



UNIVERSITÀ
DEGLI STUDI
DI BRESCIA

DOTTORATO DI RICERCA IN
PRECISION MEDICINE

settore scientifico disciplinare CHIM/08 CHIMICA FARMACEUTICA

CICLO XXXVI

DESIGN, SYNTHESIS AND OPTIMIZATION OF BIOACTIVE COMPOUNDS:
A MEDICINAL CHEMISTRY APPROACH

DOTTORANDA
ERIKA OSELLADORE

RELATORE
PROF.SSA ALESSANDRA GIANONCELLI

Table of content

Abstract (English)	V
Abstract (Italian)	VII
1 General Introduction	1
1.1 Introduction	1
1.2 Mass spectrometry.....	4
1.2.1 Introduction to mass spectrometry	4
1.2.2 Ion source	5
1.2.3 Mass analysers.....	9
1.2.4 Application of mass spectrometry	14
1.3 Nuclear magnetic resonance (NMR)	18
1.3.1 Introduction to NMR.....	18
1.3.2 Application of NMR spectroscopy.....	22
1.4 Computer aided drug design (CADD).....	27
1.4.1 Structure-based drug design (SBDD).....	28
1.4.2 Ligand based drug design (LBDD)	34
1.5 Circular dichroism (CD) spectroscopy.....	38
1.5.1 Applications of CD.....	40
2 Main project: Developing new ligands targeting G-quadruplex	43
2.1 Nucleic acids and G-quadruplex	43
2.1.1 Introduction to nucleic acids.....	43
2.1.2 G-quadruplex structure	48
2.1.3 Biological relevance of G4.....	51
2.1.4 Targeting G-quadruplex with small molecules	60
2.2 Methods to study G4-ligand interactions	71
2.2.1 ESI-MS for the study of G-quadruplex:ligand interaction	71
2.2.2 NMR spectroscopy of nucleic acids	76
2.2.3 CD spectroscopy of G-quadruplex	82
2.3 Development of ligands for G-quadruplex	84

2.3.1	Aim of the work	84
2.4	Flavonoids	85
2.4.1	Introduction	85
2.4.2	Results and discussion	88
2.5	Antraquinones	96
2.5.1	Introduction	96
2.5.2	Results and discussion	98
2.6	Anthracenes	110
2.6.1	Introduction	110
2.6.2	Results and discussion	111
2.7	Dimeric PEG-linked scaffolds	121
2.7.1	Introduction	121
2.7.2	Results and discussion	122
2.8	General conclusions and future perspectives	124
2.9	Materials and methods	127
2.9.1	Chemistry	127
2.9.2	ESI-MS binding studies	137
2.9.3	NMR titrations	138
2.9.4	2D NMR experiments	139
2.9.5	Circular dichroism (CD)	139
2.9.6	Molecular docking for rutin and quercetin ⁴⁰⁹	139
2.9.7	Molecular docking for <i>M. Pomifera</i> derived flavonoids, anthraquinones and anthracenes ^{437,463}	140
2.9.8	Molecular dynamics	141
3	Side projects	143
3.1	Selenium-containing compounds for the protection against oxidative stress	143
3.1.1	Fluoxetine derivatives	143

3.1.2	Selenoxide elimination triggers enamine hydrolysis to primary and secondary amines: a combined experimental and theoretical investigation	154
3.1.3	Conclusions.....	159
3.1.4	Materials and methods	160
3.2	Phosphodiesterase inhibitors.....	173
3.2.1	Introduction.....	173
3.2.2	Combinatorial library generation, molecular docking and molecular dynamics simulations for enhancing the isoflavone scaffold in phosphodiesterase inhibition	176
3.2.3	Virtual screening-accelerated discovery of a phosphodiesterase 9 inhibitor with neuroprotective effects in the kainate toxicity <i>in vitro</i> model	182
3.2.4	Conclusions.....	187
3.2.5	Materials and methods	189
4	Final conclusions and future perspectives	195
5	Appendix.....	197
5.1	List of abbreviations	197
5.2	Flavonoids	203
5.3	Anthraquinones	212
5.4	Anthracenes.....	225
5.5	Dimeric PEG-linked scaffolds	228
5.6	Seleno-containing compounds	245
6	References.....	294
7	Publications	317

Abstract (English)

The extensive process of drug discovery encompasses multiple stages and requires a convergence of diverse scientific disciplines and methodologies. Medicinal chemistry, which includes sciences like analytical, organic, and computational chemistry, plays a significant role in the preliminary stage of drug discovery and is pivotal in developing potentially bioactive molecules and unravelling disease mechanisms or macromolecular target structures. My PhD has been characterized and essential in enhancing the understanding and development of the use of various techniques, including analytical chemistry, chromatography, spectroscopy, X-ray diffraction, microscopy, among others. Furthermore, I have dedicated myself to optimizing organic synthesis to achieve better yields through greener and more cost-effective reactions. Throughout these years, I have focused with greater emphasis on the following techniques: mass spectrometry (MS), nuclear magnetic resonance (NMR), computational techniques, and circular dichroism (CD). My thesis, after an introduction to the theory and possible applications of these techniques, illustrates how they have been synergistically employed in various research paths.

In the primary project (Section 2), their collective use in studying small molecules' interactions with G-quadruplex (G4) highlighted their potential in anti-cancer, anti-inflammatory, anti-viral, and neuroprotective applications. Essential methodologies include ESI-MS for rapid screening and understanding interaction efficiency, NMR for 3D structure resolution and binding insights, CD spectroscopy to validate findings and assess G4 topology, and computational tools for predicting interactions and refining structures. Different scaffolds were investigated, revealing their potential in targeting DNA arrangements, particularly G4 structures, suggesting their anti-proliferative effects by inhibiting telomerase activity through G4 stabilization. Moreover, an ongoing investigation focuses on understanding the structure of a G4:ligand complex, employing NMR, molecular dynamics, and CD titrations with the aim of unravelling the interactions and the stability due to ligand addition.

Two ancillary projects (Section 3) showcased, respectively, one the development of seleno-containing compounds for neurodegeneration mitigation and the other the identification of inhibitors of phosphodiesterase (PDE) for neurodegenerative disorders;

these projects demonstrated the efficiency of integrating computational techniques with experimental assays to streamline drug discovery.

The first project presented focused on the preparation of selenofluoxetine derivatives, emphasizing their capacity to mitigate oxidative stress and offer neuroprotective effects. The investigation delved into understanding the compounds' reactions with reactive oxygen species (ROS), both in experimental and computational settings, providing crucial insights into their mechanisms and, a further development of this study, involved designing derivatives with shorter chains to trigger enamine hydrolysis, potentially creating prodrugs activated under specific conditions for drug delivery and an alternative pathway for primary and secondary amine synthesis.

The second project showed the use of computational tools for designing novel PDE inhibitors through molecular docking, and MD simulations. After a meticulous validation of the docking method, the compounds generated through combinatorial chemistry and the molecules of an internal database underwent a screening, leading to the identification of potential hits displaying significant interactions with crucial amino acids for PDE binding. These findings allowed a preliminary structure-activity relationship (SAR) study and identified crucial design features for potential PDE inhibitors. Eventually, MD simulations further validated the stability of the formed complexes, consolidating the findings and *in vitro* testing demonstrated high inhibitory activity against PDE9, comparable to a known inhibitor.

The synthesis and optimization of potential drug candidates, when coupled with a comprehensive understanding of target:ligand interactions through diverse techniques, form the crux of successful drug discovery. The seamless integration of these methodologies continues to be paramount in creating novel pharmacological agents, driving the need for further research to refine and expedite the drug discovery process.

Abstract (Italian)

Il lungo processo di scoperta di un nuovo farmaco coinvolge diverse fasi e richiede l'integrazione di diverse discipline scientifiche e metodologie. La chimica farmaceutica, che comprende settori come la chimica analitica, organica e computazionale, svolge un ruolo cruciale nella fase iniziale del processo di drug discovery ed è essenziale nello sviluppo di molecole con potenziale attività biologica nonché nella comprensione dei meccanismi delle malattie o delle strutture dei bersagli macromolecolari. Il mio percorso di dottorato è stato fondamentale per accrescere la comprensione e sviluppare l'utilizzo di diverse tecniche, tra cui: chimica analitica, cromatografia, spettroscopia, diffrazione a raggi X ed altre ancora; mi sono inoltre dedicata all'ottimizzazione della sintesi organica con lo scopo di ottenere migliori rese e se possibile di utilizzare reattivi meno pericolosi e più economici. In questi anni ho trattato con maggiore enfasi le seguenti tecniche: spettrometria di massa (MS), risonanza magnetica nucleare (NMR), tecniche computazionali e dicroismo circolare (CD); la mia tesi, dopo un'introduzione alla teoria e alle possibili applicazioni delle suddette, illustra come siano state utilizzate sinergicamente in vari ambiti di ricerca.

Nel progetto principale (Sezione 2), l'uso combinato di queste tecniche nello studio delle interazioni di piccole molecole con il G-quadruplex (G4) ha evidenziato il loro potenziale nelle applicazioni antitumorali, antinfiammatorie, antivirali e neuroprotettive. Le metodologie fondamentali includono l'ESI-MS per uno screening rapido e per una valutazione preliminare dell'efficienza dell'interazione, l'NMR per la risoluzione tridimensionale della struttura e per approfondire la natura del legame ligando:target, il CD per convalidare i risultati e valutare la topologia del G4 e le tecniche computazionali per prevedere le interazioni presenti e per perfezionare le strutture predette.

Sono stati esaminati scaffold diversi dimostrando il loro potenziale nell'interagire con vari folding degli acidi nucleici e, in particolare con le strutture G4, suggerendo possibili effetti antiproliferativi grazie all'inibizione dell'attività telomerasica mediante la stabilizzazione del G4. Attualmente, uno studio in corso mira a comprendere la struttura tridimensionale di un complesso G4:ligando, con l'utilizzo di NMR, MS e titolazioni CD, con l'obiettivo di determinare le interazioni presenti e la stabilità del complesso derivante dall'aggiunta del ligando.

I due progetti secondari presentati nella Sezione 3 mostrano, rispettivamente, uno lo sviluppo di composti contenenti selenio per ridurre la neurodegenerazione e l'altro l'identificazione di inibitori di fosfodiesterasi (PDE) per il trattamento delle malattie neurodegenerative; questi progetti hanno dimostrato l'efficacia dell'integrazione delle tecniche computazionali con saggi sperimentali nella scoperta di nuovi farmaci. Il primo progetto illustra la preparazione di derivati della selenofluoxetina, enfatizzandone la capacità di mitigare lo stress ossidativo e offrire benefici neuroprotettivi. L'indagine ha coinvolto lo studio delle reazioni di questi composti con le specie reattive dell'ossigeno (ROS), sia con metodi sperimentali che computazionali, fornendo informazioni cruciali sui loro meccanismi e, un successivo approfondimento dello studio, ha coinvolto la sintesi di derivati con catene più brevi per indurre l'idrolisi dell'enamina prodotta, creando possibili pro-farmaci attivati in condizioni specifiche e un metodo alternativo per la sintesi di ammine primarie e secondarie.

Il secondo progetto coinvolge l'uso di tecniche computazionali per la progettazione di nuovi inibitori di PDE utilizzando il docking molecolare e le simulazioni di dinamica molecolare (MD). Dopo una rigorosa validazione del metodo di docking, i composti generati tramite chimica combinatoria e quelli presenti in un database interno sono stati sottoposti ad uno screening, individuando potenziali candidati farmaci in grado di formare le interazioni significative con gli amminoacidi cruciali per il legame con le PDE. Questa individuazione ha permesso uno studio preliminare sulle relazioni struttura-attività (SAR) ed identificato le caratteristiche fondamentali per lo sviluppo di potenziali inibitori di PDE. Infine le simulazioni di MD hanno ulteriormente convalidato l'effettiva stabilità dei complessi formati consolidando i risultati, e i test *in vitro* hanno dimostrato un'elevata attività inibitoria contro la PDE9, paragonabile a quella di un noto inibitore.

La sintesi e l'ottimizzazione di potenziali candidati farmaci, insieme ad una comprensione approfondita delle interazioni bersaglio:ligando mediante l'uso di diverse tecniche strumentali, costituiscono la chiave del successo nella scoperta di nuovi farmaci. La combinazione di diverse metodologie continua ad essere fondamentale nella creazione di nuovi agenti farmacologici, sottolineando l'importanza di ulteriori ricerche per perfezionare ed accelerare il processo di drug discovery.

1 General Introduction

1.1 Introduction

Drug discovery is the process whereby compounds with activity against a specific target or function are identified, evaluated and optimized for clinical application. It is a complex and lengthy process that can take up to 15 years and cost over \$ 2 billion and it is divided in three main stages: early drug discovery which include the identification and validation of a specific target, the discovery and optimization of potential lead compounds and preclinical *in vitro* and *in vivo* testing, clinical testing that is divided in the well-known three clinical phases and approval for production, which includes the Food & Drug Administration (FDA) and European Medicines Agency (EMA) review and approval plus the post market monitoring.¹

Medicinal chemistry is a discipline that combines analytical, organic and computational chemistry, biochemistry, chemical, molecular, cell and structural biology, microbiology, pharmacology, toxicology and translational medicine and finds its application in the first early drug discovery stage.²

It is self-explanatory that, given the number of subjects involved, a medicinal chemist uses a combination of different techniques in order to develop new potential bioactive molecules or to investigate the mechanism of action of a disease or to resolve the three-dimensional structure of a macromolecular target.

Therefore, during my PhD, with the aim of developing potentially bioactive compounds, I have been focusing not only on the organic synthesis and optimization of the reactions involved to give better yield and to use cheaper and safer chemicals, but also to gain knowledge and insight of the possible application of several techniques.

The techniques available in this context comprehend a variety of methods, such as analytical chemistry, chromatography, spectroscopy, X-Ray diffraction, microscopy and more. Furthermore, as medicinal chemistry is a rapidly evolving field, new techniques are constantly being developed and as our understanding of biology and chemistry continues to grow, we can expect to see even more innovative and effective techniques being used to develop new drugs.³

The description of all the available techniques is beyond the scope of this work, therefore in the next chapters, the attention will be focused on the methods that have

been used in these three years, namely mass-spectrometry (MS), nuclear magnetic resonance (NMR), computational techniques and circular dichroism (CD).

In this first section, a detailed explanation of the principle and mechanism of the techniques with some insight on their application in the context of drug development will be provided.

A second section will present how the combination of these techniques have been used in the context of my main research project which consisted on the identification of small molecules with the aim of targeting a specific non-canonical arrangement of nucleic acids, namely G-quadruplex (G4).

Eventually, the third section will be focused on the report of two side projects in the context of neurodegenerative disorders that involved selenium-containing molecules against oxidative stress and the development of inhibitors targeting phosphodiesterase enzymes.

In particular, MS (the theory can be found in Chapter 1.2) have been used for the characterization of the synthesised compounds but also for checking their behaviour in a simulated specific physiological condition (i.e. oxidative stress) (see Chapter 3.1 for an example of this application) and for lead identification and optimization thanks to its use in binding assays performed mixing the target and binders (see Chapter 2.2.1 for more detailed information and Chapters 2.4 to 2.7 for this application).

NMR (the theory can be found in Chapter 1.3) has also been used in the context of characterization of new compounds, moreover thanks to its ability of giving structural information, the signals of the different parts of the molecules are separated, therefore it is possible to see the changes in the structure at atomic resolution both in the context of binding assay and when following a reaction (see Chapters 2.4 to 2.6 and Chapter 3.1 respectively for these applications). Moreover, NMR can be used for target identification and validation thanks to the possibility of relating given signals in its experiments with distances in space and therefore the possibility of resolving the three-dimensional structure of protein, nucleic acids and their complexes with small molecules (see Chapter 2.2.2 for a more detailed description and Chapter 2.6.2 for its application).

Computational techniques (the theory can be found in Chapter 1.4) are particularly useful because they can be applied in all the stages of early drug discovery, from studying a specific target to potential hits identification for new compounds or in the context of drug repurposing, to computing pharmacokinetic properties in the context of ADMET (absorption, distribution, metabolism, excretion and toxicity) profiles (these methods

application can be seen in Chapters 2.4, 2.5, 2.6, 3.1 and 3.2). This is particularly important because it allows to save the time of synthesising compounds that wouldn't be able to bind to the specific target because of the lack of specific chemical properties or that couldn't get to target because of undesirable physico-chemical features, which is approximately the 50% of the causes of failure in drug discovery.¹

CD (the theory can be found in Chapter 1.5) has been used to check the ability of the compounds to interact with the target and, if that happens, if the overall three-dimensional arrangement of the macromolecule is affected by the binding (see Chapters 2.5 and 2.6 for its application).

As can be understood from the different use of the various methods together with the synthesis of new potential drug candidates and its optimization, the combination of the greatest possible number of techniques is fundamental in order to gain a full understanding of the target:ligand binding which is not dependent on the experimental conditions used for the specific assay (buffer, solvent or its absence, temperature, pH) but is really due to the interactions that are involved in the binding.

Moreover, especially when a technique is used for the first time in a new application, it needs to be supported by more canonical or known assay to effectively prove that the information received is indeed significant and not fortuitous.

The work presented in this thesis is intended as an example on a possible combination of the information gained from the various assays applied to three different research lines.

1.2 Mass spectrometry

1.2.1 Introduction to mass spectrometry

Mass spectrometers are instruments that can be used to selectively detect and determine the amount of a given analyte, together with its elemental composition and some aspect of its structure. The property that gets analysed is the mass of gas-phase ions produced from the molecules of an analyte as a direct determination of the nominal mass of the analyte, but also as detection of fragments of the molecule that correspond to groups of atoms that can reveal structural features. It is important to keep in mind that it is not the atomic mass of the elements as the weighted average of the naturally occurring stable isotopes that comprise the element that gets detected, but the mass of the isotopes of the elements. Moreover, mass spectrometry doesn't directly determine the mass but the mass-to-charge ratio (m/z) of the ions, which is the mass of the ion on the atomic scale divided by the number of charges that the ion possesses. As a consequence, for the analysis to happen, it is necessary that the molecules get ionized. Ions are charged particles; hence, their position in space can be influenced with the use of electric and magnetic fields. In order to get rid of all possible contaminants, to group the ions according to their mass and number of charges, and to move them from one point to another, it is necessary to analyse them in a vacuum and, therefore, in the gas phase. Therefore, only ions can get analysed by the mass spectrometer; molecules and radicals are removed from the instrument by the continuous pumping that maintains the vacuum.⁴



Figure 1.2.1. Schematic representation of the mass spectrometer components

Mass spectrometers are composed of three main components: an ion source that produces the ions, a mass analyser that resolves ions according to their m/z , and a detector that records the signal that is produced by the impact of the ions (Figure 1.2.1).⁵

There are different techniques for producing gas-phase ions, which include electron impact (EI) and chemical ionization (CI) that produce gas-phase ions if the analytes have significant vapor pressure, or desorption/ionization (D/I) techniques, such as fast atom bombardment (FAB), electrospray ionization (ESI), and matrix-assisted laser desorption/ionization (MALDI). After ionization, ions are separated by specific types of m/z analysers: magnetic sectors, transmission quadrupoles, quadrupole ion traps (both linear and three-dimensional), time-of-flight (TOF) analysers (both linear and reflectrons);

ion cyclotron resonance mass spectrometers (magnetic ion traps) that convert thanks to Fourier transformation oscillating image currents to record the mass spectrum (FTICR) and orbitraps, which store ions using electrostatic fields and detect them through Fourier transformation of oscillating image currents.⁴

The choice of the ion source and the mass analyser depends on different factors and experimental considerations, such as the m/z range that needs to be analysed, the mass of the analyte, the required resolving power, the ability of the analyser to interface with the ion source, and the limit of detection required.⁶

1.2.2 Ion source

Electron Ionization (EI)

EI is a technique in which the sample gets thermally desorbed to the gas phase and then bombarded with free electrons emitted from a filament. This causes high ionization that fragments the molecule and turns them into positively charged ions. EI is particularly appropriate for small organic molecules (M.W. < 600) that are volatile and thermally stable.⁴ A schematic representation of EI is depicted in Figure 1.2.2a.

Chemical Ionization (CI)

CI technique is similar to EI but it uses a reagent gas previously subjected to electron impact to give reagent gas ions which further undergo ion-molecule reactions with neutral reagent molecules yielding selective ions. After the addition of the sample, the molecule of the analyte undergoes ion-molecule reactions with the ions produced from the reagent to produce sample ions. This technique is particularly useful when the molecular ion is not observed in the EI spectrum as it is considered a soft ionization technique since the energy that is transferred by the reaction is lower than the energy given by the electron impact.⁴ The scheme of a CI ion source is depicted in Figure 1.2.2b.

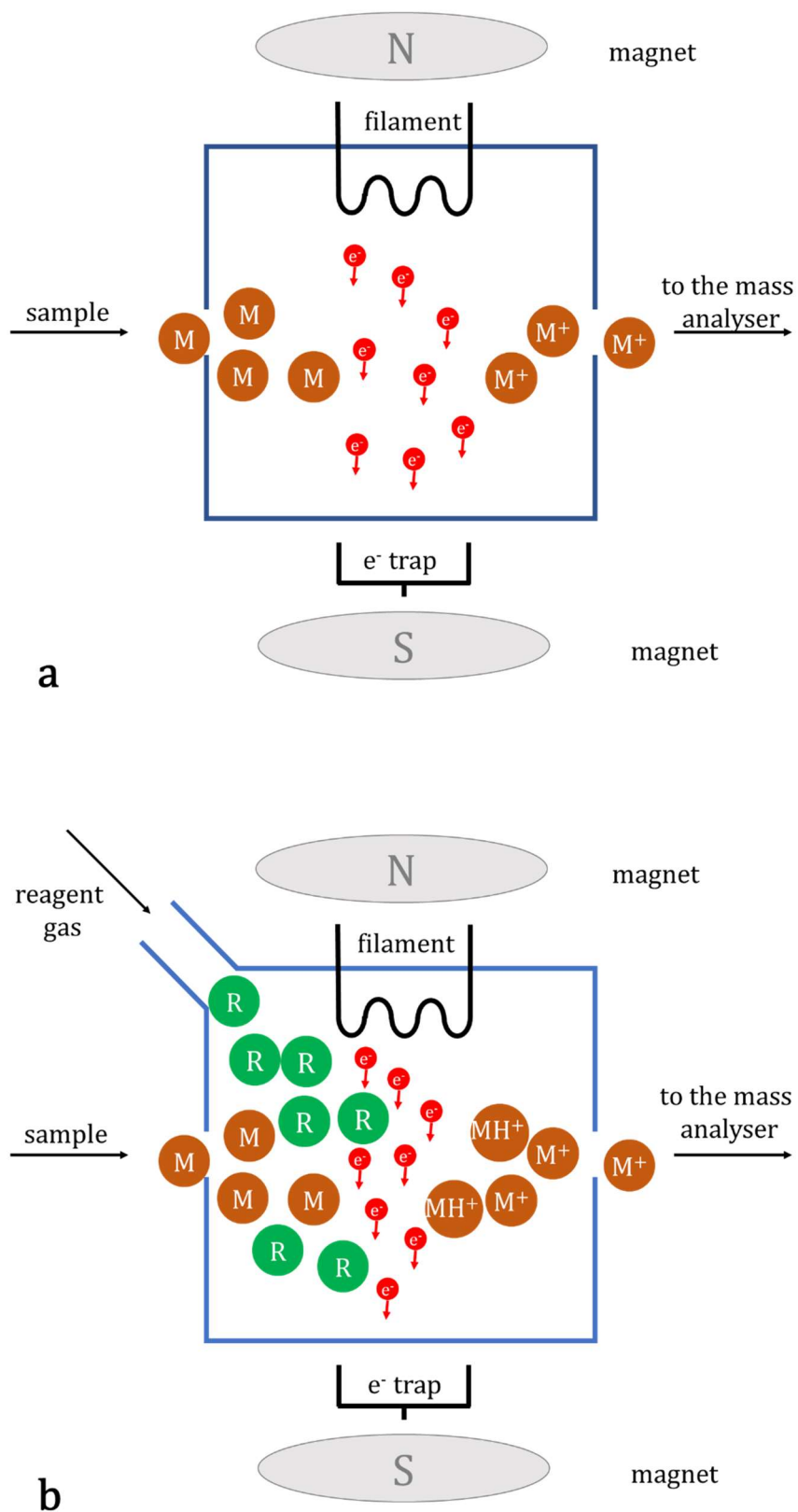


Figure 1.2.2. Scheme of (a) electron ionization (EI) and (b) chemical ionization (CI) ion sources

Electrospray Ionization (ESI)

ESI is considered a soft ionization technique which employs a high voltage applied to a capillary at atmospheric pressure, producing a spray/mist of charged droplets of the same polarity. Afterward, heat or dry gas is applied to the mist, which causes the evaporation of the solvent; therefore, it is often called desolvation gas. This evaporation causes the sequential decrease in the droplet size and the increase in the surface charge density until the coulombic repulsions of the charges are so high that the droplets “explode” (droplet fission), creating smaller, more stable droplets which substantially undergo further solvent evaporation and fission into even smaller droplets. This eventually leads to desolvated single ion formation in the gas phase. ESI is also considered a soft ionization technique and is particularly useful for large biological molecules that are not volatile.⁶ A schematic representation of ESI ion source is depicted in Figure 1.2.3a.

Matrix-assisted laser desorption/ionization (MALDI)

As the name suggests, MALDI is a technique in which a “matrix” with specific features is added in excess to the sample to induce ionization. The matrix has to be able to co-crystallize with the analyte to form a solid solution, to transfer or accept protons from the analyte, and must be chemically inert and stable under vacuum and soluble in solvents. After the co-crystallization of the sample on the plate, it gets irradiated with laser radiation, resulting in the desorption of the matrix from the plate, and the sample is therefore carried to the gas phase. The matrix prevents the analyte molecules from getting fragmented or decomposed, but the laser energy causes the ionization of the sample molecule together with its vaporization. MALDI is particularly useful with high molecular weight compounds that may be difficult to analyse through other ionization methods.⁷ A Figure 1.2.3b shows a scheme of a MALDI ion source.

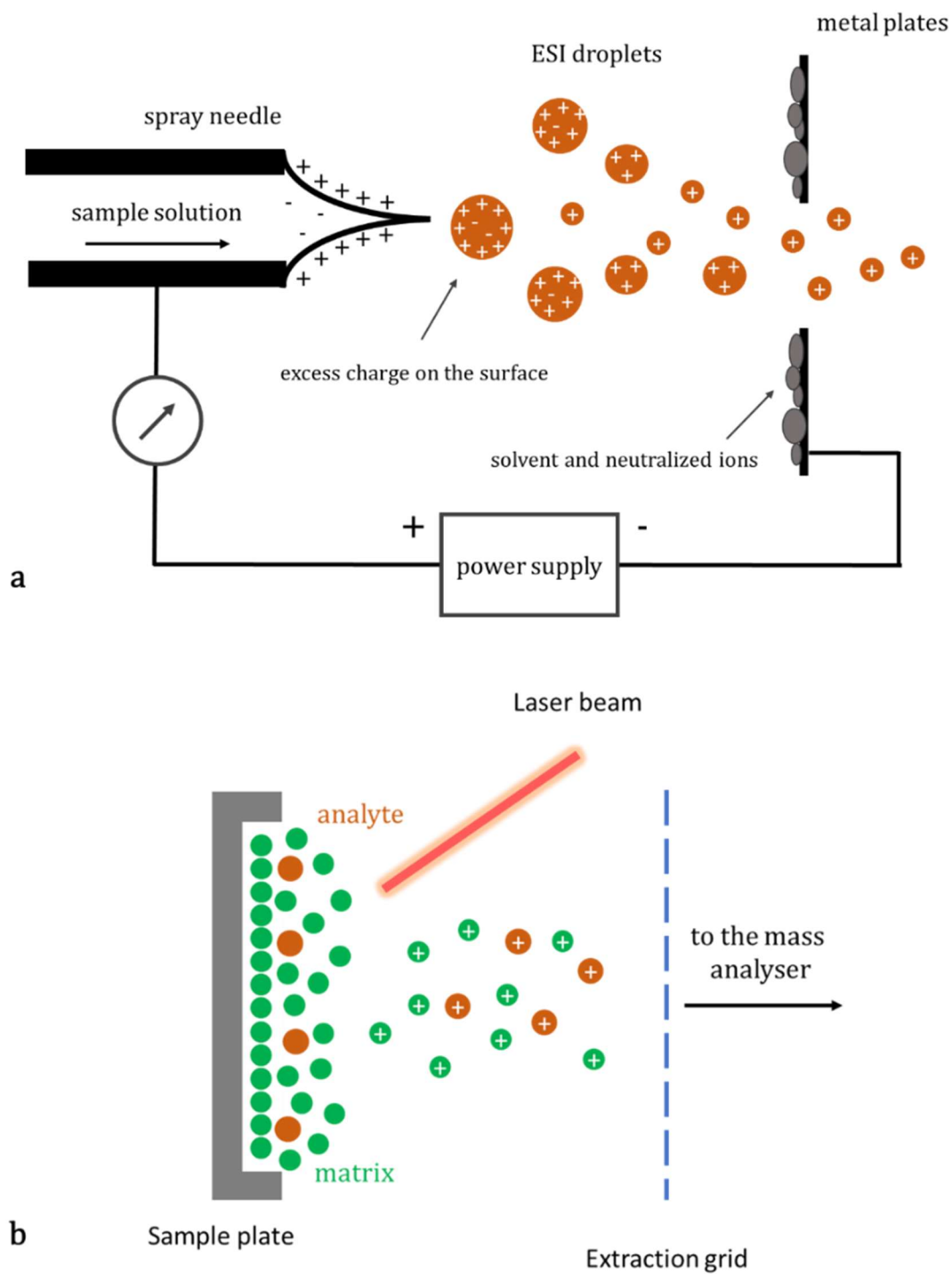


Figure 1.2.3. Schematic representations of (a) electrospray ionization (ESI) and (b) matrix-assisted laser desorption ionization (MALDI)

1.2.3 Mass analysers

Magnetic sector mass spectroscopy

Ions are separated thanks to a magnetic field that deflects them according to the mass, or more specifically to the m/z of the ion; the greater the mass of the ion, the smaller the deflection it is subjected to. Ions are accelerated out of the source region by the application of an accelerating voltage V , acquiring thereby kinetic energy. When the ion enters the magnetic field, it experiences a deflecting force at right angles to its direction of travel, which forces it to describe a circular orbit. The deflection the ions are subjected to depends on their momentum, so on the product of the mass and the velocity, so it is important that the ions enter the magnetic field with the same kinetic energy. By varying (scanning) the magnetic field strength, ions of different mass are sequentially brought into focus at this point.⁷ A scheme of the analyser is depicted in Figure 1.2.4a.

Time-of-flight

The principles of the TOF mass spectrometer involve measuring the time required for an ion to travel from an ion source to a detector (usually located 1 or 2 meters from the source). The ions are accelerated through a potential (V) and then drift down a field-free tube toward the detector. All the ions receive the same kinetic energy during the instantaneous acceleration, but, as they differ in m/z values, they have different velocities. The ions have to be produced at an accurately known start time and preferably originated from the same spatial position; therefore, TOF analysers are often conjugated with pulsed ionization techniques, such as plasma desorption and laser desorption. While the ions traverse the “field-free” region between the ion source and the detector, they separate into groups or packets according to their velocity, which is a function of their m/z values; the lower the m/z value, the faster they reach the detector. The main advantage of this technique is that there’s no upper limit for m/z values.^{4,7} A representation of how the TOF works is depicted in Figure 1.2.4b.

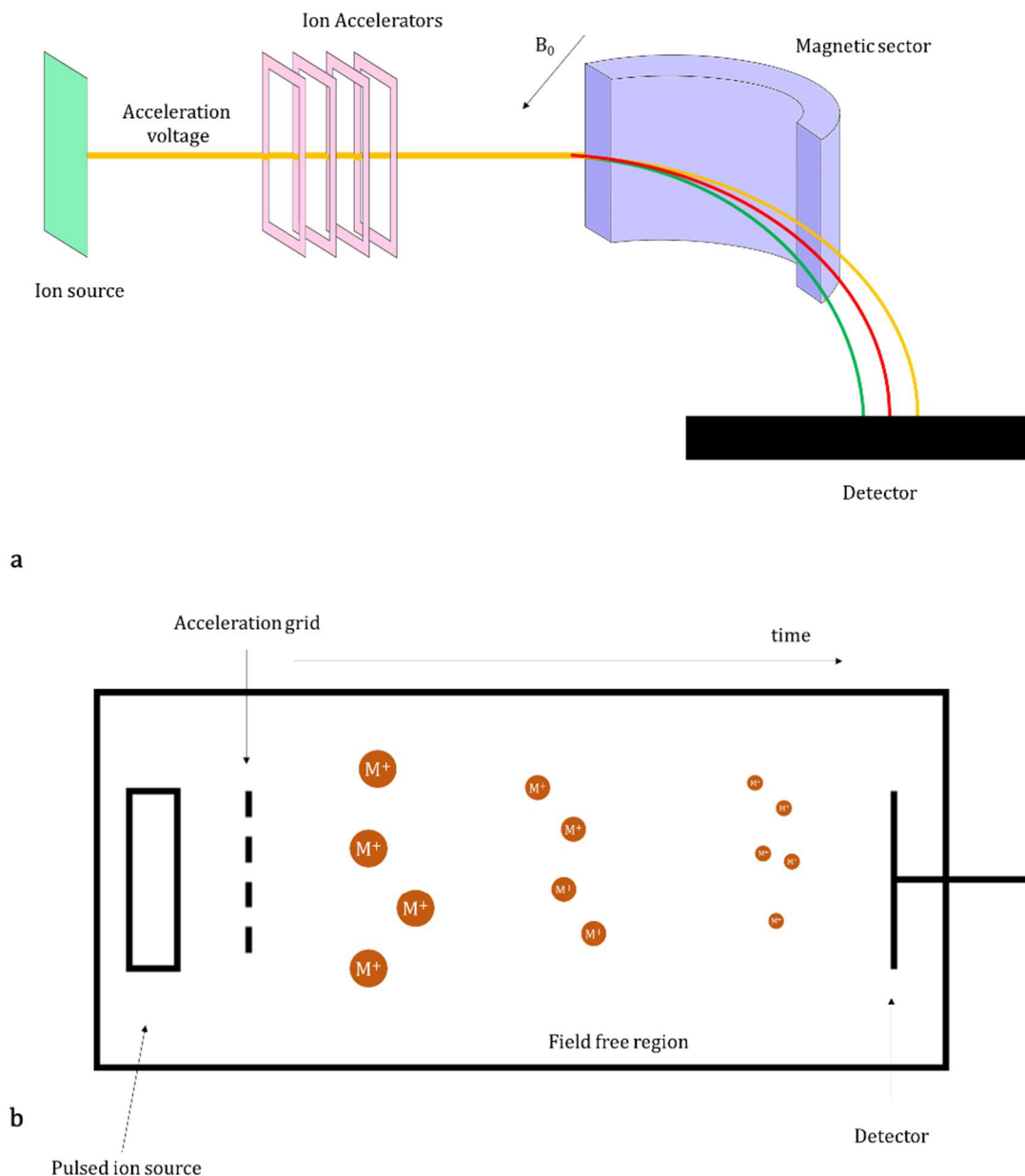


Figure 1.2.4. Schematic representation of (a) Magnetic sector and (b) time-of-flight mass analysers.

Transmission Quadrupole

The term quadrupole is used because the electric field is created between four opposing electrical poles; the shape of the electric field is a function of the geometric arrangement of these four surfaces and the magnitude of the alternating current (AC) and direct current (DC) potentials. The primary component of the quadrupole electric field that surrounds the ions is based on radio frequency (RF) potentials applied to two pairs of opposing electrodes.

The quadrupole consists of four parallel surfaces, ideally with a hyperbolic cross-section, accomplished with four longitudinally parallel round rods. Opposing surfaces are

connected together electrically and to RF and DC power sources. Extracted ions from an ion source are accelerated (5–15 V) into the central space that constitutes the quadrupole electric field along the longitudinal axis toward the detector.⁴

The combination of the RF and DC causes ions to oscillate along the x- and y-axes as they pass through the quadrupole in the z-axis. Based on the DC and the frequency of the RF, only ions with a particular m/z will have stable trajectories and reach the detector; all the others will collide into the rods and be filtered out. Varying the DC and the RF all the different m/z can be scanned.⁸ (Figure 1.2.5a) A quadrupole (or other multipoles) can also operate in a “RF-only” mode, where the DC potential is reduced and only RF is applied to the rods, allowing all ions to pass through the multipole, thus functioning as a transmission guide into another analyser, as collision cells where an inert gas is introduced and the RF is increased for the ions to undergo fragmentation which can give important structural information.⁶ The main advantages of quadrupoles are the low cost, compact shape, and size, and the requirement of little maintenance. For these kinds of analysers to work well, they need a continuous flux of ions, so they work well in coordination with GC and LC and are less suitable for pulsed ion sources, such as MALDI.^{9,10}

Quadrupole ion trap (QIT)

The QIT is a modification of the quadrupole, where two hyperbolic electrode plates facing each other have a hyperbolic ring electrode in between.¹¹ QIT can be either 2D (linear) or 3D. The 3D traps consist of two hyperbolic electrode plates facing each other and a hyperbolic ring electrode placed in between them. Using an oscillating RF field and a superimposed DC electric field, ions are trapped between the electrodes. Varying the RF potential, ions of different m/z are selectively ejected from the trap. (Figure 1.2.5b) In 2D traps, a potential field is applied to each end of the quadrupole to trap the ions within the quadrupole itself. The ejection of the ions can be either axial or radial depending on the design.⁶

QIT operates at a relatively high pressure compared to the other mass spectrometer, $\sim 10^{-1}$ Pa (10^{-4} for transmission quadrupoles, 10^{-7} for TOFs). This higher pressure is necessary for achieving sufficient resolution. The pressure is maintained by a continuous flow of buffer gas (helium or argon) that cools the ions, reducing their rotational and vibrational energies, so the displacement of their amplitude around the z-axis is diminished.⁶

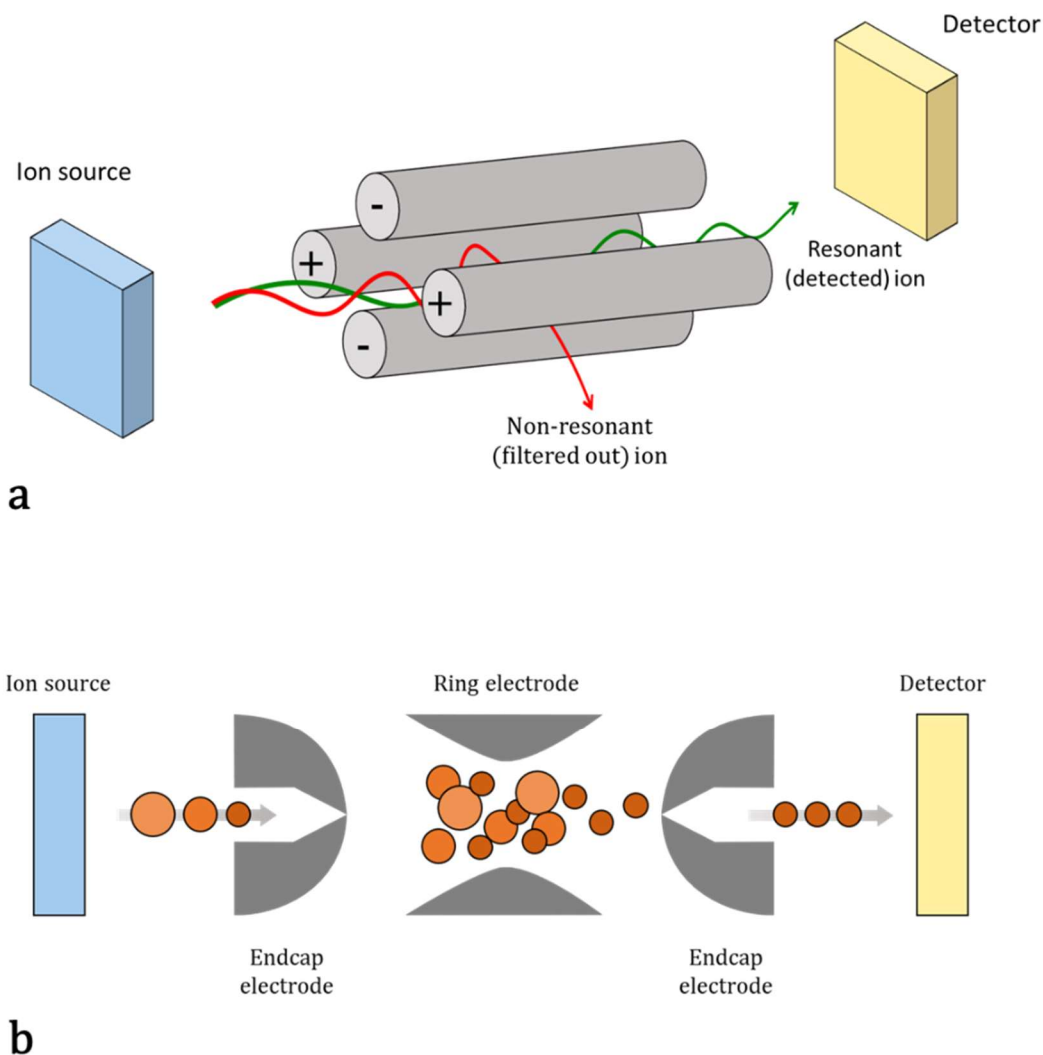


Figure 1.2.5. Scheme of (a) the transmission quadrupole and (b) the 3D quadrupole ion trap (QIT).

Fourier Transform-ion cyclotron resonance (FT-ICR)

FT ICR MS is an example of a high-performance mass spectrometer employing an ion trapping mass analyser;¹² the trap is called Penning trap, and it is similar to a 3D ion trap but it uses a combination of an electrostatic and a magnetic field. After the ions enter the trap thanks to a DC potential applied to the front and back plate and move along the z-axis, they are then subjected to a Lorentz force that makes them move circularly in the (x, y) plane, perpendicularly to the magnetic field (along the z-axis) with an angular frequency that is called cyclotron frequency and that depends on the m/z ratio and the strength of the magnetic field B_0 . The ions are not phased when they enter the trap, and that leads to small orbits that can't be detected; therefore, the ions get excited by exciting them with a limited frequency sweep of a broadband RF field.¹² The excitation synchronizes the ions' motion into higher cyclotron orbits, allowing their detection. The

detector measures the intensity of the signal (which depends on the number of ions that give that signal) over time, which is converted via a Fourier transform into a frequency that is proportional to their m/z . Figure 1.2.6a shows a schematic representation of a FT-ICR.⁶

Orbitraps

The orbitrap is a type of analyser that uses Fourier transform to convert the signal produced by ions oscillating in a trap from the time domain to frequency, mixing the technology of FT-ICR and a quadrupole trap. The oscillations are caused only by an electric field, not involving the magnetic field. It is made of three main parts: an inner spindle electrode covered by two hollow outer concave electrodes facing each other, separated by a thin ring of dielectric material. (Figure 1.2.6b) Ions are introduced tangentially into the orbitrap as a “packet” between the inner and outer electrodes. Due to the electric field between the inner and outer electrodes, the ion packet is bent towards the inner electrode while the tangential velocity of the ions creates an opposing centrifugal force. At a specific potential between the inner and outer electrodes, the ions remain in a spiral path around the inner electrode. However, due to the conical shape of the electrodes, a harmonic axial oscillation in the ions is induced. The outer electrodes also act as receiver plates that detect the back-and-forth axial harmonic motion of the ions. This signal image is digitized and transformed from the time domain to the frequency domain. Similar to FT-ICR, the axial harmonic frequencies are proportional to the m/z of the ions.^{4,6}

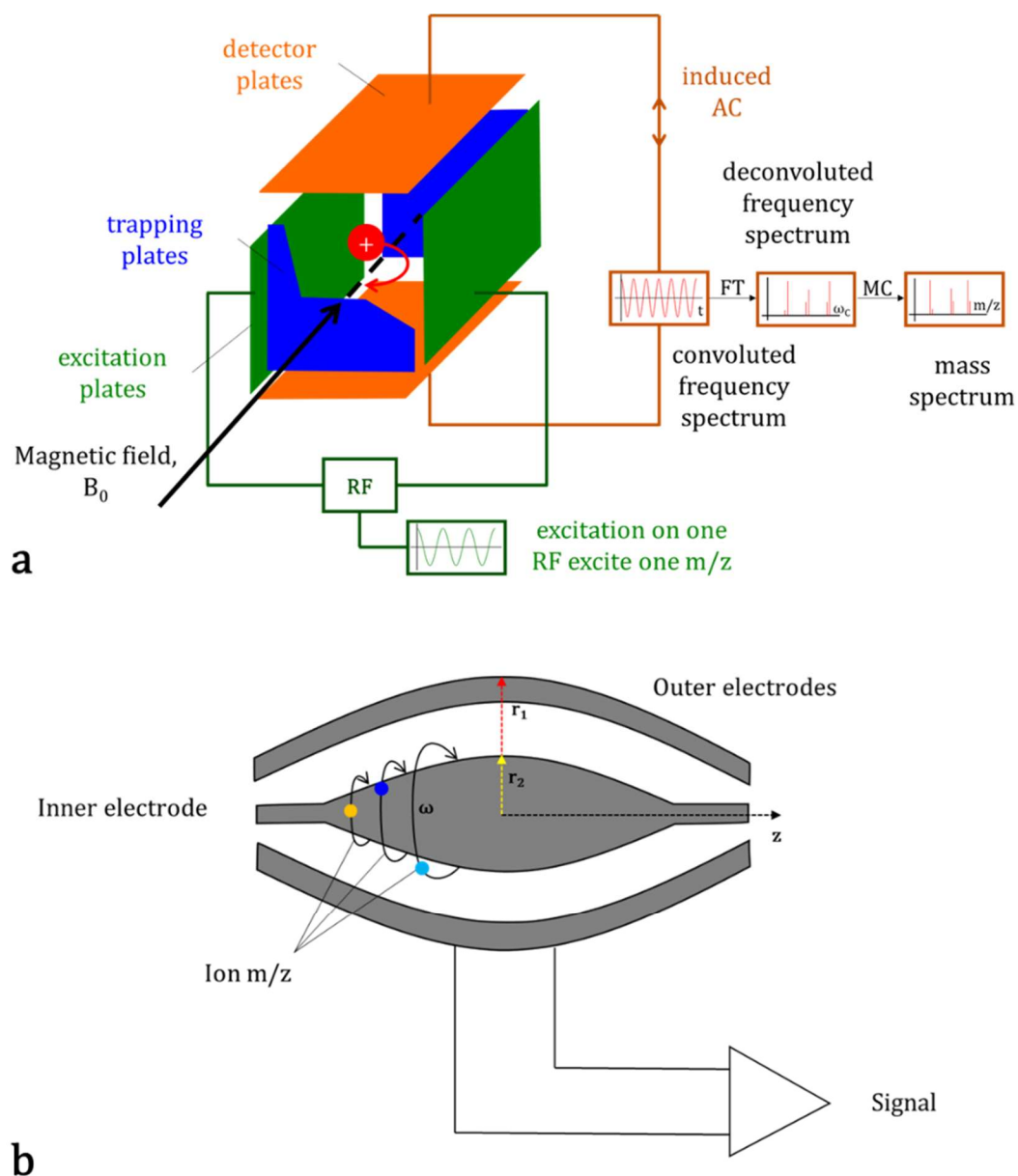


Figure 1.2.6. Schematic representation of (a) Fourier Transform-ion cyclotron resonance (FT-ICR) and (b) Orbitrap

1.2.4 Application of mass spectrometry

As previously mentioned, mass spectrometry can be used for both quantifying the amount of a given analyte but also for determining qualitatively its composition. MS can detect an almost infinitely wide range of molecules. This wide applicability, together with its high sensitivity and high throughput, results in very rich and complex data sets from which the analyte(s) of interest must be extracted.

It is a very versatile technique since it is not limited to the analysis of organic molecules; it can be used for the detection of any element that can be ionized. In this context, silicon wafers can be analysed to quantify the presence of lead and iron that can

cause the failure of a semiconductor for microprocessors, or drinking water can be sampled for determining arsenic content, or it can be used in geology or material science.⁴

During the MS experiment, a collection of ions is produced in the gas phase; the ions are separated according to their m/z ratio in vacuum and then detected to obtain the mass spectrum that displays the abundance of each ion. As previously mentioned, the ions need to be separated in a vacuum because the presence of matter would interfere with the detection. In particular, if an ion collides with neutrals in an elastic manner, its direction of travel may be altered, and the ion could not reach the detector. If the collision is inelastic, the energy transfer can cause decomposition so that the original ion could again not be detected.⁷

Until the late '80s of the past century, MS relied solely on EI and CI as ionization techniques. As described in the previous chapter, EI is an ionization technique that exploits the collision of an analyte with an electron ray, causing the ejection of one of its electrons. The ion formed has an odd number of electrons together with a positive charge, therefore its energy might be so high that it gets dissipated through the fragmentation of some of the chemical bonds unless the scaffold is particularly stable, such as those stabilized through resonance like aromatic compounds. Hence with this ionization technique, in the mass spectra, the molecular ion peak may not always be recorded. It is possible to gain valuable information from the analysis of the fragments, such as the so-called chemical "fingerprint." (Figure 1.2.7) CI works with the same principle, but the ionization is performed through the reaction of a reagent gas instead of using electrons, limiting the formation of fragments, as no radical species are involved in the process.⁴

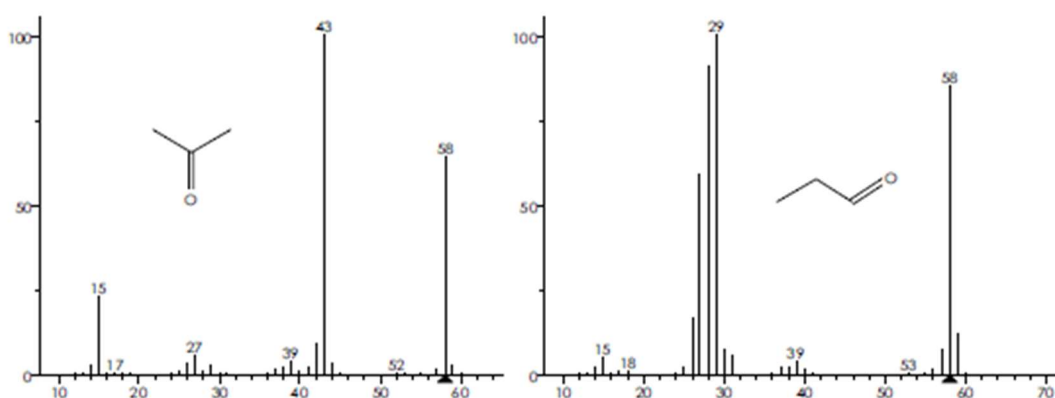


Figure 1.2.7. Example of the importance of the chemical fingerprint. EI spectra of acetone (left) and propionaldehyde (right). Both have a M.W. of 58 Da, the fragment peaks depend on the chemical structure.⁴

Both these techniques are still widely used, but as stated before, they have some limitations with the range of analytes that can be handled because they both require

molecules to be volatile and thermally stable. The analysis of small molecules is not really affected by this, and any of the ionization modes will give valuable information, but in case of analysis of biomolecules such as proteins or nucleic acids, the choice should be for ionization techniques such as ESI or MALDI thanks to their ability to analyse up to several hundred thousand Daltons without disrupting amino acid bonds, nucleic acids base pairing, and many non-covalent interactions.¹³

ESI is widely used as ionization source because it can integrate liquid chromatography prior to the mass analysis, which is especially important to remove excess salt, sample buffer, or phosphate buffer saline that may create artifacts and contribute to the chemical noise in the mass spectra. The main difference is that this kind of ion source results in protonated or deprotonated molecules and other adducts, whereas EI generates radical cations. The result is that the ions generated with ESI (protonated/deprotonated analyte, Na⁺, K⁺-adducts) depend on the characteristic of the analyte and on the experimental conditions used. The protonated/deprotonated molecular ion from ESI has limited excess energy, so few fragmentation occurs; this can be a disadvantage or an advantage. While no much structural information can be retrieved and no "fingerprint" can be recorded and compared to literature, ESI has the great advantage to record the molecular ion, which is also very important for the characterization of small molecules and moreover the "softness" of the ionization technique allows to maintain also non-covalent interaction, which are very important for the study of the binding of small molecules to their macromolecular target.⁴

Furthermore, there is a way to overcome the missing structural information due to the lack of fragmentation: the use of in-source fragmentation, tandem MS (MS/MS), and multistage MS (MSⁿ), that can induce the formation of product ions that can be used to study the structure of the analyte of interest.¹³ In this context, collision-induced dissociation (CID) is by far the most common technique. (Figure 1.2.8) MS/MS can be thought of as a way to obtain the mass spectrum of a mass spectrum; selected ions formed in the ion source get activated, and their fragmentation products are analysed. The activation of the ion is the crucial step and defines the products result.

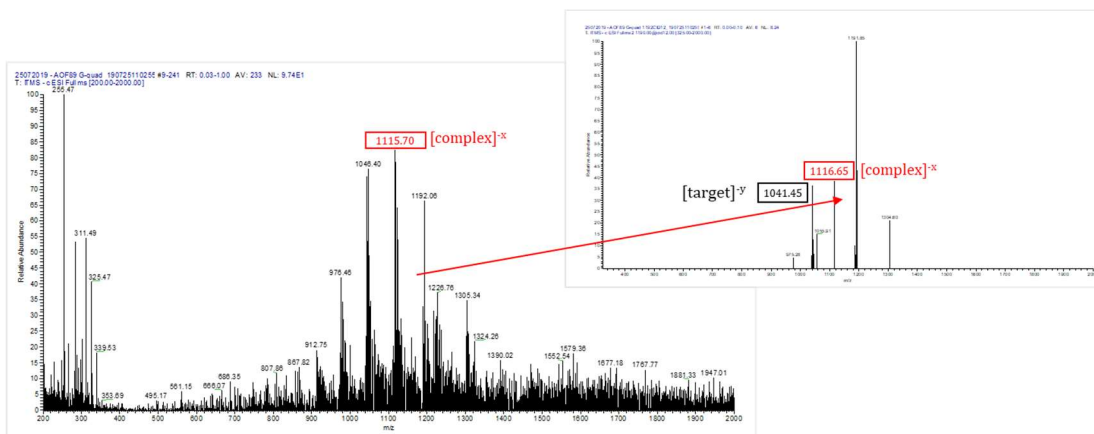


Figure 1.2.8. Example of a collision-induced dissociation (CID) experiment. The complex (red) ion on the left spectrum is isolated and fragmentation is induced in the right spectrum by the collision with an inert gas and lead to the formation of two peaks: complex (red) and target (black).

In CID experiments, the selected ions are isolated in the mass analyser and accelerated thanks to the application of an AC potential to increase their kinetic energy; the frequency of the AC depends on the m/z . The ions then are allowed to collide with an inert gas (such as He, N₂, Ar, or Xe) which causes the conversion of the kinetic energy into internal energy that results in bond breakage and the fragmentation of the molecular ion. Specific fragment ions may be isolated and undergo the same process generating an MS³ spectrum that can be further iterated to multistage fragmentation or MSⁿ, leading to the so-called "mass spectral trees".¹⁴ This kind of experiment is particularly useful both for small molecules because you can gain a molecular "fingerprint" which is even more accurate than the fragment ions obtained by EI since they are the direct consequence of the application of energy to a specific ion, *i.e.*, the molecular ion. Hence, the assignment of the specific peaks to a particular chemical entity is unambiguous. Moreover, this application is particularly interesting for the study of target-small molecule binding. The ion corresponding to the complex can be isolated and fragmented causing the breakage of the complex into target and small molecule. The energy needed to disrupt the complex can be correlated to the intensity of each peak allowing the determination of a parameter called $E_{COM}^{50\%}$, which gives an idea of the stability of the complex.⁶

1.3 Nuclear magnetic resonance (NMR)

1.3.1 Introduction to NMR

NMR spectroscopy is one of the techniques that have had the greatest impact on understanding biomolecular structure and dynamics, and it is widely used in chemistry, biochemistry, and related fields to elucidate the structure, composition, and dynamics of molecules. NMR takes advantage of the magnetic properties of certain atomic nuclei to provide detailed information about the arrangement of atoms, connectivity, chemical bonding, and even the motion of molecules in various environments. Like all other spectroscopies, it uses electromagnetic radiation to promote transitions between nuclear energy levels. In this case, the electromagnetic radiation is a RF in the range of approximately 10 to 1000 MHz to determine the transitions of nuclear spins in the presence of a static magnetic field.¹⁵

The discovery of NMR finds its roots in the work of Purcell, Pound, and Torrey of Harvard University and Bloch, Hansen, and Packard of Stanford University when, in 1945, they independently observed that magnetic nuclei, such as ^1H and ^{31}P , were able to absorb RF energy when placed in a magnetic field of a strength specific to the nucleus, and that this absorption led to the resonance of such nuclei at different frequencies for different atoms within the molecule.¹⁶

The most important property of the nuclei involved in NMR is the nuclear spin (I), whose value depends on the number of protons and neutrons and therefore that the same element isotopes adopt different values. I can assume integral or half-integral values and is equal to zero when both the number of protons and neutrons is odd. If $I = 0$, the nucleus has no spin angular momentum, thus no magnetic moment, and cannot, therefore, be analysed by NMR.

If I is not equal to zero, the nucleus generates a magnetic moment (μ) which depends on the magnetic quantum number (m) and can assume $2I + 1$ states/orientations. These states are degenerate in the absence of a magnetic field, meaning that the magnetic moments are randomly oriented. However, when atoms are placed within an external magnetic field (B_0), the magnetic moments align with the field either parallel (low energy state) or antiparallel (high energy state), and they become non-degenerate. Therefore, transitions between those states can be observed (Figure 1.3.1a). The splitting between non-degenerate spin states is determined by:

$$2\pi\nu = \omega = \gamma B_0$$

where ν is the transition frequency in s^{-1} (Hertz or Hz), ω is the transition frequency (also called Larmor frequency) in radial units (radians/s), B_0 is the strength of the magnetic field in Tesla and γ is the gyromagnetic ration, a constant that depends upon the identity of the nucleus. Therefore, the magnetic moments of each nucleus precess or rotate at a characteristic Larmor frequency (ω_0) along the direction of B_0 , conventionally z-axis. Both the lower and the higher energy states are populated, but there is an excess of nuclei that are aligned with B_0 in a parallel manner and this leads to a net magnetization M_z along the z-axis. The components along x- and y-axis are average to zero because magnetic moments are arranged evenly around the z-axis. A representation of the magnetization is depicted in Figure 1.3.1b. NMR signals are obtained when nuclei absorb energy from an RF pulse that matches their ω_0 , causing nuclear magnetic moments to flip from a lower energy state to a higher energy one. Every RF pulse is therefore characteristic for each nucleus.¹⁵

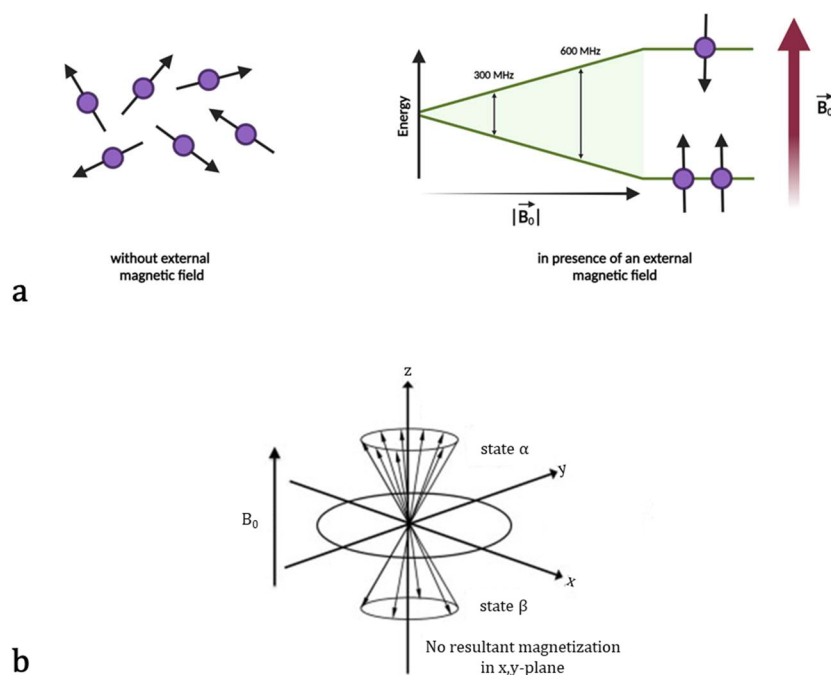


Figure 1.3.1. (a) Cartoon representation of the magnetic moments of the nuclei in absence and in presence of an external magnetic field (B_0). (b) Representation of the final net magnetization. Adapted from "Organic Chemistry" P. Yurkanis Bruice.¹⁷

During NMR experiments, the RF applied to the magnetic moments causes the spin transition and rotation of the net magnetization from the z-axis (direction of the B_0) into the transverse (x, y) plane; this is called excitation. After the RF, the nuclei begin to precess or rotate along the (x, y) plane and spiral back to the direction of the B_0 , emitting

radiofrequency signals and returning to their equilibrium. The signal that is collected is an oscillating signal with a frequency equal to the Larmor frequency of the nucleus under examination, and it gets weaker over time. Therefore, it is called free induction decay (FID) and is detected by the receiver coil. The detected FID signal is subjected to Fourier transformation, which converts it from the time domain to the frequency domain, generating the NMR spectrum.¹⁸

Electrons move in orbitals around the nucleus, surrounding every atom. Their movement as charged particles generates a magnetic field that is opposite to the B_0 , thus shielding the nucleus from the applied B_0 . The higher the electron density, the higher the shielding, and electron-donating groups shield nuclei more, causing the spin transitions to happen at a lower frequency. The opposite effect happens for electronegative or electron-withdrawing groups. Nuclei experiencing different magnetic fields are termed "inequivalent nuclei." Therefore, shifts of the NMR signal due to the chemical environment, known as chemical shift (δ) which measured in parts per million (ppm), can be observed and allows structural investigation allows for the determination of chemically inequivalent environments. The result is a spectrum featuring multiple peaks, with each peak corresponding to a unique chemical environment. (Figure 1.3.2) The area beneath each peak is directly proportional to the number of nuclei in that specific chemical environment. This means that signal integration can be used to determine the number of nuclei causing it and the ratio between them in different substances. The absolute value of δ increases with a higher applied B_0 . Therefore, different instruments would give different values for the same molecule. For this reason, it is preferable to use the relative δ , which is determined as follows:

$$\delta = \frac{\nu_{sample} - \nu_{ref}}{\nu_{ref}}$$

where ν_{sample} is the frequency specific for each nucleus and ν_{ref} is the frequency of a standard reference compound.

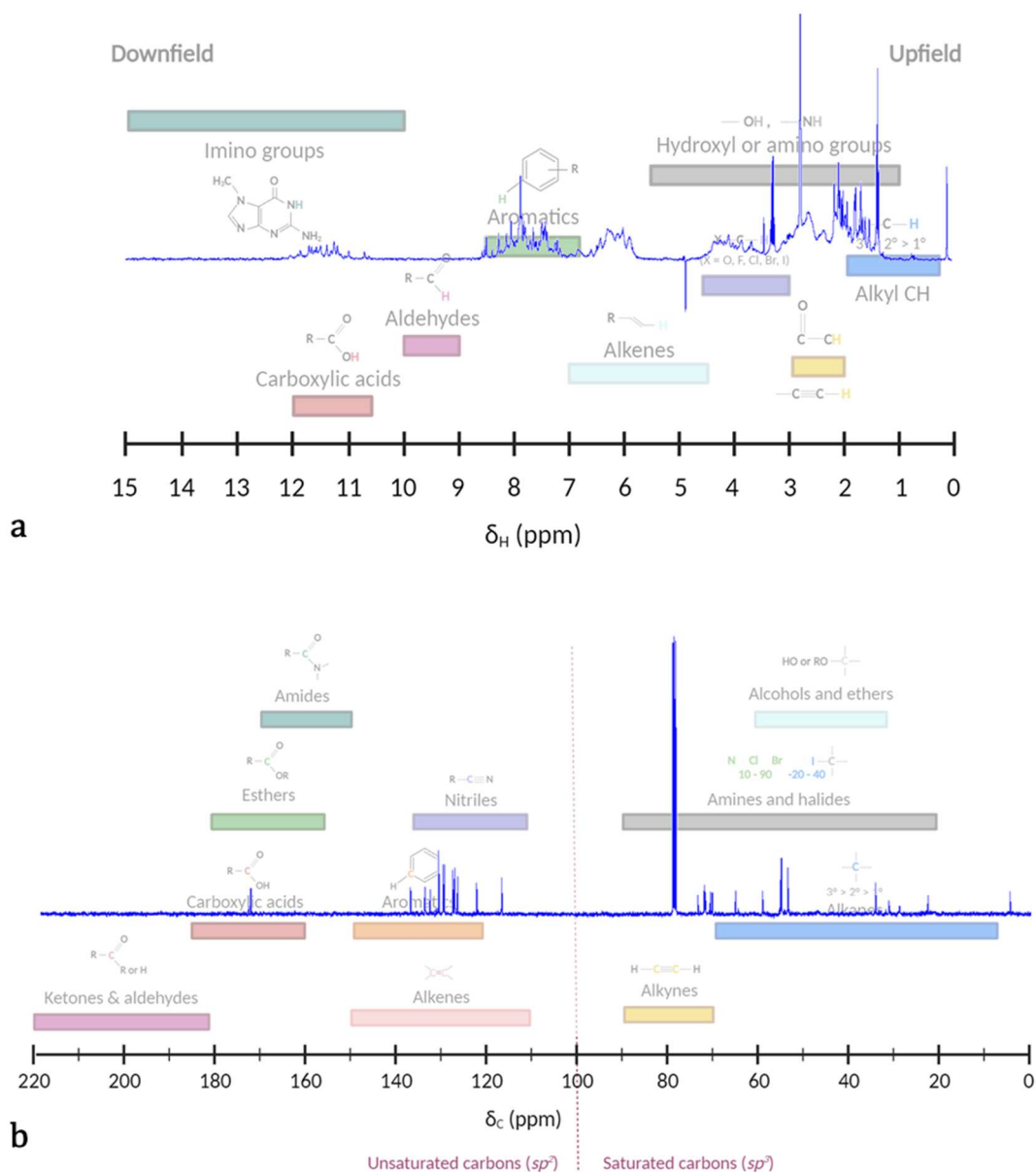


Figure 1.3.2. Representation of the common chemical shifts (δ) observed for ^1H and ^{13}C of the most common functional groups.

Nuclei of the same molecule that are not equivalent can be close enough for their magnetic fields to influence each other. This phenomenon causes the signal to be split to form the so-called "multiplet," and the interaction between the nuclei is known as scalar coupling (J), measured in Hz and corresponding to the distance in Hz of the lines in the multiplet.

The magnetic moments of each nucleus can be oriented in different ways depending on the number of possible states. For example, for atoms whose $I = 1/2$, there are two possible orientations: parallel or antiparallel to the B_0 . When two inequivalent atoms are placed together, their spins might have the same or opposite orientation, causing the field for each of them to be different. Hence, the signal is split. The number of signals that a

nucleus produces is proportional to the number of atoms to which they are coupled. The general rule is that the signal of one nucleus adjacent to n equivalent protons is split into $(n+1)$ peaks.¹⁷

Normally, coupling over one (1J), two (2J , geminal coupling), and three bonds (3J , vicinal coupling) are observed. The intensity of vicinal couplings depends on the torsion angle (θ) between the protons and follows the so-called Karplus equation:

$$^3J(\theta) = A \cos^2(\theta - 60^\circ) - B \cos(\theta - 60^\circ) + C$$

A plot of the described formula is depicted in Figure 1.3.3.

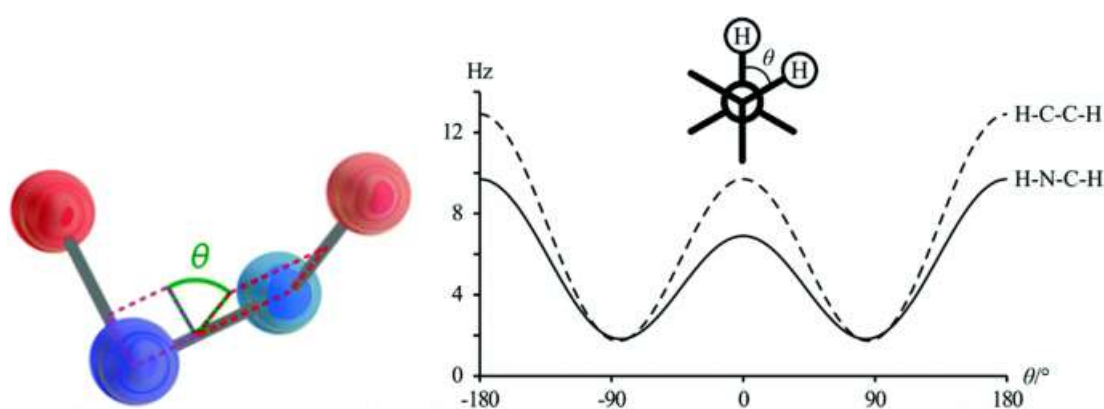


Figure 1.3.3. (left) Dihedral angle between protons (red) three bonds apart, blue may be carbons or heteroatoms. (right) Dependence of 3J -coupling constant on dihedral angle for peptide data as given by Karplus equations.¹⁹

1.3.2 Application of NMR spectroscopy

The first and most common use of this technique is the characterization of small organic molecules, thanks to the acquisition of monodimensional spectra, such as ^1H -, ^{13}C -, ^{15}N -NMR, or bidimensional spectra.¹⁵

Sample preparation is the same for all the experiments and consists of dissolving a small amount of the compound in 200 to 600 μL of solvent depending on the outer diameter of the tube used. The concentration of the sample depends on the spectrometer's resolution and the molecule's molecular weight (M.W.), but it is normally between 50 μM to 10 or 100 mM. It is necessary to work in the presence of a deuterated solvent; therefore, the analyte is normally directly dissolved in an appropriate deuterated solvent. This is necessary for two reasons: the hydrogens of the solvent are much more numerous compared to the hydrogens of the analyte, and their signal could cover the signals of the analyte. The intensity of the peaks is directly proportional to the concentration of the atoms that cause them. It is possible to suppress the signal of the

solvent with an appropriate pulse sequence. More importantly, the presence of deuterium (which is also NMR active), even in a small percentage, is essential because its signal is used by the spectrometer to "lock" the reference frequency and to check for variations in the magnetic field during the experiments.¹⁸

The easiest and fastest NMR experiment is ^1H -NMR. It consists of two sections: preparation and detection. The preparation section consists of a pulse sequence that starts with a relaxation time to allow magnetic equilibrium. Then a hard 90° pulse is applied to excite the nuclei, transferring the magnetization into the (x, y) plane. The detection is when the emission of the RF signal is recorded. The scheme of the experiment is depicted in Figure 1.3.4a.²⁰

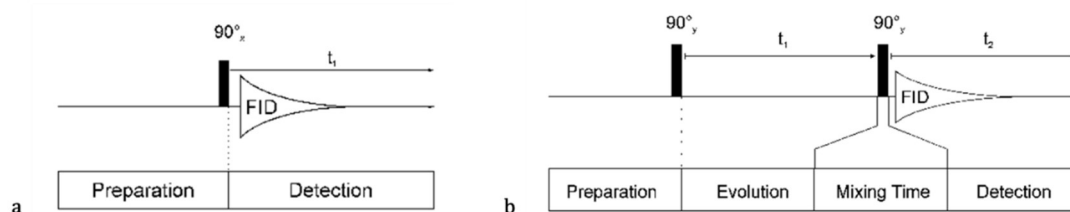


Figure 1.3.4. General scheme of (a) a 1D ^1H -NMR and (b) a 2D NMR experiments.²⁰

By analysing the number, the position, the multiplicity, and the integrals of the peaks, we can gain information about the chemical structure. A representation of the typical chemical shifts for various functional groups is reported in Figure 1.10. Moreover, if an external reference is added to the solution, we could also quantify the amount of the analysed compound.

Another common and equally easy NMR experiment is the ^{13}C -NMR, which is very similar to the ^1H experiment but with some differences. The sensitivity is lower than proton NMR for three reasons: the low isotopic abundance of ^{13}C (1.1%, whereas ^1H is $\sim 100\%$), the low γ (which is one-fourth of ^1H), and the longer relaxation time. The lower γ also has another effect on the ^{13}C nuclei: their resonance frequencies are lower than those of ^1H in the presence of the same magnetic field. For this reason, the signals of ^{13}C are observed in a completely different RF window. The chemical shift range for different ^{13}C is wider than the range for ^1H , conventionally from 0 to 220 ppm for ^{13}C whereas it is typically 0 to 15 ppm for ^1H ; therefore, the signal of each carbon is almost always seen as a distinct peak without overlapping, that on the contrary is very common in ^1H spectra.¹⁵

The low natural abundance of ^{13}C makes it unlikely for two ^{13}C atoms of the same molecule to be close to each other, hence no spin-spin coupling can be seen between two adjacent carbons. However, heteronuclear coupling between ^{13}C and hydrogen (or

fluorine) does indeed happen, and the coupling constants are really large (100-250 Hz). The multiplet would, therefore, be spread in a higher δ range, which complicates the analysis of the spectra. As a result, ^{13}C spectra are normally recorded with the so-called decoupling, in which the sample is irradiated at the frequency of one proton with continuous low-power RF. This excites all the H nuclei and cancels out the coupling pattern, leading to the formation of one singlet for each unique carbon. The larger relaxation time of ^{13}C causes NMR experiments to run longer than those of ^1H , which would not be a viable option for typical organic samples. Therefore, instead of using a 90° pulse excitation, normally a 30° or 45° pulse is used, significantly decreasing the time needed for the experiment.¹⁷

The higher the complexity of the analyte, the higher the need to develop experiments that may resolve the ambiguity of overlapping resonances. With this intent, two-dimensional NMR tries to overcome this problem by adding a second dimension to the resulting spectrum.

In addition to the preparation and detection sections, the 2D experiment adds an indirect evolution time t_1 and a mixing sequence. (Figure 1.3.4b) The first pulse given during preparation causes the precession of the nuclei that are left to precess freely for a given time t_1 during the evolution step. During this time, the magnetization is labelled with the δ of the first analysed nucleus. Then the magnetization is transferred to the second nucleus in the so-called mixing time thanks to a second RF pulse. The transfer can happen thanks to two mechanisms: scalar coupling or dipolar interaction (NOE), which will be better explained in the following paragraph. After the transfer, the data are acquired (detection), during this time the magnetization is labelled with the δ of the second analysed nucleus.²¹

The 2D spectra are composed of two frequency axes. If the two analysed nuclei are of the same isotope (homonuclear), a diagonal which corresponds to the 1D spectrum, divides the spectrum into two equal halves; outside the diagonal, symmetrically, there are the cross-signals that are the result of the second dimension and indicate interaction between the two nuclei. Common examples of homonuclear 2D experiments are COSY, TOCSY, and NOESY.

COSY (Correlation Spectroscopy) is a technique that reveals through-bond correlations between directly bonded nuclei, particularly between pairs of protons (hydrogen nuclei) that are coupled via scalar J . Protons that are more than three chemical bonds apart give no cross signal because their 4J is close to 0. Hence, only protons that

are two or more bonds apart give cross-peak signals in a COSY spectrum. It is useful for determining proton-proton connectivity in organic molecules as it can help identify which hydrogen atoms in a molecule are coupled together and elucidate the structure of complex organic compounds like natural products or pharmaceuticals by establishing spin systems.²²

TOCSY (Total Correlation Spectroscopy) is an extension of COSY that reveals correlations through multiple bonds, providing more comprehensive connectivity information. The magnetization is dispersed over a complete spin system of an amino acid or a sugar by successive scalar coupling, correlating all protons of a spin system. It is used to elucidate spin systems in complex molecules, such as proteins, nucleic acids, and carbohydrates, and to identify long-range proton-proton connectivity in large organic molecules.²³

NOESY (Nuclear Overhauser Effect Spectroscopy) uses dipolar interaction of spins (the NOE) for the transfer of magnetization. The intensity of the NOE is proportional to $1/r^6$, where r is the distance between the two nuclei. The correlation, therefore, depends on the distance between the two nuclei. Normally, a cross-peak is registered if r is smaller than 5 Å. NOESY helps determine the spatial proximity of nuclei within a molecule. It is useful for determining the three-dimensional structure of small organic molecules, especially when crystallography is not feasible, and for analysing intermolecular interactions in ligand-receptor complexes, which is crucial for drug discovery.²⁴

It is also possible to analyse two nuclei of different isotopes, which is particularly useful for the study of proteins and nucleic acids that contain many other magnetically active nuclei, such as ^{15}N or ^{13}C . The natural abundance of both species is low, but there are two strategies to increase their low sensitivity: either by isotopic enrichment of these nuclei and/or by enhancing the signal-to-noise ratio using inverse NMR experiments in which the magnetization is transferred from protons to the hetero nuclei.

HSQC (Heteronuclear Single Quantum Coherence) correlates the chemical shifts of different nuclei with 1J , typically ^1H and ^{13}C , and identifies which proton resonances correspond to which carbon atoms.

HMBC (Heteronuclear Multiple Bond Correlation) is similar to HSQC but correlates the chemical shifts of two nuclei separated by multiple bonds, revealing long-range connectivity.²⁵

The importance of NMR in drug discovery

The importance of NMR is given not only by the possibility of characterizing and investigating the chemical structure of small molecules. As will be better described in the following chapter, NMR can be used to kinetically follow some important reactions,²⁶ or also to study the interactions of the ligands with target protein or nucleic acids. NMR can be used to screen large libraries of compounds to identify those that bind to a target for example by fragment-based drug discovery.²⁷ Moreover, it can be used to study the interactions between a lead compound and its target protein, protein-protein complex, nucleic acid in detail in order to optimize the lead compound, making it more potent and selective²⁸ or to determine the three-dimensional structure of a protein or a nucleic acid both in presence and in absence of a binder. Furthermore, NMR can be used to study the absorption, distribution, metabolism, and excretion (ADME) properties of drug candidates for understanding how a drug will behave in the body and whether it is likely to be effective.²⁹

1.4 Computer aided drug design (CADD)

In the past decades, the field of computer-aided drug design (CADD) has brought a significant transformation in the way that we use to identify small molecules that target large molecular assemblies, such as proteins and nucleic acids.³⁰

CADD is a multidisciplinary field that combines principles of chemistry, biology, and computational science to expedite the process of discovering and designing new drugs. It involves the use of computational tools and techniques to predict, analyse, and optimize drug candidates, ultimately streamlining the drug discovery and development pipeline.

CADD involves four main aspects: target identification, which means identifying the molecular targets that are associated with a certain disease; hit identification, CADD uses to search chemical libraries for potential drug candidates or leads that could interact with the chosen target; lead optimization, once the leads are identified, CADD can be used to help optimize their chemical structure to enhance their efficacy and safety; preclinical and clinical testing, the promising candidates move on to preclinical and clinical trials, where their safety and effectiveness are rigorously assessed.³¹

CADD has been increasingly gaining attention because of the large benefits derived from its application. Firstly, it allows to save costs and times thanks to the more efficient identification of potential candidates. It reduces laboratory work by narrowing down the list of compounds for synthesis and testing, minimizing the need for extensive laboratory work. Furthermore, computational models are constantly improving, resulting in higher prediction accuracy and fewer experimental failures and it encourage the collaboration between chemists, biologists, pharmacologists and computational scientists.³²

CADD techniques comprehends a huge list of techniques that can be divided into different classes depending on the information that is available, in particular there are two subjects in a computational experiment, *i.e.*, the biological target and the potential binder and they can both be either known or unknown. If the information regarding both the subjects is available, we can make use of what is called structure-based drug design (SBDD), which relies on the knowledge of a target's 3D structure, typically determined through X-ray crystallography or NMR or cryo-electron microscopy (cryo-EM), and include virtual screening, molecular docking and molecular dynamics simulations in which libraries of molecules get screened to find new potential leads; if the 3D structure of the target is unknown but we have access to the information of the structure and the properties of the binders, we can make use of ligand-based drug design (LBDD), as

quantitative structure-activity relationships (QSAR), pharmacophore modelling and similarity search which exploits the information about the known ligands for the identification of new hits; when the information about the target is known but there are no or too few ligands, *de novo* drug design, in which entirely new compounds are generated computationally based on desired properties and target interactions, can be enrolled. Lately, the use of machine learning and artificial intelligence has become an always more advantageous thanks to the ability to predict biological activities, to analyse large datasets and to optimize drug candidates.³³

1.4.1 Structure-based drug design (SBDD)

As stated above, SBDD makes use of the 3D structures of biological molecules generated from either X-ray crystallography, NMR studies, cryo-EM, or homology modelling, which are then employed to screen extensive compound libraries. This approach greatly aids medicinal chemists in the initial stages of identifying and optimizing ligands that interact with the target. In this category, we can include techniques such as target protein and binding site identification, molecular docking and molecular dynamics simulations.³⁴

Target protein and binding site identification

Target structure and binding site characterization is the initial and pivotal stage within the SBDD process. The experimental determination of binding sites within the 3D structure of the target macromolecule relies on integrative structural biology techniques, such as NMR, X-ray crystallography and cryo-EM. Once the target protein's structure is resolved, the subsequent step involves the identification of the binding pocket(s).³⁵ This pocket represents a confined space where the ligand binds and exerts its therapeutic or desired effects. Moreover, even though the binding site of a specific protein is already known, identifying new potential binding pocket may be important for achieving different biological effects. To gain insights into the energy interactions and Van der Waals (vdW) forces within the binding site, various methods have been developed for binding site mapping. These methods play a critical role in pinpointing the specific regions of the target protein that interact with the appropriate functional groups on drugs and can be divided into: sequence-based approaches which exploit sequence and structural information to search in homology models conserved residues in a protein family which are assumed to be linked to the protein function; template-based approaches which

compare specific arrangement of residues that are essential for a macromolecule to perform its function; geometric methods which move the focus from the residues to the features of the binding site, they rely on the fact that a binding site is usually a cleft or a pocket, different points are assigned to the structure of the protein and of the solvent so that the pocket is defined as a region in space that contain solvent points surrounded by protein points; energy-based approaches that are based on the assumptions that a binding site is characterised by specific energetic properties which are diverse from the rest of the protein and analysing these energetic features, regions where ligands, such as drug molecules, are likely to bind can be identified.³⁶

An example of a binding site prediction is depicted in Figure 1.4.1.

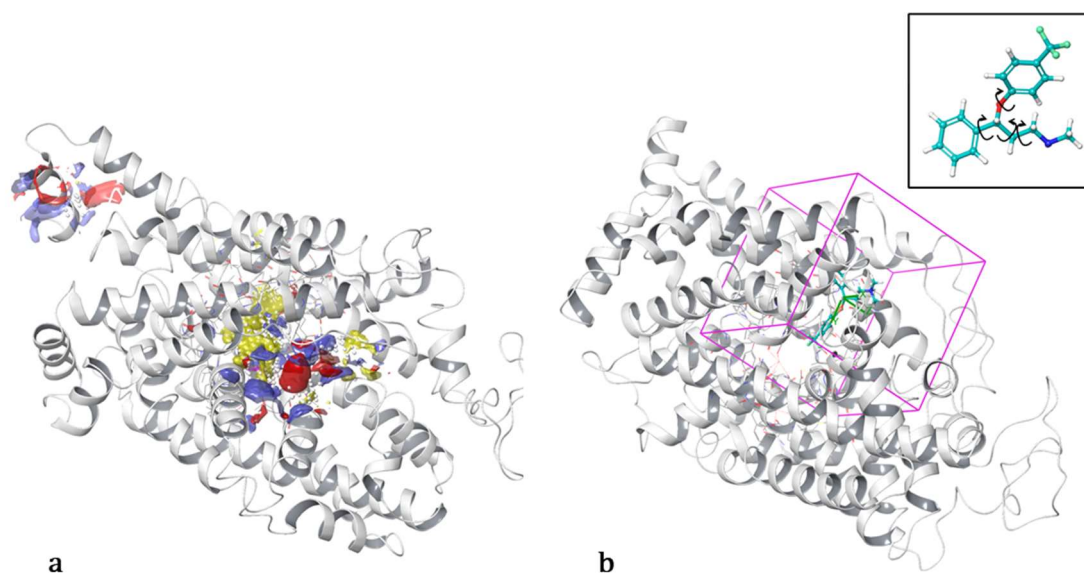


Figure 1.4.1. (a) Example of a binding site prediction on serotonin transporter (SERT) (PDB ID 5I73) highlighting multiple possible outcome; (b) example of a molecular docking within the binding site. The inset shows the chemical structure of the molecule prior the computation with few of the possible degrees of freedom.

Molecular docking

Molecular docking is a computational approach used to model the interaction between small molecules and proteins at the atomic level. It involves predicting the conformation (pose) and position of the ligand within the protein's binding site and assessing the binding affinity. The binding site may be known beforehand, enhancing docking efficiency, or it can be performed without prior assumptions ("blind docking").³⁷ Early docking methods were based on the "lock-and-key" theory, treating ligand and receptor as rigid bodies. Later, the "induced-fit" theory introduced flexibility in both ligand and receptor, resulting in more accurate binding predictions.³³ However, due to computational limitations, flexible ligand and rigid receptor docking remains prevalent. (Figure 1.4.1b)

Docking consists of two main steps: sampling ligand conformations in the active site and ranking these conformations using a scoring function. Given the extensive number of possible binding modes between molecules, various sampling techniques are employed in molecular docking. Matching algorithms, which use molecular shape and chemical features, map ligands into a protein's active site quickly.³⁸ Incremental construction methods break the ligand into fragments breaking its rotational bonds, one fragment (which may have a significant functional role) is initially placed in the active site and the remaining fragments are added incrementally.³⁹⁻⁴¹ Stochastic approaches like Monte Carlo randomly modify ligand conformations to explore conformational space. Monte Carlo generates ligand poses through bond rotation, and poses passing an energy-based criterion are saved for further modification. This process continues until the desired number of conformations is collected. Monte Carlo's main advantage is its ability to surmount energy barriers on the potential energy surface due to substantial conformational changes.^{42,43}

The scoring function's purpose is to distinguish correct poses from incorrect ones in terms of efficacy of the binding, they are normally expressed in kcal/mol and represent the binding energy of the complexes, therefore they can be used to rank the compounds efficiency of the target-ligand binding.³⁷ Scoring functions can be categorised into force-field-based, empirical, or knowledge-based functions. Force-field-based scoring functions evaluate binding energy by computing non-bonded interactions, including electrostatic and van der Waals forces. Electrostatic terms are calculated using Coulombic principles, with a distance-dependent dielectric function often applied to account for real environmental conditions. Van der Waals interactions are described by the Lennard-Jones potential function, and different parameter sets can control the "hardness" of the potential. Force-field-based functions suffer from slow computational speed, often necessitating the use of cut-off distances, which reduces accuracy for long-range effects. Extensions of these functions consider hydrogen bonds, solvation, and entropy contributions.⁴⁴⁻⁴⁶ Empirical scoring functions decompose binding energy into components like hydrogen bonding, ionic interactions, hydrophobic effects, and binding entropy. Coefficients for each component are determined through regression analysis on ligand-protein complexes with known binding affinities, resulting in relatively simple energy terms.⁴⁷⁻⁵¹ Knowledge-based scoring functions utilize statistical analysis of crystal structures of ligand-protein complexes to establish interatomic contact frequencies and distances. These interactions are converted into pairwise atom-type potentials, with scores favouring favourable contacts and penalizing repulsive interactions within a set cut-off

distance. Knowledge-based functions are computationally simple and can model rare interactions but may be limited by training data bias, especially with uncommon interactions like those involving metals or halogens. Consensus scoring combines multiple scores to evaluate docking conformations, improving the identification of strong binders during virtual screening. While it enhances pose prediction, binding energy predictions may still lack accuracy.⁵²⁻⁵⁶ To address this, physics-based scoring methods like MM-PB/SA and MM-GB/SA (MM stands for molecular mechanics, PB and GB for Poisson-Boltzmann and Generalized Born, respectively, SA for solvent-accessible surface area) are employed for rescoring and lead optimization to enhance binding affinity predictions.^{57,58}

Molecular dynamics (MD) simulations

Molecular dynamics (MD) simulations are computational methods in which the movements of single atoms belonging to a molecular system are predicted over time and are ran including a solvent in contrast with molecular docking. These simulations are based on a model of the physical forces governing interactions between atoms. They can provide valuable insights into various biological processes, including how proteins change shape, how they bind to other molecules, and how they fold into their functional structures.⁵⁹ MD simulations can also help understand how biomolecules respond at the atomic level to different conditions like mutations, chemical modifications, or changes in their environment. MD simulations are often used in combination with a wide variety of experimental structural biology techniques, including X-ray crystallography, cryo-EM, NMR, electron paramagnetic resonance (EPR), and Förster resonance energy transfer (FRET). While MD simulations have been used for several decades, they have gained significant popularity and visibility, especially among experimental molecular biologists in recent years.⁶⁰

The increased attention to MD simulations can be attributed to two main factors. First, recent breakthroughs in experimental techniques, like X-ray crystallography and cryo-EM, have provided a wealth of structural data for critical biomolecules, making them suitable for MD simulations. Second, advances in computing hardware, such as graphics processing units (GPUs), have made MD simulations more accessible and affordable, allowing researchers to perform powerful simulations on standard computers. Lastly, even though MD simulations are based on approximate physical models, they are valuable because they offer a detailed, atomic-level view of biomolecular systems, allowing precise control over simulation conditions.⁶¹

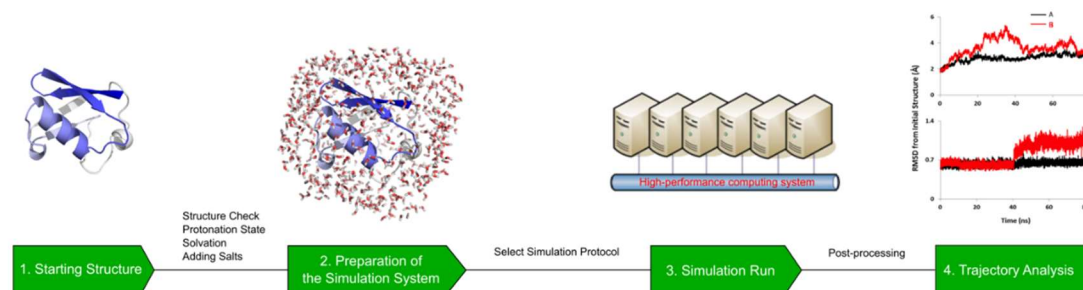


Figure 1.4.2. Representation of the steps of a molecular dynamics simulation experiment.⁶²

In order to perform a MD simulation, we need an algorithm which can take into account three components: a description of the system, a description of the system's initial condition and a defined force field (ff). The system description typically encompasses details such as the identity and coordinates of all its atoms, along with their mass, charge, and radius. It also includes information about the bonds between these atoms and particulars about the solvent molecules in the system. Establishing initial conditions necessitates specifying the initial velocities for each atom, the simulation temperature, and defining boundary conditions. The ff, a complex aspect of the simulation, substitutes classical mechanics for quantum physics in the simulation of biological systems.⁶⁰ Figure 1.4.2. shows an example of the workflow of a MD simulation.

Biological systems are often too intricate to be precisely described at the atomic level using quantum physics due to their complexity. As a result, approximations are employed, such as classical mechanics instead of quantum physics, ignoring electron movement, and using approximated ff.⁶³ In the classical simulation, atoms are represented as solid spheres with van der Waals radii, and their mass and charge are concentrated at their centres. Bonds between atoms are depicted as springs obeying Hooke's Law. The macromolecule is described as a specific set of spheres and springs. In order to simulate the movement, the position of each atom through time needs to be known and is described by the simple equation of motion where velocity is calculated in terms of acceleration and therefore related to internal energies U . The ff incorporates multiple terms that combine to calculate the total energy of the system. These terms arise from different interactions and contributions within the system, comprising bonded and non-bonded interactions. Bonded interactions include stretching which considers the change in bond length, bending that evaluates the angle formed by three-atom bonds, and torsion of bonds. Non-bonded interactions are typically represented by Coulomb and Lennard-Jones potentials, accounting for charged atoms and induced interactions such as van der Waals forces. The

sum of all the specified terms give rise to the equation of the ff which is described as follows:⁶⁴

$$U = U_{bending} + U_{stretchin} + U_{torsion} + U_{Coulomb} + U_{Lennard-Jones}$$

$$U = \frac{1}{2} \sum_{i=1}^{N_1} k_i^s (l_i - l_i^0)^2 + \frac{1}{2} \sum_{i=1}^{N_2} k_{ij}^b (\theta_{ij} - \theta_{ij}^0)^2 + \frac{1}{2} \sum_{i=1}^{N_3} V_n [1 + \cos(n\omega - \gamma)] + \sum_{i,j=1}^{N_4} \frac{q_i q_j}{r_{ij}} + 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6 \right]$$

Several software programs are available for conducting MD simulations, with popular choices like CHARMM,⁶⁵ Gromacs,⁶⁶ NAMD,⁶⁷ and Amber,⁶⁸ each offering unique features and capabilities.

MD simulations can be described as a ff application for a system that it is allowed to evolve over time from the starting conditions. This simulation takes place within a virtual environment or "box" where specific variables are held constant. These variables are typically associated with specific ensembles, such as the microcanonical ensemble (NVE), canonical ensemble (NVT), or isothermal-isobaric ensemble (NPT). The simulation can be run either on the receptor alone, on the ligand alone, or on the complex formed when the ligand binds to the receptor. Typically, the system is placed within an explicit solvent box. Depending on the chosen timescale for the simulation, various phenomena can occur and be observed and measured. One of the most significant measurements obtained by MD is the root mean square deviation (RMSD) that is crucial assess the overall stability of the system. It can be calculated separately for the receptor or the ligand and plotted in a single graph to account for the ligand movement during the simulation. The RMSD measures the average displacement of atoms at a given moment during the simulation concerning a reference structure, typically the initial frame or the crystallographic structure. It serves as a valuable tool to analyse structural changes over time⁶⁹ and is calculated as follows:

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N [r_i(t) - r_i(0)]^2}$$

where $r_i(t)$ and $r_i(0)$ are the coordinates of the i -th atom at time t and 0 , respectively, and N is the number of atoms in the domain of interest.⁷⁰

Deviations from the initial coordinates often indicates an unequilibrated simulation. However, in an equilibrated simulation, it can happen that the structure oscillates around a consistent average conformation and only a subset portion of the macromolecule is fluctuating, therefore other parameters must be taken into account, such as the radius of gyration (Rg), which measures the compactness of a structure,⁶⁰ and root mean square

fluctuations (RMSF), which are typically associated with the fluctuation of specific residues in the context of protein simulations, and is calculated as follows:

$$RMSF_i = \sqrt{\frac{1}{J} \sum_{j=0}^J [r_i(j) - \bar{r}_i]^2}$$

where $r_i(j)$ are the coordinates of the i th atom in the time step j , \bar{r}_i their average positions, and J is the time of simulations, expressed as the total number of time steps collected.⁷⁰

1.4.2 Ligand based drug design (LBDD)

Ligand-based drug design, also known as indirect drug design, is a powerful approach in the realm of pharmaceutical research and development. This methodology is particularly valuable when structural information about the three-dimensional structure of the target receptor is lacking. Instead, it leverages knowledge of molecules that interact with the biological target of interest. Ligand-based drug design aims to identify or design molecules, known as ligands, that can effectively bind to and modulate the function of the target.³³

In contrast to SBDD, LBDD does not necessitate prior knowledge of mechanisms of action but relies on structural information and bioactivity data for small molecules. LBDD is based on the principle that molecules with similar structures tend to have similar properties.⁷¹

Often, LBDD involves retrieving and preparing small molecule libraries. These chemical structures are usually represented and processed as molecular graphs, consisting of nodes and edges to represent atoms and bonds, respectively. Molecular graphs are communicated using connection tables and linear notations. Connection tables contain information about atom types, connection types, and coordinates and are found in formats like mol2, sdf, and pdb. Linear notations, like Simplified Molecular Input Line Entry Specification (SMILES) and Wiswesser line notation, use alphanumeric characters and are more compact for storing or transferring large sets of small molecules.³³

The primary LBDD techniques include molecular similarity-based search, QSAR, and pharmacophore modelling.

Molecular Similarity-Based Search

Molecular similarity-based search is a straightforward LBDD technique used to identify specific small molecules of interest. In this method, small molecule libraries are searched using molecular descriptors that are numerical values representing various characteristics of small molecules, ranging from simple physicochemical features to complex structural properties. These descriptors, derived from experiments, quantum-mechanical tools, or previous knowledge, include properties like molecular weight, atom types, bond distances, surface area, electronegativities, and more.⁷² Depending on the dimensionality, descriptors can be 1D, 2D, or 3D. 1D descriptors represent scalar physicochemical properties, while 2D descriptors are derived from molecular constitution or configuration, and 3D descriptors are derived from molecule conformation.

These molecular descriptors allow for rapid comparisons of structural and physicochemical features of small molecules. The Tanimoto coefficient (T) is a popular tool for measuring similarity between two molecules. While a T value greater than 0.85 suggests a good fit, it doesn't necessarily indicate biosimilarity.

Applications of molecular similarity-based search include the identification of novel targets based on the chemical similarity of small molecules and predicting off-target interactions and adverse drug reactions (off-target prediction).⁷³

Quantitative Structure-Activity Relationship (QSAR)

QSAR is a computational technique for establishing relationships between the chemical structures of molecules and their biological activity. It is commonly used in rational drug design, and it can also be applied to predict various physicochemical properties, hence the term quantitative structure-property relationship. QSAR is based on the assumption that molecules with similar structures are likely to have similar biological activities.⁷⁴

Data type determines the statistical approach for building QSAR models. Quantitative data are processed using regression-based methods, while graded response data are processed using classification-based methods. Common techniques for building linear QSAR models include multiple linear regression (MLR), principal component analysis (PCA), and partial least square (PLS) analysis.^{75,76}

In contrast to linear models, biological systems often exhibit nonlinear relationships between molecular descriptors and biological activities. Neural networks are widely used

for modelling these nonlinear relationships. It is essential to validate QSAR models, and this is done using both internal and external validation methods.⁷⁷

Ligand-based pharmacophore (LBP)

LBP is a favoured method when there's a lack of structural information about the therapeutic target. According to the International Union of Pure and Applied Chemistry, a pharmacophore is defined as a combination of steric and electronic attributes necessary for facilitating optimal interactions with a specific biological target structure to either activate or inhibit its biological response.⁷⁸

LBP is designed to identify a 3D pattern of features that's vital for enabling most input ligands to bind effectively to the receptor. This process becomes increasingly intricate with more input ligands and their varying degrees of flexibility. Hence, performing a conformational search is a crucial, albeit costly, step in LBP. An essential part of LBP involves identifying a "bioactive" conformation for an active molecule, serving as a reference to align the remaining molecules. In cases where a 3D bioactive conformation is unavailable, databases are scoured to find a molecule conformation resembling the input ligands. Alternatively, the most active molecule can be geometrically optimized, and its resulting minimum energy conformation is considered the bioactive conformation for aligning the other molecules. It's crucial to validate the developed LBP before further utilization. Much like QSAR, a separate set of molecules is designated as a test set to validate the predictive capabilities of the LBP.

After the selection of active and inactive molecules and the optimization of the molecular structure using an appropriate ff, the molecules are superimposed using their bioactive conformation or the minimized conformation of the most active molecules. Then, the generated pharmacophore model is used to screen databases of compounds.^{79,80}
(Figure 1.4.3)

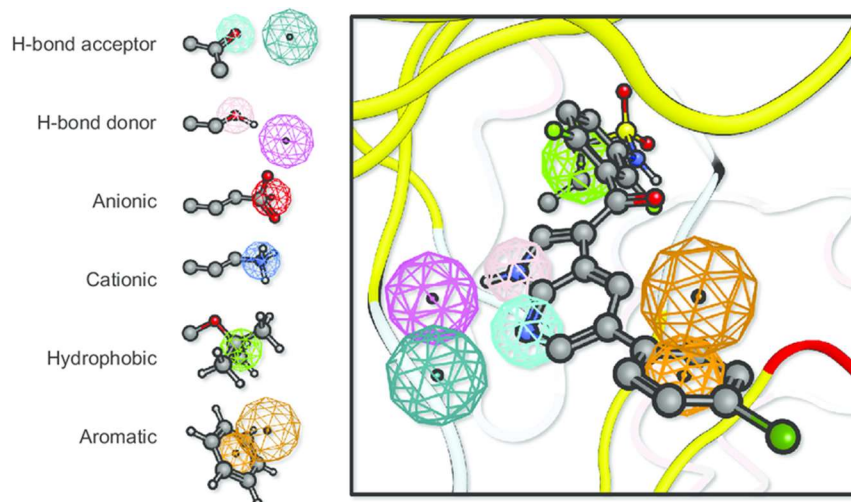


Figure 1.4.3. Pharmacophore query. Notes: A pharmacophore query is comprised of different features. The features represent molecular recognition motifs such as hydrogen bond acceptors or donors, anionic, cationic, hydrophobic, and aromatic groups.⁸¹

1.5 Circular dichroism (CD) spectroscopy

Circular dichroism is a spectroscopy that measures the difference in absorption of left and right circularly polarised light (CPL) of chiral molecules and it is defined as follows:

$$CD = A_l - A_r$$

It is used for all kind of molecules but it finds the most important application in the study of large biological molecules. For example, as protein's secondary structure is sensitive to environments changes, such as temperature or pH, CD is used to check how the arrangement change with these variables or with the interaction of other molecules.⁸²

A molecule is defined as chiral if it cannot be superimposed to its mirror image by any combination of rotation, translations and some conformational changes. The two mirror images are called enantiomers, all the physical and chemical properties of enantiomers are identical except for the way they interact with polarised light and the way that they interact with other chiral molecules.¹⁷

Linearly polarised light is an electromagnetic wave whose oscillations are confined to a single plane; (Figure 1.5.1) CPL is the result of two linearly polarised light waves of equal amplitude that are out-of-phase by a quarter-wave. This results as a helix that can be right- or left-handed and their images are non-superimposable. If the vector rotates counter clockwise when the observer looks down the axis of propagation, the light is left circularly polarized (L-CPL). If it rotates clockwise, it is right circularly polarized (R-CPL).⁸³ (Figure 1.5.1)

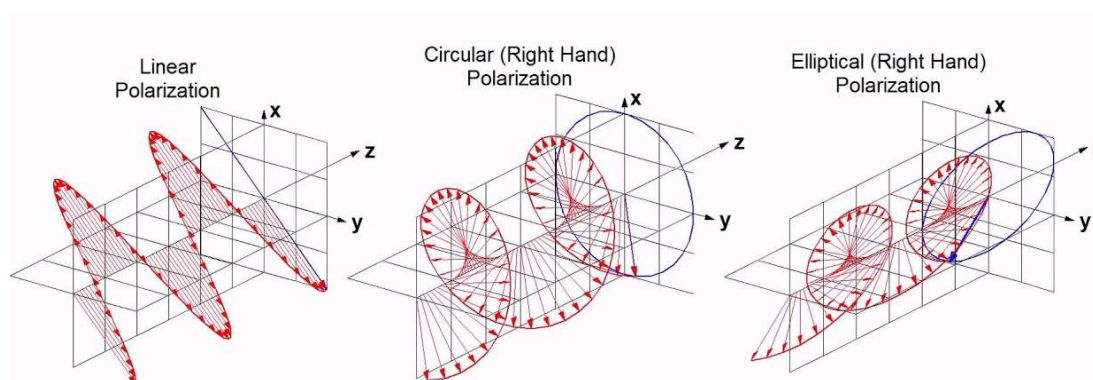


Figure 1.5.1. Representation of linearly, circularly and elliptically polarised light.⁸⁴

At the same time, linearly polarized light can be thought as a combination of L-CPL and R-CPL with the same amplitude, a solution of chiral compounds creates an anisotropic environment where L-CPL and R-CPL propagate at different velocities, therefore the phases of the two components changes and the linearly polarized wave starts

to rotate. This is called optical rotation and it is quantified with a polarimeter, that determines the optical rotatory dispersion (ORD) measuring the optical rotation as a function of wavelength.⁸⁵

For CD to occur, light has to be absorbed by the molecule, and in particular L- and R-CPL are absorbed to different extent because of the difference in indices of refraction for the two components. Therefore, CD occurs only at specific wavelengths of light and the spectrometer will record it withing a range of specified frequency so that the result is a spectrum in which the differential absorption of the components of the radiation is displayed as a function of frequency or wavelength.⁸⁶

The CD experiments (Figure 1.5.2) starts with a light source which emits unpolarised light, it passes through a monochromator which select a specific wavelength and then a linear polarizer that filters the light to ensure that all the light waves are vibrating in the same plane, making it linearly polarised. Before hitting the sample light pass through a modulator, typically a photoelastic modulator or Pockels cell, this component rapidly changes the polarization of light between R-CPL and L-CPL, resulting in a beam of light that alternates the polarization. The alternating CPL enters the sample cell where the analyte is placed inside a cylindrical cuvette and where the two components of light will be absorbed to different extent, causing a difference in absorbance. After passing through the sample, light is directed to a detector (a photomultiplier tube or a photodiode) that measures the intensity for both L-CPL and R-CPL. The difference is recorded at different wavelength and the data is plotted as a CD spectrum, showing how the absorbance difference varies across the selected wavelength range.⁸³

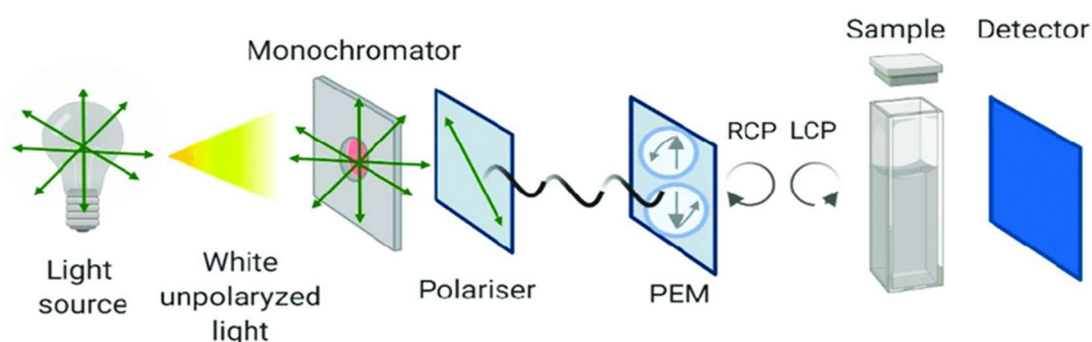


Figure 1.5.2. Schematic representation of CD experiment

CD is measured therefore as a difference in absorption of light and according to Lambert-Beer, the equation above can be rewritten as

$$CD = \Delta A = (\epsilon_l - \epsilon_r)Cb$$

where ϵ is the molar absorptivity or molar extinction coefficient that measures the probability of the electronic transition, C is the concentration of the sample and b is the cell pathlength. Rearranging the equation and dividing the CD by the concentration of the sample and the cell pathlength, we can obtain the molar circular dichroism $\Delta\epsilon$ which is expressed in $L \cdot mol^{-1} \cdot cm^{-1}$ (or M^{-1}/cm). For proteins there is another important unit for CD, the mean residue molar circular dichroism, substituting the concentration C of the protein with the mean residue concentration (CMR) obtaining the molar CD for individual protein residues instead of the whole protein molecules so that it is easier to compare proteins with vastly different size. Specifically, there are two ways to determine the CMR depending on how much is known about the protein: if the sequence is known we can obtain it by multiplication if the protein concentration in molar with the number of residues, whereas if the sequence is unknown you can estimate it by dividing the concentration of the protein in g/L per the average amino acid residue weight of 113 g/mol.⁸⁷

Historically, the unit for CD has always been the so-called *ellipticity*, expressed in degrees (or millidegrees since often the measurement is quite small) which is more relatable to optical rotation measurements as it measures the change in the polarisation state of a linearly polarised analysed beam. As previously mentioned linearly polarised light can be thought as a combination of left and right circularly polarised light with the same amplitude. When linearly polarised light pass through the optical active sample, the two components are still circularly polarised, but the magnitude is no longer equal, therefore its direction is no longer a circle but an ellipse. The ellipticity (θ) is defined as the tangent of the ratio of the minor to major elliptical axis. Molar ellipticity can be derived as $\Delta\epsilon$ to normalise it to pathlength and concentration (or CMR).⁸³

1.5.1 Applications of CD

CD spectroscopy is a non-destructive, label-free technique that provides valuable information about the chiral properties of molecules. It is widely use to probe the conformational, structural and interaction aspects of chiral compounds and mostly biomolecules, such as proteins and nucleic acids.

It can be used to gather information on the secondary structure of proteins and peptides as the CD spectra of α -helices, β -sheets, random coils and other structural elements are all characteristic for each conformation. Noteworthy with CD one can only determine the relative fraction of residues in each conformation but not where each

structural feature is located. Hence, it is used mostly to show conformational changes in molecules due to changes of pH, temperature, salt conditions, to determine how similar a wild type protein is to mutant or to gain information upon ligand binding.^{88,89}

CD spectroscopy can provide information about the secondary structure and conformational changes of nucleic acids, such as DNA and RNA. It is used to study the formation of DNA duplexes, triplexes, quadruplexes, and RNA secondary structures, and to monitor interactions with ligands or proteins. Furthermore, the CD spectrophotometer can be equipped with a temperature control system. The wavelength of the radiation can be fixed to a maximum in the CD spectrum of the analyte and temperature can be increased gradually recording the CD value at regular temperature intervals. Nucleic acids are organised (folded) into higher order structure that are held together by hydrogen bonds. As the temperature rises, thermal energy disrupts the hydrogen bonds, causing the disruption of the three-dimensional arrangement in a process called denaturation. As consequence, the CD value will gradually decrease to 0. The temperature at which this denaturation process occurs is known as the melting temperature (T_m), it can be calculated from the spectrum and it is defined as the temperature at which half of the nucleic acid molecules are denatured. The T_m should be registered upon ligand addition to trace the improved stability of the three-dimensional arrangement.⁹⁰

It is employed also in fields like organic synthesis, drug development, and natural product chemistry to analyse of chiral molecules and determining their absolute configuration and enantiomeric purity.⁹¹

2 Main project: Developing new ligands targeting G-quadruplex

2.1 Nucleic acids and G-quadruplex

2.1.1 Introduction to nucleic acids

Since the discovery of the DNA structure by Watson and Crick in 1953,⁹² research on nucleic acids has emerged as a prominent domain of scientific progress. This breakthrough started a multitude of investigations aimed at unravelling the intricate biological mechanisms occurring within cells, which led to the formulation of the central dogma of molecular biology: DNA was defined as the molecule responsible for information storage, RNA as a carrier molecule (such as messenger RNA, mRNA) or a template molecule (like ribosomal or transfer RNA, rRNA/tRNA), and proteins as the dynamically active agents within cells.⁹³ For this reason, for a substantial period of time research in medicinal chemistry was largely focused on proteins. One of the main reasons was the incorrect hypothesis that a single gene exclusively encoded a single protein (enzyme) which in turns catalyse a specific biochemical reaction, the theory is known as the One-Gene-One-Enzyme hypothesis and was postulated in the 1940s by George Beadle and Edward Tatum.⁹⁴ In reality, thanks to the discovery that many proteins contain more than one polypeptide chain specified by different genes, thanks to apprehending that one gene often give rise to more than one polypeptide since genes organise into polypeptide-encoding exons and noncoding introns, thanks to the phenomenon of alterative splicing and thanks to the discovery that some genes do not encode a polypeptide at all, but functional RNA molecules, the theory was proven wrong and nucleic acids research has increasingly played a significant role on the study of biological processes for potential therapeutic purposes.⁹⁵

Nucleic acids, and particularly DNA, were first discovered in the 1870s by the Swiss biochemist Friedrich Miescher, who isolated by precipitation a substance different from proteins and lipids for its high content of phosphorus and its resistance to proteolysis from the nuclei of white blood cells and called it “nuclein”.^{96,97} Subsequently, many scientists continued the research in this field but it is only thanks to the Russian biochemist Phoebus Levene that the composition of the nucleic acids was determined as polynucleotide model in which each nucleotide was composed by one of the four

nitrogen-containing bases, a sugar molecule and a phosphate group.⁹⁸ Levene's proposal ended with the identification of the so-called tetranucleotide structure according to which nucleotides were always linked in the same order as G-C-T-A repeats. This was also one of the causes of the lost interest on nucleic acids in drug developments because, if the tetranucleotide theory were true, DNA would have been a monomeric unit that could not encode information in its sequence, contrary to proteins that being composed of 20 different amino acids could explain better the complexity and diversity of the different life forms.⁹⁹ In 1950 the Austrian biochemist Erwin Chargaff published one of the major breakthroughs for the discovery of the structure of nucleic acids thanks to his analysis of DNA coming from different species.^{100,101} The first outcome of his research was that the nucleotide composition varies among species, hence the tetranucleotide structure was proven wrong. The second result was that all DNA maintains certain properties regardless of its origin: the so-called Chargaff's rule, whereby the amount of adenine (A) is always similar to the amount of thymine (T) and the amount of guanine (G) is always similar to the amount of cytosine (C).¹⁰²

The real proof that DNA indeed contains the genetic material was due to the contribution of Frederick Griffith and later Oswald Avery, Colin MacLeod and Maclyn McCarty. Griffith in 1928 postulated the "transforming principle", he demonstrated that genetic material can be transferred and transform living cells. He worked with two strains of the bacterium *Streptococcus pneumoniae*: a virulent (disease-causing) strain and an avirulent (non-disease-causing) strain. Killing the virulent strain and mixing its heat-killed remains with the living avirulent strain, led to observation that the avirulent strain became virulent, causing disease in laboratory mice. Therefore, genetic material could indeed be transferred.¹⁰³ In 1944, Avery, MacLeod and McCarty demonstrated that was DNA that could transform the species by isolating the crude RNA and destroying any protein, lipid, carbohydrate and RNA and by showing that the remaining DNA was still able to transform the avirulent strain whereas its degradation with DNase led to the loss of this ability.¹⁰⁴

After this major discovery, lots of scientists worked on determining the three-dimensional structure of DNA. Rosalind Franklin and Maurice Wilkins at King's College London obtained X-Ray diffraction data of DNA fibres and were the firsts proposing its helical structure,^{105,106} this research was continued by James Watson and Francis Crick in Cambridge that put together all the information and developed the well-known double helix structure of DNA in 1953 which will be described in the next paragraph.⁹²

As previously mentioned, nucleic acids are polymers of nucleotides, each composed of a pentose sugar (ribose in RNA and 2'-deoxyribose in DNA), a nitrogen-containing base or nucleobases and one or more phosphate groups. The building block for the synthesis contains three phosphate groups but two are lost during the process, meaning that in a strand each nucleotide contains one phosphate group. The nucleobases may be double-ring purines such as adenine (A) and guanine (G) or single-ring pyrimidines, as cytosine (C), thymine (T, in DNA) and uracil (U, in RNA). The carbon atoms in the sugar are numbered as 1' to 5', the nucleobase is linked to the C1' forming the N-glycosidic bond and the phosphate is bonded to the C5'. The chain is formed by the formation of the phosphodiester bond linking the hydroxyl of the C3' to the phosphate of the following nucleotide.¹⁰² A representation of the different nucleobases, of the nucleotide and of a strand can be depicted in Figure 2.1.1.

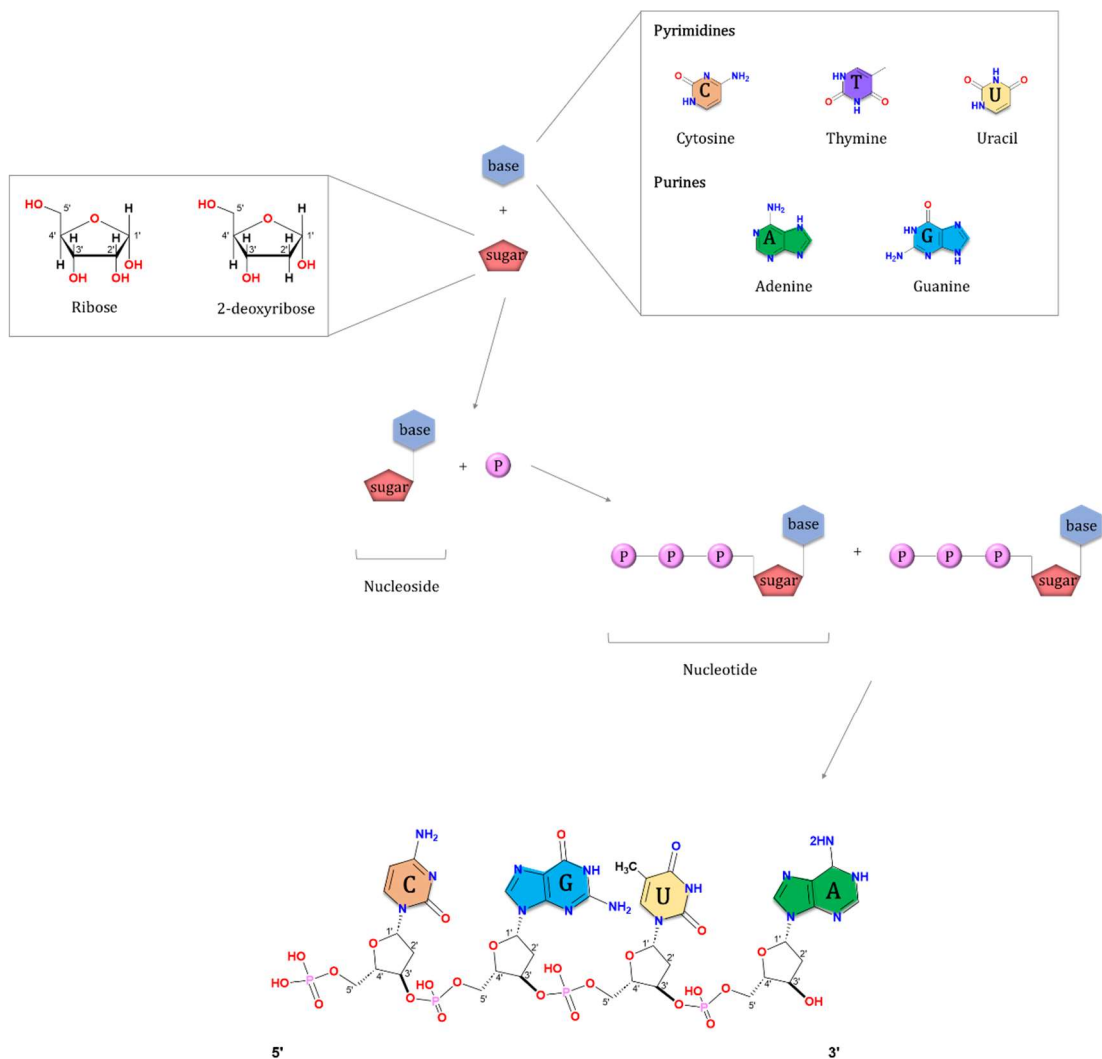


Figure 2.1.1. Schematic representation of a nucleic acid structure showing the chemical structure of the sugars, phosphate groups and nucleobases.

The pentose sugar is non-planar (puckering) and can adopt various conformations which are described using pseudo rotation parameters P and Φ_m as function of the five endocyclic torsion angles, although it has been demonstrated by Altona and Sundaralingam¹⁰⁷ that in the crystal structure of isolated nucleosides and nucleotides, there are two preferred geometries: the north conformer *C3'-endo* type and south conformers *C2'-endo* type which are depicted in Figure 2.1.2. Also, the phosphodiester backbone can be described with six torsion angles that are correlated one another.¹⁰⁸ Lastly the glycosidic torsion angle χ can also assume different conformations but it has a preference for two of them: *syn* and *anti* which are also depicted in Figure 2.1.2.¹⁰⁹

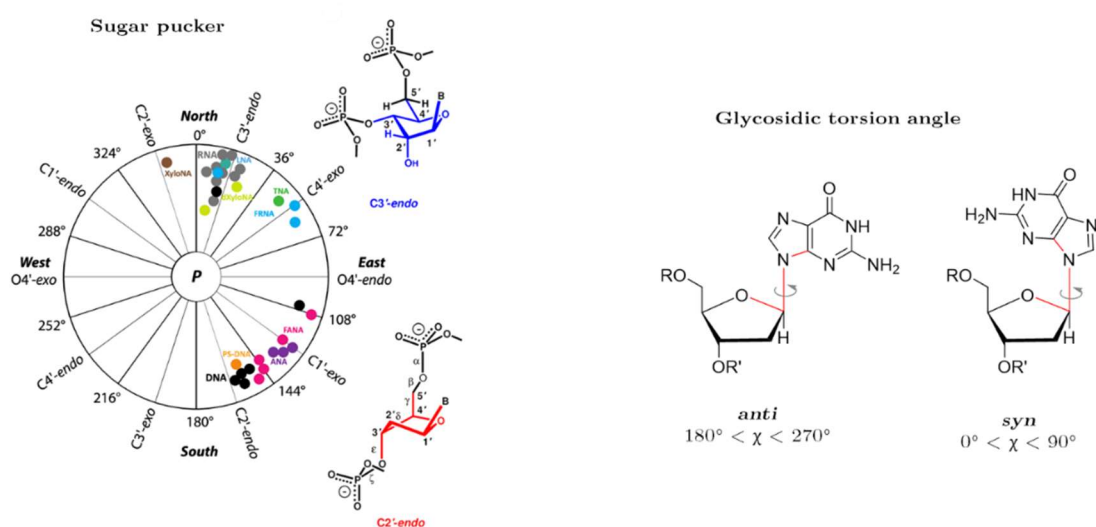


Figure 2.1.2. (left) Wheel of the sugar pucker with all the possible conformations, *C3'-endo* and *C2'-endo* are depicted as they are the two preferred conformations. Figure adapted from Egli¹¹⁰ (right) Representation of the two possible glycosidic torsion angles.

The secondary structure of nucleic acids is stabilized by base-pairing, which forms when two nucleobases interact through hydrogen bonds (H-bonds) at one of the three interaction sites of each base, which are described as Watson-Crick edge, Hoogsteen edge and sugar-edge.¹¹¹ By definition, base pairing requires the formation of at least two H-bonds and therefore Saenger et al. defined 29 different possible base pairs (with at least 2 H-bonds) taking into account not only the edges involved, but also their orientation.¹¹² The structural differences of nucleic acids range from minor variations in some local parameters of the Watson-Crick paired duplex structure to structure that are completely different, even in their essential features as handedness, base-pairing scheme or number of strands. In a recent review it has been revealed that all but four letter of the English alphabet have already been used to indicate the various structural geometries adopted by DNA.¹¹³

One of the possible base pairings is the one named Watson-Crick. The model proposed by Watson and Crick included the pairing between a purine and a pyrimidine and, more specifically, A with T/U and G with C. This base-pairing give rise to the most common secondary structure motif, the helix. Within helical structures, DNA or RNA strands engage in complementary base pairing interactions, resulting in their alignment in a coiled helical configuration that is stabilized by stacking interactions. These strands are typically arranged in an antiparallel manner. The formation of helical structures gives rise to two distinctive grooves, often referred to as the major and minor grooves, distinguished by their respective sizes.⁹²

In nature, three primary types of helices are observed: A- and B-helices (right-handed), and Z-helices (left-handed), which are depicted in Figure 2.1.3. A-helices exhibit a C3'-*endo* sugar-pucker configuration, while B-helices display a C2'-*endo* sugar-pucker configuration. In both cases, the nucleobases adopt an *anti*-configuration concerning the sugar moiety. Naturally occurring DNA is typically found in a double-stranded state, predominantly as B-DNA which is the model described by Watson and Crick (also called dsDNA). Conversely, double-stranded RNA is generally observed in the A-conformation.¹¹⁴

As previously mentioned, together with these canonical structures, also non-canonical structures are present in nature, which are termed non-B forms, and they have been demonstrated to influence critical genetic processes such as DNA replication, transcription, recombination and repair and they include the left-handed helix Z-form, *i*-motif and G4.¹¹⁵ During this work, the attention will be focused on one of these possible arrangements, namely the G4 that has been the focus of the main project.

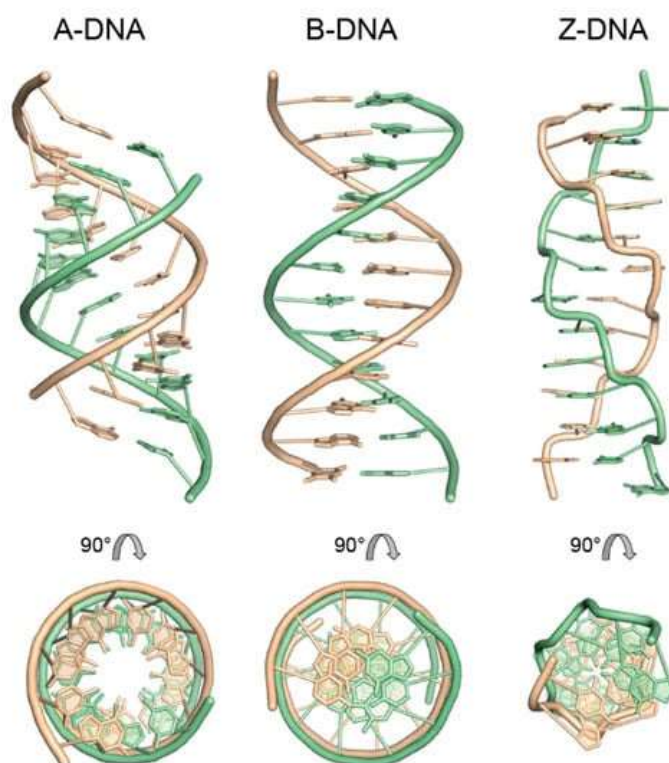


Figure 2.1.3. Different types of helices forms of DNA based on X-ray fibre diffraction analysis. Twelve-base-pair fragments of calf thymus A-DNA and B-DNA with arbitrary sequence dATCGATCGATCG and of Z-DNA with alternating sequence dCGCGCGCGCGCG. In the top row, the straight helix axis is vertical, and the view is into the minor groove at the helix center.¹¹⁶

2.1.2 G-quadruplex structure

The first difference in structure of G4 from B-DNA is the different base pairing, the Hoogsteen type which employs the so-called Hoogsteen edge that is defined by N6 and N7 in A and O6 and N7 in G. These non-canonical base pairs provide additional interaction pattern for building triplets and quartets, which are the fundamental structural units of triplex and quadruplex, respectively.

The ability of G-rich sequences to form G4 was first observed in 1910, when a concentrated solution of guanylic acid (5'-guanosine monophosphate, GMP) was demonstrated to form a gel in water at high millimolar concentrations.¹¹⁷ Only fifty years later, though, in 1962 the structure of this gel was determined by X-ray diffraction of guanylic acid fibres by Gellert, Davies and Lipsett.¹¹⁸ Thanks to Hoogsteen base pairing, four GMP units could form a square planar arrangement by cyclic interactions in which each G forms in total four H-bonds with two more G, in particular between the imino proton H1 and carbonyl oxygen O6 and between the amino proton H2 and one of the imidazole nitrogen, N7. This was called G-tetrad and is depicted in Figure 2.1.4. The stacking of more of these planes on top on each other (Figure 2.1.4) would form the

three-dimensional arrangement producing a helix that would justify the X-ray pattern and the jellification at high concentration.¹¹⁸

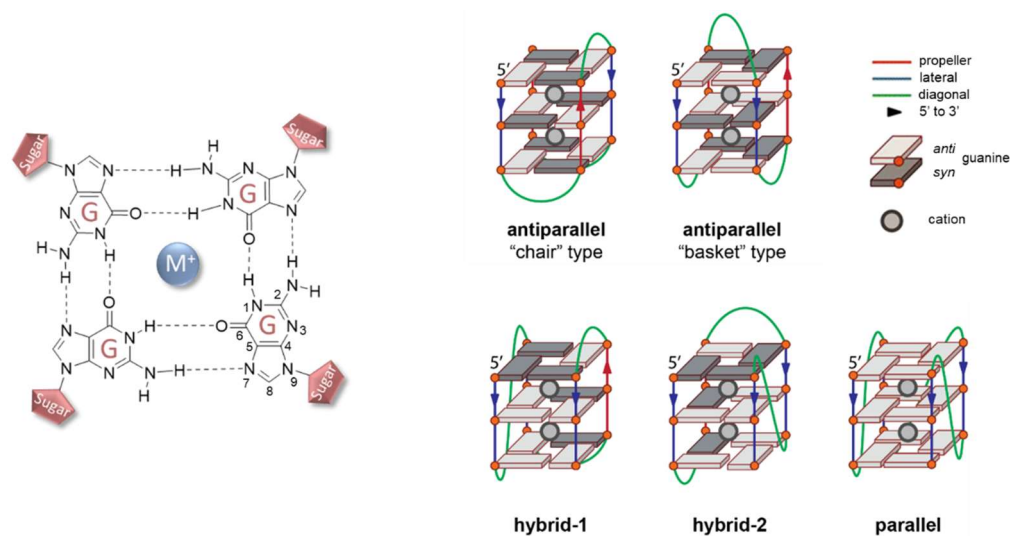


Figure 2.1.4. Representation of a G-quartet (left) and the resulting three-dimensional G4 with all the possible topologies (right)
Adapted from Luo *et al.*¹¹⁹

G4 can be formed by one nucleic acid molecule, which are known as intramolecular (monomeric), or by more than one nucleic acid molecules, thus they are defined intermolecular (multimeric). Intramolecular G4 are prevalent in biological context, mostly formed by three tetrads, they tend to form rapidly and display a remarkable diversity in their conformation, such as folding topology, loop conformation and capping structure.¹²⁰

The nucleotide in the sequence that connects the quartets are called loops and they can take on various shape: lateral (or edgewise), diagonal, and double chain reversal (or propeller). The four Gs forming the G-quartet can be differently oriented (regarding 5' to 3' polarity) and can be divided into four topologies: parallel if they have the same polarity (↑↑↑↑) often referred to as “propeller”-type, antiparallel if they have opposite polarity (↓↑↓↑ or ↑↑↓↓) on adjacent or diagonal strands (leading to two different geometries: “chair”-type and “basket”-type respectively) or hybrid with three strands share the same polarity and one has the opposite (↓↓↑↓ or ↑↓↑↑).^{121–123} (Figure 2.1.4)

Also, the glycosidic bond of the Gs can be different and as stated above it can be either *syn* or *anti* depending on the orientation of the strands. Parallel topologies contain all *anti* Gs, whereas antiparallel and hybrid structures contain a mix of *syn* and *anti*.¹²⁴ The combination of different loop shapes, strand orientations, and the number of strands involved results in a wide variety of G4 structures. It's worth noting that while there are theoretically 26 possible G4 topologies, only a limited number of them (6) have been

observed in laboratory settings.^{121,122} This suggests that some topologies may be energetically unfavourable, either thermodynamically or kinetically.

The factors that contribute to the stability of G4 structures and determine their specific folding patterns or a combination of patterns involve several key elements: stacking interactions, hydrogen bonding, solvation, and the binding of cations. The effects of cations on G4s are notably different from their effects on duplex DNA structures.¹²⁵

Similar to other nucleic acid secondary structures, the stability of G4s relies on the pi-pi stacking of consecutive guanines from different tetrads. The relative stability order of guanine stacks has been predicted, and it depends on their orientation (from 5' to 3'): *syn/anti* > *anti/anti* > *anti/syn* > *syn/syn*.¹²⁶ However, this order may not fully align with the actual structures, partly due to computational biases. Quantum mechanical calculations have provided a more consistent picture, with ΔE values of *anti/anti* > *syn/anti* > *anti/syn* > *syn/syn*.¹²⁷ The specific folding of a G4 depends on its sequence and the buffer conditions (including cation types and co-solvents). Stacking interactions are a significant source of energy gain, making it energetically favourable to increase the number of G-tetrads. The number of consecutive tetrads often matches the length of the guanine tracts, although exceptions can occur.¹²⁸ Higher-order G4 structures, such as dimers, can be formed through the stacking of external tetrads from monomeric units, contributing to additional stabilization. Most reported dimer structures exhibit a 5' to 5' interface.¹²⁹

Hydrogen bonding is another crucial element in stabilizing biomacromolecules, particularly the interaction between bases in nucleic acids. In G4, the stabilization of a tetrad involves eight H-bonds, and this cooperative effect exceeds the sum of four individual G•G pairs. This cooperation arises from charge transfer between pairs of guanines, leading to charge separation and stronger donor-acceptor orbital interactions.¹³⁰

Cations such as (K⁺, Na⁺, Li⁺, Mg²⁺, Ca²⁺, Sr²⁺, Pb²⁺ etc.) play a crucial role in the formation and stabilization of G4. These cations are coordinated by the O6 of Gs pointing towards the inner channel formed at the centre of the G-quartet. Laszlo and co-workers in 1980 demonstrated that the driving force for the assembly of GMP is actually the cation binding and not the hydrophobic interactions, therefore their presence and nature is critical to the formation of G4.¹³¹ The cavity formed at the centre of the G-tetrad is surrounded by negative charge, therefore dehydrated cations could be stabilized in it. The electrostatic repulsion between cations could be shielded by the negative charges on the O6, meaning that the cation would be positioned in between two G-quartets. Rueda et al.

demonstrated by MD simulations that the G4 cavities may indeed be empty but they are filled within few ns, so they have a short lifetime.¹³²

Regarding the nature of these cations, Wong and Wu showed by solid-state NMR studies that the stabilization induced by them on GMP self-assemblies has the following order: $K^+ > NH_4^+ > Rb^+ > Na^+ > Li^+$.¹³³

Noteworthy some bivalent cations (Ca^{2+} , Co^{2+} , Mn^{2+} , Zn^{2+} , Ni^{2+} and Mg^{2+}) demonstrated to be able to enter transiently open G4 and coordinate N7 and O6 preventing the formation of the G-quartets and therefore inducing the dissociation of G4.¹³⁴ Due to their physiological relevance, the coordination of Na^+ and K^+ is the most studied on X-ray determined structures and two G-quartets form mostly a bipyramidal geometry with the cation located between them. Thanks to its small size, Na^+ is the only cation that is able to coordinate in plane with four Gs as demonstrated in PDB: 1JB7.¹³⁵

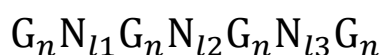
In general, Hud et al. showed that G4 coordinated with K^+ is more stable compared to G4 coordinated with Na^+ due to two main reasons: the coordination energy and the cation dehydration energy. As revealed by their melting temperature T_m , they demonstrated that there is a free energy gain from the conversion of Na^+ with K^+ due to the optimal fit of K^+ within the coordination sites formed by the G-quartets. However, considering the bigger energy difference required for the dehydration of Na^+ compared to K^+ , the preference for the latter is more probably due to the greater dehydration energy required to remove water from the Na^+ shell.¹³⁶

The nature of the stabilizing cation is a source of the polymorphism that is known for G4. Noteworthy, not only different sequences can adopt different topologies, but a given sequence can also fold into different conformation, and for example it has been demonstrated that different cations can also induce diverse G4 structures as demonstrated by Sen and Gilbert, especially for human telomeric DNA sequences.¹³⁷

2.1.3 Biological relevance of G4

Computational analyses have revealed that potential G4 motifs are not randomly distributed in the human genome and are evolutionarily conserved, suggesting their significance in various biological processes.

Early reports identified over 350,000 putative G4 sequences (PQS) using pattern-based rules, such as



where n is the number of G stacks and $l1$, $l2$ and $l3$ are the number of nucleotides belonging to loop 1, 2 and 3 respectively. The restraints for the parameters n , $l1$, $l2$ and $l3$ vary in different algorithms but the maximum length of the loop is usually set to 7 nucleotides, on the basis of molecular modelling¹³⁸ and biophysical studies¹³⁹ suggesting that an increase in the loop length has a major effect on the G4 stability. However, Yang and co-workers have recently shown that the major G4 conformation formed by the purine rich sequence involved in the regulation of the gene BCL2 contains a 13 nt-long chain reversal loop,¹⁴⁰ revealing the necessity to refine the prediction algorithm taking into account the thermodynamic stability of G4.

A distinct computational approach is the one using G4 sequencing (G4seq) that merging the genome-wide DNA polymerase-stop assay and high-throughput sequencing identified more than 716'000 G4s in the human genome in a volume (for a HeLa cell nucleus) of 0.22 pL, corresponding to a cellular G4 concentration of 6 μ M, which is quite remarkable.^{141–143}

PQS are relatively rare in the human genome compared to what would be expected by chance, about one third less suggesting evolutionary selective pressure against G4 formation in mRNA,¹⁴⁴ however they are highly concentrated in gene promoters, especially near the transcription start site (TSS).¹⁴⁵

Specifically, G4s exhibit a notable accumulation within a short distance (a few hundred bases) immediately preceding the TSS. This proximity suggests increased stability compared to G4 found elsewhere in the genome, likely due to their prevalence of one-nucleotide loops.¹⁴⁶

Importantly, over 40% of the promoter regions associated with annotated genes contain one or more PQS. DNase I hypersensitive sites (NHS) show enrichment in PQS in these regions, suggesting structural transitions from dsDNA to an unwound form that are prerequisites for transcription activation in gene promoters.¹⁴⁷

The "G4P calculator," (<http://depts.washington.edu/maizels9/G4calc.php>) assesses whether consecutive segments of 100 bases contain a minimum of four sequences, each comprising three or more Gs, irrespective of their spacing.¹⁴⁸ Maizels and colleagues introduced the concept of a "G-quadruplex genome" to describe these motifs' involvement in key processes and they adopted a different approach to quantify the potential of genomic region to form the G4 (G4P).¹⁴⁹ G-protein-coupled receptors, sensory perception, nucleosome assembly, ubiquitination and nucleic acid binding are examples of functions characterized by low G4P, while functions such as transcription

factor activity, cell signalling, muscle contraction, development, growth factors and cytokines present high G4P. Interestingly, the distribution of G4P for tumour suppressor genes was shifted from the RefSeq (RefSeq, <http://www.ncbi.nlm.nih.gov/refseq/rsg/about/>) median towards lower values, while the distribution of proto-oncogenes was shifted towards higher G4P. Analysis of the G4P frequency in genomic sequences used as reference standards for well-characterized genes shows that motifs with high potential of forming G4s (with loop size 1-12 nt) are frequent at the TSS, at the 5'-UTR and at the 5'-end of the first intron, but most of the coding regions are depleted of PQS.¹⁴⁸

The existence of G4 structures in functional parts of the genome has been a subject of debate.¹⁵⁰ Evidence for the presence of G4 at telomeres have provided using engineered antibodies. These antibodies have helped visualize G4 structures both at telomeres and outside them.^{151,152} However, the detection methods typically require cell fixation and permeabilization, which may impact cellular integrity. Alternative techniques targeting G4s in living cells have emerged, avoiding fixation and permeabilization. As an example, recently Di Antonio and co-workers demonstrated G4 formation *in vivo* in living cells through single-molecules fluorescence imaging using a fluorescent probe at a lower concentration with respect to biophysical and cellular experiments, leading to a minimized global perturbation of G4.¹⁵³ These methods have contributed to a better understanding of G4 distribution in the genome.

Overall, G4 have emerged as a fascinating area of research in genomics with potential roles in various biological processes and disease-related gene regulation.

G4 and DNA replication

G4 DNA have a dual role in DNA replication, acting as vital components of metazoan replication origins and, simultaneously, as obstacles impeding the replication process.

In mammalian genomes, numerous DNA replication origins are predicted, a significant portion of which (~ 80-90 %) contains GC-rich regions forming origin G-rich repeated elements (OGRE), with the potential to form G4 structures. A previous study demonstrated that G4 motifs within DNA replication origins are indispensable for initiating replication. Disrupting the stability of G4 structures also negatively impacts the functioning of these origins, as depicted in Figure 2.1.5a.¹⁵⁴ Moreover, a recent investigation confirmed that the removal of G4 motifs in OGRE significantly diminishes the initiation of replication in mouse cells. Interestingly, introducing a G4 sequence from

OGRE into an unrelated origin-free region can create a completely new functional replication origin.¹⁵⁵ The mechanism behind G4-induced replication initiation appears to involve the recruitment of factors responsible for starting replication at origin sites, such as the origin recognition complex, the treslin-MTBP complex, and the replication timing regulation protein Rif1, all of which have been reported to specifically bind to G4 structures, both in cell-free environments and within cellular systems.¹⁵⁶

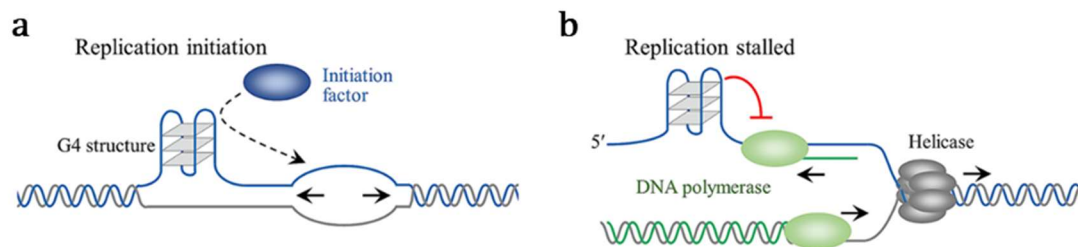


Figure 2.1.5. Possible G4 influence mechanism on replication. Adapted from Teng et al.¹⁵⁶

On the other hand, as DNA replication commences, replicative helicases are responsible for separating the two DNA strands. While the leading strand undergoes continuous synthesis, the lagging strand's synthesis follows a discontinuous pattern, causing it to become momentarily single-stranded and rendering it inherently more susceptible to the formation of G4 structures. This has the effect of stalling the DNA replication machinery, as illustrated in Figure 2.1.5b.

To enable the DNA polymerase to utilize the DNA strand as a template, G4 structures must be effectively resolved. Numerous DNA helicases have demonstrated their capacity to both bind to and unwind G4 structures in *in vitro* experiments. Among these helicases are human counterparts like WRN, BLM, FANCD1, and PIF1. Dysfunctions related to these helicases have been linked to genomic instability and an elevated risk of cancer.¹⁵⁷ In cell lines derived from human patients afflicted by Fanconi anaemia, it has been observed that the loss of the ability to unwind G4 is correlated with genomic deletions that overlap regions with high PQS.¹⁵⁸ This observation underlines the significance of the capability to resolve G4 structures in connection with human diseases.

Additionally, research by Paeschke et al. has postulated that the resolution of G4 by the PIF1 helicase serves to suppress genome instability.¹⁵⁹ This highlights the crucial role played by helicases in maintaining the stability of the genome.

G4 and gene transcription

G4s influence gene transcription in complex ways, and their impact depends on factors such as their orientation (on the template or non-template strand) and their location concerning the TSS. When G4 structures are present on the DNA template strand during transcription, they obstruct the progress of RNA polymerase (as depicted in Figure 2.1.6a). This obstruction leads to the interruption or even termination of transcription, directly reducing the expression of the target genes.^{160,161} Notably, G4s on the non-template strand of DNA can also disrupt the transcription process by forming a stable RNA/DNA hybrid with the nascent RNA, as illustrated in Figure 2.1.6b.^{162–164}

In contrast, recent research has revealed that G4s on the non-template strand can interact with and stabilize a structure called an R-loop,¹⁶⁵ facilitating transcription through a mechanism involving the repeated formation of R-loops.¹⁶⁶ However, it's essential to note that the spreading of R-loops caused by numerous stabilized G4 structures can lead to the accumulation of DNA double-strand breaks (DSBs) in human cancer cells, resulting in genome instability and triggering cell apoptosis.¹⁶⁷

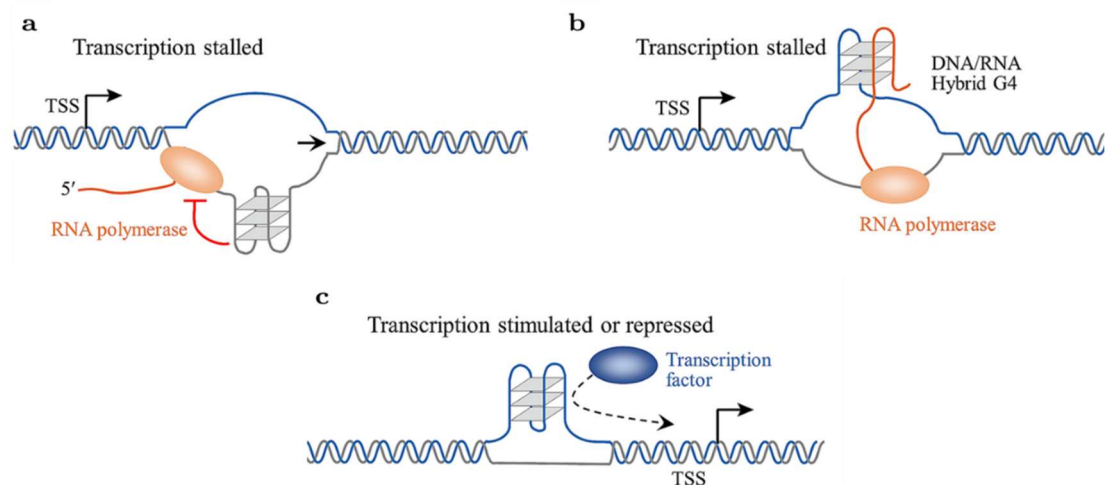


Figure 2.1.6. Possible G4 influence mechanism on transcription Adapted from Teng et al.¹⁵⁶

As reported above, G4s have been mapped and found to be enriched in cancer breakpoints and near the TSSs of highly transcribed genes in human cells, especially in oncogenes, such as c-MYC, BCL2, KRAS, c-KIT, and VEGF.¹⁵⁶

G4s found in TSSs play diverse roles during gene transcription (Figure 2.1.6c). A prominent example is the G4 in the promoter of the BCL2 gene, which encodes a protein that inhibits apoptosis and is overexpressed in various tumours and diseases.¹⁶⁸ Three promoters have been identified in BCL2: P1, P2, and M. The P1 promoter contains multiple G4 DNA sequences. Of these, Pu39 has six runs of guanine tracts and acts as a

transcriptional silencer during BCL2 expression.¹⁶⁹ Mutations in the Pu39 site have been shown to increase P1 activity, indicating its regulatory role. Several transcription factors can bind to Pu39 and relieve its inhibitory effect on BCL2 expression.^{170–172} Similar to Pu39, the G4 P1G4 and P32 also function as BCL2 transcription repressors. The regulation of G4s near the P1 promoter on BCL2 expression relies on the transcription factors recruited by these G4s. The interplay between different G4s and their cross-talk, especially for Pu39 and P1G4, is an area of further study.^{173,174}

Similar to BCL2, the oncogene KRAS also contains putative G4 sequences, each with distinct characteristics relative to the TSS. The interactions of these G4s with transcription factors influence KRAS transcription.^{175,176}

Another example is the G4 in the promoter of c-MYC, a significant driver of cancer.¹⁷⁷ The G4 sequence Pu46 contributes to most of c-MYC transcription and can form stable G4 structures with multiple topologies. It serves as both a suppressor of c-MYC expression and a recruiter of other transcription factors that regulate c-MYC transcription.^{178,179} NM23-H2, nucleolin, SP1, and CNBP are among the transcription factors that interact with Pu46. These factors either maintain the single-stranded form of Pu46 to activate c-MYC transcription or contribute to G4 formation and stabilize Pu46 to inhibit c-MYC expression.^{180–184}

G4 in telomeres

Telomeres are specialized DNA sequences, often complexed with proteins situated at the ends of linear chromosomes in eukaryotic organisms.¹⁸⁵ They serve to shield chromosome ends from homologous recombination, prevent recognition as double-strand breaks, and offer a solution to the "end-replication problem".¹⁸⁶

The structure of telomeric DNA typically consists of repetitive sequences rich in G and varies between organisms. In humans, it comprises a 10-15 kilobase double-stranded region and a 150-250 nucleotide single-stranded 3'-overhang which in humans is composed by d(TTAGGG) repeats.¹⁸⁷ This 3'-overhang plays a vital role by recruiting specific proteins called shelterin that protect telomeres and serving as a primer for telomerase-mediated extension. It can also form alternative structures such as G4 and T-loops, with the latter being proposed as a primordial mechanism for chromosome-end protection.^{188,189}

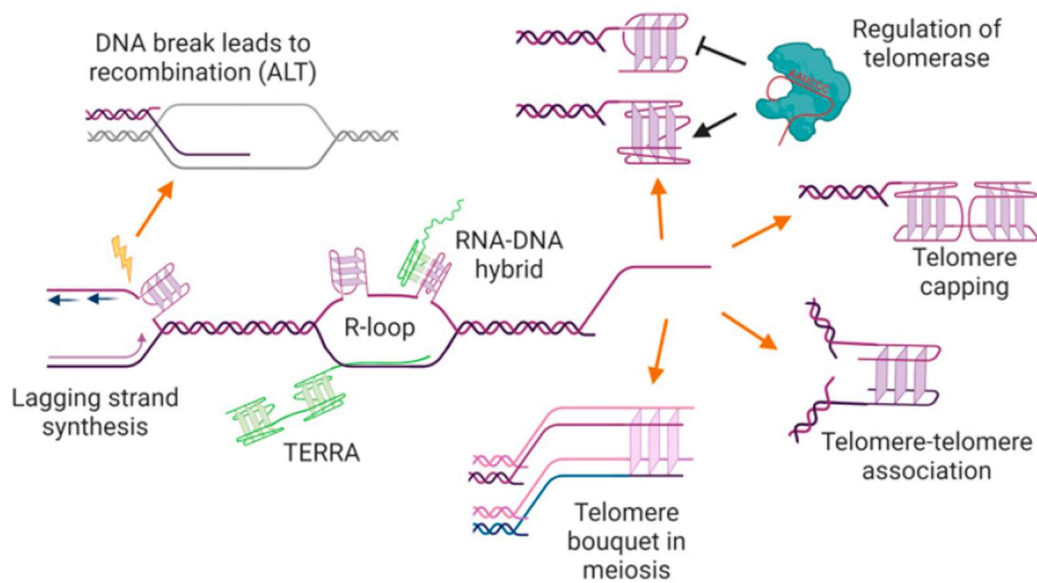


Figure 2.1.7. Influence of telomeric G4 formation on various mechanism.¹⁹⁰

G4 formation is demonstrated to play crucial role on different mechanisms involving telomeres. (Figure 2.1.7) At first, G4 structures were demonstrated to serve a crucial role as a protective cap for telomeres,¹⁹¹ preventing the degradation of the 3'-tail single-stranded section by nucleases. This function ensures the overall integrity of the telomere. However, it's important to note that G4 structures could also pose a potential threat to telomere integrity. If found within the double-stranded repetitive segment of the telomere, G4 can potentially impede DNA replication carried out by DNA polymerase. In cases where specific G4-binding proteins like helicase WRN,¹⁹² Pif1,¹⁹³ DNA2 nuclease,¹⁹⁴ and EXO1¹⁹⁵ are lacking, replication at the telomeres might slow down or even halt altogether, leading to possible collapse. Consequently, G4 structures play a dual role in the maintenance of telomere integrity. Moreover, G4 structures exert influence on the activity of telomerase.¹⁵⁶

Telomerase is the enzyme responsible for adding DNA repeats to the 3'-overhang of telomeres, effectively addressing the "end-replication problem." It is composed of TERT (telomerase reverse transcriptase) and TER (telomerase RNA) subunits, with their size and sequences varying among species.¹⁹⁶ The TERT subunit attaches to the single-stranded telomeric DNA and is recruited to the telomeres by TPP1. Telomerase helps elongate telomeres in a processive manner, with the RNA template portion pairing with single-stranded telomeric DNA. After adding one telomeric repeat, the overhang dissociates briefly from the RNA template before base pairing to add another repeat, and

so on. The precise mechanism of re-aligning the DNA overhang with the RNA template is not entirely clear. Telomerase processivity can vary under different conditions.¹⁹⁷

Telomeres play a crucial role in determining the replicative potential of a cell. In normal human cells, telomeres shorten with each division, ultimately reaching a certain size that according to the Hayflick Limit lead to cellular senescence.¹⁹⁸ Conversely, cancer cells, which need to proliferate extensively, exhibit high levels of telomerase activity to maintain telomere length and evade this limitation in 80-90 % of human carcinomas.¹⁹⁹ Notably, some tumours utilize an alternative mechanism known as the "alternative lengthening of telomeres" (ALT) pathway, which relies on homologous recombination and a different type of telomeric overhang structure.²⁰⁰

Early research revealed that the folding of telomeric DNA into G4 impedes the elongation of telomeres in the ciliate *Oxytricha nova*.²⁰¹ Furthermore, antiparallel and hybrid G4 structures formed by telomeric overhangs hinder telomerase elongation.^{202–204} Notably, the POT1-TPP1 complex has the capability to bind to and unwind all types of telomeric G4 structures, including parallel, antiparallel, hybrid, or two contiguous quadruplexes, through a mandatory unfolding process.^{205,206} The replication protein A (RPA), situated at the telomere, functions as a telomere end-binding protein and facilitates the unwinding of G4 structures. Once the POT1-TPP1 complex or RPA disrupts telomeric G4 structures, telomerase can then effectively elongate the telomere.^{207,208}

G4 in RNA

RNA G4s have been a subject of extensive discussion and scrutiny since the early 1990s.²⁰⁹ Initially, their existence and roles in cellular processes were questioned. In 2016, Bartel and colleagues used reverse-transcriptase stop assays after DMS-treatment to propose a widespread unfolding of RNA G4s in eukaryotic cells and their gradual disappearance in prokaryotic cells.²¹⁰ Recently, concrete evidence for RNA G4 formation within cells has been provided by the Balasubramanian lab, employing G4 structure-specific antibodies.²¹¹

These RNA G4 structures are found in various RNA types, including mRNAs, especially in untranslated regions (UTRs), long non-coding RNAs (lncRNAs), and telomeres. Interestingly, telomeres, once considered transcriptionally silent, are now known to produce TERRA RNAs, transcribed by RNA polymerase II, involved in chromatin remodelling and the regulation of telomerase activity.^{212–214} TERRA transcripts, originating from the C-rich strand of telomeric DNA, consist of tandem repeats of

r(UUAGGG) with variable lengths and exhibit G4 structures both *in vitro* and *in vivo*.^{215–218} TERRA acts as a potent ligand for telomerase but not as its substrate, inhibiting telomerase activity and participating in the regulation of telomere length.²¹⁹ There is also evidence suggesting that TERRA can directly interact with telomeres, forming DNA/RNA hybrid G4 structures that play a role in telomere end protection.^{220–222} Decreased TERRA levels have been observed in various telomerase-positive cancer types, implicating TERRA in age-associated neurodegenerative diseases such as Alzheimer's, Parkinson's, and amyotrophic lateral sclerosis (ALS).^{223–225} Moreover, TERRA plays a role in regulating telomeric heterochromatinization and development.²²⁶

RNA G4s, as well as G-rich regions with PQS, are known to exert specific regulatory functions, particularly in modulating translational efficiency. In cap-dependent translation initiation, the 40S ribosomal subunit scans mRNA from the 5'-cap towards the 3' direction, determining the TSS.²²⁷ G4 formed within the 5'-UTR downregulate translation. The position and thermal stability of these G4 structures influence the degree of translation inhibition. G4s located more than 50 nucleotides from the 5'-cap or with low thermal stability tend to have a limited regulatory effect.²²⁸ On the other hand, in cap-independent translation initiation, G4s can promote translation when situated close to the internal ribosome entry site (IRES), as seen in FGF2 or hVEGF.^{229,230} The IRES serves as an alternative ribosome recognition site and is utilized under various cellular conditions. Additionally, G4 structures in the 3'-UTR can influence translation efficiency. The exact mechanism remains poorly understood, but it is thought to involve mRNA maturation, localization, and interactions with auxiliary proteins, potentially linked to polyadenylation and alternative poly(A) sites.^{231–233} Polyadenylation affects mRNA stability, export, and translation by shortening the 3'-UTR, and G4 structures have been associated with increased polyadenylation efficiency and the usage of alternative poly(A) sites.^{234,235}

G4s also have indirect effects on translational efficiency, impacting mRNA maturation, localization, and interactions with other molecules.²³⁶ They can influence the localization of lncRNAs²³⁷ and microRNAs (miRNAs) in regulatory processes.²³⁸

Furthermore, G4 have been identified not only in the human genome but also in various non-human organisms, including human pathogens,²³⁹ including viruses^{240,241} like human immunodeficiency virus I (HIV),^{242–244} hepatitis C virus (HCV),²⁴⁵ Zika,²⁴⁶ Ebola,²⁴⁷ and SARS-CoV-2.²⁴⁸ Recently, researchers have even found G4 in plant genomes.²⁴⁹

Interestingly, while DNA G4 exhibit a diversity of folding topologies, RNA G4 tend to adopt an all-parallel folding with propeller loops, irrespective of the flanking

nucleotides and the experimental conditions.²⁵⁰ RNA G4s are thermodynamically more stable than DNA G4s,²⁵¹ and the structural rationale for this stability has been proposed by Neidle and co-workers on the basis of X-ray data and molecular modelling.²⁵² Their stability is attributed to interactions involving the 2'-hydroxyl group, phosphate, backbone oxygen atoms, and polar groups from the bases, which create a strong hydrogen bonding network within the structure, reducing the number of structural water molecules. These extensive intramolecular contacts confer stability to the overall structure.²⁵³

2.1.4 Targeting G-quadruplex with small molecules

Recognition of the biological significance of G4 has driven the research and development of small molecule ligands that interact with G4. The discovery of G4 structures in critical genomic regions, such as human telomeres and oncogene promoters, has provided a unique opportunity for targeted cancer-specific drug development.^{254–259} In 1997, the potential therapeutic application of targeting telomeric G4 to inhibit telomerase was first reported and has since been actively pursued.^{255–258,260} G4-ligands have also been found to hinder the ALT pathway, responsible for maintaining telomere stability in about 15% of cancer cells.^{261–264}

The regulatory potential of G4 on cancer cell growth is further substantiated by their possible presence in the promoter regions of various human genes, including retinoblastoma susceptibility,²⁶⁵ insulin,²⁶⁶ muscle-specific,²⁶⁷ vascular endothelial growth factor,²⁶⁸ hypoxia-inducible factor 1 α ,²⁶⁹ fragile X mental retardation genes,^{270–272} and oncogenes like c-MYC,^{179,273–275} K-RAS,¹⁷⁶ BCL2,^{276–278} c-KIT,^{279,280} or RET oncogenes.²⁸¹ As a result, there is ongoing research exploring novel anti-cancer therapeutic strategies with G4-DNA at their core.²⁸²

To date, there are over 1000 compounds which showed interesting G4-binding activity, and more information on them can be found in the website www.g4ldb.org.

The development of small molecules aimed at targeting G4 has been actively pursued through traditional experimental and computational screening methods, as well as rational drug design approaches. These efforts have been crucial in designing G4-specific ligands.^{283,284} Additionally, the structural information obtained from complexes formed between G4s and ligands has significantly contributed to our understanding of how small molecules recognize G4 structures, guiding the design of G4-specific compounds.^{285,286} This includes the determination of several NMR solution structures of intramolecular G4-ligand complexes, such as those involving the c-MYC^{283,287–289} and the telomeric G4,^{284,290–}

²⁹² as well as X-ray crystallographic structures of both intramolecular and intermolecular complexes formed between ligands and target.^{286,293–299}

The telomeric G4 ligand known as PIPER (N,N'-bis(2-(1-piperidino)ethyl)-3,4,9,10-perylenetetracarboxylic acid diimide), depicted in Figure 2.1.7, which has a wide hydrophobic core and two amine extensions, was initially discovered by Fedoroff and colleagues.³⁰⁰ Interestingly, Liu and co-workers had previously explored this molecule as a selective ligand for both G and C-rich DNA.³⁰¹ When interacting with the intermolecular G4 motif derived from the [dTTAGGGTT]₄ sequence, it was observed that a 1:1 complex formed initially. As more ligand was added, a 2:1 ligand to G4 ratio was reached, indicating the existence of two potential binding sites with slightly different affinities. Notably, a stable tail-to-tail G4-ligand-G4 complex was formed using the [dTTAGGG]₄ sequence. Moreover, when employing the potentially more biologically relevant [dTAGGGTTA]₄ sequence, PIPER exhibited an alternative binding mode characterized by threading intercalation at the GT step. Subsequent evaluation of PIPER's biochemical effects revealed its ability to inhibit helicase Sgs1-mediated G4 unfolding, suggesting a broader mechanism for G4 ligands.³⁰²

In 2002, researchers identified a small molecule called TMPyP4 (depicted in Figure 2.1.7), which stabilizes the G4 structure formed in the c-myc promoter. This discovery led to the inhibition of c-myc expression, highlighting the therapeutic potential of targeting promoter G4 for transcriptional regulation.^{179,303} TMPyP4, a porphyrin derivative (5,10,15,20-tetra-(N-methyl-4-pyridyl)porphine), developed by Wheelhouse and colleagues, is renowned as one of the most well-known G4 ligand.³⁰⁴ This molecule was deliberately designed to possess the appropriate shape and dimensions required to stack on top of a G4.

Initially, the cationic porphyrin TMPyP4 was shown to have the capability to bind to and stabilize G4 within human telomere sequences, consequently inhibiting telomerase activity. To gain a deeper understanding of the mechanism underlying telomerase inhibition by TMPyP4, researchers conducted a cDNA microarray analysis on cells treated with TMPyP4 and TMPyP2, a positional isomer of TMPyP4 with low affinity for G4 which is also depicted in Figure 2.1.7. The analysis of time course data from the microarray experiments revealed that TMPyP4 and TMPyP2 treatments led to alterations in the expression of several gene clusters. Among these, the c-myc oncogene, was specifically down-regulated by TMPyP4 but not by TMPyP2. The hTERT gene, responsible for encoding the catalytic subunit of telomerase, is transcriptionally regulated

by c-MYC, and TMPyP4 was found to decrease the transcripts of human telomerase reverse transcriptase. This suggests two possible mechanisms for the impact of TMPyP4 on telomerase activity. Moreover, TMPyP4, but not TMPyP2, demonstrated the ability to prolong survival and reduce tumour growth rates in two xenograft tumour models.³⁰⁵

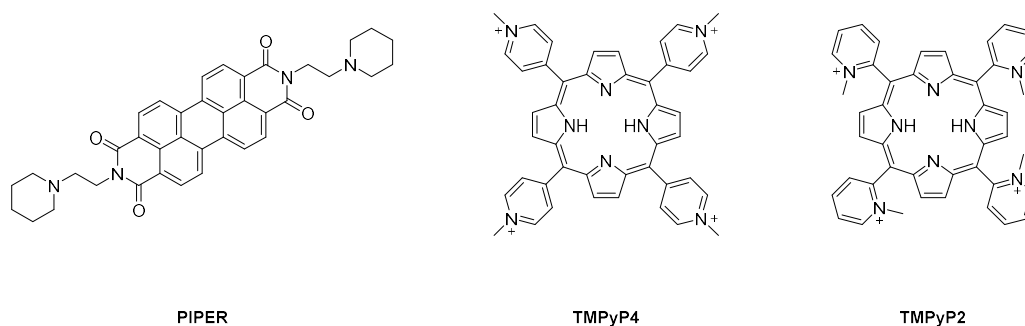


Figure 2.1.7. Chemical structure of PIPER, TMPyP4 and TMPyP2

Various categories of compounds, including quindolines and ellipticines, have been documented to hinder the transcription of the c-MYC gene by stabilizing the G4 structure within the c-MYC promoter.^{306–309} Subsequently, other compounds that stabilize G4 in gene promoters were found to repress the transcription of different oncogenes.

McLuckie and colleagues, employing a functional cell-based assay, identified two novel G4-ligands, specifically two Benzo[*a*]phenoxazines (depicted in Figure 2.1.8). These compounds demonstrated a reduction in the transcription of a luciferase reporter gene under the control of the G4-containing c-KIT gene promoter. Additionally, they showed that these molecules led to decreased expression of the endogenous c-KIT gene in a human gastric carcinoma cell line. Biophysical analysis, using surface plasmon resonance, revealed that these ligands exhibited a strong preference for binding to one of the two G4 sequences within the c-kit promoter over dsDNA.³¹⁰



Figure 2.1.8. Benzo[*a*]phenoxazine designed by McLuckie et al.³¹⁰

Wang et al. conducted a study demonstrating that the G4 structure within the human BCL2 gene was disrupted when a partial mutation changing G to A was introduced. This mutation resulted in a twofold increase in the basal transcriptional activity of the BCL2 gene promoter. Indolquinoline derivatives (depicted in Figure 2.1.9), which are highly active G4-ligands developed by their research group, were found to significantly suppress

the activation of BCL2 transcription but was less effective on the transcription of the mutated BCL2 gene. These findings provided direct evidence that the G4 structure formed in the BCL2 promoter region could function as a transcriptional repressor element, and G4-specific ligands could regulate the transcription of BCL2 by stabilizing the quadruplex structure. Moreover, the results indicated that quindoline derivatives could induce apoptosis in HL-60 tumour cells.³¹¹

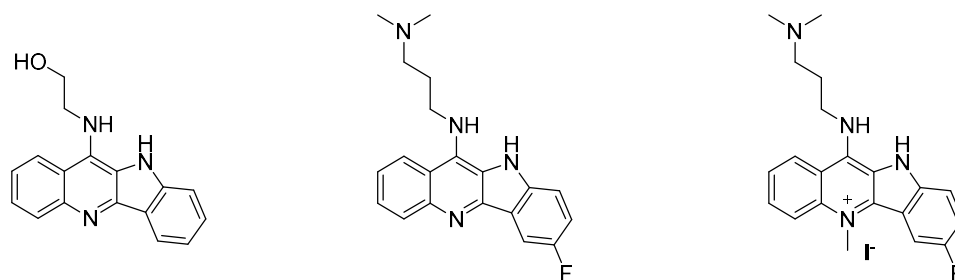


Figure 2.1.9. Quindoline derivatives synthesized by Wang et al.³¹¹

In 2013, Lavrado and colleagues reported the synthesis of 20 compounds derived from the same scaffold (named indolo[3,2-*b*]quinolines (IQb)) with mono-, di-, and trisubstituted basic side chains. The chemical structure of one of these ligands is displayed in Figure 2.1.10. Molecular modelling studies were also conducted to understand the binding orientations of these ligands with G4. The results suggested the thermal stability of different G4 DNA sequences increased upon ligand binding when they had an N5-methyl or a 7-carboxylate group and propylamine side chains. Selectivity between G4 and dsDNA appeared to be influenced by the number and relative positions of basic side chains. Among all the IQb derivatives studied, the novel trisubstituted compounds which featured a 7-(aminoalkyl)carboxylate side chain, stood out as the most promising compounds. They exhibited high G4 thermal stabilization with a preference for Hsp90A, KRas21R, and F21T, a 10-fold selectivity for G4 over dsDNA, and a 100-fold selectivity for the HCT116 cancer cell line over primary rat hepatocytes. These compounds led to decreased protein expression levels of Hsp90 and k-ras in HCT116 cancer cells.³¹²

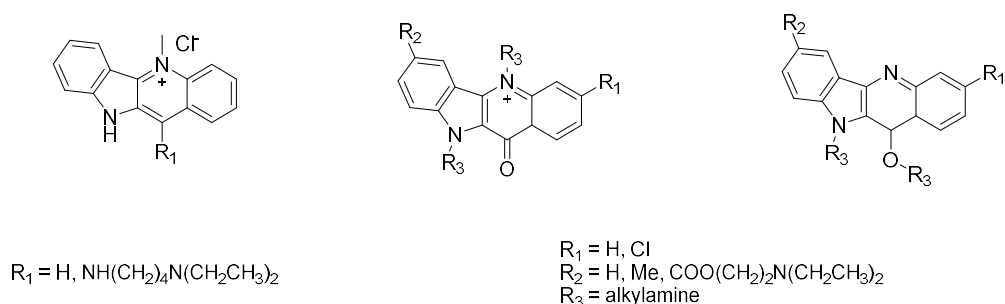


Figure 2.1.10. Indolo[3,2-*b*]quinoline derivatives designed by Lavrado and colleagues.³¹²

In 2015, the same research team proposed the synthesis of novel derivatives of indolo[3,2-*c*]quinolines (IQc) designed to target DNA and RNA G4 within the promoter and 5'-UTR mRNA of the *k-ras* gene, which are depicted in Figure 2.1.11. Biophysical experiments demonstrated that di-substituted IQc compounds effectively stabilized *k-ras* G4 and selectively inhibited the proliferation of *k-ras* mutant cancer cell lines ($0.22 < IC_{50} < 4.80 \mu M$). These compounds also reduced *k-ras* promoter activity in a luciferase reporter assay and lowered both *k-ras* mRNA and p21KRAS levels in mutant *k-ras* colon cancer cell lines. Furthermore, IQcs induced apoptosis in cancer cells, in part due to their capacity to repress KRAS gene expression.³¹³

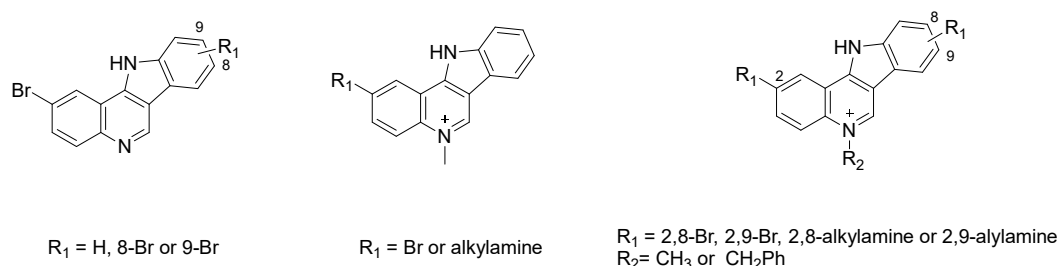


Figure 2.1.11. Indolo[3,2-*c*]quinoline derivatives designed by Lavrado and colleagues.³¹³

More recently, it has been observed that compounds known to stabilize G4 structures can induce damage to DNA, leading to genomic instability. Moreover, they exhibit a synergistic effect when combined with inhibitors or the deficiency of DNA repair mechanisms.^{192,314–317}

Salvati and their colleagues reported that the stabilization of G4 DNA by RHPS4 (3,11-difluoro-6,8,13-trimethyl-8H-quinolo[4,3,2-*kl*]acridinium methosulfate), depicted in Figure 2.1.12, results in the uncapping of telomeres. This, in turn, triggers the specific recruitment and activation of poly-adenosine diphosphate (ADP) ribose polymerase I (PARP1) at the telomeres. PARP1 forms multiple ADP-ribose polymers that co-localize with the telomeric repeat binding factor 1 protein and are inhibited by selective PARP(s) inhibitors or PARP1-specific small interfering RNAs. Knocking down PARP1 prevents the repair of RHPS4-induced telomere DNA breaks, leading to an increase in chromosome abnormalities and ultimately inhibiting the growth of tumour cells, both *in vitro* and in xenograft animal models. Remarkably, when combined with a Topoisomerase I inhibitor (a class of compounds that acts as anticancer agent interfering with DNA replication and therefore leading to cancer cell death), this multi-component treatment strategy has proven to be highly effective, resulting in complete regression of the tumour, significantly improved overall survival, and curing mice, even when treatments are

initiated at a very late stage of tumour growth. This research unveiled an uncharted connection between PARP1 and G4-ligands, demonstrating the excellent efficacy of a multi-component strategy that involves the use of PARP inhibitors in telomere-based therapy.³¹⁵

Pyridostatin (PDS) has also emerged as a promising platform for developing ligands aimed at targeting G4 as a therapeutic approach to inhibit cancer growth. The chemical structure of PDS is depicted in Figure 2.1.12. In 2013, the research groups of Balasubramanian and Jackson demonstrated that PDS induces growth arrest in human cancer cells by causing replication- and transcription-dependent DNA damage. A genome-wide analysis using chromatin immunoprecipitation sequencing of the DNA damage marker γ H2AX revealed the distribution of sites where PDS induces damage, highlighting its affinity for regions with a propensity for G4 formation. This led to the modulation of gene expression, including the proto-oncogene SRC. Notably, PDS reduced the abundance of SRC protein and suppressed SRC-dependent cellular motility in human breast cancer cells, validating SRC as a target of this drug. This unbiased approach to defining the sites of drug action within the genome sets the stage for discovering functional interactions between DNA and drugs.³¹⁶

Subsequently, McLuckie and colleagues reported that PDS stabilizes G4s within cells, provoking a DNA damage response characterized by the formation of DNA DSBs. Cell death caused by ligand-induced G4 stabilization can be enhanced in cells deficient in DNA damage repair genes. Their findings demonstrated that PDS acts synergistically with NU7441, an inhibitor of the DNA-PK kinase vital for nonhomologous end joining repair of DNA DSBs, as well as in BRCA2-deficient cells that exhibit genetic impairments in homologous recombination-mediated DSB repair. These G4-targeting ligands have a promising potential as cancer therapeutic agents, particularly when combined with the inhibition or mutation of DNA damage repair mechanisms.³¹⁷

In the context of BRCA-deficient cancers, Zimmer and colleagues illustrated that homologous recombination (HR) plays a significant role in telomere replication, stabilizing and resuming stalled replication, and repairing DSBs. They found that cells lacking HR exhibited a substantial reduction in the efficiency of replicating G-rich telomeric repeats. Furthermore, treatment with the G4-stabilizing compound PDS heightened the fragility of telomeres in BRCA2-deficient cells, indicating that G4 formation contributes to telomere instability. Significantly, PDS reduced the proliferation of HR-deficient cells by inducing DSB accumulation, activating checkpoints, deregulating

G2/M progression, and exacerbating the intrinsic replication defect associated with HR deficiency. The toxicity of PDS extended to HR-deficient cells that had developed resistance to Olaparib through the loss of 53BP1 or REV7. These findings underscored the therapeutic potential of G4-stabilizing drugs in selectively eliminating HR-compromised cells and tumors, including those that had become resistant to PARP inhibition.³¹⁸

Quarfloxin (CX-3543), an innovative drug, entered phase II clinical trials (NCT00780663) as a prospective treatment for neuro-endocrine carcinomas. This compound originally belonged to the fluoroquinolone class of compounds, known for their dual interaction with topoisomerase II and G4s. Cylene Pharmaceuticals engineered Quarfloxin (Figure 2.1.12) by optimizing fluoroquinolone-based structures to create a superior compound with potent anticancer properties, while eliminating any residual topoisomerase II inhibitory effects. Initially considered a selective binder for the myc G4 structure, subsequent investigations revealed that Quarfloxin predominantly targets G4 structures within ribosomal DNA (rDNA) in the nucleolus. This results in the inhibition of RNA polymerase I transcription. Notably, Quarfloxin induces apoptosis and significantly reduces tumour volume in pancreatic cancer xenograft models, such as MDA-MB-231 and MIA-PACA-2. Despite its initial promise, Quarfloxin was later withdrawn from clinical studies, leading to the discontinuation of its development.³¹⁹

Following this, Cylene Pharmaceuticals developed a derivative of Quarfloxin, known as CX-5461 (Figure 2.1.12). CX-5461 has demonstrated the induction of autophagy, but not apoptotic cell death, in cancer cells. When administered orally in xenograft models of pancreatic carcinoma and melanoma, CX-5461 exhibited *in vivo* antitumor activity. Currently, CX-5461 is undergoing advanced phase I clinical trials for patients with BRCA1/2-deficient tumors (NCT02719977). A recent study has shed light on the novel mechanism of action of CX-5461, showing that it blocks replication forks and causes DNA damage through the formation of single-stranded DNA gaps/breaks. Repairing these DNA lesions requires the involvement of the BRCA and NHEJ pathways. Moreover, immunofluorescence experiments using a G4-specific antibody indicated that CX-5461 primarily induces DNA damage in G4-enriched genomic sequences. Biochemical assays further demonstrated that CX-5461 specifically binds to and stabilizes these G4 structures, including those derived from myc, c-kit, and telomeres, showing a preference for G4 over ds DNA. Additionally, CX-5461 inhibits ribosomal RNA biogenesis, highlighting a multifaceted mechanism of action.³²⁰

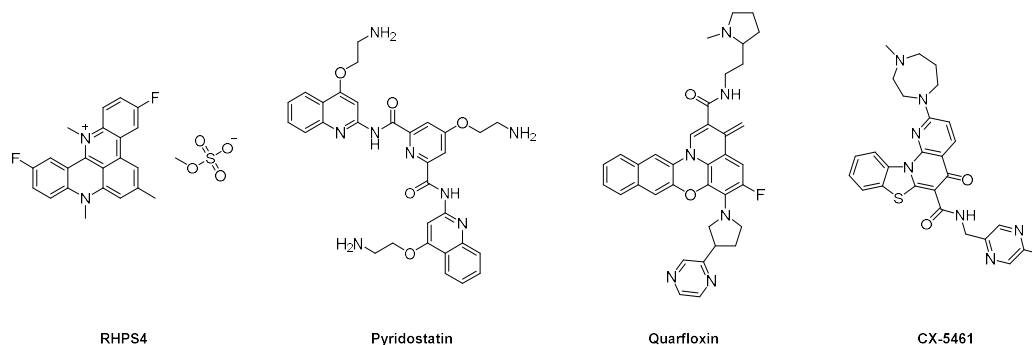


Figure 2.1.12. Chemical structure of RHSP4, Pyridostatin (PDS), Quarfloxin and CX-5461

A Japanese research group, led by Haruo Seto, uncovered a natural product known as Telomestatin, derived from *Streptomyces anulatus*. This compound emerged as a novel ligand for telomeric G4 and a telomerase inhibitor (Figure 2.1.13). In Telomerase Repeated Amplification Protocol (TRAP) assays (a common technique to determine telomerase activity in mammalian cells and tissue samples), Telomestatin exhibited a remarkable ability to inhibit telomerase activity with an IC_{50} value of 5 nM, while showing no activity against DNA polymerases such as Taq polymerase.³²¹ The structure of Telomestatin is a neutral polycyclic compound consisting of five oxazoles, two methyloxazoles, and one thiazoline ring. It is characterized by high hydrophobicity, nearly complete flatness, and an ideal size and shape that enables effective stacking on top of G-quartets. Moreover, the structural resemblance between Telomestatin and a G-tetrad suggests that its telomerase inhibition may result from its ability to either facilitate the formation of G4 structures or trap preformed G4, thus sequestering single-stranded d[TTAGGG]_n primer molecules essential for telomerase activity.

Further structural studies on Telomestatin were conducted, revealing that it promotes or stabilizes the formation of intramolecular G4, particularly at high concentrations. A simulated annealing molecular docking protocol generated models of four different Telomestatin-G4 structures, including two potential 1:1 end-stacked complexes, a 2:1 complex, and a 1:1 complex in which Telomestatin was intercalated between two G-quartets.³²² The ability of Telomestatin to stabilize telomeric G4 arrangements was assessed using a FRET assay, indicating strong stabilization effects on G4 in the presence of 100 mM Na⁺ or 10 mM K⁺. A competition experiment with double-stranded DNA under the same conditions showcased the complete selectivity of Telomestatin for G4s over dsDNA.³²³ In addition to its high affinity for telomeric overhangs, Telomestatin can disrupt telomere-specific binding proteins (TRF2 and POT1) involved in the telomere capping process. As a result, apoptosis occurs more rapidly compared to simple

telomerase inhibitors. Therefore, Telomestatin was categorized as a "telomere disrupting agent". It has proven to be an effective anti-proliferative agent, with IC_{50} values ranging from 0.1 to 5 μM in various cancer cell lines, while showing no significant effects on healthy cells.³¹⁹ Unfortunately, Telomestatin faces challenges, primarily its limited availability in large quantities through total synthesis or natural extraction. Its high hydrophobicity and poor water solubility also present obstacles to its use as a drug.³²²

In 2001, Neidle and his research team introduced a well-known G4 ligand, BRACO-19, featuring 3,6-bis(aminoalkyl) substituted acridine (Figure 2.1.13). This compound was meticulously designed through computer modelling to engage with the grooves of G4 structures while also having the capacity to stack atop a G-tetrad. BRACO-19 demonstrated remarkable inhibitory activity in the TRAP assay, displaying a 90 nM IC_{50} , and exhibited moderate cytotoxicity towards the uterine carcinoma cell line UXF1138L, with an IC_{50} of 2.5 μM . Moreover, it displayed IC_{50} values of 10.0, 10.1, and 13.0 μM in the telomerase-positive human ovarian carcinoma cell lines A2780, CH1, and SKOV-3, respectively. Its capability to stabilize the G4 arrangement resulted in a remarkable increase of 27.5 °C at a concentration of 1 μM .³²⁴⁻³²⁷

Bisquinolinium compounds have emerged as highly selective ligands for G4s. In 2007, Monchaud introduced PhenDC3, a 6,6'-disubstituted-2,2'-bipyridine and 2,9-disubstituted-1,10-phenanthroline conjugate (Figure 2.1.13). This compound's conformation is locked through the formation of intramolecular hydrogen bonding, resulting in excellent stabilization of the human telomeric F21T G4 ($\Delta T_m = 29.7$ °C), even in the presence of dsDNA. Subsequently, several 1,10-phenanthrolines, conjugated with benzimidazole moieties, were synthesized to create a ligand capable of distinguishing G4 present in the promoter regions of oncogenes, from telomeric G4. Among these compounds, Phen-Pr (Figure 2.1.13) exhibited robust stabilization of a c-MYC G4, with a ΔT_m of 20 °C. More recently, the aromatic scaffold has been further modified at positions 2 and 9 of the heterocycle, resulting in bis-triazoles and bisoxazoles derivatives. (Figure 2.1.13) Specifically, the former displayed moderate stabilization of c-kit and c-myc G4, while the latter demonstrated limited affinity for various G4 structures.³¹⁹

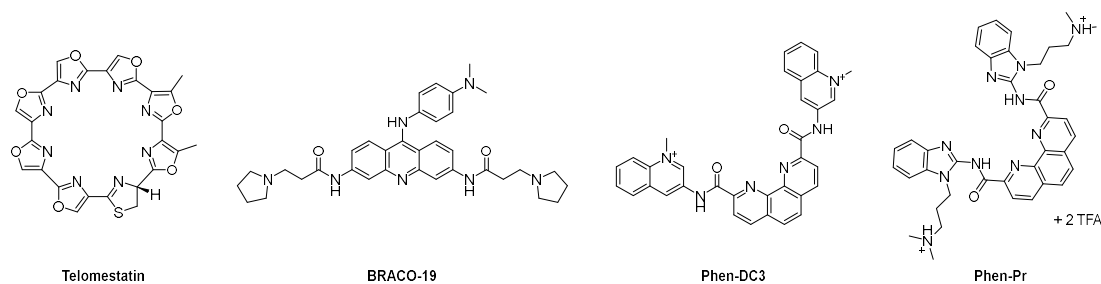


Figure 2.1.13. Chemical structure of Telomestatin, BRACO-19, Phen-DC3 and Phen-Pr.

G4-ligand selectivity at several levels

Most of G4-binding ligands share three major structural aspects, which enhance G4 affinity: an extended heteroaromatic chromophore, needed for the pi-stacking interaction with the G-quartet; extended side chains which develop into the grooves and loops of G4s, improving stability through van der Waals and electrostatic interactions; a positive charge, normally at the end of the side chain but also on the chromophore itself, which can improve the solubility of the compound in physiological conditions and strengthen the pi-stacking interaction with the G-quartet.²⁵⁹

pi-stacking occurs between the aromatic rings of the ligand and the guanine aromatic units, specifically the pyrimidine and imidazole rings.³²⁸ These interactions are electrostatic in nature and result from the spatial orientation of two similar quadrupoles. They typically manifest as edge-to-face or off-centred parallel stacking rather than face-centred parallel stacking. Electrostatic interactions involve the sugar-phosphate backbone of the G4's grooves, and they enhance the overall binding energy of a ligand to the G4. These interactions typically occur between negatively charged phosphate ions in nucleotides and positively charged moieties on the ligands, often amines. While ionic interactions favour G4 binding, they lack specificity for the G4 motif and can also increase affinity for the canonical DNA arrangement with negatively charged major and minor grooves.^{329–333}

One of the major challenges when developing DNA-binding molecules is obtaining a selectivity both versus different sequences of G4 within the genome and versus dsDNA. Although there is strong evidence that some G4s have sufficient diversity features, G4 arrangements share some structural homology that make this intent more difficult. Moreover, simultaneously targeting more oncogenes participating in several distinct signalling pathways can be therapeutically advantageous when considering large-scale profiles of genetic mutations and heterogeneity present in many human tumours. In contrast the selectivity towards G4 versus dsDNA is essential in order to avoid or at least reduce the side effects that lead to cellular toxicity. Generally, a compound is considered

highly selective when its binding energy with G4 is on the order of 10^7 M^{-1} while the binding energy with dsDNA is at least two orders of magnitude smaller.³³⁴ As mentioned earlier most of small molecules interact with G4 through quasi-external stacking, with the heteroaromatic chromophore pi-stacked onto the face of an external G-quartet and the side chains extended in the grooves. G4 versus duplex selectivity can be explained, at least in part, by the difference between the large, highly accessible surface area of a terminal quartet compared with the much smaller, less accessible A·T or G·C base pair surface area of a typical duplex DNA intercalation site.²⁵⁹

Furthermore, the objective of many studies is to target specific G4 structures among several options. However, due to the limited characterization of G4s in the human genome (less than 1%), complete G4 selectivity is challenging. Typically, studies screen a small-molecule library (4-6 members), with a small panel of well-studied G4s, occasionally revealing selectivity between some of them. Nevertheless, complete specificity for a single G4 structure is rarely found. An example can be found in the binding studies performed with the alkaloid chelerythrine, which was found to bind to three promoter G4s with similar efficiency.³³⁵ The challenge arises from the structural similarity of G4s and the architectural features shared among most G4s. Nevertheless, it's important to note that these variations are usually of lesser significance when compared to the overarching structural similarities found in the majority of G4.²⁵⁹

Finally, achieving cellular selectivity, as measured by differences in cell growth, viability, or death when treating cancer cell lines compared to normal cells, is a desirable goal. However, attributing these effects to promoter G4 binding remains complex. Telomeric G4s are exceptions, as small molecules binding to them are known to inhibit telomerase or to induce DNA damage repair in cells, with selective lethality demonstrated in cancer cells.^{315,316,336} Assessing whether a specific gene and its promoter G4 are a legitimate target can be done using appropriate assays, for instance for the compound (2-(4-(10H indolo[3,2 b]quinolin 11 yl) piperazin 1 yl)-N,N dimethylethanamine) that was found to effectively bind to the myc G4 *in vitro*, leading to myc downregulation and cell-growth inhibition. However, the myc G4 targeting wasn't directly linked to its downregulation.³³⁷ The influence of small molecules on 'G4 target' genes might therefore, be indirect and depends on the compound's nature and cell type.

These various facets of G4 ligand selectivity contribute to our understanding of how these ligands interact with G4 structures, impacting their application in cellular and therapeutic contexts.²⁵⁹

2.2 Methods to study G4-ligand interactions

2.2.1 ESI-MS for the study of G-quadruplex:ligand interaction

As previously discussed, MS functions by converting molecules into ions. In the preceding section within Chapter 1.2, the various analysers capable of achieving ionization have been extensively covered, all of which are applicable to nucleic acid analysis. However, among these methods, ESI has emerged as the preferred ion source due to its “soft” ionization process.^{338–341}

ESI has found widespread use in analysing a diverse range of biomolecules, including peptides, proteins, and nucleic acids. The earliest reports on detecting intact dsDNA by ESI-MS date back to the early days of native mass spectrometry.^{342–346}

As mentioned earlier, ESI-MS has garnered its reputation primarily because of its “soft” ionization method, characterized by minimal fragmentation. Consequently, non-covalent interactions, such as those found in ligand-target complexes, remain unperturbed during the ionization process.³⁴⁷ ESI-MS is renowned for its sensitivity and demands relatively small sample volumes. In standard setups, as little as 20–100 μL of a sample with nucleic acid concentrations ranging from 1–10 μM is sufficient for routine analysis. An example of an ESI-MS spectrum is depicted in Figure 2.2.1.

In physiological conditions, nucleic acids tend to carry an excess of negative charges on their phosphate groups (their $\text{pK}_a < 1$), while the bases are generally in a neutral state at physiological pH. As a result, the analysis of nucleic acids is typically more sensitive when conducted in the negative ion mode. Few research groups have explored the impact of both ionization modes on the detection of nucleic acid complexes.^{348–350}

In the case of noncovalent complexes, employing positive ionization mode for analysis posed certain challenges.³⁴⁹ For instance, complexes involving intercalating agents could not be reliably detected in the positive ion mode, despite being detectable in negative ionization mode where they exist in the solution and complexes featuring minor groove binders could be observed in the positive ion mode, but their relative intensities did not accurately mirror their abundance in the solution. This discrepancy arises from the nature of the charge carriers in each ionization mode. In negative ionization mode, the sprayed ions are primarily charged through the phosphate groups, similar to their state in solution. Not all phosphate groups remain negatively charged, as many are neutralized by counterions. In contrast, in the positive ion mode, the phosphate groups need to be entirely neutralized, and in addition, some nucleobases must be protonated.

Consequently, the native noncovalent bonds within the complexes are more susceptible to disruption in the positive ion mode when compared to the negative ion mode.³⁵¹

Nucleic acids rely on an adequate ionic strength to overcome repulsion between phosphate groups and attain their folded structures. Typically, mammalian cells contain around 139 mM K⁺, 12 mM Na⁺, and 0.8 mM Mg²⁺ ions.³⁵² Unfortunately, these conditions are often incompatible with electrospray ionization, as alkali or alkaline-earth metals have a strong affinity for oxygen atoms, leading to the formation of non-volatile adducts that cannot be removed by increasing the internal energy of the system. In contrast, ammonium ions (NH₄⁺) are volatile; when the system's internal energy is elevated, a proton can transfer from ammonium to the biomolecule, releasing NH₃. Consequently, ammonium acetate is a common choice in native mass spectrometry, and its concentration is adjusted to achieve the desired ionic strength, with 150 mM NH₄OAc closely mimicking physiological conditions.³⁵¹

Furthermore, G4 sequences often adopt similar structures when interacting with either ammonium or potassium, which is a biologically relevant cation.³⁵³ Recent advances in experimental conditions allow for the study of G4s in an environment that closely resembles physiological conditions and is more biologically meaningful, using potassium as the cation for sequence stabilization.³⁵⁴ The strategy entails folding the G4 in a low concentration of potassium (1 mM), which doesn't interfere with the ionization process, while employing a volatile bulky buffer like triethylammonium acetate (TEAA) to provide sufficient ionic strength for efficient ionization. Notably, TEAA is too big to fit in the G4 cavities, therefore it cannot compete with K⁺ in the folding process.³⁵⁵

In some spectrometers, samples can be directly acquired from aqueous solutions, although the addition of 10-20% of an organic cosolvent, typically methanol or isopropanol, can enhance the sample's volatility and consequently improve the signal-to-noise ratio.^{354,356,357} A reliable procedure involves adding 15-20% methanol to the samples before infusion to reduce the surface tension of the droplets. Importantly, it has been confirmed through CD analysis that the addition of methanol does not induce conformational changes or alter peak ratios.³⁵⁸ However, higher concentrations of organic solvents, such as 50% methanol, can lead to structural modifications and even trigger a shift to a different G4 topology.³⁵⁹⁻³⁶¹

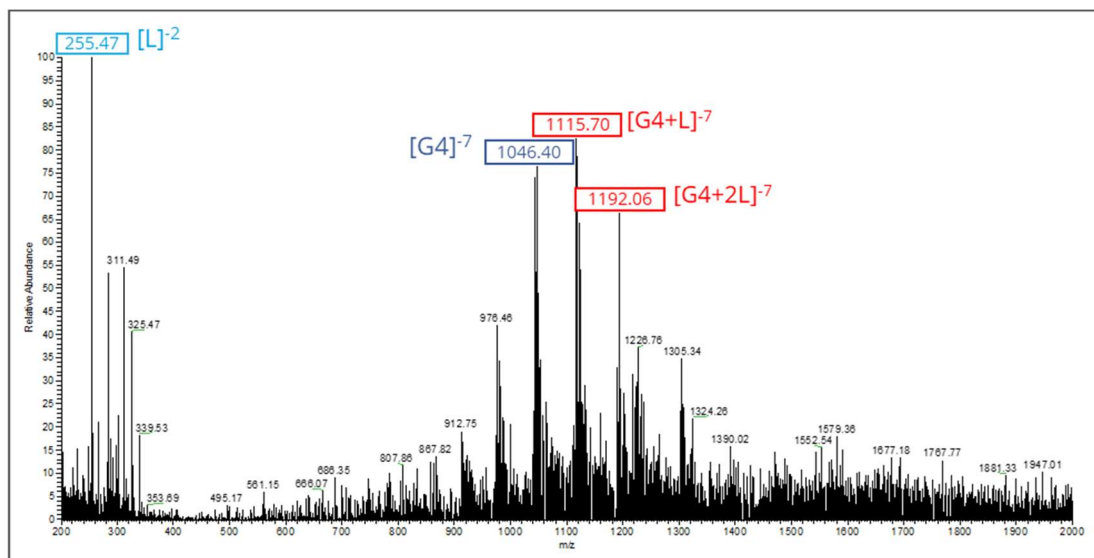


Figure 2.2.1. Example of MS spectrum of telomeric wtTel23a G4 in presence of a specific ligand. In red we can see the peaks corresponding to the complex, in blue and cyan the peaks relative to the free G4 and ligand, respectively.

In pharmacology, the effectiveness of a drug is assessed based on its ability to influence a biochemical process. This effectiveness is closely related to the fraction of the target molecules that form bound complexes with the drug. To compare the activity of different ligands targeting the same biomolecule, it's valuable to evaluate the fraction of the target molecules that form bound complexes at specific concentrations of both the total target and ligand.³⁵¹

To begin, the intensities of peaks in MS spectra are considered proportional to the concentrations of the corresponding species in the analysed solution. Therefore, these peak intensities can be used to calculate macroscopic equilibrium binding constants (K_D) and ligand binding affinity (BA). In this context, K_D is defined as the ratio of $[DNA][ligand]$ to $[1:1 \text{ complex}]$, taking into account the total amount of the 1:1 complex without specifying the number of ligand binding sites.³⁵³ However, it's essential to assume that both free and bound nucleic acid fractions exhibit the same instrumental response during ionization. This assumption implies that peak intensities accurately reflect their relative concentrations in the solution, which can be influenced by ionization efficiency and various instrumental factors. Typically, complexed and uncomplexed G4 species respond similarly to ESI, but it's necessary to sum the contributions of ion adduct peaks for each stoichiometry. Additionally, it's important to consider that species with similar m/z values have similar responses, especially when they share the same charge state. Therefore, a quantitative analysis is ideally conducted by considering the same charge state, and comparing assemblies of vastly different sizes should be avoided. It's

recommended to separately determine binding constants for each charge state and then calculate an average.³⁵⁸

Many research papers have employed K_A values determined from single mass spectra to rank ligands based on their affinities.^{349,358,362–366} ESI-MS titration, which involves collecting mass spectra at different ligand-to-nucleic acid (NA) ratios, can reveal stoichiometries that might not be apparent when observing a single concentration. A titration curve plots the concentration of the added ligand on the x-axis against the concentration of free nucleic acid and/or complexes on the y-axis.

ESI-MS is generally capable of determining association constants (K_A) with values ranging from 10^3 M^{-1} to 10^8 M^{-1} .^{364,367} The individual K_D (dissociation constant) values are then derived by numerically fitting the titration curves.^{368,369}

However, it's important to acknowledge that errors in the determination of binding constants can occur due to complex disruption during ionization or transit into the spectrometer, leading to underestimation, or due to the fragmentation of the unbound nucleic acid, resulting in overestimation.³⁵⁸

Furthermore, the intensities of the peaks representing different molecular species can be used to calculate BA, a semi-quantitative measure describing a molecule's inclination to bind with nucleic acid or a larger macromolecular target in general. This parameter is calculated as follows:

$$BA = \frac{\sum I_{G4 \text{ bound}}}{\sum I_{G4 \text{ unbound}} + \sum I_{G4 \text{ bound}}} \times 100^{338}$$

ESI-MS also serves as a valuable tool for visually assessing the relative affinity of a ligand for various nucleic acid sequences. In a typical "selectivity screening" arrangement, a test compound is simultaneously incubated with a G4 (or different G4 sequences), dsDNA, and a single-stranded sequence.

In addition to evaluating BA, the effectiveness of stabilizing the macromolecular structure can be determined through fragmentation techniques. The most common method employed is CID, which has already been described in the previous chapter. (Figure 2.2.2)

It's crucial to note that dissociation in a mass spectrometer is an irreversible process. Therefore, when ranking complexes in terms of their gas-phase kinetic stability, we refer to the dissociation kinetics rather than a dissociation equilibrium. This ranking is typically based on $E_{COM}^{50\%}$, expressed in electronvolts (eV), representing the centre-of-mass collision energy required to promote the dissociation of the complex to half of its original intensity.³⁵⁸

The dissociation reaction, measured by MS/MS, involves the ligand breaking away from the complex. The value of $E_{\text{COM}}^{50\%}$ can be directly calculated from the plot of the relative intensities of the complex that undergoes the dissociation and the applied collision energy converted with the opportune formula (that depends of the spectrometer) in eV. This plot is referred to as dissociation curve (depicted in the inset in Figure 2.2.2) and the relative intensity is determined using the formula:

$$\text{Relative Intensity \%} = \frac{I_{\text{complex}}}{I_{\text{complex}} + I_{\text{dissociation NA products}}} \times 100$$

This parameter is particularly relevant when assessing small molecules as G4 binders, since as previously stated, G4 stabilization is a critical property for triggering biological effects.

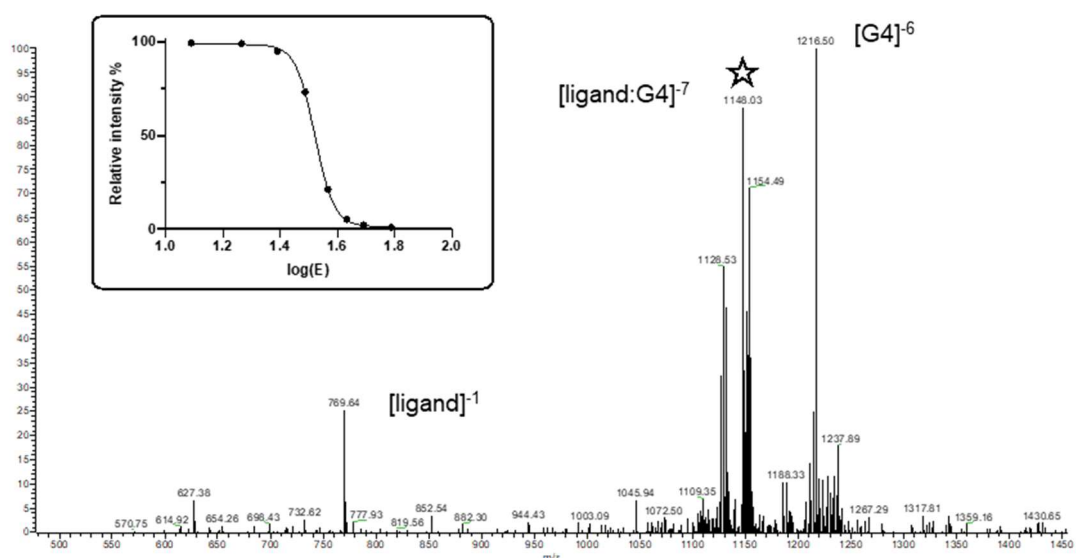


Figure 2.2.2. Example of CID study. A ligand-G4 complex (wtTel23a), marked with a star in the spectrum, undergoes fragmentation by loss of the ligand. In the inset, a dissociation curve as the plot of relative intensity of the complex against collision energy (eV) is depicted.

While ESI-MS provides mass-related information regarding complex stoichiometry and their relative intensities, it does not offer insights into the specific binding mode or interaction site. However, valuable information can be deduced from the MS spectrum, such as characterizing stoichiometry, counting cation adducts, and identifying patterns of complex fragmentation. It enables the detection of binding cooperativity and even low-abundance species with unique stoichiometries. For example, a binding stoichiometry limited to a 2:1 ratio suggests an external stacking interaction. Furthermore, the count of cation adducts can provide additional insights into the interaction, as ligand intercalation can displace one of the cations that stabilize the G4 assembly, an event clearly observed by ESI-MS. It's worth considering that in some cases, depending on the structure, ligand

interaction may lead to the entrapment of an additional cation.^{358,370} Xu et al. conducted an investigation into the CID patterns for ligand-G4 complexes, revealing that different fragmentation patterns correspond to different binding motifs. Intriguingly, complex fragmentation by the loss of the small molecule is associated with stacking interactions.³⁷¹

2.2.2 NMR spectroscopy of nucleic acids

NMR is widely used for the structure characterization of organic small molecules but also for studying nucleic acids. The simplest analysis that can be performed is 1D-¹H experiment. Nucleic acids consist as repetitive building blocks and therefore the 1D spectrum contains regions where atoms located at the same position in the nucleobase resonate. A scheme of the chemical shift dispersion for the various nucleobase region is displayed in Figure 2.2.3.

Imino protons resonate normally between 10 and 15 ppm but as they are acidic protons (their pK_a value is 9.2)³⁷² they exchange rapidly with the solvent (buffered water) so their signal will be broadened due to the detection of an ensemble of chemical shifts. When the nucleic acid folds into its secondary structure, these protons become protected from solvent exchange give rise to sharper signals which can be detected in the NMR spectrum, even though they still remain weaker and broader than the other signals. Additionally canonical base-pairs normally lead to lower field imino protons resonance with respect to non-canonical base-pairs, *i.e.*, Hoogsteen base-pairs (Figure 2.2.3a, see chapter 2.1 for further information). Aromatic protons, such as H2 for adenine, H8 in purines and H6 in pyrimidines resonate between 6 and 9 ppm, whereas H5 in pyrimidines usually resonate at lower ppm. Their exchange with the solvent is relatively slow, so their signal is usually sharp but since 1 or 2 protons for each base are observed in a relatively small region the signals are normally overlapping especially with increasing size of the nucleic acid.³⁷³

The same problem even at a larger extent applies to the sugar protons which resonate between 3.5 and 6.5 ppm, the H1' protons are normally shifted downfield compared to the remaining sugar atoms, overlapping with the aromatic H5 of pyrimidines. Therefore ¹H-1D experiments cannot be the tool for the structure determination of nucleic acids but they are normally used to study comparatively fast effects, for example in kinetic studies, or to be employed repetitively as in titration experiments.

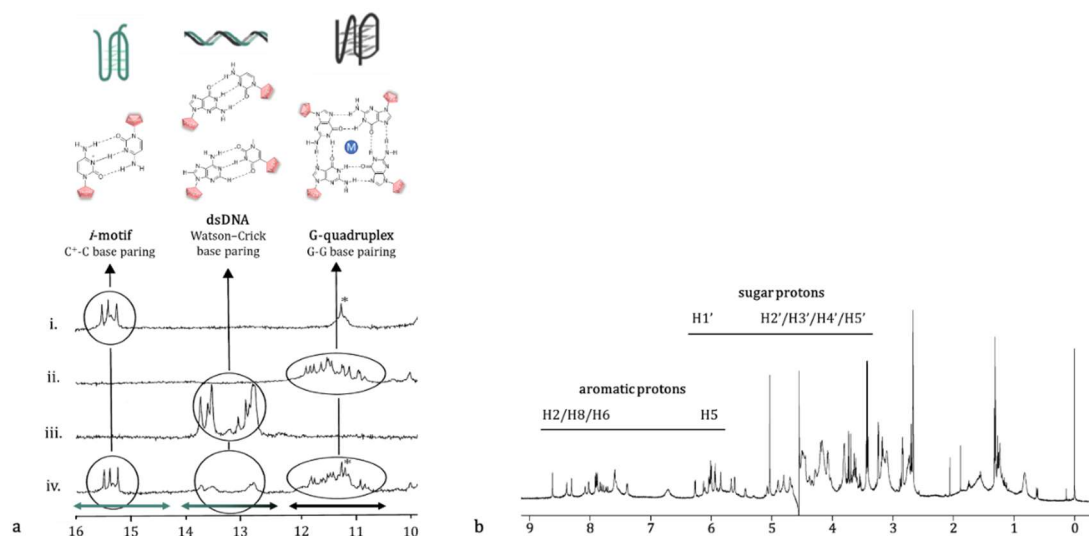


Figure 2.2.3. (a) ^1H NMR spectra of 22CT, 22AG (complementary strands) and their 1:1 mixture in different conditions to represent the chemical shift difference $\Delta\delta_{\text{H}}$ of the imino protons with different base-pairings: i.) 22CT at pH 4.5; ii.) 22AG at pH 4.5; iii.) 22CT + 22AG at pH 7; (iv.) 22CT + 22AG at pH 4.5. Imino protons of thymines in the i-motif structure are labelled with a star. Figure adapted from Phan *et al.*³⁷⁴ (b) ^1H NMR spectra of TERRA12 with annotation of the specific region of resonance of aromatic and sugar protons.

NMR experiment to study the binding of ligands

Although NMR is a very useful and widely used technique, it is important to keep in mind that its sensitivity is rather small, thus it is necessary to operate with high concentrated sample (min. $\sim 100 \mu\text{M}$) therefore a precise determination of the thermodynamic dissociation constant K_{D} for high affinity ligands (K_{D} in the nanomolar range) is rather complicated. Nonetheless, once the 1D- ^1H NMR signals have been assigned, their changes can be followed during a stepwise titration of the ligand to a nucleic acid sample, as the signals closer to the binding of the ligand will suffer the largest variation.

The binding kinetic has a large impact on the NMR signals and depends on the equilibrium between bound and free state which give rise of course to different signals. This leads to the three different case scenarios:

- fast exchange: if the binding constant is significantly larger than the difference of the resonance frequencies between bound and free states. In this case only one NMR signal per nucleus is observed at the weighted average of the chemical shifts.

- intermediate exchange: when the k_{binding} is similar to the difference of the chemical shifts, leading to a broad signal that is the most unfavourable event, because the signals can even become undetectable.

- slow exchange: for high affinity complexes, when the $k_{binding}$ is significantly lower than the chemical shifts difference, two signals are observed for each proton. Upon ligand addition the signal corresponding to the bound state will increase, whereas the signal corresponding to the free state will decrease.^{375–378}

A representation of the NMR signal changes for the three cases can be seen in Figure 2.2.4.

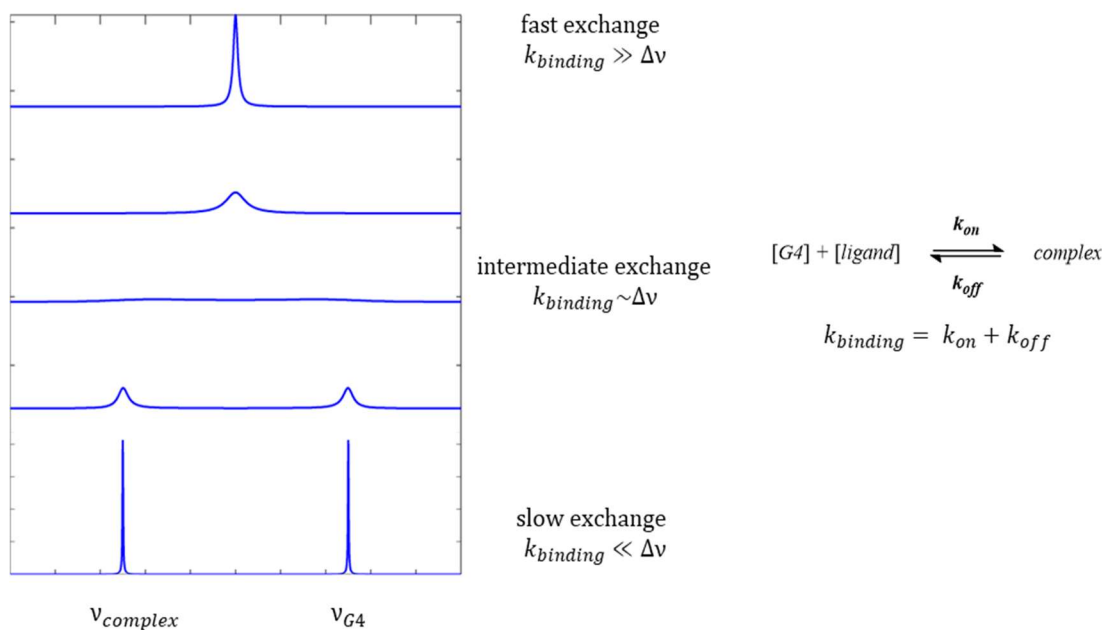


Figure 2.2.4. Schematic representation of NMR signal changes during titration with a ligand in the case of a binding process in fast, intermediate and slow exchange

Structure determination of nucleic acid

As mentioned earlier, more sophisticated methods are needed for determining the 3D structure of the nucleic acids. More specifically for the assignment of all signals a combination of different 2D NMR experiments has to be used.

Two-dimensional NOESY usually facilitates the assignment of imino and aromatic protons. The quantum chemical nuclear Overhauser effect allows the magnetization transfer by cross-relaxation of two different nuclei and depends on the spatial distance between the two nuclei but not on the chemical bonds. In the 2D spectrum, cross-signals between two signals are observed whose intensities directly related to the distance between the two atoms giving the signals.³⁷⁹ (Figure 2.2.5)

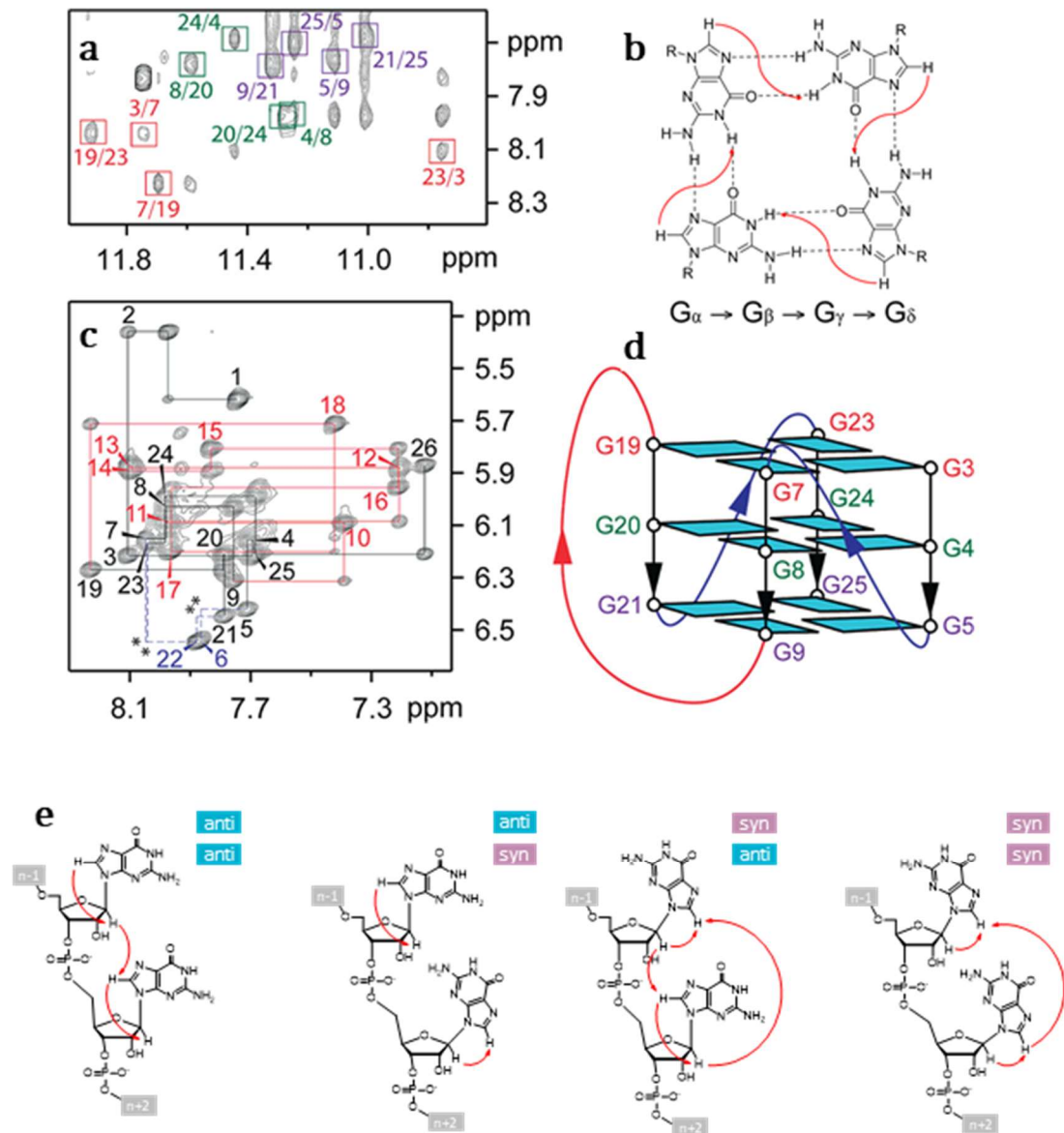


Figure 2.2.5. NOESY spectra of 26CEB (a) The imino-H8 connectivities (b) $G_{\alpha}\cdot G_{\beta}\cdot G_{\gamma}\cdot G_{\delta}$ tetrad showing the proximity of imino and H8 protons (red arrows). (c) The H8/6-H1' sequential connectivities. Intraresidue H8/6-H1' cross-peaks are labelled with residue numbers. Missing sequential connectivities are marked with asterisks. (d) Schematic structure of the 26CEB G4, guanines of the same G quartet are coloured with the same colour. (e) H8/6-H1' proximity with the various glycosidic bonds. Adapted from Amrane *et al.*³⁸⁰

NOESY spectra can be useful for several effects that can be observed:

- imino protons give rise to the so-called “imino-walk” because the distance between the imino protons of two consecutive base-pair is below 6 Å in several secondary structure (which is the upper limit distance for the formation of cross-signals), therefore cross-signal between the imino protons of two sequential nucleotides can be observed

- in G4, the formation of the quartets causes the vicinity of the aromatic H8 of one G and the imino H1 of the adjacent G on the same quartet, therefore Gs of the same quartet can be easily recognized
- the H1' of the sugar and the aromatic proton of both the same nucleotide and of the adjacent nucleotide are also sufficiently close to give a cross-signal, therefore it is possible to trace the $H_{\text{arom}}(n)$ -H1'(n)- $H_{\text{arom}}(n+1)$ sequential walk. Depending on the glycosidic torsion angle the walk may be interrupted if there is an anti-syn step, otherwise the full walk could be obtained. Similarly, the sequential walk between the aromatic protons and the H2'/H2'' of the sugar can be traced.

As previously mentioned, the NOESY cross-signals depends directly on the interatomic distance, therefore NOESY spectra can be used for determining distance restraints for structure calculations.¹²⁰

Another experiment performed for the assignment of the signals is the 2D HSQC that shows cross-signals between protons and their covalently bonded hetero atoms (either ^{13}C or ^{15}N) relying on 1st order J-coupling.^{381,382}

$^1\text{H},^{15}\text{N}$ -HSQC can be used to differentiate between the imino protons of different nucleotides, for example the imino nitrogen of uracil resonate at 15 ppm whereas in guanine it resonates around 70 ppm. Furthermore, exchange protected amino residues can be monitored with this experiment.³⁸³

$^1\text{H},^{13}\text{C}$ -HSQC also gives valuable information because it allows to distinguish between different aromatic protons, the C2 resonates around 152 ppm whereas the other aromatic carbons resonate at lower ppm, or to distinguish between the aromatic H5 of uracil/thymine with their H1' as the sugar carbon resonate at lower ppm than aromatic carbons. With sufficient resolution it is possible to assign pyrimidine C6 and purine C8 as the latter are normally found few ppm upfield, although the values are comparable.³⁸⁴ An example of a $^1\text{H},^{13}\text{C}$ -HSQC can be found in Figure 2.2.6.

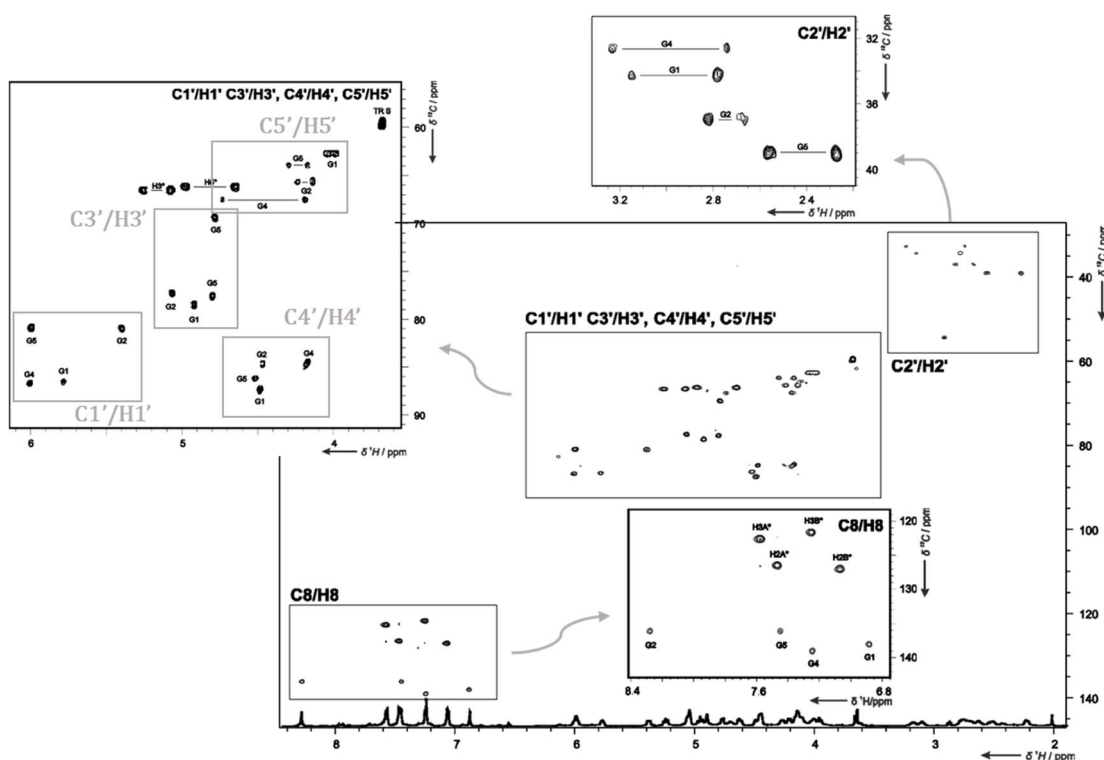


Figure 2.2.6. $^1\text{H},^{13}\text{C}$ HSQC spectrum of GG-Az1-GG construct. Zooms of the regions marked with a box are shown in panels. Figure adapted from Thevarphadam *et al.*³⁸⁵

$^{13}\text{C},^1\text{H}$ - HMBC experiments are similar to HSQC but it allows the observation of the cross-signal between the proton and heteroatoms separated by two, three or four bonds. The long-range $^3J_{\text{HC}}$ -couplings can be also exploited to correlate via the carbon C5 the imino and the aromatic proton belonging to the same residue.³⁸⁶

The sugar protons are usually assigned thanks to the combination of COSY and TOCSY experiments. As previously stated, in the first experiment a cross-signal is observed between protons that are coupled to each other and that are two or three bonds apart, whereas TOCSY shows correlations between all the protons of a given spin system, the magnetization is transferred successively to coupled protons unless there is a small or zero proton-proton coupling or in the presence of heteroatoms.

The determination of the three-dimensional structures of biomolecules by NMR has its origins in the 80s with the work of Kurt Wüthrich that resolved the structure of proteinase inhibitor IIA.³⁸⁷ The structural parameters restraints obtained by NMR experiments are utilized for the calculations which are performed by molecular dynamics that gives as result a multitude of possible structures, the arrangements with the highest overall energy are depleted whereas a subset of the most stable are selected. The quality of the simulation is evaluated in terms of RMSD values of all atoms or of heavy atoms.

The restraints are obtained by the intensity of the cross-peaks in the NOESY spectrum, in terms of integrations, that are correlated, as previously mentioned, with interatomic distances by the use of a well characterized fixed distance, such as the H5-H6 vector.³⁸⁸

Iterative computer assisted approaches such as ARIA (Ambiguous Restraints for Iterative Assignment) can help obtaining more exact distance restraints.³⁸⁹ The calculation is often repeated to remove the errors of the initial data.

As previously mentioned, the determination of the three-dimensional arrangement of the complex is particularly useful in the early drug discovery because it provides insights of the crucial interactions that takes place in the binding, giving important information about SAR relationship and mechanism of action of the binder.

2.2.3 CD spectroscopy of G-quadruplex

CD is by far the most commonly used and straightforward low-resolution method for investigating the topology of G4 structures.^{390–393} As mentioned in Chapter 1.5, CD is a spectroscopic technique that assesses the difference in the absorption of left- and right-circularly polarized light by chiral compounds or compounds in a chiral environment.

G4 are typically identified by their unique UV absorption bands in the range of 210–300 nm. These bands result from electronic transitions occurring between stacked guanine bases. More precisely, there are pi-pi* transitions within guanine exciton couples observed at 279 and 248 nm, leading to the appearance of CD bands. The specific sign and intensity of these bands depend on the relative orientation of the stacked guanines. In the case of parallel topologies, there is a prominent positive band centred at 260 nm and a negative band at 240 nm. Antiparallel topologies are marked by a positive band at 290 nm, a sometimes-shallow negative band at 260 nm, and a positive band at 245 nm. Hybrid-type structures exhibit a positive band at 290 nm and a shoulder at 260–270 nm.¹²⁵ Example of the CD spectra of the different topologies can be seen in Figure 2.2.7.

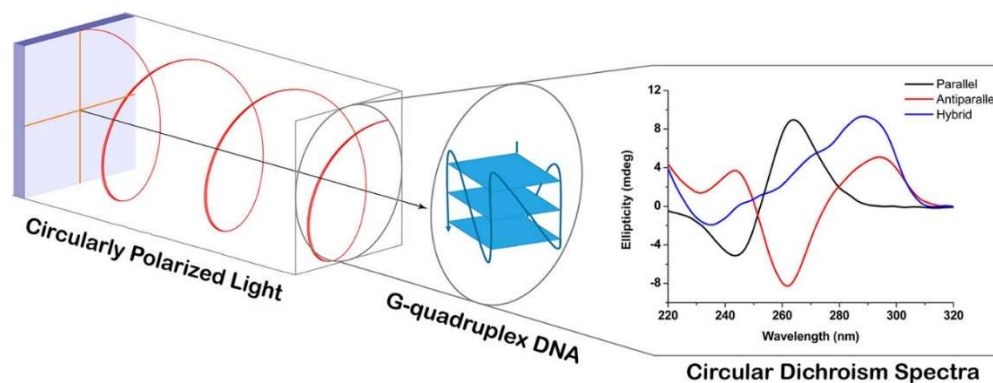


Figure 2.2.7. Difference on the CD spectra of parallel (black), antiparallel (red) and hybrid (blue) topologies.³⁹⁴

CD is a relatively fast and simple technique, but it provides limited resolution since it offers only a global view of the guanine stacks within the sample. It is challenging to differentiate mixtures of different G4 structures from pure ones. For example, spectra arising from mixtures of antiparallel and hybrid or parallel structures may resemble those of pure hybrid-type structures, leading to interpretative challenges. Furthermore, other supramolecular features, such as additional base stacking, double helices, and mismatches, can influence the overall CD spectrum and potentially lead to misinterpretations.³⁹⁵

An instance of additional guanine stacking altering the CD signature can be seen in the case of the PDB ID: 2KF8 structure.¹²⁸ Despite having a 2-tetrad antiparallel topology with tetrads of guanines stacked in a *syn/anti* fashion, its CD signature exhibits a positive band at 260 nm, which could be mistakenly interpreted as a hybrid-type topology. This is due to the presence of additional base stacking on both tetrads of triads composed of loop nucleotides, resulting in *anti/anti* stacking.³⁹⁵ Moreover, non-classical tetrads (e.g., G•C•G•C, as found in the mutant human telomeric GGGCTA motif) can also impact the CD spectra.³⁹⁶

In addition to its role in structural analysis, CD is highly valuable for investigating ligand-G4 interactions. Through titration experiments, CD provides important insights into relative binding strength and, if applicable, the emergence of changes in G4 structure. Furthermore, CD is employed to determine how a ligand influences the thermal stability of a G4 motif in melting assays.

2.3 Development of ligands for G-quadruplex

2.3.1 Aim of the work

During my PhD the attention has been focused on the design, synthesis and optimization of new ligands potentially able to bind to G4 and stabilize this arrangement. Throughout the three years different scaffolds have been explored, the following chapters will present the studied ligands divided according to their chemical class, at the beginning of each chapter a small introduction on the chemical structure and the *rationale* behind the choice will be provided together with some examples of the use of that particular moiety in literature, then the results of the study will be presented accompanied by the description of the methods used for the synthesis, the extraction and the analysis.

New potential G4 binders are normally designed after a careful literature analysis and are either extracted from plants, derivatized with simple reactions or synthesized from commercially available starting materials. After accurately designing and studying the reactions and techniques to use, further optimized while operating, the compounds and intermediates are fully characterized by analytical techniques such as ^1H - and ^{13}C -NMR spectroscopy, MS and HPLC.

The obtained compounds were then subjected to binding assays based on different techniques. Their interaction with DNA species is studied with ESI-MS using both G-rich sequences folded into G4 and dsDNA in order to calculate binding affinity and gas-phase stability for all the different arrangements and therefore their selectivity both towards G4 over dsDNA and towards one G4 sequence over the others.

Furthermore, *in silico* modelling was used to provide a preliminary insight into the three-dimensional arrangement through molecular docking and molecular dynamics (MD). Moreover, the screening of the same compounds was studied through NMR techniques, performing 1D ^1H -NMR titrations and 2D HSQC, NOESY, HMBC, TOCSY and COSY on the sequences with the most promising compounds which would lead to structural elucidation.

Thanks to the collaboration with other research groups, the compounds were also tested through different techniques that will be presented in the following.

2.4 Flavonoids

2.4.1 Introduction

Phenolic compounds are widely distributed in nature and represent an important class of secondary plant metabolism. They are involved in the mechanism of antioxidants thanks to their propensity to donate hydrogen or electrons and thanks to their composition carried by stable intermediate radicals, which block the oxidation of various food ingredients, especially lipids.³⁹⁷ Phenolic compounds are categorized into different groups depending on their chemical structure: phenols, phenolic acids (such as hydroxybenzoic acid and hydroxycinnamic acid), stilbenes, condensed and hydrolysable tannins, lignans and lignins, coumarins and flavonoids.³⁹⁸

Flavonoids were first discovered in 1930 by Albert Szent-Gyorgyino from a compound isolated from orange. It was first classified as “vitamin P”, only later it was characterized as a flavonoid and in particular as “rutin” together with the discovery of other varieties.³⁹⁹ Ever since, flavonoids have gathered a lot of attention due to both experimental and human evidences on their biological and therapeutic actions.⁴⁰⁰ To date, more than 8000 different flavonoids have been described. They comprise a group of phenolic compounds with low molecular weight polyphenolic structures and are divided into flavonols, flavones, flavanones, anthocyanins, isoflavones and flavanols.⁴⁰¹

Flavonoids are demonstrated to have beneficial effects in modulation of the enzymatic activity and inhibition of cellular proliferation.⁴⁰²

Rutin and quercetin

Rutin is defined a “multi-targeted nutraceutical agent” because of its antioxidant, neuroprotective, hepatoprotective and cardioprotective effects. It structurally consists of the flavonoid quercetin and the disaccharide rutinose (Figure 2.4.2). Nouri *et al.* very recently collected the evidences supporting its role as an anticancer agent targeting multiple signalling pathways.⁴⁰³

In 2007, rutin was also demonstrated to interact with “extended” G4 by Sun *et al.*⁴⁰⁴ but in general all the studies have been focused more on quercetin, “the most abundant dietary flavonoid” than on the glycosidic analogue.^{405–407}

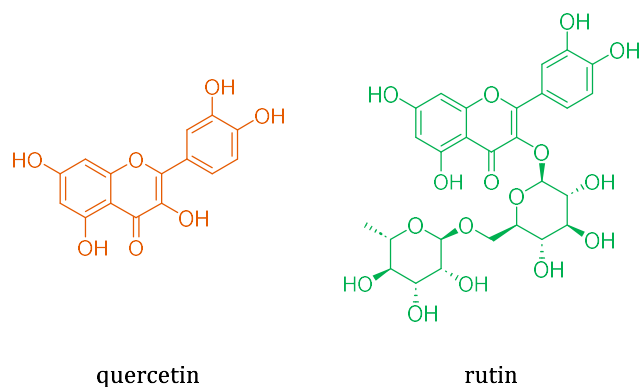


Figure 2.4.1. Chemical structures of quercetin (orange) and rutin (green)

Spectroscopic studies have been performed to investigate the interaction of quercetin with monomeric and dimeric G4. Quercetin resulted to bind to the monomeric conformation as a groove binder, whereas through pi-stacking to the dimeric conformation.⁴⁰⁸ Later, Tawani and Kumar performed detailed biophysical studies for the interaction of four representative flavonoids (luteolin, quercetin, rutin and genistein) with the human telomeric G4 sequence Tel7, which is formed by four strands of d(TTAGGGT) repeat. Among all the flavonoids they obtained the best results with quercetin and they focused the further structural study on that. They showed that it is able to stabilise the G4 structure and does not open the G-tetrad. The interaction was performed through pi-stacking at two sites, the upper and lower quartet.⁴⁰⁹ Bhattacharjee also performed a selectivity study on quercetin and showed that among all the several G4 structures, there was a “preference” for VEGF over the other G4 sequences and dsDNA.⁴¹⁰

On the other hand, the interaction of rutin with G4 was still not investigated in depth, and there was a lacking on information regarding ligand-target stoichiometry, its selectivity and its ability to stabilise the G4. In the following section, the result obtained from an ESI-MS binding study conducted on quercetin and rutin towards a telomeric G4 sequence and dsDNA, together with a docking study to investigate the possible binding mode and published in 2020 will be presented.⁴¹¹

Flavonoids derived from *M. Pomifera*

Osajin, pomiferin, scandenone and auricularin are demonstrated to possess antibacterial, anti-inflammatory, antidiabetic and antinociceptive properties.^{412–414} In particular, osajin and pomiferin were also previously reported to promote apoptosis in nasopharyngeal carcinoma cells and tumorigenic breast epithelial cells.⁴¹⁵ Their potential

to act as anti-tumorigenic and antiproliferative agents has been reported previously but is not yet fully understood. Hence, targeting human telomeric G4 DNA could be one of the mechanisms by which these flavonoids exert anticancer activity.

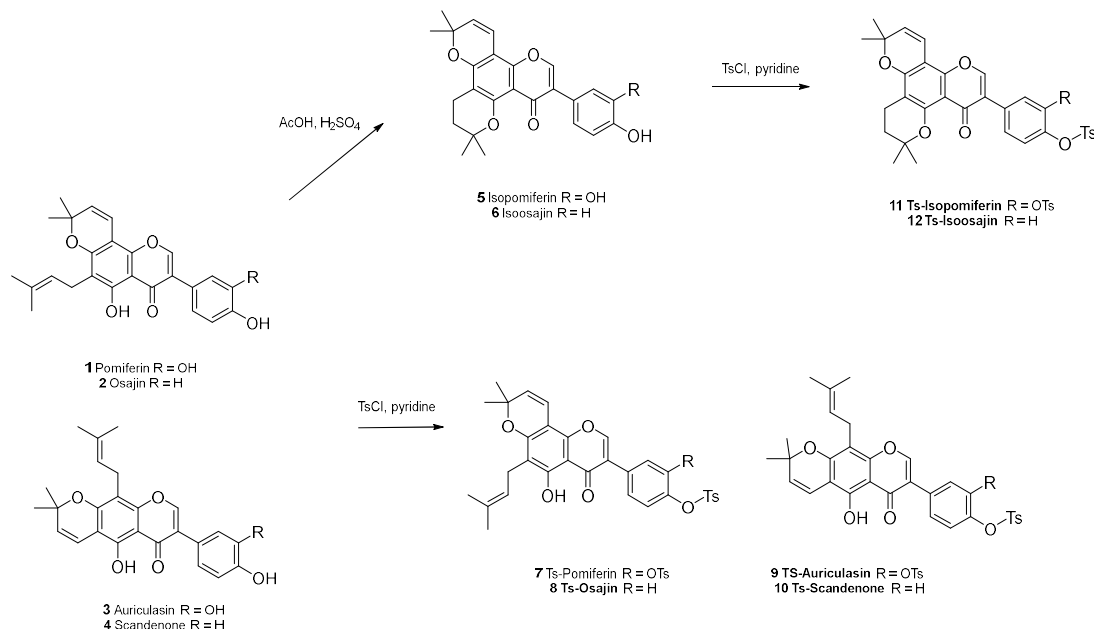


Figure 2.4.2. Chemical structure of the flavonoids derived from *M. Pomifera* and their semi-synthetic derivatization

In previous works of our research group, the flavonoids derived from *Maclura Pomifera*, namely osajin, pomiferin, scandenone and auriculasin have been tested against telomeric G4 and dsDNA using both blind molecular docking and electrospray ionization mass spectrometry (ESI-MS). They all showed a certain affinity for both sequences but, particularly for osajin and scandenone, they showed higher selectivity for G4.^{416,417} Semi-synthetic modification of osajin via B-ring derivatization (by the introduction of two different aryl sulfonate groups) was demonstrated in a previous work to lead to an overall decrease in DNA binding affinity, but the difference was more considerable for dsDNA, therefore improving the selectivity towards G4, in particular the toluenesulfonyl derivative retained satisfying G4 affinity values and stabilization properties compared to the benzenesulfonyl osajin.⁴¹⁷ In the following section, the most recent data of the study conducted on semi-synthetic flavonoids deriving from *M. Pomifera* substituted with this promising functional group (*i.e.* *p*-toluenesulfonyl) will be presented. (Figure 2.4.2)

2.4.2 Results and discussion

Rutin and quercetin

In the current study, rutin and quercetin were tested against the telomeric wtTel23a G4 and dsDNA as control using ESI-MS and molecular docking. The chemical structure of the two compounds is depicted in Figure 2.4.2. Calculating the BA for both NAs (the results are presented in Table 2.4.1), rutin and quercetin demonstrated comparable G4 BA, both forming a 1:1 ligand-target complex. However, quercetin showed an interaction also with dsDNA, whereas rutin demonstrated selectivity towards G4, as no trace of binding was detected. CID experiments were then performed to evaluate the stability of G4/flavonoid complexes. Again, the calculated $E_{COM}^{50\%}$ demonstrated that rutin is able to stabilise more efficiently the G4 arrangement (Table 2.4.1, Figure 2.4.3). The MS and CID spectra together with the characterization of the compounds can be found in the Appendix (Section 5).

Table 2.4.1 BA and $E_{COM}^{50\%}$ values determined by ESI-MS interaction assays between rutin, quercetin and wtTel23a G4 and dsDNA sequences. Quercetin results are coloured in orange, rutin results in green.

Compound	ESI-MS binding study				Molecular docking
	BA G4	BA dsDNA	Selectivity ratio (BA G4/BA dsDNA)	$E_{COM}^{50\%}$ G4 (eV)	Docking score G4 (-kcal/mol)
Quercetin	36.8	21.1	1.74	34.7	-8.8
Rutin	39.3	-	-	38.9	-9.7

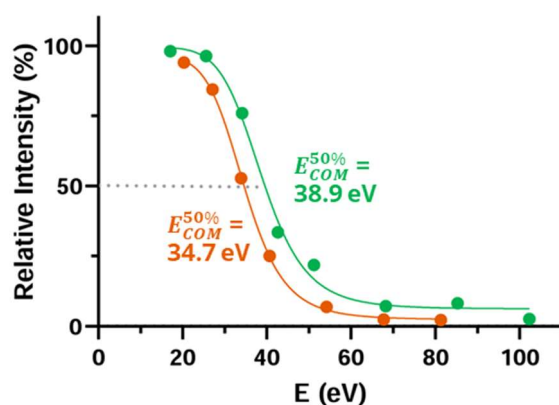


Figure 2.4.3. CID curves and calculated $E_{COM}^{50\%}$ of G4-ligand complexes for quercetin (orange) and rutin (green)

Molecular docking experiments were conducted to investigate the possible binding motif of the flavonoids with the telomeric G4. As previously mentioned, the interaction mode of rutin was never investigated at the time that the study was conducted, whereas quercetin was reported to bind to G4 *via* external binding and *via* groove binding

depending on the G4 sequence used.^{410,418} It must be pointed out that the G4 chosen for the computational study represent a parallel G4 which is reported to be the prevalent topology in cellular overcrowded solution conditions.^{419–421} According to our results (Figure 2.4.4), rutin was predicted to interact *via* stacking, while quercetin *via* groove binding. Docking results are resumed in Table 2.4.1. These findings have been published in 2020 by our research group.⁴¹¹

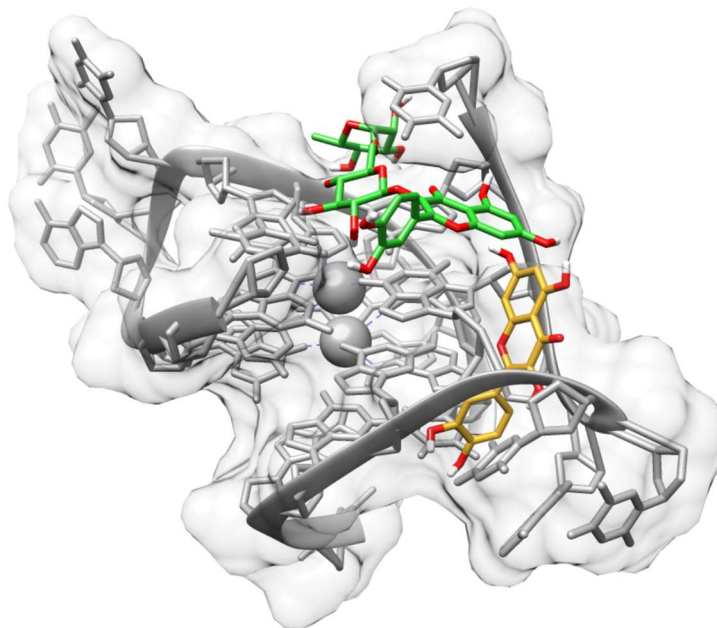


Figure 2.4.4. Predicted interaction patterns with G4 (PDB ID: 3CE5) for • quercetin and • rutin.

The major topology of wtTel23a has been demonstrated to be hybrid-2 according to the study performed by Dai et al.⁴²² Unfortunately, its three-dimensional structure has not been fully resolved by NMR or other techniques, therefore, to conduct the study of the interaction of rutin and quercetin with the telomeric G4, we decided to use wtTel26 which is resolved by NMR⁴²³ and is known to fold in the same topology.

An NMR titration of this G4 with the addition of the two ligands has been performed and the data are represented in Figure 2.4.4. Both the ligands showed indeed an interaction with the NA as can be evinced by the change of the peaks of the G4, however both the ligand showed an intermediate exchange between the bound and unbound species. This is demonstrated by the broadening of the signals, thus the interpretation to obtain information on the binding mode is not feasible, and more detailed 2D experiments to determine the three-dimensional structure would be needed (Figure 2.4.5).

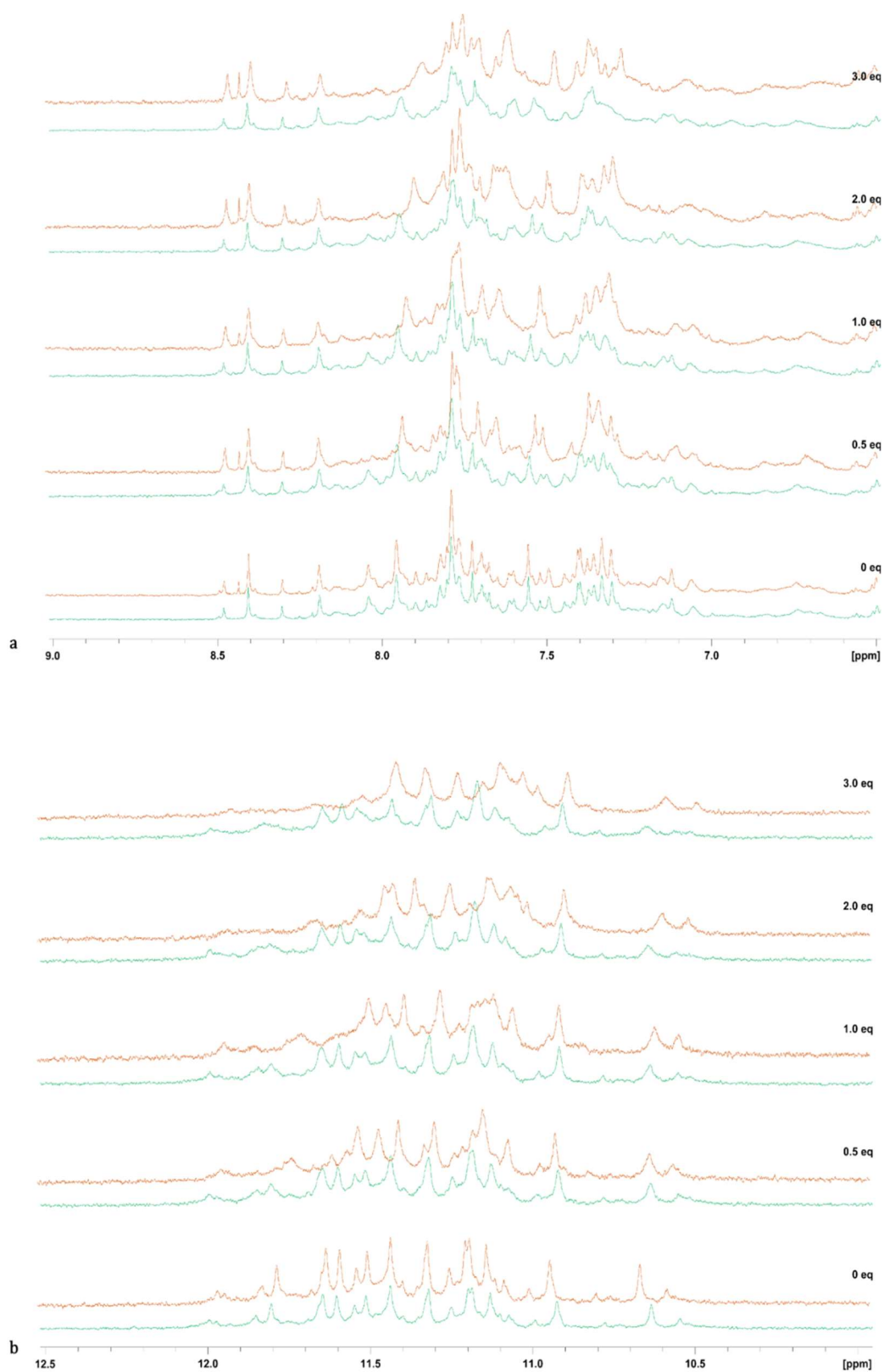


Figure 2.4.5. ^1H NMR titrations of wtTel26 with increasing amount of • quercetin and • rutin recorded at 298 K and 600 MHz with (a) excitation sculpting (aromatic region) and (b) jump-return (imino region) and water suppression.

Flavonoids derived from *M. Pomifera*

For what concerned the flavonoids deriving from *M. Pomifera*, the compounds were extracted from the fruits of the plant with ethyl acetate or diethyl ether to give pomiferin (**1**) and osajin (**2**) in high yield,^{424,425} with chloroform to gain extracts rich in auricularin (**3**) and scandenone (**4**).⁴²⁶ Isopomiferin (**5**) and isoosajin (**6**) were synthesized by heating the natural compounds in a mixture of acetic and sulfuric acid, whereas the *p*-toluenesulfonyl derivatives (**7-12**) were prepared adding at r.t. a solution of 4-methylbenzenesulfonyl chloride in pyridine to the isoflavone dissolved in pyridine.⁴²⁷

Table 2.4.2. BA and $E_{COM}^{50\%}$ values determined by ESI-MS interaction assays between compounds 1-12 and wtTel23a G4 and dsDNA sequences.

Compound	ESI-MS binding study			CID fragmentation	
	BA G4	BA dsDNA	Selectivity (BA_{G4} / BA_{dsDNA})	$E_{COM}^{50\%}$ G4	$E_{COM}^{50\%}$ dsDNA
1. Pomiferin	94	90	1.0	41.10	31.29
2. Osajin	67	21	3.2	56.61	36.98
3. Auricularin	97	99	1.0	41.09	30.33
4. Scandenone	76	38	2.0	51.97	48.96
5. Isopomiferin	26	15	1.7	40.91	50.72
6. Isoosajin	27	16	1.7	44.98	33.82
7. PomiTolSO2	-	-	-	-	-
8. OsaTolSO2	34	-	-	54.09	-
9. AuriTolSO2	34	21	1.6	44.89	40.26
10. ScandeTolSO2	19	8	2.4	45.27	48.32
11. IsopomiTolSO2	35	31	1.1	51.69	61.46
12. IsoosajinTolSO2	18	10	1.8	43.81	39.87

We investigated their behaviour thanks to an ESI-MS study in which all compounds were tested against the telomeric wtTel23a G4 sequence and dsDNA to assess whether these compounds are able to bind to the nucleic acids and how selective they are towards G4 over dsDNA. The data are presented in Table 2.4.2. Furthermore, we performed a series of molecular docking vs a G4 structure varying different parameters, such as grid dimension and precision and we carried out molecular dynamics simulations with the most represented poses to study the stability of the complexes during the simulation timeframe.

All the compounds, whose chemical structure is depicted in Figure 2.4.2, (except for **7**) showed interaction with both G4 and dsDNA but their selectivity towards G4 increased upon tosylation for most of the ligands compared to the native isoflavones, even though the BA affinity was slightly lower, as expected from a previous work.⁴¹⁷ In addition to the BA, the efficiency in stabilizing the macromolecular arrangement was determined through CID, calculating $E_{COM}^{50\%}$. The dissociation curves are depicted in

Figure 2.4.6 whereas representative MS and CID spectra together with the characterization of the compounds are reported in the Appendix (Section 5). Interestingly, the tosylated derivatives maintained almost the same values as the unsubstituted isoflavones, meaning that their ability to stabilize the G4 arrangement is preserved after the tosylation and the first event observed upon fragmentation is the loss of the ligand and the appearance of the peaks corresponding to the unbound G4. It is worth pointing out that some tosylated isoflavones showed the ability to alkylate the G4, for example **10** scandenone toluensulfonyl. Importantly, the same effect was not observed for dsDNA, therefore it may be interesting to further investigate this behaviour for a possible development of a ligand that could selectively deliver DNA damage.

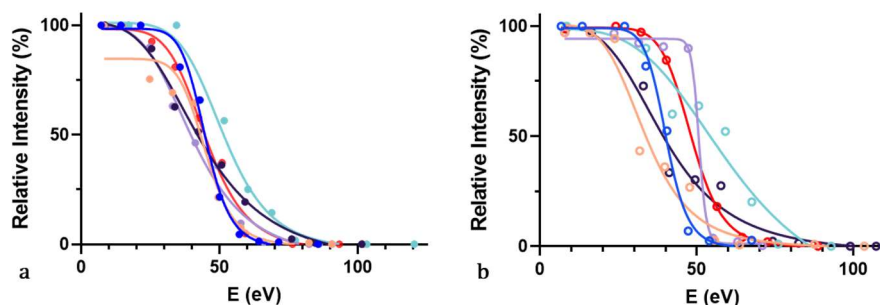


Figure 2.4.6. Plot of relative intensity of the complex (a) G4-ligand, (b) dsDNA:ligand against collision energy (eV) applied for the different compounds: • 5; • 6; • 9; • 10; • 11; • 12.

A computational study was then performed on the most promising compounds, *i.e.*, the compounds that showed the highest selectivity ratio values for G4 over dsDNA. Osajin, *p*-toluensulfonyl-osajin, scandenone and *p*-toluensulfonyl-scandenone (**2**, **4**, **8** and **10**) were subjected to a multiparametric docking process to obtain the poses for a following MD simulation using a telomeric parallel G4 structure, which is the prevalent topology in the cellular overcrowded solution conditions.⁴¹⁹ The four compounds were docked to the telomeric parallel G4 structure (PDB ID: 7KLP)⁴²⁸ using Glide and varying key parameters to obtain a complete ensemble of interaction modes for each complex, namely grid dimension (15 to 32 Å), docking precision (SP or XP), ions presence (two or three), and exhaustiveness of the sampling of conformer generation and of poses selection. This workflow produced a multitude of possible poses for every docked compound, that allowed us to select the best stacking and the best groove binding poses for each group in terms of docking score. The results are resumed in Figure 2.4.7. It is worth mentioning that the current work on tosylated derivatives of flavonoids have been performed in different years with respect to the work presented on rutin and quercetin in

the previous section, therefore the procedure adopted and the software used for the computational study has been modified and optimized.

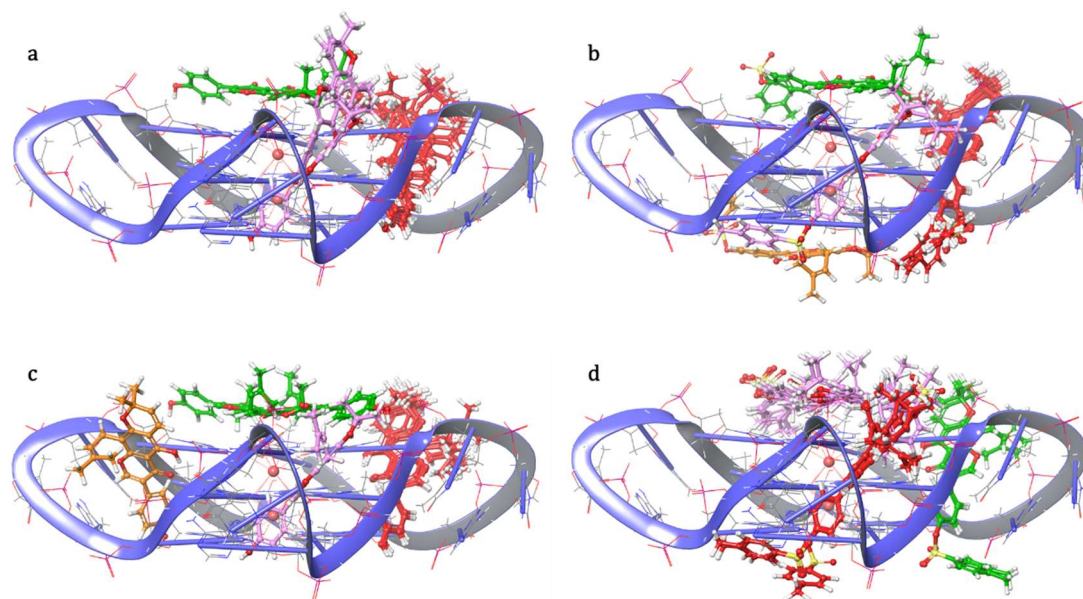


Figure 2.4.7. Poses overlap of the multiparametric molecular docking: osajin **2** (a), *p*-toluensulfonyl osajin **8** (b), scandenone **4** (c), *p*-toluensulfonyl scandenone **10** (d). The docked poses corresponding to a similar position on the G4 have been coloured with the same colour for simplicity.

A MD simulation (the details of the procedure can be found in the Material and Methods, Chapter 2.9) was then performed for each of the selected poses. The 8 trajectories obtained were then analysed and the over-time RMSD values were calculated for the ligands and for the Gs of the G4 structures. Osajin and scandenone (**2** and **4**) showed a preference for a stacking interaction as their stacking pose showed a stable trajectory with only small RMSD fluctuations, whereas the RMSD values of their tosylated derivatives (**8** and **10**) demonstrated a preference for the groove binding probably due to the steric hindrance of the tosylate moiety. Noteworthy, the groove binding pose of the native compounds showed a transition during the simulation timeframe, and the ligand move its position to stacking after 80 ns for scandenone (Figure 2.4.8 (a-d)) and after 10 ns for osajin (Figure 2.4.8 (e-h)).

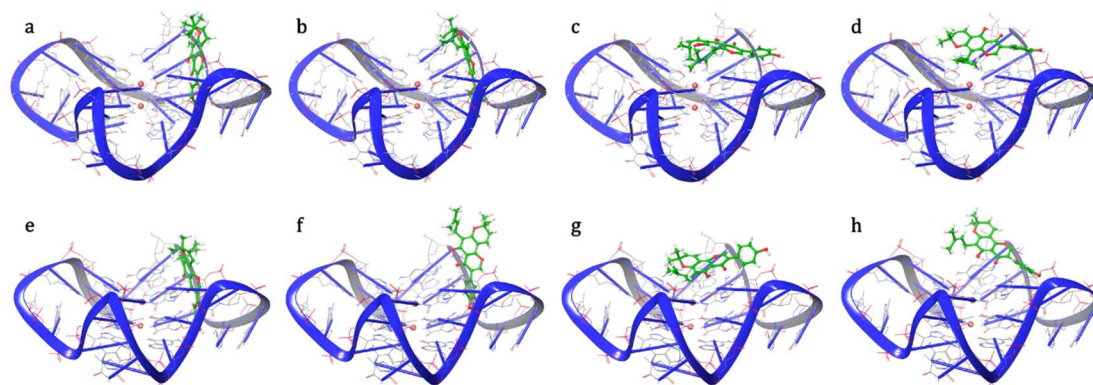


Figure 2.4.8. MD frames showing the pose transition from groove to stacking binding mode: scandenone **4** simulation time = 85 ns (a); simulation time = 88 ns (b); simulation time = 92 ns (c); simulation time = 95 ns (d), osajin **2** simulation time = 6 ns (e); simulation time = 10 ns (f); simulation time = 12 ns (g); simulation time = 15 ns (h).

Additionally, osajin and its tosylated derivative have been tested also by NMR titration against wtTel26 for the same reason as above and the data are presented in Figures 2.4.9.

The lack of changes in the NMR spectra upon addition of the ligand for both the compounds suggests that the compounds are not forming a stable complex with these particular G4s. This is probably due to the difference in the sequence between the G4s tested through NMR and ESI-MS. In order to support this hypothesis, we performed an ESI-MS binding study on wtTel26 that showed the absence of interaction between the ligand and its target, hence implying a possible sequence specificity of the wtTel23a. This aspect should be further investigated.

Thanks to the collaboration with Prof. Sissi group at the University of Padova, the compounds are currently being tested through CD titration and FRET melting.

The data presented are confidential as to date they are still unpublished.

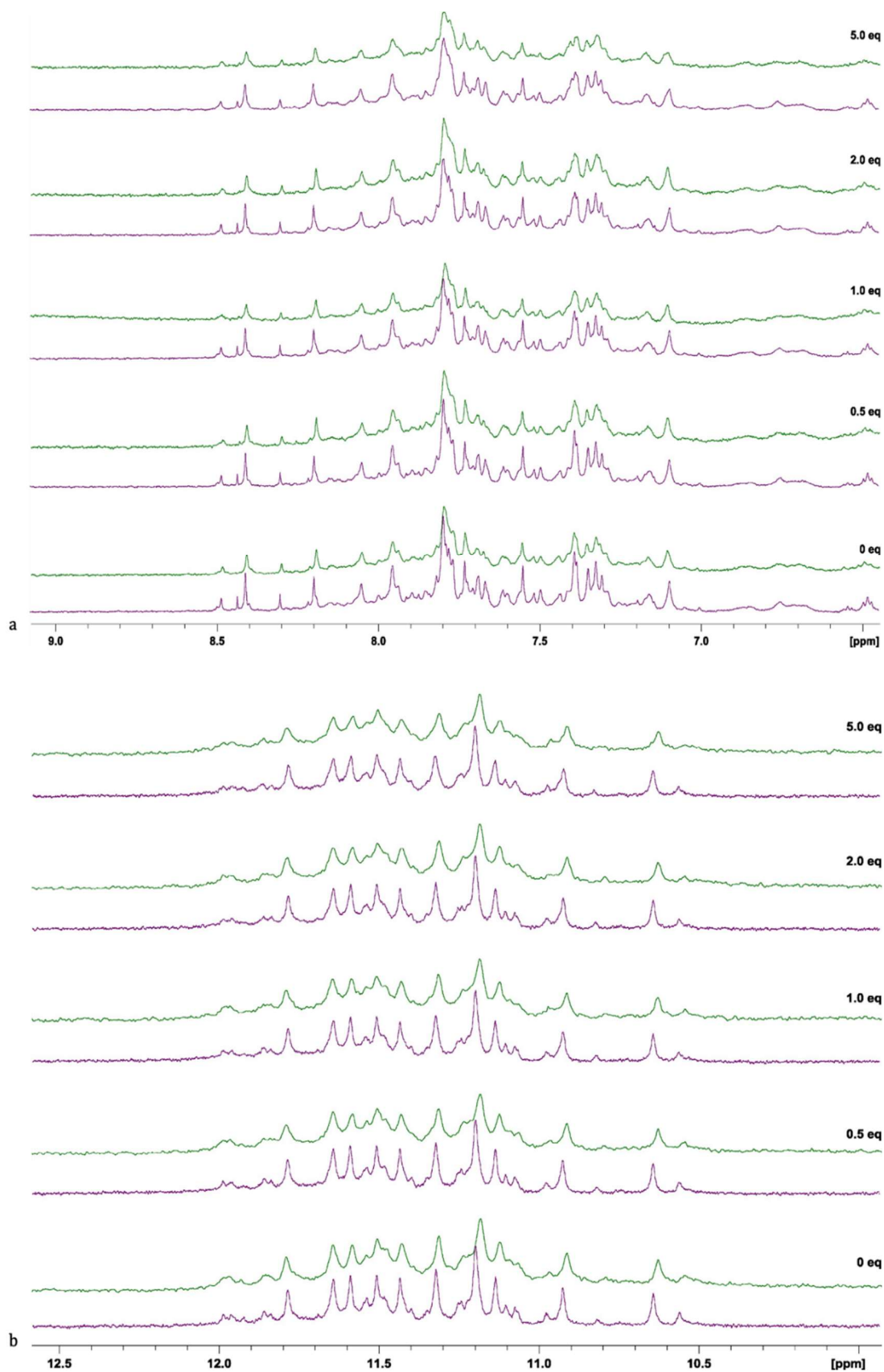


Figure 2.4.9. ^1H NMR titrations of wtTel26 with increasing amount of \bullet osajin and \bullet *p*-toluensulfonyl osajin (green) recorded at 298 K and 600 MHz with (a) excitation sculpting (aromatic region) (b) and jump-return (imino region) water suppression.

2.5 Anthraquinones

2.5.1 Introduction

Anthraquinones (AQs) represent an interesting planar scaffold for the design of ligands targeting G4. The chemical structure of the natural compounds doxorubicin and daunorubicin which are known dsDNA intercalating agents is AQ based.⁴³⁰ For this reason and thanks to their low redox potential and their ability to photosensitize through a one-electron oxidation reaction, they have been extensively used in the past.⁴³¹

In 1997, Sun and colleagues reported a 2,6-diamido AQ derivative that showed human telomerase inhibition properties and G4 binding by means of NMR thermal denaturation which indicated a drug-induced increase in the G4-DNA T_m by about 20 °C.²⁶⁰

Subsequently, Perry et al. prepared a series of 1,4- and 2,6-difunctionalized amido AQs and demonstrated their telomerase inhibition ability, the most potent having IC₅₀ levels of 4-11 μM. Moreover, a second group of AQ derivatives substituted at the position 1,5, 1,8, and 2,7 were shown to be able to interact with human telomeric G4 DNA. Their cytotoxicity appeared moderate against a panel of three ovarian cell lines and their ability to inhibit telomerase was assessed with promising IC₅₀ values in the range of 1.3–17.3 μM. For two of the compounds, an isothermal titration calorimetry (ITC) assay was performed in order to determine the binding constants to the human telomeric sequence d[AGGG(TTAGGG)₃], the values obtained were up to 7.9 x 10⁴ M⁻¹.⁴³²

In 2008, Zagotto and colleagues developed a series of AQ substituted in positions 2,6 and 2,7 with a small dipeptide (Figure 2.5.1) investigating whether the direction of the amide bond could influence G4 recognition and telomerase inhibition.^{433,434} Fluorescence melting studies revealed that introducing amino acids (AAs) with a positive charge, like lysine (Lys) or arginine (Arg), could give rise to electrostatic interaction with the phosphate groups of the backbone increasing the affinity towards the G4, whereas, adding apolar AAs, like phenylalanine (Phe), leucine (Leu) and isoleucine (Ile), would reduce the affinity for the NA. The telomerase inhibition activity was tested obtaining IC₅₀ values up to 0.8 μM in particular for the 2,6 and 2,7-AQ-βAla derivatives (derivatives whose side chain first AA was β-alanine (β-Ala) and the terminal AA was either Arg or Lys). Furthermore, all the compounds exhibited poor short-term cytotoxicity against the cell lines HeLa and 293T compared to the reference drug mitoxantrone, again the lowest value were obtained by the compounds containing Lys or Arg. However, in a senescence induction assay against HeLa cells prolonged for 6 days, a marked effect was produced by

the highly charged ligands, demonstrating their ability to trigger the envisioned mechanism and to internalize in the cells.

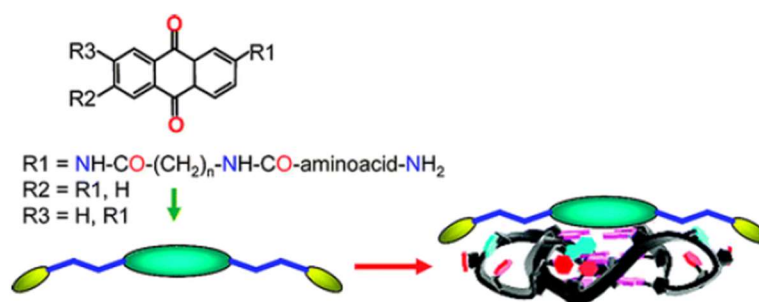


Figure 2.5.1. General scheme for the development of amino acid - anthraquinone conjugates interacting with G4 arrangement by stacking interaction through the anthraquinone rings and groove recognition by the charged side-chains.⁴³³

Moreover, in 2011 the same group proposed a novel series of aminoacyl-AQ derivatives that were symmetrically functionalized with N-terminal Lys conjugated with a Gly, Ala, or Phe spacer to study the influence of side chains with different length, flexibility, lipophilicity, and bulkiness.⁴³⁵ Moreover, different substitution positions were tested, namely 2,6, 2,7, 1,8, and 1,5, again to assess whether the position of the side chain could influence the selectivity for G4.

FRET fluorescence melting showed a general higher stabilization for the G4 DNA with respect to dsDNA with ΔT_m values up to 20.1 °C for the ligands substituted in position 2,6 and 2,7, whereas the opposite was shown for the ligands substituted at position 1,5 and 1,8. The CD study revealed that the ligands were promoting a shift on the G4 topology from a mixed type to an antiparallel topology, notably though, some variations were also observed on the spectra of dsDNA, suggesting a possible interaction also with this canonical arrangement. Furthermore, Taq polymerase and telomerase inhibition assays were conducted to determine the effects of the ligands on enzymatic processes involving DNA, they demonstrated a general inhibition of telomerase-mediated DNA elongation at lower concentrations compared to the DNA amplification by Taq polymerase. Moreover, the modulation of telomerase elongation was compared with the thermal stability obtained by the FRET melting and an almost linear dependence of the IC_{50} values for telomerase inhibition with $\Delta T_m/T_m$ was observed.

To summarise, thanks to the works of Perry *et al.* and Zagotto *et al.* it was possible to observe that the position of side chains on the central scaffold is crucial for the selectivity for specific DNA arrangements, in particular the 1,4-substitution is preferred for the

interaction with the dsDNA whereas the insertion of side chains at position 2,6 or 2,7 favours the recognition of triplex and G4 motifs.^{432,435}

In 2013, Ranjan *et al.* proposed the functionalization of various aromatic scaffold with a known aminosugar (neomycin) and the antraquinone derivative showed the highest affinity towards human telomeric G4 with IC₅₀ in the nanomolar range.⁴³⁶ Whereas Xodo and colleagues opted for the introduction of guanidinium groups on the similar anthrathiophenedione, developing a group of compounds with K_D below 0.2 μM toward hTELO and KRAS.⁴³⁷ Freccero's group took a different route, covalently attaching AQ to a quinone methide precursor, capable of generating a potent alkylating agent to form a stable covalent bond with different G4s and showing lack of interactions with dsDNA.⁴³⁸

Keeping all these information in mind, we decided to synthesize a new group of 2,6-AQ amino acid conjugates (2,6-AQs) and to test them with an ESI-MS binding assay to the telomeric tel23 G4 and dsDNA. Then thanks to the collaboration with Prof. Sissi FRET melting and CD titration were also conducted on different G4 sequences and dsDNA to gain more insights on the physico-chemical properties of the interaction and provide a preliminary sequence selectivity. Additionally, a molecular docking was performed on all the G4 topologies to investigate the possible binding motif. The work has been published in 2022.⁴³⁹

Recently, an NMR titration assay was also conducted with the best performing compound to investigate more in depth the binding and the possible selectivity for a particular G4 sequence. The data of this last study is still unpublished.

2.5.2 Results and discussion

In the current study, a series of 2,6-AQs were designed and prepared following a convergent synthetic scheme (Figure 2.5.2). The synthesis included the preparation of two 2,6-diazidoacetamido derivatives, which were used as the central scaffolds, and of 6 propargylamide derivatives used as side chains, 4 containing a single AA, namely L-Gly, Phe, Lysine and ornithine (Orn) and 2 dipeptides, i.e., Gly-Lys and βAla-Lys). The choice of the AA was dictated to investigate the influence of the side chain length and charges. Finally, a CuAAC click reaction led to the desired final compounds (**AQ3** - **AQ9**) after removal of the protecting groups.

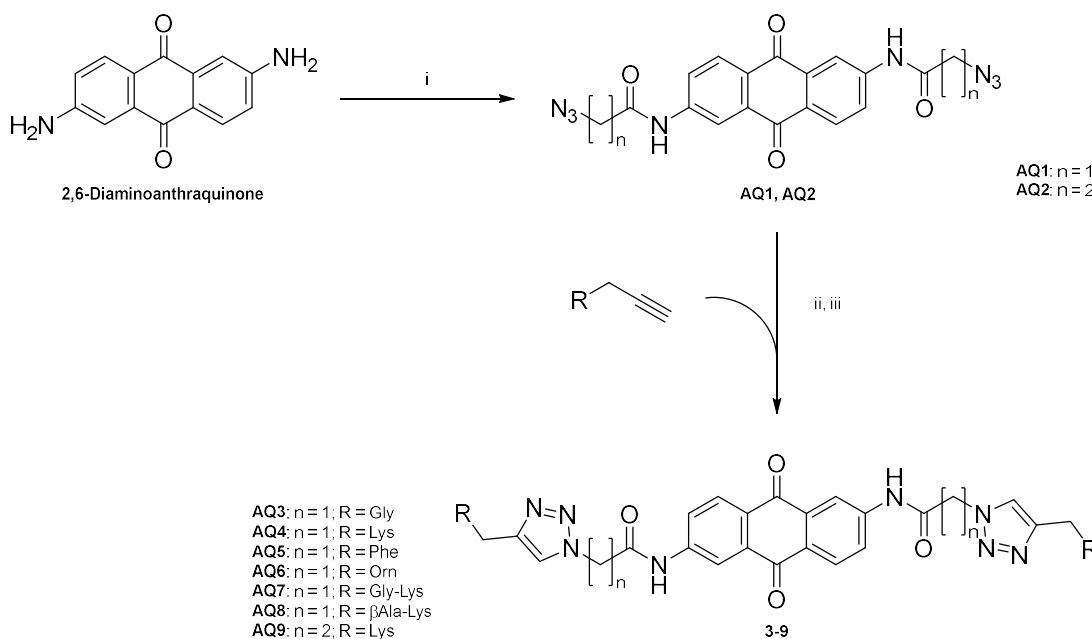


Figure 2.5.2. General scheme for the synthesis of the set of 2,6-bis-substituted anthraquinone-amino acid conjugates 3–9. i: chloroacetyl chloride, NaN₃/DMF; ii: sodium ascorbate, CuSO₄, K₂CO₃, DMSO/water, iii: TFA/DCM; 3-azidopropanoyl chloride was prepared by reacting 3-bromopropanoic acid with NaN₃/water and oxalyl chloride/DMF, DCM.

More specifically, the synthesis of bis(2-azidoacetamide) (**AQ1**) was achieved by diamidation of 2,6-diaminoanthraquinone (SM) using an excess of chloroacetyl chloride which was then subjected to a nucleophilic substitution reaction with NaN₃. The bis(3-azido propanamide) (**AQ2**) was prepared similarly but the SM was directly reacted with 3-azidopropanoyl chloride, previously prepared from 3-bromopropanoic acid according to a procedure known in literature.⁴⁴⁰ The side chains were prepared by liquid phase coupling procedures known in the literature from the propargylamine and the mono- and di- BOC-protected AA L-Gly, Phe, Lys and Orn.^{441–444} At the end, the final compounds **AQ3–AQ9** were obtained through the CuAAC coupling reaction in which thanks to a catalytic amount a pre-formed Cu(I) catalyst, obtained in situ by the reduction in water of Cu(II)SO₄ with sodium ascorbate, the central scaffolds and the side chains were coupled. Initially the reaction was conducted in neutral conditions without success, as per the mechanism proposed by Fokin (Figure 2.5.3),^{445–447} the reaction requires the deprotonation of the alkyne C-H, therefore subsequently a sub-stoichiometric amount of K₂CO₃ was added and yielded to the successful preparation of all the compounds. At the end the BOC-protecting groups were removed by the addition of TFA to a solution of the compounds dissolved in DCM leading to the TFA salts of compounds **AQ3–AQ9**.

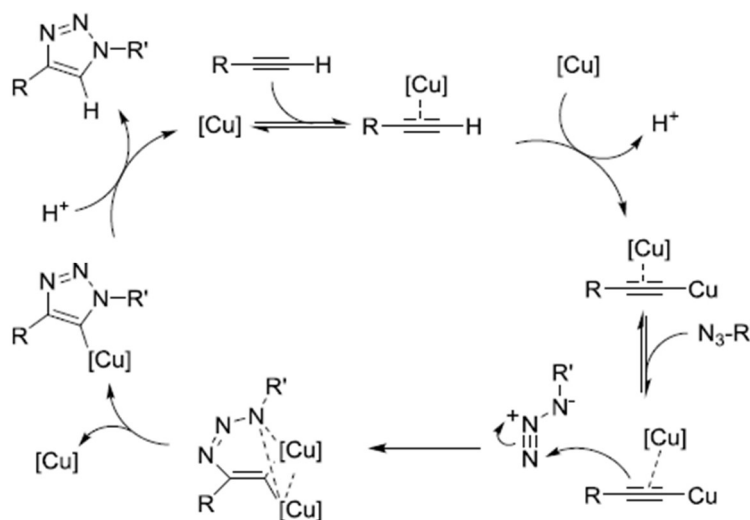


Figure 2.5.3. Mechanism of the catalytic cycle composing the CuAAC reaction between a terminal alkyne and an azide, the scheme includes the alkyne deprotonation and formation of a reactive copper acetylide which reacts with the triple bond in a [4+2] fashion that generates the aromatic triazole ring.⁴⁴⁷

As previously mentioned, ESI-MS binding studies were performed to assess the ability of compounds **AQ3-AQ9** to interact with the telomeric wtTel23a G4 and the selectivity for G4 was investigated using a dsDNA oligonucleotide as control. The samples were prepared with a 10:1 molar ratio of ligand:DNA in water with a 15-20% of MeOH. While no interaction with G4 was detected for **AQ3** and **AQ5** having Gly and Phe as side chains, all the other compounds showed a marked selectivity for G4 over dsDNA, since no binding was observed with the canonical arrangement. Furthermore, under our experimental conditions the complexes observed had a 1:1 stoichiometry. In terms of BA, compounds **AQ7** (42.5) and **AQ4** (37.1), both containing Lys, were highlighted as the most promising ligands probably due to the added positive charge of the amines which are able to intercept a higher number of negative charges present in the nucleic acid grooves. CID experiments were also performed to calculate the $E_{COM}^{50\%}$, ligands **AQ8** (46.2 eV) and **AQ9** (41.8 eV), bearing β Ala-Lys and Lys respectively as side chains, showed the highest values, whereas limited stability was observed for compound **AQ6**, in agreement with the low BA. The values obtained by this study are reported in Table 2.5.1, whereas the MS and CID spectra together with the dissociation curves and the full characterization of the compounds can be consulted in the Appendix, Section 5.

Table 2.5.1. BA and $E_{COM}^{50\%}$ (eV) values determined by ESI-MS interaction assays between compounds **AQ 3- AQ9** and the wtTel23a G4 telomeric or dsDNA sequences.

Compound	BA G4	$E_{COM}^{50\%}$ 1:1 (eV)	BA dsDNA
AQ3	-	-	-
AQ4	37.1	33.1	-
AQ5	-	-	-
AQ6	14.5	31.8	-
AQ7	42.5	36.4	-
AQ8	30.0	46.2	-
AQ9	22.8	41.8	-

Fluorescence melting analyses were performed thanks to the introduction of two functional groups (FGs) on the target DNA sequences, namely a fluorophore (6-FAM) and a quencher (Dabcyl). When the two FGs are in close proximity, *i.e.*, when the NA is folded the fluorescence signal of the FAM is quenched, upon DNA denaturation achieved by increasing the temperature, the two labelling groups fall apart and an enhancement of fluorescence is observed.^{448,449} As for CD melting experiments, the T_m is calculated upon addition of all the ligands and at different concentration and compared to the T_m of the NA without ligand to calculate the ΔT_m , giving an idea of the stabilization induced by the formation of the complex. This experiment was performed using different G4 sequences, namely Tel23, Tel24, KIT* and BCL2, and again a dsDNA sequence as control. The resulting ligand stabilization data are summarized in Figure 2.5.4. The efficiency of the tested compounds in stabilizing the G4s can divide the ligands into three main groups: compounds **AQ4**, **AQ7** and **AQ9** are the best stabilizers, **AQ6** and **AQ8** showed intermediate efficiency while **AQ3** and **AQ5** have the lowest stabilizing effect.

Summing up the information derived from the ESI-MS and the fluorescence melting, **AQ4**, **AQ7**, **AQ8** and **AQ9** were highlighted as good binders while **AQ3** and **AQ5** do not interact with G4, the ranking order was not affected by the preferred topology assumed by the NA even though the maximal ΔT_m detected can vary according to the sequences and the compounds seems to perform best in presence of parallel G4 (*i.e.*, Tel24 in presence of 20% PEG or BCL2). Furthermore, when tested in the presence of a dsDNA, the stabilizing effect was much more reduced and only compound **AQ7** significantly incremented the dsDNA T_m . This pointed to **AQ4** and **AQ9** as the most promising compounds for a further study.

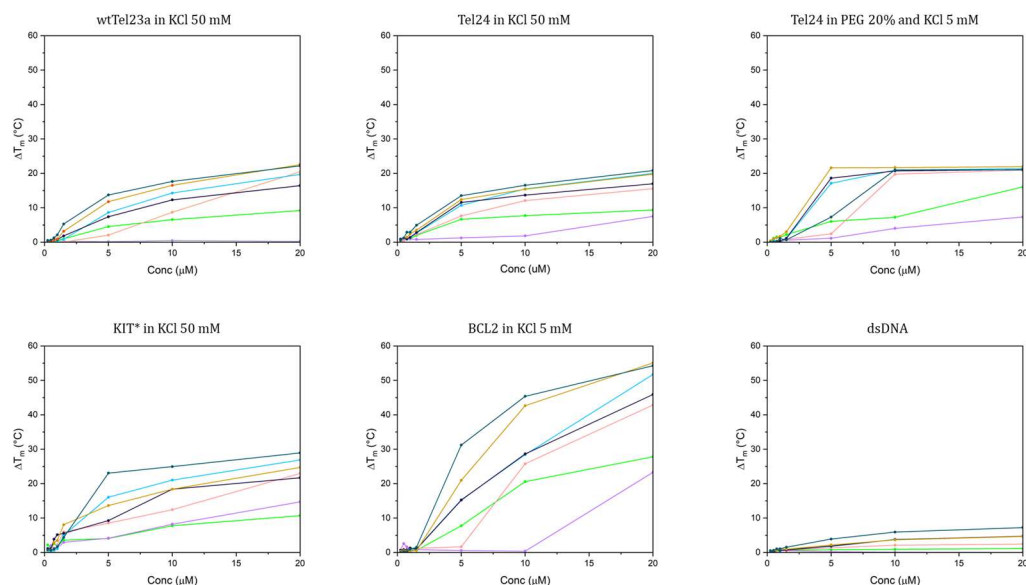


Figure 2.5.4. Values of ΔT_m of different DNA sequences with increasing concentration of ligands. Different colours indicate different ligands: • AQ3; • AQ4; • AQ5; • AQ6; • AQ7; • AQ8; • AQ9.

These derivatives were subjected to CD titrations in order to gain different information: to assess whether the folding topology is influenced upon ligand addition, to verify if that the different buffer condition used with the ESI-MS and the fluorescence melting could affect the G4 topology and to calculate the K_D . The data are presented in Figure 2.5.6. As expected, the CD spectra obtained by the target sequence wtTel23a both in K^+ or NH_4^+ are comparable both in presence and in absence of ligands. The experiment was performed both in K^+ and NH_4^+ buffer in order to assess whether the condition used in the MS experiment would lead to the same G4 topology as the one folded in a more physiological-like environment. Interestingly, our data showed an increment of the positive contribution at 290 nm and a decrease of the minimum at 260 nm. This could be explained by a partial shift of the folding topology from hybrid-2 to antiparallel, although the CD spectra of a “pure” antiparallel topology would be more symmetrical.

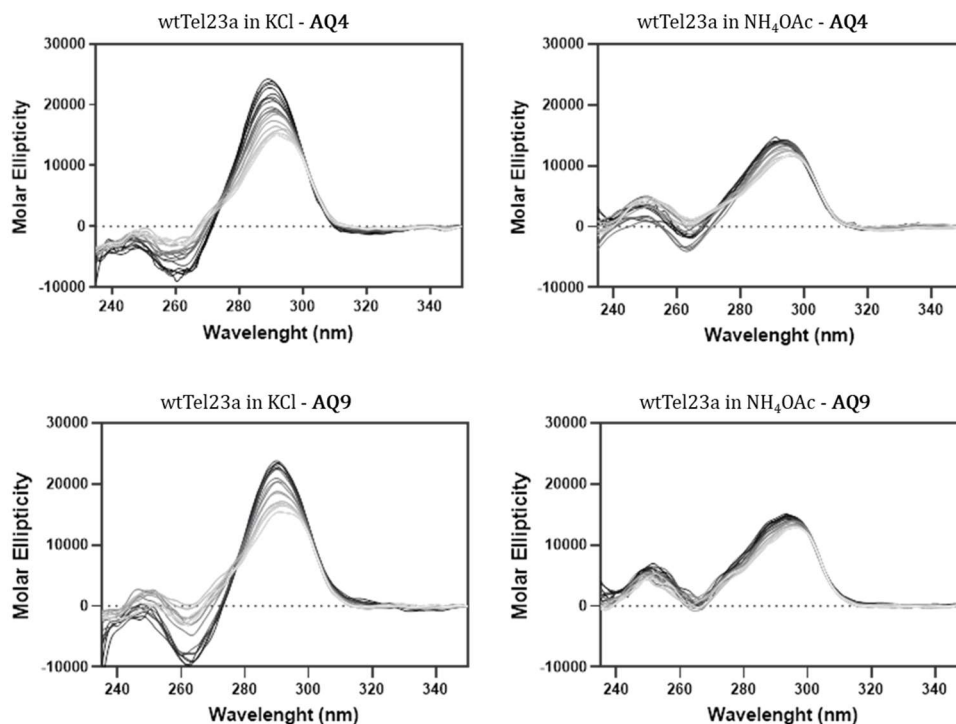


Figure 2.5.6. CD spectra of 4 μM wtTel23a in 10 mM Tris-HCl (pH 7.4) in KCl or NH_4OAc at increasing ligand/G4 molar ratios (the darker the line colour the higher molar ratio).

Finally, the variation of the CD signal was used to derive the K_D of the binding process, which are reported in Table 2.5.2. Similar values, in the range of $\sim 10 \mu\text{M}$, were observed in both the buffers. However, it must be pointed out that changes in CD signals were modest with NH_4^+ , thus leading to a reduced reliability of data analyses.

Table 2.5.2. Equilibrium dissociation constants obtained from CD experiments performed in 10 mM Tris-HCl, 1 mM EDTA, at 1:1 ligand/DNA molar ratio.

Sequence	Buffer	Ligand	K_D (μM)
wtTel23a	50 mM KCl	AQ9	8.8 ± 1.4
wtTel23a	50 mM KCl	AQ4	13.64 ± 1.2
wtTel23a	150 mM NH_4OAc	AQ9	12.04 ± 2.26
wtTel23a	150 mM NH_4OAc	AQ4	9.91 ± 1.93

Eventually, in order to investigate the binding mode of the 2,6-AQ most promising derivatives with the telomeric G4 arrangement, a computational study was performed. In particular, compounds **AQ4** and **AQ9** were docked to four G4 structures, one for each topology, namely parallel, antiparallel, hybrid-1 and hybrid-2 (PDB IDs: 1KF1,²⁹³ 143D,⁴⁵⁰ 2HY9,⁴⁵¹ 2JPZ,⁴²³ respectively), following sequence preparation and docking protocol of a previous work.⁴⁵² The pose obtained are depicted in Figure 2.5.7.

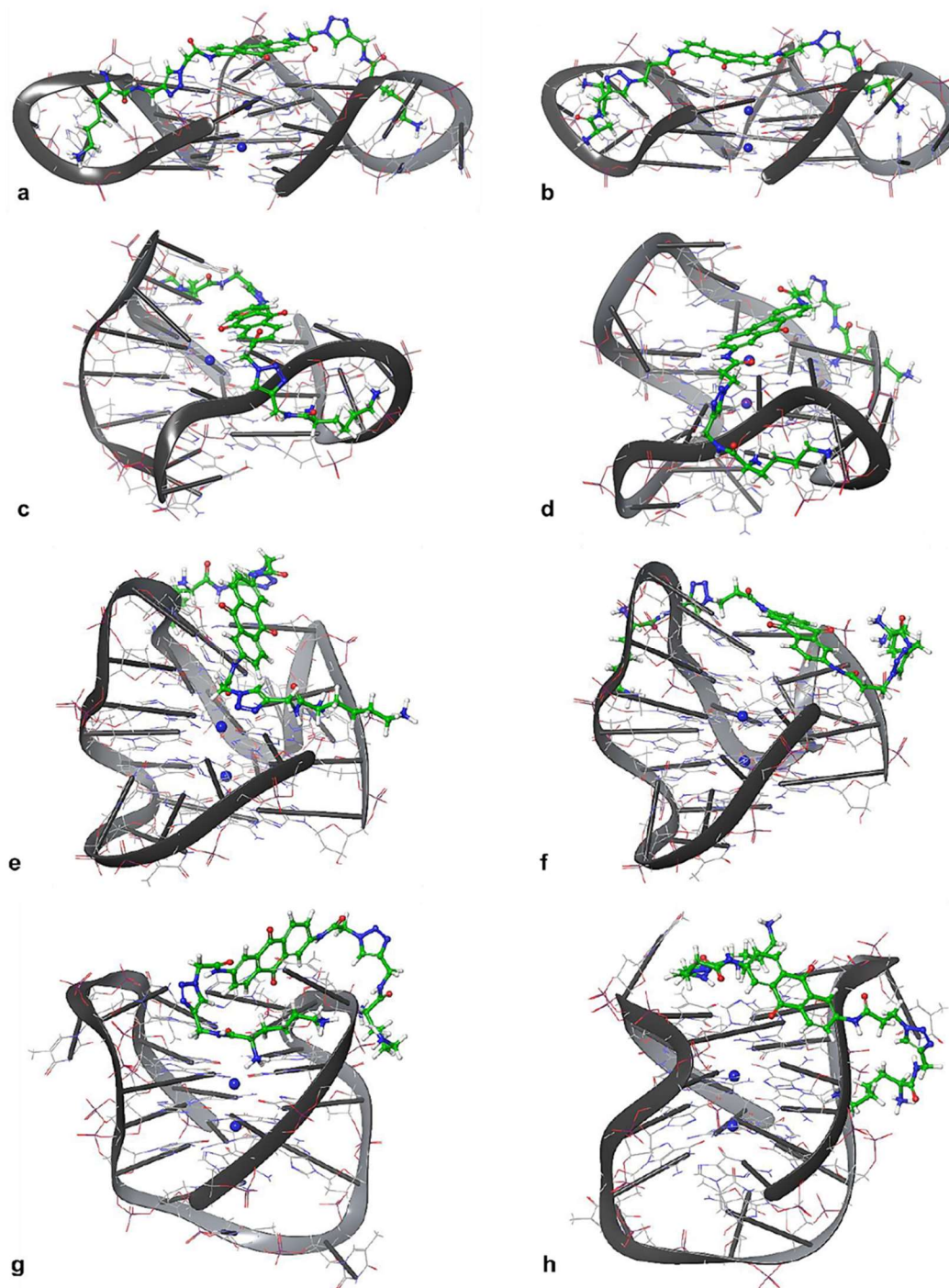


Figure 2.5.7. Predicted binding modes for the ligands with the four G4 topologies. Ligand AQ4: parallel (PDB ID: 1KF1) (a), hybrid-1 (PDB ID: 2HY9) (c), hybrid-2 (PDB ID: 2JPZ) (e), antiparallel (PDB ID: 143D) (g). Ligand AQ9: parallel (PDB ID: 1KF1) (b), hybrid-1 (PDB ID: 2HY9) (d), hybrid-2 (PDB ID: 2JPZ) (f), antiparallel (PDB ID: 143D) (h).

According to the predicted models for the parallel arrangement, as expected, our derivatives interact with G4 by pi-stacking on the upper guanine tetrad, elongating the side chains in the grooves, whereas for the other topologies the final quartets are more hindered thus less available to the pi-stacking interactions, the position that is adopted, therefore, resemble a stacking interaction and may be thought as a starting point for

following molecular dynamics simulations in which flexibility is allowed also for the receptor. Moreover, according to the CD experiments, there might be a conformational change on the hybrid topology that cannot be taken into account during molecular docking.

All the data presented and the detailed experimental procedure used for FRET melting and CD have been published in 2022. The detailed experimental procedure adopted is reported in Chapter 2.9, for the FRET melting and the CD titrations performed from the cooperation research group please refer to the publication.⁴³⁹

Given the relevant results obtained in this study, compound **AQ9** was then selected for a subsequent study using NMR for titrations on different G4 sequences, namely wtTel26, BCL2 and TERRA12. wtTel26 was chosen *in lieu of* the wtTel23a used for the ESI-MS, fluorescence and CD study, since unfortunately, to date the three-dimensional structure of the latter is still unresolved but is reported to fold as hybrid-2, equal to wtTel26.⁴²³ The BCL2 (PDB ID: 2F8U)²⁷⁶ used for the NMR is also reported to fold into hybrid-2 topology, whereas TERRA12 (PDB ID: 2KBP)²¹⁸ is a bimolecular parallel topology. The data are presented in Figure 2.5.8 and 2.5.9.

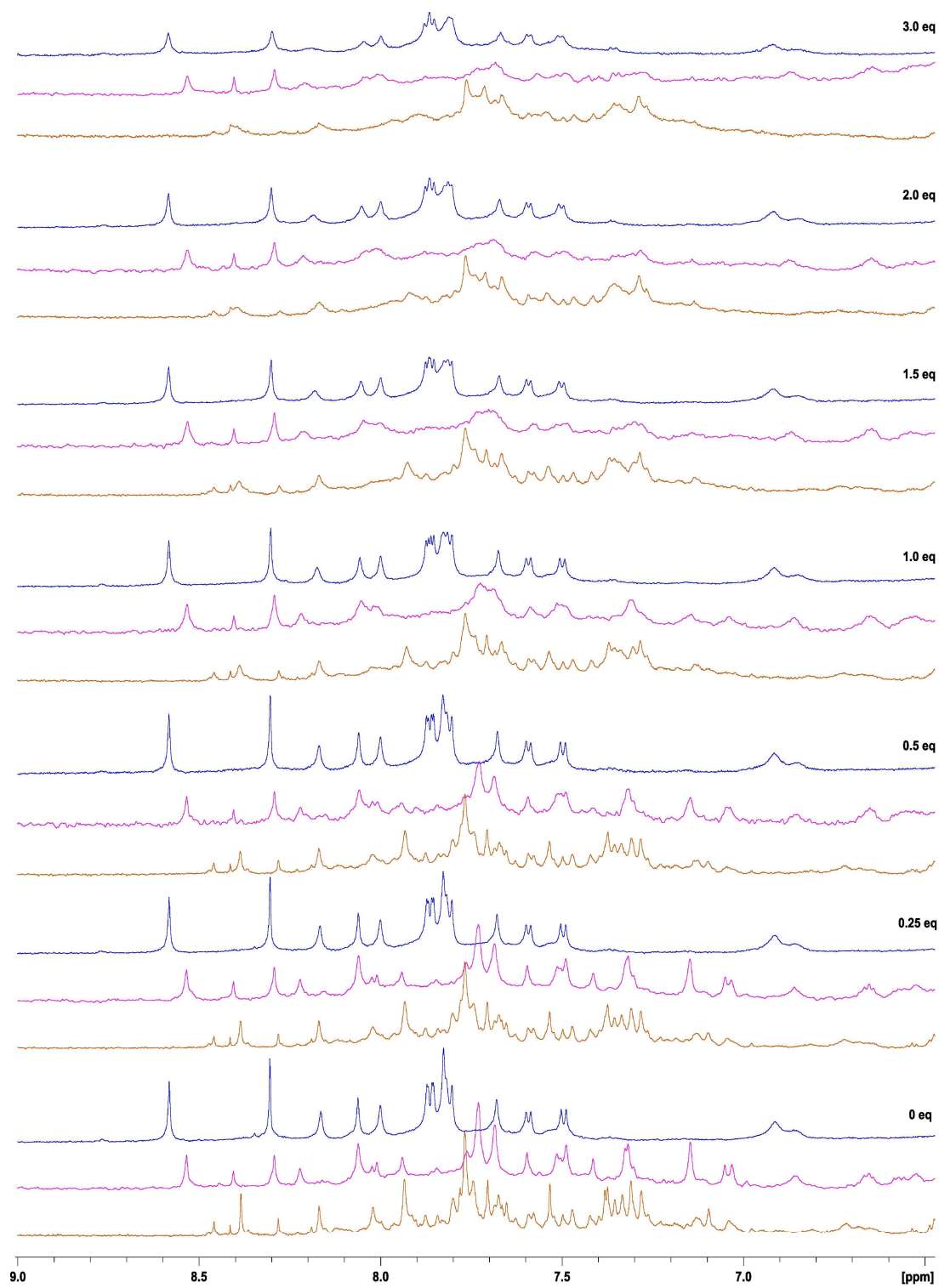


Figure 2.5.8. ¹H NMR titrations of • wtTel26, • BCL2 and • TERRA12 with increasing amount of compound AQ9 recorder at 298 K and 600 MHz with excitation sculpting water suppression (aromatic region).

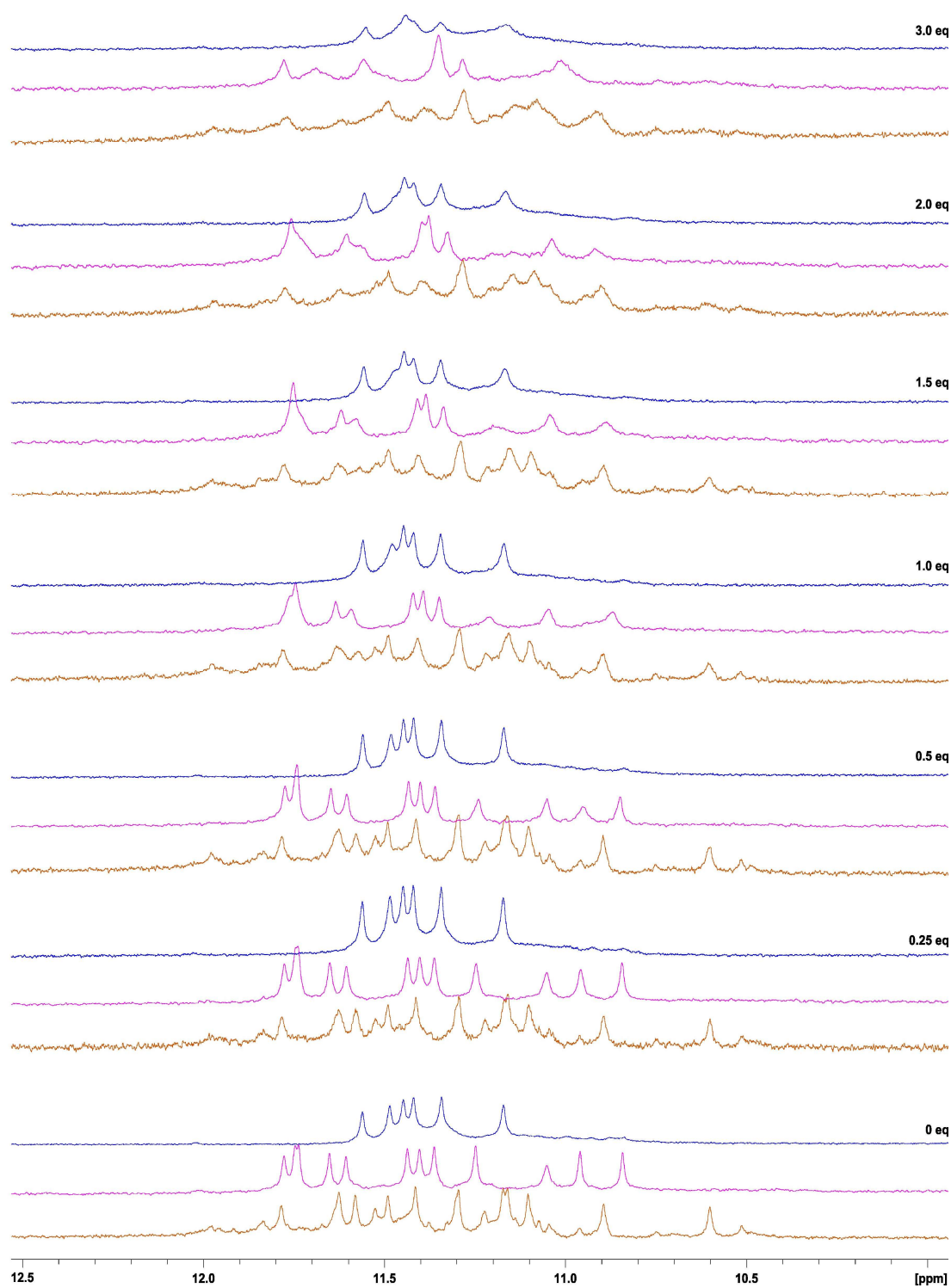
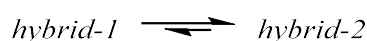


Figure 2.5.9. ¹H NMR titrations of • wtTel26, • BCL2 and • TERRA12 with increasing amount of compound AQ9 recorder at 298 K and 600 MHz with jump-return water suppression (imino region).

The ligand showed indeed binding with all the sequences with similar efficacy, although unfortunately, as for the case of rutin and quercetin, the broadening of the signal suggested an intermediate exchange which does not allow further structural studies. Noteworthy, a clear shift of the topology was not revealed by the NMR titration, even though it cannot be excluded because of the lack of sharp peaks. Therefore, we decided to investigate the binding of the ligand by CD titration also with wtTel26 to check whether the behaviour was the same as with wtTel23a. The data are reported in Figure 2.5.10a.

The initial spectra of the two sequences showed a maximum at 290 nm but the shape of the two full spectra is different. It is important to keep in mind that, when a specific topology is “assigned” to a sequence, it does not mean that it is the only topology present in solution, but that it represents the major conformation. Still, other minor conformations might still be present and it is what can lead to differences in the CD spectrum. In our study, upon ligand addition, the increase of the peak at 290 nm was detected for both the telomeric G4s and also the decrease of the signal at 260 nm, although for wtTel26 that signal was initially the maximum of the CD spectra, therefore the final shape does not resemble an antiparallel topology. However, the change in shape suggested that there could be a shift or that the ligand could preferentially bind to one of the conformations present over the others. Specifically, wtTel26 is known to fold into hybrid-2 (major) and hybrid-1 (minor). For this reason, a second CD titration was performed using Tel24, which is known to fold into hybrid-1 as major conformation. The data are presented in Figure 2.5.10b. Interestingly, upon ligand addition, no changes were detected on the shape of the CD spectra. This finding does not mean that the ligand is not binding to the NA, but that there is no difference in the folding.

The presence of two conformations in solution is described by an equilibrium, as follow:



therefore, if one of the concentrations of one of the species changes, there is normally a rearrangement of the conformations to restore the equilibrium. Hence, if the ligand would bind preferentially to hybrid-1 topology, the restore of the equilibrium will lead to a decrease of the concentration of the free hybrid-2 topology and a subsequent increase of the hybrid-1, that would be present both as free and in complex and that could explain the difference of the CD spectra. In the future, to confirm this hypothesis, an NMR titration could be performed using a sequence that folds into both topologies in a 1:1 ratio and whose signals corresponding to each topology are known, as for example Tel24.⁴⁵³

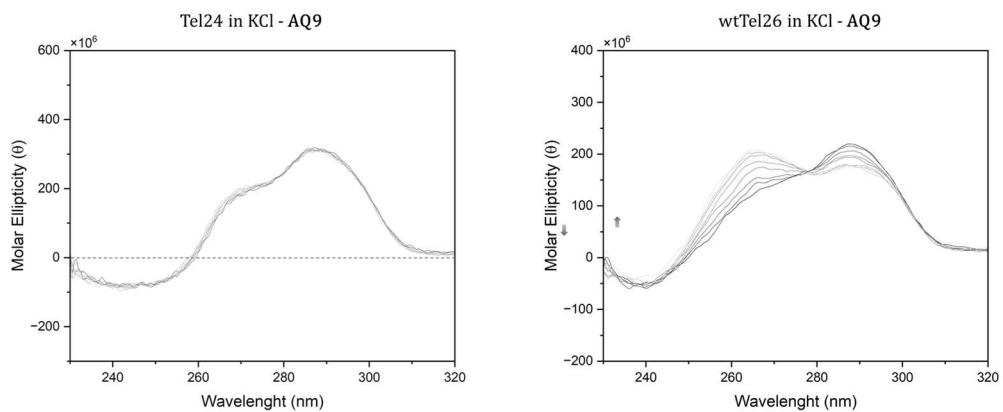


Figure 2.5.10. CD spectra of 10 μ M Tel24 and wtTel26 in 25 mM K-phosphate buffer (pH 7.0) in 70 mM KCl at increasing ligand/G4 molar ratios (the darker the line colour the higher the molar ratio).

2.6 Anthracenes

2.6.1 Introduction

Tricyclic aromatic compounds are known DNA binders, for instance drugs like doxorubicin and mitoxantrone that share an anthracenedione scaffold are thought to be able to form a ternary complex with the dsDNA and the topoisomerase II enzyme.⁴⁵⁴ Moreover, BRACO-19 which has been described in Chapter 2.1.4, has an acridine structure and is considered one of best stabilizers of the G4.⁴⁵⁵ Anthracene is also an aromatic tricyclic compound and thanks to its complete planarity, it can form pi-stacking and hydrophobic interactions, which are known features that favour the interaction with G4, as previously mentioned.⁴⁵⁶ Some examples on compounds that are demonstrated to bind to NA are presented in the following.

Anthracenyl bishydrazone (bisantrene) is an anticancer drug used clinically for years, it is known to bind strongly to dsDNA by intercalation and is an excellent topoisomerase-II poison.⁴⁵⁷ Kumar and Asuncion demonstrated that 9-aminomethylanthracene binds to DNA with high affinity,⁴⁵⁸ while Van Arman and Czamik, showed that its derivative with a second side chain that increased the amine count to a total of four enhanced the intercalation in the dsDNA.⁴⁵⁹ So far, all the presented studies showed a good possibility for the interaction with dsDNA. In 2010, Folini and colleagues reported a series of anthracene derivatives substituted with one or two 4,5-dihydro-1H-imidazol-2-yl-hydrazonic groups (the bisantrene side chain) at different positions on the aromatic system.⁴⁶⁰ The derivatives substituted in position 1,5 showed a good ability to stabilize the G4 with ΔT_m values up to 20 °C at 5 μ M concentration of the ligands. From this result, Ribaudo and colleagues developed a series of constrained bisantrene derivatives as G4 binders, and tested them through fluorometric and ESI-MS binding assays which demonstrated the tendency of the molecules to interact with G4.⁴⁶¹ Gama et al. prepared a series of anthracene-terpyridine metal complexes and investigated the role as G4 ligands achieving high selectivity for G4 over dsDNA.⁴⁶² In 2021 Coban and colleagues prepared a set of 9,10 di-substituted anthracene derivatives and probed their ability to bind to G4 by NMR spectroscopy, moreover the ligands showed good inhibitory activity on the telomerase enzyme with IC₅₀ values up to 18 μ M against HeLa cells.⁴⁶³

In a previous work of my research group, the anthracene scaffold was explored and in particular, a propargylamine side chain was added in position 9, which is protonated at physiological pH, and could be useful for targeting the negatively charged phosphates on

the NA backbone. (Figure 2.6.1a) Moreover, the addition of an alkyne moiety on the aromatic ring could increase the overall pi-conjugation of the anthracene and direct the side chain towards the G4 backbone. The compounds were studied *in silico* and their predicted pharmacokinetic properties suggested that they satisfy the main druglikeness criteria. Molecular docking showed a general preference of the molecules for the dsDNA with respect to the G4 arrangement, in term of docking score. ESI-MS and fluorescence titrations confirmed these results, however the $E_{COM}^{50\%}$ calculated through CID experiment showed a higher stability for the complexes of G4 with the ligands.⁴⁶⁴

Hence, these findings prompt us to an optimization of the scaffold to tune the selectivity towards G4. In order to do so, we thought of adding a second side chain to the scaffold in position 10 with the best performing amine, *i.e.*, hydroxy ethyl piperazine, leading to a structure that could be less favourable for the intercalation in dsDNA, leading to the preparation of compound **OAF89** (Figure 2.6.1b). The ligand was tested through ESI-MS and molecular docking on both a telomeric G4 and a dsDNA. The results have been published in a peer-reviewed journal.⁴⁶⁵

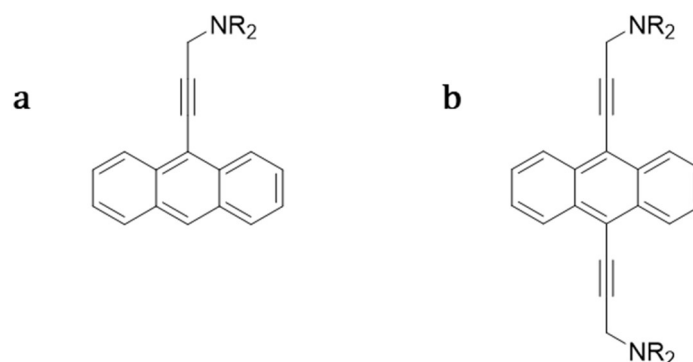
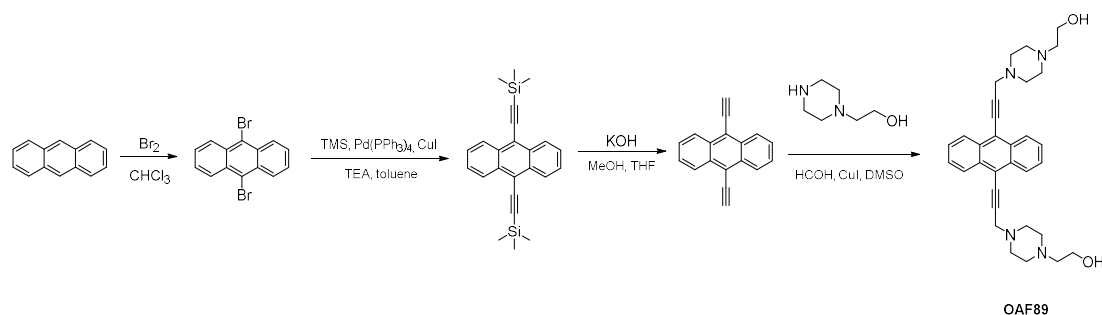


Figure 2.6.1. Chemical structure of (a) the compounds synthesised in the previous work⁴⁶⁴ substituted in position 9 and of (b) OAF89 substituted in position 9 and 10.

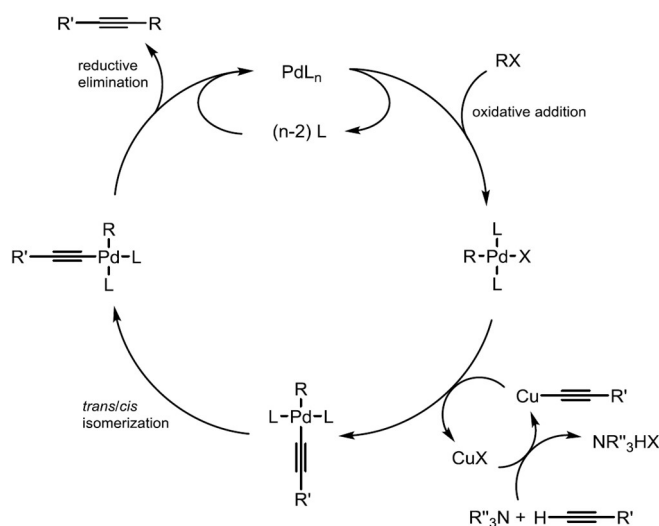
Moreover, moved by these promising results we subjected this ligand to NMR titrations and CD melting on different G4 sequences of both DNA and RNA to assess its selectivity for the topologies. Experiments are still being carried out.

2.6.2 Results and discussion

The ligand selected for this work was designed on the basis of the previous work of my research group, the anthracene scaffold was substituted with two side chains containing the amine that gave the best G4 BA in such preliminary study. In particular hydroxy ethyl piperazine can be easily mono- or di- protonated at physiological conditions and presents a primary alcohol moiety.

Figure 2.6.2. Reaction scheme of the synthesis of compound **OAF89**.

The compound **OAF89** was synthesized with a multi-step procedure reported in Figure 2.6.2, starting from anthracene, that was first di-brominated in positions 9 and 10 with elemental bromine. Subsequently, the intermediate was subjected to a double Sonogashira coupling reaction with trimethylsilyl acetylene (TMSA) on the two bromides in presence of a Pd(0) compound, Tetrakis(triphenylphosphine)palladium(0), and of CuI as the main catalyst, an exemplificative mechanism of the reaction is depicted in Figure 2.6.3.⁴⁶⁶ The TMSA groups were then deprotected in basic conditions obtaining the compounds bearing the free alkynes in position 9, and 10. The final compound **OAF89** was obtained by the Cu(I) catalysed A3 coupling reaction in the presence of formaldehyde and the secondary amine 1-(2-Hydroxyethyl)piperazine.

Figure 2.6.3. Proposed catalytic cycle for the Sonogashira reaction including Pd and Cu cycles.⁴⁶⁶

ESI-MS binding study was performed using the wtTel23a G4 sequence and dsDNA as control. The compounds showed interaction with both the NA, but the selectivity was in favour of G4 with respect to dsDNA. The G4/dsDNA selectivity ratio was calculated as 1.92, thus demonstrating that the presence of second side chain in the scaffold is able to increase the selectivity of the ligand towards the G4. CID experiments demonstrated

that the first event upon increasing energy was the loss of the small molecule indicating a possible pi-stacking binding mode. The results are reported in Table 2.6.1 (the spectra can be found in the Appendix, Section 5).

Table 2.6.1. BA values towards G4 and dsDNA, selectivity and $E_{COM}^{50\%}$, equilibrium dissociation constant K_D for G4 as calculated from the ESI-MS binding experiment and docking score.

Compound	BA G4	BA dsDNA	Selectivity ratio (BA G4/BA dsDNA)	$E_{COM}^{50\%}$ ligand:G4	K_D (μ M)	Docking score G4 (kcal/mol)
OAF89	80.3	41.8	1.92	34.39	0.69	-13.012

Molecular docking was performed of both a parallel telomeric G4 and dsDNA and revealed a good interaction of the compound with the G4 structure and moreover, confirmed that the complex with the G4 receptor is formed through an external stacking with the two side chains pointed to towards the G4 grooves (Figure 2.6.4). The calculated energy of interaction, represented by the Glide Score values which were -13.012 kcal/mol, whereas its mono-substituted analogue showed a binding energy of -10.307 kcal/mol.⁴⁶⁴ As previously mentioned, these results have been published in 2020.⁴⁶⁵

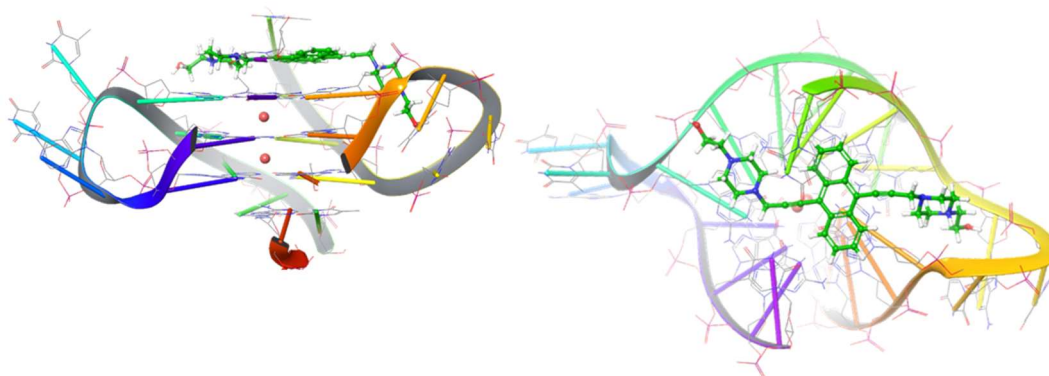


Figure 2.6.4. Docking pose obtained from molecular docking, side view (left) top view (right).

OAF89 was then tested through NMR titrations on different G4 sequences, the same used for the anthraquinones, namely DNA wtTel26, BCL2 and RNA TERRA12. The data obtained through NMR are confidential as to date they have not been published.

The compound showed overall good interaction with all the three sequences, although while for BCL2 and wtTel26, upon ligand addition there was a broadening of the signals as for the ligands previously presented (Figure 2.6.5), for TERRA12 at the beginning the signals were broadening, but with the addition of more ligand they became sharper and we could detect proper shifts of the peaks, meaning that there is a fast exchange between bound and unbound G4. (Figure 2.6.6)

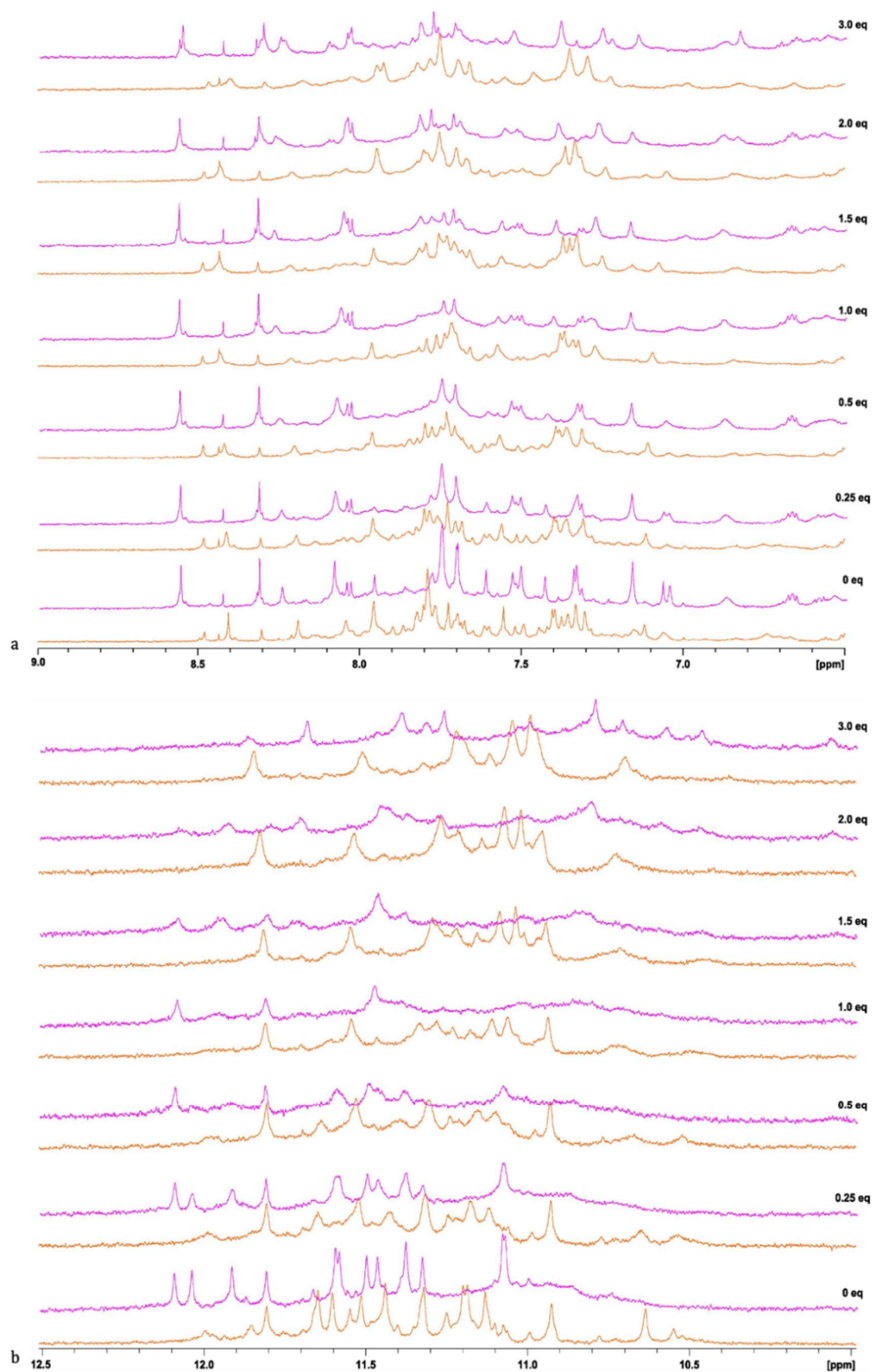


Figure 2.6.5. ^1H NMR titrations of \bullet wtTel26 and \bullet BCL2 with increasing amount of compound OAF89 recorder at 298 K and 600 MHz with (a) excitation sculpting (aromatic region) and (b) jump-return (imino region) water suppression.

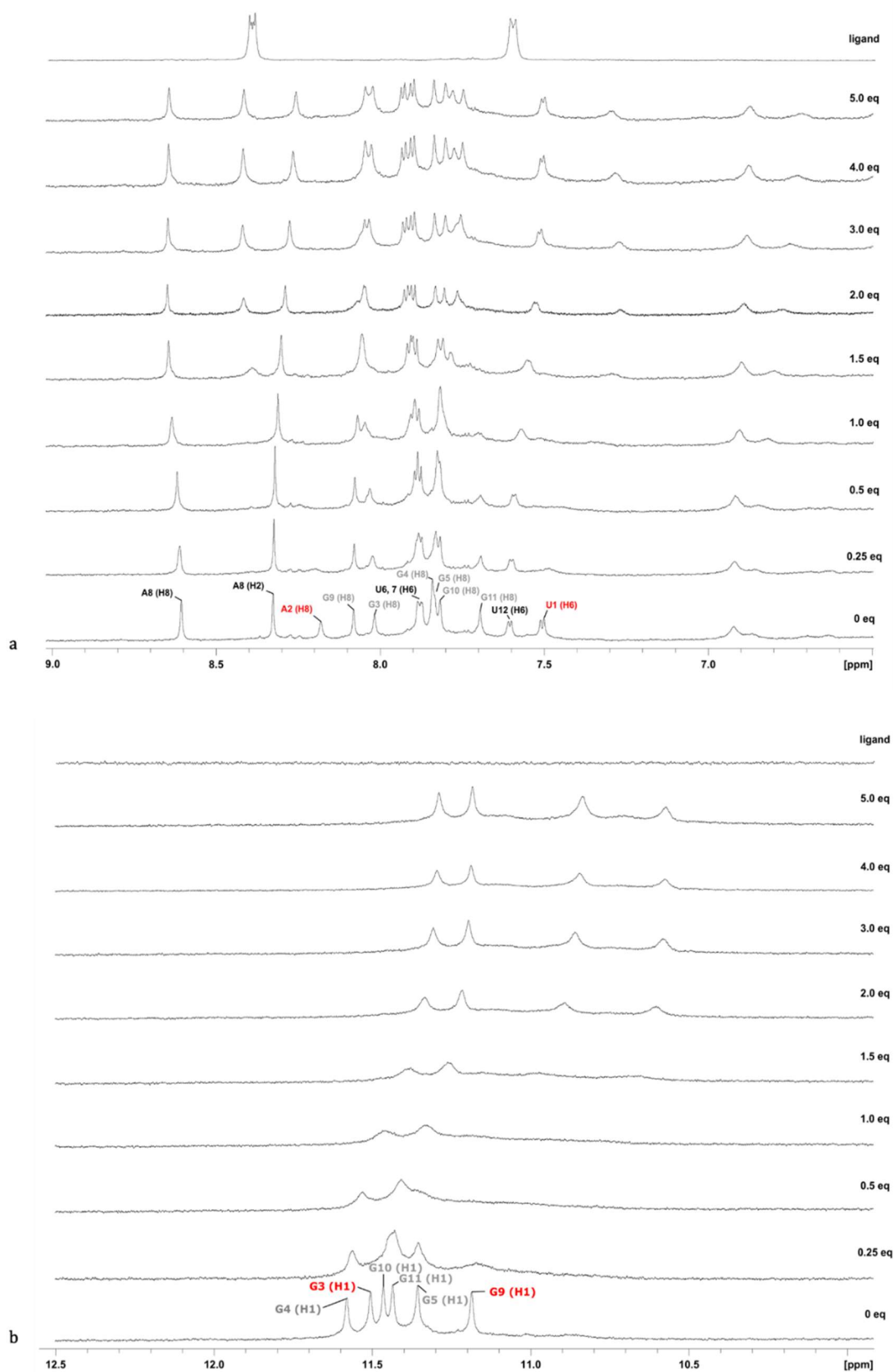


Figure 2.6.6. ^1H NMR titrations of TERRA12 with increasing amount of compound **OAF89** recorder at 298 K and 600 MHz with (a) excitation sculpting (aromatic region) and (b) with jump-return (imino region) water suppression. Marked in red the protons signals most influenced upon ligand addition, in black the protons whose signal was unvariated

As previously mentioned, when fast exchange happens, the NMR spectrum would show one peak for each proton at the average chemical shift between bound and unbound state. TERRA12 three-dimensional structure has already been resolved by NMR and published by Phan et al.²¹⁸ Transferring the assignment of the peaks to the measured spectra, we could detect and mark which of the signals are the most influenced upon ligand addition. Already by the addition of a small amount of ligand (0.25 eq) we would see some major changes of the signal corresponding to the imino proton H1 of G3 and G9 and to the aromatic H6 of U1. Moreover, also the signals corresponding to the aromatic H8 of G3 and G9 completely change throughout the titration. The scheme of the G-quadruplex is depicted in Figure 2.6.7a. These changes suggest that the ligand is interacting with the G-quadruplex structure through pi-stacking on the upper quartet as all the mentioned signals are located in this region. Moreover, the aromatic signals corresponding to the nucleotides on the groove and on the lower quartet are basically unvariated even with 5 eq of the ligand, probably meaning that **OAF89** could preferentially form a complex 1:1 with G-quadruplex and specifically bind to the upper quartet. (Figure 2.6.7b)

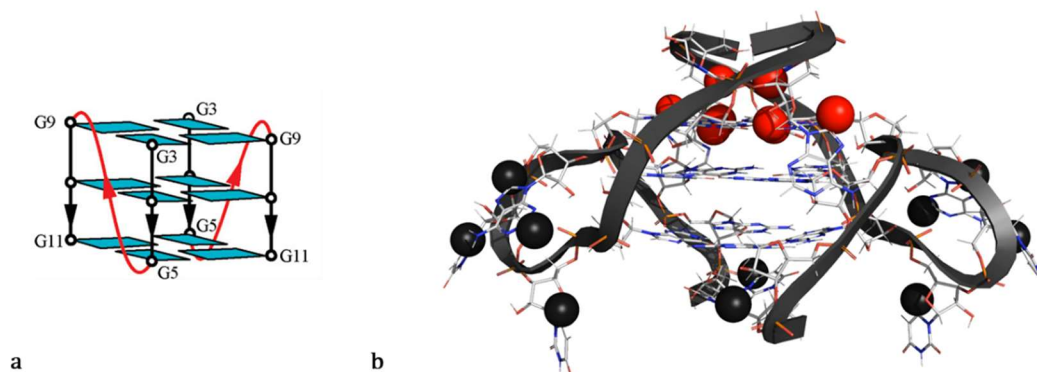


Figure 2.6.7. (a) Schematic representation of TERRA12 G-quadruplex. (b) 3D structure of TERRA12 (PDB ID: 2KBP).²¹⁸ Marked in red the protons most influenced upon ligand addition, in black the protons whose signal was unvariated.

CD melting was also performed with this ligand and TERRA12 in order to assess whether the ligand would provide an increased stability. Upon ligand addition, first the full CD spectra was recorded to determine the correct wavelength at which the arrangement has the maximum difference, then the melting curve was recorded. Interestingly, the arrangement showed a ΔT_m up to 6 °C which is in line with the best G4 binders. ΔT_m data are presented in Table 2.6.2.

Table 2.6.2. ΔT_m calculated from the MD melting curves

Conc OAF89	ΔT_m
0,5 eq	0,17
0,75 eq	0,92
1,0 eq	2,04
1,5 eq	3,19
4,0 eq	6,22
5,0 eq	6,13

Because of the promising results obtained, 2D spectra of the sample 5:1 ligand:G4 have been recorded in order to possibly calculate the 3D structure of the complex. First of all, to ensure the best conditions for analysis, we registered 1D ^1H spectra, increasing the temperature from 277 K to 318 K, and varying DMSO concentrations (5% and 10%). The results are presented in Figure 2.6.8. Focusing the attention on the imino region (~ 11.5 ppm), at lower temperature and at lower DMSO concentration only four signals are present, whereas the imino protons involved in Hoogsten base pairing which resonate in that region are six, at 318 K and 10% DMSO concentration they start appearing improving therefore signal resolution. Moreover, the signal at 7.6 ppm split into two separate signals, a duplet that is probably due to an U12 and a broader peak that is possibly due to the ligand. Hence, these designated conditions were chosen for further analysis.

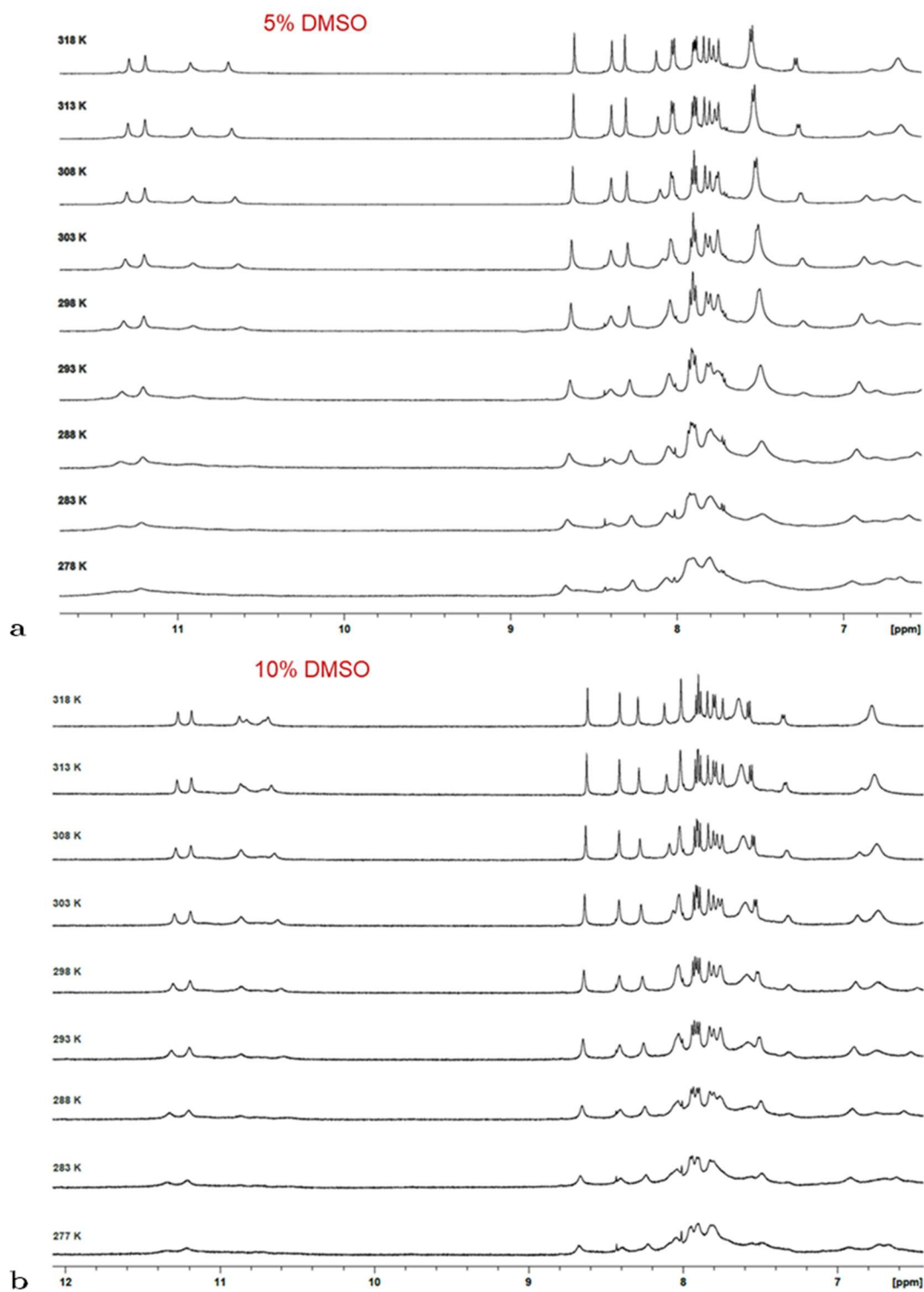


Figure 2.6.8. ¹H NMR spectra from 6.5 to 12 ppm recorded every 5 degrees from 277 K to 318 K at (a) 5% DMSO concentration and (b) 10% DMSO concentration with excitation sculping water suppression.

The 2D spectra were recorded in H₂O for imino proton assignment whose signals, thanks to their rapid exchange with the solvent, would disappear in a deuterated solvent and in D₂O for sugar proton assignment, whose signals disappear in the H₂O sample

because they resonate close to the water signal that is suppressed. Analysis of 2D ^1H - ^{13}C HSQC (Figure 2.6.9) and NOESY (Figure 2.6.10) spectra allowed the assignment of all the aromatic protons, imino protons, sugar H1' protons, and their corresponding carbons. The current focus of analysis is on interpreting COSY and TOCSY spectra to assign the signals of all other sugar protons. Namely, as previously mentioned, in the COSY experiment a cross-signal is observed between protons that are coupled to each other and that are two or three bonds apart, whereas TOCSY shows correlations between all the protons of a given spin system, the magnetization is transferred successively to coupled protons unless there is a small or zero proton-proton coupling or in the presence of heteroatoms.

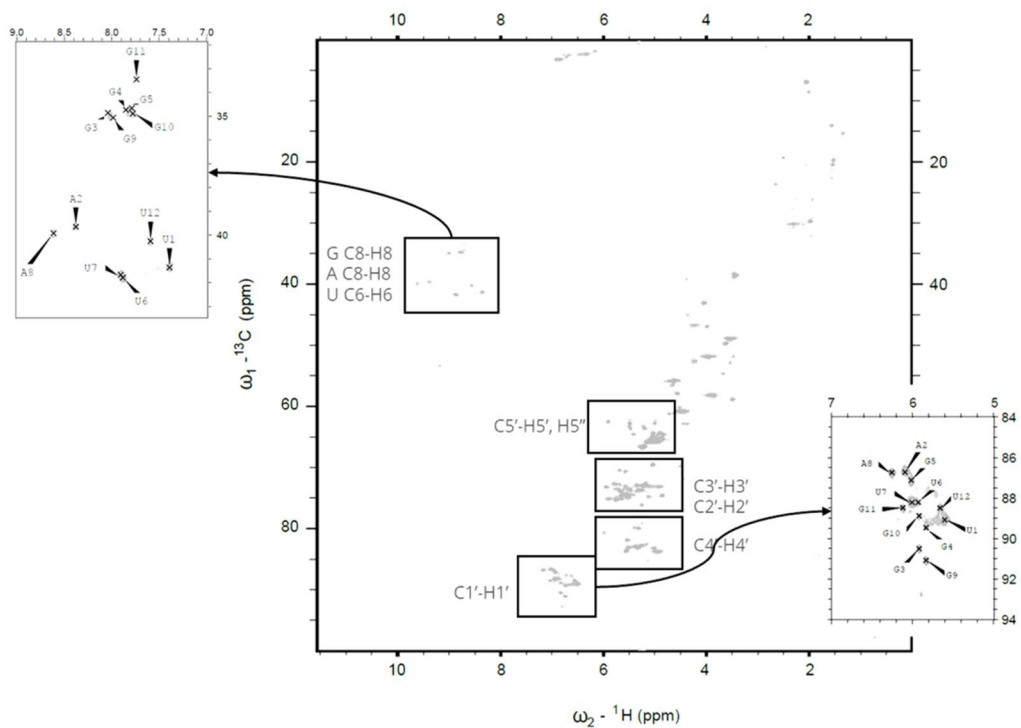


Figure 2.6.9. ^1H - ^{13}C HSQC spectra recorded at 318 K TERRA12 (1.2 mM) with 5 eq of OAF89, 25 mM of KPi, 70 mM of KCl and 10% of d6-DMSO in D_2O , detail on aromatic region and C1'-H1' region.

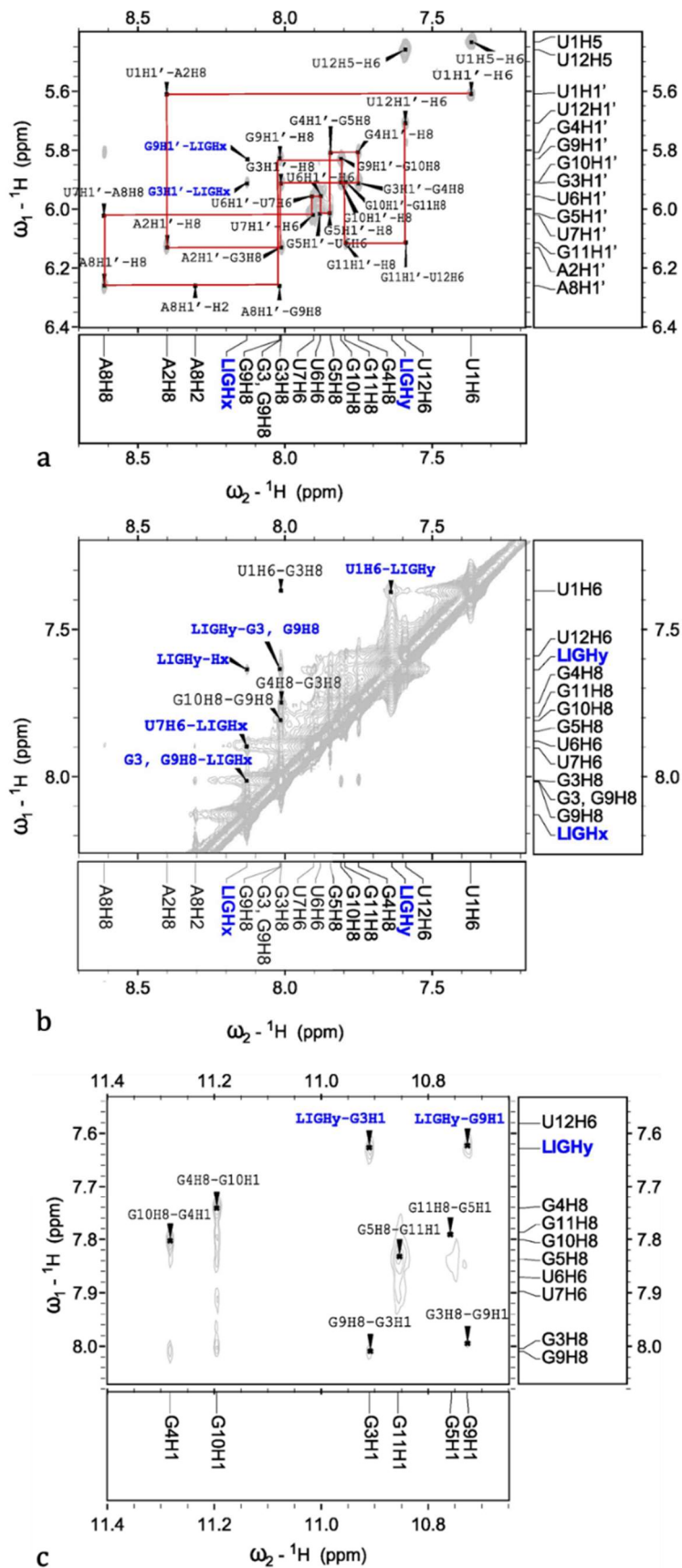


Figure 2.6.10. 2D NOESY spectra recorded at 318 K TERRA12 (1.2 mM) with 5 eq of **OAF89**, 25 mM of KP₁, 70 mM of KCl and 10% of d₆-DMSO in H₂O: (a) sugar H1'-aromatic region, (b) aromatic region, (c) imino-aromatic region.

2.7 Dimeric PEG-linked scaffolds

2.7.1 Introduction

Dimeric scaffold has demonstrated the potential to enhance the binding affinity and increase the stability of complexes with G4s.

In a study by Jain et al., the binding of a dimeric form of 1,3-phenylenebis-benzimidazole was compared to its monomer. The dimer showed superior G4 affinity, thermal stability, and telomerase inhibition activity. The BA has been studied both with prefolded G4 DNA and with ligand-induced G4 DNA. Noteworthy in presence of the ligand the G4 folding of G4 was shifted to a parallel topology. Long telomeric sequences in human genome are demonstrated to fold and stack end-to-end to form compact-stacking G4 multimers.^{423,467} They proposed that the ligand could be able to stack onto two or more consecutive G4 assemblies, stabilizing the whole structure and thus inducing the telomerase inhibition.⁴⁶⁸

Iida et al. proposed the synthesis of macrocyclic hexaoxazole dimers connected with different linker length and their ability to bind to G4 was investigated through FRET melting, showing that the linker length influence the ΔT_m . Importantly no interaction was detected with dsDNA, same as what was found for the corresponding monomers. The authors proposed that the dimer would actually bind to the G4 as a *molecular tweezer* with the two moieties stacking on the upper and lower quartets and the linker elongating on the loops because of the found 1:1 ligand:G4 stoichiometry.⁴⁶⁹

Maji and colleagues proposed the synthesis of six novel carbazole-based benzimidazole derivatives and four compounds in which these derivatives were connected with different linkers to form a dimeric molecule. The ligands ability of binding with G4 was then investigated using different spectroscopic, electrophoretic, enzymatic, and cellular studies. Finally, the possible three-dimensional structure of the complexes was further investigated with molecular modelling. All the ligands showed high G4 DNA BA with significant selectivity over the canonical arrangement. Moreover, they exhibited the potential of inhibiting telomerase and causing cancer cell selective toxicity. Noteworthy, again the dimers showed significantly higher G4 DNA binding and stabilization compared to their monomers.⁴⁷⁰

As mentioned earlier, the choice of spacer type and length can also influence the interaction with G4. The goal of the current work was to design dimeric ligands that could function as *molecular tweezers*, allowing both the aromatic moieties to stack onto the final

quartets separated by a flexible spacer that would extend into the groove. The spacer should be chosen of the correct length for the interaction with a certain topology of G4 to be selective.

As per the aromatic moieties, anthracene and anthraquinone were chosen for their ability to interact well with the G quartets thanks to their planarity which enable them to form pi-stacking, other than hydrophobic, interactions as demonstrated in the previous chapters, as described in the previous chapter. The spacer was composed of two parts: a polyethylene glycol chain to confer hydrophilicity and facilitate interactions with the backbone, and two piperazines that, through protonation, could better interact with the negatively charged phosphate groups of the backbone.

2.7.2 Results and discussion

The synthesis process, depicted in Figure 2.8.1, began with the acylation with acetyl or propionyl halides of the amine-substituted scaffolds, followed by nucleophilic substitutions of the halide with an excess of piperazine to synthesize **ANAp**, **ANPp**, **AQAp**, and **AQPp**. To obtain the final compounds, the intermediates underwent another nucleophilic substitution with the addition of half of the equivalents of di-iodo activated triethylene glycol. This led to the production of the desired dimers (**ANA3**, **ANP3**, **AQA3**, and **AQP3**), as well as the formation of monomers in which only one aromatic moiety reacted with the triethylene glycol (**ANA3m**, **ANP3m**, **AQA3m**, and **AQP3m**). After synthesis, the compounds were characterized using ESI-MS, NMR, and HPLC. To enhance the feasibility of the reactions, reduce the number of steps, and improve yields, the reaction steps were carefully optimized. Monitoring techniques, including TLC and HPLC, were employed to gain insights into side product formation and the ideal reaction termination point. This optimization was crucial in establishing a versatile synthetic pathway that allows for the preparation of dimers and monomers with various aromatic moieties and spacer lengths, potentially paving the way for future expansion of the series.

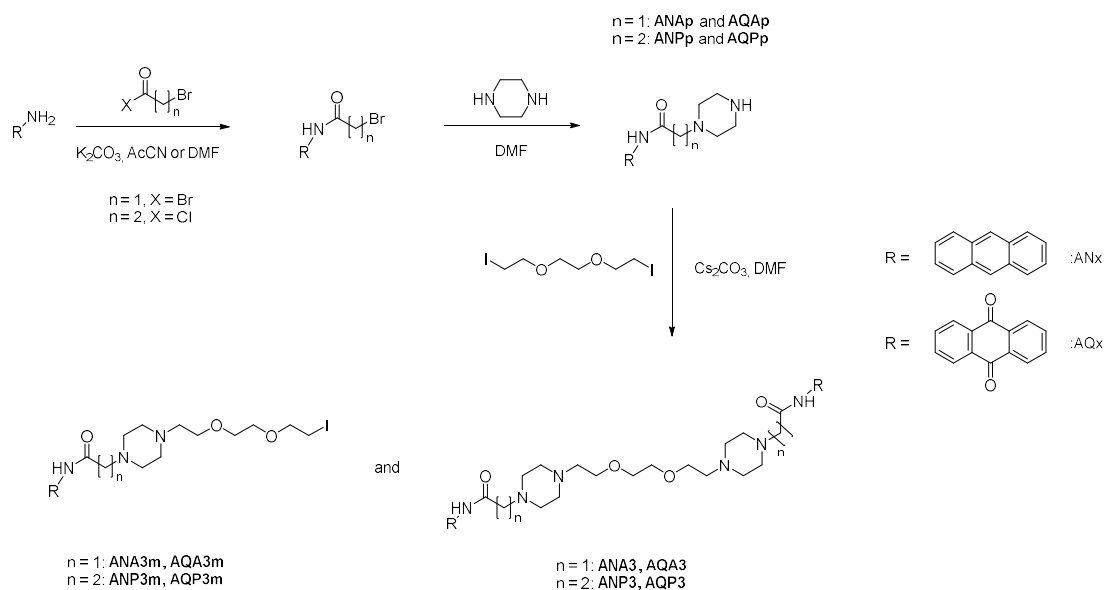


Figure 2.7.1. General scheme for the synthesis of the set of anthracene and anthraquinone dimers and monomers

The final compounds, monomers, and piperazine-derivatives underwent the ESI-MS binding assays that was described in previous reports to assess their ability to bind to dsDNA and a telomeric G4 (wtTel23a), calculating their BA and selectivity for G4 over dsDNA. Subsequent CID experiments were performed to investigate the stability of the formed complexes. The results are presented in Table 1. The most promising compound was **AQA3**, exhibiting strong BA with wtTel23a and interestingly no interaction with dsDNA. Noteworthy, increasing the spacer length by just two -CH₂- units in **AQP3** led to a loss of this ability. This observation holds promise for designing highly selective ligands, not only favouring G4 over dsDNA but also targeting specific G4 sequences. The compounds are currently being tested through fluorescence melting and CD titrations performed by the group of Prof. Sissi at the University of Padova on different G4 sequences to explore affinity and selectivity.

Table 2.7.1. BA, $E_{\text{COM}}^{50\%}$ (eV) and selectivity ratio values determined by ESI-MS interaction assays between the synthesized compounds and the wtTel23a G4 or ds DNA sequences. (“-“ is indicated when the experiment showed no interaction, the blank space indicates that the experiment was not performed, n.c. when the result of the operation was not calculable)

Compound	BA G4	$E_{\text{COM}}^{50\%}$ G4 (eV)	BA dsDNA	$E_{\text{COM}}^{50\%}$ dsDNA (eV)	Selectivity ratio BA G4 / BA dsDNA
ANAp	26	-	9	-	2.9
ANA3	-		-		
ANA3m	-		-		
AQA3p	-		-		
AQA3	30	52.4	-		n.c.
AQA3m	-		-		
AQPp	-		-		
AQP3m	-		-		
AQP3	-		-		
ANPp	-		-		
ANP3	-		-		
ANP3m	-		-		

2.8 General conclusions and future perspectives

The major aim of the study was the demonstration of the importance of the combination of different techniques in the study and understanding of the binding of small molecules with macromolecular targets. In particular, ESI-MS emerged as a valuable tool to screen large library of compounds thanks to its immediate response and the very low sample consumption. Moreover, preliminary structural insight can be gained by the analysis of the fragmentation pattern of CID experiment together with the calculation of interaction parameters, such as BA, selectivity and $E_{COM}^{50\%}$ that give an idea of the efficiency of the binding and on the stabilization on G4 arrangement provided by the ligands.

Moreover, NMR is endowed with the ability of providing insights into the binding thanks to variations of the peaks in 1D experiments and with the possibility of resolving the three-dimensional structure thanks to the integration of the cross peaks obtained in 2D experiments which can be correlated with distances.

CD spectroscopy is one of the most studied and traditional techniques for the study of G4 and is thus useful to further confirm the findings obtained with the other techniques, not to mention to gain important information of the topology assumed by the G4 and thermal stabilization.

Computational tools are also extremely valuable, not only because they are used to further refine and calculate the three-dimensional structure of the complexes, but also because they can be used priorly to other experiments to assess whether the desired interactions are indeed possible and favourable and to have a first idea of the binding energy of the complexes formed.

In particular the study of flavonoids showed that natural compounds are endowed with the ability of targeting DNA special arrangements and therefore their antiproliferative activity could be partially explained by the inhibition of telomerase activity thanks to the stabilization of G4s that interfere with its binding with the single stranded end of telomeres. Furthermore, isoflavones derived from *M. Pomifera* served as starting compounds to identify new G4 ligands and to increase the selectivity for G4 through semi-synthetic modifications. For this study, a combination of ESI-MS, NMR and extensive computational study was used to obtain structural elucidation.

In our investigation, anthraquinone was confirmed to be an important scaffold for targeting G4 and in particular the preparation of 2,6-diammido anthraquinone series demonstrated a marked ability to interact with the DNA and specifically a high selectivity

for G4 over dsDNA through ESI-MS, CD, NMR and FRET melting. Noteworthy, the use of click-chemistry approach proved to be an important improvement for the synthesis of new aminoacid-anthraquinone derivatives both for the known feasibility of the reaction that give high yields avoiding the use of hazardous chemicals and for the possibility to add a triazole moiety that add polarity of the side chains. Moreover, CD spectroscopy changes prompted us to investigate more in detail a possible switch of the topology or the preference for one of the conformations present in solutions, thus explaining the shift observed in the spectra. In this context, 1D NMR titrations needs to be performed with a sequence resolved for both the hybrid-1 and hybrid-2 conformations that could be important to provide further selectivity between various G4 sequences.

In Chapter 2.6, the optimization of the anthracene derivatives with the addition of two side chains containing an alkyne and a tertiary amine permitted to tune the selectivity towards G4 upon dsDNA. Molecular docking showed that indeed **OAF89** bind through stacking interaction and its side chains elongate on the loops thanks to H-bonds formation between the protonated amine and the negatively charged phosphate groups of the backbone. 1D NMR titrations with TERRA12 showed shift of the DNA peaks which were further investigated with 2D experiments performed after the assessment of the best analysis conditions. Currently, the analysis and assignment of the peaks is still ongoing and new titrations involving the addition of TERRA12 to a solution containing the ligand will be conducted in order to obtain unambiguous ligand signals. Once all the signals are assigned, NOESY spectra will be recorded at various mixing times to correlate the obtained area with a distance in space. These results will be used in conjunction with molecular dynamics simulations to generate the complex's structural restrains. Furthermore, new CD titrations on TERRA12 with the ligand will be needed to enable the calculations of the dissociation constant of the complex and determine the stability of the complex resulting from ligand addition.

This outcome holds particular significance as, to date, the structure of TERRA12 G4 in combination with a ligand has not been reported.

Finally, anthracene and anthraquinone scaffolds have been further studied in the context of the development of new dimeric ligands that could act as *molecular tweezer*. This new set of DNA ligands was synthesized using an efficient approach based on acylation and nucleophilic substitution. To enhance the feasibility of the reactions, reduce the number of steps, and improve yields, the reaction steps were carefully optimized. This optimization was crucial in establishing a versatile synthetic pathway that allows for the

preparation of dimers and monomers with various aromatic moieties and spacer lengths, potentially paving the way for future expansion of the series. Mass spectrometry binding experiments confirmed that some of the compounds bind to the G4 wtTel23a sequence. Interestingly, increasing the spacer length by just two -CH₂- units led to a loss of this ability. This observation holds promise for designing highly selective ligands, not only favouring G4 over dsDNA but also targeting specific G4 sequences. To verify this initial hypothesis, further studies are being conducted assessing the binding towards different G4 topologies.

2.9 Materials and methods

2.9.1 Chemistry

Commercially available chemicals were purchased from Sigma-Aldrich and used without any further purification if not specified elsewhere. NMR characterization were performed on a Bruker Avance III 400 spectrometer (frequencies: 400.13 and 100.62 MHz for ^1H and ^{13}C nuclei, respectively) equipped with a multinuclear inverse z-field gradient probe head (5 mm). For data processing, TopSpin 4.0.8 software was used, and the spectra were calibrated using solvent signal (^1H -NMR, $\delta_{\text{H}} = 7.26$ ppm for CDCl_3 , $\delta_{\text{H}} = 3.31$ ppm for CD_3OD , $\delta_{\text{H}} = 2.50$ ppm for d6-DMSO; ^{13}C -NMR, $\delta_{\text{C}} = 77.16$ ppm for CDCl_3 , 39.52 ppm for d6-DMSO, $\delta_{\text{C}} = 49.00$ for CD_3OD). Multiplicities are reported as follows: s, singlet; d, doublet; t, triplet; q, quartet; m, multiplet; b, broad; dd, doublet of doublets. Mass spectra were recorded by direct infusion ESI on a LCQ Fleet ion trap (Thermo Fisher Scientific, Waltham, MA, USA) mass spectrometer. For data processing, Qual Browser Thermo Xcalibur 4.0.27.13 software was used. ESI parameters for samples acquired in positive ionization mode: spray voltage 3.2 kV, capillary temperature 160 °C, capillary voltage 43 V. ESI parameters for samples acquired in negative ionization mode: spray voltage 5.0 kV, capillary temperature 160 °C, capillary voltage -8 V.

The characterization of the compounds can be found in the Appendix (Section 5).

Extraction of rutin⁴¹¹

The extraction procedure was adapted from the one previously reported by Paniwnyk *et al.*⁴⁷¹ Dried flower buds (20 g) were crushed and inserted in a round-bottom flask. Methanol (200 mL) was then added and the mixture was refluxed for 1 hour under vigorous stirring. After cooling to room temperature, solid components were filtered off. The obtained bright green liquid fraction was evaporated under reduced pressure. The resulting residue was treated with water, giving a pale green/yellow suspension that was washed with diethyl ether to remove chlorophyll. The aqueous suspension was then filtered under reduced pressure to give the product as a light green solid, which was further purified by crystallization from ethanol (0.8 g).

^1H -NMR (400 MHz, CD_3OD) δ_{H} : 7.67 (d, 1H, $J = 2.2$ Hz), 7.63 (dd, 1H, $J = 8.4, 2.2$ Hz), 6.87 (d, 1H, $J = 8.4$ Hz), 6.39 (d, 1H, $J = 2.2$ Hz), 6.21 (d, 1H, $J = 2.2$ Hz), 5.11 (d, 1H, $J = 7.5$ Hz), 4.52 (d, 1H, $J = 1.6$ Hz), 3.81 (dd, 1H, $J = 10.8, 1.0$ Hz), 3.64 (dd, 1H, $J = 3.4, 1.6$ Hz), 3.54 (dd, 1H, $J = 9.5, 3.4$ Hz), 3.5-3.3 (m, 5H), 3.3-3.2 (m, 2H), 1.13 (d,

3H, $J = 6.1$ Hz); ^{13}C -NMR (101 MHz, CD_3OD) δ_{C} ppm: 179.5, 166.1, 163.1, 159.5, 158.6, 149.9, 146.0, 135.8, 123.7, 123.3, 117.8, 116.2, 105.8, 104.9, 102.6, 100.1, 95.0, 78.3, 77.3, 75.9, 74.1, 72.4, 72.2, 71.5, 69.8, 68.7, 18.0. (ESI) found 609.52 ($\text{C}_{27}\text{H}_{29}\text{O}_{16}^-$, $[\text{M}-\text{H}]^-$), calc. 609.15.

Synthesis of N,N'-(9,10-dioxo-9,10-dihydroanthracene-2,6-diyl)bis(2-azidoacetamide) (AQ1)⁴³⁹

2,6-Diaminoanthraquinone (1 eq, 500 mg, 2.10 mmol) was dissolved in chloroacetyl chloride (24 eq, 5.70 g, 50 mmol) and the mixture was heated at 110 °C for 4 h until a colour change from violet to yellow was observed. The mixture was cooled to r.t. and diethyl ether was added to precipitate the solid product that was filtered and washed with diethyl ether. Without further purification the solid was dissolved in DMF and NaN_3 was added (6 eq), the mixture was stirred at r.t. for 16 h and then water was added. The resulting precipitate was filtered, washed with water and evaporated at reduced pressure to afford a yellow solid (680 mg, 80%).

^1H -NMR (400 MHz, d_6 -DMSO) δ_{H} 10.78 (s, 2H), 8.44 (d, 2H, $J = 2.2$ Hz), 8.19 (d, 2H, $J = 8.5$ Hz), 8.05 (dd, $J_1 = 8.5$ Hz, $J_2 = 2.2$ Hz, 2H), 4.15 (s, 4H); ^{13}C -NMR (101 MHz, d_6 -DMSO) δ_{C} 182.12, 170.43, 144.14, 135.03, 128.97, 128.50, 124.16, 116.42, 45.71, (ESI⁺) m/z calcd for $\text{C}_{18}\text{H}_{13}\text{N}_8\text{O}_4^+$ $[\text{M}+\text{H}]^+$: 405.10, found: 405.51.

Synthesis of N,N'-(9,10-dioxo-9,10-dihydroanthracene-2,6-diyl)bis(3-azidopropanamide) (AQ2)⁴³⁹

2,6-Diaminoanthraquinone (1 eq, 200 mg, 0.84 mmol) was refluxed in 3-azidopropionyl chloride (23 eq, 2.60 g, 19.5 mmol) at 100 °C for 4 hours. The solution changed from violet to yellow in colour. After 4 hours the mixture was cooled down to r.t. and diethyl ether (30 mL) was added to precipitate the product as a solid. The mixture was filtered under vacuum, washed with diethyl ether, and left to dry obtaining a yellow solid (258 mg, 71%).

^1H -NMR (400 MHz, d_6 -DMSO) δ_{H} 10.73 (s, 2H), 8.46 (d, 2H, $J = 1.8$ Hz), 8.19 (d, 2H, $J = 8.5$ Hz), 8.05 (dd, $J_1 = 8.5$ Hz, $J_2 = 1.8$ Hz, 2H), 3.66 (t, 2H, $J = 6.3$ Hz), 2.73 (t, 2H, $J = 6.3$ Hz); ^{13}C -NMR (101 MHz, d_6 -DMSO) δ_{C} 181.79, 170.22, 144.93, 134.87, 129.08, 128.61, 123.93, 116.31, 47.07, 36.23; (ESI⁺) m/z calcd for $\text{C}_{20}\text{H}_{17}\text{N}_8\text{O}_4^+$ $[\text{M}+\text{H}]^+$: 433.14, found: 433.62.

General synthesis of the final compounds (AQ3-AQ9)⁴³⁹

Propargylamide amino-acid derivatives were prepared as reported in literature.^{441,472-474} Compounds **AQ1** or **AQ2** (1 eq) and the opportune propargylamide amino-acid derivative (4 eq) were dissolved in DMSO (1 mL). In a separate vial, sodium ascorbate (1.2 eq), anhydrous CuSO₄ (0.5 eq) and K₂CO₃ (0.3 eq) were dissolved in distilled water (200 μL). The salts solution was added to the reaction mixture under an N₂ stream, stoppered, and left to stir overnight at room temperature. Water was added and the mixture left to stir for 10 min in an ice bath. The mixture was filtered using a Buchner funnel and the brown solid obtained were washed with water and ethyl acetate and dried to obtain a brown crystal solid. Without further purifications, the solid was dissolved in an equal amount of DCM and TFA, the solution was left to stir for 2 hours. The solvents were evaporated under reduced pressure and left to dry overnight in vacuum affording a crystalline solid.

AQ3 *N,N'*-(9,10-dioxo-9,10-dihydroanthracene-2,6-diyl)bis(2-(4-((2-aminoacetamido)methyl)-1H-1,2,3-triazol-1-yl)acetamide)

The product was obtained with the procedure described above yielding a brown solid (364 mg, 86%).

¹H-NMR (400 MHz, d₆-DMSO) δ_H 11.20 (s, 2H), 8.87 (brs, 2H), 8.48 (d, 2H, *J* = 2.1 Hz), 8.21 (d, *J* = 8.6 Hz, 2H), 8.11-7.97 (m, 9H), 5.44 (s, 4H), 4.45 (d, 4H, *J* = 4.2 Hz), 3.59 (m, 4H); ¹³C-NMR (101 MHz, d₆-DMSO) δ_C 181.07, 165.66, 165.20, 143.76, 134.27, 128.54, 128.36, 124.36, 123.55, 115.88; (ESI⁺) *m/z* calcd for C₂₈H₃₀N₁₁O₆⁺ [M+H]⁺: 629.23, found: 629.37; calcd for C₂₈H₃₁N₁₁O₆²⁺ [M+2H]²⁺: 315.12, found: 315.23.

AQ4 *(2S,2'S)*-*N,N'*-((((9,10-dioxo-9,10-dihydroanthracene-2,6-diyl)bis(azanediy))bis(2-oxoethane-2,1-diyl))bis(1H-1,2,3-triazole-1,4-diyl))bis(methylene))bis(2,6-diaminohexanamide)

The product was obtained with the procedure described above yielding a brown solid (558 mg, 92%).

¹H-NMR (400 MHz, d₆-DMSO) δ_H 11.21 (s, 2H), 8.97 (m, 2H), 8.47 (d, 2H, *J* = 2.1 Hz), 8.29-8.12 (m, 7H), 8.04 (s, 2H), 8.00 (dd, 2H, *J*₁ = 8.6 Hz, *J*₂ = 2.1 Hz), 7.73 (brs, 5H), 5.42 (s, 4H), 4.45 (m, 2H), 2.76 (m, 4H), 1.72 (m, 4H), 1.53 (m, 4H), 1.32 (m, 4H); ¹³C-NMR (101 MHz, d₆-DMSO) δ_C 181.46, 168.52, 165.50, 144.01, 134.56, 128.90, 128.70, 124.95, 123.99, 116.24, 52.40, 52.22, 38.69, 34.39, 30.61, 26.65, 21.28; (ESI⁺) *m/z*

calcd for $C_{36}H_{47}N_{14}O_6^+$ $[M+H]^+$: 771.38, found: 771.57; calcd for $C_{36}H_{48}N_{14}O_6^{2+}$ $[M+2H]^{2+}$: 386.19, found: 386.41.

AQ5 (2*S*,2'*S*)-*N,N'*-((((9,10-dioxo-9,10-dihydroanthracene-2,6-diyl)bis(*azanediy*l))bis(2-oxoethane-2,1-diyl))bis(1*H*-1,2,3-triazole-1,4-diyl))bis(methylene))bis(2-amino-3-phenylpropanamide)

The product was obtained with the procedure described above yielding a brown solid (456 mg, 89%).

1H -NMR (400 MHz, *d*₆-DMSO) δ_H 11.22 (s, 2H), 8.96 (brs, 2H), 8.49 (d, 2H, $J = 2.1$ Hz), 8.29-8.16 (m, 7H), 8.04 (s, 2H, $J_1 = 8.6$ Hz, $J_b = 2.1$ Hz), 7.39-7.11 (m, 10H), 5.43 (s, 4H), 4.40 (brs, 4H), 4.01 (m, 2H), 3.08 (m, 1H), 2.99 (m, 1H); ^{13}C -NMR (101 MHz, *d*₆-DMSO) δ_C 181.08, 167.66, 165.17, 143.78, 134.69, 134.28, 129.33, 128.55, 128.38, 124.50, 123.56, 115.90, 53.34, 52.15, 36.90, 34.08; (ESI⁺) m/z calcd for $C_{42}H_{41}N_{12}O_6^+$ $[M+H]^+$: 809.33, found: 809.46; calcd for $C_{42}H_{42}N_{12}O_6^{2+}$ $[M+2H]^{2+}$: 405.17, found: 405.33.

AQ6 (2*S*,2'*S*)-*N,N'*-((((9,10-dioxo-9,10-dihydroanthracene-2,6-diyl)bis(*azanediy*l))bis(2-oxoethane-2,1-diyl))bis(1*H*-1,2,3-triazole-1,4-diyl))bis(methylene))bis(2,5-diaminopentanamide)

The product was obtained with the procedure described above yielding a brown solid (447 mg, 93%).

1H -NMR (400 MHz, *d*₆-DMSO) δ_H 11.23 (s, 2H), 9.01 (t, 2H, $J = 5.2$ Hz), 8.47 (d, 2H, $J = 2.1$ Hz), 8.33- 8.16 (m, 8H), 8.07 (s, 2H), 8.02 (dd, 2H, $J_1 = 8.7$ Hz, $J_2 2.1$ Hz), 7.80 (brs, 6H), 5.44 (s, 4H), 4.52 (dd, 2H, $J_a = 15.0$ Hz, $J_b = 5.5$ Hz), , 4.39 (dd, 2H, $J_a = 15.0$ Hz, $J_b = 5.5$ Hz), 3.81 (m, 2H), 2.81 (m, 4H), 1.76 (m, 4H), 1.59 (m, 4H); ^{13}C -NMR (101 MHz, *d*₆-DMSO) δ_C 181.12, 168.10, 165.28, 143.80, 134.32, 128.60, 128.41, 124.68, 123.61, 115.94, 52.17, 51.55, 38.18, 34.20, 28.06, 22.49; (ESI⁺) m/z calcd for $C_{34}H_{43}N_{14}O_6^+$ $[M+H]^+$: 743.35, found: 743.62, calcd for $C_{34}H_{44}N_{14}O_6^{2+}$ $[M+2H]^{2+}$: 372.18, found: 372.52.

AQ7 (2*S*,2'*S*)-*N,N'*-((((9,10-dioxo-9,10-dihydroanthracene-2,6-diyl)bis(*azanediy*l))bis(2-oxoethane-2,1-diyl))bis(1*H*-1,2,3-triazole-1,4-diyl))bis(methylene))bis(*azanediy*l))bis(2-oxoethane-2,1-diyl))bis(2,6-diaminohexanamide)

The product was obtained with the procedure described above yielding a brown solid (584 mg, 88%).

$^1\text{H-NMR}$ (400 MHz, $d_6\text{-DMSO}$) δ_{H} 11.27 (s, 2H), 8.77 (brs, 2H), 8.57 (brs, 2H), 8.47 (brs, 2H), 8.31- 8.12 (m, 8H), 8.02 (brs, 4H), 7.83 (brs, 6H), 5.43 (s, 4H), 4.40 (m, 2H), 3.83 (m, 8H), 2.76 (m, 4H), 1.72 (m, 4H), 1.52 (m, 4H), 1.38 (m, 4H); $^{13}\text{C-NMR}$ (101 MHz, $d_6\text{-DMSO}$) δ_{C} 181.07, 168.68, 167.98, 143.80, 134.26, 128.51, 128.35, 124.60, 123.57, 115.90, 52.12, 51.83, 41.68, 38.31, 34.00, 30.22, 26.30, 20.84; (ESI⁺) m/z calcd for $\text{C}_{50}\text{H}_{43}\text{N}_{16}\text{O}_8^{2+}$ $[\text{M}+2\text{H}]^{2+}$: 443.21 found 443.65.

AQ8 (2*S*,2'*S*)-*N,N'*-((((((9,10-dioxo-9,10-dihydroanthracene-2,6-diyl)bis(*azanediy*l))bis(2-oxoethane-2,1-diyl))bis(1*H*-1,2,3-triazole-1,4-diyl))bis(methylene))bis(*azanediy*l))bis(2-oxoethane-2,1-diyl))bis(2,6-diaminohexanamide)

The product was obtained with the procedure described above yielding a brown solid (616 mg, 91%).

$^1\text{H-NMR}$ (400 MHz, $d_6\text{-DMSO}$) δ_{H} 11.24 (s, 2H), 8.53-8.45 (m, 5H), 8.27-8.09 (m, 8H), 8.07-7.96 (m, 4H), 7.89-7.71 (m, 6H), 5.42 (s, 4H), 4.35 (brs, 4H), 3.39 (m, 1H), 3.31 (m, 1H), 2.75 (brs, 4H), 2.35 (brs, 4H), 1.66 (m, 4H), 1.50 (m, 4H), 1.29 (m, 4H); $^{13}\text{C-NMR}$ (101 MHz, $d_6\text{-DMSO}$) δ_{C} 181.69, 170.55, 168.77, 165.86, 144.41, 134.88, 129.13, 128.96, 124.18, 116.51, 52.75, 52.42, 38.93, 35.84, 34.63, 30.93, 26.99, 21.56; (ESI⁺) m/z calcd for $\text{C}_{42}\text{H}_{58}\text{N}_{16}\text{O}_8^{2+}$ $[\text{M}+2\text{H}]^{2+}$: 457.23, found: 457.55.

AQ9 (2*S*,2'*S*)-*N,N'*-((((((9,10-dioxo-9,10-dihydroanthracene-2,6-diyl)bis(*azanediy*l))bis(3-oxopropane-3,1-diyl))bis(1*H*-1,2,3-triazole-1,4-diyl))bis(methylene))bis(2,6-diaminohexanamide)

The product was obtained with the procedure described above yielding a brown solid (546 mg, 94%).

$^1\text{H-NMR}$ (400 MHz, $d_6\text{-DMSO}$) δ_{H} 10.81 (s, 2H), 8.93 (brs, 2H), 8.45 (d, 2H, $J = 1.9$ Hz), 8.25-8.09 (m, 8H), 8.06-7.96 (m, 4H), 7.83-7.71 (brs, 5H), 4.67 (t, 4H, $J = 6.2$ Hz), 4.38 (m, 4H), 3.73 (m, 2H), 3.11 (t, 4H, $J = 6.2$ Hz), 2.74 (m, 4H), 1.70 (m, 4H), 1.50 (m, 4H), 1.30 (m, 4H); $^{13}\text{C-NMR}$ (101 MHz, $d_6\text{-DMSO}$) δ_{C} 181.26, 169.18, 168.25, 144.38, 134.32, 128.53, 128.07, 123.42, 115.77, 51.92, 45.10, 38.41, 36.22, 34.22, 30.36, 26.41, 21.08; (ESI⁺) m/z calcd for $\text{C}_{38}\text{H}_{51}\text{N}_{14}\text{O}_6^{2+}$ $[\text{M}+\text{H}]^{2+}$: 400.21, found: 400.60.

Synthesis of 9,10-Dibromoanthracene⁴⁶⁵

Anthracene (1 eq, 4.5 g, 25.0 mmol) was dissolved in CHCl_3 (50 mL) and a solution of bromine (1.2 eq, 3.0 mL, 58.2 mmol) in CHCl_3 (25 mL) was added dropwise, inducing the formation of a solid precipitate. The mixture was stirred at r.t. for 4 h and afterwards

the solvent was removed with a stream of N₂. The solid residue was triturated with DCM, filtered and washed with small portions of DCM to give 1 in yellow needles (4.1 g, 36%).

¹H-NMR (400 MHz, CDCl₃): δ_H (ppm) = 8.57 (m, 4H), 7.62 (m, 4H); ¹³C-NMR (100 MHz, CDCl₃): δ_C (ppm) = 131.2, 128.4, 127.6, 123.7.

Synthesis of 9,10-Bis((trimethylsilyl)ethynyl)anthracene⁴⁶⁵

9,10-Dibromoanthracene (1 eq, 236 mg, 0.70 mmol) and trimethylsilyl acetylene were dissolved in a mixture of toluene (37.0 mL) and triethylamine (8.4 mL). Tetrakis(triphenylphosphine)palladium(0) (0.02 eq, 17 mg, 0.015 mmol) and copper iodide (0.04 eq, 5.5 mg, 0.029 mmol) were then added. The mixture was purged with nitrogen and heated at reflux overnight. The solvent was removed at reduced pressure and the obtained mixture was poured in water and extracted with DCM, dried over anhydrous magnesium sulphate and evaporated under reduced pressure. The crude product was purified by column chromatography (silica gel, hexane) obtaining compound 2 as a red solid (86 mg, 36%). ¹H-NMR (400 MHz, CDCl₃): δ_H (ppm) = 8.61 (m, 4H), 7.63 (m, 4H), 0.47 (s, 18H); ¹³C-NMR (100 MHz, CDCl₃): δ_C (ppm) = 132.1, 127.2, 126.4, 117.5, 108.3, 101.7, 0.2.

Synthesis of 9,10-Diethynylanthracene⁴⁶⁵

9,10-Bis((trimethylsilyl)ethynyl)anthracene (80 mg, 0.22 mmol) was dissolved in a mixture of MeOH (6.4 mL) and THF (3.2 mL), and 10% KOH was then added (1.2 mL). The solution was stirred at r.t. for 3 h and then the solvent was partially evaporated under reduced pressure. After that, diethyl ether (20 mL) was added, the obtained mixture was washed with water (3 × 10 mL), dried over magnesium sulphate and evaporated under reduced pressure obtaining a grey solid (45 mg, 91%). ¹H-NMR (400 MHz, CDCl₃): δ_H (ppm) = 8.61 (m, 4H), 7.62 (m, 4H), 4.07 (s, 2H); ¹³C-NMR (100 MHz, CDCl₃): δ_C (ppm) = 132.5, 127.2, 127.10, 117.5, 90.1, 80.3.

Synthesis of 9,10-Bis[(4-(2-hydroxyethyl)piperazine-1-yl)prop-2-yne-1-yl]anthracene (OAF89)⁴⁶⁵

9,10-Diethynylanthracene (1 eq, 45 mg, 0.2 mmol), 37% formaldehyde (200 μL, 2.5 mmol), copper iodide (0.2 eq, 7 mg, 0.04 mmol) and 1-(2-Hydroxyethyl)piperazine (1.5 eq, 80 mg, 0.6 mmol) were dissolved in DMSO (1 mL). The solution was stirred overnight at r.t. and then DCM (30 mL) was added. The resulting mixture was washed with 0.1 M

KOH (4 × 10 mL). The crude product was purified by column chromatography (silica gel, gradient of MeOH/TEA/DCM from 2,5/0,75/96,75 to 10/2/88) obtaining a solid yellow product (25 mg, 25%). ¹H-NMR (400 MHz, d₆-DMSO): δ_H (ppm) = 8.54 (m, 4H), 7.74 (m, 4H), 4.36 (t, 2H, *J* = 5.4 Hz), 3.90 (s, 4H), 3.50 (m, 4H), 2.72 (bs, 8H), 2.51 (bs, 8H), 2.41 (t, 4H, *J* = 6.2 Hz); ¹³C-NMR (100 MHz, CDCl₃): δ_C (ppm) = 132.8, 128.7, 127.2, 117.0, 95.4, 82.2, 59.5, 55.4, 52.9, 51.6, 48.2; (ESI⁺): calcd for [C₃₂H₃₉N₄O₂]⁺ [M + H]⁺ 511.31, found 511.04

General procedure for the synthesis of the first intermediates: ANA, ANP, AQA and AQP

2-aminoanthracene (1 eq) or 1-aminoanthraquinone (1 eq) were dissolved in 3.5 mL of AcCN with K₂CO₃ (1 eq). The acyl halide (1.3 eq) was added dropwise to the flask and the reaction mixture was stirred at r.t. overnight. The reaction was monitored with TLC with the proper eluent. When the starting material appeared consumed and 10 mL of H₂O were

added to quench the reaction and the precipitate was filtered with a Büchner funnel and washed with ethanol (5 mL) and hexane (3*2 mL) in order to remove the excess of starting materials and possible by-products. The solid was characterized with ESI-MS and NMR.

ANA (*N*-(anthracen-2-yl)-2-bromoacetamide)

Yield: 121 mg, 74%. ¹H-NMR (400 MHz, d₆-DMSO) δ_H 10.58 (brs, 1H), 8.46-8.43 (m, 3H), 8.02-7.98 (m, 3H), 7.50-7.48 (m, 1H), 7.44-7.41 (m, 2H), 4.07 (s, 2H); (ESI⁺) *m/z* calcd for C₁₆H₁₃BrNO⁺ [M+H]⁺ : 314.02 and 316.02, found: 314.34 and 316.34.

ANP (*N*-(anthracen-2-yl)-3-bromopropanamide)

Yield: 142 mg, 83%. ¹H-NMR (400 MHz, d₆-DMSO) δ_H 10.32 (brs, 1H), 8.52-8.45 (m, 3H), 8.06-8.03 (m, 3H), 7.58-7.55 (m, 1H), 7.51-7.44 (m, 2H), 3.79 (t, *J* = 6.35 Hz, 2H), 3.05 (t, *J* = 6.35 Hz, 2H); (ESI⁺) *m/z* calcd for C₁₇H₁₅BrNO⁺ [M+H]⁺ : 328.03 and 330.03, found: 327.98 and 329.98.

AQA (2-bromo-*N*-(9,10-dioxo-9,10-dihydroanthracen-1-yl)acetamide)

Yield: 76 mg, 49%. ¹H-NMR (400 MHz, CDCl₃) δ_H 12.94 (brs, 1H), 9.14 (dd, *J*₁ = 1.13 Hz, *J*₂ = 8.53 Hz, 1H), 8.35-8.28 (m, 2H), 8.08 (dd, *J*₁ = 1.13 Hz, *J*₂ = 7.63 Hz, 1H), 7.83-

7.79 (m, 3H), 4.10 (s, 2H); (ESI⁺) m/z calcd for C₁₆H₁₁BrNO₃⁺ [M+H]⁺ : 343.99 and 345.99, found: 343.15 and 345.15.

AQP (*3-bromo-N-(9,10-dioxo-9,10-dihydroanthracen-1-yl)propenamide*)

Yield: 84 mg, 52%. ¹H-NMR (400 MHz, CDCl₃) δ_H 12.47 (brs, 1H), 9.14 (dd, J₁ = 1.03 Hz, J₂ = 8.56 Hz, 1H), 8.30-8.26 (m, 2H), 8.13 (dd, J₁ = 1.03 Hz, J₂ = 7.69 Hz, 1H), 7.82-7.76 (m, 3H), 3.79 (t, J = 6.35 Hz, 2H), 3.05 (t, J = 6.35 Hz, 2H); (ESI⁺) m/z calcd for C₁₇H₁₃BrNO₃⁺ [M+H]⁺ : 358.01 and 360.01, found: 358.87 and 360.87.

General procedure for the synthesis of the second intermediates: ANAp, ANPp, AQAp and AQPp

Piperazine (10 eq) was dissolved in 2 mL of DMF (2 ml) heated up to about 55 °C in order to partially dissolve piperazine. After cooling to r.t., intermediate 1 (1 eq) was added and the reaction was monitored with TLC with the proper eluent (DCM:MeOH:TEA, 92:7.5:0.5). After 3-4 hours, the starting material was consumed and 10 mL of H₂O were added to quench the reaction, the precipitate was filtered with a Büchner funnel. The product was then characterized with ESI-MS and ¹H and ¹³C NMR.

ANAp (*N-2-(anthracen-2-ylamino)-2-oxoethyl)piperazine-1-carboxylate*)

Yield: 123 mg, 99%. ¹H-NMR (400 MHz, CDCl₃): δ_H (ppm): 8.84 (s, 1H); 8.43 (s, 1H); 8.38 (s, 2H), 8.01-7.97 (m, 3H); 7.49-7.43 (m, 3H); 3.36-3.34 (m, 6H); 2.97 (t, 4H, J = 9.64 Hz, 2*CH₂); ¹³C-NMR (101 MHz, CDCl₃): δ_C (ppm): 168.8; 134.4; 132.4; 132.0; 131.3; 129.4; 129.3; 128.3; 128.1; 126.2; 125.8; 125.7; 125.2; 120.5; 115.3; 62.9; 46.4; 29.8; (ESI⁺): m/z calcd for [C₂₀H₂₂N₃O]⁺ [M+H]⁺ 320.42, found 320.25.

ANPp (*N-(anthracen-2-yl)-3-(piperazin-1-yl)propanamide*)

Yield: 136 mg, 95%. ¹H-NMR (400 MHz, CDCl₃): δ_H (ppm): 11.28 (s, 1H); 8.48 (s, 1H); 8.36 (d, 2H, J = 7.95 Hz); 7.95 (t, 3H, J = 17.14 Hz); 7.46-7.38 (m, 3H); 3.08 (m, 4H); 2.79-2.60 (m, 8H); ¹³C-NMR (101 MHz, CDCl₃): δ_C (ppm): 171.0; 135.5; 132.3; 132.3; 131.1; 129.3; 129.2; 128.3; 128.1; 126.1; 125.8; 125.6; 125.0; 120.8; 115.2; 54.5; 53.7; 46.5; 32.5. (ESI⁺): calcd for [C₂₁H₂₄N₃O]⁺ [M+H]⁺ 334.49, found 334.44.

AQAp (*N-(9,10-dioxo-9,10-dihydroanthracen-1-yl)-2-(piperazin-1-yl)acetamide*)

Yield: 79 mg, 78%. ¹H-NMR (400 MHz, CDCl₃): δ_H (ppm): 13.21 (s, 1H); 9.23 (d, 1H, J = 8.46); 8.31 (dd, 2H, J₁ = 28.12 Hz, J₂ = 6.46); 8.09 (d, 1H, J = 7.34); 7.83-7.75 (m, 3H);

3.24 (s, 2H); 3.18 (m, 4H); 2.66 (m, 4H); ^{13}C -NMR (101 MHz, CDCl_3): δ_{C} (ppm): 186.3; 183.1; 171.6; 141.3; 135.6; 134.4; 134.3; 134.3; 132.9; 127.5; 127.1; 126.4; 122.8; 118.7; 63.6; 55.1; 46.0; (ESI⁺): m/z calc for $[\text{C}_{20}\text{H}_{20}\text{N}_3\text{O}_3]^+$ [M+H]⁺ 350.40, found 350.41.

AQPp (*N*-(9,10-dioxo-9,10-dihydroanthracen-1-yl)-3-(piperazin-1-yl)propenamide)

Yield: 172 mg, 76%. ^1H -NMR (400 MHz, CDCl_3): δ_{H} (ppm): 12.39 (s, 1H); 9.11-9.09 (m, 1H); 8.28-8.26 (m, 2H); 8.07-8.04 (m, 1H); 7.82-7.74 (m, 3H); 3.03 (t, 2H, $J = 9.86$ Hz); 2.87-2.85 (m, 4H); 2.74-2.70 (m, 2H); 2.64-2.55 (m, 4H); ^{13}C -NMR (101 MHz, CDCl_3): δ_{C} (ppm): 187.2; 182.8; 171.9; 141.9; 135.8; 134.5; 134.1; 132.9; 127.5; 127.2; 126.6; 122.7; 118.1; 54.4; 53.3; 45.4; 36.3; (ESI⁺): calcd for $[\text{C}_{21}\text{H}_{22}\text{N}_3\text{O}_3]^+$ [M+H]⁺ 364.42, found 364.30.

General procedure for the synthesis of the final compounds: the dimers ANA3, ANP3, AQA3, AQP3 and the monomers ANA3m, ANP3m, AQA3m, AQP3m

The intermediates 2 (1 eq) were dissolved in 1 mL of DMF, Cs_2CO_3 (1.5 eq), and PEG3-I (0.4 eq) were added to the solution that was stirred at 100 °C. The reaction was monitored with TLC with the proper eluent (DCM:MeOH:TEA, 92:7.5:0.5). After 2 hours, 10 mL of H_2O were added to quench the reaction. Subsequently, the reaction mixture was centrifuged at 4000 g for 5 minutes, the supernatant was discarded and the procedure was repeated 3 times. The solid was then washed with a solution of MeOH:H₂O (3:7) and centrifuged with the same parameters applied before (3 times). The solid fraction was collected after removal of the liquid. The crude products were purified by column chromatography (silica gel, DCM:MeOH 9:1), the fractions corresponding to the monomer and the dimers for each compound were collected and the product were characterized with ESI-MS and ^1H and ^{13}C NMR.

ANA3 (*2,2'*-(((ethane-1,2-diylbis(oxy))bis(ethane-2,1-diyl))bis(piperazine-4,1-diyl))bis(*N*-(anthracen-2-yl)acetamide))

Yield: 7.5 mg 11%. ^1H -NMR (400 MHz, CDCl_3): δ_{H} (ppm): 9.31 (s, 2H); 8.45 (s, 2H); 8.36 (d, 4H, $J = 8.35$ Hz); 7.98-7.95 (m, 6H); 7.46-7.42 (m, 6H); 3.67-3.64 (m, 8H); 3.20 (s, 4H); 2.71-2.61 (m, 20H); ^{13}C -NMR (101 MHz, CDCl_3): δ_{C} (ppm): 168.7; 134.4; 132.4; 132.1; 131.3; 129.4; 129.3; 128.3; 128.1; 126.2; 125.8; 125.2; 120.5; 115.3; 70.5; 69.1; 62.2; 57.8; 53.9; 29.8; (ESI⁺) calcd for $[\text{C}_{46}\text{H}_{53}\text{N}_6\text{O}_4]^+$ [M+H]⁺ 753.97, found 753.42.

ANA3m (*N*-(anthracen-2-yl)-2-(4-(2-(2-(2-hydroxyethoxy)ethoxy)ethyl)piperazin-1-yl)acetamide)

Yield: 2 mg, 2%. Unfortunately, the amount was not enough for the characterization through NMR. (ESI⁺) calcd for [C₂₆H₃₄N₃O₄]⁺ 452.57, found 452.43.

ANP3 (*3,3'*-(((ethane-1,2-diylbis(oxy))bis(ethane-2,1-diyl))bis(piperazine-4,1-diyl))bis(*N*-(anthracen-2-yl)propanamide))

Yield: 13 mg, 18%. ¹H-NMR (400 MHz, MeOD): δ_H (ppm): 10.25-10.22 (m, 2H, 2*NH); 8.58-8.31 (m, 6H); 8.07-7.96 (m, 6H); 7.57-7.43 (m, 6H); 3.83-3.48 (m, 28H); 3.33-3.19 (m, 4H); 2.92-2.98 (m, 4H). Note: CH₂ signals were partially overlapped with the signal of the solvent; ¹³C-NMR (101 MHz, MeOD): δ_C (ppm): 165.3; 133.9; 133.4; 132.9; 130.9; 130.4; 129.5; 129.1; 127.5; 127.3; 127.0; 126.7; 126.5; 122.2; 117.3; 71.5; 69.2; 57.6; 52.6; 52.5; 51.5; 50.1; 37.4; 32.1; 9.5; (ESI⁺) calcd for [C₄₈H₅₇N₆O₄]⁺ [M+H]⁺ 782.02, found 781.74.

ANP3m *N*-(anthracen-2-yl)-3-(4-(2-(2-(2-iodoethoxy)ethoxy)ethyl)piperazin-1-yl)propanamide

Yield: 17 mg, 16%. ¹H-NMR (400 MHz, CDCl₃): δ_H (ppm): 10.96 (brs, 1H) 8.48 (s, 1H); 8.34 (d, 2H, *J* = 5.33 Hz); 7.95-7.92 (m, 3H); 7.45-7.38 (m, 3H); 3.69-3.62 (m, 8H); 2.86-2.60 (m, 16H). ¹³C-NMR (101 MHz, CDCl₃): δ_C (ppm): 170.7; 135.4; 132.3; 132.2; 131.1; 129.2; 128.3; 128.0; 126.1; 125.7; 125.6; 125.0; 120.8; 115.3; 72.0; 70.6; 70.4; 70.3; 69.2; 68.8; 63.6; 57.7; 53.7; 53.5; 52.1; 32.7; 21.1; (ESI⁺) calcd for [C₂₇H₃₅IN₃O]⁺ [M+H]⁺ 576.50, found 576.36.

AQA3 (*2,2'*-(((ethane-1,2-diylbis(oxy))bis(ethane-2,1-diyl))bis(piperazine-4,1-diyl))bis(*N*-(9,10-dioxo-9,10-dihydroanthracen-1-yl)acetamide))

Yield: 12.3 mg 16%, 12.3 mg. ¹H-NMR (400 MHz, CDCl₃): δ_H (ppm): 13.19 (s, 1H); 13.14 (s, 1H); 9.21-9.17 (m, 2H); 8.30-8.23 (m, 4H); 8.08-8.04 (m, 2H); 7.81-7.73 (m, 6H); 3.81-3.79 (m, 4H); 3.76-3.64 (m, 10H); 3.26 (s, 2H); 3.25 (s, 2H); 2.83 (m, 4H); 2.75 (m, 6H); 2.68-2.66 (m, 4H). ¹³C-NMR (101 MHz, CDCl₃): δ_C (ppm): 183.0; 171.5; 171.0; 155.5; 141.3; 135.7; 135.6; 134.4; 134.3; 132.8; 127.5; 127.4; 127.1; 126.4; 126.4; 122.9; 122.8; 118.6; 70.7; 70.6; 69.8; 64.7; 62.9; 58.0; 53.6; 53.5; 29.8. (ESI⁺) calcd for [C₄₆H₄₈N₆Na₂O₈]⁺ [M+2 Na]⁺ 858.90, found 857.93.

AQA3m

N-(9,10-dioxo-9,10-dihydroanthracen-2-yl)-2-(4-(2-(2-(2-hydroxyethoxy)ethoxy)ethyl)piperazin-1-yl)acetamide

Yield: 6%, 5.8 mg. ¹H-NMR (400 MHz, CDCl₃): δ_H (ppm): 13.17 (s, 1H); 9.24-9.22 (m, 1H); 8.32-8.27 (m, 2H); 8.08 (s, 1H); 7.83-7.75 (m, 3H); 3.76-3.62 (m, 12H); 3.29 (s, 2H); 2.96-2.82 (m, 8H); ¹³C-NMR (101 MHz, CDCl₃): δ_C (ppm): 186.4; 183.0; 171.3; 141.3; 135.6; 134.4; 134.3; 134.3; 132.9; 129.9; 127.4; 127.1; 126.5; 122.8; 118.6; 72.8; 70.5; 70.4; 62.7; 61.8; 29.8. (ESI⁺) calcd for [C₂₆H₃₂N₃O₆]⁺ [M+H]⁺ 482.56, found 482.76.

AQP3 (3,3'-(((ethane-1,2-diylobis(oxy))bis(ethane-2,1-diylobis(piperazine-4,1-diylobis(N-(9,10-dioxo-9,10-dihydroanthracen-1-yl)propenamide)))

Yield: 2,5 mg, 5%. Unfortunately, the amount was not enough for the characterization through NMR. (ESI⁺) calcd for [C₄₈H₅₃N₆O₈]⁺ [M+H]⁺ 841.99, found 841.71.

AQP3m

N-(9,10-dioxo-9,10-dihydroanthracen-2-yl)-3-(4-(2-(2-(2-iodoethoxy)ethoxy)ethyl)piperazin-1-yl)propanamide

Yield: 7.9 mg, 11%. ¹H-NMR (400 MHz, CDCl₃): δ_H (ppm): 12.40 (s, 1H); 9.11 (dd, 1H, *J*₁ = 8.53 Hz, *J*₂ = 1.24 Hz); 8.30-8.27 (m, 2H); 8.09-8.06 (m, 2H); 7.83-7.80 (m, 2H); 3.74 (t, 2H, *J* = 12.07 Hz); 3.70-3.60 (m, 8H); 2.88 (t, 2H, *J* = 14.06 Hz); 2.74 (t, 2H, *J* = 14.06 Hz); 2.69-2.60 (m, 10H, 5* CH₂). ¹³C-NMR (101 MHz, CDCl₃): δ_C (ppm): 187.1; 182.9; 171.9; 142.0; 135.8; 134.5; 134.2; 134.1; 133.0; 129.9; 127.4; 127.2; 126.6; 122.7; 118.1; 71.5; 70.8; 70.5; 69.3; 68.9; 57.8; 53.9; 53.5; 52.9; 42.8; 41.1; 36.4; 29.8; 29.1; (ESI⁺) calcd for [C₂₇H₃₃IN₃O₅]⁺ [M+H]⁺ 606.48, found 606.39.

2.9.2 ESI-MS binding studies

DNA sequences were obtained from Sigma-Aldrich (Milan, Italy). Samples were heat-denatured and folded in 150 mM ammonium acetate before incubation with the ligands. Stock solutions of the compounds were prepared in methanol. The final concentration of the oligonucleotide was 5-10 μM in 150 mM ammonium acetate, with a 10:1 compound/oligo ratio. Samples were acquired after an equilibration time of 30 minutes and MeOH was added to the samples to obtain a stable ESI signal. Mass spectra were recorded by direct infusion ESI on a Thermo Fisher Scientific (Waltham, MA) LCQ Fleet ion trap mass spectrometer. The instrument was set in negative ionization mode with a 3.4 kV capillary voltage, 120 °C capillary temperature and a flow rate of 5 μL/min. CID experiments were performed on the complexes by isolating the precursor ion in the trap.

The fragmentation was promoted by increasing the “normalized collision energy” parameter. For data processing, Qual Browser Thermo Xcalibur 4.0.27.13 software was used. Exact mass for the G4 sequence (5'-AGGGTTAGGGTTAGGGTTAGGGT3'): 7270.774 Da. Exact mass for dsDNA (5'-ACTATTTACGTATAATGA-3', 5'-TCATTATACGTAAATAGT-3'): 10987.922 Da.

2.9.3 NMR titrations

Sample preparation

The DNA sequences were purchased by Eurofins MWG Operon (Ebersberg, Germany) as HPSF® (High Purity Salt Free) purified oligos, TERRA12 sequence was purchased by Dharmacon (Cambridge, UK) as 2'-protected oligos, the deprotection was achieved adding 400 µL of 2'-deprotection buffers (100 mM acetic acid, adjusted to pH 3.8 with TEMED) until the RNA was completely dissolved, the solution was incubated at 60°C for 30 minutes and lyophilize to dryness. The oligomers have been HPLC purified with tetrabutylammonium acetate buffer, desalted with ultracentrifuge filtration devices (Vivacon (2 kDa cut-off for DNA, 1 kDa cut-off for RNA), VWR) and precipitated with 5 volumes of LiClO₄ (2% in Acetone (w/v)) at -20 °C over night. Subsequently, another desalting step has been performed to remove residual Li⁺ ions and acetone.

NMR experiments

The DNAs were diluted to the appropriate concentration with ddH₂O with the addition of 70 mM KCl and 25 mM of K-phosphate buffer (K₂HPO₄-KH₂PO₄) at pH 7.0 in H₂O, the folding was achieved heating the sample at 95 °C for 8 minutes and left to cool at room temperature for at least 3 hours. Subsequently 5-10% d₆-DMSO was added for the lock and to increase ligand solubility and 0.15 mM of 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) as internal reference ($\delta_{\text{H}} = 0$ ppm). The sample was transferred into a Shigemi tube for the analysis. The spectra were recorded at Bruker BioSpin GmbH at a 600 MHz or 800 MHz at 298 K with 128 scans and 16 dummy scans, for each ligand addition two spectra were recorded with gradient-assisted excitation sculpting⁴⁷⁵ or jump-return⁴⁷⁶ for water suppression.

2.9.4 2D NMR experiments

All NMR samples were referenced with DSS and prepared in 25 mM K-phosphate buffer and 70 mM KCl (pH = 7.0). The samples were prepared in 10% d₆-DMSO and in 90% either H₂O or D₂O. D₂O samples were prepared by repeated freeze-drying and dissolving in D₂O (99.98% D). The final NMR samples contained 1.2 mM RNA (0.6 mM G4) and 3.0 mM **OAF89** in NMR buffer. The 2D spectra were recorded at 700, 800 or 900 MHz

¹H,¹³C-HSQC and NOESY spectra with different water suppression schemes (presaturation for D₂O samples and excitation sculpting with gradients for H₂O samples) and different mixing times (150 and 250 ms) were recorded at 318 K.

COSY and TOCSY spectra were recorded with presaturation water suppression and phase sensitive with double quantum filter were recorded at 318 K.

All NMR data were collected, processed, and analysed using the software TopSpin 4.1.4 (Bruker) and Sparky 3.43.

2.9.5 Circular dichroism (CD)

For the CD titrations on wtTel26 and Tel24, the concentration of NA was 10 μM and the buffer conditions were the same of the NMR sample, namely 25 mM of K-phosphate buffer, 70 mM of KCl and 5 % of DMSO on a Jasco J-810 spectropolarimeter equipped with a Peltier-type temperature control system using a 1 mm path length cell. Before data acquisition, the NA sequence (4 μM) was heated at 95 °C for 5 min and left to cool at room temperature overnight. Spectra were then acquired in the absence and in the presence of increasing concentration of tested ligands. The reported spectrum of each sample represents the average of 3 scans recorded with 1 nm step resolution. Observed ellipticities were converted to mean residue ellipticity $[\theta] = \text{deg} \times \text{cm}^2 \times \text{dmol}^{-1}$ (Mol. Ellip.).

2.9.6 Molecular docking for rutin and quercetin ⁴¹¹

The structure of the macromolecular target was obtained from the RCSB Protein Data Bank (www.rcsb.org, PDB ID: 3CE5). Target and ligands were prepared for the blind docking experiments which were performed using AutoDock Vina (Molecular Graphics Laboratory, Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA, USA).⁴⁷⁷ Output data (energies, interaction patterns) were analysed and scored using UCSF Chimera molecular viewer,⁴⁷⁸ which was

also used to produce the artworks. Values are expressed in kcal/mol and refer to the most favoured predicted pose.

2.9.7 Molecular docking for *M. Pomifera* derived flavonoids, anthraquinones and anthracenes^{439,465}

G4 structure files were retrieved from the RCSB Protein Data Bank (PDB, www.rcsb.org) and were prepared with the Protein Preparation Wizard included in the Schrodinger suite,⁴²⁹ using default settings, i.e., adding hydrogens, removing waters further than 5 Å from the ligand, adjusting charges, capping termini, optimizing hydrogen bond clusters and performing a final minimization under the OPLS3e force field.⁴⁷⁹ The ligands were prepared for docking with Ligprep tool under OPLS3e force field, using Epik ionizer all possible states in the 7±2 pH range were generated, including the possible metal binding states, allowing therefore the interaction with the potassium ions. The docking protocol consisted of rigid-receptor/flexible ligand docking, which was performed with Glide, under the OPLS3e force field, with default settings, i.e., a 0.80 scaling factor for the ligand atom van der Waals radius and 0.15 for the partial charge cut-off, with generation of 1 pose per ligand. A cubic grid (30x30x30 Å) was prepared with Receptor Grid Generation tool included in Schrodinger suite and generated manually selecting, as centre the two potassium ions in the inner channel of the G4 quartets.

For *M. Pomifera* derivatives, different parameters, such as grid dimension, docking precision (standard precision is labelled as SP, whereas higher precision docking is marketed as XP) and sampling mode were changed throughout the study to investigate their influence on the computation. Five cubic grids were prepared with Receptor Grid Generation tool included in Schrodinger suite and generated manually selecting, as centre the two potassium ions in the inner channel of the G4 quartets. The dimension of the grids was varied as mentioned earlier and the third potassium ion included in the 3D structure was either left unmodified or deleted. The settings used for each calculation are summarised in table S2.9.1.

G4 structure preparation

For *M. Pomifera* derivatives we used PDB ID: 7KLP,⁴²⁸ whereas for the anthracene **OAF89** PDB ID: 1KF1.²⁹³ The grids were generated manually selecting the potassium ions in the inner channel.

For anthraquinones the choice was to retrieve PDB ID: 1KF1,²⁹³ 143D,⁴⁵⁰ 2HY9,⁴⁵¹ and 2Jpz,⁴²³ which are all telomeric arrangements formed by the sequence d[G₃(TTAGGG)₃], all head and tail caps were removed resulting in a 21mer. Two central K⁺ ions were placed between the three G-tetrads for the hybrid-1, hybrid-2 and antiparallel NMR resolved structures (PDB ID: 2HY9, 2Jpz and 143D), in the parallel structure (PDB ID: 1KF1) they were already present in the X-ray diffraction resolved structure.

Table S2.9.1. Grid settings and molecular docking parameters adopted for each computation for *M. Pomifera* derivatives

#	Grid size (Å)	Number of K ⁺ ions	Conformer generation sampling	Initial poses selection sampling	Precision
1	18x18x18	3	Default	Default	SP
2	18x18x18	3	Enhanced	Default	SP
3	18x18x18	3	Enhanced	Expanded	SP
4	18x18x18	3	Default	Default	XP
5	18x18x18	3	Default	Expanded	XP
6	18x18x18	2	Enhanced	Expanded	SP
7	18x18x18	2	Default	Expanded	XP
8	15x15x15	2	Enhanced	Expanded	SP
9	15x15x15	2	Default	Expanded	XP
10	13x13x13	2	Enhanced	Expanded	SP
11	13x13x13	2	Default	Expanded	XP
12	32x32x32	2	Enhanced	Expanded	SP
13	32x32x32	2	Default	Expanded	XP

2.9.8 Molecular dynamics

The structures of the G4-ligand complexes, obtained by the molecular docking study, were prepared for the simulations with Schrodinger application System Builder as follows: the structures were solvated in a triclinic water box (10 Å buffer) composed by 4-site model TIP4Pew water molecules,⁴⁸⁰ the total charge was neutralized by the addition of an adequate number of K⁺ ions, and a 0.20 M KCl concentration was set. The system initially prepared for Schrodinger default force field OPLS3e was converted through Viparr application to the AMBER OL15 force field specific for nucleic acids,^{481–484} the parameters set used for the ions were the ones developed by *Joung et al.*⁴⁸⁵ The ligands were parameterized with Antechamber application of AmberTools20 using GAFF2 parameters and AM1-BCC atomic charges, missing parameters were generated with Parmchk2. The simulations were conducted with Schrodinger Desmond for 100 ns each, recording 1 frame every 10 ps using the NPT ensemble at the temperature and pressure of 300 K and 1.01 bar, default settings were used for the RESPA integrator (2 fs for bonded interactions, 2 fs for near and 6 fs for far interactions), the system was relaxed prior the simulation using the default Desmond protocol.⁴⁸⁶ The trajectories were analysed after the

simulations with Simulation Event Analysis Schrodinger application, for the over-time ligand RMSD calculations the ligand atoms positions were fitted to the guanine planes.

3 Side projects

3.1 Selenium-containing compounds for the protection against oxidative stress

3.1.1 Fluoxetine derivatives

Introduction

Selective serotonin reuptake inhibitors (SSRIs) were developed in the context of major depressive disorder as a class of compounds with the same therapeutic activity of tricyclic antidepressants with the aim of limiting undesirable side effects, such as on cardiac conduction and anticholinergic activity. All currently available antidepressant drugs are based on the monoamine-deficiency hypothesis, deriving from the discovery that changes in monoamine (such as serotonin, norepinephrine and dopamine) concentrations in the forebrain and limbic system can cause mood alteration.⁴⁸⁷ SSRIs inhibit the serotonin transporter increasing the concentration of serotonin, which is proved to improve the mood together with norepinephrine, in the synaptic cleft, thus enhancing the serotonergic tone. Within all the tested compounds, fluoxetine (*N*-methyl-3-phenyl-3-[4-(trifluoromethyl)phenoxy]propan-1-amine) (Figure 3.1.1) was found to be the most potent inhibitor of serotonin reuptake in nerve endings, therefore it was approved by the FDA in 1987 and marketed by Eli Lilly as Prozac.^{488,489}

Recent studies revealed that fluoxetine is able to contrast adverse effect of different types of immune system stressor and protect against oxidative stress. Oxidative stress can be described as a perturbation of the equilibrium between the production of reactive oxygen species (ROS) like peroxides and free radicals, and the cells capacity to steadily detoxify these reactive intermediates and thus repair the damage that has been caused by them. ROS are involved in signalling event mediated by thiols residue in proteins and can regulate transcription but they are also toxic for their high reactivity towards lipids, proteins and DNA, whose oxidation results in neuronal cell death and hence in progression of cardiovascular disease, pulmonary disease, diabetes, cancer and neurodegenerative diseases.^{490,491} Antioxidant enzyme, such as superoxide dismutase (SOD), catalase (CAT) or glutathione peroxidase (GPx) are in charge of metabolising ROS into less toxic species to keep the redox homeostasis in the living organisms.⁴⁹²

Fluoxetine has showed its beneficial antioxidant effect through a combination of mechanism, such as direct ROS scavenging, stimulation of anti-oxidant enzymes and by enhancement of the serotonin antioxidant ability.^{493,494}

Computational studies have been performed on fluoxetine and its two metabolites (N-demethylated derivative known as norfluoxetine and *p*-trifluoromethylphenol derived from oxidative O-dealkylation) together with serotonin in order to better understand the mechanism of radical scavenging. The major outcome showed that serotonin is overall a better scavenger of hydroxyl and peroxy radicals compared to all the others, thus suggesting that the improvement of oxidative damage is more likely to be due to the higher concentration of free serotonin in the synaptic cleft rather than a direct effect of fluoxetine itself and its metabolites.⁴⁹⁵

On the other hand, it has been demonstrated by Herbet and colleagues that fluoxetine is able to interact with enzymes such as SOD, CAT or GPx, thus again explicating its antioxidant effect in an indirect manner.⁴⁹³

Glutathione (GSH) is an important compound used by the body to eliminate peroxides. GSH oxidation leads to glutathione disulphide (GSSG) that is reduced back using NADPH by GSSG reductase. GSH/GSSG release from tissues to extracellular space is subjected to their tissue concentrations, thus their balance can be used to estimate the oxidative stress.⁴⁹¹ GPx is the first selenoprotein discovered and its importance and wide studies is due to its ability to act against oxidative damage. Eight different GPx have been identified: in three of them the selenocysteine (Sec) of the active site has been replaced with a Cys and are less efficient in neutralizing peroxidic species transferring the oxidation of GSH, thus demonstrating the importance of the selenium atom in the catalytic triad. Sec is able to reduce hydrogen peroxide and organic peroxides leading to the formation water or the corresponding alcohols. With this reaction Sec is converted into the unstable selenic acid that reacts with a first molecule of reduced GSH to generate a selenenyl sulphide intermediate. A second molecule attack the Se-S bond producing GSSG that is enzymatically reduced back to GSH by glutathione reductase (GR)-NADPH system.⁴⁹⁶ Thus selenium-containing molecules can mimic GPx activity, thus functioning as antioxidants.⁴⁹⁷⁻⁵⁰⁰ The most known is ebselen (2-phenyl-1,2-benzisoxazol-3(2H)-one) that was used in clinical trials, but also aryl and alkyl selenides and diselenides have been synthesised and tested *in vitro* and *in vivo*.⁵⁰¹⁻⁵⁰⁵

The idea behind this work was to synthesize new compounds chemically related to fluoxetine and that could combine the antidepressant effect, thus selectively inhibiting

serotonin reuptake in the synapsis, and a direct antioxidant behaviour within the CNS. This would be provided by the introduction of a selenium atom, hence enabling a GPx-mimicking activity. Therefore, oxygen was replaced with selenium, leading to selenofluoxetine (Figure 3.1.1) and different substituents have been added on the former phenoxy-ring in para position. All the synthesized compounds were characterized by $^1\text{H-NMR}$, $^{13}\text{C-NMR}$ and ESI-MS. Moreover, their reaction with H_2O_2 was studied with $^1\text{H-NMR}$, $^{77}\text{Se-NMR}$, ESI-MS and computational studies to assess the antioxidant potential. Furthermore, their ability in inhibiting serotonin reuptake has been investigated *in vivo* in mice, with the support of molecular docking. These findings have been published in two different peer-reviewed journals.^{506,507} Figure 3.1.1 shows the chemical structure of fluoxetine, selenofluoxetine and of the other synthesized derivatives.

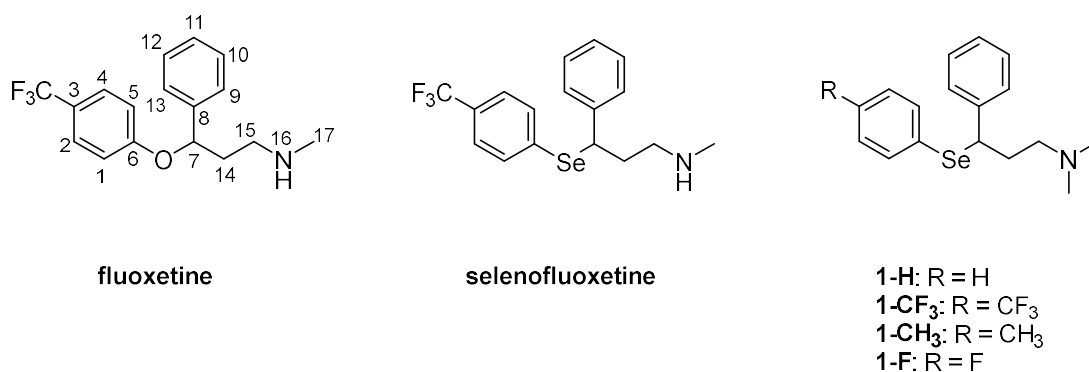


Figure 3.1.1. Chemical structure of fluoxetine, selenofluoxetine and of the derivatives.

Results and discussion

The compounds have been successfully prepared thanks to a convergent multi-step synthesis approach that is depicted in Figure 3.1.2. Initially, a *Mannich* reaction between acetophenone, dimethylamine and formaldehyde leading to the Mannich base **Se2**, then the ketone was reduced to alcohol with sodium-borohydride providing compound **Se3** that was chlorinated with thionyl chloride obtaining compound **Se4**. Lastly the chloride was substituted with the appropriated selenide that was formed *in situ* by the reduction of the corresponding diselenide with sodium borohydride, leading to the final compounds. The diphenyl diselenides (**Se5-Se7**) were obtained through the reaction of elemental selenium and the *Grignard* reagent obtained from the *p*-substituted bromo-benzene following a procedure of Sun *et al.*⁵⁰⁸

The characterization of the compounds has been reported in the Appendix, Section 5.

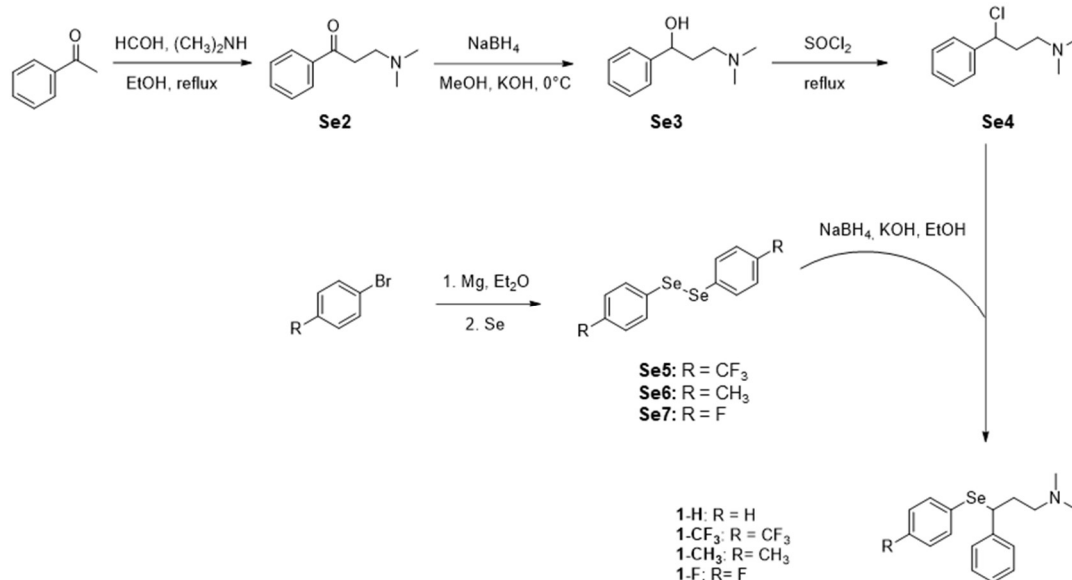


Figure 3.1.2. Schematic scheme for the preparation of selenofluoxetine derivatives.

After the synthesis, the antioxidant behaviour of the compounds was tested through their reaction with oxygen peroxide. The reaction was carried out with 1 equivalent of H₂O₂ and was followed with different techniques to gain information on the mechanism and on the formation of the products. The reaction with compound **1-H** will be discussed in detail as an example but the same behaviour was observed with the other derivatives. Firstly, we registered the ¹H-NMR spectrum of the synthesised compound in D₂O, then we added 1 eq of H₂O₂ and we registered new spectra every 2 minutes.

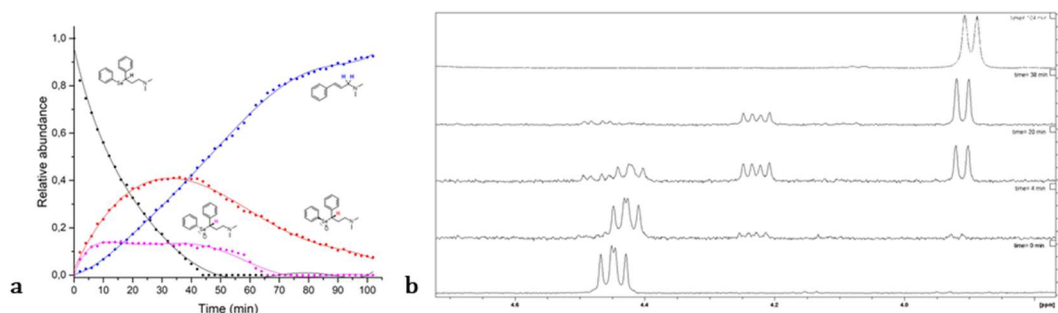


Figure 3.1.3. ¹H-NMR mechanistic study of the reaction of **1-H** with H₂O₂. The graph reported in (a) was obtained by integration of the α-H signals for the starting material and the two diastereoisomers and the allylic-hydrogens signals for the cinnamylamine in the ¹H-NMR spectra acquired at different time points (•starting material, • R-R (and its enantiomer), • R-S (and its enantiomer) • cinnamylamine). In (b), representative ¹H-NMR spectra are reported showing the variation over time during the oxidation-reaction of compound **1-H**, focusing on the region between 4.7 and 3.8 ppm (area of interest to plot the graph).

The first observation was the immediate disappearance of the signal corresponding to the proton in α-position with respect of the selenium (4.42 ppm, Figure 3.1.3), accompanied by the formation of the signal of the two diastereoisomers (4.48 and 4.23

ppm, Figure 3.1.3), thanks to the oxidation of selenium, a new chiral centre is formed. Interestingly, as the reaction proceeded, also those signals started to disappear and new peaks were found. It is reported in literature that the oxidation of selenides with protons in β -position can trigger an intramolecular *syn* elimination reaction in which a proton is transferred from the β -position to the oxygen of the selenoxide with subsequent breakage of the selenium-carbon bond and formation of the carbon-carbon double bond. The reaction is reported to be highly selective and lead to the formation of the *trans*-olefin.⁵⁰⁹ In fact the new signals could indeed be explained with the formation of an alkene and in particular with the (E)-N,N-dimethyl-cinnamylamine. The proposed mechanism of the reaction is depicted in Figure 3.1.4b and lead to the formation of selenenic acid and the olefin. The same behaviour was observed with the other two compounds although in 1- CF_3 the signal of one of the diastereoisomers partially overlap with the starting material (4.55 ppm), complicating the in-depth investigation.

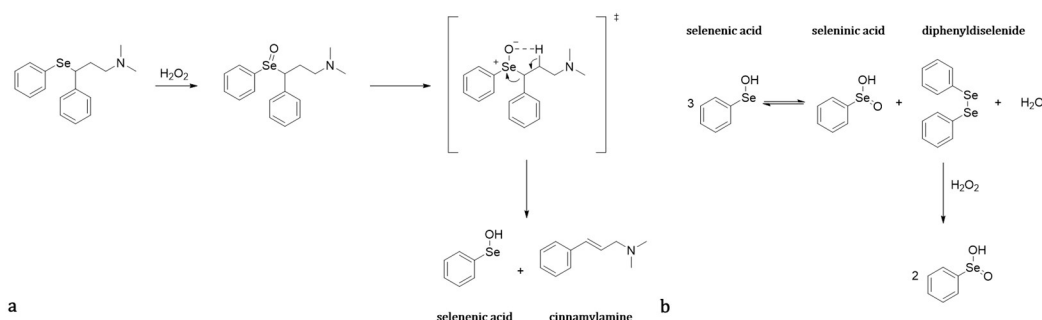


Figure 3.1.4. (a) Proposed mechanism of the selenoxide oxidation-elimination. (b) Redox equilibrium of selenenic acid

The overall reaction was relatively fast for all the compounds and a 90% conversion was observed after 36 minutes with **1-CF₃**, after 95 minutes for **1-H** and after 77 minutes for **1-CH₃**. The full $^1\text{H-NMR}$ spectra at various time points can be found in the Appendix, Section 5.

Selenenic acid is known to undergo a chemical equilibrium that involves its disproportionation into seleninic acid and diphenyl diselenide. (Figure 3.1.4b) Moreover, diphenyl diselenide can be oxidized into seleninic acid by H_2O_2 .⁵¹⁰ Both diphenyl diselenide and seleninic acid are less water-soluble, indeed we could observe a yellow precipitate at the end of the reaction. Furthermore, the presence of seleninic acid was confirmed by the presence of a signal at 1172.0 ppm in the $^{77}\text{Se-NMR}$ (Figure 3.1.5a). $^1\text{H-}^{77}\text{Se}$ HMBC in the middle of the reaction of **1-CF₃** was also recorded after dilution 1:1 in MeOD to confirm the proposed mechanism. (Figure 3.1.5b) Three signals were detected

at 447.9, 472.6 and 1211.8 ppm, corresponding respectively to seleninic acid, the selenoxide intermediates and diselenide as reported by Wang et al. study.⁵¹¹ It is important to highlight that selenium and its oxidised species have been reported to have toxic effect at high concentrations,⁵¹² nonetheless organoselenides are gaining attention as organic catalysts in environmentally friendly conditions, in particular in aqueous medium.^{513,514}

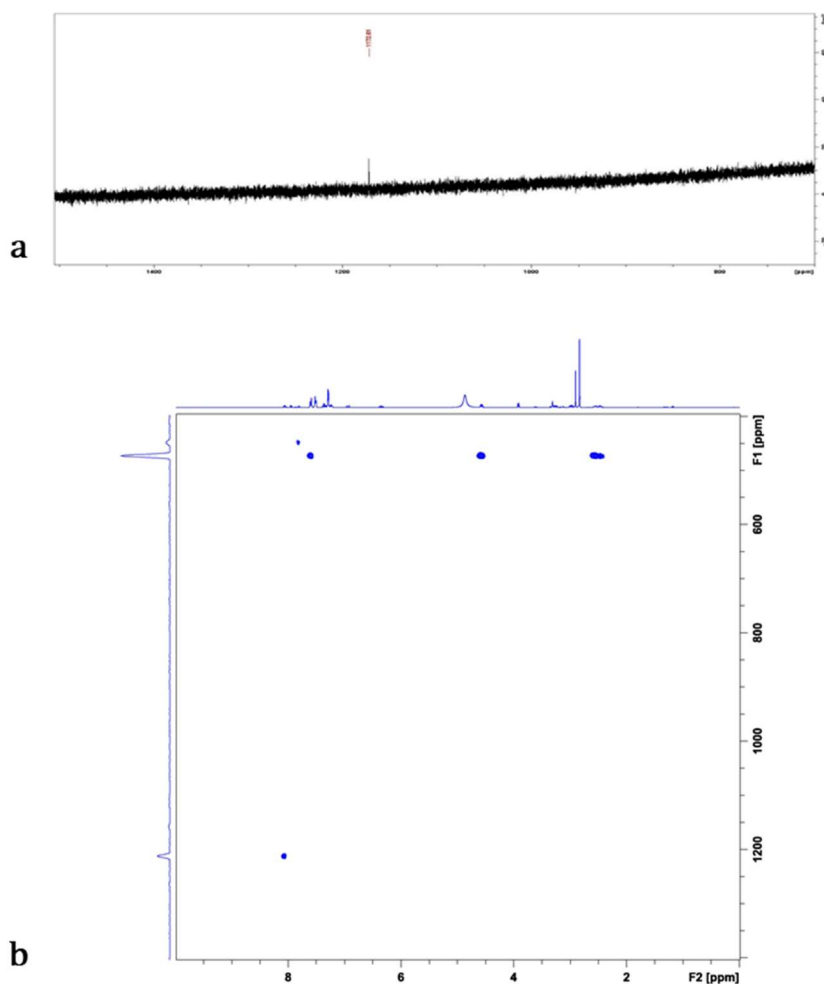


Figure 3.1.5. (a) ^{77}Se -NMR spectrum of compound **1-H** at the end of the reaction, showing the presence of seleninic acid. (b) ^1H - ^{77}Se -HMBC of compound **1-CF₃** in the middle of the reaction showing the presence of seleninic acid, the selenoxide and the diphenyl diselenide.

ESI-MS was then used to monitor the reaction similarly to the procedure used with NMR and further confirmed our findings. The reaction conditions were kept unaltered but the sample for the analysis were diluted 1:1000 in methanol before the injection in the mass spectrometer. After the addition of hydrogen peroxide, the signal corresponding to the selenoxide and cinnamylamine were detected in positive ionization mode, respectively at 334 and 162 m/z , whereas the signal corresponding of the seleninic acid at m/z 189 was detected in negative ionization mode. Moreover, MS spectra of the sample used for the NMR study after the end of the reaction were recorded in both positive and negative

ionization modes showing the presence of seleninic acid, diphenyl diselenide and cinnamylamines. The MS spectra can be seen in the Appendix, Section 5.

Thanks to the collaboration of the research group of Prof. Laura Oriani of the University of Padova, we studied the oxidation reaction of selenofluoxetine and the radical scavenging activity via HAT (Hydrogen Atom Transfer) mechanism of selenofluoxetine and of one of our derivatives through DFT analysis.

The intermediates and transition states of the oxidation with H_2O_2 were computed in the gas-phase and in water at ZORA-OLYP/TZ2P level of theory, with the latter showing a decrease of the activation energy. As previously mentioned, the attack of the peroxide can happen from both sides, therefore, both the paths of the diastereoisomers (different stereochemistry of the selenium) have been considered, even though the reaction of the two selenoxide diastereoisomers have different energy, despite the steric hindrance. Figure 3.1.6 shows the geometries and the energies of the reactants, intermediates and products.

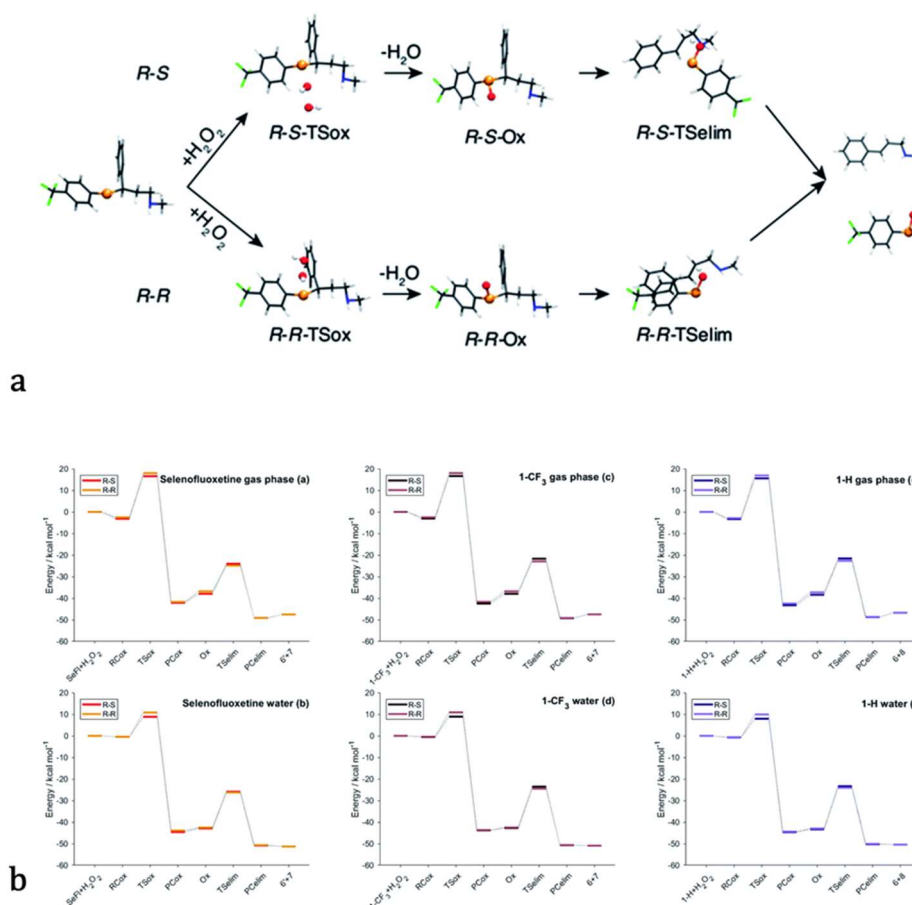


Figure 3.1.6 (a) Fully optimised geometries of intermediates and transition states of the oxidation of selenofluoxetine shown in Figure 3.1.4a. Reactant and product complexes are omitted for clarity. Level of theory: ZORA-OLYP/TZ2P. (b) Energies of the stationary points of the investigated reactions for different selenofluoxetine derivatives in the gas-phase (upper row) and in water (lower row). Level of theory: (COSMO)-ZORA-OLYP/TZ2P.

It was previously demonstrated that HAT in fluoxetine occurs from the non-aromatic sites, namely C7, C14, C15, N16 and C17 in gas phase, in water and in benzene (as example of organic solvent).⁴⁹⁵ Therefore, the same positions have been considered for selenofluoxetine and **1-F** focusing on the scavenging of $HO\cdot$ (the most reactive and electrophilic oxygen centred radical), $HOO\cdot$ and $CH_2 = CHOO\cdot$ to mimic large unsaturated peroxy radicals. The calculated free energies ΔG_{HAT} are reported in Figure 3.1.7.

The data showed that selenofluoxetine and **1-F** react C7 and C17 are the most favourable site for HAT reaction in gas phase, ΔG_{HAT}° are computed always negative for the process with $HO\cdot$ and become more negative in condensed phase making the reaction progressively more favourable in benzene and water, whereas it is mostly positive for the other radicals.

Moreover, activation free energy computed for the reaction with $HO\cdot$ on all five sites both in gas and condensed phase, showed that C7, C17 and N16 HAT scavenging are the most exergonic process and in general, all the computed values showed that HAT is thermodynamically and kinetically more favoured than in fluoxetine, meaning that seleno-derivatives have a better radical scavenging potential.

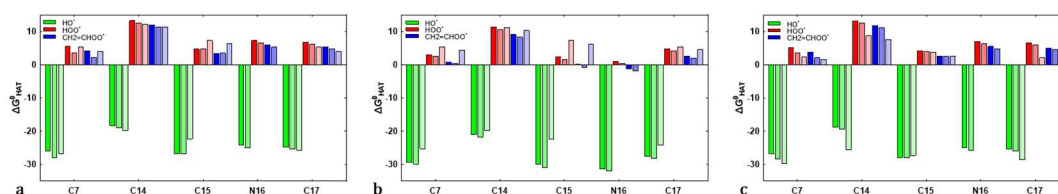


Figure 3.1.7. ΔG_{HAT}° (kcal/mol) in the gas phase (a), in water (b), and in benzene (c) for the scavenging of $HO\cdot$, $HOO\cdot$ and $CH_2 = CHOO\cdot$ from non-aromatic sites of selenofluoxetine (medium shade) and **1-F** (lighter shade). Values for fluoxetine, taken from Muraro et al.⁴⁹⁵ are showed in the darker shade for comparison. Level of theory: (SMD)-M06-2X/6-311+G(d,p)//M06-2X/6-31G(d).

Furthermore, **1-F** serotonin reuptake inhibition in CA2 region was investigated *in vivo* in mice and compared to fluoxetine thanks to the collaboration with Prof. Hashemi's Lab that recently developed a new tool for monitoring the uptake through Fast Scan Cyclic Voltammetry (FSCV) to assess the effects of escitalopram (another common SSRI).⁵¹⁵⁻⁵¹⁹ A Carbon Fiber Microelectrode (CFM) is implanted into the brain of an anesthetized mouse that allows to apply an electrochemical waveform that measures the redox potential of the molecules to identify and quantify serotonin in real time, with subsecond time resolution.^{516,520}

After 30 minutes from the i.p. injection of **1-F** and fluoxetine (when there is the maximal response in agreement with the half-life time of other common SSRI in mice),⁵²¹ of serotonin release and the $t_{1/2}$ of reuptake increased after the treatment with **1-F** and fluoxetine, although the effects on the latter were more pronounced despite the lower dose, probably due to the higher electronegativity conferred by the oxygen and because of the enhance ability of the Se-C bond to be hydrolysed compared to the O-C bond. Moreover, the maximum rate of serotonin reuptake (v_{max}) has been calculated from the FSCV acquisitions and showed a marked decrease for both the compounds, hence showing chemical efficacy for slowing serotonin reuptake. The data are reported in Table 3.1.1 and Figure 3.1.8.

Table 3.1.1. Features of *in vivo* Experimental Curves: Mean \pm SEM (n = 5 animals) of the maximum serotonin release, the clearance half time ($t_{1/2}$) and the maximum reuptake rate v_{max} ^a

	maximum release (nM)	$t_{1/2}$ of clearance (s)	v_{max} (nM/s)
control pre-fluoxetine	35.75 \pm 10.16	1.90 \pm 0.46	17.69
30 min post-fluoxetine	56.05 \pm 13.40*	4.01 \pm 0.70*	15.21
control pre-1-F	22.73 \pm 4.07	1.93 \pm 0.30	15.60
30 min post-1-F	25.84 \pm 3.41	3.38 \pm 0.41*	13.46

^a Each parameter was tested for significant difference between control and drug treatment (paired samples t-test). Significance (*) was defined as $p < 0.05$. Michaelis–Menten reuptake kinetics model for serotonin was fitted to the signal. K_m was set to 5 nM, while v_{max} was optimized to fit the average experimental trace

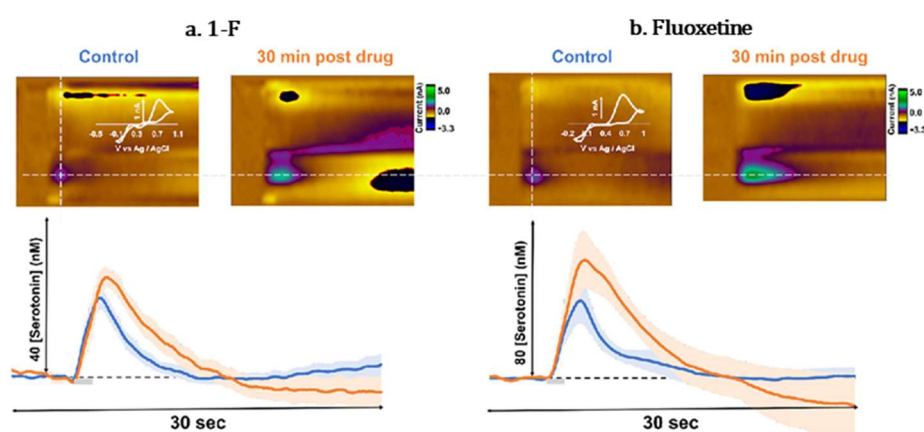


Figure 3.1.8. *In vivo* stimulated serotonin release in the CA2 region of the hippocampus. Representative colour plots for this experimental paradigm, before and 30 min into drug analysis, are found above the representative IT curves. The averaged control evoked serotonin in mixed-sex cohort can be found in blue (n = 5) with SEM calculated and outlined in the lighter shade of blue. The averaged drug files ((a) **1-F**, 50 mg kg⁻¹ and (b) Fluoxetine, 20 mg kg⁻¹) at 30 min for serotonin in mixed-sex cohort is plotted in orange (n = 5) with SEM calculated and outlined in the lighter shade of orange. Stimulation is indicated by the grey bar. Statistical analysis can be found for each data set in Table 3.1.1.

The detailed experimental procedure regarding the DFT calculation and the *in vitro* study performed from the cooperation partners can be found respectively in the two publications.^{506,507}

Finally, a molecular docking study was performed in order to investigate the binding energy and poses of the complex formed by the compounds and the serotonin transporter (SERT, PDB ID: 5I73). In particular, after the preparation of the protein using Protein Preparation Wizard embedded in Schrödinger Suite, two separate grids have been created on both the central and allosteric site, the four synthesized compounds, fluoxetine and selenofluoxetine have been docked on each grid. The results are reported in Table 3.1.2 together with the results obtained for the re-docking of S-citalopram and the RMSD of the two poses that were used as control being the co-crystallised ligand.

All the compounds showed high binding energy on the same values of S-citalopram and fluoxetine both on the central and allosteric site. Interestingly, **1-CF₃** show the highest stabilization (lowest absolute value) and the calculated binding energies were even higher than fluoxetine. Moreover, analysing the 3D complex of the central site, we can see that the predicted poses of fluoxetine and **1-CF₃** are quite similar as the interactions with the residues. (Figure 3.1.9)

The interactions that are present on fluoxetine, compound **1-CF₃** and citalopram are exactly the same include: a pi-stacking between the benzyl group and the residue Tyr176, a pi-cation interaction between the amine and the Tyr95 and a H-bond and a salt bridge between the amine and Asp98.

Table 3.1.2. Docking score of the tested compounds on the central and allosteric site of SERT (PDB ID: 5I73). Between parenthesis, the RMSD value calculated between the crystallized and the docked pose of S-citalopram

Compound	Docking Score (-kcal/mol) central site	Docking Score (-kcal/mol) allosteric site
1-CF ₃	9,377	6,939
1-CH ₃	8,547	6,058
1-F	8,104	6,483
1-H	7,778	5,874
Selenofluoxetine	8,890	6,899
Fluoxetine	9,064	6,797
S-citalopram	9,514 (1,83 Å)	6,605 (2,04 Å)

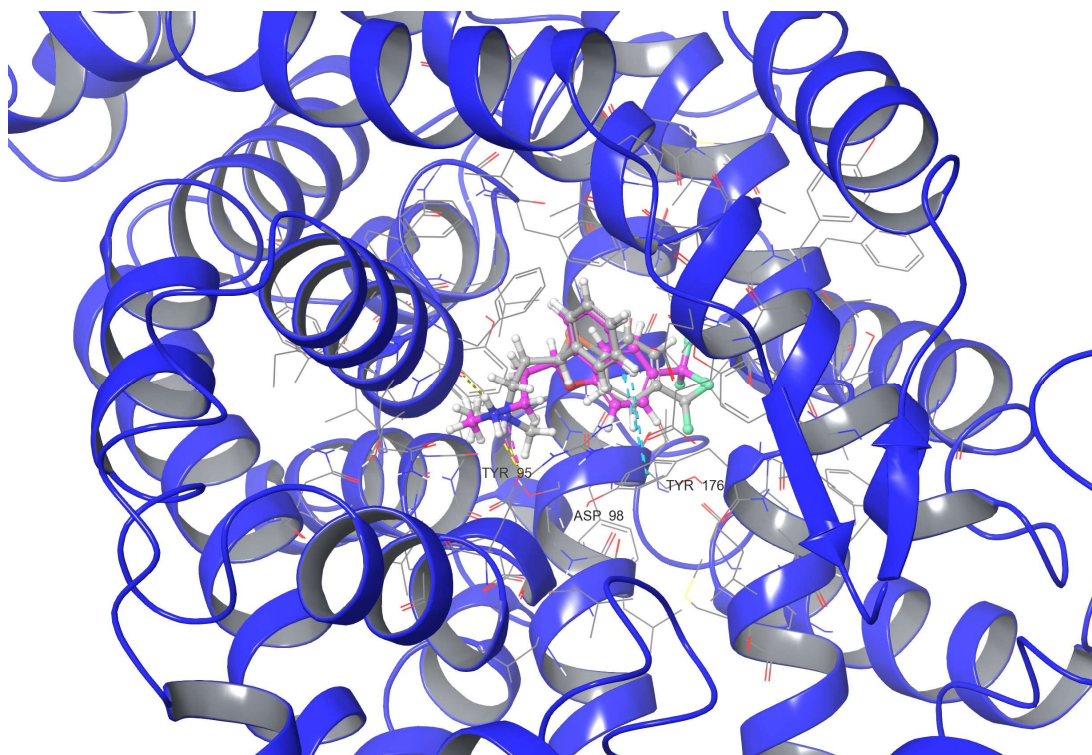


Figure 3.1.9. Overlap of the predicted docking pose of • fluoxetine and • 1-CF₃.

3.1.2 Selenoxide elimination triggers enamine hydrolysis to primary and secondary amines: a combined experimental and theoretical investigation

Introduction

A direct consequence of the work of the previous chapter has been recently published.⁵²²

In particular, considering that selenoxide elimination has been widely used for the preparation of alkenes^{523–526} using phenyl selenyl chloride or diphenyl diselenide as source of selenium that, as previously showed, is easily oxidized to selenoxide by peroxides and other ROS,^{527–529} and triggers a subsequent *syn* elimination leading to selectively trans-olefine. Primary and secondary amines are one of the most frequent functional groups in biologically relevant molecules, even though drugs containing this moiety can lead to pharmacokinetic limitations due to their poor diffusion through membrane or blood-brain barrier. Moreover, it is well known that amines undergo oxidations or dealkylations in metabolic phase I and conjugations in phase II, therefore “masking” them in prodrugs may lead to their preservation and activation under controlled conditions, such as pH, redox or enzyme dependent changes.⁵³⁰ Sun and colleagues, recently demonstrated that selenoxide elimination can be use in the context of manipulation of oxidative stress to improve the activity of anticancer drugs.⁵³¹

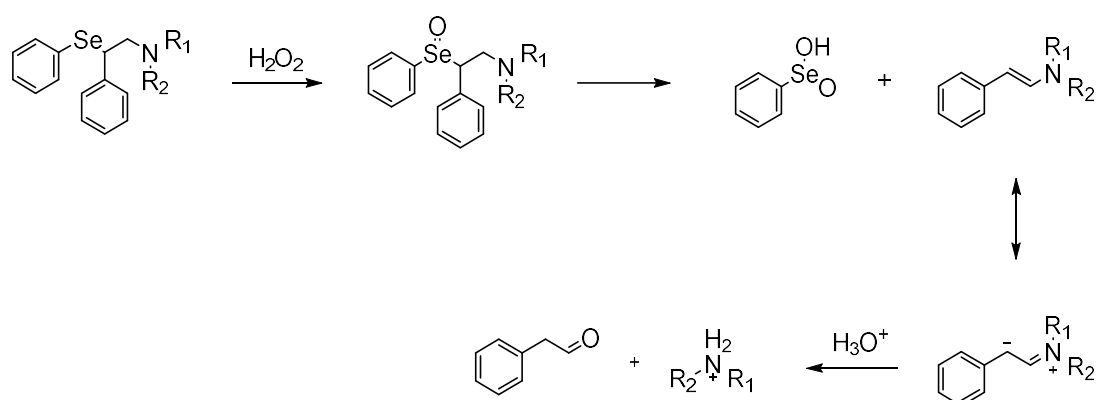


Figure 3.1.10. Proposed mechanism for selenoxide elimination-triggered enamine hydrolysis

In this context, the idea of the work was to exploit the selenoxide elimination to trigger an imine-enamine equilibrium that could lead to the hydrolysis of the latter, releasing the free amine. (Figure 3.1.10) Together with the selective release of active molecule, this work is valuable also in the context of organic synthesis, as the preparation of primary and secondary amines is not always easy and it is normally performed by *Gabriel*

synthesis,⁵³² by azidation-reduction,⁵³³ by *Leuckart* reaction,⁵³⁴ by direct alkylation,⁵³⁵ by reductive amination⁵³⁶ and by *Fukuyama* synthesis.⁵³⁷

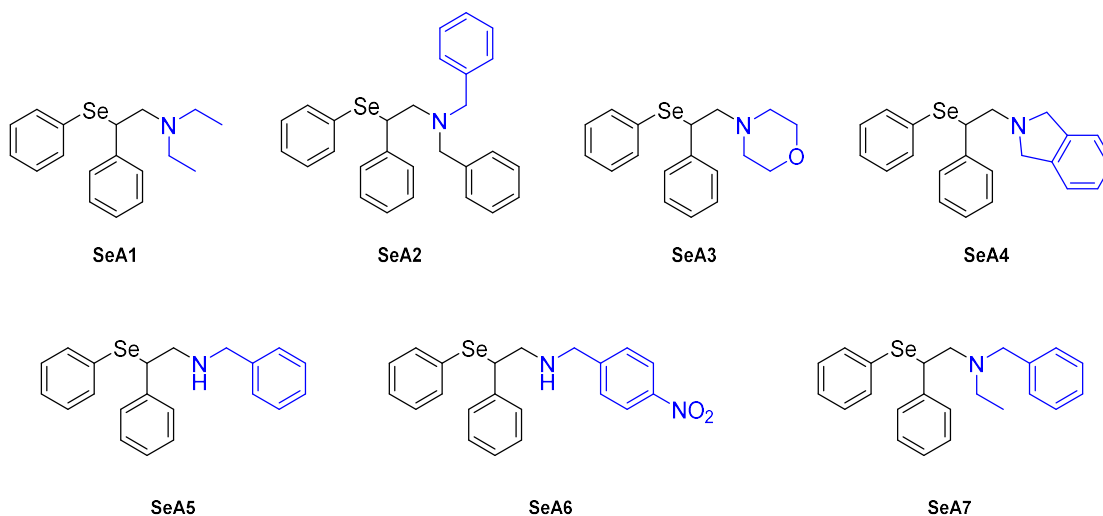


Figure 3.1.11. Chemical structure of the proposed compounds, coloured in blue the portion of the molecules providing the resulting amine.

In order to do so we synthesized a small library of seven compound (Figure 3.1.11) with a smaller chain between Se and N ($-CH_2CH_2-$) and we investigated with the same technique explained in the previous chapter their oxidation reaction with H_2O_2 .

Results and discussion

The preparation of the final compounds was obtained with a multistep approach that is depicted in Figure 3.1.12. The first intermediate **SeA8** was obtained by the reaction of styrene with the hypervalent iodine reagent (diacetoxyiodo)benzene, sodium azide and diphenyl diselenide following a procedure published by Tingoli and colleagues,⁵³⁸ then the azide was reduced with lithium aluminium hydride in THF that lead to the formation of compound **SeA9**. Different alkylation reactions were performed on this intermediate to obtain the final compounds: **SeA1-SeA4** were obtained through direct alkylation with the opportune halide, compounds **SeA5** and **SeA6** by reductive amination with the opportune aldehyde, and compound **SeA7** by direct alkylation of compound **SeA5** with iodoethane.

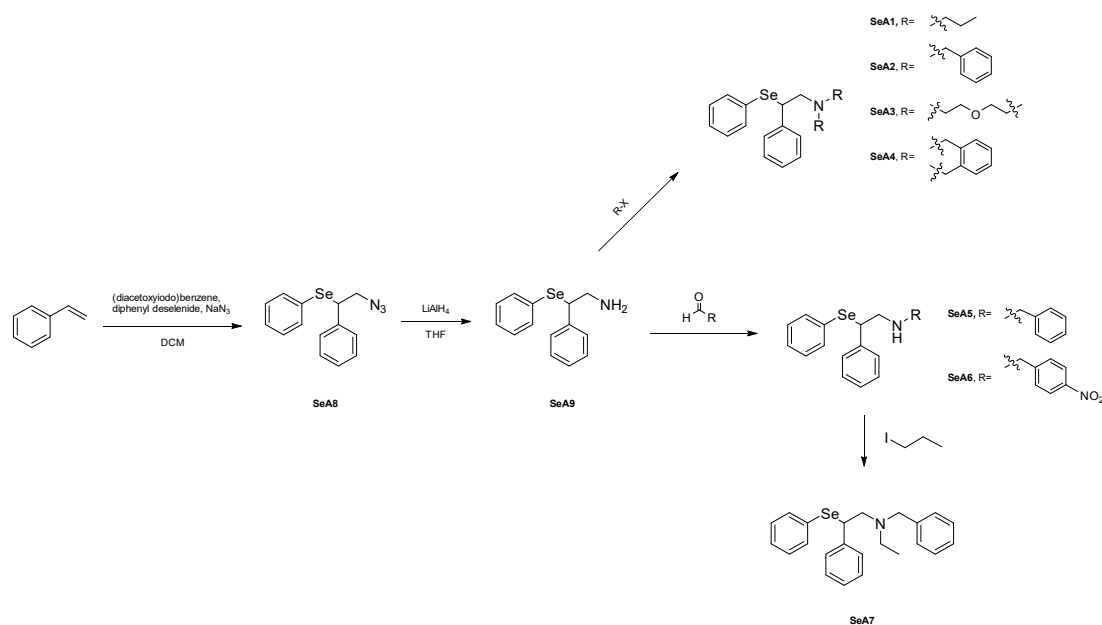


Figure 3.1.12. Reaction scheme for the synthesis of the compounds.

As for all the other works, the compounds were characterized with ESI-MS, $^1\text{H-NMR}$ and $^{13}\text{C-NMR}$ and they were subjected to oxidation with H_2O_2 . The full characterization of the final compounds and intermediates can be seen in the Appendix, Section 5.

The reaction was studied in detail similarly to what we did for **1-H**, the compound was dissolved in D_2O , its $^1\text{H-NMR}$ spectrum was recorded, then 1.2 eq of H_2O_2 was added to the solution and 1D spectra were recorded at different time points to observe the reaction ongoing. The signal of the $\alpha\text{-H}$ of the starting material gradually disappeared with the formation of the peaks corresponding to the selenoxide diastereoisomers. Then also these signals started gradually disappearing to the formation of the particular triplets corresponding to morpholine (*i.e.*, 3.25 and 390 ppm). After 70 minutes, the starting material was reduced to the 10%. (Figure 3.1.13)

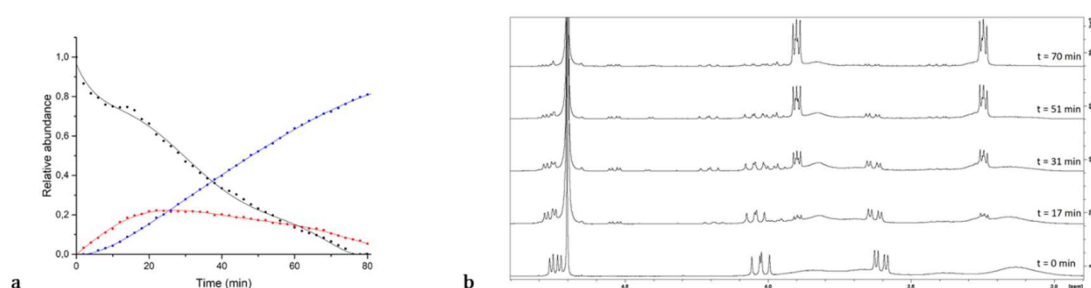


Figure 3.1.13. Reaction profiles for compound **SeA3**, intermediate and reaction products. The signals considered for relative quantification are depicted as in the following: • starting material, • intermediate, • secondary amine. Integrals were measured using acetonitrile as internal standard (a). Detailed views of representative NMR spectra showing the considered signals (b).

The reaction mechanism was studied through DFT calculation thanks to the collaboration of Prof. Orián group (Figure 3.1.14). In particular the oxidation of compound **SeA3** has an energy barrier of 14.5 kcal/mol in water, the following elimination-enamine hydrolysis was studied in three elementary steps. Firstly, an intramolecular elimination led to the breakage of the C-Se bond leading to the enamine and selenenic acid and requires 7.6 kcal/mol for the formation of the transition state TSelim but it has a free energy ΔG of -76.7 kcal/mol with respect to the starting material. Then, an ion H_3O^+ react with the enamine with an activation energy of 51.8 kcal/mol leading to an unstable immine intermediate. Eventually the C-N bond breaks to give morpholine and phenylacetaldehyde with a total ΔG of 6.4 kcal/mol with respect to the starting enamine. Interestingly, as the proton transfer from the OH to the N atom would be successfully computed in terms of more favourable energetics if more water molecules were taken into account,^{539–542} the general mechanism is not influenced by their inclusion and would complicate the conformational study. Moreover, the detailed mechanistic analysis of that reaction is beyond the aim of this work.

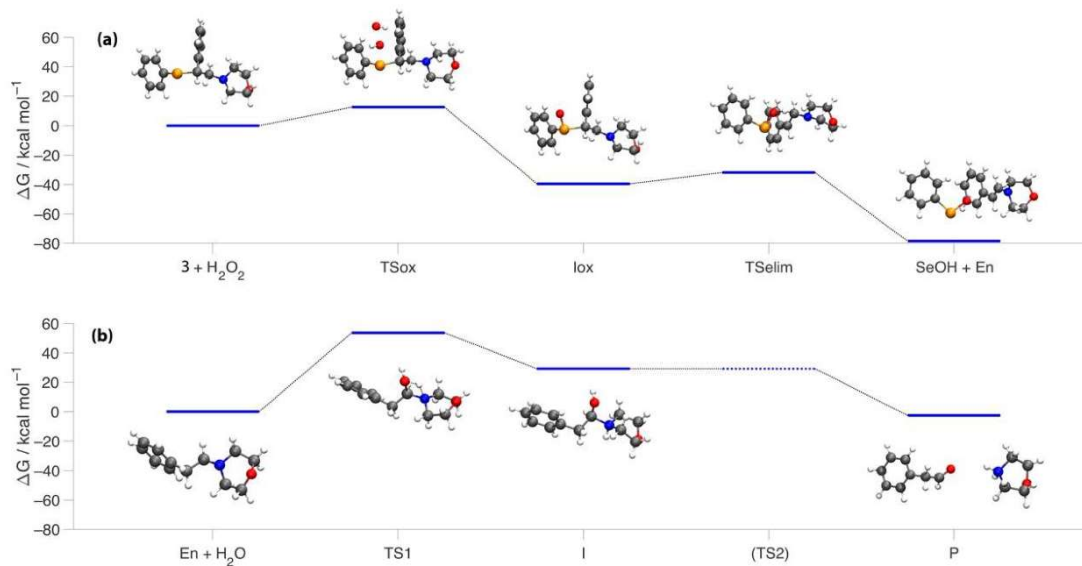


Figure 3.1.14. Gibbs free energies of the relevant intermediates and transition states characterized in water for the oxidation and subsequent elimination of compound **SeA3**: (a) corresponds to the top reaction of Figure 3.1.9 whereas (b) to the bottom one. Water and hydrogen peroxide molecules are not shown in the minimum energy structures for clarity. Level of theory: COSMO-ZORA-OLYP/TZ2P.

The same reaction was performed for all the other compounds and monitored by ¹H-NMR and ESI-MS, which allowed to further confirm the identity of the intermediates. Again, the final amines were detected with positive ionization mode, whereas the seleninic and the benzoic acids (formed from phenylacetaldehyde oxidation⁵⁴³ were confirmed with

negative ionization mode. The full NMR spectra at different time points for all the compounds can be found in Materials and Methods, Chapter 3.1.4 together with the ESI-MS study performed for compounds **SeA3**.

Additionally, the oxidation reactions were scaled up and conducted overnight with the aim of isolating the final products, which were characterised by $^1\text{H-NMR}$, $^{13}\text{C-NMR}$ and ESI-MS (see the Appendix, Section 5), and calculating the yield, that was up to the 90%, thus demonstrating the formation of secondary and primary amines in laboratory scale.

3.1.3 Conclusions

These studies were intended with the aim of underlining the importance of the combinations of different techniques for the development of new potentially active small molecules.

In particular, four selenofluoxetine derivatives were prepared after a careful optimization of a multi-step converged reaction scheme. Their ability to ameliorate oxidative stress and therefore to protect against neurodegeneration was tested with a multitude of techniques that included NMR, ESI-MS and DFT computational analysis. In particular the reaction of these compounds with H₂O₂ was studied both *in vitro* and *in silico* to gain important insight into the mechanism and the scavenging activity *via* HAT was computationally determined.

Selenoxide elimination is well known in organic chemistry and has been widely exploited for the synthesis of new molecules. In this context it is worth mentioning that the result of this reaction is cinnamylamines, whose scaffold has been found in many drugs like flunarizine and naftidine that require the use of hazardous and expensive reactant, harsh experimental conditions and do not guarantee the required trans stereoselectivity, whereas the proposed approach is efficiently achieved in aqueous medium and oxygen peroxide as oxidative agent.

NMR and ESI-MS allowed to characterize all the intermediates and products involved in the reaction and the calculation of the ΔG values allowed to confirm the overall mechanism.

The following development of the new derivatives with shorter chain to trigger an enamine hydrolysis was also valuable for the potential of preparing successfully prodrugs that could be activated only in certain conditions in the context of drug delivery but also as a valuable alternative for the synthesis of primary and secondary amines. Seven final compounds were successfully prepared through a multi-step synthesis and the oxidation reaction to give the final amine was studied both in NMR and ESI-MS sample and more importantly also on a laboratory scale and again DFT studies assisted for the understanding of the mechanism involved.

Moreover, *in vivo* studies were performed on mice with selenofluoxetine derivatives to evaluate the actual inhibitory effect on the reuptake of serotonin, thus confirming that our derivatives could still be active as SSRI, providing a potential active compound that could work on both depressive symptoms, increasing serotonin levels and therefore improving the mood, and on the protection against neurodegeneration.

3.1.4 Materials and methods

Chemistry

Commercially available chemicals were purchased from Sigma-Aldrich and used without any further purification if not specified elsewhere. NMR experiments were performed on a Bruker Avance III 400 spectrometer (frequencies: 400.13, 100.62, and 76.37 MHz for ^1H , ^{13}C , and ^{77}Se nuclei, respectively) equipped with a multinuclear inverse z-field gradient probe head (5 mm). For data processing, TopSpin 4.0.8 software was used and the spectra were calibrated using solvent signal (^1H -NMR, $\delta_{\text{H}} = 7.26$ ppm for CDCl_3 , $\delta_{\text{H}} = 2.50$ ppm for d6-DMSO, $\delta_{\text{H}} = 4.79$ ppm for D_2O ; ^{13}C -NMR, $\delta_{\text{C}} = 77.16$ ppm for CDCl_3 , $\delta_{\text{C}} = 39.52$ ppm for d6-DMSO). Multiplicities are reported as follows: s, singlet; d, doublet; t, triplet; q, quartet; m, multiplet; b, broad; dd, doublet of doublets. Mass spectra were recorded by direct infusion ESI on a Thermo Fisher Scientific LCQ Fleet ion trap mass spectrometer. The purity profile of the compounds was assayed by HPLC using a Varian Pro-Star system equipped with a Biorad 1706 UV-VIS detector and an Agilent C-18 column (5 μm , 4.6 \times 150 mm). An appropriate ratio of water (A) and acetonitrile (B) was used as mobile phase with an overall flow rate of 1 mL min $^{-1}$. The general method for the analyses is reported here: 0 min (90% A–10% B), 5 min (90% A–10% B), 25 min (10% A–90% B), 30 min (90% A–10% B), and 32 min (90% A–10% B). The purity of all compounds was $\geq 96\%$, unless otherwise stated (254 nm).

Synthesis of N,N-dimethyl-3-oxo-3-phenylpropan-1-aminium chloride (Se2)

Dimethylamine hydrochloride (2.03 g, 24.9 mmol, 1.5 eq) and paraformaldehyde (0.65 g, 21.6 mmol, 1.3 eq) were weighted in a 50 mL round-bottomed flask and dissolved in 2.5 ml of ethanol. Acetophenone (2.00 g, 16.6 mmol, 1 eq) was added to the solution together with 40 μL of concentrated hydrochloric acid. The reaction mixture was stirred at reflux and checked through TLC (DCM/MeOH/TEA 97:2.5:0.5). After 2 hours, the solution was left to cool to room temperature. A solid precipitate of N,N-dimethyl-3-oxo-3-phenylpropan-1-aminium chloride salt formed and the solid was filtered with a Buchner funnel, washed with cold acetone (3 \times 10 mL) and with hexane (1 \times 10 mL). Yield 3.48 g (98%); white solid; ^1H -NMR (400 MHz, d6-DMSO): δ_{H} (ppm) 10.57 (brs, 1H), 8.02 (d, 2H, $J = 7.2$ Hz), 7.69 (t, 1H, $J = 7.4$ Hz), 7.57 (t, 2H, $J = 7.6$ Hz), 3.63 (t, 2H, $J = 7.2$ Hz), 3.40 (t, 2H, $J = 7.2$ Hz), 2.80 (s, 6H); ^{13}C -NMR (101 MHz, d6-DMSO): δ_{C} (ppm) 196.8,

135.9, 133.7, 128.8, 128.0, 51.8, 42.2, 33.1; (ESI+) m/z calcd for $C_{11}H_{16}NO^+$ $[M+H]^+$: 178.1232; found: 178.1308.

Synthesis of 3-hydroxy-N,N-dimethyl-3-phenylpropan-1-amine (Se3)

Compound **Se2** (1.80 g, 8.39 mmol, 1 eq) was dissolved in 5 mL of distilled water and 1.2 mL of a solution 8 M of KOH was added to the mixture. A white solid precipitate formed and the mixture was extracted with DCM (4 x 20 mL). The organic phases were combined, and the solvent was evaporated under reduced pressure, yielding the free base of compound **Se2**. The oily compound was dissolved in 10 mL of methanol and a couple of drops of KOH 8 M were added in order to provide an alkaline environment. The solution was cooled to 0°C with an ice bath and sodium borohydride (0.47 g, 12.6 mmol, 3 eq) was added to the solution. When all the reactants were dissolved, the ice bath was removed and the mixture was stirred at room temperature for 1.5 hours. After said time, concentrated hydrochloric acid was added dropwise to the solution until acid pH, then the solution was again basified with KOH 8 M. The methanol was evaporated under reduced pressure and the precipitate was dissolved in 100 mL of DCM and washed with alkaline water (4 x 10 mL). The organic phase was dried with anhydrous magnesium sulphate and the solvent was evaporated under reduced pressure. Yield 1.43 g (95%); transparent oil; 1H -NMR (400 MHz, $CDCl_3$): δ_H (ppm) 7.41 – 7.31 (m, 4H), 7.28 – 7.22 (m, 1H), 4.91 (dd, $J = 7.1, 4.7$ Hz, 1H), 2.66 – 2.58 (m, 1H), 2.48 – 2.42 (m, 1H), 2.28 (s, 6H), δ_H 1.86– 1.79 (m, 2H); ^{13}C -NMR (101 MHz, $CDCl_3$): δ_C (ppm) 145.1, 128.1, 126.8, 125.5, 75.3, 58.1, 45.2, 34.7; (ESI+) m/z calcd for $C_{11}H_{18}NO^+$ $[M+H]^+$: 180.14, found: 180.11.

Synthesis of 3-chloro-N,N-dimethyl-3-phenylpropan-1-aminium chloride (Se4)

Compound **Se3** (1.50 g, 8.37 mmol) was dissolved in a small amount of diethyl ether and 5 mL of HCl 2 M in ether were added in order to obtain the hydrochloride salt. The solvent was evaporated under reduced pressure. 10 mL of thionyl chloride were added to the round-bottomed flask and the solution was stirred under reflux. The reaction was followed through TLC (DCM/MeOH/TEA 97:2.5:0.5) and was stopped after 2 hours when the starting material spot was no longer detected. The solvent was evaporated under reduced pressure obtaining the compound as hydrochloride salt. Yield 1.88 g (96%), white solid; 1H NMR (400 MHz, d_6 -DMSO): δ_H (ppm) 10.90 (brs, 1H), 7.52 – 7.49 (m, 2H), 7.45 – 7.35 (m, 3H), 5.31 (dd, $J = 9.1, 5.2$ Hz, 1H), 3.23 – 3.18 (m, 1H), 3.10 – 3.04 (m,

1H), 2.76 (s, 6H), 2.63 – 2.44 (m, 2H); ¹³C-NMR (101 MHz, d6-DMSO): δ_C (ppm) 141.0, 129.3, 129.2, 127.5, 61.1, 42.7, 42.4, 33.6; (ESI+) m/z calcd for C₁₁H₁₇ClN⁺ [M+H]⁺ : 198.10, found: 198.10.

General synthesis of diphenyl diselenides (Se5-Se7)

Under nitrogen atmosphere, magnesium chips (1 eq) were added to a solution of the para substituted bromobenzene (1 eq) in dry ether in a 50 mL three-necked round-bottomed flask. The halobenzene solution was added dropwise at gentle reflux and left stirring for another 30 minutes. Afterwards selenium powder (1 eq) was added maintaining gentle refluxing and the reaction mixture was stirred for another 30 minutes. Then the mixture was poured in a mixture of cracked ice and concentrated hydrochloric acid. The cold mixture was separated and the water phase was extracted with ether (3 x 20 mL). The combined organic layer was dried over anhydrous magnesium sulphate. The solvent was removed under reduced pressure.

1,2-bis(4-(trifluoromethyl)phenyl)diselane (Se5)

Yield 1.21 g (61%); orange oil; ¹H-NMR (400 MHz, CDCl₃): δ_H (ppm) 7.71 (d, *J* = 8.0 Hz, 4H), 7.52 (d, *J* = 8.2 Hz, 4H); ¹³C-NMR (101 MHz, CDCl₃): δ_C (ppm) 135.0, 130.9, 129.9 (q, *J* = 28.9 Hz), 126.2 (q, *J* = 3.7 Hz), 123.4 (q, *J* = 328.1 Hz). (ESI+) m/z calcd for C₁₄H₉F₅Se₂⁺ [M-F+H]⁺ :431.89, found: 431.45.

Synthesis of 1,2-Di-p-tolyldiselane (Se6)

Yield 316 mg (58%); orange oil; ¹H-NMR (400 MHz, CDCl₃): δ_H (ppm) 7.56 (d, *J* = 8.09 Hz, 4H, Ph-H), 7.12 (d, *J* = 8.09 Hz, 4H, Ph-H), 2.39 (s, 6H, Ph-CH₃); ¹³C-NMR (101 MHz, CDCl₃): δ_C (ppm) 138.0 (s, Ph-C), 132.4 (s, Ph-C), 130.0 (s, Ph-C), 126.9 (s, Ph-C), 21.2 (s, Ph-CH₃).

Synthesis of 1,2-Bis(4-fluorophenyl)diselane (Se7)

Yield 331 mg (48%); ¹H-NMR (400 MHz, d6-DMSO): δ_H (ppm) 7.68–7.63 (m, 4H, Ph-H), 7.24–7.18 (m, 4H, Ph-H).

General synthesis of the final compounds (1-CF3, 1-H, 1-CH3 and 1-F)

The diselenide **Se5- Se7** (1 eq) was introduced in a 50 mL round-bottomed flask and dissolved in ethanol, KOH (3 eq) was added and the solution was cooled in an ice bath.

Then, sodium borohydride (6 eq) was added. Once that a colour change was observed (between 30 minutes and 1 hour after the reaction began), compound **Se4** (1 eq) was added to the solution. The reaction was stirred at room temperature for 3 hours. To quench the unreacted NaBH₄, concentrated hydrochloric acid was added to the mixture until acidic pH. Afterwards KOH 8 M was added to basic pH. Ethanol was evaporated under reduced pressure and the precipitate that formed was dissolved in DCM. The solution was washed with alkaline water (3 x 20 mL), dried over magnesium sulphate, then the solvent was evaporated under reduced pressure. The product was purified by column chromatography (silica gel, DCM/MeOH/TEA, 92:7.5:0.5).

Synthesis of N,N-dimethyl-3-phenyl-3-((4-(trifluoromethyl)phenyl)selanyl)propan-1-amine (1-CF₃)

Yield 112 mg (34%); yellow oil; ¹H-NMR (400 MHz, CDCl₃): δ_H (ppm) 7.49 (d, *J* = 8.0 Hz, 2H), 7.43 (d, *J* = 8.4 Hz, 2H), 7.29 – 7.19 (m, 5H), 4.47 (m, 1H), 2.38 – 2.22 (m, 4H), 2.22 (s, 6H); ¹³C-NMR (101 MHz, CDCl₃): δ_C (ppm) 141.72, 134.9, 134.7, 129.7 (q, *J* = 32.6 Hz), 128.6, 127.9, 127.4, 125.6 (q, *J* = 3.7 Hz), 124.2 (q, *J* = 272.19 Hz), 57.8, 46.3, 45.4, 34.0; (ESI+) *m/z* calcd for C₁₈H₂₁F₃NSe⁺ [M+H]⁺ : 388.08, found: 388.09.

Synthesis of N,N-dimethyl-3-phenyl-3-(phenylselanyl)propan-1-amine (1-H)

Yield 155 mg (38%); yellow oil; ¹H-NMR (300 MHz, CDCl₃): δ_H (ppm) 7.43 – 7.40 (m, 2H), 7.28 – 7.18 (m, 8H), 4.36 (m, 1H), 2.31 – 2.19 (m, 4H), 2.17 (s, 6H); ¹³C-NMR (101 MHz, MeOD): δ_C (ppm) 140.6, 135.5, 128.8, 128.4, 128.3, 128.0, 127.3, 127.2, 56.4, 43.7, 42.1, 30.5; (ESI+) *m/z* calcd for C₁₇H₂₂NSe⁺ [M+H]⁺ : 320.09, found: 320.06.

Synthesis of N,N-Dimethyl-3-phenyl-3-(p-tolylselanyl)propan-1-amine (1-CH₃)

Yield: 152 mg (49%); yellow solid; HPLC: 96% (percentage area, 254 nm); ¹H-NMR (400 MHz, CDCl₃): δ_H (ppm) 7.30–7.18 (m, 7H), 7.03 (d, *J* = 7.9 Hz, 2H), 4.30–4.27 (m, 1H), 2.33 (s, 3H), 2.32–2.20 (m, 4H), 2.18 (s, 6H); ¹³C-NMR (101 MHz, CDCl₃): δ_C (ppm) 142.6, 138.0, 136.0, 129.7, 128.4), 127.8, 126.9, 125.7, 58.1, 46.2, 45.6, 33.9, 21.3; (ESI+) *m/z* calcd for C₁₈H₂₄NSe⁺ [M + H]⁺: 334.1068, found 334.1112.

Synthesis of 3-((4-Fluorophenyl)selanyl)-N,N-dimethyl-3-phenylpropan-1-amine (1-F)

Yield: 151 mg (47%); yellow solid; HPLC: 96% (percentage area, 254 nm); ¹H-NMR (400 MHz, CDCl₃): δ_H (ppm) 7.32–7.28 (m, 2H), 7.24–7.15 (m, 3H), 7.12–7.10 (m, 2H),

6.89–6.84 (m, 2H), 4.27–4.23 (m, 1H), 2.37–2.13 (m, 4H), 2.19 (s, 6H); ^{13}C NMR (101 MHz, CDCl_3) δ_{C} (ppm) 164.3, 138.3, 138.2, 128.5, 127.8, 127.1, 116.1, 115.9, 57.9, 46.6, 45.3, 33.4; (ESI $^+$) calcd for $\text{C}_{17}\text{H}_{22}\text{FNSe}^+$ $[\text{M} + \text{H}]^+$: 338.0818, found 338.0065.

Synthesis of compound SeA8

To a solution of styrene (1.7 eq, 0.231 g, 2.22 mmol) in DCM (20 mL) was added diphenyl diselenide (1 eq, 0.414 g, 1.33 mmol), (diacetoxyiodo)benzene (2.3 eq, 1.000 g, 3.10 mmol) and sodium azide (0.346 g, 5.32 mmol), the mixture was deaerated and purged with nitrogen. After stirring at r.t. overnight a saturated solution of NaHCO_3 (10 mL) was added, the organic phase was separated and the aqueous phase was extracted with DCM (2 x 20 mL). The organic phases were reunited and evaporated at reduced pressure. The so obtained yellow liquid was purified by column chromatography (silica gel, pure hexane then 90:10 hexane/diethyl ether) obtaining a colourless liquid.

Yield 1.010 g, 60%. ^1H -NMR (400 MHz, CDCl_3): δ_{H} (ppm) 7.50 – 7.48 (m, 2H), 7.34 – 7.24 (m, 8H), 4.40 (dd, $J = 9.7, 5.7$ Hz, 1H), 3.88 (dd, $J = 12.5, 9.8$ Hz, 1H), 4.40 (dd, $J = 12.5, 5.8$ Hz, 1H); (ESI $^+$) m/z calcd for $\text{C}_{14}\text{H}_{14}\text{N}_3\text{Se}^+$ $[\text{M} + \text{H}]^+$: 304.19, found: 304.15.

Synthesis of compound SeA9

Compound **SeA8** (1 eq, 1.010 g, 3.33 mmol) was dissolved in anhydrous THF (20 mL) and the solution was deaerated and purged with nitrogen. The system was equipped with a reflux apparatus and a solution of 1 M LiAlH_4 (1.5 eq, 5 mL, 5.01 mmol) was added portion wise to the mixture, after the addition the mixture was stirred for 2 h. When the starting material resulted disappeared by TLC (DCM/MeOH/ NH_3 , 97.5/2.0/0.5) wet diethyl ether was slowly added to the solution then distilled water was added. The so obtained bi-phasic mixture was filtered, the aqueous phase was separated and extracted with diethyl ether (2 x 20 mL). The so-obtained organic phase was dried over magnesium sulphate and evaporated at reduced pressure obtaining a viscous colourless liquid that was purified by column chromatography (silica gel, DCM/MeOH/ NH_3 , 97.5/2.0/0.5).

Yield, 503 mg, 55%. ^1H -NMR (400 MHz, CDCl_3): δ_{H} (ppm) 7.53– 7.51 (m, 2H), 7.37 – 7.27 (m, 8H), 4.43 (dd, $J = 9.6, 5.7$ Hz, 1H), 3.91 (dd, $J = 12.6, 9.6$ Hz, 1H), 3.74 (dd, $J = 12.6, 5.7$ Hz, 1H); (ESI $^+$) m/z calcd for $\text{C}_{14}\text{H}_{16}\text{NSe}^+$ $[\text{M} + \text{H}]^+$: 278.04, found: 277.80.

Synthesis of compound SeA1

Compound **SeA9** (1 eq, 100 mg, 0.36 mmol) was dissolved in iodoethane (2 mL), DIPEA (3 eq, 190 μ L, 1.08 mmol) was added to the solution that was deaerated, purged with nitrogen and refluxed under stirring at 90 °C for 2 h. When the reaction resulted concluded by TLC (DCM/MeOH/NH₃, 97.5/2.0/0.5) the mixture was cooled down to r.t. and the solvent was evaporated under reduced pressure, DCM (30 mL) was added and the so obtained solution was washed with 1 M KOH (3 x 10 mL), dried over magnesium sulphate and evaporated at reduced pressure obtaining a viscous brownish liquid which was purified by chromatographic column (silica gel, DCM/MeOH/NH₃, 97.5/2.0/0.5). The hydrochloride salt was then prepared dissolving the product in anhydrous diethyl ether and adding drop by drop a 2 M solution of HCl in diethyl ether. (119 mg, 89%).

White solid, 119 mg, 89%. ¹H-NMR (400 MHz, CD₃OD): δ_{H} (ppm) 7.63-7.61 (m, 2H), 7.46 – 7.34 (m, 8H), 4.82 (dd, $J = 10.7, 4.5$ Hz, 1H), 4.06 (dd, $J = 14.0, 10.7$ Hz, 1H), 3.52 (dd, $J = 14.0, 4.5$ Hz, 1H), 3.20 – 3.06 (m, 4H), 1.19 – 1.10 (m, 6H); ¹³C-NMR (101 MHz, CD₃OD): δ_{C} (ppm) 138.8, 137.2, 130.6, 130.4, 130.3, 129.7, 129.2, 129.0, 56.3, 41.8, 9.0, 8.5; (ESI+) m/z calcd for C₁₈H₂₄NSe⁺ [M+H]⁺ : 334.11, found: 333.95.

Synthesis of compound SeA2

Compound **SeA9** (1 eq, 100 mg, 0.36 mmol) was dissolved in EtOH (5 mL), benzyl bromide (2 eq, 87 μ L, 124 mg, 0.73 mmol) and DIPEA (3 eq, 190 μ L, 1.08 mmol) were added to the solution that was deaerated, purged with nitrogen and refluxed under stirring at 90 °C for 2 h. When the reaction resulted concluded by TLC (Hexane/diethyl ether, 6/1) the mixture was cooled down to r.t. and the solvent was evaporated under reduced pressure, DCM (30 mL) was added and the so obtained solution was washed with 1 M KOH (3 x 10 mL), dried over magnesium sulphate and evaporated at reduced pressure. The crude product was purified by column chromatography (silica gel, Hexane/DCM, 9/1) obtaining a viscous liquid. The hydrochloride salt was then prepared dissolving the product in anhydrous diethyl ether and adding drop by drop a 2 M solution of HCl in diethyl ether.

Yield 93 mg, 53 %. ¹H-NMR (400 MHz, CD₃OD): δ_{H} (ppm) 7.58– 7.20 (m, 18H), 6.80 – 6.78 (m, 2H), 4.56 – 4.43 (m, 3H), 4.31 (d, $J = 12.7$ Hz, 1H), 4.17 (d, $J = 12.6$ Hz, 1H), 3.87 (dd, $J = 13.9, 10.9$ Hz, 1H), 3.61 (dd, $J = 13.9, 5.0$ Hz, 1H); ¹³C-NMR (101 MHz, CD₃OD): δ_{C} (ppm) 136.6, 136.4, 131.5, 131.3, 130.2, 129.3, 129.2, 129.1, 128.9,

128.2, 127.3, 126.5, 58.9, 58.4, 54.1, 40.2; (ESI+) m/z calcd for $C_{28}H_{28}NSe^+$ $[M+H]^+$: 458.14, found: 458.05.

Synthesis of compound SeA3

Compound **SeA9** (1 eq, 100 mg, 0.36 mmol) was dissolved in EtOH (5 mL), bis(2-bromoethyl) ether (1 eq, 84 mg, 0.36 mmol) and DIPEA (3 eq, 190 μ L, 1.08 mmol) were added to the solution that was deaerated, purged with nitrogen and refluxed under stirring at 90 °C for 2 h. When the reaction resulted concluded by TLC (DCM/MeOH/TEA, 95/4.5/0.5) the mixture was cooled down to r.t. and the solvent was evaporated under reduced pressure, DCM (30 mL) was added and the so obtained solution was washed with 1 M KOH (3 x 10 mL), dried over magnesium sulphate and evaporated at reduced pressure. The crude product was purified by column chromatography (silica gel, DCM/MeOH/TEA, 95/4.5/0.5) obtaining a viscous liquid. The hydrochloride salt was then prepared dissolving the product in anhydrous diethyl ether and adding drop by drop a 2 M solution of HCl in diethyl ether.

White solid, 70 mg, 56 %. 1H -NMR (400 MHz, CD_3OD): δ_H (ppm) 7.52 - 7.50 (m, 2H), 7.38 - 7.28 (m, 8H), 4.92 (dd, $J = 10.4, 4.7$ Hz, 1H), 4.13 - 4.07 (m, 1H), 4.00 - 3.69 (m, 4H), 3.64 (dd, $J = 13.6, 4.7$ Hz, 1H), 3.28 - 2.99 (m, 4H); ^{13}C -NMR (101 MHz, CD_3OD): δ_C (ppm) 139.2, 136.9, 130.5, 130.2, 130.0, 129.5, 129.1, 129.1, 64.6, 62.2, 53.7, 41.4; (ESI+) m/z calcd for $C_{18}H_{22}NOSe^+$ $[M+H]^+$: 347.95, found: 347.98.

Synthesis of compound SeA4

Compound **SeA9** (1 eq, 150 mg, 0.54 mmol) was dissolved in EtOH (6 mL), α, α' -Dibromo-*o*-xylene (1 eq, 154 mg, 0.54 mmol) and K_2CO_3 (2 eq, 225 mg, 1.62 mmol) were added to the solution that was deaerated, purged with nitrogen and refluxed under stirring at 90 °C for 3 h. When the reaction resulted concluded by TLC (hexane/diethyl ether, 6/1) the mixture was cooled down to r.t. and the solvent was evaporated under reduced pressure, DCM (30 mL) was added and the so obtained solution was washed with 1 M KOH (3 x 10 mL), dried over magnesium sulphate and evaporated at reduced pressure. The crude product was purified by column chromatography (silica gel, hexane/diethyl ether, 89.75/9.75/0.5) obtaining a viscous liquid. The hydrochloride salt was then prepared dissolving the product in anhydrous diethyl ether and adding drop by drop a 2 M solution of HCl in diethyl ether.

Yield 45 mg, 20 %. ¹H-NMR (400 MHz, CDCl₃): δ_H (ppm) 7.45 – 7.43 (m, 2H), 7.30 – 7.16 (m, 12H), 4.58 (t, *J* = 7.8 Hz, 1H), 4.03 – 3.96 (m, 4H), 3.41 (dd, *J* = 7.8, 2.3 Hz, 2H); ¹³C-NMR (101 MHz, CD₃OD): δ_C (ppm) 138.9, 137.1, 134.8, 130.5, 130.3, 130.1, 129.6, 129.1, 128.7, 123.8, 60.4, 59.9, 43.2; (ESI+) *m/z* calcd for C₂₂H₂₁NSe⁺ [M+H]⁺ : 379.08, found: 380.00.

Synthesis of compound SeA5

Compound **SeA9** (1 eq, 100 mg, 0.36 mmol) was dissolved in dry MeOH (5 mL), benzaldehyde (1 eq, 37 μL, 0.36 mmol) was added to the solution that was deaerated, purged with nitrogen and stirred at r.t. for 2 h. When the starting products resulted completely disappeared by TLC (hexane/diethyl ether 6/1) the mixture was cooled in an ice bath and NaBH₄ (1.2 eq, 16 mg, 0.43 mmol) was added in portions. The ice bath was removed and the mixture was stirred at r.t. for other 2 h. When the TLC revealed the complete conversion of the imine intermediate to the secondary amine product (hexane/diethyl ether 6/1), the solution was cooled in an ice bath and first acidified to pH = 2 by the addition of 1 M HCl and the basified to pH = 14 by the addition of an 8 M KOH solution. The solvent was evaporated under reduced pressure, DCM (30 mL) was added and the so obtained solution was washed with 1 M KOH (3 x 10 mL), dried over magnesium sulphate and evaporated at reduced pressure obtaining a viscous liquid. The crude product was purified by column chromatography (silica gel hexane/EtOAc, 9/1 with 0.5% TEA). The hydrochloride salt was then prepared dissolving the product in anhydrous diethyl ether and adding drop by drop a 2 M solution of HCl in diethyl ether.

Yield 125 mg, 86 %. ¹H-NMR (400 MHz, CD₃OD): δ_H (ppm) 7.58– 7.20 (m, 18H), 6.80 – 6.78 (m, 2H), 4.56 – 4.43 (m, 3H), 4.31 (d, *J* = 12.7 Hz, 1H), 4.17 (d, *J* = 12.6 Hz, 1H), 3.87 (dd, *J* = 13.9, 10.9 Hz, 1H), 3.61 (dd, *J* = 13.9, 5.0 Hz, 1H), 3.44 – 3.41 (m, 1H); ¹³C-NMR (101 MHz, CD₃OD): δ_C (ppm) 136.6, 136.4, 131.5, 131.3, 130.2, 129.3, 129.2, 129.1, 128.9, 128.2, 127.3, 126.5, 58.9, 58.4, 54.1, 40.2; (ESI+) *m/z* calcd for C₂₁H₂₂NSe⁺ [M+H]⁺ : 368.08, found: 367.93.

Synthesis of compound SeA6

Compound **SeA9** (1 eq, 150 mg, 0.36 mmol) was dissolved in dry MeOH (6 mL), *p*-nitrobenzaldehyde (1 eq, 82 mg μL, 0.36 mmol) was added to the solution that was deaerated, purged with nitrogen and stirred o/n at r.t. The day after the reaction was checked through TLC (6/1 hexane/diethyl ether) and the starting material resulted

disappeared, therefore the mixture was cooled in an ice bath and NaBH_4 (1.2 eq, 16 mg, 0.43 mmol) was added portion wise. The ice bath was removed and the mixture was stirred at r.t. for other 2 h. When the TLC revealed the complete conversion of the imine intermediate to the secondary amine product (6/1 hexane/diethyl ether) the solution was cooled in an ice bath and first acidified to $\text{pH} = 2$ by the addition of 1 M HCl and the basified to $\text{pH} = 14$ by the addition of an 8 M KOH solution. The solvent was evaporated under reduced pressure, DCM (30 mL) was added and the so obtained solution was washed with 1 M KOH (3 x 10 mL), dried over magnesium sulphate and evaporated at reduced pressure obtaining a viscous liquid. The crude product was purified by column chromatography (silica gel hexane/EtOAc, 9/1 with 0.5% TEA). The hydrochloride salt was then prepared dissolving the product in anhydrous diethyl ether and adding drop by drop a 2 M solution of HCl in diethyl ether.

Yield 135 mg, 84 %. $^1\text{H-NMR}$ (400 MHz, DMSO-d_6): δ_{H} (ppm) 8.24 (d, $J = 8.9$ Hz, 2H), 7.75 (d, $J = 8.9$ Hz, 2H), 7.44 – 7.42 (m, 2H), 7.36 – 7.25 (m, 8H), 4.88 (dd, $J = 9.7$, 5.6 Hz, 1H), 4.26 (s, 2H), 3.74 – 3.72 (m, 2H); $^{13}\text{C-NMR}$ (101 MHz, DMSO-d_6): δ_{C} (ppm) 148.2, 139.5, 138.2, 134.8, 134.0, 131.9, 129.8, 129.2, 128.7, 128.5, 128.4, 124.0, 50.5, 49.8, 42.2; (ESI+) m/z calcd for $\text{C}_{21}\text{H}_{21}\text{N}_2\text{O}_2\text{Se}^+$ $[\text{M}+\text{H}]^+$: 413.08, found: 412.89.

Synthesis of compound SeA7

Compound **SeA5** (1 eq, 100 mg, 0.27 mmol) was dissolved in iodoethane (5 mL), the solution that was deaerated, purged with nitrogen and refluxed under stirring at 90 °C for 2 h. When the reaction resulted concluded by TLC (hexane/diethyl ether, 6/1) the mixture was cooled down to r.t. and the solvent was evaporated under reduced pressure, DCM (30 mL) was added and the so obtained solution was washed with 1 M KOH (3 x 10 mL), dried over magnesium sulphate and evaporated at reduced pressure obtaining a viscous colourless liquid. The crude product was purified by column chromatography (silica gel, hexane/EtOAc/TEA, 97/2.5/0.5) obtaining a viscous liquid. The hydrochloride salt was then prepared dissolving the product in anhydrous diethyl ether and adding drop by drop a 2 M solution of HCl in diethyl ether.

Yield 69 mg, 59 %. $^1\text{H-NMR}$ (400 MHz, CD_3OD): δ_{H} (ppm) 7.49 – 7.04 (m, 15H), 4.75 (dd, $J = 10.8$, 4.7 Hz, 0.5H), 4.59 (dd, $J = 10.5$, 4.6 Hz, 0.5H), 4.37 – 4.20 (m, 2H), 4.05 (dd, $J = 13.9$, 10.6 Hz, 0.5H), 3.86 (dd, $J = 13.9$, 10.7 Hz, 0.5H), 3.61 – 3.54 (m, 1H), 3.28 – 3.11 (m, 2H), 1.22 (m, 3H); $^{13}\text{C-NMR}$ (101 MHz, CD_3OD): δ_{C} (ppm) 138.8, 138.3, 137.5, 137.3, 132.3, 132.2, 131.4, 130.6, 130.5, 130.5, 130.4, 130.3, 130.2, 130.1, 129.7,

129.6, 129.0, 128.9, 128.8, 128.6, 58.8, 56.3, 56.1, 50.9, 50.2, 42.1, 41.4, 27.9, 8.9, 8.5; (ESI+) m/z calcd for $C_{23}H_{26}NSe^+$ $[M+H]^+$: 396.12, found: 395.94.

NMR study of the reaction with H_2O_2

In order to study the oxidation reaction of compounds **1-CF₃**, **1-H** and **1-CH₃**, their reaction with H_2O_2 was investigated by 1H -NMR in an aqueous environment at room temperature (22°C). The compounds were used as hydrochloride salts (28 mM: **1-CF₃** = 4.9 mg, 11.6 μ mol, **1-H** = 5.6 mg, 17.6 μ mol; **1-CH₃** = 5.3 mg, 14.5 μ mol,) and dissolved in D_2O , t_0 spectra were registered with 8 scans, 1 dummy scan and a delay time d_1 of 2 s. Afterwards, H_2O_2 (1.1 eq, H_2O_2 3% w/w solution: **1-CF₃** = 12 μ L, **1-H** = 18 μ L, **1-CH₃** = 15 μ L) was added and the spectra were registered every 2 or 5 minutes until the reaction was complete. The reaction was studied plotting the results of the integration of the signals corresponding to the protons bound to the carbon adjacent to the selenium atom for the starting material and the two diastereoisomers and the allylic-hydrogens for the cinnamylamine. 1H - ^{77}Se HMBC NMR spectra were acquired using a repetition delay of 1 s; a total of 300 experiments of 16–40 scans were accumulated and processed with a magnitude calculation; an evolution delay of 33.3 ms was used for 1H - ^{77}Se long-range coupling constants; the spectral width was 13 ppm in F_2 and 1000 ppm in F_1 . Zero-filling in the F_1 and F_2 dimensions, multiplication with a Gaussian function (in F_2) and a squared sine function (in F_1) were performed prior to 2D Fourier transformation.

ESI-MS study of the reaction with H_2O_2

Mechanistic ESI-MS studies were performed under the same experimental conditions (concentration and temperature) used in the NMR analysis, except for the fact that Milli-Q water was used instead of D_2O . Solutions were diluted 1:1000 in methanol before the analysis. Spectra were recorded by direct infusion ESI on a Thermo Fisher Scientific (Waltham, MA) LCQ Fleet ion trap mass spectrometer. ESI parameter for positive ionization mode are here described: 4.0 kV spray voltage, 225 °C capillary temperature, 5 μ L/min flow rate. For negative ionization mode: 5.0 kV spray voltage, 180 °C capillary temperature, 5 μ L/min flow rate.

Lab scale oxidation*Oxidation of compound SeA3*

Compound **SeA3** (1.5 g, 3.9 mmol, 1 eq) was dissolved in 120 ml of distilled H₂O in a round-bottom flask, H₂O₂ (530 μ L, 4.7 mmol, 1.2 eq) was then added to the mixture. The solution was stirred o/n and checked through TLC, the day after another 1.2 eq of H₂O₂ were added in order to consume the remained reactant. After a couple of hours and after the TLC showed that the reactant was fully consumed, the mixture was filtered on a Buchner funnel in order to remove the yellow precipitate consisting of insoluble by-products and water was evaporated under reduced pressure adding ethanol to facilitate the evaporation. Once the solvent was fully removed, the yellowish precipitate was washed with a small aliquot of acetone, forming a white solid which correspond to the desired product that was filtered on a Gooch funnel, weighted and characterized.

Compound SeA3OX (Morpholine•HCl): white solid, 435 mg, 90%; ¹H-NMR (400 MHz, CD₃OD): δ_{H} (ppm) 3.91 – 3.88 (m, 4H), 3.25 – 3.22 (m, 4H); ¹³C-NMR (101 MHz, CD₃OD): δ_{C} (ppm) 64.9, 44.6; (ESI+) m/z calcd for C₄H₁₀NO⁺ [M+H]⁺ : 88.08, found: 87.93.

Oxidation of compound SeA1, SeA4, SeA5, SeA7

The selenoamine (1 eq) is dissolved in a mixture of MeOH:H₂O 85:15 in a round-bottom flask, 1.2 eq of H₂O₂ were then added to the mixture and the solution was stirred o/n. The day after checking the reaction through TLC, another 1.2 eq were added to the mixture in order to allow the remained starting material to react, after a couple of hours the solvent was evaporated under reduced pressure. Water was added to the reaction flask in order to dissolve the desired product and eliminate most of the insoluble by-products, the following steps are the same of the previous section, leading to the formation of a white solid which was weighted and characterized.

Compound SeA1OX (Diethylamine•HCl): white solid; 55 mg, 82%; ¹H-NMR (400 MHz, CD₃OD): δ_{H} (ppm) 3.06 (q, $J = 7.3$ Hz, 4H), 1.32 (t, $J = 7.3$ Hz, 6H); ¹³C-NMR (101 MHz, CD₃OD): δ_{C} (ppm) 43.5, 11.6; (ESI+) m/z calcd for C₄H₁₂N⁺ [M+H]⁺ : 74.10, found: 73.9.

Compound SeA4OX (Isoindoline•HCl): brown solid, 69 mg, 69%; ¹H-NMR (400 MHz, CD₃OD): δ_{H} (ppm) 7.42 – 7.38 (m, 4H), 4.65 (s, 4H); ¹³C-NMR (101 MHz,

CD₃OD): δ_C (ppm) 134.4, 127.6, 122.4, 49.2; (ESI+) m/z calcd for C₈H₁₀N⁺ [M+H]⁺ : 120.08, found: 119.98.

Compound SeA5OX (Benzylamine•HCl): white solid, 183 mg, 85%; ¹H-NMR (400 MHz, CD₃OD): δ_H (ppm) 7.49 – 7.40 (m, 5H), 4.12 (s, 2H); ¹³C-NMR (101 MHz, DMSO-*d*₆): δ_C (ppm) 134.4, 130.2, 130.0, 129.9, 44.4; (ESI+) m/z calcd for C₇H₁₀N⁺ [M+H]⁺ : 108.08, found: 107.88.

Compound SeA7OX (N-ethylbenzylamine•HCl): white solid, 114 mg, 75%; ¹H-NMR (400 MHz, CD₃OD): δ_H (ppm) 7.52 – 7.44 (m, 5H), 4.19 (s, 2H), 3.12 (q, *J* = 7.3 Hz, 2H), 1.34 (t, *J* = 7.3 Hz, 3H); ¹³C-NMR (101 MHz, CD₃OD): δ_C (ppm) 132.7, 130.9, 130.7, 130.3, 52.0, 43.8, 11.5; (ESI+) m/z calcd for C₉H₁₄N⁺ [M+H]⁺ : 136.11, found: 135.97.

Oxidation of compound SeA2, SeA6

The procedure is the same of the previous section, with the only difference that after the evaporation of the mixture MeOH:H₂O and the addition of water the mixture was acidified with HCl 0.2 M until pH 2-3 and heated for a couple of hours at 40°C in order to facilitate the hydrolysis of the enamine.

Compound SeA2OX (Dibenzylamine•HCl): white solid, 47 mg, 33%; ¹H-NMR (400 MHz, CD₃OD): δ_H (ppm) 7.52 – 7.45 (m, 10H), 4.25 (s, 4H); ¹³C-NMR (101 MHz, DMSO-*d*₆): δ_C (ppm) 131.8, 130.1, 128.8, 128.5, 49.5; (ESI+) m/z calcd for C₁₄H₁₆N⁺ [M+H]⁺ : 198.13, found: 198.06.

Compound SeA6OX (p-nitrobenzylamine): white solid, 15 mg, 45%; ¹H-NMR (400 MHz, DMSO-*d*₆): δ_H (ppm) 8.78 (brs, 3H), 8.26 (d, *J* = 8.2 Hz, 2H), 7.80 (d, *J* = 8.2 Hz, 2H), 4.17 (s, 2H); ¹³C-NMR (101 MHz, DMSO-*d*₆): δ_C (ppm) 147.4, 141.7, 130.2, 123.5, 41.3; (ESI+) m/z calcd for C₇H₉N₂O₂⁺ [M+H]⁺ : 153.07, found: 152.93.

Molecular docking

SERT structure file was retrieved from the RCSB Protein Data Bank (PDB, www.rcsb.org, PDB ID: 5I73) and was prepared with the Protein Preparation Wizard included in the Schrodinger suite,⁴²⁹ using default settings, i.e., adding hydrogens, removing waters further than 5 Å from the ligand, adjusting charges, capping termini, optimizing hydrogen bond clusters and performing a final minimization under the OPLS3e force field.⁴⁷⁹ The ligand was prepared for docking with Ligprep tool under OPLS3e force field, using Epik ionizer all possible states in the 7±2 pH range were

generated, including the possible metal binding states, allowing therefore the interaction with the potassium ions. The docking protocol consisted of rigid-receptor/flexible ligand docking, which was performed with Glide, under the OPLS3e force field, with default settings, i.e., a 0.80 scaling factor for the ligand atom van der Waals radius and 0.15 for the partial charge cut-off, with generation of 1 pose per ligand. A cubic grid (30x30x30 Å) was prepared with Receptor Grid Generation tool included in Schrodinger suite and generated selecting as centre the co-crystallized ligand.

3.2 Phosphodiesterase inhibitors

3.2.1 Introduction

Phosphodiesterases (PDEs) are a family of 11 different enzymes that convert second messengers cyclic adenosine monophosphate (cAMP) and cyclic guanosine monophosphate (cGMP) into their linear forms, 5'-adenosine monophosphate (5'-AMP) and 5'-guanosine monophosphate (5'GMP) breaking their phosphodiesterase bond.⁵⁴⁴ This activity is involved in the control of several different cellular pathways.^{545,546}

PDE enzymes have been considered as promising drug targets since their discovery because, in general, to gain quicker and more pronounced variations on the concentration of ligand or second messenger, it is more efficient to interfere with degradation pathways compare to act on biosynthesis. Moreover, the low levels of cAMP and cGMP make it easier to regulate these enzymes using competitive inhibitors.

It has been demonstrated that increased levels of cAMP and cGMP can promote CREB phosphorylation and stimulate neuronal plasticity thanks to the activation of protein kinase A and protein kinase G, respectively.^{547,548}

Moreover, cAMP and cGMP have been shown to play a central role in regulating synaptic plasticity, learning, and memory as they support many neuronal functions, from regulating neurotransmitter release to the firing properties of neurons. PDE inhibition therefore modulates cyclic nucleotide signalling related to long-term potentiation processes, thus improving the storage of new memories.⁵⁴⁹⁻⁵⁵¹

In this context, PDE inhibitors are being developed and investigated to fight neurodegeneration and CNS-related diseases, even though, as previously stated, the major targets in this field are the neurotransmitter systems, such as the serotonergic, cholinergic, and monoaminergic systems.

Among all PDEs, only a few have been linked to signalling pathways involved in neurodegenerative disorders and specifically, PDE4, PDE5, and PDE9 have been linked to Alzheimer's Disease, as demonstrated by Cameron and co-workers,⁵⁵² even though recent studies have been focused on the activation of cGMP signalling (which include PDE5 and PDE9) as potential therapeutic effect on AD. Specifically the high affinity combined with the highest selectivity among the PDEs for cGMP, made the attention to be focused on PDE9 for protecting against neurodegeneration.⁵⁵³⁻⁵⁵⁵

Historically, theophylline is the first discovered PDE inhibitor, and even though it is weak and non-selective, it has found use in treating asthma and other respiratory diseases.

Sildenafil is a PDE5-selective inhibitor, and it was the first drug approved by the FDA for the treatment of erectile dysfunction (ED). Apremilast is a PDE4 selective inhibitor used as a medication to treat psoriasis and psoriatic arthritis, and it works by raising cAMP levels and partially inhibiting the production of many pro-inflammatory mediators. BAY73-6691 and PF-04447943, two selective PDE9 inhibitors, showed increased synaptic transmission and early long-term potentiation (LTP) in hippocampal slices of old rats, and improved plasticity on hippocampal synapsis and cognitive functions in rodents, respectively.^{556,557}

In recent years, new drug candidates have been synthesized and semi-synthetically obtained. Moreover, growing evidence suggested studying the PDE activity of natural products as an appealing approach.^{425,558,559} Flavones and isoflavones have been shown to be promising new classes of PDE inhibitors. In particular, luteolin, icariin, genistein, biochanin A, osajin, and pomiferin have shown interesting activity towards these enzymes. Our research group studied the activity of the prenylated isoflavones extracted from *M. Pomifera*, osajin and pomiferin, and of the different derivatives obtained using a semi-synthetic approach. The compounds were screened in vitro against PDE5A, and a docking study was performed to analyse the interaction mode with the target protein.⁴²⁵

The binding site of PDEs is composed by three pockets: the glutamine, purine-selective and hydrophobic clamp pocket (H), the solvent-filled side pocket (S) and the metal binding pocket (M). (Figure 3.1.2) Each of these pockets contains one or more residues that are important for the binding of the ligand with the enzyme and are conserved in all PDEs, namely a glutamine (Gln) is always present in the H pocket (which is always bounded through H-bond to the substrates) together with a phenylalanine (Phe) which is reported to give important pi-stacking interactions. H is also composed by a variable hydrophobic residue such as valine, leucine or isoleucine on the opposite site of the claw that are also involved in the binding.⁵⁶⁰

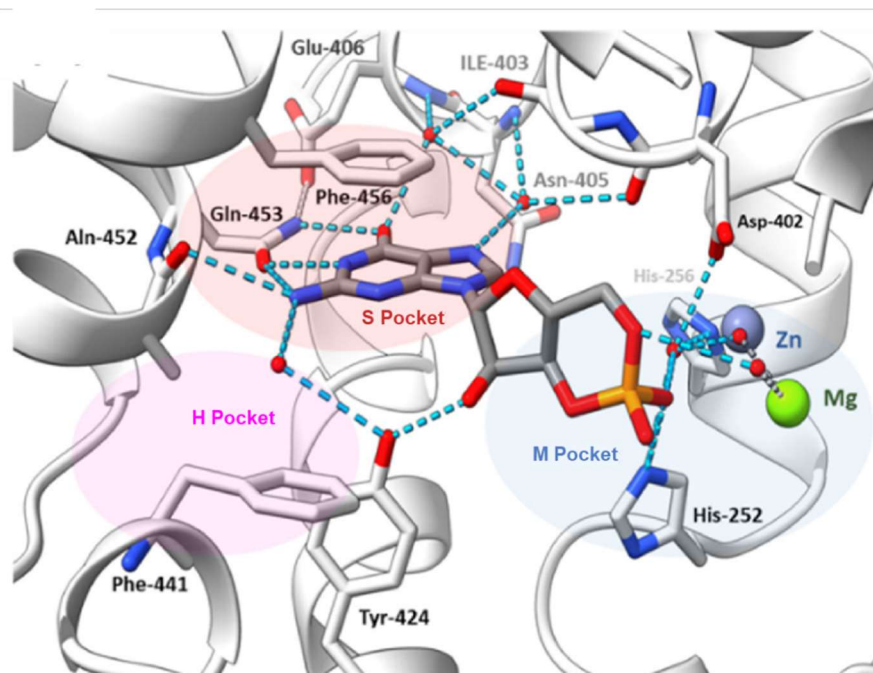


Figure 3.2.1. PDE9A crystal structure complexed with cGMP (PDB ID: 3DYI),⁵⁶¹ showing the main amino acids interactions of cGMP with PDE9 and the three major subpockets: • S pocket, • M pocket and • H pocket. Adapted from Zheng and Zhou.⁵⁶²

In a first work, we took a robust computational approach to explore the semi-synthetic chemical space offered by the isoflavones class of molecules with the goal of identifying new hit compounds obtainable with simple 1-2 step modifications starting from the natural compounds. Specifically, starting from the natural isoflavone scaffolds, a set of molecules was generated using a combinatorial chemistry approach, and then a docking protocol was designed, validated, and applied to study the interaction of these molecules with three PDE families (PDE4, PDE5, and PDE9). MD simulations were then performed to study more in detail the stability of the complexes and to confirm the interaction with the recurrent residues among the PDE inhibitors.⁵⁶³

In a second work,⁵⁶⁴ virtual screening, docking, and MD were used to identify the most suitable compound within an internal database to be tested in rat organotypic hippocampal slices exposed to kainate, a widely accepted *in vitro* epileptic model.^{565–567}

3.2.2 Combinatorial library generation, molecular docking and molecular dynamics simulations for enhancing the isoflavone scaffold in phosphodiesterase inhibition

Results and discussion

The first part of this study consisted in the design of a library of 1520 compounds all obtainable from the studied natural isoflavone scaffolds by feasible one-step reaction, specifically by alkylations, esterifications and sulphonylations of the aromatic hydroxyls. In particular, the hydroxyls in position 7 and 4' on the scaffolds of genistein, daidzein and tectorigenin, the 4' hydroxyl of prunetin, the 7 hydroxyl of formononetin, biochanin A and osajin and, finally, the 3' and 4' hydroxyls of pomiferin were considered. Modifications were designed by introducing on the scaffolds a combination of 19 chemically different moieties, (Figure 3.2.3) consisting of polar and non-polar, protonable and deprotonable molecular fragments, some of which derived from the scaffolds of known PDE inhibitors (Figure 3.2.2).

The receptor files were retrieved from RCSB PDB: the 3GWT crystal structure, representing the PDE4B, in complex with the selective quinoline inhibitor GSK256066, the 2H42 crystal structure, representing the PDE5A, in complex with the selective inhibitor sildenafil, the 3JSW crystal structure, which represents PDE9A, in complex with a specific inhibitor.

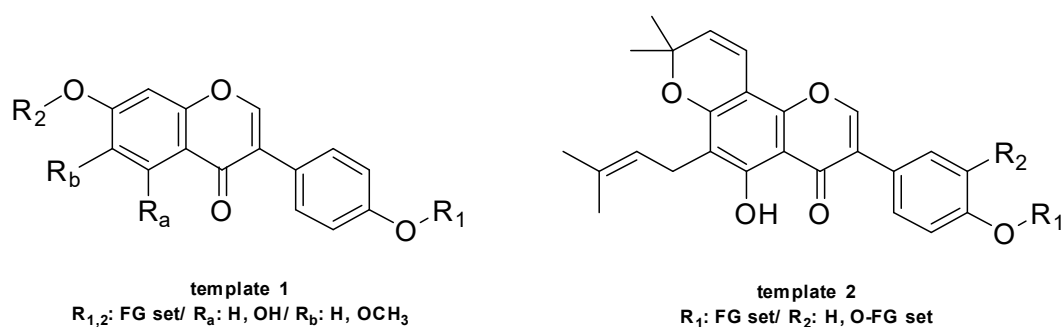


Figure 3.2.2. Molecular templates used to generate the semi-synthetically modified isoflavones set. Template 1 resumes the common scaffold of genistein, daidzein, tectorigenin, prunetin, formononetin and biochanin A. Osajin and pomiferin are based on template 2.

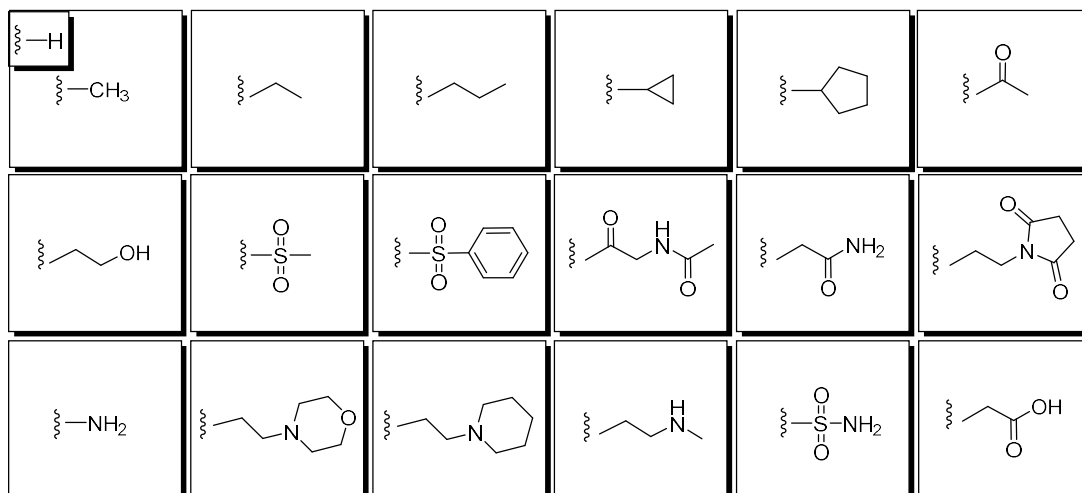


Figure 3.2.3. Molecular fragments used to generate the semi-synthetically modified isoflavones set.

First of all, the method was validated following a multi stage approach. The first step consisted in the re-docking of the crystallized ligands to their native structures and RMSD values were calculated between the two poses showing for the three systems values lower than 2.50 Å (Table 3.2.1).⁵⁶⁸

Table 3.2.1. RMSD values for the re-docking of the crystallized ligands.

Receptor	PDB ID	RMSD
PDE5A	2H42	0,59
PDE4B	3GWT	1,54
PDE9A	3JSW	0,61

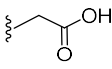
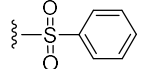
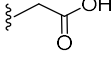
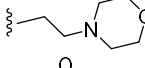


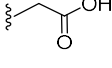





The second step consisted in an enrichment study using decoy sets, which consisted in a set of inactive or presumed inactive compounds generated based on similar physical properties but structurally different from 20 known active ligands (IC_{50} in the low micromolar range) chosen for each PDE. After ranking the decoy set by docking score, enrichment factor (EF) which can be thought as the amount of active ligands among the top-scoring compounds compared to their amount in the whole set, the efficiency (Eff) in distinguishing active ligands from inactive were calculated and Receiver Operating Characteristic (ROC) curves were plotted and represent a plot of true-positive versus false-positive rates for all compounds. The study showed that for all PDEs there is more than the 70% of chances for the active compounds to be ranked with a higher docking score than the non-actives or decoys and that , on average and for all the tested PDE isoforms, more than the 40% of the active compounds can be found in the top 5% of the ranking with an efficiency in scoring actives before decoys in an average percentage of more than 75%. Lastly, a cross docking experiment was conducted in order to determine

the ability of the protocol in identifying the right poses and interactions with the residues in the PDEs binding site in which for each PDE, five different crystal structures were considered with five different ligands that were docked on all the structures with SP and XP calculating the RMSD values showing better performance for XP docking, as expected. The validated protocol was used for the screening of the library with the intent of obtaining three separated enriched sets of compounds, one for each of the three PDE isoforms. The previously described computer-based library set composed of 1520 molecules were docked with SP precision on the receptors PDE4B (PDB ID: 3GWT), PDE5A (PDB ID: 2H42) and PDE9A (PDB ID: 3JSW) poses ranked within the range from -3 to -12 kcal/mol. Basing on the results of the enrichment study, the top 5 % of the rankings obtained for each PDE were selected. As a result, 76 molecules for each PDE were highlighted to be used in the next part of the study, which consisted of a XP docking.

The poses produced were inspected to assess the presence of specific interactions that are known in the literature for being targeted by PDEs inhibitor drugs, as previously stated namely Gln443/F446 for PDE4B (PDB ID: 3GWT), Gln817/F820 for PDE5A (PDB ID: 2H42) and Gln453/F456 for PDE9A (PDB ID: 3JSW). The poses were ordered in terms of the interaction values with each of the specified residue, rejecting the poses that did not show both these interactions, resulting in a semi-automatic inspection approach.

Analysing the poses that give interaction with both the residues, it is possible to notice some recurrences and to start tracing a structure-activity relationship (SAR) (As an example the poses obtained with XP molecular docking on PDE9 are reported in Table 3.2.2), starting with the scaffold preference that was template 1 for PDE4 and PDE9.

Table 3.2.2. Poses of the selected compounds obtained with XP docking on PDE9.

Name	GScore XP (kcal/mol)	Scaffold	R ₁	R ₂
gen.227	-11.416	Genistein (template 1)		
gen.284	-11.220	Genistein (template 1)		
tec.286	-11.215	Tectorigenin (template 1)		
gen.18	-11.033	Genistein (template 1)		
gen.117	-10.970	Genistein (template 1)		
gen.16	-10.561	Genistein (template 1)		

Name	GScore XP (kcal/mol)	Scaffold	R ₁	R ₂
gen.294	-10.240	Genistein (template 1)		
gen.303	-10.155	Genistein (template 1)		
gen.282	-9.888	Genistein (template 1)		
tec.16	-9.577	Tectorigenin (template 1)		
gen.168	-9.530	Genistein (template 1)		
tec.168	-9.447	Tectorigenin (template 1)		
gen.300	-9.359	Genistein (template 1)		
gen.73	-8.955	Genistein (template 1)		
tec.168	-8.933	Tectorigenin (template 1)		
gen.111	-8.779	Genistein (template 1)		
gen.300	-8.754	Genistein (template 1)		
gen.16	-8.681	Genistein (template 1)		
gen.282	-8.553	Genistein (template 1)		

Most poses for both PDEs interact with Phe residue through a pi-stacking with the two condensed aromatic rings, whereas the H-bond with Gln was more often with unsubstituted OH or small functional groups (FGs) as -NH₂ or -SO₂NH₂ in the same part of the molecule. Interestingly the poses selected for PDE5 share the pomiferin scaffold. It is worth pointing out that in a previous work,⁴²⁵ it was demonstrated the high affinity of pomiferin with PDE5A and these findings are in agreement, showing a preference for PDE5A for this template. The position of the molecules is similar for all the compounds, in fact both the interactions with the residues are formed between the unsubstituted -OH next to the carbonyl and Gln and between the dihydro-pyranone and Phe. In general, the substituents that appear to be more effective are polar and include protonable or deprotonable sites, especially for the part with the single aromatic ring which is mainly positioned in the pocket consisting of polar/charged residues and metals. The most

recurrent FG are $-NH_2$, $-CH_2COOH$, $-SO_2NH_2$ or $-OH$. The poses of one ligand for each PDE is depicted in Figure 3.2.4.

Lastly, with the intent of studying with greater accuracy the binding of these compounds with the considered macromolecules, MD calculations were conducted on one pose for each PDE. The poses were selected considering some peculiarities of the computed interactions such as the presence of strong or multiple H-bonds and pi stackings or of ionic bonds with the metals. The computed RMSD values showed that the poses are overall stable during the simulation time frame and that the H-bond with Gln and the pi-stacking with Phe were maintained throughout the simulation. (Figure 3.2.5)

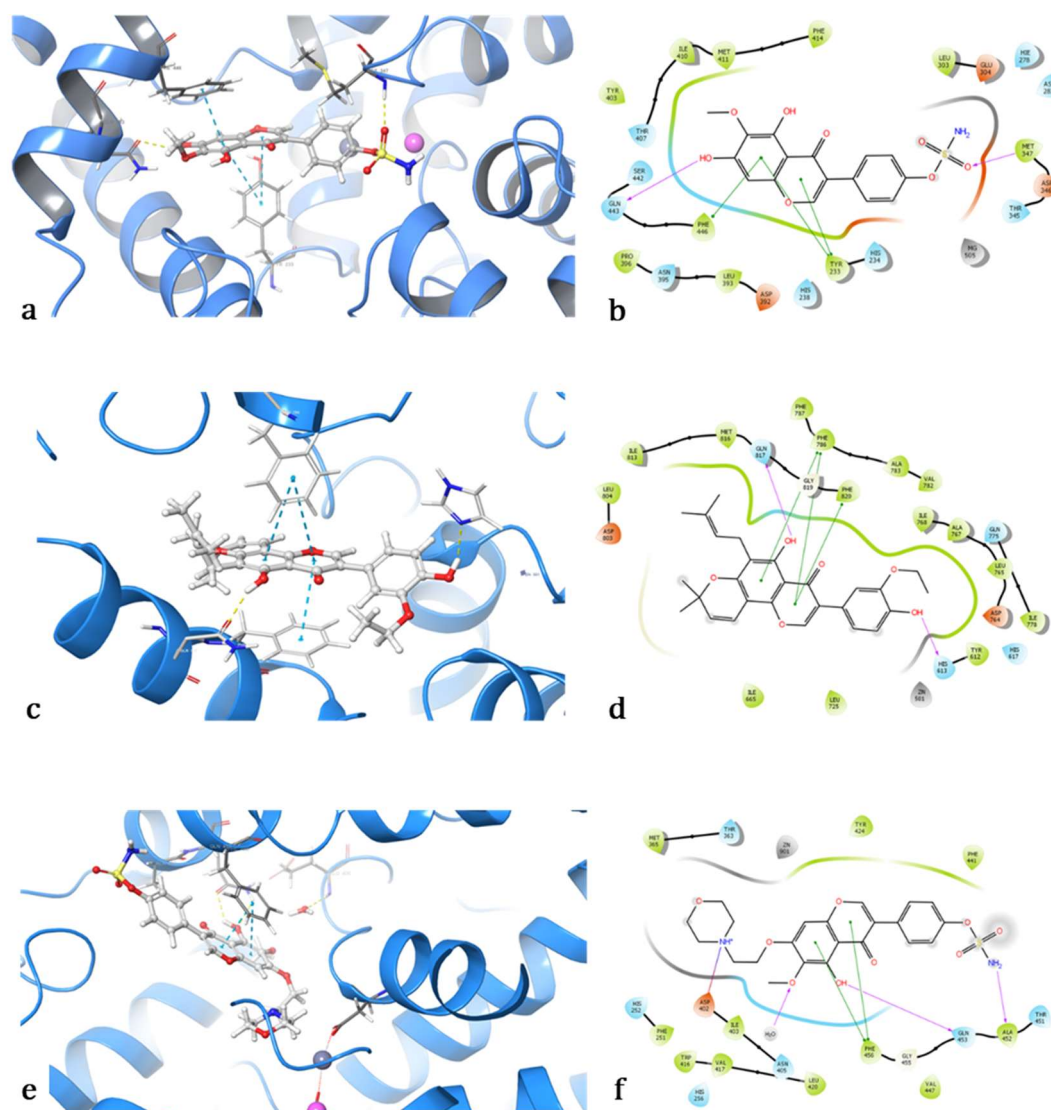


Figure 3.2.4. Docked pose of (a) compound tec.16 in PDE4B receptor (PDB ID: 3GWT) with (b) the two-dimensional interactions map, (c) compound pom.39 in PDE5A receptor (PDB ID: 2H42) with (d) the two-dimensional interactions map, (e) compound tec.168 in PDE9A receptor (PDB ID: 3JSW) with (f) the two-dimensional interactions map, showing the interactions of the complex, in particular pi-stacking with Phe and H-bond with Gln three-dimensional view of the ligand-receptor interactions.

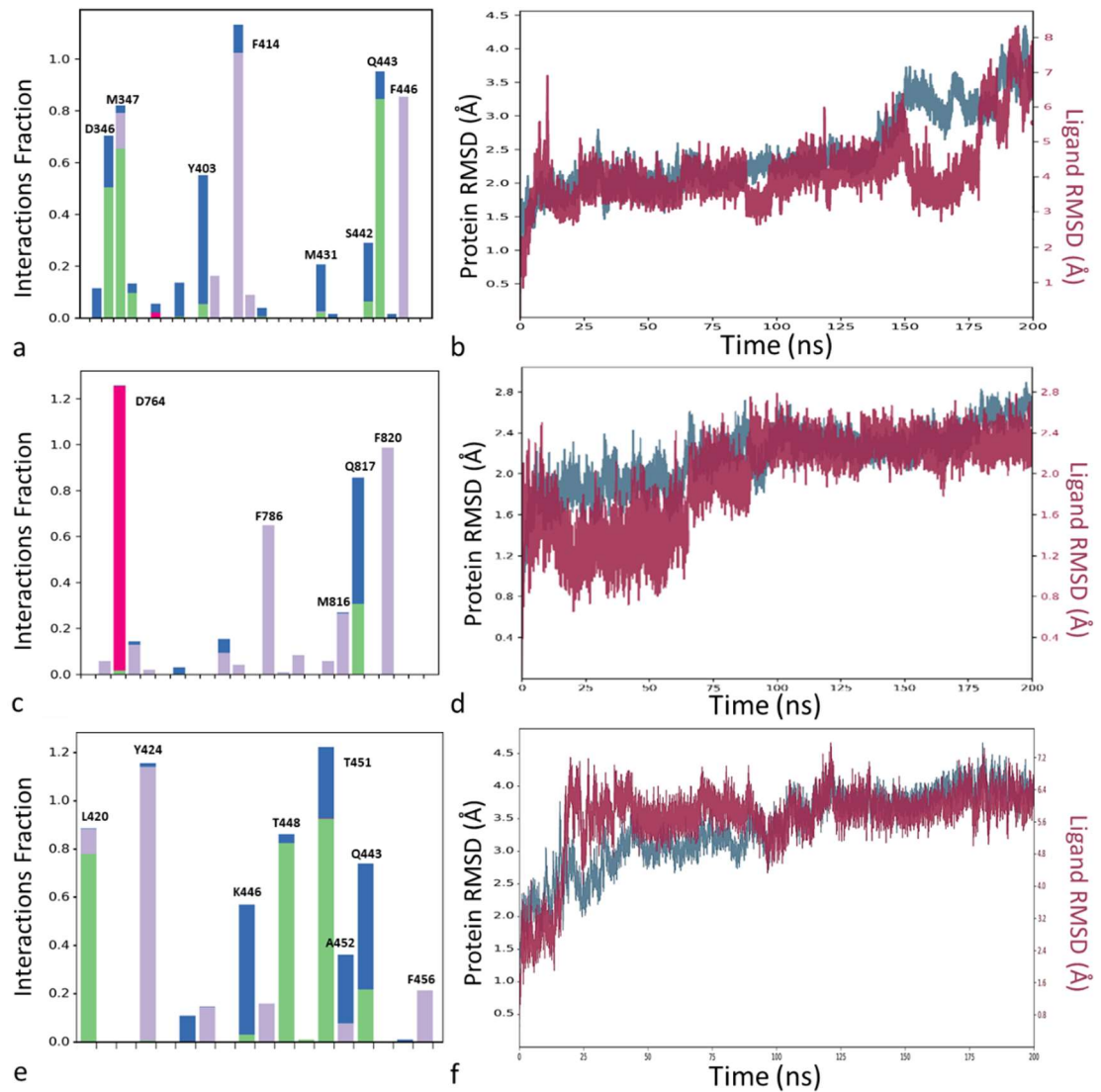


Figure 3.2.5. Interactions Fraction histograms showing the average number of interactions over the simulation time: • hydrophobic interaction, • ionic interaction, • water bridged interaction, • H-bond. Values higher than 1 in the case of multiple contacts between the ligand atoms and the protein (a: PDE4B/tec.16 – c: PDE5A/pom.39 – e: PDE9A/tec.168). RMSD trajectory plots showing the stability of the system over the simulation time: • ligand (aligned with the protein), • protein C-alphas (b: PDE4B/tec.16 – d: PDE5A/pom.39 – f: PDE9A/tec.168).

3.2.3 Virtual screening-accelerated discovery of a phosphodiesterase 9 inhibitor with neuroprotective effects in the kainate toxicity *in vitro* model

Results and discussion

In this work, we screened an internal database comprehending more than 1600 molecules that were synthesized by our research group over the years. Firstly, the druglikeness and the pharmacokinetic properties were computed and only the molecules showing maximum one violation to the “Lipinski’s rule of five” were admitted to the following step.⁵⁶⁹ In order to point the attention on natural and nature-inspired compounds that have been widely demonstrating to act against neurodegeneration, we filtered the remaining ligands using a structural query, namely a chromone scaffold with free substitution of the oxygen in position 1 using a LBDD approach, thus including in the query isoflavones, anthraquinones and naphthoquinones, that are known scaffold of many chemical compounds derived from plant extraction and have been associated with a wide range of pharmacological properties, such as anti-inflammatory, anti-cancer, diuretic, laxative, antidepressant and anti-oxidant activity and more recently for the treatment of Alzheimer disease (AD).⁵⁷⁰ The obtained molecules were docked with Schrodinger suite on a PDE9 3D structure and the 20 most promising compounds were submitted to an experimental solubility test in PBS buffer in order to build a sort of bridge between the *in silico* and the *in vitro* experiments. Considering all the results, compound DB987 was selected for further in-depth investigation. DB987 is depicted in Figure 3.2.6, it consists of the isoflavone pomiferin semi synthetically modified with the addition of mesylated functional groups.

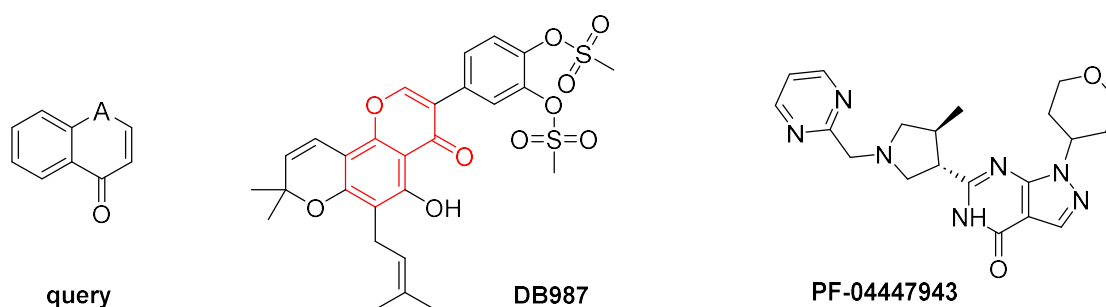


Figure 3.2.6. Chemical structure of the chromone scaffold used for the virtual screening procedure, of DB987 the compound highlighted by the screening, in which the red structure represents the part of the molecule matching query and of the reference compound, PDE9 inhibitor PF-04447943.

As previously mentioned, it is reported in literature that most PDE9 inhibitors share the interactions with two important residues, namely Gln453 and Phe456 through a H-

bond and a pi-stacking respectively. Moreover, since the PDE9 is a metalloprotein, its catalytic activity depends on the interactions with its two divalent cations, *i.e.*, Zn^{2+} and Mg^{2+} the first is known to coordinate with residues as Asp402, Asp293, His292, His256 and two water molecules, while the latter coordinates five water molecules and Asp293.^{555,571,572}

For the computational study, the structure used was in complex with PF-04447943 (PDB ID: 4E90)⁵⁷³ which is depicted in Figure 3.2.7. It is worth mentioning that the PDB file used for the docking has been changed with respect to the previous work because that particular inhibitor was supposed to be used as control in the *in vitro* study. Indeed PF-04447943 showed the important interactions with Phe456 and Gln453 previously mentioned. As validation of the method, the ligand was re-docked on the binding site and the RMSD value between the crystalized and the docked poses were measured and found 1.18 Å, below the maximum limit of 2.50 Å.⁵⁶⁸ The docked pose of DB987 showed as well the mentioned interactions, together with additional H-bonds with Tyr424 and Asn300, and a metal bond with Zn^{2+} .

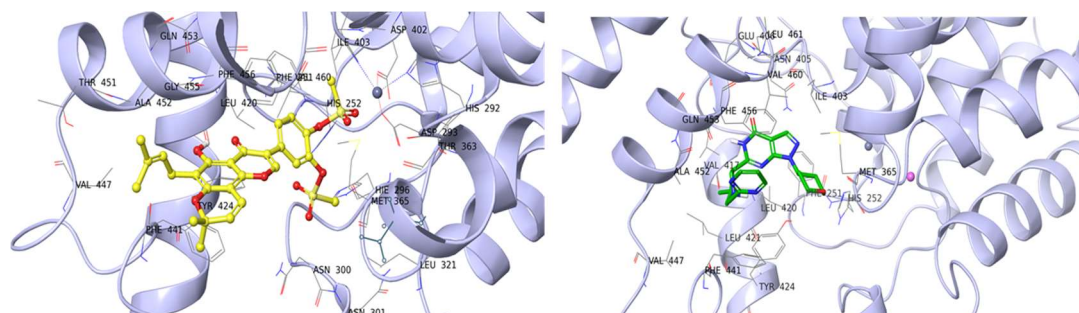


Figure 3.2.7. Docked pose of • DB987 and • PF-04447943 in complex with PDE9.

Eventually, a MD simulation was run on both the docked poses of the known inhibitor and our hit compound to assess the stability of the predicted interactions. Indeed, the mentioned interactions were maintained throughout the simulation time frame. The RMSD values calculated for the PF-04447943 (Figure 3.2.8a) was demonstrated to be stable, except for some fluctuation. The RMSF analysis showed as expected that the major fluctuations (highest RMSF values) are due to the flexibility between the pyrimidine ring and tetrahydropyran moiety and the rest of the molecule (Figure 3.2.8b) and it was further demonstrated by the absence of those fluctuations in the RMSD trajectory calculated on the pyrimidine ring atoms (Figure 3.2.8c).

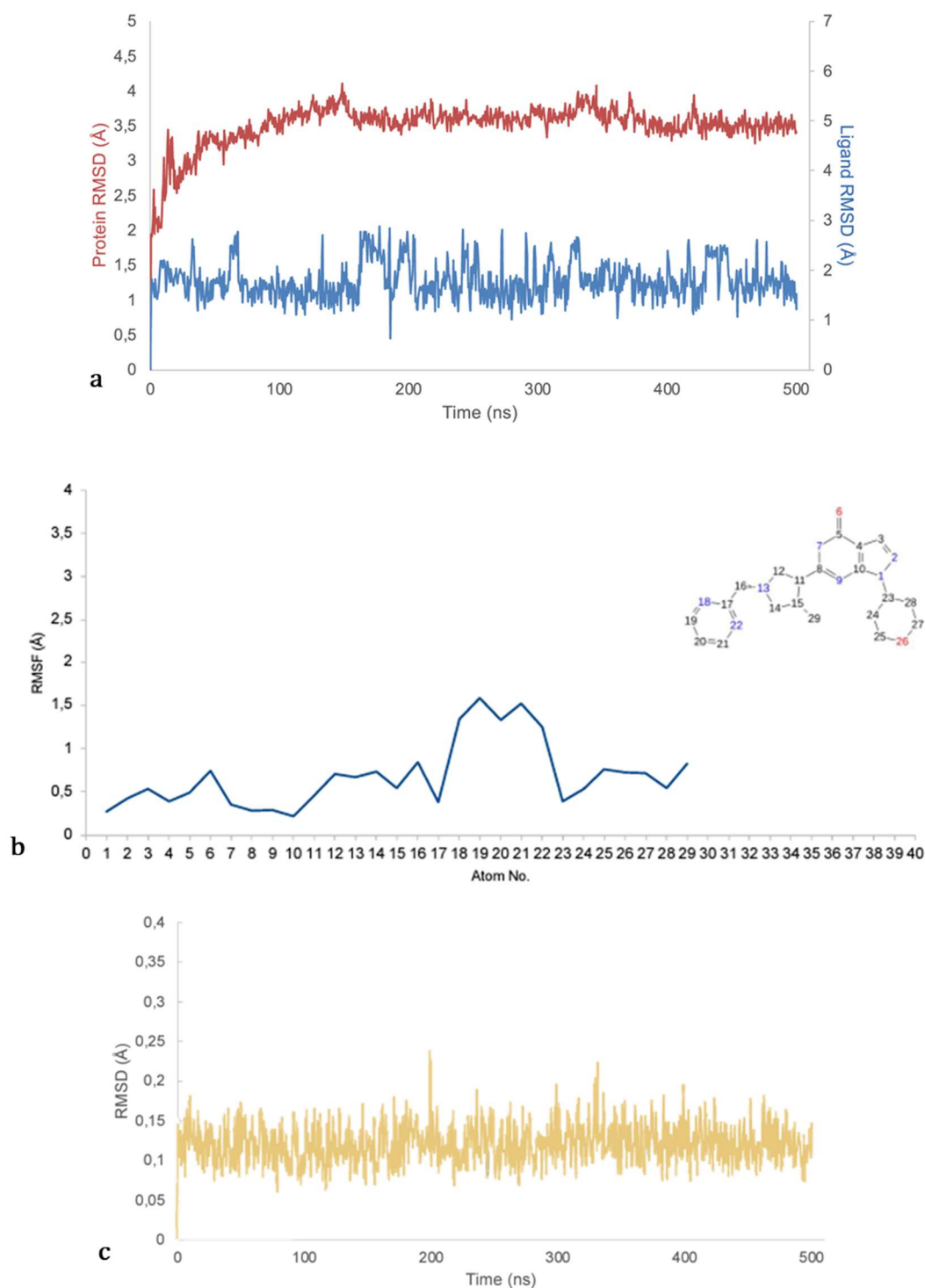


Figure 3.2.8. (a) Trajectory retrieved from MD simulations showing the RMSD of the PDE9/ PF-04447943 systems over the simulation time: • heavy atoms of the ligand (aligned with the protein), • protein C-alphas. (b) RMSF plots of PF-04447943. (c) RMSD plot of the scaffold of PF-04447943 (namely atoms 1-10).

The complex PDE9/DB987 achieved the stability after 100 ns and the RMSD trajectory (Figure 3.2.9a) displayed some fluctuations as well due to the flexibility of the 5-member alkyl chain, again confirmed with RMSF study (Figure 3.2.9b).

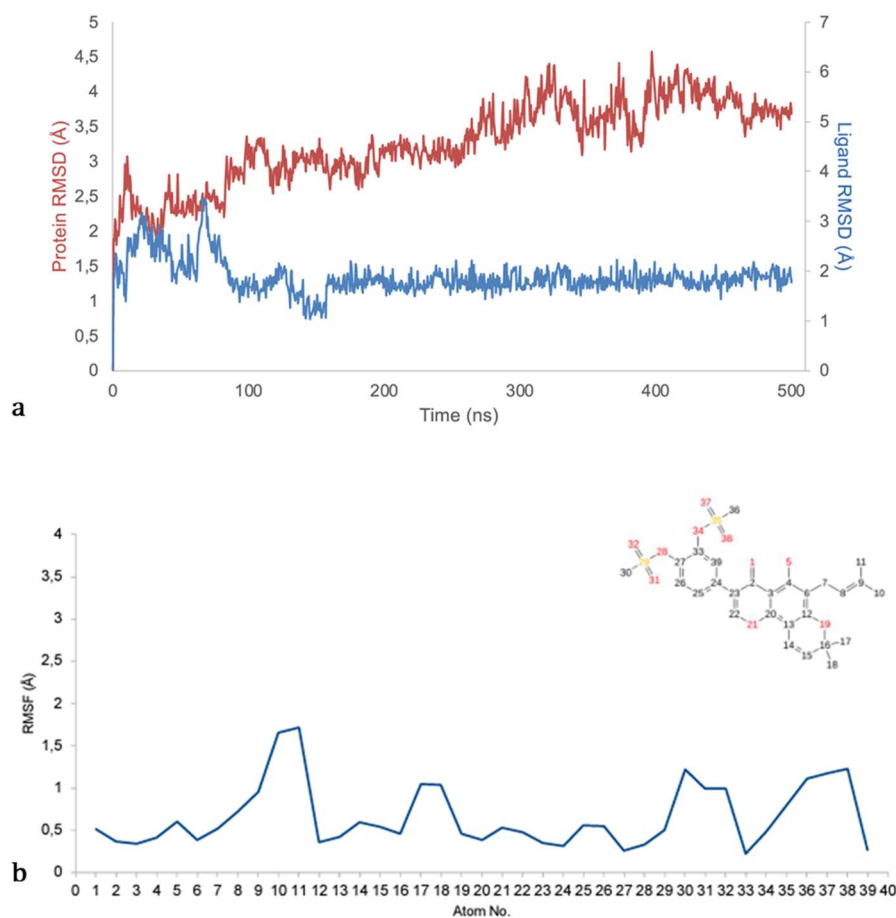


Figure 3.2.9. (a) Trajectories retrieved from MD simulations showing the RMSD of the PDE9/DB987 systems over the simulation time: • heavy atoms of the ligand (aligned with the protein), • protein C-alphas. (b) RMSF plots of DB987.

Thanks to the collaboration of the research group of Prof. Pellegrini-Giampietro from the University of Florence, first PDE9 expression was demonstrated in hippocampal slices through real time PCR and western blot analysis and its concentrations was found increased in damaged hippocampal slices treated with kainate, a potent neurotoxin,⁵⁶⁵ with immunofluorescent staining and confocal microscopy. Importantly triple immunostaining showed that, as in previous study,⁵⁷⁴ PDE9 is localized only on neuron cells since no colocalization of PDE9 expression with astrocytes and microglia in CA3 slices treated with kainate. For the details of the experiment the reader should address to the publication.^{565–567}

Secondly, the activity of DB987 on the inhibition of PDE9 was also established using an *in vitro* experiment, in accordance with previously reported procedure.⁵⁷⁵ The measured IC_{50} was 5.35 nM which is similar to the one obtained by PF-04447943 ($IC_{50} = 14$ nM) with the same method.

Lastly, the ability of the ligand to reduce neuronal death was evaluated. In particular, hippocampal slices treated with kainate for 24 hours showed significant damage in the CA3 area,^{565,567} the treatment with PDE9 inhibitor PF-0447943 was able to significantly reduce neuronal death during the kainate exposure in a dose dependent manner in CA3 and we observed similar neuroprotective effects when the slices were incubated with DB987 in accordance with the IC_{50} value. Moreover, the ligands themselves did not show to induce damage in the CA3 area of the slices. (Figure 3.2.11)

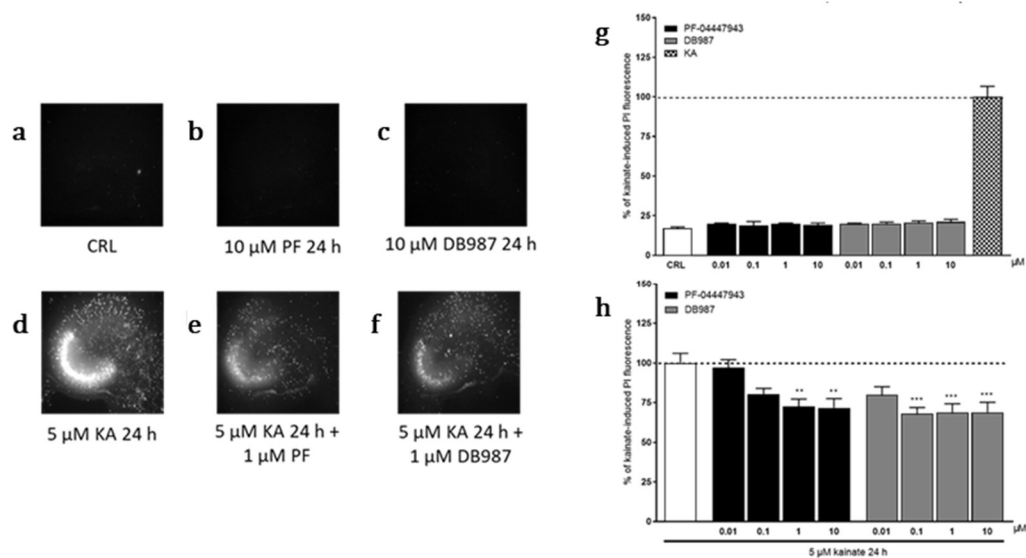


Figure 3.2.11. Qualitative and quantitative analyses of the effects of PDE9 inhibitors in rat organotypic hippocampal slices under normal conditions or exposed to kainate (called KA in the Figure). (a) Hippocampal slice under normal conditions (background PI fluorescence), (b) slice exposed to 10 μ M PF for 24 h, (c) slice exposed to 10 μ M DB987 for 24 h, (d) slice exposed to 5 μ M KA for 24 h displaying intense PI labelling in the CA3 subregion, and (e, f) CA3 damage induced by KA was attenuated by the presence of 1 μ M PF and DB987. (g) PDE9 inhibitors alone did not induce side effects. (h) PDE9 inhibitors significantly attenuated CA3 damage in a dose-dependent manner. Bars represent the mean \pm SEM of at least five experiments run in quadruplicate. ** p < 0.01 and *** p < 0.001 vs KA (one-way ANOVA plus Dunnett's test).

3.2.4 Conclusions

In the first work, described in Chapter 3.3.1, we showed the potential of using computational tools for the design of new ligands by means of combinatorial chemistry, molecular docking and MD simulations. Importantly, the docking method was carefully validated prior to the analysis, thus reassuring the reliability of the proposed protocol. Specifically, the method was validated in three different and independent steps, namely the re-docking of the co-crystalised ligand retrieved from the PDB structure on the specific target (which is the most fast and foreword technique) which showed that docking is able to predict a pose similar to the experimental one (RMSD < 2.50 Å), an enrichment study that confirmed that selecting the top ranked 5% of compounds was able to give high efficiency in finding the active compounds when screening a large library, an a cross-docking which demonstrated that XP docking was able to identify better the right poses and interactions between the target and the small molecules. Therefore, more than 1500 compounds where first screened with the low computational demanding technique SP, then the 5 % best performing compounds in terms of docking score were subjected to a second more refined XP technique, and the resulting poses were carefully semi-automatically analysed. The compounds showing the important interactions with Phe and Gln were selected for a preliminary SAR study that highlighted the important features to consider when designing new potential inhibitors.

Some important considerations are worth mentioning.

Isoflavones demonstrated to be a valuable scaffold for the development of ligands targeting PDEs and importantly even though the native derivatives were included in the study, none of them were retrieved at the end of the protocol, demonstrating indeed those semi-synthetic modifications may increase the affinity with the targets. The condensed aromatic ring was shown to give strong pi-stacking interaction with F and other hydrophobic interaction which are able to “lock” the molecule in the binding pocket, the hydroxyl group instead is able to give a H-bond with Gln or interactions with other polar groups.

Moreover, it was demonstrated that the addition of polar functional groups could favour and increase the ability of forming the afore mentioned H-bonds and metal coordinations, providing higher stability.

Finally, the overall stability of the complexes formed was demonstrated by MD simulations.

In the second work, we made use of an internal database and we used an extensive semi-computational protocol to determine the most effective compound for targeting and inhibiting PDE9. Firstly, the molecules were subjected to ADME study to determine their druglikeness and pharmacokinetic properties, allowing no more than one violation to the famous “Lipinski’s rule of five”, this is an important step that can avoid the extra cost of experimentally testing compounds that could not progress in the study due to their chemico-physical properties. Then, after validating the method re-docking the co-crystallised ligand with satisfactory RMSD, the compounds that respect a structural query that included flavonoids and other nature-inspired compounds such as anthraquinones and naphthoquinones, were subjected to a molecular docking study. Manually investigating the poses of the best compounds in terms of docking score, we progress to a further solubility test only the compounds that showed the known interaction with Phe456 and Gln453. Merging all the information, DB987 was selected for the MD simulation and the *in vitro* testing.

The MD simulation demonstrated that the predicted complex is stable is the analysed timeframe whereas the *in vitro* experiments showed high inhibitory activity with an IC₅₀ value in the nM range, comparable with the known inhibitor PF-04447943. Additionally, the compound is able to ameliorate neuronal damage in kainate treated hippocampal slices.

The outcome of these studies showed the importance of the use of *in silico* techniques for the identification of potential new hits both with *de novo* design and when screening commercial or internal databases.

In particular, the molecules selected in these works may pave the way for the development of new potential inhibitors, semi-synthetically obtained to be used against neurodegenerative diseases.

3.2.5 Materials and methods

Combinatorial library generation, molecular docking and molecular dynamics simulations for enhancing the isoflavone scaffold in phosphodiesterase inhibition⁵⁶³

Compound library generation

A set of 19 different molecular fragments and two receptor templates were created respectively in SMILES and in MDL Molfile formats with ChemSketch 2019.⁵⁷⁶ These were used as input for the software ChemT,⁵⁷⁷ that generated a set of molecules formed by all the combinatorial possibilities in SMILES format which was converted to a unique SDF file used in the following molecular docking stages.

Receptor preparation

All the receptor files were retrieved from RCSB Protein Data Bank (PDB, www.rcsb.org), for PDE4 the 3GWT crystal structure,⁵⁷⁸ representing the catalytic domain of human phosphodiesterase PDE4B was used, for PDE5 the 2H42 crystal structure,⁵⁷⁹ representing the human phosphodiesterase PDE5A, for PDE9, the 3JSW crystal structure,⁵⁸⁰ which represents the human phosphodiesterase PDE9A. The receptors were prepared with the Protein Preparation Wizard included in the Schrodinger suite,⁵⁸¹ using default settings, *i.e.* adding hydrogens, assigning disulphide bonds, removing waters further than 5 Å from the ligand, adjusting charges, capping termini, adding missing side-chains using Prime, optimizing hydrogen bond clusters and a final minimization under the OPLS3e force field.⁴⁷⁹ The receptors were accurately refined for what concerns the water molecules presence, first using the automatic removal of water molecules having less than three H-bond to non-water molecules then by superimposition of multiple structures of the same protein (PDE4B: 3GWT⁵⁷⁸-4KP6⁵⁸²-3D3P⁵⁸³-3W5E⁵⁸⁴-5OHJ⁵⁸⁵-1XLZ,⁵⁸⁶ PDE5A: 2H42⁵⁷⁹-3TGE⁵⁸⁷-6ACB⁵⁸⁸-1XOZ⁵⁸⁶-5ZZ2⁵⁸⁸-4MD6,⁴¹⁴ PDE9A: 3JSW⁵⁸⁰-4Y86⁵⁸⁹-6A3N⁵⁹⁰-4GH6⁵⁹¹-3QI4⁵⁹²-2HD1⁵⁷¹) followed by visual inspection to assess the importance of the eventually retained water molecules.

Ligands preparation

Every ligand was prepared for the docking with LigPrep application present in the Schrodinger suite under the OPLS3e force field,⁵⁹³ with Epik ionizer all the ionization

states were generated in the 7 ± 2 pH range also including the possible metal states important for the docking to metalloproteins.

Molecular docking

The docking protocol consisted in the rigid-receptor/flexible ligands docking, which was performed with Glide,⁴²⁹ under the OPLS3e force field, using the SP and XP precision with default settings, *i.e.*, 0.80 scaling factor for the ligand atoms *Van der Waals* radius and 0.15 for partial charge cut-off, generation of 1 pose per ligand. All the receptors and ligands were previously prepared as described above. The receptor grid was prepared with Receptor Grid Generation present in Schrodinger with default settings, a cubic grid was generated setting the centre in the centroid of the ligands present in the crystal structures. The dimensions of the grid depended on the dimensions of the ligands present in the crystals, spacing between 22-24 Å.

Validation method

For the three PDE receptors considered in this study (PDE4B: 3GWT, PDE5A: 2H42, PDE9A: 3JSW), the co-crystallized ligands were docked at SP precision with Glide in their own receptor using the same procedure described above for ligand and protein preparation, then the so obtained ligand poses were used for the calculations of heavy atoms RMSD respect to their native structures with Discovery Studio 2020.⁵⁹⁴

The enrichment studies were performed for each PDE seeding a so-called decoy set of inactive or presumed inactive compounds with ligands with known activity against the target in question. Three decoy sets (one for each PDE) were generated by Database of Useful Decoys: Enhanced (DUD-E) (<http://dude.docking.org/>),^{595,596} selecting for each PDE a group of 20 active ligands with an IC_{50} lower than 20 μ M from the Binding Database (<http://www.bindingdb.org/bind/index.jsp>). SP docking was performed using the same protocols described above for the preparation of the receptor, of the ligands and the docking stages.

After ranking the decoy set by docking score, EF and Eff values were calculated and ROC curves plotted using the Enrichment Calculator tool included in Schrodinger Suite. EF can be described as the concentration of the considered active ligands among the top-scoring docking hits compared to their concentration throughout the entire database and is calculated as follows:

$$EF = \left\{ \frac{N_{tot}}{N_{sel}} \right\} \left\{ \frac{Hits_{sel}}{Hits_{tot}} \right\}$$

Where N_{sel} is the number of molecules included in the top percentage considered, N_{tot} is the total number of compounds screened in the database, $Hits_{sel}$ is the number of active ligands at a specific percentage and $Hits_{tot}$ is the total number of active ligands for the target.⁵⁹⁷ Eff corresponds to the efficiency in distinguishing active ligands from decoys on an absolute scale ranging from 1 (ideal, all actives are ranked on top of decoys) to -1 (all decoys come before any actives in the ranking). ROC curves plot the sensitivity (Se) of a given docking/scoring combination against specificity (Sp), and Area Under the Curve (AUC or AU-ROC) can be calculated for comparison. The AU-ROC curve represents the probability for the active compounds to be listed in the highest part of the ranking, on top of decoy compounds. A theoretically perfect AUC value is 1.0.⁵⁹⁸

The cross-docking experiment was conducted as follows:³⁰ for each PDE five different crystal structures were considered (PDE4B: 3GWT⁵⁷⁸-4KP6⁵⁸²-3D3P⁵⁸³-5OHJ⁵⁸⁵-1XLZ,⁵⁸⁶ PDE5A: 2H42⁵⁷⁹-3TGE⁵⁸⁷-6ACB⁵⁸⁸-5ZZ2⁵⁸⁸-4MD6,⁴¹⁴ PDE9A: 3JSW⁵⁸⁰-4Y86⁵⁸⁹-6A3N⁵⁹⁰-3QI4⁵⁹²-2HD1⁵⁷¹), then each ligand belonging to each crystal structure was docked to the other 4 and heavy atoms RMSD were calculated with Discovery Studio 2020, thus verifying that the docked poses resemble the crystallized ones. The same protocols defined above were utilized for the preparation of the receptor, of the ligands and for the docking.

Molecular dynamics

The MD calculations were performed with the GPU-accelerated Desmond software,⁴⁸⁶ the protein-ligand complexes retrieved from the XP molecular docking stage were solvated using TIP3P water model, Na⁺ ions were added to neutralize the system and a concentration of 0.15 M of NaCl was simulated, an ion placement exclusion zone was created at 20 Å distance from the ligands. Prior the MD jobs the systems were relaxed with Maestro's default relaxation protocol that includes two stages of minimization (restrained and unrestrained) followed by four stages of MD runs with gradually diminishing restraints.^{599,600} All the MD runs were conducted for 200 ns in constant pressure and temperature (NPT ensemble) with Desmond default force field (OPLS_2005), after that the trajectories were analysed with Desmond computing the ligand (aligned on protein) and protein (C-alphas) RMSD trajectory plots, and the interactions pattern (Interactions Fraction).

Virtual screening-accelerated discovery of a phosphodiesterase 9 inhibitor with neuroprotective effects in the kainate toxicity *in vitro* model⁵⁶⁴

Database Search

The considered database is constituted of 1632 compounds synthesized in-house. The database file, in SDF format, was directly loaded in Schrödinger Maestro version 13.4.134, MMshare Version 6.0.134 Release 2022-4, Platform Linux-X86_64: Maestro (Schrödinger, LLC, New York, NY, USA 2021). The structures were prepared with the LigPrep tool under the OPLS4 force field using an Epik ionizer;^{601–603} all the ionization states were generated in the 7 ± 2 pH range including the possible metal states leading to a total of 3150 molecules, which were analysed with the QikProp tool to predict the pharmacokinetic properties. The molecules having more than one violation to the “Lipinski’s rule of five” were discarded, reducing the number of possible hits to 1762. These molecules were then filtered based on their chemical structure, and to perform this step, we loaded the SDF file retrieved from Schrödinger in ChemFinder Version 20.1.1 (ChemOffice, PerkinElmer Informatics, Inc. Waltham, MA, USA, 2020), and we used the chromone-based query (Figure 3.2.6) to screen the compounds, which were reduced to 441. Then, the molecules were loaded into Schrödinger to perform the docking study as described in the following paragraphs. The complexes resulting from docking experiments were visually inspected, and the 20 molecules with the best poses, also considering the docking score, were selected. The hit compounds that were physically retrieved from the database following the preliminary virtual screening were subjected to a solubility test. A 10 mM stock solution of the compounds was prepared in DMSO, and they were then diluted to 100 μ M solutions in H₂O. Then, three more 1:10 dilution steps were performed in phosphate-buffered saline (PBS, pH 7.4) up to 1 nM to assess adequate solubility of the compounds in the concentration range considered in *in vitro* test.

Molecular Docking

Human PDE9 3D structure was obtained from the RCSB Protein Data Bank (PDB, <http://www.rcsb.org>, access date 30 January 2023). PDE9 in complex with inhibitor PF-04447943, entry 4E90, was selected (resolution 2.50 Å). The preparation step was performed through the Protein Preparation Workflow tool. Default parameters were selected, for example, adding hydrogens, assigning disulphide bonds, adding missing side chains using Prime, removing waters, adjusting charges, and minimizing the convergence of heavy atoms to RMSD under the OPLS4 force field.⁶⁰² Ligands were prepared for

docking with the LigPrep tool under the OPLS4 force field. Possible ionization and metal binding states at target pH 7.0 ± 2.0 were generated with an Epik ionizer.⁶⁰³ The binding poses of PF-04447943 and DB987 in complex with PDE9 were obtained by molecular docking calculations using Glide.⁶⁰⁴ Ligands and receptors were prepared as described in the previous paragraphs. The docking protocol consisted of rigid receptor and flexible ligand sampling, which was performed under the OPLS4 force field⁶⁰¹ using standard precision (SP) with default settings like 0.80 scaling factor for the ligand atom van der Waals radius, 0.15 partial range cut-off, adding Epik state penalties to the docking score, generation of one pose per ligand, and performing post docking minimization for generated poses. The receptor grid was prepared with the Receptor Grid Generation tool included in Schrödinger, with default settings.³⁵ After the identification of the best poses, the RMSD value was used to validate the adopted method with the rmsd.py script, and a score under 2.50 Å was admitted to consider the analysis sufficiently accurate. In the case of this study, a RMSD value of 1.18 Å was obtained. The obtained poses were evaluated according to the GlideScore (GScore) scoring function, which is expressed in $-kcal/mol$, and through the inspection of the interactions established between ligands and PDE9. For interaction analysis, only residues within 5 Å in the binding pocket were considered.

Molecular Dynamics

MD simulations were performed using the GPU-accelerated Desmond tool of Schrödinger LLC (Schrödinger Release 2023-1: Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, USA, 2021. Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY, USA 2021).⁶⁰⁵ The best protein–ligand complexes retrieved from the SP molecular docking step were used to prepare the solvated system to perform MD. By using System Builder,⁶⁰⁵ all the three complexes were solvated using the explicit TIP3P water model. Periodic boundary conditions were set with an orthorhombic box shape of $10 \times 10 \times 10$ Å. During ion placement, Na^+ ions were used to neutralize the system, a concentration of 0.15 M of NaCl was set, and an exclusion zone with 20 Å from the ligand was added. A run of 500 ns was performed for each complex under the OPLS4 force field with the temperature set at 300 K and pressure at 1.0 bar. Further, the Martyna–Tuckerman–Klein chain coupling scheme with an isotropic coupling constant of 2.0 ps for the pressure control and the Nosé–Hoover chain coupling scheme for the temperature control (NPT ensemble) were used. The cut-off radius in the Coulomb interactions was 9.0 Å. A RESPA integrator was set with a time

step of 2.0, 2.0, and 6.0 fs, respectively, for bonded interactions and near and far interactions. Trajectories were saved at 500 ps intervals for analysis. To analyse interactions between ligands and PDE9, the Simulation Interaction Diagram tool and RMSD/RMSF evaluation were used. When ligand–protein contacts were evaluated, a contact strength threshold value of 30% was considered for all of the analyses. Chemicals and Biological Materials. Tissue culture reagents were obtained from Gibco-BRL (San Giuliano Milanese, MI, Italy) and Sigma (St. Louis, MO, USA). PF-0447943 was purchased from Sigma (St. Louis, MO, USA). The synthesis of DB987 was previously reported by our group.⁴²⁷

4 Final conclusions and future perspectives

The work presented in this thesis demonstrates various applications of the techniques outlined in the first section, used to develop several projects exploring the interaction of small molecules with different targets.

Specifically, the combination of plant extraction, simple semi-synthetic approaches, or organic synthesis of new compounds was merged with ESI-MS, NMR, CD, and computational techniques. This integration facilitated not only the comprehensive characterisation of the compounds but also the investigation of these ligands' interactions with G4. G4, known to play a role in various cellular processes, and its stabilization have been linked predominantly to potential anti-cancer activity, as well as anti-inflammatory, anti-viral, and neuroprotective actions.

The main project (Section 2) demonstrated that utilising a low sample-consuming high-throughput screening technique like ESI-MS, alongside a method offering atomic resolution information about the target:ligand complex, albeit requiring higher concentrations such as NMR, when combined with traditional CD spectroscopy and computational techniques, forms a valid protocol for developing potentially bioactive molecules.

The primary goal of the first side project (Chapter 3.1) was to create seleno-containing compounds through an optimised multi-step reaction scheme. These compounds were intended to act as multi-target agents to mitigate oxidative stress and combat neurodegeneration. Techniques such as NMR, ESI-MS, and DFT computational analysis were employed to track specific reactions mimicking physiological conditions. Furthermore, *in vivo* studies conducted on mice using these derivatives suggested an inhibitory effect on serotonin reuptake, demonstrating their potential as antidepressants.

The second side project (Chapter 3.2) was designed to emphasise the importance of employing computational techniques in the drug discovery process. The amalgamation of combinatorial chemistry or database searches with molecular docking and molecular dynamics enabled the identification of new inhibitors associated with neurodegenerative disorders. The results were validated through collaboration with other research groups, thanks to *in vitro* experiments. Importantly, this approach facilitated the assay testing of only one derivative instead of an entire database, thereby reducing time and costs.

The integration of all available techniques is crucial for developing new agents with potential pharmacological effects. Research efforts should continue to focus on establishing new protocols derived from this integration to streamline costs and time in the drug discovery process.

5 Appendix

5.1 List of abbreviations

Abbreviation	Definition
μ	magnetic moment
5'-AMP	Adenosine MonoPhosphate
5'-GMP	Guanosine MonoPhosphate
A	Adenine
AA	AminoAcid
AC	Alternating Current
AD	Alzheimer Disease
ADMET	Adsorption, Distribution, Metabolism, Excretion, Toxicity
ADP	Adenosine DiPhosphate
ALS	Amyotrophic Lateral Sclerosis
ALT	Alternative Lengthening of Telomeres
AQ	AnthraQuinone
Arg	Arginine
ARIA	Ambiguous Restraints for Iterative Assignment
Asp	Aspartate
AUC or AU-ROC	Area Under Receiver Operating Curve
B_0	external magnetic field
BA	Binding Affinity
BRACO-19	N,N'-(9-((4-(dimethylamino)phenyl)amino)acridine-3,6-diyl)bis(3-(pyrrolidin-1-yl)propan-amide)
BRCA	BReast CAncer gene
C	Cytosine
CADD	Computer-Aided Drug Design
cAMP	cyclic Adenosine MonoPhosphate
CAT	Catalase
CD	Circular Dichroism
CFM	Carbon Fiber Microelectrode
cGMP	cyclic Guanosine MonoPhosphate
CI	Chemical Ionization

CID	Collision-Induced Dissociation
CMR	Mean Residue Concentration
CNS	Central Nervous System
COSY	COrrrelation SpectroscopY
CPL	Circularly Polarised Light
Cryo-EM	Cryogenic Electron Microscopy
CX-3543	Quarfloxin
D/I	Desorption/Ionization
DC	Direct Current
DCM	DiChloroMethane
DFT	Density Functional Theory
DMSO	DiMethylSulphOxide
DSB	Double-Strand Breaks
DsDNA	Canonical double-strand DNA
DSS	2,2-dimethyl-2-silapentane-5-sulfonate
DUD-E	Database of Useful Decoy enhanced
ED	Erectile Dysfunction
EF	Enrichment Factor
Eff	Efficiency
EI	Electron Ionization
EMA	European Medicines Agency
EPR	Electron Paramagnetic Resonance
ESI	ElectroSpray Ionization
FAB	Fast Atom Bombardment
FDA	Food & Drug Administration
ff	force field
FG	Functional Group
FID	Free Induction Decay
FRET	Forster Resonance Energy Transfer
FSCV	Fast Scan Cyclic Voltammetry
FTICR	Fourier Transform-Ion Cyclotron Resonance
G	Guanine
G4	G-quadruplex
G4P	G4 genomic Potential

G4seq	G4 sequencing
Gln	Glutamine
Gly	Glycine
GMP	5'-Guanosine Monophosphate
GPU	Graphic Processor Unit
GPx	Glutathione Peroxidase
GR	Glutathione Reductase
GScore	Glide score
GSH	Glutathione
GSSG	oxidised glutathione
HAT	Hydrogen Atom Transfer
H-bond	Hydrogen bond
HCV	Hepatitis C Virus
His	Histidine
HIV	Human Immunodeficiency Virus
HMBC	Heteronuclear Multiple Bond Correlation
HPLC	High Performance Liquid Chromatography
HPSF®	High Purity Salt Free®
HPV	Human PapillomaVirus
HR	Homologous Recombination
HSQC	Heteronuclear Single Quantum Coherence
Ile	Isoleucine
IQb	indolo[3,2-b]quinolines
IQc	indolo[3,2-c]quinolines
IRES	Internal Ribosome Entry Site
J	coupling constant
K _A	association equilibrium constant
K _D	dissociation equilibrium constant
LBDD	Ligand-Based Drug Design
L-CPL	Left Circularly Polarised Light
Leu	Leucine
lncRNA	long non-coding RNA
LPB	Ligand-Based Pharmacophore
LTP	Long-Term Potentiation

Lys	Lysine
m	magnetic quantum number
m/z	mass to charge ratio
MALDI	Matrix-Assisted Laser Desorption/Ionization
MD	Molecular Dynamics
MeOD	deuterated methanol
MeOH	methanol
miRNA	microRNA
MLR	Multiple Linear Regression
MM-GB-SA	Molecular Mechanics – Generalized Born/Surface Area
MM-PB/SA	Molecular Mechanics- Poisson-Boltzmann/Surface Area
MS	Mass Spectrometry
NA	Nucleic Acid
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser Effect Spectroscopy
NPT	isothermal-isobaric ensemble
NVE	microcanonical ensemble
NVT	canonical ensemble
OGRE	Origin G-Rich Repeated Elements
ORD	Optical Rotatory Dispersion
Orn	Ornithine
PARP	Poly (ADP-Ribose) Polymerase
PBS	Phosphate Buffered Saline
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PDE	PhosphoDiEsterase
PDS	Pyridostatin
PEG	PolyEthyleneGlycole
Phe	Phenylalanine
Phen-DC3	3,3'-[1,10-Phenanthroline-2,9-diylbis(carbonylimino)]bis[1-methylquinolinium]

Phen-Pr	3,3'-(((1,10-phenanthroline-2,9-dicarbonyl)bis(azanediyl))bis(1H-benzo[d]imidazole-2,1-diyl))bis(N,N-dimethylpropan-1-aminium)
PIPER	N,N'-bis[2-(1-piperidino)-ethyl]-3,4,9,10-perylenetetracarboxylic diimide
PLS	Partial Least Square
PQS	Putative G4 Sequences
QIT	Quadrupole Ion Trap
QSAR	Quantitative Structure-Activity Relationship
r.t.	room temperature
R-CPL	Right Circularly Polarised Light
RF	RadioFrequency
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuations
ROC	Receiver Operating Curve
ROS	Reactive Oxygen Species
RPA	Replication Protein A
SAR	Structure-Activity Relationship
SARS-COV-2	Severe Acute Respiratory Syndrome Coronavirus 2
Sec	Selenocysteine
SBDD	Structure-Based Drug Design
SERT	SERotonin Transporter
SMILES	Simplified Molecular Input Line Entry Specification
SOD	SuperOxide Dismutase
SP	Standard Precision
SSRI	Selective Serotonin Reuptake Inhibitor
T	Thymine
TEA	TriEthylAmine
TEAA	TriEthylAmmonium Acetate
TEMED	N,N,N',N'-TEtraMethyl EthyleneDiamine
TER	Telomerase RNA
TERRA	telomeric repeat-containing RNA
TERT	Telomerase Reverse Transcriptase

THF	tetrahydrofuran
TLC	Thin Layer Chromatography
T _m	Melting Temperature
TMPyP2	5,10,15,20-tetrakis-(N-methyl-2-pyridyl)porphine
TMPyP4	5,10,15,20-tetrakis-(N-methyl-4-pyridyl)porphine
TMSA	TriMethylSilyl Acetylene
TOCSY	TOtal Correlation SpectroscopY
TOF	Time-of-flight
TRAP	Telomerase Repeated Amplification Protocol
TSS	Transcription Start Site
Tyr	Tyrosine
UTR	UnTraslated Region
V	Potential
XP	Extra Precision
β-Ala	beta-Alanine
δ	chemical shift
θ	Ellipticity
ω ₀	Larmor frequency

5.2 Flavonoids

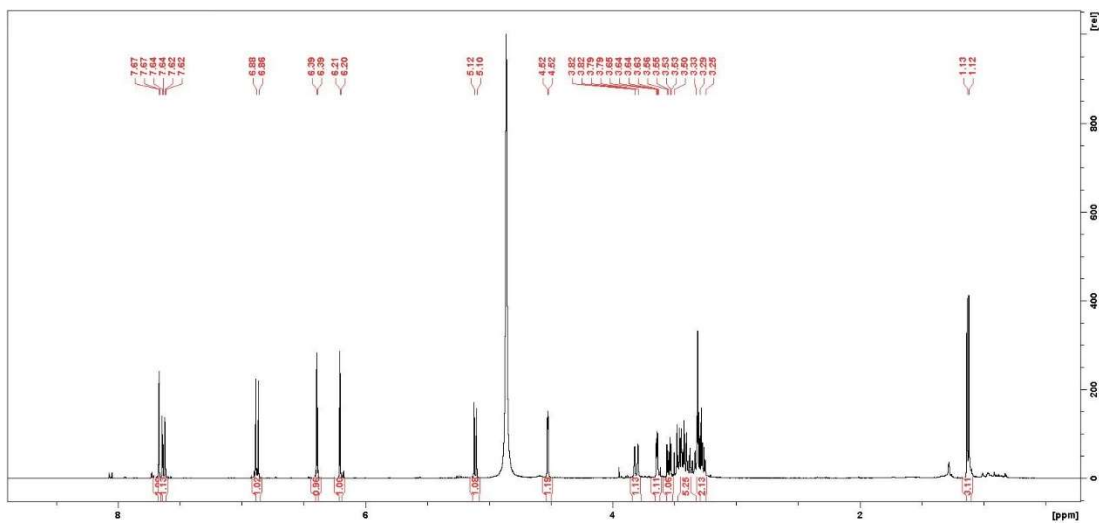


Figure 5.2.1. ¹H-NMR spectrum of rutin (CD₃OD).

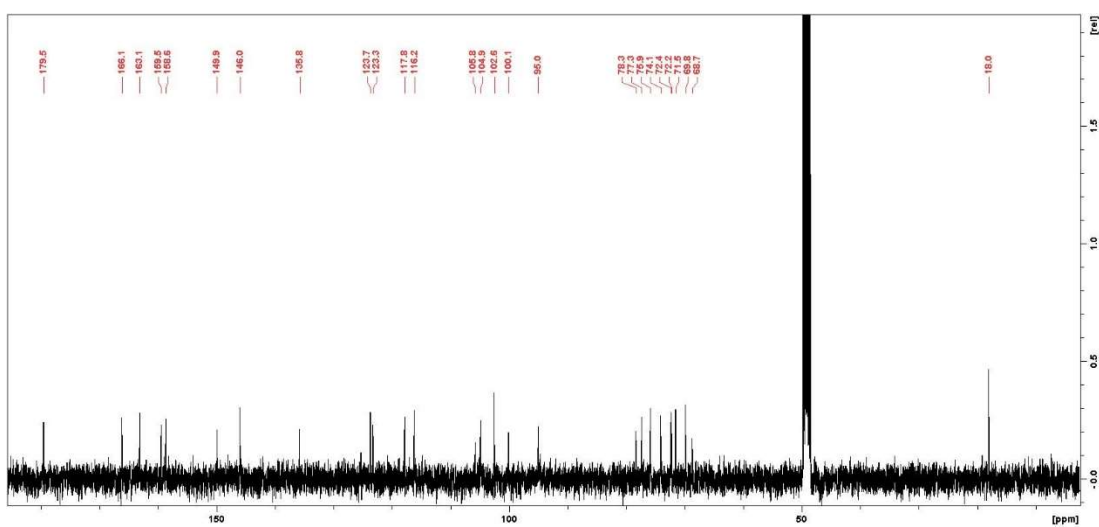


Figure 5.2.2. ¹³C-NMR spectrum of rutin (CD₃OD).

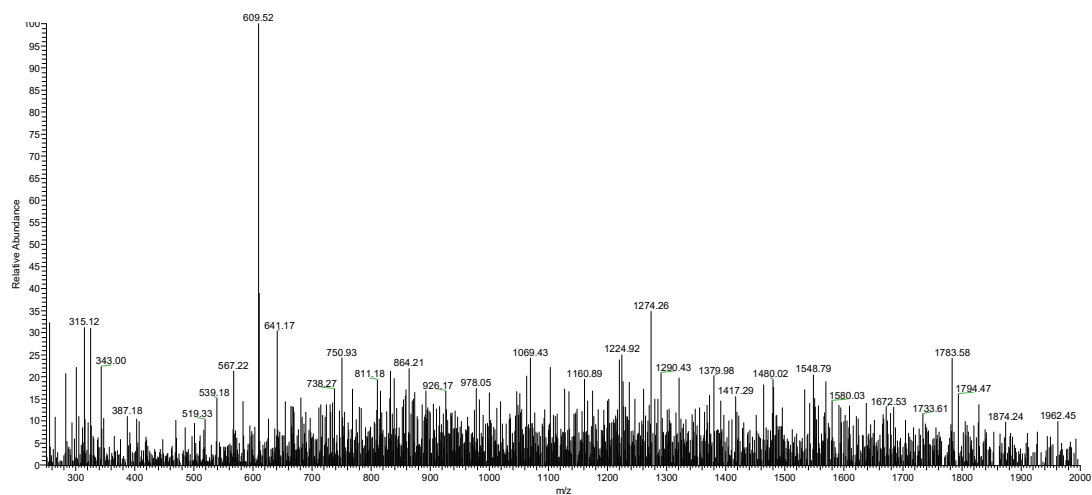


Figure 5.2.3. ESI-MS spectrum of rutin ($m/z = 609.52$, negative ionization mode).

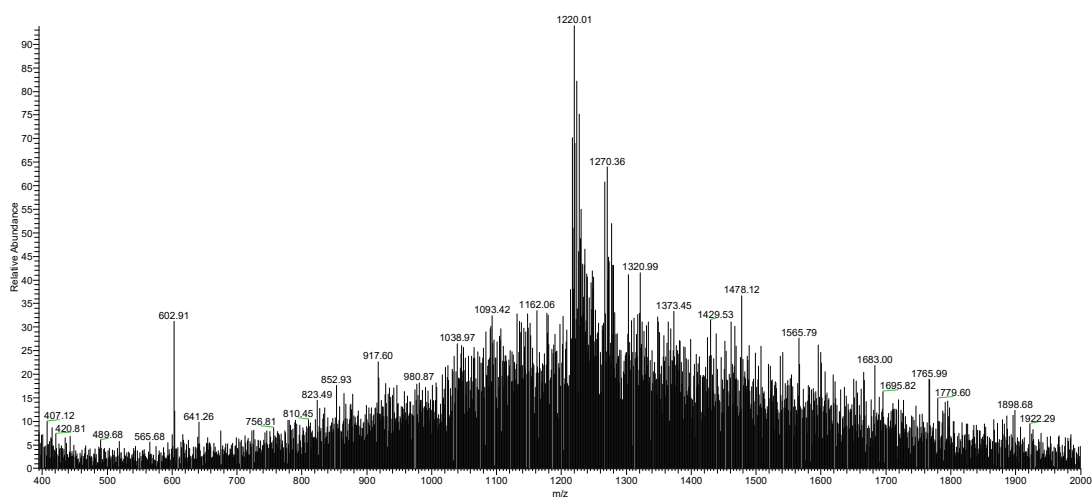


Figure 5.2.4. Mass spectrum of the G-quadruplex/quercetin binding experiment. The molecular ion of the quercetin “dimer” is present ($m/z = 602.91$). The signal at $m/z = 1220.01$ was attributed to the unbound DNA ($z = -6$). The signal at $m/z = 1270.36$ was identified as an interaction peak ($z = -6$, 1:1 G-quadruplex/rutin complex).

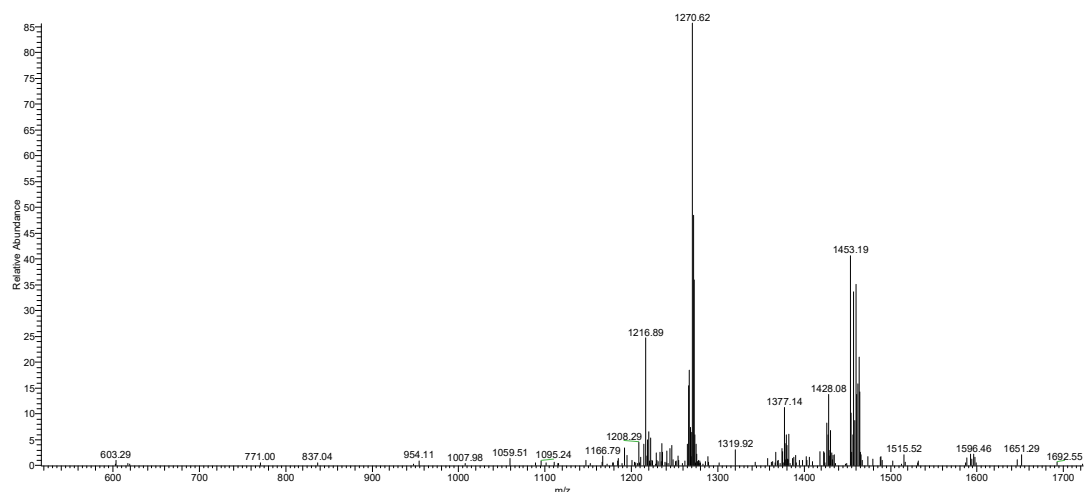


Figure 5.2.5. CID fragmentation spectrum of the G-quadruplex/quercetin complex (normalized collision energy = 10; isolated peak $m/z = 1270.62$). The signal of the complex is present (1270.62 , $z = -6$), and molecular ion of naked DNA (1453.19 , $z = -5$) and of the ligand “dimer” ($m/z = 603.29$, $z = -1$) appear after fragmentation.

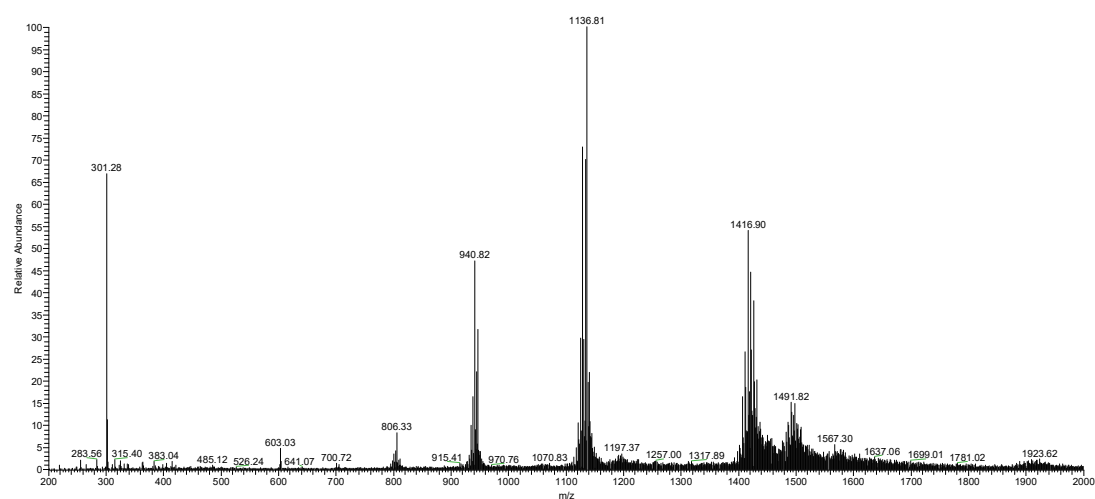


Figure 5.2.6. Mass spectrum of the dsDNA/quercetin binding experiment. The molecular ion of quercetin is present ($m/z = 301.26$), as well as the peak of the dimer ($m/z = 603.03$). The signals at $m/z = 940.82$, 1136.81 and 1418.90 were attributed to the unbound DNA ($z = -12$, $z = -10$ and $z = -8$, respectively). The signal at $m/z = 1491.82$ was identified as an interaction peak ($z = -8$, 1:1 dsDNA/quercetin complex).

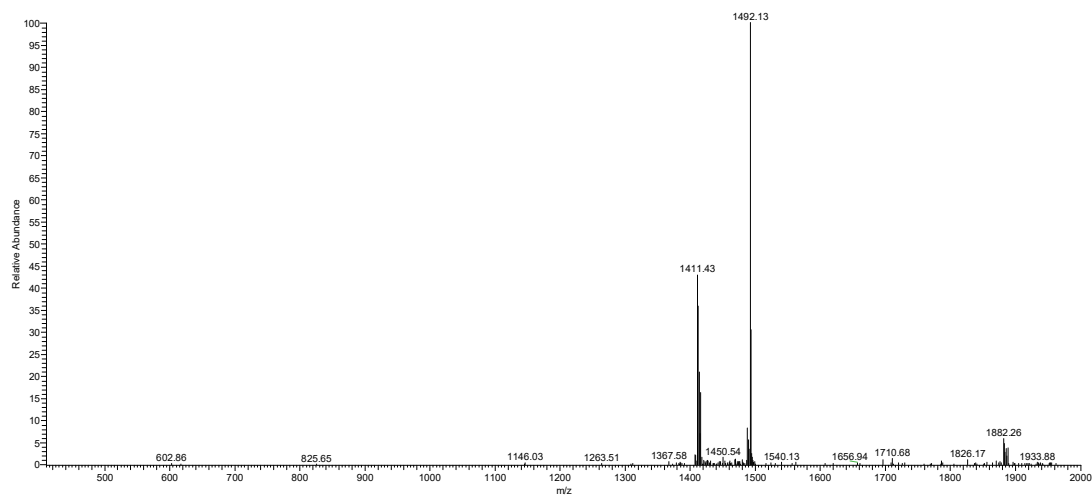


Figure 5.2.7. CID fragmentation spectrum of the dsDNA/quercetin complex (normalized collision energy = 10; isolated peak $m/z = 1492.13$). The signal of the complex is present (1492.13 , $z = -8$), and molecular ion of naked DNA (1411.43 , $z = -10$) and of the ligand “dimer” ($m/z = 602.86$, $z = -1$) appear after fragmentation.

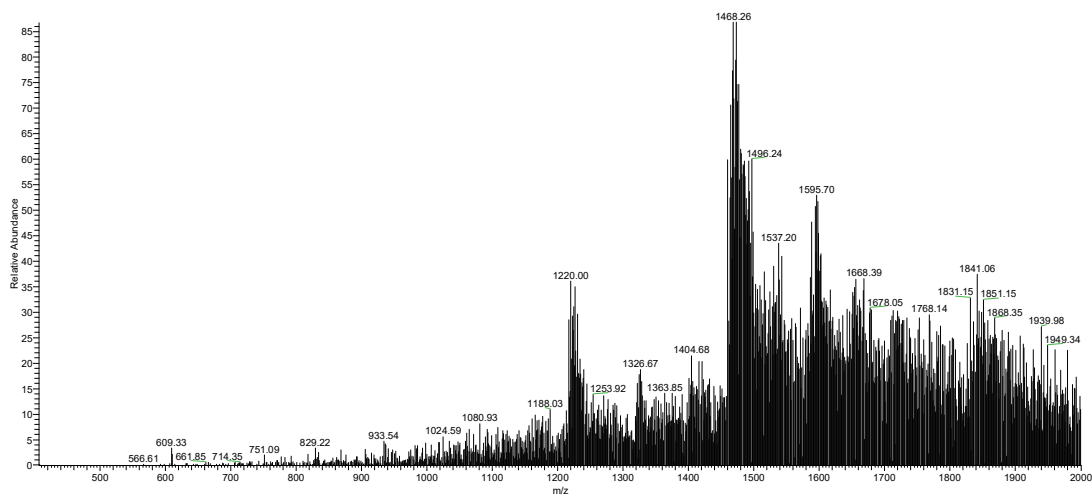


Figure 5.2.8. Mass spectrum of the G-quadruplex/rutin binding experiment. The molecular ion of rutin is present ($m/z = 609.33$). The signals at $m/z = 2220.00$, 1468.26 and 1841.06 were attributed to the unbound DNA ($z = -6$, $z = -5$ and $z = -4$, respectively). An interaction peak was detected at $m/z = 1595.70$ ($z = -5$, 1:1 G-quadruplex/rutin complex).

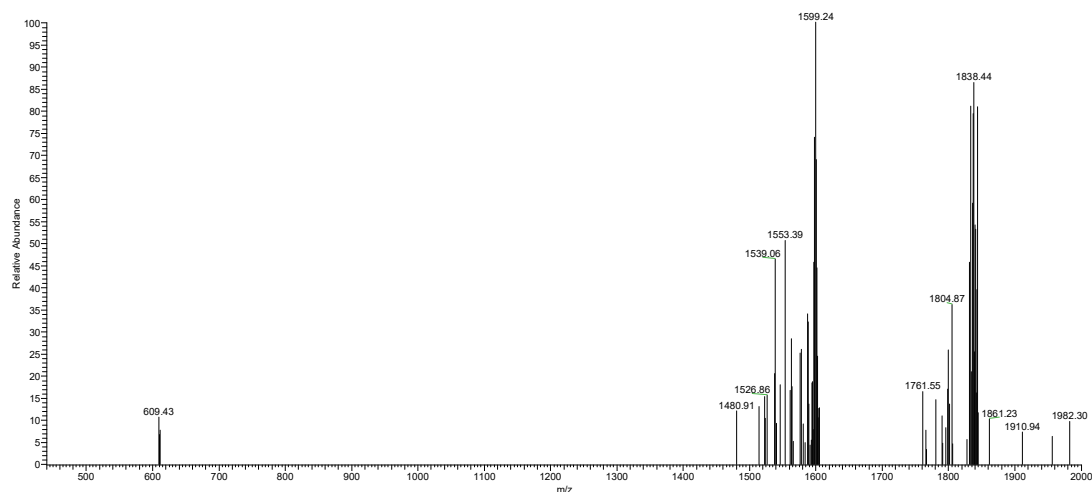


Figure 5.2.9. CID fragmentation spectrum of the G-quadruplex/rutin complex (normalized collision energy = 10; isolated peak $m/z = 1599.24$). The signal of the complex is present ($m/z = 1599.24$, $z = -5$), and molecular ion of naked DNA ($m/z = 1838.44$, $z = -4$) and ligand ($m/z = 609.43$) appear after the fragmentation.

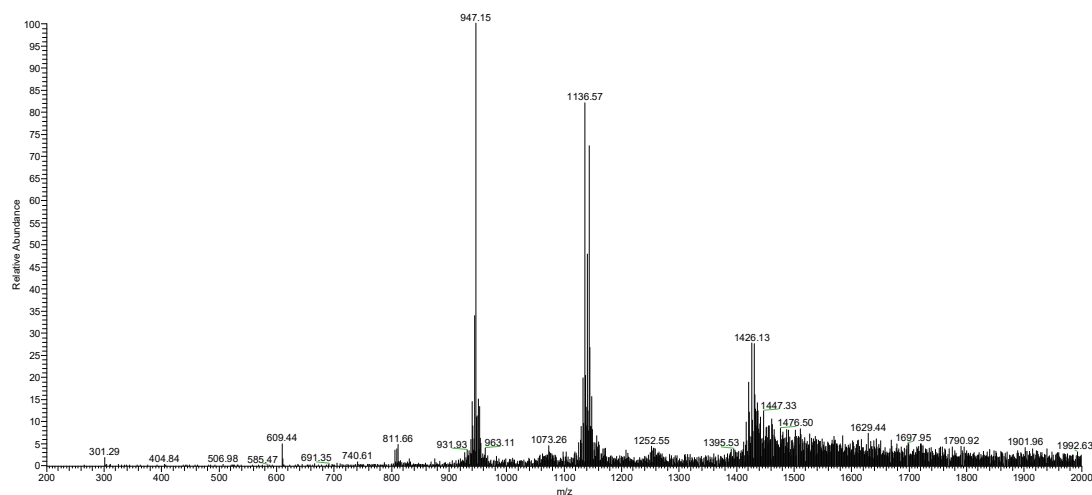


Figure 5.2.10. Mass spectrum of the dsDNA/rutin binding experiment. The molecular ion of rutin is present ($m/z = 609.44$). The signals at $m/z = 947.15$, 1136.57 and 1426.13 were attributed to the unbound DNA ($z = -12$, $z = -10$ and $z = -8$, respectively).

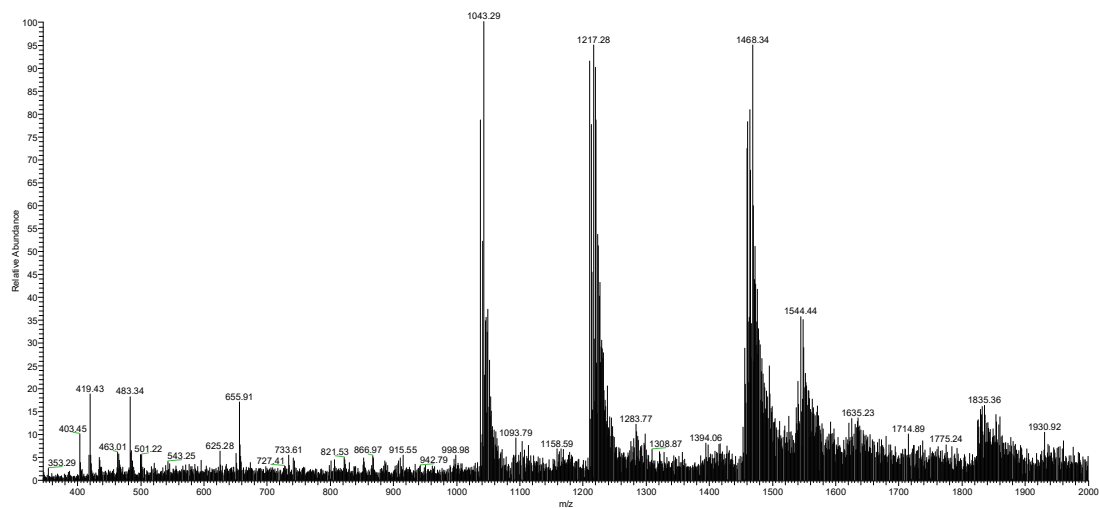


Figure 5.2.11. Mass spectrum of the G4/isoosajin binding experiment. The signals at $m/z = 1043.29$, 1217.28 and 1468.34 were attributed to the unbound DNA ($z = -7$, $z = -6$ and $z = -5$, respectively), the signal at $m/z 403.45$ was attributed to unbounded ligand. The interaction peaks were detected at $m/z = 1093.79$, 1283.77 and 1544.44 ($z = -7$, $z = -6$ and $z = -5$, respectively).

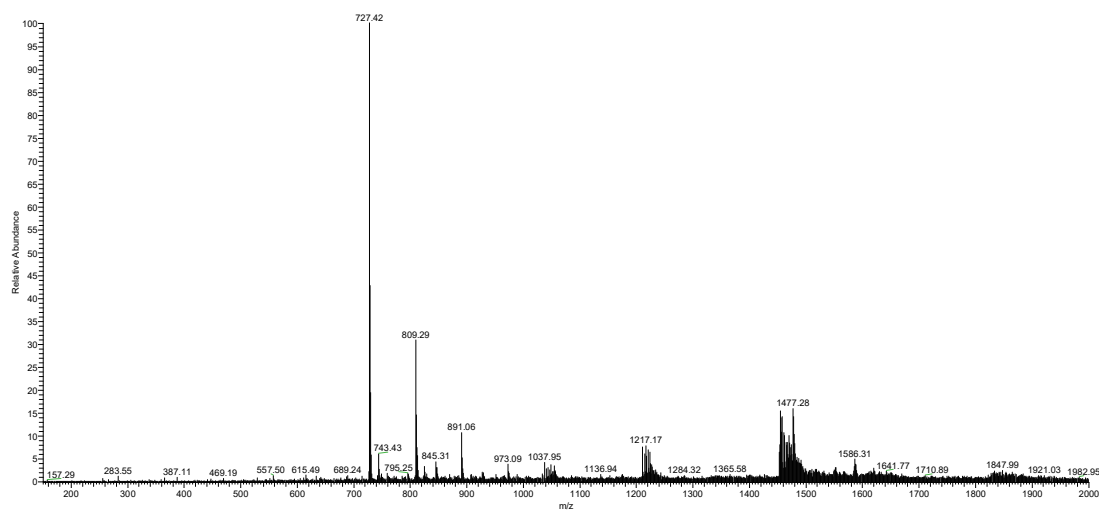


Figure 5.2.12. Mass spectrum of the G4/di-*p*-toluensulfonyl-pomiferin binding experiment. The signals at $m/z = 1037.95$, 1217.17 and 1477.28 were attributed to the unbound DNA ($z = -7$, $z = -6$ and $z = -5$, respectively), the signal at $m/z 727.42$ was attributed to unbounded ligand. No interaction peaks were detected.

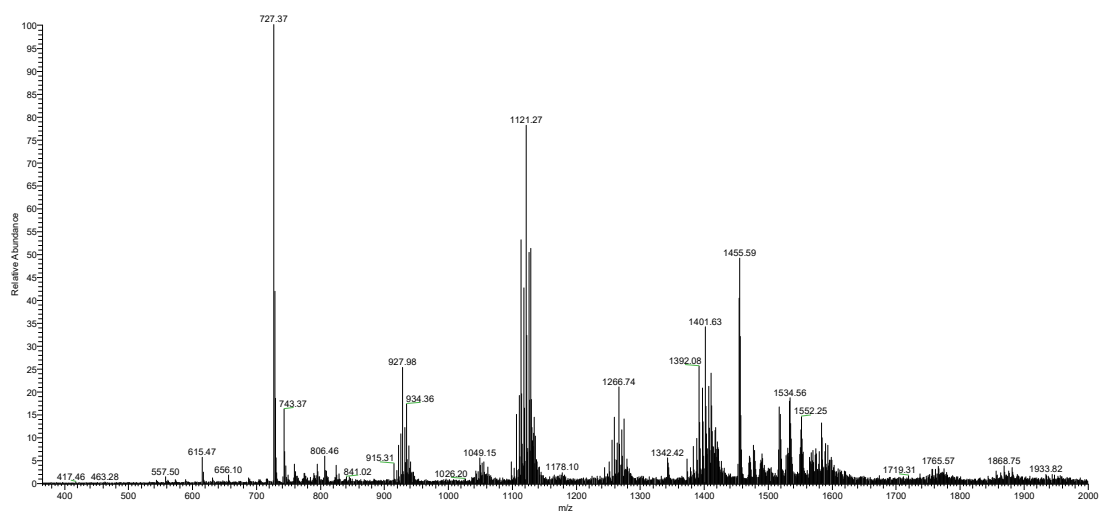


Figure 5.2.13. Mass spectrum of the dsDNA/di-*p*-toluensulfonyl-auricularin binding experiment. The signals at $m/z = 927.98$, 1121.27 and 1401.63 were attributed to the unbound DNA ($z = -12$, $z = -10$ and $z = -8$, respectively), the signal at $m/z 727.42$ was attributed to unbounded ligand. The interaction peaks were detected at $m/z = 1049.15$, 1266.74 (ligand:DNA 2:1, $z = -12$ and $z = -10$, respectively).

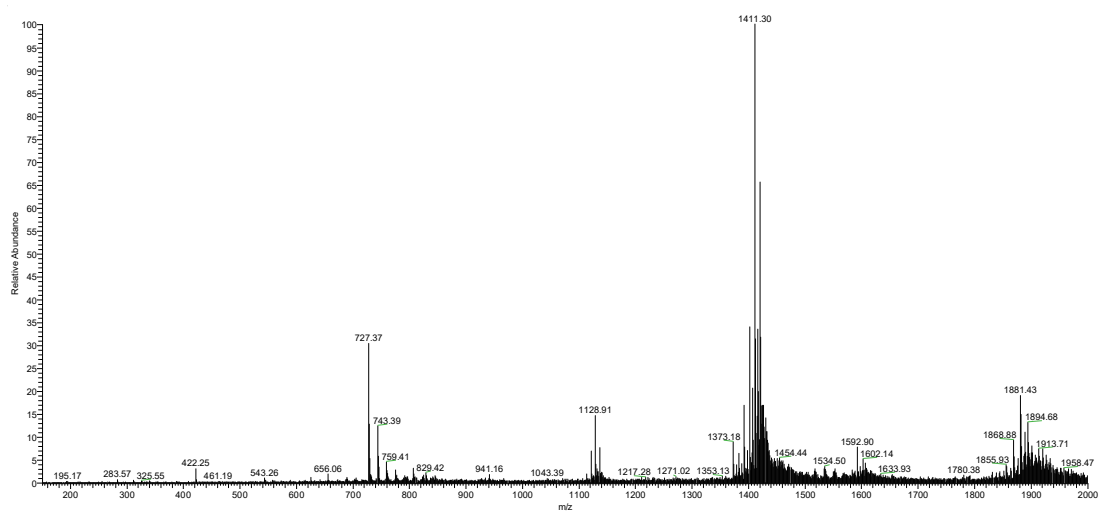


Figure 5.2.14. Mass spectrum of the dsDNA/di-*p*-toluensulfonyl-pomiferin binding experiment. The signals at $m/z = 1128.91$, 1411.30 and 1881.43 were attributed to the unbound DNA ($z = -10$, $z = -8$ and $z = -6$, respectively), the signal at $m/z 727.37$ was attributed to unbounded ligand. No interaction peaks were detected.

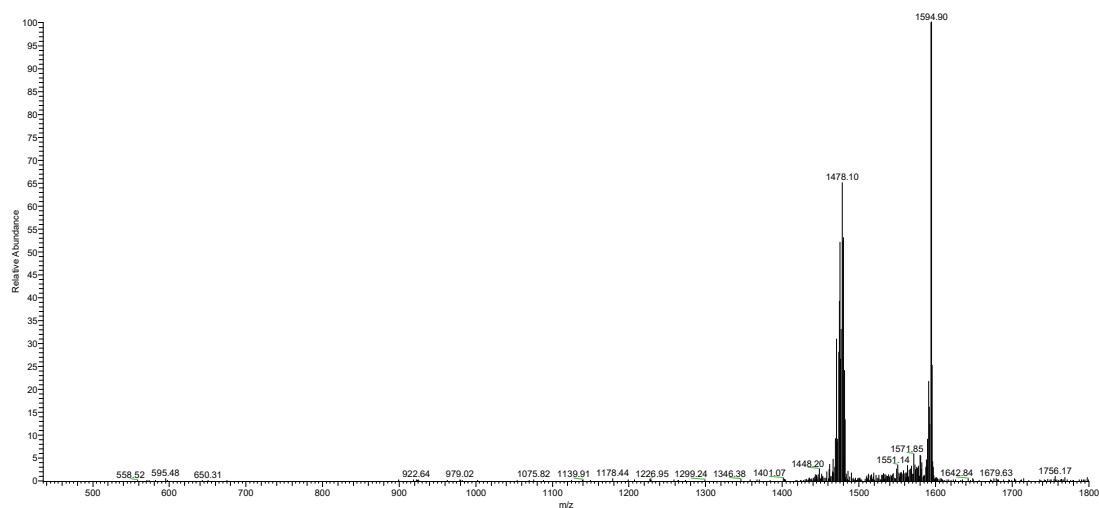


Figure 5.2.15. CID fragmentation spectrum of the G4/*p*-toluenesulfonyl-scandene complex (normalized collision energy = 10). The signal of the complex is still present ($m/z = 1594.90$, $z = -5$), the peak at $m/z = 1478.10$ was attributed to alkylated G4.

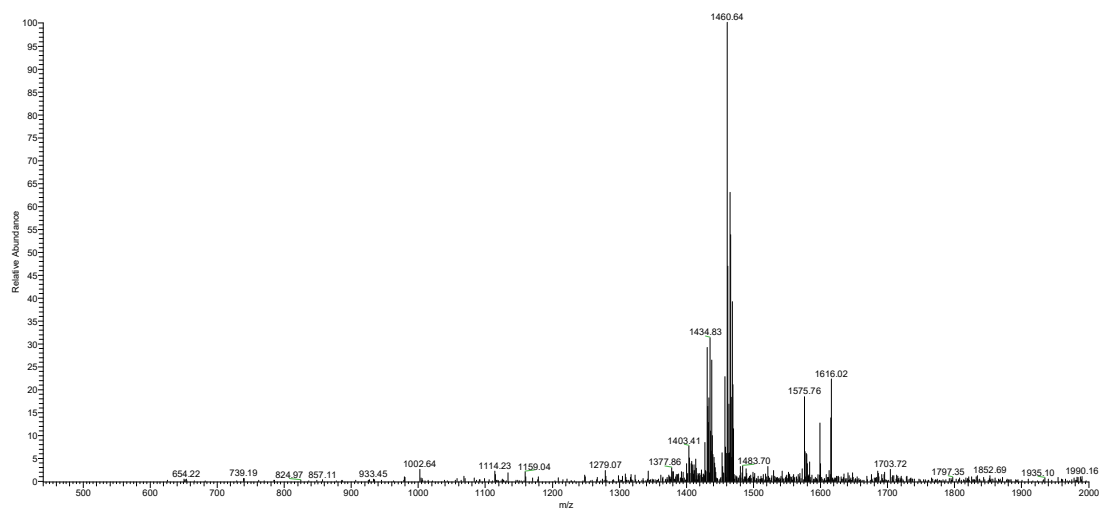


Figure 5.2.16. CID fragmentation spectrum of the G4/*p*-toluenesulfonyl-isopomiferin complex (normalized collision energy = 16). The signal of the complex is still present ($m/z = 1616.02$, $z = -5$), the peak at $m/z = 1460.64$ was attributed to unbounded G4, the signals at $m/z = 1434.83$ (-151 Da) and 1403.41 (-151 Da -136 Da) were attributed to DNA sequence after base loss.

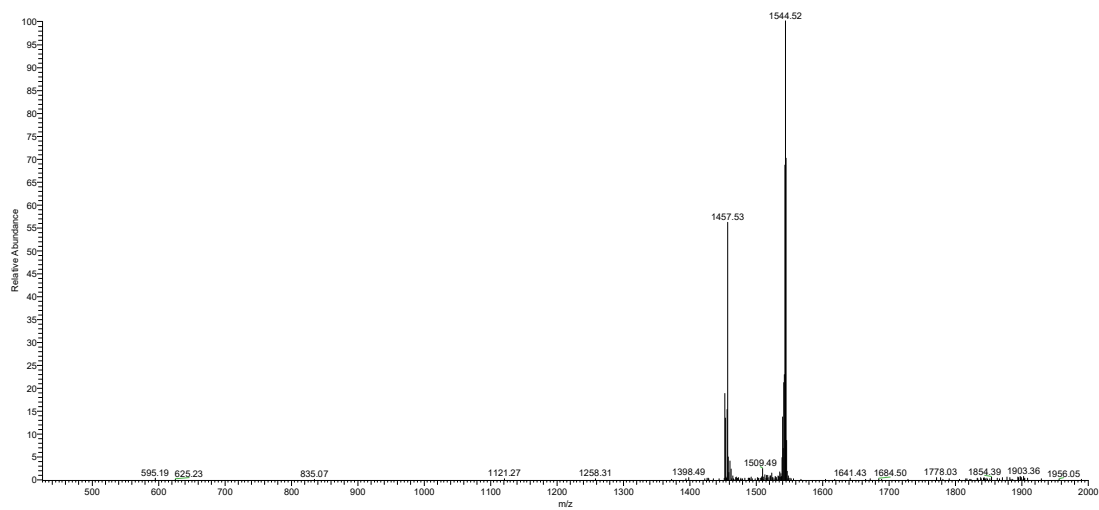


Figure 5.2.17. CID fragmentation spectrum of the G4/isopomiferin complex (normalized collision energy = 8). The signal of the complex is still present ($m/z = 1544.52$, $z = -5$), the peak at $m/z = 1457.53$ corresponding to the unbounded G4 starts to appear.

5.3 Anthraquinones

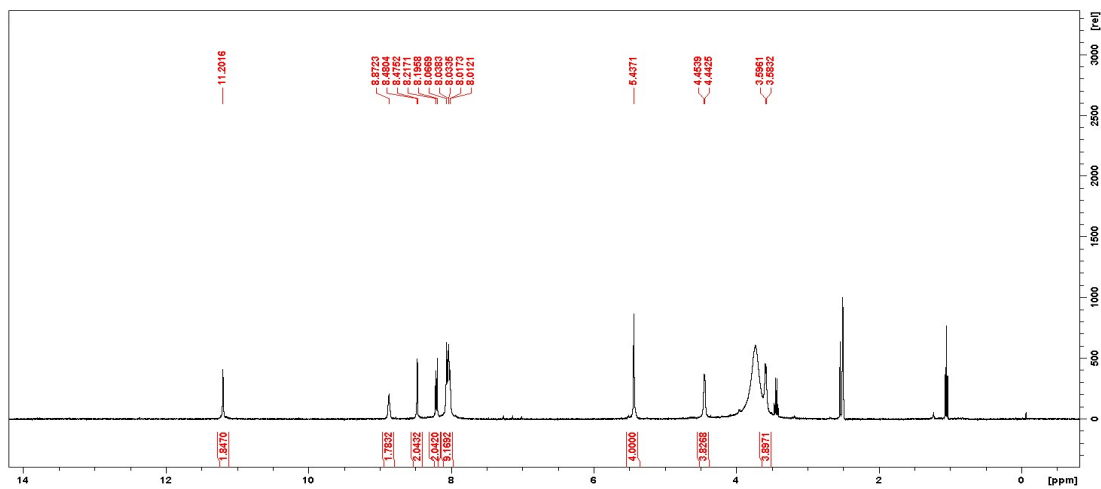


Figure 5.3.1. ¹H-NMR spectrum of compound AQ3.

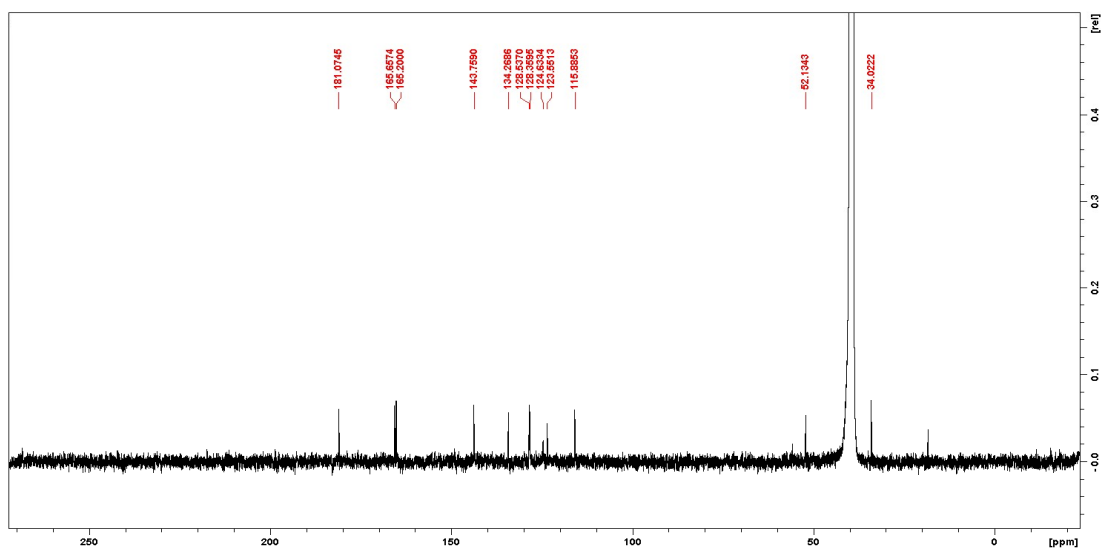


Figure 5.3.2. ¹³C-NMR spectrum of compound AQ3.

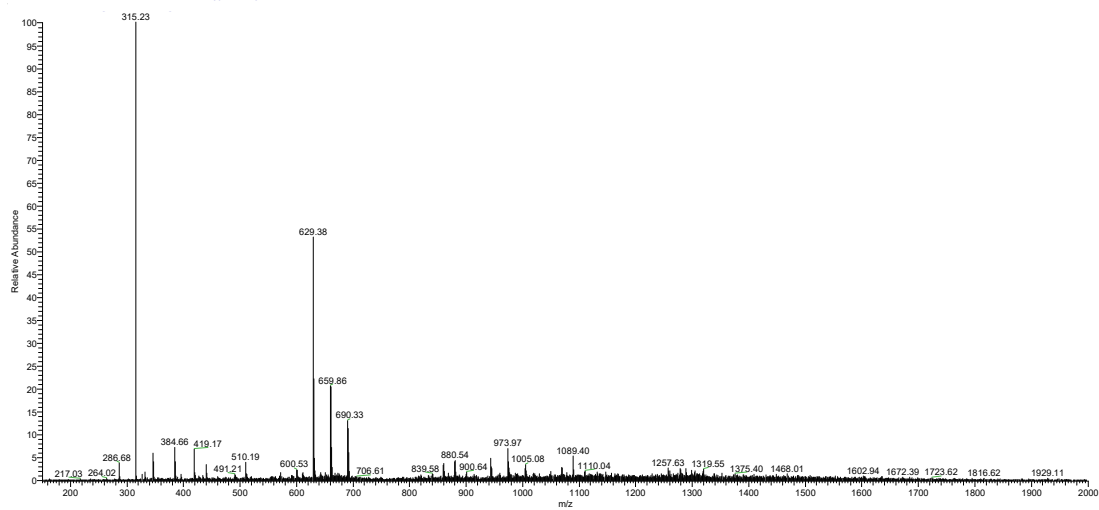


Figure 5.3.3. ESI-MS spectrum of compound AQ3.

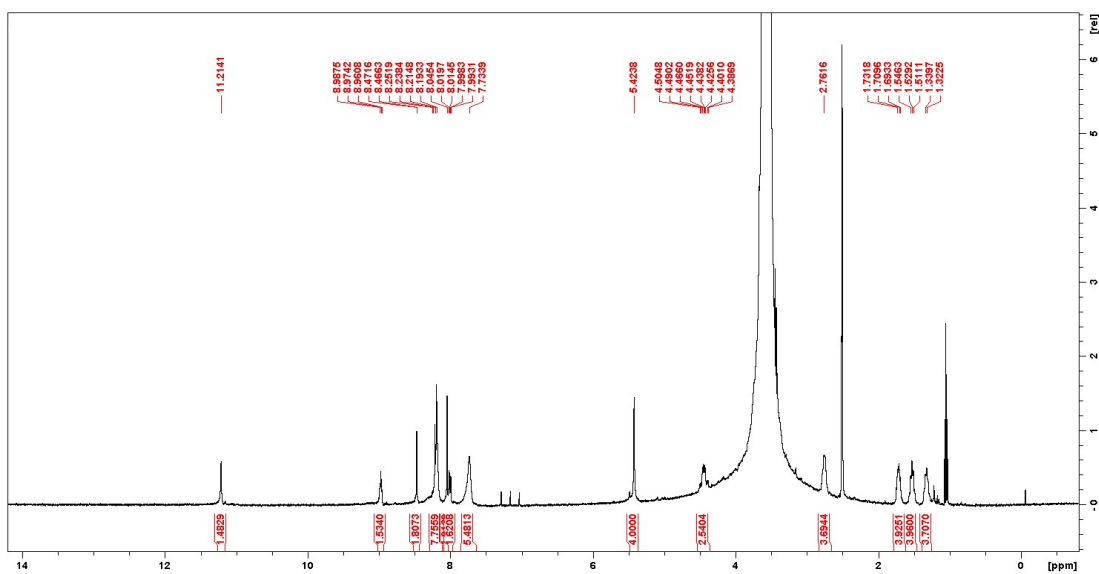


Figure 5.3.4. ¹H-NMR spectrum of compound AQ4.

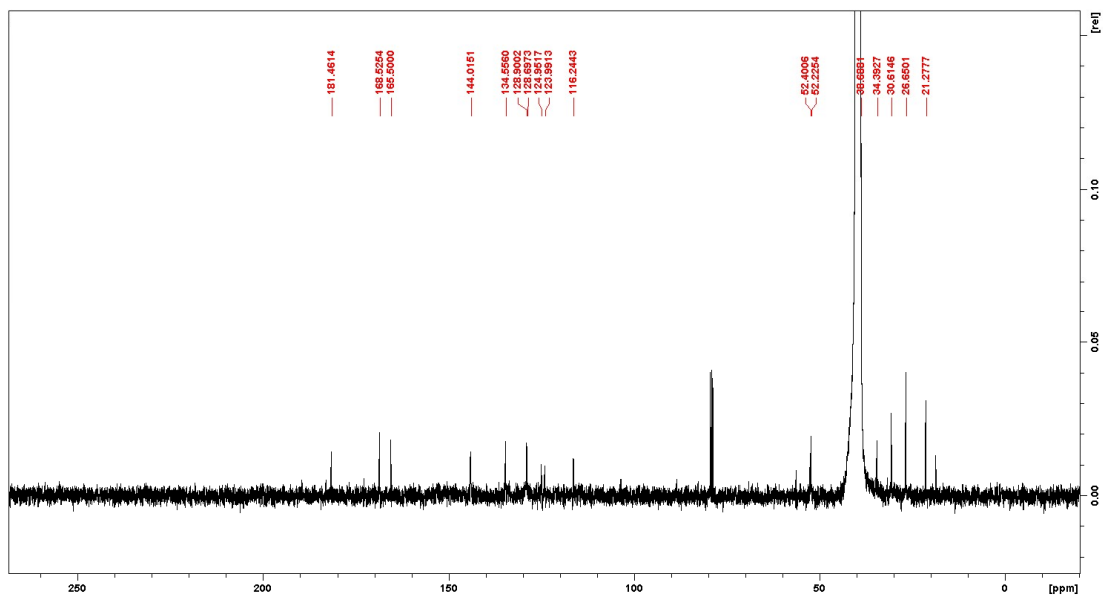


Figure 5.3.5. ¹³C-NMR spectrum of compound AQ4.

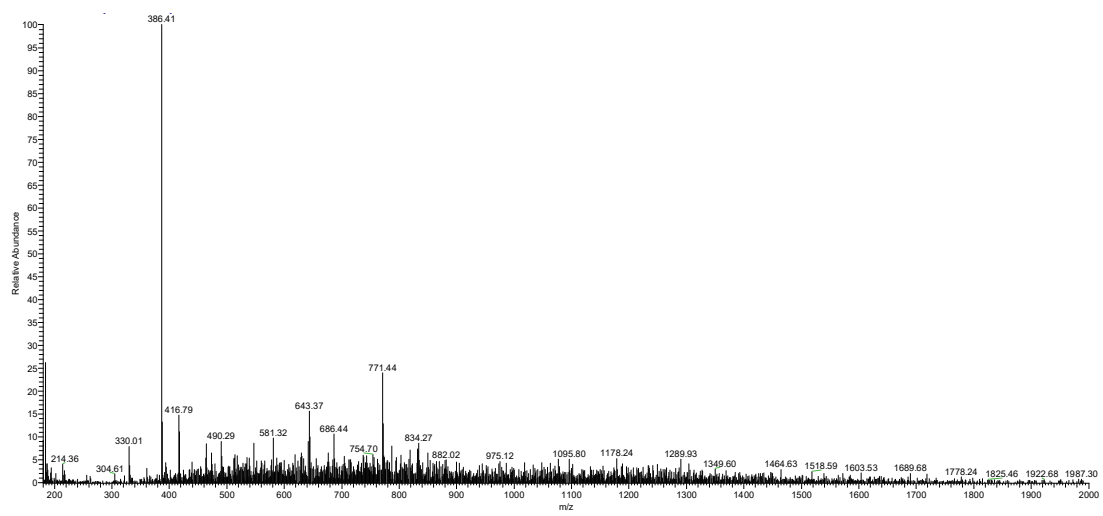


Figure 5.3.6. ESI-MS spectrum of compound AQ4.

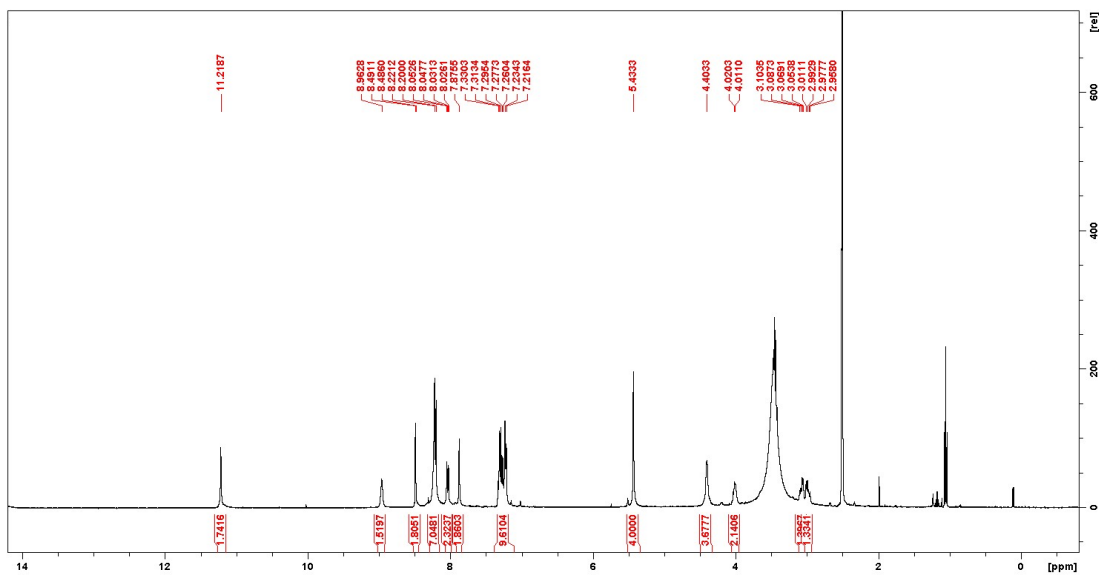


Figure 5.3.7. ¹H-NMR spectrum of compound AQ5.

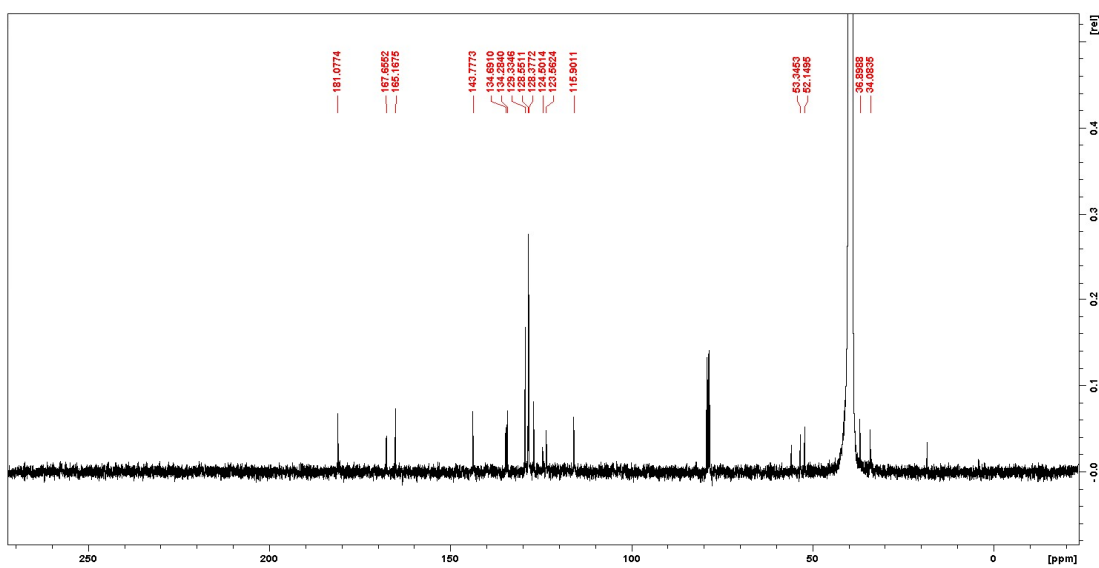


Figure 5.3.8. ¹³C-NMR spectrum of compound AQ5.

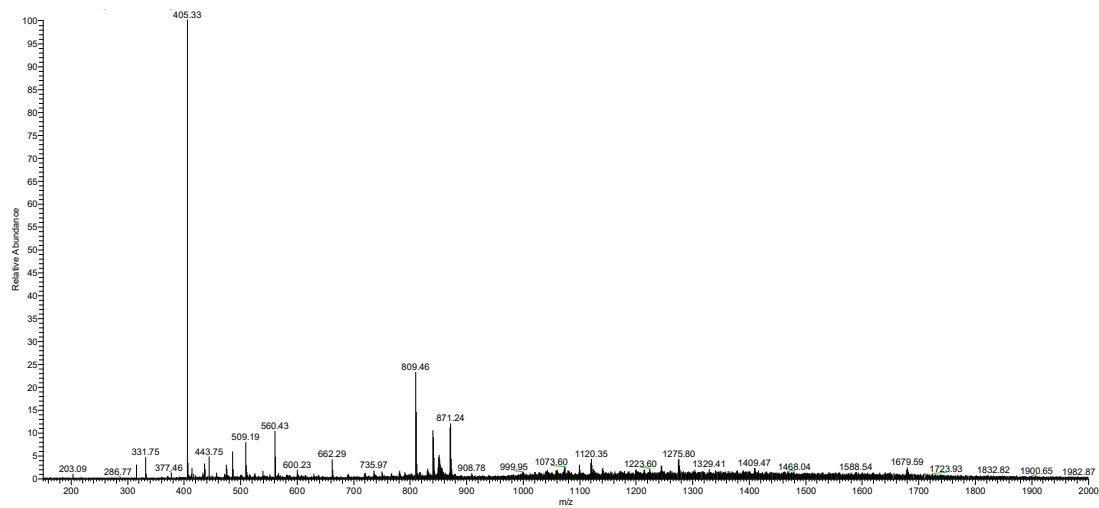


Figure 5.3.9. ESI-MS spectrum of compound AQ5.

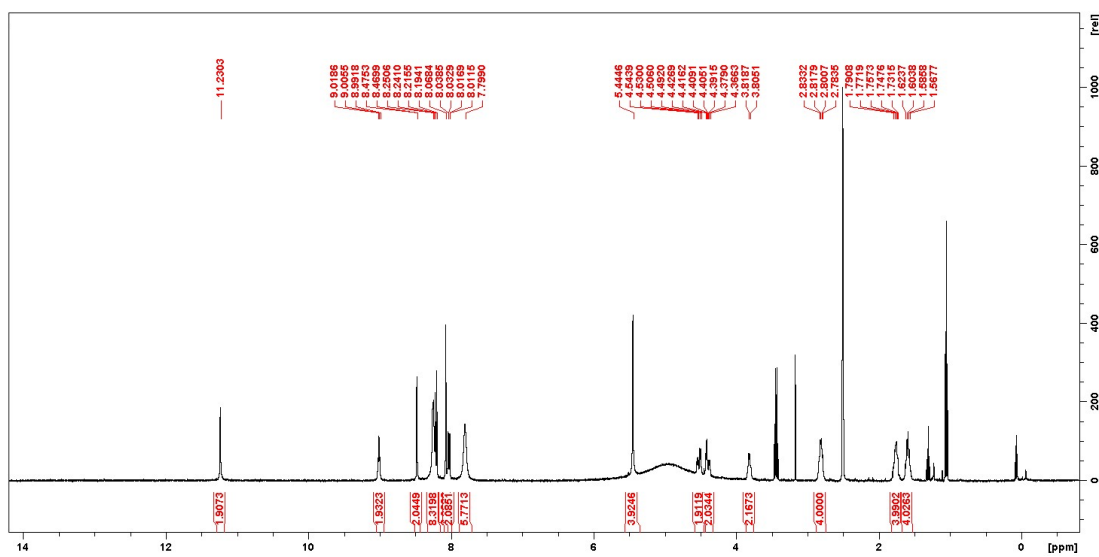


Figure 5.3.10. ¹H-NMR spectrum of compound AQ6.

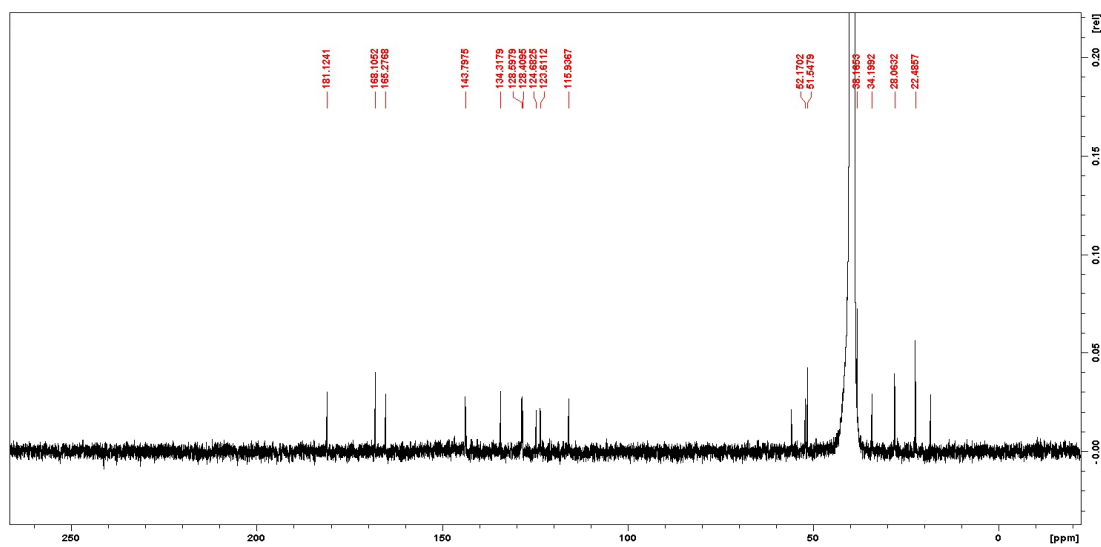


Figure 5.3.11. ^{13}C -NMR spectrum of compound AQ6.

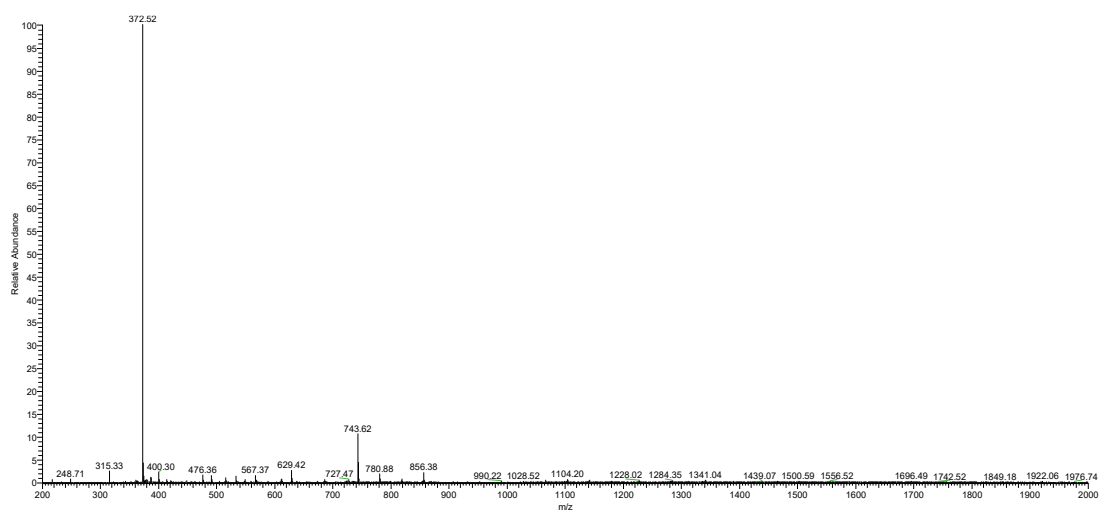


Figure 5.3.12. ESI-MS spectrum of compound AQ6.

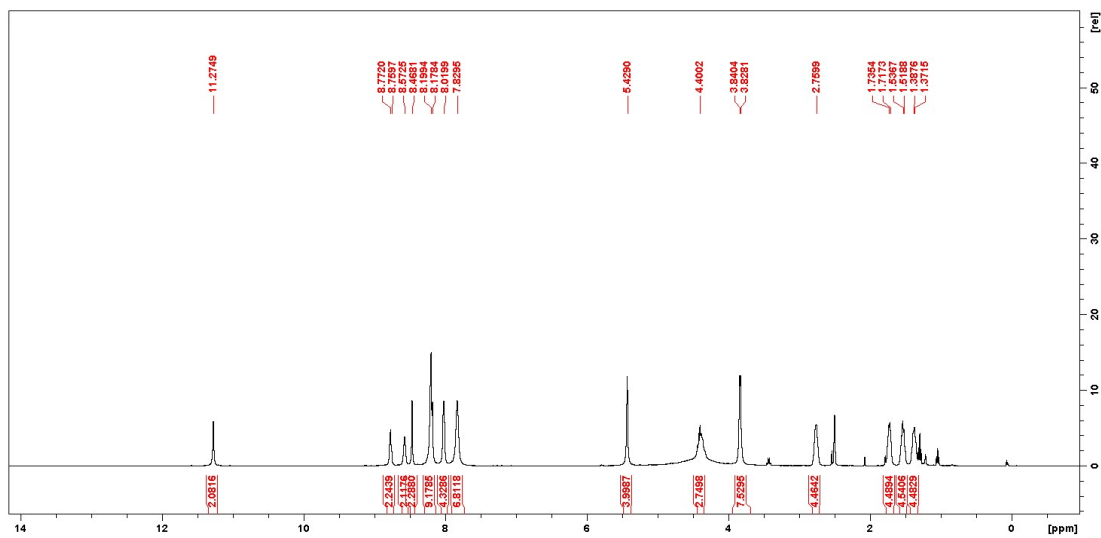


Figure 5.3.13. ¹H-NMR spectrum of compound AQ7.

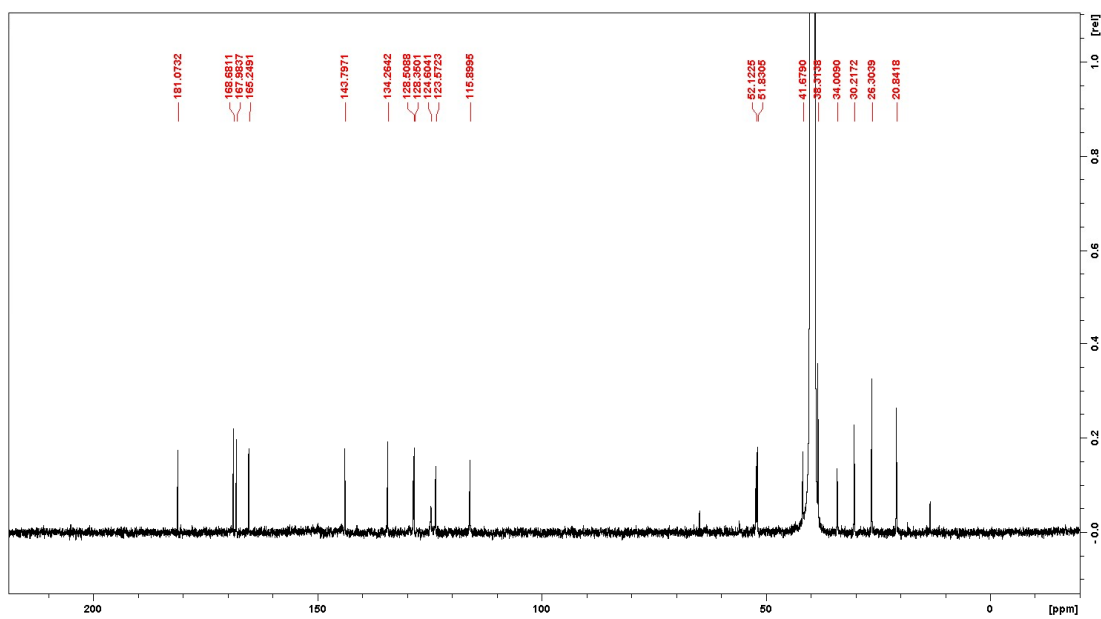


Figure 5.3.14. ¹³C-NMR spectrum of compound AQ7.

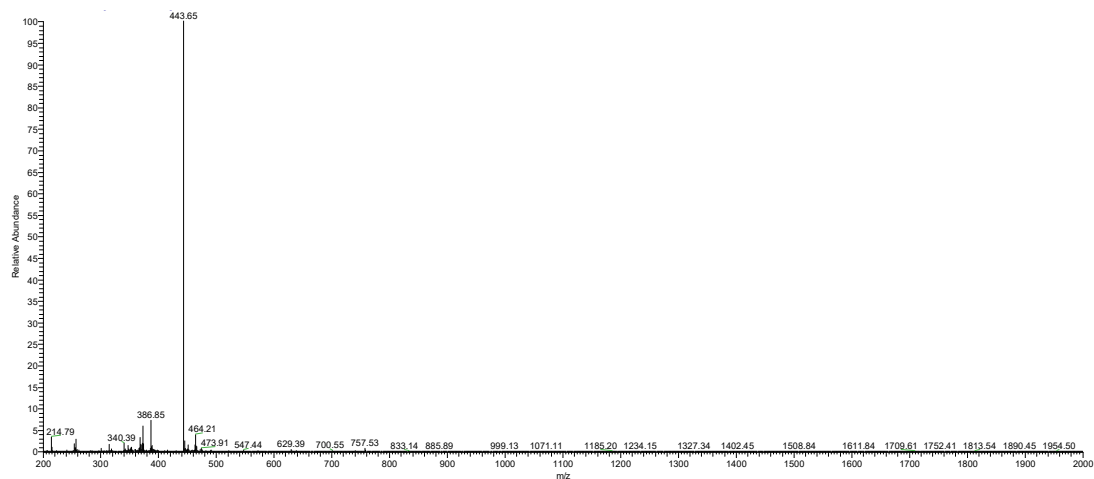


Figure 5.3.14. ESI-MS spectrum of compound AQ7.

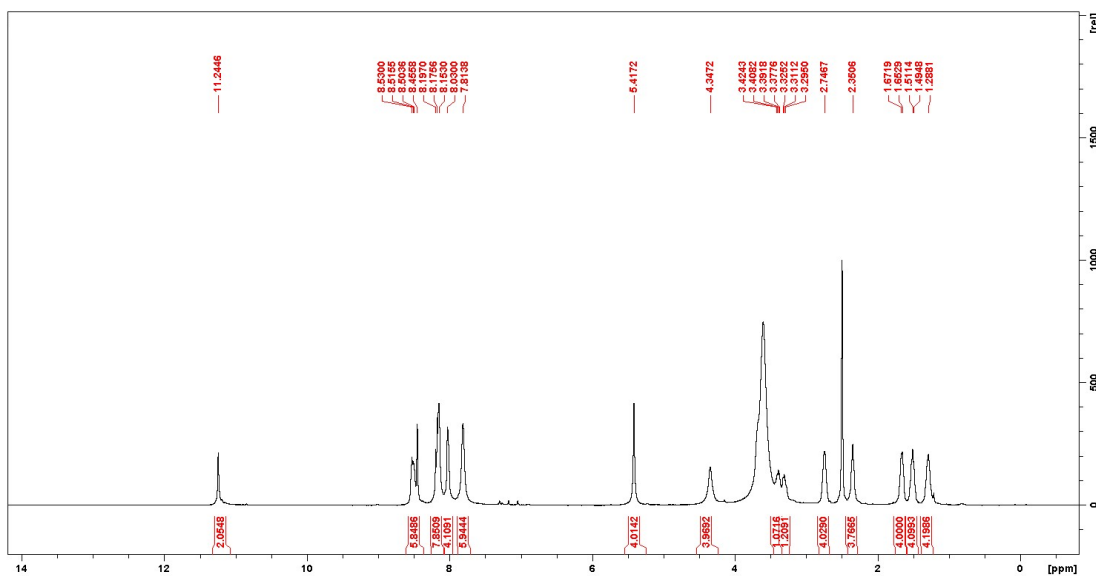


Figure 5.3.15. ¹H-NMR spectrum of compound AQ8.

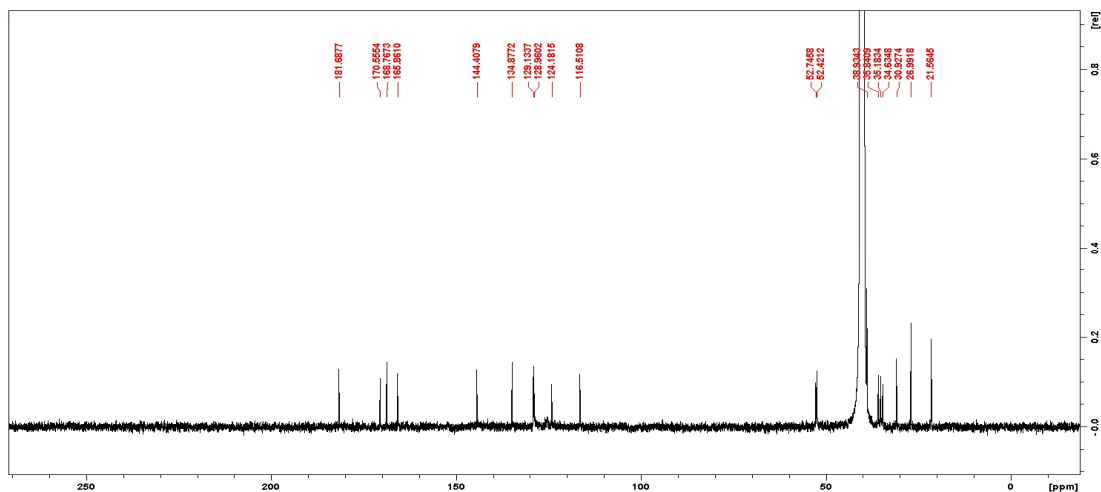


Figure 5.3.16. ¹³C-NMR spectrum of compound AQ8.

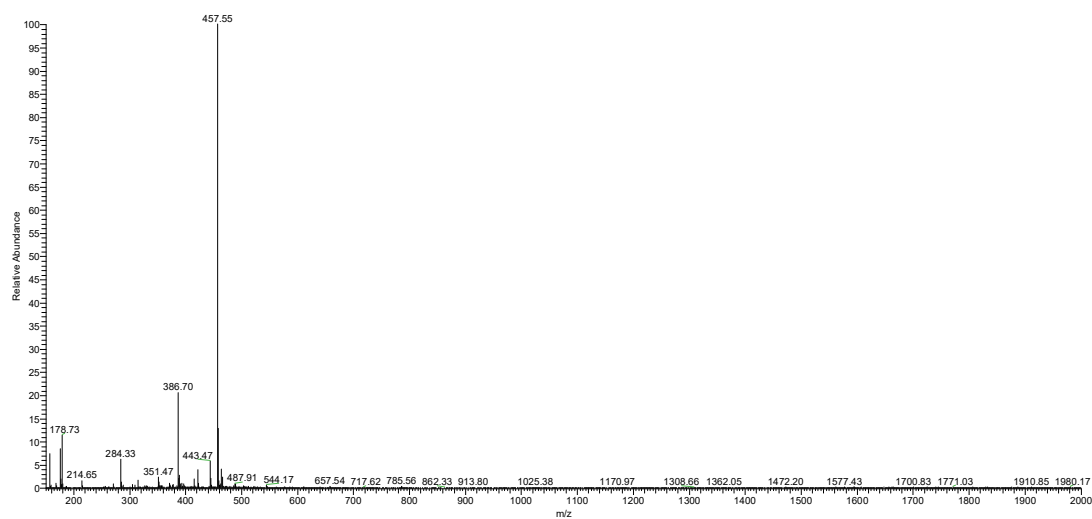


Figure 5.3.17. ESI-MS spectrum of compound AQ8.

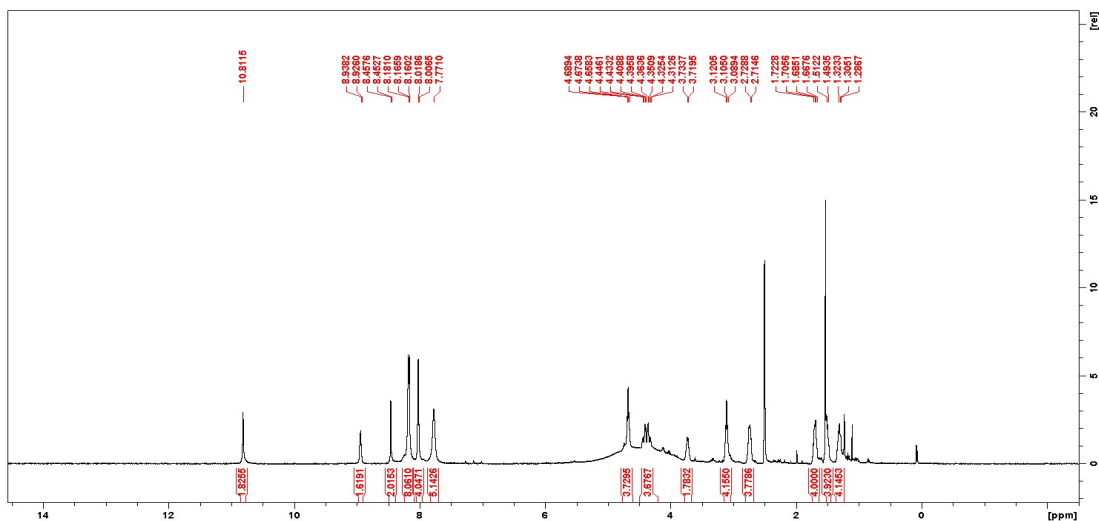


Figure 5.3.18. ¹H-NMR spectrum of compound AQ9.

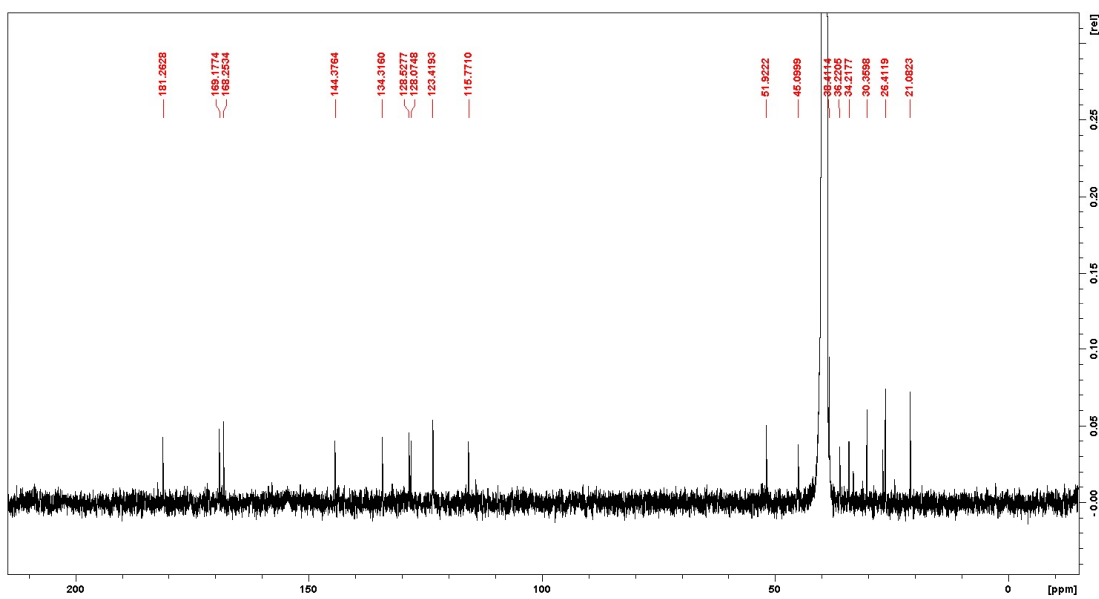


Figure 5.3.19. ¹³C-NMR spectrum of compound AQ9.

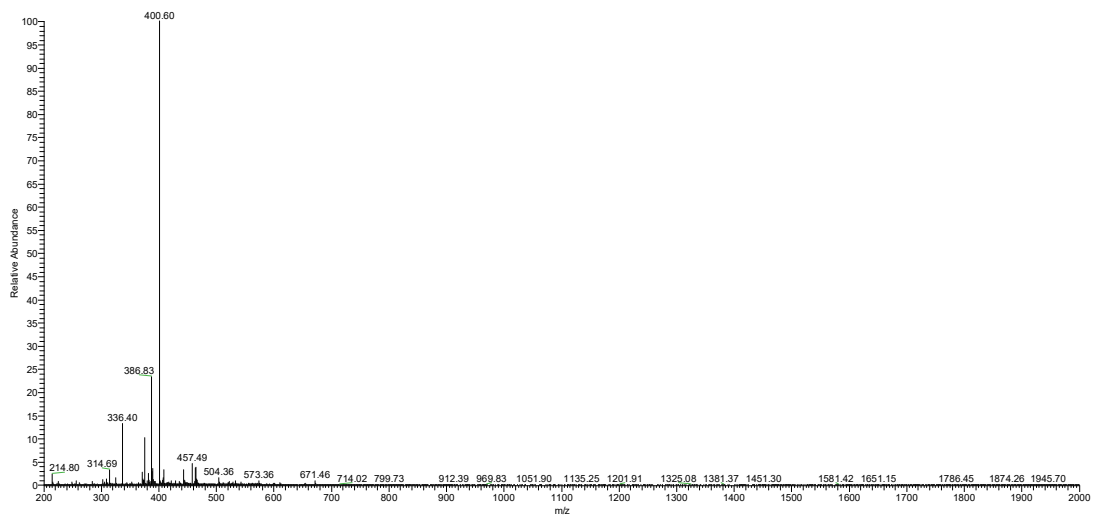


Figure 5.3.20. ESI-MS spectrum of compound AQ9.

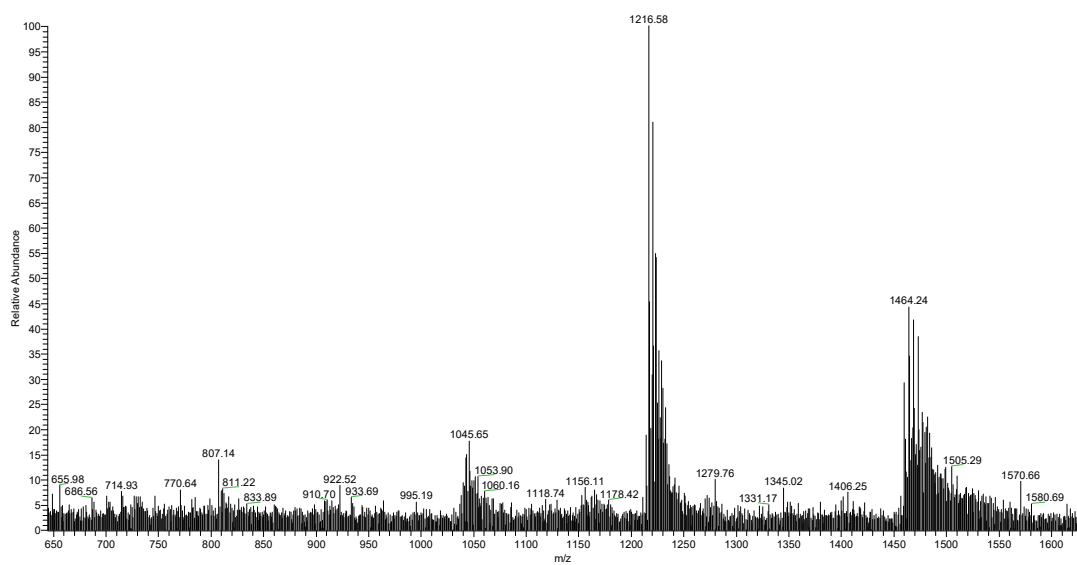


Figure 5.3.21. Mass spectrum of the G4/AQ3 binding experiment. The signals at $m/z = 1045.65$, 1216.58 , and 1464.24 were attributed to the unbound DNA ($z = -7$, $z = -6$ and $z = -5$, respectively). No interaction peaks were detected.

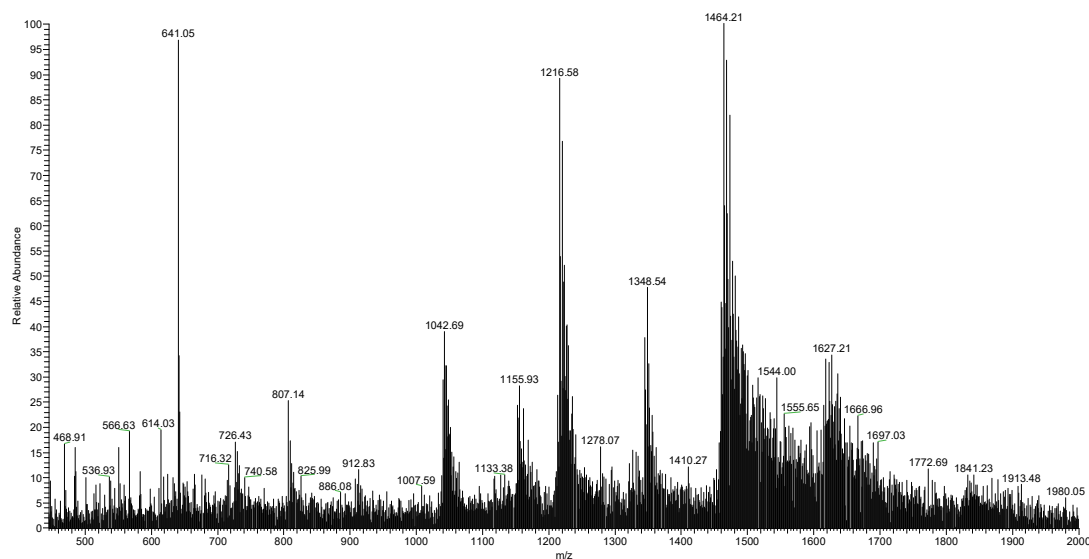


Figure 5.3.22. Mass spectrum of the G4/AQ4 binding experiment. The signals at $m/z = 1042.69$, 1216.58 , and 1464.21 were attributed to the unbound DNA ($z = -7$, $z = -6$ and $z = -5$, respectively). The interaction peaks were detected at $m/z = 1155.93$, 1348.54 and 1627.21 ($z = -7$, $z = -6$ and $z = -5$, respectively).

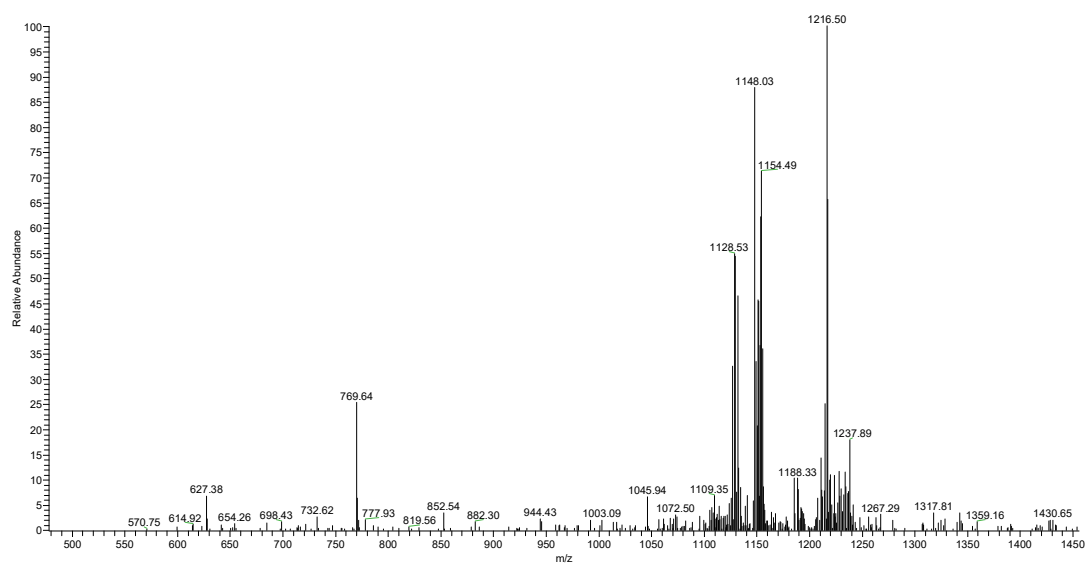


Figure 5.3.23. CID fragmentation spectrum of the G4/AQ4 complex (normalized collision energy = 12). The signal of the complex is still present ($m/z = 1154.49$, $z = -7$), and molecular ion of naked DNA ($m/z = 1216.50$, $z = -6$) and ligand ($m/z = 769.64$) start to appear.

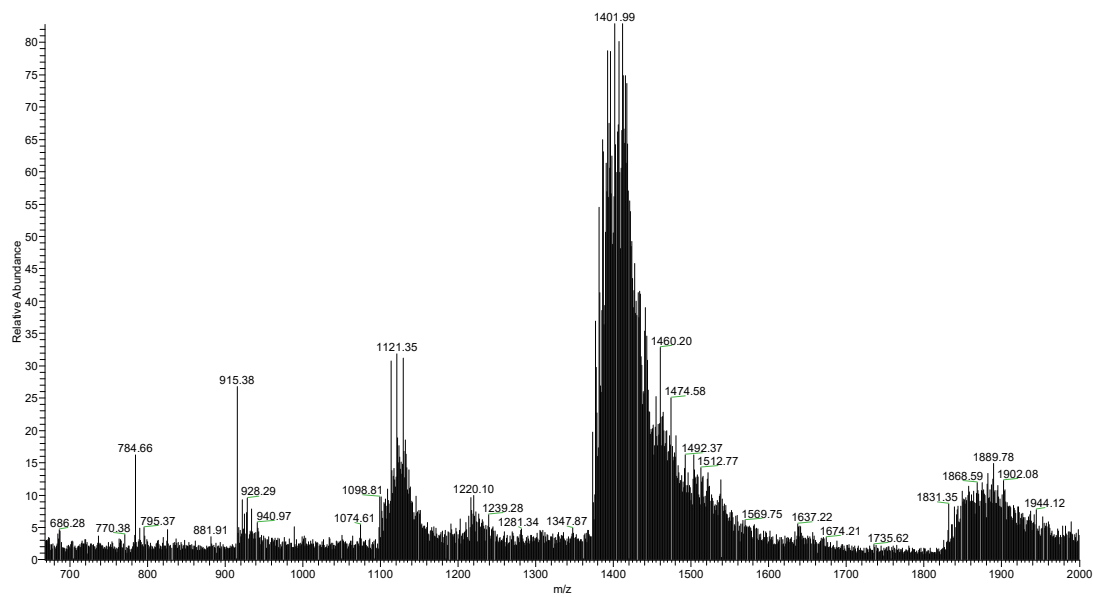


Figure 5.3.24. Mass spectrum of the dsDNA/AQ7 binding experiment. The signals at $m/z = 915.38$, 1121.35 and 1401.99 were attributed to the unbound DNA ($z = -12$, $z = -10$ and $z = -8$, respectively). No interaction peaks were detected.

5.4 Anthracenes

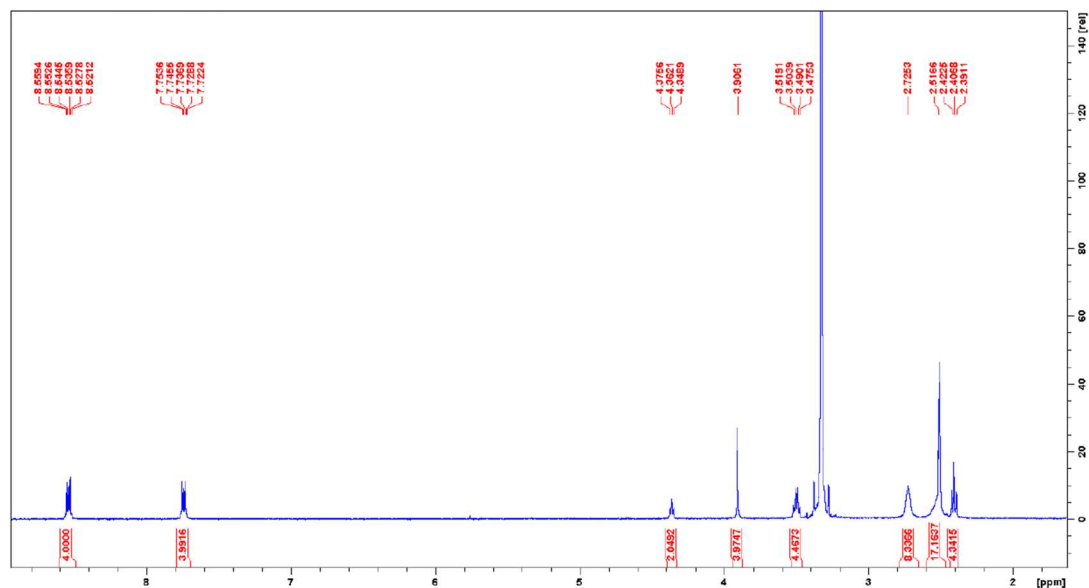


Figure 5.4.1 $^1\text{H-NMR}$ spectrum of compound **OAF89**. Note: the signal of 4 x CH_2 protons (piperazine, bs) partially overlap with the signal of the solvent (d6-DMSO).

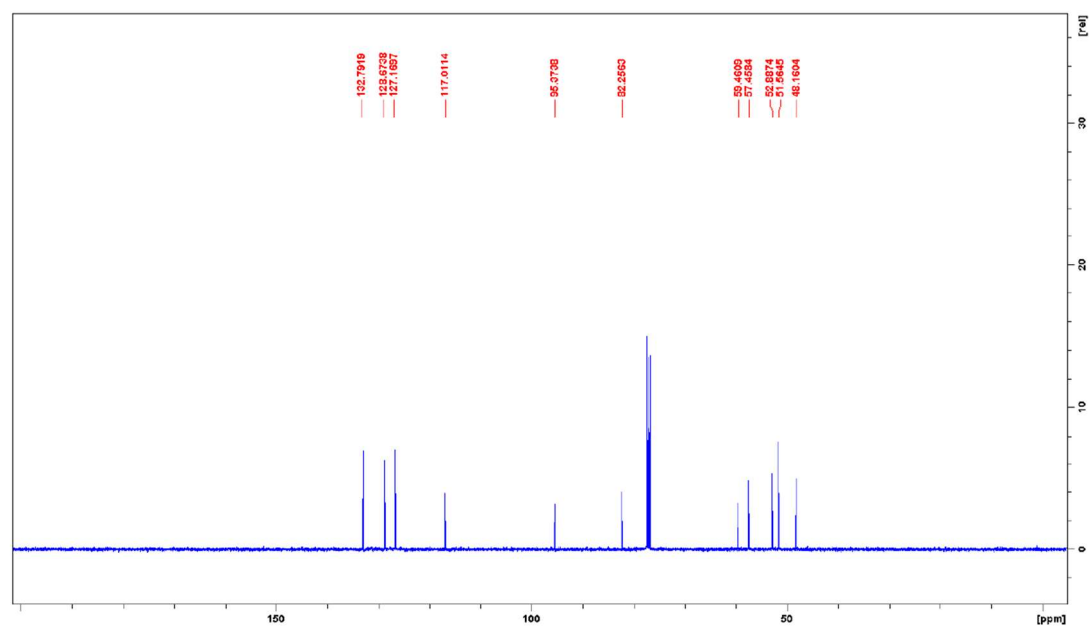


Figure 5.4.2 $^{13}\text{C-NMR}$ spectrum of compound **OAF89** (CDCl_3).

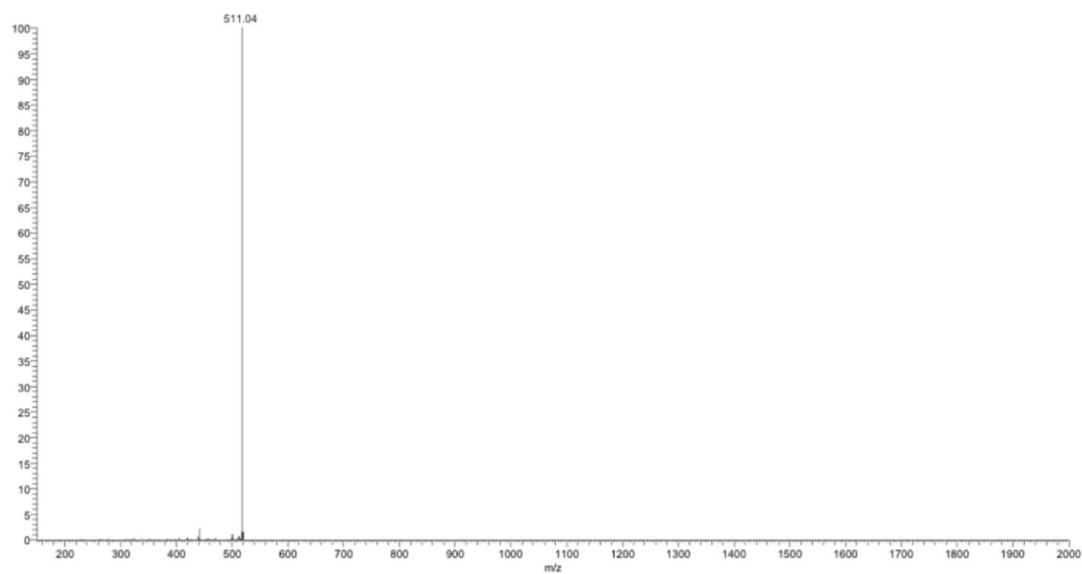


Figure 5.4.3. ESI-MS spectrum of compound **OAF89** (positive ionization mode).

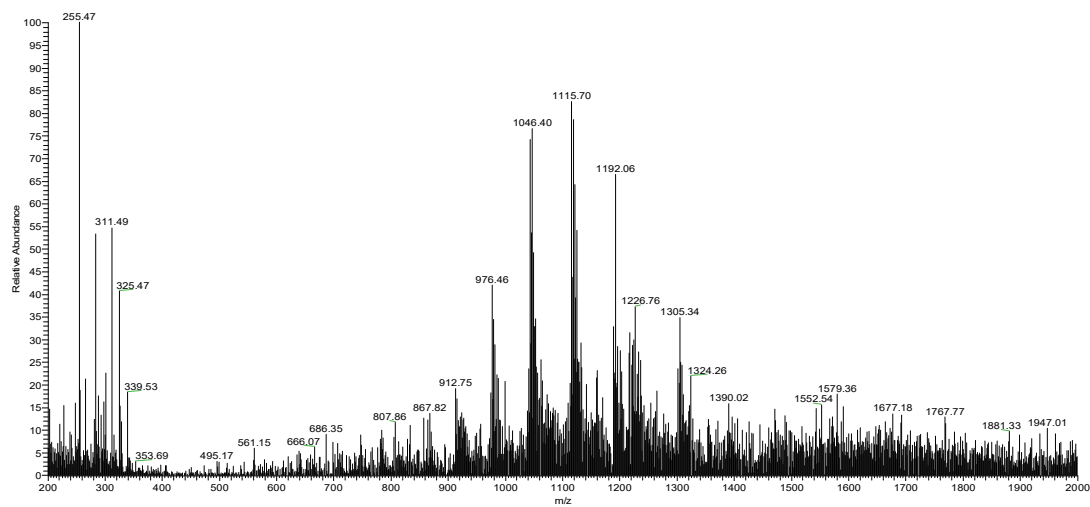


Figure 5.4.4. ESI-MS binding experiment of compound **OAF89** with G4 DNA. Unbound G4 DNA ($m/z = 1046.26$, $z = -7$), unbound ligand ($m/z = 255.47$, $z = -2$), ligand/G4 complex with 1:1 stoichiometry ($m/z = 1119.23$, $z = -7$) and ligand/G4 complex with 2:1 stoichiometry ($m/z = 1191.95$, $z = -7$) were detected.

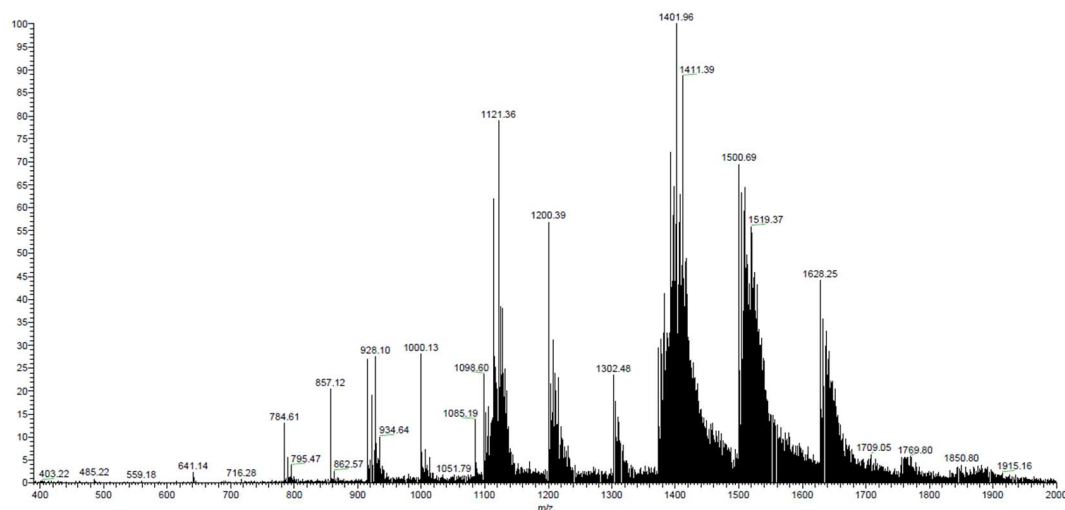


Figure 5.4.5. ESI-MS binding experiment of compound **OAF89** with dsDNA. Unbound dsDNA ($m/z = 1121.36$, $z = -9$), ligand/dsDNA complex with 1:1 stoichiometry ($m/z = 1200.39$, $z = -9$) and ligand/dsDNA complex with 2:1 stoichiometry ($m/z = 1302.48$, $z = -9$) were detected. The same behaviour was observed in the $z = -8$ charge state.

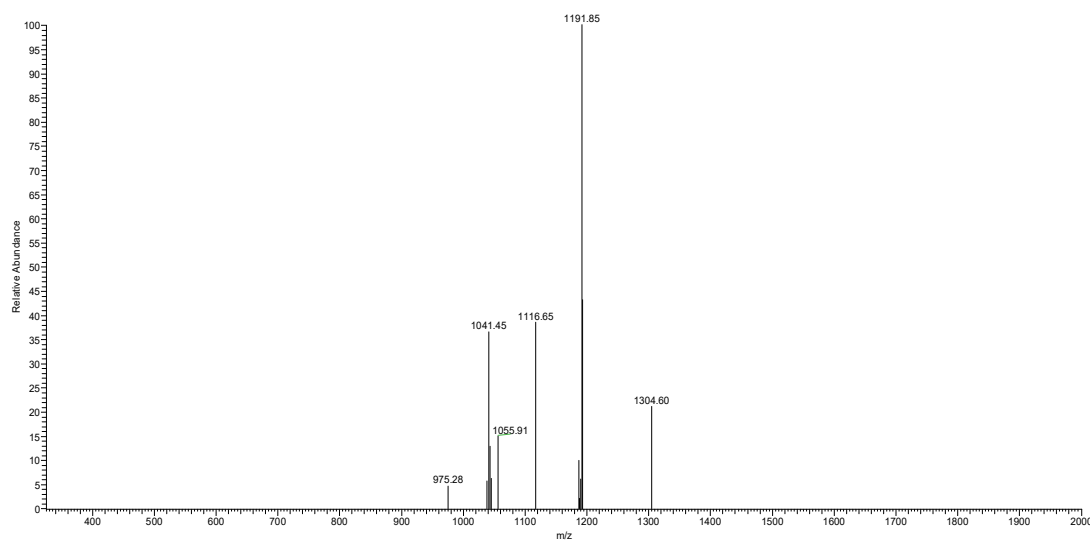
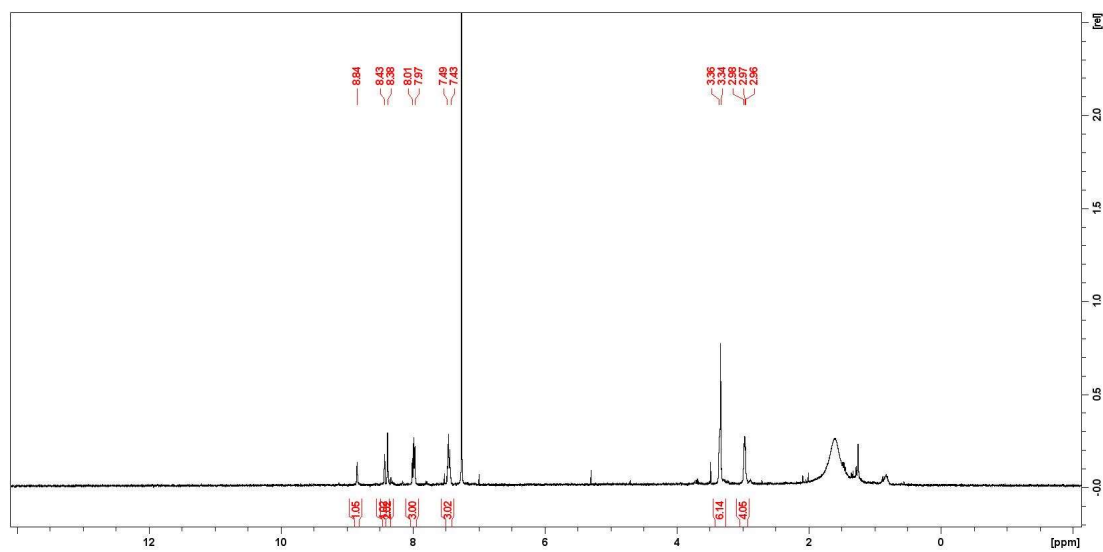
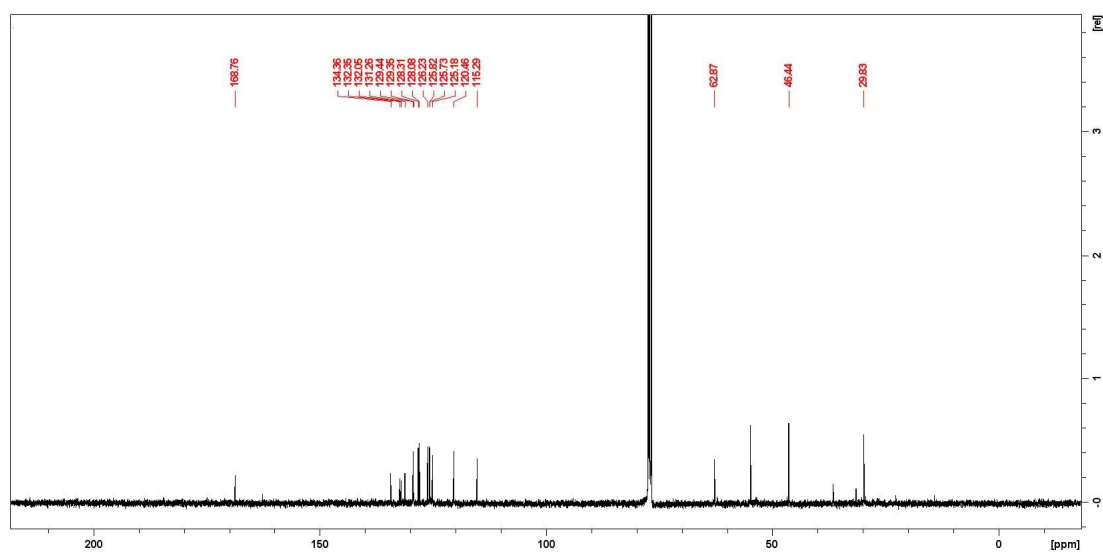


Figure 5.4.6. CID fragmentation spectrum of the ligand/G4 complex with 2:1 stoichiometry (normalized collision energy = 12). The fragmentation of the 2:1 ligand/G4 complex ($m/z = 1191.85$, $z = -7$) promoted the formation of the peaks corresponding to the 1:1 complex ($m/z = 1116.65$, $z = -7$) and to the unbound DNA ($m/z = 1041.45$, $z = -7$).

5.5 Dimeric PEG-linked scaffolds

Figure 5.5.1. ^1H -NMR spectrum of compound ANAp.Figure 5.5.2. ^{13}C -NMR spectrum of compound ANAp.

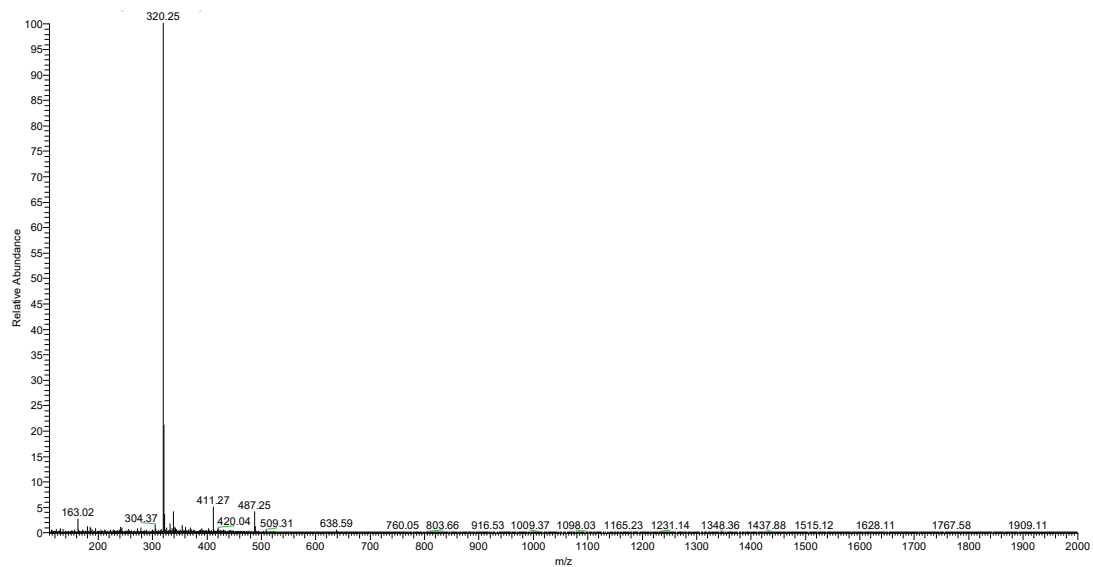


Figure 5.5.3. ESI-MS spectrum of compound ANAp.

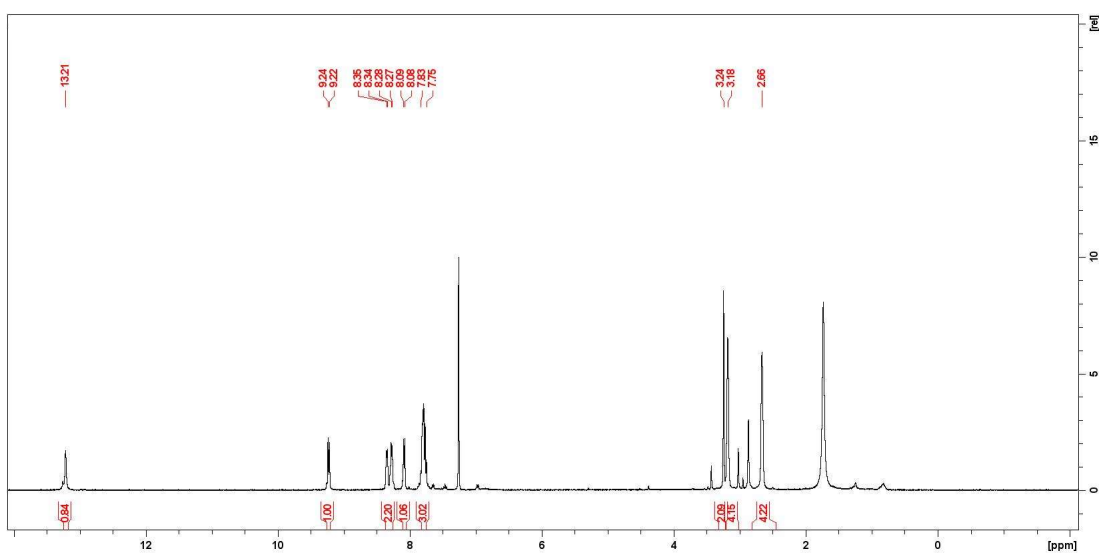
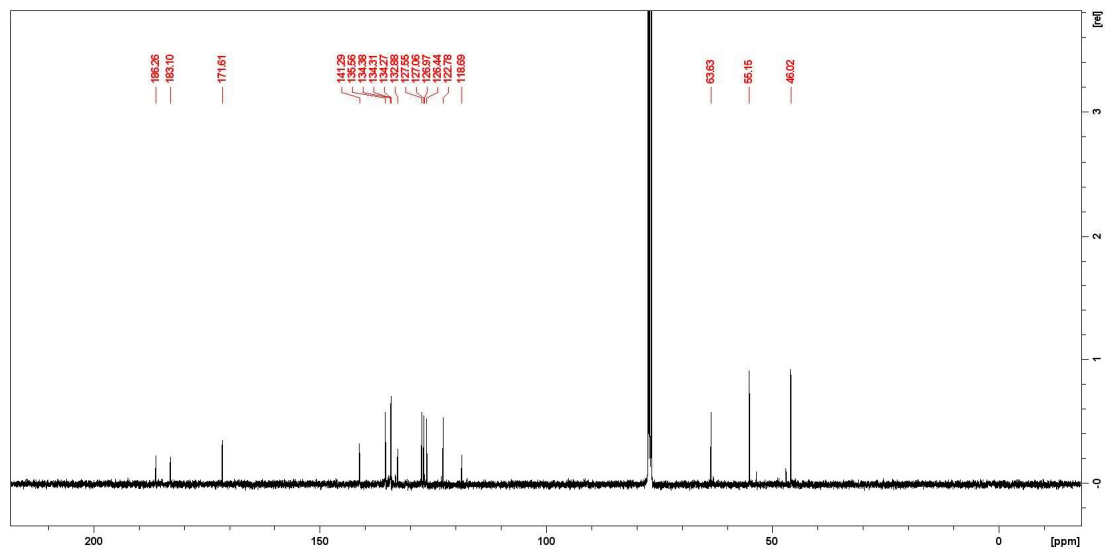
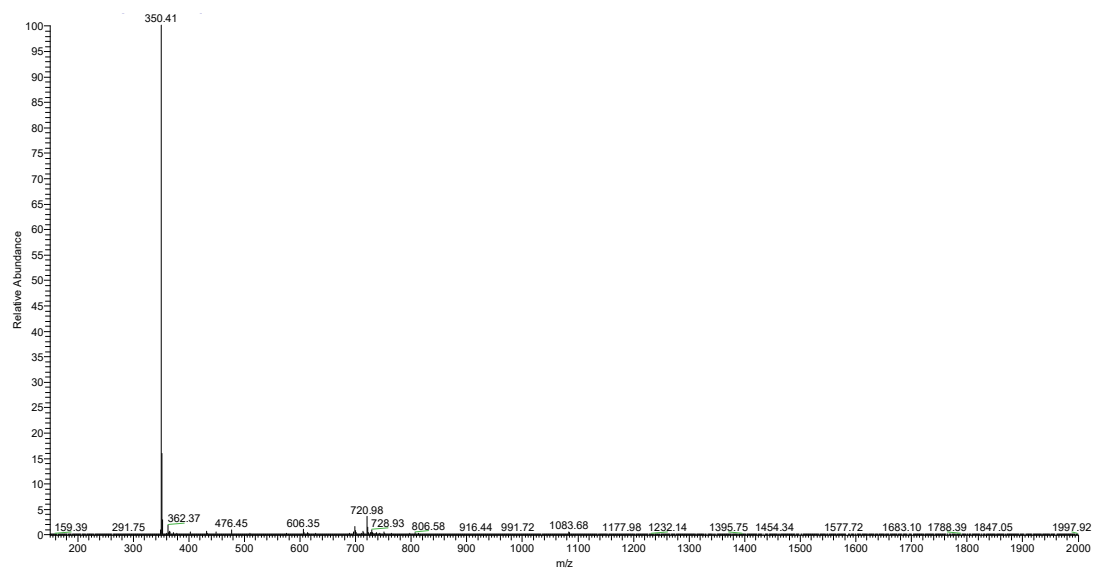


Figure 5.5.4. ¹H-NMR spectrum of compound AQAp.

Figure 5.5.5. ^{13}C -NMR spectrum of compound **AQAp**.Figure 5.5.6. ESI-MS spectrum of compound **AQAp**.

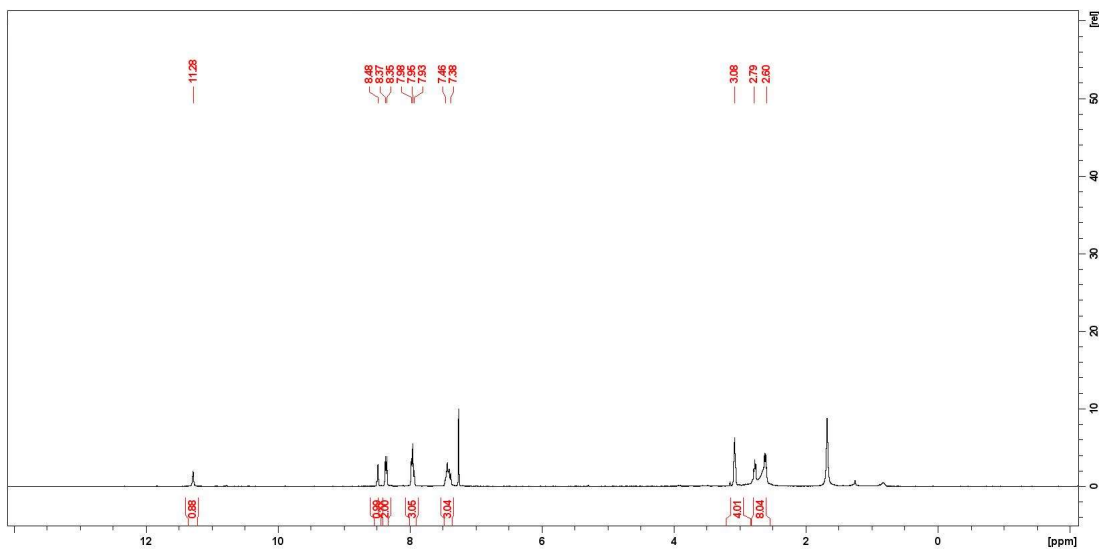


Figure 5.5.7. ^1H -NMR spectrum of compound ANPp.

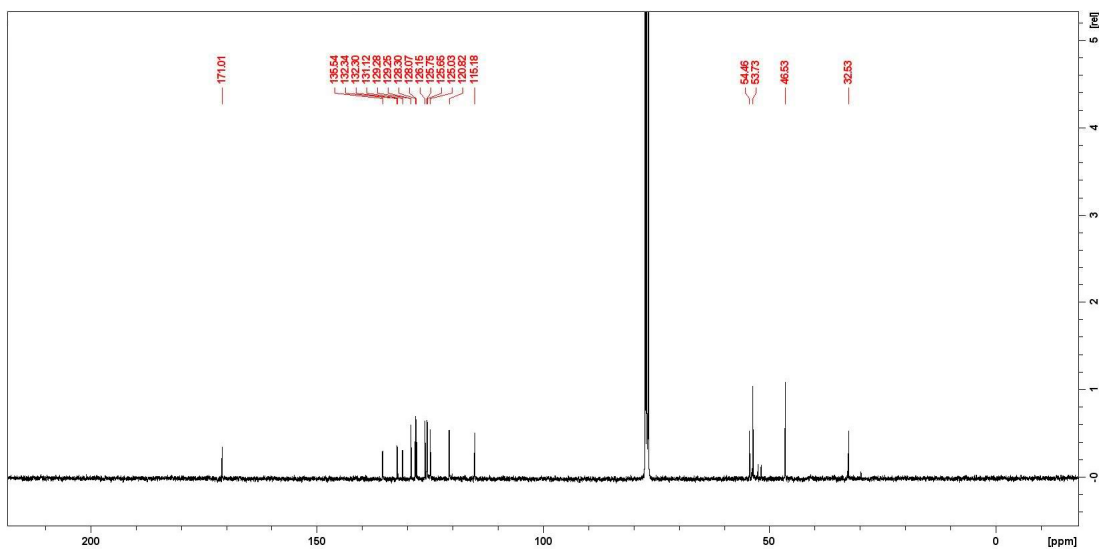


Figure 5.5.8. ^{13}C -NMR spectrum of compound ANPp.

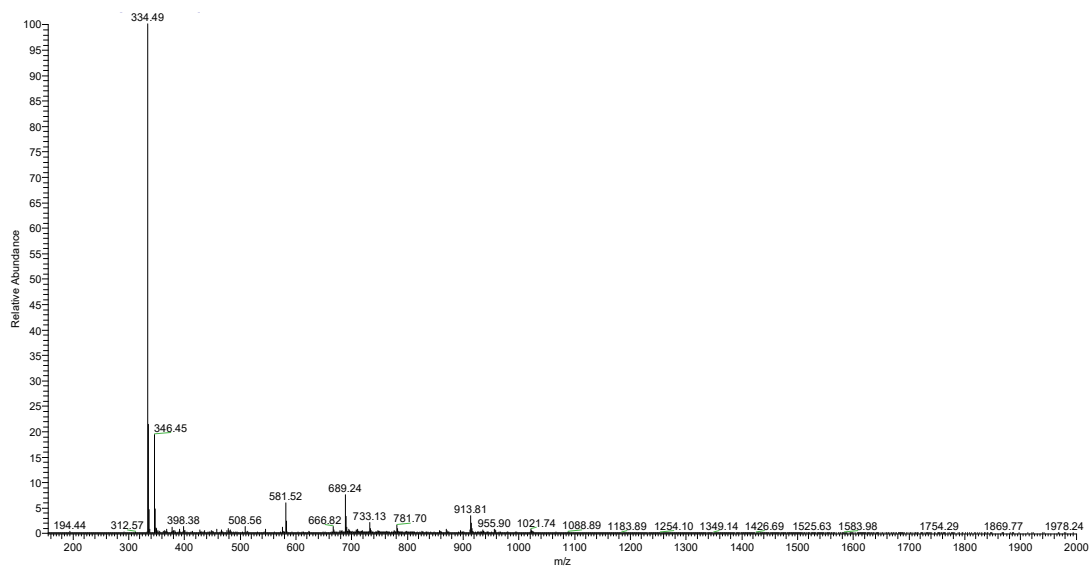
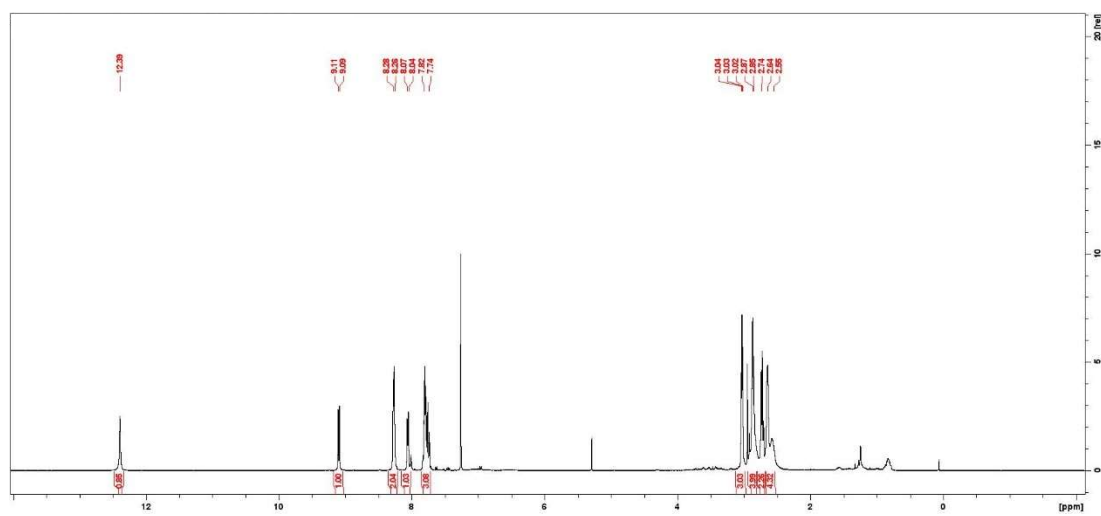


Figure 5.5.9. ESI-MS spectrum of compound ANPp.

Figure 5.5.10. ¹H-NMR spectrum of compound AQPp.

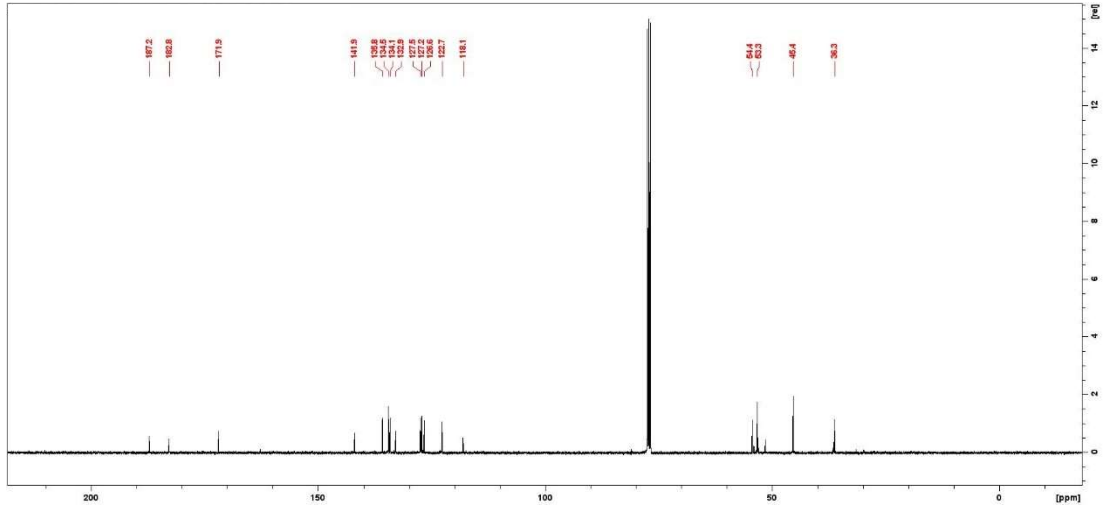


Figure 5.5.11. ¹³C-NMR spectrum of compound AQPp.

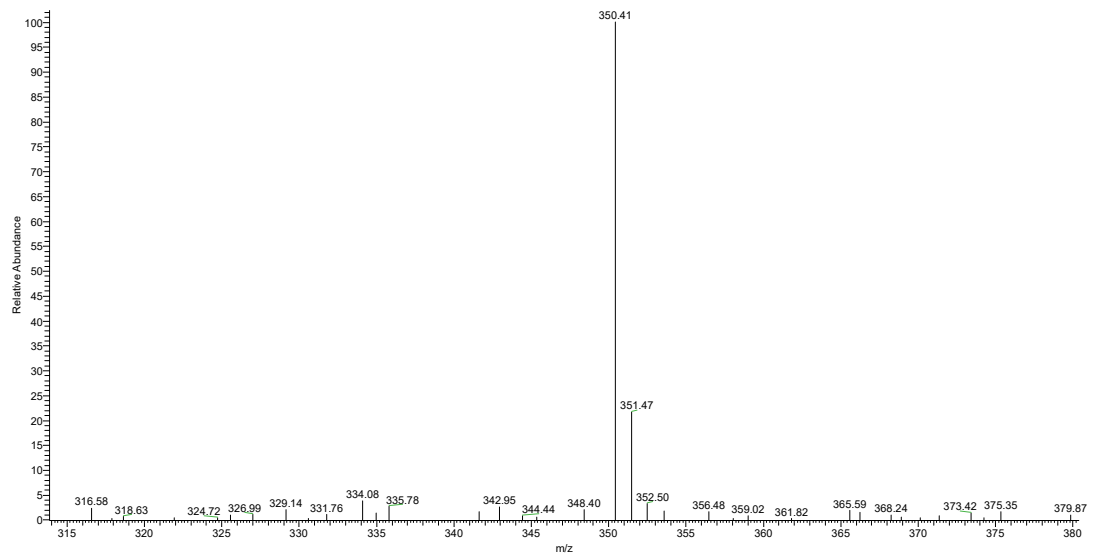
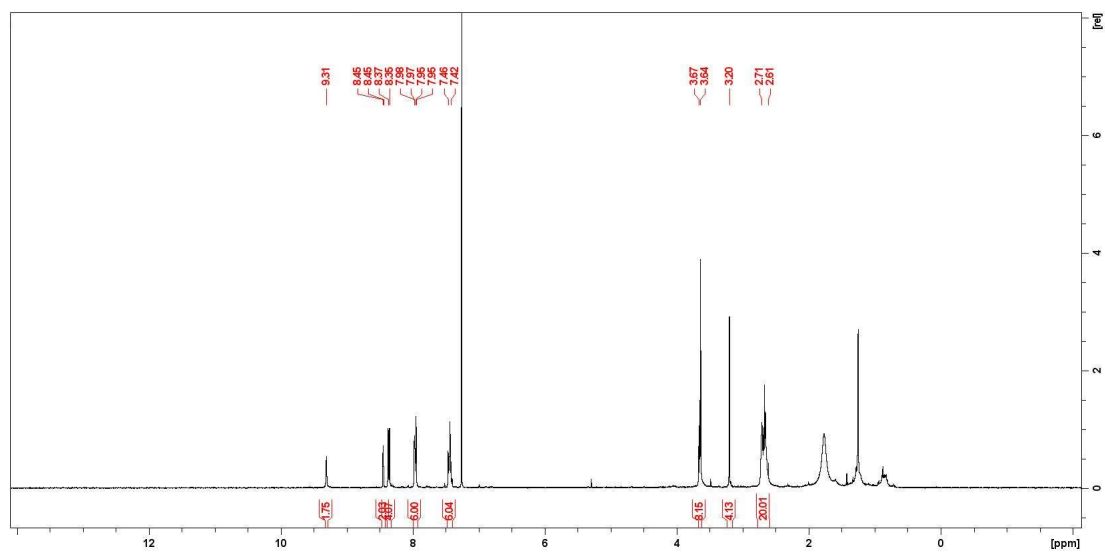
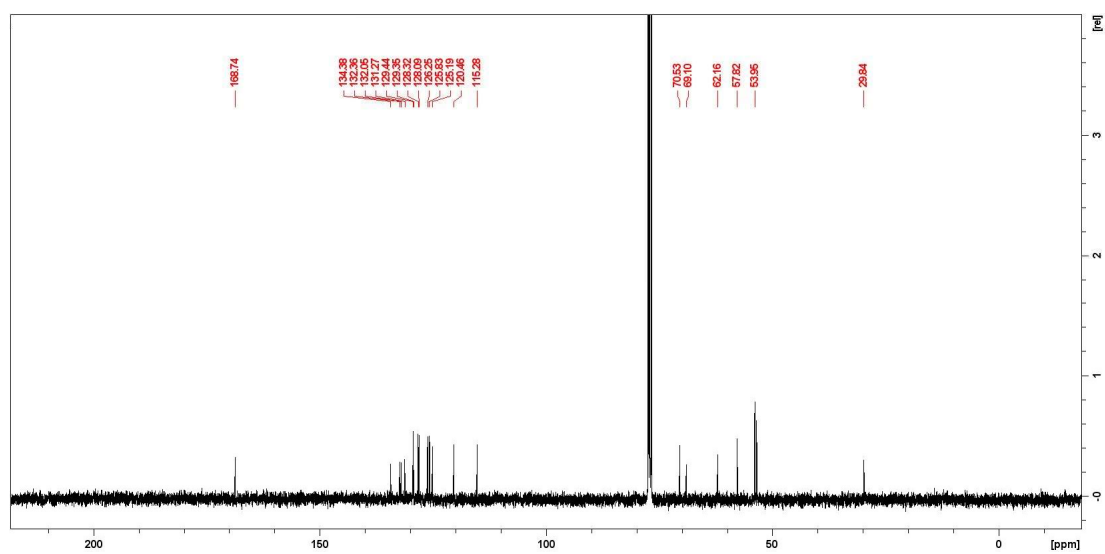


Figure 5.5.12. ESI-MS spectrum of compound AQPp.

Figure 5.5.13. ¹H-NMR spectrum of compound ANA3.Figure 5.5.14. ¹³C-NMR spectrum of compound ANA3.

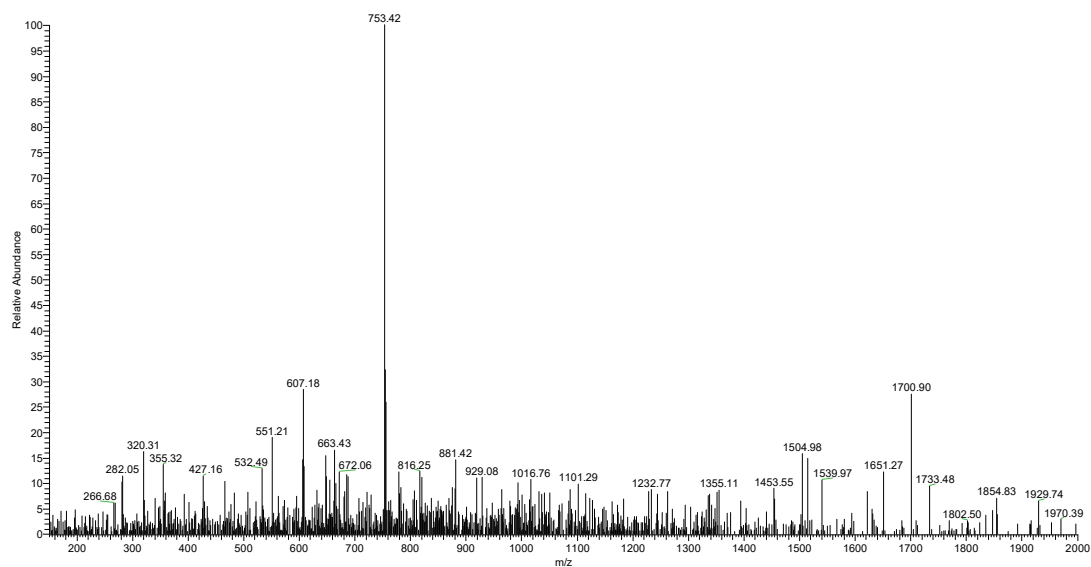


Figure 5.5.15. ESI-MS spectrum of compound ANA3.

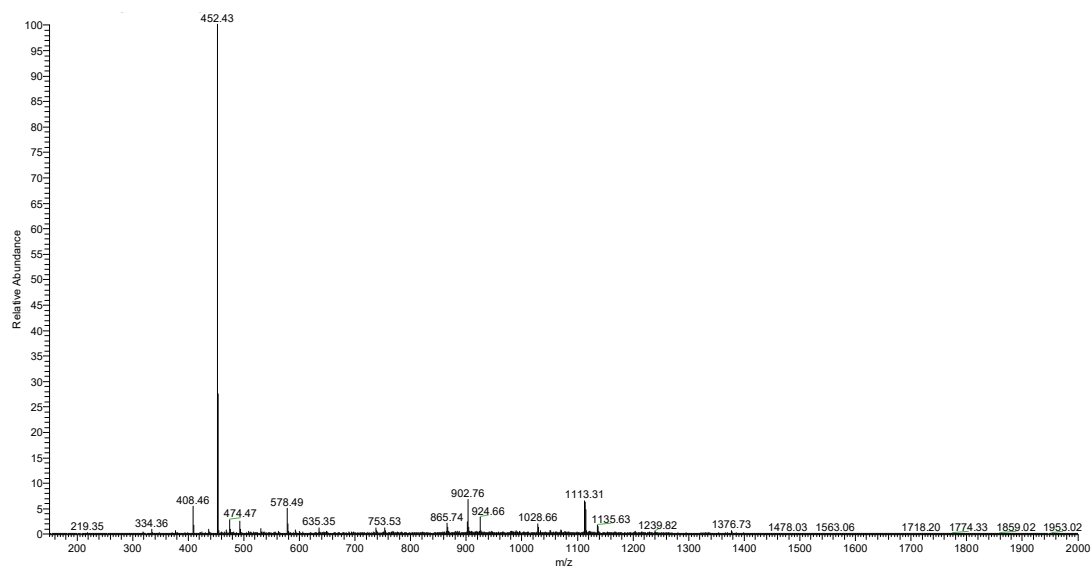
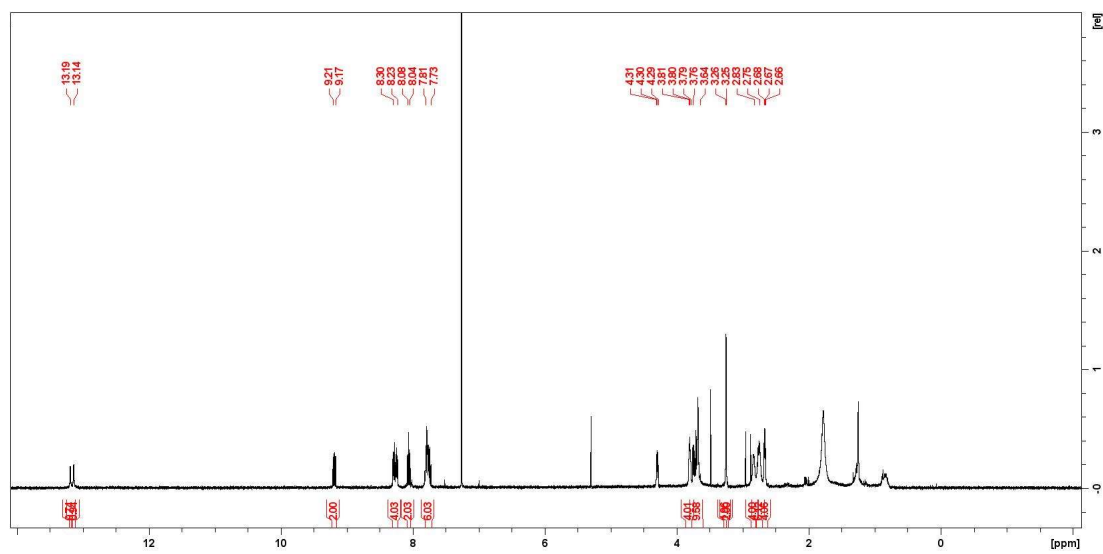
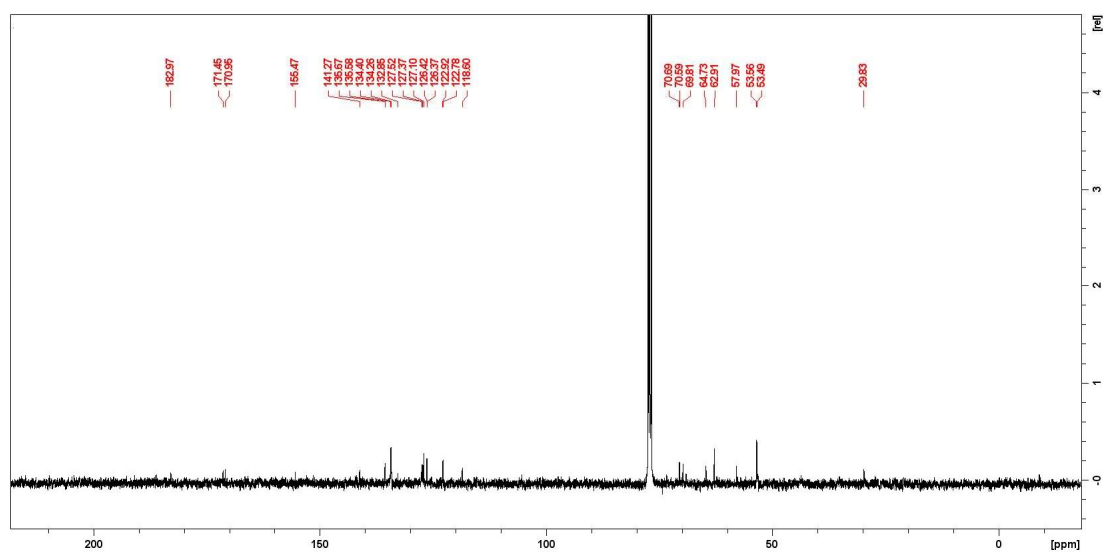


Figure 5.5.16. ESI-MS spectrum of compound ANA3m.

Figure 5.5.17. ^1H -NMR spectrum of compound **AQA3**.Figure 5.5.18. ^{13}C -NMR spectrum of compound **AQA3**.

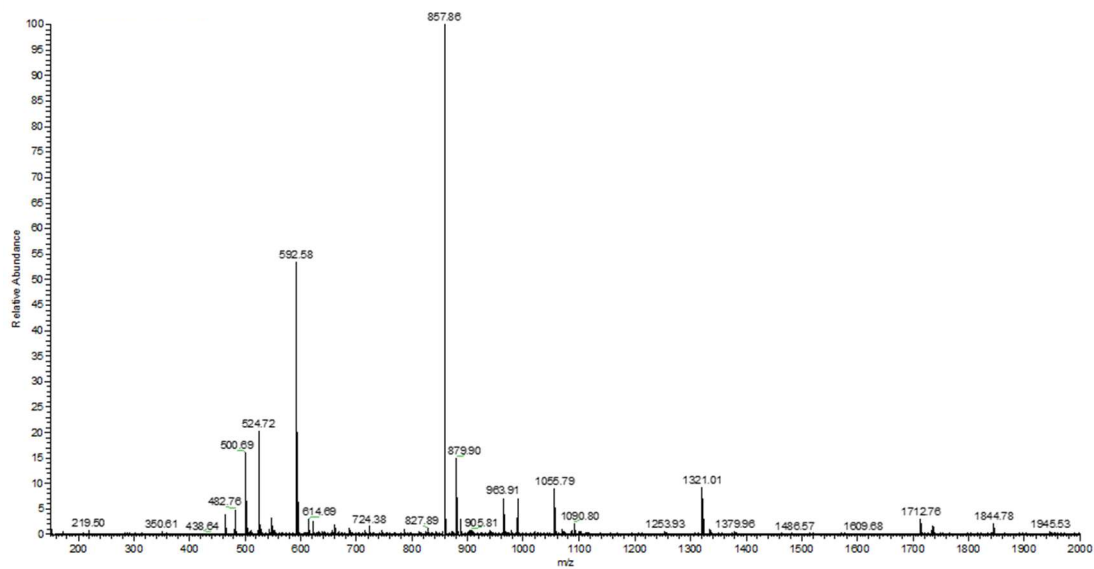


Figure 5.5.19. ES-MS spectrum of compound AQA3.

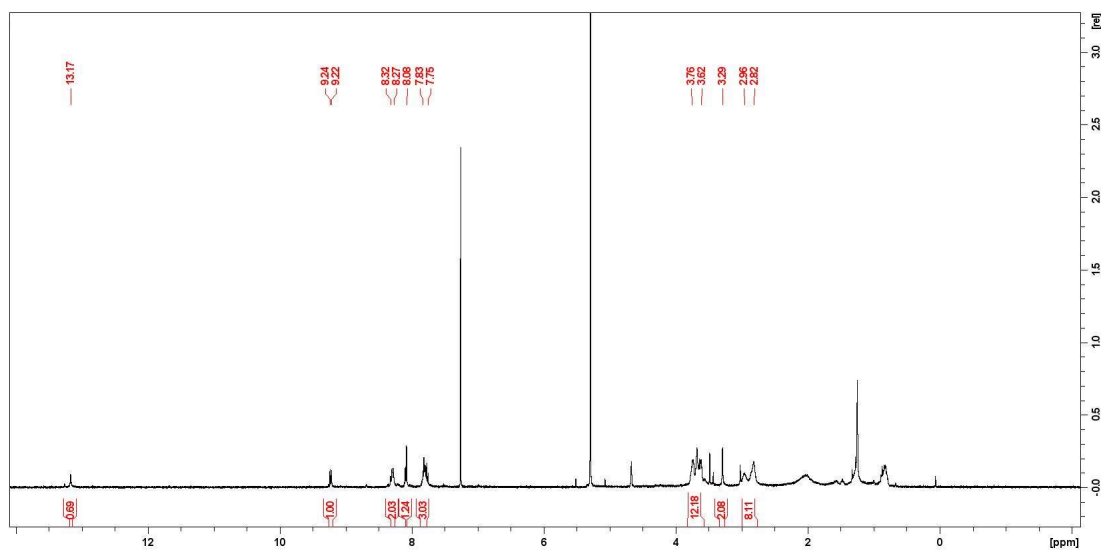
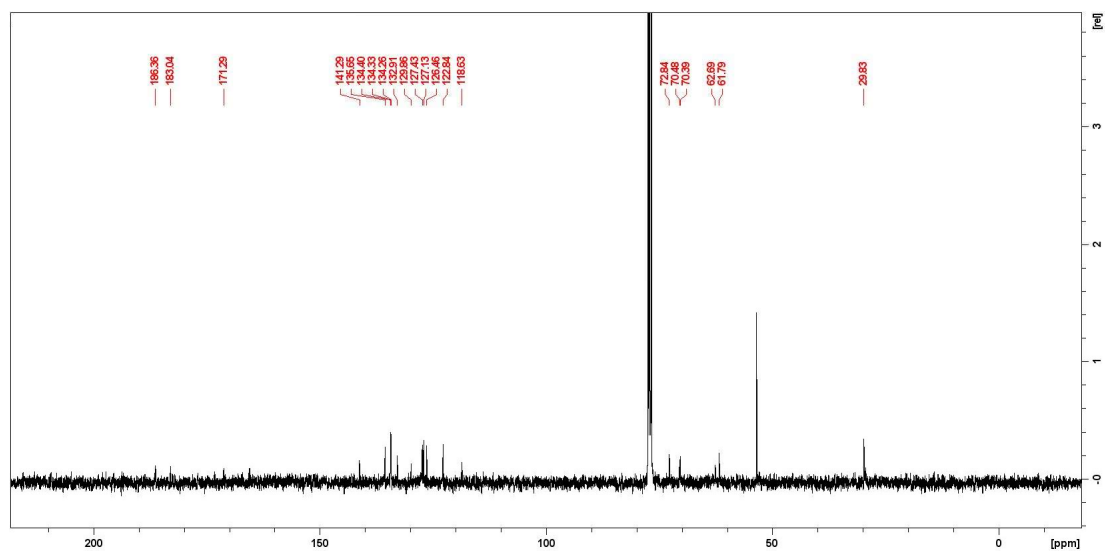
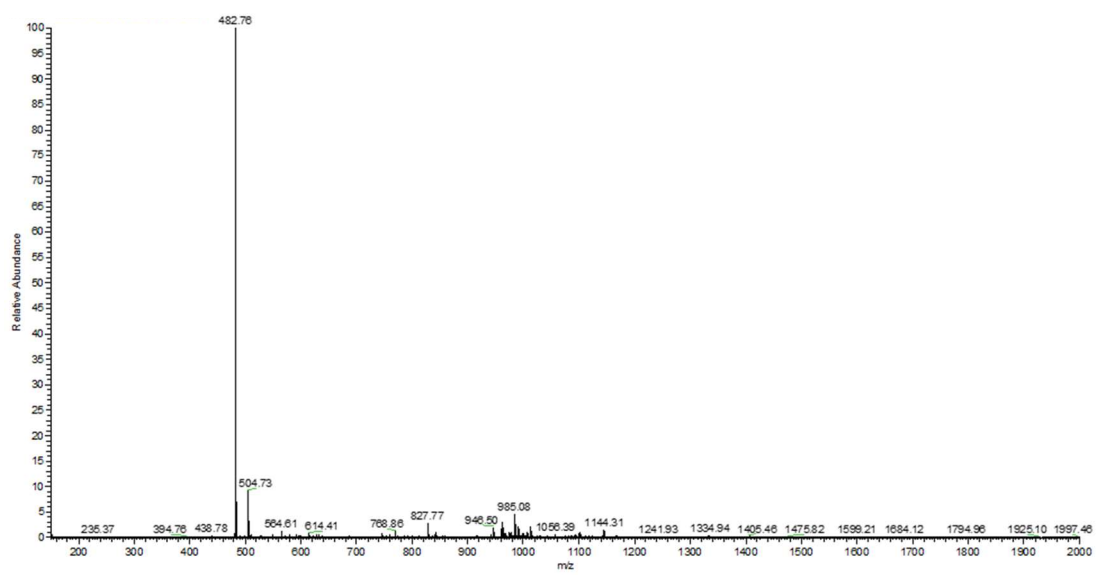


Figure 5.5.20. ¹H-NMR spectrum of compound AQA3m.

Figure 5.5.21. ^{13}C -NMR spectrum of compound **AQA3m**.Figure 5.5.22. ESI-MS spectrum of compound **AQA3m**.

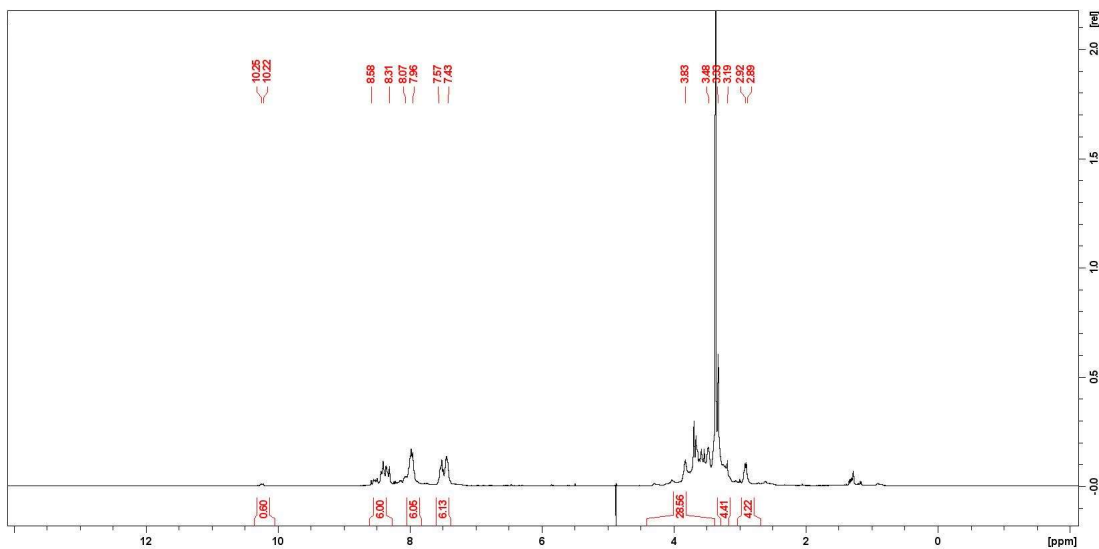


Figure 5.5.23. ¹H-NMR spectrum of compound ANP3.

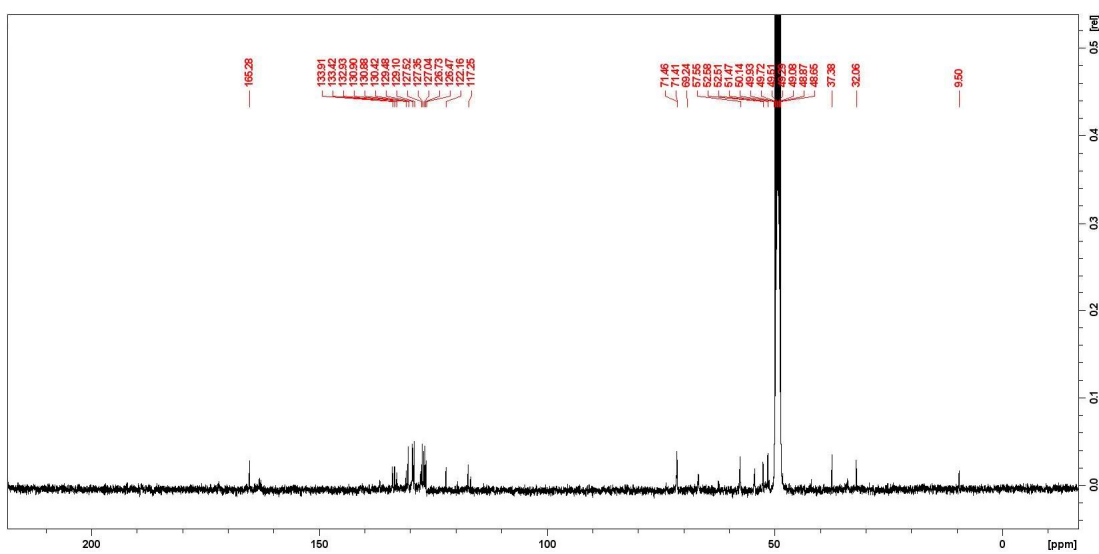


Figure 5.5.24. ¹³C-NMR spectrum of compound ANP3.

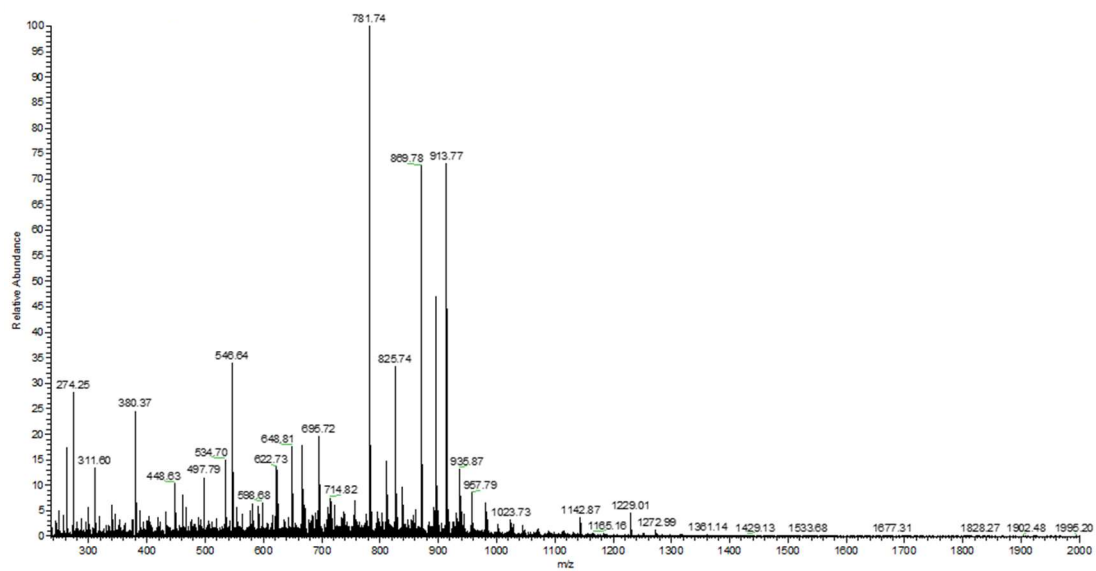
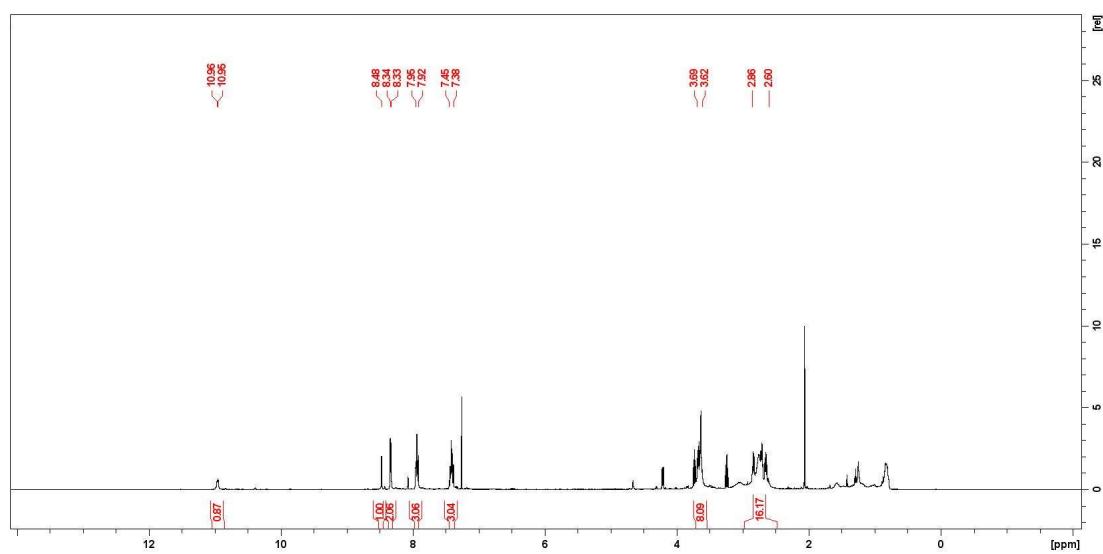


Figure 5.5.25. ESI-MS spectrum of compound ANP3.

Figure 5.5.26. ¹H-NMR spectrum of compound ANP3m.

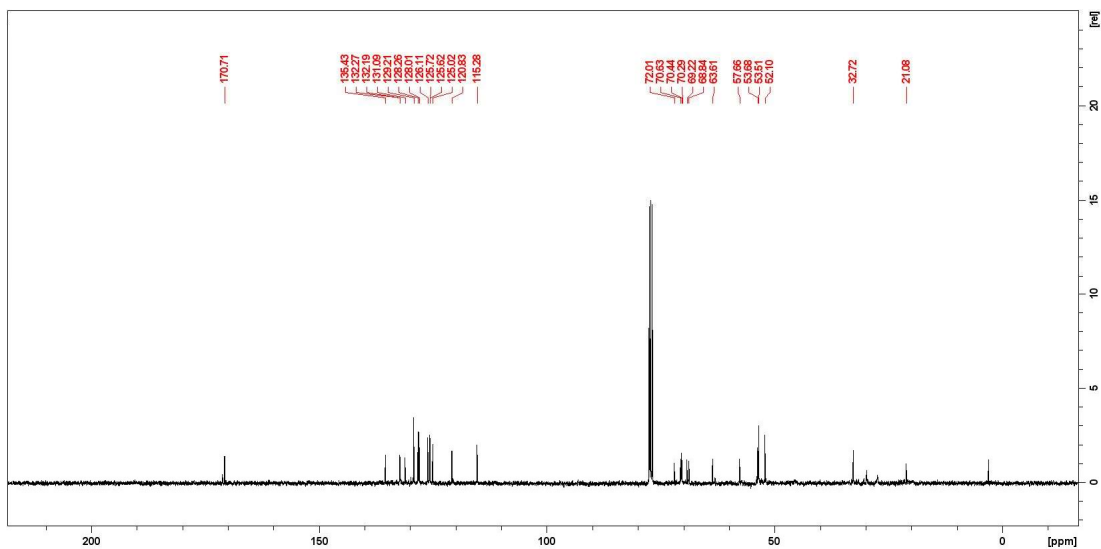


Figure 5.5.27. ¹³C-NMR spectrum of compound ANP3m.

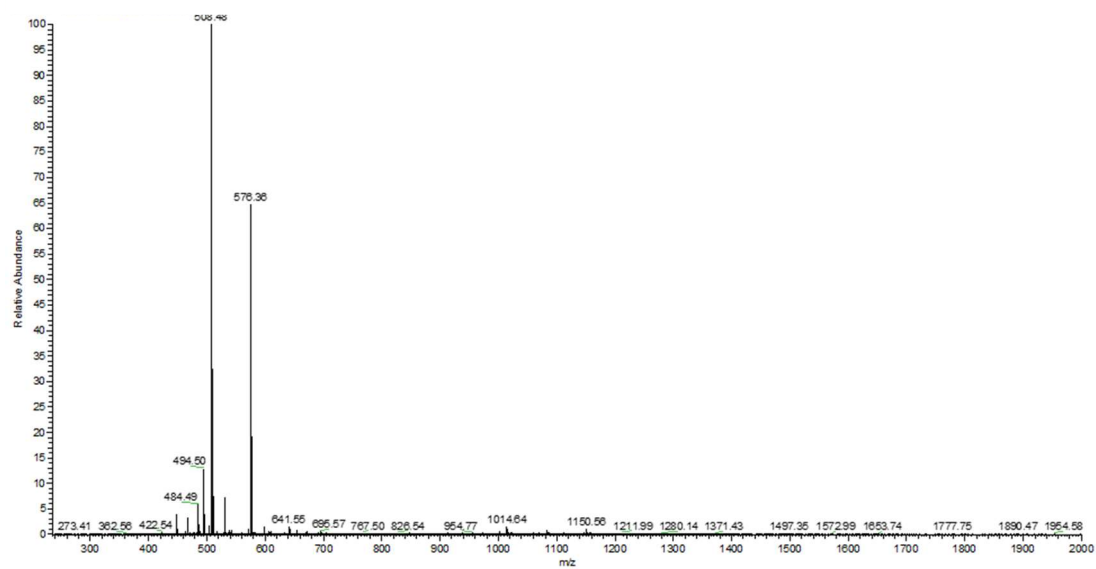


Figure 5.5.28. ESI-MS spectrum of compound ANP3m.

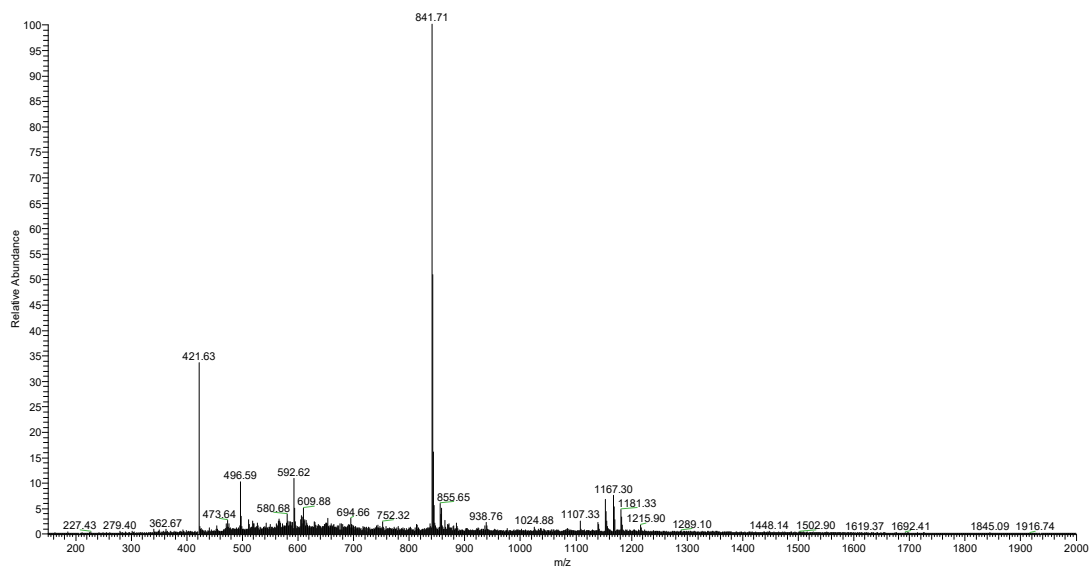
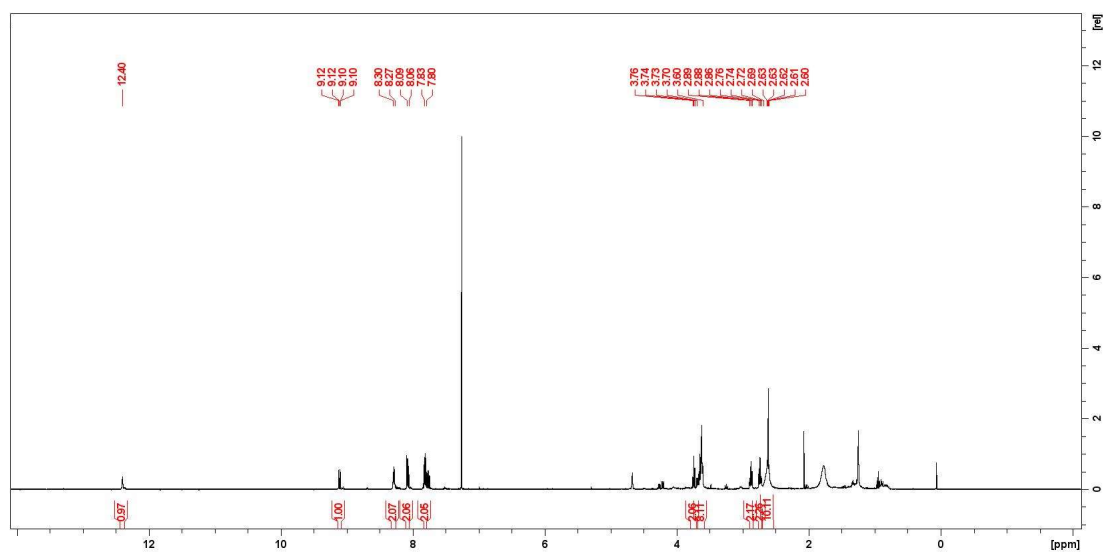


Figure 5.5.29. ESI-MS spectrum of compound AQP3.

Figure 5.5.30. ¹H-NMR spectrum of compound AQP3m.

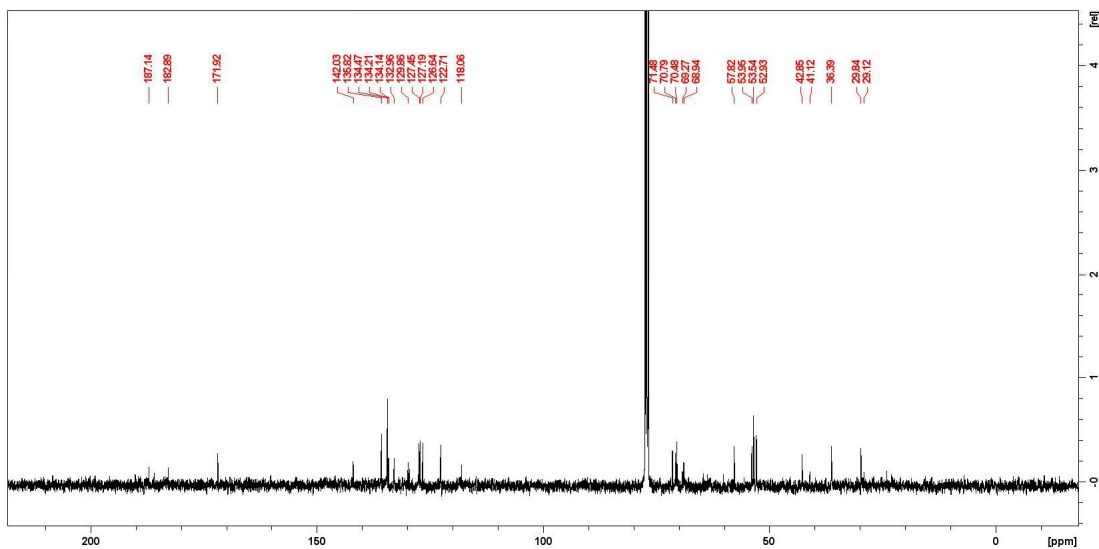


Figure 5.5.31. ^{13}C -NMR spectrum of compound **AQP3m**.

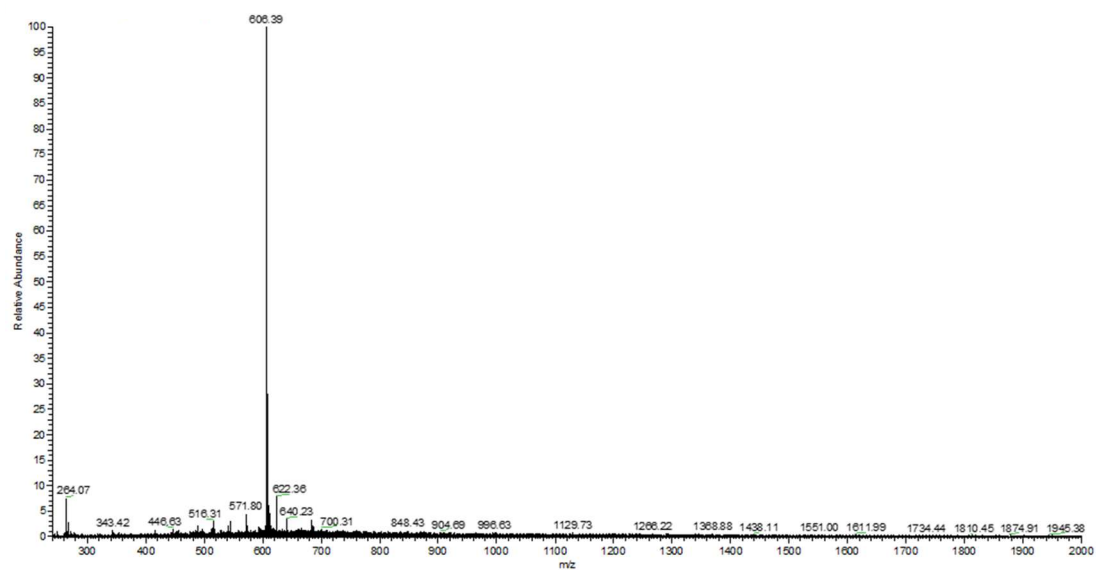


Figure 5.5.32. ESI-MS spectrum of compound **AQP3m**.

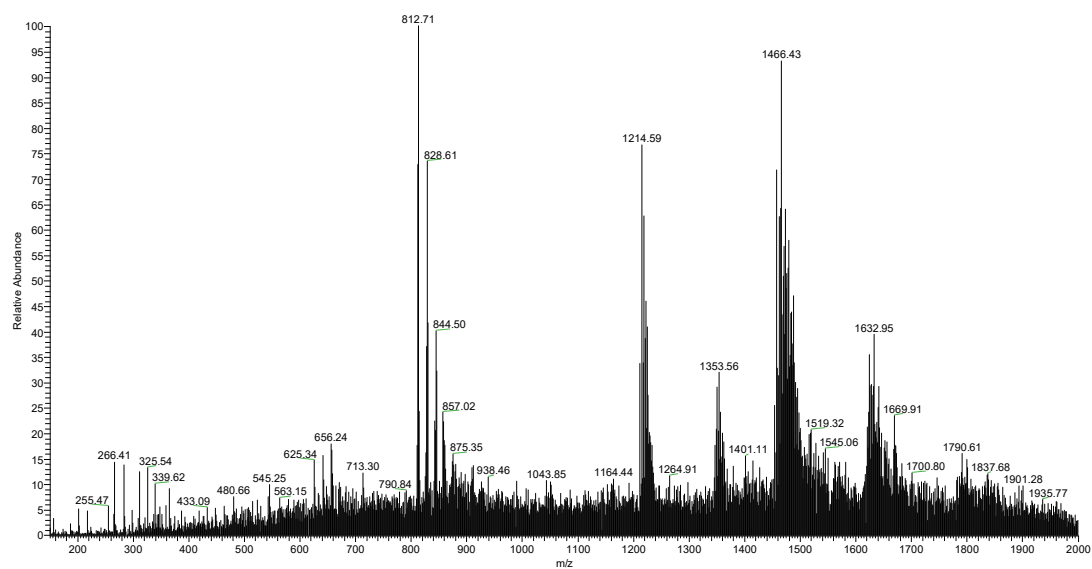


Figure 5.5.33 ESI-MS binding experiment of compound **AQA3** with G4. Unbound G4 ($m/z = 1214.61$, $z = -6$; $m/z = 1462.04$, $z = -5$) and ligand/G4 complex ($m/z = 1353.62$, $z = -6$; $m/z = 1632.97$, $z = -5$) were detected.

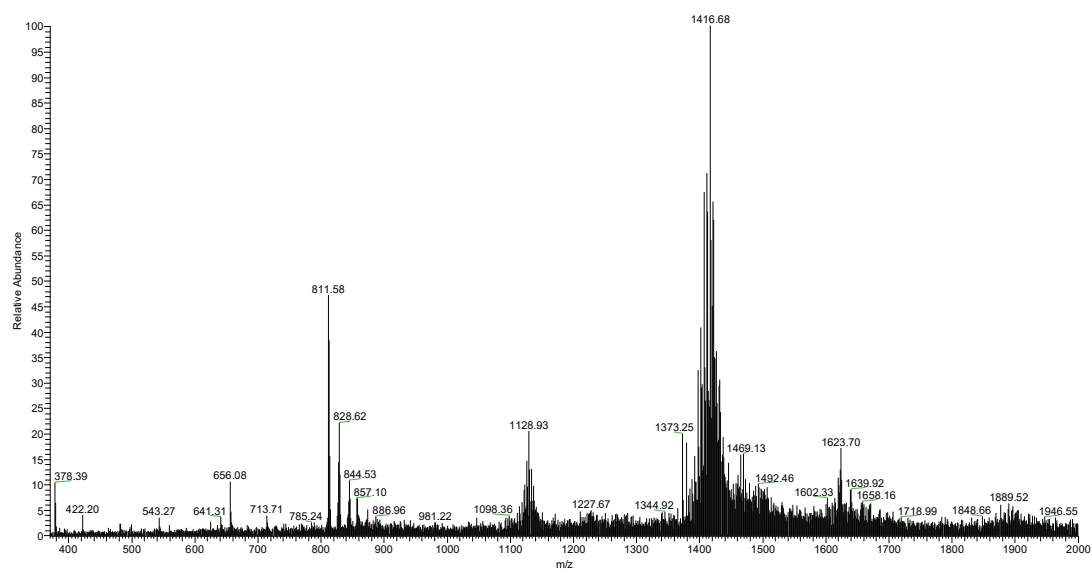


Figure 5.5.34 ESI-MS binding experiment of compound **AQA3** with dsDNA. Unbound dsDNA ($m/z = 1128.89$, $z = -10$; $m/z = 1416.73$, $z = -8$) were detected. No interaction peak was found.

5.6 Seleno-containing compounds

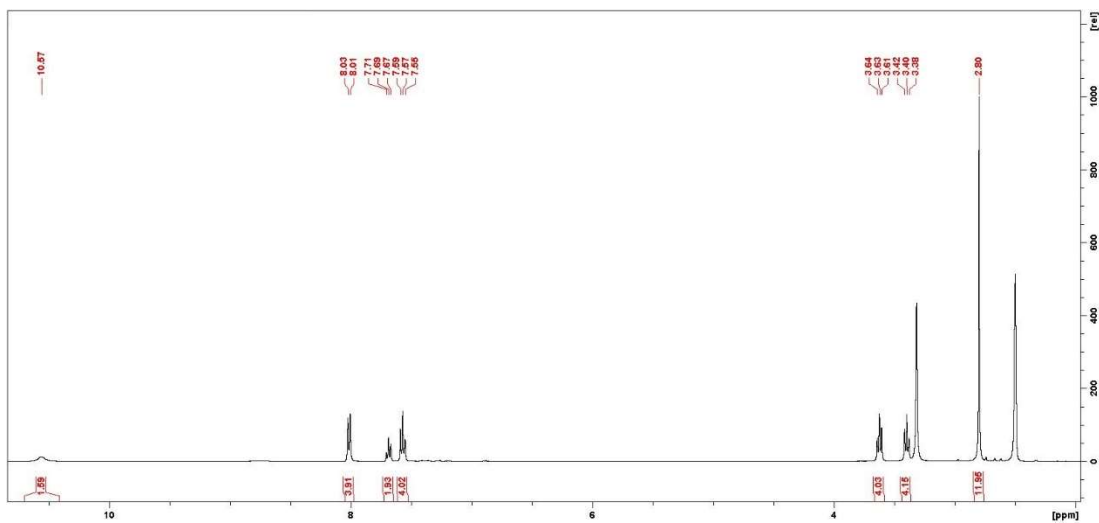


Figure 5.6.1. ^1H -NMR spectrum of compound **Se2**.

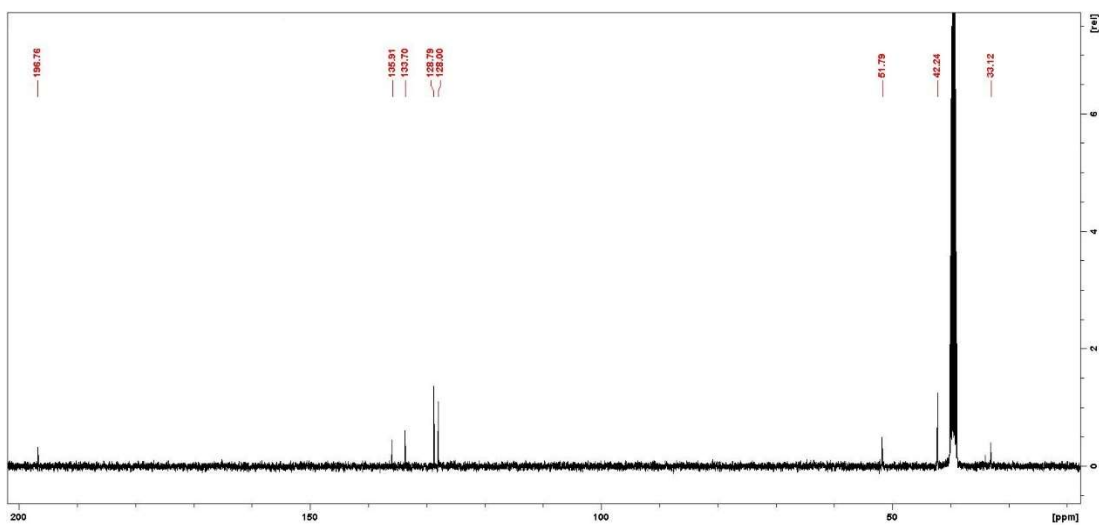


Figure 5.6.2. ^{13}C -NMR spectrum of compound **Se2**.

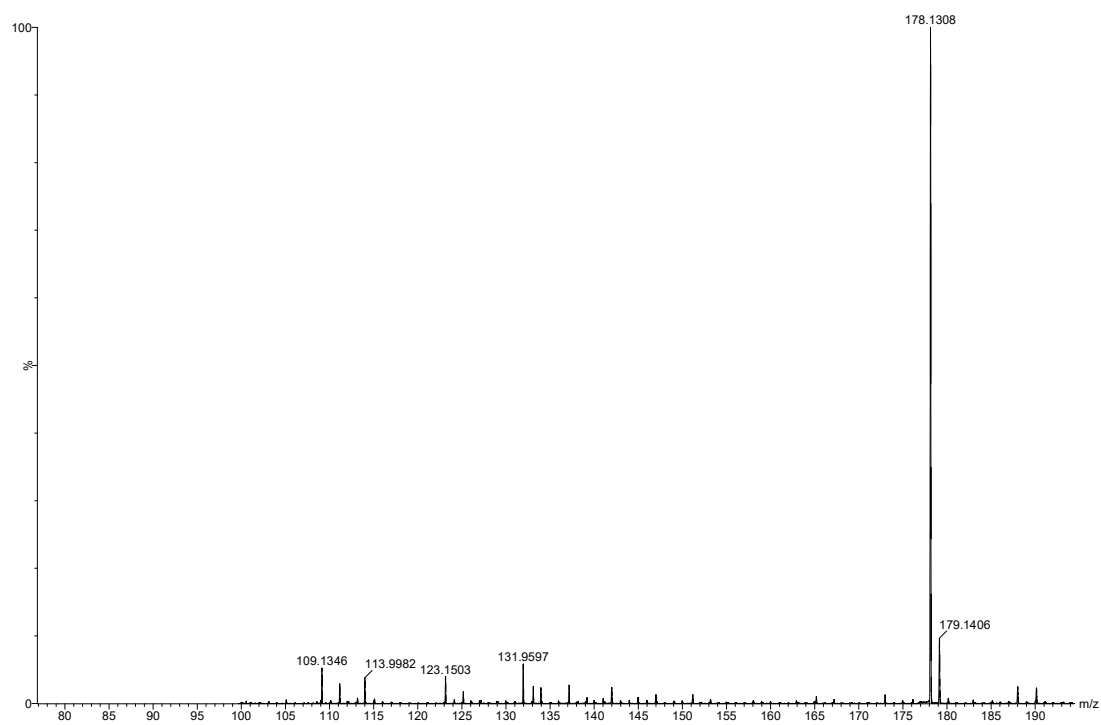
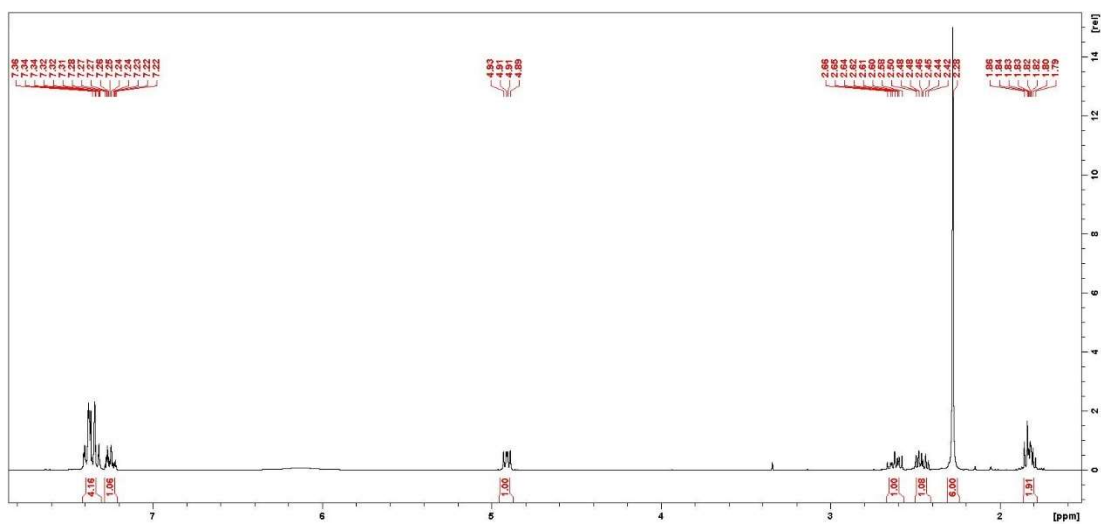


Figure 5.6.3. ESI-MS spectrum of compound Se2.

Figure 5.6.4. ¹H-NMR spectrum of compound Se3.

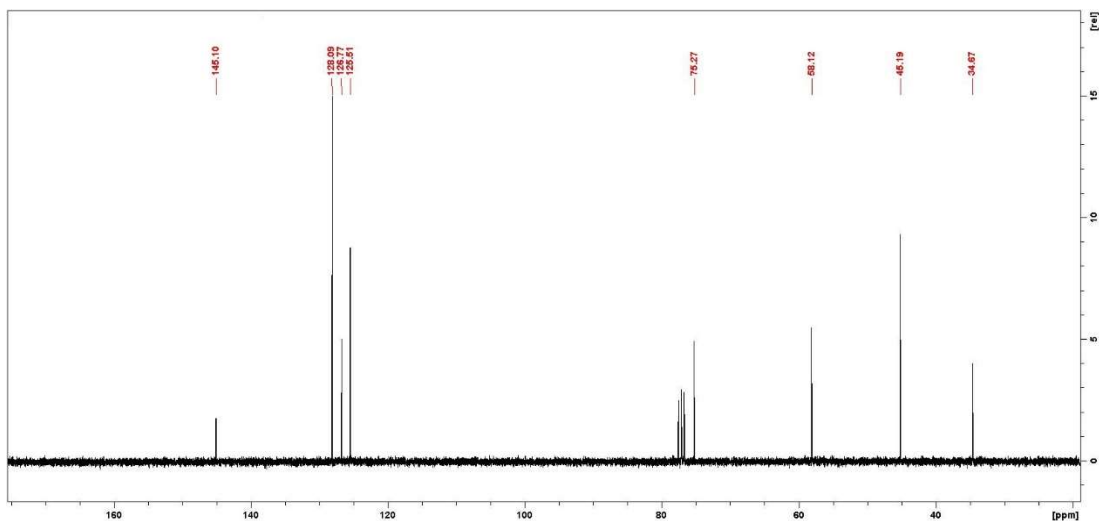


Figure 5.6.5. ¹³C-NMR spectrum of compound Se3.

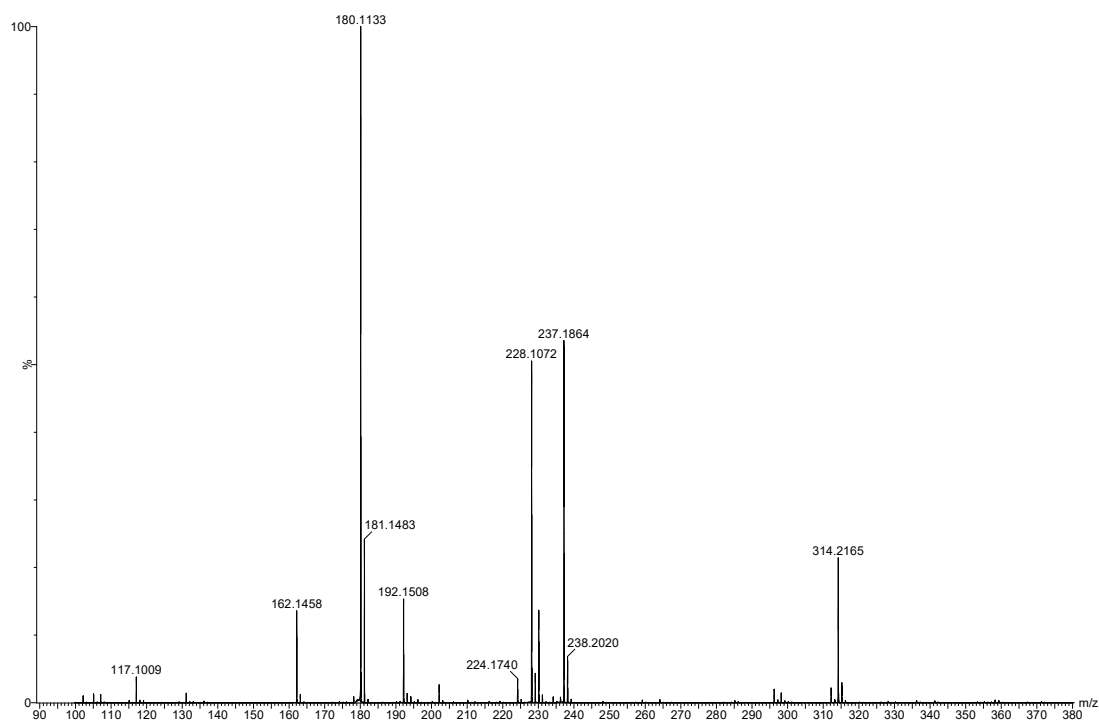
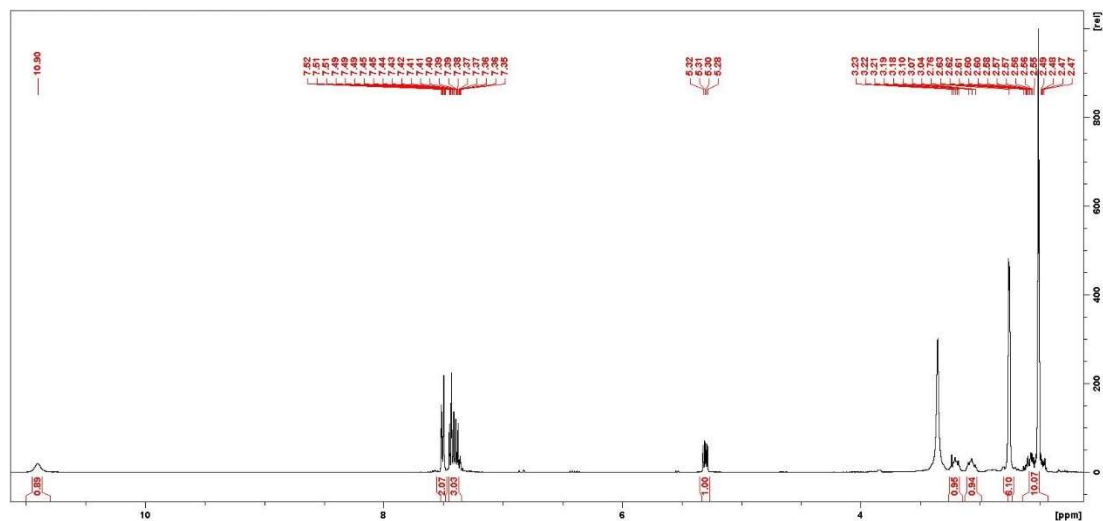
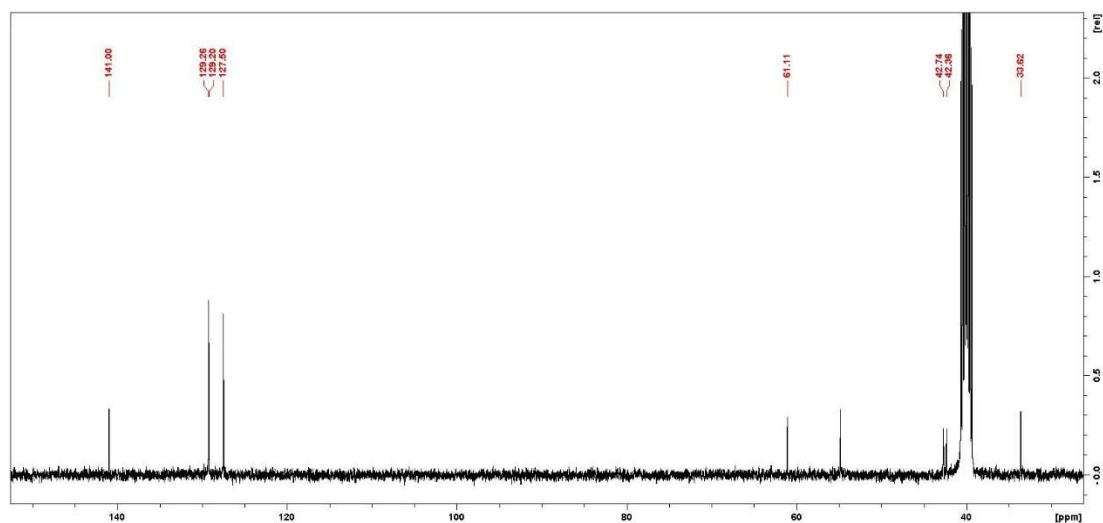


Figure 5.6.6. ESI-MS spectrum of compound Se3.

Figure 5.6.7. ^1H -NMR spectrum of compound **Se4**.Figure 5.6.8. ^{13}C -NMR spectrum of compound **Se4**.

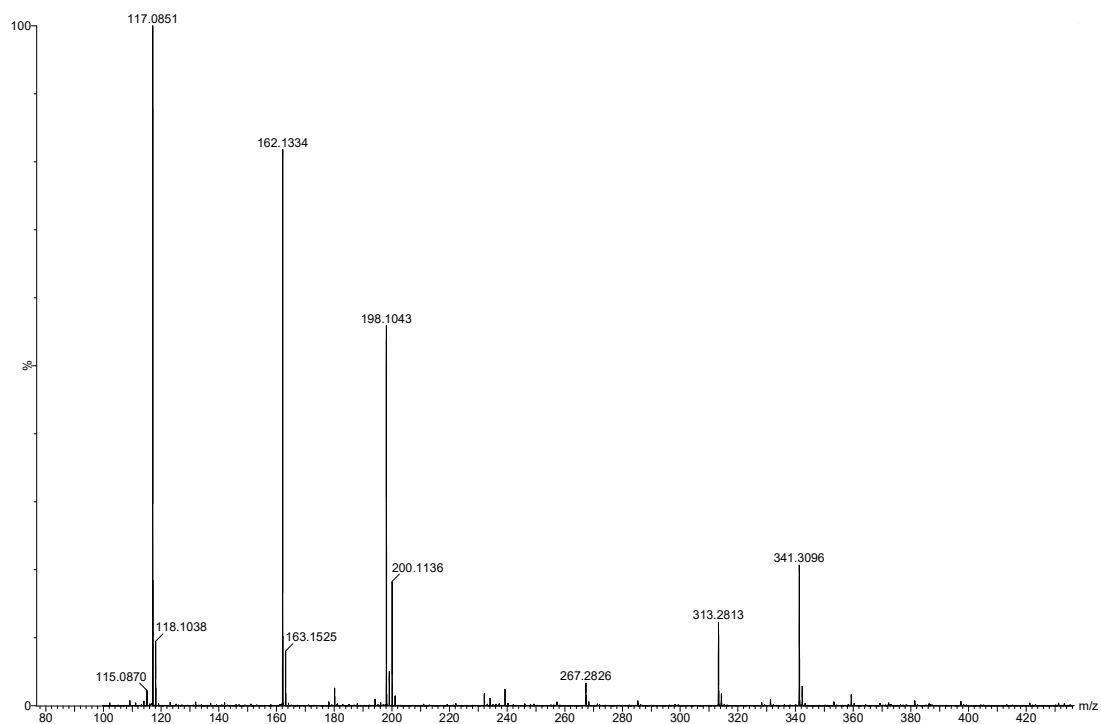


Figure 5.6.9. ESI-MS spectrum of compound **Se4**.

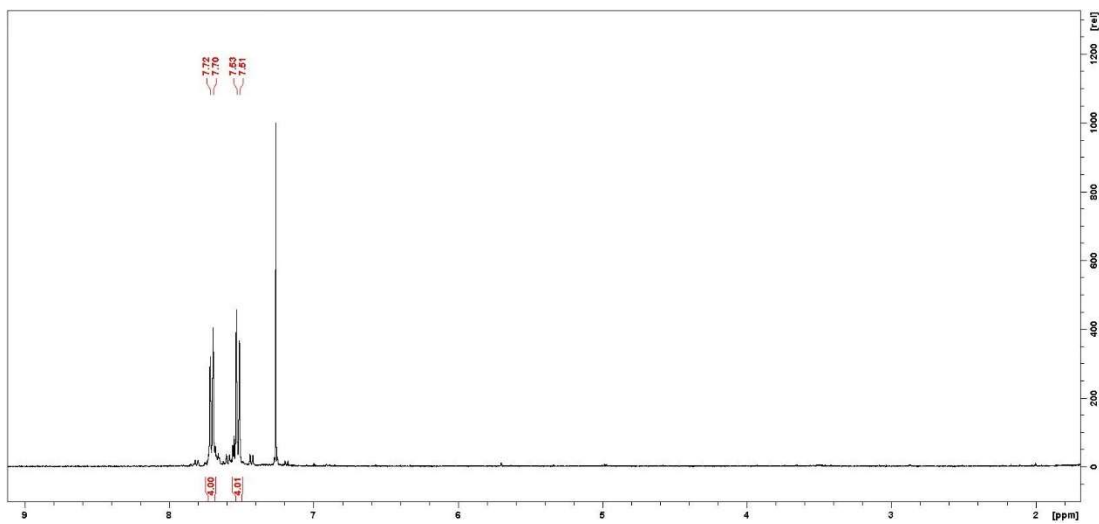


Figure 5.6.10. ¹H-NMR spectrum of compound **Se5**.

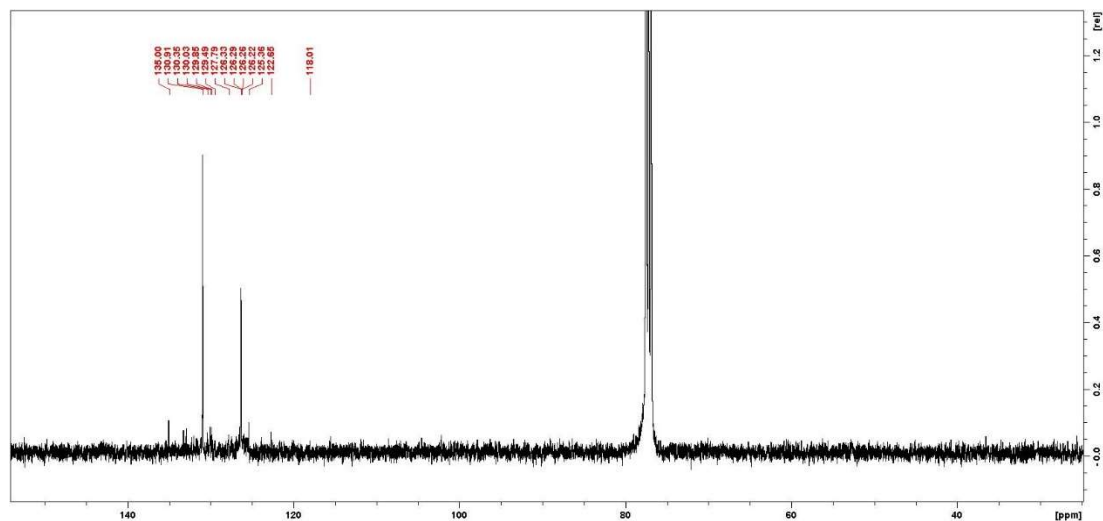
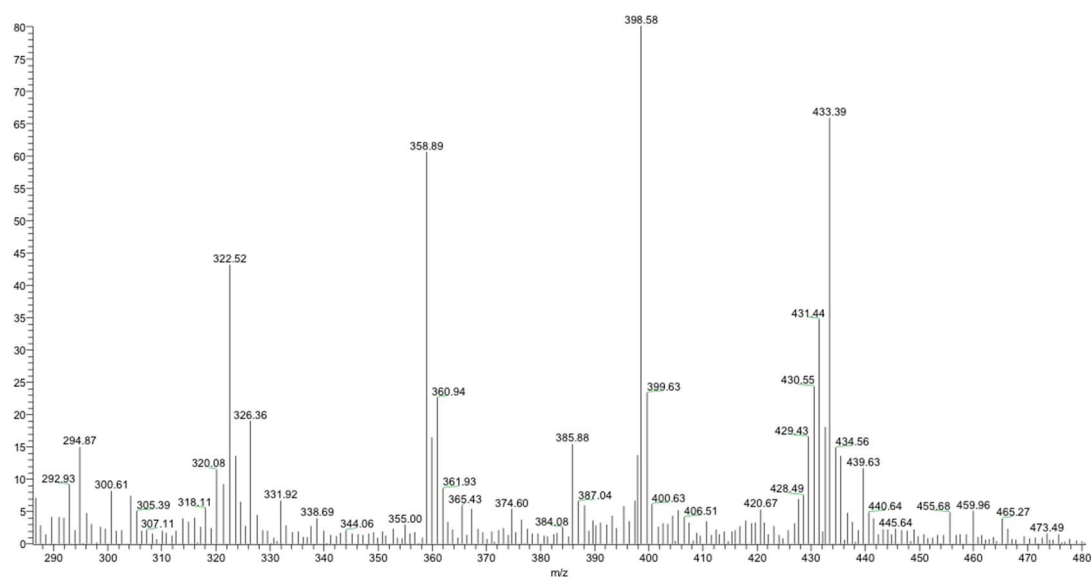
Figure 5.6.11. ¹³C-NMR spectrum of compound Se5.

Figure 5.6.12. ESI-MS spectrum of compound Se5.

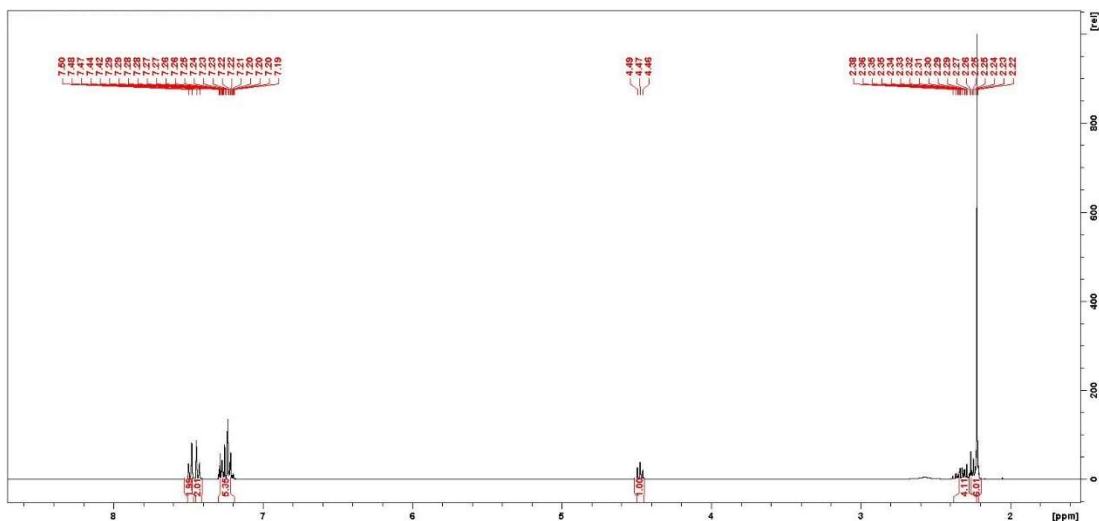


Figure 5.6.13. ¹H-NMR spectrum of compound 1-CF₃.

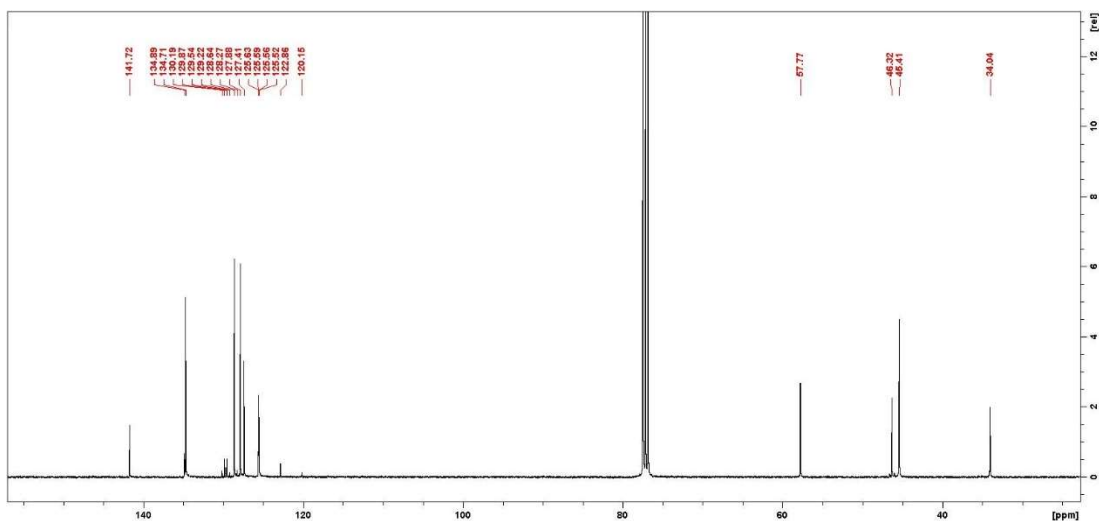
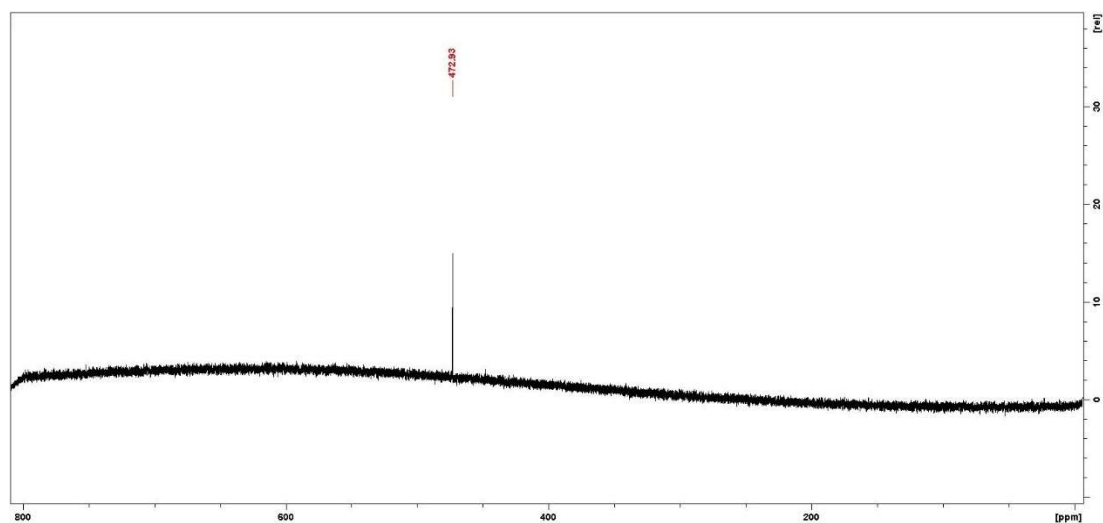
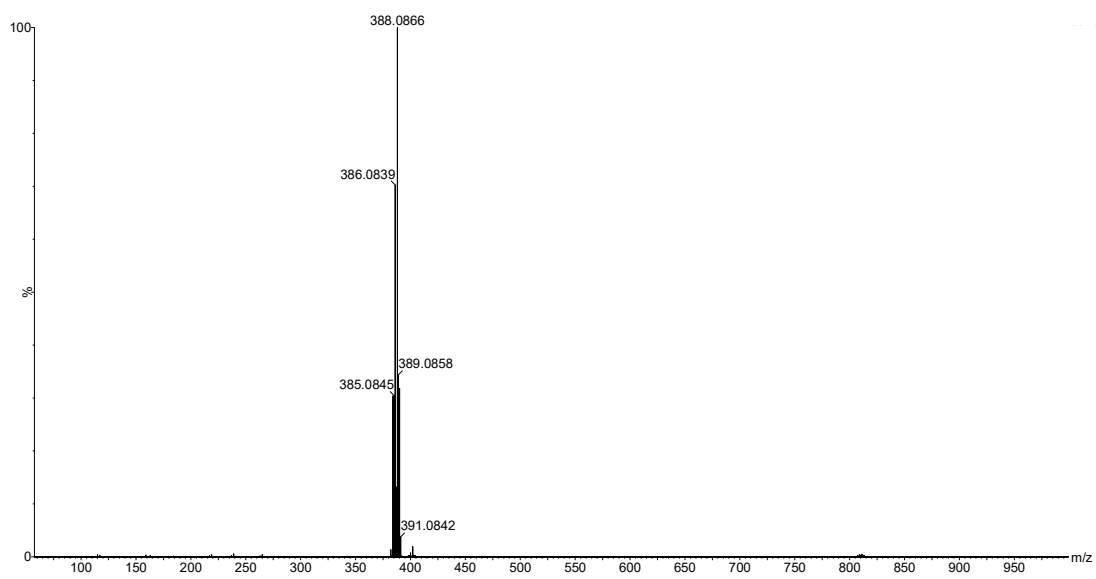


Figure 5.6.14. ¹³C-NMR spectrum of compound 1-CF₃.

Figure 5.6.15. ^{77}Se -NMR spectrum of compound **1-CF₃**.Figure 5.6.16. ESI-MS spectrum of compound **1-CF₃**.

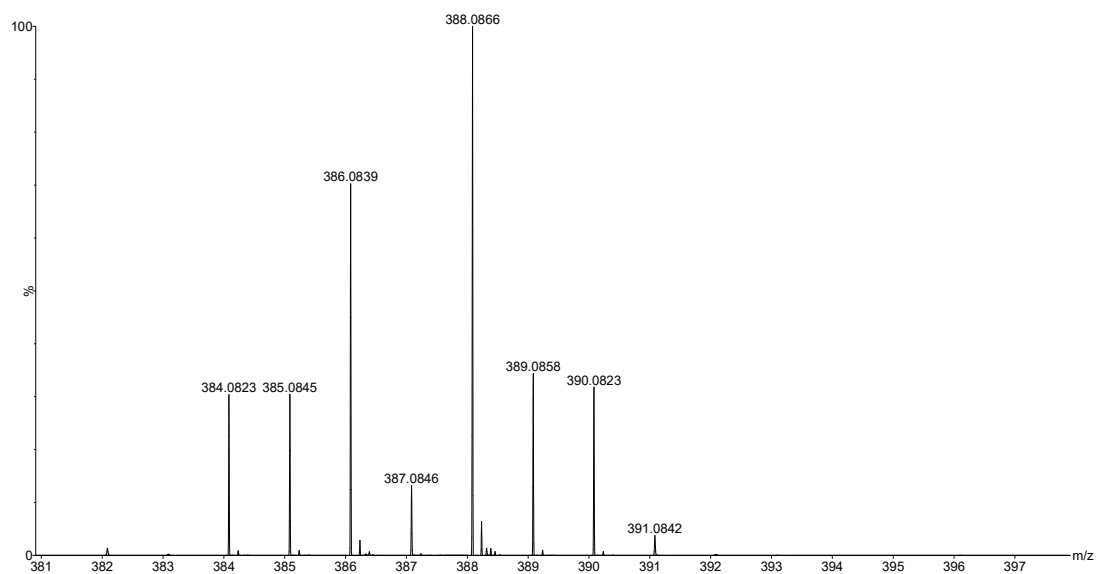


Figure 5.6.17. Detailed ESI-MS spectrum showing the isotopic distribution of compound **1-CF₃**.

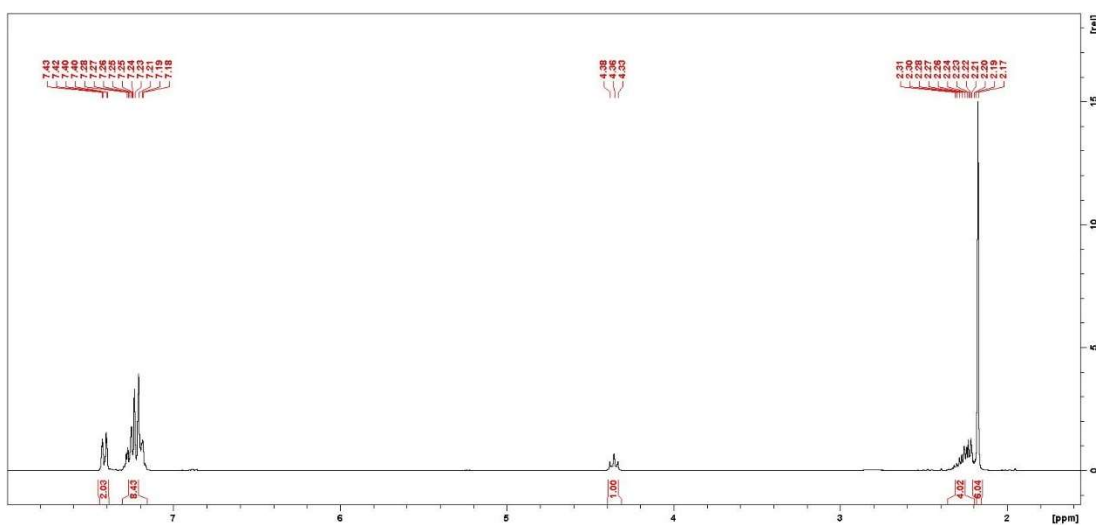
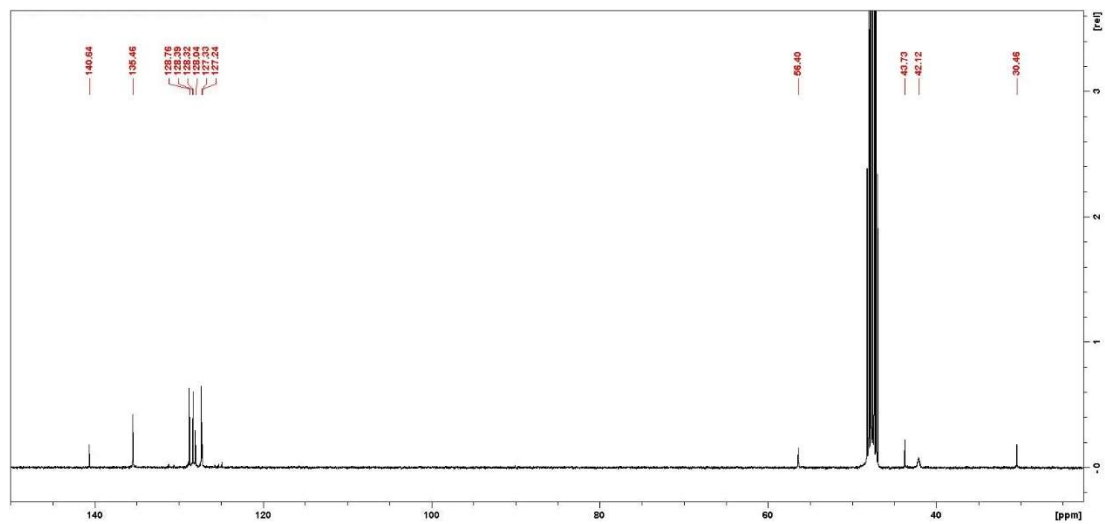
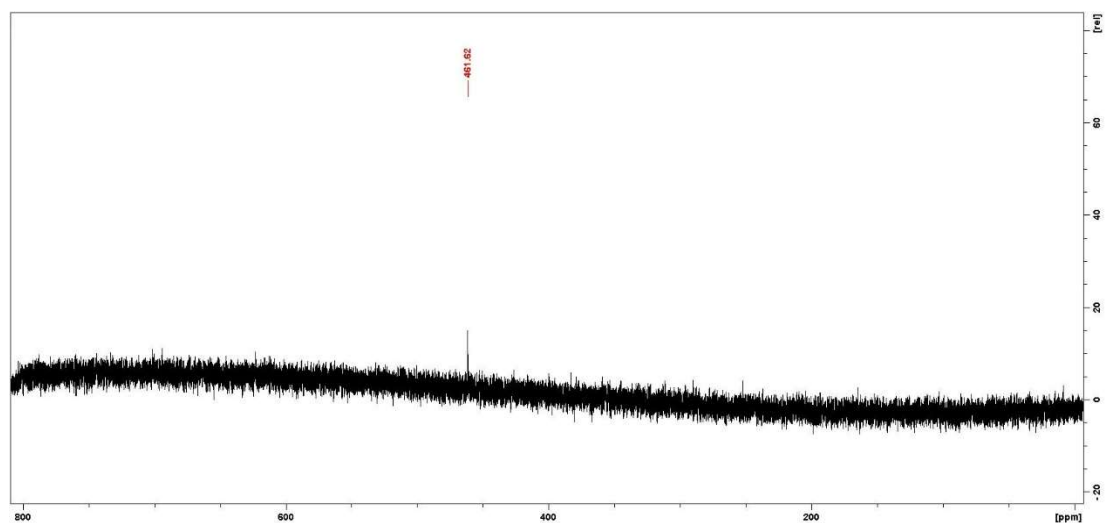


Figure 5.6.18. ¹H-NMR spectrum of compound **1-H**.

Figure 5.6.19. ¹³C-NMR spectrum of compound 1-H.Figure 5.6.20. ⁷⁷Se-NMR spectrum of compound 1-H.

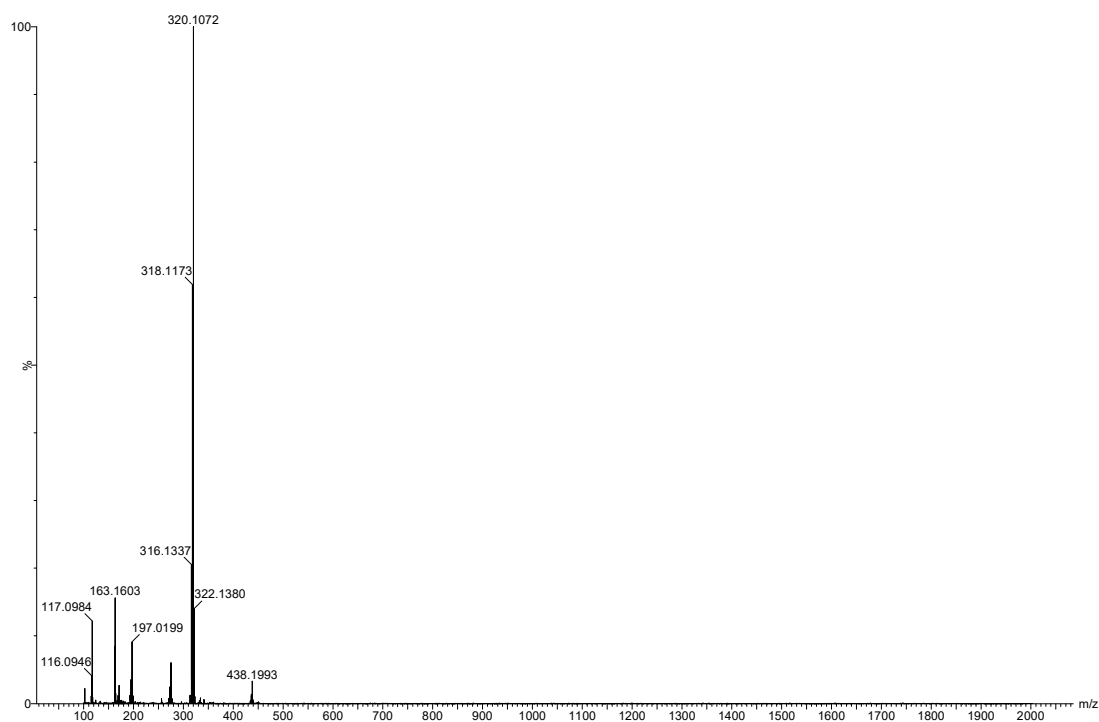


Figure 5.6.21. ESI-MS spectrum of compound **1-H**.

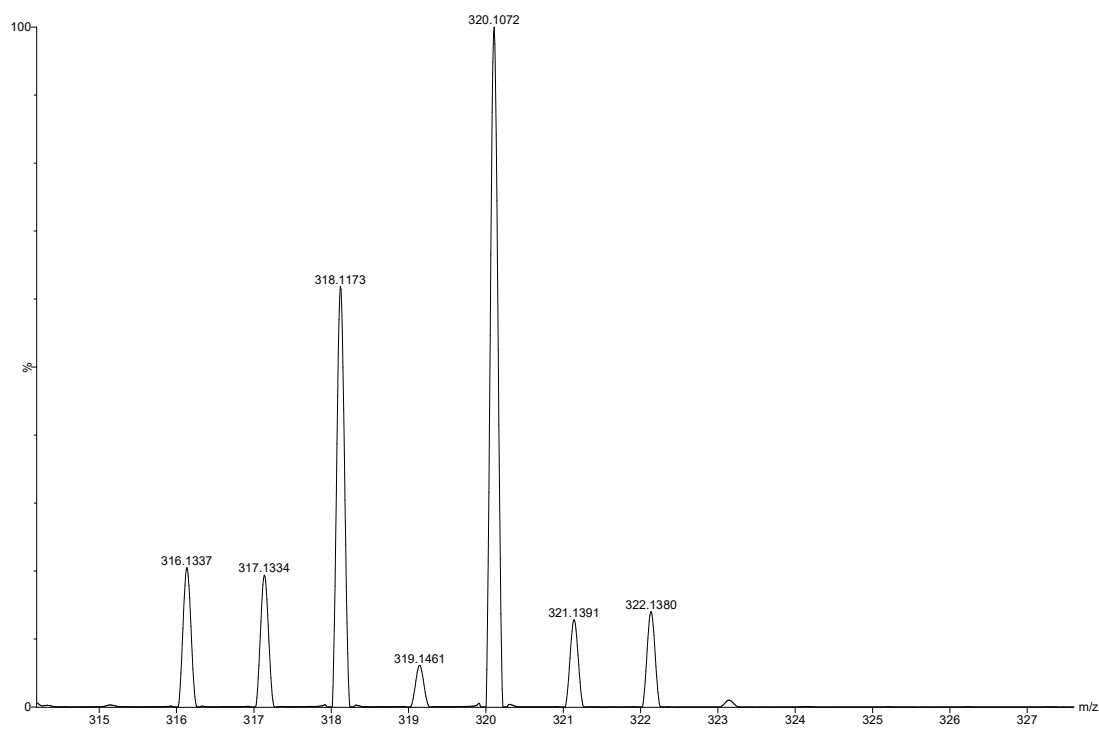
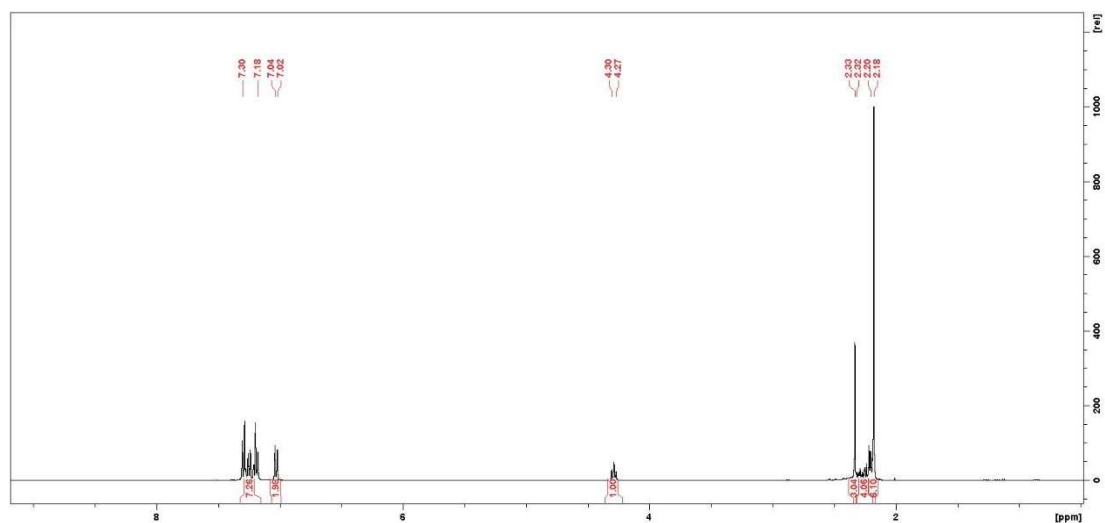
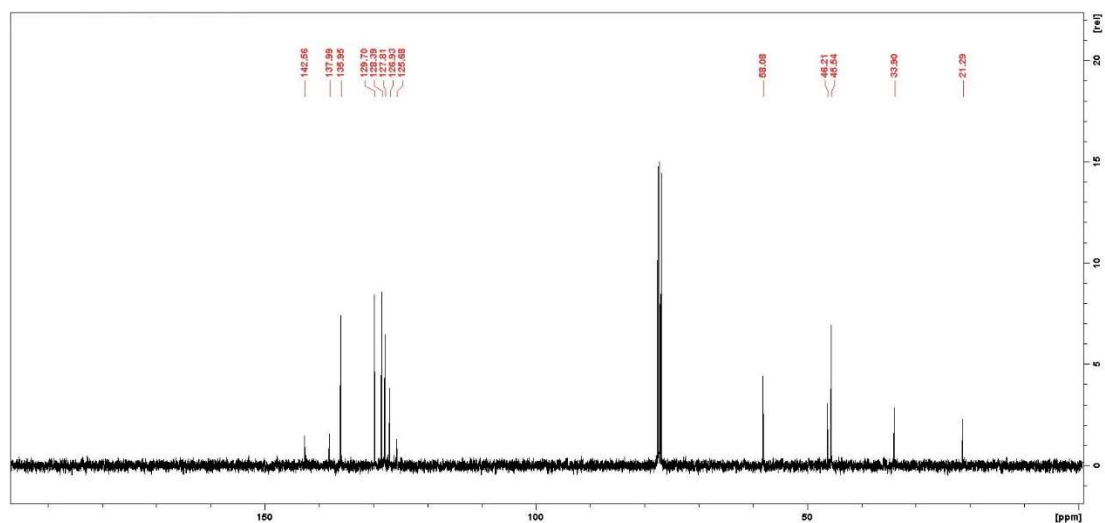


Figure 5.6.22. Detailed ESI-MS spectrum showing the isotopic distribution of compound **1-H**.

Figure 5.6.23. ¹H-NMR spectrum of compound 1-CH₃.Figure 5.6.24. ¹³C-NMR spectrum of compound 1-CH₃.

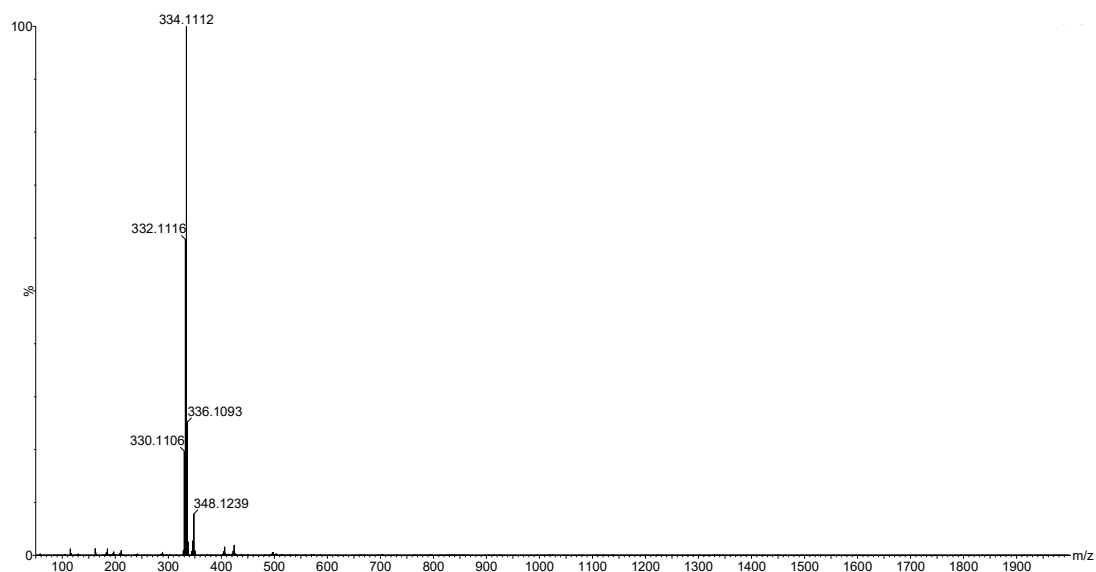


Figure 5.6.25. ESI-MS spectrum of compound 1-CH₃ (positive ionization mode).

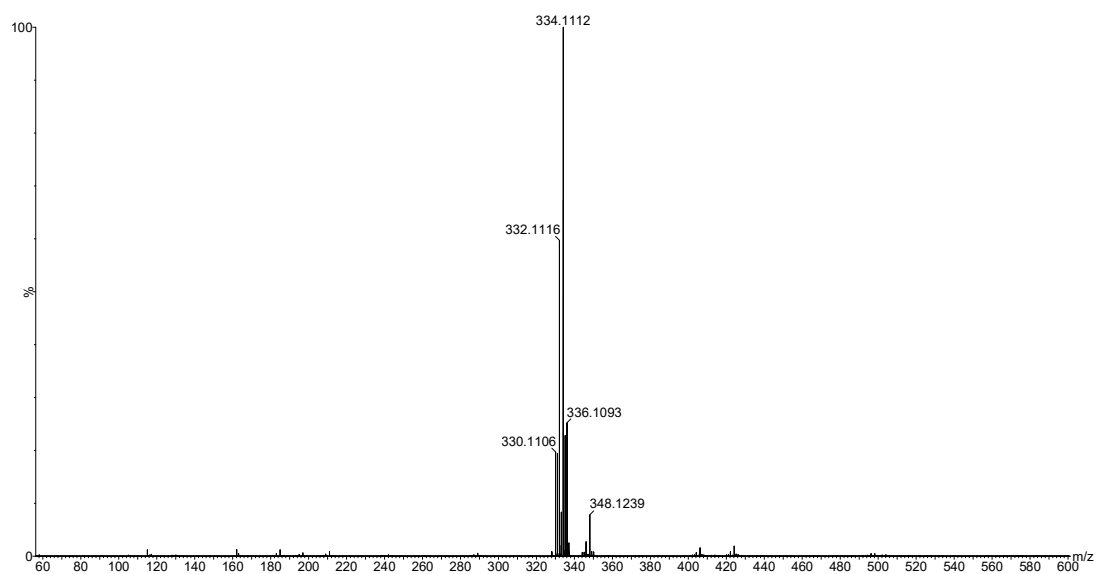
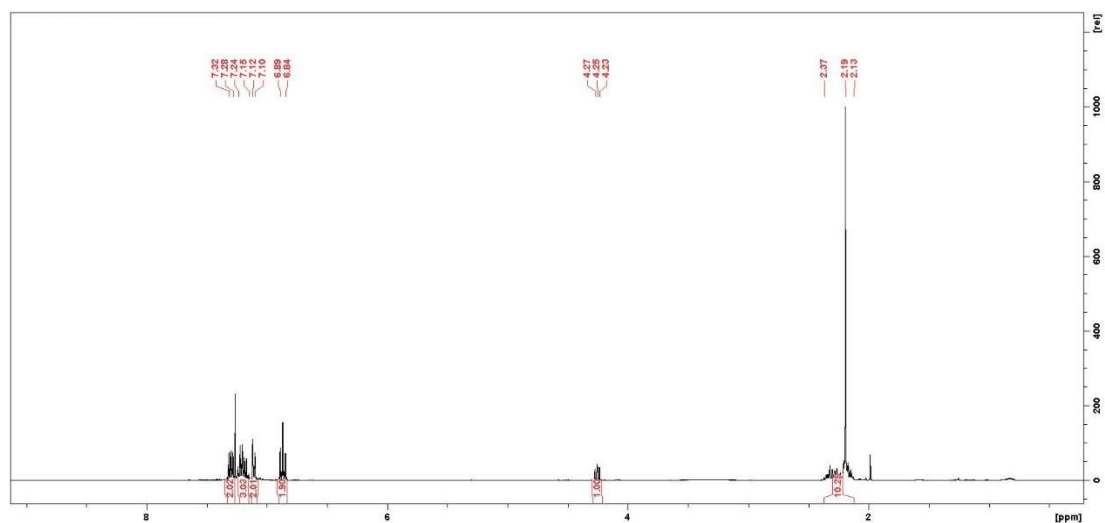
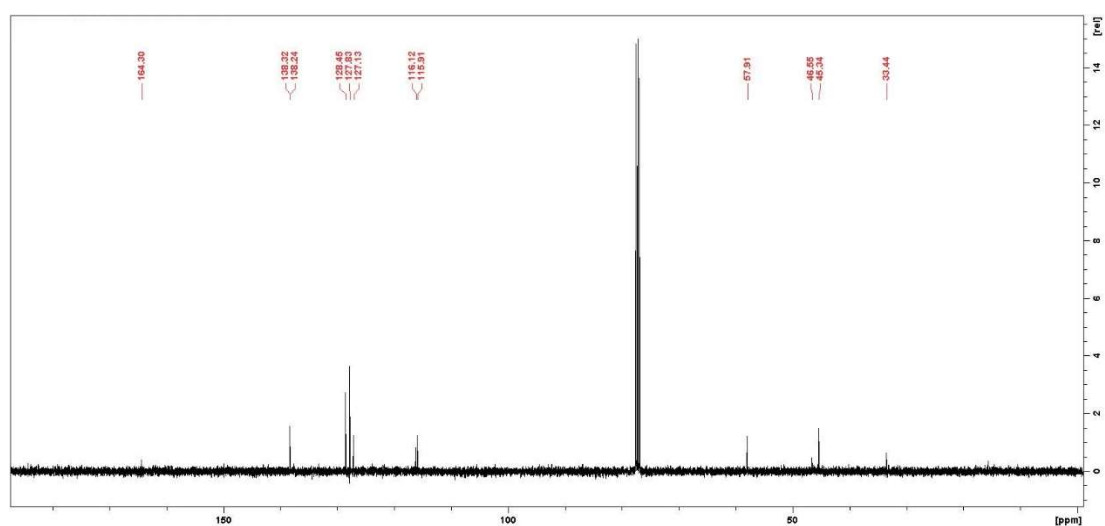


Figure 5.6.26. Detailed view of the peak of compound 1-CH₃, showing the isotopic distribution, in the ESI-MS spectrum (positive ionization mode).

Figure 5.6.27. ¹H-NMR spectrum of compound 1-F.Figure 5.6.28. ¹³C-NMR spectrum of compound 1-F.

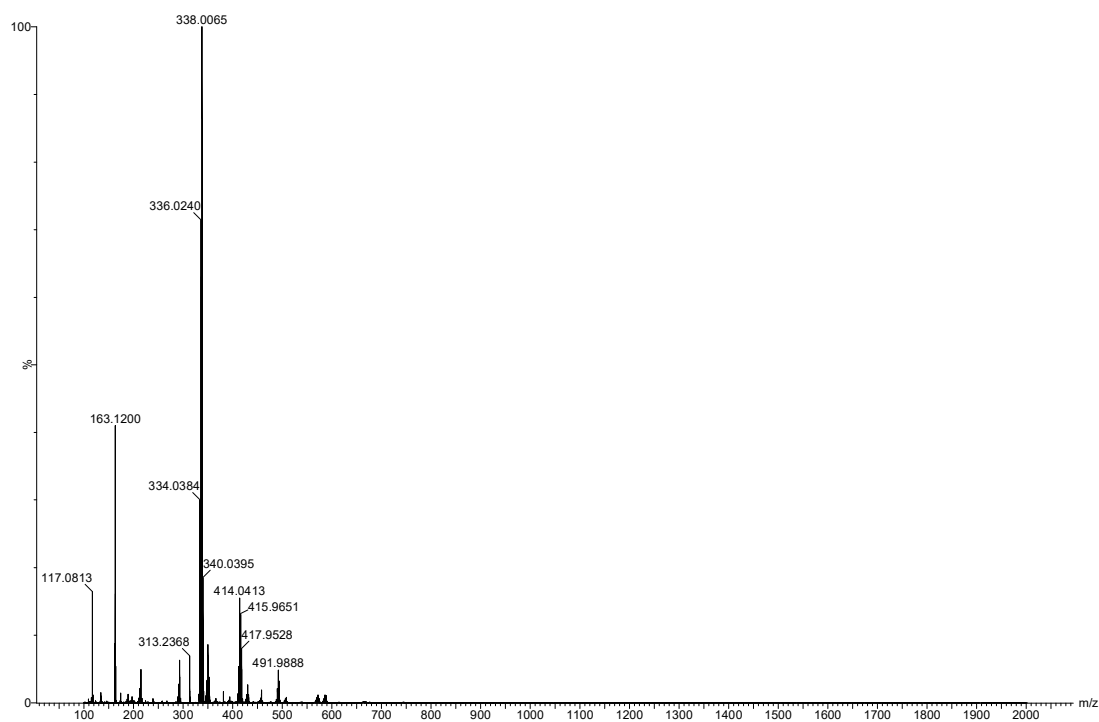


Figure 5.6.29. ESI-MS spectrum of compound 1-F (positive ionization mode).

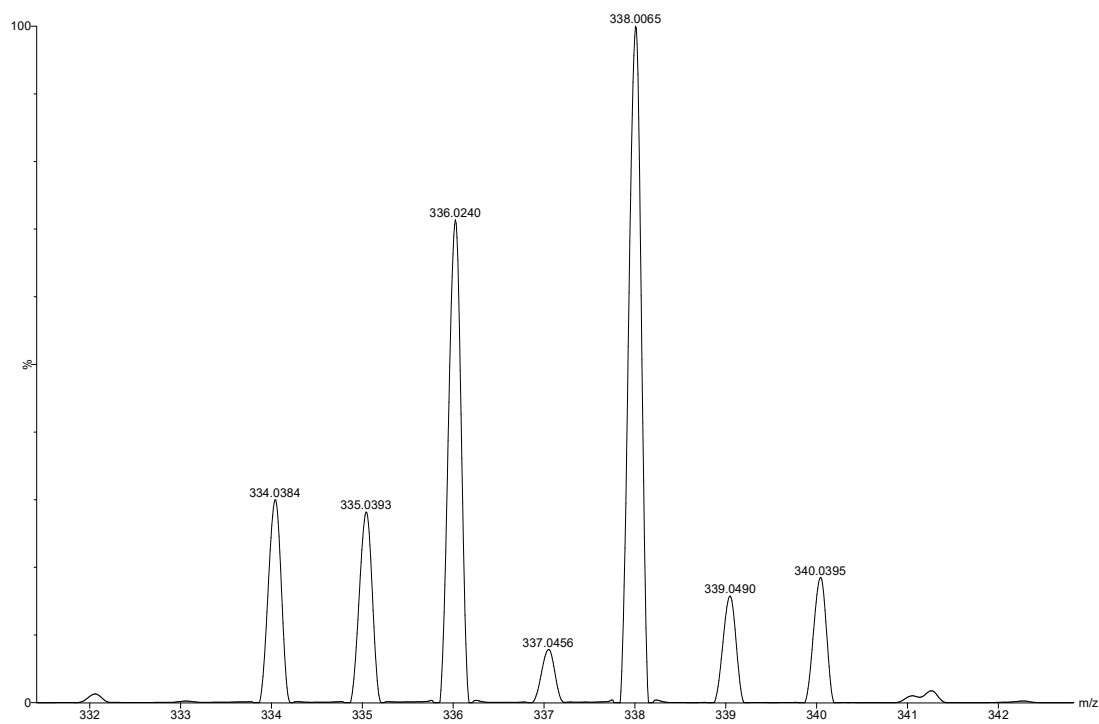
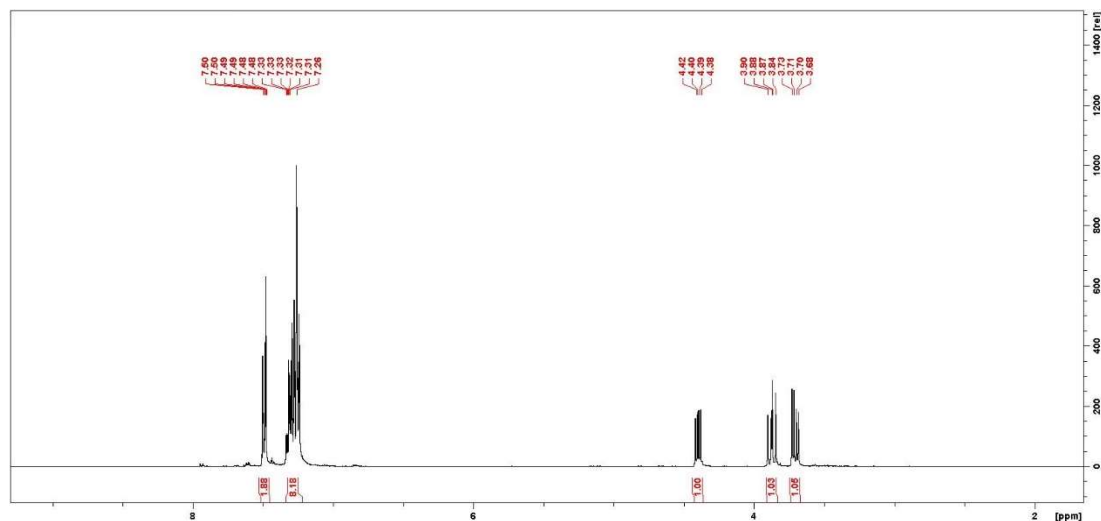
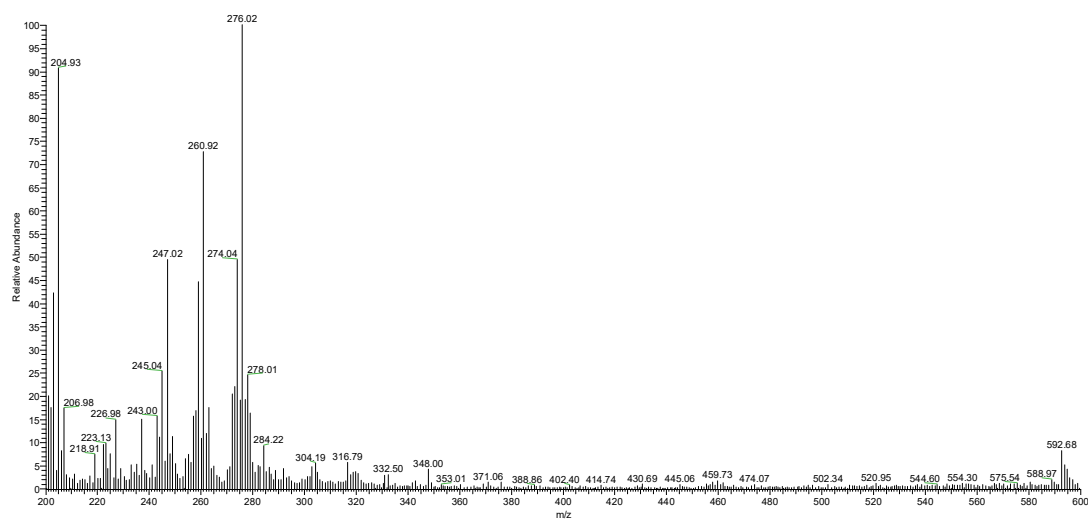


Figure 5.6.30. Detailed view of the peak of compound 1-F, showing the isotopic distribution, in the ESI-MS spectrum (positive ionization mode).

Figure 5.6.31. $^1\text{H-NMR}$ spectrum of compound **SeA8**.Figure 5.6.32. ESI-MS spectrum of compound **SeA8**. The peak at m/z 260.92 corresponds to $[\text{SeA8-N}_3+\text{H}]^+$, at m/z 276.02 to $[\text{SeA8-N}_2+\text{H}]^+$ and at m/z 284.22 to $[\text{SeA8}+\text{H}]^+$.

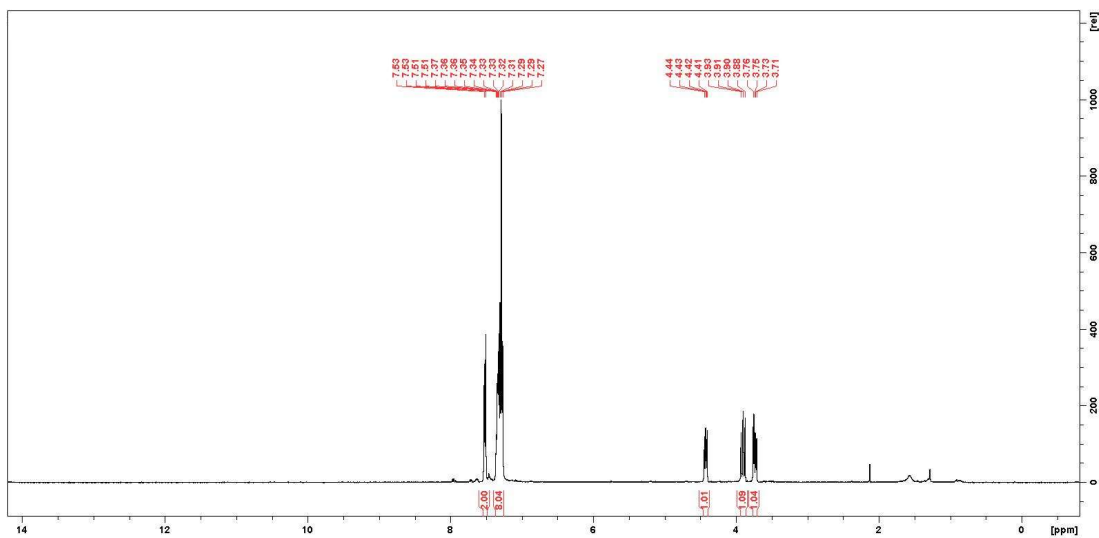


Figure 5.6.33. ¹H-NMR spectrum of compound SeA9.

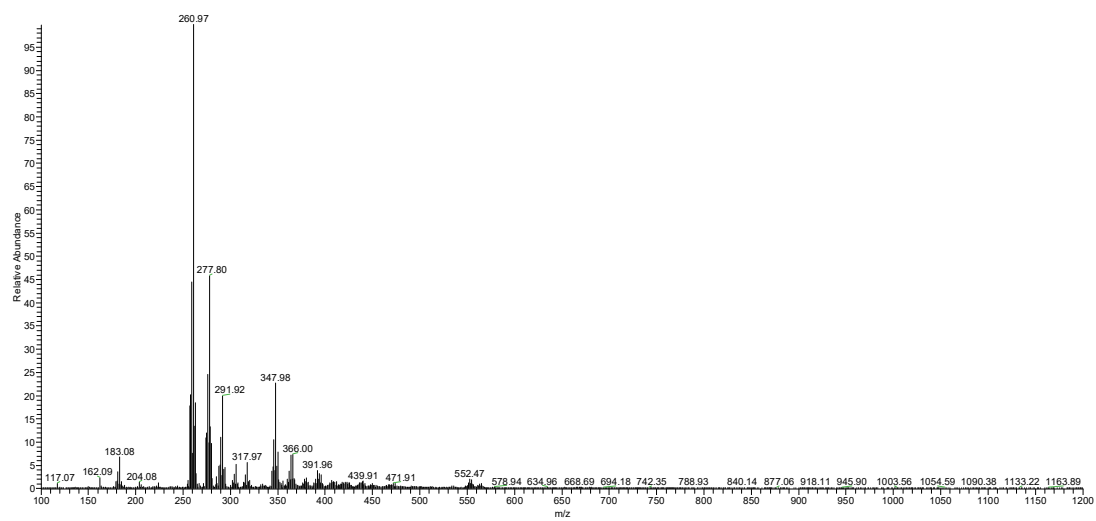
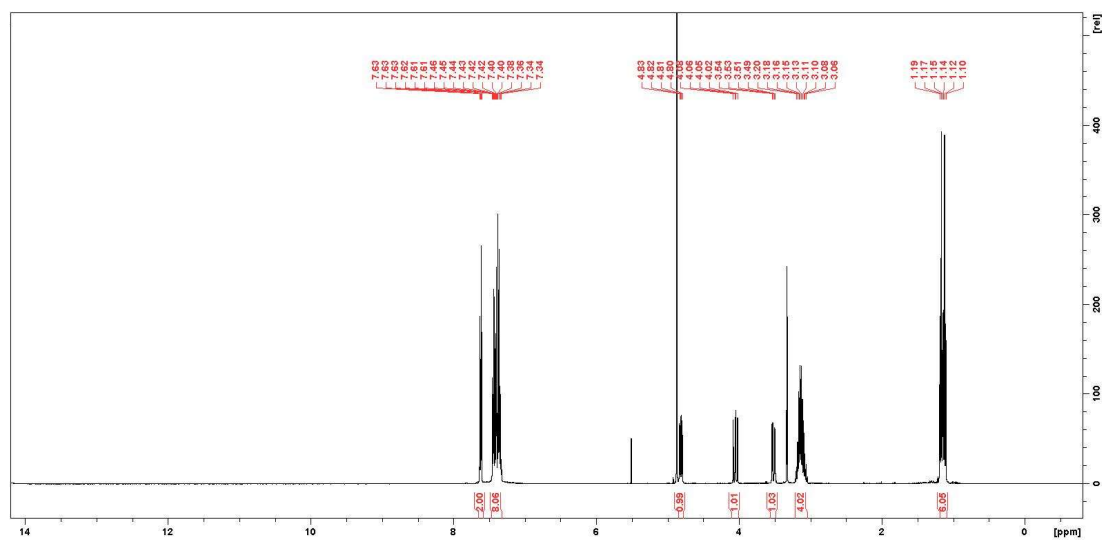
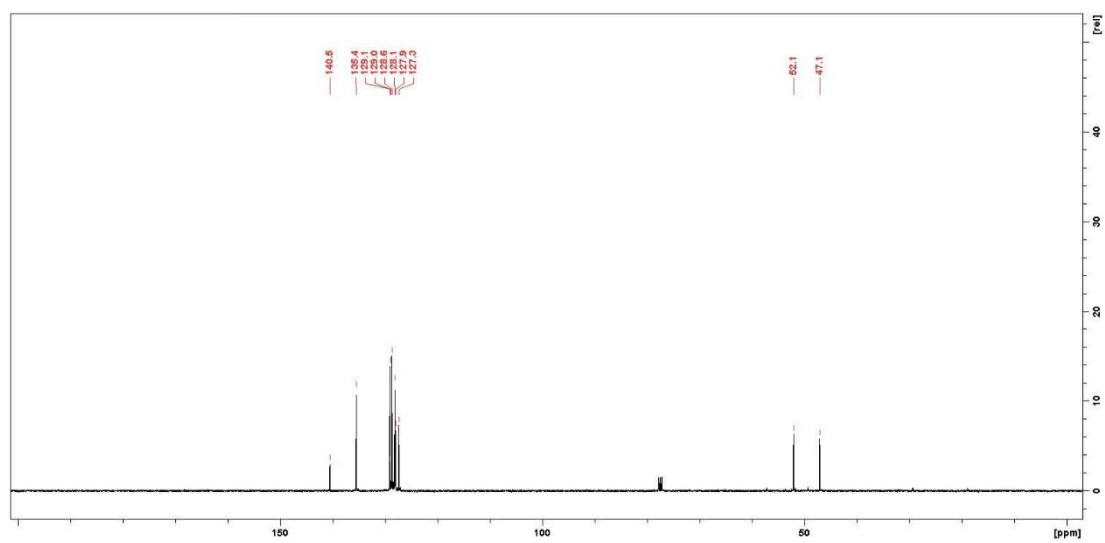


Figure 5.6.34. ESI-MS spectrum of compound SeA9. The peak at m/z 260.97 corresponds to $[\text{SeA9-NH}_2+\text{H}]^+$, at m/z 277.80 to $[\text{SeA9+H}]^+$.

Figure 5.6.35. $^1\text{H-NMR}$ spectrum of compound **SeA1**.Figure 5.6.36. $^{13}\text{C-NMR}$ spectrum of compound **SeA1**.

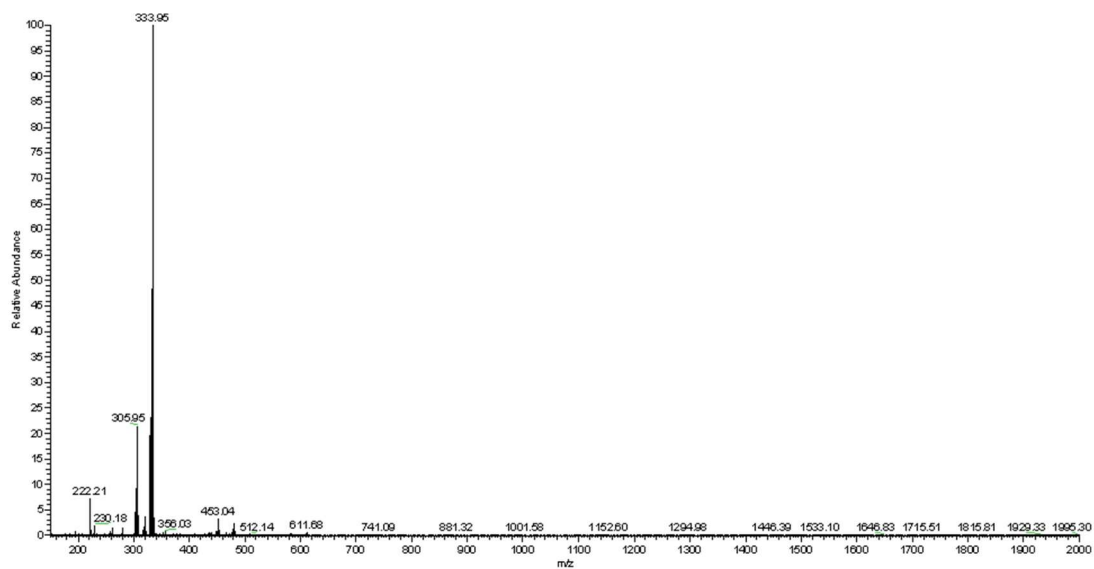


Figure 5.6.37. ESI-MS spectrum of compound SeA1.

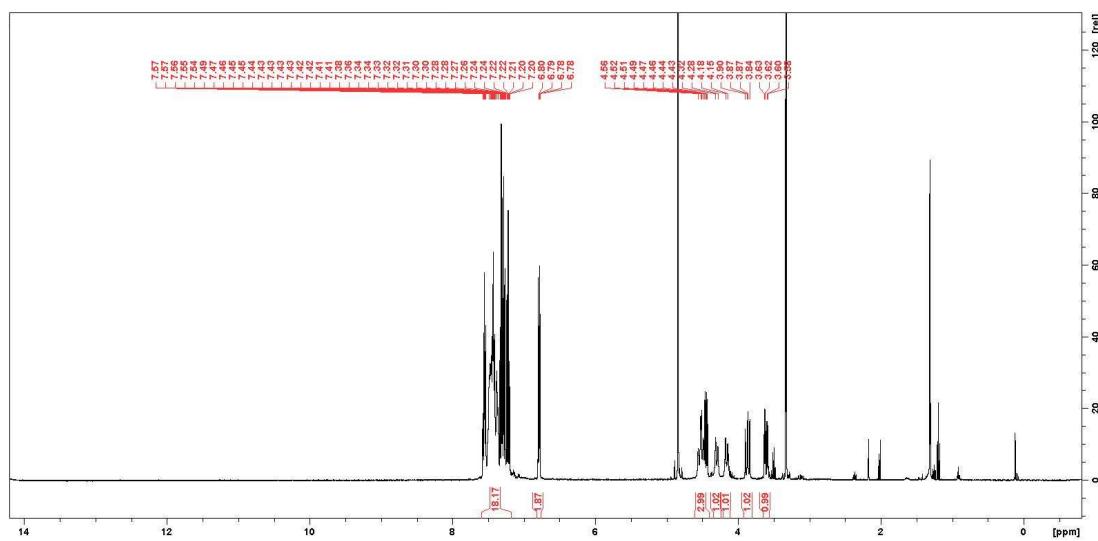
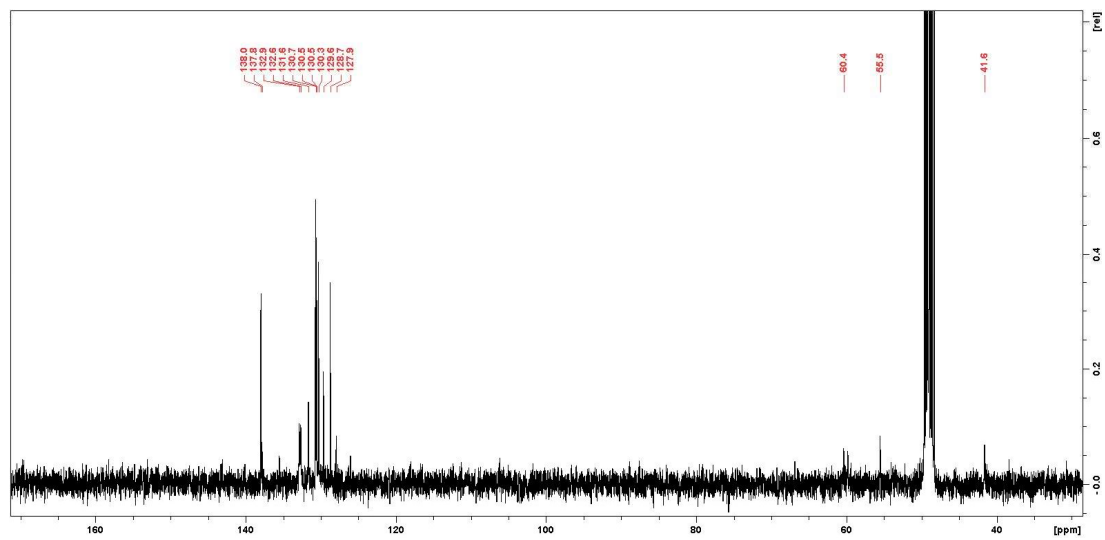
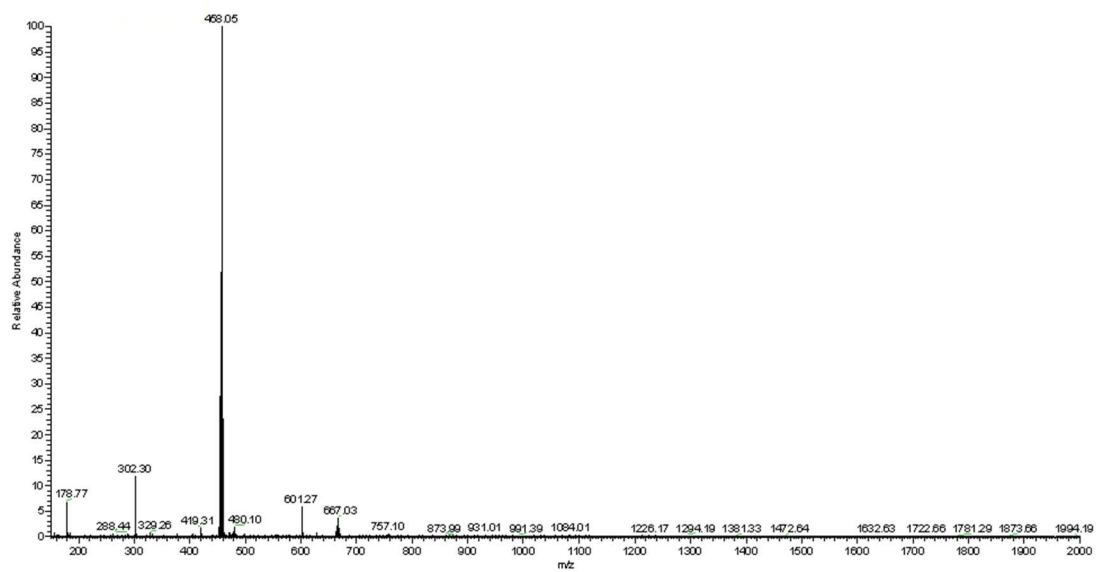


Figure 5.6.38. ¹H-NMR spectrum of compound SeA2.

Figure 5.6.39. ^{13}C -NMR spectrum of compound **SeA2**.Figure 5.6.40. ESI-MS spectrum of compound **SeA2**.

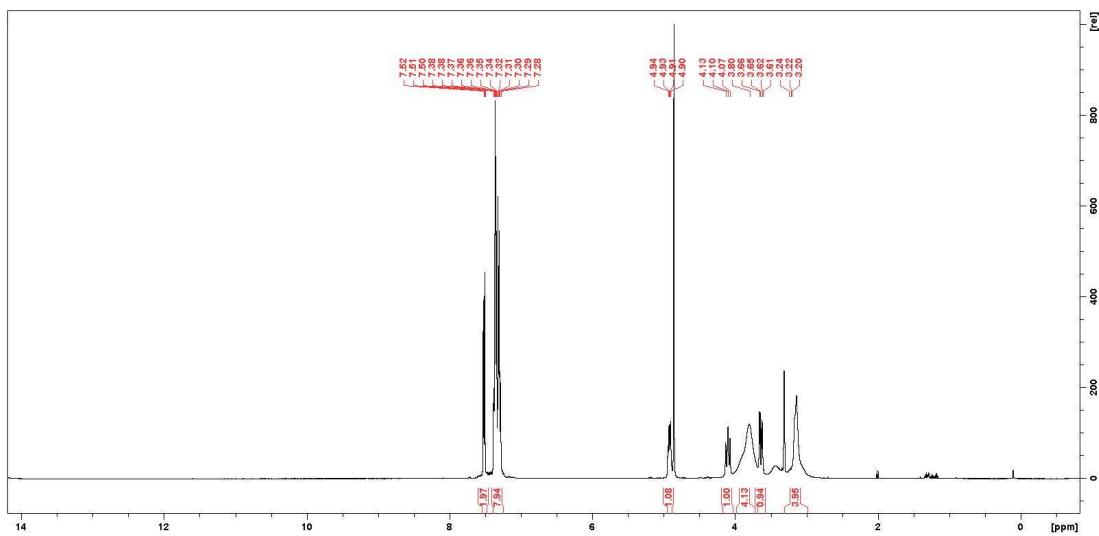


Figure 5.6.41. ^1H -NMR spectrum of compound **SeA3**.

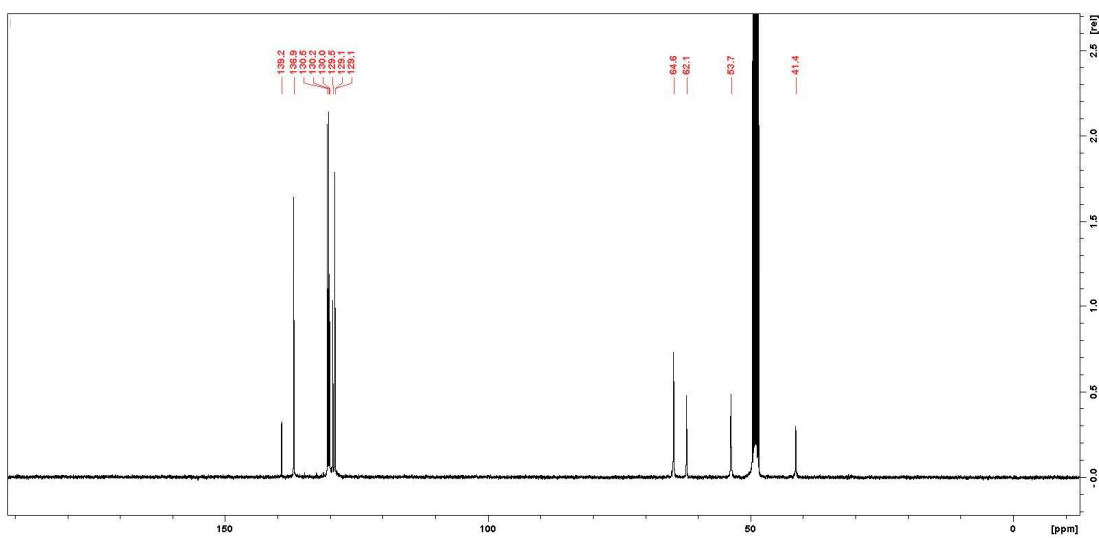
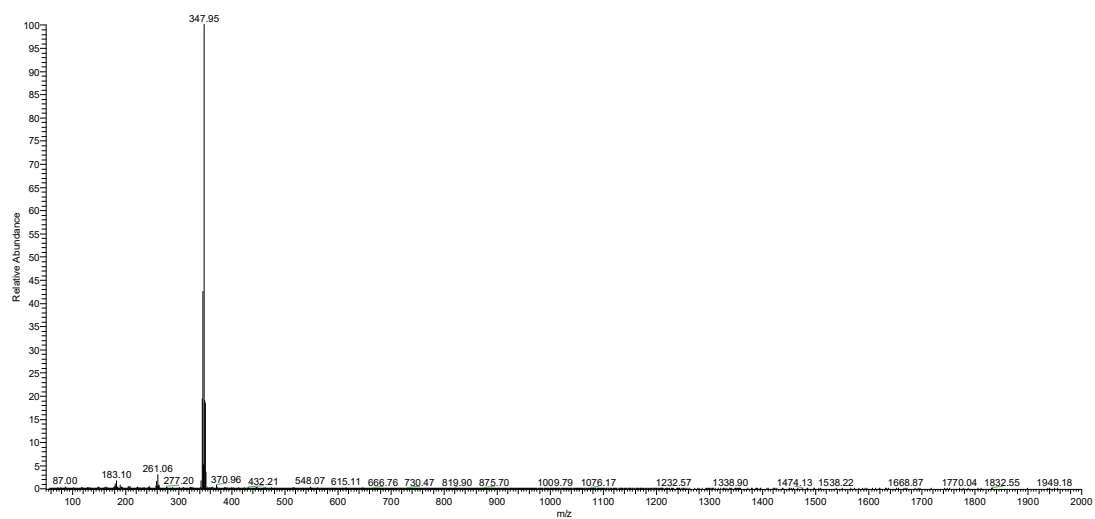
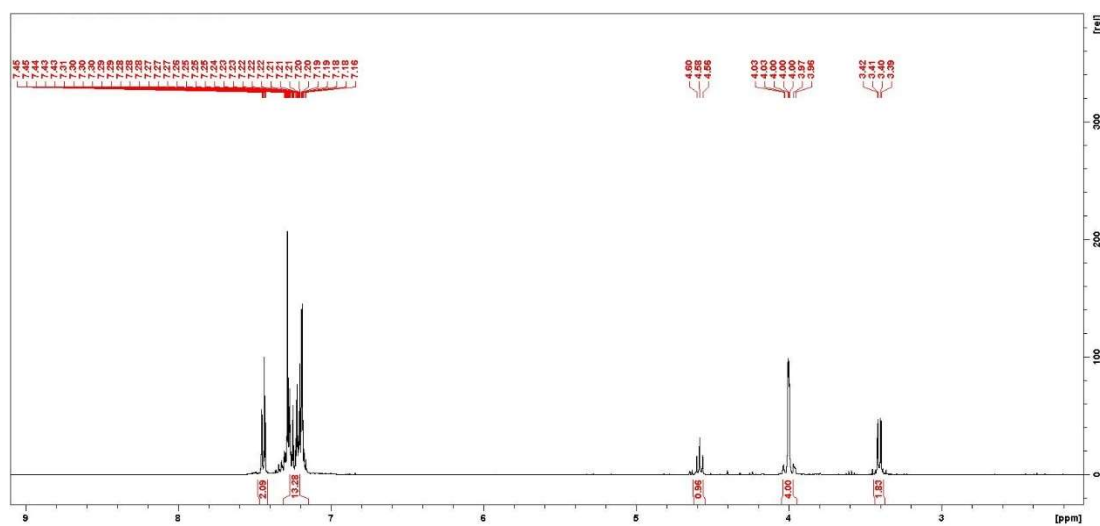


Figure 5.6.42. ^{13}C -NMR spectrum of compound **SeA3**.

Figure 5.6.43. ESI-MS spectrum of compound **SeA3**.Figure 5.6.44. ¹H-NMR spectrum of compound **SeA4**.

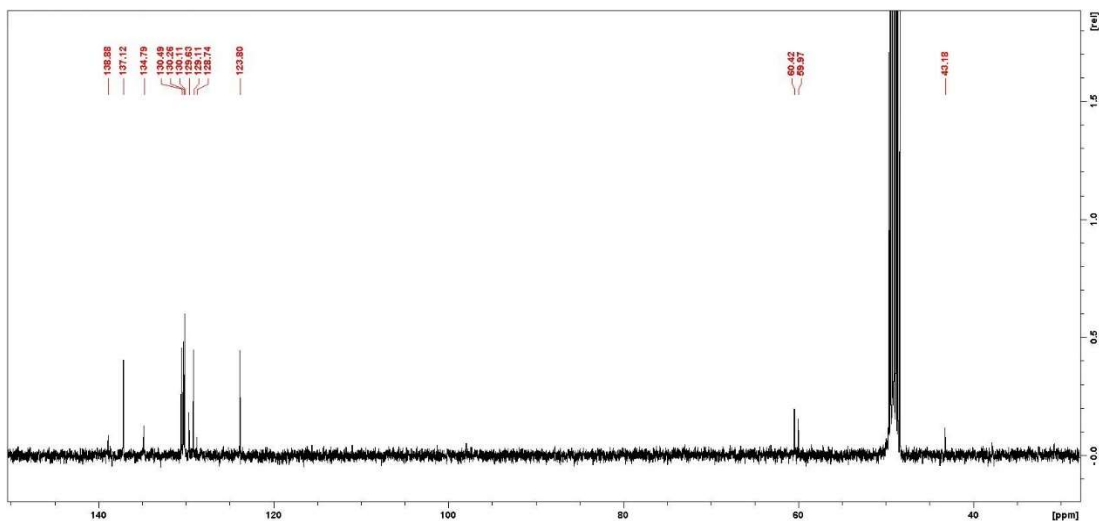


Figure 5.6.45. ¹³C-NMR spectrum of compound SeA4.

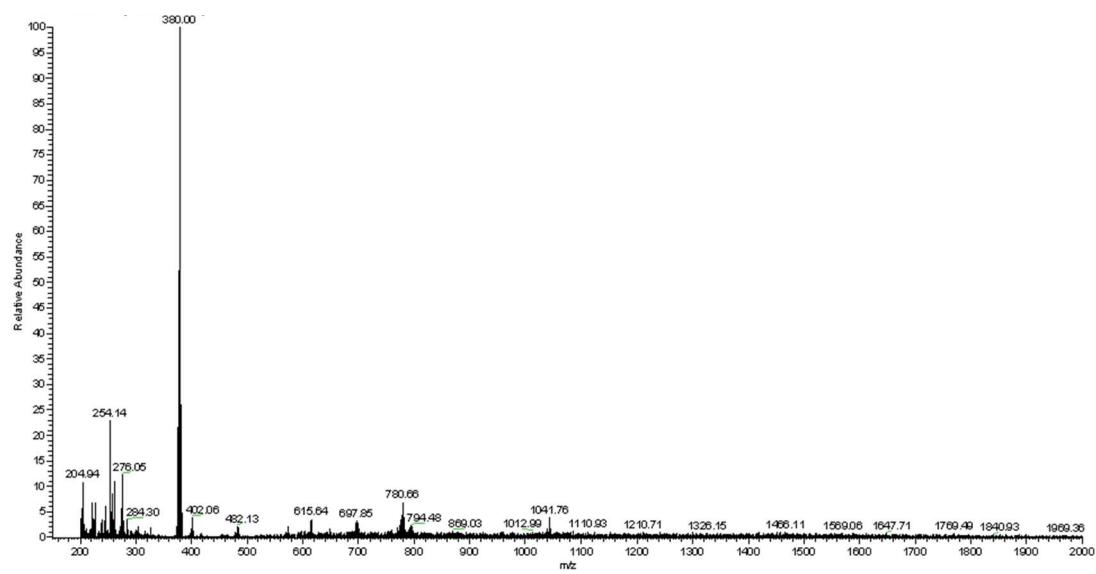
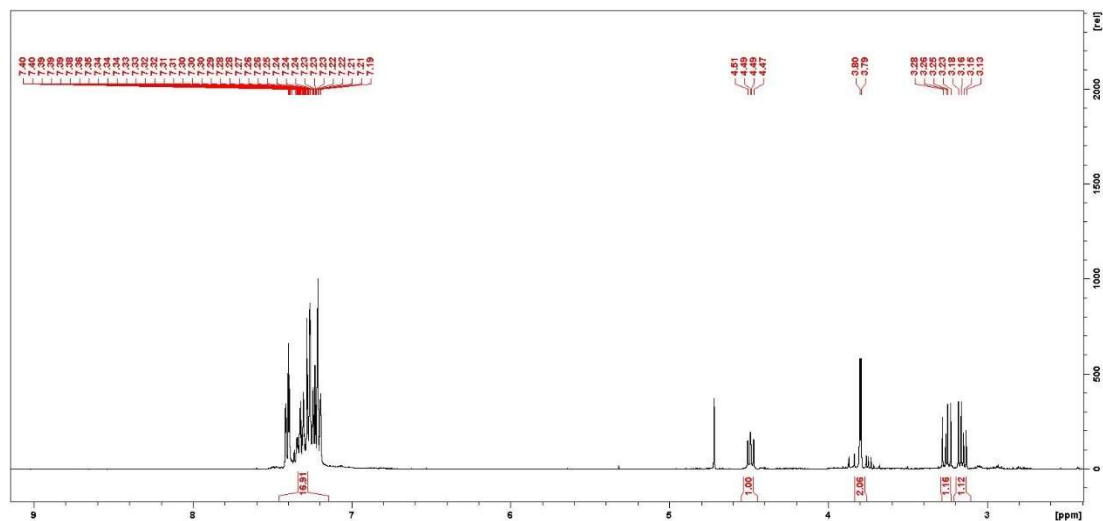
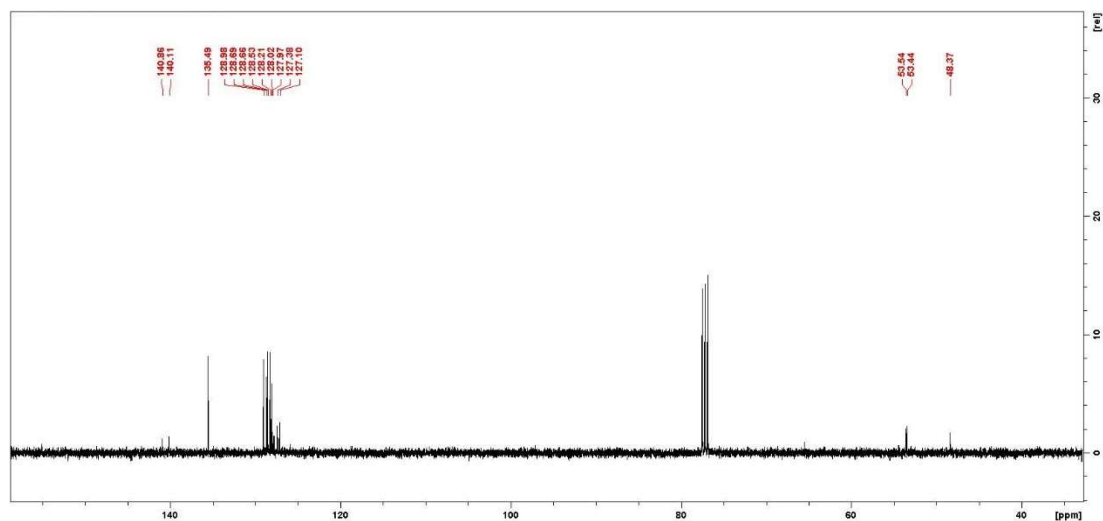


Figure 5.6.46. ESI-MS spectrum of compound SeA4.

Figure 5.6.47. ^1H -NMR spectrum of compound **SeA5**.Figure 5.6.48. ^{13}C -NMR spectrum of compound **SeA5**.

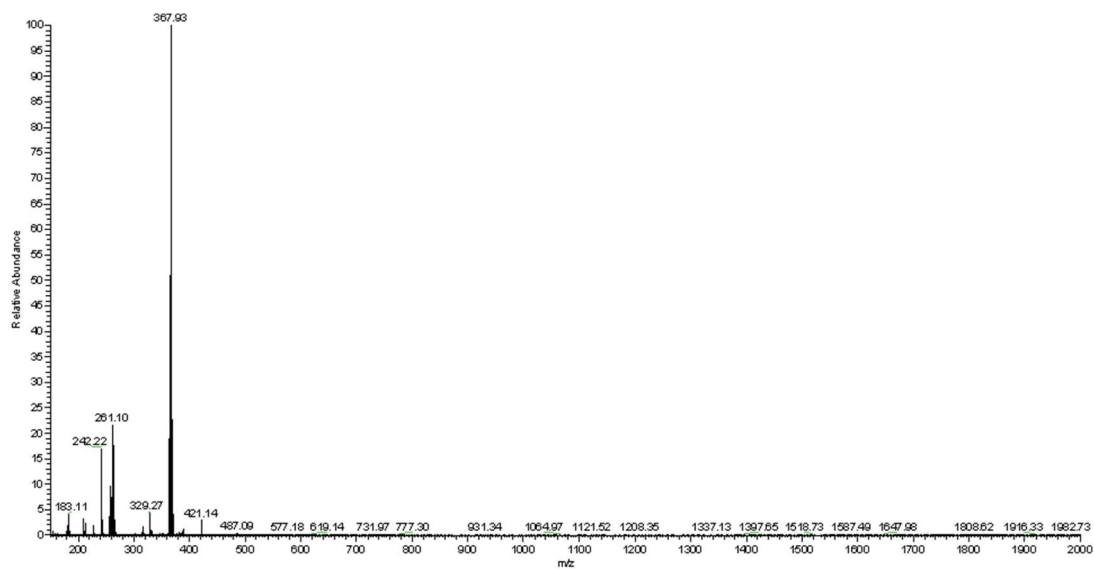


Figure 5.6.49. ESI-MS spectrum of compound SeA5.

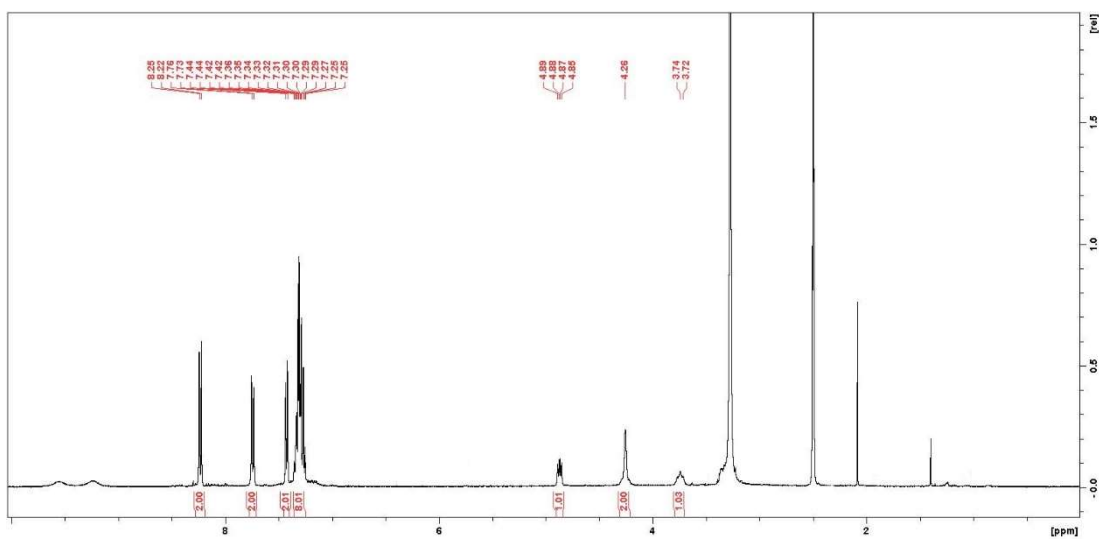
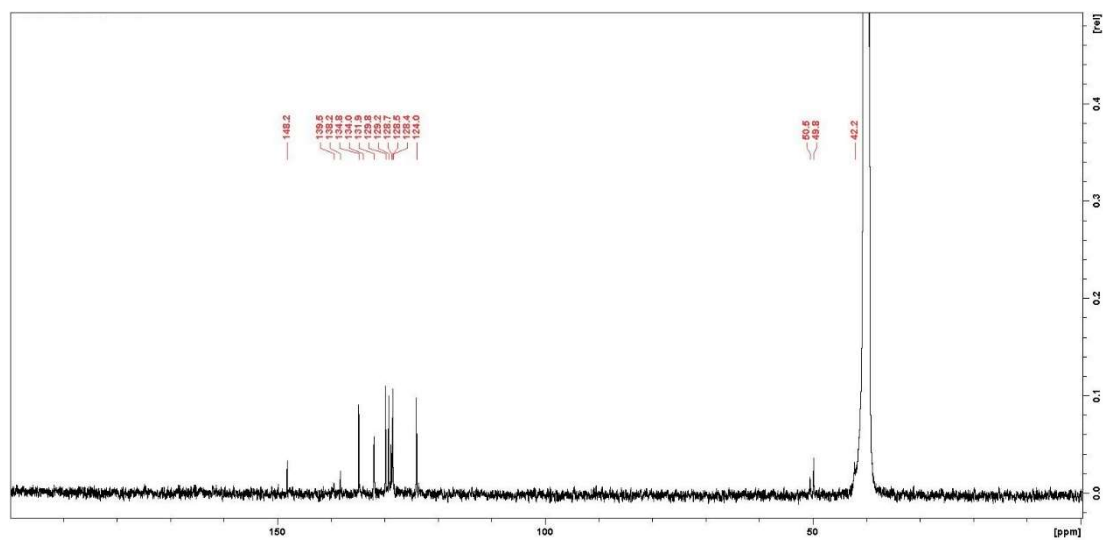
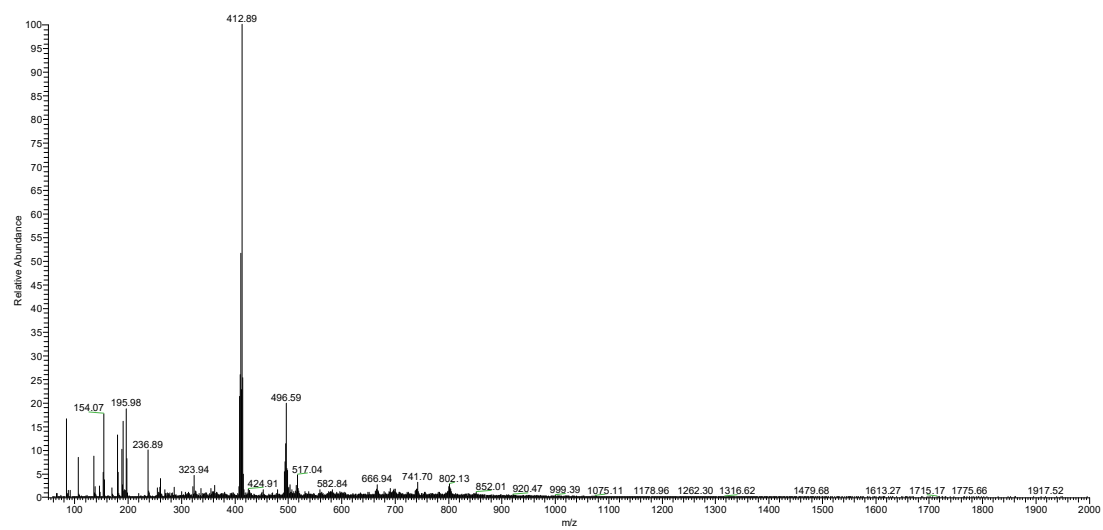


Figure 5.6.50. ¹H-NMR spectrum of compound SeA6.

Figure 5.6.51. ^{13}C -NMR spectrum of compound **SeA6**.Figure 5.6.52. ESI-MS spectrum of compound **SeA6**.

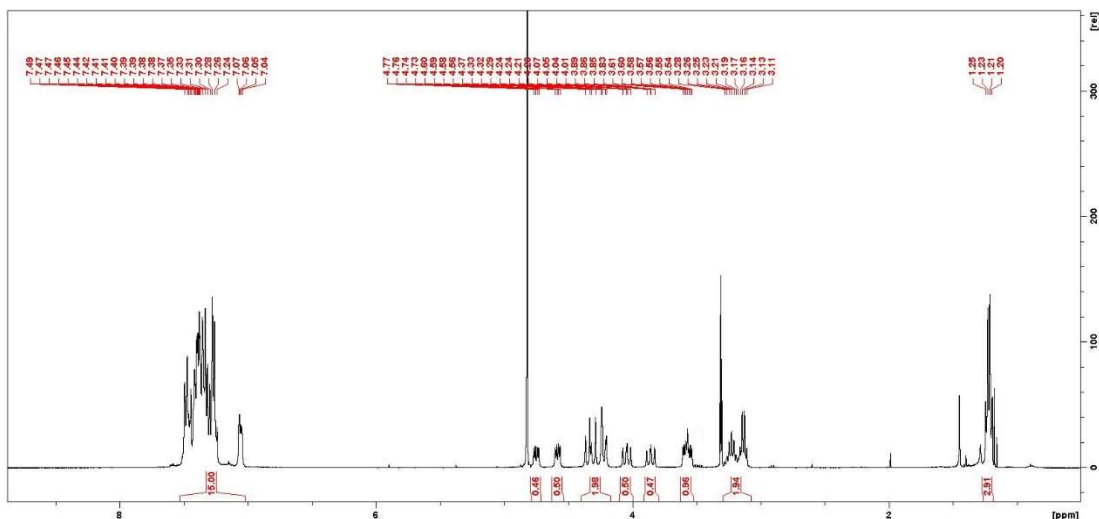


Figure 5.6.53. ^1H -NMR spectrum of compound SeA7.

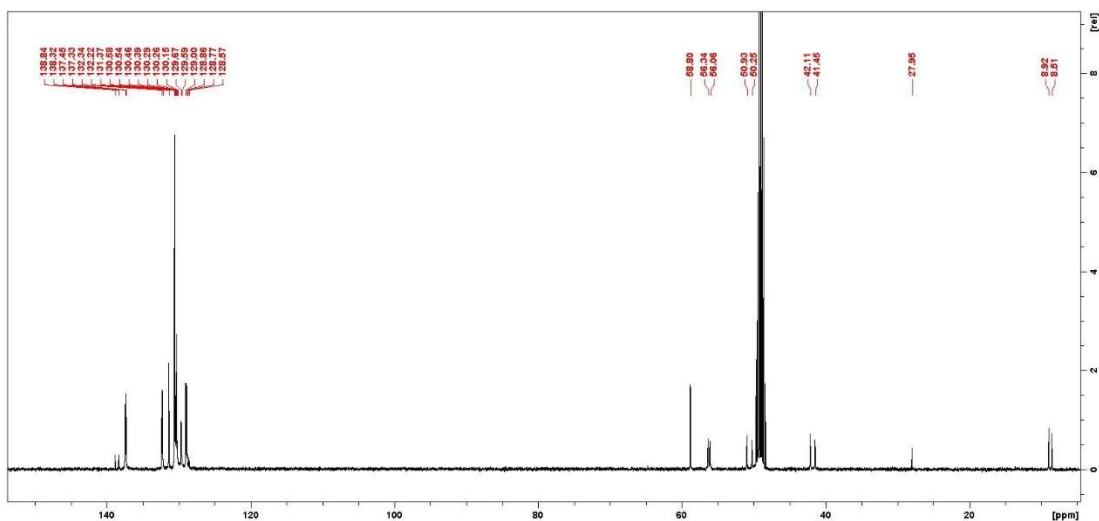
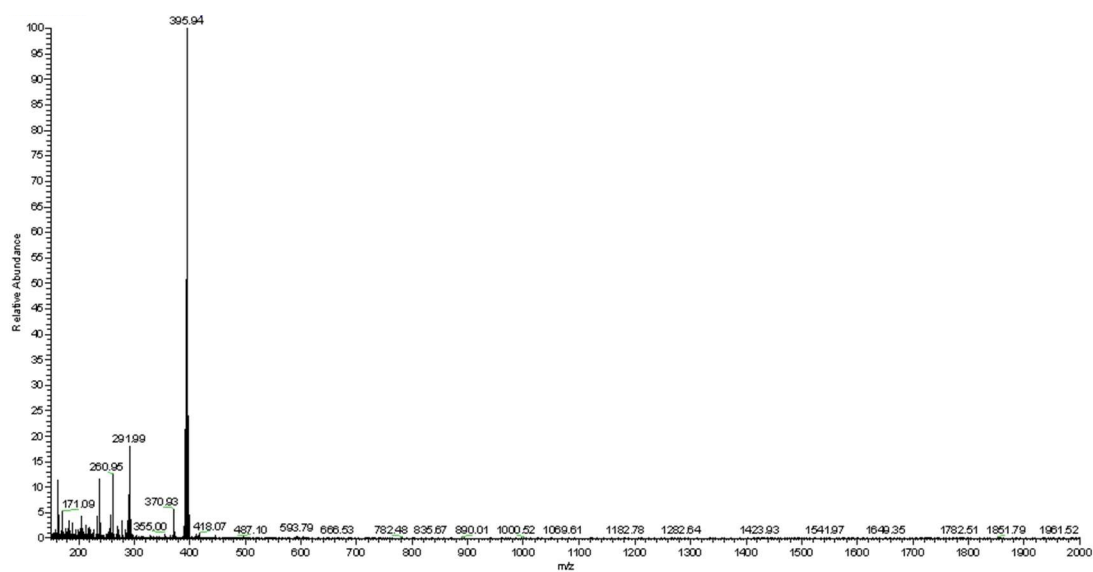
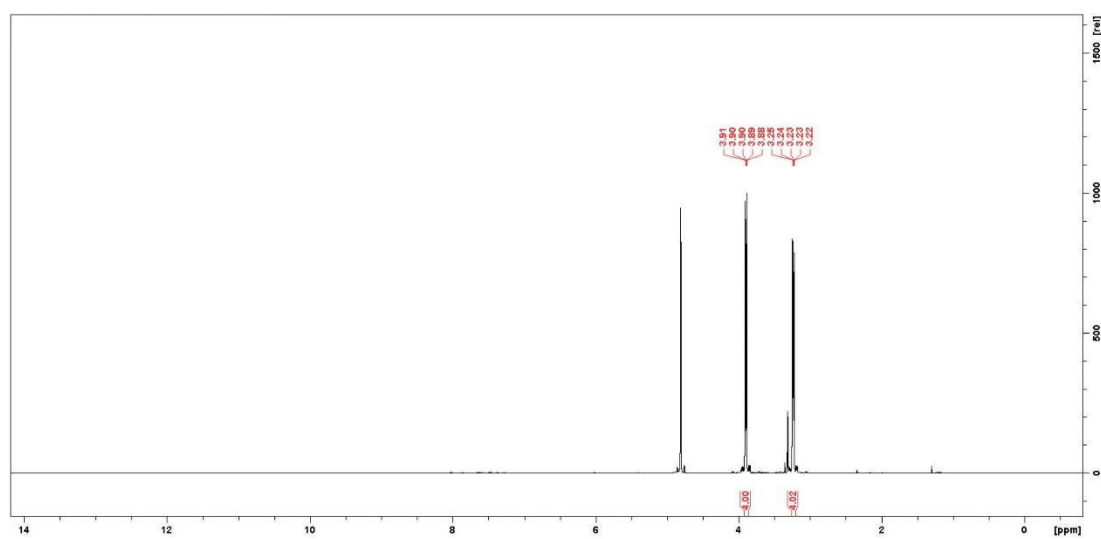


Figure 5.6.54. ^{13}C -NMR spectrum of compound SeA7.

Figure 5.6.55. ESI-MS spectrum of compound **SeA7**.Figure 5.6.56. $^1\text{H-NMR}$ spectrum of the final product of the oxidation of compound **SeA3**, *i.e.*, morpholine hydrochloride.

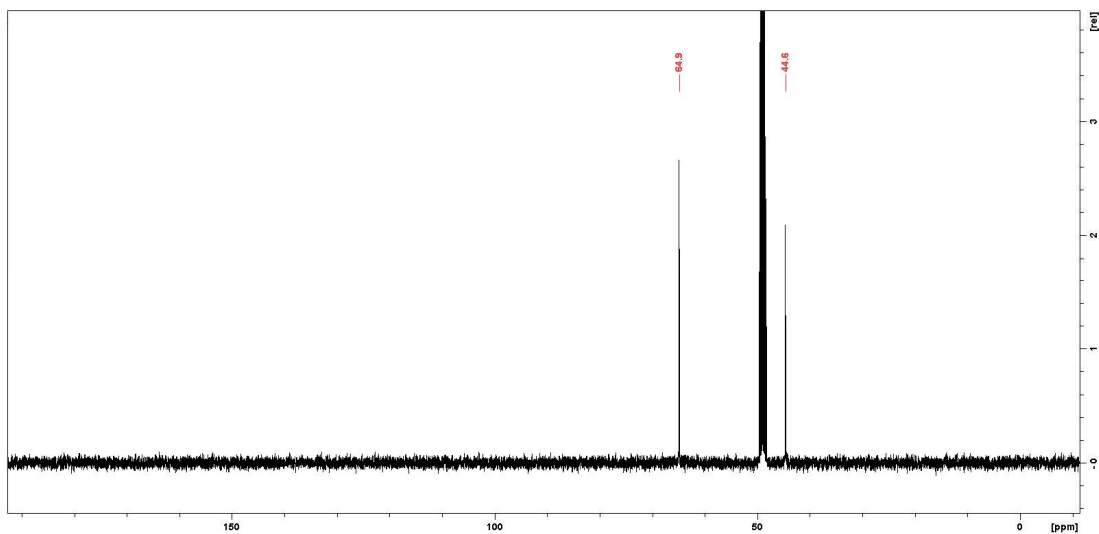


Figure 5.6.57. ¹³C-NMR spectrum of the final product of the oxidation of compound **SeA3**, *i.e.*, morpholine hydrochloride.

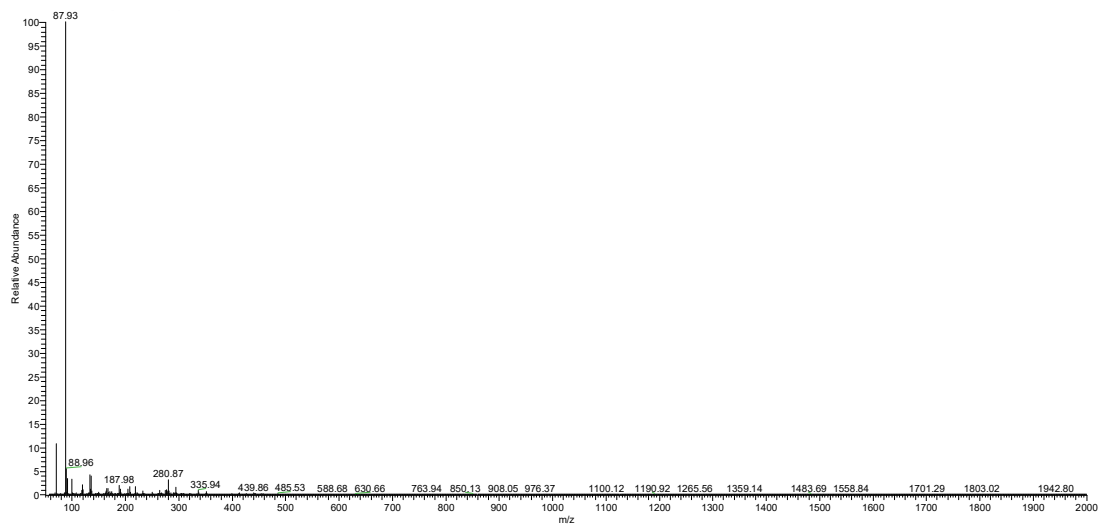


Figure 5.6.58. ESI-MS spectrum of the final product of the oxidation of compound **SeA3**, *i.e.*, morpholine hydrochloride.

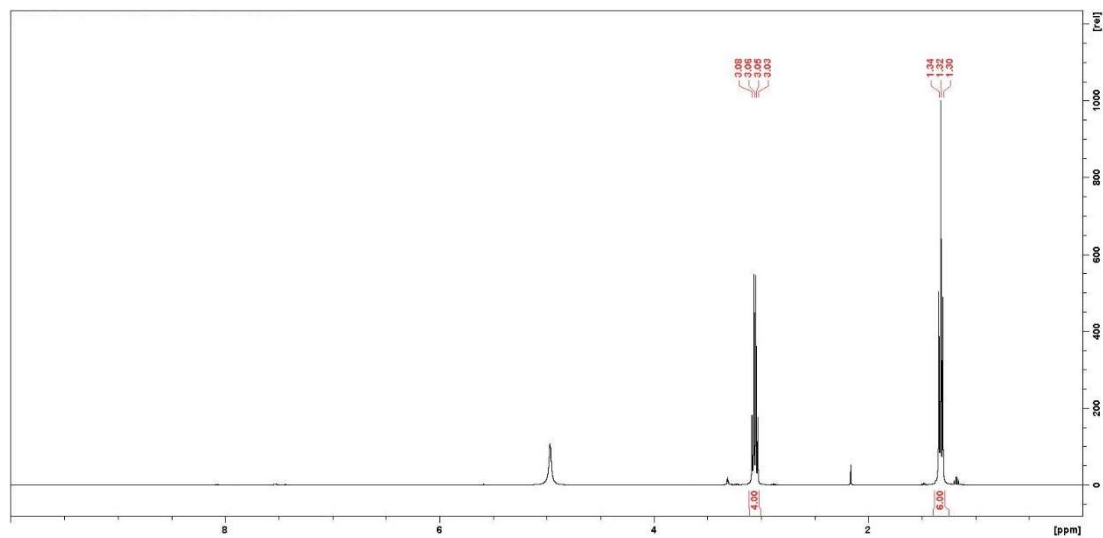


Figure 5.6.59. ¹H-NMR spectrum of the final product of the oxidation of compound **SeA1**, *i.e.*, diethylamine hydrochloride.

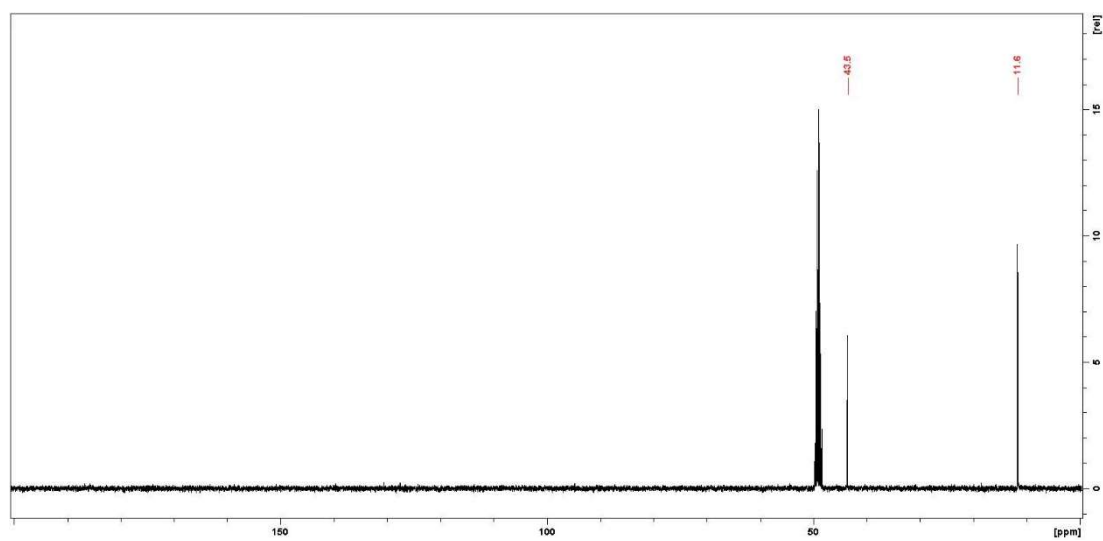


Figure 5.6.60. ¹³C-NMR spectrum of the final product of the oxidation of compound **SeA1**, *i.e.*, diethylamine hydrochloride.

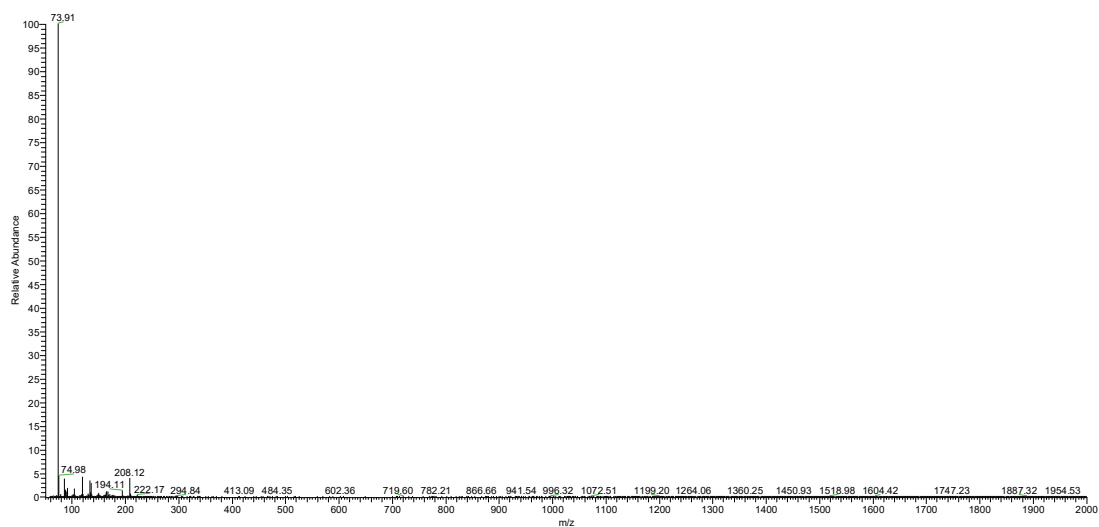


Figure 5.6.61. ESI-MS spectrum of the final product of the oxidation of compound **SeA1**, *i.e.*, diethylamine hydrochloride.

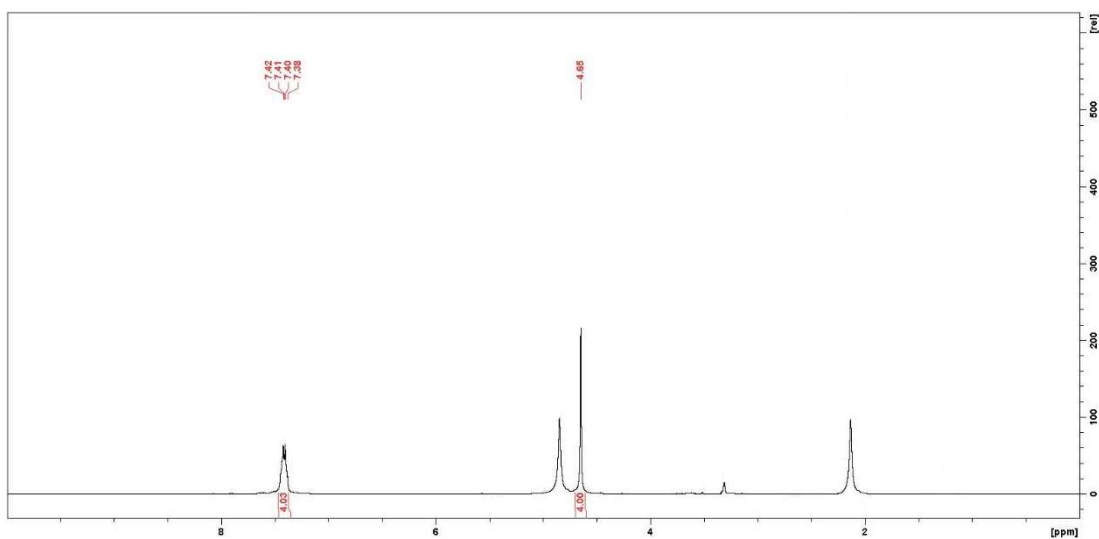


Figure 5.6.62. ¹H-NMR spectrum of the final product of the oxidation of compound **SeA4**, *i.e.*, isoindoline hydrochloride.

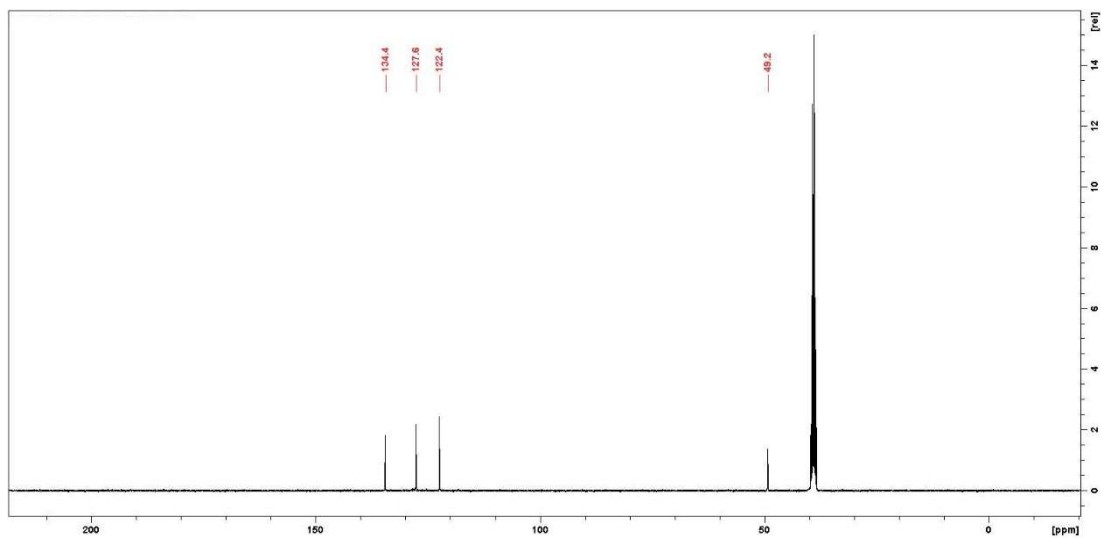


Figure 5.6.63. ¹³C-NMR spectrum of the final product of the oxidation of compound **SeA4**, *i.e.*, isoindoline hydrochloride.

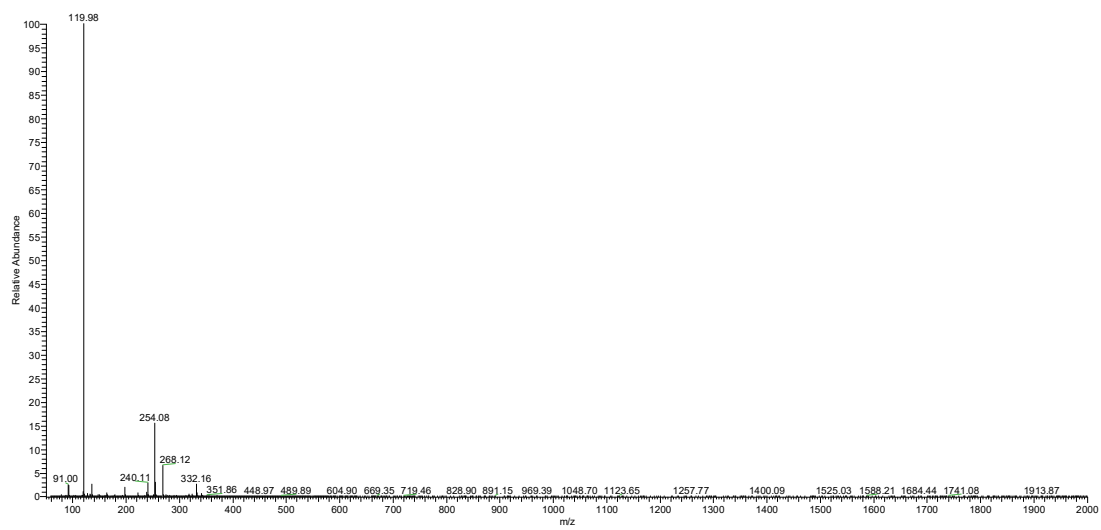


Figure 5.6.64. ESI-MS spectrum of the final product of the oxidation of compound **SeA4**, *i.e.*, isoindoline hydrochloride.

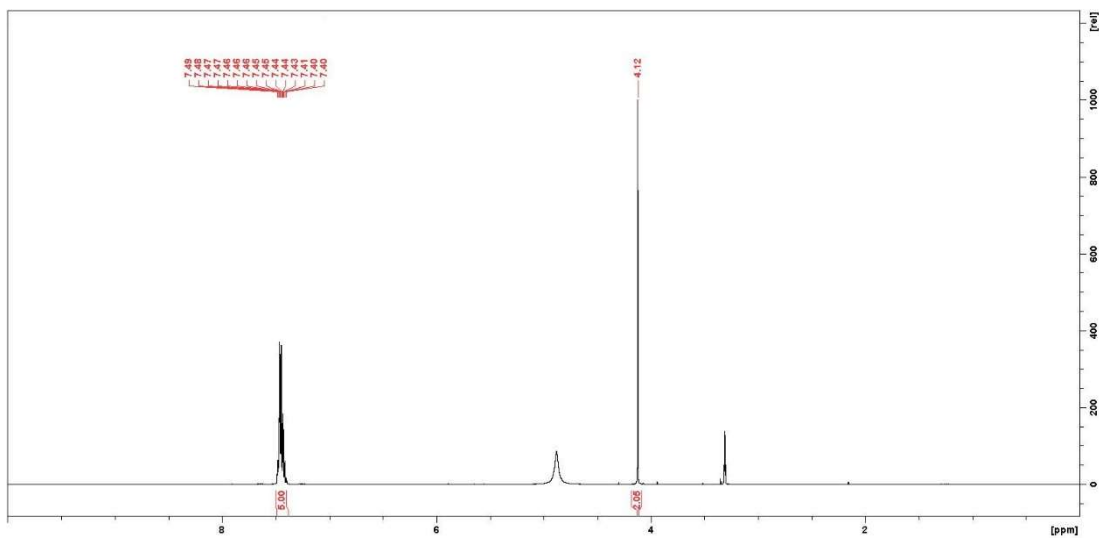


Figure 5.6.65. ^1H -NMR spectrum of the final product of the oxidation of compound **SeA5**, *i.e.*, benzylamine hydrochloride.

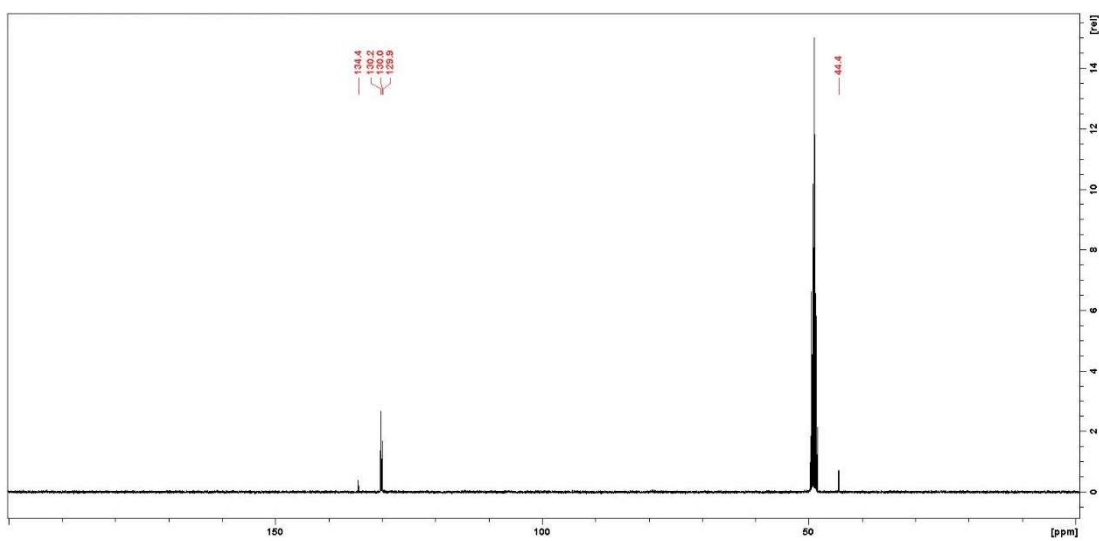


Figure 5.6.66. ^{13}C -NMR spectrum of the final product of the oxidation of compound **SeA5**, *i.e.*, benzylamine hydrochloride.

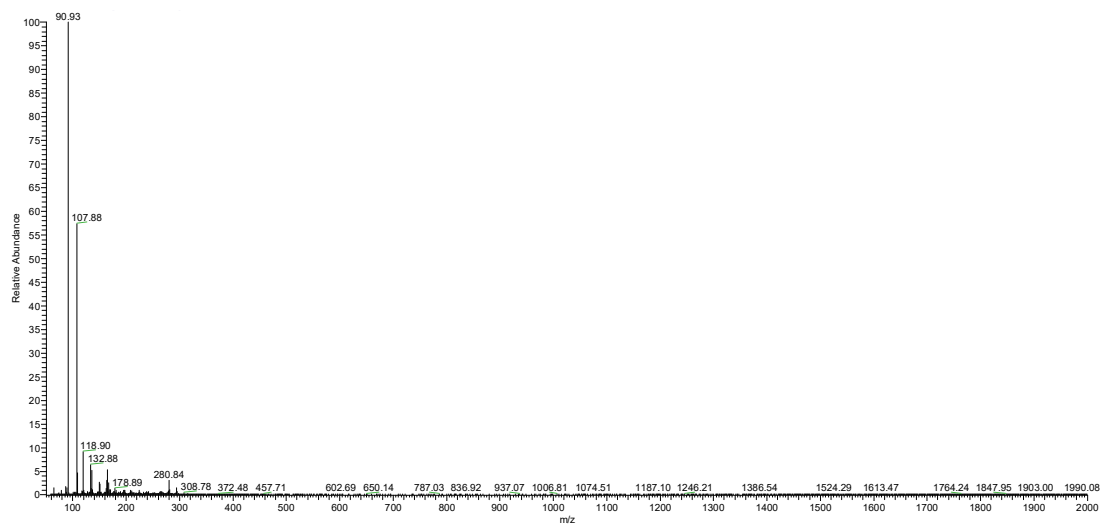


Figure 5.6.67. ESI-MS spectrum of the final product of the oxidation of compound **SeA5**, *i.e.*, benzylamine hydrochloride.

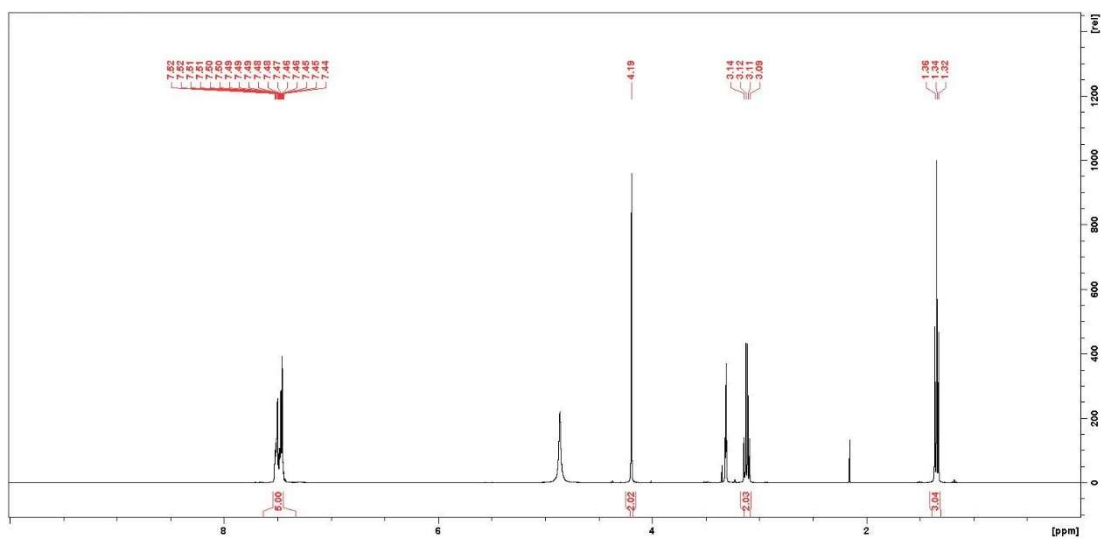


Figure 5.6.68. ¹H-NMR spectrum of the final product of the oxidation of compound **SeA7**, *i.e.*, N-ethylbenzylamine hydrochloride.

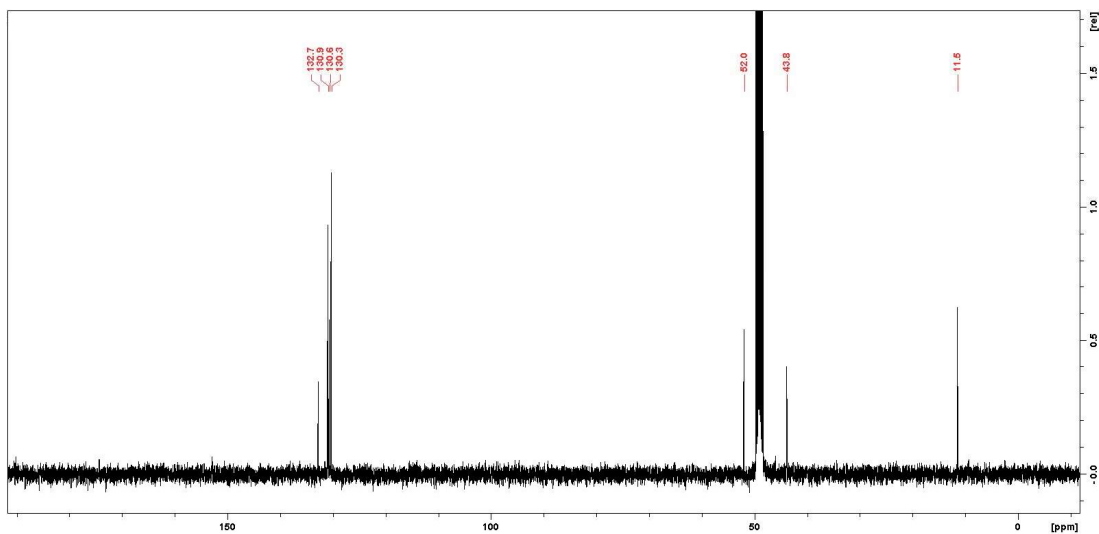


Figure 5.6.69. ^{13}C -NMR spectrum of the final product of the oxidation of compound **SeA7**, *i.e.*, N-ethylbenzylamine hydrochloride.

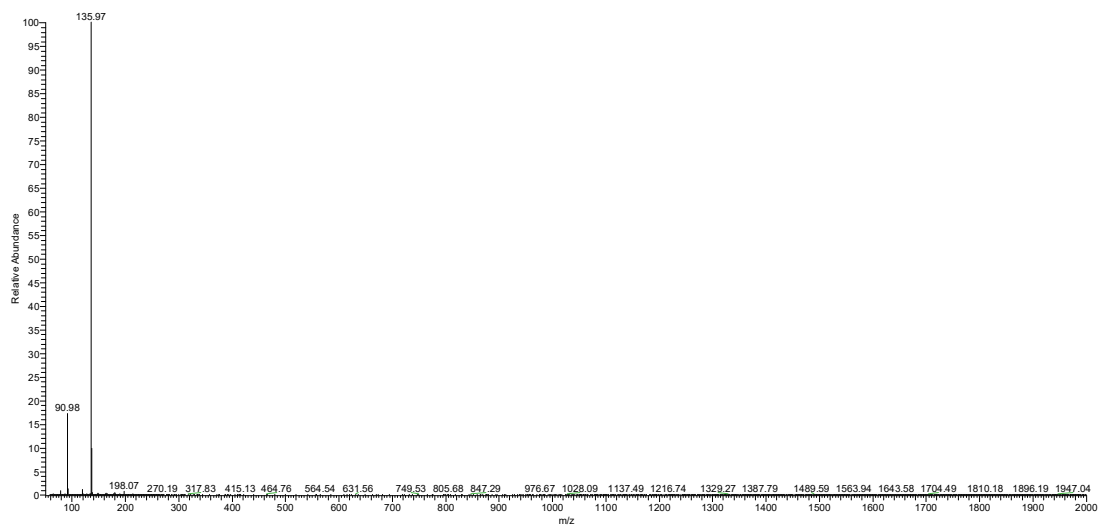


Figure 5.6.70. ESI-MS spectrum of the final product of the oxidation of compound **SeA7**, *i.e.*, N-ethylbenzylamine hydrochloride.

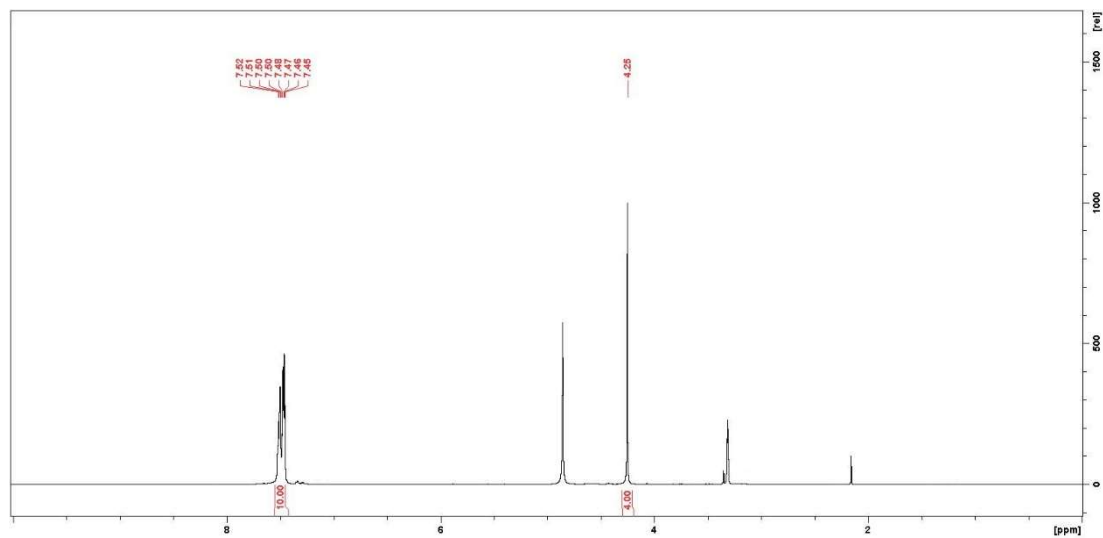


Figure 5.6.71. ¹H-NMR spectrum of the final product of the oxidation of compound **SeA2**, *i.e.*, dibenzylamine hydrochloride.

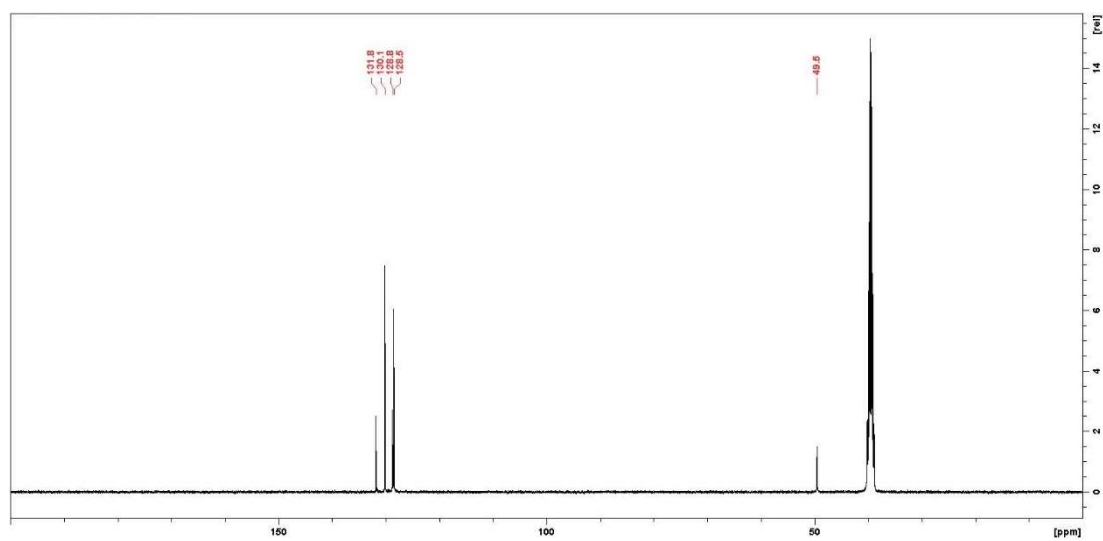


Figure 5.6.72. ¹³C-NMR spectrum of the final product of the oxidation of compound **SeA2**, *i.e.*, dibenzylamine hydrochloride.

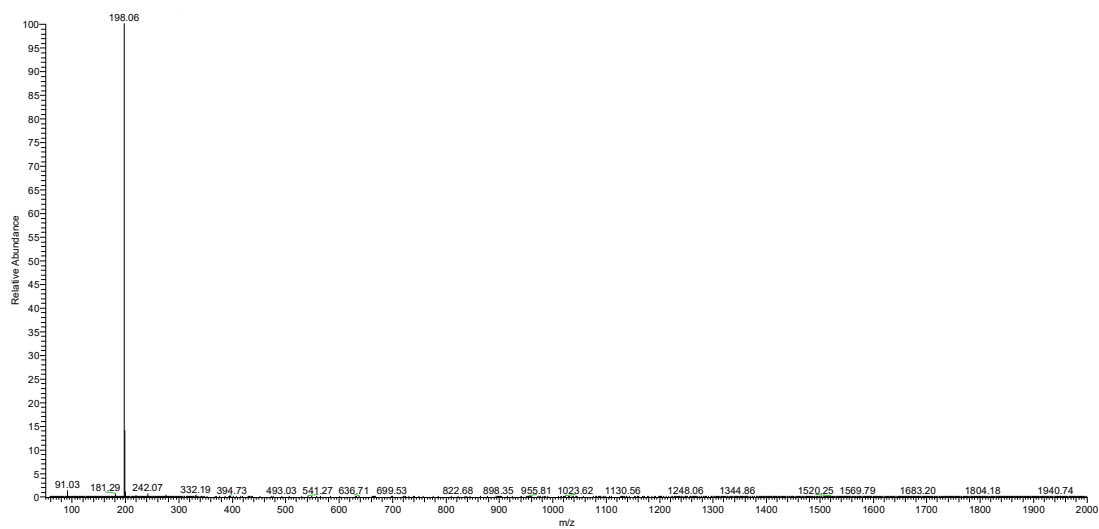


Figure 5.6.73. ESI-MS spectrum of the final product of the oxidation of compound **SeA2**, *i.e.*, dibenzylamine hydrochloride.

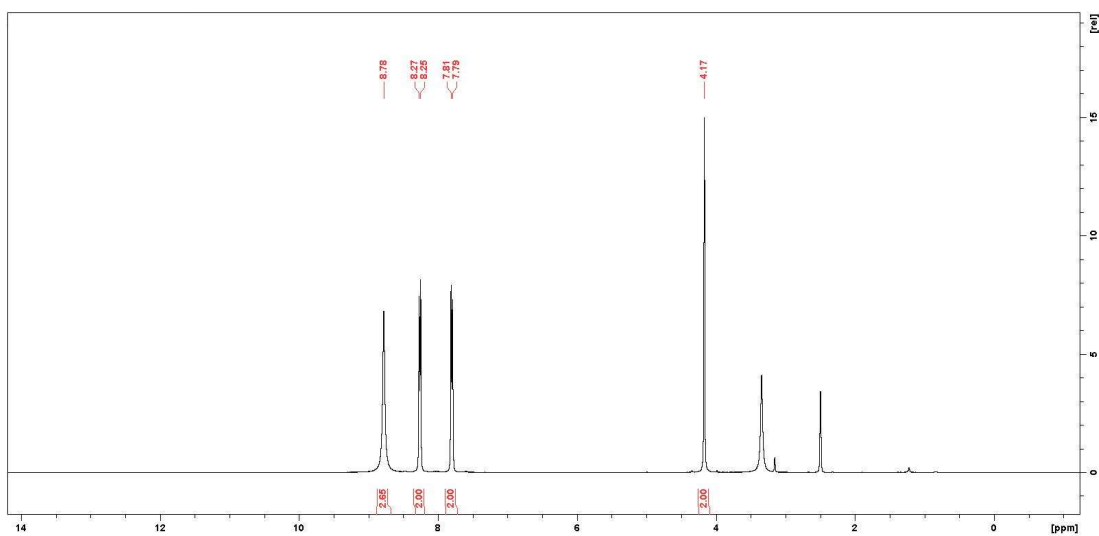


Figure 5.6.74. ¹H-NMR spectrum of the final product of the oxidation of compound **SeA6**, *i.e.*, *p*-nitrobenzylamine hydrochloride.

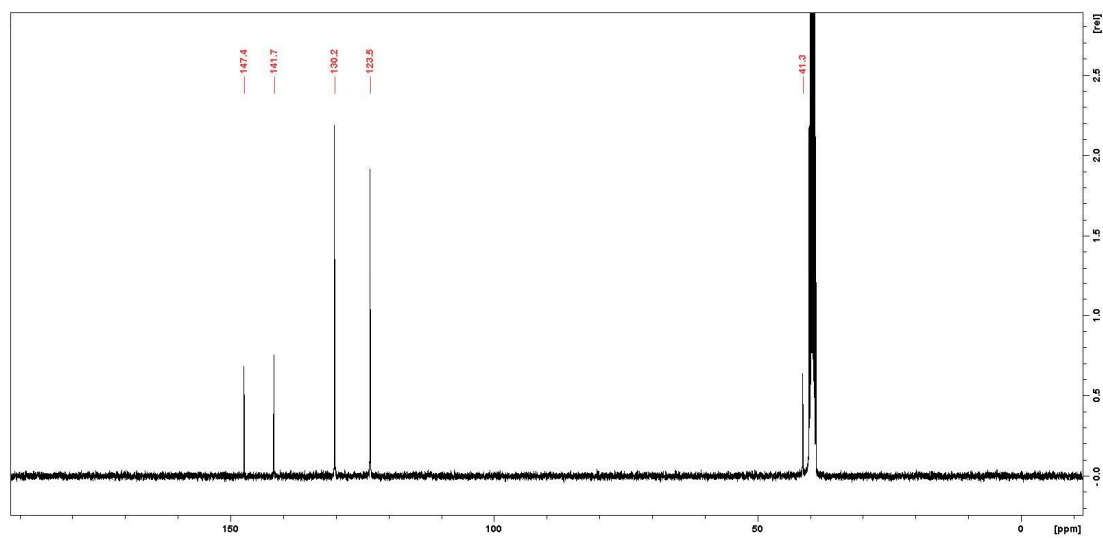


Figure 5.6.75. ^{13}C -NMR spectrum of the final product of the oxidation of compound **SeA6**, *i.e.*, *p*-nitrobenzylamine hydrochloride.

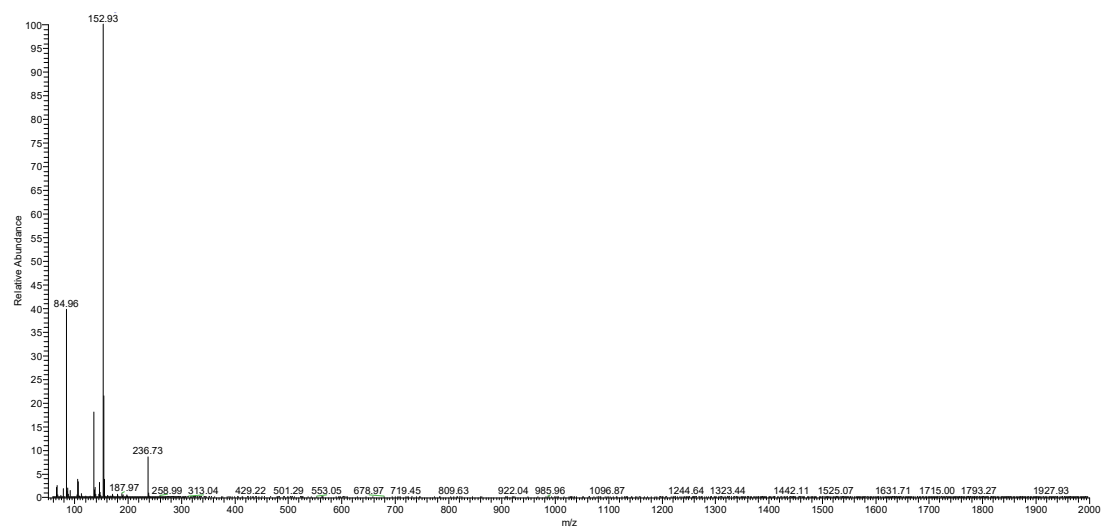


Figure 5.6.76. ESI-MS spectrum of the final product of the oxidation of compound **SeA6**, *i.e.*, *p*-nitrobenzylamine hydrochloride.

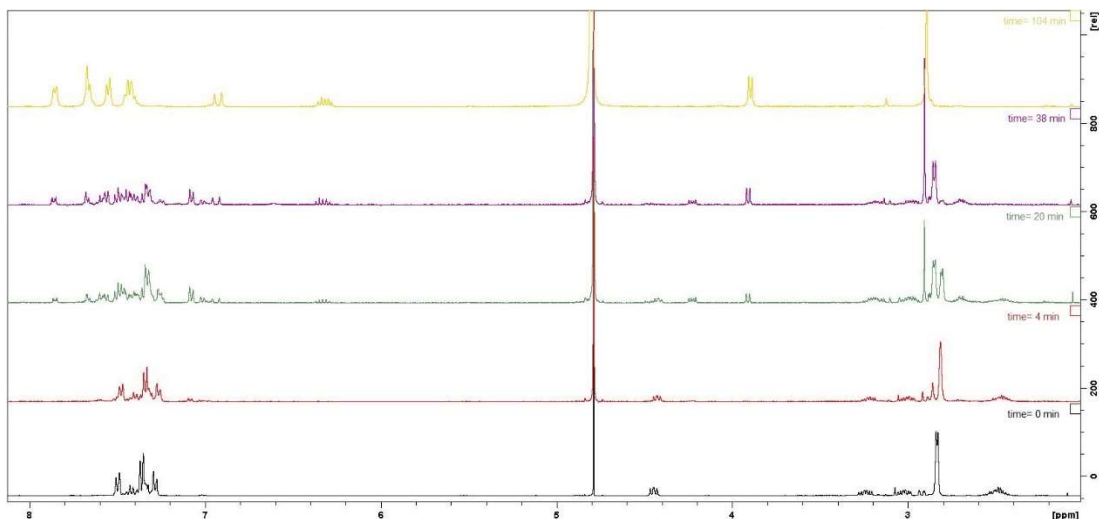


Figure 5.6.77. Variation over time of the ¹H-NMR spectra during the oxidation-reaction of compound **1-H**.

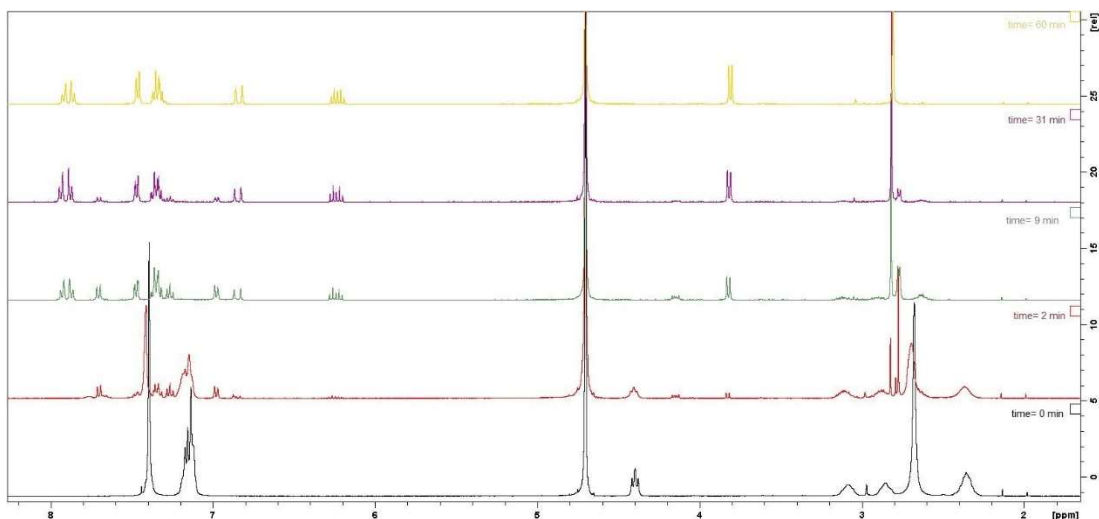


Figure 5.6.78. Variation over time of the ¹H-NMR spectra during the oxidation-reaction of compound **1-CF₃**.

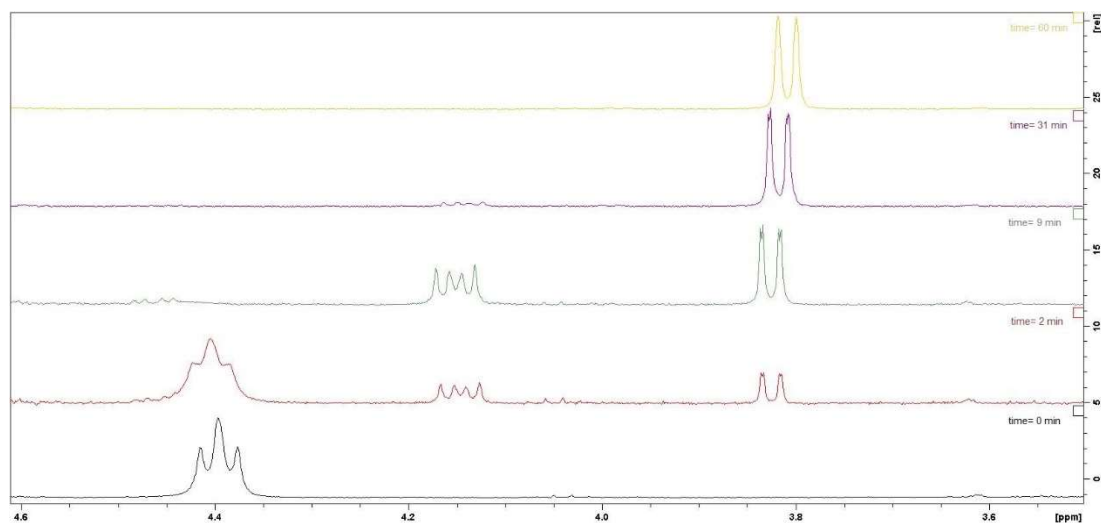


Figure 5.6.79. Variation over time of the ¹H-NMR spectra during the oxidation-reaction of compound **1-CF₃**, focus on the region between 4.6 and 3.5 ppm (area of interest to integrate for relative abundance over time calculation).

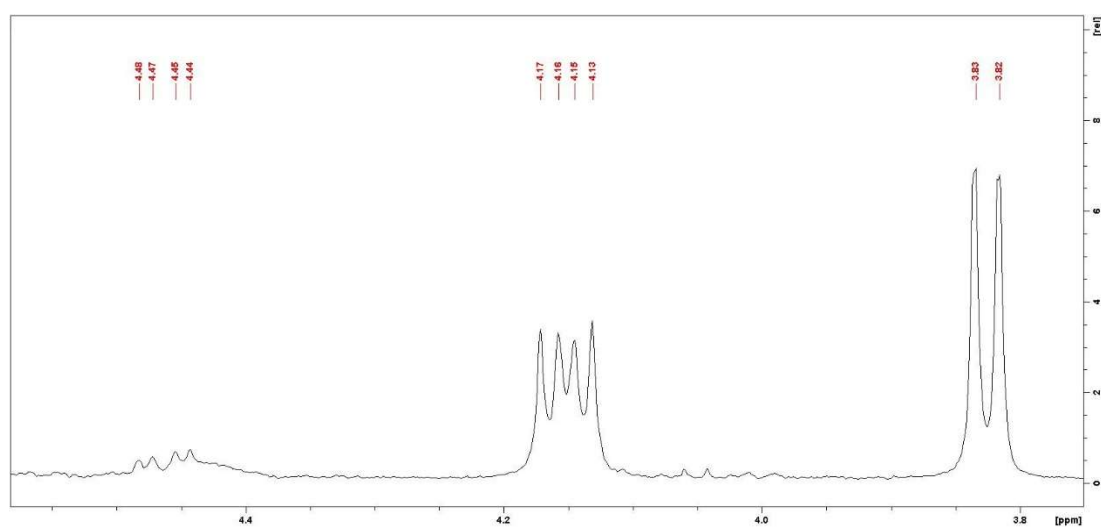


Figure 5.6.80. ¹H-NMR spectrum during the oxidation-reaction of compound **1-CF₃** at t= 9 min, focus on the region between 4.6 and 3.5 ppm (the signal corresponding to the other diastereoisomers of the selenoxide is detectable at 4.46 ppm after the disappearance of the starting material).

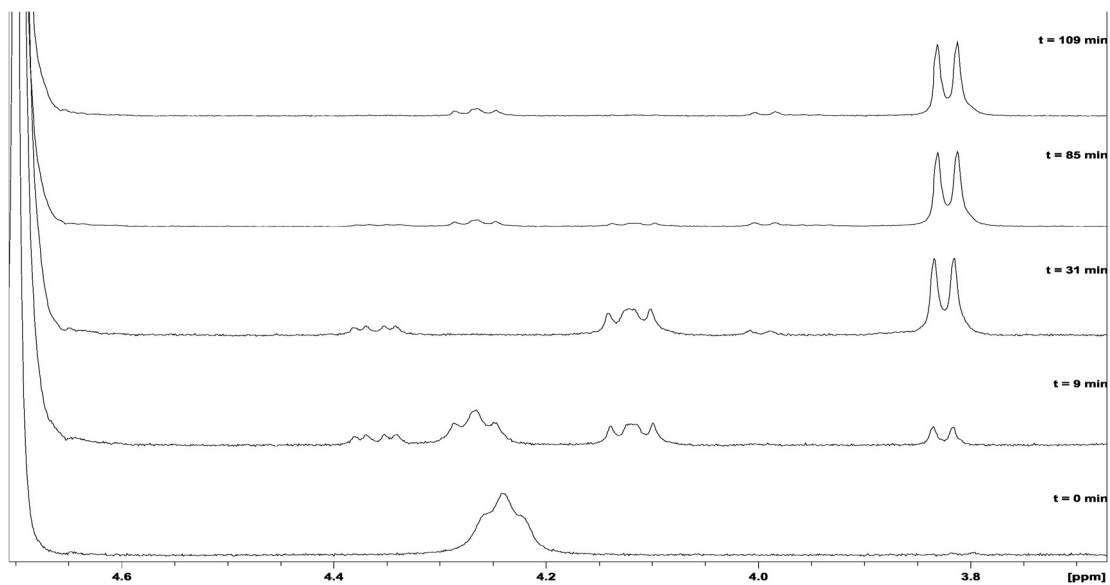


Figure 5.6.81. Variation over time of the ^1H -NMR spectra during the oxidation-reaction of compound **1-CH₃**, focus on the region between 4.7 and 3.7 ppm (area of interest to integrate for relative abundance over time calculation).

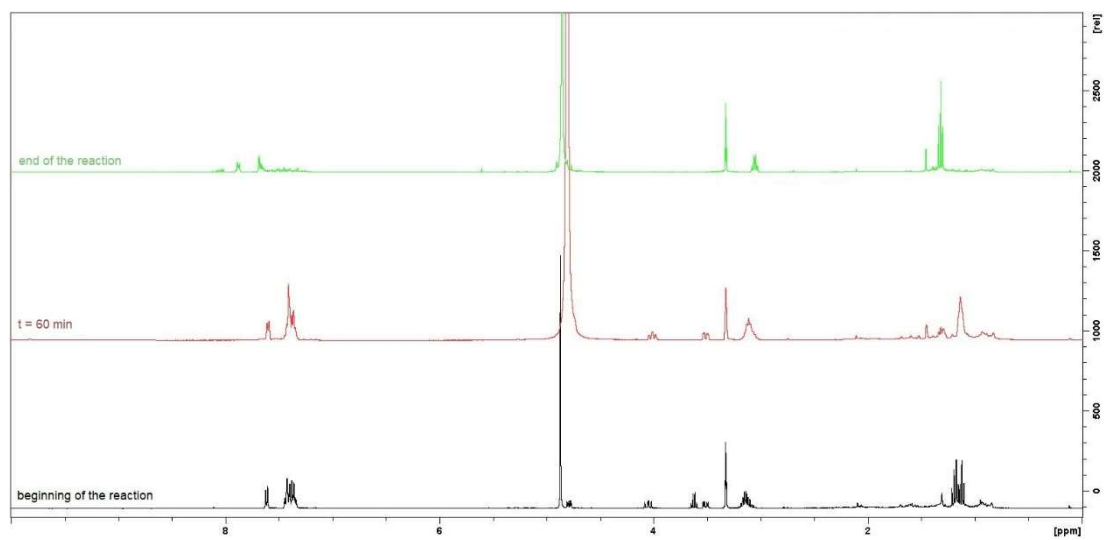


Figure 5.6.82. ^1H -NMR spectra of compound **SeA1** oxidation over time (• before the addition of H_2O_2 , • after 60 minutes from the addition of H_2O_2 , • overnight reaction).

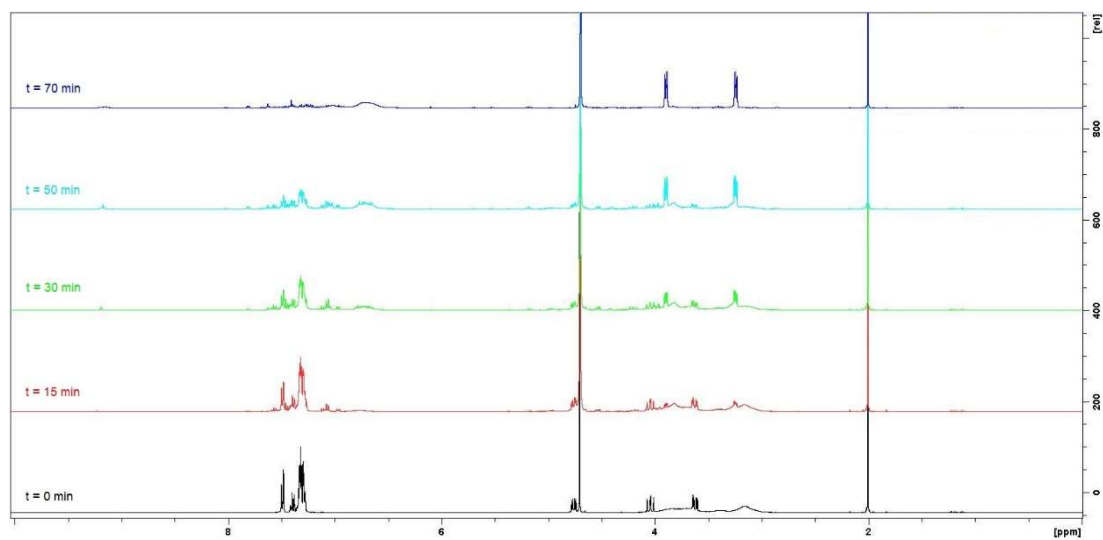


Figure 5.6.83. Full ¹H-NMR spectra over time of the oxidation of compound **SeA3**.

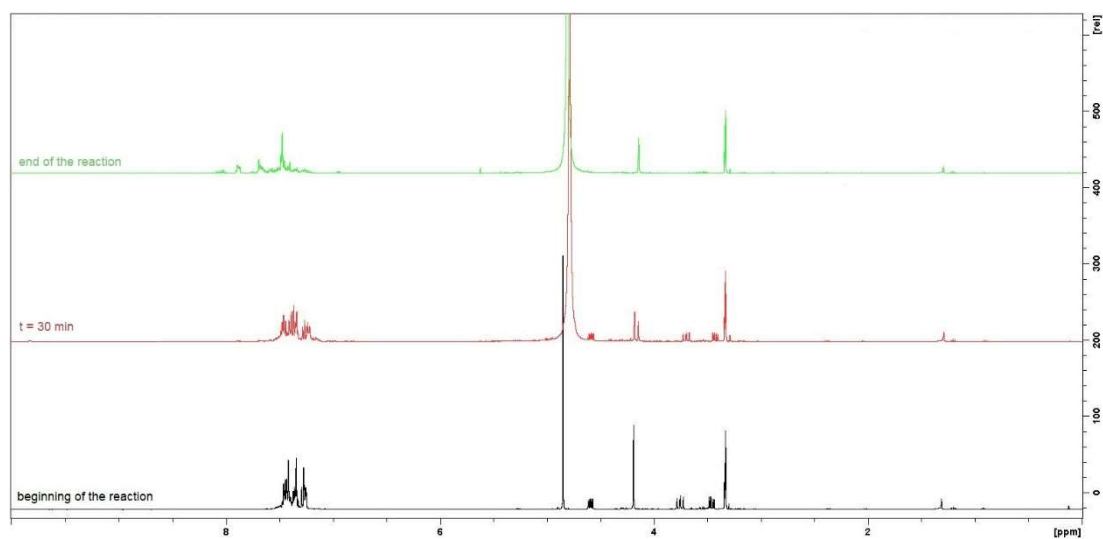


Figure 5.6.84. ¹H-NMR spectra of compound **SeA5** oxidation over time (• before the addition of H₂O₂, • after 30 minutes from the addition of H₂O₂, • overnight reaction).

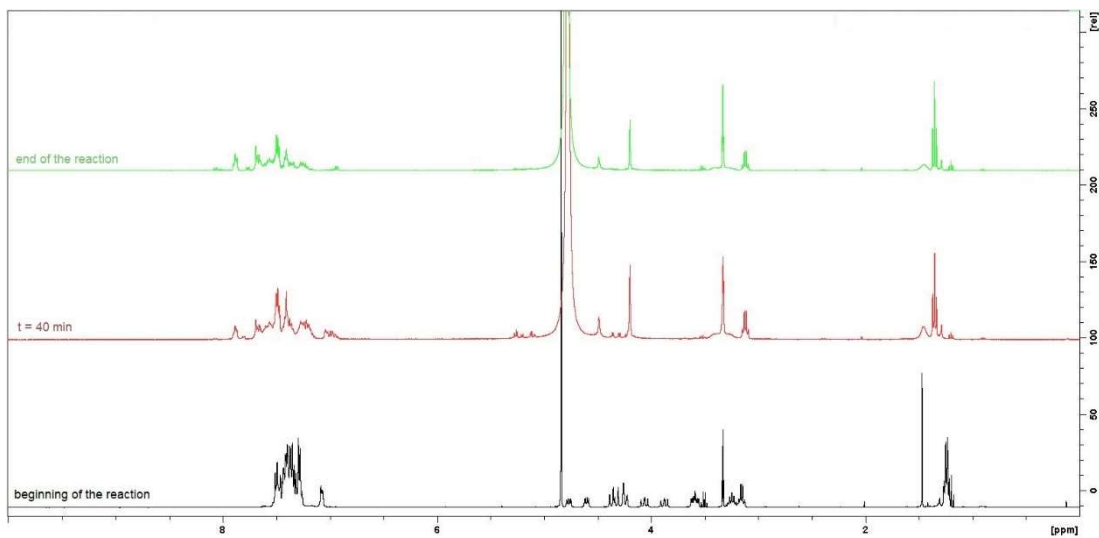


Figure 5.6.85. ^1H -NMR spectra of compound **SeA7** oxidation over time (• before the addition of H_2O_2 , • after 30 minutes from the addition of H_2O_2 , • overnight reaction).

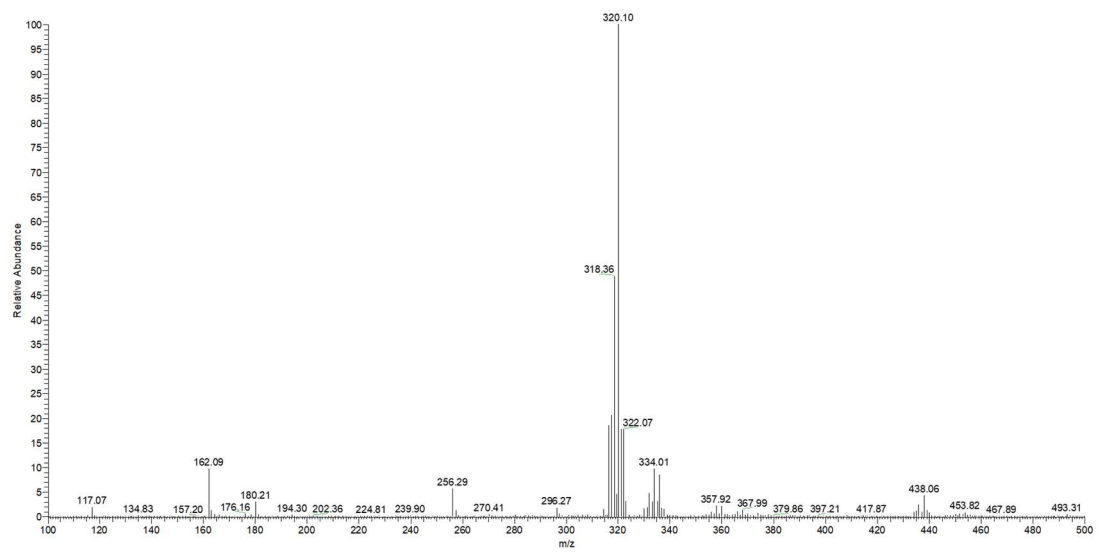


Figure 5.6.86. ESI-MS study of the reaction of compound **1-H** with H_2O_2 (positive ionization mode, $t = 5$ min). The peaks corresponding to the starting material ($m/z = 318$), the selenoxide ($m/z = 334$) and the cinnamylamine ($m/z = 162$) were detected.

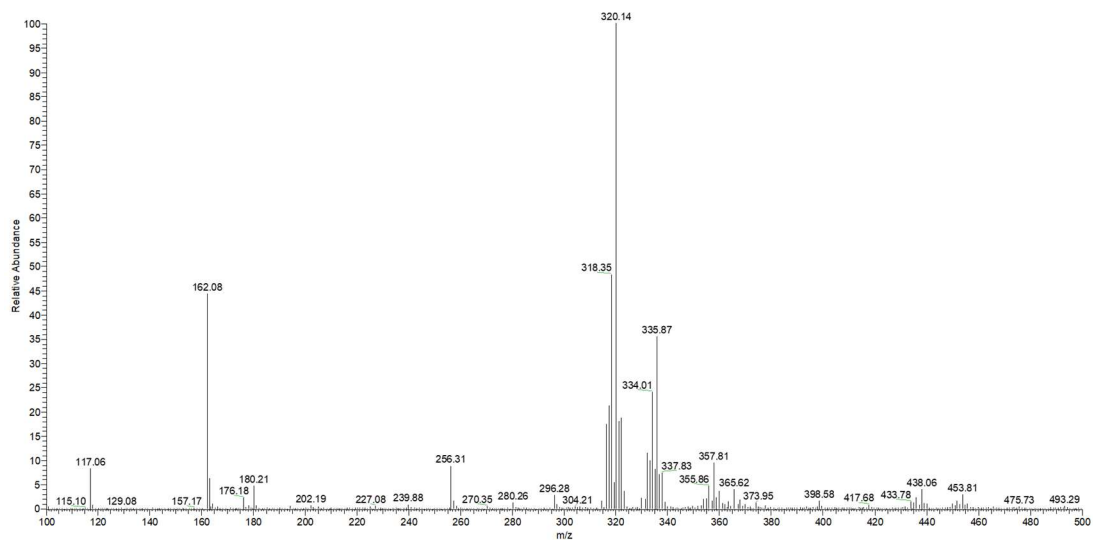


Figure 5.6.87. ESI-MS study of the reaction of compound **1-H** with H_2O_2 (positive ionization mode, $t = 20$ min). The peaks corresponding to the starting material ($m/z = 318$), the selenoxide and its sodium adduct ($m/z = 334$ and $m/z = 357$) and the cinnamylamine ($m/z = 162$) were detected.

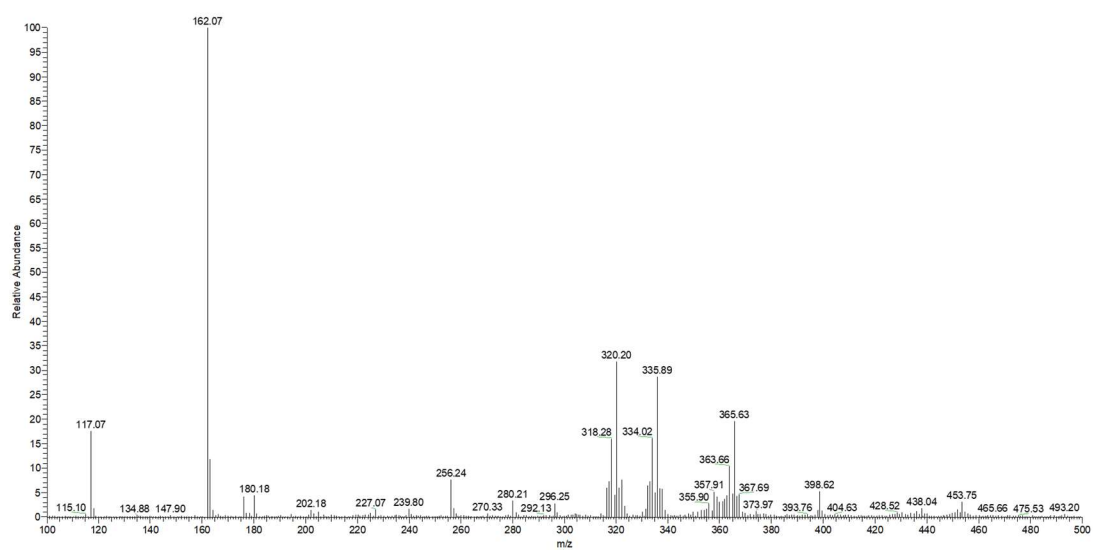


Figure 5.6.88. ESI-MS study of the reaction of compound **1-H** with H_2O_2 (positive ionization mode, $t = 60$ min). The peaks corresponding to the starting material ($m/z = 318$), the selenoxide and its sodium and methanol adduct ($m/z = 334$, $m/z = 357$, $m/z = 365$), together with the cinnamylamine ($m/z = 162$) were detected.

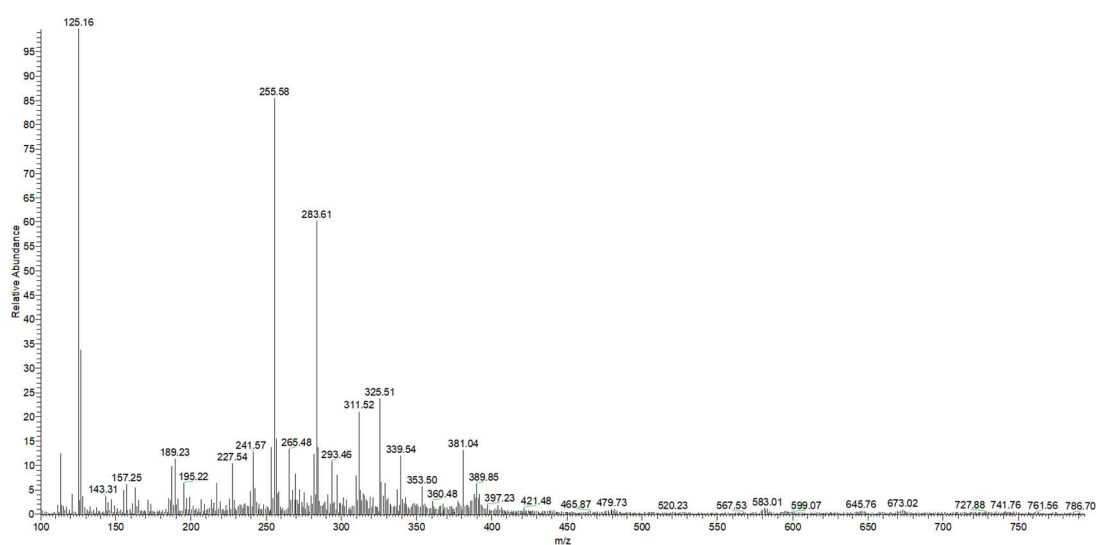


Figure 5.6.89. ESI-MS study of the reaction of compound **1-H** with H_2O_2 (negative ionization mode, $t = 10$ min). The seleninic acid ($m/z = 189$) is detectable.

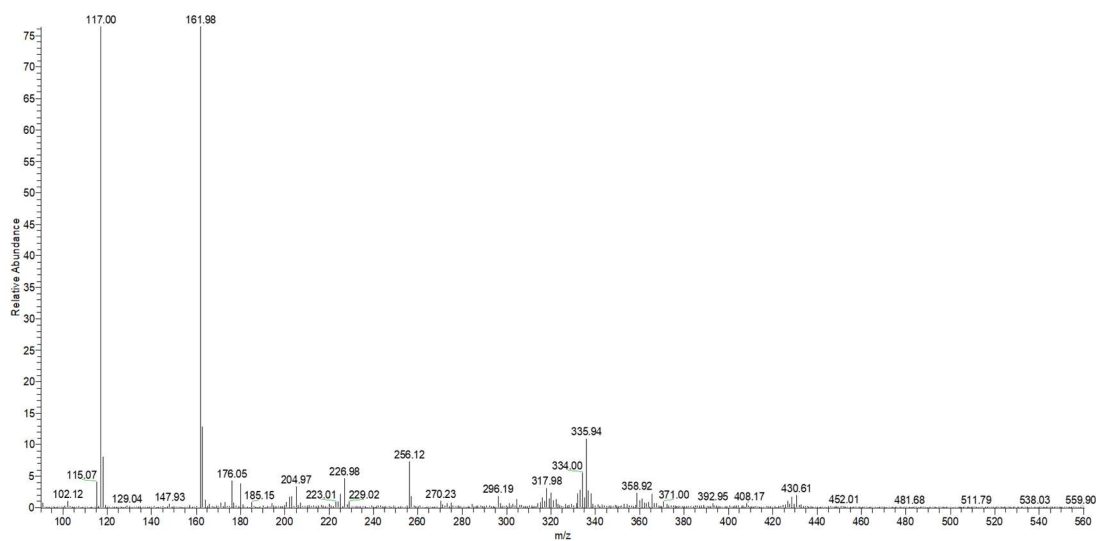


Figure 5.6.90. ESI-MS characterization of the NMR sample of compound **1-H** used for the NMR study (positive ionization mode, final timepoint). Cinnamylamine ($m/z = 162$), selenoxide ($m/z = 334$) and diphenyl diselenide ($m/z = 316$) are clearly detectable.

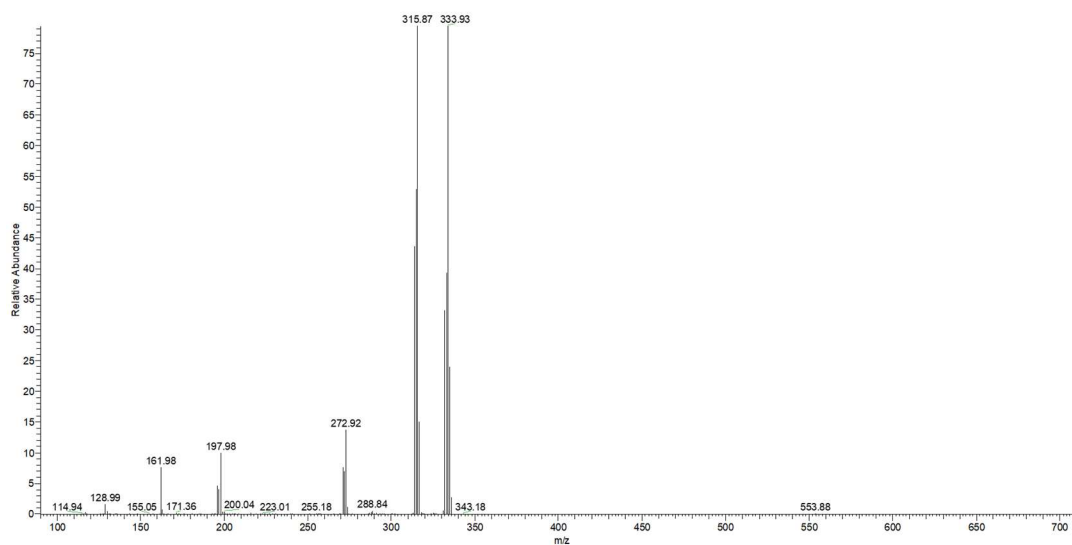


Figure 5.6.91. CID experiment showing the fragmentation pattern of the selenoxide (from compound **1-H**, positive ionization mode, parent ion $m/z = 334$, NCE = 16). The fragmentation promotes the formation of diphenyl diselenide ($m/z = 315$) and of the cinnamylamine ($m/z = 162$).

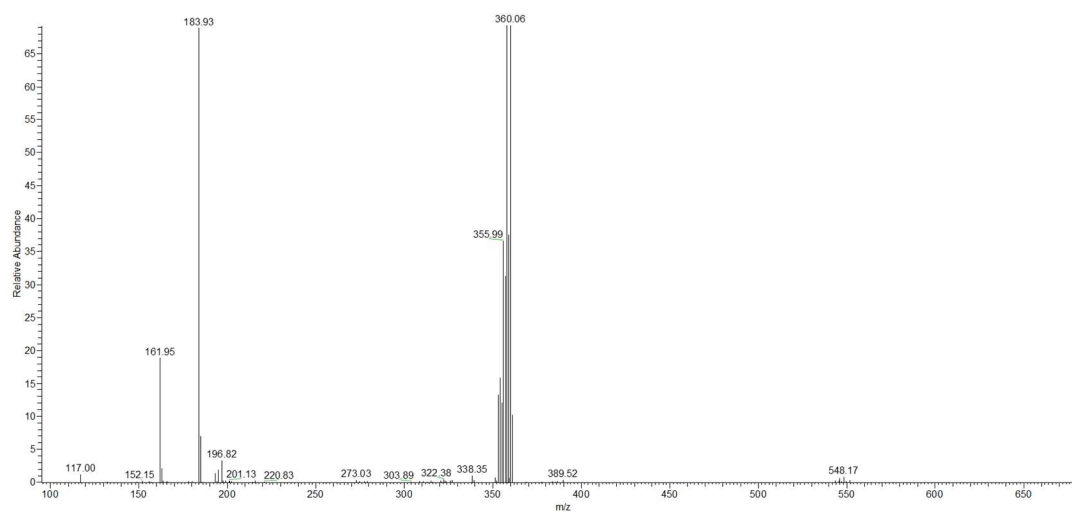


Figure 5.6.92. CID experiment showing the fragmentation pattern of the sodium adduct of the selenoxide (from compound **1-H**, positive ionization mode, parent ion $m/z = 357$, NCE = 16). The fragmentation promotes the formation of the cinnamylamine ($m/z = 162$) and of the cinnamylamine sodium adduct ($m/z = 184$).

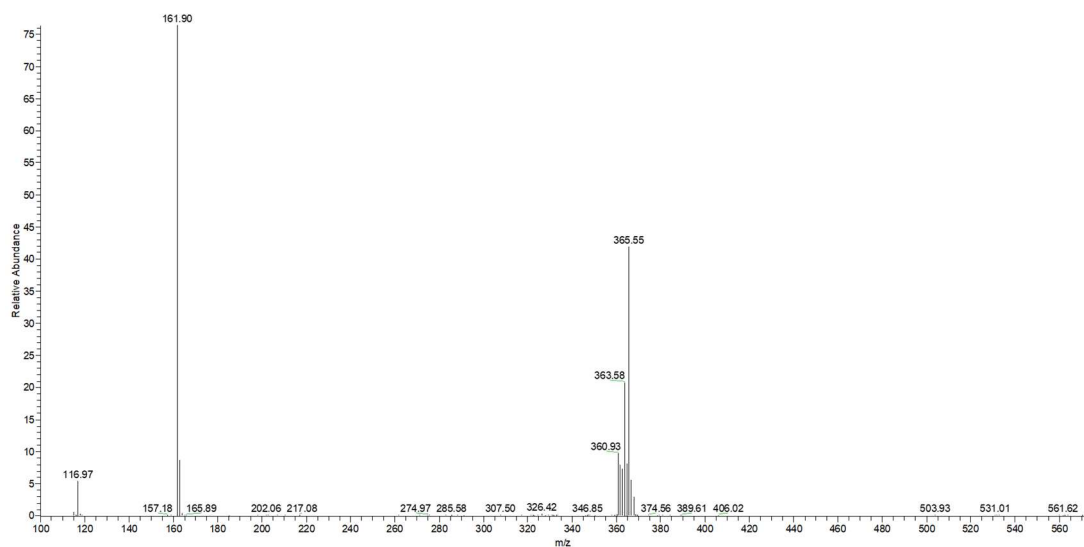


Figure 5.6.93. CID experiment showing the fragmentation pattern of the methanol adduct of the selenoxide (from compound **1-H**, positive ionization mode, parent ion $m/z = 365$, NCE = 12). The fragmentation promotes the formation of the cinnamylamines ($m/z = 162$).

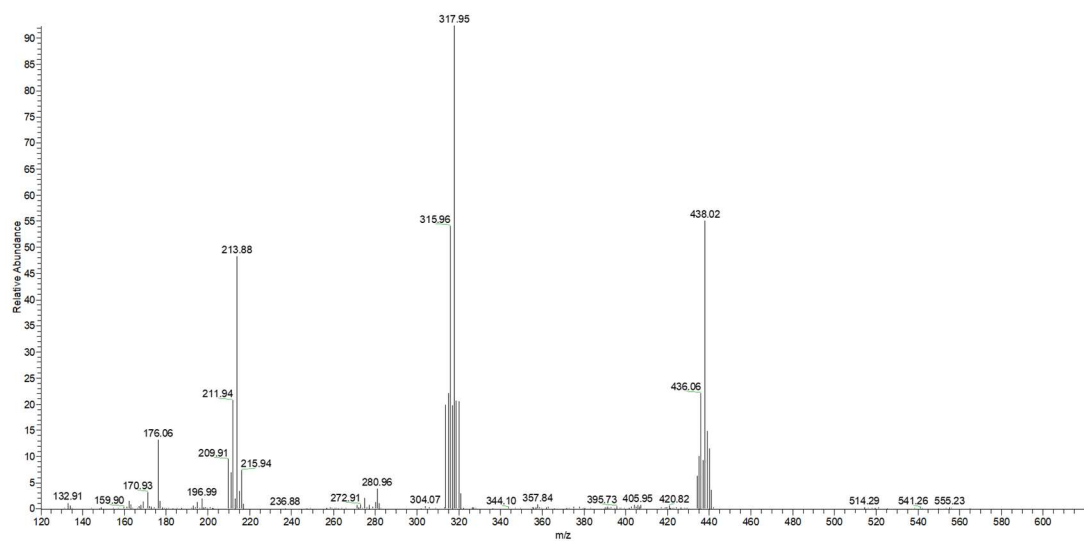


Figure 5.6.94. CID experiment showing the fragmentation pattern of the ion with $m/z = 438$ (from compound **1-H**, positive ionization mode, NCE = 24). The fragmentation promotes the formation of two Se-containing species ($m/z = 316$ and $m/z = 212$).

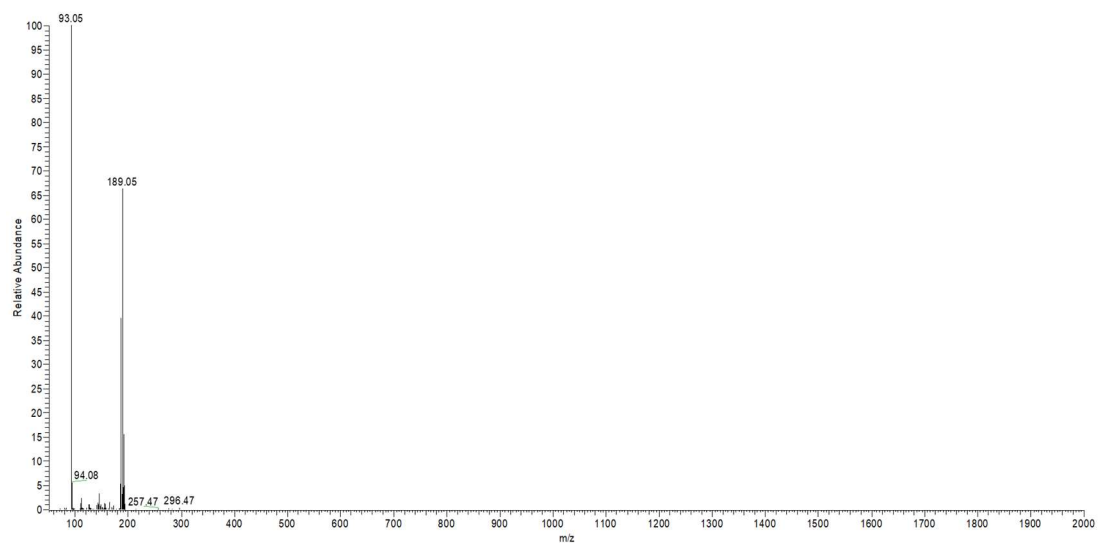


Figure 5.6.95. CID experiment showing the fragmentation pattern of the seleninic acid (from compound **1-H**, negative ionization mode, parent ion $m/z = 189$, NCE = 25). The fragmentation promotes the formation of phenol ($m/z = 93$).

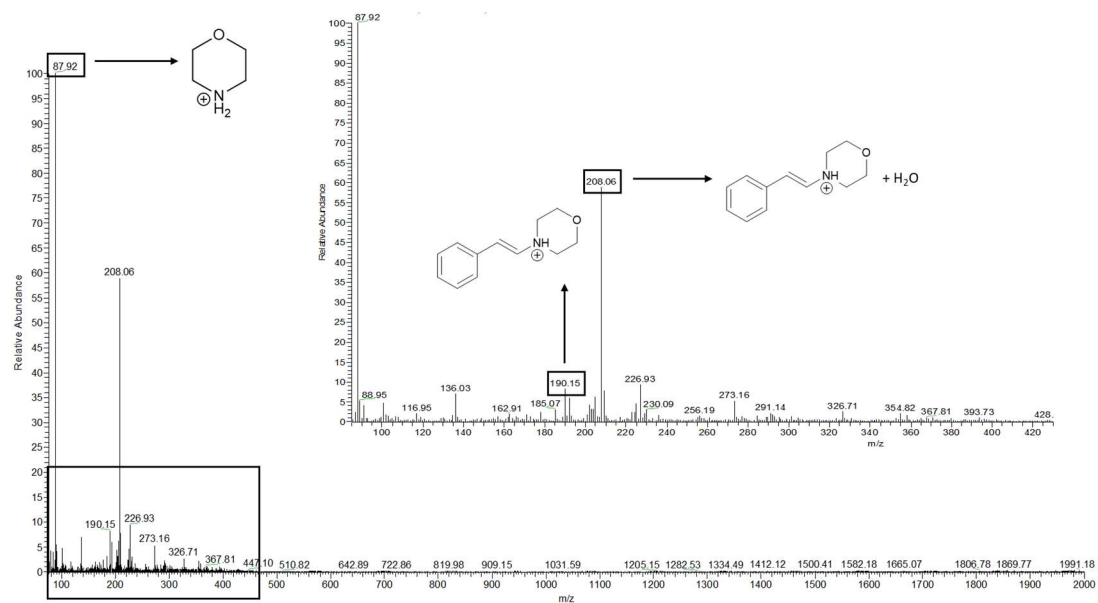


Figure 5.6.96. ESI-MS spectrum of the oxidation of compound **SeA3**. The peak at 87.92 m/z corresponds to the final product (*i.e.*, morpholine), the peaks at 190.15 and 208.06 m/z correspond respectively to the enamine and its adduct with water.

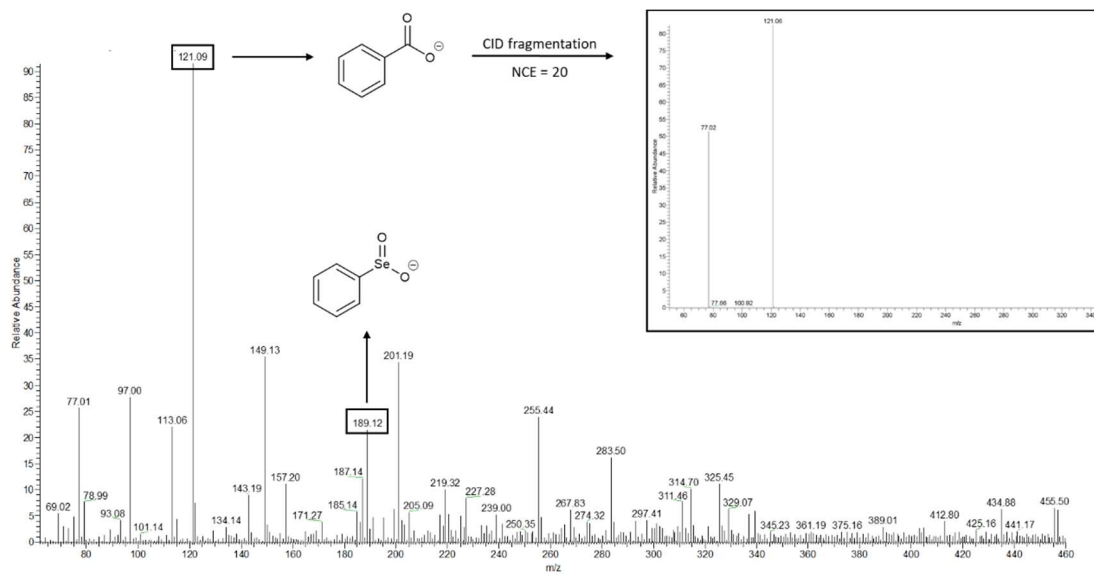


Figure 5.6.97. Negative ionization ESI-MS spectra of the oxidation of compound **SeA3**. The peak at 121.09 m/z correspond to benzoic acid (as confirmed by the Collision Induced Dissociation (CID) experiment at Normalized Collision Energy (NCE) = 20), the peak at 189.12 corresponds to the seleninic acid.

6 References

1. Zhang, Y. *et al. Int. J. Mol. Sci.* **23**, 13568 (2022).
2. Holbrook, S. Y. L. & Garneau-Tsodikova, S. *MedChemComm* **8**, 1739–1741 (2017).
3. Parascandola, J. *Studies in the history of modern pharmacology and drug therapy*. Taylor & Francis, (2023).
4. Watson, J. T. & Sparkman, O. D. *Introduction to mass spectrometry: instrumentation, applications, and strategies for data interpretation*. John Wiley & Sons, (2007).
5. Jennings, K. R. & Dolnikowski, G. G. *Methods in enzymology* vol. 193 37–61 Elsevier, (1990).
6. Mirzaei, H. & Carrasco, M. *Modern proteomics-sample preparation, analysis and practical applications*. **919**, Springer, (2016).
7. Wait, R. *Spectrosc. Methods Anal. NMR Mass Spectrom. Met. Tech.* 191–213 (1993).
8. March, R. E. *J. Mass Spectrom.* **32**, 351–369 (1997).
9. Arpino, P. J. & Guiochon, G. *Anal. Chem.* **51**, 682–701 (1979).
10. Blakley, C. & Vestal, M. *Anal. Chem.* **55**, 750–754 (1983).
11. Wong, P. S. & Graham Cooks, R. *Curr. Sep.* **16**, 85–92 (1997).
12. Comisarow, M. B. & Marshall, A. G. *Chem. Phys. Lett.* **25**, 282–283 (1974).
13. De Vijlder, T. *et al. Mass Spectrom. Rev.* **37**, 607–629 (2018).
14. Sheldon, M. T., Mistrik, R. & Croley, T. R. *J. Am. Soc. Mass Spectrom.* **20**, 370–376 (2011).
15. Keeler, J. *Understanding NMR spectroscopy*. John Wiley & Sons, (2010).
16. Akitt, J. W. & Mann, B. E. *NMR and Chemistry: An introduction to modern NMR spectroscopy*. Crc Press, (2017).
17. Bruice, P. Y. *Organic chemistry*. Pearson, (2017).
18. Pochapsky, T. C. & Pochapsky, S. *NMR for physical and biological scientists*. Garland Science, (2006).
19. Hoeck, C. R. *Solving a 3D Structural Puzzle*. Springer, (2018).
20. *Modern Biophysical Chemistry* ed. Walla, P. J. 121–148 Wiley, (2014).
21. Worsfold, P., Townshend, A., Poole, C. F. & Miró, M. *Encyclopedia of analytical science*. Elsevier, (2019).
22. Nolis, P. & Parella, T. *Magn. Reson. Chem.* **57**, S85–S94 (2019).
23. Kapaev, R. R. & Toukach, P. V. (2016).
24. Gáspári, Z. & Perczel, A. *Annu. Rep. NMR Spectrosc.* **71**, 35–75 (2010).

25. Hilton, B. D. & Martin, G. E. *J. Nat. Prod.* **73**, 1465–1469 (2010).
26. Barsukov, I. L. *Encyclopedia of Biophysics* ed. Roberts, G. C. K. 1724–1727 Springer Berlin Heidelberg, (2013).
27. Shi, L. & Zhang, N. *Molecules* **26**, 576 (2021).
28. Barile, E. & Pellicchia, M. *Chem. Rev.* **114**, 4749–4763 (2014).
29. James, A. D., Marvalín, C., Luneau, A., Meissner, A. & Camenisch, G. *Drug Metab. Dispos.* **45**, 900–907 (2017).
30. Salmaso, V. & Moro, S. *Front. Pharmacol.* **9**, 923 (2018).
31. Prieto-Martínez, F. D., López-López, E., Eurídice Juárez-Mercado, K. & Medina-Franco, J. L. *In Silico Drug Design* 19–44 Elsevier, (2019).
32. Ou-Yang, S. *et al. Acta Pharmacol. Sin.* **33**, 1131–1140 (2012).
33. Sharma, V., Wakode, S. & Kumar, H. *Chemoinformatics and Bioinformatics in the Pharmaceutical Sciences* 27–53 Elsevier, (2021).
34. Śledź, P. & Cafilisch, A. *Curr. Opin. Struct. Biol.* **48**, 93–102 (2018).
35. Halgren, T. A. *J. Chem. Inf. Model.* **49**, 377–389 (2009).
36. Harigua-Souiai, E. *et al. BMC Bioinformatics* **16**, 1–15 (2015).
37. Meng, X.-Y., Zhang, H.-X., Mezei, M. & Cui, M. *Curr. Comput. Aided Drug Des.* **7**, 146–157 (2011).
38. Brint, A. T. & Willett, P. J. *Chem. Inf. Comput. Sci.* **27**, 152–158 (1987).
39. Rarey, M., Kramer, B., Lengauer, T. & Klebe, G. *J. Mol. Biol.* **261**, 470–489 (1996).
40. DesJarlais, R. L., Sheridan, R. P., Dixon, J. S., Kuntz, I. D. & Venkataraghavan, R. *J. Med. Chem.* **29**, 2149–2153 (1986).
41. Leach, A. R. & Kuntz, I. D. *J. Comput. Chem.* **13**, 730–748 (1992).
42. Goodsell, D. S., Lauble, H., Stout, C. D. & Olson, A. J. *Proteins Struct. Funct. Bioinforma.* **17**, 1–10 (1993).
43. Hart, T. N. & Read, R. J. *Proteins Struct. Funct. Bioinforma.* **13**, 206–222 (1992).
44. Kollman, P. *Chem. Rev.* **93**, 2395–2417 (1993).
45. Åqvist, J., Luzhkov, V. B. & Brandsdal, B. O. *Acc. Chem. Res.* **35**, 358–365 (2002).
46. Carlson, H. A. & Jorgensen, W. L. *J. Phys. Chem.* **99**, 10667–10673 (1995).
47. Böhm, H.-J. *J. Comput. Aided Mol. Des.* **12**, 309–309 (1998).
48. Gehlhaar, D. K. *et al. Chem. Biol.* **2**, 317–324 (1995).
49. Verkhivker, G. M. *et al. J. Comput. Aided Mol. Des.* **14**, 731–751 (2000).
50. Jain, A. N. *J. Comput. Aided Mol. Des.* **10**, 427–440 (1996).
51. Head, R. D. *et al. J. Am. Chem. Soc.* **118**, 3959–3969 (1996).

-
52. Muegge, I. & Martin, Y. C. *J. Med. Chem.* **42**, 791–804 (1999).
 53. Mitchell, J. B., Laskowski, R. A., Alex, A. & Thornton, J. M. *J. Comput. Chem.* **20**, 1165–1176 (1999).
 54. Feher, M., Deretey, E. & Roy, S. *J. Chem. Inf. Comput. Sci.* **43**, 1316–1327 (2003).
 55. Wallqvist, A., Jernigan, R. & Covell, D. *Protein Sci.* **4**, 1881–1903 (1995).
 56. DeWitte, R. S. & Shakhnovich, E. I. *J. Am. Chem. Soc.* **118**, 11733–11744 (1996).
 57. Srinivasan, J., Cheatham, T. E., Cieplak, P., Kollman, P. A. & Case, D. A. *J. Am. Chem. Soc.* **120**, 9401–9409 (1998).
 58. Still, W. C., Tempczyk, A., Hawley, R. C. & Hendrickson, T. *J. Am. Chem. Soc.* **112**, 6127–6129 (1990).
 59. Karplus, M. & McCammon, J. A. *Nat. Struct. Biol.* **9**, 646–652 (2002).
 60. Hollingsworth, S. A. & Dror, R. O. *Neuron* **99**, 1129–1143 (2018).
 61. Salomon-Ferrer, R., Gotz, A. W., Poole, D., Le Grand, S. & Walker, R. C. *J. Chem. Theory Comput.* **9**, 3878–3888 (2013).
 62. *Molecular Dynamics Simulation - Profacgen*. <https://www.profacgen.com/molecular-dynamics-simulation.htm>.
 63. *GROMACS 4.5 Online Reference*. <https://manual.gromacs.org/archive/4.5/online.html>.
 64. Bitar, M. & Offman, M.
 65. Brooks, B. R. *et al. J. Comput. Chem.* **4**, 187–217 (1983).
 66. Hess, B., Kutzner, C., Van Der Spoel, D. & Lindahl, E. *J. Chem. Theory Comput.* **4**, 435–447 (2008).
 67. Phillips, J. C. *et al. J. Comput. Chem.* **26**, 1781–1802 (2005).
 68. Case, D. A. *et al. J. Chem. Inf. Model.* (2023).
 69. Martínez, L. *PLOS ONE* **10**, e0119264 (2015).
 70. Da Fonseca, A. M. *et al. Mol. Biotechnol.* (2023).
 71. Hendrickson, J. B. *Science* **252**, 1189–1190 (1991).
 72. Leach, A. R. & Gillet, V. J. *An introduction to chemoinformatics*. Springer, (2007).
 73. Keiser, M. J. *et al. Nat. Biotechnol.* **25**, 197–206 (2007).
 74. Akamatsu, M. *Curr. Top. Med. Chem.* **2**, 1381–1394 (2002).
 75. Wold, S., Esbensen, K. & Geladi, P. *Chemom. Intell. Lab. Syst.* **2**, 37–52 (1987).
 76. Geladi, P. & Kowalski, B. R. *Anal. Chim. Acta* **185**, 1–17 (1986).
 77. Gramatica, P. *QSAR Comb. Sci.* **26**, 694–701 (2007).

-
78. Wermuth, C.-G., Ganellin, C., Lindberg, P. & Mitscher, L. *Pure Appl. Chem.* **70**, 1129–1143 (1998).
 79. Sharma, V., Kumar, H. & Wakode, S. *RSC Adv.* **6**, 75805–75819 (2016).
 80. Al-Sha'er, M. A. & Taha, M. O. *J. Chem. Inf. Model.* **50**, 1706–1723 (2010).
 81. Voet, A. *et al. J. Recept. Ligand Channel Res.* 81 (2014).
 82. *Encyclopedia of spectroscopy and spectrometry.* eds. Lindon, J. C., Tranter, G. E. & Koppenaal, D. W. Elsevier/AP, Academic Press is an imprint of Elsevier, (2017).
 83. Gore, M. *Spectrophotometry and spectrofluorimetry: a practical approach.* **225**, OUP Oxford, (2000).
 84. (2011).
 85. Barron, L. D. *Molecular light scattering and optical activity.* Cambridge University Press, (2009).
 86. Woody, R. W. *Methods Enzymol.* **246**, 34–71 (1995).
 87. Greenfield, N. J. *Nat. Protoc.* **1**, 2876–2890 (2006).
 88. Kelly, S. M. & Price, N. C. *Biochim. Biophys. Acta BBA-Protein Struct. Mol. Enzymol.* **1338**, 161–185 (1997).
 89. Greenfield, N. J. *TrAC Trends Anal. Chem.* **18**, 236–244 (1999).
 90. Ranjbar, B. & Gill, P. *Chem. Biol. Drug Des.* **74**, 101–120 (2009).
 91. Berova, N., Di Bari, L. & Pescitelli, G. *Chem. Soc. Rev.* **36**, 914–931 (2007).
 92. Watson, J. D. & Crick, F. H. *Nature* **171**, 737–738 (1953).
 93. Crick, F. *Nature* **227**, 561–563 (1970).
 94. Beadle, G. W. & Tatum, E. L. *Proc. Natl. Acad. Sci.* **27**, 499–506 (1941).
 95. Horowitz, N. H. *et al. Genetics* **166**, 1–10 (2004).
 96. Miescher, F.
 97. Miescher, F.
 98. Levene, P. A. & Bass, L. W. *Nucleic acids.* Chemical Catalog Company New York, (1931).
 99. Dahm, R. *Hum. Genet.* **122**, 565–581 (2008).
 100. Chargaff, E. & Vischer, E. *J. Biol. Chem.* **177**, 405–416 (1949).
 101. Chargaff, E. *J. Cell. Comp. Physiol.* **38**, 41–59 (1951).
 102. Minchin, S. & Lodge, J. *Essays Biochem.* **63**, 433–456 (2019).
 103. Griffith, F. *Epidemiol. Infect.* **27**, 113–159 (1928).
 104. Avery, O. T., MacLeod, C. M. & McCarty, M. J. *Exp. Med.* **79**, 137–158 (1944).

-
105. WILKINS, M. H. F., STOKES, A. R. & WILSON, H. R. *Nature* **171**, 738–740 (1953).
 106. Franklin, R. E. & Gosling, R. G. *Nature* **171**, 740–741 (1953).
 107. Altona, C. t & Sundaralingam, M. J. *Am. Chem. Soc.* **94**, 8205–8212 (1972).
 108. Duchardt, E. *et al. J. Am. Chem. Soc.* **126**, 1962–1970 (2004).
 109. Nozinovic, S. *et al. J. Am. Chem. Soc.* **132**, 10318–10329 (2010).
 110. Egli, M. *The Excitement of Discovery: Selected Papers of Alexander Rich* vol. Volume 11 309–315 WORLD SCIENTIFIC, (2018).
 111. Miao, Z. & Westhof, E. *Annu. Rev. Biophys.* **46**, 483–503 (2017).
 112. Saenger, W. *Principles of nucleic acid structure*. Springer, (1984).
 113. Ghosh, A. & Bansal, M. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 620–626 (2003).
 114. Dickerson, R. E. & Ng, H.-L. *Proc. Natl. Acad. Sci.* **98**, 6986–6988 (2001).
 115. Asamitsu, S. *et al. Int. J. Mol. Sci.* **20**, 2884 (2019).
 116. Heinemann, U. & Roske, Y. *Symmetry* **12**, 737 (2020).
 117. Bang, I. **26**, 293–311 (1910).
 118. Gellert, M., Lipsett, M. N. & Davies, D. R. *Proc. Natl. Acad. Sci.* **48**, 2013–2018 (1962).
 119. Luo, Y. *et al. Biochimie* **214**, 5–23 (2023).
 120. Yang, D. *G-Quadruplex Nucleic Acids* eds. Yang, D. & Lin, C. vol. 2035 1–24 Springer New York, (2019).
 121. Webba da Silva, M. *Chem. Eur. J.* **13**, 9738–9745 (2007).
 122. Karsisiotis, A. I., O’Kane, C. & da Silva, M. W. *Methods* **64**, 28–35 (2013).
 123. Schultze, P., Macaya, R. F. & Feigon, J. J. *Mol. Biol.* **235**, 1532–1547 (1994).
 124. da Silva, M. W. *Methods* **43**, 264–277 (2007).
 125. Largy, E., Mergny, J.-L. & Gabelica, V. *The Alkali Metal Ions: Their Role for Life* eds. Sigel, A., Sigel, H. & Sigel, R. K. O. vol. 16 203–258 Springer International Publishing, (2016).
 126. Cang, X., Šponer, J. & Cheatham III, T. E. *Nucleic Acids Res.* **39**, 4499–4512 (2011).
 127. Šponer, J. *et al. J. Am. Chem. Soc.* **135**, 9785–9796 (2013).
 128. Lim, K. W. *et al. J. Am. Chem. Soc.* **131**, 4301–4309 (2009).
 129. Lech, C. J., Heddi, B. & Phan, A. T. *Nucleic Acids Res.* **41**, 2034–2046 (2013).
 130. Lane, A. N., Chaires, J. B., Gray, R. D. & Trent, J. O. *Nucleic Acids Res.* **36**, 5482–5515 (2008).

-
131. Borzo, M., Detellier, C., Laszlo, P. & Paris, A. *J. Am. Chem. Soc.* **102**, 1124–1134 (1980).
 132. Rueda, M., Luque, F. J. & Orozco, M. *J. Am. Chem. Soc.* **128**, 3608–3619 (2006).
 133. Wong, A. & Wu, G. *J. Am. Chem. Soc.* **125**, 13895–13905 (2003).
 134. Hardin, C. C., Perry, A. G. & White, K. *Biopolym. Orig. Res. Biomol.* **56**, 147–194 (2000).
 135. Horvath, M. P. & Schultz, S. C. *J. Mol. Biol.* **310**, 367–377 (2001).
 136. Hud, N. V., Smith, F. W., Anet, F. A. & Feigon, J. *Biochemistry* **35**, 15383–15390 (1996).
 137. Sen, D. & Gilbert, W. *Nature* **344**, 410–414 (1990).
 138. Hazel, P., Huppert, J., Balasubramanian, S. & Neidle, S. *J. Am. Chem. Soc.* **126**, 16405–16415 (2004).
 139. Bugaut, A. & Balasubramanian, S. *Biochemistry* **47**, 689–697 (2008).
 140. Agrawal, P., Lin, C., Mathad, R. I., Carver, M. & Yang, D. *J. Am. Chem. Soc.* **136**, 1750–1753 (2014).
 141. Chambers, V. S. *et al. Nat. Biotechnol.* **33**, 877–881 (2015).
 142. Kolesnikova, S. & Curtis, E. A. *Molecules* **24**, 3074 (2019).
 143. Fujioka, A. *et al. J. Biol. Chem.* **281**, 8917–8926 (2006).
 144. Huppert, J. L. & Balasubramanian, S. *Nucleic Acids Res.* **33**, 2908–2916 (2005).
 145. Huppert, J. L. & Balasubramanian, S. *Nucleic Acids Res.* **35**, 406–413 (2007).
 146. Stegle, O., Payet, L., Mergny, J.-L., MacKay, D. J. & Huppert, J. L. *Bioinformatics* **25**, i374–i1382 (2009).
 147. Crawford, G. E. *et al. Genome Res.* **16**, 123–131 (2006).
 148. Eddy, J. & Maizels, N. *Nucleic Acids Res.* **34**, 3887–3896 (2006).
 149. Maizels, N. & Gray, L. T. *PLoS Genet.* **9**, e1003468 (2013).
 150. Lipps, H. J. & Rhodes, D. *Trends Cell Biol.* **19**, 414–422 (2009).
 151. Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. *Nat. Chem.* **5**, 182–186 (2013).
 152. Schaffitzel, C. *et al. Proc. Natl. Acad. Sci.* **98**, 8572–8577 (2001).
 153. Di Antonio, M. *et al. Nat. Chem.* **12**, 832–837 (2020).
 154. Valton, A. *et al. EMBO J.* **33**, 732–746 (2014).
 155. Prorok, P. *et al. Nat. Commun.* **10**, 3274 (2019).
 156. Teng, F.-Y. *et al. Cell. Mol. Life Sci.* **78**, 6557–6583 (2021).
 157. Bochman, M. L., Paeschke, K. & Zakian, V. A. *Nat. Rev. Genet.* **13**, 770–780 (2012).

-
158. London, T. B. *et al. J. Biol. Chem.* **283**, 36132–36139 (2008).
 159. Paeschke, K. *et al. Nature* **497**, 458–462 (2013).
 160. Broxson, C., Beckett, J. & Tornaletti, S. *Biochemistry* **50**, 4162–4172 (2011).
 161. Smestad, J. A. & Maher, L. J. *BMC Med. Genet.* **16**, 1–14 (2015).
 162. Belotserkovskii, B. P. *et al. Proc. Natl. Acad. Sci.* **107**, 12816–12821 (2010).
 163. Belotserkovskii, B. P. *et al. Nucleic Acids Res.* **41**, 1817–1828 (2013).
 164. Belotserkovskii, B. P., Soo Shin, J. H. & Hanawalt, P. C. *Nucleic Acids Res.* **45**, 6589–6599 (2017).
 165. Lim, G. & Hohng, S. *Nucleic Acids Res.* **48**, 9195–9203 (2020).
 166. Lee, C.-Y. *et al. Nat. Commun.* **11**, 3392 (2020).
 167. De Magis, A. *et al. Proc. Natl. Acad. Sci.* **116**, 816–825 (2019).
 168. Zhang, L., Lu, Z. & Zhao, X. *Biochim. Biophys. Acta BBA-Rev. Cancer* **1876**, 188569 (2021).
 169. Cheng, Y. *et al. J. Biol. Chem.* **294**, 5890–5895 (2019).
 170. Heckman, C., Mochon, E., Arcinas, M. & Boxer, L. M. *J. Biol. Chem.* **272**, 19609–19614 (1997).
 171. Seto, M. *et al. EMBO J.* **7**, 123–131 (1988).
 172. Gomez-Manzano, C. *et al. Cancer Res.* **61**, 6693–6697 (2001).
 173. Onel, B. *et al. J. Am. Chem. Soc.* **138**, 2563–2570 (2016).
 174. Sun, H. *et al. Biochim. Biophys. Acta BBA-Gen. Subj.* **1840**, 3052–3057 (2014).
 175. Cogoi, S., Paramasivam, M., Spolaore, B. & Xodo, L. E. *Nucleic Acids Res.* **36**, 3765–3780 (2008).
 176. Cogoi, S. & Xodo, L. E. *Nucleic Acids Res.* **34**, 2536–2549 (2006).
 177. Madden, S. K., de Araujo, A. D., Gerhardt, M., Fairlie, D. P. & Mason, J. M. *Mol. Cancer* **20**, 1–18 (2021).
 178. Chaudhuri, R., Bhattacharya, S., Dash, J. & Bhattacharya, S. J. *Med. Chem.* **64**, 42–70 (2020).
 179. Siddiqui-Jain, A., Grand, C. L., Bearss, D. J. & Hurley, L. H. *Proc. Natl. Acad. Sci.* **99**, 11593–11598 (2002).
 180. Sengupta, P., Bhattacharya, A., Sa, G., Das, T. & Chatterjee, S. *Biochemistry* **58**, 1975–1991 (2019).
 181. Shan, C. *et al. J. Med. Chem.* **60**, 1292–1308 (2017).
 182. Wang, Y.-Q. *et al. J. Med. Chem.* **60**, 6924–6941 (2017).
 183. Chen, S. *et al. Biochim. Biophys. Acta BBA-Gen. Subj.* **1830**, 4769–4777 (2013).

-
184. Rigo, R., Palumbo, M. & Sissi, C. *Biochim. Biophys. Acta BBA-Gen. Subj.* **1861**, 1399–1413 (2017).
 185. Lu, W., Zhang, Y., Liu, D., Songyang, Z. & Wan, M. *Exp. Cell Res.* **319**, 133–141 (2013).
 186. De Lange, T. *Genes Dev.* **19**, 2100–2110 (2005).
 187. Rhodes, D. & Lipps, H. J. *Nucleic Acids Res.* **43**, 8627–8637 (2015).
 188. De Lange, T. *Nat. Rev. Mol. Cell Biol.* **5**, 323–329 (2004).
 189. Palm, W. & de Lange, T. *Annu. Rev. Genet.* **42**, 301–334 (2008).
 190. Bryan, T. M. *Molecules* **25**, 3686 (2020).
 191. Kosiol, N., Juranek, S., Brossart, P., Heine, A. & Paeschke, K. *Mol. Cancer* **20**, 1–18 (2021).
 192. Aggarwal, M., Sommers, J. A., Shoemaker, R. H. & Brosh Jr, R. M. *Proc. Natl. Acad. Sci.* **108**, 1525–1530 (2011).
 193. Kocak, E. *et al. Genetics* **213**, 835–847 (2019).
 194. Lin, W. *et al. EMBO J.* **32**, 1425–1439 (2013).
 195. Stroik, S. *et al. Nucleic Acids Res.* **48**, 4960–4975 (2020).
 196. Gao, H., Cervantes, R. B., Mandell, E. K., Otero, J. H. & Lundblad, V. *Nat. Struct. Mol. Biol.* **14**, 208–214 (2007).
 197. Jacobs, S. A., Podell, E. R. & Cech, T. R. *Nat. Struct. Mol. Biol.* **13**, 218–225 (2006).
 198. Harley, C. B., Vaziri, H., Counter, C. M. & Allsopp, R. C. *Exp. Gerontol.* **27**, 375–382 (1992).
 199. Chen, C.-H. & Chen, R.-J. *J. Formos. Med. Assoc.* **110**, 275–289 (2011).
 200. Heaphy, C. M. *et al. Am. J. Pathol.* **179**, 1608–1615 (2011).
 201. Zahler, A. M., Williamson, J. R., Cech, T. R. & Prescott, D. M. *Nature* **350**, 718–720 (1991).
 202. Oganessian, L., Moon, I. K., Bryan, T. M. & Jarstfer, M. B. *EMBO J.* **25**, 1148–1159 (2006).
 203. Moye, A. L. *et al. Nat. Commun.* **6**, 7643 (2015).
 204. Paudel, B. P. *et al. Elife* **9**, e56428 (2020).
 205. Chaires, J. B. *et al. Nucleic Acids Res.* **48**, 4976–4991 (2020).
 206. Xu, M. *et al. J. Mol. Biol.* **433**, 166846 (2021).
 207. Ray, S. *et al. Biophys. J.* **104**, 2235–2245 (2013).
 208. Fernandes, C. A. *et al. Biochim. Biophys. Acta BBA-Gen. Subj.* **1864**, 129607 (2020).
 209. Kim, J., Cheong, C. & Moore, P. B. *Nature* **351**, 331–332 (1991).

-
210. Guo, J. U. & Bartel, D. P. *Science* **353**, aaf5371 (2016).
211. Biffi, G., Di Antonio, M., Tannahill, D. & Balasubramanian, S. *Nat. Chem.* **6**, 75–80 (2014).
212. Azzalin, C. M., Reichenbach, P., Khoriauli, L., Giulotto, E. & Lingner, J. *science* **318**, 798–801 (2007).
213. Arora, R., Brun, C. M. & Azzalin, C. M. *Long Non-Coding RNAs* 65–94 (2011).
214. Luke, B. & Lingner, J. *EMBO J.* **28**, 2503–2510 (2009).
215. Bao, H.-L. & Xu, Y. *Nat. Protoc.* **13**, 652–665 (2018).
216. Martadinata, H., Heddi, B., Lim, K. W. & Phan, A. T. *Biochemistry* **50**, 6455–6461 (2011).
217. Xu, Y., Suzuki, Y., Ito, K. & Komiyama, M. *Proc. Natl. Acad. Sci.* **107**, 14579–14584 (2010).
218. Martadinata, H. & Phan, A. T. *J. Am. Chem. Soc.* **131**, 2570–2578 (2009).
219. Redon, S., Reichenbach, P. & Lingner, J. *Nucleic Acids Res.* **38**, 5797–5806 (2010).
220. Bao, H.-L. & Xu, Y. *Chem. Commun.* **56**, 6547–6550 (2020).
221. Xu, Y. *et al. Bioorg. Med. Chem.* **22**, 4419–4421 (2014).
222. Xu, Y., Kimura, T. & Komiyama, M. vol. 52 169–170 Oxford University Press, (2008).
223. Panossian, L. *et al. Neurobiol. Aging* **24**, 77–84 (2003).
224. Wang, H. *et al. Mov. Disord.* **23**, 302–305 (2008).
225. De Felice, B. *et al. J. Hum. Genet.* **59**, 555–561 (2014).
226. Deng, Z., Norseen, J., Wiedmer, A., Riethman, H. & Lieberman, P. M. *Mol. Cell* **35**, 403–413 (2009).
227. Kozak, M. *Mol. Cell. Biol.* (1989).
228. Schaeffer, C. *et al. EMBO J.* **20**, 4803–4813 (2001).
229. Morris, M. J., Negishi, Y., Pazsint, C., Schonhoft, J. D. & Basu, S. *J. Am. Chem. Soc.* **132**, 17831–17839 (2010).
230. Bonnal, S. *et al. J. Biol. Chem.* **278**, 39330–39336 (2003).
231. Zhao, X. *et al. Gene* **681**, 36–44 (2019).
232. Bolduc, F., Garant, J.-M., Allard, F. & Perreault, J.-P. *J. Biol. Chem.* **291**, 21751–21760 (2016).
233. Crenshaw, E. *et al. PloS One* **10**, e0143160 (2015).
234. Zarudnaya, M. I., Kolomiets, I. M., Potyahaylo, A. L. & Hovorun, D. M. *Nucleic Acids Res.* **31**, 1375–1386 (2003).

-
235. Beaudoin, J.-D. & Perreault, J.-P. *Nucleic Acids Res.* **41**, 5898–5911 (2013).
236. Bagga, P. S., Ford, L. P., Chen, F. & Wilusz, J. *Nucleic Acids Res.* **23**, 1625–1631 (1995).
237. Matsumura, K. *et al.* *Oncogene* **36**, 1191–1199 (2017).
238. Arachchilage, G. M., Dassanayake, A. C. & Basu, S. *Chem. Biol.* **22**, 262–272 (2015).
239. Harris, L. M. & Merrick, C. J. *PLoS Pathog.* **11**, e1004562 (2015).
240. Ruggiero, E. & Richter, S. N. *Nucleic Acids Res.* **46**, 3270–3283 (2018).
241. Métifiot, M., Amrane, S., Litvak, S. & Andreola, M.-L. *Nucleic Acids Res.* **42**, 12352–12366 (2014).
242. Sundquist, W. I. & Heaphy, S. *Proc. Natl. Acad. Sci.* **90**, 3393–3397 (1993).
243. Perrone, R. *et al.* *PloS One* **8**, e73121 (2013).
244. Piekna-Przybylska, D., Sullivan, M. A., Sharma, G. & Bambara, R. A. *Biochemistry* **53**, 2581–2593 (2014).
245. Wang, S.-R. *et al.* *Sci. Adv.* **2**, e1501535 (2016).
246. Fleming, A. M., Ding, Y., Alenko, A. & Burrows, C. J. *ACS Infect. Dis.* **2**, 674–681 (2016).
247. Krafčiková, P., Demkovičová, E. & Víglaský, V. *Biochim. Biophys. Acta BBA-Gen. Subj.* **1861**, 1321–1328 (2017).
248. Kusov, Y., Tan, J., Alvarez, E., Enjuanes, L. & Hilgenfeld, R. *Virology* **484**, 313–322 (2015).
249. Yadav, V., Hemansi, null, Kim, N., Tuteja, N. & Yadav, P. *Front. Plant Sci.* **8**, 1163 (2017).
250. Zhang, D.-H. *et al.* *Biochemistry* **49**, 4554–4563 (2010).
251. Arora, A. & Maiti, S. *J. Phys. Chem. B* **113**, 10515–10520 (2009).
252. Collie, G. W., Haider, S. M., Neidle, S. & Parkinson, G. N. *Nucleic Acids Res.* **38**, 5569–5580 (2010).
253. Di Antonio, M. *et al.* *Angew. Chem. Int. Ed.* **51**, 11073–11078 (2012).
254. Yang, D. & Okamoto, K. *Future Med. Chem.* **2**, 619–646 (2010).
255. Hurley, L. H. *et al.* *Pharmacol. Ther.* **85**, 141–158 (2000).
256. Hurley, L. H. *Nat. Rev. Cancer* **2**, 188–200 (2002).
257. Mergny, J.-L. & Hélène, C. *Nat. Med.* **4**, 1366–1367 (1998).
258. Neidle, S. & Parkinson, G. *Nat. Rev. Drug Discov.* **1**, 383–393 (2002).
259. Neidle, S. *Nat. Rev. Chem.* **1**, 0041 (2017).
260. Sun, D. *et al.* *J. Med. Chem.* **40**, 2113–2116 (1997).

-
261. Riou, J. *et al. Proc. Natl. Acad. Sci.* **99**, 2672–2677 (2002).
262. Gowan, S. M., Heald, R., Stevens, M. F. & Kelland, L. R. *Mol. Pharmacol.* **60**, 981–988 (2001).
263. Harrison, R. J. *et al. Bioorg. Med. Chem. Lett.* **14**, 5845–5849 (2004).
264. Incles, C. M. *et al. Mol. Cancer Ther.* **3**, 1201–1206 (2004).
265. Murchie, A. I. & Lilley, D. M. *Nucleic Acids Res.* **20**, 49–53 (1992).
266. Lew, A., Rutter, W. J. & Kennedy, G. C. *Proc. Natl. Acad. Sci.* **97**, 12508–12512 (2000).
267. Yafe, A., Etzioni, S., Weisman-Shomer, P. & Fry, M. *Nucleic Acids Res.* **33**, 2887–2900 (2005).
268. Sun, D. *Nucleic Acids Res.* **33**, 6070–6080 (2005).
269. De Armond, R., Wood, S., Sun, D., Hurley, L. H. & Ebbinghaus, S. W. *Biochemistry* **44**, 16341–16350 (2005).
270. Darnell, J. C. *et al. Cell* **107**, 489–499 (2001).
271. Fry, M. & Loeb, L. A. *Proc. Natl. Acad. Sci.* **91**, 4950–4954 (1994).
272. Khateb, S., Weisman-Shomer, P., Hershcó-Shani, I., Ludwig, A. L. & Fry, M. *Nucleic Acids Res.* **35**, 5775–5788 (2007).
273. Simonsson, T., Kubista, M. & Pecinka, P. *Nucleic Acids Res.* **26**, 1167–1172 (1998).
274. Ambrus, A., Chen, D., Dai, J., Jones, R. A. & Yang, D. *Biochemistry* **44**, 2048–2058 (2005).
275. Rangan, A., Fedoroff, O. Y. & Hurley, L. H. *J. Biol. Chem.* **276**, 4640–4646 (2001).
276. Dai, J., Chen, D., Jones, R. A., Hurley, L. H. & Yang, D. *Nucleic Acids Res.* **34**, 5133–5144 (2006).
277. Dexheimer, T. S., Sun, D. & Hurley, L. H. *J. Am. Chem. Soc.* **128**, 5404–5415 (2006).
278. Dai, J. *et al. J. Am. Chem. Soc.* **128**, 1096–1098 (2006).
279. Rankin, S. *et al. J. Am. Chem. Soc.* **127**, 10584–10589 (2005).
280. Fernando, H. *et al. Biochemistry* **45**, 7854–7860 (2006).
281. Guo, K. *et al. J. Am. Chem. Soc.* **129**, 10220–10228 (2007).
282. Monchaud, D. & Teulade-Fichou, M.-P. *Org. Biomol. Chem.* **6**, 627–636 (2008).
283. Dai, J., Carver, M., Hurley, L. H. & Yang, D. *J. Am. Chem. Soc.* **133**, 17673–17680 (2011).
284. Lin, C. *et al. Angew. Chem.* **130**, 11054–11059 (2018).
285. Chen, Y. & Yang, D. *Curr. Protoc. Nucleic Acid Chem.* **50**, (2012).

-
286. Neidle, S. *Curr. Opin. Struct. Biol.* **19**, 239–250 (2009).
287. Phan, A. T., Kuryavyi, V., Gaw, H. Y. & Patel, D. J. *Nat. Chem. Biol.* **1**, 167–173 (2005).
288. Chung, W. J., Heddi, B., Hamon, F., Teulade-Fichou, M. & Phan, A. T. *Angew. Chem. Int. Ed.* **53**, 999–1002 (2014).
289. Kotar, A. *et al. Angew. Chem.* **128**, 12696–12699 (2016).
290. Chung, W. J. *et al. J. Am. Chem. Soc.* **135**, 13495–13501 (2013).
291. Wirmer-Bartoschek, J. *et al. Angew. Chem.* **129**, 7208–7212 (2017).
292. Liu, W. *et al. Nat. Commun.* **9**, 3496 (2018).
293. Parkinson, G. N., Lee, M. P. & Neidle, S. *Nature* **417**, 876–880 (2002).
294. Clark, G. R., Pytel, P. D., Squire, C. J. & Neidle, S. *J. Am. Chem. Soc.* **125**, 4066–4067 (2003).
295. Haider, S. M., Parkinson, G. N. & Neidle, S. *J. Mol. Biol.* **326**, 117–125 (2003).
296. Parkinson, G. N., Ghosh, R. & Neidle, S. *Biochemistry* **46**, 2390–2397 (2007).
297. Campbell, N. H., Parkinson, G. N., Reszka, A. P. & Neidle, S. *J. Am. Chem. Soc.* **130**, 6722–6724 (2008).
298. Collie, G. W. *et al. J. Am. Chem. Soc.* **134**, 2723–2731 (2012).
299. Bazzicalupi, C., Ferraroni, M., Bilia, A. R., Scheggi, F. & Gratterer, P. *Nucleic Acids Res.* **41**, 632–638 (2013).
300. Fedoroff, O. Y. *et al. Biochemistry* **37**, 12367–12374 (1998).
301. Liu, Z.-R., Hecker, K. H. & Rill, R. L. *J. Biomol. Struct. Dyn.* **14**, 331–339 (1996).
302. Han, H., Bennett, R. J. & Hurley, L. H. *Biochemistry* **39**, 9311–9316 (2000).
303. Grand, C. L. *et al. Mol. Cancer Ther.* **1**, 565–573 (2002).
304. Han, F. X., Wheelhouse, R. T. & Hurley, L. H. *J. Am. Chem. Soc.* **121**, 3561–3570 (1999).
305. *Mol. Cancer Ther.* **2**, 208 (2003).
306. Ou, T.-M. *et al. J. Med. Chem.* **50**, 1465–1474 (2007).
307. Liu, J. *et al. Leukemia* **21**, 1300–1302 (2007).
308. Kang, H.-J. & Park, H.-J. *Biochemistry* **48**, 7392–7398 (2009).
309. Brown, R. V., Danford, F. L., Gokhale, V., Hurley, L. H. & Brooks, T. A. *J. Biol. Chem.* **286**, 41018–41027 (2011).
310. McLuckie, K. I. *et al. J. Am. Chem. Soc.* **133**, 2658–2663 (2011).
311. Wang, X.-D. *et al. J. Med. Chem.* **53**, 4390–4398 (2010).
312. Lavrado, J. *et al. ChemMedChem* **8**, 1648–1661 (2013).

-
313. Lavrado, J. *et al. Sci. Rep.* **5**, 9696 (2015).
314. Piazza, A. *et al. Nucleic Acids Res.* **38**, 4337–4348 (2010).
315. Salvati, E. *et al. Oncogene* **29**, 6280–6293 (2010).
316. Rodriguez, R. *et al. Nat. Chem. Biol.* **8**, 301–310 (2012).
317. McLuckie, K. I. *et al. J. Am. Chem. Soc.* **135**, 9640–9643 (2013).
318. Zimmer, J. *et al. Mol. Cell* **61**, 449–460 (2016).
319. Pirota, V., Stasi, M., Benassi, A. & Doria, F. *Annu. Rep. Med. Chem.* **54**, 163–196 (2020).
320. Balaratnam, S. & Schneekloth Jr, J. S. *Annual Reports in Medicinal Chemistry* vol. 54 361–407 Elsevier, (2020).
321. Shin-ya, K. *et al. J. Am. Chem. Soc.* **123**, 1262–1263 (2001).
322. Doi, T. *et al. Org. Biomol. Chem.* **9**, 387–393 (2011).
323. De Cian, A., Guittat, L., Shin-ya, K., Riou, J.-F. & Mergny, J.-L. vol. 49 235–236 Oxford University Press, (2005).
324. Read, M. *et al. Proc. Natl. Acad. Sci.* **98**, 4844–4849 (2001).
325. Gowan, S. M. *et al. Mol. Pharmacol.* **61**, 1154–1162 (2002).
326. Moore, M. J. *et al. J. Med. Chem.* **49**, 582–599 (2006).
327. Burger, A. M. *et al. Cancer Res.* **65**, 1489–1496 (2005).
328. Stasyuk, O. A., Szatyłowicz, H. & Krygowski, T. M. *Struct. Chem.* **27**, 111–118 (2016).
329. Martinez, C. & Iverson, B. (2012).
330. Hunter, C. A. & Sanders, J. K. *J. Am. Chem. Soc.* **112**, 5525–5534 (1990).
331. Chakrabarti, P. & Bhattacharyya, R. *Prog. Biophys. Mol. Biol.* **95**, 83–137 (2007).
332. Hunter, C. A. *Philos. Trans. R. Soc. Lond. Ser. Phys. Eng. Sci.* **345**, 77–85 (1993).
333. McGaughey, G. B., Gagné, M. & Rappé, A. K. *J. Biol. Chem.* **273**, 15458–15463 (1998).
334. Ohnmacht, S. A. & Neidle, S. *Bioorg. Med. Chem. Lett.* **24**, 2602–2612 (2014).
335. Jana, J. *et al. Sci. Rep.* **7**, 40706 (2017).
336. Iachettini, S. *et al. J. Exp. Clin. Cancer Res.* **32**, 1–12 (2013).
337. Boddupally, P. V. *et al. J. Med. Chem.* **55**, 6076–6086 (2012).
338. Yuan, G., Zhang, Q., Zhou, J. & Li, H. *Mass Spectrom. Rev.* **30**, 1121–1142 (2011).
339. Beck, J. L., Colgrave, M. L., Ralph, S. F. & Sheil, M. M. *Mass Spectrom. Rev.* **20**, 61–87 (2001).
340. Brodbelt, J. S. *Annu. Rev. Anal. Chem.* **3**, 67–87 (2010).

-
341. Baker, E. S., Bernstein, S. L., Gabelica, V., De Pauw, E. & Bowers, M. T. *Int. J. Mass Spectrom.* **253**, 225–237 (2006).
342. Ganem, B., Li, Y.-T. & Henion, J. D. *Tetrahedron Lett.* **34**, 1445–1448 (1993).
343. Gale, D. C., Goodlett, D. R., Light-Wahl, K. J. & Smith, R. D. *J. Am. Chem. Soc.* **116**, 6027–6028 (1994).
344. Bayer, E. *et al. Anal. Chem.* **66**, 3858–3863 (1994).
345. Light-Wahl, K. *et al. J. Am. Chem. Soc.* **115**, 803–804 (1993).
346. Ding, J. & Anderegg, R. J. *J. Am. Soc. Mass Spectrom.* **6**, 159–164 (1995).
347. Siegel, M. M. *Curr. Top. Med. Chem.* **2**, 13–33 (2002).
348. Chingin, K., Xu, N. & Chen, H. *J. Am. Soc. Mass Spectrom.* **25**, 928–934 (2014).
349. Rosu, F., Pirotte, S., De Pauw, E. & Gabelica, V. *Int. J. Mass Spectrom.* **253**, 156–171 (2006).
350. Garabedian, A. *et al. Phys. Chem. Chem. Phys.* **18**, 26691–26702 (2016).
351. Largy, E. *et al. Chem. Rev.* **122**, 7720–7839 (2021).
352. Lodish, H. *et al. Molecular Cell Biology. 4th edition* WH Freeman, (2000).
353. Monchaud, D. *et al. Biochimie* **90**, 1207–1223 (2008).
354. Scalabrin, M., Palumbo, M. & Richter, S. N. *Anal. Chem.* **89**, 8632–8637 (2017).
355. Ribaud, G., Ongaro, A., Oselladore, E., Memo, M. & Gianoncelli, A. *J. Med. Chem.* **64**, 13174–13190 (2021).
356. Pagano, B. *et al. Curr. Pharm. Des.* **18**, 1880–1899 (2012).
357. Oganessian, L., Graham, M. E., Robinson, P. J. & Bryan, T. M. *Biochemistry* **46**, 11279–11290 (2007).
358. Rosu, F., De Pauw, E. & Gabelica, V. *Biochimie* **90**, 1074–1087 (2008).
359. Li, H., Liu, Y., Lin, S. & Yuan, G. *Chem. Eur. J.* **15**, 2445–2452 (2009).
360. Fu, H., Yang, P., Hai, J. & Li, H. *Spectrochim. Acta. A. Mol. Biomol. Spectrosc.* **203**, 70–76 (2018).
361. Ma, Y., Iida, K. & Nagasawa, K. *Biochem. Biophys. Res. Commun.* **531**, 3–17 (2020).
362. Casagrande, V., Alvino, A., Bianco, A., Ortaggi, G. & Franceschin, M. *J. Mass Spectrom.* **44**, 530–540 (2009).
363. Collie, G. *et al. Chem. Commun.* 7482–7484 (2009).
364. Rosu, F., Gabelica, V., Houssier, C. & De Pauw, E. *Nucleic Acids Res.* **30**, e82–e82 (2002).
365. Rosu, F., Gabelica, V., De Pauw, E., Mailliet, P. & Mergny, J. *ChemBioChem* **9**, 849–852 (2008).

-
366. Sannes-Lowery, K. A., Griffey, R. H. & Hofstadler, S. A. *Anal. Biochem.* **280**, 264–271 (2000).
367. Griffey, R. H. *et al. J. Am. Chem. Soc.* **122**, 9933–9938 (2000).
368. Kuzmič, P. *Anal. Biochem.* **237**, 260–273 (1996).
369. Kuzmič, P. *Methods Enzymol.* **467**, 247–280 (2009).
370. Rosu, F., Gabelica, V., Poncelet, H. & De Pauw, E. *Nucleic Acids Res.* **38**, 5217–5225 (2010).
371. Xu, N. *et al. J. Mass Spectrom.* **47**, 694–700 (2012).
372. Thaplyal, P. & Bevilacqua, P. C. *Methods in Enzymology* ed. Burke-Aguero, D. H. vol. 549 189–219 Academic Press, (2014).
373. Brandes, R. & Ehrenberg, A. *Nucleic Acids Res.* **14**, 9491–9508 (1986).
374. Phan, A. T. & Mergny, J. *Nucleic Acids Res.* **30**, 4618–4625 (2002).
375. Anthis, N. J. & Clore, G. M. *Q. Rev. Biophys.* **48**, 35–116 (2015).
376. Buck, J., Fürtig, B., Noeske, J., Wöhnert, J. & Schwalbe, H. *Proc. Natl. Acad. Sci.* **104**, 15699–15704 (2007).
377. Bugaut, A., Murat, P. & Balasubramanian, S. *J. Am. Chem. Soc.* **134**, 19953–19956 (2012).
378. Harika, N. K. *et al. Nucleic Acids Res.* **44**, 4519–4527 (2016).
379. Overhauser, A. W. *Phys. Rev.* **92**, 411–415 (1953).
380. Amrane, S. *et al. J. Am. Chem. Soc.* **134**, 5807–5816 (2012).
381. Hahn, E. L. & Maxwell, D. E. *Phys. Rev.* **88**, 1070–1084 (1952).
382. Bodenhausen, G. & Ruben, D. J. *Chem. Phys. Lett.* **69**, 185–189 (1980).
383. Schnieders, R. *et al. Angew. Chem.* **131**, 9238–9242 (2019).
384. Brinson, R. G., Szakal, A. L. & Marino, J. P. *J. Virol.* **85**, 13398–13408 (2011).
385. Thevarpadam, J. *et al. Angew. Chem. Int. Ed.* **55**, 2738–2742 (2016).
386. Phan, A. T. *J. Biomol. NMR* **16**, 175–178 (2000).
387. Lipsitz, R. S. & Tjandra, N. *Annu Rev Biophys Biomol Struct* **33**, 387–413 (2004).
388. Reid, B. R., Banks, K., Flynn, P. & Nerdal, W. *Biochemistry* **28**, 10001–10007 (1989).
389. Rieping, W. *et al. Bioinformatics* **23**, 381–382 (2007).
390. Vorlíčková, M. *et al. Methods* **57**, 64–75 (2012).
391. Karsisiotis, A. I. *et al. Angew. Chem.* **45**, 10833–10836 (2011).
392. Kypr, J., Kejnovská, I., Renčíuk, D. & Vorlíčková, M. *Nucleic Acids Res.* **37**, 1713–1725 (2009).
393. Paramasivan, S., Rujan, I. & Bolton, P. H. *Methods* **43**, 324–331 (2007).

-
394. Carvalho, J., Queiroz, J. A. & Cruz, C. *J. Chem. Educ.* **94**, 1547–1551 (2017).
395. Karsisiotis, A. I. *et al. Angew. Chem. Int. Ed.* **50**, 10645–10648 (2011).
396. Lim, K. W. *et al. Nucleic Acids Res.* **37**, 6239–6248 (2009).
397. Silva, M. L. C., Costa, R. S., dos Santos Santana, A. & Koblitiz, M. G. B. *Semina Ciênc. Agrár.* **31**, 669–681 (2010).
398. Naczki, M. & Shahidi, F. *J. Chromatogr. A* **1054**, 95–111 (2004).
399. Kumar, S. & Pandey, A. K. *Sci. World J.* **2013**, (2013).
400. Nijveldt, R. J. *et al. Am. J. Clin. Nutr.* **74**, 418–425 (2001).
401. Van Acker, S. A. B. E. *et al. Free Radic. Biol. Med.* **20**, 331–342 (1996).
402. Jucá, M. M. *et al. Nat. Prod. Res.* **34**, 692–705 (2020).
403. Nouri, Z. *et al. Cancers* **12**, 2276 (2020).
404. Sun, H., Xiang, J., Tang, Y. & Xu, G. *Biochem. Biophys. Res. Commun.* **352**, 942–946 (2007).
405. Bai, L.-P. *et al. PLoS One* **8**, e53962 (2013).
406. Pavan, V., Mucignat-Caretta, C., Redaelli, M., Ribaudó, G. & Zagotto, G. *Arch. Pharm. (Weinheim)* **348**, 607–614 (2015).
407. Zanforlin, E., Zagotto, G. & Ribaudó, G. *Curr. Med. Chem.* **24**, 3749–3773 (2017).
408. Sun, H. *et al. Bioorg. Med. Chem. Lett.* **16**, 3586–3589 (2006).
409. Tawani, A. & Kumar, A. *Sci. Rep.* **5**, 17574 (2015).
410. Bhattacharjee, S., Sengupta, P. K. & Bhowmik, S. *RSC Adv.* **7**, 37230–37240 (2017).
411. Ribaudó, G. *et al. Nat. Prod. Res.* 1–5 (2020).
412. Hajimahmoodi, M. *et al. Nat. Prod. Res.* **25**, 1059–1066 (2011).
413. Ribaudó, G., Vendrame, T. & Bova, S. *Nat. Prod. Res.* **31**, 1988–1994 (2017).
414. Su, Z., Wang, P., Yuan, W., Grant, G. & Li, S. *Nat. Prod. Commun.* **12**, 1743–1745 (2017).
415. Huang, T.-T. *et al. PLoS One* **6**, e18308 (2011).
416. Ribaudó, G., Ongaro, A., Zagotto, G., Memo, M. & Gianoncelli, A. *Nat. Prod. Res.* **35**, 2583–2587 (2021).
417. Ribaudó, G., Ongaro, A., Zagotto, G., Memo, M. & Gianoncelli, A. *Nat. Prod. Res.* **36**, 405–410 (2021).
418. Tyagi, S. *et al. SN Appl. Sci.* **2**, 1–16 (2020).
419. Pagano, B. *et al. ChemMedChem* **10**, 640–649 (2015).
420. Xue, Y. *et al. J. Am. Chem. Soc.* **129**, 11185–11191 (2007).

-
421. Heddi, B. & Phan, A. T. *J. Am. Chem. Soc.* **133**, 9824–9833 (2011).
422. Dai, J., Carver, M. & Yang, D. *Biochimie* **90**, 1172–1183 (2008).
423. Dai, J., Carver, M., Punchihewa, C., Jones, R. A. & Yang, D. *Nucleic Acids Res.* **35**, 4927–4940 (2007).
424. Tsao, R., Yang, R. & Young, J. C. *J. Agric. Food Chem.* **51**, 6445–6451 (2003).
425. Ribaudó, G. *et al. Fitoterapia* **105**, 132–138 (2015).
426. Kupeli, E., Orhan, I., Toker, G. & Yesilada, E. *J. Ethnopharmacol.* **107**, 169–174 (2006).
427. Ribaudó, G. *et al. Bioorganic Chem.* **87**, 474–483 (2019).
428. Li, K., Yatsunyk, L. & Neidle, S. *Nucleic Acids Res.* **49**, 519–528 (2021).
429. *Schrödinger Release 2020-1: Glide*. Schrödinger, LLC, (2020).
430. Agudelo, D., Bourassa, P., Bérubé, G. & Tajmir-Riahi, H.-A. *Int. J. Biol. Macromol.* **66**, 144–150 (2014).
431. Ribaudó, G. *et al. Arch. Pharm. (Weinheim)* **352**, 1900030 (2019).
432. Perry, P. J. *et al. J. Med. Chem.* **41**, 3253–3260 (1998).
433. Zagotto, G. *et al. Bioorg. Med. Chem.* **16**, 354–361 (2008).
434. Zagotto, G. *et al. J. Med. Chem.* **51**, 5566–5574 (2008).
435. Zagotto, G. *et al. Bioconjug. Chem.* **22**, 2126–2135 (2011).
436. Ranjan, N., Davis, E., Xue, L. & P. Arya, D. *Chem. Commun.* **49**, 5796–5798 (2013).
437. Cogoi, S., Shchekotikhin, A. E., Membrino, A., Sinkevich, Y. B. & Xodo, L. E. *J. Med. Chem.* **56**, 2764–2778 (2013).
438. Percivalle, C. *et al. Org. Biomol. Chem.* **12**, 3744–3754 (2014).
439. Ongaro, A. *et al. ChemMedChem* **17**, e202100665 (2022).
440. Rocard, L., Berezin, A., De Leo, F. & Bonifazi, D. *Angew. Chem. Int. Ed.* **54**, 15739–15743 (2015).
441. Haridas, V. *et al. Tetrabedron* **67**, 1873–1884 (2011).
442. Vilaca, H., Ferreira, P. M. & Micaelo, N. M. *Tetrabedron* **70**, 5420–5427 (2014).
443. Ling, T. *et al. Eur. J. Med. Chem.* **102**, 9–13 (2015).
444. Späth, A. & König, B. *Tetrabedron* **66**, 6019–6025 (2010).
445. Ongaro, A. *et al. New J. Chem.* **44**, 3624–3631 (2020).
446. Hein, J. E. & Fokin, V. V. *Chem. Soc. Rev.* **39**, 1302–1315 (2010).
447. Worrell, B. T., Malik, J. A. & Fokin, V. V. *Science* **340**, 457–460 (2013).
448. Darby, R. A. *et al. Nucleic Acids Res.* **30**, e39–e39 (2002).
449. De Cian, A. *et al. Biochimie* **90**, 131–155 (2008).

-
450. Wang, Y. & Patel, D. J. *Structure* **1**, 263–282 (1993).
451. Dai, J. *et al. Nucleic Acids Res.* **35**, 2440–2450 (2007).
452. Alcaro, S. *et al. J. Med. Chem.* **56**, 843–855 (2013).
453. Bessi, I., Jonker, H. R., Richter, C. & Schwalbe, H. *Angew. Chem.* **127**, 8564–8568 (2015).
454. Asamitsu, S., Bando, T. & Sugiyama, H. *Chem. Eur. J.* **25**, 417–430 (2019).
455. Mazerski, J., Martelli, S. & Borowski, E. *Acta Biochim. Pol.* **45**, 1–11 (1998).
456. Pommier, Y., Leo, E., Zhang, H. & Marchand, C. *Chem. Biol.* **17**, 421–433 (2010).
457. Wilson, W. D., Rاتمeyer, L., Zhao, M., Strekowski, L. & Boykin, D. *Biochemistry* **32**, 4098–4104 (1993).
458. Kumar, C. & Asuncion, E. H. *J. Am. Chem. Soc.* **115**, 8547–8553 (1993).
459. Van Arman, S. A. & Czarnik, A. W. *J. Am. Chem. Soc.* **112**, 5376–5377 (1990).
460. Folini, M. *et al. Biochem. Pharmacol.* **79**, 1781–1790 (2010).
461. Ribaldo, G., Scalabrin, M., Pavan, V., Fabris, D. & Zagotto, G. *Arkivoc* **2016**, 145–160 (2016).
462. Gama, S. *et al. J. Inorg. Biochem.* **160**, 275–286 (2016).
463. Coban, T., Robertson, C., Schwikkard, S., Singer, R. & LeGresley, A. *RSC Med. Chem.* **12**, 751–757 (2021).
464. Ongaro, A., Ribaldo, G., Zagotto, G., Memo, M. & Gianoncelli, A. *ChemistrySelect* **4**, 13138–13142 (2019).
465. Ribaldo, G. *et al. Molbank* **2020**, M1138 (2020).
466. Sonogashira, K. *J. Organomet. Chem.* **653**, 46–49 (2002).
467. Haider, S., Parkinson, G. N. & Neidle, S. *Biophys. J.* **95**, 296–311 (2008).
468. Jain, A. K., Paul, A., Maji, B., Muniyappa, K. & Bhattacharya, S. *J. Med. Chem.* **55**, 2981–2993 (2012).
469. Iida, K., Tera, M., Hirokawa, T., Shin-ya, K. & Nagasawa, K. *Chem. Commun.* 6481 (2009).
470. Maji, B., Kumar, K., Muniyappa, K. & Bhattacharya, S. *Org. Biomol. Chem.* **13**, 8335–8348 (2015).
471. Paniwnyk, L., Beaufoy, E., Lorimer, J. & Mason, T. J. *Ultrason. Sonochem.* **8**, 299–301 (2001).
472. Angell, Y. & Burgess, K. *J. Org. Chem.* **70**, 9595–9598 (2005).
473. Chau, N. T. T. *et al. Appl. Catal. Gen.* **467**, 497–503 (2013).
474. Soethoudt, M. *et al. J. Am. Chem. Soc.* **140**, 6067–6075 (2018).

-
475. Hwang, T.-L. & Shaka, A. J. *Magn. Reson. A* **112**, 275–279 (1995).
476. Sklenář, V. & Bax, A. J. *Magn. Reson. 1969* **74**, 469–479 (1987).
477. Trott, O. & Olson, A. J. *J. Comput. Chem.* **31**, 455–461 (2010).
478. Pettersen, E. F. *et al. J. Comput. Chem.* **25**, 1605–1612 (2004).
479. Roos, K. *et al. J. Chem. Theory Comput.* **15**, 1863–1874 (2019).
480. Rebič, M., Laaksonen, A., Šponer, J., Uličný, J. & Mocci, F. J. *Phys. Chem. B* **120**, 7380–7391 (2016).
481. Zgarbová, M. *et al. J. Chem. Theory Comput.* **11**, 5723–5736 (2015).
482. Machireddy, B., Sullivan, H.-J. & Wu, C. *Molecules* **24**, 1010 (2019).
483. Havrila, M., Stadlbauer, P., Islam, B., Otyepka, M. & Sponer, J. *J. Chem. Theory Comput.* **13**, 3911–3926 (2017).
484. Tikhomirov, A. S. *et al. Eur. J. Med. Chem.* **159**, 59–73 (2018).
485. Joung, I. S. & Cheatham III, T. E. *J. Phys. Chem. B* **112**, 9020–9041 (2008).
486. *Schrödinger Release 2020-1: Desmond Molecular Dynamics System, 2020, D. E. Shaw Research; Maestro-Desmond Interoperability Tools, Schrödinger: New York, NY, 2020.*
487. Sontheimer, H. *Diseases of the nervous system*. Academic Press, an imprint of Elsevier, (2021).
488. Wong, D. T., Horng, J. S., Bymaster, F. P., Hauser, K. L. & Molloy, B. B. *Life Sci.* **15**, 471–479 (1974).
489. Wong, D. T., Perry, K. W. & Bymaster, F. P. *Nat. Rev. Drug Discov.* **4**, 764–774 (2005).
490. Barnham, K. J., Masters, C. L. & Bush, A. I. *Nat. Rev. Drug Discov.* **3**, 205–214 (2004).
491. Jones, D. P. *Antioxid. Redox Signal.* **8**, 1865–1879 (2006).
492. Maria Michel, T., Pulschen, D. & Thome, J. *Curr. Pharm. Des.* **18**, 5890–5899 (2012).
493. Caiaffo, V., Oliveira, B. D., de Sá, F. B. & Evêncio Neto, J. *Pharmacol. Res. Perspect.* **4**, e00231 (2016).
494. Safhi, M. M. *et al. Arch. Physiol. Biochem.* **125**, 150–155 (2019).
495. Muraro, C. *et al. Appl. Sci.* **9**, 3631 (2019).
496. Santi, C., Tidei, C., Scalera, C., Piroddi, M. & Galli, F. *Curr. Chem. Biol.* **7**, 25–36 (2013).
497. Orian, L. & Toppo, S. *Free Radic. Biol. Med.* **66**, 65–74 (2014).
498. P Wolters, L. & Orian, L. *Curr. Org. Chem.* **20**, 189–197 (2016).

-
499. Tiezza, M. D., Ribaudó, G. & Orian, L. *Curr. Org. Chem.* **23**, 1381–1402 (2019).
500. Zimmerman, M. T., Bayse, C. A., Ramoutar, R. R. & Brumaghim, J. L. *J. Inorg. Biochem.* **145**, 30–40 (2015).
501. Wendel, A., Fausel, M., Safayhi, H., Tiegs, G. & Otter, R. *Biochem. Pharmacol.* **33**, 3241–3245 (1984).
502. Mugesh, G., du Mont, W.-W. & Sies, H. *Chem. Rev.* **101**, 2125–2180 (2001).
503. Bhabak, K. P. & Mugesh, G. *Acc. Chem. Res.* **43**, 1408–1419 (2010).
504. Galant, L. S. *et al. Free Radic. Res.* **51**, 657–668 (2017).
505. Posser, T. *et al. Chem. Biol. Interact.* **164**, 126–135 (2006).
506. Ribaudó, G. *et al. RSC Adv.* **10**, 18583–18593 (2020).
507. Ribaudó, G. *et al. ACS Omega* **7**, 8314–8322 (2022).
508. Sun, J. *et al. Adv. Synth. Catal.* **351**, 920–930 (2009).
509. Sharpless, K., Young, M. & Lauer, R. *Tetrahedron Lett.* **14**, 1979–1982 (1973).
510. De Filippo, D. & Momicchioli, F. *Tetrahedron* **25**, 5733–5744 (1969).
511. Wang, L. *et al. Polym. Chem.* **10**, 2039–2046 (2019).
512. Nuttall, K. L. *Ann. Clin. Lab. Sci.* **36**, 409–420 (2006).
513. Singh, F. V. & Wirth, T. *Catal. Sci. Technol.* **9**, 1073–1091 (2019).
514. Freudendahl, D. M., Santoro, S., Shahzad, S. A., Santi, C. & Wirth, T. *Angew. Chem. Int. Ed.* **48**, 8409–8411 (2009).
515. Wood, K. M. *et al. J. Neurochem.* **130**, 351–359 (2014).
516. Hashemi, P., Dankoski, E. C., Petrovic, J., Keithley, R. B. & Wightman, R. *Anal. Chem.* **81**, 9462–9471 (2009).
517. West, A. *et al. Neurochem. Int.* **123**, 50–58 (2019).
518. Saylor, R. A. *et al. Front. Neurosci.* **13**, 362 (2019).
519. Abdalla, A. *et al. J. Neurochem.* **153**, 33–50 (2020).
520. Jackson, B. P., Dietz, S. M. & Wightman, R. M. *Anal. Chem.* **67**, 1115–1120 (1995).
521. Kreilgaard, M., Smith, D., Brennum, L. & Sanchez, C. *Br. J. Pharmacol.* **155**, 276–284 (2008).
522. Ribaudó, G. *et al. Molecules* **26**, 2770 (2021).
523. Jones, D. N., Mundy, D. & Whitehouse, R. *J. Chem. Soc. Chem. Commun.* 86–87 (1970).
524. Sharpless, K., Lauer, R. & Teranishi, A. *J. Am. Chem. Soc.* **95**, 6137–6139 (1973).
525. Heffner, R. J., Jiang, J. & Joullie, M. M. *J. Am. Chem. Soc.* **114**, 10181–10189 (1992).
526. Nicolaou, K. *et al. J. Am. Chem. Soc.* **115**, 3558–3575 (1993).

-
527. Vargas, D., Fronczek, F. R., Fischer, N. H. & Hostettmann, K. *J. Nat. Prod.* **49**, 133–138 (1986).
528. Callant, P., Ongena, R. & Vandewalle, M. *Tetrahedron* **37**, 2085–2089 (1981).
529. Waring, A. J. & Zaidi, J. H. *J. Chem. Soc. Perkin 1* 631–639 (1985).
530. Simplicio, A. L., Clancy, J. M. & Gilmer, J. F. *Molecules* **13**, 519–547 (2008).
531. Sun, C. *et al. Biomaterials* **225**, 119514 (2019).
532. Gabriel, S. *Berichte Dtsch. Chem. Ges.* **20**, 2224–2236 (1887).
533. Corey, E. & Link, J. O. *J. Am. Chem. Soc.* **114**, 1906–1908 (1992).
534. Leuckart, R. *Berichte Dtsch. Chem. Ges.* **18**, 2341–2344 (1885).
535. Hofmann, A. W. V. *Philos. Trans. R. Soc. Lond.* 93–131 (1850).
536. Bowman, W. R. & Coghlan, D. R. *Tetrahedron* **53**, 15787–15798 (1997).
537. Fukuyama, T., Jow, C.-K. & Cheung, M. *Tetrahedron Lett.* **36**, 6373–6374 (1995).
538. Tingoli, M., Tiecco, M., Chianelli, D., Balducci, R. & Temperini, A. *J. Org. Chem.* **56**, 6809–6813 (1991).
539. Yildiz, I. *J. Phys. Chem. A* **120**, 3683–3692 (2016).
540. Bayse, C. A. *J. Phys. Chem. A* **111**, 9070–9075 (2007).
541. Bayse, C. A. *Org. Biomol. Chem.* **9**, 4748–4751 (2011).
542. Antony, S. & Bayse, C. A. *Inorg. Chem.* **50**, 12075–12084 (2011).
543. Chen, H. *et al. Appl. Catal. Gen.* **588**, 117283 (2019).
544. Prickaerts, J., Heckman, P. R. & Blokland, A. *Expert Opin. Investig. Drugs* **26**, 1033–1048 (2017).
545. Ribaud, G., Ongaro, A., Zagotto, G., Memo, M. & Gianoncelli, A. *ACS Chem. Neurosci.* **11**, 1726–1739 (2020).
546. Blokland, A., Menniti, F. S. & Prickaerts, J. *Expert Opin. Ther. Pat.* **22**, 349–354 (2012).
547. Gonzalez, G. A. & Montminy, M. R. *Cell* **59**, 675–680 (1989).
548. Lu, Y.-F., Kandel, E. R. & Hawkins, R. D. *J. Neurosci.* **19**, 10250–10261 (1999).
549. Schoffelmeer, A. N., Wardeh, G. & Mulder, A. H. *Naunyn. Schmiedebergs Arch. Pharmacol.* **330**, 74–76 (1985).
550. Nishi, A. & Snyder, G. L. *J. Pharmacol. Sci.* **114**, 6–16 (2010).
551. Ribaud, G., Angelo Pagano, M., Bova, S. & Zagotto, G. *Curr. Med. Chem.* **23**, 1239–1249 (2016).
552. Cameron, R. T., Whiteley, E., Day, J. P., Parachikova, A. I. & Baillie, G. S. *FEBS Open Bio* **7**, 64–73 (2017).

-
553. Yu, Y.-F. *et al.* *ACS Chem. Neurosci.* **8**, 2522–2534 (2017).
554. Rabal, O. *et al.* *ACS Chem. Neurosci.* **10**, 4076–4101 (2019).
555. Ribaudó, G., Memo, M. & Gianoncelli, A. *Pharmaceuticals* **14**, 58 (2021).
556. van der Staay, F. J. *et al.* *Neuropharmacology* **55**, 908–918 (2008).
557. Hutson, P. *et al.* *Neuropharmacology* **61**, 665–676 (2011).
558. Dell’Agli, M. *et al.* *J. Nat. Prod.* **71**, 1513–1517 (2008).
559. Ribaudó, G., Ongaro, A. & Zagotto, G. *Molbank* **2018**, M1004 (2018).
560. Burley, S. & Petsko, G. A. *Science* **229**, 23–28 (1985).
561. Liu, S. *et al.* *Proc. Natl. Acad. Sci.* **105**, 13309–13314 (2008).
562. Zheng, L. & Zhou, Z.-Z. *Eur. J. Med. Chem.* **259**, 115682 (2023).
563. Oselladore, E. *et al.* *New J. Chem.* **44**, 19472–19488 (2020).
564. Landucci, E. *et al.* *ACS Chem. Neurosci.* **14**, 3826–3838 (2023).
565. Landucci, E. *et al.* *Neurobiol. Dis.* **174**, 105895 (2022).
566. Laurino, A. *et al.* *Thyroid* **28**, 1387–1397 (2018).
567. Landucci, E., Mazzantini, C., Buonvicino, D., Pellegrini-Giampietro, D. E. & Bergonzi, M. C. *Molecules* **26**, 1592 (2021).
568. Kuzmanic, A. & Zagrovic, B. *Biophys. J.* **98**, 861–871 (2010).
569. a Lipinski, C., Lombardo, F., Dominy, B. & Feeney, P. *Adv Drug Deliv Rev* **46**, 3–26 (2001).
570. Campora, M., Francesconi, V., Schenone, S., Tasso, B. & Tonelli, M. *Pharmaceuticals* **14**, 33 (2021).
571. Huai, Q. *et al.* *Proc. Natl. Acad. Sci.* **101**, 9624–9629 (2004).
572. Xu, R. X. *et al.* *Science* **288**, 1822–1825 (2000).
573. Berman, H. M. *et al.* *Nucleic Acids Res.* **28**, 235–242 (2000).
574. Van Staveren, W. C., Steinbusch, H. W., Markerink-van Ittersum, M., Behrends, S. & De Vente, J. *Eur. J. Neurosci.* **19**, 2155–2168 (2004).
575. Ribaudó, G. *et al.* *Eur. J. Neurosci.* **57**, 1954–1965 (2023).
576. *ACD/Structure Elucidator, version 2019.2.1, Advanced Chemistry Development, Inc., Toronto, ON, Canada, www.acdlabs.com, 2019.*
577. Abreu, R., Froufe, H. J., Daniel, P. O., Queiroz, M. J. R. & Ferreira, I. C. *SAR QSAR Environ. Res.* **22**, 603–610 (2011).
578. Woodrow, M. D. *et al.* *Bioorg. Med. Chem. Lett.* **19**, 5261–5265 (2009).
579. Wang, H. *et al.* *J. Biol. Chem.* **281**, 21469–21479 (2006).
580. Verhoest, P. R. *et al.* *J. Med. Chem.* **52**, 7946–7949 (2009).

-
581. *Schrödinger Release 2020-1: Protein Preparation Wizard; Epik, Schrödinger, LLC, New York, NY, 2020; Impact, Schrödinger, LLC, New York, NY, 2020; Prime, Schrödinger, LLC, New York, NY, 2020.*
582. Gewald, R., Grunwald, C. & Egerland, U. *Bioorg. Med. Chem. Lett.* **23**, 4308–4314 (2013).
583. Hamblin, J. N. *et al. Bioorg. Med. Chem. Lett.* **18**, 4237–4241 (2008).
584. Goto, T. *et al. Bioorg. Med. Chem. Lett.* **23**, 3325–3328 (2013).
585. Carzaniga, L. *et al. J. Med. Chem.* **60**, 10026–10046 (2017).
586. Card, G. L. *et al. Structure* **12**, 2233–2247 (2004).
587. Hughes, R. O. *et al. Bioorg. Med. Chem. Lett.* **21**, 6348–6352 (2011).
588. Wu, D. *et al. J. Med. Chem.* **61**, 8468–8473 (2018).
589. Huang, M. *et al. Mol. Pharmacol.* **88**, 836–845 (2015).
590. Wu, Y. *et al. J. Med. Chem.* **62**, 4218–4224 (2019).
591. Meng, F. *et al. J. Med. Chem.* **55**, 8549–8558 (2012).
592. Hou, J. *et al. PLOS ONE* **6**, e18092 (2011).
593. *Schrödinger Release 2020-1: LigPrep, Schrödinger, LLC, New York, NY, 2020.*
594. *Dassault Systèmes BIOVIA, Discovery Studio Modeling Environment, Release 2017, San Diego: Dassault Systèmes, 2016.*
595. Mysinger, M. M., Carchia, M., Irwin, J. J. & Shoichet, B. K. *J. Med. Chem.* **55**, 6582–6594 (2012).
596. Huang, N., Shoichet, B. K. & Irwin, J. J. *J. Med. Chem.* **49**, 6789–6801 (2006).
597. Chen, H., Lyne, P. D., Giordanetto, F., Lovell, T. & Li, J. J. *Chem. Inf. Model.* **46**, 401–415 (2006).
598. Triballeau, N., Acher, F., Brabet, I., Pin, J.-P. & Bertrand, H.-O. *J. Med. Chem.* **48**, 2534–2547 (2005).
599. Ash, J. & Fourches, D. *J. Chem. Inf. Model.* **57**, 1286–1299 (2017).
600. Kuenemann, M. A. & Fourches, D. *Mol. Inform.* **37**, 1700138 (2018).
601. Lu, C. *et al. J. Chem. Theory Comput.* **17**, 4291–4300 (2021).
602. Madhavi Sastry, G., Adzhigirey, M., Day, T., Annabhimoju, R. & Sherman, W. J. *Comput. Aided Mol. Des.* **27**, 221–234 (2013).
603. Shelley, J. C. *et al. J. Comput. Aided Mol. Des.* **21**, 681–691 (2007).
604. Friesner, R. A. *et al. J. Med. Chem.* **49**, 6177–6196 (2006).
605. Bowers, K. J. *et al.* 84-es (2006).

7 Publications

Developing new ligands targeting G-quadruplex

A. Ongaro, G. Desiderati, **E. Oselladore**, D. Auricchio, M. Memo, G. Ribaud, C. Sissi and A. Gianoncelli. Amino-Acid-Anthraquinone Click Chemistry Conjugates Selectively Target Human Telomeric G-Quadruplexes. *ChemMedChem*, **2022**, vol 17, pp e202100665 doi:[10.1002/cmdc.202100665](https://doi.org/10.1002/cmdc.202100665).

G. Ribaud, A. Ongaro, **E. Oselladore**, M. Memo and A. Gianoncelli. Combining Electrospray Mass Spectrometry (ESI-MS) and Computational Techniques in the Assessment of G-Quadruplex Ligands: A Hybrid Approach to Optimize Hit Discovery. *J. Med. Chem.*, **2021**, vol 64, pp 13174–13190 doi:[10.1021/acs.jmedchem.1c00962](https://doi.org/10.1021/acs.jmedchem.1c00962).

G. Ribaud, **E. Oselladore**, A. Ongaro, G. Zagotto, M. Memo and A. Gianoncelli. Enhanced G-quadruplex selectivity of flavonoid glycoside rutin over quercetin. *Natural Product Research*, **2020**, pp 1–5 doi:[10.1080/14786419.2020.1859505](https://doi.org/10.1080/14786419.2020.1859505).

G. Ribaud, A. Ongaro, **E. Oselladore**, G. Zagotto, M. Memo and A. Gianoncelli. 9,10-Bis[(4-(2-hydroxyethyl)piperazine-1-yl)prop-2-yn-1-yl]anthracene: Synthesis and G-quadruplex Selectivity. *Molbank*, **2020**, vol 2020, pp M1138 doi:[10.3390/M1138](https://doi.org/10.3390/M1138).

Selenium-containing compounds for the protection against oxidative stress

G. Ribaud, M. Bortoli, C. E. Witt, B. Parke, S. Mena, **E. Oselladore**, G. Zagotto, P. Hashemi and L. Orian. ROS-Scavenging Selenofluoxetine Derivatives Inhibit *In Vivo* Serotonin Reuptake. *ACS Omega*, **2022**, vol 7, pp 8314–8322 doi:[10.1021/acsomega.1c05567](https://doi.org/10.1021/acsomega.1c05567).

G. Ribaud, M. Bortoli, **E. Oselladore**, A. Ongaro, A. Gianoncelli, G. Zagotto and L. Orian. Selenoxide Elimination Triggers Enamine Hydrolysis to Primary and Secondary Amines: A Combined Experimental and Theoretical Investigation. *Molecules*, **2021**, vol 26, pp 2770 doi:[10.3390/molecules26092770](https://doi.org/10.3390/molecules26092770).

G. Ribaudo, M. Bortoli, A. Ongaro, **E. Oselladore**, A. Gianoncelli, G. Zagotto and L. Orian. Fluoxetine scaffold to design tandem molecular antioxidants and green catalysts. *RSC Adv.*, **2020**, vol 10, pp 18583–18593 doi:[10.1039/D0RA03509B](https://doi.org/10.1039/D0RA03509B).

Phosphodiesterase inhibitors

E. Landucci, G. Ribaudo, M. Anyanwu, **E. Oselladore**, M. Giannangeli, C. Mazzantini, D. Lana, M. G. Giovannini, M. Memo, D. E. Pellegrini-Giampietro et al. Virtual Screening-Accelerated Discovery of a Phosphodiesterase 9 Inhibitor with Neuroprotective Effects in the Kainate Toxicity In Vitro Model. *ACS Chem. Neurosci.*, **2023**, vol 14, pp 3826–3838 doi:[10.1021/acchemneuro.3c00431](https://doi.org/10.1021/acchemneuro.3c00431).

E. Oselladore, A. Ongaro, G. Zagotto, M. Memo, G. Ribaudo and A. Gianoncelli. Combinatorial library generation, molecular docking and molecular dynamics simulations for enhancing the isoflavone scaffold in phosphodiesterase inhibition. *New J. Chem.*, **2020**, vol 44, pp 19472–19488 doi:[10.1039/D0NJ02537B](https://doi.org/10.1039/D0NJ02537B).

Minor side projects

G. Ribaudo, M. Carotti, A. Ongaro, **E. Oselladore**, M. Scano, G. Zagotto, D. Sandonà and A. Gianoncelli. Synthesis and Evaluation of Bithiazole Derivatives As Potential α -Sarcoglycan Correctors. *ACS Med. Chem. Lett.*, **2023**, vol 14, pp 1049–1053 doi:[10.1021/acsmchemlett.3c00046](https://doi.org/10.1021/acsmchemlett.3c00046).

G. Ribaudo, X. Yun, A. Ongaro, **E. Oselladore**, J. P. L. Ng, R. K. Haynes, B. Y. K. Law, M. Memo, V. K. W. Wong, P. Coghi et al. Combining computational and experimental evidence on the activity of antimalarial drugs on papain-like protease of SARS-CoV -2: A repurposing study. *Chem Biol Drug Des*, **2023**, vol 101, pp 809–818 doi:[10.1111/cbdd.14187](https://doi.org/10.1111/cbdd.14187).

A. Mastinu, M. Anyanwu, M. Carone, G. Abate, S. A. Bonini, G. Peron, E. Tirelli, M. Pucci, G. Ribaudo, **E. Oselladore** et al. The Bright Side of Psychedelics: Latest Advances and Challenges in Neuropharmacology. *IJMS*, **2023**, vol 24, pp 1329 doi:[10.3390/ijms24021329](https://doi.org/10.3390/ijms24021329).

G. Ribaldo, A. Ongaro, **E. Oselladore**, G. Zagotto, M. Memo and A. Gianoncelli. A computational approach to drug repurposing against SARS-CoV-2 RNA dependent RNA polymerase (RdRp). *Journal of Biomolecular Structure and Dynamics*, **2022**, vol 40, pp 1101–1108 doi:[10.1080/07391102.2020.1822209](https://doi.org/10.1080/07391102.2020.1822209).

A. Ongaro, **E. Oselladore**, M. Memo, G. Ribaldo and A. Gianoncelli. Insight into the LFA-1/SARS-CoV-2 Orf7a Complex by Protein–Protein Docking, Molecular Dynamics, and MM-GBSA Calculations. *J. Chem. Inf. Model.*, **2021**, vol 61, pp 2780–2787 doi:[10.1021/acs.jcim.1c00198](https://doi.org/10.1021/acs.jcim.1c00198).

