# Synthesis, Structure and Application of Backbone Modified Nucleic Acids




Cameron Thorpe

Wolfson College, University of Oxford

A thesis submitted for the degree of Doctor of Philosophy January 2021

# Declaration

I, Cameron Thorpe, declare that the thesis titled “Synthesis, Structure and Application of Backbone Modified Nucleic Acids” and the relating research presented in this thesis was performed and written by the author under the supervision of Prof. Tom Brown at the University of Oxford. The research is original and has not been submitted in full or in part for any other degree at this or at any other university. The copyright of this thesis rests with the author.

# Acknowledgements

I am extremely grateful to everyone who has made this thesis possible. I would like to thank Sven Epple and Benjamin Woods for their hard work and contribution to the carbamate project. I would also like to thank Sven for is invaluable scientific opinion and inventive ideas. I extend my greatest thanks to Ashley Tsai for her incredible work towards the triazole project. Without her ground-breaking achievements this work would not have been possible, she is a fabulous scientist and person. I would like to express my gratitude to Ysobel Baker for her guidance over the years and invaluable contribution to my work. Thank you to Lapitrada (Pui), Nicolo and Sven for proofing this thesis and finding all those terrible spelling mistakes. Thank you to Sarah, Pui, Nicolo, Phoebe, Thanishta, Piotr, Lillian, Jack H, Jack C and all the Brown group for their support and guidance throughout my PhD. I would also like to thank Tom and Afaf for being supportive supervisors, allowing me to make mistakes, and providing a welcoming environment to develop scientific ideas. Tom's guidance as a supervisor has made my PhD an enjoyable experience and helped me to become a much better researcher.

I have been lucky enough to have a group of fantastic friends and family who have supported me throughout my studies. Thank you Ravinda, Florian, John, Laura, Tash, Ellie, Charlie, Matt, Sam, Poppy and the legends James and Jimmy for an unforgettable time in Oxford. Thank you to all the gang in Hurst street and everyone on the DTP. I would also like to thank Louise, Charlie and Tom who have been so supportive over the years and are such great friends.

Most importantly I would like to thank my family for their everlasting love and guidance. Mum and Dad have been the most supportive parents anyone could have. Submitting this PhD is not only my achievement but something I would like to share with them. Without their support it would never have happened. Finally, I would like to thank my sister Louise. An absolute gem and the funniest person I know, she is always able to bring a unique perspective and is someone whose opinion always matters.

# Abbreviations

| | |
|---|---|
| A | Adenosine/Adenine |
| AAs | Amino acids |
| Ac | Acetate |
| ADAR | Adenosine deaminases acting on RNA |
| Ago2 | Argonaute-2 |
| Am | Amide |
| AMP | Adenosine monophosphate |
| AMPS | Adenosine 5'-mono phosphorothioate |
| APS | Ammonium persulfate |
| ASOs | Antisense oligonucleotides |
| ATP | Adenosine triphosphate |
| AZT | Azidothymidine |
| B | General nucleobase |
| bp | Base pair |
| Bz | Benzoyl |
| C | Cytosine |
| CB | Carbonate |
| CBM | Carbamate |
| $CC_{1/2}$ | Correlation coefficient-half |
| CD | Circular dichroism |
| CDI | 1,1'-carbonyldiimidazole |
| CEP | Chloro(diisopropylamino)-β-cyanoethoxyphosphine |
| circRNA | Circular RNA |
| CPG | Controlled pore glass |
| CRISPR | Clustered regularly interspaced short palindromic repeats |
| CuAAC | Copper-catalysed azide–alkyne cycloadditions |
| DIPEA | N,N-Diisopropylethylamine |
| DMD | Duchenne muscular dystrophy |
| DMF | Dimethylformamide |
| DMSO | Dimethyl sulfoxide |
| DMTr | 4,4'-Dimethoxytrityl |
| DNA | Deoxyribonucleic acid |
| eIFx | Eukaryotic initiation factor (x = numerical value) |
| ELISA | Enzyme-linked immunosorbent assay |
| EMCV | Encephalomyocarditis virus |
| ESI | Electrospray ionisation |
| FBS | Foetal bovine serum |
| FDA | Food and Drug Administration |
| G | Guanine |
| GFP | Green fluorescent protein |
| HPLC | High performance liquid chromatography |
| HRMS | High resolution mass spectrometry |
| HRP | Horseradish peroxidase |
| HT | High throughput |
| IF | Initiation factor |
| I-ORF | Infinite open reding frame |
| IRES | Internal ribosome entry site |

| | |
|---|---|
| $K_D$ | Binding dissociation constant |
| LCMS | Liquid chromatography–mass spectrometry |
| LLG | Log-likelihood gain |
| LNA | Locked nucleic acids |
| MBL | Muscle blind protein |
| MES | 2-(N-morpholino)ethanesulfonic acid |
| miRNA | Micro RNA |
| MMTr | Monomethoxytrityl |
| MOE | 2-Methoxyethyl |
| MPD | 2-Methyl-2,4-pentanediol |
| MR | Molecular replacement |
| mRNA | Messenger RNA |
| MS | Mass spectrometry |
| Ms | Mesyl |
| MTPI | Methyltriphenoxyphosphonium iodide |
| NMR | Nuclear magnetic resonance |
| NTPs | Nucleotide triphosphates |
| ON | Oligonucleotide |
| P | Phosphate |
| PAGE | Polyacrylamide gel electrophoresis |
| PBS(-T) | Phosphate buffered saline (0.1% tween-20) |
| PCR | Polymerase chain reaction |
| PDB | Protein data bank |
| PG | Protecting group |
| Ph | Phenyl |
| PMO | Phosphorodiamidate Morpholino oligomer |
| PNA | Peptide nucleic acid |
| PNP | *p*-Nitrophenyl |
| PO | Phosphodiester |
| PS | Phosphorothioate |
| Ps | Polystyrene |
| PV | Poliovirus |
| PVDF | Polyvinylidene fluoride |
| qPCR | Quantitative polymerase chain reaction |
| r(N) | RNA sequence with N nucleotides |
| r.m.s(.d) | Root mean square (deviation) |
| RCT | Rolling circular translation |
| RISC | RNA-induced silencing complex |
| RNA | Ribonucleic acid |
| RNAi | Ribonucleic acid interference |
| rt | Room temperature |
| SD | Shine Dalgarno |
| SDS | Sodium dodecyl sulfate |
| SPS | Solvent purification system |
| siRNA | Small interfering RNA |
| T | Thymine |
| TBAF | Tetra-n-butylammonium Fluoride |
| TBDMS | *Tert*-butyldimethylsilane |
| TBE | Tris/Borate/EDTA buffer |

| | |
|---|---|
| TBS(-T) | Tris buffered saline (0.1% Tween-20) |
| TC | 2’-Thiomorpholine-4-carbothioate |
| TCA | Trichloroacetic acid |
| TEAB | Triethylammonium bicarbonate buffer |
| TEMED | Tetramethylethylenediamine |
| TFZ | Translation function Z-score |
| THF | Tetrahydrofuran |
| THPTA | Tris(3-hydroxypropyltriazolylmethyl) amine |
| TLC | Thin layer chromatography |
| Tm | Melting temperature |
| TMB | 3,3’,5,5’-Tetramethylbenzidine |
| tRNA | Transfer ribonucleic acid |
| Tz | Triazole |
| U | Uracil |
| UNA | Unlocked nucleic acid |
| UPLC | Ultra-performance Liquid Chromatography |
| UV | Ultraviolet |
| XRD | X-ray diffraction |

# Abstract

Modified backbones are versatile modifications which can improve biological half-lives, have good structural biocompatibility and can alter the biophysical properties of oligonucleotides. This work presents the synthesis and biophysical properties of oligonucleotides modified with carbamate-locked nucleic acids (CBM-LNA) for antisense applications, the crystallisation and structural analysis of amide-locked nucleic acids (LNA-amides) and explores peptide translation using modified triazole backbones in circular mRNAs (Tz-circRNA). From these studies it is shown that CBM-LNAs present viable templates for potential antisense applications, that LNA-amides are good structural mimics as phosphodiesters and that Tz-circRNAs are viable templates for protein expression.

Carbamate backbones are four-atom, charge neutral linkages that were previously found to be duplex destabilising and unsuitable for antisense applications. In this work, we combine the enzymatic stability of the CBM backbone with the thermodynamic stability of locked nucleic acids (LNA). CBM-LNA phosphoramidite dimers were synthesised and incorporated into oligonucleotide sequences. Biophysical characterisation using UV melting, showed that LNA-$CBM_1$-LNA modifications were able to restore duplex stability against RNA targets and that LNA-CBMs have enhanced enzymatic resistance. Compared to previous CBM modifications, LNA-CBMs now present suitable modifications for potential antisense applications.

LNA-amides have good thermodynamic properties which makes them suitable candidates for therapeutic applications. The aim of our work is to show that LNA-amides are good structural mimics as phosphodiesters and suggest the features that provide LNA-amides with good biophysical properties. XRD crystallography was used to obtain three different LNA-amide modified oligonucleotides between 2.5-2.8 Å resolution. Analysis of these structures showed that LNA-amides are structurally alike to phosphodiesters and that they cause minimal duplex disruption. The amide and LNA modifications were found to work harmoniously together. To provide good duplex stability, modified backbones should have good flexibility towards their 5'-end and should consist of four-atoms in length if they contain a carbonyl functionality.

Triazole backbones have good biocompatibility for replication and transcription mechanisms. In this work we aim to complete the central dogma and translate peptides using triazole modified circular mRNA templates. This will allow for the chemical synthesis of circular mRNAs > 100 nt without the need for enzymes, plasmid preparation or bacterial cell culture. Linear precursor RNAs were modified with 5'-azide and 3'-alkyne functional groups. These were then subjected to the CuAAC reaction to form a triazole linkage contained within the coding region of a circular mRNA. Encoded with FLAG epitopes, the triazole circular RNAs were used in cell free prokaryotic protein translation and analysed by immunoblotting techniques. From these results it is shown that triazoles can be successfully translated by the prokaryotic ribosome and yield up to 50% as much peptide as an unmodified circular RNA.

# Contents

1. Introduction to Nucleic acids and experimental techniques ..... 1

1.1 Nucleic acid structures ..... 1

1.1.1 Nucleic acid primary structures ..... 1

1.1.2 Oligonucleotide duplex structures ..... 3

1.2 Synthesis of oligonucleotides ..... 8

1.2.1 The phosphoramidite method ..... 8

1.2.2 Solid phase synthesis cycle details ..... 10

1.2.3 Synthesis of RNA oligonucleotides ..... 12

1.2.4 Oligonucleotide synthesis using the H-phosphonate method ..... 14

1.3 Ligation of oligonucleotides ..... 15

1.3.1 Enzymatic ligation ..... 16

1.3.2 Chemical ligation using the CuAAC reaction ..... 17

1.4 Biophysical analysis of oligonucleotides ..... 19

1.4.1 UV-thermal melting ..... 19

1.4.2 Circular dichroism ..... 20

1.4.3 Gel electrophoresis ..... 21

1.4.4 Antibodies and Immunoassays ..... 23

1.5 Introduction to modified backbones ..... 25

1.6 Introduction to Locked Nucleic Acids (LNA) ..... 28

1.7 Principles of X-ray diffraction (XRD) and crystallography ..... 29

1.7.1 Defining a crystal lattice ..... 29

1.7.2 XRD and obtaining an electron density map ..... 30

1.8 Aims and objectives of this work ..... 33

2 Synthesis and biophysical properties of Carbamate-Locked nucleic acids for antisense therapeutics ..... 34

2.1 Antisense, modifications and therapeutics....34
2.1.1 Antisense mechanisms....34
2.1.2 Antisense modifications and drugs....37
2.2 Aims and objectives to develop LNA-carbamates as potential antisense therapeutic modifications....42
2.3 Synthesis of LNA-$CBM_1$ dinucleotides....43
2.3.1 $CBM_1$ electrophiles....43
2.3.2 $CBM_1$ nucleophiles....45
2.3.3 $CBM_1$ phosphoramidites....46
2.4 Synthesis of LNA-$CBM_2$ dinucleotides....47
2.4.1 $CBM_2$ nucleophile....47
2.4.2 $CBM_2$ electrophiles....48
2.4.3 $CBM_2$ phosphoramidites....49
2.5 Synthesis of LNA-CBM oligonucleotides....50
2.6 Duplex hybridisation studies – $T_m$....52
2.6.1 Comparing $CBM_1$ and $CBM_2$ backbones....52
2.6.2 LNA-$CBM_1$ modifications against DNA and RNA targets....54
2.6.3 LNA-$CBM_2$ modifications....55
2.7 Circular dichroism....55
2.8 Enzymatic stability assays....56
2.9 Summary of LNA-CBM modified oligonucleotides....60
3 XRD structures of LNA-Amide modified DNA-RNA Duplexes....61
3.1 Introduction to amides as modified backbones....61
3.1.1 Amide backbones and synthesis....61
3.1.2 Amide backbone and duplex stability....63
3.2 Introduction to XRD macromolecular crystallography....64

3.2.1 Application of crystallography for oligonucleotide structures ..... 64
3.2.2 Introduction to crystal growth ..... 65
3.2.3 Data collection and processing – indexing, scaling and merging ..... 66
3.2.4 Phasing, model building and refinement ..... 67
3.3 Aims and objectives of LNA-amide crystallography ..... 67
3.4 LNA-amide oligonucleotide synthesis and initial screens ..... 68
3.4.1 Oligonucleotide sequence selection ..... 68
3.4.2 Synthesis of the LNA-amide precursor compounds and amide solid phase synthesis ..... 69
3.4.3 Crystallisation of LNA-backbone modified oligonucleotides using high throughput screens ..... 71
3.4.4 High throughput screen data collection ..... 72
3.4.5 Crystallisation condition optimisation using an unmodified duplex ..... 74
3.4.6 Obtaining crystals for LNA-Am-LNA structures ..... 77
3.4.7 Crystallisation of oligonucleotides with LNA-Am-LNA at different positions ..... 78
3.4.8 Summary of crystallisation conditions ..... 79
3.5 Data Processing of LNA-amide structures ..... 80
3.6 XRD model building ..... 81
3.7 Analysis of DNA:RNA structures with amides and LNA modifications ..... 84
3.7.1 Summary of model statistics ..... 84
3.7.2 Comparison to published structures ..... 86
3.7.3 Comparing the effect of LNA-amide modification on duplex structure ..... 88
3.7.4 Structure and orientation of the amide backbone ..... 89
3.7.5 Analysis of Amides and LNAs ..... 92
3.7.6 Analysis of base pair parameters ..... 94
3.7.7 Analysing combined LNA and amide modifications ..... 100

3.7.8 Analysis of localised hydrogen bonding....100
3.8 XRD structures with carbamate-LNA modifications....102
3.8.1 Phasing the CBM-LNA structure using molecular replacement....102
3.8.2 CBM-LNA model building....103
3.8.3 CBM-LNA duplex structure....104
3.8.4 CBM-LNA duplex crystal packing and special positions....106
3.8.5 Base pairing around the terminal junction....107
3.8.6 CBM-LNA backbone structure....108
3.9 Summary of LNA-amide and CBM-LNA modified XRD structures....111
4 Synthesis and translation of triazole modified circular RNA....114
4.1 Introduction and synthesis of circular RNAs....114
4.1.1 Introduction and biosynthesis of circRNA....114
4.1.2 In vitro synthesis of circRNA....117
4.2 RNA translation and translation of circRNA....118
4.2.1 Translation initiation....118
4.2.2 Translation elongation....120
4.2.3 Translation termination....120
4.2.4 Prokaryotic translation....121
4.2.5 Translation of circRNA....122
4.3 Triazole backbones....123
4.4 Aims and objectives of triazole circRNA....127
4.5 RNA template design and oligonucleotide synthesis....129
4.5.1 CircRNA template design....129
4.5.2 Synthesis of modified linear RNA precursors....130
4.6 Synthesis of unmodified circRNA....132
4.7 Synthesis of Tz-circRNA and Tz-LNA-circRNA....140

4.8 Cell free protein expression from triazole circRNAs in prokaryotic systems.....145
4.9 Synthesis of 5′-azides using H-phosphonates.....148
4.10 Synthesis and translation of circRNAs containing multiple triazole backbones.....152
4.11 Quantification of peptide yield from modified circRNAs.....155
4.12 Tz-circRNA future direction and applications.....162
4.13 Summary of translation from Tz-circRNA.....163
5 Conclusions and future work.....166
6 References.....170
7 Appendix and supplementary information.....184
7.1 CBM-LNA additional figures and information.....184
7.2 LNA-amide crystallization conditions and additional information.....186
7.3 Tz-circRNA supplementary information and sequences.....192
8 Experimental.....200
8.1 General procedures.....200
8.2 Oligonucleotide synthesis.....201
8.2.1 Synthesis of DNA oligonucleotides.....201
8.2.2 Synthesis of RNA oligonucleotides.....201
8.2.3 Removal of 2′-O-TBDMS protecting groups on RNA oligonucleotides.....202
8.2.4 Removal of 2′-O-TC protecting groups on RNA oligonucleotides.....202
8.2.5 Oligonucleotide 5′-Azide conversion.....202
8.2.6 Coupling of H-Phosphonates to resin bound oligonucleotides.....202
8.2.7 T4 DNA ligase templated ligation of RNA using a DNA splint.....203
8.2.8 CuAAC templated ligation of RNA using a DNA splint.....203
8.3 Purification and analysis of oligonucleotides.....203
8.3.1 HPLC Purification of oligonucleotides.....203
8.3.2 Denaturing poly acrylamide gel electrophoresis (PAGE).....204

8.3.3 Native poly acrylamide gel electrophoresis....204
8.3.4 Gel extraction and NAP gel filtration of oligonucleotides....204
8.3.5 Size exclusion spin column....204
8.3.6 Phenol: chloroform extraction....205
8.4 Biophysical studies and assays....205
8.4.1 Ultraviolet melting studies (duplex hybridisation - $T_m$)....205
8.4.2 Circular dichroism studies....205
8.4.3 Phosphodiesterase 1 assay (Snake venom)....205
8.4.4 Foetal bovine serum (FBS) assay....206
8.4.5 PURExpress cell free prokaryotic RNA translation kit (NEB PURExpress E6800) 206
8.4.6 Tris-glycine SDS-PAGE....206
8.4.7 Dot blot....206
8.4.8 Western blot....207
8.4.9 ELISA assay....207
8.5 X-Ray diffraction procedures and model refinement....208
8.5.1 Crystallisation high throughput screens....208
8.5.2 Crystallisation manual optimisation screens....208
8.5.3 Crystallisation conditions....209
8.5.4 Crystal collection, storage and transport....209
8.5.5 Data collection (Diamond light source)....210
8.5.6 Data processing and model building....210
8.6 Table of buffers....211
8.7 Synthesis and characterisation of chemical compounds....212

# 1. Introduction to Nucleic acids and experimental techniques

## 1.1 Nucleic acid structures

### *1.1.1 Nucleic acid primary structures*

Deoxyribose nucleic acid (DNA) is the molecular storage of genetic information and encodes for thousands of genes within the human genome. Nucleic acids are polymeric assemblies consisting of building blocks called nucleosides. There are four different types of DNA nucleoside: thymidine, deoxyadenosine, deoxycytidine and deoxyguanosine (Figure 1. 1). When nucleosides are combined with an additional 5'-phosphate group, they form nucleotides. These nucleotides are comprised of a deoxyribose, phosphate and nucleobase groups.

DNA has four nucleobases (Figure 1. 1). They are divided into two classes, which are the pyrimidines thymine (T) and cytosine (C), and the purines adenine (A) and guanine (G). The atomic positions within these nucleobases are defined by a numerical system (pyrimidines: 1-6, purines: 1-9) and deoxyribose sugar nomenclature in a nucleoside consists of a number and prime notation (1'-5'). The nucleobases are linked to the deoxyribose through a glycosidic bond at the C1' position. Depending on the orientation of the base relative to the deoxyribose, the nucleobase can be positioned above the sugar, in the *syn* conformation, or rotated 180 ° into the *anti*-conformation (Figure 1. 1). In DNA the *anti*-conformation normally predominates. The 1'-glycosidic position is also known as the anomeric centre, which describes the orientation of the nucleobase relative to the plane of the sugar ring. When the nucleobase is positioned on the same face as the 5'-carbon, the deoxyribose is in the β-anomeric configuration, which is the structure found in natural DNA. If the nucleobase is positioned on the opposing face to the 5'-carbon, then the nucleoside is defined as the α-anomer.

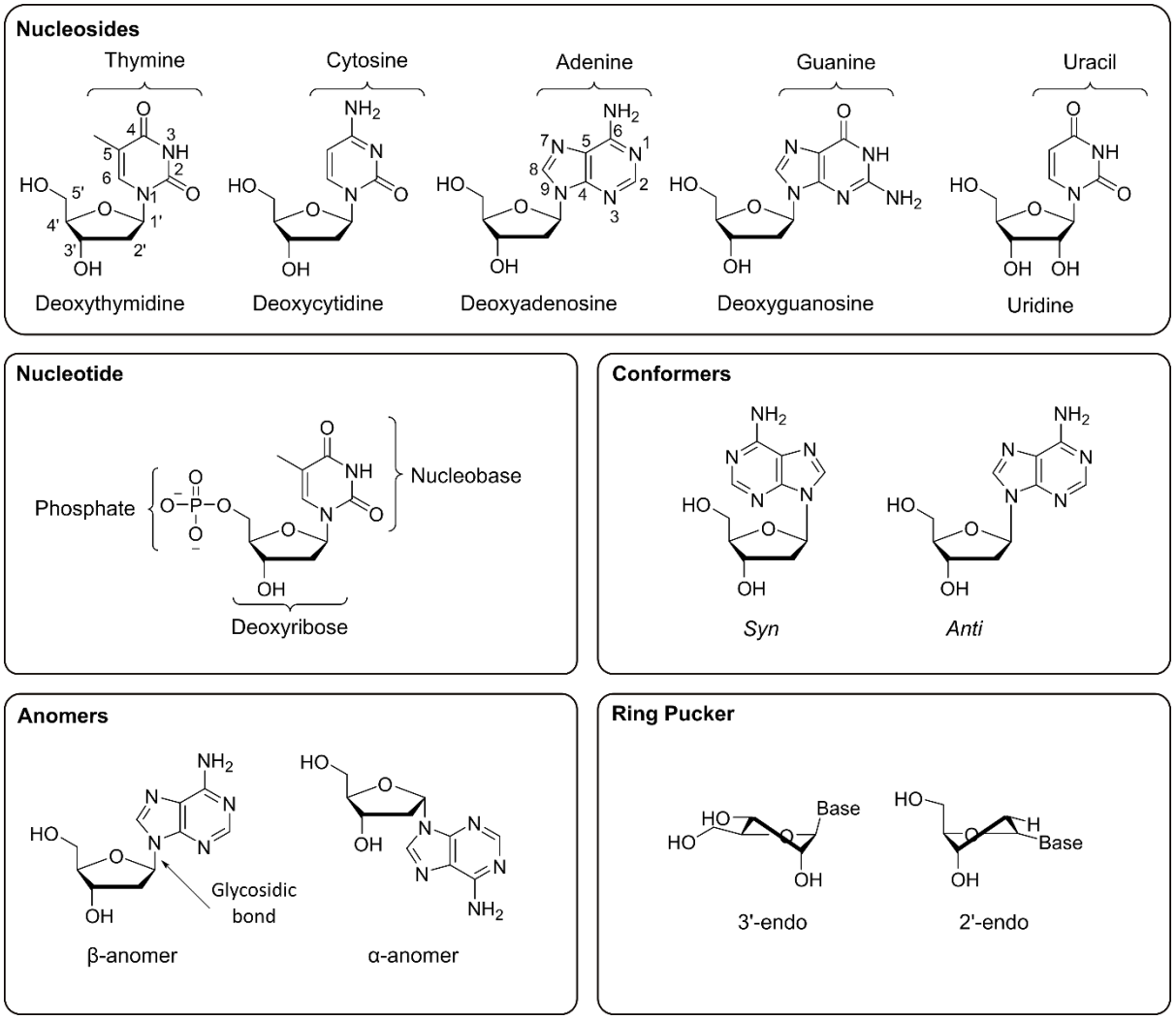



*Figure 1. 1 Overview of the primary structures of nucleic acid monomers. Nucleosides: DNA; deoxyribose nucleosides with their corresponding nucleobase, also shown is the RNA nucleoside Uracil, a defining feature of RNA with its additional 2'-hydroxyl group. Nucleotide: General structure of a nucleotide containing the three constituent parts defined as the phosphate, deoxyribose and nucleobase. Conformers: Conformations that are available to the nucleobase by rotation around the glycosidic bond. Anomers: Two anomeric forms are found for nucleic acids, the β-form is defined when the nucleobase is on the same face as the 5'-position, the α-form is when they are on opposing faces. Ring pucker: Conformations of 5-member rings showing the envelope structures of the 3'-endo and 2'-endo conformations adopted by DNA and RNA sugars.*

While DNA is the molecular storage of cellular information, Ribonucleic Acids (RNA) is responsible for cellular expression and gene regulation. RNA is divided into various types such as ribosomal RNA (rRNA), messenger RNA (mRNA) and micro RNAs (miRNA). Although structurally alike to DNA, RNA can be distinguished by the presence of an additional 2'-hydroxyl group located on the ribose. Alongside this, RNA does not contain any thymine (T) bases, which are replaced with a similar uracil (U) base. The 2'-hydroxyl group has a significant impact on the sugar conformation ("pucker"). Five-membered saturated rings form envelope structures that push one of the atomic positions out of the plane relative to the other members within the ring (Figure 1. 1). This reduces ring strain, in turn reducing the energy of

the system. Puckers that are above the ring on the same face as the 5’-position are defined as *endo*, whereas puckers on the opposing face are described as *exo*. DNA is usually in the 2’-*endo* conformation which means that the 2’-carbon is pushed out into the envelope position. Although the 2’-*endo* is the most common and energetically favoured conformation, potential energy plots of all sugar puckers, show that DNA can access many different sugar conformations, and these are often observed. By contrast, the sugars of canonical RNA are almost exclusively found in the 3’-endo conformation. This is due to the influence of the 2’-hydroxyl group.

In DNA and RNA, the 5'-phosphate group of a nucleotide is connected to the 3'-hydroxyl of another. Repetition of this process allows for the formation of DNA chains. DNA is chiral and has a directionality, which by convention is described from the 5’ to 3’ end. For example, the sequence 5’-GCAT-3’ (Figure 1. 2) defines the position of each nucleotide in the oligonucleotide tetramer chain. By using different combinations of nucleotides, a huge variety of DNA sequences can be formed, each having unique chemical and biophysical properties.

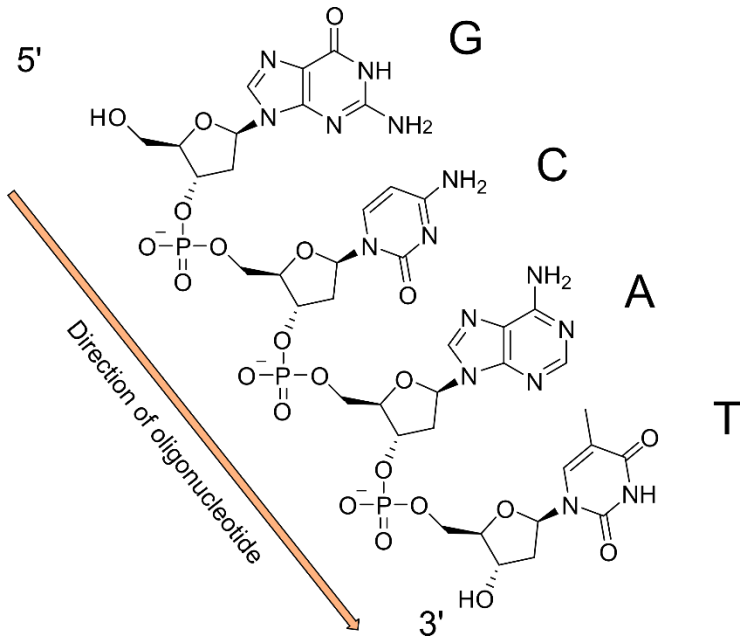



*Figure 1. 2 Oligonucleotide sequence 5’-GCAT-3’ that contains a phosphodiester backbone which links the nucleotide monomers to form an oligomeric chain. The sequence GCAT is defined as that running from the 5’ to 3’ direction.*

### *1.1.2 Oligonucleotide duplex structures*

Early analysis of DNA extracted from Sea-Urchins by Erwin Chargaff examined the molecular composition of isolated nucleic acids.[1] In this work, he determined that adenine molar ratios were always equal to thymine and that the same result was observed between cytosine and

guanine. This suggested that they are in some way linked. The duplex structure of nucleic acids was first proposed by Watson and Crick[2], and suggested that the DNA structure had two helical chains, coiled around a single axis. This was based on work by Rosalind Franklin using X-ray fibre diffraction, who found that DNA has two unique helical structures highly dependent on humidity.[3,4] Franklin's finding that DNA hydration influences DNA structure, suggests that the phosphate groups must be interacting with solvent, and therefore be positioned on the exterior of the DNA.

In Watson and Crick's structure, nucleic acid duplexes are formed by two separate chains that run anti-parallel to each other. The interior is composed of stacked complementary A-T and G-C base pairs. These interactions create a double helix and each base pair is separated by approximately 3.4 Å per step. One full helical turn is 34 Å in vertical height and consists of 10-11 bases. The exterior of the DNA helix consists of a deoxyribose sugar-phosphate backbone. The phosphodiester groups hold a formal negative charge in physiological conditions (~pH 7.4) and exist as salts.

The DNA duplex is held together by two key nucleobase interactions, hydrogen bonding and π-stacking (Figure 1. 3). As noted by Erwin Chargaff, A and T bases were always found in equal numbers, as were G and C bases, which allowed Watson and Crick to build their model with two strands held together with complementary hydrogen bonding base pairs. From these models it can be shown that purine A always pairs with pyrimidine T and purine G always pairs with pyrimidine C.

Structurally the two base pairs have a degree of pseudo symmetry to each other (Figure 1. 3). In the horizontal direction, the A-T base pair has two hydrogen bonding interactions, whereas the G-C base pair contains three. In the vertical direction, A-T and G-C base pairs are stacked upon each other with significant overlap of the aromatic rings. As a result, the aromatic π-stacking interactions between the base pairs provides enhanced stability. Electrostatic interactions also provide stability to the duplex structure. Negatively charged phosphodiesters on the exterior of the duplex interact with the surrounding ionic solvent, shielding the internal hydrophobic region. This means that oligonucleotides are highly soluble in aqueous conditions, interacting with a primary solvation shell of water molecules and metal ions.

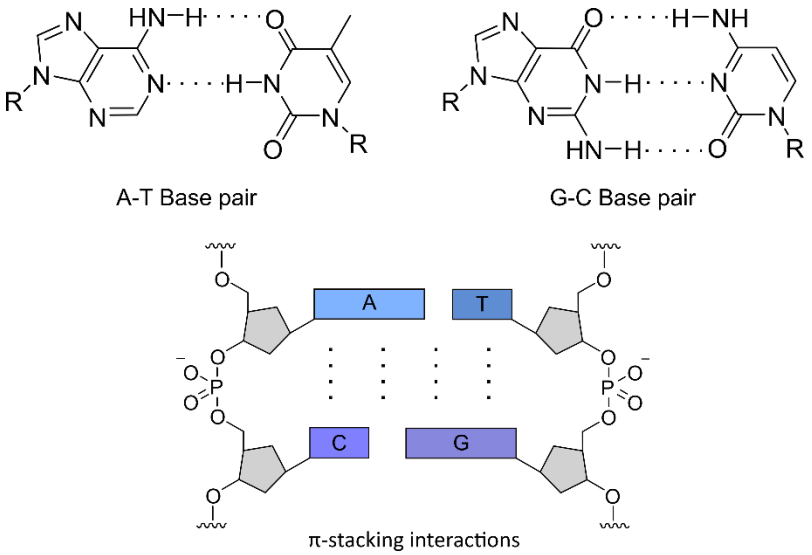



*Figure 1. 3 (Top) Watson-Crick base pair interaction showing the two hydrogen bonds in the A-T pair and the three hydrogen bonds in G-C. (Bottom) Intermolecular π-π stacking interaction between the internal aromatic rings. Interactions are approximately perpendicular to the base pair interactions.*

As first proposed by Rosalind Franklin, there are two distinct and common right-handed double stranded structures in nucleic acids known as the A and B-form.[4] In addition to this was later discovered a third Z-form which is left-handed and is less common.[5] Local helical parameters associated to the base pairs are good identifiers to distinguish between these forms. In total there are eight rotational and eight translation operations that can be used to describe the degrees of freedom associated to a base pair. Of these, the two best parameters that can be used to identify a duplex form are: inclination, which is when the base pair is tilted in the y-axis and x-displacement, which is a translation in the x-axis (Figure 1. 4A, Figure 1. 5).[6]

B-form DNA is a right-handed helix with a rise of 3.4 Å per base step, and base pairs running through the centre of the helical axis (x-displacement of zero). The base pairs have no significant inclination, meaning that they are positioned perpendicular to the helical axis. As a result, B-form DNA is narrow and long, while having a wide major groove, which is similar in depth to its minor groove (Figure 1. 5, see below). The duplex has only 10-10.5 bases per helical turn, and as a result is tightly wound with a twist angle of 34-36 ˚ per base step.[7]

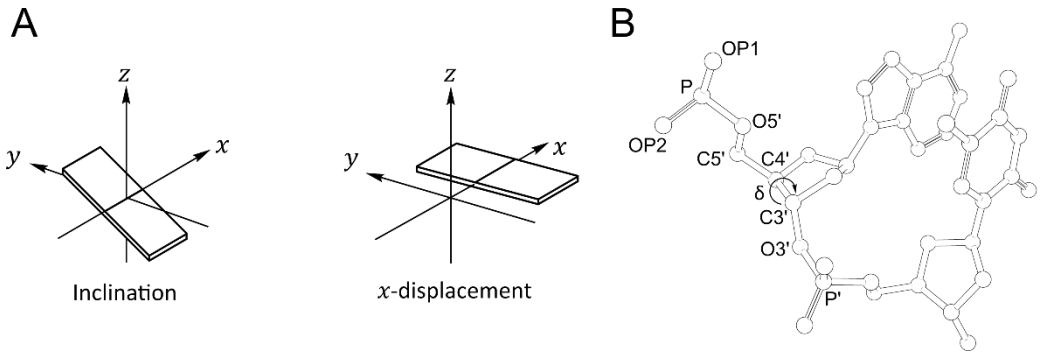


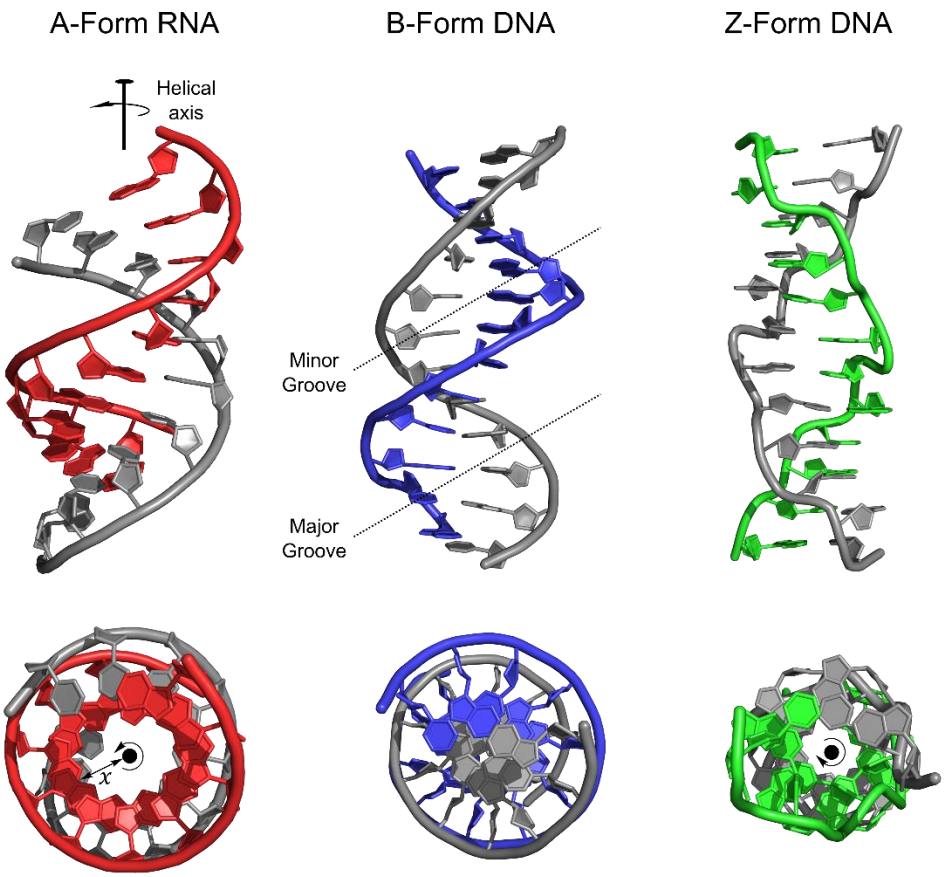

*Figure 1. 4 A) Cartesian axis showing two key transformations that aid to define a duplex form. The base pair is represented by the block. Inclination is a tilting of the base pairs in the y-axis, x-displacement is a translation of the base pair in the x-axis. A typical A-form duplex will have high inclination and x-displacement. B) Diagram showing the atomic trace of a DNA backbone between two phosphorus centres, torsion angle δ is measured and converted into a pseudo rotation to describe an oligonucleotide duplex form.*

A-form DNA is wider and vertically compressed. This is because the base pairs have an x-displacement (3-4 Å) away from the helical axis, which increases the width of the duplex, decreasing its height (2.9 Å per base step). To accommodate this change, the base pairs also incline towards the helical axis improving base stacking interactions. The A-form has a very narrow and deep major groove, whereas its minor groove is wide and shallow. The A-form has 11 bases per turn combined with a lower twist angle of 30-33 ˚ per base step.

In addition to the local base pair parameters, torsion angles of the DNA backbone can also indicate duplex form (Figure 1. 4B). The phosphate-sugar backbone runs from the phosphorus (P) of one nucleotide to the phosphorus (P') of the nucleotide below it, through the route 5'-P-O5'-C5'-C4'-C3'-O3'-P'-3'. By measuring the torsion angle between the C4'-C3' (δ) bond, a quantity known as pseudo rotation can be calculated. Pseudo rotations of 0-30 ˚ corresponds to a 3'-*endo* sugar conformation. As previously described, RNA duplexes are almost exclusively 3'-*endo* and predominantly found in the A-form duplex. Therefore, if most sugar conformations in a duplex have a pseudo rotation of 0-30 ˚ (3'-*endo*) it would suggest the structure has an A-form duplex. Alternatively, if there is large variation in pseudo rotation, it might suggest a B-form structure.[8]

Although there are several parameters for evaluating duplex forms, these are based on average values from structures obtained by various experimental methods. The reality of an individual structure might be that certain parameters fall outside of the standard form. Therefore, all parameters should be considered to determine the helical structure and a

degree of interpretation maybe required, especially if the helix contains non-canonical residues.




| Parameter | A-form | B-form | Z-form |
|---|---|---|---|
| Helix handedness | Right | Right | Left |
| $x$-displacement | Large | None | Small |
| Major groove | Deep and narrow | Wide and equal depth to minor | Shallow |
| Minor groove | Shallow and broad | Variable | Deep and narrow |
| Nucleobase conformation | *anti* | *anti* | C: *anti*, G: *syn* |
| Helical height per turn / Å | 25.3 | 34.0 | 45.6 |
| Rise per bp / Å | 2.3 | 3.4 | 3.8 |
| Bases per turn | 11 | 10.0-10.5 | 12 |
| Inclination / ˚ | 10-20 | *ca* 0 | *ca* 0 |
| Twist / ˚ | 30-33 | 34-36 | C-G: -8, G-C: -52 |
| Common biomolecules | RNA | DNA | - |

*Figure 1. 5 Oligonucleotide duplex structures showing A, B and Z-forms along with statistical measurements that differentiate between them. Highlighted on the duplexes are the helical axis, x-displacement (x) and the position of the major and minor grooves (A-form 1QCU[9], B-form 1BNA[10], Z-form 4OCB[11]).*

Compared to the structure of A and B-DNA, Z-DNA is considerably different (Figure 1. 5). Determined by x-ray crystallography[5], Z-DNA has a left-handed helix that completes one helical turn every 12 bases. Unlike the smooth curvatures of the other forms, Z-DNA bases stack directly on top of each other. This gives it a defined zig-zag structure with pronounced directional changes throughout the backbone. The structure is sequence dependent and

requires alternating C-G/ G-C base pairs. In these pairs the guanine residues are found in the less common *syn* conformation, which means that the duplex is very narrow. In general Z-DNA is less common than other forms, its biological significance is not fully understood, although more potential roles related to protein binding are being discovered.[12]

## 1.2 Synthesis of oligonucleotides

### *1.2.1 The phosphoramidite method*

DNA can be synthesised in the laboratory using solid phase synthesis. Solid supports, also known as resins, act as anchors on which the oligonucleotide chain can be built. Usually made of controlled pore glass (CPG) or polystyrene (Ps), they are porous structures with a high surface area. The first DNA monomer is attached to the resin through its 3'-hydroxyl group using a covalent linker such as a succinyl chain. From this initial monomer, the oligonucleotide is built up by successive rounds of monomer couplings. Unlike for the biosynthesis of oligonucleotides, chemical synthesis usually builds the DNA in the 3' to 5' direction, facilitated using 3'-phosphoramidite building blocks. Phosphoramidites are nucleotides that contain an activated phosphorus (III) centre. In standard solid phase synthesis, this is attached to the 3'-hydroxyl group, but can also be attached to the 5'-hydroxyl for reversed oligonucleotide synthesis. By using the phosphoramidite method developed by M. Caruthers[13], reaction between a 3'-phosphoramidite and 5'-hydroxyl group of a resin bound DNA strand, allows for oligonucleotide chain synthesis on the solid support (Figure 1. 6). The phosphoramidite monomer is first activated by protonation of a tertiary amine by a tetrazole catalyst (pKa 4.9). Nucleophilic attack of the 5'-hydroxyl on the protonated phosphoramidite promotes the loss of a diisopropylamine leaving group, which creates a phosphite ester linkage between the sugars.

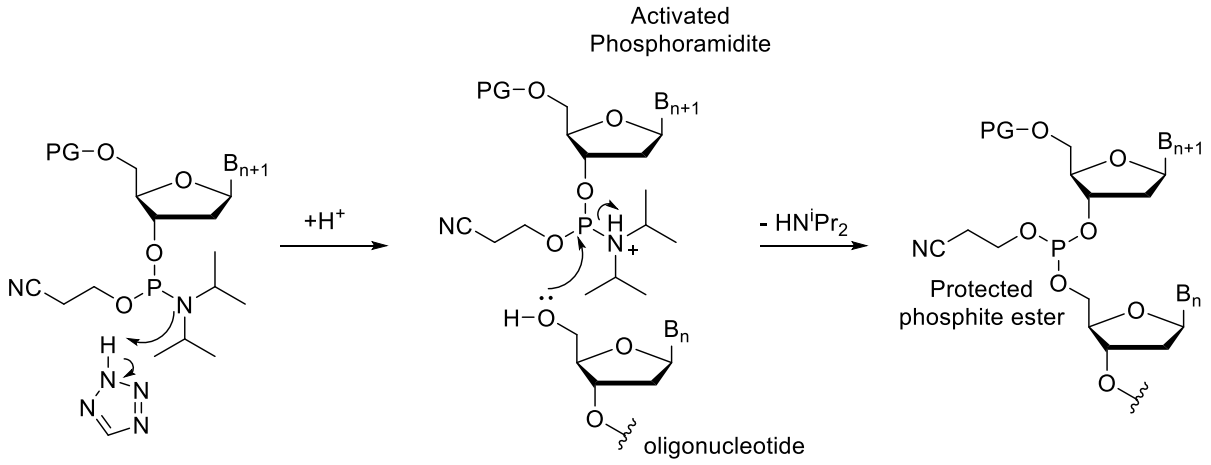



*Figure 1. 6 Phosphoramidite reaction mechanism for the coupling of a nucleoside monomer onto an oligonucleotide. The phosphoramidite is activated with a tetrazole catalyst which is then displaced by nucleophilic attack of a 5'-hydroxyl group on the oligonucleotide. This results in formation of a protected phosphite ester.*

Nucleoside phosphoramidites that are used for solid phase synthesis require protecting groups to allow for controlled oligonucleotide couplings (Figure 1. 7). Nucleobase exocyclic amines are nucleophilic and perform competing coupling reactions if left unprotected. Acyl groups are a common choice of protecting group since they are stable to both the acidic and basic conditions of the reaction cycle. The primary amines of adenine and cytosine are often protected using a benzoyl group ($A^{Bz}$ and $C^{Bz}$), whereas guanine is protected with an isobutyryl group ($G^{Ibu}$). Thymine, with its lack of primary amine, requires no additional protecting groups.

As well as nucleobase protection, the phosphoramidite 5'-hydroxyl group also requires protecting. This is to prevent successive monomer couplings from occurring in a single cycle. 5'-hydroxyl groups are protected with a dimethoxytrityl (DMTr) group. This protecting group is an acid labile, polyaromatic moiety that once removed produces a visible orange cation. Monitoring the cation using absorption at 498 nm means that each monomer coupling efficiency can be measured. In general, a phosphoramidite coupling step has a > 98% efficiency and for unmodified DNA this approaches 99.5%. Despite this, propagation of coupling errors limits the chemical synthesis of oligonucleotides to approximately 100 nt.

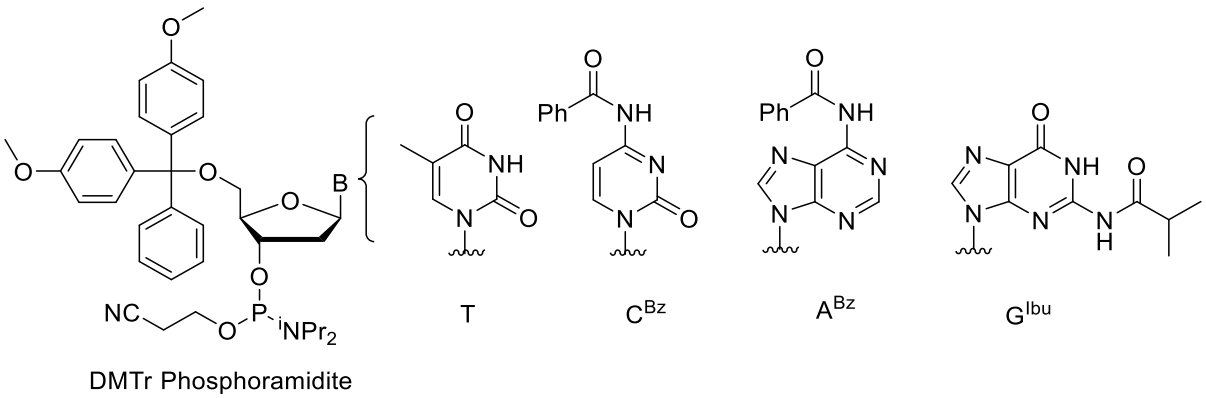



*Figure 1. 7 Phosphoramidite building blocks used for solid phase synthesis. The key structural features are: the 3'-phosphoramidite group that will form the phosphate linkage, the 5'-DMTr hydroxyl protecting group and the nucleobases with protected primary amines (Bz = Benzoyl, Ibu = isobutyryl).*

### *1.2.2 Solid phase synthesis cycle details*

Solid phase synthesis is performed as a cyclical process, with each round (cycle) resulting in the addition of a protected nucleotide onto an oligonucleotide chain. Each cycle contains a detritylation, coupling, capping and oxidation step which is repeated for the required number of cycles. The synthesis is then concluded by cleavage of the oligonucleotide from the solid support followed by deprotection of the bases and phosphodiester groups.

**1. Coupling** – The first cycle starts with a 5'-*O*-DMTr protected resin bound monomer (Figure 1. 8). The monomer is treated with trichloroacetic acid (TCA) to remove this protecting group and expose the 5'-hydroxyl. Before the phosphoramidite can be coupled to the resin bound monomer, it must first be activated. As previously shown (Figure 1. 6), the phosphoramidite is activated using a proton source and the quaternary amine becomes a good leaving group. The protonated diisopropylamine then undergoes a nucleophilic substitution by attack from 5'-hydroxyl resin-bound monomer. Excess of the phosphoramidite is used to ensure good coupling efficiency and rapid reaction kinetics. In solid phase synthesis excess reagents can be washed away, highlighting a significant advantage of solid phase synthesis over solution phase reactions.

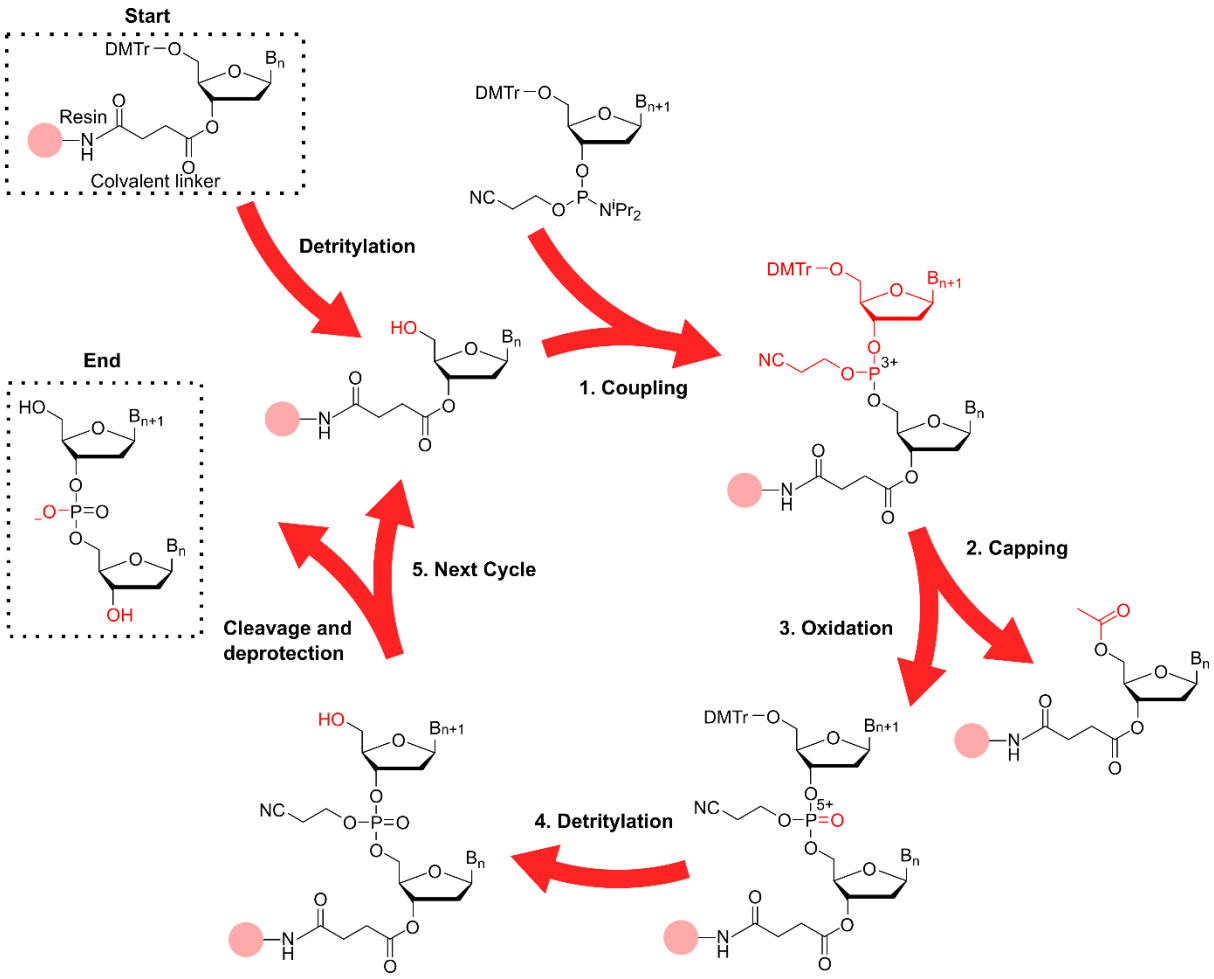



*Figure 1. 8* Solid phase synthesis cycle using phosphoramidite monomers n = residue number. The synthesis begins with detritylation of a resin bound monomer to which a phosphoramidite monomer is coupled. Sequential capping, oxidation, and detritylation completes one cycle of oligonucleotide chain extension before the addition of the next phosphoramidite. After performing the desired number of cycles, the oligonucleotide is removed from the solid support and deprotected in aqueous $NH_3$.

**2. Capping** – A large excess of phosphoramidite monomer is used, and reaction efficiency in most cases will be ~99.5%, which leaves some of the resin bound DNA chains uncoupled. Unreacted 5'-hydroxyls can react in the next cycle and therefore would create by-product sequences that are missing nucleotides (deletion mutations). To prevent these failed sequences from reacting in the next coupling step, a capping step is required. Capping is performed after each coupling step and involves a reaction with N-methylimidazole and acetic anhydride. *N*-methylimidazole activates the acetic anhydride which is displaced by a nucleophilic attack from the 5'-hydroxyl group. This produces a 5'-capped acyl oligonucleotide which cannot be coupled in following cycles.

**3. Oxidation** – After the coupling step, the phosphite-triester is in the $P^{(III)}$ oxidation state. This species is acid sensitive and cannot survive the next detritylation step. Oxidation of

phosphorus is performed in the presence of molecular iodine in basic conditions. The oxidised $P^{(V)}$ centre is acid stable and becomes a 2-cyanoethyl protected phosphodiester linkage.

**4. Detritylation** – With a stable phosphodiester bond formed, it is now possible to prepare for the next coupling cycle. Using TCA, the 5′-*O*-DMTr group is cleaved from the 5′-position of the oligonucleotide. This exposes the hydroxyl group that will be used to couple to the next phosphoramidite. Removal of the 5′-*O*-DMTr group also produces an orange cation that can be monitored to measure the coupling efficiency of the previous monomer. In most cases coupling efficiency will be > 99.5% but even with these high yields the propagation of impurities will limit the achievable oligonucleotide length.

**5. Next cycle** – By repartition of the synthetic cycle and addition of sequential phosphoramidites, nucleotides can be added to the oligonucleotide chain. After the required number of cycles are performed to assemble the desired sequence, the standard nucleobase and 2-cyanoethyl protecting groups are removed by heating in aqueous $NH_3$ at 55 ˚C for 5 hours. In parallel the oligonucleotide is also cleaved from the solid support. In total this provides an unmodified DNA oligonucleotide in solution with canonical phosphodiesters and deprotected nucleobases.

### *1.2.3 Synthesis of RNA oligonucleotides*

In general, synthesis of RNA oligonucleotides uses the same solid phase phosphoramidite cycle as DNA with several differences. Unlike DNA, RNA can undergo reversible strand migration between the 2′ and 3′-hydroxyl positions (Figure 1. 9A). Nucleophilic attack of the 2′-hydroxyl group upon the phosphorus centre forms a 2′-3′-bridged intermediate. Collapse of the intermediate then breaks the 3′ O-P bond resulting in a 5′-2′-linked backbone and overall migration between the 2′ and 3′-hydroxyl positions (Figure 1. 9A, route 1.). Alternatively, the bridged intermediate can also collapse to cleave the phosphodiester backbone. This results in the formation of a 2′-3′-bridged terminus and a cleaved sequence with a 5′-hydroxyl group (Figure 1. 9A, route 2.).

To prevent strand cleavage or migration it is necessary to protect the 2′-hydroxyl group. *Tert*-butyl dimethylsilyl (TBDMS) is a robust RNA protecting group that is routinely used in RNA synthesis. It is relatively cheap but requires deprotection in triethylamine trihydrofluoride ($NEt_3.3HF$) which often results in incomplete deprotection and loss of yield (Figure 1. 9B). An

alternative is the 2’-thiomorpholine-4-carbothioate (TC) group, which is easily deprotected using ethylenediamine but is expensive compared to TBDMS. Although both have drawbacks, they are both effective protecting groups, and are routinely used to synthesise RNA oligonucleotides.

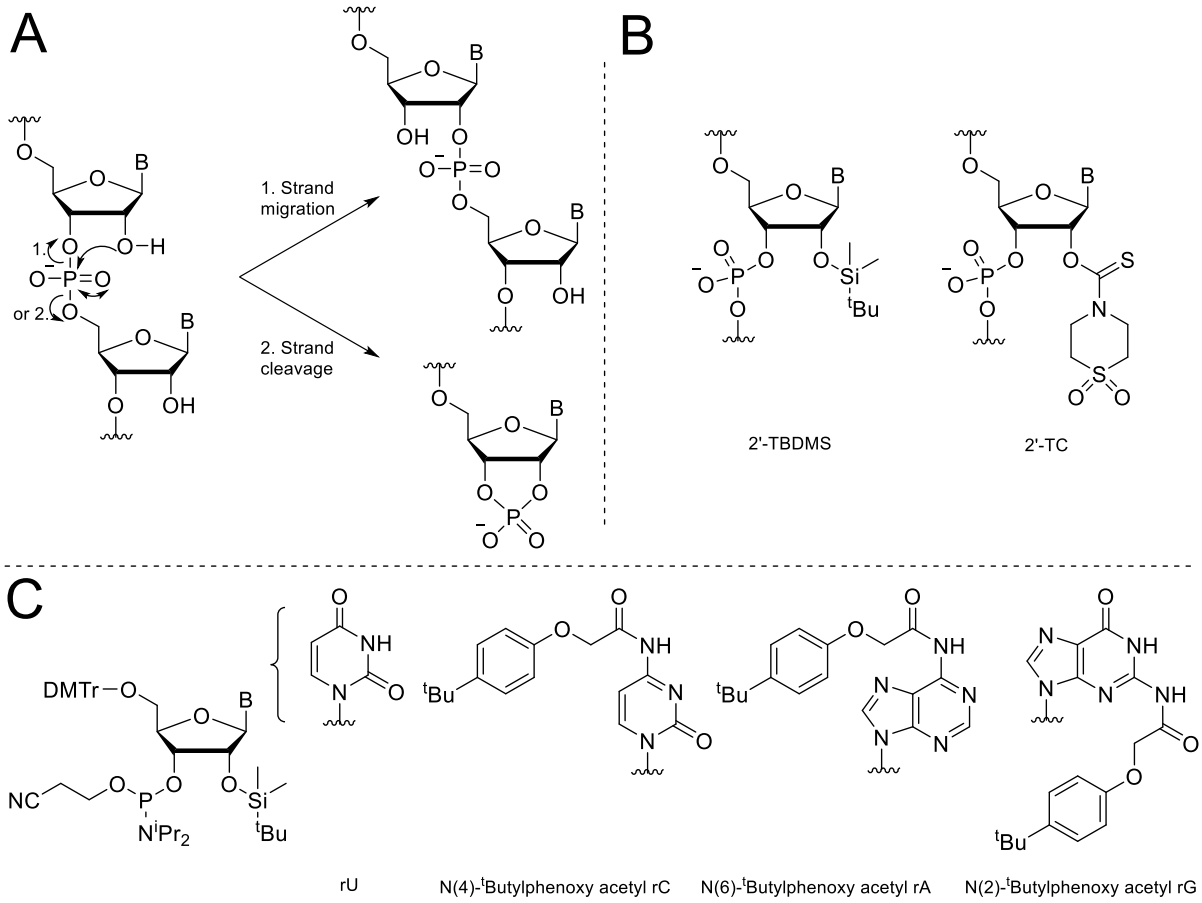



*Figure 1. 9 A) Depiction of RNA instability caused by the presence of the 2’-hydroxyl group. Nucleophilic attack of the 2’-hydroxyl results in either strand migration to the 2’-position or cleavage of RNA oligonucleotide, B) Example of 2’-protecting groups tert-butyldimethylsilane (TBDMS) and 2’-thiomorpholine-4-carbothioate (TC) that are used to prevent strand migration or cleavage, C) Example of an RNA TBDMS phosphoramidite with corresponding nucleobase protecting groups used in solid phase synthesis of RNA.*

RNA oligonucleotides even with 2’-protecting groups are not stable to deprotection by heating in aqueous ammonia. This means that the nucleobase protecting groups must be changed so that they can be selectively removed under milder conditions, without damaging the RNA. Like with the DNA phosphoramidite building blocks, the exocyclic nucleobase primary amines need to be protected. In the case of RNA, the *tert*-butylphenoxyacetyl group is used which can be removed faster and in milder conditions (Figure 1. 9C). Other changes to the RNA synthetic cycle include extended coupling times. This is to allow for a slower reaction because of the increased steric bulk caused by the 2’-protecting groups. Capping is also

performed in *tert*-butylphenoxyacetic anhydride instead of acetic anhydride to prevent exchange of the acyl nucleobase protecting groups, and the deprotection and cleavage is performed in a milder solution of EtOH:$NH_3$ (1:3, v/v) at 55 ˚C for 2.5 hours.

### *1.2.4 Oligonucleotide synthesis using the H-phosphonate method*

An alternative method to synthesise oligonucleotides is to use H-phosphonate monomers. Developed as a precursor to phosphoramidites, H-phosphonates were established for the synthesis of oligonucleotides on solid support.[14] In general, the method is less suitable for long oligonucleotide synthesis, but does have some distinct advantages. H-phosphonates contain a phosphorus (V) centre that is stable to both moisture and oxygen. This means they are much easier to prepare and store but also stable to functional groups such as azides, whereas phosphoramidites in the (III) oxidisation state are not. In the presence of azide functional groups, phosphoramidites undergo Staudinger reaction whereas H-phosphonates do not undergo this reaction and therefore allow for the synthesis of oligonucleotides that contain azide functionality.

Synthesis of oligonucleotides using H-phosphonates is shown in (Figure 1. 10).[15] Initially the H-phosphonate monomer is reacted with a carbonyl chloride to create an activated phosphorus centre. Coupling then occurs by nucleophilic attack of a 5'-hydroxyl from the resin bound oligonucleotide, which promotes the release of a carboxylate leaving group. After capping, the backbone is oxidised using $I_2$/$H_2O$ to produce the canonical phosphodiester linkage. The 5'-DMTr protecting group is then removed to begin the next cycle. After repeating the synthesis with the required number of cycles, the oligonucleotide is cleaved from the solid support and deprotected using standard solid phase synthesis protocols.

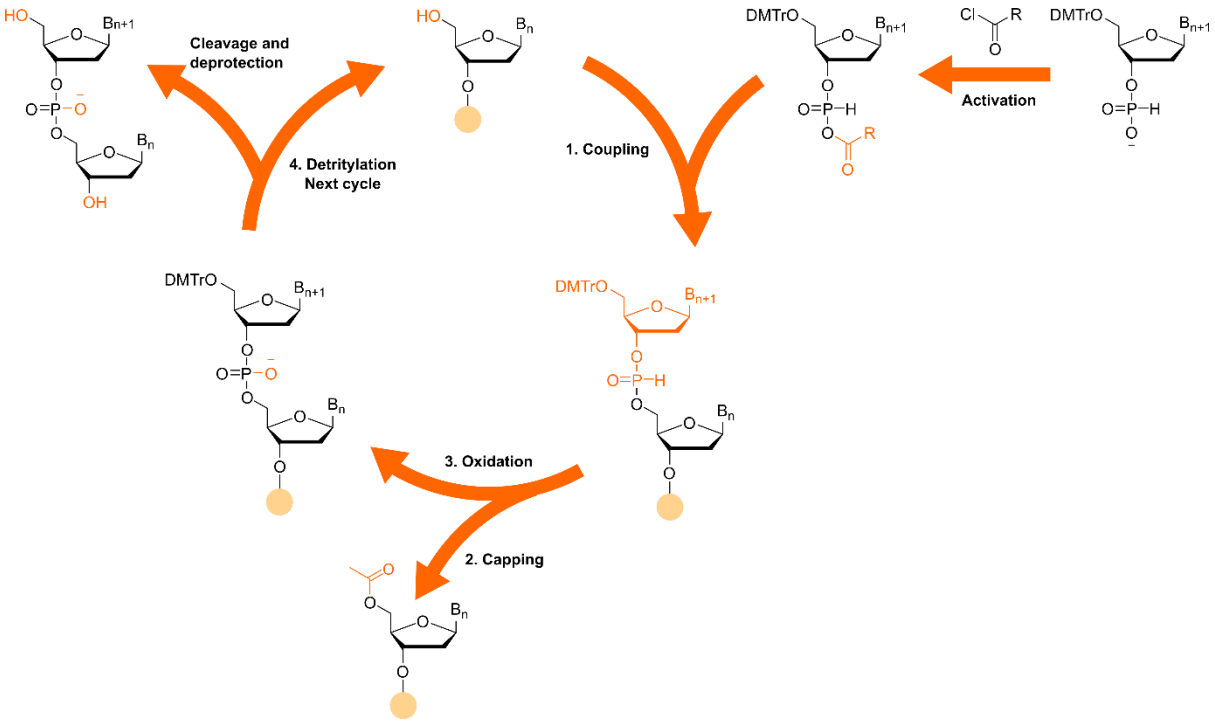



*Figure 1. 10 Solid phase synthesis of oligonucleotides using H-phosphonate monomers, R = adamantane ($C_{10}H_{15}$), n = residue number. The H-phosphonate is activated using a carbonyl chloride. Coupling of the monomer to the 5'-hydroxyl of a resin bound oligonucleotide creates the phosphonate diester linkage. Any remaining unreacted oligonucleotide is capped. Next the linkage is oxidised using $I_2/H_2O$ and then the oligonucleotide is detritylated for the next cycle. Cleavage and deprotection is performed using standard procedures (DNA: $NH_3$ aq, 5 h, 55 ˚C).*

## 1.3 Ligation of oligonucleotides

Ligation is the joining of two nucleic acid sequences by the formation of a covalent bond between them. Ligations can either be performed by enzymes such as T4 DNA ligase, or by chemical ligation using modified backbones such as phosphoramidates or triazoles.[16,17] Oligonucleotide ligations can be performed on single or double stranded sequences (Figure 1. 11). The end of a double stranded oligonucleotide can have one of two structural types. Blunt end termini have no overhanging complementary bases and do not form Watson-Crick base pairs between the two target sequences that are to be ligated. Alternatively, sticky ends have a portion of single stranded sequence that continues beyond the complementary duplex. Ligases can join both blunt and sticky ends, but in general sticky end ligations are more efficient due to Watson-Crick interactions, which hold the two duplex structures together. In both mechanisms two covalent bonds are formed, one between each strand, facilitated by the duplexes being in proximity of one another.

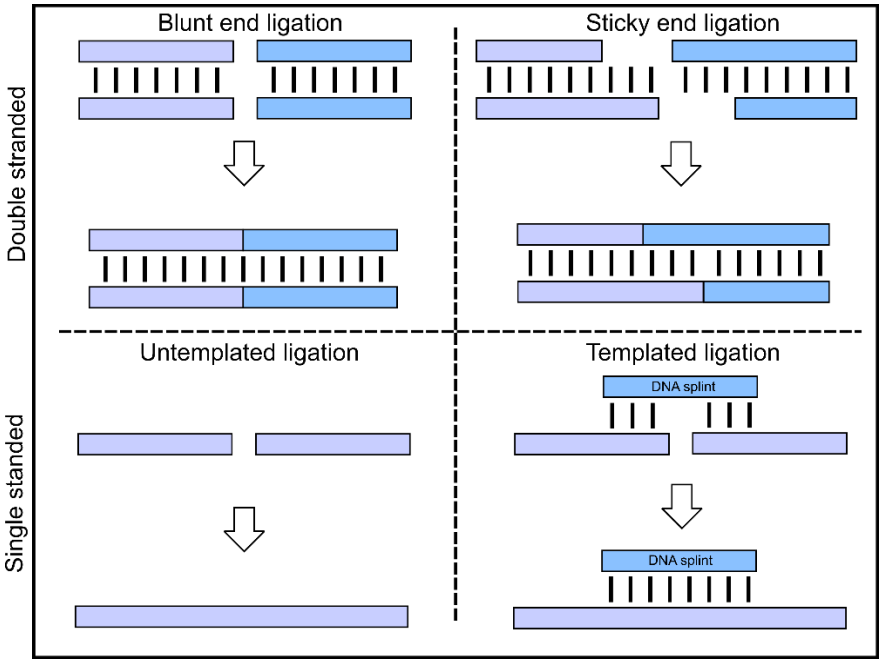



*Figure 1. 11 Ligation of oligonucleotide sequences. Double stranded sequences can be ligated using either blunt or sticky ends, single stranded ligations can be ligated using untemplated or templated ligation. Templated ligation uses a DNA splint to hold the terminal regions in proximity of one another.*

Single stranded oligonucleotides can be joined using untemplated or templated ligations. Untemplated ligation relies on two reactive terminal groups interacting in solution to form a covalent bond. The reactions are often in organic solvents such as DMF or DMSO but ligation efficiency can be limited by slow solution kinetics.

Templated ligation uses a complementary DNA splint to hybridise to the terminal regions of two oligonucleotides. This holds the two ends in proximity improving ligation yield. Ligations can be performed using enzymatic or chemical mechanisms, but requires aqueous conditions and additional salts in order to hybridise the oligonucleotide sequences. Compared to untemplated methods, templated ligation can afford better yield, but requires synthesis and purification of a DNA splint.

### *1.3.1 Enzymatic ligation*

Oligonucleotides can be ligated using enzymatic mechanisms. T4 DNA ligase is an enzyme that repair's DNA after a damage event. Using adenosine triphosphate (ATP) as a co-factor, T4 DNA ligase is able to ligate DNA duplexes or DNA:RNA hybrids. It can also ligate both blunt and templated ends, making it a useful *in vitro* tool for DNA ligation.

The T4 DNA ligase joins the 3'-hydroxyl group of one oligonucleotide to the 5'-phosphate of another (Figure 1. 12). Using a lysine residue, the enzyme cleaves pyrophosphate from ATP

and adenylates the enzyme with adenosine monophosphate (AMP). The AMP modified enzyme is then attacked by the 5′-phosphate of the DNA, transferring AMP onto the oligonucleotide. The 3′-hydroxyl group of the other oligonucleotide then performs a nucleophilic attack at the 5′-phosphate centre, releasing AMP and joining the two oligonucleotides together.

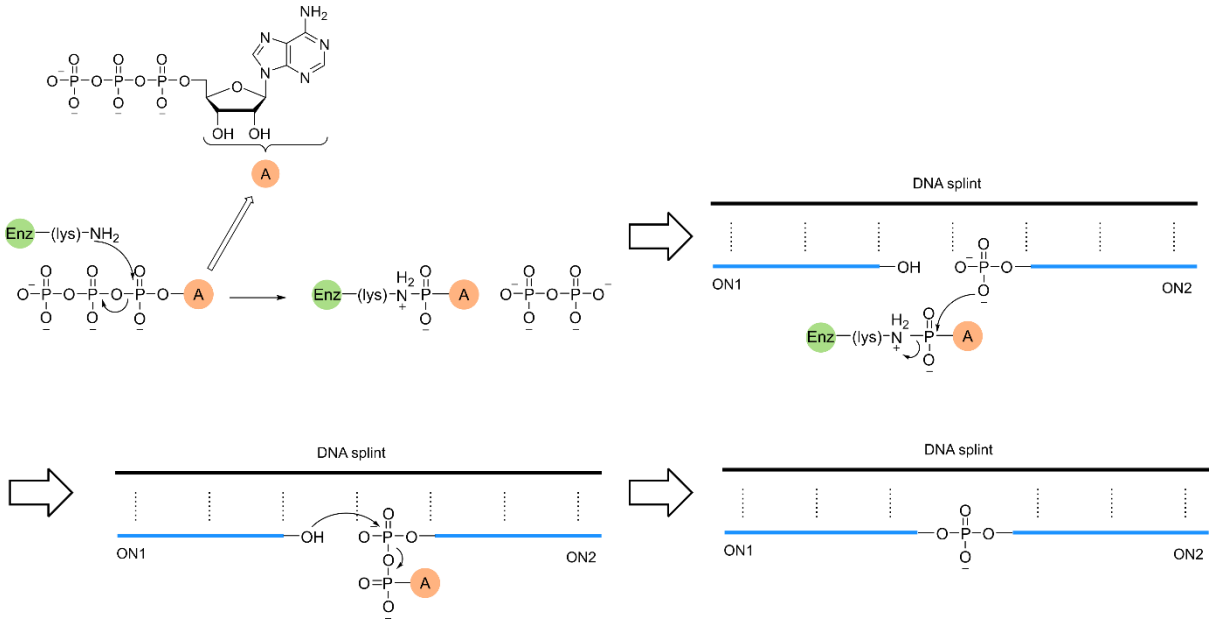



*Figure 1. 12 Enzymatic ligation of oligonucleotides using T4 DNA ligase and DNA splint. ATP is activated by the T4 DNA ligase, which is then added to the 5′-phosphate of the oligonucleotide. Nucleophilic attack of the 3′-hydroxyl, held in proximity by the DNA splint, then attacks the 5′-phosphate ligating the two oligonucleotides promoting the loss of AMP. The ligation is a catalytic process regarding the enzyme and requires ATP as a co-factor that is converted to AMP.*

### *1.3.2 Chemical ligation using the CuAAC reaction*

The $Cu^{I}$ catalysed azide-alkyne cycloaddition (CuAAC) also known as the “click” reaction, is a special example of a Huisgen 1,3-dipolar cycloaddition. By reacting azide and alkyne functional groups, mediated by a $Cu^{I}$ catalyst, it is possible to form a 5-membered 1,4-triazole ring (Figure 1. 13).[18] Compared to the uncatalyzed Huisgen cycloaddition, the $Cu^{I}$ reaction can be performed rapidly and at room temperature and is highly selective for the 1,4-isomer.

The CuAAC reaction can be performed in aqueous conditions, which makes it excellent for the labelling or ligation of biomolecules. Oligonucleotides modified with complementary azide and alkyne functional groups, can be ligated together by using the CuAAC reaction. $CuSO_4$ is often used as a source of copper but needs to be reduced to a $Cu^{I}$ species to facilitate the cycloaddition. Sodium ascorbate acts as the reductant which forms the $Cu^{I}$ in situ, allowing it to coordinate the azide and alkyne functional groups. Oligonucleotides can be ligated by

untemplated ligation in solution; however, the yield and rate of reaction can be significantly increased if a DNA splint is used (Figure 1. 14). A DNA splint is an oligonucleotide that is complementary to the terminal regions of both oligonucleotides that are being ligated. The splint acts to hold the azide and alkyne in proximity of each other, so that the $Cu^{I}$ can perform the migration-insertion. A drawback of $Cu^{I}$ ligation is its known ability to damage DNA. Reaction with molecular oxygen and other small molecules creates reactive intermediates that can cause oxidative damage to the DNA. To reduce this effect, a copper chelating ligand such as Tris(3-hydroxypropyltriazolylmethyl)amine (THPTA) is used.[19] These types of ligands coordinate to the $Cu^{I}$ stabilising the oxidation state and decreasing the number of reactions in solution with other small molecules. Despite this, some DNA damage is always inferred, so it is important to reduce reaction times and use a minimal amount of copper.[20]

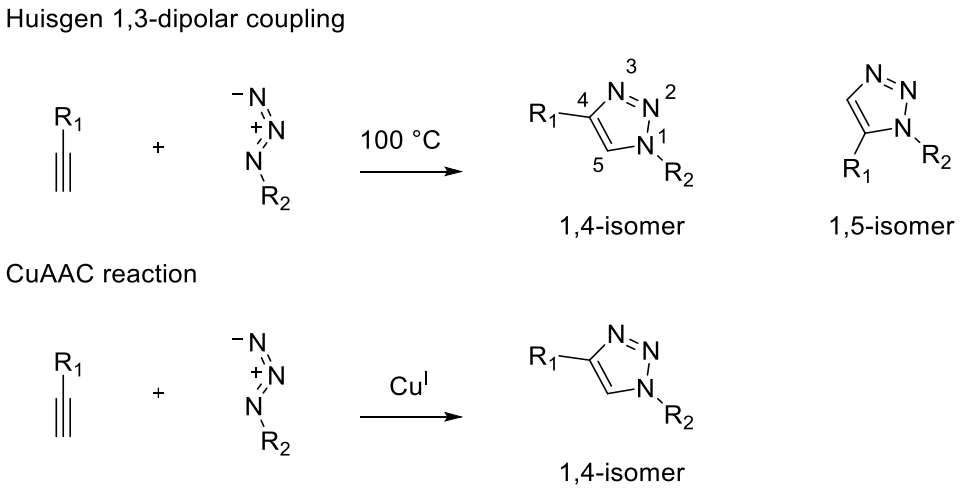



*Figure 1. 13 Comparison between a Huisgen 1,3-dipolar coupling and the copper click azide alkyne cycloaddition (CuAAC). The CuAAC reaction can be performed at room temperature in aqueous conditions without producing the 1,5-isomeric form.*

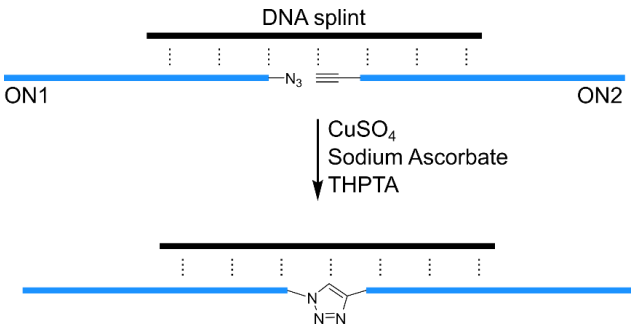



*Figure 1. 14 Chemical ligation of oligonucleotides (ON1 and ON2) using the CuAAC reaction to form a triazole linkage. Oligonucleotides are templated against a DNA splint to hold the azide and alkyne in proximity for improve ligation yield.*

## 1.4 Biophysical analysis of oligonucleotides

### *1.4.1 UV-thermal melting*

The thermal properties of an oligonucleotide duplex are an important consideration when designing therapeutics and for the analysis of the biological properties of specific sequences. The process of annealing two single stranded oligonucleotides to form a duplex is known as hybridisation. The hybridisation strength depends on the sequence, length and GC content of a duplex. GC content has a significant role in duplex stabilisation because of the additional hydrogen bond in the base pair. More hydrogen bonds require a greater amount of energy to break and therefore increase hybridisation strength.

Hybridisation strength can be measured using UV-thermal melting. In its conventional definition, melting refers to a phase transition from a solid into a liquid. In this case, melting refers to the transition of an oligonucleotide from a duplex into its constituent single strands. Starting at a low temperature to ensure the oligonucleotides are in the duplex form (10-25 ˚C), the duplex is slowly heated to 75 ˚C. As the temperature increases, the hydrogen bonds and π-stacking interactions are broken, dividing the duplex into two single stranded oligonucleotides. As this process occurs the UV absorbance increases, this is because the magnitude of UV absorbance of the two single strands is about 40% larger than when in a duplex form. The point at which half the moles of duplex have been separated into their constituent single strands is called the melting temperature ($T_m$). The $T_m$ can be obtained from a plot of absorbance *vs* temperature at the point of inflection or by plotting the differential of absorbance and obtaining the temperature at which the gradient is zero (Figure 1. 15).

Measuring the hybridisation strength of a duplex provides information about target binding but also about properties of modifications. By introducing nucleobase or backbone modifications, a change in melting temperature ($\Delta T_m$) can be measured between a modified and unmodified duplex. This shows if a modification is duplex stabilising (+ $\Delta T_m$) or duplex destabilising (- $\Delta T_m$). In general, high $T_m$ values are usually required to ensure good target affinity, however, if $T_m$ values are too high then non-specific off-targeting effects may arise.

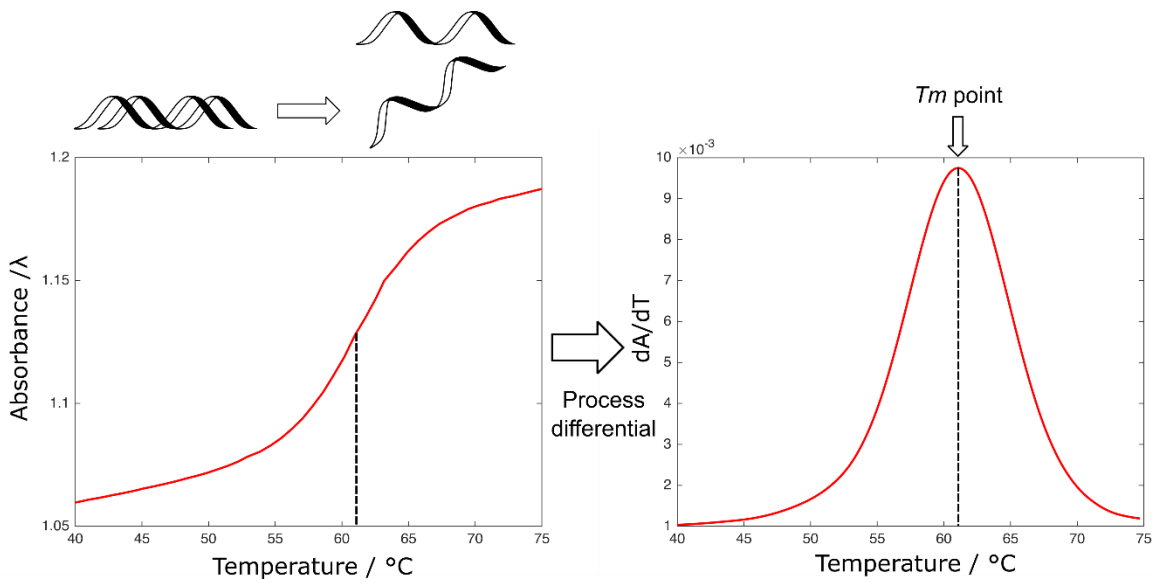



*Figure 1. 15 Example plots of an oligonucleotide UV-melting experiment. Temperature is increased from 25 to 75 ˚C while measuring the UV absorbance of the oligonucleotide duplex. As the temperature increases the duplex separates into its constituent oligoes which have a combined UV absorbance greater than in a duplex form. The point of inflection (left) is the temperature at which half of the duplexes (mol) has been separated into single stranded oligonucleotides. Processing of the differential (right) provides a melting temperature ($T_m$) at the point where gradient is zero.*

### *1.4.2 Circular dichroism*

Circular dichroism (CD) uses circular polarised electromagnetic waves (100-400 nm) to analyse the structure and conformation of molecules. Molecular species with chiral centres, absorb left and right handed components of circular polarised light differently, depending on the arrangement of the atoms.[21] This bias in absorption can be expressed as a positive or negative change in rotation, and is measured in millidegrees. Although CD cannot provide structural information about atom positions, it can provide insight into secondary structures. For oligonucleotides, CD is a sensitive technique that aids to confirm the presence, type and stability of a duplex form.[22] A and B-form structures have unique absorptions which allow them to be distinguished. To obtain a CD trace for an oligonucleotide duplex, a scan of circular polarised light is performed between 200-340 nm (Figure 1. 16). B-form DNA duplexes are characterised by broad peaks at 220, 255 and 280 nm. A-form RNA is characterised by a sharp peak at 210 nm and a broad peak at 270 nm.

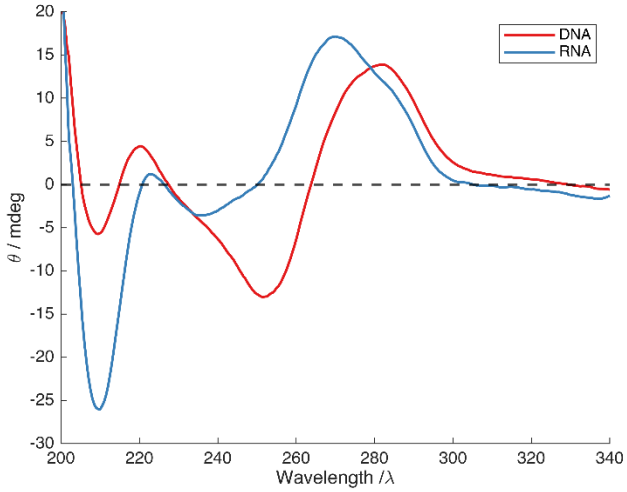



*Figure 1. 16 Typical circular dichroism traces obtained from a DNA duplex (red) and RNA duplex (blue). DNA B-form duplexes can be characterised by broad peaks at 220, 255 and 280 nm. RNA A-form duplexes are characterised by a sharp peak at 210 nm and a broad peak at 270 nm.*

Although CD is a versatile tool for the analysis of oligonucleotide secondary structure, it has some limitations. Atomic positions cannot be measured, and analysis of the curve is restricted to a qualitative interpretation, based on matching observed changes to a hypothesis. Despite this, when CD is complemented with alternative techniques, it can provide unique insight into an oligonucleotide structure.

### *1.4.3 Gel electrophoresis*

Gel electrophoresis is a technique that allows biological molecules such as DNA or proteins to be separated based on their size and charge. Polyacrylamide gel electrophoresis (PAGE) is a common method used for oligonucleotide analysis and purification. Samples are loaded into the top of a gel, submerged in a buffer containing conductive ions (Figure 1. 17). Application of a voltage draws the oligonucleotide into the gel and is facilitated by the anionic phosphodiesters along the oligonucleotide. Due to the porous composition of the gel, shorter oligonucleotides travel further per unit time. This allows them to be separated by their mass to charge ratio and be analysed by either UV-light or intercalating/external binding luminescent dyes.

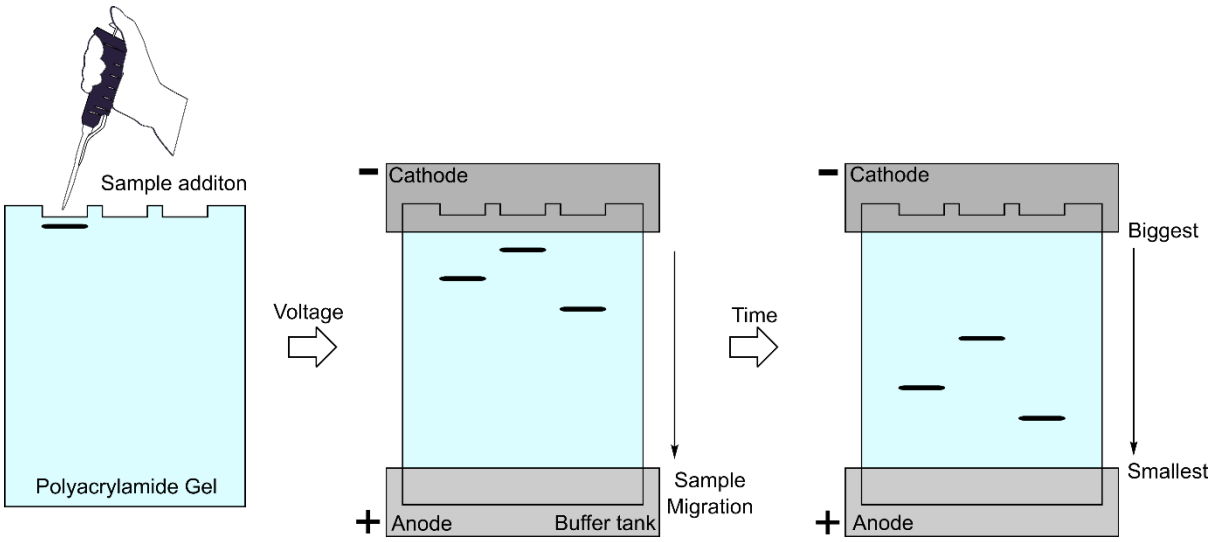



*Figure 1. 17 Experimental analysis using polyacrylamide gel electrophoresis (PAGE). Samples are loaded into the top of the gel. The gel is then submerged in a running buffer that contains conductive ions. Application of a voltage draws the negatively charged samples through the gel relative to their mass to charge ratio. Bigger molecules travel through the gel slower allowing for size separation.*

Secondary structure changes the mobility of an oligonucleotide travelling through a gel. Therefore, for consistent migration between oligonucleotide species, it is important to remove secondary structures and separate duplexes into their single strands. Denaturing PAGE is a technique that removes secondary structures and ensures consistent gel migration. The gel contains high concentrations of urea, which disrupts base pairing interactions by acting as a hydrogen bond donor/acceptor. Alongside this, samples are loaded in a 1:1 mixture of formamide:$H_2O$ to prevent duplex formation by introducing similar disruptions.

Linear and circular oligonucleotides, with the same sequence and length, migrate at different rates through a gel. This is because they have the same mass to charge ratio but different friction coefficients. Although denaturing PAGE can disrupt weak interactions and intermolecular forces, it does not break covalent bonds. As a result, denaturing electrophoresis does not remove the secondary structures caused by covalent bonding and allows circular oligonucleotides to be separated from linear sequences.

Native PAGE is an alternative method for analysis of oligonucleotide secondary structures. The gel contains no urea and samples are loaded in a non-denaturing glycerol solution. Like other methods, samples are loaded into the gel and a voltage is applied. Native PAGE is particularly useful for analysis of DNA hybridisation. The lack of base pair disruption means that oligonucleotides remain in a duplex form. However, it is important to keep the voltage below 150 V, to prevent excessive heat generation and thermal melting of the duplex.

Proteins can also be analysed by gel electrophoresis. Unlike oligonucleotides, protein isoelectric points are often dictated by amino acid residues that are usually charge neutral. When a voltage is applied, they have poor or inconsistent migration into a gel. To improve migration, surfactants can be added to coat the protein in negative charges. Sodium dodecyl sulfate (SDS) is a molecule that contains a long fatty chain and a polar sulfate group at one end. Heating a protein mixture above 95 °C in SDS causes unfolding of protein secondary structure and the surface becomes coated with the SDS negative charge. Application of the sample onto an SDS-PAGE, allows proteins to be resolved according to their size, like previously described for oligonucleotides.

### *1.4.4 Antibodies and Immunoassays*

Antibodies are Y-shaped proteins that form part of the immune response. They recognise specific targets and help to neutralise pathogens such as viruses or bacteria.[23] Formed of two types of domain known as the light and heavy chain, the terminal binding region is highly specific, often binding to only one target known as an antigen (Figure 1. 18A). For research purposes, antibodies can be produced in a range of host organisms and tailored to bind specific antigens. Many different antibodies are commercially available and can be designed to target either a protein or another antibody of interest. Antibodies can be tagged with fluorescent dyes or reactive proteins and used as diagnostic agents. Based on this, several detection or capture assays have been developed such as immunoblotting and immunoprecipitation.

Western blot is an immunoblotting technique that uses antibodies to detect the presence of a specific antigen, often a protein (Figure 1. 18B). Although antibodies can be used to bind to whole proteins, they can also be used to target short peptide sequences called epitope tags. An example of an epitope tag is the FLAG sequence (DYKDDDDK), which is a synthetic peptide that has been co-developed as an antigen for protein purification and immunoblotting with its corresponding antibodies.[24] To perform western blot analysis, the protein mixture is first resolved into its constituents by PAGE. PAGE is a non-specific technique that shows all proteins present in the sample regardless of the presence of a FLAG sequence.

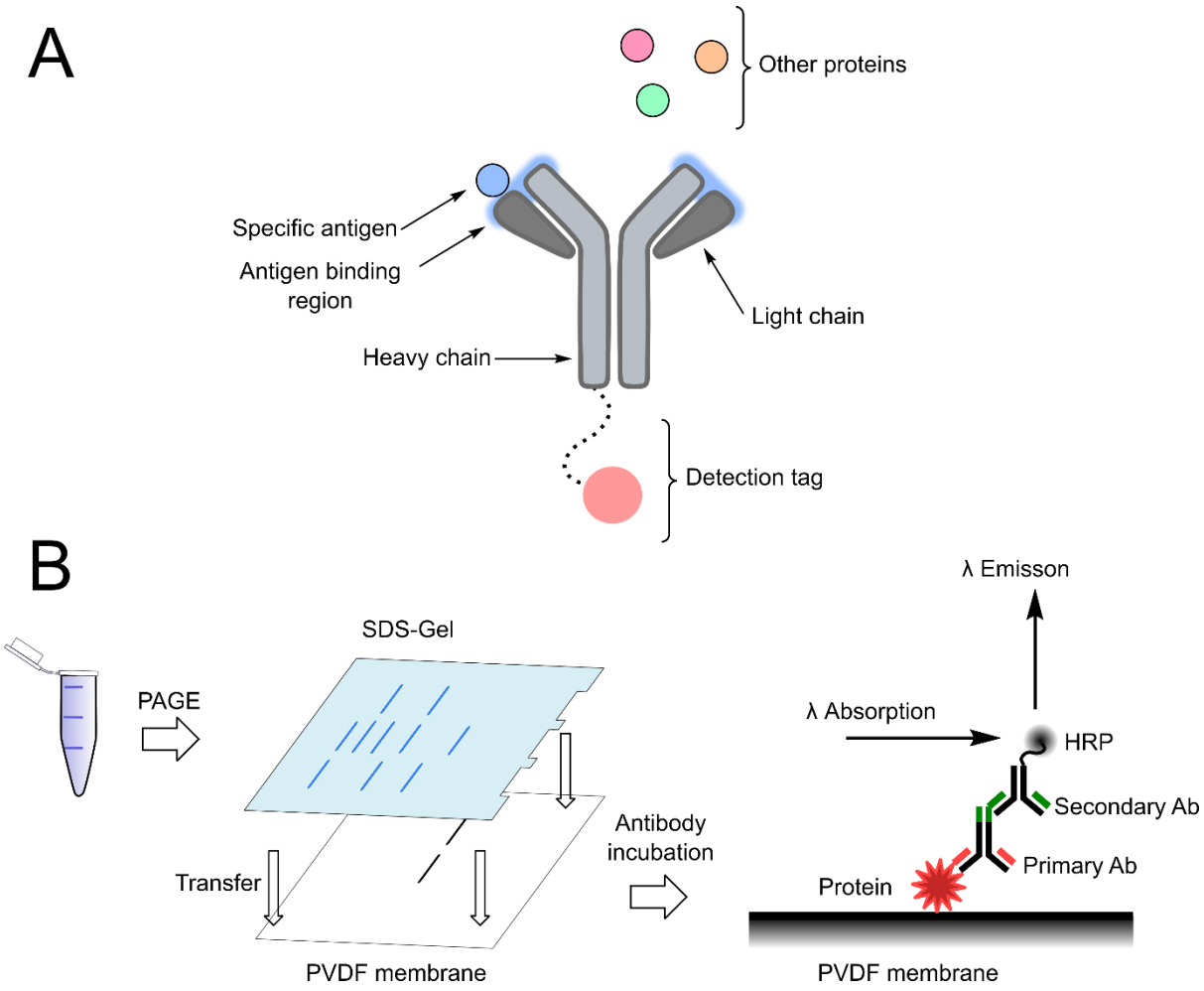



*Figure 1. 18 A) Cartoon depiction of an antibody structure showing the light and heavy chains, antigen binding region and optional detection tag for western blot analysis. Detection tags can be either a fluorophore or horseradish peroxidase enzyme. B) Example of western blot procedure. Protein samples are separated using PAGE and transferred onto a PVDF membrane. The membrane is then coated with primary and secondary antibodies, which can be analysed using an HRP tag attached to the secondary antibody. Antibody detection is specific to the target antigen (protein).*

To identify the proteins with FLAG tag sequences, specific antibodies can be used. The gel loaded with proteins is placed into a "transfer sandwich" stacked between a support and polyvinylidene fluoride (PVDF) membrane. Application of a current is then used to transfer the proteins from the gel onto the surface of the PVDF. The loaded membrane can then be incubated with an antibody that binds the FLAG sequence as an antigen, referred to as antiFLAG. This is called the primary antibody and binds directly to the target peptide. After a round of blocking, to prevent non-specific binding, a secondary antibody is added to enhance and facilitate detection. This secondary antibody specifically binds to the primary antibody. Antibodies are biosynthesised in a host organism such as a mouse, rabbit or goat. Therefore, if the primary antibody used in the western blot was produced in a mouse, it would be referred to as an antiFLAG mouse antibody. By using a secondary antibody that targets mouse

antibodies (antimouse), we can specifically bind the secondary antibody to the primary antibody.

Secondary antibodies can be modified with fluorescent molecules to enable indirect detection of the FLAG tag sequence. Fluorescent detection methods expose the PVDF membrane to a source of electromagnetic waves. The exact wavelength depends on the fluorescent molecule but is usually in the range of 200-700 nm. The fluorescent emission from the small molecule can be detected to confirm the presence of the secondary antibody.

Similarly, the secondary antibody can be conjugated to a horseradish peroxidase (HRP) enzyme which oxidises substrates using hydrogen peroxide to produce either a chemiluminescent signal or colorimetric response. When oxidised, luminol emits a chemiluminescent signal that can be recorded at 428 nm. Alternatively, 3,3',5,5'-Tetramethylbenzidine (TMB) produces a blue coulometric response which can be measured using UV-absorption.

Due to the complexity of the western blot, the technique can be time consuming and expensive. The dot blot assay provides a simplified alternative where samples do not need to be separated first by gel electrophoresis. Instead, samples are applied directly to the PVDF membrane followed by antibody detection. An advantage of this technique is its speed and simplicity. However, it offers no information about the size or ratios between proteins within the sample.

Another immunoassay is an enzyme-linked immunosorbent assay (ELISA), which can be performed in a high throughput plate reader format. Protein samples are first captured onto the surface of a plate using a capture antibody and detected using the HRP tagged antibodies. Unlike previous techniques, TMB substrate conversion reacts with the HRP in solution, which means that protein yield can be quantified by absorbance. As the solution develops, kinetic information can also be obtained by measuring multiple time points, or an absolute value can be taken as an endpoint after the addition of $H_2SO_4$.

## 1.5 Introduction to modified backbones

The phosphodiester backbone (PO) is the canonical linkage between oligonucleotide monomers (Figure 1. 19). It is a five-bond linkage, with highly flexible $sp^3$ centres. At the

centre of the backbone is the phosphate atom surrounded by two bridging and two non-bridging oxygens. The phosphate backbone is a result of biological co-evolution with enzymes that form, alter and cleave the linkage, in order to fulfil an oligonucleotides biological lifecycle. The phosphate backbone can be modified to change its properties such as to increase thermal stability, resist enzyme degradation or improve cellular uptake.

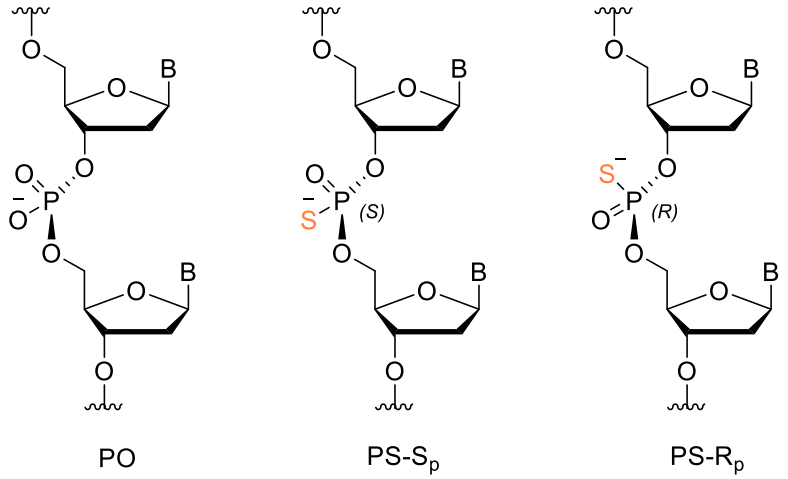



*Figure 1. 19 Structures of phosphate derived backbones including the phosphodiester (PO) and sulfur modified Phosphorothioates (PS). Addition of a non-bridging sulfur creates a chiral centre and two isomeric forms ($S_p$ and $R_p$)*

Phosphorothioates (PS) are one of the most studied modified backbones and are structurally very similar to the phosphodiester (Figure 1. 19). The difference between these two backbones is that the phosphorothioate has an additional sulfur atom that replaces one of the non-bridging oxygens. First examined as a synthesis of adenosine 5'-mono phosphorothioate (AMPS), it was noted that addition of sulfur provided significant phosphatase resistance compared to its canonical alternative, adenosine 5'-monophosphate (AMP).[25] This led to the development of an oligonucleotide phosphorothioate backbone which contains a chiral centre with two isomeric forms $R_p$ and $S_p$.[26]

The chirality of the phosphorothioate is important because it dictates the biochemistry of the backbone. When incubated with P1 or S1 nucleases, oligonucleotides synthesised with sterochemically pure $R_p$ centres, were found to be more resistant to degradation.[27] However, against enzymes such as the 3'-5'-Klenow nuclease, an enzyme that contains two ions in its catalytic site, it was found that $R_p$ centres were rapidly cleaved, whereas $S_p$ phosphorothioates were not.[28] This is due to the atomic radius of the sulfur atom. Crystal structures of the 3'-5'-Klenow enzyme, bound to a phosphorothioate modified oligonucleotide, showed that the increased atomic radius of sulfur, combined with its position

in the active site when in the $S_p$ configuration, pushes one of the metal ions out of the active site inactivating the enzyme.[29] In both cases, the presence of the phosphorothioate increases the nuclease stability of the oligonucleotide. However, the effectiveness is dependent on the specific nuclease and stereochemistry of the backbone.

Although phosphorothioates improve resistance to nuclease degradation, the increased atomic radius of sulfur causes moderate duplex destabilisation. In addition to this, controlling the stereochemistry of the phosphorus centre during synthesis is also considerably difficult[30], and it is currently unknown which configuration is best for specific applications.[31]

Carbamates (CBM, chapter 2), amides (Am, chapter 3) and triazoles (Tz, chapter 4) are the three modified backbone species contained within this thesis (Figure 1. 20). Although they are each separately reviewed within their respective sections, there are underlying features common to most modified backbones.

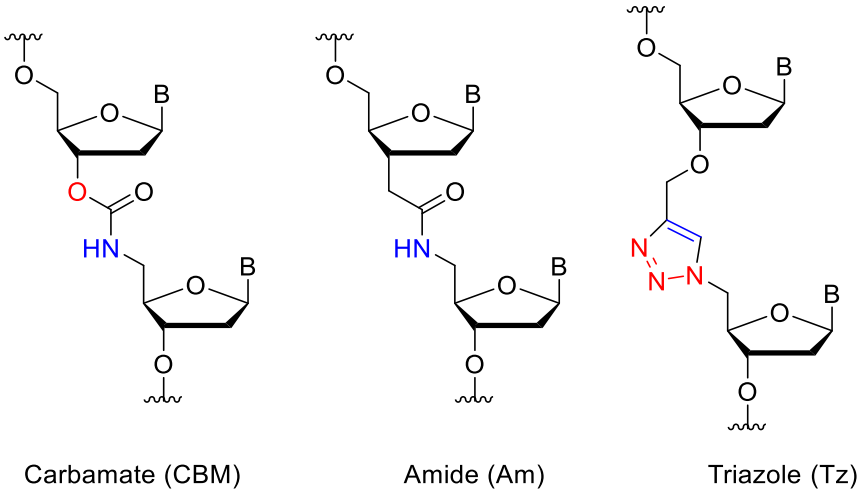



*Figure 1. 20 Structure of modified backbones featured in this work, carbamate (CBM)[32], amide (Am)[33] and triazole (Tz)[34].*

Although there are many different modified backbones within the literature, these three linkages have been selected based on several criteria. The carbamate backbone has been found to be duplex destabilising but are simple to synthesise. In contrast, amides have good thermodynamic properties but exchanging a nucleotides 3'-hydroxyl group for the amides 3'-$CH_2$ is a considerable synthetic challenge. Like the carbamate, triazole backbones moderately decrease the hybridisation strength between two oligonucleotides. Despite this, the CuAAC reaction is incredibly convenient, versatile and robust, which makes it a good backbone for applications.

In general, modified backbones are highly resistant to nuclease degradation. This is because enzymes like phosphatases have evolved to be highly specialised and only cleave P-O bonds. Addition of a modified backbone, that lacks a phosphate structure and/or charge, prevents cleavage of the linkage, increasing biological half-life.

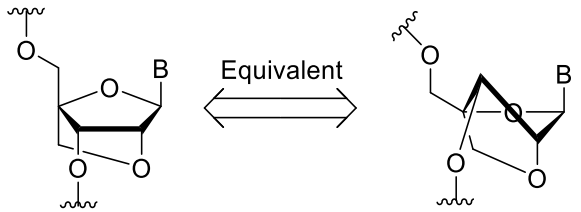
Backbone length, flexibility, charge, and shape all contribute to the biophysical properties of a modified linkage. Backbones that do not conform to the structural requirements of an oligonucleotide can disrupt hydrogen bonding or base stacking, destabilising the duplex. Because of this, backbones that can best mimic the phosphodiester provide the most stable duplexes.

One of the main strengths of modified backbones is charge neutrality. Delivery of a nucleic acid through cellular barriers is descried as a significant challenge facing oligonucleotide therapeutics. The cell membrane is composed of a hydrophilic negatively charged surface and a hydrophobic interior that is charge neutral. This means that canonical oligonucleotides with polyanionic phosphodiesters have poor uptake into the cell. Introduction of charge neutral backbones such as peptide nucleic acids (PNA), reduce the negative charge of the oligonucleotide increasing membrane permeability and cellular uptake.[35]

## 1.6 Introduction to Locked Nucleic Acids (LNA)

Locked nucleic acids (LNAs) are 2'-4'-bicyclic nucleic acids that were separately co-developed by T. Imanishi *et al.*[36,37] and J. Wengel *et al.*[38,39] (Figure 1. 21). The nature of these fused rings constrains the flexibility of the nucleic acids, locking the sugar into a 3'-*endo* conformation.[40,41] Because of this preference for the 3'-*endo* conformation, LNAs have good affinity for RNA targets and the ability to improve duplex hybridisation when incorporated into oligonucleotides (+7 °C/mod).[39] The presence of LNA in oligonucleotides is able to influence the local duplex structures, and in many cases LNAs provide duplex stability by reorganising the conformation of surrounding nucleotides.[40]

Biophysical analysis found that LNAs have good mismatch discrimination as well as extended half-lives when incubated with nuclease enzymes.[38,42] The ability to selectively bind to a target sequence with good biological stability, makes LNAs an attractive modification for therapeutic application. In some cases LNA oligonucleotides can cause hepatotoxicity but with the correct design LNAs have been proven to be non-toxic.[43,44]




Locked Nucleic Acid (LNA)

*Figure 1. 21 Structure of the bicyclic ring fused Locked nucleic acid (LNA), highlighting the 3'-endo conformation.*

LNA modifications influence a wide range of applications and have been used in antisense gapmer sequences[45] as well as being suitable substrates for enzymatic labelling.[46] They have been shown to have good compatibility with polymerases, being used in PCR amplification[47] and more recently LNAs have been applied to CRISPR RNAs, where it was found that they were able to enhance Cas9 endonuclease specificity, compared to RNA controls.[48]

## 1.7 Principles of X-ray diffraction (XRD) and crystallography

### *1.7.1 Defining a crystal lattice*

A crystal lattice is the ordered arrangement of atoms that is an infinite repeating structure with defined symmetry.[49] A lattice can be broken down into a unit cell, which is defined as the smallest repeating unit that defines all symmetry operations within a crystal lattice. The unit cell is defined using three axes (a, b, c), three angles (α, β, γ) and three faces (A, B, C) (Figure 1. 22). By translating the unit cell, it is possible to build up a repeating crystal lattice structure.

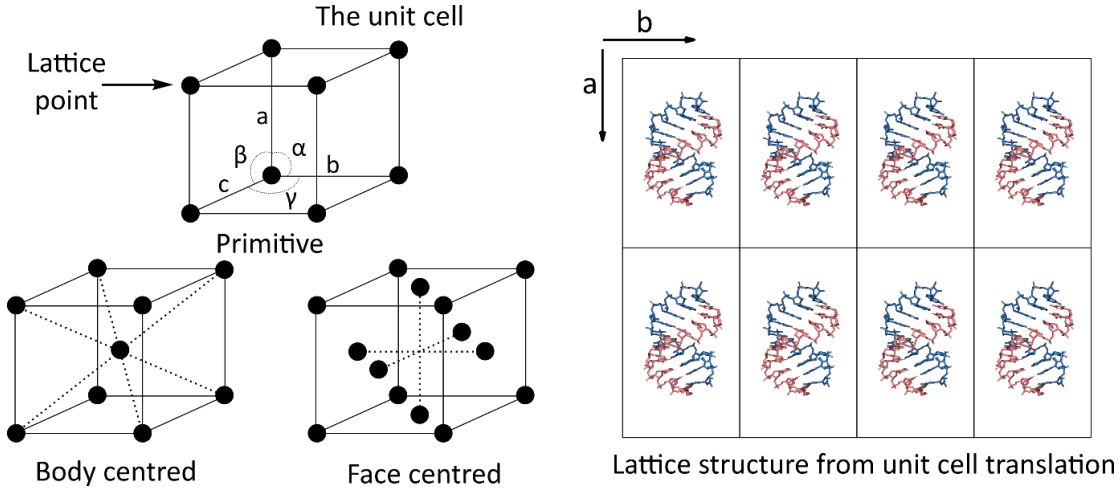



*Figure 1. 22 (Left) Structure of a unit cell, defined as the smallest repeating unit that describes all symmetry operations within a crystal lattice. Also shown is three examples of a Bravais lattice, primitive with lattice points on each corner, body centred with a central lattice point and face centred with a lattice point on each face. (Right) Translation of a unit cell along the a and b-axis, to create a repeating crystal lattice structure.*

A lattice point is a position in a lattice that an atom or molecule is placed. A defining feature of a lattice is its inherent symmetry, therefore there are a limited number of arrangements of lattice points that are allowed based on the geometry of the unit cell. In total there are seven crystal systems, which describe the geometries of the unit cell. These can be further sub divided into 14 Bravais lattices that describe the arrangements of the lattice points and the translational symmetry.[50] As an example, the cubic lattice system has unit cell dimensions of a = b = c and α = β = γ = 90 ˚. The arrangement of the lattice points within the cubic lattice system can be defined by three Bravais lattices known as primitive (P), body centred (I), and face centred (F) (Figure 1. 22). In the primitive arrangement, lattice points are positioned on each corner of the unit cell. Addition of an extra lattice point in the centre of the unit cell produces the body centred form, or addition of a lattice point to each face creates a face centred lattice. In each case, translation of these unit cells by a vector a, b or c produces an identical unit cell with overlapping lattice points. Although the Bravais lattice defines the translational symmetry of the lattice points, crystal systems also have other symmetry elements. Symmetry such as rotation or rotation translation (Screw), are also found in crystal packings. When translation and rotation symmetries are combined, they form a unique symmetry relationship that can be described as a space group. Mathematically it has been shown that there are a possible 230 different spatial arrangements with associated symmetries. These space groups can be described by a Hermann–Mauguin notation. As an example, the space group $P3_121$ is defined as belonging to the trigonal crystal system. It has a primitive (P) structure, a threefold screw axis ($3_1$) in the a-axis, a two-fold rotation parallel to the b-c-axis (2), and a one-fold rotation perpendicular to the b-c-axis (1). Although the one-fold rotation describes the lowest possible symmetry, it is added to the notation to clarify in which dimension the respective symmetry operations can be found.

### *1.7.2 XRD and obtaining an electron density map*

In X-ray crystallography the goal is to build a three-dimensional map of electron density, in order to determine the atomic arrangement of a molecule. In single crystal X-ray diffraction, diffraction images are captured on a two-dimensional surface. Therefore, to build a three-dimensional electron density map the crystal needs to be rotated to collect all the unique diffractions. Data collection is performed by placing a single crystal between an X-ray source and detector (Figure 1. 23). As the incident X-rays diffract, the crystal is rotated to image the

diffraction in different orientations. These diffraction images can then be combined to contain information of the diffraction pattern in three-dimensional space.

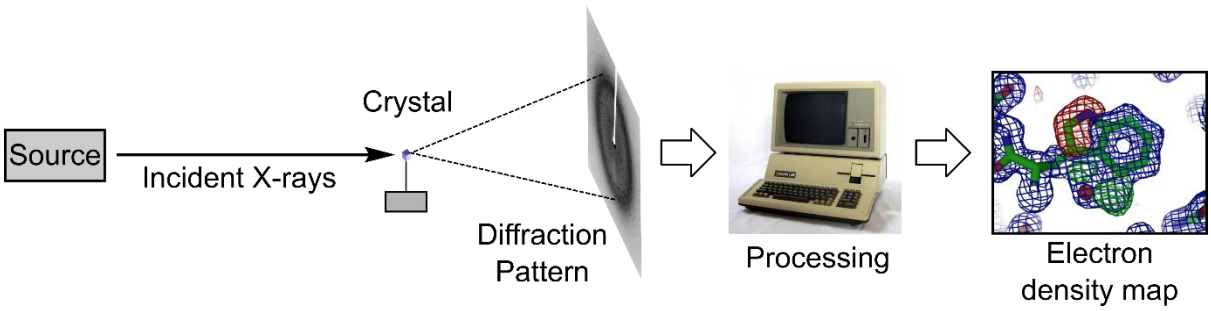



*Figure 1. 23 Overview for obtaining an electron density map using X-ray crystallography. Incident X-rays are targeted towards the crystal which is rotated to capture diffraction patterns from multiple orientations. The diffraction images are then processed into a single data set which can be used to calculate the electron density throughout a crystal lattice. (Electron density map adapted from reference[51])*

The electron density within a crystal lattice is calculated from the diffraction pattern.[52] To calculate an electron density map information about the amplitude and the phase of the diffracted wave is needed. The amplitude can be determined by measuring the pixel intensity from the diffraction pattern and the lattice geometry determined by measuring the distance between the Miller indices. However, during the experiment information about the phase cannot be recorded and in crystallography is known as the phase problem. Each wave is a product of interference with all other diffracted waves, if the phase information is obtained, then the electron density throughout the lattice can be solved.

In modern macromolecular crystallography there are two common ways in which the phase of the X-ray can be determined: experimental phasing or molecular replacement.[53] Friedel's law states that a given diffraction on a plane h, k, l will have an intensity equal to the diffraction at -h, -k, -l. However, Friedel's law is false, and therefore in some cases these intensities are not equal. Atoms that produce this effect are called anomalous scatterers. Common atoms in protein structures such as carbon, nitrogen or oxygen produce negligible anomalous scattering, but heavier atoms do. Heavy atoms such as Fe or Zn were first used to determine the phase of diffraction by breaking degeneracy between Friedel pairs.[54] However, many atomic species can be used, so protein structures are often solved using heavy atoms such as selenium[55] from selenomethionine residues or more recently the sulfur[56] atoms in methionine or cystine.

For oligonucleotides, experimental phasing is also a common method to retrieve phase information. Heavy atoms can be soaked into the crystal lattice to act as anomalous scatterers, however, it is more common to incorporate a modified residue containing bromine[57] either by transcription or solid phase synthesis. Bromine is a good anomalous scatterer because it can be easily collected at accessible X-ray energies, but unfortunately can also disrupt the crystallisation process reducing crystal quality. In an ideal situation, it is advantageous to phase a structure from a naturally abundant atom. In theory it should be possible to phase an oligonucleotide structure from the phosphorus atoms. In practice this has proven difficult, even though the strength of anomalous scattering is similar to sulfur, it is thought that the periodical stacking and movement of the phosphorus atoms in the lattice make it difficult to locate them by experimental phasing. However, some Z-DNA structures have been solved and it was suggested this is because the phosphorus atoms are stacked vertically in a more rigid structure.[11]

Alternatively, if a similar structure has already been obtained, then it is possible to solve an electron density map using molecular replacement (MR).[53] Molecular replacement uses the atomic positions from a previously solved diffraction pattern and applies this to new diffraction data to solve the new electron density map. For successful MR, the model and target structures must have > 25% sequence identity, combined with a root mean square deviation (r.m.s.d) of < 2.0 Å. For oligonucleotides specifically, the r.m.s.d of the phosphate atoms should ideally be within 1.0 Å. A common technique for MR is the maximum likelihood method,[58] which is based on a probability function. Atomic coordinates are taken from the existing model and converted into structure factors. Using r.m.s.d to evaluate the solutions, the model is rotated and translated to find the best fit for the new observed data. Once the model has been correctly placed, a Fourier transform can calculate the phase information based on the model coordinates and apply it to the reflection data to produce a new structure. Any differences between the model and actual data can then be corrected in downstream processing.

After the phase of the wave has been obtained, the final stage of crystallography is to refine the structure. This is performed by looking at the electron density map and editing the model to account for differences between the electron density and atomic model. Using a statistical

score called $R_{free}$, suitable models are then improved by multiple rounds of refinement until the fitted model matches the obtained values.

## 1.8 Aims and objectives of this work

The aim of this thesis is to design and incorporate modified backbones into oligonucleotides and to assess their properties for various applications. Within this work, we focus on three different backbone structures, which include carbamates, amide and triazole linkages. The specific aims of this work are;

Chapter 2 – To develop synthetic procedures for the synthesis of carbamate-LNA phosphoramidite dimers, to incorporate these dimers into oligonucleotides, and to assess the biophysical properties of carbamate-LNA modifications for potential antisense applications.

Chapter 3 – To assess the structural properties of LNA-amide modifications using X-ray crystallography, and link amide structural features to thermodynamic properties. Alongside this, we aim to provide a structural comparison between LNA-amide and LNA-carbamate modifications.

Chapter 4 – To use the CuAAC reactions to synthesis circular mRNA containing modified triazole backbones. To apply these circular RNAs to cell free protein expression, and to confirm successful peptide expression using immunoblotting techniques.

# 2 Synthesis and biophysical properties of Carbamate-Locked nucleic acids for antisense therapeutics

## 2.1 Antisense, modifications and therapeutics

### *2.1.1 Antisense mechanisms*

Antisense oligonucleotides (ASOs) are short ~20 mer sequences that modulate RNA processing and protein expression. Initially developed over four decades ago, these sequences bind to complementary RNA via Watson-Crick base pairing which provides sequence specific targeting.[59] Although showing exciting potential, first generation ASOs often failed to meet therapeutic targets because of fast enzymatic degradation and poor cellular uptake.[59–61] Since then there have been multiple generations of antisense templates that contain a variety of different backbone, sugar and base modifications.[62] Enhanced ASOs act through one of several mechanisms which include: steric blocking, exon-skipping, RNase-H activation or RNA interference.[63] Although all these processes originate from antisense oligonucleotide Watson-Crick base pairing, their effects on downstream functionality and regulation are very different.

Steric blocking ASOs (Figure 2. 1A) are designed to bind to complementary mRNA and prevent binding of ribosomal subunits in turn regulating gene expression.[64] They are often modified to prevent nuclease degradation and are able to regulate protein expression or alter splice switching events.[63] Like many ASO they can suffer from poor cellular uptake and in some cases high levels of modifications can result in chemical toxicity. Alongside this, the steric blocking mechanism is stochiometric, which means that one ASO blocks a single mRNA molecule. In alternative mechanisms such as RNase H induction and RNAi, the ASO is catalytic, meaning that a single sequence can act on multiple targets. As a result, the steric blocking mechanism may require higher doses for therapeutic efficacy.

ASOs can also mediate protein expression by invoking RNase H activation (Figure 2. 1B). RNase H is a nuclease enzyme that can be used to degrade the RNA strand in DNA:RNA hybrids. For this mechanism to occur, DNA ASOs bind to their complementary RNA sequence. RNase H then degrades the mRNA resulting in downregulation of protein translation. Although this catalytic mechanism is very efficient, modifications to the ASO are limited to those that are

compatible with the RNase H enzyme. The Phosphorothioate (PS) backbone is one of the few modifications tolerated by RNase H and has been regularly used within ASOs.[65,66] Modifications such as LNA do not invoke RNase H activation, however, they can be used in the terminal regions of the antisense oligonucleotide. Such gapmer designs modify the flanking regions of the DNA, while leaving the central region unmodified, which still enables RNase H activation.[44] In a review by Bennett *et. al.*[67] among the 54 ASO drugs reviewed approximately one third of them are gapmers.

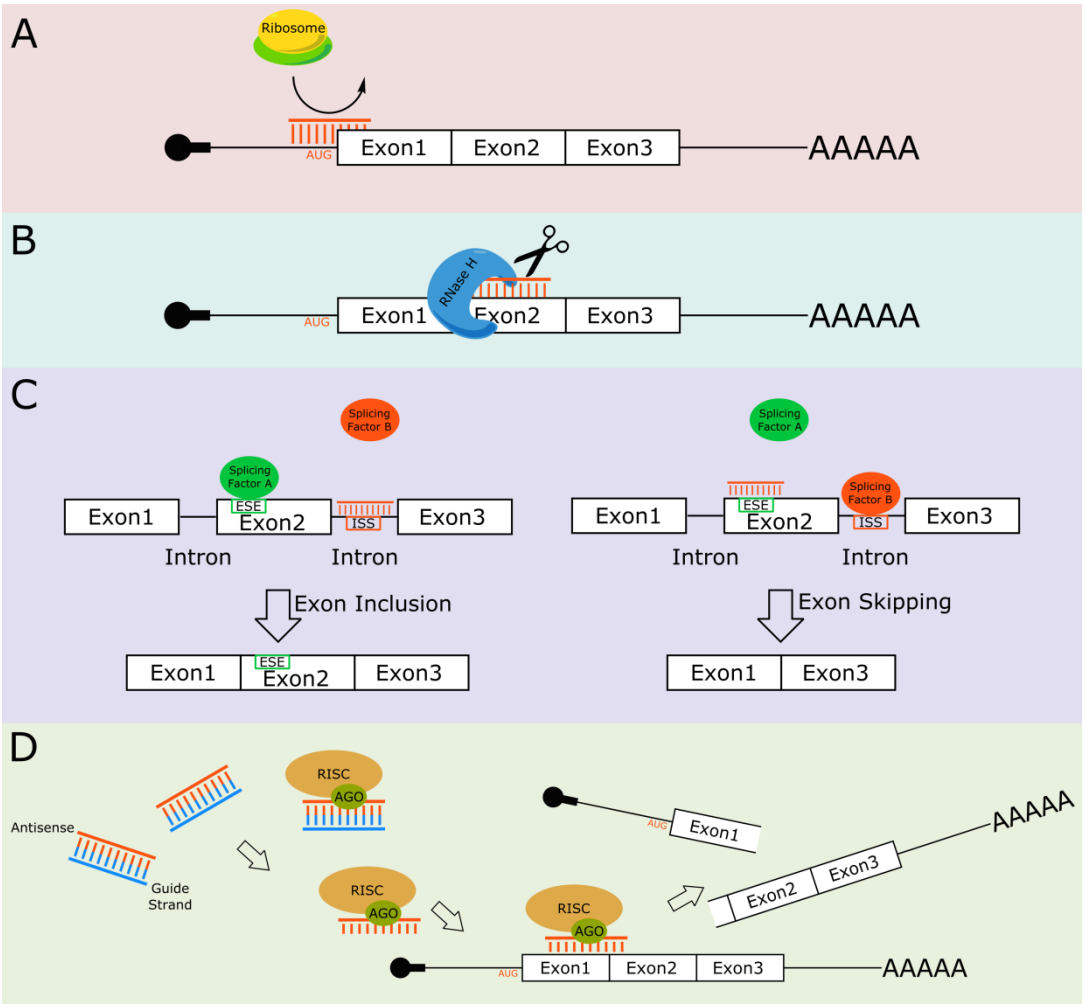



*Figure 2. 1 Overview of antisense mechanisms, A) Steric blocking ASOs, B) RNase H activated RNA cleavage, C) Exon skipping and Exon inclusion using splice switching oligonucleotides, ESE = exonic splicing enhancer, ISS = intronic splicing silencer, D) RNA interference (RNAi), (figure adapted from references.*[63,68–70]*).*

Eukaryotic pre-mRNA is expressed within the cell nucleus and contains sections of coding exon regions interspersed with non-coding introns. To convert pre-mRNA into mature mRNA, introns are spliced by a series of splicing factors which remove the non-coding regions and ligate the exons to produce a continuous coding mRNA (Figure 2. 1C).[70] Processing of pre-mRNA into mature mRNA requires high fidelity, obtained by matching the RNA sequences but

also by RNA-RNA interactions from small nuclear RNAs contained with the spliceosome protein complex. Antisense oligonucleotides can induce exon inclusion by interaction with intronic silencer regions. Sequence specific blocking of these regions prevents splicing factor binding while leaving enhancer regions open. Conversely, ASO sequences can be targeted to enhancer regions to induce exon skipping.[71]

Splice switching mechanisms are useful for their ability to alter the reading frame of mRNA, influencing downstream protein expression. Dystrophin is a protein that can be found within muscle cells and functions by linking the cell internal cytoskeleton to the outer myofiber membrane.[72] Duchene's muscular dystrophy (DMD) is a genetic X-linked disease that is caused by point mutations within the dystrophin gene. Point mutations result in a frame shift in the mRNA which when translated creates a non-functional protein instead of the original dystrophin sequence.[73] Without functional dystrophin, the myofiber membrane breaks when force is applied to the muscle fibres, which cannot be regenerated quickly and so are replaced with fat and scar tissues. Eventually degeneration of the muscle fibres causes restricted mobility and eventually patient death.

Splice switching oligonucleotides can be used for the treatment of DMD by binding to specific sites on the dystrophin gene. This promotes the removal of one or more exons which allows for the realignment of the codon sequence. This method to treat DMD is one of very few options available to patients but the mechanism has two significant drawbacks. The resultant shortened gene has reduced functionality compared to the full-length dystrophin protein. In addition to this, only a small amount of the dystrophin gene is targeted by the antisense oligonucleotide. This means that most of the dysfunctional proteins remain unaltered, along with a small percentage of truncated dystrophin produced from splicing in the presence of the antisense oligonucleotide. Since only a small fraction of total *in vivo* dystrophin is altered, it is unable to cure the condition and partial restoration only reduces the severity of DMD. This condition with reduced symptoms is called Becker's muscular dystrophy.[74]

RNA interference (RNAi) is perhaps the most promising and effective target mechanism for antisense applications (Figure 2. 1D). Pre-miRNAs are transported from the nucleus into the cytosol and processed by an endoribonuclease called Dicer.[75] Here RNAs are degraded into short 20 mer sections called miRNA. miRNAs are double stranded RNA consisting of a passenger strand and antisense guide strand. Utilising the selective nature of Watson-Crick

base pairing, the antisense strand can then be encoded to regulate gene expression by blocking mRNA. The RNA-induced silencing complex (RISC) is a combination of enzymes and cellular machinery which when combined with miRNA can be used as a scanning tool to seek out and destroy mRNA.[76] For this to occur, miRNA is loaded into the Argonaute 2 (Ago2) protein where the passenger strand is removed to expose the complementary antisense sequence. Binding using this selective sequence, allows the capture of mRNA for degradation or further processing.

Being an endogenous mechanism, RNAi presents a versatile tool that can be adapted to many needs. By introduction of exogenic miRNA known as small interfering RNA (siRNA), sequences can be used to target many genes of interest. The mechanistic details have been well defined, and a range of RNA modifications can be used to enhance siRNA efficacy. Eric Rozners and co-workers[77] have made significant progress in the use of amide backbones for siRNA both in the guide and passenger strands. Modification to the siRNA is important as it allows for improvement of biological efficacy, particularly if it enhances oligonucleotide delivery or target specificity.[78]

### *2.1.2 Antisense modifications and drugs*

For suitable efficacy, ASOs need to be modified to improve their biological properties. The 2'-*O*-methyl modification was one of the earliest sugar modifications and is naturally occurring in some RNAs post transcription *in vivo*.[79] Addition of 2'-OMe (Figure 2. 2A) to ASOs, increases RNA binding affinity,[80] and also improves resistance towards degradation by nucleases.[81] However, it does not induce RNase H cleavage. Other modifications have also been examined, such as the 2'-fluoro (Figure 2. 2B), but the most promising 2'-modification has proven to be the 2'-(2-methoxyethyl) (2'-MOE, Figure 2. 2C), which provides enhanced target selectivity compared to 2'-OMe and good nuclease stability.[82] Although it also does not enable RNase H induced cleavage, it was found that it could be used in a gapmer design. Emerging from this modification came several clinical trials and the approval of Mipomersen (Kynamro, Figure 2. 3).[83] Mipomersen has 2'-MOE wings and a phosphorothioate backbone, and is used for treatment of homozygous familial hypercholesterolemia, acting via the RNase H pathway. Due to high liver toxicity, Mipomersen was initially rejected by the FDA, but after approval in 2013 it can now be accessed for individual cases.[84]

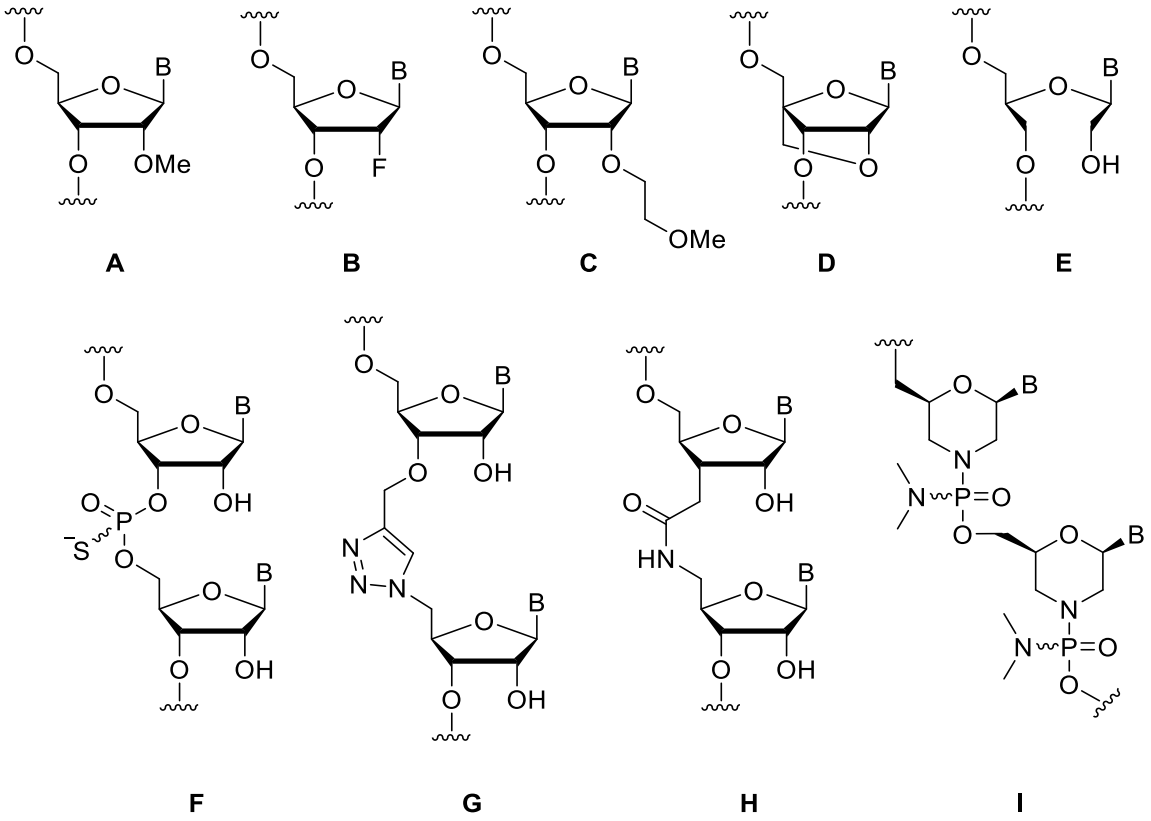



*Figure 2. 2 Common modifications used in antisense templates, A = 2'-O-Methyl, B = 2'-Fluoro, C = 2'-O-(2-Methoxyethyl) (2'-MOE), D = Locked nucleic acid (LNA), E = Unlocked nucleic acid (UNA), F = Phosphorothiolate, G = Triazole, H = Amide, I = Phosphorodiamidate morpholino oligomer (PMO).*

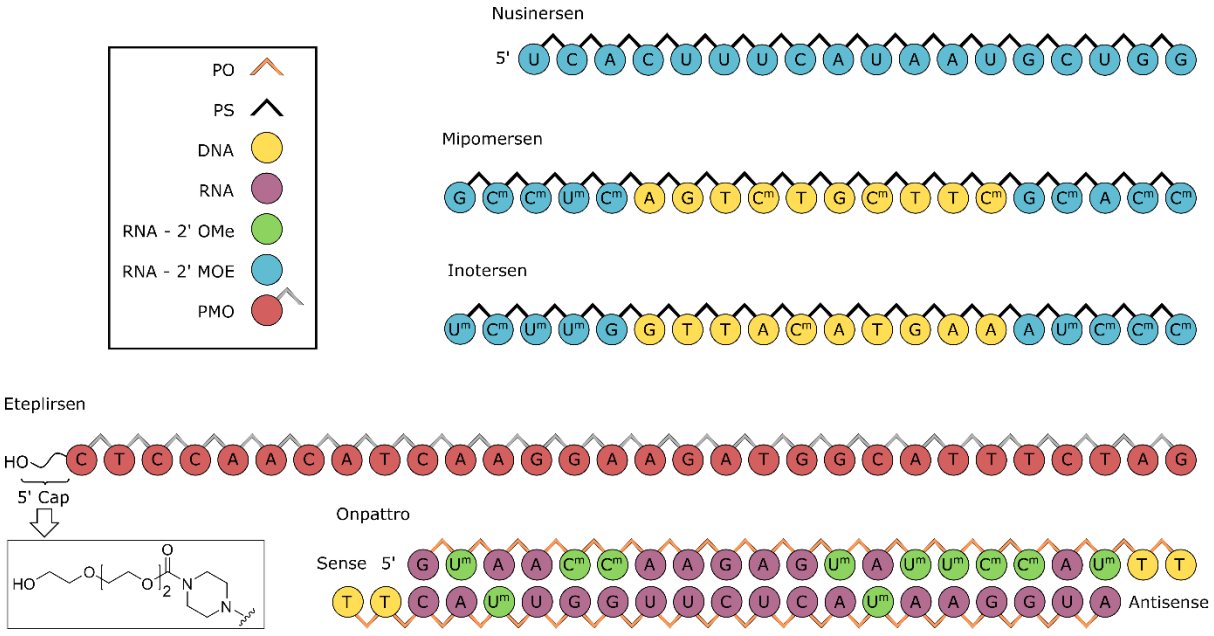



*Figure 2. 3 FDA approved antisense oligonucleotides with backbone and sugar modifications. PO = phosphodiesters, PS = phosphorothioates, OMe = 2'-O-methyl RNA, 2'-MOE = 2'-(2-methoxy ethyl), PMO = Phosphorodiamidate morpholino, $^{m}$ = methylation on the nucleobase at the 5-position.*

Locked nucleic acids in the context of ASOs have been intensely investigated (Figure 2. 2D). LNA can provide high levels of duplex stabilisation due to its constrained sugar conformation.

When compared against sequences containing 2'-MOE modification, LNA modified ASOs showed improved binding affinity. Although beneficial when used at lower drug concentrations, increased drug concentrations combined with a large number of LNA modifications can create binding to off-target sequences and hepatoxicity.[43,85] However, overall LNA is a highly promising modification for therapeutic oligonucleotide drugs. When incorporated in to ASOs it provides good sequence discrimination and work by Wengel *et al.*[44] highlights that with careful design, LNA oligonucleotides can be potent and non-toxic. Efficacy is not only limited to RNase H induction, studies have also provided evidence for good *in vivo* activity of LNA oligonucleotides as steric blockers and in RNAi.[86,87]

The unlocked nucleic acid (UNA) modification has the opposite effect to LNA; it is a highly flexible sugar modification that destabilises oligonucleotide duplexes by around -7 °C per modification (Figure 2. 2E).[88] Despite these destabilising properties, UNAs have been found to be viable for antisense applications. UNAs incorporated into various positions of the antisense oligonucleotide were found to have increased resistance to nuclease degradation, and that modification to either the sense or antisense strands can improve siRNA gene knockdown.[89] In addition to this, UNA can also reduce siRNA toxicity and potential off target effects, by preventing binding to the seed regions of other mRNA sequences.[88]

UNA is often used for RNAi applications because of its ability to reduce the competitive binding of the passenger strand. RNAi oligonucleotides are administered as duplexes consisting of an antisense guide strand and passenger sequence. When this duplex is loaded into the RNA induced silencing complex (RISC), the passenger strand is then removed. However, if the binding affinity of the passenger strand is too high, it can cause nonspecific binding and off-target effects by uptake into the RISC complex instead of the guide ASO. UNA can be used to reduce the hybridization strength of the ASO passenger, and in turn improve antisense strand selection by introducing a thermodynamic bias between the two strands. In this capacity both LNA and UNA can be used to alter the stability of the two RNAi strands.[89] In general UNA is a versatile modification and has be found to be beneficial modification when incorporated into either the antisense or passenger strands.

The phosphorothioate backbone (Figure 2. 2F) is the most studied and widely used modification in ASOs. It enhances ASO stability towards non-specific nucleases and is one of few modifications that can induce RNase H cleavage of complementary RNA.[65] As previously

discussed, phosphorothioates contain a stereocentre at the phosphorus atom. This creates two diastereoisomers at every phosphorothioate unit and when in oligonucleotides have different chemical properties.[90] The relationship between configuration at the phosphorothioate groups and ASO properties is complex, and it is suggested that the ideal isomer depends upon both target sequence and position within the ASO. Recently it was stated that ASO efficacy is influenced by distinct structural motifs such as 3'-$S_pS_pR_p$ rather than individual stereocenters.[31] Since there are $2^n$ diastereomers for each phosphorothioate group, where n = nucleotide length, a 20 mer ASO could have over 1 million configurations (distinct molecules) each with unique biophysical properties. Recent efforts have focused on the controlled synthesis of stereochemically-defined phosphorothioate stereocentres by the use of 3'-oxazaphospholidine derivatives.[31] It has also been shown that PS linkages can be synthesised with defined stereochemistry using $P^{(V)}$ and (+/-)-limonene as a chiral auxiliary.[30,71] However, coupling yields were low, which could make it difficult to produce oligonucleotides in high purity and yield.

Degenerative muscular diseases are good targets for antisense therapeutics and are sometimes deemed as “undruggable” by small molecule approaches. Approval of ASO drugs represents one of few approaches for treatment of such diseases. Nusinersen (spinraza) (Figure 2. 3) is a PS based drug that is used for the treatment of spinal muscular atrophy. Approved in 2016 by the FDA, Nusinersen promotes exon inclusion via the splice switching mechanism and corrects a frame shift in the mRNA sequence which restores biological functionality of the downstream protein.[91] Inotersen (Figure 2. 3) is another ASO that uses PS linked nucleotides which are combined with MOE modifications. It was recently approved in 2018 and is used for the treatment of hereditary transthyretin amyloidosis.[92]

The triazole backbone (Figure 2. 2G) is a charge neutral linkage that contains no chiral centres and is synthesised using the CuAAC reaction. Without additional modifications, triazoles are duplex destabilising modifications,[93] which makes them less than ideal candidates for antisense therapeutic oligonucleotides. Work by P. Kumar *et. al.*[94,95] sought to improve the triazole linkage by combining it with LNA (Figure 2. 4). A seven-bond artificial backbone linkage flanked by LNA sugars was shown to restore the stability of duplexes that contained triazole linkages. In parallel, V. Sharma *et. al*.[96] designed a five-bond triazole linkage combined with the LNA sugar modification. With this shorter linkage, the triazole was less able to

accommodate the LNA modification and caused significant disruption to the backbone. In both cases it was also observed that the position of LNA relative to the triazole backbone influenced the stability of the duplex. The use of LNA in conjunction with modified backbones needs further study, and significant insight could be obtained by investigating a variety of modified backbones beyond triazole systems.

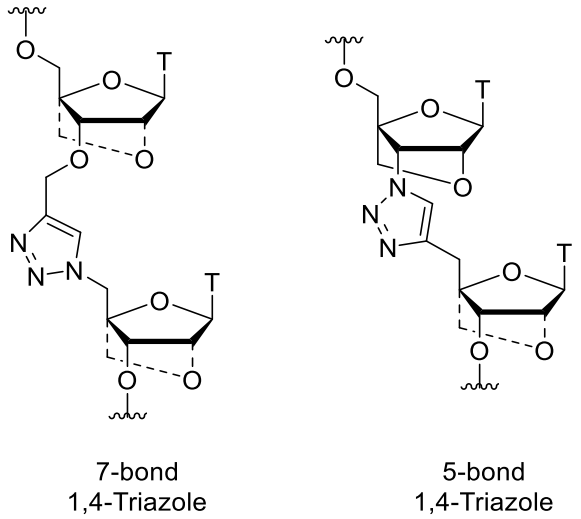



*Figure 2. 4 Comparison between two LNA-triazole modifications. (Left) Extended seven bond triazole linkage with increased flexibility toward the 5'-sugar (P. Kumar et. al.[94,95]). (Right) A short five-bond triazole linkage with reduced fixability and $sp^2$ hybridised nitrogen adjacent to the 5'-sugar (V. Sharma et. al.[96] ).*

The amide backbone (Figure 2. 2H) is a five-bond flexible linkage that has been shown to have good biocompatibility for polymerases and has similar duplex stability to an unmodified phosphodiester.[97,98] Despite this, amide modifications are still to reach maturity within the antisense field. The reason for this is the complex synthesis needed to create the amide linkage. Natural nucleic acids contain an oxygen at the 3'-position, but synthesis of an amide linkage requires a 3'-$CH_2$ group that links the sugar to a carboxylic acid. With the synthetic challenge posed by the amide backbone, functional groups with similar structures and an easier synthesis, could provide a viable alternative.

Uncharged neutral backbones such as phosphorodiamidate morpholino oligomers (PMO) (Figure 2. 2I) are particularly interesting due to their enhanced ability to enter cells.[99] One of the biggest challenges facing the antisense field is getting therapeutic ASOs through the cell membrane, out of the endosome and into the cytosol. Because of the negative charge on the canonical phosphodiester backbone (and phosphorothioates) they are repelled by net overall charge of the phospholipid fatty acids of the outer cellular membrane. Removal of

oligonucleotide charge is therefore favourable and PMO structures have been shown to have improved cell penetration.[100] Morpholino oligonucleotide analogues have also been combined with other neutral frameworks such as the triazole in an attempt to enhance hybridization compared to normal triazole backbones.[101]

Eteplirsen (Figure 2. 2I and Figure 2. 3) was the first PMO drug to be approved in the US and was accelerated for approval by the FDA.[102] Binding to enhancer regions within pre-RNA, Eteplirsen is able to promote exon skipping, particularly important for treatment of Duchene's muscular dystrophy as it allows for frame shift correction and expression of a truncated dystrophin protein that restores partial functionality.[103]

## 2.2 Aims and objectives to develop LNA-carbamates as potential antisense therapeutic modifications

The concept behind development of LNA-carbamates as a therapeutic oligonucleotide is to combine the charge neutrality and enzymatic stability of the carbamate linkage with the thermodynamic stability of LNA. The carbamate backbone is a five-bond linkage with two constitutional isomers (Figure 2. 5). Compared to the triazole, it has higher flexibility and unlike phosphorothioates does not have a chiral centre. It is also charge neutral like PMO or PNA backbones and is easy to synthesise compared to amide modifications. Carbamates for antisense applications were first explored by De Mesmaeker *et al.*[32,104] but were found to cause high levels of duplex disruption and were therefore unsuitable for clinical use. By combining them with LNA we hope to restore duplex stability and enhance the desired properties for therapeutic applications. In this chapter I outline the design and synthesis of LNA-carbamate phosphoramidite dimers, and their incorporation into DNA using standard solid phase synthesis. LNA will be added either to the 3'-position, 5'-position or to both sides flanking the modified backbone (Figure 2. 5). Oligonucleotides containing LNA-carbamates will then be analysed using ultraviolet melting hybridisation studies to assess the best modification. Structural analysis by circular dichroism and enzymatic digestion are also discussed.

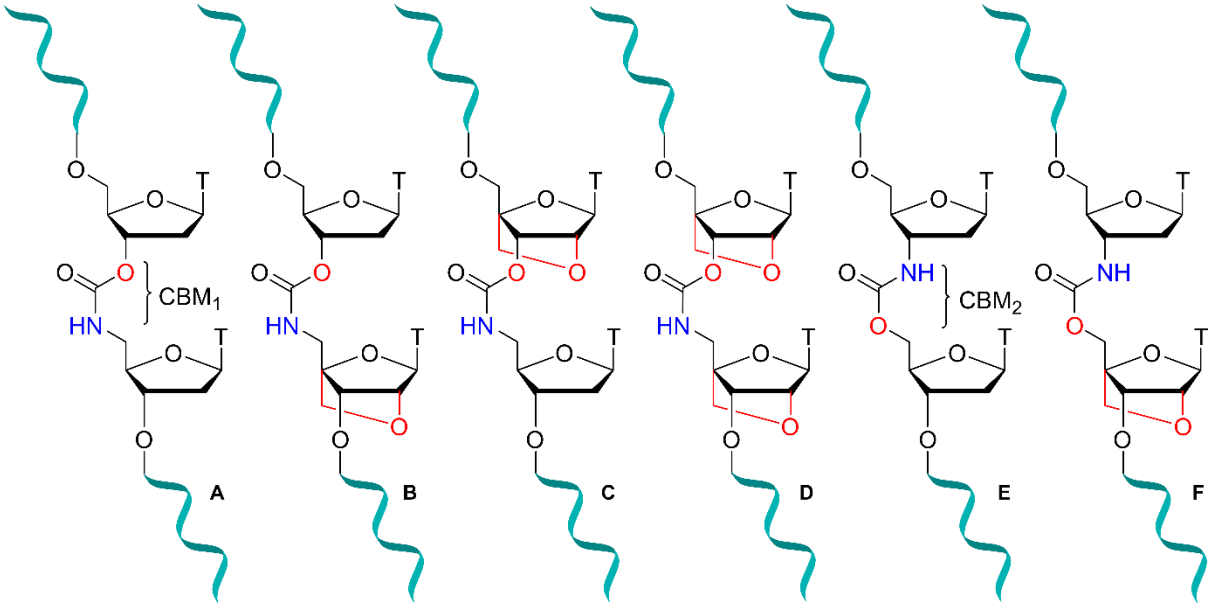



*Figure 2. 5 Modified carbamate-LNA DNA oligonucleotides. Set 1 (A – D) and set 2 (E, F) comprising of constitutional carbamate $CBM_1$ = 5'O, 3'N and $CBM_2$ = 3'O, 5'N*

## 2.3 Synthesis of LNA-$CBM_1$ dinucleotides

### *2.3.1 $CBM_1$ electrophiles*

LNA-carbamate modifications will be incorporated into oligonucleotides as dimer phosphoramidites. The dimers will comprise of a carbamate backbone flanked with either DNA or LNA sugars. To simplify the chemical synthesis, thymine nucleobases will be adjacent to the carbamate backbone, since they do not contain any exocyclic amines and therefore do not require protecting groups. From these simple thymidine building blocks, carbamate 1 ($CBM_1$) linkages can be synthesised by the ligation of a 3'-electrophilic and 5'-nucleophilic centre.

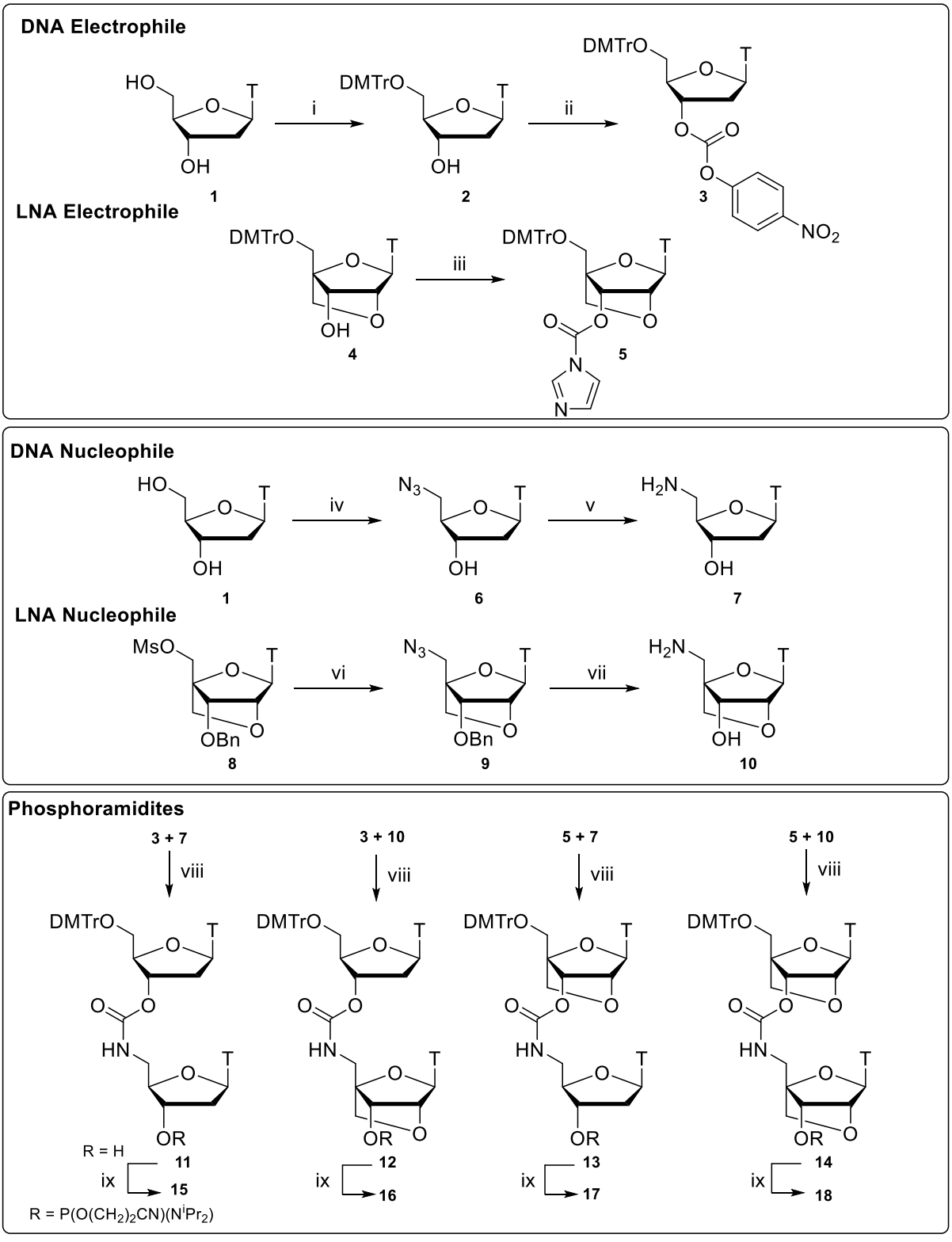



*Figure 2. 6 Synthesis of $CBM_1$-linked dinucleotides **15**–**18** using different combinations of 3'-activated carbonates **3**, **5** and 5'-amines **7**, **10**. Reagents and conditions (i) DMTrCl, pyridine, rt, 20 h, **2** 78% (ii) p-nitrophenyl chloroformate, pyridine, 80 °C, 20 h, crude **3** 60%; (iii) CDI, THF, rt, 20 h, crude **5** 84%; (iv) $NaN_3$, $PPh_3$, $CBr_4$, DMF, rt, 20 h, **6** 74% (v) $H_2$, Pd(10%)/C, MeOH, rt, 4 h, **7** 95% (vi) $NaN_3$, DMF, 65 °C, 20 h, **9** 90%; (vii) $H_2$, $Pd(OH)_2$ (20%)/C, MeOH, rt, 3 h; $NH_4HCO_2$, reflux, 4 h, **10** 87%; (viii) DMAP, pyridine, 80 °C, 20 h, **11** 62%, **12** 51%, **13** 63%, **14** 70%; (ix) Chloro(diisopropylamino)-β-cyanoethoxyphosphine, DIPEA, $CH_2Cl_2$, rt, 1 h, **15** 54%, **16** 57%, **17** 36%, **18** 88%, (compounds **2-3**, **6-7**, **9** and **11** synthesis references [32,105–108]).*

For the synthesis of a 3'-activated electrophile, the 5'-hydroxyl of thymidine **1** was protected with a 4,4'-dimethoxytrityl (DMTr) group using DMTrCl. This $S_N1$ based reaction is efficient and produced DMTr-protected thymine **2** in 78% yield (Figure 2. 6). Without 5'-DMTr protection attempted activation of the secondary 3'-hydroxy group would instead result in attack by the less hindered primary 5'-hydroxyl. DMTr is an acid labile group with good solubility in many organic solvents. Following a previous reported synthesis by Waldner *et al.*[32] the 3'-hydroxyl group of compound **2** was activated using *p*-nitrophenyl chloroformate (PNP) to produce the 3'-electrophile **3**. Although this reaction proceeded efficiently by TLC, the labile nature of the activated PNP leaving group meant that during column purification it was often difficult to purify the product due to degradation in basic solvent or by heating above 40 °C.[109] In order to avoid degradation, activated electrophiles were therefore often used without further purification with no significant effect on subsequent coupling.

Benjamin Woods (MChem, part 2) performed a synthesis of the LNA-thymidine **4** from a published synthetic route by A. A. Koshkin et. al.[110] An adaption of this route was also used to produce all subsequent LNA starting compounds. Working alongside B. Woods, compound **4** was activated using 1,1'-carbonyldiimidazole (CDI) instead of the previously used PNP carbonate to produce compound **5**. Activation using CDI produced a cleaner reaction compare to the PNP group, with higher yield and less sensitivity to degradation in solution. Because LNA is a bicyclic ribose derivative there was some concern that the rate of reaction for the addition to the 3'-hydroxyl could be slower compared to the less hindered deoxyribose. In both cases, there was not a significant reduction in rate and both reactions went to completion within 20 h monitored by TLC.

### *2.3.2 $CBM_1$ nucleophiles*

Thymidine **1** was reacted with $NaN_3$ in the presence of $PPh_3$ and $CBr_4$ using an Appel reaction to produce the 5'-azide **6**. Reduction of the azide **6** in a hydrogen atmosphere with Pd/C yielded amine **7** in 95% yield, which was then filtered and used without further purification. Although in principle an Appel reaction followed by Pd/C reduction can work, there were issues that became apparent during synthesis. The Appel reaction uses $PPh_3$ with $CBr_4$ to produce a phosphonium cation, which then activates the hydroxyl to be replaced by bromine. $NaN_3$ is then able to displace the halogen producing the azide. The issue with this is that both $PPh_3$ and $CBr_4$ are used in large excess, which means that the product is difficult to purify by

silica chromatography resulting in a reduction in yield. Even in cases where compound **6** was seen to be pure by NMR, there would often be trace amounts of $PPh_3$ that would prevent azide reduction by poisoning the Pd catalyst. Because of this, care is required to ensure that all phosphorus species are removed from the azide before reduction, which sometimes required multiple chromatography columns.

In the reported LNA synthesis of compound **4**[110], a precursor 5'-mesyl-3'-benzyl protected compound **8**[111] is also synthesised (Sven Epple, DPhil). Mesyl groups are reactive toward nucleophilic attack which allows them to be reacted with $NaN_3$ to install an azide functional group to produce compound **9** (Sven Epple). Reduction of the azide under a $H_2$ atmosphere was performed using $Pd^{2+}(OH)_2$ and addition $HCO_2NH_4$ for 3 hrs allowed for simultaneous removal of the benzyl group to give compound **10** (Sven Epple). Compared to the Appel reaction used previous, azide conversion via the mesyl group was clean and effective. The absence of $PPh_3$ also meant that reduction and deprotection of the azide and benzyl groups were unhindered. Although both methods produced products in good yield, the reliability and consistency of the mesyl reaction provides a better synthetic route for future work.

### *2.3.3 $CBM_1$ phosphoramidites*

Electrophiles **3** or **5** were combined with nucleophiles **7** or **10** to produce $CBM_1$ dimers **11**-**14**. As to be expected, coupling of LNA electrophile **5** proceeded at a slower rate, but products were obtained in comparable yield to DNA versions (DNA compound **11** = 62% *vs* LNA compound **13** = 63%).

Phosphoramidites are sensitive to hydrolysis but also readily oxidised from $P^{(III)}$ to $P^{(V)}$ under ambient conditions (atmospheric $O_2$). Because of this phosphoramidites require inert conditions as well as dry and degassed solvents. Dimers **11**-**14** were converted to phosphoramidites using chloro(diisopropylamino)-β-cyanoethoxyphosphine (CEP) to give phosphoramidites **15**-**18** (Figure 2. 7, C. Thorpe, S. Epple and B. Woods). Like the previous monomer couplings, addition of the CEP group was slowed by the presence of LNA on the 3'-sugar, but overall yield was not substantially affected. With the appropriate 5'-DMTr protecting group and 3'-phosphoramidite, the dimers are suitable for oligonucleotide synthesis on solid support.

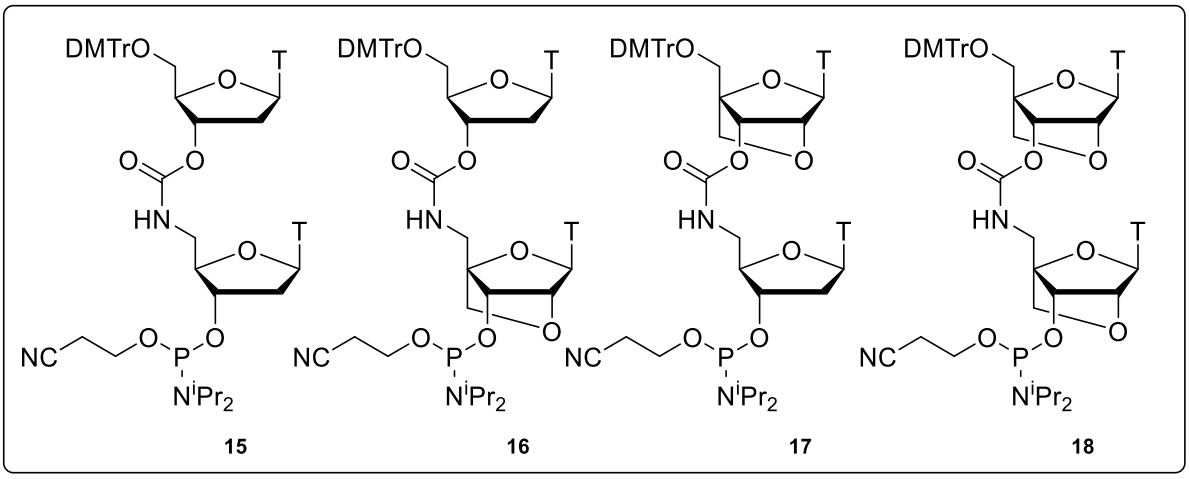



*Figure 2. 7 Structures of $CBM_1$ dimer phosphoramidites to be used in oligonucleotide solid phase synthesis*

## 2.4 Synthesis of LNA-$CBM_2$ dinucleotides

### *2.4.1 $CBM_2$ nucleophile*

$CBM_2$ linkages were synthesised using a 3'-nucleophile and 5'-electrophile, like the $CBM_1$ but with reversed connectivity (Figure 2. 8). From previous studies on carbamate linkages by Waldner *et. al.*[32] it was found that the $CBM_2$ linkages are more destabilising than the $CBM_1$. Recent biophysical experiments from our own group have shown that when using LNA-triazole backbones, addition of LNA to the 5'-position, relative to the linkage, was also highly destabilising.[95] For this reason, combined with the need for a long synthetic procedure to obtain a 3'-amine LNA monomer, we decided not to pursue the dimer with a 5'-LNA. Obtaining duplex melting temperatures ($T_m$ data) is relatively fast, so if preliminary $CBM_2$ backbones have good thermodynamic properties, we could return to the synthesis of the 5'-LNA-$CBM_2$.

For the DNA nucleophile, AZT **19** was purchased and protected with a 5'-*O*-DMTr group to produce compound **20**. This was subjected to reaction under a $H_2$ atmosphere on activated Pd/C to give nucleophile **21**. Both reactions proved to be simple and high yielding (**20** = 90%, **21** = 84%).

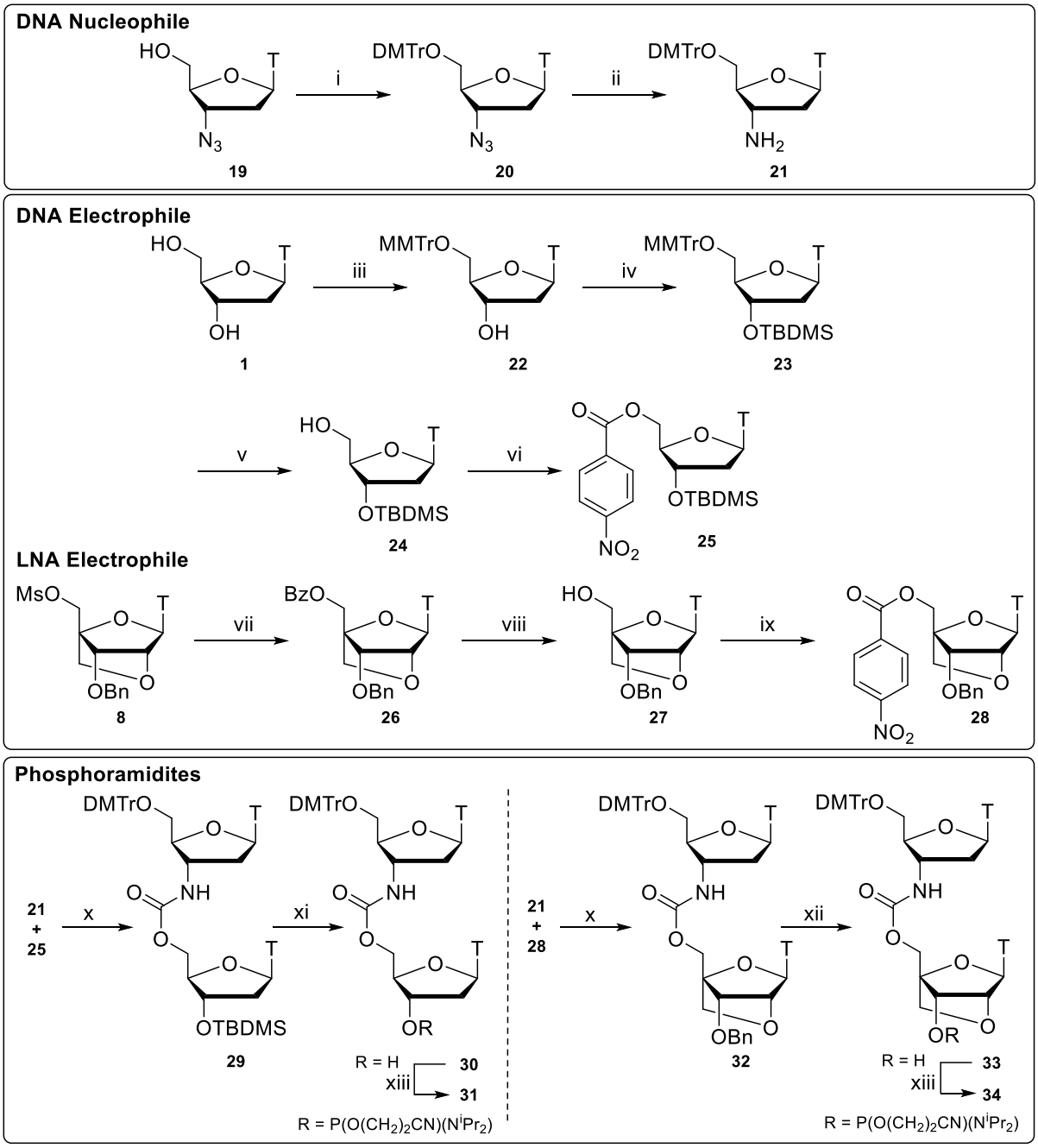



*Figure 2. 8 Synthesis of $CBM_2$ dinucleotides **31**, **34** using activated 3'-amine **21** and 5'-activated carbonates **23**, **28**. Reagents and conditions (i) DMTrCl, pyridine, rt, 20 h, **20** 90% , (ii) $H_2$, Pd(10%)/C, MeOH, rt, 3 h, **21** 84%, (iii) MMTrCl, pyridine, rt, 15 h, **22** 79%, (iv) TDBMSCl, imidazole, DMF, rt, 20 h, **23** 66%, (v) TCA, $CH_2Cl_2$, rt, 2 h, **24** 61%, (vi) p-nitrophenyl chloroformate, pyridine, 80 °C, 20 h, **25** 76%, (vii) NaOBz, DMF, 100 °C, 5 h, **26** 92%, (viii) NaOH, THF:$H_2O$ (1:1), rt, 1 h, **27** 81%, (ix) p-nitrophenyl chloroformate, pyridine, 80 °C, 20 h, **28** 73%, (x) HOBt, pyridine, 80 °C, 20 h, **29** 67%, **32** 63% (xi) TBAF, THF, 0 °C, 2 h, **30** 77%; (xii) $Pd(OH)_2$ (20%) /C, $NH_4HCO_2$, 65 °C, 2 h, **33** 97%; (xiii) Chloro(diisopropylamino)-β-cyanoethoxyphosphine, DIPEA, $CH_2Cl_2$, rt, 1 h, **31** 80%, **34** 43% (compounds **20-27** synthesis references[32,110,112–115] ).*

### *2.4.2 $CBM_2$ electrophiles*

Initially I attempted to synthesise the DNA 5'-activated electrophile by direct reaction of thymidine **1** with CDI to produce compound **35** (Figure 2. 9). However, reaction monitoring by TLC showed several products to form. Analysis by mass spectrometry and NMR showed no

evidence of formation of compound **35** (Figure 2. 9). In this case it is suggested that the 5'-intermediate could undergo degradation in basic conditions resulting in formation of compound **36** by a 6-exo-trig intramolecular cyclisation (Figure 2. 9). To overcome this, a 5'/3'-protection strategy was used, like that used by Utagawa *et. al.*.[116] Thymidine **1** was reacted with MMTrCl to give the 5'-*O*-MMTr protected compound **22**. This was then protected with TBDMSCl at the 3'-position to yield the 5'/3'-protected monomer **23**. Removal of 5'-*O*-MMTr under acidic conditions produced compound **24**, which was activated using PNP to give the electrophile **25**. Using 3'-*O*-TBDMS prevented intramolecular cyclisation from happening, but at the cost of increased synthetic length. The compounds were obtained in reasonable yield (24% over four steps, **1**→**25**) and proved to be easy to handle due to their high solubility in organic solvent provided by the protecting groups.

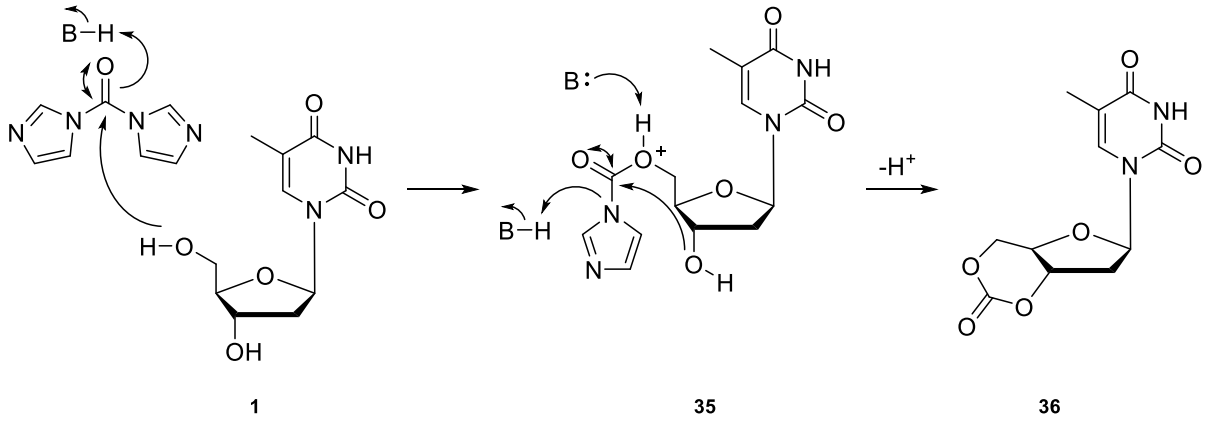



*Figure 2. 9 Mechanism of 6-exotrig 5'-3' intramolecular cyclisation of desired compound **35**. Deprotonation in basic conditions allows for nucleophilic attack and activation of the 5'-hydroxyl. This is then removed by a second intramolecular attack by the 3'-hydroxyl to form the bicycle compound **36**.*

5'-OMs-LNA compound **8** was converted to 5'-OBz compound **26** following a published procedure.[110] Treatment of **26** with NaOH at room temperature provided a simple deprotection method and compound **27** was collected in 81 % yield within 1 h. The activated LNA electrophile **28** was formed by reaction with PNP and was used without further purification.

### *2.4.3 $CBM_2$ phosphoramidites*

$CBM_2$-containing dinucleotides were obtained by addition of nucleophile **21** to either electrophile **25** or **28** in basic conditions to give DNA-dimer **29** and LNA-dimer **32** respectively. Before the dinucleosides could be converted to phosphoramidites, the 3'-protecting groups were removed. DNA-dimer **29** was treated with TBAF to facilitate the deprotection of the 3'-

TBDMS group to give compound **30**. In parallel, the LNA-dimer was reduced under a hydrogen atmosphere with $Pd(OH)_2/C$ to give compound **33**. Dimers **30** and **33** were converted to phosphoramidites using chloro(diisopropylamino)-β-cyanoethoxyphosphine (CEP) to give phosphoramidites **31** and **34** (Figure 2. 10).

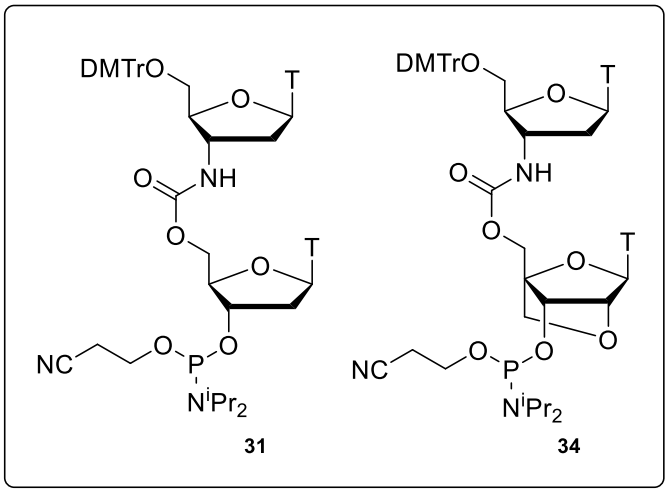



*Figure 2. 10 Structures of $CBM_2$ dimer phosphoramidites to be used in oligonucleotide solid phase synthesis*

## 2.5 Synthesis of LNA-CBM oligonucleotides

LNA-CBMs dinucleotides **15-18**, **31** and **34** were incorporated into DNA sequences using standard solid phase chemistry. DMTr cation monitoring was used to evaluate coupling efficiencies and the oligonucleotide was synthesised by successive cycles, followed by cleavage and deprotection in aqueous ammonia. For each LNA-CBM, dimers were incorporated into an oligonucleotide one time (table 2. 1, ON2.2-2.7, $T^L$ = T-LNA, **x** = $CBM_1$, **y** = $CBM_2$) and three times (ON2.8-2.13) into separate sequences. When adding the dimer an extended coupling time of 10 minutes was used on the synthesiser to ensure efficient synthesis. After solid phase synthesis oligonucleotides were purified by HPLC and their identity was confirmed by UPLC-MS (Figure 2. 11, Table 7. 1). In all cases sequences were obtained as single peaks with corresponding masses within 2.0 Da of the calculated values.

*Table 2. 1 Table of modified oligonucleotides with single or triple addition of CBM, LNA or LNA-CBM modifications x = $CBM_1$, y = $CBM_2$, $T^L$ = LNA*

| Oligonucleotide / ON | Sequence /5'→3' |
|---|---|
| ON2.1 | GCTTGCTTCGTTCC |
| ON2.2 | GCTTGCT**x**TCGTTCC |
| ON2.3 | GCTTGCT$^L$**x**TCGTTCC |
| ON2.4 | GCTTGCT**x**T$^L$CGTTCC |
| ON2.5 | GCTTGCT$^L$**x**T$^L$CGTTCC |
| ON2.6 | GCTTGCT**y**TCGTTCC |
| ON2.7 | GCTTGCT**y**T$^L$CGTTCC |
| ON2.8 | GCT**x**TGCT**x**TCGT**x**TCC |
| ON2.9 | GCT$^L$**x**TGCT$^L$**x**TCGT$^L$**x**TCC |
| ON2.10 | GCT**x**T$^L$GCT**x**T$^L$CGT**x**T$^L$CC |
| ON2.11 | GCT$^L$**x**T$^L$GCT$^L$**x**T$^L$CGT$^L$**x**T$^L$CC |
| ON2.12 | GCT**y**TGCT**y**TCGT**y**TCC |
| ON2.13 | GCT**y**T$^L$GCT**y**T$^L$CGT**y**T$^L$CC |

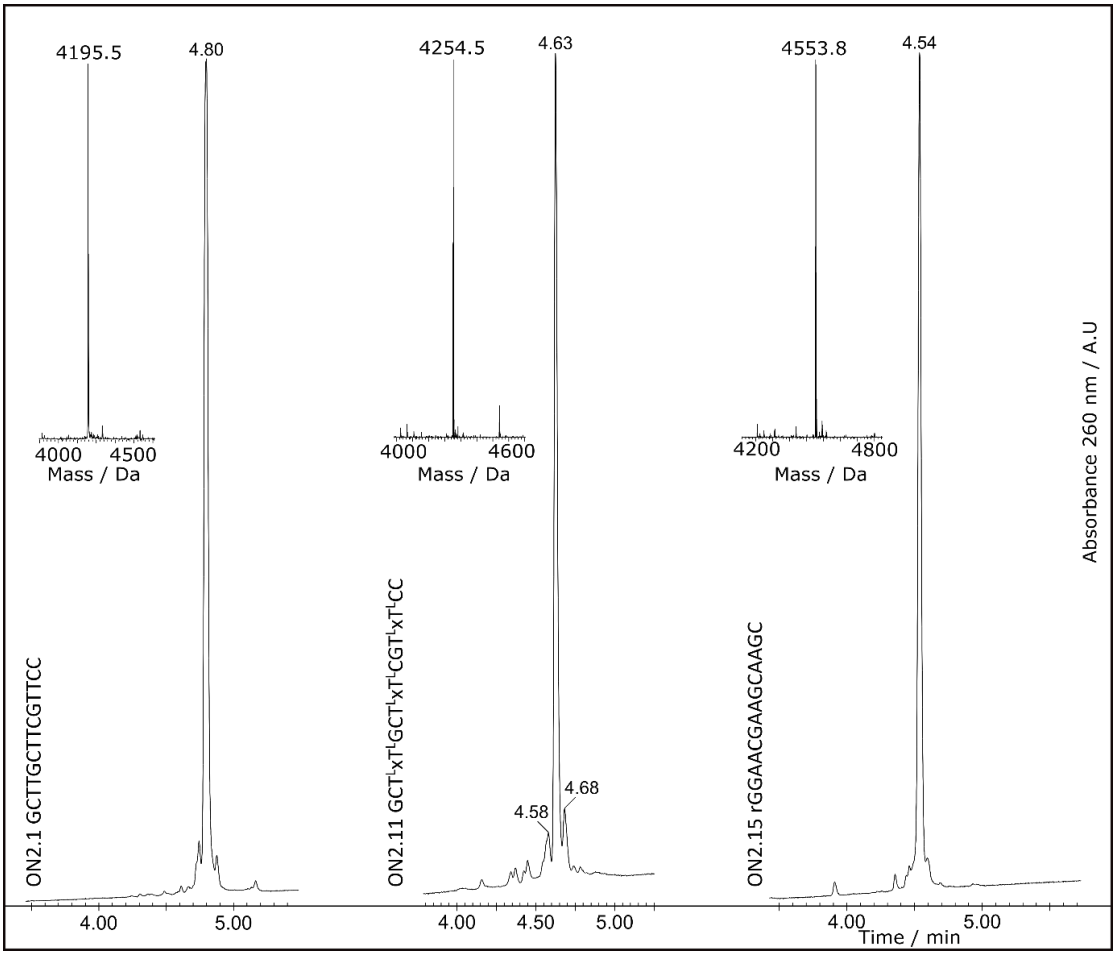



*Figure 2. 11 Example of typical UPLC-MS results obtained from solid phase synthesis and HPLC purification. Highlighted above is the unmodified DNA sequence ON2.1 (calc. mass 4197 Da), triple LNA-$CBM_1$-LNA modified sequence ON2.11 (calc. mass 4254 Da) and unmodified complementary RNA sequence ON2.15 (calc. mass 4555 Da).*

## 2.6 Duplex hybridisation studies – $T_m$

For effective antisense activity, oligonucleotide drugs are required to have high binding affinity to their complementary target sequence. When carbamates were first tested as a potential modification for antisense application, they provided good enzymatic resistance, but duplex stability was reduced relative to unmodified DNA. They were also described as unsuitable for antisense applications.[32] On average, carbamates were found to destabilise the duplex by approximately 2-5 °C per modification.[104,117] With the addition of LNA it is conceivable that this reduction in duplex stability can be mitigated. To evaluate this, we used duplex thermal melting analysis to assess the energy required to separate the duplex into its two constituent strands ($T_m$ data was obtained in collaboration with Sven Epple, DPhil).

LNA-CBM oligonucleotides (ON2.1-2.13) were added in equimolar ratios to either a complementary DNA (ON2.14) or RNA (ON2.15) to evaluate the LNA-CBM against both B-form DNA and A-form RNA targets. An important consideration is that most antisense applications involve DNA-RNA interactions, meaning that duplex stability against an RNA target could be considered as a more indicative comparison to *in vivo* efficacy.

### *2.6.1 Comparing $CBM_1$ and $CBM_2$ backbones*

As expected, single and triple incorporation of $CBM_1$ or $CBM_2$ modifications (ON2.2, ON2.8 and ON2.6, ON2.12) caused a reduction in duplex stability against DNA and RNA targets compared to the unmodified sequence (ON2.1, Table 2. 2, Table 2. 3 and Figure 2. 12). $CBM_1$ linkages were found to be less destabilising than $CBM_2$ in both duplex forms, confirming the previous results shown by Waldner *et. al*. [32] When hybridised to a DNA or RNA target, addition of $CBM_1$ modification resulted in a -2.7 °C/mod or -6.4 °C/mod reduction in stability. $CBM_2$ destabilisation was much larger, measured at -6.6 °C/mod or -8.6 °C/mod in DNA and RNA duplexes respectively. For both linkage types, CBMs were less destabilising against DNA targets, but significant reduction in affinity means that they are unsuitable as an antisense therapeutic backbone. In all cases, triple incorporation of each modification, resulted in a magnification of change in hybridisation strength compared to the single modification. For this reason, further analysis of each LNA-CBM, will be referred to by oligonucleotides with triple incorporations (ON2.8-2.13) and results for single incorporations (ON2.2-ON2.7) can be found in the supplementary (Figure 7. 1 and Figure 7. 2).

*Table 2. 2 Melting temperatures ($T_m$) from single incorporation of backbone modified LNA at a central position. x = $CBM_1$, y = $CBM_2$, $T^L$ = LNA thymidine. Values were obtained from the maxima of $dA_{260}$ vs. t using 3 mM of each oligonucleotide in a 10mM phosphate, 200 mM NaCl buffer, pH 7.0. $\Delta T_m$ is relative to the unmodified control ON2.1. Addition of LNA on the 5' site caused large reduction in $T_m$ against both targets whereas 3' addition to $CBM_1$ linkages gave the most stable modified DNA duplex and 5'/3' was the most stable against RNA.*

| | | DNA target | | RNA target | |
|---|---|---|---|---|---|
| ON | ON sequence (5'–3') | $T_m$ | $\Delta T_m$ | $T_m$ | $\Delta T_m$ |
| ON2.1 | GCTTGCTTCGTTCC | 60.2 | -- | 63.6 | -- |
| ON2.2 | GCTTGCTxTCGTTCC | 57.5 | -2.7 | 57.2 | -6.4 |
| ON2.3 | GCTTGCT$^L$xTCGTTCC | 50.5 | -9.7 | 56.1 | -7.5 |
| ON2.4 | GCTTGCTxT$^L$CGTTCC | 60.3 | 0.1 | 61.3 | -2.3 |
| ON2.5 | GCTTGCT$^L$xT$^L$CGTTCC | 53.4 | -6.8 | 62.1 | -1.5 |
| ON2.6 | GCTTGCTyTCGTTCC | 53.6 | -6.6 | 55 | -8.6 |
| ON2.7 | GCTTGCTyT$^L$CGTTCC | 51.7 | -8.5 | 57.1 | -6.5 |

*Table 2. 3 Melting temperatures ($T_m$) from triple incorporation of backbone modified LNA at central and terminal positions. x = $CBM_1$, y = $CBM_2$, $T^L$ = LNA thymidine. Values were obtained from the maxima $dA_{260}$ vs. t using 3 mM of each oligonucleotide in 10 mM phosphate, 200 mM NaCl buffer, pH 7.0. $\Delta T_m$ is relative to the unmodified control ON2.1. Incorporation of multiple CBM-LNA modifications was approximately additive compared to single modifications.*

| | | DNA target | | RNA target | |
|---|---|---|---|---|---|
| ON | ON sequence (5'–3') | $T_m$ | $\Delta T_m$ | $T_m$ | $\Delta T_m$ |
| ON2.1 | GCTTGCTTCGTTCC | 60.2 | -- | 63.6 | -- |
| ON2.8 | GCTxTGCTxTCGTxTCC | 52.1 | -8.1 | 44.8 | -18.8 |
| ON2.9 | GCT$^L$xTGCT$^L$xTCGT$^L$xTCC | <30 | <-30.2 | 40.6 | -23 |
| ON2.10 | GCTxT$^L$GCTxT$^L$CGTxT$^L$CC | 60.1 | -0.1 | 59.6 | -4.0 |
| ON2.11 | GCT$^L$xT$^L$GCT$^L$xT$^L$CGT$^L$xT$^L$CC | 41.2 | -19.0 | 61.6 | -2.0 |
| ON2.12 | GCTyTGCTyTCGTyTCC | 37.9 | -22.3 | 36.5 | -27.1 |
| ON2.13 | GCTyT$^L$GCTyT$^L$CGTyT$^L$CC | 34.7 | -25.5 | 46.8 | -16.8 |

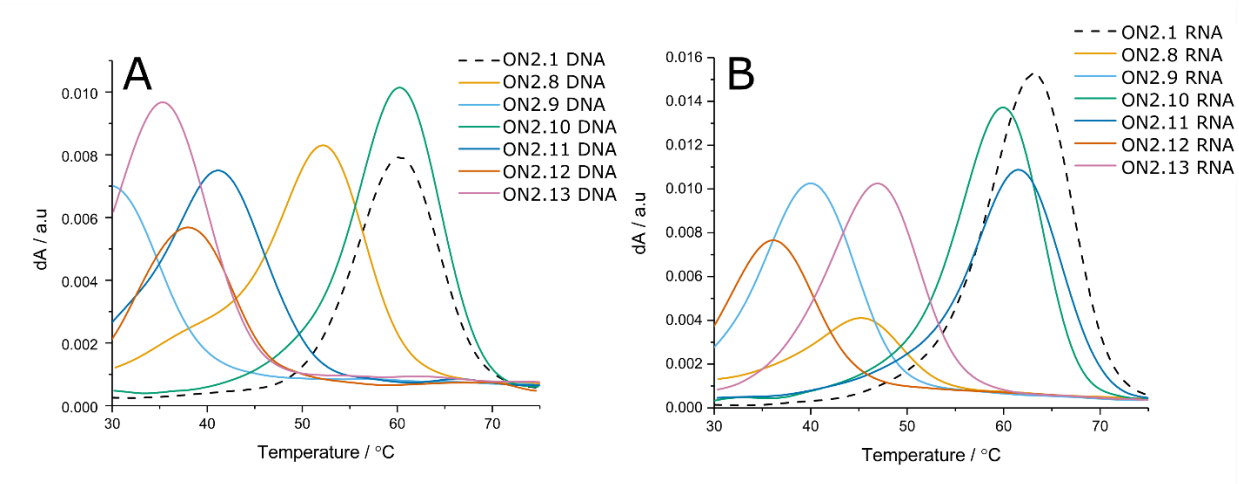



*Figure 2. 12 UV-melting data, 3 µM of oligonucleotides were added to 10 mM phosphate, 200 mM NaCl buffer, pH 7.0 and heated from 20–85 °C. $T_m$ values were determined from an average of six ramps from a smoothed plot of dA/t. (A) First derivative $T_m$ curves against DNA targets containing triple modifications of CBM-LNA, (B) First derivative $T_m$ curves against RNA targets containing triple modifications of CBM-LNA.*

### *2.6.2 LNA-CBM$_1$ modifications against DNA and RNA targets*

Locked nucleic acids are known to be stabilising modifications in unmodified duplexes. LNA was added to either the 5'-side (ON2.9), 3'-side (ON2.10) or to both sides (ON2.11) of the CBM$_1$ linkages. Against DNA targets, 5'-LNA addition (ON2.9) was found to be destabilising when compared to the unmodified (ON2.1) or CBM$_1$ (ON2.8) oligonucleotides. However, when the LNA is moved to the 3'-side (ON2.10), the modification become stabilising, restoring target binding values to those equal to that of the unmodified sequence (ON2.1). When LNA is added to both sides of the CBM$_1$ linkage, it combines the two effects of 5'-LNA and 3'-LNA to create moderate destabilisation (ON2.11).

Against RNA targets we find a difference in affinity when varying LNA position. Addition of 5'-LNA (ON2.9) again caused high levels of duplex disruption compare to unmodified (ON2.1) and CBM$_1$ (ON2.8) oligonucleotides. 3'-LNA (ON2.10) was able to mitigate duplex disruption to a significant extent but not as much as was found with DNA targets. However, when LNA is added to both sides of the CBM$_1$ linkage (ON2.11), we find that it produces stable duplexes with $T_m$ values only slightly decreased compared to the unmodified sequence (ON2.1, - 0.67 °C/mod).

Against both DNA and RNA targets it was found that addition of 5'-LNA relative to the CBM$_1$ backbone is highly destabilising. A suggestion for this is due to the conformation of LNA sugars compared to that of the DNA strand it is incorporated within. In DNA, sugar conformation is variable but tends to the 2'-*endo* conformation. The 2'-carbon is pushed out of the plane of the ring, into an envelope structure which reduces ring strain to lower conformational energy of the system. LNA being a 2'-4'-bicyclic ring system, means it cannot access this conformation and is instead found in a 3'-*endo* conformation. Because of this, the 3'-carbon pulls the backbone below it, putting strain on the linkage. In the case of phosphodiester linkages, the duplex can mitigate this change with its sp$^3$ hybridised rotational flexibility, but for the more constrained CBM$_1$ this is not available, potentially disrupting the helix structure.

In general, we find that the closer a backbone analogue is to a pure phosphodiester the greater the positive influence by a 5'-LNA above the modified backbone. Moving an LNA to the 3'-side means that the 3'-*endo* conformation is no longer acting on the modified backbone but instead on a phosphodiester linkage below it. As a result, we now see the expected relative stabilisation compared to the CBM$_1$ alone.

The biggest difference between DNA and RNA targets was seen with the introduction of LNA to both sides of the $CBM_1$ linkage. This again can be explained by the 3'-*endo* conformation of the LNA and its high affinity for RNA. In DNA:RNA duplexes, sugar conformation is pushed into the 3'-*endo* conformation. As the amount of LNA increases, the affinity of the DNA strand for the RNA target becomes greater. Although addition of a 5'-LNA relative to the backbone can impose strain on the $CBM_1$ linkage, this is compensated for by the addition of multiple LNAs, creating a stable duplex structure. Since LNA does not have increased affinity for 2'-*endo* DNA targets, the introduction of multiple LNAs does not have the same additive effect as seen against RNA targets. In summary for $CBM_1$ linkages we find that against DNA targets $TxT^L > TxT > T^LxT^L > T^LxT$ and against RNA targets $T^LxT^L > TxT^L > TxT > T^LxT$.

### *2.6.3 LNA-$CBM_2$ modifications*

Noting that addition of 3'-LNA in $CBM_1$ oligonucleotides restored duplex stability, we also synthesised a $CBM_2$ oligonucleotide with DNA sugars and an adjacent LNA at the 3'-side. $CBM_2$ modifications were found to be highly destabilising against DNA and RNA targets (ON2.6, 2.12). Opposite to the $CBM_1$, we saw that against DNA duplexes, $CBM_2$-3'-LNA further reduced hybridisation compared to the $CBM_2$ backbone (ON2.12 *vs* ON2.13). However, when hybridised to RNA, the LNA was still able to increase stability. These contrasting trends between $CBM_1$ and $CBM_2$ modified oligonucleotides, highlights possible structural and conformational differences between the two linkages. Movement of the nitrogen closer to the 5'-end of the linkage has a significant effect on duplex stability, suggesting that this region of the linkage requires increased flexibility. In summary for $CBM_2$ linkages we find that against DNA targets $TyT > TyT^L$ whereas against RNA targets $TyT^L > TyT$.

## 2.7 Circular dichroism

Circular dichroism (CD) can be used to analyse the global structure of oligonucleotides by their absorption of circular polarised light at 200-340 nm. For triple (Figure 2. 13) and single (Figure 7. 3) incorporation of CBM-LNA modifications, it was found that there were little changes in the global structure of the duplexes, and that helical structure largely remains unperturbed. DNA:DNA duplexes were found to adopt a B-form helix, determined by transitions at 220, 255 and 280 nm (Figure 2. 13A). DNA:RNA hybrids were found to adopt an A-form duplex, characterised by peaks at 210 and 270 nm (Figure 2. 13B).

Addition of CBMs (ON2.8 and ON2.12) created small changes in peak intensity compared to the unmodified duplex (ON2.1). Increasing the proportion of LNA in DNA:DNA duplexes, resulted in a hypsochromic shift towards 270 nm, whereas in DNA:RNA duplexes this was less profound. The suggestion for this is that addition of LNA causes a B→A transition to occur. As the proportion of LNA is increased, the 3'-*endo* character influences the 2'-*endo* DNA towards an A-form duplex (ON2.11).[41] However, against RNA targets this hysochromic shift is not observed because the DNA:RNA hybrid is already in the A-form. As a result, RNA transitions deviate from the unmodified control to a lesser extent compared to DNA (ON2.11 DNA, ON2.11 RNA *vs* ON2.1).

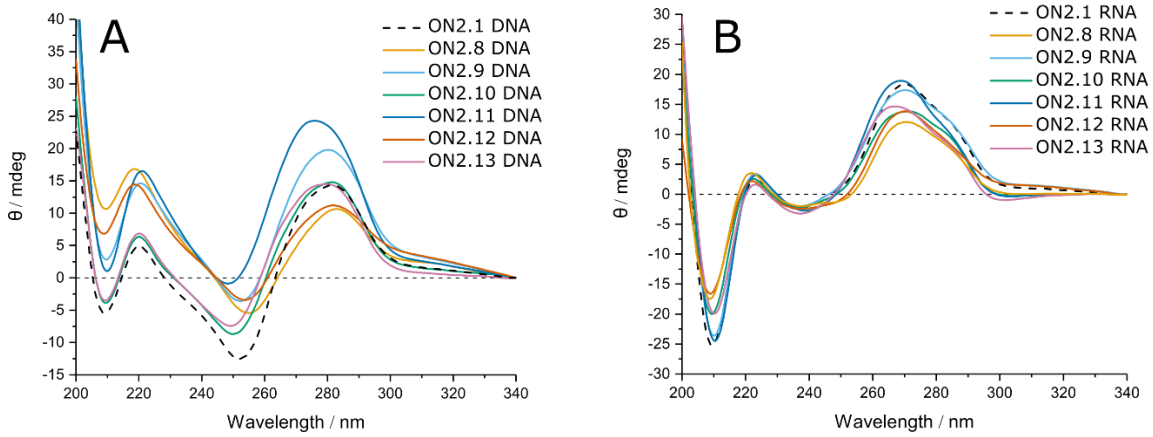



*Figure 2. 13 Circular dichroism spectra of 3 µM oligonucleotides in 10 mM phosphate, 200 mM NaCl buffer, pH 7.0. Spectra were averaged from four scans and smoothed to 20 points using a third order polynomial. (A) Triple incorporation of CBM-LNA modifications against DNA targets, (B) Triple incorporation of CBM-LNA modifications against RNA targets. Overall duplex form against both targets remained largely unchanged. Against DNA targets hypsochromic shifts were observed with addition of LNA.*

## 2.8 Enzymatic stability assays

If $CBM_1$-LNA modified oligonucleotides are to have significant efficacy *in vivo*, then they must possess good resistance towards enzymatic degradation. Phosphodiesterase 1 from the Eastern diamondback rattlesnake *C. adamnteus,* is an exonuclease that can cleave phosphodiester bonds. Phosphodiesterase 1 binds to the terminal region of the oligonucleotide and degrades in the 3' → 5' direction. Triply modified $CBM_1$-LNA oligonucleotides (ON2.8, ON2.10 and ON2.11) were incubated with phosphodiesterase 1 at 37 °C for up to 1 h. Samples were removed at set time points and frozen before analyses by 20% PAGE (Figure 2. 14, A, B). As a control, LNAs were also incorporated into oligonucleotides without $CBM_1$ (ON2.16-2.17).

Unmodified oligonucleotide (ON2.1) was degraded within 5 mins of addition of the nuclease, shown by the disappearance of any bands by UV shadowing (Figure 2. 14A, lane 1 *vs* lane 2). Addition of 5'/3'-LNA showed significant resistance to degradation and bands were still visible up to 30 mins (Figure 2. 14A, ON2.17). The bands show a reduction in intensity overtime as the oligonucleotide is degraded, but also appear to have migrated faster than the full-length starting sample (Figure 2. 14A, lane 10 *vs* 11). This suggests that the exonuclease rapidly degrades the terminal region of the sequence up until the LNA modification. Degradation is then slowed by the LNA. Addition of $CBM_1$ modifications had little effect on preventing degradation (ON2.8). However, when combined with LNA, the $CBM_1$-LNA modified oligonucleotide provided more stability than LNAs alone (Figure 2. 14, A ON2.17 *vs* B, ON2.11). For LNA-$CBM_1$-LNA modified oligonucleotides, fewer degradation products were observed, but after 5 minutes a new band with slower migration is formed (Figure 2. 14B, lane 25 *vs* 26). Gel electrophoresis migration is based on a mass to charge ratio of the DNA. As the terminal phosphodiesters are degraded the ratio of charge neutral CBMs compared to phosphodiesters increases. This means that it has a lower mass to charge ratio and migrates slower through the gel. This new band suggests that the terminal region of the oligonucleotide is degraded by the nuclease, but reaching a LNA-$CBM_1$-LNA modification, the central region of the oligonucleotide is protected, since the enzyme cannot cleave the carbonyl linkage.

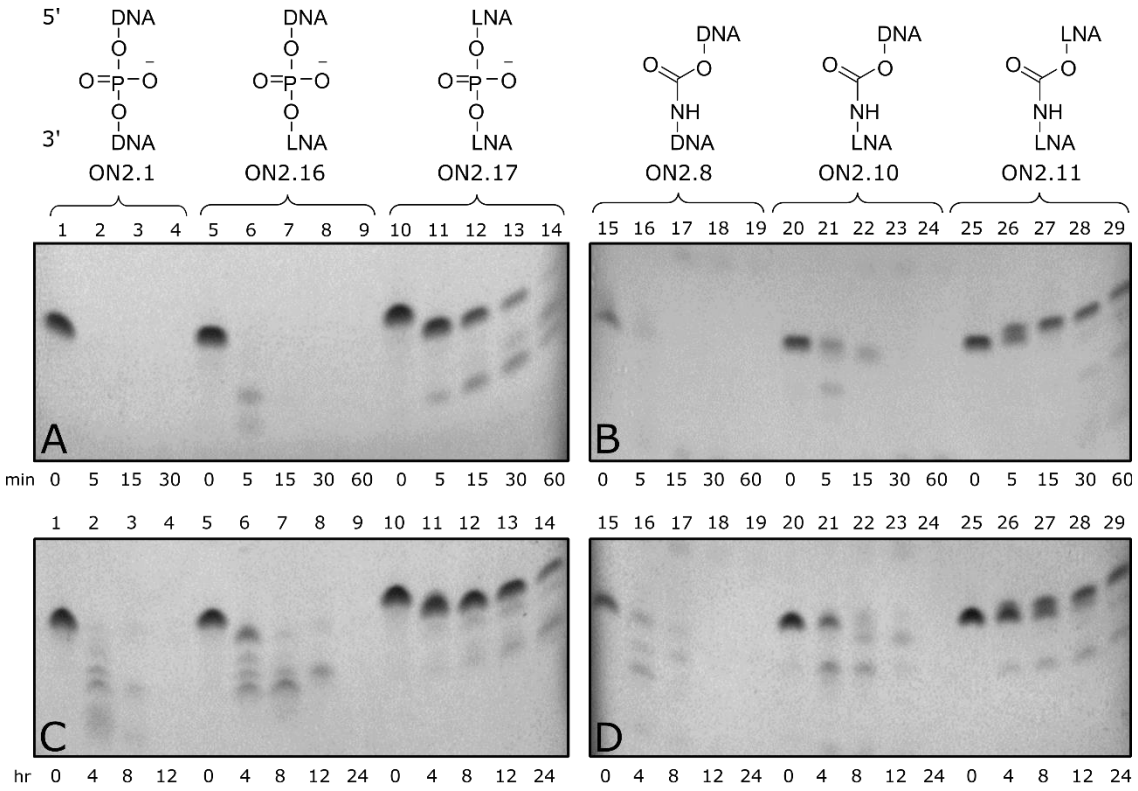



*Figure 2. 14 (Top) 20% PAGE results from phosphodiesterase 1 assay. Samples were incubated in 50 mM Tris-HCl, 10 mM $MgCl_2$ buffer pH 9.0 at 37 °C. Aliquots were removed at t = 0, 5, 15, 30 and 60 mins and frozen before analysis. (A) Gel showing unmodified and LNA controls, (B) Gel showing oligonucleotides with carbamate and combined LNA-carbamate modification. ON2.16 GCTᴸGCTTᴸCGTTᴸCC, ON2.17 GCTᴸTᴸGCTᴸTᴸCGTᴸTᴸCC, ᴸ = LNA. Results show increased nuclease resistance when introducing three separate 5'/3'-LNA-$CBM_1$ modifications (ON2.11). (Bottom) 20% PAGE results from FBS assay of modified oligonucleotides containing triple incorporation of $CBM_1$-LNA dinucleotides. Samples were incubated in Dulbecco's PBS buffer at 37 °C. Aliquots were removed at t = 0, 4, 8, 12 and 24 h and frozen before analysis. (C) Gel showing unmodified and LNA controls, (D) Gel showing oligonucleotides with carbamate and combined LNA-carbamate modification. Combined 5'/3'-LNA and $CBM_1$ modifications (ON2.11) showed increased serum resistance compared to unmodified control.*

LNA-$CBM_1$ modifications were also incubated in foetal bovine serum (FBS) to further examine enzymatic stability. FBS consists of a range of both exo and endonucleases. Modified oligonucleotides (ON2.8, ON2.10-2.11) were incubated in FBS at 37 °C for up to 24 h. Samples were removed at set time points and frozen before PAGE analysis (Figure 2. 14C and D).

Like when incubated with phosphodiesterase 1, unmodified and $CBM_1$ modified oligonucleotides were the fastest to be degraded and were cleaved within 4-8 h (Figure 2. 14, C-ON2.1, D-ON2.8). 5'/3'-LNA (ON2.17) provided good protection but the highest resistance was given by 5'/3'-LNA-$CBM_1$ (ON2.11). Previously, it was reported that CBM modifications stabilise against enzymatic degradation and can increase half-life by up to four times.[32] However, in our experiments we found that $CBM_1$ modifications offer little protection against enzymatic degradation. The reason for this is due to the size and footprint of the nucleases.[118] Although the CBM cannot be cleaved by the enzyme, when it enters the active site the

nuclease is able to skip over the isolated modified backbone to reach the adjacent phosphodiesters. When a single LNA is added (ON2.10), the speed of degradation is slowed but not stopped. To be able to prevent translocation of the enzyme beyond the modified backbone, the bulky 5'/3'-LNA modification is needed. When this is combined with the $CBM_1$ linkage, it prevents cleavage of the backbone in the active site and stops skipping to the adjacent phosphodiester. Therefore the 5'/3'-LNA-$CBM_1$ modification provides the best enzymatic resistance. In addition to this, enzymatic resistance may also be inferred by rejection of the modified oligonucleotide by the enzyme. Within the active site many nucleases use metal ions and catalytic amino acid residues to cleave the phosphate backbone. Due to the size of the 5'/3'-LNA-$CBM_1$ modification it maybe that these ions are displaced within the active site upon binding to the modified sequence. Unable to cleave the CBM backbone the oligonucleotide is eventually released.[118]

LNAs without CBM modification provides adequate protection against enzyme degradation which questions the need for CBMs combined with LNA. Although a versatile modification, LNAs do have downsides which could be mitigated with the use of the CBM modification. LNAs have been found to have significant hepatoxicity[85] in animal models and this property is probably shared by other high binding affinity oligonucleotides rather than LNAs specifically.[43] This means that LNA toxicity could be reduced by tuning LNA binding affinity with CBMs to bring it into a therapeutic window.

*In vivo* oligonucleotides are constantly degraded and recycled as mononucleotides into DNA sequences. If LNAs interspaced with phosphodiesters are used as therapeutic drugs, then there is a risk of degraded LNAs being recycled into the monophosphate-triphosphate (nucleotide triphosphates, NTPs) pool. However, the regions around the CBM linkages would be degraded to dinucleotides, which would not be converted to NTPs and cannot be incorporated into RNA or genomic DNA. This reduces incorporation of modified nucleotides into the genome.

Like with many negatively charged oligonucleotides, LNA modified therapeutics have poor cellular uptake as they do not efficiently pass through the cellular lipid bilayer. By replacement of phosphodiesters with charge neutral backbones, it has been shown that oligonucleotides can have better gymnotic uptake into cells.[119] Therefore, by creating LNA-CBM

oligonucleotides opposed to just LNA oligonucleotides, it might be possible to create an improved therapeutic construct with potentially better cellular uptake.

## 2.9 Summary of LNA-CBM modified oligonucleotides

Therapeutic oligonucleotides with LNAs combined with modified backbones are relatively new. In this chapter, we outlined the synthesis of several LNA-CBM dimer phosphoramidites that were easy to synthesise and stable to oligonucleotide synthesis, deprotection and purification. Binding affinity was analysed using UV thermal studies, where it was confirmed that the $CBM_1$ backbone is more stable than the $CBM_2$. Addition of LNA was found to be both stabilising and destabilising depending on its position, relative to the modified backbone. Accounting for LNA positioning, the most stable $CBM_1$ duplex against RNA targets contained 5'/3'-LNA-$CBM_1$ modifications and was able to restore binding affinity to a similar level as the unmodified control. We also suggest that the closer a backbone analogue is to the natural phosphodiester, the better its thermodynamic performance will be.

From CD analysis we have shown that LNA-CBM modifications have minimal effect on the global structure of the duplex. When bound to an RNA target the structures remained like the unmodified control and against DNA targets LNA was shown to influence a B→A transition. Enzymatically 5'/3'-LNA-$CBM_1$ provided the best resistance against degradation. $CBM_1$ backbones in isolation provided minor protection and LNAs contribute a significant portion of resistance.

Before this study, CBM backbones were unsuitable as templates for therapeutic applications because of their poor duplex stability. With the addition of LNA they have improved biophysical properties and possess the required stability for potential antisense applications. Compared to the amide they are easy to synthesise and the CBM linkage has no chiral centres. They have good linkage flexibility and being charge-neutral might have a potential advantage in certain applications compared to LNA alone. Future analysis of LNA-CBMs should focus on evaluating cell uptake as well as assessment of toxicology before moving to efficacy in cell culture and animal models.

# 3 XRD structures of LNA-Amide modified DNA-RNA Duplexes

## 3.1 Introduction to amides as modified backbones

### *3.1.1 Amide backbones and synthesis*

The amide backbone is a four atom, charge neutral linkage, based on carbonyl functionality. Unlike the carbamate linkage, a hydroxyl is exchanged for a $CH_2$ group with tetrahedral geometry and high rotational flexibility. This is important because unlike the carbamate, which typically is a destabilising modification, the amide can increase duplex hybridisation by as much as +3 ˚C per modification.[120] This makes amides an interesting candidate to be used as a modified backbone for biological and diagnostic applications.

Amide backbones were originally developed from a precursor modification called peptide nucleic acid (PNA) (Figure 3. 1A). PNA is a modification which removes the (deoxy)ribose in exchange for a tertiary amine which links the peptide backbone to the nucleobase. Although PNA has interesting properties due to its charge neutrality and nuclease stability, being structurally different from canonical oligonucleotides means that formation of DNA:PNA gapmers requires specially designed linkers or adaptations.[121] To retain the positive features of PNA without compromising the duplex structure, I. Idkiak *et al.*[122] (Figure 3. 1B, R = Me) developed an amide backbone spaced between (deoxy)ribose sugars. Although the effects upon duplex properties were modest, further adaptations by A. Desmaeker *et al.*[33] (Figure 3. 1B, R = H, and Figure 3. 1C) and Erik Rozners *et al.*[97] (Figure 3. 1D) showed that amides could provide increased thermodynamic and enzymatic stability.[98]

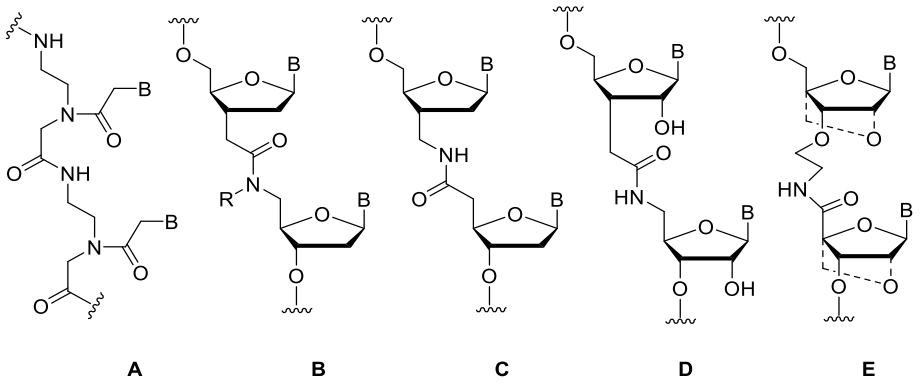



*Figure 3. 1 Published structures of PNA (A) and amide backbones (B-E), standard linkages are when the carbonyl rather than the nitrogen is closer to the 5'-sugar, (5'-C(O)N(H)-3') , reversed is defined when the amide linkage is in the opposite direction (5 '-N(H)C(O)-3'), A) PNA backbone, B) Standard four-atom linkage with DNA, R = H[33] or R = Me[122], C) Reversed four-atom linkage with DNA[33], D) Standard four-atom linkage with RNA[97], E) Reversed five-atom linkage with LNA sugars.*[123]

The structure of the amide backbone is an important factor that partly dictates its biological compatibility. In order to be accepted and processed by enzymes such as polymerases, amides need to be good bio isosteres of a phosphodiester. Work by Erik Rozners *et al.* has shown using X-ray crystallography and NMR, that amides are well tolerated in a duplex and are good structural mimics of phosphodiesters.[77] This has also been shown by several other structural studies using molecular modelling[33], X-ray crystallography[124] and NMR.[125] When applied in biological mechanisms such as RNAi, amides were able to induce a gene silencing response, even when placed within the central regions of guide strand RNA.[126] Alongside this, amides can also be used in RNase H mechanisms. Although RNase H deactivation is sensitive to modifications within the central region of the DNA strand, it has been shown that if amides are added to the wings of a DNA gapmer, then RNase H response can be maintained with increased therapeutic effect.[127] Using next generation sequencing to analyse the fidelity of different backbones as transcriptional templates, it was shown that the amide backbone outperformed other modifications such as the triazole, producing significantly fewer addition, deletion or substitution mutations in transcribed RNA products.[128]

LNA was first introduced alongside the amide backbone by A. Lauritsen and J. Wengel.[123] In this study they took the amide backbone and designed an extended 5-atom linkage combined with LNA (Figure 3. 1E). They found that LNA-amides showed enhanced thermal melting ($T_m$) properties, but like the carbamate-LNA previously discussed, they required the correct sequence design (DNA-Am-LNA). Despite finding a viable oligonucleotide modification, the work was not pursued further and the structural relationship between amide-LNAs and duplex stability remains unclear.

Although amides have good target affinity for RNA, they are difficult to synthesise. Unlike the carbamate linkage, the 3'-oxygen of the 5'-sugar needs to be converted to a $CH_2COOH$ group which is a synthetic challenge. A. De Mesmaeker *et. al.*[129] (Figure 3. 2) performed this by a radical reaction to install a 3'-alkene, followed by two oxidations to form the carboxylic acid. The route is effective but requires the use of several toxic reagents and difficult column chromatography. Modification of this procedure by V. Kotikam *et. al.*[130] involved a Wittig reaction with a carboxylic acid as a phosphonium ylide, followed by a set of reductions. Unlike the previous route, the reactions are cleaner, but the reductions can create a mixture of partly reduced compounds or promote loss of benzoyl protecting groups.

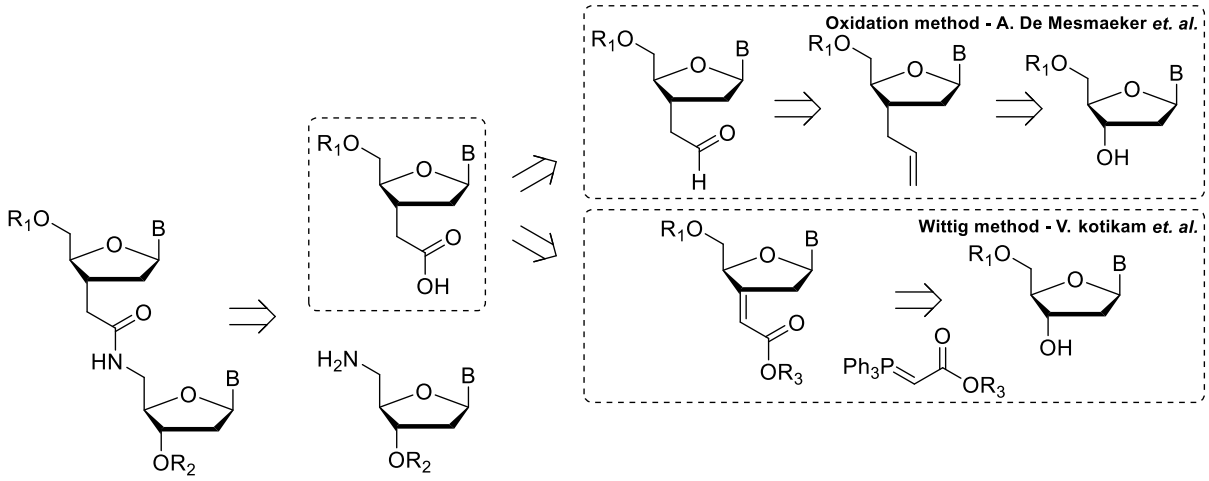



*Figure 3. 2 Retrosynthetic analysis of the amide backbone and carboxylic acid monomer ($R_1$ = DMTr, $R_2$ = $P(O(CH_2)_2CN)(N^iPr_2)$, $R_3$ = Bn). This has predominately been achieved by radical addition of an alkene to the 3'-position followed by two oxidations (Oxidation method), or by Wittig reaction to install an unsaturated alkyl chain with protected ester that can be deprotected and reduced to form the carboxylic acid (Wittig method). After synthesising the carboxylic acid, the monomers can be coupled to form the amide backbone by standard peptide coupling reagents.*

### *3.1.2 Amide backbone and duplex stability*

Chemical modifications can reduce or increase duplex thermal stability ($T_m$) and these effects are additive when including multiple incorporations. If a modification is highly destabilising, it therefore will create an unstable duplex and be unsuitable for therapeutic or *in vivo* applications.

Although structurally challenging to produce, LNA-amides have promising biophysical properties (Table 3. 1). Work by Dr Ysobel Baker (Postdoctoral Researcher) incorporated LNA-amides into a DNA oligonucleotide to assess the biophysical properties of the duplex $T_m$ against both DNA and RNA targets. The amide backbone was found to be moderately destabilising against DNA targets (-2.3 °C) and approximately like an unmodified sequence

against RNA targets (-0.3 °C). Addition of LNA is in general a stabilising modification except when placed at the 3'-side of the amide backbone. LNA-amides show good affinity for RNA targets and when LNA flanks the amide on both sides it was able to stabilise the duplex (+5.1 °C). Because of these duplex stabilising properties, LNA-amides are promising candidates for therapeutic or *in vivo* applications. Oligonucleotides can be highly modified with LNA-amides, without compromising binding affinity or target specificity.

*Table 3. 1 UV-thermal melting results obtained from oligonucleotide sequences with amide and LNA-amide modifications. $T_m$ values were measured using 3.0 µM concentrations of each strand in 10 mM phosphate buffer (pH 7.0) containing 200 mM NaCl. Am = amide linkage, $^L$ = LNA. $T_m$ values were calculated as the maximum of the first derivative of the melting curve (dA260 vs dT) and reported as the average of at least two independent experiments each with eight measurements. DNA target = GCTGCAAGCGTCG; RNA target = GCUGCAAGCGUCG. This work was performed within our group by Dr Ysobel Baker and is currently unpublished.*

| Sequence | $T_m$ ($\Delta T_m$) DNA<br>$\Delta T_m$ = modified - control | $T_m$ ($\Delta T_m$) RNA<br>$\Delta T_m$ = modified - control |
|---|---|---|
| CGACGC**$T_{Am}$T**GCAGC | 61.2 (-2.3) | 61.0 (-0.3) |
| CGACGC**$T^L_{Am}$T**GCAGC | 63.1 (-0.3) | 64.4 (+3.0) |
| CGACGC**$T_{Am}T^L$**GCAGC | 60.9 (-2.5) | 63.9 (-2.5) |
| CGACGC**$T^L_{Am}T^L$**GCAGC | 63.4 (-0.1) | 66.5 (+5.1) |
| CGACGCTTGCAGC | 63.5 | 61.4 |

## 3.2 Introduction to XRD macromolecular crystallography

### *3.2.1 Application of crystallography for oligonucleotide structures*

After Watson and Crick proposed the duplex structure of DNA aided by data from X-ray fibre diffraction, researchers looked for a method to confirm the detailed helical structure. Previously it had been shown that proteins and enzymes can be ordered into crystal structures.[131] This led to the solving of the lysozyme protein crystal structure using XRD in 1965 by C. C. F. Blake *et. al.* to a resolution of 6 Å, which would later be improved to 2 Å.[132,133] Although the principle had been proven for protein structures, the time gap between the Watson-Crick proposed structure and experimental confirmation using XRD would be considerable. In the early 1980s R. Dickerson *et. al.*[134] solved the crystal structure of a complete turn of B-form DNA which would be termed Dickerson's dodecamer.[10] Using XRD, it was shown that crystal structures of DNA could be obtained below 2 Å resolution. This confirmed the double helix structure of DNA as well as providing intricate details about Watson-Crick base pairing, backbone torsion angles and other helical parameters. Since then nucleic acid crystal structures have been deposited into the Protein Data bank (PDB) some to

resolutions less than 1.0 Å.[135] This makes crystallography a technique that can provide structural information of macromolecules up to atomic resolution.

### *3.2.2 Introduction to crystal growth*

Like all techniques, crystallography has its limitations and to produce an interpretable diffraction pattern, oligonucleotides must be in a crystalline form. Creating these crystals is often described as a "dark art" and it is not fully understood how crystals nucleate. The process of growing crystals requires screening to find a mixture of salts and buffers that provides an environment to nucleate a crystal. The probability of crystallisation can be enhanced by ensuring high oligonucleotide purity, but also by tuning environmental factors such as temperature.[136]

In order to produce a crystal, the concentration of oligonucleotide must be carefully controlled. If the concentration is too low, then no crystal will form, too high and the oligonucleotide will precipitate out as an amorphous material. This balance of conditions can be satisfied using sitting drop vapor diffusion, which allows for crystal growth by gradual dehydration across the breadth of the crystallisation phase diagram (Figure 3. 3). For this technique, two liquid phases are separated into a drop and reservoir solution in a closed environment. The drop solution (1-4 µL), contains the oligonucleotide along with a cocktail of salts and buffers. The reservoir (50 - 1000 µL), contains the same mixture at double the concentration of the drop and with no oligonucleotide. This creates a diffusion gradient between the drop and reservoir, drawing $H_2O$ out of the drop. Over time as the $H_2O$ is removed, the drop concentration increases to the point where it is high enough to nucleate a crystal. The crystal can then be removed from the drop for analysis.

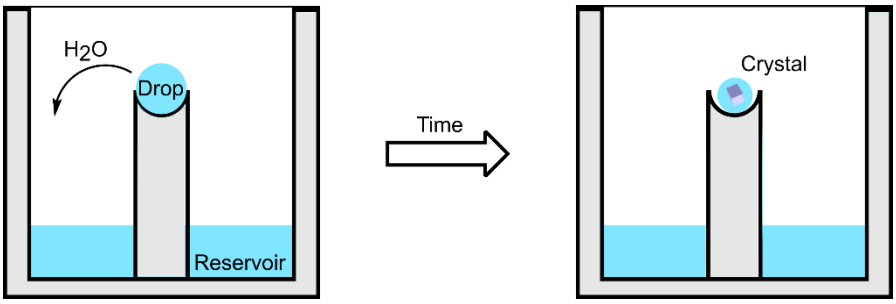



*Figure 3. 3 Sitting drop vapour diffusion – Crystallisation is afforded by diffusion of $H_2O$ from the drop solution into the reservoir. Inside the drop is contained the oligonucleotide and crystallisation buffer. The reservoir contains the same buffers without the oligonucleotide at double the concentration of the drop. This creates a diffusion gradient to concentrate the drop solution and nucleate crystals in a closed system.*

### *3.2.3 Data collection and processing – indexing, scaling and merging*

After obtaining oligonucleotide crystals, the next step is to collect the diffraction data that can be used to solve the crystal structure. The aim of X-ray crystallography is to build a three-dimensional electron density map from the diffraction pattern originating from the interaction between the electrons and incident X-rays. The electron density is a Fourier transform of the diffraction pattern, however there are various ways in which the data must be processed in order to derive a model structure.

Data collection at synchrotron radiation sources such as Diamond Light Source, is the primary method used to collect biological XRD data. The high-powered beam means that intense diffraction spots can be obtained with minimum exposure time, producing better diffraction results. Standard collection is performed by exposure of the crystal to the X-ray source and rotating the sample. Because the diffraction images are collected on a two-dimensional detector, rotating the crystal and taking an image every 0.1 ˚ ensures the best possible coverage to collect all Miller indices.

Processing the data into an electron density map is a multiple step process (Figure 3. 4). The first step is indexing and integration where the positions of each spot are marked and their intensity is assigned. This information can then be used to determine the lattice parameters and Bravais type. Due to changes in beam intensity and various other external factors, the integrations of each spot are not on a common scale. Because of this, spot intensities are normalised in a process called scaling. After the images have been scaled, they are then merged into a single data set which lists the set of unique diffractions based on their miller index values.

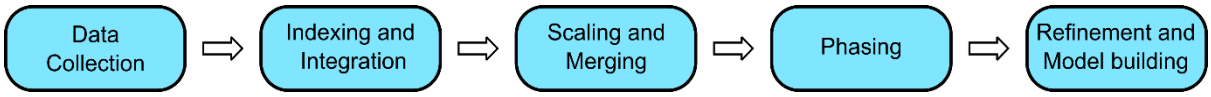



*Figure 3. 4 Overview of data processing required for determining an oligonucleotide structure using X-ray crystallography. After data collection diffraction images are analysed to find spot positions and intensity (indexing and integration). The images are then normalised and combined into a single data file (scaling and merging). After this the phase equation can be solved by using molecular replacement followed by model building and refinement.*

The procedure of data processing is largely based on satisfying statistical values, which has led to the development of automated software pipelines. Packages such as DIALS[137] and xia2 are commonly used to auto process data obtained at the synchrotron, and output files that can be directly fed into phase calculations.[138,139] The benefit of automated processing is that

it allows data to be processed without bias. However, auto processing struggles to produce the correct solution for complex data sets.

### *3.2.4 Phasing, model building and refinement*

After processing the data, it is possible to calculate the phase of the diffracted waves which will be used to generate the electron density map. Molecular replacement with programs such as phaser[58], is the most common method used to solve the phase problem for macromolecules.[140] To do this, phase information from a similar structure that has already been solved is applied to reflections of the observed diffractions. This then allows the phase to be calculated from the model and solve the electron density map for the new reflection data.[141] Molecular replacement is considered to be correct if the translational and rotational score (TFZ), is greater than eight, and if the log likelihood gain score is greater than 50. However, TFZ score is only a suggestion that a solution has been obtained and it could be that higher values maybe required for nucleic acid structures. As well as relying on statistical scores, it is also essential to look at the electron density map to confirm that it is chemically viable.

Model refinement is the process of bringing together the agreement between the model and the experimental data. Models are treated with successive rounds of refinement, spaced between editing in a graphical interface (COOT). To assess the quality of the model, statistical measures $R_{free}$ and $R_{work}$ are used. Before molecular replacement, a percentage (5-10%) of reflections are separated and remain unprocessed. By generating a set of theoretical diffraction patterns from the model, it is then possible to compare the relationship back to the original data. Typically, for a good model, $R_{free}$ should be below 0.30 and be no more than 5% higher than $R_{work}$. If $R_{work}$ is greater than $R_{free}$, it suggests that bias or over fitting of the model may have occurred.[142] However, structures with high disorder can have $R_{free} > 0.30$ and still be considered correct after inspection of the electron density map.

## 3.3 Aims and objectives of LNA-amide crystallography

Based on our own previous work on the carbamate backbone and work by Rozners *et. al.*[77] and Wengel *et. al.*[123]. It has been shown that the biophysical properties of a modified backbone or LNA-backbone are closely linked to its structural features. To date, much research into modified backbones has been through a trial and improvement method. Using

X-ray crystallography, the aim of the work reported in this chapter is to crystallise a set of modified LNA-amide backbones in order to assess their structural effects. From these structures, the amide backbone is analysed combined with LNA sugars to determine the origin of the physical properties.

We also aim to provide comparison between LNA-amide and LNA-carbamates that were previously discussed, to provide insight into why we see different biophysical properties, from two backbones that are chemically very similar. The wider objective for this work is to show that LNA-amides are an exceptionally good modifications as a phosphodiester mimic, and to show that structurally they are well tolerated within oligonucleotide duplexes.

To complete these aims, we will synthesise a set of oligonucleotides that contain different variations of LNA-amide modifications (Figure 3. 5). The modifications will be placed into the DNA strand of a DNA:RNA hybrid and various conditions will be screened to promote crystal growth. Using X-ray crystallography oligonucleotides will be analysed to provide the crystal structures of several modifications, all within identical base sequences, allowing for direct comparison. Analysis of the base pair parameters, hydrogen bonding, duplex form and backbone torsion angles will then be used to compare the structures.

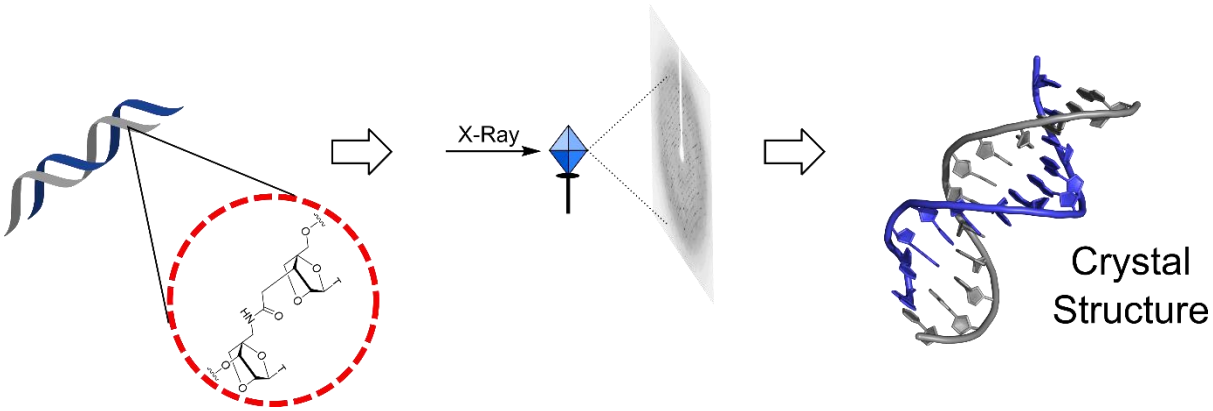



*Figure 3. 5 Overview of LNA-Amide crystal structure determination. Oligonucleotides will be modified with an amide or LNA-amide modification and crystallised using crystal screens. Resultant single crystals will then be subjected to X-ray crystallography to obtain a structure for analysis.*

## 3.4 LNA-amide oligonucleotide synthesis and initial screens

### *3.4.1 Oligonucleotide sequence selection*

Crystallisation can be difficult, especially for macromolecules such as proteins or oligonucleotides. For this reason, it is important to select a sequence that will provide the best opportunity to produce crystals. To find a promising sequence, the Protein Data Bank

was searched for an unmodified duplex that had already been crystallised and provided diffraction data below 2 Å resolution. From the initial hits, a sequence was chosen based on several key criteria; 1) the sequence must preferably have multiple thymidine steps, because the modified nucleosides have been synthesised using thymidine, 2) these thymidine positions should be remote from any crystal packing features so that any disruption caused by the artificial backbone has minimal impact on crystal formation and 3) the duplex should be a DNA:RNA hybrid. The reason for selecting a DNA:RNA duplex is due to LNAs affinity for RNA targets. The 3'-*endo* conformation of LNA matches that of RNA, which in turn means that LNAs adopt A-form duplex conformations. By targeting an RNA sequence, it will maximise the influence of the LNA. Alongside this, DNA:RNA hybrids also have biological significance in mechanisms such as RNase H RNA degradation (sequence selection performed in conjunction with Dr J. Hardwick, Dphil).

A 10 mer DNA:RNA hybrid (1PJO[143], Figure 3. 6), was selected as a model system based on the above criteria, and the coordinate file was analysed to find suitable sites for modification. Two potential modification sites were found towards the DNA 5'-end. However, in order to provide good structural insight into the effects of LNA-amides on the duplex, the modification should be placed into a central region of the helix. If the modified backbone is place too close to the terminal regions, small structural details maybe lost due to higher mobility or unwinding of the terminal base pairs. Based on this, it was decided that the best place for modification would be between the third and fourth base of the oligonucleotide, and that we would also investigate a position closer to the 5'-terminus.

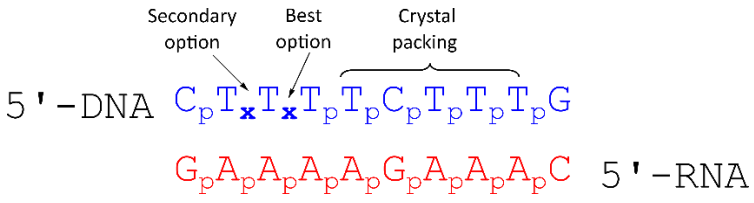



*Figure 3. 6 Sequence of DNA:RNA hybrid used for XRD studies. p = phosphodiesters, x = modified backbones. Two positions have been highlighted as good potential sites to add in the modification. This was determined by analysis of the electron density map from the unmodified published structure (1PJO). The DNA 3'-end contained potential crystal packing interactions. The best option was determined by proximity to the centre of the duplex and lack of crystal packing interactions.*

### *3.4.2 Synthesis of the LNA-amide precursor compounds and amide solid phase synthesis*

Amide coupling is a versatile synthetic reaction that is commonly used in peptide synthesis. For this reason, the amide was constructed on solid support during oligonucleotide synthesis. Synthesised by Dr Ysobel Baker, the amide building blocks are separated into carboxylic acids

**37-38** and amines **39-40** (Figure 3. 7). By choosing either the LNA or DNA version of each monomer, it is possible to synthesise a set of modified oligonucleotides varying in the position of LNA sugars relative to the amide backbone.

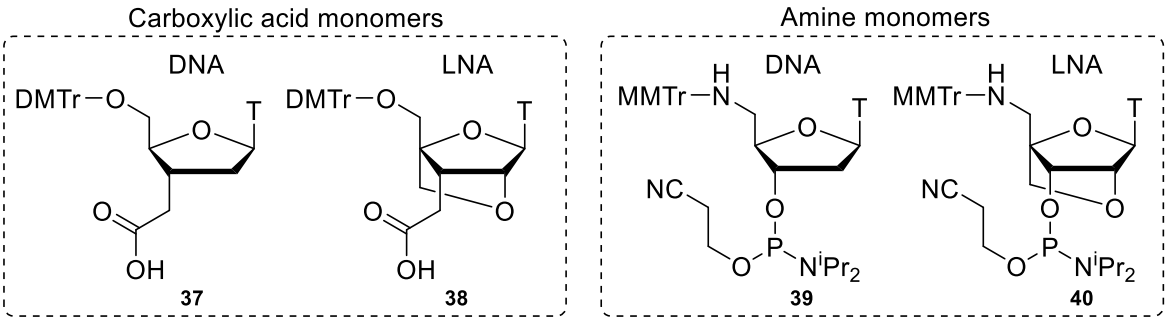



*Figure 3. 7 Modified monomers used to synthesise oligonucleotides with amide backbones. Monomers are separated into carboxylic acid and amine functionalities. Coupling between amine and carboxylic acids form the amide linkage, by choosing either DNA or LNA, linkages can be configured with various sugar modifications.*

Using the monomer units, LNA-amide modified DNA oligonucleotides were synthesised by Dr Ysobel Baker using a combination of standard solid phase synthesis and manual amide couplings (Figure 3. 8A). For the unmodified section, synthesis was performed using the standard phosphoramidite method, followed by addition of a 5'-*O*-MMTr protected amine phosphoramidite **39** or **40** in the final cycle to prepare for the amide backbone. The oligonucleotide is detritylated and then removed from the synthesiser to perform a manual peptide coupling, using the carboxylic acid monomer **37** or **38**. After forming the amide linkage, the resin bound oligonucleotide is returned to the synthesiser to continue standard solid phase synthesis. Using this procedure, a set of oligonucleotides containing LNA-amides were synthesised (Figure 3. 8B, C). For further comparison, LNA-carbamate$_1$ (LNA-CBM$_1$) dimers **16**, **18** (Figure 3. 8B, C) were also synthesised and put into equivalent sequences. If structural data could be obtained for both amides and carbamates, it should be possible to explain substantial differences in duplex stability, arising from minor changes between these chemical modifications.

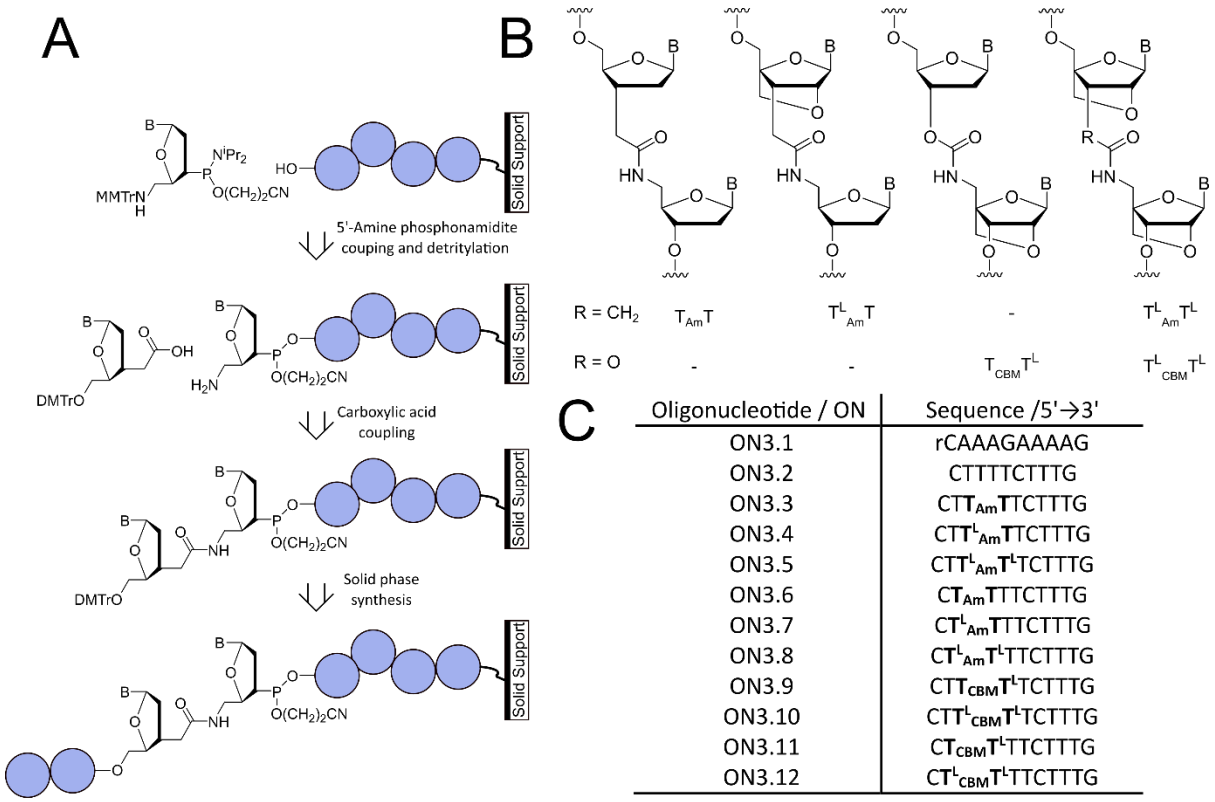



*Figure 3. 8 A) Synthetic scheme detailing the synthesis of amide modified oligonucleotides on solid support. Unmodified phosphoramidites (blue circles) are added using standard automated synthesis. The amide linkages are then coupled using manual couplings. The modified oligonucleotide can then be returned to the synthesiser for continued synthesis. B) LNA-modified backbone dimers used in sequences. $^{L}$ = LNA, Am = amide, CBM = carbamate. C) Sequences of modified oligonucleotides containing one addition of an LNA-backbone modification r = RNA sequence.*

### *3.4.3 Crystallisation of LNA-backbone modified oligonucleotides using high throughput screens*

The first stage of crystal growth is to screen the DNA:RNA hybrid in a wide range of conditions in order to establish a potential condition for crystallisation. DNA oligonucleotides (ON3.3 – ON3.12) were each combined with the complementary RNA sequence (ON3.1) and subjected to a variety of crystallisation conditions. The initial screening process is performed using a high throughput method testing up to 288 conditions per plate.

From the initial screens, it was found that amides (ON3.3-ON3.4) and carbamates (ON3.9, ON3.11-ON3.12) produced crystals in a variety of different conditions. Space group and symmetry are determined from the reciprocal lattice geometry produced from the X-ray diffraction pattern. However, it is possible to speculate on differences in crystal forms based on their physical appearance. Of all the crystals produced in the screens two morphologies were most abundant. Carbamates were found to produce hexagonal or cubic shaped crystals that were very small with defined edges (2-10 µm, Figure 3. 9A-D). Unlike these, amides

formed long needle structures (50-80 µm, Figure 3. 9E) often with many domains or intersections (Figure 3. 9F). From this it could suggest different lattice forms between amide and carbamate crystals.

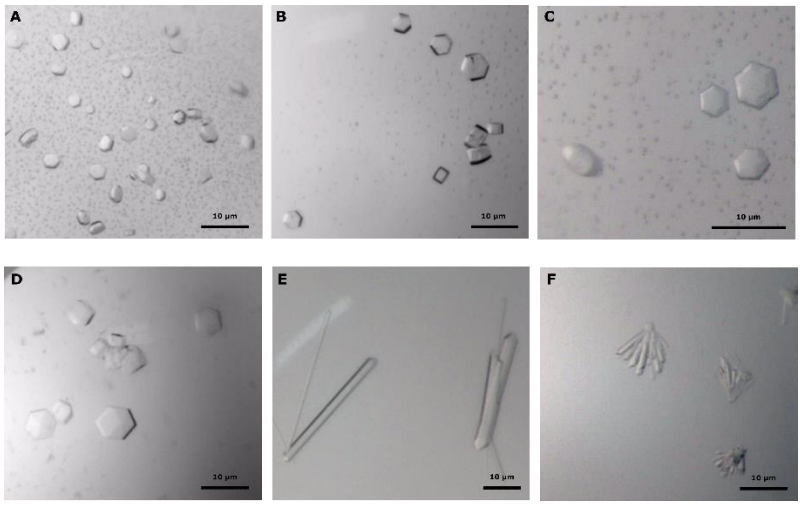



*Figure 3. 9 Summary of crystal morphologies obtained from Natrix HT screen (Hampton Research). CBM oligonucleotides mainly produced small hexagonal or cubic crystals (A-D). Amides regularly were found in needle structures (E) or as multiple domains (F). Ideally, crystals should have defined edges and single domains (A-E) a poor crystal is loosely defined to have multiple domains, but all observations are non-quantitative and true crystal quality is determined by the quality of XRD. A-B) ON3.9, C) ON3.10, D) ON3.12, E) ON3.4, F) ON3.3 (for full conditions see appendix Table 7. 2).*

### *3.4.4 High throughput screen data collection*

Crystals from modified duplexes (ON3.3-3.4, ON3.9, ON3.11-3.12) were sent to the German Electron Synchrotron and were found to diffract between 2.6 – 4.0 Å, significantly lower than expected. From the original unmodified structure by M. L. Kopka *et. al.*[143], resolutions of 1.10 Å had be reported and routinely resolutions below 1.20 Å have also been achieved.[144] Comparing the conditions used for crystallisation in these papers, versus the conditions from the high throughput screen highlighted slight differences between conditions and might explain why we did not achieve similar resolution. It was also mentioned that RNA purity was one of the key factors in producing high diffracting crystals, quoting the need for gel purification as well as HPLC. In this work oligonucleotides were purified by HPLC and the chromatograms were analysed to assess their purity (Figure 3. 10). In general RNA synthesis produces more by products reducing synthesis yield and purity. However short 10 nt sequences can be easily purified by HPLC. Analysing the purified chromatograms of RNA, DNA and modified DNA sequences, it is shown that oligonucleotides obtained from HPLC produce single sharp peaks. Minimal by-product oligonucleotide sequences remained and no contaminates were visualised in the mass spectrometry analysis. Therefore, it was

determined that gel purification is unnecessary to remove by-product oligonucleotide sequences.

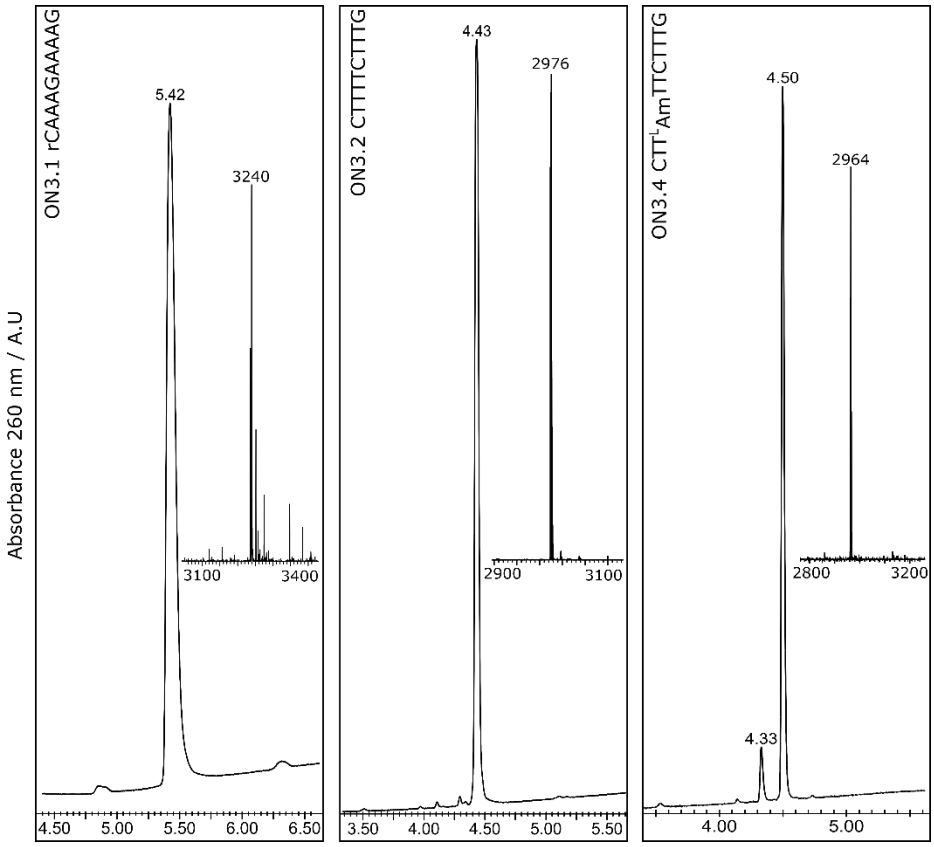



*Figure 3. 10 UPLC-MS results from purification of oligonucleotide sequences used for XRD screens and structure determination. In general samples produced single sharp peaks indicative of high purity. Calc. mass, ON3.1 = 3238 Da, ON3.2 = 2975 Da, ON3.4 = 2965 Da.*

Oligonucleotide concentration is determined by the absorbance at 260 nm using the Beer-Lambert law and determination of the extinction coefficient. An essential part of obtaining good diffracting crystals is ensuring high sample purity. In addition, there should be equal amounts of both oligonucleotide strands based on UV concentration measurements. However, if errors in pipetting or calculation resulted in an excess of one strand, then it would be unable to form a duplex, acting as a contaminant that would not be visible by HPLC analysis. To ensure a 1:1 ratio of DNA:RNA, unmodified oligonucleotide duplex (ON3.1:ON3.2) was purified by non-denaturing PAGE (Figure 7. 4). The duplex was then extracted from the gel removing excess single stranded oligonucleotide ready for future screening experiments.

Alternatively, lower diffraction resolutions between 2.6-4.0 Å might not be a consequence of purity and instead inherent to the crystal packing structures. Although addition of the modified LNA-amide steps are localised to small region of the duplex, this could have a

significant influence on crystal formation. In this case the increased oligonucleotide purity will not improve diffraction resolution, because LNA-backbone modifications could be disrupting crystal packing, preventing high resolution structures.

### *3.4.5 Crystallisation condition optimisation using an unmodified duplex*

Modified oligonucleotides are expensive and labour intensive to produce. Because of this condition optimisation was performed using an unmodified DNA sequence (ON3.2) to obtain similar structures to those published, before applying these methods to the structures with modified backbones.

Condition optimisation is performed with manual screens using a 24 well plate. Compared to high throughput methods, drop volumes are much larger (1-5 µL *vs* 0.1 µL) and assay set up is considerably more labour intensive. Plates are designed to systematically alter variables across the plate. In a one-dimensional set up, one variable is altered per row, with each row altering a different variable (Figure 3. 11). The best conditions from this can then be selected and expanded. The second round of screening then optimises within a localised area of the initial hit. Smaller variation is made to two variables, one across, and one down the plate. This two-dimensional screening process allows for fine tuning of crystallisation conditions and can be repeated as many times as required.[136]

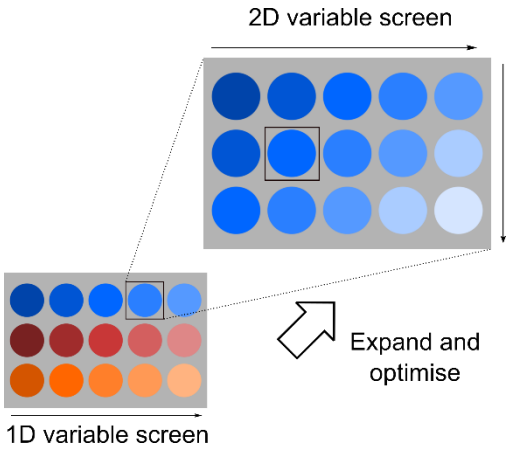



*Figure 3. 11 Outline of experimental set up for manual optimisation screens, first round consists of a one-dimensional screen where each row alters one variable. The best condition is then selected, and two variables are optimised, one condition is optimised across the row and the second down the column. The converging technique can be repeated to find the best condition.*

Using gel purified duplex (ON3.1:ON3.2), 24 well sitting drop vapour diffusion screens were set up replicating the conditions from the literature.[143] Unmodified duplex was crystallised in 12 mM MgOAc, 0.6 mM spermidine HCl, 0.075% β-octylglucoside, 12 mM sodium cacodylate

trihydrate and 12% (v/v) 2-methyl-2,4-pentanediol (MPD) against a reservoir of 1:1, MPD:$H_2O$. Conditions around these values were also tested by varying the concentration of MgOAc (Figure 3. 12, A1-A6), MPD (Figure 3. 12, B1-B6) and spermidine HCl (Figure 3. 12, C1-C6). In all cases large needle like structures were obtained (0.025-0.2 mm). In general, altering MgOAc or spermidine HCl concentrations had a minimal influence on crystal morphologies. Increased percentage of MPD reduced the number of crystals, instead creating longer and fewer needles. However, the crystal morphologies remained unchanged.

Crystals were sampled from several wells, including the reported literature conditions and data was collected at diamond light source (beamline iO3 or iO4). No significant improvements were found in resolution (2.8-5.0 Å), but crystals were consistent, and easy to produce providing a suitable condition for optimisation. In this experiment, it was found that our crystal forms were exclusively in a high symmetry $P6_1$ space group, contrasting the known space group $P2_12_12_1$. This could suggest that crystals we have obtained, have lattices with poorer diffracting properties, and that if we could influence the crystal growth by condition optimisation, it may be possible to produce better diffracting crystals.

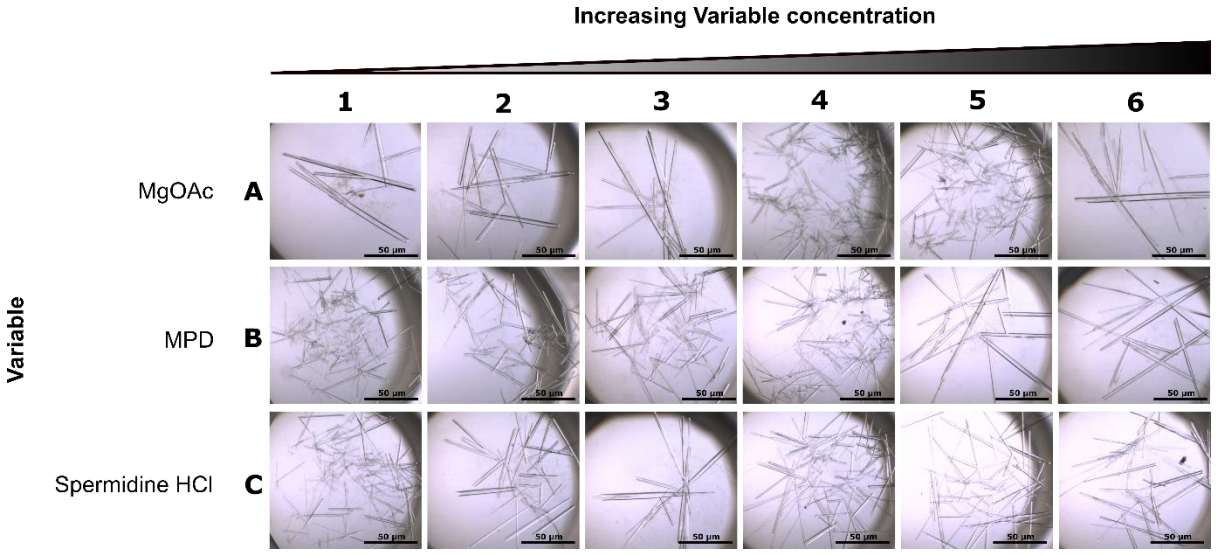



*Figure 3. 12 Manual screen of unmodified duplex ON3.1:ON3.2 to optimise crystallisation conditions. Variables were altered as a one-dimensional screen increasing from well 1 to well 6. (A1-A6) Variable = MgOAc, 6-16 mM, 2 mM intervals. (B1-B6) Variable = MPD, 6-16 mM, 2 mM intervals, (C1-C6) Variable = spermidine HCl, 0.2-1.2 mM, 0.2 mM intervals. In general, MPD had the most significant effect, reducing crystal nucleation at increased concentration with bigger crystal size.*

Expanding from the well C3 from the one-dimensional screen (Figure 3. 12) a second plate was set up varying pH and MPD (Figure 3. 13). pH values 4.8, 5.8, 6.8 and 7.8 were tested against variations in MPD from 10-22.5% in a 2D plate. pH can have substantial effect on

oligonucleotide structures allowing for changes in conformation or triplex formation.[145,146] Alternatively MPD has very little effect on the structures themselves, but as the precipitant in the screen, it is one of the key driving forces in drop diffusion. From systematic screening of these variables it was found that pH had little effect but higher percentages of MPD increased the rate of crystal nucleation. Crystals grown in ≥ 15% MPD were relatively fast to grow and formed the same needle structures as before. After 28 days, a single crystal was found at lower MPD concentrations (12.5%). Different in morphology, the crystal was cubic in appearance with well-defined edges. Collection of data from this crystal showed that the crystal quality was high, diffracting to 1.17 Å and matching that of the published structures. The space group was found to be $P3_121$ and quick processing, molecular replacement and refinement, provided satisfaction that the structure was correct. Data was also collected from other crystals from the screen with high concentrations of MPD which diffracted to 2.6 Å, like previous results.

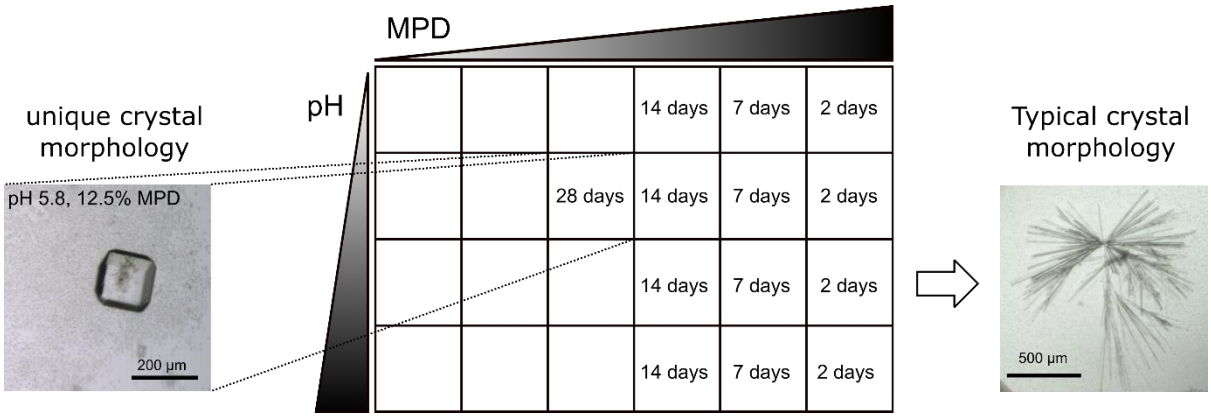



*Figure 3. 13 Optimisation screen of the unmodified duplex ON3.1:ON3.2 varying pH vs MPD concentration. pH had little effect on crystal growth but reduced MPD concentration increased crystal nucleation time. One crystal was formed at pH 5.8 in 12.5% MPD. Analysis by XRD showed that the unique crystal diffracted to a < 1.2 Å resolution, matching that of the published structure.*[143]

The crystallisation conditions from the well that produced the unique cubic crystal morphology was compared to the published conditions (Table 3. 2).[143] Other than using a higher concentration of duplex and MPD, the crystallisation conditions were found to be almost identical. This suggests that experimentally our crystallisation experiments are able to reproduce the published literature to produce results with comparable diffraction resolutions (< 1.2 Å). Despite reproducing results from the literature, the crystal obtained from this screen was found to be unique within the context of this work. Reproduction of this exact screen did not produce the same results and no cubic shaped crystal morphologies with close to atomic

resolution were obtained. These conditions were also screened with LNA amide modified duplexes with little success. This lack of consistent reproducibility highlights the difficulty inherent to crystal formation. The exact reason why this crystal was produced on a single occasion is not fully understood, but from the several thousand crystals grown throughout this project, it suggests that it is rare but not impossible to obtain.

*Table 3. 2 Conditions used to crystallise the unique crystal morphology with diffraction resolution < 1.2 Å compared to the published conditions by M. L. Kopka et. al.[143]. Despite multiple attempts with various sequences, results were not reproducible, and the crystals had alternative needle morphologies that diffracted between 2.6 – 4 Å.*

| Variable | Unmodified unique crystal conditions (*This work*) | Published Unmodified crystallisation conditions (M. L. Kopka *et. al.*[143]) |
|---|---|---|
| pH | 5.8 | 5.8 |
| MPD / % | 12.5 | 12 |
| Spermidine HCl / mM | 0.6 | 0.6 |
| MgOAc / mM | 12 | 12 |
| β-octylglucoside / % | 0.075 | 0.075 |
| Sodium cacodylate / mM | 12 | 12 |
| Duplex / mM | 0.5 | 0.3 |

### *3.4.6 Obtaining crystals for LNA-Am-LNA structures*

During the initial high throughput screens, data was collected for unmodified (ON3.2), amide (ON3.3) and LNA-amide (ON3.4) structures. However, no crystals were formed from oligonucleotides with LNA-Am-LNA (ON3.5) modifications. With the optimised conditions from the unmodified control, plates containing LNA-Am-LNA modified duplexes (0.5 mM) were set up based on the pH/MPD screen as before. No defined crystals were produced, instead micro crystals were obtained which were too small to collect data from. This could be due to rapid crystal growth or over nucleation within the drop. Possible ways to reduce this are either, to decrease oligonucleotide concentration, or use additives such as DMSO, alcohols (MeOH, EtOH), surfactants (triton X-114) or polyamines (spermine).[147] Alternatively, reduction in temperatures may allow alter the speed at which crystals nucleate forming different crystal forms.

Using repeat screening conditions, another plate was setup and incubated at 4 °C. After approximately 7 days, crystals began to nucleate that were large (0.2-1.2 mm) and consisted of several domains with undefined edges (Figure 3. 14). Data collection produced diffraction patterns to approximately 2.6 Å with high levels of twinning and statistical outliers. When

processing with phaser MR software, solutions were unconvincing with TFZ ≤ 8.0 and poor electron density maps. Analysis using Phenix.Xtriage suggested high levels of translational non crystallographic symmetry. This suggest that the crystal quality was poor, probably twinned, and unsuitable for structure determination.

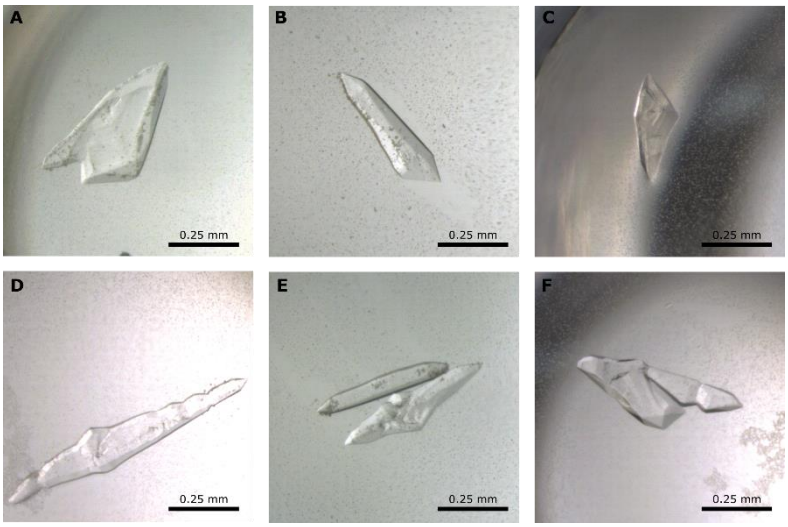



*Figure 3. 14 LNA-Am-LNA crystals grown at 4 °C were visually found to be irregular, suggesting multiple domains or twinning. Analysis by XRD found the crystals to be unsuitable for structure determination. Comparison to the unmodified crystals formed in the same conditions at rt, highlights how small structural differences between oligonucleotides can have significant effect on crystal formation.*

### *3.4.7 Crystallisation of oligonucleotides with LNA-Am-LNA at different positions*

Producing good crystals containing the LNA-Am-LNA proved to be difficult. In a new attempt, ON3.5 was resynthesized along with three other modified oligonucleotides (Figure 3. 15A, ON3.13-3.15) each containing the LNA-Am-LNA modification in separate locations. Modified DNA sequences were combined with complementary RNA and crystallisation was performed using a high throughput screen. ON3.14 (0.5 mM) initially produced a promising single crystal in a solution of 5 mM MgOAc.3$H_2O$, 25 mM 2-(N-morpholino)ethanesulfonic acid (MES) monohydrate (pH 5.6) and 1.25 M ammonium sulfate (Figure 3. 15B, a). This was then pursued in several screens, varying MgOAc (5.0-65 mM) and ammonium sulfate (0.5-2.5 M) but produced significantly worse crystals than the initial screen (Figure 3. 15B, b). This could be explained by the size of the drop. As mentioned, high throughput screens use 0.1-0.2 µL per sitting drop. In the manual screen it is up to 50 times larger (~5 uL). Because of this equilibration of salts across the drop can be much slower in the larger wells. As a result, this changes the conditions of crystallisation and means that scaling up volume does not always guarantee the same result.

Another Natrix HT plate of ON3.5 was set (drop volume ~0.1 µL). Although this screen had already been set in before in a previous experiment, it was repeated using a new batch of synthesised and HPLC purified oligonucleotide and new stock solutions. Results from this screen produced several well-defined crystals (Figure 3. 15B, c) and data collection provided a diffraction set to approximately 2.6 Å. Attempts were made to optimise crystal growth, but manual screens in identical conditions failed to produce any crystals (1-90 days).

A

| Oligonucleotide / ON | Sequence /5'→3' |
|---|---|
| ON3.5 | CT**T$^{L}_{Am}$T$^{L}$**TCTTTG |
| ON3.13 | CTT**T$^{L}_{Am}$T$^{L}$**CTTTG |
| ON3.14 | CTTTTC**T$^{L}_{Am}$T$^{L}$**TG |
| ON3.15 | CTTTTCT**T$^{L}_{Am}$T$^{L}$**G |

B

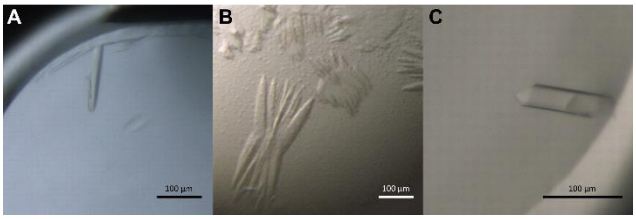



*Figure 3. 15 A) Table of synthesised oligonucleotides containing LNA-Am-LNA modifications at varying positions, B) a = ON3.14 LNA-Am-LNA crystal obtained from a high throughput screen. b = optimisation of ON3.14 crystal conditions which produced poor crystals dissimilar to the original hit. c = ON3.5 high throughput screen producing a single crystal that diffracted to approximately 2.6 Å.*

### *3.4.8 Summary of crystallisation conditions*

In total, crystals were obtained from four unique DNA:RNA structures which include: the unmodified duplex (ON3.2), amide (ON3.3), LNA-amide (ON3.4) and LNA-Am-LNA (ON3.5) modifications. Highlighted in Table 3. 3 are the final conditions used and data collection was performed from crystals grown in these conditions. In all cases, crystals were grown using MPD as the precipitate in a buffer of sodium cacodylate. Metal anion identities vary but were used in similar concentrations (12-40 mM). All crystals were grown at room temperature using sitting drop vapour diffusion and crystallised within two weeks.

*Table 3. 3 Summary of crystallisation conditions used to obtain crystals of combined LNA amide modified oligonucleotides.*

| Structure | Conditions |
|---|---|
| Unmodified (ON3.2) | **Drop solution** – 0.5 mM duplex, 12 mM $Mg(OAc)_2$, 0.6 mM Spermidine.HCl, 0.075% β-Octylglucoside, 12 mM Sodium cacodylate, 12 % (v/v) 2-Methyl-2,4-pentanediol (MPD), pH 5.8; **Well solution** – 1:1, MPD:$H_2O$ |
| Amide (ON3.3) | **Drop solution** – 0.5 mM duplex, 20 mM Lithium chloride, 40 mM Strontium chloride hexahydrate, 20 mM Sodium cacodylate trihydrate pH 7.0, 15% (v/v) 2-Methyl-2,4-pentanediol, 6.0 mM Spermine tetrahydrochloride; **Well solution** – Component concentrations are double the concentration of the drop solution and do not contain the oligonucleotide duplex |
| LNA-amide (ON3.4) | **Drop solution** – 0.5 mM duplex, 40 mM Sodium chloride, 6.0 mM Potassium chloride, 20 mM Sodium cacodylate trihydrate pH 7.0, 17.5% (v/v) 2-Methyl-2,4-pentanediol, 6.0 mM Spermine tetrahydrochloride, **Well solution** – Component concentrations are double the concentration of the drop solution and do not contain the oligonucleotide duplex |
| LNA-Am-LNA (ON3.5) | **Drop solution** – 0.5 mM duplex, 40 mM Sodium chloride, 6.0 mM Potassium chloride, 20 mM Sodium cacodylate trihydrate pH 6.0, 15% (v/v) 2-Methyl-2,4-pentanediol, 6.0 mM Spermine tetrahydrochloride; **Well solution** – Component concentrations are double the concentration of the drop solution and do not contain the oligonucleotide duplex |

## 3.5 Data Processing of LNA-amide structures

Single crystal X-ray diffraction data was collected at the German Electron Synchrotron (amide and LNA-amide) or Diamond Light Source (unmodified and LNA-Am-LNA). Raw images were processed using auto processing (xia2.DIALS, xia2.3dii or fast_dp) or by in house processing using iMosflm and scaled using Aimless.

Having processed the data into a dataset, the structures were solved by molecular replacement. Taking the original PDB (1PJO[143]) as the ensemble model, Phaser was used to calculate an electron density map from the collected reflections using the phase information from the published coordinate file. Molecular replacement was determined successful if the

TFZ score > 8.0 and LLG > 50 followed by a visual inspection of the electron density map to ensure chemical validity. Refinement of the model was then performed using refmac5 with an initial run of 150 jelly body cycles, obtaining an $R_{free}$ < 0.35 for each structure. For context, as $R_{free}$ approaches 0.50 it suggests that the model is no more significant than comparing a random model to your original data. However, producing an $R_{free}$ < 0.35 from an initial refinement, without manipulation of the model, indicates a high likelihood of the correct solution.

## 3.6 XRD model building

Until this point each model (unmodified, amide, LNA-amide and LNA-Am-LNA) was processed using the same procedure. For the unmodified control, the model was further improved by successive rounds of refinement until the structure reached the point where no further improvement in $R_{free}$ or modellable density could be made. For the modified structures containing amides and LNAs, models needed to be adapted to contain the specific modification that had been introduced during chemical synthesis, but which was not present in the model after molecular replacement. Modelling software contains libraries of parameters called restraints, which inform the program how to model quantities such as bond length and angles within a structure. For custom ligands, this dictionary of restraints does not exist, and therefore must be generated to include our unique amide dimers. The amide (Figure 3. 16, pale blue) LNA-amide (dark blue) and LNA-Am-LNA (purple) modifications were modelled as dimers containing the modified backbone and two surrounding nucleosides. These dimers were generated in Jligand which outputs the restraints that can then be used to incorporate the modified backbones into the model.[148]

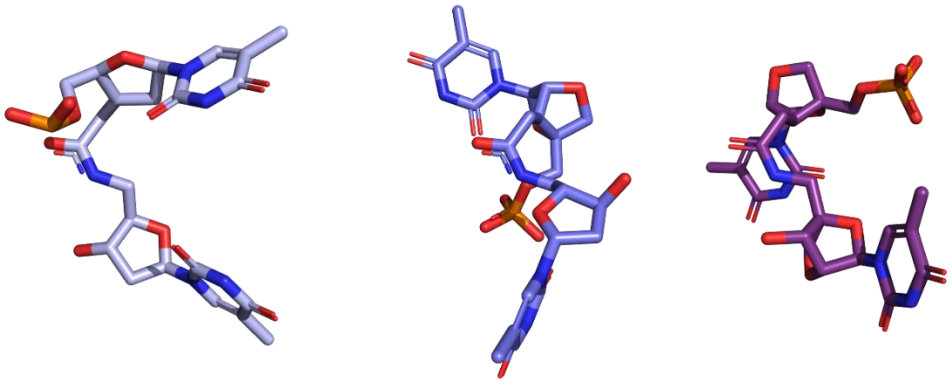



*Figure 3. 16 Molecular representations of the restraint libraries generated to incorporate the modified backbone into the duplex model. Green = DNA-Am-DNA, Blue = LNA-Am-DNA and Purple = LNA-Am-LNA. Structures have been pair aligned and orientated using the atoms of the amide backbone (N-C=O).*

Since an unmodified duplex 1PJO, was used as a model for molecular replacement, the refined models up to this point contain only unmodified backbones (Figure 3. 17, left). To incorporate our modifications into the model, the phosphodiester and surrounding bases are removed from the structure, and the required dimer restraints (Figure 3. 16) are included in the model. The dimer is fitted to the electron density and subjected to sequential rounds of refinement. PDB files are a coordinate-based format that gives the position of each atom relative to an origin in space. Software such as COOT or pymol, then interprets these files to represent covalent bonds (Figure 3. 17). Fitting the ligands was a challenge and required careful considerations to ensure the dimer identities within the structure were correct. If this was done incorrectly, then refinement would create unfeasible structures, remove links, or interpret the modified dimers as separate biomolecules from the rest of the duplex. These issues were overcome with a significant amount of manual editing of the PDB file, allowing refinement to continue without breaking the DNA chain and successfully incorporating the modification (Figure 3. 17, right).

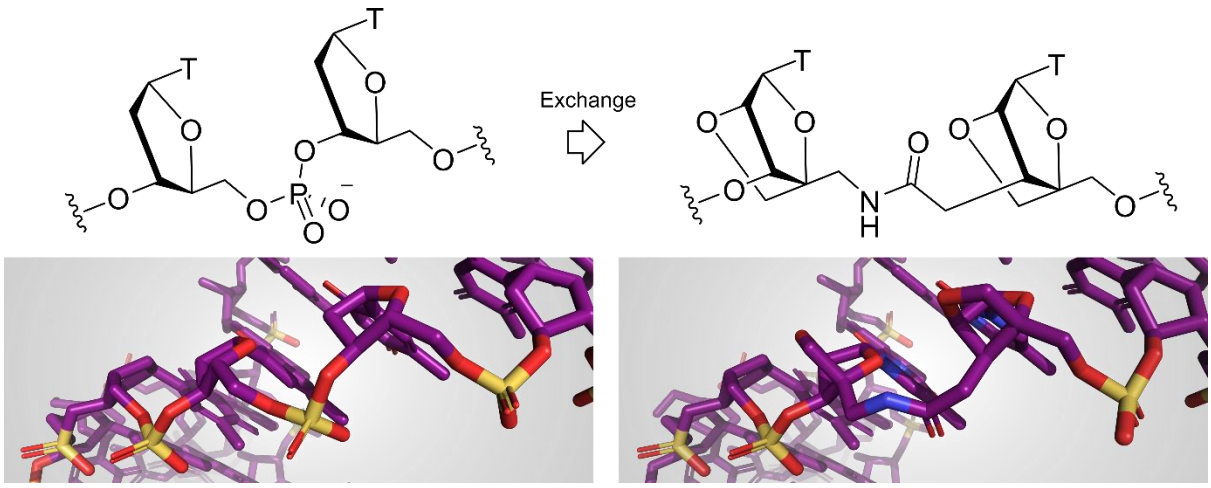



*Figure 3. 17 Example images showing an original structure from molecular replacement containing phosphodiesters (left) and the same model after addition of the LNA-Am-LNA (right). Modified restraints were generated using Jligand and then imported into the structure using COOT.*

Importantly, it must be determined if the amide is in the *cis* or *trans* conformation. When generating the modified dimer dictionaries, the software regularises the ligand into a conformation based on bond lengths and angles. The output produced *cis* conformation amides. To test if the dimer should be *cis* or *trans*, both conformations were fitted into the same electron density. In theory the wrong conformation should either be corrected during refinement or produce higher $R_{free}$ values indicating the backbone should be switched to the alternative conformation. In practice, both conformations produced the same $R_{free}$ values and the reason for this is due to the low resolution of the electron density. In order to determine if the amide was in the *cis* or *trans* conformation using refinement, higher resolution is needed. However, this information can be inferred. In general, the *trans* conformation is energetically more stable than *cis*, because it facilitates the maximum distance between bulky groups and reduces steric clash. Furthermore, peptide bonds in proteins have been analysed to evaluate the frequency of *cis* peptide bonds in protein structures. Excluding proline residues, it was found that *cis* peptides occur ~0.03% of the time when not adjacent to proline, and when adjacent to proline occurrence was ~5%.[149,150] Since the duplex forms of oligonucleotides are more relaxed than the constrained environments around proline, it suggests that a *cis* amide conformation is unlikely. This theoretical assumption has also been tested experimentally. In XRD structures,[124,126] amides were reported to be in the *trans* conformation, and NMR studies by Rozners *et al*.[125] confirmed this by NOE experiments. It is therefore highly likely that the amide backbone is almost exclusively in the *trans* conformation.

For each structure, the respective modified LNA-amide backbone was fitted into the electron density. The models were then separately refined using phenix.refine until an acceptable $R_{free}$ value was obtained that could not be reduced further.

## 3.7 Analysis of DNA:RNA structures with amides and LNA modifications

### *3.7.1 Summary of model statistics*

The unmodified, amide and LNA-amide duplexes were crystallised in a $P6_1$ space group along with an LNA-Am-LNA structure obtained in a $P3_221$ crystal packing (Table 3. 4, full conditions 8.5.3. Building of the amide structure was aided by Dr J. Hardwick and structure validation was aided by Dr James Hall (Collaborator, University of Reading, remote data collection was aided by Dr Michael McDonough, postdoctoral researcher, University of Oxford). DNA:RNA duplex structures were obtained between 2.5-2.8 Å with $CC_{1/2}$ values greater than 0.3 and signal intensities $I/\sigma(I) > 1.0$. $R_{free}$ values were refined below 0.26 and $R_{work}$ was within 0.05 of $R_{free}$. Comparing to the original structures of the unmodified duplex, published by R. Dickerson *et al.*[143], it was expected that the unmodified DNA:RNA hybrid would allow for diffraction patterns up to 1.1 Å and be found in a $P2_12_12_1$ crystal packing. After multiple screens, only one crystal was obtained to this resolution and could not be repeated using identical conditions. Comparing between the published structures and our own, we suggest that the high symmetry $P6_1$ space group is unable to produce crystals that diffract to the same resolution as the $P2_12_12_1$ group. For the modified duplexes, $P6_1$ crystals were highly abundant and lower symmetry morphologies were never obtained. This suggests that the additional modifications prevent crystallisation of the duplex in the $P2_12_12_1$ group.

*Table 3. 4 Summary of data processing for XRD structures of DNA:RNA hybrids containing amide and LNA modifications. Data was validated using pdb validation.*[151] *Each dataset was collected from a single crystal, values shown in parenthesis are for the highest resolution shell.*

| Data collection | Unmodified | Amide | LNA-Am | LNA-Am-LNA |
|---|---|---|---|---|
| Space group | $P6_1$ | $P6_1$ | $P6_1$ | $P3_221$ |
| Cell dimensions | | | | |
| *a, b, c* / Å | 54.34, 54.34, 45.88 | 55.14, 55.14, 45.96 | 53.95, 53.95, 48.69 | 49.15, 49.15, 90.96 |
| *α, β, γ* / ° | 90.00 90.00 120.00 | 90.00 90.00 120.00 | 90.00 90.00 120.00 | 90.00 90.00 120.00 |
| $R_{meas}$ | 0.081 (1.772) | 0.121 (1.124) | 0.103 (0.969) | 0.078 (3.673) |
| $I/\sigma(I)$ | 25.2 (1.7) | 8.9 (2.2) | 10.8 (2.1) | 19.8 (0.8) |
| $CC_{1/2}$ | 0.996 (0.646) | 0.943 (0.654) | 0.992 (0.753) | 0.999 (0.329) |
| Completeness /% | 99.4 (27.17-2.67) | 99.8 (47.75-2.60) | 99.8 (46.72-2.80) | 99.0 (31.08-2.57) |
| Multiplicity | 19.3 (16.0) | 7.5 (7.1) | 7.3 (7.9) | 18.7 (18.5) |
| Refinement | | | | |
| Resolution / Å | 27.17 – 2.67 | 47.75 – 2.60 | 46.72 – 2.80 | 31.08 – 2.57 |
| No. reflections | 2243 | 2517 | 2040 | 4398 |
| $R_{work}$ / $R_{free}$ | 0.187, 0.207 | 0.205, 0.232 | 0.185, 0.238 | 0.237, 0.261 |
| No. atoms | 415 | 419 | 417 | 843 |
| DNA | 197 | 196 | 198 | 400 |
| RNA | 217 | 217 | 217 | 434 |
| $Mg^{2+}$ | 1 | - | - | 2 |
| $Sr^{2+}$ | - | 2 | - | - |
| $K^+$ | - | - | 1 | - |
| Wilson *B* factor | 81.0 | 69.8 | 83.4 | 86.5 |
| r.m.s. deviations | | | | |
| Bond lengths / Å | 0.002 | 0.002 | 0.004 | 0.003 |
| Bond angles / ° | 0.49 | 0.78 | 0.63 | 1.66 |
| r.m.s.d Unmod / Å | - | 0.43 | 0.39 | 2.48 |

The asymmetric units of the unmodified, amide and LNA-amide structures contained single copies of the DNA:RNA hybrids coordinated to either one or two metal ions. Due to the resolution of the structures (2.5-2.8 Å), it is not possible to visualise interactions with water molecules. However, the distance between the observed metal ions and the nucleic acid duplexes are too far for direct interactions. This suggests that $H_2O$ molecules bridge between the polar contacts of the duplex and metal ions.

The conditions used to crystallise each DNA:RNA hybrid are unique to each structure, which means that the identity of the coordinating metal ion is dependent on the salts used in the screen. Despite this, the atomic positions of these metal ions are consistent between

structures and found far from the modified base pair step. As a result, the varying identities of coordinating ion has minimal effect on structural analysis.

The LNA-Am-LNA structure was obtained in a lower symmetry space group and the asymmetric unit contains two DNA:RNA hybrids and has larger unit cell parameters. Space groups have different crystal packing arrangements, that can introduce structural changes, altering interactions between neighbouring molecules. Although not uncommon, these effects need to be reviewed on a case by case bases. In this instance, the trigonal (P3) and hexagonal (P6) space group families share similar packing features. As a result, no significant structural differences are observed due to the change in space group.

### *3.7.2 Comparison to published structures*

Before performing a detailed analysis of the modified backbones, it needs to be confirmed that the unmodified structure agrees with the previous structure published by Dickerson *et. al.* (1PJO).[143] A structural overlay of our unmodified structure (Figure 3. 18, grey) and 1PJO (green) shows good similarity between the two duplexes. Visual inspection confirms that the duplexes are both in the A'-form and that the two duplexes are good mimics towards the 3'-DNA end of the duplex.

Looking at the 5'-DNA end we find small perturbation between the two structures in the DNA strand. The reason for this is due to differences in crystal packing arrangements between the two structures. The original duplex (1PJO) was obtained in a lower symmetry $P2_12_12_1$ space group, whereas our unmodified structure is in a $P6_1$ symmetry. Due to these differences in packing, the duplexes in the $P6_1$ crystal pack very close together compared to the lower symmetry. If the 5'-end of the DNA strand remained in the position found in the $P2_12_12_1$ group, then it would cause atomic clash between adjacent asymmetric groups in a $P6_1$ arrangement. To prevent this, the DNA strand is pushed away from the unit cell edge, increasing the gap between the two duplexes (green-blue *vs* grey-blue) and therefore the two structures deviate towards the 5'-end of the DNA strand producing an overall r.m.s.d of 3.58 Å across all atomic positions.

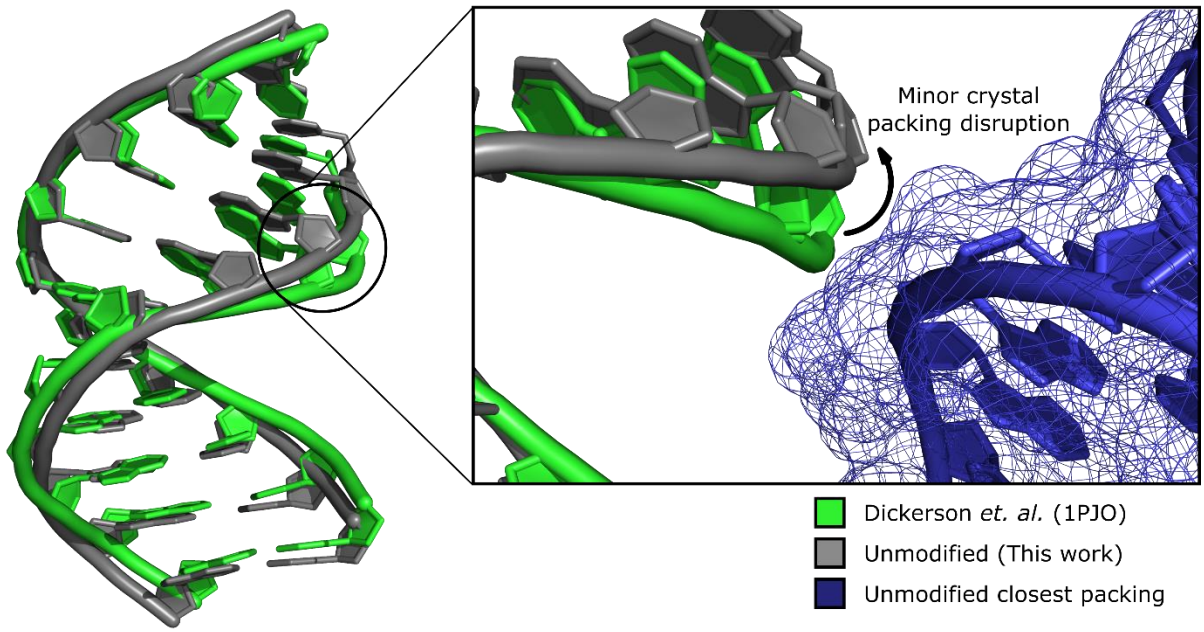



*Figure 3. 18 Comparison overlay of DNA:RNA duplexes showing structural similarity between the 1.1 Å structure 1PJO (green) and our own unmodified 2.6 Å structure. Sequences are identical but the duplexes were crystallised in different conditions with different crystal packing arrangements. The adjacent duplex (blue) is a symmetry mate of the unmodified sequence and shows the tight packing arrangement in the $P6_1$ crystal lattice. As a result of this the 5'-end of the DNA strand is displaced away from the 1PJO structure.*

Although disruption due to crystal packing is minor, it does cause changes to the base pair parameters (Figure 7. 6). For this reason, it is important to compare our modified structures to our own unmodified duplex, with similar packing arrangements. This will ensure that structural changes due to crystal packing are consistent between structures, allowing us to be confident that any observed change is directly due to the modifications and not crystal packing. Alongside this, it is better to compare crystal structures of similar quality and resolution. Comparison of a 2.8 Å modified structure to the 1.1 Å 1PJO duplex would not be appropriate because the high-resolution data is less reliant on model restraints and shows higher detail such as solvent interaction. A better comparison is to use our unmodified duplex at similar resolution and suitable space group arrangement.

In summary, we believe that the unmodified structure we have obtained is structurally consistent with previous crystal structures and differences between these can be accounted for by crystal packing effects. The unmodified structure presents a good comparison model for the modified duplexes, as it is of similar resolution and crystal packing.

### *3.7.3 Comparing the effect of LNA-amide modification on duplex structure*

The unmodified, amide, LNA-Am and LNA-Am-LNA duplexes were visualised using pymol and the structures were aligned for comparison. Highlighted in pink (Figure 3. 19) is the modified dimer step containing the specified amide modification. When comparing the modification section in overlay, we find that all structures are similar, and that there are only small deviations when compared to the unmodified duplex. Viewing the hybrid down the helical axis, the helical overlap and twist of the duplexes are structurally similar. When superimposed, the central base pair steps show significant overlap but there is structural deviation located towards the 5'-DNA end of the strand. The nucleic acid duplex becomes frayed at its terminal regions because there are fewer interactions holding the two strands together. Because of this, it is difficult to determine if differences are caused by unravelling of the duplex terminal region or due to the LNA-amide modifications. Despite this, structural differences are minor and unlikely to cause significant disruption to the duplex stability.

J. I. Gyi *et. al.*[152] found that DNA:RNA hybrids have characteristics somewhere between A and B-duplexes but are dominated by A-form character. In our structures we find that the hybrids adopt this intermediate duplex structure, denoted as A'-form throughout. On average the duplex has a twist angle of 30.7 ˚ per bp step and rise of 3.2 Å. This produces a duplex with 12 bp per helical turn, spanning a height of 38 Å. The A-form character pushes the base pairs away from the helical axis with an average x-displacement of -3.7 Å, accompanied by an average inclination of 12 ˚. Analysis of sugar puckers (Figure 7. 5), shows that most of the conformations are clustered around the 3'-*endo* conformation, a trait associated with A-form duplexes. If the duplex was closer in structure to a B-form hybrid, then we would see a larger distribution of pseudorotation angles and variation in sugar pucker.

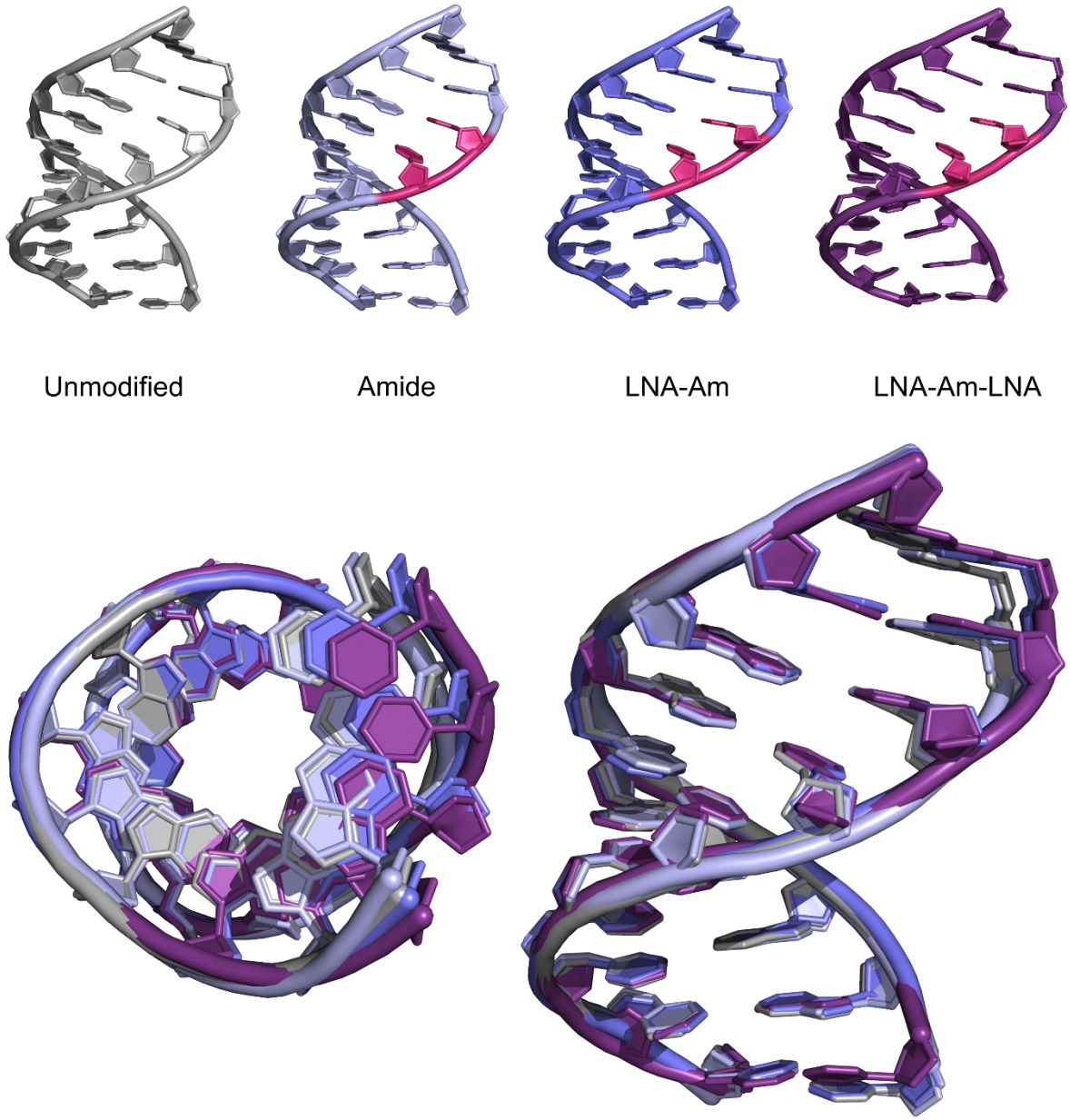



*Figure 3. 19 Duplex structures of DNA:RNA hybrids with LNA-amide modifications. Grey = unmodified, light blue = amide, dark blue = LNA-amide, purple = LNA-Am-LNA. Highlighted in pink is the modified dimer step which is incorporated at the 7-8 bp in the DNA strand. Structures were aligned in pymol to produce the comparative figure highlighting structure similarity from a top and side view.*

### *3.7.4 Structure and orientation of the amide backbone*

Overlaying the amide and unmodified duplex structures allows for comparison between backbone atomic positions of the amide and phosphate (Figure 3. 20). The sequences were aligned, and like structures obtained by Rozners *et. al.*[125] it was found that the amide linkage is a close mimic as a phosphodiester backbone. Both the amide and phosphate backbones are four atom linkages, which mean they are comparable in length, with atoms occupying similar spatial positions. The amide oxygen atom is orientated in the same direction as one of the P-

O bonds, and although the resolution of the structure is too low to observe interaction with water molecules, the amide is probably interacting with the surrounding solvent. Oligonucleotide helices are formed by the stacking of a hydrophobic core coated in an exterior of hydrophilic phosphates. Because both the amide and phosphate oxygen's have lone pairs, they can create similar hydrogen bonding interactions. However, unlike the phosphate, the amide backbone has only a single hydrogen acceptor and so is unable to form interactions that would otherwise be formed by the phosphate 'second' oxygen.

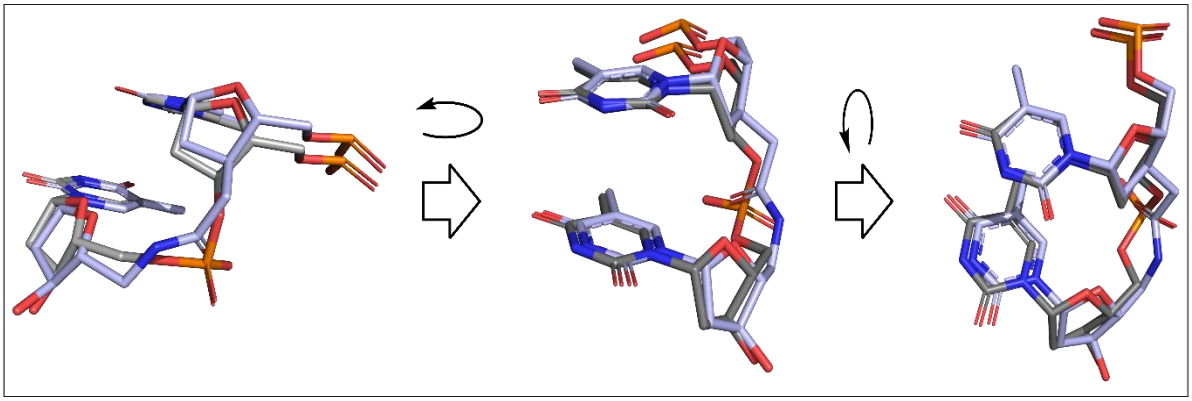


*Figure 3. 20 Overlay of the amide backbone (blue) and unmodified phosphodiester (grey) highlighting structural similarity and orientation. Key features include the direction of the amide carbonyl that is positioned in a similar direction to the P-O bond, and the alike amide backbone length compared to the phosphodiester. Sequence alignment was performed as a pair wise alignment across all atomic positions in the duplex, for visual simplicity the rest of the duplex was then removed in pymol to highlight the amide and phosphodiesters structures.*

Relative to the duplex, the amide backbone is positioned into the major groove, pointing the carbonyl group towards the internal region of the duplex (Figure 3. 21A). Crystal structures of four atom amides previously analysed have shown that amide orientation is dependent on both the linkage structure but also duplex form. In a structure obtained by Rozners *et. al.*[126] it was shown that the amide in a RNA:RNA oligonucleotide can be twisted to point the carbonyl into the duplex (Figure 3. 22, Figure 7. 7). They propose that this allows the backbone to form hydrogen bonding interactions with the adjacent H6 atom on the 3'-nucleobase. In the LNA-Am-LNA structure from this work, the distance between the base pair 3'-H6 atom and amide oxygen is too far for interaction. Measuring the distance between the C6 and amide oxygen in the LNA-Am-LNA structure produces a longer distance of 3.9 Å. Alongside this, the orientation of the amide oxygen is positioned further into the major groove, and not towards the C6 atom, suggesting minimal orbital interaction. Although with high resolution structures it is possible to directly visualise hydrogen atoms, at 2.6 Å this is not possible, therefore these measurements are estimates at best.

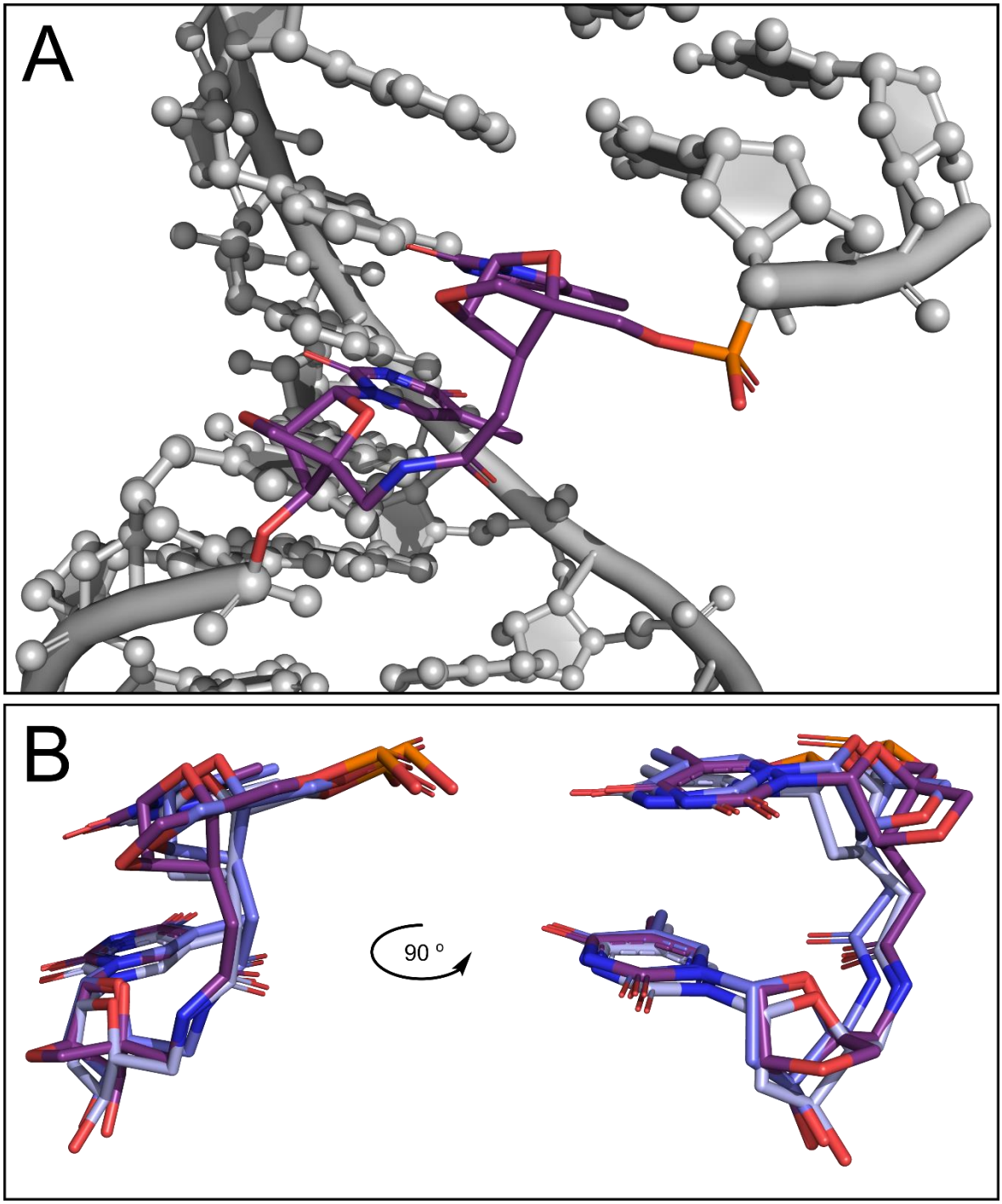



*Figure 3. 21 A) Modified duplex showing the orientation and structure of the LNA-Am-LNA step in relation to the duplex structure. The carbonyl group is directed into the major groove with the backbone in the trans conformation. As previously shown, the amide backbone is structurally like the phosphate linkage and there is structural similarity between the C-O of the amide and P-O bond of the phosphodiester. B) Overlay of all modified amide steps including amide (light blue), LNA-amide (dark blue) and LNA-Am-LNA (purple). This figure shows that amide, LNA-amide and LNA-Am-LNA modifications have similar geometry and orientations. The addition of LNA has minimal effect on backbone structure or positioning within the duplex, even when placed at the 5'-side of the backbone*

An alternative amide structure was also published by P. S. Pallan *et al*.[124] (Figure 7. 7). In this case the amide bond was adjacent to a branched tertiary carbon at the 5'-end of the backbone. The effect of this was that the amide linkage is rotated 180 ° out into the solvent

channel and away from the duplex. The influence of the branched carbon chains means that the backbone conforms to a structure that has the least steric interactions. In this case the amide backbone is still in a *trans* conformation but the ability to orientate itself in multiple positions suggests that the backbone has good flexibility. Furthermore, comparison between each of these structures suggests that the amide backbone is highly flexible. The amide has good freedom of rotation and the ability to adapt to various structural constraints, which makes it a good bioisostere and phosphate mimic. It also explains why the amide backbone causes little disruption to the duplex structure, reflected by good duplex stabilising thermodynamic properties.

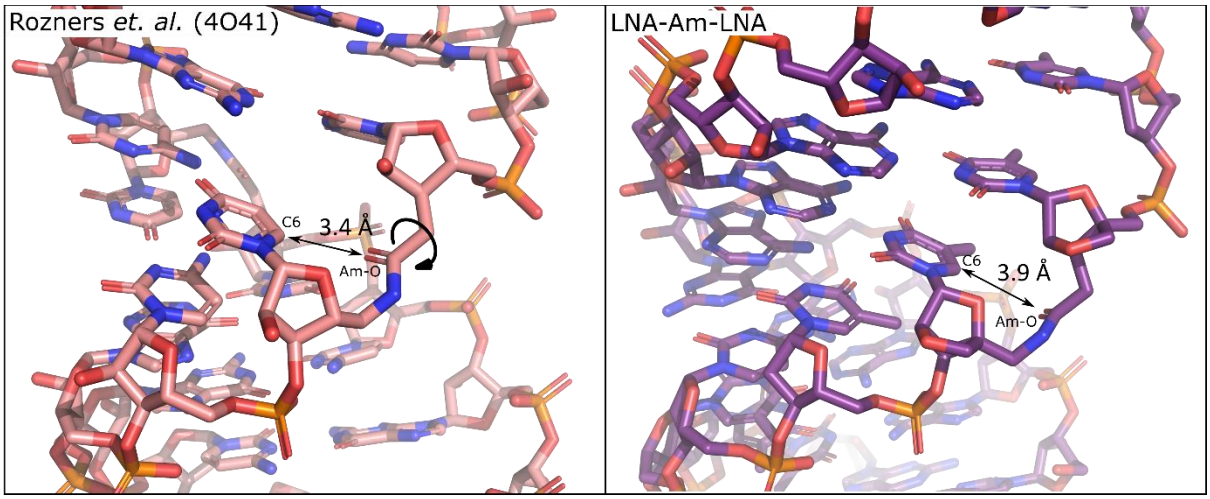



*Figure 3. 22 Comparison of amide backbone orientation and interaction with local hydrogens within the base pair step, (Left) published structure 4O41 shows an amide backbone in an RNA:RNA duplex, flanked by two RNA residues. The amide backbone is rotated into the duplex forming a hydrogen bond between the C-O carbonyl and H6 hydrogen on the nucleobase. (Right) LNA-Am-LNA modified DNA:RNA hybrid, the amide backbone is flanked by two LNA residues and unlike in the 4O41 structure the carbonyl group is facing into the major groove. Based on orbital alignment and distance to the C6 atom no hydrogen interactions are expected.*

### *3.7.5 Analysis of Amides and LNAs*

The modified backbones were overlaid to compare differences in backbone conformation and the effects of additional LNA modification (Figure 3. 21B). In each structure, the orientation of the backbone was found to be consistent, directing the amide oxygen into the major groove. Atomic positions of the backbones also show structural similarity and addition of 3'-LNA causes no significant displacement (Figure 3. 23). Addition of 5'-LNA relative to the backbone does however cause some structural displacement. The 5'-sugars in the LNA-Am and LNA-Am-LNA models are displaced outwards compared to the amide and unmodified strands. Despite this, the positioning of the amide backbone remains consistent between each structure.

As mentioned in the previous chapter of this work, LNA is thought to cause significant influence on the backbone directly at its 3'-side (Figure 3. 23).[41] Therefore, analysis of the amide torsion angles (Figure 3. 24) could provide significant insight into the structural influence of LNA on the amide backbone when placed at the 3' of the LNA.

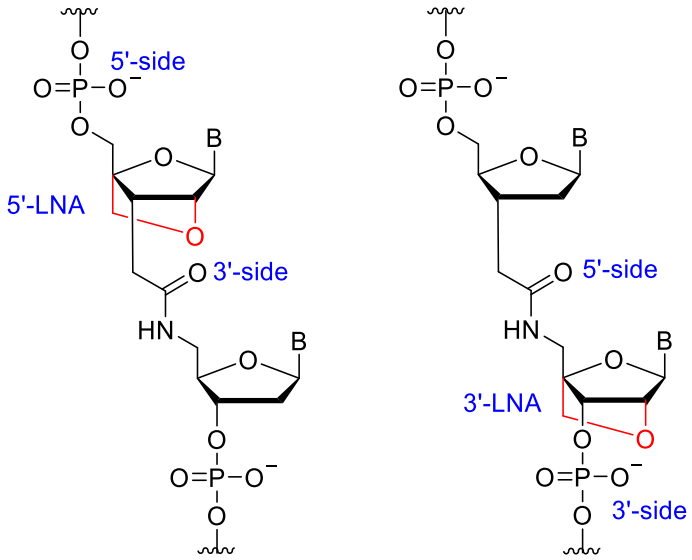



*Figure 3. 23 Structures showing the relationship between LNA and adjacent backbones. Backbones which are 5' to the LNA are defined at the 5'-side, those directly adjacent at the 3' are defined to be at the 3'-side. LNA infers greatest effect onto backbones at the 3'-side, this can therefore either be a phosphodiester or amide backbone.*

The amide backbone contains six torsion angles in the order 5'-ε, ζ, α, β, γ, δ-3'. Compared to the phosphodiester linkage, the amide linkage presents a different torsion angle profile despite having similar atomic positions as previously shown (Figure 3. 20). The amide exhibits torsion angles associated to a *trans* conformation (150 – 240 °) and this can be attributed to the $sp^2$ hybridisation of the carbonyl group. Comparing between amides with and without LNA, it is found that addition of the locked nucleic acid has a minimal effect on backbone torsion angle. When the amide backbone is placed at the 3'-side of the LNA the torsion angle varies by < 5 ° for each given angle. Alongside this, addition of consecutive LNAs (LNA-Am-LNA) also does not cause significant change in backbone torsion.

The resolution of the structures needs to be considered when measuring small variations in torsion angle. These models were obtained between 2.5-2.8 Å which is known as medium-low resolution. It is therefore inappropriate to make detailed comparison of the torsion angles between each structures, but it can be confidently concluded that the torsion angles are within the *trans or anticlinal* regions and that all amide structures exhibit similar torsion profiles with minimal disruption due to the addition of LNA.

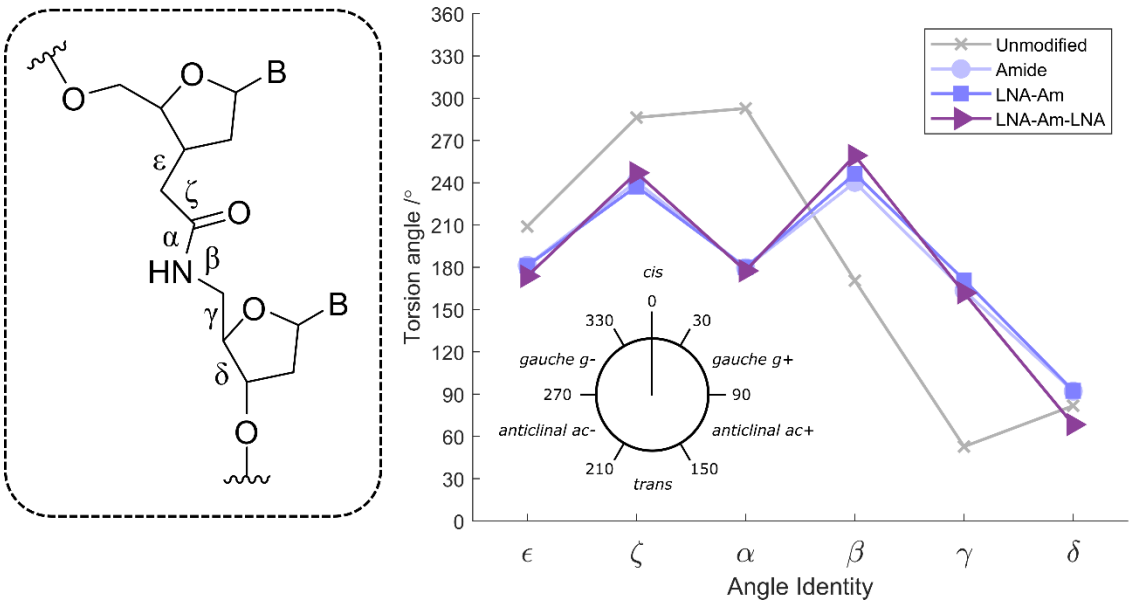



*Figure 3. 24 Torsion angle comparison between amide modified backbones. Specific torsions are given by their respective symbol (ε→δ, right). Modified backbones are located at the 7A-t base pair step. Also included is the 7A-T phosphodiester step from the unmodified duplex. In general, all amide backbones, regardless of LNA modification have similar torsion angles (data shown in Table 7. 3).*

### *3.7.6 Analysis of base pair parameters*

Base pair interactions and geometry exert major influences on duplex stability. To assess how LNA and amide modifications affect our duplex, we have identified several base pair parameters, which include inclination, propeller, rise, twist and x-displacement (Figure 3. 25). Plotting each parameter against each base pair in the duplex allows for analysis of local structural changes to see what effects our modifications have on the base pairs. The first positions have been defined as the rC-dG base pair, corresponding to the 5'-cytosine of the RNA and the 3'-guanine DNA. The modified LNA sugars are represented at the 7A-t and 8A-t positions where t is LNA or DNA residues surrounding the amide backbone which is the linkage between them.

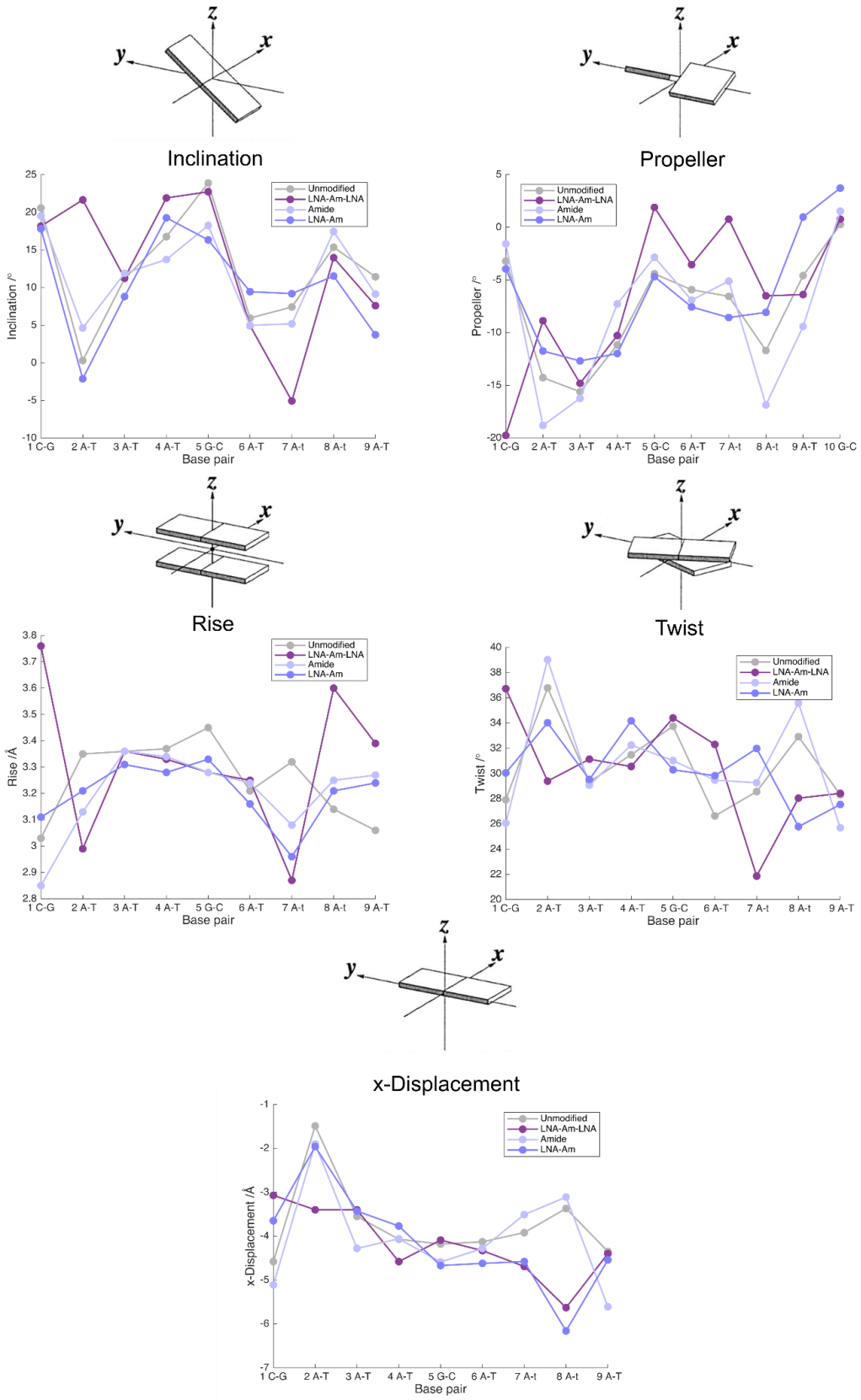



*Figure 3. 25 Scatter plots showing variation in base pair step parameters at each base pair step. X-axes show the base pair step for each data point which is plotted against various y-axis parameters. G, C, A, T are unmodified step positions and t represents the step that contains a respective (LNA)-amide modification. Modified dimers are contained within the DNA strand at positions 7 (A-t) and 8 (A-t) and position 1 is defined from the 5'-RNA strand. Above each plot is a pictorial representation of the base pair parameter, where z is parallel to the helical axis.*

Inclination describes the slope of angle between the base pair and y-axis. In general, A-form duplexes have an average inclination of +20 °.[153] Studying the profile of inclination across each base pair step in the same structure (1→10), we find that there is a high degree of variation. However, if we compare the same step across several structures, it is found that each given base pair step for each structure has similar values. In general, the LNA-Am-LNA model differs from these structures surrounding the LNA modifications at the 7A-T base pair. This suggests that the LNA sugars influence base pair stacking and therefore duplex stability. Because the amide and LNA-amide inclinations are so like the unmodified structure, it suggests that the amide backbone has minimal influence on inclination. Visualising the 7A-t step (Figure 3. 26), which varies by approximately 15 °, we find that there is no significant difference between the structures even in the LNA-Am-LNA step, and that minimal differences do not significantly perturb the of the duplex structure.

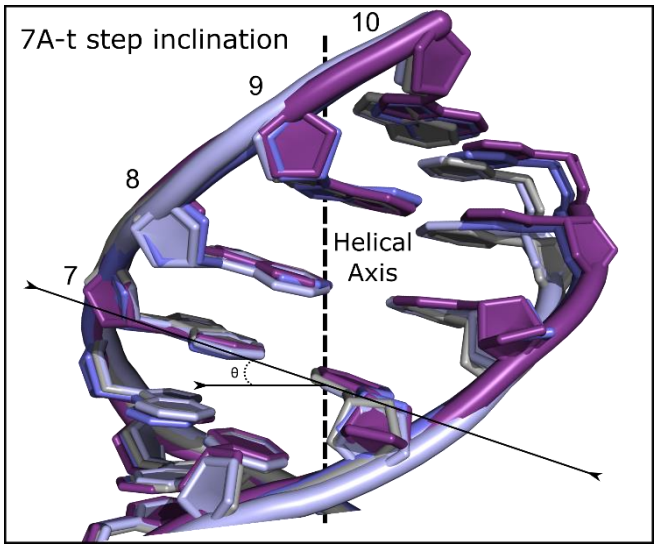



*Figure 3. 26 Inclination of base pair steps relative to the helical axis. Although compared to other steps the 7A-t shows the highest variation in inclination between structures, no significant disruption to the duplex is observed and variation is visually minimal.*

Propeller is a measure of counter rotation (+/-) between the two nucleobases in a base pair. Analysis of each model shows that each of our structures have high variations in propeller, with the central region of each duplex having positive propeller values. Due to the high levels of variation with no significant trend, it makes drawing a specific conclusion inappropriate. However, like seen with inclination, the LNA-Am-LNA structure shows the highest deviations from the unmodified duplex, and in general presents a more exaggerated profile across the duplex.

Analysis of propeller profile shows that the highest variation occurs at the 8A-t, surrounding the amide modifications. The LNA-Am-LNA 8A-t step has the smallest propeller (-6.1 °) and is approximately 10 ° less than the same base pair step in the amide duplex (Figure 3. 27). Although this is not going to significantly affect the thermodynamic properties of the duplex, higher propeller can result in weaker hydrogen bonding interactions reducing thermal stability. In general, LNA and amide modifications do not cause significant change to propeller angles.

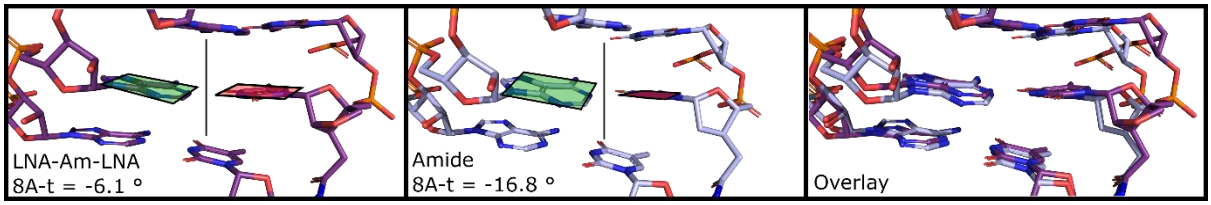



*Figure 3. 27 Comparison of base pair propeller at the 8A-t step between the LNA-Am-LNA and amide structures. The amide step has greater degrees of propeller but as shown in the overlay has minimal effect of duplex structure.*

Rise is the distance between two adjacent base pairs relative to the z-axis. Comparing between all structures, the central regions (positions 3-6) of the duplexes have consistent rise, and vary between 3.2 – 3.5 Å. The amide backbone is contained at the 7A-t step. Compared to the unmodified duplex, addition of an amide modification causes a smaller rise compared to the phosphodiester backbone. Addition of LNA in both the LNA-amide and LNA-Am-LNA structures also caused further reductions in rise and compared to the unmodified structures reduced the rise by approximately 0.5 Å (Figure 3. 28). This suggests that the amide linkage could be slightly shorter than the phosphodiester backbone reducing the gap between base pairs.

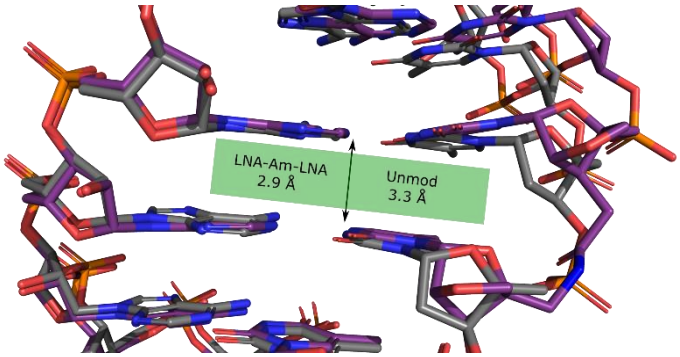



*Figure 3. 28 Base pair rise comparison between the LNA-Am-LNA and Unmodified duplex. Shown is the distance between the 7A-t and 8A-t base pairs. In general rise within the central region of the duplex was approx. 3.3 Å and LNA-amide modification reduced local base pair rise.*

Twist is a measure of the angle of rotation between adjacent base pairs about the z-axis. In general, A-form duplexes have a mean twist angle between 30-33 ° per base step. Like with previous base pair parameters, higher levels of disorder in the terminal regions of the duplex (24-39 °) compared to the central positions (27-34 °) were found. Focusing on the modified 7A-t step, it is shown that amide addition has minimal change to twist angle, again suggesting that the backbone is well tolerated within the duplex. LNA sugars are known to pre-form a duplex into the 3'-*endo* conformation and can directly influence adjacent base pairs towards the A-form structure. Based on this, it may be expected that LNA containing steps would have a twist angle close to the average for an A-form duplex. In our case we find that the LNA-Am-LNA step has a considerably lower twist angle of 22 ° and believe that this could be related to the reduced length of the amide backbone combined with a reduced ability to access some conformations that would otherwise be available for unbridged nucleic acids. Alongside this, lower twist angles could be caused by the positioning of the 5'-sugar at the 8A-t step. Comparing between the unmodified and LNA-Am-LNA structures (Figure 3. 29), it is shown that there is a displacement of the 5'-LNA at the terminal region of the DNA strand. Movement of this sugar compared to the unmodified structure could be a cause for a reduced twist angle, introduced by the LNA-Am-LNA modification. In turn the lower twist angle could affect duplex stability by altering π-orbital overlap within the duplex.

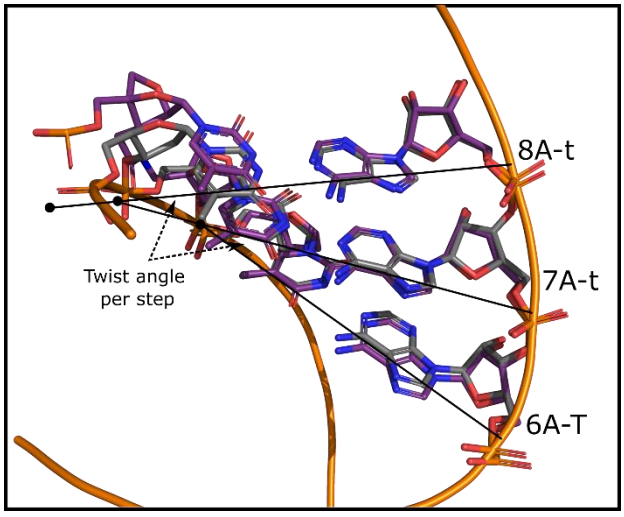



*Figure 3. 29 Depiction of twist angle between adjacent base pairs in the LNA-Am-LNA (purple) and unmodified (grey) duplexes. Highlighted in orange is the ribbon structure of the phosphate atoms in the unmodified duplex. Highlighted lines give an approximated indication of twist, which in both duplexes fluctuated between 26-36 ˚ per step. An outlier of this was the LNA-Am-LNA 7A-t step which has a reduced twist angle of 22 ˚.*

X-displacement describes base pair translation in the direction of the x-axis (Figure 3. 30). In general, higher values of x-displacement correspond to an increased distance from the helical centre. A-form duplexes have an x-displacement range from -2 to -5 Å. For our duplexes, we find that the A'-form structures conform to these x-displacement values and that the base pairs are pushed away from the helical axis. This results in a duplex with a deep major groove and shallow minor groove. Addition of the amide modification causes little change in x-displacement compared to the unmodified duplex. However, addition of LNAs in both the LNA-amide and LNA-Am-LNA structures do cause an increase in x-displacement to -6 Å. This is expected as the LNA is locked into the 3'-*endo* conformation, which has strong A-form character and A-form duplexes have large x-displacement.

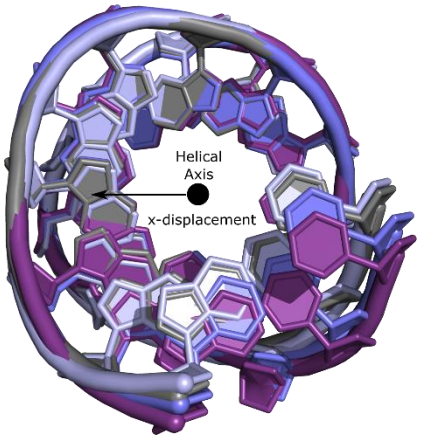



*Figure 3. 30 Top view of overlaid duplexes showing x-displacement of the nucleobases perpendicular from the helical axis which runs into the page, unmodified (grey), amide (light blue), LNA-amide (blue) and LNA-Am-LNA (purple).*

The 2A-T bp step is an unmodified section of duplex that appears to be an outlier for most parameters during analysis. Compared to the other steps, the 2A-T has higher inclination, propeller, twist and x-displacement. Sometimes outliers are introduced into a structure by crystal packing arrangement, but in this instance, this is probably not the case. In the original study published by M. Kopka *et. al.*[143] it was noted that the conformation of the adenosine in the 2A-T step was found in the 2'-*endo* conformation and that the backbone is pushed into an alternative conformation. In our 2.8 Å structures we are unable to see to this level of detail, but high levels of perturbation seen in our analysis suggests that our structure exhibits a similar disruption to the original structure.

### *3.7.7 Analysing combined LNA and amide modifications*

Unique to this work is the structural analysis we have obtained of LNA modifications combined with amide backbones. It is therefore possible to analyse the influence these two modifications have upon one another. An isolated amide without LNA confirmed that it is a good structural mimic for the phosphodiester backbone, causing minimal disruption to surrounding nucleobases. When LNA is introduced into a duplex, the local base pair step exhibits a reduced rise and twist, combined with increased x-displacement, signifying that LNA templates the duplex towards the A-form. These structural effects of LNA have been previously seen and are consistent with duplexes without amide backbones.[41,154] This suggests that the amide backbone does not restrict the properties of the modified LNAs and the amide allows the LNAs to behave as if incorporated within a duplex with canonical phosphodiesters.

Focusing on the influence of LNA on the modified backbone, we find that LNAs do not prevent the amide from being a good structural mimic for a phosphodiester. In most cases the parameter profiles of the amide and LNA-amide structures were very similar. It was only when LNA was added to both sides of the internucleoside linkage, that any significant changes in bp parameters were seen. The reason for this is most likely due to the additive properties of LNAs, which dominate the structural features surrounding the region as more are consecutively incorporated.

### *3.7.8 Analysis of localised hydrogen bonding*

LNA and backbone modifications can alter local duplex structure, but these effects drop off significantly with distance. If amide or LNA modifications cause change in hydrogen bonding, this will be visible in the closest surrounding base pair steps.

Analysis of all structures (Figure 3. 31A) shows that base pair hydrogen bonds were found between 2.6-3.2 Å, with longer interactions on the side of the major groove and shorter interactions on the side of the minor groove. Addition of amides, with or without LNA had minimal effect on hydrogen bonding and the structural overlay highlights the similarity in base pair positioning (Figure 3. 31B).

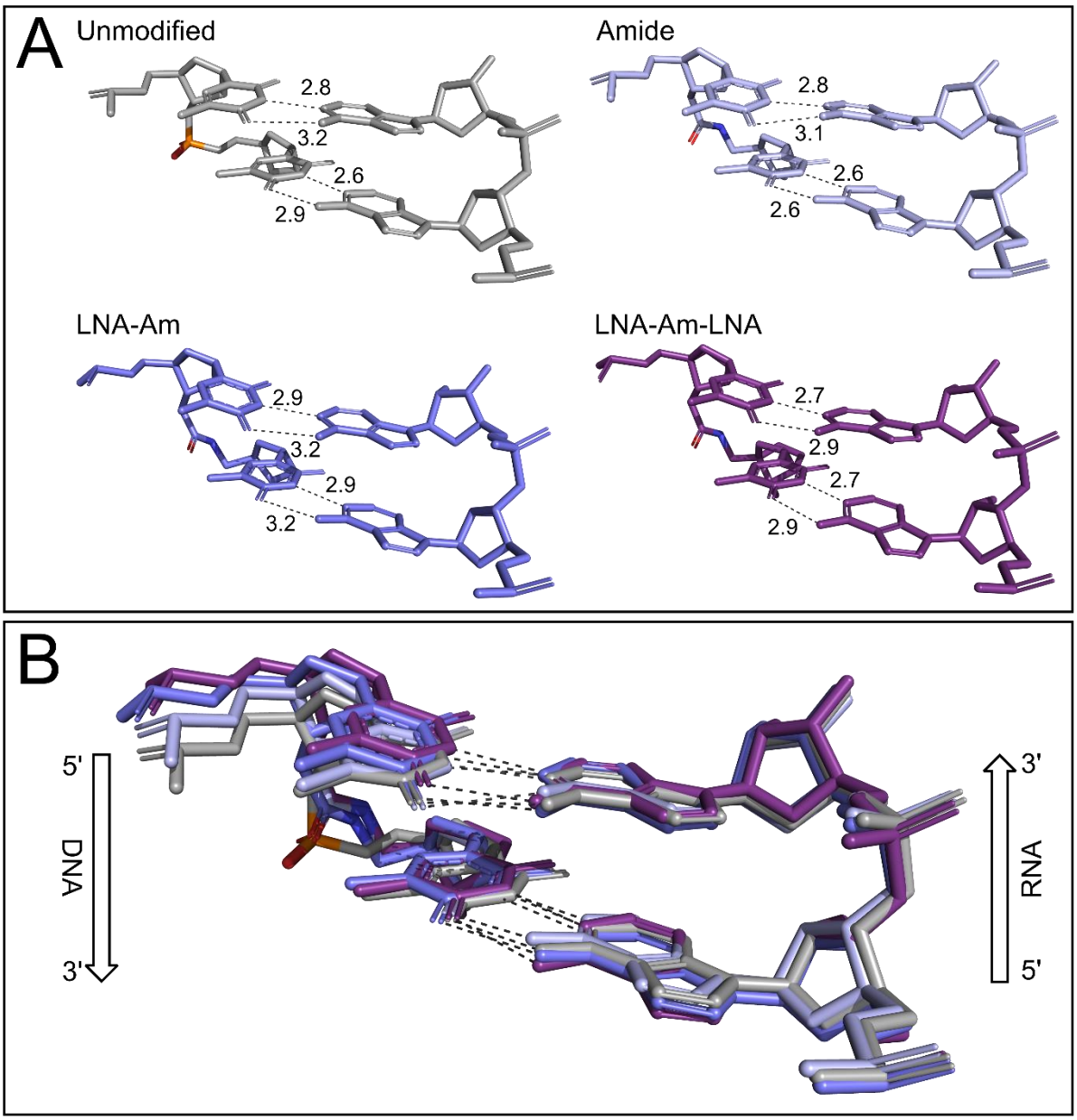



*Figure 3. 31 A) Hydrogen bonding interactions of the 7 A-t, 8 A-t base pair steps surrounding the site of modification. Unmodified = grey, amide = light blue, LNA-Am = dark blue and LNA-Am-LNA = purple. Modifications are contained within the DNA strand (Left side) and shows hydrogen bond distances between Watson-Crick base pairs. B) Overlay of above structures showing base pair step structural similarity.*

Amide and LNA modifications do not cause significant alterations to hydrogen bonding distances, and all base pairs adopt a canonical Watson-Crick structure. Linking hydrogen bonding results to the thermodynamic properties of our modifications, concludes that combined LNA and amide alterations have minimal effect upon hydrogen bonding, and that any thermodynamic improvements are not due to changes in hydrogen bonding interactions.

## 3.8 XRD structures with carbamate-LNA modifications

### *3.8.1 Phasing the CBM-LNA structure using molecular replacement*

Amides and carbamates are similar and only chemically differ at one atom position (Figure 3. 32), exchanging a $CH_2$ for oxygen. This change has a significant effect on duplex stability and the CBM backbone reduces $T_m$ on average by -4 °C per modification.[117] To compare CBMs to amides, we attempted to crystallise the carbamate backbone with LNA (ON3.9-3.12) to compare to the LNA amide structures. Of all crystals grown, carbamates produced the most abundant, well defined and best diffracting crystals (Figure 3. 9A-D). Despite this, these crystals could never be solved by molecular replacement using published structures (1PJO) or by our own unmodified duplex. Analysis using Phenix.Xtriage, suggested that the data quality was high and that it should be possible to obtain a structure to approximately 1.85 Å.

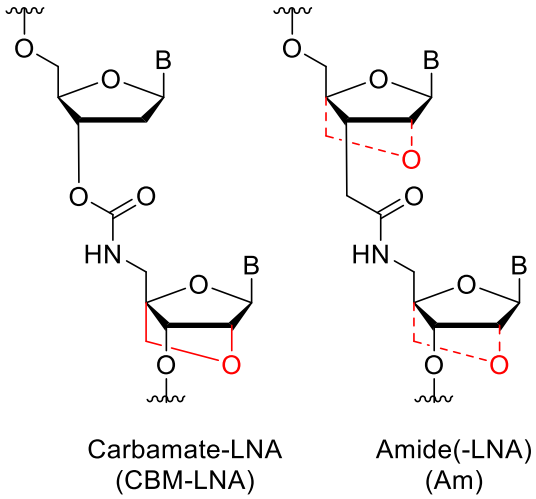



*Figure 3. 32 Comparison of carbamate and amide modifications.*

For a structure to be solved by molecular replacement, sequence identity should be >25% and for nucleic acids the r.m.s.d of the phosphate atoms must be within 1.0 Å.[140] In our case, we know that our sequence identity in terms of base pairs is 100%. Even if we assume that LNA-CBMs are interpreted as non-canonical bases, our sequence identity should be high enough to solve the structures using the same model as that used for the amide duplexes. The fact that molecular replacement consistently failed to produce a solution with multiple crystals, suggests that the model we are using, does not share enough similarity to the model we are trying to create.

Inspection of the log file during molecular replacement highlighted that the model we were using was too big to fit inside the unit cell. Evaluation of the Mathews coefficient, a parameter that estimates the number of molecules in a unit cell of given size, suggested that two duplexes are most likely to be in the unit cell based on the content percentage, unit cell size and molecular weight of the duplex. Assuming then that two duplexes were present but structurally different from the original model, several new molecules based on the original 1PJO structure were designed (Figure 3. 33, 1-6). Starting from the original model, the duplex was edited to reduce the size of the structure so that it could fit into the suggested smaller unit cell. Cutting off the ends of the chains (Figure 3. 33, 2-3) did little to improve statistical scores, as did removal of just the terminal RNA bases (Figure 3. 33, 5) or complete removal of the DNA strand (Figure 3. 33, 6). However, if a total of four base pairs were removed from the duplex, then a solution with TFZ score > 8.0 was obtained (Figure 3. 33, 4).

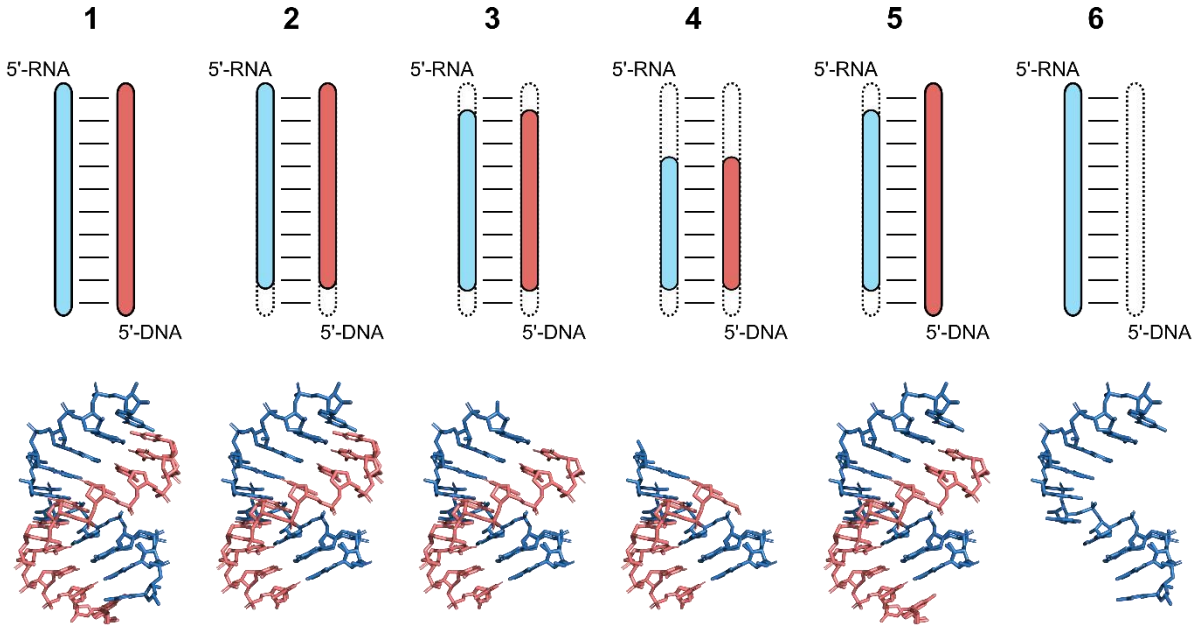



*Figure 3. 33 Search models used in molecular replacement for LNA-CBM phase determination. Regions of the model were removed from the full-length duplex 1PJO (1) to examine if removal of the terminal regions (2-5) or DNA strand (6) improve statistical likelihood search methods.*

### *3.8.2 CBM-LNA model building*

Having correctly phased the model using molecular replacement, missing sections needed to be manually rebuilt back into the oligonucleotide. Like with the amides before, the CBM backbone was modelled as a modified thymine residue. However, unlike the amide dimers previously used, the CBM was modelled as a monomer with the 5'-hydroxyl converted to a carboxylic acid functional group (Figure 7. 8). The CBM linkage was then formed during

refinement linking the 5'-carboxylic acid to the 3'-hydroxyl group of the adjacent residue. After addition of the CBM modification, the rest of the duplex was rebuilt into a canonical A'-form duplex. As the duplex was constructed, the electron density map was seen to visually improve, but the $R_{free}$ remained at ~0.35, much higher than expected. Visualisation of local crystal packing revealed that adjacent duplexes were very close and a section surrounding the CBM linkage was clashing with its symmetry mate. This clash, along with regions of unfilled electron density, suggested that the DNA:RNA duplex is not in a canonical structure and explains why we could not phase our data using the original published duplex.

The clashing elements of the duplex were removed from the structure. The CBM-LNA was then replaced but this time rotated approximately 90 ° out of the duplex towards the symmetry mate. Guided by the electron density, the new strand was then built across to the adjacent duplex, where new base pairs could be seen to invade within the symmetrical duplex. As the strand was added both the electron density and $R_{free}$ values improved. This suggested that the new model was correct, not formed of a single duplex but by a dimer of duplexes with DNA strands invading across to one another. The final dimer structure was refined to an $R_{free}$ 0.33 with a resolution of 1.82 Å (Table 7. 4).

### *3.8.3 CBM-LNA duplex structure*

The DNA:RNA duplex containing the modified CBM backbone can be structurally divided into two regions, which are the canonical A'-form base pairs towards the 5'-RNA end, and the strand invading section towards the 5'-DNA end (Figure 3. 34A). Analysis of the canonical region of the duplex, shows that on average each base pair step has an x-displacement of -4.2 Å, inclination of 28.4 ° and a rise of 3.4 Å. Although perturbed from a canonical DNA:RNA hybrid at the 5'-DNA end, this suggests that the duplex form remains close to the A'-structures previously analysed at the 5'-RNA end.

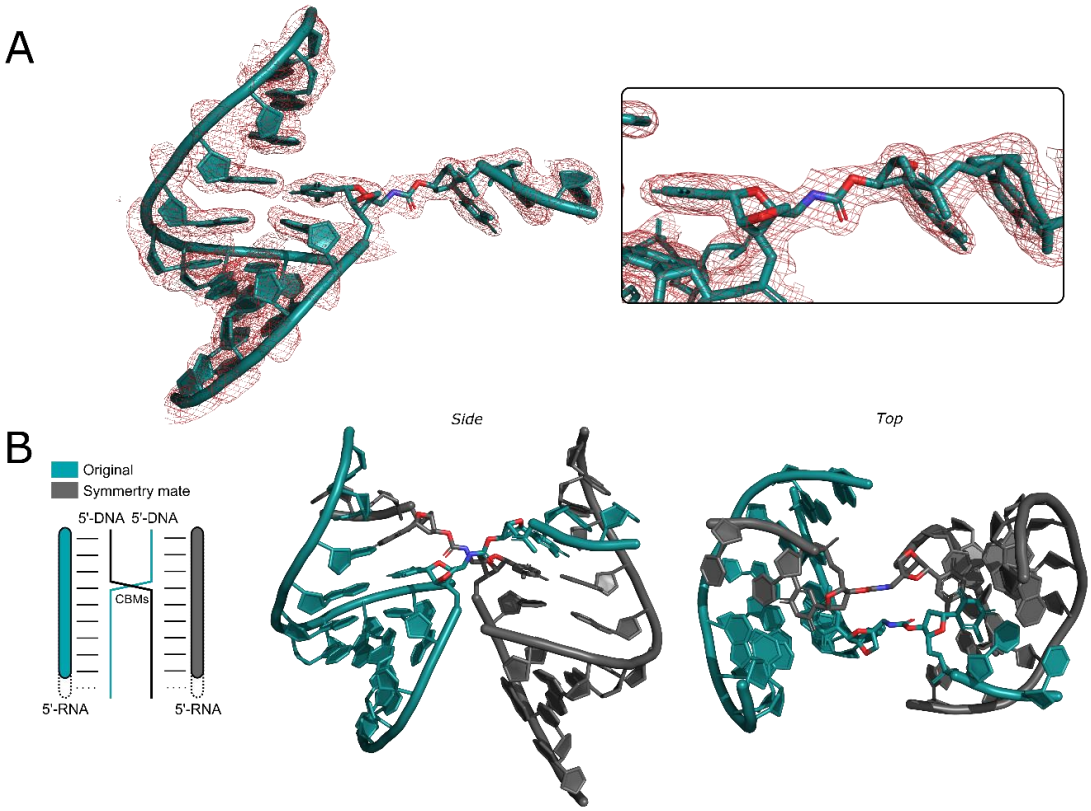



*Figure 3. 34 A) Crystal structure of a DNA:RNA duplex containing a CBM-LNA modification at the 7A-t step. Highlighted in red is the electron density (σ = 1) calculated from the diffraction pattern. The duplex follows a canonical A'-form structure until reaching the CBM-LNA modification, rotation around the CBM backbone causes the duplex to open into a non- canonical form with the DNA strand appearing to reach out into the crystal lattice. B) Crystal structure of the same DNA:RNA hybrid shown in (A) with addition of the closest packing symmetry duplex. Addition of the symmetry mate shows that the duplexes are packed as a dimer species with DNA strands invading into the opposite duplex structure, to form three canonical Watson-Crick base pairs. The strand invasion occurs at the CBM-LNA modification and was not seen in similar amide structures. Shown on the left is a two-dimensional representation of the crosslinking position and base pairing interactions. The 5'-RNA nucleobases are absent from this duplex and is addressed in sections 3.8.4 and 3.8.5.*

Towards the 5'-DNA end is the non-canonical structure where the CBM backbone reaches out of the original duplex into a new adjacent strand (Figure 3. 34A). Modelling the nearest symmetry mate, reveals that the two adjacent duplexes strand invade into one another, each forming three base pairs with the corresponding RNA strands (Figure 3. 34B). This structure is directly caused by the addition of the CBM backbone and was not observed with similar amide structures. Throughout this project, crystals that were obtained with CBM backbones were visually always found as near perfect hexagonal crystal geometries that could never be solved using the canonical MR models. Since crystal formation was abundant, and phase determination using a standard duplex was impossible, it suggests that for this sequence, the strand invading structures are the dominant crystallisation form.

Although the strand invading structure is facilitated by the CBM backbone, we do not believe that this should draw a conclusion that there are radical structural differences between

duplexes that contain amides and those with CBMs in solution. From our previous melting studies on CBM backbones (Chapter 2) and published work on amides[120] it suggests that the preferred conformation(s) of CBM are not as compatible with A or B form duplexes compared to the amide backbone. Given that CBM and amide backbones are almost isosteric and therefore are about the same length, the unusual X-ray structure which shows strand invasion is likely to occur because of the thermal instability located in the duplex where the CBM occurs. If this is the case, then it is not surprising that strand invasion occurs during the crystallisation process, as crystallisation causes adjacent strands in different duplexes to be very close to each other, and hence the local concentration is high enough to allow for the strand to invade. In this case, we believe that the strand invasion could be unique to this structure, combining the effects of the crystal packing, along with the positioning of the CBM within the duplex. If the CBM was positioned at a different base pair step, or an alternative sequence was used, then a standard canonical duplex maybe obtained.

### *3.8.4 CBM-LNA duplex crystal packing and special positions*

The original sequence is a 10 mer oligonucleotide consisting of complementary DNA and RNA strands. During manual rebuilding of the duplex, 10 DNA nucleobases were added but only nine RNA bases could be modelled into the electron density. Addition of the tenth RNA nucleobase caused significant increases in $R_{free}$ and clashes within the model. Modelling a different symmetry mate (grey, Figure 3. 35) next to the original duplex (teal) shows that the two duplexes pack towards the edge of the unit cell, revealing that the final RNA base sits directly on the unit cell axis (Figure 3. 35). Therefore, addition of the final nucleobase causes two symmetrically equivalent nucleobases to be superimposed upon one another, known as a special position. A special position is when an atom or residue sits on or crosses a symmetry element, which results in it being modelled in more than one-unit cell. To resolve this issue, it is sometimes possible to reduce the crystal packing symmetry. The CBM duplex structure was reprocessed in several P3nn space groups as well as the universal space group P1, which contains no symmetry operations. In theory this should remove the symmetry plane from the edge of the unit cell, allowing the two nucleobases to be modelled individually. However, this was unsuccessful, and the electron density remained unfilled with a high $R_{free}$ 0.33 value.

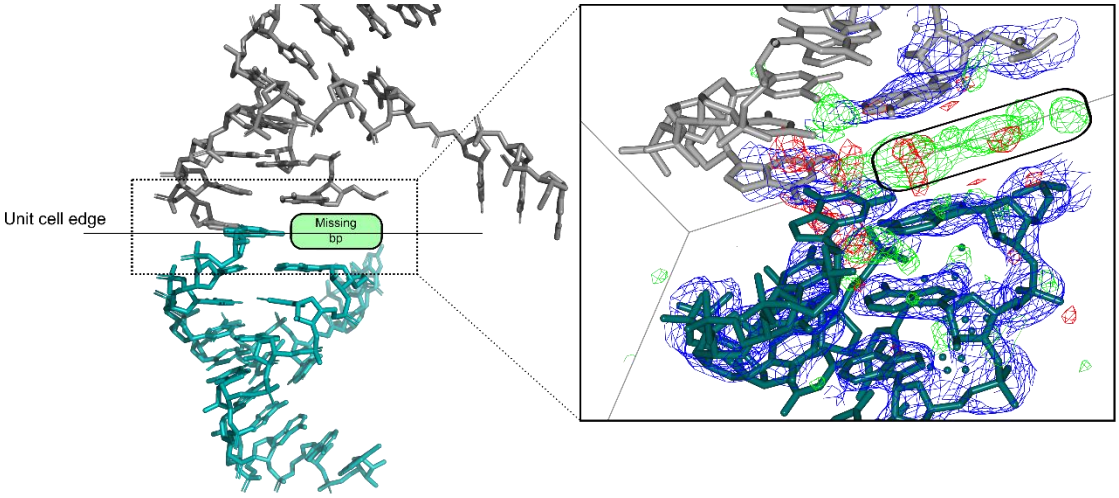



*Figure 3. 35 Crystal packing arrangement of duplexes in the CBM-LNA XRD structure. Two duplexes are shown in separate asymmetric units (teal vs grey) that meet at the junction of a symmetry plane. Either side of the unit edge sits two guanine residues and an unmodelled section of electron density. Shown on the right is the same junction with calculated electron density (σ = 1), blue represents satisfied density, red indicates disagreement between the model and electron density and green indicates the need to fill the electron density. The unfilled electron density sits directly on the unit cell axis known as a special position. Base pairs that would fill this density were omitted from the model, as they caused atomic clashes and poor refinement statistics.*

### *3.8.5 Base pairing around the terminal junction*

Although it is not possible to model all base pairs in our structure due to the special position, inspection of the partially constructed model and unfilled electron density allows us to infer potential structures. Sitting either side of the plane are two symmetrical guanosine residues. Unlike in a canonical Watson-Crick base pair, these nucleobases are rotated towards one another, which could suggest potential keto-enol hydrogen bonding between the two carbonyl groups. In relation to these guanine residues, the remaining electron density sits across from the two bases at a third position forming a triangular configuration in the plane of the base pair step, that will not fit the two cytosine nucleobases that are required to complete the model. This suggests that the occupancy of atomic positions is shared between the two DNA cytosine residues.

XRD requires atoms to be in a continuous arrangement of equal lattice positions. If atoms are mobile within the structure, then there is no resultant diffraction pattern to calculate a meaningful electron density. However, occupancy of an atom can be shared between multiple fixed positions and in these cases the electron density will represent an average of the structures. In the CBM structure, one of the DNA cytosines is flipped out of the duplex, while the second forms a hydrogen bonding interaction with one of the guanine residues within the junction (Figure 3. 36, conformation 1). Alternatively, an identical lattice position could have

the reverse structure, where the cytosines are swapped over, and the hydrogen bonding interaction is to the other guanine (conformation 2). Therefore, the overall electron density is a combination of these two structures, where the two cytosine residues are superimposed on top of one another, skewed towards opposite guanine residues. In total this creates a junction between the two duplexes where the guanines remain in fixed positions and the cytosines share the third position throughout the lattice.

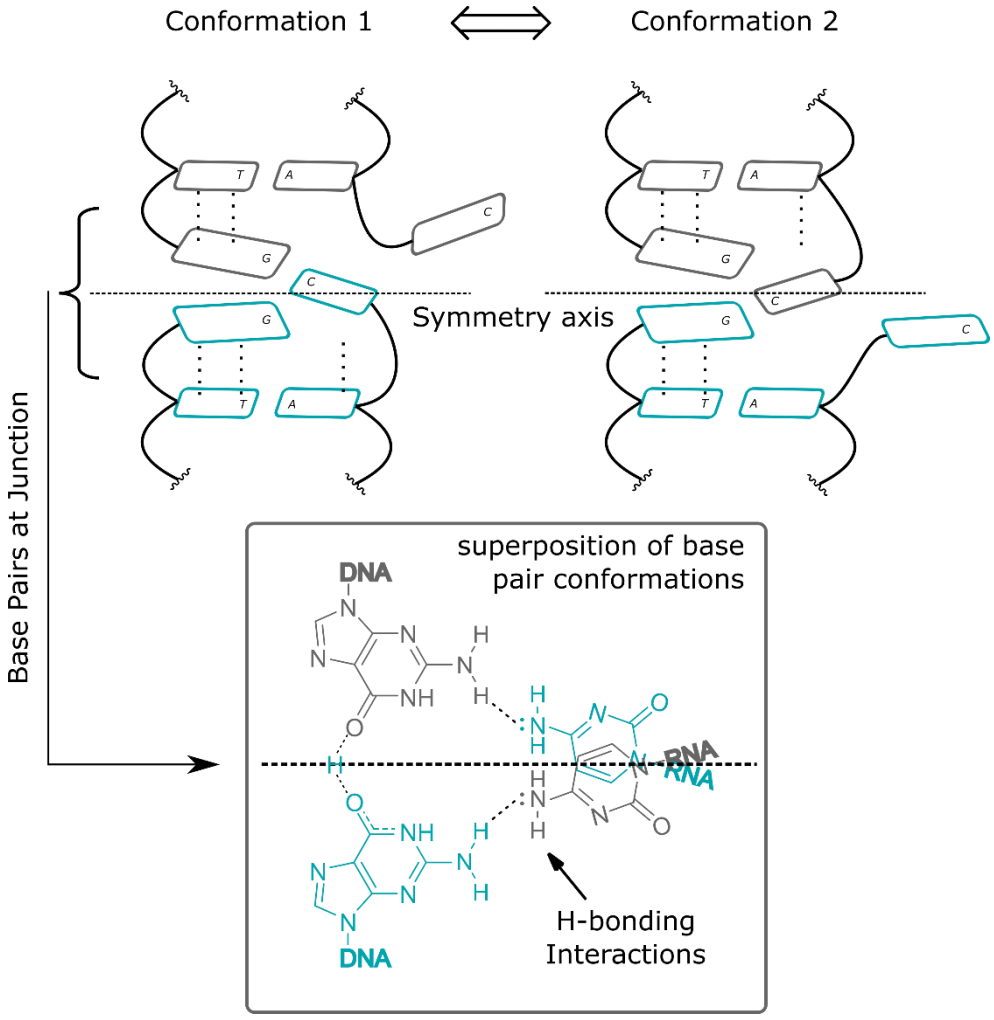



*Figure 3. 36 Block representation of the CBM-LNA duplex (teal) and adjacent duplex (grey) in the neighbouring asymmetric unit. The two duplexes meet at the edge of the unit cell at the symmetry axis with two guanines in a non-canonical interaction. Analysis of the electron density revealed that only one cytosine can be placed within the available space. Therefore, the occupancy is shared between two nucleobases and at identical lattice points throughout the crystal, the duplexes can adopt either conformation 1 or conformation 2. Shown below is a prediction of base pairing within the junction. When superimposed the two cytosines are placed over the unit cell axis interacting with the adjacent guanines.*

### *3.8.6 CBM-LNA backbone structure*

The CBM-LNA containing duplex has a non-canonical structure, which means that fine detailed comparison between the amide and CBM backbone would not provide a significant or reliable conclusion. However, we can rationalise why we see drastic differences between

structures with chemically similar backbones in identical sequences. During crystallisation screens of CBM oligonucleotides, a high number of crystals with identical morphologies were produced, and like the obtained CBM-LNA structure they could not be phased using the published canonical duplex (1PJO). This suggests that these crystals had strand invading hybridisation and that it could be the predominate structure to crystallise. Since these crystal morphologies were never obtained with amide backbones, we believe that the CBM linkage facilitates strand invasion, whereas the amide does not.

Chemically, the key difference between the two backbones is the exchange of the amide $CH_2$ for the CBM oxygen atom. This increases the $sp^2$ character of the 5'-end of the linkage and the CBM backbone has more resonance structures (Figure 3. 37). In A'-form duplexes we found that the amide backbone contains bent anticlinal torsion angles partly facilitated by the $sp^3$ hybridised $CH_2$ group. Although it is likely that the CBM backbone can adopt a similar conformation, the $sp^2$ backbone is likely to be more strained limiting the freedom of rotation around the **O-C**=O CBM bond *vs* **C**$H_2$**-C**=O amide. In our obtained structure of the CBM-LNA, we find that the backbone torsion angles (ε-γ) are all close to linear (Figure 3. 38). This means there could be an entropic factor during duplex formation that promotes strand invasion, and that the most common conformations of the CBM backbone in a single strand do not match those required for a canonical duplex. As a result, the backbone could be slower to adopt the conformation required for duplex formation and that this influences formation of a dimer structure with invading strands at high oligonucleotide concentration.

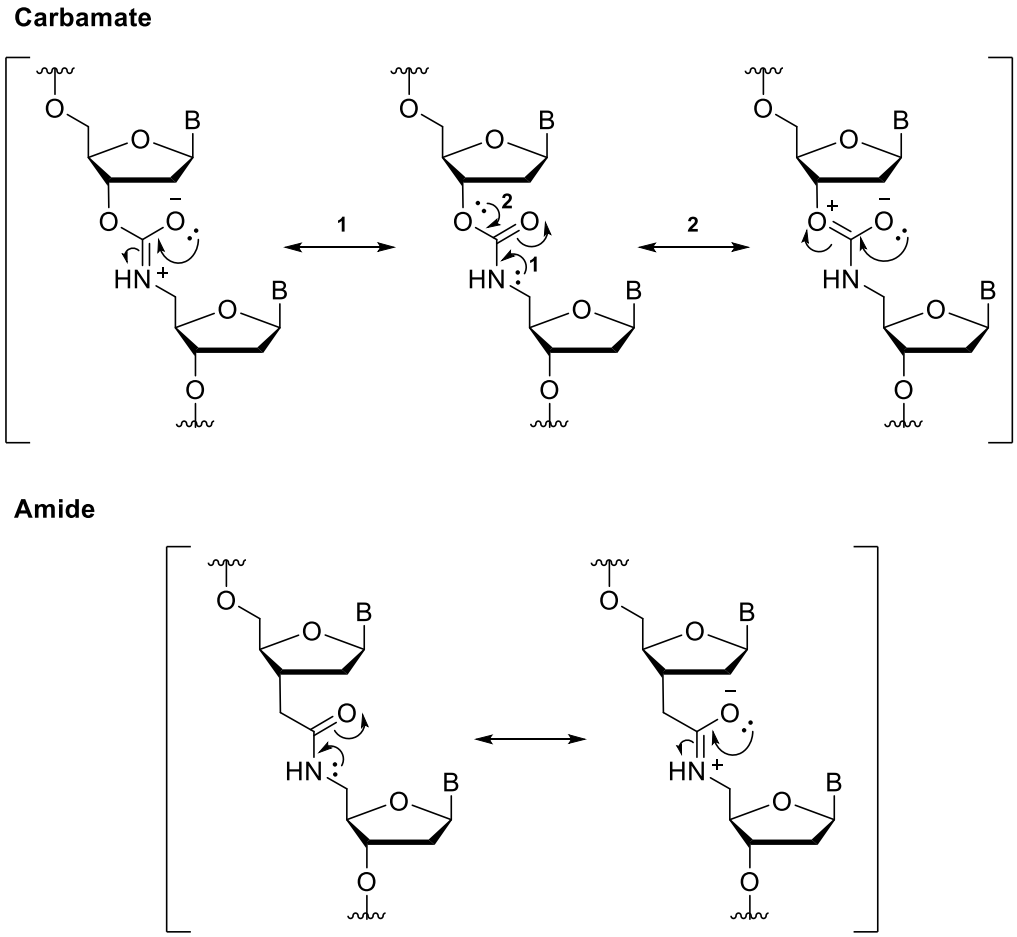



*Figure 3. 37 Comparison of resonance structures of carbamate and amide backbones. Additional electrons located on the oxygen atom in the CBM backbone allows for increased sp² character towards the 5'-sugar. This in turn reduces the degrees of rotational freedom and could contribute to strand invading structures.*

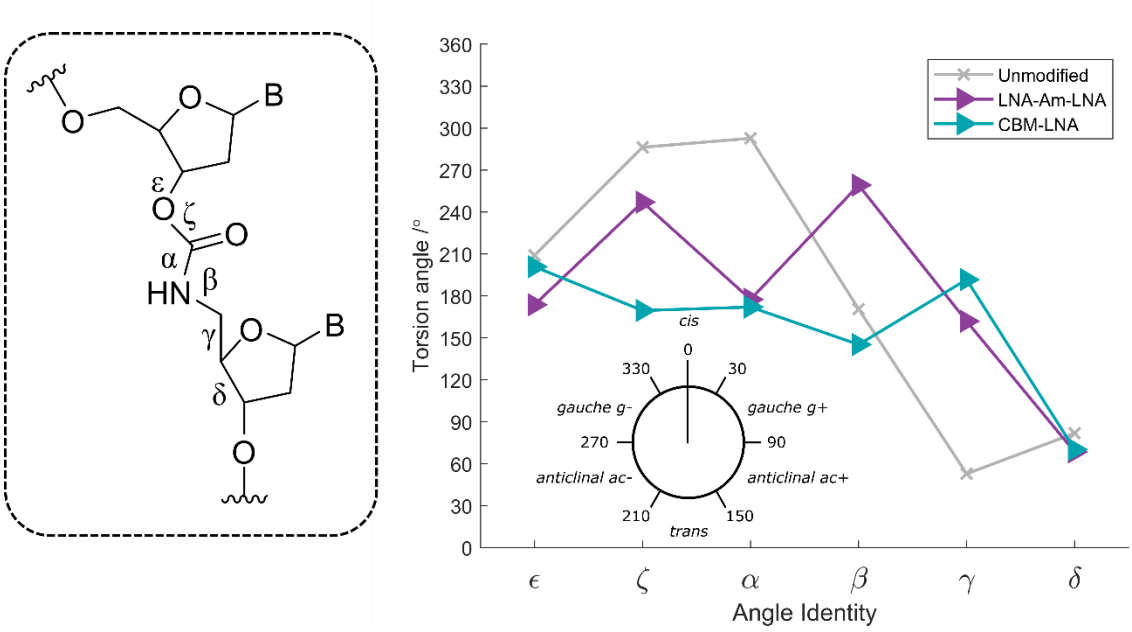



*Figure 3. 38 Torsion angle comparison between LNA-Am-LNA and CBM-LNA modified duplexes. Specific torsions are given by their respective symbol (ε→δ, right). Modified backbones are located at the 7A-t base pair step. Also included is the 7A-T phosphodiester step from the unmodified duplex. Compared to the amide, the CBM backbone sits in a linear conformation.*

Despite this, we propose that in solution CBM containing duplexes are probably in a standard A'-form structure. When the two complementary oligonucleotide strands are mixed in solution, there are two possible options for duplex formation. Either a canonical structure is formed where the CBM backbone experiences a strained conformation required to form the duplex, or a strand invaded dimer is formed with unstrained linear CBMs. The dimer is probably biologically rare, but maybe promoted by an artefact of crystallography. The nature of crystal formation is not fully understood, but once nucleated, initial crystals act as templates pulling amorphous material into the crystal lattice, extending in directions of the lattice plane. In order to be incorporated into the crystal, the incoming molecules must be structurally alike in order to fit into the lattice. If our solution drop contains a mixture of canonical duplex and dimer duplexes, but the dimer structure crystallises more readily, then it could mean that only these structures form our crystals. In addition to this, it is remarkable that making such a small chemical change in a duplex (a single O-atom for a C-atom) can have such a large effect on stability and on solid-state structure. The large structural effect is almost certainly due to the solid matrix and is unlikely to occur so readily in solution. However, in cases where strand invasion is required, placing the unstable modified backbone at a specific point in the duplex might be a good mechanism to achieve strand invasions.

## 3.9 Summary of LNA-amide and CBM-LNA modified XRD structures

The aim of this work was to investigate the structural properties of combined LNA-amide modifications and see if there was an identifiable structural relationship between these modifications and thermodynamic properties. We also aimed to compare LNA-amide backbones to structurally similar but thermodynamically different CBM-LNA modifications, to discover if there were any markedly different structural features between their duplexes. The wider scope of this work was to show that LNA-amides are suitable modifications for therapeutic applications and that they are well tolerated within an oligonucleotide duplex.

To do this we synthesised a set of DNA:RNA hybrids that contained either an amide, LNA-amide or LNA-Am-LNA modification. These were then subjected to sitting drop vapour diffusion to obtain crystals which were optimised with various screens to improve crystal quality. Using X-ray diffraction, models of each duplex were obtained between 2.5 to 2.8 Å. The structures were edited to include the relevant modifications and refined to an $R_{free}$ 0.20-0.26.

Comparative analysis between structures showed that addition of LNA had little effect on the torsion angles of the amide backbones, even when added to the 5'-side directly acting on the linkage. In general, the base pair parameters of the amide and LNA-amide structures remained similar to the unmodified duplex. However, addition of two LNAs flanking the amide backbone did change local base pair step parameters, but this was primarily caused by the consecutive addition of the LNAs, rather than an effect of LNAs being combined with amide backbones. When LNAs and amides are combined they possess the best qualities of each modification, without disrupting the function of each other. LNAs do not perturb the amide backbone or distort its structures, and likewise the amide backbone does not prevent the stabilising nature of LNAs.

Another aim of our work was to contrast the structure of an LNA-carbamate backbone with that of the LNA-amide. CBM-LNA crystals were obtained with ease but their structures could never be phased. Alterations to the molecular replacement model followed by manual processing, showed that the CBM containing duplex was not in a canonical form, but instead formed a dimer duplex with crossed invading strands. Although thought to be an artefact of crystallography, we conclude that reduced rotational flexibility at the 5'-end of the linkage due to increased $sp^2$ character increases the probability of linear conformations, promoting strand invasion. Compared to the amide, the CBM backbone might be more strained in the conformations required for A'-duplex formation and could explain why CBM containing duplexes are less stable than amide derivatives.

In summary, both amide and LNA modifications work together harmoniously due to the conformational flexibility of the amide backbone and linkage length. Although different applications require specific backbone properties, we suggest that for good duplex stability the backbone requires flexibility at the 5'-end of the linkage and should be a four-atom backbone if built around a carbonyl design. To prevent perturbation of the duplex the backbone should mimic a phosphodiester while being structurally different to prevent nuclease degradation in biological media such as cells. When combining backbone modifications with LNAs, the best combinations will result from structures which can conform to the needs of the bicyclic sugars without causing additional strain or having unnecessary linkage length. We conclude that LNA-Amide modifications are well tolerated within duplex

structures but that consecutive addition of several LNA may cause minor structural changes to the duplex.

# 4 Synthesis and translation of triazole modified circular RNA

## 4.1 Introduction and synthesis of circular RNAs

### *4.1.1 Introduction and biosynthesis of circRNA*

Circular RNAs (circRNAs) are an emerging class of biomolecules. In eukaryotes, genomic DNA is transcribed to a precursor messenger RNA (pre-mRNA) to create a sequence that contains protein coding exons spaced between non-coding introns. Prior to protein synthesis, introns are removed ("splicing") to create a linear, coding mature mRNA. Alternative splice patterns allow for various proteins to be produced from a single genome sequence. The spliced linear RNA is decorated with post-transcriptional modifications, which include a 5'-cap and a poly-A tail. Alternatively, this process can also produce circular RNAs (circRNA), through covalent bond formation between an upstream 5'-splice site and a downstream 3'-splice site ("backsplicing"). These circRNA have distinct differences compared to linear mRNA; their cyclic form confers strong resistance to exonucleases, such as RNase R, and it does not contain a 5'-cap or 3'-poly-A tail.

CircRNAs are important biomolecules that are naturally forming in nature. The Sex Determining Region Y (*Sry*) is a conserved gene found in mammals which is partly responsible for determining sex. During a short window of mouse embryonic development (days 10–13), a linear mRNA is highly expressed to induce development of the testis. After this period, the function to express protein from the mRNA decreases but the gene expression remains constant. B. Capel *et.al.*[155] found that this phenomenon is due to conversion of the RNA from the linear to the circular form, resulting in the downregulation of the *Sry* gene expression. In this study, it was suggested that circRNAs make up a small percentage of the total RNA, and general scientific opinion at this time was that circRNAs have limited function. In 2012, J. Salzman *et. al.*[156] developed a new computational method based on RNA-sequencing data that allowed the detection of circRNA. They showed that hundreds of human genes are significantly expressed in the form of circRNA, and that in many cases they are over expressed compared to linear isoforms.

With improved detection methods and evidence for high abundance of circRNAs, more research focussed on establishing how circRNAs are formed. In canonical splicing of a linear RNA, introns can be removed by self-splicing mechanisms (Figure 4. 1).[157] To initiate cleavage, the intron (I1) folds up onto itself, bringing an adenosine residue within the intron towards the end of the 5'-exon (E1). Cleavage of the 5'-exon (E1) is then facilitated by attack of an adenosine 2'-hydroxyl, which cleaves the phosphodiester bond breaking the oligonucleotide. The intron then folds back on itself to form a lariat loop (Figure 7. 9). Attack of the 3'-hydroxyl (E1) on the 5'-exon (E2), removes the intron (I1) and the 5'-exon (E1) is joined to the 3'-exon (E2).[158]

Circular RNAs use the same cellular machinery as linear RNAs for splicing (Figure 4. 1).[159] However, instead of splicing the 5'-exon (E1) to the 3'-exon (E2), the 5'-exon can self-splice (E2-E2) or backsplice to the other upstream exon (E1-E2). This backsplicing mechanism can be used to create a variety of different circular products containing single exons, introns or whole genes. To induce backsplicing, complementary sections within the intron come together to hold the RNA in a non-covalent loop.[160,161] Using the same methods that splice a linear RNA, the intron is then removed from the oligonucleotide and a covalent bond is formed, creating the circular RNA.

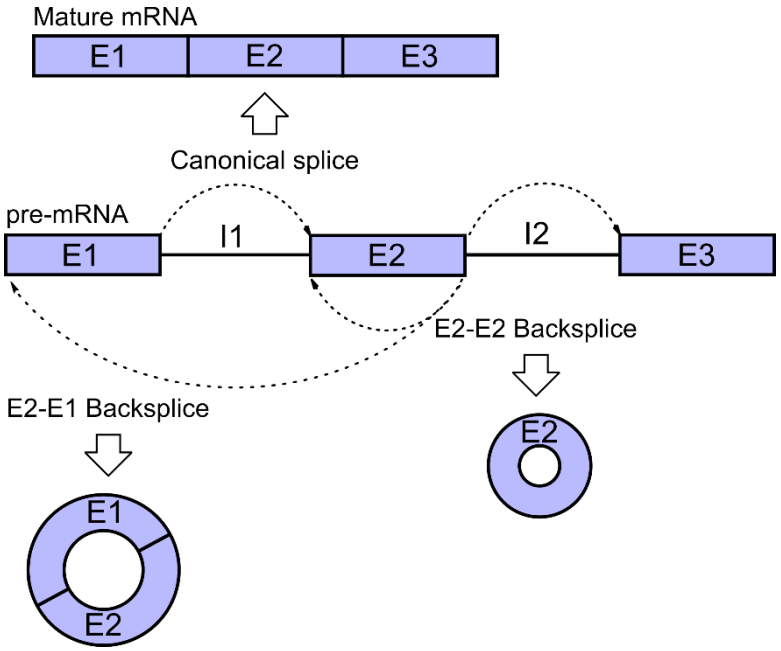



*Figure 4. 1 In vivo synthesis of mRNA in eukaryotes through canonical splicing of a linear mRNA or backsplicing to form circRNA. Introns (I) are non-coding regions in the pre-mRNA which are interspaced between the coding exon regions (E). Backsplicing can occur between various positions creating different circRNAs.*

A. Ryback-Wolf *et. al.*[162] investigated the patterns of circRNA sequences from brain tissues to confirm the mechanism of circRNA formation *in vitro*. Like previous reports, they found that circRNAs are highly abundant[163,164] but also that they are enriched within the brain.[165] A sequence analysis showed that complementary regions flanking the circRNA exon are highly conserved. Since it had been previously proposed that backsplicing is induced by base pairing between neighbouring introns, conserved sequences adjacent to circRNA suggested that the mechanism could be more ubiquitous than previously thought. If complementary sequences within the flanking introns do promote circRNA formation, then editing this region to reduce the stability of the Watson-Crick base pairs should alter circRNA expression. Adenosine Deaminases Acting on RNA (ADAR) is an enzyme that converts adenosine to inosine (A→I, Figure 4. 2A), which is a base analogue of adenosine with reduced duplex stability.[166] By comparing ADAR and circRNA concentrations, it was found that there was an inverse correlation between ADAR levels and circRNA formation. As ADAR accumulates, editing of the intron becomes more frequent, destabilising the complementary intron regions and thus reducing circRNA formation. Therefore, this suggests that the complementary intron regions and the RNA secondary structure are essential for circRNA formation.[167]

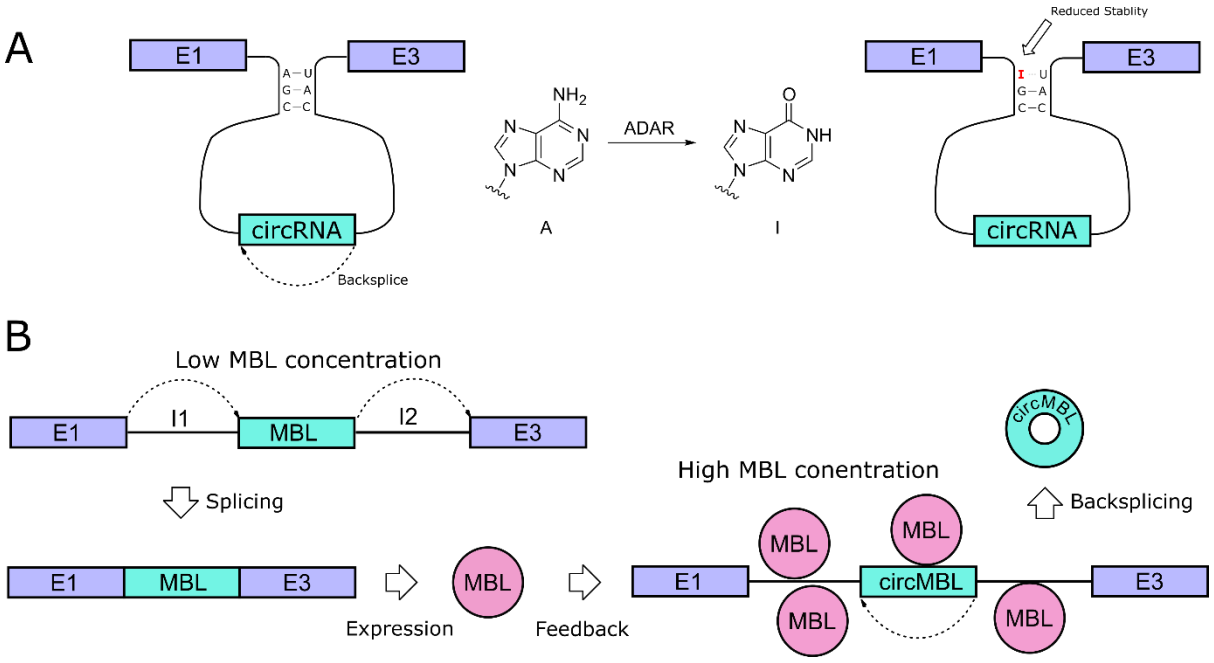



*Figure 4. 2 A) Mechanism of circRNA formation is induced by base pairing between the two introns. ADAR can down-regulate circRNA formation by converting adenosine (A) to inosine (I). The resulting complementary region has a reduced duplex stability and is less likely to promote circRNA formation, B) Up regulation of circRNA formation by a negative feedback loop. As the level of MBL protein expression from a linear RNA increases, MBL accumulates and inhibits the conical splicing of its own linear pre-mRNA, promoting the production of circRNA (Adapted from S. Barrett et. al.*[168]*)*

Muscleblind protein (MBL) is required by *Drosophila* for the development of muscles and photoreceptor cells in the eye. MBL protein is of interest because it can regulate its own expression.[168] In a cellular environment with low levels of MBL protein, a linear mRNA is expressed and spliced for production of more MBL proteins. As the concentration of MBL increases, the protein localises to the splice sites of its own pre-mRNA template, influencing the shape of the intron and hence inducing splicing of the circRNA isoform (Figure 4. 2B).[169] In this case the circRNA is unsuitable for MBL translation, allowing the MBL levels to be regulated by a negative feedback loop.

### *4.1.2 In vitro synthesis of circRNA*

CircRNA can be generated *in vitro* by using group 1 introns which are able to autocatalytically produce circRNA. However it is not feasible to synthesise long RNA greater than 100 nt *via* chemical synthesis.[170,171] Therefore, long RNA must be produced enzymatically. To do this, the target sequence is placed within flanking group 1 introns and synthesised as DNA. The DNA sequence is then cloned into a DNA plasmid (Figure 4. 3A).[172,173] Following transcription to create the long RNA, the complementary sections of the group 1 introns form a duplex to aid the formation of the circRNA as previously shown (Figure 4. 2A). Although this method is efficient and produces RNA with homogenous splice locations, it has some significant drawbacks. *In vitro* transcription requires bacterial cell culture, cloning, plasmid extraction and purification. This is labour intensive and requires specialised facilities. This cellular machinery is also limited by its ability to accept or incorporate RNA modifications that would be required to generate a therapeutic circRNA template. Finally, the cell culture process is difficult to scale up to produce enough RNA for clinical supply.

CircRNA can also be synthesised without the need for cellular machinery by using commercially available T4 DNA ligase (Figure 4. 3B). An RNA with a 5'-phosphate modification is produced via solid phase oligonucleotide synthesis. A DNA splint is added to the RNA which has a complementary sequence surrounding the ligation site. Held in proximity by the DNA splint, T4 DNA ligase ligates the 5’ and 3’-ends to produce a circular RNA.[174] An advantage of this method is that RNAs can be highly decorated with site specific modifications. However, oligonucleotide length is limited by coupling yields during solid phase synthesis. Compared to the plasmid-based method, solid-phase synthesis of RNA provides a fast and economical procedure to produce small circRNAs (< 90 nt), especially those that are difficult to cyclise.[175]

Although it can be scaled up more easily than cell-dependent methods, the amount of ligase required per mole of RNA is still high, which is expensive and problematic.

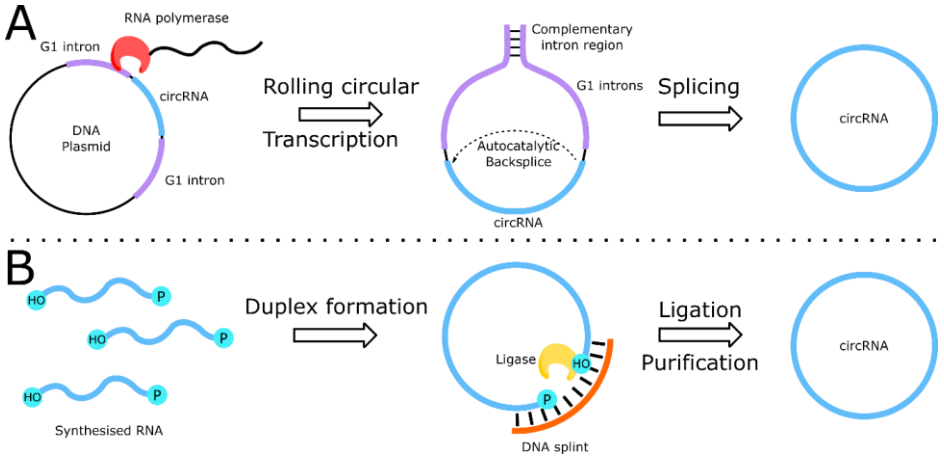



*Figure 4. 3 A) In vitro synthesis of circRNA from a DNA plasmid. The circRNA coding region is placed inside a group 1 autocatalytic self-splicing intron. After transcription the circRNA backsplices out of the template to produce circRNA. B) Cell free synthesis of circRNA using T4 DNA ligase and complementary DNA splints. 5'-phosphates are added during chemical synthesis and ligated to a 3'-OH using templated ligation.*

## 4.2 RNA translation and translation of circRNA

### *4.2.1 Translation initiation*

Eukaryotic translation is controlled by the 80S ribosome complex and up to nine eukaryotic initiation factors (eIFx, where x denotes a number, Figure 4. 4A).[176] In a simplified mechanism, translation begins with the formation of the 43S complex containing the 40S ribosomal subunit, eIF1, eIF2, eIF3, eIF5, GTP and initiation anticodon tRNA.[177] The complex is then guided to the 5'-cap of the mRNA which is bound to eIF4. For 5'-cap dependent initiation, the 43s complex then scans the 5' untranslated region (5'-UTR) until reaching the ATG start codon. To aid this recognition step, the eukaryotic mRNA also contains a conserved sequence directly upstream of the AUG codon called the Kozak sequence. This sequence was initially discovered by site-directed mutagenesis, with the consensus motif being 5'-gccRccAUGG-3' (where R is a purine, capitals are conserved bases, underlined is the start codon and lower cases are the most common bases in that position). The sequence helps locate the ribosome to the first codon for translation.[178,179] Upon GTP hydrolysis, the initiation factors are removed, and the 60S ribosomal subunit binds to the initiation complex to create the 80S eukaryotic ribosome.[176]

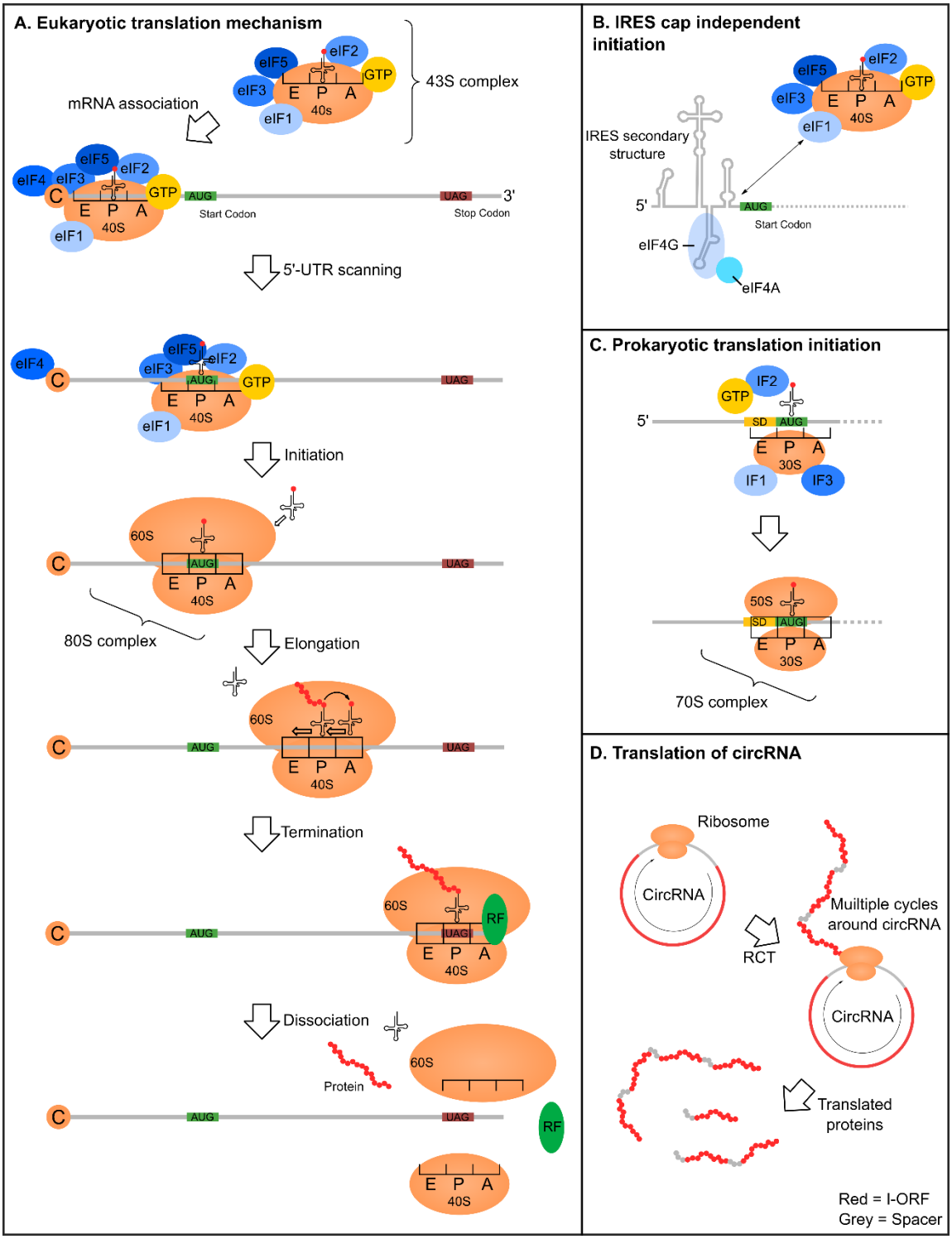



*Figure 4. 4 A) Eukaryotic translation mechanism of mRNA. 43S complex binds to the eIF4 via 5'-cap mediated initiation, triggering a scanning mechanism to locate the AUG start codon. tRNA is loaded into the P-site and elongation occurs by peptide bond transfer to the incoming tRNAs in parallel with ribosome translocation. Termination is mediated by interaction with the UAG stop codon followed by binding of the release factor (RF). The protein is released, and the ribosome dissociates for another round of translation. B) Alternative initiation mechanism using an internal ribosome entry site (IRES). eIF4G binds to secondary structures in the 5'-UTR to direct the recruitment of the 43S complex. Initiation is promoted without the need for a 5'-cap. C) Initiation of translation in prokaryotes. Fewer initiation factors (IFs) are required compared to eukaryotic translation and the ribosome is recruited directly to the AUG start codon using the Shine Dalgarno (SD) sequence. D) Translation of circRNA. The ribosome is recruited upstream of the AUG start codon. Successive rounds of translation create protein concatemers by rolling circular translation (RCT). Spacers are translated between the Infinite open reading frame (I-ORF) and contain non-specific codons for viable amino acids and initiation sequences (A, B adapted from R. Jackson et. al.[176]).*

Translation initiation can also occur by 5'-cap independent translation (Figure 4. 4B) with the use of an internal ribosome entry site (IRES). An IRES is part of the mRNA sequence located in the 5'-UTR that has specific secondary structure to promote initiation factor and ribosome binding without a 5'-cap. First discovered in the poliovirus[180] (PV) and encephalomyocarditis virus[181] (EMCV), cap independent translation has been routinely used for *in vitro* applications and also by non-viral, cellular mRNAs.[182]

In an IRES mechanism, eIF4G binds directly to the secondary structure of the mRNA 5'-UTR and can be enhanced by complexation with eIF4A.[183,184] The 43S ribosome complex can then bind directly to the IRES without performing the scanning mechanism and locates the initiation AUG codon. Comparing between different viruses and cellular mRNA, several mechanisms have been described that use various cofactors and processes.[185] Despite this, a conserved feature is that each mechanism performs mRNA recognition by scanning secondary structures in the 5'-UTR.[176]

### *4.2.2 Translation elongation*

The 80s ribosome complex contains three active sites: A, P and E. The A-site is the ribosomal entry site that accepts an incoming tRNA loaded with an amino acid for protein synthesis. The P-site is occupied by the peptidyl tRNA that contains the growing peptide chain. The E-site is the exit site for the tRNA after it has delivered the amino acid. During initiation, the initiator tRNA (tRNA-Met) is held in the P-site. Protein elongation then occurs by entry of a new tRNA into the A-site.[186] The incoming tRNA is selected by matching sequences of the tRNA anticodon to the mRNA codon. Incorrect anticodon sequences cause the tRNA to be rejected and released by the ribosome. This way, the codon sequence dictates the tRNA entry and therefore ensures that the correct amino acid is added.[187] Once the correct tRNA has entered the A-site, the peptide chain bound to the tRNA in the P-site is transferred to the new amino acid, attached to the tRNA in the A-site. At the same time, the ribosome complex translocates along the mRNA by one codon. This moves the P-site tRNA into the E-site and A-site tRNA into the P-site. By moving along the mRNA sequence, amino acids can be sequentially added to the peptide chain.

### *4.2.3 Translation termination*

Termination is promoted by the presence of a stop codon (UAG, UAA or UGA) at the end of the coding region. Unlike previous codons, stop codons do not encode for a corresponding

tRNA. In the absence of peptide chain extension, a release factor (RF) is recruited to the ribosome mRNA complex.[188] This catalyses the cleavage of the peptide bond between the peptide and the tRNA. The protein is then released, and the ribosome dissociates into its subunits, ready for another round of translation.

### *4.2.4 Prokaryotic translation*

Prokaryotic translation shares many similarities to eukaryotic translation but is in general much simpler. Made of the 30S and 50S subdomains, the prokaryotic 70S ribosome is smaller in size than the eukaryotic form, but still contains the A, P and E-sites. Translation initiation is mediated with three initiation factors (IF1, IF2 and IF3) which locate the tRNA and the 30S subdomain to the AUG start codon (Figure 4. 4C).[189] Having no 5'-cap means that initiation is dependent on localisation to the AUG codon with the aid of a Shine Dalgarno (SD) consensus sequence.[190] Like the Kozak sequence, the SD sequence is used to guide the ribosome to the correct codon to initiate translation. It is typically 6 – 8 nt in length but is less conserved than the Kozak sequence.[191] The 50S ribosomal subunit binds to the SD sequence, and the protein is elongated using a similar mechanism to eukaryotic translation. Termination is promoted by the presence of a stop codon and release factor (RF), that removes the protein and dissociates the ribosome for another round of translation. Due to the simplicity and robust nature of prokaryotic translation, it has become a powerful research tool for cell free protein expression. Research has allowed for the development of cell-free translation kits that can perform *in vitro* expression without the need for cell culture, transformation or cell lysis. Thanks to the simplicity of prokaryotic translation, it is possible to purify every enzyme required for translation using recombinant plasmid expression and combine them with the desired mRNA oligonucleotide. Commercialisation of work from Y. Shimizu *et. al.*[192] by New England Bioscience allows for cell free purified expression kits to be a simple and standard protocol. The kits are designed to contain the minimum requirements needed to express a protein from an mRNA. To develop a cell-free eukaryotic kit, the whole cell is required to be lysed and collected. Compared to the purified prokaryotic system, eukaryotic kits are complex and contain many undesired proteins which are irrelevant to translation. However, the eukaryotic kit is still simpler and more convenient than cell-based assays.

### *4.2.5 Translation of circRNA*

CircRNAs containing the sequence or structural features required for ribosome recruitment can be translated into proteins (Figure 4. 4D). Because of the cyclic nature of circRNAs, ribosomes can translocate around the circRNA multiple times. This process is called rolling circular translation (RCT), which allows a concatemer of repeat protein units to be produced, facilitated by the infinite open reading frame (I-ORF).[193] To create a singular protein unit, stop codons can be introduced into the template, or in some cases self-cleaving peptides can be used.[194,195] However, if the ribosome dissociates from the circRNA, it must reinitiate translation, which has been shown to reduce protein yield.[196]

Translation of circRNA using RCT was first proposed by R. Perriman and M. Ares[197] utilising cloning and bacterial cell culture to produce their circRNA. They encoded a green fluorescent protein (GFP), which expression generates a fluorescent signal for detection. Although they managed to confirm GFP synthesis by PAGE analysis, the concatemer species lacked the fluorescent properties shown by the monomeric GFP. Instead of using a fluorescent protein such as GFP, it is also possible to use immunoblotting for protein detection. The FLAG epitope (DYKDDDDK where D is aspartic acid; Y is tyrosine; and K is lysine) is a synthetic peptide that has been developed against antibodies in order to perform analysis or purification. By matching the FLAG antigen with a corresponding antibody, proteins can be specifically visualised by western blot or purified by immunoprecipitation. Work by N. Abe *et. al.*[198] used prokaryotic systems to show that proteins can be expressed from circRNA, but instead of using GFP as a reporter protein, they encoded the circRNA with a short FLAG tag sequence. From this template, they showed that it could be successfully translated in a cell-free protein expression kit, and that the I-ORF can create RCT products up to 250 kDa in size. This study was expanded to a cell-free eukaryotic system by editing the circRNA to contain a Kozak sequence. Due to size limitations, they did not include an IRES for many of their templates, which suggests that it is an additional but non-essential feature for circRNA expression. Translation initiation of circRNA has been observed to be the rate limiting step, and compared to linear isoforms, initiation from circRNA is much slower. Although the exact reason is unclear, comments have been made suggesting that the topography of the circRNA is a substantial influence, along with the coupled nature of prokaryotic transcription-translation and the removal of structural features from the 5'-UTR.[197,199] Despite this, circRNA are able

to express the FLAG peptide from both a cell-free eukaryotic system and HeLa cells.[200] However, the yields were substantially lower compared to prokaryotic systems, and the kit required considerably more circRNA than the previous prokaryotic experiments.

To date, few studies have tried to synthesised circular RNAs with modified backbones[16] and at the time of this study no circular RNAs have been synthesised without using enzymatic methods. If modified circular mRNAs could be synthesised using only synthetic protocols, then it would allow for a scalable supply of circRNA templates for clinical research. Due to its biophysical properties and ability to be formed in aqueous conditions, modified backbones such as triazole linkages present interesting candidates to be incorporated within circular RNAs. However, to do this, research must first focus on establishing triazoles as suitable modifications for translation and developing synthetic protocols for the formation of triazole containing circular RNAs.

## 4.3 Triazole backbones

Triazoles can be formed by using copper catalysed azide alkyne cycloadditon CuAAC 'click' chemistry. With its ability to be performed in aqueous conditions, click reactions provide a simple procedure for many applications such as the attachment of fluorescent dyes, antisense modification or DNA library synthesis.[95,201] In the literature there are several different triazole backbones that have been synthesised. 1,4-triazoles can be formed using a copper catalyst. Isobe *et. al*.[202] synthesised a 1,4-triazole linkage with five bonds (Figure 4. 5A). Thermal analysis of this backbone showed that the linkage was highly destabilising when incorporated into oligonucleotides (-8 °C/mod), most likely due to the positioning of the triazole towards the 5'-end of the backbone. However, preliminary molecular modelling showed that the backbone caused minimal disruption to the duplex structure and that the charge neutral linkage may have other useful properties. A. M. Varizhuk *et. al.*[93] developed an alternative six bond linkage (Figure 4. 5B) which proved to be advantageous towards improving thermal stability. Unlike the Isobe linkage (Figure 4. 5A), the triazole was positioned at the 3'-end of the backbone. The combination of linkage extension and 1,4-triazole positioning was less destabilising within the duplex and caused minor reductions to stability (-1.4 °C/mod).

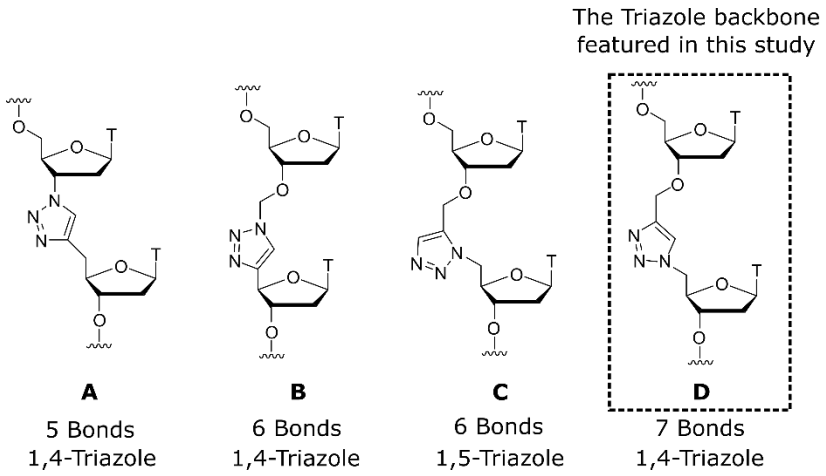



*Figure 4. 5 Structures of triazole backbones previously published A) synthesised by Isobe et.al.[202], B) Synthesised by A. M. Varizhuk et. al.[93], C) Synthesised by Y. Baker et. al.[203], and D) Synthesised by R. Lucas et. al.[34], 1,4-triazoles can be synthesised using a copper catalyst and 1,5-triazole using ruthenium. Triazole (D) is the backbone featured in this study and from here on is referred to as the 'triazole' backbone (Tz).*

1,5-Triazoles can be synthesised by replacing the copper catalyst for Ruthenium, which selectively promotes the 1,5-constitutional isomer. In a recent study by Y. Baker *et. al*.[203] the biophysical properties of 1,4-triazoles (Figure 4. 5A and B) were compared to a new 1,5-triazole (Figure 4. 5C). In this study it was found that the 1,4-triazole (Figure 4. 5B) had the best properties for RNA ASO targets, but when hybridise to DNA oligonucleotides, the 1,5-triazole (Figure 4. 5C) was less disruptive to the duplex structure and synthetically simpler to produce.

This work features a seven atom, 1,4-triazole linkage (Figure 4. 5D) first synthesised by R. Lucas *et. al*.[34] Although the backbone is slightly destabilising (-4 °C/mod) it has good biocompatibility,[128] can be easily synthesised, and has a commercially available precursor alkyne phosphoramidite. From this point onwards, this 1,4-triazole linkage (Figure 4. 5D) is referred to as the 'triazole' (Tz) linkage.

Triazoles can be incorporated into oligonucleotide sequences by DNA splint mediated ligation which enhances the click reaction (Figure 4. 6). One oligonucleotide is modified with a 5'-azide and a second is modified with a 3'-alkyne. The oligonucleotides are then held in proximity using a DNA splint, which allows for the $Cu^{I}$ click reaction to occur, ligating the two oligonucleotides together. Purification of the resultant oligonucleotide produces a single stranded sequence containing a modified triazole linkage. By ligating multiple oligonucleotide sequences together, this method could be used to produce long synthetic RNAs, potentially up to 500 nt in length.[204]

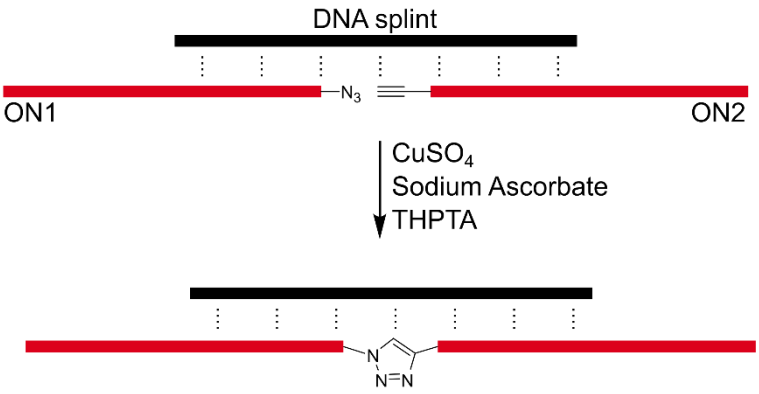



*Figure 4. 6 Oligonucleotide ligation using the CuAAC click reaction to form a triazole backbone. Oligonucleotides ON1 and ON2 are held in proximity with a DNA splint (black).*

Triazole backbones are resistant to nuclease degradation and have a neutral charge. This facilitates cell penetration and increases biological half-lives of the modified oligonucleotides.[95,96,119] As mentioned, triazole linkages in general cause duplex destabilisation and therefore can reduce affinity for target sequences.[205] To mitigate these effects, triazoles can be combined with LNA to improve duplex stability. In recent work by P. Kumar *et. al.*[95], it was shown that if LNA is added to the 3'-position relative to the triazole backbone, then duplex stability can be restored.

Triazoles are good biological mimics and can be processed by enzymes that are active within the central dogma, including read through during PCR. In a study assessing DNA replication fidelity of a triazole-containing oligonucleotide, A. Shivalingam and A. Tyburn *et. al.*[128] synthesised five different modified backbones to analyse their read through speed and template fidelity using qPCR and illumina sequencing. They found that the triazole backbones are processed slower than canonical phosphodiesters, however the best performing triazole backbones can produce viable DNA templates with few mutations during replication.

Triazoles have also been shown to be good templates for RNA polymerase and are compatible with transcription in both *E. coil* and eukaryotic systems.[206] Transcription in eukaryotes was examined by incorporating a triazole backbone into a DNA plasmid that was then transfected into human cells. After transfection the cellular machinery was able to recognise the modified template and transcribe through the triazole to produce the encoded fluorescent mCherry protein.[207] Having proven that it is possible to transcribe the triazole-modified DNA, this study was expanded to further demonstrate that a whole *iLOV* gene containing eight triazole

linkages can be efficiently transcribed.[208] These studies suggest that click ligation to generate triazole containing plasmids could potentially provide an alternative method for cloning instead of using restriction enzymes.

Translation of triazole backbones is an area of research that is not yet well studied. In a preliminary study, T. Fujino *et. al.*[209] synthesised a trinucleotide phosphoramidite containing two triazole backbones and incorporated this modified trimer into a 13 mer RNA with a total of 10 native phosphodiesters and two triazoles. Using a cell-free prokaryotic system, the 13 mer RNA was successfully translated providing evidence that it is possible to transcribe triazole-modified RNA by a prokaryotic ribosome. However, analysis of the resultant protein suggested that significant read-through errors occurred during the translation. As previously discussed, prokaryotic translation of mRNA requires features such as the Shine Dalgarno sequence in order to direct the ribosome to the correct location on the RNA. Since their template did not contain this sequence and is very short, it is difficult to verify if the read through error is the result of the triazole addition or template design. mRNA gene sequences are often much longer and extend far beyond the footprint of the ribosome, it may be the case that a short sequence is less suitable for translation. In summary of this work, it provides good evidence that triazoles can be translated by prokaryotic cellular machinery, but it does not indicate the translation efficiency, nor if a peptide or a whole protein can be expressed. In the context of mRNA therapeutics, it is still unknown if the highly complex mechanism of eukaryotic translation would be able to accept the triazole-modified RNA. This would be an important factor to determine if the triazole-modified mRNA were ever to reach the clinic.

Within our own group, Ashley Tsai (MPhil) started to investigate the triazole backbones within circRNA to establish an optimised template for translation.[210] She synthesised a set of triazole-containing circRNAs (Tz-circRNAs) and designed them to contain the triazole backbone within different codons with various GC content. From her results she found that Tz-circRNA expressed less peptide than the unmodified phosphodiester control in prokaryotic systems, and that the triazole linkage must be placed within a codon containing at least one GC base pair. It was also discovered that the triazole was best accommodated when placed adjacent to the final base in the codon surrounding the wobble position between the second and third nucleotides. In her study it appeared that there was a relationship between duplex stability and triazole translation efficiency. The exact reason was unclear, but it was suggested that

destabilisation of the mRNA interaction with the tRNA and ribosome could be a contributing factor. Based on this, if the duplex stability of the Tz-circRNA could be restored with additional chemical modifications such as LNA, there is a potential to improve translation of the triazole-modified RNA.

## 4.4 Aims and objectives of triazole circRNA

The aim of this chapter is to provide evidence that triazole backbones can be translated by prokaryotic ribosomes and produce peptides longer than three amino acids. This will prove that proteins can be translated from templates synthesised by fully chemical means, which is advantageous for large scale protein expression but also removes the need for or bacterial cell culture or T4 DNA ligase. The triazole backbone has been chosen due to its good biophysical properties, combined with fast synthesis in aqueous conditions and biocompatibility. Triazole modifications will be incorporated into the circRNAs encoding a FLAG tag peptide. Building upon the work started by Ashley Tsai, modified templates will consist of triazole backbones placed next to RNA or LNA sugars (Figure 4. 7, RNA-Tz and LNA-Tz). If duplex stability is a critical factor, translation yields compared to unmodified controls should be restored with the addition of LNA. After confirming that triazole backbones can be expressed *in vitro*, template designs will be expanded to include multiple triazoles per circRNA, develop improved protocols to synthesise azide-modified oligonucleotides using H-phosphonate monomers and quantify yields of peptides expressed from triazole-containing RNA templates.

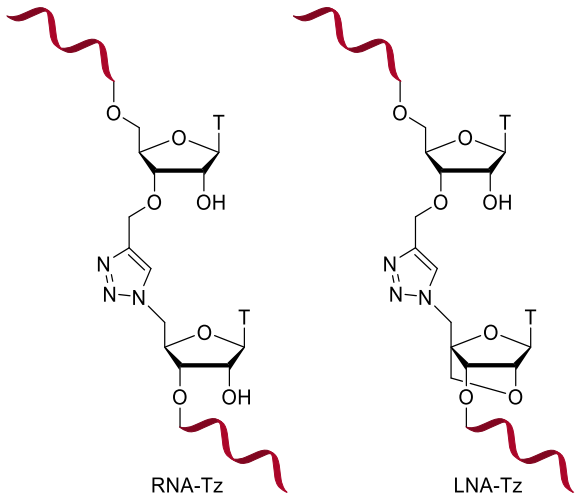



*Figure 4. 7 Structures of triazole backbones with adjacent RNA or LNA sugars. The linkage is a seven-bond backbone with a 1,4-isomeric ring structure. The RNA-Tz backbone is adjacent to two RNA sugars and the LNA-Tz linkage contains the duplex stabilising 3'-LNA modification.*

To achieve these aims, RNAs will be synthesised by solid phase synthesis with either a 3'-alkyne or 5'-azide attached to a RNA or LNA sugar (Figure 4. 8). The triazole-modified circRNA can then be formed by DNA splint-mediated ligation and subsequently purified by gel electrophoresis. Eventually circRNAs will be synthesised using two click ligations, but initial templates will be produced by a combination of click and T4 DNA ligase ligations. By producing FLAG peptides from the circRNA, immunoblotting with corresponding antibodies will be used to examine if the Tz-circRNA is a suitable template for translation and determine if LNA improves performance. If indeed a triazole is biocompatible for long peptide translation, it will mean that triazole linkages can be implemented at any stage of the central dogma, potentially enabling its use in gene therapy, anticancer peptides or diagnostics.

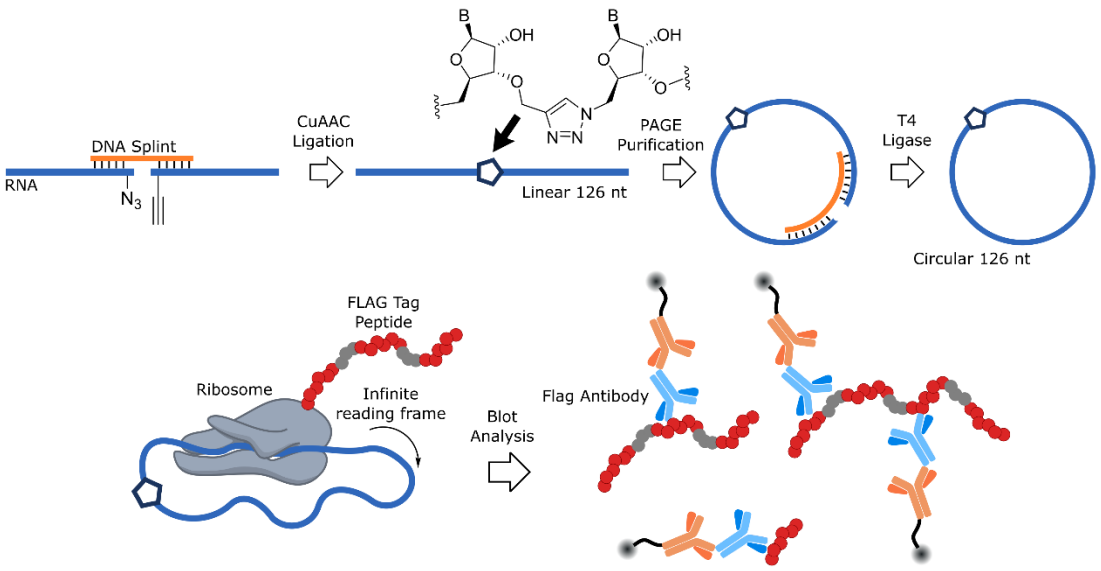



*Figure 4. 8 Overview of experimental procedure showing the synthesis of a triazole-modified circular RNA template, followed by in vitro rolling circular translation of the infinite open reading frame and peptide detection using antibody-based blot immunoassay.*

## 4.5 RNA template design and oligonucleotide synthesis

### *4.5.1 CircRNA template design*

Due to the simplicity of prokaryotic translation, Tz-circRNA expression was evaluated in a cell-free system before moving to cell or eukaryotic assays. A circRNA template (Figure 4. 9) was designed to encode three FLAG tag peptides interspaced between spacer sequences coding for non-specific amino acids excluding proline. To initiate FLAG tag expression, the Shine Dalgarno sequence was placed upstream of the start codon, and stop codons were removed to allow for continuous rolling circular translation (RCT). In total the circRNA template is 126 nucleotides long. To maximise the yield and purity, two RNAs (R1 and R2) were separately synthesised and then ligated at two ligation points (L1 and L2) using separate DNA splints. The L2 ligation position is identical for both unmodified and triazole templates whereas the L1 ligation site is moved by one base between the unmodified and triazole template. The reason for this was because the synthesis of the unmodified template had been successful in the preliminary assays performed by A. Tsai (M. Phil), so it was decided that template designs should not be altered unless required. The triazole performs best when placed within a codon with at least one GC base pair and is between the second and third positions within the codon. As a result, the ligation positions between these templates are different, but should not interfere with protein translation.

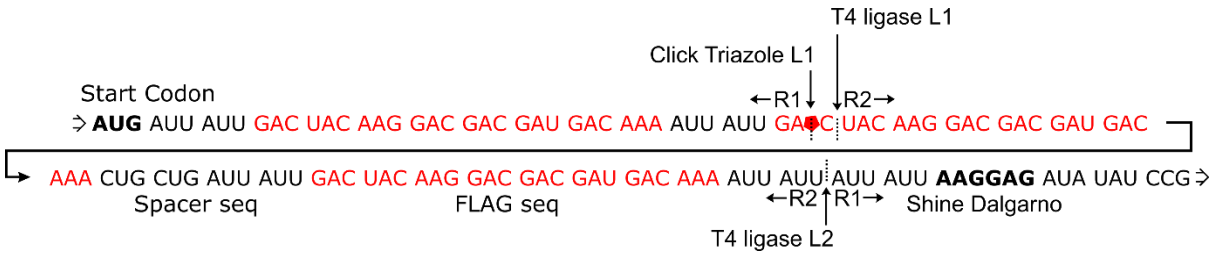



*Figure 4. 9 CircRNA sequence detailing positions of the FLAG tags (red), spacer sequences (black), start codon (bold AUG) and Shine Dalgarno (bold). The template is 126 nt and made from two oligonucleotides (R1 and R2). Ligation sites (L1 and L2) are formed using CuAAC ligation (triazole) or T4 DNA ligase to join the two RNAs to create the circular template.*

### *4.5.2 Synthesis of modified linear RNA precursors*

To create the unmodified circRNA control, linear RNA precursors must be modified with a 5'-phosphate to facilitate T4 DNA ligation (Figure 4. 10A). To do this, RNAs (ON4.1 and ON4.2, Table 7. 5) were synthesised using standard solid phase synthesis. In the final cycle, a modified phosphoramidite is used to install the phosphate group onto the 5'-end. The unmodified circRNA was then generated using templated T4 DNA ligation (section 4.6).

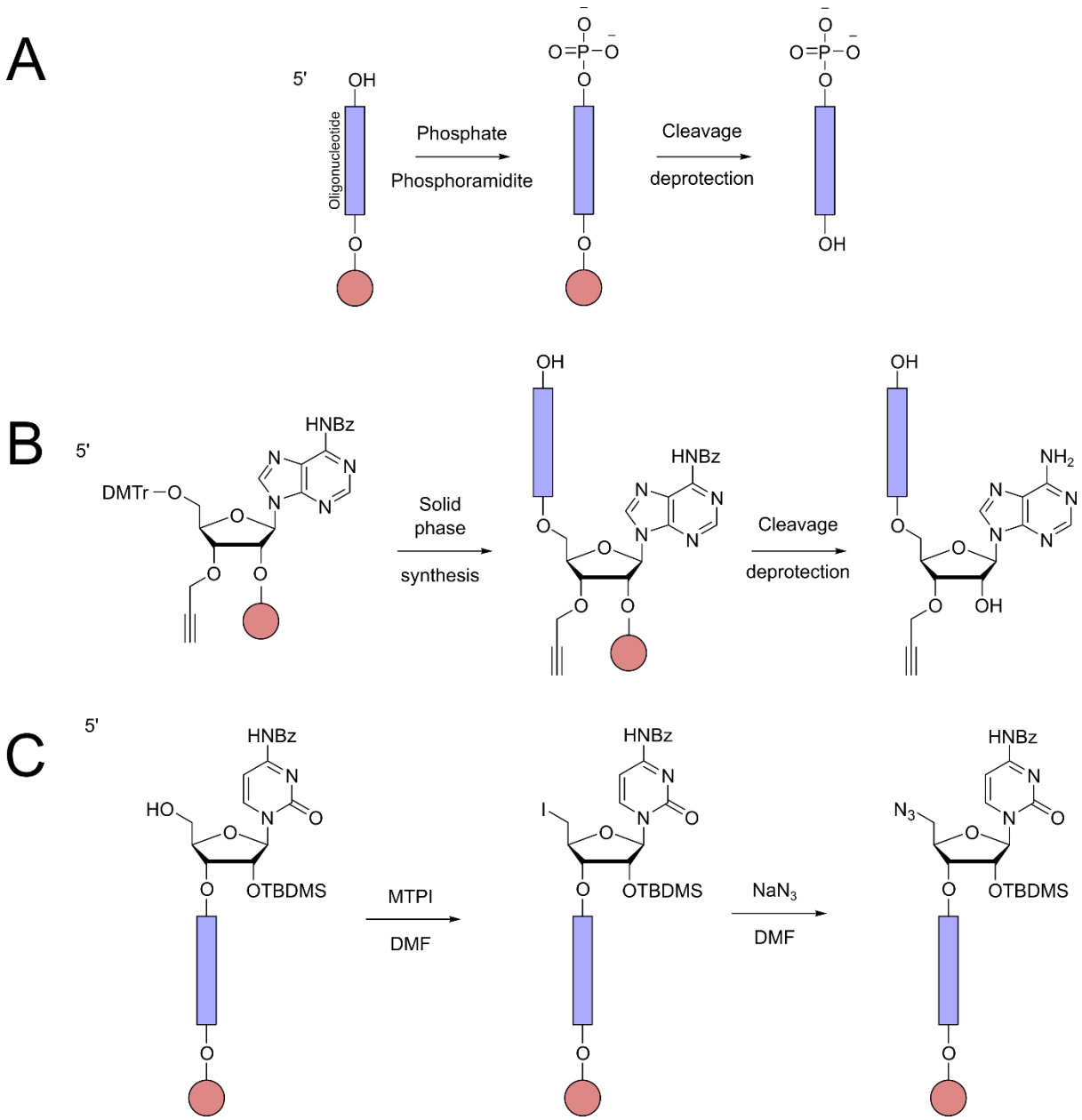



*Figure 4. 10 Synthesis of modified oligonucleotides, purple = oligonucleotide, red = solid support resin, A) Synthesis of 5'-phosphate modified oligonucleotides for unmodified circRNA. Phosphates are added during synthesis using a phosphate phosphoramidite, B) Synthesis of 3'-alkyne modified oligonucleotides for Tz-circRNA. Commercially modified alkyne monomer is connected to the resin via the 2'-position, followed by standard solid phase synthesis, C) Synthesis of 5'-azide oligonucleotides using on resin synthesis and conversion from iodine activation, MTPI = methyltriphenoxyphosphonium iodide.*

Like the unmodified control, Tz-circRNA precursors were synthesised using standard solid phase synthesis. To create the triazole linkage, one oligonucleotide was modified with a 3'-alkyne (ON4.6), while the other was modified with a 5'-azide (ON4.7, ON4.8). To install the 3'-alkyne, a commercially available resin-bound monomer was used (Figure 4. 10B). Because the monomers 3'-OH has been converted to the alkyne, this nucleotide is connected to the resin

by the 2'- position. Other than the change in linkage position, the standard cycle of synthesis was used followed by the same cleavage and deprotection conditions.

Synthesis of 5'-azide modified oligonucleotides requires multiple steps to convert the 5'-hydroxyl to an azide. First an oligonucleotide is synthesised using standard solid phase synthesis. While still on resin, the oligonucleotide is then reacted with Methyltriphenoxyphosphonium iodide (MTPI) which activates the 5'-position with an electron withdrawing and orbital diffuse iodine (Figure 4. 10C, Figure 4. 11). Being a good leaving group, iodine allows reaction with $NaN_3$, converting the 5'-iodine to the 5'-azide needed for click chemistry.

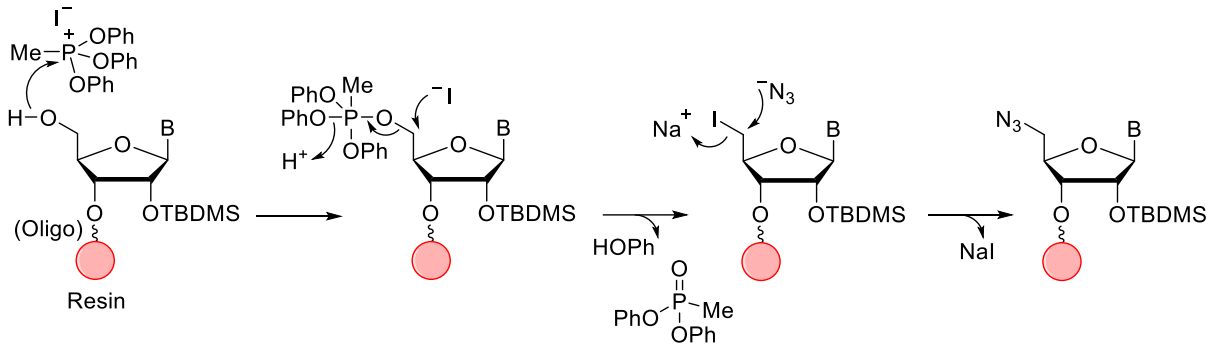



*Figure 4. 11 Mechanism showing conversion of a 5'-alcohol to 5'-azide via an iodide intermediate. MTPI is used to promote nucleophilic attack of an iodide upon the 5'-carbon. Addition of sodium azide then facilitates and SN2 reaction replacing the iodine with a 5'-azide functional group.*

To synthesise a triazole linkage with an LNA modification at the 3'-side relative to the modified backbone, commercially available LNA phosphoramidites were used. Like the normal RNA synthesis, oligonucleotides were synthesised using standard procedures and the 5'-LNA phosphoramidite was added in the final cycle. The 5'-hydroxyl was then converted to the 5'-azide using the same procedure as described previously (Figure 4. 10C). After ligation with the alkyne modified RNA, the LNA are located at the 3'-position, relative to the triazole backbone.

## 4.6 Synthesis of unmodified circRNA

Oligonucleotides (ON4.1 and ON4.2) were synthesised using standard solid phase synthesis followed by the addition of a 5'-phosphate as previously described. To assemble the unmodified linear RNA (Figure 4. 12), a precursor to the circRNA isoform, the two RNAs were held in proximity with a DNA splint (ON4.3) and were subsequently ligated using T4 DNA ligase. After incubation at 37 °C for 16 h, the ligated linear RNA was analysed and purified by PAGE. Initial attempts to form the linear 126 nt RNA were unsuccessful and no bands

corresponding to 126 nt were seen by the gel electrophoresis (Figure 4. 13A, lane 1). Considering the experimental set up, there were several components that could be at fault. RNA stability was checked by HPLC-MS (Figure 4. 16A) to confirm that the 5'-phosphate had not been degraded and that the RNA had the expected length. The MS results showed no significant degradation, leading to the assumption that RNA stability was not the issue.

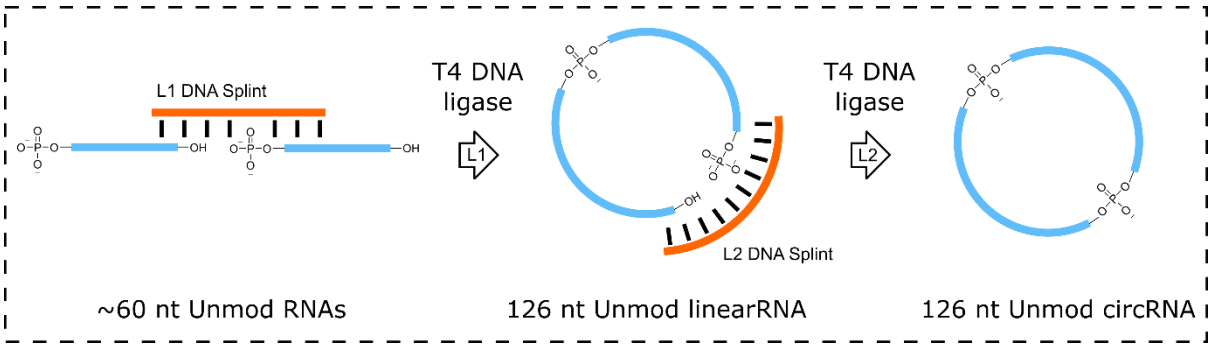



*Figure 4. 12 Synthesis of unmodified circRNA templates from 5'-phosphate modified RNA precursors. Two RNAs were ligated (L1) using a DNA splint mediated T4 DNA ligation. The linear 126 nt RNA was then purified followed by a second ligation (L2) using T4 DNA ligase in the presence of another DNA splint. This resulted in a circRNA containing conical phosphodiesters.*

Significant factors that influence RNA ligation are the length of the DNA splint, the sequence surrounding the ligation site and secondary structure of the RNA. To examine the effect of DNA splint length, a longer sequence was synthesised (ON4.4) to improve duplex hybridisation at the L1 ligation position. Another DNA splint (ON4.5) was also synthesised to hybridise to the sequences at the L2 ligation site (Figure 4. 9). Because the L2 position has a different sequence surrounding the ligation point, it could be used to determine if sequence identity and secondary structure influence ligation efficiency.

Using the new splints (ON4.4-4.5), the T4 DNA ligase-mediated ligations (L1 and L2) were incubated at 37 °C for 16 h and analysed by PAGE (Figure 4. 13A, lanes 2-3). Increasing the splint length at the L1 position produced a faint band corresponding to the 126 nt linear RNA. This suggests that for a linear ligation, increasing splint length improves duplex formation at the ligation site, facilitating ligation reaction by T4 DNA ligase. However, the band produced was still relatively faint. In retrospect the use of LNA in the DNA splint could have been beneficial for improving RNA ligation. As previously mentioned, LNA has a high affinity for RNA targets which means it could be better at penetrating potential RNA secondary structures. Once bound to the terminal regions of the two RNAs, the high affinity LNA:RNA duplex will then be more stable than a comparative DNA:RNA hybrid. It is unlikely that if the

LNA was placed within the central region of the DNA splint, that the corresponding LNA:RNA hybrid would be a suitable substrate for the T4 DNA ligase. However, if added to the wings of the splint in a gapmer design, LNA may improve ligation yields.

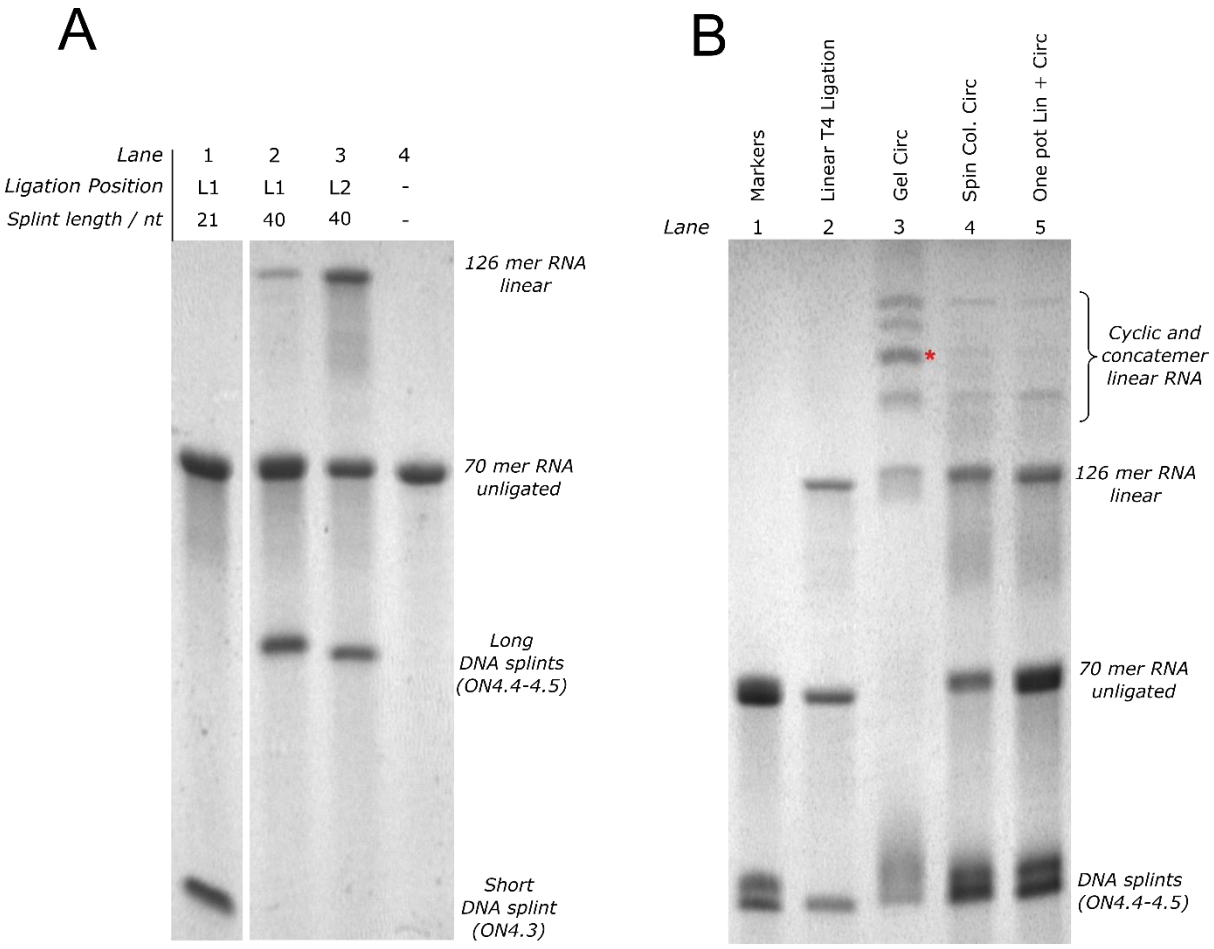



*Figure 4. 13 Polyacrylamide gels showing RNA ligations using T4 DNA ligase and DNA splints. A) Comparison of ligations using short (ON4.3) and long splints (ON4.4–4.5). The difference in ligation efficiency between L1 and L2 positions were also evaluated. Longer splints were better for linear ligation of RNAs and the L2 was more effectively ligated. RNAs were synthesised using TBDMS chemistry. B) Ligation of precursor RNAs to form 126 nt linear and circular RNAs, (Lane 2) Precursor RNA ligation to form 126 nt linear RNA. Different purification techniques after the first ligation were also compared. (lane 3) circRNA ligation after PAGE purification (the highlighted band (*) in lane 3 indicates the desired circRNA product), (lane 4) attempted circRNA formation after 10 kDa size exclusion centrifuge column, (lane 5) attempted circRNA formation by simultaneous linear and circular ligations.*

Comparison between the L1 and L2 sites showed that different ligation positions have significant effects on product yield (Figure 4. 13A, lanes 2-3). The band produced from ligation at the L2 site had a greater intensity compared to the L1 site. Since both splints used had the same length, the sequences around the L2 position must contribute to efficient ligation. However, it is unclear if this is because of the local sequence hybridisation strength or due to secondary structure around the ligation site.

Mfold is a web-based application that predicts secondary structures of RNA.[211] To determine if the disparity between L1 and L2 ligation efficiencies was the result of secondary structure, ON4.1 and ON4.2 were analysed using Mfold (Figure 7. 10). Oligonucleotide analysis using Mfold showed that there was no significant secondary structure surrounding the L1 ligation position. However, adjacent to the L2 site, ON4.1 has a region of duplex formation at the 5'-end. If secondary structure significantly impacts ligation efficiency, then the L2 should produce less ligated product. Since the opposite is found, it suggests that other factors such as secondary structure of the hybridised DNA:RNA duplex or sequence identity could be more influential factors.

The linear L2 RNA ligation using the longer RNA splint (ON4.5) was scaled up to produce sufficient amounts of materials for the circRNA generation (Figure 7. 11A, lane 1). After PAGE purification and salt removal, the circular ligation was attempted using the DNA splint (ON4.4) and T4 DNA ligase, however no circRNA product was observed (Figure 7. 11A, lane 2). It is noted by most suppliers of T4 DNA ligase, that the enzyme is sensitive to specific buffer conditions and that NaCl concentrations > 200 mM can deactivate the enzyme, most likely by exchanging divalent $Mg^{2+}$ ions for monovalent $Na^{+}$.[212] With the assumption that remaining salts from gel extraction could prevent the second cyclisation, samples were subjected to two rounds of desalting using NAP-25 columns and further desalted three more times using a 10 kDa size exclusion column. The cyclisation reaction was then repeated but still no circular RNA was observed (Figure 7. 11B).

After several repeats with no reaction, the stock of linear 126 nt RNA was consumed and so a new batch of linear RNA was prepared. The linear ligation was repeated to produce the 126 nt linear RNA precursor (Figure 4. 13B, lane 2). This was then purified by PAGE and desalted as described previously. The linear RNA was then subjected to the cyclisation using the T4 DNA ligase with a DNA splint an analysed by denaturing PAGE (ON4.4, Figure 4. 13B, lane 3). In this attempt ligation products were observed. Gel analysis produced several bands that migrated slower than the precursor linear RNA (Figure 4. 13B, lane 3). Circular RNAs are known to migrate slower than linear isoforms of identical sequence and therefore these bands are most likely to be the combination of concatemers of ligated linear RNAs and the circRNA product.[174] To distinguish between linear concatemers and the circRNA product, RNA bands were extracted from the gel and UPLC-MS was used to confirm their identity

highlighted on the figure (Figure 4. 13B, lane 3, * highlighted band, Figure 4. 14). Over the course of the study, sample concentration was optimised to promote circRNA formation. For linear ligations a much higher concentration could be used (10 pmol/µL) whereas to promote cyclisation, the samples were diluted to (1.0 pmol/µL).

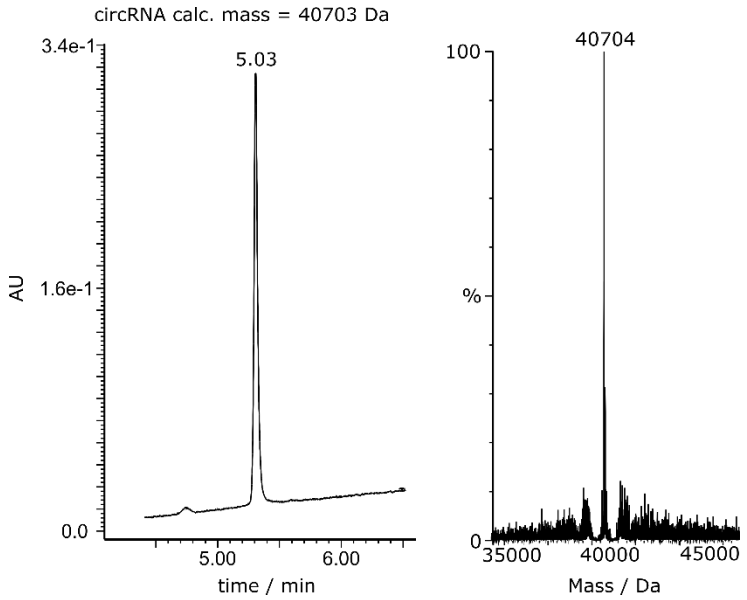



*Figure 4. 14 UPLC-MS of unmodified circular RNA extracted from the highlighted band (*) in Figure 4.* 13*B. HPLC trace shows a single peak with a mass corresponding to the calculated mass of the circular RNA.*

During recovery of the RNA from purification, it was noted that most of the RNA products could not be easily extracted from PAGE purification. As an example, the intensity of the linear RNA product before gel extraction and purification was measured to be approximately 40% compared to the unligated starting oligonucleotides when analysed using a pixel density function (ImageJ). However, after gel extraction the total yield was approximately 5%. When we consider that two ligation reactions and two rounds of PAGE purification are required to obtain the circRNA, the cumulative yield becomes extremely low (0.25%, calculated post extraction from UV adsorption 260 nm). Therefore, if the gel purification after the first ligation can be avoided, this would present a significant increase in yield (total = 2%). As an alternative to PAGE purification after the first ligation, samples were subjected to a size exclusion column to remove salts. In theory, T4 DNA ligase and the unligated RNAs should be the only contaminants remaining after desalting. The linear RNA was then cyclised using the same methods as before. In parallel, both ligations were also attempted in a single reaction to see if the circRNA could be produced by using two DNA splints simultaneously. Analysis of both

reactions (Figure 4. 13B, lane 4-5) showed only minor formation of the circRNA and that most of the products were linear concatemers.

Phenol:chloroform extraction followed by ethanol precipitation was also used to remove proteins and small molecules from oligonucleotides. This was combined with a size exclusion column to see if a combined purification would produce the same amount of circRNA as PAGE purification. Analysis using PAGE showed little circRNA was formed but also a large amount of linear by-products were generated (Figure 7. 12, lanes 2-3). From the attempts to optimise purification and reduce the number of PAGE extractions, it was found that pure RNA is the most important feature for circRNA formation and that PAGE extractions after each step is essential.

To assess how important RNA purity was for T4 DNA ligation, a batch of RNAs (ON4.1-TC and ON4.2-TC) were synthesised. Previously the RNA had been synthesised using the well-established 2'-*O*-TBDMS method of protection followed by deprotection using $NEt_3.3HF$. For the new batch, RNAs were synthesised using 2'-*O*-TC protection as the RNA could be deprotected more easily using 1,2-ethylenediamine creating a purer RNA product. The TC-RNA was purified using the same HPLC method as that used for the TBDMS-RNA. Both RNA species were then subjected to the same ligation conditions using a DNA splint (ON4.5). Comparing between TC-RNA and TBDMS-RNA, we find that both RNA protection methods can produce good ligation bands (Figure 4. 15, lane 2 *vs* lane 3). However, throughout the course of this project the quality of TBDMS-RNA was highly dependent on the ability to separate impurities by HPLC. In general, TBDMS chemistry produces more by products than the TC method. As a result, HPLC purification is harder and the quality of RNA varies between batches. Shown below are a best and worst-case example of TBDMS purified RNAs (Figure 4. 16A and C), along with a typical HPLC profile produced from TC-RNA (Figure 4. 16B) of the same sequences. From these HPLC traces TC-RNA contains minimal shoulders or additional peaks, but the highest quality obtained from TBDMS-RNA is also sufficient for ligation as shown (Figure 4. 15). In addition to this, is also shown a TBDMS-RNA HPLC profile that did not successfully produce any ligation products (Figure 4. 16C). Identical in sequence, the only difference between these samples is the purity obtained by HPLC. Based on these profiles we suggest that previous failed ligation attempts are most likely attributed to low RNA purity. It should also be noted that long RNAs (~70 nt) are difficult to purify by HPLC and each

purification is a balance between obtaining enough RNA for experiments and RNA purity. In general, it was found that additional PAGE purification of TBDMS-RNA consistently produced RNA to a high standard, comparable to TC-RNA, but that the TC method was often more convenient.

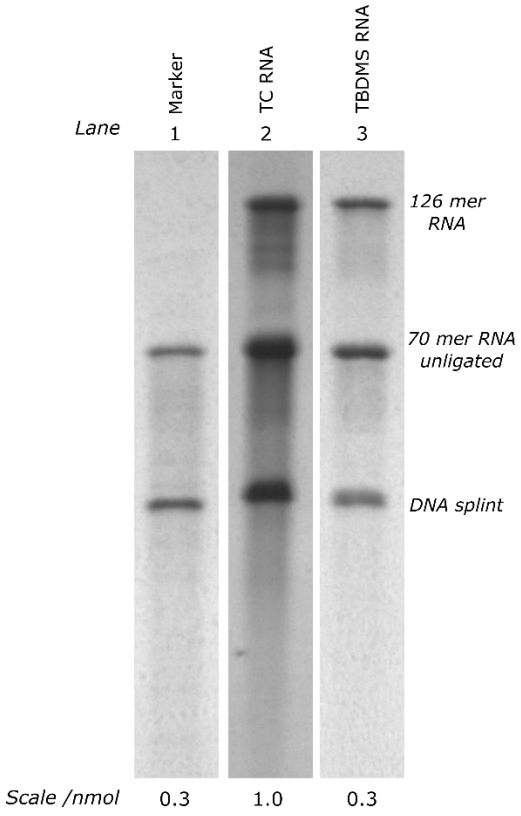



*Figure 4. 15 PAGE analysis comparing the ligation of unmodified RNAs with identical sequences synthesised by TC (lane 2) or TBDMS (lane 3) 2'-protecting groups. In general, both methods produce RNAs that can be ligated to form linear 126 nt RNA templates. However, TBDMS chemistry fidelity was found to be inconsistent between RNA batches.*

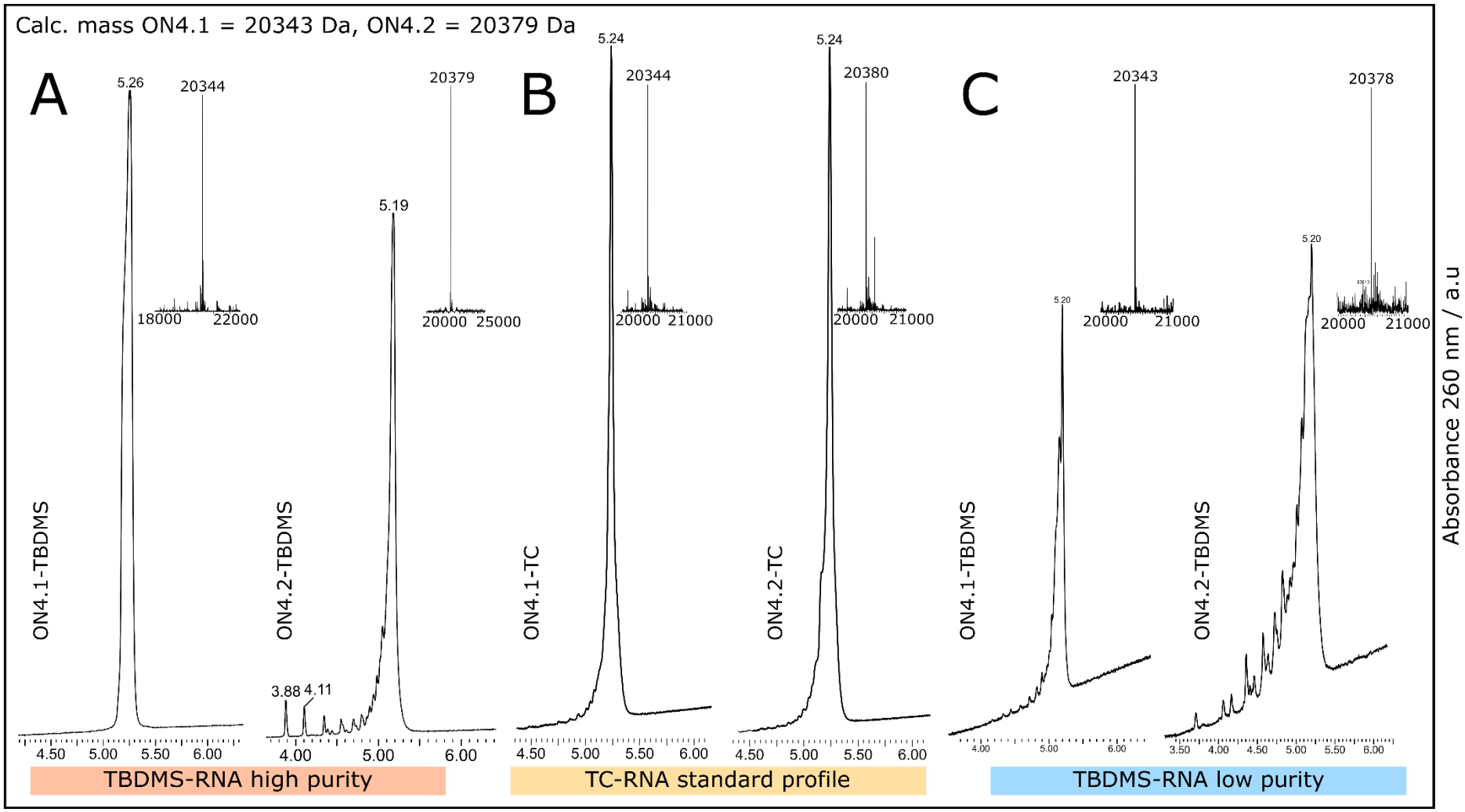



*Figure 4. 16 HPLC-MS traces of ON4.1 and ON4.2 synthesised by TBDMS or TC chemistry. A) Examples of a best case scenario after purification of long RNAs synthesised by the TBDMS method, even after HPLC small amount of failure sequences still remain, but purity is high enough for ligation reactions, B) typical profiles obtained after HPLC from RNAs synthesised by the TC method, sample purity was often higher and with good consistency between batches, C) A poor but frequent profile of TBDMS RNA after HPLC purification, RNAs of this quality are not suitable for RNA ligation.*

## 4.7 Synthesis of Tz-circRNA and Tz-LNA-circRNA

For the synthesis of a Tz-circRNA, RNA oligonucleotides require azide and alkyne modifications. For 5'-azide modification, the RNA first requires activation with iodine, which is not compatible with 2'-TC RNA protected monomers at high concentration. Because of this, RNAs that will be used for the triazole containing circRNAs must be synthesised using the TBDMS method. To maximise the purity, the RNAs were purified by PAGE followed by HPLC. The RNAs with high purity should enable successful ligations and overcome the issue of poor reactivity exhibited by TBDMS RNAs used for the unmodified control.

The Tz-circRNA was synthesised by Ashley Tsai (MPhil) and oligonucleotides (ON4.6 and ON4.7) were ligated using click chemistry with a DNA splint (ON4.4). This was then cyclised using T4 DNA ligase and DNA splint (ON4.5). The Tz-circRNA was purified by PAGE and desalted using NAP-10 filtration.

In addition to the Tz-circRNA template, a Tz-LNA-circRNA was also synthesised (ON4.6 and ON4.8) and purified by HPLC and PAGE (Figure 4. 17). Previous work by A. Tsai (MPhil) noted that triazole translation was enhanced when the modified codon contained at least one GC base pair and when the translation was performed at reduced temperatures (25 °C). This suggests that duplex stability between the ribosome, tRNA and mRNA complexes could influence translation of the modified backbone. If this hypothesis is correct, then restoration of hybridisation strength with the addition of a duplex stabilising LNA modification should improve translation performance.

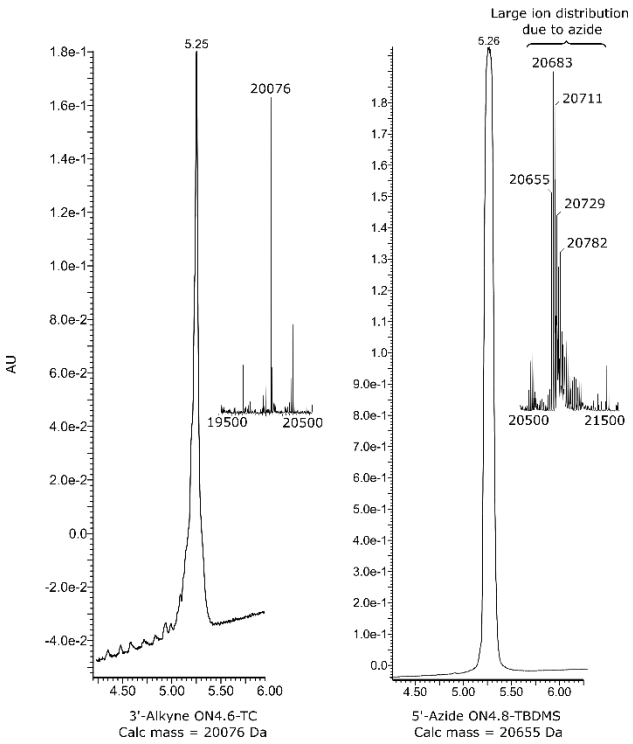



*Figure 4. 17 UPLC-MS traces showing successful purification and synthesis of ON4.6-TC and ON4.8-TBDMS. Oligonucleotides were purified by HPLC and PAGE. ON4.6 contains a 3'-alkyne and ON4.8 contains a 5'-$N_3$-LNA modification. Addition of the azide group causes significantly more ionisation of the oligonucleotide by ESI-TOF MS and as a result produces various ionic peaks in the mass spectra. Absorbance (a.u) was measured at 260 nm.*

To synthesise the Tz-LNA-circRNA template, 3'-alkyne (ON4.6) and 5'-azide-LNA (ON4.8) modified RNAs were first ligated to form the linear RNA using the CuAAC reaction and a DNA splint (ON4.4, Figure 4. 18 and Figure 4. 19A). The oligonucleotides were added to a $Cu^{I}$ solution formed *in situ* by the reduction of $CuSO_4$ by sodium ascorbate. An additional chelating ligand, THPTA, was also added to scavenge and protect the oligonucleotide from oxidative damage, caused by the co-generation of reactive species. Based on previous work,[213] the optimised procedure using 200 eq of Cu at a ratio of 1:10:7, $CuSO_4$: sodium ascorbate: THPTA was used and scaled according to the amount of RNA. The reaction was incubated for two hours at room temperature and analysed by PAGE (Figure 4. 19B, Lane 1). From the acrylamide gel, a faint band was seen corresponding to the linear RNA isoform. The low yield of ligation could be because of poor duplex formation between the DNA splint and the two RNA oligonucleotides, possibly due to secondary structures. As a result, the DNA splint failed to hold the two RNA ends in proximity for efficient CuAAC ligation. Therefore, before the addition of CuAAC reagents, the oligonucleotides were heated to 95 °C for 5 minutes to destabilise RNA secondary structures and slowly cooled (1 °C/min) to favour annealing between the DNA splint and the two RNAs.

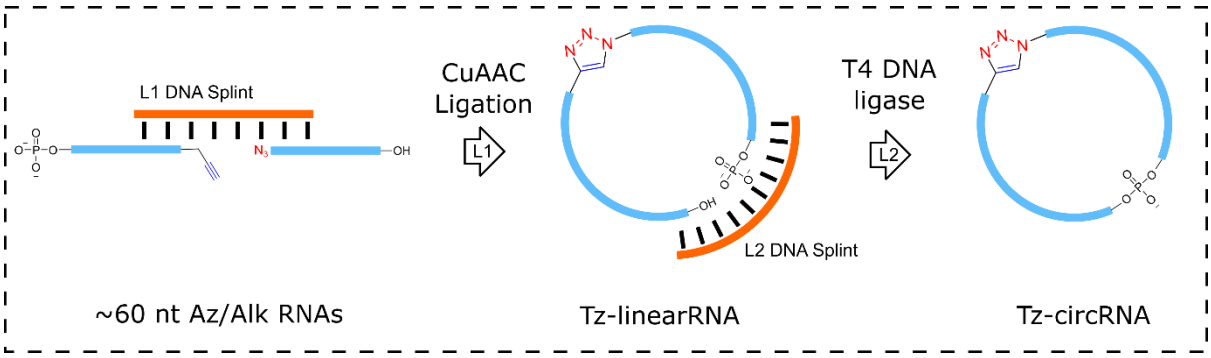



*Figure 4. 18 Synthesis of triazole-modified circRNA (Tz-circRNA and Tz-LNA-circRNA). Precursor RNAs were ligated using a templated CuAAC reaction to form the triazole backbone. After purification a second ligation was performed using T4 DNA ligase to create the Tz-circRNA.*

The CuAAC linear ligation using the Tz-LNA-RNA precursors was repeated with an annealing step before the addition of $Cu^{I}$ and analysed by PAGE (Figure 4. 19B, lane 2). Compared to ligation in the absence of annealing, the efficiency was significantly increased and a ligation band with improved intensity was produced. This suggests that low ligation efficiency was partly due to poor duplex formation between the DNA splint and RNA termini. However, still significant amounts of RNA reactants remained. As previously discussed, the formation of a triazole backbone requires a 3'-alkyne modified RNA and a corresponding 5'-azide modified RNA. Because the alkyne is at the 3'-end of the oligonucleotide, phosphoramidite solid phase synthesis in the 3' to 5' direction would not occur unless the 3'-alkyne was present on the solid support. This means that the 3'-alkyne must be present in the reaction. On the other hand, the 5'-azide was added post synthetically by conversion of the 5'-hydroxyl. It could therefore be that not all the 5'-position was converted to the azide. Theoretically this could be checked by mass spectrometry, but due to the length of the RNAs the spectra became difficult for the software to interpret. The azide group also attracted considerably more ions from the electrospray, meaning that species with similar masses such as the 5'-hydroxyl RNA could be hidden (Figure 4. 17).

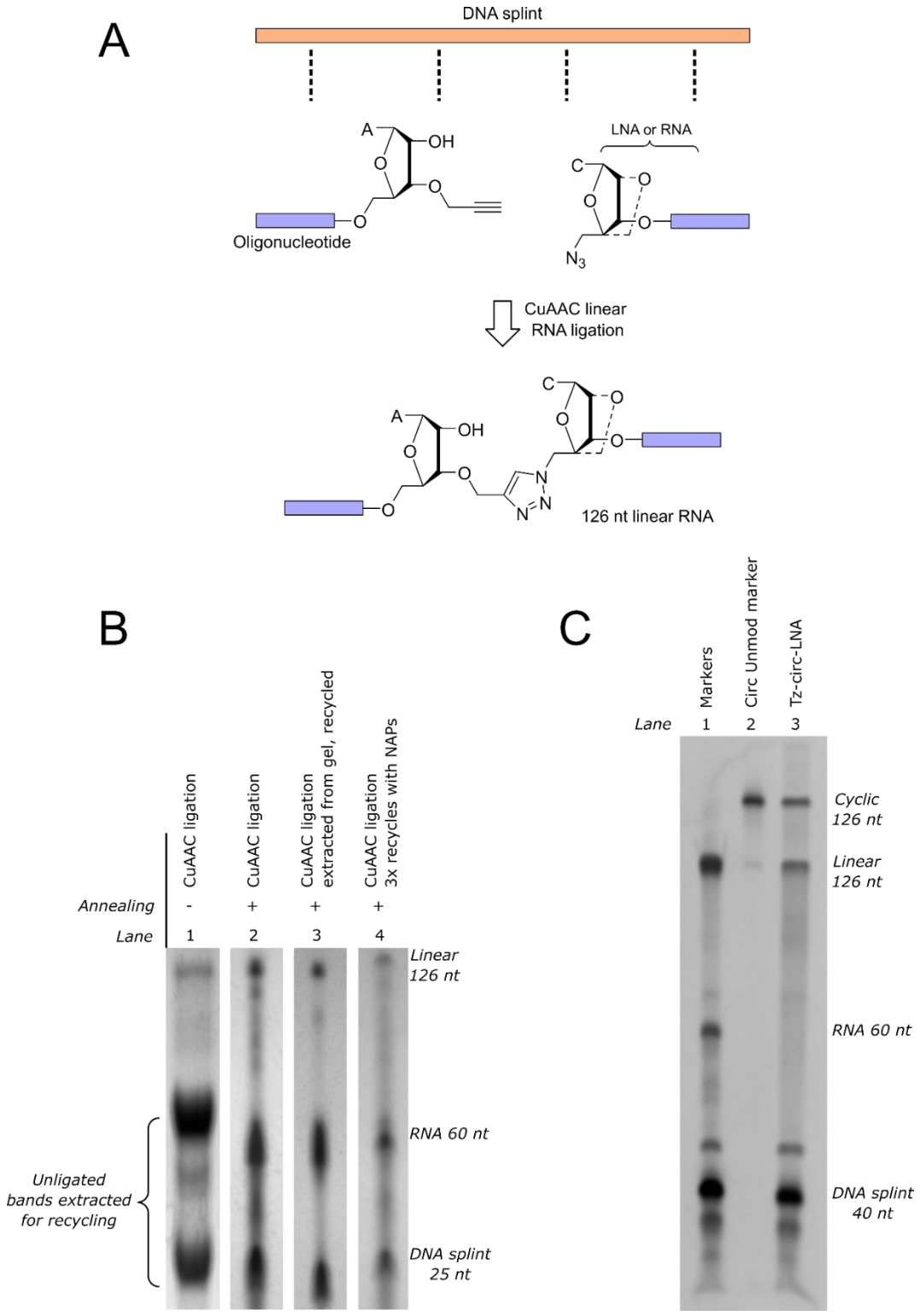



*Figure 4. 19 A) Ligation of RNA by CuAAC reaction to form a triazole-modified circRNA. The 5'-azide can either be on RNA or LNA to produce different templates. B) PAGE analysis of CuAAC ligations to form a linear 126 nt RNA, lane 1 – CuAAC linear ligation with no annealing resulted in poor ligation yield, lane 2 – CuAAC linear ligation with oligonucleotide annealing which was shown to improve ligation yield, lane 3 – ligation of unligated RNAs extracted and recycled from lane 2. This showed that the unreacted oligonucleotides were still reactive, but the reaction did not go to completion, possibly due to catalytic depletion, lane 4 – three consecutive CuAAC reactions using the same oligonucleotides. Samples were purified by NAP gel-filtration between reactions to remove small molecules and then the reaction was reset with fresh reagents, no significant consumption of starting RNAs was seen compared to the single-ligation attempts. C) cyclisation of the Tz-LNA-RNA template using T4 DNA ligase, lane 1 – linear RNA markers, lane 2 – unmodified circRNA control template, lane 3 – T4 DNA ligation of linear Tz-LNA-RNA (extracted from 4.19B, lane 2) to form Tz-LNA-circRNA.*

The reactivity of the RNAs can be tested simply by re-performing the reaction. The unreacted RNAs were extracted from the PAGE (Figure 4. 19B, lane 2) and desalted using NAP-25 gel column filtration. They were then added to a fresh mixture of copper solution and re-examined by PAGE Figure 4. 19B, lane 3). Re-performing the reaction on the recycled RNAs, produced further formation of linear RNA product, confirming that the RNAs do have the required modifications to create the triazole backbone. This therefore suggests that there are other reasons why the reaction did not go to completion.

For both T4 DNA ligase dependent and CuAAC ligations, complete consumption of the starting RNAs was not observed. Therefore, to push the reaction to completion, the reaction was recycled multiple times using the same oligonucleotides with fresh reagents. Usually, starting materials can be consumed by addition of one reactant or other reagents in large excess. However, more $Cu^{I}$ will result in excessive amounts of oxidative RNA damage, leading to the formation of truncated products.[20] To mitigate this effect, the reaction was recycled three times using 200 eq of $Cu^{I}$ for each reaction. The CuAAC reaction with modified RNAs (ON4.6 and ON4.8) and DNA splint (ON4.4) was incubated for two hours at room temperature using the standard CuAAC procedure as described before. The oligonucleotides were then purified by NAP-10 gel-filtration, freeze dried and re-dissolved in water. Two further click reactions were then performed and the reaction was analysed by PAGE (Figure 4. 19B, lane 4). The results showed the formation of a linear 126 nt RNA but with a lower intensity compared to no recycling or single recycling. This is probably because of RNA degradation from copper exposure or multiple rounds of annealing. After each cycle, it appeared that some RNA was lost or degraded during the reaction. This accumulative effect therefore reduced the total yield of the reaction and had no substantial improvement on starting material consumption.

From these optimisation reactions, it shows that driving the reaction to completion was not possible. Additional reagents were unable to produce complete ligation and further attempts (not shown) using excess of one RNA oligonucleotide did not consume the limiting RNA. Both CuAAC and T4 DNA ligase-mediated ligations appeared to stall before completion. Results by others within our group suggest that the length of an oligonucleotide affects the ligation yield. Studies on mini circular DNAs indicated that DNA ring size was inversely proportional to ligation yield (Liyiwen Yuan, DPhil, unpublished). This could therefore suggest that the length of oligonucleotide could be the limiting factor that prevents the ligation of RNAs in this work.

To improve ligation yields, reactions can be recycled one more time, but increasing the number of recycles gives no significant improvement while potentially causing damage to the RNA.

The linear 126 nt Tz-LNA-RNA (Figure 4. 19B, lane 2) was cyclised using the T4 DNA ligase in the presence of a DNA splint (ON4.5) and analysed by PAGE (Figure 4. 19C, lane 3). A single band of cyclic product was observed above the linear RNA precursor, which had comparable migration to the unmodified circRNA (Figure 4. 19C, lane 2), confirming its identity. The product was extracted from the gel and desalted using NAP-25 filtration.

## 4.8 Cell free protein expression from triazole circRNAs in prokaryotic systems

Having synthesised the unmodified, triazole (Tz) and Tz-LNA circular RNAs, their ability to serve as a template for translation was evaluated. Each circRNA template (100 ng) encoding FLAG tag peptides was added to the cell-free prokaryotic expression system (PURExpress, NEB, Figure 4. 20A). After incubation at 25 °C for 16 h, the FLAG tag peptides were analysed on a tris-glycine SDS PAGE. The proteins were transferred to a PVDF membrane to allow for western blot analysis where specific primary antibodies were added to bind to the FLAG tag peptide. A secondary antibody was then used to bind the primary antibody. The secondary antibody contains a horseradish peroxidase enzyme that reacts with a substrate to produce a chemiluminescent signal, enabling detection.

From the western blot analysis, no signal was seen from the reaction with no circRNA template (Figure 4. 20B, lane 1). This is significant because it shows that there was no non-specific binding of the antibodies to other proteins in the expression system. The unmodified circRNA produced an intense streak of products ranging from 10–250 KDa (Figure 4. 20B, lane 2). Due to the nature of rolling circular translation, the continuous expression of peptide means that translation is randomly terminated resulting in a non-specific distribution of peptide lengths. Ashley Tsai (MPhil), previously found that Tz-circRNA produced lower level of peptide expression compared to the unmodified control. However, it is unknown if this was because of poor triazole biocompatibility for the translation mechanism or reduction in duplex stability. Like those results, it was found that the triazole circular RNA expressed less peptide than the unmodified circRNA (Figure 4. 20B, lane 3). In theory the addition of LNA

should mitigate reduction in duplex stability caused by the triazole backbone. If reduced duplex stability decreases peptide translation, then LNA addition should improve peptide expression.

Tz-LNA-circRNA template was successfully translated by the prokaryotic ribosome (Figure 4. 20B, lane 4), but was unable to increase peptide expression compared to the Tz-circRNA template, producing similar amounts if not less peptide. This suggest that biocompatibility with the ribosome is probably the reason why triazole modified backbones yield less peptide compared to the unmodified circRNAs. Addition of the LNA should increase the local duplex stability between the tRNA and mRNA complex. Addition of LNA did not improve peptide expression suggesting the biocompatibility of the LNA-Tz modification has significant influence of peptide yield and that duplex stability is important but not the sole factor in peptide expression yield.

For the western blot of the unmodified circRNA, there was a significant amount of oversaturation. The reason for this is because of the detection method used by the imager. Sequential images were taken, and their intensities stacked upon each other to develop a signal over long exposure times. In order to visualise samples of weaker intensities, longer exposure times must be used, which inevitably caused over exposure of the unmodified control. When oversaturation occurs, the digital imager assigns the pixel colour to black, and all quantitative grey scaling is lost. This will be discussed in more detail (section 4.11) but it should be noted that the western blot is non-quantitative, and that there is a significant pixel intensity increase between the pre-saturation values of the Tz-circRNA and post-saturation values of the unmodified control. As a result, the band intensity shown appears much darker than is probably correct.

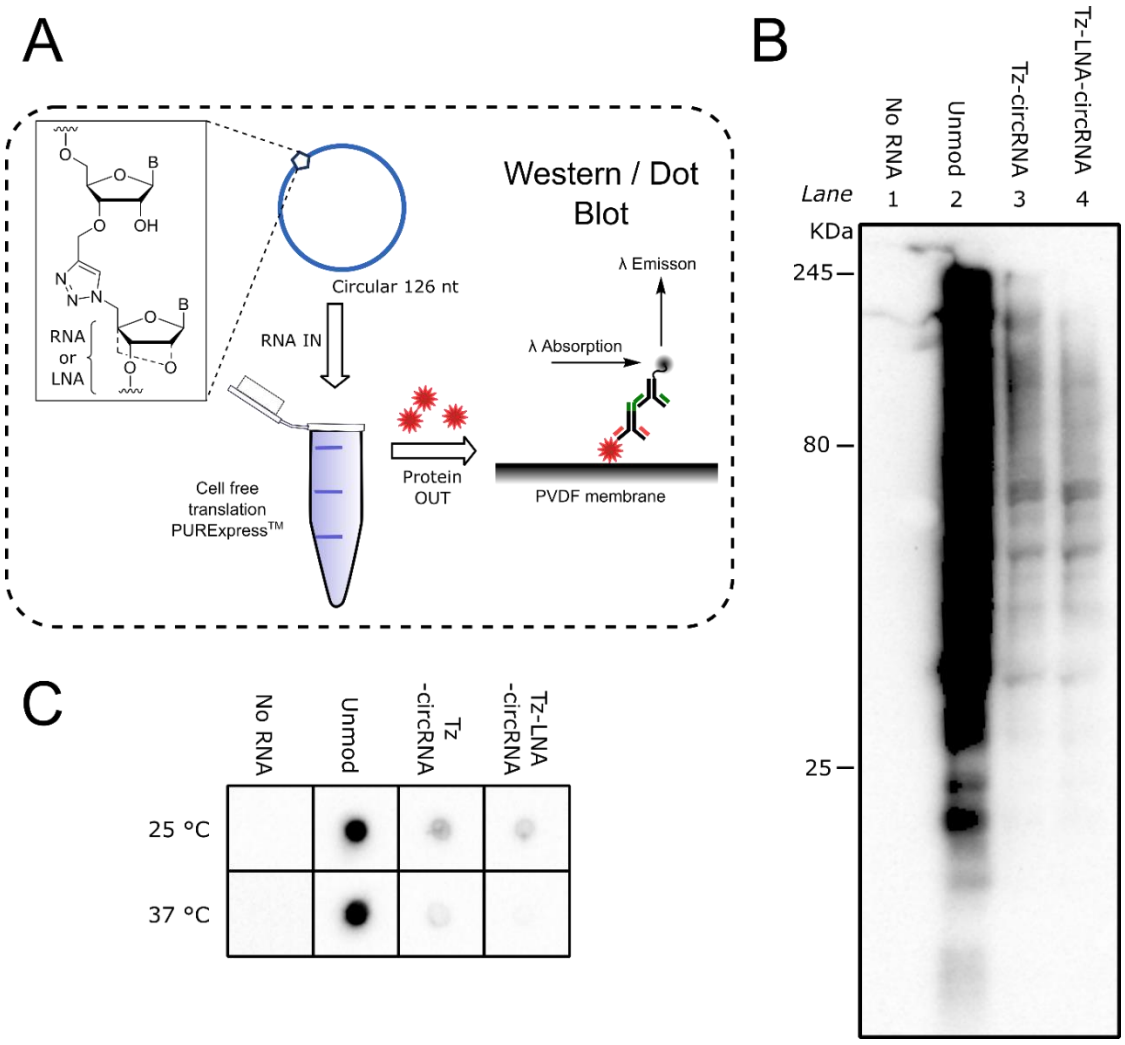



*Figure 4. 20 A) Outline of experimental workflow. Modified circRNA was added to the PURExpress prokaryotic cell-free translation kit. The expressed peptide was then either purified by tris-glycine SDS PAGE and analysed by western blot or directly added to the PVDF membrane for dot blot analysis. B) Western blot from cell-free translation of circRNA templates. Primary antibody was specific to the FLAG tag peptide and secondary antibody contained a horseradish peroxidase for visualisation, lane 1 – no circRNA template added, lane 2 – unmodified circRNA, lane 3 – Tz-circRNA, lane 4 – Tz-LNA-circRNA. No non-specific binding of antibodies was observed but LNA modification appeared to reduce expression yields. C) Dot blot analysis of circRNAs incubated at 25 °C and 37 °C, triazole-containing templates expressed better at reduced temperature and little expression was seen from the Tz-LNA-circRNA at 37 °C.*

The PURExpress system is designed to be performed at 37 °C for up to 16 h. However, from previous experiments by Ashley Tsai, it was found that minimal expression of triazole backbones occurred at the elevated temperature. When the incubation temperature was reduced to 25 °C, Tz-circRNA could be expressed as shown (Figure 4. 20B). Although Tz-LNA-circRNA was unable to improve expression yields compared to Tz-circRNA at 25 °C, it might be possible that the LNA is able to be expressed at 37 °C based on the LNAs stabilising effects. To examine this, circRNA templates were incubated for 16 h at 25 and 37 °C. Samples were then added directly to a PVDF membrane to be analysed by dot blot assay. The dot blot assay uses the same methods as the western blot, without running SDS-PAGE. Unlike the western

blot samples are not separated by size, but it allows for a faster procedure with clear results. Both Tz-circRNA and Tz-LNA-circRNA performed best when incubated at 25 °C (Figure 4. 20C). When the temperature was elevated to 37 °C, both modified templates had reduced expression compared to the lower temperature. LNA modification was unable to restore peptide expression yield compared to the unmodified template and the LNA containing template was worse than the Tz-circRNA at 37 °C. In general addition of LNA reduced peptide expression compared to the Tz-circRNA template.

In summary, Tz-LNA-circRNA offers few additional benefits compared to the Tz-circRNA template. It was found that increasing the duplex stability around the triazole containing codon did not restore protein expression and that peptide yield was reduced at higher temperatures for Tz-LNA-circRNA. This suggests that a more influential factor could be biocompatibility with the ribosome rather than mRNA: tRNA interactions. Although it is possible to translate DNA directly[214], it is likely that the RNA 2′-hydroxyl is a key but non-essential interaction needed to aid translation. Exchange for LNA, which lacks the 2′-hydroxyl group, could therefore be less favourable compared to RNA, resulting in reduced translation yields compared to Tz-circRNA.

Secondly, size and conformational flexibility could also explain why addition of LNA does not improve translation. During translation the mRNA is bent into the active sites of the ribosome, which then translocates along the mRNA strand. This process requires the RNA to be flexible and be the correct shape. If the LNA conformationally restrains the RNA in the active site or distorts interactions due to its large size, it could explain why LNA was unable to improve peptide translation. It can therefore be suggested that the more RNA-like the modification is, the more effectively it will be translated.

## 4.9 Synthesis of 5′-azides using H-phosphonates

A limiting factor towards the development of Tz-circRNAs is the synthesis of the 5′-azide modified RNA. Added post-synthetically, it was observed that not all the oligonucleotide was converted to the 5′-azide and much of the RNA was lost during purification. The downstream effect of this means that synthesis is time consuming, inefficient and expensive to synthesise larger scales. Alongside this, 5′-azide conversion has not been established with 2′-*O*-TC chemistry and therefore the lower yielding 2′-*O*-TBDMS method must be used.

An alternative method to generate the 5'-azide could be the addition of the azide during solid phase synthesis as part of a modified RNA monomer. Traditionally this is not possible due to reactivity between the azide and the phosphoramidite phosphorus (III) centre, which undergoes the Staudinger reaction, reducing the azide to an amine (Figure 4. 21).[215] However, phosphorus (V) H-phosphonates are unable to be oxidised and therefore do not cause reduction of the azide. Using the standard phosphoramidite cycle, RNA oligonucleotides will be synthesised as usual up to the penultimate monomer coupling. The resin can then be removed from the synthesiser to allow for a manual coupling of a 5'-azide modified H-phosphonate RNA monomer. The 5'-azide containing oligonucleotide can then be cleaved and deprotected under standard conditions.

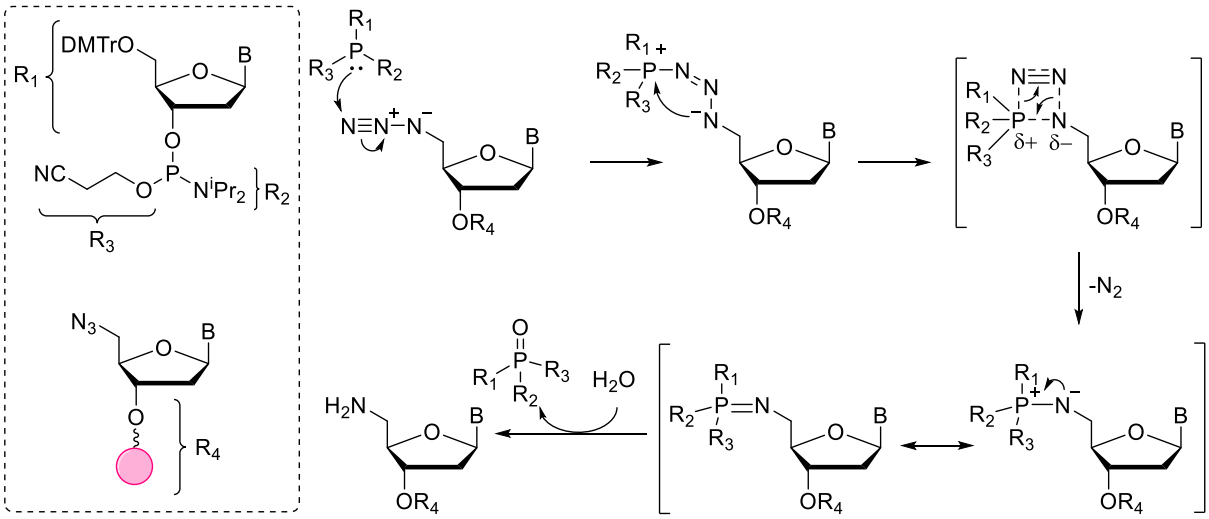



*Figure 4. 21 Mechanism of azide Staudinger reduction by reaction with a P$^{(III)}$ phosphoramidite to produce an amine and P$^{(V)}$ centre. The phosphorus lone pair attacks the azide to form a 4-membered ring intermediate species. Loss of $N_2$ promotes the formation of a P=N bond which is then subsequently cleaved in aqueous conditions forming an amine. Phosphoramidites that undergo this reaction cannot be used to synthesise oligonucleotides.*

The benefits of this method are that the synthesised oligonucleotides will be easier to purify and easily identifiable by mass spectrometry, comparing to uncoupled oligonucleotides. Synthesis of the modified oligonucleotide can also be fully automated as the H-phosphonate procedure is compatible with oligonucleotide synthesisers.[15] This makes it attractive to researchers who are not specialists in oligonucleotides or for commercial applications. Finally, it may also be possible to combine azide H-phosphonates with the 2'-*O*-TC method which would improve yield and efficiency of long RNA synthesis. However, for these initial studies the TBDMS method was used because of its known stability to conditions needed to oxidise the H-phosphonate using molecular iodine in aqueous solvent.

Since the triazole backbone is positioned between the second and third positions of a GAC codon, the 5'-azide must be on the cytidine. For the synthesis of an azide modified H-phosphonate (Figure 4. 22), benzyl-protected cytidine RNA nucleoside **41** was reacted with MsCl which resulted in a mixture of 5'-monosubstituted mesyl compound **42** and a 5'-3'-disubstituted mesyl. The resultant crude was column purified to remove the disubstituted mesyl, but due to their similar polarities both compounds often co-elute. It was debated if a 2'-3'-hydroxyl protecting group, such as an acetal, should be used to prevent mesylation of the secondary alcohols. This was attempted but it was found that addition and removal of the protecting group was less effective than just direct mesylation and column purification. The crude compound **42** was reacted with $NaN_3$ which produced 5'-azide **43**. After azide conversion the polarity difference between the 5'-azide and 5'-3'-diazide was greater than the previous 5'-mesyl and 5'-3'-dimesyl species. As a result, the 5'-3'-azide was easily removed from the reaction using silica column purification. Reaction of compound **43** with TBDMSCl facilitated the protection of the 2'-hydroxyl group. This is a critical step because it is important that the TBDMS was added to the 2'-hydroxyl position, so that during oligonucleotide synthesis the canonical 3'-5' phosphodiester linkage could be formed. Reports from the literature suggested that addition of $AgNO_3$ directs the TBDMS addition selectively to the 2'-position.[216] When TBDMSCl was added in portions with $AgNO_3$, the reaction was slow but after addition of further equivalents the starting material was observed to be consumed. Analysis of the crude LCMS and collection of separate fractions after column purification showed that the major product was the 3'-*O*-TBDMS protected by-product instead of the desired 2'-*O*-TBDMS (1:0.81). Despite this, enough of the 2'-*O*-TBDMS protected compound **44** was obtained and the protecting group position was confirmed by HSQC and COSY NMR (Figure 7. 13). Compound **44** was reacted with diphenyl phosphite followed by the addition of $H_2O$:$NEt_3$ to produce 5'-azide H-phosphonate **45** in 84% yield.

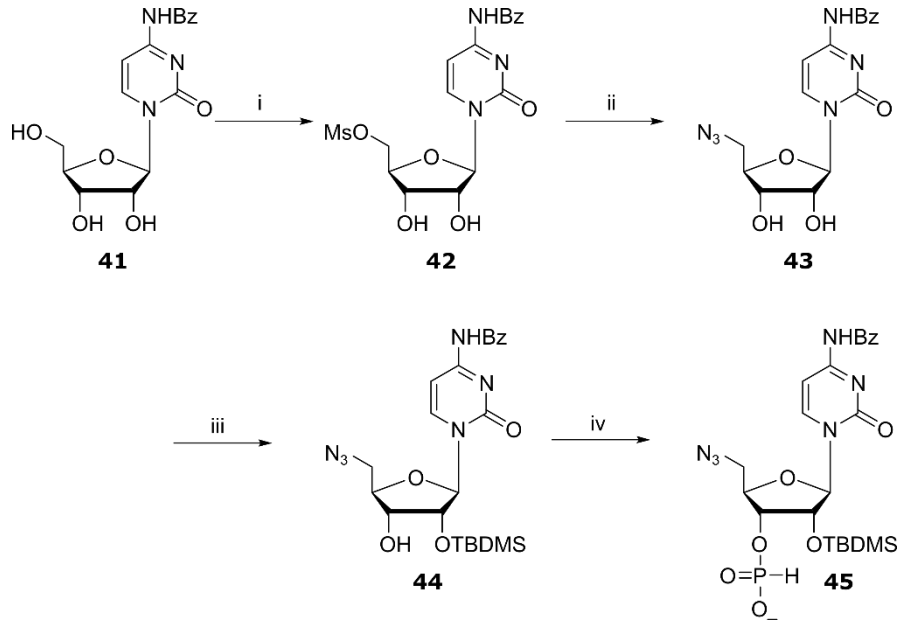



*Figure 4. 22 Synthesis of 5'-azide H-phosphonate for oligonucleotide synthesis. (i) MsCl, pyridine, rt, 1.5 h,* ***42*** *crude 35%, (ii) $NaN_3$, DMF, 70 °C, 3 h,* ***43*** *57%, (iii) TBDMSCl, $AgNO_3$, pyridine, DMF, rt, 49 h,* ***44*** *21%, (iv) diphenyl phosphite; $H_2O$: $NEt_3$ (1:1), pyridine, rt, 2 h,* ***45*** *84%.*

Since it was unknown if the 5'-azide H-phosphonate **45** will be compatible with solid phase synthesis, we decided to synthesise a DNA sequence (ON4.9) to evaluate the monomer coupling. After standard solid phase synthesis (Figure 4. 23A), a small portion of the resin-bound DNA was taken and cleaved in aqueous $NH_3$ for comparison after the 5'-azide H-phosphonate coupling. The remaining resin-bound DNA was coupled with the 5'-azide H-phosphonate **45** by incubating for 1 h at room temperature followed by oxidation in $I_2/H_2O$ to form the phosphodiester linkage. The DNA containing the 5'-azide RNA monomer was then cleaved from the solid support and analysed by UPLC-MS without deprotection of the 2'-TBDMS group. UV-trace analysis comparing the pre-coupled unmodified DNA (Figure 4. 23B, green) to the modified DNA coupled to the H-phosphonate (red) showed almost quantitative coupling of the 5'-azide H-phosphonate. Only trace amounts of unmodified DNA remained along with a small percentage of oligonucleotide with the TBDMS group removed. The UV chromatogram confirms that the H-phosphonate coupling was efficient, producing few by-products and provided an alternative method for installing 5'-azide modifications on to oligonucleotides.

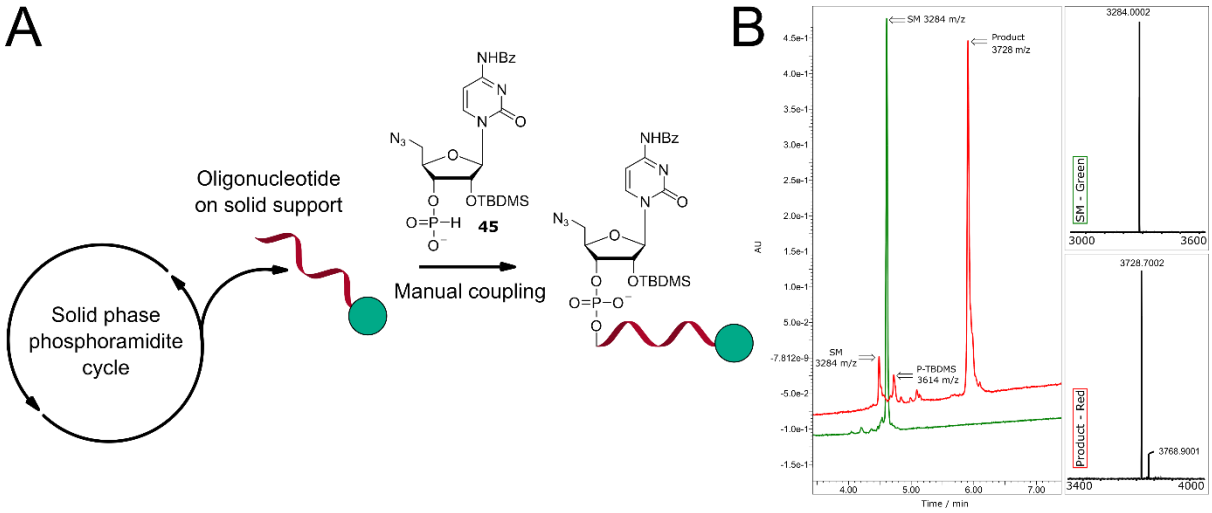



*Figure 4. 23 A) Outline of the coupling of the 5'-azide H-phosphonate monomer to the oligonucleotide. The oligonucleotide ON4.9 (**C**CGTCTACTGCT, **C** = RNA H-phosphonate **45** added to the DNA oligonucleotide) was synthesised using standard solid phase synthesis before manually coupling the H-phosphonate RNA monomer **45**. The oligonucleotide was then removed from the solid support and deprotected using aqueous $NH_3$. B) UV UPLC-MS trace showing coupling of H-phosphonate to a DNA sequence (ON4.9). Green – unmodified DNA before H-phosphonate addition, Red – DNA with 5'-azide TBDMS protected RNA monomer added by H-phosphonate coupling. Trace shows almost quantitative coupling and few by-products.*

## 4.10 Synthesis and translation of circRNAs containing multiple triazole backbones

Although we had successfully shown that the triazole backbone can be translated by the cell-free prokaryotic system, the synthesis of Tz-circRNA still required the use of T4 DNA ligase for cyclisation. To fully produce the templates by chemical means, we decided to produce a triazole template that contains two separate triazole backbones (Figure 4. 24A). This will allow for complete chemical synthesis of long circRNAs and it will be interesting to see if a template with more than one triazole can be translated.

As previously described, triazole RNAs must be modified with the 3'-alkyne and 5'-azide functional groups. In order to form two triazoles within the circRNA both precursor RNAs must have a 3'-alkyne and 5'-azide. RNAs were synthesised using the TBDMS method, and the first nucleotide attached to the resin was a 3'-modified alkyne monomer (Figure 4. 10B). The 5'-azide groups were then added to the RNA using H-phosphonate **45**. After cleavage from the solid support the oligonucleotides were deprotected to give 5'-azide-3'-alkyne modified oligonucleotide RNAs (ON4.10 and ON4.11).

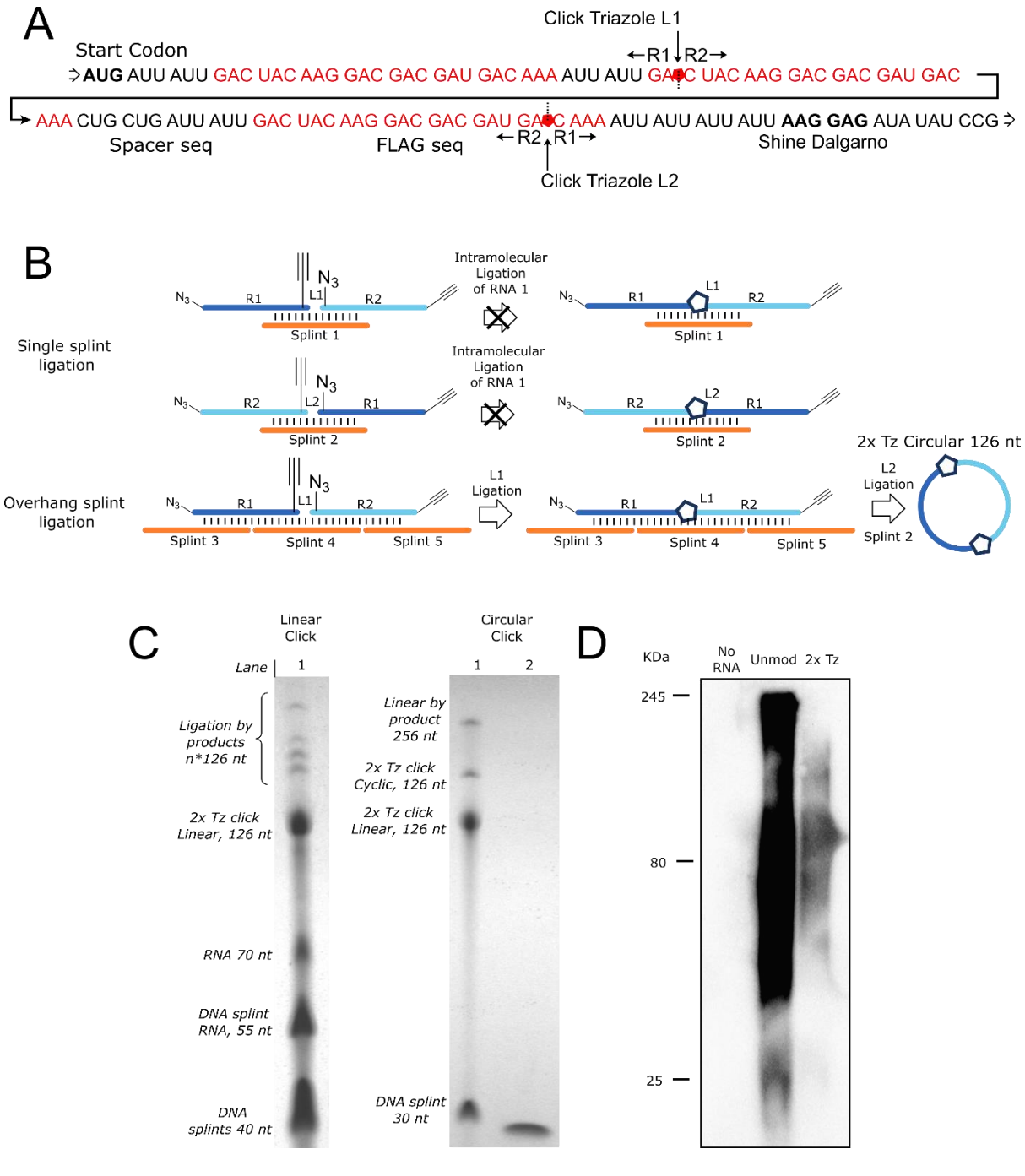



*Figure 4. 24 A) CircRNA that contained two triazole backbones and encoded three FLAG tags (red). L1 is the first ligation position and L2 is the second ligation. FLAG sequences are interspaced with spacer sequences (black) as well as the Shine Dalgarno initiation sequence and the start codon (bold). B) DNA splint-mediated ligations to generate circRNA with two triazole backbones. Ligation with a single splint spanning either the L1 or L2 ligation positions resulted in intramolecular ligation of the RNA (R1). To promote intermolecular ligation additional splints were added to the terminal regions blocking the self-ligation of R1. C) PAGE showing CuAAC-mediated linear and circular ligations to form circRNA containing two triazole linkages. D) Western blot analysis from the prokaryotic cell-free expression system comparing the unmodified (unmod) and 2x Tz-circRNA (2x Tz) as translation templates. Results showed that multiple triazoles could be successfully translated but with lower yield than the unmodified control.*

Initial attempts to ligate RNAs (ON4.10-4.11) with a single DNA splint (ON4.12, splint 1, Figure 4. 24B) were unsuccessful due to intramolecular ligation of the 70 mer RNA (R1, ON4.10). Single stranded long RNAs are relatively flexible, allowing folding upon themselves and often

bringing the terminal regions in proximity. With the ligating groups 3'-alkyne and 5'-azide at each end of the RNA, rapid intramolecular cyclisation of the RNA occurred before intermolecular ligation using a DNA splint.

By modifying the ligation protocol, it was thought that it may be possible to promote intermolecular ligation of the RNAs (ON4.10-4.11). Ligation of the L1 position was repeated using the DNA splint (ON4.12, splint 1) along with a second reaction to attempt to ligate the L2 position instead (ON4.13, splint 2, Figure 4. 24B). Analysis by PAGE (Figure 7. 14, left) showed that in both cases ligation position made no difference to R1 (ON4.10) cyclisation. Following this, attempts were also made to ligate both positions in a stepwise manner. One splint was added to the RNAs and incubated for 1 h to achieve the first ligation. After 1 h the second splint was added which should ligate the second position. In a separate reaction, splints 1 and 2 (ON4.12-4.13) were added simultaneously and annealed with the RNAs. By performing the annealing step before adding the $Cu^{I}$, the DNA splints should hybridise to the RNAs, preventing intramolecular cyclisation and allowing the circRNA to be generated. Despite all these attempts, self-cyclisation of the RNA (ON4.10) still could not be avoided (Figure 7. 14, left).

To prevent self-cyclisation of ON4.10, the reactive termini of the RNAs needed to be blocked to stop them complexing together with the $Cu^{I}$ centre. This was achieved by annealing the RNAs with three DNA splints (Figure 4. 24B, splints 3, 4 and 5, ON4.14, ON4.15, ON4.16). Splints 3 and 5 were designed to block the L2 ligation site as DNA:RNA duplexes, while the central splint 4 was re-designed as a longer sequence to replace splint 1. Alongside blocking the terminal reactive groups at the L2 position, longer sections of hybridisation decrease the flexibility of the oligonucleotide. As a result, the likelihood of the two terminal regions interacting in solution is reduced. The use of these overhanging splints for the ligation of RNAs (ON4.10-4.11) was successful, producing a ligation band corresponding to the 126-nt linear RNA and preventing the self-cyclisation of ON4.10 (Figure 7. 14, right).

Having developed an effective protocol for the ligation of RNAs containing both reactive functional groups, ON4.10 and ON4.11 were ligated at the L1 position using the overhanging splints (ON4.14 and ON4.16) and central splint (ON4.15). After ligation, the linear RNA was purified and extracted by PAGE (Figure 4. 24C) and cyclised with a second CuAAC reaction at

the L2 position using a DNA splint (ON4.13). The extracted 2xTz-circRNA template was then purified by PAGE and desalted.

Using the PURExpress cell-free prokaryotic translation system, the double triazole containing circRNA template (2xTz-circRNA) was used to express FLAG peptides which was subsequently analysed by western blot (Figure 4. 24D). From this we found that circRNAs containing multiple triazole linkages could be translated by the prokaryotic machinery. Like the single triazole containing templates, the 2xTz-circRNA produced less peptide than the unmodified circRNA, but with a distribution of protein size 60-200 kDa.

## 4.11 Quantification of peptide yield from modified circRNAs

Having qualitatively proven that the triazole-modified templates can be expressed in prokaryotes, the next goal was to quantify peptide yield. When imaging previous western blots, the unmodified control would quickly become oversaturated if the blots were exposed for enough time to observe the amount of peptide produced from the modified templates in the same image. Oversaturated pixels lose intensity information and cannot be used for quantitative yield analysis. To overcome this issue, a two-fold serial dilution of the unmodified control template was used to create a linear standard curve, which would provide a quantitative comparison without oversaturation. These samples, along with each of the triazole-containing circRNAs (Tz, Tz-LNA and 2xTz) were then used as templates in the PURExpress prokaryotic translation system and analysed by the dot blot assay (Figure 4. 25A and B). Each sample in Figure 4. 25A and B originated from the same image and was captured at the same time to enable direct comparison. Exposure time was adjusted using the gel doc imager, to ensure that the pixels were not saturated, which would be indicated by them turning red. The intensity of each dot was then measured using the imageJ software package and plotted as a function of pixel density *vs* the amount of template (Figure 4. 25B). In theory, pixel density should be linearly correlated to the amount of template. However, results from imageJ analysis (Figure 4. 25B) showed that pixel intensity was non-linear at high values. Second assessment of pixel intensity using imageJ showed that three of the samples were oversaturated (red), which previously were not indicated by the gel doc imager. This means that the standard curve cannot be used to quantify the reaction yield because information of the true pixel density relative to the samples is lost.

Following this, we questioned the method in which the samples were prepared. When performing the dot blot assay, samples were added directly to a damp PVDF membrane which was then allowed to dry. When the sample was applied to the membrane, it diffused across the surface and was highly dependent on how damp the membrane was. As the membrane became dry, the spots diffused less and were spread over a smaller area, making it difficult to control how much sample was embedded in the membrane. Because of this, it could mean that the pixel measurements across a two-dimensional surface are significantly impacted by sample distribution, invalidating the result.

Based on the finding that pixel intensity could be easily oversaturated, and that sample preparation was difficult to control, we questioned if the dot blot was a suitable assay to measure the yield of protein expression. From comparison of the triazole-containing circRNAs (Tz, Tz-LNA, 2x Tz, Figure 4. 25A) it was concluded that they each expressed similar amounts of peptide, and they all express less peptide than the unmodified control. However, the dot blot technique does not allow us to quantify how much peptide was expressed and an alternative method is required to obtain this information.

Because of the small quantities of peptides produced, it was not possible to quantify them by UV, Bradford assay or silver staining. Therefore, a highly sensitive, solution-based and specific technique was required. The enzyme-linked immunosorbent assay (ELISA), is a plate reader assay that uses antibodies like the dot blot but is quantitative due to its solution-based measurements. The reason the dot blot is not quantitative is due to the difficulty in controlling key variables that directly affect measurement values. These include the time taken to image the membrane, application and even coating of the developing solution, and suitability of the gel doc imager as imaging equipment. The ELISA assay is quantitative because the plate readers measurement range is greater than the gel doc imager and substate concentration is measured by UV in solution. This means it is sensitive enough to detect the small quantities of peptide produced from the Tz-modified templates, while at the same time not becoming oversaturated by the unmodified control. Alongside this, the solution-based measurements are controlled by using equal sample volumes as well as averaging of triplicate measurements and control of the end point by quenching the detection reaction.

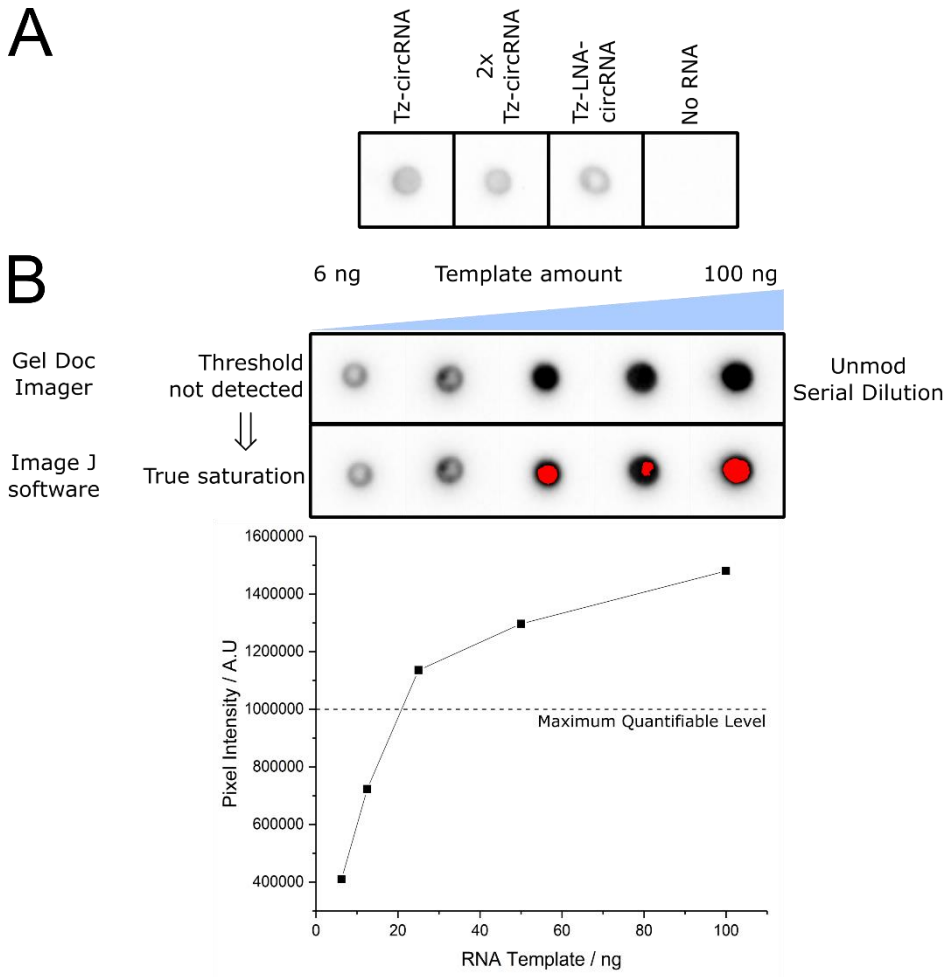



*Figure 4. 25 A) Dot blot assay showing the expression of FLAG tag peptides from triazole-modified circRNAs using prokaryotic cell-free translation. Samples include single triazole modification (Tz-circRNA), two triazole backbones (2xTz-circRNA) and triazole backbones with locked nucleic acids (Tz-LNA-circRNA). B) Dot blot assay showing a serial dilution of unmodified circRNAs from prokaryotic cell-free translation. Analysis of the pixel intensity using the gel doc imager showed no pixel saturation. However, alternative analysis using ImageJ highlighted saturated pixels (red). This prevents the serial dilution from being used as a reference to measure the amount of peptide expression. Both A and B were captured as one image to allow direct comparison between the unmodified serial dilution and triazole templates.*

In an ELISA, the peptide is first captured onto a surface using a specific capture antibody (Figure 4. 26A). The primary antibody is then bound to the peptide which recruits a secondary antibody tagged with a horseradish peroxidase enzyme (HRP). 3,3’,5,5’-tetramethylbenzidine (TMB) is a substrate for the HRP, which reacts in solution to produce a colourimetric response that can be measured in real time (655 nm) or as an end-point assay (450 nm). As a result, more peptides will bind more antibodies, and in turn produce a stronger colorimetric response.

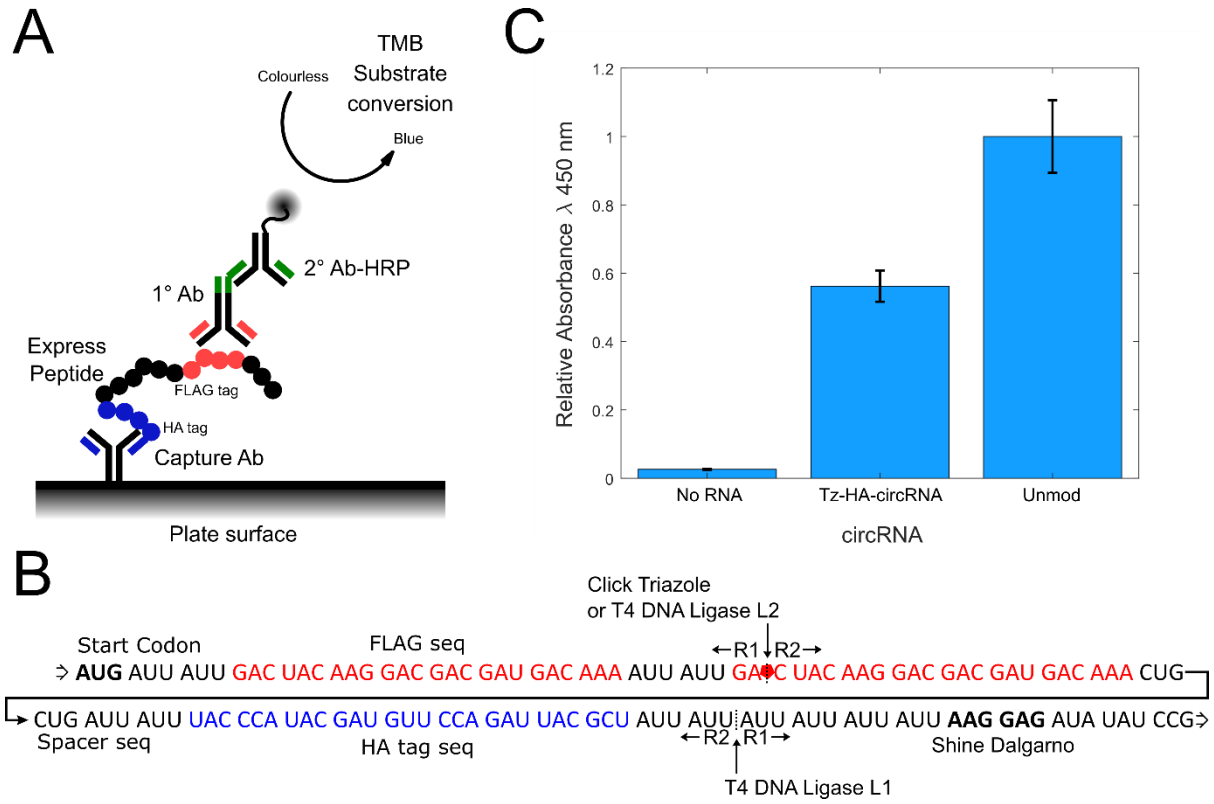



*Figure 4. 26 A) Overview of ELISA antibody binding and detection. Capture antibody specific to the HA-tag antigen was captured to the plate surface. The HA-tag then binds to the antibody which subsequently recruited the primary and secondary antibodies for analysis. B) Tz-circRNA containing two FLAG-tag sequences (red), a HA-tag capture sequence (blue), spacer sequences (black), a Shine Dalgarno sequence and the start codon (bold). C) Plot of peptide quantification by ELISA assay. Samples were measured as an absorbance at 450 nm and normalised relative to the unmodified control. Measurements were taken in triplicate and errors calculated as the standard error. Tz-HA-circRNA was found to produce approximately half the amount of peptides expressed from the unmodified control.*

In order to perform the ELISA assay, the antibodies need to be carefully selected to prevent non-specific binding to one another. The primary and secondary antibodies can be produced from the same host whereas the capture antibody must originate from a different host organism. The reason for this is because the primary antibody is produced in a mouse, and the secondary antibody binds specifically to mouse antibodies (anti-mouse). So, if the capture antibody is also generated in a mouse, the secondary antibody can bind directly to the capture antibody, producing a non-specific signal. Therefore, the capture antibody used in this study was produced in a rabbit (Thermo-Fisher scientific). Changes to the circRNA sequence were made to also prevent the capture and primary antibodies from competitively binding to the FLAG tag epitope. One of the most common epitope tags to be used with its corresponding antibodies is the His-tag sequence. However, the PURExpress assay uses His-tags to purify components in the commercial kit. Therefore, if a His-tag was encoded into the circRNA, resultant His-tag proteins will bind to the capture antibody, resulting in capture of any His-tag

containing proteins. Because of this, we opted to use a HA-tag which was chosen based on the availability of HA specific antibodies and its ability to be discriminated from the FLAG-tag and His-tag sequences by these antibodies. A new circRNA sequence was designed where one of the FLAG-tags was replaced by the HA-tag sequence (Figure 4. 26B). The expressed protein can then be captured onto the plate using the HA-tag, followed by antibody detection using FLAG-tag sequence.

An unmodified-HA and Tz-HA-circRNA were synthesised using the same ligation methods as for the previous circRNAs (Figure 7. 15). For the formation of the unmodified circRNA, ON4.17 and ON4.18 were ligated using a DNA splint (ON4.21) and T4 DNA ligase. This was followed by a second ligation, again using T4 DNA ligase and a DNA splint (ON4.22), to cyclise the RNA. Tz-HA-circRNA was formed by linear ligation of ON4.19 and ON4.20 using a DNA splint (ON4.21) and T4 DNA ligase. The RNA was then cyclised using the CuAAC reaction to form the triazole backbone, templated by a DNA splint (ON4.22) and purified by PAGE. Unlike with the previous template that contained two triazole backbones, no overhanging splints were used, since the RNAs contain orthogonal phosphate and click reacting groups that cannot ligate to one another. In addition to this, RNAs were all synthesised using the TBDMS method to ensure precursors were handled using identical chemistries.

The HA-tag containing circRNAs were subjected to the PURExpress protein synthesis kit before being added to an ELISA plate pre-prepared with HA-capture antibodies. The expression reaction and ELISA assay were performed in triplicate. After incubation with primary and secondary antibodies, the TMB substrate was added to the ELISA plate to react with the HRP attached on the secondary antibody, generating colour that could be monitored at 655 nm. The reaction was then quenched by the addition of 0.2 M $H_2SO_4$. Final ELISA plate readings were collected at 450 nm as an end point assay (Figure 4. 26C). From the ELISA assay it was found that in the absence of a circRNA template, minimal colorimetric signal was produced which confirms that there was no non-specific binding of the antibodies to each other or to the capture antibody on the plate. It was also found that the Tz-HA-circRNA yielded 48% of proteins relative to the unmodified control. This is a promising result and suggests that the triazole-containing templates can produce more peptide than was suggested by the previous dot blot quantification.

In a comparison with the previous technique, a dot blot of the Tz-HA-circRNA was performed. We found that the Tz-HA-circRNA produced significantly less proteins shown by dot blot than indicated by the ELISA assay (Figure 4. 26C vs Figure 7. 16). Although the dot blot is not fully quantitative, there should be some consistency between the two techniques. However, the results from the dot blot and ELISA are substantially different. The reason for this could be the dot blot is an inaccurate method, due to the inability to control variables such as exposure time or immunoblotting detection. But it is also important to consider possible bias originating from the ELISA assay. It is known that measurements taken on the plate reader were not saturated unlike with the dot blot assay, because the absorbance measurement was recorded below 1.0 AU (saturation ~1.3 – 1.5 AU) and all samples were measured in triplicate. This provides confidence that the physical measurements of the samples were consistent and correct, but it could be that the capture antibodies on the plate have themselves become saturated. If the amount of proteins produced by the unmodified circRNA exceeds the binding capacity of the capture antibodies bound to the plate surface, the excess proteins will be washed away. As a result, it will reduce the absorbance reading of the unmodified control, appearing to show that the Tz-circRNA template produces more peptides than it does. Therefore, further assays are needed to assess the binding capacity of the capture antibody.

To assess if the amount of proteins produced exceeds the binding capacity of the capture antibodies, the ELISA assay was run in a serial dilution. 20, 50 and 100 ng of circRNA templates were aliquoted into triplicate samples and subjected to the PURExpress translation. Using the same ELISA procedure as before, a colorimetric response of TMB conversion was analysed by UV absorbance at 450 nm. Each sample was measured three times and the triplicates were combined to give a total of nine measurements for each sample.

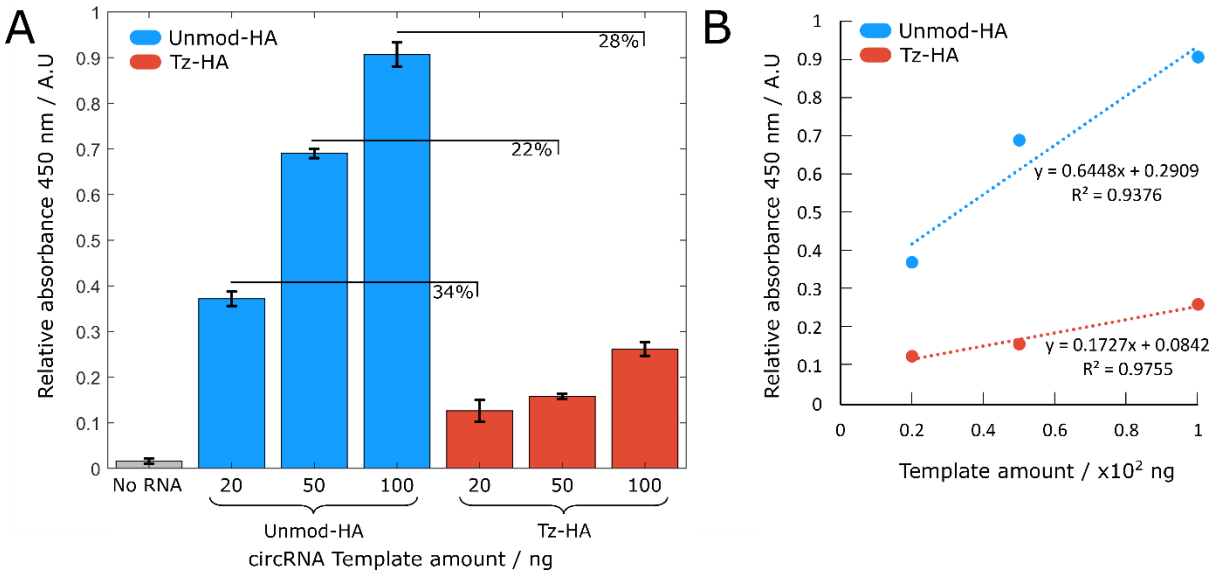



*Figure 4. 27 A) ELISA assay plot showing FLAG peptide expression from circRNA templates. Peptides were translated in vitro using a serial dilution of circRNA templates with values of 20, 50 and 100 ng. Each sample was replicated three times and UV absorbance measurements at 450 nm were recorded in triplicate (sample size n = 9). Values were taken as an average with error bars representing a standard error. Percentages highlighted on the chart are comparative yields between the unmodified and triazole-containing templates at each template dilution. B) Scatter plot showing relative absorbance vs template amount. Gradient values show that a two-fold increase in template amount does not scale at a 1:1 ratio with protein yield.*

In agreement with previous dot blot and western blot results, it was found that the unmodified template produces higher yields of peptide compared to the triazole modified template (Figure 4. 27A). Analysis of peptide expression from different quantities of circRNAs (20–100 ng) showed that RNAs containing triazole modifications produced 22–34% (28% in average) as much peptide as the unmodified control. As previously found, peptide yields using the Tz-circRNA were greater when recorded by the ELISA assay compared to the dot blot approach. Before performing this experiment, it was assumed that peptide yield is directly proportional to template amount, and a two-fold increase in template should double peptide yield. According to the gradient of the plots of template amount *vs* peptide yield (Figure 4. 27B) it was found that this was not the case. For the unmodified control the gradient increases by a factor of 0.64 and the Tz-circRNA by 0.17. This means that addition of more Tz-circRNA template has less effect on total peptide yield compare to the unmodified control.

Based on the serial dilution results, it is believed that the capture antibodies are not oversaturated by the unmodified control when using 100 ng of template. This is because each time more template is added from 20 – 100 ng total peptide yield also increases. Alongside this, the ratio of peptide produced between the unmodified control and Tz-circRNA remains

constant at each template quantity. If the capture antibodies became saturated at higher template amounts, then two trends would be observed. Firstly, it would be seen that addition of more template would not increase the amount of peptide produced, and secondly as more template is added the amount of peptide produced by the Tz-circRNA will increased relative to the unmodified control. Based on the consistencies between the two samples and the number of measurements taken, we believe that the ELISA assay produces a true reflection of peptide yield and that disparity between this assay and the dot blot can be accounted for by pixel density saturation.

## 4.12 Tz-circRNA future direction and applications

Although the ELISA assay quantifies peptide yield, it would be ideal to resolve discrepancies between the dot blot and the ELISA assay using a third technique that is not based on immunoblotting. Over the course of the project, attempts were made to quantify peptide expression using Coomassie or silver staining. However, due to the insensitive nature of these techniques no peptide was observed. Using a fluorescent assay maybe an alternative method to detect the quantities of peptide. Small molecule dyes such as FAM or BODIPY®-FL can be attached to the protein post synthetically or directly incorporated into the protein using a tRNA loaded with a fluorescent amino acid. Providing they have good sensitivity; the peptides can then be quantified using fluorescent spectroscopy or directly from a gel.

The focus of this work has been to demonstrate translation through a triazole backbone to generate polypeptide products in prokaryotic systems. If triazole-modified RNA templates are to have therapeutic applications, such as synthetic templates for *in vivo* protein expression, then their compatibility in translation will need to be confirmed in eukaryotes using *in vitro* and cellular assays. From initial experiments, beyond the scope of this thesis, it was found that the eukaryotic ribosome could translate through the triazole backbone but yields from circRNAs in general were much lower than the prokaryotic results. Alongside this, substantial off-target binding of the antibodies was observed, likely targeting specific proteins in the blood contained within the rabbit lysate. Future work on eukaryotic systems needs to focus on improving circRNA expression and screening different antibody detection methods to specifically identify the correct peptides without off target detection. This is most likely to be improved by adaption of the sequence design including features such as condensed IRES

sequences.[217] After this it would be useful to begin further cellular assays to confirmed protein expression in human cell lines such as HEK293.

CircRNAs can be expressed in eukaryotic systems and have been applied to various applications such as micro RNA sponges, immunoactivation or protein regulation.[172] Another application is gene expression for cancer treatments. The BAX protein forms self-dimers that interact with the mitochondrial cell membrane, disrupting structures and bursting it open to release cytochrome c into the cytosol.[218,219] This results in cell death and could provide an effective treatment for cancers (Figure 4. 28).

The BAX pathway is activated by a second protein BIM, that binds to the BAX dimer to activate the complex. P. E. Czabotar *et. al.*[220,221] have shown that the full-length BIM protein can be stripped back to a short peptide sequence and maintains activity. If this BIM peptide could be expressed from a Tz-circRNA, then it presents an opportunity to use *in vivo* translation to treat cancers without the concern of DNA incorporation into the genome, a concern that has held back some DNA plasmid based therapeutics, especially techniques that intend to incorporate a cell killing peptide.[222]

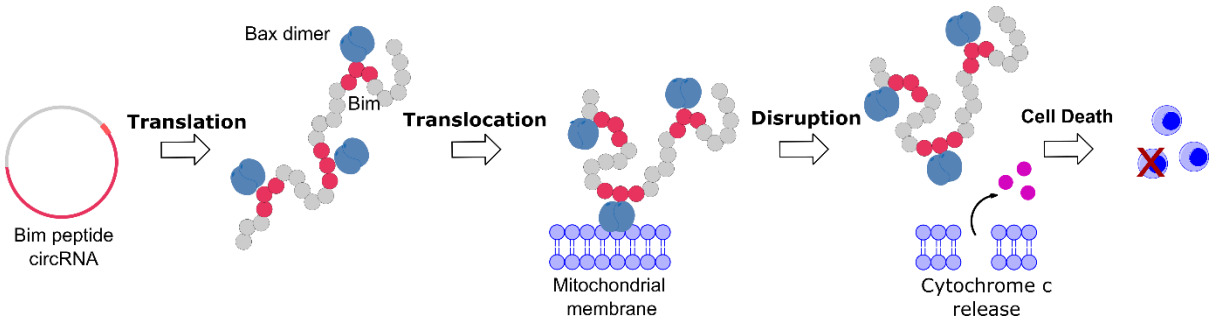



*Figure 4. 28 Outline of possible application of circRNAs for cancer cell treatment by targeting the BAX protein pathway. CircRNAs encode a BIM peptide which in turn activates the BAX dimer, triggering disruption of the mitochondrial membrane. Breakdown of the membrane releases cytochrome c into the cytosol which causes cell death.*

## 4.13 Summary of translation from Tz-circRNA

The aim of this work was to establish triazole backbones as a suitable modification for peptide expression. Triazole backbones incorporated into circRNAs allow templates to be produced by synthetic protocols without the need for enzymes or bacterial cell culture.

Triazole circRNAs were synthesised from modified RNA precursors with additional 5'-azide and 3'-alkyne groups. Using T4 DNA ligase and the CuAAC reaction, precursor RNAs were ligated together in the presence of a DNA splint to form the Tz-circRNA construct. Building on

initial experiments by A. Tsai, it was noted that reduced peptide expression from triazole-modified circRNAs could be caused by destabilisation of the ribosome, tRNA and mRNA complex during elongation. To test this, a second circRNA containing an LNA modification adjacent to the triazole was synthesised, with the aim to restore duplex stability.

The Tz and Tz-LNA circRNAs encoded with FLAG peptides were used as templates in a prokaryotic cell free *in vitro* translation. From these experiments it was shown that Tz-circRNAs can be used to express FLAG peptides but that the triazole templates yielded fewer peptides than the unmodified control. Tz-LNA-circRNA did not improve peptide translation and was found to be less effective as an mRNA template compared to both the unmodified and Tz-circRNA. This suggests that reduced translation of the modified backbone is not closely associated to duplex stability and is more likely linked to the structure and/or electronic properties of the backbone.

Having established that both triazole-containing templates can be expressed, our focus shifted to develop an optimised protocol for the synthesis and installation of 5'-azides required for the CuAAC reaction. Although converting the 5'-hydroxyl to an azide is an effective method, resulting products on long RNAs can be difficult to purify. To avoid this, a H-phosphonate approach was developed that allows a 5'-azide functionalised monomer to be added. The H-phosphonate small molecule was synthesised and incorporated onto the RNA oligonucleotide using a standard procedure. The coupling yields were high, and products were separated from the N-1 starting sequences. Compared to azide conversion, 5'-H-phosphonates offer an alternative method for installing azide functionality that enables simple production of modified oligonucleotides in large quantities.

After establishing an alternative method for the synthesis of azide-modified RNAs, a circRNA template that contained multiple triazole backbones was synthesised, removing the need for enzymatic ligations. The 2xTz-circRNA was synthesised using a strategy of overhanging DNA splints to control the ligation site. *In vitro* translation of this template showed that it was able to yield a similar amount of peptide as circRNA with a single triazole modification. However, quantification of these templates using dot blot pixel density analysis highlighted that the immunoblotting method was not quantitative and often images were oversaturated. To provide quantitative analysis, a Tz-circRNA with HA-tag epitope was synthesised, and an ELISA assay was shown to yield ~50% as much peptide as the unmodified control, as a best-case

scenario. In summary, it was shown that modified triazole backbones can be incorporated into circular mRNA templates and be successfully translated by the prokaryotic ribosome to produce the correct peptide sequence. The triazole reduces peptide expression, but in general yields 20–50% as much peptide compared to the unmodified control. Removing the need for enzymes to synthesise these templates, moves Tz-circRNA closer to becoming a viable commercial therapeutic, and future applications could allow templates to be used for gene synthesis or cancer treatments.

# 5 Conclusions and future work

Modified backbones are versatile oligonucleotide modifications that can improve therapeutic efficacy, adapt to various duplex structures and are compatible with many biological mechanisms. This thesis outlines a synthesis, a structure and an application of three modified backbones. Within this is detailed the synthesis and biophysical analysis of carbamate-LNAs, the crystal structures of amide-LNAs and the translation of triazole linkages within circular RNAs.

Previous reports suggested that the carbamate backbone was not suitable for therapeutic applications due to its duplex destabilising effects. However, they were found to have good enzymatic resistance to nuclease degradation, while being synthetically easy to produce. The aim of this work was to combine carbamate backbones with the duplex stabilising LNA modification, to produce a new template for therapeutic applications. Analysis of LNA-carbamate modifications showed that when $CBM_1$ backbones were flanked by two LNAs, RNA target affinity could be restored to a similar value as an unmodified duplex. In general, it was found that LNA addition to the 5'-side of the linkage destabilised the carbamate backbone, but that addition of more LNAs was able to mitigate these effects.

Biophysical analysis of LNA-carbamates using circular dichroism showed that global duplex structure remained largely unperturbed and enzymatic digestion was reduced when carbamate and LNA modifications were combined. In summary, this work concludes that the LNA-$CBM_1$-LNA modification produces the best template design for therapeutic use. Compared to the carbamate, LNA-carbamates now present suitable candidates for antisense applications.

The next phase of this project should focus on developing antisense assays to confirm *in vitro* efficacy of the LNA-carbamate modification. By incorporating our modifications into drug sequences, it will be possible to develop assays that allow direct comparison between the CBM-LNAs and approved FDA therapies. Building upon this, it would be interesting to begin cell uptake studies and quantify efficacy using luciferase splice switching assays, before moving toward complex *in vitro* mouse models.

Previous biophysical analysis of combined LNA and amide backbones has shown that they are duplex stabilising modifications, with potential uses as therapeutic templates. Building upon this work, the aim of the study in chapter 3 was to analyse the structural properties of LNA-amide modifications using X-ray diffraction, and to show that they are good structural mimics to replace phosphodiesters. In addition to this, the study aimed to suggest structural features that contribute to the duplex stabilising nature of a modified backbone.

A set of amide and LNA-amide modified DNA:RNA duplexes were crystallised using sitting drop vapour diffusion and analysed by XRD. As a result of the crystal screens, three modified duplexes were obtained with either an amide, LNA-amide or LNA-Am-LNA modification. In addition to this, an unmodified duplex was also crystallised, and all structures were refined to a resolution between 2.5-2.8 Å.

Modified oligonucleotides were analysed to assess the structural influence of combined LNA and amide modifications upon the duplex form. Superposition of each modified duplex showed that global structure remained largely unperturbed, but minor structural deviations were introduced by the addition of the LNA-amide and LNA-Am-LNA modifications, localised around the modification site.

In each modified structure the amide backbone was found to be orientated into the major groove with torsion angles associated to a *trans* conformation. In general, all modifications were well tolerated within the duplex and the highly modified LNA-Am-LNA step caused minimal duplex disruption.

LNA and amides are complementary modifications that have minimal restriction upon one another. Due to the length and flexibility of the amide backbone, it can adapt to the conformational needs of the duplex form. Addition of the LNA does not perturb or strain these structures and vice versa amides do not restrict the stabilising nature of the LNA. When designing future modified backbones, structures which mimic a phosphodiester linkage provide the best templates for therapeutic designs. The backbone should be highly flexible to allow it to adapt to the structural needs of the duplex and be similar in length as a phosphodiester. In summary of LNA-amide modifications, this work concludes that they are structurally good mimics as a phosphodiester, which results in minimal duplex disruption. When this is combined with their biophysical properties: including charge neutrality, duplex

stabilisation and enzymatic stability, they present promising candidates for therapeutic research.

In the final chapter of this thesis, the aim was to translate a peptide sequence from a circRNA that contains a triazole modified backbone. If successful, this would provide a fully synthetic procedure to create circRNAs without enzymatic tools, plasmid preparation or bacterial cell culture. As a result, the procedure would be scalable for therapeutic trials or for *in vitro* assays.

Preliminary studies from within our group suggested that there could be a relationship between mRNA:tRNA duplex stability and translation performance of the triazole backbone. Therefore, to investigate this, two triazole templates were designed, one with a triazole adjacent to canonical RNAs and a second with an adjacent LNA modification.

RNAs were synthesised using the phosphoramidite method and manually modified with either a 5ˈ-azide or 3ˈ-alkyne to form the functionalised precursor RNAs. The precursor RNAs were ligated using the CuAAC "click" reaction or T4 DNA ligase with a DNA splint to obtain a 126 nt linear RNA. A second ligation was then performed to create the circular RNA which contained a triazole (Tz-circRNA) or triazole-LNA (Tz-LNA-circRNA) modification. Circular RNA templates were subjected to cell free prokaryotic protein expression and analysed by immunoblotting techniques.

Western blot analysis showed that both the Tz-circRNA and Tz-LNA-circRNA were able to successfully be translated by the prokaryotic ribosome to produce peptides. However, translation yield from the modified Tz-circRNA was less than the unmodified control. In addition to this, combination with LNA did not restore expression yields compare to the unmodified template, and in fact performed worse than Tz-circRNA. These results conclude that translation affinity for the triazole was less reliant on duplex stability, and instead dictated by a modifications ability to act as a good phosphodiester bioisostere.

Following on from these results, a circRNA template containing two triazole linkages (2xTz-circRNA) was synthesised. Precursor RNAs with 3ˈ-alkynes were synthesised by the phosphoramidite method followed by addition of a 5ˈ-azide H-phosphonate monomer to functionalise the 5ˈ of the RNA. CircRNAs were synthesised using the click reaction, implementing DNA splints that extended beyond the RNA sequence to prevent self-cyclisation

of the precursor RNAs. Again, using a cell free translation experiment, the 2xTz-circRNA was used as a template and analysed by western blot. Results showed that addition of multiple triazoles did not impact translation compared to the Tz-circRNA template which demonstrates that fully synthetic circRNAs can be used as translation templates.

After providing a proof of concept, the next aim was to quantify the amount of peptide produced from the Tz-circRNA template relative to the unmodified control. Using an ELISA assay, triazole containing templates were measured in a best-case scenario to express up to 50% as much peptide as the unmodified control. ELISA measurements obtained in serial dilution, suggested that Tz-circRNAs can express between 20-50% compared to an unmodified circRNA. Although promising, some discrepancies in quantification were observed. Analysis by the dot blot assay suggests that peptide yields were much lower and quantified between 5-10%. Although the measurements were deemed inaccurate due to oversaturation, it would be ideal to confirm the ELISA quantification with a third method that does not rely on immunoblotting techniques.

In summary of peptide expression from triazole modified circRNAs, this work shows that Tz-circRNAs can be produced without enzymes by fully synthetic means. It is also shown that the Tz-circRNAs can be used to express peptides greater than three amino acids in length. Building upon this work, future experiments should focus on developing Tz-circRNAs for eukaryotic translation and explore possible applications such as cancer treatments.

In this thesis we have presented three unique projects developing the synthesis, analysis and applications of modified backbones within oligonucleotides. Based on this work significant contribution has been made towards the synthesis and biophysical characterisation of CBM-LNA backbones, which now present interesting candidates as antisense modifications. Substantial insight has been gathered into the structural influence that LNA-amides have upon oligonucleotide duplexes, and it was shown that both sugar and backbone modifications can be complementary additions. Incorporation of triazole backbones into circular RNAs showed that modified backbones can be successfully translated into extended peptide sequences. With continued development of this work, triazole circular RNA have the potential to make significant contributions to the field of gene therapy.

# 6 References


1 C. Erwin, L. Rakoma and C. Green, *J. Biol. Chem.*, 1952, **195**, 155–160.

2 F. Crick and J. Watson, *Nature*, 1953, **171**, 737–738.

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

4 R. E. Franklin and R. G. Gosling, *Nature*, 1953, **171**, 740–741.

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

6 S. Diekmann, *J. Mol. Biol.*, 1989, **205**, 787–791.

7 C. Calladine, H. Drew, B. Luisi and A. Travers, *Understanding DNA*, Elsevier Academic Press, Third., 2004.

8 R. E. Dickerson, *Int. Tables Crystallogr.*, 2012, **Vol F**, 766–773.

9 P. S. Klosterman, S. A. Shah and T. A. Steitz, *Biochemistry*, 1999, **38**, 14784–14792.

10 H. R. Drew, R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura and R. E. Dickerson, *Proc. Natl. Acad. Sci. U. S. A.*, 1981, **78**, 2179–2183.

11 Z. Luo, M. Dauter and Z. Dauter, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2014, **70**, 1790–1800.

12 A. Rich and S. Zhang, *Nat. Rev. Genet.*, 2003, **4**, 566–572.

13 M. H. Caruthers, *Science (80-. ).*, 1985, **230**, 281–285.

14 J. Jankowska, M. Sobkowski, J. Stawiński and A. Kraszewski, *Tetrahedron Lett.*, 1994, **35**, 3355–3358.

15 J. Stawinski and R. Strömberg, *Di- and Oligonucleotide Synthesis Using H-Phosphonate Chemistry. Oligonucleotide Synthesis*, Humana Press, 2005, vol. 288.

16 K. Nakamoto, N. Abe, G. Tsuji, Y. Kimura, F. Tomoike, Y. Shimizu and H. Abe, *Chem. Commun.*, 2020, **56**, 6217–6220.

17 I. R. Lehnman, *Science (80-. ).*, 1974, **186**, 790–797.

18 V. V. Rostovtsev, L. G. Green, V. V. Fokin and K. B. Sharpless, *Angew. Chemie Int. Ed.*, 2002, **41**, 2596–2599.

19 V. Hong, S. Presolski, C. Ma and M. Finn, *Angew. Chemie Int. Ed.*, 2009, **121**, 10063–10067.

20 G. R. Abel, Z. A. Calabrese, J. Ayco, J. E. Hein and T. Ye, *Bioconjug. Chem.*, 2016, **27**, 698–704.

21 R. W. Woody, *Methods Enzymol.*, 1995, **246**, 34–71.

22 J. Kypr, I. Kejnovská, D. Renčiuk and M. Vorlíč, *Nucleic Acids Res.*, 2009, **37**, 1713–1725.

23 B. Alberts, *Molecular biology of the cell*, Garland Science, 6th edn., 2015.

24 J. M. Walker and R. Rapley, *Molecular Biomethods Handbook*, Humana Press, 2nd edn., 2008.

25 A. W. Murray and M. R. Atkinson, *Biochemistry*, 1968, **7**, 4023–4029.

26 F. Eckstein, *Nucleic Acid Ther.*, 2014, **24**, 374–387.

27 B. V. L. Potter, B. A. Connolly and F. Eckstein, *Biochemistry*, 1983, **22**, 1369–1377.

28 W. B. Wan, M. T. Migawa, G. Vasquez, H. M. Murray, J. G. Nichols, H. Gaus, A. Berdeja, S. Lee, C. E. Hart, W. F. Lima, E. E. Swayze and P. P. Seth, *Nucleic Acids Res.*, 2014, **42**, 13456–13468.

29 S. H. Eom, J. Wang and T. A. Steitz, *Nature*, 1996, **382**, 278–81.

30 Y. Nukaga, K. Yamada, T. Ogata, N. Oka and T. Wada, *J. Org. Chem.*, 2012, **77**, 7913–7922.

31 N. Iwamoto, D. C. D. Butler, N. Svrzikapa, S. Mohapatra, I. Zlatev, Di. W. Y. Sah, Meena, S. M. Standley, G. Lu, L. H. Apponi, M. Frank-Kamenetsky, J. J. Zhang, C. Vargeese and G. L. Verdine, *Nat. Biotechnol.*, 2017, **35**, 845–851.

32 A. Waldner, A. De Mesmaeker and J. Lebreton, *Bioorganic Med. Chem. Lett.*, 1994, **4**, 405–408.

33 A. De Mesmaeker, C. Jouanno, R. M. Wolf and S. Wendeborn, *Bioorg. Med. Chem. Lett.*, 1997, **7**, 447–452.

34 R. Lucas, V. Neto, A. Hadj Bouazza, R. Zerrouki, R. Granet, P. Krausz and Y. Champavier, *Tetrahedron Lett.*, 2008, **49**, 1004–1007.

35 X. L. Zhao, B. C. Chen, J. C. Han, L. Wei and X. Ben Pan, *Sci. Rep.*, 2015, **5**, 1–11.

36 S. Obika, M. Onoda, J. Andoh, T. Imanishi, K. Morita and M. Koizumi, *Chem. Commun.*, 2001, 1992–1993.

37 S. Obika, D. Nanbu, Y. Hari, K. Morio, Y. In, T. Ishida and T. Imanishi, *Tetrahedron Lett.*, 1997, **38**, 8735–8738.

38 A. A. Koshkin, S. K. Singh, P. Nielsen, V. K. Rajwanshi, R. Kumar, M. Meldgaard, C. E. Olsen and

J. Wengel, *Tetrahedron*, 1998, **54**, 3607–3630.

39 S. K. Singh, P. Nielsen, A. A. Koshkin and J. Wengel, *Chem. Commun.*, 1998, **0**, 455–456.

40 M. Petersen, C. B. Nielsen, K. E. Nielsen, G. A. Jensen, K. Bondensgaard, S. K. Singh, V. K. Rajwanshi, A. A. Koshkin, B. M. Dahl, J. Wengel and J. P. Jacobsen, *J. Mol. Recognit.*, 2000, **13**, 44–53.

41 M. Petersen, K. Bondensgaard, J. Wengel and J. Peter Jacobsen, *J. Am. Chem. Soc.*, 2002, **124**, 5974–5982.

42 J. S. Jepsen, M. D. Sørensen and J. Wengel, *Oligonucleotides*, 2004, **14**, 130–146.

43 S. A. Burel, C. E. Hart, P. Cauntay, J. Hsiao, T. Machemer, M. Katz, A. Watt, H. H. Bui, H. Younis, M. Sabripour, S. M. Freier, G. Hung, A. Dan, T. P. Prakash, P. P. Seth, E. E. Swayze, C. F. Bennett, S. T. Crooke and S. P. Henry, *Nucleic Acids Res.*, 2015, **44**, 2093–2109.

44 C. Wahlestedt, P. Salmi, L. Good, J. Kela, T. Johnsson, T. Hökfelt, C. Broberger, F. Porreca, J. Lai, K. Ren, M. Ossipov, A. Koshkin, N. Jakobsen, J. Skouv, H. Oerum, M. H. Jacobsen and J. Wengel, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 5633–5638.

45 K. Fluiter, O. R. F. Mook, J. Vreijling, N. Langkjær, T. Højland, J. Wengel and F. Baas, *Mol. Biosyst.*, 2009, **5**, 838–843.

46 H. Ørum, A. Wolter and L. Kongsbak, *Lett. Pept. Sci.*, 2003, **10**, 325–334.

47 M. Kuwahara, S. Obika, J. I. Nagashima, Y. Ohta, Y. Suto, H. Ozaki, H. Sawai and T. Imanishi, *Nucleic Acids Res.*, 2008, **36**, 4257–4265.

48 C. R. Cromwell, K. Sung, J. Park, A. R. Krysler, J. Jovel, S. K. Kim and B. P. Hubbard, *Nat. Commun.*, 2018, **9**, 1448.

49 B. Souvignier, *Int. Tables Crystallogr.*, 2016, **A**, 2–11.

50 B. Souvignier, *Int. Tables Crystallogr.*, 2016, **A**, 22–41.

51 EMBL-EBI, PDBe brings electron-density viewing to the masses, https://bit.ly/2VCIMMG, (accessed 20 May 2020).

52 B. Rupp, *Biomolecular crystallography principles practice and application to structural biology*, CRC Press, 2010.

53 G. L. Taylor, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2010, **66**, 325–338.

54 W. A. Hendrickson, *Q. Rev. Biophys.*, 2014, **47**, 49–93.

55 W. A. Hendrickson, J. R. Horton and D. M. LeMaster, *EMBO J.*, 1990, **9**, 1665–1672.

56 W. A. Hendrickson and M. M. Teeter, *Nature*, 1981, **290**, 107–113.

57 M. R. Peterson, S. J. Harrop, S. M. McSweeney, G. A. Leonard, A. W. Thompson, W. N. Hunter and J. R. Helliwell, *J. Synchrotron Radiat.*, 1996, **3**, 24–34.

58 A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read, *J. Appl. Crystallogr.*, 2007, **40**, 658–674.

59 M. L. Stephenson and P. C. Zamecnik, *Proc. Natl. Acad. Sci.*, 1978, **75**, 285–288.

60 J. B. Opalinska and A. M. Gewirtz, *Nat. Rev. Drug Discov.*, 2002, **1**, 503–514.

61 C. Rinaldi and M. J. A. Wood, *Nat. Rev. Neurol.*, 2018, **14**, 9–22.

62 A. Khvorova and J. K. Watts, *Nat. Biotechnol.*, 2017, **35**, 238–248.

63 R. Kole, A. R. Krainer and S. Altman, *Nat. Rev. Drug Discov.*, 2012, **11**, 125–140.

64 B. F. Baker, S. S. Lot, T. P. Condon, S. Cheng-Flournoy, E. A. Lesnik, H. M. Sasmor and C. F. Bennett, *J. Biol. Chem.*, 1997, **272**, 11994–12000.

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

66 V. K. Sharma, R. K. Sharma and S. K. Singh, *Medchemcomm*, 2014, **5**, 1454–1471.

67 C. F. Bennett, B. F. Baker, N. Pham, E. Swayze and R. S. Geary, *Annu. Rev. Pharmacol. Toxicol*, 2017, **57**, 81–105.

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

69 S. T. Crooke, *Nucleic Acid Ther.*, 2017, **27**, 70–77.

70 M. A. Havens and M. L. Hastings, *Nucleic Acids Res.*, 2016, **44**, 6549–6563.

71 D. L. Black, *Annu. Rev. Biochem.*, 2003, **72**, 291–336.

72 K. E. Davies and K. J. Nowak, *Nat. Rev. Mol. Cell Biol.*, 2006, **7**, 762–773.

73 F. Muntoni and M. J. A. Wood, *Nat. Rev. Drug Discov.*, 2011, **10**, 621–37.

74 M. Koenig, A. H. Beggs, M. Moyer, S. Scherpf, K. Heindrich, T. Bettecken, G. Meng, C. R. Müller, M. Lindlöf, H. Kaariainen, A. de la Chapelle, A. Kiuru, M. L. Savontaus, H. Gilgenkrantz, D. Récan, J. Chelly, J. C. Kaplan, A. E. Covone, N. Archidiacono, G. Romeo, S. Liechti-Gailati, V. Schneider,

S. Braga, H. Moser, B. T. Darras, R. Murphy, U. Francke, J. D. Chen, G. Morgan, M. Denton, C. R. Greenberg, K. Wrogemann, L. A. J. Blonden, H. M. B. van Paassen, G. J. B. van Ommen and L. M. Kunkel, *Am. J. Hum. Genet*, 1989, **45**, 498–506.

75 S. M. Elbashir, J. Harborth, W. Lendeckel, A. Yalcin, K. Weber and T. Tuschl, *Nature*, 2001, **411**, 494–498.

76 G. F. Deleavey and M. J. Damha, *Chem. Biol.*, 2012, **19**, 937–954.

77 D. Mutisya, C. Selvam, B. D. Lunstad, P. S. Pallan, A. Haas, D. Leake, M. Egli and E. Rozners, *Nucleic Acids Res.*, 2014, **42**, 6542–6551.

78 M. M. Janas, M. K. Schlegel, C. E. Harbison, V. O. Yilmaz, Y. Jiang, R. Parmar, I. Zlatev, A. Castoreno, H. Xu, S. Shulga-Morskaya, K. G. Rajeev, M. Manoharan, N. D. Keirstead, M. A. Maier and V. Jadhav, *Nat. Commun.*, 2018, **9**, 723.

79 I. A. Roundtree, M. E. Evans, T. Pan and C. He, *Cell*, 2017, **169**, 1187–1200.

80 L. L. Cummins, S. R. Owens, L. M. Risen, E. A. Lesnik, S. M. Freier, D. Mc Gee, C. J. Cook and P. D. Cook, *Nucleic Acids Res.*, 1995, **23**, 2019–2024.

81 B. P. Monia, J. F. Johnston, H. Sasmor and L. L. Cummins, *J. Biol. Chem.*, 1996, **271**, 14533–14540.

82 P. Martin, *Helv. Chim. Acta*, 1995, **78**, 486–504.

83 R. S. Geary, B. F. Baker and S. T. Crooke, *Clin. Pharmacokinet.*, 2015, **54**, 133–146.

84 E. Wong and T. Goldberg, *Drug Froecast*, 2014, **39**, 119–122.

85 E. E. Swayze, A. M. Siwkowski, E. V Wancewicz, M. T. Migawa, T. K. Wyrzykiewicz, G. Hung, B. P. Monia and C. F. Bennett, *Nucleic Acids Res.*, 2007, **35**, 687–700.

86 J. Kurreck, E. Wyszko, C. Gillen and A. Erdmann, Volker, *Nucleic Acids Res.*, 2002, **30**, 1911–1918.

87 D. A. Braasch, S. Jensen, Y. Liu, K. Kaur, K. Arar, M. A. White and D. R. Corey, *Biochemistry*, 2003, **42**, 7967–7975.

88 M. A. Campbell and J. Wengel, *Chem. Soc. Rev.*, 2011, **40**, 5680–5689.

89 M. B. Laursen, M. M. Pakula, S. Gao, K. Fluiter, O. R. Mook, F. Baas, N. Langklær, S. L. Wengel, J. Wengel, J. Kjems and J. B. Bramsen, *Mol. Biosyst.*, 2010, **6**, 862–870.

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

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

92 S. J. Keam, *Drugs*, 2018, **78**, 1371–1376.

93 A. M. Varizhuk, D. N. Kaluzhny, R. A. Novikov, A. O. Chizhov, I. P. Smirnov, A. N. Chuvilin, O. N. Tatarinova, G. Y. Fisunov, G. E. Pozmogova and V. L. Florentiev, *J. Org. Chem.*, 2013, **78**, 5964–5969.

94 P. Kumar, A. H. El-Sagheer, L. Truong and T. Brown, *Chem. Commun.*, 2017, **53**, 8910–8913.

95 P. Kumar, L. Truong, Y. R. Baker, A. H. El-Sagheer and T. Brown, *ACS Omega*, 2018, **3**, 6976–6987.

96 V. Sharma, S. Singh, P. M. Krishnamurthy, J. Alterman, R. Haraszti, A. Khvorova, A. K. Prasad and J. Watts, *Chem. Commun.*, 2017, **53**, 8906–8909.

97 E. Rozners, D. Katkevica, E. Bizdena and R. Strömberg, *J. Am. Chem. Soc.*, 2003, **125**, 12125–12136.

98 A. De Mesmaeker, S. Wendeborn, C. Jouanno, V. Fritsch and R. M. Wolf, *Bioorganic Med. Chem. Lett.*, 1997, **7**, 1869–1874.

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

100 H. M. Moulton, M. H. Nelson, S. A. Hatlevig, M. T. Reddy and P. L. Iversen, *Bioconjug. Chem.*, 2004, **15**, 290–299.

101 M. J. Palframan, R. D. Alharthy, P. K. Powalowska and C. J. Hayes, *Org. Biomol. Chem.*, 2016, **6**, 3112–3119.

102 A. Aartsma-Rus and A. M. Krieg, *Nucleic Acid Ther.*, 2017, **27**, 1–3.

103 K. R. Q. Lim, R. Maruyama and T. Yokota, *Drug Des. Devel. Ther.*, 2017, **11**, 533–545.

104 A. De Mesmaeker, A. Waldner, S. Wendeborn and R. M. Wolf, *Pure Appl. Chem*, 1997, **69**, 437–440.

105 B. S. Ross, M. Han and V. T. Ravikumar, *Nucleosides, Nucleotides and Nucleic Acids*, 2006, **25**, 765–770.

106 T. H. lsamu Yamamoto, Mitsuo Sekine, *J. Chem. Soc. Perkin Trans. 1*, 1980, **0**, 306–310.

107 D. Pan, J. Sun, H. Jin, Y. Li, L. Li, Y. Wu, L. Zhang, Z. Yang and W.-L. Lu, *Chem. Commun.*, 2015,

**51**, 469–472.

108 S. Obika, O. Nakagawa, A. Hiroto, Y. Hari and T. Imanishi, *Chem. Commun.*, 2003, **3**, 2202–2203.

109 W. S. Mungall and J. K. Kaiser, *J. Org. Chem.*, 1977, **42**, 703–706.

110 A. A. Koshkin, J. Fensholdt, H. M. Pfundheller and C. Lomholt, *J. Org. Chem.*, 2001, **66**, 8504–8512.

111 C. Thorpe, S. Epple, B. Woods, A. H. El-Sagheer and T. Brown, *Org. Biomol. Chem.*, 2019, **17**, 5341–5348.

112 R. Eisenhuth and C. Richert, *J. Org. Chem.*, 2009, **74**, 26–37.

113 S. De, E. Groaz and P. Herdewijn, *European J. Org. Chem.*, 2014, **2014**, 2322–2348.

114 K. Satou, Y. Komatsu, T. Torizawa, K. Kato, I. Shimada, O. Nikaido and E. Ohtsuka, *Tetrahedron Lett.*, 2000, **41**, 2175–2179.

115 K. M. Fries, C. Joswig and R. F. Borch, *J. Med. Chem*, 1996, **38**, 2672–2680.

116 E. Utagawa, M. Sekine and K. Seio, *J. Org. Chem.*, 2006, **71**, 7668–7677.

117 S. M. Freier and K.-H. Altmann, *Nucleic Acids Res.*, 1997, **25**, 4429–4443.

118 W. Yang, *Q. Rev. Biophys.*, 2011, **44**, 1–93.

119 B. R. Meade, K. Gogoi, A. S. Hamil, C. Palm-Apergi, A. Van Den Berg, J. C. Hagopian, A. D. Springer, A. Eguchi, A. D. Kacsinta, C. F. Dowdy, A. Presente, P. Lönn, M. Kaulich, N. Yoshioka, E. Gros, X. S. Cui and S. F. Dowdy, *Nat. Biotechnol.*, 2014, **32**, 1256–1261.

120 A. De mesmaeker, A. Waldner, J. Lebreton, P. Hoffmann, V. Fritsch, R. M. Wolf and S. M. Freier, *Angew. Chemie-International Ed. English*, 1994, **33**, 226–229.

121 A. J. Debacker, V. K. Sharma, P. Meda Krishnamurthy, D. O'Reilly, R. Greenhill and J. K. Watts, *Biochemistry*, 2019, **58**, 582–589.

122 I. Idziak, G. Just, M. J. Damha and P. A. Giannaris, *Tetrahedron Lett.*, 1993, **34**, 5417–5420.

123 A. Lauritsen and J. Wengel, *Chem. Commun.*, 2002, **5**, 530–531.

124 P. S. Pallan, P. Von Matt, C. J. Wilds, K. H. Altmann and M. Egli, *Biochemistry*, 2006, **45**, 8048–8057.

125 C. Selvam, S. Thomas, J. Abbott, S. D. Kennedy and E. Rozners, *Angew. Chemie Int. Ed.*, 2011,

**50**, 2068–2070.

126 D. Mutisya, T. Hardcastle, S. K. Cheruiyot, P. S. Pallan, S. D. Kennedy, M. Egli, M. L. Kelley, A. Van Brabant Smith and E. Rozners, *Nucleic Acids Res.*, 2017, **45**, 8142–8155.

127 S. Epple, C. Thorpe, Y. R. Baker, A. H. El-Sagheer and T. Brown, *Chem. Commun.*, 2020, **56**, 5496–5499.

128 A. Shivalingam, A. E. S. Tyburn, A. H. El-Sagheer and T. Brown, *J. Am. Chem. Soc.*, 2017, **139**, 1575–1583.

129 A. De Mesmaeker, C. Lesueur, M.-O. Bévièrre, A. Waldner, V. Fritsch and R. M. Wolf, *Angew. Chemie Int. Ed. English*, 1996, **35**, 2790–2794.

130 V. Kotikam and E. Rozners, *Org. Lett.*, 2017, **19**, 4122–4125.

131 J. B. Sumner, *J. Biol. Chem.*, 1926, **69**, 435–441.

132 K. Hamaguchi and H. Sakai, *J. Biochem.*, 1965, **57**, 721–732.

133 C. Blake, D. Koenig, G. Mair, A. North, D. Phillips and V. Sarma, *Nature*, 1965, **206**, 757–761.

134 R. Wing, H. Drew, T. Takano, C. Broka, S. Tanaka, K. Itakura and R. E. Dickerson, *Nature*, 1980, **287**, 755–758.

135 P. M. Keane, J. P. Hall, F. E. Poynton, B. C. Poulsen, S. P. Gurung, I. P. Clark, I. V. Sazanovich, M. Towrie, T. Gunnlaugsson, S. J. Quinn, C. J. Cardin and J. M. Kelly, *Chem. - A Eur. J.*, 2017, **23**, 10344–10351.

136 A. McPherson and B. Cudney, *Acta Crystallogr. Sect. FStructural Biol. Commun.*, 2014, **70**, 1445–1467.

137 G. Winter, D. G. Waterman, J. M. Parkhurst, A. S. Brewster, R. J. Gildea, M. Gerstel, L. Fuentes-Montero, M. Vollmar, T. Michels-Clark, I. D. Young, N. K. Sauter and G. Evans, *Acta Crystallogr. Sect. D Struct. Biol.*, 2018, **74**, 85–97.

138 P. A. Karplus and K. Diederichs, *Science (80-. ).*, 2012, **336**, 1030–1033.

139 P. R. Evans and G. N. Murshudov, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2013, **69**, 1204–1214.

140 C. Abergel, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2013, **69**, 2167–2173.

141 G. Taylor, *Acta Cryst*, 2003, **59**, 1881–1890.

142 A. Wlodawer, W. Minor, Z. Dauter and M. Jaskolski, *FEBS J.*, 2008, **275**, 1–21.

143 M. L. Kopka, L. Lavelle, G. W. Han, H.-L. Ng and R. E. Dickerson, *J. Mol. Biol.*, 2003, **334**, 653–665.

144 Gye Won Han, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2001, **57**, 213–218.

145 S. Walsh, A. H. El-Sagheer and T. Brown, *Chem. Sci.*, 2018, **9**, 7681–7687.

146 R. W. Roberts and D. M. Crothers, *Science (80-. ).*, 1992, **258**, 1463–1466.

147 C. Sauter, J. D. Ng, B. Lorber, G. Keith, P. Brion, M. W. Hosseini, J. M. Lehn and R. Giegé, *J. Cryst. Growth*, 1999, **196**, 365–376.

148 A. A. Lebedev, P. Young, M. N. Isupov, O. V Moroz, A. A. Vagin and G. N. Murshudov, *Res. Pap. Acta Cryst*, 2012, **68**, 431–440.

149 A. Jabs, M. S. Weiss and R. Hilgenfeld, *J. Mol. Biol.*, 1999, **286**, 291–304.

150 D. E. Stewart, A. Sarkar and J. E. Wampler, *J. Mol. Biol.*, 1990, **214**, 253–260.

151 H. Berman, K. Henrick and H. Nakamura, *Nat. Struct. Biol.*, 2003, **10**, 980.

152 J. I. Gyi, A. N. Lane, G. L. Conn and T. Brown, *Biochemistry*, 1998, **37**, 73–80.

153 R. E. Dickerson, H. R. Drew, B. N. Conner, R. M. Wing, A. V Fratini and M. L. Kopka, *Science (80-. ).*, 1982, **216**, 475–483.

154 C. B. Nielsen, S. K. Singh, J. Wengel and J. P. Jacobsen, *J. Biomol. Struct. Dyn.*, 1999, **17**, 175–191.

155 B. Capel, A. Swain, S. Nicolis, A. Hacker, M. Walter, P. Koopman, P. Goodfellow and R. Lovell-Badge, *Cell*, 1993, **73**, 1019–1030.

156 J. Salzman, C. Gawad, P. L. Wang, N. Lacayo and P. O. Brown, *PLoS One*, 2012, **7**, 1–12.

157 S. M. Fica, N. Tuttle, T. Novak, N. S. Li, J. Lu, P. Koodathingal, Q. Dai, J. P. Staley and J. A. Piccirilli, *Nature*, 2013, **503**, 229–234.

158 M. M. Konarska, P. J. Grabowski, R. A. Padgett and P. A. Sharp, *Nature*, 1985, **313**, 552–557.

159 L. Szabo and J. Salzman, *Nat. Rev. Genet.*, 2016, **17**, 679–692.

160 T. G. Truscott and R. S. Sinclair, *Advances in experimental medicine and biology - Circular RNAs*, Springer Singapore, 1983, vol. 44.

161 A. Costello, N. T. Lao, N. Barron and M. Clynes, *Trends Biotechnol.*, 2020, **38**, 217–230.

162 A. Rybak-Wolf, C. Stottmeister, P. Glažar, M. Jens, N. Pino, M. Hanan, M. Behm, O. Bartok, R. Ashwal-Fluss, M. Herzog, L. Schreyer, P. Papavasileiou, A. Ivanov, M. Öhman, D. Refojo, S. Kadener and N. Rajewsky, *Mol. Cell*, 2014, **58**, 870–885.

163 L. L. Chen, *Nat. Rev. Mol. Cell Biol.*, 2016, **17**, 205–211.

164 W. R. Jeck, J. A. Sorrentino, K. Wang, M. K. Slevin, C. E. Burd, J. Liu, W. F. Marzluff and N. E. Sharpless, *RNA*, 2013, **19**, 141–157.

165 L. Szabo, R. Morey, N. J. Palpant, P. L. Wang, N. Afari, C. Jiang, M. M. Parast, C. E. Murry, L. C. Laurent and J. Salzman, *Genome Biol.*, 2015, **16**, 1–26.

166 D. J. Wright, J. L. Rice, D. M. Yanker and B. M. Znosko, *Biochemistry*, 2007, **46**, 4625–4634.

167 A. Ivanov, S. Memczak, E. Wyler, F. Torti, H. T. Porath, M. R. Orejuela, M. Piechotta, E. Y. Levanon, M. Landthaler, C. Dieterich and N. Rajewsky, *Cell Rep.*, 2015, **10**, 170–177.

168 S. P. Barrett and J. Salzman, *Dev.*, 2016, **143**, 1838–1847.

169 R. Ashwal-Fluss, M. Meyer, N. R. Pamudurti, A. Ivanov, O. Bartok, M. Hanan, N. Evantal, S. Memczak, N. Rajewsky and S. Kadener, *Mol. Cell*, 2014, **56**, 55–66.

170 E. Ford and M. Ares, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 3117–3121.

171 S. Petkovic and S. Müller, *Nucleic Acids Res.*, 2015, **43**, 2454–2465.

172 R. A. Wesselhoeft, P. S. Kowalski and D. G. Anderson, *Nat. Commun.*, 2018, **9**, 1–10.

173 Y. G. Chen, M. V. Kim, X. Chen, P. J. Batista, S. Aoyama, J. E. Wilusz, A. Iwasaki and H. Y. Chang, *Mol. Cell*, 2017, **67**, 228-238.e5.

174 C. Y. Chen and P. Sarnow, *Science (80-. ).*, 1995, **268**, 415–417.

175 L. Wang and D. E. Ruffner, *Nucleic Acids Res.*, 1998, **26**, 2502–2504.

176 R. J. Jackson, C. U. T. Hellen and T. V Pestova, *Nat. Rev. Mol. Cell Biol.*, 2010, **11**, 113–127.

177 A. G. Hinnebusch and J. R. Lorsch, *Cold Spring Harb. Perspect. Biol.*, 2012, **4**, 1–25.

178 M. Kozak, *J. Cell Biol.*, 1989, **108**, 229–241.

179 M. Kozak, 1987, **15**, 1543–1558.

180 J. Pelletier and N. Sonenberg, *Nature*, 1988, **334**, 320–325.

181 S. K. Jang, H. G. Kräusslich, M. J. Nicklin, G. M. Duke, A. C. Palmenberg and E. Wimmer, *J. Virol.*, 1988, **62**, 2636–2643.

182 M. Mokrejs, V. Vopálenský, O. Kolenaty, T. Masek, Z. Feketová, P. Sekyrová, B. Skaloudová, V. Kríz and M. Pospísek, *Nucleic Acids Res.*, 2006, **34**, 125–130.

183 S. De Breyne, Y. Yu, A. Unbehaun, T. V Pestova and C. U. T. Hellen, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 9197–9202.

184 A. C. Gingras, B. Raught and N. Sonenberg, *Annu. Rev. Biochem.*, 1999, **68**, 913–963.

185 Y. Yang and Z. Wang, *J. Mol. Cell Biol.*, 2019, **11**, 911–919.

186 M. M. Yusupov, G. Z. Yusupova, A. Baucom, K. Lieberman, T. N. Earnest, J. H. D. Cate and H. F. Noller, *Science (80-. ).*, 2001, **292**, 883–896.

187 B. Alberts, A. Johnson, J. Lewis, D. Morgan and and M. Raff, *Molecular Biology of The Cell*, Taylor and Francis, 6th edn., 2014.

188 L. Frolova, X. Le Goff, H. H. Rasmussen, S. Cheperegin, G. Drugeon, M. Kress, I. Arman, A. L. Haenni, J. E. Celis, M. Phllippe, J. Justesen and L. Kisselev, *Nature*, 1994, **372**, 701–703.

189 C. W. P. Donald Voet, Judith G. Voet, *Fundamentals of biochemistry : life at the molecular level*, Hoboken, New Jersey : Wiley, 5th edn., 2016.

190 J. Shine and L. Dalgarno, *J. Mol. Biol.*, 1973, **75**, 57–72.

191 V. Vimberg, A. Tats, M. Remm and T. Tenson, *BMC Mol. Biol.*, 2007, **8**, 1471–2199.

192 Y. Shimizu, A. Inoue, Y. Tomari, T. Suzuki, T. Yokogawa, K. Nishikawa and T. Ueda, *Nat. Biotechnol.*, 2001, **19**, 751–5.

193 A. Costello, N. T. Lao, N. Barron and M. Clynes, *Metab. Eng.*, 2019, **52**, 284–292.

194 J. H. Kim, S. R. Lee, L. H. Li, H. J. Park, J. H. Park, K. Y. Lee, M. K. Kim, B. A. Shin and S. Y. Choi, *PLoS One*, 2011, **6**, 1–8.

195 N. R. Pamudurti, O. Bartok, M. Jens, R. Ashwal-Fluss, C. Stottmeister, L. Ruhe, M. Hanan, E. Wyler, D. Perez-Hernandez, E. Ramberger, S. Shenzis, M. Samson, G. Dittmar, M. Landthaler, M. Chekulaeva, N. Rajewsky and S. Kadener, *Mol. Cell*, 2017, **66**, 9–21.

196 K. A. Underwood, J. R. Swartz and J. D. Puglisi, *Biotechnol. Bioeng.*, 2005, **91**, 425–435.

197 R. Perriman and M. Ares, *RNA*, 1998, **4**, 1047–1054.

198 N. Abe, M. Hiroshima, H. Maruyama, Y. Nakashima, Y. Nakano, A. Matsuda, Y. Sako, Y. Ito and H. Abe, *Angew. Chemie Int. Ed.*, 2013, **52**, 7004–7008.

199 I. Iost and M. Dreyfus, *EMBO J.*, 1995, **14**, 3252–3261.

200 N. Abe, K. Matsumoto, M. Nishihara, Y. Nakano, A. Shibata, H. Maruyama, S. Shuto, A. Matsuda, M. Yoshida, Y. Ito and H. Abe, *Sci. Rep.*, 2015, **5**, 16435.

201 H. Xiong and F. Seela, *Bioconjug. Chem.*, 2012, **23**, 1230–1243.

202 H. Isobe, T. Fujino, N. Yamazaki, M. Guillot-Nieckowski and E. Nakamura, *Org. Lett.*, 2008, **10**, 3729–3732.

203 Y. R. Baker, D. Traoré, P. Wanat, A. Tyburn, A. H. El-Sagheer and T. Brown, *Tetrahedron*, 2020, **76**, 5–9.

204 M. Kollaschinski, J. Sobotta, A. Schalk, T. Frischmuth, B. Graf and S. Serdjukow, *Bioconjug. Chem.*, 2020, **31**, 507–512.

205 A. H. El-Sagheer and T. Brown, *J. Am. Chem. Soc.*, 2009, **131**, 3958–3964.

206 A. P. Sanzone, A. H. El-Sagheer, T. Brown and A. Tavassoli, *Nucleic Acids Res.*, 2012, **40**, 10567–10575.

207 C. N. Birts, A. P. Sanzone, A. H. El-Sagheer, J. P. Blaydes, T. Brown and A. Tavassoli, *Angew. Chemie - Int. Ed.*, 2014, **53**, 2362–2365.

208 M. Kukwikila, N. Gale, A. H. El-Sagheer, T. Brown and A. Tavassoli, *Nat. Chem.*, 2017, **9**, 1089–1098.

209 T. Fujino, T. Suzuki, K. Okada, K. Kogashi, K. I. Yasumoto, K. Sogawa and H. Isobe, *J. Org. Chem.*, 2016, **81**, 8967–8976.

210 A. Tsai, *Assembly of Nucleic Acids with Site-Specific Modifications*, 2018.

211 M. Zuker, *Nucleic Acids Res.*, 2003, **31**, 3406–3415.

212 Promega, T4 DNA Ligase Protocol, https://bit.ly/3p6FcHK, (accessed 6 November 2020).

213 A. H. El-Sagheer, A. P. Sanzone, R. Gao, A. Tavassoli and T. Brown, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 11338–11343.

214 R. D. Ricker and A. Kaji, *Nucleic Acids Res.*, 1991, **19**, 6573–6578.

215 F. L. Lin, H. M. Hoyt, H. Van Halbeek, R. G. Bergman and C. R. Bertozzi, *J. Am. Chem. Soc.*, 2005,

**127**, 2686–2695.

216 G. H. Hakimelahi, Z. A. Proba and K. K. Ogilvie, *Tetrahedron Lett.*, 1981, **22**, 4775–4778.

217 G. C. Owens, S. A. Chappell, V. P. Mauro and G. M. Edelman, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 1471–1476.

218 J. G. Valero, L. Sancey, J. Kucharczak, Y. Guillemin, D. Gimenez, J. Prudent, G. Gillet, J. Salgado, J. L. Coll and A. Aouacheria, *J. Cell Sci.*, 2011, **124**, 556–564.

219 D. Mérino, M. Giam, P. D. Hughes, O. M. Siggs, K. Heger, L. A. O'Reilly, J. M. Adams, A. Strasser, E. F. Lee, W. D. Fairlie and P. Bouillet, *J. Cell Biol.*, 2009, **186**, 355–362.

220 P. E. Czabotar, D. Westphal, G. Dewson, S. Ma, C. Hockings, W. D. Fairlie, E. F. Lee, S. Yao, A. Y. Robin, B. J. Smith, D. C. S. Huang, R. M. Kluck, J. M. Adams and P. M. Colman, *Cell*, 2013, **152**, 519–531.

221 J. M. Brouwer, P. Lan, A. D. Cowan, J. P. Bernardini, R. W. Birkinshaw, M. F. van Delft, B. E. Sleebs, A. Y. Robin, A. Wardak, I. K. Tan, B. Reljic, E. F. Lee, W. D. Fairlie, M. J. Call, B. J. Smith, G. Dewson, G. Lessene, P. M. Colman and P. E. Czabotar, *Mol. Cell*, 2017, **68**, 659–672.

222 Z. Wang, P. J. Troilo, X. Wang, T. G. Griffiths, S. J. Pacchione, A. B. Barnum, L. B. Harper, C. J. Pauley, Z. Niu, L. Denisova, T. T. Follmer, G. Rizzuto, G. Ciliberto, E. Fattori, N. L. Monica, S. Manam and B. J. Ledwith, *Gene Ther.*, 2004, **11**, 711–721.

223 G. N. Murshudov Alexe, A. Vagin and E. J. Dodson, *Acta Cryst*, 1997, **53**, 240–255.

224 W. Kabsch, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2010, **66**, 133–144.

225 P. R. Evans, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2011, **67**, 282–292.

226 M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin and K. S. Wilson, *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 2011, **67**, 235–242.

227 D. Liebschner, P. V. Afonine, M. L. Baker, G. Bunkoczi, V. B. Chen, T. I. Croll, B. Hintze, L. W. Hung, S. Jain, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, B. K. Poon, M. G. Prisant, R. J. Read, J. S. Richardson, D. C. Richardson, M. D. Sammito, O. V. Sobolev, D. H. Stockwell, T. C. Terwilliger, A. G. Urzhumtsev, L. L. Videau, C. J. Williams and P. D. Adams, *Acta Crystallogr. Sect. D Struct. Biol.*, 2019, **75**, 861–877.

228 P. M. Reddy and T. C. Bruice, *J. Am. Chem. Soc.*, 2004, **126**, 3736–3747.

# 7 Appendix and supplementary information

## 7.1 CBM-LNA additional figures and information

*Table 7. 1 Oligonucleotide identities and sequences. Found mass results were obtained using UPLC-MS and compared to a calculated value. Modified oligonucleotides include, x = $CBM_1$, y = $CBM_2$, $^L$ = LNA. RNA sequences are denoted by a lowercase (r) at the start of the sequence.*

| ON Code | Sequence / 5'-3' | Calculated /Da | Found /Da |
|---|---|---|---|
| ON2.1 | GCTTGCTTCGTTCC | 4197 | 4196 |
| ON2.2 | GCTTGCTxTCGTTCC | 4160 | 4160 |
| ON2.3 | GCTTGCT$^L$xTCGTTCC | 4188 | 4187 |
| ON2.4 | GCTTGCTxT$^L$CGTTCC | 4188 | 4188 |
| ON2.5 | GCTTGCT$^L$xT$^L$CGTTCC | 4216 | 4215 |
| ON2.6 | GCTTGCTyTCGTTCC | 4160 | 4158 |
| ON2.7 | GCTTGCTyT$^L$CGTTCC | 4188 | 4188 |
| ON2.8 | GCTxTGCTxTCGTxTCC | 4086 | 4086 |
| ON2.9 | GCT$^L$xTGCT$^L$xTCGT$^L$xTCC | 4170 | 4167 |
| ON2.10 | GCTxT$^L$GCTxT$^L$CGTxT$^L$CC | 4170 | 4169 |
| ON2.11 | GCT$^L$xT$^L$GCT$^L$xT$^L$CGT$^L$xT$^L$CC | 4254 | 4255 |
| ON2.12 | GCTyTGCTyTCGTyTCC | 4086 | 4085 |
| ON2.13 | GCTyT$^L$GCTyT$^L$CGTyT$^L$CC | 4170 | 4170 |
| ON2.14 | GGAACGAAGCAAGC | 4331 | 4330 |
| ON2.15 | rGGAACGAAGCAAGC | 4555 | 4554 |
| ON2.16 | GCTT$^L$GCTT$^L$CGTT$^L$CC | 4281 | 4280 |
| ON2.17 | GCT$^L$T$^L$GCT$^L$T$^L$CGT$^L$T$^L$CC | 4365 | 4365 |

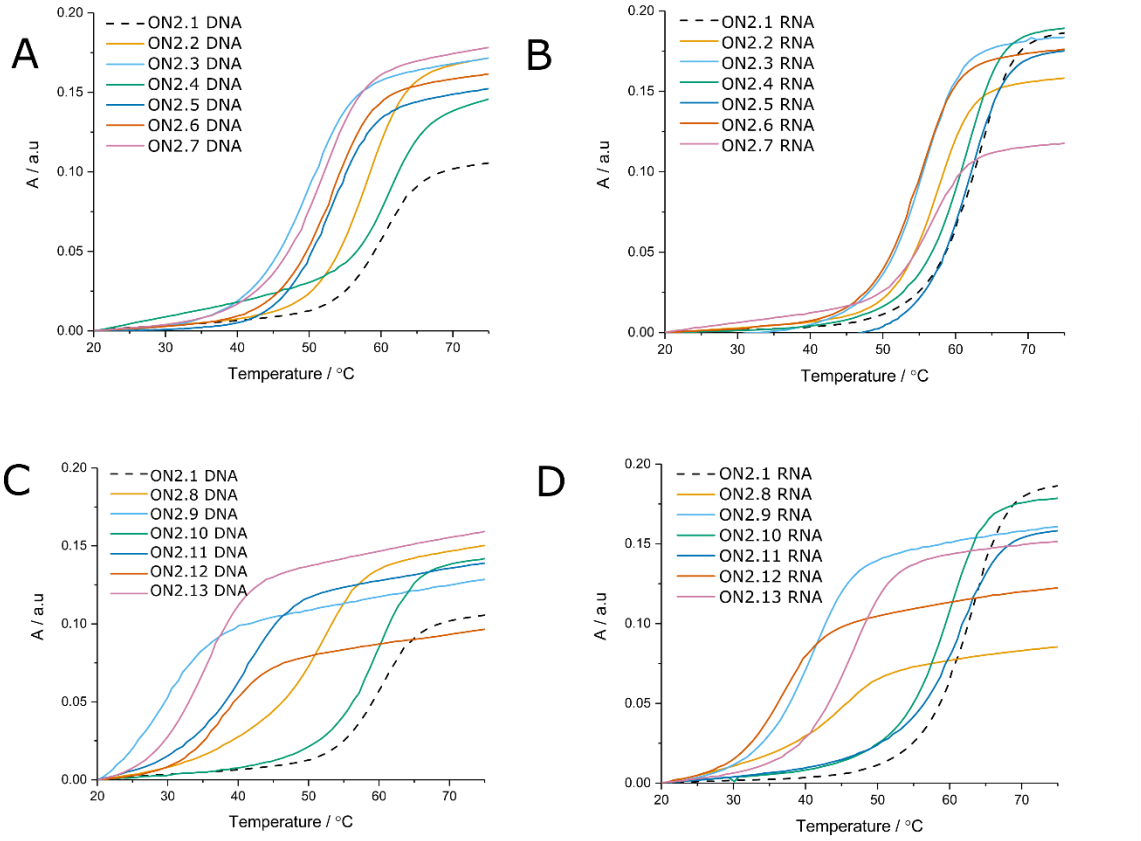



*Figure 7. 1 UV-melting curves of CBM-LNA modified oligonucleotides containing 3 µM oligonucleotide in 10 mM phosphate, 200 mM NaCl, pH 7.0 buffer. Samples were taken as an average of six ramps from 20 to 85 °C. (A) single incorporation of backbone modifications in DNA duplex, (B) single incorporation of backbone modifications in DNA:RNA hybrid, (C) triple incorporation of backbone modifications in DNA duplex, (D) triple incorporation of backbone modifications in DNA:RNA hybrid.*

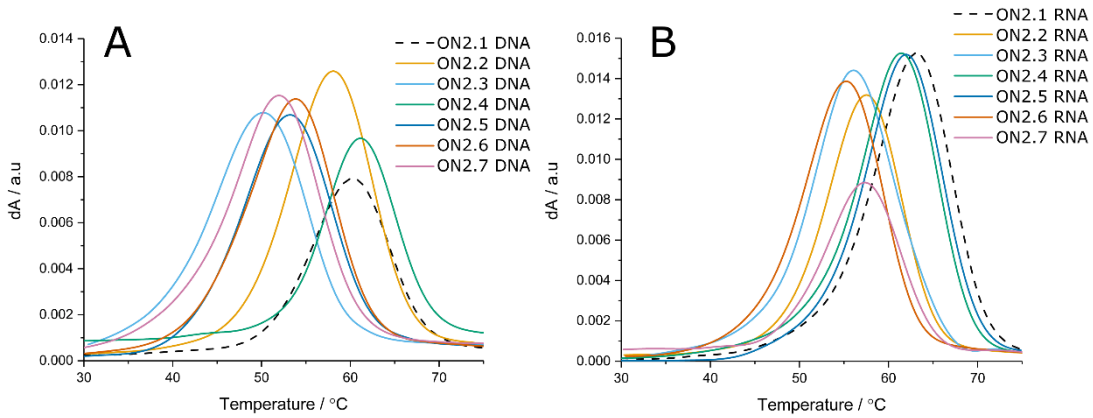



*Figure 7. 2 First derivatives of UV-melting curves containing 3 µM oligonucleotide with single modified carbamate backbones in 10 mM phosphate, 200 mM NaCl, pH 7.0 buffer. Samples were taken as an average of six ramps from 20 to 85 °C. (A) Modifications in DNA duplex single addition, (B) modifications in DNA:RNA hybrid single addition.*

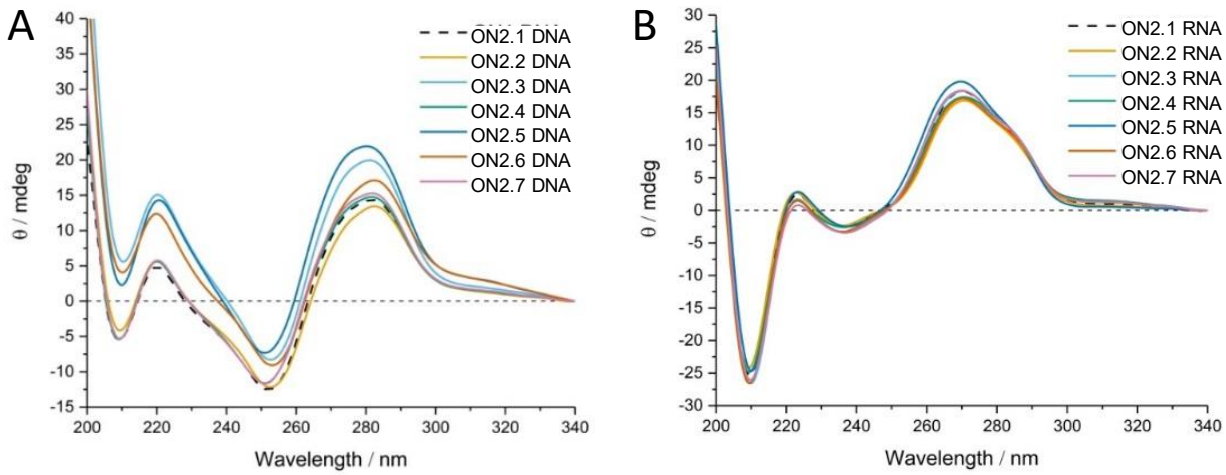



*Figure 7. 3 Circular dichroism of 3 µM sample oligonucleotides in 10 mM phosphate, 200 nM NaCl pH = 7.0 buffer. Spectra were averaged from four scans and smoothed to 20 points using a third order polynomial. (A) Single incorporation of CBM-LNA modifications against DNA target, (B) Single incorporation of CBM-LNA modifications against RNA target.*

## 7.2 LNA-amide crystallization conditions and additional information

*Table 7. 2 Table of conditions used to produce crystals shown in Figure 3. 9*

| Image | Oligonucleotide | Crystallisation Conditions |
|---|---|---|
| A | ON3.9 | 0.01 M $MgCl.6H_2O$, 0.02 M sodium cacodylate trihydrate pH 5.5, 20% v/v MPD, 0.001 M Heammine cobalt (III) chloride |
| B | ON3.9 | 0.04 M NaCl, 0.02 M sodium cacodylate trihydrate pH 5.5, 20% v/v MPD, 0.006 M spermine tetrahydrochloride, 0.01 M $MgCl_2.6H_2O$ |
| C | ON3.10 | 0.02 M $Mg(OAc)_2.4H_2O$, 0.025 M sodium cacodylate trihydrate pH 5.5, 15% v/v MPD, |
| D | ON3.12 | 0.025 M $MgSO_4$, 0.025 M Na HEPES pH 7.0, 0.8 M $Li_2SO_4.H_2O$ |
| E | ON3.4 | 0.04 M NaCl, 0.02 M sodium cacodylate trihydrate pH 7.0, 17.5% v/v MPD, 0.006 M spermine tetrahydrochloride |
| F | ON3.3 | 0.01 M $MgCl.6H_2O$, 0.025 M MOPS pH 7.0, 1.0 M Ammonium sulfate, 0.00025 M spermine |

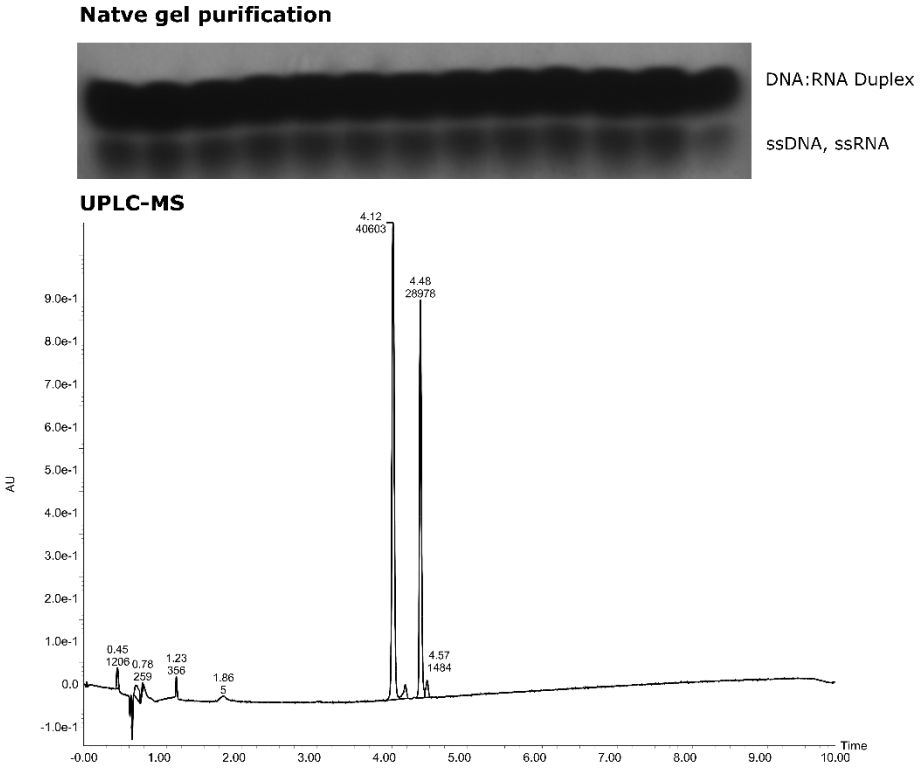



*Figure 7. 4 20% native gel purification of unmodified RNA:DNA (ON3.1:ON3.2) duplex to separate excess single stranded oligonucleotides (ssDNA and ssRNA). The duplex band was extracted from the gel and checked UPLC-MS to ensure duplex purify and confirm the presence of both strands.*

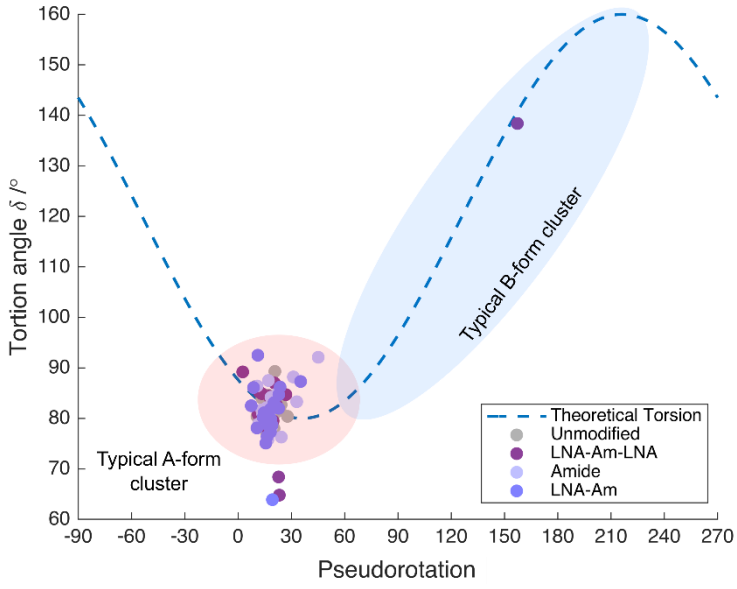



*Figure 7. 5 Pseudorotation vs Torsion angles of angle δ used to define sugar pucker conformations. Torsion data points were calculated using w3DNA 2.0 software and each point represents a single sugar conformation within a corresponding duplex. Each duplex has 20 data points and the clustering of these points can be interpreted to determine duplex form. The theoretical torsion angle is represented by equation ($\delta = 40\,cos(P + 144) + 120$). In general A-form duplexes have consistent pseudorotations 0-60 ° known as 3'-endo conformation. B-form duplexes have a larger distribution of pseudorotations 0-240 °*

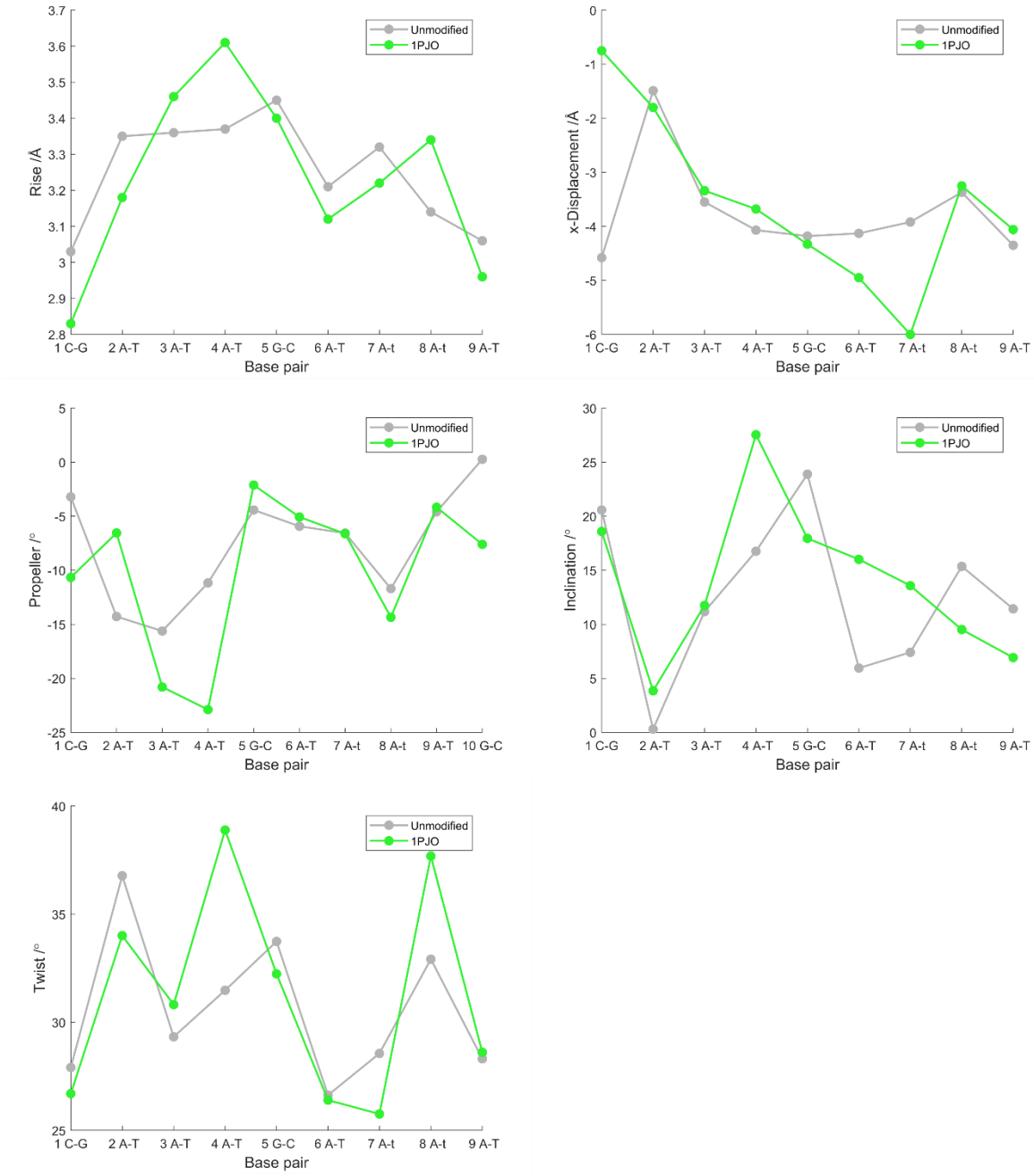



*Figure 7. 6 Comparison of base pair parameters from the published structure 1PJO (green) and our own unmodified duplex (grey). Parameters include, rise, x-displacement, propeller, inclination and twist. Base pair 1 is defined as the base pair at the 5'-RNA end. In general plots show good similarity between structures, however differences are observed due to crystal packing distortions at positions 6-10.*

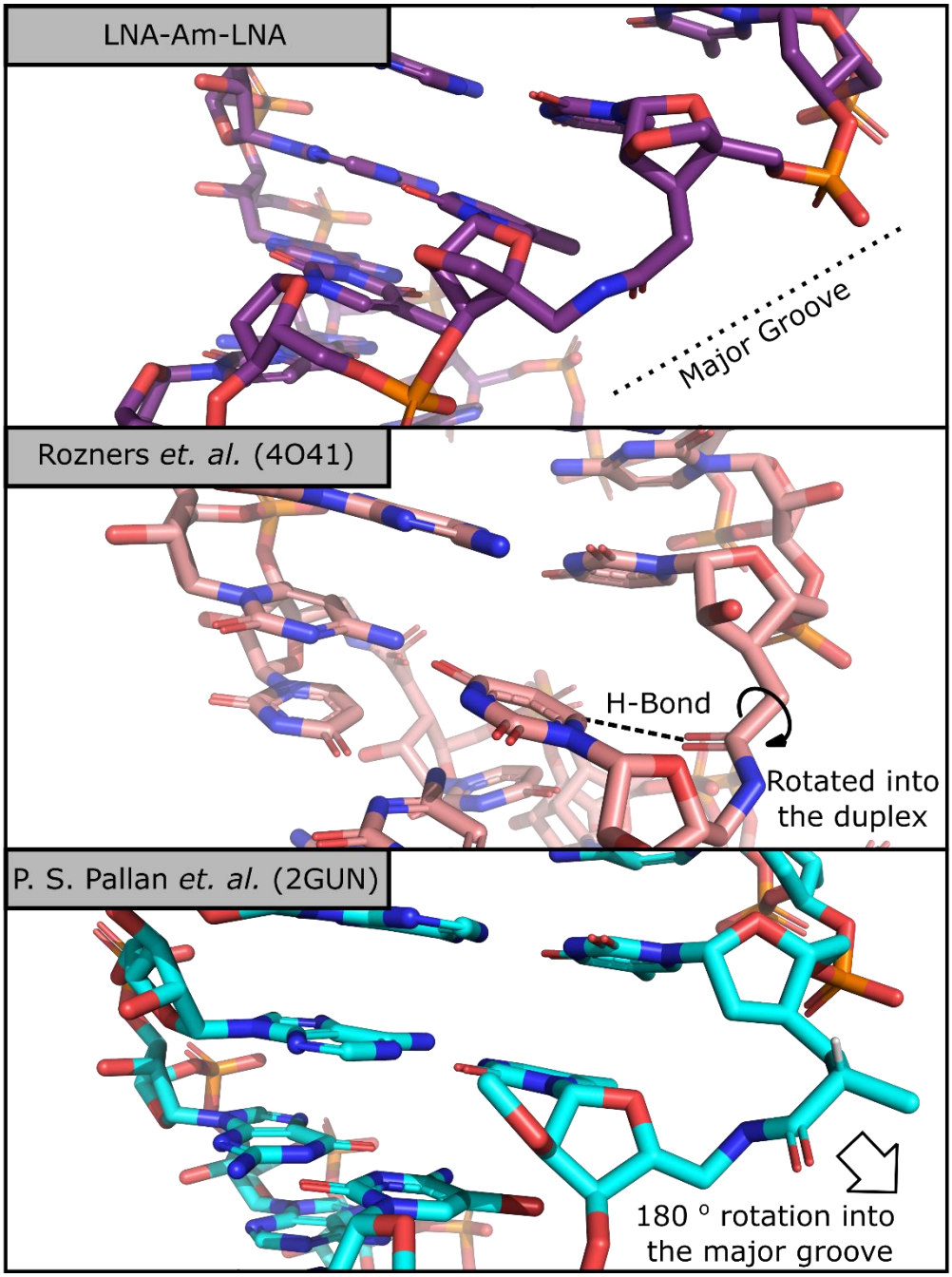



*Figure 7. 7 Structure comparison of different amide backbones within an oligonucleotide step. Shown is the LNA-Am-LNA modification which has a trans conformation and a carbonyl bond directed into the major groove. Also shown is two structures from the literature. Rozners et. al*[126] *(4O41) contains an amide backbone flanked by two RNA sugars. The carbonyl bond is rotated into the duplex forming a hydrogen bonding interaction to the H-6 atom of the 3'-nucleobase. P.S Pallan et. al.* [124] *(2GUN) structure contains a 4 atom branched amide in a trans conformation. The additional alkyl chain causes the amide backbone to be pushed outwards into the major groove.*

*Table 7. 3 Summary of torsion angles from w3DNA-2.0 analysis used to generate Figure 3. 24.*

| Torsion angle /° | Unmodified | Am | LNA-Am | LNA-Am-LNA |
|---|---|---|---|---|
| ε: C4'-C3'-O3'-P(n+1) | -151.0 | -178.6 | -179.0 | 173.7 |
| ζ: C3'-O3'-P(n+1)-O5'(n+1) | -73.7 | -119.2 | -122.6 | -113.0 |
| α: O3'(n-1)-P-O5'-C5' | -67.3 | 179.7 | 179.9 | 177.5 |
| β: P-O5'-C5'-C4' | 170.6 | -119.7 | -113.7 | -100.7 |
| γ: O5'-C5'-C4'-C3' | 52.8 | 163.5 | 171.0 | 161.9 |
| δ: C5'-C4'-C3'-O3' | 81.8 | 92.1 | 92.5 | 68.4 |

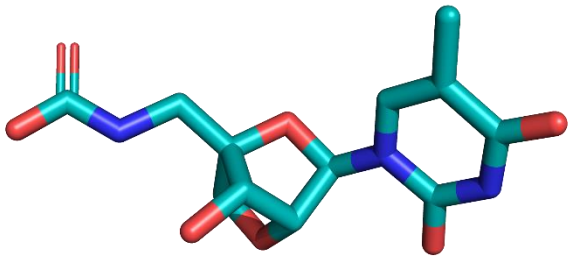


*Figure 7. 8 Visual representation of the ligand dictionary used to incorporate the CBM-LNA modification into the DNA:RNA duplex. Restrains were calculated using JLigand and imported using COOT molecular viewer. The monomer contains a 5'-COOH group that is used to form the CBM linkage during model refinement.*

*Table 7. 4 Summary of data processing for XRD structures of DNA:RNA hybrids containing CBM-LNA modification. Data was validated using pdb validation.*[151] *The dataset was collected from a single crystal, values shown in parenthesis are for the highest resolution shell.*

| Data collection | CBM-LNA |
|---|---|
| Space group | $P6_1\,2\,2$ |
| Cell dimensions | |
| *a, b, c* / Å | 41.08, 41.08, 109.33 |
| *α, β, γ* / ° | 90.00 90.00 120.00 |
| $R_{meas}$ | 0.107 (6.316) |
| $I/\sigma(I)$ | 1.40 (at 1.82 Å) |
| $CC_{1/2}$ | 1.000 (0.306) |
| Completeness / % | 99.8 (35.60-1.82) |
| Multiplicity | 35.2 (31.2) |
| Refinement | |
| Resolution / Å | 35.60 – 1.82 |
| No. reflections | 1309 |
| $R_{work}$ / $R_{free}$ | 0.318, 0.328 |
| No. atoms | 388 |
| DNA | 182 |
| RNA | 200 |
| $Mg^{2+}$ | 1 |
| Wilson *B* factor | 47.9 |
| r.m.s. deviations | |
| Bond lengths / Å | 0.0092 |
| Bond angles / ° | 2.16 |

## 7.3 Tz-circRNA supplementary information and sequences

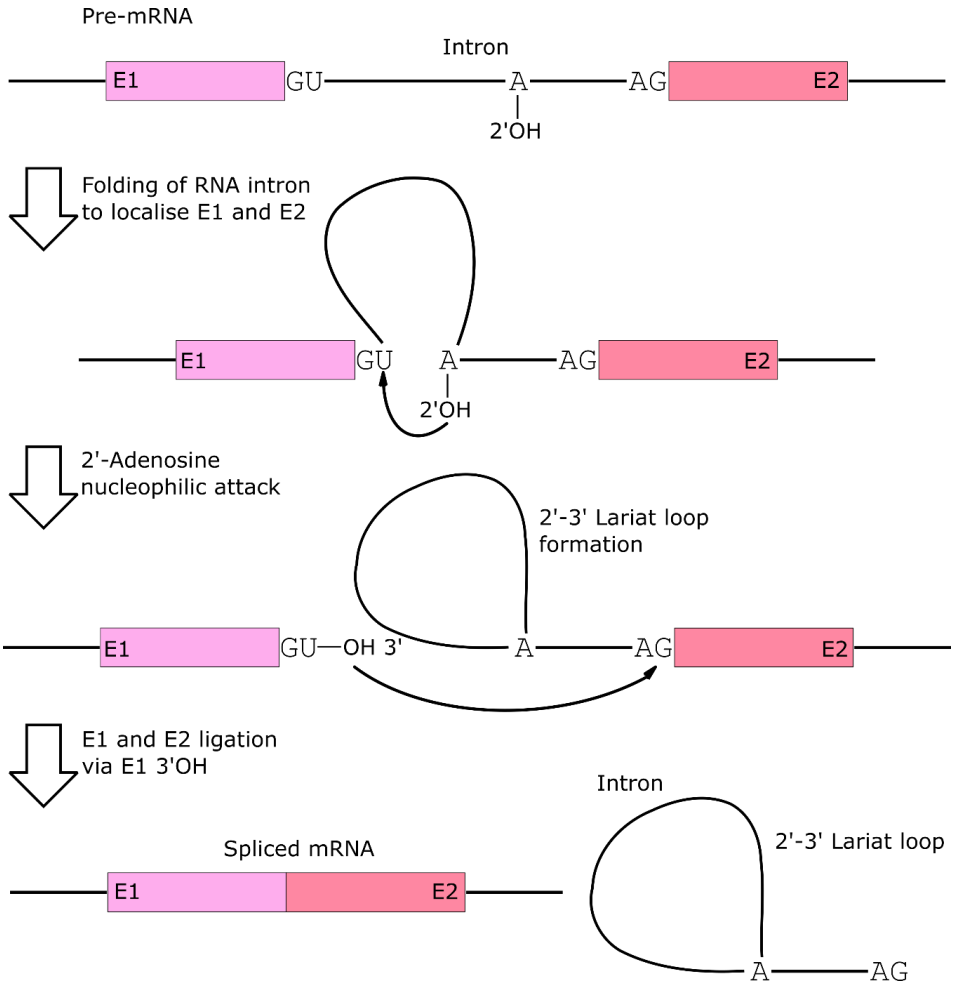



*Figure 7. 9 RNA self-splicing mechanism to remove introns from pre-mRNA. Secondary structures within the intron form a loop structure that localises the terminal regions of the exons (E1 and E2) together. Nucleophilic attack of an adenosine 2′-OH located within the intron forms a 2′-3′Lariat loop which is then spliced out of the RNA by joining the 3′-OH of the E1 to the 5′-end of the E2 exon.*

*Table 7. 5 Oligonucleotide sequences outlined in circRNA translation work. ON = oligonucleotide number, Modification name = RNA abbreviation name, RNA 2'= protection strategy during synthesis, length = number of bases given in nucleotides, 5' = functionality or modification at the 5' of the oligonucleotide, sequence = identity of the sequence, 3' = functionality or modification at the 3' of the oligonucleotide. Also included is calculated and found mass results from UPLC-MS after synthesis and purification. Multiple mass peaks represent ionisation of the oligonucleotide [M+Na] = +24 Da. P = Phosphate, OH = hydroxyl group, Alk = alkyne, $N_3$ = azide, $^L$ = LNA, (C) = nucleotide added via H-phosphonate.*

| ON | Modification Name | RNA 2' | Length /nt | 5' | Sequence | 3' | Calc Mass /Da | Found Mass /Da |
|---|---|---|---|---|---|---|---|---|
| ON4.1 | Unmod-1 | TBDMS or TC | 63 | P | AUUAUUAAGGAGAUAUAUCCGAUGAUUAUUGACUACAAGGACGAC GAUGACAAAAUUAUUGAC | OH | 20343 | 20342 |
| ON4.2 | Unmod-2 | TBDMS or TC | 63 | P | UACAAGGACGACGAUGACAAACUGCUGAUUAUUGACUACAAGGACG ACGAUGACAAAAUUAUU | OH | 20379 | 20379 |
| ON4.3 | Splint-12-short | - | 21 | OH | GTCCTTGTAGTCAATAATTTT | OH | 6401 | 6402 |
| ON4.4 | Splint-12-long | - | 40 | OH | TTGTCATCGTCGTCCTTGTAGTCAATAATTTTGTCATCGT | OH | 12224 | 12223 |
| ON4.5 | Splint-21-long | - | 40 | OH | GGATATATCTCCTTAATAATAATAATTTTGTCATCGTCGT | OH | 12249 | 12250 |
| ON4.6 | Tz-1, Tz-LNA-1 | TBDMS | 62 | P | AUUAUUAAGGAGAUAUAUCCGAUGAUUAUUGACUACAAGGACGAC GAUGACAAAAUUAUUGA | Alk | 20076 | 20077 |
| ON4.7 | Tz-2 | TBDMS | 64 | $N_3$ | CUACAAGGACGACGAUGACAAACUGCUGAUUAUUGACUACAAGGAC GACGAUGACAAAAUUAUU | OH | 20630 | 20629 20654 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ON4.8 | GAC23-LNA-2 | TBDMS | 64 | $N_3$-LNA | $C^{L}$UACAAGGACGACGAUGACAAACUGCUGAUUAUUGACUACAAGGACGACGAUGACAAAAUUAUU | OH | 20655 | 20655 |
| ON4.9 | DNA H-Phos test | - | 11 | OH | CGTCTACTGCT | OH | 3283 | 3284 |
| ON4.10 | GAC23-2Tz-1, RNA 1 | TBDMS | 72 | $N_3$ | (C)AAAAUUAUUAUUAUUAAGGAGAUAUAUCCGAUGAUUAUUGACUACAAGGACGACGAUGACAAAAUUAUUGA | Alk | 23194 | 23197 |
| ON4.11 | GAC23-2Tz-2, RNA 2 | TBDMS | 54 | $N_3$ | (C)UACAAGGACGACGAUGACAAACUGCUGAUUAUUGACUACAAGGACGACGAUGA | Alk | 17491 | 17493 |
| ON4.12 | Splint 1 L1 | - | 30 | OH | TCGTCGTCCTTGTAGTCAATAATTTTGTCA | OH | 9153 | 9153 |
| ON4.13 | Splint 2 L2 | - | 30 | OH | ATAATAATAATTTTGTCATCGTCGTCCTTG | OH | 9161 | 9161 |
| ON4.14 | Splint 3 | - | 57 | OH | TCAATAATCATCGGATATATCTCCTTAATAATAATAATTTTGACTGCATTCGATTCT | OH | 17423 | 17424 |
| ON4.15 | Splint 4 L1 | - | 40 | OH | TGTCATCGTCGTCCTTGTAGTCAATAATTTTGTCATCGTC | OH | 12209 | 12209 |
| ON4.16 | Splint 5 | - | 39 | OH | CGCTTAGAACGTCGCTCATCGTCGTCCTTGTAGTCAATA | OH | 11909 | 11909 |
| ON4.17 | HA-Unmod 1 | TBDMS | 68 | P | AUUAUUAUUAUUAAGGAGAUAUAUCCGAUGAUUAUUGACUACAAGGACGACGAUGACAAAAUUAUUGA | OH | 21921 | 21921 |
| ON4.18 | HA-Unmod 2 | TBDMS | 67 | P | CUACAAGGACGACGAUGACAAACUGCUGAUUAUUUACCCAUACGAUGUUCCAGAUUACGCUAUUAUU | OH | 21430 | 21431 |
| ON4.19 | Tz-HA 1 | TBDMS | 68 | P | AUUAUUAUUAUUAAGGAGAUAUAUCCGAUGAUUAUUGACUACAAGGACGACGAUGACAAAAUUAUUGA | Alk | 21959 | 21960 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| ON4.20 | Tz-HA 2 | TBDMS | 67 | $N_3$ | CUACAAGGACGACGAUGACAAACUGCUGAUUAUUUACCCAUACGAU GUUCCAGAUUACGCUAUUAUU | OH | 21376 | 21402 + Na |
| ON4.21 | HA-splint 1 | - | 35 | OH | CTTAATAATAATAATAATAATAGCGTAATCTGGAA | OH | 10794 | 10792 |
| ON4.22 | HA-splint 2 | - | 24 | OH | TCGTCCTTGTAGTCAATAATTTTG | OH | 7324 | 7324 |

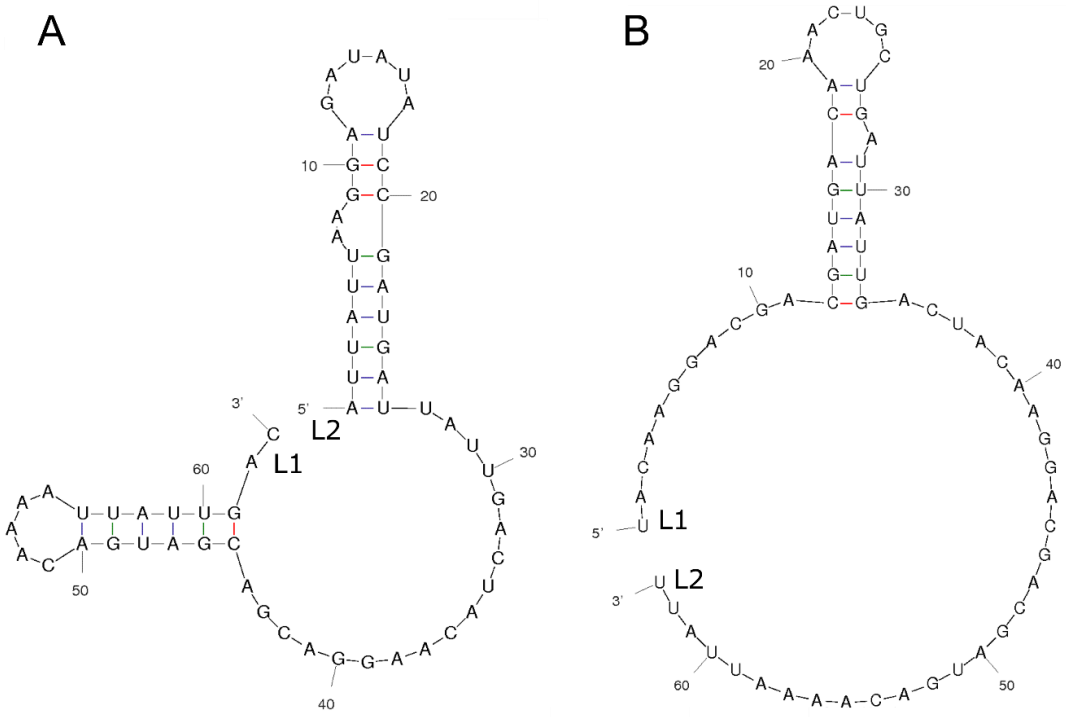



*Figure 7. 10 RNA folding and secondary structures calculated using Mfold web application[211]. A) ON4.1 showing duplex forming secondary structure around the 5'-L2 ligation end, ΔG = -5.7 kcal/mol, B) ON4.2 showing limited to no secondary structure around the ligation sites, ΔG = -3.3 kcal/mol.*

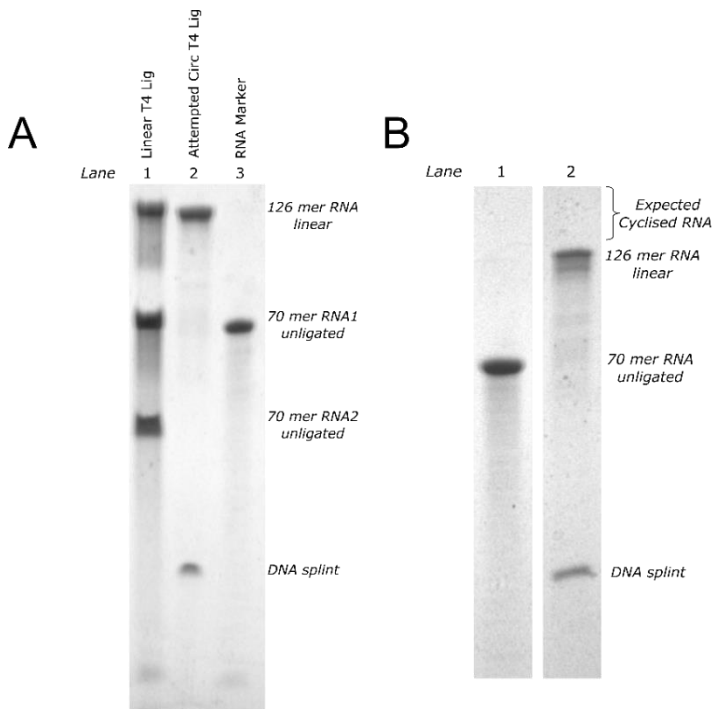



*Figure 7. 11 A) 10% page showing scale up of linear RNA precursor (lane 1) and attempted ligation of circRNA (lane 2). No ligation was seen. Possible reasons were high salt concentrations, impure RNA or denatured T4 ligase. B) 10% PAGE showing attempted cyclisation after desalting with two NAP gel filtrations and 10 kDa size exclusion washes. (lane 1) 70 mer RNA precursor marker, (lane 2) attempted cyclisation with no circRNA product observed.*

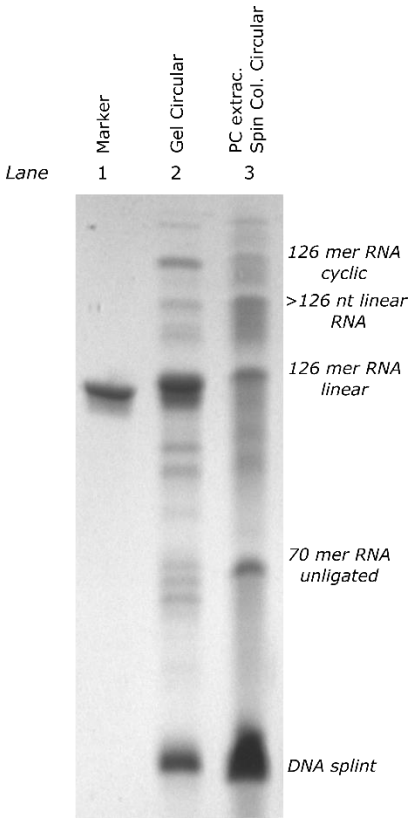



*Figure 7. 12 Gel comparing circRNA synthesise and purification methods: (lane 2) PAGE purification, (lane 3) phenol:chloroform extraction followed by desalting. PAGE purification must be used as phenol:chloroform extraction and desalting alone resulted in no significant cyclisation and the major products were linear RNAs. (lane 1) 126 nt linear RNA marker.*

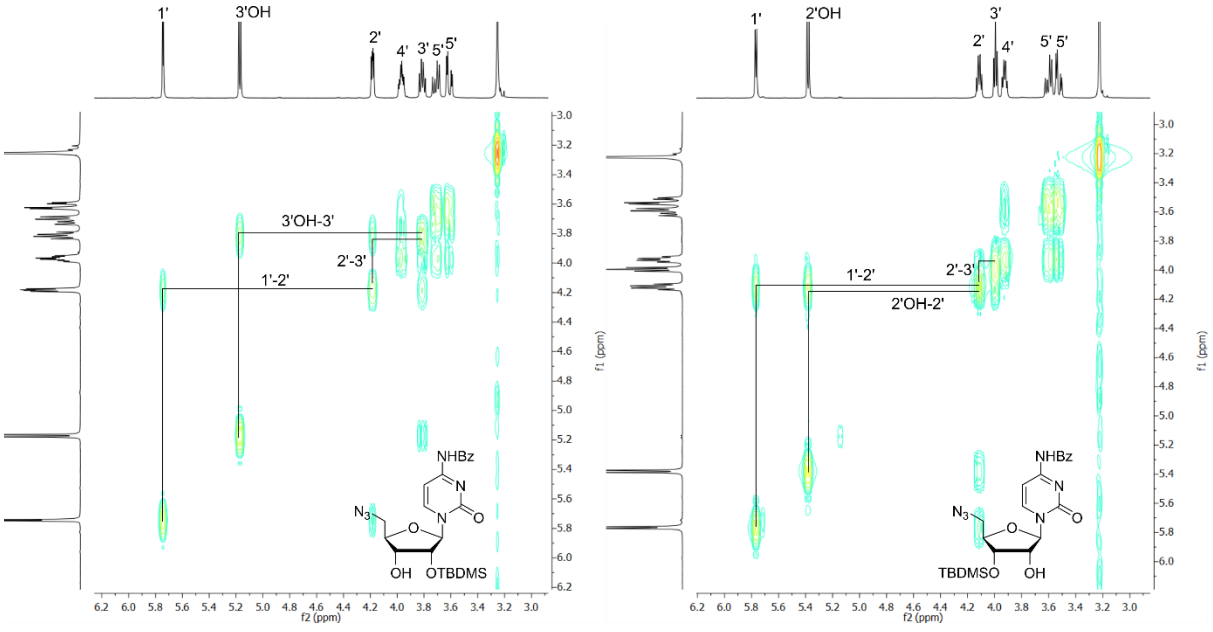



*Figure 7. 13 Proton COSY NMR spectra of TBDMS protected nucleotides to determine constitutional isomers. (Left) Desired 2'-TBDMS protected azide characterised by the 3'H-3'OH interaction (compound **44**), (Right) 3'-TBDMS protected by-product characterised by the 2'H-2'OH. Full characterisation of compound **44** can be found in the experimental.*

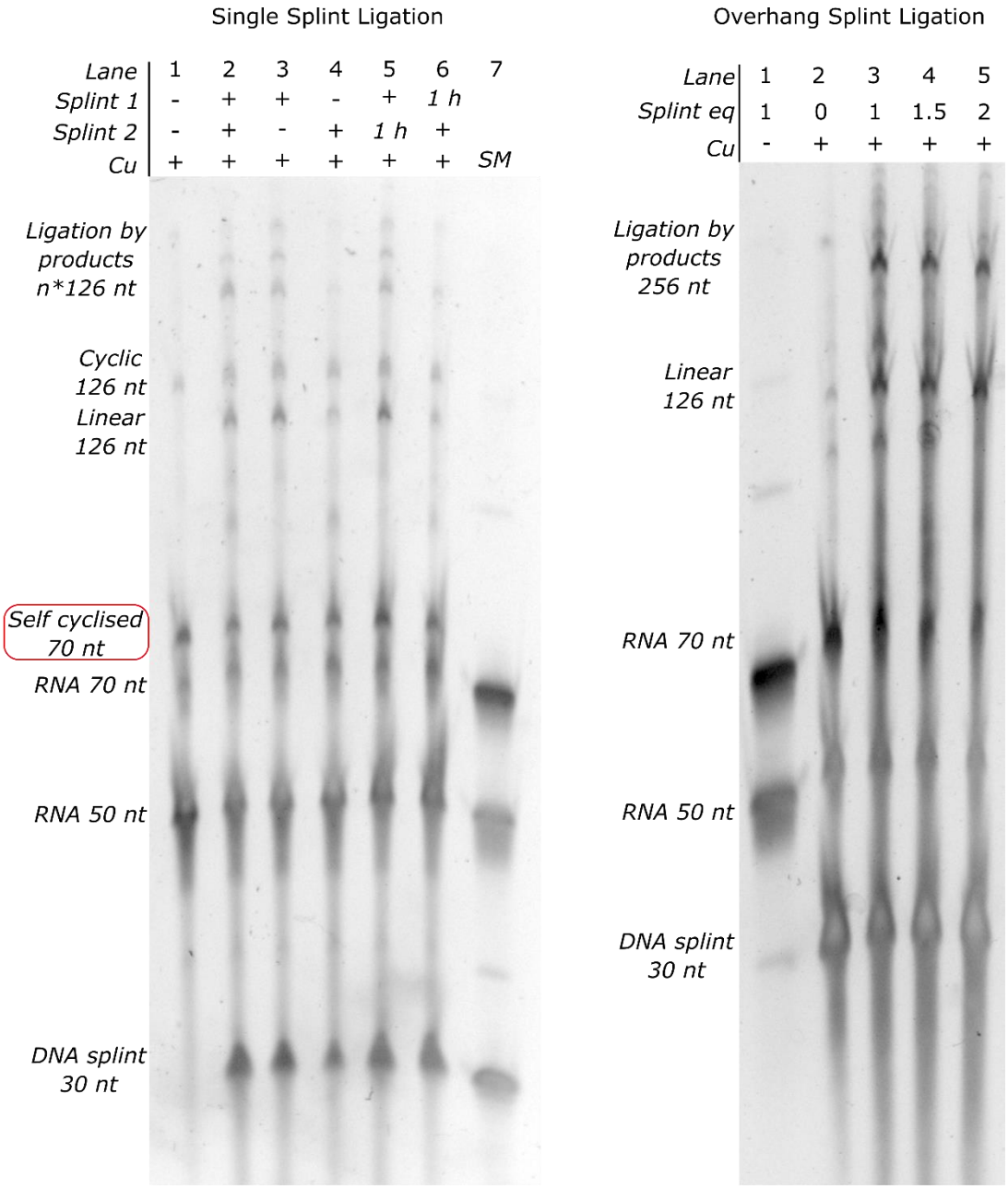



*Figure 7. 14 Optimisation of CuAAC linear ligations of RNAs ON4.10 and ON4.11 which both contain 3'-alkyne and 5'-azide reactivity. (Left) CuAAC ligation using single splints with no overhanging regions beyond the RNA, Lane 1 - no DNA splints, Lane 2 – splint 1 (ON4.12) and splint 2 (ON4.13) added at the same time to attempt both ligations simultaneously, Lane 3 – splint 1 only, ligation at the L1 position, Lane 4 – splint 2 only, ligation at the L2 position, Lane 5 – splint 1 added at t = 0 and splint 2 added after t = 1 h, Lane 6 – splint 2 added at t = 0 and splint 1 added at t = 1 h, Lane 7 – starting RNAs and splints, no copper. In all cases, regardless of splint addition order or times, 70 mer RNA (ON4.10) was observed to self-cyclise preventing linear ligation of the two RNA oligonucleotides. (Right) CuAAC ligation using three splints with 5' and 3' overhanging DNA sequences to prevent self-cyclisation, Lane 1 – RNAs (ON4.10 and ON4.11) and DNA splints 3, 4 and 5 (ON4.14, ON4.15 and ON4.16), Lane 2 – RNAs with copper and no splints, Lane 3 – RNAs with copper and 1.0 equivalent of DNA splints, Lane 4 – RNAs with copper and 1.5 equivalent of DNA splints, Lane 5 – RNAs with copper and 2.0 equivalent of DNA splints. Addition of splints overhanging the RNA prevented the self-cyclisation of the 70 mer RNA (ON4.10) allowing for ligation of the RNA oligonucleotides. Gels were visualised using SYBR gold intercalator, SM = starting materials.*

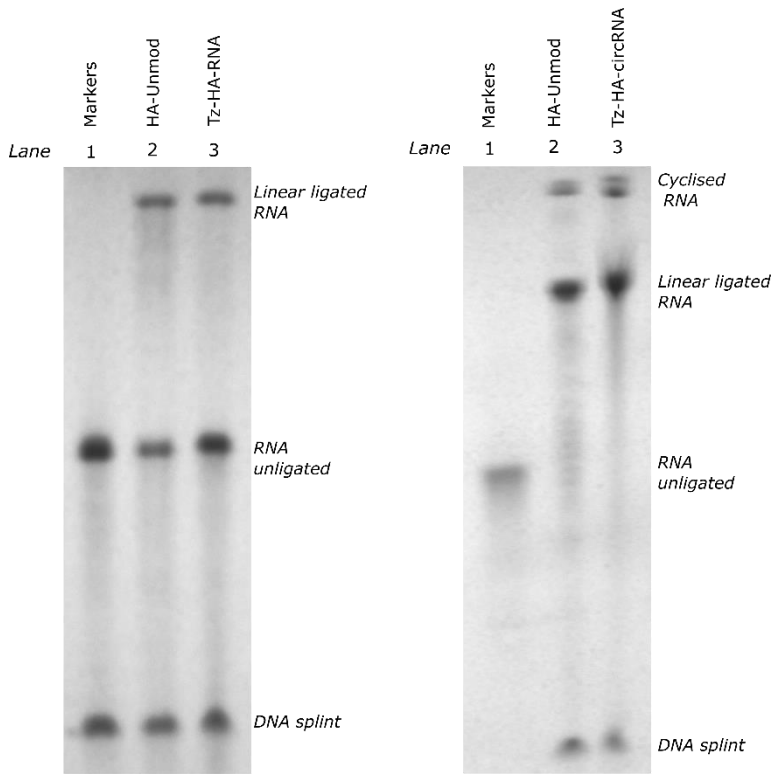



*Figure 7. 15 Synthesis of unmodified and triazole-modified circRNA containing HA-epitope tags for ELISA plate quantification. (Left) Ligation of RNAs to form linear precursors, (1) unligated RNA marker 70 mer, (2) ligation of ON4.17 and ON4.18 using T4 DNA ligase to form the unmodified linear RNA, (3) ligation of ON4.19 and ON4.20 using T4 DNA ligase to form linear Tz RNA. (Right) Cyclisation of HA-circRNAs, (1) unligated RNA marker 70 mer, (2) cyclisation of linear RNA extracted from lane 2 gel using T4 ligase to form unmodified HA-circRNA, (3) cyclisation of linear RNA using the CuAAC reaction to form the Tz-HA circular RNA.*

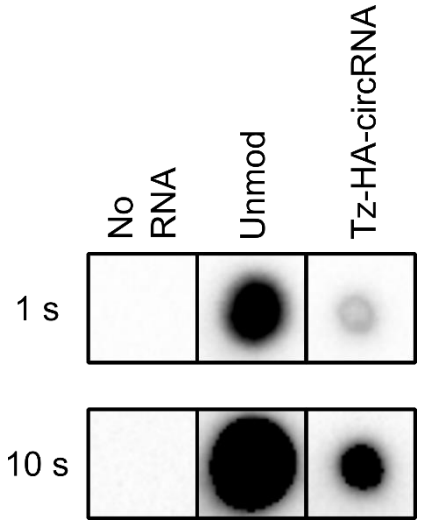



*Figure 7. 16 Dot blot assay of circRNAs containing HA-epitope tags. Two versions of the same image with exposure times of 1 and 10 seconds are shown. Exposure time of 1 s shows minimal amount of peptide expression from the Tz-HA-circRNA and oversaturation of the unmodified circRNA. Increased exposure time to 10 s causes oversaturation of both samples. This highlights the difficulties obtaining quantification from the dot blot assay. Despite this there should be some similarity between the dot blot and ELISA assay.*

# 8 Experimental

## 8.1 General procedures

All reagents were purchased from Sigma-Aldrich, Acros Organics, Fluka and Fisher Scientific, FluoroChem, Jena Bioscience or Alfa Aesar and used without further purification. Dry solvents (pyridine, $Et_3N$, $CH_2Cl_2$, THF, MeCN) were obtained using an MBraun SPS Bench Top solvent purification system (SPS). All air/moisture sensitive reactions were carried out under inert atmosphere (argon) in oven-dried glassware. Solvents used for phosphitylation reactions and purification were also degassed by saturating the solvents with argon for 20 min. Reactions were monitored by thin layer chromatography (TLC) using Merck Kieselgel 60 F24 silica gel plates (0.22 mm thickness, aluminium backed). The compounds were visualized by UV irradiation at 254/265 nm and by staining in *p*-anisaldehyde or ninhydrin solution followed by gentle heating. Silica column chromatography was performed using Merck Geduran® 60 Å (40–62 µm) or using a Biotage Isolera™ Prime auto column (10-100 g SNAP ultra-cartridges). $^1$H NMR (400 MHz), $^{13}$C NMR (101 MHz) and $^{31}$P NMR (162 MHz) were recorded using a Bruker DPX 400 (AVIIIHD 400) or Bruker AVII 500 spectrometer ($^1$H-500 MHz, $^{13}$C-126 MHz) and referenced to an appropriate deuterated solvent. Assignment was aided by DEPT-135, COSY, HSQC-DEPT and HMBC spectra. Data was then processed using MestreNova software. Low resolution mass spectrometry (LRMS) using electrospray ionisation positive (ESI+) and negative (ESI-) modes was recorded on a Waters ZMD quadrupole mass spectrometer in HPLC grade MeOH/MeCN. High resolution mass spectrometry (HRMS) using electrospray ionisation was recorded on a Bruker APEX III FT-ICR mass spectrometer. Oligonucleotide mass spectrometry was recorded on a UPLC-MS Waters XEVO G2-QTOF (ESI-) using an ACQUITY UPLC oligonucleotide BEH C18 column, 130 Å (1.7 µm, 2.1 mm × 50 mm). Data was then deconvoluted using MassLynx v4.1. A gradient of MeOH in $Et_3N$ and hexafluoroisopropanol (HFIP) was used (buffer A, 8.6 mM $Et_3N$, 200 mM HFIP in 5% MeOH/$H_2O$ (v/v); buffer B, 20% buffer A in MeOH). Buffer B was increased from 0–70% over 8 min, at a flow rate of 0.2 mL min$^{-1}$.

## 8.2 Oligonucleotide synthesis

### *8.2.1 Synthesis of DNA oligonucleotides*

Standard phosphoramidites, solid supports and reagents were purchased from Link Technologies and Applied Biosystems. LNA phosphoramidites were obtained from Exiqon. Automated solid phase synthesis of oligonucleotides (trityl off) was performed on an Applied Biosystems 394 synthesiser. Synthesis was performed on 1.0 µmole scale involving cycles of acid catalysed detritylation, coupling, capping, and iodine oxidation. Standard DNA phosphoramidites were coupled for 60 s while extended coupling time of 10 min was used for LNA phosphoramidites. Coupling efficiencies and overall synthesis yields were determined by the inbuilt automated trityl cation conductivity monitoring facility and were ≥ 98.0% in all cases. The oligonucleotides were then cleaved from the solid support and protecting groups from the nucleobase and backbone were removed by exposure to conc. aq ammonium hydroxide for 60 min at room temperature, followed by heating in a sealed tube for 5 h at 55 °C.

### *8.2.2 Synthesis of RNA oligonucleotides*

2'-*O*-TBDMS or 2'-*O*-TC protected RNA phosphoramidite monomers with *t*-butylphenoxyacetyl protection of the A, G and C nucleobases were used to assemble RNA oligonucleotides. Benzylthiotetrazole (BTT) was used as the coupling agent, *t*-butylphenoxyacetic anhydride as the capping agent and 0.1 M iodine as the oxidizing agent (Sigma-Aldrich). A coupling time of 10 min was used and coupling efficiencies of > 97% were obtained. Cleavage of oligonucleotides from the solid support and protecting groups from the nucleobase and backbone were removed by exposure to concentrated aqueous ammonia/ethanol (3/1 v/v) for 2 h at room temperature followed by heating in a sealed tube for 2 h at 55°C. For the synthesis of 3'-Alkyne modified RNAs, resin bound 3'-Alkyne monomer (ChemGenes: N-9882-10) was used followed by the standard RNA cycle. 5'-Phosphates were synthesised using the standard RNA cycle, in the final round of synthesis, chemical phosphorylation reagent (Link, 2101-F100) was used with a coupling time of 10 min. The RNA was then deprotected using standard conditions.

### *8.2.3 Removal of 2'-O-TBDMS protecting groups on RNA oligonucleotides*

Oligonucleotides were concentrated under reduced pressure until formation of a white turbidity and freeze dried overnight. The residue was re-dissolved in DMSO (300 µL) and triethylamine trihydrofluoride (300 µL) was added followed by incubation at 65 °C for 2.5 h. Sodium acetate (3 M, 50 µL) and butanol (3 mL) were added and the samples were stored at –80 °C for a minimum of 1 h. After thawing, the suspension was pelleted under centrifuge (13,000 rpm, 4 °C, 10 min) and the supernatant decanted. The precipitate was then washed with ethanol (2 × 750 µL) then dried under vacuum.

### *8.2.4 Removal of 2'-O-TC protecting groups on RNA oligonucleotides*

600 µL of a 1:1 dry toluene:dry ethylenediamine (v/v) solution was slowly added to the RNA synthesis column with no agitation and incubated for 6 hr at rt. The column was then washed twice with 1 mL dry toluene and dried with argon. RNA was eluted from the resin in the column by slowly pushing 1 mL of water through the column and repeating with another 1 mL of water.

### *8.2.5 Oligonucleotide 5'-Azide conversion*

After solid-phase synthesis, the protected resin-bound 5'-OH RNA was treated with methyltriphenoxyphosphonium iodide (0.5 M) in anhydrous DMF (1 mL) for 1 h at rt. The solid support was washed with dry DMF (3 × 1 mL) and dried with argon. A saturated solution of sodium azide was then prepared by heating sodium azide (100 mg) in dry DMF (2 mL) for 10 min at 70 °C. After cooling to rt, the resin was treated with the azide solution for 5 h at 55 °C. The resin was then washed with DMF (3 × 1 mL), acetonitrile (3 × 1 mL) and dried with argon. The 5'-azido-RNA was then cleaved from solid support and deprotected using standard conditions.

### *8.2.6 Coupling of H-Phosphonates to resin bound oligonucleotides*

Resin bound oligonucleotide was washed with $CH_2Cl_2$ (3 × 150 µL) followed by pyridine:MeCN (3 × 150 µL, 1:3). To a separate vial was added 1-adamantane chloride (9.04 mg, 45.5 µmol) which was dissolved in pyridine:MeCN (150 µL, 1:3). To this solution was added H-phosphonate **45** (5.00 mg, 9.10 µmol) which was mixed for 15 seconds. The activated solution was then added to the resin bound oligonucleotide and rotated at rt for 1 h. The resin was washed with pyridine:MeCN (3 × 150 µL, 1:3) before being incubated in 2% iodine (6.00 mg, 300 µL, pyridine:$H_2O$, 49:1) for 30 mins at rt. The resin was washed with pyridine:MeCN (4 ×

150 µL, 1:3), $CH_2Cl_2$ (3 × 150 µL) and dried under vacuum. The oligonucleotide was then cleaved using the standard DNA or RNA deprotection procedure (8.2.3 or 8.2.4).

### *8.2.7 T4 DNA ligase templated ligation of RNA using a DNA splint*

RNA oligonucleotides (500 pmol) were combined with the DNA splint (500 pmol) in nuclease free water (20 µL). To this was added 10x T4 DNA ligase buffer (5µL) and T4 DNA ligase (5 µL, 5 units/µL, thermofisher). The final volume was made up to 50 µL with nuclease free water and the reaction was incubated at 37 °C for 16 h. For linear ligations, RNA concentration was 10 pmol/µL and for intramolecular cyclisation the concentration was 1.0 pmol/µL. The reaction was scaled according to the number of RNA moles.

### *8.2.8 CuAAC templated ligation of RNA using a DNA splint*

25 mg of $CuSO_4.5H_2O$ and 100 mg of sodium ascorbate were each dissolved separately in 1 mL of 200 mM NaCl. To 3 mg of THPTA was added 10 µL of $CuSO_4.5H_2O$ solution and 20 µL sodium ascorbate solution. The reagent solution was then degassed under argon for 5 mins.

RNA alkyne (1 nmol) and RNA azide (1 nmol) were combined with the DNA splint (1 nmol) and 2 M NaCl to make a total volume of 94 µL and 200 mM NaCl. The oligonucleotides were annealed at 95 °C for 5 mins and cooled to rt (1 °C/min). The oligonucleotide solution was degassed under argon for 5 mins before addition of 6 µL of the prepared reagent solution (200 eq $Cu^{I}$, $CuSO_4.5H_2O$: sodium ascorbate: THPTA, 1:10:7). The reaction was mixed, degassed under argon for a further 5 mins and incubated at rt for 2 h. For linear ligations, RNA concentration was 10 pmol/µL and for intramolecular cyclisation the concentration was 1.0 pmol/µL. The reaction was scaled according to the number of RNA moles.

## 8.3 Purification and analysis of oligonucleotides

### *8.3.1 HPLC Purification of oligonucleotides*

Oligonucleotides were purified using high performance liquid chromatography (HPLC) performed on a Gilson 118 UV detector, with 306 pumps, 805 manometric module, and 811c dynamic mixer, equipped with a Phenomenex Luna 10 µm C8 column (100 Å pore; 10 × 250 mm). A gradient of acetonitrile in 0.1 M TEAB buffer (from 10:90 to 40:60 acetonitrile:TEAB) was used to elute the ONs at rt over 20 mins at a flow rate of 4 mL/min. Elution was monitored by UV absorbance at 285–300 nm. Collected fractions were freeze-dried overnight and re-suspended in 1 mL deionised water.

### *8.3.2 Denaturing poly acrylamide gel electrophoresis (PAGE)*

To the urea containing denaturing gel composition buffer containing the desired acrylamide percentage (see table of buffers) was added 56 µL TEMED and 560 µL of 10% ammonium persulfate. This was then poured into the gel electrophoresis unit (Hoefer, SE660) and allowed to polymerise for 30 mins. The gel was then pre-run for 30 mins (20 W) in 1x TBE buffer, followed by washing of the loading wells with fresh buffer. Samples were mixed with 50% formamide and applied to the electrophoresis wells. The gel was then run between 1.5 and 5 h at 20 W. For 20% gels the load was reduced to 400 V to improve DNA band migration. Oligonucleotides were visualised with UV shadowing (260 nm) or SYBR gold intercalator stain (Ex= 495 nm, Em= 537 nm).

### *8.3.3 Native poly acrylamide gel electrophoresis*

To the native gel composition buffer containing the desired acrylamide percentage (see table of buffers) was added 56 µL TEMED and 560 µL of 10% ammonium persulfate. This was then poured into the gel electrophoresis unit (Hoefer, SE660) and allowed to polymerise for 30 mins. The gel was then pre-run for 30 mins in 1x TBE buffer, followed by washing of the loading wells with fresh buffer. Samples were mixed with 50% glycerol and applied to the electrophoresis wells. The gel was then run for up to 20 h at 120 V at 4 °C. Oligonucleotides were visualised with UV shadowing (260 nm) or SYBR gold intercalator stain (Ex= 495 nm, Em= 537 nm).

### *8.3.4 Gel extraction and NAP gel filtration of oligonucleotides*

Desired DNA bands were cut from the PAGE gel under UV light and ground up using a glass rod. The gel containing DNA was then added to 1 – 10 mL of $H_2O$ and incubated for 16 h at 37 °C, 600 rpm. The gel was filtered, and the supernatant was desalted with sequential NAP-10g and NAP-25g gel filtration columns. For small volumes, the NAP-25g was exchanged for three washes in a size exclusion spin column. RNA extraction was performed using the same method as described, however, the incubation was run in RNA gel extraction buffer instead of $H_2O$ (see table of buffers).

### *8.3.5 Size exclusion spin column*

The oligonucleotide sample (500 µL) was added to a 10 kDa size exclusion column (Amicon, Sigma Aldrich) and centrifuged (14000 rpm, 10 mins). The sample was then washed with $H_2O$

(450 µL) and centrifuge for a further 10 mins. Two subsequent washes were performed followed by elution of the oligonucleotide into a clean vial (1000 rpm, 2 mins).

### *8.3.6 Phenol: chloroform extraction*

5 µL of reaction mixture was added to an equal volume of phenol: chloroform: isoamyl alcohol, (25:24:1) and vortexed for 1 min. The sample was centrifuged (16000 rpm, 5mins) before removal of the top layer. The removed aqueous layer was then diluted in $H_2O$ (300 µL) and the buffer exchanged into water (3x 500 µL) using a size exclusion spin filtration column.

## 8.4 Biophysical studies and assays

### *8.4.1 Ultraviolet melting studies (duplex hybridisation - $T_m$)*

UV oligonucleotide melting curves were recorded in a Cary 4000 Scan UV-Visible Spectrophotometer using 3 µM of each oligonucleotide in a 10 mM phosphate buffer containing 200 mM NaCl at pH 7.0. Samples were annealed by heating to 85 °C (10 °C/min) and then slowly cooling to 20 °C (1 °C/min). As these six successive cycles (heating and cooling) were performed at a gradient of 1 °C/min, the change in UV absorbance at 260 nm was recorded. The melting temperature was calculated from the first derivative of the melting curve using in built software.

### *8.4.2 Circular dichroism studies*

A solution of 3 µM sample oligonucleotide and complementary sequence was added to 10 mM phosphate buffer with 200 mM NaCl (pH 7.0). Samples were incubated for 2 mins at 85 °C and cooled to rt (1 °C/min). Circular dichroism was performed on a Chirscan Plus spectrometer using a quartz cuvette (L = 10 mm). Scans were taken at 20 °C from 200–340 nm with 0.5 a step and 1.0 s time point intervals. An average of four scans was taken and smoothed to 20 points using a third order polynomial (OriginPro 2017). The spectra were then baseline corrected to the θ-value at 340 nm.

### *8.4.3 Phosphodiesterase 1 assay (Snake venom)*

5 nmol of oligonucleotide was dissolved in 100 µL of buffer (50 mM Tris-HCl, 10 mM $MgCl_2$, pH 9.0). 20 µL of sample was removed (t = 0 control) and 2 µL of snake venom phosphodiesterase 1 from the Eastern diamondback rattlesnake *C. adamanteus* (Sigma-Aldrich, cat P3243, 0.45 units, dissolved in 7 mL $H_2O$) was added to the remaining volume. The reaction was incubated at 37 °C and 20 µL aliquots were removed at set times, these were

then mixed with formamide (v/v) and stored at -80 °C. Samples were analysed by 20% denatured polyacrylamide gel electrophoresis (400 V) and viewed under short wave UV-light.

### *8.4.4 Foetal bovine serum (FBS) assay*

5 nmol of oligonucleotide was dissolved in 50 µL of Dulbecco's PBS and 50 µL of FBS (Gibco, standard sterile filtered). The reaction was vortexed and 20 µL was removed (t = 0 control). The control was mixed with formamide (v/v) and frozen at –80 °C. The remaining solution was incubated at 37 °C and 20 µL aliquots were removed at set times, these were then mixed with formamide (v/v) and stored at –80 °C. Samples were analysed by 20% denatured polyacrylamide gel electrophoresis (400 V) and viewed under short wave UV-light.

### *8.4.5 PURExpress cell free prokaryotic RNA translation kit (NEB PURExpress E6800)*

CircRNA template (100 ng) was freeze dried, re-dissolved in nuclease free water (1.5 µL) and combined with solution A (2 µL). To this was added solution B (1.5 µL) and the reaction was mixed by pipette before incubating at 25 °C for 20 h. To analyse protein expression the reaction (2 µL) was subjected to tris-glycine SDS-PAGE (8.4.6) and immunoblotting. The reaction was scaled accordingly, depending on the amount of protein required (https://bit.ly/3gStZXc).

### *8.4.6 Tris-glycine SDS-PAGE*

Novex™ 10-20% tris-glycine mini gels (Thermo Fisher Scientific) and mini gel tank (Thermo Fisher Scientific) were used as electrophoresis equipment. Protein samples were prepared by mixing with 2x SDS-PAGE loading buffer (v/v) and boiling at 100 °C for 10 min. Colour Prestained Protein Standard, Broad Range (New England Biolabs) was used as a marker and loaded directly onto the gel. The gel was submerged in 1x tris-glycine gel running buffer and run at 80 V until the sample reached the resolving gel, after which the voltage was increased to 180 V.

### *8.4.7 Dot blot*

PVDF membranes (AmershamTM Hybond®, 0.2 µm pore size) were prepared by soaking in MeOH (15 seconds), $H_2O$ (2 mins) and PBS-T (10 mins). While still damp, samples (1.5-2 µL) were pipetted onto the membrane and allowed to soak in. Once the membrane had dried it was washed with PBS (2 x 10 mL) and the membrane was blocked by incubation in 5% skimmed milk powder in PBS-T (w/v, 30 mins, rt, 300 rpm). Primary antibody (mouse anti-

FLAG, dilution 1:2000, Sigma-F3165) was added to the blocking solution and the membrane was incubated overnight at 4 °C and 300 rpm. The membrane was rinsed with PBS (2x 10 mL) and washed with PBS-T (3x 10 mL, 3x 5 mins). In a solution of 5% skimmed milk powder in PBS-T (w/v) was added secondary antibody (rabbit anti-mouse IgG with peroxidase HRP conjugate, dilution 1:250000, Sigma) and incubated for 1 h at rt, 300 rpm. The wash steps were repeated followed by two further washes with PBS (10 mL). To visualise the blot, SuperSignal™ West Femto Maximum Sensitivity Substrate (Thermo Fisher Scientific) was added to the membrane and imaged with a Gel Doc imager (1 → 30 s exposure accumulation, chemi channel, Bio-rad).

### *8.4.8 Western blot*

Before performing the western blot, samples were separated using tris-glycine SDS-PAGE (8.4.6). PVDF membranes (AmershamTM Hybond®, 0.2 µm pore size) were prepared by soaking in MeOH (15 seconds), $H_2O$ (2 mins) and 1x transfer buffer (10 mins). The transfer sandwich was set up as follows: anode support grid (red), sponge, filter paper, PVDF membrane, SDS-gel, filter paper, sponge, cathode support grid (black). The sandwich was then placed into the omniPAGE Mini tank with 1x western blot transfer buffer and run at 400 mA for 1.5 hr. Ice was placed around the tank to cool the temperature. After rinsing the membrane with PBS, the blocking, primary antibody, secondary antibody, and visualisation steps were the same as described above for dot blots (8.4.7).

### *8.4.9 ELISA assay*

To each well was added coating antibody (concentration 5 µg/mL, rabbit anti-HA, Thermo Fisher Scientific-71-5500) in 1x coating buffer (150 µL) and incubated overnight at 4 °C and 300 rpm. The wells were washed with PBS-T (3x 150 µL, 3 x 5 mins) and blocking was performed using 5% skimmed milk powder in PBS-T (w/v, 150 µL) for 1 h at rt and 300 rpm. The wells were then washed with PBS-T (3x 150 µL, 3x 5 mins). Protein samples (3 µL) were diluted in PBS-T (97 µL), mixed and then incubated in the wells for 1.5 h at rt and 300 rpm. The wells were then washed with PBS-T (3x 150 µL, 3x 5 mins). Primary antibody (concentration 1 µg/mL, mouse anti-FLAG, Sigma-F3165) was added in 5% skimmed milk powder in PBS-T (w/v, 100 µL) and incubated for 1 h at rt and 300 rpm. The wells were then washed with PBS-T (3x 150 µL, 3x 5 mins). Secondary antibody (concentration 0.1 µg/mL, rabbit anti-mouse IgG with peroxidase HRP conjugate, Sigma) was added in 5% skimmed milk

powder in PBS-T (w/v, 100 µL) and incubated for 1 h at rt and 300 rpm. The wells were then rinsed with PBS-T (4x 150 µL) followed by three washes (3x 150 µL, 3x 5 mins). To visualise the ELISA, TMB core+ (100 µL, BioRad) was added to each well and allowed to develop in the dark (10-30 mins). Data points were recorded as a kinetic assay at 655 nm. Alternately $H_2SO_4$ (0.2 M, 100 µL) was added and an endpoint assay was measured at 450 nm.

## 8.5 X-Ray diffraction procedures and model refinement

### *8.5.1 Crystallisation high throughput screens*

Crystals were obtained by sitting drop vapour diffusion. DNA and RNA oligonucleotides were purified by HPLC, desalted by gel filtration (NAP-10) and then freeze dried. 2 mM stock solutions of each oligonucleotide were made in KCl (10 mM). A sample of 1 mM duplex (50 µL) was made from equal moles of complementary DNA and RNA. High throughput conditions were tested using a Natrix HT screen (Hampton Research, HR2-131). A CrystalMation Intelli-Plate 96-3 low-profile plate (Hampton Research, HR3-119) was placed into the loading slot of an Art Robbins Phoenix automatic liquid handler. 80 µL of solution from each well in the Natrix HT screen was then dispensed into the plate reservoirs. To set up the crystallization drop, 0.1 µL of Natrix HT solutions were dispensed into the drop well, followed by 0.1 µL of oligonucleotide duplex (1 mM), run in triplicate. The plate was removed from the dispenser and sealed (final drop vol. 0.2 µL, duplex concentration 0.5 mM). The plate was then incubated at rt, crystals usually formed within one week (range 2-90 days, crystal size < 0.01-0.2 mm).

### *8.5.2 Crystallisation manual optimisation screens*

Crystals were obtained by sitting drop vapour diffusion. DNA and RNA oligonucleotides were purified by HPLC, desalted by gel filtration (NAP-10) and then freeze dried. 2 mM stock solutions of each oligonucleotide were made in KCl (10 mM). A sample of 1 mM duplex (60 µL) was made from equal moles of complementary DNA and RNA. The drop volume was 4 µL and consisted of 0.5 mM duplex, 12 mM MaOAc, 0.6 mM spermidine.HCl, 0.075% (w/v) β-octylglucoside, 12 mM sodium cacodylate and 12% MPD. This was crystallised against a reservoir of $H_2O$:MPD, 1:1 (400 µL). To screen conditions, components of the drop were varied as specified in the text. The plates were then sealed and incubated at rt, crystals usually formed within one week (range 2-90 days, crystal size < 0.1-1.0 mm).

### *8.5.3 Crystallisation conditions*

The unmodified duplex structure was crystallised using a 24 well manual optimisation screen. Conditions were adapted based on a procedure by M. Kopka *et. al.*[143] All other structures were obtained using a NatrixHT screen. Samples were crystallised at rt unless specified.

*Table 8. 1 Table of crystallisation conditions used to produce crystals for each final structure*

| Structure | Conditions |
|---|---|
| Unmodified | Drop solution – 0.5 mM duplex, 12 mM $Mg(OAc)_2$, 0.6 mM Spermidine.HCl, 0.075% β-Octylglucoside, 12 mM Sodium cacodylate, 12 % (v/v) 2-Methyl-2,4-pentanediol (MPD), pH 5.8; Well solution – 1:1, MPD:$H_2O$ |
| Amide | Drop solution – 0.5 mM duplex, 20 mM Lithium chloride, 40 mM Strontium chloride hexahydrate, 20 mM Sodium cacodylate trihydrate pH 7.0, 15% (v/v) 2-Methyl-2,4-pentanediol, 6.0 mM Spermine tetrahydrochloride; Well solution – Component concentrations are double the concentration of the drop solution and do not contain the oligonucleotide duplex |
| LNA-amide | Drop solution – 0.5 mM duplex, 40 mM Sodium chloride, 6.0 mM Potassium chloride, 20 mM Sodium cacodylate trihydrate pH 7.0, 17.5% (v/v) 2-Methyl-2,4-pentanediol, 6.0 mM Spermine tetrahydrochloride, Well solution – Component concentrations are double the concentration of the drop solution and do not contain the oligonucleotide duplex |
| LNA-Am-LNA | Drop solution – 0.5 mM duplex, 40 mM Sodium chloride, 6.0 mM Potassium chloride, 20 mM Sodium cacodylate trihydrate pH 6.0, 15% (v/v) 2-Methyl-2,4-pentanediol, 6.0 mM Spermine tetrahydrochloride; Well solution – Component concentrations are double the concentration of the drop solution and do not contain the oligonucleotide duplex |
| CBM-LNA | 40 mM Magnesium acetate tetrahydrate, 25 mM Sodium cacodylate trihydrate pH 6.5, 15% w/v Polyethylene glycol |

### *8.5.4 Crystal collection, storage and transport*

Sample wells were opened and cryo protectant 20% glycerol in reservoir solution (v/v, 2 µL) was added. Crystals were collected under microscope using cryoloops (0.01-0.05 mm) and immediately submerged in liquid $N_2$ (77 K). While under liquid $N_2$, samples were transferred

into a holding vial. The holding vial was placed in long term storage (up to 6 months) in a liquid $N_2$ filled dewar (77 K). For transport to the synchrotron, samples were transferred under liquid $N_2$ into a sample puck. The puck was then dried shipped in a dewar to the synchrotron (77 k).

### *8.5.5 Data collection (Diamond light source)*

Data collection was performed at Diamond light source beamline i03 or i04. Beam transmission was set to 100% and the sample was rotated for collection. Images were taken at 0.1 ˚/image and a minimum of 360 ˚ was obtained. The detector distance was set to collect a maximum resolution 0.5 Å greater than the expected diffraction pattern to maximise spot separation.

### *8.5.6 Data processing and model building*

Data collected at Diamond light source was auto processed using either fast_dp, xia2_dials or xia2_3dii. Aimless reports from auto processing were checked to ensure $CC_{1/2}$ > 0.3 and completeness > 90%, crystal data quality was reviewed using Phenix.Xtriage. Molecular replacement was performed using Phaser 2.8.2 to obtain a TFZ score > 8.0 and LLG > 50. Refinement was performed using refmac5 and Phenix.Refine, in conjunction with model building using winCOOT.[223] Modified ligand restraint files were generated using Jligand or Acedrg and these were added in winCOOT. Final model building was continued until the model electron density was satisfied and the $R_{free}$ values would not reduce further. $R_{free}$ < 0.3 was considered as a correct solution. Structures were visualised using Pymol graphic software.

In some cases, data was manually reprocessed. Indexing, cell refinement and integration was performed using iMosflm or XDS.[224] The data was then scaled and merged using Aimless.[225] Data processing and model building was then the same as above.

Software packages and project management was handled using CCP4[226] and Phenix[227].

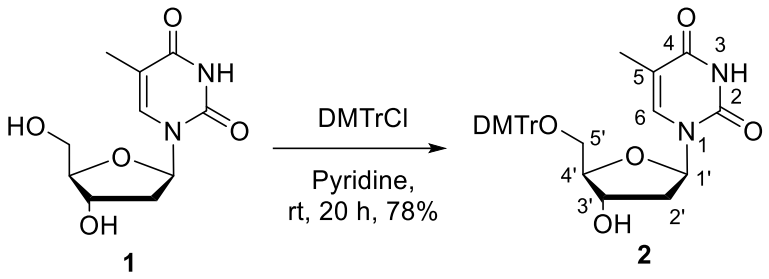
## 8.6 Table of buffers

*Table 8. 2 Table detailing buffers used for various applications*

| Buffer | Components |
|---|---|
| 10x TBE | 109 g Tris, 55 g Boric acid, 9.2 g EDTA |
| Denaturing PAGE gel composition | 29.4 g urea, 7 mL 10x TBE, x mL 40% Bis acrylamide solution, (40 – x mL) $H_2O$ |
| Native PAGE gel composition | 7 mL 10x TBE, x mL 40% Bis acrylamide solution, (63 – x mL) $H_2O$ |
| RNA gel extraction buffer | 25 mM NaCl, 50 mM Tris-HCl pH 7.5 |
| 1x TBS-T | 20 mM Tris, 150 mM NaCl, 0.1% (w/v) Tween 20 |
| 2x SDS-PAGE loading buffer | 125 mM Tris-HCl pH 6.8, 30% glycerol, 2% SDS (w/v), 5% β-mercaptoethanol (v/v), 0.03% bromophenol blue (w/v) |
| 10x Tris-glycine gel running buffer | 1.92 M glycine, 0.25 M Tris base, 1% SDS (w/v); adjust pH to 8-8.3 |
| 1x PBS | 1 PBS Dulbecco A tablet (typical formula in g/L: NaCl 8.0, KCl 0.2, $Na_2HPO_4$ 1.15, $KH_2PO_4$ 0.2, pH 7.3 ± 0.2 at 25 °C; Oxoid) per 100 mL deionised water |
| 1x PBS-T | PSB with 0.1% (v/v) Tween 20 |
| 10x stock transfer buffer | 1.92 M glycine, 0.25 M Tris base; adjust to pH 8.3 |
| 1x western blot transfer buffer | 10 % 10x stock transfer buffer, 20% MeOH, 70% deionised $H_2O$ |
| 1x Coating buffer | 1.5 g anhydrous $Na_2CO_3$, 2.93 g anhydrous $NaHCO_3$, distilled water to 1 L, pH 9.6 |

## 8.7 Synthesis and characterisation of chemical compounds

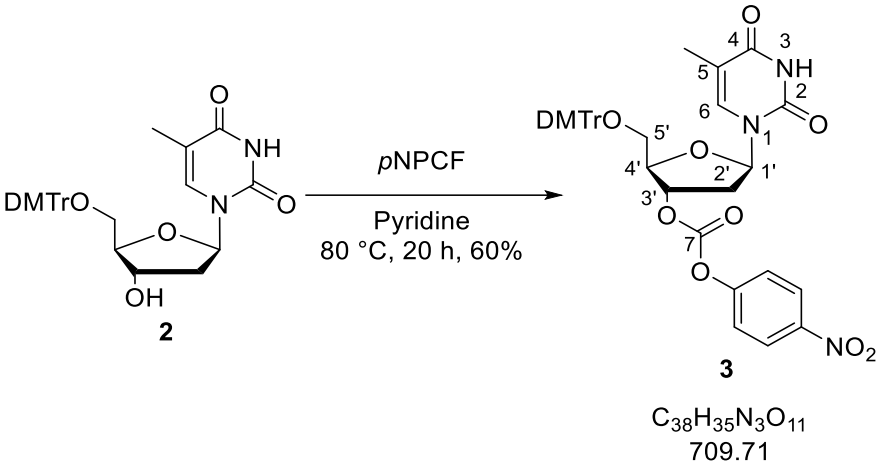
5′-*O*-(4′′-Dimethoxytriphenylmethyl)thymidine (**2**)[105]




$C_{31}H_{32}N_2O_7$
544.60

Thymidine **1** (3.00 g, 12.38 mmol) was co-evaporated in dry pyridine (3 x 10.0 mL) and dissolved in dry pyridine (60.0 mL). To this was added DMTrCl (5.03 g, 14.84 mmol) and stirred at rt for 20 h. This was then concentrated under reduced pressure and purified by silica column chromatography ($CH_2Cl_2$: MeOH, 100:0 to 19:1) to produce compound **2** (5.26 g, 9.66 mmol, 78%). as a white foam). $R_f$ 0.24 ($CH_2Cl_2$:MeOH, 19:1); $^1$H NMR (400 MHz, $CDCl_3$) δ 9.03 (s, 1H, H-3), 7.52 (q, *J* = 1.2 Hz, 1H, H-6), 7.36 – 7.28 (m, 2H, H-Ar), 7.26 – 7.13 (m, 7H, H-Ar), 6.80 – 6.72 (m, 4H, H-Ar), 6.36 (dd, *J* = 7.9, 5.9 Hz, 1H, H-1′), 4.53 – 4.47 (m, 1H, H-3′), 4.00 (appq, *J* = 3.0 Hz, 1H, H-4′), 3.71 (s, 6H, H-OMe), 3.39 (dd, *J* = 10.6, 3.0 Hz, 1H, H-5′), 3.29 (dd, *J* = 10.6, 3.0 Hz, 1H, H-5′), 2.35 (m, 1H, H-2′), 2.24 (m, 1H H-2′), 1.38 (d, *J* = 1.2 Hz, 3H, $CH_3$); $^{13}$C NMR (101 MHz, $CDCl_3$) δ 163.9 (C-4), 158.7 (C-Ar), 150.5 (C-2), 144.33 (C-Ar) 135.7 (C-6), 135.40 (C-Ar), 135.4 (C-Ar), 130.1 (C-Ar), 128.1 (C-Ar), 128.0 (C-Ar), 127.2 (C-Ar), 113.3 (C-Ar), 86.3 (C-4′), 84.8 (C-1′), 72.5 (C-3′), 63.6 (C-5′), 55.3 (C-OMe), 41.0 (C-2′), 11.85; LRMS (ESI-TOF) m/z: $[M+Na]^+$ 567 (49%), $[2M+Na]^+$ 1111 (100%).

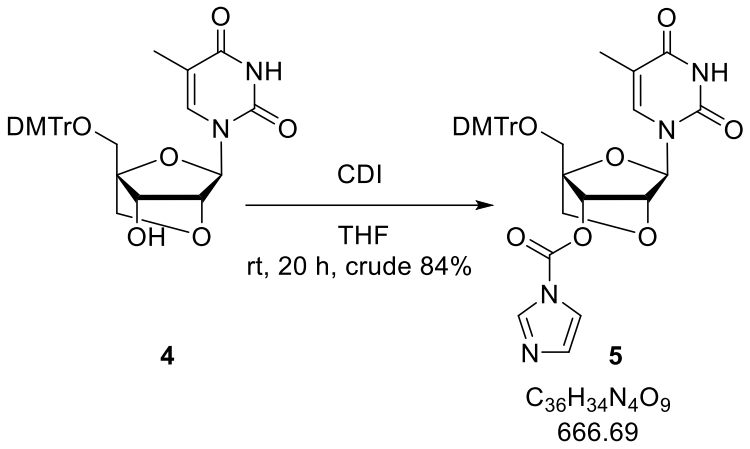
5′-O-(4′′-Dimethoxytriphenylmethyl)-3′-O-(4-Nitrophenyl formate)-thymidine (**3**)[32]




$C_{38}H_{35}N_3O_{11}$
709.71

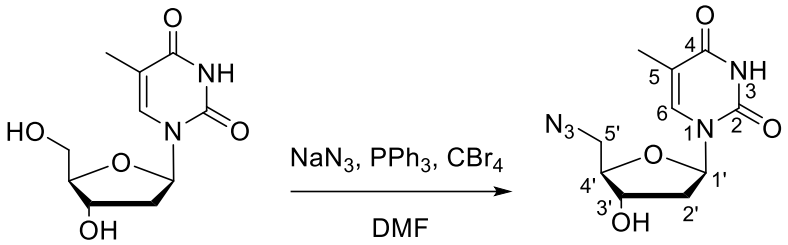
To a solution of 5′-O-(4′′-Dimethoxytriphenylmethyl)thymidine **2** (1.17 g, 2.15 mmol) in dry pyridine (20.0 mL) was added 4-Nitrophenyl chloroformate (520mg, 2.58 mmol) and stirred at 80 °C for 20 h. This was concentrated under reduced pressure and purified by silica column chromatography (EtOAc: Petroleum ether, 1:9 to 10:0) to produce compound **3** (910 mg, 1.28 mmol, 60%) as a white solid. $R_f$ 0.66 ($CH_2Cl_2$:MeOH, 9:1); $^1$H NMR (400 MHz, $CDCl_3$) δ 8.74 (s, 1H, H-3), 8.26 – 8.18 (m, 2H, H-Ar), 7.55 (d, *J* = 1.3 Hz, 1H, H-6), 7.35 – 7.28 (m, 4H, H-Ar), 7.28 – 7.16 (m, 7H, H-Ar), 6.81 – 6.73 (m, 4H, H-Ar), 6.45 (dd, *J* = 9.2, 5.3 Hz, 1H, H-1′), 5.40 – 5.34 (m, 1H, H-3′), 4.25 (appq, *J* = 1.8 Hz, 1H, H-4′), 3.72 (s, 6H, H-OMe), 3.51 (dd, *J* = 10.6, 2.8 Hz, 1H, H-5′), 3.41 (dd, *J* = 10.6, 2.4 Hz, 1H, H-5′), 2.60 (m, 1H, H-2′), 2.45 (m, 1H, H-2′), 1.35 (d, *J* = 1.2 Hz, 3H, $CH_3$-*thymidine*); $^{13}$C NMR (101 MHz, $CDCl_3$) δ 163.5 (C-4), 158.9 (C-Ar), 155.1 (C-Ar), 151.9 (C-7), 150.4 (C-2), 145.6 (C-Ar), 144.1 (C-Ar), 135.1 (C-6), 135.0 (C-Ar), 130.1 (C-Ar), 130.01 (C-Ar), 128.1 (C-Ar), 128.1 (C-Ar), 127.3 (C-Ar), 125.4 (C-Ar), 121.7 (C-Ar), 113.4 (C-Ar), 111.9 (C-5), 87.4 (C-$CAr_2$Ph), 84.3 (C-1′), 83.5 (C-3′), 80.4 (C-4′), 63.7 (C-5′), 55.3 (C-OMe), 37.9 (C-2′), 11.71 (C-$CH_3$-*thymidine)*; LRMS (ESI-TOF) m/z: $[M+Na]^+$ 732 (100%)

3′-*O*-Carbonylimidazole-5′-*O*-dimethoxytrityl-LNA-thymidine (**5**) (*Synthesised by Benjamin Woods)*




Compound **4** (1.11 g, 1.94 mmol) and 1,1'-carbonyldiimidazole (0.630 g, 3.88 mmol) were separately co-evaporated in dry THF (3 × 5.00 mL), and each re-dissolved in dry THF (20.0 mL). The solution of compound 4 was then added dropwise to a solution of CDI over 1 h and stirred for further 20 h at rt. The solvent was removed under reduced pressure and re-dissolved in EtOAc (100 mL), washed with 5% $KH_2PO_4$ (2 × 100 mL) and dried over $Na_2SO_4$. Removal of the solvent under reduced pressure gave intermediate **5** (1.08 g, 1.61 mmol, crude 84%) as a white foam which was used without further purification. $R_f$ 0.33 ($CH_2Cl_2$:MeOH, 19:1); HRMS–ESI (*m/z*): $[M + H]^+$ calcd for $C_{36}H_{35}N_4O_9$, 667.2399; found, 667.2393.

5′-Azido-5′-deoxythymidine (**6**)[106]



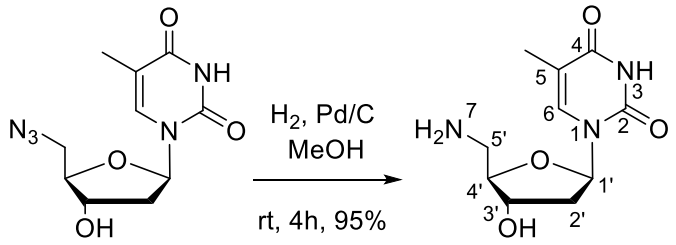

**1**

**6**

$C_{10}H_{13}N_5O_4$
267.25

Thymidine **1** (5.00 g, 21.0 mmol) was co-evaporated in pyridine (2x 20.0 mL) and dissolved in DMF (80.0 mL). To this was added $NaN_3$ (4.03 g, 62.0 mmol) and $PPh_3$ (7.08 g, 27.0 mmol) followed by $CBr_4$ (10.3 g, 31.0 mmol). After stirring at rt for 20 h the reaction was concentrated under vacuum and purified by silica column chromatography solid ($CH_2Cl_2$: MeOH, 100:0 to 90:10) to produce compound **6** as a white (4.17 g, 15.6 mmol, 74%). $R_f$ 0.47 ($CH_2Cl_2$: MeOH, 9:1); IR ($cm^{-1}$): 3384 (br m, OH), 3182 (m, CH), 2098 (s, $N_3$), 1718 (s, CO), 1651 (s CO); $^1$H NMR (400 MHz, DMSO-$d_6$) δ 11.35 (s, 1H, H-3), 7.50 (d, *J* = 1.2 Hz, 1H, H-6), 6.21 (dd, *J* = 7.6, 6.5 Hz, 1H, H-1′), 5.42 (d, *J* = 4.1 Hz, 1H, H-7), 4.20 (dq, *J* = 7.5, 4.1 Hz, 1H, H-3′), 3.85 (m, 1H, H-4′), 3.56 (d, *J* = 5.3 Hz, 2H, H-5′), 2.26 (m, 1H, H-2′), 2.09 (m, 1H, H-2′), 1.80 (d, *J* = 1.2 Hz, 3H, $CH_3$-*thymidine*); $^{13}$C NMR (101 MHz, DMSO-$d_6$) δ 164.1 (C-4), 150.9 (C-2), 136.6 (C-6), 110.3 (C-5), 85.0 (C-4′), 84.3 (C-1′), 71.2 (C-3′), 52.14 (C-5′), 38.5 (C-2′), 12.6 (C-$CH_3$ *thymidine*); LRMS (ESI-TOF) m/z: $[M+K]^+$ 306 (100%), $[M-H]^-$ 266 (53%), $[M+HCOO^-]^-$ 312 (100%)

5′-Amino-5′-deoxythymidine (**7**)[107]




**6**

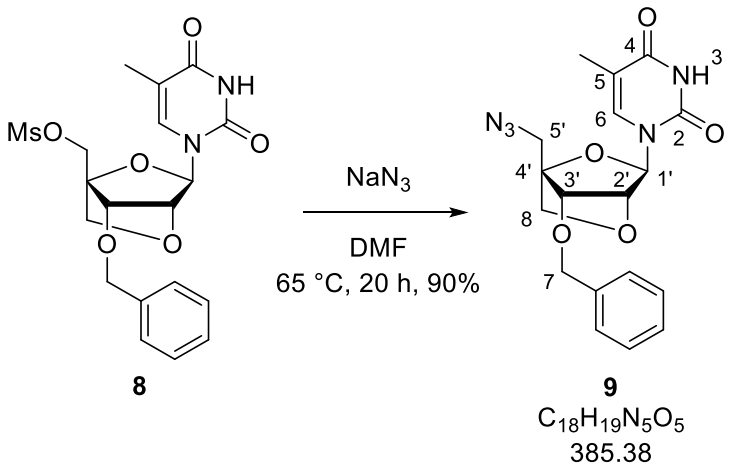
**7**

$C_{10}H_{15}N_3O_4$

241.25

To a solution of 5′-azido-5′-deoxythymidine **6** (2.10 g, 7.86 mmol) in MeOH (100 mL) was added 10% Pd/C (420 mg, 0.390 mmol) and flushed with an argon atmosphere. The solution was then degassed with hydrogen for 20 mins and stirred under 1 atmosphere of hydrogen at rt for 4 h. The reaction was filtered through celite and the filtrate concentrated under vacuum to produce compound **7** as a white solid (1.76 g, 7.30 mmol, 95%). $R_f$ 0.35 ($CH_2Cl_2$:MeOH, 19:1); $^1$H NMR (400 MHz, DMSO-$d_6$) δ 7.65 (d, *J* = 1.3 Hz, 1H, H-6), 6.14 (dd, *J* = 7.6, 6.3 Hz, 1H, H-1′), 4.20 (dt, *J* = 6.5, 3.3 Hz, 1H, H-3′), 3.66 (td, *J* = 5.3, 3.3 Hz, 1H, H-4′), 3.40 (br s, 2H, H-7), 2.74 (dd, *J* = 5.3, 1.7 Hz, 2H, H-5′), 2.14 (m, 1H, H-2′), 2.04 (m, 1H, H-2′), 1.79 (d, *J* = 1.3 Hz, 4H, $CH_3$-*thymidine*); $^{13}$C NMR (101 MHz, DMSO-$d_6$) δ 164.2 (C-4), 150.9 (C-2), 136.8 (C-6), 110.0 (C-5), 88.0 (C-4′), 83.9 (C-1′), 71.2 (C-3′), 44.0 (C-5′), 39.3 (C-2′), 12.6 (C-$CH_3$-*thymidine*); LRMS (ESI-TOF) m/z: $[M+HCOO^-]^-$ 286 (100%).

5′-Azido-3′-benzyl-5′-deoxy-LNA-thymidine (**9**)[108] (*Synthesised by Sven Epple)*



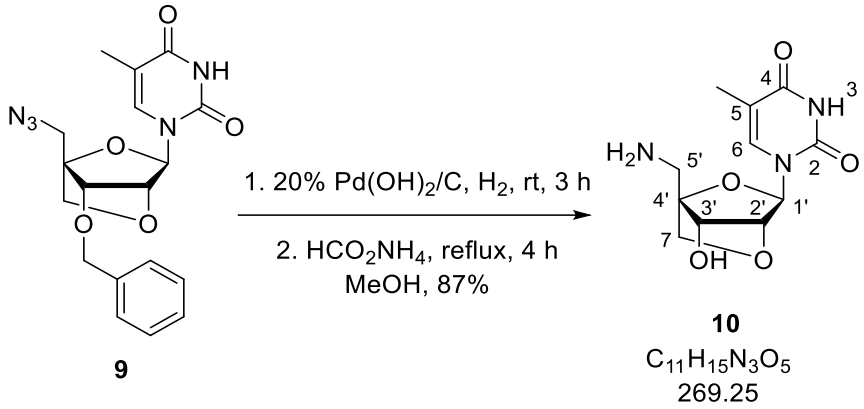

**8** **9**

$C_{18}H_{19}N_5O_5$

385.38

To a solution of compound **8** (1.00 g, 2.28 mmol) in DMF (7.00 mL) was added sodium azide (1.01 g, 15.5 mmol) and stirred at 65 °C for 20 h. The solvent was removed under reduced pressure, re-suspended in EtOAc (70.0 mL) and filtered over silica. Evaporation of the solvent gave an oil which was purified by silica column chromatography (Petroleum ether:EtOAc, 3:7 to 0:10) to give azide **9** (0.790 g, 2.05 mmol, 90%) as a white solid. $R_f$ 0.60 ($CH_2Cl_2$:MeOH, 9:1); IR ($cm^{-1}$): 2102; $^1$H NMR (400 MHz, $CDCl_3$) δ 8.88 (bs, 1H, H-3), 7.40 (d, *J* = 1.2 Hz, 1H, H-6), 7.39 – 7.27 (m, 5H, H-Ar), 5.64 (s, 1H H-1′), 4.66 (d, *J* = 11.4 Hz, 1H, H-8), 4.58 (s, 1H, H-2′), 4.54 (d, *J* = 11.4 Hz, 1H, H-8), 4.03 (d, *J* = 7.9 Hz, 1H, H-7), 3.87 (s, 1H, H-3′), 3.84 (d, *J* = 7.9 Hz, 1H, H-7), 3.75 (d, *J* = 13.7 Hz, 1H, H-5′), 3.67 (d, *J* = 13.7 Hz, 1H, H-5′), 1.94 (d, *J* = 1.2 Hz, 3H, $CH_3$); $^{13}$C NMR (101 MHz, $CDCl_3$) δ 163.7 (C-4), 149.8 (C-2), 136.7 (C-Ar), 134.4 (C-6), 128.7 (C-Ar), 128.5 (C-Ar), 128.0 (C-Ar), 110.8 (C-5), 87.8 (C-1′), 86.9 (C-4′), 76.8 (C-2′), 76.4 (C-3′), 72.5 (C-7), 72.5 (C-8), 47.5 (C-5′), 12.9 ($CH_3$); HRMS–ESI (*m/z*): $[M+H]^+$ calcd for $C_{18}H_{20}N_5O_5$, 386.1459; found, 386.1460.

5′-Amino-5′-deoxy-LNA-thymidine (**10**) (*Synthesised by Sven Epple)*

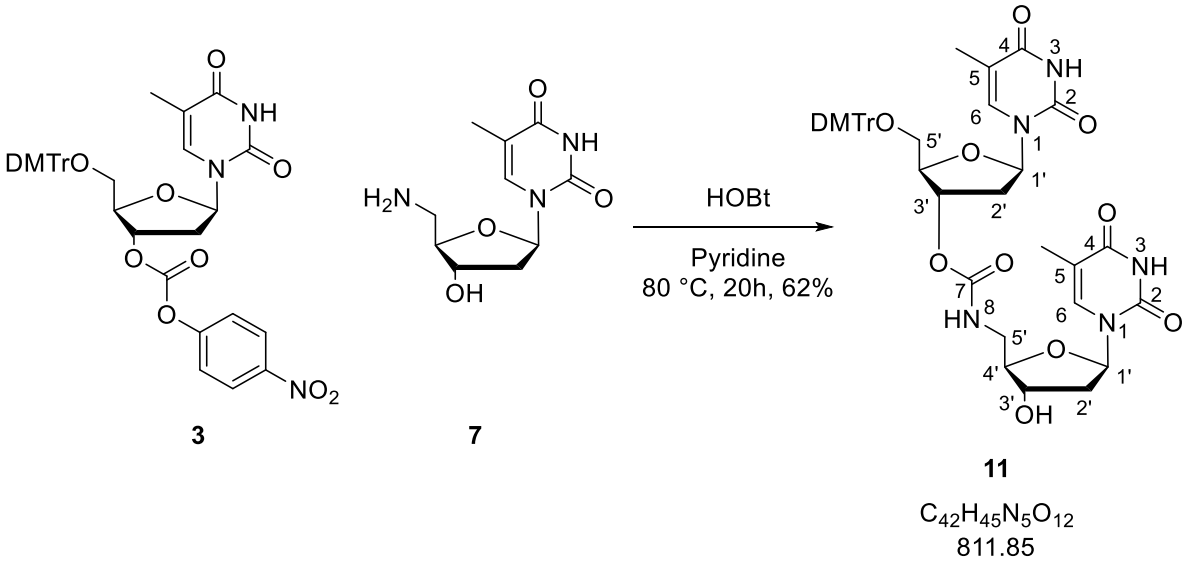



**9**

**10**
$C_{11}H_{15}N_3O_5$
269.25

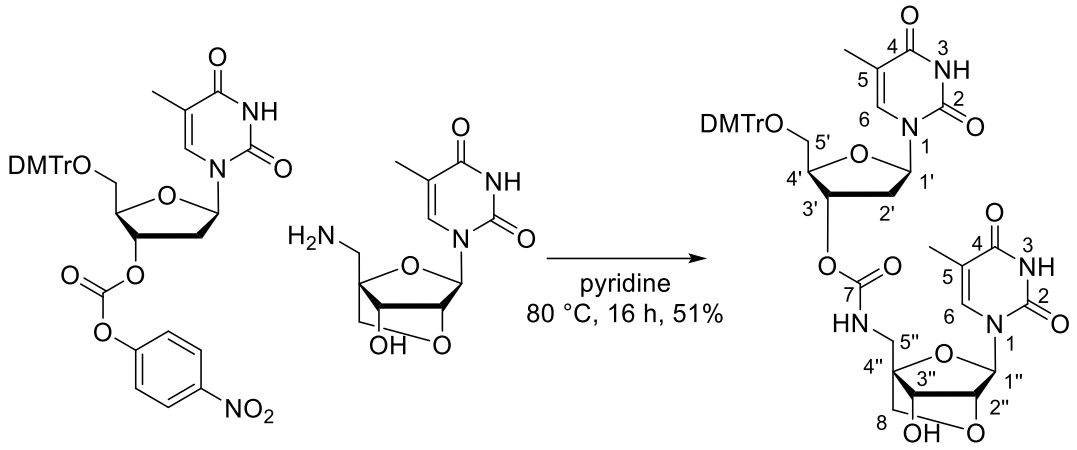
To a solution of azide **9** (1.50 g, 3.90 mmol) in MeOH (70.0 mL) was added 20% $Pd(OH)_2/C$ (338 mg, 2.41 mmol) and stirred under $H_2$ atmosphere at rt for 3 h. After complete consumption of the starting material by TLC ($CH_2Cl_2$:MeOH, 17:1), ammonium formate (1.84 g, 29.2 mmol) was added and the reaction was refluxed under argon for 4 h. The reaction mixture was diluted with $CH_2Cl_2$ (30.0 mL) and MeOH (30.0 mL), filtered over celite and concentrated under reduced pressure. The residue was purified by silica column chromatography ($CH_2Cl_2$:MeOH, 97:3 to 19:1 + 1% $Et_3N$) to give amine **10** (0.960 g, 3.58 mmol, 87%) as a white solid. $R_f$ 0.25 ($CH_2Cl_2$:MeOH:$Et_3N$, 84:15:1); $^1$H NMR (400 MHz, $D_2O$) δ 7.53 (d, *J* = 1.4 Hz, 1H, H-6), 5.66 (s, 1H, H-1′), 4.50 (s, 1H, H-2′), 4.20 (s, 1H, H-3′), 4.02 (q, *J* = 8.6 Hz, 2H, H-7), 3.37 (d, *J* = 14.5 Hz, 1H, H-5′), 3.29 (d, *J* = 14.3 Hz, 1H, H-5′), 1.88 (d, *J* = 1.2 Hz, 3H, $CH_3$); $^{13}$C NMR (101 MHz, $D_2O$) δ 166.7 (C-4), 151.2 (C-2), 136.0 (C-6), 110.6 (C-5), 86.9 (C-4′), 86.9 (C-1′), 79.4 (C-2′), 71.6 (C-7), 70.4 (C-3′), 37.0 (C-5′), 11.6 ($CH_3$); HRMS–ESI (*m/z*): $[M + H]^+$ calcd for $C_{11}H_{16}N_3O_5$, 270.1085; found, 270.1085.

5′-O-(4′′-Dimethoxytriphenylmethyl)thymidine carbamate dimer (**11**) (DNA-CBM$_1$-DNA)[32]




To a solution of 5′-O-(4′′-Dimethoxytriphenylmethyl)-3′-O-(4-Nitrophenyl formate)-thymidine **3** (400 mg, 0.560 mmol) and HOBt (7.60 mg, 56.2 µmol) in dry pyridine (20.0 mL) was added 5′-Amino-5′-deoxythymidine **7** (200 mg, 0.820 mmol) and stirred at 80 °C for 20 h. This was purified by silica column chromatography ($CH_2Cl_2$:MeOH, 100:0 to 9:1) to give dimer **11** a white solid (300 mg, 0.370 mmol 66%). $R_f$ 0.43 ($CH_2Cl_2$:MeOH, 9:1); $^1$H NMR (400 MHz, DMSO-$d_6$) δ 11.39 (s, 1H, H-3), 11.30 (s, 1H, H-3), 7.61 (s, 1H, H-8), 7.56 – 7.51 (m, 1H, H-6), 7.50 (d, *J* = 1.3 Hz, 1H, H-6), 7.42 – 7.28 (m, 4H, H-Ar), 7.29 – 7.20 (m, 4H, H-Ar), 6.90 (dt, *J* = 9.0, 2.0 Hz, 4H, H-Ar), 6.24 (dd, *J* = 8.7, 5.8 Hz, 1H, H-1′), 6.14 (dd, *J* = 8.0, 6.0 Hz, 1H, H-1′), 5.31 (d, *J* = 4.3 Hz, 1H, H-3′OH), 5.24 (d, *J* = 5.9 Hz, 1H, H-3′), 4.14 (t, *J* = 4.3 Hz, 1H, H-3′), 4.05 (br appq, *J* = 3.2 Hz, 1H, H-4′), 3.74 (s, 6H, H-OMe), 3.30 – 3.16 (m, 4H, H-4′, H-5′), 2.48 – 2.39 (m, 2H, H-2′, H-5′), 2.29 (m, 1H, H-2′), 2.15 – 1.98 (m, 2H, H-2′), 1.77 (d, *J* = 1.3 Hz, 3H, $CH_3$-*thymidine*), 1.42 – 1.37 (m, 3H, $CH_3$-*thymidine*); $^{13}$C NMR (101 MHz, DMSO-$d_6$) δ 164.2 (C-4), 164.1 (C-4), 158.7 (C-Ar), 156.1 (C-Ar), 150.9 (C-2), 150.9 (C-2), 145.0 (C-Ar) , 136.5 (C-6), 135.9 (C-Ar), 135.5 (C-6), 130.2 (C-Ar), 128.4 (C-Ar), 128.1 (C-Ar), 113.8 (C-Ar), 110.4 (C-5), 86.6 (C-Ar2Ph), 85.4 (C-4′), 84.4 (C-1′),75.1 (C-3′), 71.6 (C-3′), 55.5 (C-OMe), 38.8 (C-2′), 12.5 (C-$CH_3$-*thymidine)*, 12.1 (C-$CH_3$-*thymidine*; LRMS (ESI-TOF) m/z: $[M-H]^-$ 810 (100%)

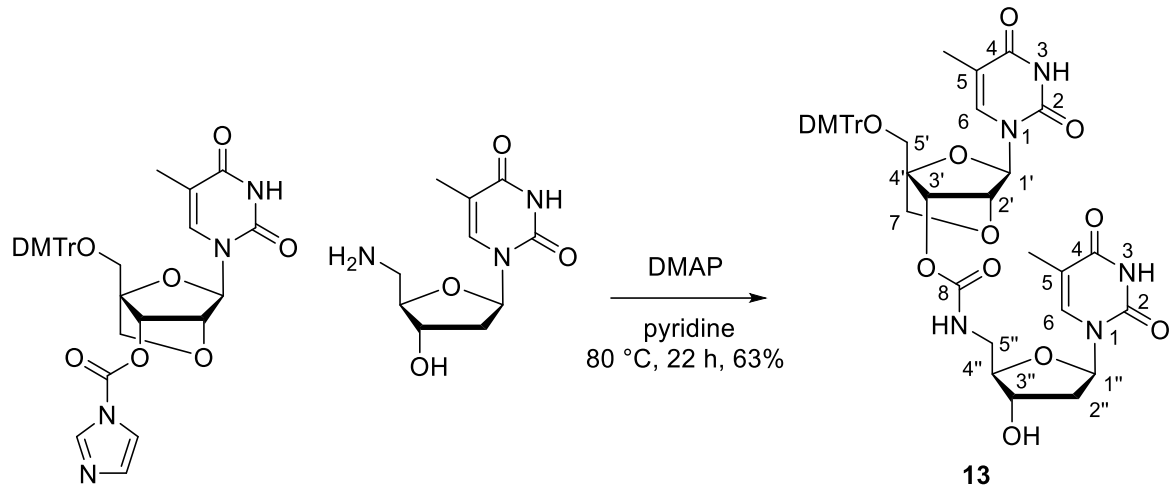
5′-*O*-(Dimethoxytrityl)-DNA-CBM$_1$-LNA-thymidine dimer (DNA-CBM$_1$-LNA) (**12**) (*Synthesised by Sven Epple)*




Amine **10** (282 mg, 1.05 mmol) and carbonate **3** (750 mg, 1.06 mmol) were separately co-evaporated with dry pyridine (3 × 5.00 mL) and each dissolved in dry pyridine (5.00 mL). The solution of carbonate **3** was added dropwise to amine **10** and stirred at 80 °C for 16 h. The reaction was concentrated under reduced pressure and purified by silica column chromatography (EtOAc:MeOH, 49:1 to 25:1 + 1% $Et_3N$). This gave a white solid which was re-dissolved in $CHCl_3$ : $^i$PrOH, [3/2, (v/v), 50 mL] and washed with sat. aq $NaHCO_3$ until the aq layer stayed colourless. The organic layer was dried over $Na_2SO_4$ and filtered over Celite. Evaporation of the solvents gave dimer **12** (452 mg, 0.540 mmol, 51%) as a white solid. $R_f$ 0.38 ($CH_2Cl_2$:MeOH, 9:1); $^1$H NMR (400 MHz, DMSO-$d_6$) δ 11.39 (s, 1H, H-3), 11.37 (s, 1H, H-3), 7.79 (t, *J* = 6.2 Hz, 1H, NH CBM), 7.52 (d, *J* = 1.0 Hz, 1H, H-6), 7.42 (d, *J* = 1.2 Hz, 1H, H-6), 7.40 – 7.19 (m, 9H, H-Ar), 6.89 (dq, *J* = 9.5, 2.5 Hz, 4H, H-Ar), 6.25 (dd, *J* = 8.6, 5.9 Hz, 1H, H-1′), 5.57 (d, *J* = 4.1 Hz, 1H, OH), 5.41 (s, 1H, H-1′′), 5.27 (dd, *J* = 5.5, 3.3 Hz, 1H, H-3′), 4.12 (s, 1H, H-2′′), 4.05 (q, *J* = 3.3 Hz, 1H, H-4′), 3.89 – 3.82 (m, 2H, H-8, H-3′′), 3.73 (s, 6H, $OCH_3$), 3.62 (d, *J* = 8.0 Hz, 1H, H-8), 3.54 (t, *J* = 6.2 Hz, 2H, H-5′′), 3.31 (d, *J* = 3.9 Hz, 1H, H-5′), 3.21 (dd, *J* = 10.4, 3.1 Hz, 1H, H-5′), 2.45 (q, *J* = 10.0 Hz, 1H, H-2′), 2.29 (dq, *J* = 10.0 , 4.7, 3.3 Hz, 1H, H-2′), 1.76 (d, *J* = 1.2 Hz, 3H, $CH_3$), 1.40 (d, *J* = 1.0 Hz, 3H, $CH_3$); $^{13}$C NMR (101 MHz, DMSO-$d_6$) δ 164.3 (C-4), 164.1 (C-4), 158.7 (C-Ar), 158.7 (C-Ar), 156.3 (C-7), 150.9 (C-2), 150.4 (C-2), 145.1 (C-Ar), 135.9 (C-Ar), 135.8 (C-6), 135.6 (C-Ar), 135.1 (C-6), 130.2 (C-Ar), 128.4 (C-Ar), 128.1 (C-Ar), 127.3 (C-Ar), 113.7 (C-Ar), 110.4 (C-5), 109.0 (C-5), 87.6 (C-4′′), 86.8 (C-1′′), 86.5, 84.1 (C-1′), 83.8 (C-4′),

79.5 (C-2''), 75.1 (C-3'), 71.9 (C-8), 70.3 (C-3''), 64.2 (C-5''), 55.5 ($OCH_3$), 37.7 (C-5'), 37.2 (C-2'), 12.8 ($CH_3$), 12.1 ($CH_3$); HRMS–ESI (*m/z*): $[M + Na]^+$ calcd for $C_{43}H_{45}N_5O_{13}Na$, 862.2906; found, 862.2906.

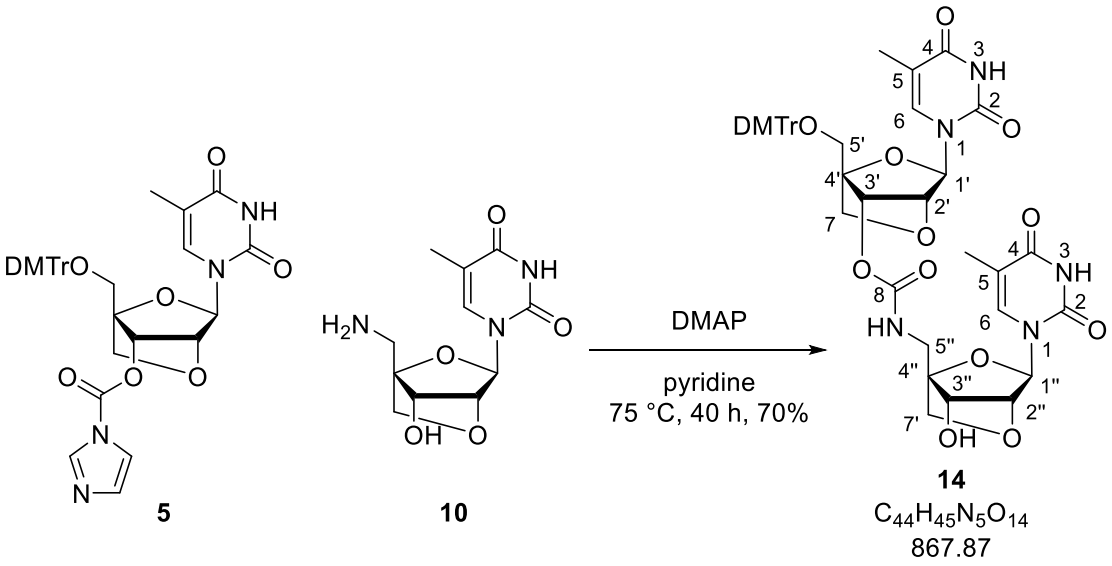
5'-*O*-(Dimethoxytrityl)-LNA-$CBM_1$-DNA-thymidine dimer (LNA-$CBM_1$-DNA) (**13**) (*Synthesised by Benjamin Woods)*




Carbamate **5** (542 mg, 0.813 mmol) and amine **7** (196 mg, 0.813 mmol) were separately co-evaporated in dry pyridine (2 × 10.0 mL) and each re-dissolved in dry pyridine (5.00 mL). To the solution of amine **7** was added carbamate **5** and stirred under argon at 80 °C for 20 h. A catalytic amount of DMAP (9.93 mg, 0.0813 mmol) was added and the reaction was stirred for a further 2 h at 80 °C. The solvent was removed under reduced pressure, re-dissolved in $CH_2Cl_2$ (70.0 mL) and extracted with 5% $KH_2PO_4$. The aq layer was back extracted with $CH_2Cl_2$ (2 × 70.0 mL) and dried over $Na_2SO_4$ and purified by silica column chromatography (EtOAc:MeOH, 99:0 to 90:9 + 1% $Et_3N$) to give dimer **13** (0.430 g, 0.510 mmol, 63%) as a white solid. $R_f$ 0.16 ($CH_2Cl_2$:MeOH, 93:7); $^1$H NMR (400 MHz, DMSO-$d_6$) δ 11.57 (s, 1H, H-3), 11.37 (s, 1H, H-3), 7.82 (t, *J* = 6.0 Hz, 1H, NH CBM), 7.62 (d, 1H, *J* = 1.2 Hz, H-6), 7.56 (d, 1H, *J* = 1.2 Hz, H-6), 7.50 – 7.44 (m, 2H, H-Ar), 7.43 – 7.30 (m, 7H, H-Ar), 6.97 (dd, *J* = 9.0, 2.1 Hz, 4H, H-Ar), 6.20 (dd, *J* = 7.7, 6.4 Hz, 1H, H-1''), 5.59 (s, 1H, H-1'), 5.38 (d, J = 4.0 Hz, 1H, OH), 5.18 (s, 1H, H-3'), 4.50 (s, 1H, H-2'), 4.21– 4.16 (m, 1H, H-3''), 3.88 (d, *J* = 8.2 Hz, 2H, H-7), 3.80 (s, 7H, H-4'', $OCH_3$), 3.45 (2 × d *J* = 11.5 Hz, 2H, H-5'), 3.34 – 3.22 (m, 2H, H-5''), 2.14 – 2.08 (m, 2H, H-

2''), 1.82 (d, *J* = 1.2 Hz, 3H, $CH_3$), 1.65 (d, *J* = 1.2 Hz, 3H, $CH_3$); $^{13}$C NMR (101 MHz, DMSO-$d_6$) δ 164.2 (C-4), 158.6 (C-Ar), 155.3 (C-8), 150.9 (C-2), 150.4 (C-2), 145.0 (C-Ar) , 136.5 (C-6), 135.6 (C-Ar), 135.3 (C-Ar), 134.2 (C-6), 130.2 (C-Ar), 130.1 (C-Ar), 128.5 (C-Ar), 128.0 (C-Ar), 127.3 (C-Ar), 113.8 (C-Ar), 110.1 (C-5), 109.4 (C-5), 86.9 (C-1''), 86.7 (C-4'), 86.4 ($C_{quat}$DMTr), 85.4 (C-4''), 84.6 (C-1'), 78.0 (C-2'), 72.2 (C-7), 71.6 (C-3''), 71.3 (C-3'), 55.5 ($OCH_3$), 43.6 (C-5''), 38.9 (C-2''), 12.9 ($CH_3$), 12.3 ($CH_3$); HRMS–ESI (*m/z*): $[M+Na]^+$ calcd for $C_{43}H_{45}N_5O_{13}Na$, 862.2906, found 862.2906.

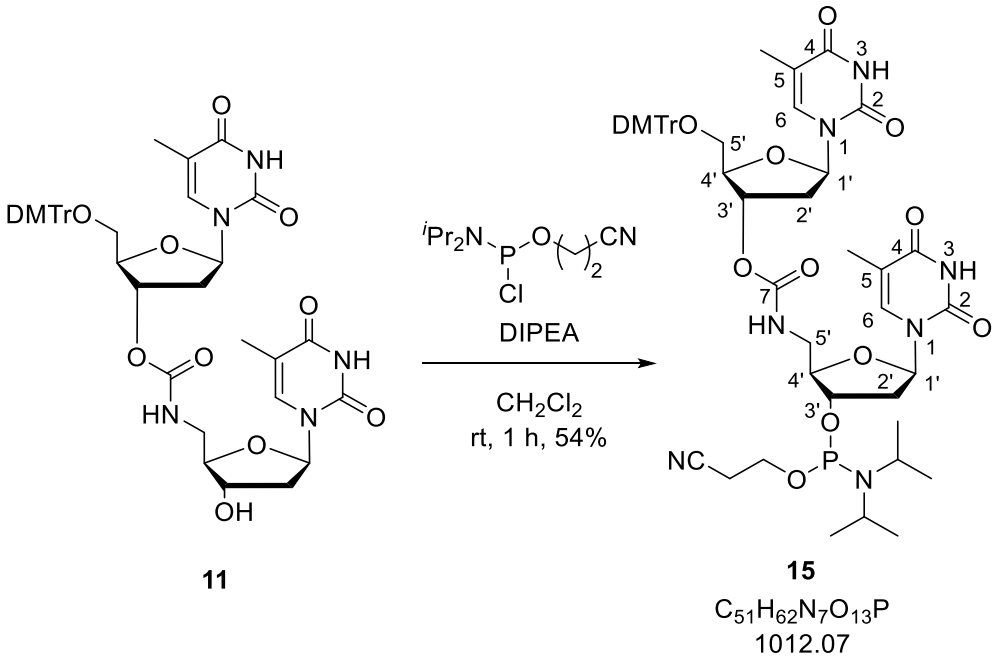
5'-*O*-(Dimethoxytrityl)-LNA-$CBM_1$-LNA-thymidine dimer (LNA-$CBM_1$-LNA) (**14**) (*Synthesised by Sven Epple)*




Amine **10** (202 mg, 0.750 mmol) and carbamate **5** (500 mg, 0.750 mmol) were separately co-evaporated with dry pyridine (3 × 5.00 mL) and each dissolved in dry pyridine (5.00 mL). Carbamate **5** was added dropwise to the solution of amine **10** and the reaction was stirred at 80 °C for 16 h. A catalytic amount of DMAP (9.00 mg, 0.08 mmol) was added and the mixture was stirred at 80 °C for another 24 h. The reaction was then cooled to rt and the solvents were evaporated under reduced pressure. The remaining solid was re-dissolved in $CH_2Cl_2$ (70.0 mL) and the organic phase was washed with a sat. aq $NaHCO_3$ (2 × 70.0 mL), brine (50.0 mL) and dried over $MgSO_4$. The crude was concentrated under reduced pressure and purified by silica column chromatography (EtOAc:MeOH, 100:0 to 96:4 + 1% $Et_3N$) to give dimer **14** (453 mg,

520 µmol, 70%) as a white solid. $R_f$ 0.41 (EtOAc:MeOH:$Et_3N$, 25:4:1); $^1$H NMR (400 MHz, DMSO-$d_6$) δ 11.50 (s, 1H, H-3), 11.39 (s, 1H, H-3), 7.91 (t, *J* = 6.1 Hz, 1H, NH CBM), 7.63 – 7.48 (m, 1H, H-6), 7.45 – 7.17 (m, 10H, H-Ar, H-6)), 6.96 – 6.81 (m, 4H, H-Ar), 5.62 (d, *J* = 4.1 Hz, 1H, OH), 5.53 (s, 1H, H-1′ or H-1′′), 5.39 (s, 1H, H-1′ or H-1′′), 5.18 (s, 1H, H-3′), 4.47 (s, 1H, H-2′), 4.11 (s, 1H, H-2′′), 3.91 – 3.76 (m, 5H, H-7, H-7′, H-3′′), 3.72 (d, *J* = 4.1 Hz, 6H, $OCH_3$), 3.65 – 3.47 (m, 2H, H-5′ or H-5′′), 3.47 – 3.38 (m, 2H, H-5′ or H-5′′), 1.72 (d, J = 1.1 Hz, 3H, $CH_3$), 1.55 (d, J = 1.1, 3H, $CH_3$); $^{13}$C NMR (101 MHz, DMSO-$d_6$) δ 163.8 (C-4), 163.7 (C-4), 158.2 (C-Ar), 158.2 (C-Ar), 155.0 (C-8), 150.0 (C-2), 149.9 (C-2), 144.5 (C-Ar), 135.0 (C-Ar), 134.9 (C-Ar), 134.8 (C-6), 133.8 (C-6), 129.7 (C-Ar), 128.0 (C-Ar), 127.6 (C-Ar), 126.9 (C-Ar), 113.3 (C-Ar), 109.0 (C-5), 108.5 (C-5), 87.1 ($C_{quat}$DMTr), 86.4 (C-1′ or C-1′′), 86.3 (C-1′ or C-1′′), 85.9 (C-4′, C-4′′), 79.0 (C-2′′), 77.5 (C-2′), 71.7(C-7, C-7′) 71.4 (C-5′), 70.9 (C-3′), 70.2 (C-3′′), 55.1 ($OCH_3$), 55.0 ($OCH_3$), 12.4 ($CH_3$), 12.2 ($CH_3$); HRMS–ESI (m/z): $[M + Na]^+$ calcd for $C_{44}H_{45}N_5O_{14}Na$, 890.2855; found, 890.2857.

5′-O-(4′′-Dimethoxytriphenylmethyl)thymidine carbamate phosphoramidite dimer (DNA-$CBM_1$-DNA) (**15**)



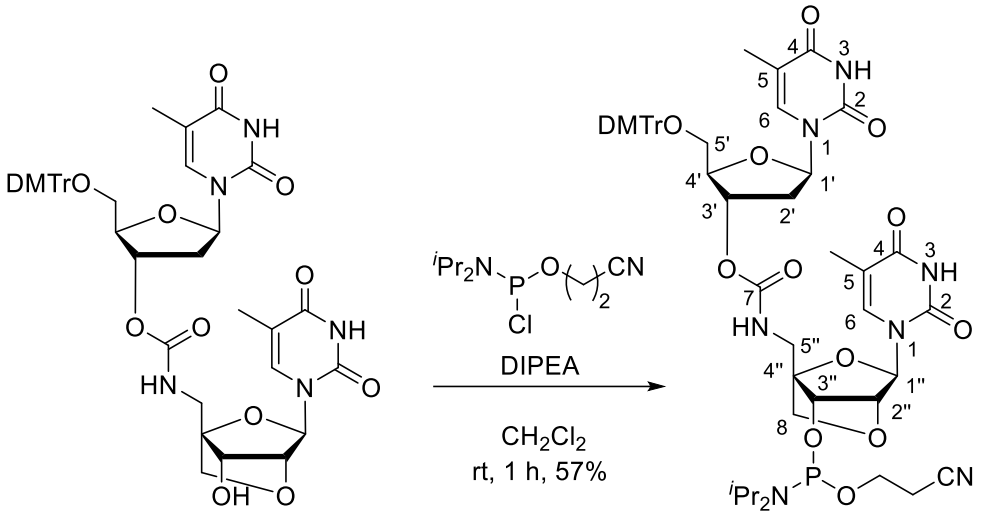

**11**

**15**
$C_{51}H_{62}N_7O_{13}P$
1012.07

5′-O-(4′′-Dimethoxytriphenylmethyl)thymidine carbamate dimer **11** (300 mg, 0.370 mmol) was co-evaporated with degassed $CH_2Cl_2$ (3x 2.00 mL) and dissolved in dry degassed $CH_2Cl_2$

(5.00 mL). To this was added degassed DIPEA (160 µL, 0.919 mmol) followed by 2-Cyanoethyl *N,N*-diisopropylchlorophosphoramidite (99 µL) and stirred under argon for 1 h. The reaction was diluted in degassed $CH_2Cl_2$ (5.00 mL), washed with sat KCl solution (10 mL) and dried over $Na_2SO_4$. The organic layer was concentrated and purified by silica column chromatography using degassed solvent (EtOAc) to give compound **15** as a white solid (200 mg, 0.198 mmol, 54%). $R_f$ 0.61 (EtOAc); $^{31}P$ NMR (162 MHz, $CDCl_3$) δ 149.4, 148.9; HRMS–ESI (*m/z*): $[M + H]^+$ calcd for $C_{51}H_{62}N_7O_{13}P$, 1012.42160, found, 1012.42181.

5′-*O*-(Dimethoxytrityl)-DNA-CBM$_1$-LNA-thymidine phosphoramidite dimer (DNA-CBM$_1$-LNA) (**16**) (*Synthesised by Sven Epple)*



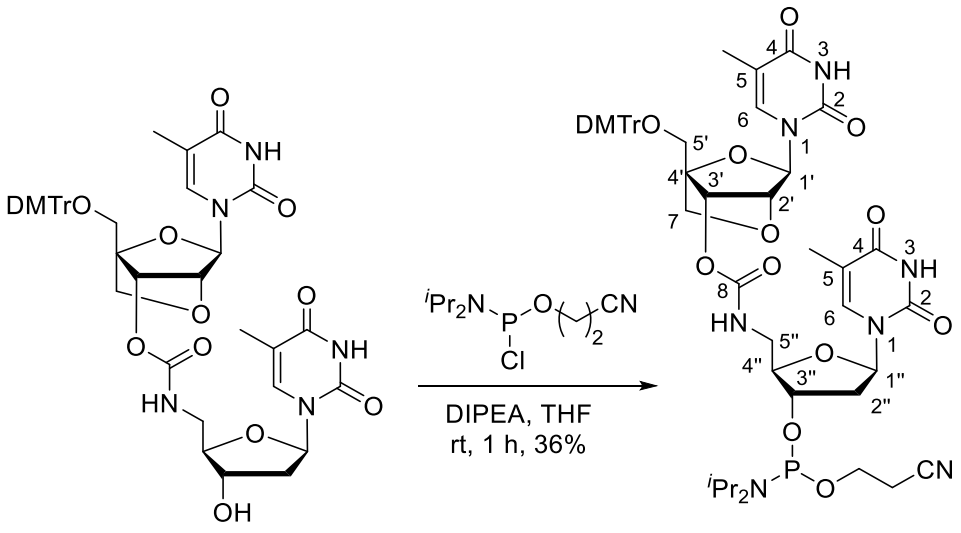

**12**

**16**
$C_{52}H_{62}N_7O_{14}P$
1040.08

To a solution of dimer **12** (480 mg, 0.570 mmol) in dry, degassed $CH_2Cl_2$ (5.00 mL), was added degassed DIPEA (250 µL, 1.44 mmol) followed by chloro(diisopropylamino)-*β*-cyanoethoxyphosphine (180 µL, 0.810 mmol) dropwise over 10 min. After 1 h at rt, a second portion of chloro(diisopropylamino)-*β*-cyanoethoxyphosphine (120 µL, 0.540 mmol) and DIPEA (120 µL, 0.690 mmol) was added and stirred for another 1 h. The reaction mixture was diluted with dry $CH_2Cl_2$ (10.0 mL) and washed with sat. aq KCl (10.0 mL). The organic layer was dried over $Na_2SO_4$ and the solvents were evaporated under reduced pressure. The solid was re-dissolved in dry, degassed $CH_2Cl_2$ (2.50 mL) and precipitated by addition of degassed hexane (250 mL). The suspension was cooled to 4 °C and the supernatant was decanted. The

remaining slurry was dried under reduced pressure to give a white solid. Purification by column chromatography (Hexane: EtOAc, 1:10 + 1% pyridine) gave dimer **16** (338 mg, 0.320 mmol, 57%) as a white solid. $R_f$ 0.51 (hexane:EtOAc:pyridine, 10:100:1); $^{31}$P NMR (162 MHz, $CDCl_3$) δ 149.5, 149.1; HRMS–ESI (*m/z*): $[M + H]^+$ calcd for $C_{52}H_{63}N_7O_{14}P$, 1040.4165, found, 1040.4183.

5′-*O*-(Dimethoxytrityl)-LNA-CBM$_1$-DNA-thymidine phosphoramidite dimer (LNA-CBM$_1$-DNA) (**17**) (*Synthesised by Benjamin Woods)*

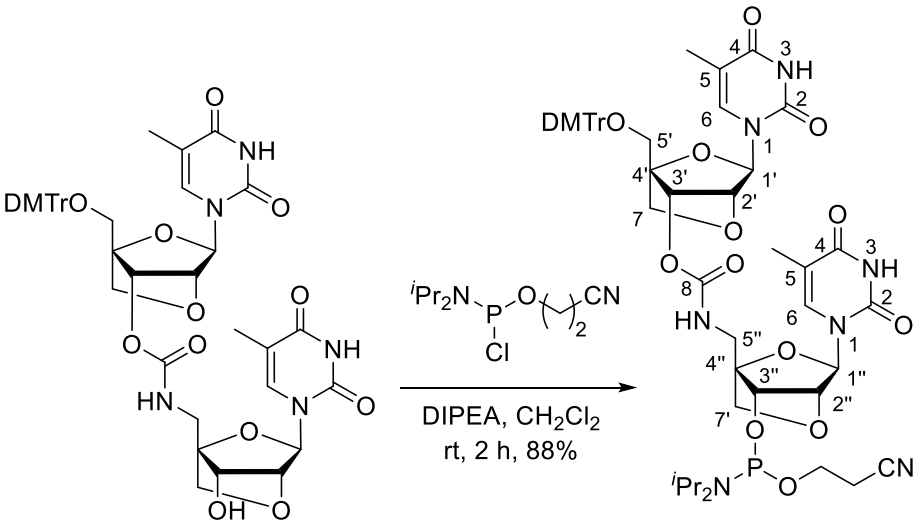



**13** **17**
$C_{52}H_{62}N_7O_{14}P$
1040.08

Carbamate dimer **13** (303 mg, 0.360 mmol) was dissolved in degassed, anhydrous THF (5.00 mL). To this was added degassed DIPEA (199 µL, 1.14 mmol) and chloro(diisopropylamino)-*β*-cyanoethoxyphosphine (128 µL, 0.571 mmol) and stirred at rt under argon for 1 h. The solvents were removed under reduced pressure and the residue was re-dissolved in $CH_2Cl_2$ (5.00 mL). This was washed with sat. aq KCl (10.0 mL), dried over $Na_2SO_4$ and concentrated under reduced pressure to form a glassy white solid. The solid was then re-dissolved in dry $CH_2Cl_2$ (1.00 mL) and stirred vigorously. Upon addition of dry degassed hexane (30.0 mL), the product precipitated as a white solid. The precipitation was repeated five times, and the solid was purified with silica column chromatography (EtOAc:Hexane, 85:14 +1% $Et_3N$) using degassed dry solvents to give phosphoramidite **17** (136 mg, 0.130 mmol, 36%) as a white solid.

$R_f$ 0.73 (EtOAc:MeOH:$Et_3N$, 95:4:1); $^{31}P$ NMR (162 MHz, DMSO-$d_6$) 146.9, 146.9; HRMS–ESI (*m/z*): $[M + H]^+$ calcd for $C_{52}H_{63}N_7O_{14}P$, 1040.4174, found 1040.4165.

5′-*O*-(Dimethoxytrityl)-LNA-CBM$_1$-LNA-thymidine phosphoramidite dimer (LNA-CBM$_1$-LNA) (**18**) (*Synthesised by Sven Epple)*

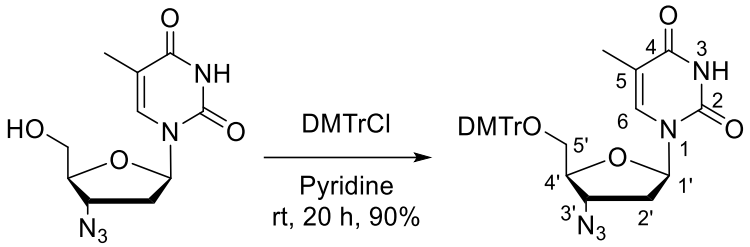



**14**

**18**
$C_{53}H_{62}N_7O_{15}P$
1068.09

To a solution of dimer **14** (334 mg, 0.380 mmol) in dry degassed $CH_2Cl_2$ (3.50 mL) was added degassed DIPEA (165 µL, 950 µmol). Chloro(diisopropylamino)-*β*-cyanoethoxyphosphine (120 µL, 540 µmol) was added dropwise over 10 min and the reaction was stirred at rt for 1 h. After 1 h, a second portion of chloro(diisopropylamino)-*β*-cyanoethoxyphosphine (100 µL, 450 µmol) was added and stirred for another 1 h. The reaction mixture was diluted with dry degassed $CH_2Cl_2$ (10.0 mL) and washed with sat. aq KCl (10.0 mL) under an argon atmosphere. The organic layer was dried over $Na_2SO_4$ and the solvents were evaporated under reduced pressure. The solid was re-dissolved in dry degassed $CH_2Cl_2$ (2.50 mL) and precipitated by addition of hexane (150 mL). The supernatant was decanted and the solid dried under reduced pressure. The precipitation was repeated followed by silica column chromatography (Hexane:EtOAc, 3:20 + 1% pyridine) under argon atmosphere giving phosphoramidite **18** (356 mg, 0.330 mmol, 88%) as a white solid. $R_f$ 0.57 (Hexane:EtOAc:pyridine, 9:90:1); $^{31}P$ NMR (162 MHz, DMSO-$d_6$) δ 148.1, 147.5; HRMS–ESI (*m/z*): $[M + H]^+$ calcd for $C_{53}H_{63}N_7O_{15}P$, 1068.4114, found, 1068.4109.

5′-O-(4′′-Dimethoxytriphenylmethyl)-3′-azido-3′-deoxythymidine (**20**) [112]

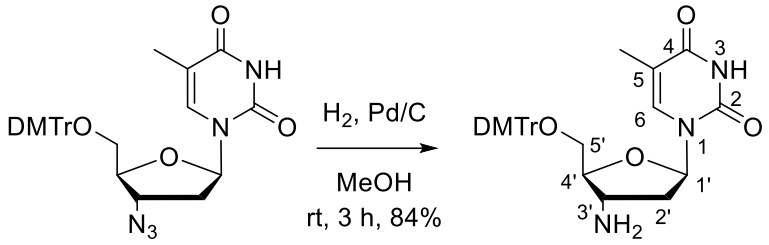



**19** **20**

$C_{31}H_{31}N_5O_6$
569.62

To a solution of AZT **19** (500 mg, 1.87 mmol) in dry pyridine (10.0 mL) was added DMTrCl (820 mg, 2.43 mmol) and stirred at rt for 20 h. The reaction was then concentrated under reduced pressure and resuspended in $CH_2Cl_2$ (50.0 mL). The organic layer was washed with $H_2O$ (3x 50.0 mL), dried over $Na_2SO_4$ and purified by silica column chromatography (EtOAc: Petroleum ether, 0:100 to 40:60) to produce compound **20** as a yellow solid (960 mg, 1.69 mmol, 90%); $R_f$ 0.30 (EtOAc: petroleum ether, 40:60); IR ($cm^{-1}$): 2931 (m, CH), 2835 (m, CH), 2101 (s, $N_3$), 1685 (CO); $^1$H NMR (400 MHz, DMSO-$d_6$) δ 11.37 (s, 1H, H-3), 7.53 (d, *J* = 1.2 Hz, 1H, H-6), 7.43 – 7.37 (m, 2H, H-Ar), 7.34 – 7.24 (m, 7H, H-Ar), 6.93 – 6.87 (m, 4H, H-Ar), 6.14 (dd, *J* = 7.0, 5.7 Hz, 1H, H-1′), 4.57 (dt, *J* = 7.6, 5.8 Hz, 1H, H-3′), 3.89 (dt, *J* = 5.8, 3.8 Hz, 1H, H-4′), 3.74 (s, 6H, OMe), 3.25 (d, *J* = 3.8 Hz, 2H, H-5′), 2.50 (m, 1H, H-2′), 2.36 (m, 1H, H-2′), 1.55 (d, *J* = 1.2 Hz, 3H, $CH_3$-*thymidine*); $^{13}$C NMR (101 MHz, DMSO-$d_6$) δ 164.1 (C-4), 158.6 (C-Ar), 150.8 (C-2), 145.1 (C-Ar), 136.3 (C-6), 135.7 (C-Ar), 135.7 (C-Ar), 130.2 (C-Ar), 128.4 (C-Ar), 128.1 (C-Ar), 127.3 (C-Ar), 113.7 (C-Ar), 110.2 (C-5), 86.4 (C-$Ar_2$Ph), 83.8 (C-1′), 82.5 (C-4′), 63.5 (C-5′), 60.3 (C-3′), 55.5 (C-OMe), 36.5 (C-2′), 12.3 (C-$CH_3$-*thymidine*); LRMS (ESI-TOF) m/z: $[M-H]^-$ 568 (50%), $[M+Cl]^-$ 604 (100%)

5′-O-(4′′-Dimethoxytriphenylmethyl)-3′-amino-3′-deoxythymidine (**21**)[112]

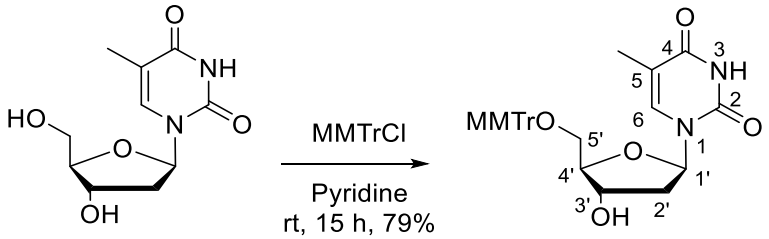



**20** **21**

$C_{31}H_{33}N_3O_6$
543.62

To a solution 5′-O-(4′′-dimethoxytriphenylmethyl)-3′-azido-3′-deoxythymidine **20** (940 mg, 1.65 mmol) in MeOH (40.0 mL) was added 10% Pd/C (88.0 mg, 82.5 µmol) and flushed with an argon atmosphere. The solution was then degassed by bubbling with hydrogen for 10 mins and stirred under 1 atm hydrogen atmosphere at rt for 3 h. The reaction was filtered through celite and the filtrate was concentrated under vacuum and purified by silica column chromatography ($CH_2Cl_2$: MeOH, 100:0 to 85:15) to produce compound **21** a white foam (750 mg, 1.38 mmol, 84%). $R_f$ 0.47 ($CH_2Cl_2$: MeOH, 1:9); $^1$H NMR (400 MHz, DMSO-$d_6$) δ 7.52 (d, *J* = 1.3 Hz, 1H, H-6), 7.44 – 7.37 (m, 2H, H-Ar), 7.35 – 7.21 (m, 7H, H-Ar), 6.89 (dd, *J* = 9.0, 1.9 Hz, 4H, H-Ar), 6.15 (dd, *J* = 6.9, 5.3 Hz, 1H, H-1′), 3.74 (m, 6H, H-OMe), 3.70 (m, 1H, H-4′), 3.51 (q, *J* = 6.6 Hz, 1H, H-3′), 3.34 (s, 2H, H-N$\underline{H}_2$ ), 3.22 (dd, *J* = 3.7, 1.6 Hz, 2H, H-5′), 2.21 (m, 1H, H-2′), 2.10 – 1.97 (m, 1H, H-2′), 1.48 (d, *J* = 1.3 Hz, 3H, $CH_3$-*thymidine*); $^{13}$C NMR (101 MHz, DMSO-$d_6$) δ 164.2 (C-4), 158.6 (C-Ar), 150.8 (C-2), 145.3 (C-Ar), 136.2 (C-Ar), 136.0 (C-6), 135.9 (C-Ar), 130.2 (C-Ar), 128.3 (C-Ar), 128.2 (C-Ar), 127.2 (C-Ar), 113.7 (C-Ar), 109.7 (C-5), 86.2 – 86.1 (C-4′, C-$\underline{C}$Ar$_2$Ph), 84.1 (C-1′), 64.1 (C-5′), 55.5 (C-OMe), 52.1 (C-3′), 41.2 (C-2′), 12.3 (C-$CH_3$-*thymidine*); LRMS (ESI-TOF) m/z: $[M+Cl]^-$ 578 (100%).

5′-O-(4′′-Methoxytriphenylmethyl)thymidine (**22**)[113]



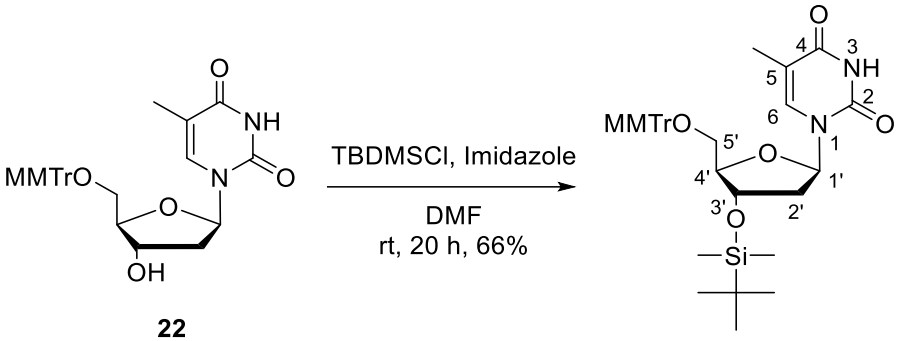

**1**

**22**

$C_{30}H_{30}N_2O_6$
514.58

Thymidine **1** (15.0 g, 62.0 mmol) was co-evaporated with dry pyridine (3x 100 mL) and then dissolved in dry pyridine (100 mL). To this was added MMTrCl (28.9 g, 93.0 mmol) in dry pyridine (200 mL) and stirred at rt. After 15 h the reaction was concentrated under vacuum and re-dissolved in $CH_2Cl_2$ (300 mL). This was extracted with $H_2O$ (3x 300 mL) and dried over $Na_2SO_4$. The organic layer was concentrated under vacuum and purified by silica chromatography (1:99 to 1:19, MeOH: $CH_2Cl_2$) to produce compound **22** as a white solid (25.2 g, 49.0 mmol, 79%).[228] $R_f$ 0.71 (19:1 $CH_2Cl_2$: MeOH); IR ($cm^{-1}$): 3057 (m, CH), 1684 (CO), 1438 (m, C=C); $^1$H NMR (400 MHz, $CDCl_3$) δ 9.56 (s, 1H, NH-3), 7.61 (d, *J* = 1.4 Hz, 1H, H-6), 7.46 – 7.38 (m, 4H, Ar-PhOMe), 7.35 – 7.20 (m, 8H, Ar-Ph), 6.89 – 6.82 (m, 2H, Ar-Ph), 6.46 (dd, *J* = 7.9, 5.8 Hz, 1H, H-1′), 4.59 (dd, *J* = 6.2, 2.8 Hz, 1H, H-3′), 4.10 (q, *J* = 2.8 Hz, 1H, H-4′), 3.79 (s, 3H, O-$CH_3$), 3.46 (dd, *J* = 10.5, 3.1 Hz, 1H, H-5′), 3.42 – 3.34 (m, 1H, H-5′), 2.46 (ddd, *J* = 13.7, 5.8, 2.8 Hz, 1H, H-2′), 2.33 (ddd, *J* = 13.7, 7.9, 6.2 Hz, 1H, H-2′), 1.46 (s, 3H, $CH_3$-*thymidine*); $^{13}$C NMR (101 MHz, $CDCl_3$) δ 164.0 (C-2), 158.7 (C-ArOMe), 150.6 (C-6), 143.9 (C-ArOMe), 143.8 (C-ArOMe), 135.7 (C-4), 134.9 (C-ArOMe), 130.4 (C-Ar), 128.4 (C-Ar), 128.0 (C-Ar), 127.2 (C-Ar), 113.3 (C-ArOMe), 111.3 (C-5), 87.1 (C-Ar), 86.3 (C-1′), 84.8 (C-4′), 72.4 (C-3′), 63.7 (C-5′), 55.2 (C-OMe), 40.9 (C-2′), 11.9 (C-$CH_3$-*thymidine*); LRMS (ESI-TOF) m/z: $[M+Cl]^-$ 549 (100%).

5′-O-(4′′-Methoxytriphenylmethyl)-3′-*O*-(*tert*-butyldimethylsilyl)thymidine (**23**)[114]



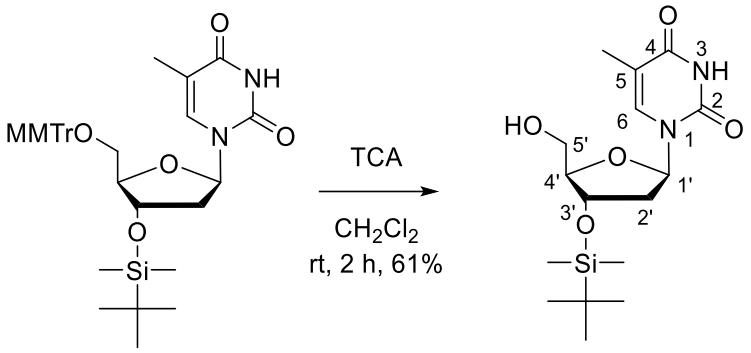

**22**

**23**

$C_{36}H_{44}N_2O_6Si$
628.84

To a solution of compound **22** (3.00 g, 5.83 mmol) in DMF (40.0 mL) was added imidazole (830 mg, 12.2 mmol) and TBDMSCl (1.05 g, 7.00 mmol) and stirred for 20 h at 0 °C. After 20 h a further portion of imidazole (79.0 mg, 1.17 mmol) and TBDMSCl (176 mg, 1.17 mmol) was added and stirred for 1 h at rt. The solvent was removed under vacuum and the crude was purified by silica column chromatography ($CH_2Cl_2$: MeOH, 100:0 to 95:5) to produce compound **23** as a white solid (2.41 g, 2.80 mmol, 66%). $R_f$ 0.50 ($CH_2Cl_2$: MeOH, 9:1); $^1$H NMR (400 MHz, DMSO-$d_6$) δ 11.34 (s, 1H, H-3), 7.53 (d, *J* = 1.2 Hz, 1H, H-6), 7.44 – 7.20 (m, 12H, H-Ar), 6.93 – 6.86 (m, 2H, H-Ar), 6.16 (t, *J* = 6.6 Hz, 1H, H-1′), 4.47 (dt, *J* = 6.7, 4.5 Hz, 1H, H-3′), 3.82 (td, *J* = 4.5, 4.5, 3.0 Hz, 1H, H-4′), 3.73 (s, 3H, OMe), 3.25 (dd, *J* = 10.7, 3.0 Hz, 1H, H-5′), 3.16 (dd, *J* = 10.7, 4.5 Hz, 1H, H-5′), 2.36 – 2.26 (m, 1H, H-2′), 2.14 (m, 1H, H-2′), 1.52 (d, *J* = 1.2 Hz, 3H, $CH_3$-*thymidine*), 0.78 (s, 9H, $^t$Bu-Si), 0.00 (s, 3H, Si-$CH_3$), -0.06 (s, 3H, Si-$CH_3$); $^{13}$C NMR (101 MHz, DMSO-$d_6$) δ 164.2 (C-4), 158.8 (C-Ar), 150.8 (C-2), 144.5 (C-Ar), 144.4 (C-Ar), 136.2 (C-Ar), 135.1 (C-6), 130.5 (C-Ar), 128.4 (C-Ar), 127.5 (C-Ar), 113.7 (C-Ar), 110.0 (C-5), 86.7 (C-$\underline{C}Ph_2Ar$), 85.6 (C-4′), 84.2 (C-1′), 72.0 (C-3′), 63.5 (C-5′), 55.5 (C-O$\underline{C}H_3$), 40.1 (C-2′), 26.1 (C-SiC($\underline{C}H_3)_3$), 18.08 (C-Si$\underline{C}(CH_3)_3$), 12.35 (C-$CH_3$-*thymidine*), -4.32(C-Si$\underline{C}H_3$), -4.63 (C-Si$\underline{C}H_3$); HRMS-ESI (*m/z*): $[M+Na]^+$: calcd. for $C_{36}H_{44}N_2O_6NaSi$ 651.2861; found, 651.2860.

3′-*O*-(*tert*-Butyldimethylsilyl)thymidine (**24**)[115]

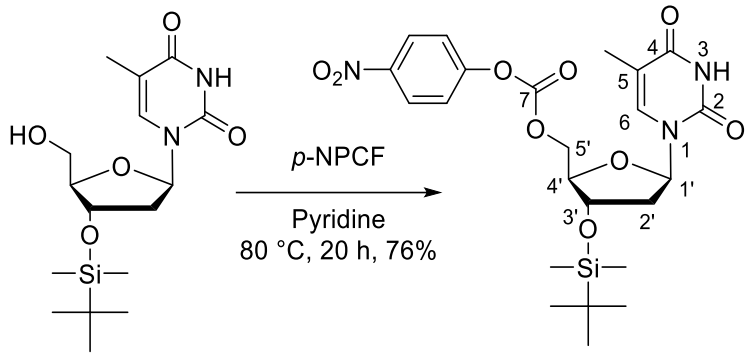



**23**

**24**
$C_{16}H_{28}N_2O_5Si$
356.49

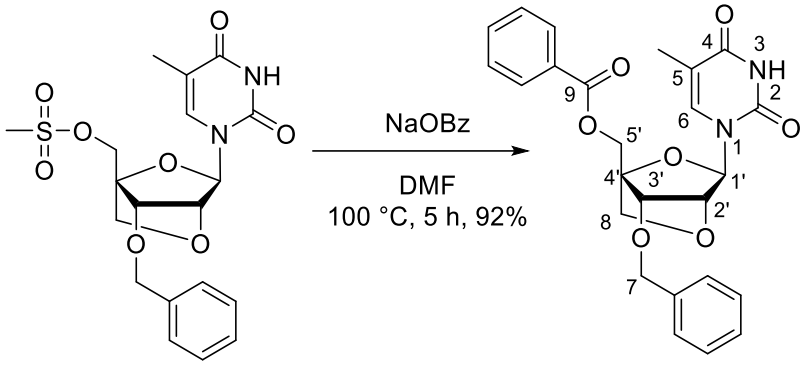
To a solution of compound **23** (2.00 g, 3.18 mmol) in $CH_2Cl_2$ (10.0 mL) was added a solution of TCA:$CH_2Cl_2$ (100 mL, 1:9) and stirred for 2 h at rt. The reaction was concentrated under reduced pressure and purified by silica column chromatography ($CH_2Cl_2$: MeOH, 100:0 to 9:1) to produce compound **24** as a white foam (690 mg, 1.94 mmol, 61%); $R_f$ 0.30 ($CH_2Cl_2$: MeOH, 9:1); $^1$H NMR (400 MHz, DMSO-$d_6$) δ 11.21 (s, 1H, H-3), 7.58 (d, *J* = 1.2 Hz, 1H, H-6), 6.07 (dd, *J* = 7.8, 5.9 Hz, 1H, H-1′), 4.33 (dt, *J* = 5.9, 2.7 Hz, 1H, H-3′), 3.67 (td, *J* = 4.0, 2.7 Hz, 1H, H-4′), 3.48 (qd, *J* = 11.9, 4.0 Hz, 2H, H-5′), 2.10 (m, 1H, H-2′), 2.03 – 1.88 (m, 1H, H-2′), 1.69 (d, *J* = 1.2 Hz, 3H, $CH_3$-*thymidine*), 0.79 (s, 9H, $^t$Bu-Si), 0.00 (s, 6H, Si-$CH_3$); $^{13}$C NMR (101 MHz, DMSO-$d_6$) δ 166.4 (C-4), 153.2 (C-2), 138.7 (C-6), 112.1 (C-5), 90.1 (C-4′), 86.4 (C-1′), 74.9 (C-3′), 63.7 (C-5′), 42.6 (C-2′), 28.4 (C-Si-C($\underline{C}H_3)_3$), 20.4 (C-Si-$\underline{C}(CH_3)_3$), 15.0 (C-$CH_3$-*thymidine*), -2.1 (C-Si$\underline{C}H_3$), -2.16 (C-Si$\underline{C}H_3$); LRMS (ESI-TOF) m/z: $[M+Na]^+$ 379 (100%).

5′-*O*-(4-Nitrophenyl formate)-3′-*O*-(*tert*-butyldimethylsilyl)thymidine (**25**)[32]




To a solution of compound **24** (600 mg, 1.68 mmol) in anhydrous pyridine (20.0 mL) was added 4-Nitrophenyl chloroformate (410 mg, 2.02 mmol) at rt and heated to 80 °C for 20 h. The reaction was concentrated under reduced pressure and purified by silica column chromatography ($CH_2Cl_2$: MeOH, 100:0 to 90:10) to produce compound **25** as a white foam (670 mg, 1.28 mmol, 76%). $R_f$0.66 ($CH_2Cl_2$: MeOH, 19:1); $^1$H NMR (400 MHz, DMSO-$d_6$) δ 11.39 (s, 1H, H-3), 8.41 – 8.32 (m, 2H, H-Ar), 7.63 – 7.57 (m, 2H, H-Ar), 7.56 (d, *J* = 1.3 Hz, 1H, H-6), 6.24 (t, *J* = 7.0 Hz, 1H, H-1′), 4.58 – 4.40 (m, 3H, H-3′, H-5′), 4.06 (m, 1H, H-4′), 2.37 (m, 1H, H-2′), 2.16 (m, 1H, H-2′), 1.82 (d, *J* = 1.3 Hz, 3H, $CH_3$-*thymidine*), 0.92 (d, *J* = 1.4 Hz, 9H, $^t$Bu-Si), 0.14 (t, *J* = 1.4 Hz, 6H, Si-$CH_3$); $^{13}$C NMR (101 MHz, DMSO-$d_6$) δ 164.1 (C-4), 155.6 (C-7), 152.3 (C-Ar), 150.0 (C-2), 145.7 (C-Ar), 136.7 (C-6), 125.9 (C-Ar), 123.0 (C-Ar), 110.3 (C-5), 84.6 (C-1′), 83.6 (C-4′), 72.2 (C-3′), 68.7 (C-5′), 38.5 (C-2′), 26.1 (C-Si-C($\underline{C}H_3$)), 18.1 (C-Si-$\underline{C}$($CH_3$)), 12.6 (C-$CH_3$-*thymidine*), -4.3 (C-Si$\underline{C}H_3$), -4.53 (C-Si$\underline{C}H_3$); HRMS-ESI (*m/z*): $[M+H]^+$ calculated for $C_{23}H_{31}N_3O_9Si$, 522.1902, found 522.1902.

5′-*O*-(Benzoyl)-3′-*O*-(benzyl)-LNA-thymidine (**26**)[110]

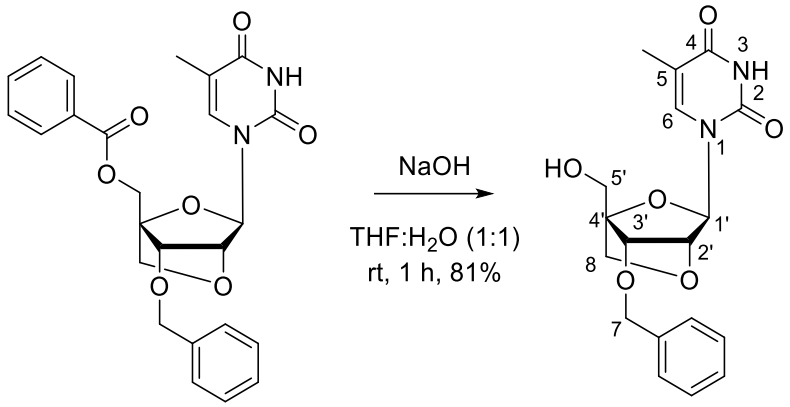



**8** **26**

$C_{25}H_{24}N_2O_7$
464.47

To a solution of 5′-*O*-(mesyl)-3′-*O*-(benzyl)-LNA-thymidine **8** (2.00 g, 4.60 mmol) in DMF (100 mL) was added sodium benzoate (1.33 g, 9.20 mmol) and stirred at 100 °C for 5 h. The reaction was concentrated under reduced pressure and dissolved in EtOAc (100 mL). This was then washed with $H_2O$ (3x 100 mL) and dried over $Na_2SO_4$ before concentrating under vacuum to produce compound **26** a white solid (1.96 g, 4.20 mmol, 92%). $R_f$ 0.74 (EtOAc), $^1$H NMR (400 MHz, $CDCl_3$) δ 9.36 (s, 1H,H-3), 7.87 (dd, *J* = 8.4, 1.4 Hz, 2H, H-Ar), 7.59 – 7.50 (m, 1H, H-Ar), 7.43 – 7.34 (m, 2H, H-Ar), 7.24 – 7.17 (m, 7H, H-Ar), 7.16 (q, *J* = 1.2 Hz, 1H, H-6), 5.58 (br s, 1H, H-1′), 4.77 (d, *J* = 12.7 Hz, 1H, H-5′), 4.66 (d, *J* = 11.8 Hz, 1H, H-7), 4.62 (s, 1H, H-2′), 4.50 (d, *J* = 12.7 Hz, 1H, H-5′), 4.46 (d, *J* = 11.8 Hz, 1H, H-7), 4.11 (d, *J* = 7.8 Hz, 1H, H-8), 3.90 (d, *J* = 7.8 Hz, 1H, H-8), 3.85 (s, 1H, H-3′), 1.51 (d, *J* = 1.2, 3H, $CH_3$-*thymidine*); $^{13}$C NMR (101 MHz, $CDCl_3$) δ 165.7 (C-9), 163.9 (C-4), 149.8 (C-2), 136.5 (C-Ar), 133.8 (C-6), 133.7 (C-Ar), 129.6 (C-Ar), 128.7 (C-Ar), 128.6 (C-Ar), 128.3 (C-Ar), 127.9 (C-Ar), 110.5 (C-5), 87.6 (C-1′), 86.3 (C-4′), 76.6 (C-2′), 75.8 (C-3′), 72.2 (C-7), 72.1 (C-8), 59.0 (C-5′), 12.3 (C-$CH_3$-*thymidine*); LRMS (ESI-TOF) m/z: $[M+Na]^-$ 487 (72%), $[2M+Na]^+$ 951 (100%).

3′-*O*-(benzyl)-LNA-thymidine (**27**)[110]

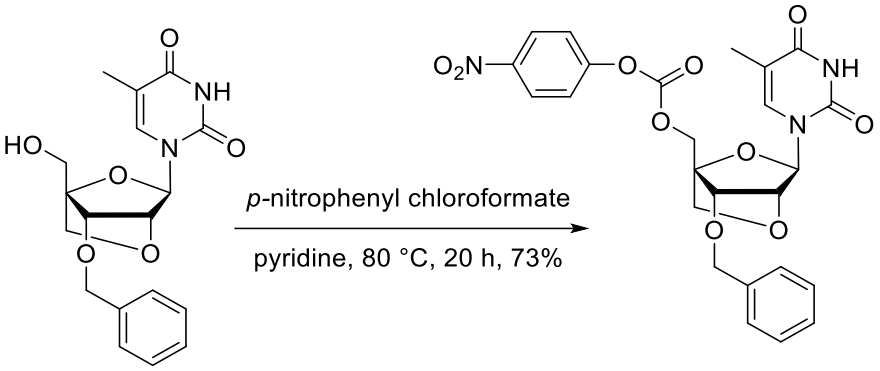



**26** **27**

$C_{18}H_{20}N_2O_6$
360.37

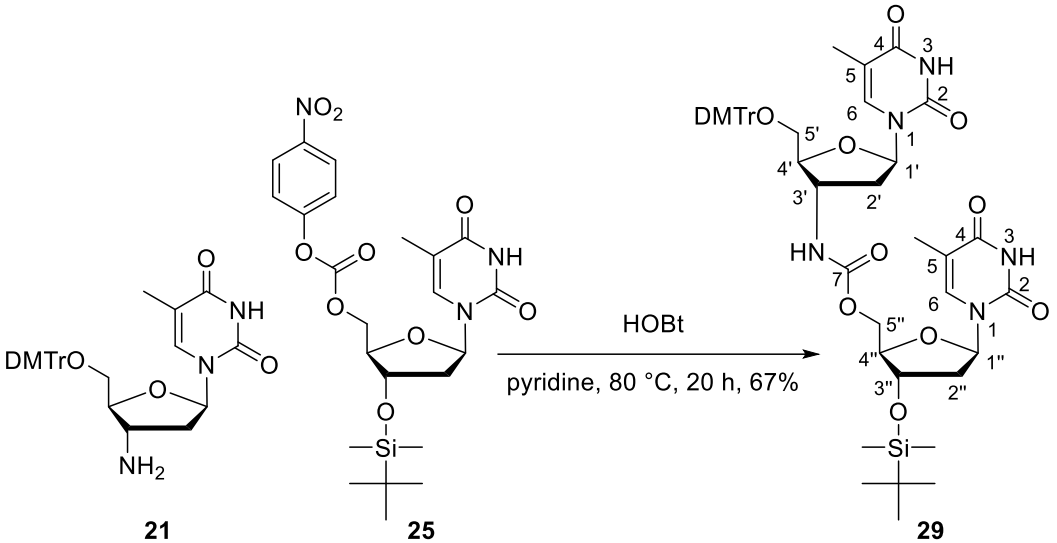
To a solution of compound **26** (1.96 g, 4.20 mmol) in THF:$H_2O$ (1:1, 30.0 mL) was added 1M NaOH (8 mL, 8.00 mol) and stirred at rt for 1 h. The reaction was concentrated under reduced pressure and redissolved in $CH_2Cl_2$ (100 mL), washed with $NaHCO_3$ (3x 100 mL) and dried over $Na_2SO_4$. The crude was then purified by silica column chromatography ($CH_2Cl_2$:MeOH, 100:0 to 49:1) to produce compound **27** as a white solid (1.24 g, 3.40 mmol, 81%). $R_f$ ($CH_2Cl_2$:MeOH, 33:1) 0.24; $^1$H NMR (400 MHz, $CDCl_3$) δ 9.16 (s, 1H, H-3), 7.39 (q, *J* = 1.2 Hz, 1H, H-6), 7.32 – 7.15 (m, 5H, H-Ar), 5.59 (d, *J* = 0.7 Hz, 1H, H-1′), 4.60 (d, J = 11.5 Hz, 1H, H-7), 4.50 (d, J = 11.5 Hz, 1H, H-7), 4.47 (s, 1H, H-2′), 3.98 (d, J = 7.9 Hz, 1H, H-8), 3.91 (2xd, J = 10.5, 2H, H-5′, H-5′′), 3.88 (s, 1H, H-3′),3.76 (d, *J* = 7.9 Hz, 1H, H-8), 2.31 (t, *J* = 5.6 Hz, 1H, H-5′OH), 1.82 (d, *J* = 1.2 Hz, 3H, $CH_3$-*thymidine*); $^{13}$C NMR (101 MHz, $CDCl_3$) δ 163.9 (C-4), 149.8 (C-2), 136.9 (C-Ar), 134.7 (C-6), 128.6 (C-Ar), 128.2 (C-Ar), 127.8 (C-Ar), 127.8 (C-Ar), 110.3 (C-5) , 88.2 (C-4′), 87.3 (C-1′), 76.9 (C-2′), 75.8 (C-3′), 72.3 (C-7), 72.0 (C-8), 57.6 (C-5′), 12.7 (C-$CH_3$-*thymidine*); LRMS (ESI-TOF) m/z: $[M+Na]^+$ 383 (37%), $[2M+Na]^+$ 743 (100%).

3′-*O*-(Benzyl)-5′-*O*-(4-nitrophenyl formate)-LNA-thymidine (**28**)




To a solution of alcohol[110] **27** (1.24 g, 3.40 mmol) in anhydrous pyridine (20.0 mL) was added *p*-nitrophenyl chloroformate (830 mg, 4.10 mmol) at rt and then heated to 80 °C. After 20 h, a second portion of *p*-nitrophenyl chloroformate (89.0 mg, 0.340 mmol) was added and stirred at 80 °C for another 3 h. The reaction was concentrated under reduced pressure and purified by silica column chromatography ($CH_2Cl_2$:EtOAc, 50:50) to give carbonate **28** (1.30 g, 2.40 mmol, crude 73%) as a white foam. $R_f$ 0.68 ($CH_2Cl_2$:MeOH, 19:1); HRMS–ESI (*m/z*): $[M + Na]^+$ calcd for $C_{25}H_{23}N_3O_{10}Na$, 548.1276, found 548.1266.

5′-*O*-(Dimethoxytrityl)-3′-*O*-(*tert*-butyldimethylsilyl) thymidine dimer (DNA-CBM$_2$-DNA) (**29**)

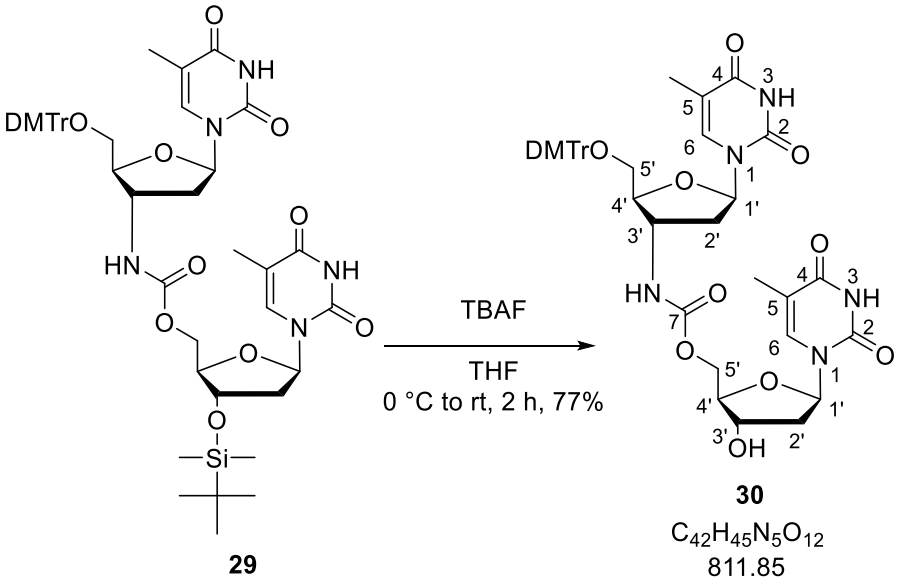



**21** **25** **29**

$C_{48}H_{59}N_5O_{12}Si$
926.11

To a solution of carbonate **25** (570 mg, 1.09 mmol) and HOBt (140 mg, 1.18 mmol) in dry pyridine (35.0 mL) was added amine **21** (580 mg, 1.07 mmol) and stirred at 80 °C for 20 h. The reaction was then concentrated under reduced pressure and the residue was diluted with EtOAc (100 mL). The solution was washed with $H_2O$ (2 × 100 mL), brine (100mL) and dried over $Na_2SO_4$. Purification by silica column chromatography (EtOAc:Petroleum ether, 1:9 to 10:0) gave dimer **29** (660 mg, 0.710 mmol, 67%) as a white foam. $R_f$ 0.46 (EtOAc); $^1$H NMR (400 MHz, $CDCl_3$) δ 9.30 (2 × s, 2H, H-3), 7.54 (d, *J* = 1.4 Hz, 1H, H-6), 7.37 – 7.30 (m, 2H, H-Ar), 7.26 – 7.07 (m, 8H, H-Ar, H-6), 6.80 – 6.72 (m, 4H, H-Ar), 6.41 (d, *J* = 7.5 Hz, 1H, NHCBM), 6.33 (t, *J* = 6.8 Hz, 1H, H-1′ or H-1′′), 6.00 (t, *J* = 6.5 Hz, 1H, H-1′ or H-1′′), 4.45 (br s, 1H, H-3′′), 4.31 (d, *J* = 5.8 Hz, 1H, H-3′), 4.19 (m, 2H, H-5′′), 3.94 (m, 2H, H-4′, H-4′′), 3.71 (d, *J* = 5.9 Hz, 6H, $OCH_3$), 3.38 (dd, *J* = 10.7, 2.7 Hz, 1H, H-5′), 3.32 (dd, *J* = 10.5, 2.7 Hz, 1H, H-5′), 2.33 – 2.16 (m, 4H, H-2′, H-2′′), 1.78 (app s, 3H, $CH_3$), 1.33 (app s, 3H, $CH_3$), 0.81 (s, 9H, $^t$Bu-Si), 0.00 (2 × s, 6H, Si-$CH_3$); $^{13}$C NMR (101 MHz, $CDCl_3$) δ 164.1 (C-4), 158.7-158.6 (C-Ar, C-7), 150.8 (C-2), 147.4 (C-Ar), 144.3 (C-Ar), 136.7 (C-Ar), 135.9 (C-6), 130.1 (C-Ar), 129.1 (C-Ar), 128.2-127.0 (7 × C-Ar, C-6), 113.1 (C-Ar), 110.9 (C-5), 87.0 (C-1′ or C-1′′), 84.5 ($C_{quat}$DMTr, C-4′, C-4′′, [C-1′ or C-1′′]), 71.6 (C-3′), 64.2 (C-5′′), 63.7 (C-5′), 55.3 ($OCH_3$), 52.4 (C-3′′) 40.2 (C-2′ or C-2′′), 38.4 (C-2′ or C-2′′), 25.7 (C-Si-C($\underline{C}H_3$), 17.9 (C-Si-$\underline{C}(CH_3)$), 12.5 ($CH_3$), 11.7 ($CH_3$), –4.7 (C-Si$\underline{C}H_3$), –4.9 (C-Si$\underline{C}H_3$); HRMS–ESI (*m/z*): $[M + Na]^+$ calcd for $C_{48}H_{59}N_5O_{12}SiNa$, 948.3822 found 948.3830.

5′-O-(4′′-Dimethoxytriphenylmethyl)thymidine carbamate dimer (**30**) (DNA-CBM$_2$-DNA)




**29**

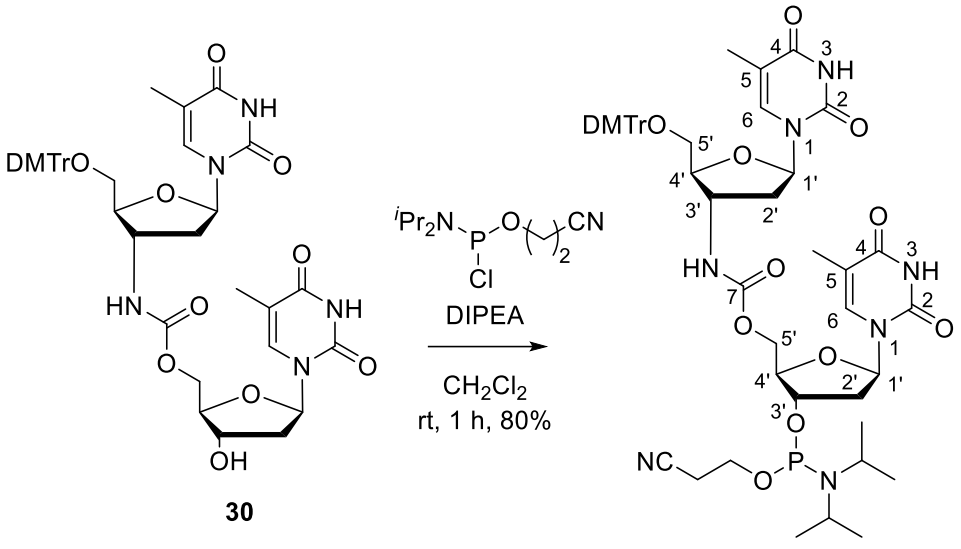
**30**

$C_{42}H_{45}N_5O_{12}$

811.85

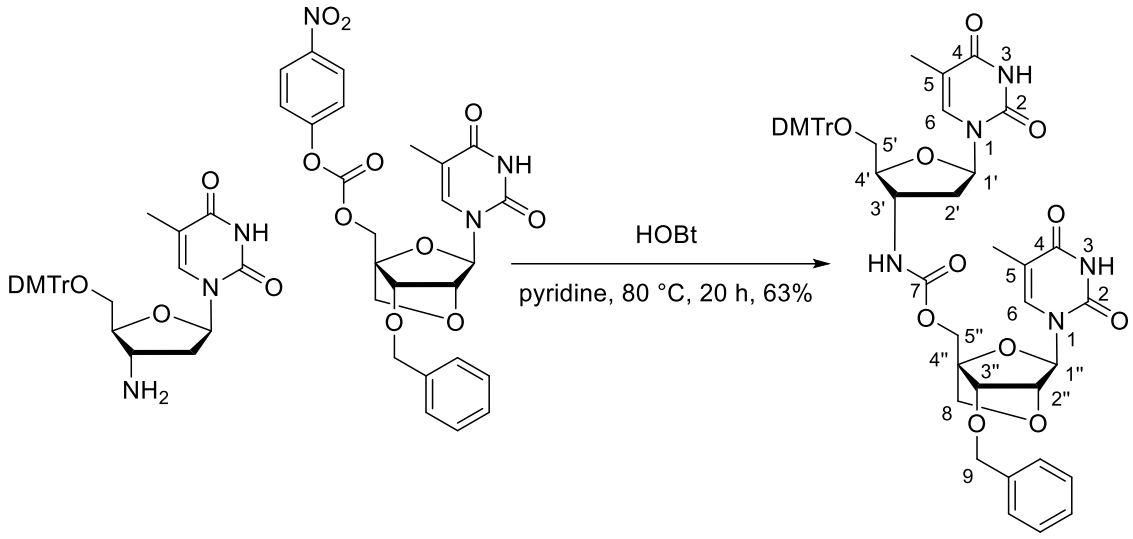
To a solution of dimer **29** (620 mg, 0.669 mmol) in THF (30.0 mL) was added TBAF (1.34 mL, 1.338 mmol) at 0 °C over 5 mins. After 10 mins the reaction was warmed to rt and stirred for 2 h. This was the concentrated under reduced pressure and diluted in $CH_2Cl_2$ (100 mL), washed with $H_2O$: Brine (19:1, 3x100 mL), $Na_2CO_3$ (2x100 mL) and dried over $Na_2SO_4$. This was then purified by silica column chromatography solid ($CH_2Cl_2$: MeOH, 100:0 to 4:1) to produce dimer **30** as a while solid (420 mg, 0.712 mmol, 0.517 mmol, 77%). $R_f$ 0.34 ($CH_2Cl_2$: MeOH, 9:1); $^1$H NMR (500 MHz, $CDCl_3$) δ 10.18 (s, 2H, H-3), 7.62 (s, 1H, H-6), 7.40 (m, 2H, H-Ar), 7.32 – 7.20 (m, 8H, H-Ar), 6.97 (s, 1H, H-6), 6.83 (d, *J* = 8.6 Hz, 4H, H-Ar), 6.37 (s, 1H, H-1′), 6.22 (s, 1H, H-1′), 4.56 – 3.98 (m, 6H, H-3′, H-4, H-5′), 3.76 (s, 6H, H-OMe), 3.41 (2x br s, 2H, H-5′), 2.39 (s, 2H, H-2′), 2.27 (s, 2H, H-2′), 1.81 (s, 3H, $CH_3$-*thymidine*), 1.38 (s, 3H, $CH_3$-*thymidine*); $^{13}$C NMR (126 MHz, $CDCl_3$) δ 164.4 (C-4), 164.2 (C-4), 158.7 (C-Ar), 158.7 (C-Ar), 156.0 (C-7), 151.2 (C-2), 150.9 (C-2), 144.3 (C-Ar), 136.2 (C-6), 135.6 (C-6), 135.4 (C-Ar), 135.3 (C-Ar), 130.1 (C-Ar), 128.2 (C-Ar), 128.0 (C-Ar), 127.1 (C-Ar), 113.3 (C-Ar), 111.7 (C-5), 111.1 (C-5), 85.7 (C-1′) , 84.5 (C-1′, C-$CAr_2$Ph), 71.5 (C-3′ or C-4′), 64.6 (C-5′), 64.0 (C-5′), 55.3 (C-OMe), 52.7 (C-3' or C-4′), 39.7 (C-2′), 38.2 (C-2′), 12.5 (C-$CH_3$-*thymidine*), 11.7 (C-$CH_3$-*thymidine*); HRMS-ESI (*m/z*): $[M+Na]^+$ calcd. for $C_{42}H_{45}N_5O_{12}Na$ 834.2957, found 834.2962.

5′-O-(4′′-Dimethoxytriphenylmethyl)thymidine carbamate phosphoramidite dimer (**31**) (DNA-$CBM_2$-DNA)




Dimer **30** (200 mg, 0.246 mmol) was co-evaporated with degassed $CH_2Cl_2$ (3x 2.00 mL) and dissolved in dry degassed $CH_2Cl_2$ (5.00 mL). To this was added degassed DIPEA (107 µL, 0.616 mmol) followed by 2-Cyanoethyl *N,N*-diisopropylchlorophosphoramidite (65.0 µL, 0.295 mmol) and stirred under argon for 30 mins. After 30 mins a further portion of DIPEA (21 µL, 0.123 mmol) and 2-Cyanoethyl *N,N*-diisopropylchlorophosphoramidite (27 µL, 0.123 mmol) were added and stirred for another 30 mins. The reaction was diluted in degassed $CH_2Cl_2$ (5.00 mL), washed with sat KCl solution (10 mL) and dried over $Na_2SO_4$. The organic layer was concentrated and purified by silica column chromatography (EtOAc + 0.1% DIPEA) under argon using degassed solvent to produce dimer **31** as a white solid (200 mg, 0.198 mmol, 80%). $R_f$ (EtOAc) 0.61; $^{31}P$ NMR 149.1, 148.6; LRMS-ESI (*m/z*): $[DMTr+H]^+$ 303.20 (83%), $[2xDMTr+H]^+$ 607.79 (13%), $[M+Na]^+$ 1034.64 (3%).

3′′-*O*-(Benzyl)-5′-*O*-(dimethoxytrityl)-DNA-CBM$_2$-LNA-thymidine dimer (DNA-CBM$_2$-LNA) (**32**)

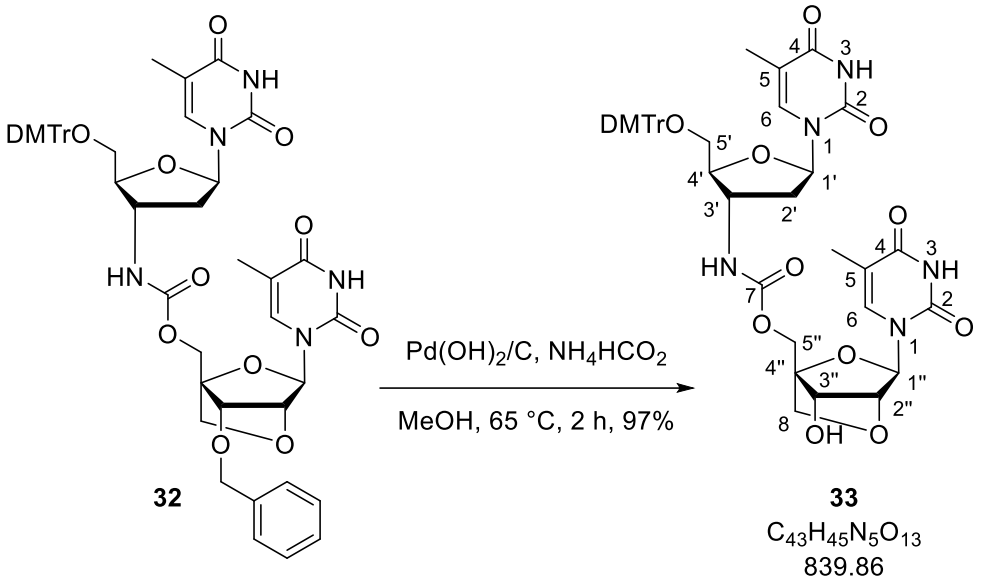



**21** **28** **32**

$C_{50}H_{51}N_5O_{13}$
929.98

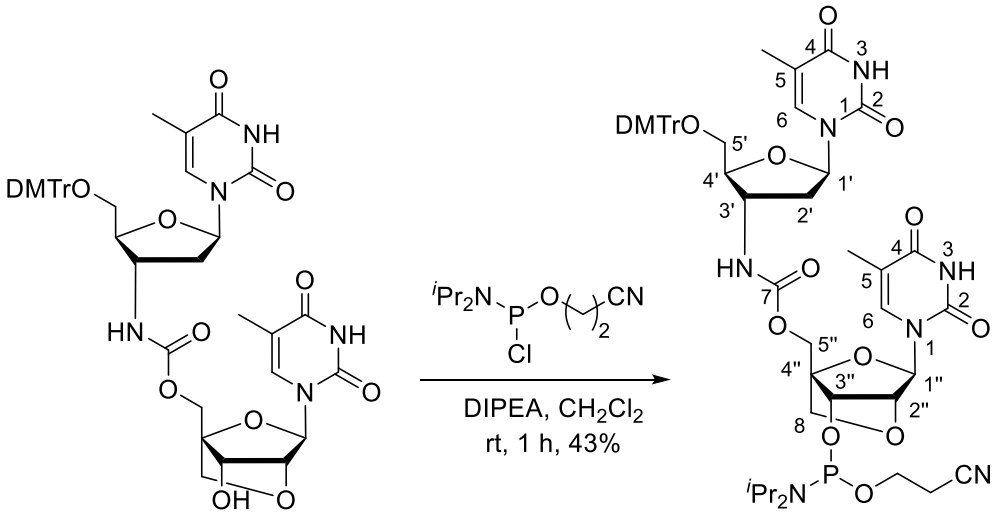
To a solution of carbonate **21** (0.960 g, 1.90 mmol) and HOBt (25.0 mg, 0.190 mmol) in dry pyridine (60.0 mL) was added amine **28** (1.03 g, 1.90 mmol) and stirred at 80 °C for 20 h. The reaction was concentrated under reduced pressure and purified by silica column chromatography ($CH_2Cl_2$:MeOH, 99:0 to 19:1 + 1% pyridine) to give dimer **32** (1.12 g, 1.20 mmol, 63%) as a white solid. $R_f$ 0.68 ($CH_2Cl_2$:MeOH, 19:1); $^1$H NMR (400 MHz, DMSO-$d_6$) δ 11.39 (2 × s, 2H, H-3), 7.86 (d, *J* = 8.0 Hz, 1H, NH CBM), 7.61 – 7.56 (m, 1H, H-6), 7.43 – 7.15 (m, 15H, H-6, H-Ar), 6.91 – 6.82 (m, 4H, H-Ar), 6.17 (t, *J* = 6.3 Hz, 1H, H-1′), 5.53 (s, 1H, H-1′′), 4.68 –4.46 (m, 5H, H-2′′, H-8, H-9), 4.35 (app p, *J* = 7.4 Hz, 1H, H-3′), 3.97 (s, 1H, H-3′′), 3.93–3.90 (m, 2H, H-4′, H-5′′), 3.75 (d, 1H, H-5′′), 3.71 (2 × s, 6H, $OCH_3$), 3.27-3.17 (m, 2H, H-5′), 2.43–2.34 (m, 1H, H-2′), 2.26–2.17 (m, 1H, H-2′), 1.74 (d, *J* = 1.1 Hz, 3H, $CH_3$), 1.49 (d, *J* = 1.2 Hz, 3H, $CH_3$); $^{13}$C NMR (101 MHz, DMSO-$d_6$) δ 164.3 (C-4), 164.2 (C-4), 158.6 (C-Ar), 155.7 (C-7), 150.8 (C-2), 150.4 (C-2), 145.2 (C-Ar), 138.2 (C-Ar), 136.3 (C-6), 135.8 (C-Ar), 135.7 (C-Ar), 134.8 (C-6), 130.2 (C-Ar), 128.7 (C-Ar), 128.3 (C-Ar), 128.1 (C-Ar), 127.9 (C-Ar), 127.2 (C-Ar), 113.6 (C-Ar), 110.0 (C-5), 109.2 (C-5), 87.1 (C-1′′), 86.2–86.0 (C-4′′, $C_{quat}$DMTr), 84.0 (C-1′), 83.1 (C-4′), 77.2–77.1 (C-2′′, C-3′′), 72.1 (C-5′′), 71.6 (C-8), 63.5 (C-5′), 60.1 (C-9), 55.5–55.4 ($OCH_3$), 51.1 (C-3′), 37.5 (C-2′), 12.9 ($CH_3$), 12.3 ($CH_3$); HRMS–ESI (*m/z*): $[M + Na]^+$ calcd for $C_{50}H_{51}N_5O_{13}Na$, 952.3376 found 952.3379.

5′-*O*-(Dimethoxytrityl)-DNA-CBM$_2$-LNA-thymidine dimer (DNA-CBM$_2$-LNA) (**33**)




To a suspension of dimer **32** (560 mg, 0.602 mmol) and 20% Pd(OH$_2$)/C (50.0 mg, 360 µmol) in MeOH (50 mL) was added ammonium formate (266 mg, 4.22 mmol) and stirred for 2 h at 65 °C. The reaction was filtered through celite and concentrated under reduced pressure. The residue was re-dissolved in $CH_2Cl_2$ (100 mL) and washed with $H_2O$ (3 × 100 mL). The organic phase was dried over $Na_2SO_4$, filtered and concentrated under reduced pressure to give **33** (490 mg, 0.58 mmol, 97%) as a white solid. $R_f$ 0.12 ($CH_2Cl_2$:MeOH, 19:1); $^1$H NMR (400 MHz, DMSO-$d_6$) δ 11.36 (s, 2H, H-3), 7.83 (d, *J* = 8.0 Hz, 1H, NH CBM), 7.63 – 7.53 (m, 1H, H-6), 7.45 – 7.35 (m, 3H, H-6, H-Ar), 7.35 – 7.15 (m, 7H, H-Ar), 6.93 – 6.78 (m, 4H, H-Ar), 6.16 (br t, *J* = 6.3 Hz, 1H, H-1'), 5.86 (s, 1H, OH), 5.46 (br s, 1H, H-1''), 4.53 (d, *J* = 12.7 Hz, 1H, H-8), 4.43 (d, *J* = 12.7 Hz, 1H, H-8), 4.35 (t, *J* = 7.4 Hz, 1H, H-3'), 4.17 (s, 1H, H-2''), 3.97 – 3.86 (m, 3H, H-3'', H-5'', H-4'')), 3.72 (d, *J* = 1.6 Hz, 6H, $OCH_3$), 3.69 (d, *J* = 8.1 Hz, 1H, H-5''), 3.29 – 3.14 (m, 2H, H-5'), 2.38 (m, 1H, H-2'), 2.21 (m, 1H, H-2'), 1.75 (d, *J* = 1.1 Hz, 3H, $CH_3$), 1.49 (d, *J* = 1.1 Hz, 3H, $CH_3$); $^{13}$C NMR (101 MHz, DMSO-$d_6$) δ 164.3 (C-4), 164.2 (C-4), 158.6 (C-Ar), 158.6 (C-Ar), 155.8 (C-7), 150.8 (C-2), 150.4 (C-2), 145.2 (C-Ar), 136.3 (C-6), 135.8-134.9 (C-6, 3 × C-Ar), 128.3 (C-Ar), 128.1, (C-Ar) 127.2 (C-Ar), 113.6 (C-Ar), 110.0 (C-5), 109.2 (C-5), 86.9 ($C_{quat}$DMTr), 86.8 (C-1''), 86.2 (C-4''), 84.0 (C-1'), 83.0 (C-4'), 79.6 (C-2''), 71.4 (C-5''), 70.2 (C-3''), 63.5 (C-5'), 60.4 (C-8), 55.4 ($OCH_3$), 51.0 (C-3'), 37.5 (C-2'), 12.7 ($CH_3$), 12.3 ($CH_3$); HRMS–ESI (*m/z*): $[M + Na]^+$ calcd for $C_{43}H_{45}N_5O_{13}Na$, 862.2906 found 862.2909.

5′-*O*-(Dimethoxytrityl)-DNA-CBM$_2$-LNA thymidine phosphoramidite dimer (DNA-CBM$_2$-LNA) (**34**)

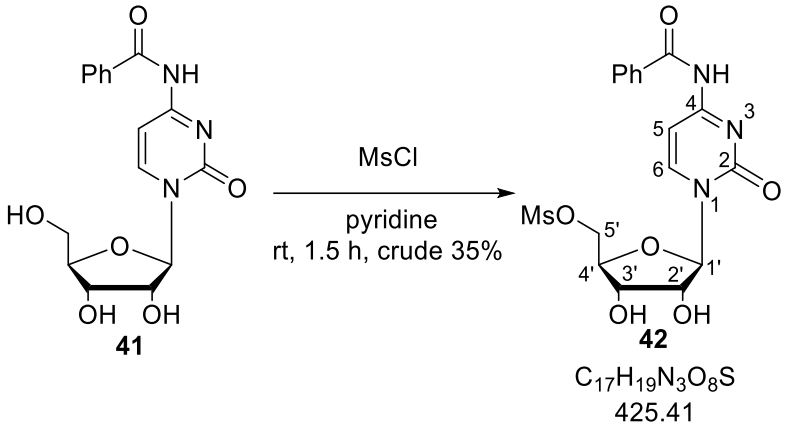



**33**

**34**
$C_{52}H_{62}N_7O_{14}P$
1040.08

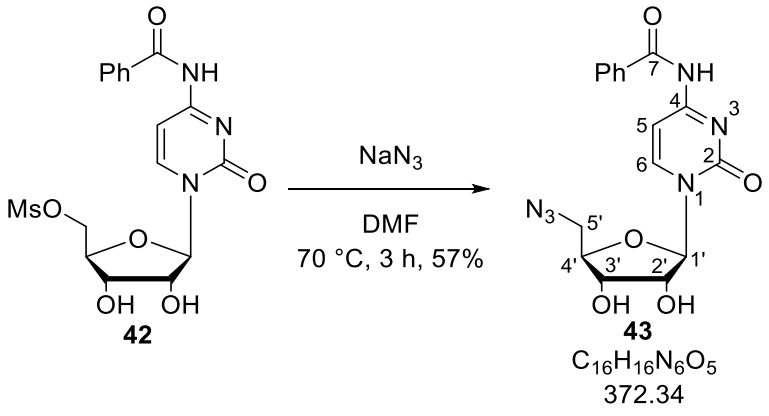
To a solution of dimer **33** (490 mg, 580 µmol) in dry degassed $CH_2Cl_2$ (5.00 mL) was added dry degassed DIPEA (254 µL, 1.50 mmol) followed by chloro(diisopropylamino)-*β*-cyanoethoxyphosphine (196 µL, 880 µmol) and stirred at rt under argon for 1 h. The reaction was diluted with degassed $CH_2Cl_2$ (5.00 mL), washed with sat. aq KCl (5.00 mL) and dried over $Na_2SO_4$. The organic layer was concentrated under reduced pressure and purified by silica column chromatography (degassed EtOAc) to give phosphoramidite **34** (260 mg, 250 µmol 43%). $R_f$ 0.56 (EtOAc); $^{31}$P NMR (162 MHz) 149.1, 147.8; HRMS–ESI (*m/z*): $[M + Na]^+$ calcd for $C_{52}H_{62}N_7O_{14}PNa$ 1062.3985 found 1062.3975

5′-Mesyl-5′-deoxy-$N^4$-benzoylcytidine (**42**)




$N^4$-benzoylcytidine **41** (2.00 g, 5.76 mmol) was refluxed in dry pyridine (100 mL) until fully dissolved and cooled to rt. MsCl (0.560 mL, 5.76 mmol) in dry pyridine (5.00 mL) was added dropwise over 1 h at rt. After 30 mins, pyridine was removed under reduced pressure and the residue was dissolved in $CH_2Cl_2$:MeOH (19:1, 50.0 mL). The organic layer was washed with $NaHCO_3$ (3 x 25.0 mL), dried over $MgSO_4$ and concentrated under reduced pressure. Silica column chromatography ($CH_2Cl_2$:MeOH, 49:1 to 19:1) yielded a mixture of compound **42** and 3′,5′-mesyl-3′,5′-deoxy-$N^4$-benzoylcytidine (95:5, 0.860 g, 2.00 mmol, crude 37%) as a white solid. $R_f$ = 0.27 ($CH_2Cl_2$:MeOH, 47:3), LRMS-ESI (*m/z*): $[M+H]^+$ 426.64 (100%).

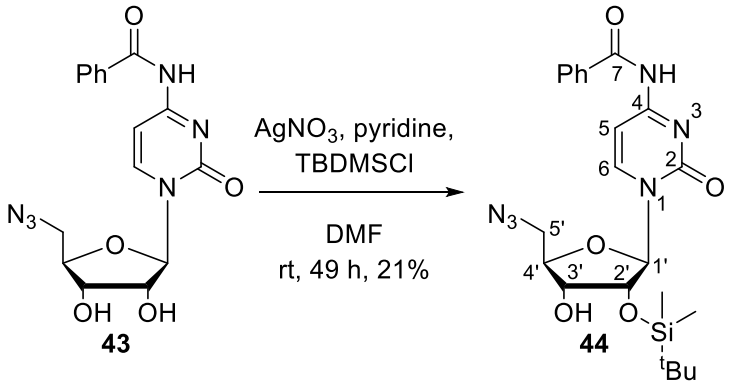
5′-Azido-5′-deoxy-$N^4$-benzoylcytidine (**43**)




To a solution of crude compound **42** (0.760 g, 1.80 mmol) in dry DMF (10 mL) was added $NaN_3$ (0.350 g, 5.40 mmol) and heated at 70 °C for 3 h. DMF was removed under reduced pressure and the residue was dissolved in EtOAc:MeOH (9:1, 25.0 mL). The organic layer was washed

with $H_2O$ (3 x 25.0 mL). The aqueous layer was back extracted with EtOAc (25.0 mL) and the combined organic layers were dried over $MgSO_4$ and reduced under pressure. The residue was purified using silica column chromatography ($CH_2Cl_2$:MeOH, 97:3 to 19:1) to produce compound **43** (0.380 g, 1.00 mmol 57%) as a white solid. $R_f$ 0.31 ($CH_2Cl_2$:MeOH, 47:3), IR ($cm^{-1}$): 3327, 2104, 1700, 1646; $^1$H NMR (500 MHz, DMSO-$d_6$) δ 11.29 (s, 1H, NH), 8.20 (d, *J* = 7.6 Hz, 1H, H-6), 8.02 (m, 2H, H-Ar), 7.67 – 7.60 (m, 1H, H-Ar), 7.53 (t, *J* = 7.8 Hz, 2H, H-Ar), 7.41 (d, *J* = 7.6 Hz, 1H, H-5), 5.86 (d, *J* = 3.6 Hz, 1H, H-1'), 5.61 (d, *J* = 5.3 Hz, 1H, H-2'OH), 5.30 (d, *J* = 6.0 Hz, 1H, H-3'OH), 4.14 (td, *J* = 5.3, 3.6 Hz, 1H, H-2'), 4.03 (td, *J* = 6.4, 3.3 Hz, 1H, H-4'), 3.93 (q, *J* = 6.0 Hz, 1H, H-3'), 3.73 (dd, *J* = 13.4, 6.4 Hz, 1H, H-5'), 3.66 (dd, *J* = 13.4, 3.3 Hz, 1H, H-5). $^{13}$C NMR (126 MHz, DMSO-$d_6$) δ 167.8 (C-7), 163.7 (C-4), 155.0 (C-2), 146.1 (C-6), 133.5 (C-Ar), 133.2 (C-Ar), 128.9 (C-Ar), 128.9 (C-Ar), 97.0 (C-5), 91.6 (C-1'), 82.4 (C-4'), 74.0 (C-2'), 70.7 (C-3'), 52.0 (C-5'); HRMS-ESI (*m/z*): $[M + H]^+$ calcd for $C_{16}H_{17}N_6O_5$, 373.1255; found, 373.1255

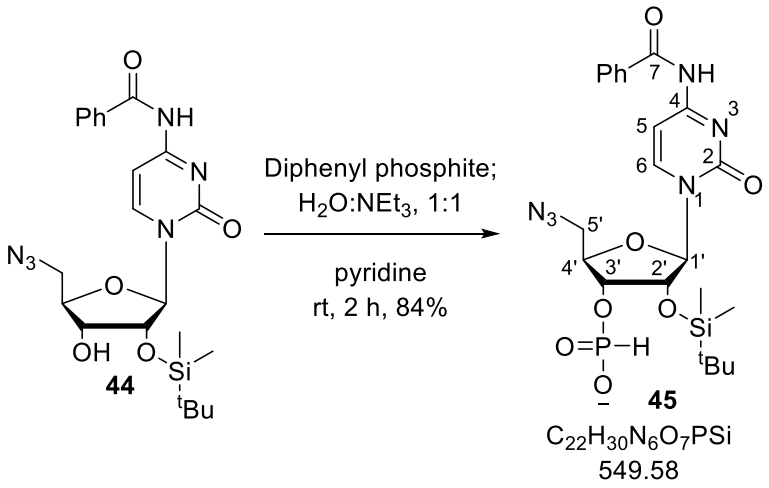
3'-*O*-(*tert*-butyldimethylsilyl)-5'-Azido-5'-deoxy-N$^4$-benzoylcytidine (**44**)




To a solution of compound **43** (0.360 g, 0.967 mmol) in dry DMF (10 mL) was added dry pyridine (0.313 mL, 3.87 mmol) and $AgNO_3$ (0.210 g, 1.26 mmol). The reaction was covered from light and stirred at rt for 30 mins. TBDMSCl (0.160 g, 1.06 mmol) was added and the reaction stirred at rt. At 16 h and 20 h further portions of TBDMSCl (0.029 g, 0.193 mmol) were added. At 48 h a further potion of TBDMSCl (0.146 g, 0.967 mmol) and $AgNO_3$ (0.164 mL, 0.967 mmol) were added. After a total of 49 h the starting material had been consumed. The reaction was concentrated under reduced pressure at rt and the residue was dissolved in $CH_2Cl_2$:MeOH (99:1, 25.0 mL). The organic layer was washed with $H_2O$ (3 x 25.0 mL), dried

over $MgSO_4$ and concentrated under reduced pressure. The residue was purified by silica column chromatography (petroleum ether: EtOAc, 1:10 to 13:9) to produce compound **44** (0.100 g, 0.206 mmol, 21%) as a white solid. $R_f$ 0.54 (Petroleum ether: EtOAc, 1:1); $^1$H NMR (400 MHz, DMSO-$d_6$) δ 11.30 (s, 1H, NH), 8.24 (d, *J* = 7.5 Hz, 1H, H-6), 8.09 – 8.01 (m, 2H, H-Ar), 7.72 – 7.62 (m, 1H, H-Ar), 7.60 – 7.51 (m, 2H, H-Ar), 7.44 (d, *J* = 7.5 Hz, 1H, H-5), 5.86 (d, *J* = 3.0 Hz, 1H, H-1'), 5.28 (d, *J* = 6.0 Hz, 1H, H-3'OH), 4.30 (dd, *J* = 4.9, 3.0 Hz, 1H, H-2'), 4.08 (td, *J* = 6.6, 3.2 Hz, 1H, H-4'), 3.97 – 3.88 (m, 1H, H-3'), 3.82 (dd, *J* = 13.5, 6.6 Hz, 1H, H-5'), 3.73 (dd, *J* = 13.5, 3.2 Hz, 1H, H-5'), 0.91 (s, 9H, H-SiCC$H_3$), 0.12 (d, *J* = 3.0 Hz, 6H, H-SiC$H_3$), $^{13}$C NMR (126 MHz, DMSO-$d_6$) δ 167.9 (C-7), 163.7 (C-4), 154.9 (C-2), 145.8 (C-6), 133.6 (C-Ar), 133.2 (C-Ar), 129.0 (C-Ar), 128.9 (C-Ar), 96.9 (C-5) , 91.7 (C-1'), 82.2 (C-3'), 76.0 (C-2'), 70.3 (C-4'), 51.8 (C-5'), 26.2 (C-Si-C(<u>C</u>$H_3$), 18.4 (C-Si-<u>C</u>($CH_3$), -4.3 (C-Si<u>C</u>$H_3$), -4.5 (C-Si<u>C</u>$H_3$); HRMS-ESI (*m/z*): $[M + H]^+$ calcd for $C_{22}H_{31}N_6O_5Si$, 487.2120; found, 487.2116.

3′-*O*-(*tert*-Butyldimethylsilyl)-4′-*O*-(hydrogenphosphonate)-5′-azido-5′-deoxy-N$^4$-benzoylcytidine (**45**)




To a solution of compound **44** (80.0 mg, 0.164 mmol) in dry pyridine (2.00 mL) was added diphenyl phosphite (157 µL, 0.822 mmol) and stirred at rt for 2 h. $NEt_3$:$H_2O$ (1:1, 0.500 mL) was added and the reaction stirred for 15 mins before concentrating under reduced pressure. The residue was dissolved in EtOAc (10.0 mL) and washed with $NaHCO_3$ (3 x 10 mL). The aqueous layer was extracted with EtOAc (10.0 mL), the combined organic layers were dried over $MgSO_4$ and reduced under pressure. The residue was purified by silica column

chromatography ($CH_2Cl_2$:MeOH:$NEt_3$, 99:0:1 to 91:8:1) to produce H-phosphonate **45** (90.0 mg, 0.138 mmol, 84%) as a white solid. $R_f$ 0.25 ($CH_2Cl_2$:MeOH:$NEt_3$, 91:8:1); $^{31}P$ NMR (162 MHz, DMSO-$d_6$) 3.19; HRMS-ESI (*m/z*): $[M + H]^+$ calcd for $C_{22}H_{31}N_6O_7PSi$, 549.1688; found, 549.1683.