UNIVERSITY OF SOUTHAMPTON

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES

Department of Chemistry

Volume 1 of 1

Improving gene silencing oligonucleotides by incorporation of Peptide Nucleic Acids

by

Alexandre J. Debacker

Thesis for the degree of Doctor of Philosophy

September 2017

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES

Chemistry

Thesis for the degree of Doctor of Philosophy

IMPROVING GENE SILENCING OLIGONUCLEOTIDES BY INCORPORATION OF PEPTIDE NUCLEIC ACIDS

Alexandre Jean-Marie Debacker

The use of PNAs in therapeutics is limited by its mechanism of action. PNA (peptide nucleic acid) acts as steric blocker and therefore one copy per target of these therapeutic oligonucleotides is needed. While this mechanism is very interesting in splice-switching therapeutics it falls short of Ago2 or RNase H dependent gene silencing. Although the failure of PNA to recruit these enzymes to cleave their target could be a deal breaker, the high nuclease stability, neutral backbone and high affinity of PNAs are features that could enhance efficacy of siRNAs and antisense oligonucleotides (ASOs). First, the present work discussed the design, synthesis and properties of PNAs and LNA-modified oligonucleotides, which were used to switch off a silencing modified small non-coding RNA $MicA_{stab}$. Then, usage of PNAs in tandem with highly modified siRNA is discussed. The silencing activity of siRNAs containing PNA sense strand as RNA:PNA duplex was investigated. Association of PNA and siRNA was further studied and silencing activity and biophysical properties of PNA-Peptide carrier for siRNA delivery is shown. Finally, the optimisation of DNA-PNA chimeras was investigated. The synthesis of the monomers as well as the oligomerisation of LNA-DNA-PNA is described. The biophysical properties of chimeras and their ability to efficiently knock down *MALAT1* RNA in cells are shown in this thesis.

Table of Contents

List of Tables

List of Figures

List of Schemes

DECLARATION OF AUTHORSHIP

I, ALEXANDRE J. DEBACKER declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

IMPROVING GENE SILENCING OLIGONUCLEOTIDES BY INCORPORATION OF PEPTIDE NUCLEIC ACIDS

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission

Signed: .....

Date: .....

Acknowledgements

I would like to thank first and foremost Dr Jonathan Watts to have given the amazing opportunity to carry-out this PhD. The numerous advices and direction he provided me along the road as well as the trust and freedom he gave me to conduct my research have been a driving force throughout my studies. Thank you for the never ending support and positivity.

I would also like to thank ESPRC and the Marie Skłodowska-Curie Actions without whom none of this research could have happened; these funds gave us the opportunity to produce this work.

I thank the University of Southampton as well as UMASS Medical School for both providing an outstanding environment to carry out research and giving me the opportunity to meet so many amazing scientists and people.

Thanks to the whole ATD Bio team in Southampton and to Eugen Stulz for lending me their DNA synthesizers to begin my research, and even more importantly to share their invaluable experience with me.

Special thanks to my fellow lab members Hannah, Mike, Hassan, Vivek and Pranathi. Working with you was a blast.

Thank you to all the undergraduate students I had the immense pleasure to accompany in the lab: Dan, Jack, Rachel and Adam, every single one of you are amazing and working with you all taught me so much.

Particular thanks to all my collaborators. Thanks to Dr Charlotte Henderson, Dr Louise Butt and Dr Anastasia Callaghan from the University of Portsmouth to have given me the chance to collaborate on your very exciting projects. Thanks to Julia Altermann, Emily Knox and Professor Anastasia Khvorova from the RTI for sharing my excitement on our mutual projects and for the innumerable inputs on siRNA and science.

Last but not least I would like to thank my family and friends for your never-ending support.

Definitions and Abbreviations

°C	degrees Celsius
A:	adenine
aa:	amino acid
ACN:	acetonitrile
AcOH:	acetic acid
aeg:	*N*-(2-aminoethyl) glycine
aq:	aqueous phase (extractions)
ASO:	antisense oligonucleotide
B:	nucleobase
Bhoc:	benzhydryloxycarbonyl
Boc:	tert-butyloxycarbonyl
Bz:	benzoyl
C:	cytosine
Cbz:	benzyloxycarbonyl
$CDCl_3$:	deuterated chloroform
CPG:	controlled pore glass
d:	doublet
DBU:	1,8-diazabicyclo(5.4.0)undec-7-ene
DCM:	dichloromethane
dd:	doublet of doublets
DIC:	N,N'-diisopropylcarbodiimide
DIPEA:	diisopropylethylamine
DMF:	dimethylformamide

DMT:	bis(4-methoxyphenyl)phenylmethyl
DMAP:	dimethylaminopyridine
DNA:	deoxyribonucleic acid
eq:	equivalents
ESI-:	electrospray negative
ESI+:	electrospray positive
Et_2O:	diethyl ether
Et_3SiH:	triethylsilane
EtOAc:	ethylacetate
EtOH:	ethanol
Fmoc:	9-fluorenylmethyloxycarbonyl
G:	guanine
GalNAc:	N-acetylgalactosamine
HATU:	O-(7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HBTU:	O-(benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HOBt:	1-hydroxybenzotriazole
HPLC:	high-performance liquid chromatography
h:	hours
J:	coupling constant (Hz)
LRMS:	low resolution mass spectroscopy
m/z:	mass / charge ratio
M:	mole/L
m:	multiplet
min:	minutes

mmol:	millimole(s)
mp:	melting point
MMT:	monomethoxytrityl
MsCl:	methanesulfonyl chloride
NEt_3:	trimethylamine
NMM:	*N*-methylmorpholine
ppm:	parts per million
PyBOP:	benzotriazolyl-1-oxytris(dimethylamino)phosphonium hexafluorophosphate
PS:	phosphorothioate linkage
q:	quartet
qn:	quintet
R_f:	retardation factor
RNA:	ribonucleic acid
RP:	reverse phase
rt:	retention time (min) – HPLC
RT:	room temperature
s:	singlet
T:	Thymine
t:	triplet
T3P:	propylphosphonic anhydride
T_a:	temperature of annealing
TIC:	total ion chromatography
TEG:	tetraethylene glycol
TLC:	thin layer chromatography

T_m: melting temperature

δ: chemical shift (ppm)

Chapter 1: **Roles and structure of nucleic acids**

1.1 Discovery and structure of nucleic acids

While trying to determine the chemical building blocks of cells, Friedrich Miescher discovered what he named nuclein in 1869. This discovery was the first stepping stone toward what would become the biggest revolution in understanding the common element to all life forms. Later, during the last decades of the 19th century, Albrecht Kossel was able to identify and name the five nucleobases that would be recognised as the alphabet of genetic information (**Figure 1-1**).

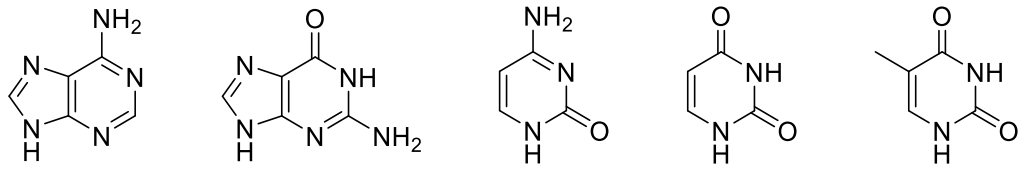


Figure 1-1: Naturally occurring nucleobases: purines (adenine and guanine) and pyrimidines (cytosine, uracil and thymine).

Since those foundational discoveries, everything we have learned about nucleic acids has not ceased to amaze. One of the most important milestones in molecular biology was the publication of the DNA structure elucidated by Watson and Crick in 1953 using X-ray data from the work of Rosalind Franklin, Maurice Wilkins and Raymond Gosling.[1]

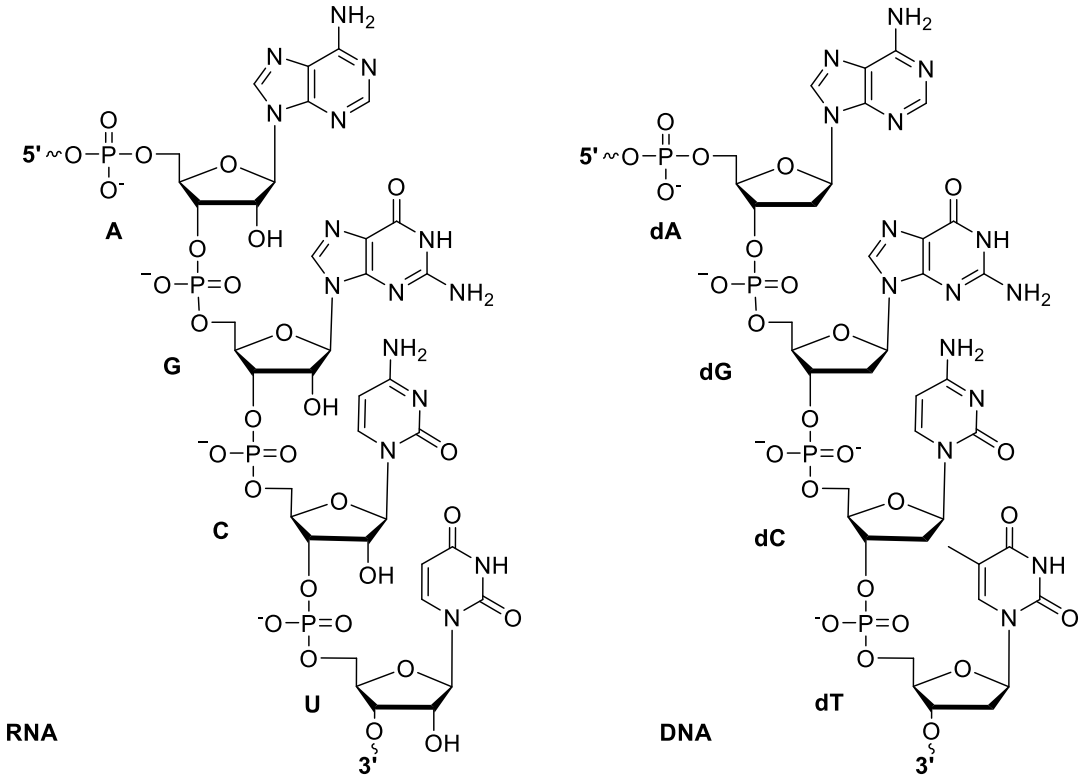


Figure 1-2: Primary structure of nucleic acids.

The primary structure (**Figure 1-2**) of nucleic acids is a sequence of ribose (RNA) or deoxyribose (DNA) sugars linked together at the 3’ and 5’ positions by phosphodiester groups. To this sugar-phosphate backbone, nucleobases are attached at the sugar 1' position.

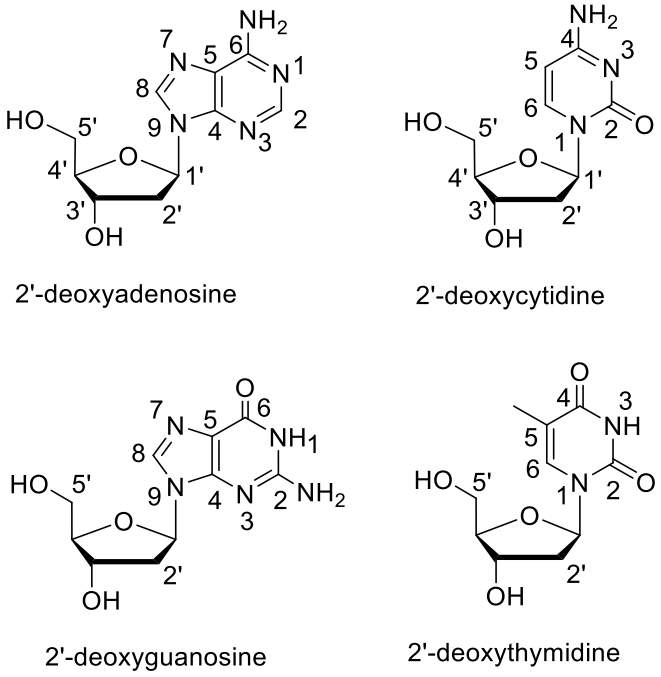


Figure 1-3: Numbering of deoxynucleotides.

Besides the presence or absence of a 2’ hydroxyl group, RNA and DNA differ by the fact that DNA normally contains thymine while it is replaced by uracil in RNA.

The presence of the 2’-hydroxyl group in RNA leads to an important structural and conformational difference compared to DNA. The sugar ring pucker of a nucleoside is described using two parameters; the pseudorotational phase angle P which is calculated from the endocyclic torsional angles, and the degree of maximum puckering which is the degree of torsion on the ring.

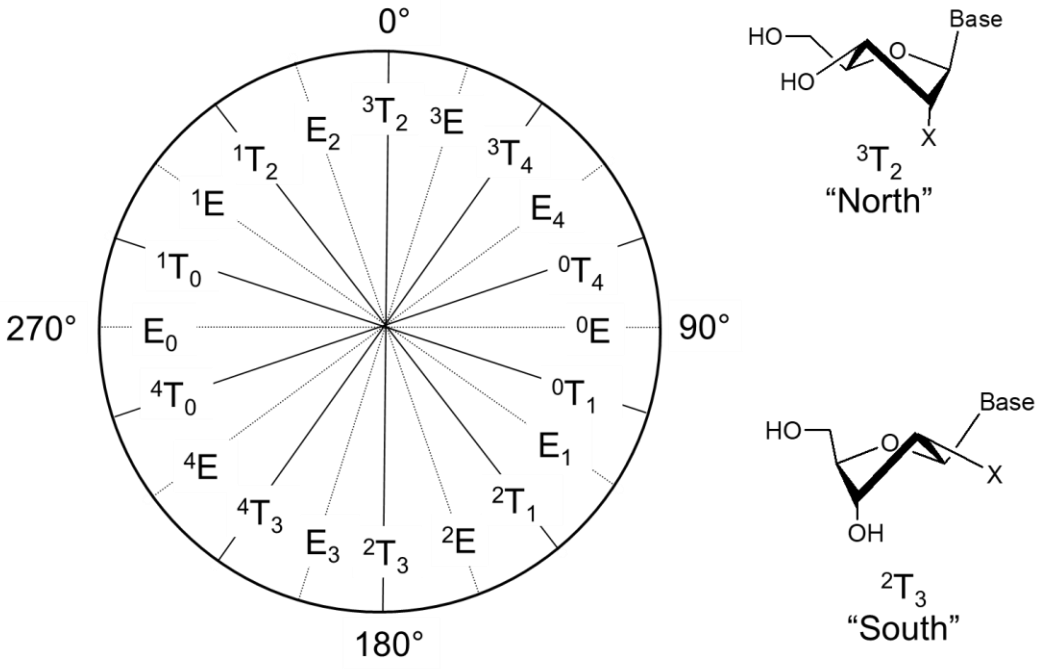


Figure 1-4: Pseudorotational wheel that describes the conformations of furanosides. T (twist) or E (envelope) furanose conformation.[2]

The two major conformations of unmodified DNA and RNA nucleotides are in the north (0°-36°) and the south (144°-180°), shown as grey zones in **Figure 1-4**. The absence of a 2’-hydroxyl group in DNA makes the conformational equilibrium favour the south configuration while the distribution of north/south in ribonucleosides is fairly even.[3]

The secondary structure of nucleic acids is driven by the nucleobases’ ability to bind each other in pairs[1]. Adenine forms two hydrogen bonds with thymine or uracil, and guanine forms three hydrogen bonds with cytosine (**Figure 1-5**). This base pairing drives nucleic acid strands to bind in an anti-parallel double-helical structure.

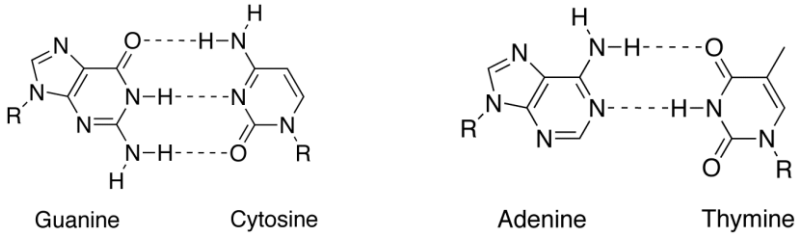
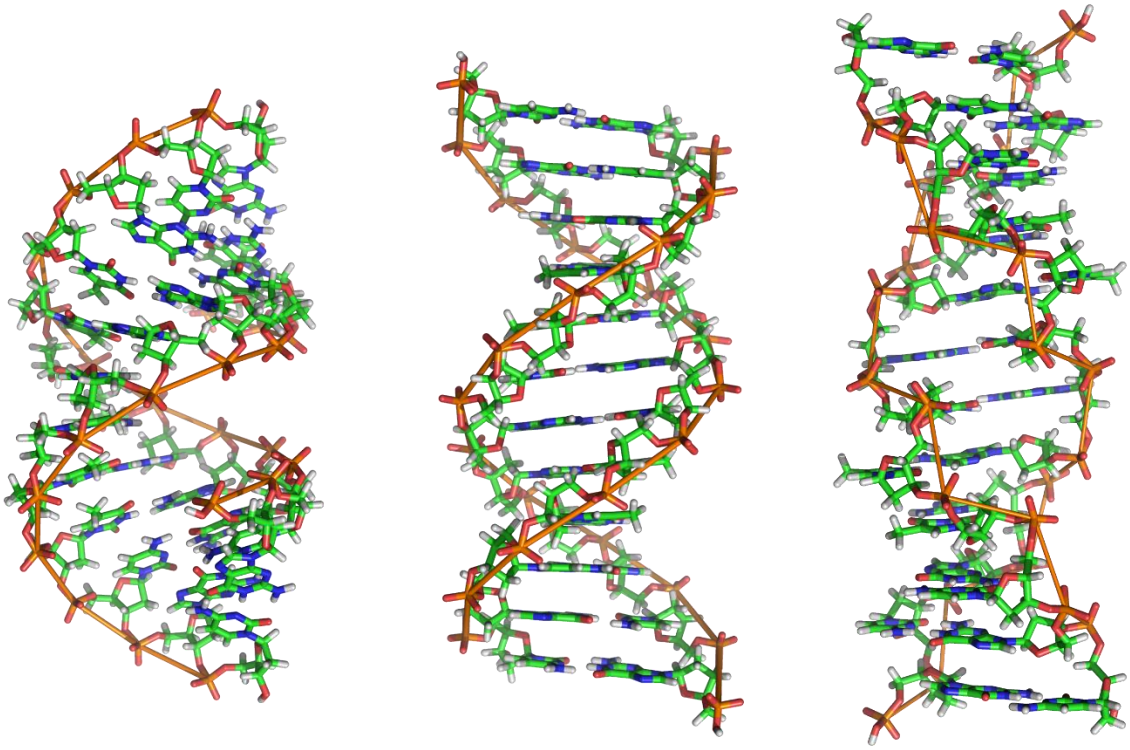


Figure 1-5: Watson and Crick base pairs, dashed lines are non-covalent hydrogen bonds.

The duplex can adopt three different helical forms: A-DNA, B-DNA and Z-DNA. A and B-DNA are both right-handed helices while Z-DNA is left handed. DNA exists predominantly in the B form inside cells, however it can adopt the A form during crystallographic experiments in conditions of low hydration and high salt. The RNA double-helix cannot adopt the B form and is nearly always found as an A form helix.

Both DNA and RNA can also be observed in the Z form. Z-DNA presents a unique structure where the sugars alternatively present radial and tangential positions. This form is rare but can be promoted by factors like an high GC content.[4]



Characteristic	A-DNA	B-DNA	Z-DNA
Helix sense	Right-handed	Right-handed	left-handed
Sugar pucker	C3'-endo	C2'-endo	C3'-endo(syn)
Bp/turn	11	10	12
Inclination bp to axis	+19 °	-1.2 °	-9 °
Rise per base pair	2.56 Å	3.3-3.4 Å	3.7 Å
Diameter	23 Å	20 Å	18 Å
Base tilt (τ°)	20 °	-6 °	-7 °
Major groove: width	2.7 Å	11.7 Å	8.8 Å

depth	13.5 Å	8.8 Å	3.7 Å
Minor groove: width	11.0 Å	5.7 Å	2.0 Å
depth	2.8 Å	7.5 Å	13.8 Å

Figure 1-6: Helical conformations of A-DNA, B-DNA and Z-DNA. Image taken from Wikimedia commons and used with permission, data from references.[4-7]

Inside cells, DNA is nearly always in duplex form. In contrast, RNA is commonly found in various structural forms in nature. RNA carries different functions in gene expression and metabolism, requiring different three-dimensional forms presenting structured recognition domains, for information transfer or enzymatic activity. The RNA secondary structure is driven by base pairing interactions and RNAs often form stem-loops or hairpins, but they can further assemble into complex three-dimensional structures like pseudoknots or the characteristic cloverleaf structure which gives tRNAs their biological functions.

DNA and RNA can also come together to form heteroduplexes, that are a very important feature in biology. They are formed when reverse transcriptase makes the DNA complement of viral RNA, when RNA polymerases transcribe DNA into mRNA and in primer regions during DNA replication. In biotechnology, they are found when a single stranded antisense DNA binds a complementary RNA target. The formation of this type of duplex is very important in gene regulation as it causes the cleavage of the bound RNA by the action of ribonuclease H (RNase H; the H stands for hybrid)[8, 9]. As the DNA and RNA strands of the heteroduplex have different conformations, the structure of the RNA:DNA duplex is different from DNA-DNA or RNA-RNA duplexes. They adopt an A-like conformation similar to DNA and RNA but the DNA sugar has an unusual east (O4'-endo) conformation.[10]

Soon after their discovery, nucleic acids were suspected of being a wide source of information. But the key to decrypt it was missing. The postulate that RNA was a template was proposed early by Dounce.[11] Since then the role of messenger RNA, transfer RNA and ribosomal RNA have been well established and the main mechanisms of the transcription of DNA into mRNA and the translation of mRNA into proteins are well understood.

1.2 Polynucleotides and their functions

1.2.1 DNA

DNA is found in every living cell, under various forms. DNA is most commonly found in the form of chromosomes in eukaryotes and prokaryotes. Every cell of a multicellular organism will have the same DNA sequences bearing the same genetic information. However, the DNA will not be shaped and read in the same way, allowing two cells with the same genetic material to have

different structures and functions. Although the DNA is a library of information and is generally somewhat inert as a molecule, DNA has a lot of interactions with proteins and RNAs.

In eukaryotes, the DNA is localised in the nucleus, where it is compacted by interactions with structural proteins including histones. The interactions of structural proteins with DNA are mainly governed by the positively charged amino acids of the protein which interact with the negatively charged phosphate of the DNA. These interactions are tuned by modification of theses amino acids: acetylation, methylation, and phosphorylation are all modifications that open or close a gene for reading and transcription, or inactivate it by increasing or decreasing the histone/DNA binding.[12, 13]

The other main types of proteins interacting with DNA are enzymes, including nucleases, polymerases and topoisomerases. Nucleases function is to cut polynucleotides by hydrolysis of the phosphodiester bond. Endonucleases are cutting in the middle of strands while exonucleases begin hydrolysing from the 5' or 3' end of a sequence. Ligases are, on the contrary, enzymes that catalyse the creation of phosphodiester bonds. Nucleases and ligases are important parts of the repair mechanism of DNA but also of all mechanisms of recombination. Recombination is a process by a gene is moved to a different locus of a same or different chromosome or plasmid. While DNA repair is the key for the safekeeping of genetic information, recombination is an important process of evolution. These two categories of enzymes are extensively used in molecular biology.

The last group of important enzymes interacting with DNA are the many enzymes taking part in the process of transcription. Transcription is the process by which the genetic code is translated into a phenotype. During the transcription and replication, the double stranded DNA is opened by topoisomerases and helicases to allow polymerases to synthesize a new strand of DNA or RNA.

1.2.2 Coding RNA

Messenger RNAs (mRNAs) are the most famous products of transcription of DNA. The mRNAs produced by transcription act as a vector to transport genetic information from the nuclear DNA to the protein factories of the cell, the ribosomes. In prokaryotic cells, the mRNA is active for translation immediately after transcription, but in eukaryotic cells the newly transcribed mRNAs (pre-mRNA) go through a process of maturation before being able to be translated.

In this maturation process, the pre-mRNA will be first capped on its 5', the 5'-cap being a 7-methylguanosine which is linked to the 5' of the mRNA by a 5'-5' triphosphate linkage. The 3' end of the pre-mRNA is then processed with the addition of a poly adenosine tail (poly(A) tail) by polyadenylate polymerase. The poly(A) tail serves to protect the mRNA from exonucleases and

binds translation factors. During the process of maturation, the pre-mRNA is also spliced. The pre-mRNA is made of introns and exons: the exons are the part of the sequence that will be carried on into the mature mRNA while the introns are the portions of the pre-mRNA that are spliced out and degraded during formation of the mature mRNA.

After processing, the mRNA is recognized as mature and transported into the cytoplasm. The coding mRNA can then be translated by the ribosome. The genetic code is based on triplet codons: Each codon is formed of a sequence of three nucleotides therefore $4^3 = 64$ different codons exist, which code for the 21 natural amino acids. Several codons can code for one amino acid and some particular codons will initiate or terminate translation (called start and stop codons, respectively).

The ribosome is a complex cellular machine formed of ribosomal RNAs (rRNA) and dozens of proteins. rRNAs are specialized RNA which have an enzymatic activity. Translation occurs as follows: first, the small ribosome subunit attaches to the 5' end of the mRNA and then it moves across the mRNA until it finds a start codon. There a transfer RNA (tRNA) and the large ribosomal subunit join to form the ribosome. The tRNA has a complementary triplet called the anticodon which is complementary to the codon on the mRNA; it transports the amino acid corresponding to this codon, which is attached to the tRNA by an ester bond. At that point a second tRNA associates with the following codon, and the ribosome catalyses the formation of a peptide bond between the amino acids on the two tRNAs. The first tRNA is then released and a third tRNA complementary to the third codon is bound; the growing polypeptide chain is transferred to each incoming aminoacyl tRNA in turn. When the ribosome encounters a stop codon, it releases the polypeptide. The polypeptide will fold into a complex three-dimensional structure (often with the help of chaperone proteins) to form a functional protein.

1.2.3 Non-Coding RNA

Although in the early days of nucleic acids research the community thought that most of the RNA transcribed was coding for protein, it has since been established that only 1.5% of the human genome is actually coding for protein.[14] Nevertheless 60 to 90% of the genome is transcribed, and many genes are also transcribed into both sense and antisense transcripts.[15, 16] All these transcribed RNAs have multiple roles and functions and while most of it were until recently considered non-active and superfluous the aforementioned rRNAs and tRNAs involved in translation taught us than RNAs can have enzymatic properties and play a different role than just transport information from the DNA to protein. The discovery of catalytic RNA was rewarded with the 1989 Nobel Prize in Chemistry to Tom Cech (University of Colorado) and Sidney Altman (Yale).

Transcripts that do not code for protein are called non-coding RNA (ncRNAs). Besides the historic classes of noncoding RNA (transfer RNAs and ribosomal RNA), multiple classes of ncRNAs have been identified more recently, including small nucleolar RNAs (snoRNAs), microRNAs (miRNAs), small nuclear RNA (snRNAs), enhancer RNAs (eRNAs), natural antisense transcripts (NATs) and long non-coding RNAs (lncRNAs). Some of the ncRNAs have well defined functions but for many of them the details of their mechanisms are still unknown.[17] However, many ncRNAs seem to be involved in the regulation of transcription and translation; as such, they are found to interact with the promoter region of genes or with the untranslated regions of mRNAs, recruiting proteins or otherwise inhibiting transcription and translation. They are therefore of interest in the therapeutic regulation of gene expression: either through mimicking the roles of regulatory ncRNAs with modified oligonucleotide drugs or through inhibiting their effects.

1.3 RNA interference: siRNA and the RISC machinery

The different mechanisms of regulation induced by non-coding RNAs have been a source of inspiration for the development of new aspects of the oligonucleotide therapeutics toolbox, and vice versa. The story of RNAi interference is an example that went both ways: an advance in technology led to a greater understanding of endogenous mechanisms, and exploration of the endogenous mechanisms continues to inform progress in the field. In 1998, Fire and Mello observed that double-stranded RNA induced gene silencing more potently than single-stranded RNA (either sense or antisense).[18] This surprising observation earned them the 2006 Nobel Prize in Physiology or Medicine, but it also led to the discovery of the cellular machinery called the RISC (RNA-induced silencing complex). The RISC is the engine of the RNA interference pathway (RNAi) and uses small RNA guides, both endogenous, microRNAs (miRNA) or exogenous small interfering RNA (siRNA).

The RNAi is triggered when the ribonuclease Dicer cuts a double stranded or hairpin RNA to form an RNA duplex of 21-25 nucleotides long, with 3'-overhangs of two nucleotides. This duplex is then loaded into the RISC, aided by Dicer and R2D2. These two proteins are core of the RLC (RISC loading complex) which will unite the duplex and Ago2 to form the RISC. The duplex is then unwound and the sense strand is released. The RISC will then bind a RNA that is complementary to the antisense strand that is still loaded in the RISC. When the RISC finds a fully complementary mRNA, the Ago2 catalytic domain cleaves a phosphodiester bond of the mRNA opening the path for nucleases to degrade the rest of the mRNA. This process is catalytic and after cleaving its mRNA target the RISC will stay complexed and bind to a new copy of the mRNA target. In animal if a miRNA is loaded in RISC, it will be able to bind a non-fully matched complementary sequence,

generally on the 3' untranslated region (3'-UTR) of the mRNA target. When binding to the 3'-UTR the RISC will stay bound and prevent the ribosome to bind to the mRNA and initiate translation.

Nowadays siRNA and miRNAs are used routinely in chemical biology to study gene expression and a lot of resources are invested to make drugs out of this technology.[19-21] The RISC complex is able to support a lot of chemical modifications made on the siRNA.[22-24] The development of chemistry has allowed the siRNA which when unmodified is degraded in minutes in the bloodstream[25] to became stable in vivo. Single stranded siRNAs[26, 27] which are made of only the antisense strand as well as asymmetric siRNAs which have a shorter sense strand are both demonstrated to be taken up by RISC and inducing silencing of a targeted mRNA.[28, 29]

1.4 RNase H

RNase H is a ubiquitous enzyme found in eukaryotes, prokaryotes, and even some viruses. RNase H is a nuclease which cleaves the RNA strand of DNA/RNA duplexes.[30] The RNase H hydrolyses the 3'-O-P bond of RNA in DNA/RNA duplexes generating a 3'-OH and a 5'-Phosphate terminus. This enzyme is involved in DNA replication and could be responsible for cleaving the RNA primers used in DNA replication.[31] It also plays a role in genome defence, including the common task of removing R-loops (regions where an RNA strand displaces one of the strands of DNA).[32]

The most studied prokaryotic RNase H is the *E. coli* RNase H1, for which the active site and the substrate/enzyme interaction have been thoroughly studied. The RNase H active site usually contains three negatively charged amino acids, two aspartic acids and one glutamic acid. The two aspartic acids are chelating the divalent cation while the glutamic acid activates a molecule of water.[33] Moreover, the *E. coli* RNase H1 active site shares strong homology with the human RNase H1.[34] Despite the similarity in the catalytic site, the cleavage pattern of the *E. coli* RNase H1 is very different than the human enzyme (**Figure 1-7**).

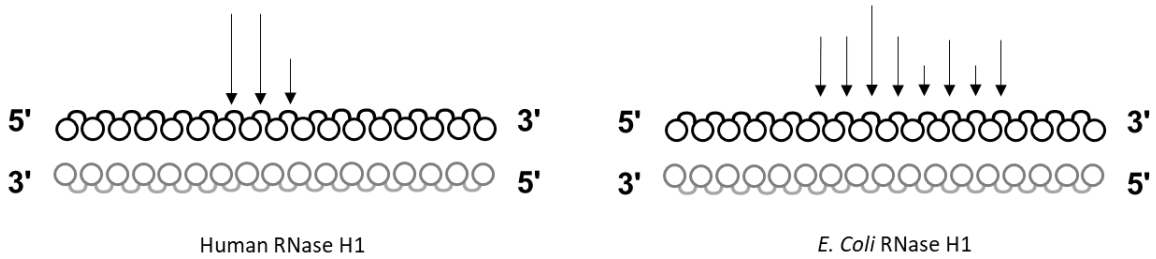


Figure 1-7: Cleavage pattern of Human and E. coli RNase H1.[33, 35]The arrows show the sites of cleavage, the length of the arrow is proportional to the amount of cleavage observed in one site.

The difference in cleavage is due to the presence in the human RNase H of a RNA binding domain. The RNA binding domain recognises the first base from the 5' of the RNA of a RNA/DNA duplex

(**Figure 1-8**). However, the cleavage pattern for the *E. Coli* is to some extent sequence dependent and can present a more specific profile.[36] The human RNase H1 possesses a spacer between the RNA binding domain and the catalytic domain, inducing the specific cleavage pattern of the human RNase H1. The RNA binding domain is very specific to DNA/RNA duplexes: in case of gapmer ASOs (**Figure 1-8**) the RNA binding domain will not bind to 2'-MOE/RNA, 2'-OMe/RNA, or LNA/RNA base pairs (modification described in section 1.6).[37]

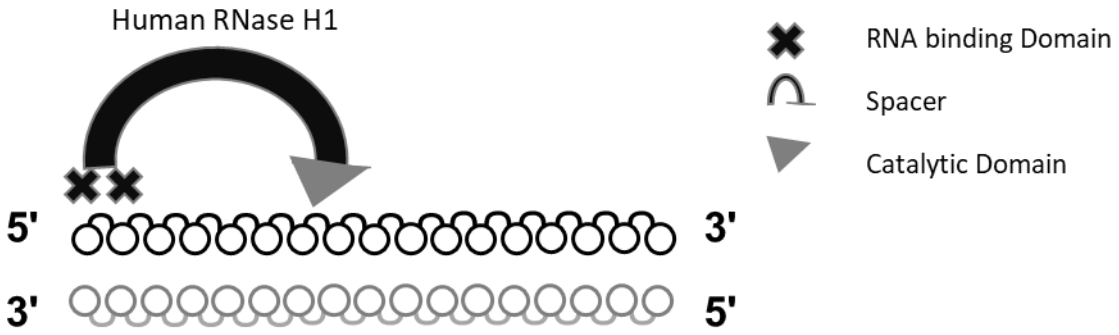


Figure 1-8: Schematic of Human RNase H1 binding a RNA/DNA duplex.[35]

ASO have been developed to recruit RNase H, they are modified "DNA like" oligonucleotides (including phosphorothioate DNA and gapmer designs, described in section 1.6). ASO that recruit RNase H are catalytic: the ASO is not cleaved by RNase H and it is released by the enzyme after cleavage of the target RNA. The ASO can thus bind and destroy other RNAs after the first RNA target has been cleaved. Several examples of this technology are being investigated in clinical trials[38] and two RNase H-based ASOs have been approved by the US Food and Drug Administration, fomivirsen (approved in August 1998 for the treatment of AIDS-related cytomegaloviral eye infections, then removed from market in 2005 because the appearance of antiretroviral HIV treatment rendered the drug nearly obsolete), and mipomersen (approved in January 2013 for the treatment of familial hypercholesterolemia).

1.5 Steric blocker ASOs: splice switching, miRNA inhibition and more

Some ASOs are not designed to recruit any cellular machinery to achieve gene regulation. This type of ASO is designed to bind its DNA or RNA target with high affinity, to prevent the target from interacting with other proteins or nucleic acids. These ASOs can engage in multiple mechanisms: for example, they can bind the promoter region of genes to block transcription,[39-41] or bind to the translation initiation site within a mRNA to prevent translation.[42, 43] However, steric blocker ASOs have been mainly studied in two other mechanisms: splice switching and miRNA inhibition (anti-miR).

Splice switching ASOs target a splice signal or exon-intron junction within pre-mRNA in order to mask the intron from the splicesome which cannot splice it out of the mRNA. This is particularly interesting for proteins which are misfolded due to a mutation in a particular intron but that keep all or part of their activity when the intron that is containing the mutation is not translated. This class of oligonucleotides has now achieved its first clinical success, as two splice-switching oligonucleotides received FDA approval in 2016: eteplirsen, a morpholino oligonucleotide (PMO) that triggers the exclusion of a mutant exon within dystrophin mRNA for the treatment of Duchene muscular dystrophy,[44] and nusinersen, a PS-MOE oligonucleotide that modulates the splicing of SMN2 mRNA to treat spinal muscular atrophy.[45]

Steric blockers can also target microRNAs: these ASOs are called anti-miRs.[46] MicroRNAs are small non-coding RNAs that intervene in the regulation of translation. Anti-miRs are usually short oligonucleotides heavily modified with LNA, 2'-F-RNA, and 2'-OMe.[47]

1.6 Chemical modifications of oligonucleotides

Synthetic DNA or RNA, both single stranded and double-stranded, is very unstable in cells and *in-vivo*, because of the high concentration of nucleases in serum, cytosol and endosomes. Furthermore, the delivery of naked oligonucleotides is poor. Naked, unmodified oligonucleotides do not cross the cell membrane.[48] Finally, unmodified oligonucleotides exert innate immune responses that can be problematic to patients.[49] In order to make oligonucleotides active as research tools and therapeutics, a large number of ingenious chemical modifications have been developed.[50]

The chemical modification of oligonucleotides has been focused along three main axes: sugar modifications, phosphate modifications and conjugates.

1.6.1 Sugar-modified nucleotides

The ribose sugar can be modified in many ways.[51] This section will focus on two major classes of sugar-modified nucleotides: those with 2'-modifications and those in which the 2' and 4' positions are linked to create a bicyclic sugar.

Functionalising or replacing the 2'-OH group often increases the binding affinity of the modified oligonucleotide while strongly enhancing the nuclease-stability of the oligonucleotides. For example, in unmodified RNA, the 2'-OH is involved in both base-catalysed and nuclease-catalysed hydrolysis. When the stereochemistry is intact at the 2'-position, the ribose maintains the North conformation of the parent RNA.

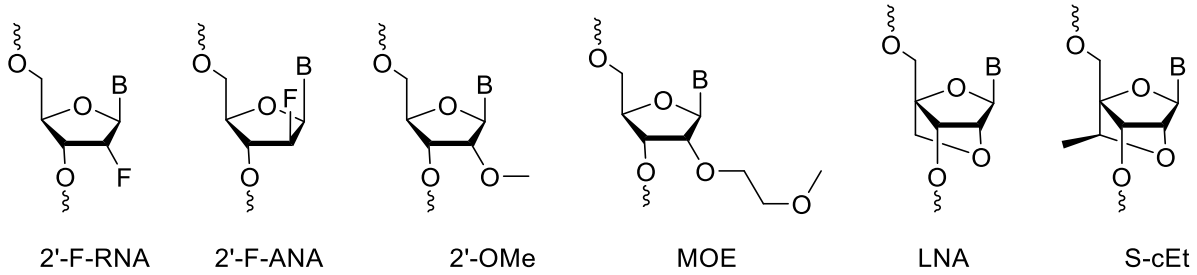


Figure 1-9: Important 2'-sugar modifications and 2', 4' bridged sugar modifications.

The 2'-*O*-alkyl modification is the substitution of 2'-hydroxyl of RNA by an alkylether, which leads to increased stability of modified oligonucleotides against nucleases. For example, the naturally-occurring 2'-OMe modification improves binding affinity by ~1.3°C per modified residue.[52] A large series of additional 2'-alkyl, alkylamino and alkoxyalkyl modifications have been developed, of which 2'-methoxyethyl (MOE) is the most widely used (within two FDA-approved drugs). It improves both binding affinity (increasing the T_m by about +2°C per modified nucleotide) and stability against nucleases.[53]

Substitution of the 2'-hydroxyl group by a fluorine gives 2'-F-RNA, which possesses strong binding affinity for its complementary strand (T_m +1.8°C per residue).[52]

Inverting the stereochemistry of the 2'-fluorine yields 2'-fluoroarabinonucleic acid (2'-F-ANA), a nucleotide with a predominantly south "DNA like" sugar conformation which gives fully-2'-F-ANA-modified oligonucleotides the ability to recruit RNase H.[54]

Another interesting family of sugar modifications involves the formation of a bridge between the 2'-oxygen and the 4' carbon. This bridge locks the nucleoside in an RNA-like conformation[55] which gives it very high affinity for complementary strands. Thus Locked Nucleic Acid (LNA) shows very high binding affinity (+1 to +8°C per base[56]) and good nuclease stability.[57] (*S*)-constrained ethyl

(cEt), a methylated derivative of the LNA nucleotide, was developed by Isis Pharmaceuticals shows a 5 fold increase in potency compared to a 2'-MOE gapmer.[58]

With the exception of 2'F-ANA, these 2'-modified nucleotides are all RNA-like, and none of them are able to recruit RNase H to cleave their target RNA. However, they can be combined with DNA into chimeric strands containing a region of DNA flanked by high affinity RNA-like modified nucleotides.[59] The central 'gap' of DNA has given the name 'gapmer' to this class of oligonucleotides. LNA and MOE have been extensively used in gapmer ASOs, often using roughly 3-9-3 and 5-9-5 patterns respectively, 9 being the length of the DNA gap while 3 and 5 are the length of the modified wings.

These 2'-modifications are also widely used in steric blocker applications. For example, LNA-2'-OMe chimeric oligonucleotides have been reported as a particularly potent design for anti-miRs.[60]

Finally, an alternating pattern of 2'-OMe and 2'-F-RNA has been used with great success in siRNA designs. The introduction of these modifications into siRNA duplexes has been a very important milestone in terms of the ability of siRNA to be active in vivo.[22, 23, 61, 62]

1.6.2 Backbone modifications

A major focus for chemical modification has been the phosphodiester linkage, since most of the degradation of oligonucleotides is happening by hydrolysis of the phosphodiester linkage.

One of the first backbone modifications was the substitution of one of the non-bridging oxygen atoms of the phosphate by sulphur to create a phosphorothioate (PS) linkage (**Figure 1-10**).[63, 64] This modification is still the most widely used today due to its straightforward and affordable synthesis as well as three attractive properties. Firstly, PS linkages possess very high resistance to nucleases which increases their half-life in cells and tissues. Secondly, PS-modified oligonucleotides show significantly higher binding to plasma proteins and cell-surface trafficking proteins, which increases their plasma half-life and cellular uptake[65] and improves their intracellular trafficking.[66] Thirdly, PS-modified oligonucleotides are compatible with RNase H and RISC machineries. Fully-PS-modified DNA are able to elicit RNase H cleavage of complementary RNA.[67] Moreover, fully modified PS siRNAs are not as efficacious as partially modified ones, but partially PS siRNAs are much better in vivo than siRNAs with unmodified backbones. These three properties and the relative ease of synthesis in standard oligonucleotide synthesizers make PS backbone one of the most useful tools in RNA silencing.

In spite of these three advantages, PS modifications also have two significant disadvantages: PS-modified oligonucleotides have reduced affinity for their targets,[68] and the increased protein binding of PS oligonucleotides can also lead to toxicity when undesired proteins are bound.[69]

A chiral centre is created on the phosphorus[65] at the sulfurization step of the oligonucleotide synthesis. In most PS oligonucleotide synthesis, each phosphorothioate linkage is created as a random mixture of diastereomers. The affinity of each diastereoisomer has been studied and the [all-*R*p]-PS-DNA has a higher affinity for its target than the [all-*S*p]-PS-DNA.[70, 71] Wave Life Sciences recently published a robust method of synthesis of stereopure PS-oligonucleotides adaptable to pharmaceutical scale. They show that careful control of the stereochemistry of the ASO, and the use of specific patterns of *S*p and *R*p linkages, can substantially improve the properties of PS ASOs.[72]

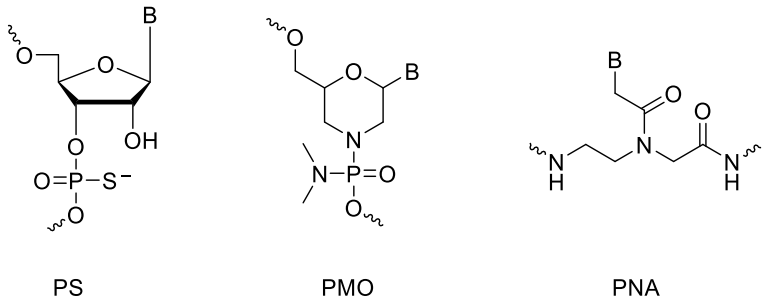


Figure 1-10: Main oligonucleotide backbone modifications.

In parallel with the modification of the phosphodiester linkage, some oligonucleotides have been designed with a totally different backbone. The two main examples are phosphorodiamidate morpholino oligomers (PMO) and peptide nucleic acid (PNA) (**Figure 1-10**). These two designs were developed in order to obtain a neutral oligomer which in theory should have better permeability across cell membranes and higher binding affinity with their target than wild type (polyanionic) oligonucleotides. PMO are chemically modified oligomers where the ribose is replaced by a morpholine pseudosugar and the charged phosphate linkages are replaced with by neutral phosphorodiamidate linkages.[73]

PNA is radically different than both RNA and DNA and will be discussed in detail in Chapter 2 of this thesis.

Both PNA and PMO have been investigated as steric blocker ASOs: they are able to inhibit transcription,[74] regulate translation,[75-77] inhibit miRNAs[78] and affect splice switching.[79]

1.6.3 Oligonucleotide conjugates

Naked ASO and siRNA have different delivery profiles. While single stranded ASO are taken up into cells[80] by endocytosis and are able to slowly cross the lipid bilayer of the endosome membrane to enter the cytosol,[66, 81, 82] double stranded siRNA stay trapped in the endosome without means to escape. Attaching a conjugate to the oligonucleotide has been a main focus of the research to improve delivery and one of the most promising.[48]

One of the most exciting conjugates is N-acetylgalactosamine (GalNAc) which has shown excellent results for delivery of both siRNAs[83] and ASOs[84] to hepatocytes. Ten GalNAc conjugates, based on various classes of oligonucleotides, have entered clinical trials.[48]

Conjugated peptides for ASO and siRNA delivery have both been studied but face challenging synthetic obstacles. Nevertheless, several peptide conjugates that enhance ASO delivery in eukaryotic and prokaryotic cells have been identified for PNA and PMO. Hydrophobic conjugates of oligonucleotides (*e.g.* conjugated cholesterol or fatty acids) are also currently being studied and show improved uptake and activity in cells. However, there is an unmet need for conjugates with comparable potency to GalNAc for other target tissues than the liver.

1.7 Synthesis of oligonucleotides

Oligonucleotides are routinely synthesized on solid support with an automated synthesizer.[85, 86] The use of solid support allows numerous advantages over liquid phase chemistry: it allows the use of large excess of reagents to ensure fast completion of every step, it allows simple washing between each step to get rid of impurities instead of complicated purification, and it lends itself readily to automation.

The most widely used method for oligonucleotides synthesis is the phosphoramidite method developed by Marvin Caruthers (University of Colorado).[85, 87] Each successive nucleotide is added by a cycle of four chemical reactions (**Figure 1-11**). First, the protected 5'-OH of the support-bound nucleotide is deprotected. In the second step, the desired nucleotide phosphoramidite is added in the presence of a mildly acidic activator and the coupling takes place to form a phosphite triester. The third step is the capping of any sites that failed to couple, which facilitates purification of the full-length product. The fourth step is the oxidation or sulfurization of the successful coupled phosphite triester to a phosphate (or phosphorothioate) triester. Then the cycle starts again with the deprotection of the nucleotide.

As described, the phosphoramidite method generally grows the oligonucleotide in the 3' to 5' direction, though the opposite can be achieved by varying the monomer structure. The solid

phase used is usually a controlled pore glass (CPG) or polystyrene resin functionalized via a succinic acid linker with the first nucleotide of the sequence. After the synthesis, the oligonucleotide is both cleaved from the resin and deprotected by an amine solution. Concentrated aqueous ammonia is commonly used but, when compatible with the chemistry of the oligonucleotide, a 1:1 mixture of concentrated aqueous ammonia and 40% $MeNH_{2\,(aq)}$ is preferred.

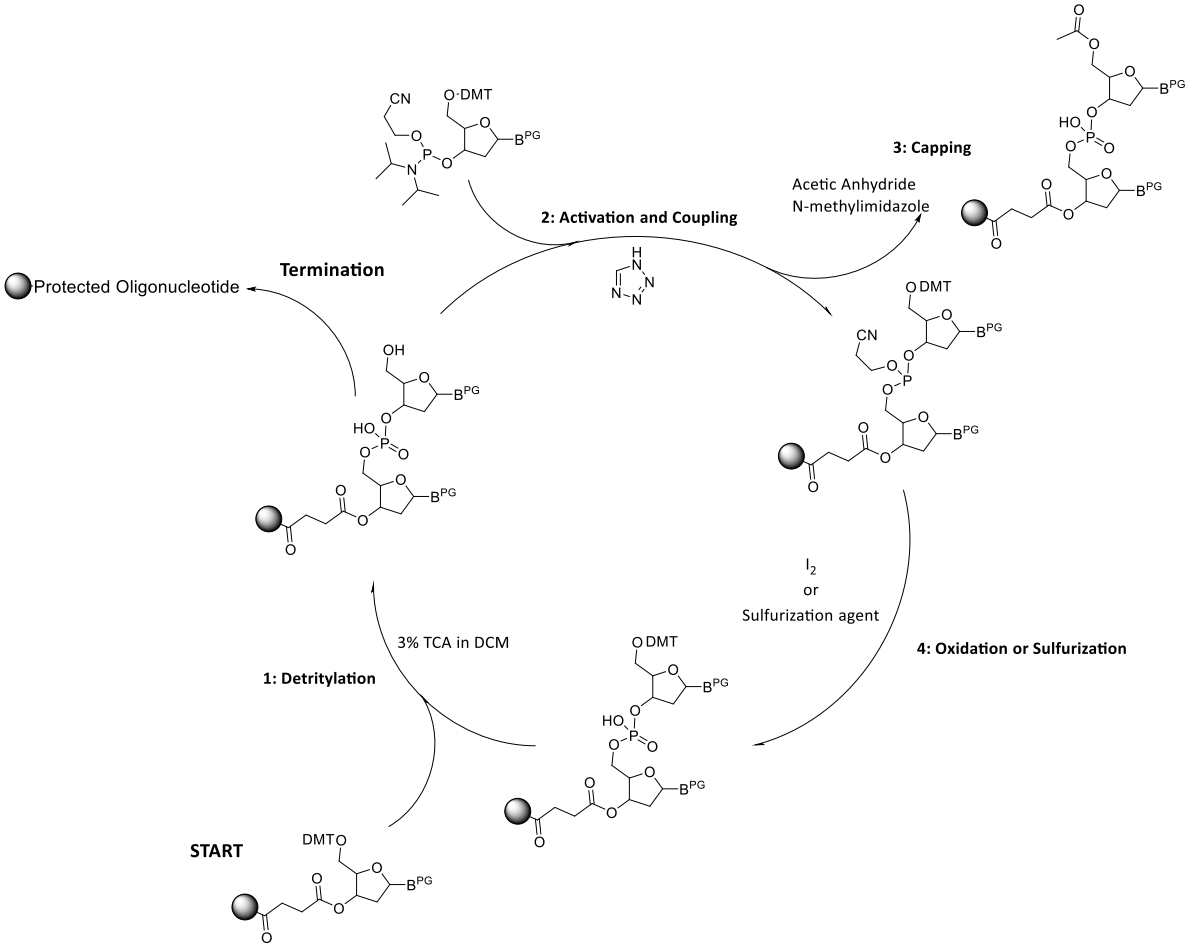


Figure 1-11: Synthesis cycle of solid-phase phosphoramidite-based oligonucleotide synthesis.

Oligonucleotide purification from other macromolecules (such as failure sequences generated by incomplete coupling) is usually carried out by HPLC or, less commonly, by preparative denaturing PAGE. For removing small molecule impurities, size-exclusion chromatography on Sephadex or ethanol precipitation is commonly used.

1.8 Aim of the thesis

This project is focusing on studying how PNA can be applied to improve existing oligonucleotide therapeutics.

The second chapter of this thesis will focus on unmodified PNA around three axes. The first part describes the use of a PNA strand as an oligoswitch for gene regulation. As such, the effect of PNAs to reverse the effects of an engineered unnatural small regulatory RNA is studied in comparison with LNA-modified ASOs. In a second part, the compatibility of the PNA chemistry with siRNAs is investigated. Specifically, we present the properties and efficacy of a fully chemically modified siRNA containing a PNA sense strand. In a third section we describe the design, properties and efficacy of PNAs able to be used as peptide carriers for siRNA delivery.

The third chapter of this thesis is focused on the development of gapmer ASOs containing a PNA wing. We first discuss the design and rationale for these chimeric ASOs, then present the synthesis of the custom monomers needed for their synthesis. Following this, we present the solid phase synthesis unique to this class of constructs.

Finally, in Chapter 4, we discuss the biophysical properties and gene silencing efficacy in cells of these chimeras.

Lastly, Chapter 5 concludes and point to future directions in these projects and Chapter 6 gives the details of the experimental methods we used for all aspects of the work.

Chapter 2: **New usages of peptide nucleic acids**

2.1 Background on peptide nucleic acids

2.1.1 Discovery and structure

Peptide nucleic acids (PNAs) are oligonucleotide mimics with a radically altered structure. By attaching nucleobases onto an aminoethylglycine (aeg) backbone, Buchardt, Nielsen, Egholm and Berg[88-90] formed a strand able to hybridize with DNA and RNA, creating a hybrid molecule between peptide and oligonucleotide that possessed properties from both but was in a class of its own. PNA can bind to any complementary oligonucleotide following the Watson-Crick rules. PNA:PNA duplexes have their own helical structure (which has been called P form[91]), which is distinct from DNA or RNA and somewhat more flexible. When hybridized with DNA or RNA, PNAs adopt a structure depending on the nature of the complementary strand, an A-like form for PNA:RNA and B-like form for PNA:DNA.[92]

PNA monomers are made of a non–sugar backbone on which the nucleobase is attached (**Figure 2-1**). PNA monomers are assembled by peptide bond formation and PNA sequences are conventionally written like peptides, from N-terminus to C-terminus. Conveniently, this also aligns with the conventions of oligonucleotide notation (which are written 5' to 3'), since for the purposes of parallel / antiparallel duplex nomenclature, the N-terminus is considered equivalent to the 5'-end of an oligonucleotide. Although PNAs are able to bind to their complementary sequence of PNA, DNA or RNA in both parallel and anti-parallel fashion, the anti-parallel duplex is favored.

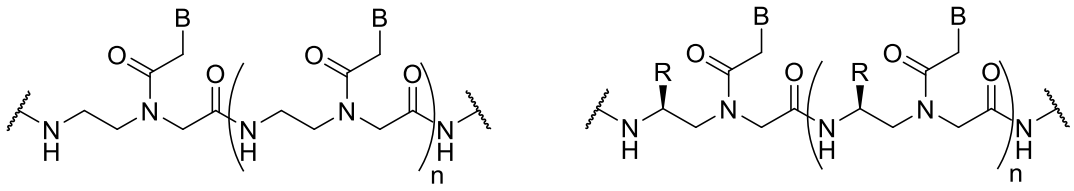


Figure 2-1: Standard aeg-PNA oligomer structure (left) and L-γ-PNA structure (right).

Over the years, many variations of the structure of PNAs have been investigated: in a recent review Amwasti *et al.*[93] identified 27 backbone modifications in addition to the standard aeg backbone and several other base modifications. Two of the main focuses of the backbone modification of PNA were creation of chirality and addition of features to raise the solubility of PNA. The L-γ-PNA has been a chiral PNA investigated thoroughly[94, 95] and the gamma position was

also used to add hydrophilicity modifier such as mini-PEG.[96, 97] Nonetheless, the aeg backbone is still the most widely used due to its greater availability.

2.1.2 Properties and applications of PNAs

As PNAs are neither oligonucleotides nor peptides, they are neither degraded by nucleases nor proteases.[98] PNA:PNA duplexes shows very high affinity to each other (T_m ~75°C for a 14mer) and PNA:RNA and PNA:DNA duplexes have better binding affinity than their RNA:RNA and DNA:DNA counterparts. Generally each PNA monomer will raise the melting temperature by 1°C compared to DNA or RNA.[99] PNA has been shown to strand-invade duplexes and bind structured targets more rapidly and effectively than charged-backbone oligonucleotides.[100-102]

The high enzyme and serum stability[98] of PNAs positioned them early as candidates for therapeutics. However, PNA:RNA duplexes are not able to recruit RNase H and are neither able to be loaded in RISC as antisense. Therefore, PNAs have been investigated for aspects of the regulation of gene expression that do not require the recruitment of a nuclease. This includes miRNA inhibition,[103] splice switching,[104] and inhibition of transcription[40] or translation.[105] Currently, PNA ASOs are actively being investigated for their potential applications in regulation of gene expression. The investigation of PNA as steric blocker ASOs shown an ability at stopping the transcription[74] and regulating the translation.[75-77] In particular, PNA-based inhibitors of translation have been widely studied as antibacterial drugs and for deactivation of parasites.[106]

Another major application of PNA in biological research is fluorescence *in-situ* hybridization (FISH). PNA technology has been used for more than a decade in routine FISH experiments with success.[107, 108] Additional applications of PNA technology in the area of diagnostics include applications in microarray approaches[109-111] and as PCR clamps (small PNA targeting a small region of RNA to block its amplification during PCR to identify specific sequences or mutation at the femtomole level in a sample).[112, 113]

2.1.3 Synthesis of PNA

Several synthetic strategies for PNA monomers have been described.[114-116] Most of the synthetic methods for PNA oligomers are made on solid-phase and four main strategies can be identified. These strategies are based on the protective group strategy chosen for the monomers (**Figure 2-2**). Historically the first PNAs were synthesized with Boc/Bz strategy on Merrifield resin,[88] then Fmoc/Bhoc, MMT/Acyl and Dde/MMT strategies were developed. The most common strategy is using Fmoc/Bhoc protection and polystyrene resin on a peptide synthesizer. In 2010 Fabani *et*

al.[115] reported high yield and purity of relatively long PNA monomers using this chemistry on a microwave peptide synthesizer using Rink-Amide ChemMatrix solid support.

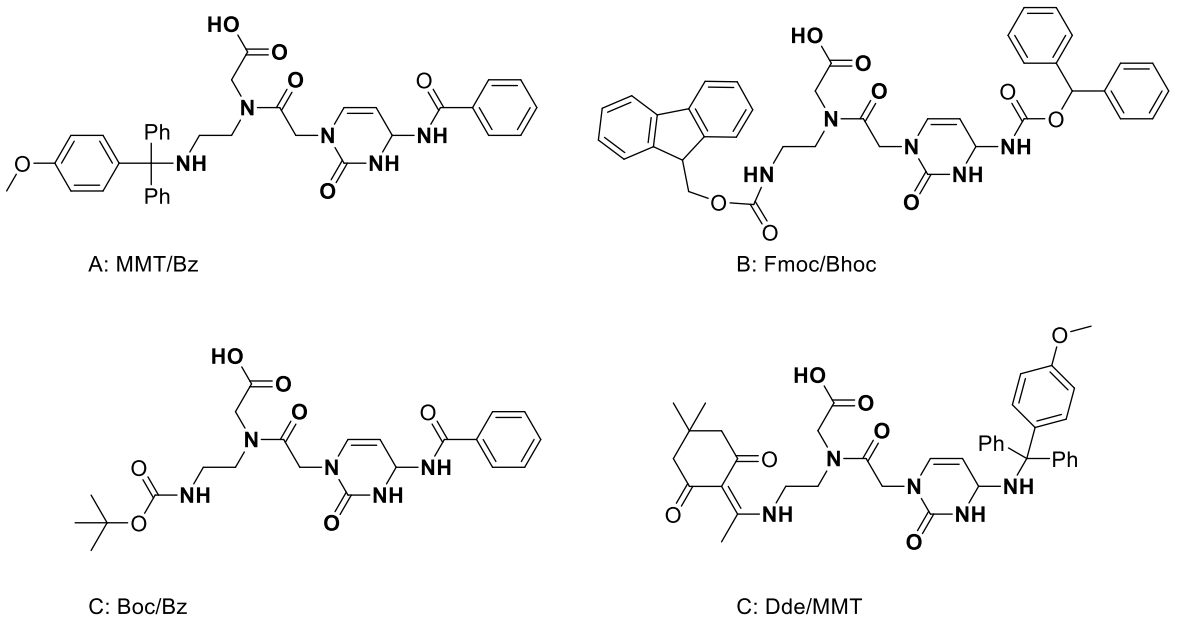


Figure 2-2: Protecting groups used for PNA synthesis. A: MMT/Bz; B: Fmoc/Boc; C: Boc/Bz; D: Dde/MMT. In each case, only cytosine the cytosine nucleobase is shown. Bolded parts define the core monomer structure.

The strategy used in this work was dictated by the synthesizer available to us. We used Fmoc/Bhoc PNA monomers that are available commercially and the automated solid phase protocol has been optimized on the Expedite 8909 synthesiser giving good yield and a straightforward way to make PNA strands at the 2 µmole scale.

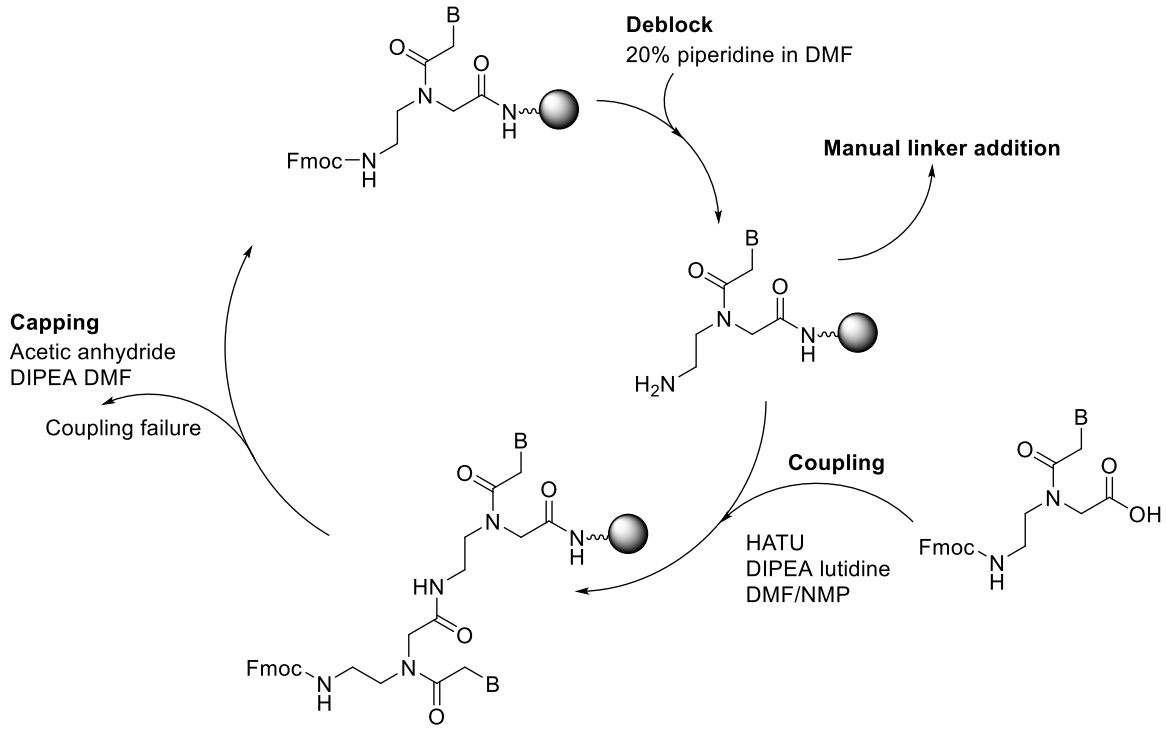


Figure 2-3: Fmoc/Bhoc PNA synthetic cycle on Expedite 8909.

Step	Reagent	Time
Wash	DMF	—
Deblock	20% Piperidine in DMF	2.5min
Wash	DMF	—
Activate and couple	5eq DIPEA; 7.5eq Lutidine; 5eq monomer; 4.5eq HATU	8.25min
Wash	DMF	—
Cap	5% acetic anhydride 6% DIPEA in DMF	3min

Table 2-1: Details of the PNA solid phase synthesis from Fmoc/Bhoc protected monomers on the Expedite 8909 synthesizer.

The resin used in this solid phase synthesis is an Fmoc-PAL-PEG polystyrene resin (**Figure 2-4**). The deprotection of the Bhoc group and the cleavage from the resin were achieved in neat TFA and PNAs were recovered after precipitation in cold ether.

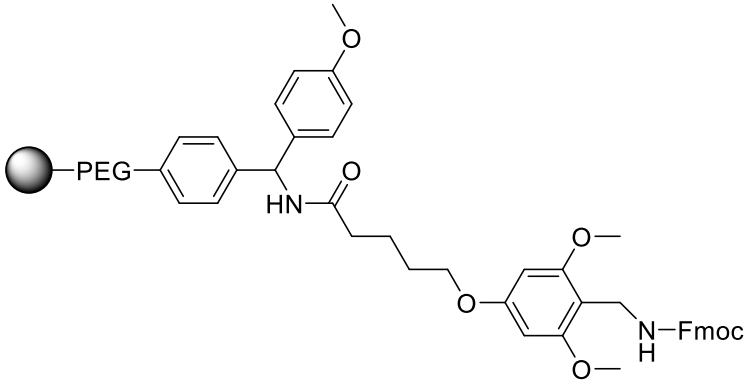


Figure 2-4: PAL (Peptide Amine Linker) resin structure

2.1.4 PNA delivery: cell-penetrating peptides

One of the factors that limit the applications of PNAs is their generally low uptake in cells. PNA do not cross the cell membrane easily and show a very slow release from the endosomes.[117-120]

The development of delivery agent was needed and the research focused on two parallel strategies of PNA-conjugates. The principle is to conjugate a small molecule or peptide which can interact with cell surfaces or endosomes to enhance PNA uptake. Cell-penetrating peptides (CPP) are ~3-30 aa peptides that cross the cell membrane and can be used to improve the internalization of various biomolecules. CPPs have been conjugated or co-injected with plasmids, ASOs, siRNAs and PNAs to improve their delivery.

Most CPPs are either derived from an active domain of a viral or other uptake protein, or designed to be a highly charged peptide that serves as a "proton-sponge", swelling and breaking endosomes as the pH decreases. Historically, the 18-mer peptide "penetratin" derived from the fly homeobox protein antennapedia and a 12-mer peptide derived from the HIV transactivator of transcription (TAT) were the first to show improved uptake by conjugation and co-injection.[121] Many additional conjugated PNA-CPPs have been tested since. In one of the most powerful applications CPP-PNAs were able to inhibit gene expression of intracellular bacteria in cultured cells and in *C. Elegans,*[122] demonstrating the powerful potential of this technique.

2.2 MicA silencing: Oligoswitch

Our collaborators, the group of Dr Anastasia Callaghan at the University of Portsmouth, developed a modified small non coding RNA, $MicA_{stab}$ to block the translation of a bacterial gene called OmpA (Outer membrane protein A) mRNA.[123] Together, we developed a PNA-based ON-OFF switch for this small non-coding RNA.

2.2.1 MicA background

MicA is a small non-coding RNA that plays a key role in post-transcriptional regulation in bacteria. It interacts with the ribosome binding site (RBS) on the 5'-UTR of the messenger RNA to activate or deactivate its translation.[124] MicA regulates the expression of OmpA (Outer membrane protein A) by binding OmpA mRNA[125] and blocking its translation. The pairing of the MicA with the mRNA prevents translation by occluding the ribosome binding site of the mRNA[125] and producing a double strand which is a substrate for RNase III.[126, 127] The OmpA protein has been most deeply studied in *E.coli,* but it has been found in all gram-negative bacteria. The various bacterial proteins are structurally diverse on the outer membrane part of the protein but highly conserved at the

intramembrane β-barrel. OmpA is suspected to play a role in the structural integrity of the outer membrane.[128] OmpA is also very important for the stress resistance of the bacteria and a defect in the β-barrel structure leads to very poor stress resistance.[128] MicA is transcribed from one of the genes induced by the σ^E pathway in response to environmental stress. MicA has two states: one is inactive (in which the sRNA forms a hairpin containing part of the OmpA mRNA complementary sequence) and the other is active (in which MicA changes its shape, displacing the hairpin to uncover the OmpA mRNA binding site)[125] (**Figure 2-5**).

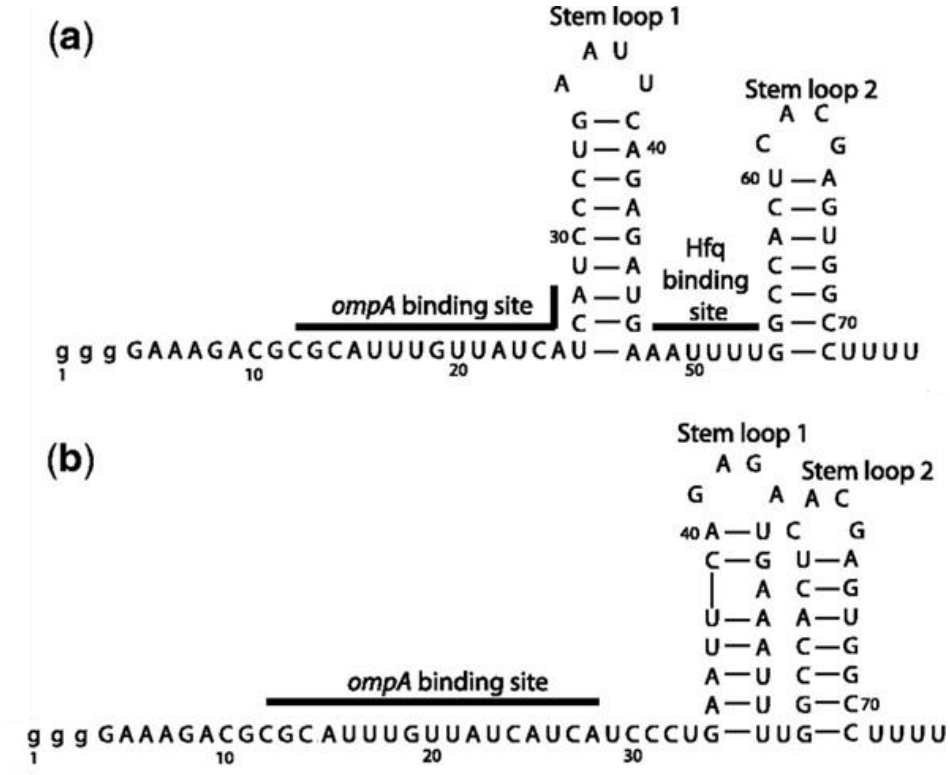


Figure 2-5: MicA structure determined by Udekwu et al.[125] a) native state b) bound to OmpA mRNA. Reproduced with permission from Henderson et al. 2013.[123]

2.2.2 $MicA_{stab}$ structure and mechanism

Our collaborator Anastasia Callaghan and her group at the University of Portsmouth developed a modified sRNA called $MicA_{stab}$ which switches off the translation by binding the ribosomal binding site of the OmpA 5'-UTR.[123] By modifying the sequence between the 34th and the 52nd nucleotides raising the GC content, they restructured the stem loops of MicA making the OmpA binding site constitutively available. Their study shows that $MicA_{stab}$ has greater affinity for OmpA mRNA and can efficiently switch off the translation of OmpA mRNA.

Figure 2-6: $MicA_{stab}$ structure predicted by mfold[129]

In order to validate the switching model, the project is to develop a second oligoswitch that will be able to bind the modified sRNA to free the mRNA. This would enable the system to activate or deactivate the translation of the mRNA. This ability to deactivate the modified sRNA can be a very important feature in order to validate this kind of modified sRNA for therapeutic applications.

The efficiency of the ASOs was assessed by our collaborators Dr Charlotte Henderson. The ASOs were first tested in-vitro by Electrophoretic Mobility Shift Assay (EMSA) and then *in*-vivo in *E. coli.* by a fluorescent reporter assay developed by our collaborators.

2.2.3 Oligonucleotide designs

To disrupt the interaction between the mRNA and the modified sRNA various parameters must be taken into consideration. The binding affinity of the ASO with the targeted sRNA has to be higher than the binding affinity between sRNA and mRNA. The ASO must be stable in a biological environment (high nuclease stability), and the toxicity at the effective concentration has to be low or non-existent. In addition to these standard limitations, the ASO has to be able to switch ON/OFF the sRNA in a concentration-dependent manner: in this way the deactivation of the sRNA could be made reversible by adding more sRNA. Because of these factors, the silencing approach in this project will use ASOs acting as steric blockers. We chose to investigate two types of chemistry with very high affinity, namely PNAs and LNA-DNA phosphorothioate mixmers. A mixmer is an ASO with a fully phosphorothioate backbone, made of DNA with LNA nucleotides distributed every 2-3 bases. This mixture of LNA and DNA is unable to recruit RNase H and so serves as a steric blocker.

In order to target $MicA_{stab}$,[123] three antisense sequences complementary to the OmpA binding site of the sRNA were designed (**Table 2-2**). They were then analysed with OligoAnalyzer 3.1[130]. Key to the design was to maximize affinity for the target, while minimizing hairpin formation or self-assembly of the oligonucleotides.

Name	Sequence
ASO-A	5'-TGA TAA CAA ATG CGC G-3'
ASO-B	5'-GGG ATG ATG ATA ACA A-3'
ASO-C	5'-ATA ACA AAT GCG CGT C-3'

Table 2-2: ASO sequences for the $MicA_{stab}$ switch.

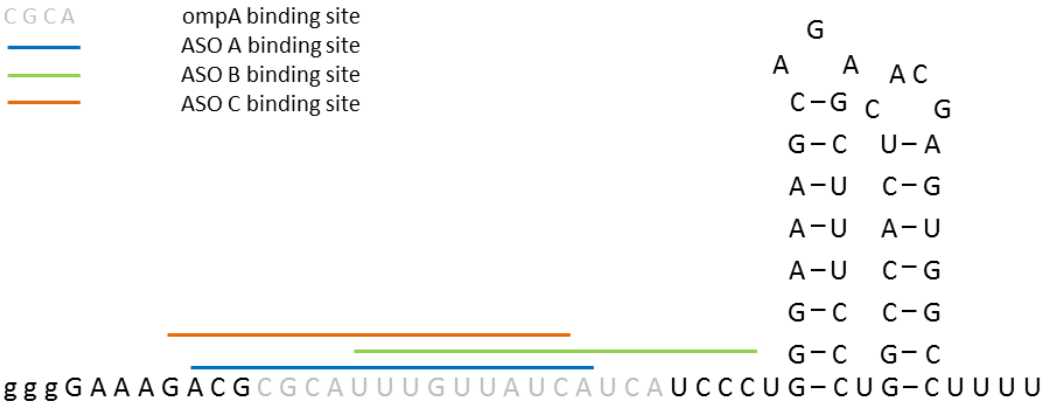


Figure 2-7: $MicA_{stab}$ and binding sites of designed oligoswitch.

PNA oligonucleotides were designed with standard aminoethyl glycine monomers with non-modified nucleobases. One lysine (K) was added at each extremity of the PNA strand to improve solubility.[88]

The design of the LNA mixmers was based around two parameters, the nature of the backbone and the ratio and position of the LNA nucleotides. A phosphorothioate backbone was selected for the nuclease resistance it provides.[63, 64, 131, 132] The positioning of the LNA bases was directed by the LNA/LNA binding affinity, pairing between LNA nucleotides has to be considered because the LNA affinity for LNA is very high and can easily form self-dimer or hairpins. Furthermore, LNA nucleotides were positioned in the 5' and 3' ends to enhance further the exonuclease resistance. Theoretical properties of designs were calculated with the help of the online software OligoAnalyzer 3.1[130] and the following designs were selected.

Name	Sequence
LNA-A	$+T_s+G_sdA_s+T_sdA_sdA_s+C_sdA_sdA_s+A_sdT_s+G_sdC_sdG_sdC_s+G$
LNA-B	$+G_s+G_sdG_sdA_s+T_sdG_sdA_sdT_s+G_sdA_sdT_s+A_sdA_sdC_s+A_s+A$
LNA-C	$+A_s+T_sdA_sdA_s+C_sdA_sdA_sdA_s+T_sdG_sdC_s+G_sdC_sdG_s+T_s+C$
PNA-A	K-tgataacaaatgcgcg-K
PNA-B	K-gggatgatgataacaa-K
PNA-C	K- ataacaaatgcgcgtc-K

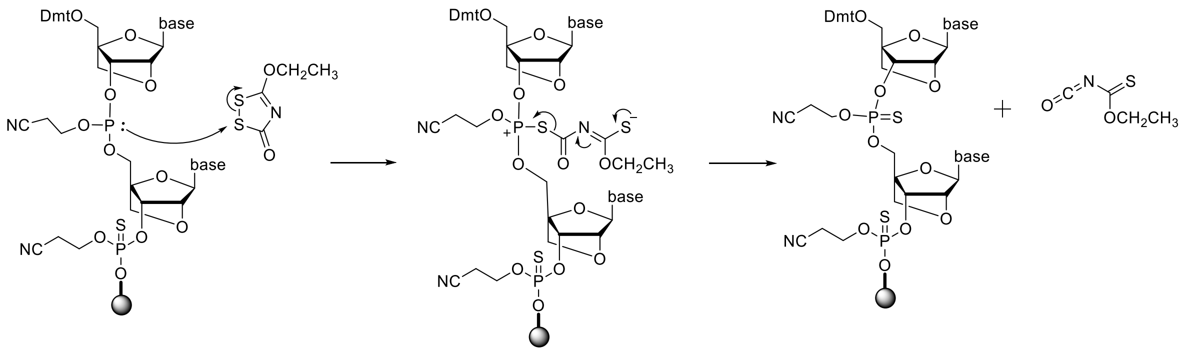
Table 2-3: PNA and LNA designs for the $MicA_{stab}$ switch. LNA is "+N", DNA is "dN" is DNA, phosphorothioate linkage is "$_s$", PNA is "n". Sequences are given from 5' to 3' and N-terminus to C-terminus respectively.

2.2.4 Oligoswitch synthesis for in-vitro experiment

2.2.4.1 PS-LNA mixmers

The LNA mixmers solid phase synthesis was conducted on an ABI 394 DNA synthesizer. LNA phosphoramidites were synthesized in house and were dissolved in acetonitrile at 0.15M and filtered if necessary. LNA-C was dissolved in 3:1 Acetonitrile:THF and filtered. The steps of deprotection, activation, coupling, and capping followed the usual method in solid phase synthesis. The sulfurization step was carried out with EDITH (3-Ethoxy-1,2,4-dithiazoline-5-one), using two sulfurization steps of 1 min each and a solution of EDITH at 0.05M in Acetonitrile.

The mechanism of sulfurization is shown in **Scheme 2-1**. The lone pair from the phosphorus attacks the disulphide bond in the EDITH and opens it. Then the charges migrate to give the uncharged isocyanate and the desired phosphorothioate.



Scheme 2-1: The mechanism of sulfurization with EDITH

Following standard work up and deprotection with ammonium hydroxide the oligonucleotides were desalted with Nap-10 (Sephadex) columns. The oligonucleotides were recovered in good yield (**Table 2-4**).

Name	Yield	Expected MW	Observed MW
LNA-A	78%	5365.31	5364.5
LNA-B	89%	5415.33	5414.5
LNA-C	60%	5339.31	5339.5

Table 2-4: Crude yield of the LNA modified oligonucleotides and their observed mass.

The purity of the product was assessed by analytical PAGE and HPLC-MS.

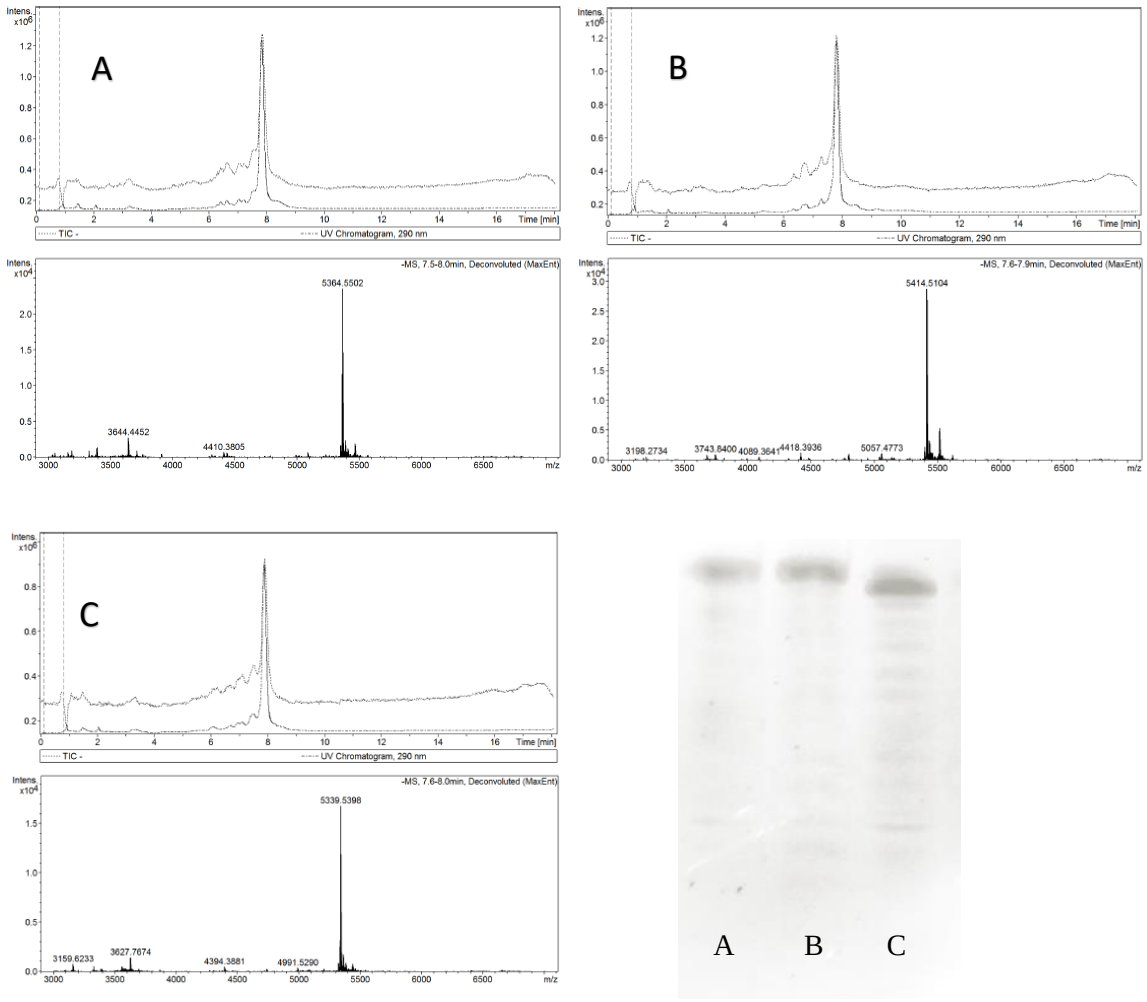


Figure 2-8: HPLC-MS and analytical PAGE of LNA-A, LNA-B and LNA-C.

The mass spectrometry results confirmed that we had obtained the desired product. The 20% polyacrylamide analytical gel and HPLC trace show small amounts of impurities (primarily failure sequences with shorter molecular weight than the full-length product) along with the desired oligonucleotides. However, the amount of impurities was judged sufficiently small that the mixmers could be used as it was in this set of experiments.

2.2.4.2 PNA

PNA synthesis was achieved using standard Fmoc/Bhoc protection with standard protocol on an Expedite 8909. As PNA synthesis is known to be challenging we followed a set of rules for the synthesis. During synthesis, the third or more purine in a row was double-coupled; the same rule applies for the third and more pyrimidine in a row. Lysine at the extremities is always double coupled. In addition, for PNA longer than 15mers, every base beyond this length was double-coupled.

Name	Yield	Stepwise Yield	Expected MW	Observed MW
PNA-A	63%	97%	4642.52	4642.1
PNA-B	58%	97%	4706.57	4706.0
PNA-C	16%	90%	4602.50	4602.0

Table 2-5: Crude yield of the PNA synthesis and calculated and observed mass of synthesized PNAs.

PNA-C is recovered with a very low yield, another attempt to synthesize it has been done and the recovered yield was of the same order. Since the LNA-C shows a lower yield as well relative to the other two LNA sequences, it suggests that the sequence itself is more difficult to synthesize. Mass spectrometry analysis confirmed that the correct products were obtained (**Table 2-5** and **Figure 2-9**). PNAs were desalted by Nap-10 (Sephadex) and sent without further purification to our collaborators.

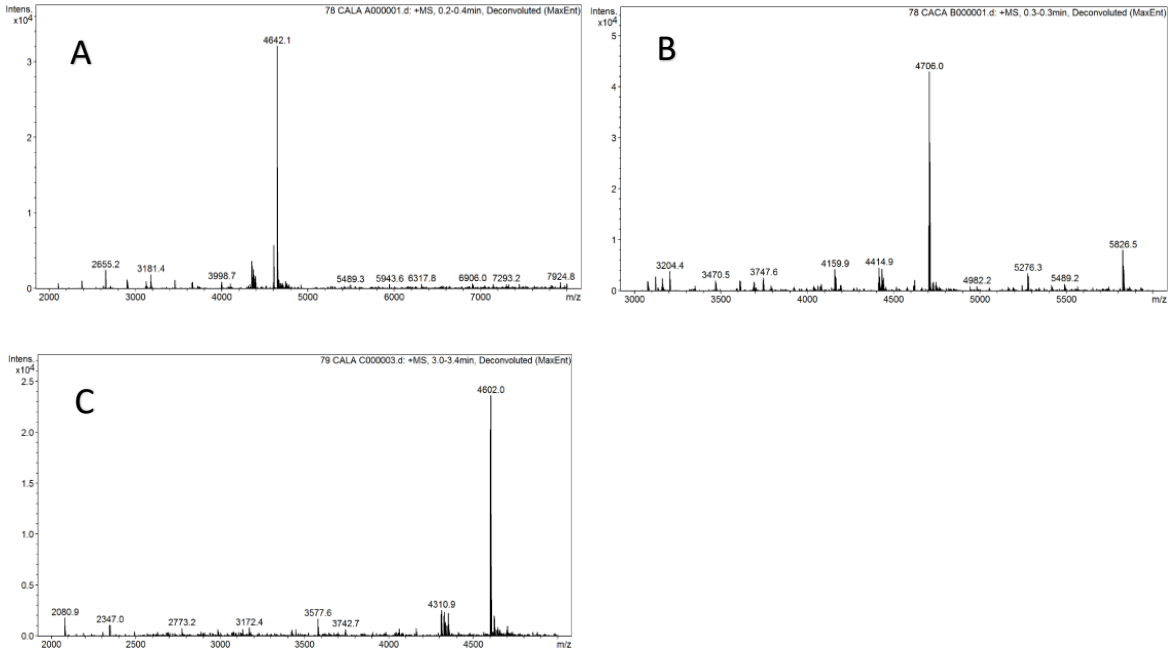


Figure 2-9: Mass spectrometry spectra for PNA-A (A), PNA-B (B) and PNA-C (C). Calculated mass shown in **Table 2-5**.

2.2.5 Oligoswitch in vitro efficacy

Our collaborators tested the efficacy of the two different chemistries with the three different sequences by EMSA (Electrophoretic Mobility Shift Assay). The ompA RNA (150nM), ASO (0 to 350 nM), and $MicA_{stab}$ (60nM) were added together and incubated in buffer (10 mM Tris pH 8, 50 mM NaCl, 50 mM KCl and 10 % glycerol) at 37°C for 30 minutes. After incubation, the oligonucleotides complexes were analysed by electrophoresis on 8% native polyacrylamide gel (**Figure 2-10**).

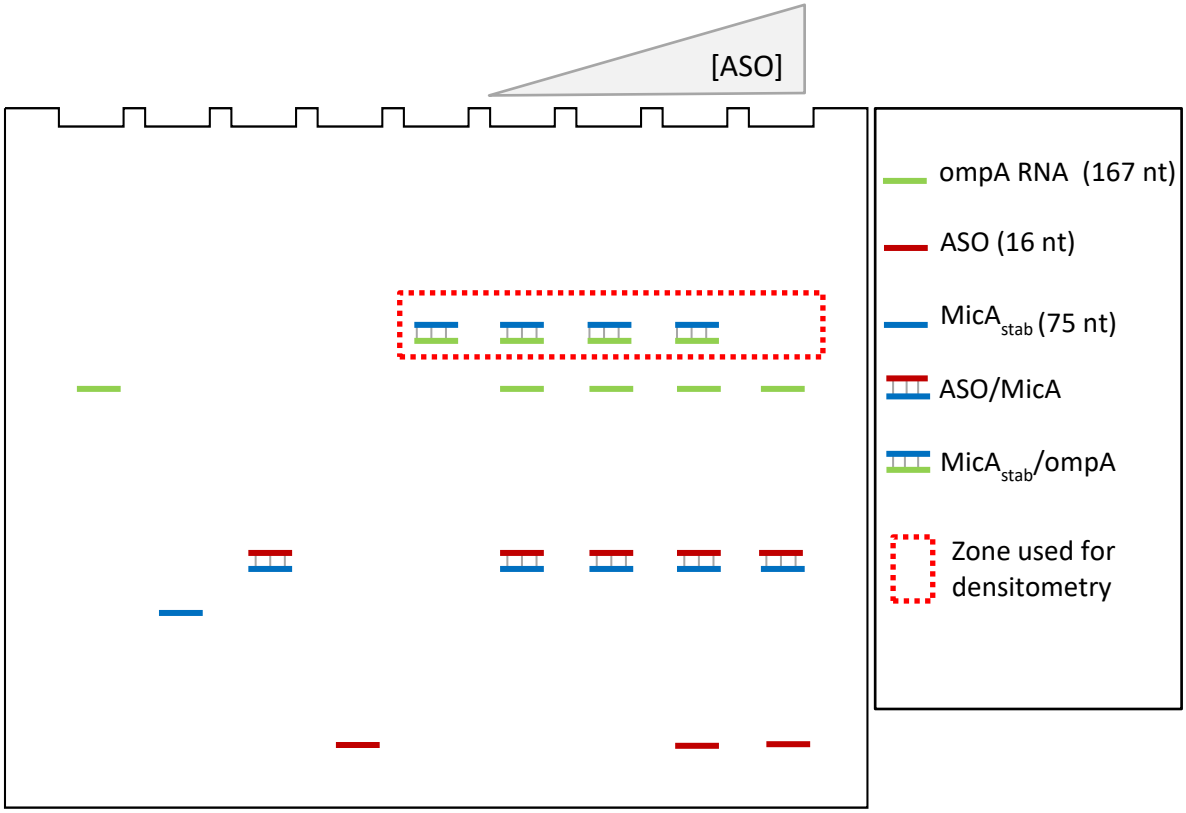


Figure 2-10: Schematic of EMSA experiment for oligoswitch in vitro validation. Well 1 is loaded with ompA RNA only, well 2 is loaded with control with $MicA_{stab}$ only, well 3 is loaded with duplex ASO/$MicA_{stab}$, well 4 is loaded with ASO only (in case of PNA ASO well 4 is irrelevant), well 5 is duplex $MicA_{stab}$/ompA RNA, wells 6 to 11 are loaded with $MicA_{stab}$/ompA RNA and an increasing concentrations of ASO. The gel bands presented in **Figure 2-11**, **Figure 2-12**, **Figure 2-13** and **Figure 2-14** are images of the delimited red box.

The gels were revealed with SYBR Gold and read in a gel imager. Densitometry of the bands of interest (red dotted box) was measured with ImageJ and plotted as the percentage of disrupted $MicA_{stab}$/ompA duplexes.

The assay was performed with two different sets of conditions. The first set of conditions was labelled as direct and the second set was labelled as preformed. The direct assay was designed to check the ability for the ASO to compete with the ompA RNA for the binding of $MicA_{stab}$. The "preformed" assay was performed under the same condition as the direct assay but in the first place the ASO was not incubated with the $MicA_{stab}$ and the ompA RNA. $MicA_{stab}$ and RNA were incubated first at 37°C for 30 minutes then ASO was added and the solution was incubated again at 37°C for 30 minutes. This set of conditions was designed to check the ability of the ASO to disrupt an already formed $MicA_{stab}$ /ompA RNA duplex.

It was surprising to us that the gels of the EMSAs (**Figure 2-11**, **Figure 2-12**, **Figure 2-13** and **Figure 2-14**) also contained signal on the edges; nevertheless, this is consistent with our collaborators'

published EMSA gels in this system.[123] The changes in the band densities from sample to sample are clear and can be reproducibly quantitated using densitometry.

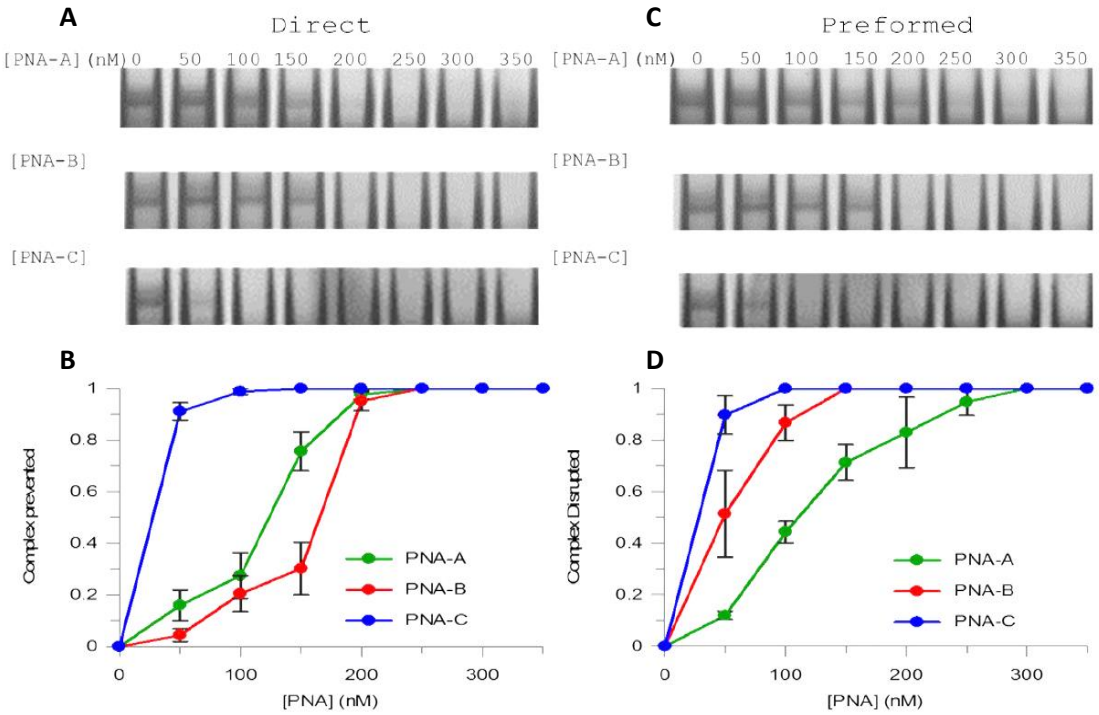


Figure 2-11: Results of direct (ompA, ASO and $MicA_{stab}$ are incubated together) and preformed (mpA and $MicA_{stab}$ are duplexed before addition of ASO) EMSAs showing the ability of the PNAs to bind $MicA_{stab}$. Reproduced from T. Caddick thesis [133] with permission. The graphs show quantitation of the amount of disruption, based on densitometry analysis of the gels. See **Figure 2-10** for details of the experiment.

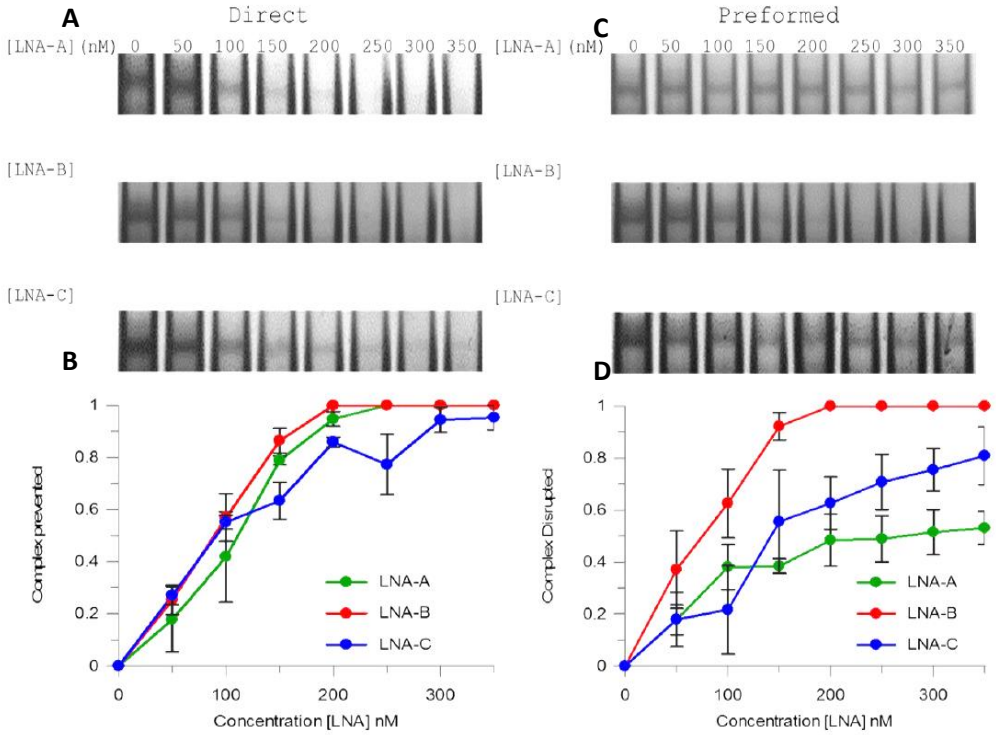


Figure 2-12: Results of direct (ompA, ASO and $MicA_{stab}$ are incubated together) and preformed (mpA and $MicA_{stab}$ are duplexed before addition of ASO) EMSAs showing the ability of the LNA mixmers to bind $MicA_{stab}$. Reproduced from T. Caddick's thesis[133] with permission. The graphs show quantitation of the amount of disruption, based on densitometry analysis of the gels. See **Figure 2-10** for details of the experiment.

All ASO designs show an ability to prevent or displace the $MicA_{stab}$/RNA duplex. However, most of the oligonucleotides need to be in large excess compared to $MicA_{stab}$ to achieve this result. LNA mixmers need a larger excess than PNA to be efficacious: at least 150nM for nearly complete prevention of duplexes in competed set up, which is equivalent to ~ 3-fold excess relative to $MicA_{stab}$ (**Figure 2-12-B**) is needed using LNA-A and LNA-B. However, the best PNA, PNA-C shows nearly complete prevention of duplexes by 50 nM concentration (around one equivalent of $MicA_{stab}$) (**Figure 2-11-B**).

In the preformed assay the advantages of the PNAs is shown again with similar results. The best PNA design, PNA-C, shows nearly complete displacement of duplexes by 50 nM concentration (around 1 equivalent relative to $MicA_{stab}$) (**Figure 2-11-D**) while the best LNA mixmers, LNA-B, needs 150 nM for displacing most of the duplexes (around three equivalents relative to $MicA_{stab}$) (**Figure 2-12-D**).

The advantage of PNA is more obvious in the preformed assay where all PNA are better than the LNA, while in direct assay only one PNA (PNA-C) is better than the LNA mixmers. This observation could be explained by a better ability of PNA to invade preformed duplex due to their neutral nature.

The same direct and preformed EMSA assays were performed in the same conditions but this time, Hfq protein was added to the ompA RNA:$MicA_{stab}$ duplex. This set of experiment is interesting before going to in vivo studies because Hfq is highly expressed in the model used for the in vivo studies. It has been observed that even if the natural site of binding to Hfq has been removed from $MicA_{stab}$, the Hfq protein binds to $MicA_{stab}$. The results of the second set of assays are very different than the first. When the Hfq is bound to $MicA_{stab}$ the LNA mixmers are more proficient to compete or to displace $MicA_{stab}$ from the ompA RNA than the PNA designs (**Figure 2-13** and **Figure 2-14**).

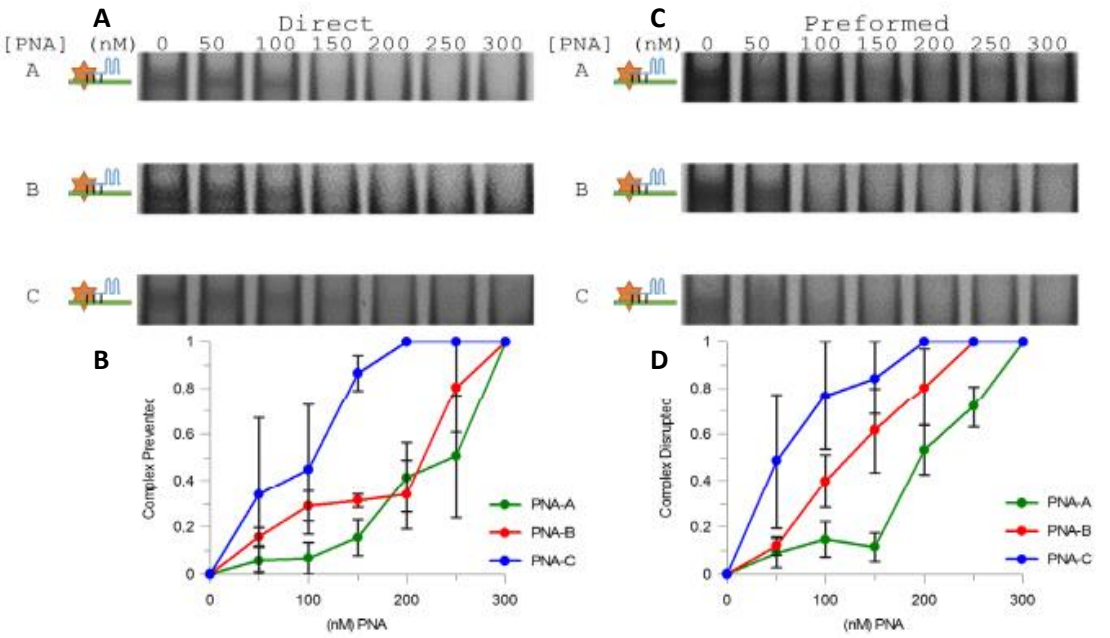


Figure 2-13: Results of direct (ompA, ASO and $MicA_{stab}$ are incubated together) and preformed (mpA and $MicA_{stab}$ are duplexed before addition of ASO) EMSAs showing the ability of the PNAs to bind $MicA_{stab}$ in presence of Hfq protein. Reproduced from T. Caddick thesis [133] with permission. The graphs show quantitation of the amount of disruption, based on densitometry analysis of the gels. See **Figure 2-10** for details of the experiment.

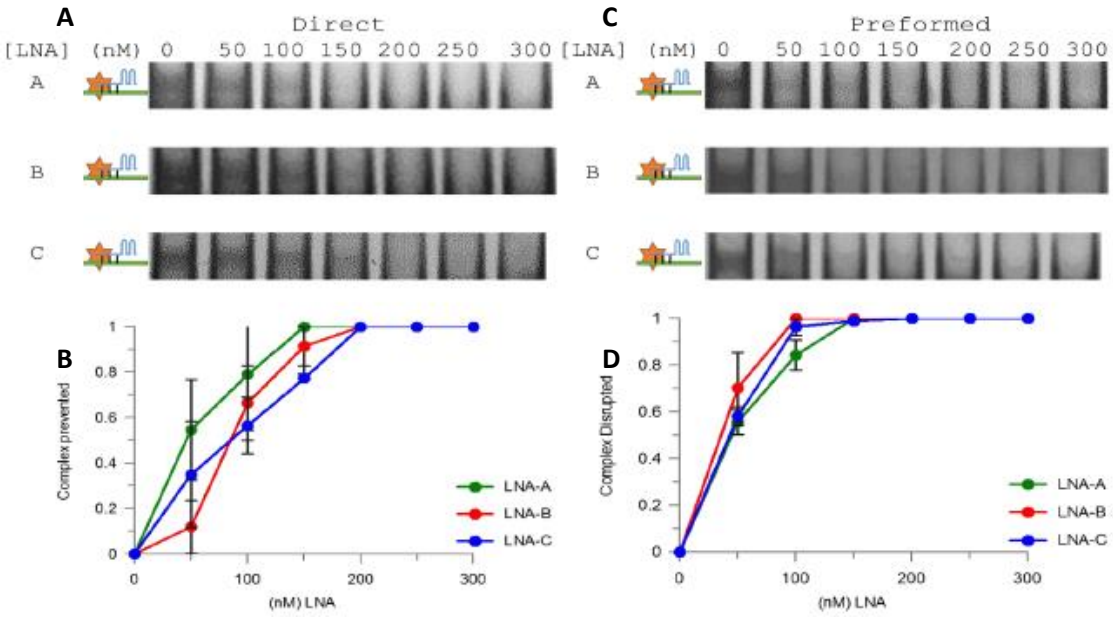


Figure 2-14: Results of direct (ompA, ASO and $MicA_{stab}$ are incubated together) and preformed (mpA and $MicA_{stab}$ are duplexed before addition of ASO) EMSAs showing the ability of the LNA mixmers to bind $MicA_{stab}$ in presence of Hfq protein. Reproduced from T. Caddick thesis [133] with permission. The graphs show quantitation of the amount of disruption, based on densitometry analysis of the gels. See **Figure 2-10** for details of the experiment.

The difference of efficacy seen in the presence or absence of Hfq protein suggests an interaction between the ASOs and the protein.

2.2.6 Interaction of PNA with Hfq protein

Our collaborators in Portsmouth used microscale thermophoresis (MST) to study the interaction of the ASOs with the Hfq protein. Microscale thermophoresis is a technique which measures the directed movement of particles in a microscopic temperature gradient. The thermophoretic movement is measured by monitoring the fluorescence distribution of the fluorescently labeled molecule of interest. The molecules are placed in buffer in a capillary and the micro temperature changes are made with the help of an infrared laser. Our collaborators made a fluorescent Hfq protein and tested LNA-B and PNA-B binding by MST. The choice of the sequence B was made because both of the ASO designs show similar efficacy in the EMSA with or without the presence of Hfq protein.

LNA-B shows a K_d <50nM and PNA-B had a K_d of 38µM. Thus while the K_d of the LNA mixmer is on the same order as the K_d value reported for RNA[123], the PNA shows very low affinity binding with the Hfq protein. Thus, the better efficacy of the LNA mixmers in presence of Hfq could be due to a synergy between the ASO and the protein.

2.2.7 Oligoswitch design and synthesis for *in vivo* work

Our collaborators in the Callaghan lab developed a reporter assay to monitor the repression of the translation of the ompA RNA. In essence, the ompA sequence is fused to a green fluorescent protein (GFP) sequence, so that the translation of ompA can be monitored by fluorescence. $MicA_{stab}$ completely suppressing the expression of GFP, if the ASOs are successful a restoration of the fluorescence that depends on the concentration of ASO will be observed. Our collaborators selected the sequence B for in vivo experiments.

PNA are known to have poor uptake. It was then necessary to design the PNA ASOs with a means to enter the cell. It has been demonstrated that the attachment of a lysine- and/or arginine-rich peptide strongly improves the uptake of PNA in *E. Coli.*[134]

The coupling of a peptide onto PNA can be performed following several strategies. The strategy used in this work was a horizontal in-line synthesis of the PNA and peptide on the Expedite synthesizer, which is a straightforward option since both PNA and peptide components are assembled using Fmoc peptide chemistry. PNA and peptides were linked by a hydrophilic spacer, AEEA (2-(2-(2-aminoethoxy)ethoxy)acetic acid) (**Figure 2-15**). This type of linker improves solubility while physically separating the peptide from the PNA, circumventing some of the steric interactions that might hinder duplex formation of the PNA with its target.

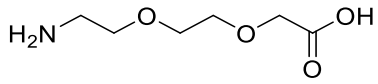


Figure 2-15: Structure of 2-(2-(2-aminoethoxy)ethoxy)acetic acid (known as AEEA or O-linker), used as a spacer in PNA-Peptide synthesis

The PNA-peptides were synthesized on an Expedite 8089 with standard PNA protocol. All amino acids and the AEEA linker were double-coupled. Syntheses of the PNA-peptide with the peptide at the NH_2 and at the COOH end were both attempted. Synthesis with the peptide at the C-terminus of the PNA failed: the recovered product resulted from an incomplete synthesis (^{Ac}ta aca a-O-KFFKFFKFFK). However, the syntheses with the peptide on the N-terminus of the PNA afforded the desired products (**Figure 2-7**).

	Sequence	Yield	Expected Mass	Observed Mass
PNA-B-CPP-O	KFFKFFKFFK-O-ggg atg atg ata aca a	5%	5990.77	5990.6
PNA-B-CPP-OO	KFFKFFKFFK-O-O-ggg atg atg ata aca a	14%	6136.27	6135.7

Table 2-6: Oligomers used in this study, K is lysine, F is phenylalanine, lowercase is PNA, O is 2-(2-(2-aminoethoxy)ethoxy)acetic acid) (AEEA).

The insertion of a second AEEA linker between the peptide and the PNA part afforded the desired product with a higher yield than the one with only one AEEA linker. The increase in yield may be at least partly explained by increased PNA solubility, but in addition, the increased distance between the PNA and the peptide could make the solid phase synthesis easier. The PNA-B-CPP-OO was sent to our collaborators along with peptide alone and a non-targeting PNA-peptide for in vivo toxicity and efficacy study.

2.2.8 Design and synthesis of Cy3 tagged PNA

We also synthesized the PNA sequences with a fluorescent tag. Cy3 was chosen as the fluorescent label to be coupled to the PNA strand. Cy3 and Cy3-NHS (**Figure 2-16**) were used to couple the dye to the PNA strand. The NHS active ester is reactive toward primary alkyl amine groups without the need for a coupling reagent.

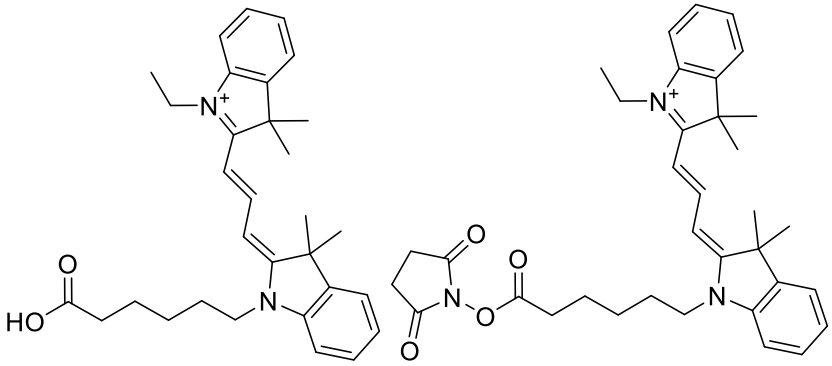


Figure 2-16: Cy3 structure (on the right) and Cy3-NHS structure (on the left).

Cy3 can be coupled on the PNA strand by peptide coupling. The Cy3-NHS ester can be coupled to the PNA strand following several strategies: it can be coupled on solid phase at the N-terminus after the synthesis and before deprotection, it can be added on the N-terminal in solution after deprotection or it can be added on an amino acid side chain containing a terminal amine (**Table 2-7**). Cy3 can be coupled in-line in the synthesizer but this method is not preferred because the volume of solution needed to prime the synthesizer is a waste of product and Cy3 is very expensive.

Name	Sequence	Expected Mass	Observed Mass
PNA-B-Cy3	Cy3-OO-ggg atg atg ata aca a	5180.00	**5178.2**
PNA-A-K-Cy3	^{Ac}tga taa caa atg cgc g-OO-K(Cy3)	5299.65	**5299.2**
PNA-B-K-Cy3	^{Ac}ggg atg atg ata aca a-OO-K(Cy3)	5364.40	**5363.3**

Table 2-7: Synthesis results for PNA-Cy3 syntheses. N-terminal Cy3 have been added on solid phase by synthesizer, C-terminal Cy3 has been added in solution with use of Cy3-NHS. K is Lysine, F is Phenylalanine, lowercase is PNA, O is 2-(2-(2-aminoethoxy)ethoxy)acetic acid).

The coupling of the NHS-Cy3 on the PNA was performed after deprotection of the PNA, the PNA was capped with acetic anhydride to avoid reaction of the N-terminal with the NHS-ester. The NHS-ester is reacting sufficiently faster with primary amine than with aromatic amine so the reaction can be performed even if the amines on the nucleobases are deprotected. However, the nucleobase can still react with the NHS ester so the reaction has to be timed to avoid multiple Cy3 couplings on one PNA strand. The first method attempted to add the Cy3 to the PNA strand was taking place in 0.1 M $NaHCO_{3\ aq}$ pH=8.3, but the PNA strand precipitated out of the solution. The PNA was then dissolved in 4:1 DMF:DMSO (200µL) with DIPEA as a base (5µL). The Cy3-NHS ester was added to the solution and the mixture was left at room temperature. The timing of the reaction was followed by HPLC.

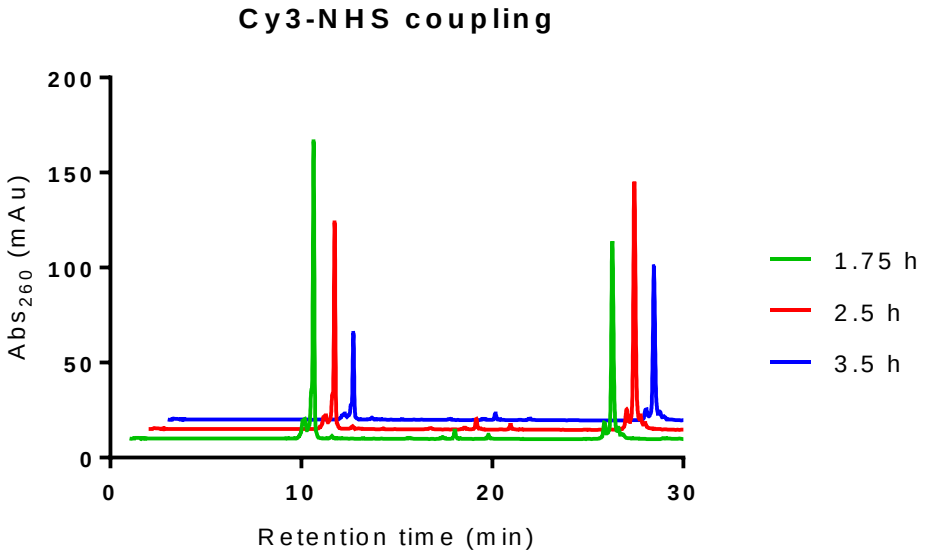


Figure 2-17: RP-HPLC traces showing the absorbance at 260nm at different time points of the Cy3-NHS ester with the side chain amine of a lysine in PNA.

From the HPLC traces shown in **Figure 2-17** it can be seen that the reaction is very slow. The ratio of the desired product (peak at 25 min) to the starting material (peak at 9.7 min) is increasing slowly over time. In addition, throughout the reaction course, some non-desired product accumulates (retention time > 30min, presumably multiply-labelled PNAs, not shown).

This method was able to afford the desired product reproducibly on different PNA sequences, but the yields obtained by this method were low: 10 – 30 nmol of tagged PNA were recovered from a 2 µmol synthesis. The use of a different base and a higher excess of NHS-ester could be investigated for better yield as multiple additions of Cy3 in the same PNA strand haven't been observed. However, optimisation wasn't performed as a large enough quantity of the desired compounds was obtained for the in vivo experiment.

The syntheses of oligoswitches have been performed with success. While the first batch of mixmers and PNAs was limited due to lack of instruments for satisfactory synthesis, purification and analyses. Methods were in a second time developed for better synthesis and purification of PNA and were successfully used to afford the CyDye and peptide tagged PNAs. The preliminary results from our collaborators for in vitro strand displacement were encouraging. However, several obstacles appeared while running the experiment in vivo: some toxicity has been observed for the PNAs, delivery of peptide alone and comparison with naked and tagged PNA is underway and we are in close communication to improve the design by changing the chemistry or the peptide sequence if needed.

2.3 PNA modified siRNAs (siPNAs)

2.3.1 Aim and design

A lot of chemical modifications for therapeutic siRNA have been studied over the last 20 years. The inclusion of PS linkages and sugar-modified nucleotides including 2'-OMe-RNA, LNA, 2'-F-RNA and 2'-MOE-RNA have been studied and efficient and stable designs combining several chemistries have been reported.[135, 136] The general idea was to make the sense and antisense strands resistant to nucleases, reduce immune stimulation and other off-target effects, and improve delivery of the siRNA without losing the interaction with the RISC complex. It has been observed that modifications on the sense strand are often more readily tolerated than modifications on the antisense strand. Our collaborators at UMASS medical school, the group of Dr Anastasia Khvorova, developed an asymmetric siRNA that shows very high efficiency in vitro and in vivo (**Figure 2-18**).[22, 137-140]

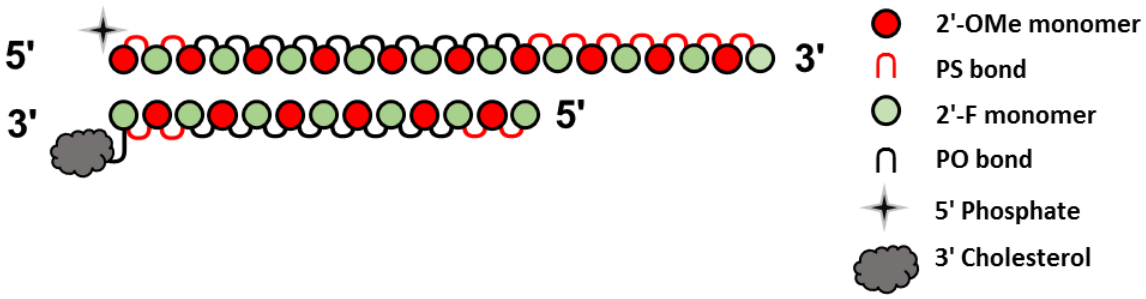


Figure 2-18: siRNA design used by the Khvorova group.

The effect of PNA chemistry on siRNA technology has barely been investigated in siRNA technology. In 2008, Potenza *et al.*[141] made an unmodified 21mer duplex siRNA with different chemically modified 3' TT overhangs. Working with an unmodified siRNA scaffold, they compared the stability of duplexes containing 3'-overhangs of DNA or PNA, in the sense strand, the antisense strand or both. They observed very similar RNAi activity for the all unmodified and PNA-modified designs, and a small improvement in serum stability when having the PNA overhang in the sense strand. Nevertheless, the overhangs of a siRNA are very tolerant to modifications, and to the best of our knowledge there has been no exploration of PNA in the body (duplex region) of an siRNA.

The present study is investigating the effect of a PNA sense strand coupled with a chemically modified siRNA antisense strand. The PNA with its neutral backbone could have the ability to facilitate cellular uptake by "dilution" of charge while enhancing the stability of the siRNA. However, the possible advantage of the PNA sense in delivery would only be worthwhile if the duplex is efficiently taken up by the RISC complex.

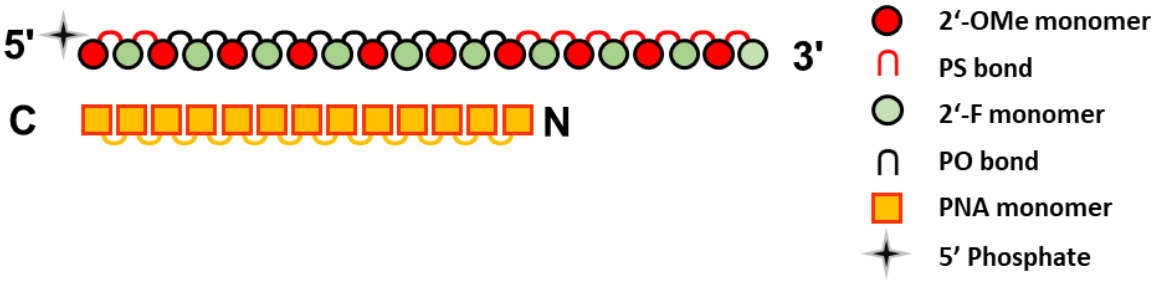


Figure 2-19: Design of the siPNA.

The efficiency of siPNA was tested in cultured human cell lines for their ability to silence two ubiquitously expressed RNA targets, peptidyl-prolyl cis-trans isomerase B (PPIB) and huntingtin (HTT) messenger RNAs.

2.3.2 Synthesis of siPNA

The synthesis of the antisense strand was performed by D. Echeverria of the Khvorova lab using standard oligonucleotide synthesis conditions. The PNA strands were synthesized using standard PNA conditions on an Expedite 8909. The 14mer PNAs were synthesized without addition of lysine to avoid any disruption of interactions with the RISC complex. PNAs have very high affinity to RNA, to get around the possibility of a deactivation of the siPNAs due to the inability for the RISC to separate the duplex, every PNA sequence was synthesized both fully complementary and comprising one mismatch. All oligonucleotides were HPLC-purified and desalted with Sephadex columns.

Name	Sequence (N to C or 5' to 3')	Expected Mass	Observed Mass
PNA_s_HTT	agt aaa gag att aa	3891.72	3891.6
PNA_s_HTT_mm	agt aaa **t**ag att aa	3866.71	3866.6
PNA_s_PPIB	^{Ac}aaa ttc cat cgt ga	3836.67	3836.5
PNA_s_PPIB_mm	^{Ac}aaa tt**a** cat cgt ga	3860.70	3860.5
mf_HTT_as*	P-U_sU_sA AUC UCU UUA $C_sU_sG_s$ $A_sU_sA_s$ U_sA	6620.40	6619.7
mf_PPIB_as*	P-U_sC_sA CGA UGG AAU $U_sU_sG_s$ $C_sU_sG_s$ U_sU	6730.50	6729.7
mf_HTT_s*	Cy3-C_sA_sG UAA AGA GAU U_sA_sA-TegChol	6394.58	6394.7
mf_PPIB_s*	C_sA_sA AUU CCA UCG U_sG_sA-TegChol	5654.20	5653.2

Table 2-8: Sequences and chemistry of the nucleotides used in the siPNA study. Lower case is PNA monomer, ^{Ac} is acetyl cap, P is a 5' phosphate, green letters are 2'-Fluoro-RNA nucleotides, red letters are 2'-OMe-RNA nucleotides, Cy3 is a Cyanine 3 dye and TegChol is a cholesterol linked by a tetraethylene glycol linker, bold and underlined PNA bases represent mismatched positions. Compounds marked with * were synthesized by our collaborator.

For annealing the duplexes, the PNA and RNA strands were combined following a standard procedure of adding equimolar amounts of PNA and RNA to form a 150 µM solution in water, heating to 95˚C for 1 min, then slowly cooling down to room temperature. However, this protocol led to the formation of precipitate for three of the four siPNA duplexes (all except the siPNAs-mm^{PPIB},**Table 2-9**). This precipitate was seen even when the solution was at 95˚C. Diluting the duplexes to 30 µM in 1X PBS allowed complete solubilisation of the precipitate. Duplex formation was checked by LC/MS; the main ion peak in all cases showed the correct duplex mass (**Figure 2-20**).

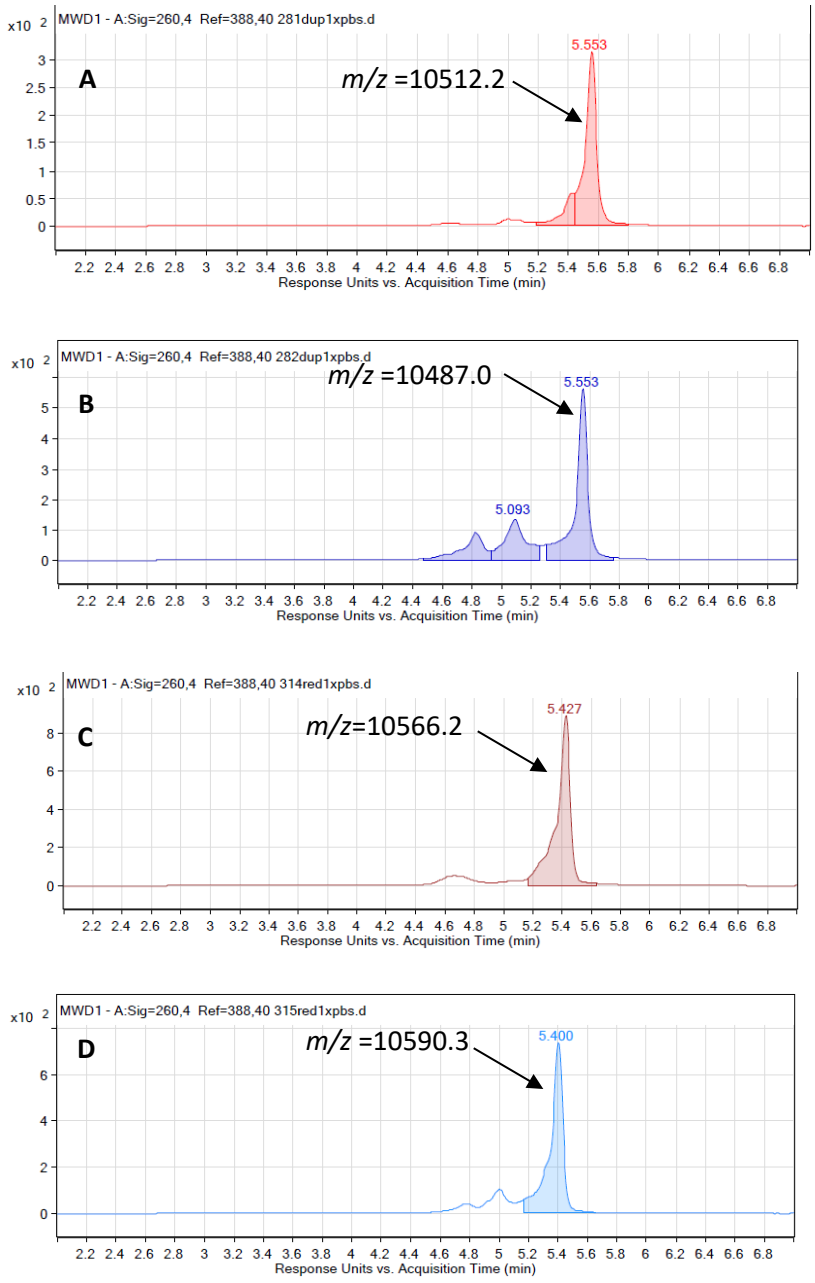


Figure 2-20: siPNA duplexes are correctly annealed as shown by LC/MS traces; the indicated peaks show the correct duplex mass. A: siPNAs^{HTT}; B: siPNAs-mm^{HTT}; C: siPNAs^{PPIB}; D: siPNAs-mm^{PPIB}.

2.3.3 Biophysics of PNA:RNA duplexes

Before transfecting the duplexes into cells, the biophysical properties of the PNA:RNA duplexes were investigated.

Name	Composition	T_m (°C)	Stdev
siPNAs^{HTT}	PNA_s_HTT /mf_HTT_as	76.9	0.6
siPNAs-mm^{HTT}	PNA_s_HTT_mm /HTT_as	56.6	0.4
siPNAs^{PPIB}	PNA_s_PPIB /PPIB-as	84.6	0.4
siPNAs-mm^{PPIB}	PNA_s_PPIB_mm /PPIB-as	70.9	0.5

Table 2-9: Melting temperatures of the siPNAs duplexes at 1 µM, in 1xPBS are calculated by derivative method. The mean and standard deviation of the T_m were calculated from n=3.

The PNA:RNA duplexes show a high T_m. The introduction of a mismatch dramatically lowers the T_m of the duplexes (by 20˚C for the HTT sequence and 14 ˚C for the PPIB sequence). However, the T_m for the mismatched duplexes is still very high (**Table 2-9**) and demonstrates that these duplexes bind with high affinity.

2.3.4 Silencing by siPNAs

The efficacy of the siPNAs was tested by our collaborator Julia Alterman in HeLa cells. The duplexes were transfected with Lipofectamine RNAiMAX. The duplexes efficacies were tested in a range from 1 pM to 1 µM with the standard asymmetric chemically modified siRNA design developed in the Khvorova lab as reference.

Results on silencing of *HTT* are shown in **Figure 2-21** and the results summarized in Table 2-10. The siPNAs were able to silence *HTT* mRNA expression. Silencing by siPNA is two orders of magnitude less potent than silencing by the benchmark siRNA^{HTT} (**Table 2-10**), but more potent than silencing by the antisense strand alone (here termed ‘single-stranded siRNA’ or ss-siRNA).

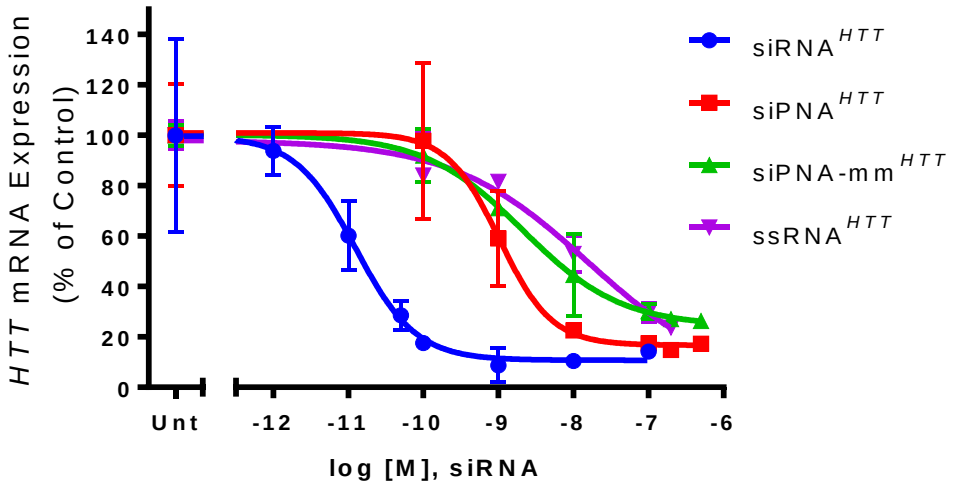


Figure 2-21: Silencing of HTT mRNA. Dose-response of the HTT siPNAs in HeLa cells in 1:1, OptiMEM:DMEM + 6% FBS, plated at 10 000 cells per well in 96 well-plate and transfected using Lipofectamine RNAiMAX. Cells were lysed after 72 h. Quantification of mRNAwas performed using the Quantigene 2.0 assay. Data is normalized to housekeeping gene (PPIB) and presented as % of untreated control. Error bars represent mean ± SD (n=3 wells).

	siRNA^{HTT}	siPNAs^{HTT}	siPNAs-mm^{HTT}	ss-siRNA^{HTT}
IC 50 [nM]	0.012	1	2	15

Table 2-10: IC_{50} of the HTT PNA modified siRNAs (calculated by non-linear fitting using GraphPad Prism).

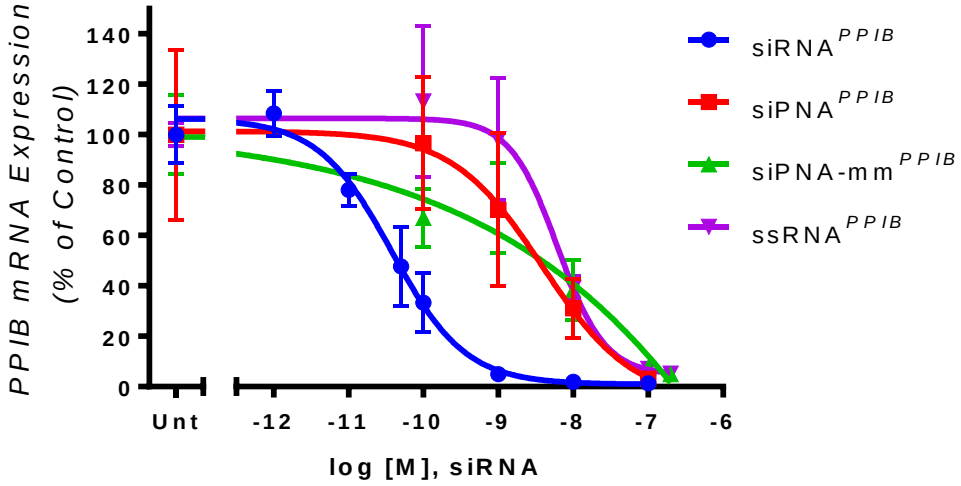


Figure 2-22: Silencing of PPIB mRNA. Dose-response of the PPIB siPNAs in HeLa cells in 1:1, OptiMEM:DMEM + 6% FBS, plated at 10 000 cells per well in 96 well-plate and transfected using Lipofectamine RNAiMAX. Cells were lysed after 72 h. Quantification of mRNA was performed using the Quantigene 2.0 assay. Data is normalized to gene (HTT) and presented as % of untreated control. Error bars represent mean ± SD (n=3 wells).

	siRNA^{PPIB}	siPNAs^{PPIB}	siPNAs-mm^{PPIB}	ss-siRNA^{PPIB}
IC 50	37 pM	4 nM	NA	6 nM

Table 2-11: IC_{50} of the PPIB siPNAs. (Calculated by non-linear fitting using GraphPad Prism).

Silencing was also observed by the second siPNA sequence (**Figure 2-22** and **Table 2-11**). In this case, the fully matched siPNA^{PPIB} showed dose-dependent silencing of PPIB, to a similar extent as the antisense strand alone (ss-siRNA). The mismatched siPNA^{PPIB} dose response was oddly shaped and did not allow us to calculate an IC_{50} value. As with the HTT sequence, silencing with the siPNA^{PPIB} is 2 orders of magnitude lower than the control siRNA^{PPIB} (**Table 2-11**).

In both cases, the mismatched siPNAs were not significantly different than the fully complementary siPNAs.

The use of PNA sense strands in siRNA seems to be tolerated by RISC but the silencing activity is 10 fold lower than optimized siRNA which makes the siPNA design very unlikely to be a candidate in silencing. The lipid-mediated delivery used in the experiment allowed us to draw conclusion only about the RISC compatibility of the construct and it is not possible to conclude if the siPNA is offering any advantages in cellular uptake.

2.4 PNA-peptide tail for siRNA (PPT-siRNA)

Our study on siPNAs, while not uncovering a potent design, opened the way to another concept. The fact that the PNA sense strand doesn't seem to disrupt siRNA activity led us to consider the use of PNA as a peptide carrier for targeted delivery of siRNAs.

2.4.1 Design and synthesis

The use of PNA as peptide carrier for RISC mediated silencing has not been widely investigated. It has been tried without much success on an unmodified short hairpin (shRNA) by addition of an overhang to the hairpin to hybridize with the PNA carrier.[142] For siRNA delivery, PNA carriers have been investigated with non-modified ss-siRNA[143] showing little uptake however the ss-siRNA was unmodified and compared to transfection by Lipofectamine. PNA carriers have also been successfully used as a cell-specific activity switch of siRNA, where a PNA-peptide that is a substrate for Cathepsin B blocks the activity of siRNA in cells not expressing the protease.[144]

We designed siRNA with PNA-peptide tails, called PNA PeptideTail-siRNA (PPT-siRNA), in collaboration with Julia Alterman, a PhD student in Dr Khvorova group. The antisense strand, as

for the siPNA study described in section 1.3, was a 20mer oligonucleotide with an alternating pattern of 2'-OMe and 2'-Fluoro monomers, and the 5' end of the strand was phosphorylated. The two first linkages as the 5' end were phosphorothioated as well as the seven linkages at the 3' end. The sense strand was made of the same pattern of alternating modified nucleotides; in both strands, this pattern of sugar modifications enhances nuclease stability and activity of the siRNA[22]. The two last linkages at the 3' end were joined by phosphorothioate linkages, which both confer nuclease stability and enhance the uptake of the siRNA. In the benchmark asymmetric siRNA design used in the Khvorova lab, the sense strand is a 15mer oligonucleotide with two PS linkages at the 5' end; the sense strand thus has a 5-nucleotide overhang.

This overhang is the target for hybridization with the PNA peptide carrier (**Figure 2-23**). However, a 5 base pair duplex is unlikely to be sufficiently stable for biological use. Thus, four PNA lengths were investigated, the sense strand being shortened when the PNA strand was lengthened. The different designs were therefore: 15mer RNA : 5mer PNA | 14mer RNA : 6mer PNA | 13mer RNA : 7mer PNA and 12mer RNA / 8mer PNA. The siRNA was expected to keep most of its activity even with the RNA-like portion of its sense strand shortened to 12 nt.

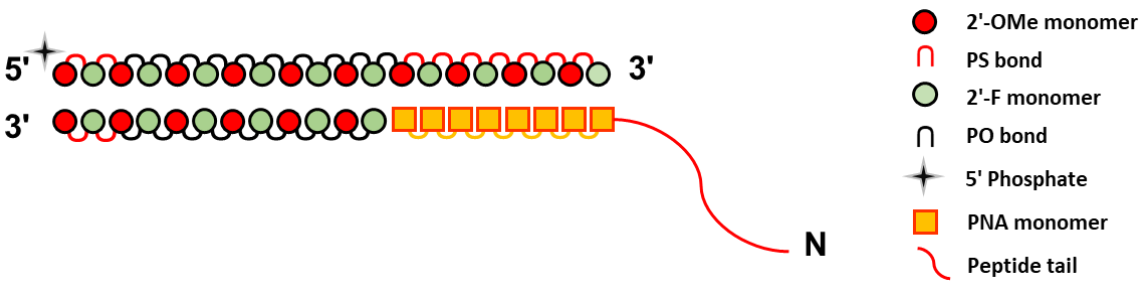


Figure 2-23: Design of the PPT-siRNA. N represents the N-terminus of the peptide. The PNA length is shown here as 8 nucleotides but varied from 5 to 8 nucleotides.

Four PNAs of appropriate length and sequence to be exactly complementary to the single-stranded 3' end of the antisense strand were therefore synthesized using standard PNA synthesis conditions on an Expedite 8909. The PNAs were annealed to an 8mer RNA of the same chemistry and sequence as the 3' end of the antisense strand (in 1xPBS at a concentration of 1µM). Melting studies were carried out on the duplexes. Melting temperature for the PNAs were fairly low (**Table 2-12**) but the 8mer was estimated to have a strong enough binding for the desired design.

Size	Sequence	T_m
8mer	aac agc aa	39.5
7mer	aac agc a	34.5
6mer	aac agc	19.5
5mer	aac ag	NA

Table 2-12: Melting temperature of the PNA of different length. Calculation of T_m was done by the derivative method. The mean and standard deviation of the T_m were calculated from 3 experiments (1 µM, in 1xPBS).

The peptide chosen for this study was a $(RXR)_4$ peptide, R being arginine and X being aminocaproic acid (6-aminohexanoic acid). This peptide has been reported as a good cell penetrating peptide (CPP) showing limited toxicity while enhancing delivery to a variety of cells.[122, 145] Our synthesis protocol involves dissolving Fmoc-protected amino acid monomers in NMP or DMF at 0.2 mM, however, the Fmoc protected aminocaproic acid we used was precipitating in DMF and was insoluble in NMP. Aminocaproic acid was therefore replaced by AEEA PNA linker (O), this change is expected to enhance the hydrophilicity of the PNA-Peptide while keeping the other properties. An AEEA linker was also used to separate the PNA from the peptide part (**Table 2-13**).

Name	Sequence	Expected Mass	Observed Mass
P^{p}	RORRORRORROR-O-aac agc aa	4163.34	4162.0
P	aac agc aa	2187.01	2186.9
S	AAU UCC AUC GU$_s$G	3868.45	3867.6
S^{c}	AAU UCC AUC GU$_s$G-TegChol	4624.42	4624.05
S^{15c}	C$_s$A$_s$A AUU CCA UCG U$_s$G$_s$A-TegChol	5654.20	5653.16
A	P-U$_s$C$_s$A CGA UGG AAU U$_s$U$_s$G$_s$ C$_s$U$_s$G$_s$ U$_s$U	6730.50	6729.72

Table 2-13: Sequence and mass of the oligonucleotides used in the sPtail study. Lower case is PNA monomer, P is a 5' phosphate, green letters are 2'-Fluoro-RNA nucleotides, red letters are 2'-OMe-RNA nucleotides, Cy3 is a Cyanine 3 dye and TegChol is a cholesterol linked by a tetraethylene glycol linker, O is 2-(2-(2-aminoethoxy)ethoxy)acetic acid) and R is arginine.

Based on the binding affinity results, only the 8mer PNA was used to form three-component duplexes. The duplexes were annealed at 50 µM by adding equimolar concentration of the PNA and heated for 1 minute at 95°C before being cooled at room temperature. The integrity of the duplexes was checked in the first place by electrophoresis (native PAGE at room temperature) but the positively charged peptide wasn't compatible with this experiment.

The three-component duplexes were then analysed on HPLC under "native" conditions. The HPLCs were run at 15°C on a C18 column with 50mM TEAA in water as buffer and an increasing gradient of acetonitrile. These conditions were suitable for analysis of the duplexes without a cholesterol group conjugated onto the sense strand.

Name	Peptide	PNA	Cholesterol	Length sense strand
A-S-P^{p}	✓	✓	x	12
A-S^{c}-P^{p}	✓	✓	✓	12
A-S-P	x	✓	x	12
A-S^{c}-P	x	✓	✓	12
A-S^{c}	x	x	✓	12
A-S	x	x	x	12
A-S^{15c}	x	x	✓	15

Table 2-14: Table of the 3-parts duplexes used in the experiment. A is antisense strand, S is sense -strand, ^{c} is cholesterol, P is PNA and ^{p} is peptide, see **Table 2-13** for sequence of each component of the duplex.

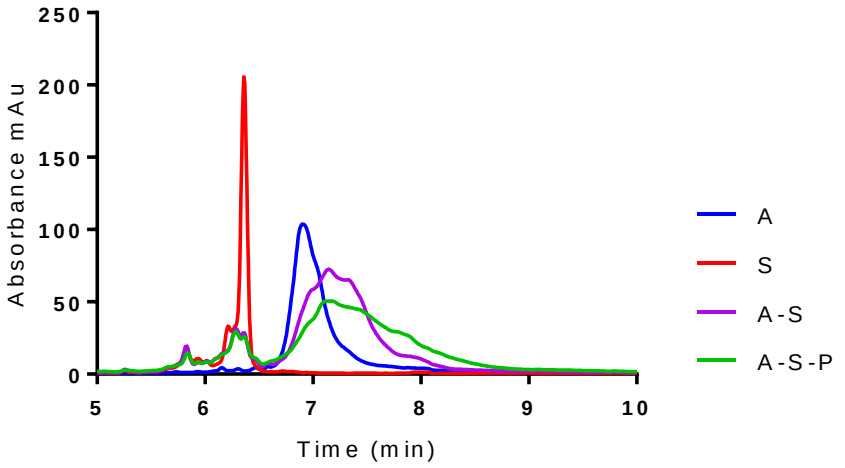


Figure 2-24: RP-HPLC trace of the A-S-P 3-parts duplex.

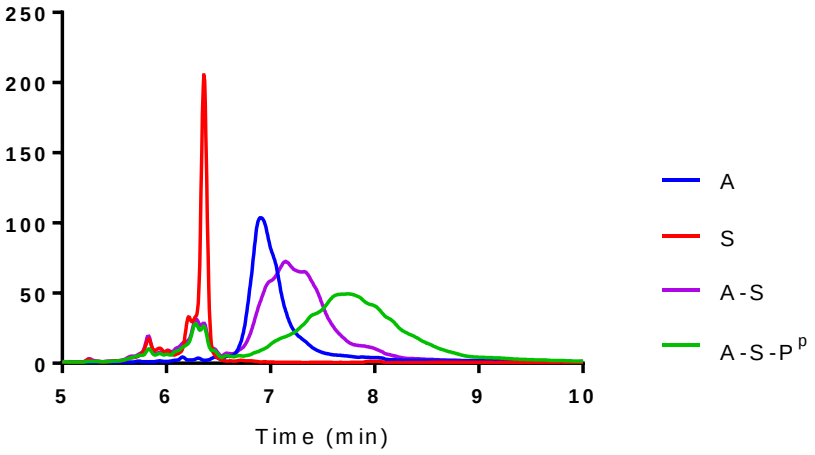


Figure 2-25: RP-HPLC trace of the A-S-P^{p} 3-parts duplex.

The green lines in **Figure 2-24** and **Figure 2-25** represent the 3-component duplexes while the purple lines show the duplex without the PNA or PNA-peptide associated. In both cases, the duplex peak is shifted to the right, with a greater shift visible for the complex including the PNA-peptide. This longer retention time can be partly explained as being due to the added mass (since larger molecules tend to migrate more slowly, when polarity is kept constant), but the peak was also somewhat more sharply defined, consistent with a conformational change or other effect from electrostatic interactions.

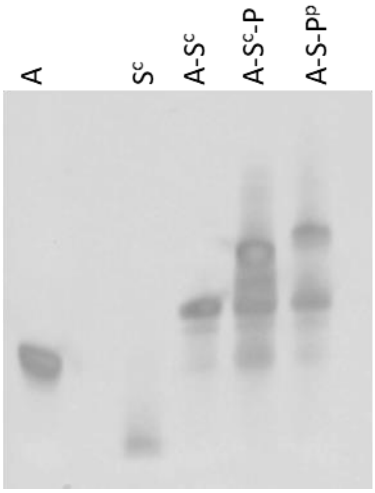


Figure 2-26: 20% native PAGE of the cholesterol 3-part duplexes. Electrophoresis run at room temperature, and gel was visualized with Sybr Gold. A is antisense strand, S^{c} is cholesterol conjugated sense strand, P is PNA, and P^{p} is PNA-Peptide. A-S^{c} is duplex, A-S^{c}-P and A-S^{c}-P^{p} are 3 part duplexes as descried in **Table 2-14**. The gel shows partial formation of 3 part duplex under the gel conditions.

Figure 2-26 shows the partial formation of the 3-component duplexes containing the cholesterol-modified sense RNA. Formation of the 3component duplexes is apparent but appears to be in equilibrium between the 3-component duplex comporting PNA and the 2 parts duplex seems to be present. The compounds were however, declared fit for carrying out the experiments.

2.4.2 Physical properties of PPT-siRNAs

The biophysical properties of the 3-component duplexes were investigated by UV thermal melting and circular dichroism experiments. The presence of the PNA doesn't change the melting temperature of the duplexes and the melting curves all had the standard shape. T_m results are summarized in (**Table 2-15**). These results are expected as the melting temperature for the antisense/sense duplex is higher than the PNA tail/antisense duplex. However the PNA strand dissociates earlier showing a more gradual melting curve representing several state in dissociation and not a classic 2-state model (Appendix C). A-S^c-P^p present a large standard deviation and a low T_m, this could be due to an instability of the 3 parts duplex containing the peptide and the cholesterol, one bulky and one charged molecule could have an additive effect on the duplex by hindering its formation and/or its dissociation. **Figure 2-27** shows A-S^c-P^p with the lowest height at 265 nm which could confirm this hypothesis.

Name	T_m	stdev
A-S-P^p	66.3	0.4
A-S^c-P^p	62.2	1.6
A-S-P	63.7	0.3
A-S^c-P	64.8	0.6
A-S^c	65.9	0.3
A-S	65.6	0.3

Table 2-15: Melting temperature of the PPT-siRNA 3-component duplexes. Calculation of T_m was done by the hyperchromicity method. The mean and standard deviation of the T_m were calculated from 3 experiments (1 µM, in 1xPBS).

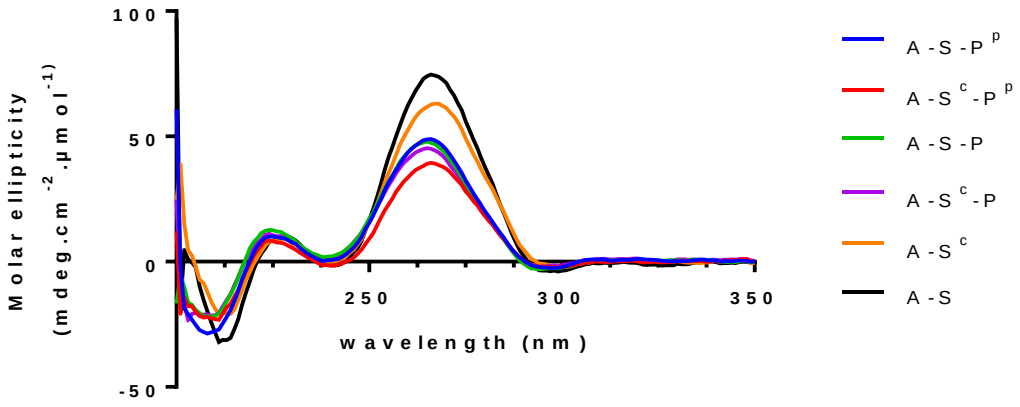


Figure 2-27: CD spectra of PPT-siRNAs, taken at 4°C in 1xPBS.

The CD spectra of the 3-component duplexes are all comparable, showing a classic A-form profile, and the peptide and the cholesterol do not seem to have any influence on the overall helical structure. However, the height of the 265 nm peak is significantly lower for all the 3-component duplexes relative to the 2-component duplex controls (A-S^{c} and A-S). This is surprising since the 3-component duplexes contain more stacked bases, and the intensity of the ~260 nm CD peak often correlates with the extent of base stacking. Clearly, as observed by other authors,[146] the interpretation of CD spectra is more complex than can be conveyed by comparing peak heights.

2.4.3 In vitro activity of PPT-siRNAs

The efficacy of the PPT-siRNA design was tested by targeting PPIB in HeLa cells. **Figure 2-28** shows the knockdown of PPIB mRNA by PPT-siRNA for one biological replicate by passive delivery. These are preliminary results and still need to be confirmed by repetition. The dark blue line shows the activity of the benchmark compound used by our collaborator (A-S^{15c}): this compound is essentially the same scaffold as the one we used for the PPT-siRNA except that the sense strand is a 15mer with a TEG-cholesterol on the 3'-terminus, and neither peptide nor PNA are included.

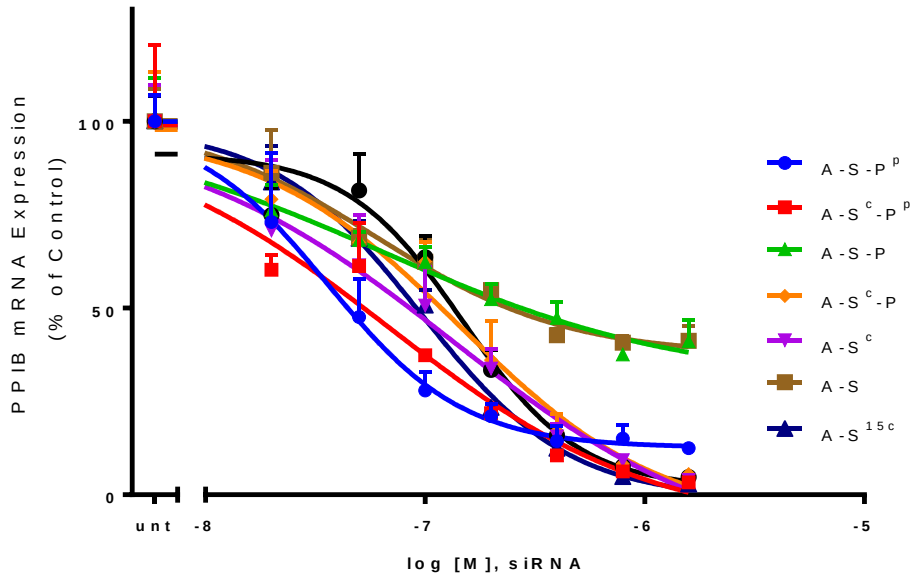


Figure 2-28: Silencing of PPIB mRNA by PPT-siRNAs. Dose-response of the PPIB PPT-siRNA in HeLa cells in 1:1, OptiMEM:DMEM + 6% FBS, plated at 10 000 cells per well in 96 well-plate. Cells were lysed after 72 h. Quantification of mRNA using the Quantigene 2.0 assay. Data is normalized to gene (HPRT) and presented as a % of untreated control. Error bars represent mean ± SD (n=3 wells).

In this experiment both cholesterol and peptide are expected to enhance cell uptake and neither of them is expected to have a poor uptake therefore poor silencing activity. Thus, as expected, the compounds A-S-P and A-S have poor activity. Duplex A-S-P, which includes the PNA 8mer is comparable to A-S, suggesting that the PNA strand in and of itself has no effect in terms of uptake or silencing activity.

All compounds possessing cholesterol show good silencing activity as expected. More interestingly, all peptide tagged compounds, including those without cholesterol, also show good silencing activity. For example, the activity of compound A-S-P^p shows silencing in the same order of magnitude as the benchmark compound. This is very promising for the PPT-siRNA platform and justifies the synthesis and testing of additional PNA-peptides, siRNA sequences, and biological systems including animal models.

1.4.4. PPT-siRNA summary

An 8-base PNA, alone or with a CPP conjugate, can be annealed to the 3'-end of the antisense strand of an asymmetric siRNA without compromising siRNA function. The simplicity of PNA-peptide synthesis, in the context of this PPT-siRNA strategy, would increase the throughput of screening peptides for targeted delivery of siRNA. The fact that a highly cationic peptide can be used in our assay is particularly promising: cationic peptides can form aggregates or cause other problems when complexed or conjugated with negatively charged oligonucleotides, and we expect that neutral or anionic peptides will be compatible with this platform given that the more challenging cationic peptides were successful. Cationic, neutral and anionic peptides are all promising for the purposes of targeted delivery.[147]

Chapter 3: **Synthesis of DNA-PNA chimeras**

3.1 Background on DNA-PNA chimeras

3.1.1 Interest of PNA-DNA chimeras

The discovery of PNA in 1991 by Nielsen *et al.*[88] and the further development of this class of oligonucleotide showed very interesting properties and unfortunate limitations. Some of these limitations are the inability of PNA to recruit RNase H and their poor membrane permeability.[117, 148] In 1995 Petersen *et al*. published the first DNA-PNA chimeras:[149] the concept of the chimera was based on the idea that combining DNA with PNA would provide the advantages from the two models while limiting their drawbacks. The most common designs of DNA-PNA chimeras investigated have been the 5'-DNA-3'-PNA-$CONH_2$, the NH_2-PNA-5'-DNA-3', and the 5'-DNA-3'-PNA-5'-DNA-3' (**Figure 3-1**).

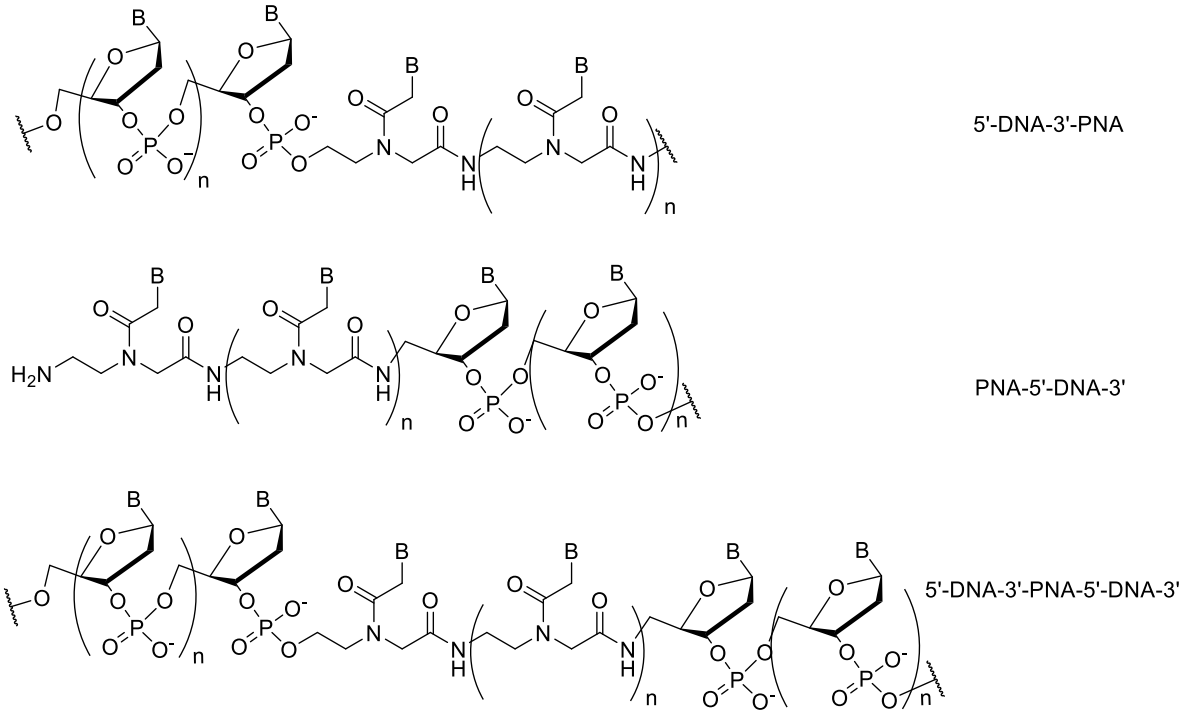


Figure 3-1: Most common designs of DNA-PNA chimeras.

All of these designs of chimeras are able to recruit RNase H to efficiently knock down a targeted RNA.[114, 150] Another very interesting feature is the excellent resistance against 3'-exonucleases that the PNA in the 3' position confers to the ASO. Such design of chimera is fifty times more resistant to nucleases than their DNA equivalent.[151] These two features make them very interesting for eventual ASO therapeutic applications.

Synthesis of DNA-PNA chimeras

Other types of chimeras have been investigated showing interesting properties but lacking the combination of recruiting RNase H and stability to nucleases. An alternating DNA/PNA[152] showed good stability in serum as well as interference with reverse transcription and disturbance of Protein/DNA interaction but this type of design can only be used for splice switching or steric blocking technology. Another chimera design made with the insertion of one or more PNA monomer in the middle of a DNA strand have been investigated.[153] It shows a big drop in binding affinity for its RNA target (ΔT_m -10.5°C for one base), but does show RNase H activity.

To be relevant as antisense oligonucleotides, DNA-PNA chimeras have to be taken up by cells. In 1998 Uhlmann et al. looked at the uptake of DNA-PNA chimeras by Vero cells. The 5' end of the oligonucleotides was labelled with 35S-γ-ATP and the oligonucleotides were incubated with the cells for 24 h at 37°C. The cells were then washed and the uptake was measured by scintillation counting. This method gives the relative amount of cell-bound oligonucleotide but not functional uptake: it did not distinguish between membrane-bound, endosome-entrapped and cytosolic oligomers. This study showed similar but lower uptake of the DNA-PNA chimeras compared to DNA oligonucleotide.

3.1.2 Affinity: The drawback of existing PNA-DNA chimeras

The binding affinity of the different designs of chimeras was studied extensively in the early days. The binding affinity is usually reported as comparison between the DNA/DNA duplex and a PNA/DNA duplex.[151] However, three main factors influence the melting temperature of chimeras with their counterpart. They are: the ratio of PNA to DNA in the chimera, the chemical structure of linker between the DNA and PNA parts, and the sequence of nucleobases.[114]

The melting temperatures reported in the early study about DNA-PNA chimeras are very diverse.[114, 149, 154-157] This is due at least partly to fact that historically the PNA parts of the first chimeras synthesized were exclusively made with pyrimidine bases.[149, 154, 156, 157] However, a poly-T PNA/poly-A DNA duplex is an excellent acceptor for a third poly-T PNA strand which ultimately forms PNA/DNA/PNA triplexes after invasion of the double strand.[88] Some of the early reports of melting of chimeras are likely to have encountered this phenomenon, potentially without recognizing it, which could explain the disparity of the results observed.[114, 149, 154, 156, 158] The effect of the chemistry of the linker on the binding affinity has been investigated by Greiner *et al.* in 1999 and 2002.[159, 160] In 1999 they published a study testing the influence of the nature of the atom linking the 3'-P of the DNA to the PNA part on the binding affinity.[159] They showed that the best affinity is obtained by linking the PNA part to the DNA part with oxygen instead of nitrogen or sulphur[159] (**Table 3-1** and **Figure 3-2**). Then in 2002 they showed that instead of using an aminoethylglycine based linker which is the best linker for DNA-PNA chimeras (**Table 3-1**), using

an aminobutylglycine based PNA linker in 2'-OMe-PNA chimeras could raise the binding affinity (+2°C compared to ethyl linker).

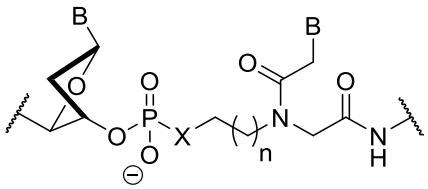


Figure 3-2: DNA-PNA chimera junction.

X	n	T_m DNA (°C)	T_m RNA (°C)
O	1	54.2	54.0
N	1	51.6	53.0
S	1	52.1	52.7
O	2	55.5	54.8
O	3	54.0	53.8
O	4	53.8	53.8

Table 3-1: Melting temperature of 6-6 DNA-PNA chimeras against complementary DNA and RNA. Chimera sequence 5'-ACA TCA tgg tcg-$CONH_2$. Buffer: 140mM KCl, 10mM NaH_2PO_4, 0.1 mM Na-EDTA, pH 7.4. Data are from Greiner *et al.* 1999 and 2001.

These different studies highlight the fact than the melting temperature of chimeras with their targets is dependant of the combination of the chemistry of the two nucleotides making the junction between the DNA and the PNA part. It is interesting to note than what is improving the affinity for DNA target is not necessarily improving affinity for RNA target.

One of the hypotheses about the relatively low melting temperature of chimeras is that the helical structure of PNA is not matching with the helical structure of DNA or RNA and the difference causes tension in the strand. Supporting this hypothesis, it is interesting to note that a mismatch at the linker position destabilizes less than in any other position[151, 161] and that chimeras have a better affinity for RNA than DNA. Furthermore PNA/RNA duplexes are structurally closer to RNA/RNA duplexes than RNA/DNA hybrid duplexes[141, 154, 162], the two parts of the chimeras are thus striving to get different helicoid duplex structure. The linkage thus plays an important role in the field of research in order to enhance the properties of the chimeras.

3.1.3 Prior approaches to synthesis of PNA-DNA chimeras

Syntheses of chimeras have been reported following three different strategies that will be discussed below. The first strategy is direct in-line synthesis on solid support. This strategy requires PNA monomers that contain protecting groups compatible with standard phosphoramidite DNA synthesis. The second one is to do a two-step in-line synthesis with protective group swap on solid phase which permits the use of commercially available PNA

monomers. The third strategy is to do a templated synthesis of chimeras (ligating PNA and DNA parts that were synthesized separately).

The first and the most documented strategy to make DNA-PNA chimera is the on solid-phase in-line synthesis using compatible chemistry between the PNA and DNA. This strategy usually uses monomethoxytrityl/benzoyl protective groups as the deprotection and synthesis condition are compatible with standard DNA synthesis.

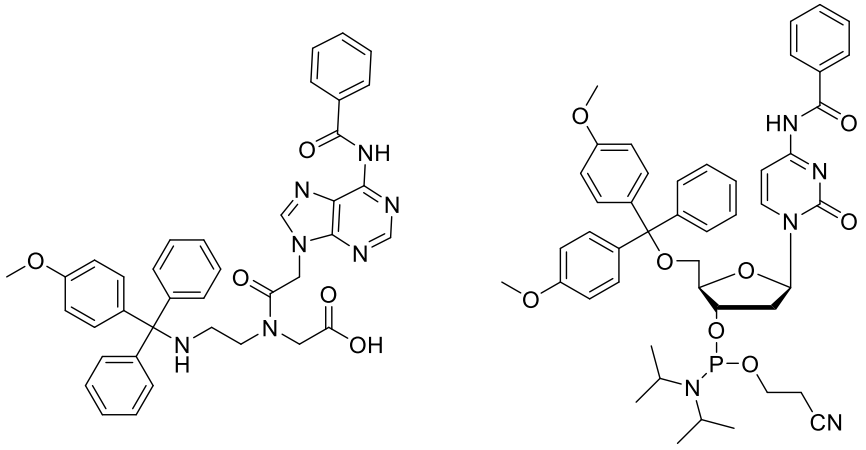
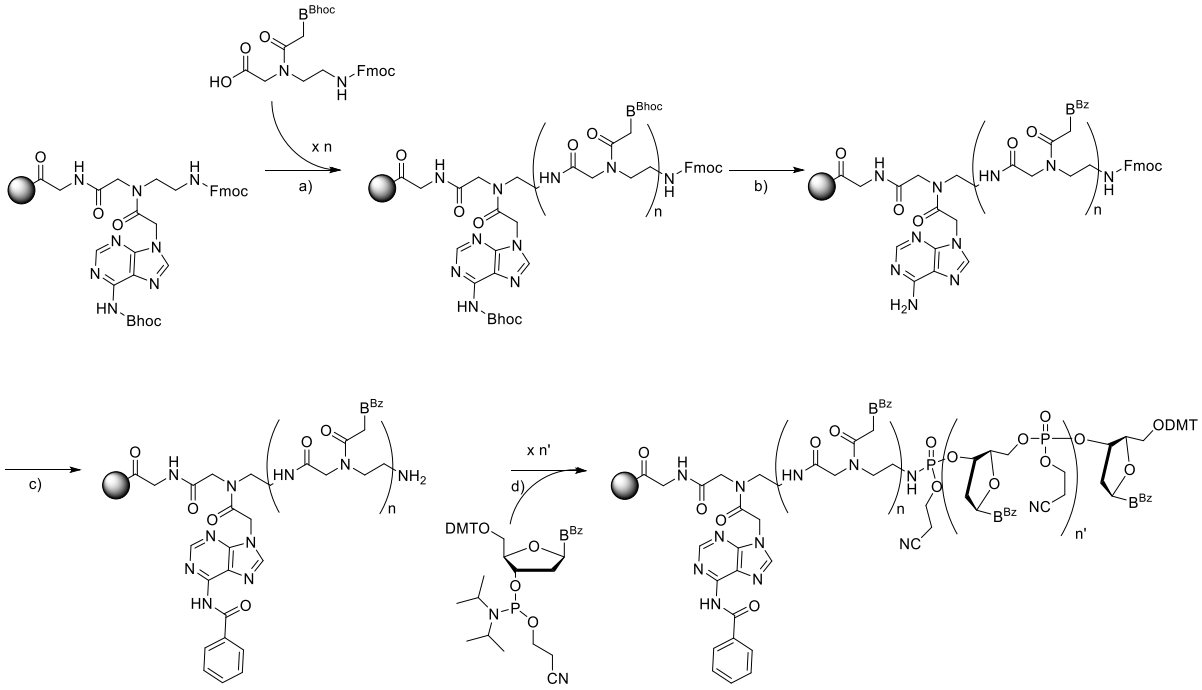


Figure 3-3: PNA-A and DNA-C protected monomers used in chimeras in-line synthesis.

The monomethoxytrityl (MMT) strategy can be carried out by several ways. The PNA part of the chimeras can be assembled manually,[163] on peptide synthesiser or on DNA synthesizer (Expedite 8909, Eppendorf Biotronik Ecosyn D-300 DNA[160] or ABI 394[161]). The DNA part is then grown on a DNA synthesiser using standard conditions. The solid phases used to grow the chimeras are CPG-OH,[149, 151] Tentagel-OH,[164] and HMBA.[161, 165] These solid phases are linked by an ester to either a glycine or aminohexyl succinic acid linker. Comprehensive comparisons of these solid supports have not been done and their advantages relative to other are unclear. Several coupling conditions have been investigated in the literature for the PNA synthesis and the combination of HATU and DIPEA has been taken as the most efficient and reliable.

The MMT-based PNA cycle consists of the following steps. First the resin is washed with DCM and then the MMT group is deprotected by a solution of 3% TCA in DCM and washed again by DCM. Then the solid phase is washed with DMF. HATU, DIPEA and PNA monomers are premixed before being pushed onto the column. The reaction mixture is in contact with the resin for 10 to 20 min and washed with DMF. Then, unreacted PNA strands are capped with a solution of acetic anhydride, DIPEA and lutidine in DMF. The resin is washed again with DMF and the cycle start again for the next PNA base. When the PNA part is finished the solid phase is taken as it is to be used in standard DNA synthesis condition. Cleavage and deprotection of the complete chimera are usually done by treatment with ammonium hydroxide at 55°C overnight.

The second strategy was proposed by Capasso *et al.* (**Scheme 3-2**) in 2001.[166] Standard commercially available Fmoc/Bhoc PNA monomers are used to grow the pseudo 3' of the PNA on solid-phase using standard PNA cycle. However, the Bhoc protecting group is labile during standard DNA synthesis condition. Capasso *et al.* showed that it was possible to deprotect the PNA part with a solution of 75% TFA in DCM and replace these with acetyl or benzoyl groups. To accomplish the protection on solid phase, the resin was treated with a solution of acetic anhydride in pyridine (2:3, v/v, 3 h) or with a solution of benzoyl chloride in pyridine (3:7, v/v, 6 h) if the PNA contains G. The big advantage of this strategy is the availability of the PNA monomers thus making it an affordable method even at mmol scale. The biggest drawback is the limitation in modification of that one can add to the structure of the PNA monomers. Furthermore, we attempted to follow this procedure but failed to afford the desired compound.



Scheme 3-1: In-line synthesis of PNA-DNA chimeras with protective group swap. a) regular automated PNA cycle. b) 75% TFA in DCM (w/w) for 1h at room temperature, then BzCl in pyridine (3/7 : v/v) for 6h at rt. c) Fmoc deprotection with 20% piperidine in DMF. d) Regular automated phosphoramidite DNA cycle.[166]

Thirdly the templated strategy was published in 1998 by Marcus Koppitz *et al.*[167] The study investigates different condition to make a chimera with the help of template. The 3'-phosphoramidate bond formation as well as the 5' phosphoramidate bond formation are performed with preactivation of the DNA strand as a phosphoimidazolide using EDC in the presence of imidazole. Different concentrations (0 eq, 0.1 eq, 1 eq, 10 eq) of the template, for the DNA part and the PNA part have been tested. They observed that the presence of template is necessary for the reaction to happen and that the best yield (80%) is obtained with 10 equivalents

Synthesis of DNA-PNA chimeras

of DNA template, 1 equivalent of DNA and 10 equivalents of PNA. The chimeras were made of 6mer DNA and 8mer PNA.

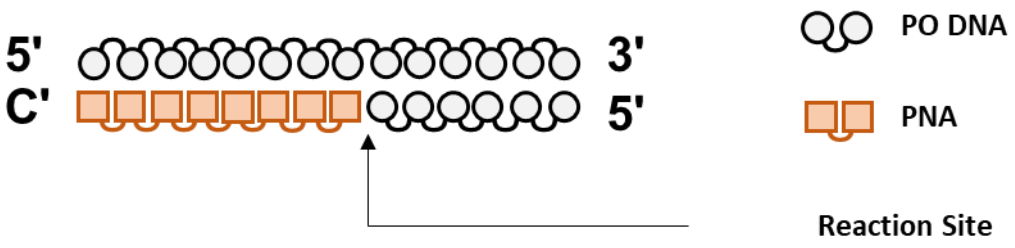


Figure 3-4: Templated synthesis of PNA-DNA chimera. 10eq of DNA template and PNA are used for 1eq of DNA. 80% yield at 4˚C after 2 days at 60nmol scale.

This method is robust and works very well for small scale. However, this method requires synthesizing and purifying three different oligonucleotides to make a fourth one which needs purification as well. This method is material and time consuming and limited in scale. Due to the limitation in scale I didn't try this method to synthesize chimeras, though it proved useful in the following preliminary studies by Dr. Watts.

3.1.4 Preliminary data: a possible solution to the binding affinity problem

Preliminary computational studies performed by Dr. Watts showed that the addition of one LNA at the 3' end of the DNA strand linked to PNA is structurally compatible with chimeras. Melting studies on DNA-LNA-PNA chimeras which have been synthesized by template ligation[167] (method 3 above) showed an increase in the ΔG when inserting an LNA at the junction.

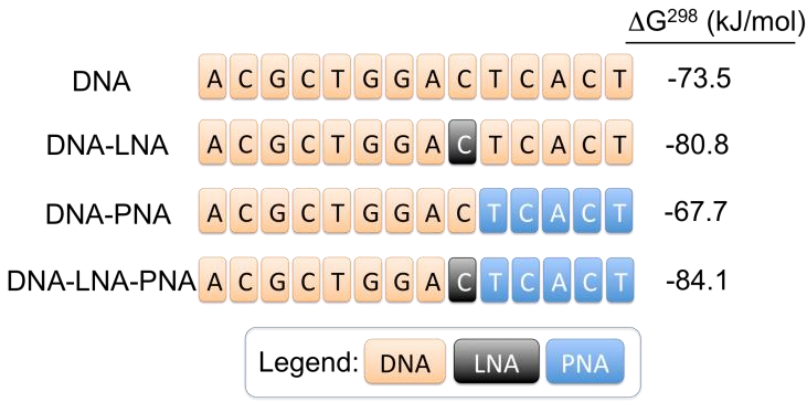


Figure 3-5: Preliminary binding studies for DNA-LNA-PNA.

3.1.5 Project objectives

The aim of this project is to investigate the possibilities of optimized PNA-containing chimeras as antisense oligonucleotide therapeutics. As stated previously, the synthesis of DNA-PNA chimeras

has been achieved following different routes and several studies have reported their binding properties. Chimeras are able to recruit RNase H and their very good serum stability and apparent ability to be taken up by cells has been published. However, knockdown of RNA in cells or animals has never been shown before. The present work presents the tailoring of the DNA-PNA chimeras to achieve efficient knockdown of non-coding RNA. To achieve this goal the MMT/benzoyl synthesis strategy was chosen. Therefore, synthesis of the PNA monomers and PNA linkers needs be achieved in house due to their unavailability on the market. Synthesis of monomers and oligomerisation of the chimeras are optimized at multiple steps. New designs of chimeras are investigated in comparison with the published work. The chimeras will be developed in the gapmer fashion with a small wing of LNA on the 5'-end of the oligonucleotide. The junction between the DNA and PNA in the strand are a focus of our investigation as it has been identified as a critical point for the stability of chimeras with their RNA targets. The preliminary data obtained by Dr. Watts suggested that using a LNA nucleotide at the 3' part of the DNA to link with the PNA could enhance the affinity. Therefore, different combinations of PNA linkers and modified DNA nucleotides are investigated. The different designs are compared in terms of their binding affinity to complementary RNA along with their ability to recruit RNase H. Following, the knockdown of MALAT1 in cell culture is reported.

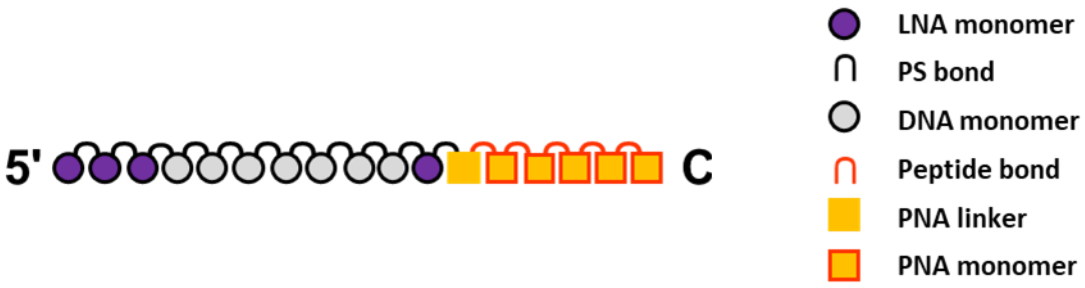
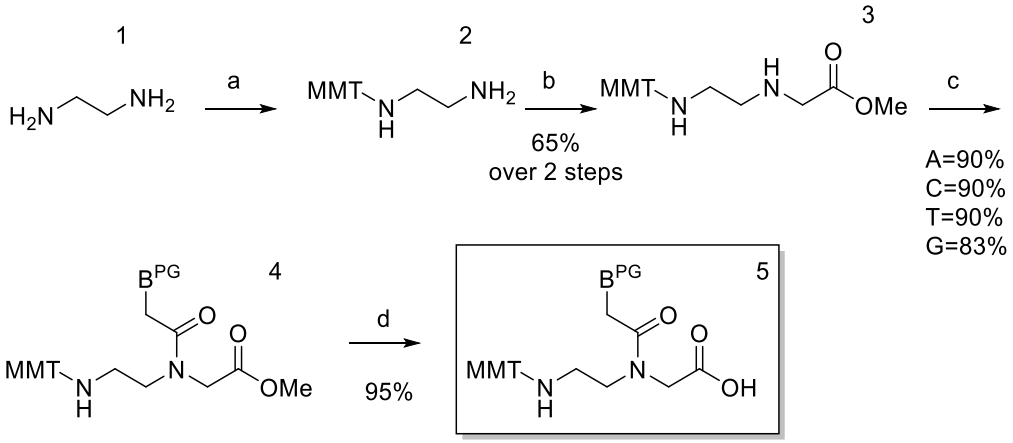


Figure 3-6: Design of an optimized design for DNA-PNA chimeras

3.2 PNA monomer synthesis

3.2.1 General strategy

To synthesise the PNA monomers compatible with in-line DNA synthesis, the work published by Musumeci *et al.*[168] in 2004 has been taken as a starting point.



Scheme 3-2: The route of Musumeci *et al.* for PNA monomer synthesis: a) MMTCl, pyridine, DCM; b) methyl bromoacetate, pyridine, DCM; c) nucleobase acetic acid, DIC/HOBt, NMM, DMF; d) Tetrabutyl ammonium hydroxide; B^{PG} represents an appropriately protected nucleobase: T, C^{Bz}, A^{Bz}, G^{iBu}.

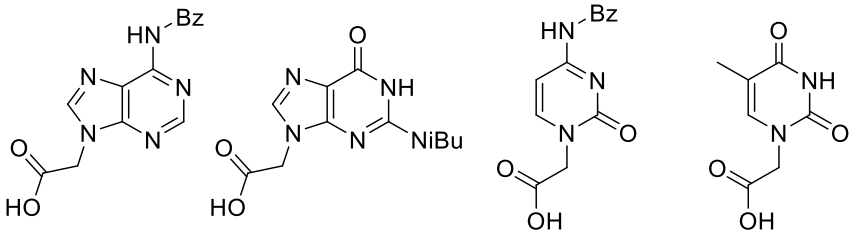
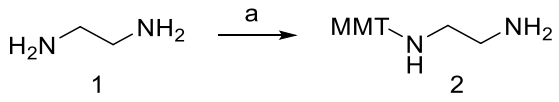


Figure 3-7: Protected nucleobases acetic acids used in the synthesis of PNA monomers.

3.2.2 Backbone preparation

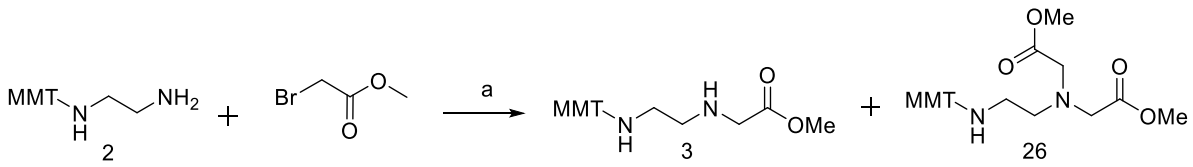
The first step is the selective protection of ethylene-1,2-diamine with monomethoxytrityl chloride. The reaction was achieved in dichloromethane with a nine-fold excess of 1,2-diaminoethane. The monomethoxytrityl chloride in pyridine is added dropwise into the reaction mixture at 0°C. Usually the reaction was then run overnight before being extracted with DCM and washed with water. The excess of ethylenediamine and extended time for the addition were identified as crucial to limit the amount of diprotected side product obtained.



Scheme 3-3: MMT protection of ethane-1,2-diamine. a) MMT-Cl in DCM.

The product (2) was afforded with 56% yield after chromatography on silica gel. However, routinely the crude was directly carried over to next step after extraction.

The second step of the backbone synthesis is the S_N2 substitution of the *N*-monomethoxytrityl-ethane-1,2-diamine on the methylbromoacetate. Due to reduced steric hindrance, the nucleophilic attack is preferentially performed by the terminal amine. However, the alkylated secondary amine present in the desired product is also nucleophilic enough to react once again with methyl bromoacetate producing the dialkylated side product (**Scheme 3-4**) which is removed by column chromatography. The steric bulk of the 4-methoxytriphenylmethyl group prevents the nucleophilic attack by the other secondary amine. In order to minimize the amount of side product, methylbromoacetate was added dropwise at 0ºC. The desired product (3) was recovered after chromatography on silica gel with 57% yield.



Scheme 3-4: Second step PNA backbone synthesis: 3-desired product, 26-dialkylated side product. a) NEt_3 in DCM.

3.2.3 Coupling with nucleobases

We initially attempted to follow the protocol of Musumeci[168] for nucleobase coupling. As such, the monomethoxytrityl-*N*-(2-aminoethylglycine)-methyl ester 3 was coupled with the nucleobase acetic acid with DIC/HOBt and NMM. However, the yields obtained by this method were poor and the side product, diisopropylurea, was very difficult to separate from the desired product. Both precipitation and silica gel chromatography failed to afford the pure product. Therefore, alternative conditions were investigated (**Table 3-2**). The use of DIC/HOBt, HBTU, EDC/HOBt and T3P were compared. All reactions were carried out in DMF overnight at room temperature.

Coupling Reagent	Base	Nucleobase	Yield [(a)]	Purity (NMR)
DIC/HOBt	NMM	Thymine	65%	< 57%
DIC/HOBt	NMM	Cytosine	73%	< 66%
DIC/HOBt	DIPEA	Thymine	71%	< 47%
EDC.HCl/HOBt	DIPEA	Thymine	54%	> 95%
T3P	DIPEA	Thymine	77%	> 95%
T3P	DIPEA	Cytosine	61%	> 95%
T3P	DIPEA	Adenine	86%	> 95%
T3P	DIPEA	Guanine	34%	> 95%
HBTU	DIPEA	Thymine	46%	< 70%

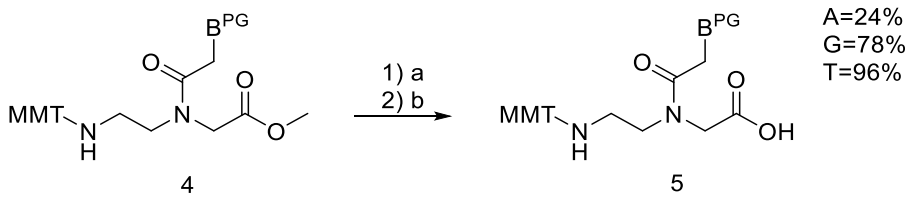
Table 3-2: Amide coupling conditions investigated for coupling of nucleobase acetic acid on PNA backbone and their results. [(a)]: isolated yield of purified product.

EDC/HOBt affords the desired product with high purity by precipitation from hexane but with a poor yield. HBTU affords the desired product but purification by column chromatography was necessary to afford a clean product. The yield obtained by HBTU was unsatisfactory. Finally, T3P (propylphosphonic anhydride) as a 50% solution in DMF was used. This reagent was previously used for Boc protected PNA monomer synthesis by Breipohl in 1997.[169] The yields were better than the other coupling conditions investigated and the pure desired product was obtained by simple precipitation from hexane.

3.2.4 Carboxylic acid deprotection

The deprotection of the methyl ester was initially achieved in dioxane with a solution of tetrabutylammonium hydroxide[168]. The reactive mixture was then neutralised with an aqueous solution of 1M $KHSO_4$. However, this method was troublesome as the slightest mistake in monitoring the pH would deprotect the monomethoxytrityl, the resulting side product being difficult to separate by chromatography. Then another method based on Will *et al.*[164] using an aqueous solution of sodium hydroxide in methanol and further neutralization by a milder acid, Dowex pyridinium was been investigated and gave satisfying results. Dowex pyridinium was obtained by treating Dowex H^+ with pyridine to saturate the benzene sulfonic acid on the surface of the resin with pyridinium as a counter-cation. Using this resin to neutralize the solution offered

sufficiently mild conditions to selectively neutralize the reaction mixture without deprotecting the MMT group.



Scheme 3-5: Deprotection of the ester group of PNA monomers. a) NaOH in MeOH; b) Dowex pyridinium.

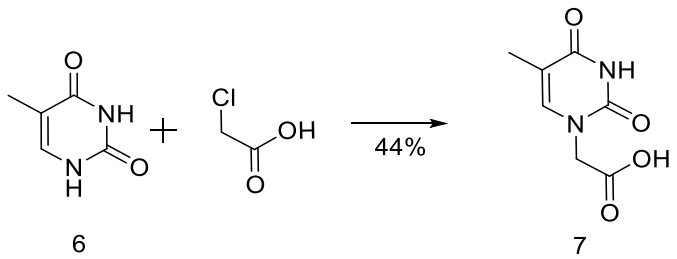
All monomers but C have been obtained after column chromatography ready to be used in solid phase synthesis, though at low yield. The use of aqueous NaOH raises the danger of deprotection of the benzoyl protecting group on A, G and C, and so the reaction must be monitored very carefully. This point was particularly troublesome when working with C. The conditions used to deprotect C were saponification with the help of tetrabutylamonium hydroxide and neutralization by Dowex pyridinium. This method proved more reliable and was used for the deprotection during the linker synthesis.

3.2.5 Nucleobase preparation

In order to couple the nucleobase to the backbone **3** the synthesis of nucleobases with an acetic acid on N^1 for the pyrimidine and on N^9 for the purine has to be achieved. Furthermore, a suitable protecting group for PNA and DNA solid phase synthesis has to be selected for exocyclic amines of adenine, guanine and cytosine.

3.2.6 Synthesis of thymine acetic acid

The thymine is generally the easiest nucleobase to work with, because it doesn't need a protective group in solid phase synthesis and also because the N^1 position is the most nucleophilic and thus easy to functionalise.

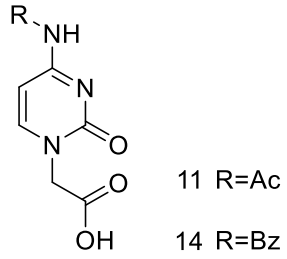


Scheme 3-6: Synthesis of thymine acetic acid. Reaction conditions: 3.6M KOH_{aq}.

Thymine acetic acid (7) was obtained with 44% yield by reaction with chloroacetic acid in alkaline conditions following procedure from Jones *et al.*[170] (**Scheme 3-6**). The product was afforded by precipitation at pH 3-4 and was purified by recrystallization from water.

3.2.7 Synthesis of N4-protected-1-carboxymethyl-cytosine

The target molecule is N-4 protected cytosine with an acetic acid on N1. The usual protecting groups for cytosine in solid phase synthesis are *N*-acetyl or *N*-benzoyl.

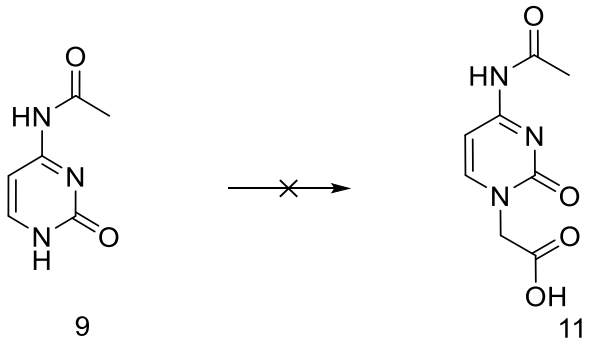


11 R=Ac

14 R=Bz

Figure 3-8: N4-protected-1-carboxymethyl-cytosine

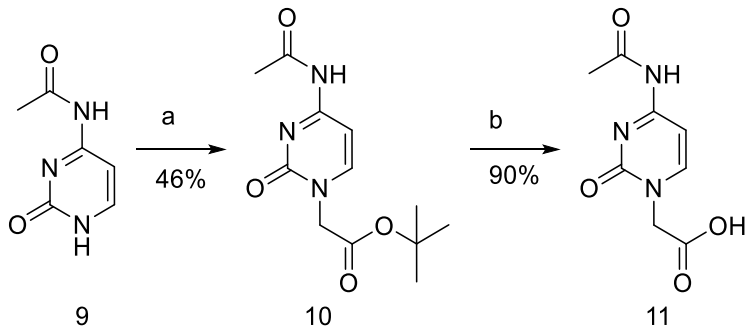
The *N*-acetyl protection of cytosine was achieved following the procedure from Okabe *et al.*[171] The cytosine was placed in pyridine with acetic anhydride and heated at reflux to afford the desired product 9 with 89% yield.



Scheme 3-7: N4-actyl cytosine carboxymethylation. Reagent and conditions: KOH $_{aq}$, chloroacetic acid.

Functionalisation of the N1 position was first carried out under similar conditions as those described for the thymine. However, the basicity of the reaction mixture led to deprotection of the N4 acetyl protecting group, therefore the desired product was not obtained (**Scheme 3-7**).

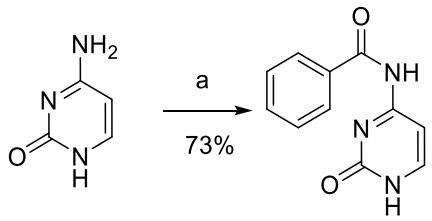
In a second attempt to synthesize N4-acetyl-carboxymethyl-cytosine (**Scheme 3-8**) we followed the procedure from Ferrer[172] by adding the tert-butyl bromoacetate to the acetyl cytosine in presence of DBU. The desired N4-acetyl-N1-tert-butoxycarbonylmethyl-cytosine (10) was recovered with 46% yield (reported 56%). DIPEA and K_2CO_3 were also investigated as alternative bases to DBU but the yields were lower (36% and 37% respectively).



9 10 11

Scheme 3-8: N4-acetyl-carboxymethyl-cytosine (Following Ferrer's[172] strategy): a) tert-butyl bromoacetate/DBU, b) neat TFA

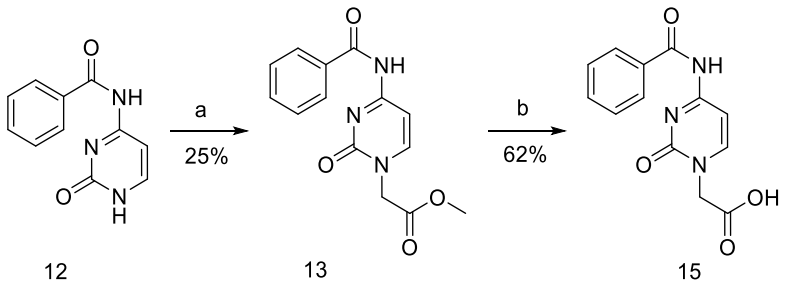
The deprotection of the tert-butyl ester to afford the acid (11) was achieved in neat TFA with 90% yield. The desired N4-acetyl-carboxymethyl-cytosine is obtained from cytosine with an overall yield of 34%. However, the last step involving TFA is too acidic for the N4-acetyl protecting group and it was not possible to avoid a partial deprotection of the cytosine. Furthermore, the N4-deacylated cytosine has proven difficult to separate from the desired product. The coupling of the N4-acetyl-carboxymethyl-cytosine with the backbone would adversely be affected by any presence of free amine.



8 12

Scheme 3-9: Cytosine benzoylation. Reagent and condition: BzCl in pyridine.

Therefore, another protecting strategy was investigated: a benzoyl group was used to protect the N4-amino group (**Scheme 3-10**). Benzoylation of cytosine was achieved with benzoyl chloride in pyridine to afford the desired product (12) with 73% yield (reported 62% by Finn156).



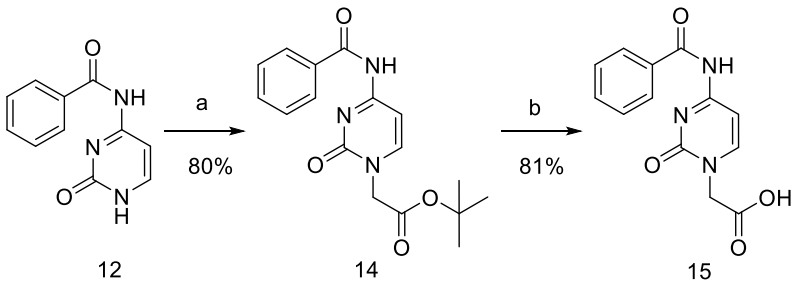
12 13 15

Scheme 3-11: N4-benzoyl-carboxymethyl-cytosine (Following Finn's strategy[156]): a) methylbromoacetate/K_2CO_3, b) NaOH.

Then Finn's strategy[156] was adapted to afford the N4-benzoyl-carboxymethyl-cytosine. Addition of methyl bromoacetate in presence of potassium carbonate was achieved with 25% yield (13) (**Scheme 3-11**). The poor yield is likely due because of side reactions. Namely, traces of debenzoylated desired product have been identified in LC/MS, suggesting that the conditions are partially deprotecting the base. Furthermore, NMR of the concentrated filtrate from the recrystallization suggests production of a by-product with addition of a second methylbromoacetate on the molecule.

The deprotection of the methyl ester to the carboxylic acid was performed with sodium hydroxide and neutralised by acid work up which afforded the desired product (15) in 62% yield. (2 steps yield 15%, reported 45%).

The overall yield of this strategy was not sufficient in order to use it routinely to make PNA-C monomer, therefore an adaptation of the method by Ferrer[172] for the N4-acetyl cytosine was attempted (**Scheme 3-12**).



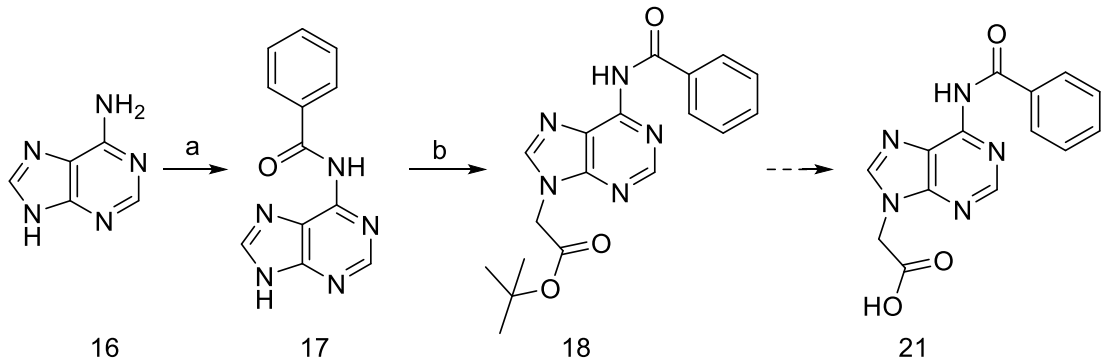
Scheme 3-12: Final strategy for N4-benzoyl-carboxymethyl-cytosine: a) tert-butyl bromoacetate, K_2CO_3, Cs_2CO_3 ; b) 4M HCl_{aq} in dioxane.

Tert-butyl bromoacetate was added to N4-benzoyl cytosine in DMF with K_2CO_3 and catalytic Cs_2CO_3 to afford the desired product (15) after recrystallization from EtOAc in 80% yield. Use of DIPEA instead of K_2CO_3/Cs_2CO_3 resulted in much lower yield (24%). Then the deprotection of the carboxylic acid was achieved in DCM by treatment with a solution of 4M HCl in dioxane. After precipitation in DCM, the desired product (15) was afforded in high purity in 91% yield. These conditions were giving the best yield with very easy purification and this route was therefore preferred to synthesise the protected cytosine acetic acid building block (15) for all future batches of monomer synthesis.

3.2.8 Synthesis of 6-N-(Benzyloxycarbonyl)-9-carboxymethyl adenine

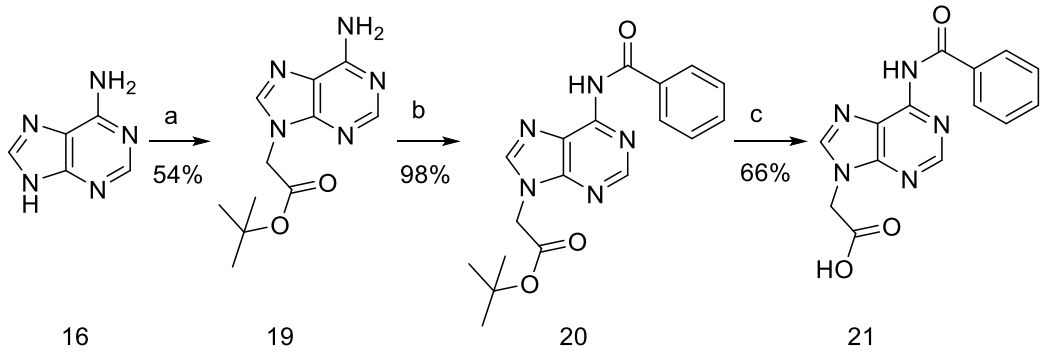
The preparation of the adenine monomer was investigated with the help of Dan O'Reilly, an MChem student in our laboratory at the time. In order to synthesize the benzoyl protected carboxymethyl adenine, the method from Thomson[173] was adapted (**Scheme 3-13**). The

benzoylation of N6 was achieved with benzoyl chloride in pyridine and the desired N-6-benzoyladenine (17) was recovered with 70% yield. Then the N-9 position was functionalised with *tert*-butyl bromoacetate in presence of NaH. However, the N-7 and N-9 isomers as well as second addition on N-6 were recovered. Isolation of the desired product (18) was reported to be possible by recrystallization from EtOAc/hexane[173] but our attempts at purification were unsuccessful.



Scheme 3-13: Thomson's[173] 6-*N*-(benzyloxycarbonyl)-9-carboxymethyl adenine strategy. Reagents and conditions: a) BzCl/pyridine, b) *tert*-butyl bromoacetate/NaH.

Another synthetic strategy inspired by Dueholm's protocol[174] (**Scheme 3-14**) was investigated to make the desired product. The adenine was first treated by *tert*-butyl bromoacetate in the presence of NaH. The 9-*tert*-butoxycarbonylmethyl adenine was recovered after recrystallization with a yield of 54% (19). Then the benzoylation was achieved with benzoyl chloride in pyridine to afford a mixture of mono and di-benzoylated adenine (20 and 20'). The product was treated for deprotection of the *tert*-butyl with TFA in presence of triethylsilane. The triethylsilane is acting here as a cation scavenger to avoid the loss of the second benzoyl group. The desired 6-*N*-(benzyloxycarbonyl)-9-carboxymethyl adenine was afforded with 44% yield over 2 steps from 19 (21).

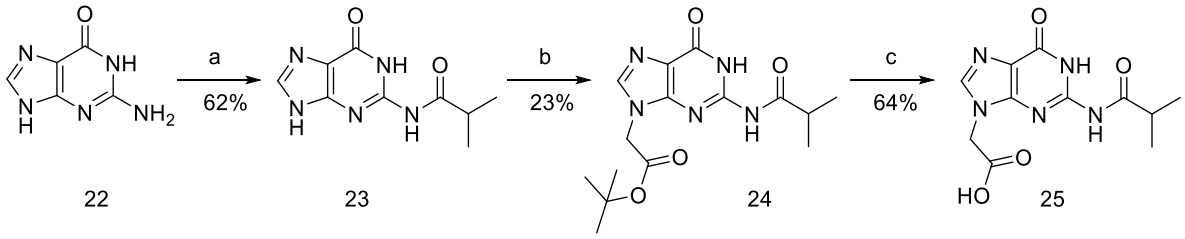


Scheme 3-14: 6-*N*-(Benzyloxycarbonyl)-9-carboxymethyl adenine strategy; a) *tert*-butyl bromoacetate/NaH, b) BzCl/pyridine, c)TFA, Et_3SiH.

These conditions were judged satisfactory for any further synthesis of monomers.

3.2.9 Synthesis of 9-(Carboxymethyl)-2-*N*-isobutyryl)guanine

In order to prepare the PNA-G monomer we adapted the approaches of Liu[175] and Timar.[176] The protecting group chosen for guanine's exocyclic amine, N2, was isobutyryl, consistent with standard oligonucleotide synthesis approaches since the 1960s.[177] A suspension of guanine in DMF was treated with isobutyric anhydride to afford the protected product in 64% yield (23). Then 2-*N*-isobutyrylguanine was reacted with *tert*-butyl bromoacetate to afford the desired N-9 isomer by precipitation with 23% yield (24). Finally, the removal of the *tert*-butyl group was achieved by treatment with TFA and triethylsilane. The desired 9-(carboxymethyl)-2-*N*-(isobutyryl) guanine was afforded with 65% yield (25).



Scheme 3-15: 9-(Carboxymethyl)-2-*N*-(isobutyryl) guanine synthesis strategy; a) isobutyric anhydride in DMF; b) NaH, *tert*-butyl bromoacetate; c) TFA, Et_3SiH.

These conditions were judged satisfactory for any further synthesis.

3.3 Synthesis of PNA hydroxyl linkers

The chemistry of the linker has an influence on the final thermal stability of the chimeras (see introduction on chimeras). Nevertheless, the need of a hydroxyl PNA linker is driven by the difficulty of synthesizing phosphoramidate. The terminal NH_2 of the PNA monomer is potentially protonated during the reaction with the nucleotide phophoramidite due to the acidity of the activators, the protonated amine is a terrible nucleophile, therefore the reaction is slow. Furthermore, the newly formed phophoramidite is very sensitive to nucleophile until oxidation.

Three lengths of backbone have been selected for investigation and a hydroxyl group has been chosen as reactive group for the following coupling with the sugar part of the chimera.

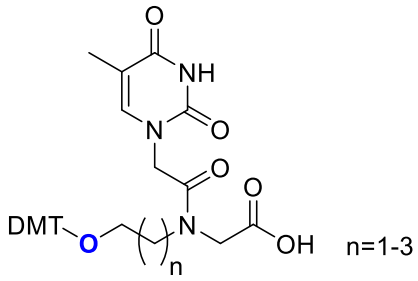
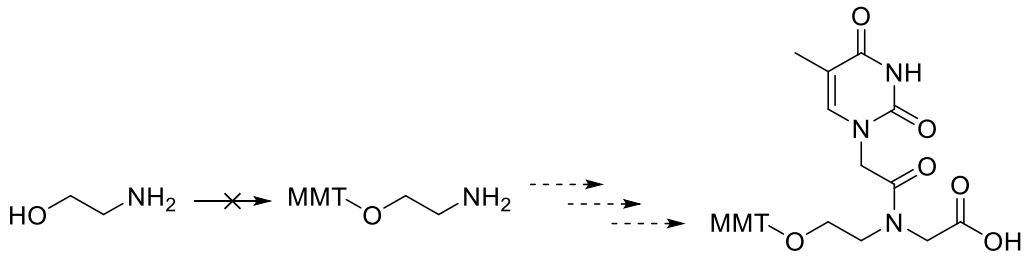


Figure 3-9: Structure of the PNA linker used in this study.

In order to synthesize the different linkers, four synthetic routes were investigated

3.3.1 Strategy 1: Tritylation first

The first strategy investigated to make the ethyl PNA linker was the method from Musumeci.[168] The first step was to protect the hydroxyl group of ethanolamine with a trityl group. We attempted to do this by using ethanolamine hydrochloride with MMT-Cl in pyridine. However, the MMT group reacted with the amino group instead of the hydroxyl group. The same protocol was attempted using ethanolamine instead of ethanolamine hydrochloride but the afforded product was the *N*-monomethoxytrityl-aminoethan-1-ol. The reaction was attempted again with prior treatment of ethanolamine by Dowex H^+ before adding the pyridine and the MMT-Cl. However, *N*-monomethoxytrityl-aminoethan-1-ol was recovered again.



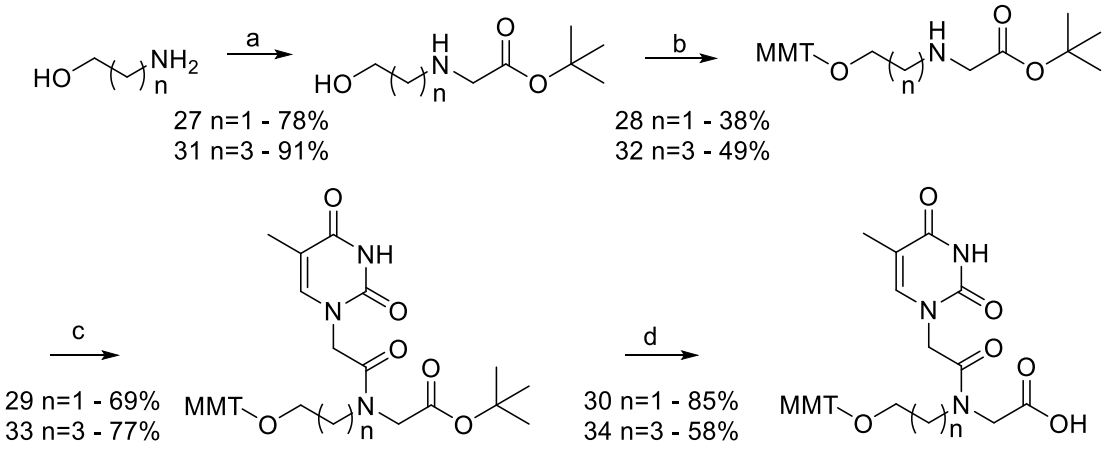
Scheme 3-16: General strategy from Musumeci *et al.*

This strategy was abandoned after the unsuccessful attempt to make the desired product.

3.3.2 Strategy 2: Nucleophilic substitution first

The second strategy attempted was modified from the approach described by Breipohl *et al*.[169] (**Scheme 3-17**). While Breipohl *et al.* added the trityl protecting group as the last step, in our own work the tritylation of the hydroxyl was done as the second step. This gave two advantages: it allows using a larger variety of coupling agents for the addition of the nucleobase (since the hydroxyl is protected) and it facilitates the monitoring of the reactions by TLC. Two different protecting groups for the carboxylic acid were tested: methyl ester and *tert*-butyl ester.

However, reaction with methyl bromoacetate failed to afford the expected methyl ester product. No identifiable products were recovered after extraction with ethyl acetate and washing with water even though a TLC of reaction mixture was showing formation of new compounds. Our hypothesis is that the product was lost during rotary evaporation.

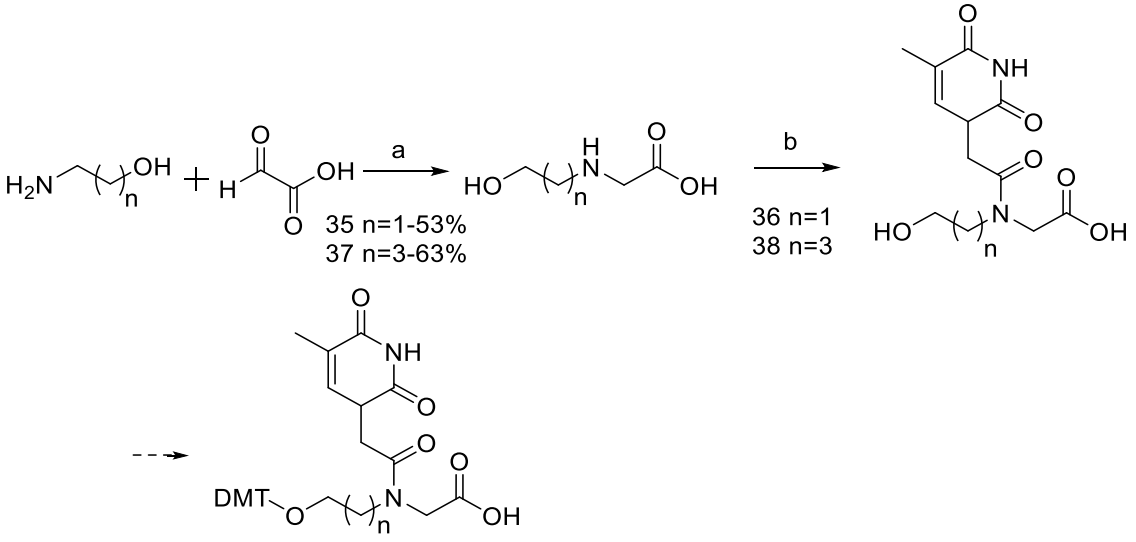


Scheme 3-17: Synthesis of PNA linker a) *tert*-butyl bromoacetate/NEt_3, b) MMT-Cl/Pyridine, c) T3P or EDC-HOBt /DIPEA/thymine acetic acid, d) NaOH/dowex pyridinium.

The synthesis route using the *tert*-butyl ester was performed in tandem with project student Rachel Greenhill (who carried out the synthesis for n=3, i.e. the "butyl PNA linker"). This strategy afforded the desired ethyl and butyl PNA linkers with very similar yield (**Scheme 3-17**). The overall yield for 4 steps was 16% for n=1 (30) and 20% for n=3 (34). Coupling yields using T3P were similar to those using EDC/HOBt.

3.3.3 Strategy 3: Reductive amination

The reductive amination method of Greiner *et al.*[159, 160] was attempted for the synthesis of the PNA linkers by Rachel Greenhill (**Scheme 3-18**).[178] The route begins by reductive amination of 2-aminobutanol with glyoxylic acid in water with H_2 as reducing agent and Pd/C as catalyst. As such, the *N*-(4-hydroxybutyl) backbone was successfully produced with a yield of 63% (37).



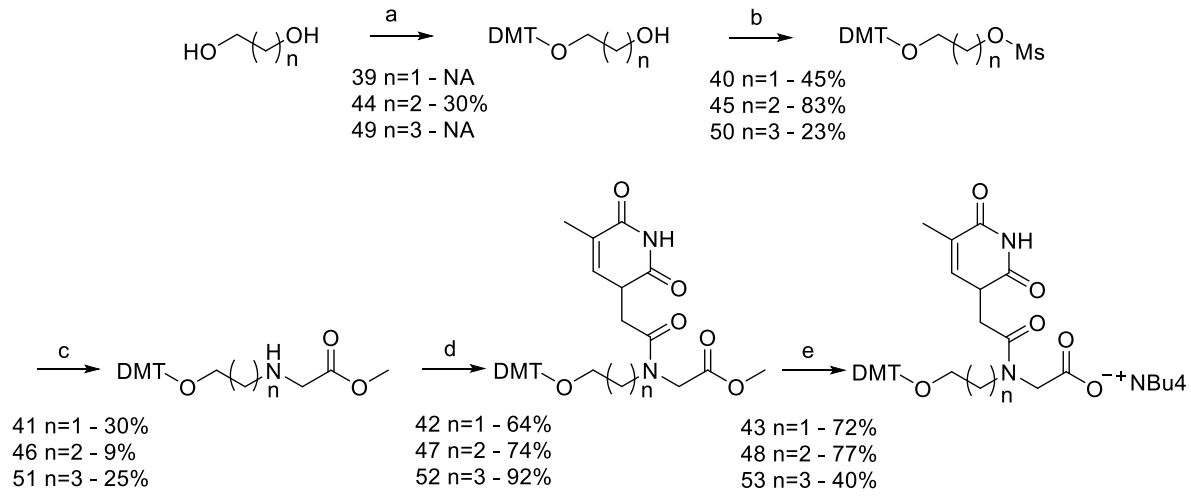
Scheme 3-18: PNA linker, reductive amination. Reagent and condition: a) H_2, Pd/C; b) EDC/HOBt/ DIPEA in DMF; c) DMT-Cl/pyridine.

The product was then coupled to thymine acetic acid using TBTU/DMF/NEt_3 (n=1 - 36) or EDC/HOBt/DIPEA (n=3 - 38) but in both cases only traces of the desired product were recovered. This route was not investigated further, as other successful strategies were used.

3.3.4 Strategy 4: Displacement of a mesylate of ethylene glycol

The last strategy investigated to synthesize the PNA linkers has been published by Spurling *et al.*[179] for synthesis of the n=1 linker. The starting material is the diol of the corresponding length, ethan-1,2-diol, propan-1,3-diol, or butan-1,4-diol. The diol in large excess in DCM is protected with DMT-Cl. Obtaining the mono-protected alcohol as main product is possible because of the large excess of diol. The product of the reaction is then extracted with DCM/water and used directly for the next reaction. The protected product was reacted with mesyl chloride in DCM in the presence of triethylamine, then concentrated, taken up in triethylamine, and reacted with glycine methyl ester. The linker backbone was then purified by chromatography on silica gel (**Scheme 3-19**).

The three linker backbones have been obtained by this strategy (41, 46, 51). However, we noticed that it is beneficial to carry out purification after the first step or second step of the route as the substitution of the mesylate by the glycine methyl ester can fail to afford the desired product if too much side product is present in the starting material. Also the propyl linker was difficult to obtain by this route: our hypothesis is that the mesyl intermediate can be subject to β-elimination in the presence of triethylamine at reflux. The use of a milder base like pyridine instead of triethylamine could bypass this problem but the experiment has not been carried out.



Scheme 3-19: Spurling synthesis of PNA linker backbone. Reagents and conditions: a) DMT-Cl/Pyridine, b) MsCl/NEt_3, c) NEt_3 reflux, glycine methyl ester d) T3P/DIPEA, thymine acetic acid e) NBu_4OH.

Synthesis of DNA-PNA chimeras

The coupling of the thymine on the backbone has been carried out with the condition selected for the monomer synthesis. Coupling of the thymine acetic acid on the linker backbone by T3P/DIPEA afforded the desired compound in high yield. The deprotection of the methyl ester was then carried out in water with a solution of tetrabutyl ammonium hydroxide solution (0.5M in dioxane). The reaction was neutralized with Dowex pyridinium and the desired linker was afforded after automated reverse phase column chromatography.

3.3.5 Summary of PNA linker synthesis routes

Of the four routes above, the fourth strategy route seems to be the most robust for synthesis of the three lengths of linker. Nevertheless, if using of a milder base during the substitution by the glycine methyl ester for the propyl linker does not improve yield, the second strategy route should be followed to make it.

3.4 Solid support

3.4.1 Sate of the art

Three main types of solid phases have been used in PNA-DNA chimera synthesis. The first PNA synthesis by MMT-protecting group strategy by Will in 1995[164] aminotentagel™ or aminopropyl-CPG have been coupled with an aminohexyl spacer and a base cleavable succinyl linker[159 179] (**Figure 3-10**).

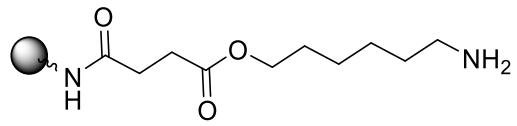


Figure 3-10: Succinyl aminohexanol linker

More recently in 2007 Moggio *et al.*[152] coupled Fmoc glycine to hydroxy tentagel. In 2008 Potenza *et al.* applied the same strategy from CPG-OH[141] (**Figure 3-11**).

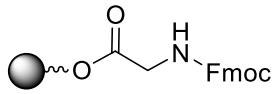


Figure 3-11: Solid phase Gly-Fmoc

Thirdly a polystyrene type of solid phase has been used by Vinayak *et al.* and Van Der Laan in 1997[161, 165]. The highly crosslinked polystyrene beads have a hydroxymethyl benzoic acid linker and were functionalised with Fmoc glycine. The resin was deprotected and coupled to the first PNA monomer prior to use in automated synthesis (**Figure 3-12**).

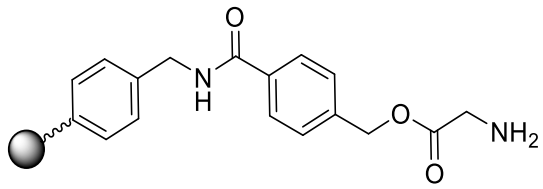


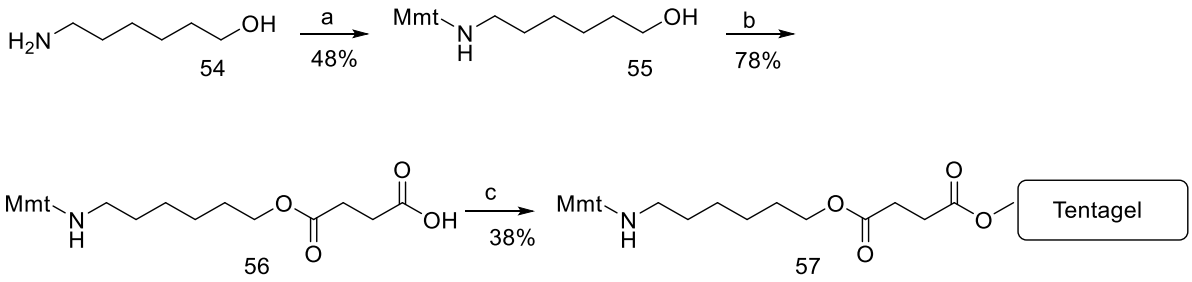
Figure 3-12: Highly cross linked polystyrene solid phase

3.4.2 Preparation of functionalized solid support

Both the aminohexyl succinyl linker on tentagel the glycine linker on tentagel were used in the present work.

The amino hexyl succinyl linker needed to be synthesized and was prepared (**Scheme 3-20**). Firstly, the terminal amine of aminohexanol was protected with the monomethoxytrityl group

(53), and then the hydroxyl group was reacted with succinic anhydride in order to form the monomethoxytritylamino-hex-l-y1 hemisuccinate (54)**.**



Scheme 3-20: Preparation of succinyl aminohexanol linker; a) MMT-Cl/pyridine; b) succinic anhydride/pyridine/DMAP; c) HATU/NMM/tentagel-OH.

Finally, the linker was coupled to the resin by HATU with NMM, by gently shaking the suspension overnight. The loading obtained 94 µmol/g, was measured by UV measurement of deprotected MMT.

Tentagel-O-Glycine-Fmoc was also synthesized under the same conditions as c), allowing a loading of 120 µmol/g measured by UV absorbance of the deprotected Fmoc.

Methods to calculate loading are presented in the general section of the experimental chapter.

3.4.3 Comparison of solid supports for the synthesis of chimeras

In order to identify if another solid support would be better than tentagel for the purpose of growing chimeras we investigated NittoPhase and amino hybrid CPG. The solid phases were prepared by coupling in DMF of the solid phase and the MMT aminohexyl succinic acid in presence of HATU and DIPEA. The hybrid CPG was loaded at 10 µmol/g and the Nittophase was loaded at 25 µmol/g. The resulting loadings were too low to permit a 2 µmol synthesis but the column was filled at the ¾ of the total volume and 2umol standard PNA cycle on Expedite 8909 was run (see next for detail of synthesis). Synthesis on Tentagel was performed in the same batch with the same sequence for comparison. Two syntheses of the PNA part of the same sequence on the same solid phase were synthesis at the time. At the end of the PNA synthesis one of them was deprotected with 40% aqueous $MeNH_2$ / NH_4OH, 1:1 ratio, for 90 minutes at rt and analysed. The other was carried over on an ABI394 with PS cycle. The oligomers were then deprotected with 40% aqueous $MeNH_2$ / NH_4OH, 1:1 ratio, for 90 minutes at RT, purified and analysed. This experiment setup does not permit to compare yield but LC and MS analysis are able to give hint on the efficiency of the synthesis. The following sequence was used for this experiment: $T_sG_sT_sA_s$ $C_sG_sT_s$ $A_sT_sG_s$ G_s+T_s et ac atc, (“N” is DNA; “n” is PNA; et is ethyl PNA linker; $_s$ is phosphorothioate linkage).

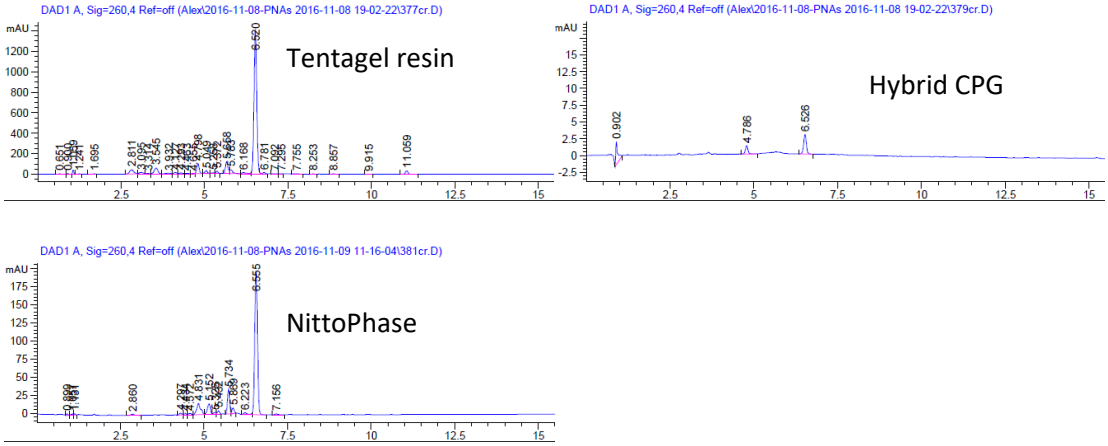


Figure 3-13: Analytical HPLC trace of crude PNA part. Retention time of the expected product is 6.5min

Tentagel and Nittophase seem to have similar behaviour in PNA synthesis affording predominantly the desired product in big majority. Hybrid-CPG didn't seem to have worked at all.

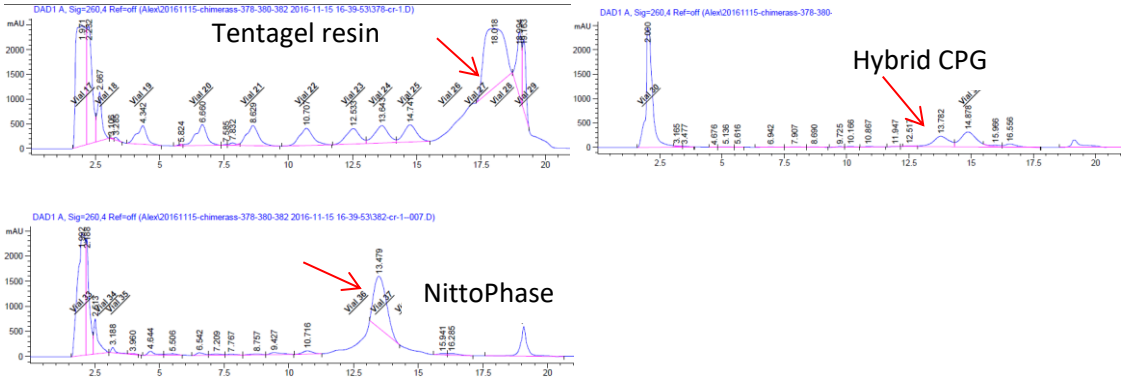


Figure 3-14: Semi-Preparative HPLC traces of the full chimeras. Expected product indicated with red arrows.

Chimeras were purified by HPLC and analysed in LC/MS. From the purification trace, Tentagel and NittoPhase have made possible a clean synthesis. The amount recovered from hybrid-CPG was very low as it was expected from the PNA part analysis.

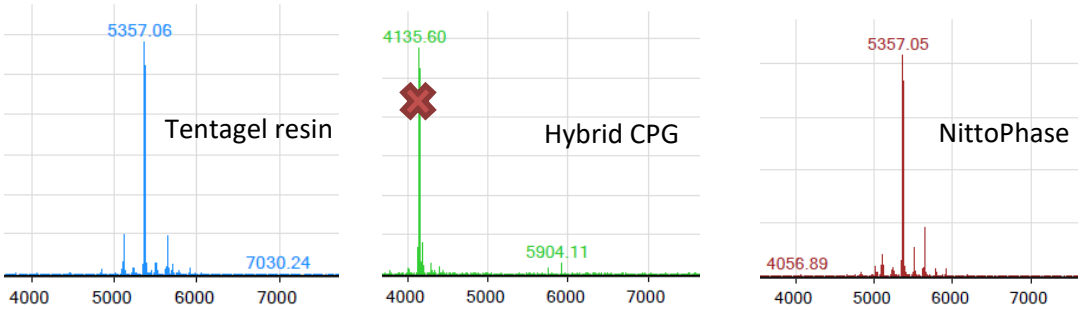


Figure 3-15: MS spectra of a chimera synthesis on 3 different supports. Calculated mass is MW=5357.64. Observed mass with tentagel resin M_{obs}=5357.06; with hybrid CPG $M_{obs=}$4135.60; with NittoPhase M_{obs}=5357.05.

The desired chimeras were recovered from the Tentagel and NittoPhase experiment but not from the hybrid-CPG. The results were similar for these two solid phases. If NittoPhase can be loaded with a higher amount it will be interesting to use for routine chimera synthesis as it is known to be efficient for standard oligonucleotide synthesis. The use of hybrid-CPG does not seem to be recommended but the low loading obtained may be an important factor in the poor results obtained.

3.4.4 Manual coupling SPPS of the PNA end of chimeras

The manual solid phase synthesis of PNA oligomers was carried out at 10 umol scale in a 10 mL syringe equipped with a frit (**Figure 3-16**). The different solutions were prepared under Argon atmosphere in sealed and dried RBF. Monomers were dissolved at 0.1 M in NMP, HATU was used as activator in solution in DMF at 0.1M. The base mix used was 0.1 M of lutidine and 0.1 M DIPEA in NMP. The deblock solution was 3% TCA in DCM. Four equivalents of activated PNA monomer were used at each step. Between all steps, three washes with DCM/NMP 1/1 solution were performed.

Figure 3-16: Reactor used for manual solid phase of PNA.[180]

The solid phase was washed with deblock solution to remove the MMT group until the solution was colourless. The first base and every cytosine were always double coupled. Following the synthesis of the PNA wing at 4 µmol scale by this method, half of the resin was taken and the PNA linker was coupled using the same cycle as the PNA monomers. Finally, the solid phase was put into a synthesizer column and the DNA part of the chimeras was grown on an ABI394 DNA synthesizer following standard condition. The obtained chimeras were then deprotected with ammonium hydroxide (55°C for 1 hour and room temperature overnight) to afford the crude chimeras.

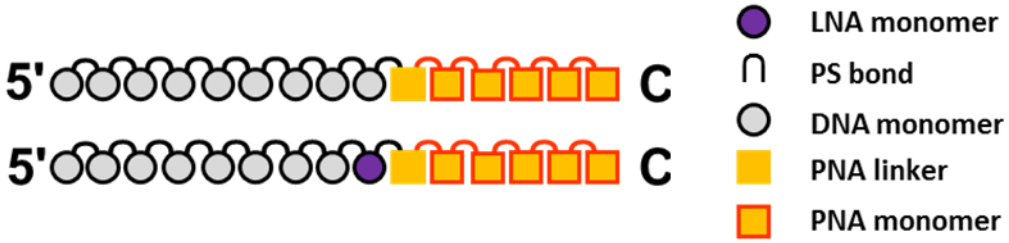


Figure 3-17: Design of the first batch of chimeras synthesized by manual solid phase synthesis.

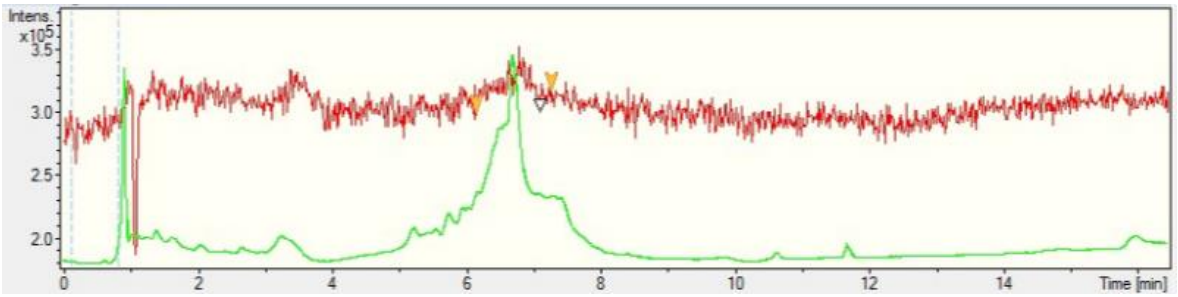


Figure 3-18: LC/MS trace of crude dAdCdG dTdAdT dGdGdC et tc aac chimeras (green: absorbance at 260, red: TIC)

The crude LC trace (**Figure 3-18**) shows the low efficiency of this synthesis method. The low efficiency of the method is further highlighted by the difference of between crude and pure yield (**Table 3-3**). Another interesting thing to note is the poor resolution of the chimeras on reverse phase HPLC, this property of the chimeras makes reverse phase a poor choice to purify and analyze them (see section **3.4.6)**.

The chimeras were purified by 20% denaturing PAGE (**Figure 3-17** and **Table 3-3**). The amount of each oligo afforded after purification was between 5 and 20 nmol. This is a very low yield: purified amounts recovered from chimera synthesis when the PNA part is made on the Expedite synthesizer are typically between 80 and 400 nmol. Therefore, the manual method is not optimal for making chimeras. Nevertheless, we isolated sufficient pure material for melting studies, which was the goal of this set of syntheses.

Name		Crude quantity (nmol)	Quantity after gel purification (nmol)	Expected Mass	Observed Mass
A_Ch_eDNA	dAdCdGdTdAdTdG dGdA et aca tc	23	12	4484.35	4484.2
A_Ch_eLNA	dAdCdGdTdAdTdG dG +A et aca tc	21	7	4512.36	4512.2
A_Ch_pDNA	dAdCdGdTdAdTdG dGdA pt aca tc	29	6	4498.38	4497.3
A_Ch_pLNA	dAdCdGdTdAdTdG dG +A pt aca tc	45	6	4526.39	4526.2
A_Ch_bDNA	dAdCdGdTdAdTdG dGdA bt aca tc	21	8	4512.41	4512.4
A_Ch_bLNA	dAdCdGdTdAdTdG dG +A bt aca tc	17	6	4540.42	4540.2
B_Ch_eDNA	dAdCdGdTdAdTdG dGdC et tca ac	24	12	4460.33	4459.2
B_Ch_eLNA	dAdCdGdTdAdTdG dG +C et tca ac	33	15	4502.37	4501.2
C_Ch_eDNA	dAdCdGdTdAdTdG dGdT et ttt tt	47	20	4488.35	4892.7
C_Ch_eLNA	dAdCdGdTdAdTdG dG+T et ttt tt	33	18	4515.35	4514.2
C_Ch_pDNA	dAdCdGdTdAdTdG dGdT pt ttt tt	51	10	4501.37	4500.1
C_Ch_pLNA	dAdCdGdTdAdTdG dG+T pt ttt tt	81	9	4529.38	4529.1
C_Ch_bDNA	dAdCdGdTdAdTdG dGdT bt ttt tt	53	6	4515.40	4515.2
C_Ch_bLNA	dAdCdGdTdAdTdG dG+T bt ttt tt	88	12	4543.31	4543.3

Table 3-3: Table of chimeras synthesized by manual solid phase synthesis. “dN’ is DNA, “+N” is LNA, “n” is PNA, et is ethyl linker, pt is propyl linker and bt is butyl linker. All DNA-DNA and DNA-PNA bonds are phosphodiester.

3.4.5 Expedite synthesis of the PNA end of chimeras

The synthesis of the PNA 3’ end of the chimeras was performed on solid phase at 2 µmole scale on an automated Expedite 8909 synthesizer with standard PNA protocol. Double coupling was performed for the first base. 3%TCA in DCM was used as deblock solution. HATU (0.19 M solution in DMF) was used as activator. Monomers were dissolved to 0.2 M in DMF. Acetic anhydride with DIPEA in DMF was used as capping solution. The base mix was lutidine plus DIPEA in DMF.

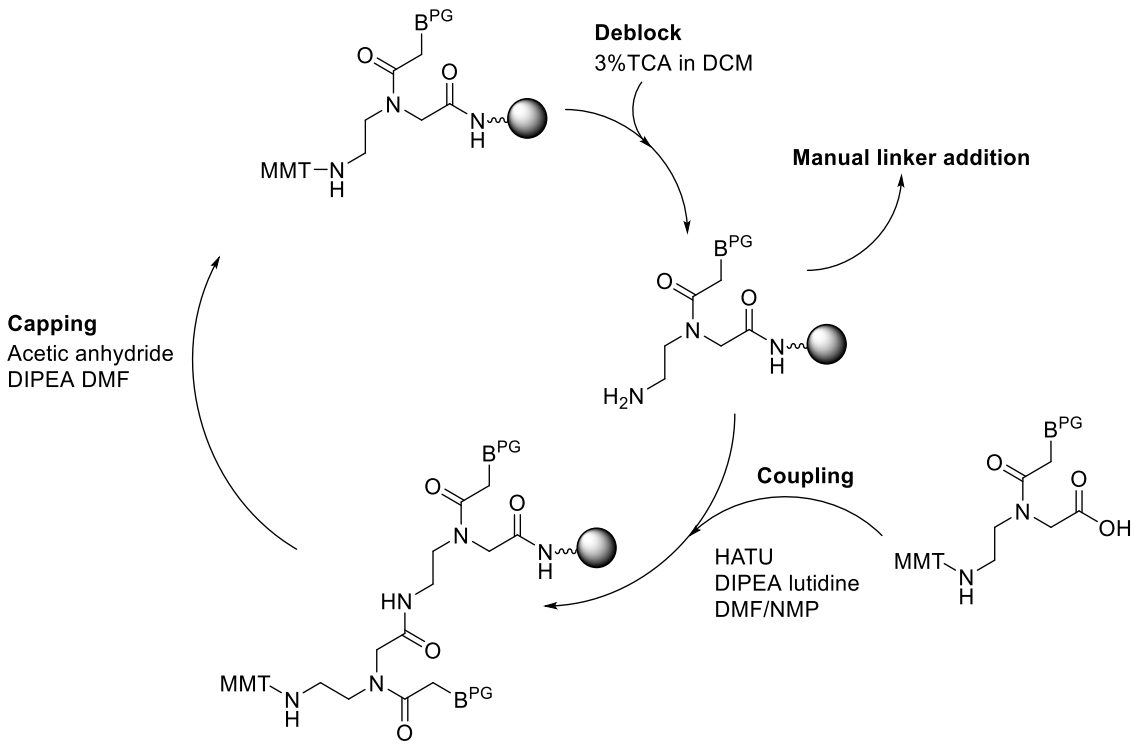


Figure 3-19: MMT protected PNA automated synthesis cycle.

Step	Reagent	Time
Wash	DMF	—
Deblock	3% TCA in DCM	2.5min
Wash	DMF	—
Activate and couple	5eq DIPEA 7.5eq Lutidine 5eq monomer 4.5eq HATU	8.25min
Wash	DMF	—
Cap	5% acetic anhydride 6% DIPEA in DMF	3min

Table 3-4: Details of the MMT-PNA cycle.

The synthesis was ended leaving the final DMT group attached and the column was taken to carry on the synthesis on ABI synthesis at 1 µmol scale.

3.4.6 Elongation of Chimeras: adding nucleotides

The elongation of the nucleotide part of the chimeras was performed on an ABI394 oligonucleotide synthesizer with standard phosphoramidite chemistry. BTT was used as activator

and the coupling time for DNA bases was one minute. Modified nucleotides and the first nucleotides were coupled for 10 minutes.

At one point, because of the relatively low yield obtained during chimera synthesis, we attempted to extend the coupling time of every base to 10 minutes. However, we observed multiple coupling of G (**Figure 3-20**) which is likely derived from deprotection of the newly coupled DMT group, during the coupling step, by the slightly acidic BTT activator, followed by coupling of another phosphoramidite.

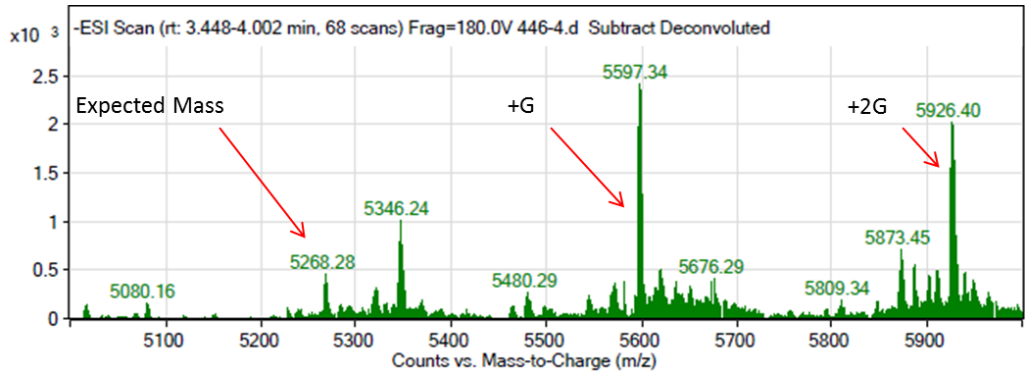


Figure 3-20: Deconvoluted MS spectrum of +G+T+A CGT ATG G FANAT pt ac atc chimeras main peak. "+N" is LNA, "dN' is DNA, "n" is PNA, "FANAT" is 2'-F-ANA-T, "et" is ethyl linker, "pt" is propyl linker and "bt" is butyl linker. All non PNA-PNA bonds are phosphodiester.

Deprotection of DMT can be slow on Tentagel; complete deprotection of the first bases was monitored by looking at the trityl colour on synthesiser and deprotection was extended if necessary. When using α-L-LNA and 2'F-ANA nucleotides, the addition was made manually as the amount of monomer was limited: manual coupling avoids the dead volume and prime volume required by the synthesizer, therefore allowing less material to be used for one coupling. The manual addition was performed on the column with similar conditions to the synthesizer. The coupling time was 15 minutes. The chimeras that were made with a manual coupling have lower yield than the other made in all automated synthesis, therefore if possible an all automated synthesis is preferred.

Chimera cleavage and deprotection were performed with equal success by treatment with $NH_4OH_{(aq)}$ at 55°C for 16 h or with $NH_4OH_{(aq)}$ at RT for 48 h. Purification of the chimeras was more challenging than we expected. As discussed above, we had previously observed that RP-HPLC was an inefficient method to purify chimeras (**Figure 3-18**). Therefore, we purified the chimeras by ion-exchange HPLC with apparent success (**Figure 3-21**). LC/MS analysis performed on these chimeras showed a single peak with the correct mass, but the column used for this analysis was very low resolution and, as it turned out, hid multiple compounds under a single peak. Melting experiments using this material showed surprising hysteresis (T_m-T_a = 1-13°C). After it was brought

to our attention that this hysteresis could be caused by low purity of the oligonucleotides, further analyses were conducted. The chimeras were analysed by reverse phase HPLC (**Figure 3-22**) and the same samples that appeared pure by anion-exchange HPLC were shown to contain multiple products. Analysis of the different peaks by LC/MS allowed us to identify the main product as the desired product as well as the second largest product the n-1 from the solid phase synthesis.

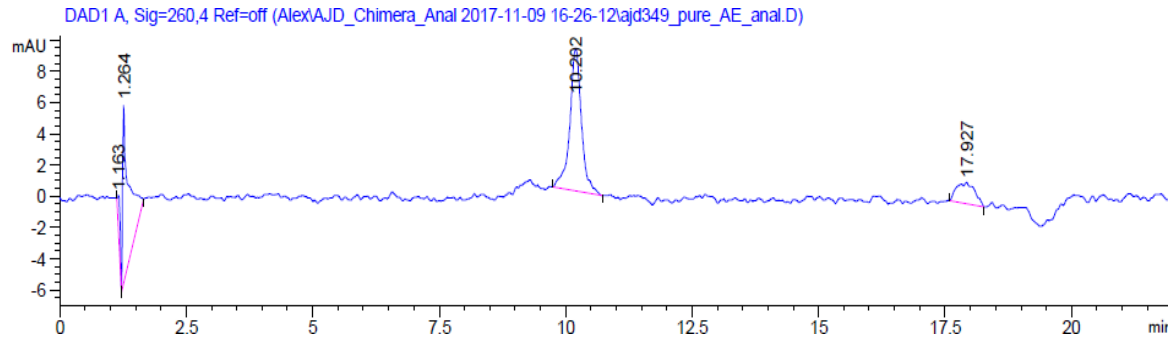


Figure 3-21: UV spectrum of anion-exchange analytical HPLC of ptLNA chimera after anion exchange HPLC purification. Retention time of the desired chimera is 10.2 min. The peak at 1.2 min is an injection peak, not a contaminant. The HPLC was performed with a gradient of 2% to 75% of B in A (A: 30%ACN in water; B: 30%ACN in 1M $NaClO_{4\ aq}$). This spectrum is representative of the entire batch of chimeras.

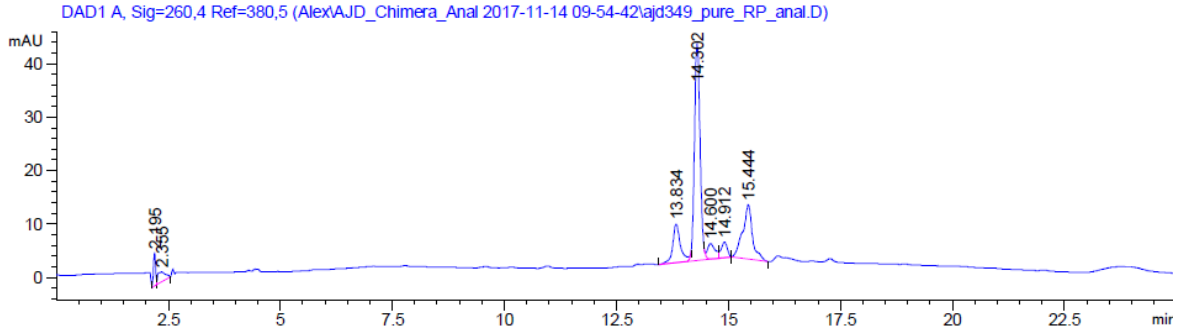


Figure 3-22: UV spectrum of reverse phase analytical HPLC of ptLNA chimera after anion exchange HPLC purification. Retention time of the desired chimera is 14.3 min. The peak at 2.2 min is an injection peak. The HPLC was performed with a gradient of 35% to 100% of B in A (A: 0.1M TEAA $_{aq}$; B: 20%ACN in 0.1M TEAA $_{aq}$). This spectrum is representative of the entire batch of chimeras.

The chimeras were therefore purified a second time: Gel purification was performed on all chimeras presented on **Table 4-1** after their initial purification by anion-exchange HPLC. LC/MS analysis was performed on these doubly-purified chimeras using a high definition column (**Figure 3-23**).

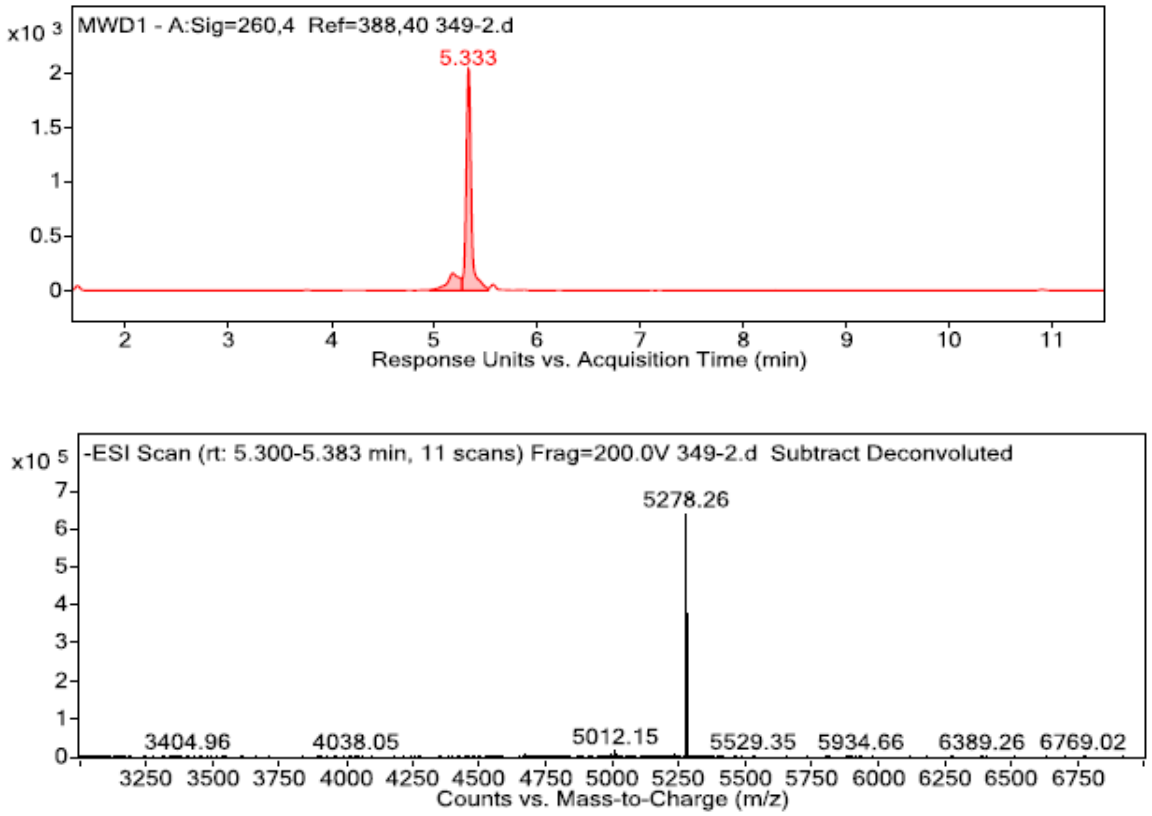


Figure 3-23: LC/MS spectra of pLNA chimeras after second purification. MW expected=5278.5. These spectra are representative of the entire batch of chimeras. HPLC was undertaken via an Agilent Poroshell 120 EC-C18 column (4.5mm x 50mm 2.7µm). Gradient 2% B to 100% B over 10 min at a flow rate of 1 ml/min. Buffer A: 100mM HFIP with 9 mM NEt_3 in water; Buffer B: 100mM HFIP with 9 mM NEt_3 in MeOH.

In summary, we have found that two methods of purification are required to obtain pure PNA chimeras. We previously observed (**Figure 3-18**) that reverse phase HPLC alone is not sufficient to purify PNA chimeras. Furthermore, anion-exchange HPLC alone is not sufficient to obtain pure chimeras (**Figure 3-21** and **Figure 3-22**). The use of two purification techniques in tandem allowed us to obtain a pure product. We chose to perform gel purification after anion-exchange HPLC. However, we believe that anion exchange HPLC followed by reverse phase HPLC would likely achieve similar results.

Chapter 4: **Properties of chimeras**

4.1 Biophysics of DNA-PNA chimeras

4.1.1 Design

We focused our optimization of the biophysical properties of chimeras on modifications to the DNA-PNA junction. We conceptually divided the junction in three different parts (**Figure 4-1**), the 3'-nucleotide, the phosphate linkage and the first PNA monomer (called the PNA linker since it contains a terminal oxygen instead of a nitrogen). According to the state of the art, the linkage that will be used would be a phosphodiester linkage for *in vitro* biophysical studies, and a phosphorothioate linkage for *in cellulo* and *in vivo* studies).

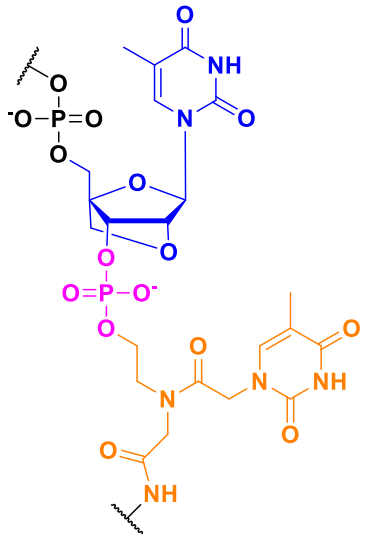


Figure 4-1: Three areas of focus for improving the DNA-PNA junction: the nucleotide (blue), the the phosphate linkage between DNA and PNA (violet), and the first PNA monomer, called the PNA linker (orange).

Three different nucleotides are being investigated at the junction: DNA, LNA and 2'F-ANA (**Figure 4-2**). These nucleotides will be tested in combination with three different types of PNA linker that differ in terms of the length of their backbone (Figure 3-3). The basic backbone for PNA linkers is a hydroxyethyl-glycine, and in this study, we include hydroxypropyl and hydroxybutyl glycine as well as the basic linker. We hypothesised that a longer linker would give more flexibility at the junction, and that varying the length might yield appropriate combinations in combination with nucleotides of different conformation and flexibility.

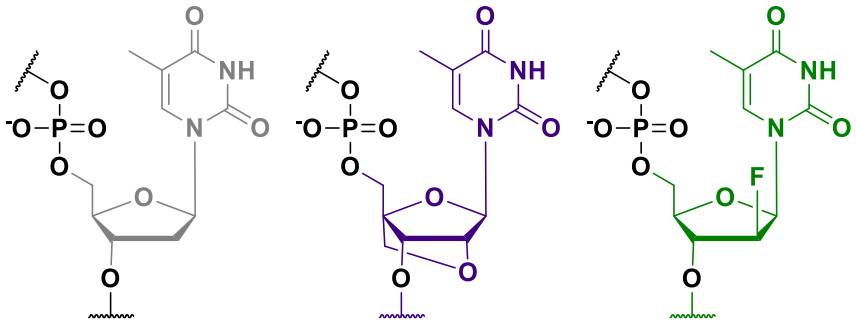


Figure 4-2: The chemistries of the monomer at the DNA side of the junction. DNA, LNA, and 2'F-ANA.

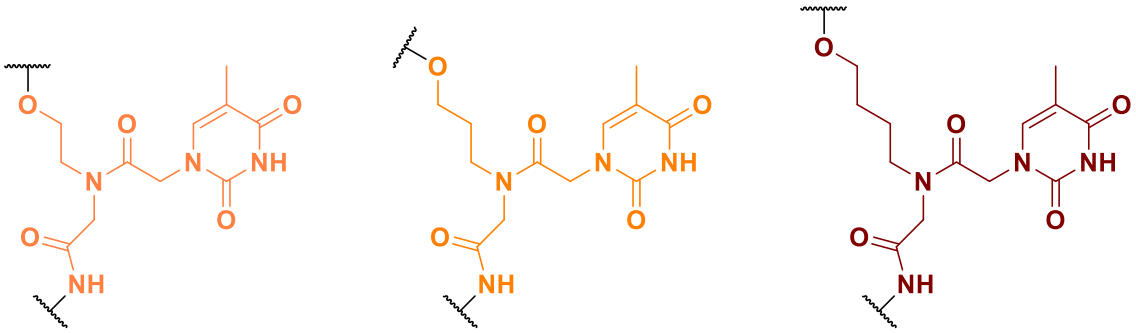


Figure 4-3: Structure of the different PNA linkers. From left to right ethyl T (et), propyl T (pt) and butyl T (bt) linkers. The phosphate would extend from the top of the molecule as drawn, and the remaining PNA chain from the bottom of the monomer.

The general design of the chimeras used in the biophysical experiments is a 17mer oligomer. At the 5' end is a wing of three LNA nucleotides; the core is made of a gap of seven DNA nucleotides to recruit RNase H, and the nucleotide after the 3' end of the DNA gap is the junction nucleotide, which we modulated to explore how different conformations would affect binding affinity and other properties. The PNA region is then made of one PNA linker and 5 aeg-PNA monomers. All the PNA used in this study have an aminohexyl tail at the C terminal due to the solid support used for the synthesis.

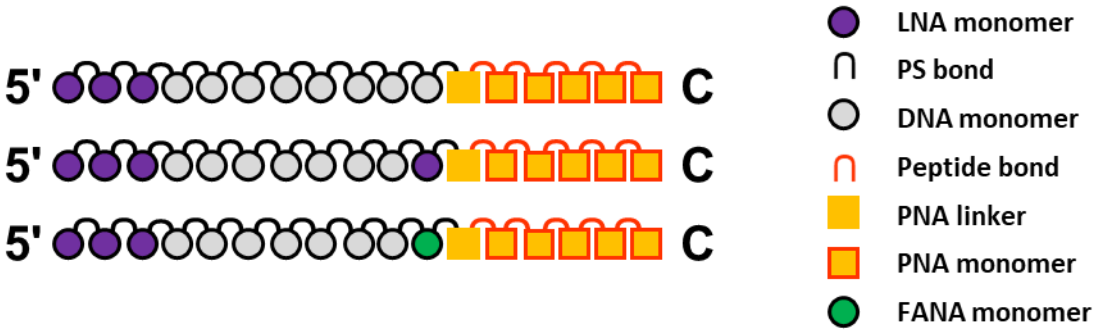


Figure 4-4: A schematic of the PNA chimera designs used in this study.

The rationale for our designs is based on the idea that the junction between the DNA part and the PNA part is the key to obtain high affinity chimera/RNA duplexes. All measurements are made on

fully complementary chimera/RNA duplexes. The chimeras differ by the structure of the 3'-DNA base and by the structure of the 5' PNA base.

Name	Sequence	Expected Mass (g/mol)	Observed Mass (g/mol)
PO DNA	dGdTdAdCdG dTdAdTdGdGdTdTdAdCdAdTdC	5200.4	5200.9
PO RNA	UGAUGUAACCAUACGUACU	6006.6	6005.9
PS DNA	$dG_sdT_sdA_sdC_sdG_sdT_sdA_sdT_sdG_sdG_sdT_sdT_sdA_sdC_sdA_sdT_sdC$	5457.5	5456.5
LNA Gap	+G+T+AdCdGdTdAdTdGdGdTdTdAdC+A+T+C	5639.6	5638.5
LNA Gap^s	$+G_s+T_s+A_sdC_sdG_sdT_sdA_sdT_sdG_sdG_s+T_s+T_s+A$	4365.5	4364.4
eDNA	+G+T+A dCdGdT dAdTdG dG dT et ac atc	5236.9	5236.5
pDNA	+G+T+A dCdGdT dAdTdG dG dT pt ac atc	5251.0	5250.4
bDNA	+G+T+A dCdGdT dAdTdG dG dT bt ac atc	5265.0	5264.4
eLNA	+G+T+A dCdGdT dAdTdG dG +T et ac atc	5265.0	5264.4
pLNA	+G+T+A dCdGdT dAdTdG dG +T pt ac atc	5279.0	5278.5
bLNA	+G+T+A dCdGdT dAdTdG dG +T bt ac atc	5293.0	5292.5
eFANA	+G+T+A dCdGdT dAdTdG dG fT et ac atc	5254.6	5254.5
bFANA	+G+T+A dCdGdT dAdTdG dG fT bt ac atc	5283.0	5282.5

Table 4-1: Sequence of the oligonucleotides used for melting, CD and RNase H experiments. “N” is RNA, “+N” is LNA, “dN’ is DNA, “n” is PNA, et is ethyl linker, pt is propyl linker, bt is butyl linker and “$_s$” is phosphorothioate bond.

4.1.2 Melting studies

In order to study the affinity of the chimeras for complementary RNA we carried out melting studies on the chimeras. The absorbance at 260 nm is recorded over time while the temperature is increased; when the two strands dissociate, the absorbance increases. This phenomenon is called hyperchromicity and is used to determine the melting temperature (T_m) of the duplex. The melting curve is sigmoidal, and the point of inflection of the curve represents the point when 50% of the duplexes are dissociated. The temperature at this point of inflection on the ascending curve is the T_m for the duplex.

Our first set of melting experiments (**Table 4-2:** Melting temperature of the different chimeras designs against complementary RNA. A, B and C are different sequences. Chimera sequences are shown in **Table 3-3.** Duplex at 1 µM in 1xTRIS buffer with 140 mM NaCl. *Tm* and ΔG310 are reported as mean of n=3. Samples were heated at 1 °C/min and temperatures were calculated with the hyperchromicity method from Agilent Thermal software.

) were conducted on the chimeras made of 8 DNA nucleotides at the 5'-end, one additional nucleotide (either DNA or LNA) at the junction, then a PNA linker and 5 normal PNA monomers at the 3'-end (**Figure 3-17**, **Table 3-3**). Three different sequences were synthesized to confirm that the effect of the junction on affinity was independent of the oligonucleotide sequence. UV

Properties of chimeras

experiments were conducted on fully complementary RNA/chimera duplexes. The nomenclature code can be interpreted as follows: A, B, and C refer to the sequences. PO_DNA and PO_RNA represent fully DNA and RNA sequences, respectively, used as controls. Ch is chimera; e, p and b represent the ethyl, propyl and butyl PNA linker at the junction while the DNA, LNA or FANA in these sequences refers to the chemistry of the nucleotide at the junction.

Sequence A	T_m (°C)	Stdev	ΔG^{310} (kJ/mol)	Sequence B	T_m (°C)	Stdev	ΔG^{310} (kJ/mol)	Sequence C	T_m (°C)	Stdev	ΔG^{310} (kJ/mol)
A_PO_DNA	**41.1**	0.1	**-42.7**	B_PO DNA	**46.6**	0.5	**-49.3**	C_PO DNA	**38.5**	0.2	**-39.4**
A_PO_RNA	**55.1**	0.3	**-52.2**	B_PO RNA	**57.3**	0.0	**-55.5**	C_PO RNA	**44.6**	0.2	**-46.4**
A_Ch_eDNA	**39.5**	0.9	**-39.2**	B_Ch_eDNA	**45.4**	0.3	**-42.6**	C_Ch_eDNA	**42.4**	0.9	**-40.3**
A_Ch_pDNA	**43.4**	0.1	**-46.5**					C_Ch_pDNA	**47.3**	0.2	**-42.3**
A_Ch_bDNA	**43.3**	0.7	**-45.9**					C_Ch_bDNA	**46.9**	0.7	**-41.6**
A_Ch_eLNA	**44.6**	0.6	**-45.3**	B_Ch_eLNA	**50.3**	1.0	**-46.6**	C_Ch_eLNA	**50.2**	0.3	**-43.9**
A_Ch_pLNA	**44.2**	0.3	**-45.2**					C_Ch_pLNA	**49.8**	0.2	**-43.0**
A_Ch_bLNA	**41.6**	0.6	**-43.5**					C_Ch_bLNA	**47.4**	0.3	**-42.2**

Table 4-2: Melting temperature of the different chimeras designs against complementary RNA. A, B and C are different sequences. Chimera sequences are shown in **Table 3-3**. Duplex at 1 µM in 1xTRIS buffer with 140 mM NaCl. T_m and ΔG^{310} are reported as mean of n=3. Samples were heated at 1 °C/min and temperatures were calculated with the hyperchromicity method from Agilent Thermal software.

Table 4-2 shows that modifications have the same impact on the melting temperature, regardless of the sequence. The lowest melting temperature is the design that has been widely studied in the past (Ch_eDNA). The substitution of the DNA by LNA at the junction greatly improves the T_m (more than +5°C) when ethyl linker is used. However, the substitution of DNA by LNA when using propyl and butyl linker has little to no effect (+/- 2°C). As reported by Greiner *et al.*[159, 160] (**Table 3-1**) we can observe an improvement in T_m by using propyl linker (+3-5°C) and butyl linker (+4°C) in combination with DNA. These results underline the importance of compatibility between the PNA linker and the nucleotide at the junction. They have to be approached as one entity, since the best linker for one nucleotide can be destabilizing for another type of nucleotide.

Melting studies were also performed on the chimera gapmer design shown **Figure 4-4** which possesses a wing of LNA at the 5'-end (**Table 4-1** and **Table 4-3**). The variations observed in the T_m of the chimera gapmers are similar to the ones observed for the previous design with similar junctions (**Table 4-2:** Melting temperature of the different chimeras designs against complementary RNA. A, B and C are different sequences. Chimera sequences are shown in **Table 3-3.** Duplex at 1 µM in 1xTRIS buffer with 140 mM NaCl. *Tm* and ΔG310 are reported as mean of n=3. Samples were heated at 1 °C/min and temperatures were calculated with the hyperchromicity method from Agilent Thermal software.

). The substitution of DNA by LNA in combination with an ethyl linker (eDNA and eLNA) greatly increased the T_m (+3°C). The use of propyl and butyl linkers with DNA (pDNA and bDNA) improved

the T_m (+2°C) while it slightly lowered the T_m (-1-2°C) if it was used with LNA (pLNA and bLNA compared to eLNA). When coupled with FANA the ethyl linker shows a T_m equivalent to eDNA. However, coupled with butyl linker, bFANA shows a high T_m (55.7°C), nearly as high as eLNA. The best linker/nucleotide combinations are pDNA, eLNA and bFANA. This underlines again that the junction has to be considered as a whole when investigating the best chemistry for stability. It is interesting that the combination used in most previous studies (eDNA) shows the lowest T_m.

Design	T_m (°C)	StDev T_m	ΔG^{310} (kJ/mol)	StDev ΔG^{310}
PO DNA	**49.3**	0.8	**-46.5**	1.2
PO RNA	**63.9**	0.1	**-59.6**	1.3
PS DNA	**48.0**	0.5	**-47.2**	0.7
LNA Gap	**62.9**	0.1	**-59.6**	2.9
eDNA	**52.4**	0.6	**-56.0**	0.7
pDNA	**54.7**	0.2	**-55.6**	0.7
bDNA	**52.7**	0.2	**-47.4**	0.7
eLNA	**56.5**	0.4	**-53.4**	1.2
pLNA	**55.2**	0.1	**-53.8**	0.4
bLNA	**53.0**	0.3	**-51.4**	0.8
eFANA	**53.1**	0.2	**-55.6**	1.3
bFANA	**55.7**	0.1	**-56.1**	1.2

Table 4-3: Melting temperatures of chimera:RNA duplexes at 1µM in 1xPBS. Steps are 0.5˚C/min against RNA complementary strand. T_m, and ΔG^{298} were obtained from hyperchromicity method from Agilent Thermal software. Mean and standard deviation (StDev) were calculated from n=3. Full sequences are given in **Table 4-1**.

The conditions of the first melting experiment and the second differ due to a change in our general laboratory procedures in order to have better consistency across our different projects. The buffering agent used in the two experiments was different (TRIS vs phosphate) but the concentration of salt was similar. Both buffers are considered standard conditions for melting experiment. The temperature ramp was also slower in the second set (0.5˚C/min compared to 1˚C/min) to afford better accuracy. ΔG values are all in the same bracket with the exception of bDNA that is higher which is suggesting a weak stability of the bDNA/RNA duplex.

In summary, the eLNA and bFANA chimeras showed the highest binding affinity. The eLNA design was selected for *in cellulo* gene silencing experiments due to the fact that we had a good stock of LNA amidites of all four bases, which was not the case for FANA.

4.1.3 CD studies

In order to investigate the helical structure of the duplexes formed by chimeras and RNA, we studied the duplexes by circular dichroism (CD). The chimeras used in these studies are the same ones used for the second melting experiment (**Table 4-1**).

All the designed chimera:RNA duplexes adopt an A-form helical structure, with CD spectra intermediate between an RNA:RNA helix and a A-like RNA:DNA helix as seen in **Figure 4-5**.

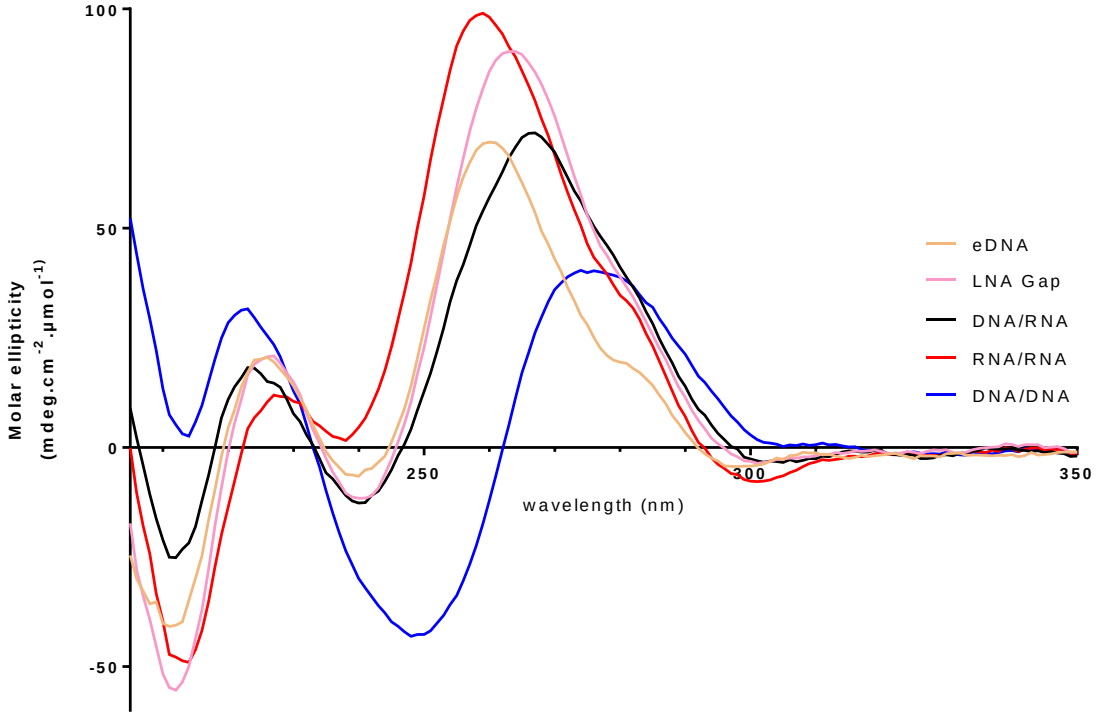


Figure 4-5: CD spectra of eDNA chimera/RNA duplex and references DNA:DNA, RNA:RNA, LNA gapmer:RNA duplexes. Measurements are taken in three replicates in 1xPBS at 4°C.

The RNA:RNA duplex shows a positive peak at 260nm and a negative peak at 210nm that are characteristic of an A-form helix. The DNA:RNA duplex show a positive peak at 265 and an negative peak at 210nm which is expected of the modified "A-like helix" formed by this kind of hybrid duplex. The DNA:DNA duplex shows a positive peak at 275nm and a negative peak at 250nm, which is characteristic of a B-form helix. All chimeras show a positive peak at 260nm and a negative peak of absorbance at 210nm which indicates the chimera designs form an A-form or A-like helix with complementary RNA.

The following CD spectra of chimeras are grouped to keep the PNA linker constant and vary the nucleotide at the PNA-oligonucleotide junction (**Figure 4-6** to **Figure 4-8**), then a figure where the nucleotide and the PNA linker is varied (**Figure 4-9**).

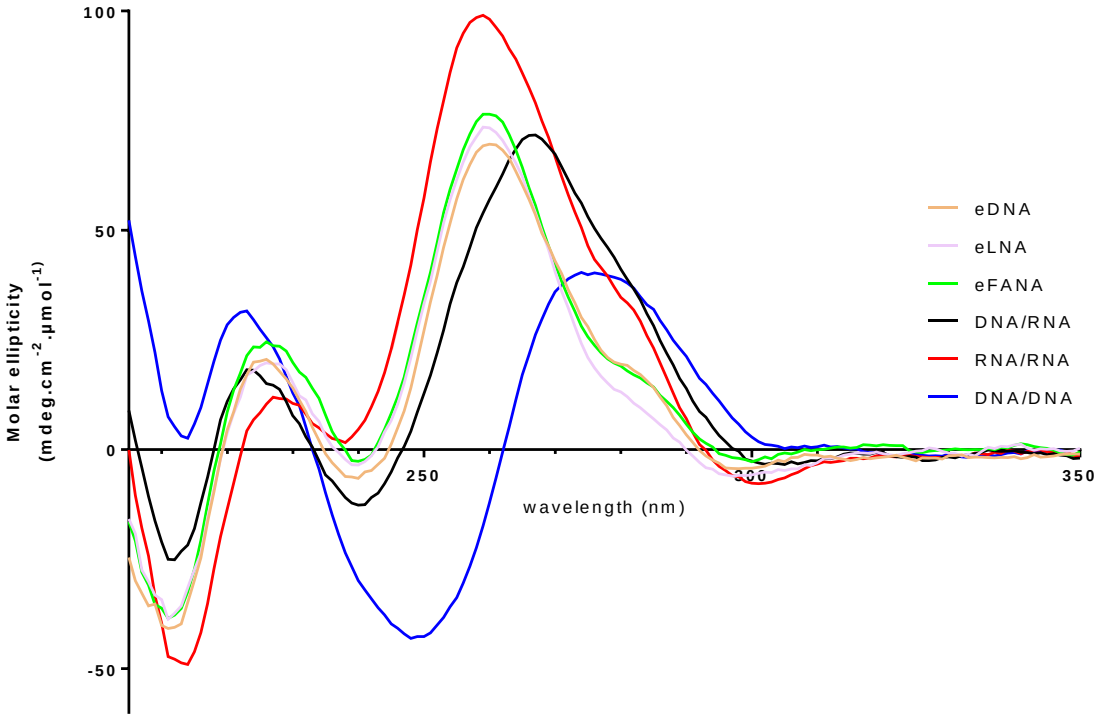


Figure 4-6: CD spectra of chimeras containing ethyl PNA linker.

The effect of the nature of the nucleotide at the junction with an ethyl PNA linker on the general conformation of the duplexes is shown in **Figure 4-6**. It is apparent that the combination of an ethyl-linker with DNA, LNA or 2'F-ANA nucleotides has very little effect on the helical structure of the chimeras, as these three chimeras show nearly identical CD spectra.

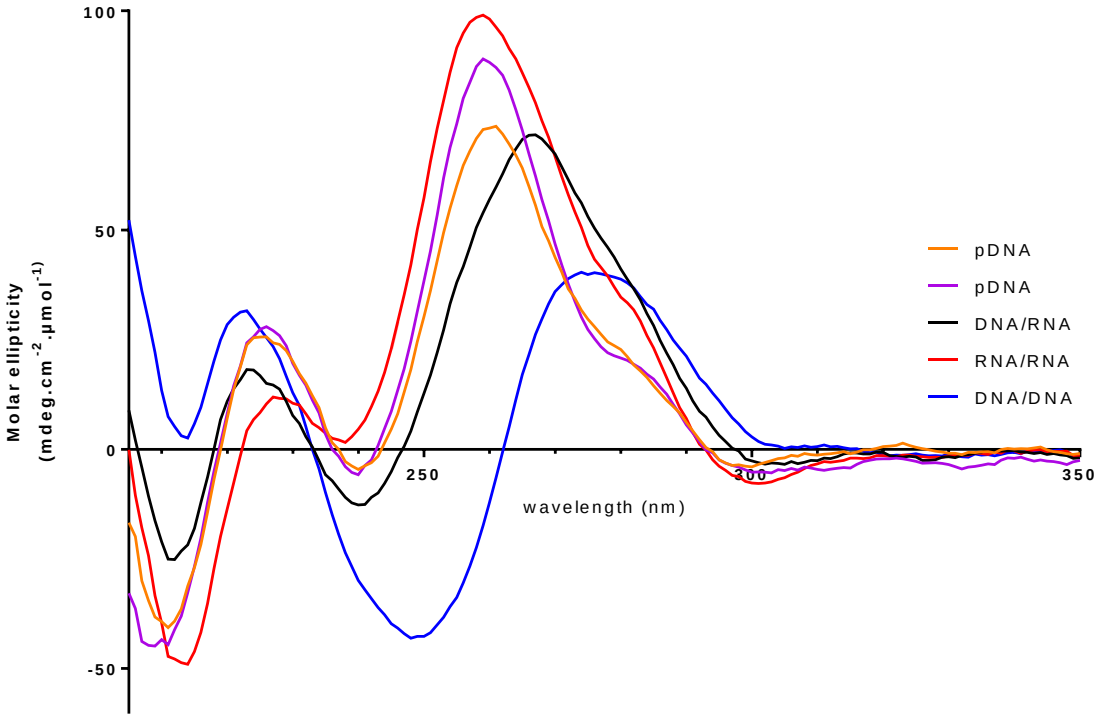


Figure 4-7: CD spectra of chimeras containing propyl PNA linker.

Properties of chimeras

Similarly, **Figure 4-7** and **Figure 4-8** show the chimeras with propyl and butyl PNA linkers, respectively, and it is clear that the terminal nucleotide has no significant structural effect in the context of propyl and butyl PNA linkers.

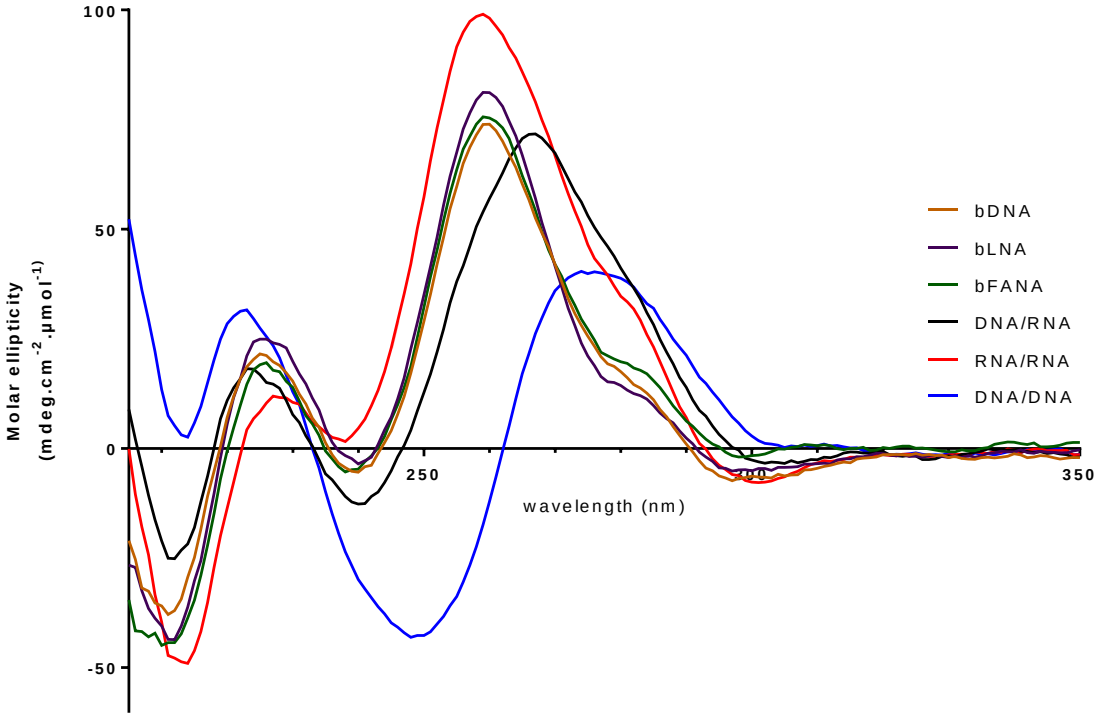


Figure 4-8: CD spectra of chimeras containing butyl PNA linker

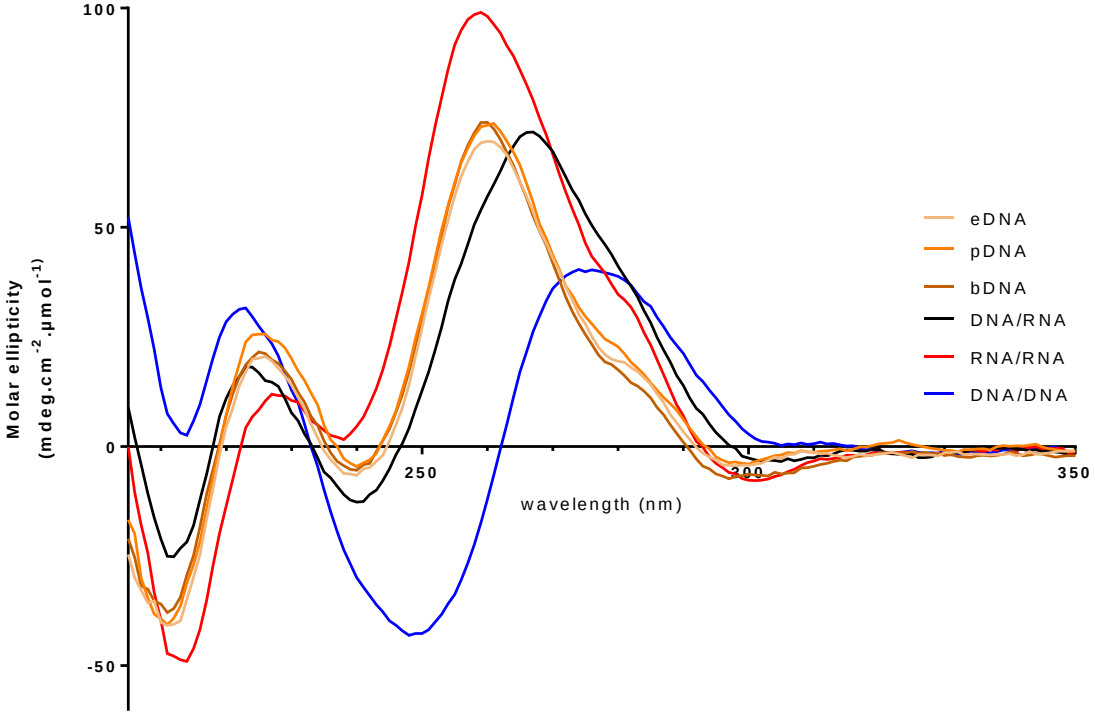


Figure 4-9: CD spectra of chimeras containing DNA linking nucleotide

Figure 4-9 compares the different PNA linkers with constant nucleotide chemistry at the junction. Once again, the three chimeras present similar spectra. The other comparisons (made with LNA or 2'F-ANA nucleotides constant) show similar results.

Thus, for all the junctions we investigated, the effect of changing the junction on the overall chimera/RNA duplex is not marked enough to be seen by this type of circular dichroism experiment. The overall structure of the duplex is the same, reflecting the idea that both the PNA:RNA and DNA:RNA parts of the chimera:RNA duplexes are similar in all cases. The fact that the chimera/RNA duplexes have CD spectra close to the structure of a RNA/DNA duplex and of a Gapmer/RNA duplex is encouraging in terms of the ability of RNase H to recognize chimera/RNA duplexes with high efficiency.

4.1.4 RNase H Assays

The different chimera designs were tested for their ability to elicit RNase H cleavage of a complementary RNA target. The classic design of chimera has been reported as being able to recruit RNase H, and our different designs were expected to be able to recruit RNase H as well. The different designs were compared to each other in their ability to induce cleavage of a complementary RNA under different conditions. The RNase H assays were performed with single purified chimeras (ion-exchange HPLC purification) the chimeras used in the assay are therefore mainly the described product with ~20% of the n-1 (from the 5' end) product and ~10% of undesired product (see discussion in section **3.4.6** page 103) we believe that the results obtained in these assays are representing the behaviour of the studied chimeras.

The RNase H assay monitors the degradation of a fully complementary FAM-labelled RNA target over time. A sample of the reaction is taken at different time points and run on a denaturing PAGE gel after the assay. The gels were read with a fluorescence imager and the intensities of the bands were quantitated using ImageJ and normalized to the intensity of the band at t=0 to obtained the percentage of RNA left at each time point. The results are shown as gel images and as a bar graph representing the intensity of the full-length RNA band. Error bars are large in some cases but in the range of expectation for this kind of measurement and assay (**Figure 4-10** to **Figure 4-16**).

It has been shown that RNase H needs a footprint of seven base pairs[33] and the gap in our chimeras is only 7 or 8 thus straining the system and hopefully amplifying the observed differences between design.

Properties of chimeras

The first assay had as its primary objective to verify the ability of the different designs to cleave their target. The different chimeras were placed in the presence of a single stranded, fully complementary RNA, in a 1:1 ratio (**Figure 4-10** and **Figure 4-11**).

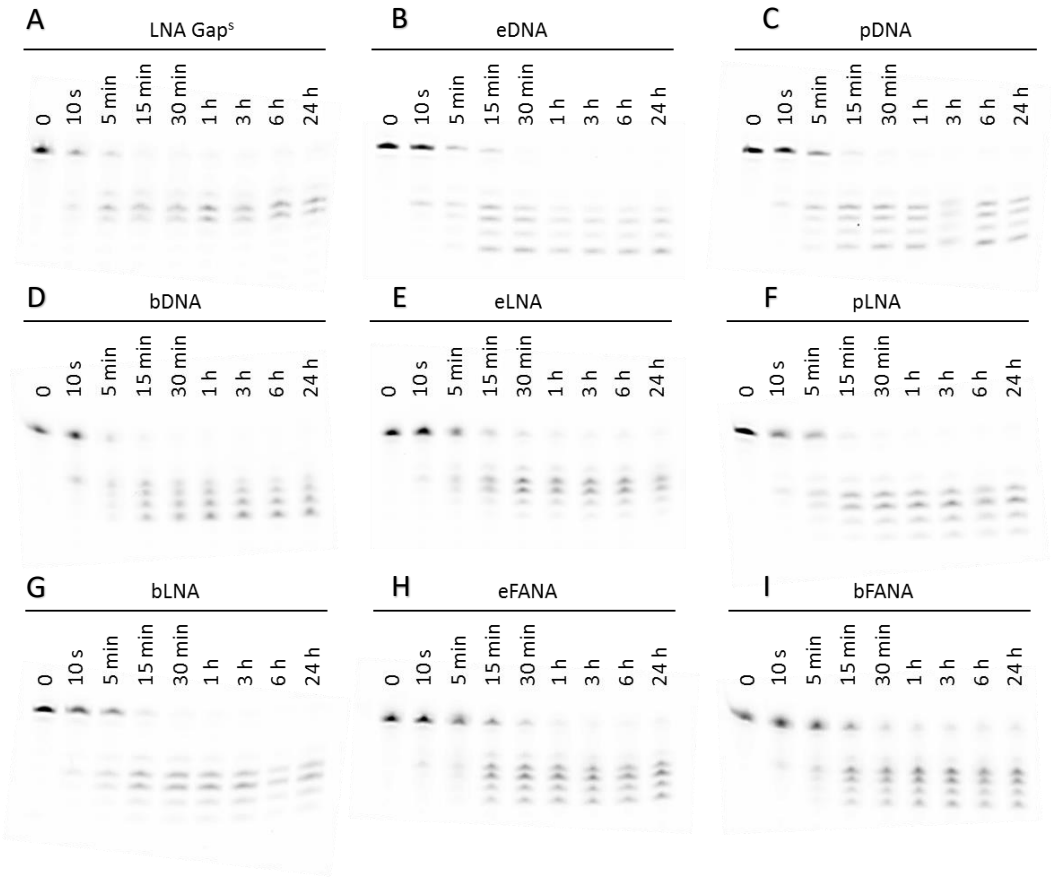


Figure 4-10: 1/1 RNase H assay, ASO with single strand FAM-RNA complementary target. 1 equivalent of FAM labelled RNA was incubated at RT with 1 equivalent of ASO and RNase H1 in 1x RNase H buffer. 10 µL of the assay mixture was taken out at each time point and fixed in formamide and kept at -20 ˚C. The results were run on a 20% PAGE and the gel was observed with a gel imager revealing the FAM-RNA and its cleaved products.

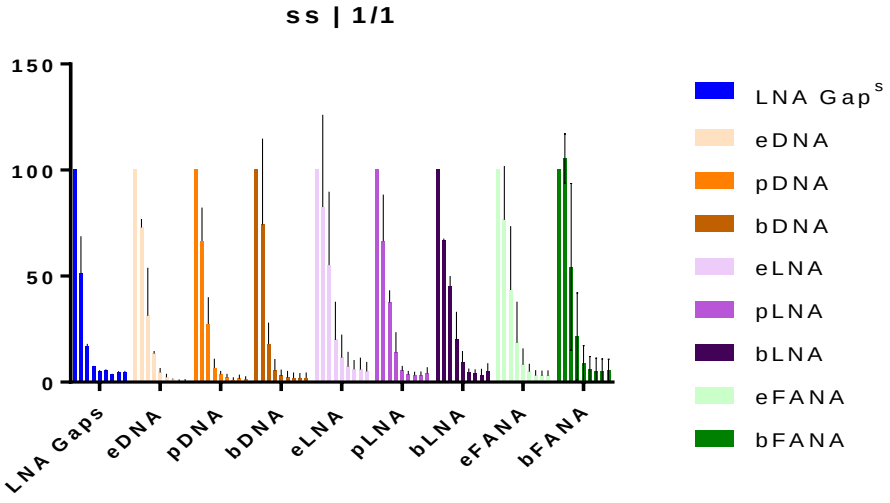


Figure 4-11: Quantitation of full-length RNA remaining after the RNase H assay shown in **Figure 4-10** with ImageJ. Time points are (0, 10 s, 5min, 15min, 30min, 1 h, 3 h, 6h, 24h), mean of 2.

Properties of chimeras

Under these conditions, (**Figure 4-10** and**Error! Reference source not found.**) all chimeras show complete or nearly complete cleavage of the RNA target in less than an hour. The LNA gapmer is used as positive control and performs better than the chimeras. This is most likely due to a bigger gap able to recruit RNase H. The difference observed between the different designs is not big enough to conclude anything; under this set of conditions all chimeras perform similarly.

The second assay was performed under the same conditions against the same RNA target but the ratio of the ASO to RNA was reduced to 1 ASO for 5 RNA. These conditions were designed to examine the catalytic properties of the different designs under multiple turnover conditions. **Figure 4-12** and **Figure 4-13** show that all chimeras have the ability to act catalytically with RNase H. All chimeras have cleaved most of the target by three hours. The LNA gapmer is able to elicit cleavage of all the substrate within the first 1 h **Figure 4-12-A**).

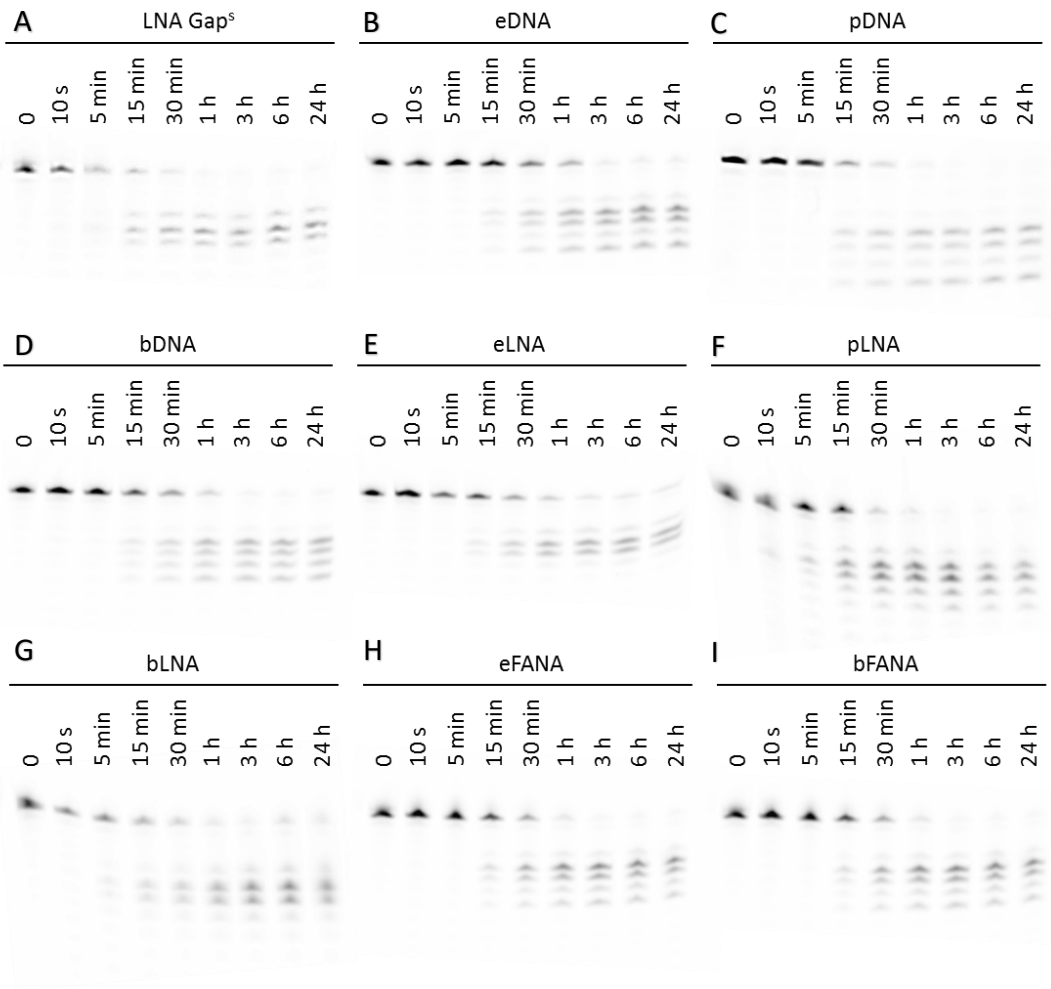


Figure 4-12: 1/5 RNase H assay, ASO with single strand FAM-RNA complementary target. 5 equivalents of FAM labelled RNA was incubated at RT with 1 equivalent of ASO and RNase H1 in 1x RNase H buffer. 10 µL of the assay mixture was taken out at each time point, fixed in formamide and kept at -20 ˚C. The results were run on a 20% PAGE and the gel was observed with a gel imager revealing the FAM-RNA and its cleaved products.

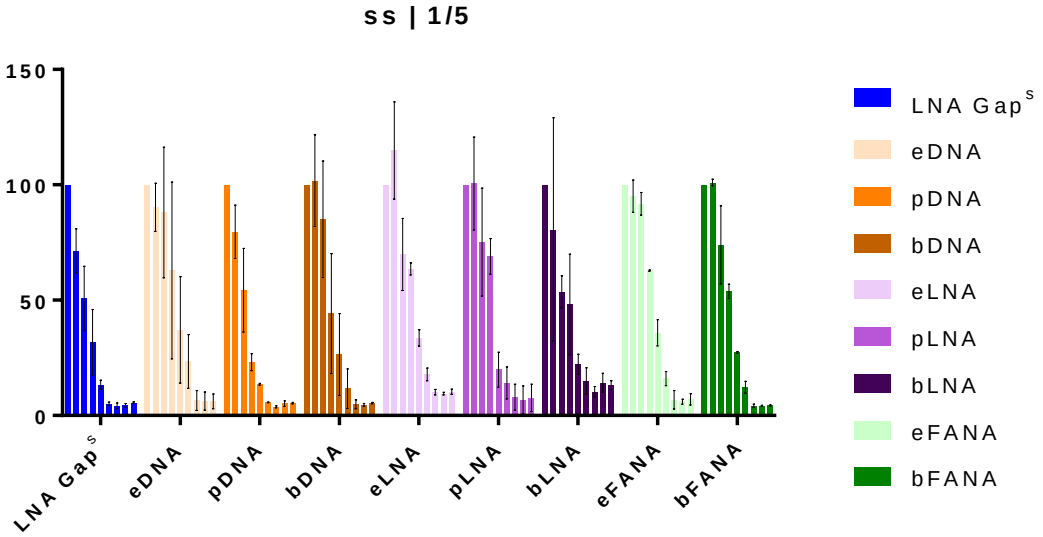


Figure 4-13: Quantitation of full-length RNA remaining after the RNase H assay shown in **Figure 4-12** with ImageJ. Time points are (0, 10 s, 5min, 15min, 30min, 1 h, 3 h, 6h, 24h), mean of 2.

Properties of chimeras

The third assay was designed to study the ability of the chimeras to invade a structured RNA target. As such, it was performed under the same conditions as the first experiments with a 1 to 1 ration of ASO to RNA. The RNA target sequence was designs in a way that it forms a hairpin (**Figure 4-14)**.

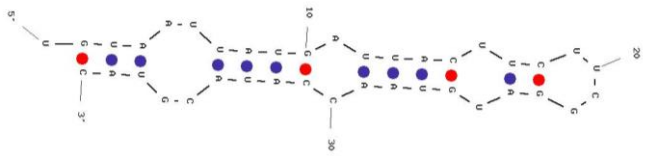


Figure 4-14: Structure of the RNA used as a target in the third RNase H assay. Predicted structure obtained from the mfold server.[129]

Under the conditions of the assay, none of the chimeras was able to cleave all of the targeted RNA after 24 h. Relatively better efficiency were obtained with bDNA and eLNA (**Figure 4-16-D** and **E)** but the assay does not allow us to draw conclusion from this. It is interesting to observe that the standard LNA gapmer design does not show complete cleavage of the hairpin after 24 h either.

Overall, the different designs of chimeras show similar efficiency in these RNase H assay. Within error the designs containing DNA at the junction (eDNA, pDNA and bDNA) seem to perform slightly better but it is difficult to know if it is due to a better conformation or due to the fact that having DNA at the linkage effectively increases the length of the gap able to be recognised by RNase H1 from 7 to 8 nucleotides. We can observe as well that despite not having an extra deoxynucleotide, eLNA shows an efficiency similar to the DNA chimeras. The assay does not allow us to claim that any chimera designs is better than the other and while chimeras are able to recruit RNase H in a catalytic fashion and be able to bind and elicit cleavage of structured target, these designs do not present significant advantages compared to a gapmer in these conditions.

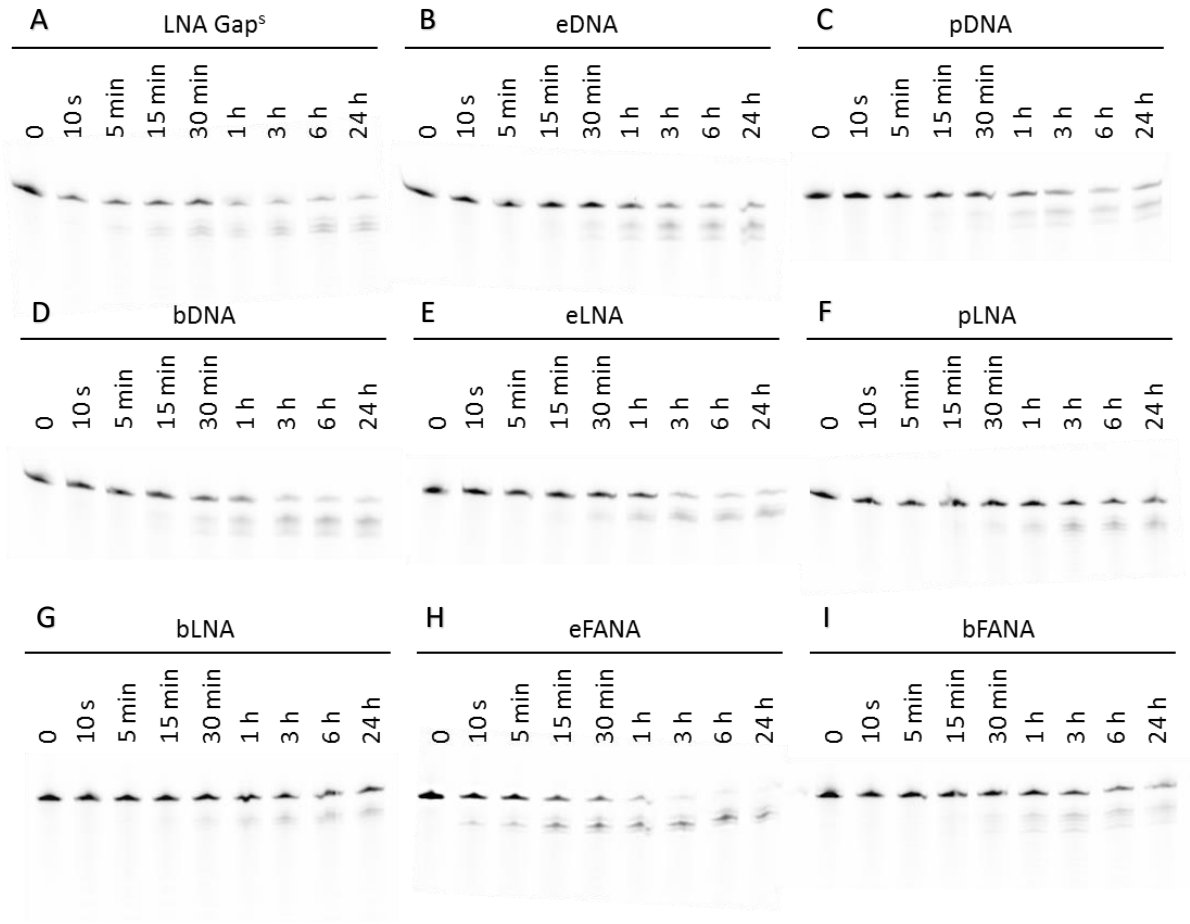


Figure 4-15: 1/1 RNase H assay, ASO with a hairpin FAM-RNA complementary target. 1 equivalent of FAM labelled RNA was incubated at RT with 1 equivalent of ASO and RNase H1 in 1x RNase H buffer. 10 µL of the assay mixture was taken out at each time point and fixed in formamide and kept at -20 ˚C. The results were run on a 20% PAGE and the gel was observed with a gel imager revealing the FAM-RNA and its cleaved products.

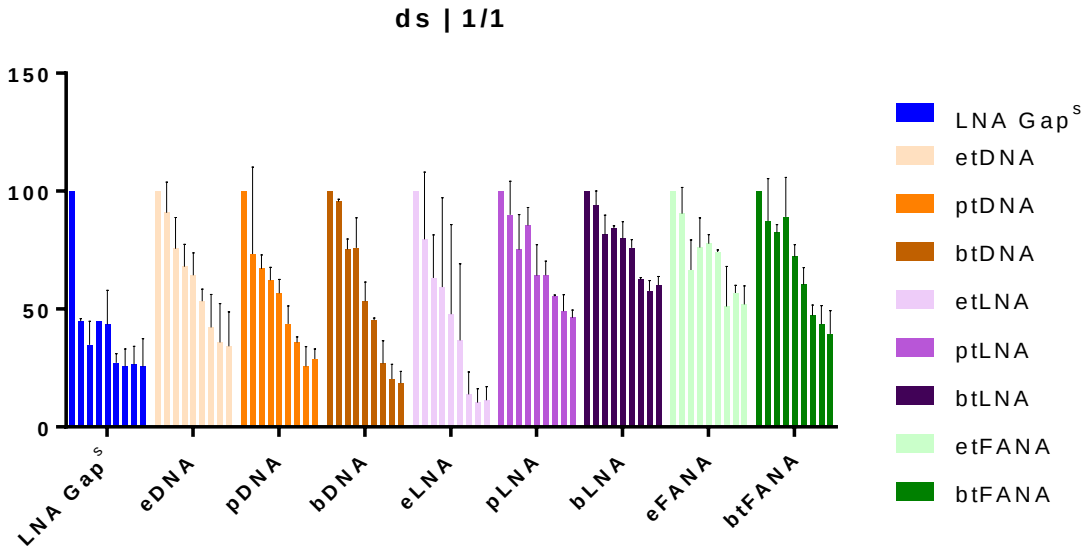


Figure 4-16: Quantitation of full-length RNA remaining after the RNase H assay shown in **Figure 4-15** with ImageJ. Time points are (0, 10 s, 5min, 15min, 30min, 1 h, 3 h, 6h, 24h), mean of 2.

4.2 Biological activity of DNA-PNA Chimeras

4.2.1 Targeting MALAT1: design and sequences of chimeras

After validation of the ability of the chimeras to recruit RNase H, their silencing efficacy was investigated in cultured cells. The RNA target chosen for the study was MALAT1, a noncoding RNA expressed at high levels in most cell types and tissues. MALAT1 has therapeutic relevance as it is associated with metastatic tumours, and the knockdown of MALAT1 has been shown to reduce the incidence of metastasis.[181] Another reason for choosing this target is that it is a non-coding RNA mainly localized in the nucleus.[182] One of our hypotheses about the advantages of chimeras is that they would be able to retain some of the ability of PNA to enter the nucleus and target structured RNA and DNA. The sequence used for the study was an adaptation of the MALAT1 ASO2 sequence published by Zhang *et al.*[183] The reported sequence showed good silencing in vivo without signs of toxicity. The original ASO was a 5-9-5 fully PS MOE gapmer (**Figure 4-17**), but the sequence was shortened to make an LNA gapmer as positive control for the experiment (**Table 4-4**).

MALAT 1ASO2 5'-GGG TCA GCT GCC AAT GCT AG-3'

Figure 4-17: Gapmer sequence from *Zhang et al.*[183] Underlined bases are 2'-MOE and the gapmer is fully modified with a phosphorothioate backbone.

The PNA chimera sequences were designed from this sequence with the constraint that the PNA linker available has to be a T (sequence shown in **Table 4-4**).

Only two types of chimeras were synthesised for this study: since the RNase H assays didn't show any significant differences between most of the chimera designs, we focused on only two designs (either LNA or DNA at the junction coupled with the ethyl PNA linker, **Table 4-4**). The DNA chimera (MALAT1_Ch_DNA) can be used as reference to compare with prior work on chimeras and the LNA chimera (MALAT1_Ch_LNA) is one of the most promising designs from the biophysical study and one of the more available in terms of the amount of starting material we had available. The chimeras were fully phosphorothioate to ensure nuclease stability to the DNA gap in the chimera.

Name	Sequence	Expected Mass	Observed Mass
MALAT1_Ch_LNA	$+G_s+G_s+T_sC_sA_sG_sC_sT_sG_sC_sC_sA_s+A_s$ et gctag	6059.07	6059.0
MALAT1_Ch_DNA	$+G_s+G_s+T_sC_sA_sG_sC_sT_sG_sC_sC_sA_sA_s$ eT gctag	6031.06	6031.0
MALAT1_Gap_LNA	$+G_s+G_s+T_sC_sA_sG_sC_sT_sG_sC_sC_sA_s+A_s+T_s+G$	4985.90	4985.5
NTC_Gap_LNA	$+C_s+T_s+T_sC_sC_sC_sT_sG_sA_sA_sG_sG_s+T_s+T_s+C$	5907.00	5907.0
NTC_Ch_DNA	$+C_s+T_s+T_s\ C_sC_sC_sT_sG_sA_sA_sG_sG_sT_s$ eT cctcc	4939.90	4939.5

Table 4-4: Sequences of the ASOs used for the MALAT1 knockdown study. Targeted sequences and non-targeting sequences (NTC) were adapted from *Zhang et al.*[183] "N" is DNA; "+N" is LNA, et is ethyl PNA linker, "n" is PNA and "$_s$' is phosphorothioate linkage.

The silencing of MALAT1 by the chimeras was investigated in mouse embryonic fibroblasts (MEFs) by lipid transfection and in mouse hepatocellular carcinoma (Hepa1-6) cells by passive delivery. I designed the experiment and processed the data. However, the cell culture, transfections and QuantiGene assays were performed by Pranathi Meda.

4.2.2 Lipid transfection

First the efficacy of the chimeras was investigated by lipid transfection. In this experiment, cationic lipids are used to form a complex with the ASOs and enable them to cross the cellular membrane.

We first treated Hepa 1-6 cells with the ASOs (0 to 50 nM) and observed that cells treated with active ASO showed a strong morphology change and associated toxicity (decrease in cell number). Cells that were given the chimera NTC or gapmer NTC by lipid transfection did not show any change in phenotype, suggesting that the morphology change was not due to toxicity from the chemistry of the ASO. Therefore, the change could have been related to the silencing of MALAT1. Hepa 1-6 cells may depend on MALAT1 expression for growth.[184] Therefore, the dose response experiment was repeated in mouse embryonic fibroblast (MEF) cells and no morphology change or toxicity was observed.

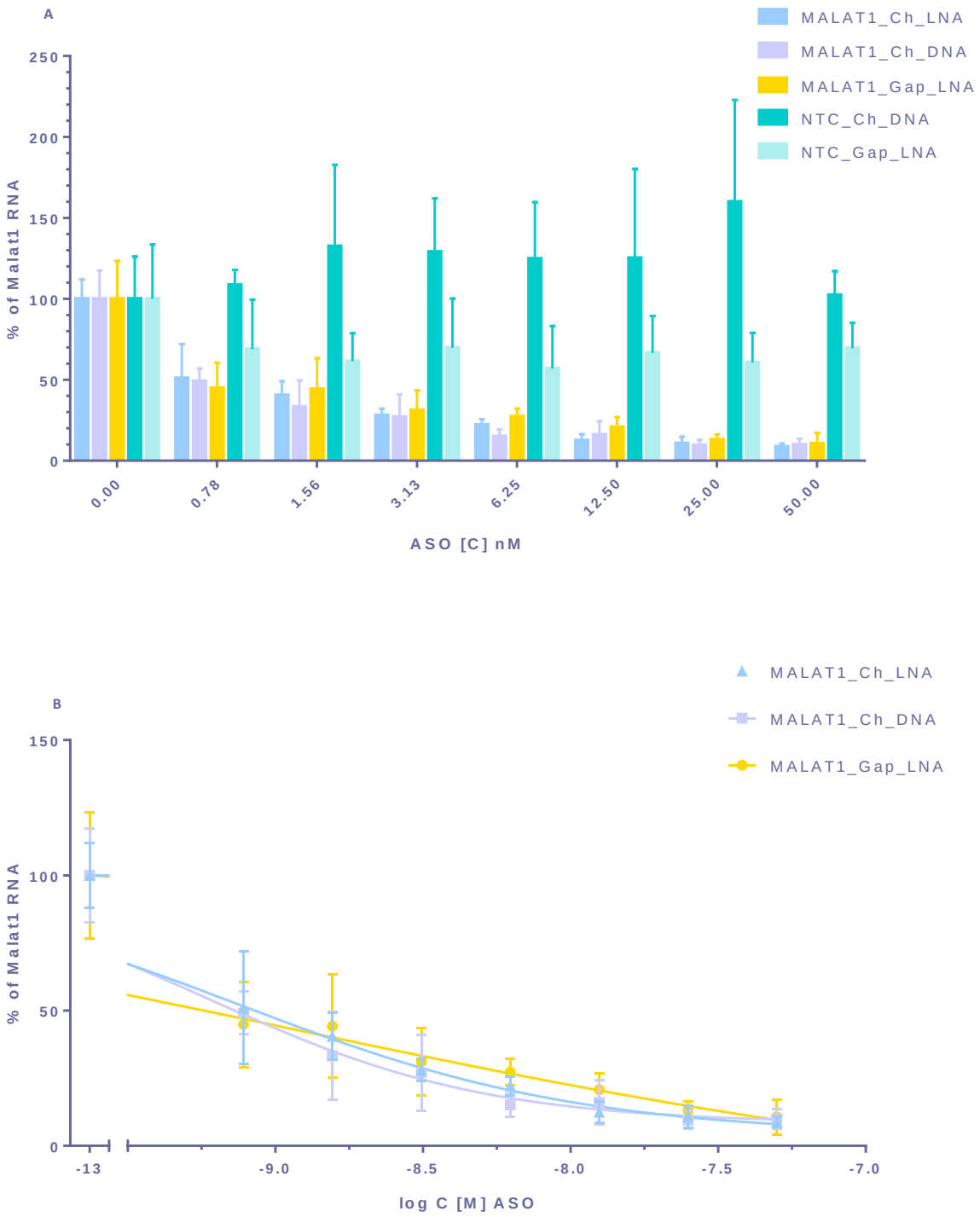


Figure 4-18: (A) Level of MALAT1 RNA observed in MEF cells 16 h after transfection of MALAT1 ASO. (B) Dose response curve with logarithmic x-axis, used to obtain the IC_{50} values.

Figure 4-18-A shows that the silencing activity of chimeras is as potent as the LNA gapmer. 90% silencing is achieved at 25 nM concentration and 50% of silencing is still achieved at 0.8 nM. The difference of silencing observed between the DNA and the LNA-containing chimera designs is too small to differentiate, and similarly, no significant difference in silencing between the LNA gapmer and the chimera can be observed. These data show that the PNA wing is not reducing the ability of the ASO to find its target, bind to it and recruit RNase H when being in the cell.

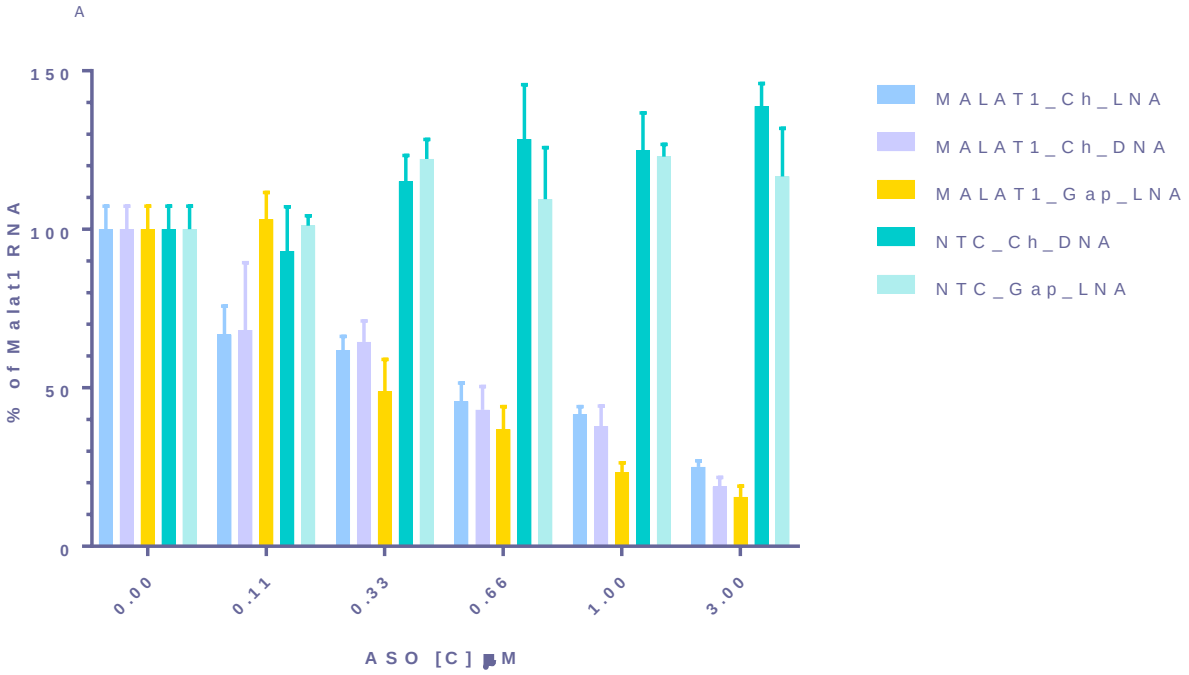
ASO	IC_{50} [nM]	R square
MALAT1_Ch_LNA	0.8	0.92
MALAT1_Ch_DNA	0.6	0.89
MALAT1_Gap_LNA	1.2	0.82

Table 4-5: IC_{50} of the ASO for knockdown of MALAT1 RNA with lipid transfection at 16 h. Calculated with interpolation of the level of RNA observed by the log of the concentration.

The **Table 4-5** shows that the calculated IC_{50} of the three ASO tested is comparable. Thus when the size of the DNA gap is maintained, our LNA-DNA-PNA chimeras are of similar efficacy as LNA gapmers when transfected into cells. An IC_{50} of 1 nM is considered excellent which is very promising for the chimeras.

4.2.3 Gymnotic delivery

After having demonstrated their efficacy by lipid transfection, the efficacy of chimeras was investigated by gymnotic delivery. In gymnotic delivery, the ASO is added into the growth media of the cells and taken up into the cells by endocytosis. In this experiment, Hepa 1-6 cells were treated with ASO 1 day after seeding and grown for 6 days before harvest and quantification of the level of RNA. The cells didn't show any morphological changes or cell death for any of the ASO chimeras or gapmer, targeted or non-targeted.



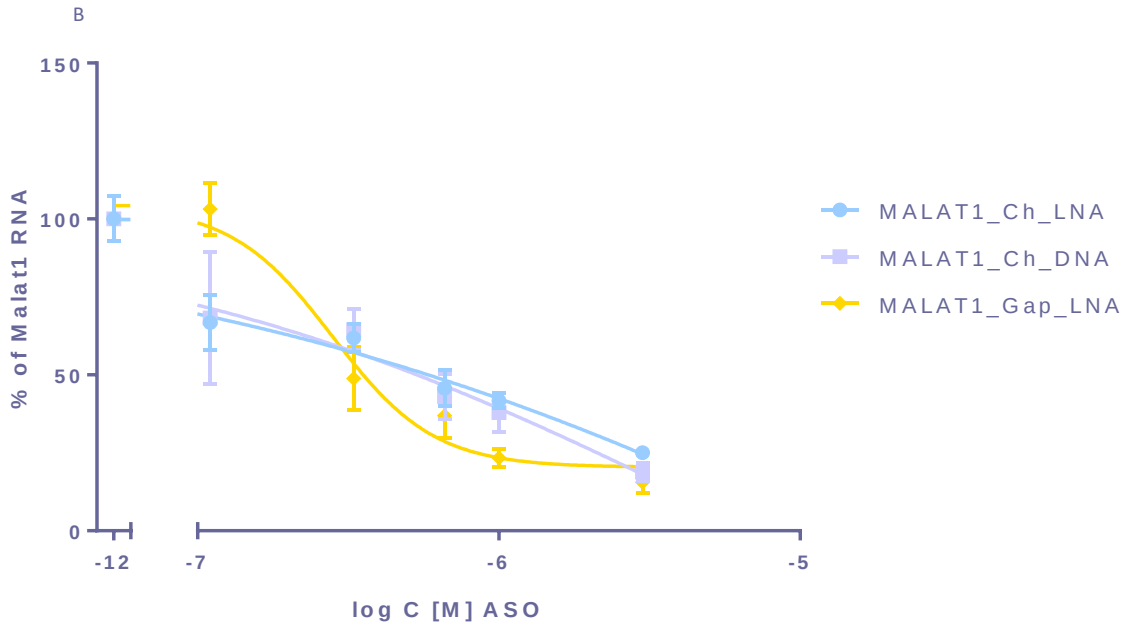


Figure 4-19: (A) Level of MALAT1 RNA observed in HEPA 1-6 cells 6 days after passive delivery of MALAT1 ASO. (B) Dose response curve with logarithmic x-axis, used to obtain the IC50 values.

The **Figure 4-19** shows that the silencing activity of the chimeras is of the same order as the LNA gapmer. However, the LNA gapmer seems to be more efficacious at higher dose while chimeras retain silencing activity at lower dose when the gapmer lose all efficacy. Chimeras exerted over 70% silencing of RNA level at 3 µM while the gapmer showed 80% silencing (**Figure 4-19**). Meanwhile, at 0.1 µM, the gapmer has lost all efficacy but the chimeras still show 25% knockdown (**Figure 4-19**).

ASO	IC_{50} [µM]	R^2
MALAT1_Ch_LNA	NA	NA
MALAT1_Ch_DNA	4.6	0.88
MALAT1_Gap_LNA	0.3	0.95

Table 4-6: IC_{50} of the ASO for knockdown of MALAT1 RNA by gymnotic delivery (harvested at 6 days).

The calculated IC_{50} values shown in **Table 4-6** indicated a ten-fold potency difference between the DNA chimeras and gapmer. However, the results should be interpreted carefully, as neither of the chimeras showed the sigmoidal dose response that we would expect for this kind of dose response curve. The program wasn't able to calculate the IC_{50} for the LNA chimera, but its curve is very similar to the DNA chimera. The IC_{50} value of 0.3 µM for the gapmer is excellent in the context of gymnotic delivery, and an order of 5 µM for the chimeras is still in a typical range for ASO-mediated gene silencing by gymnosis. The difference in the profile of the dose response of the chimeras and the gapmer could be a hint that the mechanism of uptake is different for the two different chemistries, or the endosmal escape could happen differently.

In both lipid-mediated and gymnotic delivery, the efficacy shown by the chimeras is considered very good and justify further exploration of PNA-containing chimeric ASOs as therapeutics. In both types of gene silencing experiments, the difference between the chimeras containing LNA or DNA at the DNA-PNA junction is insignificant.

Chapter 5: **Conclusions and future work**

This work has explored new ways to use peptide nucleic acids in gene regulation. Our focus has been on the production of modified PNA to use in tandem with other established silencing oligonucleotides.

The first part explored the use of PNA and LNA as an off-switch for a synthetic small RNA that regulates the translation of a mRNA. Our work focused on design and synthesis of optimised sequences to be used in this application. Three sequences were synthesized using PNA and LNA-DNA mixmer designs to bind and inhibit this small RNA. One of the sequences showed promising results in vitro and we developed PNA-peptide conjugates for in vivo work. The cell penetrating peptide chosen for the in vivo study seems to have a slight toxicity, so the design could be optimized by better tailoring of the peptide to the application.

The second approach investigated in this thesis was the use of a PNA as a sense strand for siRNA. Our work showed that the PNA sense strand neither enhanced nor inhibited uptake of the antisense strand by the RISC complex. While it is of some interest that the antisense strand uptake is not inhibited, this type of construct does not present significant interest for improving gene silencing.

The third approach to use PNA in tandem with existing therapeutic oligonucleotide technology described the use of PNA as a peptide carrier for delivery of chemically modified siRNA. The present work described the synthesis and physical properties of “three-component duplexes” called PNA-peptide tail siRNA (PPT-siRNA). We obtained promising preliminary in vitro silencing data for the PPT-siRNA, which needs to be repeated for confirmation. The methods to ascertain the formation of the three-component duplexes should also be optimized. And last but not least, the use of different peptides, especially tissue-specific receptor-targeting peptides, should be investigated.

In the second half of this thesis, the development of a potent LNA-DNA-PNA chimeric ASO was described. We described the synthesis of the custom monomers needed to synthesise these chimeras. A robust synthesis is presented for each monomer affording the desired monomers with high purity. However, the synthesis of the monomers needs to be optimised to be able to routinely afford the monomers in scales over 10 g. Methods for solid phase synthesis of the chimera on Expedite 8909 and ABI 394 instruments are discussed and afforded the desired chimeras in high purity. The limited yield of the synthesis is at least partly related to the limitations of the synthesizers used in this work.

Conclusions and future work

The biophysical properties of 8 different chimera designs were investigated showing very little difference in the shape and activity of the chimeras even in the case of very distinct chemistry at the junction of the PNA and DNA part in the chimera.

Finally, two selected chimera designs were tested for silencing efficacy in cells by lipid transfection and gymnotic (lipid-free) delivery. ASOs were designed to target MALAT1 RNA and the silencing activity of the chimeras was found to be comparable to that of an LNA gapmer. These results are very promising for the use of our chimera design in therapeutics. The uncharged nature of PNA might confer significantly different tissue distribution properties than fully charged ASOs, and future work will include an exploration of the ability of PNA chimeras to silence gene expression in mice.

Chapter 6: **Experimental section**

6.1 Organic chemistry

6.1.1 General

When inert conditions were required, all apparatus was oven-dried overnight and argon gas was used with standard Schlenk and syringe or cannula techniques. For heating, an oil bath set to the desired temperature was used. Reactions were monitored by thin-layer chromatography using Merck TLC silica gel 60 F_{254} plates. All chemicals required were obtained from Fisher, Sigma Aldrich, VWR, Chem-Impex or Fluorochem and were used without further purification. ^{1}H, ^{13}C, ^{31}P NMR spectra were recorded on Bruker DPX or Bruker AV operating at 400, 101 and 162 MHz respectively. Automated column chromatography was carried out using an Isolera Four instrument from Biotage.

ESI+ mass spectrometry experiments were performed on a Waters (Manchester, UK) TQD mass spectrometer equipped with a triple quadrupole analyser. Samples were introduced to the mass spectrometer via an Acquity H-Class quaternary solvent manager (with TUV detector at 254nm, sample and column manager). Ultra-performance liquid chromatography was undertaken via a Waters BEH C18 column (50mm x 2.1mm 1.7µm). Gradient 20% acetonitrile (0.2% formic acid) to 100% acetonitrile (0.2% formic acid) in five minutes at a flow rate of 0.6 mL/min.

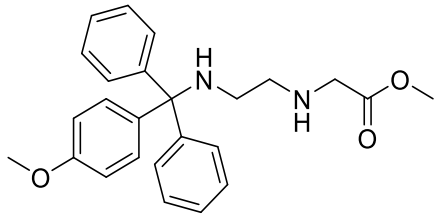
ESI- mass spectrometry experiments were performed on an Agilent 6530 Q-TOF LC/MS spectrometer with electrospray and time of flight in positive ionization mode. High Performance Liquid Chromatography was undertaken via an Agilent Poroshell 120 EC-C18 column (4.5 mm x 50 mm 2.7 µm). Gradient 2% B to 100% B over 10 min at a flow rate of 1 ml/min. Buffer A: 100 mM hexafluoroisopropanol with 10 mM triethylammonium acetate; Buffer B: 20 mM triethylammonium acetate and acetonitrile.

Quantifications of Fmoc loaded resin were carried out by deprotecting 10 mg of resin in 10 mL of 20% piperidine solution in DMF for 10 min. The resin was then filtered out and the absorbance at 300 nm of the deprotecting solution was taken in a UV/Vis spectrometer. The calculation of the loading was performed using the Beer-Lambert law and the extinction coefficient=6566.

Quantifications of MMT loaded resin were carried out by deprotecting 10 mg of resin in 10 mL of 3% TCA in DCM for 10 min. The resin was then filtered out and the absorbance at 470 nm of the

deprotecting solution at 1/100 dilution was taken in a UV/Vis spectrometer. The calculation of the loading was performed using the Beer-Lambert law and the extinction coefficient=56000.

6.1.2 Procedures



***N*-[2-(4-Methoxyphenyl)diphenylmethylaminoethyl] glycine methyl ester (3)**

A solution of 4-methoxytriphenylchloromethane (4.5 g, 14.6 mmol, 1 eq) in pyridine (30 mL) was added dropwise to a solution of ethane-1,2-diamine (9.74 mL, 145.8 mmol, 9 eq) in DCM (40 mL) at 0°C. The mixture was stirred for 15 h and 10 mL of MeOH was added. Then mixture was concentrated by rotary evaporation and solubilised in DCM with 1% of NEt_3. The solution was washed with water and the solvent was removed by rotary evaporation. R_f =0.9 (8:2 DCM/MeOH). The recovered yellow oil was solubilised in dry DCM, and NEt_3 (12 mmol, 1.7 mL, 0.8 eq) was added. Then, methyl bromoacetate (12 mmol, 1.83 g, 0.8 eq) was added dropwise while stirring at 0°C. The reaction mixture was stirred at RT for 16 h. When the reaction was completed, a solution of 20% $NaHCO_{3\,(aq)}$ (20 mL) was added and the mixture was extracted by DCM (3x20 mL). The combined organic phases were dried over $MgSO_4$, the solvent was removed by rotary evaporation, and the residue purified by automated silica gel chromatography using increasing amounts (15-50%) of Ethyl acetate in hexane as eluent, to afford **3** as a yellow oil (1.61 g, 3.9 mmol, 28%)

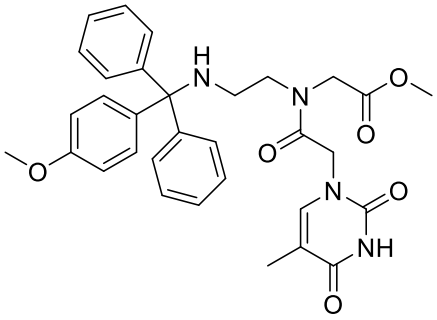
R_f=0.1 (7:3 Hexane/EtOAc)

^{1}H NMR (300 MHz, $CDCl_3$) δ 7.15 - 7.51 (m, 12 H, MMTH), 6.78 - 6.86 (m, 2 H, MMTH), 3.79 (s, 3 H, MMTOCH_3), 3.73 (s, 3 H, COOCH_3), 3.36 (s, 2 H, NCH_2CO), 2.74 (t, *J*=5.8 Hz, 2 H, CH_2NHCH_2), 2.28 (t, *J*=5.8 Hz, 2 H, MMTNHCH_2), 1.71 - 1.95 (m, 2 H, NH)

^{13}C NMR (75 MHz, $CDCl_3$) δ 173.1 157.3 146.4 138.3 129.8 128.6 127.8 126.1 113.1 70.2 55.2 51.7 50.6 49.8 43.1

LRMS (ESI) calcd for $C_{25}H_{28}N_2O_3$+Na $[M+Na]^+$ *m/z*=427.2 found 427.1

Analyses are consistent with the literature[164]



_N_-(2-(4-Methoxyphenyl)diphenylmethylaminoethyl)-_N_-(1-thyminyl)acetyl glycine methyl ester (4T)[153]

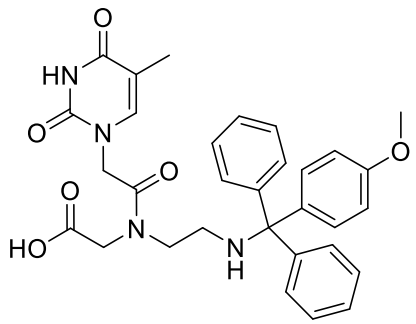
Under a dry argon atmosphere, to thymine-acetic acid **(7)** (300 mg, 1.63 mmol, 1 eq) was added a solution of *N*-[2-(4-methoxyphenyl)diphenylmethylaminoethyl]glycine methyl ester **(3)** (658 mg, 1.63 mmol, 1 eq) in DMF (5 mL). Then DIPEA (0.85 mL, 3.26 mmol, 2 eq) was added and the solution was cooled to 0°C before T3P (1.43 mL, 2.45 mmol, 1.5 eq) was added dropwise. The solution was stirred for 20 h at RT. After completion, the mixture was concentrated by rotary evaporation and dissolved in ethyl acetate. Product was washed with saturated $NaHCO_{3aq}$ and water (2x), dried over $MgSO_4$ and concentrated by rotary evaporation to afford a beige solid. The product was then dissolved in 2 mL of DCM and precipitated in Petroleum ether to afford **4T** as a white solid (759 mg, 77%, 1.33 mmol).

R_f: 0.05 in DCM:MeOH 99:1.

^{1}H NMR (400 MHz, DMSO-d_6) δ 11.29 (s, 1 H, NH-3), 7.15 - 7.41 (m, 13 H, MMTH and H-6), 6.85 (d, *J*=8.3 Hz, 2 H, MMTH), 4.86 (ma.) and 4.50 (mi.) (s, 2 H, NCH_2CO), 4.45 (mi.) and 4.02 (ma.) (s, 2 H, CH_2COOCH_3), 3.72 (s, 3 H, $MMTOCH_3$), 3.57 (m, 3 H, $COOCH_3$), 3.48 (ma.) and 3.36 (mi.) (m, 2 H, $CONCH_2CH_2$), 2.16 (ma.) and 2.09 (mi.) (m, 2 H, MMT-N -CH_2), 1.75 (s, 3 H, CH_3-5)

LRMS (ESI) calcd for $C_{32}H_{34}N_4O_6$ $[M+H]^+$ *m/z*=570.25 found 571.1

Analyses are consistent with the literature[153]



***N*-(2-(((4-Methoxyphenyl)diphenylmethyl)amino)ethyl)-*N*-(2-(5-methyl-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)acetyl)glycine (5T)[153]**

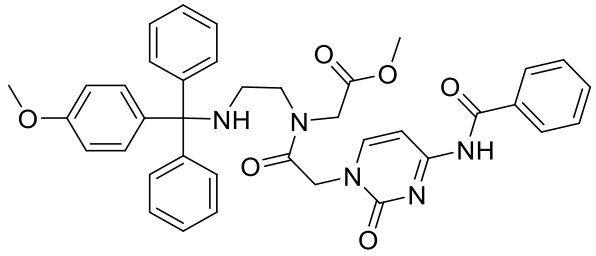
N-(2-(4-methoxyphenyl)diphenylmethylaminoethyl)-*N*-(1-thyminyl)acetyl glycine methyl ester **(4T)** (624 mg, 1.09 mmol) in 30 mL of MeOH is stirred for 2 h with 2M NaOH solution at RT. Then Dowex® pyridinium was added in portions until the solution reached pH 7. The resin was removed by filtration and the filtrate was dried by rotary evaporation. The compound was purified by column chromatography on SiO_2 (eluted with 8:1:1 to 7:2:1 EtOAc/MeOH/H_2O) to afford compound **5T** as a beige solid (587mg, 97%, 1.06 mmol).

R_f =0.25 in EtOAc:MeOH:H_2O 7:2:1.

^{1}H NMR (400 MHz, DMSO-d_6) δ 11.25 (br s, 1 H, NH-3), 7.07 - 7.48 (m, 13 H, MMTH and H-6), 6.84 (d, *J*=8.4 Hz, 2 H, MMTH), 4.81 (mi.) and 4.46 (ma.) (s, 2 H, NCH_2CO), 3.91 (mi.) and 3.80 (ma.) (s, 2 H, CH_2COOH), 3.71 (s, 3 H, $MMTOCH_3$), 3.34 (m, 2 H, NCH_2CH_2NH), 2.17 (mi.) and 2.11 (ma.) (m, 2 H, $NHCH_2CH_2N$), 1.73 – 1.75 (m, 3 H, CH_3-5)

LRMS (ESI) calcd for $C_{31}H_{32}N_4O_6$+Na $[M+Na]^+$ *m/z*=579.2 found 580.3

Analyses are consistent with the literature[153]



M**ethyl *N*-(2-(4-benzamido-2-oxopyrimidin-1(2H)-yl)acetyl)-*N*-(2-(((4-methoxyphenyl)diphenylmethyl)amino)ethyl)glycinate (4C)[156, 185]**

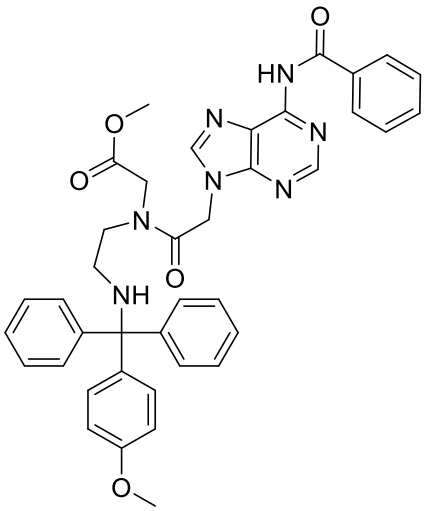
Under a dry atmosphere, to 2-(4-benzamido-2-oxopyrimidin-1(2H)-yl) acetic acid **(15)** (383 mg, 1.03 mmole, 1 eq) was added a solution of *N*-[2-(4-methoxyphenyl)diphenylmethylaminoethyl] glycine methyl ester **(3)** (419 mg, 1.03 mmol, 1 eq) in DMF (5 mL). Then DIPEA (0.54 mL, 2.06

mmole, 2 eq) was added, the solution was cooled to 0°C and T3P (0.9 mL, 1.55 mmol, 1.5 eq) was added dropwise. The solution was stirred for 20 h. After completion, the mixture was concentrated by rotary evaporation and dissolved in ethyl acetate. The product was washed with saturated $NaHCO_{3\,(aq)}$ and H_2O (2x), dried over $MgSO_4$ and concentrated under vacuum. The residue was dissolved in DCM (2 mL) and precipitated in petroleum ether to afford **4C** as a white solid (416 mg, 61%, 0.63 mmol).

^{1}H NMR (400 MHz, DMSO-d_6) δ 11.22 (br s, 1,) 6.74 - 8.19 (m, 21 H, ArH, CH-6, CH-5) 5.04 (ma.) and 4.73 (mi.) (s, 2H, NCH2CON) 4.53 (mi.) and 4.05 (ma.) (s, 2 H) 3.73 (s, 3 H, MMTOCH_3) 3.71 (mi.) and 3.58 (ma.) (s, 3 H, COOCH_3) 3.54 (ma.) and 3.40 (mi.) (br t, *J*=5.7 Hz, 2 H, NHCH_2CH_2N) 3.12 (ma.) and 2.78 (mi.) (t, *J*=7.8 Hz, 1 H, MMTNH) 2.21(ma.) and 2.12 (mi.) (m, 1 H, MMTNHCH_2)

LRMS (ESI) calcd for $C_{38}H_{39}N_5O_6$ $[M+H]^+$ *m/z*=659.27 found 660.3

Analyses are consistent with the literature of the corresponding ethyl ester[156, 185]



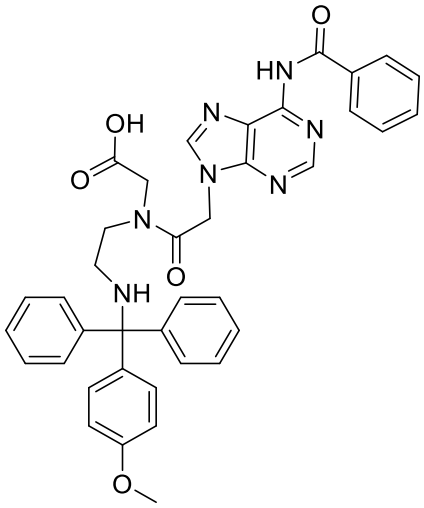
Methyl *N*-(2-(6-benzamido-9H-purin-9-yl)acetyl)-*N*-(2-(((4-methoxyphenyl)diphenylmethyl)amino)ethyl)glycinate (4A)

Under argon atmosphere, to 2-(4-benzamido-2-oxopyrimidin-1(2H)-yl) acetic acid **(21)** (29 mg, 0.09 mmol, 1 eq) was added a solution *N*-(2-(4-Methoxyphenyl)diphenylmethylaminoethyl)-*N*-(1-thyminyl)acetyl glycine methyl ester **(3)** (43 mg, 0.1 mmol, 1.1 eq) in DMF (1 mL). Then DIPEA (0.046 mL, 0.18 mmol, 2 eq) was added, the solution was cooled to 0°C and T3P (0.082 mL, 0.14 mmol, 1.5 eq) was added dropwise. The solution was stirred for 20 h until TLC indication completion of the reaction then, concentrated by rotary evaporation. The residue was dissolved in Ethyl acetate, washed with a saturated solution of $NaHCO_3$ and twice with H_2O, dried over $MgSO_4$ and concentrated under vacuum to afford **4A** as white foam (53 mg, 86%, 0.07).

^{1}H NMR (400 MHz, DMSO-d_6) δ 11.17 (br s, 1 H, NHCOBz), 8.71 (ma.) and 8.69 (mi.) (s, 1 H, CH-8), 8.36 (ma.) and 8.34 (mi.) (s, 1 H, CH-2), 8.07 (m, 2 H, BzH), 7.10 - 7.73 (m, 15 H, Ar-H), 6.81 - 6.93 (m, 2 H, MMTH), 5.59 (ma.) and 5.25 (mi.) (s, 2 H, NCH_2CON), 4.59 (mi.)and 4.08 (ma.) (s, 2 H, NCH_2COOCH_3), 3.72 (m, 3 H, MMTOCH_3), 3.74 (mi.) and 3.65 (s, 3 H, COOCH_3), 3.65 (ma.) and 3.41 (mi.) (m, 2 H, HNCH_2CH_2N), 2.25 (ma.) and 2.12 (mi.) (m, 2 H, HNCH_2CH_2N)

LRMS (ESI) calcd for $C_{39}H_{37}N_7O_5$+H $[M+H]^+$ *m/z*=684.29 found 684.1

Analyses are consistent with the literature[156]

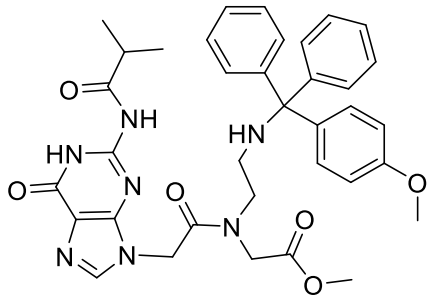


_N_-(2-(6-Benzamido-9H-purin-9-yl)acetyl)-_N_-(2-(((4-methoxyphenyl)diphenylmethyl)amino)ethyl)glycine (5A)

To a solution of methyl *N*-(2-(6-benzamido-9H-purin-9-yl)acetyl)-*N*-(2-(((4-methoxyphenyl)diphenylmethyl)amino)ethyl)glycinate **(4A)** (732 mg, 1.22 mmol, 1 eq) in of MeOH (40 mL) was added aqueous NaOH (20 mL of 2M) and stirred for 4.5 h at rt. Then Dowex® pyridinium was added in portions until the solution reached pH 7. The resin was removed by filtration and washed with MeOH. The combined filtrates were concentrated by rotary evaporation and the residue purified by chromatography on SiO2 (eluent 8:1:1 to 7:2:1 EtOAc/MeOH/H_2O) to afford **5A** as a white foam (166 mg, 20%, 0.25 mmol).

^{1}H NMR (400 MHz, DMSO-d_6) δ 8.69 (mi.) and 8.64 (ma.) (s, 1 H, CH-8), 8.35 (mi.) and 8.30 (ma.) (s, 1 H, CH-2), 7.96-8.10 (m, 2 H, BzH), 7.10-7.73 (m, 15 H, Ar-H), 6.76-6.92 (m, 2 H, MMTH), 5.55 (mi.) and 5.16 (ma.) (s, 2 H, NCH_2CON), 3.90 (ma.) and 3.79 (mi.) (s, 2 H, NCH_2COOH), 3.71 (s, 3 H, MMTOCH_3), 3.60 (mi.) and 3.38 (ma.) (m, 2 H, HNCH_2CH_2N) 2.23 (mi.) and 2.10 (ma.) (m, 2H, NCH_2CH_2N)

R_f: 0.4 in 8/1/1.



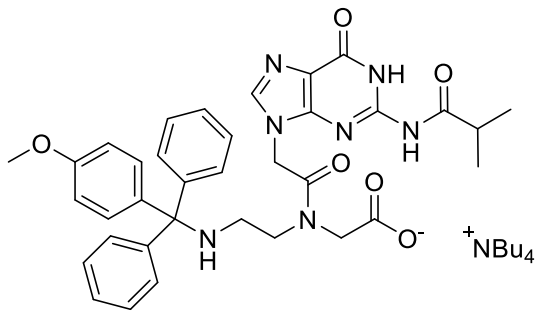
Methyl *N*-(2-(2-isobutyramido-6-oxo-1,6-dihydro-9H-purin-9-yl)acetyl)-*N*-(2-(((4-methoxyphenyl)diphenylmethyl)amino)ethyl)glycinate (4G)[164]

Under a dry argon atmosphere, to 2-(2-isobutyramido-6-oxo-1,6-dihydro-9h-purin-9-yl)acetic acid **(25)** (300 mg, 1.07 mmol, 1 eq) was added a solution of *N*-[2-(4-methoxyphenyl)diphenylmethylaminoethyl]glycine methyl ester **(3)** (433 mg, 1.07 mmol, 1 eq) in DMF (5 mL). Then DIPEA (0.557 mL, 2.15 mmol, 2 eq) was added, the solution was cooled to 0°C, and T3P (0.939 mL, 1.61 mmol, 1.5 eq) was added dropwise. The solution was stirred for 3.5 h, and then concentrated by rotary evaporation. The residue was taken up in ethyl acetate, washed with a saturated solution of $NaHCO_3$ and with H_2O, dried over $MgSO_4$ and concentrated under vacuum. The residue was then dissolved in a minimal volume of hot DCM, precipitated in petroleum ether and filtered to afford **4G** beige sticky solid (239 mg, 34%, 0.36 mmol).

^{1}H NMR (400 MHz, DMSO-d_6) δ 12.08 (brs, 1 H, NHCO$_{ibu}$), 11.61 (brs, 1 H, NH-1), 7.82 and 7.80 (s rotamers, 1 H, H-8), 7.06 - 7.43 (m, 12 H, MMTH), 6.80 - 6.91 (m, 2 H, MMTH), 5.21 (ma.) and 4.98 (mi.) (s, 2 H, NCH$_2$CON), 4.48 (mi.) and 4.07 (ma.) (s, 2 H, CH$_2$COOCH$_3$), 3.71 (br s, 3 H, MMTOCH$_3$), 3.58 (s, 3H, COOCH$_3$), 3.55 (ma.) and 3.39 (mi.) (m, 2 H, CONCH$_2$CH), 3.80 (ma.) and 3.17 (mi.) (br t, *J*=7.9 Hz, 1 H, MMT-NH) 2.76 (m, 1 H, CH$_{ibu}$), 2.27 (ma.) and 2.09 (mi.) (m, 2H, MMTNHCH$_2$), 1.11 (d, *J*=6.9 Hz, 6 H, CH$_{3ibu}$)

LRMS (ESI) calcd for $C_{36}H_{39}N_7O_6$ $[M+Na]^+$ *m/z*=688.29 found 689.5

Analyses are consistent with the literature[169]



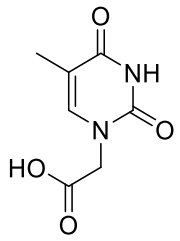
_N_-(2-(2-Isobutyramido-6-oxo-1,6-dihydro-9H-purin-9-yl)acetyl)-_N_-(2-(((4-methoxyphenyl)diphenylmethyl)amino)ethyl)glycinate (5G)[169]

To a stirred solution of methyl N-(2-(2-isobutyramido-6-oxo-1,6-dihydro-9H-purin-9-yl)acetyl)-N-(2-(((4-methoxyphenyl)diphenylmethyl)amino)ethyl)glycinate (1.3 g, 1.9 mmol, 1 eq) in dioxane (10 mL), was added a solution of 0.5M tetrabutylammonium hydroxide 30% hydrate in dioxane until pH12. The reaction was stirred for 1.5 h then concentrated by rotary evaporation and purified by automated reverse phase chromatography (SNAP-Ultra C18, 10g column), eluted with a gradient of acetonitrile in water. The product was isolated as white foam (1.4g, 78%, 1.5 mmol).

^{1}H NMR (500 MHz, DMSO-d_6) δ 12.01 (br s, 1 H, NHCO$_{ibu}$) 7.77 (s, 1 H, CH-8) 7.13 - 7.43 (m, 12 H, MMTH) 6.79 - 6.86 (m, 2 H, MMTH) 5.17 (mi.) and 4.92 (ma.) (s, 2 H, NCH$_2$CON) 3.77 (s, 2 H, NCH$_2$COO) 3.71 (s, 3 H, MMTOCH$_3$) 3.50 (mi.) and 3.38 (ma.) (m, 2 H, NHCH$_2$CH$_2$N) 3.11 - 3.25 (m, 8 H, NCH$_2$CH$_2$CH$_2$CH$_3$) 2.73 - 2.78 (m, 1 H, CH$_{ibu}$) 2.22 (mi.) and 2.08 (ma.) (m, 2 H, NHCH$_2$CH$_2$N) 1.50 - 1.63 (m, 8 H, NCH$_2$CH$_2$CH$_2$CH$_3$) 1.30 (m, 8 H, NCH$_2$CH$_2$CH$_2$CH$_3$) 1.10 (d, J=6.9 Hz, 6 H, CH$_{3ibu}$) 0.93 (t, J=7.3 Hz, 12 H, NCH$_2$CH$_2$CH$_2$CH$_3$)

HRMS (ESI) calcd for $C_{35}H_{38}N_7O_6$ $[M-H]^-$ m/z=650.27 found 650.28

Analyses are consistent with the literature description of the corresponding acid[169]



2-(1'-Thyminyl)acetic acid (7)[170, 186]

To a stirred suspension of thymine (2.5 g, 19.82 mmol, 1eq) in water (40 ml) was added a solution of KOH (3.6M, 158.56 mmol, 44 mL, 8 eq). When the solution was clear, chloroacetic acid (7.49 g, 79.28 mmol, 4 eq) was added and the mixture was heated at reflux for 16 h. Then reaction was cooled to room temperature and the pH was acidified to pH 1 and allowed to stir for 1 h. The solid

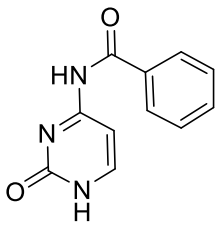
was then filtered off and placed under high vacuum for 1 h to afford **7** (1.6 g 44%, 8.7 mmol) of white solid.

^{1}H NMR (400 MHz, d_6-DMSO): δ 13.10 (broad s, 1H, COOH), 11.33 (s, 1H, NH-3), 7.49 (s, 1H, H-6), 4.36 (s, 2H, CH_2), 1.75 (s, 3H, CH_3)

^{13}C NMR (400 MHz, d_6-DMSO): δ 169.7 164.4 151.0 141.8 108.3 48.4 11.9

LRMS (ESI) calcd for $C_7H_9N_2O_4$ $[M+H]^+$ m/z=185.1 found 184.8

Analyses are consistent with the literature[180]



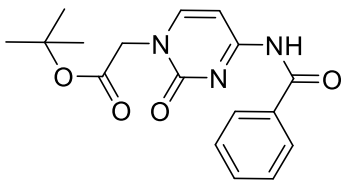
N^4-Benzoyl cytosine (12)[187]

To a stirred suspension of cytosine (5 g, 45 mmol, 1 eq) in pyridine (200 mL) was added benzoyl chloride (10.5 mL, 90 mmol, 2 eq) dropwise over 3 h. The reaction was stirred at RT for 3 h. Then 4M HCl was added to bring the solution to pH 5 and the reaction was stirred for 14 h and the remaining solid was collected by filtration and washed with EtOH and dried under high vacuum for 12 h to afford **12** as a white powder (7.15 g, 73%, 33 mmol).

^{1}H NMR (400 MHz, DMSO-d_6) δ 8.00 (m, 2 H, BzH), 7.87 (d, J=7.0 Hz, 1 H, CH-6), 7.61 (m, 1 H, BzH), 7.50 (m, 2 H, BzH), 7.18 (m, 1 H, CH-6)

LRMS (ESI) calcd for $C_{11}H_9N_3O_2$ $[M+H]^+$ m/z=216.1 found 216.0

Analyses are consistent with the literature[187]



Tert-butyl 2-(4-benzamido-2-oxopyrimidin-1(2H)-yl)acetate (14)[188]

To a suspension of N^4-benzoyl cytosine **(12)** (10 g, 40.8 mmol, 1 eq) in DMF (160 mL) at RT was added K_2CO_3 (5.63 g, 40.7 mmol, 0.9 eq) and $CsCO_3$ (1.33 g, 4.1 mmol, 0.1 eq), then *tert*-butyl bromoacetate (6.33 mL, 43 mmol, 1.05 eq) was added dropwise then mixture was stirred at RT for

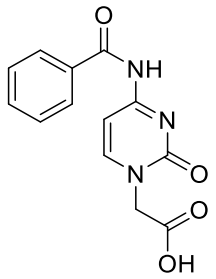
54 h. A small amount of solid remained in suspension, which was removed by filtration. The filtrate was concentrated by rotary evaporation and then recrystallized from EtOAc to afford **14** as a white solid (10.73 g, 80%, 32.6 mmol)

R_f=0.12 (1:1 Hexane/EtOAc +2% MeOH)

^{1}H NMR (400 MHz, DMSO-d_6) δ 11.22 (br s, 1 H, NHCO), 8.11 (m, 1 H, CH-6), 8.01 (d, J=7.7 Hz, 2 H, BzH), 7.57 - 7.68 (m, 1 H, BzH), 7.52 (t, J=7.7 Hz, 2 H, BzH), 7.34 (m, 1 H, CH-5), 4.55 (s, 2 H, NCH$_2$CO), 1.40 - 1.45 (m, 9 H, CH$_{3tBu}$)

^{13}C NMR (101 MHz, DMSO-d_6) δ 167.3 167.0 163.6 155.2 150.7 133.1 132.7 128.4 95.8 81.8 51.3 27.7

LRMS (ESI) calcd for $C_{17}H_{19}N_3O_4$ $[M+H]^+$ m/z=330.15 found 330.2



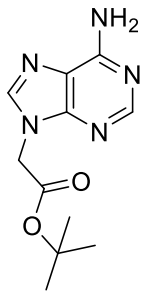
2-(4-Benzamido-2-oxopyrimidin-1-(2H)-yl)acetic acid (15)[189]

Tert-butyl 2-(4-benzamido-2-oxopyrimidin-1-(2H)-yl) acetate **(14)** (400 mg, 1.21 mmol) was dissolved in TFA (1.3 mL) and stirred at RT for 1.3 h. Then the reaction mixture was precipitated by pouring into diethyl ether (150 mL). The precipitate was filtered and dried under vacuum to afford **15** as a white solid (300 mg, 91%, 1.10 mmol).

^{1}H NMR (400 MHz, DMSO-d_6) δ 13.04 (br s, 1 H, COOH) 11.23 (br s, 1 H, NHCO) 8.12 (br d, J=7.0 Hz, 1 H, CH-6) 8.01 (br d, J=7.5 Hz, 2 H, BzH) 7.56 - 7.73 (m, 1 H, BzH) 7.47 - 7.55 (m, 2 H, BzH) 7.32 (br d, J=7.0 Hz, 1 H, CH-5) 4.58 (s, 2 H, CH$_2$COOH)

LRMS (ESI) calcd for $C_{13}H_{11}N_3O_4$ $[M+H]^+$ m/z=273.07 found 273.9

Analyses are consistent with the literature[189]



Tert-butyl-2-(6-amino-9H-purin-9-yl) acetate (19)[190]

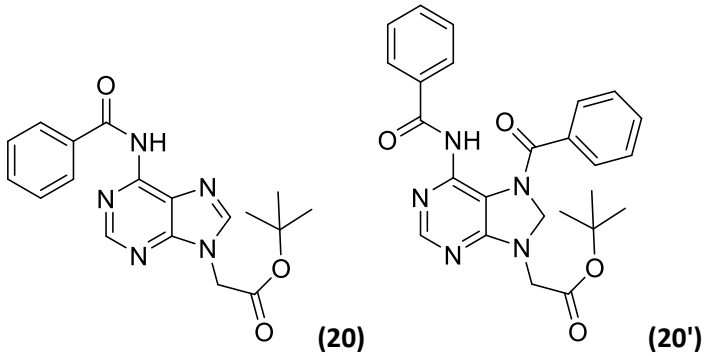
Adenine (10 g, 74 mmol, 1 eq) was suspended in DMF. Next NaH (3.6 g, 88 mmol, 1.2 eq) was added at 0 °C. The suspension was allowed to stir at room temperature for 1 h. Next *tert*-butyl bromoacetate (14.3 ml, 148 mmol, 1.2 eq) was added dropwise. How about "After 1.5 h, TLC (5% MeOH in DCM) indicated completion and the reaction was quenched with water (100 mL). The solvents were removed under vacuum to afford **(19)** as a beige solid (10.06 g, 54%, 40 mmol).

^{1}H NMR (400 MHz, DMSO-d_6): δ 8.13 (s, 1 H, C<u>H</u>-2) 8.09 (s, 1 H, C<u>H</u>-8) 7.24 (s, 2 H, N<u>H$_2$</u>) 4.94 (s, 2 H, C<u>H$_2$</u>) 1.41 (s, 9 H, C<u>H$_3$</u>)

^{13}C NMR (101 MHz, DMSO-d_6): δ 167.0 155.9 152.6 149.7 141.3 118.2 82.0 44.5 27.7

LRMS (ESI) calcd for $C_{11}H_{16}N_5O_2$ $[M+H]^+$ *m/z*=250.13 found 249.9

Analyses are consistent with the literature[190]



(20) **(20')**

Tert-butyl 2-(6-benzamido-9H-purin-9-yl) acetate (20)[191]

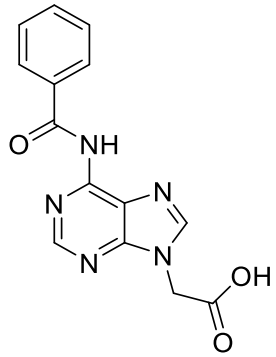
Tert-butyl 2-(6-amino-9H-purin-9-yl) acetate (790 mg, 3.17 mmol, 1 eq) was suspended in pyridine (3 mL) and dried by rotary evaporation. The residue was resuspended in pyridine (10 mL) and benzoyl chloride (0.44 mL, 3.8 mmol, 1.2 eq) was added dropwise with stirring over 2 h. The mixture was then stirred for 5 h. After that, water (50mL) was added and the product was extracted 3 times with DCM. The organic layers were washed 3 times with aqueous $NaHCO_3$, and once with brine, then dried over $MgSO_4$ and concentrated to afford a mixture of **20** and **20'**

analogue as orange foam (776 mg). The mixture was used without further purification in the next reaction.

(20) LRMS (ESI) calcd for $C_{18}H_{19}N_5O_3$+H $[M+H]^+$ *m/z*=354.16 found 354.0

(20') LRMS (ESI) calcd for $C_{25}H_{23}N_5O_4$+H $[M+H]^+$ *m/z*=458.18 found 458.1

Analysis are consistent with the corresponding methyl ester[192]



2-(6-Benzamido-9H-purin-9-yl)acetic acid (21)

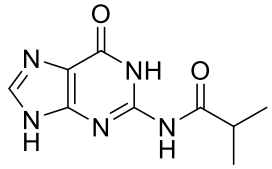
The mixture of **20** and **20'** (776 mg) was dissolved in neat TFA (2.35 mL) and stirred for 1.25 h. The mixture was then added dropwise to diethyl ether with vigorous stirring. Then precipitate was filtered and washed with diethyl ether. The solid was taken up in EtOH and then filtered over SiO_2 to afford **21** as a white powder (430 mg, 1.44 mmol, 45% over 2 steps from **19**).

^{1}H NMR (400 MHz, DMSO-d_6): δ 8.74 (s, 1 H, CH-2), 8.47 (s, 1 H, CH-8), 8.06 (d, *J*=7.7 Hz, 2 H, BzH), 7.65 (s, 1 H, BzH), 7.53 - 7.59 (m, 2 H, BzH), 5.12 (s, 2 H, CH_2)

^{13}C NMR (100 MHz, DMSO-d_6): δ 166.9 155.9 152.6 151.9 141.3 118.2 107.7 82.0 44.5 27.6

LRMS (ESI) calcd for $C_{14}H_{11}N_5O_3$+H $[M+H]^+$ *m/z*=298.09 found 298.0

Analyses are consistent with the literature[193]



***N*-(6-Oxo-6,9-dihydro-1H-purin-2-yl)isobutyramide (23)[194]**

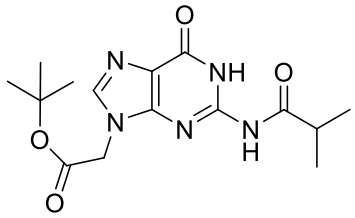
Guanine (2 g, 13.29 mmol) was suspended in anhydrous DMF (45 mL). Isobutyric anhydride (6.6 mL, 39.87 mmol, 3 eq) was then added and the suspension was stirred for 2.5 h. Then the reaction

was quenched with MeOH (10 mL) and concentrated by rotary evaporation. The product was purified by recrystallization from EtOH/H_2O (1:1) to afford compound **23** (1.9 g, 62%, 8.5 mmol).

^{1}H NMR (500 MHz, DMSO-d6) δ 13.05 (br s, 1 H, NH-7/9), 12.06 (s, 1 H, NH-1), 11.52 (s, 1 H, NHCO), 8.02 (s, 1 H, CH-8), 2.75 (sept, J=6.7 Hz, 1 H, CH_{ibu}), 1.12 (d, J=6.7 Hz, 6H, CH_3)

LRMS (ESI) calcd for $C_9H_{12}N_5O_2$ $[M+H]^+$ m/z=222.1 found 221.9

Analyses are consistent with the literature[195]



Tert-butyl 2-(2-isobutyramido-6-oxo-1,6-dihydro-9H-purin-9-yl)acetate (24)[175]

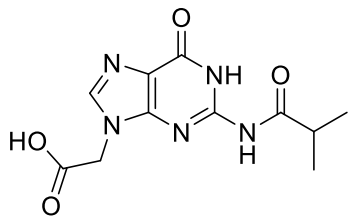
N-(6-oxo-6,9-dihydro-1H-purin-2-yl)isobutyramide **(23)** (1.9 g, 8.59 mmol, 1 eq) was suspended in dry DMF (25 mL) and then NaH (60% suspended in oil) (0.378 g, 9.45 mmol, 1.1 eq) was added in portions, then the mixture was stirred at RT for 1 h. Afterward *tert*-butyl bromoacetate (2.0 g, 10.31 mmol, 1.2 eq) was added dropwise at 0°C and stirred for 1 h. The reaction was quenched with MeOH and concentrated by rotary evaporation. The recovered solid was triturated in a small amount of hot ethyl acetate and cold filtered to recover **24** as a white solid (666 mg, 23%, 1.99 mmol).

^{1}H NMR (400 MHz, $CDCl_3$) δ 12.02 (br s, 1 H, $NHCO_{ibu}$), 8.45 (br s, 1 H, NH-1), 7.70 (s, 1 H, CH-8), 4.69 (s, 2 H, CH_2CO), 2.67 (spt, J=6.9 Hz, 1 H, CH_{ibu}), 1.42 - 1.53 (m, 9 H, CH_{3tBu}), 1.29 (d, J=6.9 Hz, 6 H, CH_{3ibu})

^{13}C NMR (101 MHz, $CDCl_3$) δ 178.1 165.8 155.5 148.5 147.5 139.3 120.8 83.7 45.2 36.6 28.0 18.9

LRMS (ESI) calcd for $C_{15}H_{21}N_5O_4$+Na $[M+Na]^+$ m/z=358.2 found 358.1

Analyses are consistent with the literature[196]



(2-Isobutyramido-6-oxo-1,6-dihydro-9H-purin-9-yl)acetic acid (25)[175]

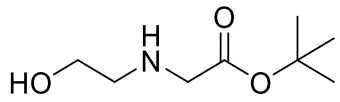
To a suspension of *tert*-butyl 2-(2-isobutyramido-6-oxo-1,6-dihydro-9H-purin-9-yl)acetate **(24)** (666 mg, 1.99 mmol, 1 eq) in DCM (4 mL), was added triethylsilane (1.6 mL, 10 mmol, 5 eq). The reaction was stirred at 0°C and TFA (5 mL) was added dropwise over 5 min. Solution was stirred at RT for 1.5 h. Reaction mixture was precipitated by pouring into stirred diethyl ether (200 mL), filtered, and rinsed with diethyl ether to afford **25** as a white powder (356 mg, 64%, 1.2 mmol).

^{1}H NMR (400 MHz, DMSO-d_6) δ 13.35 (brs, 1 H, COOH), 12.08 (s, 1 H, NHCO$_{ibu}$), 11.67 (s, 1 H, NH-1), 7.95 (s, 1 H, NH-8), 4.90 (s, 2 H, CH$_2$CO), 2.76 (spt, *J*=6.9 Hz, 1 H, CH$_{ibu}$), 1.11 (d, *J*=6.9 Hz, 6 H, CH$_{3ibu}$)

^{13}C NMR (101 MHz, DMSO-d_6) δ 180.2 166.6 154.8 148.9 148.1 140.3 119.6 82.3 44.8 34.6 27.7 18.9

LRMS (ESI) calcd for $C_{11}H_{13}N_5O_4$ $[M+H]^+$ *m/z*=281.1 found 280.0

Analyses are consistent with the literature[175]



Tert-butyl (2-hydroxyethyl)glycinate (27)

In a 50-mL round-bottomed flask, ethanolamine (16.4 mmol, 0.91 mL, 3 eq) was dissolved in dry DMF (20 mL) and DIPEA (0.88 mL, 5 mmol, 1 eq) was added. *Tert*-butyl bromoacetate (0.76 mL, 5 mmol, 1 eq) was added dropwise at RT with stirring. The reaction was stirred at RT overnight then, concentrated by rotary evaporation to afford yellow oil. The residue was dissolved in ethyl acetate and washed with brine, and then it was back-extracted 3 times with ethyl acetate. The combined organic layers were dried over $MgSO_4$, concentrated by rotary evaporation, and the residue purified by column chromatography (eluent DCM/MeOH 95:5) to afford compound **27** as a yellow oil (0.68 g, 78%, 3.8 mmol).

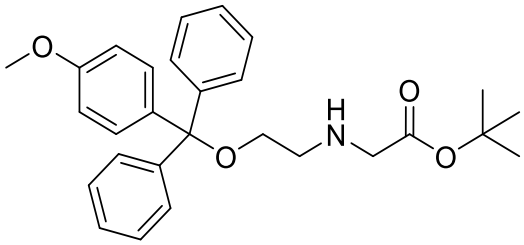
R_f = 0.14 (95:5 DCM/MeOH)

^{1}H NMR (500 MHz, $CDCl_3$) δ 3.61 (m, 2 H, HOC<u>H$_2$</u>), 3.31 (s, 2 H, C<u>H$_2$</u>COO), 2.78 (m, 2 H, CH$_2$C<u>H$_2$</u>NH), 1.46 (s, 9 H, C<u>H$_3$</u>)

^{13}C NMR (126 MHz, $CDCl_3$) δ 172.1 81.4 60.9 51.2 51.0 28.1

LRMS (ESI) calcd for $C_8H_{17}NO_3$+H $[M+H]^+$ *m/z*=176.13 found 176.1

Analyses are consistent with literature[169]



Tert-butyl (2-((4-methoxyphenyl)diphenylmethoxy)ethyl)glycinate (28)[185]

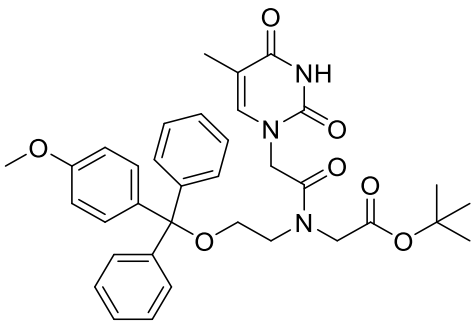
To a stirred solution of *tert*-butyl (2-hydroxyethyl)glycinate **(27)** (1 g, 5.67 mmol, 1 eq) in pyridine (12 mL) was added monomethoxytrityl chloride (2.1 g, 6.81 mmol, 1.2 eq) with stirring, in portions, at RT. After 6.5 h, MeOH (2 mL) was added, the reaction was concentrated by rotary evaporation, and the residue was dissolved in MeOH and precipitated in hexane. The precipitate was filtered and purified by column chromatography on SiO2 eluted by DCM to afford **28** as colourless oil (1 g, 39%, 2.23 mmol).

^{1}H NMR (400 MHz, DMSO-d_6) δ 7.19 - 7.41 (m, 12 H, MMT<u>H</u>), 6.90 (d, *J*=7.7 Hz, 2 H, MMT<u>H</u>) 3.74 (s, 3 H, MMTOC<u>H$_3$</u>) 3.20 (s, 2 H, C<u>H$_2$</u>COO) 3.00 (t, *J*=5.6 Hz, 2 H, C<u>H$_2$</u>COO) 2.70 (t, *J*=5.6 Hz, 2 H, C<u>H$_2$</u>NH) 1.42 (s, 9 H, C<u>H$_3$</u>)

^{13}C NMR (101 MHz, DMSO-d_6) δ 172.0 158.6 145.0 135.8 130.4 128.4 128.3 127.3 113.6 86.0 80.5 63.5 55.5 51.6 48.7 28.3

LRMS (ESI) calcd for $C_{28}H_{33}NO_4$+Na $[M+Na]^+$ *m/z*=470.23 found 470.2

Analyses are consistent with the reported corresponding DMT protected methyl ester [185]

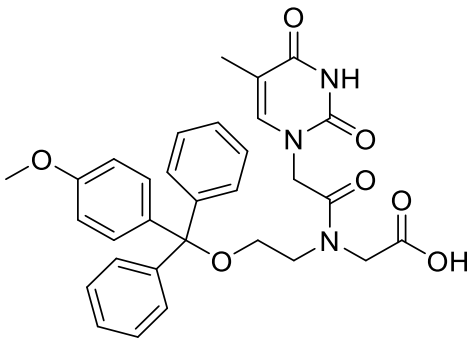


Tert-butyl *N*-(2-((4-methoxyphenyl)diphenylmethoxy)ethyl)-*N*-(2-(5-methyl-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)acetyl)glycinate (29)

To a stirred solution of thymine acetic acid **(7)** (548 mg, 2.90 mmol, 1.3 eq) in DMF (10 mL) was added EDC (598 mg, 3.12mmol, 1.4 eq) and HOBt (451.7 mg, 3.34 mmol, 1.5 eq) and the solution was stirred for 30 min at RT. Then DIPEA (0.777 mL, 4.46 mmmol, 2 eq) and *tert*-butyl (2-((4-methoxyphenyl)diphenylmethoxy)ethyl) glycinate **(28)** (1 g, 2.23 mmol, 1 eq) were added and stirred at RT for 6 h. The reaction was concentrated by rotary evaporation, taken up in Ethyl acetate and washed with saturated aqueous $NaHCO_3$ (25mL) and brine (25mL). The organic phase was evaporated under reduced pressure to afford **29** as a yellow solid (940 mg, 69%, 1.53 mmol).

LRMS (ESI) calcd for $C_{35}H_{39}N_3O_7$+Na $[M+Na]^+$ *m/z*=636.27 found 636.3

Analyses are consistent with the reported corresponding DMT protected methyl ester[185]



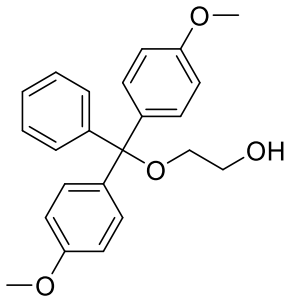
***N*-(2-((4-methoxyphenyl)diphenylmethoxy)ethyl)-*N*-(2-(5-methyl-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)acetyl)glycine (30)**

Tert-butyl *N*-(2-((4-methoxyphenyl)diphenylmethoxy)ethyl)-*N*-(2-(5-methyl-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)acetyl)glycinate **(29)** (200 mg, 0.33 mmol, 1 eq) was dissolved in MeOH (30mL) and a solution of 2M $NaOH_{(aq)}$ was added until the solution reach pH12. The mixture was stirred at RT for 4.5 h until TLC indicated reaction completion and it was then neutralised by Dowex® pyidinium. The resin was removed by filtration and the filtrate was concentrated by rotary evaporation. The solid was then purified by column chromatography (eluted with

EtOAc:MeOH:H_2O (increasing from 9:1 to 8:1:1, then 7:2:1) to afford compound **30** as a yellow foam (156 mg, 85%, 0.28 mmol).

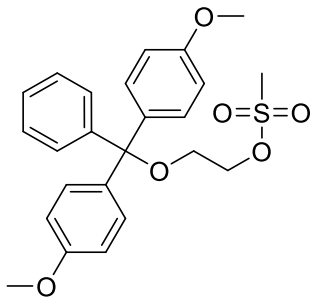
LRMS (ESI) calcd for $C_{31}H_{31}N_3O_7$+Na $[M+Na]^+$ *m/z*=580.21 found 580.3

See **43** for full characterisation of the corresponding DMT-protected tetrabutylammonium salt.



Dimethoxytritylethan-1-ol (39)[179]

In a dry round bottom flask under argon atmosphere, a solution of DMT-Cl (15 g, 44 mmol, 1 eq) in DCM (100 mL) and pyridine (3.5 mL) was added dropwise to a solution of ethylene glycol (50mL, 880mmol, 20 eq) in DCM (100 mL) and pyridine (3.5 mL). The reaction was stirred for 6 hrs at RT. The reaction mixture was then concentrated by rotary evaporation, and the residue was taken up in EtOAc (150 mL) and washed with saturated $NaHCO_{3\,(aq)}$ (100 mL), water (2 x 50 mL) and brine (100 mL). The organic phase was dried over Na_2SO_4 and concentrated by rotary evaporation. The crude product (17 g, quantitative yield) was used directly for the next reaction.

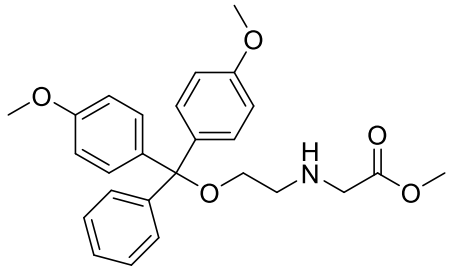


Dimethoxytrityl-ethyl methanesulfonate (40)[179]

In a dry round bottom flask under argon atmosphere, a solution of 2-(bis(4-methoxyphenyl)(phenyl)methoxy)ethan-1-ol **(39)** (17 g, 44 mmol, 1 eq) was solubilized in DCM (100 mL) and NEt_3 (12.3 mL, 88 mmol, 2 eq). The solution was stirred at -5˚C and methanesulfonyl chloride (3.4mL, 44 mmol, 1 eq) was added dropwise. The reaction mixture was then stirred at RT for 2.5 h and quenched with MeOH. The reaction was concentrated by rotary evaporation, and the resulting yellow oil was purified was purified by automated column chromatography using a

SNAP-Ultra 100g silica column. Concentration by rotary evaporation afforded **40** as a yellow oil (9.1 g, 45% from **39**, 20 mmol).

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.18-7.46 (m, 9 H, DMTH) 6.74 - 6.91 (m, 4 H, DMTH) 4.30 - 4.38 (m, 2 H, CH_2OS) 3.79 (s, 6 H, DMTOCH_3) 3.36 - 3.42 (m, 2 H, DMTOCH_2CH_2) 3.05 (s, 3 H, SO_2CH_3)

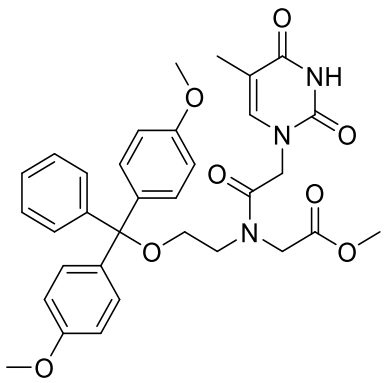


Dimethoxytrityl-ethyl glycinate (41)[179]

In a dry 250-mL round bottom flask under argon atmosphere, to a solution of 2-(bis(4-methoxyphenyl)(phenyl)methoxy)ethyl methanesulfonate **(40)** (9.1 g, 20.6 mmol, 1 eq) in of NEt_3 (80 mL), was added a solution of glycine methyl ester hydrochloride (3.87g, 30.8 mmol, 1.5eq) in 15 mL of DMF. The reaction mixture was heated to the boiling point of NEt_3 (105˚C) and refluxed for 5 h. The reaction was then concentrated by rotary evaporation and the residue taken up in DCM. The precipitate was then filtered out and the filtrate was concentrated and purified by automated chromatography using a SNAP-Ultra 50g silica column. Concentration of the appropriate fractions afforded **41** as a yellow oil (2.61 g, 30%, 6.0 mmol).

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.11 - 7.60 (m, 9 H, DMTH) 6.65 - 6.93 (m, 4 H, DMTH) 3.78 (s, 6 H, DMTOCH_3) 3.74 (s, 3 H, COOCH_3) 3.43 (s, 2 H, NCH_2COO) 3.21 (t, *J*=5.4 Hz, 2 H, OCH_2CH_2N) 2.82 (t, *J*=5.4 Hz, 2 H, OCH_2CH_2N)

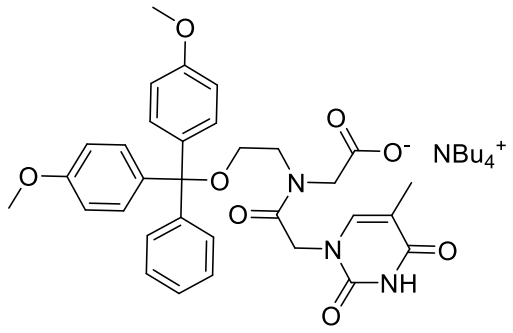
Analyses are consistent with the reported corresponding allyl ester[169]



Methyl N-(2-(bis(4-methoxyphenyl)(phenyl)methoxy)ethyl)-N-(2-(5-methyl-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)acetyl)glycinate (42)[185]

In a dry 250-mL round-bottom flask under argon atmosphere, dimethoxytrityl-ethyl glycinate **(41)** (2.61 g, 5.9 mmol, 1 eq) was dissolved in DMF (50 mL). Thymine acetic acid (1.63 g, 8.85 mmol, 1.5 eq) and DIPEA (2.3 mL, 12.98mmol, 2.2 eq) were then added to the solution. T3P (5.63 g, 8.85 mmol, 1.5 eq, as a 50% solution in DMF) was then added and the reaction mixture was stirred at RT for 16 h. The reaction mixture was concentrated by rotary evaporation and taken up in of ethyl acetate (100 mL). The solution was washed twice with 10% $NaHCO_{3aq}$ and once with brine, dried over Na_2SO_4, and concentrated by rotary evaporation. The crude product was then purified by automated column chromatography using a SNAP-Ultra 25g silica column, to afford **42** as a beige solid (2.32 g, 64%, 3.86 mmol).

^{1}H NMR (400 MHz, DMSO-*d6*) δ 11.31 (s, 1H, N<u>H</u>-3) 7.09 - 7.45 (m, 9 H, DMT<u>H</u>) 7.02 (m, 1 H, C<u>H</u>-6) 6.82 - 6.93 (m, 4 H, DMT<u>H</u>) 4.66 (ma.) and 4.51 (mi.) (s, 2 H, NC<u>H$_2$</u>CON) 4.44 (mi.) and 3.97 (ma.) (s, 2 H, NC<u>H$_2$</u>COOCH$_3$) 3.73 (s, 3 H, DMTOC<u>H$_3$</u>) 3.64 (mi.) and 3.57 (ma.) (m, 2 H, COOC<u>H$_3$</u>) 3.03 - 3.29 (m, 2 H, OC<u>H$_2$</u>CH$_2$N) 3.27 (ma.) and 3.05 (mi.) (m, 2 H, OCH$_2$C<u>H$_2$</u>N) 1.74 (mi.) and 1.69 (ma.) (s, 3 H, C<u>H$_3$</u>-5)



_N_-(2-(Bis(4-methoxyphenyl)(phenyl)methoxy)ethyl)-N-(2-(5-methyl-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)acetyl)glycine (43)

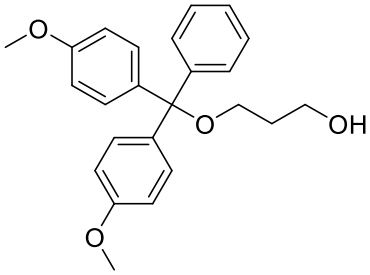
In a round-bottom flask, methyl *N*-(2-(bis(4-methoxyphenyl)(phenyl)methoxy)ethyl)-*N*-(2-(5-methyl-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)acetyl)glycinate (2.32g, 3.8 mmol, 1 eq) was dissolved in dioxane (50 mL) and 1M tetrabutylammonium aqueous solution was added until the reaction was pH=12. The reaction was stirred at RT for 2.5 h, then concentrated by rotary evaporation and purified by automated reverse-phase column chromatography (SNAP-Ultra C18, 30g), eluted with gradients of acetonitrile in water, to afford **43** as white foam (2.25 g, 72%, 2.71 mmol).

^{1}H NMR (500 MHz, DMSO-d_6) δ 11.25 (s, 1 H, NH-3), 7.15 - 7.38 (m, 10 H, DMTH) 6.84 - 6.93 (m, 4 H, DMTH), 4.59 (mi.) and 4.47 (ma.) (s, 2 H, NCH_2CON), 3.73 (s, 6 H, $DMTOCH_3$)3.71 (ma.) and 3.67 (mi.) (s, 2 H, NCH_2COO) 3.49 (mi.) and 3.32 (ma.) (m, 2 H, $NHCH_2CH_2N$), 3.11 - 3.20 (m, 8 H, $NCH_2CH_2CH_2CH_3$), 3.05 - 2.98 (m, 2 H, $NHCH_2CH_2N$), 1.74 (ma.) and 1.68 (mi.) (s, 3 H, CH_3-5), 1.61 - 1.51 (m, 8 H, $NCH_2CH_2CH_2CH_3$), 1.36 - 1.25 (m, 8H, $NCH_2CH_2CH_2CH_3$) 0.93 (t, *J*=7.32 Hz, 12 H, $NCH_2CH_2CH_2CH_3$)

^{13}C NMR (126 MHz, DMSO-d_6) δ 169.5 167.9 164.9 158.5 151.5 145.5 142.7 136.2 130.1 128.3 128.1 127.0 113.7 108.3 86.5 85.9 61.9 58.0 55.5 53.9 48.1 23.5 19.7 14.0 12.4

LRMS (ESI) calcd for $C_{32}H_{32}N_3O_8$ $[M-H]^-$ *m/z*=586.22 found 586.22

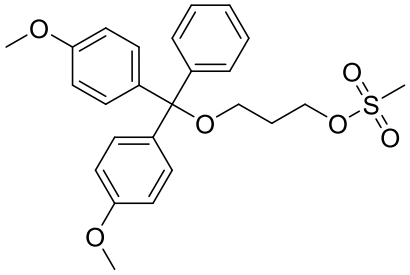
Analyses are consistent with literature[169]



3-(bis(4-methoxyphenyl)(phenyl)methoxy)propan-1-ol (44)[197]

In a dry argon atmosphere, in a round-bottom flask at RT, 1,3-propanediol (30 mL, 415mmol, 14 eq) was dissolved in DCM (100 mL). To this solution was added dropwise a solution of DMT-Cl (10g, 29.5 mmol, 1 eq) in DCM (100 mL) and pyridine (10 mL). The reaction was stirred at RT for 16 h, and then concentrated by rotary evaporation. The resulting yellow oil was filtered through a short pad of silica with DCM/MeOH 9:1 to afford a mixture of the mono and di-protected products (9.9 g total), which was used without further purification in the next reaction.

R_f=0.47 (2:1:1 Pentane/EtOAc/DCM)

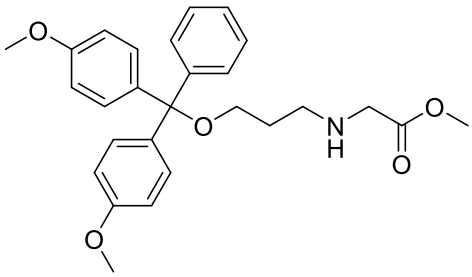


3-(bis(4-methoxyphenyl)(phenyl)methoxy)propyl methanesulfonate (45) [179]

In a dry and under argon atmosphere in a round-bottom flask, **44** was dissolved in DCM (100 mL) and NEt_3 (7.72mL, 52.4mmol, 2.1 eq) and stirred at 0°C. Methanesulfonyl chloride (2.03 mL, 26.2 mmol, 1 eq) was added at 0°C and then the reaction was stirred at RT for 3 h. The reaction was then washed with 10% aqueous $NaHCO_3$ (100 mL), water (2 x 100 mL) and brine (100 mL). The organic phase was dried over Na_2SO_4 and concentrated by rotary evaporation. The resulting oil was purified by automated column chromatography (50 g SNAP-Ultra column with a gradient of pentane:EtOAc from 95:5 to 20:80) to afford **45** as a yellow oil (9.92 g, 73% from **44**, 21.7 mmol).

R_f=0.35 (DCM)

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.14 - 7.34 (m, 9 H, DMTH), 6.79 - 6.85 (m, 4 H, DMTH), 4.40 (t, J=6.45 Hz, 2 H, $CH_2OSO_2CH_3$), 3.79 (s, 6 H, DMTOCH_3), 3.21 (t, J=5.86 Hz, 2 H, DMTOCH_2CH_2), 2.91 (s, 3 H, SO_2CH_3), 1.94 - 2.03 (m, 2 H, $CH_2CH_2CH_2$)

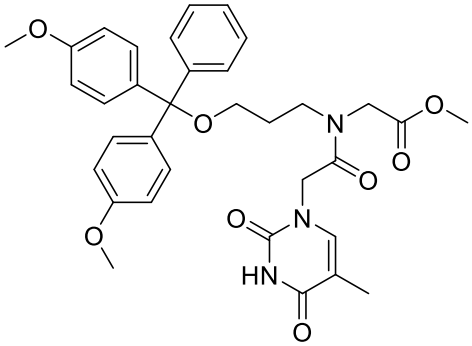


Methyl (3-(bis(4-methoxyphenyl)(phenyl)methoxy)propyl)glycinate (46) [179]

In a dry round bottom flask under argon atmosphere, 3-(bis(4-methoxyphenyl)(phenyl)methoxy)propyl methanesulfonate **45** (9.92 g, 21 mmol, 1 eq) was solubilized in 100 mL of NEt_3 and 10 mL of DMF. Glycine methyl ester hydrochloride (7.9 g, 63 mmol, 3 eq) was added and the solution was heated at reflux of NEt_3 for 16 h. The reaction mixture was then concentrated by rotary evaporation and taken up in chloroform. The precipitate was filtered out and the filtrate was concentrated by rotary evaporation. The recovered oil was purified by automated column chromatography using a 50 g SNAP-Ultra silica column with a gradient of chloroform/acetone (100:0 to 40:60) to afford **46** as yellow oil (1.01 g, 9%, 2 mmol).

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.17 - 7.49 (m, 9 H, DMTH) 6.77 - 6.87 (m, 4 H, DMTH) 3.79 (s, 6 H, DMTOCH_3) 3.71 (s, 3 H, COOCH_3) 3.40 (s, 2 H, NCH_2COO) 3.13 (t, *J*=6.06 Hz, 2 H, OCH_2CH_2) 2.72 (t, *J*=7.03 Hz, 2 H, CH_2CH_2N) 1.73 - 1.87 (m, 2 H, $CH_2CH_2CH_2$)

^{13}C NMR (101 MHz, $CDCl_3$) δ 172.8, 158.3, 145.2, 136.4, 126.6, 113.0, 85.5, 61.6, 55.2, 51.7, 50.9, 130.0, 128.1, 127.7, 47.2, 30.4



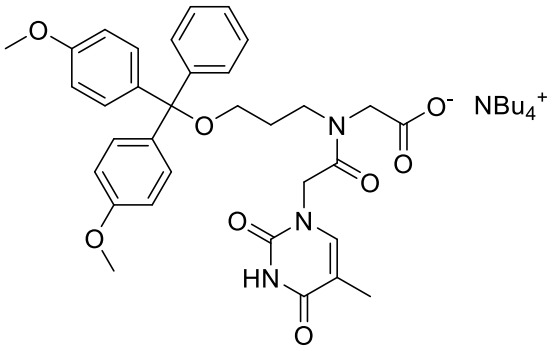
Methyl *N*-(3-(bis(4-methoxyphenyl)(phenyl)methoxy)propyl)-*N*-(2-(5-methyl-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)acetyl)glycinate (47)

In a dry 100-mL round bottom flask under argon, compound **46** (1.07 g, 2.3 mmol, 1 eq) was solubilized in 50 mL of DMF. T3P 50% in DMF (2.1 mL, 3.5 mmol, 1.5 eq) and DIPEA (0.75 mL, 4.3 mmol, 2.2 eq) were then added to the solution. Thymine acetic acid **7** (645 mg, 3.5 mmol, 1.5 eq) was added after 10min and the reaction mixture was stirred at RT for 16 h. The reaction mixture

was concentrated by rotary evaporation and taken up in chloroform (100 mL). The solution was washed twice with 10% aqueous $NaHCO_3$ and once with brine. The organic phase was dried over Na_2SO_4 and concentrated by rotary evaporation. The crude product was purified by automated column chromatography with a SNAP-Ultra 25g silica column using a gradient of Chloroform / methanol, to afford **47** as a beige solid (1.06 g, 74%, 1.7 mmol).

R_f=0.55 (95:5 $CHCl_3$/MeOH)

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.14 - 7.46 (m, 10 H, DMTH and H-6), 6.80 - 6.88 (m, 4 H, DMT-H), 4.55 (ma.) and 4.37 (mi.) (s, 2 H, NCH_2CON), 4.09 (mi.) and 3.97 (ma.) (s, 2 H, $NCOCH_2N$), 3.81 (m, 6 H, $DMTOCH_3$), 3.82 (mi.) and 3.71 (ma.) (s, 3 H, $COOCH_3$), 3.50 - 3.59 (m, 2 H, OCH_2CH_2), 3.22 (ma.) and 3.11 (mi.) (m, 2 H, CH_2CH_2N), 1.76 - 1.94 (m, 5 H, CH_3-5 and $CH_2CH_2CH_2$)



_N_-(3-(Bis(4-methoxyphenyl)(phenyl)methoxy)propyl)-_N_-(2-(5-methyl-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)acetyl)glycinate (48)

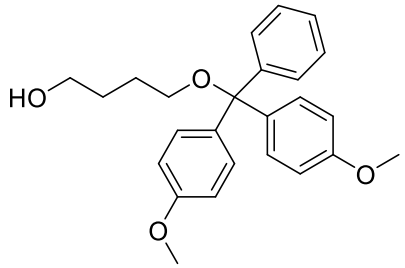
In a 50-mL round-bottom flask, methyl *N*-(3-(bis(4-methoxyphenyl)(phenyl)methoxy)propyl)-N-(2-(5-methyl-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)acetyl)glycinate **47** (1.62 g, 2.6 mmol, 1 eq) was dissolved in dioxane (10 mL) and a solution of 0.5M tetrabutylammonium hydroxide 30% hydrate in dioxane was added until the reaction was pH=12. The reaction was stirred at RT for 16h, then concentrated by rotary evaporation and purified by automated reverse phase chromatography (SNAP-Ultra C18, 30g column), eluted with a gradient of acetonitrile in water. The product **(48)** was isolated as white foam (1.7 g, 77%, 2.0 mmol).

^{1}H NMR (500 MHz, DMSO-d_6) δ 11.22 (br s, 1 H, NH-3), 7.15 - 7.40 (m, 10 H, DMTH, H-6), 7.22 (m, 4 H, DMTH), 4.54 (mi.) 4.40 (ma.) (s, 2 H, $NCOCH_2N$), 3.70 - 3.75 (m, 6 H, $DMTOCH_3$), 3.48 (mi) 3.47 (ma.) (s, 2 H, NCH_2COO), 3.25 - 3.30 (m, 2 H, CH_2CH_2N), 3.12 - 3.22 (m, 8 H, $NCH_2CH_2CH_2CH_3$), 3.01 (mi.) 2.92 (ma.) (m, 2 H, OCH_2CH_2), 1.73 - 1.80 (m, 3 H, CH_3-5). 1.64 - 1.71 (m, 2 H, $OCH_2CH_2CH_2N$), 1.48 - 1.62 (m, 8 H, $NCH_2CH_2CH_2CH_3$), 1.36 - 1.25 (m, 8 H, $NCH_2CH_2CH_2CH_3$), 0.93 (t, *J*=7.32 Hz, 12 H, $NCH_2CH_2CH_2CH_3$)

^{13}C NMR (126 MHz, DMSO-d_6) δ 12.4, 14.0, 19.7, 23.5, 28.1, 45.7, 48.2, 52.9, 55.5, 58.0, 61.45, 85.7, 113.6, 127.0, 128.3, 130.0, 136.4, 142.8, 145.6, 158.4

LRMS (ESI) calcd for $C_{33}H_{34}N_3O_8$ $[M-H]^-$ *m/z*=600.24 found 600.24

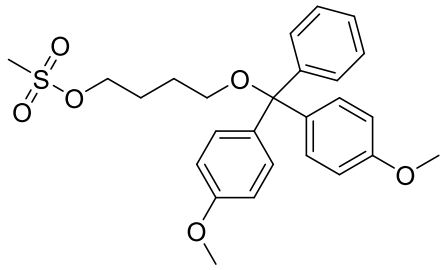
Analyses are consistent with the literature[160]



4-(Bis(4-methoxyphenyl)(phenyl)methoxy)butan-1-ol (49)[198]

In a 50-mL round-bottom flask under a dry argon atmosphere, was added and stirred at 1,4-butanediol (60 mL, 677 mmol, 30 eq). To the reaction mixture was added dropwise a solution of DMT-Cl (10g, 29.5 mmol, 1 eq) in 40mL of DCM and 5 mL of pyridine. Then the reaction was stirred at RT for 16 h and concentrated by rotary evaporation. 12.17 g of a mixture of mono and di-protected product was afforded and was used without further purification in the next reaction.

R_f=50% (90:10 DCM: EtOAc)



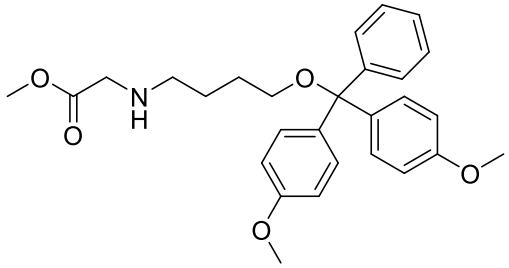
4-(Bis(4-methoxyphenyl)(phenyl)methoxy)butyl methanesulfonate (50)

In a round-bottom flask under a dry argon atmosphere, 3-(bis(4-methoxyphenyl)(phenyl)methoxy)butan-1-ol **(49)** (12.17 g, 31 mmol, 1 eq) was dissolved in of DCM (150 mL) and NEt_3 (8.7 mL, 62 mmol, 2 eq) and stirred at 0°C. Methanesulfonyl chloride was added at 0°C and then the reaction was allowed to come to RT and stirred for 4.5 h. The reaction was concentrated by rotary evaporation and the residue was taken up in ethyl acetate and washed with 10% $NaHCO_{3\,(aq)}$ (100 mL), water (100 mL) and brine (100 mL). The organic phase was dried over Na_2SO_4 and concentrated by rotary evaporation. The recovered oil was purified by automated column chromatography using a 50 g SNAP-Ultra silica column with a gradient of pentane:EtOAc (98:2 to 20:80) to afford **50** as yellow oil (3.04 g, 23%, 7 mmol).

R_f=0.47 in DCM

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.15 - 7.48 (m, 9 H, DMTH), 6.75 - 6.90 (m, 4 H, DMTH), 4.21 (t, J=6.5 Hz, 2 H, $SO_2OCH_2CH_2$), 3.79 (s, 6 H), 3.10 (t, J=6.3 Hz, 2 H, CH_2ODMT), 2.96 (s, 3 H, CH_2SO_2), 1.81 - 1.92 (m, 2 H, $CH_2CH_2SO_2$), 1.65 - 1.77 (m, 2 H, CH_2CH_2ODMT)

^{13}C NMR (101 MHz, $CDCl_3$) δ 158.4, 145.1, 136.3, 130.0, 128.1, 127.8, 126.7, 113.0, 85.8, 70.0, 62.3, 55.2, 37.3, 26.3, 26.0

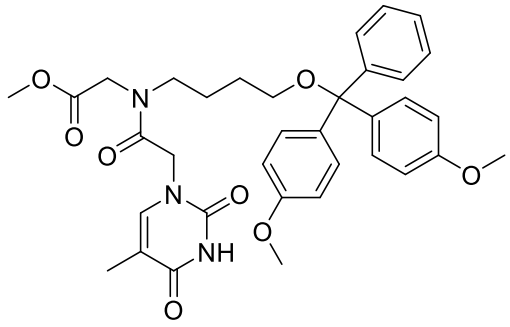


Methyl (4-(bis(4-methoxyphenyl)(phenyl)methoxy)butyl)glycinate (51)

In a dry round-bottom flask under argon atmosphere, 4-(bis(4-methoxyphenyl)(phenyl)methoxy)butyl methanesulfonate **(50)** (6.52 g, 14.3 mmol, 1 eq) was dissolved in NEt_3 (100 mL) and DMF (10 mL). Glycine methyl ester hydrochloride (5.37 g, 42.8 mmol, 3 eq) was added and the solution was heated at reflux for 5 h. The reaction mixture was then concentrated by rotary evaporation and the residue was taken up in chloroform and washed with 10% $NaHCO_{3aq}$ (100 mL), water (100 mL) and brine. The organic phase was dried over Na_2SO_4 and concentrated by rotary evaporation. The recovered oil was purified by automated column chromatography using a 25 g SNAP-Ultra silica column with a gradient of chloroform:acetone (100:0 to 40:60) to afford **51** as a yellow oil (1.67 g, 25%, 3.6 mmol).

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.10 - 7.53 (m, 9 H, DMTH), 6.75 - 6.86 (m, 4 H, DMTH), 3.78 (s, 6 H, $DMTOCH_3$), 3.72 (s, 3 H, $COOCH_3$), 3.40 (s, 2 H, CH_2COO), 3.05 (t, J=6.1 Hz, 2 H, CH_2ODMT), 2.58 (t, J=7.0 Hz, 2 H, CH_2NH), 1.53 - 1.72 (m, 4 H, $NCH_2CH_2CH_2CH_2O$)

^{13}C NMR (101 MHz, $CDCl_3$) δ 172.9, 158.3, 145.3, 136.6, 123.0, 128.2, 127.7, 126.6, 113.0, 85.7, 63.1, 55.2, 51.8, 50.8, 49.6, 27.8, 27.0

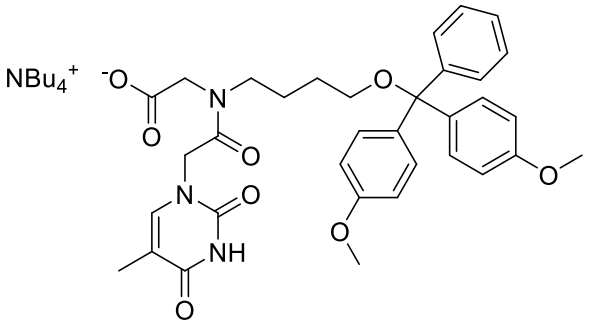


Methyl *N*-(4-(bis(4-methoxyphenyl)(phenyl)methoxy)butyl)-*N*-(2-(5-methyl-2,6-dioxo-1,2,3,6-tetrahydropyridin-3-yl)acetyl)glycinate (52)

In a dry 250-mL round-bottom flask under an argon atmosphere, methyl (4-(bis(4-methoxyphenyl)(phenyl)methoxy)butyl)glycinate **(51)** (1.67 g, 3.6 mmol, 1 eq) was dissolved in DMF (50 mL). Thymine acetic acid (995 mg, 5.4 mmol, 1.5 eq) and DIPEA (0.64 mL, 7.6 mmol, 2.2 eq) were then added to the solution. T3P (3.44 g, 5.4 mmol, 1.5 eq; 50% in DMF) was then added and the reaction mixture was stirred at RT for 16 h. The reaction mixture was concentrated by rotary evaporation and the residue was taken up in ethyl acetate (100 mL) and washed with 10% $NaHCO_{3\ (aq)}$ (2 x 100 mL) once brine (100 mL). The organic phase was dried over Na_2SO_4 and concentrated by rotary evaporation to afford the crude product **(52)** was recovered as beige solid and use without further purification (2.1 g, 92%, 3.3 mmol).

R_f=0.22 (1:1 DCM/EtOAc)

^{1}H NMR (400 MHz, DMSO-d_6) δ 11.31 (br s, 1 H, NH-3) 7.16 - 7.48 (m, 10 H, DMTH and CH-6) 6.79 - 6.94 (m, 4 H, DMTH) 4.57 (ma.) and 4.46 (mi.) (s, 2 H, NCH_2CON) 4.30 (mi.) and 4.03 (ma.) (m, 2 H, NCH_2COOCH_3) 3.72 (m, 6 H, $DMTOCH_3$) 3.69 (mi.) and 3.60 (ma.) (s, 3 H, $COOCH_3$) 3.21 - 3.39 (m, 2 H, NCH_2CH_2) 3.00 - 2.98 (s, 2 H, OCH_2CH_2) 1.74 (s, 3 H, CH_3-5) 1.38 - 1.70 (m, 4 H, $NCH_2CH_2CH_2CH_2O$)



_N_-(4-(Bis(4-methoxyphenyl)(phenyl)methoxy)butyl)-_N_-(2-(5-methyl-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)acetyl)glycine (53)[160]

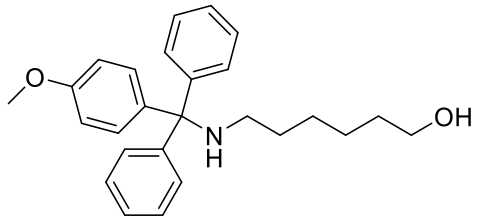
In a 50-mL round-bottom flask, methyl *N*-(4-(bis(4-methoxyphenyl)(phenyl)methoxy)butyl)-N-(2-(5-methyl-2,6-dioxo-1,2,3,6-tetrahydropyridin-3-yl)acetyl)glycinate **(52)** (2.1 g, 3.3 mmol, 1 eq) was dissolved in dioxane (10 mL) and a solution of tetrabutylammonium hydroxide (0.5M, 30% hydrate in dioxane) was added until the reaction was pH=12. The reaction was stirred at RT for 3h, then concentrated by rotary evaporation the residue was purified by automated reverse phase chromatography (30 g SNAP-Ultra C18 column, 0 to 100% acetonitrile in water), to afford the expected product **53** as a white foam (1.12 g, 40%, 1.3 mmol).

^{1}H NMR (400 MHz, DMSO-d_6) δ 11.20 (s, 1 H, NH-3) 7.15 - 7.40 (m, 10 H, DMTH and CH-6) 6.85 - 6.91 (m, 4 H, DMTH) 4.47 (mi.) 4.41 (ma.) (s, 2 H, NCH_2CO) 3.73 (m, 6 H, $DMTOCH_3$) 3.63 (mi.) and 3.53 (ma.) (s, 2 H, NCH_2COO) 3.11 - 3.27 (m, 10 H, $NCH_2CH_2CH_2CH_2O$ and $NCH_2CH_2CH_2CH_3$) 2.89 - 2.98 (m, 2 H, OCH_2CH_2) 1.70 - 1.77 (m, 3 H, CH_3-5) 1.52 - 1.61 (m, 8 H, $NCH_2CH_2CH_2CH_3$) 1.39 - 1.50 (m, 4 H, $NCH_2CH_2CH_2CH_2O$) 1.24 - 1.36 (m, 8 H, $NCH_2CH_2CH_2CH_3$) 0.87 - 0.96 (m, 12 H, $NCH_2CH_2CH_2CH_3$)

^{13}C NMR (126 MHz, DMSO-d_6) δ 169.7 164.9 167.5 158.4 151.5 145.7 142.9 136.5 130.0 128.3 112.4 128.1 127.0 113.6 108.2 85.6 63.2 58.0 55.5 27.5 24.4 23.5 19.7 14.0

HRMS (ESI) calcd for $C_{34}H_{36}N_3O_8$ $[M-H]^-$ *m/z*=614.25 found 614.25

Analyses are consistent with the literature[160]



6-(((4-Methoxyphenyl)diphenylmethyl)amino)hexan-1-ol (55)

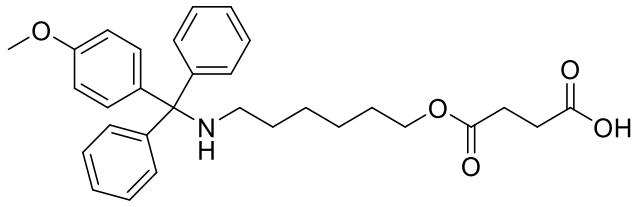
6-amino-hexan-1-ol (1.00 g, 8.55 mmol, 1.05 eq) was dissolved in pyridine (7 mL) and NEt_3 (0.2 mL) and stirred at RT. A solution of MMT-Cl (2.50 g, 8.12 mmol, 1 eq) in pyridine (9 mL) was added dropwise. The solution was stirred at RT for 20 h and concentrated by rotary evaporation. The brown oil recovered was then solubilised in EtOAc and saturated aqueous $NaHCO_3$ solution. Then it was extracted with EtOAc (2x 15mL) and washed with water (1x 15mL), and brine (1x 15mL). Organic phase was dried over $MgSO_4$ and concentrated by rotary evaporation. Purification was achieved by column chromatography on SiO_2 using PetEt/EtOAc (4:2 to 1:1) as eluent. Concentration of the eluent yielded compound **54** as a colourless oil (1.53 g, 48%, 3.9 mmol).

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.08 - 7.64 (m, 12 H, MMTH), 6.72 - 6.92 (m, 2 H, MMTH), 3.79 (s, 3 H, MMTOCH_3), 3.62 (t, *J*=6.60 Hz, 2 H, OCH_2), 2.13 (t, *J*=7.03 Hz, 2 H, NCH_2), 1.3 - 1.65 (m, 8 H, $CH_2CH_2CH_2$)

^{13}C NMR (101 MHz, $CDCl_3$) δ 157.7 146.6 138.5 129.8 128.5 127.7 126.1 113.0 70.3 63.0 55.2 43.5 32.7 30.8 27.1 25.7

LRMS (ESI) calcd for $C_{26}H_{31}NO_2$+Na $[M+Na]^+$ *m/z*=412.23 found 412.1

Analysis are consistent with literature[164]



4-((6-(((4-Methoxyphenyl)diphenylmethyl)amino)hexyl)oxy)-4-oxobutanoic acid (56)

To a solution of 6-(((4-methoxyphenyl)diphenylmethyl)amino)hexan-1-ol **(54)** (1.0 g, 2.57 mmol, 1 eq) in anhydrous pyridine (10 mL) was added succinic anhydride (287 mg, 2.57 mmol, 1 eq) and DMAP (31.3 mg, 0.26 mmol, 0.1 eq) and the solution was stirred at 50°C for 4 h. Then DMAP (62.6 mg, 0.52 mmol, 0.2 eq) and 6-(((4-methoxyphenyl)diphenylmethyl)amino)hexan-1-ol (0.1 eq) were added and the solution was stirred at RT for 18 h. The reaction was concentrated by rotary evaporation, taken up in Ethyl acetate and washed with ice cold 5% citric acid (aqueous solution),

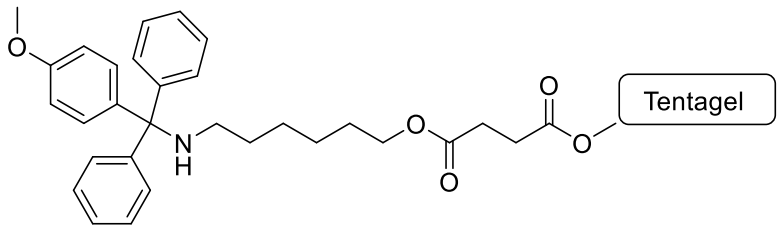
extracted with Ethyl acetate (2 x 20mL), dried over $MgSO_4$ and concentrated by rotary evaporation. The resulting yellow oil was purified by column chromatography on SiO_2 eluted by DCM/EtOAc (1:1) to DCM/MeOH (95:5) and concentrated to yield **55** as yellow oil (980 mg, 78%, 2 mmol).

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.16 - 7.60 (m, 12 H, MMTH), 6.81 (d, *J*=8.93 Hz, 2 H, MMTH) 4.04 (t, *J*=6.66 Hz, 2 H, OCH_2) 3.79 (s, 3 H, MMTOCH_3) 2.58 (m, 4 H, $COCH_2CH_2CO$) 2.14 (t, *J*=6.97 Hz, 2 H, NCH_2) 1.28 - 1.65 (m, 8 H, $CH_2CH_2CH_2$)

^{13}C NMR (101 MHz, $CDCl_3$) δ 176.9 173.2 157.8 146.4 138.3 129.8 128.6 27.7 113.0 70.5 64.4 55.2 43.5 30.7 30.6 30.2 28.5 26.9 25.9

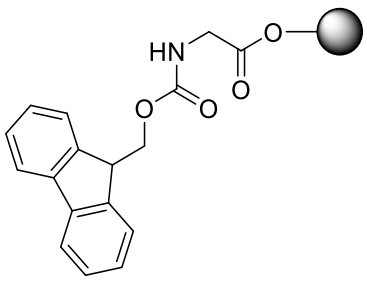
LRMS (ESI) calcd for $C_{30}H_{35}NO_5$+Na $[M+Na]^+$ *m/z*=512.24 found 512.1

Analysis are consistent with literature[164]



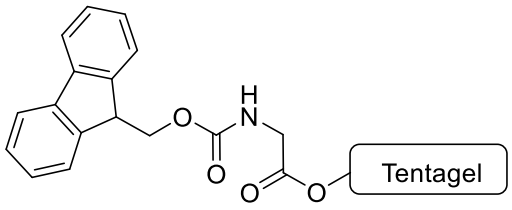
6-((4-Methoxyphenyl)-diphenylmethylamino)-hex-l-y1 succinylamido-Tentagel (57)

Hydroxy-tentagel (500 mg, 0.3 mmol/g) was swollen for 10 minutes in NMM (100 µL) and DMF (5 mL), then filtered. The resin was added to a solution of 4-((6-(((4-methoxyphenyl)diphenylmethyl)amino)hexyl)oxy)-4-oxobutanoic acid (780 mg, 11 eq) with NMM (378 µL, 21 eq), HATU (693 mg, 1.82 mmol, 11 eq) in DMF (13 mL) and shaken at RT for 21 h. The solution was filtered and the resin was washed with DMF (3 x 3 mL), DCM (3 x 1 mL) and Et2O (3 x 1 mL). The resin was dried under high vacuum for 30 min, then capped with CapA (1 mL) and CapB (1 mL) solutions by agitation at RT. Filtering and washing of the resin were achieved by DCM (3x3 mL) and Et2O (3x3 mL). The resin loading was measured by taking out 1 mg of the resin and deprotecting with solution of 3% TFA in DCM (10mL), the MMT ion concentration was then measured on a UV-Visible spectrophotometer by taking the absorbance at 478 nm of a dilution 1/100 of the previous deprotection solution. The final loading was 94 µmol/g.



Fmoc-glycine Nittophase (58)

In a dry 10 mL peptide vessel under argon atmosphere, HL Nittophase native (1.0 g, 566 µmol/g) was washed with DCM (5 mL) and DMF (2 x 5 mL). The resin was then activated with DMF/NMM (5:1, 6 mL). A solution of HATU (1.7g, 4.5 mmol, 8 eq), NMM (548 µL, 5.0 mmol, 8.8 eq) and glycine (1.35g, 4.5 mmol, 8 eq) in DMF (7 mL) was added to the support and mixed by mechanical rotation of the peptide vessel at RT for 4 h. The support was then washed with DMF (2 x 5 mL) and DCM (2 x 5 mL). The support was then reacted again with a solution of HATU (1.7g, 4.5 mmol, 8 eq), NMM (548 µL, 5.0 mmol, 8.8 eq) and glycine (1.35g, 4.5 mmol, 8 eq) in 7 mL of DMF and mixed at RT overnight. The support was then washed with DMF (2 x 10 mL) and DCM (3 x 10 mL) and dried under vacuum in a desiccator for 2 h. A sample of resin was taken aside to measure the loading by measuroffing the absorbance of the deprotected Fmoc. The final loading was found to be 122 µmol/g.



Fmoc-glycine Tentagel (59)

In a dry 10 mL peptide vessel under argon atmosphere, Tentagel-OH native (100 mg, 300 µmol/g) was washed once with 5 mL DCM and twice with 5 mL DMF. The resin was then activated with 6 mL of 5:1 DMF:NMM. A solution of HATU (91 mg, 240 µmol, 8 eq) NMM (29 µL, 2.6 µmol, 8.8 eq) and glycine (135 mg, 2.4 µmol, 8 eq) in 7 mL of DMF was added to the support and mixed by mechanical rotation of the peptide vessel at RT for 4 h. The support was then washed with DMF (2x5 mL) and DCM (2x5 mL). The support was then reacted again with a solution of HATU (91mg, 240 µmol, 8 eq) NMM (29µL, 2.6 µmol, 8.8 eq) and glycine (135 mg, 2.4 µmol, 8 eq) in 7 mL of DMF and mixed at RT overnight. The support was then washed with DMF (2x10 mL) and DCM (3x10 mL) and dried under vacuum in a desiccator for 2 h. A sample of resin was taken aside to measure the loading by measure of absorbance of the deprotected Fmoc. Final loading 45 µmol/g.

6.2 PNA methods

PNA oligomers were synthesized at 2µmol scale on Fmoc-PAL-PEG-PS solid support (Applied Biosystems) using an Expedite 8909 synthesizer. Fmoc/Bhoc-protected PNA monomers (Link Technologies) were dissolved to 0.2M in anhydrous *N*-methylpyrrolidinone; amino acid monomers (Sigma Aldrich) and AEEA linker (Link Technologies) were dissolved to 0.2 M in anhydrous dimethylformamide. Coupling time was 8.5 min using HATU (Alfa Aesar) as activator; double coupling was performed on selected bases.

PNAs were cleaved and deprotected with 400 µL of 19:1 TFA:*m*-Cresol for 90 min at room temperature and the resin was separated with a PTFE centrifugal filter. PNAs were then precipitated from cold diethyl ether and resuspended in MilliQ water.

PNAs were purified by HPLC with an Agilent 1260 system using a Waters XSelect CSH C18 5µm column (4.6x150mm). Buffer A was 0.1%TFA in water, Buffer B was 0.1%TFA in ACETONITRILE, and purification was performed with a gradient of 5%B to 20%B over 18min at 60°C.

PNAs were characterized on an Agilent 6530 Q-TOF LC/MS system with electrospray ionization and time of flight ion separation in positive ionization mode. The data were analyzed using Agilent Mass Hunter software. The LC/MS eluent was a gradient of acetonitrile in water containing 0.1% of formic acid.

6.3 Oligonucleotide methods

Oligonucleotides were synthesized at 1 µmole scale on an Applied Biosystems 394 DNA synthesizer. BTT or ETT (0.25 M in acetonitrile, ChemGenes) was used as activator. 0.02 M iodine in THF/water/pyridine (TEDIA or ChemGenes) was used as oxidizer. Sulfurization was accomplished with DDTT or EDITH (0.1 M, ChemGenes) using a 1 min wait time. 3% TCA in DCM (TEDIA) was used as deblock solution.

LNA phosphoramidites were synthesized in-house from the 3'-hydroxy nucleoside precursor (Rasayan Inc.) and were dissolved in acetonitrile to 0.15 M. LNA-C^{Bz} was solubilized in DCM to 0.15 M. DNA phosphoramidites (ChemGenes) were dissolved in acetonitrile to 0.1 M. Coupling time was 10 min for RNA and LNA and was 40 sec for DNA.

In most cases, oligonucleotides were grown on 1000 Å CPG functionalized with Unylinker (~42 µmol/g).

Oligonucleotides not containing RNA were deprotected with concentrated aqueous NH_3 (1 mL) at 55 °C for 16 h. Then the oligonucleotides were cooled to RT, left open for several minutes to allow some reduction of the NH_3 pressure, then it was snap-freeze and evaporated to dryness in a centrifugal evaporator and resuspended in 1 mL RNase-free water.

RNA-containing oligonucleotides were deprotected with 40% aqueous $MeNH_2$ for 1 h at 65°C, then deprotection of the TBDMS group was achieved with DMSO/$NEt_3 \cdot 3HF$ (4:1) solution (500 µL) at 65 °C for 3 h. RNA oligonucleotides were then recovered by precipitation in 3M NaOAc (25 µL) and n-BuOH (1 mL), and the pellet was washed with cold 70% EtOH and resuspended in 1 mL RNase-free water.

Purification of oligonucleotides were carried out by high performance liquid chromatography using a 1260 infinity system with an Agilent PL-SAX 1000 Å column (150 x 7.5 mm, 8 µm). Buffer A: 30% ACETONITRILE in water; Buffer B: 30% acetonitrile in 1M $NaClO_{4\,aq}$. Excess salt was removed with Sephadex Nap-25 column.

Oligonucleotides were diluted to 20µM concentration and analyzed on a Bruker MicrOTOF Ultimate 3000 spectrometer with electrospray and time of flight in negative ionization mode. The data was analyzed using Compas DataAnalysis software. Buffer A: 100mM hexafluoroisopropanol with 10mM triethylammonium acetate; Buffer B: 20mM triethylammonium acetate and acetonitrile.

6.4 RNase H assays

RNase H assays were conducted in 200-µL PCR tubes with E. *coli* RNase H (NEB, 5000u/mL). The solution of RNA (5uM in water) and ASO (5uM in water) were prepared beforehand. A solution of 1/10 Enzyme in 1x reaction Buffer (75 mM KCl, 50 mM Tris-HCl, 3 mM $MgCl_2$, 10 mM dithiothreitol) was prepared and kept on ice.

For 1:5 ASO:RNA studies

On ice, 65 uL of water was added to a 200 µL PCR tube with 8.4 µL of 10x reaction buffer. Then 1.6 µL of ASO 5µM and 8 µL of prepared 1/10 enzyme solution were added. The addition of the 8 µL of RNA is the start of the assay, the RNA (5µM) was added and 10 µL of the reactive solution was taken out to be mixed with 10 µL of formamide and kept at -20°C. The reaction was kept at RT protected from light and 10 µL were taken out at every time point and mixed with formamide and kept at -20°C. After the reaction, the samples at -20°C were taken and rapidly heated to 75°C for 5 min to ensure total enzyme inactivation. For each time point, 10 µL was taken out to be mixed with 10 µL of RNA Sample Loading Buffer (Sigma Aldrich). 15 µL of the sample was loaded on a denaturing 20% polyacrylamide gel at 400V. The gel was then read with a GE -Typhoon FLA 9500 gel imager on the FAM channel. The relative intensity of the gel bands was measured with ImageJ.

For 1:1 ASO:RNA studies

On ice, 58.6 µL of water was added to a 200 µL PCR tube with 8.4 µL of 10x reaction buffer. Then 8 µL of ASO (5µM) and 8 µL of prepared enzyme solution were added. After this point, the experiment was carried out as described for the 1:5 ASO:RNA protocol above.

6.5 Cell culture and transfection

The mouse embryonic fibroblast (MEF) WT cells and mouse hepatoma cells (Hepa 1-6) were purchased from ATCC and cultured in regular DMEM media supplemented with 10 % FBS . Cells were maintained in 5% CO_2 humidified incubator.

For Lipid transfection, MEF cells were plated in 96 well plates at 8 K cells/well at least 16 h prior to transfection. Cells were transfected with non-targeting controls and chimeras and LNA gapmers targeting MALAT 1 using Lipofectamine RNAiMAX (Life technologies). Cells were harvested 16 h post transfection and *Malat1* and *Hprt* mRNA levels were measured using the Quantigene 2.0 assay (Affymetrix,#QS0011) following, the manufactures protocol. IC_{50} concentrations were obtained by dose-response analysis using interpolation of the log of concentration with GraphPad Prism.

For gymnotic delivery of chimeras and LNA gapmers , Hepa 1-6 cells were plated at 11k cells/well in 12 well tissue-culture-treated plates. Cells were treated with oligonucleotides at 5 different concentrations in complete media. Cells were harvested after 5 d and assayed for *Malat1* and *Hprt* mRNA levels using the Quantigene 2.0 assay kit (Affymetrix, # QS0011) following the manufacturers protocol. IC_{50} concentrations were obtained by dose-response analysis.

Probes used for the Quantigene assay were Mouse *Hprt*: SB-15463 and Mouse *Malat1*: SB-26581.

Bibliography

1. J. D. Watson and F. H. Crick, *Nature*, 1953, **171**, 737-738.
2. C. Altona and M. Sundaralingam, *J. Am. Chem. Soc.*, 1972, **94**, 8205-8212.
3. D. B. Davies, *Prog. Nucl. Magn. Reson. Spectrosc.*, 1978, **12**, 135-225.
4. G. M. Blackburn and M. J. Gait, *Nucleic acids in chemistry and biology*, IRL Press at Oxford University Press, Oxford; New York, 1990.
5. P. S. Ho, *Proceedings of the National Academy of Sciences*, 1994, **91**, 9549-9553.
6. A. Rich, A. Nordheim and A. H. J. Wang, *Annu. Rev. Biochem.*, 1984, **53**, 791-846.
7. R. R. Sinden, *DNA structure and function*, Academic Press, San Diego, 1994.
8. J. Minshull and T. Hunt, *Nucleic Acids Res.*, 1986, **14**, 6433-6451.
9. R. Y. Walder and J. A. Walder, *Proceedings of the National Academy of Sciences*, 1988, **85**, 5011-5015.
10. A. N. Lane, S. Ebel and T. Brown, *Eur. J. Biochem.*, 1993, **215**, 297-306.
11. A. L. Dounce, *Nature*, 1953, **172**, 541.
12. T. Jenuwein and C. D. Allis, *Science*, 2001, **293**, 1074-1080.
13. T. Ito, in *Protein Complexes that Modify Chromatin*, ed. J. L. Workman, Springer Berlin Heidelberg, Berlin, Heidelberg, 2003, DOI: 10.1007/978-3-642-55747-7_1, pp. 1-22.
14. E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, et al., *Nature*, 2001, **409**, 860-921.
15. M. Guttman, I. Amit, M. Garber, C. French, M. F. Lin, D. Feldser, et al., *Nature*, 2009, **458**, 223-227.
16. P. Kapranov, J. Cheng, S. Dike, D. A. Nix, R. Duttagupta, A. T. Willingham, et al., *Science*, 2007, **316**, 1484-1488.
17. J. S. Mattick and I. V. Makunin, *Hum. Mol. Genet.*, 2006, **15**, R17-R29.
18. A. Fire, S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver and C. C. Mello, *Nature*, 1998, **391**, 806-811.
19. G. J. Hannon and J. J. Rossi, *Nature*, 2004, **431**, 371-378.
20. O. A. Kent and A. M. MacMillan, *Org. Biomol. Chem.*, 2004, **2**, 1957-1961.
21. S. W. Jones, P. M. de Souza and M. A. Lindsay, *Curr. Opin. Pharm.*, 2004, **4**, 522-527.
22. C. R. Allerson, N. Sioufi, R. Jarres, T. P. Prakash, N. Naik, A. Berdeja, et al., *J. Med. Chem.*, 2005, **48**, 901-904.

23. T. P. Prakash, C. R. Allerson, P. Dande, T. A. Vickers, N. Sioufi, R. Jarres, et al., *J. Med. Chem.*, 2005, **48**, 4247-4253.

24. G. F. Deleavey, J. K. Watts, T. Alain, F. Robert, A. Kalota, V. Aishwarya, et al., *Nucleic Acids Res.*, 2010, **38**, 4547-4557.

25. J. M. Layzer, A. P. McCaffrey, A. K. Tanner, Z. Huang, M. A. Kay and B. A. Sullenger, *RNA*, 2004, **10**, 766-771.

26. D. Yu, H. Pendergraff, J. Liu, H. B. Kordasiewicz, Don W. Cleveland, Eric E. Swayze, et al., *Cell.*, 2012, **150**, 895-908.

27. Walt F. Lima, Thazha P. Prakash, Heather M. Murray, Garth A. Kinberger, W. Li, Alfred E. Chappell, et al., *Cell.*, 2012, **150**, 883-894.

28. D. Grimm, *Gene Ther.*, 2009, **16**, 827.

29. H. M. Pendergraff, A. J. Debacker and J. K. Watts, *Nucleic Acid Ther.*, 2016, **26**, 216-222.

30. H. Stein and P. Hausen, *Science*, 1969, **166**, 393-395.

31. S. M. Cerritelli and R. J. Crouch, *FEBS J.*, 2009, **276**, 1494-1505.

32. W. F. Lima, H. M. Murray, S. S. Damle, C. E. Hart, G. Hung, C. L. De Hoyos, et al., *Nucleic Acids Res.*, 2016, **44**, 5299-5312.

33. S. T. Crooke, *Antisense drug technology: principles, strategies, and applications*, CRC Press, 2007.

34. H. Wu, W. F. Lima and S. T. Crooke, *Antisense Nucleic Acid Drug Dev.*, 1998, **8**, 53-61.

35. H. Wu, W. F. Lima and S. T. Crooke, *J. Biol. Chem.*, 2001, **276**, 23547-23553.

36. O. Plashkevych, Q. Li and J. Chattopadhyaya, *Molecular bioSystems*, 2017, **13**, 921-938.

37. W. F. Lima, H. Wu, J. G. Nichols, T. P. Prakash, V. Ravikumar and S. T. Crooke, *J. Biol. Chem.*, 2003, **278**, 49860-49867.

38. J. K. Watts and D. R. Corey, *J. Pathol.*, 2012, **226**, 365-379.

39. R. L. Beane, R. Ram, S. Gabillet, K. Arar, B. P. Monia and D. R. Corey, *Biochemistry*, 2007, **46**, 7572-7580.

40. J. Hu and D. R. Corey, *Biochemistry*, 2007, **46**, 7581-7589.

41. B. A. Janowski, K. Kaihatsu, K. E. Huffman, J. C. Schwartz, R. Ram, D. Hardy, et al., *Nat. Chem. Biol.*, 2005, **1**, 210-215.

42. B. F. Baker, S. S. Lot, T. P. Condon, S. Cheng-Flournoy, E. A. Lesnik, H. M. Sasmor, et al., *J. Biol. Chem.*, 1997, **272**, 11994-12000.

43. M. Faria, D. G. Spiller, C. Dubertret, J. S. Nelson, M. R. H. White, D. Scherman, et al., *Nat Biotech*, 2001, **19**, 40-44.

44. J. R. Mendell, L. R. Rodino-Klapac, Z. Sahenk, K. Roush, L. Bird, L. P. Lowes, et al., *Ann. Neurol.*, 2013, **74**, 637-647.

45. A. Aartsma-Rus, *Nucleic Acid Ther.*, 2017, **27**, 67-69.

46. K. A. Lennox and M. A. Behlke, *Gene Ther.*, 2011, **18**, 1111.

47. K. A. Lennox and M. A. Behlke, *Pharm. Res.*, 2010, **27**, 1788-1799.

48. S. F. Dowdy, *Nat Biotech*, 2017, **35**, 222-229.

49. M. Robbins, A. Judge, L. Liang, K. McClintock, E. Yaworski and I. MacLachlan, *Mol. Ther.*, 2007, **15**, 1663-1669.

50. I. Lebedeva and C. A. Stein, *Annu. Rev. Pharmacol. Toxicol.*, 2001, **41**, 403-419.

51. K. R. Ito and S. Obika, in *Comprehensive Medicinal Chemistry III*, eds. D. Rotella and S. E. Ward, Elsevier, Oxford, 2017, DOI: https://doi.org/10.1016/B978-0-12-409547-2.12420-5, pp. 216-232.

52. Isis Pharmaceuticals Inc).

53. C. F. Bennett and E. E. Swayze, *Annu. Rev. Pharmacol. Toxicol.*, 2010, **50**, 259-293.

54. M. J. Damha, C. J. Wilds, A. Noronha, I. Brukner, G. Borkow, D. Arion, et al., *J. Am. Chem. Soc.*, 1998, **120**, 12976-12977.

55. J. T. Nielsen, K. Arar and M. Petersen, *Nucleic Acids Res.*, 2006, **34**.

56. D. A. Braasch and D. R. Corey, *Chem. Biol.*, 2001, **8**, 1-7.

57. C. Wahlestedt, P. Salmi, L. Good, J. Kela, T. Johnsson, T. Hökfelt, et al., *Proceedings of the National Academy of Sciences*, 2000, **97**, 5633-5638.

58. P. P. Seth, A. Siwkowski, C. R. Allerson, G. Vasquez, S. Lee, T. P. Prakash, et al., *Nucleic Acids Symp. Ser.*, 2008, **52**, 553-554.

59. B. P. Monia, E. A. Lesnik, C. Gonzalez, W. F. Lima, D. McGee, C. J. Guinosso, et al., *J. Biol. Chem.*, 1993, **268**, 14514-14522.

60. Z. Li and T. M. Rana, *Nat. Rev. Drug Discov.*, 2014, **13**, 622-638.

61. D. V. Morrissey, K. Blanchard, L. Shaw, K. Jensen, J. A. Lockridge, B. Dickinson, et al., *Hepatology*, 2005, **41**, 1349-1356.

62. A. Khvorova, *New Engl. J. Med.*, 2017, **376**, 4-7.

63. F. Eckstein, *Nucleic Acid Ther.*, 2014, **24**, 374-387.

64. F. Eckstein and H. Sternbach, *Biochimica et Biophysica Acta (BBA) - Enzymology*, 1967, **146**, 618-619.

65. N. Dias and C. A. Stein, *Mol. Cancer Ther.*, 2002, **1**, 347-355.

66. S. T. Crooke, S. Wang, T. A. Vickers, W. Shen and X.-h. Liang, *Nat Biotech*, 2017, **35**, 230-237.

67. P. J. Furdon, Z. Dominski and R. Kole, *Nucleic Acids Res.*, 1989, **17**, 9193-9204.

68. C. A. Stein, C. Subasinghe, K. Shinozuka and J. S. Cohen, *Nucleic Acids Res.*, 1988, **16**, 3209-3221.

69. W. Shen, X. H. Liang, H. Sun and S. T. Crooke, *Nucleic Acids Res.*, 2015, **43**, 4569-4578.

70. M. Boczkowska, P. Guga and W. J. Stec, *Biochemistry*, 2002, **41**, 12483-12487.

71. W. J. Stec, C. S. Cierniewski, A. Okruszek, A. Kobylanska, Z. Pawlowska, M. Koziolkiewicz, et al., *Antisense Nucleic Acid Drug Dev.*, 1997, **7**, 567-573.

72. N. Iwamoto, D. C. D. Butler, N. Svrzikapa, S. Mohapatra, I. Zlatev, D. W. Y. Sah, et al., *Nat Biotech*, 2017, **35**, 845-851.

73. J. Summerton and D. Weller, *Antisense Nucleic Acid Drug Dev.*, 1997, **7**, 187-195.

74. P. E. Nielsen, M. Egholm, R. H. Berg and O. Buchardt, *Anti-Cancer Drug Des.*, 1993, **8**, 53.

75. J. C. Hanvey, N. J. Peffer, J. E. Bisi, S. A. Thomson, R. Cadilla, J. A. Josey, et al., *Science*, 1992, **258**, 1481-1485.

76. A. Nasevicius and S. C. Ekker, *Nat. Genet.*, 2000, **26**, 216-220.

77. J. S. Eisen and J. C. Smith, *Development*, 2008, **135**, 1735-1743.

78. M. M. Fabani, C. Abreu-Goodger, D. Williams, P. A. Lyons, A. G. Torres, K. G. Smith, et al., *Nucleic Acids Res.*, 2010, **38**, 4466-4475.

79. B. Lebleu, H. M. Moulton, R. Abes, G. D. Ivanova, S. Abes, D. A. Stein, et al., *Adv Drug Deliv Rev*, 2008, **60**, 517-529.

80. G. J. Doherty and H. T. McMahon, *Annu Rev Biochem.*, 2009, **78**, 857-902.

81. C. A. Stein, J. B. Hansen, J. Lai, S. Wu, A. Voskresenskiy, A. Høg, et al., *Nucleic Acids Res.*, 2010, **38**, e3-e3.

82. D. Castanotto, M. Lin, C. Kowolik, L. Wang, X. Q. Ren, H. S. Soifer, et al., *Nucleic Acids Res.*, 2015, **43**, 9350-9361.

83. J. K. Nair, J. L. S. Willoughby, A. Chan, K. Charisse, M. R. Alam, Q. Wang, et al., *J. Am. Chem. Soc.*, 2014, **136**, 16958-16961.

84. T. P. Prakash, M. J. Graham, J. Yu, R. Carty, A. Low, A. Chappell, et al., *Nucleic Acids Res.*, 2014, **42**, 8796-8807.

85. M. Caruthers, *Science*, 1985, **230**, 281-285.

86. G. Alvarado-Urbina, G. M. Sathe, W. C. Liu, M. F. Gillen, P. D. Duck, R. Bender, et al., *Science*, 1981, **214**, 270-274.

87. S. L. Beaucage and M. H. Caruthers, *Tetrahedron Lett.*, 1981, **22**, 1859-1862.

88. P. E. Nielsen, M. Egholm, R. H. Berg and O. Buchardt, *Science*, 1991, **254**, 1497-1500.

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

90. M. Egholm, O. Buchardt, L. Christensen, C. Behrens, S. M. Freier, D. A. Driver, et al., *Nature*, 1993, **365**, 566-568.

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

92. P. E. Nielsen, *Peptide nucleic acids : protocols and applications*, Horizon Bioscience, Wymondham, 2004.

93. C. Sharma and S. K. Awasthi, *Chem. Biol. Drug Des.*, 2017, **89**, 16-37.

94. E. A. Englund and D. H. Appella, *Angew. Chem. Int. Ed.*, 2007, **46**, 1414-1418.

95. A. Manicardi and R. Corradini, *Artificial DNA, PNA & XNA*, 2014, **5**, e1131801.

96. A. Manna, S. Rapireddy, R. Bahal and D. H. Ly, *Methods Mol. Biol.*, 2014, **1050**, 1-12.

97. B. Sahu, I. Sacui, S. Rapireddy, K. J. Zanotti, R. Bahal, B. A. Armitage, et al., *The Journal of organic chemistry*, 2011, **76**, 5614-5627.

98. V. V. Demidov, V. N. Potaman, M. Frank-Kamenetskil, M. Egholm, O. Buchard, S. H. Sönnichsen, et al., *Biochem. Pharmacol.*, 1994, **48**, 1310-1313.

99. R. G. Kuimelis, A. C. van der Laan and R. Vinayak, *Tetrahedron Lett.*, 1999, **40**, 7671-7674.

100. B. A. Armitage, *Drug Discov. Today*, 2003, **8**, 222-228.

101. K. K. Jensen, H. Ørum, P. E. Nielsen and B. Nordén, *Biochemistry*, 1997, **36**, 5072-5077.

102. N. Dias, C. Sénamaud-Beaufort, E. l. Forestier, C. Auvin, C. Hélène and T. Ester Saison-Behmoaras, *J. Mol. Biol.*, 2002, **320**, 489-501.

103. H. Kim, K. H. Lee, K. B. Kim, Y. S. Park, K. S. Kim and D. E. Kim, *Bull. Korean Chem. Soc.*, 2013, **34**, 735-742.

104. Y. M. Bae, M. H. Kim, G. S. Yu, B. H. Um, H. K. Park, H.-i. Lee, et al., *J. Controlled Release*, 2014, **175**, 54-62.

105. H. Knudsen and P. E. Nielsen, *Nucleic Acids Res.*, 1996, **24**, 494-500.

106. I. Amit-Avraham, G. Pozner, S. Eshar, Y. Fastman, N. Kolevzon, E. Yavin, et al., *Proceedings of the National Academy of Sciences*, 2015, **112**, E982-E991.

107. S. M. Misenko and S. F. Bunting, *Journal of visualized experiments : JoVE*, 2014, DOI: 10.3791/51806.

108. M. D. Genet, I. M. Cartwright and T. A. Kato, *Molecular Cytogenetics*, 2013, **6**, 42.

109. Z.-C. Liu, D.-S. Shin, M. Shokouhimehr, K.-N. Lee, B.-W. Yoo, Y.-K. Kim, et al., *Biosensors Bioelectron.*, 2007, **22**, 2891-2897.

110. F. R. Raymond, H. A. Ho, R. Peytavi, L. Bissonnette, M. Boissinot, F. J. Picard, et al., *BMC Biotechnol.*, 2005, **5**, 10.

111. O. Brandt, J. Feldner, A. Stephan, M. Schroder, M. Schnolzer, H. F. Arlinghaus, et al., *Nucleic Acids Res.*, 2003, **31**, e119.

112. H. Orum, P. E. Nielsen, M. Egholm, R. H. Berg, O. Buchardt and C. Stanley, *Nucleic Acids Res.*, 1993, **21**, 5332-5336.

113. P. E. Nielsen, *Peptide nucleic acids: protocols and applications*, Garland Science, 2004.

114. E. Uhlmann, A. Peyman, G. Breipohl and D. W. Will, *Angew. Chem. Int. Ed.*, 1998, **37**, 2796-2823.

115. M. M. Fabani, C. Abreu-Goodger, D. Williams, P. A. Lyons, A. G. Torres, K. G. C. Smith, et al., *Nucleic Acids Res.*, 2010, **38**, 4466-4475.

116. L. Bialy, J. J. Diaz-Mochon, E. Specker, L. Keinicke and M. Bradley, *Tetrahedron*, 2005, **61**, 8295-8305.

117. P. Wittung, J. Kajanus, K. Edwards, P. Nielsen, B. Norden and B. G. Malmstroem, *FEBS Lett.*, 1995, **365**, 27-29.

118. M. A. Bonham, S. Brown, A. L. Boyd, P. H. Brown, D. A. Bruckenstein, J. C. Hanvey, et al., *Nucleic Acids Res.*, 1995, **23**, 1197-1203.

119. G. D. Gray, S. Basu and E. Wickstrom, *Biochem. Pharmacol.*, 1997, **53**, 1465-1476.

120. W. M. Pardridge, R. J. Boado and Y. S. Kang, *Proc. Natl. Acad. Sci. U. S. A.*, 1995, **92**, 5592-5596.

121. F. Heitz, M. C. Morris and G. Divita, *Br. J. Pharmacol.*, 2009, **157**, 195-206.

122. M. F. N. Abushahba, H. Mohammad, S. Thangamani, A. A. A. Hussein and M. N. Seleem, 2016, **6**, 20832.

123. C. A. Henderson, H. A. Vincent, C. M. Stone, J. O. Phillips, P. D. Cary, D. M. Gowers, et al., *Nucleic Acids Res.*, 2013, **41**, 3386-3397.

124. S. Gottesman and G. Storz, *Cold Spring Harb. Perspect. Biol.*, 2011, **3**.

125. K. I. Udekwu, F. Darfeuille, J. Vogel, J. Reimegaard, E. Holmqvist, E. Gerhart, et al., *Genes Dev.*, 2005, **19**, 2355-2366.

126. K. I. Udekwu, *PLoS One*, 2010, **5**, No pp. given.

127. S. C. Viegas, I. J. Silva, M. Saramago, S. Domingues and C. M. Arraiano, *Nucleic Acids Res.*, 2011, **39**, 2918-2930.

128. Y. Wang, *Biochem. Biophys. Res. Commun.*, 2002, **292**, 396-401.

129. M. Zuker, *Nucleic Acids Res.*, 2003, **31**, 3406-3415.

130. PrimerQuest® program, IDT, Coralville, USA. Retrieved 12 December, 2012. http://www.idtdna.com/Scitools).

131. F. Eckstein, *Tetrahedron Lett.*, 1967, **8**, 1157-1160.

132. F. Simeoni, M. C. Morris, F. Heitz and G. Divita, *Nucleic Acids Res.*, 2003, **31**, 2717-2724.

133. T. H. Caddick, Master of Philosophy, University of Portsmouth, 2017.

134. L. Good, S. K. Awasthi, R. Dryselius, O. Larsson and P. E. Nielsen, *Nat Biotech*, 2001, **19**, 360-364.

135. J. K. Watts, G. F. Deleavey and M. J. Damha, *Drug Discov. Today*, 2008, **13**, 842-855.

136. A. Khvorova and J. K. Watts, *Nat Biotech*, 2017, **35**, 238-248.

137. R. A. Haraszti, L. Roux, A. H. Coles, A. A. Turanov, J. F. Alterman, D. Echeverria, et al., *Nucleic Acids Res.*, 2017, **45**, 7581-7592.

138. M. Nikan, M. F. Osborn, A. H. Coles, A. Biscans, B. Godinho, R. A. Haraszti, et al., *Bioconj. Chem.*, 2017, **28**, 1758-1766.

139. J. F. Alterman, L. M. Hall, A. H. Coles, M. R. Hassler, M. C. Didiot, K. Chase, et al., *Mol Ther Nucleic Acids*, 2015, **4**, e266.

140. M. R. Hassler, A. A. Turanov, J. F. Alterman, R. A. Haraszti, A. H. Coles, M. F. Osborn, et al., *Nucleic Acids Res.*, 2018, **46**, 2185-2196.

141. N. Potenza, L. Moggio, G. Milano, V. Salvatore, B. Di Blasio, A. Russo, et al., *International Journal of Molecular Sciences*, 2008, **9**, 299-315.

142. M. Kitamatsu, T. Kubo, R. Matsuzaki, T. Endoh, T. Ohtsuki and M. Sisido, *Bioorg. Med. Chem. Lett.*, 2009, **19**, 3410-3413.

143. P. E. Nielsen and T. Shiraishi, *Artificial DNA, PNA & XNA*, 2011, **2**, 90-99.

144. E. K. Lee, C. W. Kim, H. Kawanami, A. Kishimura, T. Niidome, T. Mori, et al., *RSC Advances*, 2015, **5**, 85816-85821.

145. A. El-Sayed, S. Futaki and H. Harashima, *The AAPS Journal*, 2009, **11**, 13-22.

146. J. Kypr, I. Kejnovská, D. Renčiuk and M. Vorlíčková, *Nucleic Acids Res.*, 2009, **37**, 1713-1725.

147. T. Lehto, K. Ezzat, M. J. A. Wood and S. El Andaloussi, *Adv. Drug Del. Rev.*, 2016, **106**, 172-182.

148. M. Ardhammar, B. Nordén, P. E. Nielsen, B. G. Malmström and P. Wittung-Stafshede, *J. Biomol. Struct. Dyn.*, 1999, **17**, 33-40.

149. K. H. Petersen, D. K. Jensen, M. Egholm, P. E. Nielsen and O. Buchardt, *Bioorg. Med. Chem. Lett.*, 1995, **5**, 1119-1124.

150. A. C. van der Laan, P. Havenaar, R. S. Oosting, E. Kuyl-Yeheskiely, E. Uhlmann and J. H. van Boom, *Bioorg. Med. Chem. Lett.*, 1998, **8**, 663-668.

151. E. Uhlmann, D. W. Will, G. Breipohl, D. Langner and A. Ryte, *Angew. Chem. Int. Ed.*, 1996, **35**, 2632-2635.

152. L. Moggio, A. Romanelli, R. Gambari, N. Bianchi, M. Borgatti, E. Fabbri, et al., *Biopolymers*, 2007, **88**, 815-822.

153. B. M. Patureau, R. H. Hudson and M. J. Damha, *Bioconj. Chem.*, 2007, **18**, 421-430.

154. F. Bergmann, W. Bannwarth and S. Tam, *Tetrahedron Lett.*, 1995, **36**, 6823-6826.

155. Z. Bajor, G. Sagi, Z. Tegyey and L. Otvos, *Nucleosides Nucleotides Nucleic Acids*, 2003, **22**, 1215-1217.

156. P. J. Finn, N. J. Gibson, R. Fallon, A. Hamilton and T. Brown, *Nucleic Acids Res.*, 1996, **24**, 3357-3363.

157. D. A. Stetsenko, E. N. Lubyako, V. K. Potapov, T. L. Azhikina and E. D. Sverdlov, *Tetrahedron Lett.*, 1996, **37**, 3571-3574.

158. E. Uhlmann, *Biol. Chem.*, 1998, **379**, 1045-1052.

159. B. Greiner, G. Breipohl and E. Uhlmann, *Helv. Chim. Acta*, 1999, **82**, 2151-2159.

160. B. Greiner, G. Breipohl and E. Uhlmann, *Helv. Chim. Acta*, 2002, **85**, 2619-2626.

161. A. C. vanderLaan, R. Brill, R. G. Kuimelis, E. KuylYeheskiely, J. H. vanBoom, A. Andrus, et al., *Tetrahedron Lett.*, 1997, **38**, 2249-2252.

162. G. Barone, L. De Napoli, G. Di Fabio, E. Erra, C. Giancola, A. Messere, et al., *Nucleosides Nucleotides Nucleic Acids*, 2003, **22**, 1089-1091.

163. E. Ferrer, A. Shevchenko and R. Eritja, *Biorg. Med. Chem.*, 2000, **8**, 291-297.

164. D. W. Will, G. Breipohl, D. Langner, J. Knolle and E. Uhlmann, *Tetrahedron*, 1995, **51**, 12069-12082.

165. R. Vinayak, A. C. vanderLaan, R. Brill, K. Otteson, A. Andrus, E. KuylYeheskiely, et al., *Nucleosides Nucleotides*, 1997, **16**, 1653-1656.

166. D. Capasso, L. De Napoli, G. Di Fabio, A. Messere, D. Montesarchio, C. Pedone, et al., *Tetrahedron*, 2001, **57**, 9481-9486.

167. M. Koppitz, P. E. Nielsen and L. E. Orgel, *J. Am. Chem. Soc.*, 1998, **120**, 4563-4569.

168. D. Musumeci, G. N. Roviello, M. Valente, R. Sapio, C. Pedone and E. M. Bucci, *Biopolymers*, 2004, **76**, 535-542.

169. G. Breipohl, D. W. Will, A. Peyman and E. Uhlmann, *Tetrahedron*, 1997, **53**, 14671-14686.

170. A. S. Jones, P. Lewis and S. F. Withers, *Tetrahedron*, 1973, **29**, 2293-2296.

171. M. Okabe, R. C. Sun and G. B. Zenchoff, *J. Org. Chem.*, 1991, **56**, 4392-4397.

172. E. Ferrer and R. Eritja, *Lett. Pept. Sci.*, 2001, **7**, 195-206.

173. S. A. Thomson, J. A. Josey, R. Cadilla, M. D. Gaul, C. Fred Hassman, M. J. Luzzio, et al., *Tetrahedron*, 1995, **51**, 6179-6194.

174. K. L. Dueholm, M. Egholm, C. Behrens, L. Christensen, H. F. Hansen, T. Vulpius, et al., *The Journal of Organic Chemistry*, 1994, **59**, 5767-5773.

175. Z.-C. Liu, D.-S. Shin, K.-T. Lee, B.-H. Jun, Y.-K. Kim and Y.-S. Lee, *Tetrahedron*, 2005, **61**, 7967-7973.

176. Z. Timar, L. Kovacs, G. Kovacs and Z. Schmel, *Perkin 1*, 2000, DOI: 10.1039/A907832K, 19-26.

177. P. T. Gilham and H. G. Khorana, *J. Am. Chem. Soc.*, 1958, **80**, 6212-6222.

178. R. Greenhill, 2015.

179. T. L. Spurling, R. L. Eoff and K. D. Raney, *Bioorg. Med. Chem. Lett.*, 2006, **16**, 1816-1820.

180. Torviq, PPsyringe, http://www.torviq.com/images/PPsyringe1.jpg).

181. G. Arun, S. Diermeier, M. Akerman, K. C. Chang, J. E. Wilkinson, S. Hearn, et al., *Genes Dev.*, 2016, **30**, 34-51.

182. M. N. Cabili, M. C. Dunagin, P. D. McClanahan, A. Biaesch, O. Padovan-Merhar, A. Regev, et al., *Genome Biol.*, 2015, **16**, 20.

183. B. Zhang, G. Arun, Y. S. Mao, Z. Lazar, G. Hung, G. Bhattacharjee, et al., *Cell Reports*, 2012, **2**, 111-123.

184. C. Li, L. Chang, Z. Chen, Z. Liu, Y. Wang and Q. Ye, *Int. J. Mol. Med.*, 2017, **39**, 347-356.

185. V. A. Efimov, M. V. Choob, A. A. Buryakova, A. L. Kalinkina and O. G. Chakhmakhcheva, *Nucleic Acids Res.*, 1998, **26**, 566-575.

186. J. Fangkangwanwong, N. Sae-liang, C. Sriworarat, A. Sereemaspun and S. Chirachanchai, *Bioconj. Chem.*, 2016, **27**, 2301-2306.

187. J. Schmidt, B. Eschgfaeller and S. A. Benner, *Helv. Chim. Acta*, 2003, **86**, 2937-2958.

188. 1997.

189. V. Kumar, V. Kesavan and K. V. Gothelf, *Org. Biomol. Chem.*, 2015, **13**, 2366-2374.

190. T. Watanabe, T. Hoshida, J. Sakyo, M. Kishi, S. Tanabe, J. Matsuura, et al., *Org. Biomol. Chem.*, 2014, **12**, 5089-5093.

191. G. Ceulemans, A. VanAerschot, B. Wroblowski, J. Rozenski, C. Hendrix and P. Herdewijn, *Chem.-Eur. J.*, 1997, **3**, 1997-2010.

192. V. A. Efimov and O. G. Chakhmakhcheva, in *Oligonucleotide Synthesis*, ed. P. Herdewijn, Humana Press, Totowa, NJ, 2005, DOI: 10.1385/1-59259-823-4:147, pp. 147-163.

193. P. Li, J. Sun, M. Su, X. Yang and X. Tang, *Org. Biomol. Chem.*, 2014, **12**, 2263-2272.

194. Y. Hari, T. Kodama, T. Imanishi and S. Obika, in *Therapeutic Oligonucleotides: Methods and Protocols*, ed. J. Goodchild, Humana Press, Totowa, NJ, 2011, DOI: 10.1007/978-1-61779-188-8_3, pp. 31-57.

195. L. Yang, J. Li, R. Simionescu and H. Yan, *Magn. Reson. Chem.*, 2013, **51**, 60-64.

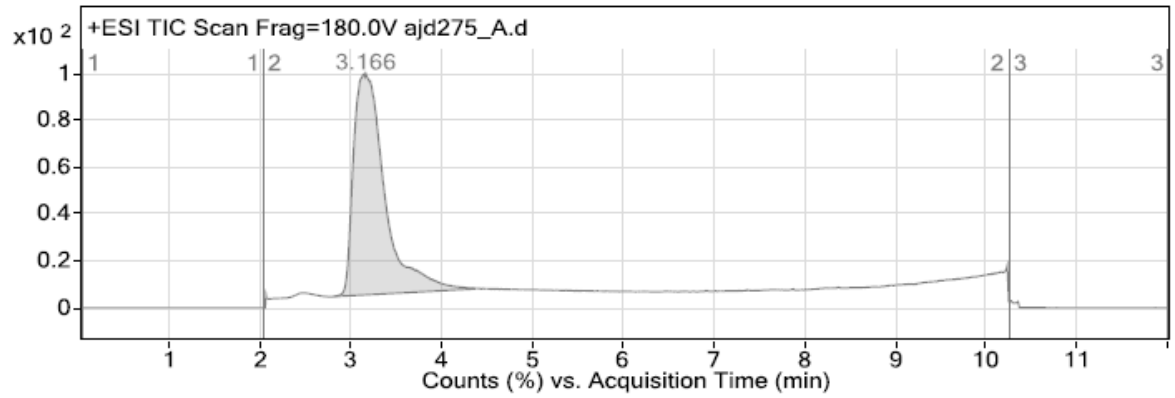
196. Z. Timar, L. Kovacs, G. Kovacs and Z. Schmel, *Journal of the Chemical Society, Perkin Transactions 1*, 2000, DOI: 10.1039/A907832K, 19-26.

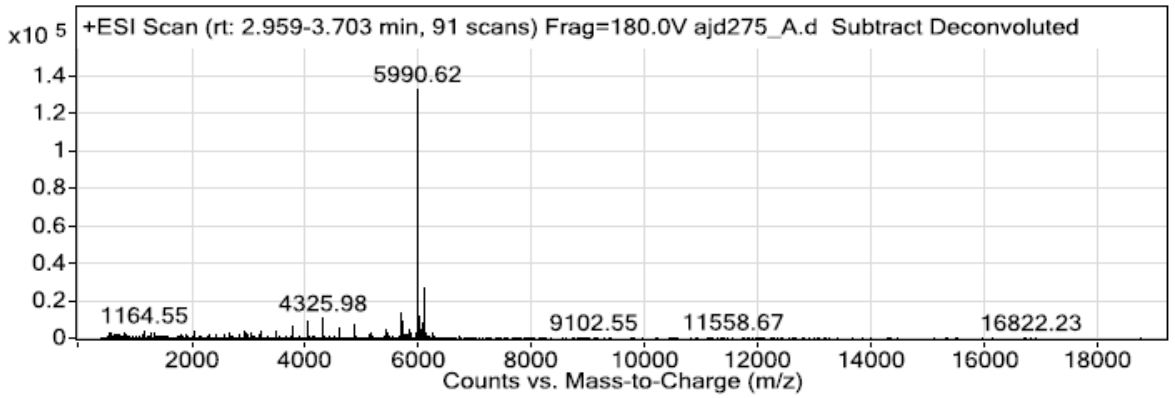
197. K. Wawrzyniak-Turek and C. Hobartner, *Chem. Commun.*, 2014, **50**, 10937-10940.

198. W. A. Velema and E. T. Kool, *J. Am. Chem. Soc.*, 2017, **139**, 5405-5411.

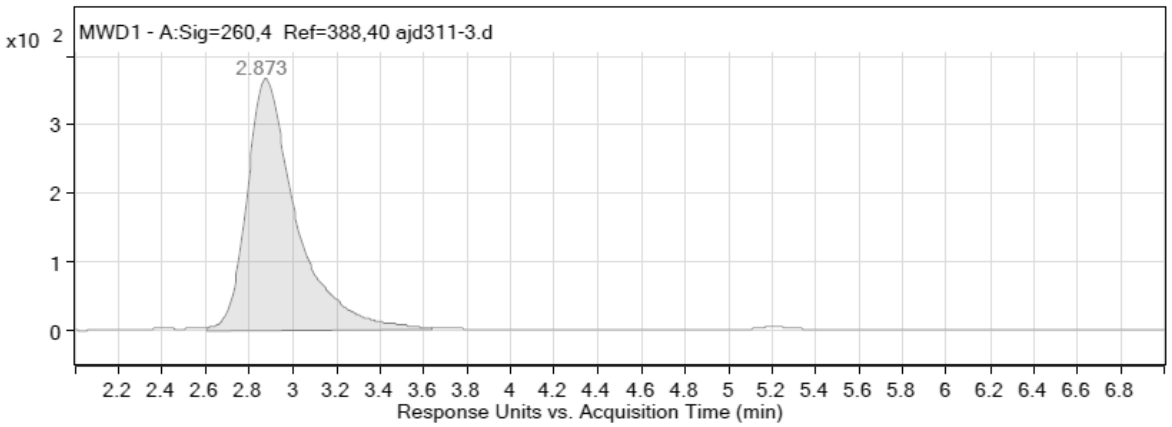
Appendix A LC/MS Spectra of oligonucleotides

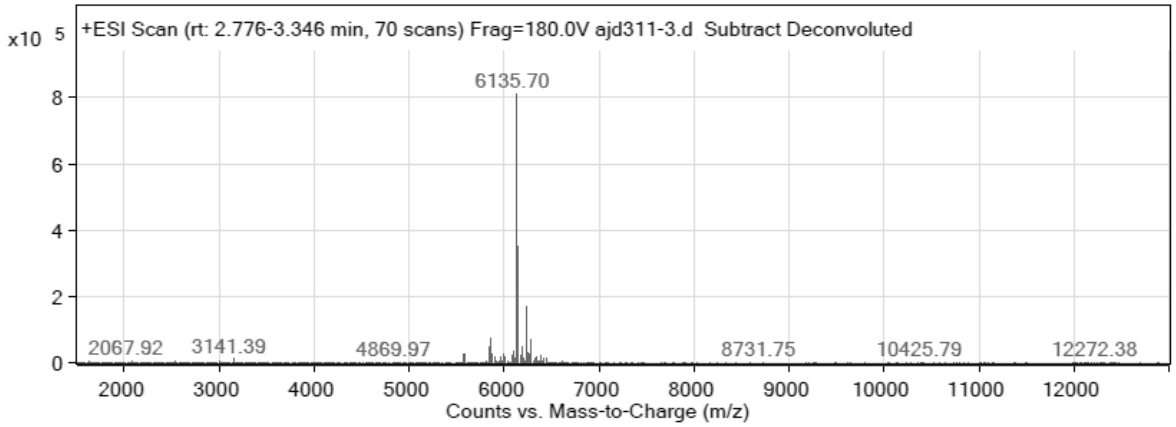
PNA-B-CPP-O



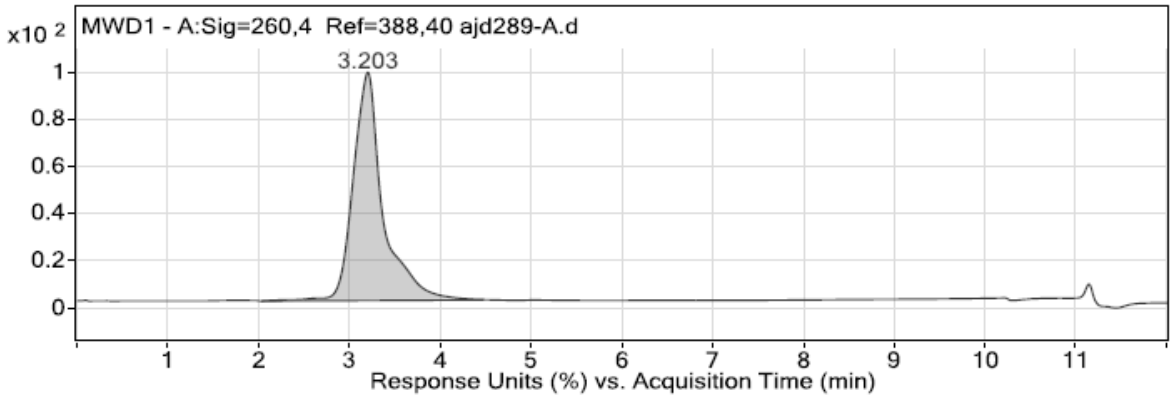


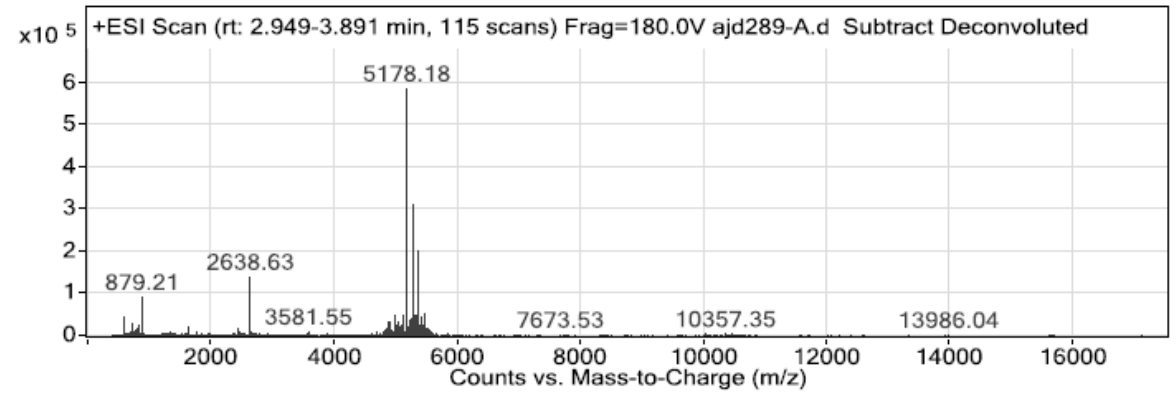
PNA-B-CPP-OO



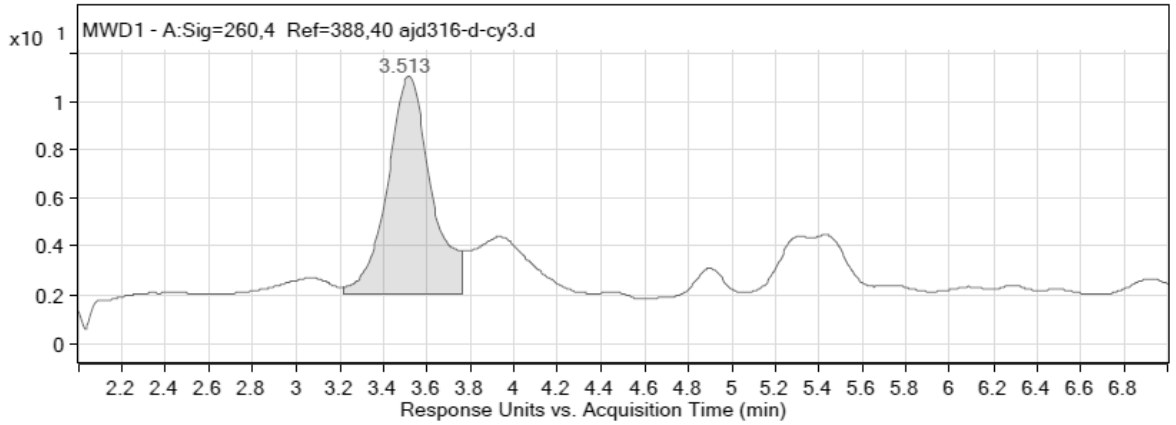


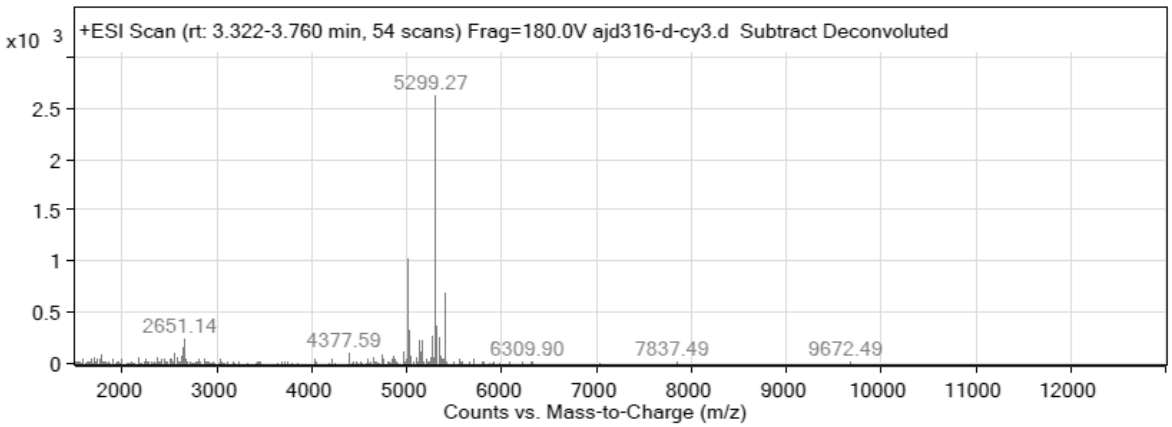
PNA-B-Cy3



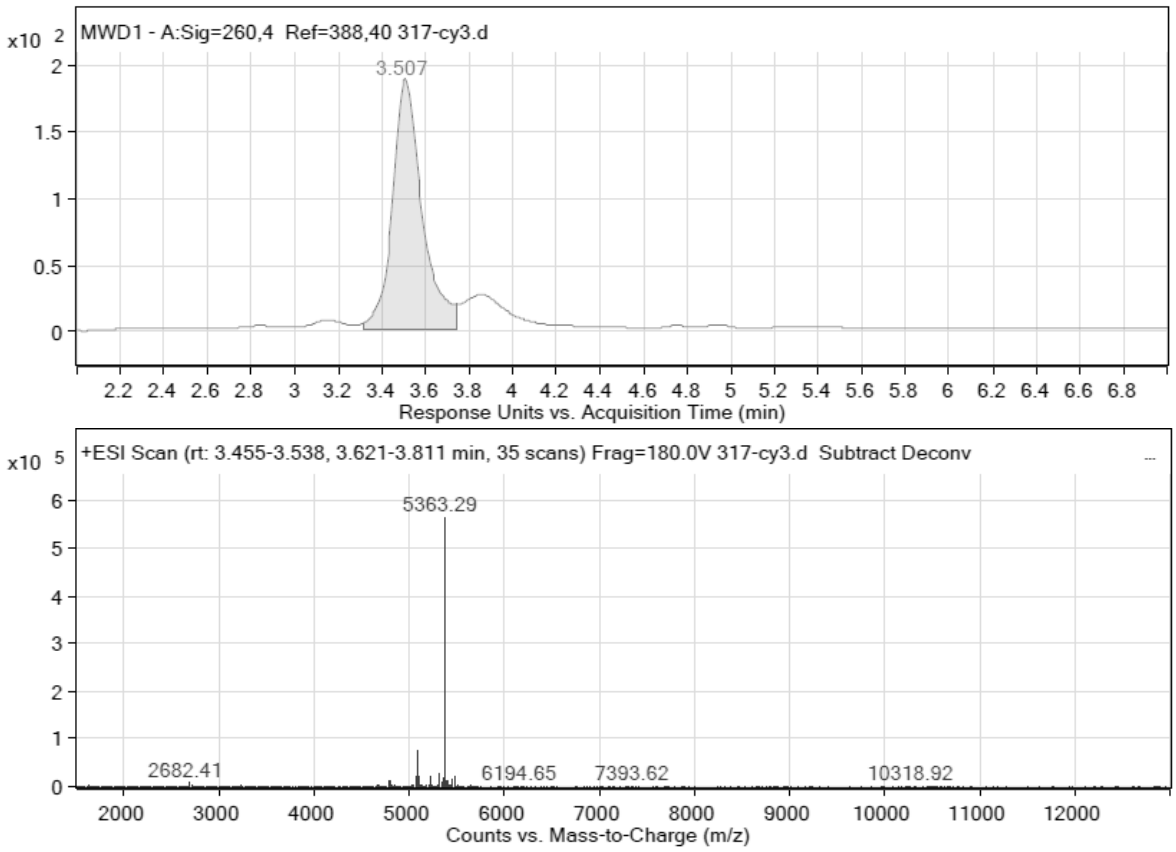


PNA-A-K-Cy3

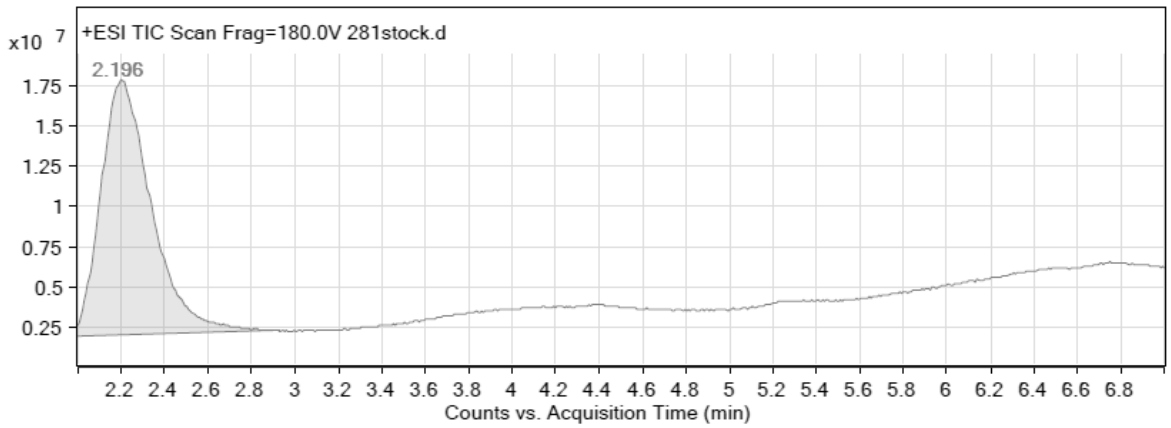


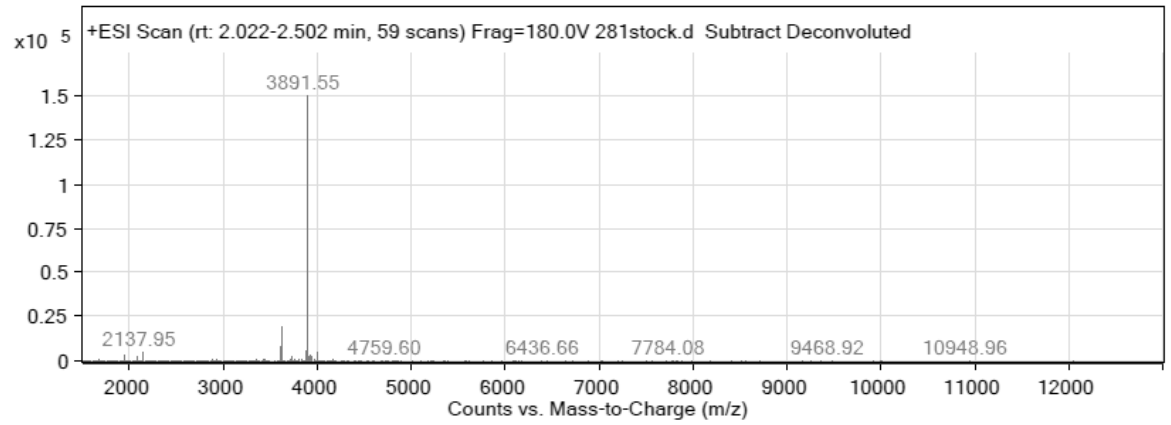

+ESI Scan (rt: 3.322-3.760 min, 54 scans) Frag=180.0V ajd316-d-cy3.d Subtract Deconvoluted
x10 3
5299.27
2651.14
4377.59
6309.90
7837.49
9672.49
Counts vs. Mass-to-Charge (m/z)

PNA-B-K-Cy3

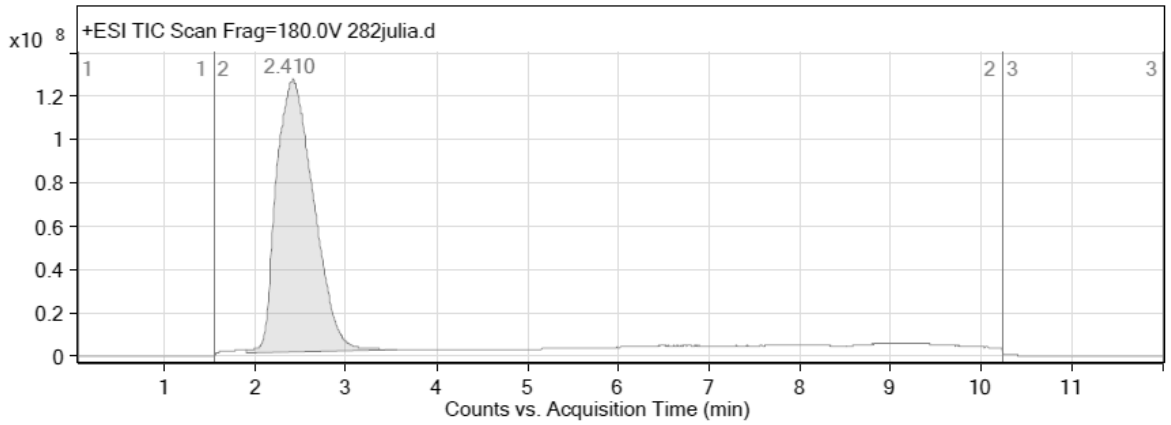

MWD1 - A:Sig=260,4 Ref=388,40 317-cy3.d
x10 2
3.507
Response Units vs. Acquisition Time (min)
+ESI Scan (rt: 3.455-3.538, 3.621-3.811 min, 35 scans) Frag=180.0V 317-cy3.d Subtract Deconv
x10 5
5363.29
2682.41
6194.65
7393.62
10318.92
Counts vs. Mass-to-Charge (m/z)

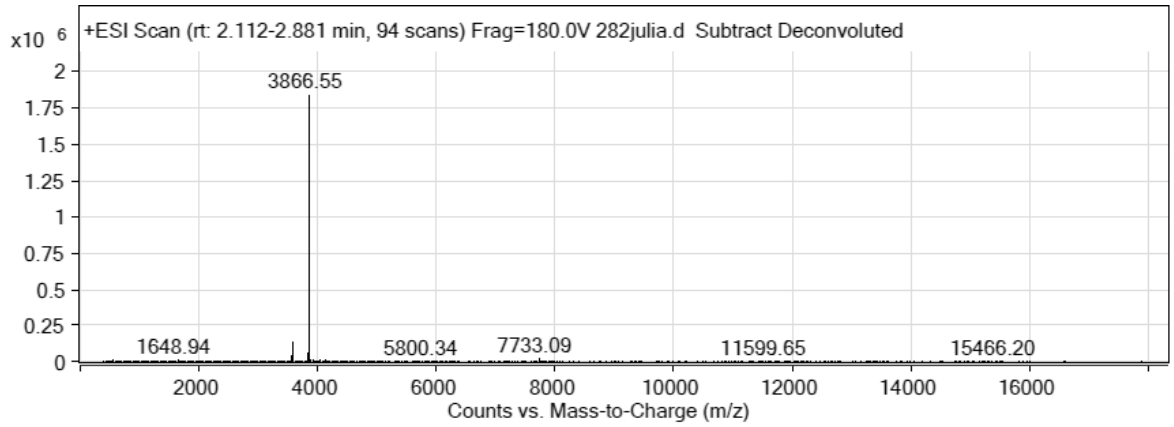
PNA_s_HTT



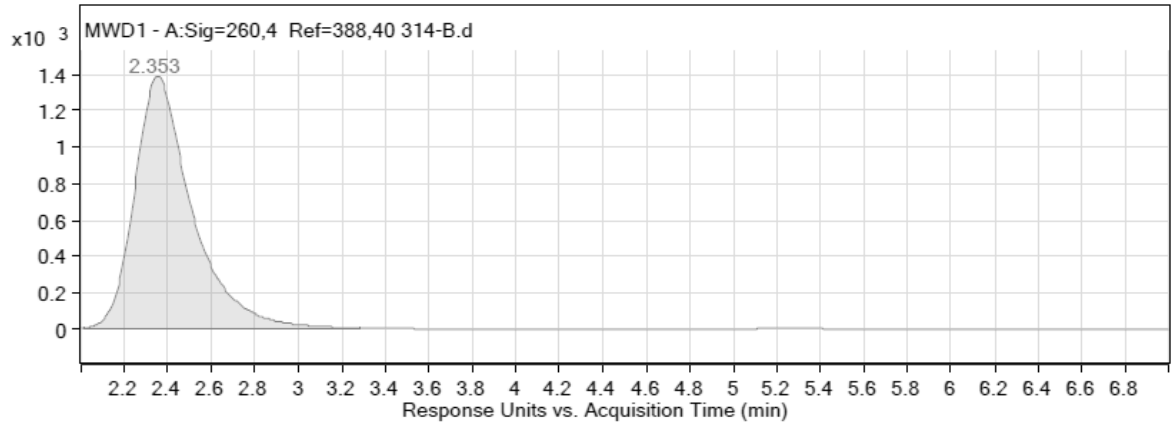


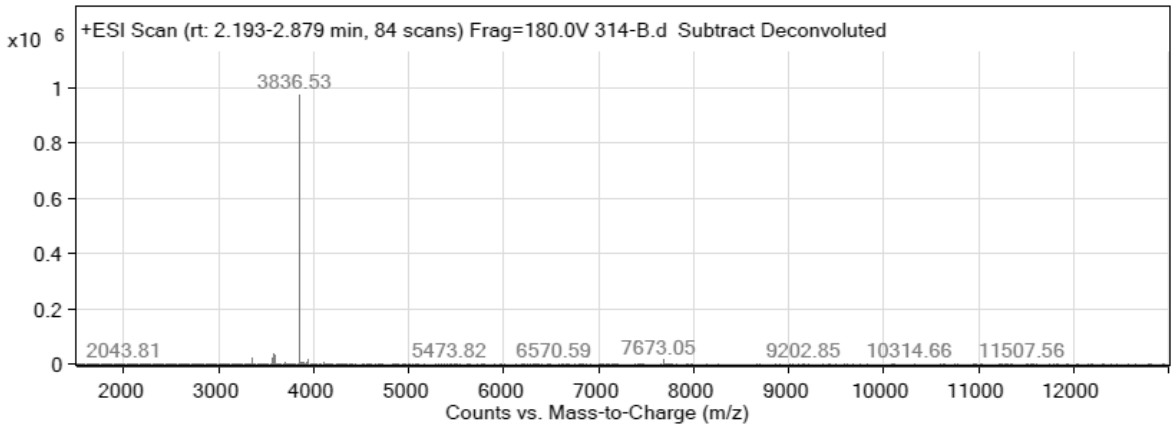
PNA_s_HTT_mm



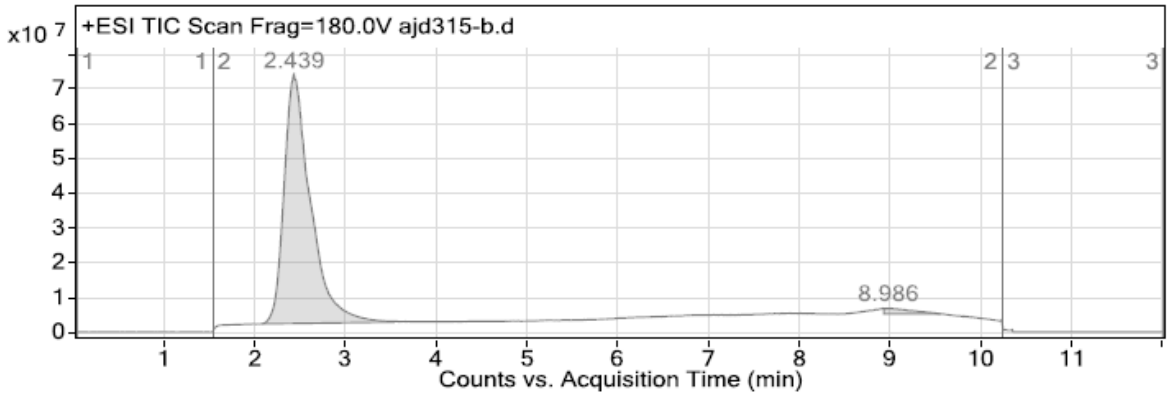

+ESI Scan (rt: 2.112-2.881 min, 94 scans) Frag=180.0V 282julia.d Subtract Deconvoluted
x10 6
3866.55
1648.94
5800.34
7733.09
11599.65
15466.20
Counts vs. Mass-to-Charge (m/z)

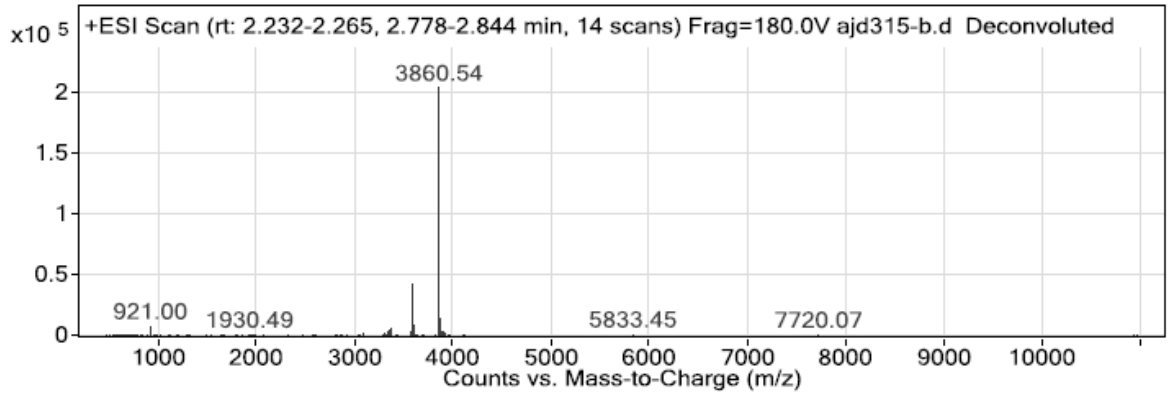
PNA_s_PPIB


MWD1 - A:Sig=260,4 Ref=388,40 314-B.d
x10 3
2.353
Response Units vs. Acquisition Time (min)

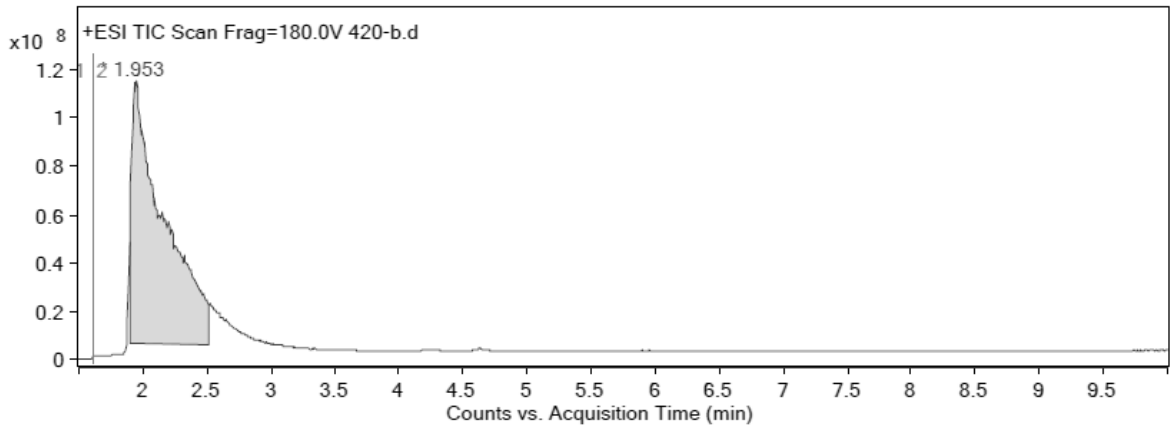

+ESI Scan (rt: 2.193-2.879 min, 84 scans) Frag=180.0V 314-B.d Subtract Deconvoluted
x10 6
3836.53
2043.81
5473.82
6570.59
7673.05
9202.85
10314.66
11507.56
Counts vs. Mass-to-Charge (m/z)

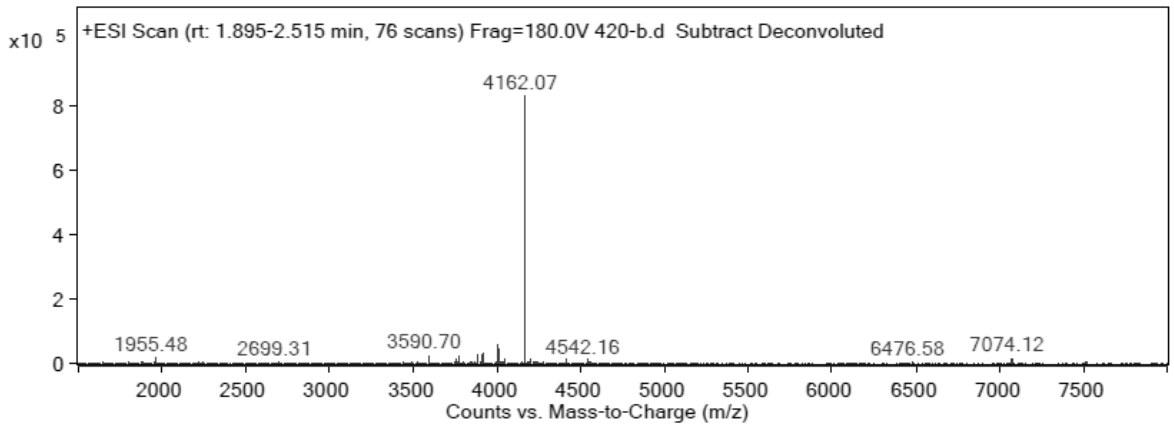
PNA_s_PPIB_mm



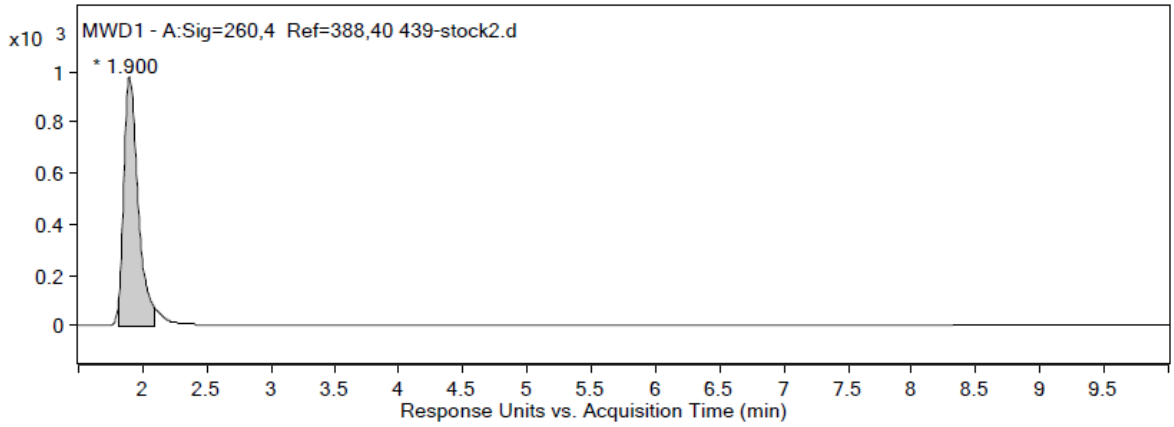


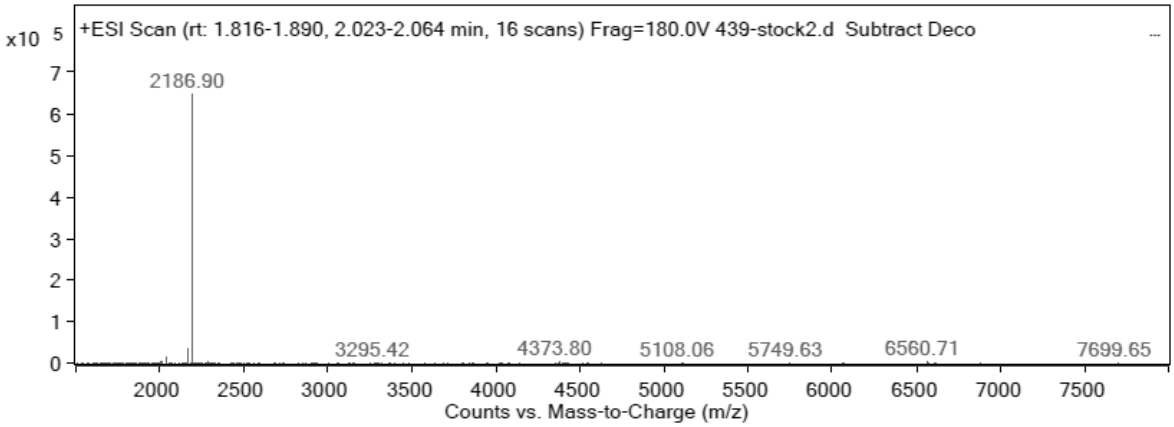
P^{p}



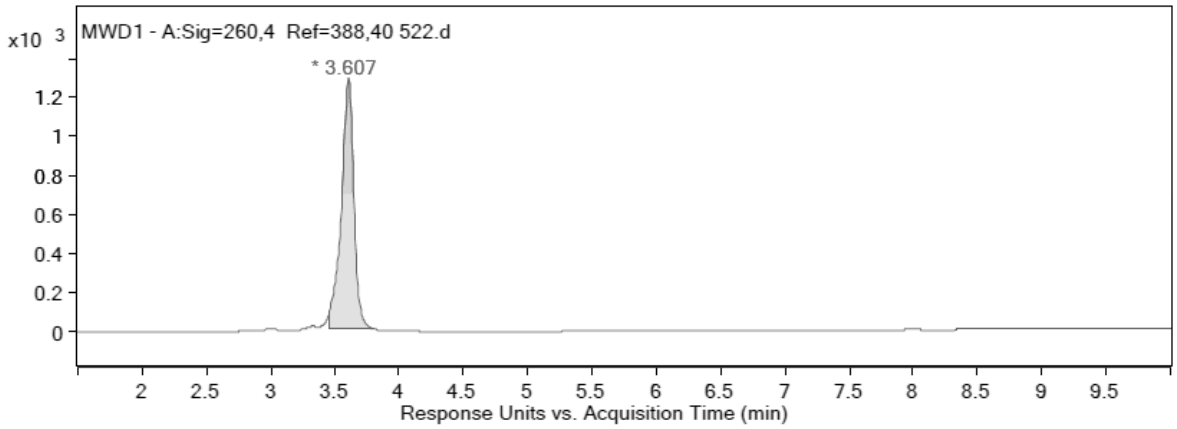

x10 5
+ESI Scan (rt: 1.895-2.515 min, 76 scans) Frag=180.0V 420-b.d Subtract Deconvoluted
4162.07
1955.48
2699.31
3590.70
4542.16
6476.58
7074.12
Counts vs. Mass-to-Charge (m/z)

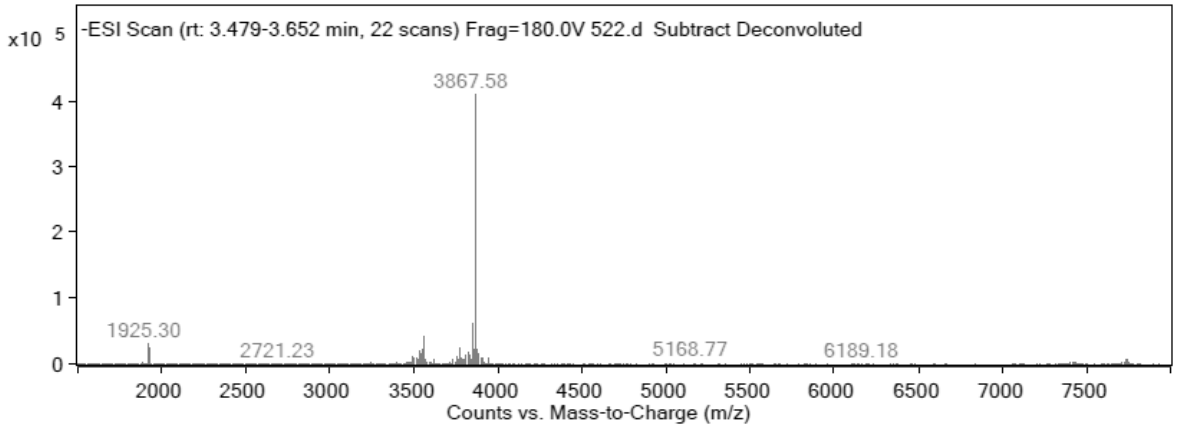
P


x10 3
MWD1 - A:Sig=260,4 Ref=388,40 439-stock2.d
* 1.900
Response Units vs. Acquisition Time (min)

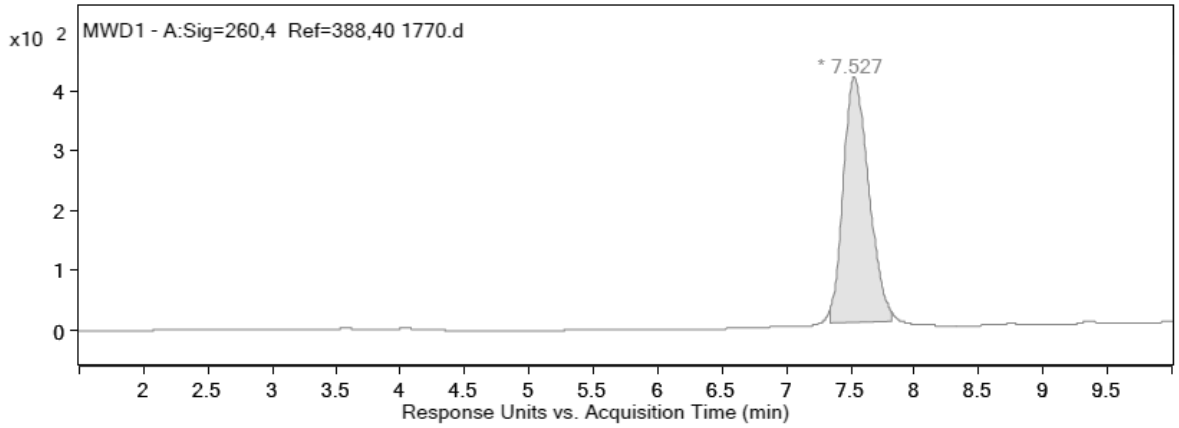

x10 5
+ESI Scan (rt: 1.816-1.890, 2.023-2.064 min, 16 scans) Frag=180.0V 439-stock2.d Subtract Deco
2186.90
3295.42
4373.80
5108.06
5749.63
6560.71
7699.65
Counts vs. Mass-to-Charge (m/z)

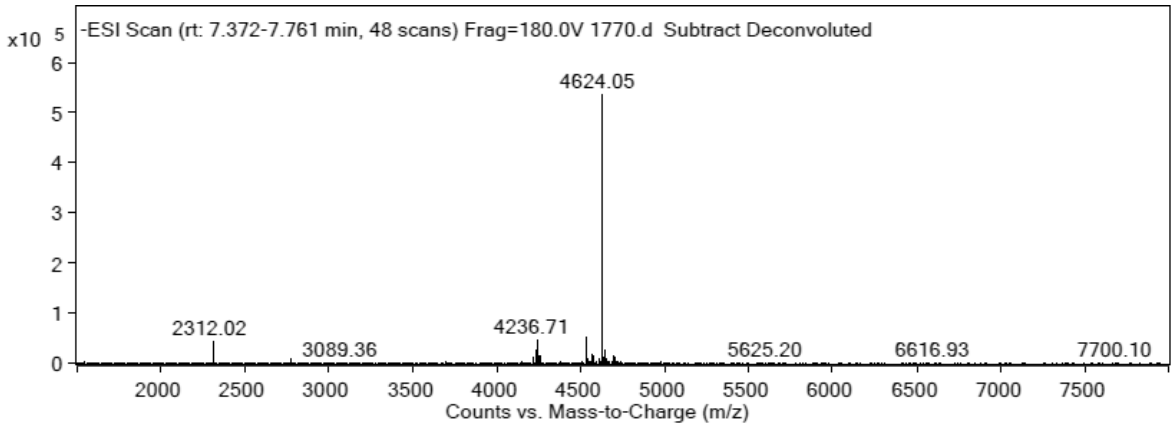
S


x10 3
MWD1 - A:Sig=260,4 Ref=388,40 522.d
* 3.607
1.2
1
0.8
0.6
0.4
0.2
0
2 2.5 3 3.5 4 4.5 5 5.5 6 6.5 7 7.5 8 8.5 9 9.5
Response Units vs. Acquisition Time (min)

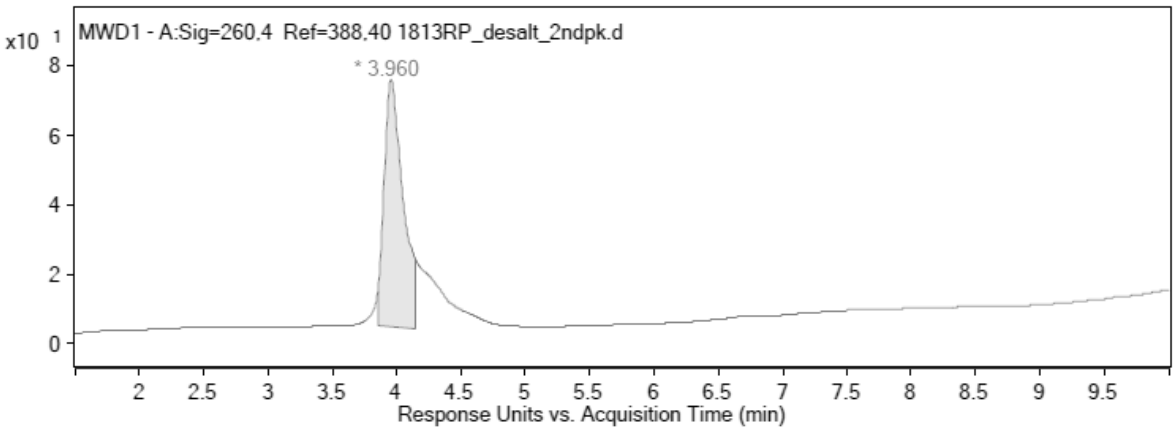

x10 5
-ESI Scan (rt: 3.479-3.652 min, 22 scans) Frag=180.0V 522.d Subtract Deconvoluted
3867.58
4
3
2
1
1925.30
2721.23
5168.77
6189.18
0
2000 2500 3000 3500 4000 4500 5000 5500 6000 6500 7000 7500
Counts vs. Mass-to-Charge (m/z)

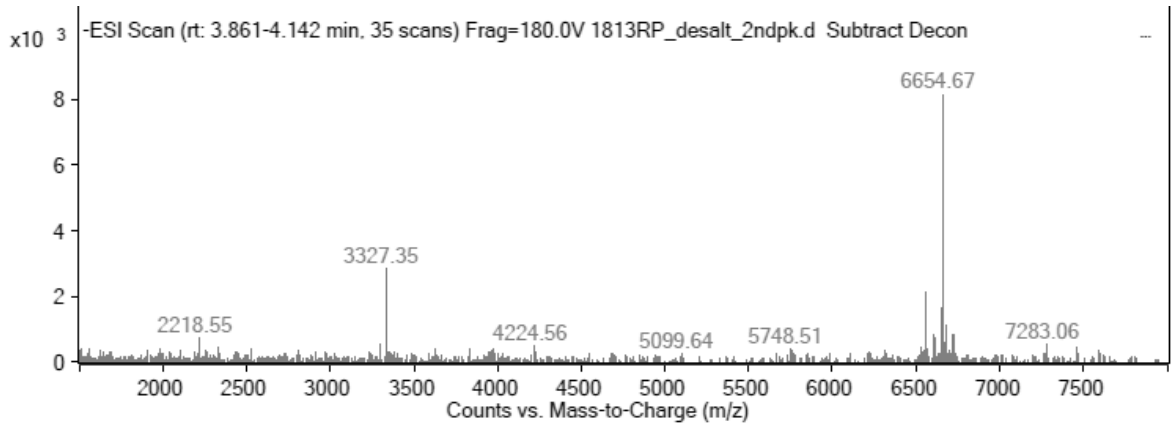
S^{c}


x10 2
MWD1 - A:Sig=260,4 Ref=388,40 1770.d
* 7.527
4
3
2
1
0
2 2.5 3 3.5 4 4.5 5 5.5 6 6.5 7 7.5 8 8.5 9 9.5
Response Units vs. Acquisition Time (min)

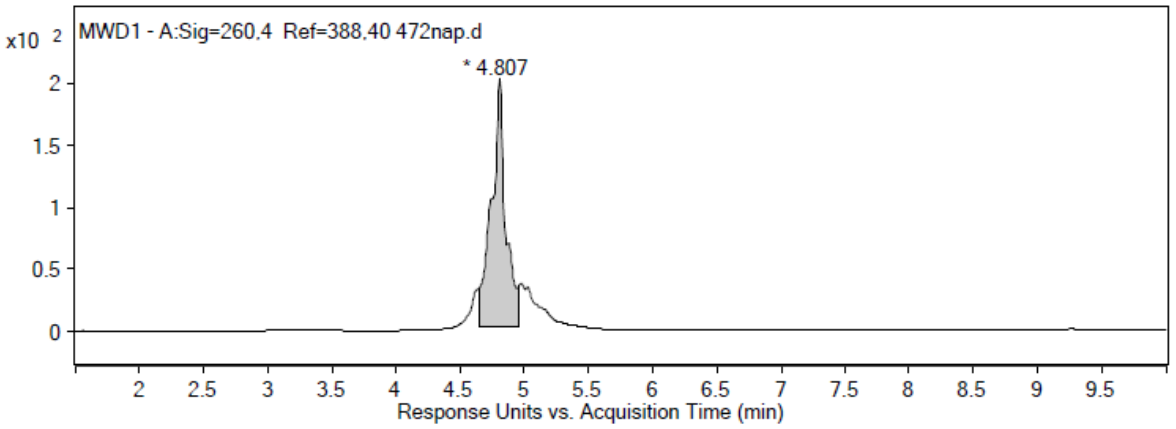

-ESI Scan (rt: 7.372-7.761 min, 48 scans) Frag=180.0V 1770.d Subtract Deconvoluted
x10 5
4624.05
2312.02
3089.36
4236.71
5625.20
6616.93
7700.10
Counts vs. Mass-to-Charge (m/z)

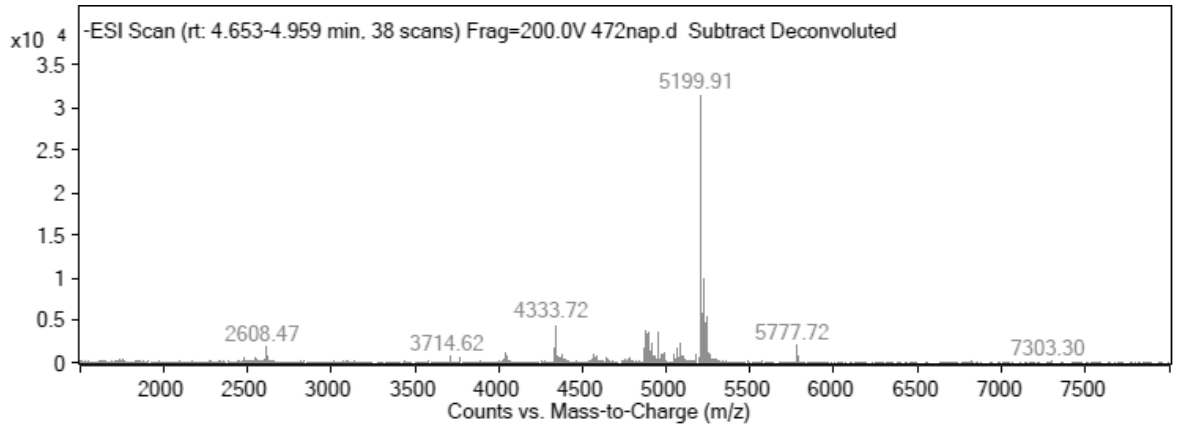
S[15c]


MWD1 - A:Sig=260,4 Ref=388,40 1813RP_desalt_2ndpk.d
x10 1
* 3.960
Response Units vs. Acquisition Time (min)

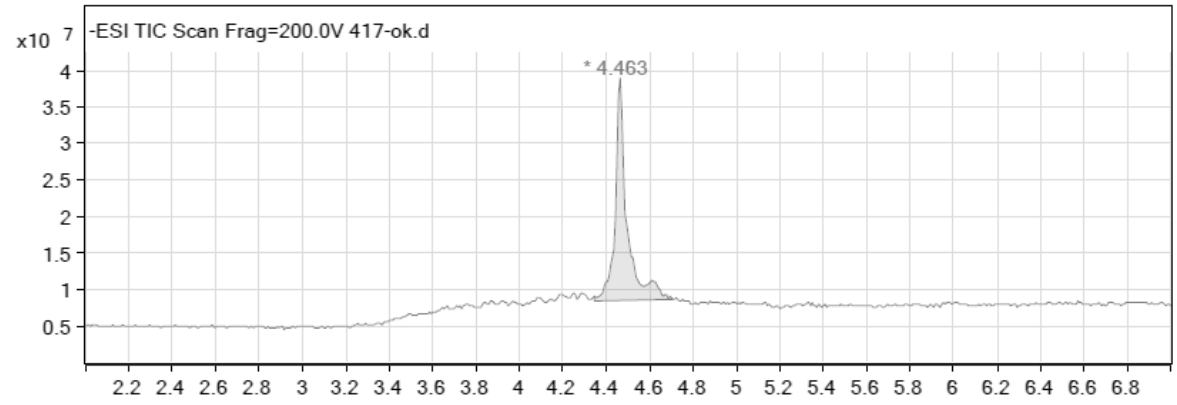

-ESI Scan (rt: 3.861-4.142 min, 35 scans) Frag=180.0V 1813RP_desalt_2ndpk.d Subtract Decon
x10 3
6654.67
3327.35
2218.55
4224.56
5099.64
5748.51
7283.06
Counts vs. Mass-to-Charge (m/z)

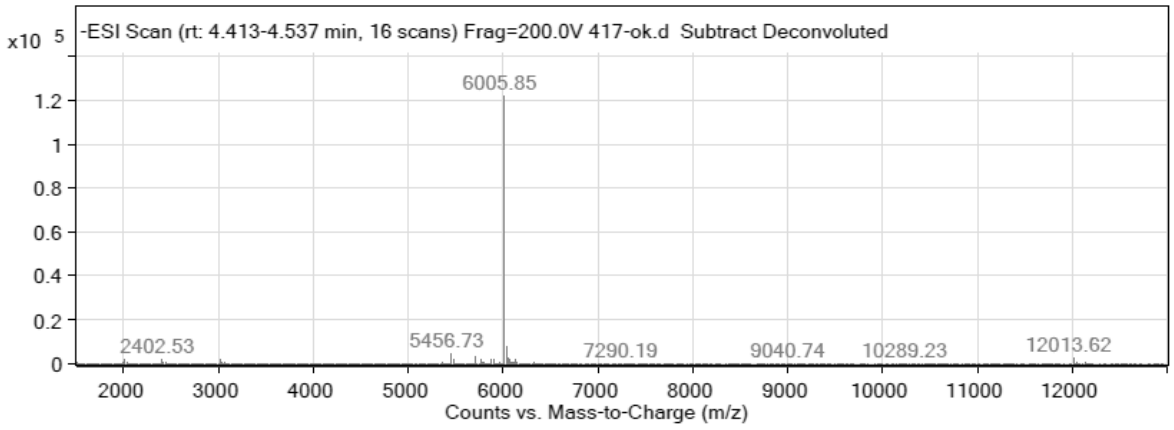
PO DNA



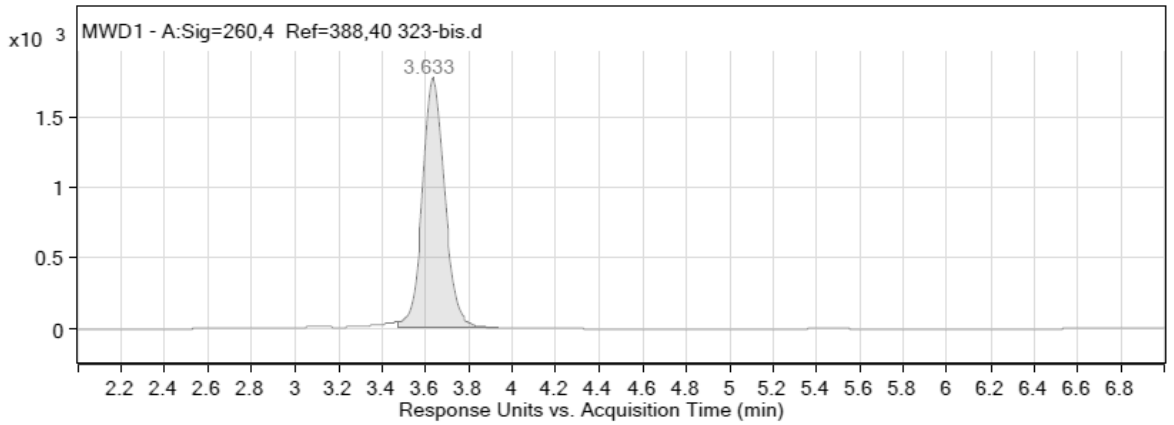


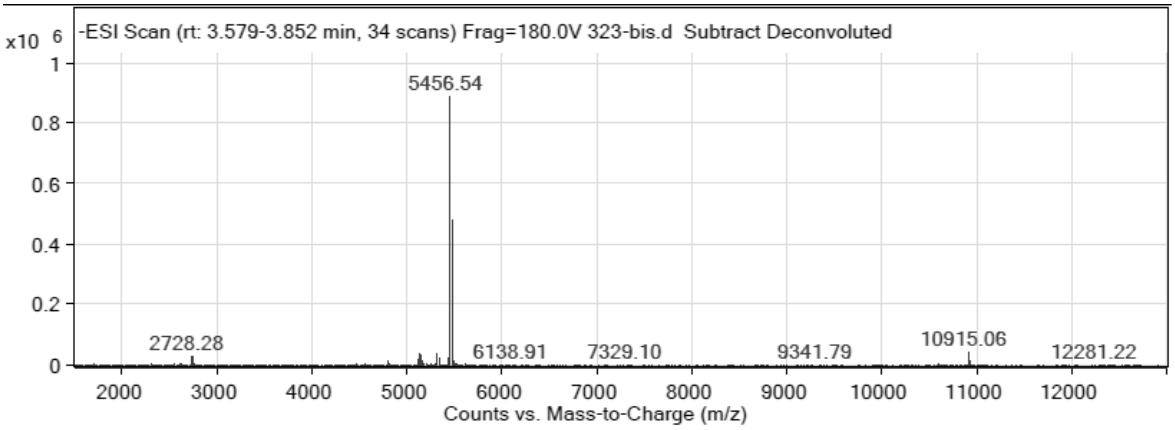
PO RNA 417



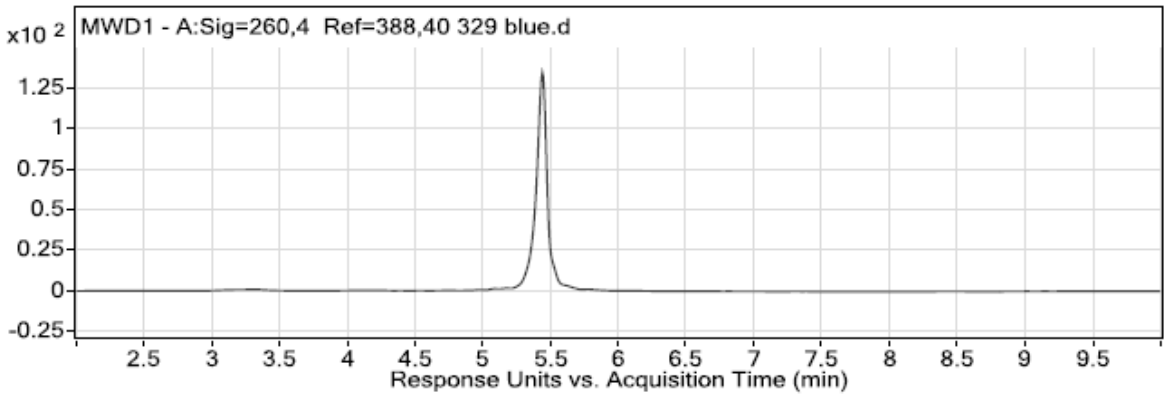

-ESI Scan (rt: 4.413-4.537 min, 16 scans) Frag=200.0V 417-ok.d Subtract Deconvoluted
x10 5
6005.85
2402.53
5456.73
7290.19
9040.74
10289.23
12013.62
Counts vs. Mass-to-Charge (m/z)

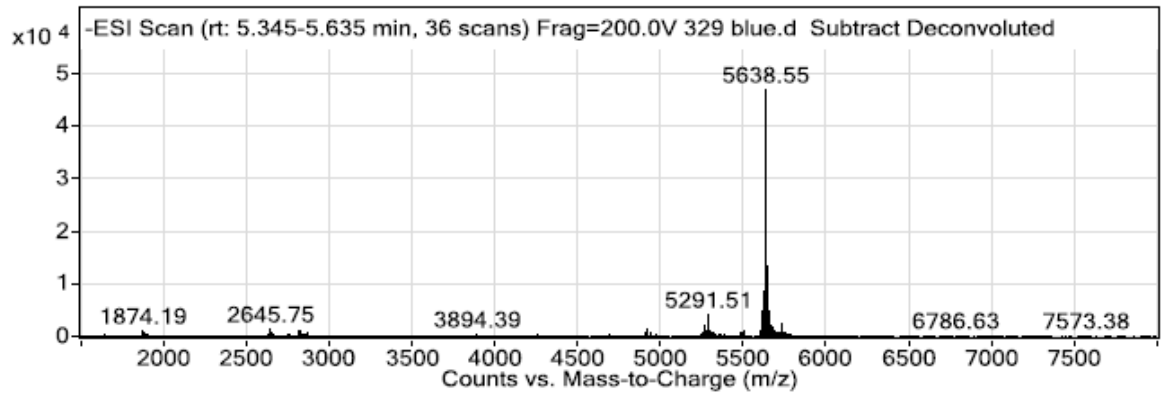
PS DNA


MWD1 - A:Sig=260,4 Ref=388,40 323-bis.d
x10 3
3.633
Response Units vs. Acquisition Time (min)

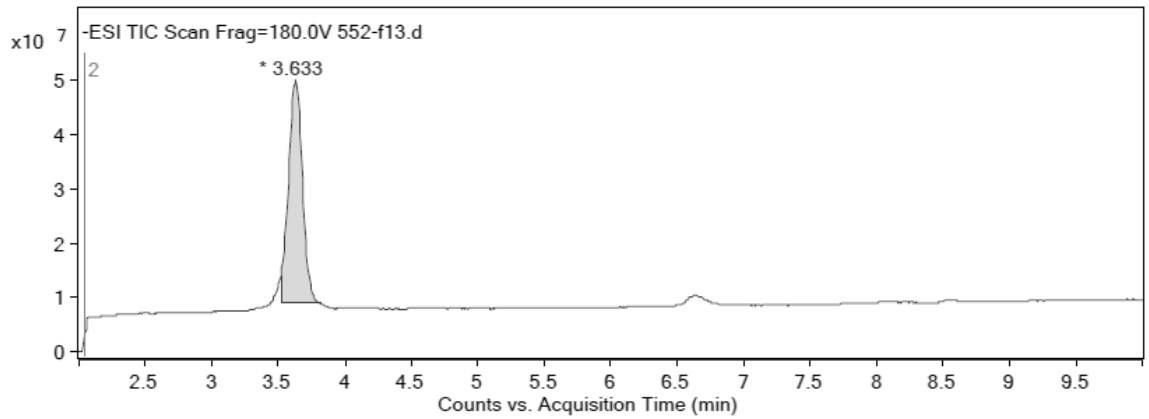

-ESI Scan (rt: 3.579-3.852 min, 34 scans) Frag=180.0V 323-bis.d Subtract Deconvoluted
x10 6
5456.54
2728.28
6138.91
7329.10
9341.79
10915.06
12281.22
Counts vs. Mass-to-Charge (m/z)

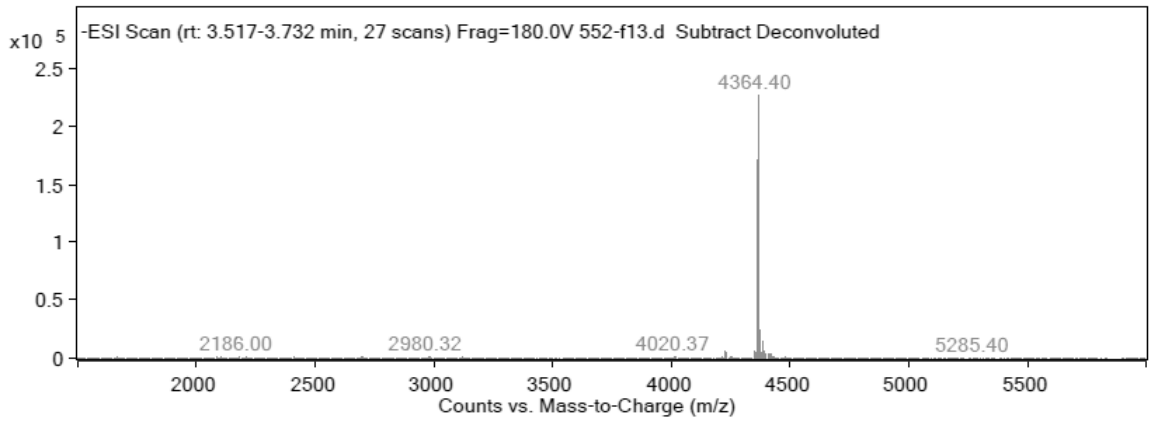
LNA Gap



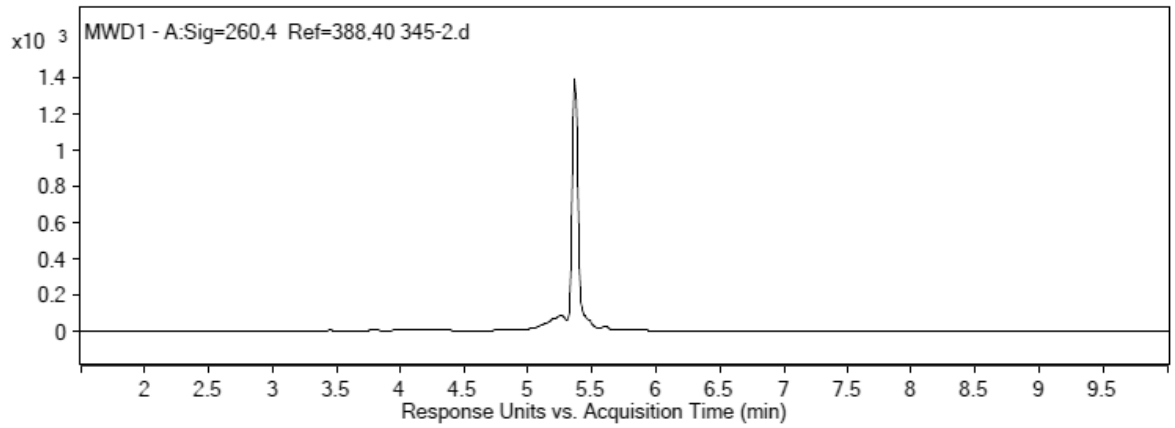


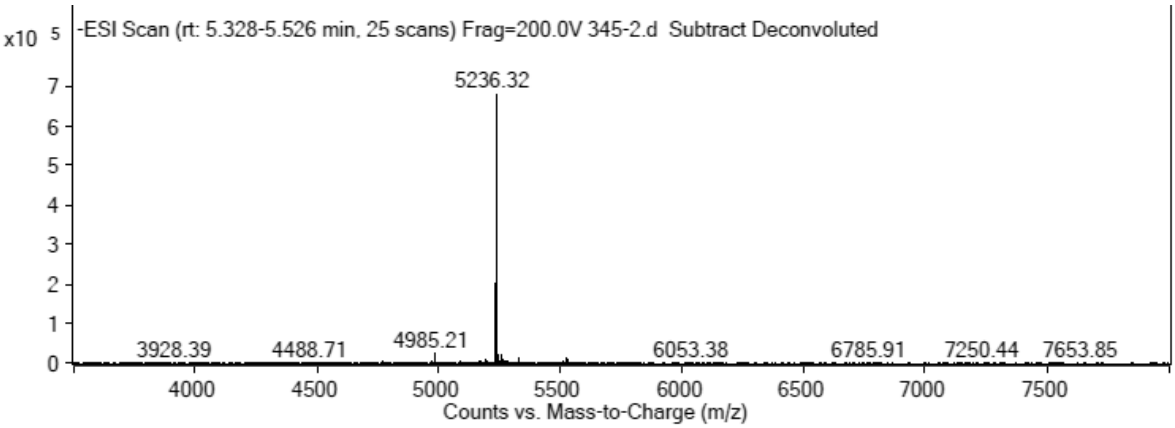
LNA Gap[5]



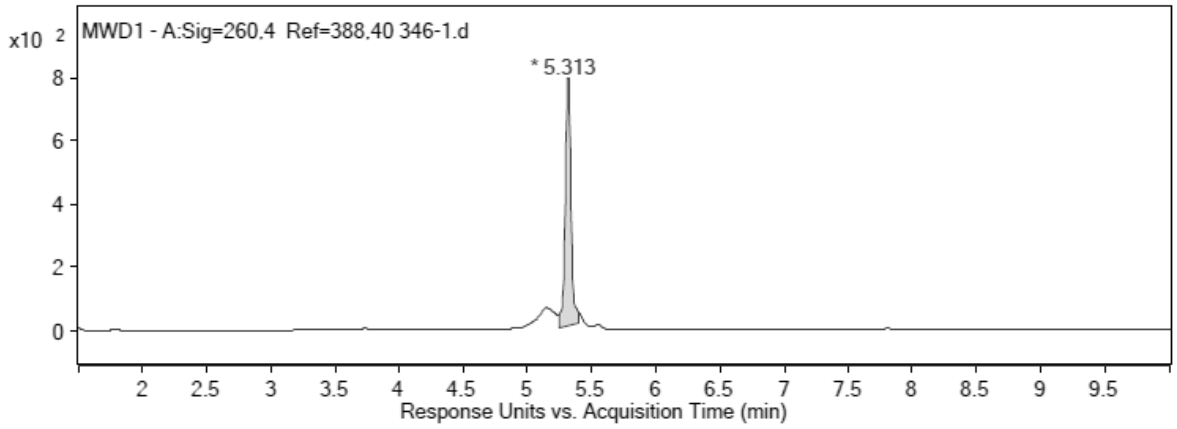

-ESI Scan (rt: 3.517-3.732 min, 27 scans) Frag=180.0V 552-f13.d Subtract Deconvoluted
x10 5
2.5
2
1.5
1
0.5
0
4364.40
2186.00
2980.32
4020.37
5285.40
2000
2500
3000
3500
4000
4500
5000
5500
Counts vs. Mass-to-Charge (m/z)

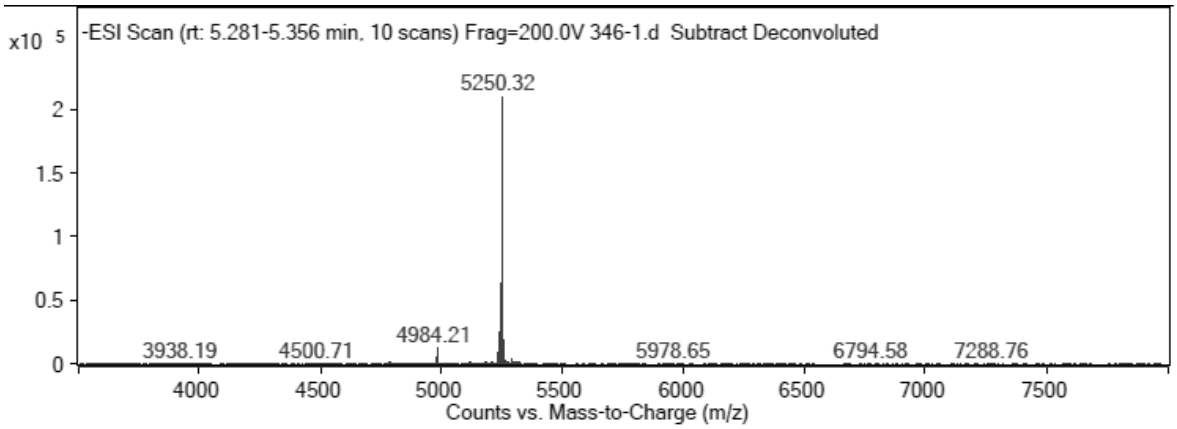
eDNA


MWD1 - A:Sig=260,4 Ref=388,40 345-2.d
x10 3
1.4
1.2
1
0.8
0.6
0.4
0.2
0
2
2.5
3
3.5
4
4.5
5
5.5
6
6.5
7
7.5
8
8.5
9
9.5
Response Units vs. Acquisition Time (min)

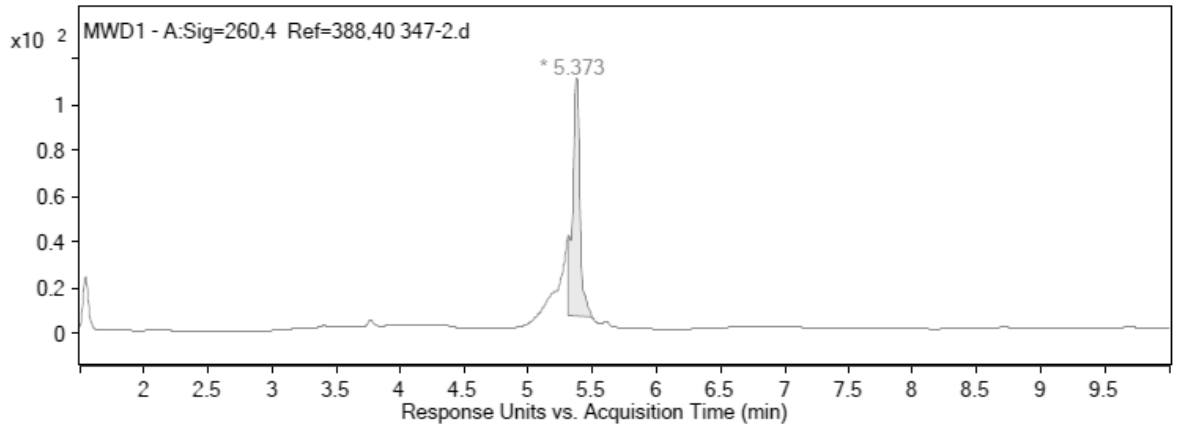

-ESI Scan (rt: 5.328-5.526 min, 25 scans) Frag=200.0V 345-2.d Subtract Deconvoluted
x10 5
7
6
5
4
3
2
1
0
5236.32
3928.39
4488.71
4985.21
6053.38
6785.91
7250.44
7653.85
4000
4500
5000
5500
6000
6500
7000
7500
Counts vs. Mass-to-Charge (m/z)

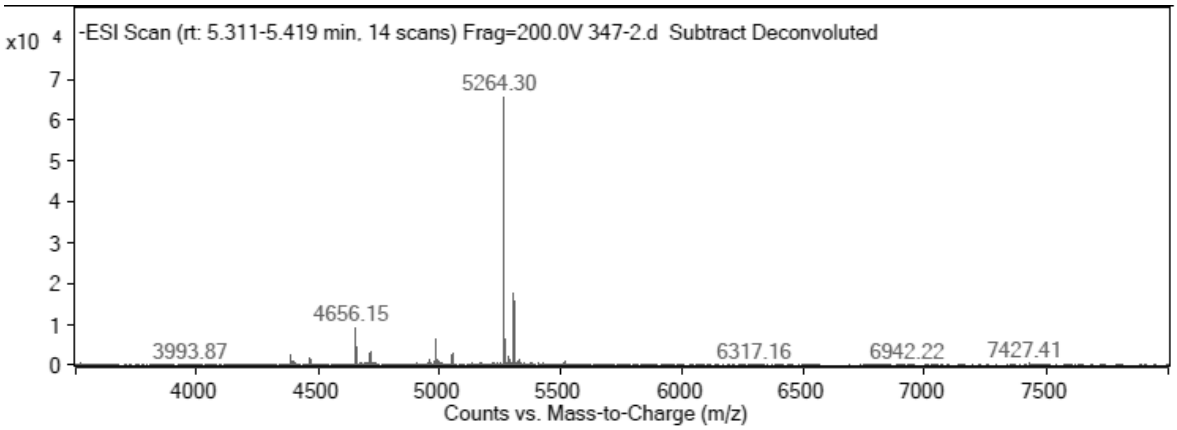
pDNA


x10 2
MWD1 - A:Sig=260,4 Ref=388,40 346-1.d
* 5.313
Response Units vs. Acquisition Time (min)

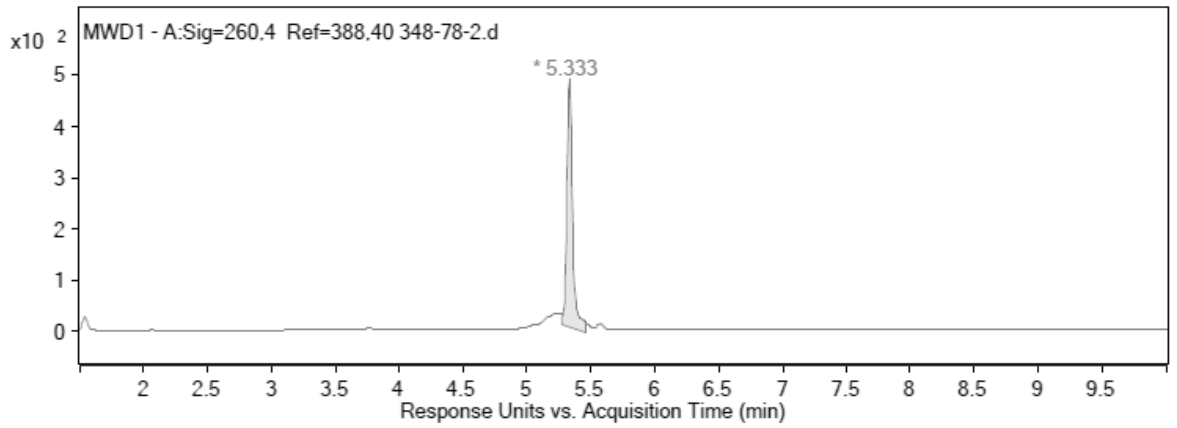

x10 5
-ESI Scan (rt: 5.281-5.356 min, 10 scans) Frag=200.0V 346-1.d Subtract Deconvoluted
5250.32
3938.19
4500.71
4984.21
5978.65
6794.58
7288.76
Counts vs. Mass-to-Charge (m/z)

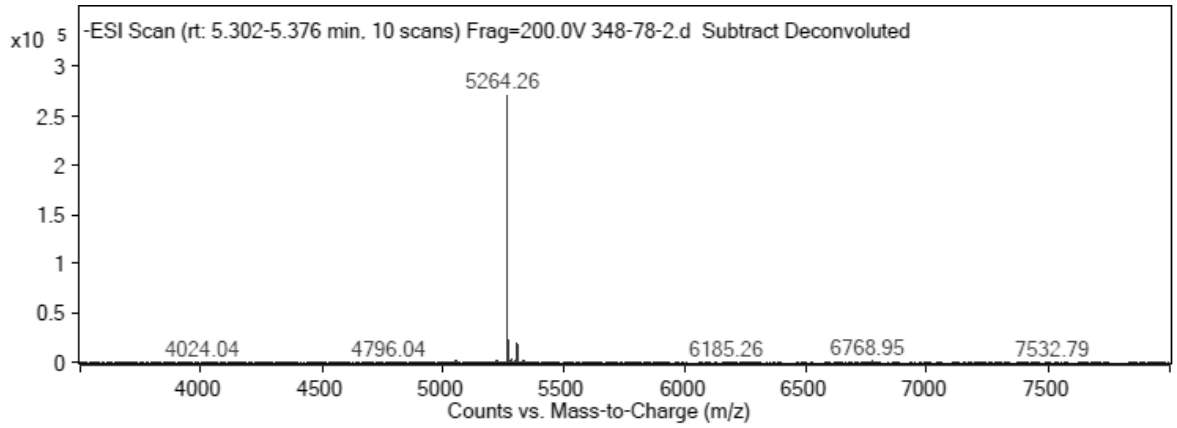
bDNA


x10 2
MWD1 - A:Sig=260,4 Ref=388,40 347-2.d
* 5.373
Response Units vs. Acquisition Time (min)

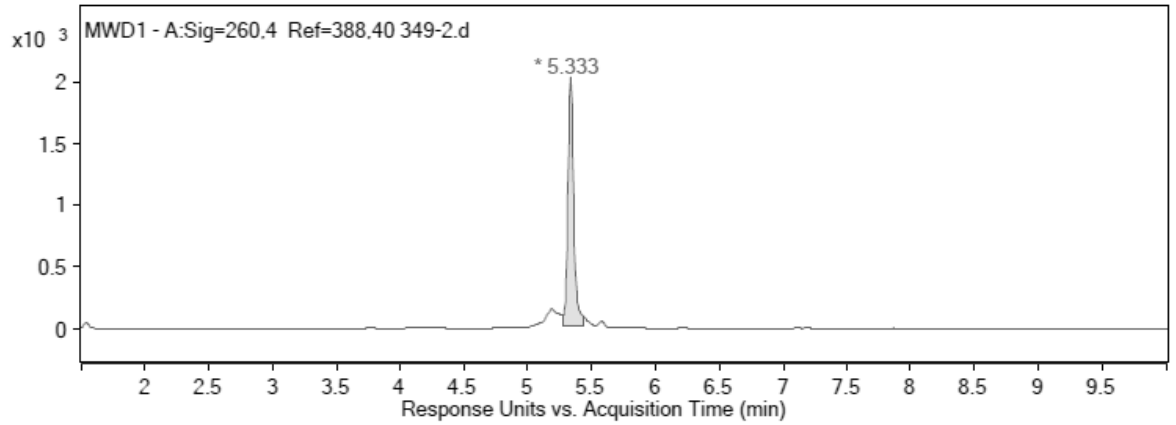

-ESI Scan (rt: 5.311-5.419 min, 14 scans) Frag=200.0V 347-2.d Subtract Deconvoluted
x10 4
5264.30
4656.15
3993.87
6317.16
6942.22
7427.41
Counts vs. Mass-to-Charge (m/z)

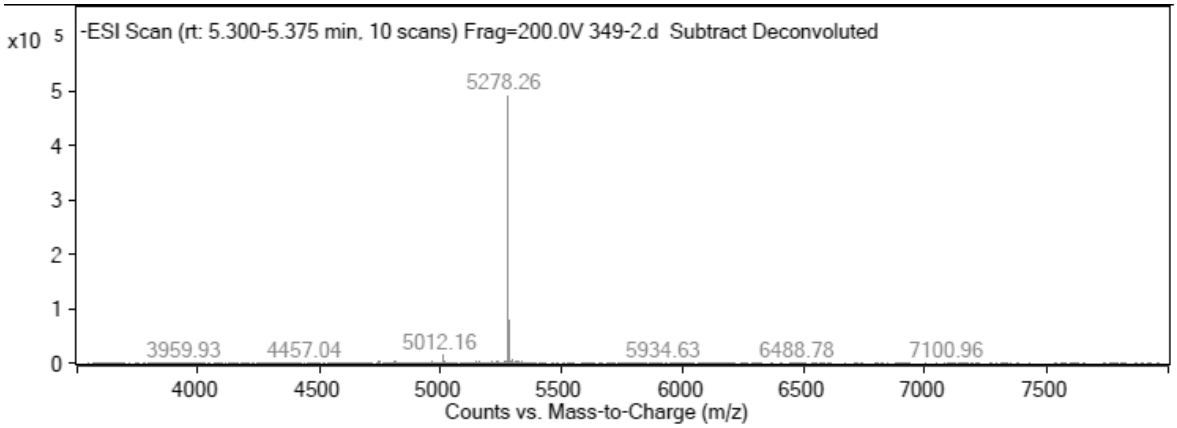
eLNA


MWD1 - A:Sig=260,4 Ref=388,40 348-78-2.d
x10 2
* 5.333
Response Units vs. Acquisition Time (min)

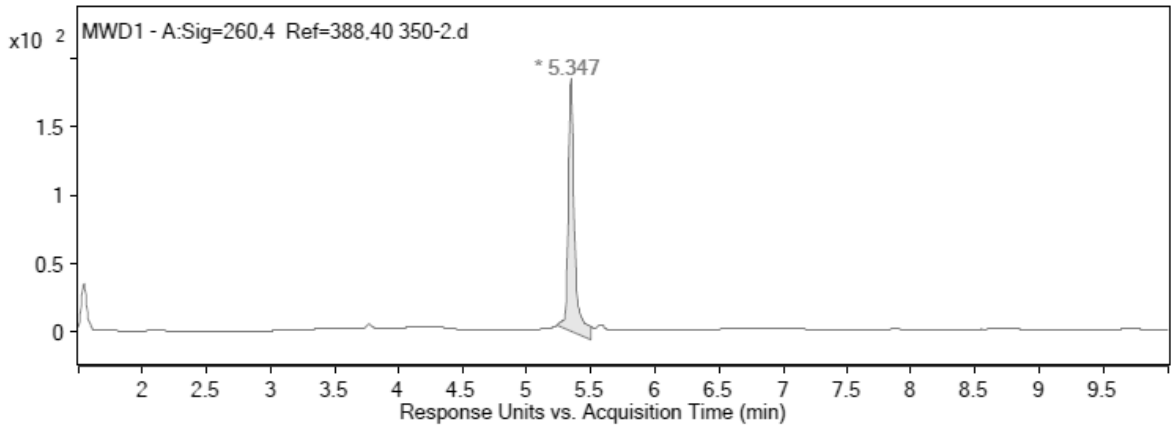

-ESI Scan (rt: 5.302-5.376 min, 10 scans) Frag=200.0V 348-78-2.d Subtract Deconvoluted
x10 5
5264.26
4024.04
4796.04
6185.26
6768.95
7532.79
Counts vs. Mass-to-Charge (m/z)

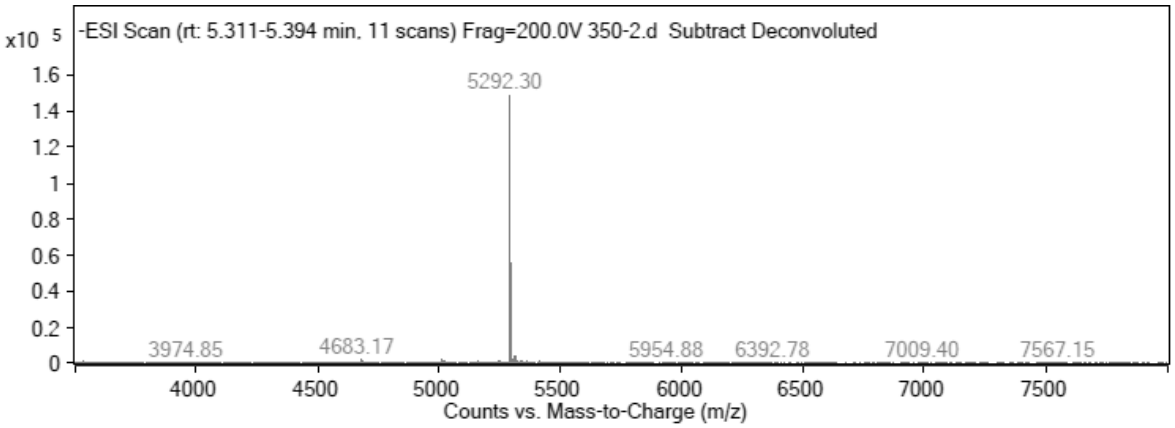
pLNA


MWD1 - A:Sig=260,4 Ref=388,40 349-2.d
x10 3
* 5.333
Response Units vs. Acquisition Time (min)

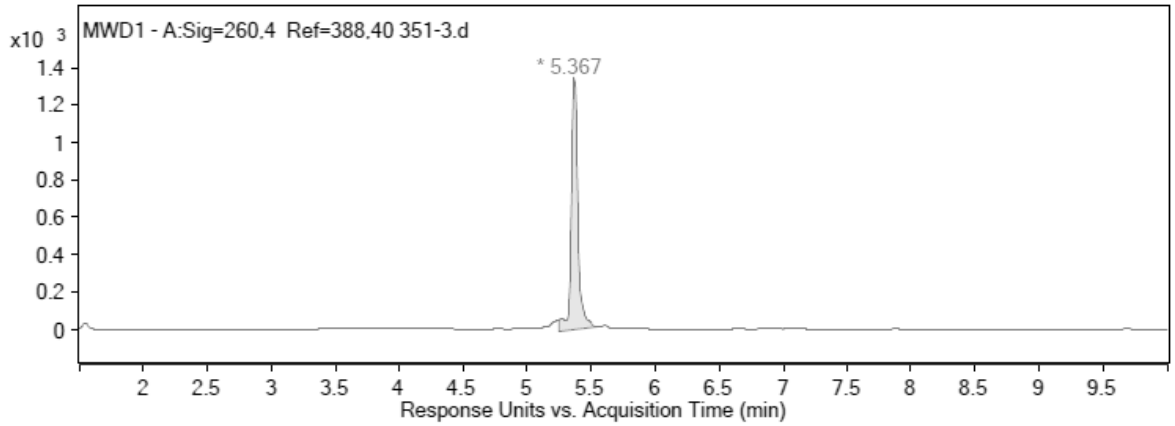

-ESI Scan (rt: 5.300-5.375 min, 10 scans) Frag=200.0V 349-2.d Subtract Deconvoluted
x10 5
5278.26
3959.93
4457.04
5012.16
5934.63
6488.78
7100.96
Counts vs. Mass-to-Charge (m/z)

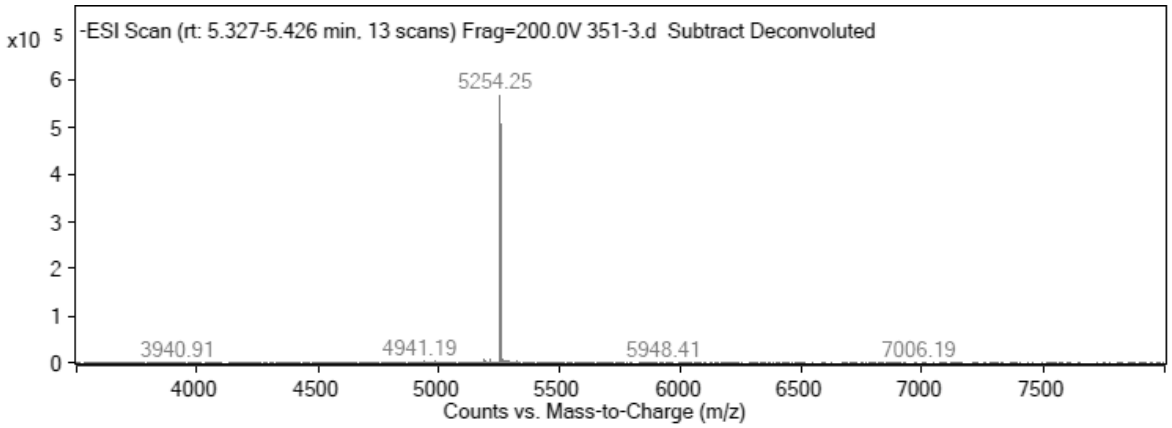
bLNA


MWD1 - A:Sig=260,4 Ref=388,40 350-2.d
x10 2
* 5.347
Response Units vs. Acquisition Time (min)

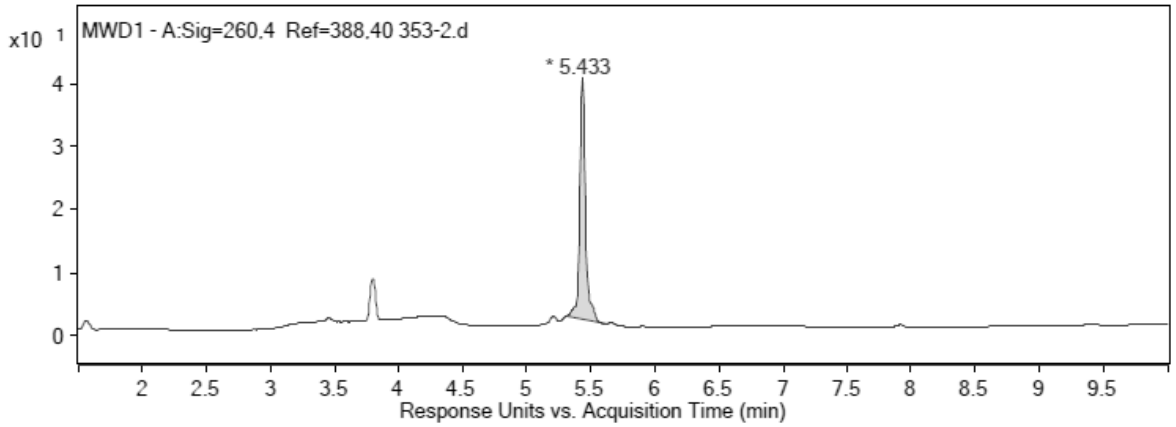

-ESI Scan (rt: 5.311-5.394 min, 11 scans) Frag=200.0V 350-2.d Subtract Deconvoluted
x10 5
5292.30
3974.85
4683.17
5954.88
6392.78
7009.40
7567.15
Counts vs. Mass-to-Charge (m/z)

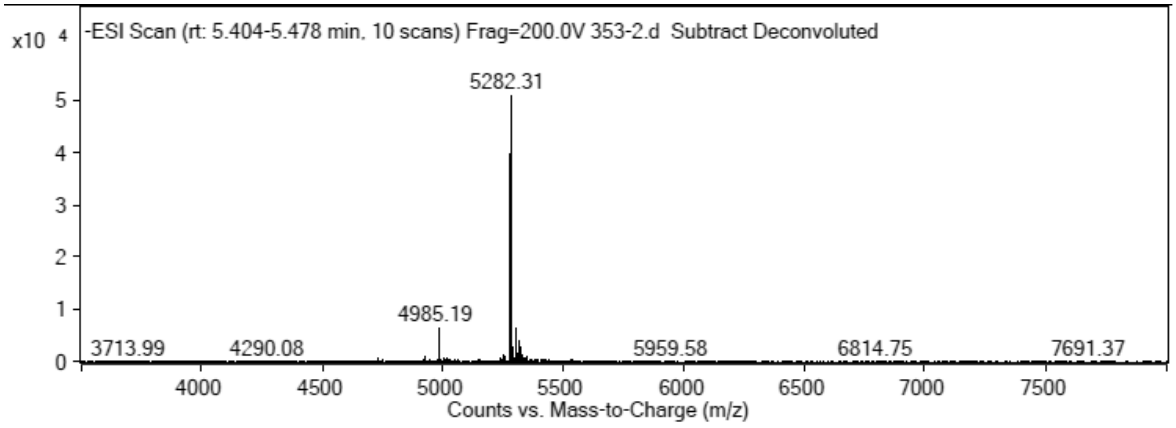
eFANA


MWD1 - A:Sig=260,4 Ref=388,40 351-3.d
x10 3
* 5.367
Response Units vs. Acquisition Time (min)

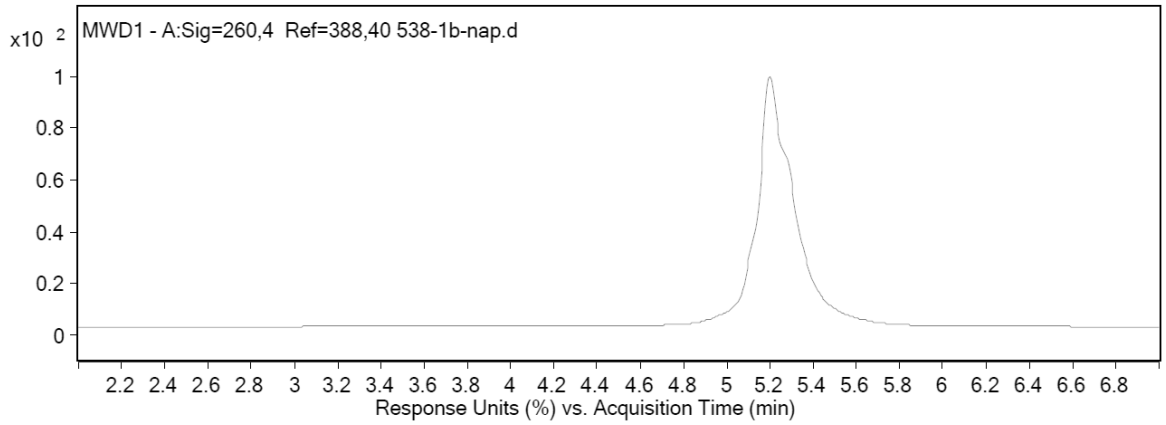

-ESI Scan (rt: 5.327-5.426 min, 13 scans) Frag=200.0V 351-3.d Subtract Deconvoluted
x10 5
5254.25
3940.91
4941.19
5948.41
7006.19
Counts vs. Mass-to-Charge (m/z)

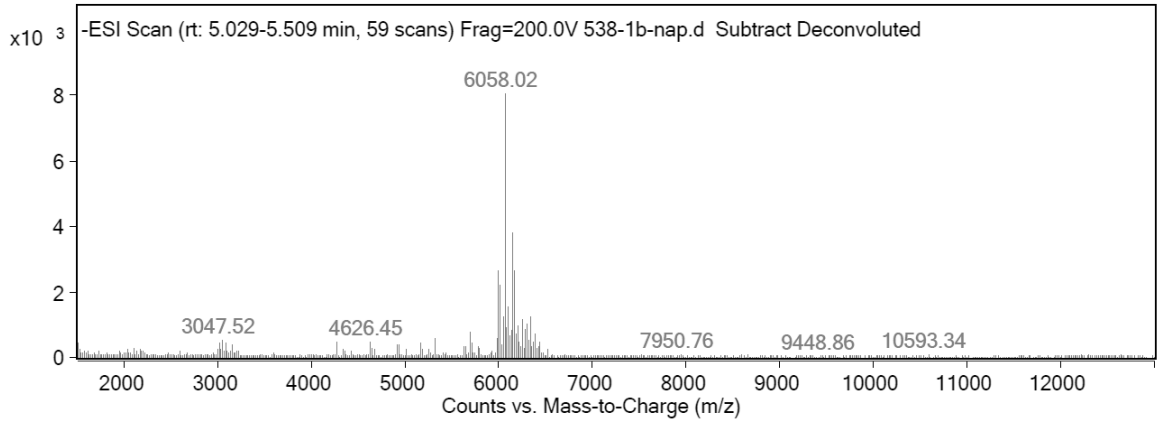
bFANA



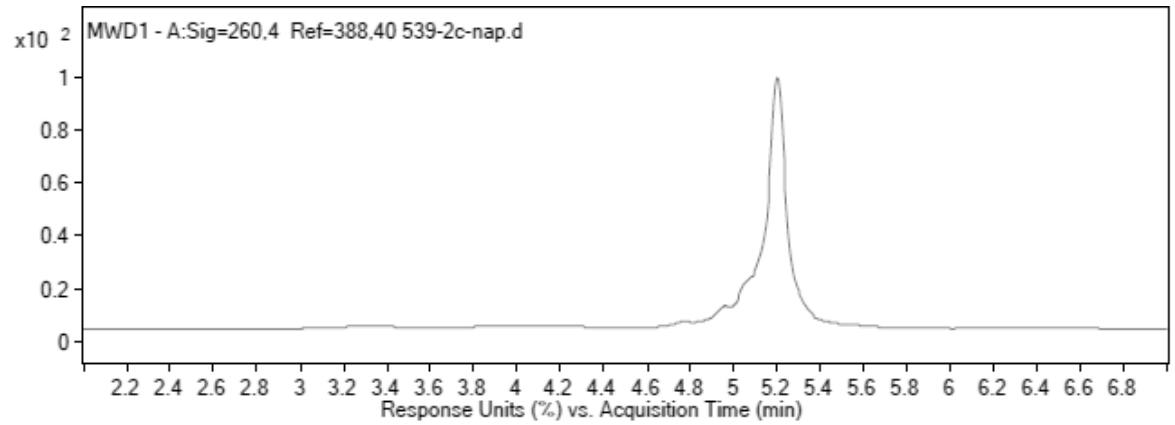


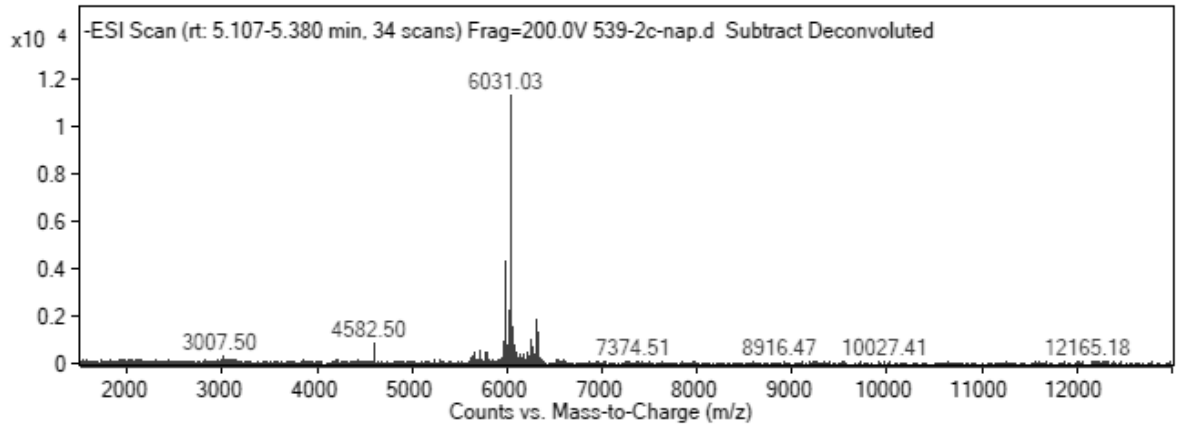
MALAT1_Ch_LNA



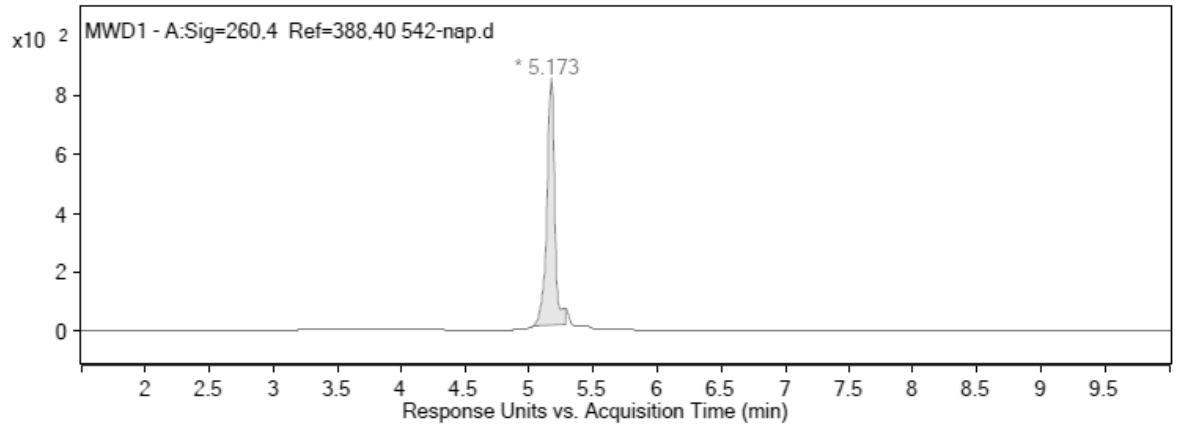


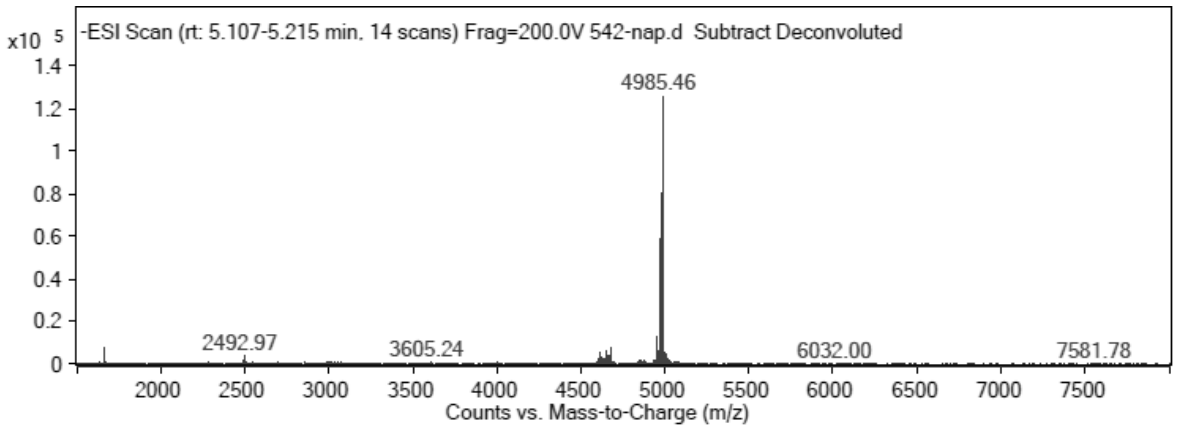
MALAT1_Ch_DNA



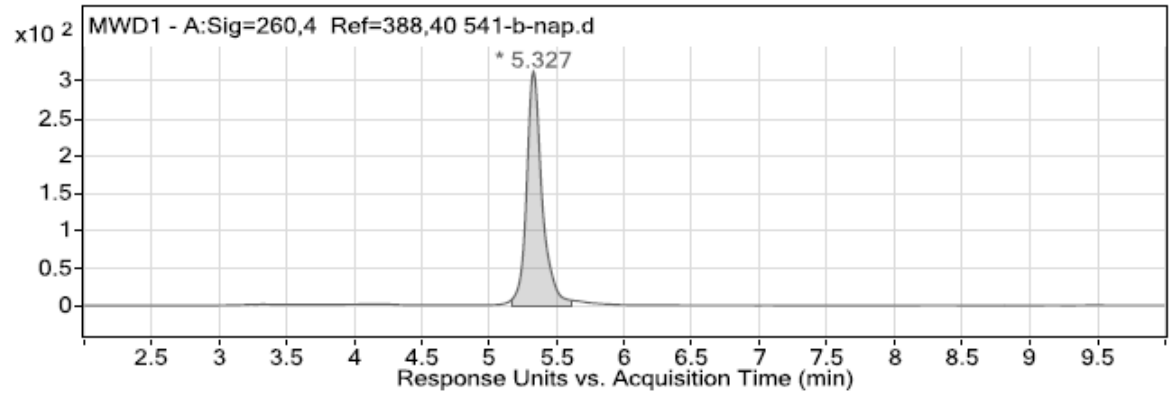


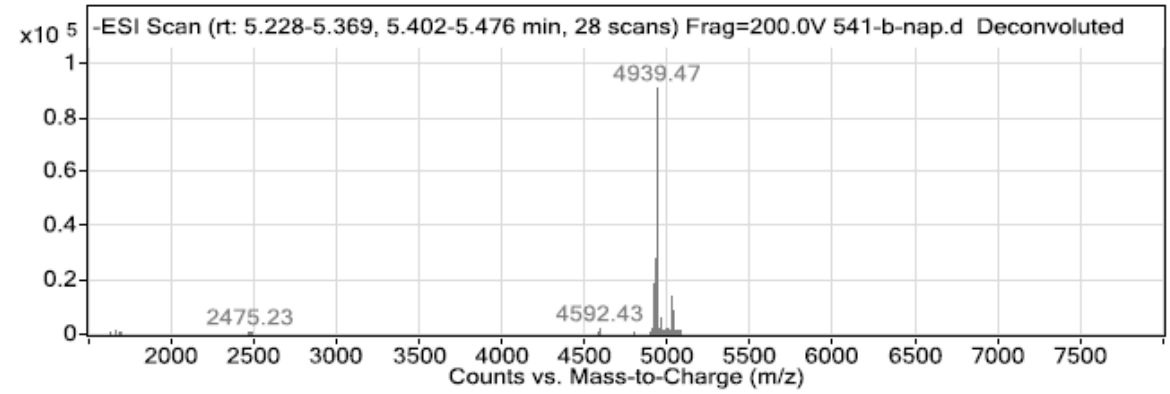
MALAT1_Gap_LNA



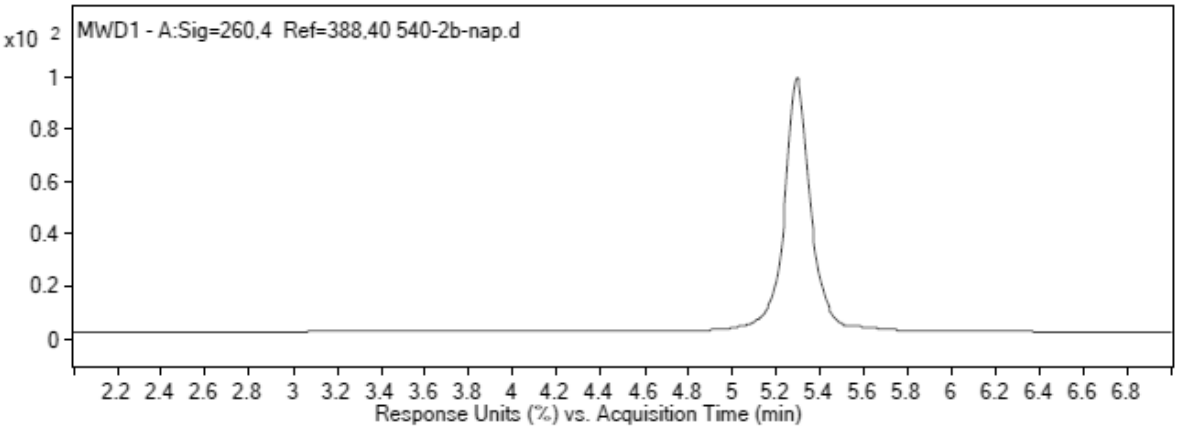


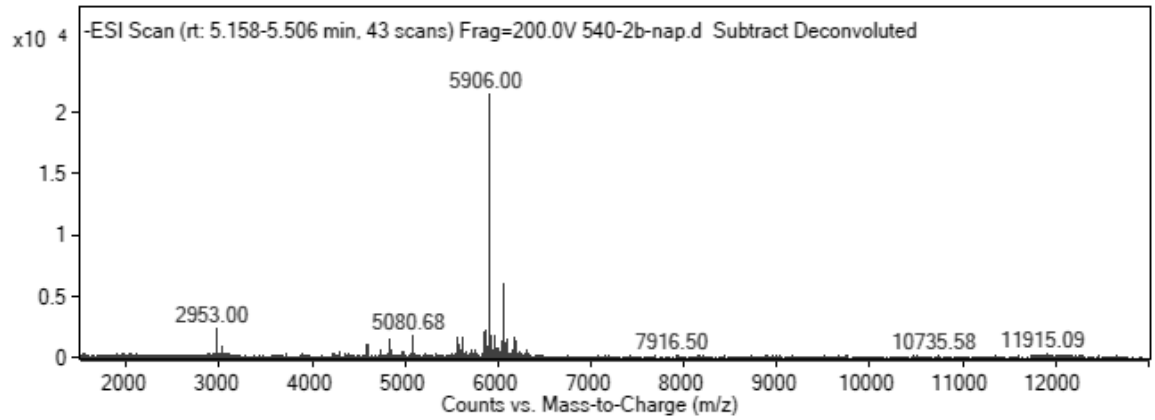
NTC_Gap_LNA




-ESI Scan (rt: 5.228-5.369, 5.402-5.476 min, 28 scans) Frag=200.0V 541-b-nap.d Deconvoluted
x10 5
4939.47
2475.23
4592.43
Counts vs. Mass-to-Charge (m/z)

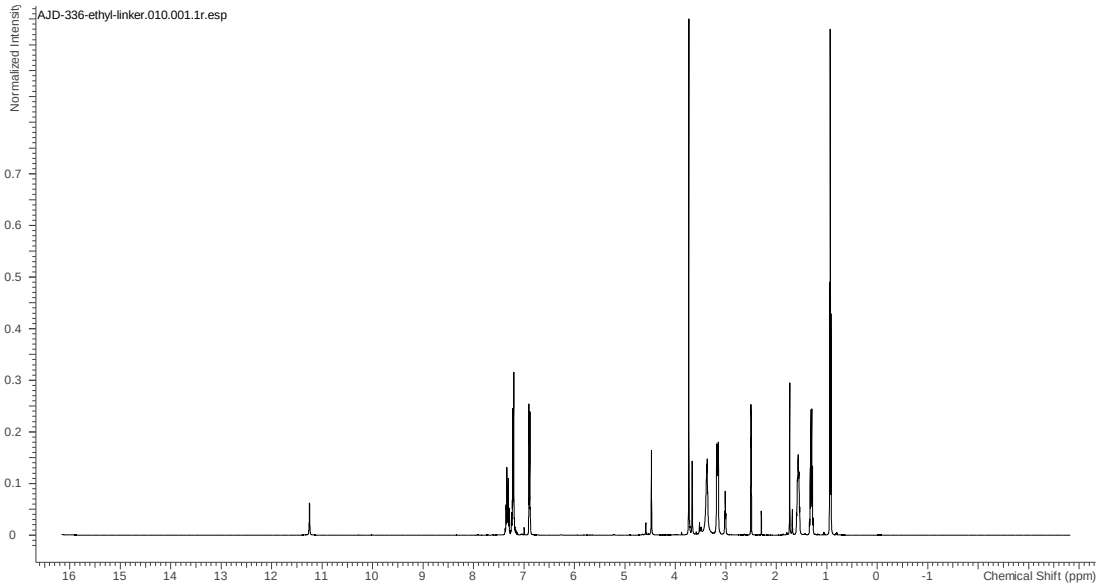
NTC_Ch_LNA


MWD1 - A:Sig=260,4 Ref=388,40 540-2b-nap.d
x10 2
Response Units (%) vs. Acquisition Time (min)

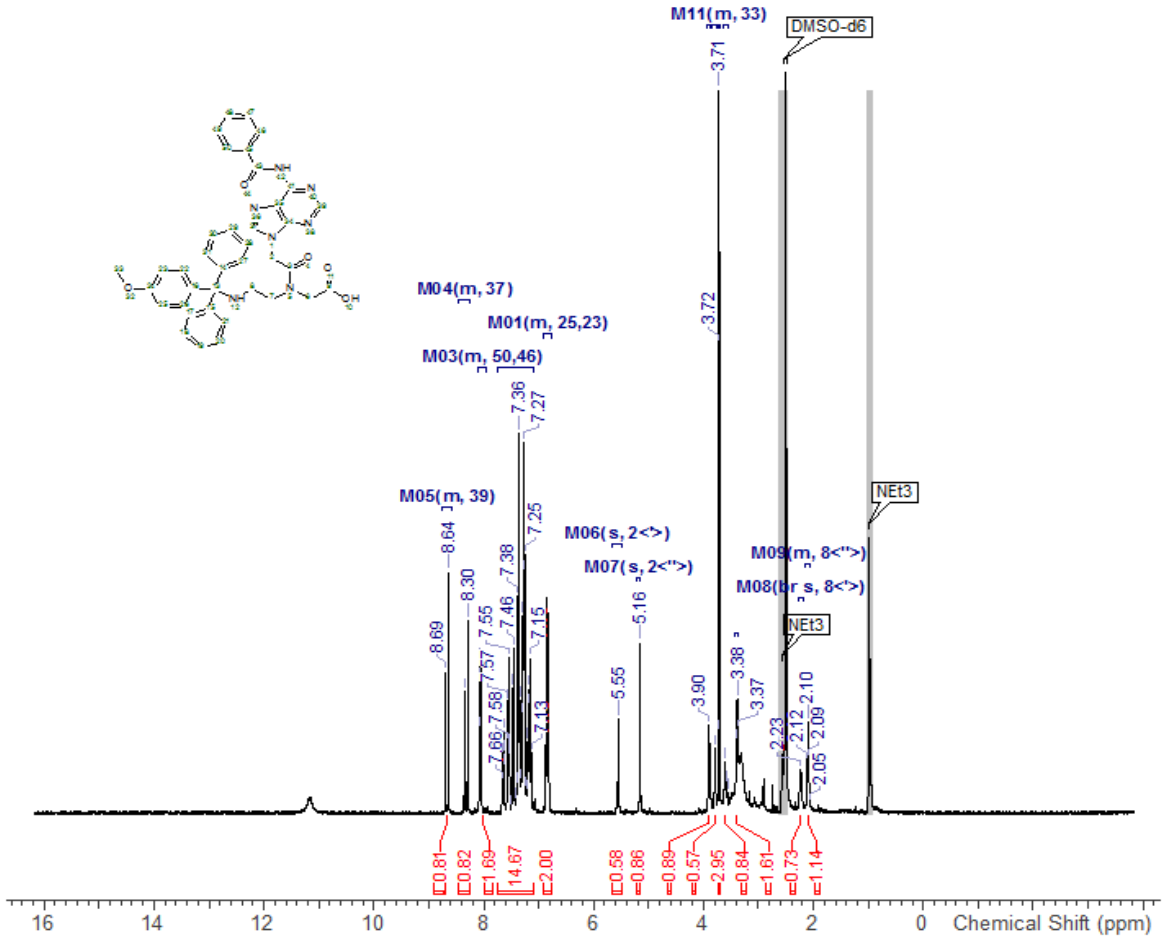

-ESI Scan (rt: 5.158-5.506 min, 43 scans) Frag=200.0V 540-2b-nap.d Subtract Deconvoluted
x10 4
5906.00
2953.00
5080.68
7916.50
10735.58
11915.09
Counts vs. Mass-to-Charge (m/z)

Appendix B NMR and MS spectra of small molecules

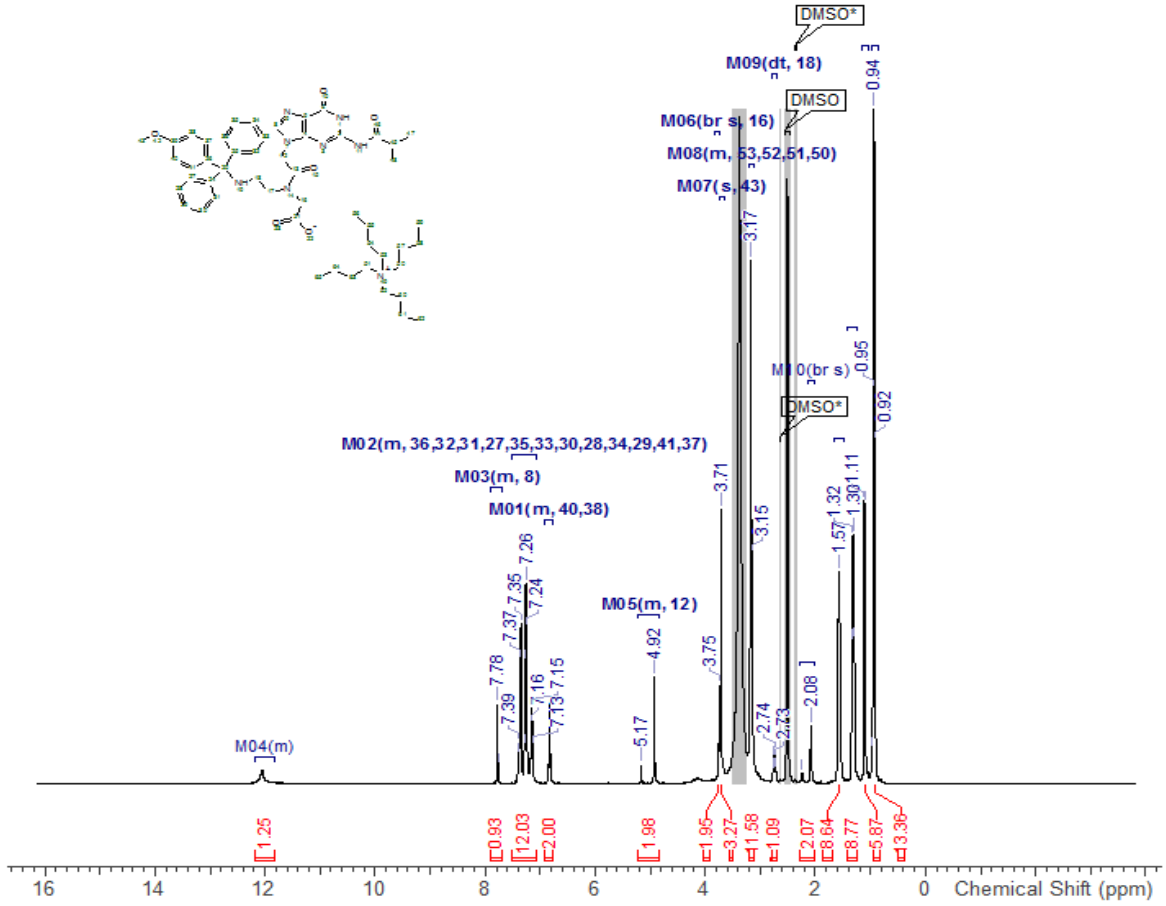
N-(2-(((4-Methoxyphenyl)diphenylmethyl)amino)ethyl)-N-(2-(5-methyl-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)acetyl)glycine (5T)

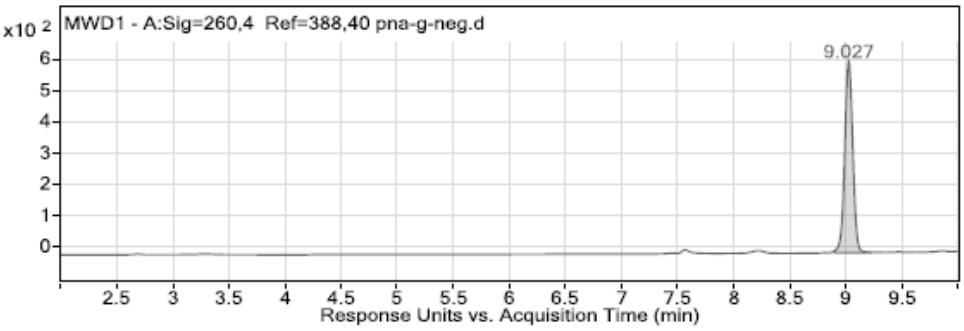


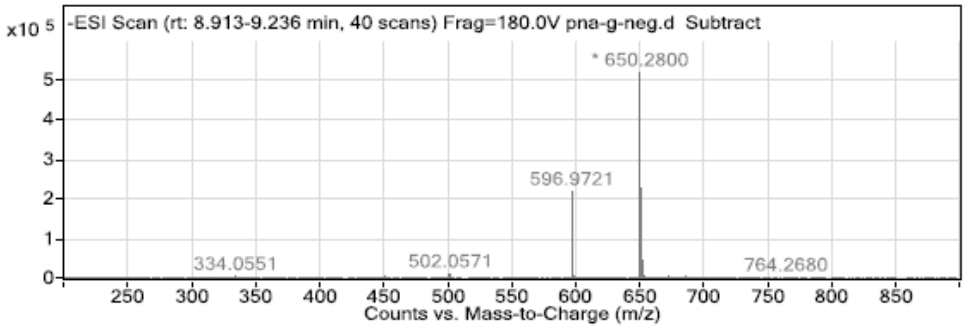
N-(2-(6-Benzamido-9H-purin-9-yl)acetyl)-N-(2-(((4-methoxyphenyl)diphenylmethyl)amino)ethyl)glycine (5A)



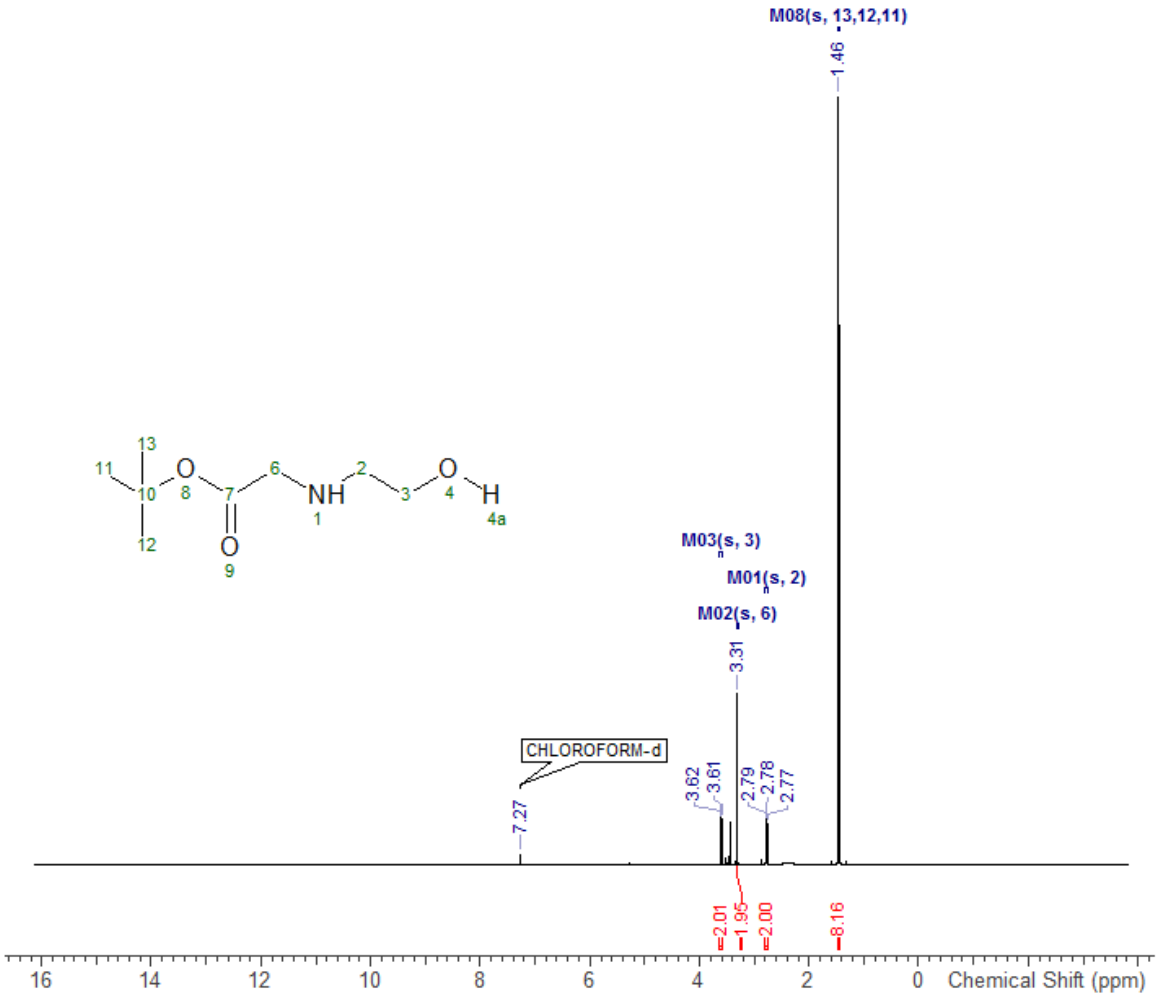
N-(2-(2-Isobutyramido-6-oxo-1,6-dihydro-9H-purin-9-yl)acetyl)-N-(2-(((4-methoxyphenyl)diphenylmethyl)amino)ethyl)glycinate (5G)



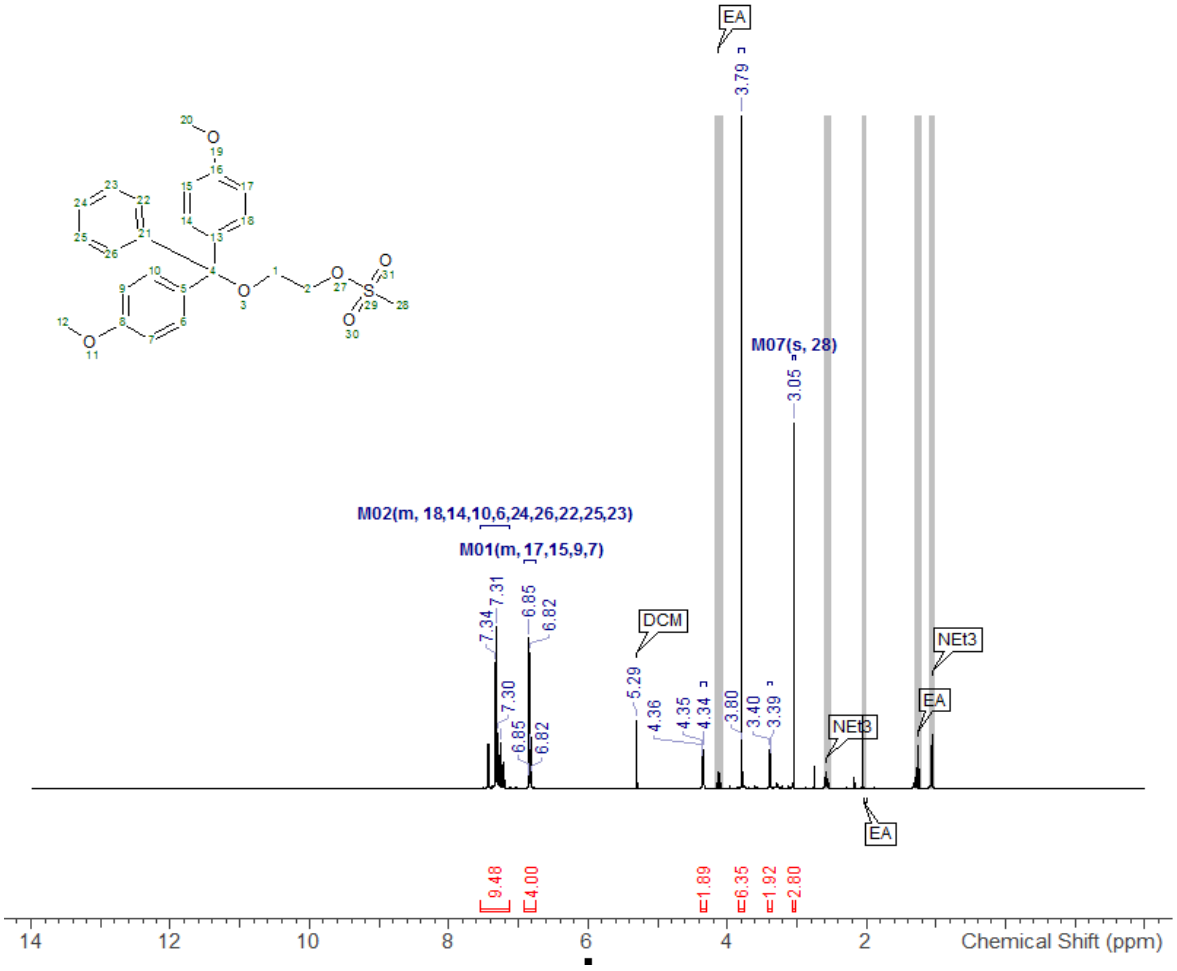




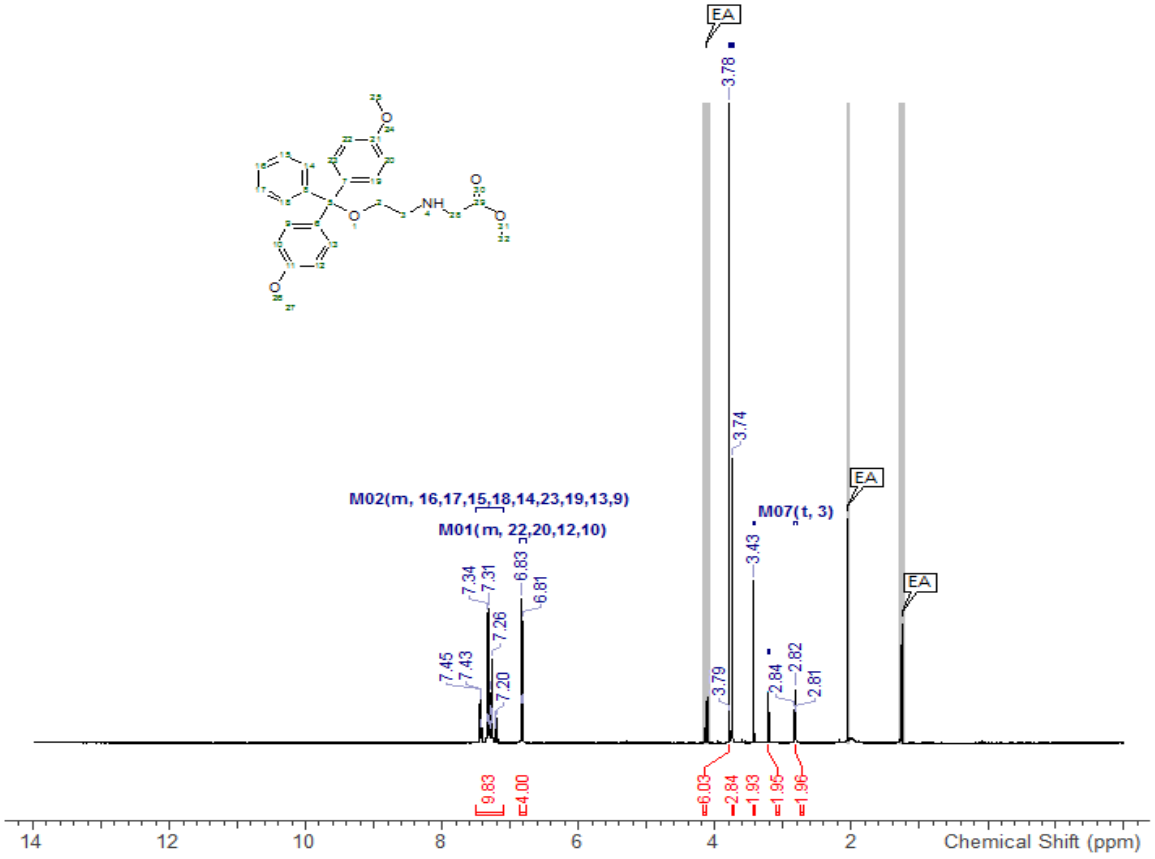
Tert-butyl (2-hydroxyethyl)glycinate (27)



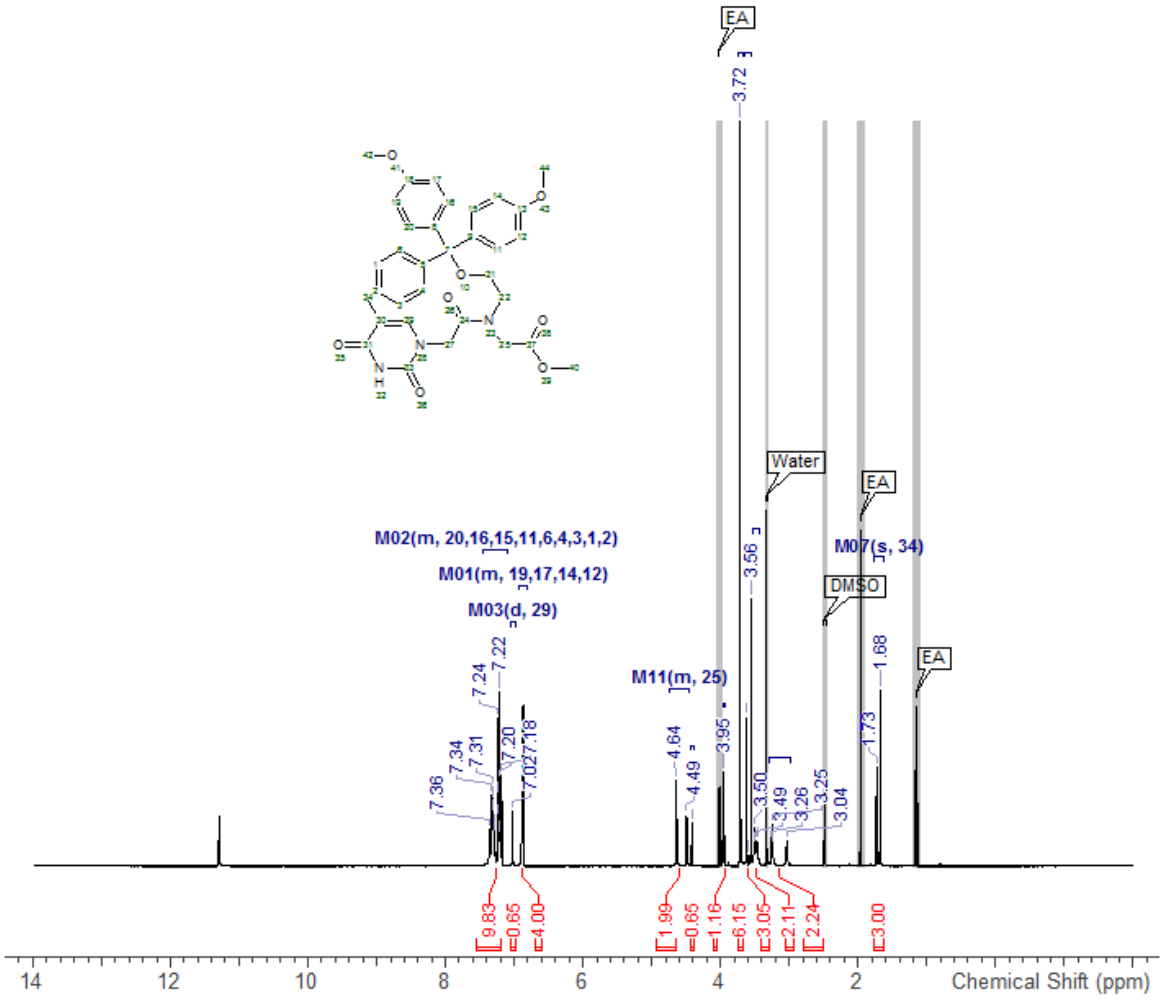
Dimethoxytrityl-ethyl methanesulfonate (40)



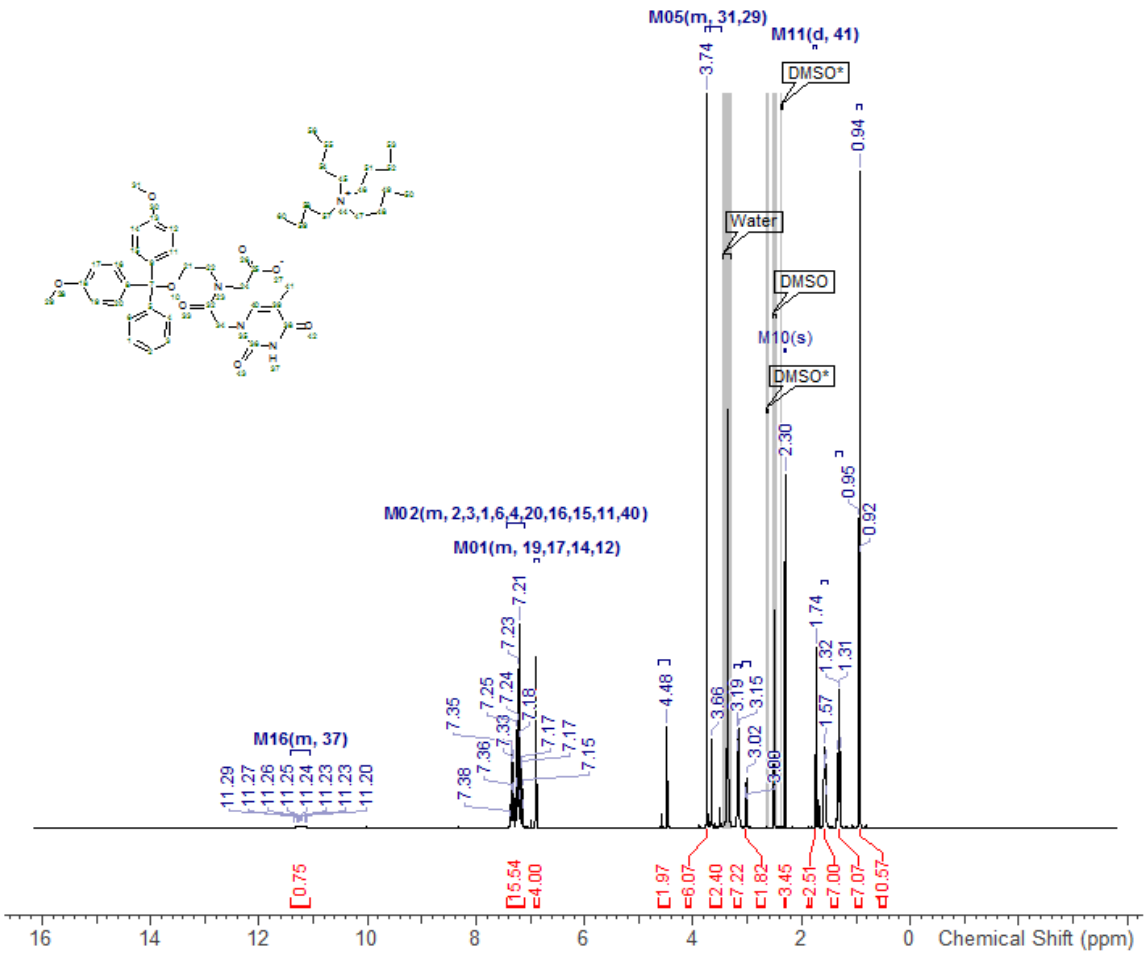
Dimethoxytrityl-ethyl glycinate (41)

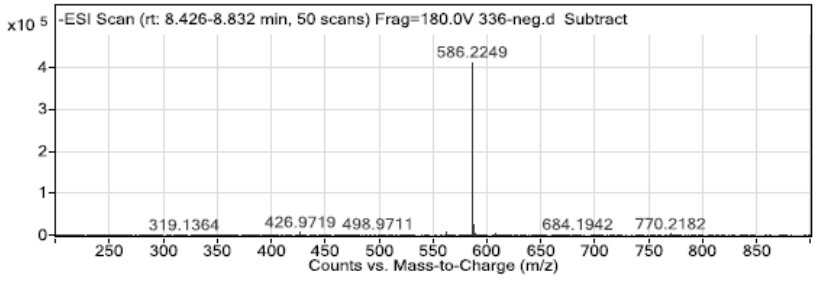


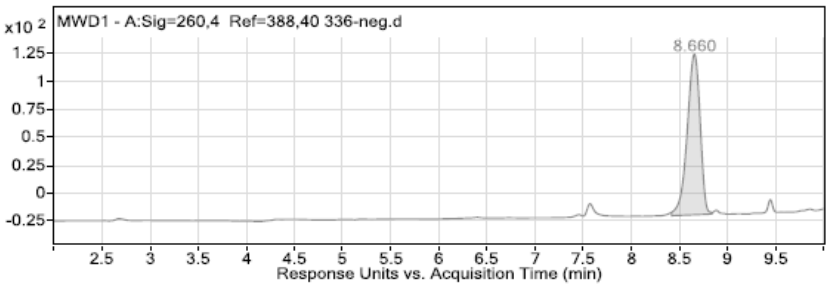
Methyl N-(2-(bis(4-methoxyphenyl)(phenyl)methoxy)ethyl)-N-(2-(5-methyl-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)acetyl)glycinate (42)



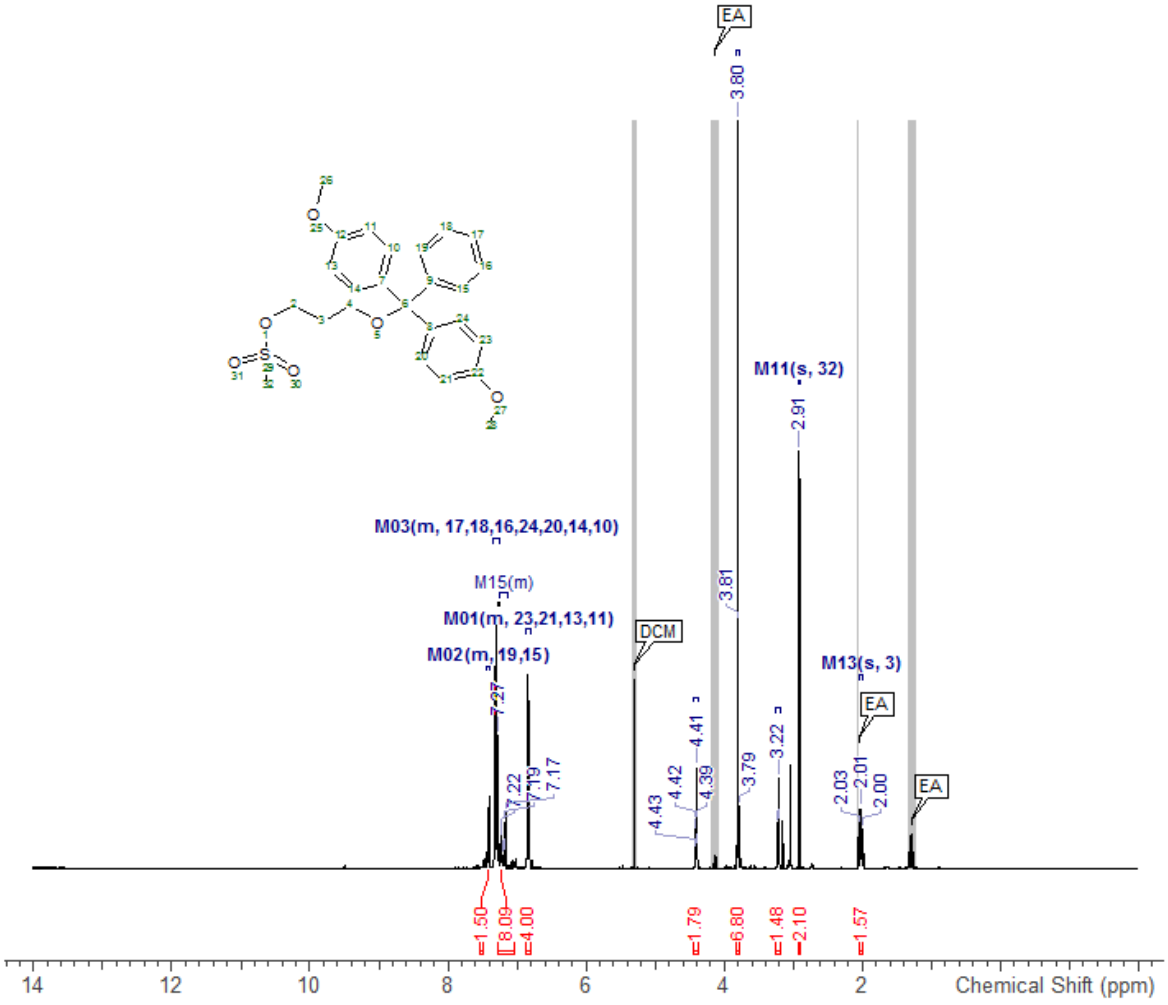
***N*-(2-(Bis(4-methoxyphenyl)(phenyl)methoxy)ethyl)-N-(2-(5-methyl-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)acetyl)glycine (43)**



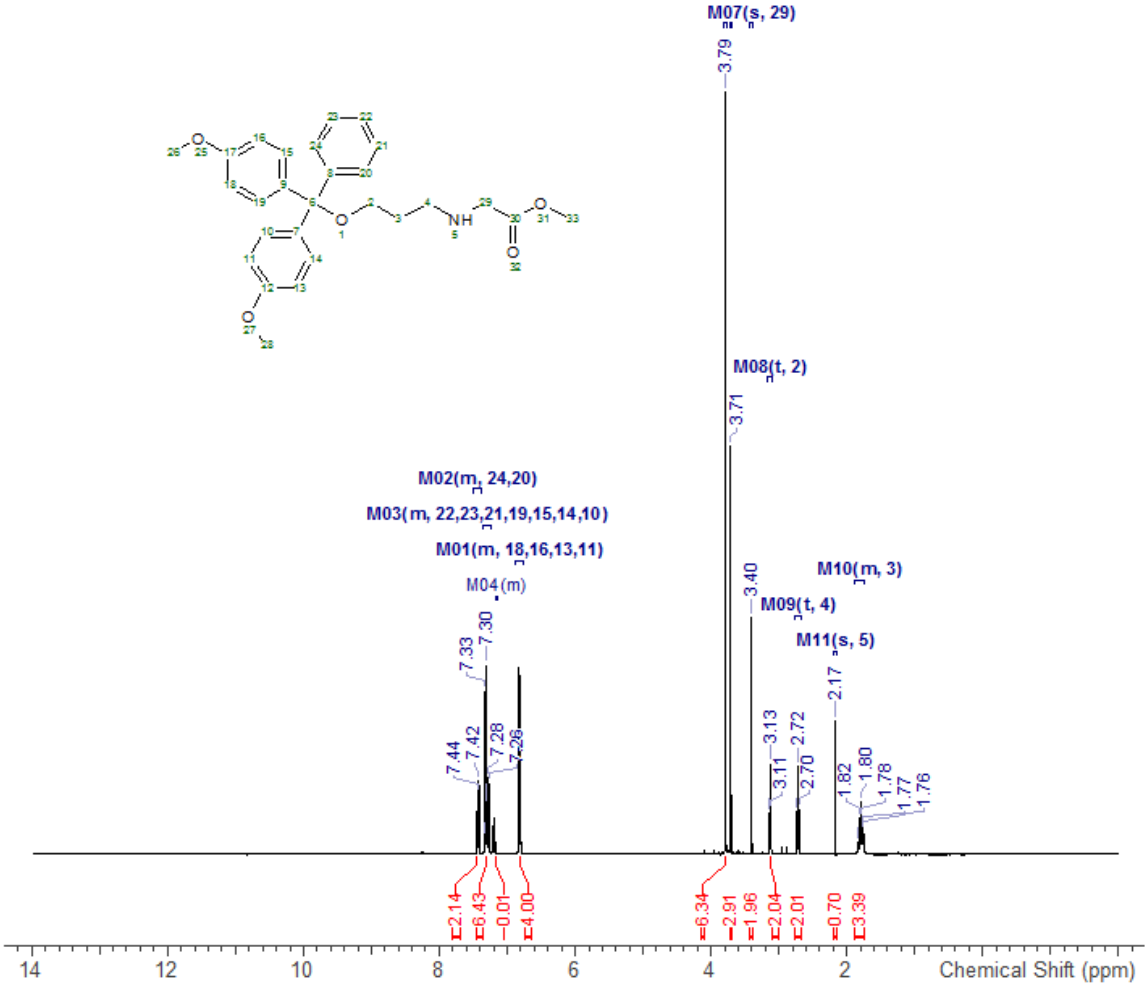




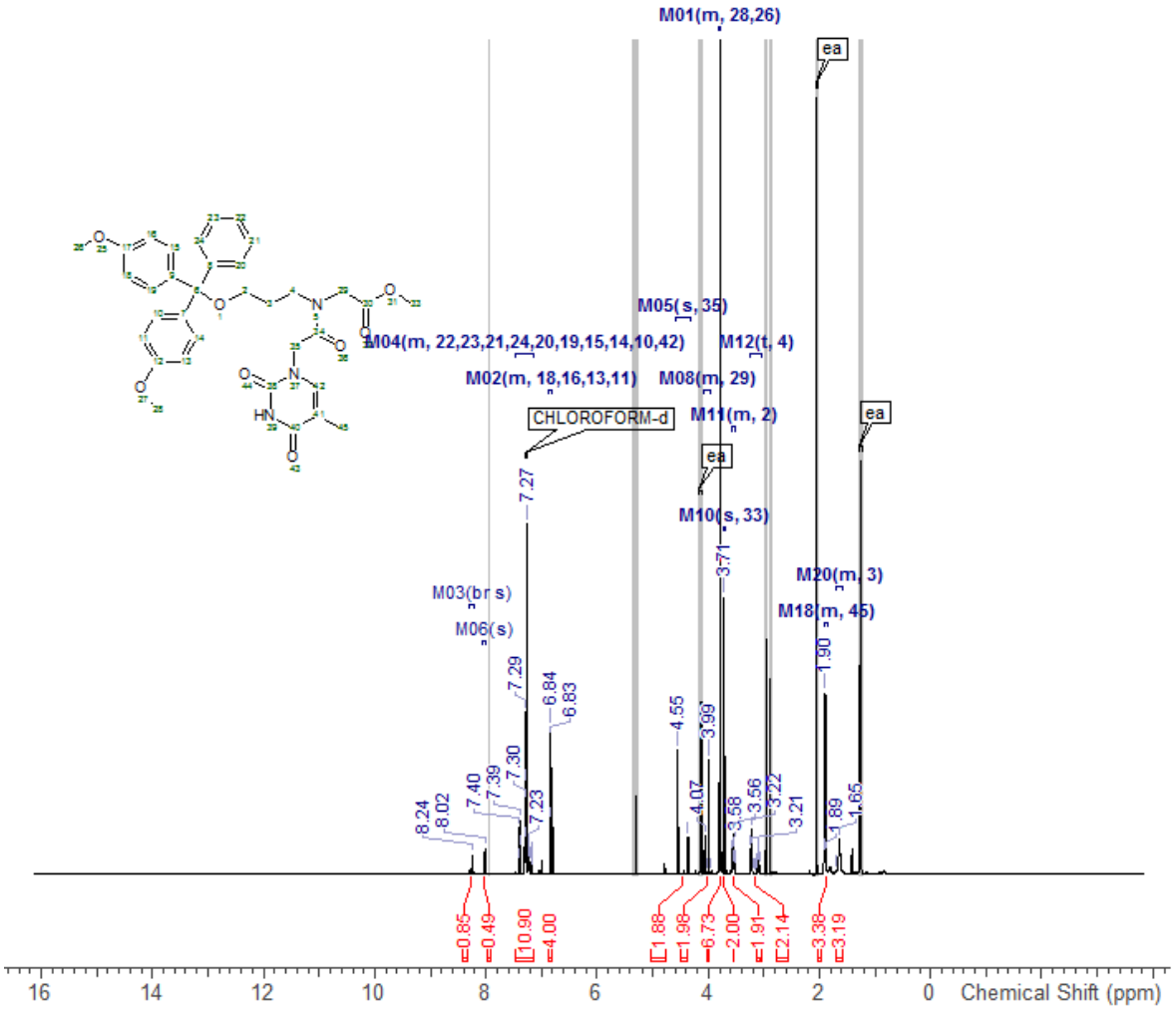
3-(bis(4-methoxyphenyl)(phenyl)methoxy)propyl methanesulfonate (45)



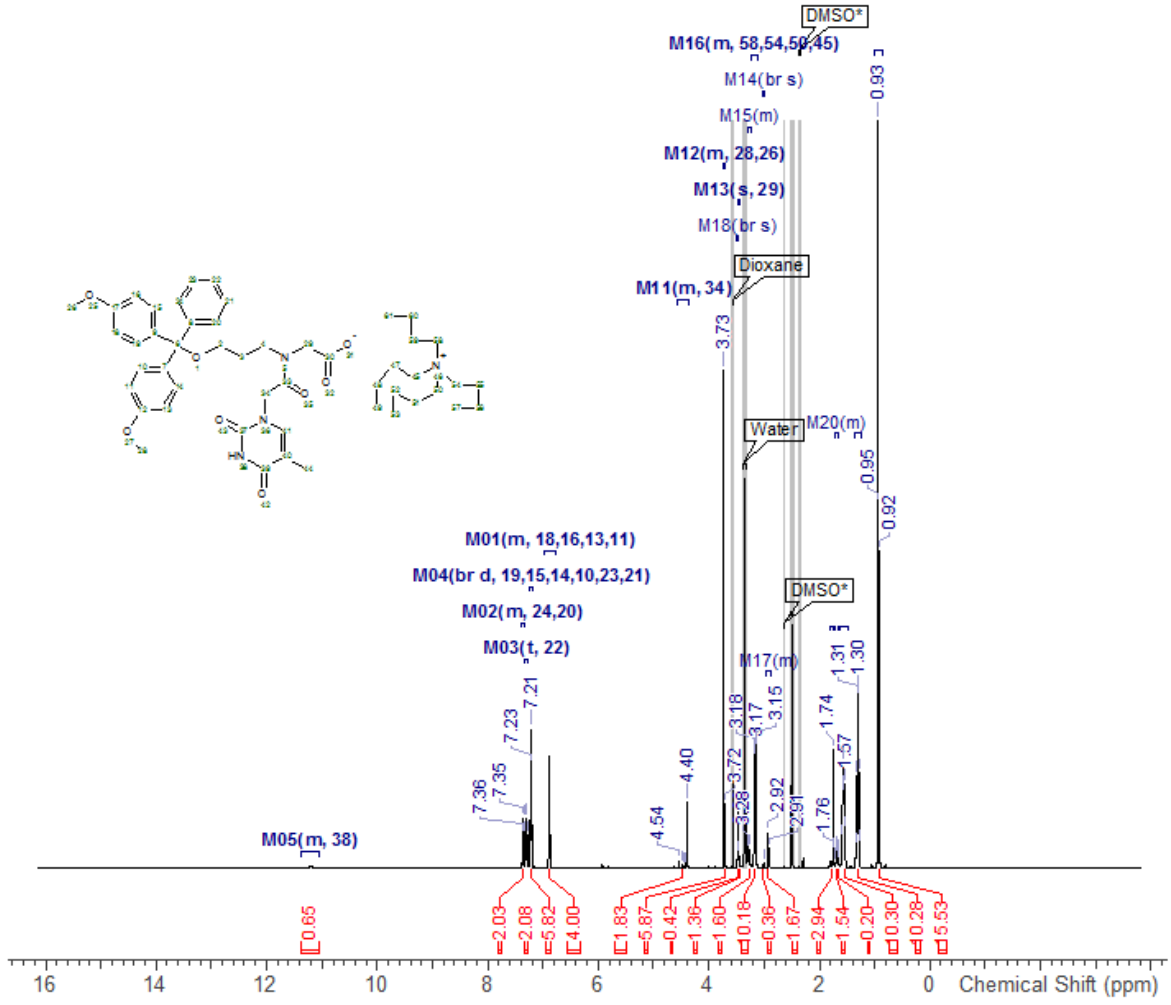
Methyl (3-(bis(4-methoxyphenyl)(phenyl)methoxy)propyl)glycinate (46)

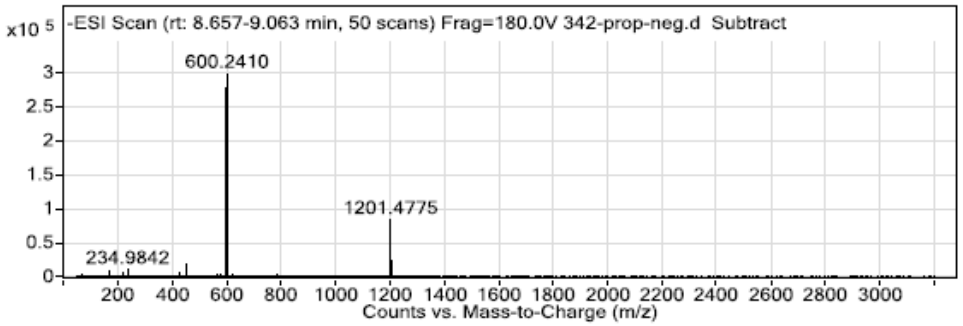


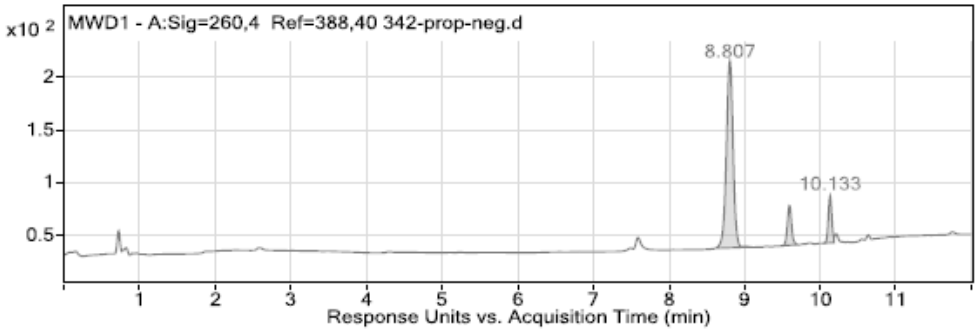
Methyl *N*-(3-(bis(4-methoxyphenyl)(phenyl)methoxy)propyl)-*N*-(2-(5-methyl-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)acetyl)glycinate (47)



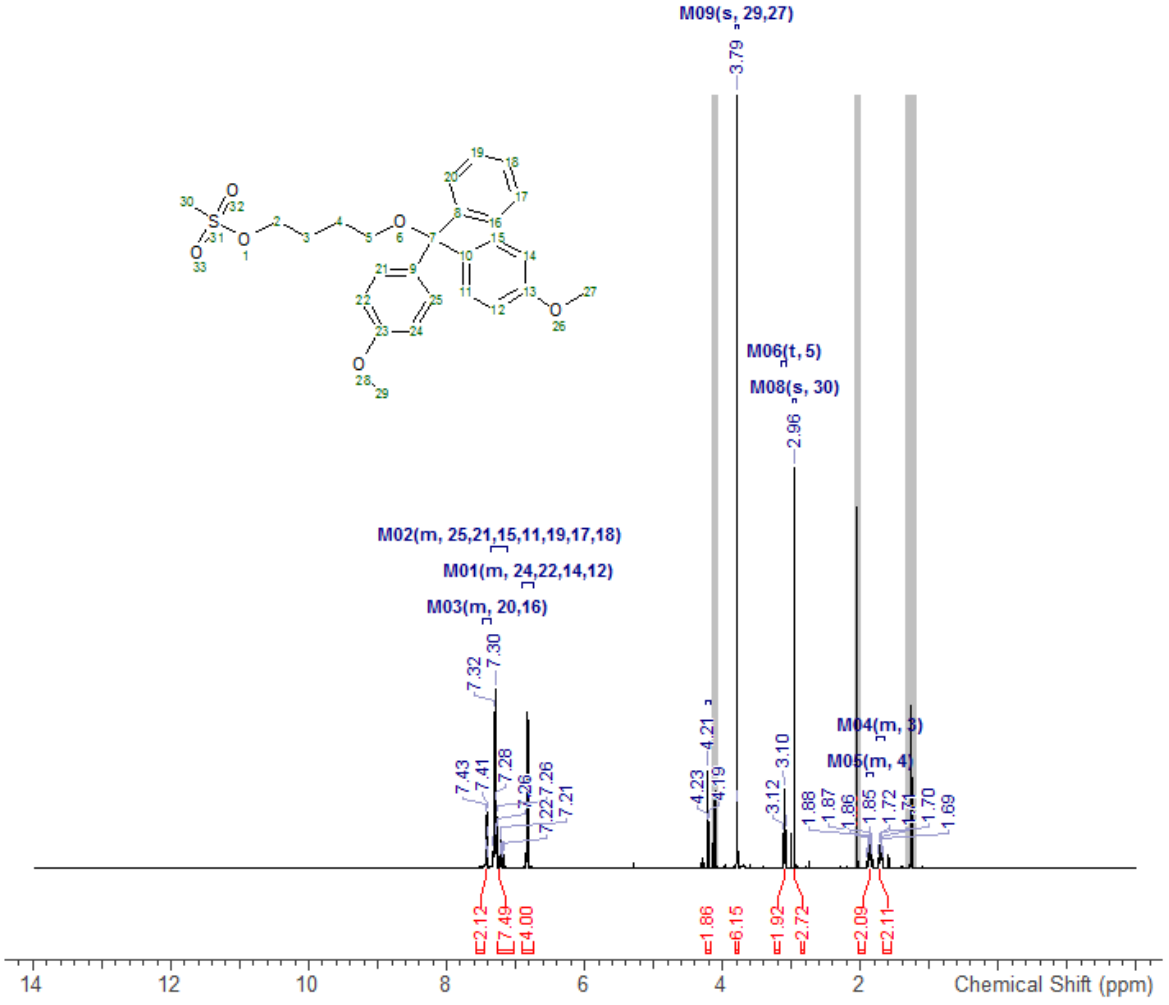
***N*-(3-(Bis(4-methoxyphenyl)(phenyl)methoxy)propyl)-*N*-(2-(5-methyl-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)acetyl)glycinate (48)**



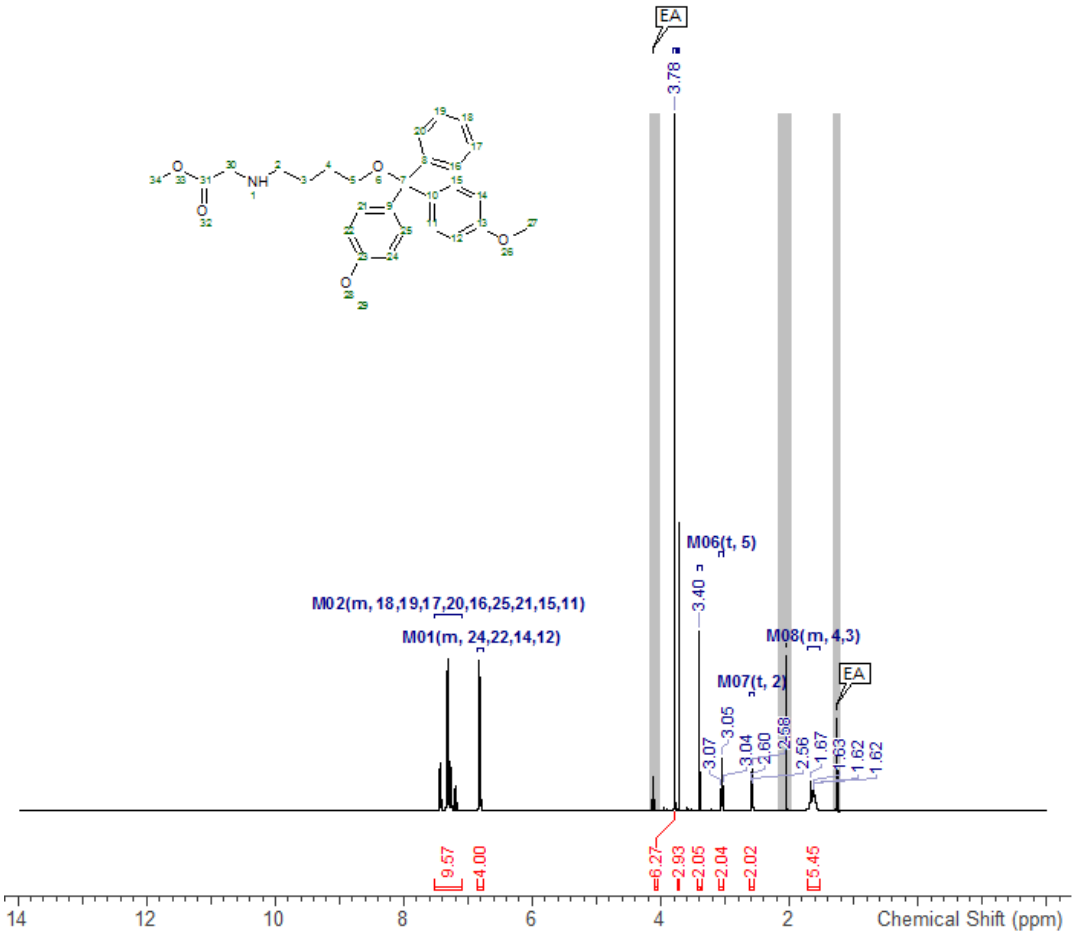




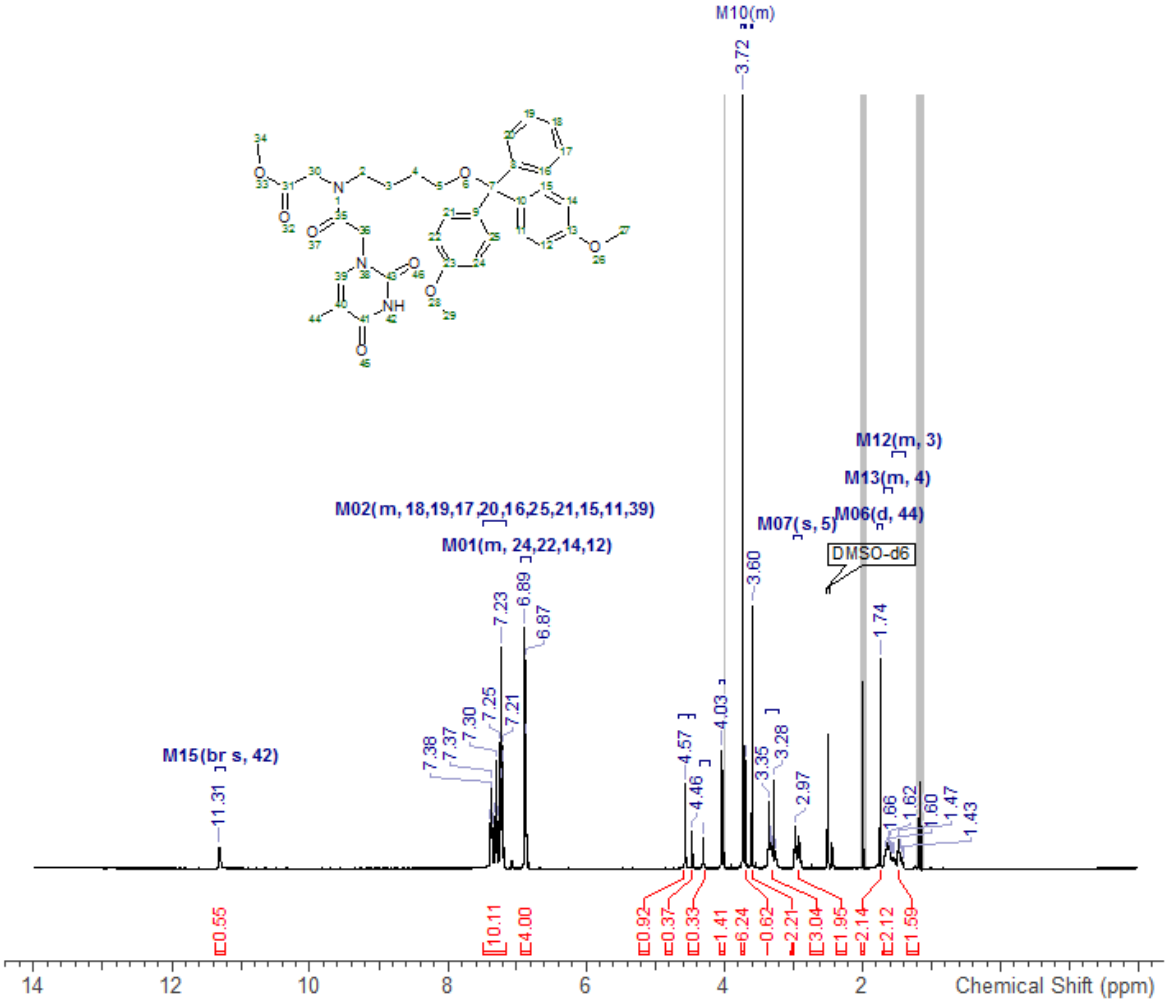
4-(Bis(4-methoxyphenyl)(phenyl)methoxy)butyl methanesulfonate (50)



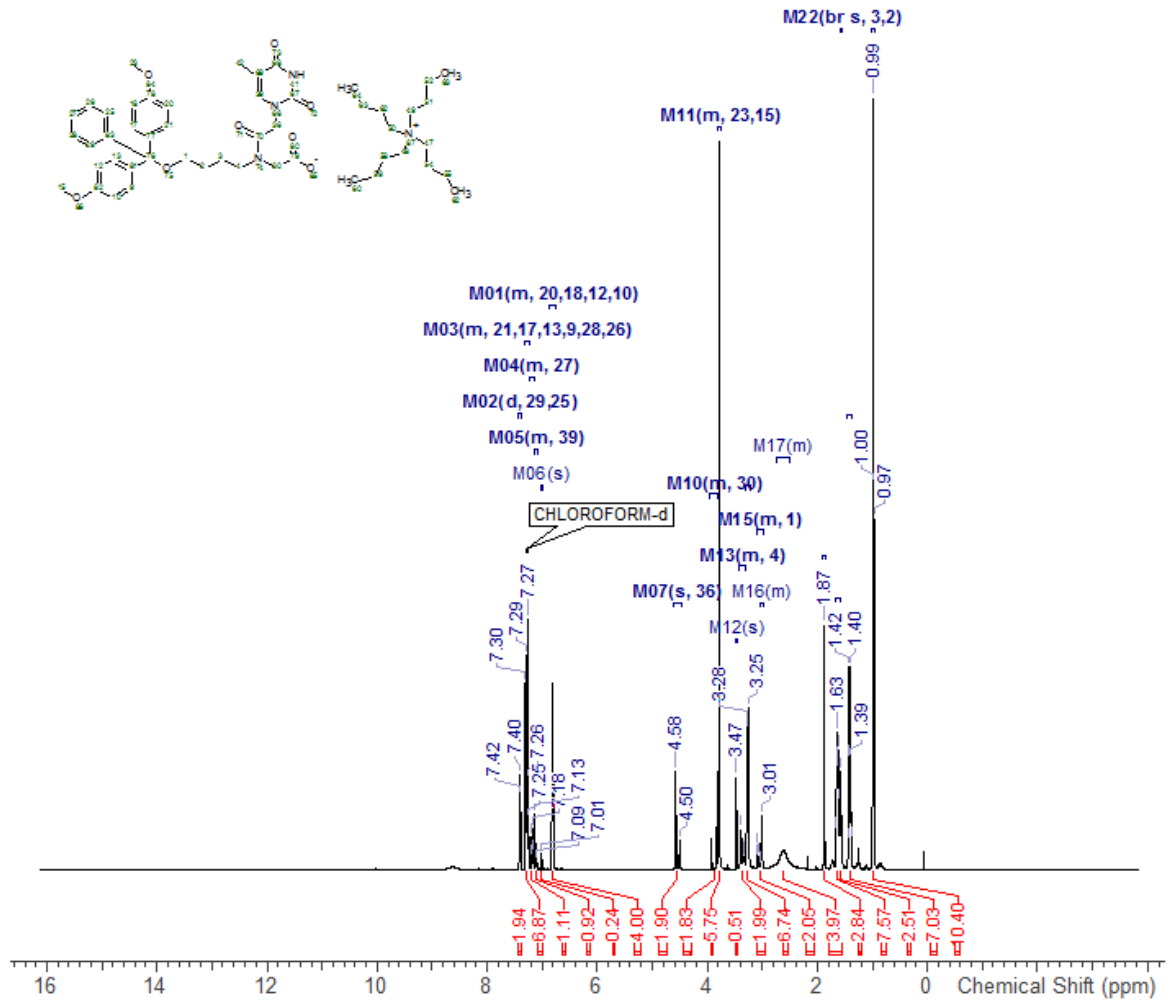
Methyl (4-(bis(4-methoxyphenyl)(phenyl)methoxy)butyl)glycinate (51)

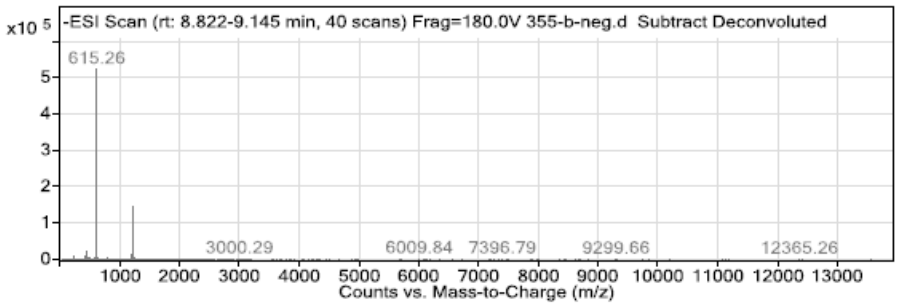


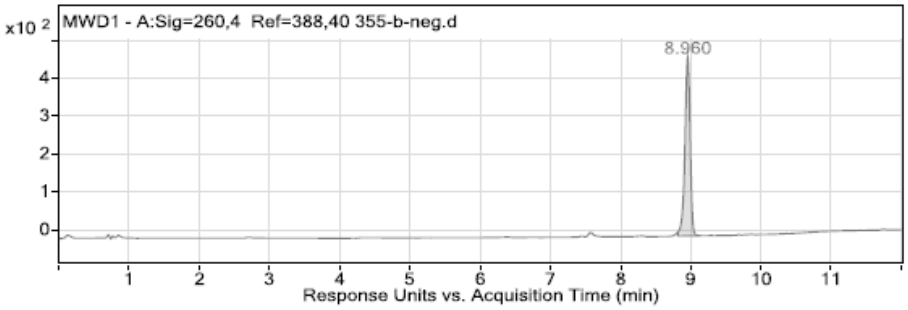
Methyl *N*-(4-(bis(4-methoxyphenyl)(phenyl)methoxy)butyl)-*N*-(2-(5-methyl-2,6-dioxo-1,2,3,6-tetrahydropyridin-3-yl)acetyl)glycinate (52)



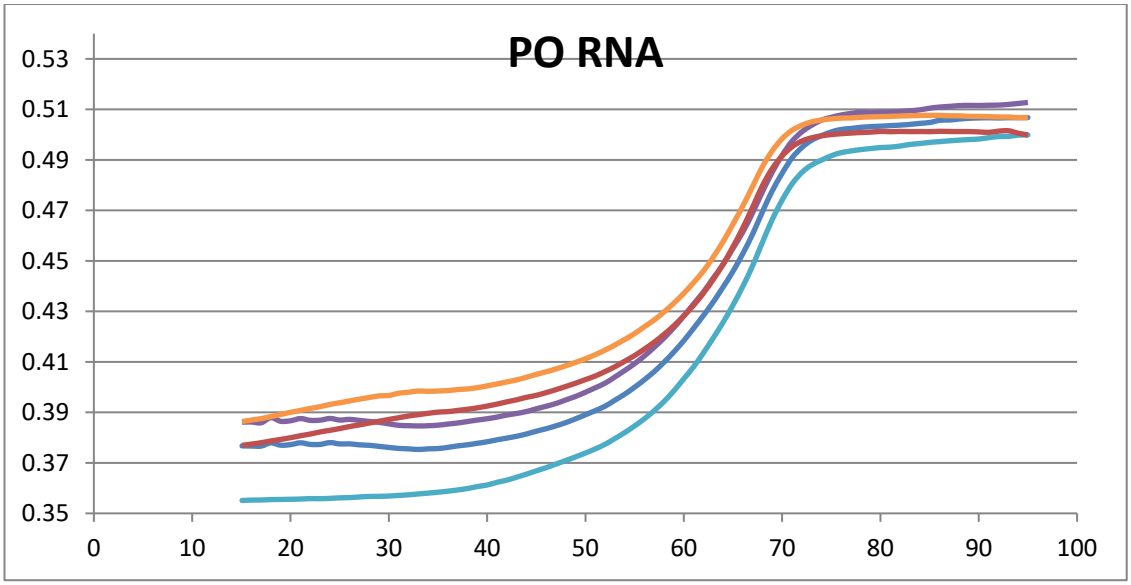
***N*-(4-(Bis(4-methoxyphenyl)(phenyl)methoxy)butyl)-*N*-(2-(5-methyl-2,4-dioxo-3,4-dihydropyrimidin-1(2H)-yl)acetyl)glycine (53)**

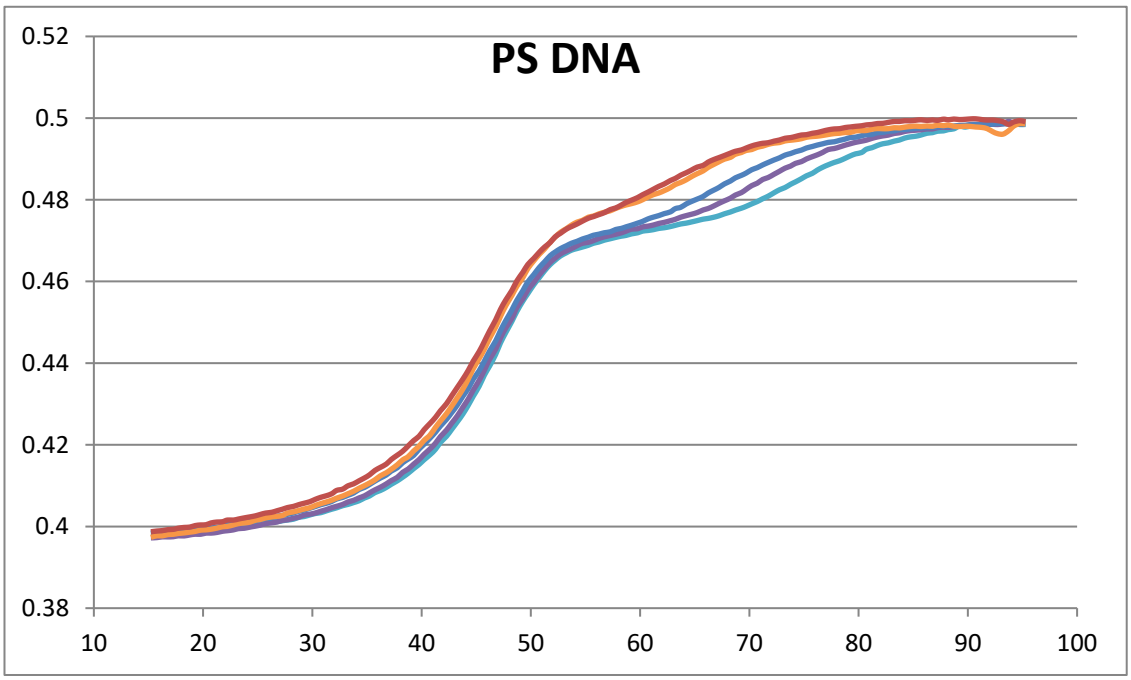


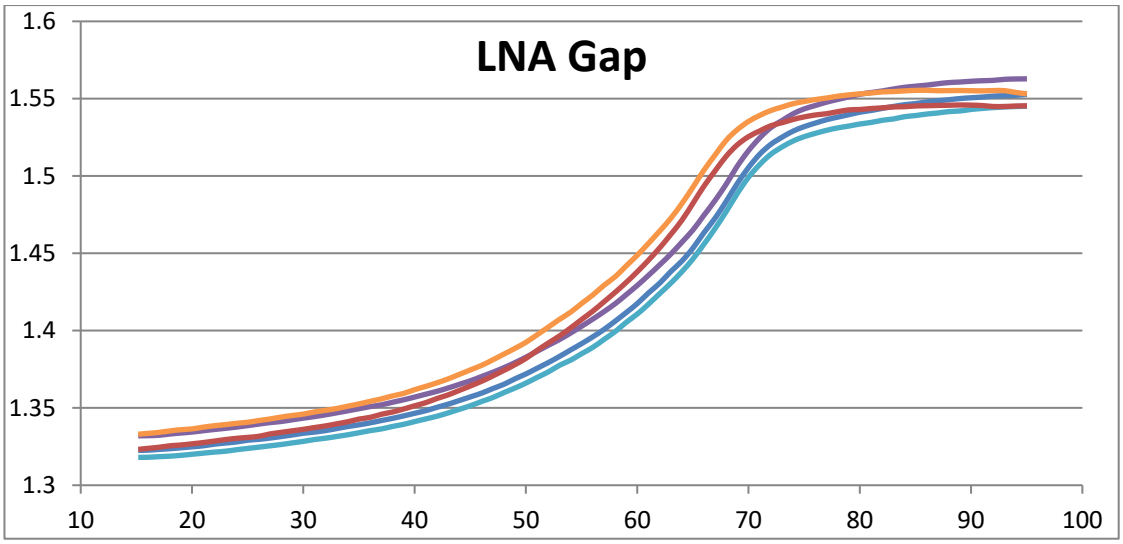


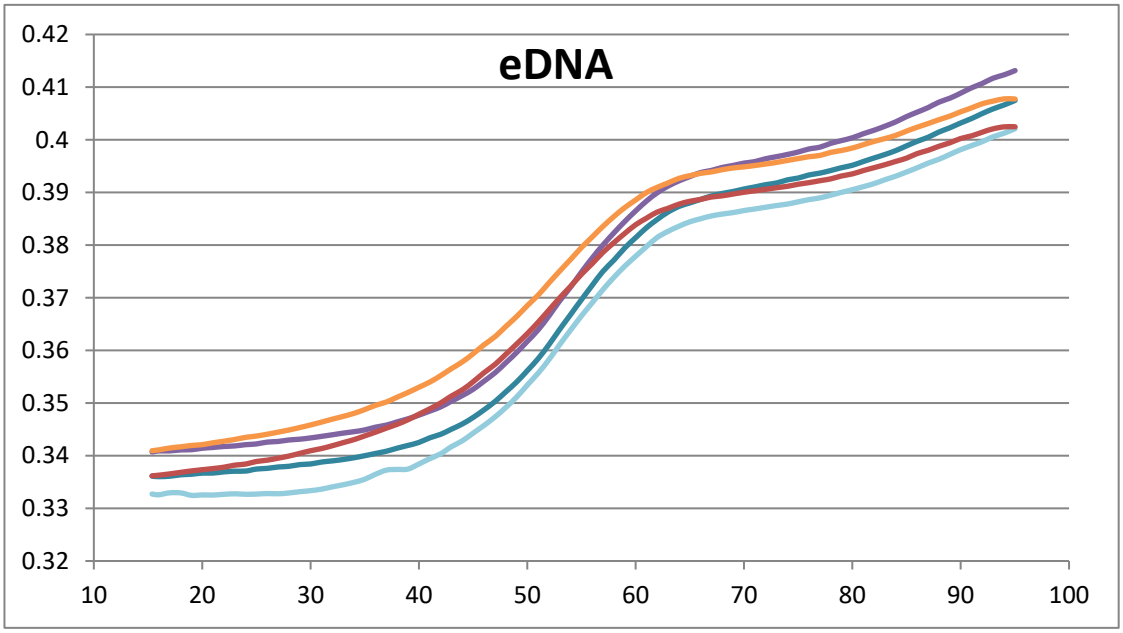


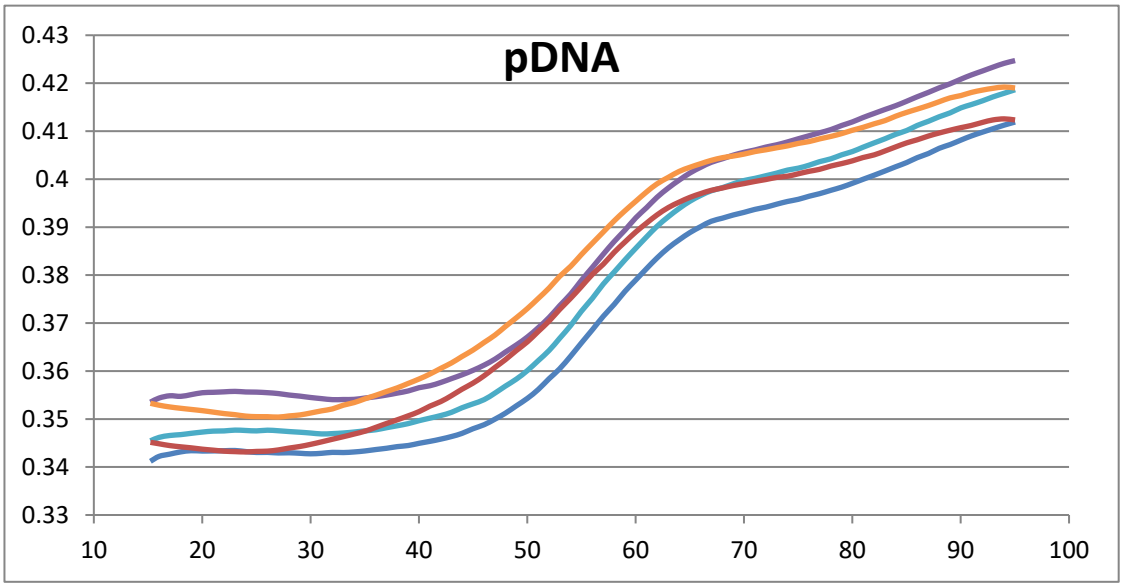
Appendix C Melt curves

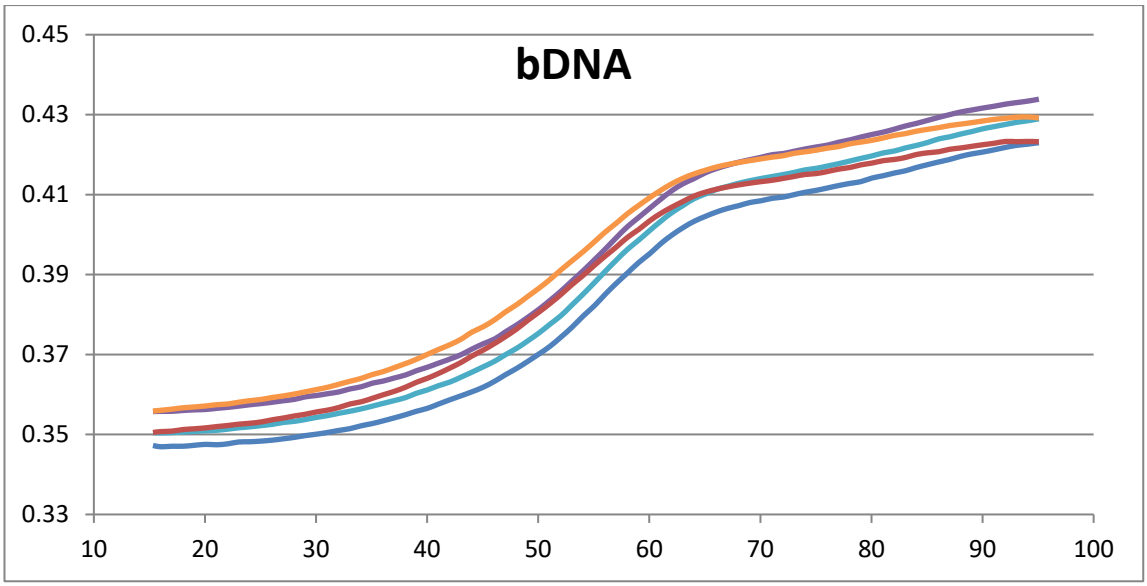


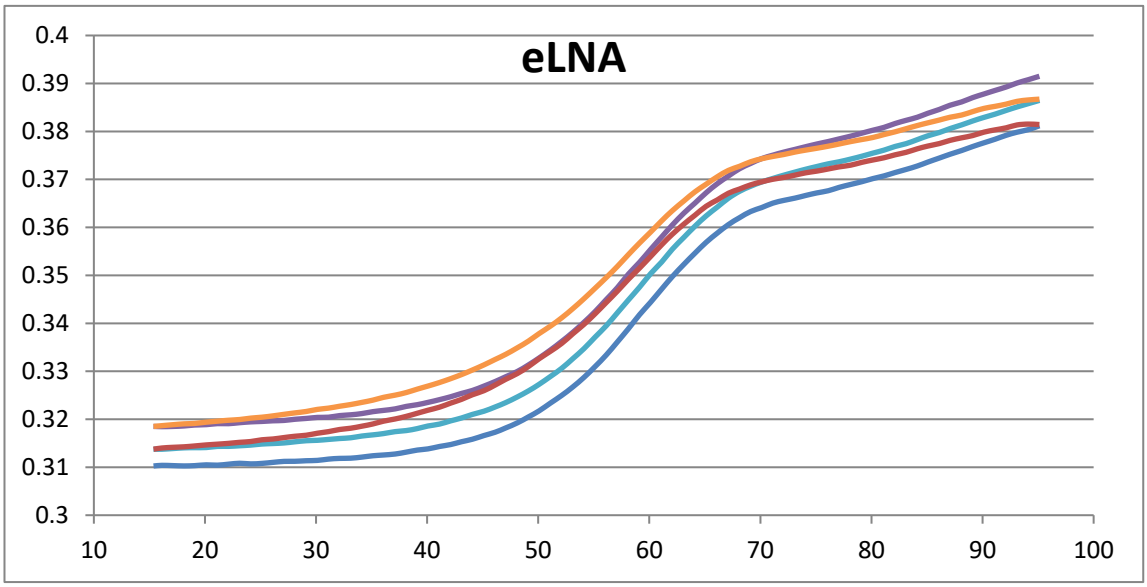


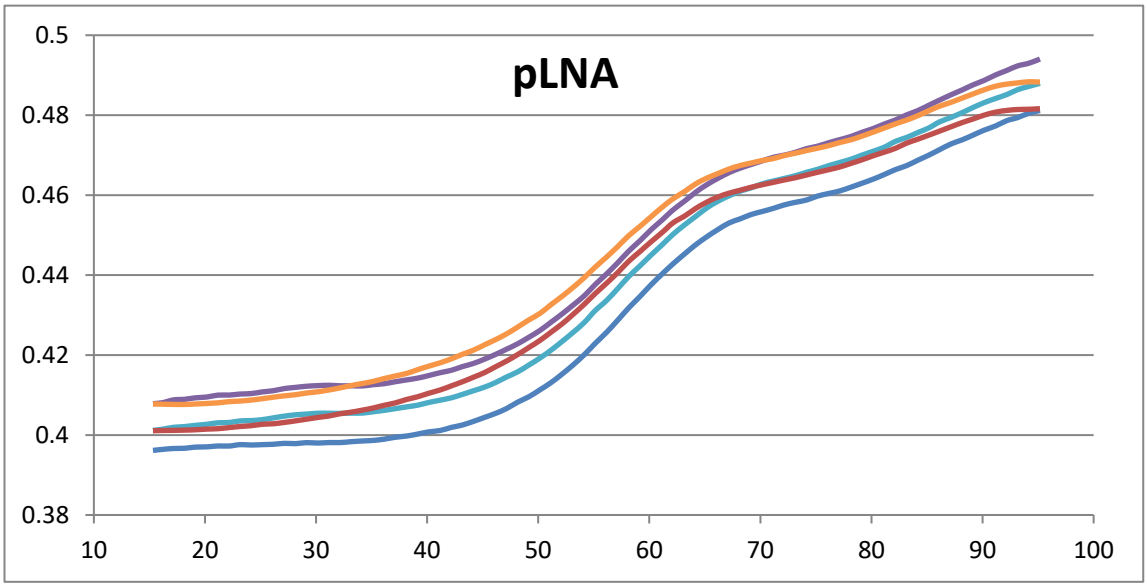

LNA Gap
1.6
1.55
1.5
1.45
1.4
1.35
1.3
10
20
30
40
50
60
70
80
90
100

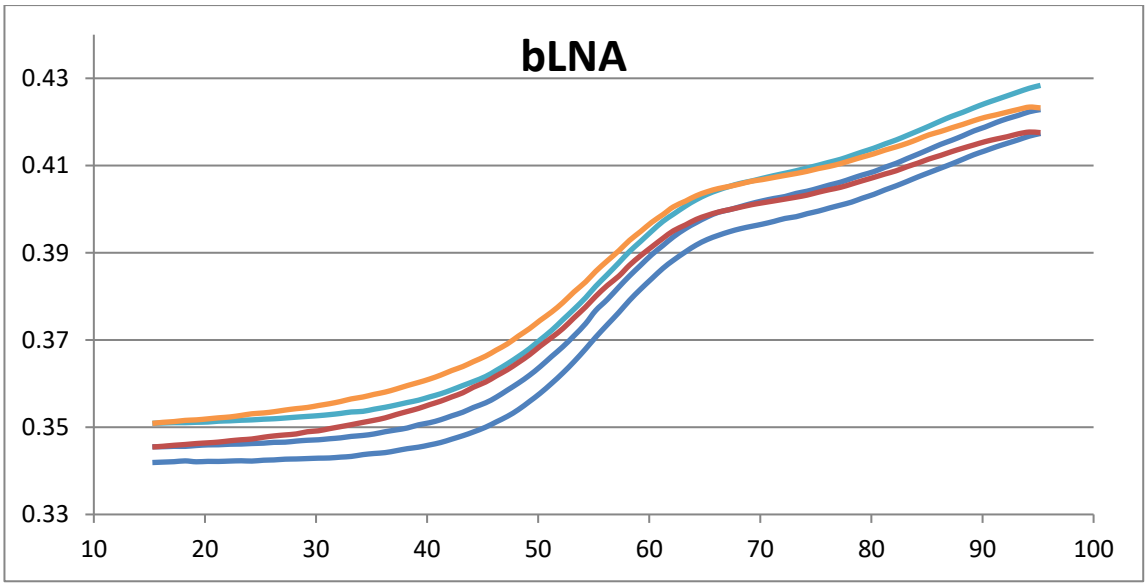

eDNA
0.42
0.41
0.4
0.39
0.38
0.37
0.36
0.35
0.34
0.33
0.32
10
20
30
40
50
60
70
80
90
100

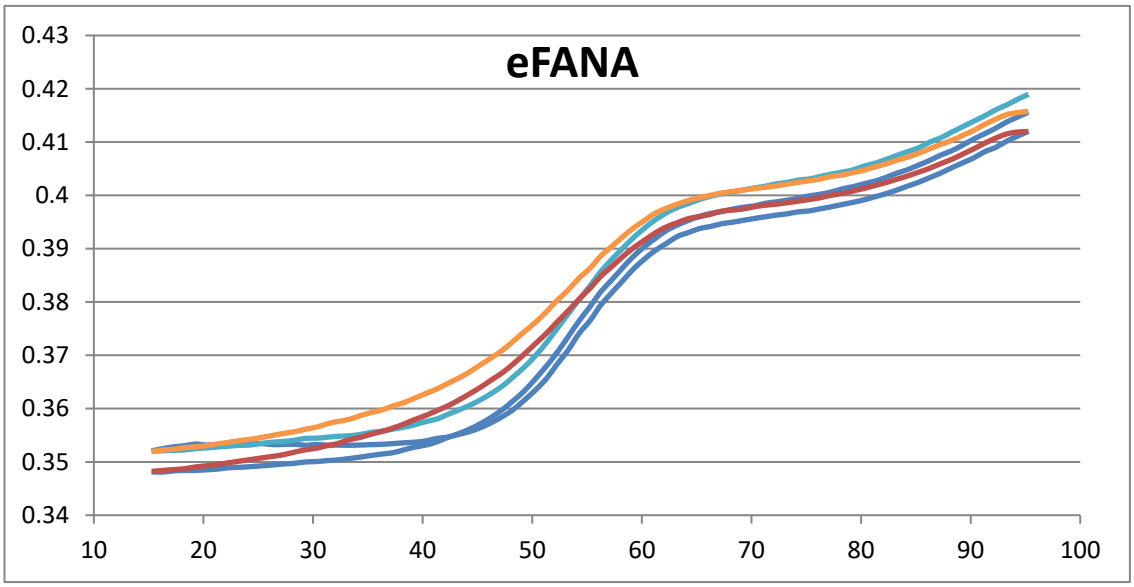

pDNA
0.43
0.42
0.41
0.4
0.39
0.38
0.37
0.36
0.35
0.34
0.33
10
20
30
40
50
60
70
80
90
100

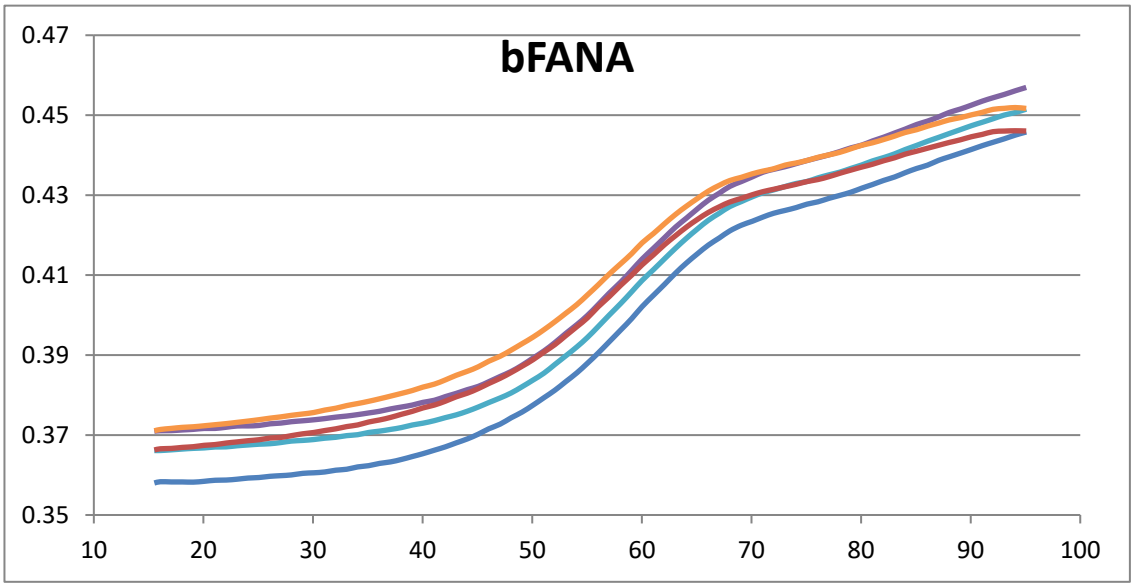

bDNA
0.45
0.43
0.41
0.39
0.37
0.35
0.33
10
20
30
40
50
60
70
80
90
100

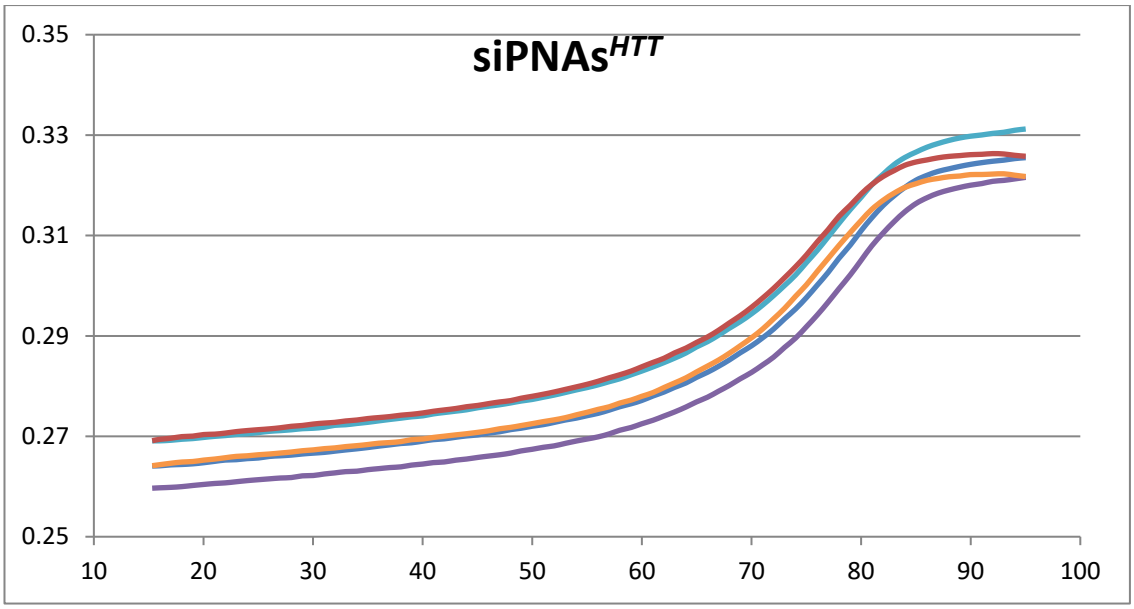

eLNA
0.4
0.39
0.38
0.37
0.36
0.35
0.34
0.33
0.32
0.31
0.3
10
20
30
40
50
60
70
80
90
100

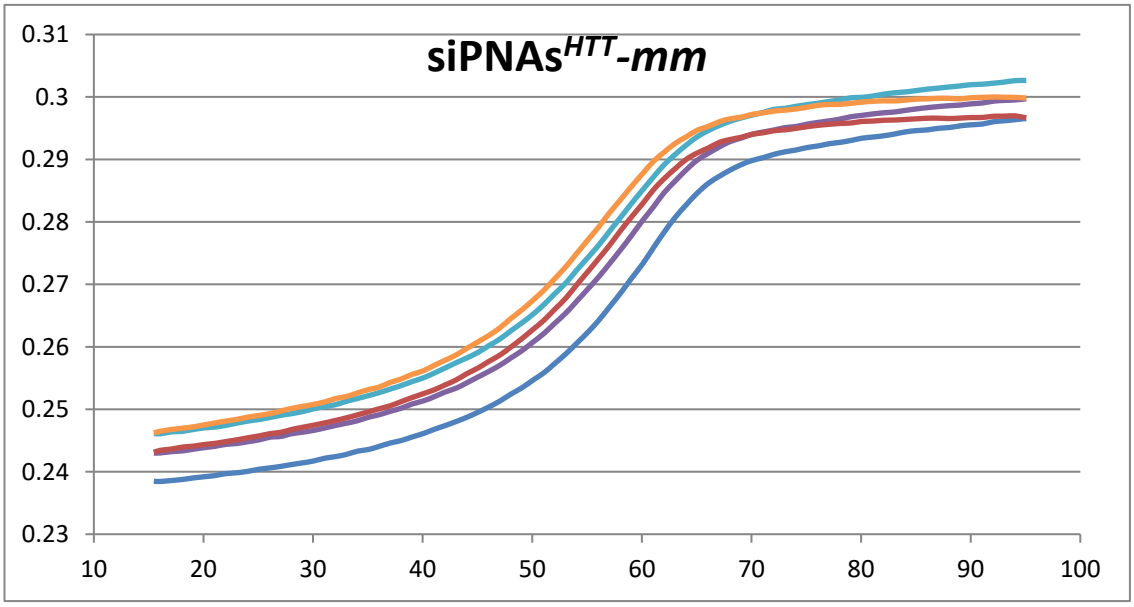

pLNA
0.5
0.48
0.46
0.44
0.42
0.4
0.38
10
20
30
40
50
60
70
80
90
100

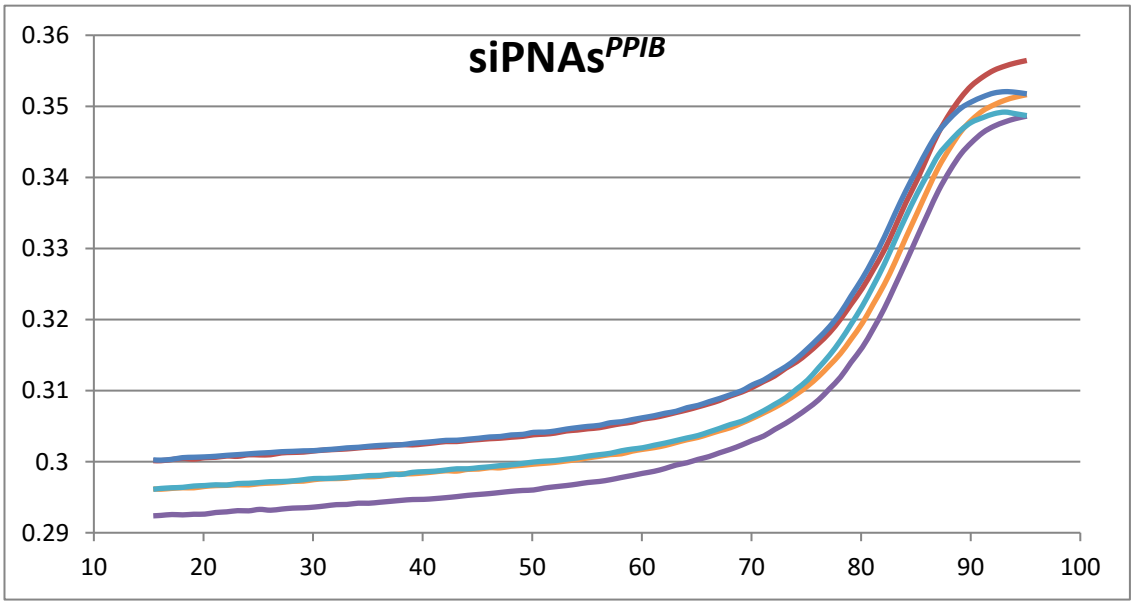

bLNA
0.43
0.41
0.39
0.37
0.35
0.33
10
20
30
40
50
60
70
80
90
100

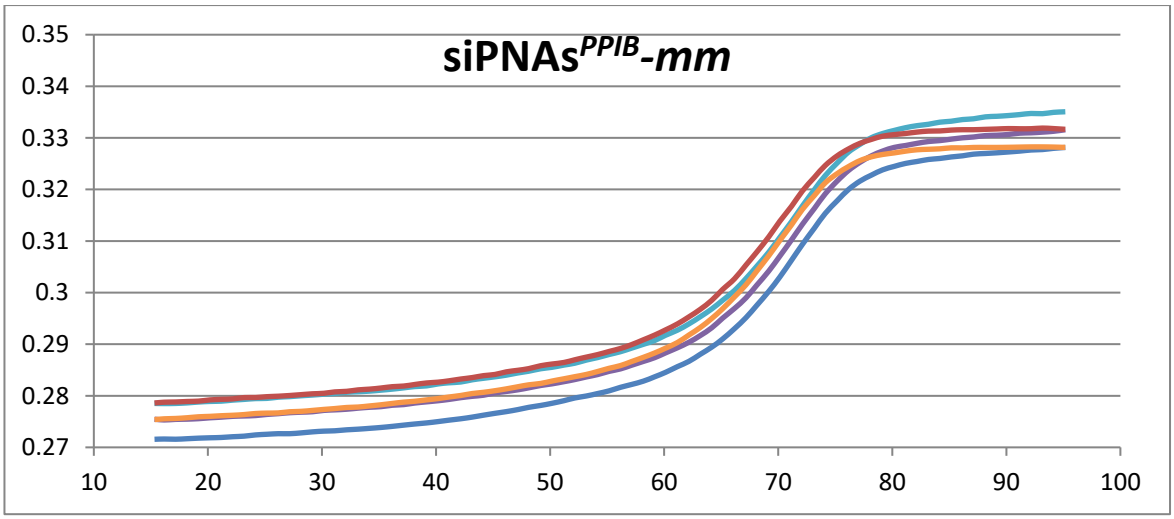

eFANA
0.43
0.42
0.41
0.4
0.39
0.38
0.37
0.36
0.35
0.34
10
20
30
40
50
60
70
80
90
100

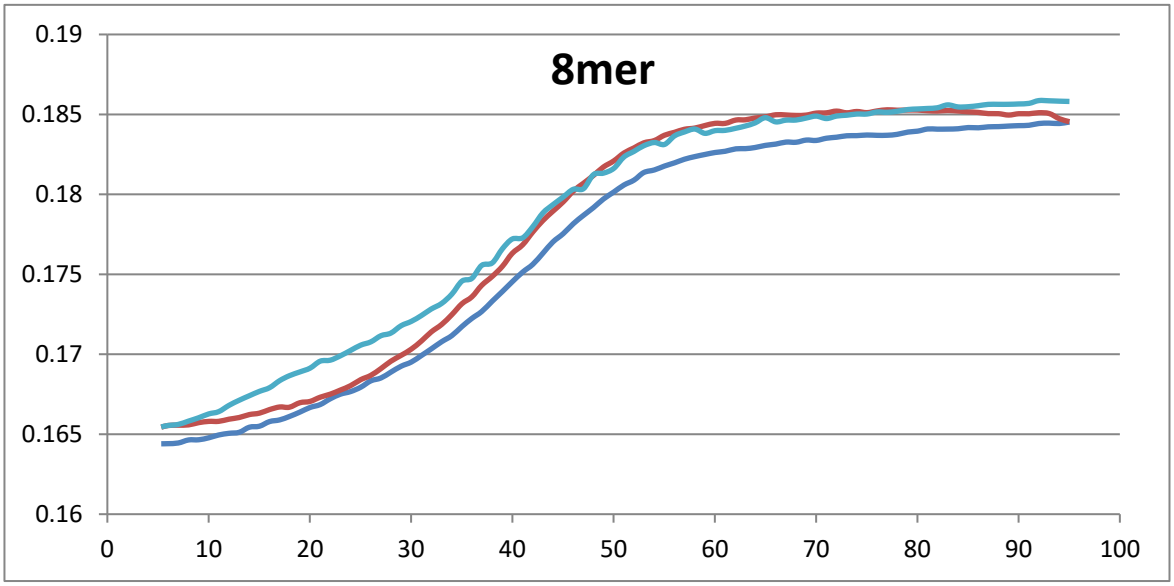

bFANA
0.47
0.45
0.43
0.41
0.39
0.37
0.35
10
20
30
40
50
60
70
80
90
100

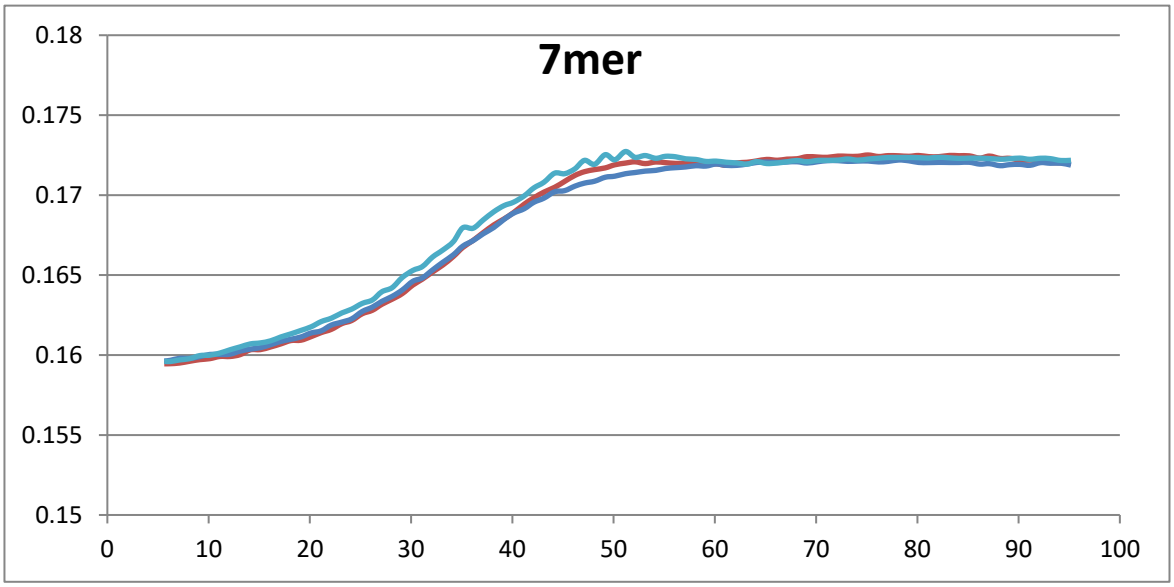

siPNAs^HTT
0.35
0.33
0.31
0.29
0.27
0.25
10
20
30
40
50
60
70
80
90
100

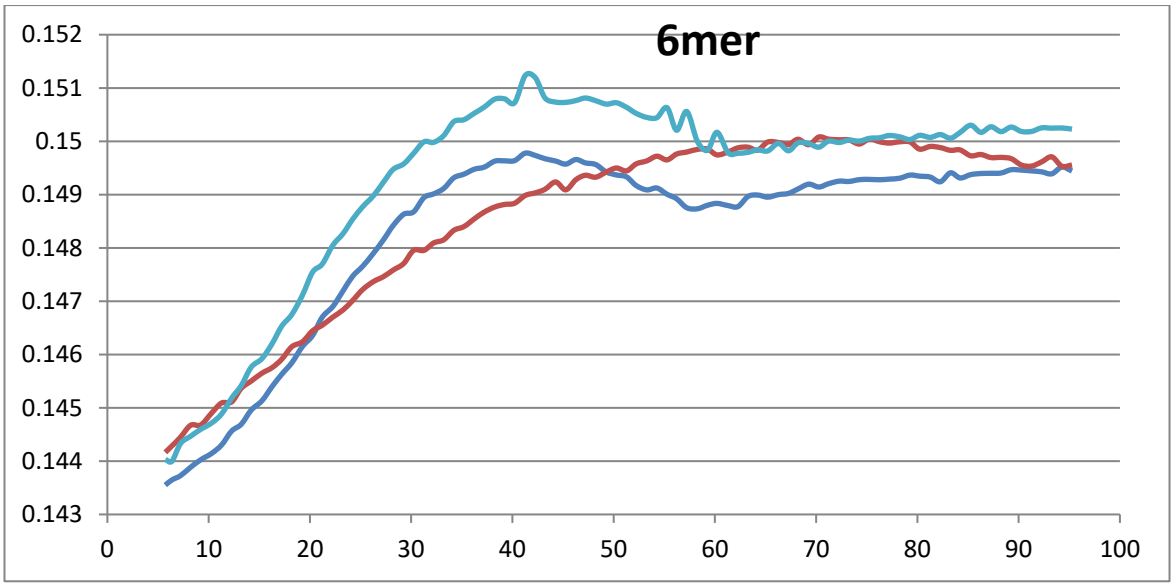

siPNAs^HTT-mm
0.31
0.3
0.29
0.28
0.27
0.26
0.25
0.24
0.23
10
20
30
40
50
60
70
80
90
100

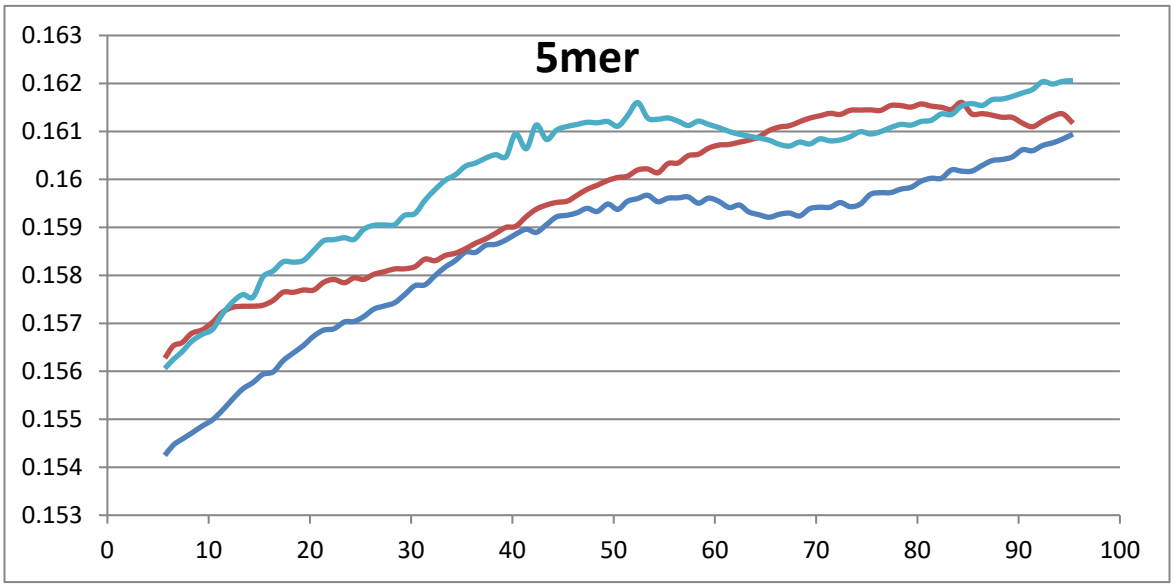

siPNAs^PPIB
0.36
0.35
0.34
0.33
0.32
0.31
0.3
0.29
10
20
30
40
50
60
70
80
90
100

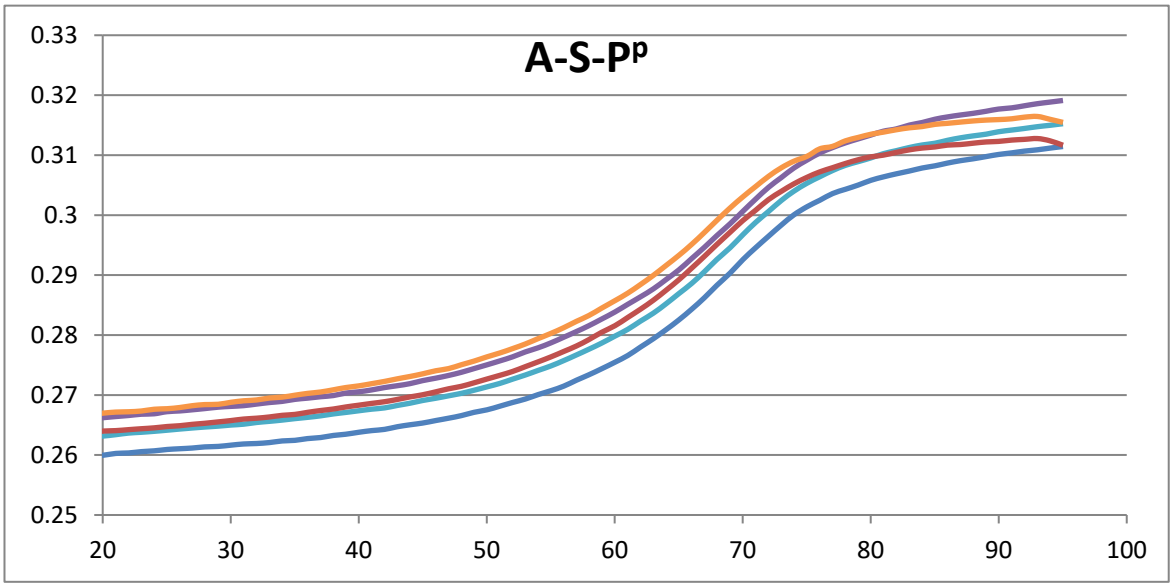

siPNAs^PPIB-mm
0.35
0.34
0.33
0.32
0.31
0.3
0.29
0.28
0.27
10
20
30
40
50
60
70
80
90
100

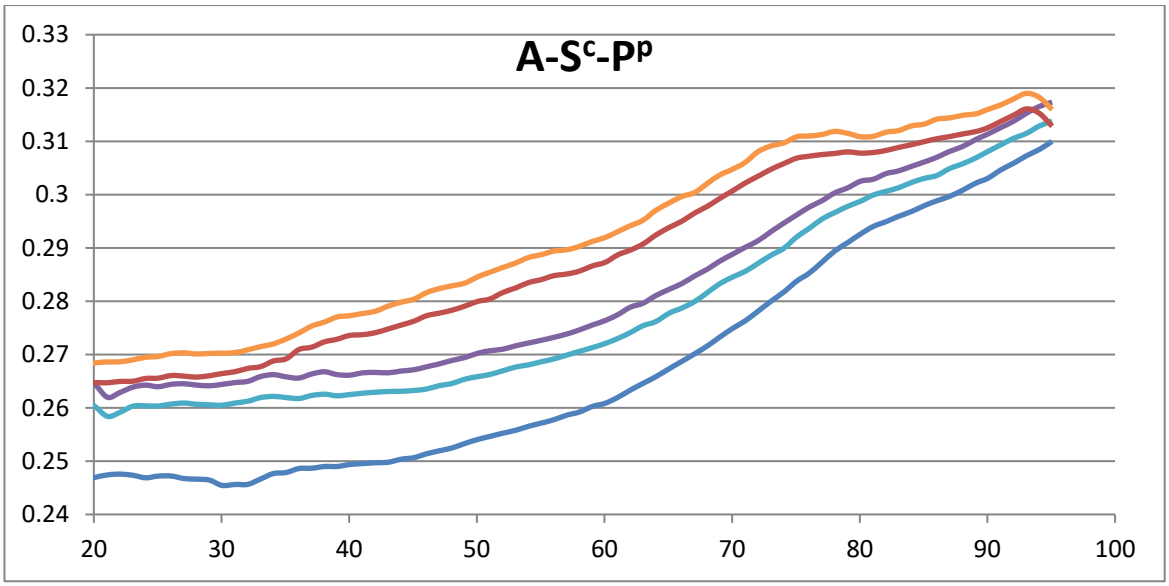

8mer
0.19
0.185
0.18
0.175
0.17
0.165
0.16
0
10
20
30
40
50
60
70
80
90
100

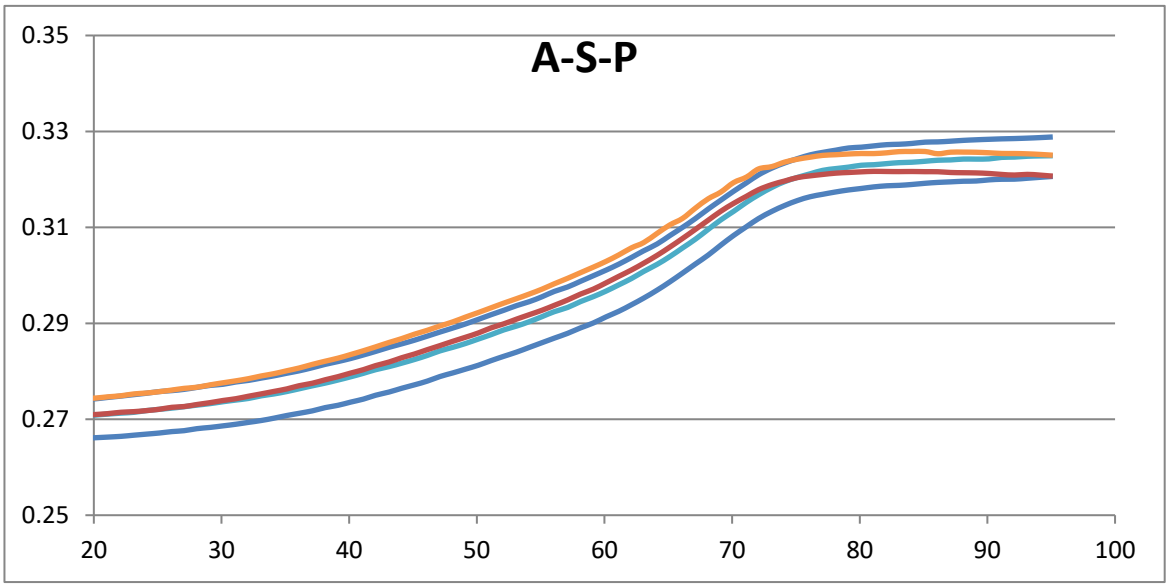

7mer
0.18
0.175
0.17
0.165
0.16
0.155
0.15
0
10
20
30
40
50
60
70
80
90
100

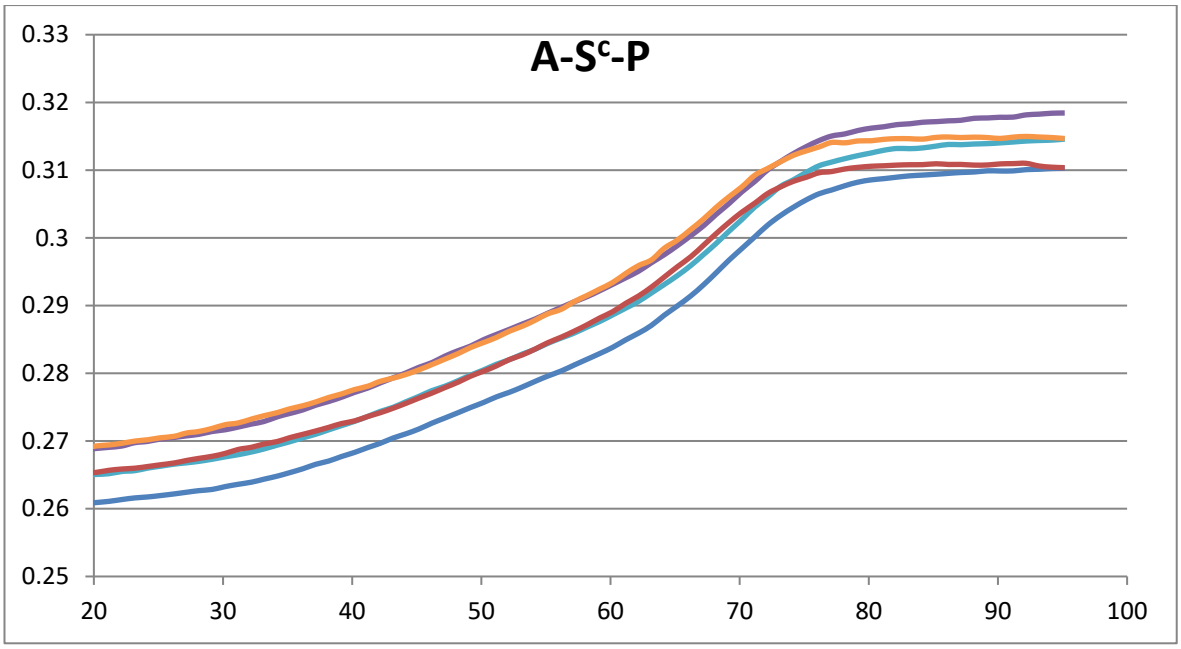

6mer
0.152
0.151
0.15
0.149
0.148
0.147
0.146
0.145
0.144
0.143
0
10
20
30
40
50
60
70
80
90
100

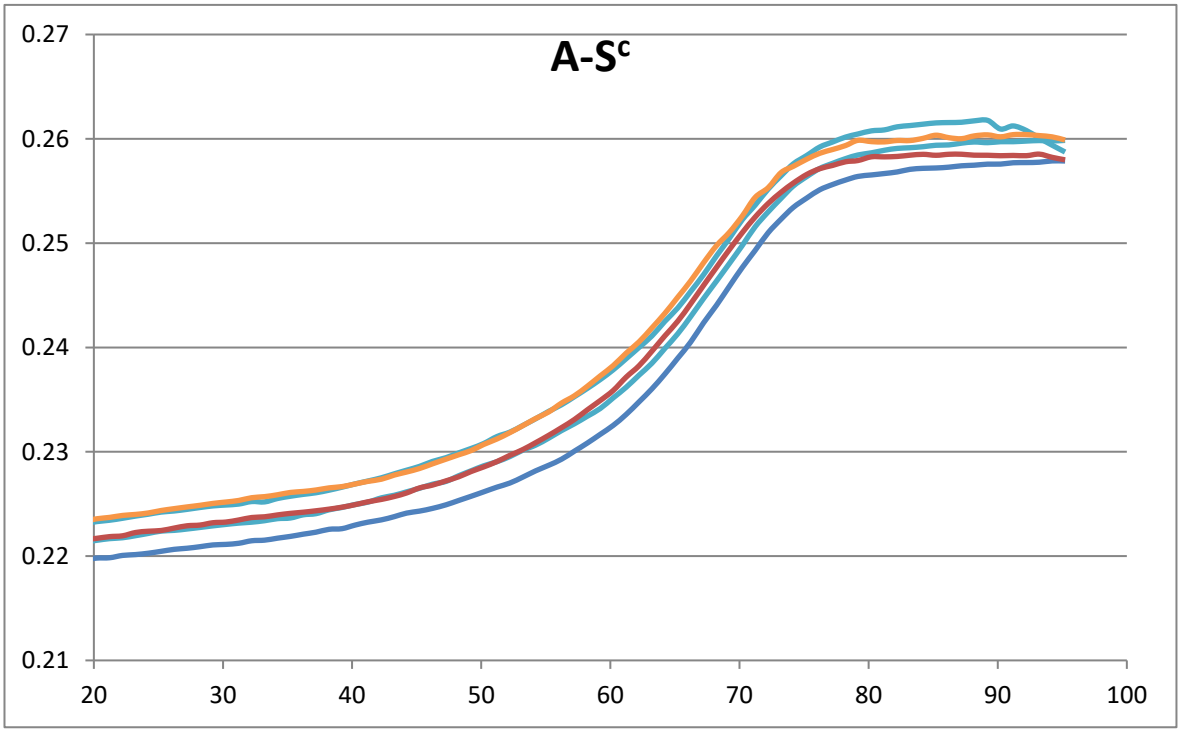

5mer
0.163
0.162
0.161
0.16
0.159
0.158
0.157
0.156
0.155
0.154
0.153
0
10
20
30
40
50
60
70
80
90
100

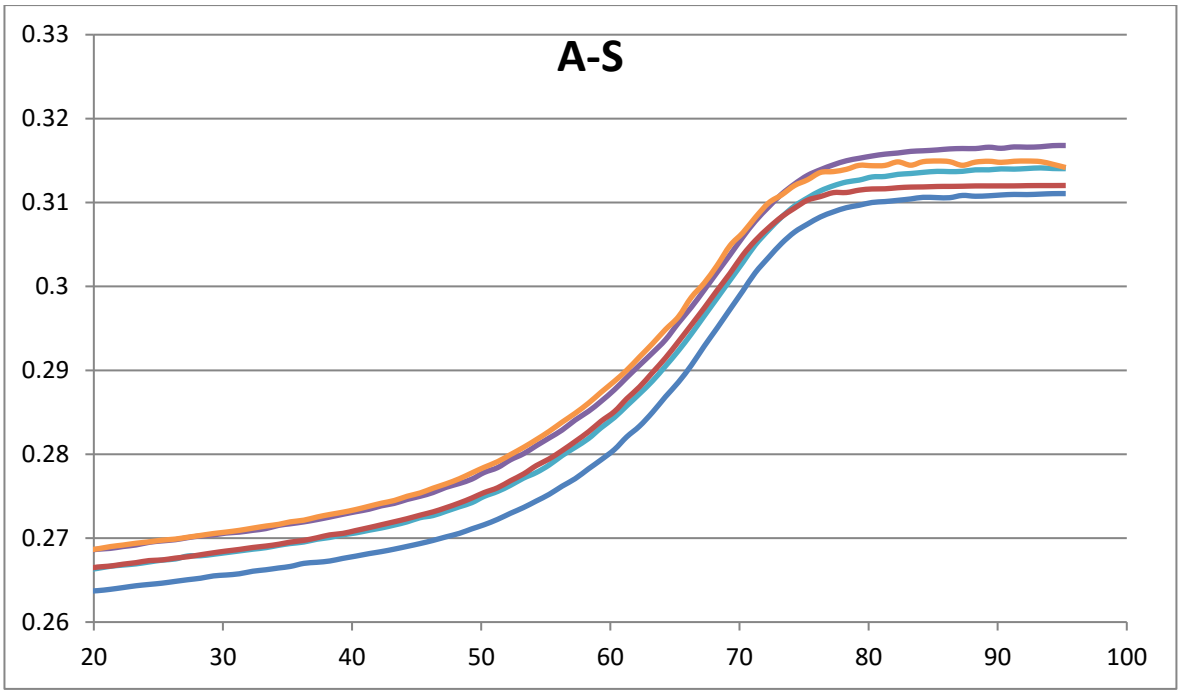

A-S-Pp
0.33
0.32
0.31
0.3
0.29
0.28
0.27
0.26
0.25
20
30
40
50
60
70
80
90
100


A-S^c-P^p
0.33
0.32
0.31
0.3
0.29
0.28
0.27
0.26
0.25
0.24
20
30
40
50
60
70
80
90
100


A-S-P
0.35
0.33
0.31
0.29
0.27
0.25
20
30
40
50
60
70
80
90
100


A-Sc-P
0.33
0.32
0.31
0.3
0.29
0.28
0.27
0.26
0.25
20
30
40
50
60
70
80
90
100


A-Sc
0.27
0.26
0.25
0.24
0.23
0.22
0.21
20
30
40
50
60
70
80
90
100


A-S
0.33
0.32
0.31
0.3
0.29
0.28
0.27
0.26
20
30
40
50
60
70
80
90
100