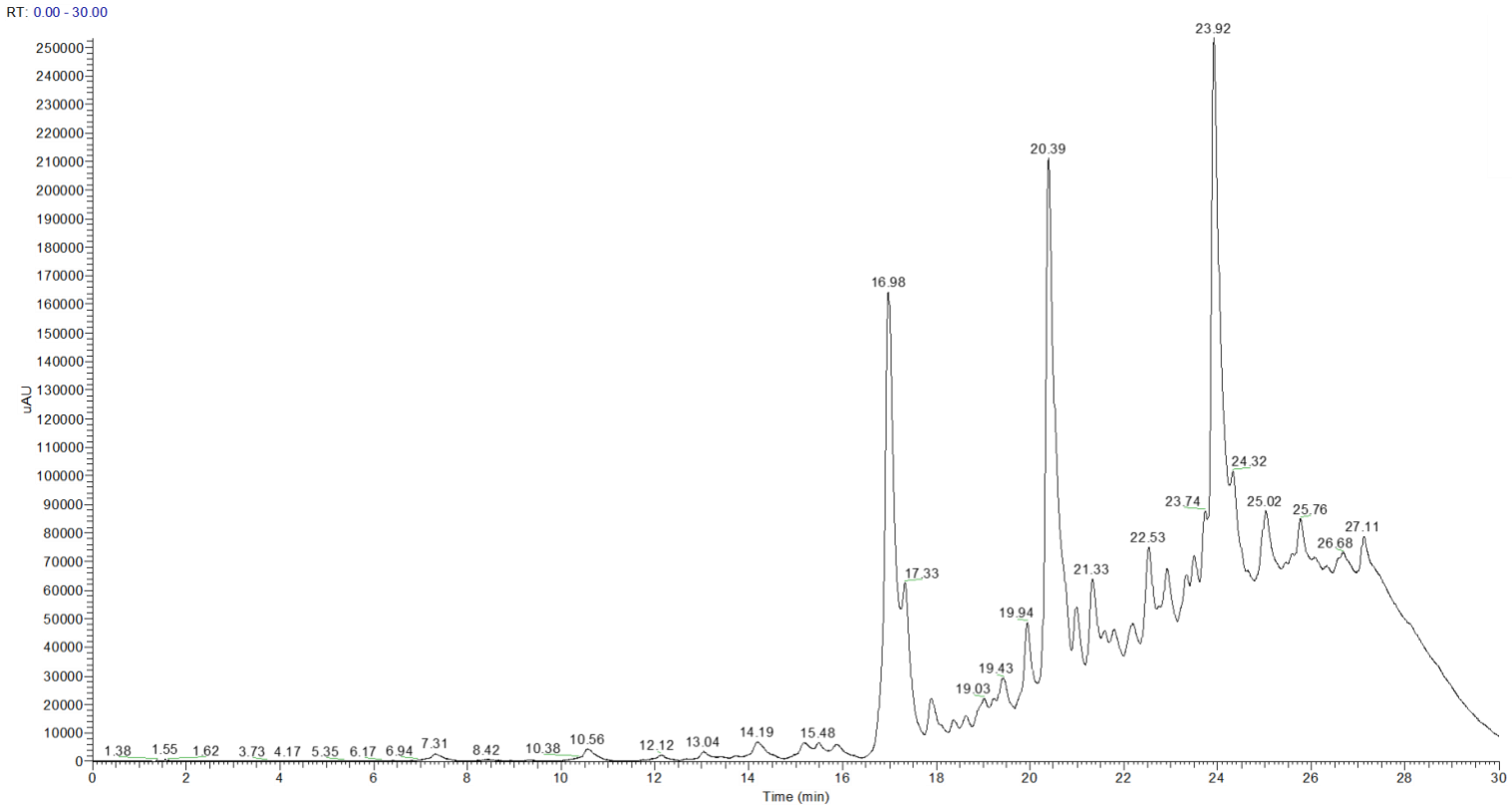
S2-I: HPLC-UV-ESI-MS/MS of oxidative coupling crude of ethyl cinnamate and ethyl cinnamate

S3. Synthesis of ethyl *para*-coumarate

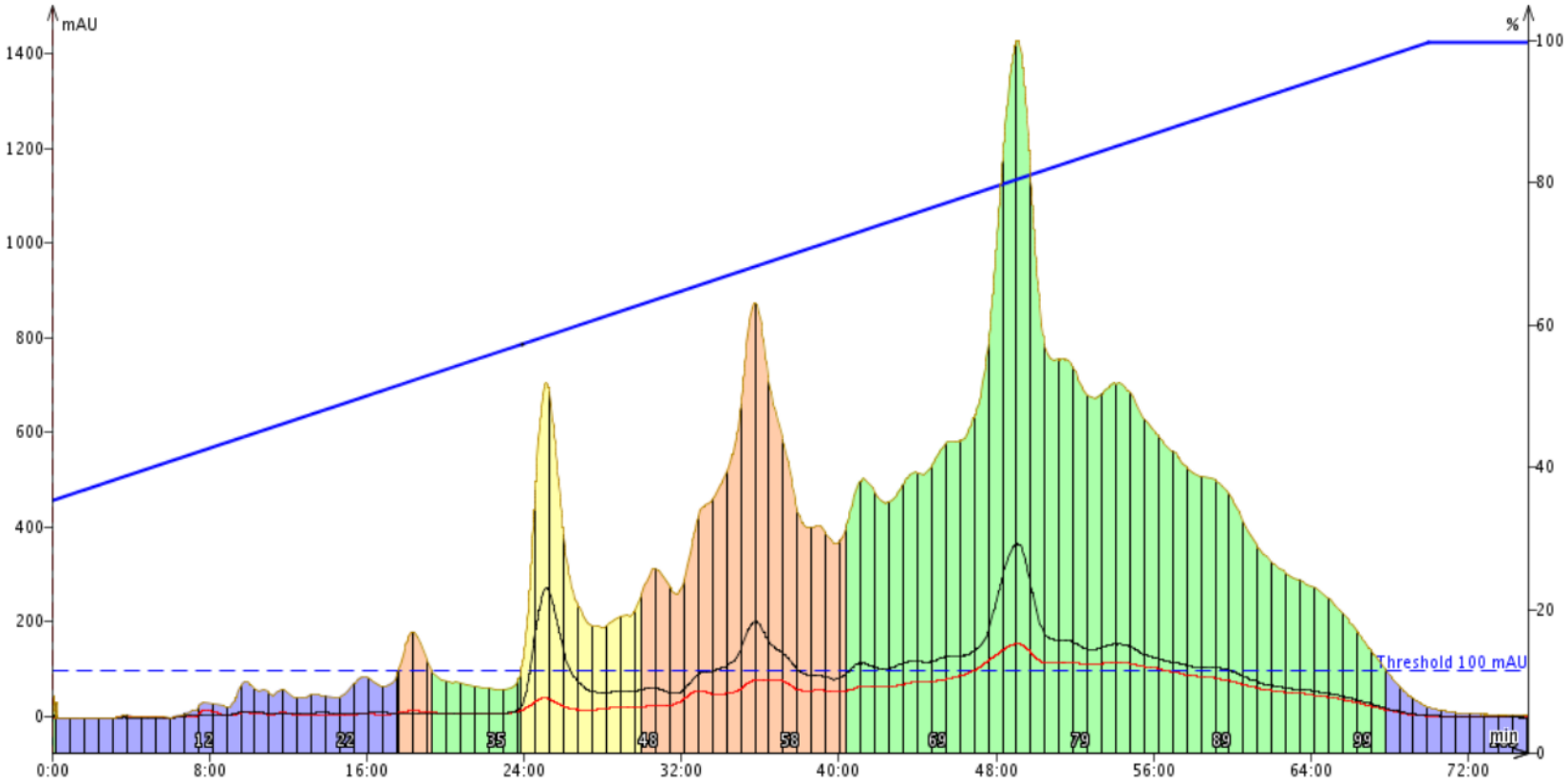


S3-I: HPLC-UV-ESI-MS/MS of ethyl *para*-coumarate

S4. Oxidative coupling of ethyl *para*-coumarate and *para*-coumaric acid.

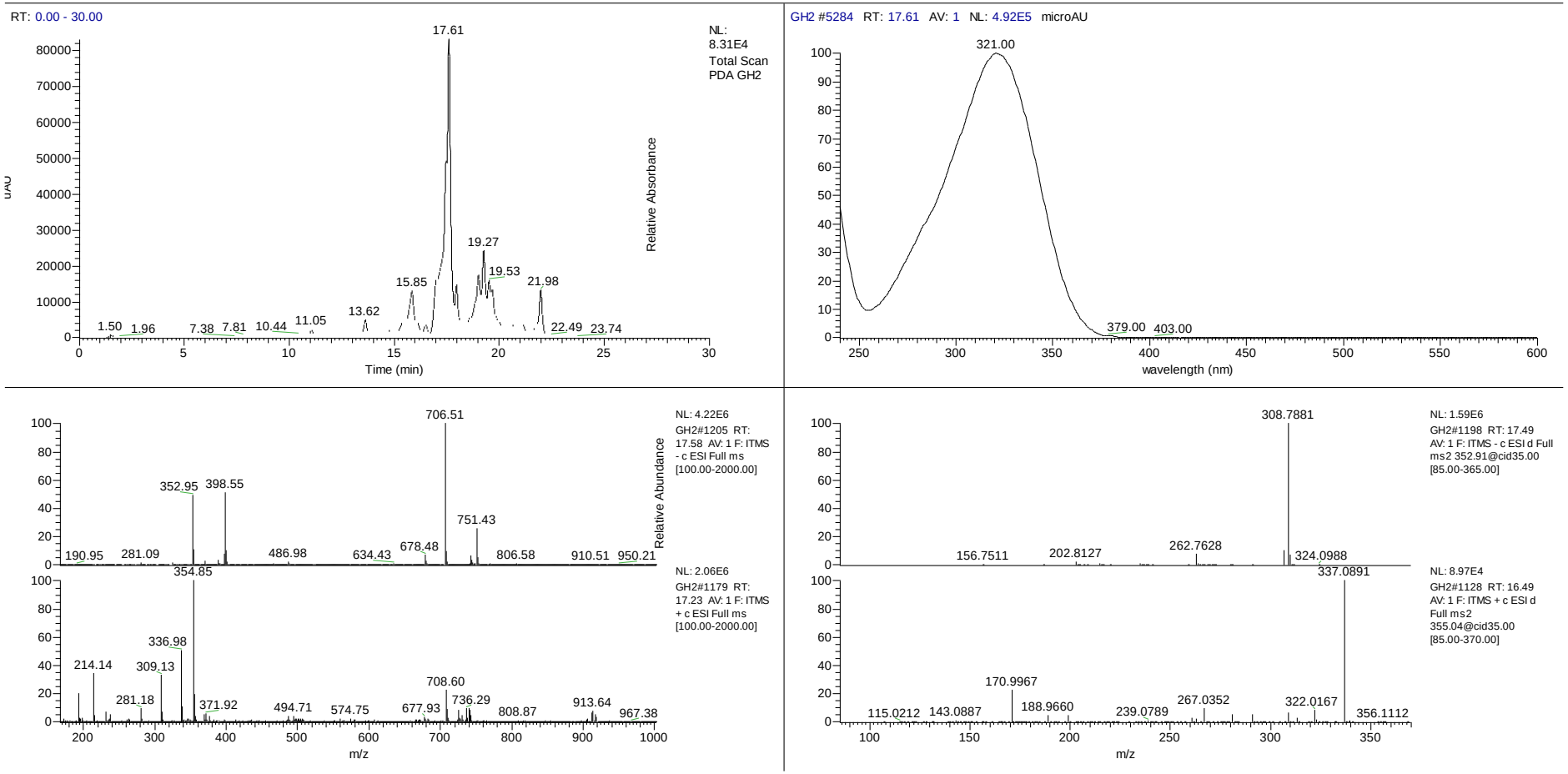


S4-I: HPLC -UV-ESI-MS/MS of oxidative coupling crude of ethyl *para*-coumarate and *para*-coumaric acid.



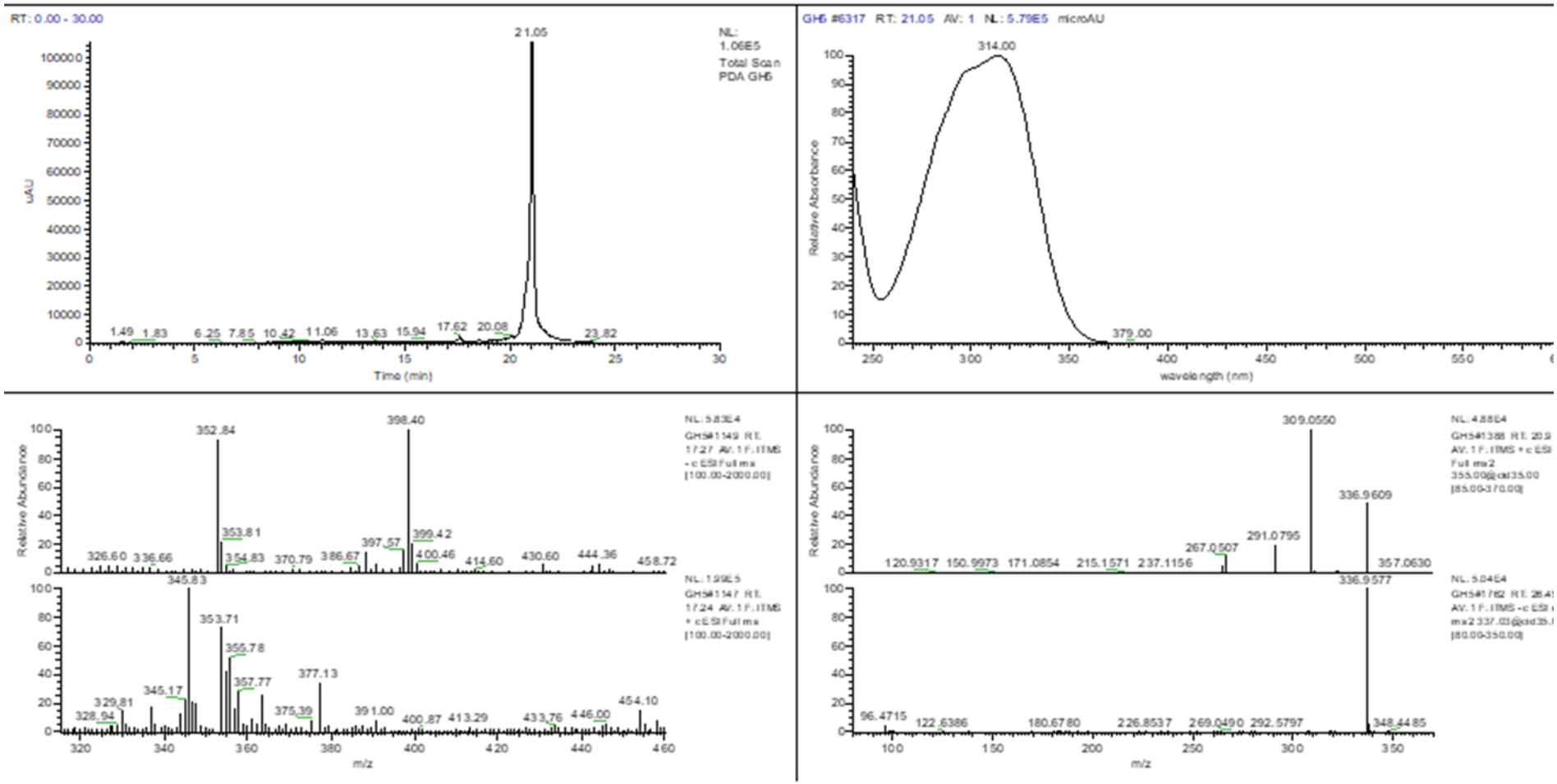
S4-II: UV chromatogram of Biotage Isolera Flash Chromatography – purification of oxidative crude of ethyl *para*-coumarate and *para*-coumaric acid.

S5. Compound **1**

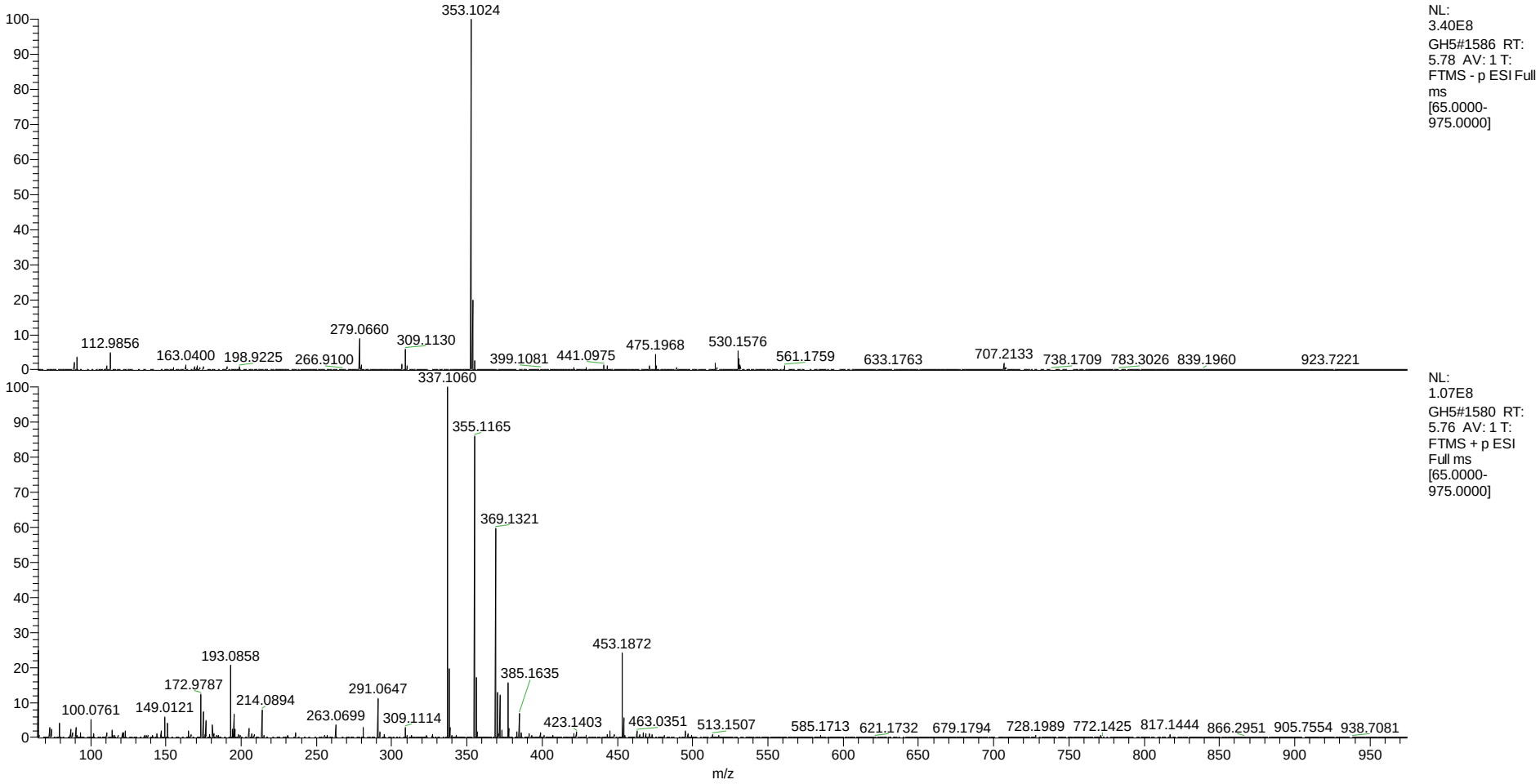


S5-I: HPLC-UV-ESI-MS/MS of compound **1.**

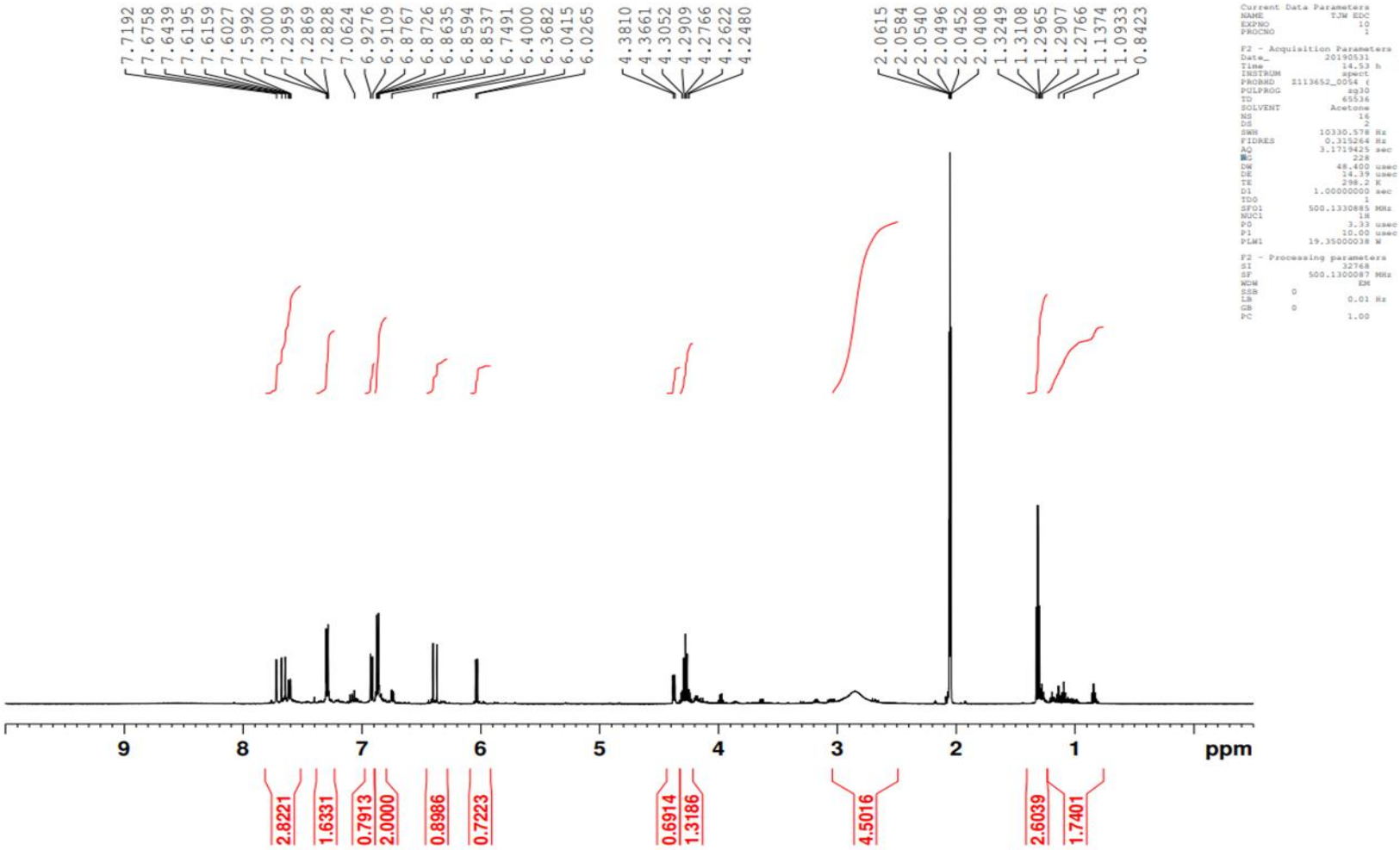
S6. Compound **2**



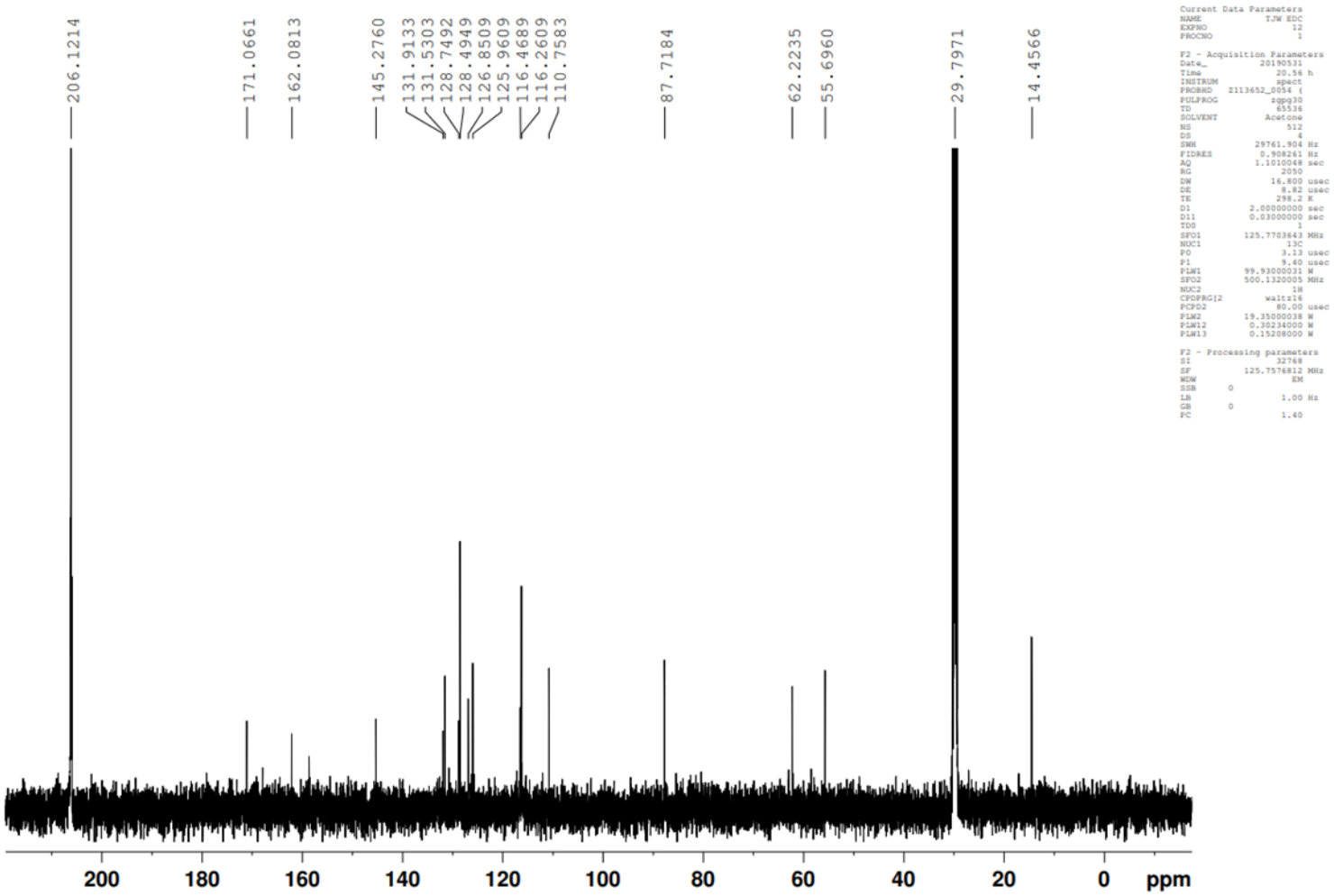
S6-I: HPLC-UV-ESI-MS/MS of compound **2**



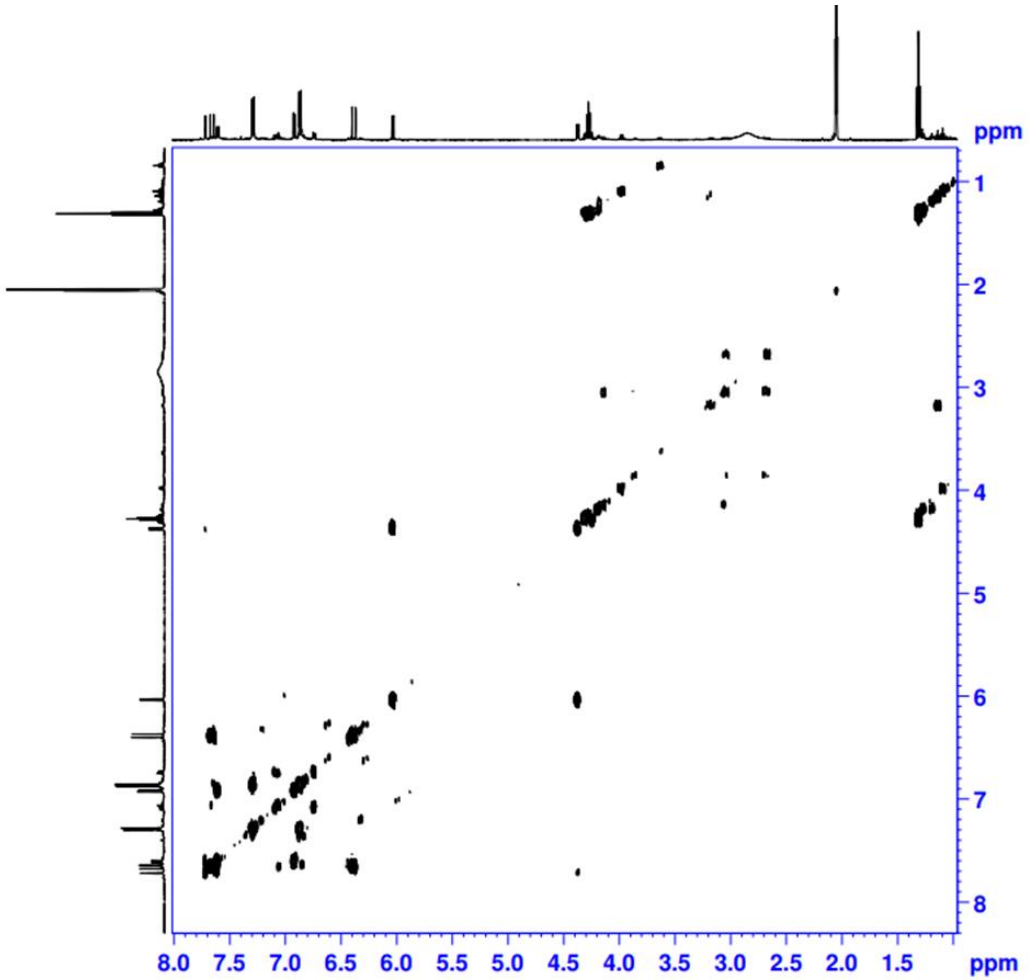
S6-II: HPLC Accurate Mass of compound **2**.



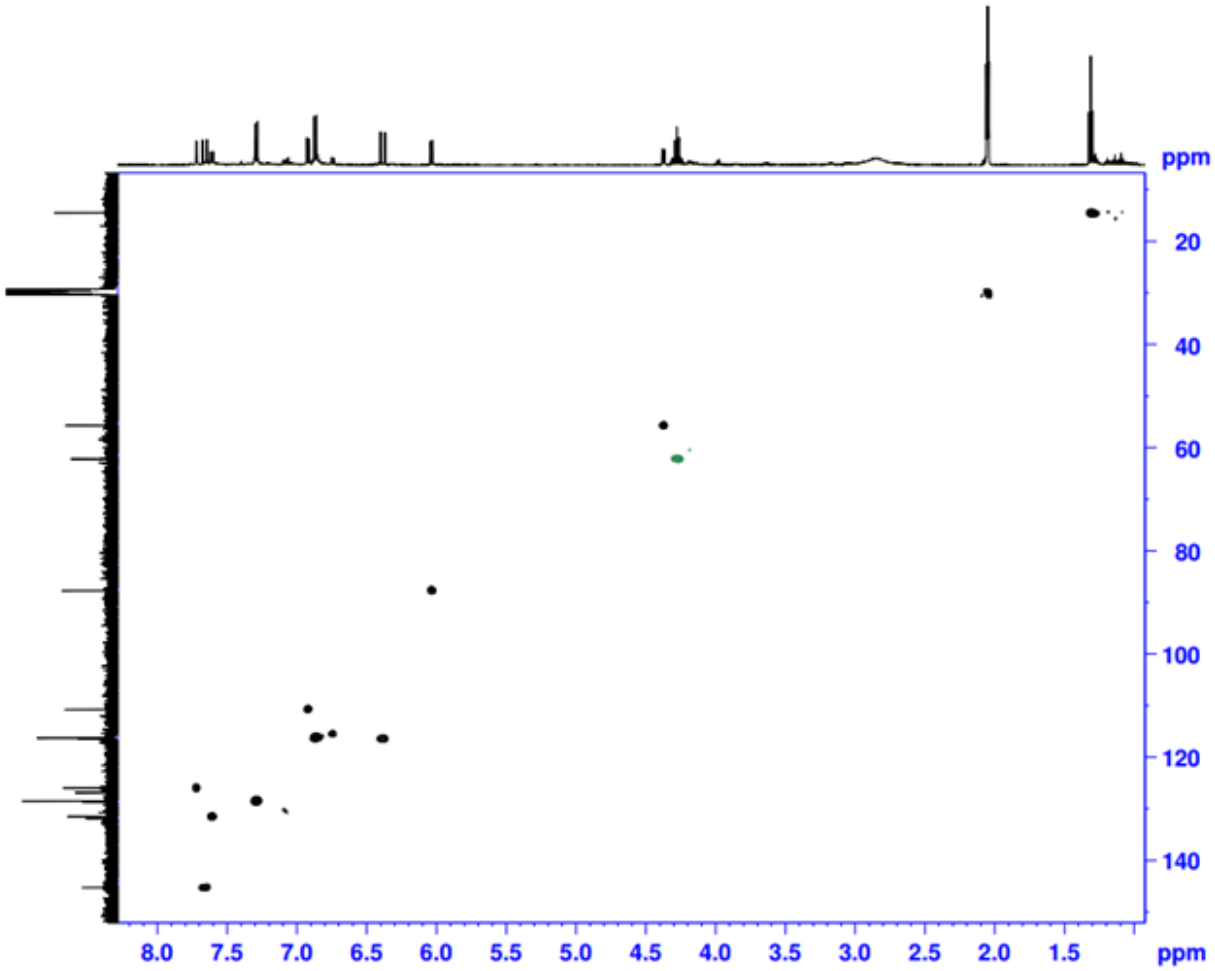
S6-III: 1H NMR spectrum of compound **2**.



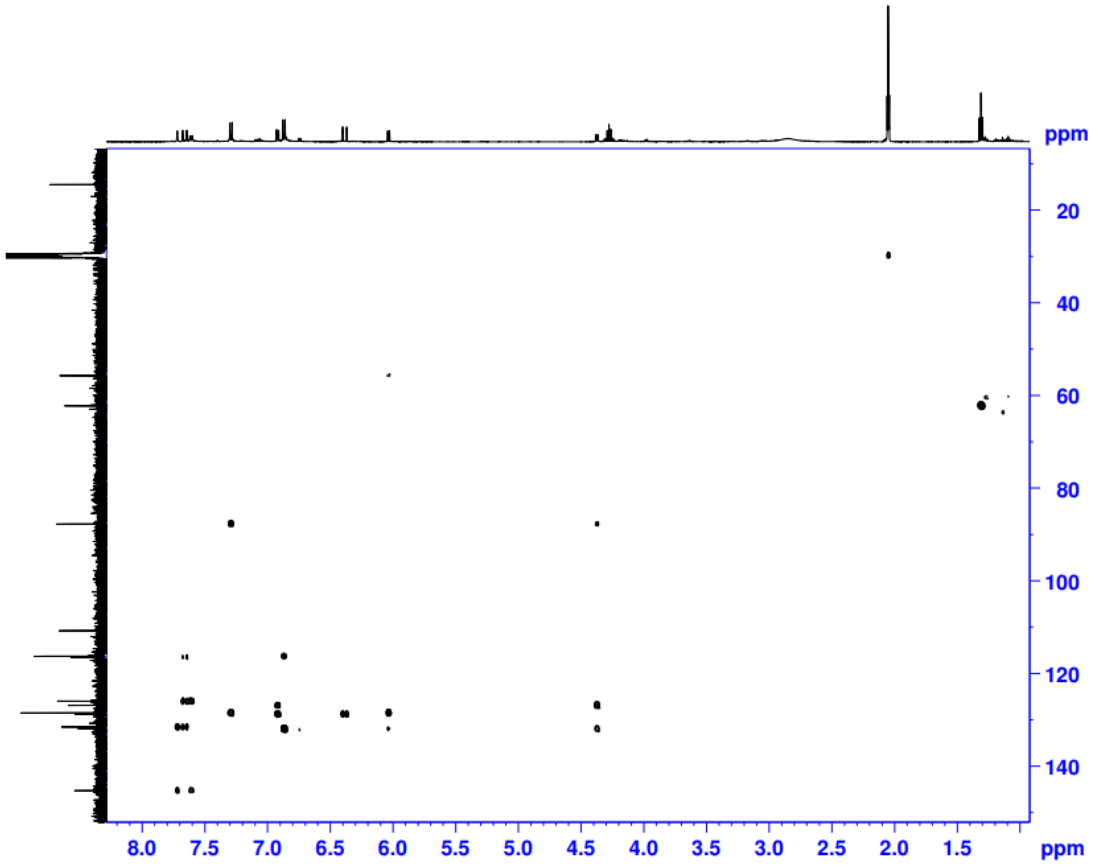
S6-IV: ^{13}C NMR spectrum of compound **2.**



S6-V: COSY spectrum of compound **2**

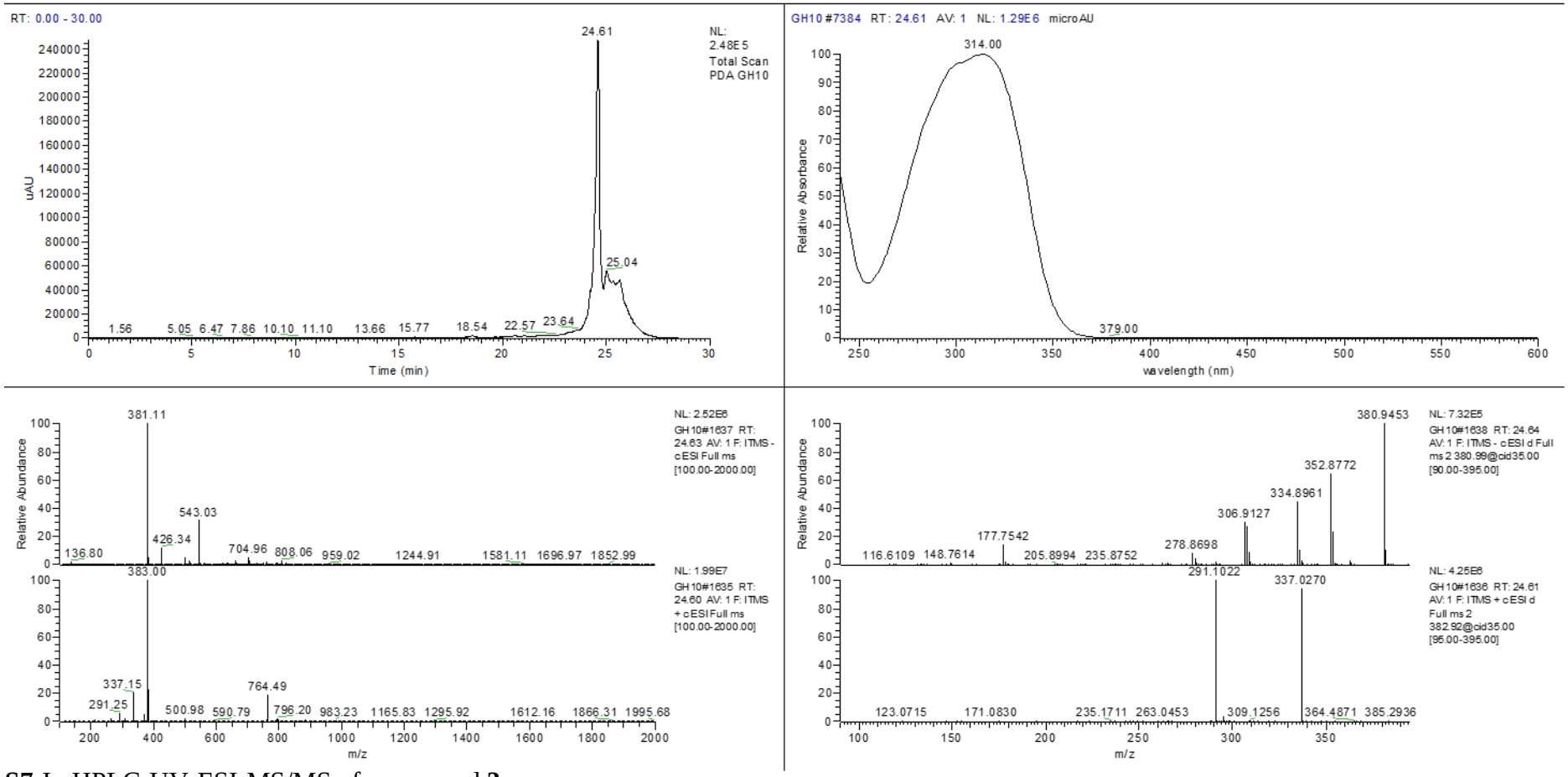


S6-VI: HSQC spectrum of compound **2**

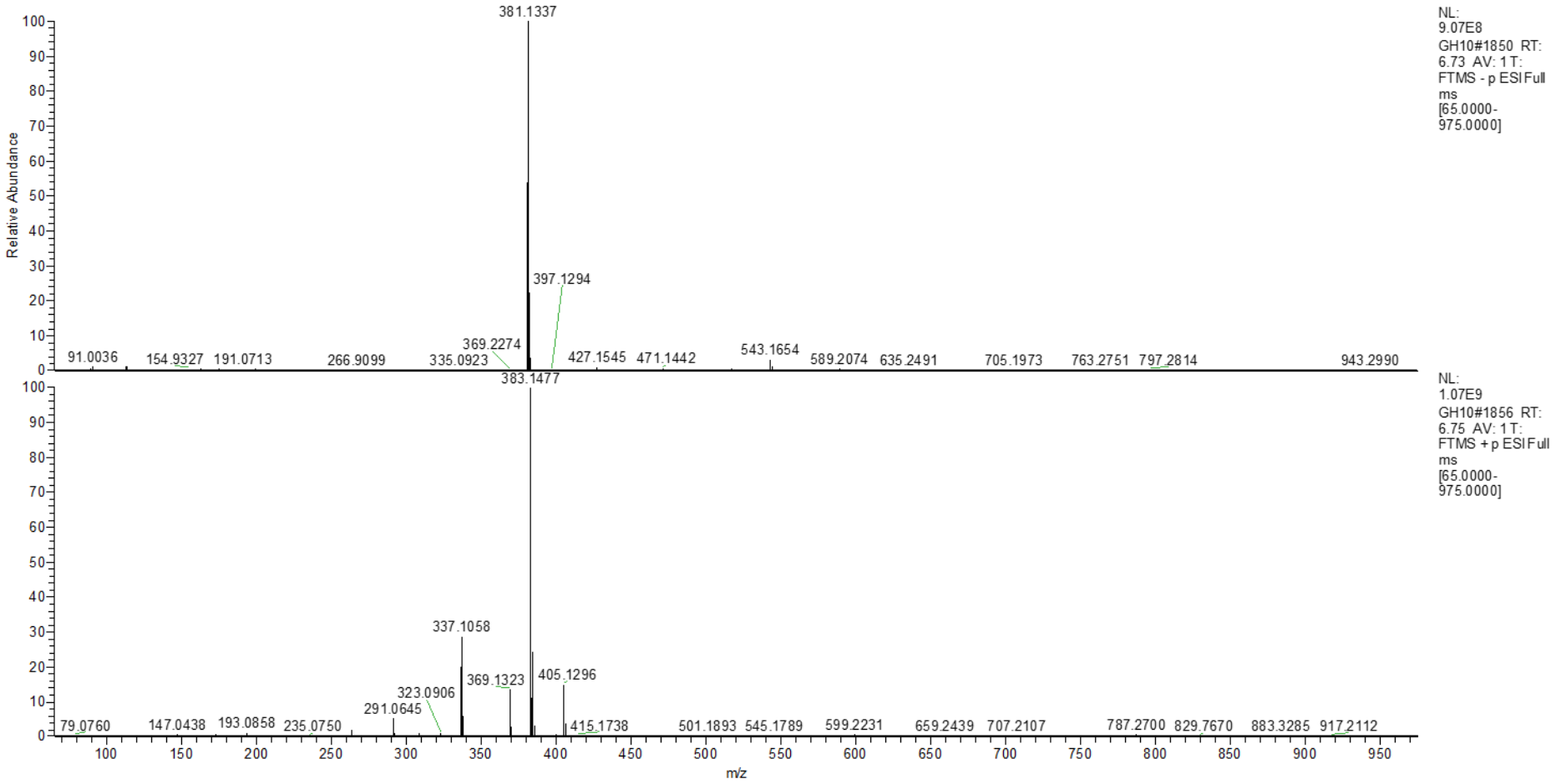


S6-VII: HMBC spectrum of compound **2**

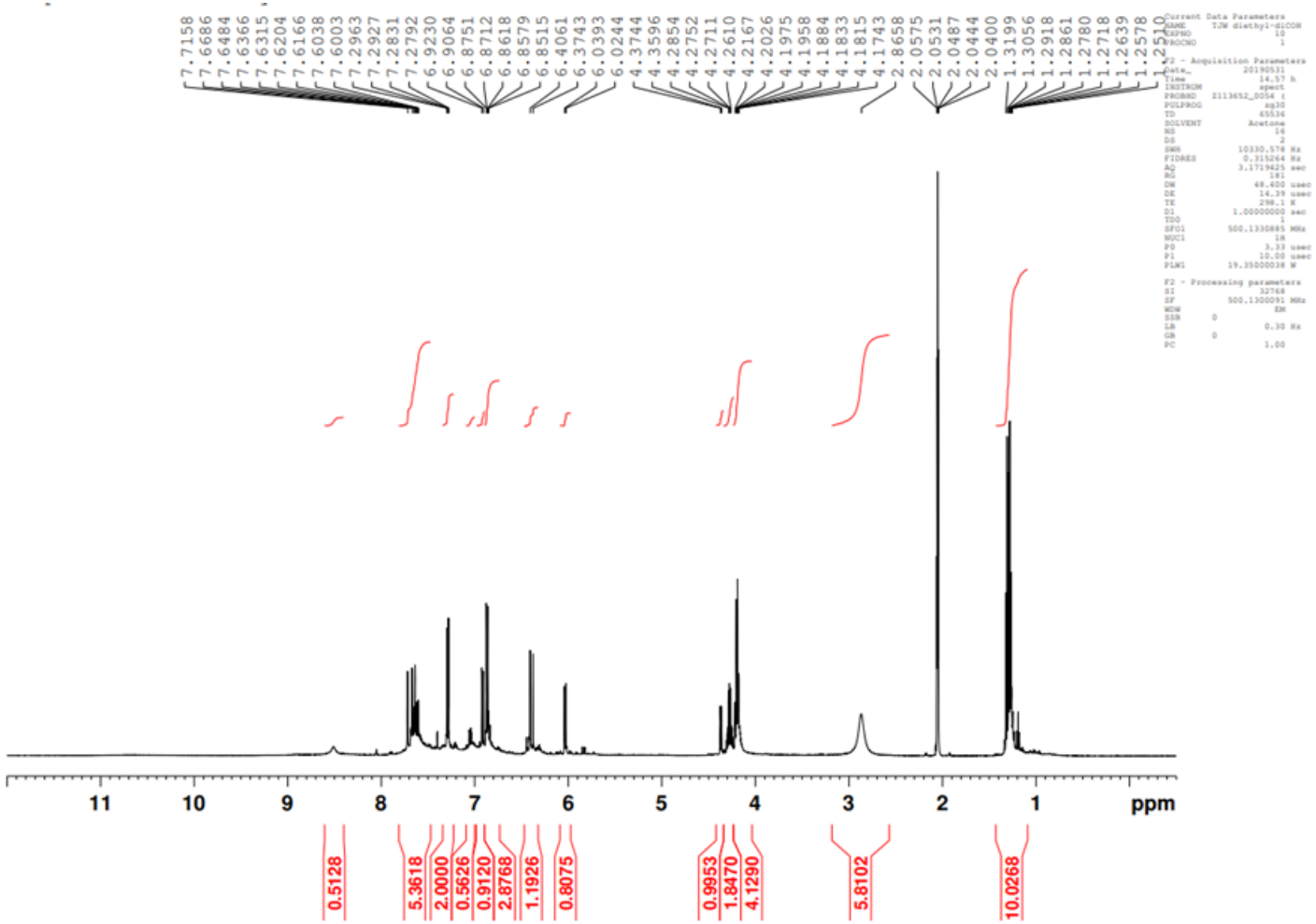
S7. Compound **3**



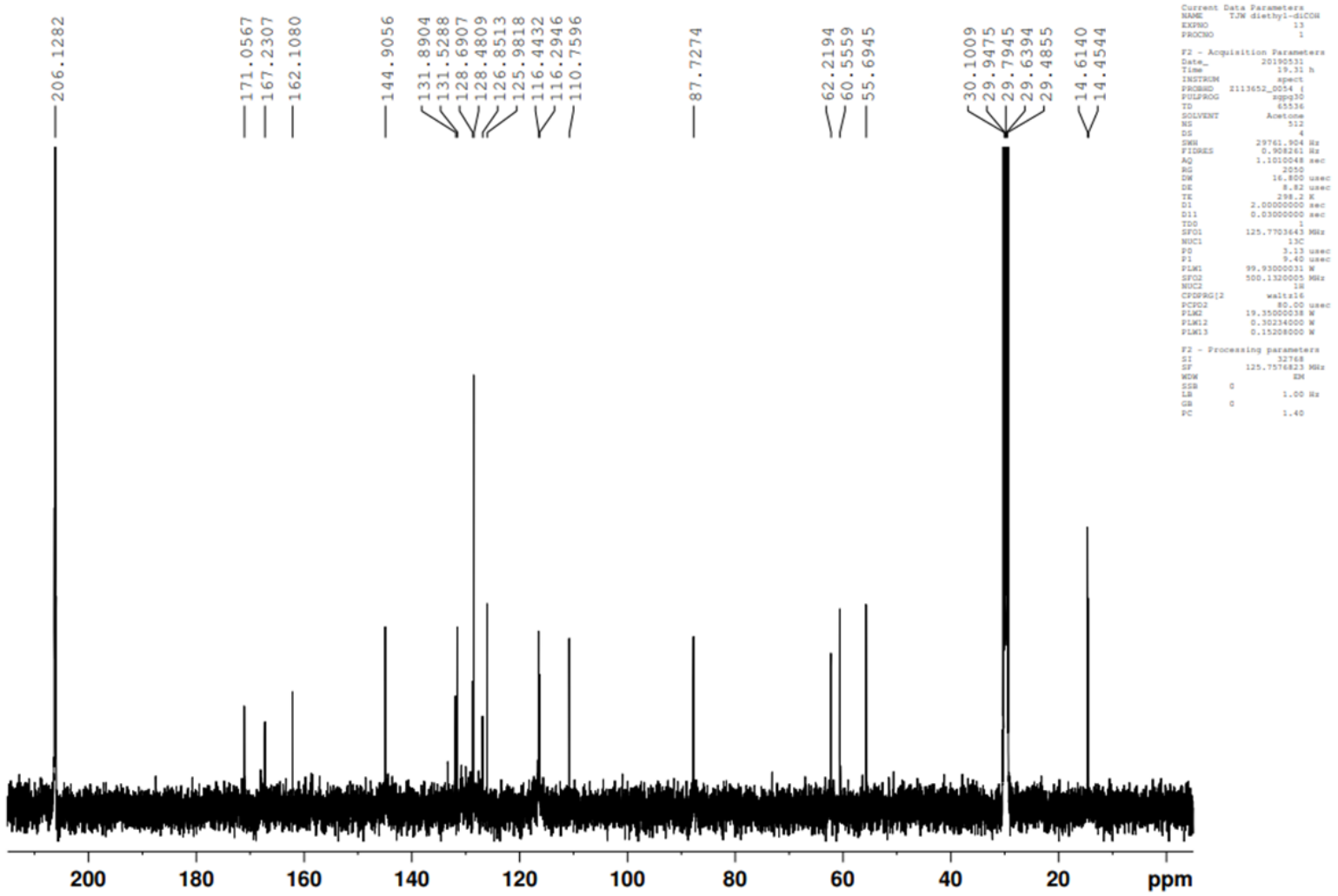
S7-I: HPLC-UV-ESI-MS/MS of compound **3**



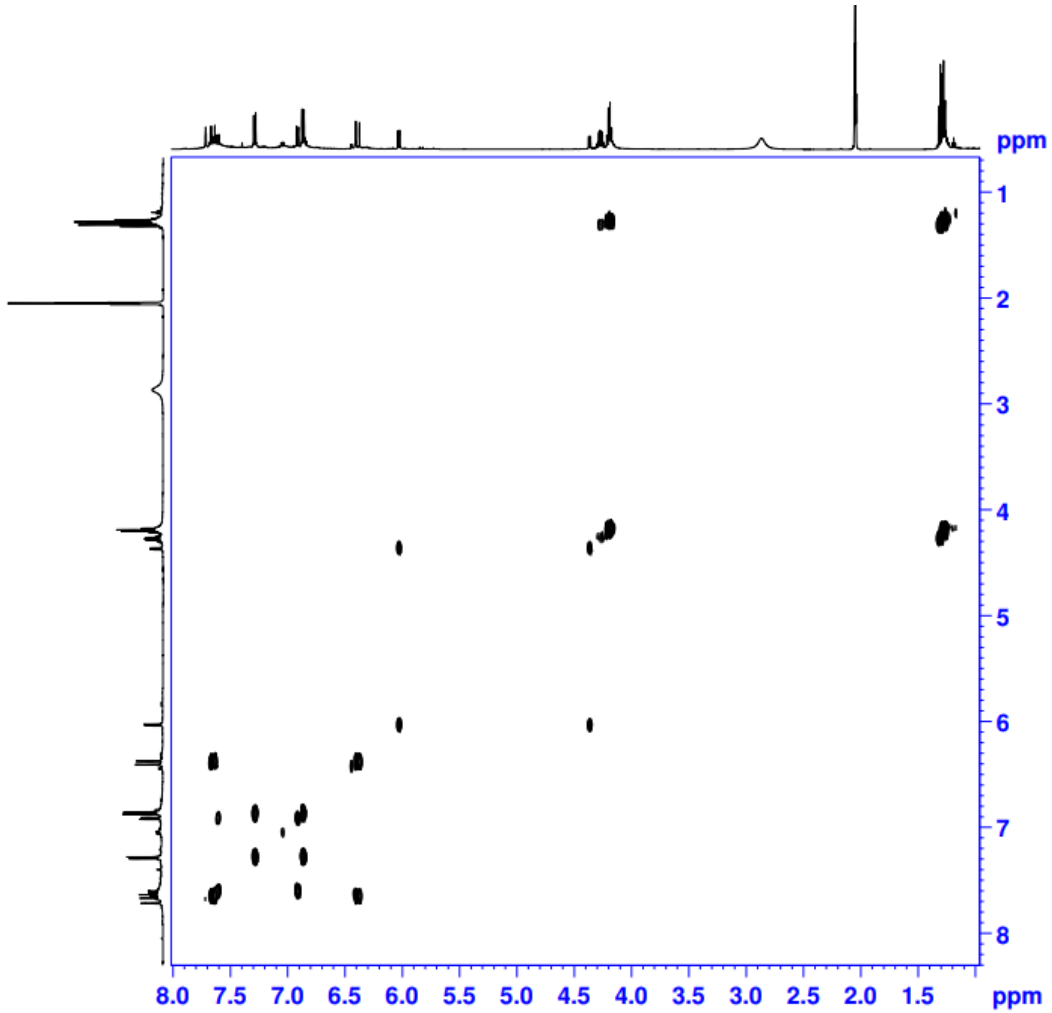
S7-II: Accurate Mass of compound **3**



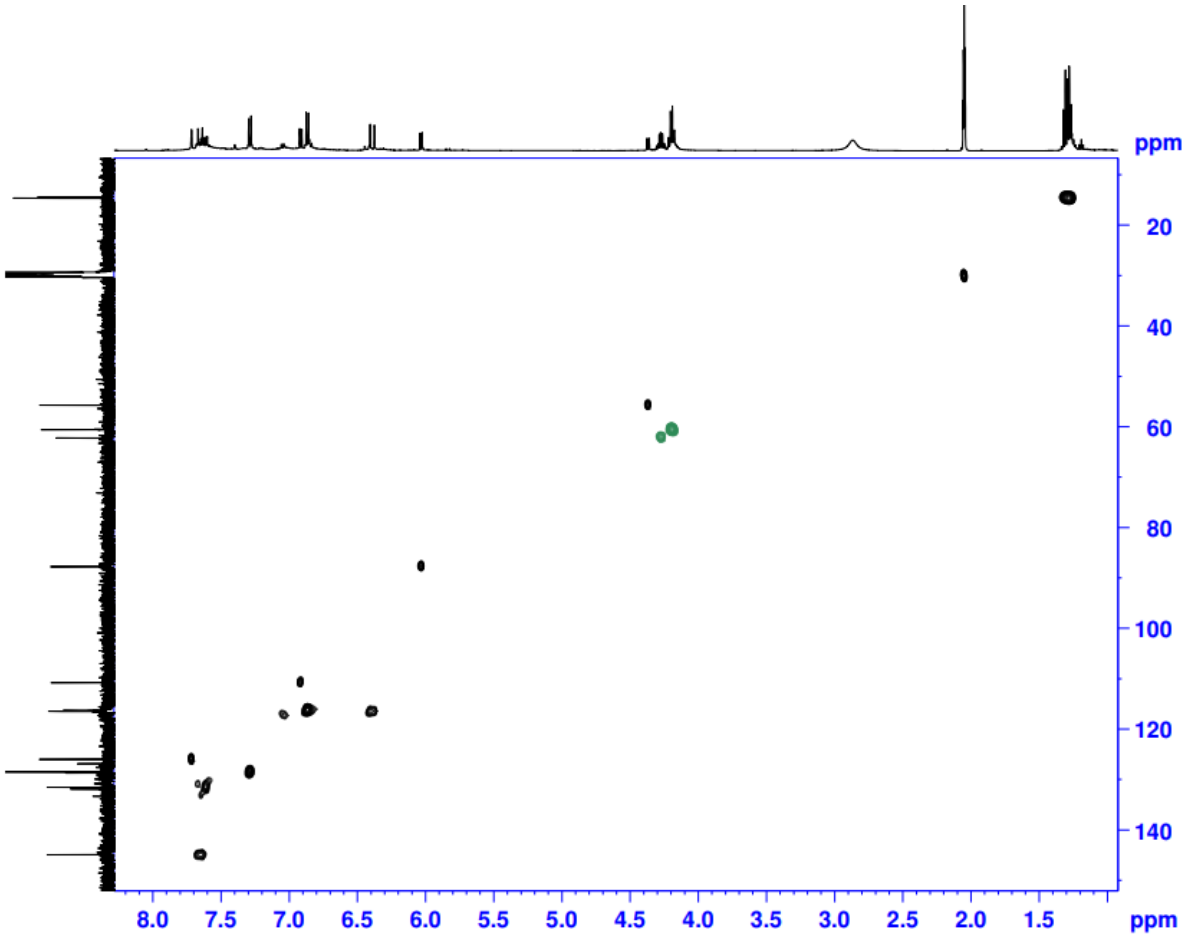
S7-III: 1H NMR spectrum of compound **3**



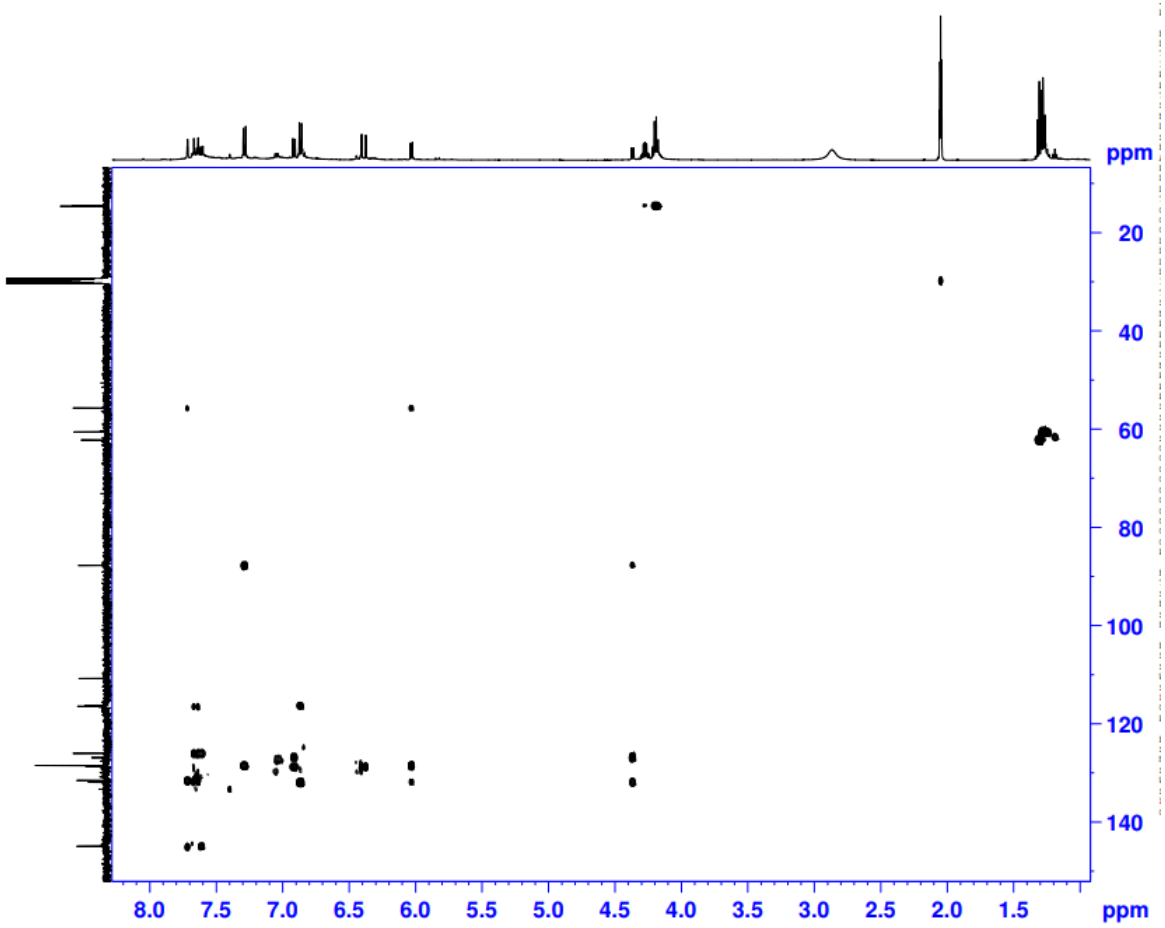
S7-IV: ^{13}C NMR spectrum of compound **3**



S7-V: COSY spectrum of compound **3**

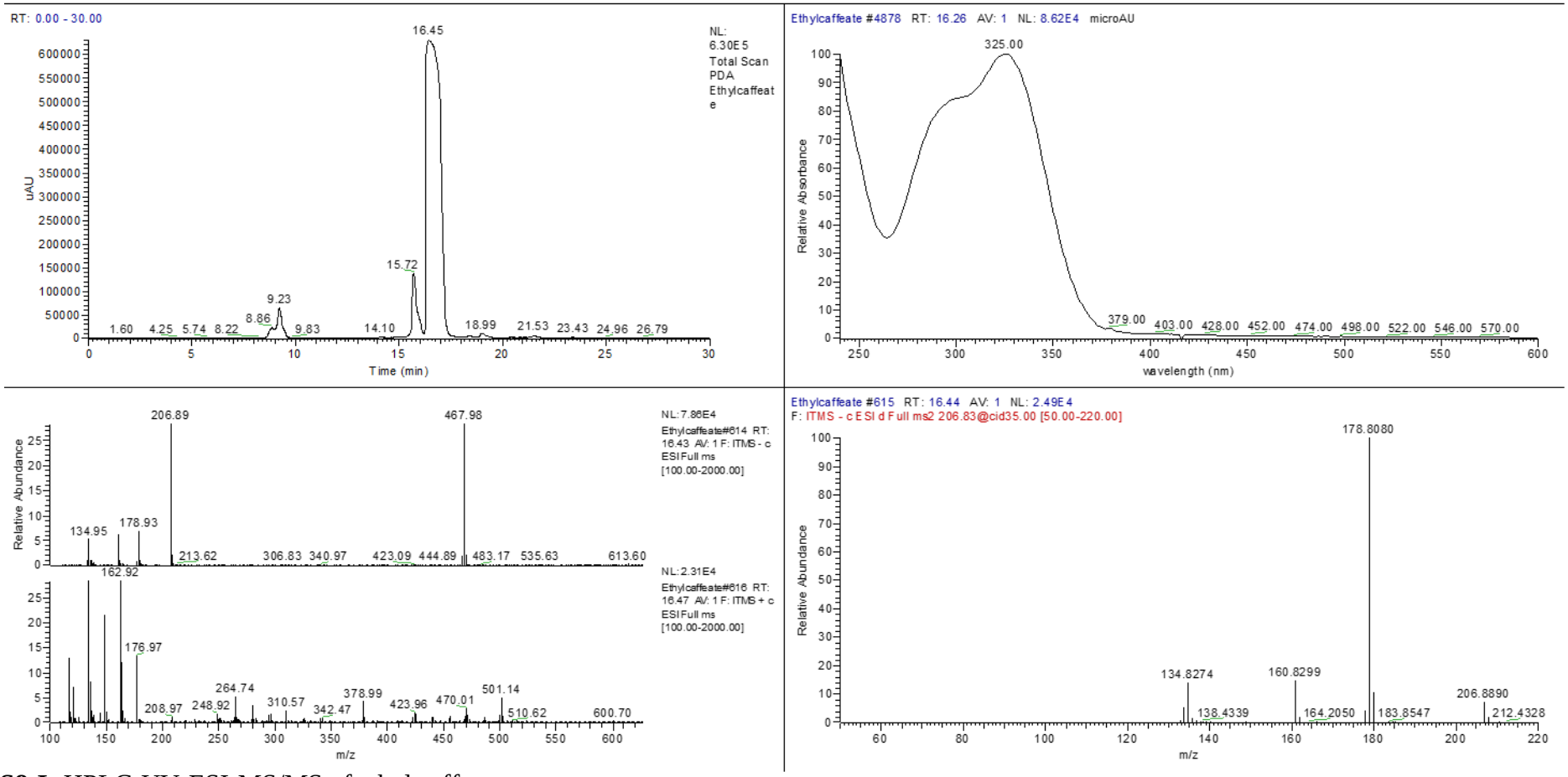


S7-VI: HSQC spectrum of compound **3**

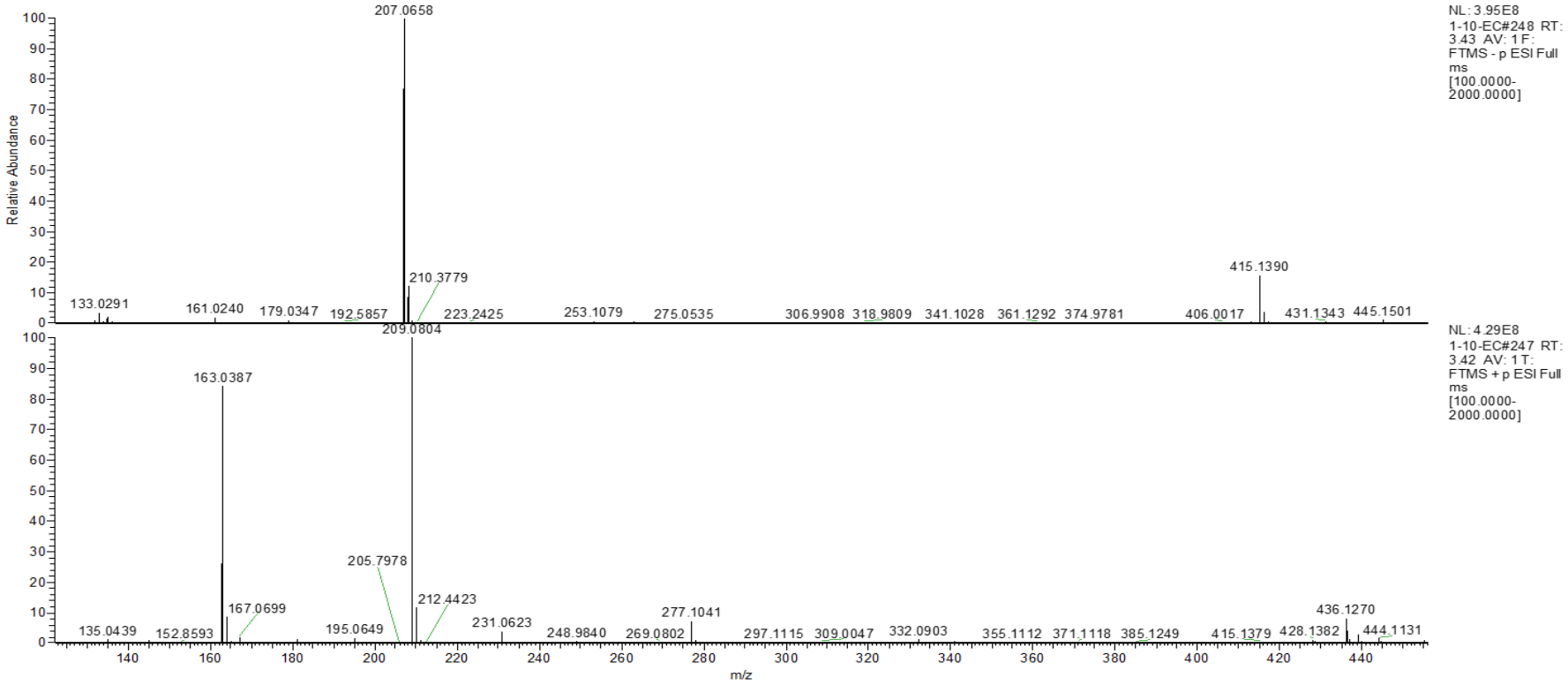


S7-VII: HMBC spectrum of compound **3**

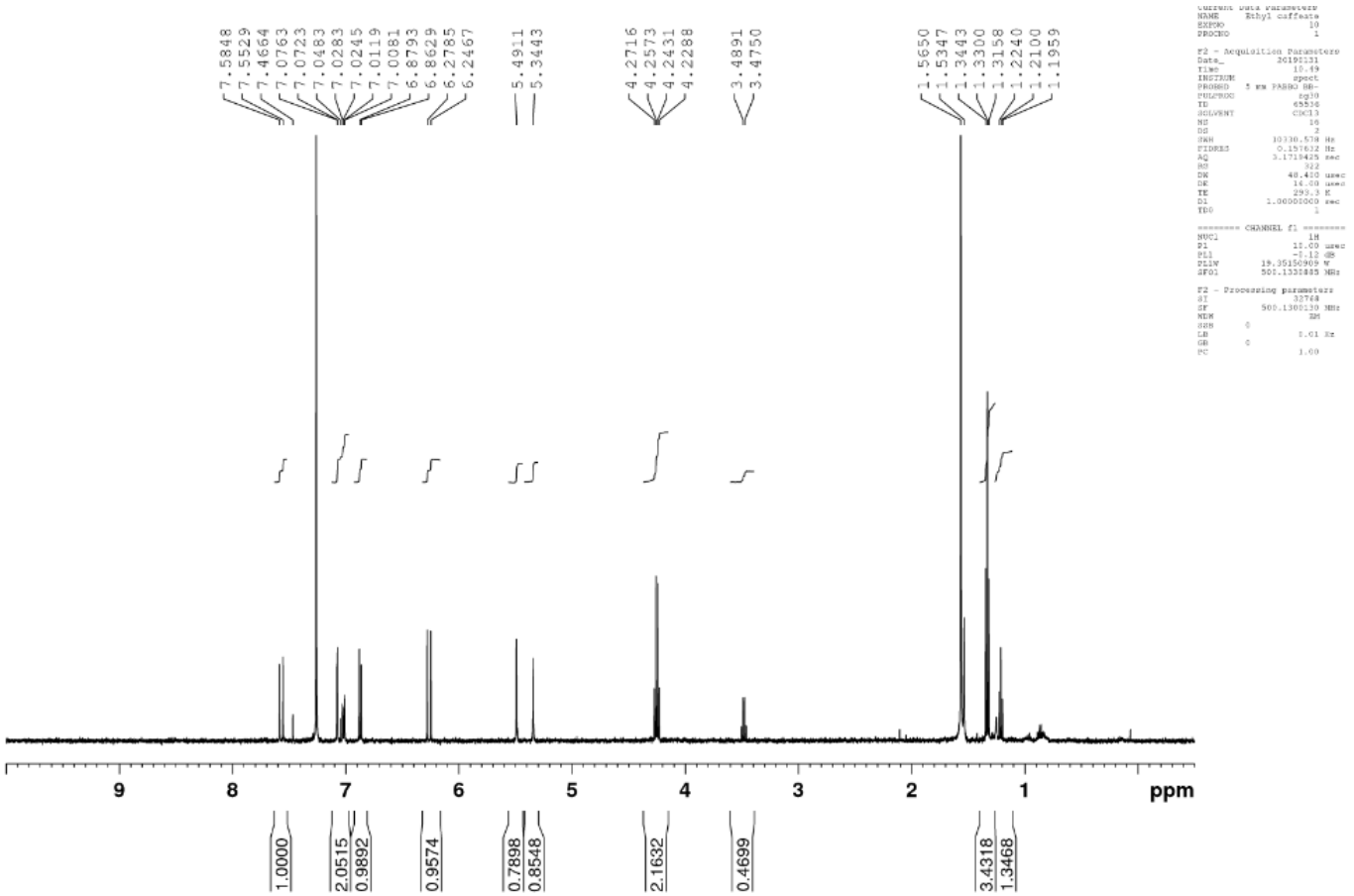
S8. Synthesis of ethyl caffeate



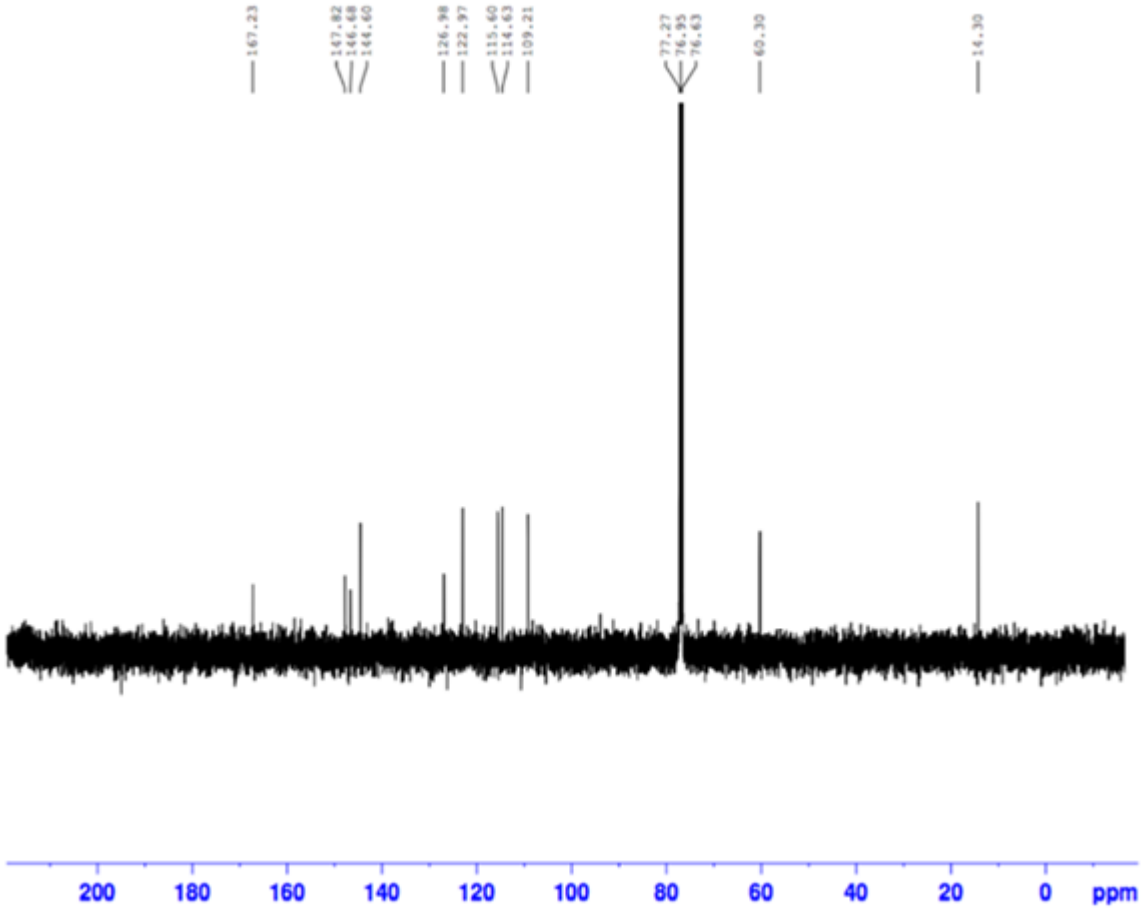
S8-I: HPLC-UV-ESI-MS/MS of ethyl caffeate



S8-II: Accurate mass of ethyl cinnamate

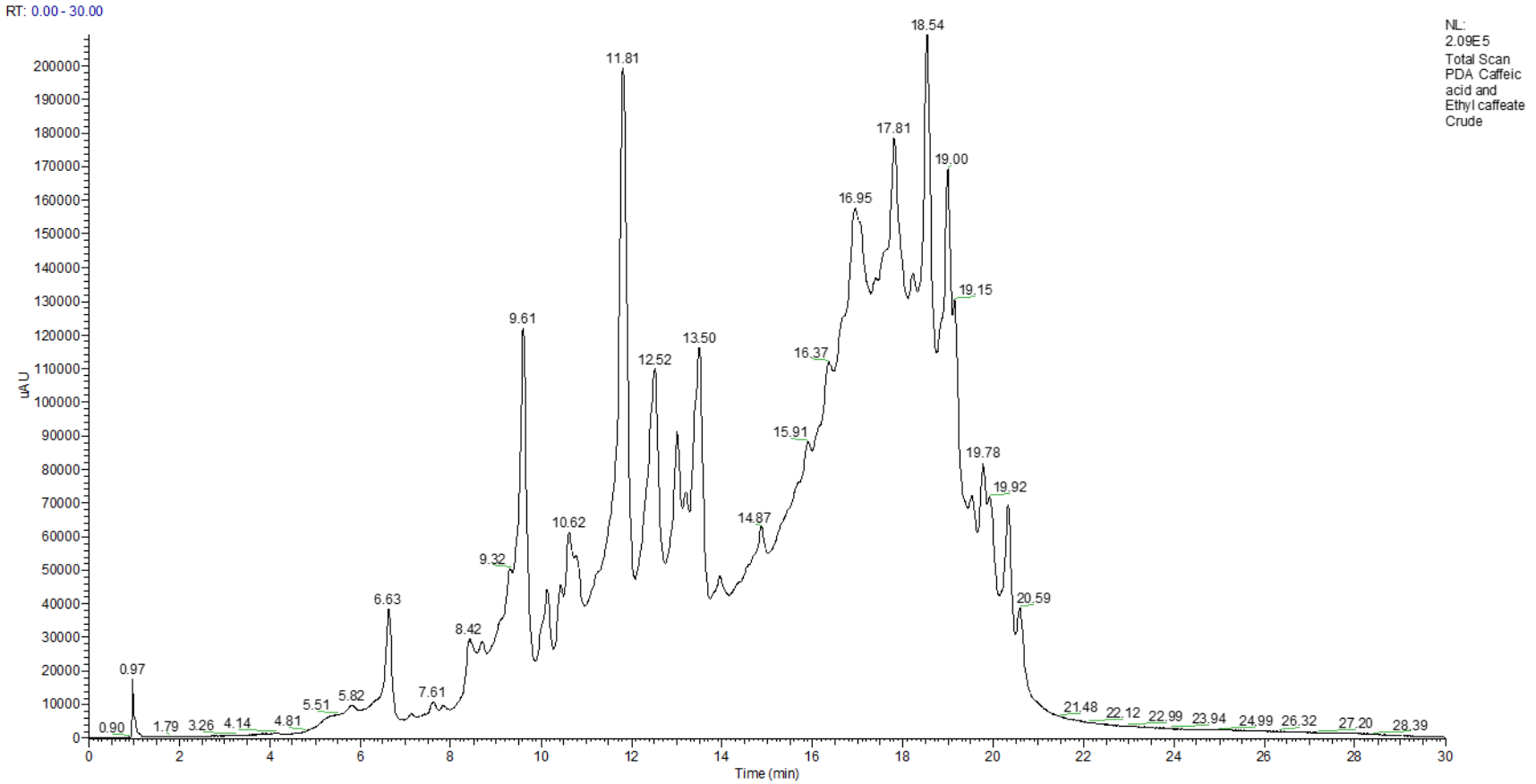


S8-III: 1H NMR spectrum of ethyl caffeate

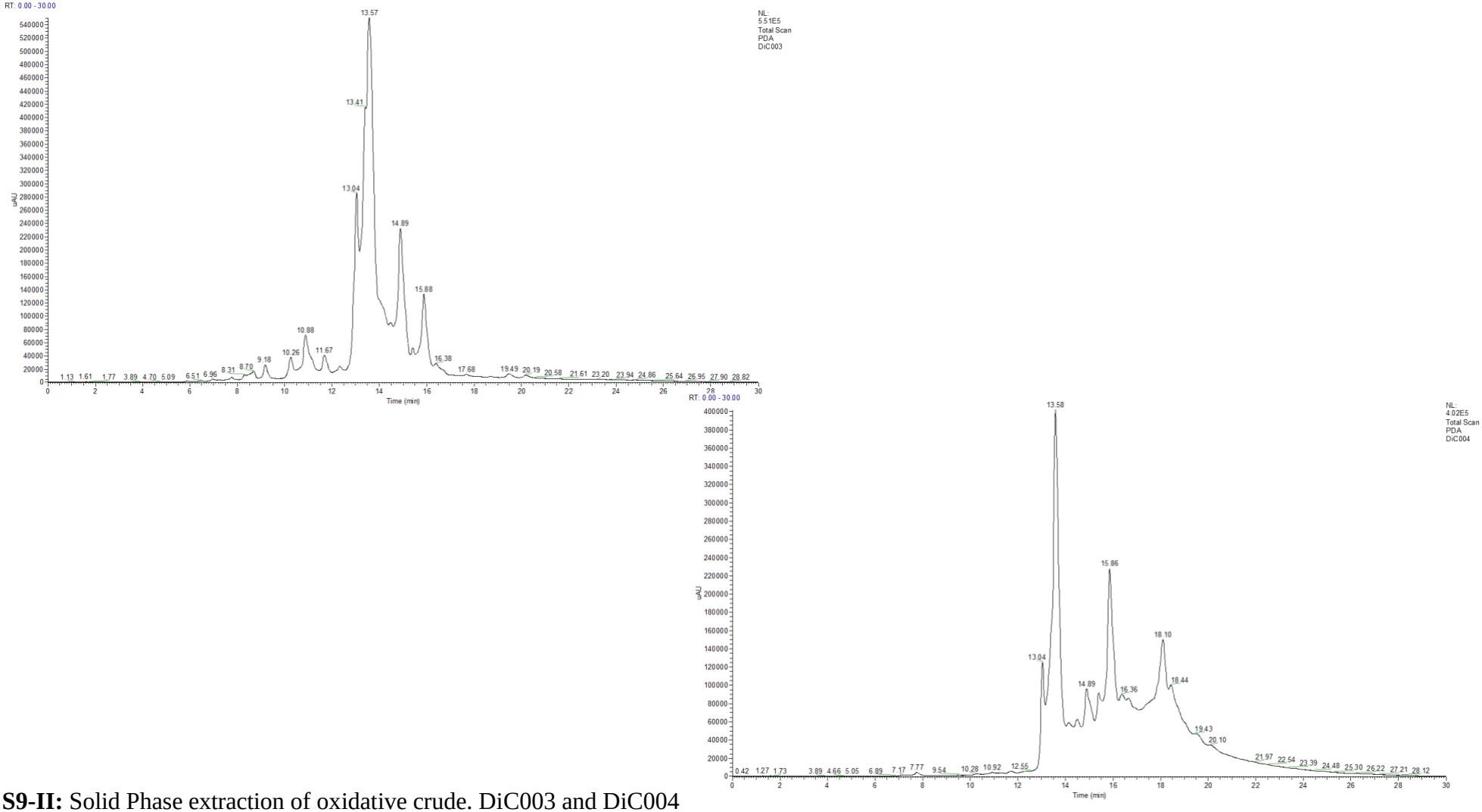


S8-IV: ^{13}C NMR spectrum of ethyl caffeate

S9. Oxidative coupling of ethyl caffeate and caffeic acid.

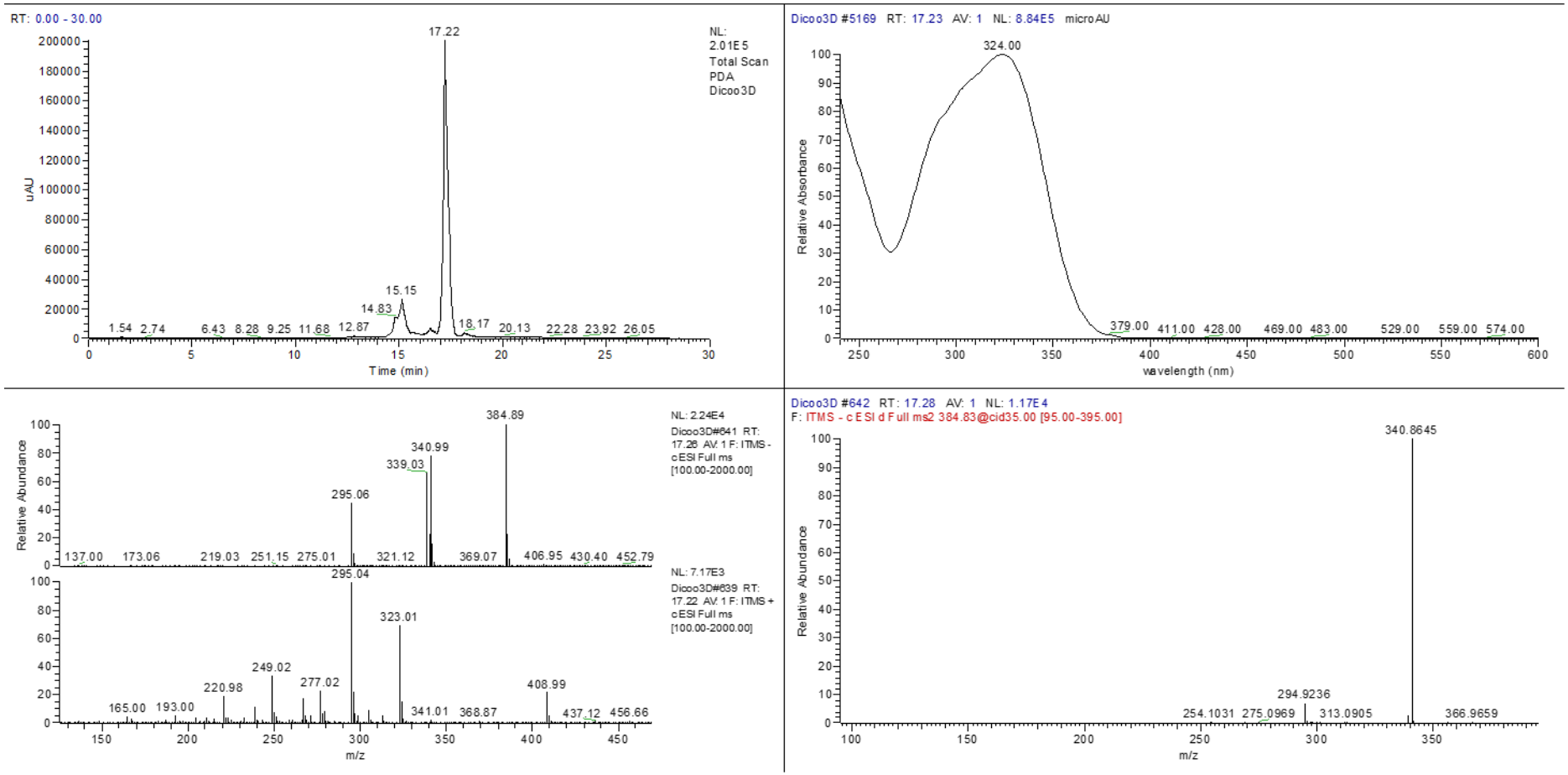


S9-I: HPLC-UV-ESI-MS/MS of oxidative coupling of ethyl caffeate and caffeic acid crude.

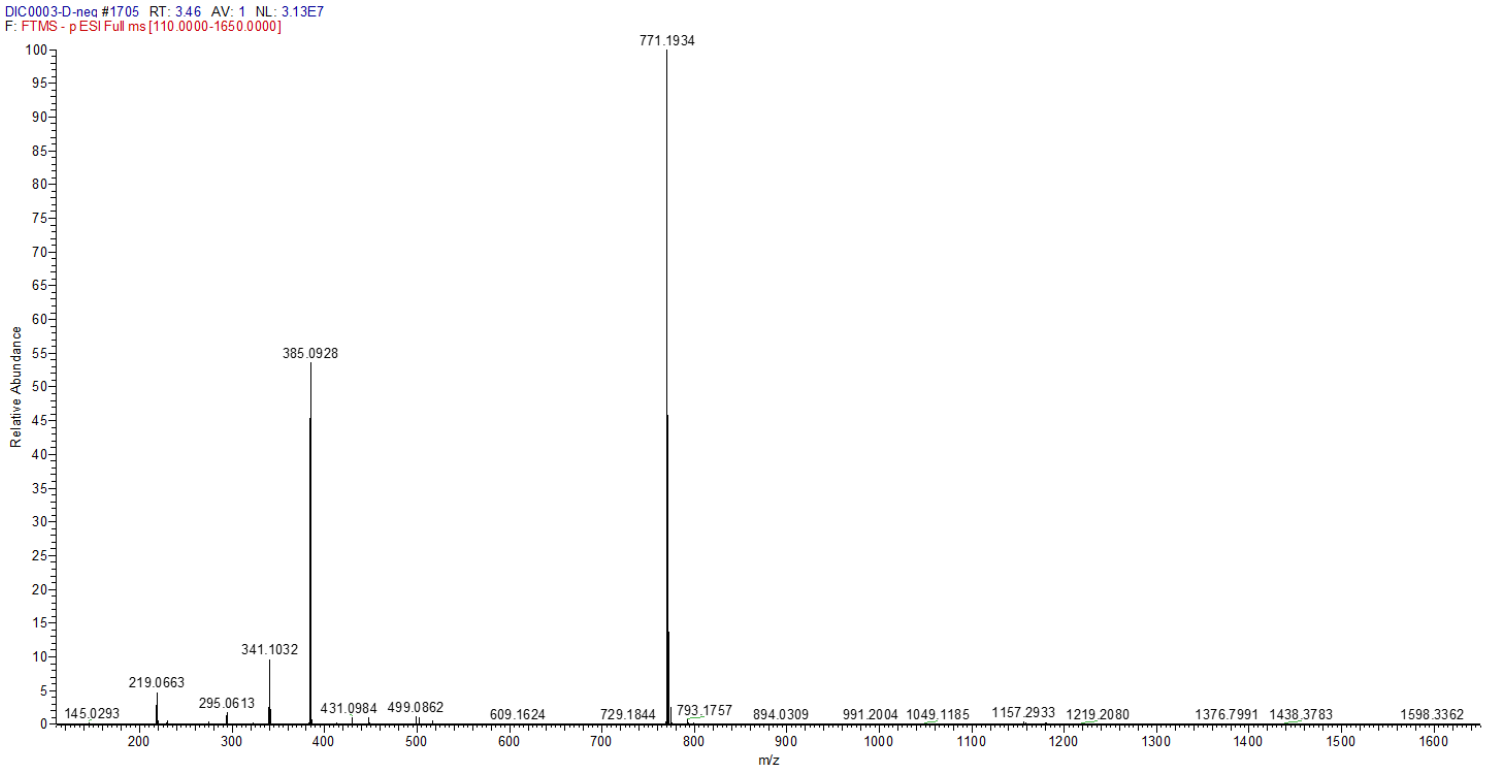


S9-II: Solid Phase extraction of oxidative crude. DiC003 and DiC004

S10. Compound **4**

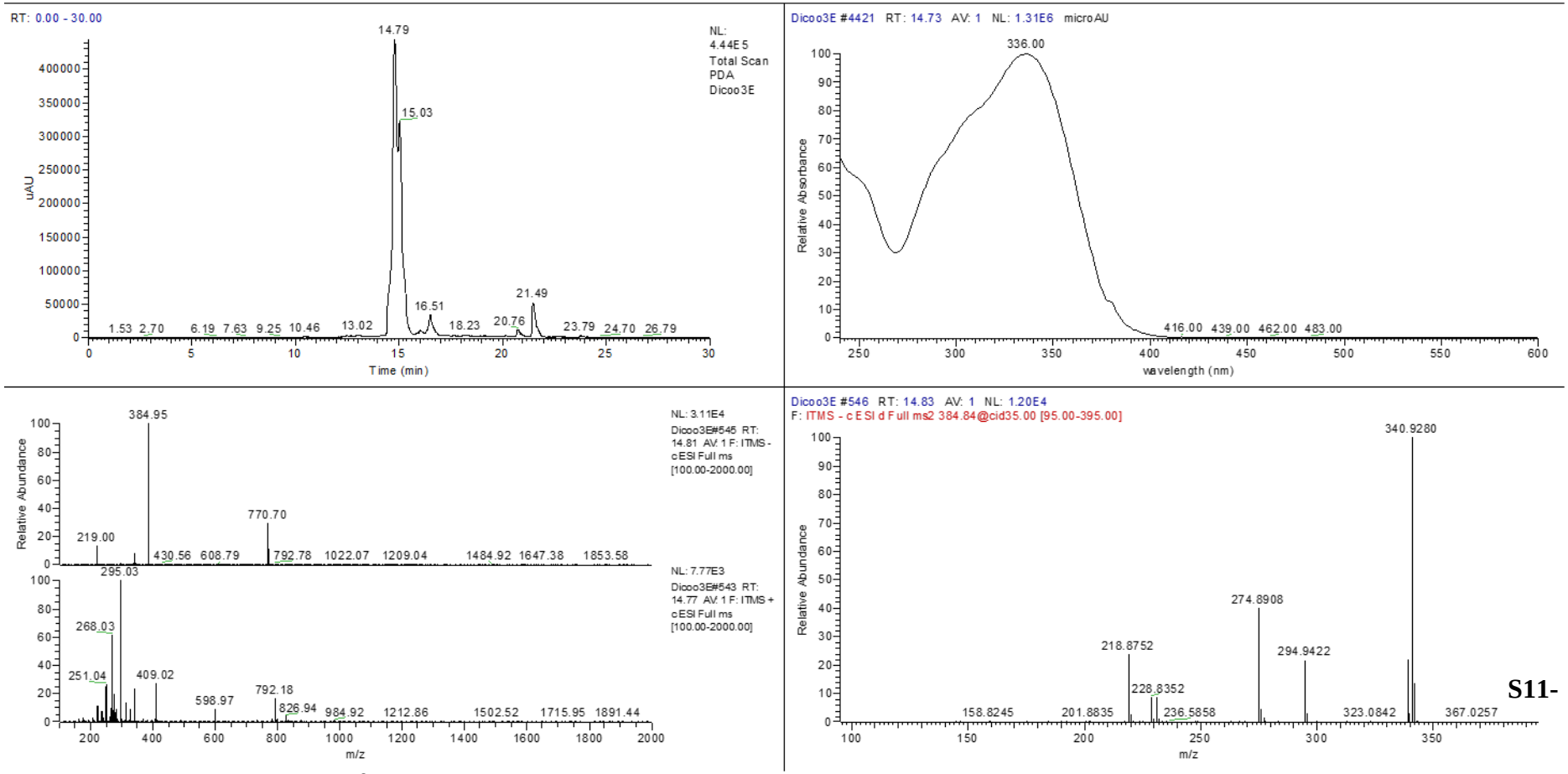


S10-I: HPLC-UV-ESI-MS/MS of compound **4**.



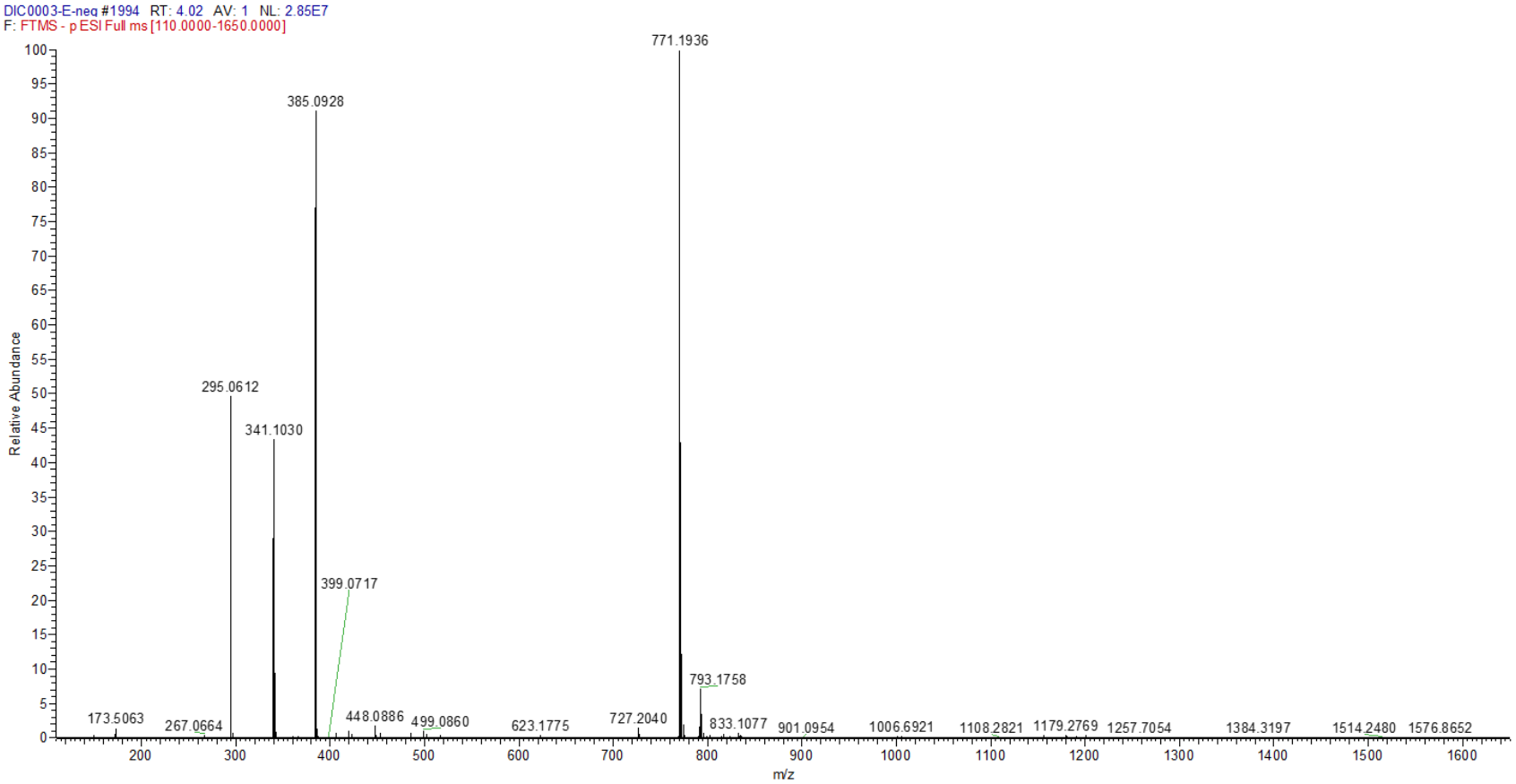
S10-II: Accurate mass of compound **4**.

S11. Compound **5**.



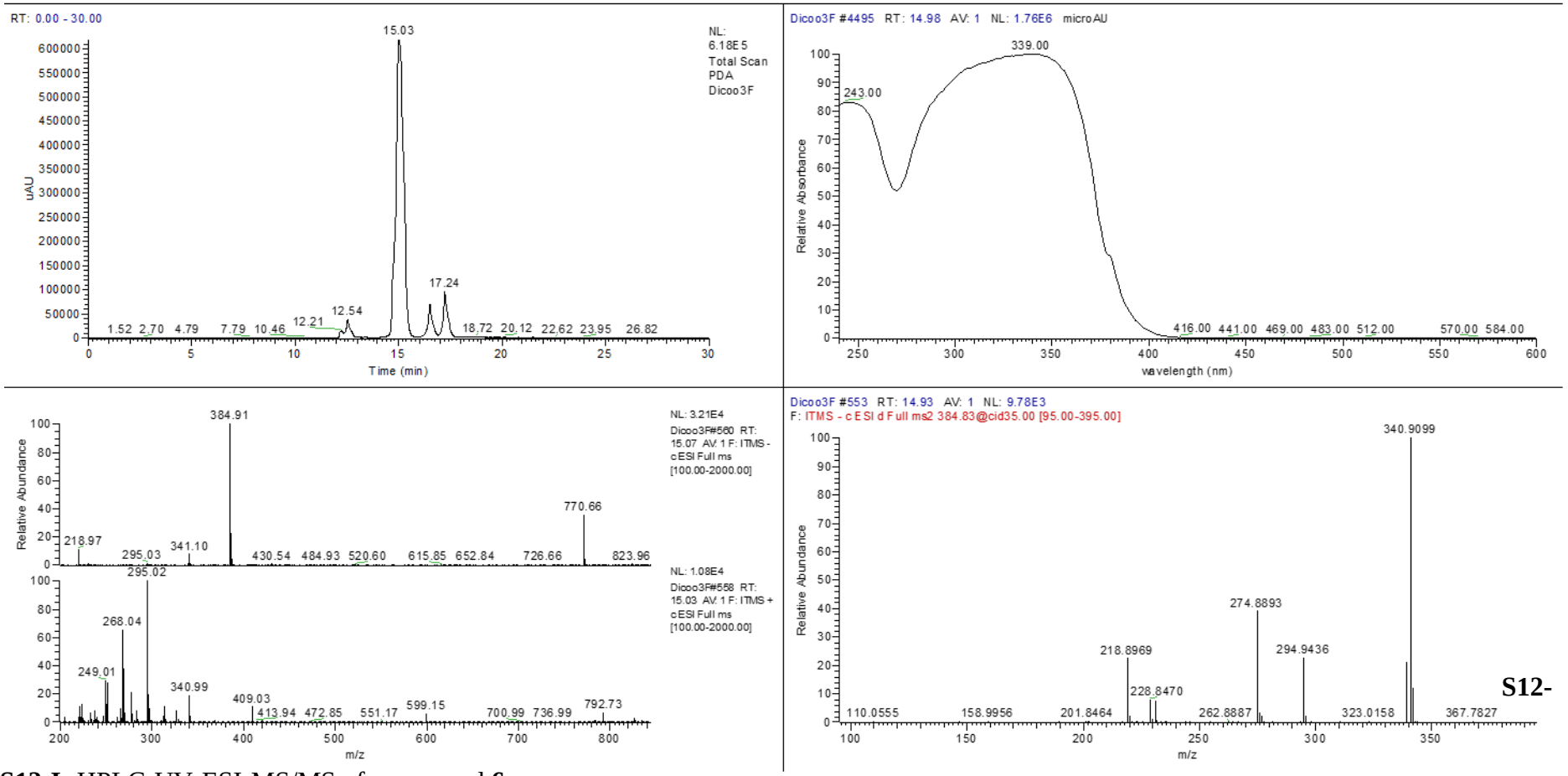
S11-I: HPLC-UV-ESI-MS/MS of compound **5**

S11-



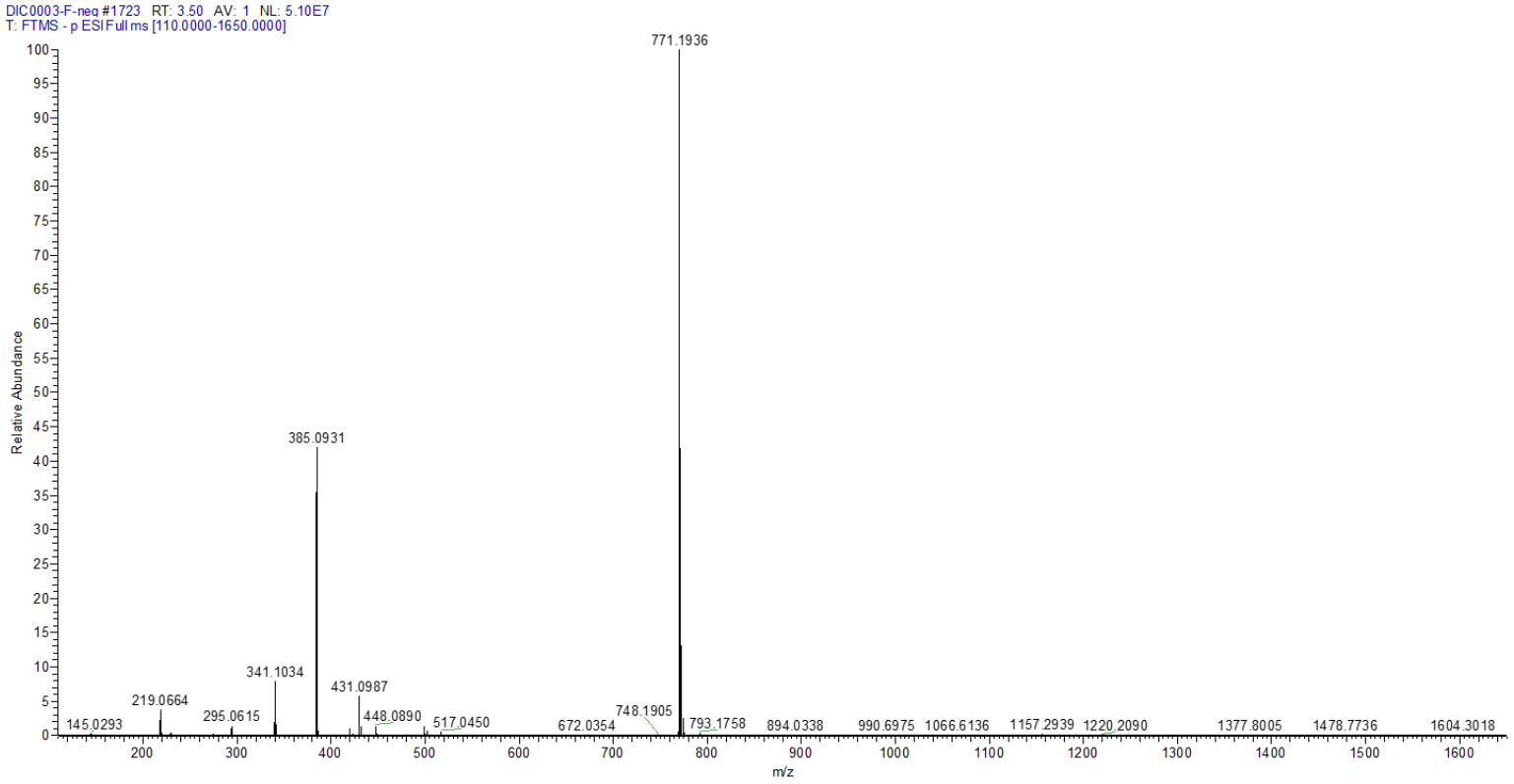
S11-II: Accurate mass of compound **5**.

S12. Compound **6**



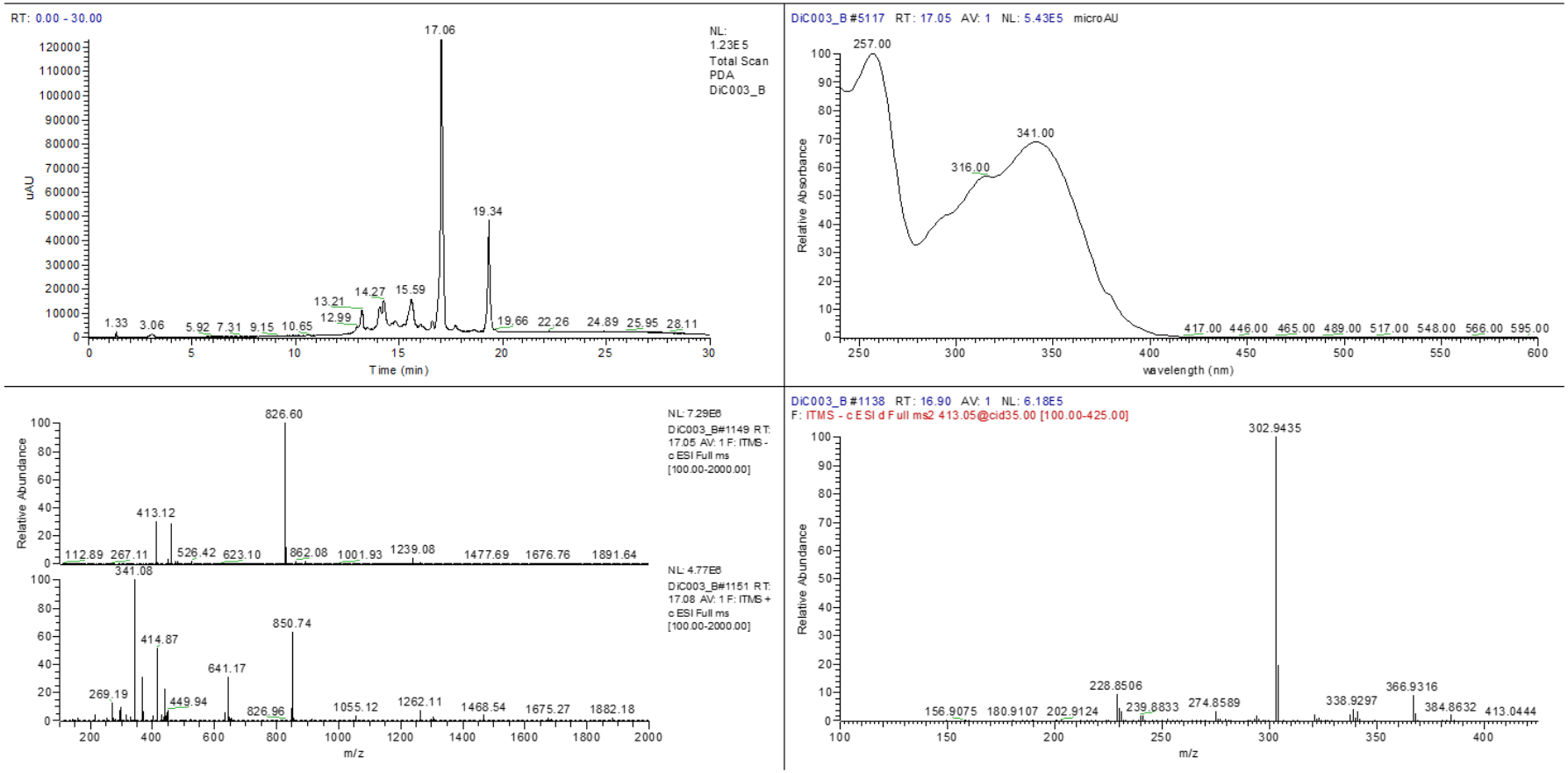
S12-I: HPLC-UV-ESI-MS/MS of compound **6**

S12-

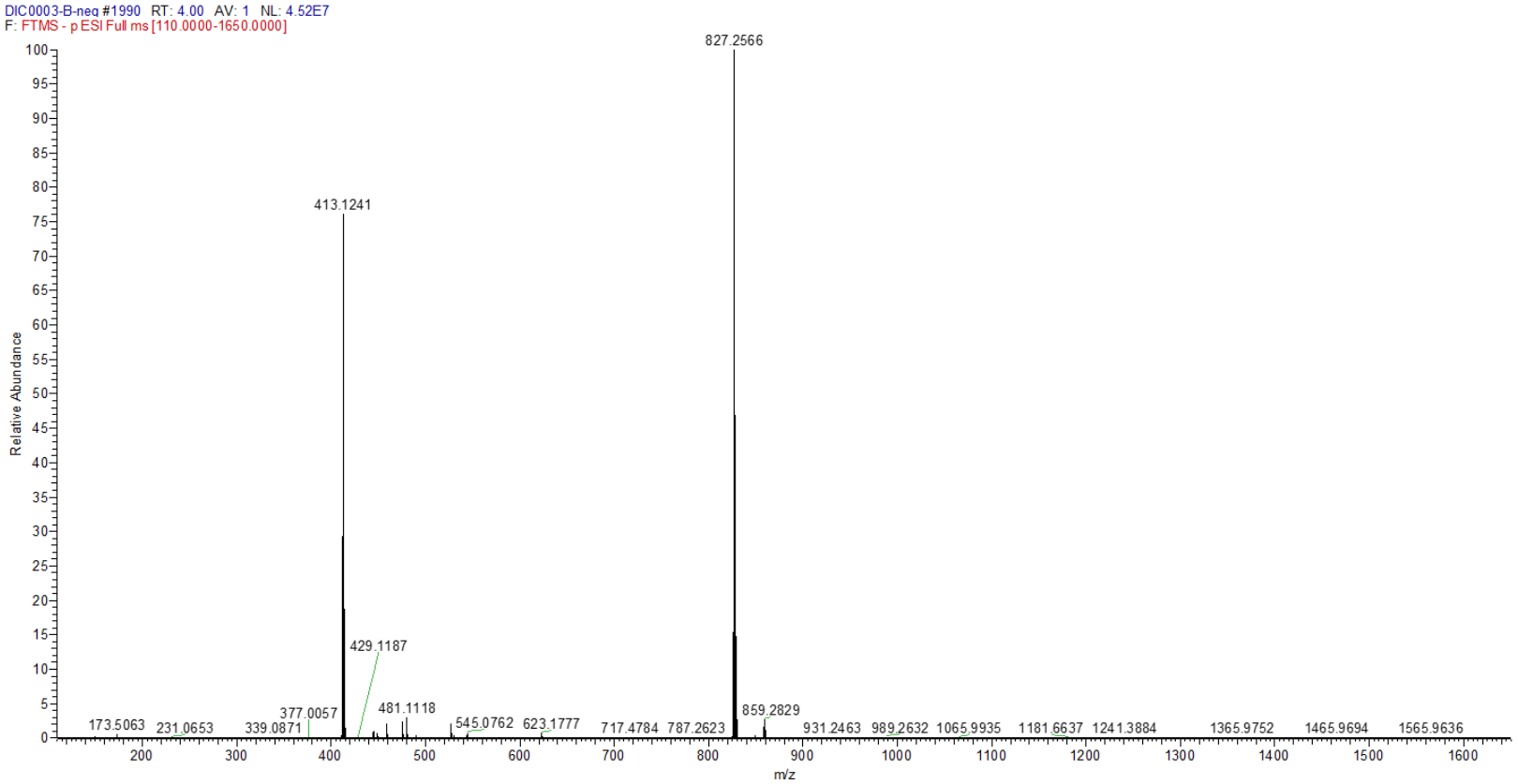


S12-II: Accurate mass of compound **6**

S13. Compound **7**

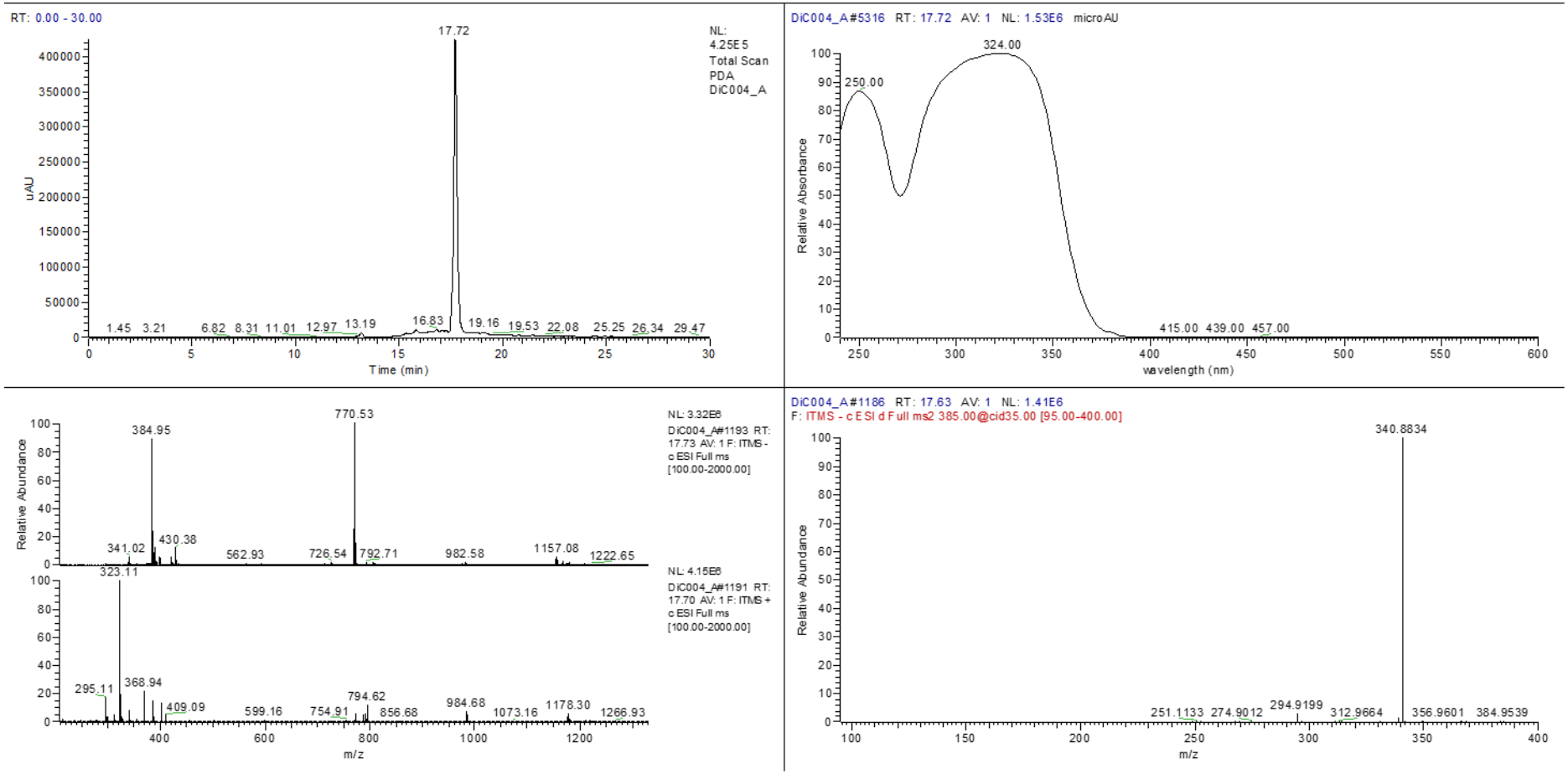


S13-I: HPLC-UV-ESI-MS/MS of compound **7.**

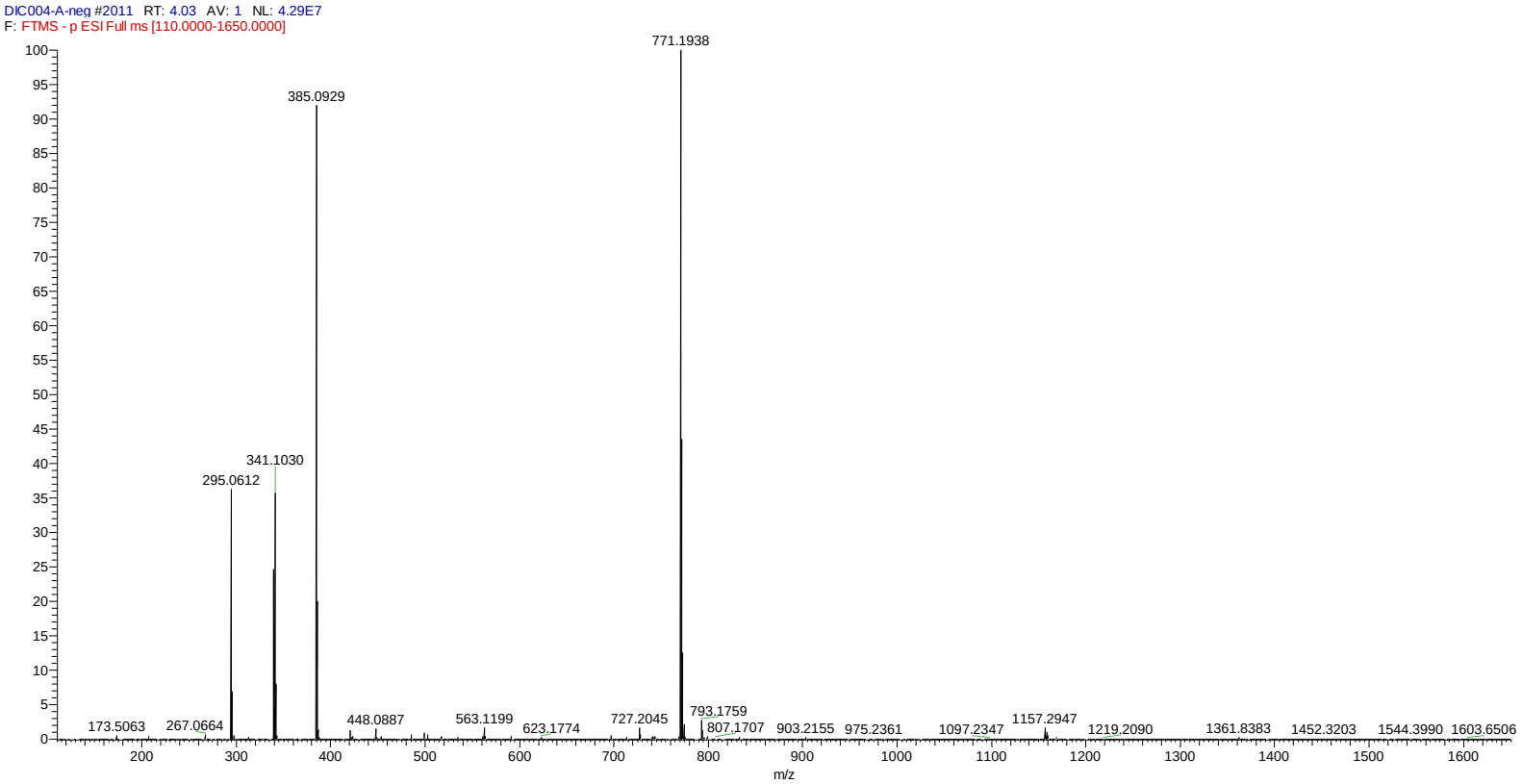


S13-II: Accurate mass of compound **7**

S14. Compound **8**

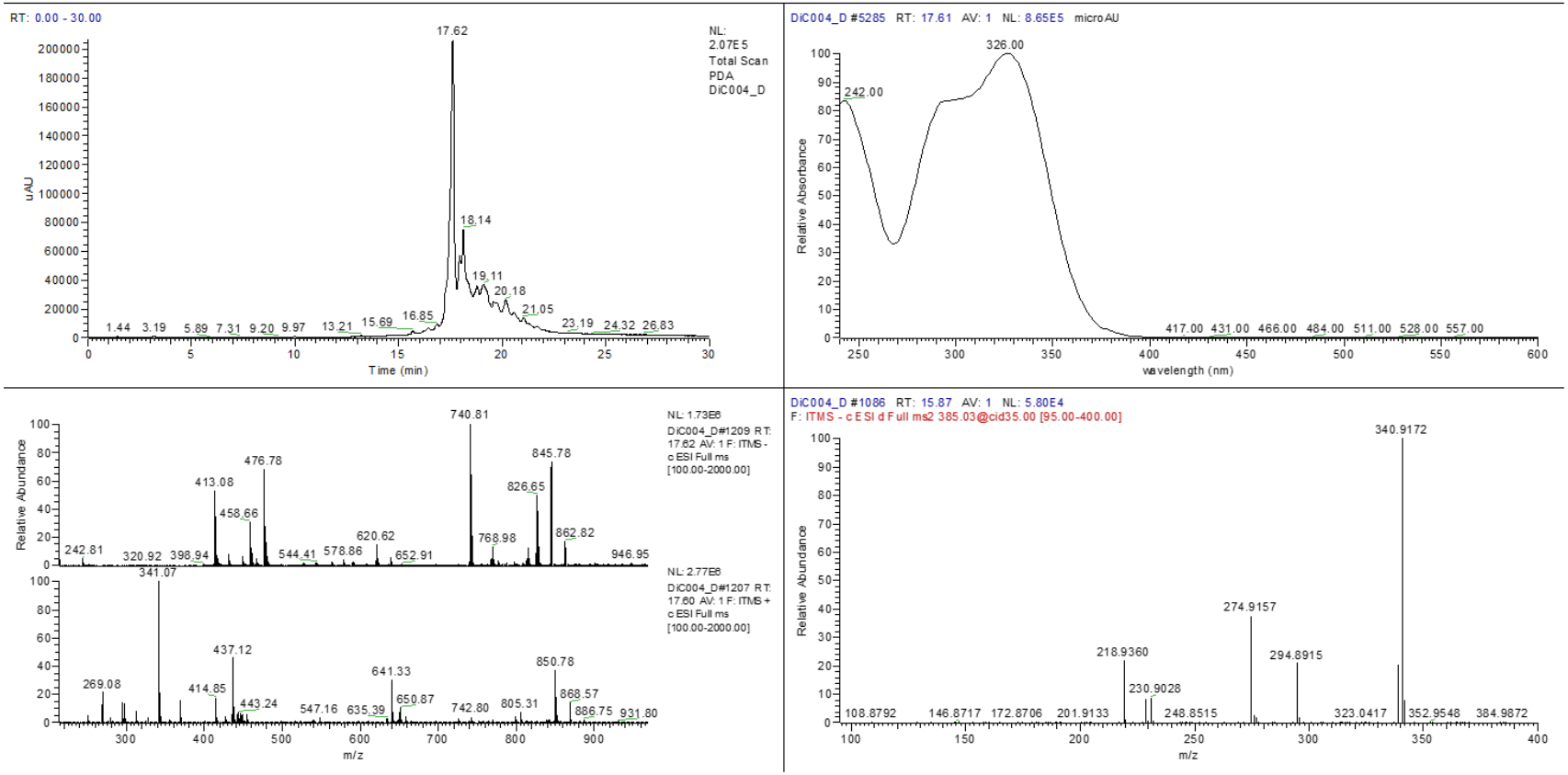


S14-I: HPLC-UV-ESI-MS/MS of compound **8**

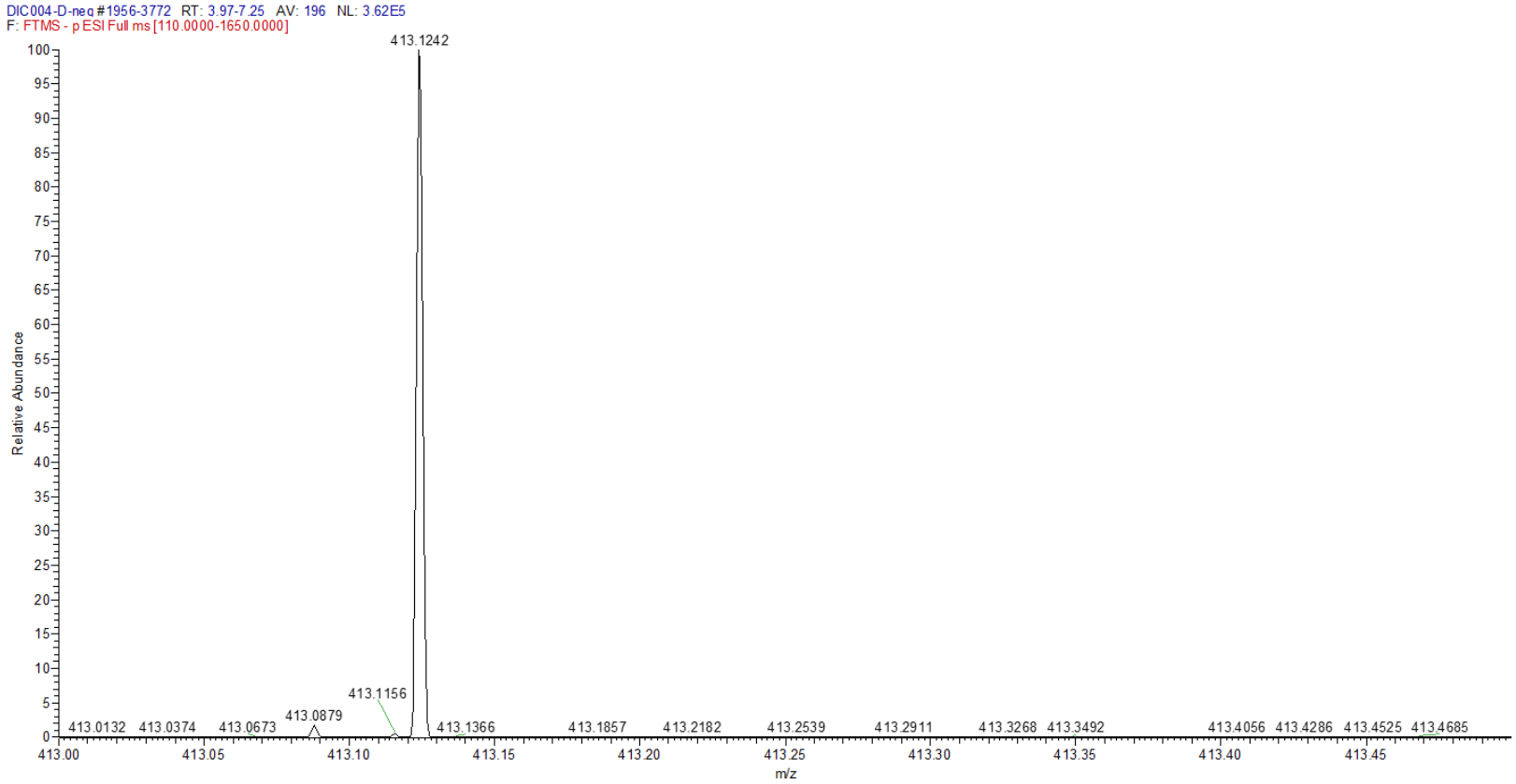


S14-II: Accurate mass of compound **8**

S15. Compound **9**

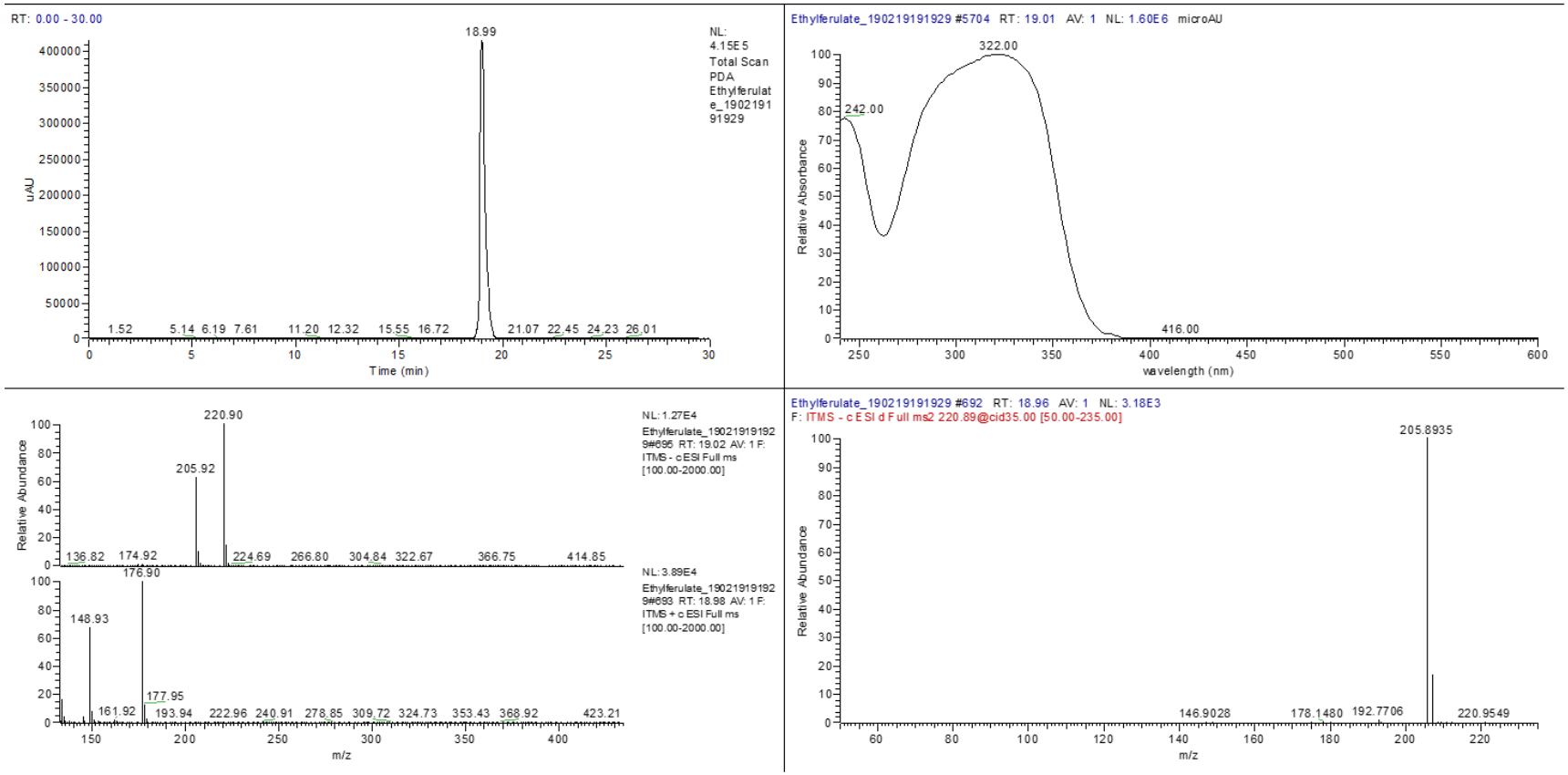


S15-I: HPLC-UV-ESI-MS/MS of compound **9**.

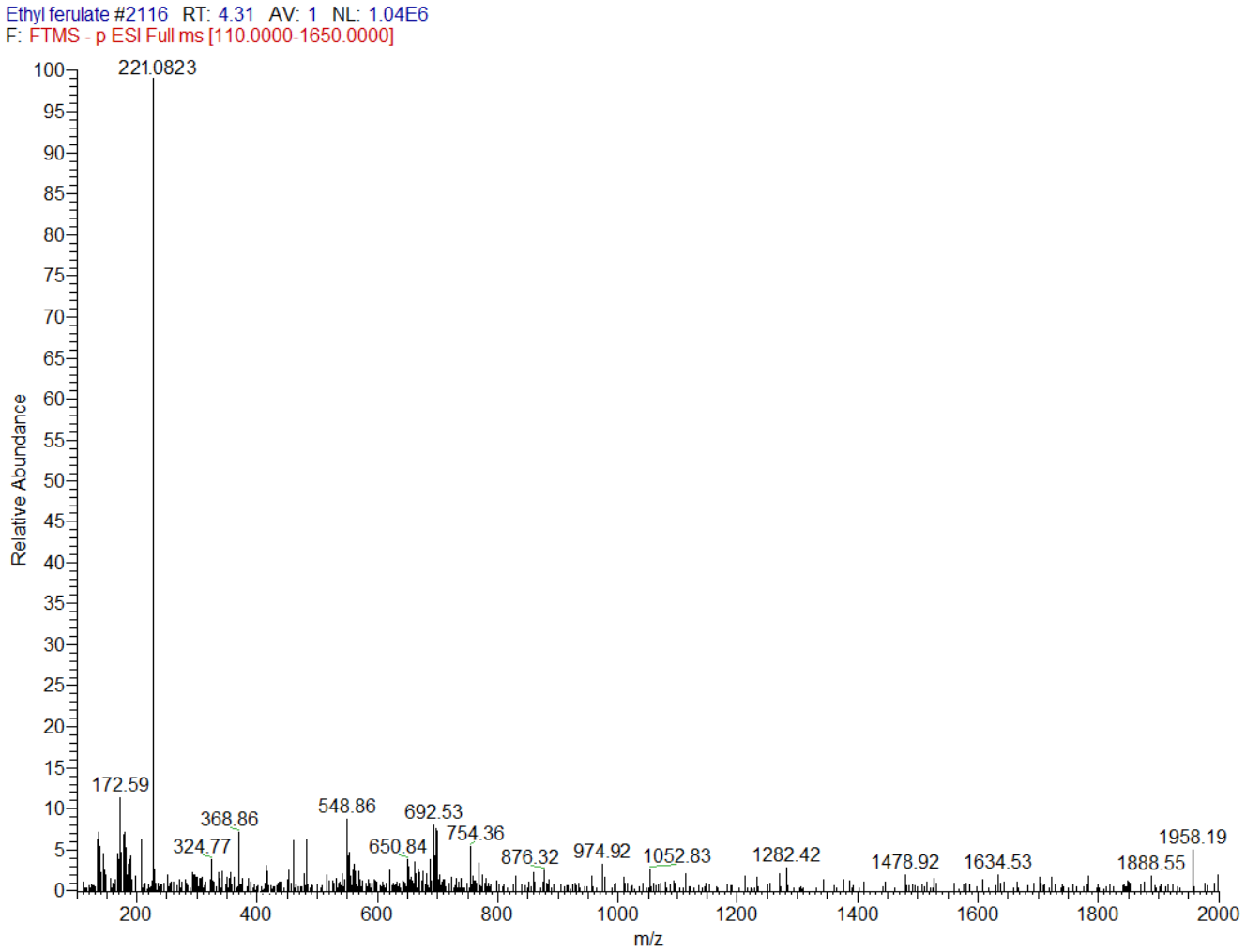


S15-II: Accurate mass of compound **9.**

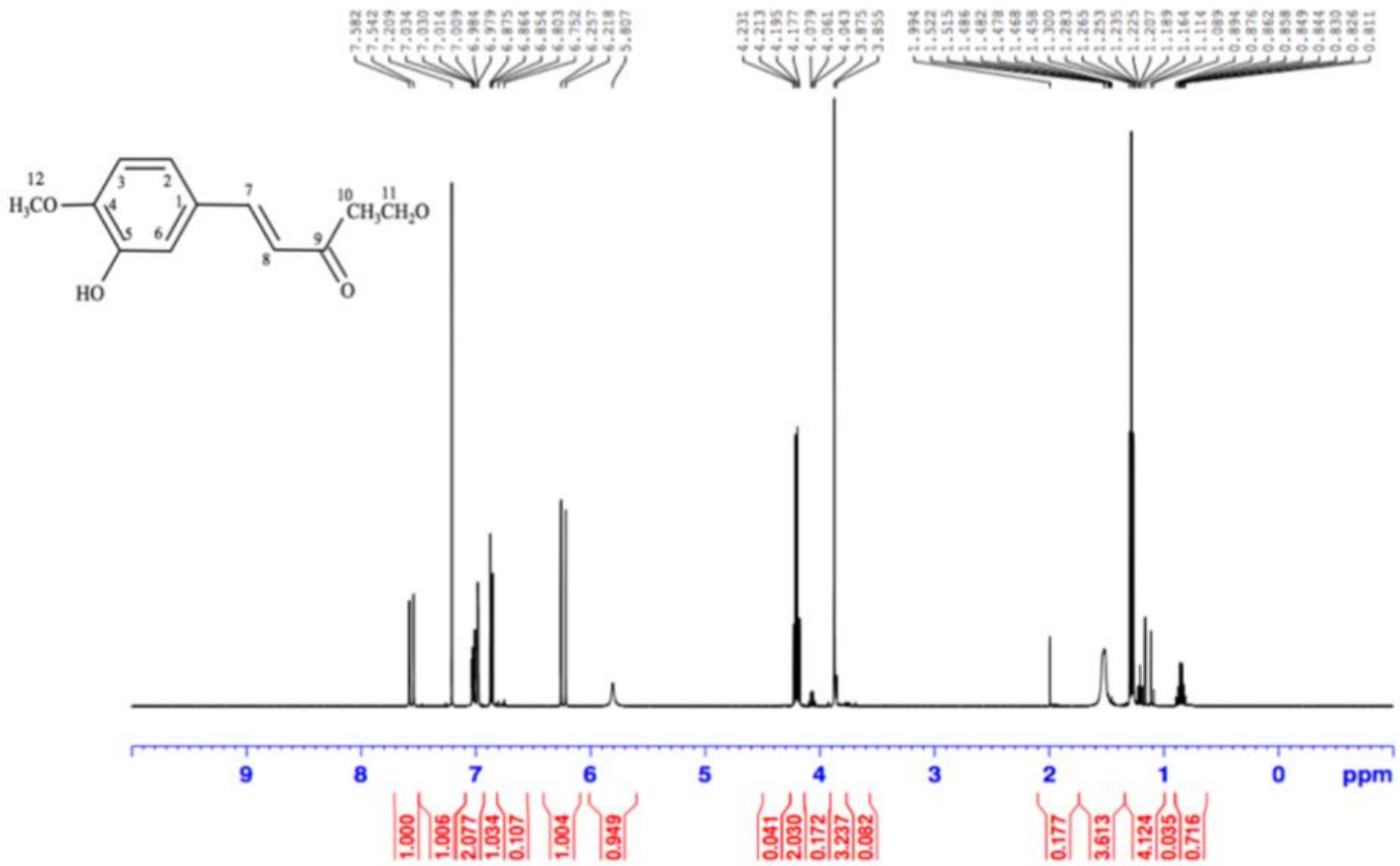
S16**.** Synthesis of ethyl ferulate



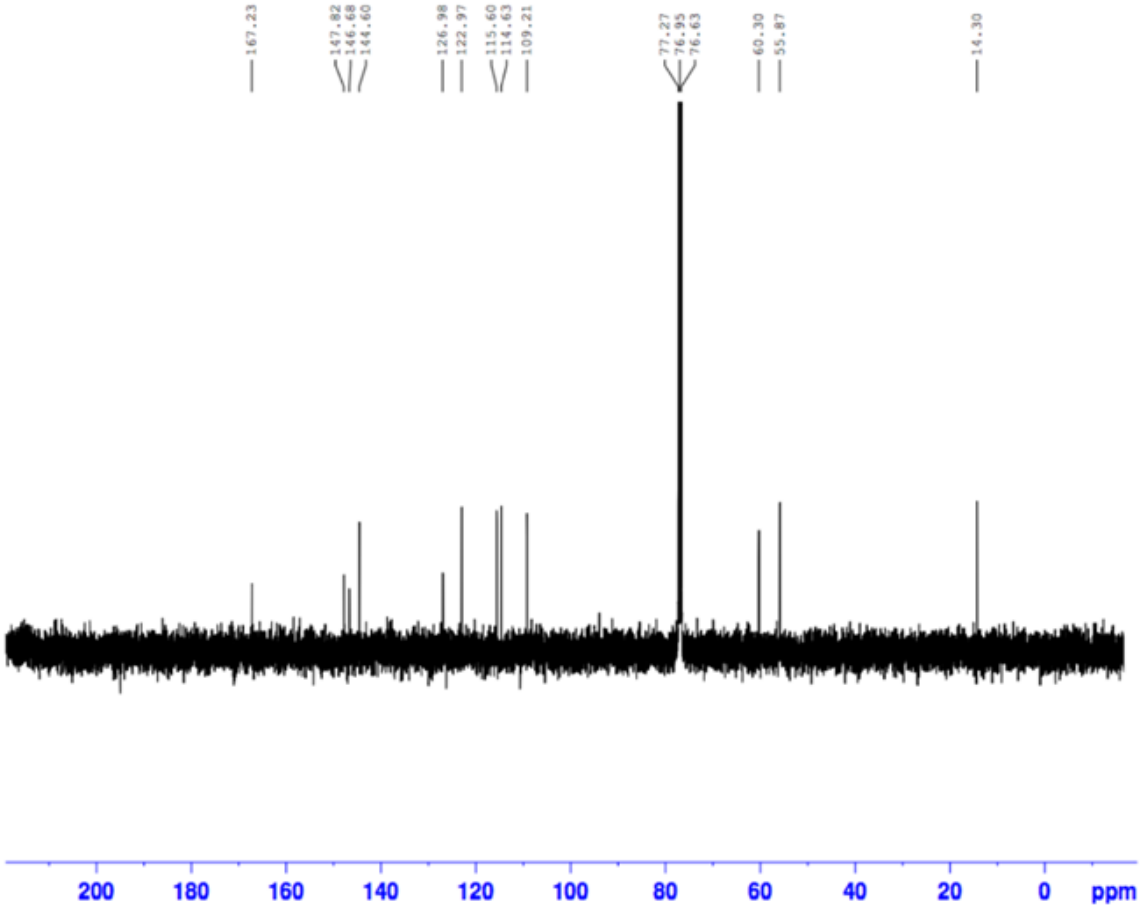
S16-I: HPLC-UV-ESI-MS/MS of ethyl ferulate



S16-II: Accurate mass of ethyl ferulate

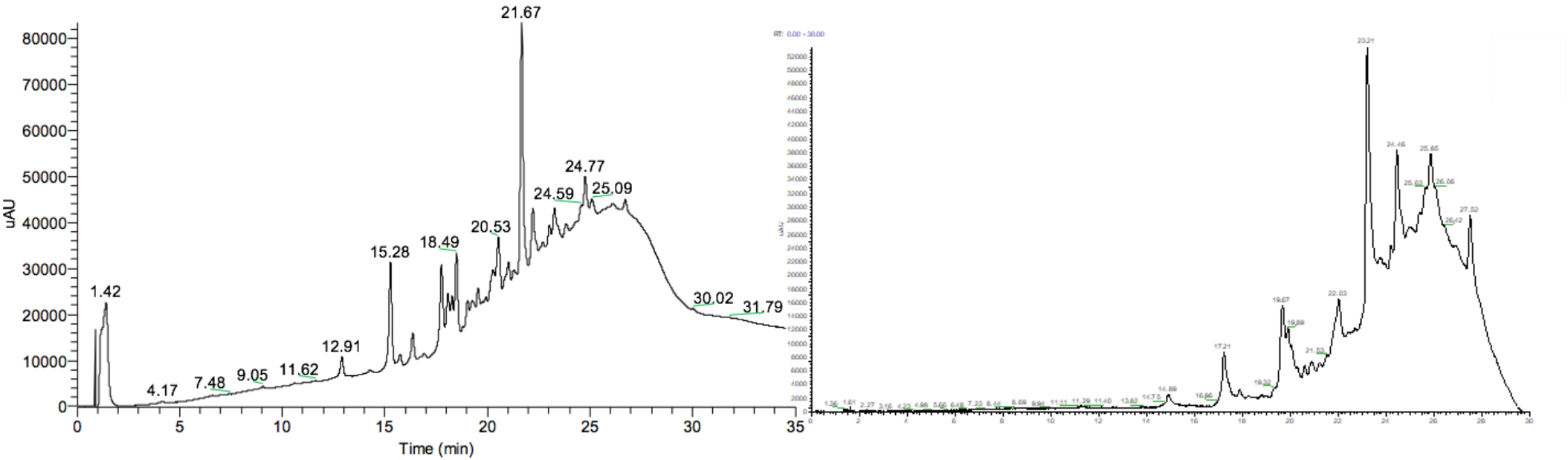


S16-III: 1H NMR spectrum of ethyl ferulate

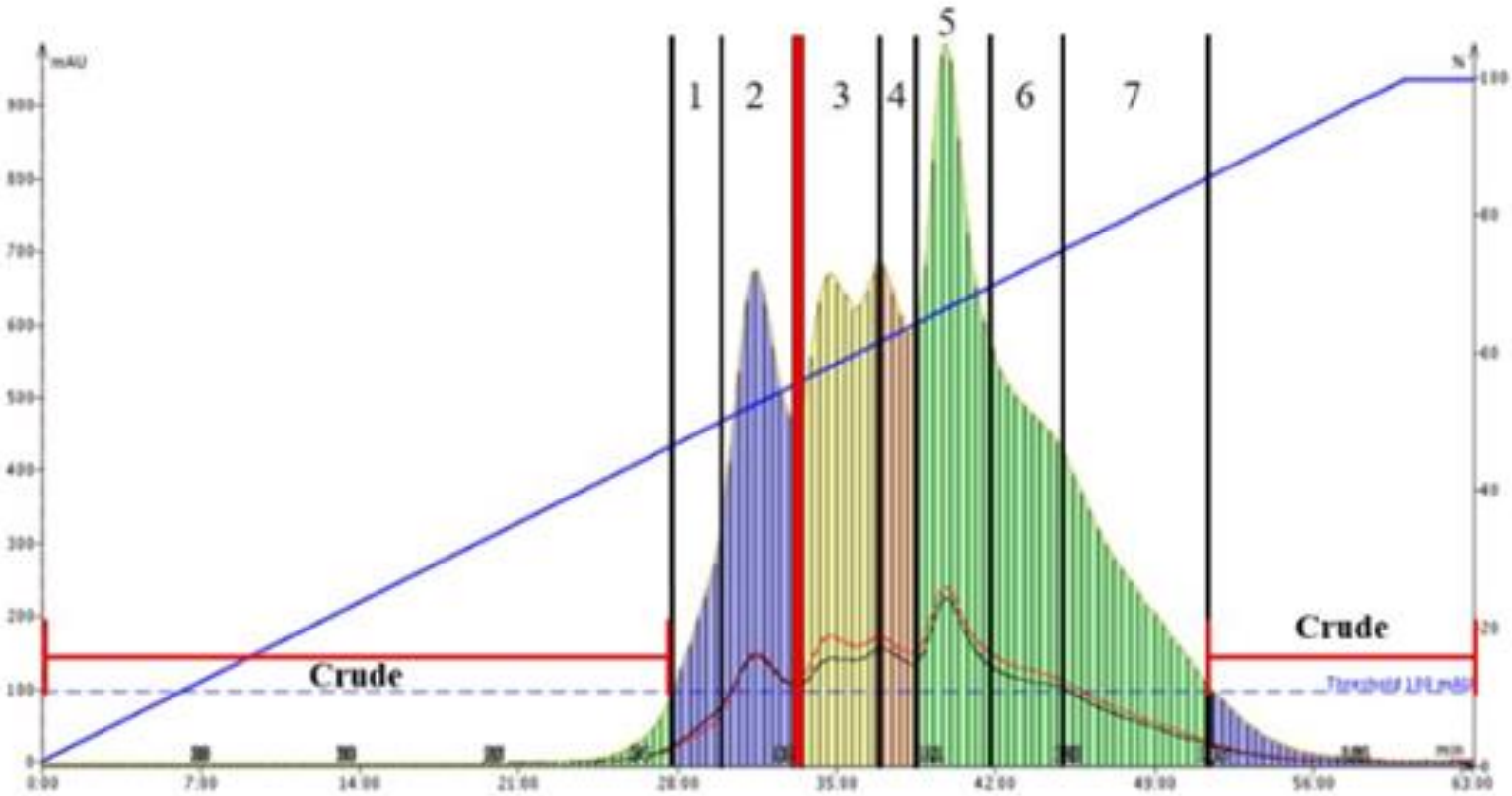


S16-IV: ^{13}C NMR spectrum of ethyl ferulate

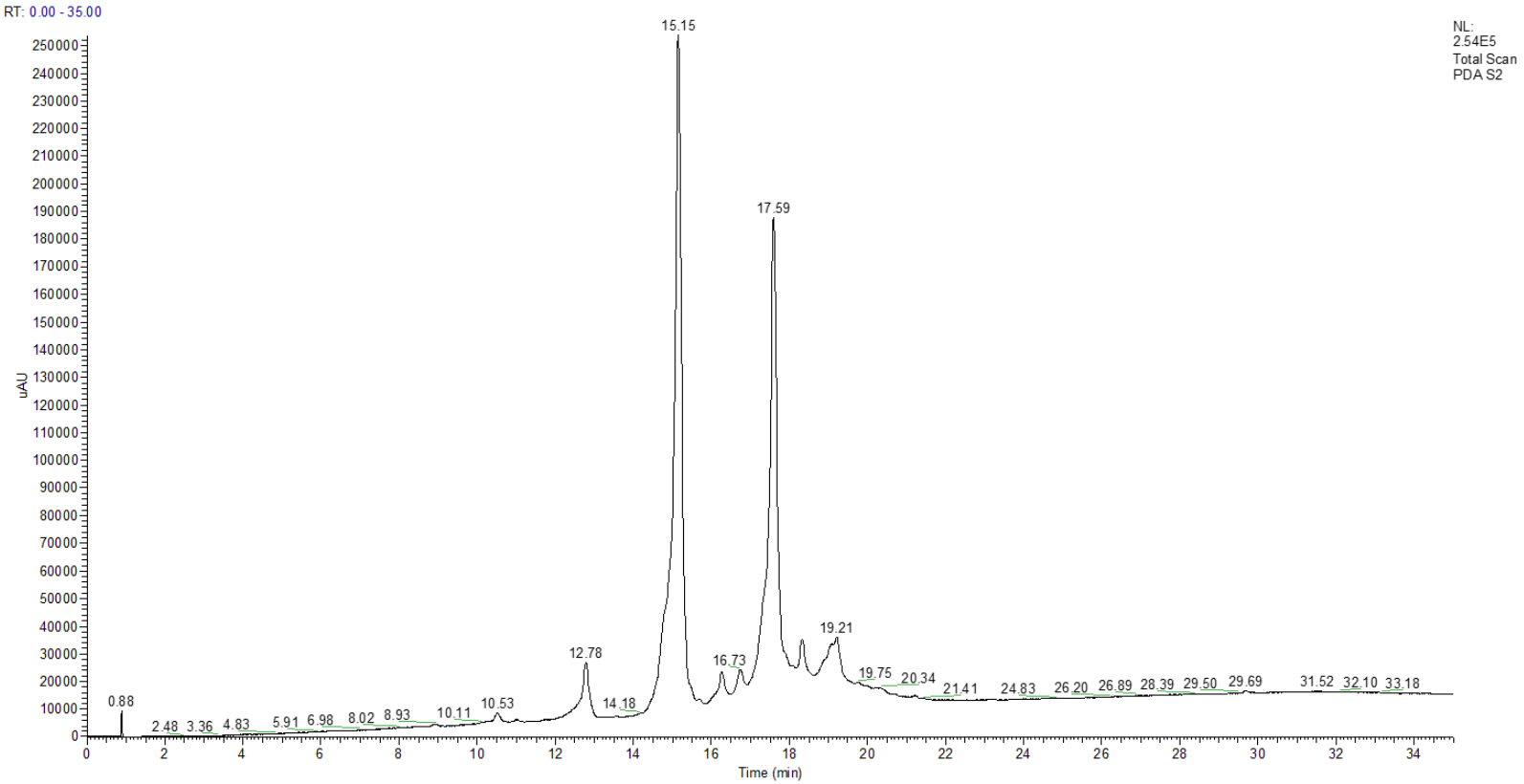
S17. Oxidative coupling of ethyl ferulate and ferulic acid.



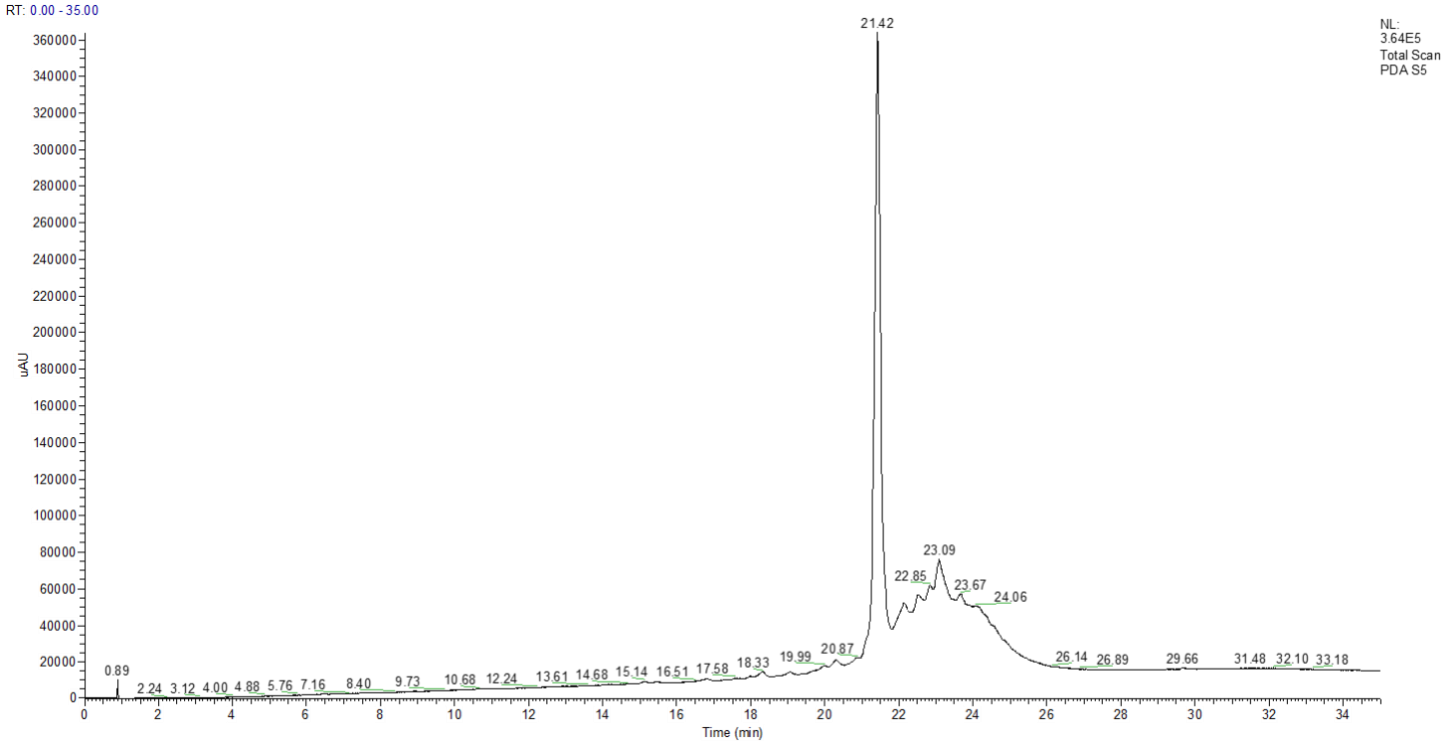
S17-I: HPLC-UV-ESI-MS/MS of both crudes of ethyl ferulate and ferulic acid.



S17-II: UV chromatogram of Biotage Isolera Flash Chromatography.

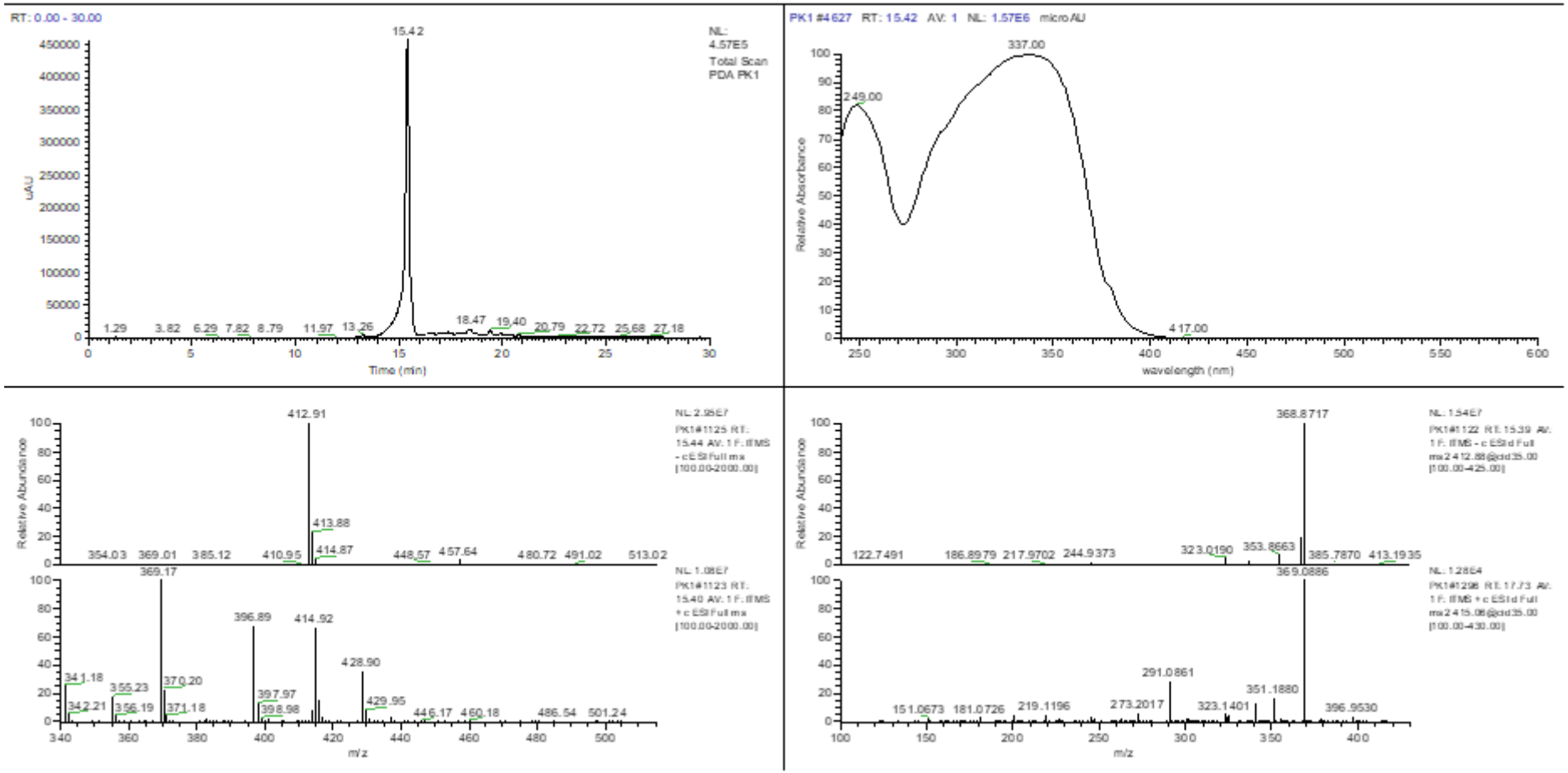


S17-III: HPLC-UV-ESI-MS/MS of fractions S2.

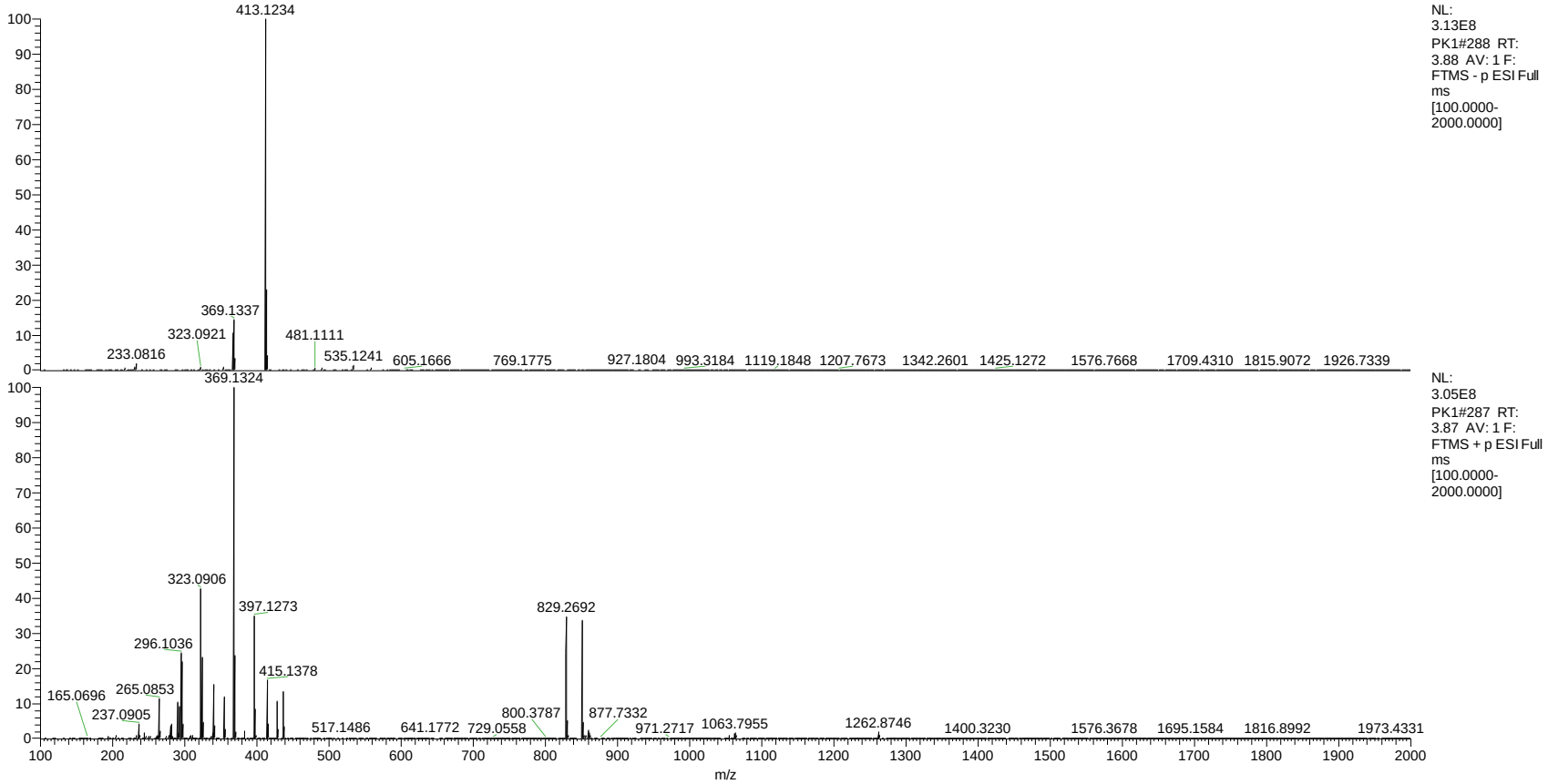


S17-III: HPLC-UV-ESI-MS/MS of fractions S5.

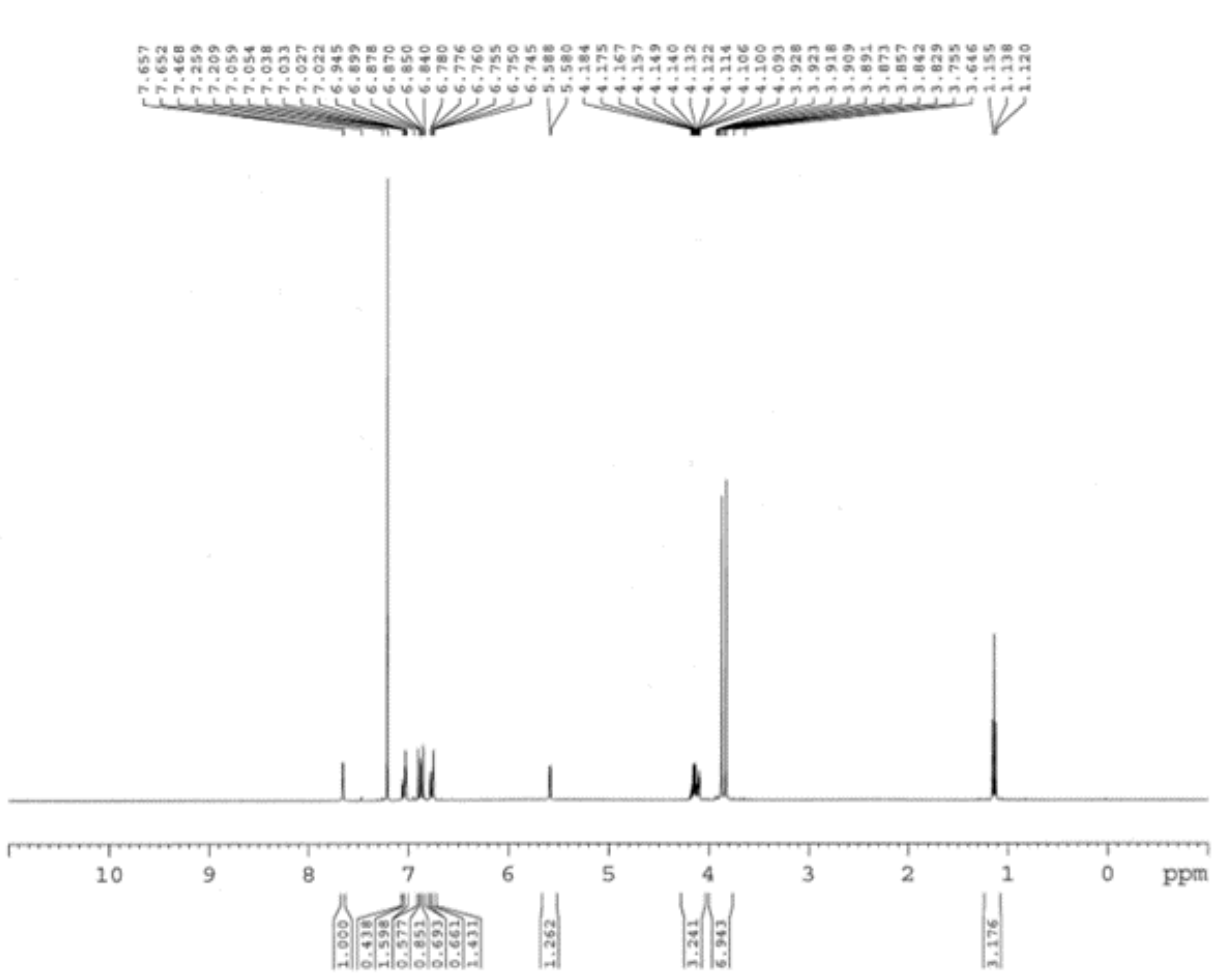
S18. Compound **10**



S18-I: HPLC-UV-ESI-MS/MS of compound **10.**

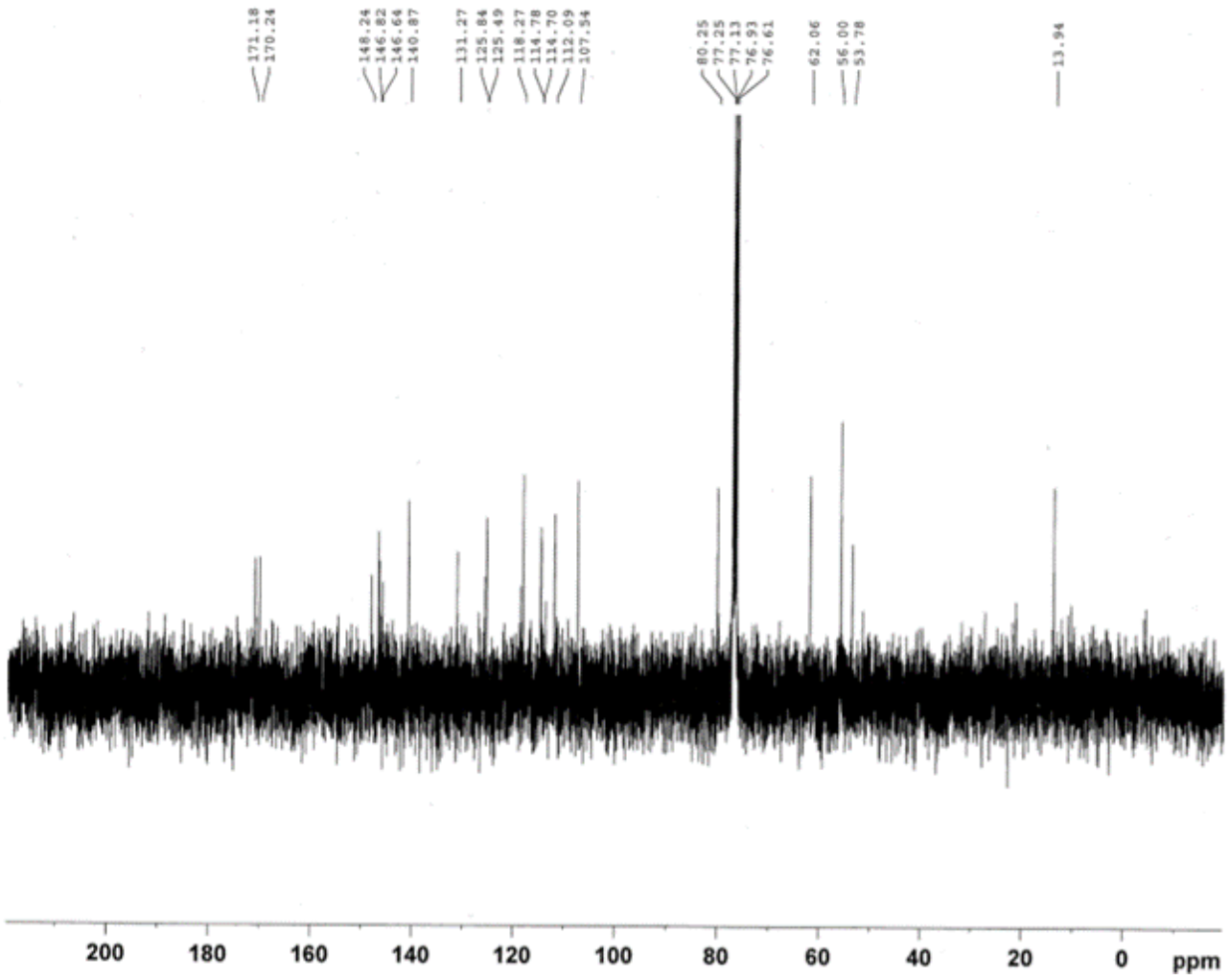


S18-II: Accurate mass of compound **10.**

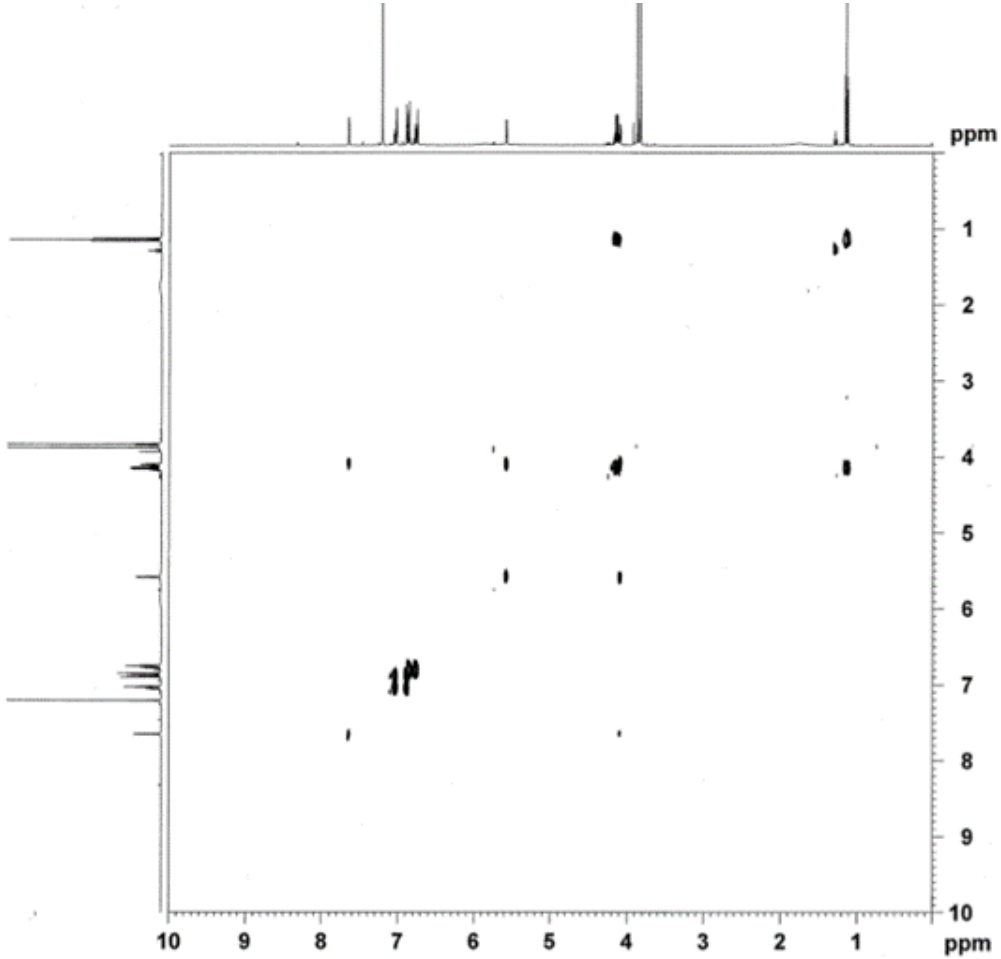


PROBHD	5 mm PABBO BB-
PULPROG	zg30
TD	65536
SOLVENT	CDCl3
NS	16
DS	2
SWH	8223.685 Hz
FIDRES	0.125483 Hz
AQ	3.9846387 sec
RG	287
DW	60.800 usec
DE	17.48 usec
TE	298.2 K
D1	1.00000000 sec
TD0	1
======== CHANNEL f1 ========	
NUC1	1H
P1	11.90 usec
PL1	-1.00 dB
PL1W	12.26963711 W
SFO1	400.0824707 MHz
SI	65536
SF	400.0800000 MHz
WDW	EM
SSB	0
LB	0.20 Hz
GB	0
PC	1.00

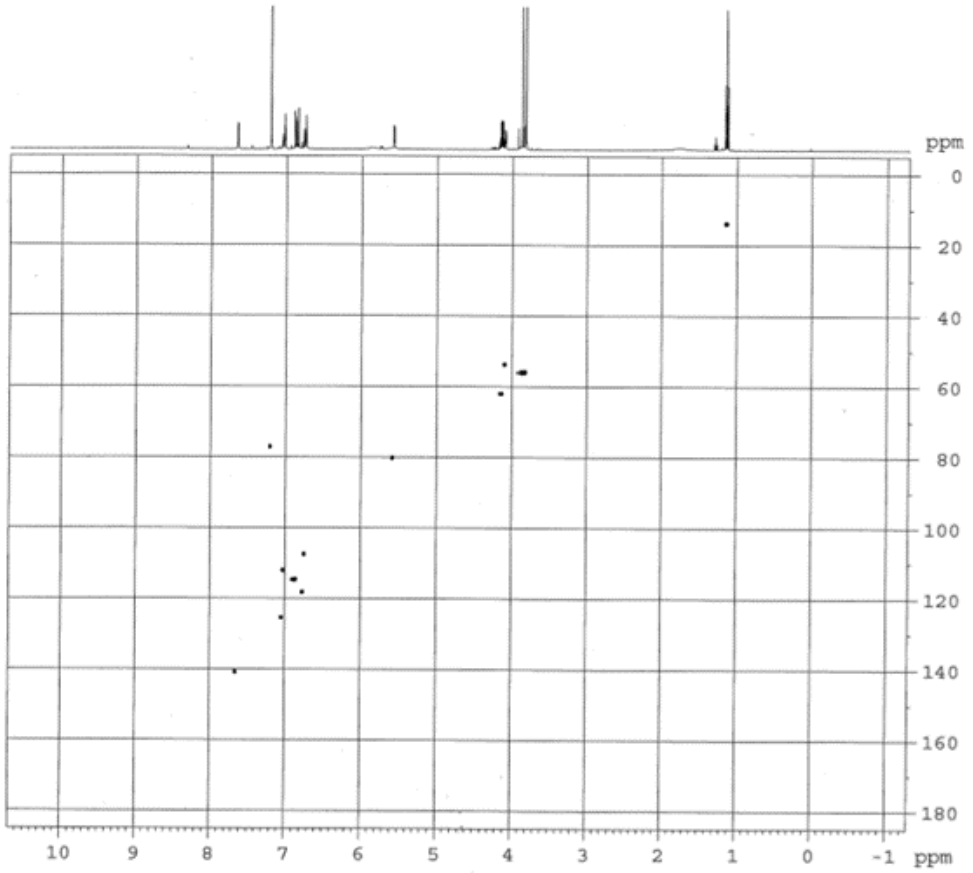
S18-III: 1H NMR spectrum of compound **10**



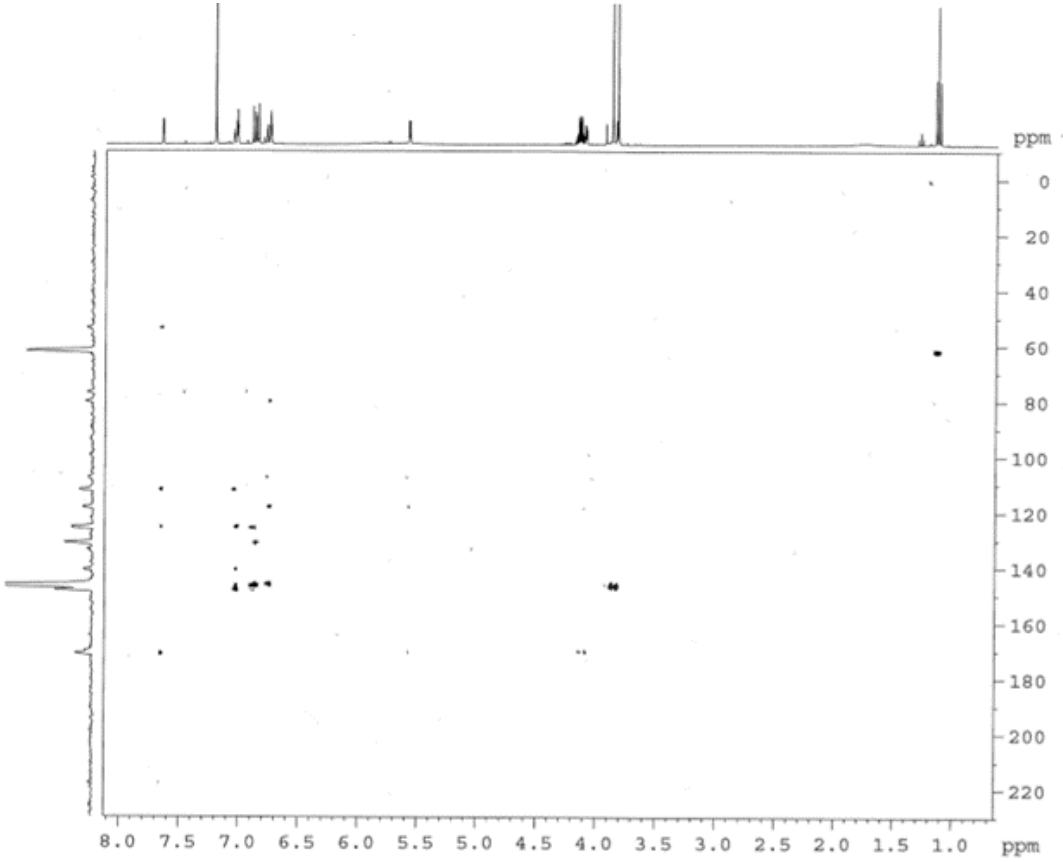
S18-IV: ^{13}C NMR spectrum of compound **10.**



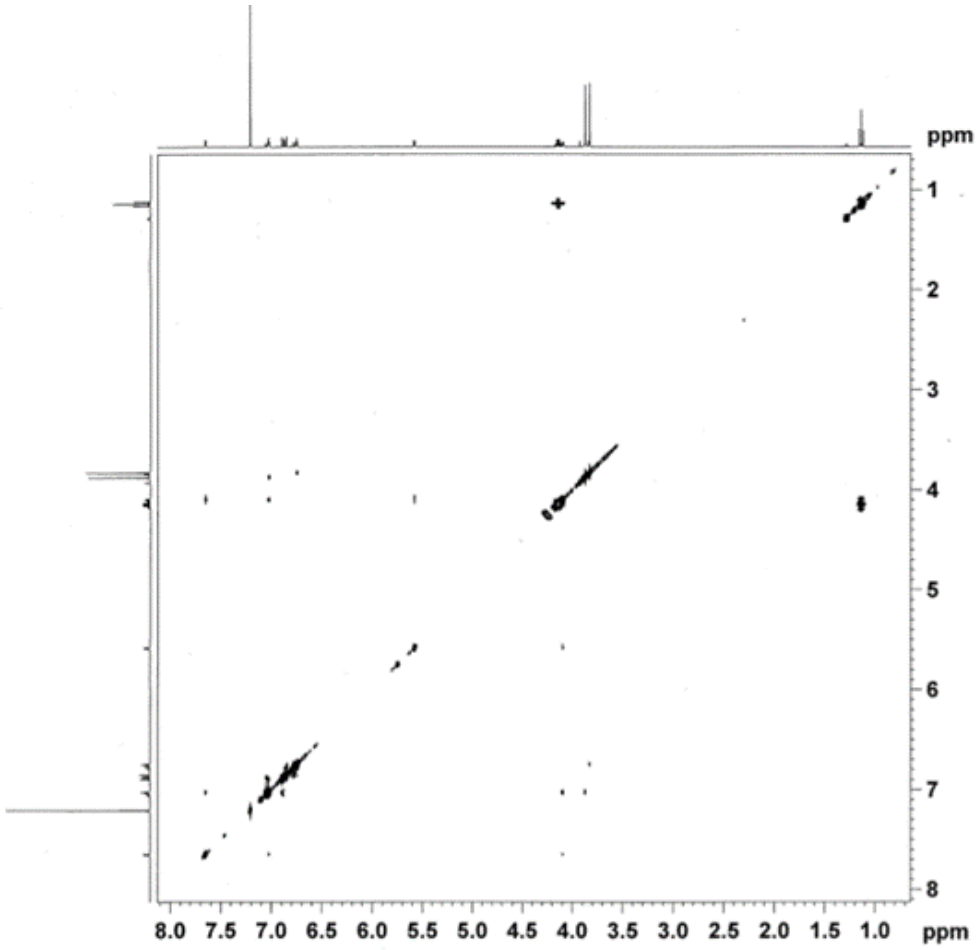
S18-V: COSY spectrum of compound **10.**



S18-VI: HSQC spectrum of compound **10.**

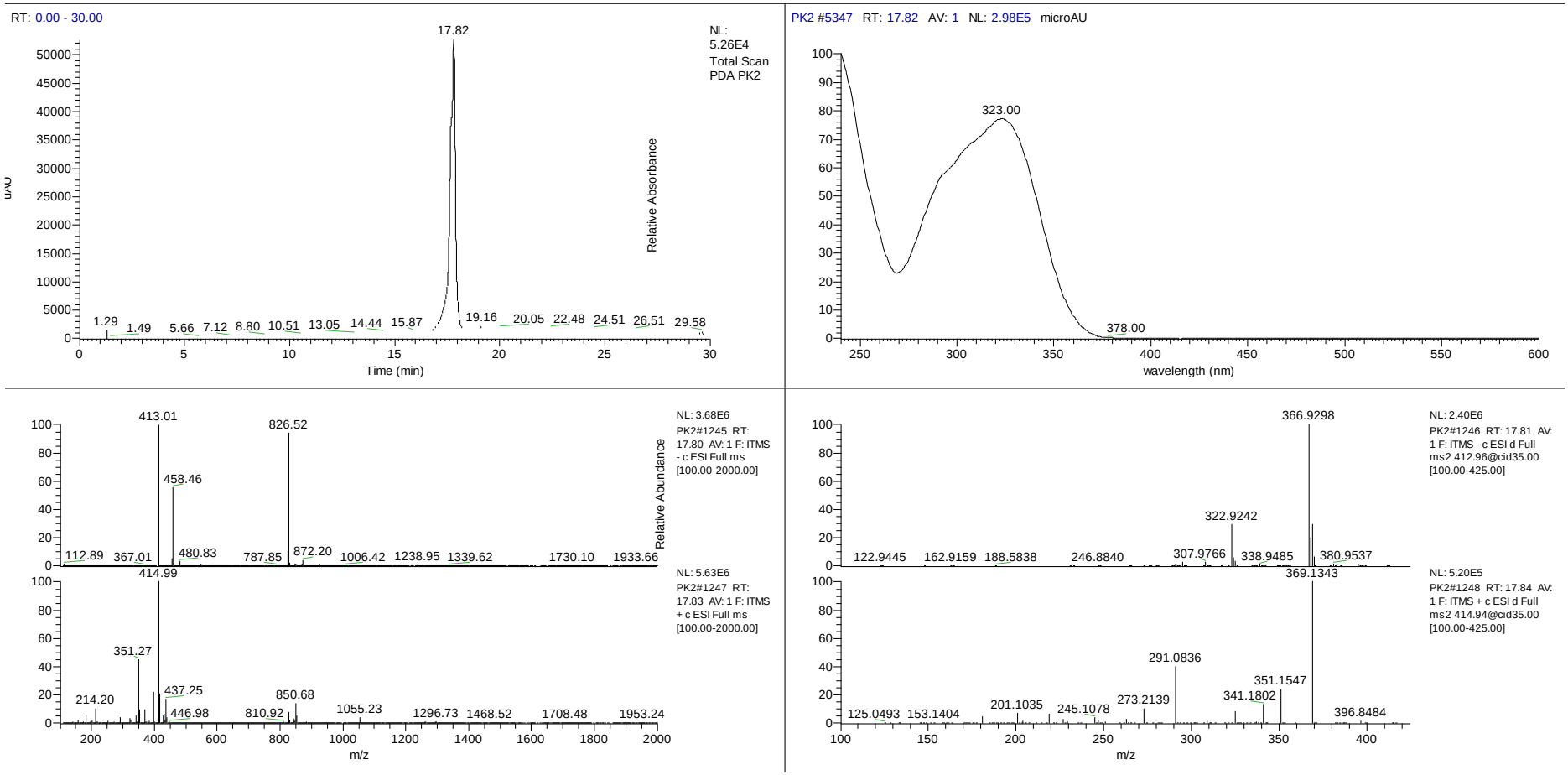


S18-VII: HMBC spectrum of compound **10.**

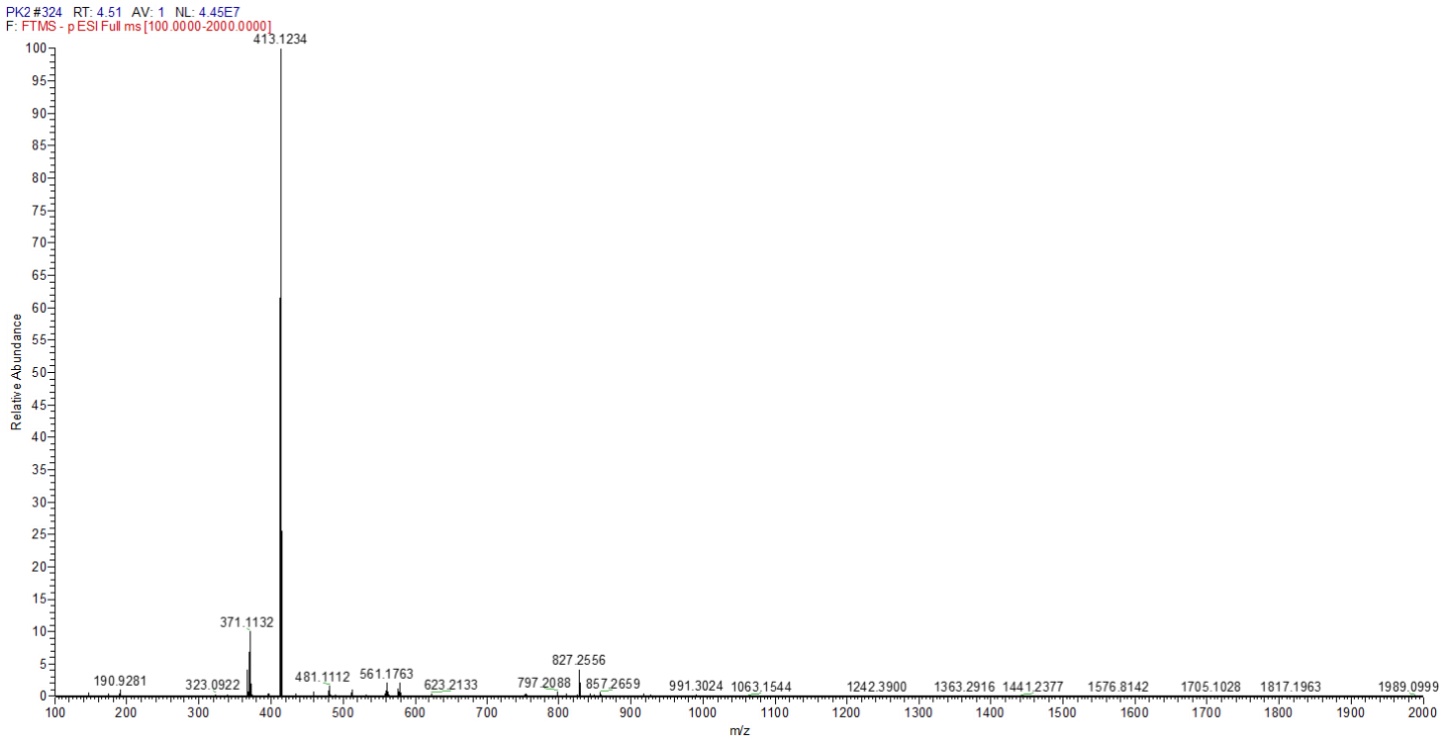


S18-VIII: NOESY spectrum of compound **10.**

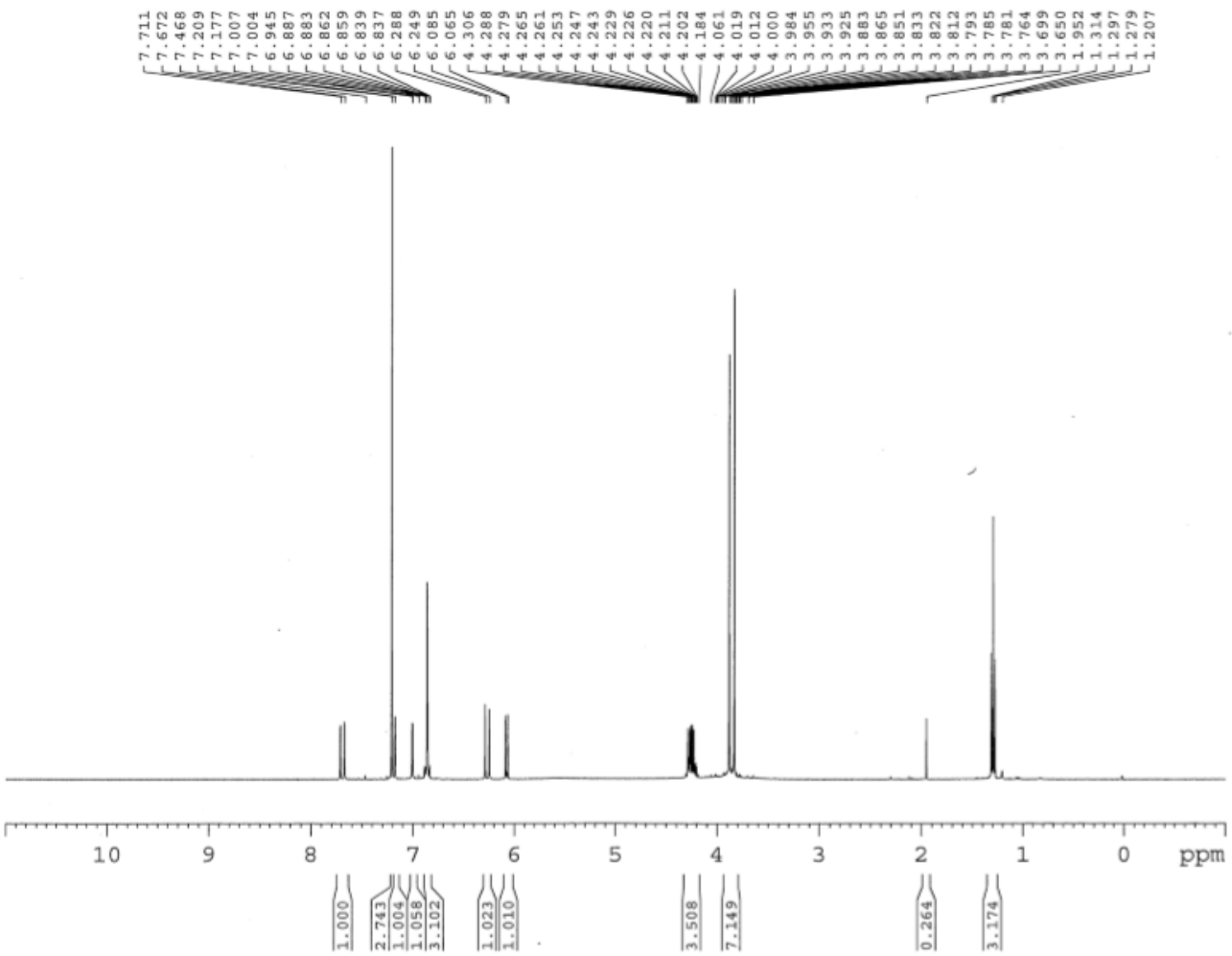
S19. Compound **11**



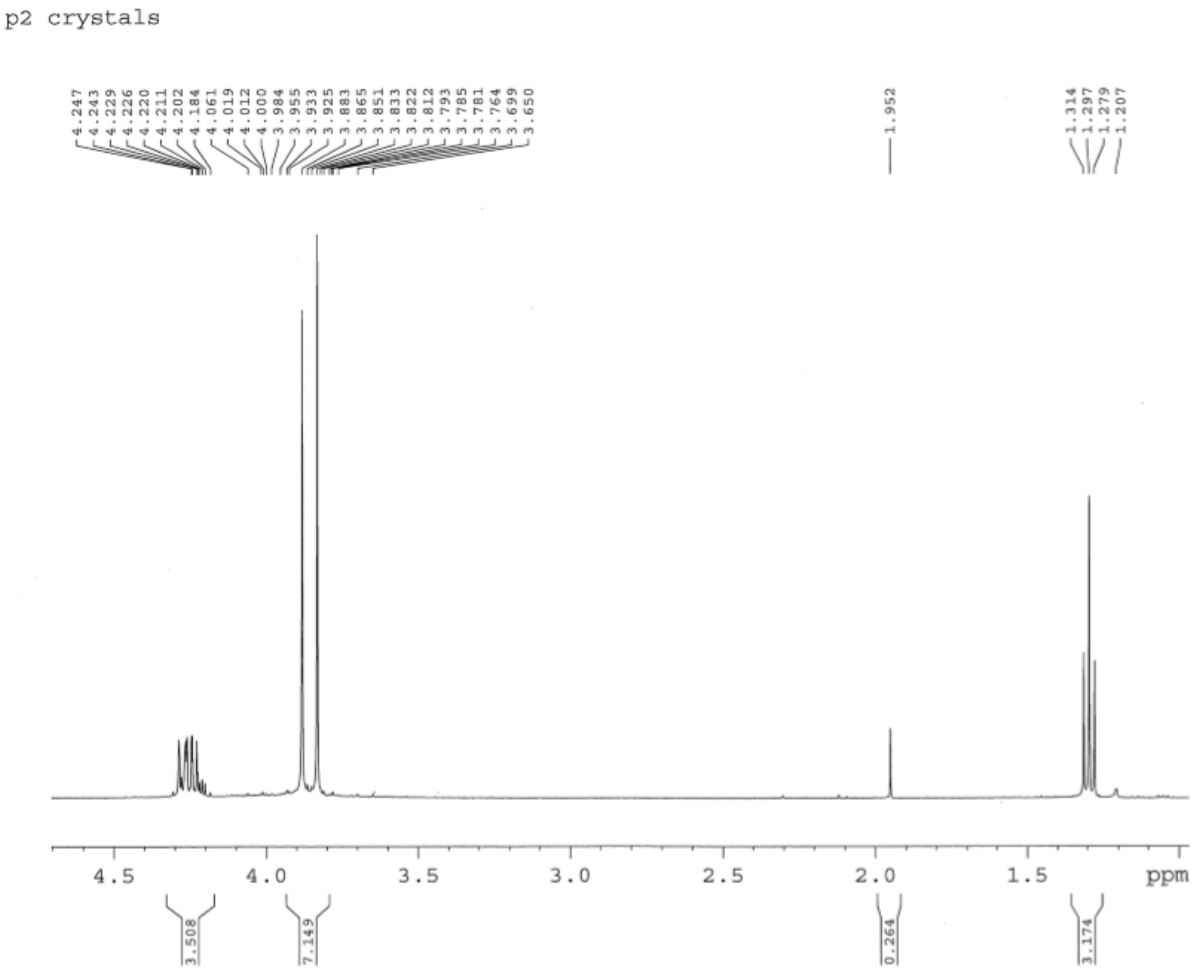
S19-I: HPLC-UV-ESI-MS/MS of compound **11.**



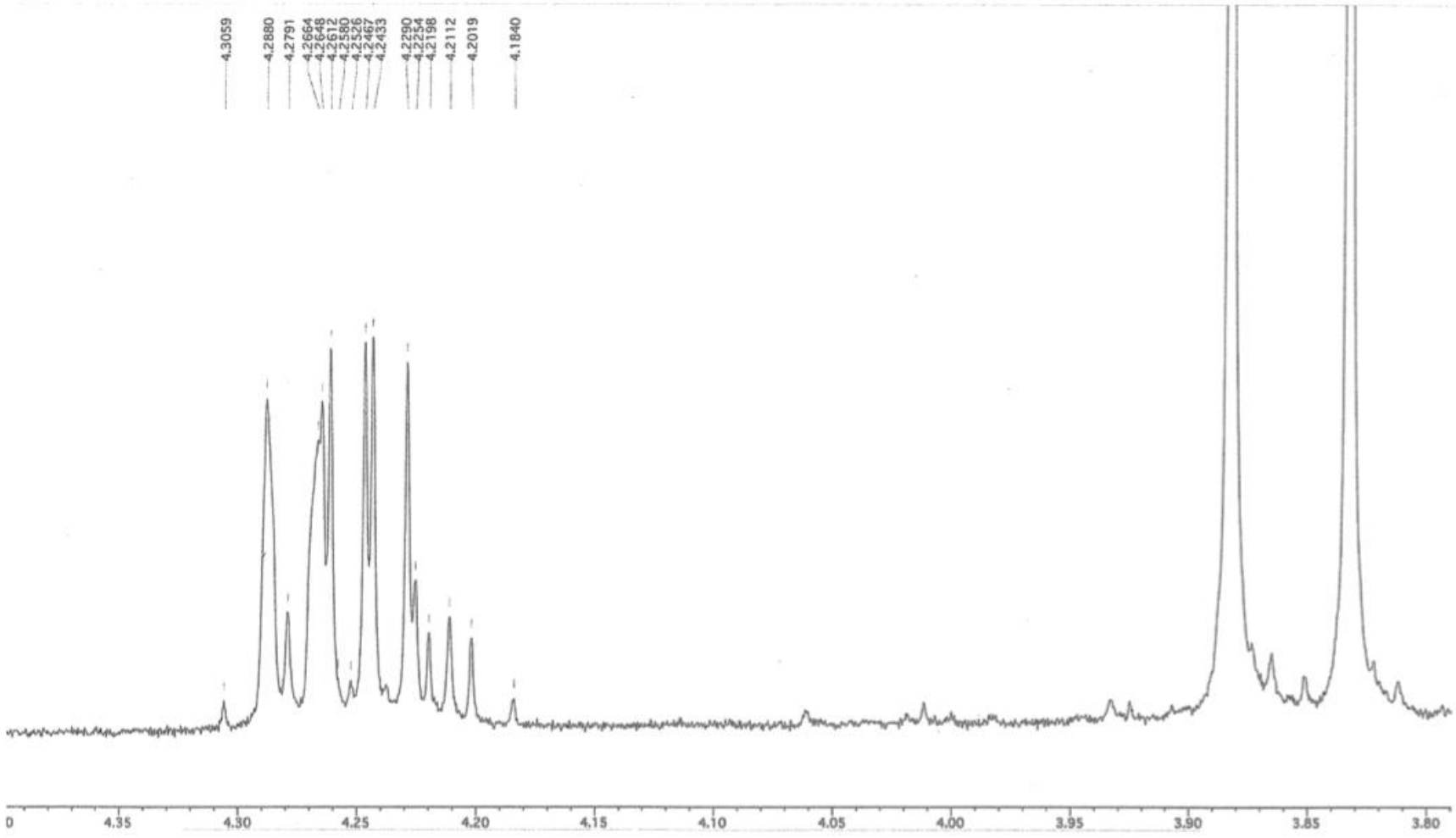
S19-II: Accurate mass of compound **11.**



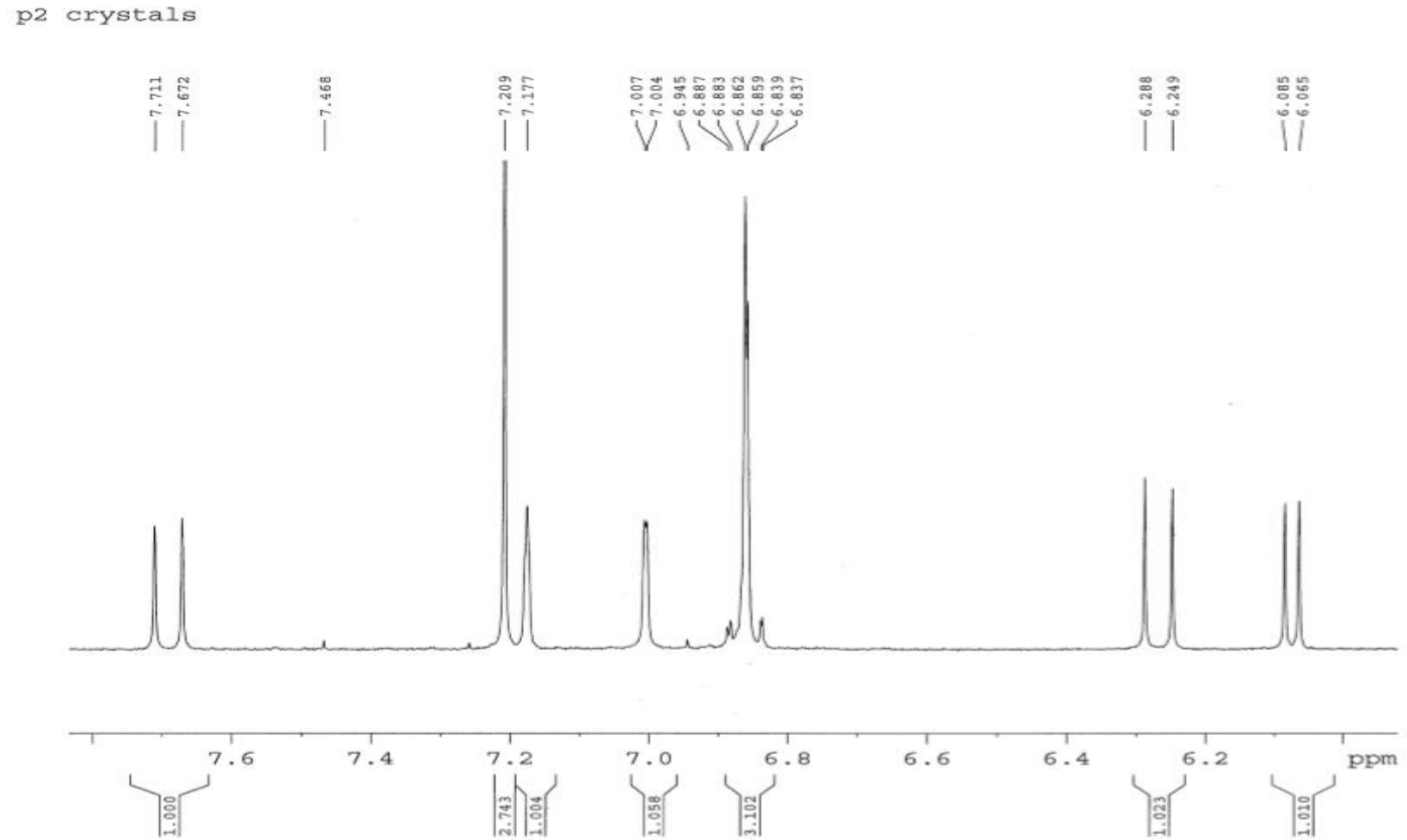
S19-III: 1H NMR spectrum of compound **11.**



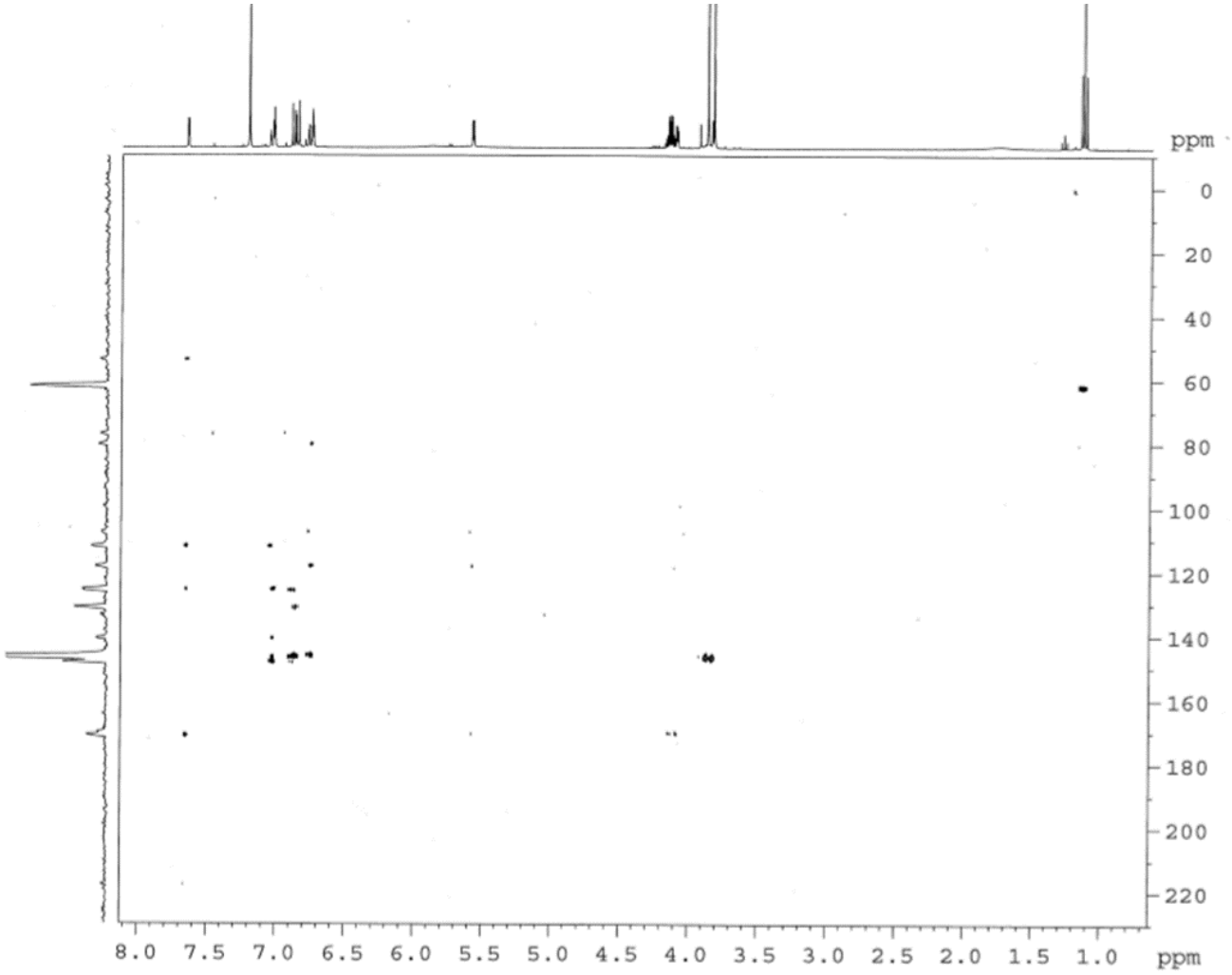
S19-IV: $_1$H NMR expansion A spectrum for compound **11.**



S19-V: 1H NMR expansion B spectrum for compound **11.**

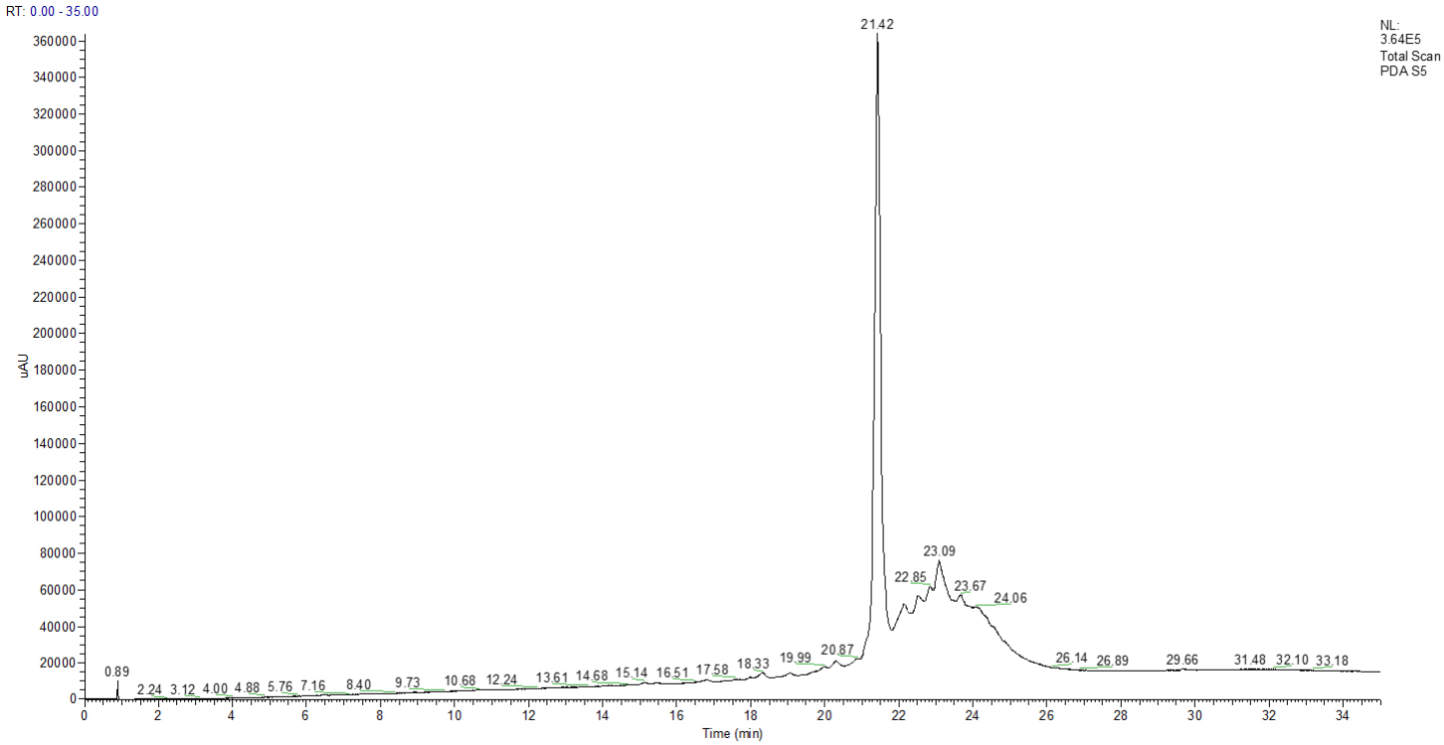


S19-VI: $_1$H NMR expansion C spectrum for compound **11.**

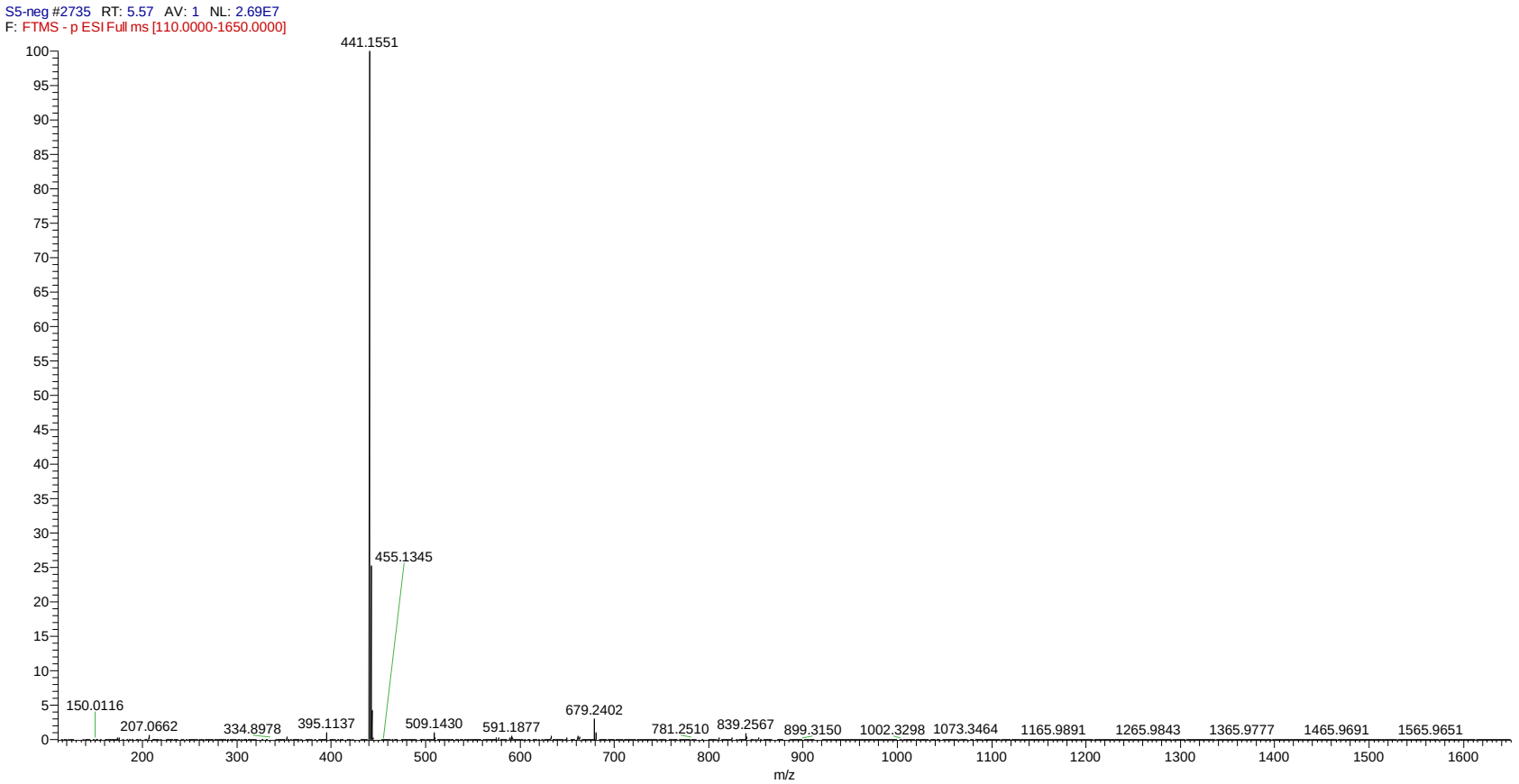


S19-VII: HMBC correlation for compound **11.**

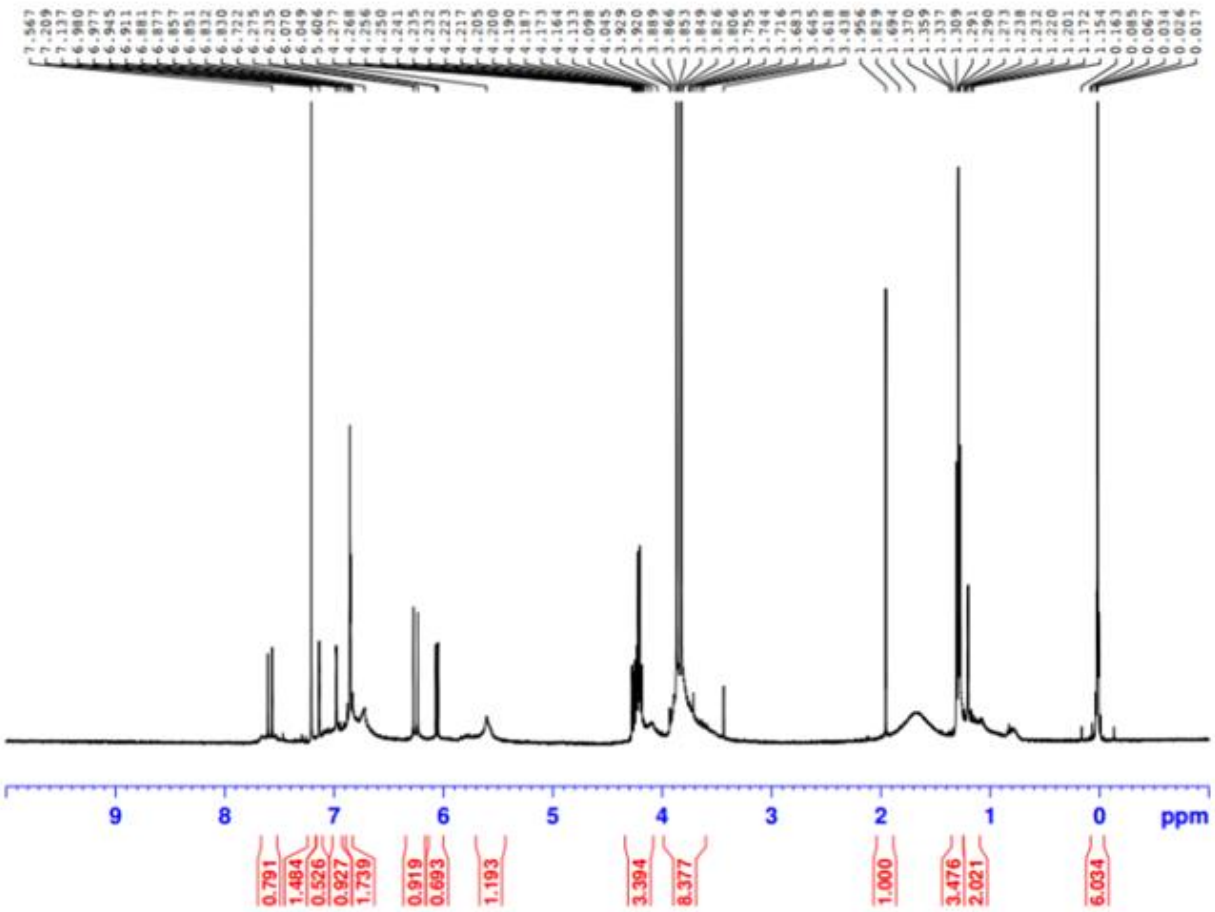
S20. Compound **12**



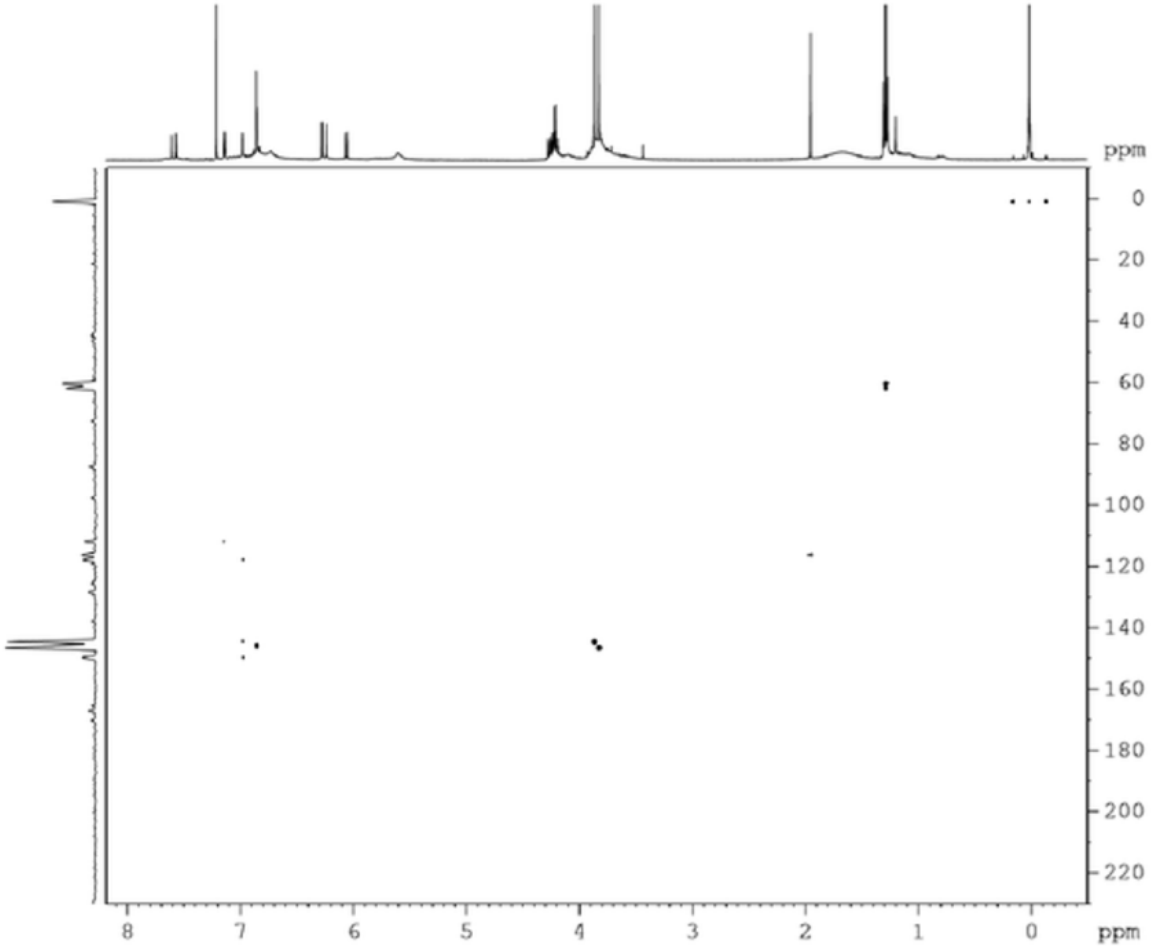
S20-I: HPLC-UV-ESI-MS/MS of compound **12.**



S20-II: Accurate mass of compound **12**.

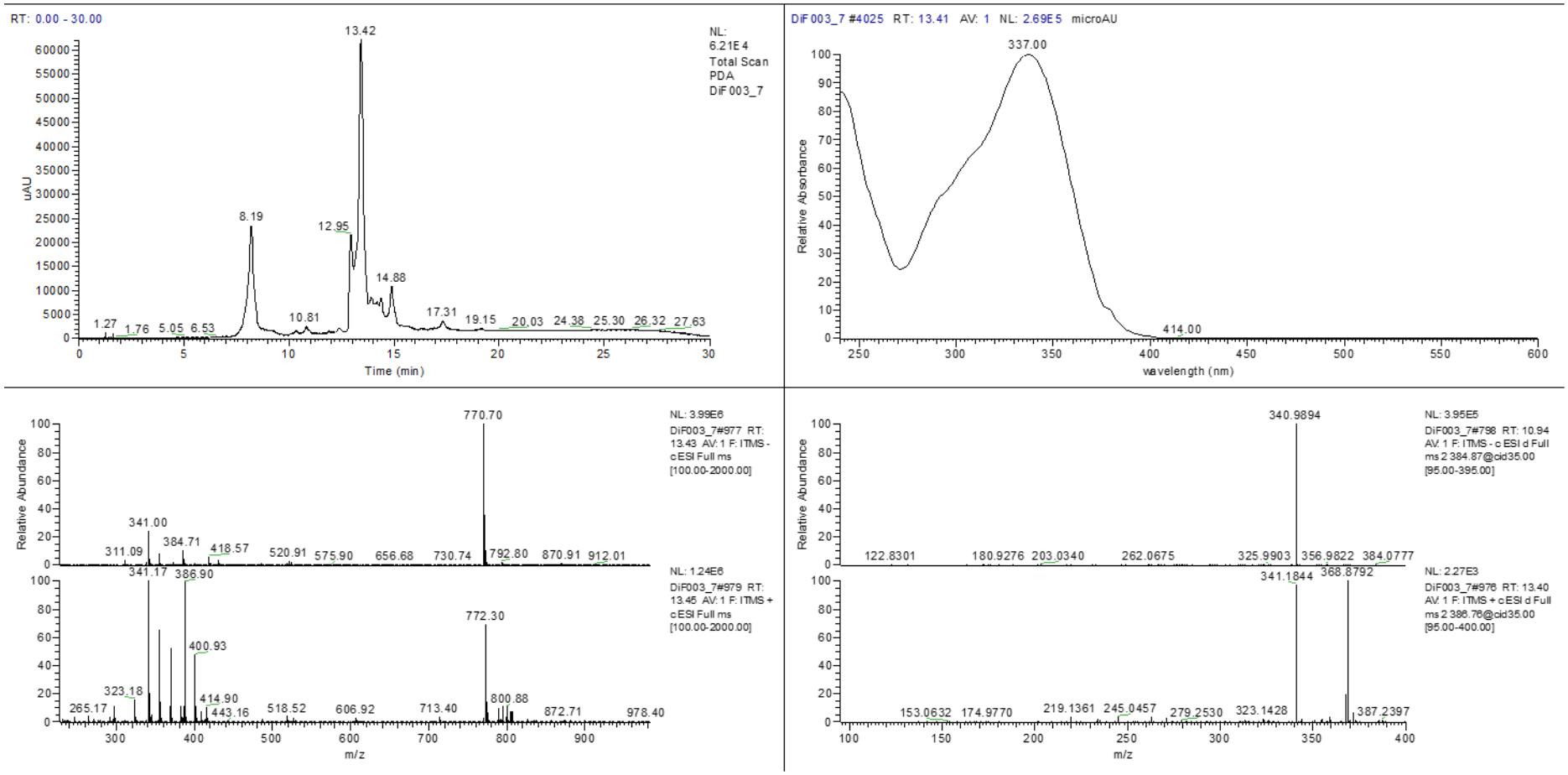


S20-III: 1H NMR spectrum of compound **12.**



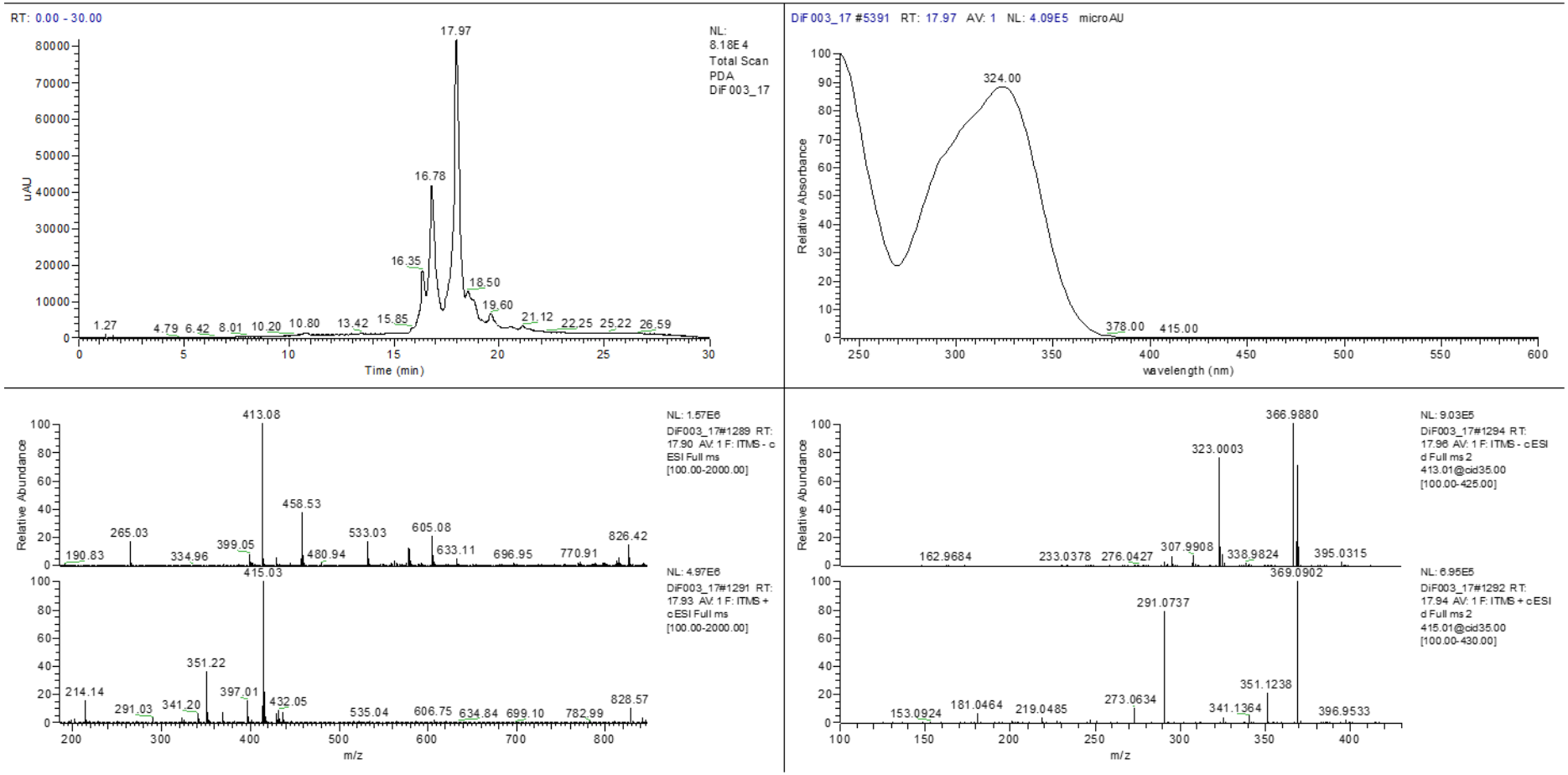
S20-IV: HMBC correlations for compound **12**.

S21. Compound **13**



S21. HPLC-UV-ESI-MS/MS of compound **13.**

S22. Compound 14



S22-I: HPLC-UV-ESI-MS/MS of compound **14.**