A new Peptide Nucleic Acid (PNA) structure with potential for high affinity duplex and triplex binding

Victoria Kamperi

A thesis submitted for the Degree of Master of Science by Research

University of East Anglia

School of Pharmacy

September 2022

Abstract

Peptide Nucleic Acids (PNA) are synthetic DNA analogues, first discovered by Prof Peter E. Nielsen and his team in 1991. The deoxyribose phosphate backbone of DNA is replaced by a 2-aminoethyl glycine unit to which the nucleobases are linked through a methyl carbonyl linker. The lack of charge in the PNA's backbone, allows for the formation of PNA-DNA or PNA-RNA chimeras with significantly higher affinity and selectivity than the relevant homo- or heteroduplexes. This tight binding of the DNA mimics to DNA or RNA, allows them to inhibit processes such as transcription and translation, leading to their potential as antisense or antigene agents.

Certain limitations with the initially suggested structure by Nielsen led to further investigations into synthesizing partially altered PNA structures that could overcome them, including PNAs forming triplex structures with double-stranded DNA. Topham et al. recently disclosed a theoretical modified PNA monomeric residue unit comprised of a 2-cis olefin carbonyl side-chain linker that is optimised to target thymine or uracil. Homopyrimidine PNAs have the ability to strand invade double-stranded homopurine DNA, leading to the formation of highly stable PNA·DNA-PNA triple helix structures via Watson-Crick and Hoogsteen base pairing.

In this project, we investigated the synthesis of the PNA analogue described by Topham et al. Most of the initial synthetic pathway was successfully completed. However, a number of different directions had to be explored, due to complications with the synthesis of certain intermediate compounds, which are discussed and described in this report.

Table of Contents

List of Figures

Abbreviations

A	Adenine
AMD	Age-related macular degeneration
ApoB100	Apolipoprotein B100
ASO	Antisense oligonucleotide
Boc	Tert-butyloxycarbonyl protecting group
C	Cytosine
CMV	Cytomegalovirus
DIBAL	Diisobutylaluminium hydride
DMAP	4-dimethylaminopyridine
DMF	Dimethylformamide
DMP	Dess-Martin periodinane
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dsDNA	Double-stranded nucleic acid
E	3-oxo-2,3-dihydropyridazine
EtOAc	Ethyl acetate
EtOH	Ethanol
ETS2	ETS proto-oncogene 2, transcription factor
G	Guanine
HER2	Human epidermal growth factor receptor 2
HIV	Human immunodeficiency virus
IBX	2-iodoxybenzoic acid
IER2	Immediate-early response 2 protein
J	Pseudoisocytosine
LDL-C	Low-density lipoprotein cholesterol
LNA	Locked nucleic acid
M	2-aminopyridine
MeOH	Methanol
mRNA	Messenger ribonucleic acid
NMR	Nuclear magnetic resonance
PCC	Pyridinium chromate
PDC	Pyridinium dichromate
PNA	Peptide nucleic acid
ppm	Parts per million

pu	Purine
py	Pyrimidine
RNA	Ribonucleic acid
RT	Room temperature
SMA	Spinal muscular atrophy
SMN	Survival of motor neuron
ssDNA	Single-stranded nucleic acid
ssRNA	Single stranded ribonucleic acid
T	Thymine
TFO	Triplex-forming oligonucleotide
THF	Tetrahydrofuran
TLC	Thin layer chromatography
T*m*	Melting temperature
tRNA	Transfer ribonucleic acid
U	Uracil
UV	Ultraviolet (spectroscopy)
VEGF	Vascular endothelial growth factor

Acknowledgements

I would like to thank my supervisors Prof Mark Searcey and Dr Andrew Beekman for giving me this opportunity to carry out this project and be part of their team. Their guidance, support and endless patience really made this experience very valuable to me. I would also like to give special thanks to Dr Zoe Goddard, her day-to-day help both inside and outside the lab truly made my experience better than I had imagined.

I would also like to say an enormous thank you to both the Searcey and Beekman research groups for welcoming me so whole-heartedly and making my Master's degree such a good experience.

Last but not least, I would like to thank my family, for the endless love and support they have given me throughout my life, and for always believing in me.

Chapter 1: Introduction

1.1 Discovery of DNA/RNA structure and orientation

Nucleic acids are macromolecules made up of monomers, called nucleotides, which carry genetic information. There are two main classes, deoxyribonucleic acids (DNA) and ribonucleic acids (RNA). Each nucleotide is made up of three parts: a sugar, a phosphate group and a nitrogen-containing ring structure, called a base. Nucleotides contain one of four bases, namely adenine (A), cytosine (C) or guanine (G) -which are found in both DNA and RNA-, with the fourth base option being thymine (T) in DNA and uracil (U) in RNA[1]. Cytosine, thymine and uracil contain a single ring made up of carbon and nitrogen atoms, and are known as pyrimidines (py), as seen in figure **1A**. Adenine and guanine, also known as purines (pu), are made up of two carbon-nitrogen rings (figure **1A**). The bases are attached to the sugar through a bond connecting the 1' carbon of the sugar with the nitrogen found at position 9 of the purines and position 1 of the pyrimidines. The phosphate group, binds to the 5' carbon of the sugar, as seen in figure **1A**. DNA chains are formed by the binding of nucleotides through phosphodiester bonds between the 5' carbon of the phosphate group of one nucleotide and the 3' carbon of the sugar of the next nucleotide, thus forming polynucleotides[1].

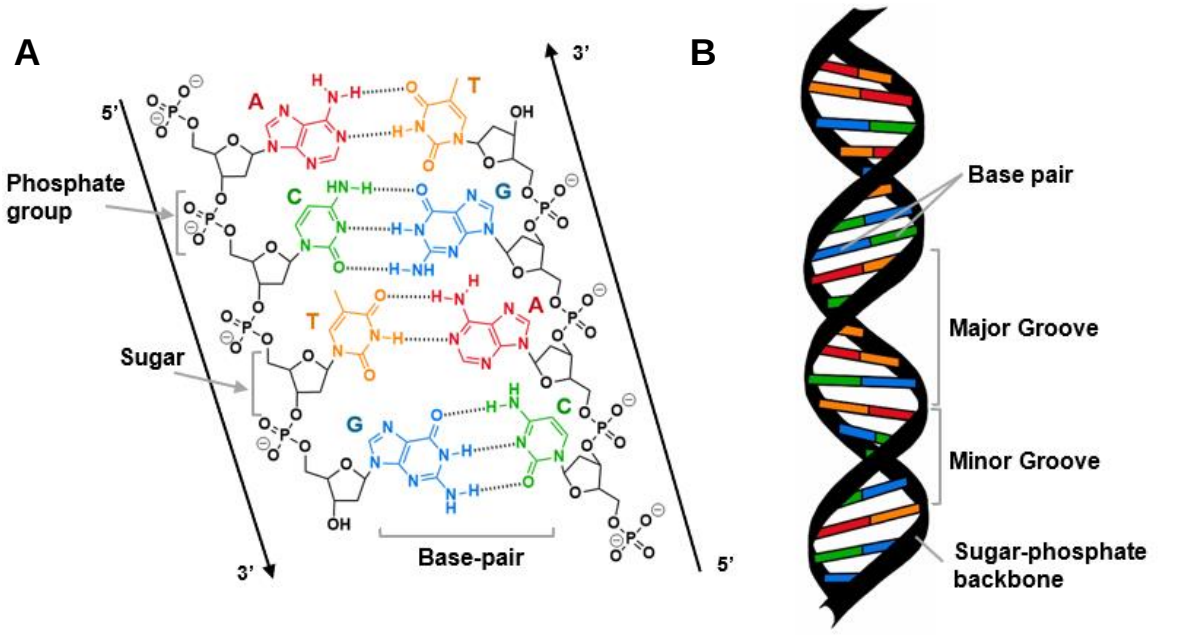


Figure 1: A. *Structure of a double-stranded DNA molecule, made up of nucleotides. A is always bound to T through two hydrogen bonds whereas C and G are held together through three hydrogen bonds.* ***B.*** *dsDNA helix discovered by Watson and Crick*[2,3]

The structure of DNA was discovered in 1953 by James Watson and Francis Crick, based upon experimental observations by Rosalind Franklin and Maurice Wilkins[2–4]. They

presented a double-stranded helical molecule made up of two chains with sugar-phosphate backbones, held together through hydrogen-bonded base pairs. The sugar-phosphate parts of the nucleotides form a spine sitting on the outside of the helix, with the bases found flat on the inside facing each other and the pairs being stacked on top of each other. When looking at the helix a major and minor groove can be detected (**figure 1B**)**,** which is important for protein interactions and DNA processes.

In the context of a helix, a purine base only interacts with a pyrimidine, due to the structures of the polynucleotides, where it has been found that A always binds to T through 2 hydrogen bonds, and C to G through 3 hydrogen bonds, making the latter a stronger interaction[5]. Therefore A-T and C-G have come to be known as complementary base pairs, while any other combination does not meet the requirements that allow both the helix to form and the correct alignment of the hydrogen bonds. Because of the selectivity of the bases, the two strands are described as complementary to each other and therefore one can be used to determine the sequence of the other. This feature is vital and is used to carry and express genetic information, as well as passing it along to new cells following cell division. Certain conditions can break the hydrogen bonds and separate the two strands, such as heat, certain chemicals and some enzymes, when undertaking processes such as DNA replication[5].

1.2 DNA processes (replication, transcription, translation)

1.2.1 DNA replication

DNA replication can be described as the process by which a double-stranded DNA molecule (dsDNA) is used as a template to produce two identical DNA molecules (figure **2**). This ensures that, whenever a cell divides, the two daughter cells contain the same genetic information as the original cell. This process is semi-conservative, since the newly-formed dsDNA molecule consists of two strands, one from the original molecule, which is used as a template, and a newly synthesized strand, complementary to the first[5,6]. The process starts with the unwinding of the double helix, which occurs at specific positions in the sequence known as origins, where helicase enzymes bind and begin unwinding the helix in both directions simultaneously; by breaking the hydrogen bonds between the bases. Once the two strands have been separated, enzymes called DNA polymerases come in and begin replicating the DNA by matching bases in a complementary way, always in the 5’ $\rightarrow$ 3’ direction. As DNA polymerase cannot initiate replication, an RNA primer is first synthesized and binds to the template strand, thus creating an attachment point for DNA polymerase, which begins the replication process by adding nucleotides to the 3’ end of the primer. The

two strands run in opposite directions. Since the DNA polymerase can only build the strands in the 5' → 3' direction, only one of the original DNA strands can be copied continuously. The enzyme builds towards the direction of the unwinding, by constantly adding nucleotides to the new sections, thus creating a new strand, known as the leading strand. In contrast, the other original strand is copied discontinuously as the DNA polymerase builds away from the helicase, therefore having to return to copy newly separated regions of the DNA. The fragments built, known as Okazaki fragments, are finally joined together by an enzyme named DNA ligase, which makes up the new strand, which in this case is known as the lagging strand[7]. Due to the importance of this process, DNA polymerases also check that the bases are correctly matched, through a mechanism known as proofreading.

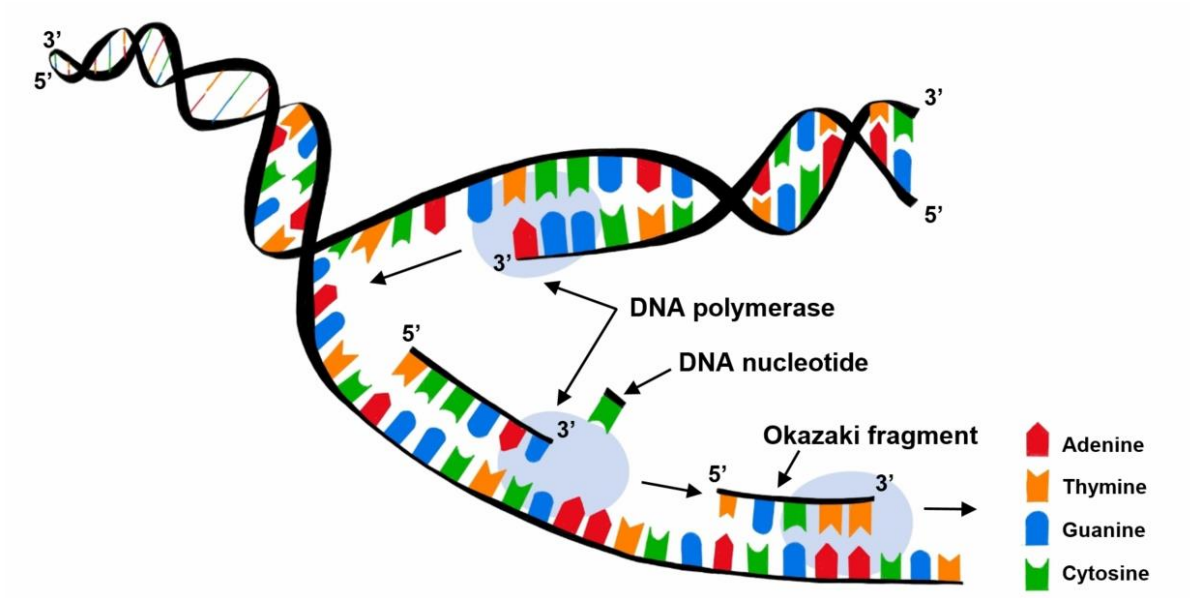


Figure 2: *The process of DNA replication*[5,6]

1.2.2 DNA transcription

With regard to gene expression, this can be broken down into two major steps, namely transcription and translation. During transcription[8–11], enzymes known as RNA polymerases use the specific sequence of bases in DNA as a template and copy it into RNA, synthesizing messenger RNA (mRNA)[12,13]. A promoter sequence is found in front of every gene, which RNA polymerases recognise and bind to, using it as the starting point for the unwinding of the double-helix, which is then followed by the synthesis of the RNA strand in a 5' → 3' direction. One of the DNA strands acts as the template strand that the RNA polymerase binds to, which means that the synthesized RNA strand carries the same information as the non-template strand, the only difference being that thymine (T) is replaced by uracil (U), (figure **3**). Once the mRNA strand has been generated it undergoes a process known as

splicing, where unwanted regions are removed, called introns. The parts that remain are characterised as exons, developing the premature mRNA to mature mRNA.

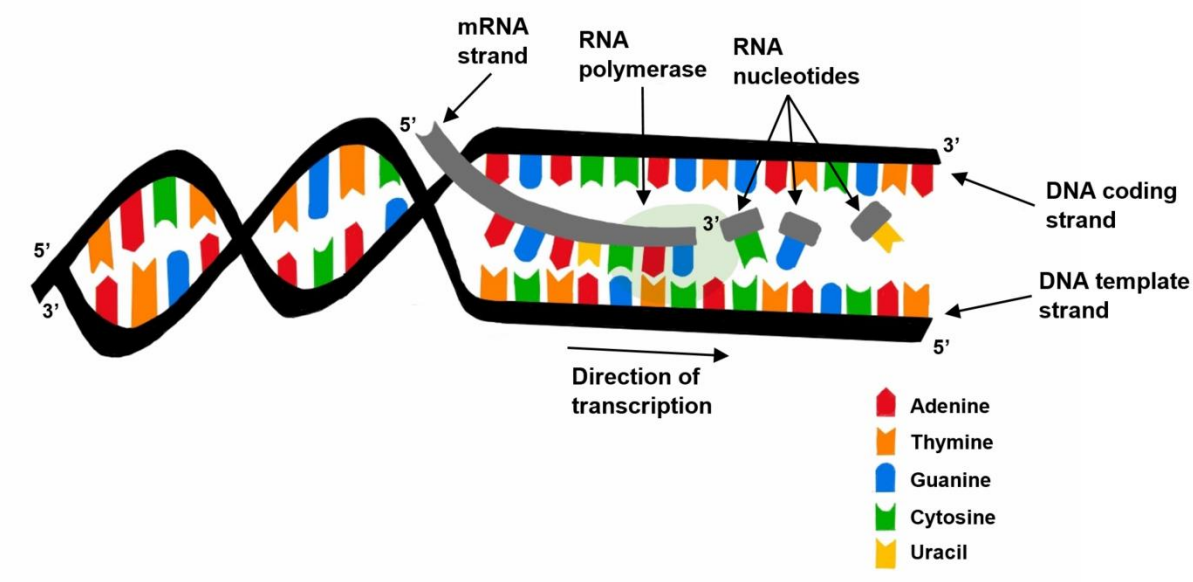


Figure 3: *The process of DNA transcription*[8–11]

1.2.3 DNA translation

Once the mRNA strand is in the cytoplasm, the translation process begins, whereby the genetic information carried is "decoded" and used to build a polypeptide[1,14]. The mRNA strand is read in groups of three nucleotides, called codons, each corresponding to an amino acid found in the genetic code. This process takes place on ribosomes[15], and always begins at AUG, the initiation codon. Transfer RNA molecules (tRNA) act as temporary carriers of amino acids, using the mRNA sequence to bring the correct amino acids to the ribosome[16]. Each tRNA contains an anti-codon, which binds to the complementary codon in the mRNA sequence, whilst carrying its respective amino acid to the ribosome. An enzyme called peptidyl transferase forms peptide bonds between the amino acids, thus creating polypeptide sequences (figure **4**). Following the release of the polypeptide chain, other processes occur to fold the chain and, often, to incorporate non-RNA template based post-translational modifications to generate the final working protein.

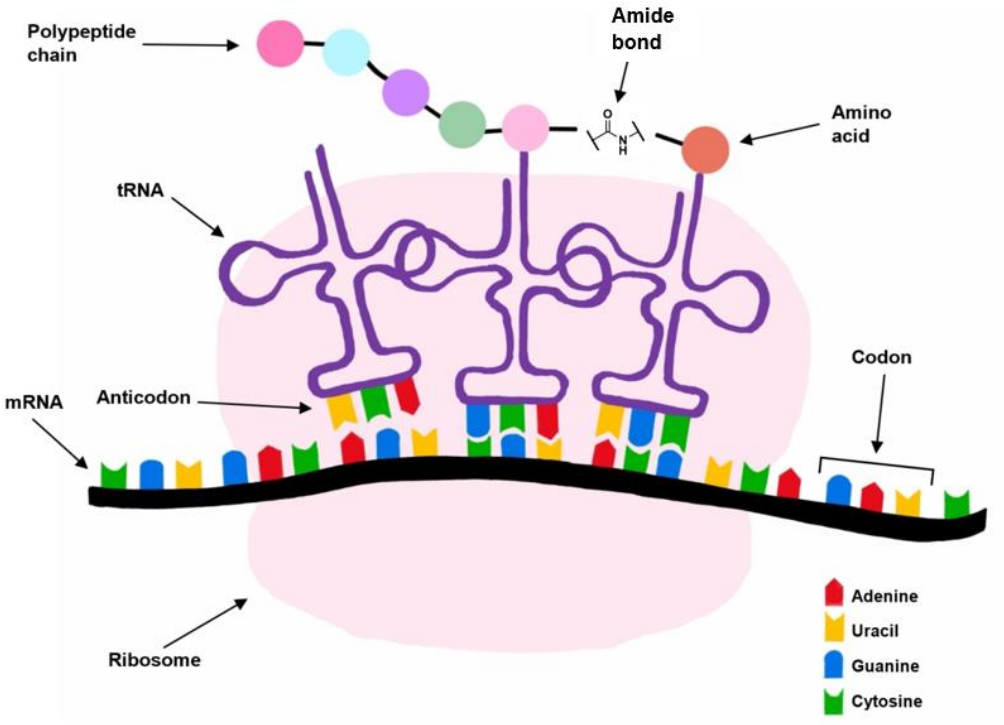


Figure 4: *The process of DNA translation*[14]

1.3 Alternative forms of dsDNA (A-form, B-form, Z-form)

1.3.1 B-form of DNA

In addition to the right-handed double helix form of DNA described by Watson, Crick and Franklin, also known as B-form[2,3], DNA has been reported to exist in a number of different forms, the main ones being A-form, B-form and Z-form. As described above, the B-DNA is made up of two strands, each forming a right-hand helix, which are held together through hydrogen bonds between the bases (figure **5B**). Since the two strands are wound around the same axis, the base pairs are stacked almost on top of each other. The bases are in anti-conformation and the two strands are antiparallel, as the bases of the one strand are found in 5' to 3' orientation and bound to the complementary bases of the opposite strand, found in 3' to 5' orientation.

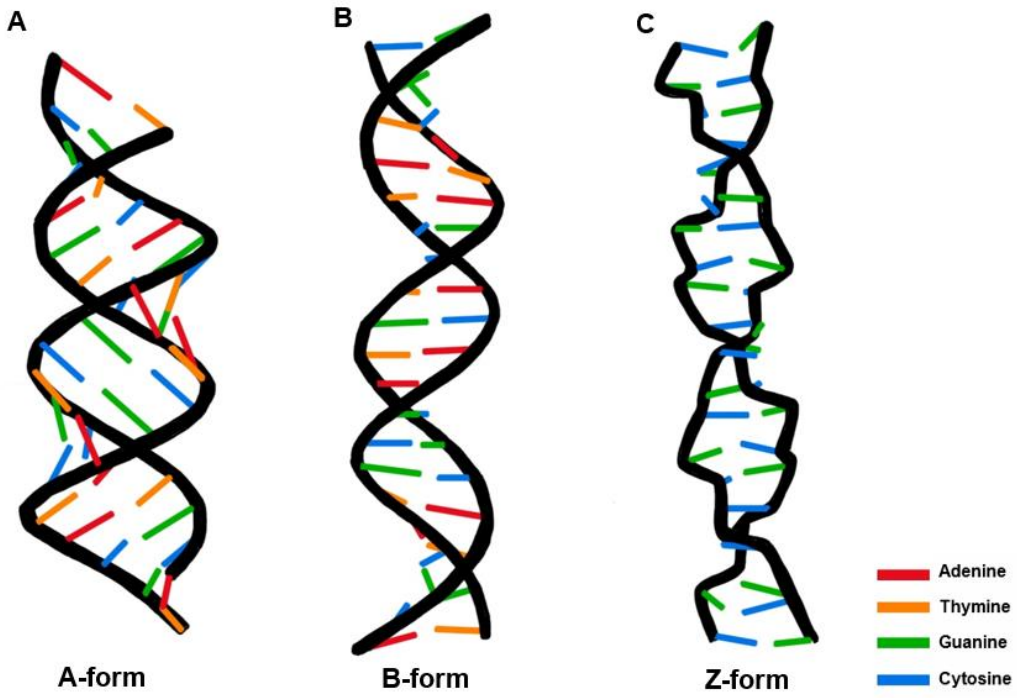


Figure 5: *Different DNA forms* ***A.*** *A-form* ***B.*** *B-form* ***C.*** *Z-form*[2,17–19]

1.3.2 A-form of DNA

A-form is also made up two strands forming a right-hand duplex, similar to B-form[20]. In the B-form, the backbone ring is in the C2' endo-conformation, which means that the structure is positioned so that C2 faces upwards, in the same direction as the base (figure **6**). Conversely, the sugar in the A-form is in the C3' endo-conformation, therefore the sugar is arranged so that C3' faces upwards in this case, in the same direction as the base (figure **6**)[21,22]. This DNA form is preferred by RNA-DNA and RNA-RNA duplexes, as the sugar's orientation allows more room for the 2' –OH group present in RNA. The helix formed in A-form is wider and shorter than that in B-form, as the base pairs within the duplex are not placed directly on the central axis, therefore, unlike the B-form, the pairs are stacked a little off-centre (figure **5A**). This arrangement of the pairs creates a hole within the duplex, a feature absent in both B and Z-form, making it a less stable structure. It has been observed that A-DNA prefers to form in purine stretches found in DNA, and that as few as 4 consecutive purines suffice for a local A-DNA helix to form; it should however be noted that the likelihood of this occurring differs between sequences. As RNA-DNA hybrid molecules prefer the A-form, it has been suggested that a change of B-DNA to A-DNA occurs during transcription, where the mRNA molecule forms complementarily to one of the DNA strands. A-form has also been associated to non-RNA related processes, one example being its presence in long terminal repeats (LTRs) of transposable elements, which have the ability to readily form A-DNA as they often contain purine stretches.

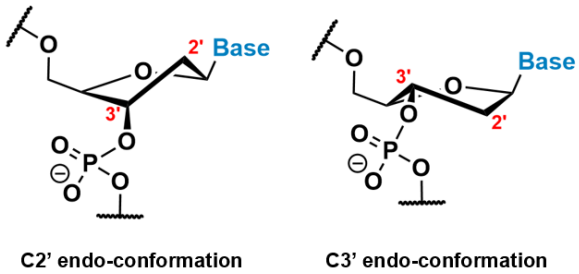


Figure 6: *Two different conformations of the backbone ring, C2' endo-conformation and C3' endo-conformation*[21,22]

1.3.3 Z-form of DNA

The third type of DNA structure is the Z-form, which was discovered in 1984 by Rich, Nardheim and Wang[18,23]. In contrast to A-DNA and B-DNA, the helix formed by the two DNA strands in Z-form is left-handed, and the helix is narrower and more elongated compared to the other two, with the two strands antiparallel to each other, similar to the B-form (figure **5**). Z-DNA forms in stretches of alternating purines and pyrimidines, for example GCGCGC, and can be identified by its characteristic zigzag pattern within the backbone (figure **5C**). This occurs as a result of the purines found in a different conformation in the Z-form, known as syn-conformation, placing them back over the sugar ring, which is in C3' endo-conformation, as in the A-form. However, the pyrimidines remain in anti-conformation, and their sugars in C2' endo-conformation[21,22]. This leads to the helical structure being made up of dinucleotide repeats, in which anti and syn-conformations of the bases alternate in succession along the chain forming the zigzag pattern of the helix, from which the DNA type got its name. In terms of its biological involvements, Z-DNA has been postulated to play a role in terminal differentiation and as a transcription enhancer.

Other than the factors mentioned above that favour a specific form of DNA, there are also certain conditions that determine what structure the DNA molecule is going to take. The first is related to the ionic or hydration environment, which has the ability to facilitate conversion between the different helical forms of DNA. For example, A-form favours low hydration, whereas Z-form favours high salt environments[21]. Secondly, the form the DNA adopts also depends on the DNA sequence. For example, A-form favours certain stretches of purines, Z-form is most readily formed by alternating purine-pyrimidine regions and B-form is adopted in mixed sequences. Another determining condition is the presence of certain proteins, which have the ability to bind to DNA molecules that have adopted one helical conformation and force the molecules to change to a different one. For example, certain proteins that bind to B-DNA can drive the molecule to adopt either an A- or Z-conformation[21,24].

1.4 DNA targeting *via* small molecules

1.4.1. DNA intercalation

The discovery of the significant features of DNA molecules described in **section 1.2**, was shortly followed by exploring application of DNA-ligand complexes. Small molecules have been developed that interact with nucleic acids either covalently or non-covalently and disrupt their natural biological functions[25–27]. In this thesis, the focus will be on non-covalent interactions.

One of the ways non-covalent groups binds to DNA is through intercalation, whereby small molecules are reversibly inserted between the planar adjacent base pairs of double-stranded DNA, thus stabilizing the dsDNA helix[28–30]. In order to accommodate the intercalator, the DNA chain lengthens and partially unwinds through a decrease of the helical twist, which allows the molecule to sit between the base pairs[31–33]. The local deformation of the DNA backbone results in both structural and mechanical changes to the DNA chain, which disrupt certain processes, such as DNA replication and transcription, and often this can lead to cell death. DNA intercalator molecules typically contain planar polyaromatic rings, allowing π-π interactions with the base pairs[34]. Additional interactions that are often used to help maintain the binding are Van der Waals forces and hydrophobic interactions. As DNA intercalators are capable of inhibiting cell division and cell growth, they can be used as anti-cancer compounds, anti-parasitic compounds and anti-microbial agents[31,35].

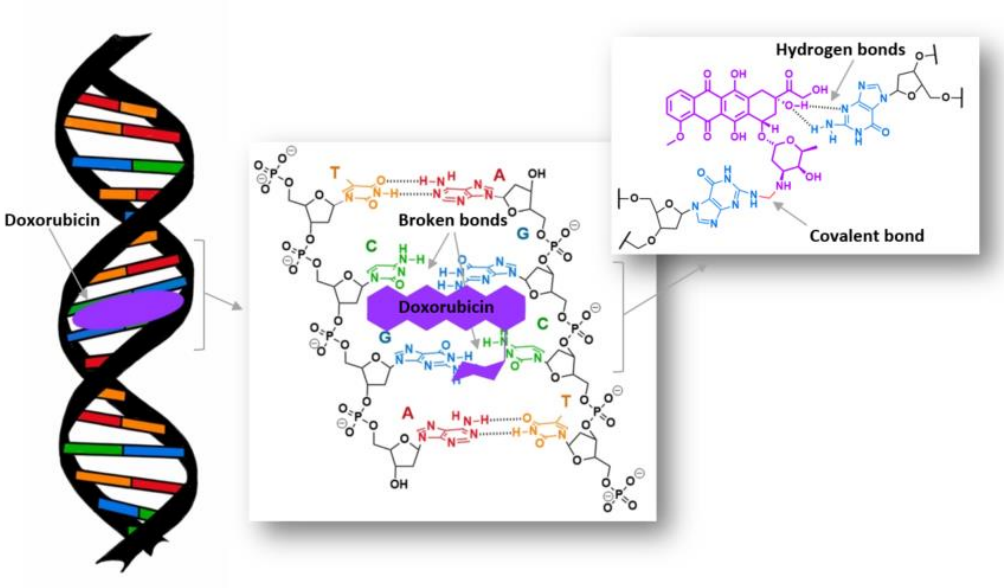


Figure 7: *Doxorubicin binding to DNA through DNA intercalation*[31,33,36,37]

A striking example of a DNA intercalator is the compound known as doxorubicin, a molecule used extensively in chemotherapy for the treatment of several types of cancer[36,38].

Doxorubicin belongs to a class of compounds called anthracyclines, it is a planar molecule that has the ability to intercalate between neighbouring DNA base pairs, anchoring on one side by one or more sugar moieties located in the minor groove. The compound forms a covalent bond with guanine mediated by formaldehydes, as seen in figure **7**, whilst forming hydrogen bonds with another guanine base on the opposite strand[39,40]. The act of Doxorubicin intercalating into DNA, leads to the inhibition of Topoisomerase II, an enzyme active during DNA replication[41]. Inhibition of topoisomerase II in dividing cells ultimately leads to cell death. As uncontrolled cell growth and division are key characteristics of cancer cells, doxorubicin significantly slows this process down, thus aiding chemotherapy treatments.

1.4.2 DNA minor groove binding

Another method through which small molecules bind to DNA non-covalently is through groove binding. Multiple sites for interaction can be found in the major groove, which provide easy access for bulky compounds and constitute strong binding sites for proteins, oligonucleotides and peptide nucleic acids (*vide infra*)[42,43]. On the contrary, although the minor groove has fewer potential interactions due to its size, it is as the main target for non-covalent antibiotic and anticancer small molecule drugs to attack and bind to[44,45]. Furthermore, the shape of the minor groove provides a tighter fit for molecules, as it is deeper and narrower than the major groove[34,46]. Minor groove binders normally have a crescent-shaped aromatic framework with both electron-donating and electron-accepting groups that are capable of hydrogen bonding, with a preference for A-T rich sequences. Other important types of interactions for minor groove binding include electrostatic and Van der Waals interactions.

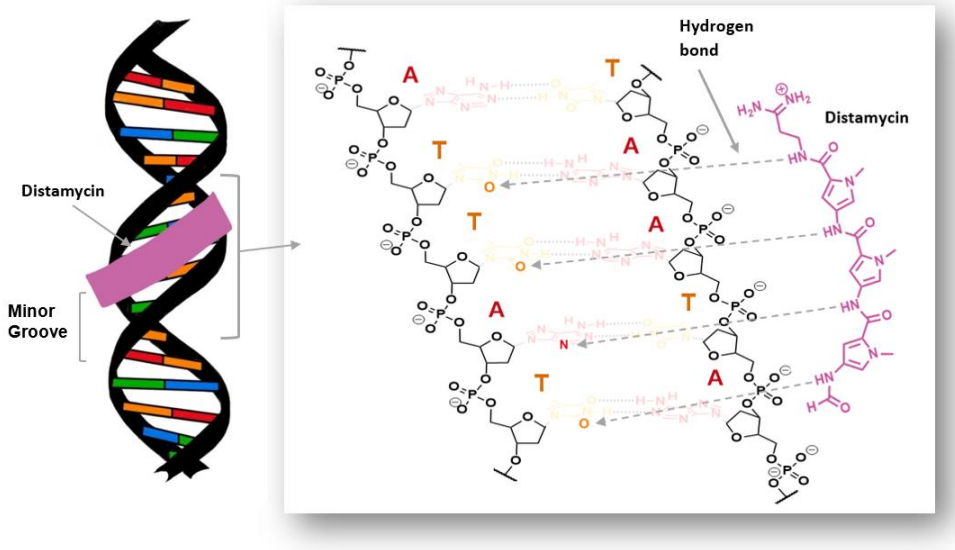


Figure 8: *Distamycin binding to dsDNA's minor groove*[44,45,47,48]

A minor groove binding drug that has been studied excessively for its anticancer activity is distamycin. Distamycin reversibly binds to the minor groove region of double-stranded DNA through electrostatic interactions, hydrogen bonding and Van der Waals interactions with specificity to A-T rich regions[47,48]. It is a natural antibiotic that has been found to have anticancer activity, however it is too toxic to be used in cancer therapy. Therefore, it has served as the basis for a number of studies and for the formation of a group of heteroaromatic compounds called lexitropins, which are based on its structure. Distamycin is made up of 3 pyrrole rings, which are joined through amide bonds, and an amidino side chain, which is an important group for the electrostatic interactions between the molecules and the duplex DNA (figure **8**).

1.5 Nucleic acid therapeutics

1.5.1 Introduction of field

The field of nucleic acid therapeutics, where nucleic acids themselves or closely related molecules are used to treat diseases has developed over the last 50 years. Contrasting to conventional drugs, which normally target key proteins in diseases, nucleic acid therapeutics aim to cure or achieve long-lasting effects in disease through gene regulation or modification. This idea dates back to 1978, when Zamecnik published a report describing how an oligonucleotide can be used to control gene expression[49]. Zamecnik's findings indicated the possibility of controlling the expression of any gene through synthetic nucleic acids. However, numerous questions arose concerning the delivery of the molecules into the cells, the tissues that could be targeted, the complications with large scale production and more. The amount of uncertainty about key factors regarding nucleic acid therapeutics led to slow progression in this field. In addition, numerous barriers had to be overcome, such as the minimization of off-target effects, metabolic stability and safety features[50].

1.5.2 Antisense oligonucleotides (ASOs)

One of the most promising applications of oligonucleotides is their use as antisense agents, which are compounds that have the ability to bind to RNA sequences in a complementary manner. The idea behind antisense oligonucleotides (ASOs) is that they can bind complementarily to mRNA sequences through sequence-specific hybridization, resulting in the cleavage or disablement of the mRNA and therefore blocking the expression of the gene of interest[51,52]. There are three generations of ASOs. First generation oligonucleotides are synthesized by replacing one of the non-bridging oxygen atoms in the phosphate group. The oxygen atom is replaced with either a methyl group (methyl phosphonates), a sulphur

group (phosphorothioates) or an amine (phosphoramidates)[53,54]. Second generation ASOs are characterised by alkyl modifications at the 2' position of the ribose, with the aim to improve hybridization stability and binding affinity with the target mRNA sequence. Finally, third generation ASOs were developed through chemical modifications of the nucleotide, more specifically to the furanose ring in the backbone[51,53,54].

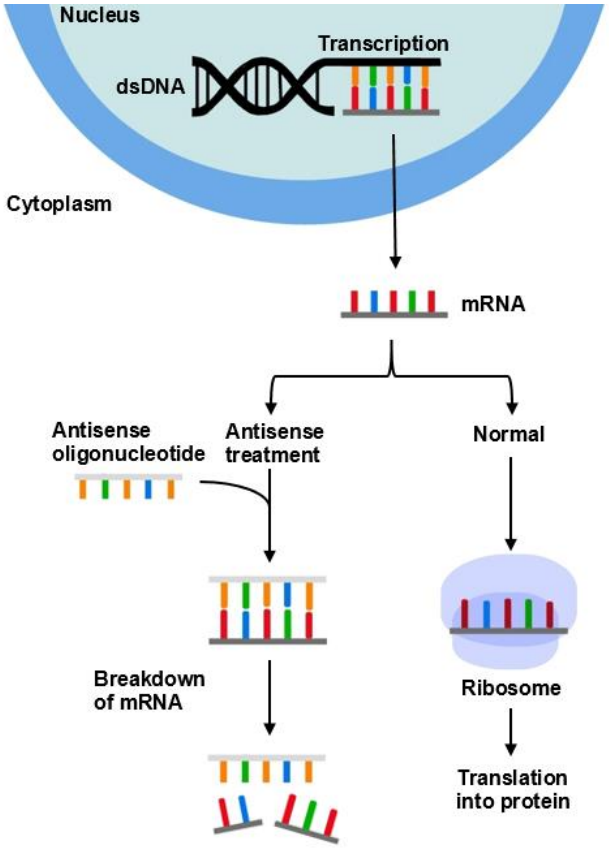


Figure 9: *Process of antisense oligonucleotide treatment*[49,52,53]

1.5.3 First oligonucleotide drugs released

The first oligonucleotide drug brought to the market was Vitravene; it was released in 1998 for the treatment of cytomegalovirus (CMV) retinitis in carriers of human immunodeficiency virus (HIV)[55]. The oligonucleotide was designed to bind in a complementary way to CMV mRNA transcripts of the major immediate-early region 2 of CMV, which encodes proteins that regulate viral gene expression.

Therefore, through the binding of Vitravene to the mRNA strand, the expression of the proteins was prevented ultimately blocking the spread of the virus in the body[55]. The drug was very successful in clinical trials, showing activity against CMV strains that drug alternatives at the time could not affect. However, Vitravene could be described as a scientific landmark but a commercial disappointment. Despite the initial success of the novel

drug, it was withdrawn in 2001, as the rising of a different kind of therapy significantly decreased the number of relevant cases, therefore rendering Vitravene redundant. Nevertheless, from a scientific perspective, the clinical success of the drug paved the way for future treatments based on antisense oligonucleotides.

After Vitravene, the next nucleic acid drugs released to the market were Macugen and Kynamro, in 2004 and 2013 respectively. Macugen was approved for the treatment of neovascular age-related macular degeneration (AMD)[56,57]. It is a 28-base RNA aptamer – a single-stranded oligonucleotide that folds into a defined architectures and binds to targets- that selectively binds to the 165th amino acid of vascular endothelial growth factor (VEGF), blocking its activity. VEGF is known to be a key molecule in the development of AMD, therefore making it a good therapeutic target for the treatment of AMD. Additionally, Kynamro was approved as a therapy for homozygous familial hypercholesterolemia, a genetic condition characterized by extremely high levels of low-density lipoprotein-cholesterol (LDL-C)[58,59]. The latter is a 20-base second generation antisense agent, it binds to the mRNA that codes for apolipoprotein B100 (ApoB100), a protein that plays an essential role in the production of LDL-C, therefore preventing the production of ApoB100 ultimately lowers LDL-C levels. Although Macugen and Kynamro were both met with initial success, they were both side-lined due to the emergence of other therapies. Nevertheless, they are both considered significant milestones in nucleic acid therapeutics, as they proved that systemically administered nucleic acids could engage with their intended target and, as a result, control the expression of the gene involved in the relevant disease.

Finally, the drug Spinraza was released in 2016, which was the first drug intended to treat spinal muscular atrophy (SMA), a fatal childhood disease[60]. SMA is a disease caused by the lack of the survival of motor neuron protein (SMN) which in turn leads to the death of motor neurons and ultimately leads to muscle wasting. The SMN protein is produced predominantly by the SMN1 gene in healthy humans, which produced functional SMN proteins[61]. In contrast, the SMN2 gene also found in humans only expresses a small amount of functional SMN protein due to an mRNA splicing error, specifically the mistaken removal of exon 7. Patients with SMA lack the SMN1 gene therefore do not produce enough SMN protein from solely the SMN2 gene. Spinraza is an antisense oligonucleotide that binds sequence-selectively near the exon 7 in the premature mRNA, preventing its removal. This leads to a significant increase of functional SMN protein by the SMN 2 gene. The profoundly positive data gathered from the relevant testing led to the release of this drug to the market, forever shifting the foundation of nucleic acid therapeutics. The drug could be administered to the patients through the central nervous system; it then entered the associated cells, and subsequently controlled gene expression in the target cells. When the drug was

administered to children suffering from this fatal disease, significant improvements were noted in both motor function and survival rates[60]. The overwhelming success of Spinraza led the way for the approval of several other nucleic acid-based treatments, including antisense oligonucleotides, small interfering RNAs and, more recently, mRNA vaccines[62–66]. The recent success of the latter further validated the impact nucleic acid therapeutics can have in the pharmaceutical world, making long-term investments in this field a worthwhile practice.

1.6 Triplex DNA molecules

1.6.1 Discovery and Hoogsteen binding

Due to the importance of DNA with regard to the genetic information of living organisms, extensive research has been performed in order to fully understand its function and structure, two elements directly linked to one another. Numerous studies have demonstrated that DNA can form a number of different arrangements, the most important being the double helix. However, it was also discovered that DNA can form multi-stranded helices, through the folding of the two strands or through the association of multiple DNA strands, ranging from two to four. Only four years after the discovery of the double-helical structure of DNA, Felsenfeld and Rich reported in 1957 that nucleic acids also have the ability to form triple helices[67]. Triple helices are helical structures made up of 3 DNA strands, with the 3rd strand binding to the major groove of a DNA double helix through hydrogen bonds, known as Hoogsteen bonds (**figure 5, 9**)[68]. The structure of purines carries the potential for hydrogen bond formation with an additional strand, as such the third strand binds to the purine-rich strand of the duplex DNA.

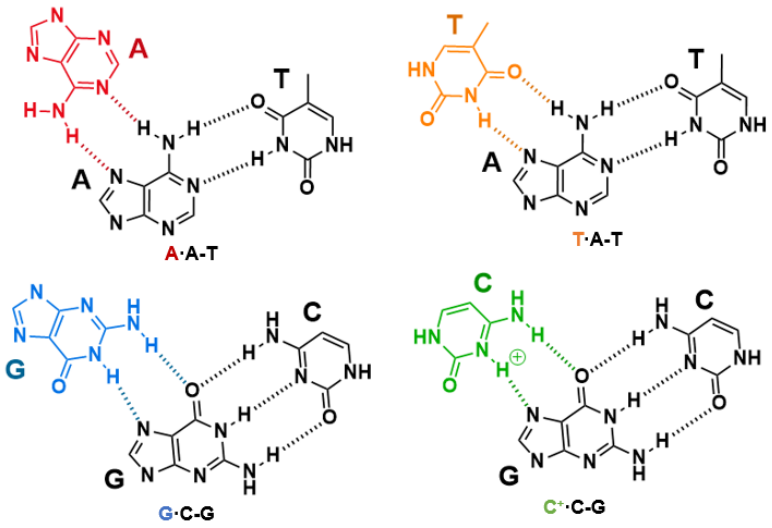


Figure 10: *Triplex formation by Watson-Crick and Hoogsteen pairing, the latter highlighted in colour*[68]

Triplex DNA molecules can be divided into two major categories, intramolecular and intermolecular, depending on the formation of the triplex. An intramolecular triplex, or H-DNA, occurs naturally in areas where mirror repeat symmetry is present between the strands, whereby one half of the "*tract"* can dissociate into single strands by using the energy provided by supercoiling[69]. The triplex forms when one of the newly created single strands swivels its backbone parallel to the purine-rich strand in the remaining duplex, where it forms a three-stranded helix, therefore leaving the complementary strand of the other half unpaired. Due to the mirror symmetry, four isomers can exist for the same region, depending on which strand acts as the third strand when forming the triplex.

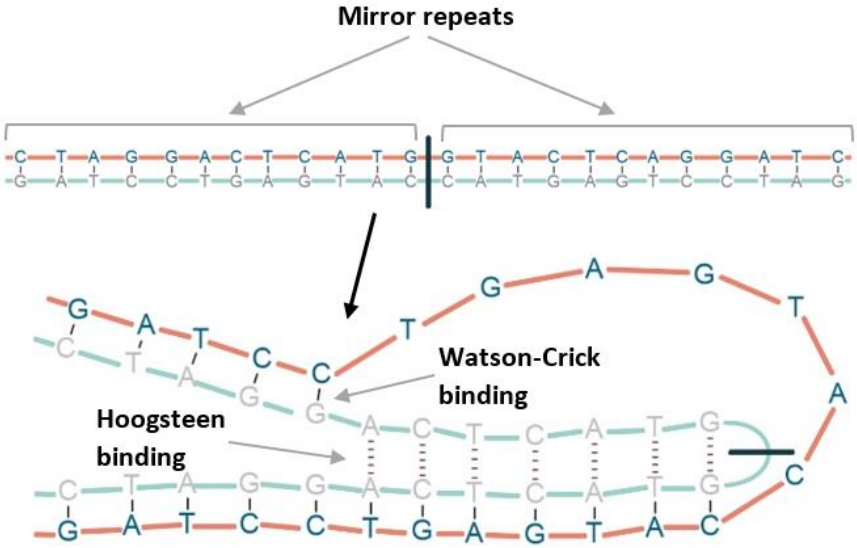


Figure 11: *Intramolecular formation of triplex DNA with mirror repeats*[69]

1.6.2 Triplex-forming oligonucleotides (TFOs)

Intermolecular triplexes are formed between a DNA double helix and a second DNA molecule, called a triplex-forming oligonucleotide (TFO), which provides the third strand (figure **12**)[70]. TFOs are major groove ligands that bind to DNA molecules in a sequence-specific manner through the formation of hydrogen bonds[25,71]. These binding agents have significant therapeutic potential as they can inhibit gene expression, target genes linked to disease for inactivation and stimulate DNA repair.

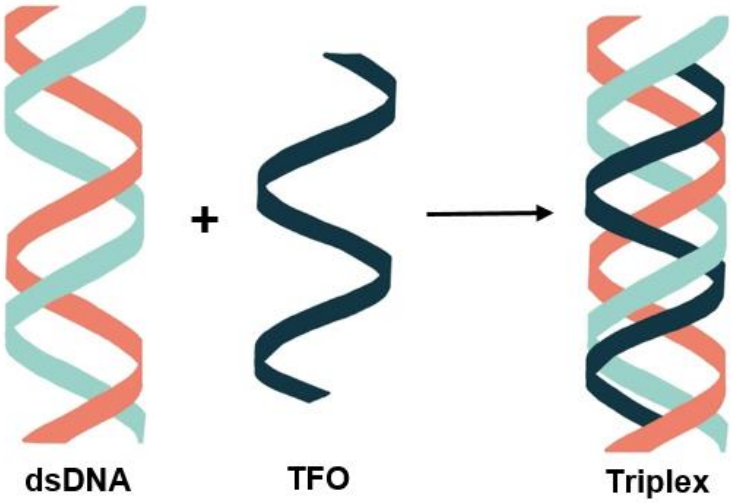


Figure 12: *Intermolecular triplex formation*[25,71]

Two triple helical DNA motifs have been discovered, differing in base composition and in the orientation of the additional strand[72]. In both types, the third strand follows a path through the major groove of the duplex DNA and requires a homopurine-homopyrimidine sequence region. The motifs are known as the pyrimidine and purine binding motifs, or the parallel and antiparallel binding motifs respectively. In the pyrimidine binding motif, the homo-pyrimidine TFO binds to the polypurine strand of the dsDNA via Hoogsteen hydrogen bonding in a parallel l fashion, forming C^{+}·G-C and T·A-T triplets[73–75]. Similarly, in the purine binding motif, the homo-purine TFO binds to the polypurine strand of the dsDNA molecule via reverse-Hoogsteen hydrogen bonding in an anti-parallel fashion, forming G·G-C, A·A-T and T·A-T triplets[74]. In the pyrimidine motif, protonation of the cytidine is required for the formation of the triplex (**figure 7**).

As TFOs have the ability to bind to duplex DNA with high affinity and specificity, they are considered ideal compounds for targeting specific genes, in order to modulate their function and/or structure in the genome. Therefore, triplex technology holds the potential to be used to regulate processes, such as cell signalling, cell proliferation and carcinogenesis, by identifying and targeting unique TFO binding sites in the associated genes[76–79]. However, the formation of both intramolecular and intermolecular triplexes is dependent on a number of factors, including base composition, length of sequence, ionic conditions and temperature[80].

In order to confirm the potential of TFOs in therapeutics, it is important to investigate the abundance of TFO target sequences in mammalian genomes. More specifically, triplex technology requires unique purine-rich regions of approximately 15-30 bases to be available in the DNA target molecules to allow TFO molecules to bind and form stable triplex helices.

An algorithm was initially developed by Gaddis and co-workers, which enabled scientists to search for TFO target sites in both strands of entire human and mouse genomes[81]. The results indicated over 2 million possible regions in these mammalian genomes that hold binding potential for TFOs. Analysis of these results showed that the in most annotated genes of both genomes, at least one unique TFO binding site was identified. These sites were mostly located in the promotor and/or transcribed gene regions, highlighting the potential of TFO based DNA technology. In 2009, Jenjaroenpun and Kuznetsov also developed a database with the ability to predict and annotate potential TFO target sequences and other regulatory DNA elements in the human genome[82]. These tools can today be used to identify sequences ideal for the formation of triplex helices in genes of interest.

1.6.3 Modifications to Triplex-Forming Oligonucleotides and their applications

It has been reported that TFOs can compete for transcription factor binding sites, inhibit initiation of transcription and cause premature termination of elongations[83]. Many groups have demonstrated this *in vitro*[84,85], however there only few examples *in vivo* due to a number of restrictions. Certain conditions can directly influence the effect of TFOs *in vivo*[83]*.* These include neutral pH, the presence of single-strand nucleases, low concentrations of magnesium and high potassium levels. Additionally, they also have to compete against other non-specific nucleic acid binding proteins, such as histones, and even sequence-specific binding proteins, such as transcription factors, affecting TFO activity. The targeted chemical modification of TFOs can help overcome a number of these barriers. For example, various alterations to the backbone can play a significant role in increasing the stability of the oligonucleotide[86–89]. Additionally modifying the pyrimidine bases can help address issues linked to pH-dependent hydrogen bonding between the strands[83,90]. The production of cationic oligonucleotides by modifying the phosphodiester linkages between the nucleosides leads to resistance against self-aggregation at physiological potassium concentrations and allows improved binding with the duplex DNA molecule[91].

Other approaches aiming to improve the efficacy of TFOs in biological systems include the chemical modification of certain parts of the molecules, most importantly the bases, the backbone, the 5’ and/or 3’ ends of the molecule, and the sugar moiety. Modifying the targeted areas mentioned above has been linked to improvements in the binding affinity, selectivity and stability of oligonucleotides. For example, it was observed that replacing

guanine with 6-thioguanine[92,93], and adenine with 7-dezaxanthine[94], prevented the formation of unwanted intramolecular secondary structures within TFOs (figure **13**).

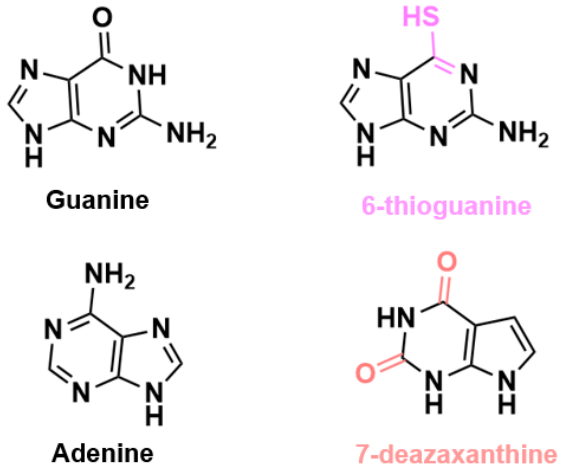


Figure 13: *Modified nucleobases, guanine and a guanine modification 6-thioguanine, adenine and a modified adenine nucleobase 7-deazaxanthine*[92–94]

With regard to the potential of TFOs in gene modulation, Hogan and co-workers in 1988 demonstrated the inhibition of transcription by TFO binding[71]. Since then, a number of groups have reported using TFOs as a tool to inhibit the expressions of various genes in different systems. This is accomplished through the binding of TFOs to promoter regions or other targets in or near the genes of interest. A number of publications report chemically modified TFOs inhibiting the transcription of a variety of genes, which include several HIV genes, the HER2 gene related to breast cancer, and the ETS2 gene found in prostate cancer cells[95–97]. These discoveries further highlight the promise these compounds hold as therapeutic agents against various diseases and types of cancer.

In addition, TFOs have also been investigated as tools for directing site-specific DNA damage, by using TFOs to target DNA damaging agents. TFO-directed DNA damage leads to the stimulation of recombination, mutation and DNA repair at the sites of interest, differing from TFO-directed transcription inhibition[77]. Therefore, instead of regulating gene expression, this approach aims to directly inactivate genes.

Chemically modifying the backbone of triplex-forming oligonucleotides has led to the discovery of two types of oligonucleotides that hold great promise, namely locked nucleic acids (LNAs) and peptide nucleic acids (PNAs), with the latter being the main focus of this project.

1.6.4 Locked Nucleic Acids

A family of well-known modified oligonucleotides are locked nucleic acids (LNAs), which were introduced as a result of studying the physical recognition properties of DNA and RNA[98–100]. One is known as entropy, which can be described as a measure of how much the energy of molecules or atoms becomes more spread out in a process. The hybridization of an oligonucleotide to a single-stranded nucleic acid is entropically unfavourable, caused by the "freezing out" of any free internal bond rotations present in either molecule[101,102]. It was believed that the conformational flexibility of the oligonucleotide could be a factor affecting binding affinity to its target. Therefore, it was suggested that a more conformationally rigid molecule would cause less entropy changes. A chemically modified type of oligonucleotide created following this theory are LNAs, which are conformationally locked oligonucleotide analogues made up of modified RNA monomers. LNAs are also known as bridged nucleic acids (BNAs), due to the oxymethylene bridge bond found between the C2' and the C4' of ribose sugar moiety (figure **13**)[98,103]. The bridge restricts the flexibility of the ring, therefore locking the monomer into a rigid bicyclic formation. An important feature of LNA is its ability to bind to DNA and RNA molecules with a significantly high binding affinity. Moreover, LNAs also exhibit improved Watson-Crick base-pairing selectivity compared to unmodified nucleic acids. LNA oligonucleotides have been and continue to be examined as a regulator of gene expression, with successful studies being carried out on LNAs in relation to gene silencing technologies[104–107].

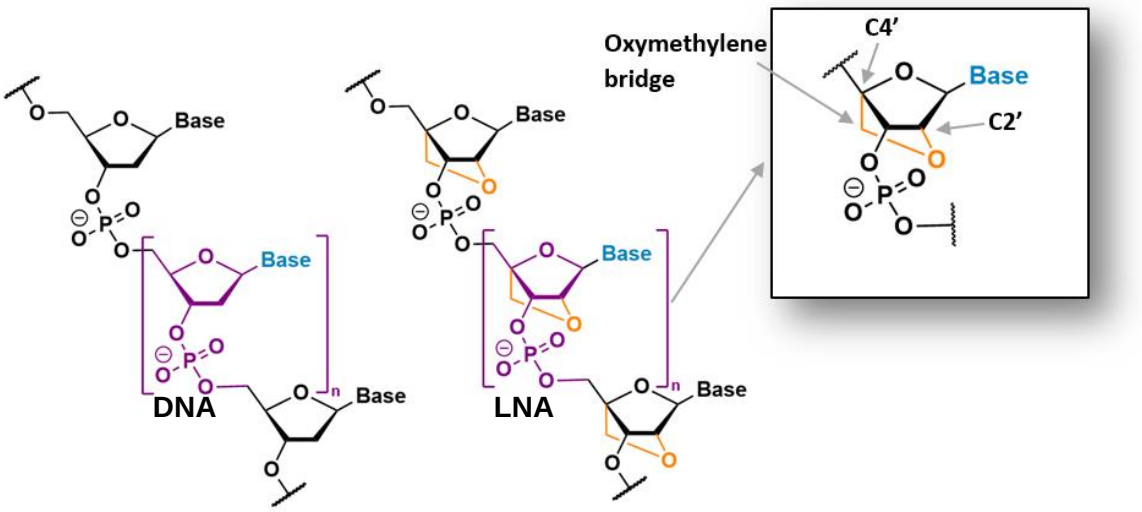


Figure 14: *Structure of a locked nucleic acid (LNA)*[98]

1.7 Peptide Nucleic Acids (PNAs)

1.7.1 Discovery and structure of PNAs

A recent chemically modified oligonucleotide that shows great promise is the peptide nucleic acid (PNA) oligonucleotide analogue, introduced by Nielsen and co-workers in 1991[43]. At that point, synthetic peptides had not been successfully used to bind sequence-selectively to target DNA sequences and oligonucleotides were difficult to prepare. However, since oligonucleotides had been shown to have the ability to bind to dsDNA through Hoogsteen binding, Nielsen and co-workers thought to the design a polyamide backbone that could recognise dsDNA through Hoogsteen-like base pairing, overcoming these obstacles. The team aimed to replace the phosphate backbone of DNA with a polyamide homomorphous to it, maintaining features such as the number of bonds in the backbone, and the distance between the latter and the nucleobase[43]. They chose thymine as the nucleobase for the PNA unit due to some promising factors, including the fact that it was the easiest to synthesize and there was already data proving it could take part in the triplex formation of oligonucleotides via Hoogsteen pairing[25,90,108–111]. Computational studies concluded that a 2-aminoethylglycine unit, in replacement of the sugar backbone, linked to thymine through a methylene carbonyl linker, was the best fit, shown in **figure 10**[43].

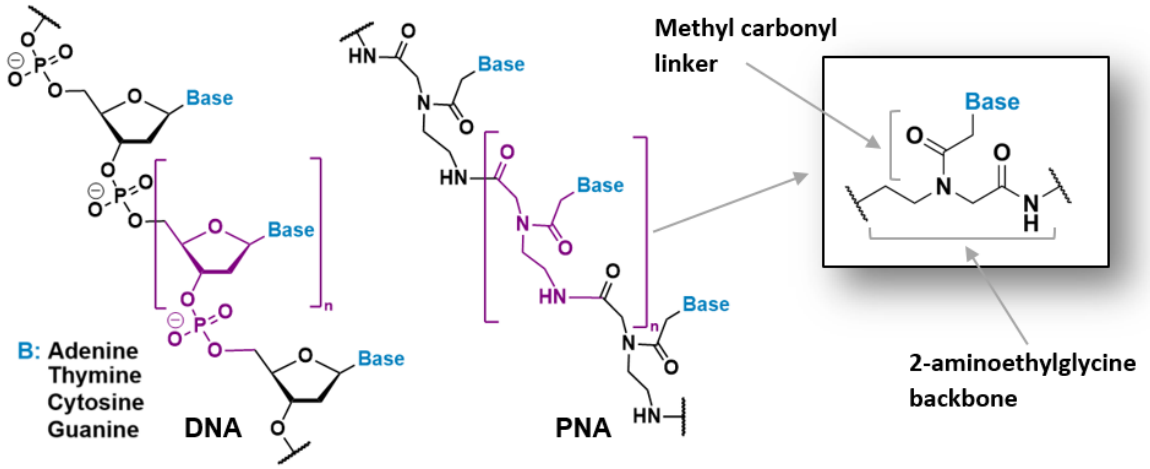


Figure 15: *The structure of a Peptide Nucleic Acid*[43]

Once the synthesis was completed, PNA sequences generated by repeated PNA monomers were incubated with dsDNA, made up of one strand that was complementary to the PNA strand and another that was complementary to the DNA strand, in order to examine binding. Under certain conditions, the homopyrimidine PNA strand had the ability to displace the non-complementary homopyrimidine DNA strand in order to bind to the complementary homopurine one, thus forming a PNA-DNA hybrid chain[43]. Due to the absence of 3' to 5' polarity in the PNA strand, the latter can bind to DNA/RNA in both parallel and antiparallel fashion. However, the antiparallel orientation is preferred, since T*m* values of antiparallel

PNA-DNA/RNA chimeras were calculated at 15-20°C higher than those measured when the strands were bound parallel to each other[112–115]. Additionally, in spite of the high stability of the dsDNA sequence, it was reported that the PNA-DNA chimera was more thermodynamically favoured and showed higher stability than the duplex DNA molecule; a speculation on why is discussed further in this section.

Another advantage of PNA sequences is that they can be easily synthesized using standard Boc or Fmoc protocols for solid-phase peptide synthesis[116,117]. Either protocol can be carried out straightforwardly, so as to synthesize PNA strands either manually or automatically.

A strand displacement complex was proposed in 1993 by Nielsen and co-workers, as seen in **Figure 16**[118]. It was suggested that once the two strands of the duplex DNA molecule were separated and the PNA strand bound to its complementary DNA strand, the high stability of the hybrid molecule PNA-DNA meant that the DNA strand was kinetically trapped and remained stable[43,119]. Finally a second PNA strand could bind to the same DNA strand through Hoogsteen binding, leading to the formation of a triplex molecule, which is even more stable than the chimera, trapping the DNA strand between the two PNA strands, and forming a PNA·DNA-PNA (py·pu-py) complex . Therefore, this system can be described as a four-stranded complex, whereby the target DNA strand is Watson-Crick paired to the one PNA strand and Hoogsteen paired to a second PNA strand and the displaced DNA strand remains as a single strand. This feature of peptide nucleic acids has been extensively investigated, leading to the conclusion that there are certain factors that favour strand invasion and others that disfavour strand invasion. Some of the favouring factors include secondary structures present at target, supercoiling, A-T rich regions, low ionic strength and potential for triplex pairing. Alternatively, factors that disfavour strand invasion include relaxed target DNA, high salt and the presence of divalent cations[119–125].

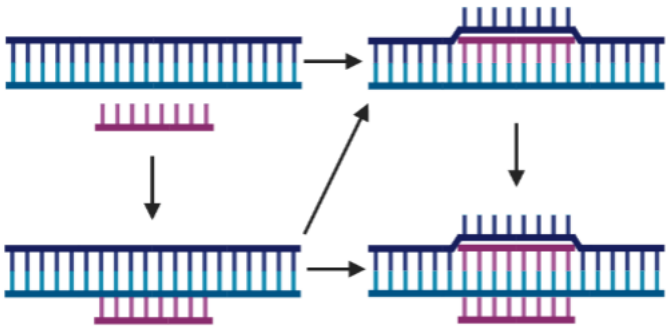


Figure 16: *The strand displacement complex suggested by Nielsen*[43,118]

1.7.2 PNA-DNA hybrid stability

The polyamide-based synthetic DNA mimic is more chemically stable than DNA and RNA. PNA has been found to be highly stable in acidic conditions and moderately stable in a basic environment and at high temperatures[126]. Furthermore, from a biological perspective, we observe that they are resistant to enzymes/chemical such as proteases and nucleases, unlike their biological counterparts DNA and RNA, which are highly sensitive to these enzymes[127]. This means that they are not degraded in cells and display high biochemical stability, a quality significant for their involvement in *in vivo* studies. One theory for their resistance to enzymatic degradation is that the amide bonds holding the PNA monomers together are not found in natural peptides, and are therefore not recognized by the enzymes. Another significant feature of PNA is its sensitivity to the existence of a single mismatched base pair when forming the PNA-DNA duplex[128–132]. It was reported by Egholm in 1993 that the presence of a single mismatch within a PNA-DNA molecule led to a reduction in the *T*m by up to 15°C[132]. Due to such high levels of discrimination being noted, this characteristic has been used as the basis for the development of various PNA-based strategies for the analysis of point mutations[133–137]. Also of importance for *in vivo* work is the fact that PNA seems not to interact with proteins sequence-specifically or otherwise, as well as DNA or RNA molecules do. This attribute is a significant advantage, since it allows PNA to avoid taking part in unwanted interactions with different proteins, thus reducing any off-target effects[138].

1.7.3 PNA binding techniques

PNA strands can bind to DNA strands *via* both Watson-Crick binding and Hoogsteen binding, allowing the formation of various helical DNA structures (figure **17**)[43,115,132,139]. Simple Hoogsteen binding leads to the PNA·DNA-DNA complex. The strand displacement complex described above (figure **16**) can lead to the formation of a PNA-DNA hybrid molecule, alongside the displaced complementary DNA strand. Additionally, another PNA strand can bind to the PNA-DNA hybrid through Hoogsteen binding forming a PNA·DNA-PNA complex. Finally, certain conditions regarding the sequence of the dsDNA and the PNA strand, two PNA strands can invade the DNA molecule and bind to one DNA strand each, creating two PNA-DNA chimeras, a stable molecule however only possible if the PNA strands contain modified nucleobases.

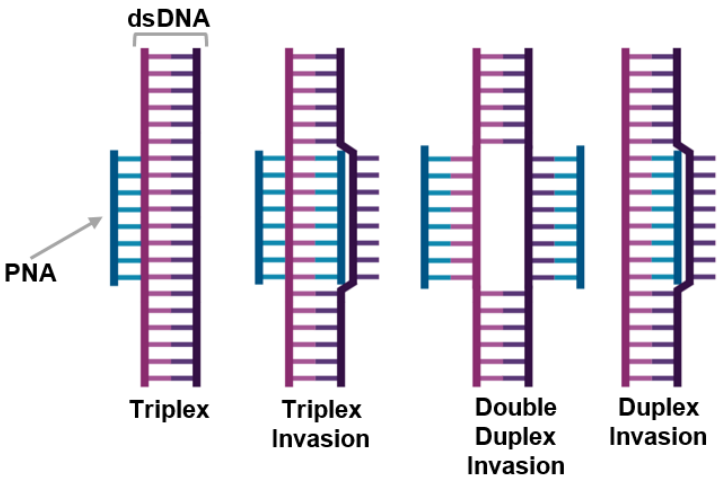


Figure 17: *Different ways PNA strands can bind to dsDNA*[139]

1.7.4 PNA modifications to overcome limitations

Even though PNAs can evade enzymatic degradation, other issues arise with their insertion and behaviour in cells. A major challenge in PNA technology relates to their poor cellular uptake[126]. Significant efforts have focused on identifying promising modifications that can be made to the molecule in order to tackle this issue[140]. A variety of solutions have been reported that indicate an increased uptake of the PNA into the cell, including coupling the latter to DNA oligomers or receptor ligands. An even more efficient way to improve the delivery of PNA strands into cells is through binding to specific peptides, such as cell penetrating peptides (CPPs)[141]. CPPs are a group of cationic peptides that are used as vectors for delivery of large cargo molecules into cells.

Aside of cellular uptake, other challenges include poor membrane permeability, low aqueous solubility and ambiguity in the DNA binding orientation of PNAs[142]. Thus, a number of chemical modifications were explored to address these issues, both concerning the backbone and the nucleobases. Regarding backbone modifications, the addition of cationic functional groups led to the improvement of aqueous solubility and alterations to the linker or the nucleobases resulted in improved control of DNA/RNA binding at physiological conditions[126].

1.7.5 PNA-DNA and PNA-PNA duplexes

As mentioned above, the PNA-DNA hybrid formed via the binding of a PNA strand to its complementary DNA strand found in a dsDNA molecule is significantly more stable than the DNA duplex, thought to be due to a number of reasons. A major factor that can explain this high stability is the lack of electrostatic repulsion between the DNA and the PNA

backbones[43,132]. Since DNA is negatively charged but PNA lacks charges, during the hybridization of PNA with DNA, PNA lacks electrostatic repulsion with its target DNA strand, thus offering PNA respective hybridizations. A penta-decamer sequence was used and the Tm values recorded for PNA-PNA, PNA-DNA and DNA-DNA were 67°C, 51°C and 33.5°C respectively[115]. These data points clearly indicate that PNA-PNA is the most stable structure, and that the hybrid DNA-PNA is more stable than the "common" DNA-DNA duplex. In addition, when studying the melting temperatures of the triplexes formed, namely PNA-DNA-PNA or $(PNA)_2$-DNA, these structures show a Tm > 70°C, which explains the high stability of these complexes[119,143–145].

Before PNA-PNA duplex formation was confirmed, Wittung and co-workers carried out a study investigating such a possibility in 1994[115]. The team designed two complementary PNA decamers that included all four nucleobases, which were not self-complementary. Through circular dichroism spectroscopy, it was shown that the two PNA strands could hybridize to one another to form a helical duplex. This suggested that a (deoxy)ribose backbone is not essential for the formation of DNA-like helices or, in other words, that a helical structure can be achieved through the combination of base stacking interactions and hydrogen bonding, even when in the presence of a less constrained more flexible backbone lacking of electrostatic repulsive forces, such as a PNA backbone[115]. It was also noted that, whereas duplex DNA helicity is determined by the deoxyribose residues, duplex PNA is determined by the cooperative arrangement of the nucleobases.

It therefore becomes apparent, owing to the unique features of PNAs discussed above with regard to their physical and chemical properties, that they hold potential in various fields for application. Studies have already confirmed their potential as biomolecular tools, antisense and antigene agents, biosensors and molecular probes. Only a few years after Nielsen *et al*. had presented this DNA mimic, it had already been successfully used for hybridization *in vitro* for the quantification of telomere length, the isolation of transcriptionally active DNA, screening for mutation, and for the inhibition of human telomerase.

1.7.6 PNA technology applications

One of the most promising discoveries regarding PNAs is their medicinal use as antisense or antigene agents[123,126,142,146]. Antisense agents are described as oligonucleotides that have the ability to complementarily bind to RNA molecules, blocking their function, whereas antigene agents are designed to bind to dsDNA *via* the formation of triplexes. As PNA molecules have exhibited the ability to bind to both dsDNA and RNA molecules, they hold great promise in this field. Peptide nucleic acids are one of the most commonly used third-

generation ASOs. *In vitro* studies utilizing PNA technology reported that PNA could complementarily bind to DNA or RNA with high affinity and specificity and inhibit transcription and translation[147–149].

PNA technology has also been explored as a tool in *in vitro* diagnostic assays for clinical testing and environmental monitoring. Some of the *in vitro* diagnostic methods developed are briefly described below, and include chromosomal analysis, Fluorescence *in situ* hybridization (FISH), microarray and polymerase chain reaction (PCR). Due to PNA's significant hybridization features, there is a keen interest in using PNA-based technology for human chromosome investigations in different tissues through *in situ* detection[139,150–152]. Regarding PCR, it has been reported that PNA-based techniques have the ability to modulate PCR, therefore detecting genomic mutations through the capture of nucleic acids[153,154]. Such an example is a modified PNA-PCR method designed by Yu *et al.* aimed to screen a specific known cancer mutation[155]. The PCR technique was altered in a way that allowed it to screen cancer patients' genetic information with high accuracy and sensitivity and identify patients who carry the mutation of interest. Furthermore, the modified method was also able to quantify mutant DNA in order to predict the patients' response to treatment and could be used to monitor disease progression. FISH is a procedure used to identify and locate a specific DNA region in a chromosome. When this method is combined with PNA technology, the relative hydrophobic character of PNA allows for a better diffusion of the PNA probes through the cell wall of the mycobacteria[156–159]. Finally, another diagnostic tool that PNA technology has been combined with is a DNA microarray, a method normally used to simultaneously measure the expression levels of numerous genes. PNA microarray can be used to conduct a highly sensitive high-throughput analysis of samples through the hybridization of the synthesized PNA molecules with the targets[160–162]. Overall, the unique hybridization properties of PNAs have led to their introduction in a number of nucleic acid biosensors in an effort to enhance detection sensitivity.

1.8 Synthetic nucleobases

1.8.1 J base, M base

To overcome certain issues with natural nucleobases in triplex formation, such as the steric hindrance of the 5' methyl group and the protonation of cytosine, scientists explored the option of chemically modifying natural bases. For example, due to the unfavourable cytosine protonation necessary for the formation of the C^{+}·G-C triplet formation at physiological conditions, researchers examined the notion of developing neutral analogues of the cytisine[163–165]. Protonation of C in acidic conditions is essential, due to the fact that the extra

hydrogen provides a second position for hydrogen bonding in C-G Hoogsteen pairing, (figure **10)**. In 1991, Ono *et al*. presented a structure, now known as pseudoisocytosine or J base (figure **18A**), which has the ability to bind to guanine through Hoogsteen pairing and form a stable complex at physiological pH[166,167].

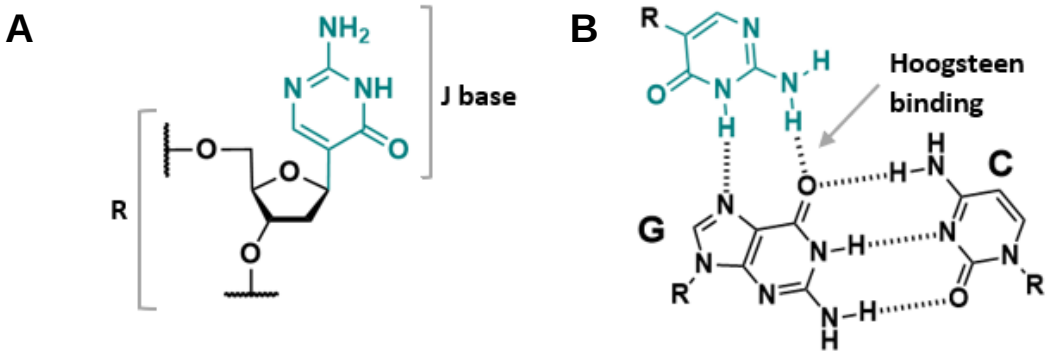


Figure 18: A. *DNA monomer with modified nucleobase, J base,* ***B.*** *Triplex binding of dsDNA with J base via Hoogsteen base pairing*

The neutral replacement of the protonated cytosine introduced the prospect of triplex formation in living cells, which was not favoured with the natural base, as cellular pH is normally above pH 7. When looking to overcome the same issue for PNA triplex formation with dsRNA, Zengeya *et al*. suggested that cytosine be replaced with a different heterocycle, namely 2- aminopyridine, which is known as M base (figure **19A**)[168].

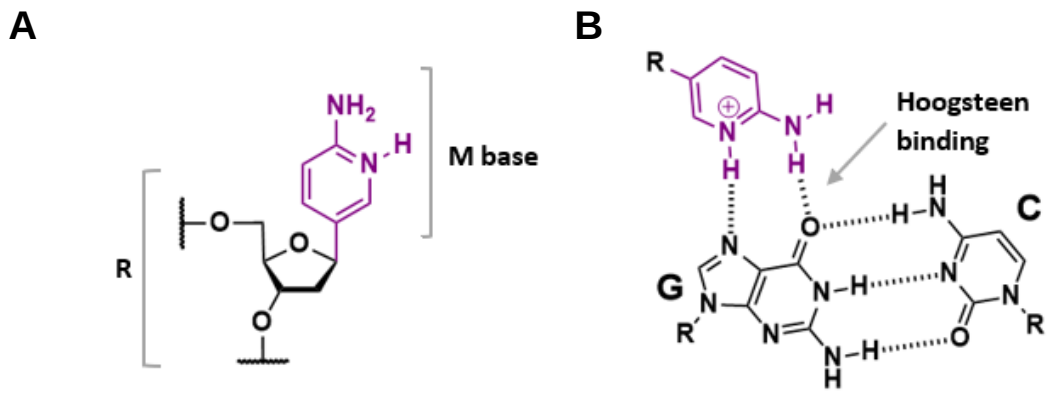


Figure 19: A. *DNA monomer with the modified base M base,* ***B.*** *Triplex binding of dsDNA with M base via Hoogsteen base pairing*

The team chose this ring as it is more basic than cytosine, which means that the structure would be more stable in physiological conditions. As a result, the M base allowed the formation of triple helices with high stability and sequence-selectivity with dsRNA under physiological conditions. When compared to the approach that uses the neutral J base, where the two bases were compared as to their ability to form triplexes with dsDNA and ds

RNA, as well as duplexes with ssDNA and ssRNA, the M base exhibited significantly higher affinity for all of the structures used[168]. Moreover, the M base also displayed improved cellular uptake and sequence selectivity when compared to the J base, thus leading to its characterization as a superior modified base.

1.8.2 E base

A different kind of synthetic nucleobase was disclosed by Nielsen *et al.* in 1997, aiming to selectively recognise thymine[169]. This idea was based on the fact that triplex targeting is mainly limited to homopurine stretches, therefore a synthesis of altered nucleobases that recognize T (-A) and C (-G) base pairs is required to expand the range of recognised sequences. At the time, several nucleobases had been created that could recognise cytosine, however there were no reported nucleobases that targeted thymine. The team decided on certain features that were required for the nucleobase to behave the way they intended. Firstly, the presence of a heterocycle was necessary, that would be connected to the PNA polyamide backbone through a linker long enough to avoid the steric hindrance of the 5'methyl group in thymine. Additional requirements included a hydrogen donor that would bind to the 4-oxo group in thymine, as well as a hydrogen acceptor aimed to form a hydrogen bond with one of the hydrogen atoms of N6 in adenine, if possible. Using computational model building, the ideal structure suggested was 3-oxo-2,3-dihydropyridazine that would connect to the PNA backbone through a β-alanine linker, the latter containing two additional atoms when compared to the original methylene carbonyl linker of the original PNA monomer, as seen in figures **20, 21**. The structure also lacked a hydrogen atom N1, aimed to further reduce any steric interference with the 5' methyl group. However, results indicated that this structure did not fully solve the steric hindrance issue with thymine, as E bound to U with stronger affinity than T, when uracil was introduced instead of thymine to assess the issue.

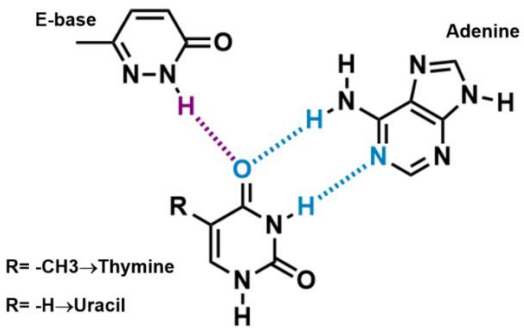


Figure 20: *Representation of an E·T-A complex*

1.8.3 New PNA monomer suggested by Topham

The idea presented by Nielsen *et al*. held great promise, however certain improvements needed to be made to the structure[169]. Topham *et al*. published a paper in 2021, in which they presented computational work postulating an improved linker for the E-base to the PNA backbone, which would solve the steric issue with thymine's 5-methyl group[170]. More specifically, the team aimed to develop a better linker for the E-base structure, which would enable improved recognition of the pyrimidine in T-A inversions. This feature is believed to unlock new potential in recognising pyrimidines in the major groove and targeting them with purine PNA sequences, ultimately using PNA technology for pu/py mixed regions. The team chose quantum chemical methods, as previous studies had proven that these techniques are more suitable when studying nucleic acids or their analogues. Quantum chemical methods are used to assess the conformational and energetic properties of complex molecules and possess features that involve various parameters required for these investigations that are harder to include when using molecular mechanics. The quantum chemical modelling techniques applied by Topham et al considered data regarding spatial constrains gathered from a hetero PNA·DNA-PNA (py·pu-py) triplex X-ray crystallographic structure[170]. Their purpose was to investigate the geometrical effects of the flexibility of different linker designs on the presumed binding interactions with T/U-A base pairs. Previous linkers presented by Nielsen et al, such as β-alanine and the 3' *trans* olefin side-chain E-base linker, were investigated in both the duplex and triplex structure models.

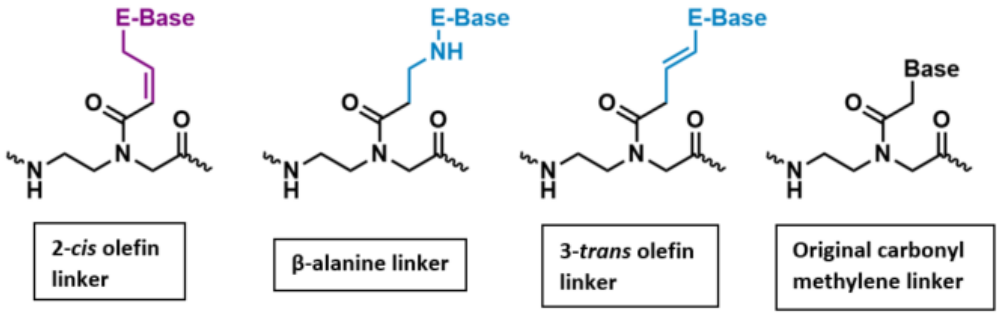


Figure 21: *The different linkers investigated by Topham compared to the original methylene linker introduced by Nielsen*[43,169,170]

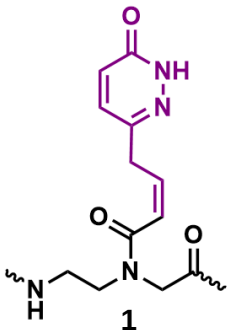
The results gathered through geometry optimisation and energy calculations agreed with the thermal stability data, leading to the proposal of an altered *cis* olefin side-chain linker design.

The study of the β-alanine linker showed that the addition of 2 atoms to the original PNA linker design provided significant advantages for both thymine recognition in DNA and uracil recognition in RNA. However, the geometry optimisation approach indicated that the

flexibility of the linker was strained due to the formation of an internal hydrogen bond interaction located between the carbonyl and amino groups, separated by two carbons.

Following an investigation of the more rigid side-chain linker proposed by Nielsen, the 3' *trans* double bond in the linker was not found to be compatible with the formation of a hydrogen bond between the O4 atom of thymine/uracil and the E-base. According to the above-mentioned data, the quantum chemical techniques pointed at replacing the 3' *trans* double bond with a 2' *cis*, postulated to allow the interactions between the E-base and the T-A inversion as proposed originally. It was also noted that, with this modification, the parameters regarding the 3D orientation of the E·T base pair were not affected, and predicted the same values as those gathered from the PNA·DNA-PNA triplex X-ray data. On the other hand, the parameters were found to be altered when the β-alanine linker was used[170].

In conclusion, the three linker designs analysed by Topham et al were all found to be partially rigid, with the 2' *cis* side-chain linker design being the most promising for a number of reasons. Namely, it enables the recognition of the 4' oxo group in thymine which is located in the major groove, it allows intra-strand base-pair stacking interactions found in PNA·DNA-PNA triplexes to be maintained, and finally, due to the arrangement of the linker, there is no steric hindrance issue with the 5' methyl group in thymine.



1

Figure 22: *Structure of the PNA monomer suggested by Topham, made up of an E-base as the nucleobase, attached to the 2-aminoethylglycine backbone through a 2-cis olefin side-chain linker*[170]

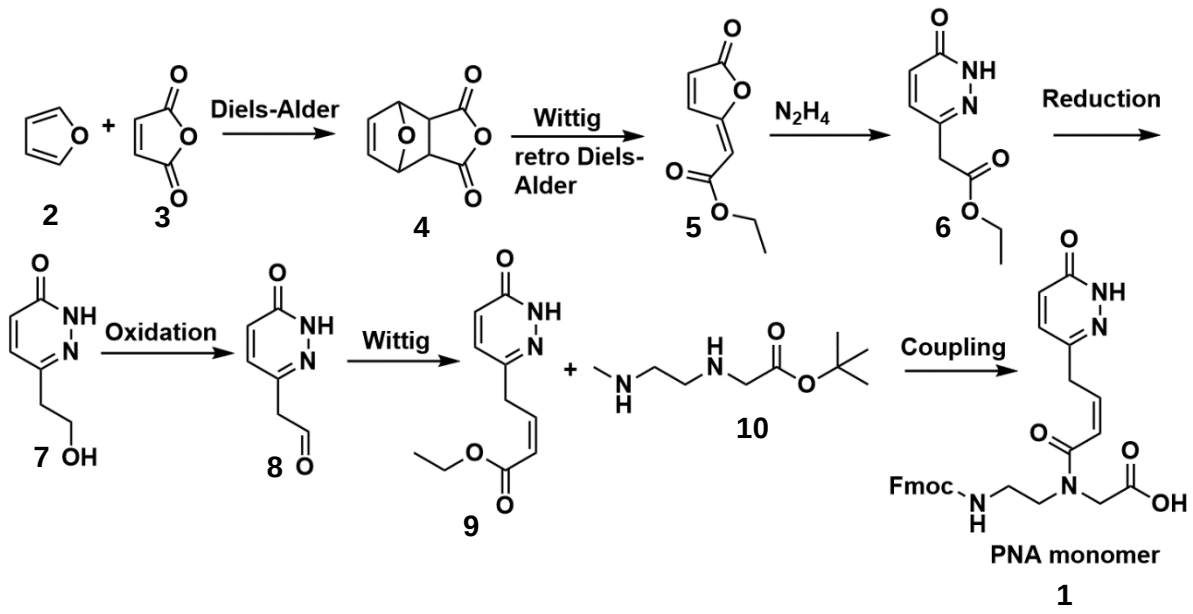
1.9 Aims of study

In this project, our aim is to design and complete a synthetic pathway to create the modified PNA monomer suggested by Topham et al. the PNA unit is made up of the standard polyamide backbone (2-aminoethylglycine), connected to an E-base, as the nucleobase, through a 2' *cis* olefin linker. From the information mentioned above, we believe this PNA monomer will have the ability to bind to T-A inversions in DNA through Hoogsteen-like pairing, forming the triplex E·T-A structure.

Chapter 2: Original synthesis of PNA monomer

2.1 Overall synthesis

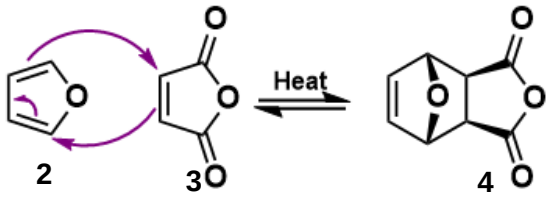
The original synthesis plan seen in scheme **1**, is partially based on a synthetic pathway published by Tomori[171]. In this paper the team continued studying the promising pyrimidine base analogues for the development of new synthetic oligonucleotides. They described synthetic pathways for four PNA monomers, two containing cytosine analogues and two thymine analogues. One of the thymine analogues is the structure described earlier as E base. The main difference between the E base PNA monomer described by Tomori and the one by Topham is the linker located between the PNA backbone and the E base. Therefore the synthetic pathway by Tomori was partially followed, carrying out the same steps in order to prepare the base, and modifying the steps involving the preparation of the linker.



Scheme 1: *Overall synthetic pathway to form a PNA monomer*

2.2 Diels-Alder step

The first step of the synthesis is a Diels-Alder reaction between maleic anhydride and furan. The Diels-Alder reaction was discovered by Otto Diels and Kurt Alder in 1928[172]. It is one of the most famous examples of pericyclic reactions, which are reactions where the electrons involved move around a “circle” in a concerted fashion and no intermediates are formed. The reaction often occurs easily through a single step by heating the reagents. It is a type of cycloaddition where two π molecules react, a conjugated diene and an alkene, creating a new cyclic structure with two new σ bonds. The reaction is characterised as a [4+2]-cycloaddition, as in a Diels-Alder reaction 4 π-electrons of the diene and 2 π-electrons of the alkene were involved.

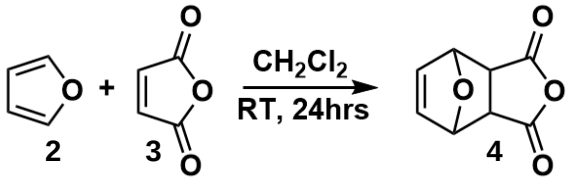


Scheme 2: *Mechanism of Diels-Alder reaction*

The structure of the diene can vary in several aspects, it can be cyclic or open-chain and have different kinds of substituents[173]. The one characteristic that is essential for the reaction to occur, is the component's conformation, which must be s-*cis*. If the diene is in s-*trans* conformation, and cannot adopt an s-*cis* conformation, then it cannot participate in a Diels-Alder reaction. This is due to the fact that the two ends of the double bonds of the diene aren't close enough to react with an alkene, and also because the 6-membered ring that forms from the reaction cannot have a *trans* double bond in it[174].

Two products can form from this reaction, an exo- product an endo- product, depending on the reaction. If the reaction is reversible, meaning it is thermodynamically controlled, the exo- product forms. In contrast, the endo- product is preferred when the reactions is irreversible, even though it is less stable than the exo product[174].

Here, maleic anhydride was successfully reacted with furan, undertaking an exo-selective cycloaddition, as seen in scheme **3**[175,176].

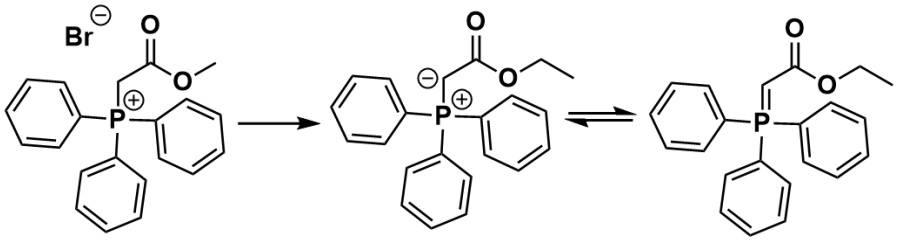


Scheme 3*: Diels-Alder reaction of maleic anhydride and furan*[175,176]

Furan, maleic anhydride and CH_2Cl_2 were stirred at room temperature (RT) for 24 hours, until all the starting material was consumed. TLC stained with vanillin was used to monitor the reaction, a stain commonly used to observe a range of nucleophiles, aldehydes and ketones. Once the reaction reached completion, the product was isolated using filtration, reaching a moderate yield of 58%. ^{1}H NMR confirmed the formation of the desired product, as all 3 peaks expected according to literature were present[175,176]. Due to the symmetry of the compound only 3 proton environments appear, with all of them integrating correctly to 2. The appearance of a singlet at 3.17ppm, corresponding to the protons alpha to the carbonyls, confirmed the synthesis of compound **4**.

2.3 Wittig and retro-Diels-Alder reaction

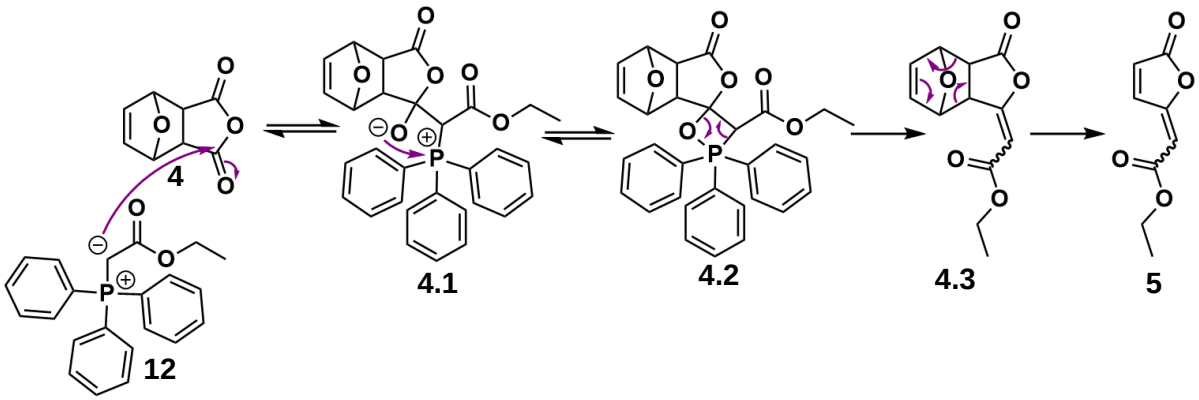
The next step was completed via a Wittig reaction followed by a retro-Diels-Alder reaction. A Wittig reaction is a convenient and relatively well studied way to form alkenes. The reaction happens between a carbonyl group and phosphonium ylide, which is the compound that forms from the deprotonation of a phosphonium salt with a strong base[177,178]. Ylide can also be represented as doubly bonded species, known as phosphoranes, as seen in scheme **4**.



Phosphonium salt 11 **Phosphonium ylide 12** **Phosphorane 13**

Scheme 4: *Example of phosphonium salt, phosphonium ylide and phosphorane*[178]

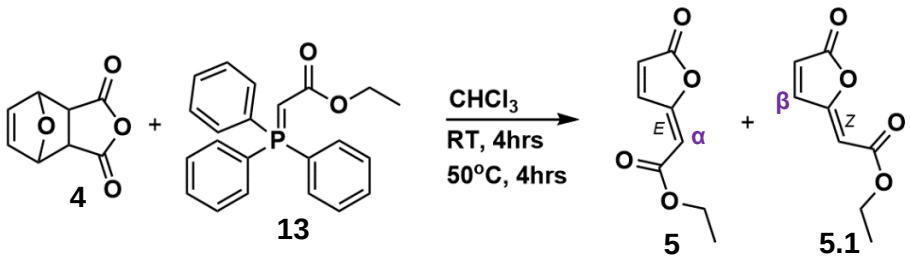
A Wittig reaction is made up three steps to this reaction. Firstly, the nucleophile in this reaction is the carbanion part of the ylide, which attacks the carbonyl group of the other molecule. Next, the newly negatively charged oxygen attacks the positively charged phosphorus, generating a four-membered intermediate ring, called an oxaphosphetane. Lastly, due to the newly formed ring being highly unstable, it undergoes elimination forming an alkene and generating triphenylphosphine oxide as a by-product. The mechanism described is presented in the first three steps of the reaction scheme **5** below, up to the formation of compound **4.2**. Wittig reactions can be complicated by their stereochemical outcome[178]. Conditions can control when a Wittig reaction is E isomer selective or Z isomer selective, commonly controlled by the substituent linked to the carbon atom of the ylide. *E* selective ylides are normally those whose anions are stabilized by further conjugation, commonly through the presence of a carbonyl group. A prime example is the reaction described here, where a carbonyl can be seen directly next to the carbanion.



Scheme 5: *The mechanism of a Wittig reaction followed by a retro-Diels-Alder reaction*

The second reaction that occurs alongside the Wittig, is a retro-Diels-Alder reaction[174]. Where the compound formed in step 1 of the synthesis eliminates the 5-membered ring, regenerating furan as a by-product. It is unclear in which part of the synthesis the retro-Diels-Alder reaction occurs, as the part of the molecule that undergoes it is unaffected by the Wittig reaction. The reaction was carried out at RT initially, followed by a higher temperature of 50°C. It is likely that the increased temperature drove the retro-Diels-Alder reaction, so that it occurs after the Wittig[174,178]. Additionally, if it were not essential for the Wittig reaction to occur prior to the retro-Diels-Alder, then the standard Diels-Alder in the first step would not have been necessary, as the Wittig reagent could have directly been reacted with maleic anhydride. This all suggests that the latter is not feasible, therefore rendering necessary the Diels-Alder in the first step and the retro-Diels-Alder in the following step, this theory will be revisited further in this section.

The Wittig reaction of compound **4** with (carbethoxymethylene) triphenylphosphorane followed by a retro-Diels-Alder was successfully carried out following a procedure published by Costas-Lago, as seen in scheme **6**[179,180].

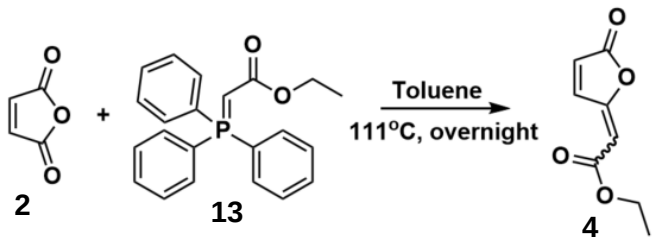


Scheme 6: *Wittig reaction between compound 3 and (carbethoxymethylene) triphenylphosphorane, followed by a retro-Diels-Alder reaction*

Compound **3** and the Wittig reagent were stirred in $CHCl_3$ at RT for 3 hours, followed to 50°C for another 3 hours. The reaction was monitored by TLC, showing the formation of two isomers, E and Z, as a result of the Wittig reaction, which were isolated using flash

chromatography. A significantly larger amount of E isomer, compound **5**, is produced, with a high yield of 89%, in comparison to the Z isomer, compound **5.1**, which is harder to isolate and gives a low yield of 5%. ^{1}H NMR can be used to identify which isomer has been isolated. Five proton environments are visible in both isomers, sharing three of the signals. The peaks corresponding to the proton found at the beta-position to the ring carbonyl (compound **5.1**, scheme **6**) and the proton located at the alpha-position to the carbonyl on the side chain (compound **5**, scheme **6**) can be used to confirm that the desired isomer has been isolated. According to literature, a doublet of doublets appears at 8.38 ppm corresponding to the vinylic proton located in the beta-position to the ring carbonyl in the E isomer, whereas this proton in the Z isomer appears as a doublet of doublets at 7.43 ppm[179]. Therefore, this double of doublets was used to distinguish between the two isomers.

Due to the Diels-Alder reaction performed in scheme **3**, being immediately reversed with a retro-Diels-Alder reaction in scheme **6**, as mentioned above. Reacting maleic anhydride with the Wittig reagent was briefly explored as a potentially simpler method[181].



Scheme 7: *Wittig reaction of maleic anhydride and (carbethoxymethylene) triphenylphosphorane*

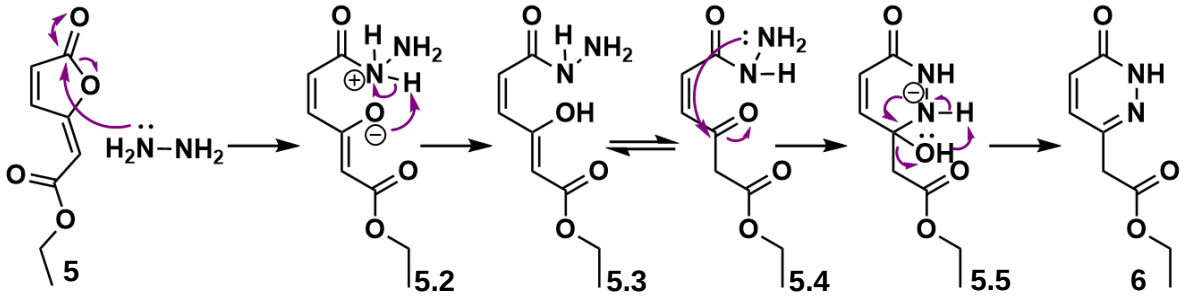
Maleic anhydride, (carbethoxymethylene) triphenylphosphorane were dissolved in toluene and refluxed overnight. TLC monitoring indicated the formation of the desired product, however when flash chromatography was used to purify it, a low of yield of 10% was reached.

Maleic anhydride is a soft nucleophile as it is a conjugated system. The presence of an α,β-unsaturated carbonyl in maleic anhydride, making it a Michael acceptor, leads to more possible side reactions with the Wittig reagent. As well as the Wittig reagent attacking the carbonyl, which is the desired mechanism, it can also perform a 1,4-addition, through a Michael reaction, or a 1,2-addition, attacking the other vinylic carbon. As the Wittig reagent is also a soft nucleophile, it will likely prefer to attack the vinylic carbon to the carbonyl, as the electrons of the latter are held more tightly due to the electronegativity of the oxygen. On the other hand, when reacting compound **4** with the Wittig reagent, the favoured location for it to attack is the carbonyl, which leads to the desired product and perhaps explains the higher yields reached through this reaction.

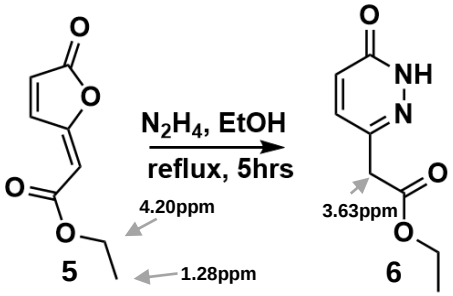
2.4 Hydrazine step

The next step of the synthesis was carried out using both isomers individually to observe any differences justifying the use of the E isomer in literature[171], however both isomers led to the same product. This concluded that the E isomer is chosen to move forward with due to the significantly higher yields reached and the fact that it is easier to purify and isolate.

In the following step, compound **5**, was reacted with hydrazine, which acts as a nucleophile, attacking the carbonyl of the α,β-unsaturated lactone, compound **5**, leading to its transformation to a 6-membered cyclic hydrazide, compound **6**[171].



Scheme 8: *The mechanism for the hydrazine reacting with compound **5***

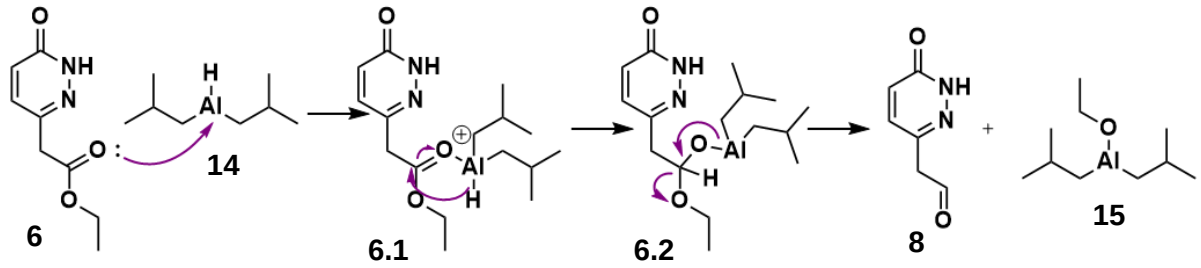


Scheme 9: *Reaction of α, β-unsaturated lactone with hydrazine to give 6-membered cyclic hydrazide*

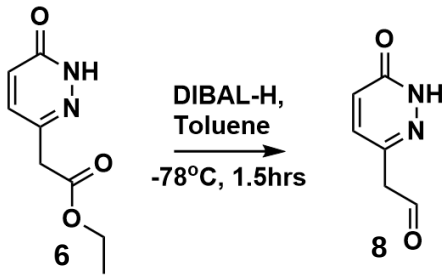
The procedure followed has been previously been described by Costas-Lago[179]. Compound **5**, hydrazine and ethanol were stirred under reflux, monitored by TLC until the reaction reached completion. Flash chromatography was used to purify the desired product, which was confirmed through ^{1}H NMR, with the data matching the literature[179]. The appearance of a singlet at 3.63ppm, which integrates to two protons, corresponds to the proton located on the carbon found between the ring and the ester, confirming the removal of the double bond and therefore the correct structure of compound **6**. A quartet at 4.20ppm and a triplet at 1.28ppm are seen, matching the protons of the ethyl group attached to the ester, -CH_2- and –CH_3 accordingly, confirming no other change to the side chain has been made.

2.5 Reduction of ester to alcohol

Initial synthetic plans required the formation of an aldehyde, and as such partial reduction of the ester was attempted using DIBAL-H, a strong reagent commonly used for the reduction of esters to aldehydes[182]. This reagent was chosen as it is known for partial reduction of ester when one equivalence is used and is expected not to affect the amide on the ring. DIBAL-H exists as a bridged dimer, becoming a reducing agent once it has formed a Lewis acid-base complex, therefore expected to reduce electron-rich carbonyls rapidly, following the mechanism described in scheme **10** below[182].



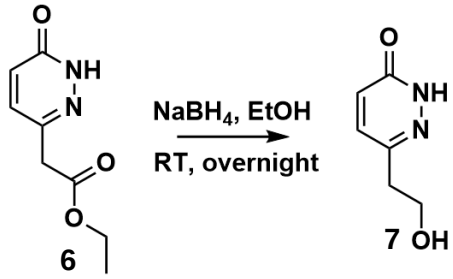
Scheme 10: *The mechanism of ester reduction to an aldehyde using DIBAL*



Scheme 11: *Reduction of ester to aldehyde with DIBAL*

The procedure for this was that of Silva[183]. DIBAL-H, compound **6** and THF were stirred under nitrogen, distilled water was later added and EtOAc was used to extract the product. ^{1}H NMR analysis of the product showed the same peaks as compound **6**, meaning the ester was not reduced and the rest of the molecule was left unaffected.

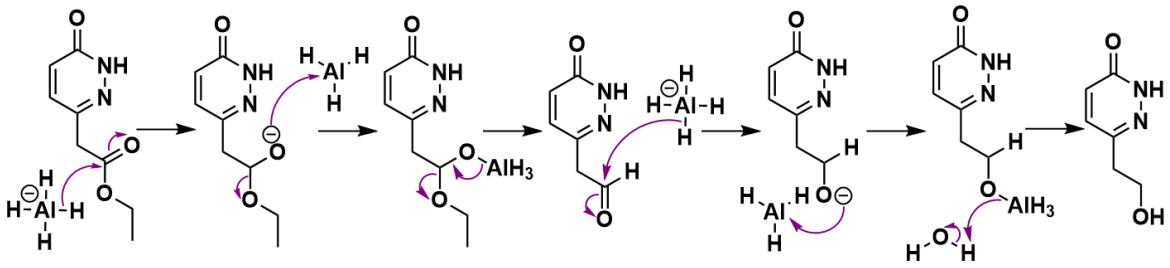
Furthermore, complete reduction of the ester to an alcohol was investigated using $NaBH_4$, a mild reducing agent, not commonly used to reduce esters. However literature suggests that carrying out the reaction in methanol or ethanol can lead to the partial reduction of an ester, and therefore was explored as an option, following a procedure described by Zhu[184].



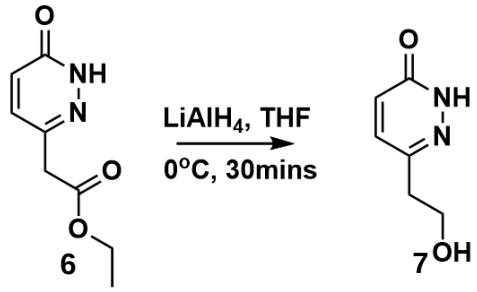
Scheme 12: *Reduction of ester to alcohol with $NaBH_4$*

Sodium borohydride, compound **6** and EtOH were stirred at RT for 2 hours, MeOH was added to the solution as compound **6** did not dissolve, leaving the reaction to stir for 17 hours. TLC was used to try and monitor the reaction, but nothing could be seen under UV. The reaction was acidified and extracted with ethyl acetate. The solid isolated was insoluble in deuterated chloroform, methanol, water and acetone, not allowing ^{1}H NMR analysis. Due to the difficulty in analysing the product, lithium aluminium hydride was tried as an alternative.

Both reducing agents serve as a hydride source due to the presence of a polar metal-hydrogen bond. What makes $LiAlH_4$ stronger than $NaBH_4$, is the fact that aluminium is less electronegative that boron, therefore the hydrides of the B-H bonds are less electron dense and cannot act as easily as nucleophiles as the hydrides of the Al-H bonds[185]. The mechanism by which $LiAlH_4$ reduces an ester, primarily to an aldehyde and finally to an alcohol is described in scheme **13**. Due to the latter being such a strong reagent the process cannot be interrupted once the ester has been reduced to an aldehyde, therefore always reaching the alcohol. Once the alcohol had been formed, it was thought that partial oxidation of the latter to an aldehyde would be carried out, continuing as planned with the rest of the synthesis.



Scheme 13: *Mechanism of ester reduction to an alcohol using $LiAlH_4$*



Scheme 14: *Reduction of compound **6** to compound **7** using $LiAlH_4$*

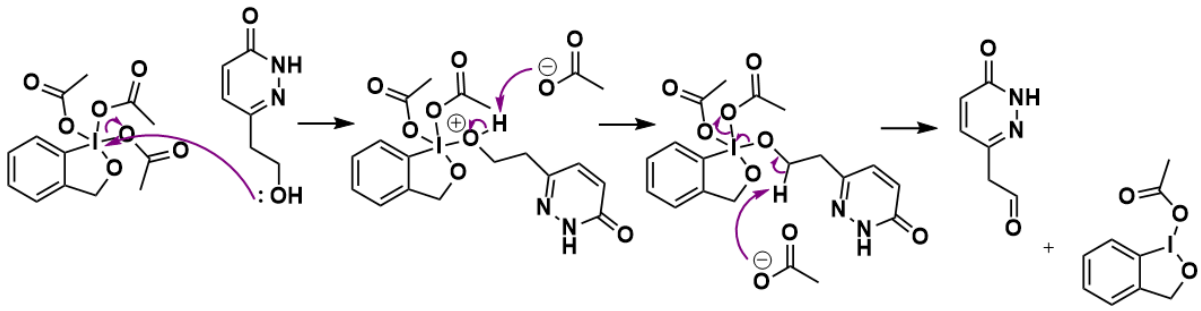
The procedure followed was that previously described by Goldberg[186]. Compound **6** was dissolved in THF and cooled to 0°C, the $LiAlH_4$ was added using a dropping funnel after diluting it with some THF, the reaction was stirred and quenched with cooled HCl. It was identified that the product does not fully dissolve in ethyl acetate, therefore during the extraction only part of the product was pulled from the aqueous layer. Due to the compound being insoluble in a number of solvents, including CH_2Cl_2 and THF, and being highly soluble in methanol, EtOAc was used as it is a highly polar solvent that is not miscible in water. Increasing the number of extractions necessary was found to be the most effective way of gaining more product. This was the method followed, even though the work up is time consuming it was found to be the best way to reduce the ester to an alcohol and purify it.

^{1}H NMR analysis was used to confirm the formation of the desired product. It was dissolved in deuterated methanol, and the spectrum showed all 4 expected proton environments with the correct integrations. The two doublets corresponding to the two aromatic protons were slightly shifted, compared to ^{1}H NMR of compound **6**. The doublet previously found at 7.33ppm in compound **6**, shifted to 7.46ppm and the doublet located originally at 6.96ppm only shifted to 6.92ppm, both integrating correctly to one proton. The main indication that the reaction was successful was associated with the rest of the peaks. From the ^{1}H NMR of compound **7**, the lack of the triplet at 1.28ppm and the quartet at 4.20ppm confirms the removal of that ethyl group found on the side chain, and the absence of the singlet found at 3.64ppm, suggesting further modifications near the other side of the ester. The presence of two new triplets located at 3.83ppm and 2.82ppm suggested the presence of two $–CH_2$- groups, with one of them near an electron withdrawing group, shielding the protons and shifting the peak to the right. All the changes seen in the ^{1}H NMR confirm the successful reduction of the ester to an alcohol. Although the yields were not optimal, consistently around 35%, this remained the most beneficial method to carry on the synthesis at this point.

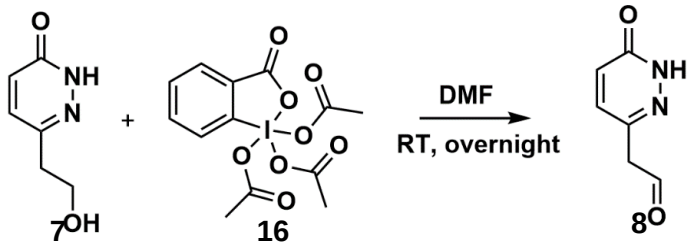
2.5 Oxidation of the alcohol to an aldehyde

The next step is the partial oxidation of the alcohol to the aldehyde. The main issue when working with compound **7** was its insolubility in several solvents, its high solubility in MeOH was not useful, as this was a problematic solvent for the next step as any oxidation technique would potentially also oxidise the methanol –OH group.

The first method used a reagent commonly used for partial oxidation of an alcohol to an aldehyde, Dess-Martin periodinane (DMP), a mild and selective oxidizing agent of primary alcohols to aldehydes expected to work through the mechanism described in scheme **15**[187].



Scheme 15: *The mechanism by which Dess-Martin periodinane oxidises an alcohol to an aldehyde*



Scheme 16: *Oxidation of alcohol **7** to aldehyde **8** using DMP*

The procedure followed was described by Boeckman[188]. Excess DMP and compound **7** were dissolved in DMF. The reaction is commonly performed in CH_2Cl_2 however compound **7** does not dissolve in this solvent. After the reaction was allowed to stir overnight, the purification process was extensive, including a celite plug, followed by flash chromatography and a silica plug, however no product was successfully isolated. This method was repeated on an NMR scale to monitor the reaction more efficiently, the reaction was left for 24 h and ^{1}H NMR readings were recorded at 1, 2, 4 and 24 hours. A peak around 10.62ppm appeared at 1 h, suggesting formation of an aldehyde, however the later ^{1}H NMRs recorded show the peak becoming smaller and eventually disappearing when recorded at 24 hours, showed in **Figure 23**. In the presence of the 10.62ppm peak, all peaks from the starting material can still be observed, however, when the aldehyde peak disappears, one of the triplets from compound **7**, marked with an arrow around 3.79 ppm

on **Figure 23**, also starts to fade. This either suggests that the potential aldehyde peak being formed is not the desired one, and is just part of a side reaction, or that compound **8** is forming at first but eventually degrades. It is also a possibility that the structure we are trying to form, compound **8**, is highly unstable, and breaks down quickly after it has been formed due to unfavourable conditions.

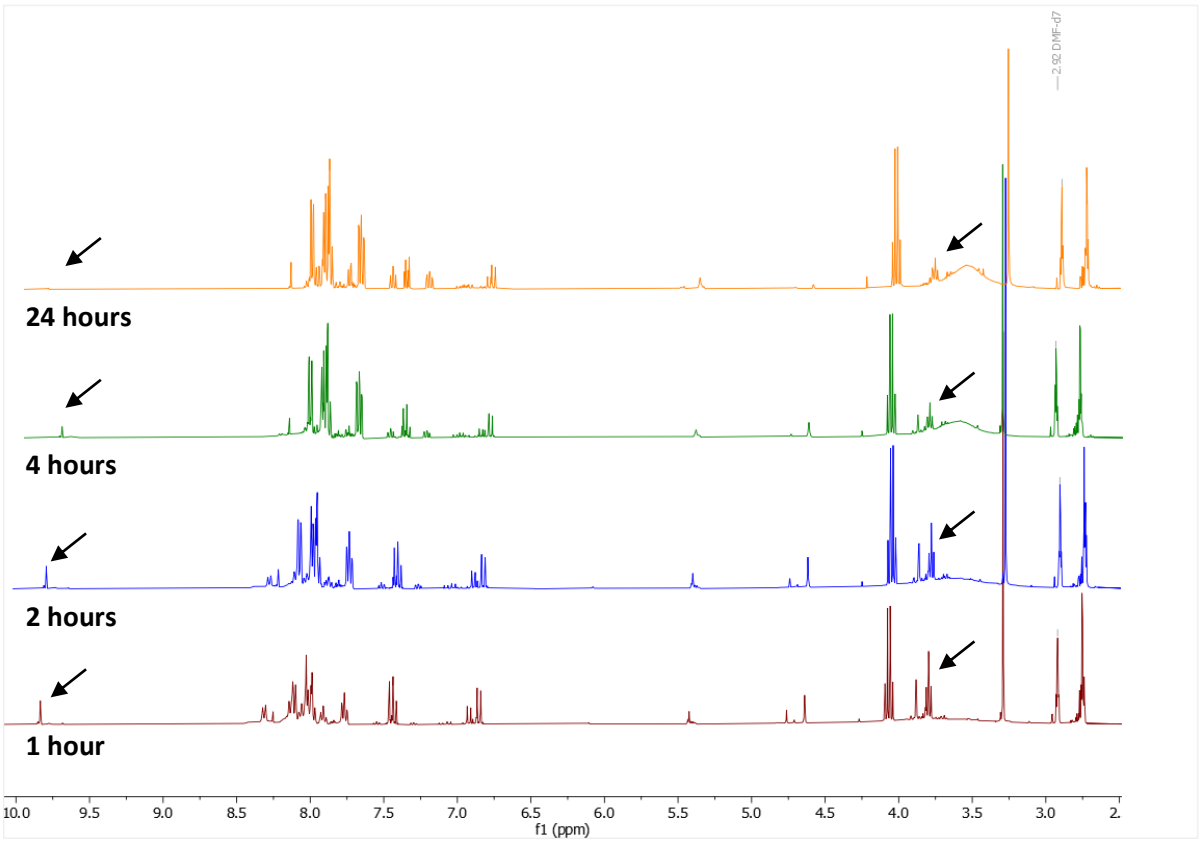
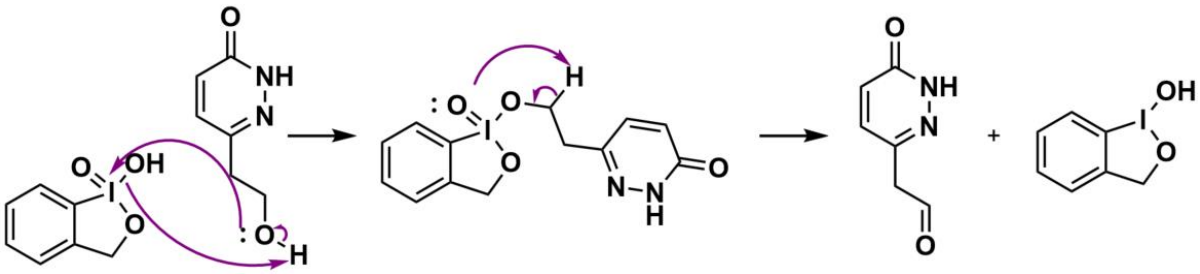


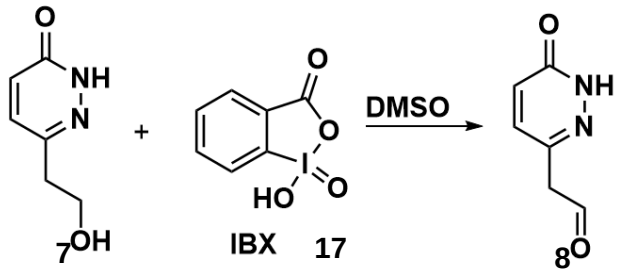
Figure 23: *Dess-Martin oxidation ^{1}H NMR time-lapse*

To eliminate the possibility of the desired aldehyde forming early on during the DMP reaction, but degrading as time passed, 2-iodoxybenzoic acid (IBX) was used an alternative, as it is a milder oxidizing agent[187].



Scheme 17: *The mechanism of alcohol oxidation to an aldehyde using IBX*

IBX is the intermediate created during the synthesis of the Dess-Martin periodinane, commonly used to oxidize primary alcohols to their corresponding aldehydes under mild conditions.



Scheme 18: *Oxidation of compound to compound using IBX*

Compound **7** and IBX were dissolved in dimethyl sulfoxide (DMSO), the reaction was run on an NMR scale to allow direct comparison with the DMP reaction, collecting ^{1}H NMR data at 1, 2, 4 and 24 hours. ^{1}H NMR analysis showed no change to the starting material through the progression of time, using the two triplets marked with an arrow in **Figure 24** to observe any changes to the starting material, as well as no aldehyde peak appearing at any measured point, suggesting IBX is not a strong enough oxidizing agent to have an effect on compound **7**. The ^{1}H NMR data collected at different points of time can be seen in **Figure 24**, highlighting the triplets that were used to monitor the reaction over time.

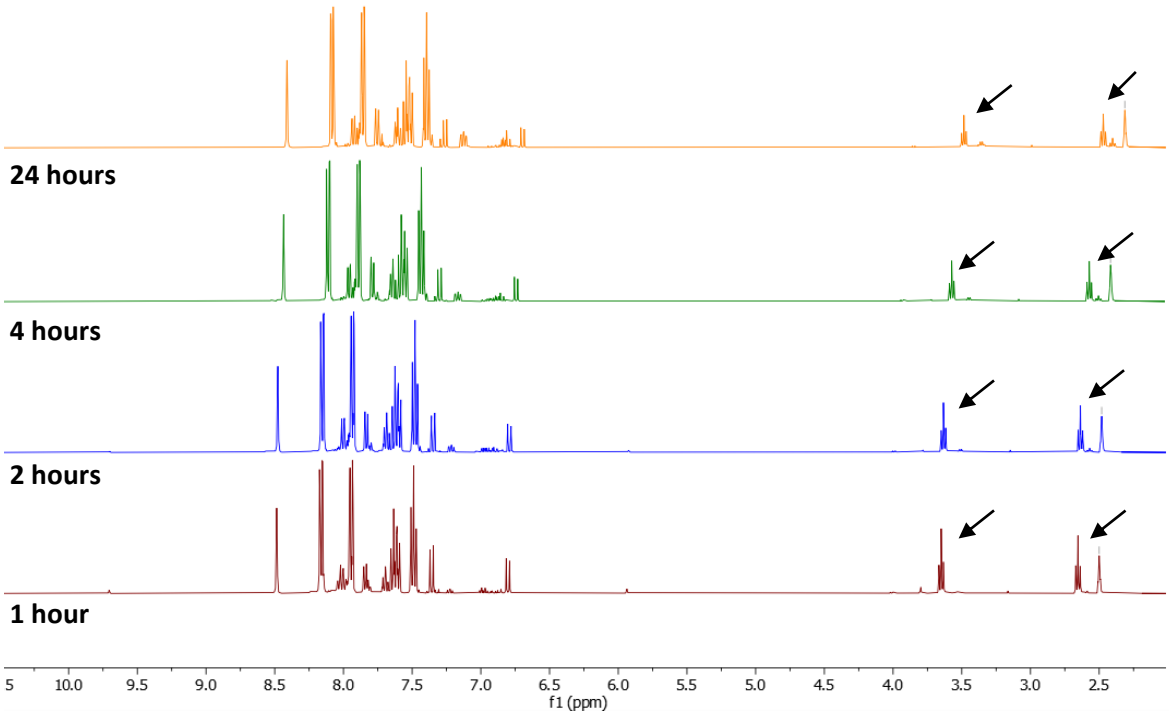
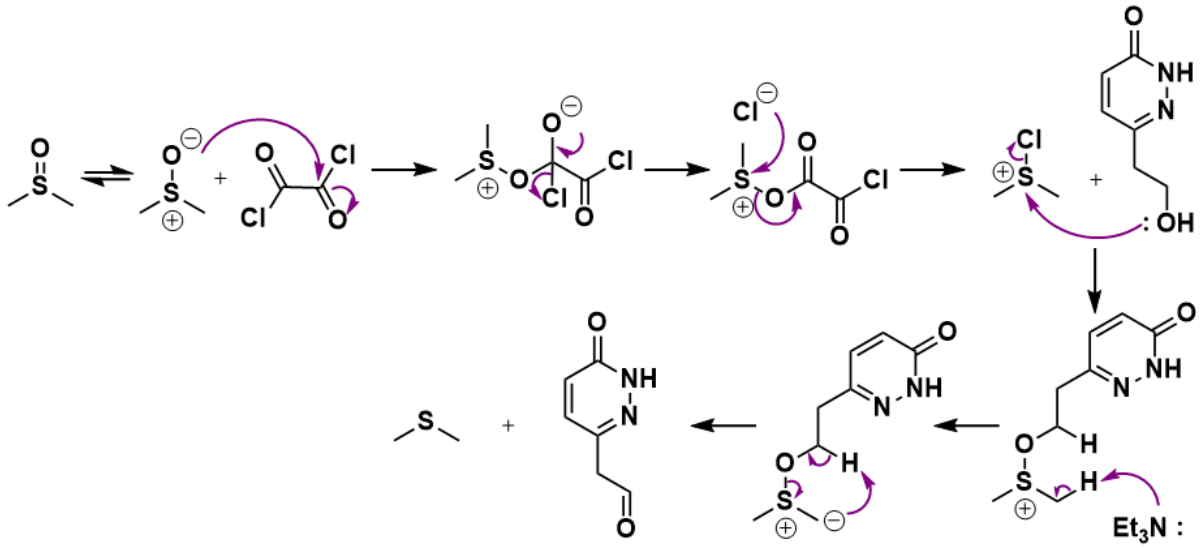
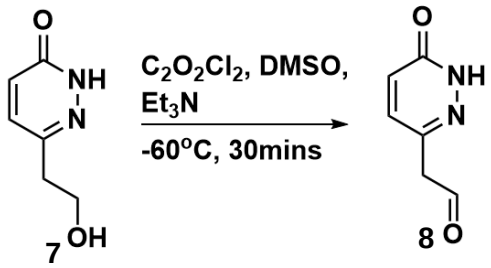


Figure 24: *IBX oxidation ^{1}H NMR time-lapse*

Another way to oxidise an alcohol to the aldehyde is a Swern oxidation, described in scheme **19** below[187]. In the first part of the mechanism, DMSO reacts with oxalyl chloride, producing an electrophilic sulphur compound. The newly negatively charged oxygen in turn reacts with oxalyl chloride, which in turn attacks the positively charged sulphur atom. As a result, a leaving group is expelled which breaks into CO_2, CO and a chloride ion. Up to this point the alcohol is not involved in the reaction. Once the chlorosulfonium ion has been formed, the alcohol attacks the sulphur forming a sulfonium salt, which is deprotonated by triethylamine forming an ylide. Finally, a proton transfer to the carbanion leads to the aldehyde being formed, with the DMSO reduced to a dimethylsulfide (DMS)[187].



Scheme 19: *The mechanism of a Swern oxidation*

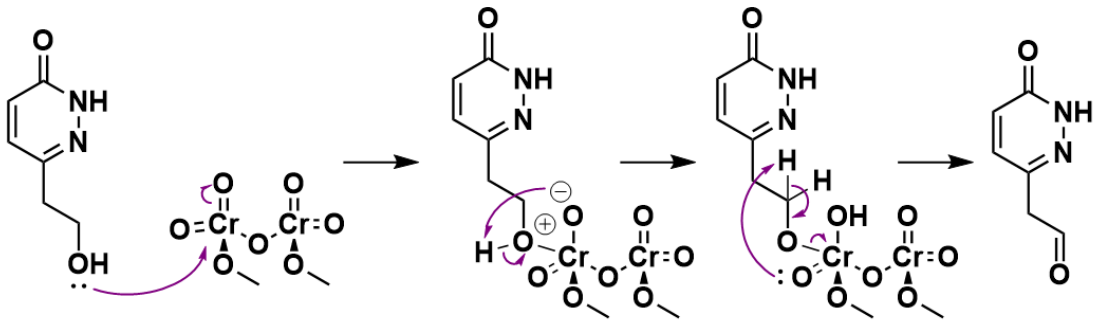


Scheme 20: *Oxidation of compound to compound by Swern oxidation*

The procedure followed for the Swern oxidation is that described by Souto[189]. Oxalyl chloride, anhydrous CH_2Cl_2 and DMSO were cooled to -61°C and treated with compound **7**, partially dissolved in THF, EtOAc and acetone and finally, treated with triethylamine. After quenching the reaction, CH_2Cl_2 was used for extraction followed by EtOAc, however no compound was found. $NaHCO_3$ was slowly added to the aqueous layer, making solution

more basic gradually, when pH of the solution reached 10, a small amount of starting material was identified through ^{1}H NMR, suggesting the reaction had not been successful as no aldehyde proton peaks could be seen around 10ppm.

The next method that was tried to complete the partial oxidation was the use of pyridinium dichromate. Pyridinium chlorochromate (PCC)[190,191] and pyridinium dichromate (PDC)[192] are reagents commonly used for the oxidation of alcohols. Due to difficulties associated with the removal of polymeric chromium byproducts from the solutions, these reagents have commonly been used in conjunction with celite or silica gel, both easily removed using filtration. Both compounds can be used to oxidize a primary alcohol to an aldehyde, however PDC is less reactive than PCC and therefore was preferred, due to uncertainty of how the amide on the ring will be affected by the reaction[193].



Scheme 21: *The mechanism by which PDC oxidises an alcohol to an aldehyde*

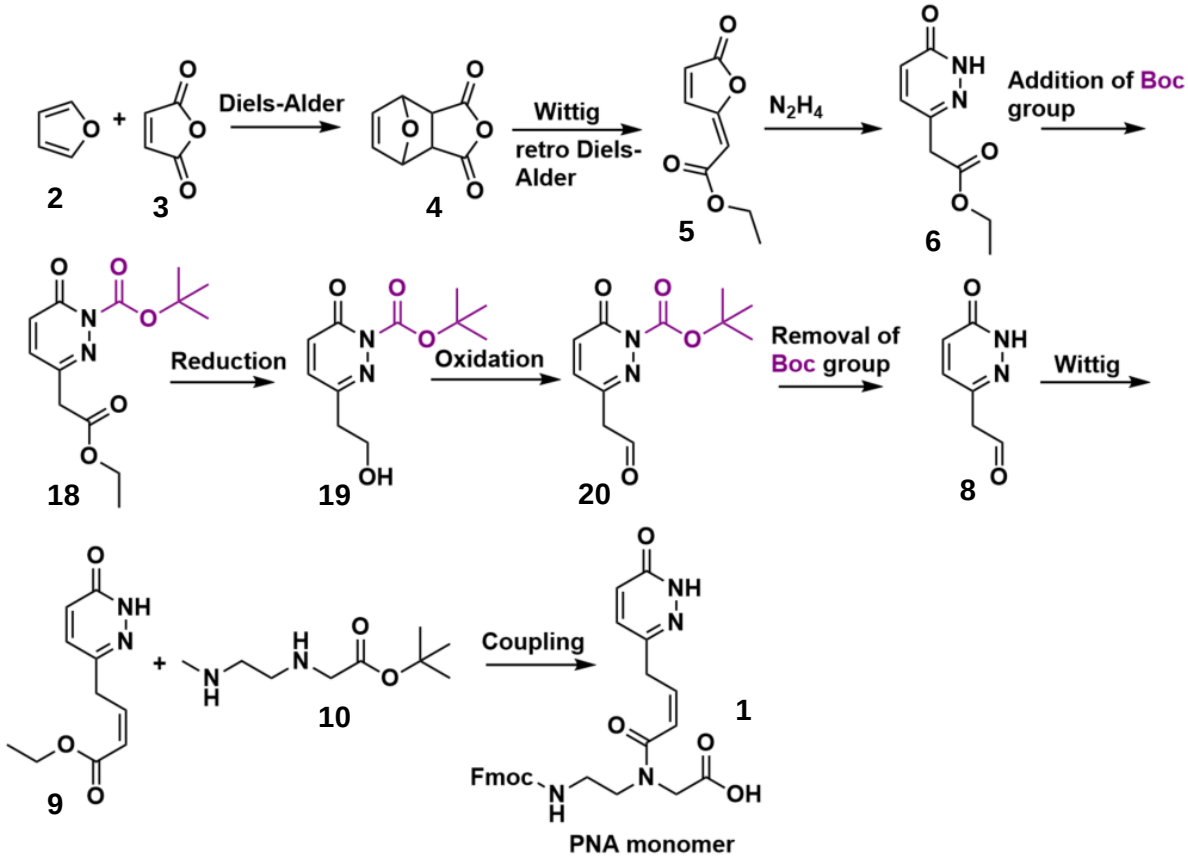
Scheme 22: *Oxidation of compound to compound with Pyridinium Dichromate*

The procedure followed was modified from that previously described by Luzzio, replacing PCC with PDC[194]. PDC, grounded onto silica gel, and compound **7** were dissolved in DMSO and allowed to stir. Filtration removed the chromium by-products, extraction with EtOAc gave no product, suggesting that the reaction was unsuccessful.

Chapter 3: Alternative approaches to complete the synthesis

3.1 Boc protecting group

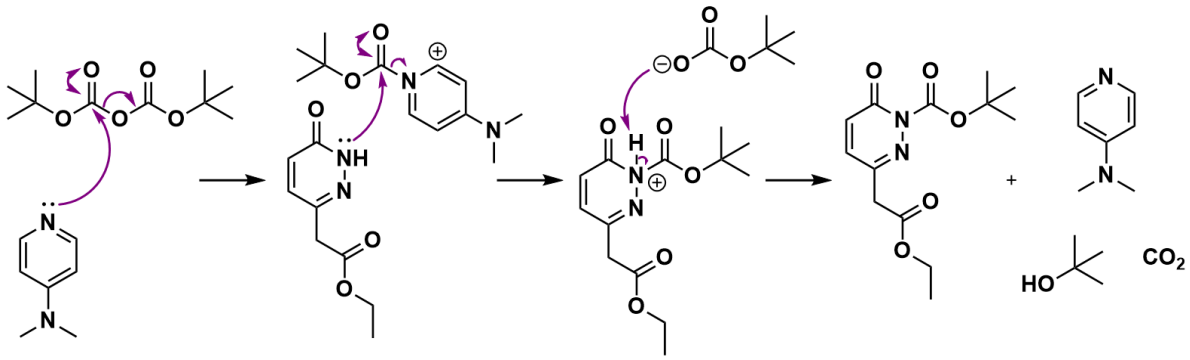
One of the main difficulties with the oxidation of the alcohol is its solubility, which led to difficulties in finding appropriate reaction conditions. The alcohol is soluble in polar solvents, suggesting it is a highly polar compound. In order to allow it to be soluble in less polar solvents the overall polarity of the compound needs to be lowered. A way to potentially solve this issue is the addition of a Boc group onto the amide nitrogen, which will reduce the polarity of the compound and potentially increase its solubility in more solvents[187,195]. The new synthetic pathway proposed with the addition of the Boc group is presented in scheme **23** below.



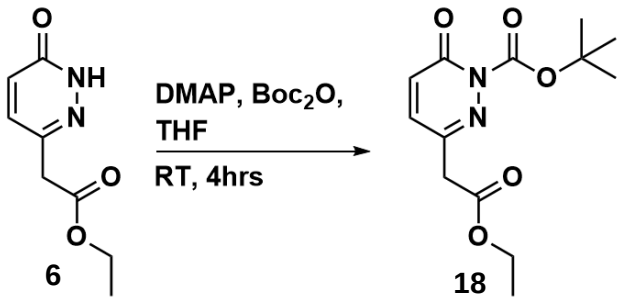
Scheme 23: *New synthetic approach including a Boc-protection*

The Boc group was chosen as a protecting group as it is commonly used to protect amines and is stable towards most nucleophiles and bases. The nitrogen being next to two carbonyl groups may make it less reactive and the addition of a large non-polar group will lower the polarity of the compound. If Boc anhydride is reacted with compound **7** it is very likely both the lone pair of the alcohol oxygen and the lone pair of the amide nitrogen will act as

nucleophiles and two Boc groups will be added onto the molecule, making it hard to selectively then remove the one from the alcohol. Therefore, Boc anhydride was reacted with compound **6**. The process expected to occur is presented in scheme **24** below.



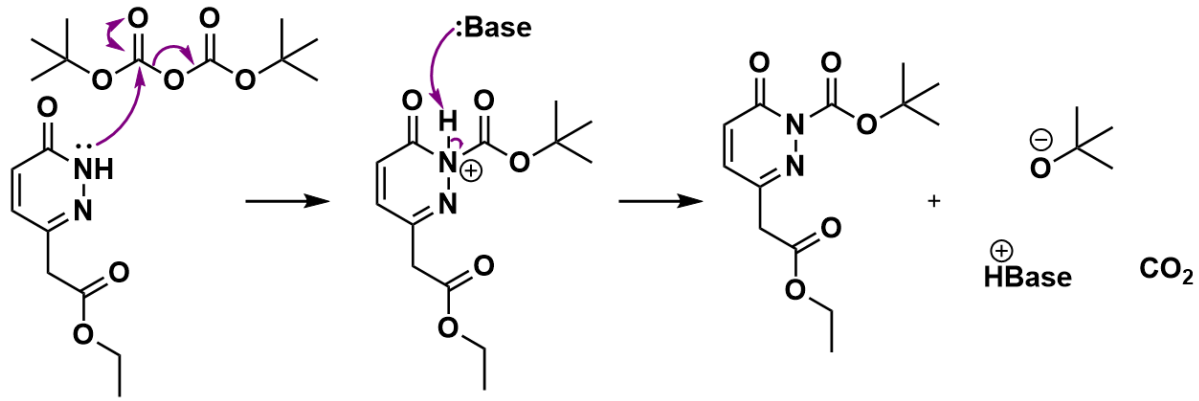
Scheme 24: *The mechanism by which the amide is Boc-protected using Boc anhydride and DMAP*



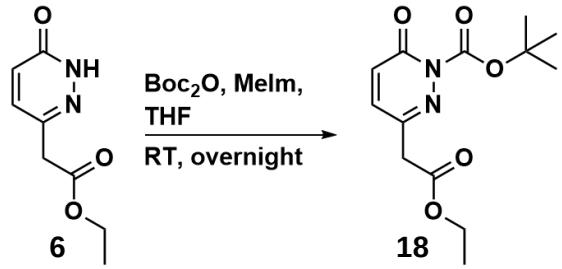
Scheme 25: *Boc protection of amide using Boc_2O and DMAP*

Compound **6** was used for the Boc protection, the procedure followed was that of Stephenson[196]. An excess of Boc anhydride (1:3), a catalytic amount of DMAP and compound **6** were dissolved in THF. A silica gel plug and flash chromatography were carried out to purify the product, removing some starting material that was identified, two compounds were isolated. ^{1}H NMR analysis was performed for both compounds, and the two compounds shared the same proton peaks, however the large peak at 1.50 ppm, corresponding to the protons found on the Boc group, integrated to 17.45 for the one compound, suggesting the potential binding of 2 Boc groups. Furthermore, in the second compound isolated, a similar peak appeared at 1.51 ppm integrating to 18, as well as a second peak at 1.59 ppm integrating to 9, suggesting the binding of 3 Boc groups, as all excess Boc had been removed. However, looking at compound **6** there are no two/three visible atoms that can act as nucleophiles to attach the Boc groups. This suggests a different mechanism taking place, rearranging the compound in a way it can bind to more than one Boc groups. However, no mechanism has been established yet, explaining the binding of multiple Boc groups.

A different method for attaching a Boc group was explored, where DMAP was not used, and Boc anhydride was added along with a base. According to literature, 1-methylimidazole and triethylamine are often used in this reaction, following the mechanism showed below.



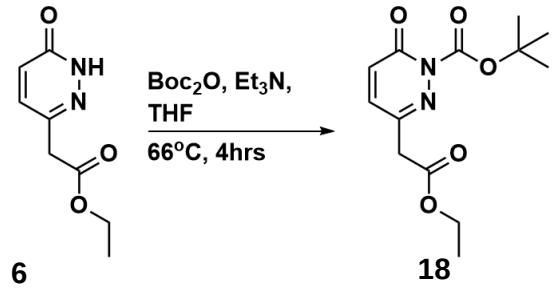
Scheme 26: *The mechanism by which an amide can be Boc protected using Boc anhydride and a base*



Scheme 27: *Boc protection of the amide using Boc anhydride and 1-methylimidazole*

The first base tried was 1-methylimidazole. Compound **6**, Boc anhydride, 1-methylimidazole were dissolved in THF and allowed to stir overnight, monitored by TLC. Once the reaction reached was complete, flash chromatography was used to isolate the product, however some impurities remained. Through ^{1}H NMR, the peaks seen when analysing compound **6** were still visible, suggesting the compound structure had not been altered, and a new peak at 1.63 ppm appeared, which integrated to 9, suggesting the addition of one Boc group. However, the presence of a number of smaller peaks remained after further purification using different conditions of flash chromatography. Due to the purification issues on a small scale, the purification process would be even more complicated when scaling up the reaction.

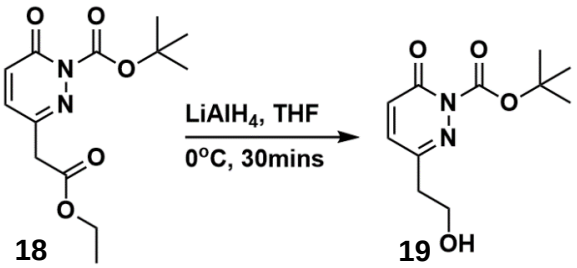
Simultaneously, a method using triethylamine as the base for the addition of the Boc group was followed, previously described by Cominetti[197].



Scheme 28: *Boc protection of amide using Boc anhydride and triethylamine*

Excess Boc, triethylamine (1:1) and compound **6** were dissolved in THF and stirred overnight at RT, TLC monitoring showed no change, indicating the reaction conditions were not favourable. The reaction was allowed to stir under reflux (66°C), where after 8 hours, TLC indicated the reaction had reached completion. ^{1}H NMR analysis showed the expected signals, maintaining the proton peaks of the ring protons, and the 3 proton peaks from the 3 proton environments of the side chain, with the addition of a large peak found at 1.59 ppm, integrating to 9, as well as a peak at 1.51 ppm integrating to 27, suggesting excess Boc was still in the reaction. This was expected due to the 1:3 equivalence of Boc anhydride added, the excess Boc was easily removed using flash chromatography, leading to the removal of the peak found at 1.51 ppm. As the purification of this reaction was significantly easier when using triethylamine, this base was used to carry out the rest of the synthesis.

The next step in the synthesis, was the reduction of the ester to the alcohol using the Boc-protected compound, the same procedure performed to reduce compound **6** was carried out to reduce compound **18**, using $LiAlH_4$ as the reducing agent, expected to behave as shown in scheme **13**.



Scheme 29: *Reduction of compound to compound using $LiAlH_4$*

The procedure[186] described above in scheme **14** was utilised and extraction led to the isolation of a small amount of product from the aqueous layer which was purified using flash chromatography. A mixture of compounds was isolated, ^{1}H NMR showed an impure mixture most likely including the desired product, compound **19**, however was found in a very small amount with a very low yield of 1.2%. Amongst several peaks due to impurities, two doublets at 7.23 ppm and 6.95 ppm were identified, believed to match the two protons found on the

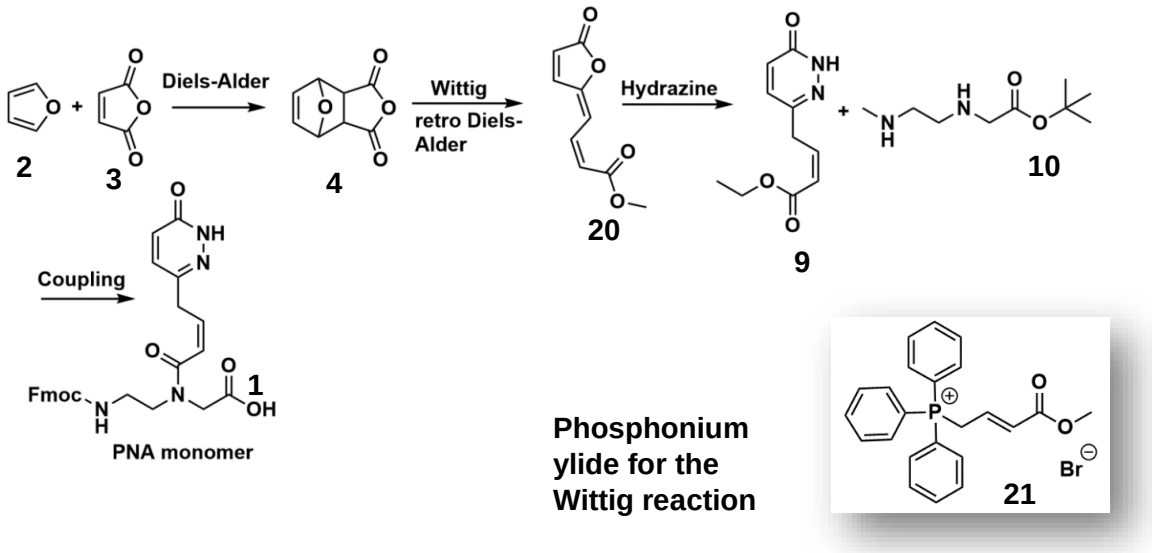
ring, two triplets at 3.97 ppm and 2.86 ppm were recognised, agreeing with the those seen in ^{1}H NMR of compound **7**. However, due to the very low yield, this method of reduction is not feasible when scaling up.

A second product was isolated from the flash chromatography, where ^{1}H NMR showed peaks matching those of compound **6**, missing the peak corresponding to the Boc group, suggesting the reaction of compound **18** with $LiAlH_4$ led to the removal of the Boc group. Since the ester was not reduced to the alcohol, the $LiAlH_4$ showed preference to the carbamate group, de-protecting it, rather than reducing the ester.

3.2 Modifying the Wittig reaction

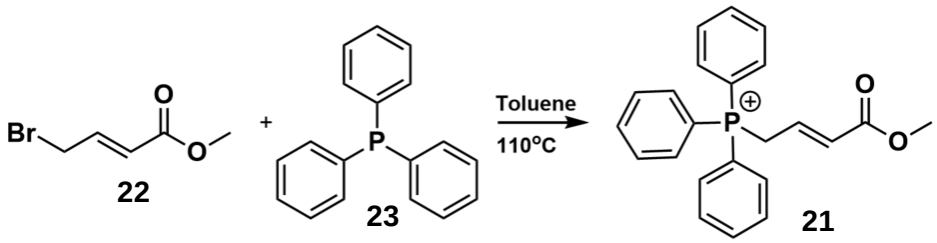
Due to the length of the synthesis, it was not feasible to continue the alternate method discussed in section **3.1**. The maximum yield by the reduction step was significantly low, and therefore an extremely large amount of starting material would be necessary to complete the pathway. As there is usually loss of product during every step, it was unlikely that the end of the pathway described in scheme **23** could be reached with a sufficient amount to be tested further. As a result, it was necessary to develop an alternative route to the PNA monomer. It was therefore suggested to go back to the initial steps of the synthesis, altering it from earlier on, specifically to the first Wittig reaction.

In the original synthesis proposed, two Wittig reaction were envisaged (scheme **1**), one in the second step and one before the coupling of the compound to the backbone. Due to the ease of the Wittig reaction in step 2 and the high yields achieved, it was proposed to combine the two steps into one Wittig reaction, by using an altered Wittig reagent with a longer side chain, specifically the sidechain seen attached to the ring in compound **9**. The new synthetic pathway is presented below.



Scheme 30: *New synthetic pathway designed through the modification of the Wittig reaction step*

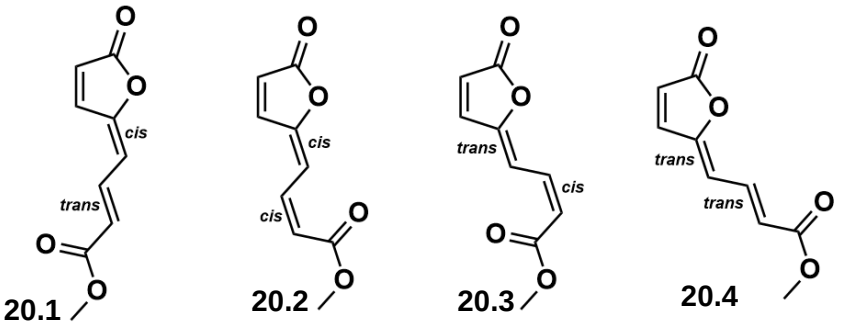
The formation of a Wittig salt is a common procedure, through the reaction of triphenylphosphine with an alkyl halide. In this case triphenylphosphine was reacted with 4-bromocrotonate[198].



Scheme 31: *The synthesis of a Wittig salt from 4-bromocrotonate and triphenylphosphine*

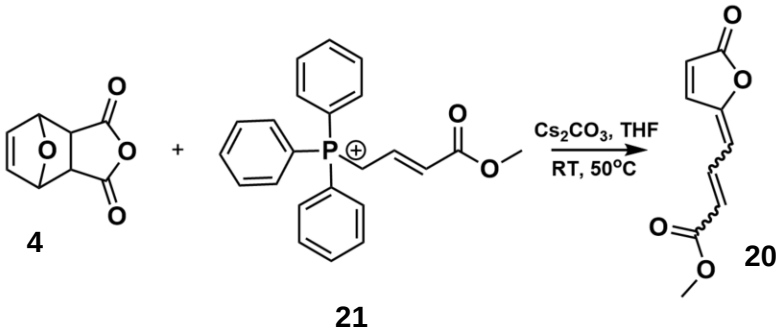
Triphenylphosphine and 4-bromocrotonate were dissolved in toluene and stirred at 110°C, forming a white precipitate which was washed with toluene. Once the precipitate was dried, it was confirmed to be the desired product through ^{1}H NMR, needing no further purification. The ^{1}H NMR analysis showed a doublet of doublets and a multiplet at 6.49 ppm and 6.73 ppm, respectively. Both signals integrate to one, corresponding to the two vinylic protons adjacent to the alkene. Additionally, a signal around 5.24 ppm was identified, specifically a doublet of doublets, integrating to two protons, corresponding to the two protons located on the carbon attached to the bromide. These signals, as well as the rest, match those in literature for this structure. Due to the benzene rings attached to the phosphorus, the 15 aromatic protons of the three rings give rise to a mixture of signals between 7.5 ppm and 8 ppm.

Next, the Wittig salt could be used to react with compound **4** to give compound **21**. In section **2.3**, where the Wittig reaction was described, it was mentioned that E and Z isomers of the product are formed from this reaction, depending on the conformation of the double bond, *cis* or *trans*, formed between the ring and side chain. The product formed from this new Wittig reagent has two double bonds, therefore there are two conformations for each double bond, leading to 4 possible combinations, *cis-cis*, *cis-trans, trans-cis* or *trans-trans*. The desired conformation of the product is *trans-cis*, with the first corresponding to the double bond connecting the side chain to the ring and the latter to the double bond closest to the ester group. However, it should be taken into consideration that the Wittig reaction product will be reacted with hydrazine, but as described in the reaction scheme **9,** it should not affect the conformation of the side chain.



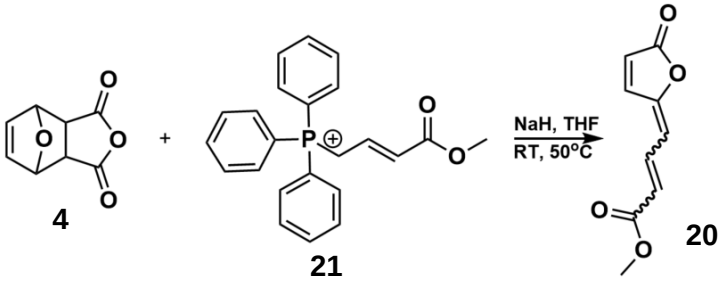
Scheme 32: *Four possible isomers of the product of the Wittig reaction*

A variety of bases were explored to carry out the Wittig reaction, which should deprotonate the phosphonium salt, therefore leaving the phosphonium ylide to react with compound **4**. The first base used was cesium carbonate (Cs_2CO_3), following a procedure by Ye[199].



Scheme 33: *Wittig reaction using cesium carbonate as a base*

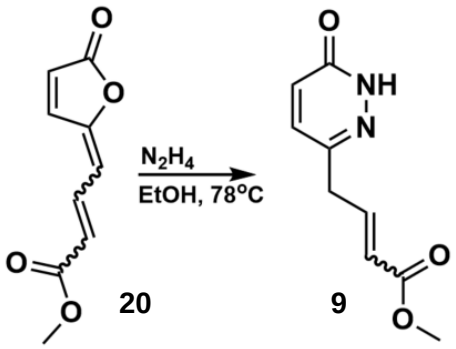
The Wittig salt was dissolved in THF, the Cs_2CO_3 was added dropwise, turning the solution bright yellow from the first drop added, allowing it to stir at RT. The solution turned orange and was then heated to 50°C, eventually turning the solution brown. TLC was used to monitor the reaction, a number of new spots were visible using the vanillin stain. Purification of the product was carried out using flash chromatography, however only side products were isolated and triphenylphosphine oxide. Crude ^{1}H NMR analysis showed indication that the desired product had formed, however the large amount of aromatic signals due to the triphenylphosphine oxide, potentially drowning out any smaller signals. The next base used, was sodium hydride, following a procedure previously published by Matsuo and Kende[200].



Scheme 34: *Wittig reaction using sodium hydride as the base*

Compound **4** was dissolved along with Wittig salt in THF and cooled to 0°C, prior to the addition of NaH due to its high reactivity. NaH was added and the solution was allowed to warm up to RT and then stirred. It was then heated to 50°C and, triphenylphosphine oxide precipitated. Purification of the product was followed by analysis using ^{1}H NMR. Eight doublets were seen around 7.5-8 ppm, likely corresponding to the aromatic protons of the ring in compound **20**. Due to the possibility of multiple isomers forming, the eight doublets are explained as two corresponding to one of the four isomers, the different stereochemistry of them leading to the same signals located in slightly altered positions in that region. However, only 6μg of product was gathered containing a mixture of all isomers, therefore it was not feasible to further purify it and try and separate the isomers.

Nevertheless, the small amount of mixed product was reacted with hydrazine, to investigate the effects on the isomers.



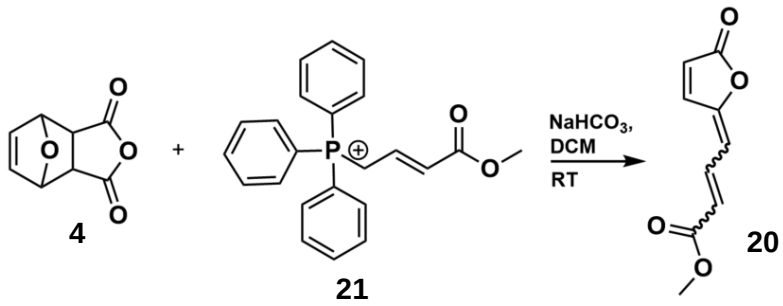
Scheme 35: *Reaction of compound with hydrazine*

The procedure followed was that described previously in the hydrazine step, shown in scheme **9**[179]. ^{1}H NMR analysis of the crude product gave some preliminary indications around the success of the step, although, due to the small amount of 6μg starting material used, peaks that were difficult to identify. In the aromatic region only two doublets could be seen, potentially suggesting that when the different isomers reacted with the hydrazine they all undertook the same conformation. This effect was also seen in the previous hydrazine reaction, however that was due to the lack of double bond on the sidechain after the reaction. It is likely that in this the case two isomers were produced, differing in conformation of the bond highlighted in the product of scheme **36**, but only one can be identified from the ^{1}H NMR due to selectivity of that conformation, rendering the signals of the other isomer too low to be seen in such a small amount.

The Wittig reaction was repeated on a larger scale, where in none of the compounds isolated through purification could any isomer aromatic peaks be identified. This suggests that either some other side reaction occurs when performing it on a larger scale preventing

the formation of the desired product, or the purification method required adjustments. Impurities, including triphenyl phosphine oxide, made identification of products difficult.

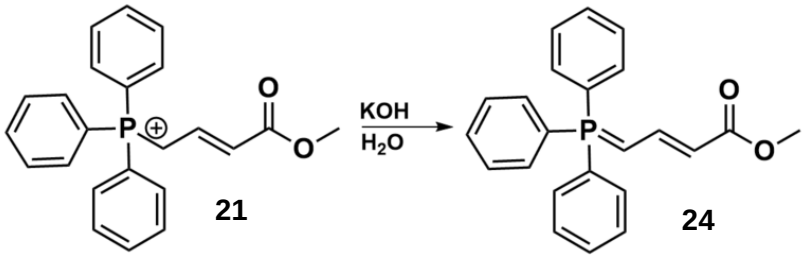
Use of weaker base was then investigated, sodium bicarbonate ($NaHCO_3$). The procedure followed was that of Frei[201].



Scheme 36: *Wittig reaction of compound with Wittig salt and saturated $NaHCO_3$ as the base*

The Wittig salt was dissolved in CH_2Cl_2, the saturated $NaHCO_3$ was added leaving the solution to stir. Compound **4** was added and left to stir for 2 days at RT. The product gathered through extraction and purification had the same peaks as the Wittig salt, suggesting the reaction had not worked, leaving the salt unaffected.

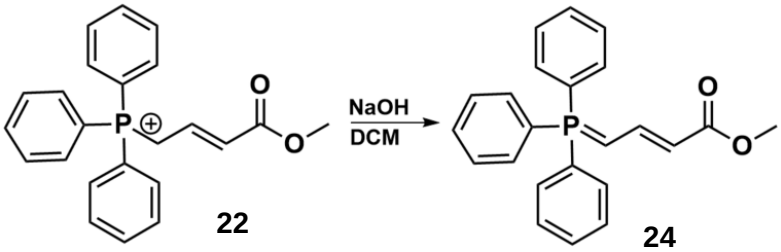
Due to the difficulties arising with the Wittig salt and matching it to the right base, the idea of isolating the intermediate phosphonate was explored. Literature suggested the double bond could be formed through the use of KOH[198].



Scheme 37: *Reaction of Wittig salt with KOH to give the relative phosphorane*

The Wittig salt was dissolved in distilled water and 1M KOH was added until pH 9 was reached, turning the solution bright yellow. A yellow precipitate formed which was filtered out and ^1H NMR data was collected. A mixture of signals could be seen however the doublet of doublets at 5.24 ppm, corresponding to the protons attached to the carbon bound to the phosphorus atom, still integrated to 2, meaning the double bond had not formed.

Literature reports suggested that NaOH has also been used for the reaction above for the transformation of the Wittig salt to a phosphorane Wittig agent[202].



Scheme 38: *Transformation of Wittig salt to the relative phosphorane using NaOH*

The salt was dissolved in CH_2Cl_2 and NaOH was added, immediately turning the solution yellow. After stirring briefly, the solution had turned orange and was diluted with water and extracted with CH_2Cl_2. However, ^{1}H NMR showed that the reaction had not worked, once again the signal at 5.24 ppm integrating to two, confirming the double bond formation had not occurred.

Conclusion

The main aim of this project was the synthesis of a modified PNA monomer based on a theoretical structure published by Topham et al. The molecule is made up of a 2-aminoethylglycine backbone, a 2-*cis* olefin linker connecting the backbone to a modified nucleobase, known as E base. The synthesis of the PNA monomer was to be followed by the generation of a PNA sequence made up of repeated units of the monomer, in order to test its binding affinity with dsDNA. The synthesis of the E-base and the linker were to be achieved by a preparative route partially based on a published article by Tomori. The final compound would then be coupled to the glycine backbone generating the PNA monomer. A large part of the synthesis was successfully completed, reaching moderate to high yields for all synthesized compounds. Certain complications with the route led to the exploration of various different paths explored in order to bypass synthetic issues. These were partially completed successfully, however hold great promise for follow-up research.

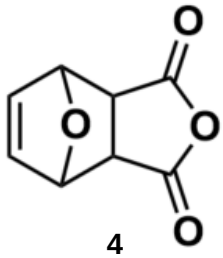
General procedures

The reagents used were used from the following named suppliers: Sigma Aldrich, Acros Organics, Fluorochem, Thermo Fisher. All reactions requiring anhydrous conditions were performed under an atmosphere of nitrogen. Water used in all reactions was distilled. All ^{1}H and ^{13}C NMR spectra were recorded in Fourier transform mode operating at a ^{1}H NMR frequency of 400 MHz and a ^{13}C NMR frequency of 100 MHz, using the specified deuterated solvent. ^{1}H- and ^{13}C- NMR were collected using a Bruker spectrometer operating at 400 MHz (^{1}H) or100 MHz (^{13}C) using the specified deuterated solvent. Chemical shifts were recorded in ppm and were referenced to residual solvent peak of MeOH (3.31 ppm) or chloroform (7.26 ppm). Multiplicities in the NMR spectra are described as s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, and combinations thereof. Coupling constants are reported in Hz. Infrared spectra were obtained on a Perkin Elmer Spectrum Two. Organic solutions which had been in contact with water were dried over sodium sulphate unless otherwise specified. Thin-layer chromatography was performed on aluminium plates coated with 0.20 mm silica gel 60 with florescent indicator UV_{254}. After elution, the TLC plates were visualized under UV light, unless staining prior to visualisation was specified. Flash chromatographic separations were performed on silica gel for column chromatography. Detection wavelength was 254 nm and 280 nm. All figures and schemes containing structures were created on ChemDraw, whilst the rest of the schematics were drawn freehand on a Samsung Galaxy Tab S4.

Experimental

Synthesis of 3a,4,7,7a-tetrahydro-4,7-epoxyisobenzofuran-1,3-dione (4)

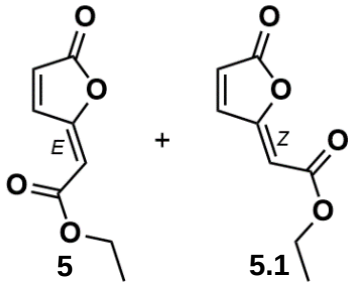
Compound **4** was prepared following the literature procedure (reference). Maleic anhydride (5.19 g, 53.02 mmol) and furan (4.94 mL, 67.4 mmol) were dissolved in CH_2Cl_2 (20 mL) under argon atmosphere and stirred at room temperature for 24 h. A white precipitate formed which was collected by filtration and washed twice with diethyl ether (58%, 4.98g, 0.030mol). Obtained data matched that previously reported[175]:



4

Rf = 0.4 (50:50 EtOAc/Hexane), m.p. = 117-118^{o}C, ^{1}H NMR ($CDCl_3$, 400 MHz) δ 6.58 (t, J=0.92 Hz, 2H, HC=), 5.46 (t, J=0.96 Hz, 2H, HCO), 3.17 (s, 2H, CHCO). ^{13}C NMR ($CDCl_3$, 100 MHz) δ 170.0 (2H, CO), 137.1 (2H, HC=), 82.3 (2H, HCO), 48.8 (2H, CHCO). IR (cm^{-1}) 1857.6, 1780.4, 1019.7, 878.8

Synthesis of ethyl (2E)-(5-oxo-2(5H)-furanylidene) acetate (5)



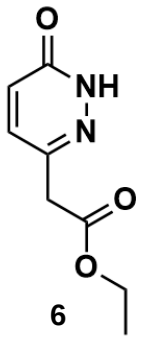
5 **5.1**

Compound **4** (2.99 g, 18.05 mmol) and (carbethoxymethylene) triphenylphosphorane (6.91 g, 19.8 mmol) were chloroform (26 mL), the reaction was stirred for 3 hours at room temperature and 3 h at 50^{o}C. The reaction mixture was concentrated in vacuo to 10 ml and dried onto silica gel. Purification of the residue by silica gel column chromatography gave E isomer at 10% EtOAc/hexane (85%, 2.55g, 0.015mol) and Z isomer at 50% EtOAc/hexane (5%, 0.134g, 0.0008mol). Obtained data matched that previously reported[179]:

E isomer: Rf = 0.6 (50:50 EtOAc/Hexane), m.p. = 78-80^{o}C, ^{1}H NMR ($CDCl_3$, 400 MHz) δ 8.38 (dd, J=5.68, 0.76 Hz, 1H, H3), 6.46 (dd, J=5.68, 1.8 Hz, 1H, H4), 5.93 (dd, J=1.8, 0.76 Hz, 1H, C*H*CO_2), 4.26 (q, J=7.12 Hz, 2H, CH_2), 1.33 (t, J=7.28 Hz, 3H, CH_3). ^{13}C NMR ($CDCl_3$, 100 MHz) 167.9, 165.0, 160.4, 142.1 (C3), 124.5 (C4), 102.8 (C*H*CO_2), 61.3 (CH_2), 14.3 (CH_3). IR (cm^{-1}) 3157.1, 3119.6, 2990.2, 1786.8, 1722.6, 1644.7, 1063.3, 837.9

Z isomer: Rf = 0.5 (50:50 EtOAc/Hexane), m.p. = 113-115^{o}C, ^{1}H NMR ($CDCl_3$, 400 MHz) δ 7.43 (d, J=5.52 Hz,, 1H, H3), 6.45 (dd, J=5.4, 0.72 Hz, 1H, H4), 5.46 (s, 1H, C*H*CO_2), 4.28 (q, J=7.24 Hz, 2H, CH_2), 1.34 (t, J=7.04 Hz, 3H, CH_3). ^{13}C NMR ($CDCl_3$, 100 MHz) 168.2, 163.2, 156.3, 144.8 (C3), 123.9 (C4), 102.2 (C*H*CO_2), 61.5 (CH_2), 14.3 (CH_3). IR (cm^{-1}) 3093.9, 2992.3, 1785.3, 1703.3, 1665.1, 1155.8, 857.1

Synthesis of ethyl 2-(6-oxo-1,6-dihydropyridazin-3-yl) acetate (6)

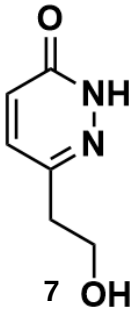


6

To a solution of compound **5** (1.88 g, 11.1 mmol) in 8 ml ethanol, hydrazine monohydrate (765 µL, 15.2 mmol) was added and the reaction was stirred under reflux (78^{o}C) for 4 h. The solution was concentrated in vacuo and dried onto silica gel. The residue was purified by silica gel column chromatography (1% MeOH/EtOAc) to give a yellow solid (78%, 1.45g, 0.008mol). Obtained data matched that previously reported[179]:

Rf = 0.4 (5% MeOH/EtOAc), m.p. = 106-108^{o}C, ^{1}H NMR ($CDCl_3$, 400 MHz) δ 11.70 (s, 1H, NH), 7.32 (d, J=9.72 Hz, 1H, H4), 6.95 (d, J=9.68 Hz, 1H, H5), 4.20 (q, J=7.36 Hz, 2H, C*H*$_2$$CH_3$), 3.64 (s, 1H, -$CH_2$-), 1.28 (t, J=6.88 Hz, 3H, CH_2C*H*$_3$). ^{13}C NMR ($CDCl_3$, 100 MHz) 169.6 (*C*O_2Et), 161.8 (C6), 142.5 (C3), 134.6 (C4), 130.1 (C5), 61.6 (*C*H_2CH_3), 40.1 (CH_2), 14.2 (CH_3). IR (cm^{-1}) 3045.6, 2989.7, 1722.3, 1655.5, 1594.3, 1171.2, 810.7

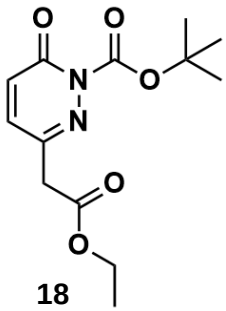
Synthesis of 6-(2-hydroxyethyl)-3(2H)-pyridazinone (7)



Compound **6** (0.200 g, 1.097 mmol) was cooled to 0°C and dry THF (20 mL) were added under nitrogen and $LiAlH_4$ (1M in THF, 6.4 mL, 0.152mol) was added dropwise and solution was stirred for 30mins at 0°C under nitrogen. The solution was quenched with 1M HCl which was added dropwise to reach pH 2 and was then extracted with EtOAc (20 mL x20), dried with Na_2SO_4 and concentrated in *vacuo*. Purification of the residue by silica gel column chromatography (100% EtOAc then 1-10% MeOH/EtOAc) gave a beige solid (40%, 0.07g, 0.50mmol).

Rf = 0.1 (5% MeOH/EtOAc), m.p. = 148-150 °C, ^{1}H NMR (MeOD, 400 MHz) δ 7.45 (d, J=8.92 Hz, 1H, H3), 6.91 (d, J=9.32 Hz, 1H, H4), 3.84 (t, J=5.84 Hz, 2H, C*H*$_2$CH$_2$OH), 2.81 (t, J=6.04 Hz, 2H, C*H*$_2$OH). ^{13}C NMR (MeOD, 100 MHz) 163.5, 149.0, 136.5, 130.2, 61.2, 38.3. IR (cm^{-1}) 3153.4, 2918.3, 1669.8, 1651.5, 1598.9, 1050.7, 1006.1, 841.7

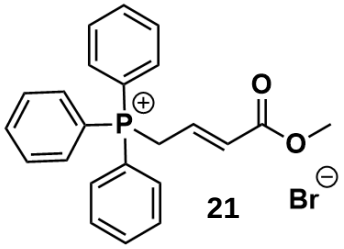
Synthesis of Boc-protected ethyl 2-(6-oxo-1,6-dihydropyridazin-3-yl) acetate (18)



To a solution of compound **6** (0.100 g, 0.548 mmol) in 5ml THF, Et_3N (0.540 g, 0.540 mmol) and Boc_2O (0.351 mg, 1.60 mmol) were added and stirred under reflux (66°C) overnight. The solution was concentration in *vacuo* and dried onto silica gel. The residue was purified by flash chromatography (20-50% EtOAc/Hexane) to remove excess Boc, giving a white solid (81%, 0.125g 0.44mmol).

Rf = 0.7 (5% MeOH/EtOAc), m.p. = 130°C. ^{1}H NMR ($CDCl_3$, 400 MHz) δ 7.23 (d, J=9.6 Hz, 1H), 6.89 (d, J=9.72 Hz, 1H), 4.17 (q, J=7.2 Hz, 2H, C*H*$_2$CH$_3$), 3.65 (s, 2H, CH$_2$CO$_2$), 1.60 (s, 9H, 3xCH$_3$), 1.25 (t, J=7.16 Hz, 3H, CH$_2$CH$_3$). ^{13}C NMR ($CDCl_3$, 100 MHz) 169.2, 158.1, 150.3, 141.9, 133.9, 131.8, 86.5, 61.6(*C*H$_2$CH$_3$), 40.1 (-CH$_2$-) 27.7 (3xCH$_3$), 14.1 (CH$_3$). IR (cm^{-1}) 3045.9, 2989.1, 2935.8, 1722.2, 1655.5, 1594.7, 1170.3, 1035.1, 855.2, 810.5

Synthesis of phosphonium [(2E)-4-methoxy-4-oxo-2-buten-1-yl] triphenyl- bromide (21)



Methyl 4-bromocrotonate (0.300g, 1.67 mmol) and triphenylphosphine (1:1.1, 0.474 g, 1.80 mmol) were dissolved in 15 ml toluene and stirred at 110°C for 3 h. The toluene was pipetted out and the precipitate was washed

with toluene (20 ml, x3). The solution was concentrated in *vacuo* and gave a pure white solid (93%, 0.685 g, 1.55 mmol).

Rf = 0.55 (5% MeOH/EtOAc), m.p. = 120°C ^{1}H NMR ($CDCl_3$, 400 MHz) δ 7.65-7.93 (m, 15H), 6.72 (m, 1H), 6.46 (dd, J = 15.4 Hz, 5.0 Hz, 1H), 5.23 (dd, J = 16.4 Hz, 7.6 Hz, 2H), 3.65 (s, 3H). ^{13}C ($CDCl_3$, 100MHz) 135.1, 134.2, 130.2, 77.5, 77.1, 76.7. IR (cm^{-1}) 1713.0, 1434.4, 1327.5, 1110.5, 980.9, 721.6

References

1 H. Fletcher and I. Hickey, *Genetics*, Fourth., 2012.

2 J. D. Watson and F. H. C. Crick, *Nat. 1953 1714356*, 1953, **171**, 737–738.

3 R. E. Franklin, *Nature*, 1950, **165**, 71–72.

4 M. H. F. Wilkins, A. R. Stokes and H. R. Wilson, *Nat. 1953 1714356*, 1953, **171**, 738–740.

5 J. D. Watson and F. H. C. Crick, *Genetical implications of the structure of deoxyribonucleic acid*, 1953.

6 M. Meselson and F. W. Stahl, *Proc. Natl. Acad. Sci.*, 1958, **44**, 671–682.

7 R. Okazaki, T. Okazaki, K. Sakabe, K. Sugimoto and A. Sugino, *Proc. Natl. Acad. Sci. U. S. A.*, 1968, **59**, 598.

8 N. F. Lue and R. D. Kornberg, *Proc. Natl. Acad. Sci. U. S. A.*, 1987, **84**, 8839–8843.

9 H. Boeger, J. Griesenbeck, J. S. Strattan and R. D. Kornberg, *Mol. Cell*, 2003, **11**, 1587–1598.

10 P. Cramer, D. A. Bushnell and R. D. Kornberg, *Science (80-.).*, 2001, **292**, 1863–1876.

11 R. J. Kelleher, P. M. Flanagan and R. D. Kornberg, *Cell*, 1990, **61**, 1209–1215.

12 F. Gros, H. Hiatt, W. Gilbert, G. Kurland, R. W. Risebrough and J. D. Watson, *Unstable ribonucleic acid revealed by pulse labelling of Escherichia Coli*, Academic Press, 1961, vol. 1.

13 S. Brenner, F. Jacob and M. Meselson, *An unstable intermediate carrying information from genes to ribosomes for protein synthesis*, Harwell, 1957, vol. 186.

14 F. H. C. Crick, L. Barnett, S. Brenner and R. J. Watts-Tobin, *Nature*, 1961, **192**, 1227–1232.

15 G. E. Palade, *J. Biophys. Biochem. Cytol.*, 1955, **1**, 59–67.

16 M. B. Hoagland, M. L. Stephenson, J. F. Scott, L. I. Hecht and P. C. Zamecnik, *J. Biol. Chem.*, , DOI:10.1016/S0021-9258(19)77302-5.

17 R. E. Franklin and R. G. Gosling, *Nature*, 1953, **172**, 156–157.

18 A. H. J. Wang, G. J. Quigley, F. J. Kolpak, J. L. Crawford, J. H. Van Boom, G. Van Der Marel and A. Rich, *Nature*, 1979, **282**, 680–686.

19 G. Wang and K. M. Vasquez, *Front. Biosci.*, 2007, **12**, 4424–4438.

20 R. E. Franklin and R. G. Gosling, *Acta Crystallogr.*, 1953, **6**, 673–677.

21 R. E. Dickerson, H. R. Drew, B. N. Conner, R. M. Wing, A. V. Fratini and M. L. Kopka, *Science*, 1982, **216**, 475–485.

22 R. E. Dickerson, *Methods Enzymol.*, 1992, **211**, 67–111.

23 G. Wang and K. M. Vasquez, *Front. Biosci.*, 2007, **12**, 4424–4438.

24 J. M. Vargason, K. Henderson and P. S. Ho, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 7265.

25 H. E. Moser and P. B. Dervan, *Science (80-.).*, 1987, **238**, 645–650.

26 A. Klinakis, D. Karagiannis and T. Rampias, *Cell. Mol. Life Sci. 2019 774*, 2019, **77**, 677–703.

27 T. M. Geel, M. H. J. Ruiters, R. H. Cool, L. Halby, D. C. Voshart, L. Andrade Ruiz, K. E. Niezen-Koning, P. B. Arimondo and M. G. Rots, *Philos. Trans. R. Soc. B Biol. Sci.*, , DOI:10.1098/RSTB.2017.0077.

28 L. S. Lerman, *J. Mol. Biol.*, 1961, **3**, 18–30.

29 W. C. Tsai, N. Bhattacharyya, L. Y. Han, J. A. Hanover and M. M. Rechler, *Endocrinology*, 2003, **144**, 5615–5622.

30 V. Luzzati, F. Masson and L. S. Lerman, *J. Mol. Biol.*, 1961, **3**, 634–639.

31 A. Mukherjee and W. D. Sasikala, *Adv. Protein Chem. Struct. Biol.*, 2013, **92**, 1–62.

32 A. S. Biebricher, I. Heller, R. F. H. Roijmans, T. P. Hoekstra, E. J. G. Peterman and G. J. L. Wuite, *Nat. Commun. 2015 61*, 2015, **6**, 1–12.

33 H. M. Sobell and S. C. Jain, *J. Mol. Biol.*, , DOI:10.1016/0022-2836(72)90259-8.

34 A. Soni, P. Khurana, T. Singh and B. Jayaram, *Bioinformatics*, 2017, **33**, 1488–1496.

35 F. P. Guengerich, *Mol. Life Sci.*, 2018, 45–46.

36 H. Taymaz-Nikerel, M. E. Karabekmez, S. Eraslan and B. Kırdar, *Sci. Reports 2018 81*, 2018, **8**, 1–14.

37 F. Yang, S. S. Teves, C. J. Kemp and S. Henikoff, *Biochim. Biophys. Acta*, 2014, **1845**, 84.

38 F. Yang, S. S. Teves, C. J. Kemp and S. Henikoff, *Biochim. Biophys. Acta*, 2014, **1845**, 84.

39 D. J. Taatjes, G. Gaudiano, K. Resing and T. H. Koch, *J. Med. Chem.*, 1997, **40**, 1276–1286.

40 D. J. Taatjes, G. Gaudiano, K. Resing and T. H. Koch, *J. Med. Chem.*, 1996, **39**, 4135–4138.

41 A. K. McClendon and N. Osheroff, *Mutat. Res.*, 2007, **623**, 83.

42 A. A. Shokeir, M. K. al-Hussaini and I. A. Wasfy, *Br. J. Ophthalmol.*, 1969, **53**, 263–266.

43 P. E. Nielsen, M. Egholm, R. H. Berg and O. Buchardt, *Science (80-.).*, 1991, **254**, 1497–1500.

44 M. L. Kopka, C. Yoon, D. Goodsell, P. Pjura and R. E. Dickerson, *Proc. Natl. Acad. Sci. U. S. A.*, 1985, **82**, 1376.

45 S. Husale, W. Grange and M. Hegner, *Single Mol.*, 2002, **3**, 91–96.

46 G. S. Khan, A. Shah, Zia-Ur-Rehman and D. Barker, *J. Photochem. Photobiol. B Biol.*, 2012, **115**, 105–118.

47 M. Asagi, A. Toyama and H. Takeuchi, *Biophys. Chem.*, 2010, **149**, 34–39.

48 A. Sischka, K. Toensing, R. Eckel, S. D. Wilking, N. Sewald, R. Ros and D. Anselmetti, *Biophys. J.*, 2005, **88**, 404.

49 P. C. Zamecnik and M. L. Stephenson, *Proc. Natl. Acad. Sci. U. S. A.*, 1978, **75**, 280.

50 D. R. Corey, M. J. Damha and M. Manoharan, *Nucleic Acid Ther.*, 2022, **32**, 8–13.

51 P. C. Zamecnik, *Antisense Ther.*, 1996, 1–11.

52 C. F. Bennett, N. Dean, D. J. Ecker and B. P. Monia, *Antisense Ther.*, 1996, 13–46.

53 M. Mansoor and A. J. Melendez, *Gene Regul. Syst. Bio.*, 2008, **2**, 275.

54 J. Kurreck, *Eur. J. Biochem.*, 2003, **270**, 1628–1644.

55 R. S. Geary, S. P. Henry and L. R. Grillone, *Clin. Pharmacokinet.*, 2002, **41**, 255–260.

56 S. A. Vinores, *Int. J. Nanomedicine*, 2006, **1**, 263.

57 J. Zhou and J. Rossi, *Nat. Rev. Drug Discov.*, 2017, **16**, 181.

58 S. T. Crooke and R. S. Geary, *Br. J. Clin. Pharmacol.*, 2013, **76**, 269.

59 P. B. Duell, R. D. Santos, B. A. Kirwan, J. L. Witztum, S. Tsimikas and J. J. P. Kastelein, *J. Clin. Lipidol.*, 2016, **10**, 1011–1021.

60 A. Aartsma-Rus, *https://home.liebertpub.com/nat*, 2017, **27**, 67–69.

61 A. K. Gubitz, W. Feng and G. Dreyfuss, *Exp. Cell Res.*, 2004, **296**, 51–56.

62 X. Shen and D. R. Corey, *Nucleic Acids Res.*, 2018, **46**, 1584–1600.

63 S. T. Crooke, J. L. Witztum, C. F. Bennett and B. F. Baker, *Cell Metab.*, 2018, **27**, 714–739.

64 R. L. Setten, J. J. Rossi and S. ping Han, *Nat. Rev. Drug Discov. 2019 186*, 2019, **18**, 421–446.

65 M. Alberer, U. Gnad-Vogt, H. S. Hong, K. T. Mehr, L. Backert, G. Finak, R. Gottardo, M. A. Bica, A. Garofano, S. D. Koch, M. Fotin-Mleczek, I. Hoerr, R. Clemens and F. von Sonnenburg, *Lancet*, 2017, **390**, 1511–1520.

66 M. W. Tenforde, M. M. Patel, A. A. Ginde, D. J. Douin, H. K. Talbot, J. D. Casey, N. M. Mohr, A. Zepeski, M. Gaglani, T. McNeal, S. Ghamande, N. I. Shapiro, K. W. Gibbs, D. C. Files, D. N. Hager, A. Shehu, M. E. Prekker, H. L. Erickson, M. C. Exline, M. N. Gong, A. Mohamed, D. J. Henning, J. S. Steingrub, I. D. Peltan, S. M. Brown, E. T. Martin, A. S. Monto, A. Khan, C. T. Hough, L. Busse, C. C. ten Lohuis, A. Duggal, J. G. Wilson, A. J. Gordon, N. Qadir, S. Y. Chang, C. Mallow, H. B. Gershengorn, H. M. Babcock, J. H. Kwon, N. Halasa, J. D. Chappell, A. S. Lauring, C. G. Grijalva, T. W. Rice, I. D. Jones, W. B. Stubblefield, A. Baughman, K. N. Womack, C. J. Lindsell, K. W. Hart, Y. Zhu, S. M. Olson, M. Stephenson, S. J. Schrag, M. Kobayashi, J. R. Verani, W. H. Self and F. the I. and O. V. in the A. I. (IVY) Network, *medRxiv*, 2021, 2021.07.08.21259776.

67 G. Felsenfeld and A. Rich, *Biochim. Biophys. Acta*, 1957, **26**, 457–468.

68 K. Hoogsteen, *Acta Crystallogr.*, 1959, **12**, 822–823.

69 A. K. Shchyolkina, E. N. Timofeev, O. F. Borisova, I. A. Il'icheva, E. E. Minyat, E. B. Khomyakova and V. L. Florentiev, *FEBS Lett.*, 1994, **339**, 113–118.

70 A. K. Shchyolkina, E. N. Timofeev, Y. P. Lysov, V. L. Florentiev, T. M. Jovin and D. J. Arndt-Jovin, *Nucleic Acids Res.*, 2001, **29**, 986.

71 M. Cooney, G. Czernuszewicz, E. H. Postel, S. J. Flint and M. E. Hogan, *Science (80-.).*, 1988, **241**, 456–459.

72 S. Buchini and C. J. Leumann, *Curr. Opin. Chem. Biol.*, 2003, **7**, 717–726.

73 P. A. Beal and P. B. Dervan, *Science (80-.).*, 1991, **251**, 1360–1363.

74 A. G. Letai, M. A. Palladino, E. Fromm, V. Rizzo and J. R. Fresco, *Biochemistry*, 1988, **27**, 9108–9112.

75 R. H. Durland, D. J. Kessler, S. Gunnell, M. E. Hogan, M. Duvic and B. M. Pettitt, *Biochemistry*, 1991, **30**, 9246–9255.

76 F. A. Buske, J. S. Mattick and T. L. Bailey, *RNA Biol.*, 2011, **8**, 427–439.

77 A. Mukherjee and K. M. Vasquez, *Biochimie*, 2011, **93**, 1197.

78 S. B. Boulware, L. A. Christensen, H. Thames, L. Coghlan, K. M. Vasquez and R. A. Finch, *Mol. Carcinog.*, 2014, **53**, 744–752.

79 M. Z. Akhter and M. R. Rajeswari, *J. Biomol. Struct. Dyn.*, 2016, **35**, 689–703.

80 M. D. Frank-Kamenetskii and S. M. Mirkin, *Annu. Rev. Biochem.*, 1995, **64**, 65–95.

81 S. S. Gaddis, Q. Wu, H. D. Thames, J. Digiovanni, E. F. Walborg, M. C. MacLeod and K. M. Vasquez, *Oligonucleotides*, 2006, **16**, 196–201.

82 P. Jenjaroenpun, C. S. Chew, T. P. Yong, K. Choowongkomon, W. Thammasorn and V. A. Kuznetsov, *Nucleic Acids Res.*, 2015, **43**, D110–D116.

83 K. M. Vasquez and J. H. Wilson, *Trends Biochem. Sci.*, 1998, **23**, 4–9.

84 K. Hyung-Gyoon and D. M. Miller, *Biochemistry*, 1995, **34**, 8165–8171.

85 L. James Maher, *Biochemistry*, 1992, **31**, 7587–7594.

86 H. G. Kim and D. M. Miller, *Biochemistry*, 1998, **37**, 2666–2672.

87 H. G. Kim, J. F. Reddoch, C. Mayfield, S. Ebbinghaus, N. Vigneswaran, S. Thomas, D. E. Jones and D. M. Miller, *Biochemistry*, 1998, **37**, 2299–2304.

88 C. Escudé, C. Giovannangeli, J. S. Sun, D. H. Lloyd, J. K. Chen, S. M. Gryaznov, T. Garestier and C. Hélène, *Proc. Natl. Acad. Sci.*, 1996, **93**, 4365–4369.

89 L. Xodo, M. Alunni-fabbroni, G. Manzini and F. Quadrifoglio, *Nucleic Acids Res.*, 1994, **22**, 3322.

90 R. H. Durland, T. S. Rao, G. R. Revankar, J. H. Tinsley, M. A. Myrick, D. M. Seth, J. Rayford, P. Singh and K. Jayaraman, *Nucleic Acids Res.*, 1994, **22**, 3233.

91 J. M. Dagle and D. L. Weeks, *Nucleic Acids Res.*, 1996, **24**, 2143.

92 T. S. Rao, R. H. Durland, D. M. Seth, M. A. Myrick, V. Bodepudi and G. R. Revankar, *Biochemistry*, 1995, **34**, 765–772.

93 W. M. Olivas and L. J. Maher, *Nucleic Acids Res.*, 1995, **23**, 1936.

94 J. F. Milligan, S. H. Krawczyk, S. Wadwani and M. D. Matteucci, *Nucleic Acids Res.*, 1993, **21**, 327–333.

95 N. Bianchi, C. Rutigliano, M. Passadore, M. Tomassetti, L. Pippo, C. Mischiati, G. Ferioho and R. Gambari, *Biochem. J.*, 1997, **326**, 919.

96 G. M. Carbone, E. M. McGuffie, A. Collier and C. V. Catapano, *Nucleic Acids Res.*,

2003, **31**, 833.

97 S. B. Noonberg, G. K. Scott, C. A. Hunt, M. E. Hogan and C. C. Benz, *Gene*, 1994, **149**, 123–126.

98 P. Nielsen, H. M. Pfundheller, C. E. Olsen, J. Wengel, J. Chem, V. E. Marquez, M. A. Siddiqui, A. Ezzitouni, P. Russ, J. Wang, R. W. Wagner, M. D. Matteucci, R. D. Youssefyeh, J. P. H Verheyden and J. G. Moffatt, *Chem. Commun.*, 1998, **90**, 455–456.

99 S. Obika, D. Nanbu, Y. Hari, J. I. Andoh, K. I. Morio, T. Doi and T. Imanishi, *Tetrahedron Lett.*, 1998, **39**, 5401–5404.

100 P. Nielsen, H. M. Pfundheller and J. Wengel, *Chem. Commun.*, 1997, 825–826.

101 A. O. Schmitt and H. Herzel, *J. Theor. Biol.*, 1997, **188**, 369–377.

102 L. Jen-Jacobson, L. E. Engler and L. A. Jacobson, *Structure*, 2000, **8**, 1015–1023.

103 A. A. Koshkin, S. K. Singh, P. Nielsen, V. K. Rajwanshi, R. Kumar, M. Meldgaard, C. E. Olsen and J. Wengel, *Tetrahedron*, 1998, **54**, 3607–3630.

104 A. Grünweller, E. Wyszko, B. Bieber, R. Jahnel, V. A. Erdmann and J. Kurreck, *Nucleic Acids Res.*, 2003, **31**, 3185.

105 J. Kurreck, E. Wyszko, C. Gillen and V. A. Erdmann, *Nucleic Acids Res.*, 2002, **30**, 1911.

106 J. Roberts, E. Palma, P. Sazani, H. Ørum, M. Cho and R. Kole, *Mol. Ther.*, 2006, **14**, 471–475.

107 E. E. Swayze, A. M. Siwkowski, E. V. Wancewicz, M. T. Migawa, T. K. Wyrzykiewicz, G. Hung, B. P. Monia and C. F. Bennett, *Nucleic Acids Res.*, 2007, **35**, 687.

108 J. C. François, T. Saison-Behmoaras and C. Hélène, *Nucleic Acids Res.*, 1988, **16**, 11431–11440.

109 S. A. Strobel and P. B. Dervan, *Science*, 1990, **249**, 73–75.

110 L. C. Griffin and P. B. Dervan, *Science*, 1989, **245**, 967–971.

111 O. S. Fedorova, D. G. Knorre, L. M. Podust and V. F. Zarytova, *FEBS Lett.*, 1988, **228**, 273–276.

112 J. Klysik, K. Rippe and T. M. Jovin, *Biochemistry*, 1990, **29**, 9831–9839.

113 J. H. Van De Sande, N. B. Ramsing, M. W. Germann, W. Elhorst, B. W. Kalisch, E. V. Kitzing, R. T. Pon, R. C. Clegg and T. M. Jovin, *Science (80-.).*, 1988, **241**, 551–557.

114 H. Rasmussen, S. J. Kastrup, J. N. Nielsen, J. M. Nielsen and P. E. Nielsen, *Nat. Struct. Biol. 1997 42*, 1997, **4**, 98–101.

115 P. Wittung, P. E. Nielsen, O. Buchardt, M. Egholm and B. Nordén, *Nature*, 1994, **368**, 561–563.

116 K. L. Dueholm, M. Egholm, C. Behrens, L. Christensen, H. F. Hansen, T. Vulpius, K. H. Petersen, R. H. Berg, P. E. Nielsen, O. Buchardt and P. E. Proc Natl Acad, *J. Am. Chem. Soc*, 1994, **59**, 1481.

117 S. A. Thomson, J. A. Josey, R. Cadilla, M. D. Gaul, C. Fred Hassman, M. J. Luzzio,

A. J. Pipe, K. L. Reed, D. J. Ricca, R. W. Wiethe and S. A. Noble, *Tetrahedron*, 1995, **51**, 6179–6194.

118 P. E. Nielsen, M. Egholm and O. Buchardt, *Bioconjug. Chem.*, 1994, **5**, 3–7.

119 M. Egholm, O. Buchardt, P. E. Nielsen and R. H. Berg, *J. Am. Chem. Soc.*, 1992, **114**, 1895–1897.

120 T. Bentin and P. E. Nielsen, *Biochemistry*, 1996, **35**, 8863–8869.

121 N. J. Peffer, J. C. Hanvey, J. E. Bisi, S. A. Thomson, C. F. Hassman, S. A. Noble and L. E. Babiss, *Proc. Natl. Acad. Sci. USA*, 1993, **90**, 10648–10652.

122 P. E. Nielsen and L. Christensen, *J. Am. Chem. Soc.*, 1996, **118**, 2287–2288.

123 J. C. Hanvey, N. J. Peffer, J. E. Bisi, S. A. Thomson, R. Cadilla, J. A. Josey, D. J. Ricca, C. F. Hassman, M. A. Bonham, K. G. Au, S. G. Carter, D. A. Bruckenstein, A. L. Boyd, S. A. Noble and L. E. Babiss, *New Ser.*, 1992, **258**, 1481–1485.

124 X. Zhang, T. Ishihara and D. R. Corey, *Nucleic Acids Res.*, 2000, **28**, 3332.

125 P. Wittung, P. Nielsen and B. Nordén, *J. Am. Chem. Soc.*, 1996, **118**, 7049–7054.

126 K. R. B. Singh, P. Sridevi and R. P. Singh, *Eng. Reports*, , DOI:10.1002/ENG2.12238.

127 V. V. Demidov, V. N. Potaman, M. D. Frank-Kamenetskil, M. Egholm, O. Buchard, S. H. Sönnichsen and P. E. Nlelsen, *Biochem. Pharmacol.*, 1994, **48**, 1310–1313.

128 A. Kiliszek, K. Banaszak, Z. Dauter and W. Rypniewski, *Nucleic Acids Res.*, 2016, **44**, 1937.

129 G. L. Igloi, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 8562.

130 F. P. Schwarz, S. Robinson and J. M. Butler, *Nucleic Acids Res.*, 1999, **27**, 4792.

131 K. Kilså Jensen, H. Ørum, P. E. Nielsen and B. Nordén, *Biochemistry*, 1997, **36**, 5072–5077.

132 M. Egholm, O. Buchardt, L. Christensen, C. Behrens, S. M. Frelert, D. A. Drivert, R. H. Bergt, S. K. Kim, B. Norden and P. E. Nielsen II, *Nature*, 1993, 566–568.

133 J. Wang, E. Palecek, P. E. Nielsen, G. Rivas, X. Cai, H. Shiraishi, N. Dontha, D. Luo and P. A. M. Farias, *J. Am. Chem. Soc.*, 1996, **118**, 7667–7670.

134 C. Thiede, E. Bayerdörffer, R. Blasczyk, B. Wittig and A. Neubauer, *Nucleic Acids Res.*, 1996, **24**, 983.

135 C. Carlsson, M. Jonsson, B. Norden, M. T. Dulay, R. N. Zare, J. Noolandi, P. E. Nielsen, L. C. Tsui and J. Zielenski, *Nat. 1996 3806571*, 1996, **380**, 207–207.

136 H. Perry-O'Keefe, X. W. Yao, J. M. Coull, M. Fuchs and M. Egholm, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 14670–14675.

137 S. A. Webb and P. Hurskainen, *SLAS Discov.*, 1996, **1**, 119–121.

138 D. R. Corey, *Trends Biotechnol.*, 1997, **15**, 224–229.

139 F. Pellestor, *Hum. Reprod.*, 2004, **19**, 1946–1951.

140 R. Marchelli, R. Corradini, A. Manicardi, S. Sforza, T. Tedeschi, E. Fabbri, M. Borgatti, N. Bianchi and R. Gambari, in *Targets in Gene Therapy*, InTech, 2011.

141 P. E. Nielsen and T. Shiraishi, *Artif. DNA. PNA XNA*, 2011, **2**, 90.

142 H. J. Larsen, T. Bentin and P. E. Nielsen, *Biochim. Biophys. Acta - Gene Struct. Expr.*, 1999, **1489**, 159–166.

143 M. Eriksson, L. Christensen, J. Schmidt, G. Haaima, L. Orgel and P. E. Nielsen, *New J. Chem.*, 1998, **22**, 1055–1059.

144 M. Egholm, O. Buchardt, P. E. Nielsen and R. H. Berg, *J. Am. Chem. Soc.*, 2002, **114**, 9677–9678.

145 M. Egholm, C. Behrens, L. Christensen, R. H. Berg, P. E. Nielsen and O. Buchardt, *J. Chem. Soc. Chem. Commun.*, 1993, 800–801.

146 A. Gupta, A. Mishra and N. Puri, *J. Biotechnol.*, 2017, **259**, 148–159.

147 C. Pesce, F. Bolacchi, B. Bongiovanni, F. C.-A. research and undefined 2005, *Elsevier*.

148 C. Gambacorti-Passerini, L. Mologni and C. Bertazzoli, .

149 L. Good, P. N.-N. biotechnology and undefined 1998, *nature.com*, 1998, **16**, 355–358.

150 P. Paulasova and F. Pellestor, *Ann. Génétique*, 2004, **47**, 349–358.

151 F. Pellestor and P. Paulasova, *Chromosoma*, 2004, **112**, 375–380.

152 C. Chen, B. lin Wu, T. Wei, M. Egholm and W. M. Strauss, *Mamm. Genome 2000 115*, 2000, **11**, 384–391.

153 S. Jeong, J. O. Kim, S. H. Jeong, I. K. Bae and W. Song, *J. Microbiol. Methods*, 2015, **113**, 4–9.

154 F. Pellestor and P. Paulasova, *Eur. J. Hum. Genet. 2004 129*, 2004, **12**, 694–700.

155 S. Yu, J. Wu, S. Xu, G. Tan, B. Liu and J. Feng, *Cancer Biol. Ther.*, 2012, **13**, 314–320.

156 R. S. Santos, N. Guimarães, P. Madureira and N. F. Azevedo, *J. Biotechnol.*, 2014, **187**, 16–24.

157 D. Hnedzko, D. W. McGee and E. Rozners, *Bioorg. Med. Chem.*, 2016, **24**, 4199–4205.

158 A. Machado, J. Castro, T. Cereija, C. Almeida and N. Cerca, *PeerJ*, , DOI:10.7717/PEERJ.780/SUPP-1.

159 M. W. Vandersea, R. W. Litaker, B. Yonnish, E. Sosa, J. H. Landsberg, C. Pullinger, P. Moon-Butzin, J. Green, J. A. Morris, H. Kator, E. J. Noga and P. A. Tester, *Appl. Environ. Microbiol.*, 2006, **72**, 1551.

160 H. Kim, J. jin Choi, M. Cho and H. Park, *BioChip J. 2012 61*, 2012, **6**, 25–33.

161 F. Yang, B. Dong, K. Nie, H. Shi, Y. Wu, H. Wang and Z. Liu, *ACS Comb. Sci.*, 2015, **17**, 608–614.

162 Y. Q. Wu, F. P. Yang, H. Y. Wang, J. X. Liu and Z. C. Liu, *J. Nanosci. Nanotechnol.*, 2013, **13**, 2061–2067.

163 S. Hildbrand, A. Blaser, S. P. Parel and C. J. Leumann, *J. Am. Chem. Soc.*, 1997, **119**, 5499–5511.

164 S. A. Cassidy, P. Slickers, J. O. Trent, D. C. Capaldi, P. D. Roselt, C. B. Reese, S. Neidle and K. R. Fox, *Nucleic Acids Res.*, 1997, **25**, 4891–4898.

165 P. J. Bates, C. A. Laughton, T. C. Jenkins, D. C. Capaldi, P. D. Roselt, C. B. Reese and S. Neidle, *Nucleic Acids Res.*, 1996, **24**, 4176–4184.

166 M. Egholm, L. Christensen, K. L. Dueholm, O. Buchardt, J. Coull and P. E. Nielsen, *academic.oup.com*, 1995, **23**, 217–222.

167 A. Ono, P. O. P. Ts'o and L. sing Kan, *J. Org. Chem.*, 1992, **57**, 3225–3230.

168 T. Zengeya, P. Gupta and E. Rozners, *Angew. Chemie*, 2012, **124**, 12761–12764.

169 A. B. Eldrup, O. Dahl and P. E. Nielsen, *J. Am. Chem. Soc.*, 1997, **119**, 11116–11117.

170 C. M. Topham and J. C. Smith, *J. Comput. Aided. Mol. Des.*, 2021, **35**, 355–369.

171 T. Tomori, Y. Miyatake, Y. Sato, T. Kanamori, Y. Masaki, A. Ohkubo, M. Sekine and K. Seio, , DOI:10.1021/acs.orglett.5b00522.

172 O. Diels and K. Alder, *Justus Liebigs Ann. Chem.*, 1928, **460**, 98–122.

173 C. Dockendorff, S. Sahli, M. Olsen, L. Milhau and M. Lautens, *J. Am. Chem. Soc.*, 2005, **127**, 15028–15029.

174 J. Clayden, N. Greeves and S. Warren, *Organic Chemistry*, New York, Second., 2001.

175 A. Köse, M. Kaya, N. H. Kishalı, A. Akdemir, E. Şahin, Y. Kara and G. Şanlı-Mohamed, *Bioorg. Chem.*, , DOI:10.1016/J.BIOORG.2019.103421.

176 Y. W. Goh, B. R. Pool and J. M. White, , DOI:10.1021/jo7018575.

177 G. Wittig and K. Torkssell, *Acta Chem. Scand.*, 1953, 1293–1301.

178 J. Clayden, N. Greeves and S. Warren, *Organic Chemistry*, New York, Second., 2001.

179 M. C. Costas-Lago, P. Besada, F. Rodríguez-Enríquez, D. Viña, S. Vilar, E. Uriarte, F. Borges and C. Terán, *Eur. J. Med. Chem.*, 2017, **139**, 1–11.

180 C. F. Ingham, R. A. Massy-Westropp, G. D. Reynolds and W. D. Thorpe, *Aust. J. Chem.*, 1975, **28**, 2499–2510.

181 A. Skrzyńska, P. Drelich, S. Frankowski and Ł. Albrecht, *Chem. - A Eur. J.*, 2018, **24**, 16543–16547.

182 J. Clayden, N. Greeves and S. Warren, *Organic Chemistry*, New York, Second., 2001.

183 L. Silva, S. Quintiliano, M. Craveiro, F. Vieira and H. Ferraz, *Synthesis (Stuttg).*, 2007, **2007**, 355–362.

184 Y. Q. Zhu, L. W. Hui and S. B. Zhang, *Adv. Synth. Catal.*, 2021, **363**, 2170–2176.

185 J. Clayden, N. Greeves and S. Warren, *Organic Chemistry*, New York, Second., 2001.

186 D. R. Goldberg, T. Butz, M. G. Cardozo, R. J. Eckner, A. Hammach, J. Huang, S. Jakes, S. Kapadia, M. Kashem, S. Lukas, T. M. Morwick, M. Panzenbeck, U. Patel, S. Pav, G. W. Peet, J. D. Peterson, A. S. Prokopowicz, R. J. Snow, R. Sellati, H. Takahashi, J. Tan, M. A. Tschantz, X.-J. Wang, Y. Wang, J. Wolak, P. Xiong and N.

Moss, , DOI:10.1021/jm020446l.

187 J. Clayden, N. Greeves and S. Warren, *Organic Chemistry*, New York, Second., 2001.

188 R. K. Boeckman, J. Pengcheng Shao, J. J. Mullins and K. P. Minbiole, *Org. Synth.*, 2000, **77**, 141.

189 J. A. Souto, D. Román, M. Domínguez and Á. R. de Lera, *European J. Org. Chem.*, 2021, **2021**, 6057–6070.

190 G. Piancatelli, A. Scettri and M. D'Auria, *Synthesis (Stuttg).*, 1982, **1982**, 245–258.

191 H. S. Kasmai, S. G. Mischke and T. J. Blake, *J. Org. Chem.*, 1995, **60**, 2267–2270.

192 E. J. Corey and J. W. Suggs, *Tetrahedron Lett.*, 1975, **16**, 2647–2650.

193 J. Clayden, N. Greeves and S. Warren, *Organic Chemistry*, New York, Second., 2001.

194 F. A. Luzzio, R. W. Fitch, W. J. Moore and K. J. Mudd, *Lab. 974 J. Chem. Educ.* •.

195 U. Ragnarsson and L. Grehn, *RSC Adv.*, 2013, **3**, 18691–18697.

196 M. J. Stephenson, L. A. Howell, M. A. O, K. R. Fox, C. Adcock, J. Kingston, H. Sheldrake, K. Pors, S. P. Collingwood and M. Searcey, , DOI:10.1021/acs.joc.5b01373.

197 M. M. M. D. Cominetti, Z. R. Goddard, C. E. Howman, M. A. O'Connell and M. Searcey, *Tetrahedron Lett.*, 2021, **72**, 153058.

198 I. M. Sakhautdinov, A. M. Gumerov, A. F. Mukhamet'yanova, A. B. Atangulov and M. S. Yunusov, *Chem. Nat. Compd. 2018 543*, 2018, **54**, 622–623.

199 L. W. Ye, X. Han, X. L. Sun and Y. Tang, *Tetrahedron*, 2008, **64**, 8149–8154.

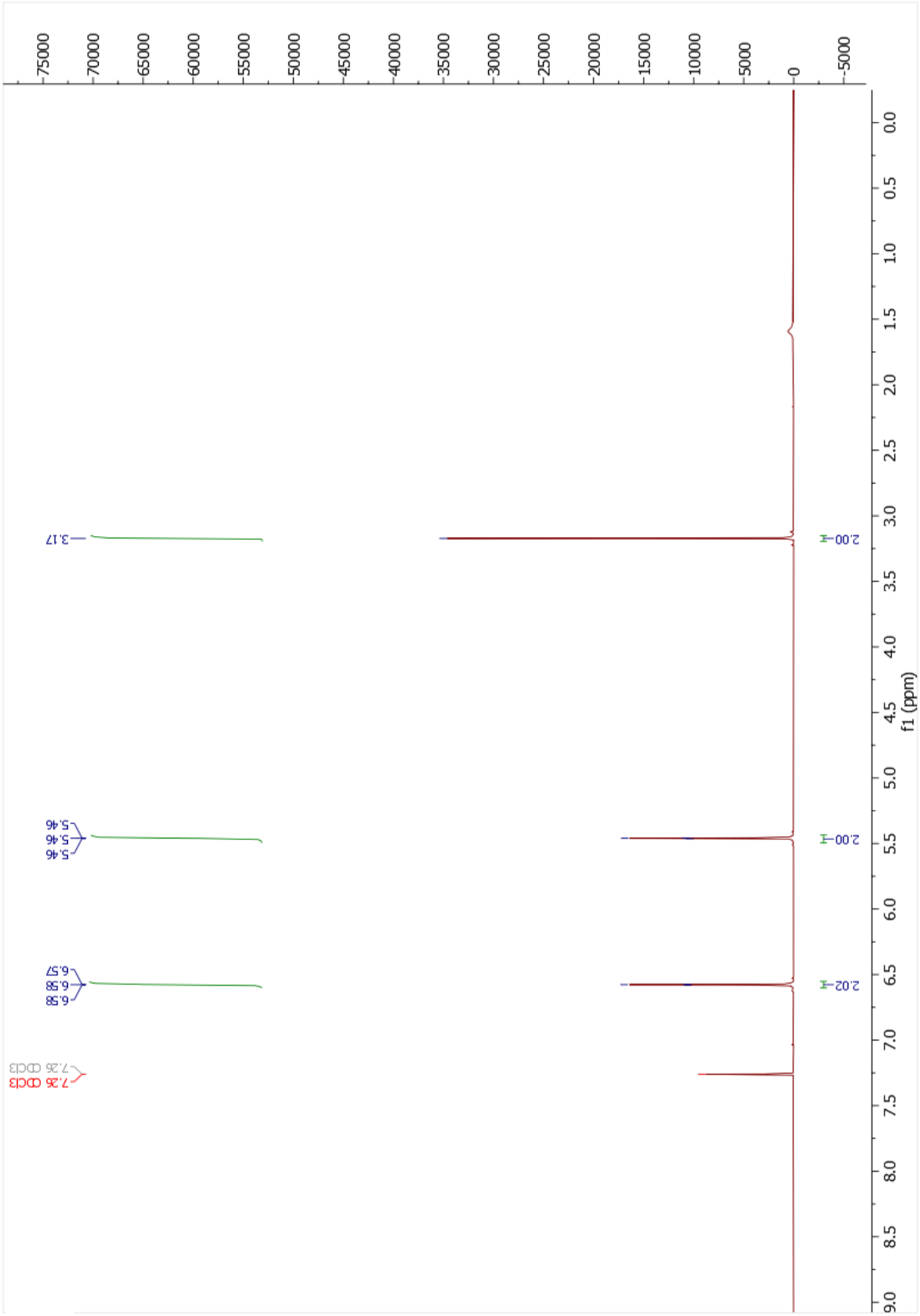
200 N. Matsuo and A. S. Kende, *Izv. Akad. Nauk. SSSR Ser. Khim*, 1988, **53**, 454.

201 A. Frei, B. Spingler and R. Alberto, *Chem. – A Eur. J.*, 2018, **24**, 10156–10164.

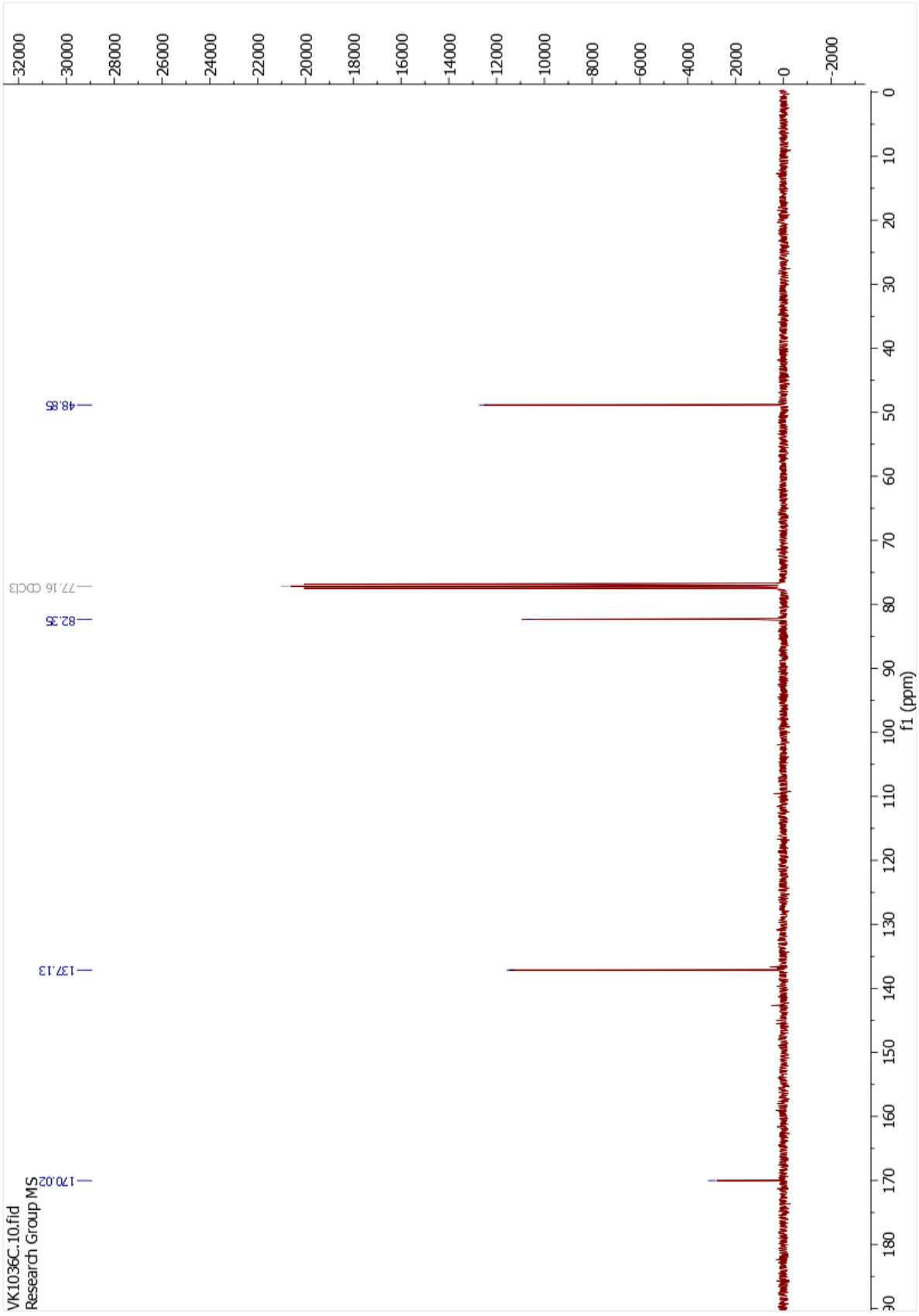
202 C. Granchi, S. Fortunato, S. Meini, F. Rizzolio, I. Caligiuri, T. Tuccinardi, ⊥ Hyang, Y. Lee, P. J. Hergenrother and F. Minutolo, , DOI:10.1021/acs.jafc.7b01668.

Data

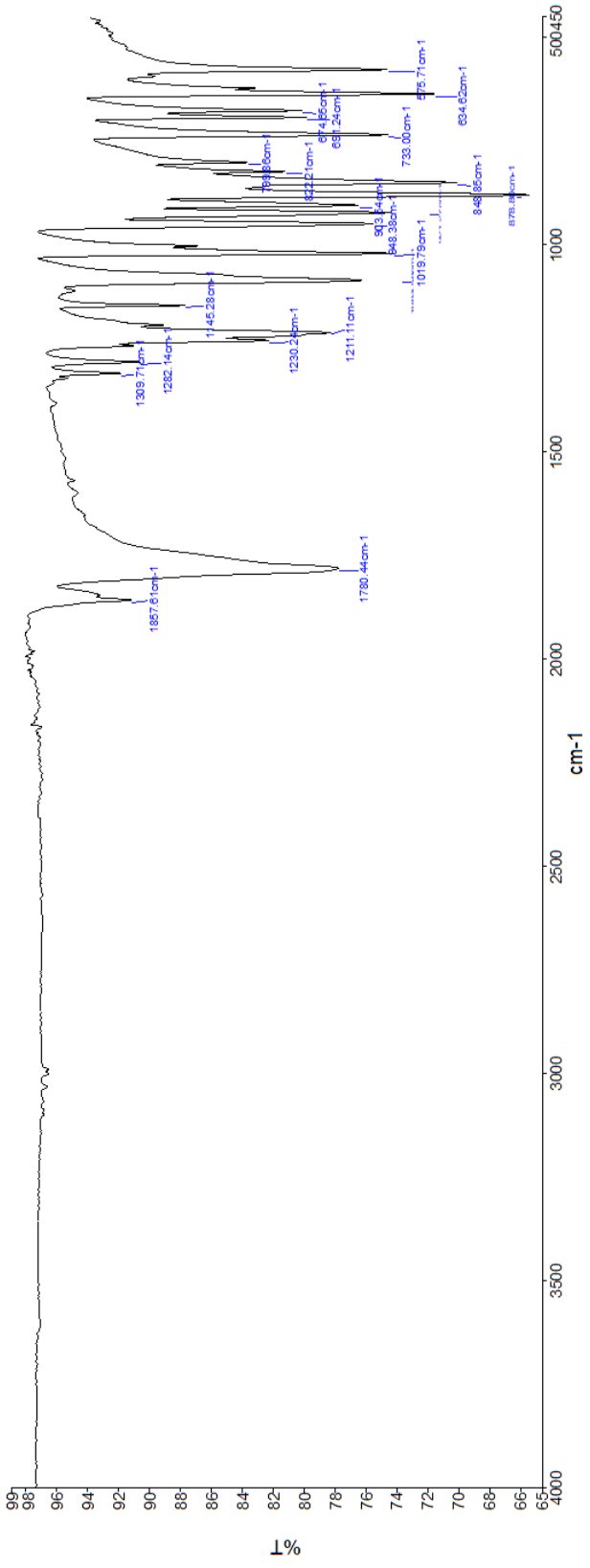
^{1}H NMR of compound **4**



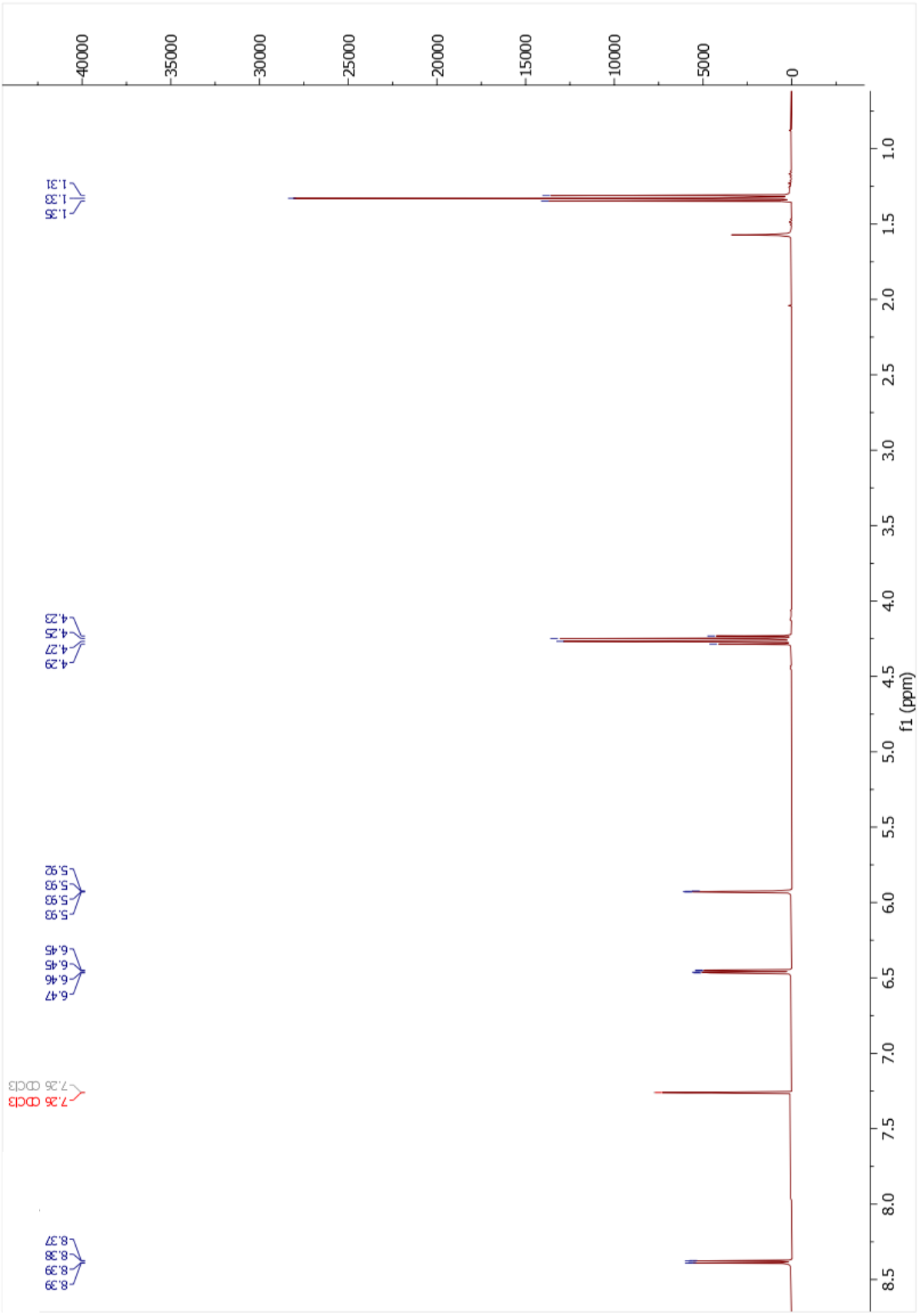
^{13}C NMR of compound **4**



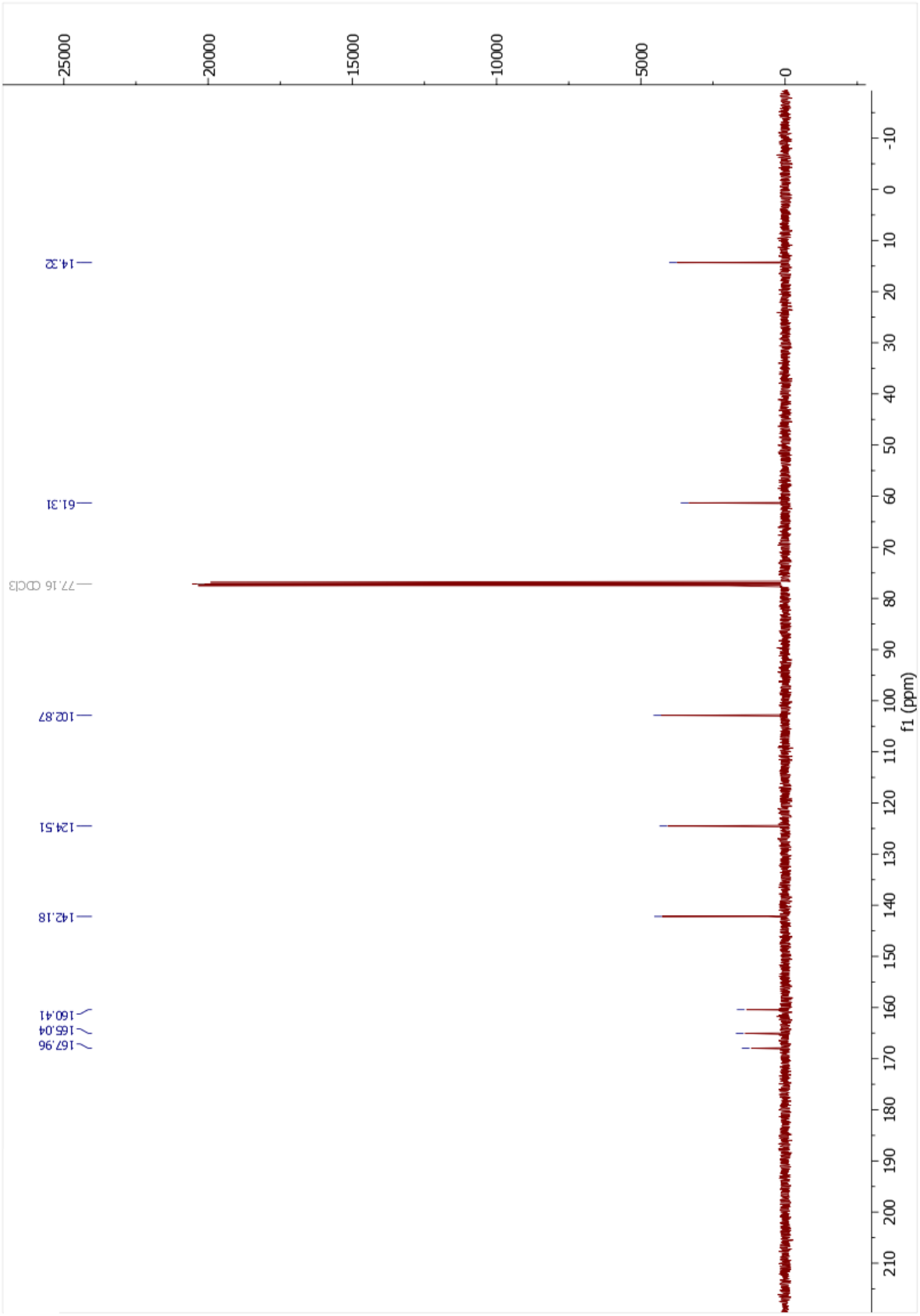
IR spectrum of compound **4**



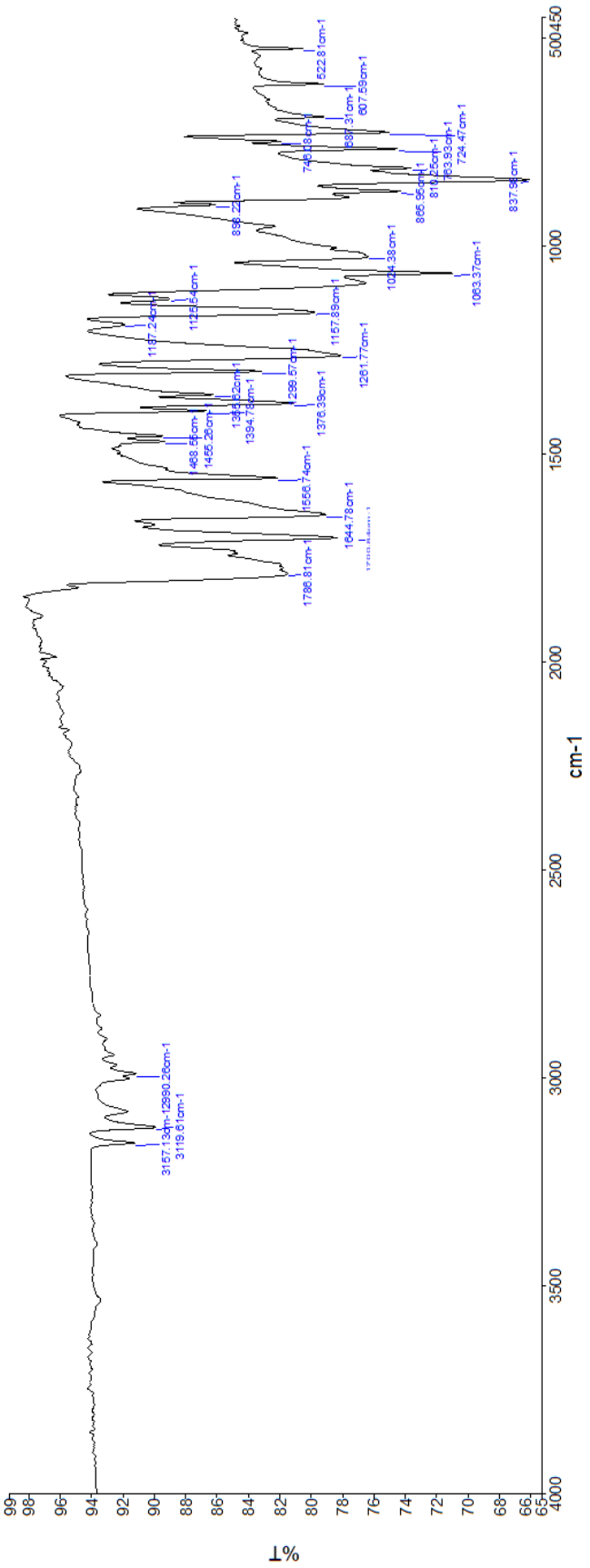
^{1}H NMR of compound **5**



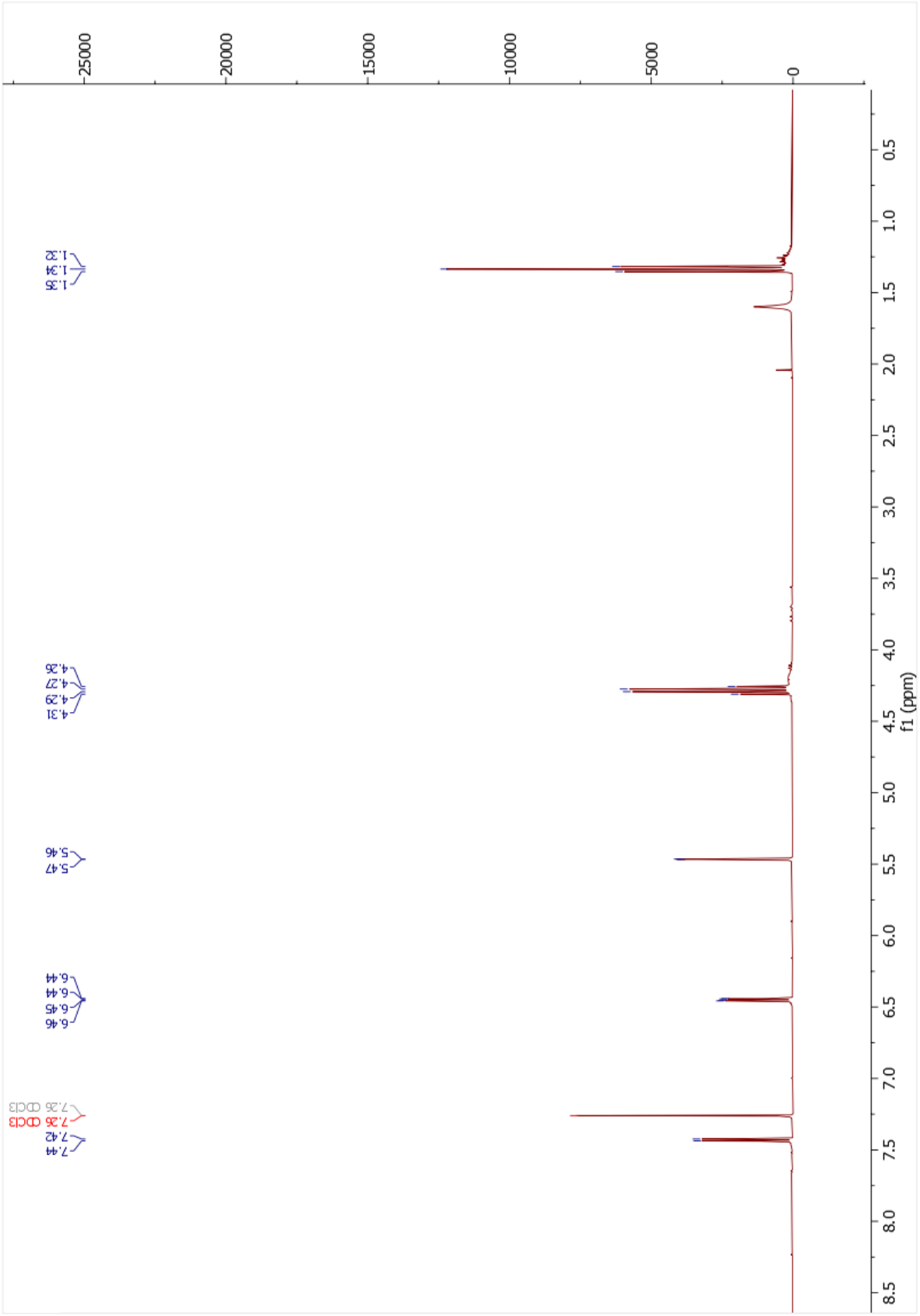
^{13}C NMR of compound **5**



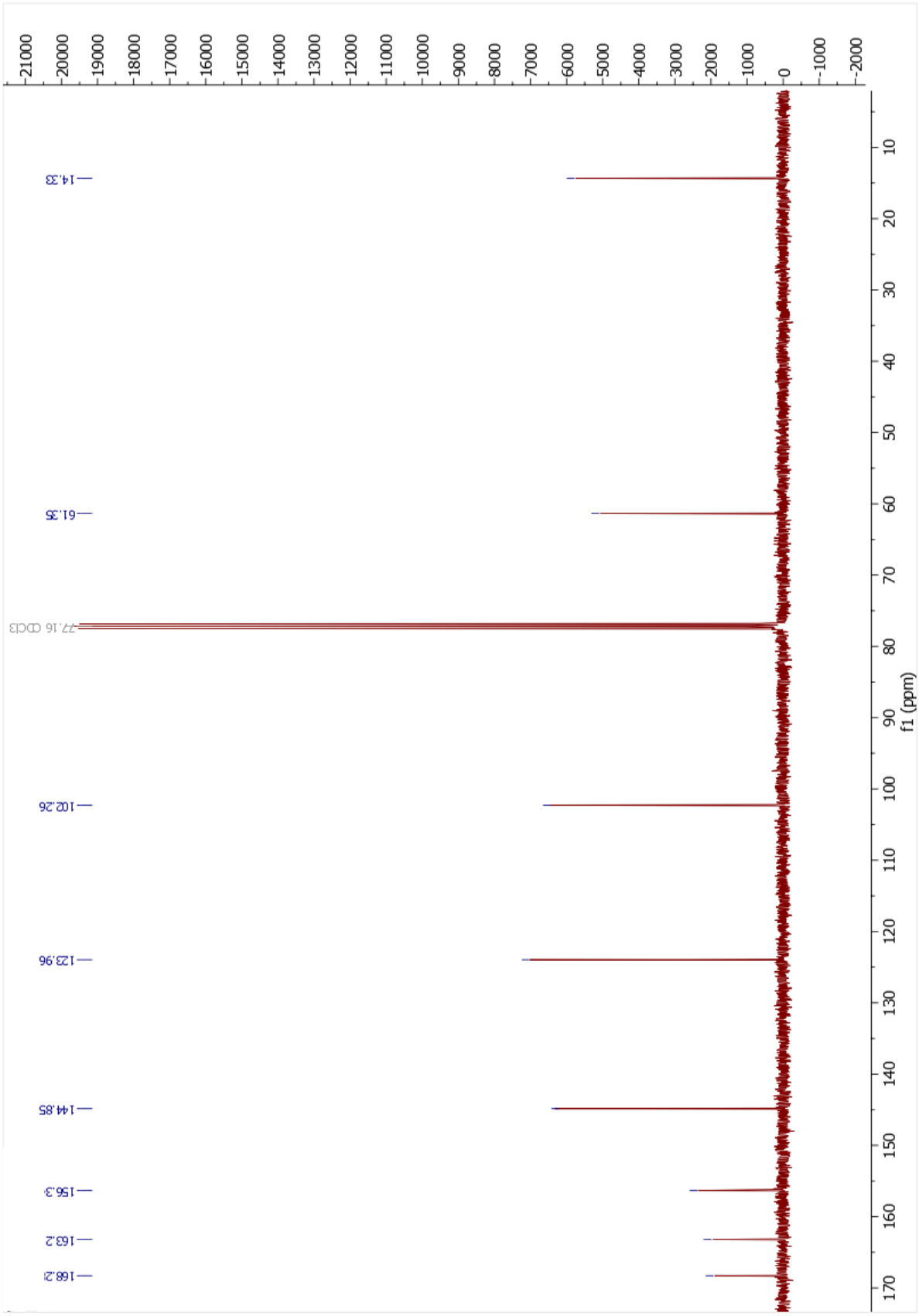
IR spectrum of compound **5**



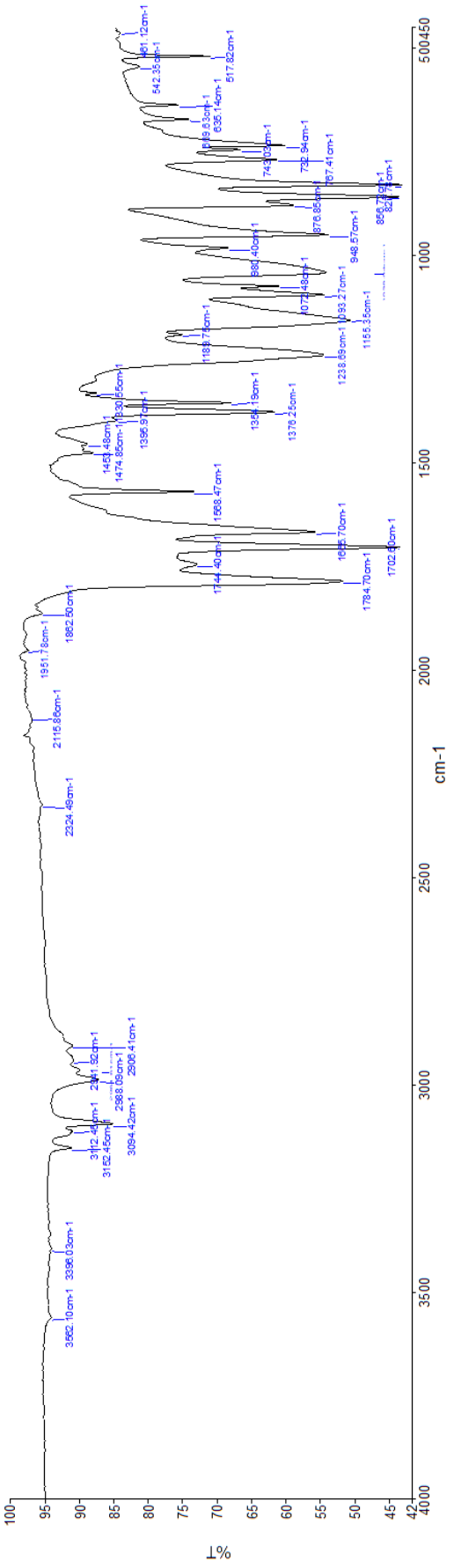
^{1}H NMR of compound **5.1**



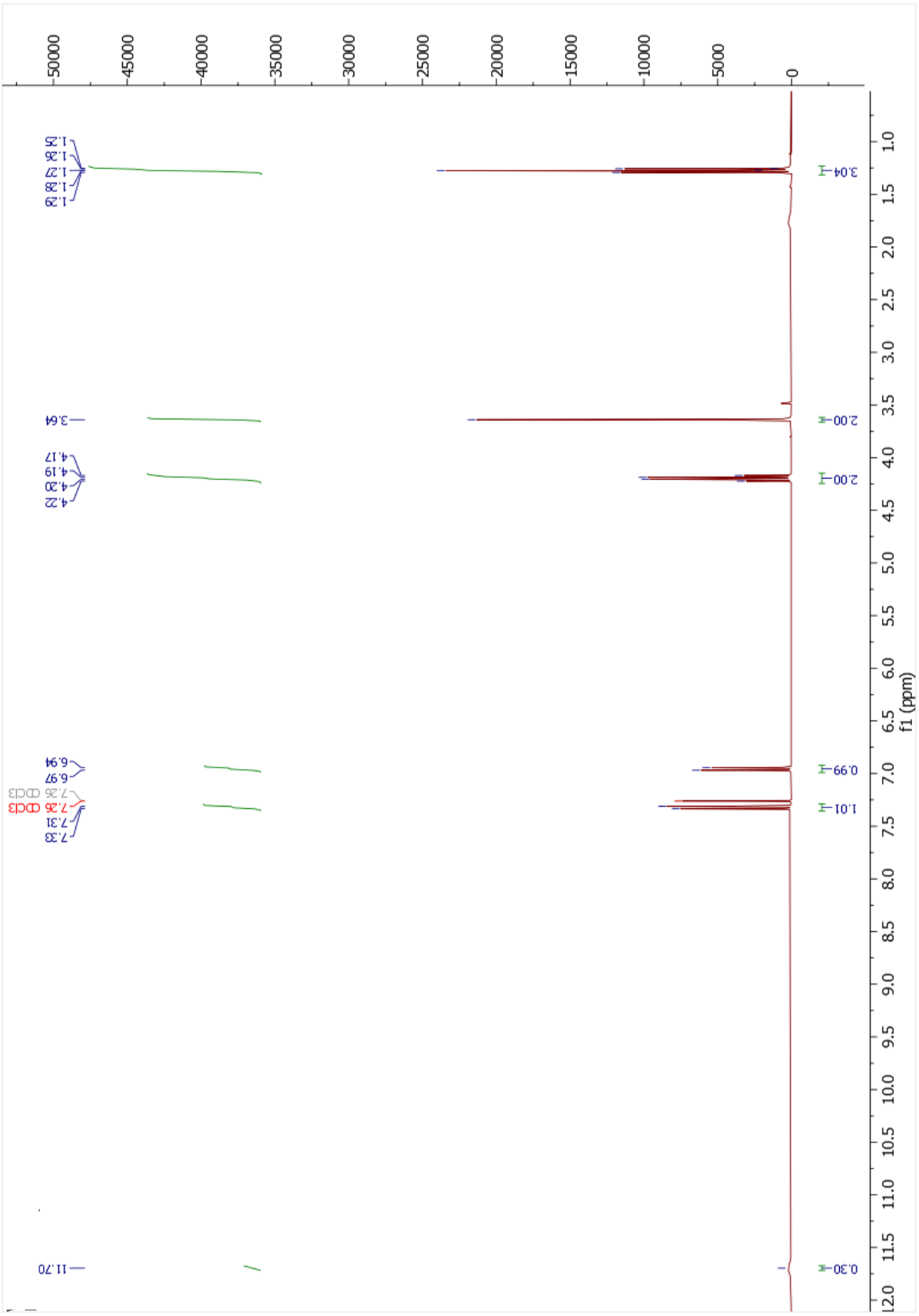
^{13}C NMR of compound **5.1**



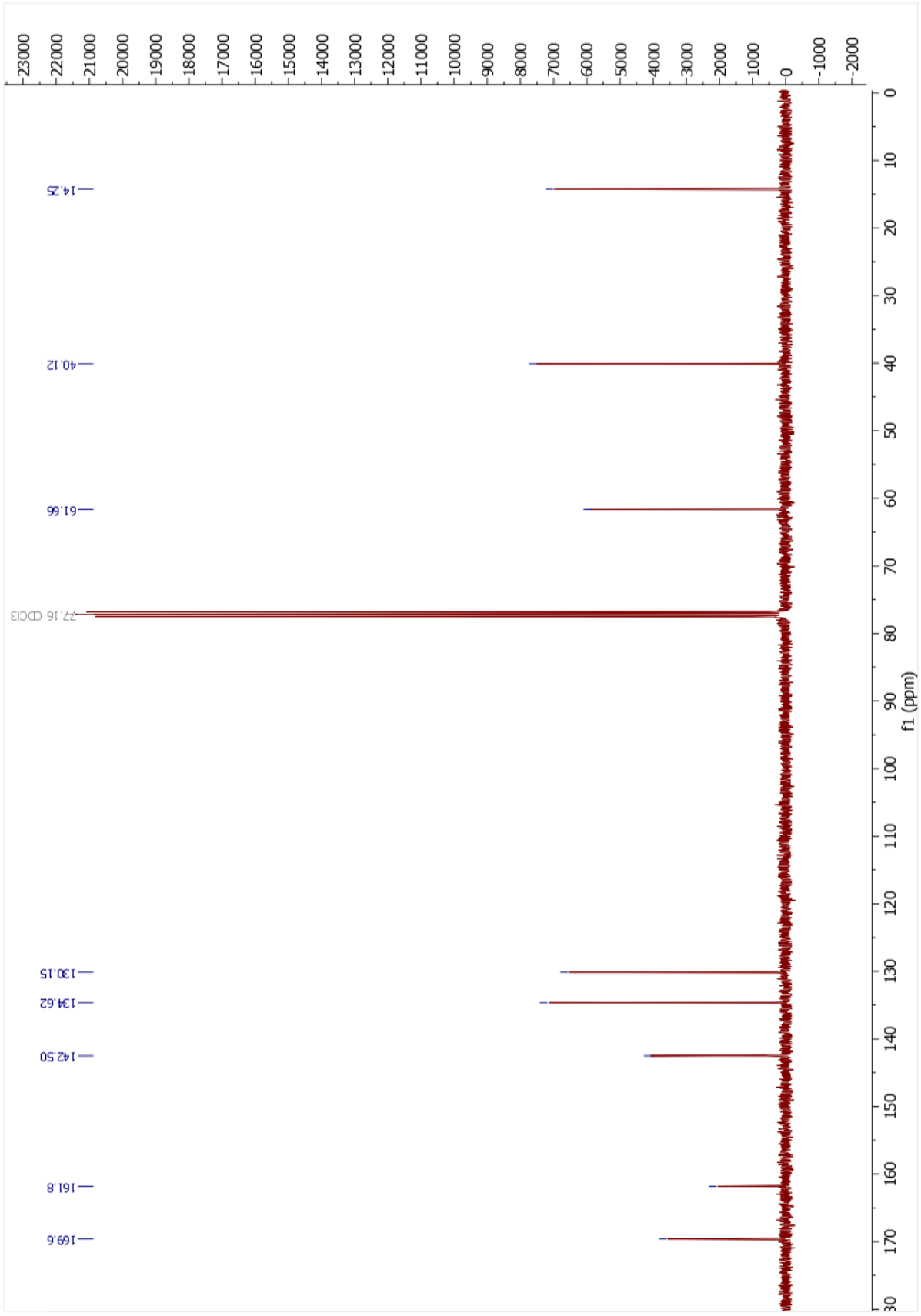
IR spectrum of compound **5.1**



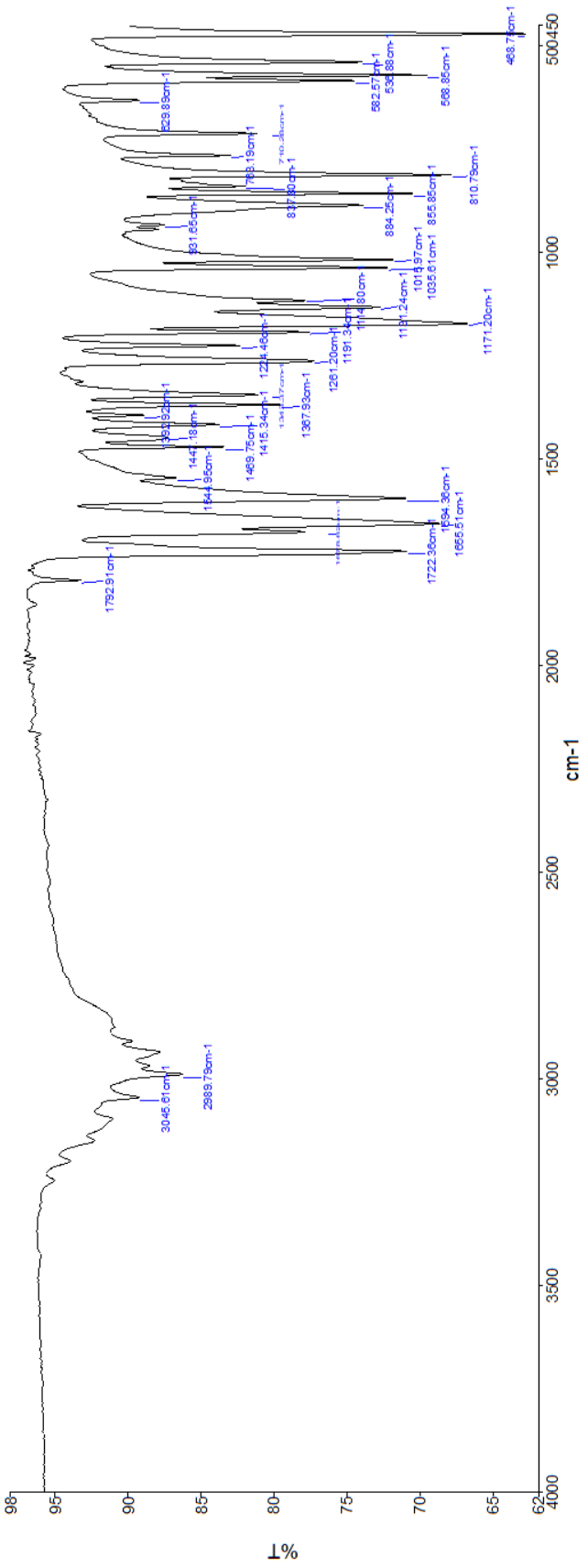
^{1}H NMR of compound **6**



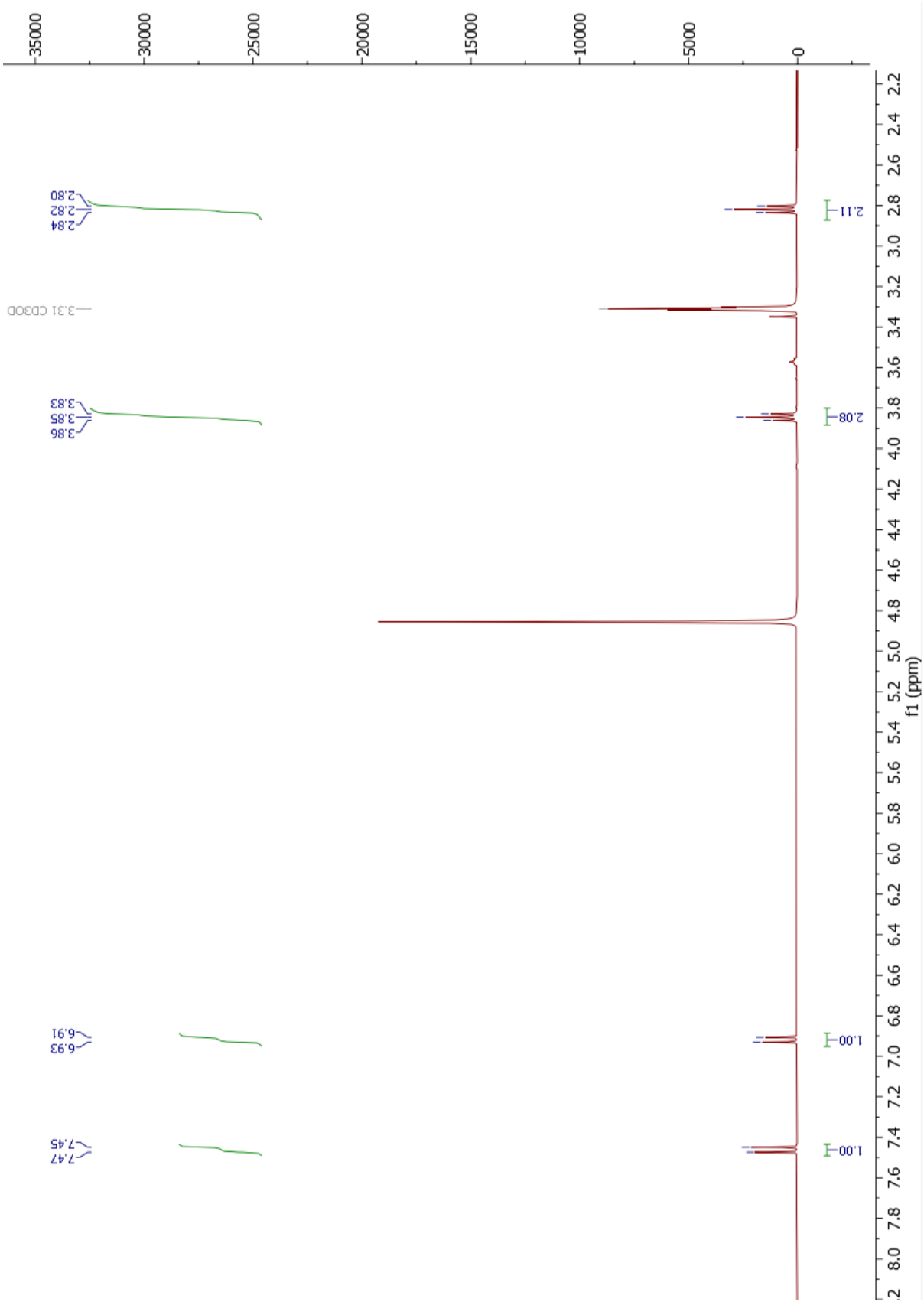
^{13}C NMR of compound **6**



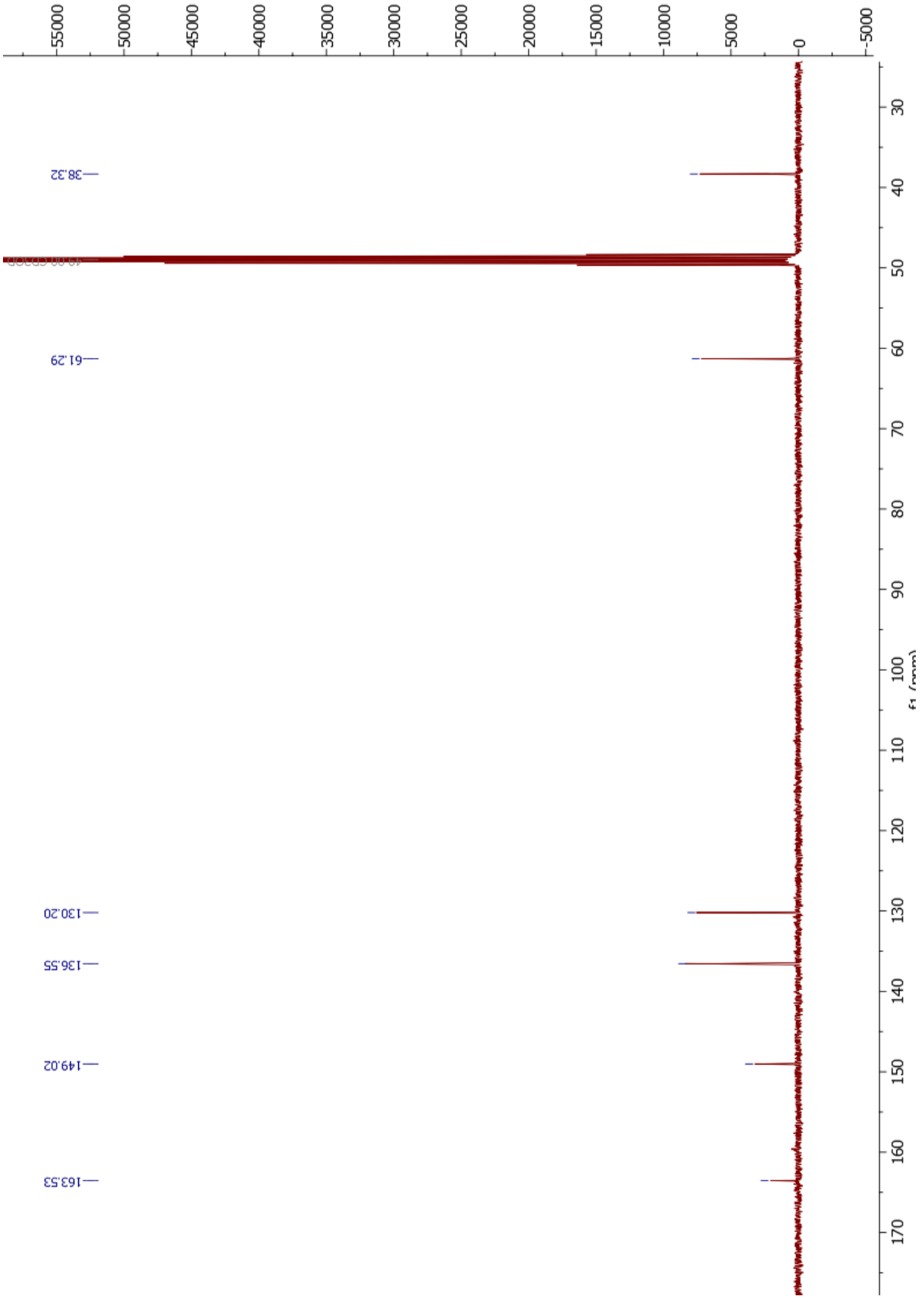
IR spectrum of compound **6**



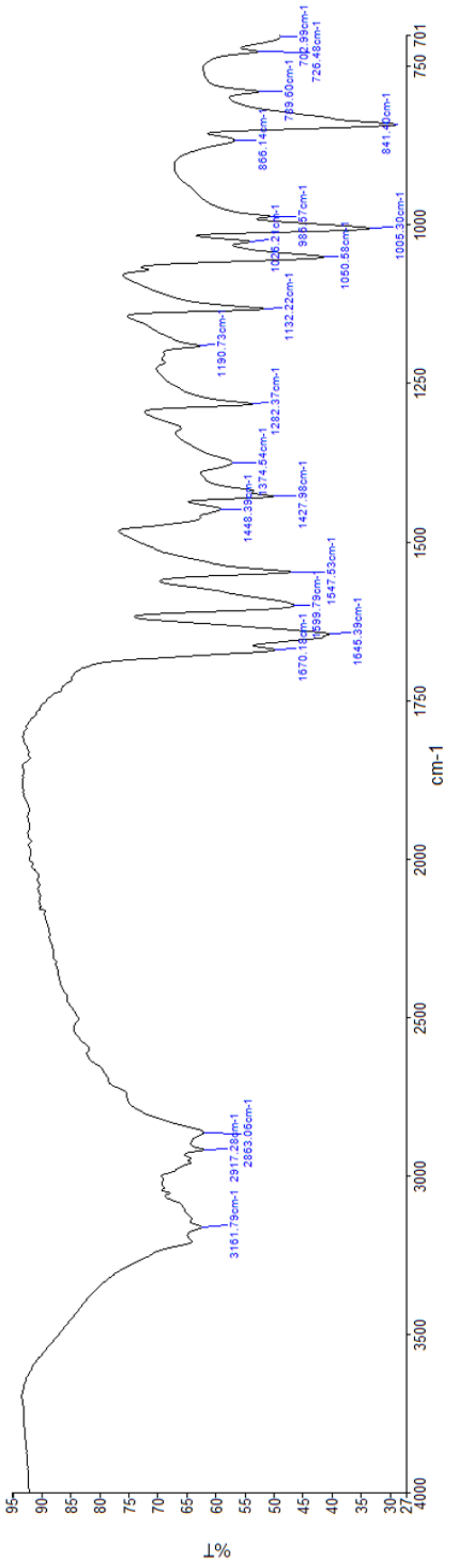
^{1}H NMR of compound **7**



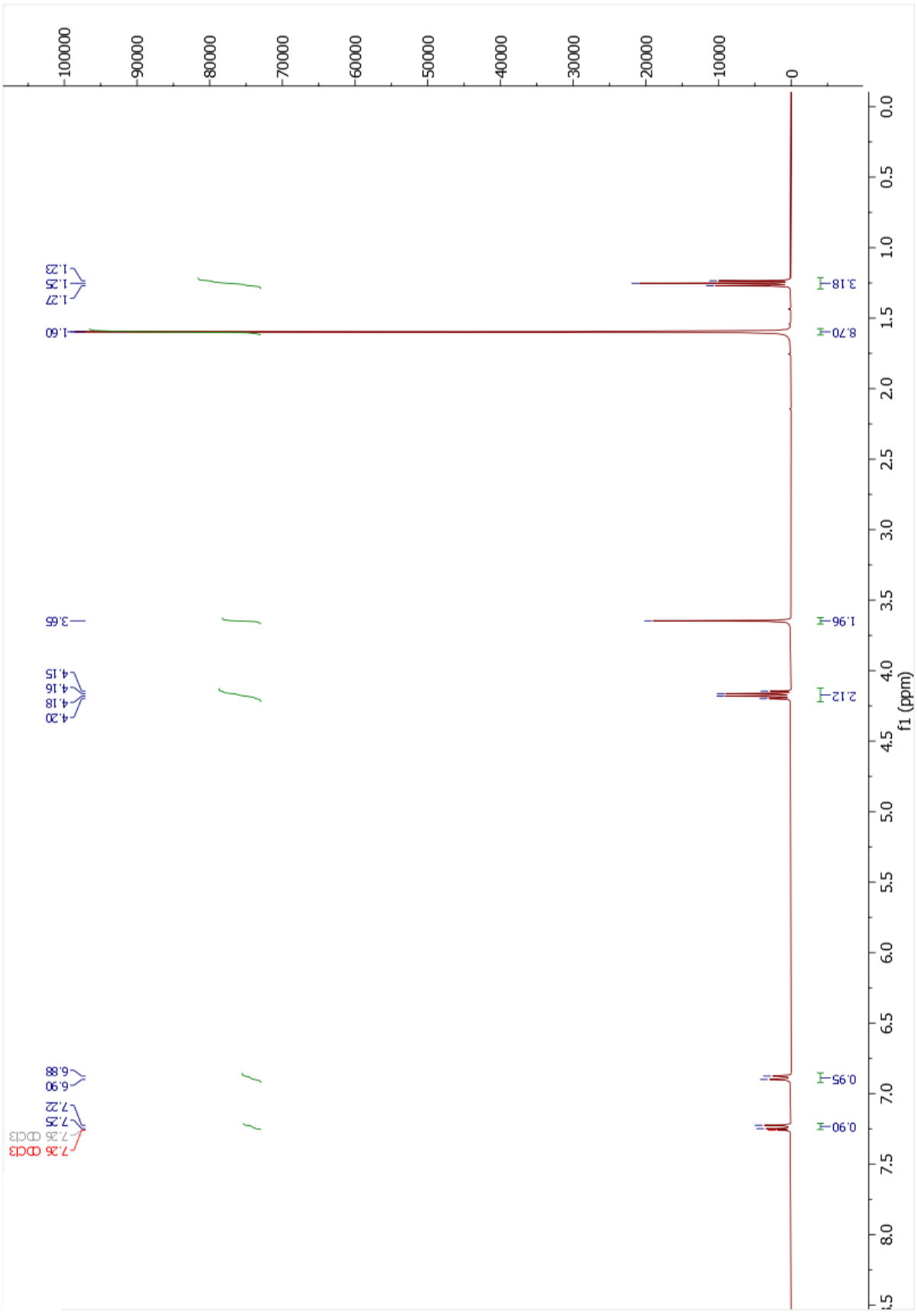
^{13}C NMR of compound **7**



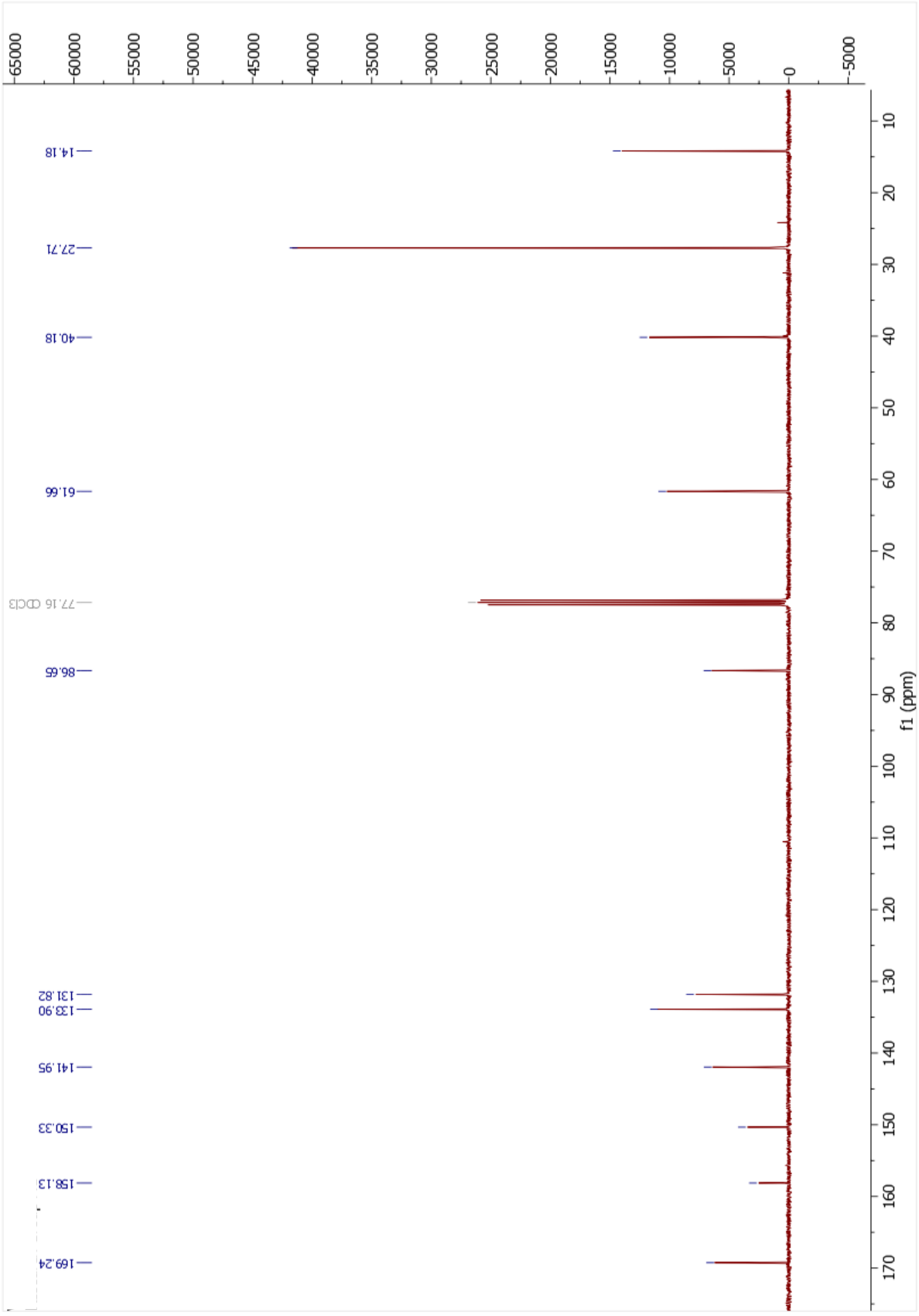
IR spectrum of compound **7**



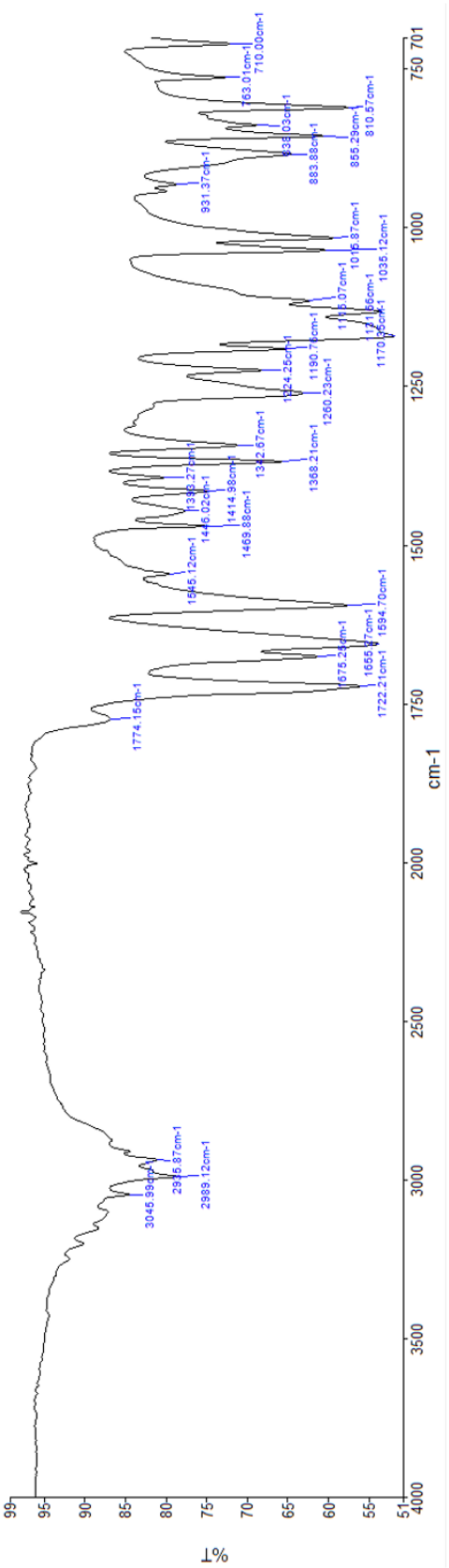
^{1}H NMR of compound **18**



^{13}C NMR of compound **18**

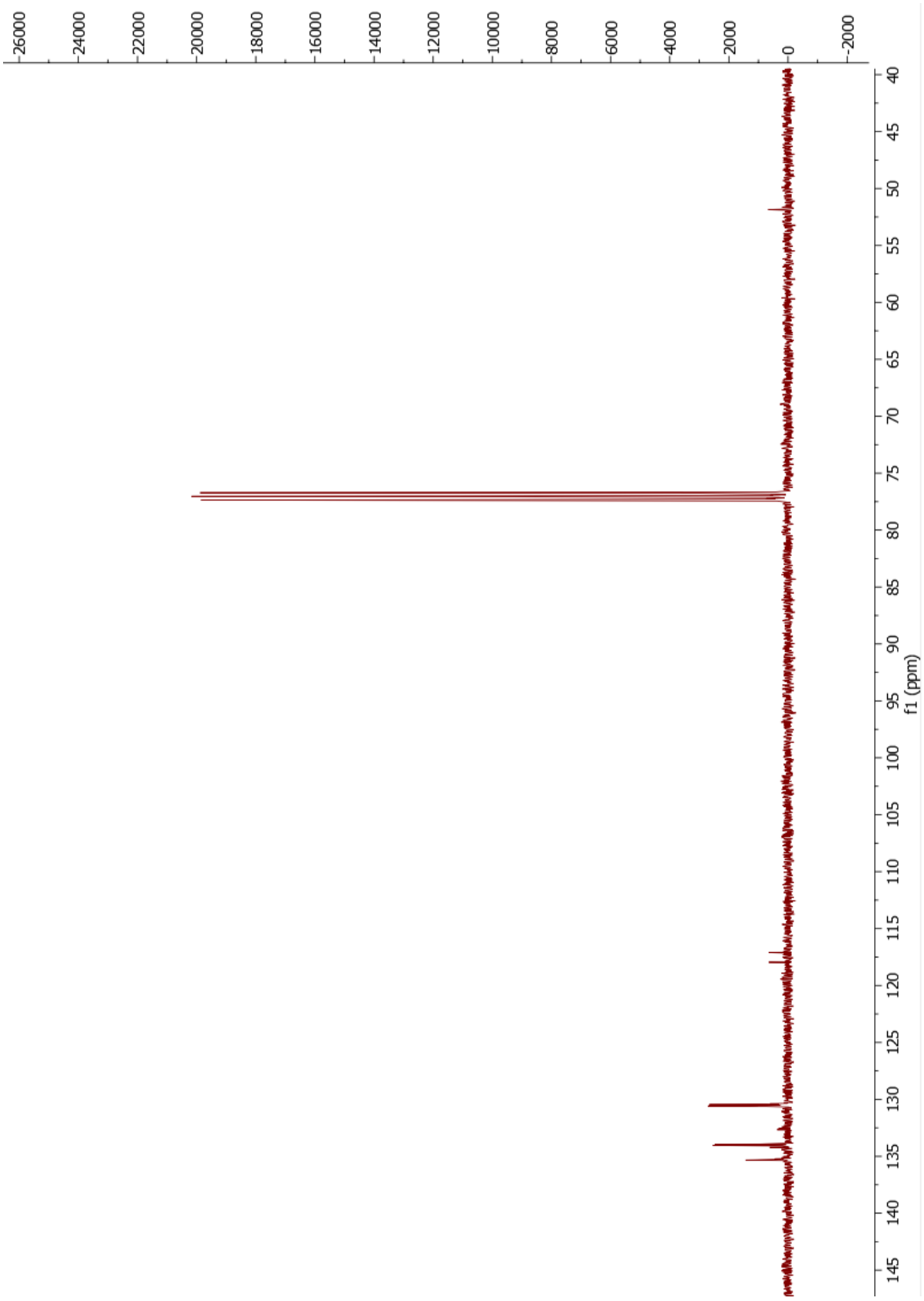


IR spectrum of compound **18**



^{1}H NMR of compound **21**

^{13}C NMR of compound **21**



IR spectrum of compound **21**