Synthesis and properties of nucleic acid analogues containing artificial backbones

A thesis submitted to the Board of Mathematical, Physical and Life Sciences Division, Department of Chemistry in fulfilment for the degree of

Doctor of Philosophy

Agnès E. S. Tyburn

Linacre College, University of Oxford, Hilary 2016

Table of contents

Abstract

Synthesis and properties of nucleic acid analogues containing artificial backbones

Thesis for the degree of Doctor of Philosophy

Agnes Tyburn, Linacre College, Hilary 2016

Nucleic acid analogues are very attractive molecules for therapeutic applications as in principle they can specifically inhibit the expression of virtually any gene through antisense or RNA interference pathways. For this purpose, nucleic acid analogues with modified backbones have advantageous properties compared to their natural counterparts. In particular they can provide better resistance to nucleases as well as improving cellular uptake. However, radically changing the nucleic acid backbone while maintaining a good affinity for a mRNA or DNA target remains challenging. Developing synthetic strategies to conveniently access new modified backbones that could be used in antisense oligonucleotides is therefore essential. With this in mind a new solid-phase method to efficiently functionalise an amine-modified DNA backbone has been developed. Using this approach, three new oligonucleotide modified backbones have been synthesised, and their hybridisation properties have been investigated.

Modified DNA backbones are also potentially useful in the development of new methods for chemical-ligation of oligonucleotides. This approach could allow the large-scale *de novo* synthesis of genes containing site-specific chemical modifications including epigenetic signals. Such constructs represent challenging targets in medicinal and biotechnology applications. An important class of modified DNA linkages containing a triazole ring developed in our research group has been further investigated through the synthesis of two new triazole modified backbones. The synthesis of a DNA amide backbone with favourable duplex stability was also performed. A detailed study of the biophysical properties of the above modified backbones including their performance as templates for various DNA polymerases has been carried out.

Declaration of authorship

I, Agnès E. S. Tyburn declare that this thesis and the work presented in it are my own and have been generated by me as the result of my own original research.

I confirm that:

- This work was done wholly while in candidature for a research degree at this University;
- Where I have consulted the published work of others, this is always clearly attributed;
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;

Acknowledgements

Over my PhD years, I have acquired intellectual and personal debts to people whom in one way or another contributed in the completion of this thesis. This PhD would not have been possible without the support and guidance of my supervisor Prof Tom Brown. I would like to thank him for imparting so much knowledge, endless ideas and enthusiasm in all my PhD projects. I will always be thankful for all the scientific skills and knowledge of the "Nucleic Acid world" I was able to develop under his supervision. Many of these will be, and are already valuable assets in my professional life. I would also like to acknowledge ATDBio Ltd for the generous funding over these four years of PhD. Thank you to Prof Afaf El-Sagheer, who have been my mentor and provided selfless care, time and kept me motivated throughout these years.

Particular thanks are due to Dr Arun Shivalingam for his contribution in the 1,2,3-triazole DNA backbone project and who has been extremely generous in sharing his knowledge and his time to proofread this thesis. I also would like to thank Dr Sara De Ornellas for always answering my questions, even the most stupid ones, for giving me so much confidence in my work and for the help and advices she provided in writing up this manuscript. I would like to thank Angus Danton for being such an exemplary and motivated student, it was a pleasure to work with him. I do not forget the other members of the Brown's group, and I would like to thank in no particular order, Mei, Anna, Carolyn, Pietro, Adam, Joanna, Pui and Ellia for providing such a strong and enjoyable scientific environment. Thank you for all the dinners, drinks, tea breaks with delightful cakes, dance in the lab (only on the week-end), comforting conversations and for simply making this PhD a much more fun experience.

I would like to thank my family and the Malassis family for they kind thoughts, words of encouragement and much more send over the English Channel and the 8000 km of Atlantic Ocean separating us. *Maman, Papa, Arnaud, Gabi et Etienne, merci pour tout.*

A big cheers to all the friends and amazing people I had the chance to meet here in the UK; Luc and Chacha, Stephanie, Phoebe, Thomas, all the members of the Linclau's group and Prof. Bruno Linclau himself.

Finally I would like to thank Dr Julien Malassis, for his invaluable help and patience, for being so supportive especially in the most doubtful and difficult moments, and for simply being here.

“Just don’t give up trying to do what you really want to do. Where there is love and inspiration, I don’t think you can go wrong.”

Ella Fitzgerald

Abbreviations

app.	apparent
aq.	aqueous
AIBN	Azobisisobutyronitrile
AZT	Azidothymidine 3'-Azido-3'-deoxythymidine
br	broad
bp	base pair
Bz	Benzoyl
calc.	calculated
conc.	concentrated
CPG	Controlled Pore Glass
COSY	^{1}H-^{1}H correlation spectroscopy
CuAAC	Copper-Azide-Alkyne Cycloaddition
°C	Degree Celsius
d	doublet
dd	double doublet
ds	double strand
dt	double triplet
DCM	Dichloromethane
DCE	1,2-Dichloroethane
DEPT	Distortionless Enhancement through Polarisation Transfer
DIAD	Diisopropyl azodicarboxylate
DIPEA	Diisopropylethylamine
DMAP	4-Dimethylaminopyridine
DMF	*N,N*'-Dimethylformamide
DMSO	Dimethyl sulfoxide
DMT	Dimethoxytrityl
DMT-Cl	4,4'-Dimethoxytrityl chloride
DNA	Deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
EDTA	Ethylendiaminetetraacetic acid

eq.	equivalent
ESI	Electrospray ionisation
EtOAc	Ethyl acetate
Et_2O	Diethyl ether
Et_3N	Triethylamine
EtOH	Ethanol
FMOC	9-fluorenylmethyl
g	gram
HATU	1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate
HBTU	*O*-(Benzotriazol-1-yl)-*N*,*N*,*N*',*N*'-tetramethyluronium hexafluorophosphate
H-bond	Hydrogen bond
H-Bonding	Hydrogen bonding
HEG	Hexaethylene glycol
HMBC	Heteronuclear Multiple Bond Correlation
HMQC	Heteronuclear Multiple Quantum Coherence
HOBt	hydroxybenzotiazole
HPLC	High performance liquid chromatography
HRMS	High Resolution Mass Spectrometry
Hz	Hertz
*i*PrOH	isopropanol
kb	kilobase pair
LRMS	Low Resolution Mass Spectrometry
M	Molar concentration
m	multiplet (NMR)
MALDI-TOF	Matrix assisted laser desorption ionisation
mg	Milligram
mL	Milliliter
mm	Millimeter
MeC	5-methylcytosine
MHz	megahertz
MeCN	Acetonitrile

MeOH	Methanol
MS	Mass spectrometry
Ms	Mesyl group
Mw	Molecular weight
m/z	mass/charge
NHS	N-hydroxysuccinimide
nm	Nanometre
nM	Nanomolar concentration
µM	Micromolar
NMR	Nuclear Magnetic Resonance
ODN	Oligodeoxynucleotide
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase Chain Reaction
pH	potential of hydrogen
ppm	parts per million (NMR)
PPh_3	Triphenylphosphine
py	pyridine
q	quartet (NMR)
rt	room temperature
RNA	Ribonucleic acid
s	singlet (NMR)
ss	single strand
t	triplet (NMR)
TBAF	Tetrabutylammonium fluoride
TBDMS	*tert*-butyldimethylsilyl
TCA	Trichloroacetic acid
THF	Tetrahydrofuran
td	triple doublet
TEA	Triethylamine
TLC	Thin layer chromatography
UV	Ultraviolet

Chapter 1

General introduction

1.1 Nucleic acid structure

1.1.1 Structure of components

DNA and RNA are long thread-like polymers made of a linear array of monomers called nucleotides. Each nucleotide is constituted of thee components, namely a nitrogen heterocyclic base, a pentose sugar and a phosphate residue.

The four bases in DNA are either pyrimidines for cytosine (C) and thymine (T), or bicyclic purines for adenine (A) and guanine (G). In RNA the pyrimidine thymine is replaced by the heterocyclic base uracil (U) (*Figure 1-1*).

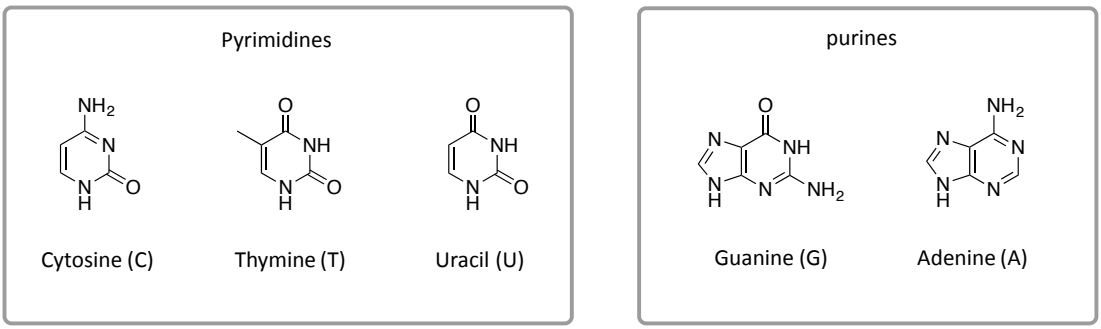


Figure 1-1: Structures of heterocyclic bases of DNA and RNA

The purine or pyrimidine base is joined to the 1'-carbon of the pentose sugar *via* a N-glycosidic bond. The pentose is D-ribose in ribonucleic acid and 2-deoxy-D-ribose in deoxyribonucleic acid; both sugars are locked into a five-membered furanose ring (*Figure 1-2*). In nucleic acids, all nucleosides are in a β-configuration, which is defined by the position of the base on the same side of the sugar ring as the 5'-hydroxyl group. The other anomer is defined as the α-anomer (*Figure 1-3*)

Figure 1-2: Example of structures of a deoxynucleoside and a ribonucleoside. The N-glycosidic bond is represented in red.

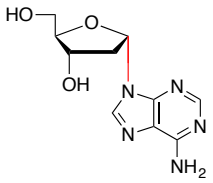


Figure 1-3: Structure of the α-anomer of deoxyadenosine

The phosphate esters of nucleosides are named nucleotides and the phosphate group can be present on either the 5’ or 3’ position of the sugar. The simplest nucleotides have a single phosphate monoester and by analogy, nucleotides with two or three phosphate monoester functions are defined as nucleoside diphosphate and nucleoside triphosphate respectively (*Figure 1-4*).

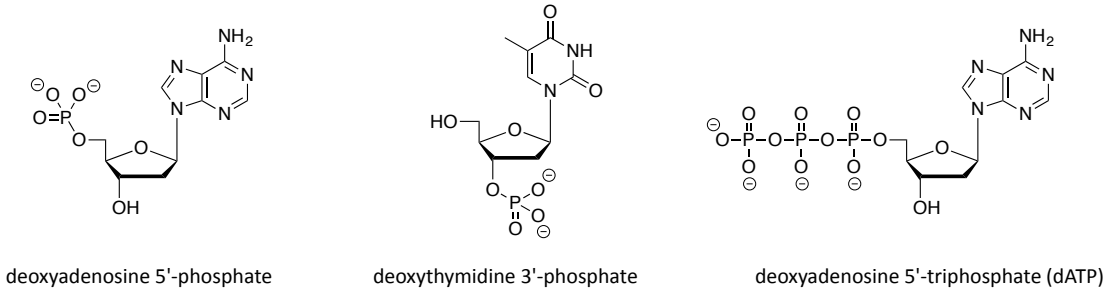


Figure 1-4: structure of some common nucleotides

In nucleic acids each nucleotide is covalently linked to its two neighbours through its 5'-phosphate and 3'-hydroxyl, the alternating sugar-phosphate chain formed, constitutes the DNA or RNA backbone. At physiological pH (~7.4) the phosphodiester bonds are negatively charged and nucleic acids are water-soluble polyanionic molecules. The succession of nucleotides interspaced by phosphodiester linkages defines the primary structure of nucleic acid, which depends solely on the sequence of the bases in the linear chain.

1.1.2 Structure of nucleic acid duplexes

One of the most fundamental discoveries on DNA was the elucidation of its 3-dimensional structure by the Nobel laureates James Watson and Francis Crick in 1953.[1] This totally reshaped the concept of the biological role of DNA, which became recognised as the molecule holding all the information for the function, development and reproduction of all living organisms (the blueprint of life). The Watson and Crick structure was based on the important experimental work of Franklin *et al.*, showing through X-ray diffraction of DNA fibres, evidence for two crystalline and helical forms of DNA, an A structure at low humidity and a B structure at high humidity.[2,3] Franklin concluded that this particular behaviour would require the phosphate groups to be exposed to water and to therefore lie on the outside of the structure. Based on these results, Watson and Crick established that the structure must adopt a double stranded helix whose two chains run in opposite direction. The evidence for the A:T and G:C base pairing was supported by the findings of Erwin Chargaff, stipulating that the molar amount of adenine in DNA was always equal to that of thymine and was also valid for guanine and cytosine (A=T, G=C).[4]

The well-known structure of the DNA duplex, as described by Watson and Crick is as follows. The double helix consists of two strands, which are complementary, and the sequence of one strand precisely defines the sequence of the other. The two strands are antiparallel and coil around each other to form a double helix, with a hydrophobic interior constituted from two base pairs, A:T and G:C linked through hydrogen bonds (*Figure 1-5*). The planar heterocyclic bases stack

one on another and the separation between successive base pairs along the helix axis is around 3.4 Å. The external hydrophilic backbone is composed of the sugars and anionic phosphodiesters. A full 360° turn of the double helix is repeated every ten to eleven base pairs. Stability of the duplex structure arises from both hydrogen bonding and base stacking interactions.

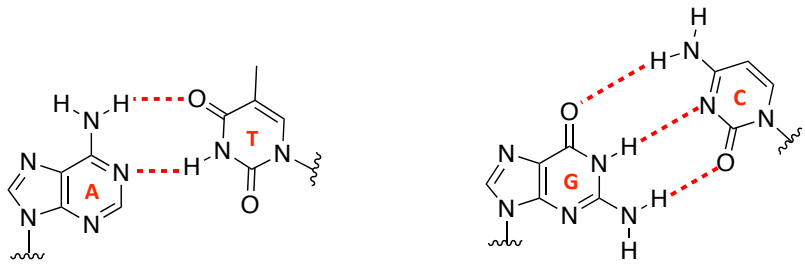


Figure 1-5: Watson-Crick base-pairing. Hydrogen bonds are represented in dashed red lines.

As discussed above, the first phase of investigation of DNA secondary structure by Franklin and co-workers, identified two distinct conformations for the DNA double helix. The B-DNA form is favored at high humidity while the A-DNA form is predominant at low humidity. Since these findings, hundred of different oligonucleotide structures have been solved and they provide detailed high-resolution information on nucleic acid structure conformation.

The A-form and B-form are the major standard conformations of nucleic acid duplexes and are right-handed helixes; the backbone at the front of the molecule facing the observer slopes down from top right to bottom left. In B-DNA form, the base pairs are nearly perpendicular to the helix axis so that the wide major groove and the narrow minor groove are both of moderate depth (*Figure 1-6*). The A-form has 11 base pairs per turn instead of 10.5 base pairs as in B-form. The base pairs are tilted sideways to enhance base stacking and pulled away from the center of the helix, which results in a fairly stiff helix with a deep major groove and a very shallow minor groove (*Figure 1-6*). The structure preferentially adopted by DNA is the B-form whereas the structure adopted by RNA is the A-form.

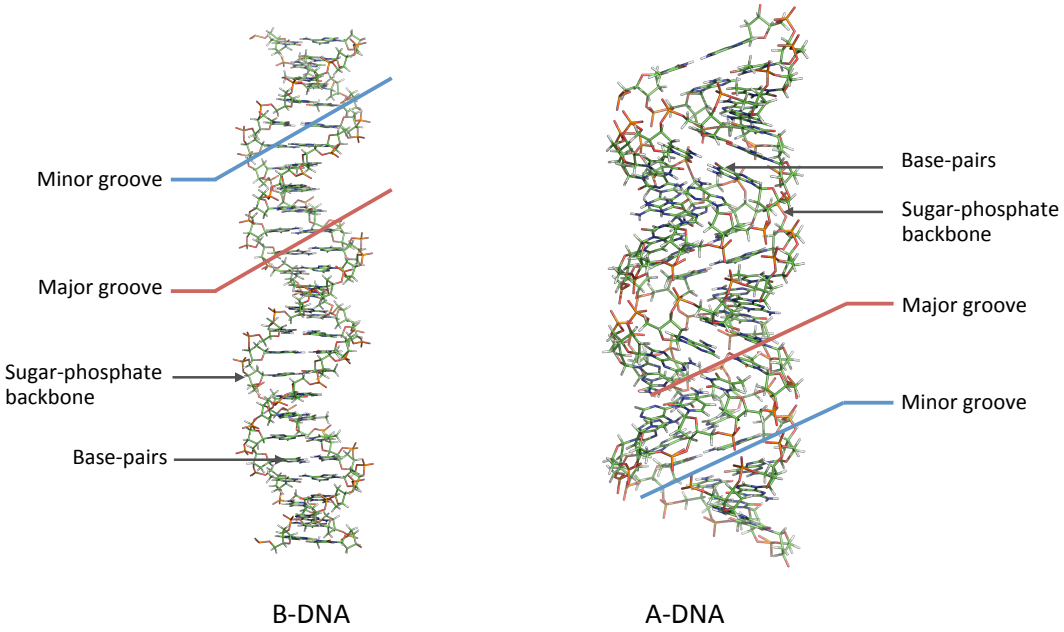


Figure 1-6: Three-dimensional structure of B-DNA and A-DNA double helixes. Sugar-phosphate backbone is represented in green/orange and the bases-pairs in blue

1.1.3 Helix coil transition of nucleic acid duplexes

The Watson and Crick model, first describes DNA as a rigid double helix. It is nowadays known, that nucleic acids are not stiff or static molecules, but dynamic structures that can hybridize, denature and anneal. These features of the behaviour of nucleic acids are essential for their *in vivo* biological functions and applications in various areas including DNA sequencing,[5] forensic analysis,[6] genome engineering,[7,8] and nanotechnology.[9]

The separation of two nucleic acid strands, by the disruption of the hydrogen bonds and base stacking, can be caused by various conditions such as heat, change of pH or use of specific organic solvents. These structural transitions from double helix to separate strands can be monitored by various physical methods, which include UV light absorption. Unpairing of two single DNA or RNA strands results in an increase of UV light absorption due to the disruption of base stacking. This hyperchromicity can usually be monitored at 260 nm and an increase of UV absorption of about 40 per cent can typically be observed. The dissociation of the

double helix to strand-separated coil upon increase of the temperature gives a sharp cooperative transition, with a midpoint defined as melting temperature (T_m) (*Figure 1-7*).[10]

The melting temperature of a given nucleic acid duplex depends on various parameters such as, chain length, percentage of GC base pairs, base sequence, oligonucleotide concentration, as well as pH of the solvent, salt concentration and nature of the counter-ions.

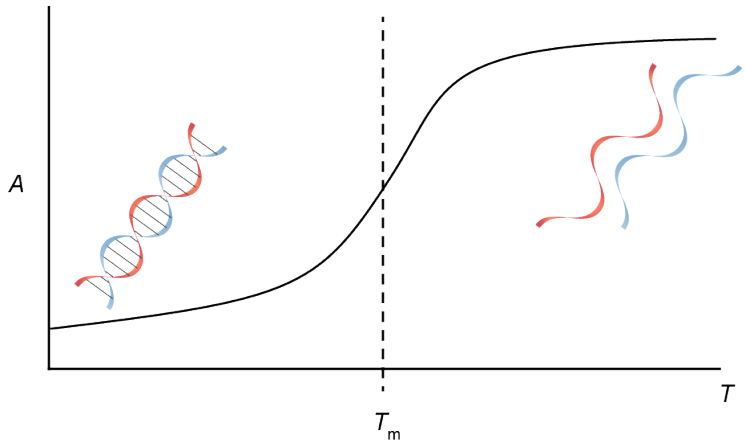


Figure 1-7: UV thermal denaturation of a nucleic acid duplex

1.2 Chemical oligonucleotides synthesis

Oligonucleotides are short single strands composed of a defined sequence of nucleotides and this term is commonly used to describe any chemically synthesised nucleic acid strands. Oligonucleotides have two hydroxyl extremities defined as 5'- and 3' ends and which give the nucleic acid chain its direction. By convention, the oligonucleotide sequence is written from the 5' to the 3' end (*Figure 1-8*).

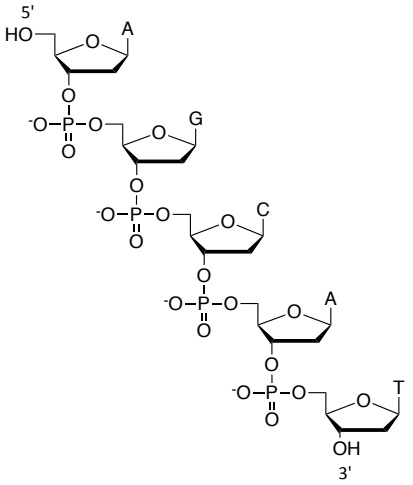


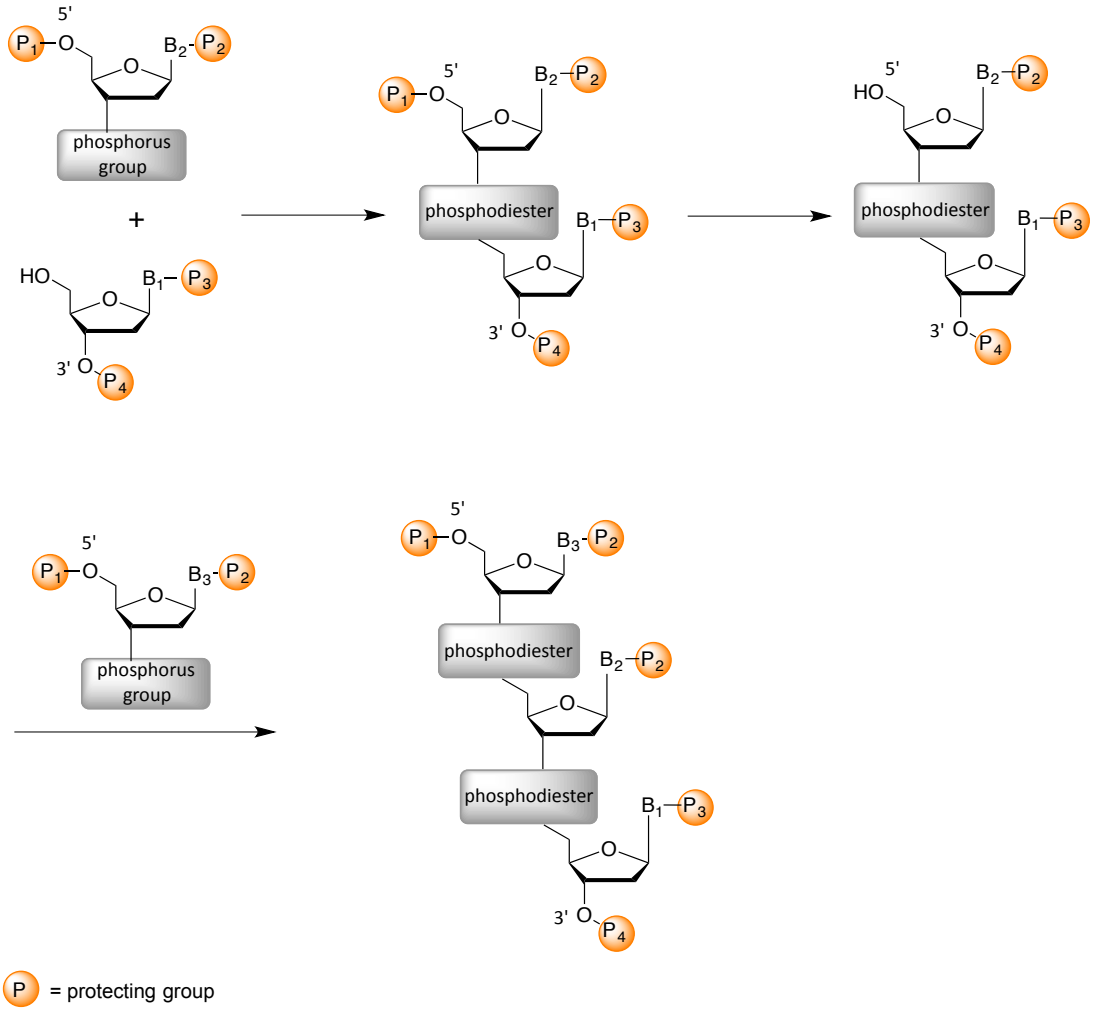
Figure 1-8: Structure of oligonucleotide of sequence AGCAT

Such molecules are nowadays routinely used in laboratories for diagnostic and genomic applications and the development of effective chemical methods for their synthesis, has led to great advances in molecular biology.[11] Chemical oligonucleotides synthesis has enabled the synthesis of a wide range of modified oligonucleotides and oligonucleotide analogues that have been used in gene silencing and gene synthesis applications, as will be discussed later.

1.2.1 Overall strategy for oligonucleotides synthesis

The main challenge in oligonucleotide synthesis is the selective and sequential formation of the phosphodiester internucleosidic linkage. In all oligonucleotide synthesis strategies developed, this involves the protection of other reactive groups that could interfere in the formation of the linkage. The primary

amines of the heterocyclic bases have to be protected as well as the 5′-hydroxyl of one building block and the 3′-hydroxyl of the other building block (*Scheme 1-1*)



Scheme 1-1: Strategy for oligonucleotide synthesis. The heterocyclic bases must be protected and one of the terminal hydroxyl groups has to be selectively deprotected to allow chain elongation. In solid-phase synthesis P_4 is replaced by a solid support, the synthesis is usually performed in 3′ to 5′ direction.

The chemical synthesis of oligonucleotides has a particular advantage over polymerase-based synthesis as the chain extension can be carried out in 5′ to 3′ or 3′ to 5′ directions. To extend the chain, one of the protecting groups on the 5′- or 3′-hydroxyl has to be selectively removed to allow coupling of the next protected building unit (*Scheme 1-1*). In solid-phase oligonucleotide synthesis, which will be

described later, the synthesis is usually performed in the 3' to 5' direction and the 3'-hydroxyl is attached to the solid support, which serves as a protecting group.

1.2.2 Making the internucleosidic phosphodiester linkage: the phosphoramidite approach

The evolution of oligonucleotide synthesis saw the development of several methods for the formation of the internucleosidic phosphate linkage. The early experiments of Khorana and co-workers utilised nucleoside phosphodiester building blocks. Todd's H-phosphonate approach and the phosphotriester methods were also applied to solid phase synthesis between the late 1970's and early 1980's (*Figure 1-9*).[12–14] All these methods presented drawbacks and were quickly supplanted by the Caruthers phosphoramidite approach,[15] which became the method of choice for the solid-phase synthesis of oligonucleotides.

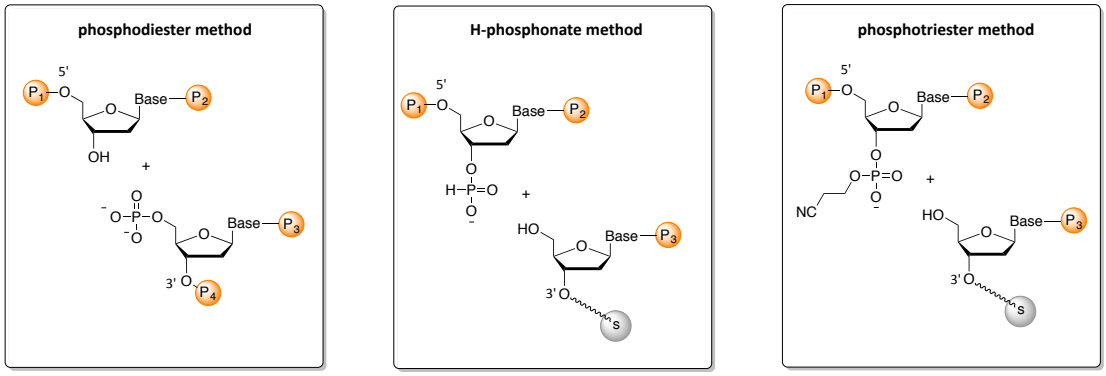


Figure 1-9: Previous methods for the synthesis of the phosphodiester linkage

The advantage of this method over the others mentioned above is the high yield (> 98 %) of the coupling reaction to form the phosphite triester bond. This reaction requires the presence of an activator to protonate the diisopropylamine function that is then rapidly displaced by attack of the 5'-hydroxyl group of the nucleoside attached to the solid support (*Figure 1-10*). The Caruthers method requires the use of preformed protected nucleoside phosphoramidite building blocks that will be described in the following section.

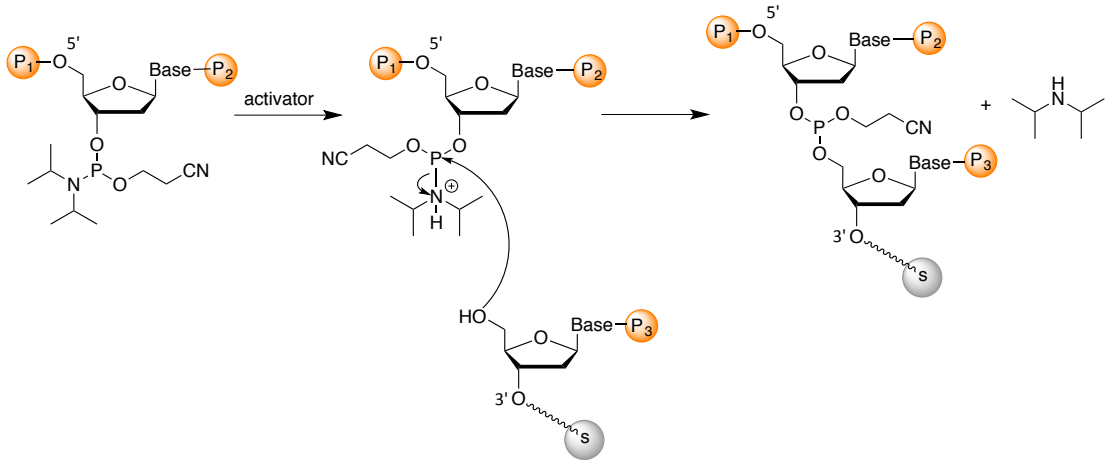


Figure 1-10: mechanism of phosphoramidite coupling

1.2.3 Synthesis of the nucleoside phosphoramidite building blocks

The four phosphoramidite monomers used in DNA synthesis are prepared from A, G, C, and T deoxynucleosides (*Figure 1-11*).

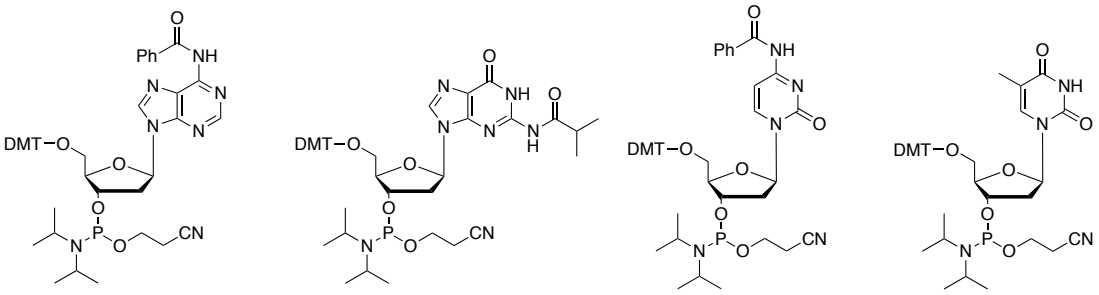
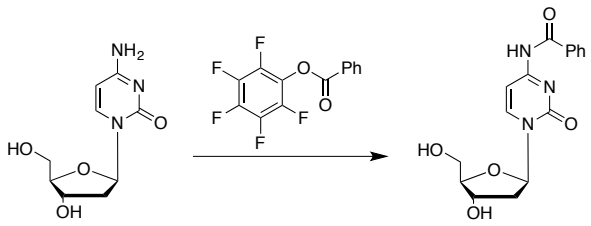


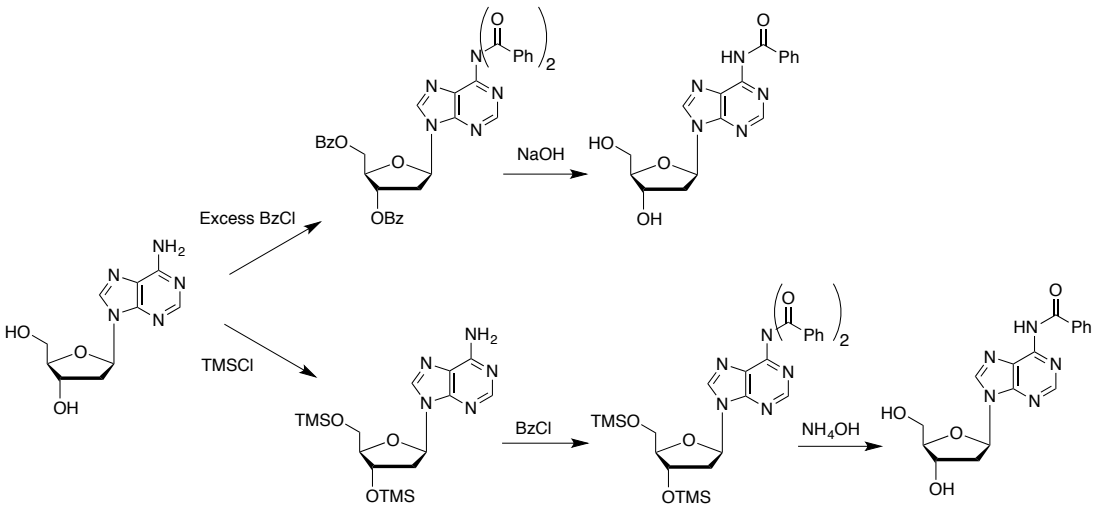
Figure 1-11: Structures of the four phosphoramidite monomers

First, the primary amines of adenine, guanine and cytosine bases have to be protected. Acyl protecting groups are usually chosen, as they remain stable under the acid and basic conditions used for the oligonucleotides synthesis. The benzoyl group is commonly used to protect both adenine and cytosine and the *iso*-butyryl group is used to protect guanine. For synthesis of the deoxycytidine monomer, selective acylation of the primary amine is performed using pentafluorophenyl benzoate (*Scheme 1-2*).



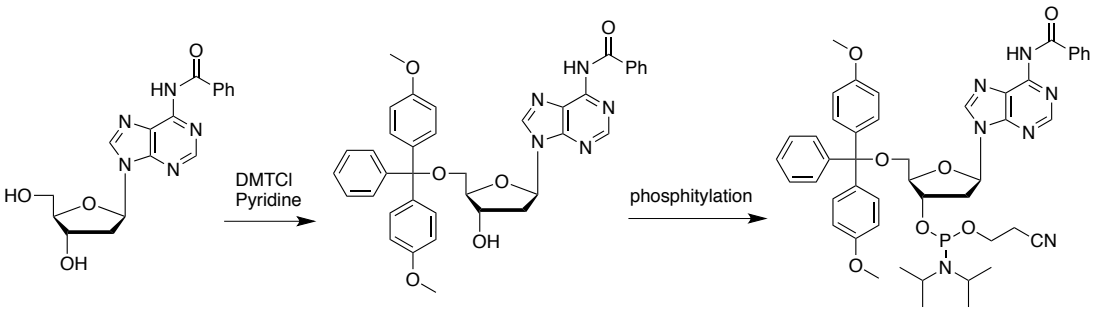
Scheme 1-2: Benzoylation of deoxycytidine

The amino groups of adenine and guanine are too weakly nucleophilic and the sugar hydroxyl groups can compete during the protection reaction. Thus, two common procedures are used for acylation of these primary amines. The first one uses an excess of acylating agent (benzoyl chloride), followed by selective deacylation of the hydroxyl groups by treatment with sodium hydroxide. The second route consists in the protection of the sugar hydroxyls as silyl ethers. Benzoylation then give the dibenzoyl derivative, which is deacylated upon basic treatment. The silyl protecting groups are also removed during this last step (*Scheme 1-3*). Protection of guanine is performed using either route with isobutyric anhydride as acylating reagent. Similar procedures are sometimes used for benzoyl dC.



Scheme 1-3: Benzoyl protection of deoxyadenosine

The nucleosides are then protected on the 5'-position using dimethoxytrityl chloride in the presence of a mildly basic catalyst such as pyridine. Following the purification of the DMT-nucleosides the 3'-position is phosphitylated using 2-cyanoethyl diisopropylaminophosphorochloridite in the presence of a non-nucleophilic base such as diisopropylethylamine (DIPEA) (*Scheme 1-4*). The DMT-protected phosphoramidite nucleosides obtained are stable for years when dry and stored in anhydrous and oxygen-free conditions. They are stable for days when dissolved in anhydrous acetonitrile.



Scheme 1-4: tritylation and phosphitylation of deoxyadenosine

1.2.4 Solid-phase oligonucleotide synthesis

Solid-phase synthesis involves a heterogeneous reaction between a residue bound to a solid support and reactants in solution. This method presents several advantages over solution synthesis. First, a large excess of reagent can be used to drive reactions quickly to completion. Second, impurities and excess reagents are easily washed away which avoids purification after each synthetic steps.[16] Two types of solid support are commonly used in oligonucleotide synthesis, which are controlled pored glass (CPG) and polystyrene.

Solid-phase oligonucleotide synthesis usually proceeds in the 3' to 5' direction. Four distinct resins pre-functionalized with one of the four deoxynucleosides (A, G, C and T) through the 3'-hydroxyl, are commercially

available. Solid-phase phosphoramidite oligonucleotide synthesis consists of a series of steps where one nucleotide is added during each synthesis cycle (*Figure 1-12*).

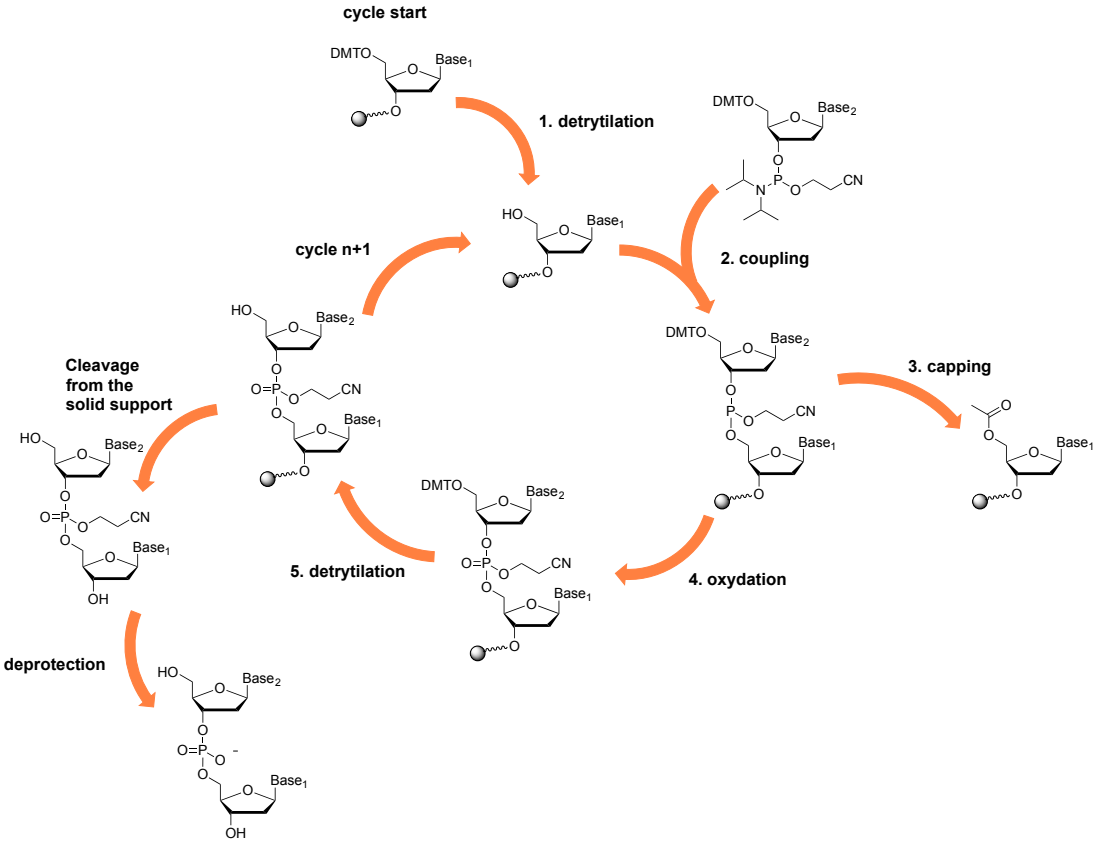


Figure 1-12: Oligonucleotide solid phase synthesis cycle

At the beginning of the synthesis, the 5'-hydroxyl group of the first nucleoside attached to the resin is deprotected by treatment with trichloroacetic acid (TCA). The orange colour coming from the dimethoxytrityl cation liberated is monitored in order to obtain the coupling efficiency (step 1).

Following this, the support-bound nucleoside is coupled with the next phosphoramidite, which is activated with tetrazole. This compound is sufficiently acidic enough to protonate the phosphoramidite monomer without causing the loss of the 5'-DMT group (step 2).

The capping step is a safety step designed to limit the number of failure oligonucleotide sequences. The 5'-hydroxyls groups of the oligonucleotide chains that did not react during the previous coupling reaction are therefore blocked by acylation using acetic anhydride (step 3). The oxidation of the phosphite-triester formed during the coupling reaction, is subsequently achieved with iodine and water (step 4).

This cyclic process is then repeated until the correct oligonucleotide sequence is obtained. The last 5'-DMT group is removed using the same detritylating agent used during the synthesis cycles. The oligonucleotide is then cleaved from the solid support, and the remaining protecting groups are removed by treatment with concentrated ammonium hydroxide. Synthesis of oligonucleotides up to 150 nucleotides in length can be obtained using solid phase phosphoramidite chemistry. The robustness of this method makes it easily applicable to automation and this solid phase oligonucleotide synthesis is currently used in most of commercial DNA synthesisers.[17]

1.3 Oligonucleotide analogues with modified backbones and their applications

The development of the chemical synthesis of oligonucleotides greatly contributed to opening up the scope of nucleic acid chemistry. This quickly led to the synthesis of numerous oligonucleotide analogues that have been used as tools for DNA-based nanotechnology[18], nucleic acid detection[19,20], sequencing[21] and aptamer design and synthesis[22].

Different positions on the nucleic acid can be altered and this includes the base, the pentose sugar and the phosphate backbone (*Figure 1-13*).[23] This section will focus on oligonucleotide analogues containing modified backbones with an emphasis on their therapeutic and gene synthesis applications.

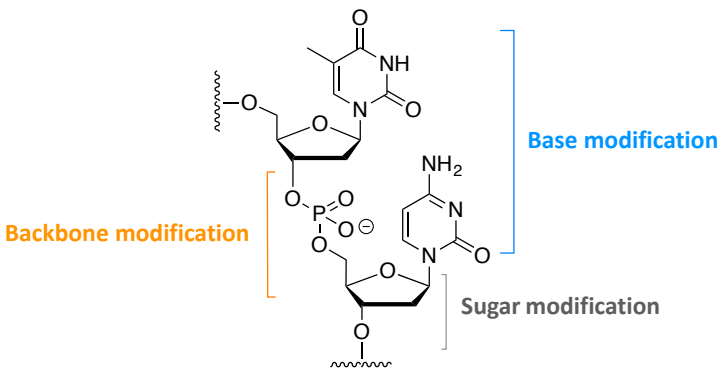


Figure 1-13: modification sites on DNA

1.3.1 Modified DNA backbone for gene silencing therapeutic applications

Since the discovery of the potential therapeutic use of oligonucleotides through gene silencing mechanisms, such as antisense, RNAi or exon-skipping,[24] modifications have been made on nucleic acids in order to improve their intrinsic properties. Indeed, the use of natural DNAs or RNAs as therapeutic agent is limited. These molecules would first be rapidly degraded through the hydrolytic action of cellular nucleases on the sugar-phosphate backbone. Moreover, the pharmacokinetics of these large polyanionic molecules had to be improved mainly

by enhancing their cellular uptake. Optimising oligonucleotides properties through alteration of their structure without reducing their functional activity as gene inhibitors remains a great challenge. One of the most effective and widely used strategies is to modify the oligonucleotide backbone,[25,26]and the most frequently used and promising analogues involved in pharmaceutical trials[27] are described below.

1.3.1.1 Phosphorothioate oligonucleotides

Phosphorothioate (PS) oligonucleotides are modified DNA in which one of the two non-bridging oxygen atoms is replaced by a sulphur atom.[28,29] They can be easily synthesised through automated phosphoramidite oligonucleotide solid phase synthesis. The synthesis proceeds as previously described except for the iodine oxidation step, which is replaced by sulfurisation using for example tetraethylthiuram disulphide. Since the phosphitylation of the monomers is not stereospecific, a mixture of phosphorothioate oligonucleotide diasteroisomers is therefore obtained (*Figure 1-14*). Other methods for the chemical synthesis of PS oligonucleotides have been reviewed by Guga *et al.*[30]

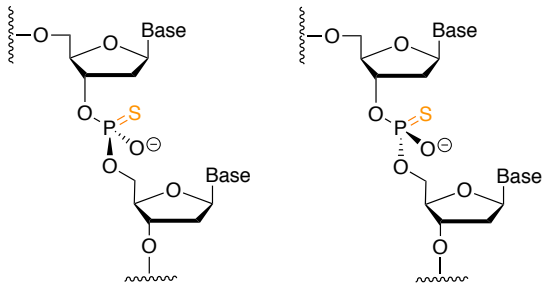


Figure 1-14: Diastereoisomers of phophorothioate backbone

The great utility of phosphorothioate analogues lies in their resistance to degradation by nucleases. They are compatible with the RNAi biological machinery and are consequently able to activate RNase H degradation of the target mRNA.[31] This modification is the most widely used in RNA-targeted drugs being currently

studied in clinical trials or already approved by the US Food and Drug Administration (FDA).[32–35] For instance, the antisense drug (Vitraven) is a 21-mer phosphorothioate, directly injected into the eye, to treat cytomegalovirus-induced retinitis.[36] This complicated mode of administration is partly necessitated by the low permeability of these modified oligonucleotides through the cell membrane. Interesting reviews focus on strategies that have been developed to improve this poor property *via* conjugation of PS oligonucleotides with other organic molecules such as cell penetrating peptides (CPP) or cholesterol.[37,38]

Another drawback of the phosphorothioate modification is that it destabilises the duplex formed with a complementary DNA or RNA strand. This is mainly due to the bulkiness of sulphur relative to oxygen, and it is usually overcome by the addition of other thermally stabilising modifications (e.g. 2'-OMe sugar modifications).[32] Phosphorothioate oligonucleotides where all the chiral centres are in Rp configuration, proved to have a higher binding affinity than pure Sp or mixed analogues.[39,40] The Rp oligomer was also found to better induce RNase-H activation than Sp or diastereomeric mixtures of oligonucleotides.[39] However, the large scale synthesis of enantiopure phosphorothioate oligonucleotides still remains a challenge despite considerable efforts by many research group.[41–44]

1.3.1.2 Methylphosphonate and phosphoramidates oligonucleotides

Methylphosphonate (MP) modified oligonucleotides were first introduced by Miller *et al* in 1971.[45] In MP oligonucleotides, one of the non-bridging oxygen is replaced by a methyl group, giving a neutral backbone with a high stability towards nucleases (*Figure 1-15*). It was envisaged that the reduction of the overall charge of the oligonucleotide, with the introduction of neutral linkages would facilitate the penetration through the cell membrane. Unfortunately this was not the case and internalisation of MP oligonucleotides is reduced compared to phosphorothioate oligonucleotides.[46] This modification also severely affects the hybridisation properties of the oligonucleotide to its mRNA target. As for PS oligonucleotides, the melting temperature of MP oligonucleotides can be increased by using a chiraly pure oligonucleotide with the Rp configuration.[47] However the synthesis of such

enantiopure oligonucleotides is too challenging to be a viable commercial option. This last point, in addition to the poor solubility of these neutral modified DNA has impeded their extensive use in antisense applications.

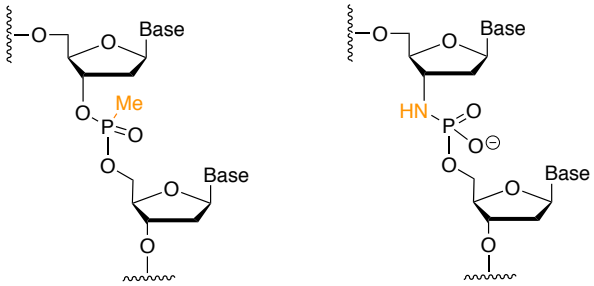


Figure 1-15: Structure of methylphosphonate and N3'-P5' phosphoramidate modified DNA backbones

N3'-P5' phosphoramidate oligonucleotides are modified DNA in which the 3'-oxygen of the phosphodiester internucleosidic linkage has been replaced by an amino group (*Figure 1-15*). These modified oligonucleotides are resistant to nuclease degradation and they hybridize specifically to the mRNA target with a higher affinity compared to the aforementioned phosphorothioate and methylphosphonate oligonucleotides.[48] In this regard, they are better antisense drug candidates. As for MP oligonucleotides, these analogues do not induce RNase H activity and are used in other gene silencing mechanisms that do not employ this mode of antisense inhibition.[49] A disadvantage of this modified backbone is that it is labile in acidic conditions. Strategies to circumvent this problem reside in the introduction of electron-withdrawing groups near the backbone or by the substitution of one of the non-bridging oxygen for a sulfur atom to form N3'-P5' thiophosphoramidate modified backbones.[50,51]

1.3.1.3 Boranophosphate oligonucleotides

Boranophosphate is a slightly more recent DNA backbone modification. In this linkage, one of the non-bridging phosphodiester oxygen is replaced by a BH_3 borane moiety (*Figure 1-16*).[52] As for PS oligonucleotides, boranophosphate oligonucleotides are able to induce RNase H activity.[53] They are twice as resistant to nucleases and could present a reduced toxicity compared to PS oligonucleotides. While the boranophosphate linkage retains a negative charge, it present an increased lipophilicity compared to phosphodiester and phosphorothioate linkages.[54]

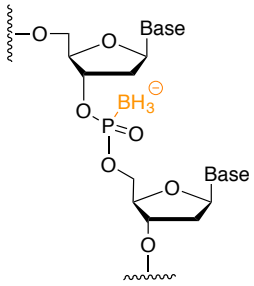


Figure 1-16: Structure of boranophosphate modified oligonucleotide

As for methylphosphonate and phosphorothioate modified backbones, introduction of the BH_3 moiety leads to a racemic mixture of phosphorus derivatives. However, enzymatic and chemical methods were successfully applied to the synthesis of enantiopure boranophosphate oligonucleotides.[54,55] Shaw and co-workers have demonstrated that siRNAs containing boranophosphate modifications towards the ends showed improved potency compared to PS siRNA and natural phosphodiester siRNA.[54] Besides, a single stranded siRNA containing boranophosphate modifications was able to silence gene expression as efficiently as non-modified double stranded siRNA.[56] Although significantly less studied than other backbone modifications, boranophosphate oligonucleotides display interesting *in vitro* gene silencing properties and it would be interesting to use this modification in more thorough *in vivo* antisense studies.[57]

1.3.1.4 Amide modified backbone

The amide linkage was first introduced into oligonucleotides by De Mesmaeker and co-workers (*Figure 1-17*), and amide modified ssDNA hybridises with a higher affinity to a complementary RNA (+ 0.9 °C per modification) compared to unmodified oligonucleotides and slightly destabilises DNA duplexes (- 1.5 °C per modification) [54,55,60] This heavily modified backbone also provides an improved resistance to 3'-exonucleases.[58] The amide bond is compatible with solid phase phosphoramidite synthesis and is readily accessible by simple synthetic methods. An efficient solid-phase synthesis of amide-linked RNA has also been developed by Iwase and co-workers.[61]

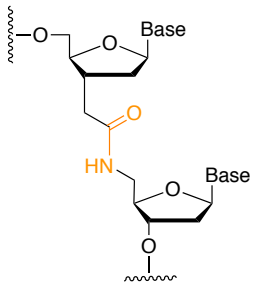


Figure 1-17: Structure of amide-modified backbone

Rozners *et al.* found that the amide linkage associated with either 2'-OH or 2'-O-Me sugars were well accommodated in RNA duplexes and demonstrated with thermodynamic and NMR structural studies that the amide linkage had very little effect on the A-type conformation, thermal stability and hydration of the RNA duplex.[59] They have latter showed using X-ray crystallography that amide-modified RNA forms a typical A-form duplex with the amide carbonyl group pointing into the major groove in a similar orientation as the phosphodiester bond and suggest that this carbonyl could mimic one of the non-bridging P-O bond.[62] Studies by Iwase and co-workers have showed that amide backbone modifications were successfully incorporated in the 3'-overhangs of an siRNA[61] and Rozners *et al.* proved that it was also tolerated at internal positions of both sense and antisense strands.[62] The silencing activity was increased when placed near the 5'-end of the passenger strand

and a siRNA containing eight amide modifications presented a better silencing activity compared to its non-modified siRNA counterpart.[62] Desaulniers and co-workers have also showed that siRNAs containing a PNA-derived amide linkage was well tolerated in the 3'-overhang of the passenger strand.[63] Amide-modified oligonucleotides could be a promising new class of therapeutic oligonucleotides, and future *in vivo* siRNA or antisense studies will be decisive to reach this goal.[57]

1.3.1.5 Peptide nucleic acid (PNA)

Peptide nucleic acids have a backbone similar to that of poly-amino acids (peptides). The phosphodiester backbone has been replaced by a 2-aminoethyl-glycine and the nucleotide bases are connected through a methyl carbonyl linker to the backbone amino groups (*Figure 1-18*).[64] PNA oligomers can be easily prepared using standard solid phase peptide synthesis.[65] Despite all these structural variations from natural nucleic acids, PNA hybridize with DNA and RNA tighter and with greater specificity and avidity compared to phosphodiester and phosphorothioate oligonucleotides, partly because they are not anionic.[66] They are unsurprisingly also fully resistant to endo- and exonuclease activity.[67,68]

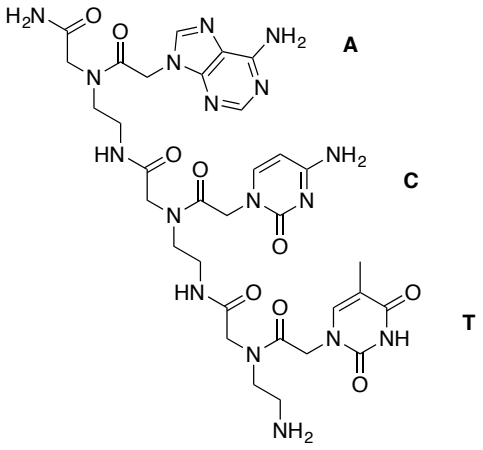


Figure 1-18: Structure of PNA analogue ACT

PNA is not compatible with the RNAi biological machinery but has been used in the development of gene-targeted drugs applying antigene or antisense strategies.[69] They have been also successfully used in siRNA overhangs[70] and in numerous other biological pathways including splice-switching activities.[71,72] Nielsen and co-workers have shown that PNA was capable of selectively preventing the translation process by forming PNA:RNA duplexes or $(PNA)_2$:RNA triplexes. Such complexes create steric hindrance that either blocks the assembly of the ribosome initiation complex or stops the ribosome elongation.[73] PNA can also be used to inhibit DNA replication through strand invasion of the targeted DNA duplex.[74–76]

The major hurdle in the use of PNA constructs as drugs is the development of an efficient *in vivo* delivery method.[77] Indeed, these molecules are not easily taken up by pro-or eukaryotic cells and have poor *in vivo* pharmacokinetic properties due to rapid excretion through the kidneys.[78] Efforts to overcome this problem have been made through the conjugation of PNAs with cell penetrating peptides or other molecules;[72,79–81] however more efficacy and pharmacokinetic studies are needed towards the development of an efficient PNA-based therapeutic agent.

1.3.2 Modified DNA backbones for gene synthesis applications

1.3.2.1 Enzymatic-based gene synthesis methods

Current methods of gene synthesis typically use enzymes in order to assemble chemically synthesised oligonucleotides. DNA assembly methods that are usually used to construct relatively small DNA fragment (~2 kb), can be divided into two groups, ligation-based assembly and PCR-based assembly.

Ligation-mediated methods are among the earliest gene synthesis strategies and use DNA ligase enzymes to assemble synthetic oligonucleotides (≤150 bp). Shot-gun ligation produces the desired gene *en masse* using a pool of oligonucleotides and DNA ligase in a one-pot reaction.[82] The reaction usually requires extended time and although it has been improved with the use of thermostable DNA ligase, it was quickly supplanted by the ligase chain reaction (LCR).[83] The first method of LCR developed uses 5'-phosphorylated oligonucleotides with overlapping sequences

designed to span the entire length of the targeted dsDNA.[84] This method involves several cycles of temperature variations similar to PCR. Oligonucleotides are first mixed together, denatured by heat, and then cooled down to promote annealing and ligation by the thermostable ligase. The product formed can then serve as template in the next ligation events and the cycles are repeated until the targeted gene is assembled. This method is often accompanied by a final PCR step to amplify the gene fragment. Recent advances in LCR gene synthesis have been made by Pachuk *et al.* and involve the use of bridge oligonucleotides.[85] Each bridge oligonucleotide is complementary to the ends of the two DNA strands to be ligated and serve as catalyst in a product-driven reaction (*Figure 1-19*). The recent optimisation of this method by Kok and co-workers enabled the efficient one-step assembly of a 20 kb DNA construct.[86]

Another widely used gene synthesis technique relies on the use of PCR to mediate assembly of a DNA fragment from shorter oligonucleotides. Most PCA (polymerase cycling assembly)-based protocols involve two main steps. In the first step overlapping oligonucleotides are annealed and recursively elongated with a thermostable polymerase until the full-length target sequence is obtained. The second step consists in the amplification of the full-length product and is achieved by addition of extension primers in another round of PCR (*Figure 1-20*).[87] This process is compromised by occurrence of errors, mainly due to chemical oligonucleotide synthesis and to a lesser extent, to the subsequent polymerase extension reactions. Strategies have been developed for error corrections in gene synthesis and have been reviewed by Tian and co-workers.[88]

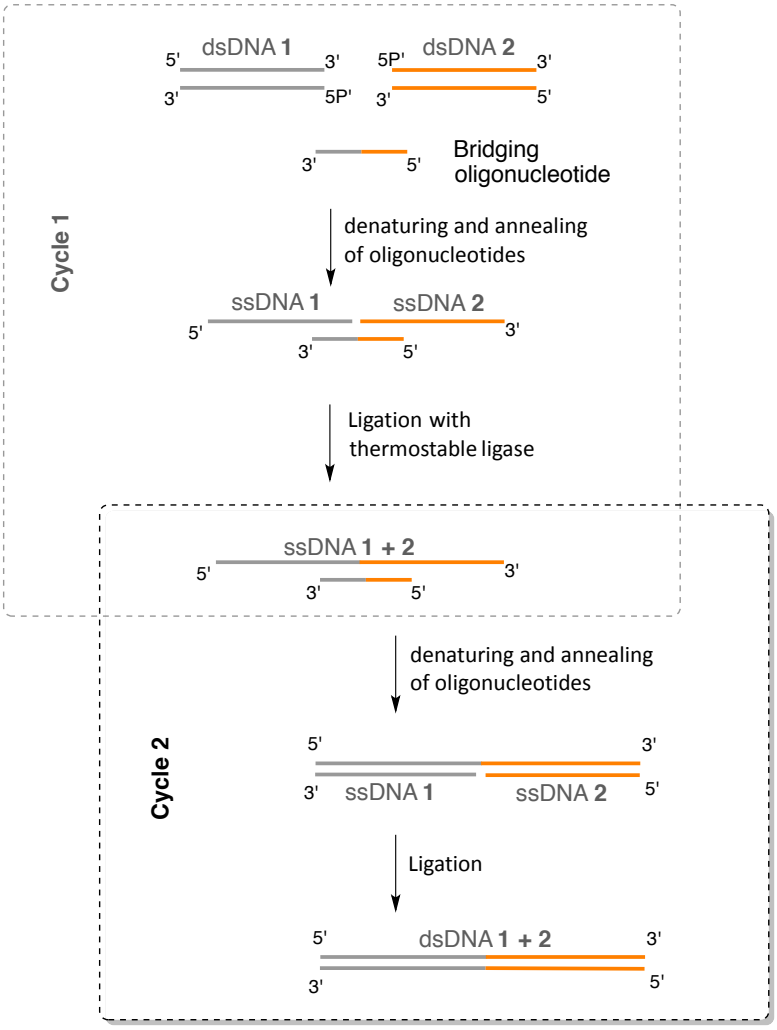


Figure 1-19: DNA assembly *via* Pachuk *et al.* LCR method. After denaturation at 94 °C, neighbouring DNA parts (5'-phosphorylated) are annealed at 55 °C on both halves of the bridging oligonucleotides. The thermostable ligase then joins the DNA *via* a phosphodiester bond. In the second and following cycles, the newly assembled strand serves as a template for ligation of the other strands.

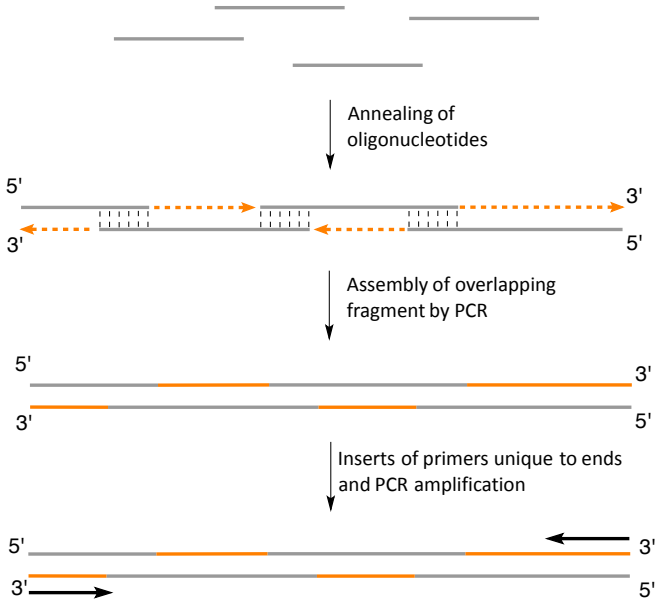


Figure 1-20: basic principle of gene assembly using PCA

A number of new DNA assembly methods have been developed built upon the PCA methods, such as Gibson assembly and SLIC (Sequence and Ligation independent cloning), and are mainly applicable to the synthesis of bigger DNA constructs (several hundred kb). Alternative methods to ligation and PCA mediated assemblies also exist and employ standardised restriction enzyme protocols like the Golden gate method. All these recent methods have been exhaustively reviewed by Zaho *et. al.*[89]

1.3.2.2 Chemical synthesis of genes: basic concept

In the past decades, enzymatic approaches for DNA assembly enabled to go far beyond the mere synthesis of genes[90] to the synthesis of entire organisms genomes.[91,92] However, the main limitation intrinsic to these kind of method, is that it is not possible to incorporate various site-specific modifications throughout the targeted dsDNA. Some base modifications, such as 5-Methylcytosine and 5-hydroxymethylcytosine are of capital importance for a comprehensive understanding of epigenetic mechanisms and their impact in human diseases.[93,94] Therefore, there is a clear need of methods that could synthesise DNA stretches containing selective insertions of one or a combination of epigenetic and mutagenic bases. The synthesis of more exotic genes, containing for example unnatural base

pairs would also be interesting and could lead to the development of artificial organisms an to further advancement in synthetic biology.[95–97]

A promising alternative approach to current enzymatic gene synthesis, resides in the sole use of chemical reactions for the total synthesis and assembly of genes and genomes. This would require the development of efficient chemical reactions to couple very pure oligonucleotide fragments (*Figure 1-21*). To do so, two strategies could be adopted. The first one would enable the ligation of oligonucleotides through a phosphodiester linkage whereas the second would exploit a modified internucleotide bond presenting similar properties as the natural phosphodiester backbone.

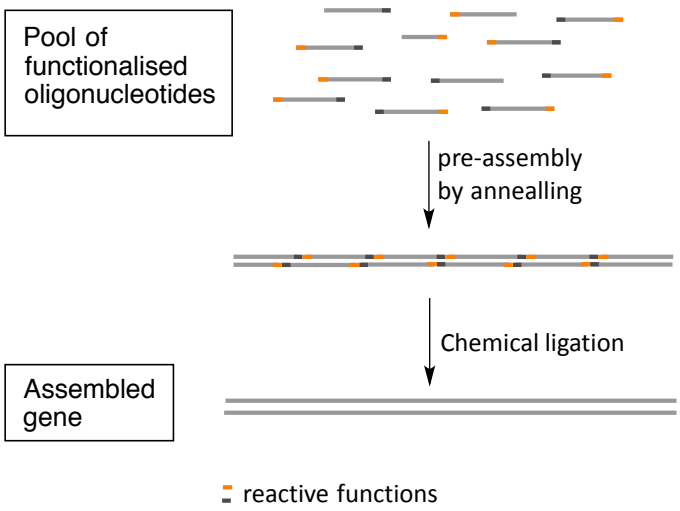


Figure 1-21: Concept of chemical-mediated gene assembly

1.3.2.3 DNA ligation through phosphodiester backbone

In the first methods of chemical DNA ligation, oligonucleotides were joined using a water-soluble condensing agent that activated the terminal phosphate group and facilitated attack of the hydroxyl functional group on the other end. Carbodiimides, such as 1,1′-carbonyldiimidazole or 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) have been first reported as condensing

reagents.[98,99] The reactions had to be performed in presence of a template that hold the two reactive extremities in close proximity (*Figure 1-22*).

Although the ligation reaction was more efficient when the phosphate group was in the 3'-position and the hydroxyl group in the 5'-position, this method suffered from several drawbacks. First, the reaction was extremely slow and usually required 6 to 8 days to reach completion. Second the use of carbodiimides usually induced the formation of by-products by reacting with thymine and guanine bases.[100] These detrimental results were greatly improved by the use of cyanogen bromide as condensing agent.[101] The reaction indeed proceeded within one to two minutes at 0 °C with 96% yield of ligated products. Using these conditions, up to 15 multiple ligations of a decanucleotide could be obtained.[98] A similar approach using *N*-cyanoimidazole and $ZnCl_2$, was also employed to ligate a 3.7 kb blunt-ended linear DNA duplex.[102]

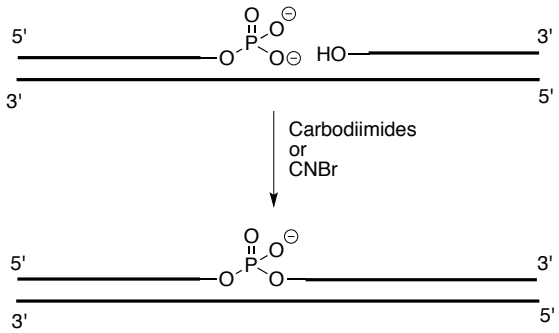


Figure 1-22: Chemical ligation of oligonucleotides to form a phosphodiester backbone

1.3.2.4 DNA ligation through modified backbones

The phosphoramidate linkage was one of the first modified internucleosidic bonds employed for oligonucleotide ligation. This phosphoramidate modified backbone was obtained *via* reaction between a terminal amino group and a phosphate group using carbodiimide or cyanogen bromide as coupling reagent (*Figure 1-23*). [101,103] As for the chemical formation of the phosphodiester bridge,

ligation using a 5'-amino group was more efficient than ligation using a 3'-amino group.

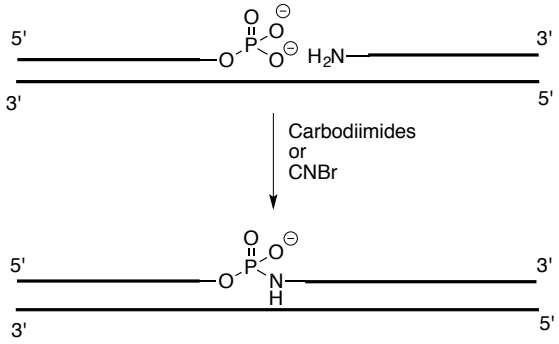


Figure 1-23: Templated-ligation of oligonucleotides through N5'-P3' phosphoramidate backbone

The phosphoramidate linkage is more stable to nucleases degradation compared to its natural counterpart, and it was previously demonstrated that DNA polymerases could accept 5'-amino-dideoxynucleoside triphosphate as substrate.[104] Further enzymatic studies, including replication through this modified backbone, would be interesting for its potential use in chemical gene synthesis.

Letsinger and co-workers have also joined oligonucleotides terminated with a phosphorothioate or bromoacetyl group to oligonucleotides containing a phosphorothioate extremity.[105,106] Although the reaction proceeded effectively and exclusively in the presence of a template, the extended internucleosidic bridges formed could be disadvantageous (*Figure 1-24*); they would cause significant distortion in the DNA duplex and might not be accepted by DNA polymerases.

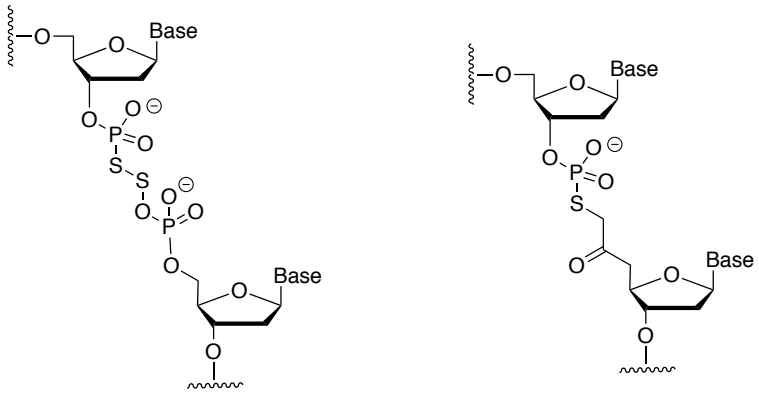
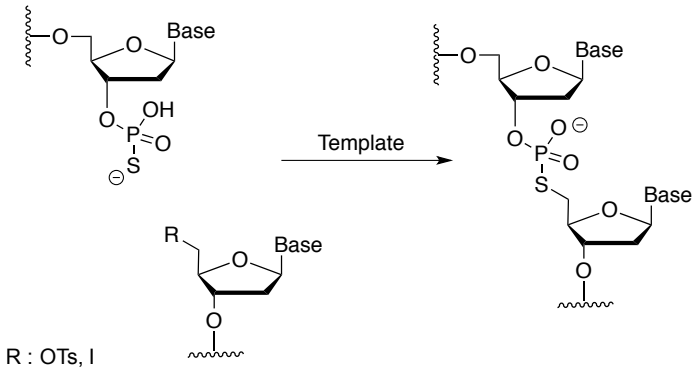


Figure 1-24: First backbones used in Letsinger's oligonucleotides chemical ligation

Letsinger *et al* circumvented this last drawback with the introduction of a bridging 5'-*S*-phosphorothioester linkage, obtained by the reaction of a phosphorothioate on a terminal 5'-iodo or 5'-tosylate group (*Scheme 1-5*).[107] Xu and co-workers later demonstrated that oligonucleotide templates containing this modified backbone were successfully read through by Klenow DNA polymerase without apparent pausing at the modified linkage.[108] However the chemical ligation approach proposed by Letsinger would hardly be applicable to chemical gene synthesis. First, the 5'-iodo or 5'-tosylate oligonucleotide extremities are not very stable chemical functions, which complicates their synthesis. More importantly, there is no efficient mean to trigger selective ligation only when all participating oligonucleotides are correctly assembled.



Scheme 1-5: Templated ligation to make a phosphorothioate backbone

Chapter 2

Synthesis and properties of amine and *N*-functionalised amine modified oligonucleotide analogues

This chapter begins by presenting previous reports of amine-containing oligonucleotide analogues followed by a description of molecules and functional groups that stabilise DNA duplexes through non-covalent interactions. The synthesis of oligonucleotides containing an artificial amine backbone is then described, with the development of a functionalisation strategy to access pyrene, anthraquinone and guanidine conjugated oligonucleotide analogues. UV melting studies of the latter are presented with a discussion of the influence of the different modifications on the DNA duplex stability. This chapter also include the synthesis of a thymidine phosphoramidite dimer with an amide internucleoside linkage for use in oligonucleotide synthesis. This backbone could be a useful addition to our current pool of oligonucleotide artificial backbones and can potentially be used in chemical-based gene assembly applications.

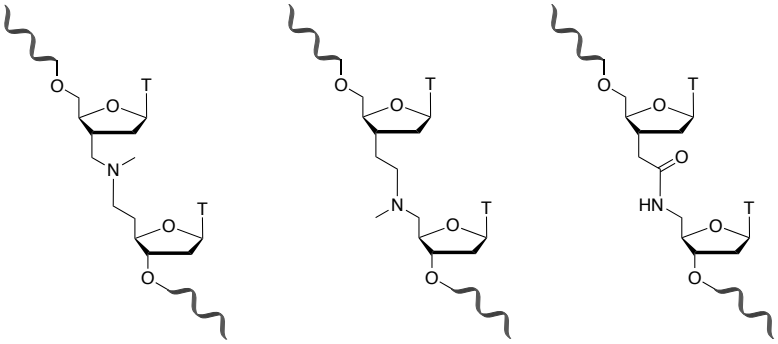
2.1 Introduction

2.1.1 Previous study on oligonucleotides containing amine modified DNA backbones

Oligonucleotides containing amine-modified internucleosidic linkages were first studied by De Mesmaeker and co-workers.[109] *N*-methyl functionalised amine and amide thymidine dinucleotides were synthesised and incorporated into a range of DNA sequences **I** to **III** (*Table 2-1*) using phosphoramidite chemistry. The thermal stability of these modified single-stranded oligonucleotides with natural RNA or DNA complements was then determined and compared to corresponding unmodified duplexes.

Table 2-1: Comparison of T_m values of *N*-methylamine and amide duplexes formed with DNA or RNA complements. X represents the incorporation site of *N*-methylamine and amide modified backbone **2.I** to **2.III**. [a]ΔTm per modifications = T_m(modified duplex) - T_m(unmodified duplex)/number of modifications. From De Mesmaeker *et al.*[109]

			ΔTm per modifications (°C)[a]		
	Sequences (5' to 3')	complementary strand	Backbone 2.I	Backbone 2.II	Backbone 2.III
I	CTCGTACCTxTTCCGGTCC	RNA	-2.5	-3.8	+0.9
II	GCGTxTTxTTxTTxTTxTGCG	RNA	-2.6	-3.3	-0.3
III	GCGTxTTxTTxTTxTTxTGCG	DNA	-3.8	-4.4	-1.5



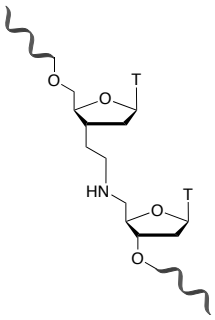
X: **2.I** **2.II** **2.III**

It was observed that a single addition of both tertiary amine modifications (**2.I** and **2.II**) in DNA:RNA duplexes of sequence **I** was thermally destabilising (*Table 2-1*).

A similar trend was observed for duplexes of sequences **II** and **III** with five additions of each amine-modified internucleosidic linkages. Although it should be noted that DNA:DNA and DNA:RNA duplexes formed using oligonucleotide sequences **II** and **III** are relatively unstable as they contain ten consecutive thymidines. When five alternating *N*-methylamine modifications **2.I** and **2.II** were present, duplexes with DNA and RNA complements could still be formed with an average drop in melting temperature of *ca.* 15 and 20 °C for DNA:RNA and DNA:DNA duplexes respectively. The greater destabilisation of DNA:DNA duplexes compared to DNA:RNA duplexes suggests that duplexes containing these tertiary amine modifications might preferentially adopt an A-type helix (*Table 2-1*).

Interestingly, the thermal stability of *N*-methylamine backbone **2.I** was slightly higher than its structural isomer **2.II**, implying that the stabilisation depends on the position of the heteroatom within the linkage. De Mesmaeker *et al.* attributed the thermal instability of amine backbones **2.I** and **2.II**, to the increased conformational flexibility of the linkages compared to the natural phosphodiester counterpart. This hypothesis was supported through direct comparison with the more rigid amide linkage **2.III**, which in fact has a positive effect on the thermal stability in the case of a single incorporation within the duplex (*Table 2-1*)

In order to further explain the loss in thermal stability induced by the amine backbone, Lynn *et al.* later studied a hexameric DNA duplex containing one central unalkylated secondary amine internucleosidic linkage (backbone **2.IV**, *Figure 2-1*).[110] In this case, the decrease in thermal stability was of 17 °C compared to the unmodified DNA duplex. Measurement of the thermodynamic parameters showed an unfavourable change in ΔH of 7 kcal.mol^{-1} and a favourable increase in ΔS of 15 eu.mol^{-1}. The former provides evidence that the global structure of the duplex is destabilised compared to its natural counterpart. The latter was postulated to be due to a change in water molecules binding to the modified backbone. This is in accordance with the fact that the ethylamine is more hydrophobic than the natural phosphodiester linkage and interacts less with water molecules.



2.IV

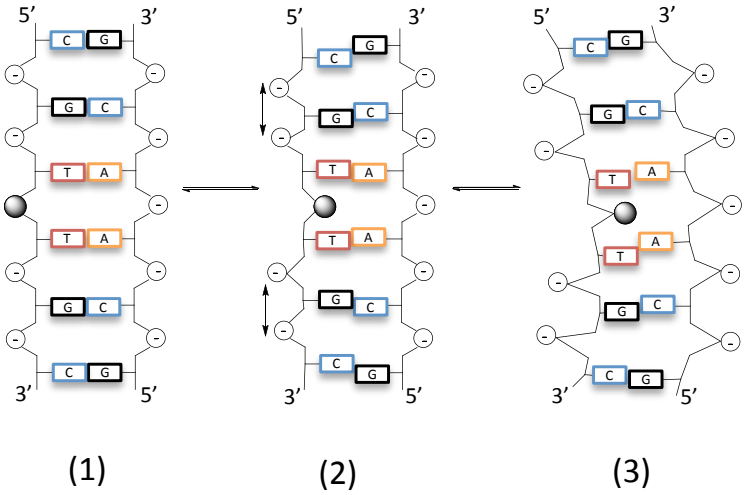


Figure 2-1: Model proposed by Lynn *et al.* for the destabilisation induced by the amine modification **2.IV** (grey ball) in a DNA duplex of sequence CGTTGC.[110] Phosphodiester linkages are represented by circled negative charges. The amine linkage collapses into the hydrophobic interior of the duplex. The two flanking phosphodiesters are then forced to reduce the distance separating them, causing backbone bending and loss of optimal H-bonding and π-stacking interactions.

NMR structural analysis showed that the modified duplex formed a B-DNA helix with canonical base pairing. However, significant interactions of the ethylamine linkage with the hydrophobic core of the duplex formed by the sugar rings and the base edges could be observed. It was therefore hypothesised that the destabilisation induced by the amine modification is due to the collapse of the flexible linkage into the interior of the duplex (*Figure 2-1*, (**1**)). This movement accentuates the gap between the phosphodiester groups flanking the amine modification (*Figure 2-1*, (**2**)), forcing a DNA curvature to shorten the distance separating these phosphodiester groups. This bending disrupts the optimum

geometry for base stacking and H-bonding, thus lowering the overall thermal stability of the duplex (*Figure 2-1*, (**3**)).

2.1.2 Increasing the thermal stability of nucleic acid duplexes

A broad range of chemical species such as small organic molecules, ions and metal complexes can reversibly interact with nucleic acids. These reversible interactions play an important role in nucleic acids configuration and generally stabilise the duplex structure. Their mode of interaction can generally be divided into three categories, namely groove-binders, intercalators and molecules that bind through electrostatic interactions.

2.1.2.1 Groove-binding molecules

These molecules bind to nucleic acids by interacting with the edges of base pairs in either of the major or minor groove of the duplex. Proteins usually exhibit a binding specificity for the major groove of DNA while small organic molecules preferentially bind to the minor groove.[111]

These molecules are usually composed of several aromatic rings such as pyrrole, furan, benzimidazole and indole derivatives, linked by bonds with torsional freedom. Most groove binders *e.g.* netropsin, distamycin or DAPI (*Figure 2-2*) can, with the appropriate twist, fit tightly in the helical curve of the DNA minor groove, form hydrogen bonds and electrostatic interactions and effectively increase the duplex thermal stability (*Figure 2-3*).[112,113] Minor groove-binders such as distamycin derivatives[114] and Hoechst 33258[115] (*Figure 2-2*) have been covalently attached to DNA in order to achieve greater stabilisation of duplex structures. Bruice and co-workers have also described the conjugation of Hoechst 33258 to a pentameric thymidine oligonucleotide where all phosphodiester linkages have been replaced by a guanidine backbone.[116] This tethered deoxyribonucleic guanidine (DNG) enhanced the stability of a triplex formed with a 30-mer dsDNA by 13 °C compared to the triplex formed with an untethered DNG. The stability of a duplex formed between DNG and a ssDNA was also improved through the binding of the Hoechst 33258 derivative in the minor groove of the duplex. These types of conjugates, combining positively charged and nuclease resistant modified backbone with a cell penetrating

DNA groove-binder, are interesting conceptually and potentially useful for antisense or antigene applications.[117]

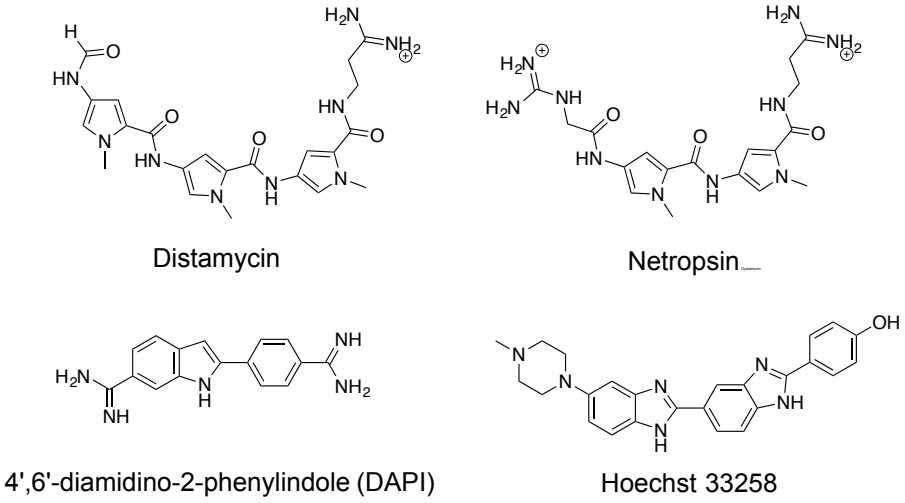


Figure 2-2: Structure of four DNA minor groove-binders

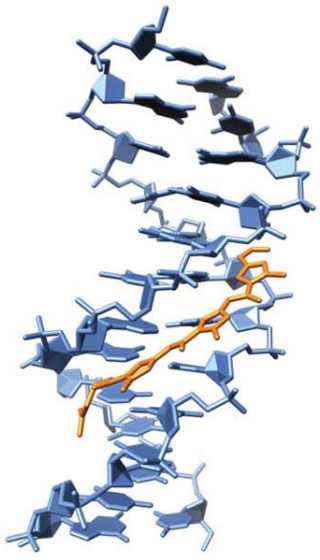


Figure 2-3: X-ray structure of the minor groove binder distamycin (orange) in complex with the DNA dodecamer d(CGCAAATTTGCG) (blue).[118]

2.1.2.2 DNA intercalating molecules

Lerman and co-workers first introduced the principle of "intercalation" to describe the interaction of planar aromatic molecules with DNA (*Figure 2-4*).[119] Unlike groove binding described above, intercalation requires changes in the nucleic acid duplex structure. The sugar-phosphate torsional angles have to be modified in order to locally unwind adjacent base pairs and allow insertion of the intercalating agent.[120] Generation of the intercalation site usually induces a lengthening of the double helix and could also cause other distortions such as bending.[121]

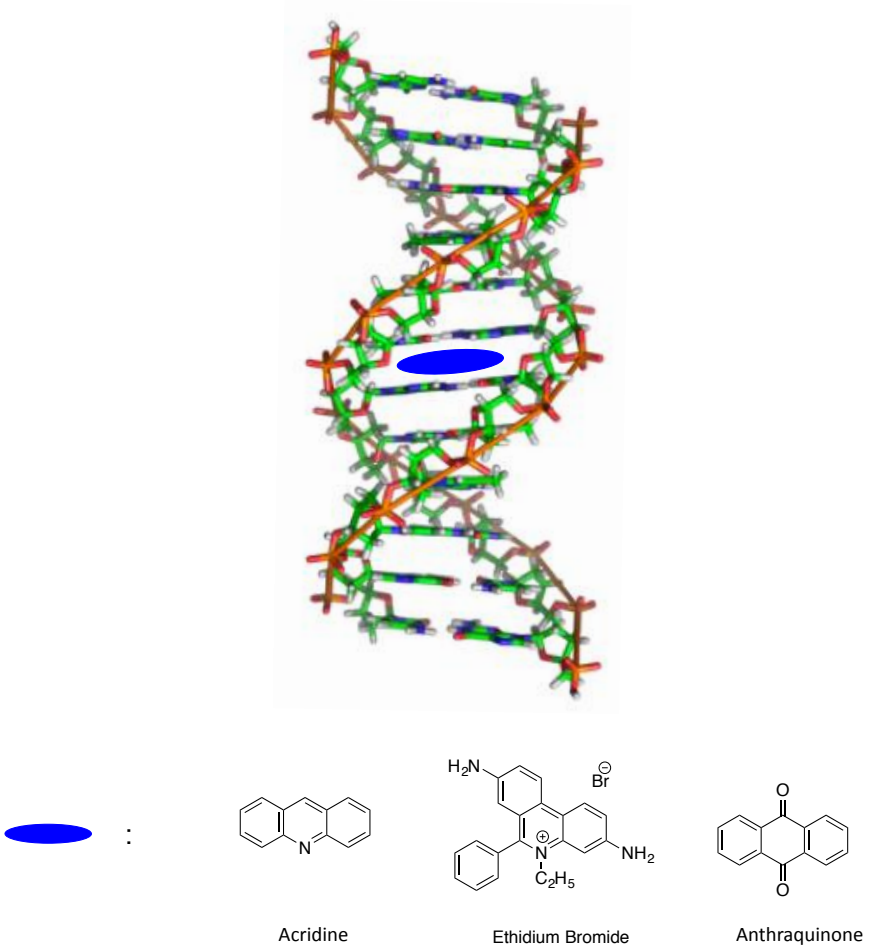


Figure 2-4: Intercalation of a non-exhaustive set of examples of aromatic molecules

Furthermore, due to their aromaticity, intercalators such as ethidium bromide or SYBR dyes, often display interesting fluorescent properties, which have been widely applied to aid nucleic acid visualisation.[122] Another interesting application of these molecules resides in their ability to stabilise the duplex structure, partly by π-stacking interactions with the adjacent base pairs. Hélène and co-workers were one of the first to demonstrate that it was feasible to covalently attach an acridine intercalator to a DNA strand. These conjugated oligonucleotides displayed an enhanced binding affinity for their complementary strands and could effectively be used to inhibit translation of gene 32-encoded mRNA from bacteriophage T4. This gene codes for a protein that plays a crucial role in the phage DNA replication, recombination, and repair.[123]

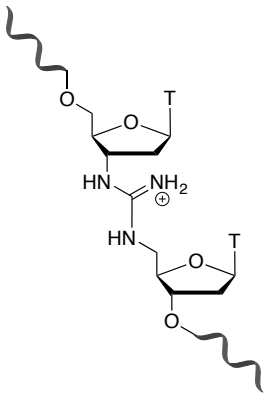
2.1.2.3 Electrostatic interactions and the stability of nucleic acid duplexes

In the past decades, zwitterionic or positively charged oligonucleotides have been designed and used to form more stable constructs with their negatively charged natural complementary DNA or RNA strand by electrostatic interactions.[124–126] In particular, positively charged modifications have been introduced into oligonucleotide backbones for the synthesis of more potent antisense agents.

Bruice *et al.* have reported the solid phase synthesis of oligonucleotide analogues in which all phosphodiester linkages were replaced by positively charged guanidinium moieties (deoxyribonucleic guanidine **DNG**) **2.V** (*Figure 2-5*).[127,128] Octameric and pentameric thymidine deoxyribonucleic guanidine (DNG-T_8 and DNG-T_5 respectively) formed triplexes (DNG:DNA:DNG) upon annealing with an unmodified octameric or pentameric deoxyadenosyl phosphodiester strand (DNA-A_8 and DNA-A_5 respectively). The DNG-T_8:DNA-A_8:DNG-T_8 triplex displayed a considerably improved thermal stability compared to its triplex unmodified counterpart ($\Delta T_m \approx +60$ °C). This is probably due to the opposite charge attraction between the positively charged DNG strands and the anionic DNA strand. To examine whether the high affinity of DNG for a complementary DNA strand was not detrimental to the base pairing specificity, the influence of mismatches in the DNA target was investigated.[127] It is indeed important for the design of DNG antisense or siRNA oligonucleotides, to determine if the primary binding mode of these

analogues to a DNA or RNA target occurs through base-pairing. This reduces the risk of off-target hybridisation, which is related to *in vivo* toxicity.[129] To investigate the fidelity of base recognition of modified guanidine DNA analogues, UV-melting studies on DNG:DNA:DNG triplexes containing mismatched base pairs have been carried out. Mismatched bases at the ends of the triplex induce a decrease in melting temperature of 8 °C, whereas a single mismatch at the centre causes a drop of 15 °C. However, comparison of these data with unmodified phosphodiester DNA triplexes were not provided and it is difficult to conclude if DNG base pairing specificity is as high as natural oligonucleotides.

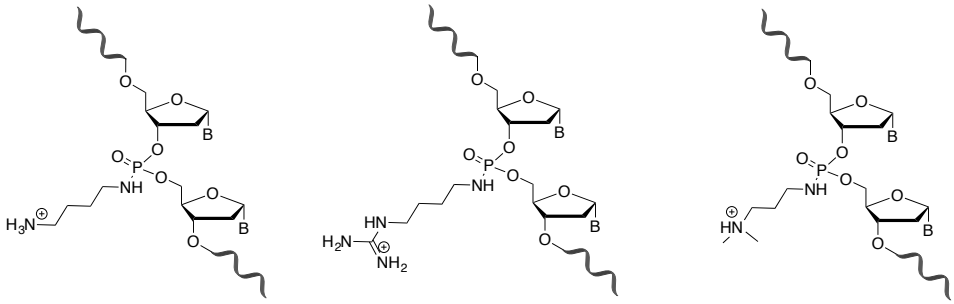
The stabilising effect of the guanidinium-modified backbone **2.V** was far less pronounced when introduced in oligonucleotides to form mixed-backbones (phosphodiester/guanidine) DNA chimeras. The guanidinium modification was in fact destabilising at physiologically relevant NaCl concentration ($\Delta T_m \approx -2$ °C).[130] Moreover, The preparation of modified nucleosides (A, C, T and 7-deaza-G) required for the synthesis of fully modified or chimeric DNA/DNG oligonucleotides is lengthy,[128] involves the use of toxic reagent ($HgCl_2$) and formation of potentially toxic byproducts (mercury sulfide),[131] which can present another limitation in the use of DNG as therapeutic oligonucleotide analogues (antisense, siRNA).



2.V

Figure 2-5: Structure of guanidine modified oligodeoxyribonucleotides

Vasseur and co-workers later developed an efficient strategy for the guanylation of α-oligonucleotides containing amino modifications (**α-ODN-1**) to access new α-guanidinobutylphosphoramidate oligonucleotides **α-ODN-2** (*Figure 2-6*).[132] They previously demonstrated that α-anomeric oligonucleotides are nuclease resistant and form parallel-stranded duplexes with complementary phosphodiester β-anomeric DNA and RNA strands. It was observed that the combination of the inversion of the anomeric configuration of the sugar (from β to α) with backbone modifications (methylphosphonate, phosphoramidate, and N-alkylphosphoramidates) generally increases the binding affinity for a DNA or RNA target.



α-ODN-1 **α-ODN-2** **α-ODN-3**

Figure 2-6: Structures of cationic modified oligonucleotide backbones designed by Vasseur *et al*. Sequence of modified oligonucleotides: TCTTAACCCACA. All phosphodiester linkages were replaced by the corresponding modifications. All bases were in α-configuration.[132]

The hybridisation properties of fully modified oligonucleotides **α-ODN-1** and **α-ODN-2** were compared to the previously synthesised **α-ODN-3** (*Figure 2-6*) and to natural phosphodiester oligonucleotides. All positively charged backbones formed much more stable duplexes compared to the natural phosphodiester β-anomeric duplex, especially in the case of amino modified **α-ODN-1** and guanidine modified **α-ODN-2** which displayed an increased stability of 29°C. Molecular dynamics simulations confirmed that electrostatic interactions between the cationic backbone of the α-oligonucleotides and the anionic backbone of the nucleic acid targets were

partially involved in the huge stabilisation observed. Another advantage of these positively charged backbones resides in the fact that they can improve cell penetration of therapeutic oligonucleotides. This was observed in the case of guanidinium **α-ODN-2** for which the cellular uptake was six time higher than negatively charged phosphorothioate oligonucleotides.[133]

2.2 Aim of the study

Although the amine internucleosidic linkage is duplex-destabilising, the nucleophilicity of the group makes it an interesting and versatile functional handle that could expand the scope of oligonucleotide-modified backbones. We therefore envisaged that it could be exploited to access novel oligonucleotide backbones that might display improved physicochemical properties compared to the natural phosphodiester backbone (*Figure 2-7*, A). It was consequently decided to functionalise the secondary amine with chemical groups that are known for their nucleic acid duplex stabilisation properties. The synthesis and hybridisation studies of *N*-substituted amine backbones bearing positively charged or intercalating modifications will be presented in the following sections.

A thymidine phosphoramidite dimer containing an amide internucleosidic linkage was also synthesised and was used for oligonucleotide synthesis (*Figure 2-7*, B). The aim was to investigate the fidelity of replication through a DNA template featuring an amide-modified backbone by four DNA polymerases (KOD XL, GoTaq, Klenow large fragment, 9 °N). These results will be described in *Chapter 4*.

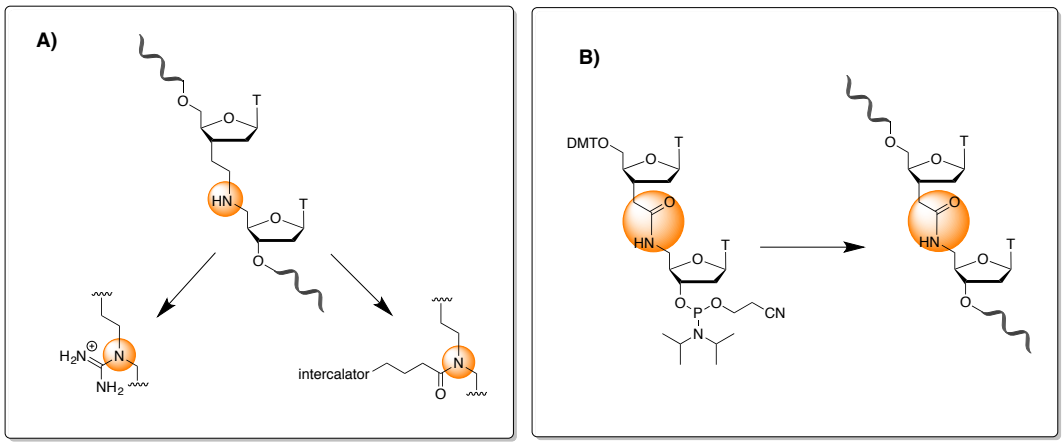
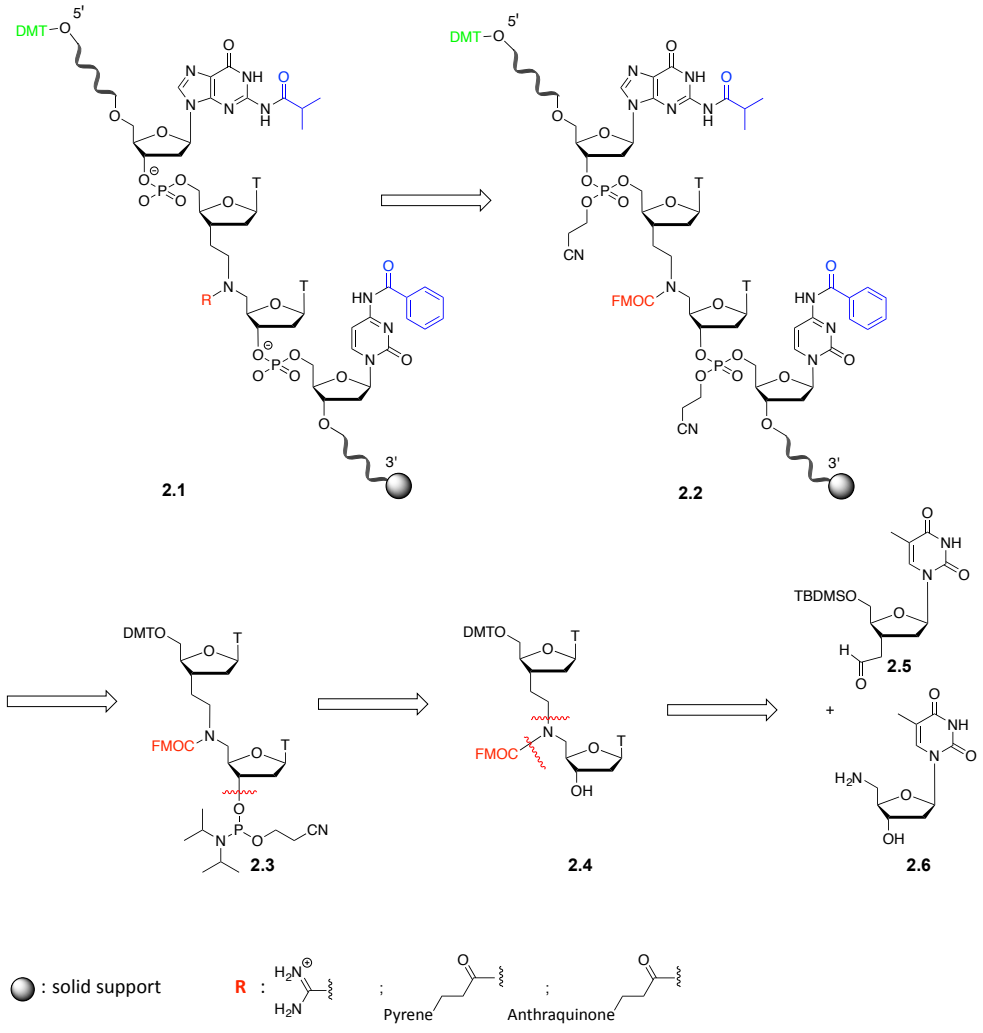


Figure 2-7: A) Functionalisation of secondary amine to access novel modified DNA backbones. B) Synthesis of amide T-T phosphoramidite dimer for oligonucleotide synthesis.

2.3 Synthesis of amine and *N*-functionalised amine oligonucleotide analogues

2.3.1 Retrosynthetic strategy for the synthesis of *N*-substituted amine modified backbones

Targeted oligonucleotide analogues **2.1** were obtained after deprotection and functionalisation of the secondary amine of oligonucleotides **2.2**. These steps were performed directly on the solid support to allow easy removal of any excess reagents. Protected amine oligonucleotides **2.2** were obtained *via* automated solid phase oligonucleotide synthesis using the key thymidine dimer phosphoramidite **2.3**, which was accessed by phosphitylation of intermediate **2.4.** This precursor was synthesised *via* reductive amination between the 3′-oxoethyl nucleoside synthon **2.5** and the 5′-amino nucleoside synton **2.6** (*Scheme 2-1*).

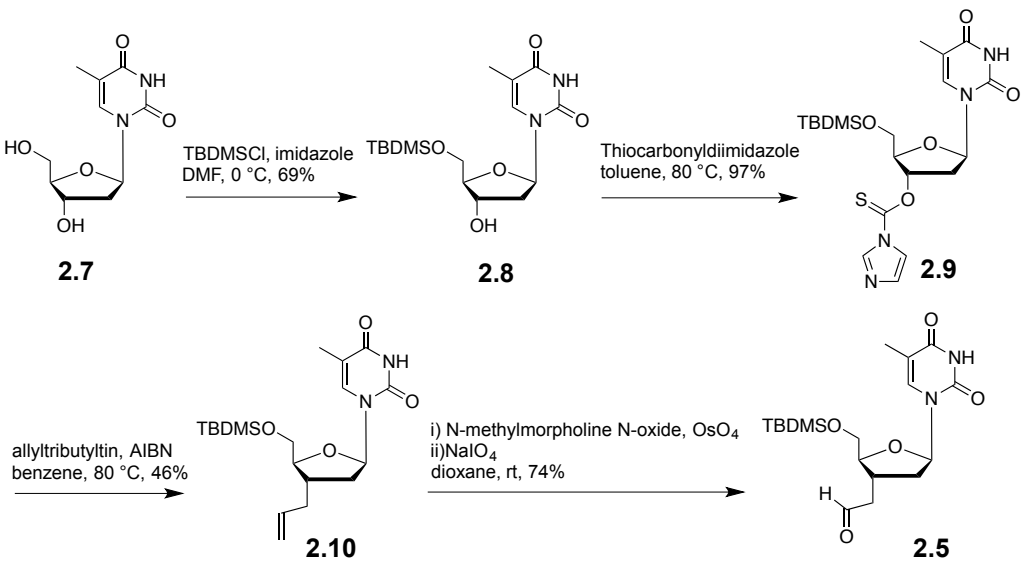


Scheme 2-1: Retrosynthetic strategy to access amine modified oligonucleotides

The secondary amine of the backbone must be protected to prevent side reactions during oligonucleotide synthesis. The FMOC protecting group is a sensible choice, as it is stable to the conditions used for oligonucleotide synthesis. It can also be selectively cleaved in presence of DMT, isobutyryl and benzoyl protecting groups, thus allowing selective functionalisation of the secondary amine linkage (*Scheme 2-1*).

2.3.2 Synthesis of the 3′-oxoethyl thymidine nucleoside synthon

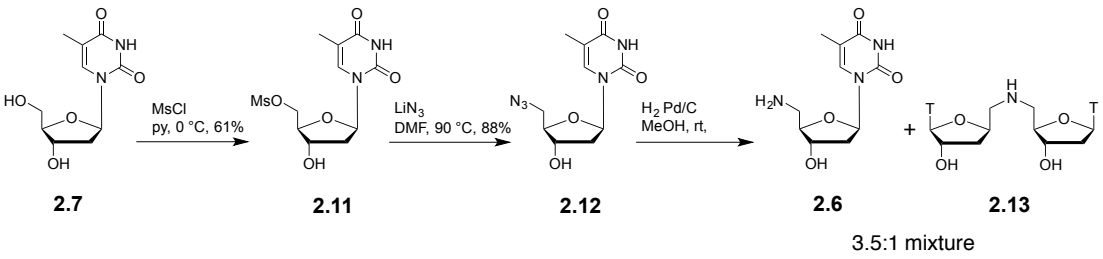
The 3′-oxoethyl nucleoside **2.5** was obtained following the procedure of Fiandor and co-workers.[134] The 5′-hydroxyl group of thymidine **2.7** was first protected as a *tert*-butyldimethylsilyl ether in 69% yield. Subsequent treatment of nucleoside **2.8** with thiocarbonyldiimidazole gave the 3′-thioacylimidazole nucleoside **2.9** in an excellent yield of 97%. The radical allylation step was then achieved using AIBN and allyltributylstannane as initiator, giving the expected 3′-allyl nucleoside **2.10** in a moderate yield of 46%. Following this, nucleoside **2.10** was dihydroxylated with a catalytic amount of osmium tetroxide and the resulting diol subsequently treated with sodium periodate to give the final 3′-oxoethyl thymidine nucleoside **2.5** in 74% yield (*Scheme 2-2*).



Scheme 2-2: Synthesis of 3′-oxoethyl nucleoside **2.5**

2.3.3 Synthesis of the 5'-amino-5'-deoxythymidine nucleoside synthon

The synthesis of the targeted 5'-amino nucleoside **2.6** was achieved as described in *Scheme 2-3* following published procedures.[135] In the first step, thymidine **2.7** was converted to the 5'-mesylate intermediate **2.11** in 61% yield. Then, treatment of nucleoside **2.11** with lithium azide enabled formation of 5'-azido nucleoside **2.12** in an excellent yield of 88%. This precursor was then reduced *via* a hydrogenation to give a 3.5:1 mixture of the desired 5'-amino-5'-deoxythymidine nucleoside **2.6** and secondary amine dimer impurity **2.13**[136], which was used in the next synthetic step without further purification.

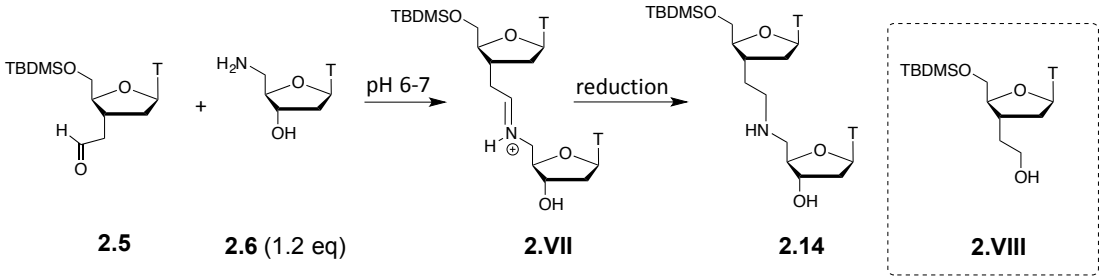


Scheme 2-3: Synthesis of 5'-amino-5'-deoxythymidine nucleoside **2.6**

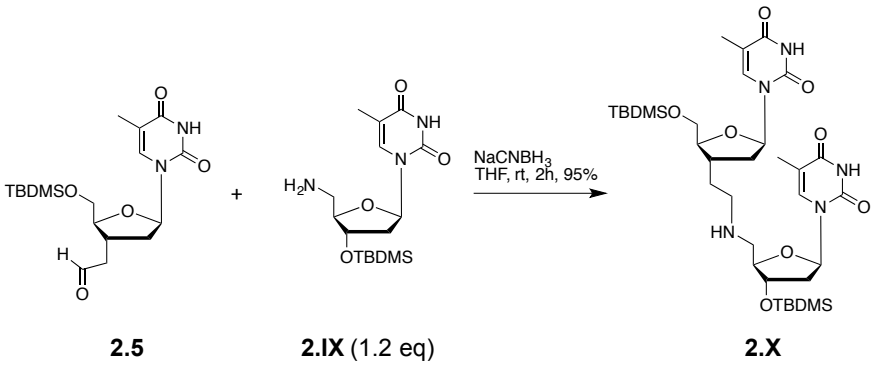
2.3.4 Synthesis of T-T dimer phosphoramidite containing amine backbone and modified oligonucleotides

With 3'-oxoethyl and 5'-amino thymidine nucleosides **2.5** and **2.6** in hand, the key reductive amination step was investigated (*Table 2-2*). The reaction involves the formation of the iminium intermediate **2.VII**, whose formation requires slightly acidic conditions (pH 6 – 7).[137] The use of sodium cyanoborohydride as reducing agent was initially examined, since it was used by Lynn and co-workers for the reductive amination reaction of similar substrates (*Scheme 2-4*).[110]

Table 2-2: Optimisation of reductive amination reaction for synthesis of 5'-*O*-TBDMS thymidine amine dimer **2.14.** [a]Isolated yield.



Entry	Reducing agent	Solvent	reaction time (h)	Yield (%)
1	$NaCNBH_3$ (1.5eq)	THF	2	traces
2	$NaCNBH_3$ (3 eq)	THF	2	16[a]
3	$NaBH(OAc)_3$ (2 eq)	DCE	12	0
4	$NaBH(OAc)_3$ (2 eq)	Isopropanol	2	83[a]

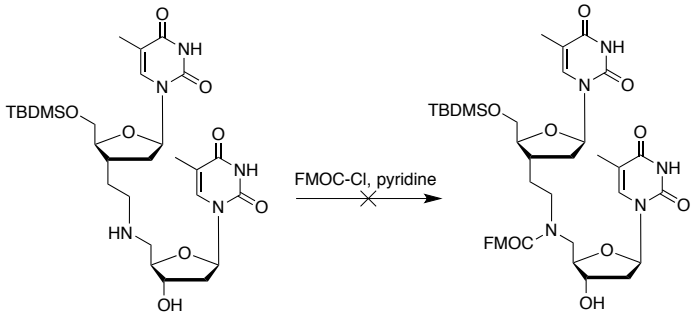


Scheme 2-4: Synthesis of Bis-*O*-TBDMS-5'-thymidinylaminoethylthymidine **2.X** described by Lynn *et al.*[110]

Surprisingly applying these conditions on the 5'-amino thymidine nucleoside **2.6**, which does not contain a silyl protecting group on the 3'-hydroxyl group, led to the recovery of the starting materials (*Table 2-2*, Entry 1). Increasing the excess of $NaCNBH_3$ and lowering the pH of the reaction enabled isolation of the desired amine T-T dinucleoside **2.14** in a low yield (Entry 2). Given the poor yield obtained in these conditions, the use of sodium triacetoxyborohydride as an alternative reducing agent and DCE as solvent was envisaged, as described by Abdel-Magid *et al* (Entry 3).[138] However, this led to the almost total reduction of 5'-aldehyde nucleoside **2.5**

to the 5'-*O*-TBDMS-3'-deoxythymidine-3'-ethanol **2.VIII**, and no formation of the desired product **2.14** could be observed (Entry 3). Assuming that the previous reaction outcome was due to a low solubility of the 5'-amino nucleoside **2.6** in DCE, the reductive amination was attempted in isopropanol, keeping the same reducing agent (Entry 4). This time, the formation of the desired secondary amine **2.14** could be observed by ^{1}H NMR, and was isolated in excellent yield by column chromatography (83 %, Entry 4).

The subsequent FMOC protection of the secondary amine of T-T dinucleoside **2.14** proved to be cumbersome as formation of the desired compound **2.15** could not be observed by TLC (*Scheme 2-5*). This could be explained by the proximity of the bulky TBDMS protecting group, hindering nucleophilic attack of the secondary amine on FMOC-Cl.

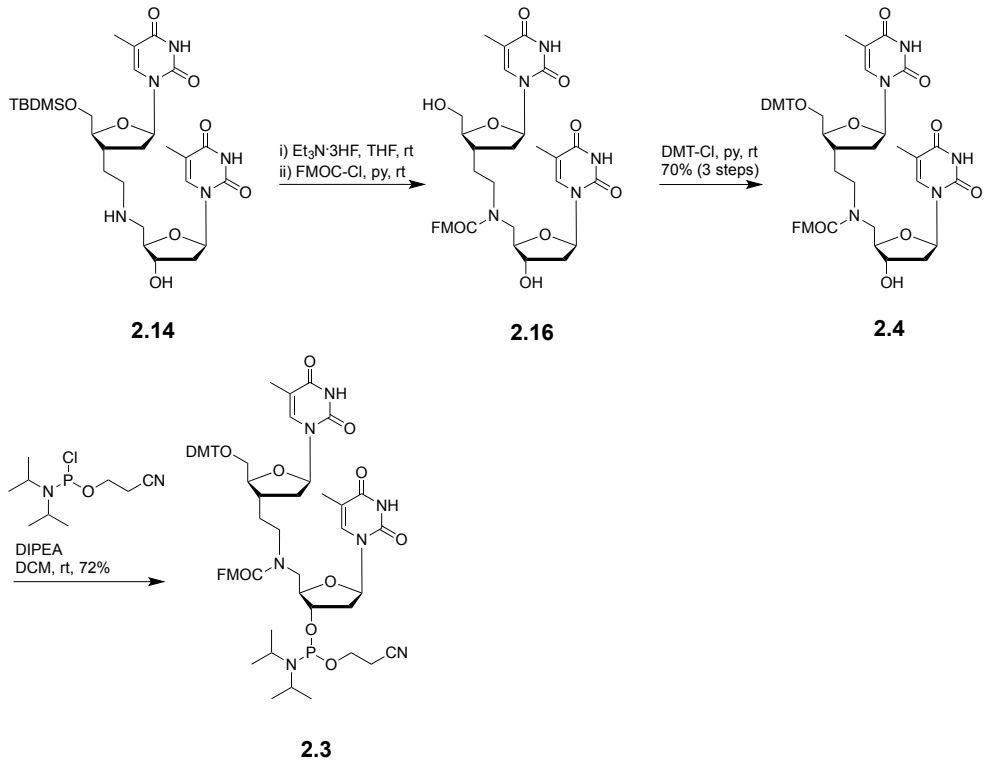


2.14 **2.15**

Scheme 2-5: Attempt of FMOC protection of the secondary amine of 5'-*O*-TBDMS thymidine amine dimer **2.14**

Based on these results, another approach to access the targeted FMOC-amine T-T dimer phosphoramidite **2.3** was developed (*Scheme 2-6*). The TBDMS protecting group of amine thymidine dinucleoside **2.14** was first cleaved off by treatment with triethylamine trihydrofluoride and the resulting intermediate was immediately protected with FMOC-Cl. The crude mixture **2.16** was then treated with DMT-Cl, which enabled isolation of intermediate **2.4** in 70% yield over 3 steps. Following this, the final phosphitylation step proceeded successfully to afford the

targeted T-T phosphoramidite dimer **2.3** in 72% yield.

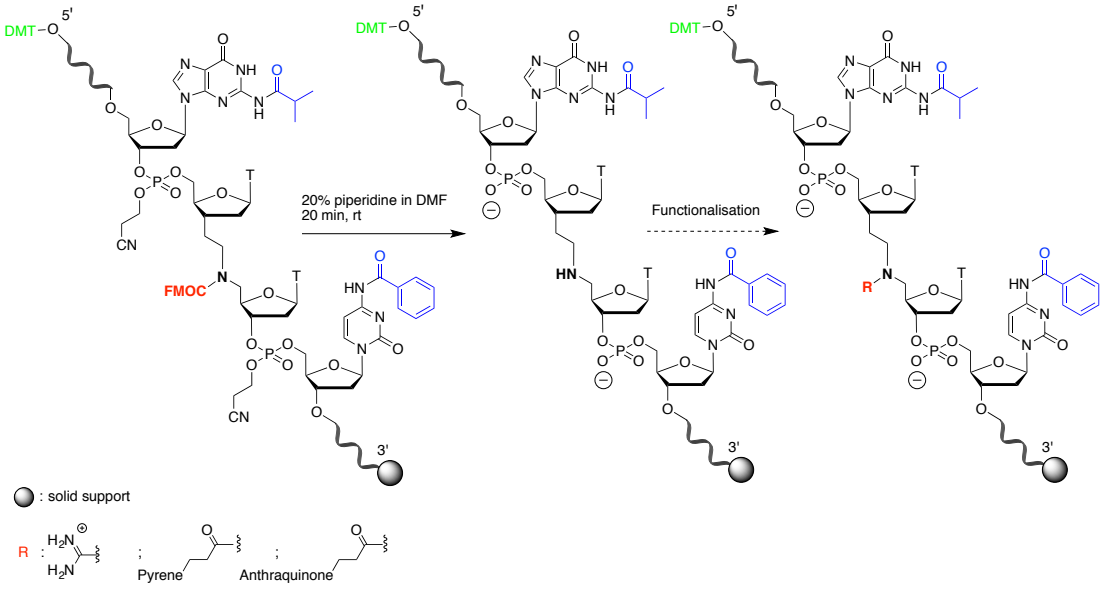


Scheme 2-6: Synthesis of T-T phosphoramidite dimer containing FMOC-amine backbone

Amine modified oligonucleotides of different sequences (**ODN-2.1** to **ODN-2.3**), were then prepared by standard solid phase oligonucleotide synthesis, using phosphoramidite dimer **2.3** (*Table 2-3*). These oligonucleotides were then treated on the CPG solid support with piperidine to selectively cleave the FMOC protecting group in presence of the 5'-DMT and nucleobases protecting groups, and allow functionalisation with pyrene, anthraquinone and guanidinium motifs (*Scheme 2-7*).

Table 2-3: Sequences of FMOC protected amine-modified oligonucleotides. These oligonucleotides were deprotected on the solid support to afford secondary amine-modified oligonucleotides. *a* represents the positions that have been replaced by the secondary amine internucleosidic linkage (*Scheme 2-7*).

code	Sequences (5' to 3')
ODN-2.1	CGTTCGCT$_a$TGCAGC
ODN-2.2	ACCCATTCCT$_a$TCGTTCCACT
ODN-2.3	ACCCAT$_a$TCCTTCGT$_a$TCCACT

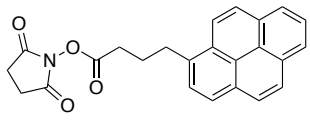


Scheme 2-7: Post-oligonucleotide synthesis deprotection of the modified internucleosidic linkage secondary amine.

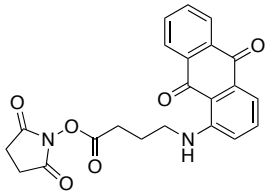
2.3.5 Optimisation of the solid phase synthesis of pyrene and anthraquinone modified oligonucleotides

Functionalisation of the secondary amine internucleosidic linkage of oligonucleotides **ODN-2.1** to **ODN-2.3** with pyrene and anthraquinone moieties was first optimised using oligonucleotide **ODN-2.1** (*Table 2-3*). These modifications were introduced through amide bond formation using the *N*-hydroxysuccinimide active

ester (NHS) of pyrene[a] and anthraquinone[b] intercalators (*Figure 2-8*). Active NHS esters have been extensively used for the post-synthetic labelling of oligonucleotides with fluorophores and intercalating dyes; this is commonly accomplished using reactive primary amino groups introduced at one of the extremities of the oligonucleotide (5' or 3'-ends) or on the nucleobases (reviewed by *Davies* et al.).[139] This conjugation strategy was chosen for the solid-phase functionalisation of the secondary amine artificial linkage of oligonucleotides **ODN-2.1** to **ODN-2.3** as it presents two main advantages. First, the amide bond formed is stable to the basic conditions used for oligonucleotide cleavage from the solid support and deprotection of the nucleobases (A, G and C). Second, given the diverse range of NHS esters available commercially or that have known synthetic routes,[140] this methodology could potentially greatly expand the oligonucleotide backbone diversity.



1-pyrenebutyric acid *N*-hydroxysuccinimide ester



N-(anthraquinonyl)amino butyric acid *N*-hydroxysuccinimide ester

Figure 2-8: Structure of NHS active esters of pyrene and anthraquinone

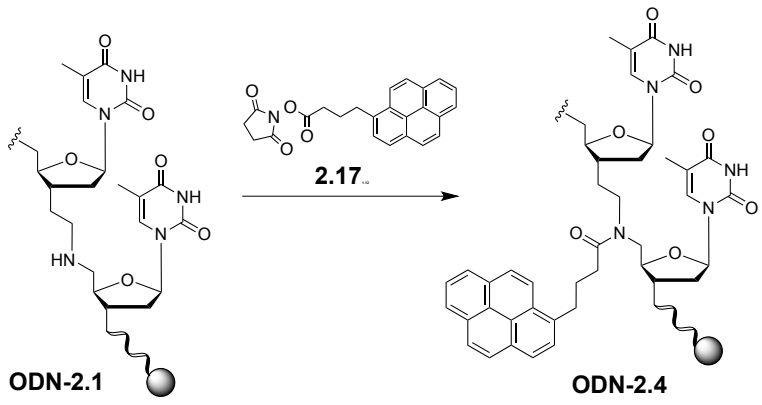
Oligonucleotide **ODN-2.1** was acylated directly on the CPG solid support, using a large excess of pyrene and anthraquinone active NHS esters (~80 eq) in order to quickly drive the reaction to completion. The acylation reactions were

[a] 1-pyrenebutyric acid *N*-hydroxysuccinimide ester was prepared by Anna Dysko.[193]

[b] The synthesis of N-(anthraquinonyl)amino butyric acid *N*-hydroxysuccinimide ester was achieved by Angus Danton (internship) under direct supervision following published procedures.[194]

carried out in the presence of an organic base to prevent unwanted 5'-DMT deprotection and avoid protonation of the backbone secondary amine.

Optimisation of the acylation conditions with pyrene NHS ester **2.17** (*Scheme 2-8*) is presented in *Table 2-4*. Performing the reaction using DMF as a solvent and triethylamine as a base (*Table 2-4*, Entry 2) instead of pyridine (*Table 2-4*, Entry 1) gave a better conversion although the reaction conversion was low (*Table 2-4*, Entry 1 and 2). This could be due to the fact that the active NHS ester has a higher solubility in DMF compared to pyridine. Next the effect of increasing the reaction temperature from 55 °C to 70 °C was considered; this did not have any effect on the efficiency of the reaction as showed in *Table 2-4*, Entry 3. Finally, the reaction time was increased from 5 to 20 hours, which significantly improved the reaction conversion up to 75% (*Table 2-4*, Entry 4). These optimised conditions were therefore used for the acylation of oligonucleotides **ODN-2.2** and **ODN-2.3** with pyrene NHS ester and anthraquinone NHS ester (*Chapter 6, Sections 6.2.2.2* and *6.2.2.3*). In the case of oligonucleotide **ODN-2.3** (*Table 2-3*), which contains two amine modifications, the desired diacylated oligonucleotide was obtained in addition to a minor amount of monoacylated oligonucleotides, which were successfully separated using RP-HPLC



Scheme 2-8: Acylation of oligonucleotide **ODN-2.1** with 1-pyrenebutyric acid N-hydroxysuccinimide ester.

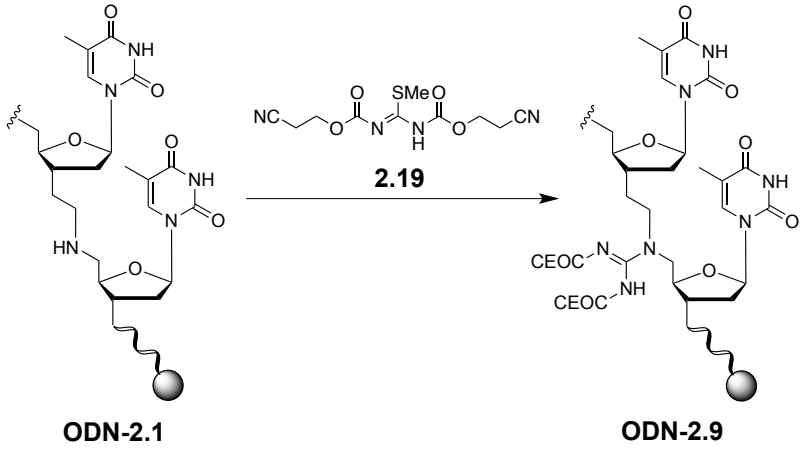
Table 2-4: Conditions used for the optimisation of acylation of **ODN-2.1** with pyrene NHS ester. [a]Estimated by HPLC-MS from the relative integrations of starting material and product after cleavage from the CPG resin.

Entry	Solvent/base	T (°C)	Reaction time (h)	Conversion (%)[a]
1	Pyridine	55	5	0
2	DMF/10% TEA	55	5	10
3	DMF/10% TEA	70	5	10
4	DMF/10% TEA	55	20	75

2.3.6 Optimisation of the solid phase synthesis of guanidinium modified oligonucleotides

Guanylating reagents are defined as compounds that are able to form a guanidine moiety *via* a chemical reaction with an amine functional group. Important classes of these reagents include thioureas, isothioureas, carbodiimides and pyrazole-1-caboxamidines, the use of which have been reviewed by Katritzky *et al.*[141] The guanylation of the secondary amine internucleosidic linkage was first optimised using oligonucleotide **ODN-2.1** coupled to the CPG solid support using *N*,*N*'-bis-(2-cyano)ethoxycarbonyl)-2-methyl-2-thiopseudourea **2.19** as a guanylating reagent [c] (*Scheme 2-9*). This reagent was previously used by Manoharan and co-workers for the synthesis of 2'-*O*-(2-guanidinium) ethyl modified oligonucleotides from 2'-aminoethoxy oligonucleotides.[124] The use of this guanylating reagent over other reagents presents several advantages. First, the two electron-withdrawing *N*-(2-cyano)ethoxycarbonyl (CEOC) groups on positions conjugated with the isothiourea carbon reactive centre increases the electrophilic character of the latter. This reagent is therefore more reactive than these without or with only one electron-withdrawing group, which is beneficial for the guanylation of the hindered secondary amine linkage of the modified oligonucleotide. The CEOC groups also enhance the solubility of the guanylating agent in the organic solvents (DMF) typically used for solid phase reactions. Finally these groups are labile in basic conditions and can be conveniently removed using standard oligonucleotides deprotection conditions.

[c] *N*,*N*'-bis-(2-cyano)ethoxycarbonyl)-2-methyl-2-thiopseudourea was previously synthesised by Simon Gerrard following the procedure of Manoharan *et al.*[124]



Scheme 2-9: guanylation of the secondary amine of oligonucleotide **ODN-2.1**

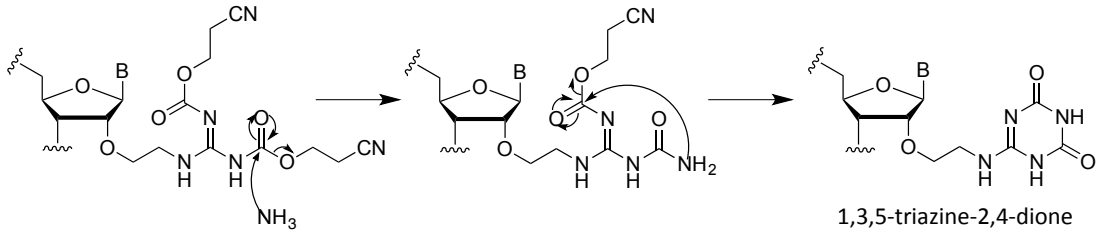
The guanylation reaction proved to be very efficient when performed at 55 °C for 20 hours, resulting in the almost total conversion of amine modified oligonucleotide **ODN-2.1** into the guanylated oligonucleotide product (*Table 2-5*, Entry 1). Furthermore, increasing the temperature from 55 to 70 °C enabled the reaction time to be decreased to 5 hours (*Table 2-5*, Entry 2). These conditions were consequently used for guanylation of amine-modified oligonucleotides **ODN-2.2** and **ODN-2.3**. All the reactions were very efficient, including **ODN-2.3**, which contains two amine functions that were successfully guanylated almost quantitatively.

Table 2-5: Conditions used for the optimisation of guanylation of oligonucleotide **ODN-2.1**.[a]Estimated by HPLC-MS from the relative integrations of starting material and product after cleavage from the CPG solid support.

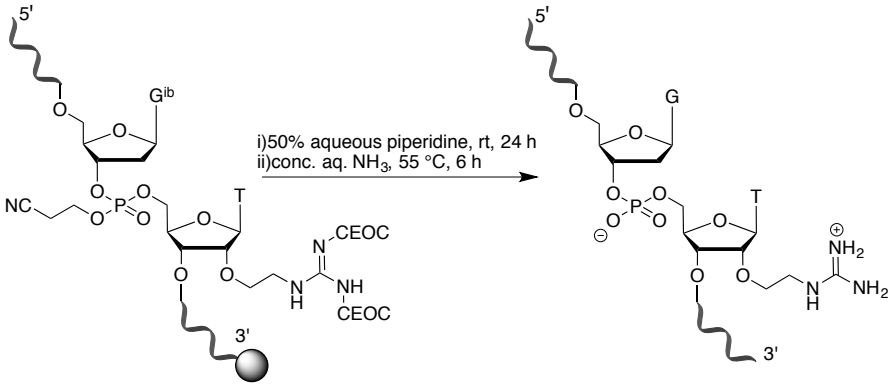
Entry	Solvent/base	T(°C)	Reaction time (h)	Conversion (%)[a]
1	DMF/10% TEA	55	20	>99%
2	DMF/10% TEA	70	5	>99%

Manoharan and co-workers reported the formation of a major byproduct upon cleavage of the CEOC groups of the 2'-*O*-(2-guanidinium) ethyl modified oligonucleotides by treatment with concentrated aqueous ammonia at 55 °C. The formation of a triazine derivative was suspected through the mechanism depicted in

Scheme 2-10. Consequently they employed a two steps deprotection procedure to first cleave the CEOC protecting group using piperidine as a non-nucleophilic base. Cleavage of the oligonucleotide from the solid support and deprotection of the internucleosidic phosphodiester cyanoethyl protecting groups was also performed during this first step. The oligonucleotide was then heated up in aqueous ammonia at 55 °C to complete deprotection of the amino group of the nucleobases (*Scheme 2-10*).



Scheme 2-10: Formation of 1,3,5-triazine-2,4-dione upon aqueous ammonia deprotection of oligonucleotides with a CEOC-protected guanidinium group at the 2'-position. The mechanism starts with the nucleophilic attack of the nitrogen of ammonia at one of the carbonyls bearing the 2-cyanoethoxy group followed by an intramolecular cyclisation to afford the triazine derivative.[124]



Scheme 2-11: two-step procedure for deprotection of the 2'-CEOC guanidine protecting groups, oligonucleotide solid-support cleavage and deprotection developed by Manoharan *et al.*[124]

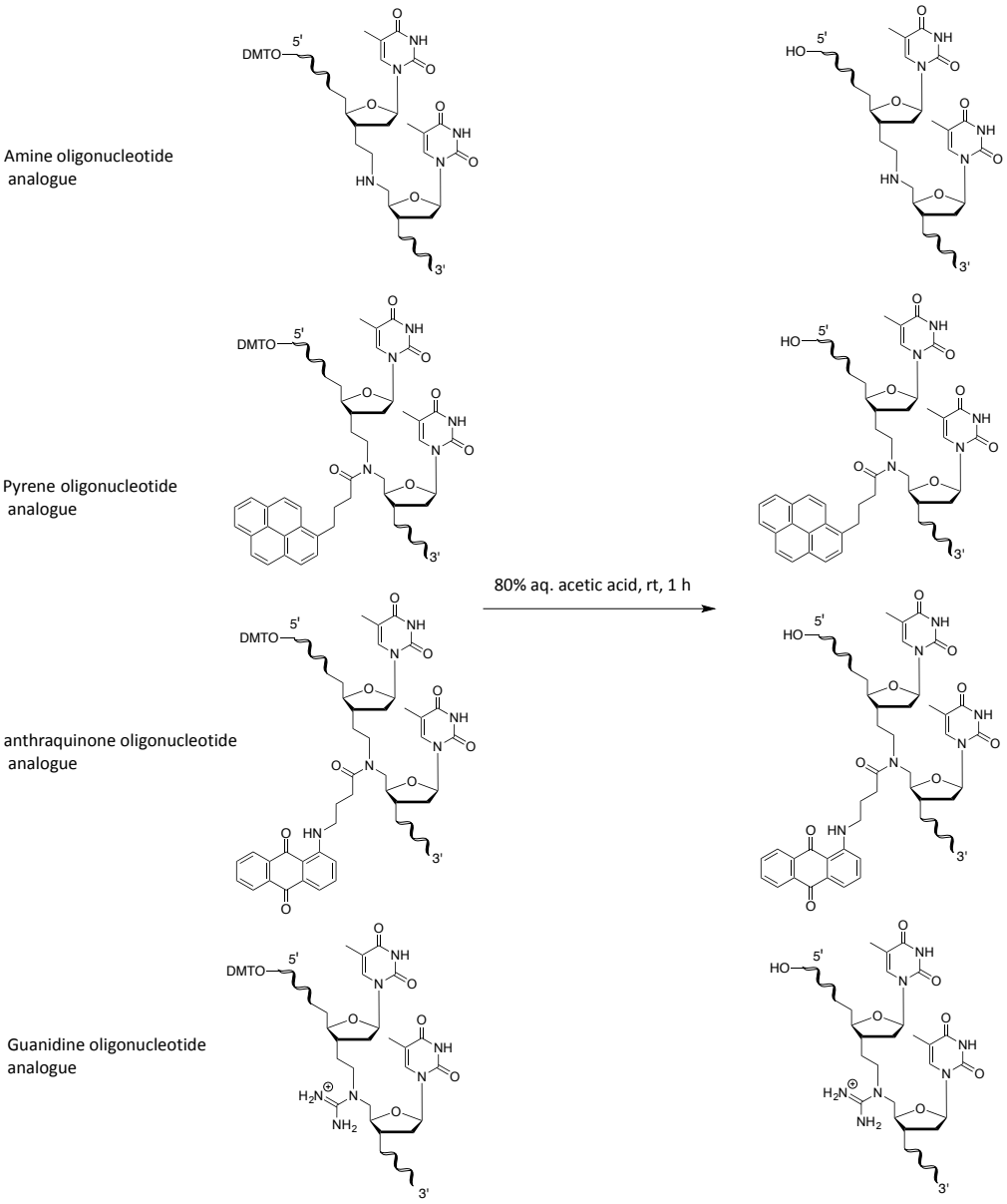
Pleasingly, upon treatment of the guanylated oligonucleotides **ODN-2.1** to **ODN-2.3** with concentrated aqueous ammonia, the formation of the above-mentioned triazine derivative was not observed. This enabled the convenient one-pot deprotection of the *N*-(2-cyano)ethoxycarbonyl groups and cleavage of the oligonucleotides under standard conditions.

2.3.7 Completion of the synthesis of amine and *N*-functionalised amine oligonucleotides

After deprotection of the nucleobases and cleavage from the solid support pyrene, anthraquinone and guanidine oligonucleotide analogues were purified by RP-HPLC. The 5'-dimethoxytrityl protecting group was then removed by treatment with acetic acid (*Scheme 2-12*). A summary of all modified oligonucleotides is presented in *Table 2-6* with their characterisation in *Chapter 6, section 6.2.2*.

Table 2-6: Summary of modified oligonucleotides. Insertion of amine, pyrene, anthraquinone and guanidine modified backbones (*Scheme 2-12*) are represented respectively by: *a*, *py*, *an* and *g.*

Code	Sequences (5' to 3')
ODN-2.1	CGTTCGCT$_a$TGCAGC
ODN-2.2	ACCCATTCCT$_a$TCGTTCCACT
ODN-2.3	ACCCAT$_a$TCCTTCGT$_a$TCCACT
ODN-2.4	CGTTCGCT$_{py}$TGCAGC
ODN-2.5	ACCCATTCCT$_{py}$TCGTTCCACT
ODN-2.6	ACCCAT$_{py}$TCCTTCGT$_{py}$TCCACT
ODN-2.7	ACCCATTCCT$_{an}$TCGTTCCACT
ODN-2.8	ACCCAT$_{an}$TCCTTCGT$_{an}$TCCACT
ODN-2.9	CGTTCGCT$_g$TGCAGC
ODN-2.10	ACCCATTCCT$_g$TCGTTCCACT
ODN-2.11	ACCCAT$_g$TCCTTCGT$_g$TCCACT



Scheme 2-12: Manual 4,4'-dimethoxytrytil (DMT) deprotection of the purified amine, pyrene, anthraquinone and guanidine oligonucleotide analogues.

Following the successful synthesis of the partially backbone modified oligonucleotides **ODN-2.1** to **ODN-2.11**, their ability to hybridise with a complementary unmodified DNA strand was studied by UV melting. In all cases, these results were compared to control unmodified DNA duplexes containing the same sequence.

2.3.8 Hybridisation properties of oligonucleotides containing intercalating backbone modifications

The presence of one amine backbone modification within the DNA duplex **ODN-2.1** led to a significant decrease in thermal stability (ΔT_m = -10.1 °C) compared to the unmodified control DNA duplex **ODN-2.12**, which is in accordance with the results previously obtained by Lynn and co-workers (*Figure 2-9*).[110] Pleasingly, the introduction of the pyrene functional group on the amine linkage induced a remarkable increase in melting temperature compared to the parent amine dsDNA **ODN-2.1** (ΔT_m = +7.7 °C), thereby restoring the thermal stability to a level comparable with that of the unmodified control DNA duplex **ODN-2.12** (ΔT_m = -2.4 °C).

Table 2.6

code	Sequences (5' to 3')	ΔT_m [a]
ODN-2.12	CGTTCGCTTGCAGC	
ODN-2.1	CGTTCGCT$_a$TGCAGC	-10.1
ODN-2.4	CGTTCGCT$_{py}$TGCAGC	-2.4
ODN-2.13	ACCCATTCCTTCGTTCCACT	
ODN-2.2	ACCCATTCCT$_a$TCGTTCCACT	-8.3
ODN-2.5	ACCCATTCCT$_{py}$TCGTTCCACT	-3.9
ODN-2.7	ACCCATTCCT$_{an}$TCGTTCCACT	-3.5

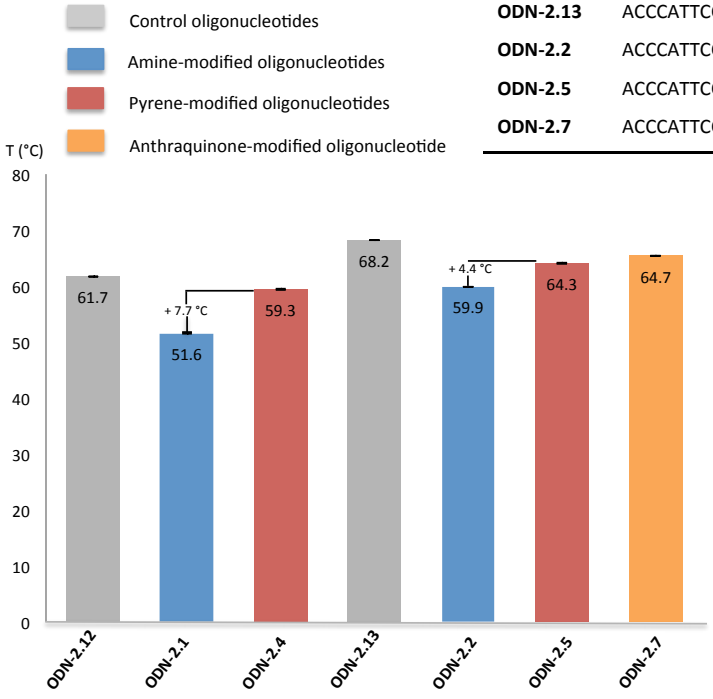


Figure 2-9: Melting temperatures of DNA duplexes with single amine, pyrene and anthraquinone modifications compared to unmodified phosphodiester DNA duplexes. Calculations of ΔT_m are presented in ***Table 2.6*** with $^a\Delta T_m$ = T_m (modified DNA) - T_m (unmodified DNA).

The melting temperatures measured for the amine dsDNA **ODN-2.2** and its modified pyrene analogue **ODN-2.5** followed a similar trend. However, the gain in thermal stability induced by the pyrene substituent was less significant in this case (ΔT_m = +4.4 °C). Oligonucleotides **ODN-2.4** and **ODN-2.5** have different sequences suggesting that the thermal stability induced by the pyrene modification, might be context dependent. In the future, it could be interesting to perform fluorescence-melting experiments using oligonucleotides **ODN-2.4** and **ODN-2.5** in order to confirm whether the stabilisation brought by the pyrene aromatic modification is due to an intercalation process (*Figure 2-9*).

The influence of another covalently linked DNA intercalator (anthraquinone) on the thermal stabilisation of the modified double-stranded DNA was also studied. The anthraquinone moiety has indeed a different aromatic surface compared to the pyrene moiety, and possesses two carbonyl functions which depending on the position of the DNA functional groups in the vicinity, could be involved in hydrogen bonding formation. It was observed that the DNA duplex formed with oligonucleotide **ODN-2.7**, which contains a single anthraquinone backbone-modification, has a melting temperature close to the value measured for the pyrene-modified DNA duplex **ODN-2.5**. This suggests that despite of their structural differences, the pyrene and anthraquinone intercalators have a similar stabilizing effect on the duplex containing the parent amine-modified backbone (*Figure 2-9*). This implies that their mode of action is similar and potentially intercalative

The introduction of two pyrene and anthraquinone backbone modifications in DNA duplexes formed with oligonucleotides **ODN-2.6** and **ODN-2.8** respectively, lead to a significant decrease in melting temperature compared to the unmodified control DNA duplex **ODN-2.13** (*Figure 2-10*). However, for both duplexes a similar gain in thermal stability compared to the parent amine-modified dsDNA was measured ($\Delta T_m \approx$ +4°C, *Figure 2-10*). As for oligonucleotides **ODN-2.5** and **ODN-2.7,** which contains respectively a single pyrene and anthraquinone modification (*Figure 2-9*), the aromatic moieties are able to improve the thermal stability of the DNA duplex containing amine-modified backbones but cannot completely compensate for the strong loss in thermal stability induced by this modification.

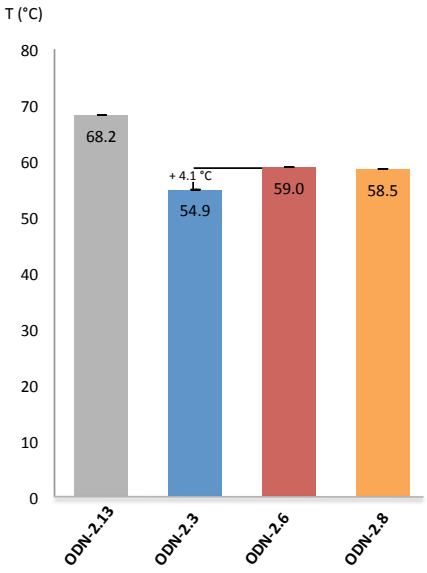


Table 2.7

code	sequences	ΔT_m [a]
ODN-2.13	ACCCATTCCTTCGTTCCACT	
ODN-2.3	ACCCAT$_a$TCCTTCGT$_a$TCCACT	-13.3
ODN-2.6	ACCCAT$_{py}$TCCTTCGT$_{py}$TCCACT	-9.2
ODN-2.8	ACCCAT$_{an}$TCCTTCGT$_{an}$TCCACT	-9.7

Figure 2-10: Melting temperatures of DNA duplexes with two amine, pyrene and anthraquinone modifications compared to unmodified phosphodiester DNA duplexes. Calculations of ΔT_m are presented in ***Table 2.7*** with $^a\Delta T_m = T_m$ (modified DNA) - T_m (unmodified DNA).

2.3.9 Hybridisation properties of oligonucleotides containing guanidinum backbone modifications

Amine-modified oligonucleotides were then converted to their corresponding guanidinium analogues, with the aim of increasing the thermal stability of DNA duplexes through electrostatic and hydrogen-bonding interactions with the complementary DNA strand anionic phosphodiesters. Unfortunately, this modification only had a moderate stabilising effect relative to the parent-amine modified dsDNA with the gain in melting temperature being 3.7 and 2.3 °C (**ODN-2.1** *vs*. **ODN-2.9** and **ODN-2.2** *vs.* **ODN-2.10** respectively, *Figure 2-11*). A possible explanation for this could be that the secondary amine linkage ($pK_a \approx 11$) is protonated at the pH of the buffer used for UV-melting experiments (pH = 7.0). Thus, introduction of the cationic guanidinium modification ($pK_a = 12.5$) might not have a strong additional stabilizing effect on the amine-modified DNA duplex. Nonetheless the subtle increase in melting temperature compared to the parent amine-modified backbone could be due to the protrusion of the guanidinium

functional group, which may bring it closer to the negatively charged phosphodiester of the complementary strand.

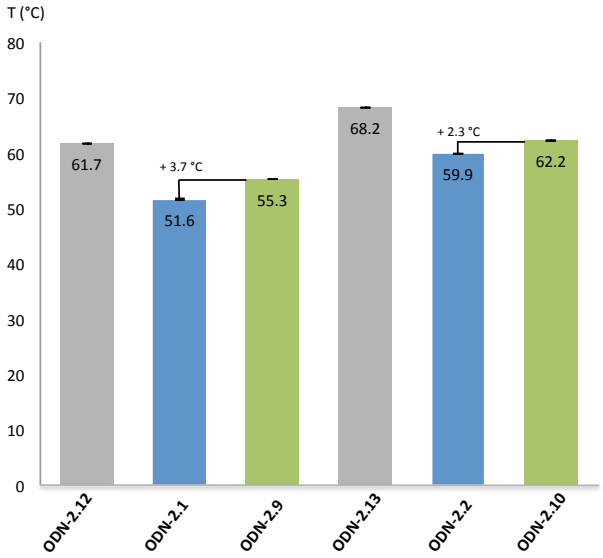


Table 2.8

code	Sequences (5′ to 3′)	ΔT_m [a]
ODN-2.12	CGTTCGCTTGCAGC	
ODN-2.1	CGTTCGCT$_a$TGCAGC	-10.1
ODN-2.9	CGTTCGCT$_g$TGCAGC	-6.4
ODN-2.13	ACCCATTCCTTCGTTCCACT	
ODN-2.2	ACCCATTCCT$_a$TCGTTCCACT	-8.3
ODN-2.10	ACCCATTCCT$_g$TCGTTCCACT	-6

Figure 2-11: Melting temperatures of DNA duplexes with a single amine and guanidine modifications compared to unmodified phosphodiester DNA duplexes. Calculations of ΔT_m are presented in ***Table 2.8*** with $^a\Delta T_m = T_m$ (modified DNA) - T_m (unmodified DNA).

The introduction of two guanidinium modifications within the DNA duplex **ODN-2.11** was also investigated. Unfortunately, **ODN-2.11** was found to display a lower T_m compared to oligonucleotide **ODN-2.10**, which contains only one guanidine group (*Figure 2-11* and *Figure 2-12*). It is believed that the structural perturbation induced by the additional modified internucleosidic linkage is stronger that the gain in electrostatic interaction potentially induced by the guanidinium cation.

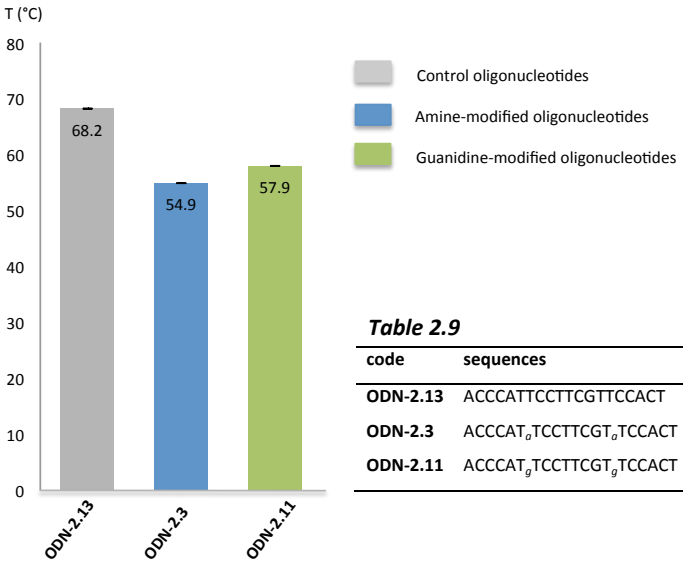


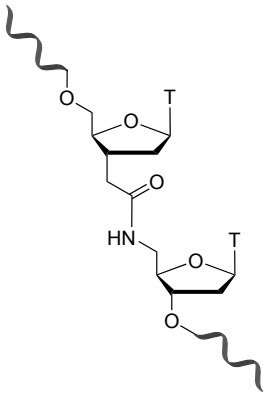
Table 2.9

code	sequences	$\Delta T_m{}^a$
ODN-2.13	ACCCATTCCTTCGTTCCACT	
ODN-2.3	ACCCAT$_a$TCCTTCGT$_a$TCCACT	-13.3
ODN-2.11	ACCCAT$_g$TCCTTCGT$_g$TCCACT	-10.3

Figure 2-12: Melting temperatures of DNA duplexes with two amine and guanidine modifications compared to unmodified phosphodiester DNA duplex. Calculations of ΔT_m are presented in ***Table 2.9*** with ${}^a\Delta T_m = T_m$ (modified DNA) - T_m (unmodified DNA).

2.4 Synthesis of modified oligonucleotides with amide-modified backbone

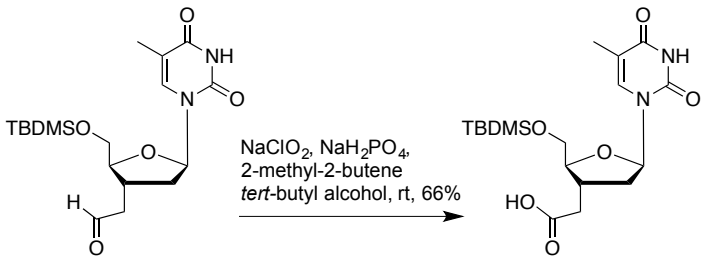
The preparation of an additional internucleosidic linkage **2.III** featuring an amide bond (*Figure 2-13*) was also evaluated, given its ease of access from 3′-oxoethyl and 5′-amino nucleoside precursors **2.5** and **2.6**. This backbone modification previously synthesised by De Mesmaeker and co-workers generally stabilises DNA:RNA (ΔT_m = +0.9 °C) and slightly destabilises DNA:DNA duplexes (ΔT_m = -1.5 °C).[58] Rozners *et al.* have demonstrated using X-ray crystallography that this amide backbone is an excellent mimic of the phosphodiester internucleosidic linkage in RNA.[59,62] Amide-modified RNA forms a typical A-form duplex with the amide carbonyl group pointing into the major groove in a similar orientation as the phosphodiester bond and suggest that this carbonyl could mimic one of the non-bridging P-O bond. It was also shown that a siRNA containing height 3′-CH_2-CO-NH-5′ amide-backbone modifications displayed a better gene silencing activity compared to its unmodified siRNA counterpart.[62]



2.III

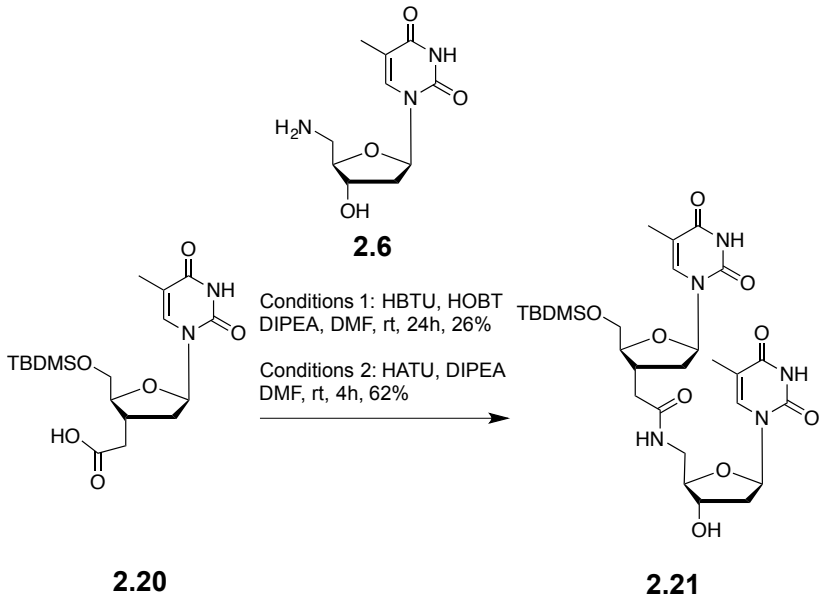
Figure 2-13: Structure of amide modified backbone previously synthesised by De Mesmaeker.[58]

In a first step, oxidation of 3'-oxoethyl-thymidine nucleoside **2.5** afforded the 5'-*O*-TBDMS-3'-deoxythymidine-3'-acetic acid **2.20** in 66% yield (*Scheme 2-13*). Following this, the first attempt in the coupling of 5'-amino-thymidine **2.6** with the functionalised carboxylic acid nucleoside **2.20** was carried out, using the peptide coupling reagent HBTU (*O*-(Benzotriazol-1-yl)-*N*,*N*,*N*',*N*'-tetramethyluronium hexafluorophosphate) associated with hydroxybenzotiazole (HOBt) (*Scheme 2-14*).[59] The amide thymidine dimer **2.21** was obtained in a poor yield of 26% and was contaminated with HOBt impurities that proved difficult to remove from the desired 5'-*O*-TBDMS thymidine amide dimer **2.21** by column chromatography. Consequently, the coupling reagent was changed to HATU which is a more efficient reagent than HBTU[142], and pleasingly afforded intermediate **2.21** in a reasonable yield of 62% (*Scheme 2-14*).



2.5 **2.20**

Scheme 2-13: Synthesis of 5′-*O*-tert-butyldimethylsilyl-3′-deoxythymidine-3′-acetic acid



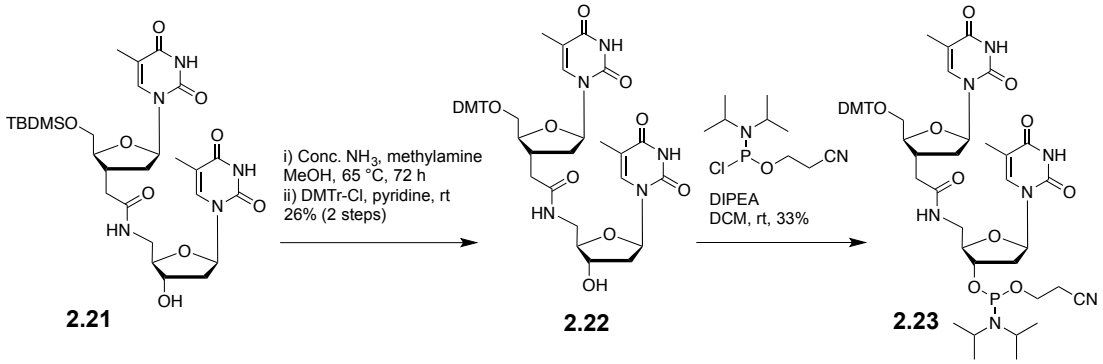
2.6

2.20 **2.21**

Scheme 2-14: Optimisation of amide coupling between 5′-amino-5′-deoxythymidine **2.6** and 5′-*O*-TBDMS-3′-deoxythymidine-3′-acetic acid **2.20**.

Subsequent silyl deprotection of the 5′-hydroxyl group of thymidine amide dimer **2.21** using an excess of tetrabutylammonium fluoride, as reported by *Lynn et al.* was challenging as the polar deprotected amide dimer intermediate was contaminated with residual tetrabutylammonium salts.[110] Removal of this impurity by column chromatography was cumbersome and another method of deprotection using tetrabutylammonium fluoride loaded on silica gel was attempted. Contamination of the deprotected T-T amide dimer with tetrabutylamonium salts was also observed in this case. It was therefore decided to cleave the silyl protecting

group of the thymidine amide dinucleoside **2.21** under basic conditions using concentrated aqueous ammonia and methylamine. This was immediately followed by tritylation to give the 5′-DMT thymidine amide dimer **2.22** in 26% yield over 2 steps (*Scheme 2-15*). Precursor **2.22** was then phosphitylated to afford 5′-DMT thymidine amide dimer phosphoramidite **2.23**, which was used in the synthesis of oligonucleotide **ODN-4.8** (*Chapter 6, section 6.2.4*).



Scheme 2-15: Completion of the synthesis of 5′-*O*-4,4′-dimethoxytrityl thymidine amide dimer phosphoramidite **2.23**

As previously mentioned the amide-backbone modification is a good mimic of the phosphodiester internucleosidic bond in RNA. It would therefore be interesting to study if this modified backbone is tolerated during DNA replication or other polymerase-based process. To this end, Kuwahara and co-workers reported a brief study of the ability of DNA polymerases, (KOD(exo-), Vent(exo-) and Taq, to read through this linkage.[143] When the 3′-CH_2-CO-NH-5′ amide modification was inserted on the 3′-end of the primer, polymerase extension could not proceed and largely terminated after incorporation of the first nucleotide. Primer extension reactions using amide-modified DNA templates was also investigated and it was observed by PAGE gel analysis that the extension reactions did not proceed efficiently. The authors concluded that even though DNA polymerases could read through the modified backbone that its negative effect on downstream extension was too significant.

However, as will be discussed in Chapter 4, their conclusion appeared to be premature; the amide backbone still has potential for use in PCR and oligonucleotide chemical ligation. Indeed, the amine and carboxylic acid functions could be easily installed on one or both ends of oligonucleotides. These functions are chemically stable and selective ligation of the functionalised oligonucleotides can be triggered by the use of water-soluble amide coupling agents, such as 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC).[144] In order to demonstrate the possible use of this backbone in chemical DNA assembly for the *de novo* synthesis of genes, its ability to be accepted by various DNA polymerases will be investigated and discussed in Chapter 4.

2.5 Conclusion and outlook

In summary, three novel oligonucleotide backbones, namely the pyrene, anthraquinone and guanidinium modified backbones, were successfully synthesised from a parent amine backbone. The solid phase synthetic strategy developed to access these backbones was simple, effective, and could be easily extended to other types of substituents.

It was first demonstrated that introduction of aromatic modifications (pyrene and anthraquinone) increased the thermal stability of the amine-modified backbone but could not fully compensate the loss of stability induced by this very flexible modification. It was observed that pyrene and anthraquinone functional groups, although being structurally different, induced a similar stabilising effect on DNA duplexes containing the parent amine-modified internucleosidic linkage. It will be interesting in the future to perform fluorescence melting of pyrene and anthraquinone modified DNA duplexes as well as modeling studies in order to confirm that the reason for this is intercalation of these groups.

The correct intercalation of these covalent aromatic groups between the duplex base pairs strongly depends on the linker length and on the type of bond used for the tethering on the modified amine backbone. Here pyrene and anthraquinone moieties were attached through an amide bond using NHS active esters. In the future it would be interesting to investigate other methods of attachment such as a reductive amination using aldehyde derivatives of pyrene, anthraquinone and other intercalators. This will give access to oligonucleotides bearing functionalised tertiary amine modified backbones that would be protonated over a wide range of pH, which might have a positive effect on DNA duplex stability. Variation of the length of the linker separating the aromatic function and the backbone should also be studied.

It was also shown that the positively charged guanidinium modification only had a subtle effect on the thermal stability of the amine modified duplex. It is thought that the amine-modified linkage is protonated at the pH of the UV-melting experiment and thus the gain in thermal stability by adding another cationic

modification is moderate. However, the slight increase in melting temperature brought by the guanidinium backbone to the parent ammonium-modified oligonucleotide could be due to the protruding of the guanidinium functional group. It will therefore be interesting to also increase the linker length between the guanidine moiety and the amine backbone in order to position the positive charge closer to the negatively charged phosphodiester.

One way to increase the DNA duplex stability is to reduce the number of negative charges within the modified oligonucleotide strand in order to decrease the negative charge to charge repulsion between the two strands of the duplex. This can also improve the cell uptake of modified oligonucleotides and increase the potency of antisense or siRNA based therapeutics. It would therefore be relevant and fascinating to introduce several guanidinum modifications to obtain zwitterionic or positively charged oligonucleotides.

Finally, an additional amide modified backbone **2.III** was synthesised from similar precursors used in the synthesis of the amine modified backbone. The enzymatic properties of this backbone will be discussed further in chapter 4.

Chapter 3

Synthesis and study of the hybridisation properties of DNA containing triazole modified backbones

This chapter first reviews our previous studies of chemical ligation using the 1,4-disubstituted 1,2,3-triazole backbone as a phosphate mimic via copper (I) catalysed azide-alkyne cycloaddition. The synthesis of novel dinucleotide thymidine phosphoramidites containing two different 1,2,3-triazole internucleosidic linkages is then described. Finally study of the hybridisation properties of these modified oligonucleotides will be reported.

3.1 Background

3.1.1 Click chemical ligation of oligonucleotides

"Click chemistry" was introduced by Sharpless *et al.* to describe reactions that are "modular, wide in scope, give very high yields, generate only inoffensive by-products that can be removed by non-chromatographic methods, and that are stereospecific".[145] An excellent example of such a reaction is the copper (I) catalysed azide-alkyne cycloaddition (CuAAC)[146], which was exploited by Brown *et al.* as a novel chemical strategy to join long nucleic acid fragments (*Figure 3-1*). The so-called "click ligation" of oligonucleotides was employed in numerous applications such as RNA synthesis,[147] nanotechnology,[148,149] and could be used as an alternative to current PCR-based gene synthesis methods. This is due to the finding that the triazole linkage is read through by DNA and RNA polymerases. In this context, several triazole backbone analogues have been reported by our group and will be discussed below.

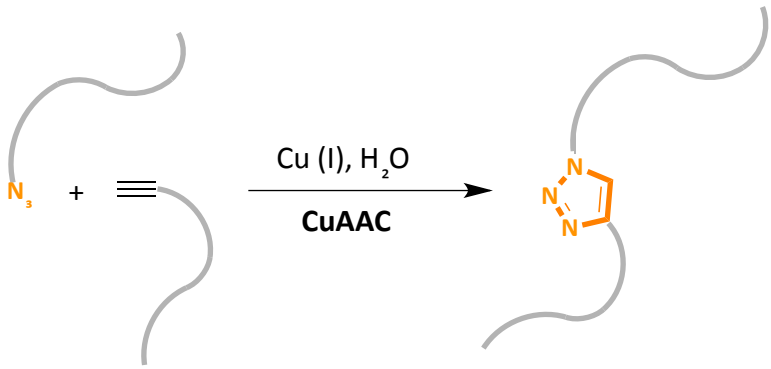


Figure 3-1: Click ligation of DNA through the formation of a 1,4-disubstituted 1,2,3-triazole linkage.

3.1.2 First generation of triazole modified DNA backbone

Triazole backbone **3.I** can be obtained by ligation of oligonucleotides functionalised with 3'-azido and 5'-propargylamido groups (*Figure 3-2*). The rate of the reaction is greatly enhanced if the two reactive groups are held together in close

proximity by an annealed complementary template strand (sometimes referred to as a splint) but can occur without a template if longer reaction times and higher oligonucleotide concentrations are used. The DNA strand containing the modified linkage **3.I** was successfully amplified by PCR reaction, but replication of one thymidine was consistently missed at the ligation site. These results were nonetheless encouraging as they provided the first example of efficient chemical DNA strand ligation combined with PCR amplification through an unnatural DNA backbone linkage. It also gave hope that a perfect phosphodiester linkage mimic could be designed if the limitations of triazole linkage **3.I** could be rationalised.

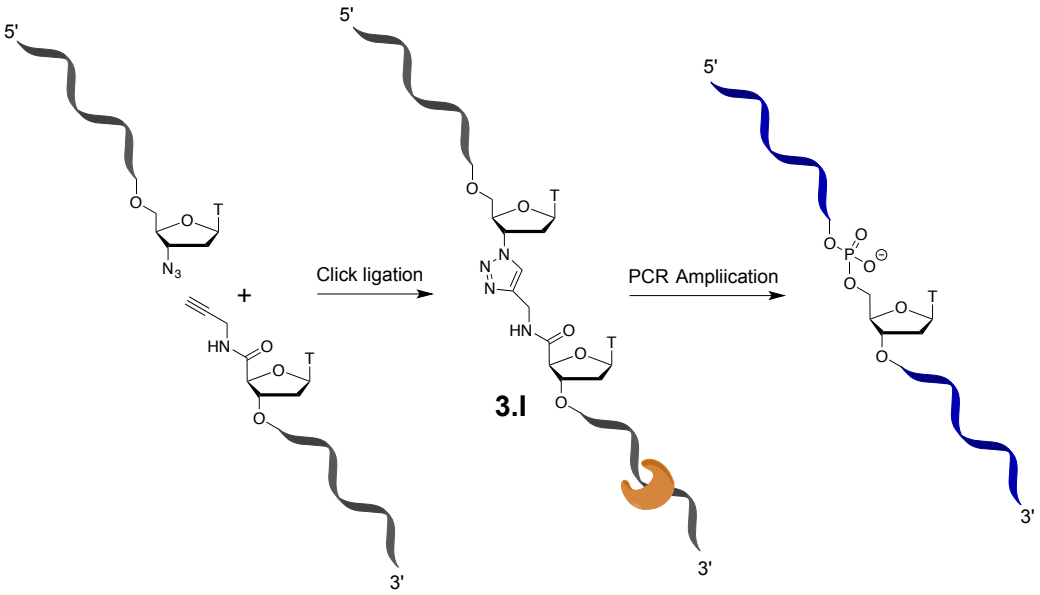


Figure 3-2: Ligation and PCR of oligonucleotides containing triazole backbone **3.I**.

3.1.3 Second generation of triazole modified DNA backbone

A possible explanation for the incorrect copying of the thymidine adjacent to the triazole linkage **3.I** could be the presence of the amide bond. Indeed, assuming that the *trans*-amide configuration is the predominant form, this might force one of the thymidines in the artificial linkage to be flipped out of position in the template strand, consequently facilitating its bypass during replication (*Figure 3-3*).[150]

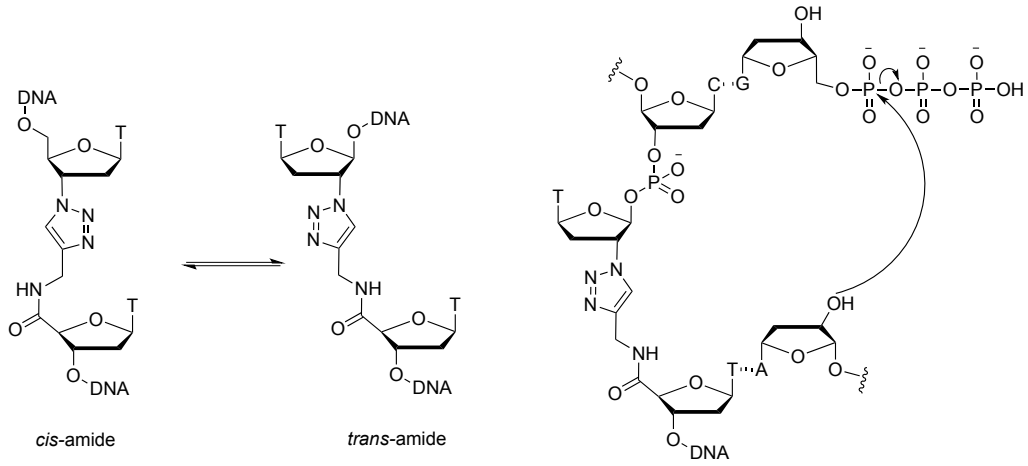


Figure 3-3: Proposed model for the flipping out of thymidine in triazole containing DNA template during the replication process.[151]

Triazole backbone **3.II** was therefore designed to maintain the flexibility of the 5′-methylene functional group from canonical DNA. It can be obtained by a CuAAC reaction between 5′-azido and 3′-propargyl oligonucleotides, which are synthetically accessible through standard phosphoramidite chemistry (*Figure 3-4*). Oligonucleotides bearing a 3′-alkyne extremity were obtained by solid phase synthesis using a 3′-propargyl MeC functionalised resin (synthesised in 3 steps (14%) from thymidine).[152] The 5′-azide DNA strand was conveniently made in two steps from the corresponding unmodified 5′-hydroxyl DNA congener. The utility of click ligation using triazole linkage **3.II** for the assembly of large DNA molecules was demonstrated through the successful ligation of three 100-mer alkyne and azide functionalised oligonucleotides to form a 300-mer oligonucleotide (*Figure 3-4*). The 1,2,3-triazole **3.II** proved to be entirely biocompatible as PCR templates containing one or more triazole linkages were amplified with high fidelity and a gene containing this modified linkage was functional in *E. coli*.[153] It was later showed that the triazole linkage **3.II** was the first example of unnatural DNA backbone to be fully functional in eukaryotic cell.[154]

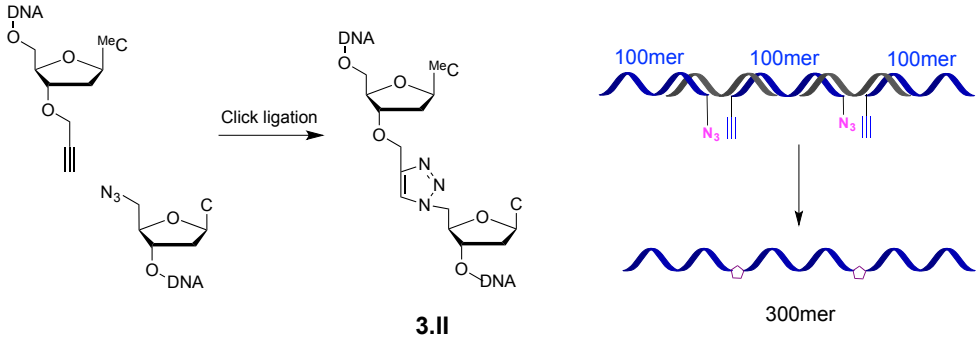


Figure 3-4: Templated ligation of 100-mer oligonucleotides *via* triazole backbone **3.II**.[153]

Despite the fact that backbone **3.II** appears to be a perfect surrogate of the natural phosphodiester linkage *in vitro* and *in vivo*, the modification does thermally destabilise the resultant duplex by 8 °C. Yet, the NMR solution structure of a 13-mer duplex containing **3.II** showed that all base pairs (T20:A07, T19:A08) around the triazole site are correctly formed and that the overall B-shape of the duplex is maintained (*Figure 3-5*). The N3 of the triazole ring occupies a similar position to that of the non-bridging oxygen atoms of the phosphodiester bond and might therefore be a good substitute H-bond acceptor. This could partially explain the exceptional tolerance of this backbone by DNA polymerases. However, to accommodate its extra length and conformational flexibility, the linkage displaces the 5′-CH_2 group 3.2 Å from its original position, which might explain the loss of thermal stability.[155]

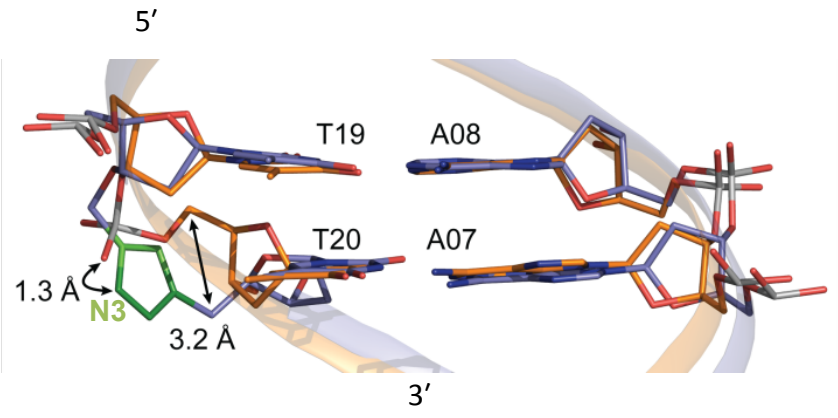


Figure 3-5: Overlay of the average structure of natural duplex (orange) and modified duplex (blue) with triazole ring represented in green.[155]

3.1.4 Enhancement of the thermodynamic stability of second generation of triazole backbone with a G-clamp cytosine analogue

An approach to circumvent the decrease in thermal stability of duplexes containing triazole internucleosidic linkage **3.II**, is the introduction of an additional stabilising modification that would cause minimal structural perturbation in order to retain the biocompatibility of the modified triazole linkage. The tricyclic cytosine analogue 9-(2-aminoethoxy)-phenoxazine carries an additional aminoethoxy group on the rigid phenoxazine scaffold and is termed "G-clamp" as this provide a high affinity for a complementary guanine nucleobase. It was demonstrated that a single incorporation of G-clamp enhanced the DNA duplex stability with a ΔTm of up to 18 °C relative to a control DNA containing 5MedC. G-clamp provides an even higher discrimination between the perfect match and mismatched sequences compared to control 5MedC. The increased affinity of G-clamp is due to the combination of extended intrahelical base stacking and an additional specific hydrogen bonded contact with the O6 atom of the complementary guanine (*Figure 3-6*).[156]

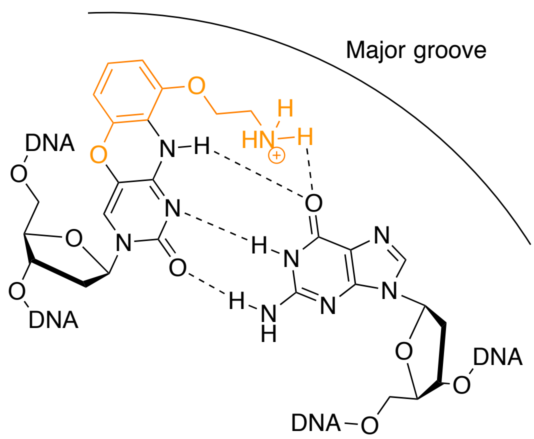


Figure 3-6: Model of G-clamp-guanine H-bonding interactions within a DNA duplex showing the 3-(2-aminoethoxy)phenoxy (in orange) additional steric bulk compared to cytosine.

Therefore, combination of the G-clamp cytosine analogue with the triazole internucleosidic bridge (*Figure 3-7*) was studied as a mean of overcoming the destabilising effect of the triazole linkage, the idea being that the 3-(2-aminoethoxy)phenoxy group protruding into the major groove would not affect the linkage biocompatibility (*Figure 3-6*). Ultra-violet melting studies of a model 13-mer oligonucleotide containing a triazole G-clamp dinucleotide modification with complementary unmodified DNA and RNA were performed. The G-clamp/triazole modified backbone combination improved the melting temperature of the duplex by 6.5 °C compared to the unmodified duplex and by 12 °C compared to the triazole modified duplex containing a cytosine instead of the G-clamp. The increase in thermal stability induced by the additional G-clamp was even more potent for DNA:RNA duplexes. The same trend in duplex thermal stability was observed for duplexes containing a T-triazole-G-clamp step, *i.e*. the thermal stability of G-clamp/triazole was superior to the stability of canonical DNA:RNA, which was superior to the one of triazole DNA:RNA (*Figure 3-8*).

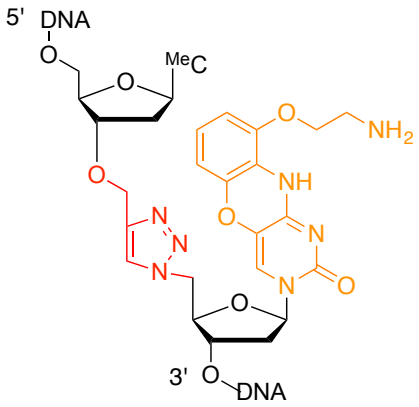


Figure 3-7: Combined G-clamp (orange) and triazole internucleosidic modification (red) within a oligodeoxyribonucleotide.[157]

Linear copying and PCR experiments using Klenow large fragment and GoTaq DNA polymerases respectively followed by Sanger sequencing of the PCR amplicons showed that the G-Clamp/triazole was faithfully read-through by DNA polymerases without apparent mutagenesis. This suggests that the G-clamp/triazole could be

used for *in vivo* applications with the small disadvantage that the base at the 3'-side of the triazole will be imposed. It will therefore be preferential to circumvent the extra base modification by designing a biocompatible linkage that would also display improved hybridisation properties compared to the 1,2,3-triazole linkage **3.II**.

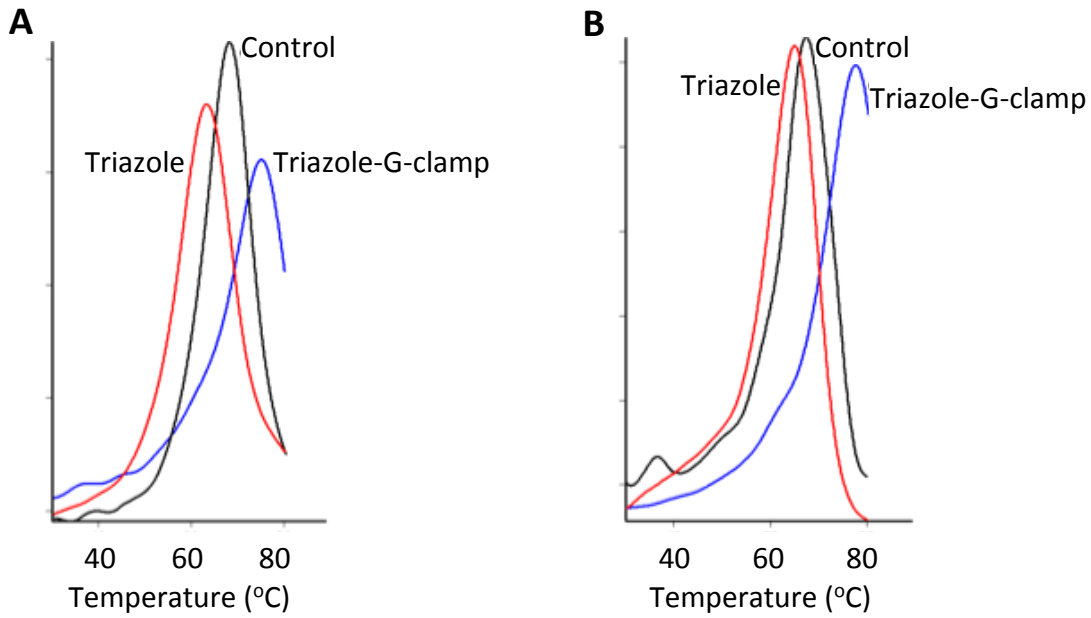


Figure 3-8: first derivative curves of oligonucleotides containing discretely, triazole, triazole/G-clamp, and unmodified phosphodiester (control) linkages with targeted complementary DNA (**A**) or RNA (**B**).[157]

3.1.5 Previous synthesis of an analogue of 1,2,3-triazole linkage 3.II

An interesting structural variant of triazole linkage **3.II** is the DNA modified 1,4-disubstituted 1,2,3-triazole linkage **3.III**, which was previously used by Isobe et al. for the design of a "triazole-linked analogue of DNA (^{TL}DNA)" featuring triazole linkages in place of all phosphodiester bonds (*Figure 3-9*).[53,54] This involved the synthesis of the key monomer **3.VI**, featuring an alkyne functional group at the 5'-position and an azide function at the 3'-position.

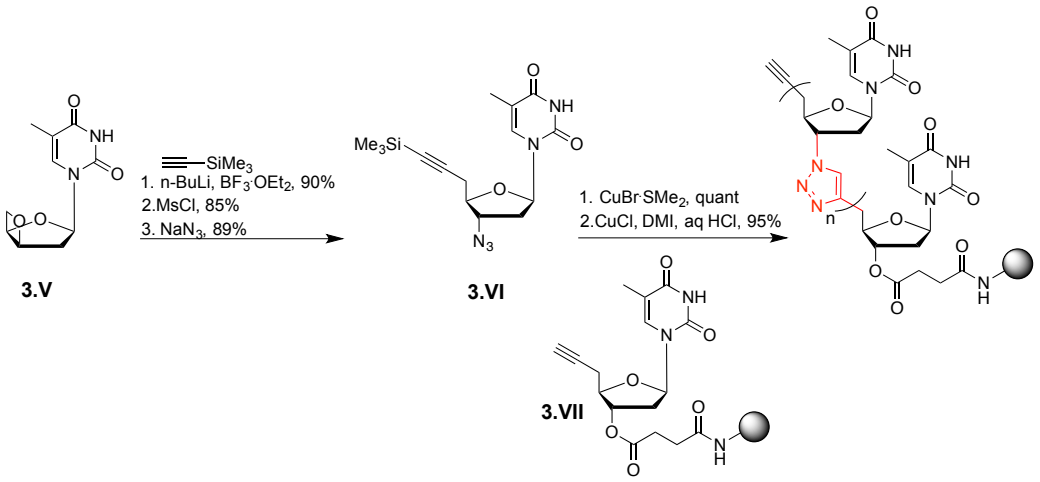


Figure 3-9: Solid phase synthesis of ^{TL}DNA containing triazole linkages **3.III** (in red) described by Isobe *et al.*[158]

Study of a model 10-mer oligothymidine ^{TL}DNA showed that this DNA analogue formed a much more stable B-DNA duplex compared to the unmodified control DNA duplex (ΔT_m = + 41.1 °C). It was hypothesised that this substantial increase in T_m is due to the neutral backbone that does not have any repulsive interactions with the anionic phosphate backbone of the DNA complementary target. Molecular modelling showed that the six-bond periodicity of ^{TL}DNA seemed to be important for the duplex formation.[158] Further studies of the behaviour of the triazole linkage **3.III** in a different DNA context were however not reported.

3.2 Aim of the study

In order to investigate the relationship between polymerases replication and 1,2,3-triazole linkages, the most direct method is to study structural analogues of the most successful triazole internucleosidic bridge **3.II**. With this idea in mind the aim was to synthesise two 1,2,3-triazole backbones **3.III** and **3.IV** (*Figure 3-10*), which for convenience will be referred to as "short-triazole" and "methyl-triazole" respectively. Triazole linkage **3.III** is two bonds shorter than the initial analogue **3.II** which positions the 1,2,3-triazole ring higher compared to triazole backbone **3.II**. This linkage might be better accommodated within the DNA duplex, which might have an effect on the linkage thermodynamic stability and biocompatibility. Backbone **3.IV** is the methylated version of triazole linkage **3.II** and considering that hydrogen bonded contact with the N3 nitrogen atom of the triazole might be involved in its acceptance by DNA polymerases, it could be interesting to study the impact of the methyl substituent on the backbone biocompatibility.

The synthetic strategies developed to access triazole modified backbones **3.III** and **3.IV** will be presented in the next sections, followed by a study of their hybridisation properties.

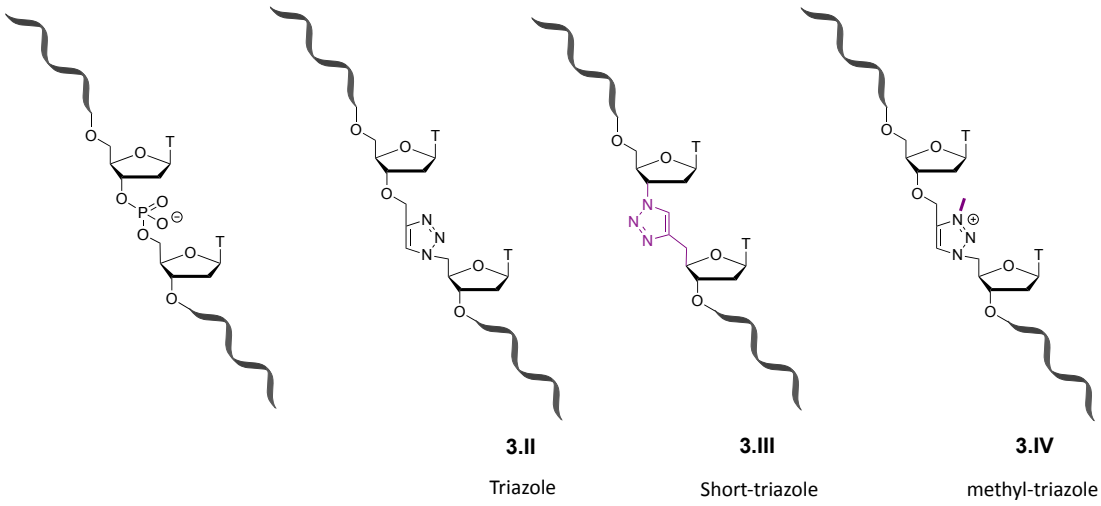


Figure 3-10: Structures of two new triazole **3.II** variants (**3.III** and **3.IV**) in comparison with the structure of unmodified phosphodiester linkage.

3.3 Synthesis of phosphoramidite nucleoside dimers with triazole linkages and modified oligonucleotides

A possible approach to access oligonucleotides containing short-triazole linkages **3.III**, could be through a variation of the methodology developed to synthesise triazole (**3.II**) linked DNA. This typically requires two oligonucleotides; one with a 5′-alkyne function and the other with a 3′-azide functional group, followed by click ligation to join the two oligonucleotides (*Figure 3-11*, A). The DNA strand with a 5′-terminal alkyne function could be obtained *via* standard oligonucleotide solid phase synthesis using a 5′-ethynyl-3′-phosphoramidite nucleoside (*Figure 3-11*, C). However, the synthesis of the 3′-azide oligonucleotide would require the preparation of a 3′-*O*-mesyl-5′-phosphoramidite nucleoside[160] for reverse direction oligonucleotide synthesis. The 3′-*O*-mesyl function could be then converted to a 3′-azide functional group post-oligonucleotide synthesis (*Figure 3-11*, B).

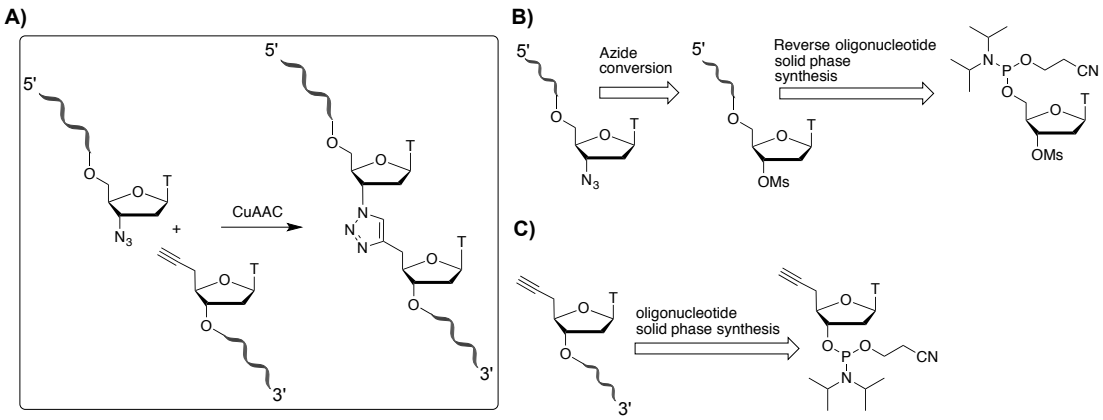


Figure 3-11: **A)** Possible strategy to obtain short-triazole linked oligonucleotides *via* click ligation. **B)** Proposed retrosynthetic strategy to access 3′-azide oligonucleotide and **C)** 5′-alkyne oligonucleotide

A more facile route to access oligonucleotide analogues containing triazole linkages **3.III** was chosen and involve the incorporation of a dinucleoside 3′-phosphoramidite building block during standard automated oligonucleotide solid phase synthesis (*Figure 3-12*). The most straightforward approach was to have the

bases on either side of the triazole linkage as thymines. Unlike adenine, cytosine and guanine, this nucleobase does not require additional protecting groups and the building blocks required for the synthesis of the phosphoramidite dimer can be easily prepared from commercially available thymidine. Methodologies to access other combinations of dinucleosides will be investigated in the future.[161]

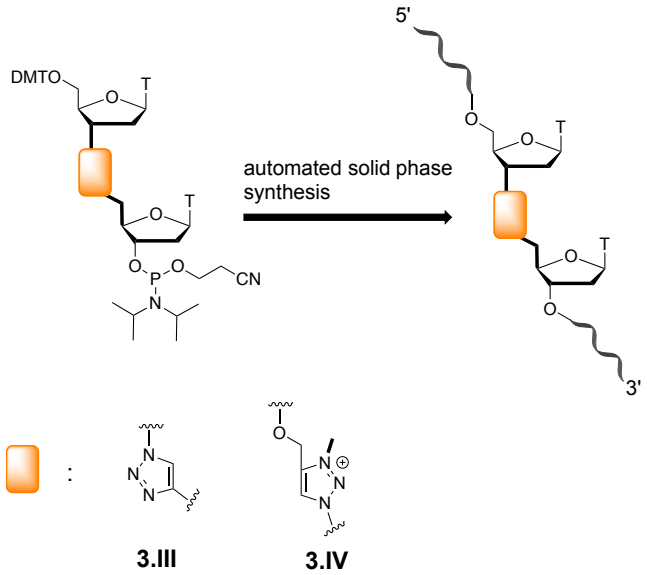
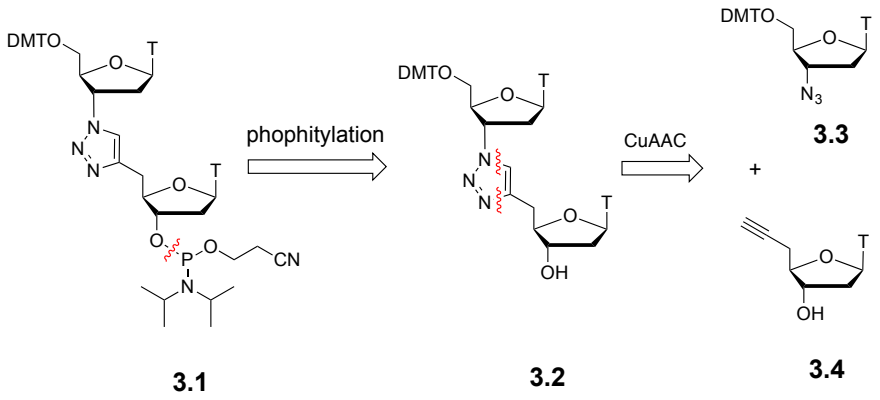


Figure 3-12: Synthetic strategy for the preparation of modified oligonucleotides containing triazole linkages **3.III** (short-triazole) and **3.IV** (methyl-triazole).

It is important to notice that the click ligation of oligonucleotides *via* formation of a 1,2,3-triazole linkage **3.II**, followed by methylation to access the methyl-triazole **3.IV** is not possible as this will lead to alkylation of the nucleobases and degradation of the DNA strand. Consequently the same strategy as described above was adopted to access oligonucleotides with methyl-triazole linkage **3.IV** (*Figure 3-12*).

3.3.1 Retrosynthesis for preparation of T-T dimer 3′-phosphoramidite with short triazole linkage

Dinucleoside 3′-phosphoramidite **3.1** bearing modified triazole backbone **3.III** can be obtained from the 5′-dimethoxytrityl-1,2,3-triazole thymidine dimer **3.2** *via* phosphitylation. This precursor will arise from a CuAAC reaction between the 3′-azido-5′-DMT-3′-deoxythymidine and 5′-ethynyl-5′-deoxythymidine fragments **3.3** and **3.4** (*Scheme 3-1*).



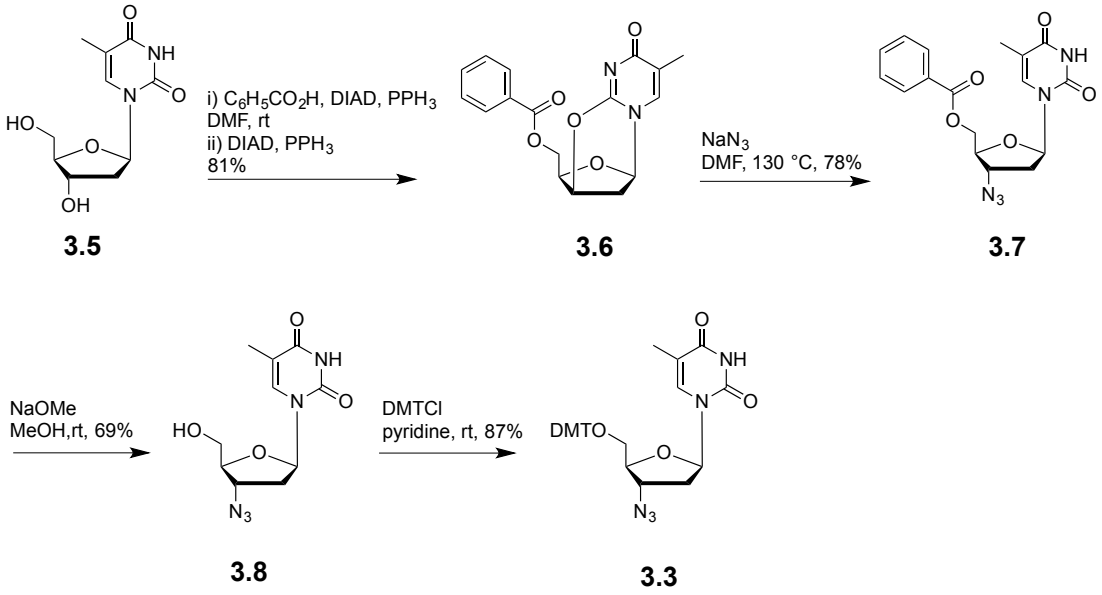
Scheme 3-1: Retrosynthetic analysis of 5′-DMT T-short-triazole-T dimer 3′-phosphoramidite building block **3.1**

3.3.2 Synthesis of 3′-azido-5′-DMT-3′-deoxythymidine fragment 3.3

The synthesis of the 3′-azido thymidine fragment **3.3** was achieved following published procedures.[162,163] At first, intermediate **3.6**[d] was obtained from thymidine **3.5** *via* a one-pot transformation involving a tandem Mitsunobu reaction.[164] The first reaction enabled the protection of the 5′-hydroxyl group as a benzoate ester then the second Mitsunobu reaction allowed an intramolecular cyclisation to obtain precursor **3.6**. Following this, ring opening of the 2,3′-anhydro-5′-*O*-benzoylthymidine **3.6** by treatment with sodium azide furnished the 5′-benzoyl protected 3′-azido thymidine intermediate **3.7** in a good yield of 78%. This precursor was then deprotected with sodium methoxide to afford the 3′-azido-3′-deoxythymidine (AZT) **3.8** in 69% yield. Finally, protection of the 5′-hydroxyl group

[d] Previously synthesised by A. Halfar following the procedure of Czernecki *et al.*[164]

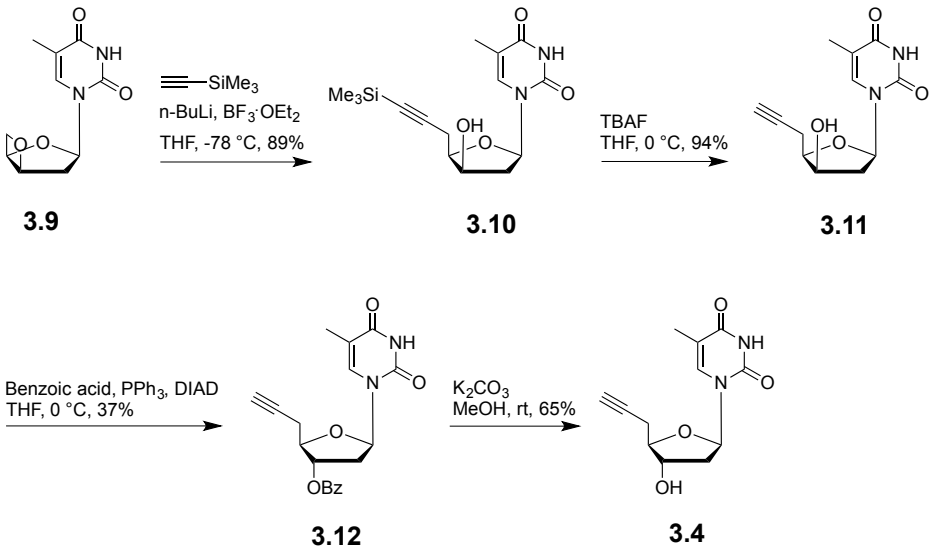
of AZT **3.8** as a dimethoxytrityl ether, gave the targeted 3′-azido nucleoside fragment **3.3** in an excellent yield of 87% (*Scheme 3-2*).



Scheme 3-2: Synthetic route to 3′-azido-5′-DMT-3′-deoxythymidine **3.3**.

3.3.3 Synthesis of 5′-ethynyl-5′-deoxythymidine fragment 3.4

5′-ethynyl thymidine **3.4** was synthesised following the procedure described by Isobe *et al.*[158] In a first step, ring opening of oxetane thymidine derivative **3.9** by treatment with trimethylsilylacetylene, n-butyl lithium and $BF_3{\cdot}OEt_2$ provided the 5′-(trimethylsilyl)ethynyl nucleoside **3.10** in a good yield of 89%. This precursor was subsequently subjected to deprotection by treatment with TBAF to give nucleoside **3.11** in an excellent yield of 94%. The latter was then treated under Mitsunobu conditions, to afford 5′-ethynyl nucleoside **3.12** with the desired configuration at the 3′-hydroxyl position. Deprotection of alkyne nucleoside **3.12** *via* methanolysis afforded the desired alkyne fragment **3.4** in 65% yield (*Scheme 3-3*).

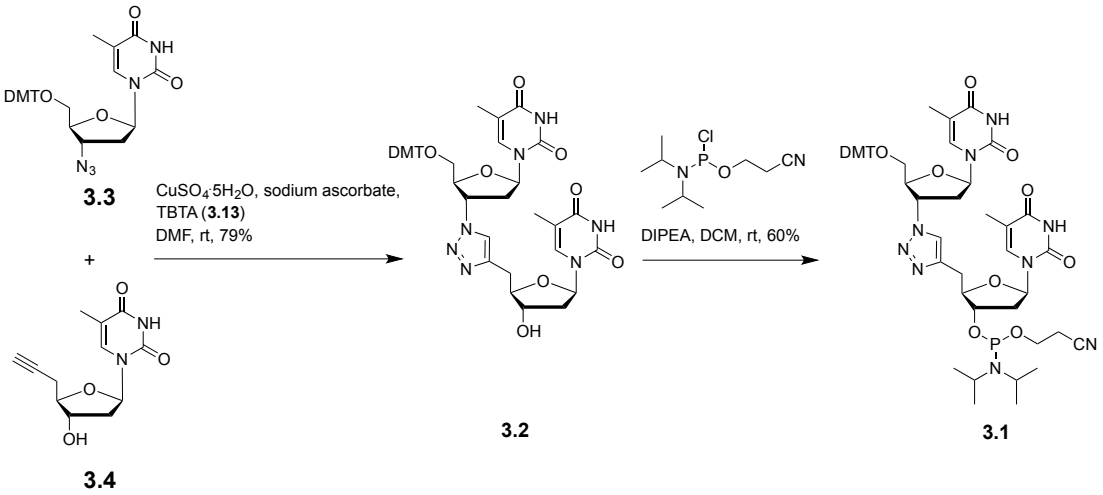


Scheme 3-3: Synthetic route to 5′-ethynyl-5′-deoxythymidine fragment **3.4**

3.3.4 Completion of the synthesis of T-short-triazole-T dimer 3′-phosphoramidite 3.1

The synthesis of triazole dimer phosphoramidite **3.1** was performed as described in *Scheme 3-4*. The 5′-DMT T-short-triazole-T dimer **3.2** was first obtained in a good yield of 79% *via* a CuAAC reaction between 3′-azido and 5′-ethynyl nucleosides **3.3** and **3.4**. Sodium ascorbate was used as an *in situ* reducing agent to generate the Cu(I) catalyst from the Cu(II) source. Polytriazoles as tris-hydroxypropyl triazole or tris-(benzyltriazolylmethyl)amine (TBTA) ligand **3.13** are often used in CuAAC reactions. The former is more suitable for reactions performed in aqueous solutions, while the latter is more desirable for use in organic solvents. The use of the TBTA ligand improves the reaction by binding to the copper (I), which stabilises the +1 oxidation state and prevents disproportionation to Cu(II) or Cu (0).[165]

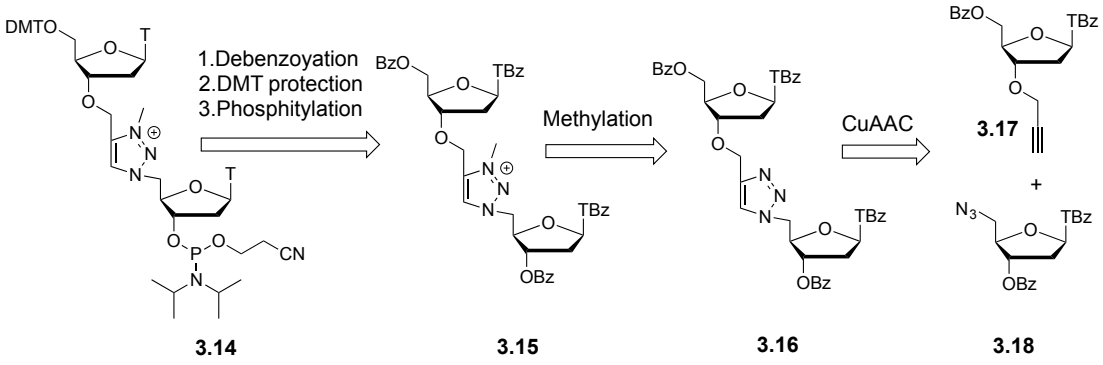
Following this, 5′-DMT protected T-short-triazole-T dimer **3.2** was successfully phosphitylated to afford the targeted dinucleoside phosphoramidite **3.1**, in a moderate yield of 60%.



Scheme 3-4: Synthesis of 5′-DMT T-short-triazole-T dimer 3′-phosphoramidite **3.1**

3.3.5 Retrosynthesis for preparation of T-T dimer phosphoramidite with methyl-triazole linkage

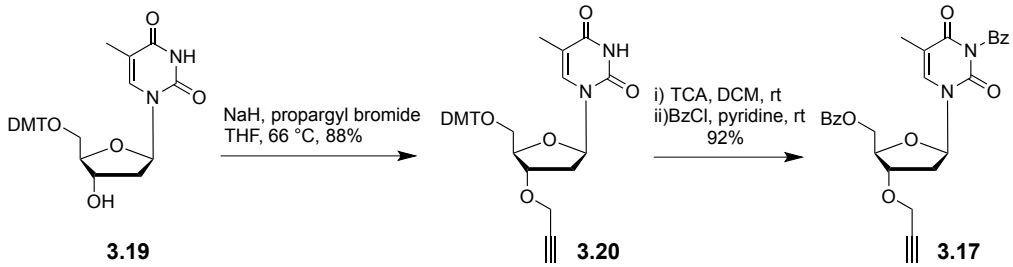
The synthesis of T-methyl-triazole-T dimer 3′-phosphoramidite **3.14** was envisaged from precursor **3.15** through debenzoylation, protection of the 5′-hydroxyl group as a dimethoxytrityl ether and phosphitylation. The methyl triazolium intermediate **3.15** could be synthesised *via* alkylation on the 1,2,3-triazole aromatic ring of thymidine dimer **3.16**. Protection of the 5′ and 3′ alcohols, as well as of the thymine ring N-3 atoms of dimer **3.16**, was required to avoid methylation at these positions. These protecting groups were installed on each 3′-*O*-propargyl and 5′-azido nucleoside fragments, **3.17** and **3.18**, prior to the CuAAC reaction that enabled formation of intermediate **3.16** (***S**cheme 3-5*).



***S**cheme 3-5:* Retrosynthetic strategy for 5′-DMT T-methyl-triazole-T dimer 3′-phosphoramidite building block **3.14**

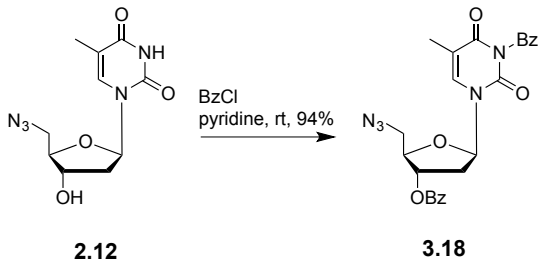
3.3.6 Synthesis of benzoyl protected 3′-*O*-propargyl and 5′-azido thymidine fragments 3.17 and 3.18.

Synthesis of the desired 3-*N*-benzoyl-3′-*O*-propargyl-5′-*O*-benzoyl-thymidine **3.17** was performed as described in *Scheme 3-6*. In a first step, 5′-dimethoxytrityl-thymidine **3.19** was alkylated to afford the 3′-*O*-propargyl nucleoside **3.20** in a good yield of 88%. Acid mediated cleavage of the DMT protecting group of nucleoside **3.20** was then carried out, followed by immediate protection with benzoyl chloride to give the targeted compound **3.17** in an excellent yield of 92%.



Scheme 3-6: Synthetic route to 3-*N*-benzoyl-3′-*O*-propargyl-5′-*O*-benzoyl-thymidine **3.17**.

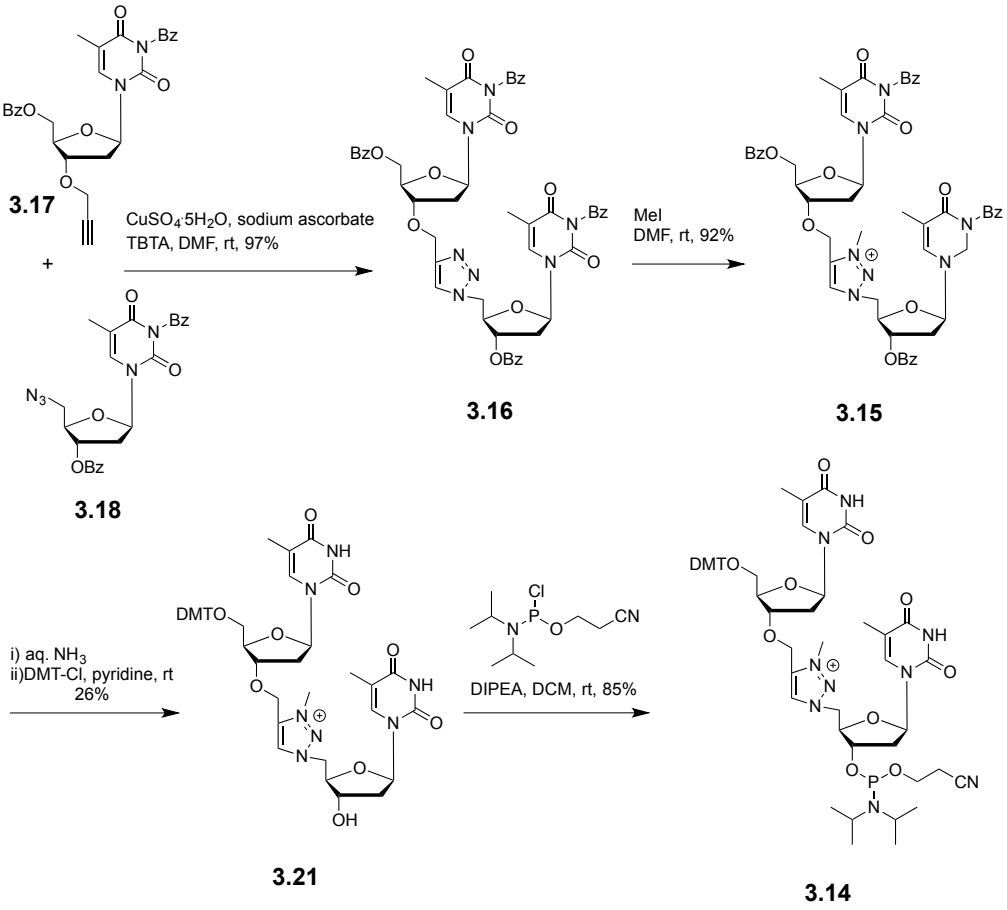
The 5′-azido thymidine precursor **2.12** was synthesised following the route described in Chapter 2 section 2.3.3. The latter was subsequently treated with an excess of benzoyl chloride to afford the targeted benzoyl protected 5′-azido thymidine nucleoside **3.18** in an excellent yield of 94% (*Scheme 3-7*).



Scheme 3-7: Synthesis of 3-*N*-benzoyl-3′-*O*-benzoyl-5′-azido-5′-deoxythymidine **3.18**

3.3.7 Synthesis of T-methyl-triazole-T 3'-phosphoramidite dimer 3.14

With 3'-*O*-propargyl and 5'-azido thymidine fragments **3.17** and **3.18** in hand, the CuAAC reaction step could be performed, which resulted in the formation of T-triazole-T dimer **3.16** in an excellent yield of 97%. Then, treatment of the dinucleoside **3.16** with an excess of methyl iodide afforded the methyl triazolium dimer **3.15** in 92% yield. The benzoyl protecting groups of dinucleoside **3.15** were first cleaved off by treatment with concentrated aqueous ammonia, and the resulting compound was immediately protected at the 5'-position with DMT-Cl to give intermediate **3.21** in 26% yield over 2 steps. The low yield of these synthetic steps is due to the difficult purification of the highly polar methyl triazolium dimer **3.21.** This caused a significant loss of compound on column chromatography and will be optimised in the future. Finally, phosphitylation of DMT protected dimer **3.21** gave the targeted T-methyl-triazole-T 3'-phosphoramidite dimer **3.14** in a good yield of 85% (*Scheme 3-8*).

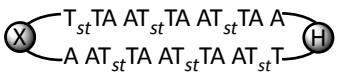


Scheme 3-8: Synthetic route to T-methyl-triazole-T 3'-phosphoramidite dimer **3.14**

3.3.8 Synthesis of triazole-backbone modified oligonucleotides analogues

Triazole modified oligonucleotides of different sequences were synthesised using phosphoramidite dimer building blocks **3.1** and **3.14**. Oligonucleotides **ODN-3.1** to **ODN-3.3** were designed to study the effect of a single 1,2,3-triazole modification **3.III** and **3.IV** on the structure and thermodynamic stability of the corresponding duplexes and compare it to our previous triazole linkage **3.II** (*Table 3-1*). Oligonucleotides containing several additions of short-triazole linkage **3.III** (**ODN-3.4** and **ODN-3.6**) were also synthesised for reasons that will be developed in section 3.4.2.

Table 3-1: Sequences of oligonucleotides used for biophysical studies. Insertions of triazole backbone **3.II**, short-triazole backbone **3.III** and methyl-triazole backbone **3.IV** are represented respectively by: *t, st* and *Met*; X: triazole linkage; H: hexaethylene glycol

Code	oligonucleotide sequence (5' to 3')
ODN-3.1	CGACGT$_t$TTGCAGC
ODN-3.2	CGACGT$_{st}$TTGCAGC
ODN-3.3	CGACGT$_{Met}$TTGCAGC
ODN-3.4	hexynol-T$_{st}$TAAT$_{st}$TAAT$_{st}$TAA-**H**-T$_{st}$TAAT$_{st}$TAAT$_{st}$TAA-aminoC7
ODN-3.6	

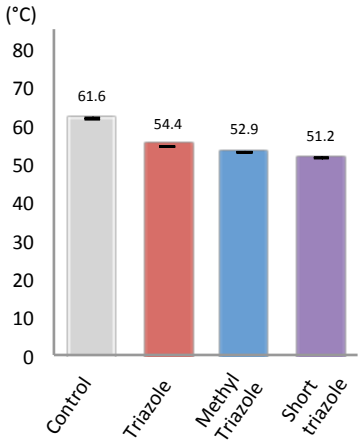
3.4 Biophysical studies on triazole modified oligonucleotides

3.4.1 Comparison of thermal stability and duplex structure of oligonucleotides containing, triazole, short-triazole and methyl-triazole linkages

The melting temperatures (T_m) of duplexes formed between oligonucleotides **ODN-3.1** to **ODN-3.3** with a complementary DNA strand was measured and compared to that of the control unmodified DNA duplex **ODN-3.I** (*Figure 3-13*). The presence of the triazole linkage **3.II** within the DNA duplex induced a lost in thermal stability of 7.2 °C compared to control DNA duplex **ODN-3.I**, which was consistent with our previously published results.[155] Introduction of a positive charge on this

triazole backbone, as in the case of methyl-triazole backbone **3.IV**, did not dramatically change the duplex stability and DNA duplexes **ODN-3.3** and **ODN-3.1** have fairly similar melting temperatures (ΔTm = -1.5 °C, *Figure 3-13*).

As previously mentioned, Isobe *et al.* reported that a fully modified short-triazole DNA strand had an improved binding affinity for a DNA target (+ 41.1 °C) compared to a canonical DNA strand. However, in the case of oligonucleotide **ODN-3.2** the short triazole linkage **3.III** induced a destabilisation of -10.4 °C compared to the unmodified DNA duplex **ODN-3.I** (*Figure 3-13*). The stronger structural perturbation induced by the single incorporation of the short-triazole linkage in the middle of oligonucleotide **ODN-3.2** might explain this difference in thermal stability results.



Code	Type of modification	ΔTm (°C)
ODN-3.I	Phosphodiester	
ODN-3.1	Triazole (**3.II**)	-7.2
ODN-3.3	Methyl triazole (**3.IV**)	-8.7
ODN-3.2	Short triazole (**3.III**)	-10.4

Figure 3-13: Comparison of thermal stability of oligonucleotides containing 1,2,3-triazole modified-backbones against control unmodified DNA. ΔTm = T_m (modified oligonucleotide) - T_m (unmodified oligonucleotide)

Despite the relative thermodynamic instability caused by these three triazole modifications, DNA duplexes **ODN-3.1** to **ODN-3.3** adopt a classic B-DNA helix structure, as confirmed by circular dichroism analysis (*Figure 3-14*).

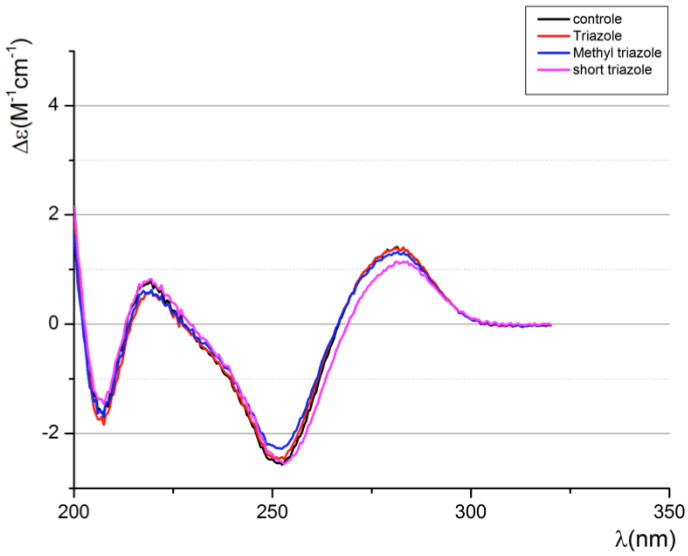


Figure 3-14: Circular dichroism spectra of unmodified and backbone-modified triazole, short-triazole and methyl-triazole duplexes. All triazole-backbone modified oligonucleotide analogues adopt the B-DNA conformation.

3.4.2 Ultra-violet melting study of oligonucleotides containing multiple additions of short-triazole linkage

Based on the previous observation that a single incorporation of the short-triazole linkage **3.III** is detrimental for the DNA duplex stability, it was of interest to study if several additions would on the contrary improve the binding affinity for a complementary ssDNA. The effect of alternating triazole/phosphodiester linkages was therefore investigated with the idea that the structural periodicity of the alternating pattern would be beneficial for the duplex stability. Moreover, such chimeric oligonucleotide structures could be potentially used in antisense or siRNA applications as they display a reduced overall charge, which can increase the binding affinity for a nucleic acid target and improve penetration through the cell membrane. The alternating triazole backbone should also enhance protection against 3' and 5' exonucleases while maintaining the phosphodiester linkages required to trigger degradation of the targeted mRNA upon binding, *via* a RNase H dependent mechanism. [166] Oligonucleotide **ODN-3.4** was thus synthesised to form an intramolecular hairpin duplex containing alternating short-triazole linkages (*Table 3-2*). The 3'-amino linker of this oligonucleotide was then labeled with azidohexanoic NHS ester, to obtain a 3'-azido functionality. Both extremities of

oligonucleotide **ODN-3.4** were subsequently joined *via* a copper click reaction, thus leading to the formation of cyclic duplex **ODN-3.6** (*Chapter 6, section 6.2.3.1*). It was indeed envisaged that the cyclisation of the alternating triazole modified oligonucleotide **ODN-3.6** would bring extra stabilisation to the duplex structure.[167]

Unfortunately the instability caused by the short-triazole linkages **3.III** was very important in the alternating triazole/phosphodiester hairpin oligonucleotide **ODN-3.4** and the very low melting temperature of this chimeric construct could not be measured. The contribution in stabilisation usually brought by the cyclisation of this hairpin, did not compensate the large loss of stability induced by the short-triazole modifications **3.III**. UV-melting studies of oligonucleotide **ODN-3.6** showed a non-cooperative transition suggesting that a stable DNA duplex could not be formed (*Table 3-2*). Interestingly, upon addition of the minor groove DNA binder distamycin (1 eq.) the cyclic short-triazole oligonucleotide **ODN-3.6** was able to form a duplex with a much higher melting temperature (48.6 °C).

This preliminary study was performed using AT sequences and improvement of the above results may still be achieved by varying the type of sequences (increased GC content) and the number and position of short-triazole modifications.

Table 3-2: Comparison of the thermodynamic stability of hairpin and cyclic short-triazole (**3.III**) duplexes with their unmodified analogues. *st*: short triazole backbone; H: hexaethylene glycol; X: triazole linkage

Code	Sequences	Tm (°C)
ODN-3.4	5'–T$_{st}$TA AT$_{st}$TA AT$_{st}$TA A–(H) 3'–A AT$_{st}$TA AT$_{st}$TA AT$_{st}$T–	<15
ODN-3.5	5'–TTAATTAATTAA–(H) 3'–AATTAATTAATT–	57.9
ODN-3.6	(X)–T$_{st}$TA AT$_{st}$TA AT$_{st}$TA A–(H) A AT$_{st}$TA AT$_{st}$TA AT$_{st}$T–	<15
ODN-3.7	(X)–TTAATTAATTAA–(H) AATTAATTAATT–	72.5

3.5 Conclusion and outlook

To conclude, short-triazole and methyl-triazole linkages **3.III** and **3.IV** have been purposely designed to be structural variations of our previous triazole linkage **3.II**, which is an excellent phosphodiester mimic *in vitro* and *in vivo*. Two synthetic routes have been successfully developed to access two novel phosphoramidite thymidine dimers featuring internucleosidic linkages **3.III** and **3.IV**. The use of these phosphoramidite dinucleoside building blocks allowed the straightforward synthesis of oligonucleotides containing triazole modifications **3.III** and **3.IV** at desired positions.

Ultra-violet melting studies revealed that the two structural variants, short-triazole **3.III** and methyl-triazole **3.IV**, did not dramatically change the duplex thermal stability compared to triazole linkage **3.II**. The latter appeared to be the least duplex destabilising, closely followed by the methyl-triazole linkage **3.IV** and finally by the short-triazole linkage **3.III.** Circular dichroism analysis confirmed that all modified triazole DNA adopted a standard B-DNA shape. The relationship between the biophysical properties of these 1,2,3-triazole linkages and their ability to be correctly read-through by DNA polymerase will be investigated in the next chapter.

A preliminary investigation of the hybridisation properties of oligonucleotides containing alternating short-triazole linkages **3.III** revealed that these specific constructs could not form stable duplexes. These types of chimeric oligonucleotides could nonetheless be useful for antisense and siRNA applications and it will be interesting to pursue this investigation with other short-triazole modified oligonucleotide sequences.

Chapter 4
Study of the biocompatibility of modified DNA backbones for gene synthesis applications

In order to assess the biocompatibility of DNA modifications, such as artificial internucleoside linkages, DNA polymerase read through efficiency and fidelity must be tested. This chapter thus begins by briefly presenting past and current methods used for the precise determination of the DNA primary structure. Enzymatic studies, including linear copy and Real-Time PCR using two modified oligonucleotide templates containing respectively a 1,4-disubstituted 1,2,3-triazole and a 1,3,4-trisubstituted 1,2,3-triazole is then presented and compared to templates containing a triazole modification previously utilised in our group. This is followed by a report of the sequencing results obtained using DNA templates containing each one of the aforementioned modified triazole backbones and an additional amide-modified backbone.

4.1 DNA sequencing: from early methods to next-generation sequencing

As described in the Introduction *Chapter 1*, development of novel chemical DNA ligation methods for gene synthesis requires the formation of an artificial internucleosidic linkage mimicking the natural phosphodiester bond. This last feature is crucial for both *in vitro* and *in vivo* use of such synthetic genes. For preliminary validation of a new artificial linkage, the products from replication of the linkage should be sequenced, the details of which are described in this chapter.

DNA sequencing involves the precise determination of DNA primary structure *i.e.* the order of the four bases, A, G, C, and T that constitute a nucleic acid strand. Since the development of the first practical DNA sequencing methods in the mid-1970, DNA sequencing has rapidly evolved and has greatly accelerated discoveries in biology and medicinal research; it has become indispensable in numerous fields such as medical diagnosis[168], virology[169,170] genomics[171] and forensics[172,173], and is of crucial importance for the development of new chemical gene synthesis strategies[88].

4.1.1 First sequencing methods

One of the first widely used methods for DNA sequencing was developed in 1977 by Maxam, Gilbert and co-workers.[174] This method was based on a two-step process involving preliminary chemical modification of the nucleobases followed by DNA cleavage at the site of modification. In order to detect these DNA fragments, Maxam-Gilbert sequencing first requires radioactive ^{32}P labelling of 5′ end of the DNA to be sequenced. After purification of the radiolabelled DNA, the sample is subject to four different reactions that modify various combinations of nucleobases. The first reaction methylates the N7 position of guanine (G) using dimethyl sulfate; the second reaction induces depurination of A and G using formic acid; the third reaction involves pyrimidines (C and T) hydrolysis using hydrazine; the final reaction selectively cleaves cytidine using sodium chloride and hydrazine. Following these reactions, piperidine treatment catalyses the cleavage of the phosphodiester internucleosidic bond at the abasic site. The different reactions can be then loaded on a polyacrylamide gel and the labelled fragments resolved by electrophoresis.

Elucidation of the base sequence is achieved by the interpretation of the gel banding pattern for the four chemical reactions. For example, a band in the lanes for "G only reaction" and "G + A reaction" will correspond to G. On the other hand, if a band is present in the "G + A reaction" but not in the "G only reaction" this will correspond to an A base. The same rational can be used for the two other chemical reactions (*Figure 4-1*). 200 to 300 bases could thus be determined by Maxam-Gilbert sequencing, however this method has several drawbacks, namely the use of a large amount of radioactive and toxic material as well as complex technical methods and chemical transformations.

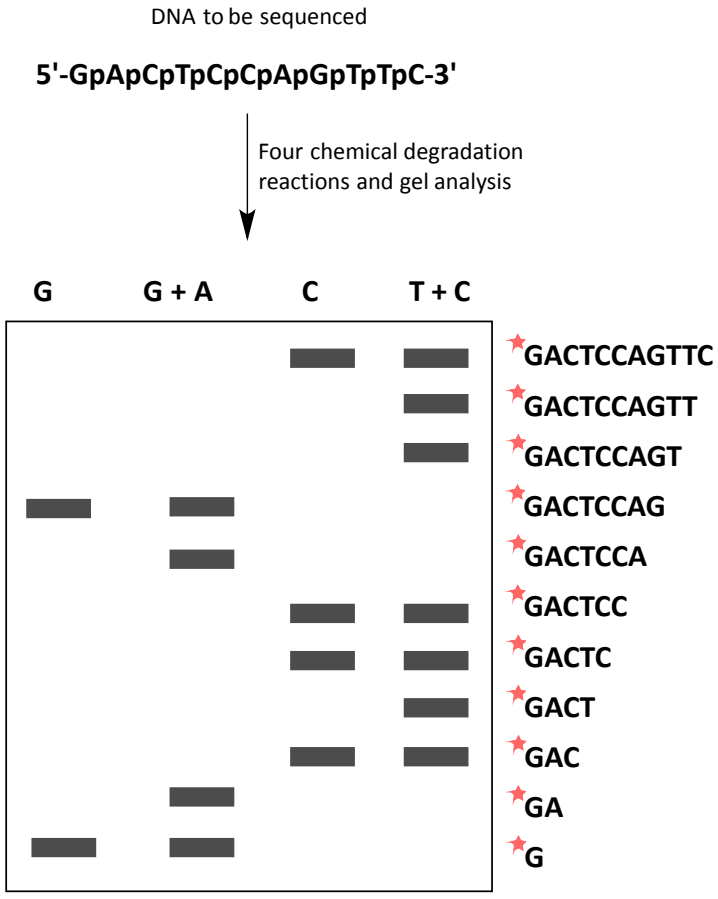


Figure 4-1: Overview of the Maxam-Gilbert sequencing method.[174] A star represents the radioactive marker.

Around the same time, Frederick Sanger and co-workers developed an alternative "chain-termination" sequencing method. Thanks to its ease of use and

reproducibility, this technique quickly became the sequencing method of choice until the mid 2000s. Sanger sequencing is based on the extension of a primer by a DNA polymerase, which is then terminated by the insertion of a dideoxynucleoside triphosphate (ddNTP) (*Figure 4-2*).[175]

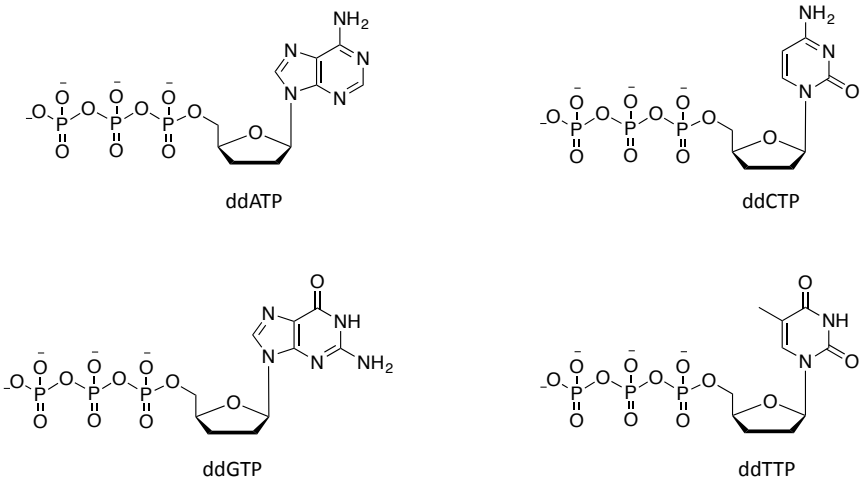


Figure 4-2: The structure of dideoxynucleoside triphosphates used in Sanger sequencing.

During the Sanger sequencing process, the DNA to be sequenced and its complementary labelled primer, are divided into four sequencing reactions. Each reaction contains a specific dideoxynucleoside triphosphate, ddATP, ddCTP, ddGTP or ddTTP alongside normal dNTPs. When a ddNTP is incorporated by the DNA polymerase, chain elongation is stopped as there is no 3'-hydroxyl for further reaction. This ultimately leads to truncated DNA fragments of different length, each of which is terminated by one of the four ddNTPs. These fragments can then be resolved by polyacrylamide gel electrophoresis and the primary sequence of the DNA inferred (*Figure 4-3*).

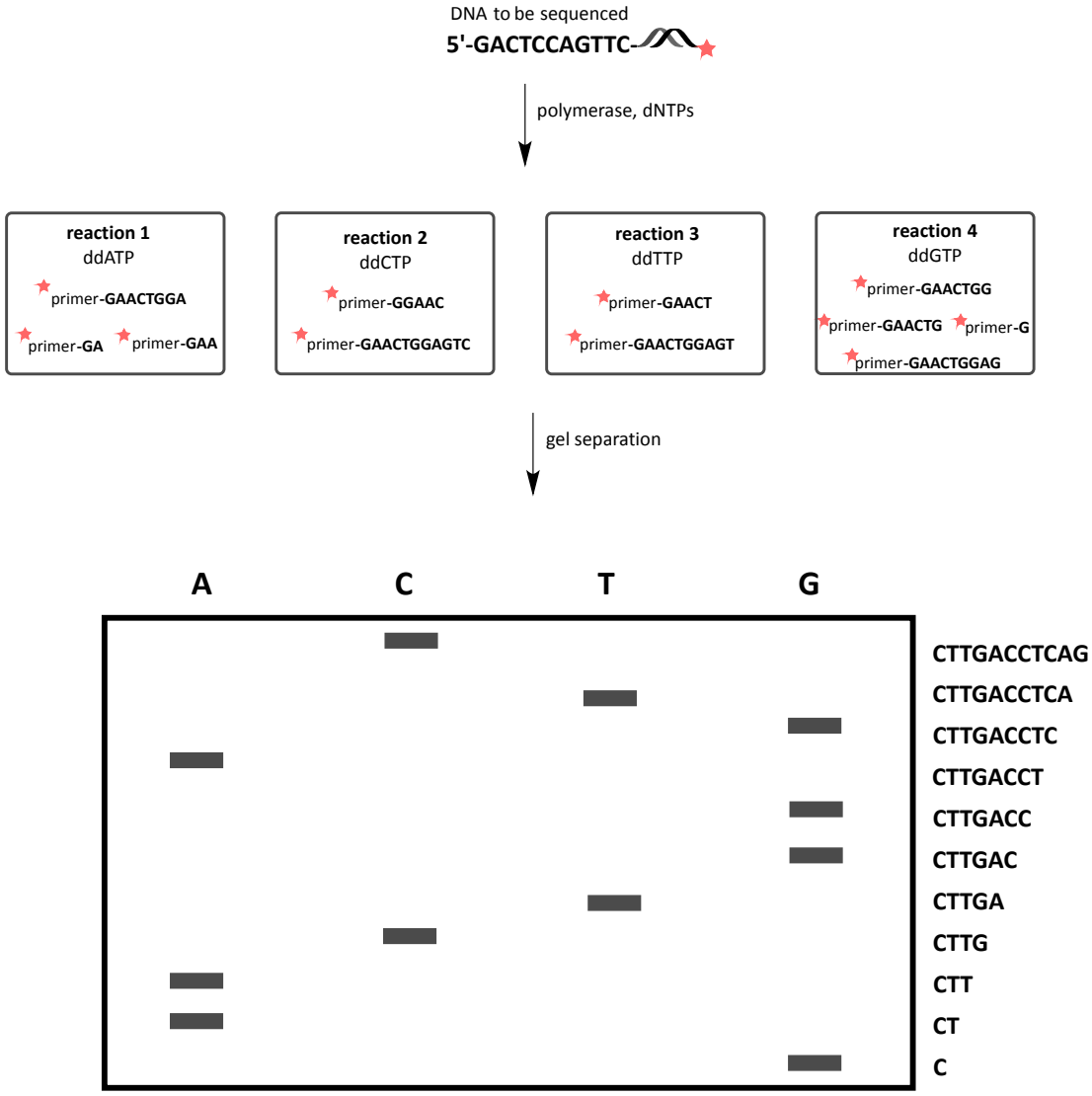


Figure 4-3: Sequencing of DNA using the Sanger method. The sequence fragments loaded on the gel are the complements of the actual template.

An important development in chain termination technology involved the use of fluorescently labeled ddNTPs, thus reducing the number of sequencing reactions to one, as each ddNTP termination could be easily identified by a specific color.[176] Further refinements involved analysis of the fragments by capillary gels instead of flatbed gels, which facilitated higher throughput (*Figure 4-4*).[177] Used in combination, these improvements set the stage for automated, high-throughput DNA sequencing, which allowed much more efficient sequencing at a lower cost ultimately enabling, in 2001, sequencing of the human genome for the first time.[178,179]

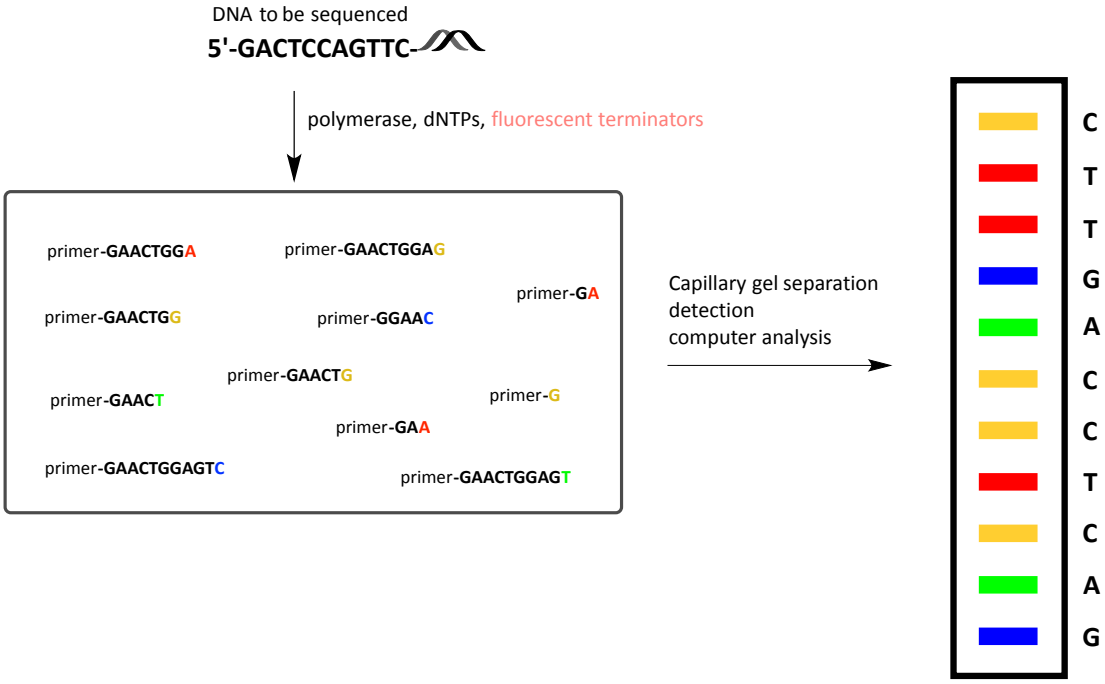


Figure 4-4: Fluorescent automated Sanger sequencing

4.1.2 Next generation sequencing: Illumina® technology

With the desire to reduce the cost of DNA sequencing, new technologies that could produce thousands or millions of sequencing reads in parallel were developed in the early 2000s,[5, 180] the most widely adopted of which is the Illumina® platform, originally known as Solexa sequencing.

For this methodology, purified DNA must be fragmented into smaller ~300-500 bp pieces and tagged with specific 5' and 3' adapters. These 'libraries' are then attached to the Illumina® flow cell through hybridisation of the 5' or 3' adaptor to a complementary oligonucleotide pre-bound to a solid support. Following this step, cluster generation begins through a process called 'bridge amplification'. A DNA polymerase moves along the attached DNA strand creating a dsDNA, which is then denatured. The original template is washed away while the newly synthesised strand remains attached to the solid support (*Figure 4-5*). This strand folds over and binds to the other 5' or 3' complementary oligonucleotide present on the flow cell.

The DNA replication process occurs once more and after denaturation of the dsDNA bridge, two single stranded DNA copies are now tethered on the flow cell, the first one being the forward strand and the second one being the reverse strand. This process is repeated multiple times and occurs simultaneously for millions of clusters resulting in the clonal amplification of the DNA fragments (*Figure 4-6*).

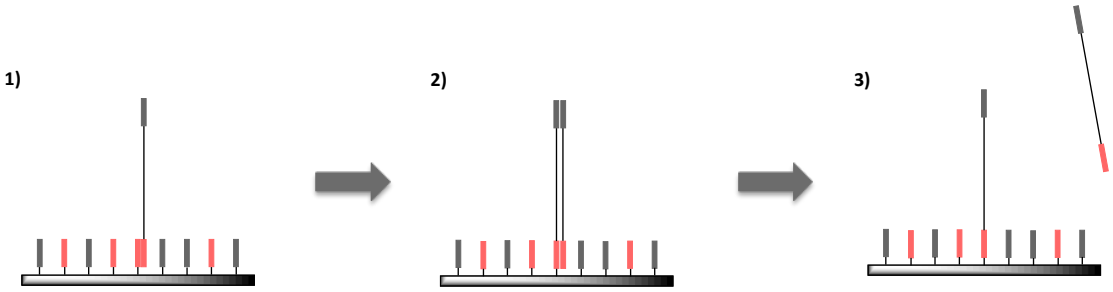


Figure 4-5: Tethering of the oligonucleotides to the flow cell. 1) The oligonucleotide fragments bind to the oligonucleotides attached to the flow cell. 2) The bound DNA strands are replicated. 3) The dsDNA is denatured and the original strand is washed away.

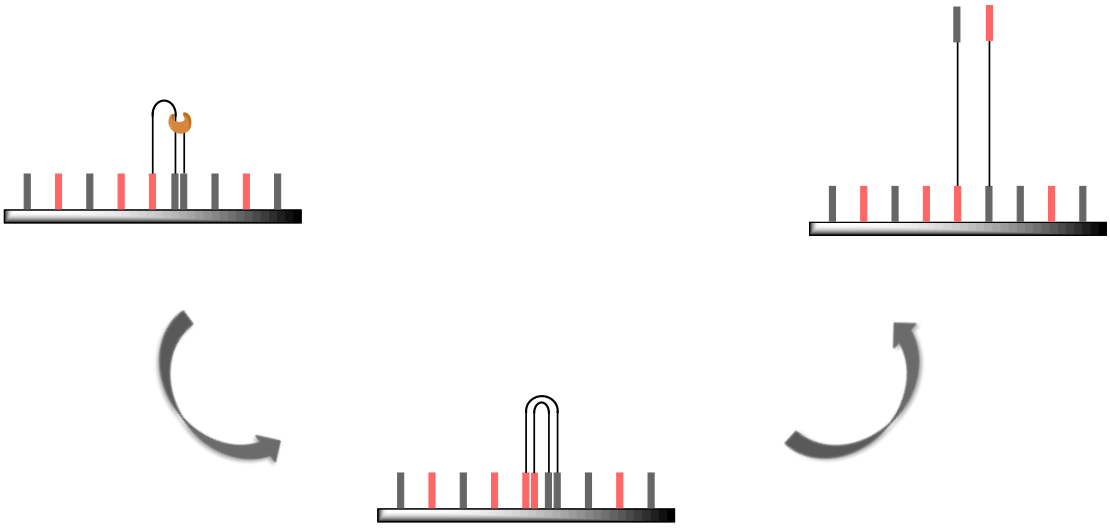


Figure 4-6: The bridge amplification process. The DNA strand bends over and binds to an oligonucleotide attached to the flow cell. The DNA is replicated to form a dsDNA bridge, which is denatured to obtain to single DNA strands bound to the support.

Following cluster formation, sequencing begins by the binding of a sequencing primer. One of four distinct fluorescent dNTP is then added depending

upon base complementarity. These dNTPs contains a reversible 3'-terminator such that only one nucleotide can be incorporated. The characteristic emission wavelength of the added base is then detected and the reversible terminator is removed allowing the addition of the next nucleotide. This cycle is repeated until the maximum length of sequencing, currently *ca.* 300 bases, is reached (*Figure 4-7*). During this process, multiple strands are read simultaneously and millions of clusters are sequenced in a massively parallel process.

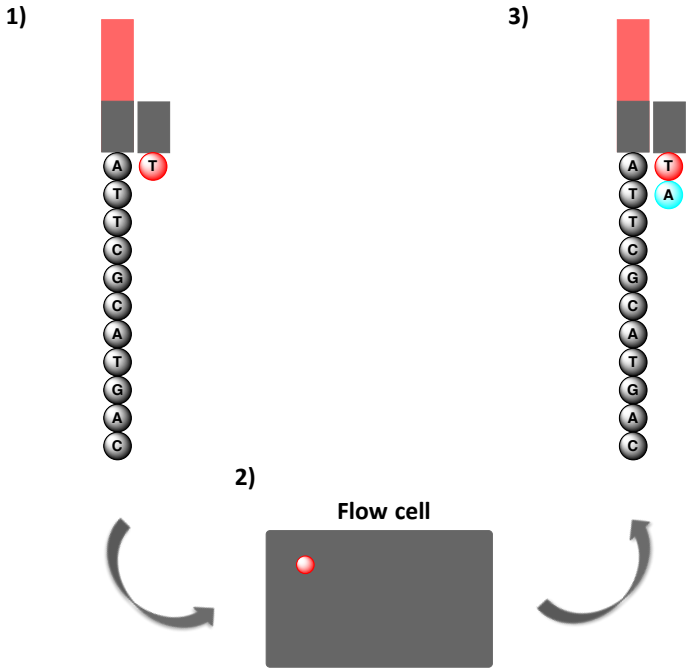


Figure 4-7: The sequencing by synthesis process. 1) Sequencing primer binds to the adapter and the first nucleotide is added. 2) The clusters are excited with a laser and the fluorescence emitted is detected with a camera. 3) The 3'-terminator is cleaved and the next nucleotide is added.

4.2 Aim of the study

Following the synthesis of oligonucleotides containing triazole modifications **3.III** and **3.IV** as well as of amide-containing oligonucleotide analogues (*Figure 4-8*), the objective was to correlate the previous biophysical studies with their potential ability to be read-through by DNA polymerases. Enzymatic experiments involving linear extension and polymerase chain reaction (PCR) were therefore carried out. The products obtained from these enzymatic reactions were then sequenced *via* Illumina® next-generation sequencing to further evaluate the recognition of these modified backbones (**3.II** to **2.III**) by different DNA polymerases. The aim was to compare these enzymatic results with our previously published results on the biocompatibility of 1,2,3-triazole linkage **3.II.** It was of interest to find out if structural variations of this 1,2,3 triazole linkage changes the polymerase reading-through ability. Secondly, it was relevant to study if the unusual biocompatibility of the latter was intrinsic to its particular chemical structure. The investigation of the polymerase reading-through of the amide linkage, which displays a very dissimilar chemical structure was investigated for this purpose. All together the aim was to provide further insight on how such a heavily modified internucleosidic linkage (**3.II**) is correctly accepted in the replication machinery.

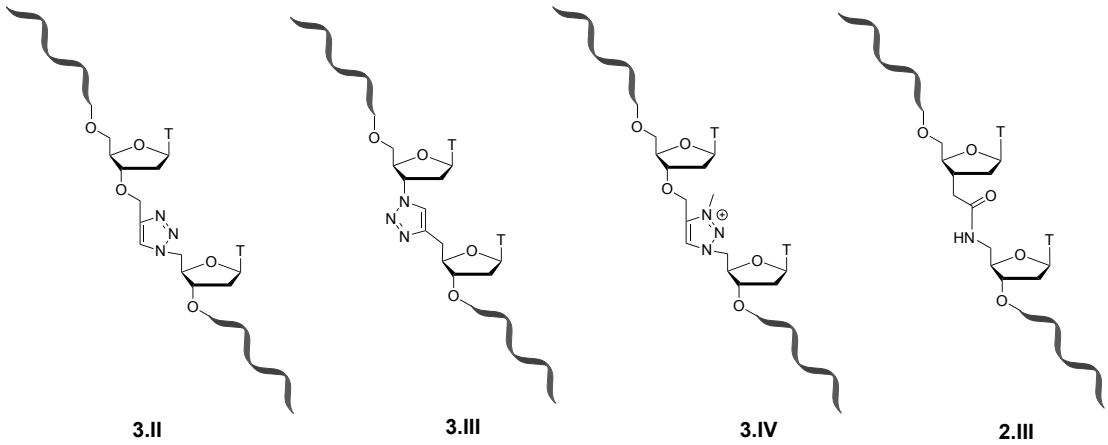


Figure 4-8: Structures of triazole (**3.II**), short-triazole (**3.III**), methyl-triazole (**3.IV**) and amide backbone (**2.III**) modifications.

4.3 Studies on the accuracy and efficiency of read-through of modified DNA backbones by DNA polymerases

First, in order to evaluate the enzymatic properties of the different modified linkages, oligonucleotides containing these modifications were synthesised (*Table 4-1*). Oligonucleotides **ODN-4.1** to **ODN-4.3** have identical sequences but contain different triazole linkages, and were used as templates in linear copy experiments. The oligonucleotide primer for these studies, **P-4.1**, contains a 6-carboxyfluorescein functional group attached to its 5'-end for ease of visualisation during gel electrophoresis. Canonical and modified oligonucleotides **ODN-4.4** to **ODN-4.8** were used as PCR templates and contain different sequences at their forward and reverse primer binding regions, in order to avoid cross contamination and false positive results during PCR amplification. These distinct primer sequences also serves as additional barcodes for analysis of the sequencing data obtained on the Illumina® MiSeq platform.

Table 4-1: Sequence of oligonucleotides used for enzymatic reactions. Insertions of triazole **3.II**, short-triazole **3.III**, methyl-triazole **3.IV** and amide **2.III** backbones, are represented respectively by: *t, st, Met* and *am.* Primer regions are represented in red and green; FAM is 6-carboxyfluorescein.

Code	oligonucleotide sequence (5' to 3')
P-4.1	FAM-GGTTATGTGTGTCGGCAG
ODN-4.1	CTCAGTC**T$_t$T**CTGTACTGCCGACACACATAACC
ODN-4.2	CTCAGTC**T$_{st}$T**CTGTACTGCCGACACACATAACC
ODN-4.3	CTCAGTC**T$_{Met}$T**CTGTACTGCCGACACACATAACC
ODN-4.4	ACGTTAGCACGAAGAGGCATCTTAGCACACAATCTCACAC TCTGGAATTCACACTGACAATACTCGCGAACACACCCAAT
ODN-4.5	GCTTACGACGAAGAACGGATC**T$_t$T**AGCACACAATCTCAC ACTCTGGAATTCACACTGACAATACCTCGGCCAATACACACA
ODN-4.6	GCGTACGAGAACCGAATGATC**T$_{st}$T**AGCACACAATCTCACAC TCTGGAATTCACACTGACAATACGTCGCACACAACAATCC
ODN-4.7	GCATTCGAGCAACGTAAGATC**T$_{Met}$T**AGCACACAATCTCAC ACTCTGGAATTCACACTGACAATACTGCCGACACACATAACC
ODN-4.8	AGCAACAGCTGGTAGACGATC**T$_{am}$T**AGCACACAATCTCACAC TCTGGAATTCACACTGACAATACACCACAAATCCGTGCCA

4.3.1 PCR and real-time PCR experiments of DNA templates containing triazole modified backbones using GoTaq DNA polymerase

In a first experiment, the biocompatibility of the methyl triazole linkage was evaluated by PCR amplification of template **ODN-4.7**. Analysis of the reaction by agarose gel electrophoresis showed a positive amplification band (*Chapter 6, section 6.2.4.2*). In early PCR cycles, replication of the modified methyl-triazole template would give rise to an unmodified DNA product, which could be exponentially amplified in subsequent cycles. Since the reaction is only monitored at the end point of PCR, it is possible that the reaction may appear efficient even if read-through of the artificial linkage (in the first PCR cycles) is a rare event. The effect of the modified triazole linkages (**3.II** to **3.IV**) on the replication by GoTaq® polymerase were consequently monitored more rigorously by real-time PCR and compared to an unmodified control template.

In a real time PCR experiment, fluorescence is used to monitor the accumulation of the DNA PCR product with each cycle of amplification. Common fluorescent reporter molecules include dyes that bind to the double-stranded DNA (SYBR Green) or sequence specific probes that generate fluorescence (Molecular Beacons or TaqMan® Probes).[181] A positive real-time PCR amplification curve can be divided into three major phases (*Figure 4-9*). The linear ground phase also called the PCR initiation phase occurs during the first PCR cycles where the fluorescence generated is too low and cannot be distinguished from the baseline. The amount of fluorescence then reach a threshold where it is significantly higher than background level, which shows the beginning of the exponential phase. During this phase, there is an exponential increase in fluorescence reflecting the doubling of PCR product after each cycle. Finally, the plateau stage is reached when reaction components become limited. The cycle threshold (C_T) is determined by the intersection of the PCR amplification curve with a fluorescence threshold line and correlate to the number of DNA copies in the reaction sample.

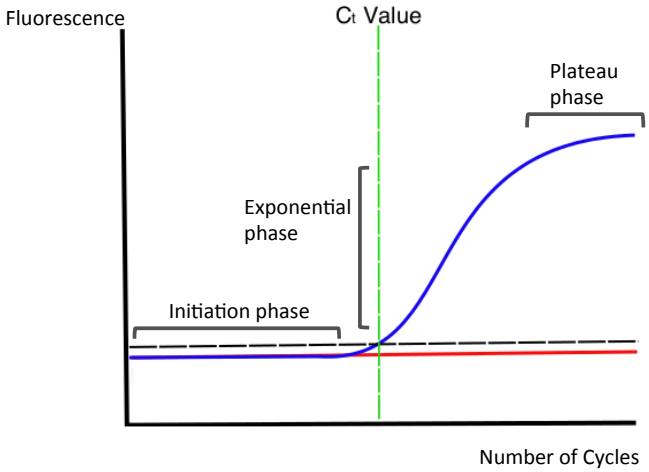


Figure 4-9: Example of a real-time PCR amplification curve of a DNA sample showing the phases of the process.

In a first experiment, four real-time PCR reactions were simultaneously performed with the same input (5 ng) of unmodified and modified templates, **ODN-4.4** to **ODN-4.7.** An arbitrary fluorescence threshold was fixed for all reactions and was chosen such that it is sufficiently higher than background fluorescence and enabled the measure of a C_T value for each PCR reactions (*Table 4-2*).

Table 4-2: Measured C_T values from real-time PCR amplification curves (*Figure 4-10*) using triazole-modified and unmodified templates.

Code	Modification	C_T
ODN-4.4	None	10.1
ODN-4.5	Triazole (**3.II**)	14.3
ODN-4.6	Short triazole (**3.III**)	12.5
ODN-4.7	Me-triazole (**3.IV**)	15.5

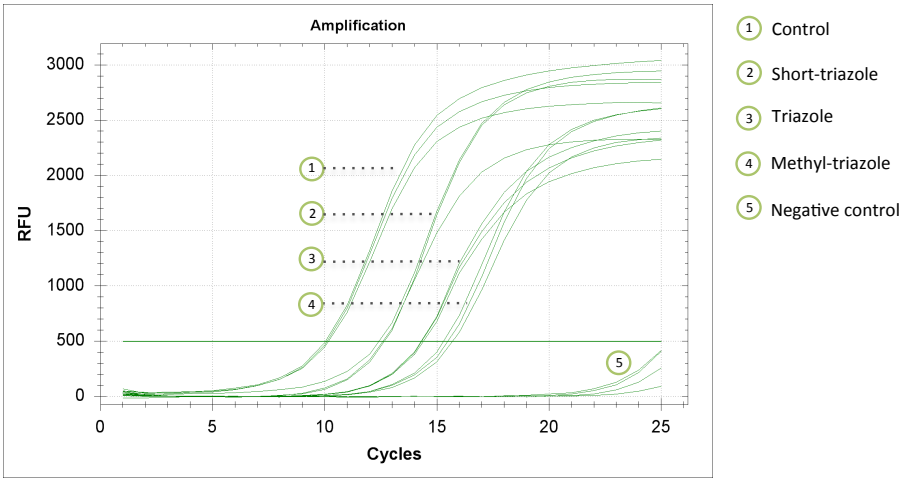


Figure 4-10: PCR amplification curves of modified triazole and un-modified templates. Three technical replicates of each PCR reactions were performed. The fluorescence threshold was set at 497.84 RFU and C_T values were calculated by averaging cycle threshold of three amplification curves.

A rightward shift in the amplification curves of all modified templates compared to the unmodified template curve is observed. This shift could originate from a slower replication through all templates containing triazole-modified linkages at the beginning of the PCR process. The observed C_T value for amplification of the template containing the triazole backbone **3.II** was higher than the value obtained for triazole backbone **3.III**, suggesting that reading through the former linkage was slower than reading through the short triazole. It could also be observed that replication through the methyl triazole was the most significantly inhibited (*Table 4-2*). However, these are preliminary results that require further validation by varying the amount of modified and unmodified templates used.

It was also observed that the amplification curves for the modified templates (curves 2, 3 and 4, *Figure 4-10*) showed similar exponential phases as the control amplification curve. Suggesting that despite the fact that the GoTaq replication through the modified triazole template **ODN-4.5** to **ODN-4.7** is slowed down at the beginning of the PCR reaction, unmodified DNA copies are still produced, which lead to a dilution of the original modified template after each PCR cycle. This will ultimately results in the successfully exponential synthesis of unmodified DNA copies.

In the future, it might be possible to reduce the prolonged lag phase observed at the beginning of the PCR amplification of all modified triazole templates (*Figure 4-10*), by varying reaction parameters such as magnesium ions concentrations, time intervals and temperatures of the PCR steps.

4.3.2 Modified DNA template linear copying experiments

To further assess the read-through efficiencies of the different modified triazole backbones, primer extension experiments using templates **ODN-4.1** to **ODN-4.3** were performed. It was of interest to determine if commercially available DNA polymerases could tolerate the different modified backbones. Two A-family polymerases were selected, the mesophilic polymerase Large Klenow fragment and the thermophilic polymerase from *Thermus aquaticus* (GoTaq®). A B-family DNA polymerase, *Thermococus kodakaraensis* (KOD XL) was also chosen.

The large Klenow fragment of E.Coli DNA polymerase I retains a 3'➔5' exonuclease activity but has lost 5'➔3' exonuclease activity. Linear copy experiments using Klenow large fragment DNA polymerase were performed at 37 °C, which is its optimum temperature of activity. Polyacrylamide gel electrophoresis (PAGE) analysis and mass spectrometry, confirmed that the triazole linkage **3.II** was successfully read through and that the correct full-length product was obtained. These results corroborate our previous findings,[153] though the reaction was much slower and required 30 minutes to reach completion (*Figure 4-11* B, Lane 3). Previous results where the triazole linkage **3.II** was surrounded by two cytidines reported full extension of the primer in five minutes. Conversely, linear extension of primer **P-4.1** using the template containing the short-triazole linkage **3.III** proceeded very quickly and was completed within five minutes (*Figure 4-11* A, Lane 2). The extension product obtained was one nucleotide shorter than the full-length product and mass analysis confirmed that this product was missing one adenine nucleotide.

The presence of the methyl-triazole linkage **3.IV** within the template greatly inhibited replication by Klenow large fragment DNA polymerase. The replication was mainly stopped before or at the triazole insertion site (confirmed by

mass spectroscopy, *Figure 4-11*, C) and did not proceed further even after 3 hours (*Figure 4-11* B, Lanes 1).

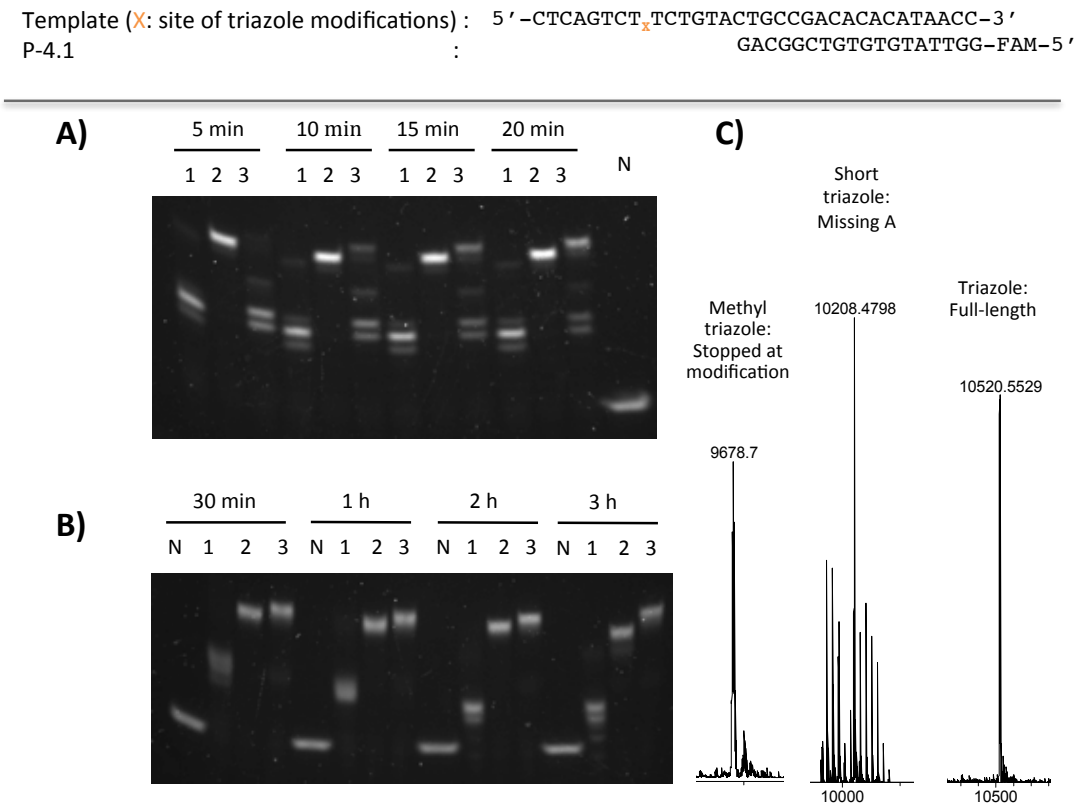


Figure 4-11: Linear extension using Klenow large fragment DNA polymerase at 37 °C analysed by denaturing PAGE A) Primer extension reaction after 5, 10, 15 and 20 minutes. B) Primer extension reaction after 30 min, 1, 2 and 3 hours. Lanes 1 - methyl triazole template (**3.IV**); Lanes 2 - short-triazole template (**3.III**); Lanes 3 - triazole template (**3.II**); Lanes N - primer **P-4.1**. C) Mass analysis of extension product for methyl-triazole, short-triazole and triazole linkages.

Linear extension experiments with KOD XL and GoTaq® DNA polymerases were both performed at 60 °C, which is 15 °C lower than the enzymes optimum temperature for activity but is required due to the lower T_m of the primers. KOD XL is a blend of KOD polymerase and a mutant form of KOD which is deficient in 3'➔5' exonuclease activity. Primer extension reactions with KOD XL showed similar results compared to the results obtained with Klenow large fragment. An exception could be observed for the linear extension of the template containing the short-triazole

backbone **3.III**, which this time showed a less intense band corresponding to the full-length product (*Figure 4-12* A, Lanes 2).

Finally for GoTaq® DNA polymerase, primer extension results were noticeably different to the ones presented above. All three templates required an extended reaction time of 20 hours to reach completion (*Figure 4-12* B) with a significant accumulation of the truncated product that stops before the triazole insertion site at time points up to 3 hours. Interestingly, for template **ODN-4.3** that contains the methyl-triazole linkage (**3.VI**), incorporation of bases past the triazole linkage could be observed, although the extension product is shorter than the full-length extension product (*Figure 4-12* B, Lanes 1h, 2h and 3h).

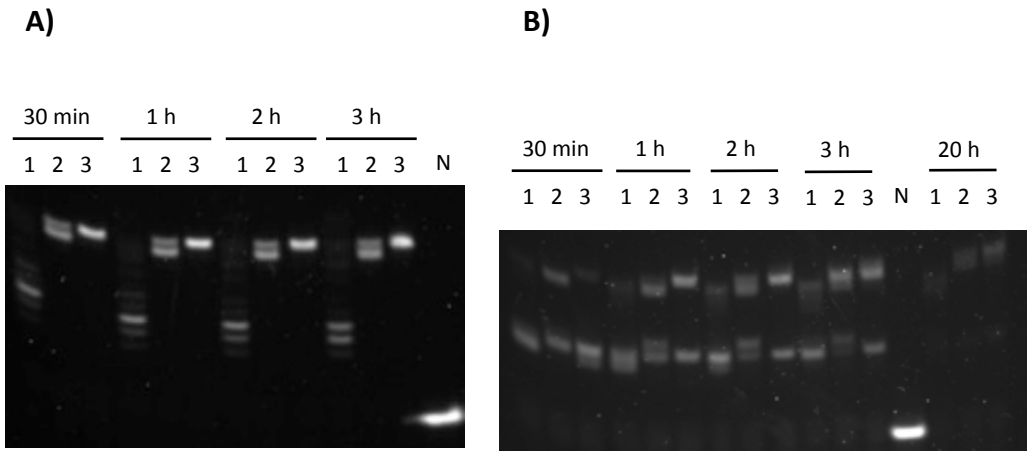


Figure 4-12: Linear extension reactions using KOD XL (A) and GoTaq® (B) DNA polymerases at 60 °C analysed by denaturing PAGE. Lanes 1 - methyl-triazole template (**3.IV**); Lanes 2 - short-triazole template (**3.III**); Lanes 3 - triazole template (**3.II**); Lanes N - primer **P-4.1**.

4.3.3 Summary of linear copying results

The triazole linkage **3.II** is successfully read through by Klenow large fragment and KOD XL DNA polymerases, and in all cases the full-length product was obtained. However the reaction seems to be generally much slower than our previous reports.[153] As previously mentioned in Chapter 3, the triazole backbone **3.II** when surrounded by two thymidines, induces a loss in thermal stability of 7.2 °C.

Considering that Klenow large fragment and KOD XL DNA polymerases both exhibit 3'➔5' exonuclease activity, it could be envisaged that when these polymerases incorporate a dATP in front of the first thymidine before the triazole, the destabilisation caused by this modification induces the transfer of the extended primer from the polymerase domain to the 3'➔5' exonuclease domain. After excision of the most recently incorporated nucleotide, the polymerase activity will be resumed but will be again followed by excision due to the destabilisation of the primer/modified-template duplex.[182] This back and forth process could account for the observed slow down of the linear copy reactions and could end when the modified duplex adopts a more favourable conformation. This hypothesis is supported by previously published results obtained using a template where the triazole linkage **3.II** is flanked by two cytidines instead of two thymidines. The stronger G:C base pairing around the triazole modification increases the stability of the latter and consequently improved the speed of the primer extension reaction.[153] For the linear extension using the template containing triazole backbone **3.II** and GoTaq® DNA polymerase, the full-length product was also obtained but the reaction was much slower than with the above-mentioned polymerases. Unlike Klenow large fragment and KOD XL, Taq DNA polymerases do not exhibit 3'➔5' exonuclease activity. A possible explanation for the significant slowdown of the primer extension reaction could be thermodynamic or structural destabilisation generated by the triazole backbone modification **3.II** which leads to the dissociation of the enzyme from the template-primer duplex.[183]

Linear copying experiments showed that the short-triazole linkage **3.III** was quickly read through by Klenow large fragment and KOD XL polymerases. These results were in accordance with these obtained by real-time PCR in section 4.3.1 as amplification of the template containing the short-triazole backbone **3.III** proceeded faster than with modified templates containing triazole and methyl-triazole linkages. However, it should be noted that in the case of linear extension reactions performed with Klenow large fragment, only an extension product missing one adenine nucleotide could be detected. Linear copying experiments using KOD XL DNA polymerase afforded a mixture of the full-length product and of the shorter product missing one adenine nucleotide.

Finally, the presence of the methyl-triazole linkage **3.IV** within the DNA template generally induced pausing of the primer extension at the modification site. On very rare occasion, Klenow large fragment and GoTaq® polymerases could replicate past this modification to afford a minor amount of extension product missing few nucleotides.

4.4 DNA sequencing results obtained with triazole modified templates

Following the above PCR and linear copying experiments of templates containing triazole modifications (**3.II to 3.IV**), the accuracy of replication through these modified linkages was evaluated by next-generation sequencing. The GoTaq® PCR amplification products obtained from the modified templates **ODN-4.5** to **ODN-4.7** were therefore sequenced on an Illumina® MiSeq instrument. In order to determine the consistency of these results and assess the fidelity of other more natural polymerases, linear extension products using Klenow large fragment and modified templates **ODN-4.5** and **ODN-4.6** were also sequenced. Sequencing of the replicated strand revealed the errors associated with copying by Klenow large fragment through the modified linkage whilst sequencing of the modified template strand demonstrated the errors associated with the Illumina 9° N polymerase used in cluster generation on the MiSeq sequencing platform (*Figure 4-13*). The same experiments were carried out using the amide-modified template **ODN-4.8,** and the unmodified template **ODN-4.4** that served as a control for comparison of sequencing results. It should be noted that primer extension using methyl-triazole template **ODN-4.7** and Klenow large fragment polymerase could not be performed as the efficiency of the reaction was too low; the template strand was therefore directly sequenced on the Illumina Miseq by annealing it to a synthesised complementary strand.

Libraries were generated using a PCR-free approach thus enabling direct comparison of different polymerases described above. Library preparation and sequencing data analysis were performed by Dr Arun Shivalingam in our research group.

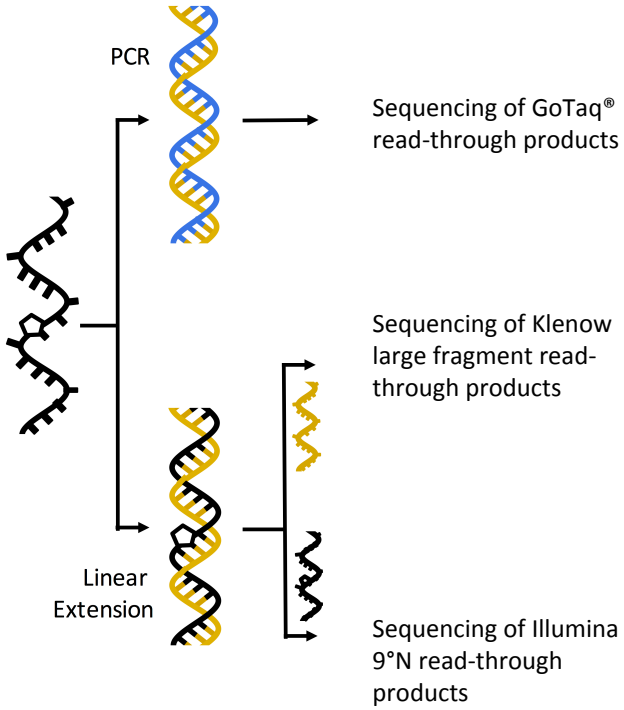


Figure 4-13: Sequencing strategy to assess fidelity of three DNA polymerases in processing triazole modified linkages. Libraries for sequencing were prepared using an Illumina® TruSeq PCR-free kit.

4.4.1 Sequencing of GoTaq® PCR products, Klenow large fragment linear copy and 9°N polymerase read through products obtained with the unmodified control template

Figure 4-14 shows global alignment of all sequence reads to a known template. The frequency of the observed bases is then highlighted at sites adjacent to the modified (or in this case control unmodified) internucleosidic linkage. As expected, the control phosphodiester unmodified template **ODN-4.4** was correctly read through by the three DNA polymerases. From these results, it can be seen that thymidine 22 (and other bases in the oligonucleotide) are not always observed 100% of the time. The origin of these errors may be due to the fact that the oligonucleotides used in this study are prepared by solid phase synthesis as opposed to naturally derived DNA fragments. The coupling yield of each phosphoramidite monomer to an extending oligonucleotide chain is not 100% and will lead to a small percentage of base deletion. These small errors cannot be detected using mass spectrometry or other analytical techniques (HPLC, capillary gel electrophoresis) but

are uncovered by sequencing due to the large number of reads (~20,000-100,000 reads per PCR product). On average, for the control oligonucleotide, the expected base was only observed 99.7% of the time for all three polymerases.

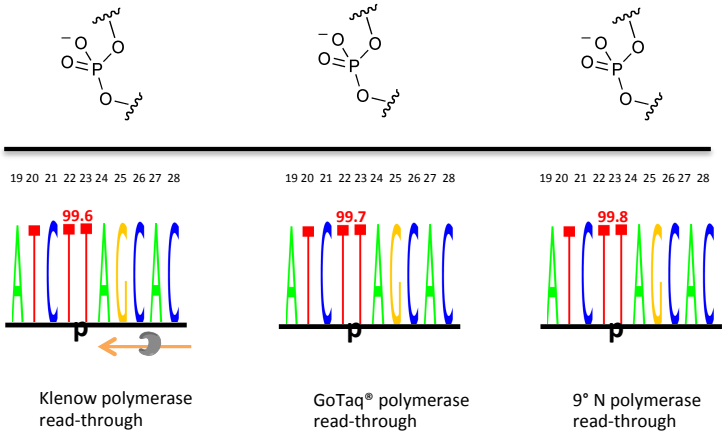


Figure 4-14: Sequencing results obtained with unmodified template **ODN-4.4** and putting in evidence error related with oligonucleotide automated solid phase synthesis. The arrow indicates the direction of the DNA polymerase extension of the original template.

4.4.2 Sequencing of GoTaq® PCR products and Klenow large fragment linear copy products

The 1,2,3-triazole linkage **3.II** was correctly read through on most occasions by GoTaq® DNA polymerase, which is in accordance with the linear copying results presented above in section 4.3.2. It was observed that the fidelity of replication was reduced compared to the unmodified phosphodiester linkage, as a mismatch incorporation of a cytidine after the triazole was observed in 2.6% of cases and one of the thymidine around the triazole linkage was skipped by the polymerase in 10.2% of the products (*Figure 4-15*).

Concerning the short-triazole backbone **3.III**, copying of one thymidine at the modified-triazole site was missed in 93% of the cases. In the remaining 7%, this thymidine was faithfully copied. It also appears that the presence of this linkage affects replication two bases downstream from the triazole position (*Figure 4-15*).

Replication through the methyl-triazole linkage **3.IV** gave rise to a major fraction of PCR products were the two thymidines surrounding the methyl-triazole linkage were not copied. In the remaining fraction the two bases after the T-methyl-triazole-T modification, were also not copied (*Figure 4-15*).

Surprisingly, the DNA template **ODN-4.8** containing amide-modified linkage **2.III** was successfully read through with a fidelity similar to that observed for the natural phosphodiester linkage, contradicting the results previously obtained by Kuwahara and co-workers (*Figure 4-15*).[143] Indeed they reported that primer extension reactions using KOD(exo-) DNA polymerase and a modified template containing the 3'-CH_2-CO-NH-5' amide modification **2.III** did not proceed efficiently as the full length product could not be obtained. They concluded that even though the DNA polymerase could pass beyond the modified backbone that its negative effect on downstream extension was too significant.

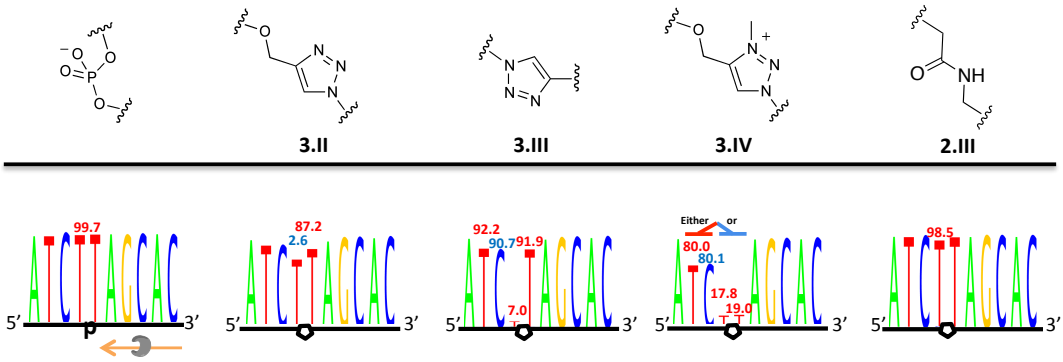


Figure 4-15: Sequencing results of PCR products obtained with GoTaq® polymerase using the Illumina® MiSeq platform. The arrow indicates the direction of the DNA polymerase extension of the original templates **ODN-4.4** to **ODN-4.8.**

Replication of the triazole linkage **3.II** by Klenow large fragment (*Figure 4-16*) was more accurate than with GoTaq® polymerase (*Figure 4-15*). One thymidine around the triazole linkage was not copied in 1.5% of the PCR products, the remainder were identical copies of the template (*Figure 4-16*) .

The short-triazole modification **3.III** was read as a single thymidine more frequently than for GoTaq® reactions, and more promisingly the bases past this modification were successfully replicated (*Figure 4-16*).

Replication of the amide linkage **2.III** using Klenow large fragment DNA polymerase was slightly less accurate than with GoTaq® polymerase and a thymidine surrounding the triazole linkage was skipped in 5.3% of the PCR products (*Figure 4-15* and *Figure 4-16*).

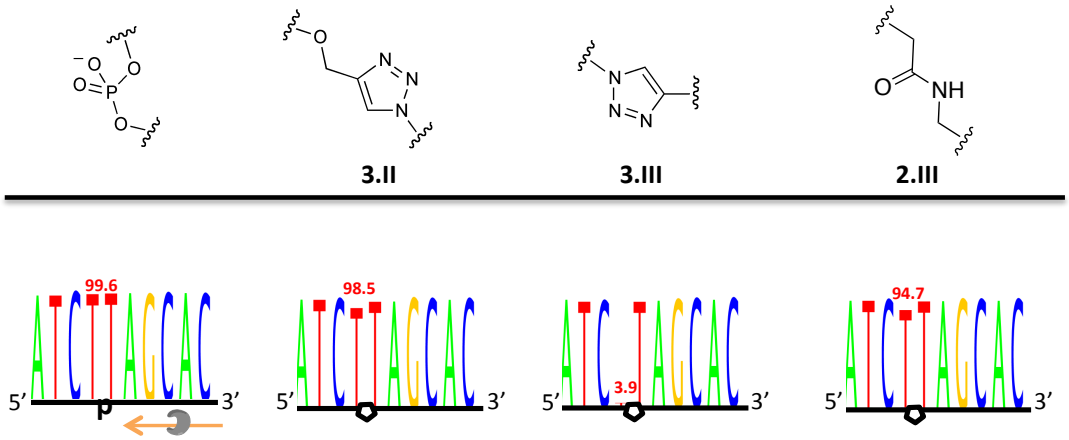


Figure 4-16: Sequencing of products obtained *via* linear extension with Klenow large fragment using the Illumina® MiSeq platform. The arrow indicates the direction of the polymerase extension using templates **ODN-4.4** to **ODN-4.8.**

4.4.3 Sequencing of products read-through by Illumina® 9 °N polymerase

In order to obtain these sequencing results, double-stranded DNA products generated through Klenow large fragment linear extension, are ligated to the Illumina® adaptors (provided in the TruSeq DNA library preparation PCR free kit). One strand is the modified DNA template whilst the other is the unmodified Klenow-generated strand. The dsDNA are then tethered to the flow cell and clonally amplified through bridge amplification using a 9°N DNA polymerase and proprietary undisclosed extension conditions. Sequencing of the modified-template strand DNA

copies is a direct measure of the errors associated with the replication of the Illumina® 9°N polymerase through the triazole and amide modified linkages.

It was observed that the 1,2,3-triazole linkage **3.II** was read through by the Illumina® 9°N polymerase with a reduced fidelity compared to Klenow large fragment (*Figure 4-16*), with one of the thymidine around the triazole being correctly incorporated in 86.3% of the cluster DNA copies (*Figure 4-17*). However, replication of this linkage by the 9°N polymerase was more accurate than with GoTaq® polymerase (*Figure 4-15*) as mutagenesis was not observed in this case (*Figure 4-17*).

Interestingly, the short-triazole linkage **3.III** appeared to be more tolerated by the 9°N polymerase than with the two previously studied DNA polymerases. Indeed, in this case the two thymidines surrounding the short-triazole linkage were faithfully copied in 65.2 % of the products (*Figure 4-17*), whereas these two bases were correctly incorporated in 7% of the DNA copies for GoTaq® and in 3.9% for Klenow large fragment (*Figure 4-15* and *Figure 4-16*).

The read-through sequencing results obtained for methyl-triazole linkage **3.IV** was similar to the one obtained with GoTaq® with the two thymidines around the methyl-triazole linkage being missed for around 86% of the replication products. Bases further away from the methyl-triazole linkage are this time correctly replicated (*Figure 4-17*).

Finally, the amide linkage **2.III** was also well accepted by the 9°N polymerase and 1.7% of deletion of one thymidine was observed (*Figure 4-17*).

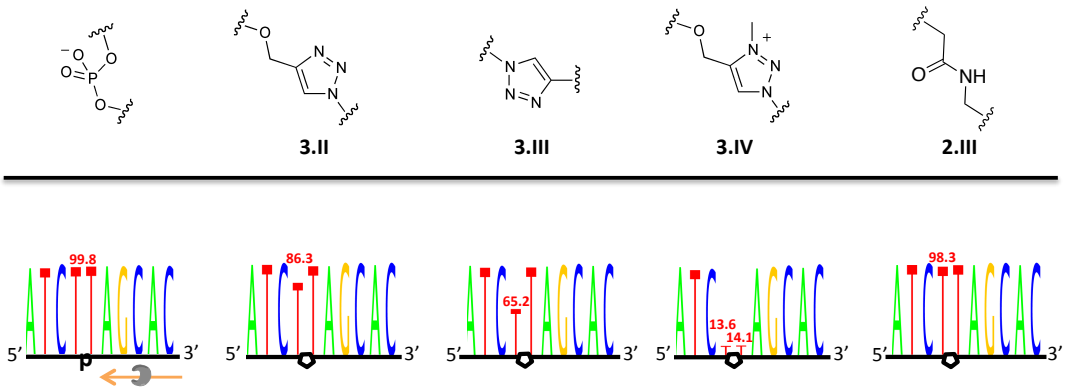


Figure 4-17: Sequencing of 9°N DNA products obtained during Illumina® MiSeq bridge amplification of modified and unmodified DNA templates **ODN-4.4** to **ODN-4.8**. Direction of 9°N polymerase extension is represented by an arrow.

4.5 Discussion of modified DNA read-through studies using different DNA polymerases

Based on the enzymatic studies (real-time PCR, linear extension) and the sequencing results, a general trend in the polymerase read-through properties of the modified triazole linkages can be elucidated. The triazole linkage **3.II** is read-through by all studied polymerases with a reduced speed compared to the unmodified phosphate linkage or the short-triazole linkage **3.III**. However this is compensated by good fidelity in the replication process, especially with the Klenow large fragment DNA polymerase.

The short-triazole linkage **3.III** is quickly read-through and one of the thymidine around the triazole linkage is repeatedly skipped. Rationalisation of this phenomenon is difficult as several factors could be involved. The first one could be the thermodynamic stability of the template-primer duplex that might be improved by pushing out one of the thymidines around the triazole linkage out of the DNA duplex. This base may then no longer serve as a template during replication and would be missed. Indeed as demonstrated in Chapter 3, the short-triazole is the most duplex-destabilising modification among the three 1,2,3-triazole linkages and induces a loss of stability of around 10.4°C. Another factor could be related to the intrinsic characteristic of the DNA polymerases. The sequencing and linear copying

results clearly demonstrated that KOD XL and 9°N polymerases, were better at replicating through the short-triazole linkage than other polymerases studied. It was previously showed that these archaeal polymerases belonging to sequence family B, are more efficient in utilising modified nucleotides than are DNA polymerases from sequence family A, such as Taq DNA polymerase.[184]

One of the most interesting results arises from the methylation of the triazole **3.II**. It is clear that introduction of this small modification on the nitrogen N3 completely destroys the biocompatibility of this triazole linkage. Indeed, replication through methyl-triazole linkage **3.IV** was a very rare event with all DNA polymerases tested (GoTaq®, KOD XL and Klenow) and always led to deletion of the two thymidines surrounding the triazole backbone. This supports our previous proposal that the N3 triazole nitrogen is a hydrogen-bond acceptor surrogate for the oxygen of the natural phosphate linkage (*Figure 4-18*). Indeed, it was previously demonstrated through NMR solution analysis that this nitrogen occupies a similar position to that of one of the non-bridging oxygen. [155]

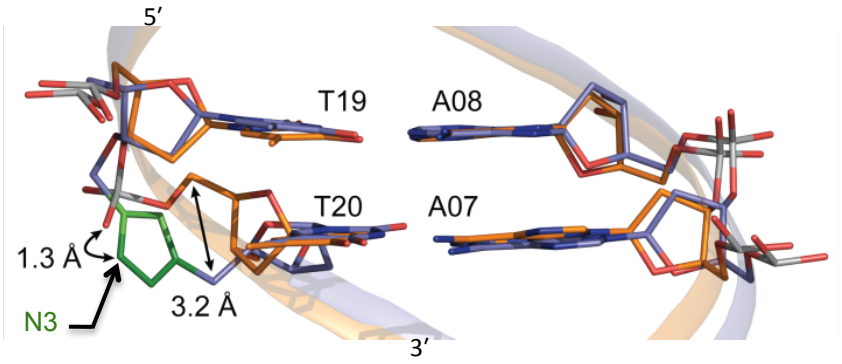


Figure 4-18: overlay of the average structure of natural duplex (orange) and modified duplex (blue) determined by NMR solution. From A. Dallmann, T. Brown *et al.*[155]

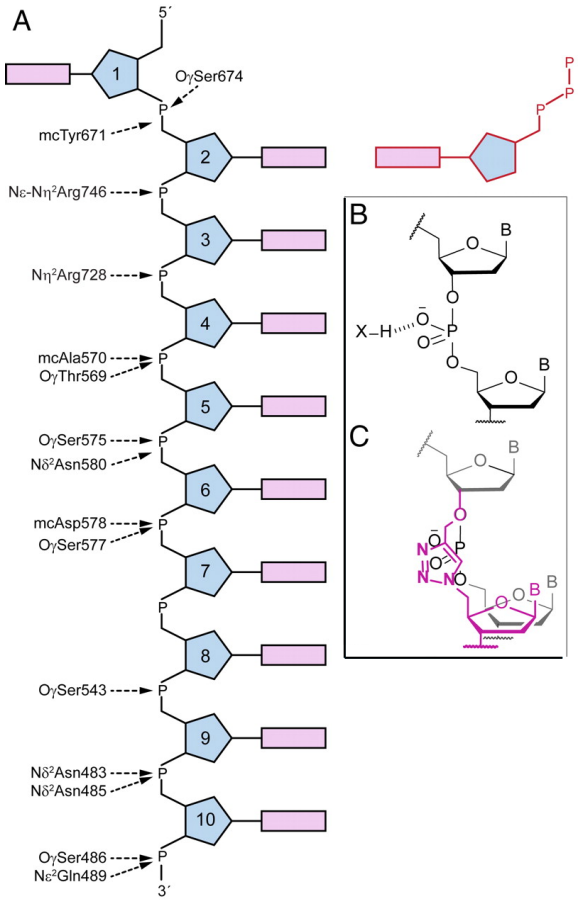


Figure 4-19: Schematic diagram of interactions between KlenTaqI and the DNA template in the closed form.[153] A) only interactions with the phosphodiester linkage are shown and only the template strand is represented. B) The majority of the interactions with the polymerase involves the branched phosphate oxygen, and none of the bridging oxygen atoms; C) Overlay of natural DNA with triazole showing that nitrogen atoms N2 and N3 could substitute for the phosphate oxygen atoms.

It has been previously demonstrated that one or more hydrogen bonds between the oxygen of the phosphate and DNA polymerases is required for the binary DNA-enzyme complex to bind to the incoming nucleoside triphosphate (*Figure 4-19*).[21, 23, 24]

The suppression of these interactions, by using a ethyl phosphotriester backbone for example, has been shown to slow down the DNA polymerase reading-through compared to the unmodified phosphodiester.[185] The methyl-triazole linkage further demonstrates the importance of the contact of the phosphate oxygen with the DNA polymerase during the replication process. However, it is likely that the loss

of hydrogen bonding is not the only chemical effect involved in the reduced biocompatibility of the methyl-triazole linkage. Steric effects should also be considered and this will require a more thorough study of the structure of DNA duplexes containing this linkage.

Finally sequencing results have surprisingly demonstrated that the amide linkage **2.III** can be successfully read through by GoTaq®, Klenow large fragment and 9°N DNA polymerases with quite high fidelity. This backbone modification is known to only slightly destabilise dsDNA (*see Chapter 2*) and as importantly, can form hydrogen bonds via the carbonyl H-bond acceptor and the NH H-bond donor of the amide function. These results consequently further highlights the parameters that seems to be necessary for the acceptance of a modified backbone by DNA polymerases, which are namely a good thermal stability and the ability to form hydrogen-bonds in the DNA-enzyme complex.

4.6 Conclusion and outlook

In summary the biocompatibility of two newly synthesised 1,2,3-triazole linkages and of an amide linkage has been evaluated and compared to the previously published triazole backbone **3.II**. It is important to notice that for synthetic practical reasons, all these studies were performed with a T-triazole-T linkage, which will be the most destabilising configuration of bases. Previous studies of triazole linkage **3.II** were performed with MeC-triazole-T and MeC-triazole-C and consequently showed a better thermal stability and faster reading through by the DNA polymerase (GoTaq and Klenow).[151]

Among the 1,2,3-triazole backbones, triazole **3.II** appears to be the closest mimic of the phosphodiester linkage, as it is the most faithfully replicated by all the DNA polymerases that have been tested. This backbone is a good candidate for chemical gene synthesis applications and strategies to improve the read-through efficiency are currently under investigation. Surprisingly, the amide modification has shown a remarkably good biocompatibility with DNA polymerases and seems to be accepted almost as well as the natural phosphodiester linkage. It is a potentially useful linkage for DNA ligation provided that the amide bond can be efficiently made in a templated reaction between a carboxylic acid and an amine modified oligonucleotides. It is also important that the coupling reagent does not cause mutagenesis by reacting with the nucleobases, even at a low level. All this parameters will have to be investigated in the future.

From the enzymatic studies on the short-triazole linkage, it also appears that family B DNA polymerases can better tolerate this chemical modification, which is the most duplex-destabilising among the three 1,2,3-triazole linkages. Recent advances in polymerase engineering associated with the need of developing artificial genetic systems for applications in medicine, biotechnology and nanotechnology,[187] could soon give rise to classes of polymerases that could fully accept this type of backbone modifications. Furthermore the behaviour of the methyl-triazole linkage is one more step in understanding how polymerases accommodate artificial triazole linkages. In future it will be informative to study the structures of DNA duplexes containing the above backbone modifications in detail.

The most suitable techniques are X-ray crystallography, and NMR. Crystallisation of DNA complexed to polymerase enzymes should also be feasible and would provide further insights into the mechanisms of the replication machinery.[184] This will enable the range of artificial backbones available for gene synthesis applications to be greatly extended.

Chapter 5
Conclusion

A wide variety of oligonucleotides containing artificial linkages were successfully synthesised (Figure 5-1). A solid-phase strategy to access novel pyrene, anthraquinone and guanidine modified artificial DNA backbones from a common amine internucleosidic linkage was developed for the first time. All these newly introduced functionalities strongly increased the thermal stability of DNA duplexes containing the parent amine-modified backbone. Importantly, hybridisation studies of the modified analogues have shown that aromatic intercalating moieties can be more efficient at stabilising dsDNA containing amine modified backbone than positively charged modifications. These findings give some useful insight into how the thermodynamic stability of a given modified backbone can be increased. This is important information in the quest to broaden the range of backbone modifications that can be used in antisense oligonucleotides. The various amine substitutions did not fully compensate for the loss of stability induced by replacing the phosphodiester group with an amine modification, and the modified oligonucleotides generally displayed a decreased T_m compared to the unmodified control duplexes. Although other functionalities should be investigated to further improve duplex stability, the pyrene, guanidine and anthraquinone modified analogues might still display useful antisense properties. Indeed these non-canonical DNA backbones should be fully resistant to nucleases, and the neutral or positively charged linkages might improve the cellular uptake of oligonucleotides. Evaluating these analogues in antisense assays will be a stepping-stone to the design of therapeutic oligonucleotides with improved properties.

A synthetic route to prepare oligonucleotides with short-triazole linkage **3.III** *via* phosphoramidate chemistry was successfully developed and the synthesis of oligonucleotides with modified methyl-triazole internucleoside bridge **3.IV** was presented for the first time (Figure 5-1). These 1,2,3-triazole modified backbones (**3.III** and **3.IV**) were designed to be structural variants of our previously published triazole backbone **3.II** to further understand its unprecedented acceptance by the cell replication machinery (E.Coli and MCF-7 human cells).[153, 154] Circular dichroism spectroscopy showed that the modified 1,2,3-triazole linkages (**3.II** to **3.IV**) did not alter the overall three-dimensional structure of the modified DNA duplexes, which all adopted a standard B-DNA shape. Ultraviolet duplex melting studies revealed

that the newly synthesised short-triazole and N-methylated triazole backbones induced a similar decrease in thermal stability as our previously published triazole backbone **3.II**. The latter appeared to be the least duplex destabilising, closely followed by the methyl-triazole linkage **3.IV** and finally by the short-triazole linkage **3.III**.

A set of commercially sourced DNA polymerases (GoTaq®, Large Klenow fragment and KOD XL) was used to study the *in vitro* replication of DNA templates containing short-triazole and methyl-triazole modified backbones. The DNA polymerases replication through 1,2,3-triazole backbones **3.III** and **3.IV** was compared to our previously published triazole backbone **3.II,** using real-time PCR, PAGE gel analysis and Illumina® DNA sequencing. The results confirmed that triazole linkage **3.II** is successfully read through with a good degree of fidelity by all the DNA polymerases tested. This backbone is a good candidate for *de novo* chemical gene synthesis and optimisation is currently undertaken in this direction. Methylation of the triazole backbone **3.II** to form the methyl-triazole backbone (**3.IV**) prevented correct replication by all DNA polymerases. This further confirmed our previous hypothesis that hydrogen bonding to the N3 nitrogen atom of the 1,2,3-triazole ring (**3.II**) is involved in the recognition of the template by DNA polymerases. The short-triazole linkage **3.III**, which is the most thermally destabilising among the studied modified triazole backbones, was read through by DNA polymerases with a consistent skipping of one nucleobase surrounding the modified short-triazole linkage. This could be attributed to the fact that during replication, the thermodynamic stability of the template-primer duplex might be improved by pushing out one of the bases (thymidines) around the short-triazole linkage out of the DNA duplex. This base may then no longer serve as a template and would be missed.

It was also interesting to investigate the biocompatibility of another modified DNA backbone, namely the amide backbone **2.III,** which have an entirely different structure compared to the 1,2,3-triazole linkages. This backbone modification only slightly destabilise dsDNA (-1.5 °C per modification)[188] and as importantly, can form hydrogen bonds *via* the carbonyl H-bond acceptor and the NH

H-bond donor of the amide function. The amide modification **2.III** was pleasingly well tolerated by DNA polymerases and could be very useful in chemical gene synthesis assuming the successful development of a templated oligonucleotide "amide-ligation", which is currently being investigated. The efficient polymerase read-through of the amide linkage also further demonstrated that increased DNA duplex stability and hydrogen bonding are important parameters for correct recognition by DNA polymerases.

These results are further steps towards understanding the parameters that influence the correct replication through modified internucleosidic linkages that can be used as phosphodiester mimic in the complete chemical assembly of artificial genes or genomes. It will be important in the future to supplement these findings with modelling and crystal structures studies of these modified templates with both family A and B DNA polymerases.

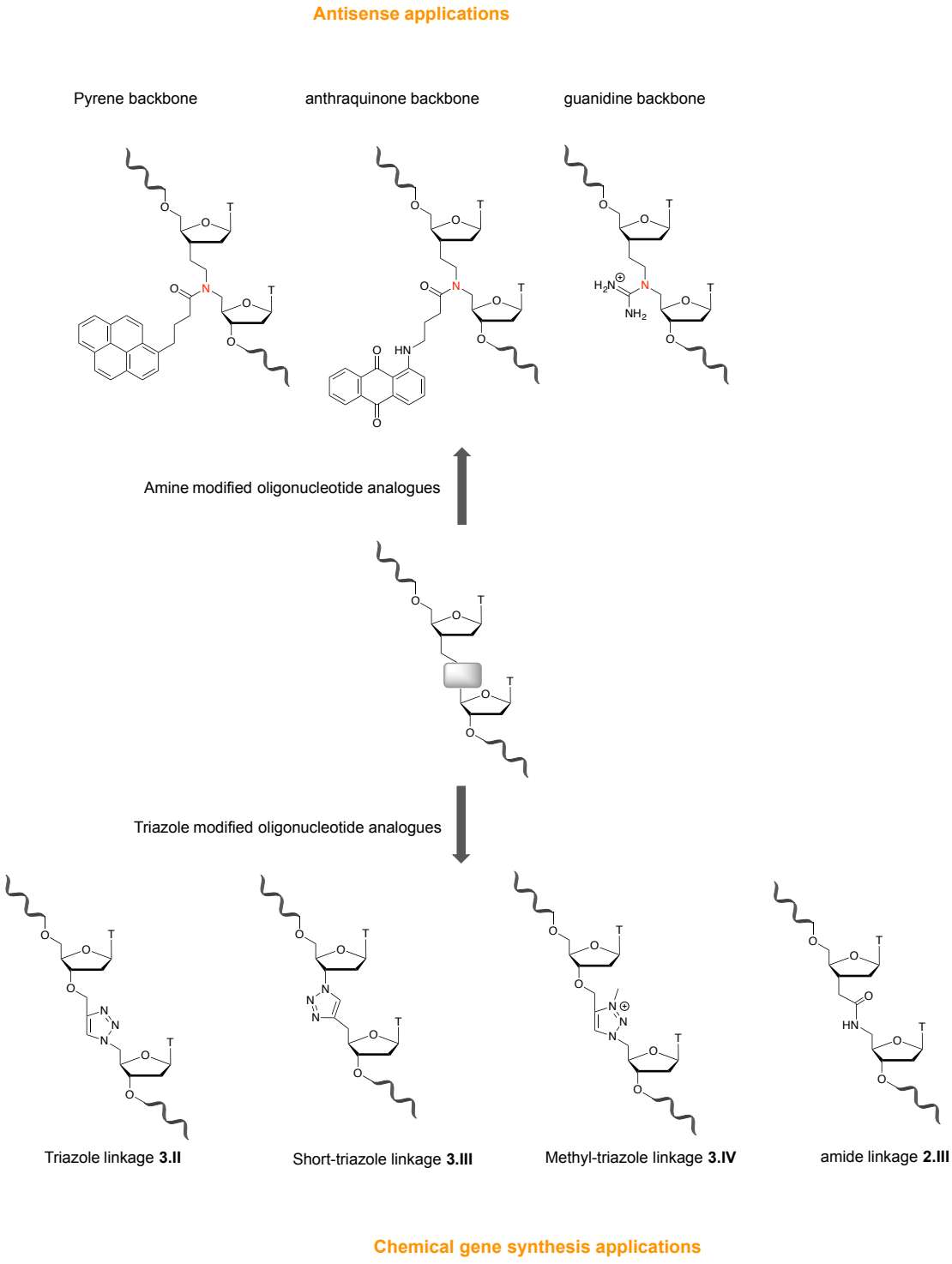


Figure 5-1: Summary of modified oligonucleotide backbones synthesised with some of their potential future applications indicated

Chapter 6

Experimental

6.1 Synthesis of chemical compounds

6.1.1 General conditions

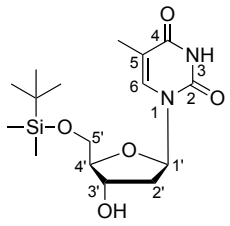
All chemical reagents were purchased from Sigma-Aldrich, Alfa Aesar, Acros Organics and Fisher Scientific and used without further purification. THF (from Na and benzophenone), DCM, TEA, DIPEA and pyridine (from CaH_2) were distilled prior to use, or dried over 3 Å activated molecular sieves for 24 h. When mentioned, deoxygenated solvents were prepared by bubbling argon through the solvent for 15 min. All air/moisture sensitive reactions were carried out under inert atmosphere (Ar) in flame-dried glassware. 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 visualised by UV irradiation at 254/365 nm and by staining in *p*-anisaldehyde solution or $KMnO_4$ (10% aq.). Column chromatography was carried out using Merck Geduran 60 Å (particle size 40-63 micron) silica.

^{1}H, ^{13}C, and ^{31}P NMR spectra were recorded using Bruker AV300 (300, 75 and 121 MHz respectively), AVIII 400 (400 and 101 MHz respectively) and AVII 500 (500, 126 and 202 MHz respectively) spectrometers. Chemical shifts (δ) are given in ppm and were internally referenced to the appropriate residual solvent signal, all coupling constant (*J)* are quoted in Hertz (Hz). Assignment of compounds was aided by COSY, HSQC, DEPT-135 and HMBC experiments.

Low-resolution mass spectra were measured on a Waters LCT premier mass spectrometer for +/- electrospray ionisation (ESI), and on a Waters MALDI Micro MX mass spectrometer in positive linear mode using 2,5-dihydroxybenzoic acid (0.1 M in CH_3CN) as a matrix for MALDI. High-resolution mass spectra were recorded on a Bruker 9.4T FT-ICR-MS mass spectrometer. Samples were run in HPLC grade MeOH or MeCN.

6.1.2 Synthesis and properties of amine and N-functionalised amine modified oligonucleotide analogues

5'-*O*-*tert*-butyldimethylsilylthymidine 2.8[189]



$C_{16}H_{28}N_2O_5Si$
(356.49)

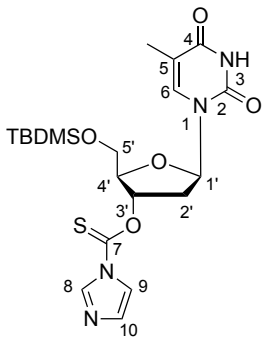
A solution of thymidine (30.0 g, 124 mmol) and imidazole (18.6 g, 270 mmol) in DMF (300 mL) was cooled to 0 ˚C and *tert*-butyldimethylsilyl chloride (21.0 g, 140 mmol) was added. The reaction was stirred overnight at 0 ˚C, the solvent was removed under vacuum and the resulting mixture was slowly poured into cold water. The white solid obtained was recrystallised from chloroform/toluene (10:1) to give **2.8** as a white crystalline solid (30.7 g, 86.1 mmol, 69%).

^{1}H NMR (CD_3OD, 300 MHz): δ 7.63 (d, *J* = 1.1 Hz, 1 H, H-6), 6.27 (dd, *J* = 8.1, 6.0 Hz, 1 H, H-1'), 4.37 (app. dt, *J* = 5.9, 2.6 Hz, 1 H, H-3'), 3.99 - 3.96 (m, 1 H, H-4'), 3.91 (dd, *J* = 11.3, 2.6 Hz, 1 H, H-5'), 3.84 (dd, *J* = 11.3, 3.4 Hz, 1 H, H-5'), 2.28 (ddd, *J* = 13.7, 6.0, 2.6 Hz, 1 H, H-2'), 2.13 (ddd, *J* = 13.7, 8.1, 5.9 Hz, 1 H, H-2'), 1.88 (d, *J* = 1.1 Hz, 3 H, CH_3-*thymine*), 0.94 (s, 9 H, CH_3-*tert-butyl*), 0.14 (s, 6 H, ($Si(CH_3)_2$)

^{13}C NMR (CD_3OD, 75 MHz): δ 166.4 (C-4), 152.3 (C-2), 137.5 (C-6), 111.5 (C-5), 89.0 (C-4'), 86.5 (C-1'), 72.8 (C-3'), 64.8 (C-5'), 41.7 (C-2'), 26.6 (CH_3 - *tert-butyl*), 19.4 (C_q-*tert-butyl*), 12.9 (CH_3-*thymine*), -5.1 ($Si(CH_3)_2$)

LRMS (ESI-TOF) m/z (%): $[M+Na]^+$ 379.2 (100)

5′-*O*-*tert*-butyldimethylsilyl-3′-*O*-(imidazolyl)thiocarbonyl)-thymidine 2.9[189]



$C_{20}H_{30}N_4O_5SSi$
466.63

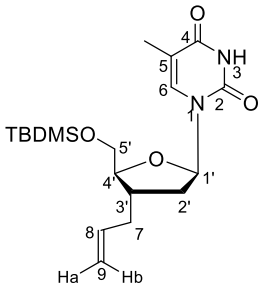
To a solution of 5′-*O*-*tert*-butyldimethylsilylthymidine **2.8** (2.50 g, 7.01 mmol) in toluene (20 mL) was added 1′,1-thiocarbonyldiimidazole (1.75 g, 9.82 mmol). After 3 h at 80 ˚C, the solvent was removed under vacuum. The residue was purified by column chromatography on silica gel (40:60 to 90:10 EtOAc/PETROLEUM ETHER) to afford compound **2.9** as a white crystalline solid (3.20 g, 6.86 mmol, 97%).

^{1}H NMR ($CDCl_3$, 300 MHz): δ 8.69 (br s, 1 H, NH-3), 8.40 (app. t, *J* = 1.0 Hz, 1 H, H-8), 7.65 (dd, *J* = 1.7, 1.0 Hz, 1 H, H-9), 7.59 (d, *J* = 1.2 Hz, 1 H, H-6), 7.09 (dd, *J* = 1.7, 1.0 Hz, 1 H, H-10), 6.46 (dd, *J* = 9.5, 5.3 Hz, 1 H, H-1′), 5.96 (app. d, *J* = 6.2 Hz, 1 H, H-3′), 4.42 - 4.39 (m, 1 H, H-4′), 4.10 - 4.06 (m, 1 H, H-5′), 4.00 (dd, *J* = 11.5, 1.9 Hz, 1 H, H-5′), 2.72 (app. dd, *J* = 14.7, 5.3 Hz, 1 H, H-2′), 2.39 - 2.27 (m, 1 H, H-2′), 1.95 (d, *J* = 1.2 Hz, 3 H, CH_3-*thymine*), 0.97 (s, 9 H, CH_3-*tert-butyl*), 0.18 (s, 6 H, $Si(CH_3)_2$)

^{13}C NMR ($CDCl_3$, 75 MHz): δ 183.3 (C-7), 164.1 (C-4), 150.8 (C-2), 137.4 (C-8), 134.8 (C-6), 131.3 (C-9), 118.1 (C-10), 111.7 (C-5), 85.1 (C-4′), 85.0 (C-1′), 84.5 (C-3′), 64.0 (C-5′), 38.3 (C-2′), 26.1 (CH_3- *tert-butyl*), 18.5 (C_q-*tert-butyl*), 12.7 (CH_3-*thymine*), -5.1 ($Si(CH_3)_2$)

LRMS (ESI-TOF) m/z (%): $[M+Na]^+$ 489.2 (100)

3'-allyl-5'-*O*-*tert*-butyldimethylsilyl-3'-deoxythymidine 2.10[190]



$C_{19}H_{32}N_2O_4Si$
(380.55)

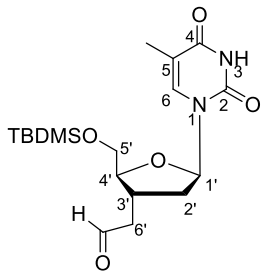
A solution of 5'-*O*-*tert*-butyldimethylsilyl-3'-*O*-(imidazolyl)thiocarbonyl)-thymidine **2.9** (500 mg, 1.07 mmol) and allyltributylstannane (1.9 mL, 6.4 mmol) in anhydrous benzene (4 mL) was degassed with argon. The solution was refluxed and a solution of AIBN (60.0 mg, 0.428 mmol) in benzene (1 mL) was added dropwise. After 3 h the solvent was removed under vacuum. The crude was purified by column chromatography on silica gel with 10% K_2CO_3 (10:90 to 80:20 EtOAc/hexane) to afford compound **2.10** as a white solid (189 mg, 0.497 mmol, 46%).

^{1}H NMR ($CDCl_3$, 500 MHz): δ 9.08 (br s, 1 H, NH-3), 7.59 (s, 1 H, H-6), 6.09 (app. t, *J* = 6.5 Hz, 1 H, H-1'), 5.74 (ddt, *J* = 17.0, 10.1, 6.9 Hz, 1 H, H-8), 5.09 (dd, *J* = 17.0, 0.9 Hz, 1 H, H-9b), 5.06 (dd, *J* = 10.1, 0.9 Hz, 1 H, H-9a), 4.00 (dd, *J* = 11.7, 2.2 Hz, 1 H, H-5'), 3.80 - 3.67 (m, 2 H, H-4' and H-5'), 2.36 (app. sxt, *J* = 6.5 Hz, 1 H, H-3'), 2.26 (app. dt, *J* = 13.6, 6.5 Hz, 1 H, H-2'), 2.20 - 2.07 (m, 3 H, H-2' and H-7), 1.92 (s, 3 H, CH_3-*thymine*), 0.93 (s, 9 H, CH_3-*tert-butyl*), 0.11 (s, 6 H, $Si(CH_3)_2$)

^{13}C NMR ($CDCl_3$, 126 MHz): δ 164.2 (C-4), 150.6 (C-2), 135.9 (C-6), 135.6 (C-8), 117.4 (C-9), 110.4 (C-5), 86.0 (C-4'), 85.1 (C-1'), 63.4 (C-5'), 39.1 (C-7), 37.1 (C-3'), 36.7 (C-2'), 26.2 (CH_3-*tert-butyl*), 18.7 (C_q-*tert-butyl*), 12.9 (CH_3-*thymine*), -5.1 ($Si(CH_3)_2$)

LRMS (ESI-TOF) m/z (%): $[M+Na]^+$ 403.5 (100)

3'-oxoethyl-5'-*O*-*tert*-butyldimethylsilyl-3'-deoxythymidine 2.5[191]



$C_{18}H_{30}N_2O_5Si$
(382.53)

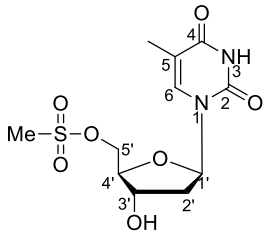
To a mixture of alkene nucleoside **2.10** (500 mg, 1.31 mmol) and *N*-methylmorpholine-*N*-oxide (190 mg, 1.62 mmol) in dioxane (8.3 mL) was added osmium tetroxide (4% wt. in H_2O, 0.4 mL, 0.070 mmol). After 2 h at rt, sodium periodate (450 mg, 2.10 mmol) was added in small portions. The reaction mixture was stirred for 2 h, diluted with EtOAc and filtered through Celite. The filtrate was concentrated under vacuum and purified by column chromatography on silica gel (40:60 to 60:40 EtOAc/PETROLEUM ETHER). Compound **2.5** was obtained as a white solid (370 mg, 0.967 mmol, 74%).

^{1}H NMR ($CDCl_3$, 400 MHz): δ 9.68 (s, 1 H, COH), 9.47 (s, 1 H, NH-3), 7.43 (d, *J* = 1.3 Hz, 1 H, H-6), 6.02 (dd, *J* = 6.8, 5.1 Hz, 1 H, H-1'), 3.85 (dd, *J* = 11.3, 2.7 Hz, 1 H, H-5'), 3.77 - 3.57 (m, 2 H, H-5' and H-4'), 2.73 - 2.56 (m, 2 H, H-3' and H-6'), 2.48 (dd, *J* = 19.6, 9.4 Hz, 1 H, H-6'), 2.20 (ddd, *J* = 13.4, 7.8, 5.1 Hz, 1 H, H-2'), 1.97 (dt, *J* = 13.4, 6.8 Hz, 1 H, H-2'), 1.81 (d, *J* = 1.3 Hz, 3 H, CH_3-*thymine*), 0.80 (s, 9 H, CH_3-*tert-butyl*), 0.00 (s, 6 H, $Si(CH_3)_2$)

^{13}C NMR ($CDCl_3$, 101 MHz): δ 199.9 (COH), 164.2 (C-4), 150.6 (C-2), 135.5 (C-6), 110.6 (C-5), 85.4 (C-4'), 84.7 (C-1'), 63.3 (C-5'), 46.7 (C-6'), 38.6 (C-2'), 32.3 (C-3'), 26.0 (CH_3-*tert-butyl*), 18.5 (C_q-*tert-butyl*), 12.6 (CH_3-*thymine*), -5.4 ($Si(CH_3)_2$)

LRMS (ESI-TOF) m/z (%): $[M-H]^-$ 381.0 (13)

5'-*O*-methanesulfonylthymidine 2.11[135]



$C_{11}H_{16}N_2O_7S$
(320.32)

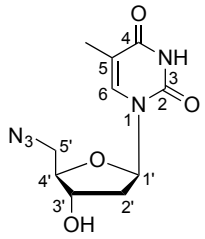
To a solution of thymidine (10.0 g, 41.3 mmol) in anhydrous pyridine (140 mL) at -10 ˚C, methanesulfonyl chloride (3.2 mL, 41.3 mmol) was added dropwise. The reaction mixture was warmed to 0 ˚C and stirred overnight under an argon atmosphere. MeOH was subsequently added to the reaction mixture and the solvent was removed *in vacuo*. The crude was purified by column chromatography on silica gel (2:98 to 4:96 MeOH/DCM) to give **2.11** as a white solid (8.02 g, 25.0 mmol, 61%)

^{1}H NMR ($(CD_3)_2SO$, 300 MHz): δ 11.32 (s, 1 H, NH-3), 7.48 (d, *J* = 0.8 Hz, 1 H, H-6), 6.22 (app. t, *J* = 6.8 Hz, 1 H, H-1'), 5.49 (br s, 1 H, OH-3'), 4.40 (dd, *J* = 10.9, 3.5 Hz, 1 H, H-5'), 4.35 (dd, *J* = 10.9, 5.1 Hz, 1 H, H-5'), 4.26 (app. dt, *J* = 6.4, 3.5 Hz, 1 H, H-3'), 3.96 (app. dt, *J* = 5.1, 3.5 Hz, 1 H, H-4'), 3.23 (s, 3 H, CH_3-*mesyl*), 2.28 - 2.01 (m, 2 H, H-2'), 1.78 (d, *J* = 0.8 Hz, 3 H, CH_3-*thymine*)

^{13}C NMR ($(CD_3)_2SO$, 75 MHz): δ 163.7 (C-4), 150.4 (C-2), 135.9 (C-6), 110.2 (C-5), 84.1 (C-1'), 80.6 (C-4'), 79.5 (C-3'), 68.5 (C-5'), 37.7 (C-2'), 36.9 (CH_3-*mesyl*), 12.1 (CH_3-*thymine*)

LRMS (ESI-TOF) m/z (%): $[M+TFA-H]^-$ 433.0 (100)

5′-azido-5′-deoxythymidine 2.12[135]



$C_{10}H_{13}N_5O_4$
(267.24)

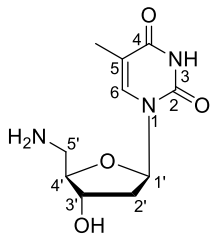
5′-*O*-methanesulfonylthymidine **2.11** (8.00 g, 25.0 mmol) was dissolved in DMF (65 mL) under an argon atmosphere. A solution of lithium azide (20% wt. in H_2O, 35 mL, 142 mmol) was subsequently added and the reaction mixture was stirred for 4 h at 90 ˚C. After removal of the solvent under vacuum, the crude was purified by column chromatography on silica gel (1:99 to 15:85 MeOH/DCM) to afford **2.12** as a white crystalline solid (5.86 g, 21.9 mmol, 88%)

^{1}H NMR ($(CD_3)_2SO$, 300 MHz): δ 11.32 (br s, 1 H, NH-3), 7.49 (d, *J* = 0.9 Hz, 1 H, H-6), 6.20 (app. t, *J* = 6.8 Hz, 1 H, H-1′), 5.40 (br s, 1 H, OH-3′), 4.26 - 4.14 (m, 1 H, H-3′), 3.84 (dt, *J* = 5.2, 3.8 Hz, 1 H, H-4′), 3.56 (d, *J* = 5.2 Hz, 2 H, H-5′), 2.25 (app. dt, *J* = 13.6, 6.8 Hz, 1 H, H-2′), 2.08 (ddd, *J* = 13.6, 6.8, 3.8 Hz, 1 H, H-2′), 1.79 (d, *J* = 0.9 Hz, 3 H, CH_3-*thymine*)

^{13}C NMR ($(CD_3)_2SO$, 75 MHz): δ 163.8 (C-4), 150.6 (C-2), 136.2 (C-6), 110.0 (C-5), 84.7 (C-1′), 84.0 (C-4′), 70.9 (C-3′), 51.8 (C-5′), 38.2 (C-2′) 12.2 (CH_3-*thymine*)

LRMS (ESI-TOF) m/z (%): $[M-H]^-$ 266.0 (84)

5′-amino-5′-deoxythymidine 2.6[135]



$C_{10}H_{15}N_3O_4$
(241.24)

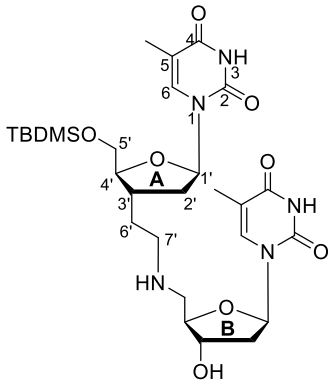
To a solution of 5′-azido-5′-deoxythymidine **2.12** (5.86 g, 21.9 mmol) in MeOH (30 mL), Pd/C (10 % wt., 1.45 g, 0.140 mmol) was added. The resultant mixture was flushed with H_2 and stirred overnight under a hydrogen atmosphere at rt. The palladium was removed by filtration and the reaction mixture was concentrated under vacuum to afford 4.37 g of white flakes containing 78% of the desired product **2.6** (3.40 g, 14.1 mmol, 64%) and secondary amine dimer thymidine impurity **2.13**.

^{1}H NMR ($(CD_3)_2SO$, 300 MHz): δ 7.65 (d, *J* = 1.1 Hz, 1 H, H-6), 6.14 (app. t, *J* = 7.0 Hz, 1 H, H-1′), 5.11 (br s, 1 H, OH-3′), 4.19 (app. dt, *J* = 6.2, 3.4 Hz, 1 H, H-3′), 3.64 (dt, *J* = 5.2, 3.4 Hz, 1 H, H-4′), 3.36 (br. s, 2 H, NH_2), 2.72 (d, *J* = 5.2 Hz, 2 H, H-5′), 2.13 (ddd, *J* = 13.2, 7.0, 6.2 Hz, 1 H, H-2′), 2.03 (ddd, *J* = 13.2, 7.0, 3.4 Hz, 1 H, H-2′), 1.78 (d, *J* = 1.1 Hz, 3 H, CH_3-*thymine*)

^{13}C NMR ($(CD_3)_2SO$, 126 MHz): δ 165.3 (C-4), 151.7 (C-2), 136.4 (C-6), 109.9 (C-5), 88.0 (C-1′), 83.7 (C-4′), 71.1 (C-3′), 44.0 (C-5′), 40.0 (C-2′), 12.7 (CH_3-*thymine*)

LRMS (ESI-TOF) m/z (%): $[M+Cl]^-$ 276.1 (100)

5'-*O*-tert-butyldimethylsylil-5'-thymidinylaminoethylthymidine 2.14[110]



$C_{28}H_{45}N_5O_8Si$
(607.78)

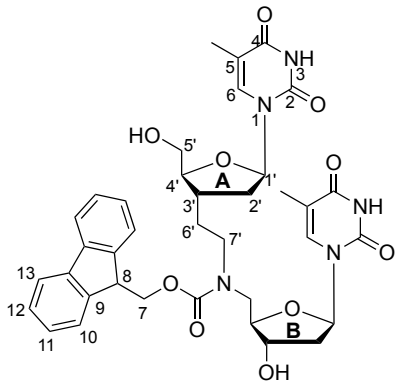
Sodium triacetoxyborohydride (1.50 g, 7.07 mmol) was added to a solution of 3'-oxoethyl-5'-*O*-*tert*-butyldimethylsilyl-3'-deoxythymidine **2.5** (1.37 g, 3.58 mmol) and 5'-amino-5'-deoxythymidine **2.6** (1.03 g, 4.27 mmol) in isopropanol (20 mL). The reaction mixture was subsequently stirred at rt under an argon atmosphere for 2 h. The solvent was removed under vacuum and the residue was purified by column chromatography on silica gel (0:100 to 15:85 MeOH/DCM) to give 2.01 g of a white solid containing 90% of compound **2.14** (1.81 g, 2.98 mmol, 83%).

^{1}H NMR ($(CD_3)_2SO$, 500 MHz): δ 7.59 (d, *J* = 1.0 Hz, 1 H, H-6), 7.48 (d, *J* = 1.0 Hz, 1 H, H-6), 6.15 (app. t, *J* = 6.8 Hz, 1 H, H-1'), 5.99 (dd, *J* = 7.3, 3.6 Hz, 1 H, H-1'), 5.38 - 5.23 (m, 1 H, OH-3'), 4.23 - 4.15 (m, 1 H, H-3'B), 3.85 (dd, *J* = 11.5, 3.4 Hz, 1 H, H-5'A), 3.84 - 3.79 (m, 1 H, H-4'B), 3.70 (dd, *J* = 11.5, 3.4 Hz, 1 H, H-5'A), 3.65 (app. dt, *J* = 7.4, 3.4 Hz, 1 H, H-4'A), 2.93 - 2.77 (m, 2 H, H-5'B), 2.74 - 2.62 (m, 2 H, H-7'), 2.28 - 2.11 (m, 3 H, H-3'A, H-2'), 2.10 – 1.96 (m, 2 H, H-2'), 1.79 (d, *J* = 1.0 Hz, 3 H, CH_3-*thymine*), 1.77 (d, *J* = 1.0 Hz, 3 H, CH_3-*thymine*), 1.73 - 1.63 (m, 1 H, H-6'), 1.53 - 1.39 (m, 1 H, H-6'), 0.90 - 0.85 (m, 9 H, CH_3-*tert-butyl*), 0.07 - 0.06 (m, 6 H, $Si(CH_3)_2$)

^{13}C NMR ($(CD_3)_2SO$, 126 MHz): δ 163.8 163.7 (C-4), 150.5 (C-2), 136.5 (C-6), 135.8 (C-6), 109.6 (C-5), 108.9 (C-5), 85.4 (C-4'), 83.8 (C-1'), 71.3 (C-3'B), 63.1 (C-5'A), 50.4 (C-5'B), 47.5 (C-7'), 38.5 (C-2'), 37.5 (C-2'), 35.3 (C-3'A), 30.7 (C-6'), 25.8 (CH_3-*tert-butyl*), 18.1 (C_q-*tert-butyl*), 12.3 (CH_3-*thymine*), 12.2 (CH_3-*thymine*), -5.4 ($Si(CH_3)_2$)

LRMS (ESI-TOF) m/z (%): $[M+H]^+$ 608.3 (100)

5'-thymidinyl-*N*-FMOC-aminoethylthymidine 2.16



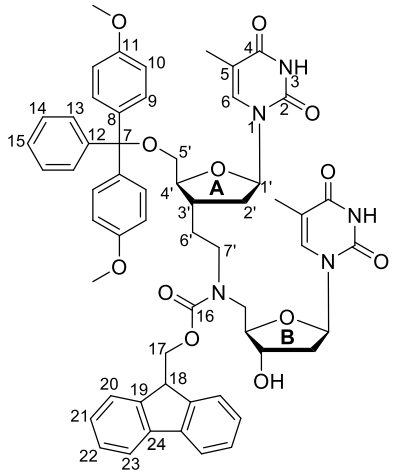
$C_{37}H_{41}N_5O_{10}$
(715.76)

A solution of amine thymidine dimer **2.14** (0.453 g, 0.745 mmol) in THF (10 mL) was treated with an excess of $Et_3N{\cdot}3HF$ (0.990 g, 1 mL, 6.1 mmol) for 12 h at rt. The reaction mixture was concentrated under vacuum and co-evaporated with anhydrous pyridine. After dissolution of the foamy residue in anhydrous pyridine (10 mL), a solution of FMOC-Cl (0.234 g, 0.9 mmol) in DCM (2 mL) was added dropwise. The reaction was left to stir for 2 h at rt with molecular sieves (3 Å). The solvent was removed under vacuum and the residue was co-evaporated with DCM. The residue was then dissolved in MeOH/DCM mixture (15:85) and filtered through a pad of silica to afford 0.592 g of a white solid which was used in the next step without further purification. NMR indicated a mixture of the FMOC rotamers **1** and **2** in a 6:4 ratio.

^{1}H NMR ($(CD_3)_2SO$, 400 MHz): δ 11.41 - 11.18 (m, 2 H, NH-3), 7.96 - 7.78 (m, 3 H, ArH, H-6), 7.70 - 7.52 (m, 3 H, ArH, H-6), 7.47 - 7.28 (m, 4 H, ArH), 6.23 - 6.03 (m, 1 H, H-1'), 6.03 - 5.87 (m, 1 H, H-1'), 5.34 (d, *J* = 4.0 Hz, 0.6 H, OH-3' **1**), 5.21 (m, 0.4 H, OH-3' **2**), 5.18 - 5.05 (m, 1 H, OH-5' **1**), 4.69 - 4.51 (m, 1.2 H, H-7 **1**), 4.47 - 4.22 (m, 1.8 H, H-7 **2**, H-8), 4.10 - 3.90 (m, 1 H, H-3'B), 3.90 - 3.77 (m, 1 H, H-4'), 3.76 - 3.48 (m, 2 H, H-5'), 3.29 - 3.16 (m, 3 H, H-4', H-5'), 2.95 - 2.76 (m, 2 H, H-7'), 2.21 - 1.90 (m, 3 H, H-3'A, H-2'), 1.80 - 1.66 (m, 6 H, CH_3-*thymine*), 1.60 - 1.28 (m, 2 H, H-2'), 1.10 - 0.78 (m, 2 H, H-6')

LRMS (ESI-TOF) m/z (%): [$FMOCNH_2$] 239.2 (100), [M-thymine]$^+$ 590.2 (1), [M+Na]$^+$ 738.2 (0.5)

5′-*O*-4,4′-dimethoxytrityl-5′-thymidinyl-*N*-FMOC-aminoethylthymidine 2.4



$C_{58}H_{59}N_5O_{12}$
(1018.13)

To a crude mixture of 5′-thymidinyl-*N*-FMOC-aminoethylthymidine **2.16** (0.400 g) was added a solution of 4,4′-dimethoxytrityl chloride (0.454 g, 1.34 mmol) in anhydrous pyridine (5 mL). After 2 h MeOH (5 mL) was added and the reaction mixture was left to stir for 30 min. The solvent was removed under vacuum and the residue was taken up in DCM and washed with a saturated aqueous solution of $NaHCO_3$. The organic phase was dried over Na_2SO_4 and the solvent was removed under vacuum before purification by column chromatography on silica gel (0:100 to 7:93 MeOH/DCM with 0.5% pyridine). Compound **2.4** was obtained as a beige foam (0.426 g, 0.417 mmol, 70% over 3 steps). NMR indicated a mixture of the FMOC rotamers **1** and **2** in a 7:3 ratio.

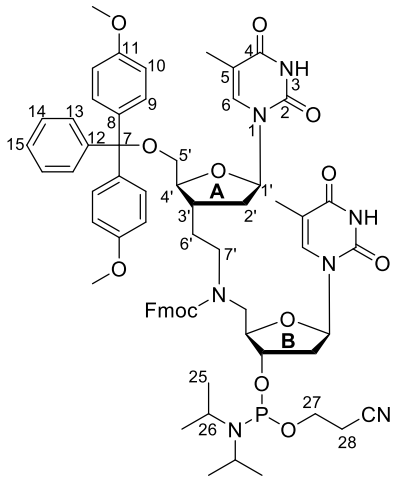
^{1}H NMR ($(CD_3)_2SO$, 500 MHz): δ 11.37, 11.33, 11.29 (3s, 2 H, NH-3), 7.88 - 7.83 (m, 0.6 H, H-23 **2**), 7.81 (t, *J* = 6.6 Hz, 1.4 H, H-23 **1**), 7.64 - 7.56 (m, 1.4 H, H-20 **1**), 7.55 - 7.51 (m, 0.6 H, H-20 **2**), 7.48 (s, 1 H, H-6), 7.46 (s, 1 H, H-6), 7.41 - 7.16 (m, 13 H, H-9, H-13, H-14, H-15, H-21, H-22), 6.89 - 6.79 (m, 4 H, H-10), 6.11 (dd, *J* = 8.2, 6.0 Hz, 1 H, H-1′), 6.01 - 5.96 (m, 0.3 H, H-1′ **2**) 5.95 - 5.91 (m, 0.7 H, H-1′ **1**), 5.37 - 5.32 (m, 0.3 H, OH-3′ **2**), 5.31 (d, *J* = 3.8 Hz, 0.7 H, OH-3′ **1**), 4.55 (dd, *J* = 10.7, 4.7 Hz, 0.7 H, H-17 **1**), 4.52 (dd, *J* = 10.7, 5.0 Hz, 0.7 H, H-17 **1**), 4.37 - 4.26 (m, 0.6 H, H-17 **2**), 4.25 - 4.16 (m, 1 H, H-18), 4.05 - 3.97 (m, 1 H, H-3′B), 3.84 -3.68 (m, 7 H, H-4′, OCH_3), 3.45 (m, 1 H, H-4′*partially overlapped with H_2O peak*), 3.27 - 3.09 (m, 3.3 H, H-5′ *partially overlapped with H_2O peak*), 2.96 - 2.90 (m, 0.7 H, H-5′ **1**), 2.76 (m, 2 H, H-7′), 2.32 - 2.19 (m, 0.6 H,

H-2' **2**), 2.11 – 1.91 (m, 3.4 H, H-3'A, H-2'), 1.69 (s, 3 H, CH_3-*thymine*), 1.63 - 1.52 (m, 1 H, H-2'), 1.49 - 1.42 (s, 3 H, CH_3-*thymine*), 1.21 - 1.11 (m, 1 H, H-6'), 1.01 - 0.89 (m, 0.7 H, H-6' **2**), 0.89 - 0.80 (m, 0.3 H, H-6' **1**)

^{13}C NMR ($(CD_3)_2SO$, 126 MH) δ 173.6 (C-16), 164.2 164.1 (C-4 **1, 2**), 158.6 155.9 (C-11 **1, 2**), 150.8 150.1 (C-2 **1, 2**), 145.2 144.5 144.3 (C-19 **1, 2**), 141.3 139.9 (C-24 **1, 2**) , 137.9 136.6 136.0 135.8 (C-6 **1, 2**), 130.1 129.4 128.3 128.1 127.9 127.8 127.5 127.2 125.0 124.4 121.9 120.5 120.4 (C-9, C-13, C-14, C-15, C-21, C-22), 113.6 (C-10), 110.2 110.1 109.5 (C-5 **1, 2**), 86.1 (C-7), 84.5 84.4 84.1 83.1 (C-1', C-4'), 71.9 (C-3'B), 66.0 (C-17), 63.6 (C-5'), 55.5 55.4 (OCH_3) , 49.7 (C-5'), 47.4 (C-18), 45.8 (C-7'), 41.8 38.7 37.9 (C-2'), 36.1 35.7 (C-3'A), 30.5(C-6'), 12.4 12.3 (CH_3-*thymine* **1, 2**)

LRMS [Maldi-TOF] m/z (%): $[M+Na]^+$ Calcd for $C_{58}H_{59}N_5O_{12}Na$ 1040.41 (100), 1041.41 (70); Found 1040.41 (100), 1041.41 (66)

5'-*O*-4,4'-dimethoxytrityl-5'-thymidinyl-*N*-FMOC-aminoethyl-thymidine-3'-(2-cyanoethyl N,N-diisopropyl)-phosphoramidite 2.3



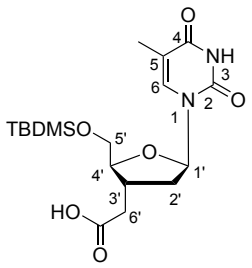
$C_{67}H_{76}N_7O_{13}P$
(1218.35)

To a solution of amine dimer **2.4** (300 mg, 0.294 mmol) in anhydrous DCM (3 mL) was added DIPEA (0.13 mL, 0.725 mmol) followed by the dropwise addition of 2-cyanoethyl-*N*,*N*-diisopropylchlorophosphoramidite (0.15 mL, 0.672 mmol). The reaction was stirred at rt for 3 h under an argon atmosphere with molecular sieves (3 Å). The reaction mixture was diluted with anhydrous DCM (10 mL) and washed with degassed saturated aqueous KCl (10 mL). The organic phase was dried over Na_2SO_4 and concentrated under vacuum before purification by precipitation in anhydrous hexane with a minimum amount of anhydrous DCM. The precipitation process was repeated 3 times to afford phosphoramidite **2.3** as a mixture of diastereomers and as a beige solid (257 mg, 0.211 mmol, 72%).

^{1}H NMR (CD_2Cl_2, 500 MHz): δ 8.96 (br s, 2 H, NH), 7.80 - 7.70 (m, 2 H, ArH), 7.60 - 7.16 (m, 16 H, ArH, H-6), 6.88 - 6.72 (m, 4 H, H-10), 6.23 - 6.10 (m, 1 H, H-1'), 6.09 - 6.01 (m, 1 H, H-1'), 4.76 - 2.77 (m, 23 H, H-17, H-18, H-3'B, H-4', H-26, H-27, OCH3, H-5'), 2.73 - 1.92 (m, 6 H, H-7', H-28), 1.90 - 0.89 (m, 23 H, H-2', H-6', H-25, CH_3-*thymine*)

^{31}P NMR (CD_2Cl_2, 202 MHz): δ 150.09, 149.93, 149.62, 149.28

5′-*O*-*tert*-butyldimethylsilyl-3′-deoxythymidine-3′-acetic acid 2.20[110]

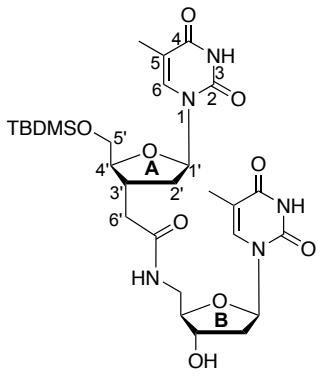


$C_{18}H_{30}N_2O_6Si$
(398.53)

To a solution of 3′-oxoethyl- 5′-*O*-*tert*-butyldimethylsilyl-3′-deoxythymidine **2.5** (2.71 g, 7.08 mmol) in *tert*-butyl alcohol (25 mL) and 2-methyl-2-butene (13 mL) was added a solution of $NaClO_2$ (2.16 g, 23.9 mmol) and NaH_2PO_4 (2.15 g, 17.9 mmol) in water (10 mL). After 3 h the reaction mixture was diluted with water (20 mL), acidify until pH 4 with an aqueous solution of HCl (1 M), and extracted with EtOAc (4 X 100 mL). The organic layer was dried over Na_2SO_4 and concentrated *in vacuo*. The crude was purified by column chromatography on silica gel (1:99 to 5:95 MeOH/DCM) to give the 3′-carboxylic acid thymidine nucleoside **2.20** as a white solid (1.86 g, 4.67 mmol, 66%).

^{1}H NMR ($CDCl_3$, 300 MHz): δ 9.44 (br s, 1 H, COOH), 7.48 (d, *J* = 1.3 Hz, 1 H, H-6), 5.99 (dd, *J* = 6.7, 4.5 Hz, 1 H, H-1′), 3.87 (dd, *J* = 11.4, 2.3 Hz, 1 H, H-5′), 3.73 - 3.61 (m, 2 H, H-5′ and H-4′), 2.65 - 2.43 (m, 2 H, H-3′ and H-6′), 2.32 - 2.16 (m, 2 H, H-6′ and H-2′), 2.22 - 1.95 (m, 1 H, H-2′), 1.81 (d, *J* = 1.3 Hz, 3 H, CH_3-*thymine*), 0.80 (s, 9 H, CH_3-*tert-butyl*), 0.00 (s, 6 H, $Si(CH_3)_2$)

5'-*O*- *tert*-butyldimethylsilyl amide dimer 2.21[110]



$C_{28}H_{43}N_5O_9Si$
(621.76)

5'-*O-tert*-butyldimethylsilyl-3'-deoxythymidine-3'-acetic acid **2.20** (1.76 g, 4.42 mmol) was dissolved in DMF (4 mL), HATU (1.68 g, 4.42 mmol) was added followed by DIPEA (0.85 mL, 4.88 mmol). After 1 h amine **2.6** (1.07 g, 4.43 mmol) was added. The orange solution was stirred for 4 h at rt and the solvent was removed under vacuum. The residue was taken up in water (20 mL) and a saturated aqueous solution of Na_2CO_3 (10 mL) was added. The aqueous layer was extracted with EtOAc (3 X 150 mL). The combined organic layers were washed with a saturated aqueous solution of KCl (50 mL), dried over Na_2SO_4 and concentrated under vacuum. The residue was purified by column chromatography on silica gel (0:100 to 12:88 MeOH/DCM) to give dimer **2.21** as a white solid (1.71 g, 2.75 mmol, 62%).

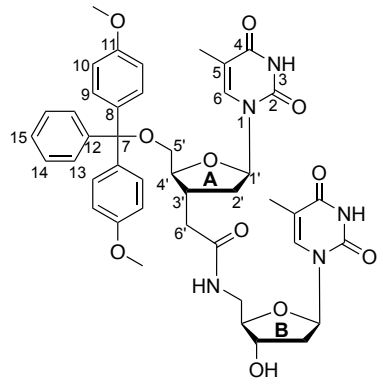
^{1}H NMR (CD_3OD, 400MHz): δ 7.65 (d, *J* = 1.5 Hz, 1 H, H-6), 7.51 (d, *J* = 1.5 Hz, 1 H, H-6), 6.17 (app. t, *J* = 6.6 Hz, 1 H, H-1'B), 6.09 - 6.01 (m, 1 H, H-1'A), 4.24 (app. dt, *J* = 7.1, 3.7 Hz, 1 H, H-3'B), 4.01 - 3.87 (m, 2 H, H-4'B and H-5'A), 3.85 - 3.68 (m, 2 H, H-4'A and H-5'A), 3.53 (dd, *J* = 13.8, 7.6 Hz, 1 H, H-5'B), 3.42 (dd, *J* = 13.8, 3.5 Hz, 1 H, H-5'B), 2.71 (m, 1 H, H-3'A), 2.45 (ddd, *J* = 14.1, 6.6, 3.7 Hz, 1 H, H-2'B), 2.35 - 2.11 (m, 5 H, H-2' and H-6'), 1.89 (d, *J* = 1.5 Hz, 3 H, CH_3-*thymine*), 1.87 (d, *J* = 1.5 Hz, 3 H, CH_3-*thymine*), 0.94 0.91 (s, 9 H, *tert-butyl*), 0.13 0.07(s, 6 H, $Si(CH_3)_2$)

^{13}C NMR (CD_3OD, 101 MHz) δ 174.0 (CO-*amide)*, 166.3 (C-4), 152.3 (C-2), 138.4 137.6 (C-6), 111.8 111.1 (C-5), 87.5 (C-1'B), 87.2 (C-4'A), 86.9 (C-4'B), 86.4 (C-1'A), 73.1 (C-3'B), 64.5 (C-5'A), 42.5 (C-5'B), 40.1 (C-2'B), 39.3 (C-6'), 38.8 (C-2'A), 36.4 (C-

3'A), 26.5 (CH_3-*tert-butyl)*, 19.4 (C_q-*tert-butyl)*, 12.8 (CH_3-*thymine*), 12.5 (CH_3-*thymine*), -5.1 ($Si(CH_3)_2$), -5.2($Si(CH_3)_2$)

LRMS (ESI-TOF) m/z (%): $[M+Na]^+$ 644.2 (89)

5'-*O*-4,4'-dimethoxytrityl thymidine amide dimer 2.22[110]



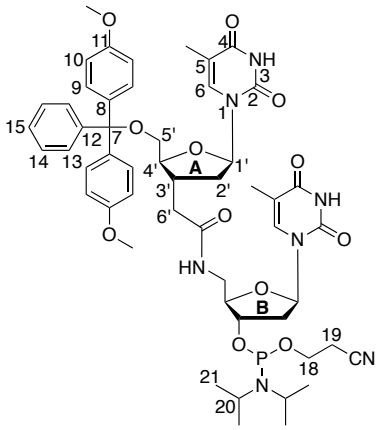
$C_{43}H_{47}N_5O_{11}$
(809.87)

Amide dimer **2.21** (1.71 g, 2.75 mmol) was disolved in a minimum amount of MeOH (6 mL) and a concentrated solution of ammonia (28% in H_2O, 4 mL) was added followed by methylamine (8 mL). The reaction was stirred in a sealed tube for 72 h at 65 °C. The solvent was removed under vacuum and the residue was co-evaporated with ether (3 X 10 mL). The crude product obtained (0.743 g) was dissolved in anhydrous pyridine (4 mL) and molecular sieves (3 Å) were added. A solution of 4,4'-dimethoxytrityl chloride (0.744 g, 2.19 mmol) in pyridine (5 mL) was subsequently added. After 3 h at rt, MeOH (15 mL) was added and the reaction mixture was poured in 60 mL of EtOAc. The organic layer was washed with a saturated aqueous solution of $NaHCO_3$, followed by H_2O and brine. The organic layer was dried over Na_2SO_4 and concentrated under vacuum. The crude was purified by column chromatography on silica gel (0:100 to 8:92 MeOH/DCM with 0.5 % pyridine) to afford dimer **2.22** as a white solid (0.583 g, 0.720 mmol, 26% over 2 steps).

^{1}H NMR (CD_3OD, 400MHz): δ 7.74 (d, *J* = 1.0 Hz, 1 H, H-6), 7.48 (s, 3 H, H-6 and ArH), 7.18 - 7.38 (m, 7 H, ArH), 6.86 - 6.84 (dd, *J* = 8.8, 1.3 Hz, 4 H, H-10), 6.15 (app. t, *J* = 6.8 Hz, 1 H, H-1'B), 6.08 (dd, *J* = 6.6, 3.5 Hz, 1 H, H-1'A), 4.22 (app. dd, *J* = 8.6, 5.0 Hz, 1 H, H-3'B), 3.90 - 3.87 (m, 2 H, H-4'), 3.75 (s, 6 H, OMe), 3.53 - 3.43 (m, 2 H, H-5'), 3.30 - 3.23 (m, 2 H, H-5'), 2.99 - 2.81 (m, 1 H, H-3'A), 2.39 - 2.12 (m, 6 H, H-2' and H-6'), 1.87 (s, 3 H, CH_3-*thymine*), 1.40 (d, *J* = 1.0 Hz, 3 H, CH_3-*thymine*)

^{13}C NMR (CD_3OD, 101MHz): δ 173.8 (CO-*amide*), 166.2 (C-4), 160.3 (ArC_q), 152.2 (C-2), 146.1 (ArC_q), 138.4 (C-6), 137.6 (C-6), 136.8 (ArC_q), 131.4 (ArCH), 129.4 (ArCH), 128.9 (ArCH), 128.0 (ArCH), 114.2 (C-10), 111.6 (C-5), 111.1 (C-5), 87.8 (C-7), 87.1 (C-1'B), 86.54 (C-4'), 86.46 (C-4'), 86.2 (C-1'A), 73.1 (C-3'B), 64.2 (C-5'), 55.7 (OMe), 42.4 (C-5'), 40.0 (C-2'B), 39.6 (C-6'), 38.9 (C-2'A), 36.3 (C-3'A) , 12.4 (CH_3-*thymine*), 12.1 (CH_3-*thymine*)

5'-*O*-4,4'-dimethoxytrityl thymidine amide dimer phosphoramidite 2.23[110]



$C_{52}H_{64}N_7O_{12}$
(1010.09)

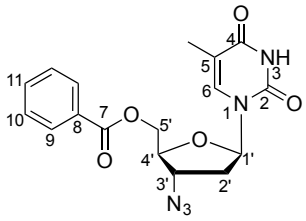
To a solution of amide dimer **2.22** (583 mg, 0.720 mmol) in anhydrous DCM (3 mL) was added DIPEA (0.312 mL, 1.79 mmol) followed by the dropwise addition of 2-cyanoethyl-*N*,*N*-diisopropylchlorophosphoramidite (0.225 mL, 1.01 mmol). The reaction was stirred at rt for 3 h under an argon atmosphere with molecular sieves (3 Å). The reaction mixture was diluted with anhydrous DCM (10 mL) and washed with degassed saturated aqueous KCl (10 mL). The organic phase was dried over Na_2SO_4 and concentrated under vacuum, before purification by column chromatography under argon pressure eluting with degassed EtOAc and 0.5 % pyridine. Phosphoramidite **2.23** was obtained as a 1:1 mixture of diastereomers **1** and **2** and as a white solid (0.240 g, 0.238 mmol, 33%).

^{1}H NMR (CD_3CN, 500 MHz): δ 9.08 9.04 (br s, 2 H, NH-3), 7.49 (s, 1 H, H-6), 7.48 - 7.42 (m, 2 H, ArH), 7.37 - 7.27 (m, 8 H, H-6 and ArH), 6.89 - 6.84 (m, 4 H, H-10), 6.68 (m, 1 H, NH-*amide*), 6.12 - 6.02 (m, 2 H, H-1'), 4.45 - 4.36 (m, 1 H, H-3'B), 4.03 (dt, *J* = 7.0, 3.5 Hz, 0.5 H, H4'B **1**), 3.97 (dt, *J* = 6.6, 3.5 Hz, 0.5 H, H4'B **2**), 3.86 - 3.78 (m, 1 H, H-4'A), 3.76 (s, 6 H, OCH_3), 3.75 - 3.68 (m, 2 H, H-18), 3.62 (spt, *J* = 6.6 Hz, 1 H, H-20 **1**), 3.61 (spt, *J* = 6.6 Hz, 1 H, H-20 **2**), 3.52 (dd, *J* = 14.5, 7.0 Hz, 0.5 H, H-5'B **1**), 3.47 (dd, *J* = 14.5, 7.0 Hz, 0.5 H, H-5'B **2**), 3.34 (dt, *J* = 10.7, 2.2 Hz, 1 H, H-5'A), 3.31 - 3.23 (m, 2 H, H-5'A H-5'B), 2.82 - 2.72 (m, 1 H, H-3'A), 2.66 (t, *J* = 6.0 Hz, 1 H, H-19 **1**), 2.65 (t, *J* = 6.0 Hz, 1 H, H-19 **2**), 2.34 (ddd, *J* = 13.9, 6.3, 3.2 Hz, 0.5 H, H-2'), 2.31 - 2.16 (m, 3.5 H, H-2'), 2.13 - 2.06 (m, 2 H, H-6'), 1.82 (s, 3 H, CH_3 -*thymine*) 1.48 (s, 3 H, CH3-*thymine*), 1.12 - 1.20 (m, 12 H, H-21)

^{31}P NMR (CD_3CN, 202 MHz) δ 148.4, 148.1

6.1.3 Synthesis and study of the hybridisation properties of DNA containing triazole modified backbones

3′-azido-5′-*O*-(benzoyl)-3′-deoxythymidine 3.7[162]



$C_{17}H_{17}N_5O_5$
(371.35)

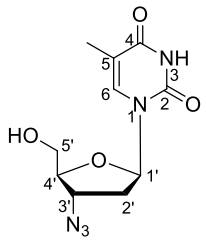
A mixture of 2,3′-anhydro-5′-*O*-benzoylthymidine **3.6** (2.00 g, 6.09 mmol) and sodium azide (1.27 g, 19.5 mmol) in DMF (40 mL), was stirred under an argon atmosphere for 15 h at 130 °C. The mixture was poured into a saturated aqueous solution of NH_4Cl (200 mL) and the resulting yellow precipitate was extracted with EtOAc (3 X 200 mL). The combined organic layers were washed with H_2O and brine, dried over $MgSO_4$ and concentrated *in vacuo*. The residue was purified by column chromatography on silica gel (30:70 to 65:35 EtOAc/hexane). Compound **3.7** was obtained as a white powder (1.75 g, 4.73 mmol, 78%).

^{1}H NMR ($(CD_3)_2CO$), 400 MHz): δ 10.10 (br s, 1 H, NH-3), 8.15 - 8.03 (m, 2 H, H-9), 7.74 - 7.61 (m, 1 H, H-11), 7.56 (dd, *J* = 8.4 Hz, 2 H, H-10), 7.47 (d, *J* = 1.2 Hz, 1 H, H-6), 6.29 (app. t, *J* = 6.6 Hz, 1 H, H-1′), 4.83 - 4.65 (m, 2 H, H-5′ and H-3′), 4.62 (dd, *J* = 12.2, 4.2 Hz, 1 H, H-5′), 4.27 (app. dt, *J* = 5.2, 4.2 Hz, 1 H, H-4′), 2.66 (ddd, *J* = 14.0, 7.6, 6.6 Hz, 1 H, H-2′), 2.56 (ddd, *J* = 14.0, 6.6, 5.3 Hz, 1 H, H-2′), 1.65 (d, *J* = 1.2 Hz, 3 H, CH_3-*thymine*)

^{13}C NMR ($(CD_3)_2CO$), 126 MHz): δ 165.6 (C-7), 163.3 (C-4), 150.3 (C-2), 135.5 (C-6), 133.4 (C-10), 129.8 (C-8), 129.5 (C-9), 128.7 (C-11), 110.4 (C-5), 84.5 (C-1′), 81.5 (C-4′), 63.8 (C-5′), 60.8 (C-3′), 36.67 (C-2′), 11.4 (CH_3-*thymine*)

LRMS: [ESI-TOF] m/z (%): $[M-H]^-$ 370.0 (100)

3′-azido-3′-deoxythymidine (AZT) 3.8[192]



$C_{10}H_{13}N_5O_4$
(267.24)

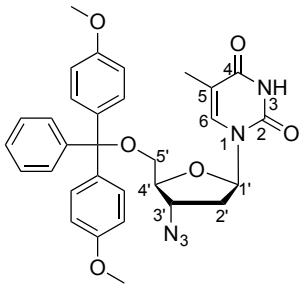
To a solution of azide **3.7** (2.71 g, 7.30 mmol) in dried MeOH (36 mL) was added NaOMe (0.572 g, 10.6 mmol). The mixture was stirred at rt under an argon atmosphere for 21 h and H_2O (45 mL) was added. The reaction mixture was washed with Et_2O (3 X 45 mL) and stirred with DOWEX-H+ resin 50WX8 for 30 min. The resin was removed by filtration and the residual aqueous solution was concentrated *in vacuo* to give compound **3.8** as a white powder (1.34 g, 5.01 mmol, 69%).

^{1}H NMR ($(CD_3)_2SO$, 300 MHz): δ 11.33 (s, 1 H, NH-3), 7.68 (s, 1 H, H-6), 6.09 (app. t, *J* = 6.4 Hz, 1 H, H-1′), 5.23 (s, 1 H, OH-3′), 4.40 (dt, *J* = 7.3, 4.7 Hz, 1 H, H-3′), 3.81 (app. q, *J* = 4.7 Hz, 1 H, H-4′), 3.65 (dd, *J* = 12.0, 4.7 Hz, 1 H, H-5′), 3.59 (dd, *J* = 12.0, 4.7 Hz, 1 H, H-5′), 2.46 - 2.19 (m, 2 H, H-2′), 1.78 (s, 3 H, CH_3-*thymine*)

^{13}C NMR ($(CD_3)_2SO$, 75 MHz): δ 163.7 (C-4), 150.4 (C-2), 136.1 (C-6), 109.5 (C-5), 84.0 (C-4′), 83.4 (C-1′), 60.8 (C-5′), 60.2 (C-3′), 36.2 (C-2′), 12.2 (CH_3-*thymine*)

LRMS: [ESI-TOF] m/z (%): $[M-H]^-$ 266.0 (100)

3′-azido-5′-*O*-(4,4′-dimethoxytrityl)-3′-deoxythymidine 3.3[163]



$C_{31}H_{31}N_5O_6$
(569.62)

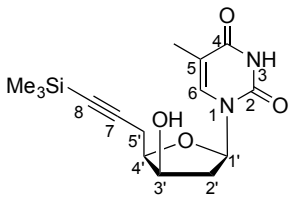
To a solution of 3′-azido-3′-deoxythymidine **3.8** (0.400 g, 1.50 mmol) in dry pyridine (8 mL) was added 4,4′-dimethoxytrityl chloride (1.00 g, 2.95 mmol). The reaction was stirred for 15 h with molecular sieves (3 Å) at rt. MeOH (10 mL) was subsequently added and the solution was poured into EtOAc (30 mL). The mixture was washed with a saturated aqueous solution of $NaHCO_3$, H_2O and brine. The combined organic layers were dried over Na_2SO_4 and concentrated *in vacuo*. The residue was purified by column chromatography on silica gel (5:95 to 25:75 EtOAc/DCM with 0.5 % pyridine) to give compound **3.3** as a white powder (0.744 g, 1.31 mmol, 87%).

^{1}H NMR ($CD(Cl)_3$, 300 MHz): δ 8.59 (br s, 1 H, NH-3), 7.54 (d, *J* = 1.2 Hz, 1 H, H-6), 7.40 - 7.31 (m, 2 H, ArH), 7.27 - 7.20 (m, 7 H, ArH), 6.85 - 6.75 (m, 4 H, ArH), 6.23 (app. t, *J* = 6.4 Hz, 1 H, H-1′), 4.30 (app. dt, *J* = 6.9, 4.9 Hz, 1 H, H-3′), 4.00 (app. dt, *J* = 4.9, 2.8 Hz, 1 H, H-4′), 3.75 (s, 6 H, OCH_3), 3.53 (dd, *J* = 10.9, 2.8 Hz, 1 H, H-5′), 3.29 (dd, *J* = 10.9, 2.8 Hz, 1 H, H-5′), 2.49 - 2.30 (m, 2 H, H-2′), 1.46 (d, *J* = 1.2 Hz, 3 H, CH_3-*thymine*)

^{13}C NMR ($CD(Cl)_3$, 75 MHz): δ 163.5 (C-4), 158.8 (ArC_q), 150.1 (C-2), 144.2 (ArC_q), 135.19 (C-6), 135.17 (ArC_q), 135.12 (ArC_q), 130.0 (ArCH), 128.1 (ArCH), 127.2 (ArCH), 113.3 (ArCH), 111.3 (C-5), 87.1 (CPh_3), 84.4 (C-1′), 83.5 (C-4′), 62.8 (C-5′), 60.6 (C-3′), 55.3 (OCH_3), 38.0 (C-2′), 11.9 (CH_3-*thymine*)

LRMS: [ESI-TOF] m/z (%): $[M-H]^-$ 568.2 (100)

5′-(trimethylsilyl)ethynyl-3′-*epi*-5′-deoxythymidine 3.10[158]



$C_{15}H_{22}N_2O_4Si$
(322.44)

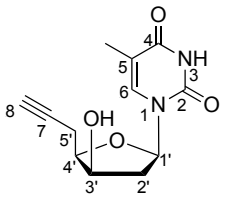
A solution of trimethylsilylacetylene (10.8 mL, 76.4 mmol) in THF (40 mL) was cooled down to -78 °C under an argon atmosphere and treated dropwise with n-BuLi (2.3 M in hexane, 31.4 mL, 72.2 mmol). The solution was stirred for 30 min at -78 °C and for 45 min at 0 °C. $BF_3{\cdot}Et_2O$ (9.0 mL, 72.9 mmol) was added at -78 °C and the solution was stirred for 10 min. Following this, a solution of 1-(3,5-anhydro-2-deoxy-β-D-*threo*-pentofuranosyl*)*thymidine (3.00 g, 13.4 mmol) in THF (200 mL) was added. The reaction was stirred for 20 h at 0 °C and saturated aqueous $NaHCO_3$ (200 mL) was added, the mixture was warmed to rt and extracted with $CHCl_3$ (3 X 200 mL). The organic layers were washed with brine (200 mL), dried over $MgSO_4$ and concentrated *in vacuo*. The crude was purified by column chromatography on silica gel (50:50 to 90:10 EtOAc/$CHCl_3$) to give the desired compound **3.10** as a white powder (3.85 g, 11.9 mmol, 89%).

^{1}H NMR ($CD(Cl)_3$, 300 MHz): δ 8.11 (br s, 1 H, NH-3), 7.59 (d, *J* = 1.2 Hz, 1 H, H-6), 6.14 (dd, *J* = 8.9, 2.8 Hz, 1 H, H-1′), 4.53 - 4.44 (m, 1 H, H-3′), 3.84 (ddd, *J* = 8.8, 6.0, 2.9 Hz, 1 H, H-4′), 2.84 - 2.59 (m, 4 H, H-2′, H-5′, OH), 2.13 (app. dd, *J* = 15.5, 2.8 Hz, 1 H, H-2′), 1.94 (d, *J* = 1.2 Hz, 3 H, CH_3-*thymine*), 0.18 (s, 9 H, CH_3-*trimethylsilyl*)

^{13}C NMR ($CD(Cl)_3$, 75 MHz): δ 163.4 (C-4), 150.3 (C-2), 137.2 (C-6), 110.9 (C-5), 101.5 (C-7), 87.8 (C-8), 85.3 (C-1′), 81.8 (C-4′), 70.7 (C-3′), 40.1 (C-2′), 20.2 (C-5′), 12.6 (CH_3-*thymine*), -0.05 (CH_3-*trimethylsilyl*)

LRMS: [ESI-TOF] m/z (%): $[M-H]^-$ 321.0 (100)

5′-ethynyl-3′-*epi*-5′-deoxythymidine 3.11[158]



$C_{12}H_{14}N_2O_4$
(250.25)

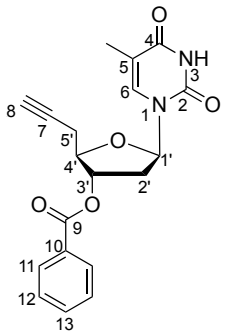
To a solution of 5′-(trimethylsilyl)ethynyl-3′-*epi*-5′-deoxythymidine **3.10** (2.90 g, 8.99 mmol) in THF (25 mL) at 0 °C was added tetra-n-butylammonium fluoride (1 M in THF, 9.9 mL, 9.9 mmol). After 2 h the mixture was filtered through a pad of silica gel and concentrated *in vacuo*. The residue was purified by column chromatography on silica gel (3:97 MeOH/DCM) to give **3.11** as a white powder (2.12 g, 8.47 mmol, 94%).

^{1}H NMR ($CD(Cl)_3$, 300 MHz): δ 9.38 (br s, 1 H, NH-3), 7.57 (d, *J* = 1.2 Hz, 1 H, H-6), 6.08 (dd, *J* = 8.2, 2.3 Hz, 1 H, H-1′), 4.33 (ddd, *J* = 7.5, 5.3, 3.7 Hz, 1 H, H-3′), 4.06 (app. td, *J* = 7.5, 2.9 Hz, 1 H, H-4′), 3.26 (d, *J* = 3.7 Hz, 1 H, OH), 2.85 - 2.67 (m, 2 H, H-5′), 2.63 (ddd, *J* = 15.2, 8.2, 5.3 Hz, 1 H, H-2′), 2.35 (dd, *J* = 15.2, 2.3 Hz, 1 H, H-2′), 2.09 (t, *J* = 2.9 Hz, 1 H, H-8), 1.88 (d, *J* = 1.2 Hz, 3 H, CH_3-*thymine*)

^{13}C NMR ($CD(Cl)_3$, 75MHz): δ 164.0 (C-4), 150.7 (C-2), 137.5 (C-6), 110.2 (C-5), 86.1 (C-1′), 82.7 (C-4′), 79.8 (C-7), 70.4 (C-3′), 70.1 (C-8), 40.6 (C-5′), 18.8 (C-2′), 12.5 (CH_3-*thymine*)

LRMS: [ESI-TOF] m/z (%): $[M-H]^-$ 249.0 (100)

3'-*O*-benzoyl-5'-ethynyl-5'-deoxythymidine 3.12[158]



$C_{19}H_{18}N_2O_5$
(354.36)

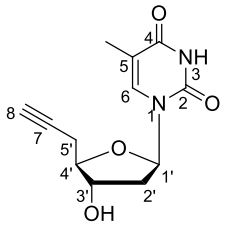
5'-ethynyl-3'-*epi*-5'-deoxythymidine **3.11** (0.500 g, 2.00 mmol) and benzoic acid (0.270 g, 2.21 mmol) was dissolved in dry THF (15 mL). Triphenylphosphine resin (1.6 mmol/g, 1.90 g, 3.04 mmol) and diisopropyl azodicarboxylate (0.43 mL, 2.18 mmol) were added and the reaction mixture was stirred at 0 °C under an argon atmosphere. After 3 h, the resin was removed and the filtrate was washed with a saturated aqueous solution of $NaHCO_3$ (15 mL), brine (15 mL) and dried over $MgSO_4$. The crude was purified by column chromatography on silica gel (5:95 to 30:70 EtOAc/DCM) to afford compound **3.12** as a white powder (0.262 g, 0.739 mmol, 37%)

^{1}H NMR ($CD(Cl)_3$, 300 MHz): δ 8.22 (br s, 1 H, NH-3), 8.09 - 8.03 (m, 2 H, H-11), 7.68 (d, *J* = 1.2 Hz, 1 H, H-6), 7.66 - 7.58 (m, 1 H, H-13), 7.53 - 7.44 (m, 2 H, H-12), 6.43 (dd, *J* = 9.0, 5.6 Hz, 1 H, H-1'), 5.49 (app. dt, *J* = 6.6, 2.2 Hz, 1 H, H-3'), 4.29 (app. td, *J* = 4.3, 2.2 Hz, 1 H, H-4'), 2.90 (ddd, *J* = 17.1, 4.3, 2.6 Hz, 1 H, H-5'), 2.81 (ddd, *J* = 17.1, 4.3, 2.6 Hz, 1 H, H-5'), 2.63 (ddd, *J* = 14.9, 5.6, 2.2 Hz, 1 H, H-2'), 2.43 (ddd, *J* = 14.9, 9.0, 6.6 Hz 1 H, H-2'), 2.17 (t, *J* = 2.6 Hz, 1 H, H-8), 1.97 (d, *J* = 1.2 Hz, 3 H, CH_3-*thymine*),

^{13}C NMR ($CD(Cl)_3$, 75 MHz): δ 166.1 (C-9), 163.3 (C-4), 150.2 (C-2), 135.1 (C-6), 133.7 (C-13), 129.7 (C-11), 129.1 (C-10), 128.6 (C-12), 111.5 (C-5), 84.4 (C-1'), 81.9 (C-4'), 80.2 (C-7), 76.6 (C-3'), 71.6 (C-8), 37.7 (C-5'), 23.3 (C-2'), 12.6 (CH_3-*thymine*)

LRMS: [ESI-TOF] m/z (%): $[M-H]^-$ 353.0 (100)

5'-ethynyl-5'-deoxythymidine 3.4[158]



$C_{12}H_{14}N_2O_4$
(250.25)

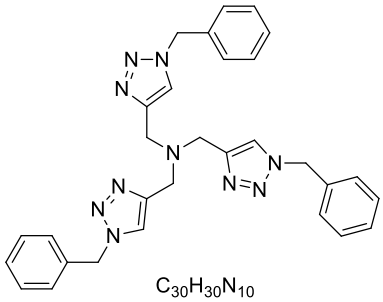
To a solution of 3'-*O*-benzoyl-5'-ethynyl-5'-deoxythymidine **3.12** (0.100 g, 0.282 mmol) in MeOH (10 mL) was added K_2CO_3 (0.058 g, 0.419 mmol) and the reaction mixture was stirred at rt for 15 h. The solvent was subsequently removed *in vacuo* and DCM was added (35 mL). The solution was washed with a saturated aqueous solution of $NaHCO_3$ (3 X 20 mL), dried over $MgSO_4$ and concentrated *in vacuo*. The crude was purified by column chromatography on silica gel (MeOH/DCM 5:95 to 8:92) to give compound **3.4** as a white solid (0.046 g, 0.184 mmol, 65%).

^{1}H NMR (CD_3OD, 300 MHz): δ 7.60 (d, *J* = 1.3 Hz, 1 H, H-6), 6.15 (app. t, *J* = 6.8 Hz, 1 H, H-1'), 4.26 (app. td, *J* = 5.0, 3.3 Hz, 1 H, H-3'), 3.86 (app. td, *J* = 5.0, 3.3 Hz, 1 H, H-4'), 2.59 (ddd, *J* = 17.2, 5.0, 2.6 Hz, 1 H, H-5'), 2.51 (ddd, *J* = 17.2, 5.0, 2.6 Hz, 1 H, H-5'), 2.39 (t, *J* = 2.6 Hz, 1 H, H-8), 2.18 (app. dd, *J* = 6.8, 5.0 Hz, 2 H, H-2'), 1.78 (d, *J* = 1.3 Hz, 3 H, CH_3-*thymine*)

^{13}C NMR (CD_3OD, 75 MHz): δ 166.3 (C-4), 152.3 (C-2), 137.7 (C-6), 111.6 (C-5), 86.1 (C-1'), 85.5 (C-4'), 81.3 (C-7), 74.0 (C-8), 72.4 (C-3'), 40.6 (C-2'), 23.5 (C-5'), 12.4 (CH_3-*thymine*)

LRMS: [ESI-TOF] m/z (%): $[M-H]^-$ 249.0 (100)

Tris-(benzyltriazolylmethyl)amine (TBTA) 3.13[165]



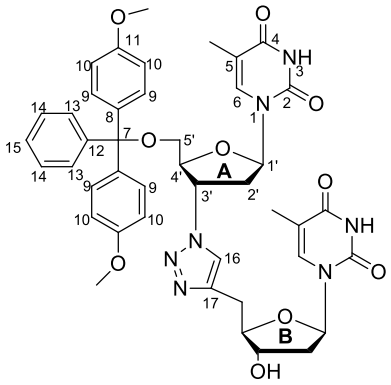
To a solution of tripropargylamine (1.8 mL, 12.7 mmol) in acetonitrile (25 mL) was added benzyl azide (9.4 mL, 75.2 mmol) followed by 2,6-lutidine (1.5 mL, 12.9 mmol) and tetrakis(acetonitrile)copper(I) hexafluorophosphate (0.121 g, 0.326 mmol). The reaction mixture was stirred at rt for 3 days and a white solid precipitated out from the reaction mixture. The solid was filtered and washed with cold acetonitrile to give **3.13** as a white solid (2.44 g, 4.60 mmol, 36%).

^{1}H NMR ($CD(Cl)_3$, 300 MHz): δ 7.66 (s, 3 H), 7.42 – 7.30 (m, 9 H), 7.28 - 7.25 (m, 6 H), 5.51 (s, 6 H), 3.71 (s, 6 H)

^{13}C NMR ($CD(Cl)_3$, 75 MHz,): δ 134.7, 129.0 (CH), 128.6 (CH), 127.9 (CH), 123.7 (CH), 54.0 (CH_2), 47.1 (CH_2)

LRMS: [ESI-TOF] m/z (%): $[M+H]^+$ 531.2 (100)

5'-*O*-(4,4'-dimethoxytrityl)-triazole thymidine dimer 3.2



$C_{43}H_{45}N_7O_{10}$
(819.32)

To a solution of 3'-azido-5'-*O*-(4,4'-dimethoxytrityl)-3'-deoxythymidine **3.3** (0.750 g, 1.32 mmol), 5'-ethynyl-5'-deoxythymidine **3.4** (0.300 g, 1.20 mmol) and TBTA **3.13** (0.320 g, 0.603 mmol) in DMF (5 mL), was added a solution of sodium ascorbate (0.480 g, 2.42 mmol) in H_2O (4 mL). Following this, a solution of $CuSO_4 \cdot 5H_2O$ (0.150 g, 0.601 mmol) in H_2O (4 mL) was added and the reaction was stirred at rt for 2 h. The solvent was subsequently removed *in vacuo* and the residue was dissolved in a 10:90 MeOH/DCM (55 mL) mixture. The solution was washed with a 5% aqueous solution of Na_2EDTA (3 X 55 mL), H_2O and brine. The organic layer was then dried over Na_2SO_4 and concentrated *in vacuo*. The residue was purified by column chromatography on silica gel (4.75:0.5:94.75 to 7:0.5:92.5 MeOH/pyridine/DCM). Compound **3.2** was isolated with 1 eq. of residual pyridine as a beige solid (0.780 g, 0.952 mmol, 79%).

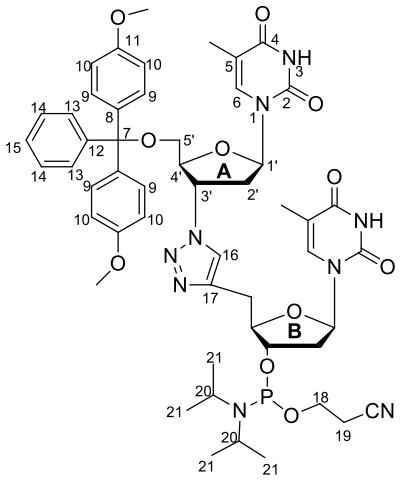
^{1}H NMR ($CD(Cl)_3$, 400 MHz): δ 9.52 (br s, 1 H, NH-3), 9.40 (br s, 1 H, NH-3), 7.61 (s, 1 H, H-6), 7.50 (s, 1 H, H-16), 7.44 - 7.34 (m, 2 H, H-13), 7.33 - 7.21 (m, 7 H, H-15, H-14, H-9), 7.17 (s, 1 H, H-6), 6.93 - 6.78 (m, 4 H, H-10), 6.48 (app. t, *J* = 5.8 Hz, 1 H, H-1'A), 6.17 (app. t, *J* = 6.9 Hz, 1 H, H-1'B), 5.37 (app. dt, *J* = 8.6, 5.8 Hz, 1 H, H-3'A), 4.49 (br s, 1 H, OH-3'), 4.46 - 4.37 (m, 2 H, H-4'A, H-3'B), 4.07 (app. dt, *J* = 7.2, 5.5 Hz, 1 H, H-4'B), 3.80 (s, 6 H, OCH_3), 3.67 (dd, *J* = 10.8, 3.0 Hz, 1 H, H-5'A), 3.38 (dd, *J* = 10.8, 2.8 Hz, 1 H, H-5'A), 3.16 (dd, *J* = 15.3, 5.5 Hz, 1 H, H-5'B), 3.08 (dd, *J* = 15.3, 7.2 Hz, 1 H, H-5'B), 3.00 (dt, *J* = 14.1, 5.8 Hz, 1 H, H-2'A), 2.75 (ddd, *J* = 14.1, 8.6, 5.8 Hz, 1 H, H-2'A), 2.45 (app. dt, *J* = 13.4, 6.9 Hz, 1 H, H-2'B), 2.33 (app. dt, J = 13.4, 6.9 Hz, 1 H, H-2'B), 1.91 (s, 3 H, CH_3-*thymine*), 1.58 (s, 3 H, CH_3-*thymine*)

^{13}C NMR ($CD(Cl)_3$, 101 MHz): δ 163.8 (C-4), 163.7 (C-4), 158.8 (C-11), 150.5 (C-2), 150.4 (C-2), 144.1 (C-12), 143.8 (C-17), 135.3 (C-6A, C-6B), 135.1 135.0 (C-8), 130.0 (C-9), 128.1 128.0 (C-14, C-13), 127.3 (C-15), 121.8 (C-16), 113.4 (C-10), 111.7 (C-5), 111.2 (C-5), 87.2 (C-7), 85.3 (C-1'A), 85.0 (C-1'B), 84.3 (C-4'A), 83.7 (C-4'B), 73.7 (C-3'B), 62.7 (C-5'A), 60.2 (C-3'A), 55.3 (OCH_3), 39.5 (C-5'B), 38.3 (C-2'A), 29.4 (C-2'B), 12.6 (CH_3-*thymine*), 12.0 (CH_3-*thymine*)

LRMS: [ESI-TOF] m/z (%): $[M+Na]^+$ 842.32 (100)

HRMS [ESI-TOF] m/z: $[M+H]^+$ Calcd for $C_{43}H_{46}N_7O_{10}$ 820.3301; Found 820.3299

5′-*O*-(4,4′-dimethoxytrityl)-triazole thymidine dimer phosphoramidite 3.1



$C_{52}H_{62}N_9O_{11}P$
(1020.08)

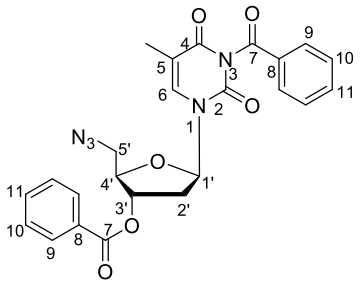
5′-*O*-(4,4′-dimethoxytrityl)-triazole thymidine dimer **3.2** (0.400 g, 0.488 mmol) was co-evaporated with dry pyridine (3 X 10 mL) and dissolved in anhydrous DCM (5 mL) under an argon atmosphere. DIPEA (0.22 mL, 1.26 mmol) was added, followed by the dropwise addition of 2-cyanoethyl-*N*,*N*-diisopropylchlorophosphoramidite (0.14 mL, 0.633 mmol). The reaction was stirred at rt for 2 h and diluted with DCM (10 mL). The solution was washed with degassed saturated aqueous KCl (10 mL), dried over Na_2SO_4 and concentrated under vacuum. The residue was purified by column chromatography under argon pressure eluting with degassed EtOAc. 5′-*O*-(4,4′-dimethoxytrityl)-triazole thymidine dimer phosphoramidite **3.1** was obtained as a 1:1 mixture of diastereomers **1** and **2** and as a white solid (0.298 g, 0.292 mmol, 60%).

^{1}H NMR (CD_3CN, 300MHz): δ 9.22 (s, 2 H, NH-3), 7.68 (d, *J* = 8.5 Hz, 1 H, H-15), 7.55 (t, *J* = 1.5 Hz, 1 H, H-16), 7.46 (s, 1 H, H-6), 7.43 (s, 1 H, H-6), 7.41 - 7.19 (m, 8 H, H-14, H-13, H-9), 6.92 - 6.83 (m, 4 H, H-10), 6.43 (app. t, *J* = 6.8 Hz, 1 H, H-1′A), 6.19 (app. t, *J* = 7.0 Hz, 0.5 H, H-1′B **1** or **2**), 6.15 (app. t, *J* = 6.8 Hz, 0.5 H, H-1′B **1** or **2**) 5.46 - 5.35 (m, 1 H, H-3′A), 4.57 - 4.44 (m, 1 H, H-3′B), 4.36 (app. td, *J* = 6.8, 3.3 Hz, 0.5 H, H-4′ **1** or **2**), 4.34 (app. td, *J* = 6.6, 3.3 Hz, 0.5 H, H-4′ **1** or **2**), 4.29 - 4.22 (m, 0.5 H, H-4′ **1** or **2**), 4.22 - 4.14 (m, 0.5 H, H-4′ **1** or **2**) 3.88 - 3.72 (m, 2 H, H-18), 3.78 (s, 6 H, OCH_3), 3.65 (spt, *J* = 6.6 Hz, 1 H, H-20), 3.56 (spt, *J* = 7.0 Hz, 1 H, H-20), 3.45 (dd, *J* = 11.0, 3.2 Hz, 1 H, H-5′A), 3.32 (dd, *J* = 11.0, 2.9 Hz, 1 H, H-5′A), 3.14 (ddd, *J* = 15.1,

4.5, 1.5 Hz, 1 H, H-5'B), 3.04 (ddd, *J* = 15.1, 7.5, 1.5 Hz, 1 H, H-5'B), 2.86 (ddd, *J* = 13.4, 6.8, 3.1 Hz, 0.5 H, H-2'A **1** or **2**), 2.83 (ddd, *J* = 9.5, 6.8, 3.1 Hz, 0.5 H, H-2'A **1** or **2**), 2.75 (ddd, *J* = 9.5, 6.8, 2.7 Hz, 0.5 H, H-2'A **1** or **2**), 2.72 (ddd, *J* = 13.4, 6.8, 3.20 Hz, 0.5 H, H-2'A **1** or **2**), 2.67 (t, *J* = 5.9 Hz, 2 H, H-19), 2.47 - 2.21 (m, 2 H, H-2'B), 1.83 (d, *J* = 1.2 Hz, 3 H, CH_3-*thymine*), 1.61 (s, 3 H, CH_3-*thymine*), 1.27 - 1.10 (m, 12 H, H-21)

^{31}P NMR (121 MHz, ACETONITRILE-d_3): δ 149.32, 142.24

3-*N*-benzoyl-3′-*O*-benzoyl-5′-azido-5′-deoxythymidine 3.18



To a solution of 5′-azido-5′-deoxythymidine **2.12** (7.51 g, 28.1 mmol) in anhydrous pyridine (200 mL) benzoyl chloride (6.53 mL, 56.2 mmol) was added dropwise under an argon atmosphere. The reaction was stirred for 15 h and the solvent was removed *in vacuo*. The residue obtained was dissolved in DCM (100 mL) and the organic layer was washed with water (100 mL), saturated aqueous solution of $NaHCO_3$ (100 mL) and dried over Na_2SO_4. The crude was purified by column chromatography on silica gel (10:90 to 70:30 EtOAc/hexane) to afford compound **3.18** as a white solid (12.6 g, 26.5 mmol, 94%)

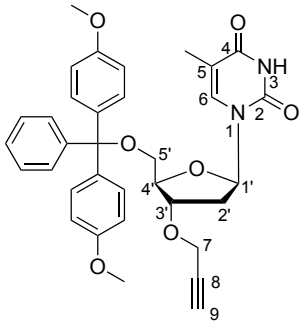
^{1}H NMR ($(CD_3)_2SO$, 500MHz): δ 8.02 – 7.99 (m, 4 H, H-9), 7.89 (d, *J* = 1.0 Hz, 1 H, H-6), 7.78 (tt, *J* = 7.6, 1.1 Hz, 1 H, H-11), 7.67 (tt, *J* = 7.6, 1.1 Hz, 1 H, H-11), 7.63 - 7.57 (m, 2 H, H-10), 7.56 - 7.50 (m, 2 H, H-10), 6.30 (dd, *J* = 7.0, 6.3 Hz, 1 H, H-1′), 5.45 (app. dt, *J* = 7.0, 2.8 Hz, 1 H, H-3′), 4.34 (ddd, *J* = 5.9, 4.3, 2.8 Hz, 1 H, H-4′), 3.78 (dd, *J* = 12.9, 5.9 Hz, 1 H, H-5′), 3.74 (dd, *J* = 12.9, 4.3 Hz, 1 H, H-5′), 2.76 (app. dt, *J* = 14.6, 7.0 Hz, 1 H, H-2′), 2.55 (ddd, *J* = 14.6, 6.3, 2.8 Hz, 1 H, H-2′), 1.92 ppm (d, *J* = 1.0 Hz, 3 H, CH_3-thymine)

^{13}C NMR ($(CD_3)_2SO$, 126MHz): δ 169.4 (C-7), 165.3 (C-7), 162.4 (C-4), 149.1 (C-2), 137.2 (C-6), 135.5 (C-11), 133.7 (C-11), 131.1 (C-8), 130.4 (ArCH), 129.5 (ArCH), 129.2 (C-8), 128.7 (ArCH), 110.0 (C-5) , 84.9 (C-1′), 82.1 (C-4′), 74.8 (C-3′), 51.6 (C-5′), 35.4 (C-2′), 12.0 (CH_3-*thymine*)

LRMS: [ESI-TOF] m/z (%): $[M+Na]^+$ 498.0 (100)

HRMS: [ESI-TOF] m/z: $[M + Na]^+$ Calcd for $C_{24}H_{21}N_5O_6Na$ 498.1384; Found 498.1381

3′-*O*-propargyl-5′-*O*-(4,4′-dimethoxytrityl)-thymidine 3.20[147]



$C_{34}H_{34}N_2O_7$
(582.65)

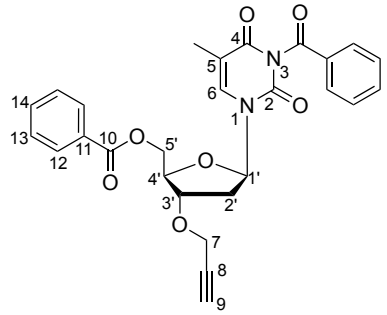
5′-*O*-(4,4′-dimethoxytrityl)-thymidine (11.0 g, 20.2 mmol) was dissolved in dry THF (500 mL) under an argon atmosphere. Sodium hydride (60 % in mineral oil, 2.03 g, 50.7 mmol) was added by small portion at 0 ˚C, the reaction mixture was warmed to rt and stirred for 45 min. A solution of propargyl bromide (80 wt. % in toluene, 3.27 mL, 29.3 mmol) was added dropwise and the reaction mixture was refluxed for 15 h. The solvent was removed under vacuum and the crude mixture was dissolved in DCM (300 mL). The organic layer was washed with water (300 mL), brine (300 mL) and dried over Na_2SO_4. The crude was purified by column chromatography on silica gel (0:100 to 5:95 MeOH/DCM) to afford compound **3.20** as white crystals (10.4 g, 17.8 mmol, 88%).

^{1}H NMR ($(CD_3)_2SO$, 500MHz): δ 11.36 (s, 1 H, NH-3), 7.50 (d, *J* = 1.0 Hz, 1 H, H-6), 7.42 - 7.37 (m, 2 H, ArH), 7.34 – 7.31 (m, 2 H, ArH), 7.29 - 7.20 (m, 5 H, ArH), 6.94 – 6.89 (m, 4 H, ArH), 6.12 (dd, *J* = 7.9, 6.0 Hz, 1 H, H-1′), 4.46 (app. dt, *J* = 5.7, 2.9 Hz, 1 H, H-3′), 4.24 (dd, *J* = 16.4, 2.5 Hz, 1 H, H-6′), 4.20 (dd, *J* = 16.4, 2.5 Hz, 1 H, H-6′), 4.04 (app. dd, *J* = 5.0, 2.9 Hz, 1 H, H-4′), 3.74 (s, 6 H, OMe), 3.52 (t, *J* = 2.5 Hz, 1 H, H-7), 3.26 (dd, *J* = 10.5, 5.0 Hz, 1 H, H-5′), 3.17 (dd, *J* = 10.5, 2.9 Hz, 1 H, H-5′), 2.36 (ddd, *J* = 13.8, 6.0, 2.9 Hz, 1 H, H-2′), 2.30 (ddd, *J* = 13.8, 7.9, 5.7 Hz, 1 H, H-2′), 1.43 (d, *J* = 1.0 Hz, 3 H, CH_3-*thymine*)

^{13}C NMR ($(CD_3)_2SO$, 126MHz): δ 163.6 (C-4), 158.2(Cq), 150.3 (C-2), 144.6 (Cq), 135.6 (Cq), 135.4 (C-6), 135.2 (Cq), 129.7 (ArCH), 128.0 (ArCH), 127.6 (ArCH), 126.8 (ArCH), 113.3 (ArCH), 109.8 (C-5), 86.0 (CPh_3), 83.8 (C-1′), 82.6 (C-4′), 80.0 (C-8) 78.1 (C-3′), 77.5 (C-9), 63.6 (C-5′), 55.9 (C-7), 55.1 (OMe), 36.2 (C-2′), 11.7 (CH_3-thymine)

LRMS: [ESI-TOF] m/z (%): $[M-H]^-$ 581.1 (29)

3-*N*-benzoyl-3′-*O*-propargyl-5′-*O*-benzoyl-thymidine 3.17



$C_{27}H_{24}N_2O_7$
(488.50)

3′-*O*-propargyl-5′-*O*-(4,4′-dimethoxytrityl)-thymidine **3.20** (4.70 g, 8.07 mmol) was dissolved in a solution of 3% TCA in DCM (100 mL). The reaction mixture was stirred for 1 h at room temperature and a solution of NH_3 (28% in H_2O) was added dropwise until the reaction mixture reached pH 7. The reaction mixture was extracted with water (3 X 33 mL) and the combined aqueous layers were concentrated *in vacuo*. The residue was co-evaporated with dry pyridine (3 X 20 mL) and dried under vacuum for 15 h. The white solid obtained was dissolved in anhydrous pyridine (40 mL) and benzoyl chloride (5.90 mL, 50.8 mmol) was added dropwise at 0 ˚C. The reaction mixture was stirred at rt for 2 days and the solvent was removed *in vacuo*. The residue was dissolved in DCM (100 mL) and the organic layer was washed with water (100 mL) then with a saturated aqueous solution of $NaHCO_3$ (100 mL) and dried over Na_2SO_4. The crude was purified by column chromatography on silica gel (10:90 to 70:30 EtOAc/hexane) to afford compound **3.17** as a white solid (3.63 g, 7.43 mmol, 92%)

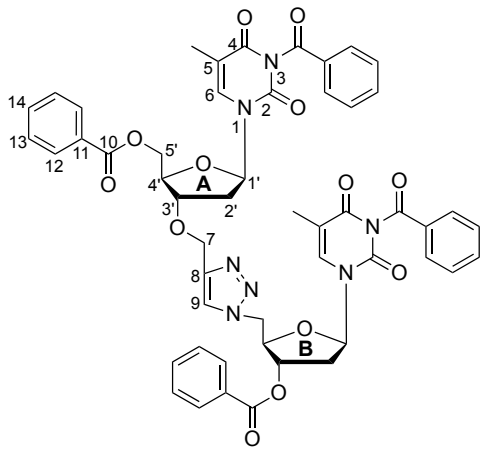
^{1}H NMR (CD_2Cl_2, 500 MHz): δ 7.99 - 7.93 (m, 2 H, H-12), 7.84 - 7.78 (m, 2 H, H-12), 7.60 (tt, *J* = 7.3, 1.4 Hz, 1 H, H-14), 7.54 (tt, *J* = 7.3, 1.2 Hz, 1 H, H-14), 7.46 - 7.37 (m, 4 H, H-13), 7.27 (d, *J* = 1.3 Hz, 1 H, H-6), 6.16 (dd, *J* = 7.9, 6.0 Hz, 1 H, H-1′), 4.56 (dd, *J* = 12.2, 3.0 Hz, 1 H, H-5′), 4.46 (dd, *J* = 12.2, 3.0 Hz, 1 H, H-5′), 4.38 (app. dt, *J* = 6.3, 3.0 Hz, 1 H, H-3′), 4.32 (app. q, *J* = 3.0 Hz, 1 H, H-4′), 4.18 (dd, *J* = 16.0, 2.4 Hz, 1 H, H-7), 4.12 (dd, *J* = 16.0, 2.4 Hz, 1 H, H-7), 2.49 (ddd, *J* = 14.2, 6.0, 3.0 Hz, 1 H, H-2′), 2.43 (t, *J* = 2.4 Hz, 1 H, H-9), 2.12 (ddd, *J* = 14.2, 7.9, 6.3 Hz, 1 H, H-2′), 1.58 (d, *J* = 1.3 Hz, 3 H, CH_3-*thymine*)

^{13}C NMR (CD_2Cl_2, 126 MHz,): δ 169.2 (C-10), 166.0 (C-10), 162.6 (C-4), 149.2 (C-2), 135.2 (C-14), 134.9 (C-6), 133.6 (C-14), 131.6 (C-11), 130.3 (C-12), 129.6 (C-11), 129.4 (C-12), 129.2 (C-13), 128.7 (C-13), 111.2 (C-5), 85.6 (C-1'), 82.5 (C-4'), 79.0 (C-8), 78.4 (C-3'), 75.1 (C-9), 64.3 (C-5'), 57.0 (C-7), 37.6 (C-2'), 12.1 (CH_3-*thymine*)

LRMS: [ESI-TOF] m/z (%): $[M+Na]^+$ 511.0 (57)

HRMS [ESI-TOF] m/z: $[M + Na]^+$ Calcd for $C_{27}H_{24}N_2O_7Na$ 511.1476; Found 511.1474

3-*N*-benzoyl-3',5'-*O*-benzoyl-methoxytriazolyl-thymidine dimer 3.16



$C_{51}H_{45}N_7O_{13}$
(963.96)

3-*N*-benzoyl-3'-*O*-propargyl-5'-*O*-benzoyl-thymidine **3.17** (3.52 g, 7.20 mmol), **3**-*N*-benzoyl-3'-*O*-benzoyl-5'-azido-5'-deoxythymidine **3.18** (4.11 g, 8.64 mmol), and TBTA **3.13** (1.90 g, 3.58 mmol) were dissolved in DMF (50 mL). An aqueous solution of sodium ascorbate (5 mL, 2.80 g, 14.1 mmol) was added followed by the addition of an aqueous solution of $CuSO_4 \cdot 5H_2O$ (2 mL, 0.897 g, 3.59 mmol). The reaction was stirred at rt for 3 h under an argon atmosphere. The solvent was subsequently removed *in vacuo* and the residue was dissolved in a 10:90 MeOH/DCM mixture (50 mL). The solution was washed with water, dried over Na_2SO_4 and concentrated *in vacuo*. The crude was purified by column chromatography on silica gel eluting with 40:60 to 100:0 EtOAc/hexane to afford compound **3.16** as a white solid (6.70 g, 6.95 mmol, 97%).

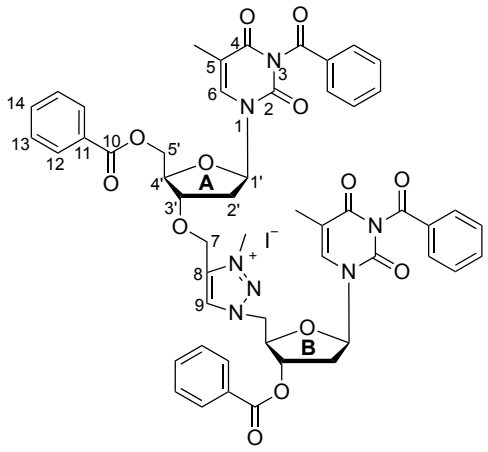
^{1}H NMR ($(CD_3)_2SO$, 500MHz): δ 8.21 (s, 1 H, H-9), 8.04 - 7.93 (m, 8 H, H-12), 7.82 - 7.73 (m, 3 H, H-14, H-6), 7.71 - 7.64 (m, 3 H, H-14, H-6), 7.62 - 7.47 (m, 8 H, H-13), 6.23 (app. t, *J* = 6.9 Hz, 1 H, H-1'B), 6.16 (app. t, *J* = 6.9 Hz, 1 H, H-1'A), 5.55 (app. dt, *J* = 6.9, 3.0 Hz, 1 H, H-3'B), 4.87 (dd, *J* = 14.4, 5.3 Hz, 1 H, H-5'B), 4.82 (dd, *J* = 14.4, 7.6 Hz, 1 H, H-5'B), 4.66 (s, 2 H, H-7), 4.62 - 4.53 (m, 2 H, H-4'B, H-5'A), 4.52 - 4.42 (m, 2 H, H-3'A, H-5'A), 4.36 - 4.29 (m, 1 H, H4'A), 2.69 (app. dt, *J* = 14.6, 6.9 Hz, 1 H, H-2'B), 2.54 (ddd, *J* = 14.6, 6.9, 3.0 Hz, 1 H, H-2'B), 2.48 - 2.41 (m, 2 H, H-2'A), 1.91 (d, *J* = 0.9 Hz, 3 H, CH_3-*thymine*), 1.68 (d, *J* = 0.8 Hz, 3 H, CH_3-*thymine*)

^{13}C NMR ($(CD_3)_2SO$, 126MHz): δ 169.5 169.4 165.5 165.1 (C-10), 162.4 162.3 (C-4), 149.0 148.9 (C-2), 143.9 (C-8), 137.3 136.6 (C-6), 135.5 133.7 133.6 (C-14), 131.0 (C-11), 130.40 130.38 129.5 129.4 (C-12), 129.2 129.1 128.9 128.7 (C-13), 124.8 (C-9), 110.0 109.8 (C-5), 85.3 (C-1'B), 85.0 (C-1'A), 81.7 (C-4'A), 81.5 (C-4'B), 78.4 (C-3'A), 74.8(C-3'B), 64.4 (C-5'A), 62.2(C-7), 51.0 (C-5'B), 36.3 (C-2'A), 35.4 (C-2'B), 12.0 (CH_3-*thymine*), 11.8 (CH_3-*thymine*)

LRMS: [ESI-TOF] m/z (%): $[M+Cl]^-$ 998.0 (100)

HRMS [ESI-TOF] m/z: $[M + Na]^+$ Calcd for $C_{51}H_{45}N_7O_{13}Na$ 986.2968; Found 986.2936

3-*N*-benzoyl-3',5'-*O*-benzoyl-methoxytriazolium iodide thymidine dimer 3.15



$C_{52}H_{48}N_7O_{13}I$
(1105.90)

To a solution of 3-*N*-benzoyl-3',5'-*O*-benzoyl-methoxytriazolyl-thymidine dimer **3.16** (2.00 g, 2.07 mmol) in DMF (4 mL) methyl iodide (4 mL, 64.2 mmol) was added at rt. After stirring for 15 h at 55 ˚C, the solvent was removed under vacuum and the crude was directly purified by column chromatography on silica gel (4:96 to 6:94 MeOH/DCM). Compound **3.15** was isolated with 1 eq of residual DMF and as a yellowish solid (2.11 g, 1.91 mmol, 92%).

^{1}H NMR ($(CD_3)_2CO$, 500 MHz) δ 9.29 (s, 1 H, H-9), 8.29 (s, 1 H, H-6), 8.15 - 8.10 (m, 2 H, H-12), 8.10 - 8.04 (m, 4 H, H-12), 8.04 - 8.01 (m, 2 H, H-12), 7.79 - 7.73 (m, 2 H, H-14), 7.72 (d, *J* = 1.2 Hz, 1 H, H-6), 7.71 - 7.64 (m, 2 H, H-14), 7.63 - 7.49 (m, 8 H, H-13), 6.45 (dd, *J* = 7.9, 6.5 Hz, 1 H, H-1'B), 6.32 (app. t, *J* = 6.9 Hz, 1 H, H-1'A), 5.68 (app. dt, *J* = 7.9, 2.9 Hz, 1 H, H-3'B), 5.61 (dd, *J* = 14.1, 7.6 Hz, 1 H, H-5'B), 5.44 (dd, *J* = 14.1, 3.9 Hz, 1 H, H-5'B), 5.30 (s, 2 H, H-7), 4.91 (app. dt, *J* = 6.4, 3.4 Hz, 1 H, H-3'A), 4.75 (app. dd, *J* = 7.6, 3.9 Hz, 1 H, H-4'B), 4.75 – 4.71 (m, 1 H, H-5'A), 4.67 (dd, *J* = 12.1, 4.6 Hz, 1 H, H-5'A), 4.56 (app. dt, *J* = 4.6, 3.4 Hz, 1 H, H-4'A), 4.52 (s, 3 H, CH_3-triazole), 3.27 (app. dt, *J* = 14.7, 7.9 Hz, 1 H, H-2'B), 2.70 (ddd, *J* = 14.7, 6.5, 2.9 Hz, 1 H, H-2'B), 2.68 - 2.61 (m, 2 H, H-2'A), 2.02 (d, *J* = 1.2 Hz, 3 H, CH_3-*thymine*), 1.70 (d, *J* = 1.2 Hz, 3 H, CH_3-*thymine*)

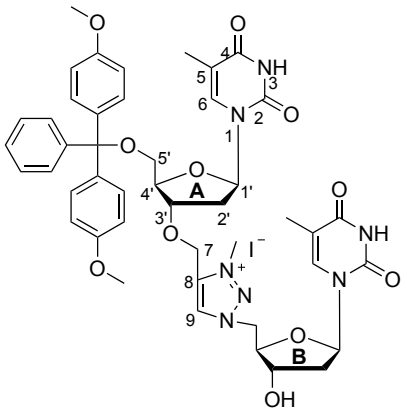
^{13}C NMR ($(CD_3)_2CO$, 126 MHz) δ 169.5 169.4 165.84 165.77 (C-10), 162.5 161.9 (C-4), 149.5 149.3 (C-2), 140.8 (C-8), 138.0 136.1 (C-6), 135.2 135.0 133.6 133.4 (C-14), 132.0 131.9 (C-11), 130.9(C-9), 130.45 130.4 129.7 129.5 (C-12), 129.4 129.3 128.8

128.6 (C-13), 111.0 110.5 (C-5), 86.5 (C-1'B), 85.4 (C-1'A), 82.2 (C-4'A), 81.3 (C-4'B), 80.0 (C-3'A), 74.5 (C-3'B), 64.4 (C-5'A), 59.4 (C-7), 54.4 (C-5'B), 38.9 (CH_3-*triazole*), 37.0 (C-2'A), 35.4 (C-2'B), 11.52 (CH_3-*thymine*), 11.46 (CH_3-*thymine*)

LRMS: [ESI-TOF] m/z (%): $[M]^+$ 978.2 (100)

HRMS [ESI-TOF] m/z: $[M]^+$ Calcd for $C_{52}H_{48}N_7O_{13}$ 978.3305; Found 978.3267

5'-*O*-(4,4'-dimethoxytrityl)-methoxytriazolium iodide thymidine dimer 3.21



$C_{45}H_{50}N_7O_{11}I$
(991.84)

3-*N*-benzoyl-3',5'-*O*-benzoyl-methoxytriazolium iodide thymidine dimer **3.15** (1.64 g, 1.48 mmol) was dissolved in DCM (6 mL) and a solution of NH_3 (28% in H_2O, 12 mL) was added. The reaction mixture was stirred in a sealed tube for 20 h at rt and the solvent was removed under vacuum. The remaining aqueous solution was washed with DCM (3 X 12 mL) and with EtOAc (3 X 12 mL) and concentrated *in vacuo.* The white solid obtained was dissolved in dry pyridine (8 mL) and 4,4'-dimethoxytrityl chloride (0.881 g, 2.60 mmol) was added. After 3 h at rt, MeOH (10 mL) was added and the solution was stirred for an additional 30 min before removal of the solvent under vacuum. The residue was dissolved in DCM (30 mL) and washed with a saturated aqueous solution of $NaHCO_3$, followed by H_2O and brine. The organic layer was dried over Na_2SO_4 and concentrated *in vacuo*. The crude was purified by column chromatography on silica gel (4.75:0.5:94.75 MeOH/pyridine/DCM then 10:10:30 MeOH/NH_3/DCM). Compound **3.21** was isolated with 1 eq. of residual pyridine as a white solid (0.379 g, 0.382 mmol, 26%).

^{1}H NMR ($(CD_3)_2SO$, 400 MHz): δ 8.93 (s, 1 H, H-9), 7.49 (s, 2 H, H-6), 7.41 - 7.35 (m, 4 H, ArH), 7.26 - 7.24 (m, 5 H, ArH), 6.91 - 6.88 (m, 4 H, ArH), 6.26 - 6.07 (m, 2 H, H-1'), 5.01 (dd, *J* = 14.2, 4.2 Hz, 1 H, H-5'B), 4.96 (dd, *J* = 14.2, 8.0 Hz, 1 H, H-5'B), 4.85 (s, 2 H, H-7), 4.45 - 4.39 (m, 1 H, H-3'), 4.35 (dt, *J* = 7.2, 4.3 Hz, 1 H, H-3'), 4.21 (s, 3 H, CH_3-triazole), 4.16 (dt, *J* = 8.0, 4.2 Hz, 1 H, H-4'B), 4.12 (dd, *J* = 5.9, 3.2 Hz, 1 H, H-4'A), 3.74 (s, 6 H, OMe), 3.26 - 3.11 (m, 2 H, H-5'A), 2.46 - 2.27 (m, 3 H, H-2'), 2.14 (ddd, *J*

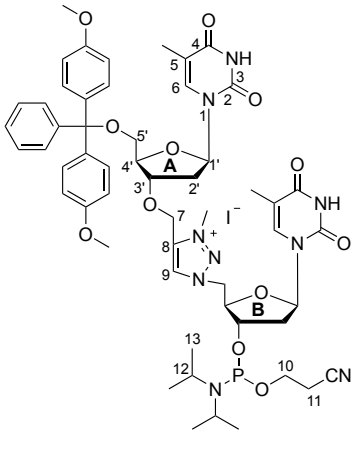
= 13.3, 6.8, 4.4 Hz, 1 H, H-2'), 1.80 (s, 3 H, CH_3-*thymine*), 1.44 ppm (s, 3 H, CH_3-*thymine*)

^{13}C NMR ($(CD_3)_2SO$, 101 MHz): δ 158.7 (Cq), 150.1 (C-4), 140.7 (C-8), 137.2 (Cq), 136.6 (C-6), 135.8 (Cq), 130.5 (ArCH), 130.2 (ArCH), 128.5 (ArCH), 128.2 (ArCH), 124.4 (C-9), 113.8 (CHAr), 110.4 (C-5), 110.2 (C-5), 86.6 (Cq-*trityl*), 84.3 (C-1'), 83.1 (C-4'), 80.9 (C-3'), 70.9 (C-3'), 64.4 (C-5'), 59.0 (C-7), 55.6 (OMe), 38.7 (C-2'), 38.0 (C-2'), 12.7 (CH3-*thymine*) , 12.2 (CH3-*thymine*)

LRMS: [ESI-TOF] m/z (%): $[M]^+$ 864.3 (100)

HRMS [ESI-TOF] m/z: $[M]^+$ Calcd for $C_{45}H_{50}N_7O_{11}$ 864.3563; Found 864.3553

5′-*O*-(4,4′-dimethoxytrityl)-methoxytriazolium iodide thymidine dimer phosphoramidite 3.14



$C_{54}H_{67}N_9O_{12}PI$
(1192.06)

5′-*O*-(4,4′-dimethoxytrityl)-methoxytriazolium iodide thymidine dimer **3.21** (0.252 g, 0.254 mmol) was co-evaporated with dry pyridine (10 mL) and dissolved in a 1:5 mixture of anhydrous DMF/DCM (6 mL) under an argon atmosphere. DIPEA (0.26 mL, 1.49 mmol) was added, followed by the dropwise addition of 2-cyanoethyl-*N*,*N*-diisopropylchlorophosphoramidite (0.11 mL, 0.636 mmol). The reaction was stirred at rt for 3 h and the solvent was removed under vacuum. The residue was diluted with anhydrous DCM (10 mL), washed with degassed saturated aqueous KCl (10 mL), dried over Na_2SO_4 and concentrated *in vacuo*. The crude product was precipitated 3 times consecutively in anhydrous degassed hexane (100 mL) from anhydrous degassed DCM (8 mL). The product was co-evaporated with distilled DCM (5 mL) and dried *in vacuo* to afford phosphoramidite **3.14** as a 1:1 mixture of diastereisomer **1** and **2** and as a white solid (0.256 g, 0.215 mmol, 85%).

^{1}H NMR (CD_2Cl_2, 400 MHz): δ 9.21, 9.08, 9.05 (3s, 2 H, NH-3), 8.92 (m, 1 H, H-9), 7.56 - 7.06 (m, 11 H, H-6, ArH), 6.76 - 6.74 (m, 4 H, ArH), 6.16 - 6.14 (m, 1 H, H-1′), 5.81 - 5.78 (m, 1 H, H-1′), 5.69 - 5.27 (m 1 H, H-5′), 5.22 - 4.64 (m, 3 H, H-7, H-5′), 4.52 - 4.10 (m, 6 H, H-3′, H-4′, CH_3-*triazole*), 4.10 - 3.96 (m, 1 H, H-4′), 3.93 - 3.44 (m, 10 H, H-10, H-12, OMe), 3.43 - 3.12 (m, 2 H, H-5′), 2.93 - 2.67 (m, 2 H, H-2′), 2.67 - 2.32 (m, 2 H, H-12), 2.32 - 2.25 (m, 2 H, H-2′), 1.85 - 1.64 (m, 3 H, CH_3-*thymine*), 1.44 - 1.34 (m, 3 H, CH_3-*thymine*), 1.31 - 0.95 (m, 12 H, H-13)

^{31}P NMR (CD_2Cl_2, 162 MHz): δ 149.34 , 148.79

6.2 Oligonucleotides synthesis and studies

6.2.1 General conditions: oligonucleotide synthesis and purification

DNA reagents including standard DNA phosphoramidites and solid supports (CPG resin) were purchased from Link Technologies Ltd. Oligonucleotides were synthesised on an Applied Biosystems 394 automated DNA/RNA synthesiser. Phosphoramidite cycles, including acid-catalysed detritylation, coupling, capping and iodine oxidation steps, were undertaken in 1.0 µmole scale. Standard DNA phoshoramidites were used for all the oligonucleotide sequences. Coupling efficiencies and overall oligonucleotide yields were determined by the automated trityl cation conductivity monitoring facility of the synthesiser and were ≥98.0% for all cases. Standard phosphoramidite monomers were dissolved in anhydrous acetonitrile to a concentration of 0.1 M immediately prior to use. The coupling time for A, G, C and T monomers, was set to 25 s. The cleavage of oligonucleotides from the solid support (CPG resin) and subsequent deprotection was undertaken by suspending the resin in concentrated aqueous ammonia for 60 min at rt followed by heating in a sealed tube for 5 h at 55 °C.

Oligonucleotides were purified by RP-HPLC using a Gilson HPLC system with ABI Aquapore C8 column (8 mm x 250 mm, pore size 300 Å), with a gradient of acetonitrile in triethylammonium bicarbonate (TEAB) (0 to 50% buffer B over 20 min, flow rate 4 mL/min). The following elution buffers were used: buffer A (0.1 M triethylammonium bicarbonate, pH 7.5), buffer B (0.1 M triethylammonium bicarbonate with 50% acetonitrile, pH 7.5). The fractions from HPLC were evaporated without additional desalting.

All oligonucleotides were characterized by negative-mode electrospray HPLC-mass spectrometry, using a Bruker micrOTOF mass spectrometer and an Acquity UPLC system with a BEH C18 1.7 µm column (Waters). Raw data were processed and deconvoluted using the Data Analysis function of the Bruker Daltronics Compass^{TM} 1.3 software package.

6.2.2 Synthesis and properties of amine and *N*-functionalised amine modified oligonucleotide analogues

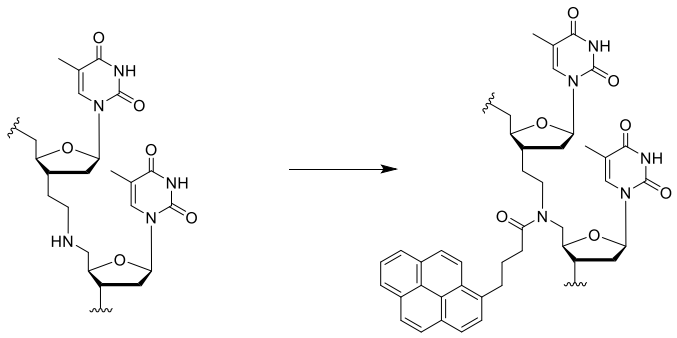
6.2.2.1 Oligonucleotide synthesis and purification

Oligonucleotides were synthesised following the above procedure with some alterations. The modified phosphoramidite monomer (**2.3**) was dissolved in anhydrous DCM/MeCN mixture (3:1) to a concentration of 0.1 M immediately prior to use. The coupling time for A, G, C and T monomers was set to 25 s, however, the coupling time for the modified monomer was extended to 360 s. After synthesis, oligonucleotides were kept on the solid support with 4,4'-dimethoxytrityl protecting group on the 5' hydroxyl.

Modified oligonucleotides on resin were treated with a solution of piperidine (20% in DMF) for 20 min for selective deprotection of FMOC protecting group. After functionalisation, modified oligonucleotides were cleaved from the solid support (CPG resin) and deprotected by suspending the resin in concentrated aqueous ammonia followed by heating in a sealed tube for 5 h at 55 °C. Manual removal of 5'-dimethoxytrityl protecting group, was performed by treating the freeze dried oligonucleotides with an aqueous solution of acetic acid (80%, 1 mL) for 1 h at rt. The acid was then neutralised by addition of ammonia and the oligonucleotide was desalted using NAP-25 gel filtration columns (GE Healthcare).

Oligonucleotides were purified by RP-HPLC using a Gilson HPLC system with ABI Aquapore C8 column (8 mm x 250 mm, pore size 300 Å). Oligonucleotides with 5'-dimethoxytrityl protecting group were purified using a gradient of acetonitrile in triethylammonium bicarbonate (TEAB) (0 to 50% buffer B over 20 min, flow rate 4 mL/min). The following elution buffers were used: buffer A (0.1 M triethylammonium bicarbonate, pH 7.5), buffer B (0.1 M triethylammonium bicarbonate with 80% acetonitrile, pH 7.5). The fractions from HPLC were evaporated without additional desalting.

6.2.2.2 Pyrene functionalisation of amine modified backbone oligonucleotides



Resin containing amine modified oligonucleotide (**ODN-2.1, ODN-2.2, ODN-2.3**) was transferred into a glass vial (5 mg, ~ 0.16 µmol) and treated with 1-pyrenebutyric acid *N*-hydroxysuccinimide ester (5 mg, 12.3 µmol) in a solution of triethylamine (10% in DMF, 30 µL). The resin was then heated at 55 °C for 20 h and washed with MeCN (3 X 150 µL) before cleavage of oligonucleotide from the solid support, deprotection and RP-HPLC purification following general conditions.

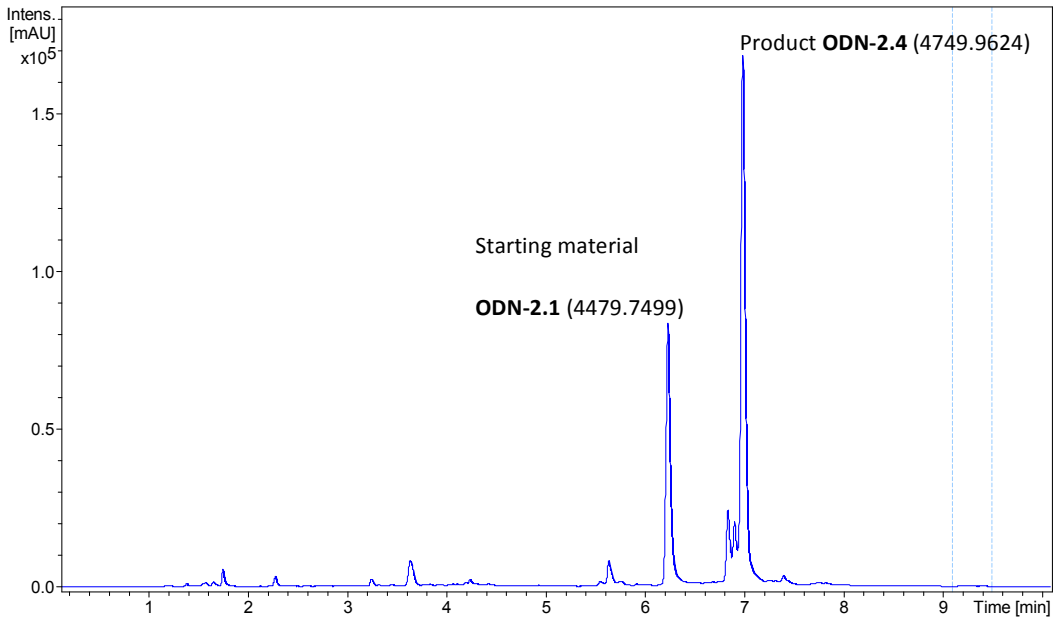
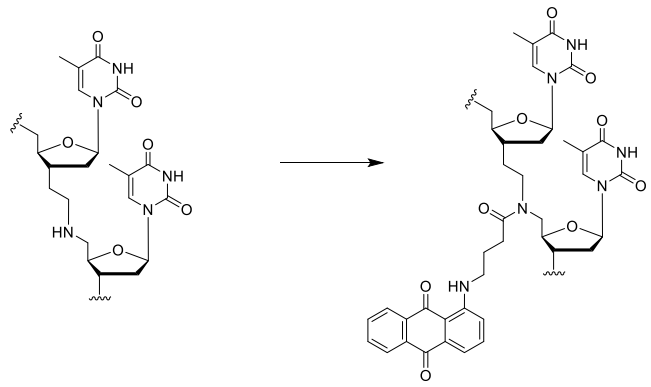


Figure 6-1: HPLC trace of functionalisation of **ODN-2.1** with pyrene

6.2.2.3 Anthraquinone functionalisation of amine modified backbone oligonucleotides



Resin containing amine modified oligonucleotide (**ODN-2.2** and **ODN-2.3)** was transferred into a glass vial (5 mg, ~ 0.16 µmol) and treated with N-(anthraquinonyl)amino butyric acid *N*-hydroxysuccinimide ester (5 mg) in a solution of triethylamine (10% in DMF, 30 µL). The resin was heated at 55 °C for 20 h, washed with MeCN (3 X 150 µL) before cleavage of oligonucleotide from the solid support, deprotection and RP-HPLC purification following general conditions.

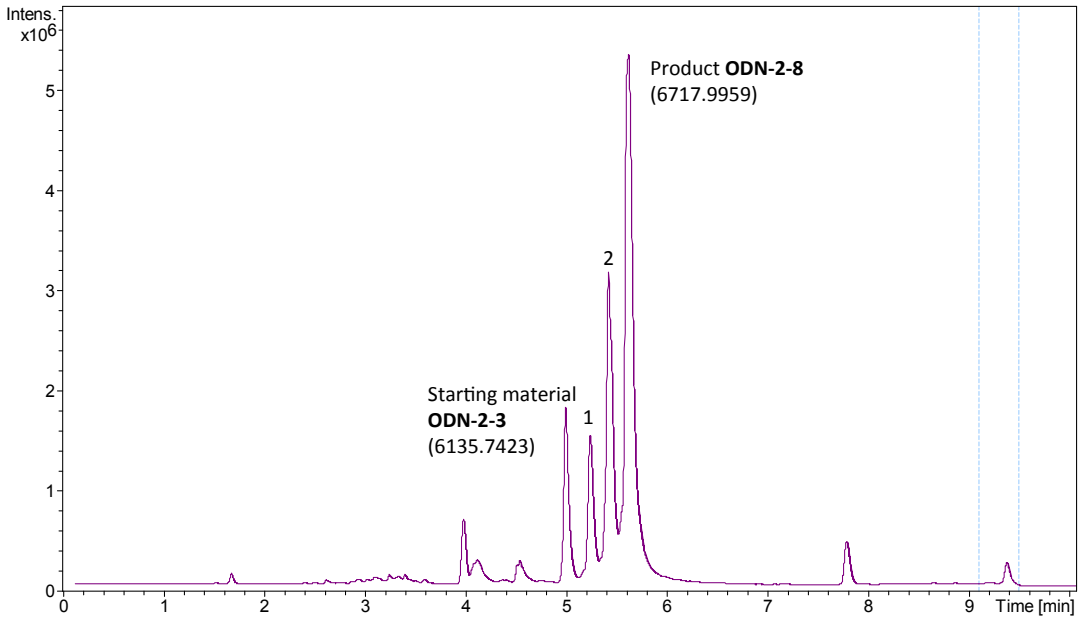
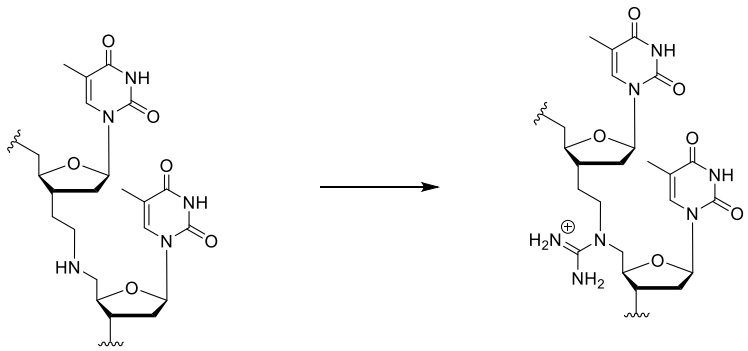


Figure 6-2: HPLC trace of functionalisation of **ODN-2.3** with anthraquinone. A small amount of mono-functionalised oligonucleotides (1 and 2) was also formed**.**

6.2.2.4 Guanidine functionalisation of amine modified backbone oligonucleotides



Resin containing amine modified oligonucleotide (**ODN-2.1**, **ODN-2.2** or **ODN-2.3**) was transferred into a glass vial (5 mg, ~ 0.16 µmol) and treated with N,N'-BIS-(2-cyano)ethoxycarbonyl)-2-methyl-2-thiopseudourea (67 mg) in a solution of triethylamine (10% in DMF, 30 µL). The oligonucleotide on resin was then heated at 70 °C for 5 h. The resin was washed with DMF (3 X 150 µL) followed by MeCN (3X 150 µL) before cleavage of oligonucleotide from the solid support, deprotection and RP-HPLC purification following general conditions.

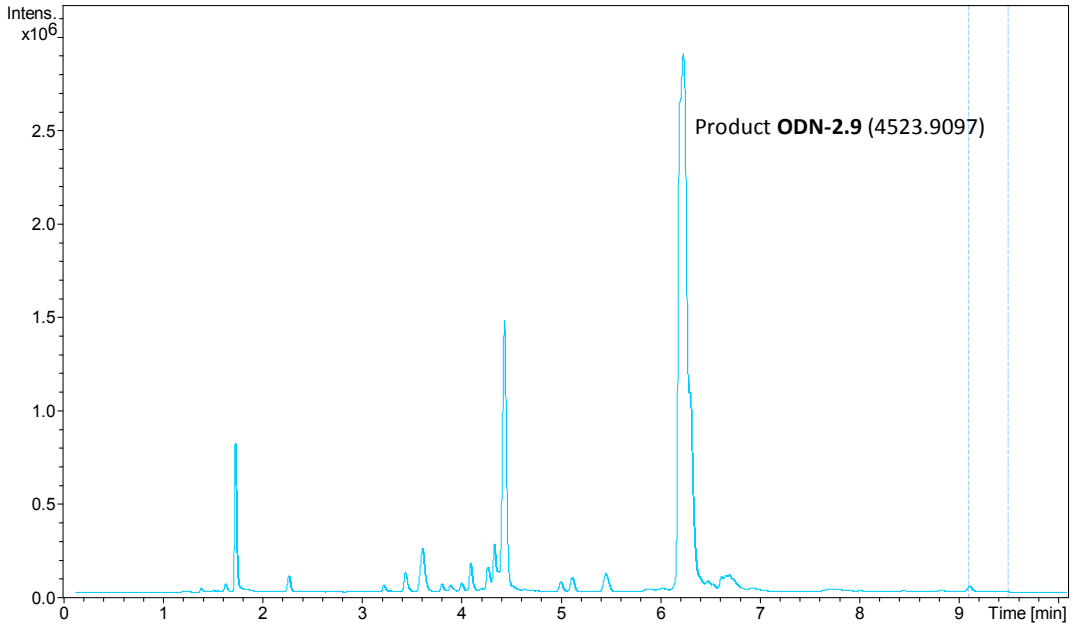


Figure 6-3: HPLC trace of functionalisation of **ODN-2.1** with guanidine

Table 6-1: Characterisation of modified oligonucleotides. Insertion of amine, pyrene, anthraquinone and guanidine modified backbones are represented respectively by: *a, py, an* and *g.*

		DMT-ON		DMT-OFF	
Code	**Sequences (5' to 3')**	**$Mw_{Calc.}$**	**$Mw_{found.}$**	**$Mw_{Calc.}$**	**$Mw_{found.}$**
ODN-2.1	CGTTCGCT$_a$TGCAGC	4481	4480	4179	4178
ODN-2.4	CGTTCGCT$_{py}$TGCAGC	4751	4750	4449	4448
ODN-2.9	CGTTCGCT$_g$TGCAGC	4524	4524	4222	4221
ODN-2.2	ACCCATTCCT$_a$TCGTTCCACT	6189	6188	5887	5886
ODN-2.10	ACCCATTCCT$_g$TCGTTCCACT	6232	6231	5930	5929
ODN-2.5	ACCCATTCCT$_{py}$TCGTTCCACT	6459	6459	6157	6156
ODN-2.7	ACCCATTCCT$_{an}$TCGTTCCACT	6480	6480	6178	6177
ODN-2.3	ACCCAT$_a$TCCTTCGT$_a$TCCACT	6137	6135	5835	5833
ODN-2.11	ACCCAT$_g$TCCTTCGT$_g$TCCACT	6221	6221	5919	5919
ODN-2.6	ACCCAT$_{py}$TCCTTCGT$_{py}$TCCACT	6675	6676	6373	6373
ODN-2.8	ACCCAT$_{an}$TCCTTCGT$_{an}$TCCACT	6717	6718	6415	6415

6.2.2.5 UV melting studies of amine modified oligonucleotides and analogues

Measurements were made on a Cary 4000 UV-Vis spectrometer with Cary temperature controller. Cary Win UV Thermal software was used with an absorption wavelength of 260 nm. Samples were analysed in 1 mL cuvettes and were made to 1.5 µM oligonucleotide concentration in phosphate buffer (10 mM) with NaCl (200 mM) at pH 7.0. The temperature was varied between 20 and 85 °C at 1 °C/min. T_m values were calculated with Cary Win UV Thermal application Software, and averaged from six successive T_m values.

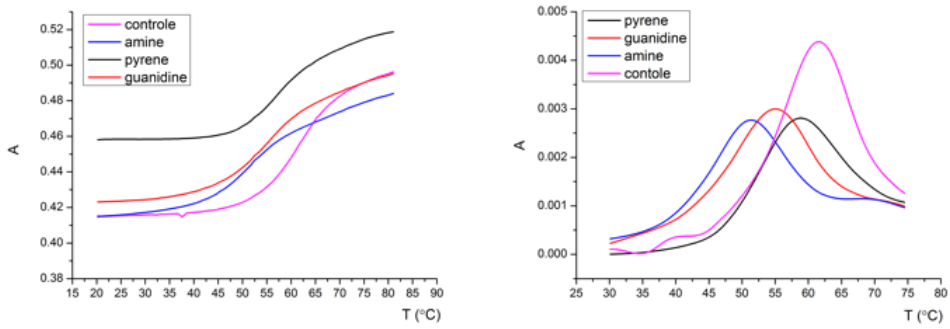


Figure 6-4: Melting and first derivative curves of amine oligonucleotide **ODN-2.1** along with pyrene **ODN-2.4**, and guanidine **ODN-2.9** oligonucleotide analogues.

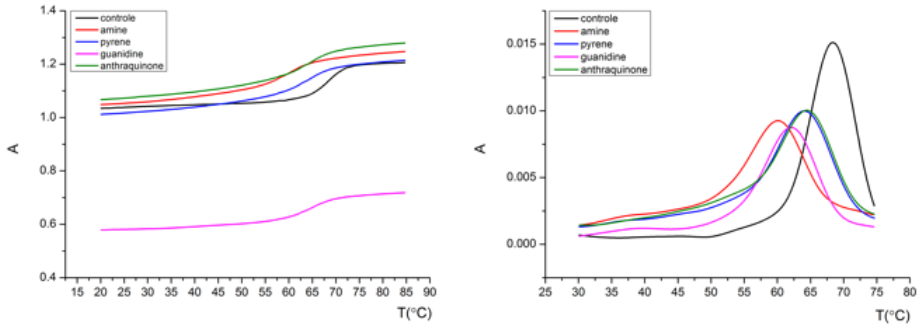


Figure 6-5: Melting and first derivative curves of amine oligonucleotide **ODN-2.2** along with pyrene **ODN-2.5**, anthraquinone **ODN-2.7** and guanidine **ODN-2.10** oligonucleotide analogues.

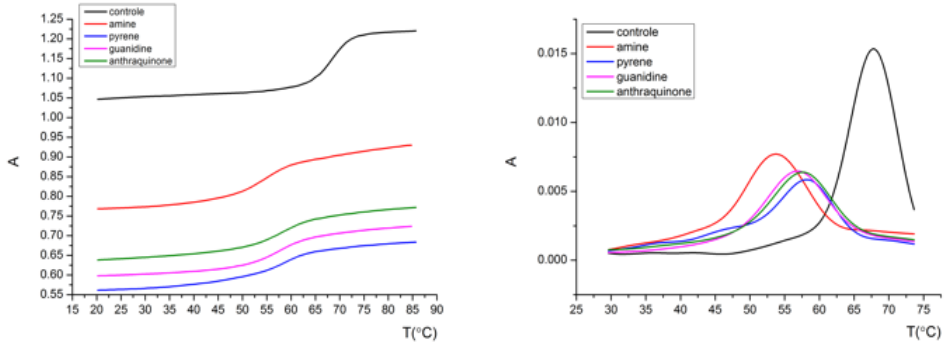


Figure 6-6: Melting and first derivative curves of amine oligonucleotide **ODN-2.3** along with pyrene **ODN-2.6**, anthraquinone **ODN-2.8** and guanidine **ODN-2.11** oligonucleotide analogues.

6.2.3 Synthesis and study of the hybridisation properties of DNA containing triazole modified backbones

6.2.3.1 Azide labeling and cyclisation of oligonucleotides ODN-3.4 and ODN-3.5

4-azidobutyrate NHS ester (2 mg) was added post synthetically to each 1.0 µmol synthesis of the amino-modified oligonucleotides **ODN-3.4** and **ODN-3.5** in 120 µL of DMSO/0.5 M $NaHCO_3$ buffer (1:2) at pH 8.75 (4 h, rt). The fully-labelled oligonucleotides were desalted using NAP-10 Sephadex columns (GE Healthcare) and purified by RP-HPLC. Following this the oligonucleotides in NaCl solution (1.0 mL, 200 mM) were heated at 80 °C for 5 min, cooled down slowly and added to a degassed solution of sodium ascorbate (25.0 µmol in 50.0 µL 200 mM NaCl), tris-hydroxypropyl triazole ligand (17.5 µmol in 3.9 mL 200 mM NaCl) and $CuSO_4 \cdot 5\,H_2O$ (2.5 µmol in 25.0 µL 200 mM NaCl) . The reaction mixture was kept under argon at room temperature for 3 h, and two disposable NAP-25 Sephadex columns were used to remove reagents. The cyclic oligonucleotides were purified by RP-HPLC following the general conditions.

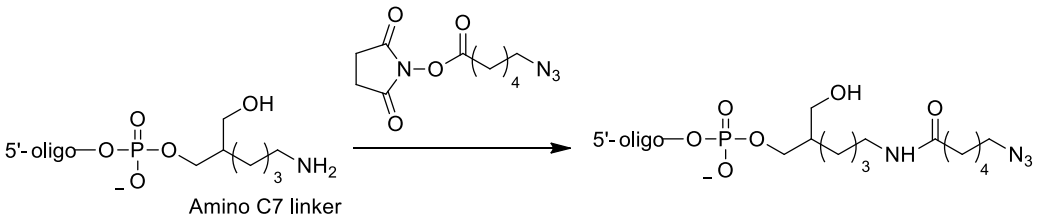


Figure 5.1: Post-oligonucleotide synthesis azide labelling

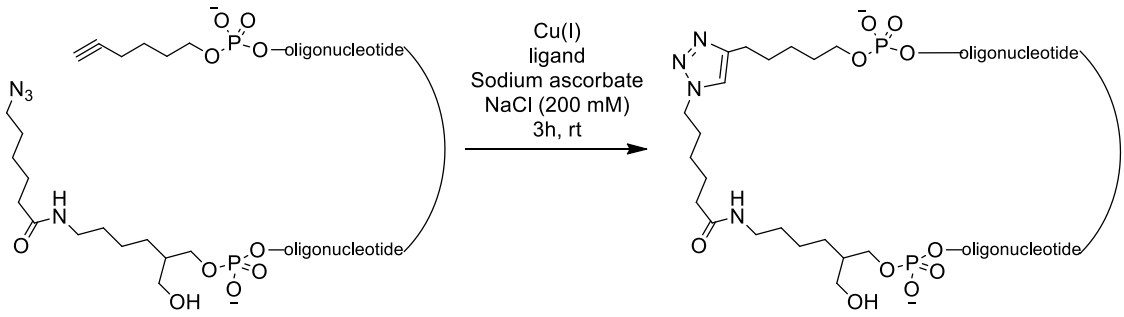
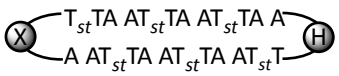
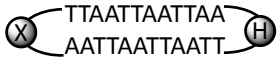


Figure 5.2: CuAAC click mediated oligonucleotide cyclisation

Table 6-2: Characterisation of oligonucleotides used for thermodynamic studies. Insertions of triazole backbone **3.II**, short triazole backbone **3.III** and methyl triazole backbone **3.IV** are represented respectively by: *t, st* and *Met*; X: triazole linkage; H: hexaethylene glycol

Code	oligonucleotide sequence (5' to 3')	$Mw_{Calc.}$	Mw_{Found}
ODN-3.1	CGACGT$_t$TTGCAGC	3952	3951
ODN-3.2	CGACGT$_{st}$TTGCAGC	3922	3921
ODN-3.3	CGACGT$_{Met}$TTGCAGC	3966	3966
ODN-3.4	hexynol-T$_{st}$TAAT$_{st}$TAAT$_{st}$TAA-**H**-T$_{st}$TAAT$_{st}$TAAT$_{st}$TAA-aminoC7	7885	7886
ODN-3.5	hexynol-TTAATTAATTAA-**H**-TTAATTAATTAA-aminoC7	8058	8059
ODN-3.6		7885	7886
ODN-3.7		8058	8059

6.2.3.2 UV melting analysis

UV melting analysis was monitored by optical absorption at 260 nm using a Varian Cary 4000 Scan UV-Visible spectrophotometer. The solutions had a total duplex concentration from 1 to 3 µM, with 10 mM phosphate buffer, 200 mM NaCl at pH 7.0. The optical path length was 1 cm, the temperature was varied between 20 and 85 °C at 1 °C/min. T_m values were calculated using Cary Win UV thermal application software and averaged from six successive T_m values.

6.2.3.3 Circular dichroism

Circular dichroism spectra were recorded on a Jasco J-720 spectropolarimeter. Samples were made to 10 µM oligonucleotide concentration in 10 mM phosphate buffer,200 mM NaCl at pH 7.0 Spectra were recorded between 200-320 nm at a rate of 100 nm/min, with step resolution of 0.2 nm, bandwidth of 1.0 nm and sensitivity of 50 mdeg. Ten successive spectra were recorded and an average was taken. A blank (buffer) baseline was subtracted from each spectrum.

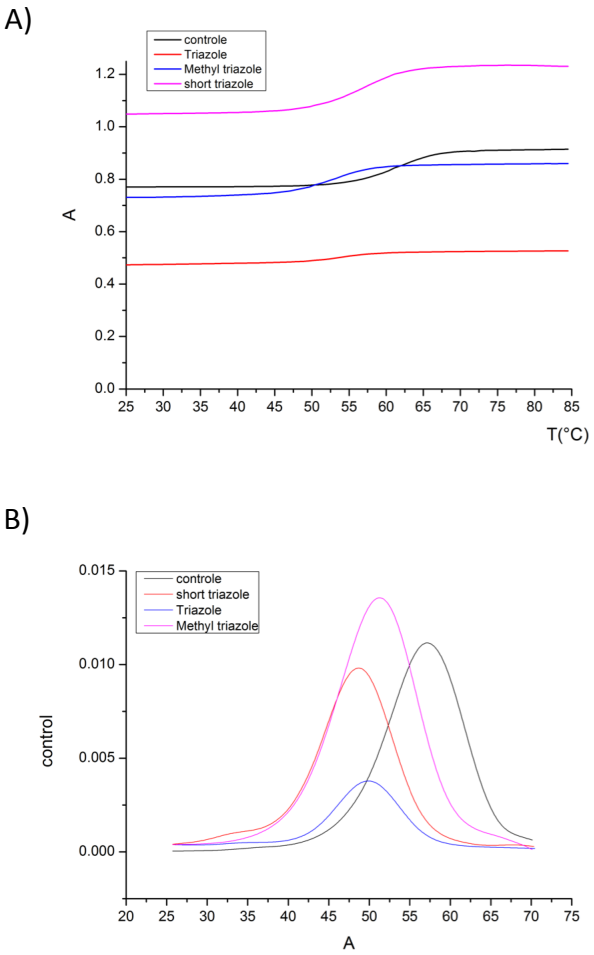


Figure 6-7: **A)** Melting curves and **B)** first derivative curves of triazole **ODN-3.1**, short-triazole **ODN-3.2**, methyl-triazole **ODN-3.3**, and unmodified control **ODN-3.I** oligonucleotides.

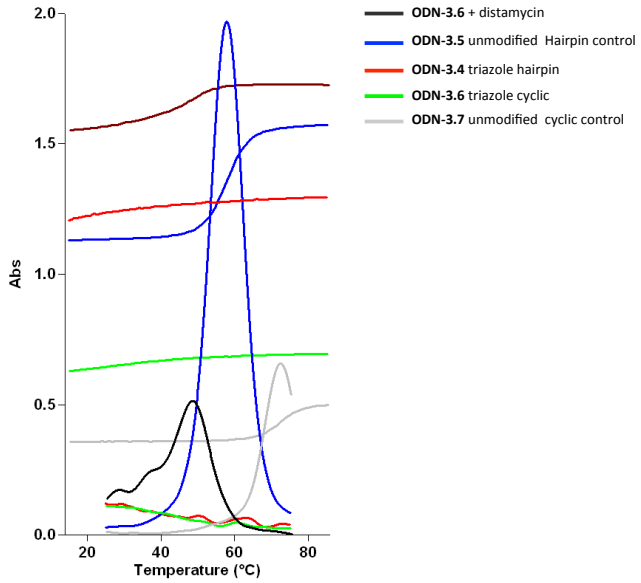


Figure 6-8: Melting and first derivative curves of oligonucleotides containing alternating short-triazole **3.III** in comparison with unmodified control oligonucleotides.

6.2.4 Study of the biocompatiblity of modified backbones for gene synthesis applications

Table 6-3: Characterisation of oligonucleotides for enzymatic studies. Insertions of triazole backbone 3.II, short triazole backbone 3.III, methyl triazole backbone 3.IV and amide triazole backbone 2.III are represented respectively by: *t, st, Met* and *am.* FAM is 6-carboxyfluorescein.

Code	oligonucleotide sequence (5' to 3')	$MW_{Calc.}$	$MW_{Found.}$
P-4.1	FAM-GGTTATGTGTGTCGGCAG	6138	6138
ODN-4.1	CTCAGTCT$_t$TCTGTACTGCCGACACACATAACC	9665	9664
ODN-4.2	CTCAGTCT$_{st}$TCTGTACTGCCGACACACATAACC	9635	9635
ODN-4.3	CTCAGTCT$_{Met}$TCTGTACTGCCGACACACATAACC	9680	9679
P-4.2	ACGTTAGCACGAAGAGGC	5558	5557
P-4.3	GCTTACGACGAAGAACGG	5558	5557
P-4.4	GCGTACGAGAACCGAATG	5558	5557
P-4.5	GCATTCGAGCAACGTAAG	5533	5533
P-4.6	ATTGGGTGTGTTCGCGAG	5602	5601
P-4.7	TGTGTGTATTGGCCGAGG	5602	5601
P-4.8	GGATTGTTGTGTGCGACG	5602	5601
P-4.9	GGTTATGTGTGTCGGCAG	5602	5601
P-4.10	AGCAACAGCTGGTAGACG	5558	5557
P-4.11	TGGCACGGATTTGTGGTG	5602	5601
ODN-4.4	ACGTTAGCACGAAGAGGCATCTTAGCACACAATCTCACACTCTGG AATTCACACTGACAATACTCGCGAACACACCCAAT	24451	24451
ODN-4.5	GCTTACGACGAAGAACGGATCT$_t$TAGCACACAATCTCACACTCTG GAATTCACACTGACAATACCTCGGCCAATACACACA	24453	24453
ODN-4.6	GCGTACGAGAACCGAATGATCT$_{st}$TAGCACACAATCTCACACTCTG GAATTCACACTGACAATACGTCGCACACAACAATCC	24423	24423
ODN-4.7	GCATTCGAGCAACGTAAGATCT$_{Met}$TAGCACACAATCTCACACTCT GGAATTCACACTGACAATACTGCCGACACACATAACC	24442	24442
ODN-4.8	AGCAACAGCTGGTAGACGATCT$_{am}$TAGCACACAATCTCACACTCT GGAATTCACACTGACAATACACCACAAATCCGTGCCA	24413	24412

6.2.4.1 Linear extension of modified template ODN-4.1 to ODN-4.3

In a total volume of 20 µL, 66 pmol of FAM labeled DNA primer (**P-4.1**), 66 pmol of template (**ODN-4.1**, **ODN-4.2** or **ODN-4.3**) and 0.2 mM dNTP, were mixed with the polymerase enzyme (2 unit for GoTaq G2 and Klenow, 1 unit for KOD XL) and 1X buffer (supplied with the enzymes). The reaction mixtures were heated at 60 °C for KOD XL and GoTaq G2 and at 37 °C for Klenow. The incubation was stopped after each time point and 20 µL formamide was added before analysis by 10% denaturing PAGE in 1X TBE buffer. Gel images were taken using Syngene G: BOX with the gene-snap imaging software.

6.2.4.2 Real time-PCR of oligonucleotides ODN-4.4 to ODN-4.7 for Ct measurement

In a total volume of 20 µL, 5 pg of template, 0.5 µM of each primer, 0.2 mM dNTP and SYBR Green (0.6 µL, 5X) were mixed with 0.5 unit of GoTaq® G2 DNA polymerase and 4 µL of 5X GoTaq® Reaction Buffer. Amplification was performed on a real-time PCR instrument Bio-Rad CFX96 using the following procedure: 95 °C (initial denaturation) for 2 min, followed by 25 cycles of 95 °C (denaturation) for 15 s, 54 °C (annealing) for 20 s, 72 °C (extension) for 30 s. Further extension was carried out at 72 °C for 5 min. Cycle threshold values C_T were averaged from 3 C_T values and using a single fluorescence threshold at 497.84.

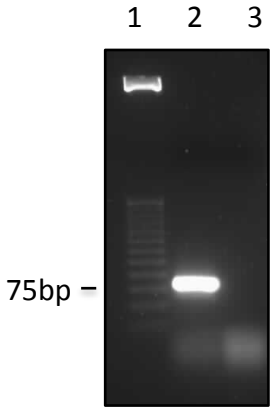


Figure 6-9: PCR amplification of oligonucleotide **ODN-4.7**. Lane 1: 25 bp ladder, Lane 2: PCR using template **ODN-4.7**, Lane 3: negative control PCR without template **ODN-4.7**.

6.2.4.3 Linear extension of oligonucleotides ODN-4.4 to ODN-4.8 for sequencing

Linear copy products were prepared using 60 pmol of template and 60 pmol of primer in a total reaction volume of 60 µL with 6 µL of 10X NEBuffer 2, 0.2 mM dNTP and 1.2 µL of DNA Polymerase I, Large (Klenow) Fragment (5u/µL). The reaction mixture was left at 37 °C for 1 h then the DNA polymerase was inactivated by heating at 70 °C for 20 min.

6.2.4.4 PCR of oligonucleotides for sequencing

PCR products were generated using 0.5 unit of GoTaq® G2 DNA polymerase with 4 µL of 5X Green GoTaq® Reaction Buffer in a total reaction volume of 20 µL with 5 ng of the DNA template, 0.5 µM of each primer and 0.2 mM dNTP. PCR amplification was performed on a BIO-RAD T100 Thermal Cycler using the following cycling conditions: 95 °C (initial denaturation) for 2 min then 25 cycles of 95 °C (denaturation) for 15 s, 54 °C (annealing) for 20 s and 72 °C (extension) for 30 s, and 72 °C (final extension) for 2 min.

Chapter 7
References

(1) WATSON, J. D.; CRICK, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, *171* (4356), 737–738.

(2) Franklin, R. E.; Gosling, R. G. The Structure of Sodium Thymonucleate Fibres. I. The Influence of Water Content. *Acta Crystallogr.* **1953**, *6* (8), 673–677.

(3) FRANKLIN, R. E.; GOSLING, R. G. Evidence for 2-Chain Helix in Crystalline Structure of Sodium Deoxyribonucleate. *Nature* **1953**, *172* (4369), 156–157.

(4) CHARGAFF, E.; LIPSHITZ, R.; GREEN, C. Composition of the Desoxypentose Nucleic Acids of Four Genera of Sea-Urchin. *J. Biol. Chem.* **1952**, *195* (1), 155–160.

(5) Metzker, M. L. Sequencing Technologies - the next Generation. *Nat. Rev. Genet.* **2010**, *11* (1), 31–46.

(6) Børsting, C.; Morling, N. Next Generation Sequencing and Its Applications in Forensic Genetics. *Forensic Sci. Int. Genet.* **2015**, *18*, 78–89.

(7) Cong, L.; Ran, F. A.; Cox, D.; Lin, S.; Barretto, R.; Habib, N.; Hsu, P. D.; Wu, X.; Jiang, W.; Marraffini, L. A.; Zhang, F. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **2013**, *339* (6121), 819–823.

(8) Doudna, J. A.; Charpentier, E. The New Frontier of Genome Engineering with CRISPR-Cas9. *Science (80-.).* **2014**, *346* (6213), 1258096–1258096.

(9) Lu, Y.; Liu, J. Functional DNA Nanotechnology: Emerging Applications of DNAzymes and Aptamers. *Curr. Opin. Biotechnol.* **2006**, *17* (6), 580–588.

(10) Mandel, M.; Marmur, J. *Nucleic Acids, Part B*; Methods in Enzymology; Elsevier, 1968; Vol. 12.

(11) Beaucage, S. L.; Jain, H. V. *The Chemical Synthesis of DNA and RNA Oligonucleotides for Drug Devellopment and Synthetic Biology Applications*; Elsevier, 2016.

(12) Michelson, A. M.; Todd, A. R. Nucleotides Part XXXII. Synthesis of a Dithymidine Dinucleotide Containing a 3’: 5'-Internucleotidic Linkage. *J. Chem. Soc.* **1955**, 2632.

(13) Letsinger, R. L.; Ogilvie, K. K. Nucleotide Chemistry. XIII. Synthesis of Oligothymidylates via Phosphotriester Intermediates. *J. Am. Chem. Soc.* **1969**, *91* (12), 3350–3355.

(14) Letsinger, R. L.; Mahadevan, V. Stepwise Synthesis of Oligodeoxyribonucleotides on an Insoluble Polymer Support. *J. Am. Chem. Soc.* **1966**, *88* (22), 5319–5324.

(15) Caruthers, M. H. Chemical Synthesis of DNA and RNA Analogs. *Acc. Chem. Res.* **1991**, *24* (9), 278–284.

(16) Merrifield, R. B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* **1963**, *85* (14), 2149–2154.

(17) Caruthers, M. Gene Synthesis Machines: DNA Chemistry and Its Uses. *Science (80).* **1985**, *230* (4723), 281–285.

(18) Pinheiro, A. V; Han, D.; Shih, W. M.; Yan, H. Challenges and Opportunities for Structural DNA Nanotechnology. *Nat. Nanotechnol.* **2011**, *6* (12), 763–772.

(19) Válóczi, A.; Hornyik, C.; Varga, N.; Burgyán, J.; Kauppinen, S.; Havelda, Z. Sensitive and Specific Detection of microRNAs by Northern Blot Analysis Using LNA-Modified Oligonucleotide Probes. *Nucleic Acids Res.* **2004**, *32* (22), 175.

(20) Kloosterman, W. P.; Wienholds, E.; de Bruijn, E.; Kauppinen, S.; Plasterk, R. H. A. In Situ Detection of miRNAs in Animal Embryos Using LNA-Modified Oligonucleotide Probes. *Nat. Methods* **2006**, *3* (1), 27–29.

(21) Chen, F.; Dong, M.; Ge, M.; Zhu, L.; Ren, L.; Liu, G.; Mu, R. The History and Advances of Reversible Terminators Used in New Generations of Sequencing Technology. *Genomics. Proteomics Bioinformatics* **2013**, *11* (1), 34–40.

(22) Shigdar, S.; Macdonald, J.; O'Connor, M.; Wang, T.; Xiang, D.; Al.Shamaileh, H.; Qiao, L.; Wei, M.; Zhou, S.-F.; Zhu, Y.; Kong, L.; Bhattacharya, S.; Li, C.; Duan, W. Aptamers as Theranostic Agents: Modifications, Serum Stability and Functionalisation. *Sensors* **2013**, *13* (10), 13624–13637.

(23) Deleavey, G. F.; Damha, M. J. Designing Chemically Modified Oligonucleotides for Targeted Gene Silencing. *Chem. Biol.* **2012**, *19* (8), 937–954.

(24) Bennett, C. F.; Swayze, E. E. RNA Targeting Therapeutics: Molecular Mechanisms of Antisense Oligonucleotides as a Therapeutic Platform. *Annu. Rev. Pharmacol. Toxicol.* **2010**, *50* (1), 259–293.

(25) Freier, S. The Ups and Downs of Nucleic Acid Duplex Stability: Structure-Stability Studies on Chemically-Modified DNA:RNA Duplexes. *Nucleic Acids Res.* **1997**, *25* (22), 4429–4443.

(26) Bell, N. M.; Micklefield, J. Chemical Modification of Oligonucleotides for Therapeutic, Bioanalytical and Other Applications. *Chembiochem* **2009**, *10* (17), 2691–2703.

(27) Sharma, V. K.; Sharma, R. K.; Singh, S. K. Antisense Oligonucleotides: Modifications and Clinical Trials. *Med. Chem. Commun.* **2014**, *5* (10), 1454–1471.

(28) Eckstein, F. Nucleoside Phosphorothioates. *J. Am. Chem. Soc.* **1966**, *88* (18), 4292–4294.

(29) Eckstein, F. Nucleoside Phosphorothioates. *J. Am. Chem. Soc.* **1970**, *92* (15), 4718–4723.

(30) Guga, P.; Koziołkiewicz, M. Phosphorothioate Nucleotides and Oligonucleotides - Recent Progress in Synthesis and Application. *Chem. Biodivers.* **2011**, *8* (9), 1642–1681.

(31) Agrawal, S.; Kandimalla, E. R. Antisense Therapeutics: Is It as Simple as Complementary Base Recognition? *Mol. Med. Today* **2000**, *6* (2), 72–81.

(32) Geary, R. S.; Baker, B. F.; Crooke, S. T. Clinical and Preclinical Pharmacokinetics and Pharmacodynamics of Mipomersen (kynamro(®)): A Second-Generation Antisense Oligonucleotide Inhibitor of Apolipoprotein B. *Clin. Pharmacokinet.* **2015**, *54* (2), 133–146.

(33) Pan, W.-H.; Clawson, G. A. Antisense Applications for Biological Control. *J. Cell. Biochem.* **2006**, *98* (1), 14–35.

(34) Corey, D. R. RNA Learns from Antisense. *Nat. Chem. Biol.* **2007**, *3* (1), 8–11.

(35) Wheeler, T. M.; Leger, A. J.; Pandey, S. K.; MacLeod, A. R.; Nakamori, M.; Cheng, S. H.; Wentworth, B. M.; Bennett, C. F.; Thornton, C. A. Targeting Nuclear RNA for in Vivo Correction of Myotonic Dystrophy. *Nature* **2012**, *488* (7409), 111–115.

(36) Marwick, C. First "Antisense" Drug Will Treat CMV Retinitis. *JAMA* **1998**, *280* (10), 871.

(37) Juliano, R. L.; Ming, X.; Nakagawa, O. The Chemistry and Biology of Oligonucleotide Conjugates. *Acc. Chem. Res.* **2012**, *45* (7), 1067–1076.

(38) Winkler, J. Oligonucleotide Conjugates for Therapeutic Applications. *Ther. Deliv.* **2013**, *4* (7), 791–809.

(39) Yu, D.; Kandimalla, E. R.; Roskey, A.; Zhao, Q.; Chen, L.; Chen, J.; Agrawal, S. Stereo-Enriched Phosphorothioate Oligodeoxynucleotides: Synthesis, Biophysical and Biological Properties. *Bioorg. Med. Chem.* **2000**, *8* (1), 275–284.

(40) Bachelin, M.; Hessler, G.; Kurz, G.; Hacia, J. G.; Dervan, P. B.; Kessler, H. Structure of a Stereoregular Phosphorothioate DNA/RNA Duplex. *Nat. Struct. Biol.* **1998**, *5* (4), 271–276.

(41) Oka, N.; Wada, T. Stereocontrolled Synthesis of Oligonucleotide Analogs Containing Chiral Internucleotidic Phosphorus Atoms. *Chem. Soc. Rev.* **2011**, *40* (12), 5829–5843.

(42) Wilk, A.; Grajkowski, A.; Phillips, L. R.; Beaucage, S. L. Deoxyribonucleoside Cyclic N -Acylphosphoramidites as a New Class of Monomers for the Stereocontrolled Synthesis of Oligothymidylyl- and Oligodeoxycytidylyl-Phosphorothioates. *J. Am. Chem. Soc.* **2000**, *122* (10), 2149–2156.

(43) Oka, N.; Wada, T.; Saigo, K. An Oxazaphospholidine Approach for the Stereocontrolled Synthesis of Oligonucleoside Phosphorothioates. *J. Am. Chem. Soc.* **2003**, *125* (27), 8307–8317.

(44) Nawrot, B.; Rębowska, B.; Cieślińska, K.; Stec, W. J. New Approach to the Synthesis of Oligodeoxyribonucleotides Modified with Phosphorothioates of Predetermined Sense of P-Chirality. *Tetrahedron Lett.* **2005**, *46* (39), 6641–6644.

(45) Miller, P. S.; Fang, K. N.; Kondo, N. S.; Ts'o, P. O. P. Conformation and Interaction of Dinucleoside Mono- and Diphosphates. V. Syntheses and Properties of Adenine and Thymine Nucleoside Alkyl Phosphotriesters, the Neutral Analogs of Dinucleoside Monophosphates. *J. Am. Chem. Soc.* **1971**, *93* (24), 6657–6665.

(46) Zhao, Q.; Matson, S.; Herrera, C. J.; Fisher, E.; Yu, H.; Krieg, a M. Comparison of Cellular Binding and Uptake of Antisense Phosphodiester, Phosphorothioate, and Mixed Phosphorothioate and Methylphosphonate Oligonucleotides. *Antisense Res. Dev.* **1993**, *3* (1), 53–66.

(47) Lesnikowski, Z. J.; Jaworska, M.; Stec, W. J. Octa(thymidine Methanephosphonates) of Partially Defined Stereochemistry: Synthesis and Effect of Chirality at Phosphorus on Binding to Pentadecadeoxyriboadenylic Acid. *Nucleic Acids Res.* **1990**, *18* (8), 2109–2115.

(48) Gryaznov, S.; Chen, J.-K. Oligodeoxyribonucleotide N3'→P5' Phosphoramidates: Synthesis and Hybridization Properties. *J. Am. Chem. Soc.* **1994**, *116* (7), 3143–3144.

(49) Faria, M.; Spiller, D. G.; Dubertret, C.; Nelson, J. S.; White, M. R. H.; Scherman, D.; Helene, C.; Giovannangeli, C. Phosphoramidate Oligonucleotides as Potent Antisense Molecules in Cells and in Vivo. **2001**, *19* (1), 40–44.

(50) Pongracz, K.; Gryaznov, S. Oligonucleotide N3'→P5' Thiophosphoramidates: Synthesis and Properties. *Tetrahedron Lett.* **1999**, *40* (43), 7661–7664.

(51) Schultz, R. G.; Gryaznov, S. M. Oligo-2'-Fluoro-2'-Deoxynucleotide N3'-P5' Phosphoramidates: Synthesis and Properties. *Nucleic Acids Res.* **1996**, *24* (15), 2966–2973.

(52) Shaw, B. R.; Sergueev, D.; He, K.; Porter, K.; Summers, J.; Sergueeva, Z.; Rait, V. *Boranophosphate Backbone: A Mimic of Phosphodiesters, Phosphorothioates, and Methylphosphonates*; Methods in Enzymology; Elsevier, 2000; Vol. 313.

(53) Rait, V. K.; Shaw, B. R. Boranophosphates Support the RNase H Cleavage of Polyribonucleotides. *Antisense Nucleic Acid Drug Dev.* **1999**, *9* (1), 53–60.

(54) Hall, A. H. S.; Wan, J.; Shaughnessy, E. E.; Ramsay Shaw, B.; Alexander, K. A. RNA Interference Using Boranophosphate siRNAs: Structure-Activity Relationships. *Nucleic Acids Res.* **2004**, *32* (20), 5991–6000.

(55) Iwamoto, N.; Oka, N.; Wada, T. Stereocontrolled Synthesis of Oligodeoxyribonucleoside Boranophosphates via Stereodefined H-Phosphonate Intermediates. *Nucleic Acids Symp. Ser. (Oxf).* **2009**, No. 53, 9–10.

(56) Hall, A. H. S.; Wan, J.; Spesock, A.; Sergueeva, Z.; Shaw, B. R.; Alexander, K. A. High Potency Silencing by Single-Stranded Boranophosphate siRNA. *Nucleic Acids Res.* **2006**, *34* (9), 2773–2781.

(57) Sazani, P.; Gemignani, F.; Kang, S. H.; Maier, M. A.; Manoharan, M.; Persmark, M.; Bortner, D.; Kole, R. Systemically Delivered Antisense Oligomers Upregulate Gene Expression in Mouse Tissues. *Nat. Biotechnol.* **2002**, *20* (12), 1228–1233.

(58) De Mesmaeker, A.; Waldner, A.; Lebreton, J.; Hoffmann, P.; Fritsch, V.; Wolf, R. M.; Freier, S. M. Amides as a New Type of Backbone Modification in Oligonucleotides. *Angew. Chemie Int. Ed. English* **1994**, *33* (2), 226–229.

(59) Selvam, C.; Thomas, S.; Abbott, J.; Kennedy, S. D.; Rozners, E. Amides as Excellent Mimics of Phosphate Linkages in RNA. *Angew. Chemie-International Ed.* **2011**, *50* (9), 2068–2070.

(60) Lebreton, J.; Waldner, A.; Fritsch, V.; Wolf, R. R.; Demesmaeker, A. Comparison of 2 Amides as Backbone Replacement of the Phosphodiester Linkage in Oligonucleotides. *Tetrahedron Lett.* **1994**, *35* (29), 5225–5228.

(61) Iwase, R.; Toyama, T.; Nishimori, K. Solid-Phase Synthesis of Modified RNAs Containing Amide-Linked Oligoribonucleosides at Their 3'-End and Their Application to siRNA. *Nucleosides Nucleotides & Nucleic Acids* **2007**, *26* (10-12), 1451–1454.

(62) Mutisya, D.; Selvam, C.; Lunstad, B. D.; Pallan, P. S.; Haas, A.; Leake, D.; Egli, M.; Rozners, E. Amides Are Excellent Mimics of Phosphate Internucleoside Linkages and Are Well Tolerated in Short Interfering RNAs. *Nucleic Acids Res.* **2014**, *42* (10), 6542–6551.

(63) Gong, W.; Desaulniers, J. P. Gene-Silencing Properties of siRNAs That Contain Internal Amide-Bond Linkages. *Bioorg. Med. Chem. Lett.* **2012**, *22* (22), 6934–6937.

(64) Egholm, M.; Buchardt, O.; Nielsen, P. E.; Berg, R. H. Peptide Nucleic Acids (PNA). Oligonucleotide Analogs with an Achiral Peptide Backbone. *J. Am. Chem. Soc.* **1992**, *114* (5), 1895–1897.

(65) Christensen, L.; Fitzpatrick, R.; Gildea, B.; Petersen, K. H.; Henrik, F.; Koch, T.; Egholm, M.; Buchardt, E.; Nielsen, P. E.; Coull, J.; Berg, R. H. Solid-Phase Synthesis of Peptide Nucleic Acid. **1995**, *3*, 175–183.

(66) Shakeel, S.; Karim, S.; Ali, A. Peptide Nucleic Acid (PNA) — a Review. *J. Chem. Technol. Biotechnol.* **2006**, *81* (6), 892–899.

(67) Egholm, M.; Buchardt, O.; Christensen, L.; Behrens, C.; Freier, S. M.; Driver, D. A.; Berg, R. H.; Kim, S. K.; Norden, B.; Nielsen, P. E. PNA Hybridizes to Complementary Oligonucleotides Obeying the Watson-Crick Hydrogen-Bonding Rules. *Nature* **1993**, *365* (6446), 566–568.

(68) Nielsen, P. E. *Peptide Nucleic Acids: Protocols and Applications*; Garland Science, **2004**.

(69) Hanvey, J. C.; Peffer, N. J.; Bisi, J. E.; Thomson, S. A.; Cadilla, R.; Josey, J. A.; Ricca, D. J.; Hassman, C. F.; Bonham, M. A.; Au, K. G.; Carter, S. G.; Bruckenstein, D. A.; Boyd, A. L.; Noble, S. A.; Babiss, L. E. Antisense and Antigene Properties of Peptide Nucleic Acids. **1992**, *258* (November), 1481–1485.

(70) Potenza, N.; Moggio, L.; Milano, G.; Salvatore, V. RNA Interference in Mammalia Cells by RNA-3 ' -PNA Chimeras. *Int. J. Mol. Sci.* **2008**, *9*, 299–315.

(71) Nielsen, P. E. Peptide Nucleic Acids (PNA) in Chemical Biology and Drug Discovery. *Chem. Biodivers.* **2010**, *7* (4), 786–804.

(72) Sazani, P.; Kang, S. H.; Maier, M. A.; Wei, C. F.; Dillman, J.; Summerton, J.; Manoharan, M.; Kole, R. Nuclear Antisense Effects of Neutral, Anionic and Cationic Oligonucleotide Analogs. *Nucleic Acids Res.* **2001**, *29* (19), 3965–3974.

(73) Knudsen, H.; Nielsen, P. E. Antisense Properties of Duplex- and Triplex-Forming PNAs. *Nucleic Acids Res.* **1996**, *24* (3), 494–500.

(74) Nielsen, P.; Egholm, M.; Berg, R.; Buchardt, O. Sequence-Selective Recognition of DNA by Strand Displacement with a Thymine-Substituted Polyamide. *Science (80).* **1991**, *254* (5037), 1497–1500.

(75) Nielsen, P. E.; Egholm, M.; Buchardt, O. Sequence-Specific Transcription Arrest by Peptide Nucleic Acid Bound to the DNA Template Strand. *Gene* **1994**, *149* (1), 139–145.

(76) Smulevitch, S. V; Simmons, C. G.; Norton, J. C.; Wise, T. W.; Corey, D. R. Enhancement of Strand Invasion by Oligonucleotides through Manipulation of Backbone Charge. *Nat. Biotechnol.* **1996**, *14* (13), 1700–1704.

(77) Mayer, G. *The Chemical Biology of Nucleic Acids*; John Wiley & Sons, 2011.

(78) Hamzavi, R.; Dolle, F.; Tavitian, B.; Dahl, O.; Nielsen, P. E. Modulation of the Pharmacokinetic Properties of PNA: Preparation of Galactosyl, Mannosyl, Fucosyl, N-Acetylgalactosaminyl, and N-Acetylglucosaminyl Derivatives of Aminoethylglycine Peptide Nucleic Acid Monomers and Their Incorporation into PNA Oligomers. *Bioconjug. Chem.* **2003**, *14* (5), 941–954.

(79) Simmons, C. G.; Pitts, A. E.; Mayfield, L. D.; Shay, J. W.; Corey, D. R. Synthesis and Membrane Permeability of PNA-Peptide Conjugates. *Bioorg. Med. Chem. Lett.* **1997**, *7* (23), 3001–3006.

(80) Shiraishi, T.; Nielsen, P. E. Nanomolar Cellular Antisense Activity of Peptide Nucleic Acid (PNA) Cholic Acid ("umbrella") and Cholesterol Conjugates Delivered by Cationic Lipids. *Bioconjug. Chem.* **2012**, *23* (2), 196–202.

(81) Joshi, R.; Mishra, R.; Su, W.; Engelmann, J. CPP or Cholesterol Conjugated Antisense PNA for Cellular Delivery. *J. Pept. Sci.* **2008**, *14* (S1), 47–194.

(82) Eren, M.; Swenson, R. Chemical Synthesis and Expression of a Synthetic Gene for the Flavodoxin from Clostridium MP. *J. Biol. Chem.* **1989**, *264* (25), 14874–14879.

(83) Xiong, A.-S.; Peng, R.-H.; Zhuang, J.; Liu, J.-G.; Gao, F.; Chen, J.-M.; Cheng, Z.-M.; Yao, Q.-H. Non-Polymerase-Cycling-Assembly-Based Chemical Gene Synthesis: Strategies, Methods, and Progress. *Biotechnol. Adv.* **2008**, *26* (2), 121–134.

(84) Au, L. C.; Yang, F. Y.; Yang, W. J.; Lo, S. H.; Kao, C. F. Gene Synthesis by a LCR-Based Approach: High-Level Production of Leptin-L54 Using Synthetic Gene in Escherichia Coli. *Biochem. Biophys. Res. Commun.* **1998**, *248* (1), 200–203.

(85) Pachuk, C. J.; Samuel, M.; Zurawski, J. A.; Snyder, L.; Phillips, P.; Satishchandran, C. Chain Reaction Cloning: A One-Step Method for Directional Ligation of Multiple DNA Fragments. *Gene* **2000**, *243* (1-2), 19–25.

(86) Kok, S. De; Stanton, L. H.; Slaby, T.; Durot, M.; Holmes, V. F.; Patel, K. G.; Platt, D.; Shapland, E. B.; Serber, Z.; Dean, J.; Newman, J. D.; Chandran, S. S. Rapid and Reliable DNA Assembly via Ligase Cycling Reaction. *ACS Synth. Biol.* **2014**, *3* (2), 97–106.

(87) Hughes, R. A.; Miklos, A. E.; Ellington, A. D. Gene Synthesis: Methods and Applications. *Methods Enzymol.* **2011**, *498*, 277–309.

(88) Ma, S.; Saaem, I.; Tian, J. Error Correction in Gene Synthesis Technology. *Trends Biotechnol.* **2012**, *30* (3), 147–154.

(89) Chao, R.; Yuan, Y.; Zhao, H. Recent Advances in DNA Assembly Technologies. *FEMS Yeast Res.* **2014**, *15* (1), 1–9.

(90) Kodumal, S. J.; Patel, K. G.; Reid, R.; Menzella, H. G.; Welch, M.; Santi, D. V. Total Synthesis of Long DNA Sequences: Synthesis of a Contiguous 32-Kb Polyketide Synthase Gene Cluster. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (44), 15573–15578.

(91) Gibson, D. G.; Smith, H. O.; Hutchison, C. A.; Venter, J. C.; Merryman, C. Chemical Synthesis of the Mouse Mitochondrial Genome. *Nat. Methods* **2010**, *7* (11), 901–903.

(92) Cello, J.; Paul, A. V; Wimmer, E. Chemical Synthesis of Poliovirus cDNA: Generation of Infectious Virus in the Absence of Natural Template. *Science* **2002**, *297* (5583), 1016–1018.

(93) Booth, M. J.; Raiber, E.-A.; Balasubramanian, S. Chemical Methods for Decoding Cytosine Modifications in DNA. *Chem. Rev.* **2015**, *115* (6), 2240–2254.

(94) Moore, L. D.; Le, T.; Fan, G. DNA Methylation and Its Basic Function. *Neuropsychopharmacology* **2012**, *38* (1), 23–38.

(95) Hwang, G. T.; Romesberg, F. E. Unnatural Substrate Repertoire of A, B, and X Family DNA Polymerases. *J. Am. Chem. Soc.* **2008**, *130* (44), 14872–14882.

(96) Malyshev, D. A.; Dhami, K.; Lavergne, T.; Chen, T.; Dai, N.; Foster, J. M.; Corrêa, I. R.; Romesberg, F. E. A Semi-Synthetic Organism with an Expanded Genetic Alphabet. *Nature* **2014**, *509* (7500), 385–388.

(97) König, H.; Frank, D.; Heil, R.; Coenen, C. Synthetic Genomics and Synthetic Biology Applications between Hopes and Concerns. *Curr. Genomics* **2013**, *14* (1), 11–24.

(98) Shabarova, Z. A.; Dolinnaya, N. G.; Drusta, V. L.; Melnikova, N. P.; Purmal, A. A. DNA-like Duplexes with Repetitions. III, Efficient Template-Guided Chemical Polymerization of d(TGGCCAAGCTp). *Nucleic Acids Res.* **1981**, *9* (23), 6407–6419.

(99) Harada, K.; Orgel, L. In Vitro Selection of Optimal DNA Substrates for Ligation by a Water-Soluble Carbodiimide. *J. Mol. Evol.* **1994**, *38* (6).

(100) Ho, N. W. Y.; Gilham, P. T. The Reversible Chemical Modification of Uracil, Thymine, and Guanine Nucleotides and the Modification of the Action of Ribonuclease on Ribonucleic Acid. **1966**.

(101) Sokolova, N. I.; Ashirbekova, D. T.; Dolinnaya, N. G.; Shabarova, Z. A. Chemical Reactions within DNA Duplexes Cyanogen Bromide as an Effective Oligodeoxyribonucleotide Coupling Agent. *FEBS Lett.* **1988**, *232* (1), 153–155.

(102) Luebke, K. J.; Dervan, P. B. Nonenzymic Sequence-Specific Ligation of Double-Helical DNA. *J. Am. Chem. Soc.* **1991**, *113* (19), 7447–7448.

(103) Dolinnaya, N. G.; Sokolova, N. I.; Gryaznova, O. I.; Shabarova, Z. A. Site-Directed Modification of DNA Duplexes by Chemical Ligation. *Nucleic Acids Res.* **1988**, *16* (9), 3721–3738.

(104) Wolfe, J. L. Synthesis and Polymerase Incorporation of 5’-Amino-2',5'-Dideoxy-5'-N-Triphosphate Nucleotides. *Nucleic Acids Res.* **2002**, *30* (17), 3739–3747.

(105) Gryaznov, S. M.; Letsinger, R. L. Template Controlled Coupling and Recombination of Oligonucleotide Blocks Containing Thiophosphoryl Groups. *Nucleic Acids Res.* **1993**, *21* (6), 1403–1408.

(106) Herrlein, M. K.; Letsinger, R. L. Selective Chemical Autoligation on a Double-Stranded DNA Template. *Nucleic Acids Res.* **1994**, *22* (23), 5076–5078.

(107) Herrlein, M. K.; Nelson, J. S.; Letsinger, R. L. A Covalent Lock for Self-Assembled Oligonucleotide Conjugates. *J. Am. Chem. Soc.* **1995**, *117* (40), 10151–10152.

(108) Xu, Y. Chemical and Enzymatic Properties of Bridging 5'-S-Phosphorothioester Linkages in DNA. *Nucleic Acids Res.* **1998**, *26* (13), 3159–3164.

(109) Demesmaeker, A.; Waldner, A.; Sanghvi, Y. S.; Lebreton, J. Comparison of Rigid and Flexible Backbones in Antisense Oligonucleotides. *Bioorg. Med. Chem. Lett.* **1994**, *4* (3), 395–398.

(110) Luo, P. Z.; Leitzel, J. C.; Zhan, Z. Y. J.; Lynn, D. G. Analysis of the Structure and Stability of a Backbone-Modified Oligonucleotide: Implications for Avoiding Product Inhibition in Catalytic Template-Directed Synthesis. *J. Am. Chem. Soc.* **1998**, *120* (13), 3019–3031.

(111) Hamilton, P. L.; Arya, D. P. Natural Product DNA Major Groove Binders. *Nat. Prod. Rep.* **2012**, *29* (2), 134–143.

(112) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. The Molecular Origin of DNA-Drug Specificity in Netropsin and Distamycin. *Proc. Natl. Acad. Sci. U. S. A.* **1985**, *82* (5), 1376–1380.

(113) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. Binding of an Antitumor Drug to DNA. *J. Mol. Biol.* **1985**, *183* (4), 553–563.

(114) Ghosh, S.; Defrancq, E.; Lhomme, J. H.; Dumy, P.; Bhattacharya, S. Efficient Conjugation and Characterization of Distamycin-Based Peptides with Selected Oligonucleotide Stretches. *Bioconjug. Chem.* **2004**, *15* (3), 520–529.

(115) Wiederholt, K.; Rajur, S. B.; Giuliano, J.; O'Donnel, M. J.; McLaughlin, L. W. DNA-Tethered Hoechst Groove-Binding Agents: Duplex Stabilization and Fluorescence Characteristics. *J. Am. Chem. Soc.* **1996**, *118* (30), 7055–7062.

(116) Reddy, P. M.; Bruice, T. C. Solid-Phase Synthesis of Positively Charged Deoxynucleic Guanidine (DNG) Tethering a Hoechst 33258 Analogue: Triplex and Duplex Stabilization by Simultaneous Minor Groove Binding. *J. Am. Chem. Soc.* **2004**, *126* (12), 3736–3747.

(117) Jain, M. L.; Bruice, P. Y.; Szabó, I. E.; Bruice, T. C. Incorporation of Positively Charged Linkages into DNA and RNA Backbones: A Novel Strategy for Antigene and Antisense Agents. *Chem. Rev.* **2012**, *112* (3), 1284–1309.

(118) Boer, D. R.; Canals, A.; Coll, M. DNA-Binding Drugs Caught in Action: The Latest 3D Pictures of Drug-DNA Complexes. *Dalton Trans.* **2009**, No. 3, 399–414.

(119) Lerman, L. S. Structural Considerations in the Interaction of DNA and Acridines. *J. Mol. Biol.* **1961**, *3* (1), 18–IN14.

(120) Hayashi, M.; Harada, Y. Direct Observation of the Reversible Unwinding of a Single DNA Molecule Caused by the Intercalation of Ethidium Bromide. *Nucleic Acids Res.* **2007**, *35* (19), e125.

(121) Williams, L.; Egli, M. DNA Intercalation: Helix Unwinding and Neighbor-Exclusion. *Struct. Funct.* **1992**.

(122) Neto, B. A. D.; Lapis, A. A. M. Recent Developments in the Chemistry of Deoxyribonucleic Acid (DNA) Intercalators: Principles, Design, Synthesis, Applications and Trends. *Molecules* **2009**, *14* (5), 1725–1746.

(123) Toulmé, J. J.; Krisch, H. M.; Loreau, N.; Thuong, N. T.; Hélène, C. Specific Inhibition of mRNA Translation by Complementary Oligonucleotides Covalently Linked to Intercalating Agents. *Proc. Natl. Acad. Sci. U. S. A.* **1986**, *83* (5), 1227–1231.

(124) Prakash, T. P.; Püschl, A.; Manoharan, M. N,N'-Bis-(2-(cyano)ethoxycarbonyl)-2-Methyl-2-Thiopseudourea: A Guanylating Reagent for Synthesis of 2'-O-[2-(Guanidinium)ethyl]-Modified Oligonucleotides. *Nucleosides, Nucleotides and Nucleic Acids* **2007**, *26* (2), 149–159.

(125) Maier, M. A.; Barber-Peoc'h, I.; Manoharan, M. Postsynthetic Guanidinylation of Primary Amino Groups in the Minor and Major Grooves of Oligonucleotides. *Tetrahedron Lett.* **2002**, *43* (42), 7613–7616.

(126) Lam, C. H.; Perrin, D. M. Introduction of Guanidinium-Modified Deoxyuridine into the Substrate Binding Regions of DNAzyme 10–23 to Enhance Target Affinity: Implications for DNAzyme Design. *Bioorg. Med. Chem. Lett.* **2010**, *20* (17), 5119–5122.

(127) Linkletter, B. A.; Szabo, I. E.; Bruice, T. C. Solid-Phase Synthesis of Deoxynucleic Guanidine (DNG) Oligomers and Melting Point and Circular Dichroism Analysis of Binding Fidelity of Octameric Thymidyl Oligomers with DNA Oligomers. *J. Am. Chem. Soc.* **1999**, *121* (16), 3888–3896.

(128) Reddy, P. M.; Bruice, T. C. Solid-Phase Synthesis of Positively Charged Deoxynucleic Guanidine (DNG) Oligonucleotide Mixed Sequences. *Bioorg. Med. Chem. Lett.* **2003**, *13* (7), 1281–1285.

(129) Lindow, M.; Vornlocher, H.-P.; Riley, D.; Kornbrust, D. J.; Burchard, J.; Whiteley, L. O.; Kamens, J.; Thompson, J. D.; Nochur, S.; Younis, H.; Bartz, S.; Parry, J.; Ferrari, N.; Henry, S. P.; Levin, A. A. Assessing Unintended Hybridization-Induced Biological Effects of Oligonucleotides. *Nat. Biotechnol.* **2012**, *30* (10), 920–923.

(130) Barawkar, D. A.; Bruice, T. C. Synthesis, Biophysical Properties, and Nuclease Resistance Properties of Mixed Backbone Oligodeoxynucleotides Containing Cationic Internucleoside Guanidinium Linkages: Deoxynucleic guanidine/DNA Chimeras. *Proc. Natl. Acad. Sci.* **1998**, *95* (19), 11047–11052.

(131) Liu, J.; Shi, J.-Z.; Yu, L.-M.; Goyer, R. A.; Waalkes, M. P. Mercury in Traditional Medicines: Is Cinnabar Toxicologically Similar to Common Mercurials? *Exp. Biol. Med. (Maywood).* **2008**, *233* (7), 810–817.

(132) Michel, T.; Debart, F.; Vasseur, J.-J. Efficient Guanidination of the Phosphate Linkage towards Cationic Phosphoramidate Oligonucleotides. *Tetrahedron Lett.* **2003**, *44* (35), 6579–6582.

(133) Deglane, G.; Abes, S.; Michel, T.; Prévot, P.; Vives, E.; Debart, F.; Barvik, I.; Lebleu, B.; Vasseur, J.-J. Impact of the Guanidinium Group on Hybridization and Cellular Uptake of Cationic Oligonucleotides. *Chembiochem* **2006**, *7* (4), 684–692.

(134) Fiandor, J.; Tam, S. Y. Synthesis of 3’-Deoxy-3'-(2-Propynyl)thymidine and 3'-Cyanomethyl-3'-Deoxythymidine, Analogs of AZT . *Tetrahedron Lett.* **1990**, *31* (5), 597–600.

(135) Cui, H.; Carrero-Lérida, J.; Silva, A. P. G.; Whittingham, J. L.; Brannigan, J. A.; Ruiz-Pérez, L. M.; Read, K. D.; Wilson, K. S.; González-Pacanowska, D.; Gilbert, I. H. Synthesis and Evaluation of α-Thymidine Analogues as Novel Antimalarials. *J. Med. Chem.* **2012**, *55* (24), 10948–10957.

(136) Lange, M.; Pettersen, A. L.; Undheim, K. Synthesis of Secondary Amines by Reductive Dimerization of Azides. *Tetrahedron* **1998**, *54* (21), 5745–5752.

(137) Borch, R. F.; Bernstein, M. D.; Durst, H. D. Cyanohydridoborate Anion as a Selective Reducing Agent. *J. Am. Chem. Soc.* **1971**, *93* (12), 2897–2904.

(138) Abdel-Magid, A. F.; Carson, K. G.; Harris, B. D.; Maryanoff, C. A.; Shah, R. D. Reductive Amination of Aldehydes and Ketones with Sodium Triacetoxyborohydride. Studies on Direct and Indirect Reductive Amination Procedures1. *J. Org. Chem.* **1996**, *61* (11), 3849–3862.

(139) Davies, M. J.; Shah, A.; Bruce, I. J. Synthesis of Fluorescently Labelled Oligonucleotides and Nucleic Acids. *Chem. Soc. Rev.* **2000**, *29* (2), 97–107.

(140) Rescifina, A.; Zagni, C.; Varrica, M. G.; Pistarà, V.; Corsaro, A. Recent Advances in Small Organic Molecules as DNA Intercalating Agents: Synthesis, Activity, and Modeling. *Eur. J. Med. Chem.* **2014**, *74*, 95–115.

(141) Katritzky, A. R.; Rogovoy, B. V. Recent Developments in Guanylating Agents. *Arkivoc* **2005**, 49–87.

(142) Carpino, L. A.; Imazumi, H.; Foxman, B. M.; Vela, M. J.; Henklein, P.; El-Faham, A.; Klose, J.; Bienert, M. Comparison of the Effects of 5- and 6-HOAt on Model Peptide Coupling Reactions Relative to the Cases for the 4- and 7-Isomers † , ‡. *Org. Lett.* **2000**, *2* (15), 2253–2256.

(143) Kuwahara, M.; Takeshima, H.; Nagashima, J.; Minezaki, S.; Ozaki, H.; Sawai, H. Transcription and Reverse Transcription of Artificial Nucleic Acids Involving Backbone Modification by Template-Directed DNA Polymerase Reactions. *Bioorg. Med. Chem.* **2009**, *17* (11), 3782–3788.

(144) Kachalova, A. V.; Stetsenko, D. A.; Gait, M. J.; Oretskaya, T. S. Synthesis of Oligonucleotide 2'-Conjugates via Amide Bond Formation in Solution. *Bioorg. Med. Chem. Lett.* **2004**, *14* (3), 801–804.

(145) Kolb, H. C.; Finn, M. G.; Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angewandte Chemie - International Edition*. 2001, pp 2004–2021.

(146) Rostovtsev, V. V; Green, L.; Sharpless, K. B. Copper-Catalyzed Cycloaddition of Azides and Acetylenes. *Abstr. Pap. Am. Chem. Soc.* **2002**, *224*, U186–U186.

(147) El-Sagheer, A. H.; Brown, T. New Strategy for the Synthesis of Chemically Modified RNA Constructs Exemplified by Hairpin and Hammerhead Ribozymes. *Proc. Natl. Acad. Sci.* **2010**, *107* (35), 15329–15334.

(148) Kumar, R.; El-Sagheer, A.; Tumpane, J.; Lincoln, P.; Wilhelmsson, L. M.; Brown, T. Template-Directed Oligonucleotide Strand Ligation, Covalent Intramolecular DNA Circularization and Catenation Using Click Chemistry. *J. Am. Chem. Soc.* **2007**, *129* (21), 6859–6864.

(149) Lundberg, E. P.; El-Sagheer, A. H.; Kocalka, P.; Wilhelmsson, L. M.; Brown, T.; Nordén, B. A New Fixation Strategy for Addressable Nano-Network Building Blocks. *Chem. Commun. (Camb).* **2010**, *46* (21), 3714–3716.

(150) El-Sagheer, A. H.; Brown, T. Synthesis and Polymerase Chain Reaction Amplification of DNA Strands Containing an Unnatural Triazole Linkage. *J. Am. Chem. Soc.* **2009**, *131* (11), 3958–3964.

(151) El-Sagheer, A. H.; Brown, T. Click Nucleic Acid Ligation: Applications in Biology and Nanotechnology. *Acc. Chem. Res.* **2012**, *45* (8), 1258–1267.

(152) El-Sagheer, A. H.; Brown, T. New Strategy for the Synthesis of Chemically Modified RNA Constructs Exemplified by Hairpin and Hammerhead Ribozymes. *Proc. Natl. Acad. Sci.* **2010**, *107* (35), 15329–15334.

(153) El-Sagheer, A. H.; Sanzone, A. P.; Gao, R.; Tavassoli, A.; Brown, T. Biocompatible Artificial DNA Linker That Is Read through by DNA Polymerases and Is Functional in Escherichia Coli. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (28), 11338–11343.

(154) Birts, C. N.; Sanzone, A. P.; El-Sagheer, A. H.; Blaydes, J. P.; Brown, T.; Tavassoli, A. Transcription of Click-Linked DNA in Human Cells. *Angew. Chemie Int. Ed.* **2014**, *53* (9), 2362–2365.

(155) Dallmann, A.; El-Sagheer, A. H.; Dehmel, L.; Mugge, C.; Griesinger, C.; Ernsting, N. P.; Brown, T. Structure and Dynamics of Triazole-Linked DNA: Biocompatibility Explained. *Chem. Eur. J.* **2011**, *17* (52), 14714–14717.

(156) Lin, K.-Y.; Matteucci, M. D. A Cytosine Analogue Capable of Clamp-Like Binding to a Guanine in Helical Nucleic Acids. *J. Am. Chem. Soc.* **1998**, *120* (33), 8531–8532.

(157) El-Sagheer, A. H.; Brown, T. Combined Nucleobase and Backbone Modifications Enhance DNA Duplex Stability and Preserve Biocompatibility. *Chem. Sci.* **2014**, *5* (1), 253–259.

(158) Isobe, H.; Fujino, T.; Yamazaki, N.; Guillot-Nieckowski, M.; Nakamura, E. Triazole-Linked Analogue of Deoxyribonucleic Acid ((TL)DNA): Design, Synthesis, and Double-Strand Formation with Natural DNA. *Org. Lett.* **2008**, *10* (17), 3729–3732.

(159) Fujino, T.; Yamazaki, N.; Isobe, H. Convergent Synthesis of Oligomers of Triazole-Linked DNA Analogue ((TL)DNA) in Solution Phase. *Tetrahedron Lett.* **2009**, *50* (28), 4101–4103.

(160) Crooks, P. A.; Reynolds, R. C.; Maddry, J. A.; Rathore, A.; Akhtar, M. S.; Montgomery, J. A.; Secrist, J. A. Nucleoside Sultones: Synthons for the Preparation of Novel Nucleotide Analogs. 1. Synthesis and Ring-Opening Reactions. *J. Org. Chem.* **1992**, *57* (10), 2830–2835.

(161) Lewkowicz, E. S.; Iribarren, A. M. Nucleoside Phosphorylases. *Curr. Org. Chem.* **2006**, *10* (11), 1197–1215.

(162) Becouarn, S.; Czernecki, S.; Valéry, J.-M. Efficient Transformation of Thymidine into 2',3'-Didehydro-2',3'-Dideoxy-Thymidine (D4T) Involving Opening of a 2,3'-Anhydro Derivative by Phenylselenol. *Nucleosides and Nucleotides* **1995**, *14* (6), 1227–1232.

(163) Gogoi, K.; Gunjal, A. D.; Phalgune, U. D.; Kumar, V. A. Synthesis and RNA Binding Selectivity of Oligonucleotides Modified with Five-Atom Thioacetamido Nucleic Acid Backbone Structures. *Org. Lett.* **2007**, *9* (14), 2697–2700.

(164) CZERNECKI, S.; VALERY, J. An Efficient Synthesis of 3'-Azido-3'-Deoxythymidine (AZT). *SYNTHESIS-STUTTGART* **1991**, No. 3, 239–240.

(165) Chan, T. R.; Hilgraf, R.; Sharpless, K. B.; Fokin, V. V. Polytriazoles as copper(I)-Stabilizing Ligands in Catalysis. *Org. Lett.* **2004**, *6* (17), 2853–2855.

(166) Dias, N.; Stein, C. A. Antisense Oligonucleotides: Basic Concepts and Mechanisms. *Mol. Cancer Ther.* **2002**, *1* (5), 347–355.

(167) El-Sagheer, A. H.; Kumar, R.; Findlow, S.; Werner, J. M.; Lane, A. N.; Brown, T. A Very Stable Cyclic DNA Miniduplex with Just Two Base Pairs. *Chembiochem* **2008**, *9* (1), 50–52.

(168) Boyd, S. D. Diagnostic Applications of High-Throughput DNA Sequencing. *Annu. Rev. Pathol.* **2013**, *8*, 381–410.

(169) Li, Y.; Hui, H.; Burgess, C. J.; Price, R. W.; Sharp, P. M.; Hahn, B. H.; Shaw, G. M. Complete Nucleotide Sequence, Genome Organization, and Biological Properties of Human Immunodeficiency Virus Type 1 in Vivo: Evidence for Limited Defectiveness and Complementation. *J. Virol.* **1992**, *66* (11), 6587–6600.

(170) Radford, A. D.; Chapman, D.; Dixon, L.; Chantrey, J.; Darby, A. C.; Hall, N. Application of next-Generation Sequencing Technologies in Virology. *J. Gen. Virol.* **2012**, *93* (9), 1853–1868.

(171) Frampton, G. M.; Fichtenholtz, A.; Otto, G. A.; Wang, K.; Downing, S. R.; He, J.; Schnall-Levin, M.; White, J.; Sanford, E. M.; An, P.; Sun, J.; Juhn, F.; Brennan, K.; Iwanik, K.; Maillet, A.; Buell, J.; White, E.; Zhao, M.; Balasubramanian, S.; Terzic, S.; Richards, T.; Banning, V.; Garcia, L.; Mahoney, K.; Zwirko, Z.; Donahue, A.; Beltran, H.; Mosquera, J. M.; Rubin, M. A.; Dogan, S.; Hedvat, C. V; Berger, M. F.; Pusztai, L.; Lechner, M.; Boshoff, C.; Jarosz, M.; Vietz, C.; Parker, A.; Miller, V. A.; Ross, J. S.; Curran, J.; Cronin, M. T.; Stephens, P. J.; Lipson, D.; Yelensky, R. Development and Validation of a Clinical Cancer Genomic Profiling Test Based on Massively Parallel DNA Sequencing. *Nat. Biotechnol.* **2013**, *31* (11), 1023–1031.

(172) Roewer, L. DNA Fingerprinting in Forensics: Past, Present, Future. *Investig. Genet.* **2013**, *4* (1), 22.

(173) Yang, Y.; Xie, B.; Yan, J. Application of next-Generation Sequencing Technology in Forensic Science. *Genomics. Proteomics Bioinformatics* **2014**, *12* (5), 190–197.

(174) Maxam, A. M.; Gilbert, W. A New Method for Sequencing DNA. *Proc. Natl. Acad. Sci.* **1977**, *74* (2), 560–564.

(175) Sanger, F.; Nicklen, S.; Coulson, A. R. DNA Sequencing with Chain-Terminating Inhibitors. *Proc. Natl. Acad. Sci.* **1977**, *74* (12), 5463–5467.

(176) Prober, J.; Trainor, G.; Dam, R.; Hobbs, F.; Robertson, C.; Zagursky, R.; Cocuzza, A.; Jensen, M.; Baumeister, K. A System for Rapid DNA Sequencing with Fluorescent Chain-Terminating Dideoxynucleotides. *Science (80).* **1987**, *238* (4825), 336–341.

(177) Karger, B. L.; Guttman, A. DNA Sequencing by CE. *Electrophoresis* **2009**, *30* (S1), S196–S202.

(178) 2004 Release: IHGSC Describes Finished Human Sequence http://www.genome.gov/12513430 (accessed Dec 26, 2015).

(179) Human Genome Sequencing Consortium International. Finishing the Euchromatic Sequence of the Human Genome. *Nature* **2004**, *431* (7011), 931–945.

(180) Maitra, R. D.; Kim, J.; Dunbar, W. B. Recent Advances in Nanopore Sequencing. *Electrophoresis* **2012**, *33* (23), 3418–3428.

(181) Navarro, E.; Serrano-Heras, G.; Castaño, M. J.; Solera, J. Real-Time PCR Detection Chemistry. *Clin. Chim. Acta.* **2015**, *439*, 231–250.

(182) Khare, V.; Eckert, K. A. The Proofreading 3'→5' Exonuclease Activity of DNA Polymerases: A Kinetic Barrier to Translesion DNA Synthesis. *Mutat. Res. Mol. Mech. Mutagen.* **2002**, *510* (1-2), 45–54.

(183) Sikorsky, J. A.; Primerano, D. A.; Fenger, T. W.; Denvir, J. DNA Damage Reduces Taq DNA Polymerase Fidelity and PCR Amplification Efficiency. *Biochem. Biophys. Res. Commun.* **2007**, *355* (2), 431–437.

(184) Bergen, K.; Betz, K.; Welte, W.; Diederichs, K.; Marx, A. Structures of KOD and 9°N DNA Polymerases Complexed with Primer Template Duplex. *Chembiochem* **2013**, *14* (9), 1058–1062.

(185) Tsujikawa, L.; Weinfield, M.; Reha-Krantz, L. J. Differences in Replication of a DNA Template Containing an Ethyl Phosphotriester by T4 DNA Polymerase and Escherichia Coli DNA Polymerase I. *Nucleic Acids Res.* **2003**, *31* (17), 4965–4972.

(186) Li, Y. Crystal Structures of Open and Closed Forms of Binary and Ternary Complexes of the Large Fragment of Thermus Aquaticus DNA Polymerase I: Structural Basis for Nucleotide Incorporation. *EMBO J.* **1998**, *17* (24), 7514–7525.

(187) Loakes, D.; Holliger, P. Polymerase Engineering: Towards the Encoded Synthesis of Unnatural Biopolymers. *Chem. Commun. (Camb).* **2009**, No. 31, 4619–4631.

(188) Demesmaeker, A.; Waldner, A.; Lebreton, J.; Hoffmann, P.; Fritsch, V.; Wolf, R. M.; Freier, S. M. Amides as a New-Type of Backbone Modification in Oligonucleotides. *Angew. Chemie-International Ed. English* **1994**, *33* (2), 226–229.

(189) Horton, D.; Chen, K. M.; No, Z.; Lee, H. C. Synthesis of 3’-C-Substituted Thymidine Derivatives by Free-Radical Techniques: Scope and Limitations. *Carbohydr. Res.* **2007**, *342* (2), 259–267.

(190) Chu, C. K.; Doboszewski, B.; Schmidt, W.; Ullas, G. V.; Van Roey, P. Synthesis of Pyrimidine 3’-Allyl-2',3'-Dideoxyribonucleosides by Free-Radical Coupling. *J. Org. Chem.* **1989**, *54* (11), 2767–2769.

(191) Butterfield, K.; Thomas, E. J. Synthesis of Analogues of Nucleotides with All-Carbon Backbones: Synthesis of N-Protected C-Linked Dinucleotides. *J. Chem. Soc. Perkin Trans. 1* **1998**, No. 4, 737–746.

(192) Kruszewska, H.; Chmielowiec, U.; Bednarek, E.; Witowska-Jarosz, J.; Dobrowolski, J. C.; Misicka, A. Spectroscopic Identification of AZT Derivative Obtained from Biotransformation of AZT by Stenotrophomonas Maltophilia. *J. Mol. Struct.* **2003**, *651*, 645–650.

(193) Wu, Z. Synthesis and Characterization of Active Ester-Functionalized Fluorescent Polymers: New Materials for Protein Conjugation. *J. Appl. Polym. Sci.* **2008**, *110* (2), 777–783.

(194) Gibson, D.; Binyamin, I.; Haj, M.; Ringel, I.; Ramu, A.; Katzhendler, J. Anthraquinone Intercalators as Carrier Molecules for Second-Generation Platinum Anticancer Drugs. *Eur. J. Med. Chem.* **1997**, *32* (10), 823–831.