Ferrocene-containing DNA assemblies

By

Yifeng Ma

UNIVERSITY OF BIRMINGHAM

A thesis submitted to the

University of Birmingham

For the degree of

DOCTOR OF PHILOSOPHY

School of Chemistry

College of Engineering and Physical Sciences

University of Birmingham

March 2024

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

In recent decades, significant achievements have been made in the development of modified nucleic acids and nucleic acid analogues. As far as metal-containing systems are concerned, ferrocene has emerged as a popular choice due to its stability, ease of functionalisation and well-understood electrochemistry. It can also be readily incorporated into oligomeric strands using standard automated synthesis.

In this thesis, the scope of ferrocene-containing nucleic acids has been expanded in two main ways. Firstly the design, assembly and electrochemical characterisation of ferrocene-containing DNA tetrahedra are reported. These nanostructures, which contain ferrocene units located at up to four of their vertices, have been shown to form stable self-assembled monolayers (SAMs) on gold surfaces and exhibit reversible electrochemistry.

Secondly a series of hybrid oligomer strands containing DNA and ferrocene nucleic acid (FcNA) units have been synthesised. The synthetic chemistry of the FcNA monomeric components has been optimised so that units with either two thymine or two cytosine units can be made in sufficient quantities for oligomerisation. Further evidence is provided using high field NMR spectroscopy for cross-duplex H-bonding between FcNA and DNA nucleobases within complementary duplexes. Furthermore the FcNA unit has been to shown to be compatible for templating the insertion of polymerase-mediated nucleotide triphosphates into growing DNA strands, providing the first example of a functional biological use for FcNA-containing oligonucleotides.

Acknowledgements

I extend my heartfelt gratitude to Prof Jim Tucker for entrusting me with the opportunity to undertake this project. Thank you for your dedicated mentorship and selfless support. Your expertise and patient guidance have been invaluable, enabling me to continuously improve and progress. Similar gratitude is extended to Dr. Sarah Horswell, who embraced the role of my second supervisor with enthusiasm.

I would like to express my gratitude to my colleagues, Francia, Liyao, Jack, Aldrich, Jack, and La, who have been my companions throughout this academic journey. Their camaraderie, shared experiences, and insightful discussions have enriched my learning and fostered a supportive environment for growth.

Special thanks are owed to the staff of the analytical facilities at the University of Birmingham, particularly John, Chris, and Allen. Their exceptional service and wealth of knowledge have not only facilitated our work but have also been instrumental in resolving numerous challenges.

I express my gratitude to my collaborators at the Institut Pasteur, Dr. Marcel Hollenstein, Chiara, and Fabiane, for their assistance in conducting experiments. I am thankful for their hospitality during my time in Paris, and I thoroughly enjoyed our collaboration.

Finally, and most importantly, I want to convey my deepest gratitude to my family. Their steadfast support, love, and encouragement have been my anchor throughout this

journey. Their sacrifices and unwavering belief in me have propelled me forward, even during the most challenging times.

Contents

Abbreviations

A	Adenine
Ac	Acetyl
aq.	Aqueous
Bu	Butyl
Bz	Benzoyl
C	Cytosine
C.E.	Counter electrode
CBS	Corey-Bakshi-Shibata
Cp	Cyclopentadienyl
CV	Cyclic voltammogram
DCM	Dichloromethane
DEAD	Diethylazodicarboxylate
DIAD	Diisopropylazodicarboxylate
DIBAL	Diisobutylaluminium hydride
DIPEA	Diisopropylethyamine
DMF	Dimethylformamide
DMT	Dimethoxytrityl
DNA	Deoxyribonucleic acid
E	Potential / V
E_p^a	Anodic peak potential / V
E_p^c	Cathodic peak potential / V
ESI	Electrospray Ionisation
e.e.	Enantiomeric excess
eq.	Equivalents
ES	Electrospray ionisation
Et	Ethyl
Fc	Ferrocene
FcNA	Ferrocene nucleic acid
G	Guanine
H-bond	Hydrogen bond

HPLC	High performance liquid chromatography
hr	Hour
IR	Infrared
Me	Methyl
MeOH	Methanol
MeCN	Acetonitrile
min	Minutes
MS	Mass spectrometry
n-BuLi	n-Butyllithium
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
PAGE	Polyacrylamide gel electrophoresis
ppm	Parts per million
pr	pyridine
T	Thymine
TBAF	Tetrabutylammonium fluoride
TBDPS	tert-butyldiphenylsilyl
TBE	Tris-borate-EDTA
TEA	Triethylamine
THF	Tetrahydrofuran
TLC	Thin-layer chromatography
TMEDA	Tetramethylethylenediamine
TMS	Trimethylsilyl
TOF	Time-of-flight
U	Uracil
UV	Ultraviolet
W.E	Working electrode

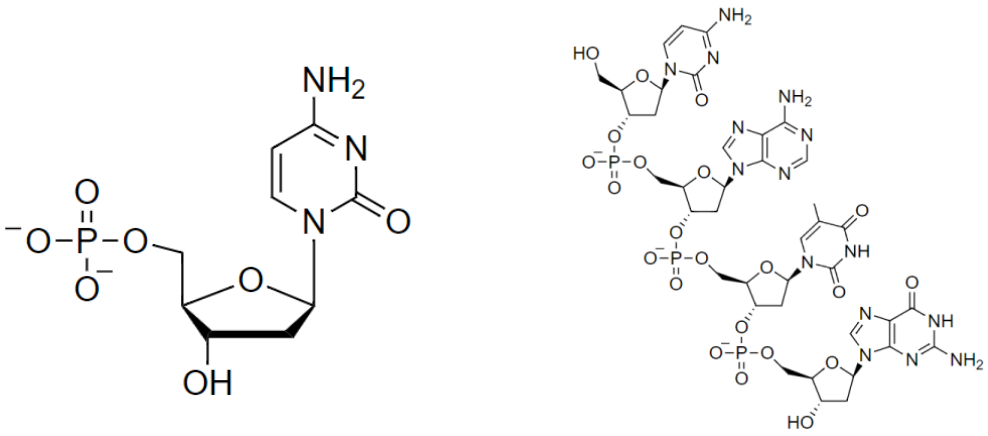
Chapter 1 – Introduction

1.1 Deoxyribonucleic Acid (DNA)

Nucleic acids, essential biopolymers for all living organisms, are composed of monomers called nucleotides. In 1953, James Watson and Francis Crick revealed the double-helix secondary structure of DNA. Their work, which was published in the journal *Nature*, is often referred to as the revelation of the "secret of life," and they were awarded the Nobel Prize in 1962 for their remarkable achievement. This discovery remains one of the greatest breakthroughs in modern biology.[1]

1.1.1 Structure of DNA and RNA

The structure of DNA consists of nucleotides strung together to form an oligonucleotide, with each monomeric nucleotide unit consisting of three components: a phosphomonoester, a ribose sugar, and an *N*-heterocyclic base (see Fig.1.1). The phosphomonoester is attached at the 5' position of the sugar, while the *N*-heterocyclic base is attached at the 1' position.[2]



1.1 Structure of a nucleotide on the left and an oligo ssDNA on the right

A dinucleotide (dimer) in DNA or RNA is created by chemically bonding the 5'-phosphate group of one nucleotide to the 3'-hydroxyl group of another, forming a phosphodiester bond. An oligonucleotide (oligomer) is produced when multiple such bonds are established. Naturally occurring nucleic acids are extended, high-molecular-weight molecules of this type. The phosphomonoester group in nucleotides becomes deprotonated under physiological conditions, contributing to the acidity and negative charge of nucleic acids.

The sugar in a nucleotide can be either ribose, or deoxyribose, which does not have a hydroxyl group at the 2' position of the sugar, giving either ribonucleic acid (RNA) or deoxyribonucleic acid (DNA). The nucleotides' *N*-heterocyclic bases can exist in two forms: monocyclic, known as pyrimidines, and bicyclic, known as purines. DNA and RNA each have two purines and two pyrimidines as shown in Figure 1.2, with thymine found in DNA and uracil found in RNA. When labeling sequences of nucleic acids, the first letter of each base corresponds to each nucleobase: A for adenine, T for thymine, C for cytosine, G for guanine and U for uracil.[2]

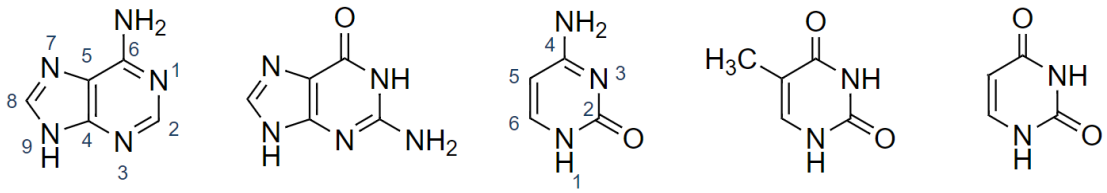


Figure 1.2 *N*-heterocyclic bases in nucleic acids. U is found in RNA and T in DNA.

The fundamental structure of an oligonucleotide is a linear single strand. Hydrogen bonds form between nucleobases, either within the same strand or between different

strands, leading to the creation of more complex structures within oligonucleotides. Cytosine pairs with guanine through three hydrogen bonds, while adenine pairs with thymine (in DNA) or uracil (in RNA) through two hydrogen bonds (Figure 1.3).

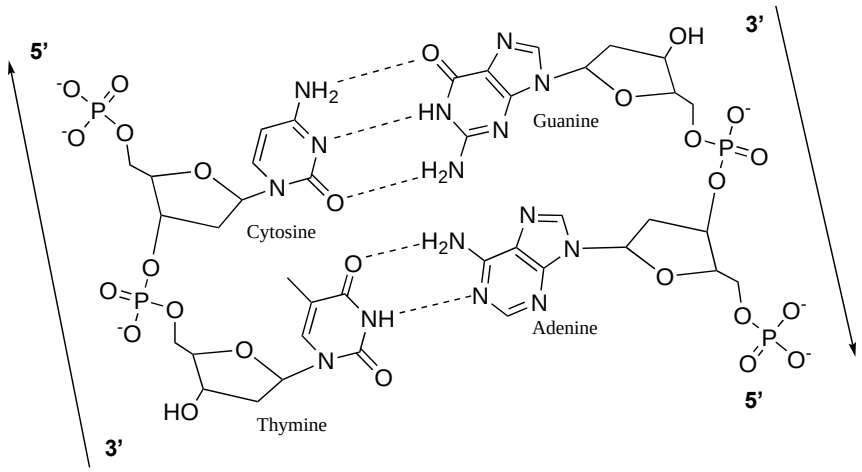


Figure 1.3 Complementary base pairing in DNA

When two oligonucleotide strands possess complementary base sequences, they combine to form a double helix structure. This helix consists of complementary oligonucleotide strands running in opposite directions, where the 3' end of one strand forms pairs with the 5' end of the other. Subsequent studies have revealed the dynamic nature of this helix, indicating variations in dimensions and shapes. Despite these variations, three primary conformations have been identified (Figure 1.4).

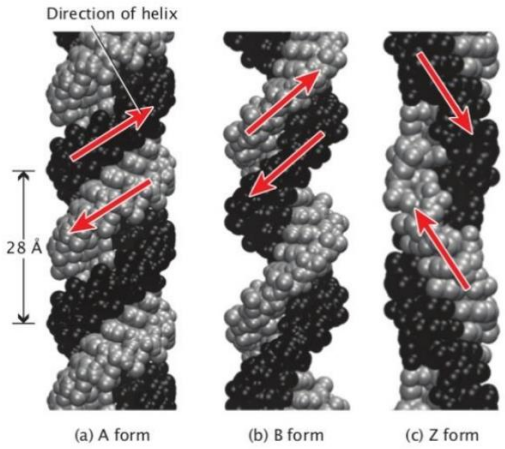


Figure 1.4 Structures of a) A-DNA b) B-DNA and c) Z-DNA

B-DNA, which appears as a right-handed helix, represents the most common DNA structure in the natural environment. It displays a standard base-pair spacing of 0.34 nm. In contrast, A-DNA adopts a more compact right-handed helical arrangement, especially under dehydrated conditions, characterized by a reduced base-pair stacking distance of 0.28 nm. Meanwhile, the Z-DNA conformation assumes a left-handed helical structure, shaped by the abundance of alternating guanine and cytosine bases.[3][4][5]

1.2 DNA Modifications

Over the past few decades, there have been numerous advances in developing nucleic acid analogues. The simplest approach to achieve such structures is to modify one or more components of the repeating nucleotide unit that comprises the structure of a natural nucleic acid. As stated earlier, the nucleotide unit is made up of three elements: a ribose sugar, a phosphomonoester, and an *N*-heterocyclic base. The following three sub-sections describe some examples of organic modifications of each of these elements in turn.

1.2.1 Sugar Backbone Modification

The first strategy is sugar backbone modification. Sugar modifications are employed to enhance the resistance of oligonucleotides to thermal and nuclease degradation. These modifications are typically introduced at the C’2 and C’4 positions of the sugar molecule. They are more commonly observed in nucleosides based on pyrimidine

because these are more susceptible to nuclease degradation.[6][7]

Some modifications have been widely used in the oligonucleotide industry, especially in oligonucleotide therapeutics.[7] For example, in the 1960s, modifications such as 2'-fluorothymidine (2'-F),[8] and 2'-O-methylated polyadenylic acid were first reported (Figure 1.5).[9] These 2'-F and 2'-O-Me modifications play significant roles in certain medications. For example, they are utilised in MACUGEN[10], an anti-VEGF aptamer used for treating wet age-related macular degeneration, as well as in Givosiran[10], an siRNA for the treatment of acute hepatic porphyria. Additionally, the 2'-O-Me modification can be found in Patisiran[11], which was the first FDA-approved RNAi therapeutic in 2018, used for the treatment of hereditary transthyretin-mediated polyneuropathy. Recently, 2'-O-Me/2'-F modified oligonucleotides are still represented and include Inclisiran, an siRNA targeting hypercholesterolaemia that gained approval from the US FDA in 2022.[12]

2'-O-(2-Methoxyethyl)-RNA (MOE-RNA) is another successful modification. The synthesis of MOE-modified building blocks and oligonucleotides was initially reported in 1995[13] and, over approximately 20 years, MOE chemistry has progressed from laboratory research to practical clinical applications. The first of three MOE-based therapeutics to receive FDA approval was KYNAMRO, designed for the treatment of homozygous familial hypercholesterolaemia[14]

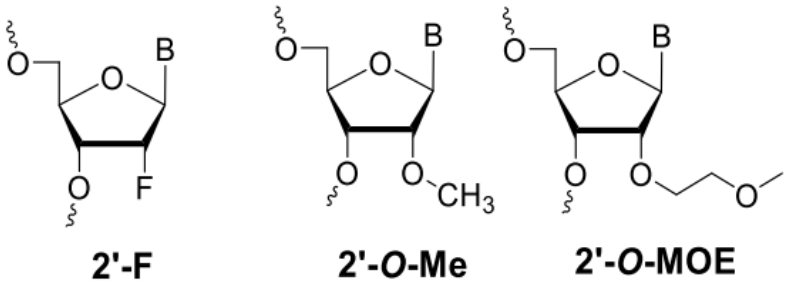


Figure 1.5 Sugar modifications used commercially

1.2.1.1 Threose nucleic acid

A type of unnatural oligonucleotide called TNA, threose nucleic acid, has been designed.[15] The structure of TNA is illustrated in Figure 1.6. These TNAs have closely connected phosphodiester bridges and can participate in base pairing when their strands are arranged in an antiparallel strand orientation. TNA stands out as one of the simplest structures among all the nucleic acid alternatives of the oligonucleotide type studied thus far. This is because TNA is derived from a sugar that contains only four carbon atoms. Furthermore, TNAs have the capability to create hybrid pairs with both RNA and DNA.

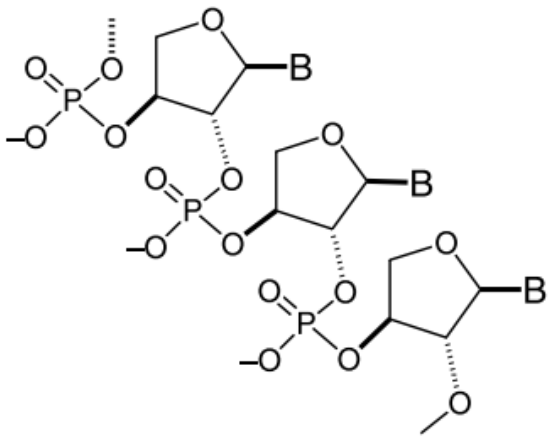


Figure 1.6 Structure of threose nucleic acid

1.2.1.2 Hexitol nucleic acids

Another example of a unnatural modified oligonucleotide is hexitol nucleic acid (HNA), which is formed from a pyranose rather than a furanose (Figure 1.7).[16] In light of the favorable energetic properties of pyranose oligomers over furanose oligomers, due to a smaller change in entropy during duplex formation, researchers have explored various pyranose derivatives.[17-20] In one case, an additional methylene group inserted between the base moiety and the ring oxygen atom results in a significantly less flexible tetrahydropyran skeleton. The increase of the melting temperature value of 3 to 5.5 degrees Celsius per base pair illustrates the strong interaction of these HNAs with their complementary strands.[21]

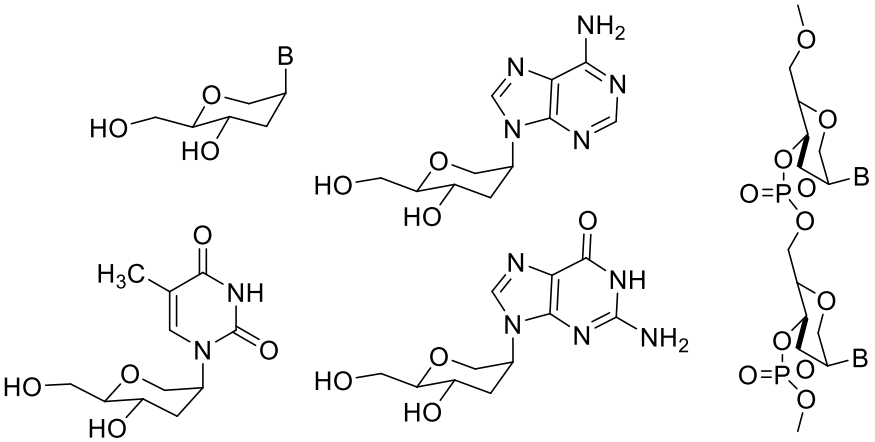


Figure 1.7 Structure of various hexitol nucleic acids

1.2.1.3 Bicyclo-DNA and Tricyclo-DNA

Bicyclo-DNA, a modified oligonucleotide analogue, belongs to a group of molecules with restricted conformations. [22] Bicyclo-DNA sequences composed primarily of homopurine bases have been observed to form duplex structures that are entropically stabilized. These duplexes show a preference for adopting the Hoogsteen modes of

base pairing, which differ from the typical association mode seen in regular DNA.[23] Tricyclo-DNA, a second-generation derivative of bicyclo-DNA, incorporates an additional cyclopropane ring to enhance the nucleoside's structural stability.[24] Tricyclo-DNA forms highly stable Watson-Crick base pairs, showing increased thermal stability when paired with complementary DNA, with a melting temperature increase of up to +1.2 °C. When paired with complementary RNA, the stability increase is even higher, at +2.4 °C. In self-pairing, tricyclo-DNA exhibits a remarkable stability increase of +3.1 °C per base pair compared to regular DNA.[25]

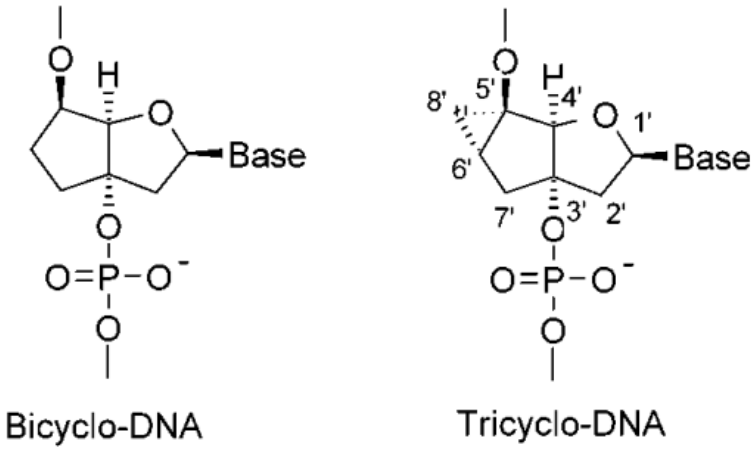


Figure 1.8 Structure of Bicyclo-DNA Tricyclo-DNA.[25]

1.2.1.4 Locked nucleic acids

Imanishi [26-27] and Wengel [28-29] independently created locked nucleic acids (LNAs), a unique type of nucleic acid with a 2'-4'-bicyclic structure (Figure1.9). LNAs restrict nucleic acid flexibility by locking the sugar molecule into a specific conformation. They are highly effective in binding to RNA targets and enhancing duplex stability when used in oligonucleotides.[30-31] LNA oligonucleotides can be synthesized with various tags commonly used for nucleic acids, and they have been employed in different applications. These applications include using Cy5-labelled LNA oligonucleotides as

PCR primers[32] and anthraquinone-tagged LNA oligonucleotides for covalent attachment to polymer surfaces with UV light.[33] Additionally, LNA oligonucleotides have been tested as substrates for various enzymes, including those for radioactive labelling and PCR amplification.[34] They have also been evaluated for their stability against nucleases, particularly in therapeutic contexts. However, excessive stability can lead to non-specific binding, so LNA oligonucleotides require careful engineering to balance their thermodynamic properties. [35-36]

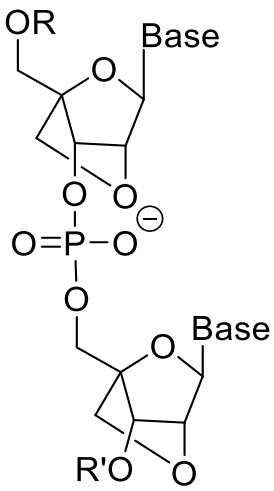


Figure 1.9 Structure of locked nucleic acids (LNAs).

1.2.2 Phosphate Linker Modification

The linker between monomers in oligonucleotides is the phosphodiester backbone (PO). The connection involves five bonds and sp^3 hybridised centres that provide flexibility. At the core of the backbone lies the phosphate atom, surrounded by two bridging and two non-bridging oxygen atoms. Alterations to the phosphate backbone can improve nucleic acid properties, including increased resistance to enzymatic degradation, enhanced thermal stability, and greater cellular uptake.[37]

1.2.2.1 Phosphorothioate linkers

Phosphorothioates (PS, shows in Figure 1.10) are structurally similar to phosphodiesters and stand out as one of the extensively studied modified backbone structures.[38] The primary distinction between these two backbones lies in the fact that phosphorothioate includes a sulfur atom in place of a non-bridging oxygen atom. VITRAVENE, one of the first oligonucleotide therapeutic to gain clinical approval, is an antisense drug used for the treatment of cytomegalovirus retinitis. It is a 21-mer oligo-2-deoxynucleotide that has been entirely modified with phosphorothioate (PS) groups.[39-40] While this modification enhances resistance to phosphatase enzymes, it also disrupts the oligonucleotide duplex structure due to the larger size of sulfur atoms. Furthermore, controlling stereochemistry during synthesis is challenging, and the ideal configuration for therapeutic purposes remains uncertain.[41][42]

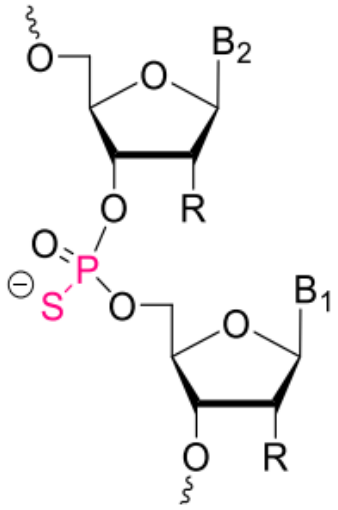


Figure 1.10 Structure of Phosphorothioates modification[7]

A different phosphate modification option involves using methylphosphonates, which replace one of the non-bridging oxygen atoms with a methyl group, so creating a neutral linkage.[43] Researchers have also developed nonionic RNA analogues by

replacing the whole phosphate diester linkers with various sulfone and sulfide-based linkages (e.g. Figure 1.11).[44] Short oligonucleotide analogues with a specific sulfone linkage were stable and could form Watson-Crick pairs but longer analogues lost their pairing ability and showed unusual behaviours, including the potential to catalyse reactions.

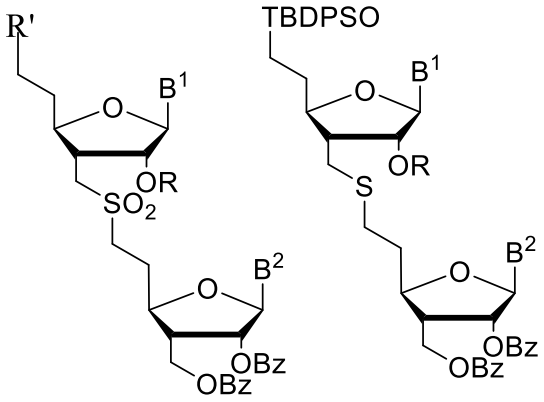


Figure 1.11 Structure of sulfone and sulfide-based linkages.

1.2.2.2 Triazole and amide linkers

Triazole-containing backbones are playing an increasingly important role in the design of artificial nucleic acids. Their formation involves a so-called “click reaction” based on the cycloaddition reaction between an azide and alkyne. [46] Various triazole linkages have been developed to replace native phosphate backbones in DNA analogues.[45-47] In the polymerase chain reaction (PCR), this triazole backbone linkage is correctly recognized by DNA polymerase and functions effectively in E*scherichia coli*.[48][49] At the same time, amide linkers have continued to attract interest, for example in modulating the function of siRNAs.[74]

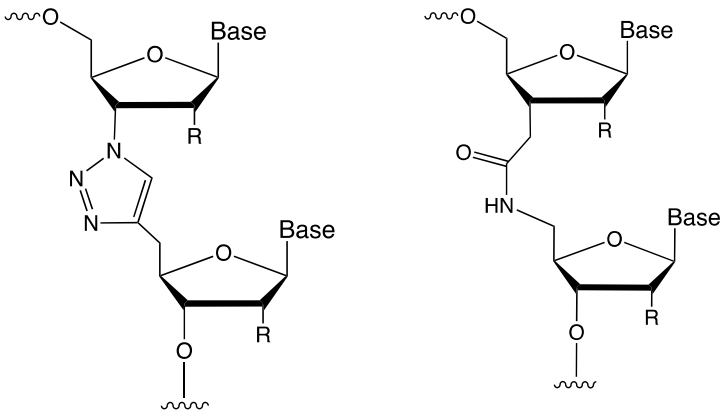


Figure 1.12 Structures of triazole and amide backbone linkers in nucleic acids.[73][74]

1.2.2.3 Peptide nucleic acid (PNA)

Peptide nucleic acid, developed by Nielsen in the 1990s,[75] is an extensively used DNA analogue. In this system, the conventional phosphate-sugar backbone is entirely substituted with repeating *N*-(2-aminoethyl)glycine monomers, and different bases are attached through methylene carbonyl linkers. PNA serves as a DNA mimic, and its spatial arrangement and conformational flexibility allow it to specifically hybridize with natural DNA. Complementary base pairs form conventional hydrogen bonds, resulting in a PNA/DNA duplex.[76] PNA shows promising biomedical potential, particularly as an antisense agent. It can selectively target and regulate gene expression by binding to both DNA and RNA, despite existing challenges in delivery.[77]

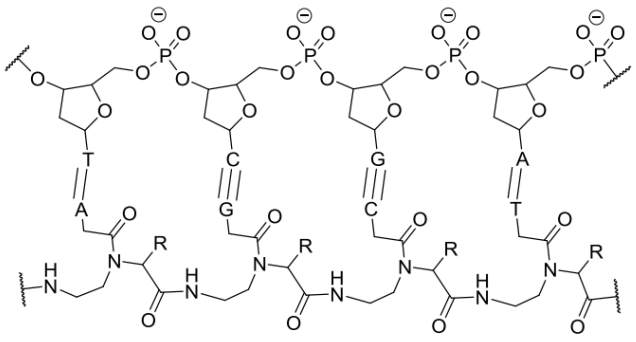


Figure 1.13 Structure of PNA and its H-bonding to DNA.

1.2.3 Nucleobase modifications

Nucleobase modifications are commonly used to enhance the binding affinity and specificity of aptamers.[7] Unmodified oligos utilise the original five nucleobases A, T, C, G and U. These modifications are typically added to the C5 position of pyrimidines shows in figure 1.14 and the N7/C8 positions of purines because they are more accessible for synthesis and do not disrupt base pairing or other hydrogen bonding interactions. However, even slight chemical alterations to these bases, such as methylation, can significantly affect the function of DNA, and more dramatic changes, *e.g.* substituting them with unnatural base pairs, have the potential to expand its capabilities.[51]

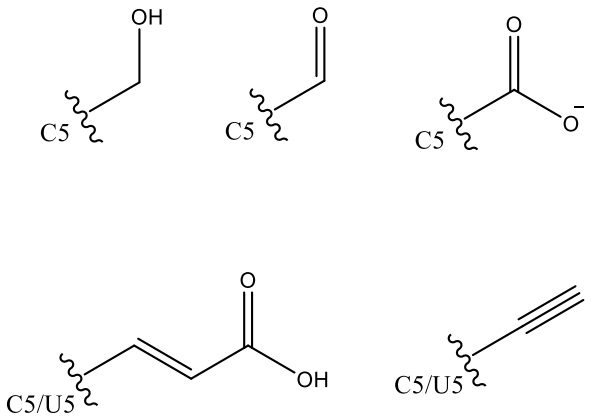


Figure 1.14 Structures of C-5 modifications

While one approach focuses on a more fundamental strategy involves developing unnatural base pairs (UBPs).[50] These UBPs, shown for example in Figure1.15, expand the genetic alphabet and enable the storage of more information within a DNA molecule. Benner pioneered this effort by focusing on unnatural nucleotides with hydrogen bonding patterns different from those found in natural base pairs.[52] Subsequently, researchers, including the Kool laboratory and the Hirao laboratory,[53]

explored alternative forces, such as hydrophobic interactions and packing interactions, to control UBP formation. Hirao's work involved modifying natural purine and pyrimidine structures to create "shape complementary" UBPs[54], while another approach focused on nucleotides with nucleobase analogues that had little to no similarity to their natural counterparts.

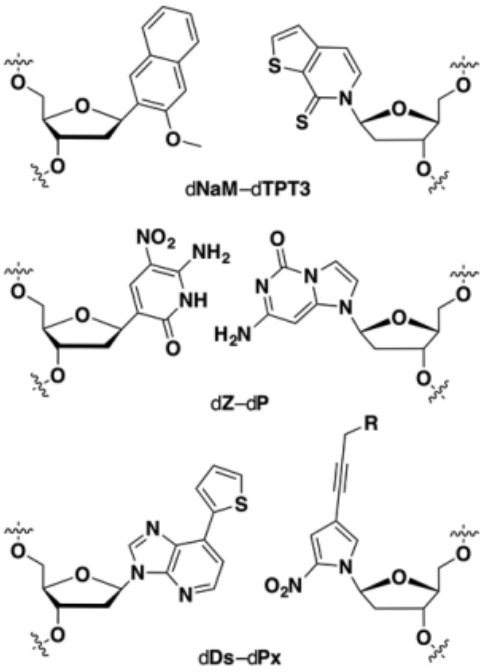


Figure 1.15 Examples of unnatural base pairs used in modified nucleic acids

1.3 Metal-containing nucleic acids

In recent years, metal-mediated DNA base pairs, composed of two ligand-type artificial nucleobases and a bridging metal ion, have gained growing attention due to their unique base pairing mode distinct from natural base pairing. In 1962, Katz demonstrated the first metal-base pair by stabilizing a T/T mismatch using Hg^{2+}.[55] A similar discovery was made in 2007 by Ono, who employed Ag^{+} ions to stabilize a C/C mismatch, creating the C-Ag^{+}-C metal-base pair.[56] In both cases, the DNA duplex maintains its usual shape, with the linear metal complex complementing the flat geometry of traditional Watson-Crick base pairing.

Using only the naturally occurring DNA bases, interactions with metals are limited, resulting in very few possible metal-base pairs. Moreover, interference from the standard Watson-Crick base pairing is a common issue. To address these limitations, researchers sought an entirely independent system in which the DNA bases would be replaced with specially designed ligands. This substitution would allow for the optimization of metal complexation properties in terms of both selectivity and strength, significantly increasing the number of possible metal-base pairs while minimizing interference from Watson-Crick pairing. To achieve this goal, Shionoya developed various chelator-type bases (Figure 1.16), expanding the scope and enhancing the characteristics of these metal-base pairs. [57-59]

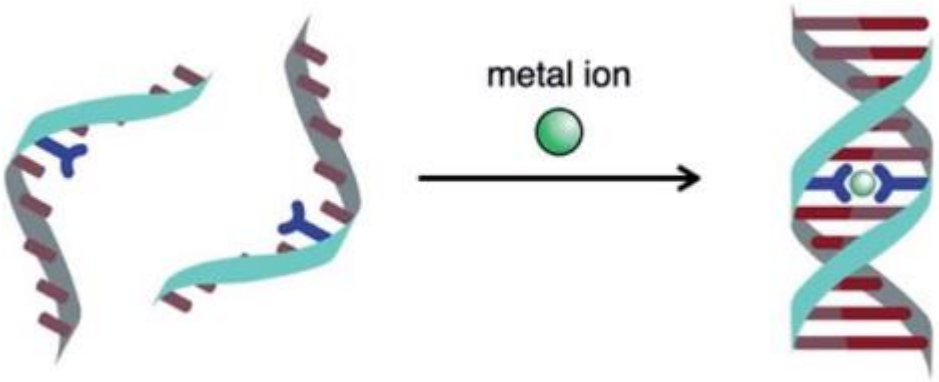


Figure 1.16 Schematic diagram showing formation of a DNA duplex containing one metal-mediated base pair.

Various artificial nucleobases with ligand-like structures have been created by modifying simple ligand frameworks such as pyridine, maltol, and pyrimidine. Among these, imidazole, one of the smallest scaffolds, has been widely explored.[60-62] Modified imidazole nucleobases tend to form metal-mediated base pairs, particularly with Ag(I), through N-Ag(I)-N coordination. This metal-binding behavior can be adjusted by introducing additional coordination sites. For instance, imidazole-4-carboxylate nucleobases can form base pairs through Cu(II) and Ag(I)-mediated coordination. These carboxylate groups carry negative charges that help to neutralize the positive charge of the bridging metal ion, making the imidazole-Cu pair become one of the most stabilising artificial base pairs. Notably, these base pairs can also interact with other metals such as Ni(II) and Co(II). imidazole-Cu pairing has been successfully applied in the metal-dependent regulation of DNAzymes[63-65] and in PCR amplification. [66]

1.2.3 Ferrocene-containing nucleic acids

Over the course of several decades, numerous inorganic systems have been explored wherein metal complexes are connected to nucleobases, with ferrocene emerging as a favoured option due to its stability, relative ease of functionalization and well-understood electrochemistry.[78-80] The inaugural instances of ferrocenyl-nucleobase compounds, encompassing a range of adenine derivatives, were disclosed by Chen in 1980.[67] Letsinger et al. found that the phosphoramidite approach could be easily applied, building upon earlier research that incorporated different tags, like biotin or fluorescein. [68] This approach focused on the coupling of a ferrocenyl-phosphoramidite to a thymine through solid-phase DNA synthesis.

The development of ferrocene-based electrochemical DNA sensors has made significant progress in recent years. Pioneering work[69] introduced a ferrocene tag in a three-strand system, where the electrochemical signal was only detectable upon hybridization of the target DNA with both an unmodified and a ferrocene-modified oligonucleotide. Plaxco's molecular beacon system[70] utilized a ferrocene-modified DNA probe attached to an Au surface, allowing reliable detection through changes in electrochemical signals upon target DNA binding. Further advancements were made by incorporating a ratiometric system on surface[71], using methylene blue as a detector and ferrocene as an internal control. This innovative approach enhanced the accuracy of the sensor.

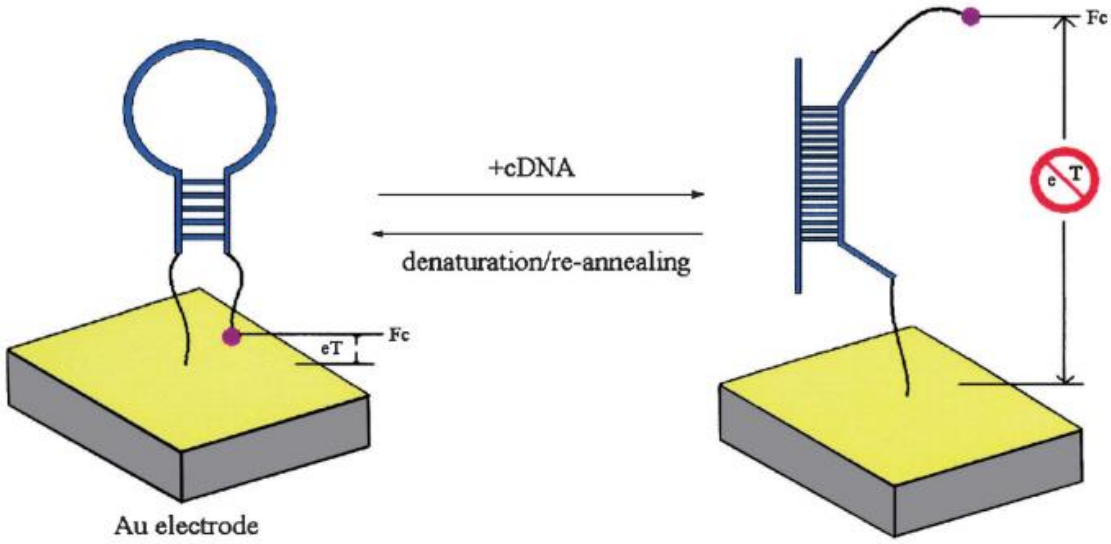


Figure 1.17 Electrical sensing of DNA conformational changes upon target detection. [70]

1.4 Thesis outline

Ferrocenes are widely used in DNA sensing because its unique electrochemical properties. Notably, the distance between two Cp units of ferrocene closely resembles that between adjacent base pairs in duplex DNA. Based on this similarity, the Tucker group developed a ferrocene-containing analogue of DNA, called ferrocene nucleic acid (FcNA) [72], which served as a structural mimic of DNA, this is discussed in more in the following chapters.

In this thesis, building upon our previous research, we have conducted an in-depth investigation into ferrocene-modified oligonucleotides, including those containing FcNA units. In Chapter 3, the synthesis of FcNA monomers containing thymine and cytosine nucleobases is described. In Chapter 4, the primary emphasis is on exploring the formation and electrochemical properties of self-assembled DNA structures

containing ferrocene. In Chapter 5, FcNA-DNA hybrid sequences and their duplexes are investigated, with their unique properties explored both in terms of structure and biological function.

References

1. Watson, J. D.; Crick, F. H. *Nature*, **1953**, *171*(4356), 737-738.
2. Michelson, A. M.; Todd, A. R. *Journal of the Chemical Society*, **1955**, 2632-2638.
3. Kool, E. T.; Morales, J. C.; Guckian, K. M. *Angewandte Chemie International Edition*, **2000**, *39*(6), 990-1009.
4. Kool, E. T. *Chemical reviews*, **1997**, *97*(5), 1473-1488.
5. Pray, L. A. *Nature education*, **2008**, *1*(1), 204.
6. Iadevaia, G.; Hunter, C. A. *Accounts of Chemical Research*, **2023**, *56*(6), 712-727.
7. Egli, M.; Manoharan, M. *Nucleic Acids Research*, **2023**, *51*(6), 2529-2573.
8. Codington, J. F.; Doerr, I.; Van Praag, D.; Bendich, A.; Fox, J. J. *Journal of the American Chemical Society*, **1961**, *83*(24), 5030-5031.
9. Furukawa, Y.; Kobayashi, K.; Kanai, Y.; Honjo, M. *Chemical and Pharmaceutical Bulletin*, **1965**, *13*(11), 1273-1278.
10. Scott, L. J. *Drugs*, **2020**, *80*(3), 335-339.
11. Akinc, A.; Maier, M. A.; Manoharan, M.; Fitzgerald, K.; Jayaraman, M.; Barros, S.; Cullis, P. R. *Nature nanotechnology*, **2019**, *14*(12), 1084-1087.
12. Migliorati, J. M.; Jin, J.; Zhong, X. B. *Trends in pharmacological sciences*, **2022**, *43*(5), 455-456.
13. Martin, P. *ChemInform*, **1995**, *26*(31), no-no.
14. Geary, R. S.; Watanabe, T. A.; Truong, L.; Freier, S. U. E.; Lesnik, E. A.; Sioufi, N. B.; Levin, A. A. *Journal of Pharmacology and Experimental Therapeutics*, **2001**, *296*(3), 890-897.
15. Schoning, K. U.; Scholz, P.; Guntha, S.; Wu, X.; Krishnamurthy, R.; Eschenmoser, A. *Science*, **2000**, *290*(5495), 1347-1351.
16. Verheggen, I.; Van Aerschot, A.; Van Meervelt, L.; Rozenski, J.; Wiebe, L.; Snoeck, R.; Claes, P. *Journal of medicinal chemistry*, **1995**, *38*(5), 826-835.
17. Eschenmoser, A. *Pure and Applied Chemistry*, **1993**, *65*(6), 1179-1188.
18. Augustyns, K.; Van Aerschot, A.; Urbanke, C.; Herdewijn, P. *Bulletin des Sociétés Chimiques Belges*, **1992**, *101*(2), 119-130.
19. Augustyns, K.; Vandendriessche, F.; Van Aerschot, A.; Busson, R.; Urbanke, C.;

Herdewijn, P. *Nucleic acids research*, **1992**, *20*(18), 4711-4716.

20. Hendrix, C.; Rosemeyer, H.; Verheggen, I.; Van Aerschot, A.; Seela, F.; Herdewijn, P. *Chemistry–A European Journal*, **1997**, *3*(1), 110-120.
21. Aerschot Van, A.; Verheggen, I.; Hendrix, C.; Herdewijn, P. *Angewandte Chemie International Edition*, **1995**, *34*(12), 1338-1339.
22. Tarköy, M.; Bolli, M.; Leumann, C. *Helvetica chimica acta*, **1994**, *77*(3), 716-744.
23. Bolli, M.; Trafelet, H. U.; Leumann, C. *Nucleic acids research*, **1996**, *24*(23), 4660-4667.
24. Steffens, R.; Leumann, C. J. *Journal of the American Chemical Society*, **1999**, *121*(14), 3249-3255.
25. Renneberg, D.; Leumann, C. J. *Journal of the American Chemical Society*, **2002**, *124*(21), 5993-6002.
26. Obika, S.; Onoda, M.; Morita, K.; Andoh, J. I.; Koizumi, M.; Imanishi, T. *Chemical Communications*, **2001**, (19), 1992-1993.
27. Obika, S.; Nanbu, D.; Hari, Y.; Morio, K. I.; In, Y.; Ishida, T.; Imanishi, T. *Tetrahedron Letters*, **1997**, *38*(50), 8735-8738.
28. Koshkin, A. A.; Singh, S. K.; Nielsen, P.; Rajwanshi, V. K.; Kumar, R.; Meldgaard, M.; Wengel, J. *Tetrahedron*, **1998**, *54*(14), 3607-3630.
29. Singh, S. K.; Koshkin, A. A.; Wengel, J.; Nielsen, P. *Chemical communications*, **1998**, (4), 455-456.
30. Petersen, M.; Nielsen, C. B.; Nielsen, K. E.; Jensen, G. A.; Bondensgaard, K.; Singh, S. K.; Jacobsen, J. P. *Journal of Molecular Recognition*, **2000**, *13*(1), 44-53.
31. Petersen, M.; Bondensgaard, K.; Wengel, J.; Jacobsen, J. P. *Journal of the American Chemical Society*, **2002**, *124*(21), 5974-5982.
32. Choleva, Y.; Nørholm, M.; Pedersen, S.; Mouritzen, P.; Høiby, P. E.; Nielsen, A. T.; Kongsbak, L. *JALA: Journal of the Association for Laboratory Automation*, **2001**, *6*(4), 92-97.
33. Ørum, H.; Jakobsen, M. H.; Koch, T.; Vuust, J.; Borre, M. B. *Clinical chemistry*, **1999**, *45*(11), 1898-1905.
34. Ørum, H.; Wolter, A.; Kongsbak, L. *Letters in Peptide Science*, **2003**, *10*, 325-334.
    a) Burel, S. A.; Hart, C. E.; Cauntay, P.; Hsiao, J.; Machemer, T.; Katz, M.; Henry,

S. P. *Nucleic acids research*, **2016**, *44*(5), 2093-2109.

35. Burel, S. A., Hart, C. E., Cauntay, P., Hsiao, J., Machemer, T., Katz, M., Watt, A., Bui, H.-h., Younis, H., & Sabripour, M. *Nucleic Acids Res*. 2016, 44 (5), 2093-2109.
36. Wahlestedt, C.; Salmi, P.; Good, L.; Kela, J.; Johnsson, T.; Hökfelt, T.; Wengel, J. *Proceedings of the National Academy of Sciences*, **2000**, *97*(10), 5633-5638.
37. Wang, X.; Feng, M.; Xiao, L.; Tong, A.; Xiang, Y. *ACS chemical biology*, **2016**, *11*(2), 444-451.
38. Grünweller, A.; Wyszko, E.; Bieber, B.; Jahnel, R.; Erdmann, V. A.; Kurreck, J. *Nucleic acids research*, **2003**, *31*(12), 3185-3193.
39. Crooke, S. T.; Witztum, J. L.; Bennett, C. F.; Baker, B. F. *Cell metabolism*, **2018**, *27*(4), 714-739.
40. Shen, X.; Corey, D. R. *Nucleic acids research*, **2018**, *46*(4), 1584-1600.
41. Nukaga, Y.; Yamada, K.; Ogata, T.; Oka, N.; Wada, T. *The Journal of Organic Chemistry*, **2012**, *77*(18), 7913-7922.
42. Iwamoto, N.; Butler, D. C.; Svrzikapa, N.; Mohapatra, S.; Zlatev, I.; Sah, D. W.; Verdine, G. L. *Nature biotechnology*, **2017**, *35*(9), 845-851.
43. Levis, J. T.; Miller, P. S. *Antisense Research and Development*, **1994**, *4*(4), 223-230.
44. Richert, C.; Roughton, A. L.; Benner, S. A. *Journal of the American Chemical Society*, **1996**, *118*(19), 4518-4531.
45. Fujino, T.; Yamazaki, N.; Isobe, H. *Tetrahedron Letters*, **2009**, *50*(28), 4101-4103.
46. Isobe, H.; Fujino, T.; Yamazaki, N.; Guillot-Nieckowski, M.; Nakamura, E. *Organic letters*, **2008**, *10*(17), 3729-3732.
47. El-Sagheer, A. H.; Brown, T. *Quarterly Reviews of Biophysics*, **2015**, *48*(4), 429-436.
48. Birts, C. N.; Sanzone, A. P.; El‐Sagheer, A. H.; Blaydes, J. P.; Brown, T.; Tavassoli, A. *Angewandte Chemie International Edition*, **2014**, *53*(9), 2362-2365.
49. El-Sagheer, A. H.; Sanzone, A. P.; Gao, R.; Tavassoli, A.; Brown, T. *Proceedings of the National Academy of Sciences*, **2011**, *108*(28), 11338-11343.
50. Feldman, A. W.; Romesberg, F. E. *Accounts of chemical research*, **2018**, *51*(2), 394-403.

51. Shivalingam, A.; Brown, T. *Biochemical Society Transactions*, **2016**, *44*(3), 709-715.
52. Benner, S. A.; Karalkar, N. B.; Hoshika, S.; Laos, R.; Shaw, R. W.; Matsuura, M.; Moussatche, P. *Cold Spring Harbor perspectives in biology*, **2016**, *8*(11), a023770.
53. Moran, S.; Ren, R. X. F.; Kool, E. T. *Proceedings of the National Academy of Sciences*, **1997**, *94*(20), 10506-10511.
54. Hirao, I.; Kimoto, M.; Yamashige, R. *Accounts of chemical research*, **2012**, *45*(12), 2055-2065.
55. Katz, S. *Biochimica et Biophysica Acta (BBA)-Specialized Section on Nucleic Acids and Related Subjects*, **1963**, *68*, 240-253.
56. Ono, A.; Cao, S.; Togashi, H.; Tashiro, M.; Fujimoto, T.; Machinami, T.; Tanaka, Y. *Chemical communications*, **2008**, (39), 4825-4827.
57. Tanaka, K.; Tengeiji, A.; Kato, T.; Toyama, N.; Shionoya, M. *Science*, **2003**, *299*(5610), 1212-1213.
58. Takezawa, Y.; Maeda, W.; Tanaka, K.; Shionoya, M. *Angewandte Chemie*, **2009**, *121*(6), 1101-1104.
59. Naskar, S.; Guha, R.; Mueller, J. *Angewandte Chemie International Edition*, **2020**, *59*(4), 1397-1406.
60. Johannsen, S.; Megger, N.; Böhme, D.; Sigel, R. K.; Müller, J. *Nature chemistry*, **2010**, *2*(3), 229-234.
61. Petrovec, K.; Ravoo, B. J.; Müller, J. *Chemical Communications*, **2012**, *48*(97), 11844-11846.
62. Hensel, S.; Megger, N.; Schweizer, K.; Müller, J. *Beilstein Journal of Organic Chemistry*, **2014**, *10*(1), 2139-2144.
63. Takezawa, Y.; Hu, L.; Nakama, T.; Shionoya, M. *Angewandte Chemie International Edition*, **2020**, *59*(48), 21488-21492.
64. Flamme, M.; Figazzolo, C.; Gasser, G.; Hollenstein, M. *Metallomics*, **2021**, *13*(4), mfab016.
65. Kobayashi, T.; Takezawa, Y.; Sakamoto, A.; Shionoya, M. *Chemical Communications*, **2016**, *52*(19), 3762-3765.
66. Kaul, C.; Müller, M.; Wagner, M.; Schneider, S.; Carell, T. *Nature chemistry*, **2011**,

3(10), 794-800.

67. Chen, S. C. *Journal of Organometallic Chemistry*, **1980**, *202*(2), 183-189.
68. Mucic, R. C.; Herrlein, M. K.; Mirkin, C. A.; Letsinger, R. L. *Chemical Communications*, **1996**, (4), 555-557.
69. Ihara,T.; Nakayama, M.; Murata, M.; Nakano, K.; Maeda, M. *Chemical Communications*, **1997**, (17), 1609-1610.
70. Fan, C.; Plaxco, K. W.; Heeger, A. J. *Proceedings of the National Academy of Sciences*, **2003**, *100*(16), 9134-9137.
71. Du, Y.; Lim, B. J.; Li, B.; Jiang, Y. S.; Sessler, J. L.; Ellington, A. D. *Analytical chemistry*, **2014**, *86*(15), 8010-8016.
72. Nguyen, H. V.; Zhao, Z. Y.; Sallustrau, A.; Horswell, S. L.; Male, L.; Mulas, A.; Tucker, J. H. *Chemical communications*, **2012**, *48*(100), 12165-12167.
73. Shivalingam, A.; Tyburn, A. E.; El-Sagheer, A. H.; Brown, T. *Journal of the American Chemical Society*, **2017**, *139*(4), 1575-1583.
74. Kotikam, V.; Rozners. *Accounts of Chemical Research*, **2020**, 53(9), 1782-1790.
75. Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O. *Science*, **1991**, *254*(5037), 1497-1500.
76. Nielsen, P. E. *Pure and applied chemistry*, **1998**, *70*(1), 105-110.
77. Pandey, V. N.; Upadhyay, A.; Chaubey, B. *Expert opinion on biological therapy*, **2009**, *9*(8), 975-989.
78. Duprey, J. L. H.; Tucker, J. H. *Chemistry Letters,* **2014**, *43*(2), 157-163.
79. Pike A R, Ryder L C, Horrocks B R, et al. *Chemistry–A European Journal*, 2002, 8(13): 2891-2899.
80. Pike A R, Ryder L C, Horrocks B R, et al. *Chemistry–A European Journal*, 2005, 11(1): 344-353.

Chapter 2 – Techniques

2.1 Oligonucleotide Synthesis

Every year, millions of oligonucleotides are synthesised globally for laboratory applications. The majority of DNA requirements for various experiments are met through oligonucleotide synthesis, typically conducted on the 40 nmol scale or lower.[1] This scale proves sufficient for most biochemical and biological studies. In contrast, biophysical investigations, such as NMR and X-ray crystallography, require larger quantities of DNA, often exceeding 10 μmol. Solid-phase methods have been developed to facilitate the synthesis of multi-kilogram quantities of oligonucleotides, particularly for applications as drug molecules, exemplified by antisense oligonucleotides. Solid-phase synthesis finds widespread utility in various fields, including peptide synthesis, oligonucleotide synthesis. The inception of solid-phase chemical synthesis dates back to the 1960s, credited to Bruce Merrifield, whose groundbreaking work in this area earned him the Nobel Prize for Chemistry in 1984. [1]

Numerous strategies for the chemical synthesis of oligonucleotides have been devised over time. The inception can be traced back to the H-phosphonate method introduced by Todd in the 1950s.[2] However, the phosphoramidite method, pioneered by Caruthers in the early 1980s[3] and subsequently refined through the incorporation of solid-phase technology and automation, has emerged as the preferred methodology.

The research pioneered by Caruthers employs a reactive 3′-O-(N,N-diisopropyl phosphoramidite) substituent attached to the 3′ hydroxyl group of individual

nucleosides. To reduce side reactions and enhance coupling efficiencies, protective measures are applied to the other functional groups, namely the nucleobases and the 5′ hydroxyl group, using cleavable entities. The dimethoxytrityl (DMT) group shields the 5′ hydroxyl group from involvement in side reactions during synthetic processes and can be readily cleaved under acidic conditions. The prevalent base-protecting groups (Figure 2.1) are routinely eliminated under alkaline conditions, typically employing a 30% ammonia solution.

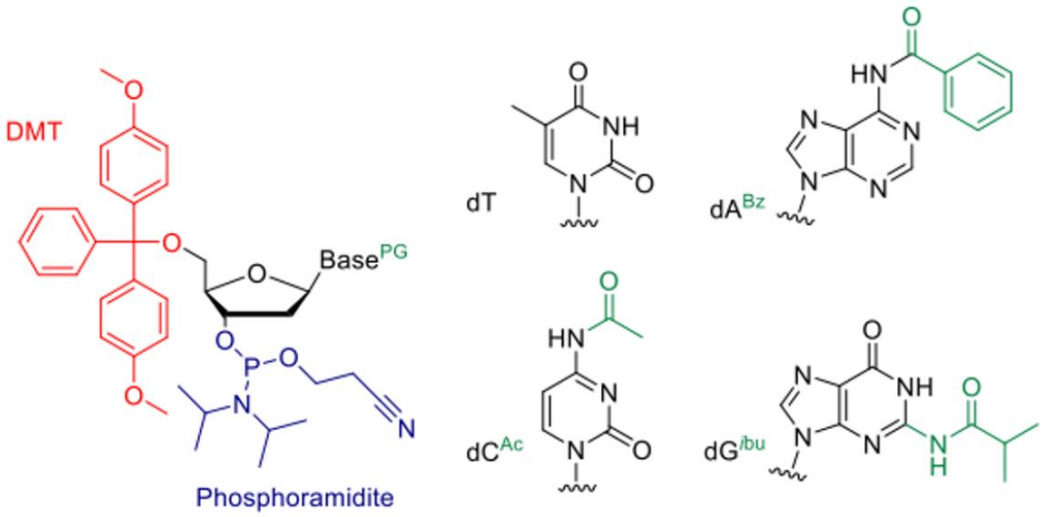


Figure 2.1 Protected nucleosides for automated DNA synthesis

2.1.1 DNA synthesis process

Automated DNA synthesis proceeds in a 3′ → 5′ direction, wherein the chain elongates from the 3′ hydroxyl group of the nucleoside. This direction contrasts with the natural biosynthesis of DNA. The synthesis cycle（Figure 2.2）encompasses five primary steps, with acetonitrile utilised for resin washing between each step. [4]

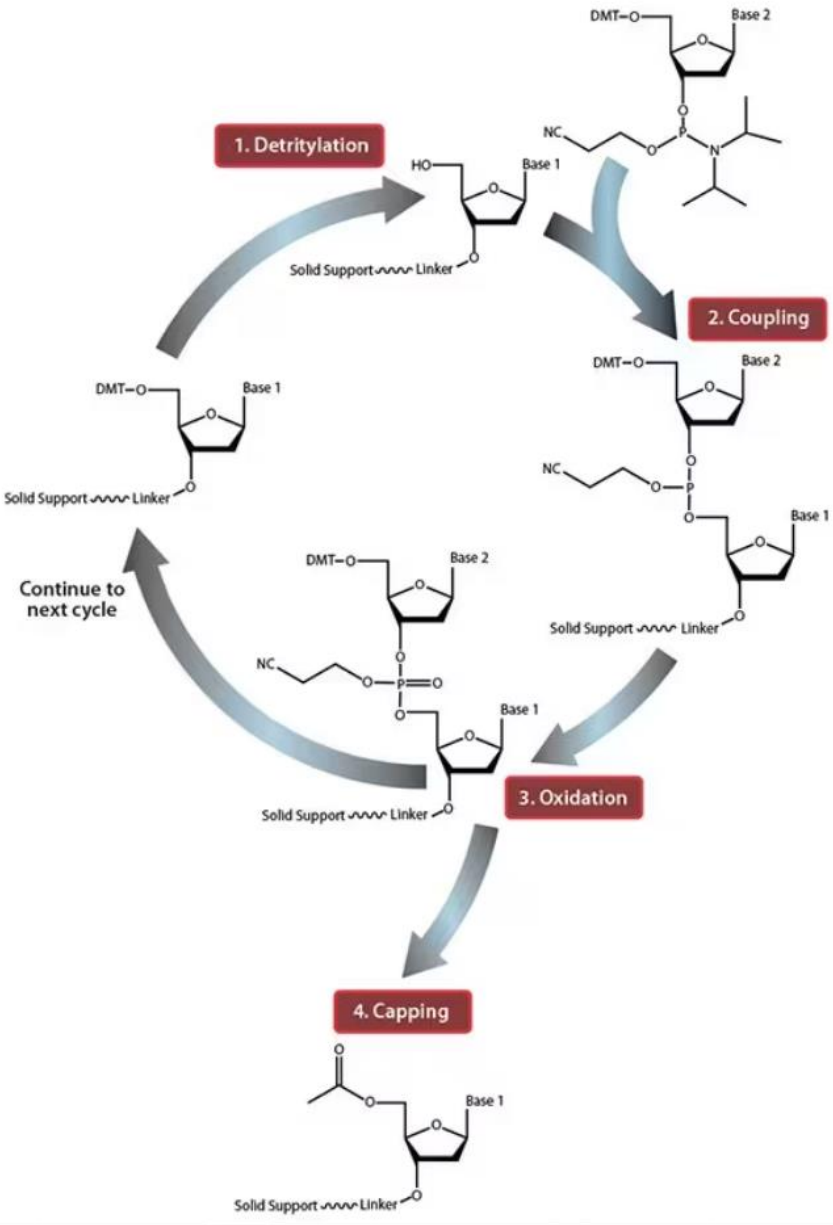


Figure 2.2 Solid-phase oligonucleotide synthesis cycle, taken from Sigma. [4]

2.1.2 Activation and Coupling

Before the initial coupling, the nucleotide attached to the solid support undergoes treatment with trichloroacetic acid to eliminate the 5′ hydroxyl-protecting DMT group, resulting in the formation of a free hydroxyl group at the 5′ terminus (see Figure 2.3). The orange-coloured trityl is employed throughout the synthesis for monitoring coupling efficiencies at each step with UV-vis absorbance. [5]

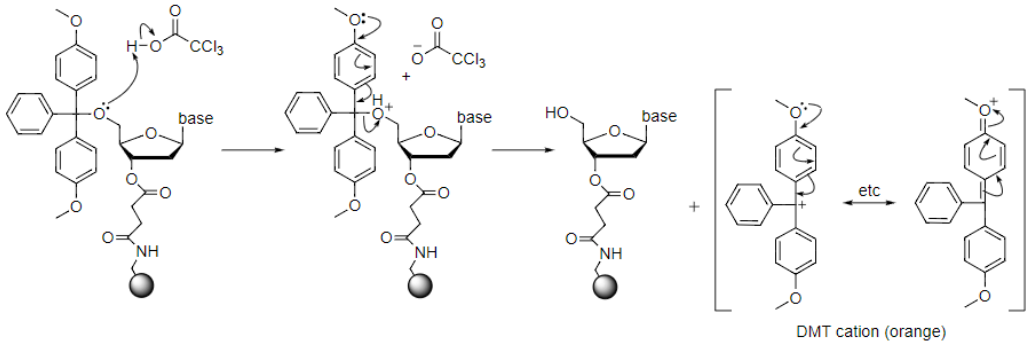


Figure 2.3 Mechanism of acid-catalysed detritylation of a DMT-protected nucleoside phosphoramidite.[1]

After detritylation, the support-bound nucleotide interacts with the next nucleoside phosphoramidite, introduced in substantial excess along with an activator, typically tetrazole. These reagents are delivered in acetonitrile. The activator, by protonating the phosphoramidite, generates an effective leaving group, subsequently displaced by the reactive support-bound 5′ hydroxyl group. This hydroxyl group attacks the phosphorous atom, resulting in the formation of a phosphite triester linkage between the two nucleosides (Figure 2.4).

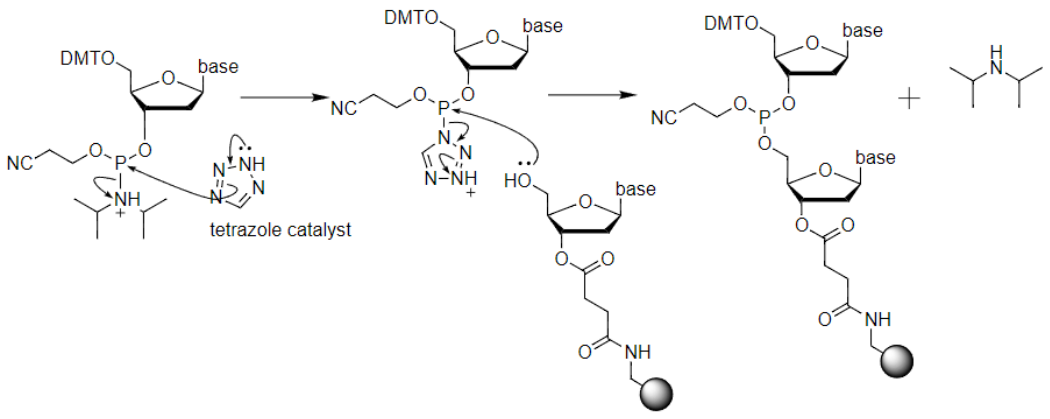


Figure 2.4 Mechanism of the phosphoramidite coupling reaction.[1]

2.1.3 Capping

While individual coupling steps frequently achieve yields of up to 99.5%, even the most efficient processes invariably incur a notable overall yield loss. To preclude the participation of uncoupled 5′ hydroxyl groups linked to the resin in subsequent coupling steps, a process known as 'capping' is implemented (see Figure 2.5). This procedural measure ensures the facile separation of the desired final oligonucleotide from unsuccessful truncated sequences. The capping reagents comprises a blend of acetic anhydride and *N*-methylimidazole dissolved in THF, with small quantities of pyridine. This reaction mixture acetylates the unbound hydroxyl groups, with pyridine serving to maintain a sufficiently basic pH, thereby preventing the cleavage of the DMT protecting group.

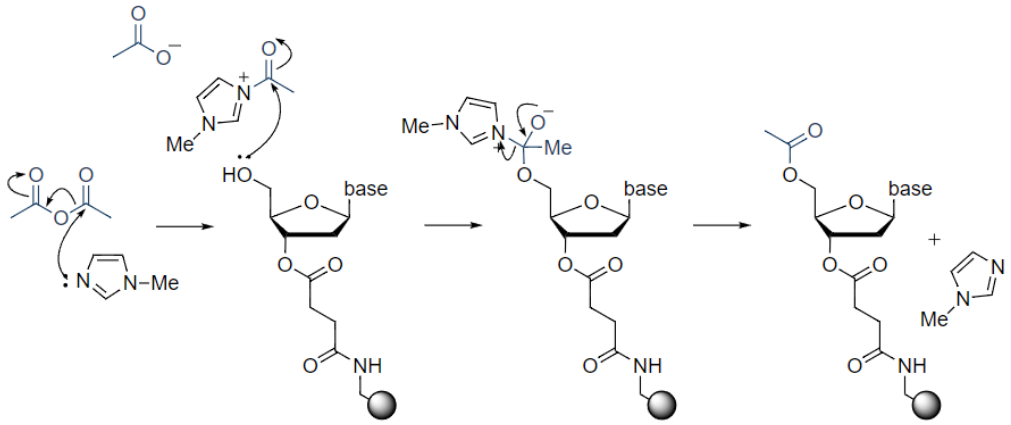


Figure 2.5 Phosphoramidite Capping Mechanism of unreacted 5′ OH groups[1]

2.1.4 Oxidation

The oxidation stage serves to transform the unstable phosphite triester produced in step 1, initially in the P(III) oxidation state, into a P(V) species (see Figure 2.6). This conversion is achieved using a solution of iodine in water and pyridine. The resultant phosphotriester essentially mirrors the configuration in the DNA backbone, but featuring a cyanoethyl protecting group.

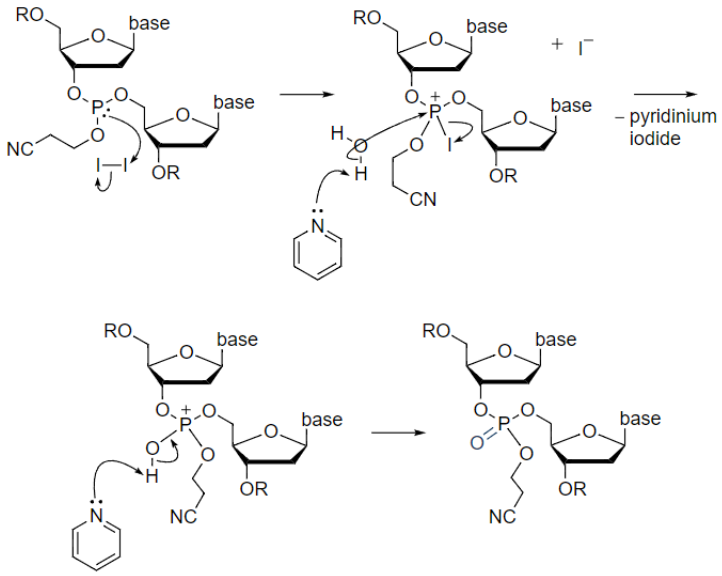


Figure 2.6 Phosphoramidite Oxidation Mechanism using iodine.[1]

2.1.5 Detritylation

The coupled nucleosides are prepared for extension. To facilitate this, the removal of the DMT group on the newly coupled nucleoside is necessary, allowing the next incoming phosphoramidite to engage in a reaction. This procedural step mirrors the one conducted prior to the initial activation and coupling.

2.1.6 Cleavage

The sequence is iteratively reproduced by adding one nucleoside at a time until the desired oligonucleotide is achieved. Upon the addition of the final nucleoside, the last DMT group is eliminated using trichloroacetic acid (TCA). Subsequently, a concentrated aqueous ammonia solution is introduced to liberate the strands from the resin and eliminate cyanoethyl protecting groups (Figure 2.7). The ammonia solution is then heated to 60 °C for 6 h to remove acyl protecting groups from the nucleobases. Finally, the solvent is removed, leaving the targeted DNA sequence, alongside short capped strands and other impurities. The resulting mixture is then purified to isolate the desired product. [6]

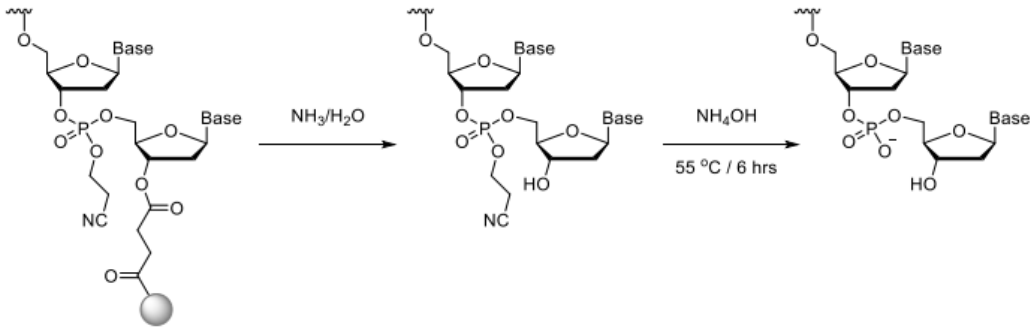


Figure 2.7 Cleavage and deprotection of the oligonucleotide.

2.2 Oligonucleotide purification

Following solid-phase DNA synthesis, the crude product comprises a mixture of truncated sequences, byproducts from deprotection steps and various impurities, alongside the desired DNA sequences. The two primary methodologies frequently employed for isolating the desired strands are high-performance liquid chromatography (HPLC) and polyacrylamide gel electrophoresis (PAGE). HPLC proves suitable for purifying strands of up to 50 bases, ensuring high sample recovery. Conversely, PAGE purification is more effective for longer sequences, although it often results in a low yield.[7]

2.2.1 High performance liquid chromatography (HPLC)

HPLC accomplishes the separation of mixtures by selectively adsorbing each component onto a stationary phase of adsorbent material. Subsequently, individual components are desorbed from the stationary phase and eluted by adjusting solvent (mobile phase) conditions. A typical HPLC system includes a sample injector, an interchangeable column (stationary phase), a solvent delivery pump, a solvent mixing system for gradient elutions and a detector for analysis (UV/vis, fluorescent, refractive index), all interconnected by a computer. Reversed-phase HPLC (RP-HPLC) is applied for the purification of chemically synthesized DNA, separating components in a mixture based on their hydrophobicity. The mixture traverses a hydrophobic stationary phase, commonly C18, and components are eluted with a polar solvent such as acetonitrile, methanol or water. Longer DNA strands, with more hydrophobic nucleobases, exhibit increased retention on the column and consequently elute later than shorter

strands.[8][9]

2.2.2 Polyacrylamide gel electrophoresis (PAGE)

Polyacrylamide gel electrophoresis (PAGE) segregates the targeted DNA product from unsuccessful sequences based on their electrophoretic mobility, a property determined by the molecule's mass, size, conformation, and charge. The hydrogels employed are composed of acrylamide and bisacrylamide crosslinkers, typically containing 5-25% acrylamide content. Additionally, a buffer, commonly tris-borate-EDTA, is integrated to stabilize the pH of the gel.

Gels can be used in either native conditions to preserve DNA conformations and interactions or denaturing conditions with the introduction of a denaturant like urea, disrupting all non-covalent interactions. Wells are created in the gel for sample loading, and an electrical potential gradient is applied to induce the migration of oligos through the gel. The gel matrix offers resistance to strand movement, thereby improving separation. Notably, shorter strands tend to exhibit higher electrophoretic mobilities, enabling them to traverse the gel more swiftly. Individual strands are identified by staining so that they are visualized as distinct bands within the gel. The band corresponding to the desired strand is precisely excised and the DNA is subsequently extracted for further analysis.[10]

2.3 Oligonucleotide Characterisation by Mass Spectrometry

Mass spectrometry is widely recognised as a highly suitable and efficient method for characterising oligonucleotides and other biomolecules. This sophisticated technique analyses ions based on their mass-to-charge ratios (*m/z*) and involves three integral components: an ionisation source, a mass analyser, and an ion detector, all controlled by a computer. The ionisation source plays a crucial role and employs techniques such as electrospray ionisation (ESI) or matrix-assisted laser desorption/ionisation (MALDI). ESI generates ions by spraying a sample solution through a high-voltage field, while MALDI uses laser pulses on the sample surface to induce ionisation. The diversity of mass analysers, including time-of-flight (TOF), ion traps, and quadrupoles, each with unique principles, contributes to the precision of the analysis. The mass analyser distinguishes ions based on their flight time or other characteristics. Various ion detectors, such as electron multipliers and ion multipliers, play a pivotal role in capturing and amplifying the ion signals. Despite its destructive nature, this technique requires minimal material (1 mg) for analysis.[11]

The commonly employed ionisation method for oligonucleotide analysis is electrospray ionization (ESI). In negative mode ESI, the sample is dissolved in a volatile solvent and introduced into the instrument as a fine spray under the influence of a strong electric field. Ionisation occurs through deprotonation, and the resulting droplets are conveyed to a high vacuum region for solvent removal. This process may lead to multiple

deprotonations within the same strand, generating a variety of species with multiple negative charges.

The raw data generated through ESI does not display the parent ion directly. Instead, numerous signals emerge corresponding to strands with various multiple negative charges. To calculate the mass of the parent ion, the measured signal is multiplied by the charge of the species, which is determined from the inverse of the spacing between the signals in a cluster, and the magnitude of the charges is subsequently added to that value (accounting for protons removed in ionisation). Software applies the calculation to the raw data, yielding a spectrum that highlights the parent ion peak.

2.4 Electrochemistry

Electrochemistry is the study of the impact of electrical outputs on a chemical process or the reciprocal relationship. An electrochemically active or redox-active species is capable of undergoing reduction (electron gain) or oxidation (electron loss) under specific conditions, either by applying a specific voltage or through a chemical reaction. The flow of electrons, known as current, can be quantified and utilized as electrical energy or analysed to provide information on mechanisms or quantities of analyte.[12] Ferrocene is a commonly used analyte as it renowned for its precisely defined electrochemistry, displaying a reversible, single-electron redox conversion leading to the formation of the positively charged ferrocenium(III) ion.

2.4.1 Self-Assembled Monolayers (SAMs) on Electrodes

This thesis primarily employs electrochemical investigation of surface-bound species. The introduction of a disulfide group at the 3′ end of the DNA probe enables the anchoring of the species of interest onto the surface of a gold electrode in a predictable and well-established manner through the formation of a self-assembled monolayer (SAM). In forming a SAM, the DNA strand with its disulfide attachment first undergoes exposure to tris(2-carboxyethyl)phosphine (TCEP), selectively reducing S-S bonds to thiols. As the free thiol is unstable, keeping the end group in its oxidised (disulfide) form until it is required allows for the safe storage of the probe in a solution. Although the exact nature of alkanethiol chemistry on gold is still a subject of ongoing research and discussion,[15] it is clear that DNA-containing SAMs continue to be highly effective for sensing and other nanotechnological applications.[13]

2.4.2 Cyclic Voltammetry

Redox-active analytes are commonly investigated using cyclic voltammetry (CV), where an applied potential is swept cyclically between limits at a uniform rate, and the resulting current from the electro-active species is recorded. This technique stands out as one of the most widely employed techniques for electrochemical analysis, primarily due to the wealth of information that can be obtained using relatively inexpensive equipment and straightforward, rapid procedures.

A commonly employed setup for CV involves a three-electrode system, comprising a working electrode (W.E.), a counter electrode (C.E.), and a reference electrode (R.E.). (Figure 2.8) These three electrodes are interconnected via a potentiostat, serving the dual purpose of controlling the potential between the W.E. and R.E. and monitoring the resulting current. A potential difference (voltage) is applied between the W.E. and the R.E. To avoid changes to the composition of the R.E., a high impedance in the instrument ensures no current directly flows between these two electrodes. The current instead passes between the W.E. and the C.E., and its variation is recorded as a function of the applied voltage.

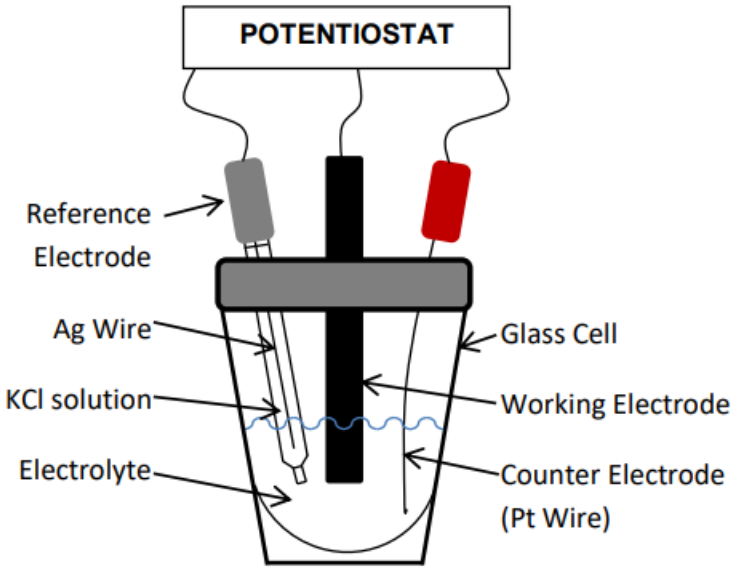


Figure 2.8 A three electrode cell set-up used in cyclic voltammetry.[9]

The potential undergoes a linear scan within a defined range and is then reversed to complete a cycle (Figure 2.9). The resulting current is plotted against the applied potential, revealing peaks corresponding to the oxidation and reduction of the redox-active species under examination.[14-15]

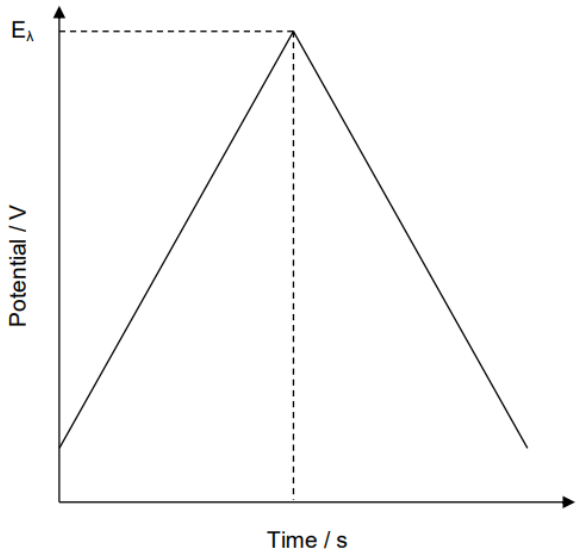


Figure 2.9 Cyclic potential sweep showing *E vs* time.

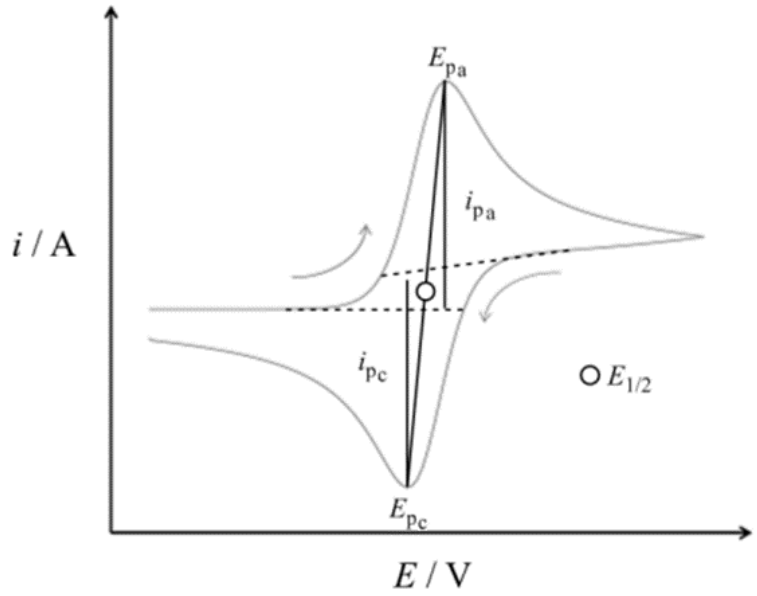


Figure 2.10 A typical cyclic voltammogram of a reversible redox process.

A voltammogram such as the one shown in Figure 2.10 can provide substantial information over the electrochemical properties of the analyte. The anodic peak maximum enables the determination of anodic peak potential (E_{Pa}) and peak current (I_{Pa}), while the reduction or cathodic peak minimum provides the cathodic peak potential(E_{Pc}) and cathodic peak current(I_{Pc}).

$E_{1/2}$, denoting the midpoint between the anodic and cathodic peak potentials in relation to the reference electrode, is a valuable parameter for the comparative analysis of electrochemical behaviours of different species.

Equation 2.1

$$E_{½} = ½ (E_{pa} + E_{pc})$$

In a fully reversible system, the number of electrons oxidized must be equal to the number of electrons being reduced. Therefore, for a reversible system with similar

diffusivities of reacting species, the ratio of I_{Pa} to I_{Pc} should be approximately equal to 1.[16]

While the data acquisition method remains consistent, the configuration of cyclic voltammograms differs notably depending on whether the analyte is present in a solution or bound to the electrode surface, for example in a SAM. In solution-based electrochemistry, as the analyte is oxidised or reduced, its surface concentration is depleted and it must be replaced *via* diffusion. The balance between increasing rate of reaction (which increases the current) and the time-dependent change in the concentration profile results in a peak in the current. For a fully reversible system, the resulting peak separation for a one-electron process is typically around 59 mV, called Nernstian behaviour.

Investigating the current response (I_{Pa}) to changes in the scan rate, which represents the time taken to scan from one potential to another, allows the determination of information about the diffusion rates of the species of interest. This is achieved using the Randles-Sevcik equation (Equation 2).[17-18] For a reversible electron transfer process in solution, the peak position remains constant with the scan rate, and the relationship of the current is proportional to the square root of the scan rate.

Equation 2.2

$$I_P = 2.69 \times 10^5 A\ n^{3/2}\ C\ D^{1/2}\ v^{1/2}$$

Where:

I_p = peak current (A)

n = number of electrons transferred per molecule

A = surface area of electrode (cm^2)

D = diffusion coefficient of solution species ($cm^2\ s^{-1}$)

C = bulk concentration ($mol\ dm^{-3}$)

v = scan rate ($V\ s^{-1}$)

Conversely, electrochemical analysis of surface-bound redox-active species is no longer dependent on diffusion for electron transfer. As a result, anodic and cathodic peaks should exhibit a peak separation very close to 0 mV. The current should exhibit a linear increase with the scan rate (not the square root of the scan rate), as indicated by the following relationship:

Equation 2.3

$$I_p = \frac{n^2F^2}{4RT} v A_{sur} \Gamma$$

Where:

I_p = peak current (A)

n = no of electrons transferred per molecule in redox process

F = Faraday's constant

R = gas constant

T = temperature (often assumed to be 25 °C)

v = sweep rate ($V\ s^{-1}$)

A_{sur} = microscopic electrode surface area (cm^2)

Γ = surface coverage

The total charge obtained from under the peak in a cyclic voltammogram of a surface-bound species can be used to obtain the surface coverage of that species.

References

1. Brown, T.; Brown, T. Jr. *Nucleic Acids Book*. https://atdbio.com/nucleic-acids-book
2. Michelson, A. M.; Todd, A. R. *Journal of the Chemical Society*, **1955**, 2632-2638.
3. Beaucage, S. L.; Caruthers, M. H. *Tetrahedron Letters*, **1981**, *22*, 1859-1862.
4. DNA Oligonucleotide Synthesis. https://www.sigmaaldrich.com/GB/en/technical-documents/technical-article/genomics/pcr/dna-oligonucleotide-synthesis
5. Berner, S.; Mühlegger, K.; Seliger, H. *Nucleic Acids Research*, **1989**, *17*(3), 853-864.
6. Carr-Smith, J. *PhD thesis*, University of Birmingham, **2015**.
7. Allabush,F., *PhD Thesis*, University of Birmingham, **2018**.
8. Method of Oligonucleotide Purification. Biosyn.com.
9. Holly, R. *PhD thesis*, University of Birmingham, **2018**.
10. A Brief Overview and History of Gel Electrophoresis. *Nucleic Acid Gel Electrophoresis—Overview and History | Thermo Fisher Scientific – UK*.
11. Fleming, I.; Williams, D. H. *Spectroscopic Methods in Organic Chemistry*; McGraw-Hill: New York, **1966**; p 56.
12. Bard, A. J.; Faulkner, L. R.; White, H. S. *Electrochemical Methods: Fundamentals and Applications*; John Wiley & Sons, **2022**.
13. Fan, C.; Plaxco, K. W.; Heeger, A. J. *Proceedings of the National Academy of Sciences*, **2003**, *100*(16), 9134-9137.
14. Eckermann, A. L.; Feld, D. J.; Shaw, J. A.; Meade, T. J. *Coordination Chemistry*

Reviews, **2010**, *254*(15-16), 1769-1802.

15. Inkpen, M. S.; Liu, Z. F.; Li, H.; Campos, L. M.; Neaton, J. B.; Venkataraman, L. *Nature Chemistry*, **2019**, *11*(4), 351-358.

Chapter 3 – Ferrocene Nucleic Acid (FcNA) Monomer Synthesis

3.1 Introduction

3.1.1 Structure of Ferrocene

Initially called "yellow sludge" by Union Carbide workers when found in pipe blockages,[1] dicyclopentadienyliron (ferrocene) soon became a subject of fascination in organometallic chemistry due to its unique properties. Pauson accidentally discovered some orange crystals while working with $FeCl_3$ and a cyclopentadienyl Grignard reagent.[2] These crystals were remarkably stable, dissolving in concentrated sulfuric acid without decomposition, and were stable in water and bases but decomposed above 300 °C. Pauson and Kealy attributed this stability to the cyclopentadienyl group's tendency to become "aromatic" through the acquisition of a negative charge, which could be evenly distributed through resonance.[2] After that, Miller independently reproduced this serendipitous discovery, confirming the compound's molecular formula as $C_{10}H_{10}Fe$. This marked the beginning of extensive research on ferrocene.[3]

Wilkinson proposed the 'sandwich' structure of ferrocene on the basis of X-ray analysis, the compound's diamagnetic properties, the absence of a dipole moment, and a unique C–H absorption frequency. Despite Wilkinson illustrating the molecule in staggered and eclipsed conformations (see Fig.3.1), subsequent studies demonstrated facile ring rotation. [4-5]

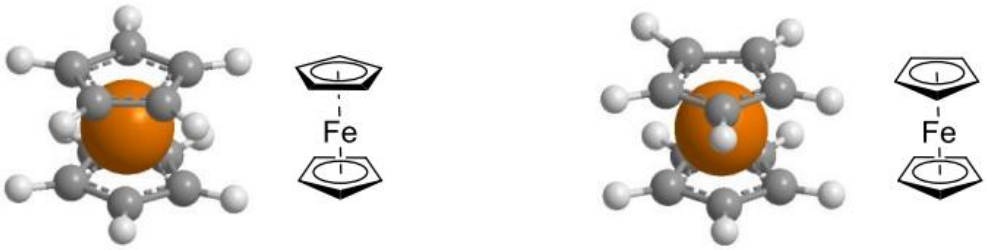


Figure 3.1 Structure of ferrocene showing its eclipsed and staggered conformations. [6]

The planar cyclopentadienyl rings exhibit anionic character, each possessing 6 π-electrons and intrinsic aromaticity. These symmetric plates 'sandwich' the central Fe^{2+} ion, ensuring a net neutral charge. The 6 electrons from each cyclopentadienyl ring combine with the 6 *d*-electrons of the central metal ion, establishing a stable 18-electron configuration. This type of structure is also found in other organometallic transition metal complexes, which collectively constitute a family known as metallocenes.[7]

3.1.2 Functionalisation of Ferrocene

With two negatively charged aromatic rings, ferrocene displays chemical reactivity akin to an electron-rich benzene, providing insight into the origin of its name. This makes it amenable to electrophilic substitution reactions, such as Friedel-Crafts acylation/alkylation, facilitating effective functionalisation.[8-9] Significantly, the reaction of ferrocene with butyllithium reagents (BuLi) allows for deprotonation, yielding a metallated pseudo-carbanion. This versatile species can then interact with various electrophiles, resulting in a diverse array of functionalized products. Notably, while *t*-BuLi typically favours the mono-lithiated product, *n*-BuLi in conjunction with TMEDA tends to promote the formation of di-lithiated species, with one lithium on each

cyclopentadienyl (Cp) ring. This property enhances the synthetic versatility of ferrocene, enabling a broader spectrum of tailored modifications and applications.[10-11]

Ferrocene has been widely explored in DNA research, with a focus on electrochemical DNA sensing, as mentioned before. The Tucker group has contributed to this field by creating a Ferrocene Nucleic Acid (FcNA) oligomer as an structural mimic of DNA, comprising eight ferrocene units and sixteen thymine nucleobases (Figure 3.2). This involved the initial synthesis of a tetra-substituted ferrocene monomer with two thymine bases attached to it, followed by oligomer formation through solid-phase synthesis using phosphoramidite chemistry.[12]

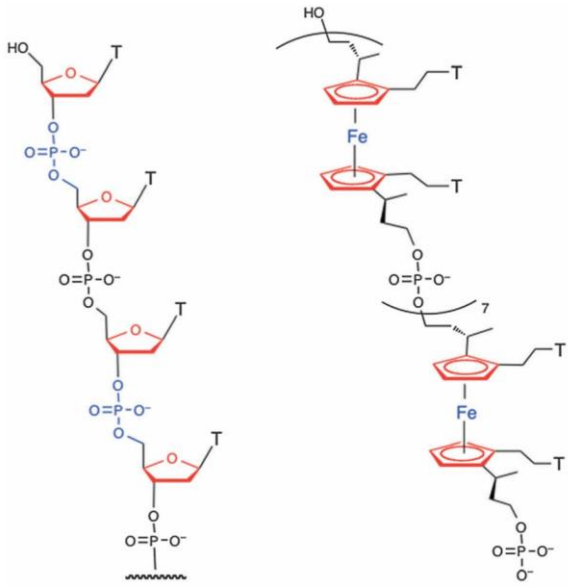


Figure 3.2 Structure of an 8-mer strand of FcNA, and its comparison with DNA.[12]

3.2 Aim

In research on FcNA by the Tucker group, achievements related to an oligomer containing the nucleobase thymine were successfully published [12], but in-depth investigations into its properties have yet to be completed. This study aims to synthesise more ferrocene monomers, but not only the one incorporating two thymine (**FcTT**) but also one incorporating two cytosines (**FcCC**). Although there has been some previous research on the synthesis of **FcCC** within the group [15], the overall yield was too low to be suitable for detailed studies on the resulting oligomers. In this chapter, the synthesis of phosphoramidite monomers of both **FcTT** and **FcCC** is described, which includes a focus on improving the efficiency of their synthesis to maximise the chances of obtaining sufficient amounts of product for studying the resulting oligomers using a variety of techniques.

The synthetic route towards both FcNA target molecules involves several steps. In the following sections, these are separated into various stages, as outlined in Figure 3.3. The first three stages are common to both **FcTT** and **FcCC**. The synthesis of a control monomer, **FcHH** is also described in this chapter.

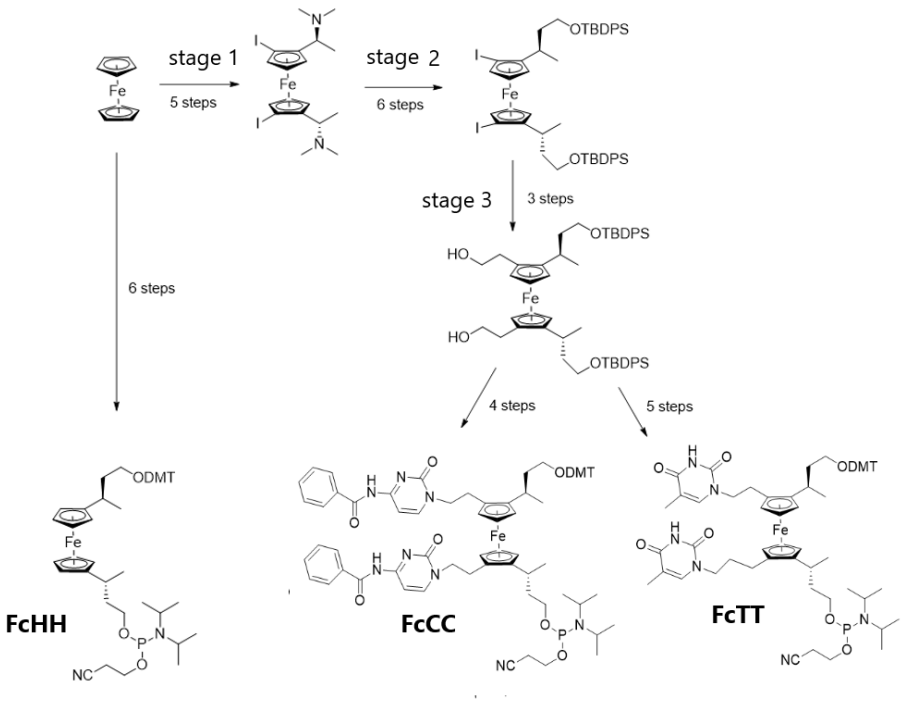
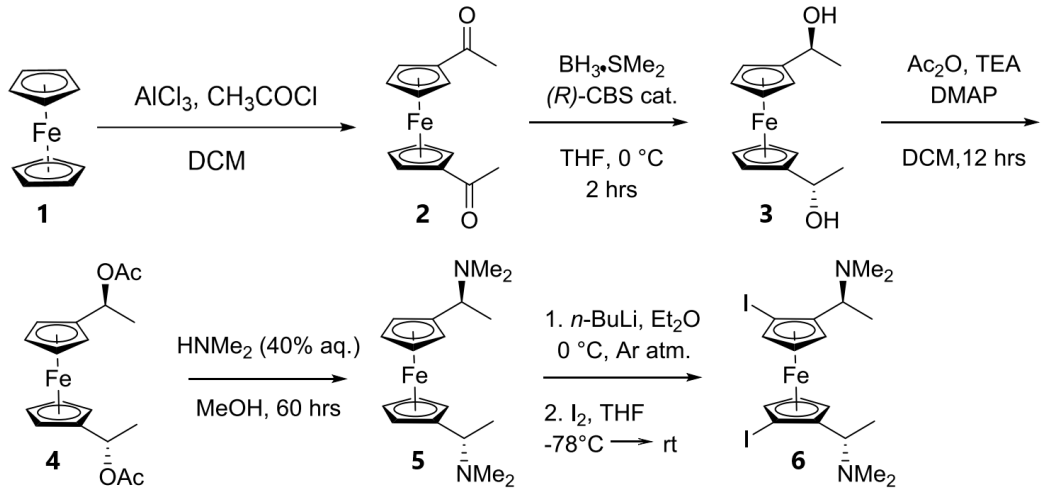


Figure 3.3 The three target phosphoramidite monomers, **FcHH**, **FcTT** and **FcCC** (all with *R*,*R* stereochemistry and for FcTT and FcCC with S_p,S_p stereochemistry also).

3.3 Results and Discussion

3.3.1 Synthesis of Tetrasubstituted Ferrocene Monomers

3.3.1.1 Stage 1 - Tetrasubstituted bis-iodo ferrocene (compound 6)



Scheme 3.1 Synthesis route to compound **6** from ferrocene **1** (Stage 1).

Control over stereochemistry is crucial in the synthesis of FcNA monomers, as the target molecules contain not only two chiral centres in the form of two asymmetric carbons attached to the Cp rings (here called point chirality), but also two planar chiral centres. Planar chirality is typically elucidated by applying the Cahn-Ingold-Prelog rules to assign priorities. When observing the plane of interest from above, (R_p) denotes the enantiomer with a priority increase in a clockwise direction, while (S_p) designates the opposite enantiomer with an anticlockwise priority arrangement (Fig. 3.4).[13]

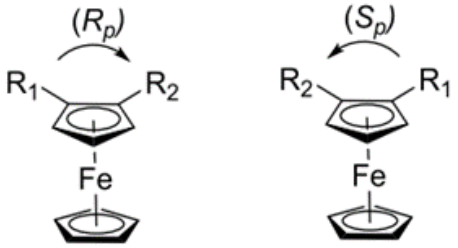


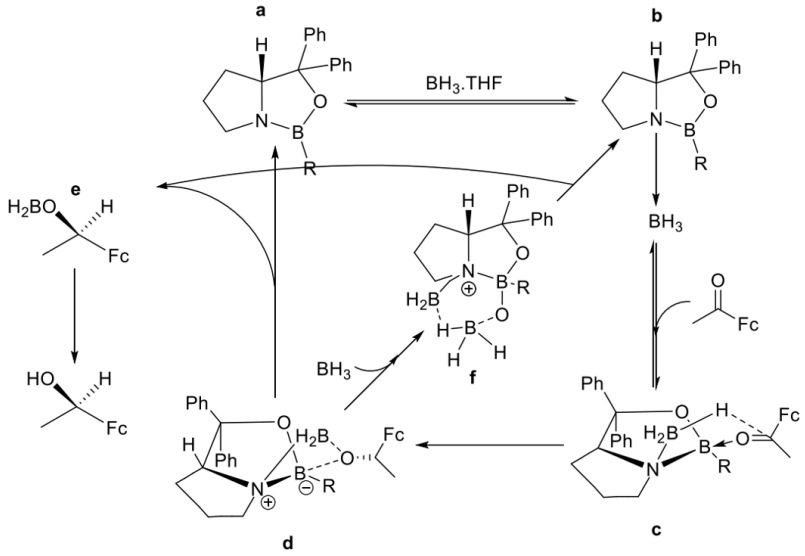
Figure 3.4 Planar chiral designations (R_p) and (S_p) of 1,2-disubstituted ferrocenes.[6]

Synthesis of Compound 2

The synthetic pathway to compound **2** commences with the acylation of ferrocene via a Friedel-Crafts reaction, a procedure designed to introduce electrophiles to aromatic rings.[14] This reaction was carried out with aluminium chloride and acetyl chloride in DMC, to yield 1,1'-diacetylferrocene **2** in a yield of 85%. The mechanism begins with the generation of an acylium ion resulting from the interaction between acetyl chloride and aluminum chloride, where aluminum chloride serves as a Lewis acid. Following this, ferrocene interacts with the acylium ion, initiating an attack by ferrocene on the electrophilic species. This interaction leads to the formation of bis-ketone ferrocene and the release of HCl. The singlet peak observed at 2.36 ppm in the ^{1}H NMR spectrum indicated the presence of two methyl groups in the compound, confirming that the desired product has been successfully synthesized (see experimental chapter for more details).

Synthesis of Compound 3

Strong reducing agents such as $LiAlH_4$ or $NaBH_4$ are typically used to reduce ketones to alcohols. Non-symmetrical ketones such as compound **2** are prochiral, leading to two possible enantiomers (*R* or *S*) depending on which face of the carbonyl group is attacked. This means that without any stereochemical control over the reaction, a racemic product is produced as there is no differentiation between the two faces. Recent advances in chiral catalysis have enabled the production of chiral ferrocenes with exceptionally high enantioselectivity and relative ease. Knochel, for instance, employed the Corey-Bakshi-Shibata catalyst to induce chirality, conducting a Corey-Bakshi-Shibata (CBS) reduction of diacetylferrocene to yield the diol with a decent yield and an enantiomeric excess of 97%.[16-17] Advances in this method further enhanced the enantioselectivity of the reduction, the mechanism as shown in Scheme 3.2. The doublet peak at 1.4 ppm in the ^1H NMR spectrum of compound **3** integrating to six protons indicated the presence of two CH_3 groups adjacent to a CH group, which itself gave rise to quartet at 4.65 ppm.



Scheme 3.2 Mechanism of CBS reduction.[16]

Synthesis of Compound 5

In seeking control of the planar chirality in ferrocene chemistry, diverse stereoselective methods have been devised to aid ortho lithiation on the cyclopentadienyl (Cp) rings of ferrocene. A notable success in this endeavour emerged in the early 1970s through the work of Ugi.[16] The chiral amine he utilized, [1-(dimethylamino)ethyl]ferrocene, to guide lithiation in a diastereoselective manner is now recognized by his name due to its ready availability and effectiveness in directing this reaction.

The underlying rationale is as follows. In the lowest energy configuration, the relatively bulky methyl and dimethylamine groups associated with the 'Ugi amine' strategically position themselves with minimal steric hindrance to the iron centre. These groups extend beyond the plane of the Cp ring, residing on the outer periphery of the molecule. Consequently, the amine group exhibits a preferential spatial arrangement proximal to a single carbon atom. The subsequent complexation with the Li^+ ion serves to stabilize the metal in this specific ortho position. Knochel's investigations affirmed the equal efficacy of this approach when concurrently applied to both rings, maintaining a commendable level of diastereoselective > 98%.[16]

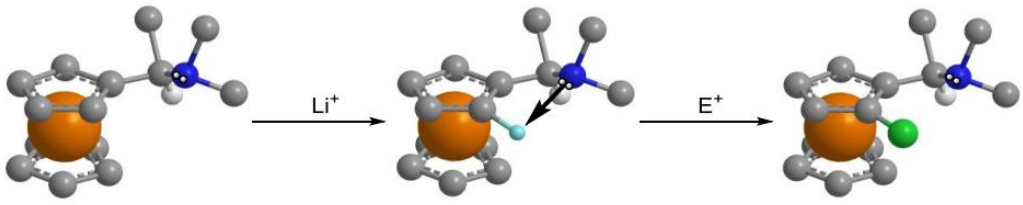


Figure 3.5 Directed ortho-lithiation using the 'Ugi amine' [(*R*)-1-(dimethylamino)ethyl]ferrocene (carbon atoms are dark grey, nitrogen atoms are blue, hydrogen atoms are white, and the Lithium atom is cyan).

The hydroxyl groups in compound **3** were subsequently transformed into acetate groups in compound **4**, (*S*,*S*)-1,1'-bis-(α-hydroxyethyl)-ferrocene, through acetylation with acetic anhydride. The singlet peak at 2.04 ppm and the doublet peak at 1.53 ppm indicated the presence of four methyl groups in two different chemical environments within the compound.

After the successful substitution with dimethylamine, resulting in the formation of difunctionalised version of the 'Ugi amine', compound **5**, in a yield of 70%. The singlet peak observed at 2.08 ppm (Fig. 3.6) indicated the presence of 12 equivalent protons belonging to two sets of two methyl groups each bonded to one nitrogen atom.

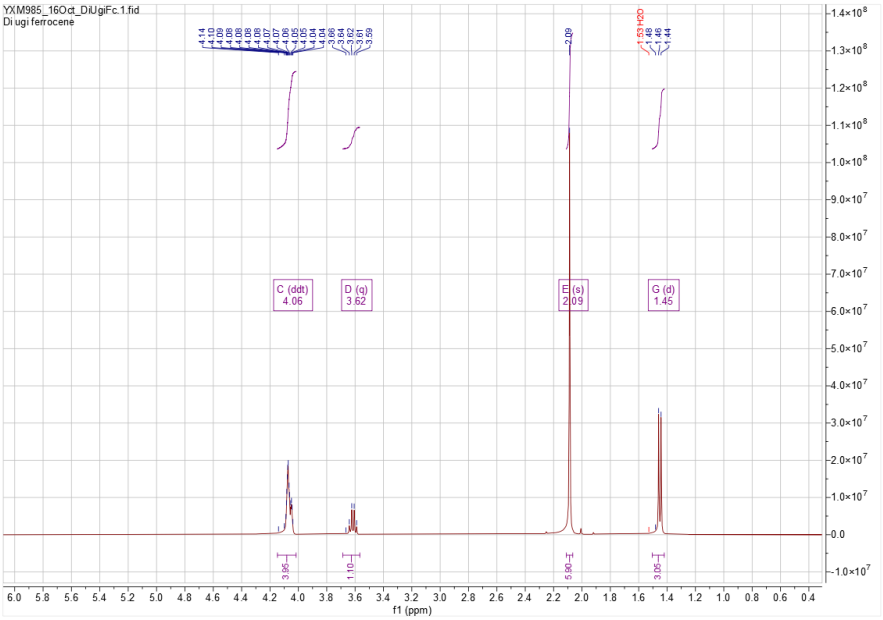


Figure 3.6 ^{1}H NMR spectrum of compound **5** in Chloroform-d.

To improve the yield and minimise the loss of the product during the purification process, especially considering the instability of the acetyl groups in the column, on repeating Stage 1, no purification steps were carried out after work-up in the process from 1,1'-diacetylferrocene (compound **2**) to the Ugi amine (compound **5**). This improved the overall yield of these three steps to 70% from 47%.

Next, the difunctionalised Ugi amine was employed to guide the diastereoselective iodination of both Cp rings on the ferrocene. An ortho-lithiation was conducted using *n*-BuLi, and the resulting carbanion was reacted with iodine to produce compound **6** in a yield of 69% following the Knochel method.[16] Compound **6** exhibits a similar ^{1}H NMR spectrum to compound **5**, with the main difference being the number of protons on the Cp rings (i.e. six rather than eight).

After column purification and characterization by NMR and mass spectrometry, the chiral purity of **6**-(*SSR*$_p$*R*$_p$) was checked against a sample of the opposite enantiomer using chiral HPLC (OD column with isocratic eluent system of 1.5% IPA in hexane). Even though the peaks were broad, the peaks of the two enantiomers were well separated (Fig. 3.7). No traces of the opposite enantiomer were observed in the sample of **6**-(*SSR*$_p$*R*$_p$).

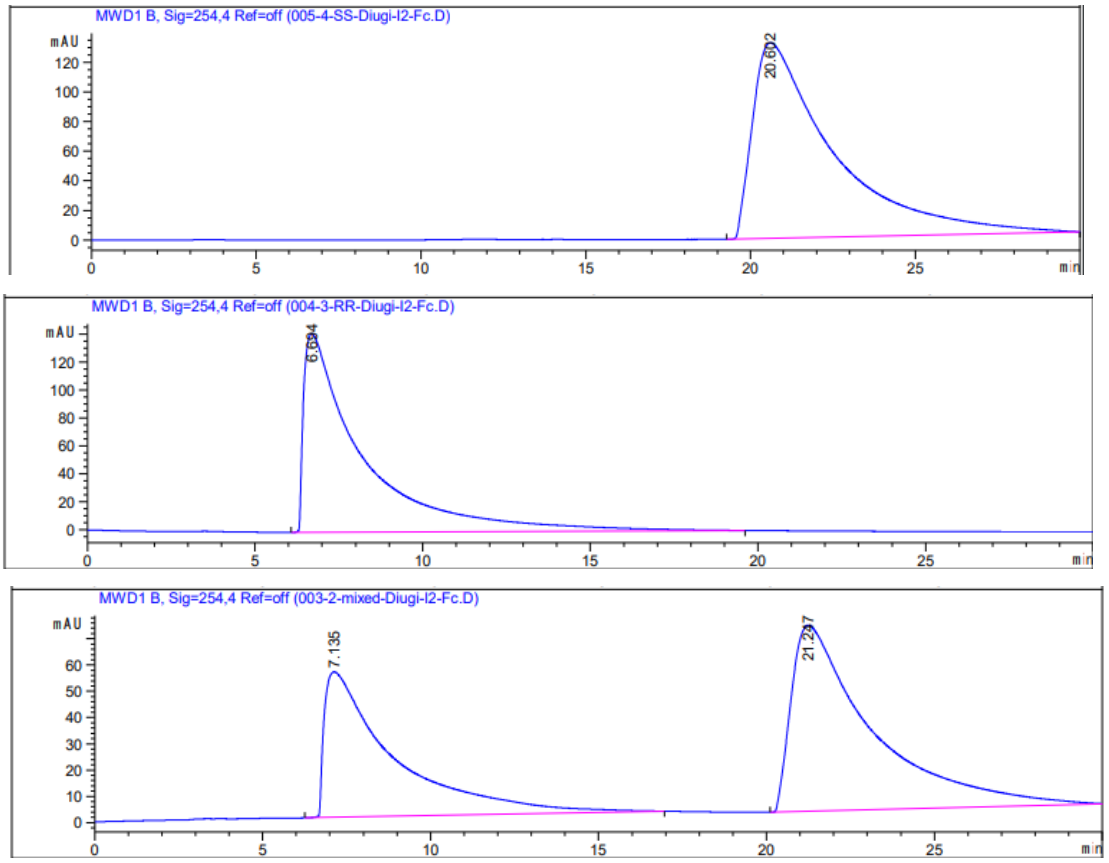
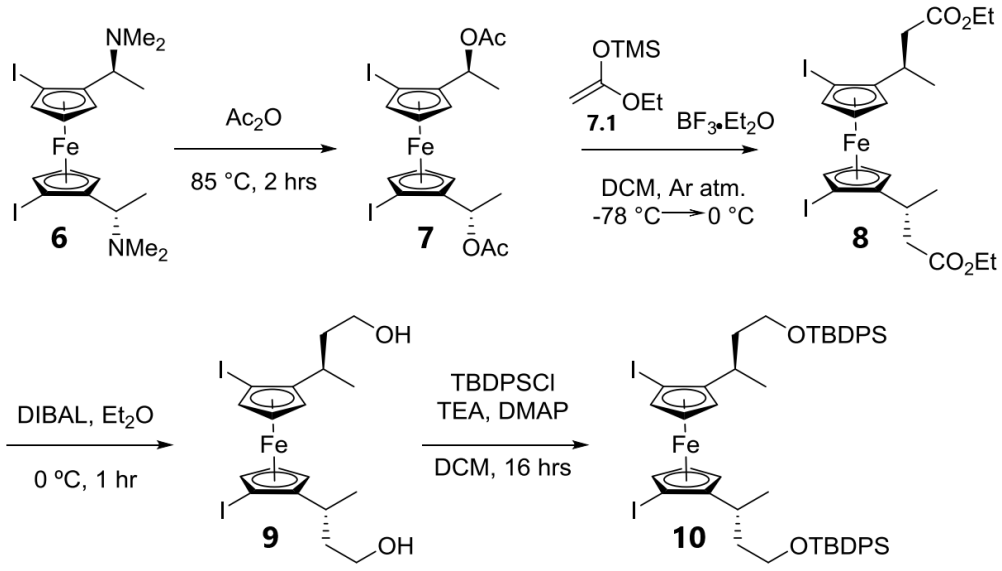


Figure 3.7 Chiral HPLC traces of compound **6**. A: **6**-(*SSR*$_p$*R*$_p$); B: **6**-(*RRS*$_p$*S*$_p$); C: Mixture of two enantiomers (OD column with isocratic eluent system of 1.5% IPA in hexane).

3.3.1.2 Stage 2: Linker extension of the 3-carbon phosphate linker

The next stage of the synthesis focused on producing the two three-carbon linkers that were to serve as the phosphodiester linkers (Scheme 3.3). This intricate process comprised four steps, starting with acetylation, followed by ester formation through a silyl enol ether, then ester reduction, and concluding with a protective step.



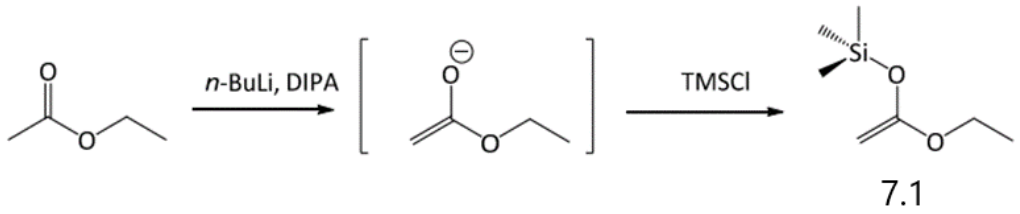
Scheme 3.3 Linker extension of tetrasubstituted ferrocene (Stage 2).

Synthesis of Compound 7

The acetylation procedure employed acid anhydride at 85 °C for 2 h. The desired **7** was achieved with analysis of the ^{1}H NMR spectrum indicating a singlet at 2.02 ppm, belong to the methyl groups of the acetate. Unfortunately, the OAc group proved to be unstable during column chromatography, necessitating purification through crystallization from DCM and hexane, then used immediately in the next step.

Synthesis of Compound 8

To form **8**, it was imperative to synthesise a crucial reagent: the silyl precursor 1-ethoxy-1-trimethylsilyloxyethylene (compound **7.1**). Typically, this precursor is obtained from a ketone using a strong base and trimethylsilyl chloride. Following the procedure outlined in Scheme 3.4,[18] which involves *n*-BuLi, diisopropylamine, trimethylsilyl chloride and ethyl acetate, the method for creation of **7.1** involves deprotonating ethyl acetate and then capturing the resulting enolate with trimethylsilyl chloride. The crude product was purified through vacuum distillation. Initially, hexane and ethyl acetate fractions were collected, followed by the silyl precursor. Separation posed a challenge due to the closely matched boiling points of the silyl precursor and ethyl acetate, but the yield (60%) could be calculated through NMR integration.



Scheme 3.4 Synthesis of the silyl precursor **7.1**.

The silyl precursor was then reacted with compound **7**. Substitution by the enolate nucleophile was employed to extend the carbon skeleton. The reaction proceeded through a nucleophilic attack of the silyl enol ether on the sp^2 hybridized carbon atom, displacing the acetate group in the presence of boron trifluoride, which acts as a Lewis acid by coordinating with the oxygen atom of the acetate group. As a consequence of substituting OAc with CH_2COOEt, this step changed the stereochemical description of

the point chirality on the two α-carbons from *SS* to *RR* due to a change in priority.

Synthesis of Compound 9

Brisset employed $LiAlH_4$ reduction on the non-iodinated version of compound 8, bis-[(2-ethyloxycarbonyl)ethyl]-ferrocene, giving the alcohol in a high yield.[19] However, $LiAlH_4$, being a strong reducing agent, not only reduces the ester but also facilitates a metal-halogen exchange reaction, which could eliminate the iodo groups. As a result, the mild reducing agent DIBAL was used. DIBAL has commonly been employed to reduce esters to aldehydes under low temperature conditions, with a standard stoichiometry of 1 equivalent of DIBAL per ester. In this case, the reaction was carried out for 6 h at room temperature, utilizing an excess of DIBAL: three equivalents per ester group.

Synthesis of Compound 10

The final step of Stage 2 involves the protection of the diol. Silyl ethers consist of a Si atom covalently bonded to an alkoxy group. Due to their inert nature towards many reagents that typically react with alcohols, silyl ethers find utility as protective groups in organic synthesis. Out of various silyl groups, TBDPS exhibits high stability in different environments and can be readily removed using TBAF. Following the procedure outlined by Hanessian in 1975, [20] the diol underwent treatment with TBDPSCl, TEA, and DMAP in DCM overnight, resulting in the formation of the TMBDPS-protected product with an 89% yield. Two multiplets in the aromatic region

were observed, integrating to 20H and corresponding to the phenyl groups. The *t*-butyl substituent was identified as a singlet at 1.05 ppm, corresponding to six CH_3 groups.

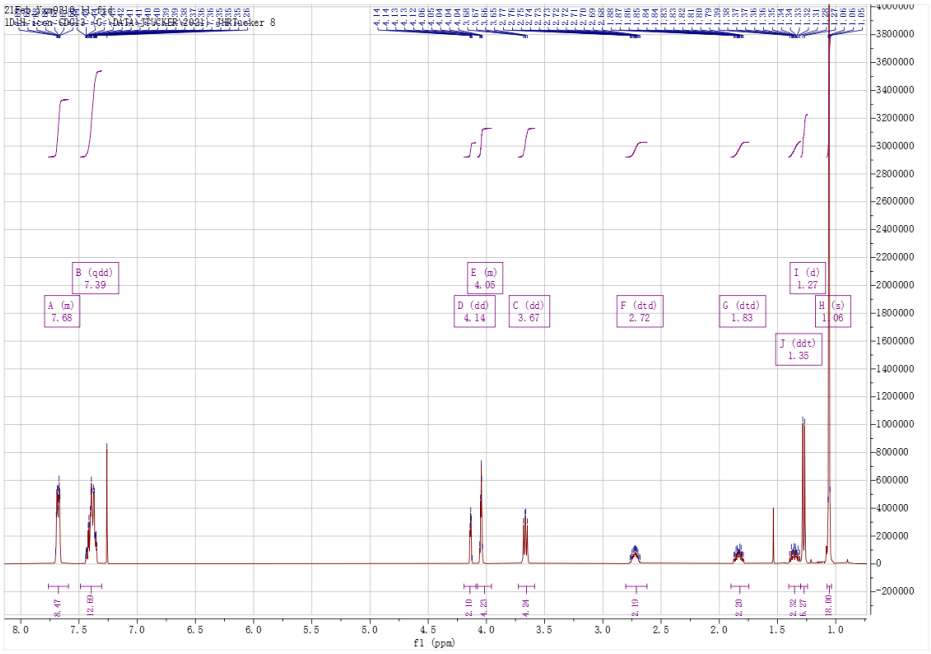
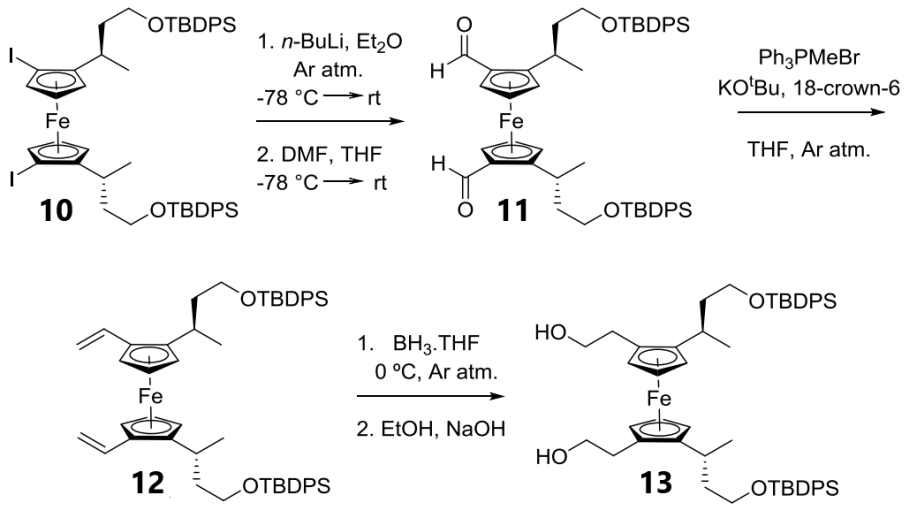


Figure 3.8 ^{1}H NMR spectrum of compound **10** in Chloroform-d.

3.3.1.3 Stage 3: Linker extension of the 2-carbon nucleobase linker

The next stage involved the synthesis on the two-carbon linkers that were to be connected to the nucleobases. This involved three steps: formylation, followed by a Wittig reaction, and then hydroboration and oxidation (Scheme 3.5).



Scheme 3.5 Continued synthesis of tetra-substituted ferrocene (Stage 3).

Synthesis of Compound 11

The initial step of this process is the formylation of the tetra-substituted diiodoferrocene **10**. This formylation process is a two-step process, commencing with a halogen-lithium exchange, analogous to the earlier-discussed directed ortho-metallation step. Subsequently, a lithium-halogen exchange occurs, resulting in the formylation. The exchange takes place in diethyl ether in the presence of *n*-BuLi. Following the formation of dilithioferrocene, DMF is then added. It was crucial to maintain dry and

low-temperature conditions throughout the process. A red color indicated the successful formation of the aldehyde; (R,R,R_p,R_p)-CHO, **compound 11**, which was obtained with a 89% yield. The two aldehyde protons appeared as one singlet at 10.05 ppm in the ^{1}H NMR spectrum.

Synthesis of Compound 12

The next phase in the reaction involves the transformation of the formyl ferrocene into a terminal alkene through Wittig olefination. The key reagent in this reaction is the phosphorous ylide, formed by deprotonating the methyl group adjacent to the PPh_3 moiety of the phosphonium salt with the tert-butoxide base. The incorporation of 18-crown-6 in this strategy streamlines the deprotonation step by coordinating with the potassium counterion. This coordination intensifies the negative charge of the tert-butoxide anion, thereby enhancing its basicity. Although the stability of the resulting phosphorous ylide could be exploited to govern stereochemistry in alkene products, in this case, where no defined stereochemistry is required for a terminal alkene, this factor is not a consideration.

The reaction exhibited good conversion according to TLC analysis; however, the yield was low after column purification. It was postulated that the ferrocene moiety experienced partial oxidation in the presence of air, as indicated by a color change from yellow to green in the column. To address this concern, the eluent system for purification underwent optimization, reduced the necessity for silica and minimized the product's exposure to air during column chromatography. This modification led to a

successful purification, with the product obtained in an excellent yield of 84% on the second attempt. The disappearance of the aldehyde protons' singlet at 10.05 ppm in the ^{1}H NMR spectrum and the appearance of three doublets of doublets at 5.08 ppm, 5.25 ppm and 6.48 ppm, representing protons on the alkene, indicated the success of the reaction.

Synthesis of Compound 13

The hydroboration and oxidation reaction of bis-divinyl ferrocene **12**, conducted in the presence of a borane tetrahydrofuran solution and hydrogen peroxide, led to the formation of bis-silyl protected bis-dihydroxyethyl ferrocene **13**. Its ^{1}H NMR spectrum is shown in the Figure 3.9. Its chiral purity was checked against a sample of the opposite enantiomer through chiral HPLC (Figure 3.10). A cellulose column with an isocratic elution system consisting of 10% H_2O in MeCN was employed, with the flow rate set at 1.0 mL/min. The synthesised compound **13-(*R*,*R*,*S*$_p$,*S*$_p$)** shows a large peak at 8.43 mins, with its opposite enantiomer having a slightly longer retention time at 9.1 mins.

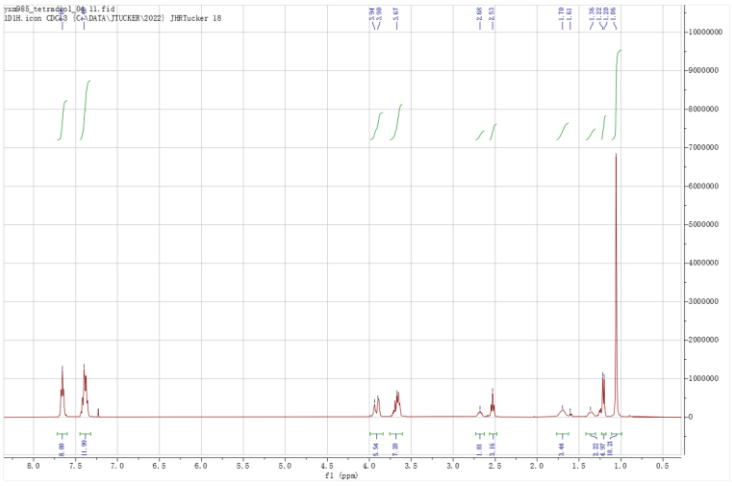
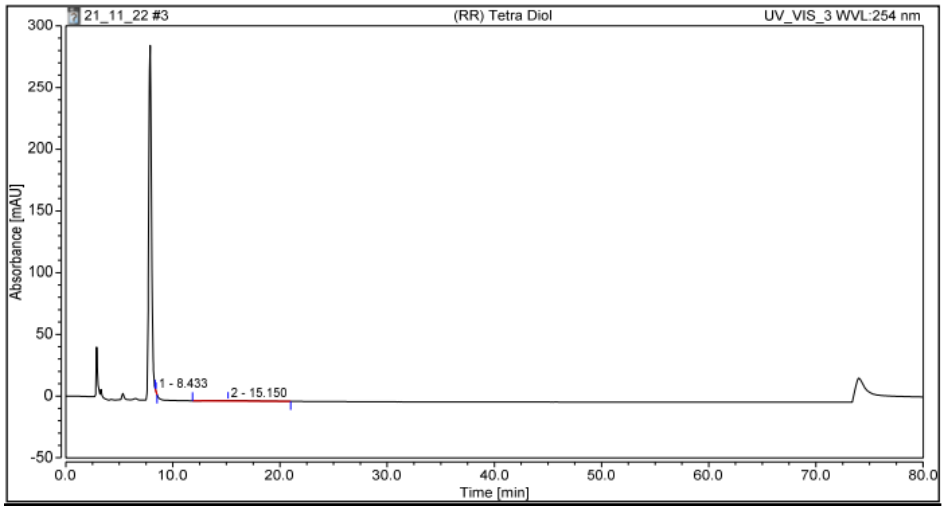
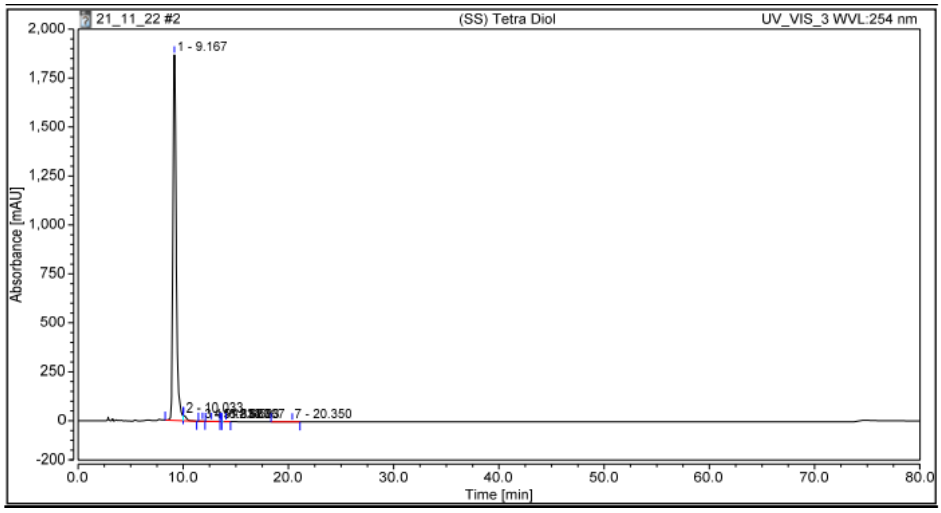


Figure 3.9 ^{1}H NMR spectrum of compound **13** in Chloroform-d.

A)



B)



C)

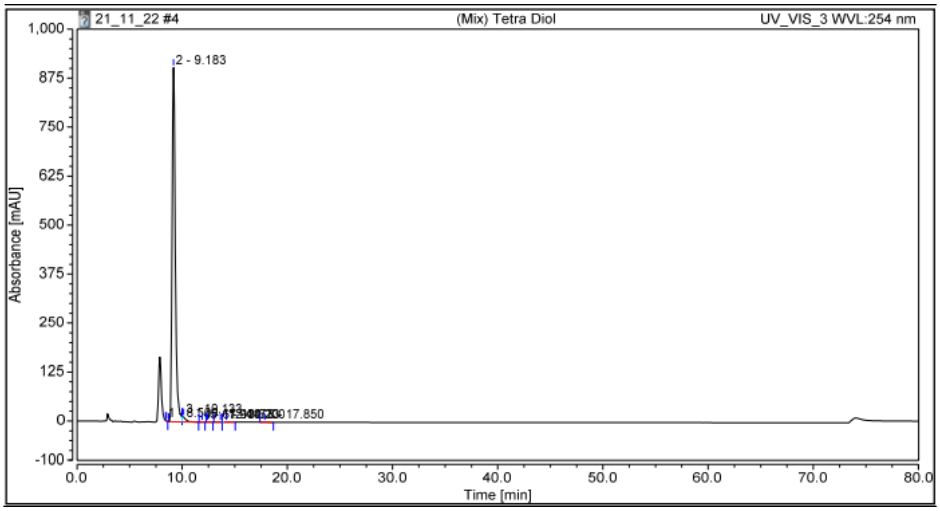
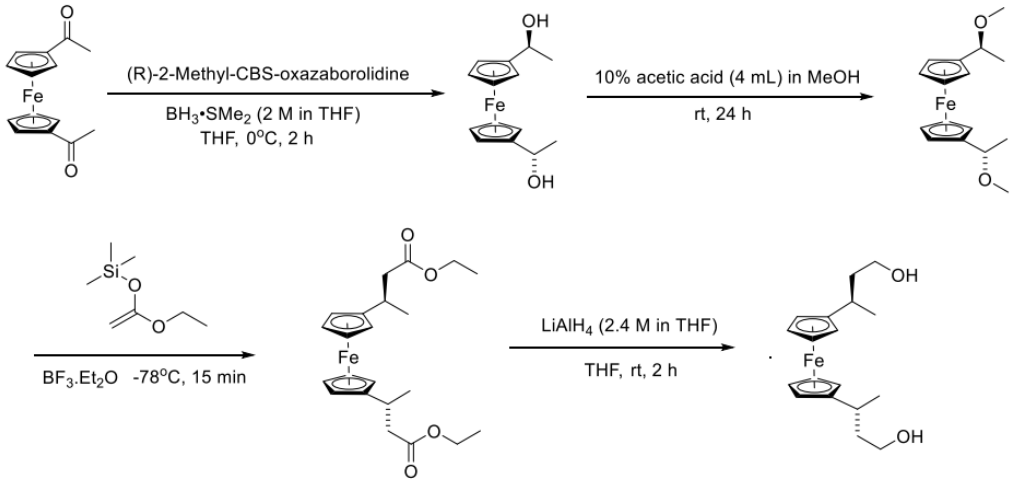


Figure 3.10 Chiral HPLC trace of tetra-diol, compound 13. A) **13-(*R*,*R*,*S*$_p$,*S*$_p$),** B) **13-(*S*,*S*,*R*$_p$,*R*$_p$)** C) Mixture of two Isomers (cellulose column with isocratic eluent system of 10% H_2O in MeCN).

3.3.1.4 Synthesis of the control compound FcHH

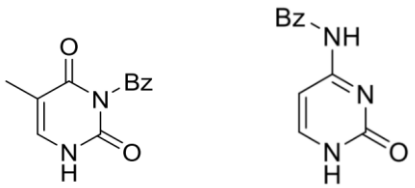
The compound **FcHH** is an excellent control compound for this study. It shares the same stereochemistry as the **FcTT**/**CC** compounds but lacks any additional bases. Ultimately, it can be incorporated into the same DNA sequences. Its synthesis route is illustrated in Scheme 3.6, which resembles Stage 2 of the tetra-substituted ferrocene monomer synthesis (Section 3.3.1.2), but starting from the bis-aldehyde rather than the Ugi amine. This synthesis was carried out by master student Joe Parry. [24]



Scheme 3.6 Synthesis of **FcHH-(*R*,*R*)**.

3.3.2 Coupling to Nucleobases

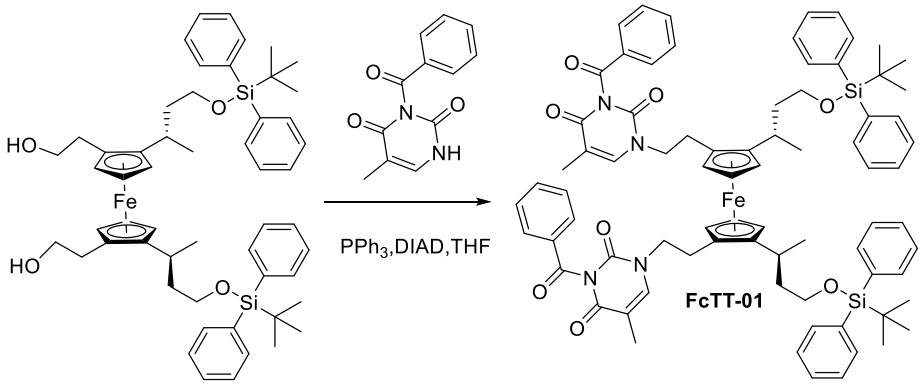
Two pyrimidines, thymine (T) and cytosine (C), were selected to be attached to compound **13**. In order to influence the selectivity of this reaction and enhance the compound's solubility, thymine was subjected to a two-step process for selective protection at N3 with a benzoyl group. Initially, thymine underwent a reaction with benzoyl chloride in pyridine at both N1 and N3 (Figure 3.11). Subsequently, the mixture was dissolved in a 1:1 dioxane and H_2O solution, and selective deprotection at N1 was achieved using K_2CO_3 to yield the final product. For cytosine, a mono-protection approach was employed. [15] This entailed reacting with 5 equivalents of benzenecarbonyl chloride, and subsequent treatment with NaOH. Following addition of HCl, the compound was recrystallised from ethanol, yielding a white solid.



14.1 **14.2**

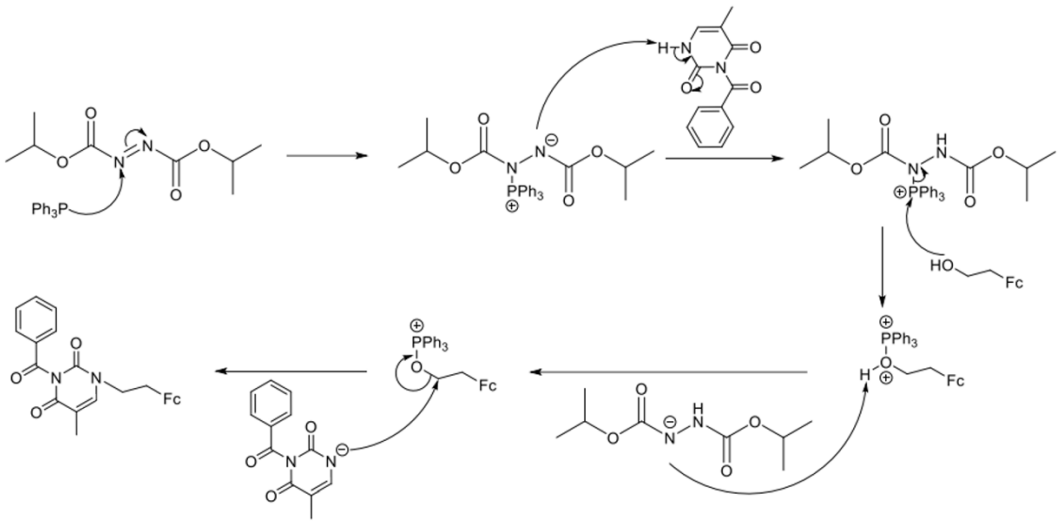
Figure 3.11 benzoyl protected base N-3-benzoylthymine (14.1) and N-4-benzoylcytosine(14.2).

1. Synthesis of FcTT-01



Scheme 3.7 The Mitsunobu reaction to give **FcTT-01** from compound **13**.

The Mitsunobu reaction, named after its originator Mitsunobu in 1967, involves the reaction of an alcohol and a nucleophile using triphenylphosphine and diisopropyl azodicarboxylate (DIAD) or diethyl azodicarboxylate (DEAD) (Scheme 3.7) [22]. The mechanism includes the attack of triphenylphosphine on the azodicarboxylate, leading to the formation of a nitrogen anion. Subsequent steps involve alcohol deprotonation, nucleophile deprotonation, and nucleophile attack on a phosphorus derivative, resulting in the desired product (Scheme 3.8). In the context of the reaction undertaken, the ferrocene compound acts as the electrophile, and the nucleobase serves as the nucleophile. The MS of **FcTT-01** matched calculated, with the ^{1}H spectrum giving a characteristic singlet at 6.82 ppm and singlet at 1.78 ppm (CH_3) for the proton on the thymine unit.

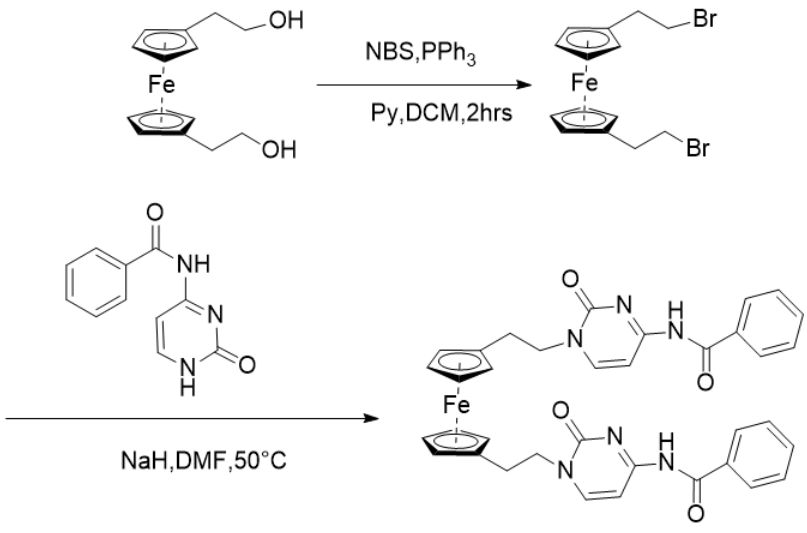


Scheme 3.8 Mechanism for Mitsunobu Reaction.

2. Synthesis of FcCC-02

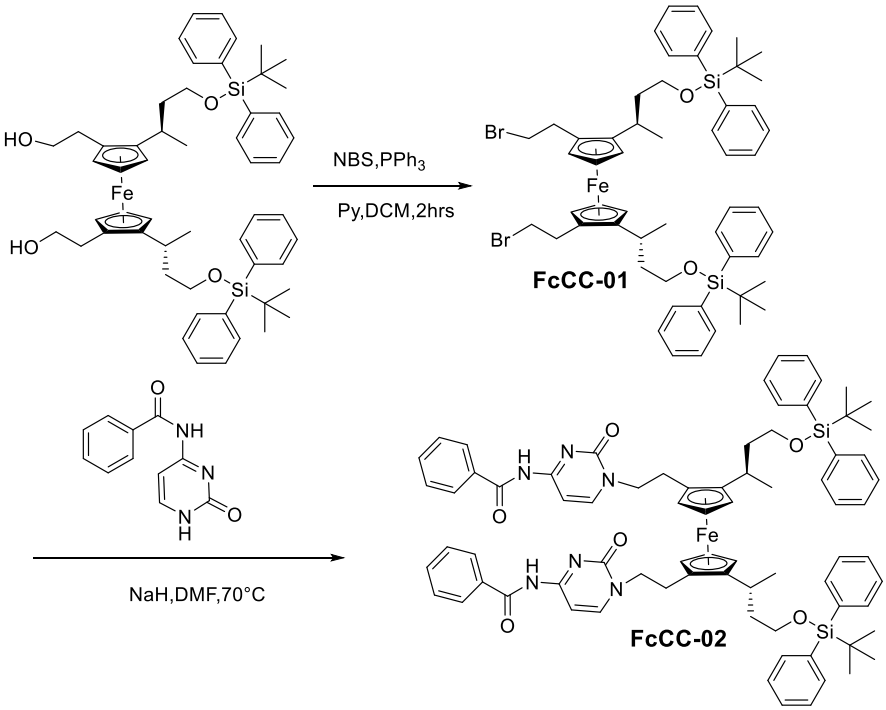
The same reaction was tried with cytosine. However, the reaction did not proceed effectively due to the low solubility of cytosine in almost all types of solvents commonly used in the laboratory. Even in DMSO, it exhibited poor solubility. Taking inspiration from the work of Howarth and Wakelin on α-PNA, a new method was applied to FcCC synthesis. [23]

The process was applied to 1,1'-bis-(2-hydroxyethyl) ferrocene, resulting in the successful generation of the corresponding bromides with high yields of 80% (Scheme 3.9). The reaction begins with the activation of triphenylphosphine through its reaction with N-bromosuccinimide. This is followed by the nucleophilic attack of the alcohol's oxygen on the phosphorus atom, forming an oxyphosphonium intermediate. Subsequently, the oxygen is converted into a leaving group, enabling an SN2 displacement, resulting in the formation of the dibrominated ferrocene product.



Scheme 3.9 Trial reaction for cytosine coupling using 1,1'-bis-(2-hydroxyethyl) ferrocene.

The cytosine coupling reaction involves a nucleophilic substitution, which can be inferred based on several factors. Firstly, the ferrocene substrate is unhindered being a primary bromide, facilitating a one-step reaction (i.e. S_N2 mechanism) where the nucleophile requires room to displace the leaving group by breaking the bromine carbon bond. Additionally, bromide functions as an effective leaving group. Cytosine undergoes deprotonation by sodium hydride before attack on the ferrocene bromide. Due to its polarity and aprotic nature, dimethylformamide promotes nucleophilic substitution in this reaction mechanism.



Scheme 3.10 Synthesis of **FcCC-01** from compound **13** and subsequent reaction with compound **14.2** to form **FcCC-02**.

In a prior initial study by the Tucker group that employed the method outlined in Scheme 3.10, the yield of **FcCC-02** from the dibromide precursor was very low (16%). Consequently, the reaction procedure underwent optimisation, which involved

introducing increasing the molar equivalents of N4-Bz-cytosine **14.2** to eight. Furthermore, the reaction conditions were changed by raising the temperature to 70 °C and extending the reaction time to 24 h. In addition, the deprotonation of cytosine by sodium hydride required time, with it being important to wait for the reaction to complete before introducing the ferrocene reagent, with adding an excess of NaH possibly leading to the failure of the reaction. This systematic optimisation collectively contributed to a significant improvement in the yield from 16% to 54%. The presence of two NH protons at 8.7 ppm and extra peaks in the aromatic region corresponding to benzoylcytosine in figure 3.12 indicate the desired product has been successfully synthesized with MS data also aligning with expectations.

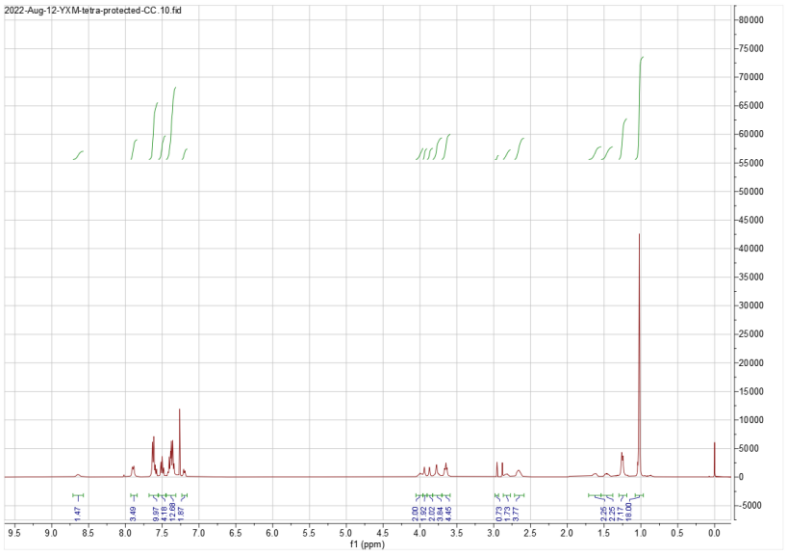
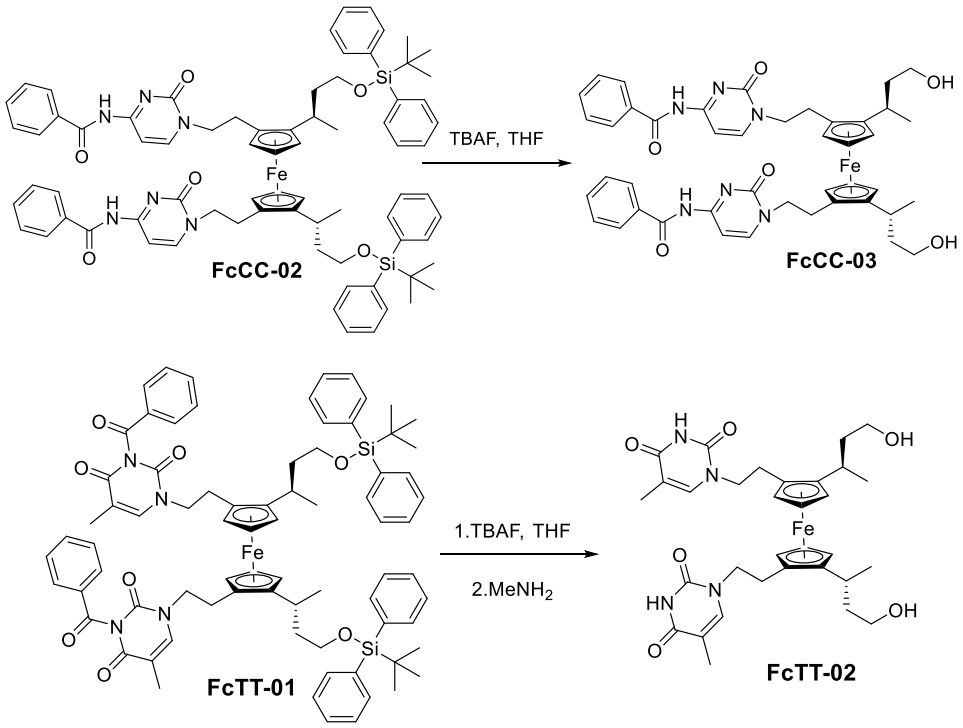


Figure 3.12 ^{1}H NMR spectrum of **FcCC-03** in Chloroform-d.

3.Deprotection of TBDPS



Scheme 3.11 Deprotection of **FcCC-02** and **FcTT-01** to form the compounds **FcCC-03** and **FcTT-02** respectively (all with ***R,R,S$_p$,S$_p$*** stereochemistry).

As shown in Scheme 3.11, the TBDPS groups on the cytosine-substituted ferrocene **FcCC-02** were successfully removed, giving the product **FcCC-03** in a yield of 69%. The 1H NMR spectrum showed the disappearance of the 18H singlet peak that belonged to six CH_3 protons on the protecting groups (Figure 3.13).

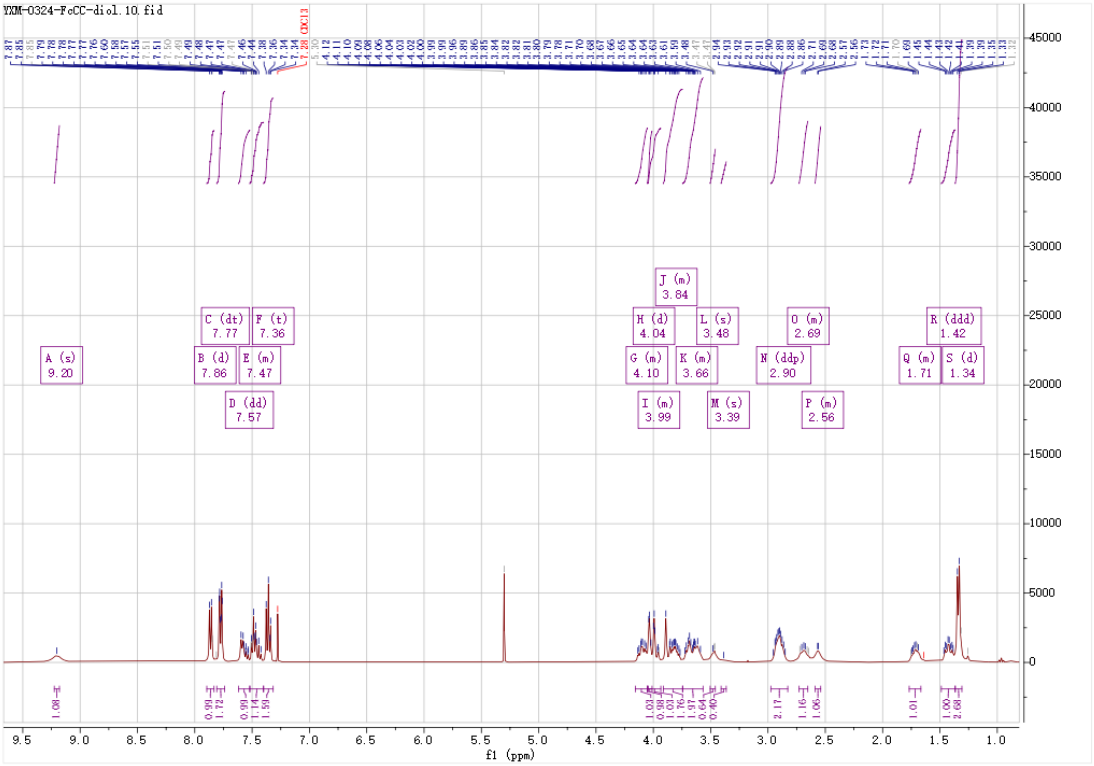
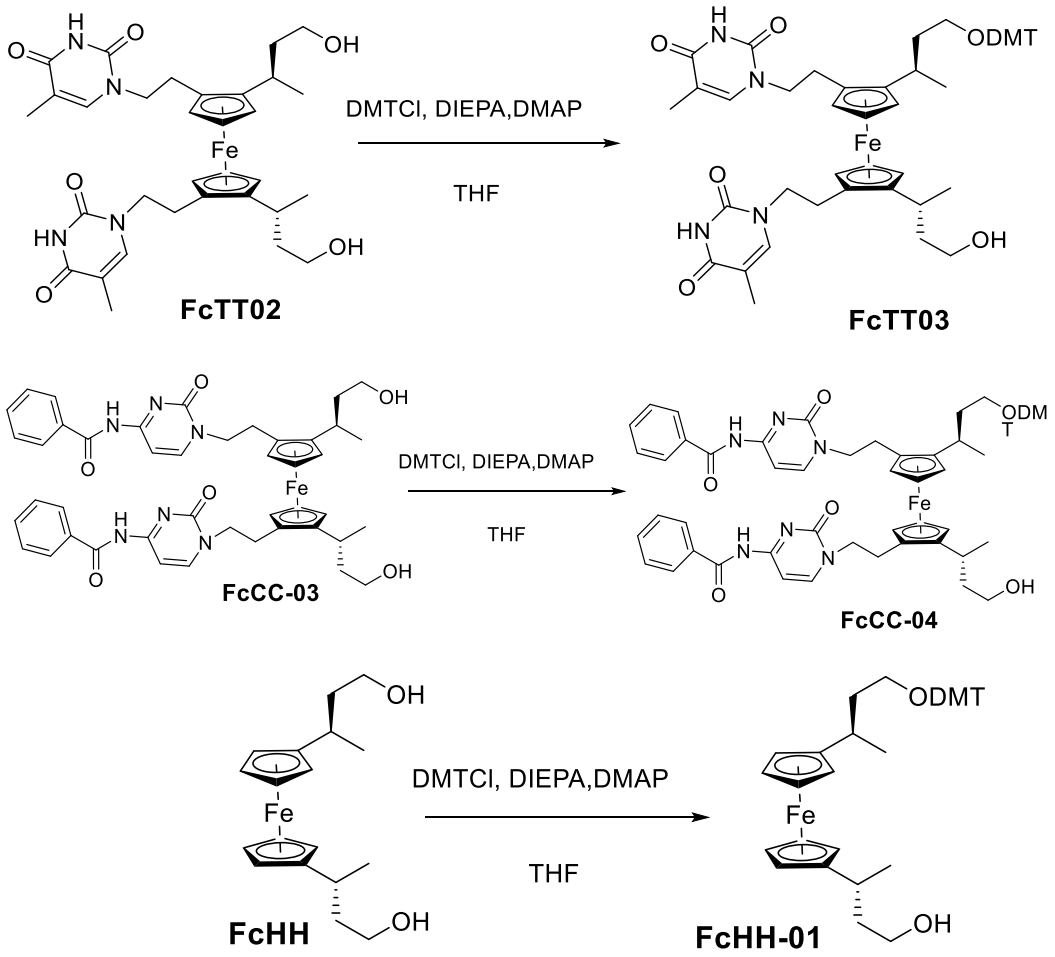


Figure 3.13 ^{1}H NMR spectrum of **FcCC-03** in Chloroform-d.

However, in the case of the thymine-substituted ferrocene **FcTT-02**, multiple spots were observed in the TLC after treatment with TBAF. After analysis, it was found that the benzoyl groups on the thymine residues had also been partially removed. Consequently, methylamine was then added to cleave the benzoyl groups. The NMR spectra and MS of **FcTT-02** matched that in the literature.[12]

3.3.3 Preparation of Monomers for DNA synthesis

1. Mono-DMT protection of the ferrocene diols.



Scheme 3.12 DMT protection of the three ferrocene target compounds.

Two further steps are required to make the required three phosphoramidites for oligomer synthesis. The first step involves protecting one hydroxyl group with a dimethoxytrityl (DMT) group (Scheme 3.12) before proceeding to phosphitylation. DMT serves as the effective temporary protecting group in DNA synthesis due to its sensitivity to mild acidic conditions, setting it apart from other protective agents. This allows for the sequential removal of the DMT group in each coupling step during the

DNA synthesis cycle, as outlined in Chapter 2. Additionally, the distinctive orange color of the liberated DMT carbocation serves as a visual cue for monitoring nucleoside coupling on the DNA synthesizer.

In practical terms, obtaining the mono-protected species proved somewhat challenging, as TLC analysis revealed a statistical mixture of starting material, the desired mono-protected product, and the undesired bis-protected product. It was determined that utilizing 1.1 molar eq. of DMT-Cl would give close to the optimal amount of the mono-protected product, with the yield obtained for **FcHH-02** as 37%. Similar values were obtained for **FcTT-03** and **FcCC-04**. The bis-protected product could be treated with acid to remove the DMT group, allowing recovery of the starting material. As shown in Figure 3.14, which shows the ^{1}H NMR spectrum of **FcCC-04**, the characteristic doublet peak at 6.79 ppm with the correct integration indicated a successful synthesis, supported by MS data.

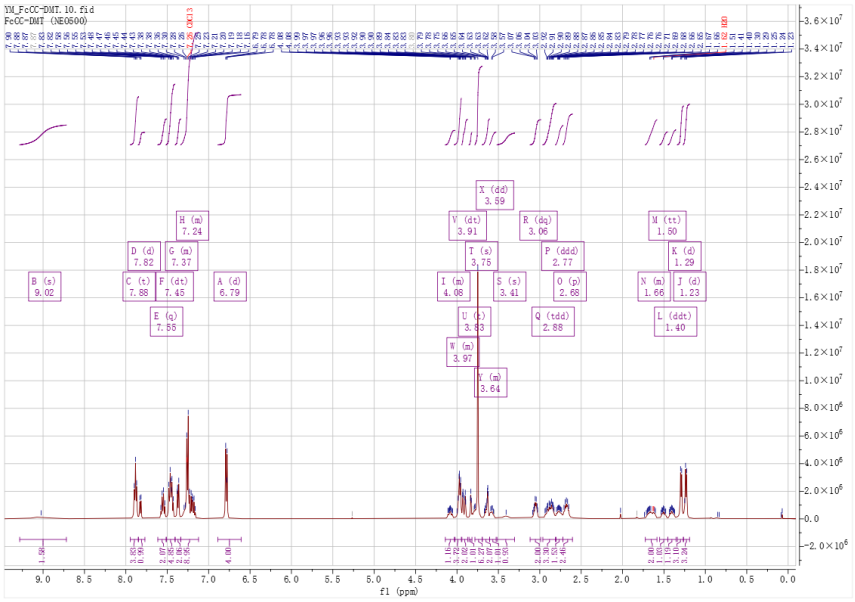
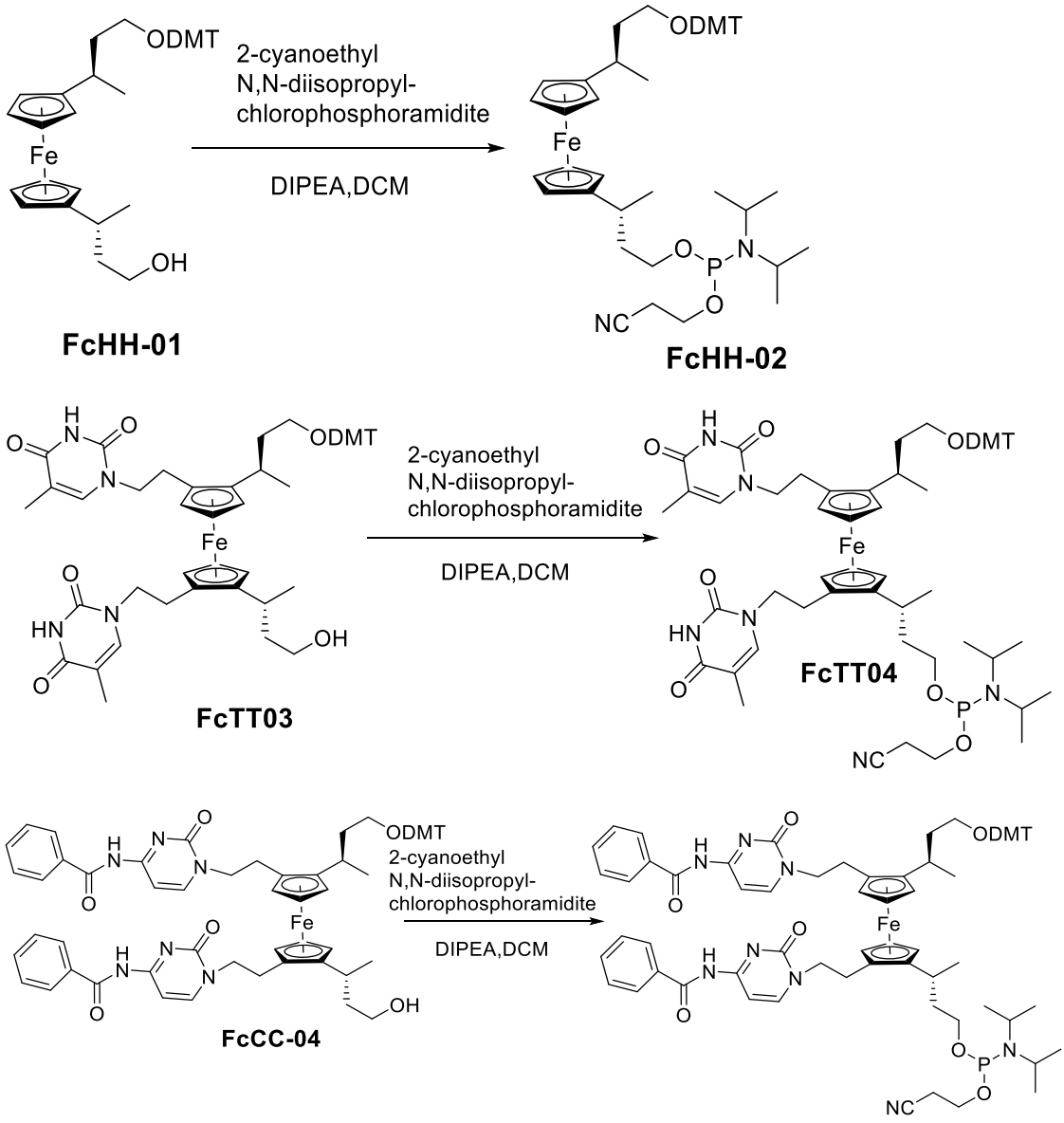


Figure 3.14 ^{1}H NMR spectrum of **FcCC-04** in Chloroform-d.

2. Phosphitylation of the ferrocene target compounds



Scheme 3.13 Phosphitylation of the three ferrocene target compounds.

The phosphitylation stage marks the completion of the ferrocene monomer synthesis, involving the introduction of a phosphoramidite group onto the modified ferrocene molecule (Scheme 3.13). The process is relatively uncomplicated, progressing through a nucleophilic attack of the alcohol group to replace the chloride attached to the phosphorus centre.

The ^{31}P NMR analysis of **FcTT-04-(*R,R,S*$_p$,*S*$_p$)** exhibited two prominent singlet peaks at 147.2 ppm and 146.9 ppm, indicating the success of the reaction. However, for the cytosine analogue, additional peaks at 10.6 ppm and 17.9 ppm were also observed, which were judged to correspond to oxidized products. Attempts were made to purify the product through precipitation, dissolve in 2 mL of EtOAc, then add it into 50 mL Hexane, without significant improvements. Nevertheless, based on the integration, at least half of the sample containing the monomer, which was deemed acceptable. Its instability meant that is essential to proceed with DNA synthesis as soon as possible.

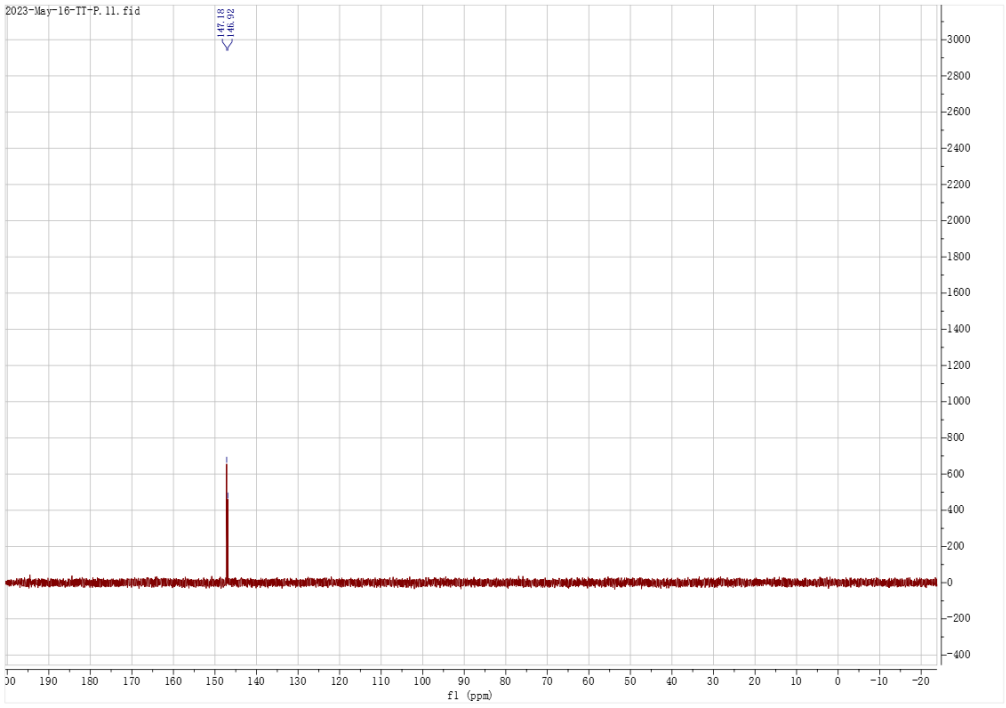


Figure 3.15 ^{31}P NMR of **FcTT-04-(*R,R,S*$_p$,*S*$_p$)** in Chloroform-d.

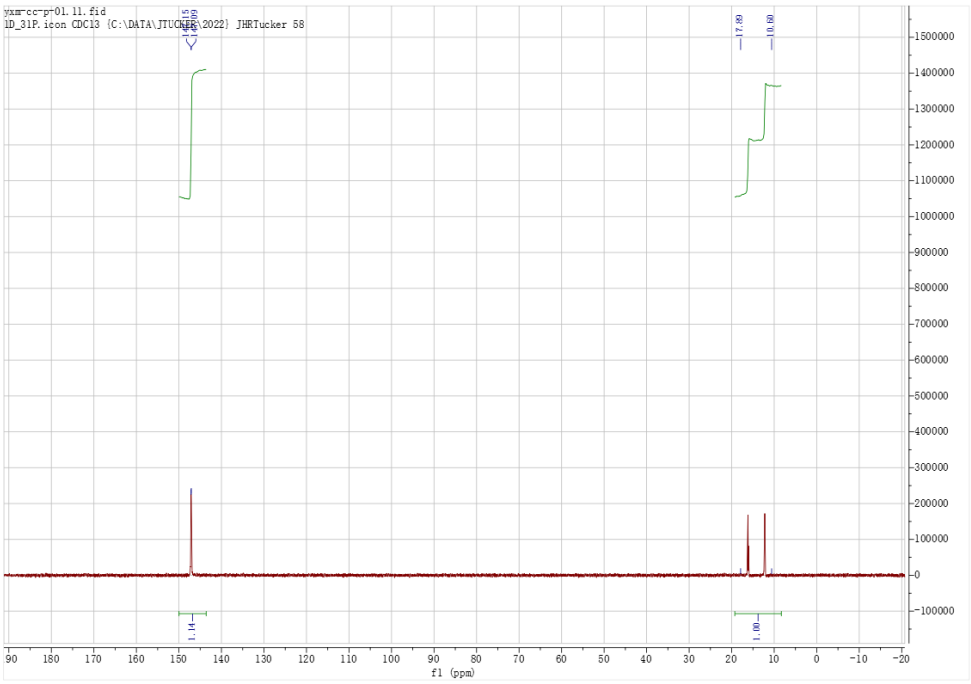


Figure 3.16 ^{31}P NMR of **FcCC-05-(*R,R,S*$_p$,*S*$_p$)** in Chloroform-d.

3.4 Conclusion

The synthesis of two tetra-substituted FcNA phosphoramidite monomers with either two thymine or two cytosine nucleobases was successfully accomplished starting from ferrocene. Additionally, a control chiral ferrocene without bases was also synthesized. The procedures for the tetra-substituted ferrocenes were optimised, resulting in acceptable yields and sufficient amounts for DNA synthesis, which would enable various analytical studies, including NMR spectroscopy.

References

1. Werner, H. *Angewandte Chemie International Edition*, **2012**, *51*(25), 6052-6058.
2. Kealy, T. J.; Pauson, P. L. *Nature*, **1951**, *168*(4285), 1039-1040.
3. Miller, S. A.; Tebboth, J. A.; Tremaine, J. F. *Journal of the Chemical Society*, **1952**, 632-635.
4. Wilkinson, G.; Rosenblum, M.; Whiting, M. C.; Woodward, R. B. *Journal of the American Chemical Society*, **1952**, *74*(8), 2125-2126.
5. Abel, E. W.; Long, N. J.; Orrell, K. G.; Osborne, A. G.; Šik, V. *Journal of organometallic chemistry*, **1991**, *403*(1-2), 195-208.
6. Kedge, J.; *PhD thesis*, University of Birmingham, **2016**.
7. Fischer, V. E.; Pfab, W. *Zeitschrift für Naturforschung B* **1952**, *7*(7), 377-379.
8. Woodward, R. B.; Rosenblum, M.; Whiting, M. C. *Journal of the American Chemical Society*, **1952**, *74*(13), 3458-3459.
9. Broadhead, G. D.; Osgerby, J. M.; Pauson, P. L. *Journal of the Chemical Society*, **1958**, 650-656.
10. Roemer, M.; Nijhuis, C. A. *Dalton Transactions*, **2014**, *43*(31), 11815-11818.
11. Rebiere, F.; Samuel, O.; Kagan, H. B. *Tetrahedron Letters*, **1990**, *31*(22), 3121-3124.
12. Nguyen, H. V.; Zhao, Z. Y.; Sallustrau, A.; Horswell, S. L.; Male, L.; Mulas, A.; Tucker, J. H. *Chemical communications*, **2012**, *48*(100), 12165-12167.
13. A. Togni, T. Hayashi, *Materials Science*, **1995**, 173 – 174.
14. Woodward, R. B.; Rosenblum, M.; Whiting, M. C. *Journal of the American Chemical Society*, **1952**, 74(13), 3458-3459.

15. Sallustrau, A. *PhD thesis*, University of Birmingham, **2014**.
16. Schwink, L.; Knochel, P. *Chemistry–A European Journal*, **1998**, 4(5), 950-968.
17. Wright, J.; Frambes, L.; Reeves, P. *Journal of organometallic chemistry*, **1994**, 476(2), 215-217.
18. Mikami, K.; Matsumoto, S.; Ishida, A.; Takamuku, S.; Suenobu, T.; Fukuzumi, S. *Journal of the American Chemical Society*, **1995**, 117(45), 11134-11141.
19. Navarro, A. E.; Spinelli, N.; Moustrou, C.; Chaix, C.; Mandrand, B.; Brisset, H. *Nucleic acids research*, **2004**, 32(17), 5310-5319.
20. Hanessian, S.; Lavallee, P. *Canadian Journal of Chemistry*, **1975**, 53(19), 2975-2977.
21. Media S. *PhD thesis*, University of Birmingham, **2013**.
22. Mitsunobu, O.; Yamada, M. *Bulletin of the Chemical Society of Japan*, **1967**, 40(10), 2380-2382.
23. Howarth, N. M.; Wakelin, L. P. *The Journal of Organic Chemistry*, **1997**, 62(16), 5441-5450.
24. Parry, J. *MSci thesis*, University of Birmingham, **2014**.

Chapter 4 – Ferrocene-Containing DNA Tetrahedron Assemblies

4.1 Introduction

4.1.1 The Turberfield DNA Tetrahedron

Nucleic acids offer a means to bring multiple molecular components together in a programmable and predictable manner to form higher-ordered assemblies and frameworks. This methodology is at the core of DNA nanotechnology and DNA origami.[1][2] Various three-dimensional structures have been created, including triangular prisms, cubes, pentagonal prisms and hexagonal prisms.[3][4]

The DNA tetrahedron **T1**, initially described by the Turberfield research group, is perhaps the most straightforward example of a non-biological nucleic acid assembly.[5] The synthetic approach, as illustrated in Figure 4.1,[5] involves constructing tetrahedra using four 55-base oligonucleotides. Each of the tetrahedron's six edges is formed by a 17-base "edge subsequence" hybridised to its complement. Each strand contains three of these subsequences (or their complements), separated by two-base-pair "hinges" that remain unhybridized. These hinges are included in the design to allow the vertices of the structure to flex and accommodate a 60° angle between adjacent edges.

The four component oligonucleotides are designed to self-assemble into a regular tetrahedron with double-stranded edges connected through these hinges. Each oligonucleotide runs along one of the four faces and is hybridized to three oligonucleotides running along the neighboring faces at the shared edges. Each vertex

forms a nicked three-arm junction. Since the edges have distinct sequences, each tetrahedron is chiral, with two possible enantiomers. Its applications include drug delivery and its use as a surface scaffold for sensing.

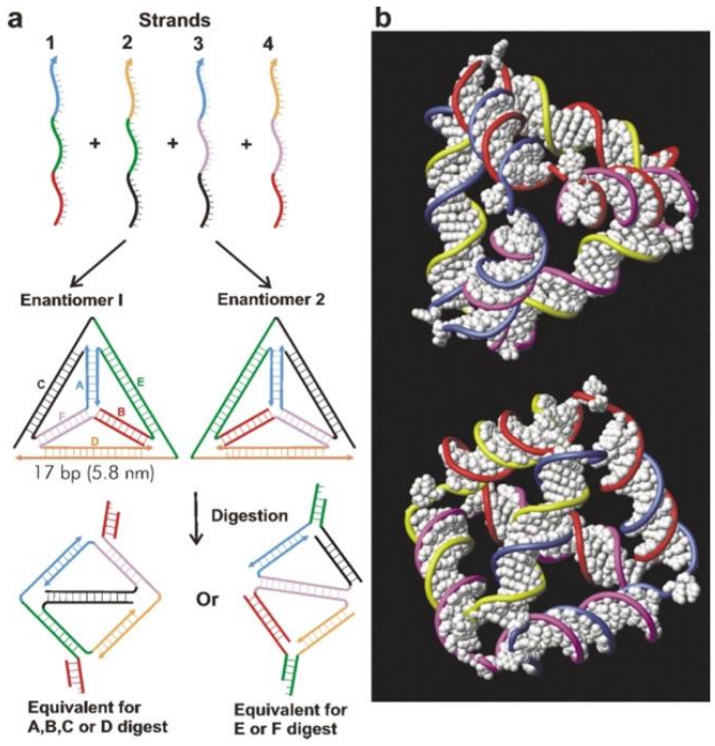


Figure 4.1 Turberfield DNA tetrahedron, taken from reference[5].

4.1.2 Applications of the DNA Tetrahedron

The application of DNA tetrahedra in drug delivery has significant potential for several reasons. Firstly, the 3D structure and dimensions of DNA nanostructures can be readily controlled by altering the sequence and length of the strands. Specific drugs can be incorporated into the three-dimensional structure as “guest” molecules.[6] Furthermore, medications can be directly linked to the tetrahedron before spontaneous assembly takes place: the drugs, primarily nucleic acids, are typically attached in advance to

either the 5' or 3' end of individual strands.[8] Secondly, DNA polymers are not easily influenced by antibodies within the human body, as nucleic acids and DNA derivatives are non-toxic and pose a low physiological risk. An illustrative example of a DNA drug delivery system was pioneered by Kim,[7] wherein a DNA tetrahedron encapsulated Doxorubicin, a potent anticancer drug.

Another application for the DNA tetrahedron is as a scaffold for sensing, as demonstrated by the Fan group.[9] They developed a structure to modify biosensing interfaces effectively using precisely defined DNA tetrahedral nanostructures of varying sizes (Figure 4.2). Thiol-modified tetrahedra were securely affixed to gold electrode surfaces with excellent reproducibility, serving as a sturdy framework for the specific attachment of a wide range of biomolecular probes. The size of these tetrahedra could be tailored to finely adjust the kinetics and thermodynamics of biomolecular recognition.

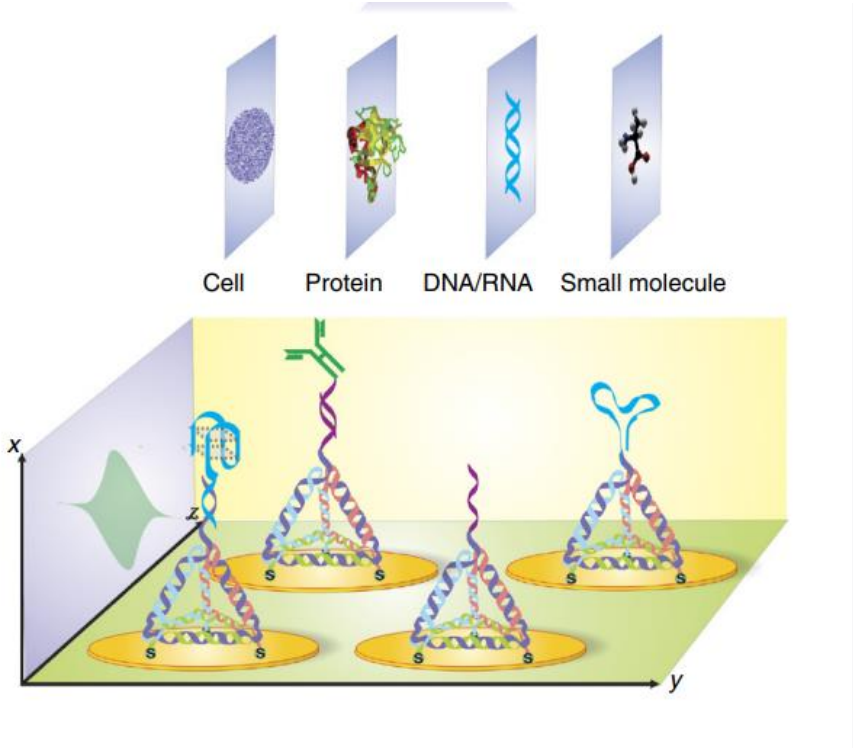


Figure 4.2 Fan's DNA tetrahedron sensor.[9]

This tetrahedral-based biosensing platform significantly amplified target accessibility and sensitivity when detecting diverse molecular targets, encompassing DNA, RNA, proteins, small molecules, and even entire cells.[10-15]

4.1.3 Ferrocene-containing Tetrahedron

Despite the popularity of ferrocene-modified DNA as mentioned in Chapter 1, there are relatively few instances of higher-order nucleic acid assemblies incorporating ferrocene units that are more complex than single strands and duplexes.[16][17] This stands in stark contrast to the numerous examples of analogous self-assembled supramolecular structures containing ferrocene derived from chemical components, as opposed to nucleic acids.[27]

In one of those rare cases where ferrocene has been employed in this manner, Fan and colleagues designed a series of tetrahedron assemblies (Figure 4.3). The aim of creating these assemblies was to explore the mechanisms of charge transport facilitated by DNA, both through the duplex structure and through the surrounding space. This investigation involved the attachment of a single ferrocene tag at specific positions above the surface of a gold electrode.[16][17]

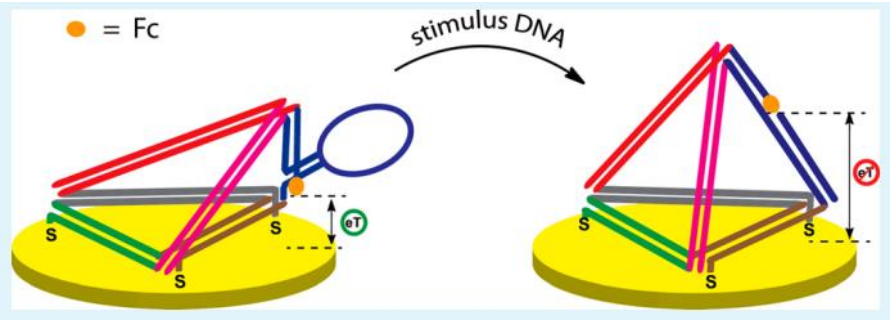


Figure 4.3 An example of a ferrocene-appended DNA tetrahedron.[17]

4.2 Aim

In order to broaden the range of redox-active nucleic acid assemblies, and to continue the ongoing interest in exploring metal-modified oligonucleotides containing ferrocene [18] and other [19-20] components, this chapter reports on the synthesis and analysis of self-assembled redox-active DNA tetrahedra on surfaces. These tetrahedra incorporate one or more ferrocene groups into the backbone of DNA at their vertices (up to four). The objective was to create assemblies that would establish evenly spaced and robust surface anchors on electrodes (Figure 4.4). Subsequently, these anchors could serve as platforms for affixing probes at the apex for electrochemical sensing applications.

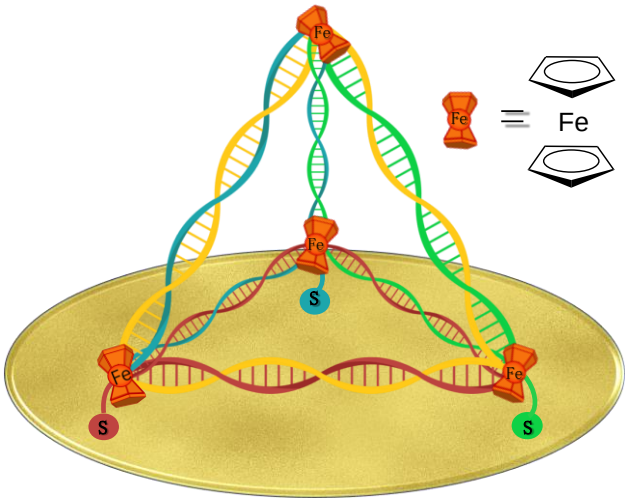


Figure 4.4 A ferrocene-containing nucleic acid assembly on a gold surface.

4.3 Results and Discussion

4.3.1 Synthesis of components of the tetrahedron assembly

As previously mentioned, Turberfield's DNA tetrahedron **T1** is composed of four 55-mer unmodified strands (A1, A2, A3, and A4, as indicated in Table 4.1), which can be annealed together in a one-step process. Each strand is subdivided into three segments that form duplexes, and each segment is separated by two unhybridized bases, functioning as "hinges." In the design of our metal-modified systems, we opted to position a ferrocene unit at one of these two hinges, between these two bases. This is exemplified in strands B_1-B_4 (Table 4.1), where ferrocene is situated at the first hinge of each strand (closest to the 5' end). Additionally, a commercially available disulfide linker is affixed to the 3' end of each strand, facilitating the attachment of the tetrahedra to a gold surface. We also formulated two types of control strands (C1-C4 and D1-D4) that exclusively contain either the disulfide linker or the ferrocene unit, respectively. Strands A_1-A_4 were bought commercially.

Table 4.1. The sequences used as components for the Fc-functionalised tetrahedra.

A_1	5′-ACATTCCTAAGTCTGAAACATTACAGCTTGCTACACGAGAAGAGCCGCCATAGTA-3′
A_2	5′-TATCACCAGGCAGTTGACAGTGTAGCAAGCTGTAATAGATGCGAGGGTCCAATAC-3′
A_3	5′-TCAACTGCCTGGTGATAAAACGACACTACGTGGGAATCTACTATGGCGGCTCTTC-3′
A_4	5′-TTCAGACTTAGGAATGTGCTTCCCACGTAGTGTCGTTTGTATTGGACCCTCGCAT-3′
B_1	5′-ACATTCCTAAGTCTGAAA**Fc**CATTACAGCTTGCTACACGAGAAGAGCCGCCATAGTA-C6-SS-3′
B_2	5′-TATCACCAGGCAGTTGAC**Fc**AGTGTAGCAAGCTGTAATAGATGCGAGGGTCCAATAC-C6-SS-3′
B_3	5′-TCAACTGCCTGGTGATAA**Fc**AACGACACTACGTGGGAATCTACTATGGCGGCTCTTC-C6-SS-3′
B_4	5′-TTCAGACTTAGGAATGTG**Fc**CTTCCCACGTAGTGTCGTTTGTATTGGACCCTCGCAT-C6-SS-3′
C_1	5′-ACATTCCTAAGTCTGAAACATTACAGCTTGCTACACGAGAAGAGCCGCCATAGTA-C6-SS-3′
C_2	5′-TATCACCAGGCAGTTGACAGTGTAGCAAGCTGTAATAGATGCGAGGGTCCAATAC-C6-SS-3′
C_3	5′-TCAACTGCCTGGTGATAAAACGACACTACGTGGGAATCTACTATGGCGGCTCTTC-C6-SS-3′
C_4	5′-TTCAGACTTAGGAATGTGCTTCCCACGTAGTGTCGTTTGTATTGGACCCTCGCAT-C6-SS-3′
D_1	5′-ACATTCCTAAGTCTGAAA**Fc**CATTACAGCTTGCTACACGAGAAGAGCCGCCATAGTA-3′
D_2	5′-TATCACCAGGCAGTTGAC**Fc**AGTGTAGCAAGCTGTAATAGATGCGAGGGTCCAATAC-3′
D_3	5′-TCAACTGCCTGGTGATAA**Fc**AACGACACTACGTGGGAATCTACTATGGCGGCTCTTC-3′
D_4	5′-TTCAGACTTAGGAATGTG**Fc**CTTCCCACGTAGTGTCGTTTGTATTGGACCCTCGCAT-3′

A known[21] ferrocene phosphoramidite with two propyl linkers was selected as the insertion group for the hinge (Figure 4.5). This compound was synthesised in six steps from ferrocene and subsequently incorporated into DNA for strands B1-B4 and D1-D4 using standard automated solid-phase synthesis.

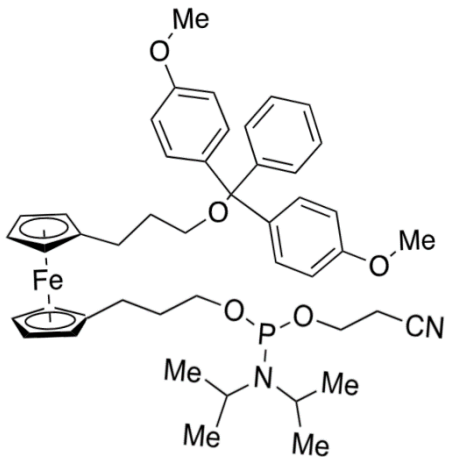


Figure 4.5 Ferrocene phosphoramidite monomer incorporated into the strands (labelled as Fc).

4.3.2 Purification of the tetrahedron assembly components

Typically, DNA strands are purified in the Tucker group using high-performance liquid chromatography (HPLC). However, the length of these oligonucleotides proved too challenging for chromatographic separation, especially with the added complexity of ferrocene and disulfide modifications.

As a result, denaturing polyacrylamide gel electrophoresis was employed for strand purification.[22] This method functions by applying an electric field to biomolecules migrating through the gel, segregating them based on their electrical charge, size, and shape. The negatively charged phosphate backbone of DNA makes it well-suited for this technique. Urea was used as a denaturing agent to disrupt hydrogen bonds and prevent unwanted aggregation.

The gels were visualized under UV light (see Figure 4.6), revealing the presence of multiple strong bands in some cases. This observation was attributed to a combination of incomplete strand synthesis and some strands containing residual protecting groups. To remove these protective groups, further heating in an ammonia solution and increasing the duration time could be employed in the future to enhance this process.

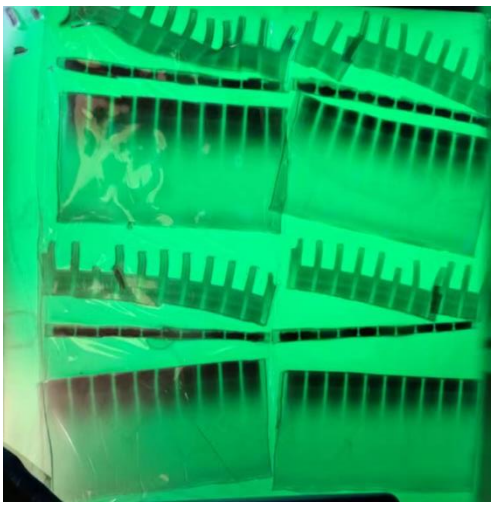


Figure 4.6 Visualisation of bands obtained using gel electrophoresis purification. The photograph shows the cut bands.

In each case, the slowest band at the top, corresponding to the desired product, was excised. The cut bands were then crushed and extracted. They were frozen at -80 ºC in liquid nitrogen and subsequently reheated to 90 ºC. Afterwards, they were dissolved in water, filtered, desalted, and concentrated. The successful synthesis of each strand was confirmed using mass spectrometry (see Chapter 6) and native polyacrylamide gel electrophoresis (Figure 4.7).

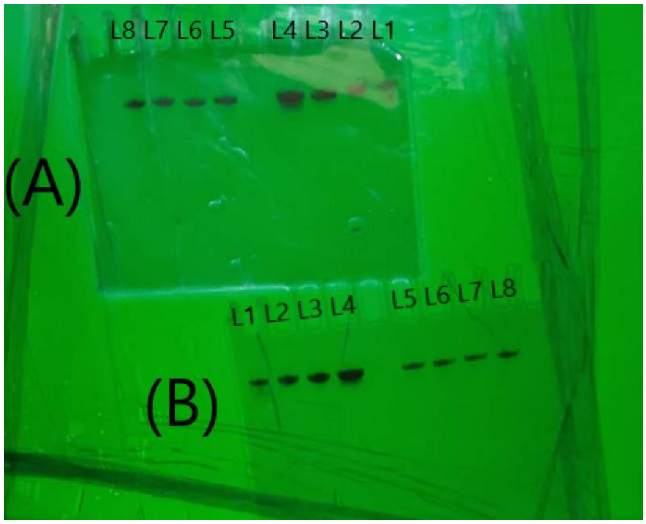


Figure 4.7 Purity check using denaturing electrophoresis gel of the individual strands. (A) Lanes L1-L4: A_1, A_2, A_3, A_4 respectively; L5-L8: B_1, B_2, B_3, B_4 respectively; (B) Lanes L1–L4: A_1, A_2, A_3, A_4 respectively; L5: D_1, D_2, D_3, D_4 (15% PAGE, 1 × TBE buffer, 100 V, 2 h at 4 °C).

4.3.3 Synthesis of Fc-modified tetrahedra

Lastly, four strands from various combinations of strands 1-4 (selected from groups A, C, B, or D) were combined at 1 μM each in a single step to create a set of tetrahedra. Each strand included three sections that were complementary to the three other strands. In each instance, the ferrocene units were positioned at the hinges, precisely between the two unhybridized bases that remain so during tetrahedron assembly. The resulting eight tetrahedra, **T1**-**T8**, are shown in Figure 4.8, with the sulphur units shown in their reduced forms, as used for monolayer formation.

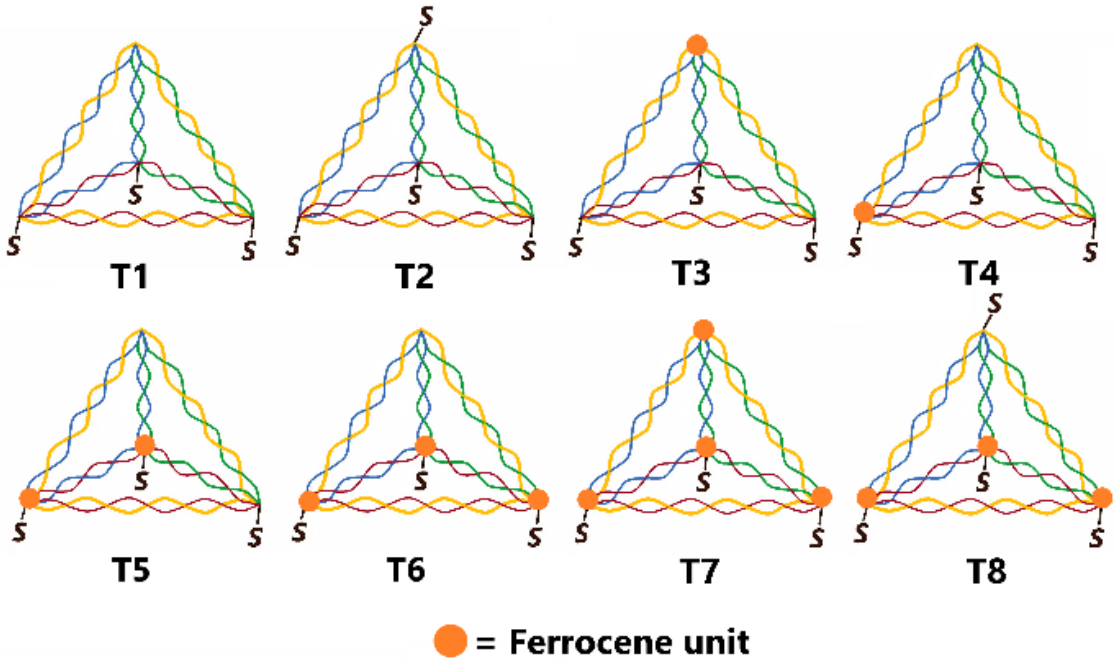


Figure 4.8 The DNA tetrahedra made in this study with the stand combinations in brackets: **T1** (A_1 + A_2 + A_3 + A_4); **T2** (C_1 + C_2 + C_3 + C_4); **T3** (A_1 + C_2 + C_3 + B_4); **T4** (A_1 + C_2 + B_3 + C_4); **T5** (A_1 + B_2 + B_3 + C_4); **T6** (D_1 + B_2 + B_3 + C_4); **T7** (D_1 + B_2 + B_3 + B_4); **T8** (B_1 + B_2 + B_3 + B_4).

To create each DNA tetrahedron, four DNA components were initially prepared at a concentration of 100 μM and then 1 μl of each of the four strands was mixed in a Tris

base buffer at a final concentration of 1 μM for each strand. The solution was subsequently heated to 95 °C using a heating block for 10 min and then rapidly cooled in ice water for 1 min, leading to the formation of the tetrahedron structure.

Native polyacrylamide gel electrophoresis (PAGE) was employed to scrutinize the tetrahedron assembly process at a strand concentration of 1 μM. An illustrative instance, demonstrating the creation of tetrahedron **T8**, is presented in Figure 4.9. The most slowly migrating band exclusively emerges when all four complementary strands are present (lane L9, Figure 4.9), whereas the fastest bands are exclusively observed in the single-strand loading lanes (L1-L4). This observed pattern, along with the two- and three-component assemblies noted in the other lanes, is consistent with findings in existing literature[5][8] and indicates the successful formation of the tetrahedron. The static band at the upper end of L9 (adjacent to the wells) is attributed to higher-order DNA misfolds. It was detected only when all four strands were present and was even observed when a lower strand concentration of 15 nM was used. The original Turberfield tetrahedron **T1** ($A_1 + A_2 + A_3 + A_4$), made from four commercial strands, shows a similar pattern (Figure 4.10).

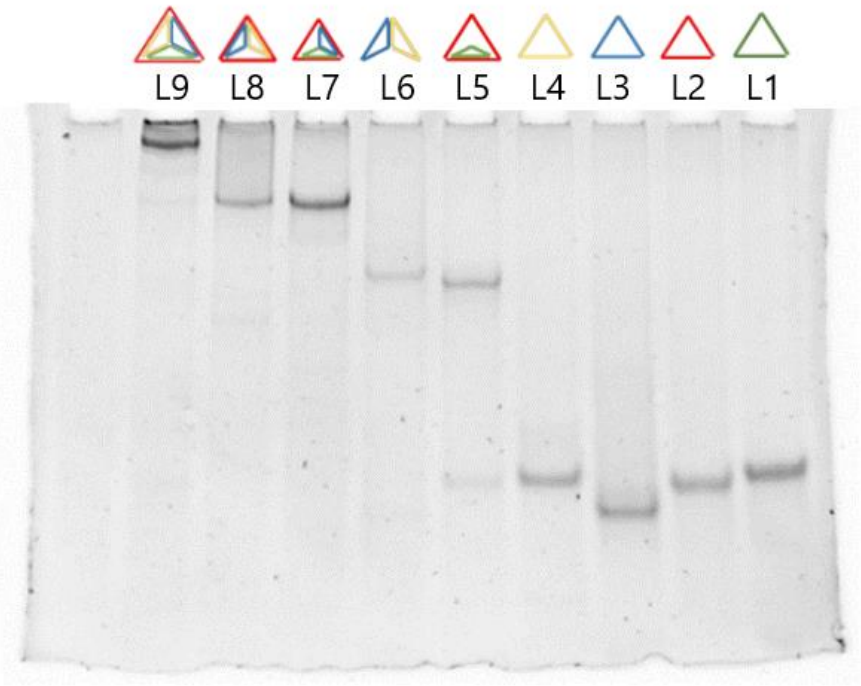


Figure 4.9 Native electrophoresis gel of the individual and assembled strand components of tetrahedron **T8**. Lanes L1–L4: B_1, B_2, B_3, B_4 respectively; L5: B1 + B2 together; L6: B3 + B4 together; L7: B1 + B2 + B3 together; L8: B2 + B3 + B4 together; L9: B1 + B2 + B3 + B4. Conditions: 10% PAGE, TBE buffer (pH = 8.3), 100 V, run for 1.5 h at 4 °C, [strand] = 1 μM.

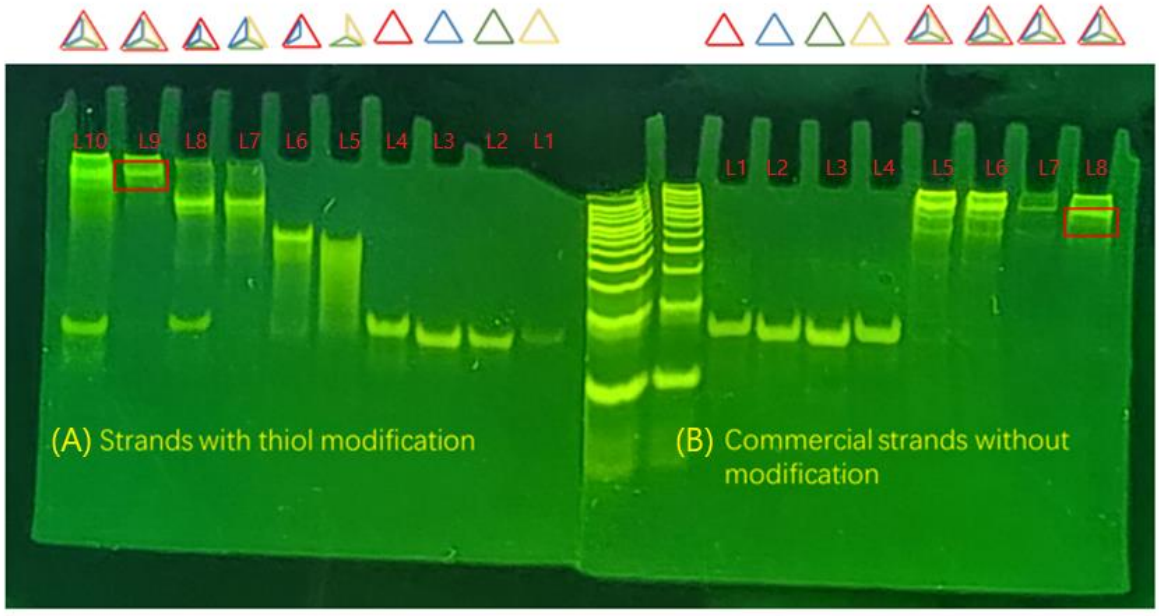


Figure 4.10 Native electrophoresis gel of the individual and assembled strand components of DNA tetrahedra (A) Tetrahedron **T2**, Lanes L1-L4: C_1, C_2, C_3, C_4 respectively; L5: C_1 + C_2 together; L6: C_3 + C_4 together; L7: C_1 + C_2 + C_3 together; L8: B_2 + B_3 + B_4 together; L9: C_1 + C_2 + C_3 + C_4; (B) Tetrahedron **T1**, Lanes L1-L4: A_1, A_2, A_3, A_4 respectively; L8: A_1 + A_2 + A_3 + A_4 (10% PAGE, 1 × TBE buffer, 100 V, 1.5 h at 4 °C [strand] = 1 μM.

Next, we compared the gel behavior of the seven modified tetrahedra to the unmodified DNA tetrahedron, as depicted in Figure 4.11. These results consistently demonstrate that, in each case, the proper combination of equimolar amounts of four strands results in the intended assembly. Moreover, the introduction of ferrocene modifications does not have a detrimental impact on tetrahedron formation.

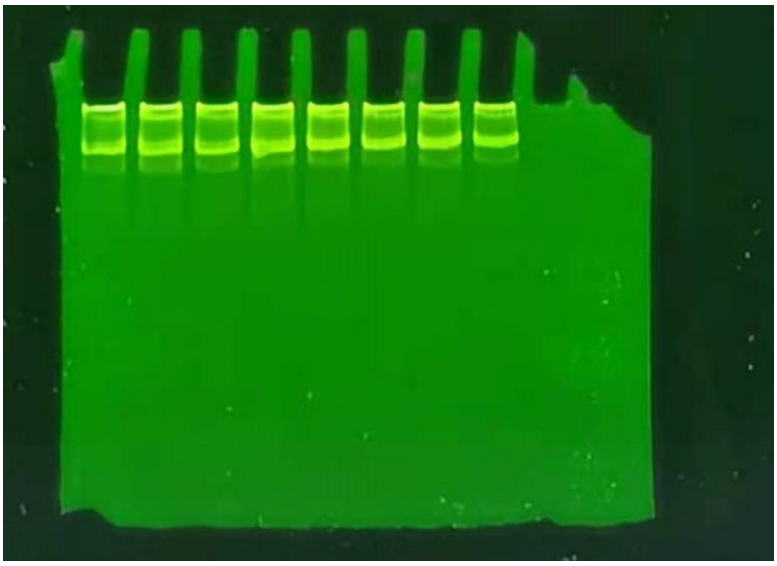


Figure 4.11 Native electrophoresis gel of eight DNA tetrahedra made in this study, 100 V, 1.5 h at 4 °C. Lane from left to right, (1) **T1** (2) **T2** (3) **T3** (4) **T4** (5) **T5** (6) **T6** (7) **T7** (8) **T8**, (10% PAGE, 1 × TBE buffer, 100 V, 1.5 h at 4 °C [strand] = 1 μM

Another technique commonly employed in the literature to investigate this assembly process is Atomic Force Microscopy (AFM).[9] However, thus far, it has not yielded any useful data.

4.3.4 Electrochemical properties of ferrocene-modified DNA tetrahedra

With the ferrocene-modified tetrahedron assemblies now prepared, the next stage was to create self-assembled monolayers (SAMs) of these assemblies on gold electrodes to facilitate electrochemical investigations. The process entailed reducing the disulfide linkers using tris(2-carboxyethyl)phosphine (TCEP), followed by the reaction of the thiol groups with the gold surface to establish S-Au bonds, as described by others previously.[10-15]

Cyclic voltammetry (CV) experiments were conducted on the ferrocene-modified assemblies at various scan rates, ranging from 1000 mV s^{-1} to 20 mV s^{-1}. In the case of the tetrahedron **T3**, which consists of component strands A1, C2, C3, and B4, and has one ferrocene unit at one vertex, no detectable electrical current signal was observed (Figure 4.12). A previous study conducted by Fan's research group [9] involved attaching a ferrocene tag externally to the edge of a tetrahedron, revealing that the efficiency of electron transfer was significantly influenced by the distance between the tag and the electrode surface. In the **T3** system described here, it seems that integrating the ferrocene unit within such a structurally rigid framework, anchored by three thiol units, effectively fixes the ferrocene at a distance estimated to be between 4-5 nm from the electrode surface. As a result, no discernible electron transfer occurs under these experimental conditions.

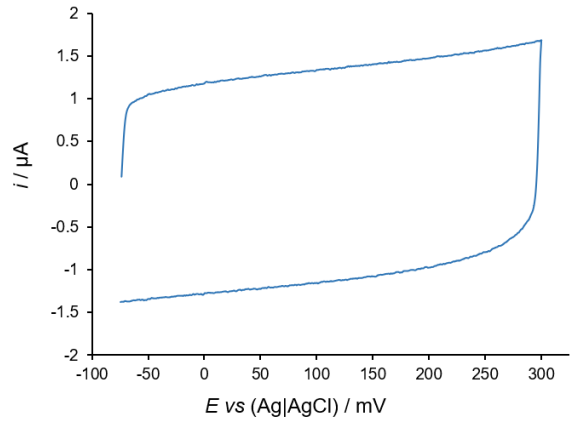


Figure 4.12 CV of tetrahedron **T3** in 10 mM phosphate buffer (pH 7.0), 1 M $NaClO_4$, scan rate: 1000 mV s^{-1}.

Based on the above findings, as expected, the **T4** system, which had one ferrocene unit positioned closer to the electrode surface, indeed produced a signal at 1000 mV s^{-1}, albeit relatively weak (Figure 4.13). However, this signal was even weaker than that of the individual component strand **B3** under the same experimental conditions (Figure 4.14). One possible explanation for this discrepancy could be the limited surface coverage resulting from the bulkiness of the nanostructures, although a comprehensive analysis of such a small, broad peak is challenging.

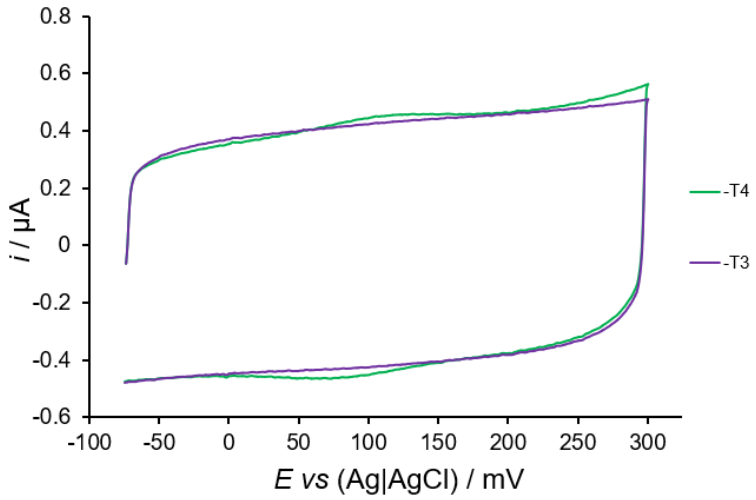


Figure 4.13 CVs of tetrahedra **T3** & **T4** in 10 mM sodium phosphate buffer (pH 7.0), 1 M $NaClO_4$, scan rate: 1000 mV s^{-1}.

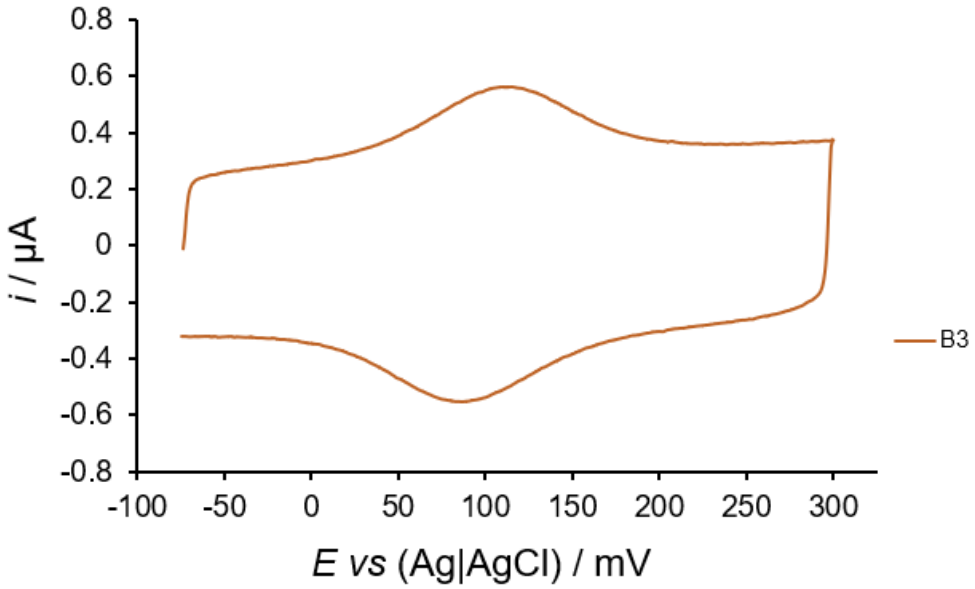


Figure 4.14 CVs of the single ferrocene-modified DNA strand **B_3** in 10 mM sodium phosphate buffer (pH 7.0), 1 M $NaClO_4$, scan rate: 1000 mV s^{-1}.

The precise mechanism of electron transfer through (or from) double-stranded DNA anchored to an electrode is not yet fully understood. One proposed mechanism is the "through-DNA" mechanism [25][26][27] and some scientific literature has posited that DNA duplexes serve as highly efficient charge carriers, thanks to π−π* interactions between closely stacked base pairs. However, in the present case, the tetrahedron unit does not form a traditional DNA duplex. As mentioned in the introduction, each oligonucleotide includes two bases serving as 'hinges', with the ferrocene molecule incorporated between these two bases to allow for adequate flexibility between adjacent edges. Consequently, the two adjacent bases of ferrocene remain unhybridized, a factor that could potentially influence the efficiency of electron transfer through the DNA structure.

However, another critical factor to consider is the difference in rigidity between the individual strand and the tetrahedron affixed to the surface. The lower three edges of the pyramid structure lie flat, in close proximity to and perfectly aligned with the electrode surface. The inherent rigidity of this facet may impede its ability to approach the electrode surface within the range of tunnelling distance. According to Plaxco and coworkers,[28] this rigidity of the system could prevent it from making close contact with the surface, which, in turn, results in a weaker signal. In contrast, the single strand **B3** exhibits greater flexibility and dynamism within its self-assembled monolayer (SAM), potentially allowing the ferrocene unit to approach the surface more closely. Consequently, this flexibility leads to a more prominent and pronounced signal.

A plot of peak current against potential scan rate is normally used to confirm that a redox process is confined to the surface. As the weak signals did not allow the currents to be accurately determined in assemblies containing only one ferrocene unit, the focus shifted to assemblies with multiple units. It was expected that an increase in the number of ferrocene units within the assembly would lead to a higher current signal. With this expectation in mind, the investigation turned to the **T7** system, which featured four ferrocene units (comprising component strands D1, B2, B3, and B4).

The results showed a considerably stronger signal in a series of cyclic voltammograms (Figure 4.15). Additionally, a linear dependence of peak current on potential scan rate was observed (Figure 4.16), suggesting the presence of surface-bound ferrocene

species and indicating the successful formation of a self-assembled monolayer (SAM).

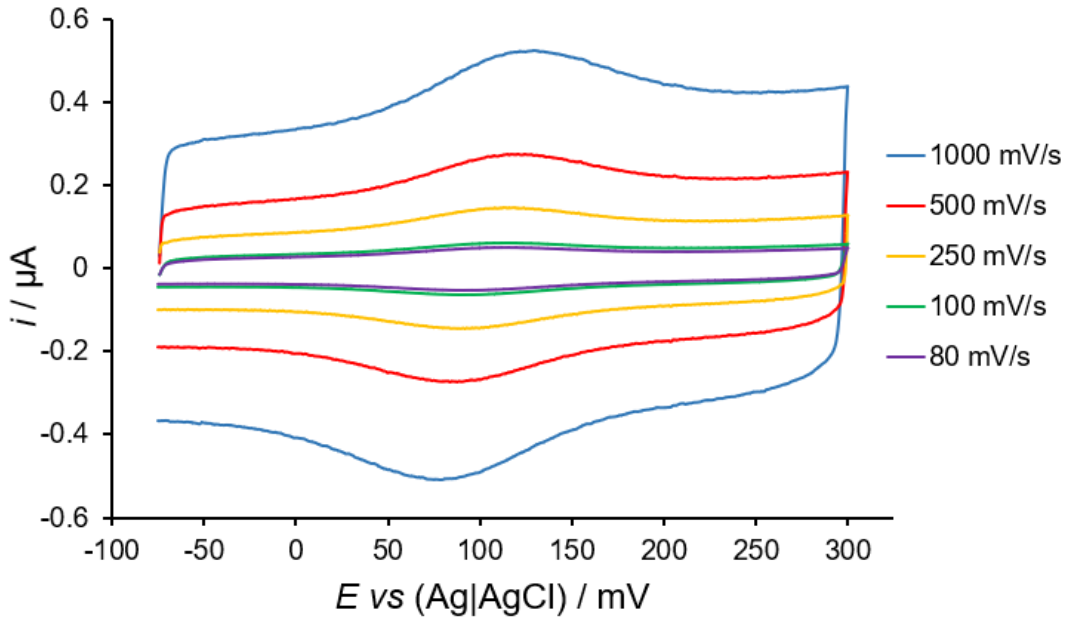


Figure 4.15 CVs of tetrahedron **T7** at various scan rates in 10 mM sodium phosphate buffer (pH 7.0), 1 M $NaClO_4$

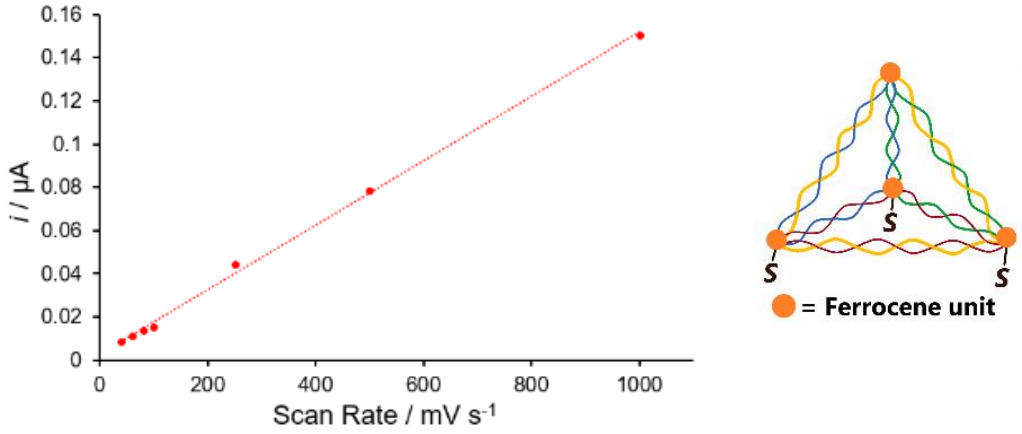


Figure 4.16 Plot of cathodic peak current *vs* scan rate for the **T7** system.

It is important to highlight that in surface-based electrochemistry, the anodic and cathodic peaks for electrochemically reversible processes should exhibit a peak separation (ΔE) very close to 0 mV, unlike solution-based electrochemistry where the value for a one-electron process is approximately 59 mV.[11] In this system, the ΔE_p

values varied from 9 mV to 30 mV (see Table 4.2), which are notably smaller than typically observed for freely diffusing species, especially at lower scan rates. The fact that the peak separations are greater than 0 mV may suggest slower electron transfer kinetics in these systems.

Table 4.2 Electrochemical data for T7.

Scan Rate / mV s^{-1}	Anodic Peak Current, I_{pa} / µA	Cathodic Peak Current, I_{pc} / µA	Anodic Peak Potential, E_{pa} /mV	Cathodic Peak Potential, E_{pc} /mV	ΔE_p *vs* Ag|Ag/Cl / mV	$E_{1/2}$ *vs* Ag|Ag/Cl / mV
1000	0.082	0.0989	113	83	30	98
500	0.0546	0.0548	113	88	25	100.5
250	0.0243	0.0351	108	94	14	101
100	0.0111	0.0128	108	96	12	102
80	0.01	0.0109	109	100	9	104.5

Next, the modified tetrahedra with two or three Fc units were studied, **T5** (Figures 4.17 and 4.18) and **T6** (Figures 4.19 and 4.20). As expected, a strong signal and linear relationship between peak current and potential scan rate was observed in both of them, once again indicating the successful formation of surface-bound assemblies.

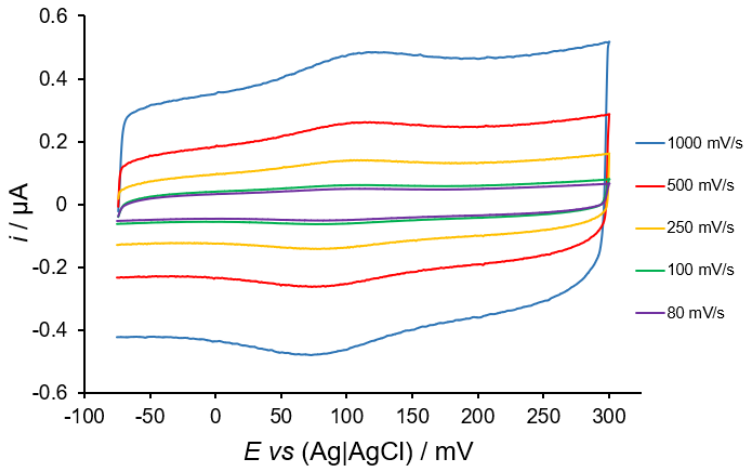


Figure 4.17 CVs of tetrahedron **T5** at various scan rates in 10 mM phosphate buffer (pH 7.0), 1 M $NaClO_4$.

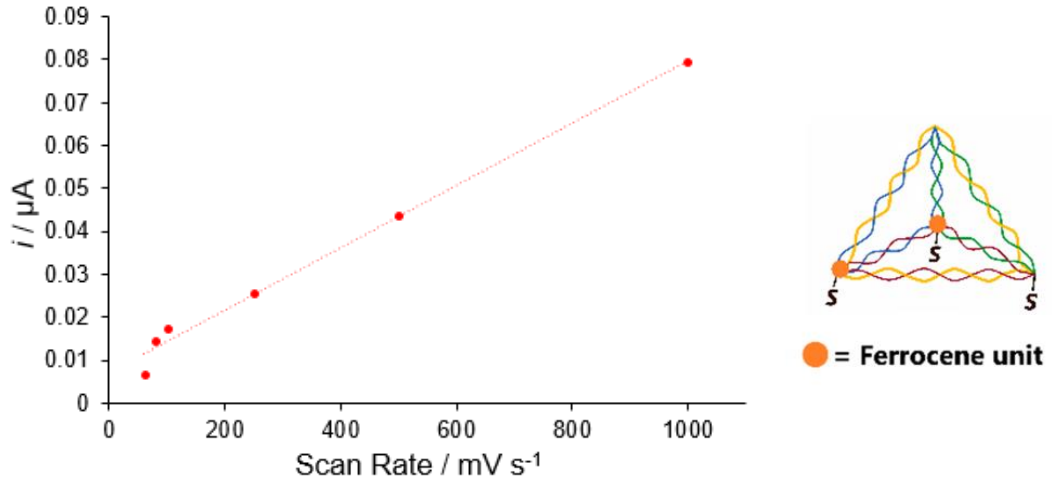


Figure 4.18 Plot of cathodic peak current *vs* scan rate for the **T5** system.

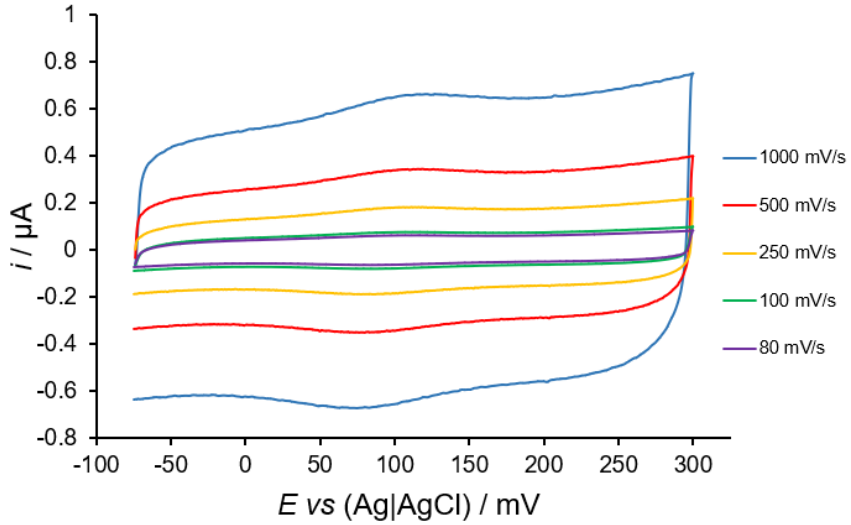


Figure 4.19 CVs of tetrahedron **T6** at various scan rates in 10 mM phosphate buffer (pH 7.0), 1 M $NaClO_4$.

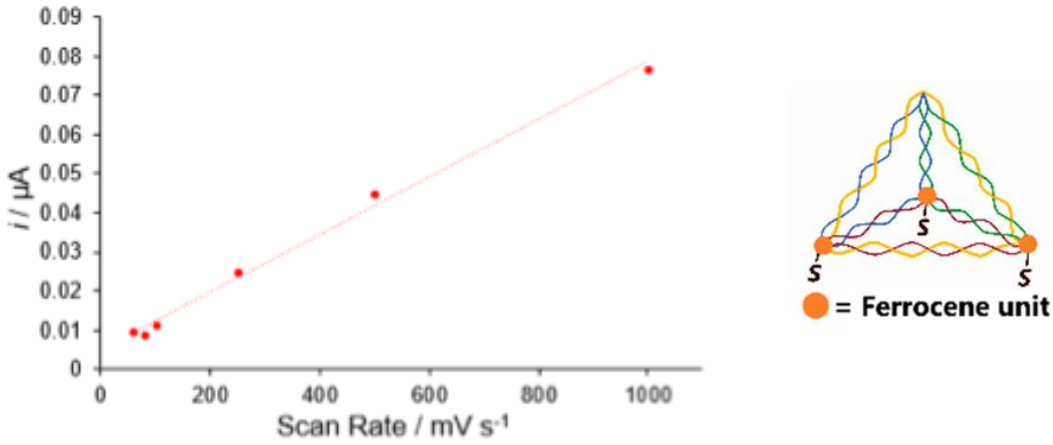


Figure 4.20 Plot of cathodic peak current vs scan rate for the **T6** system.

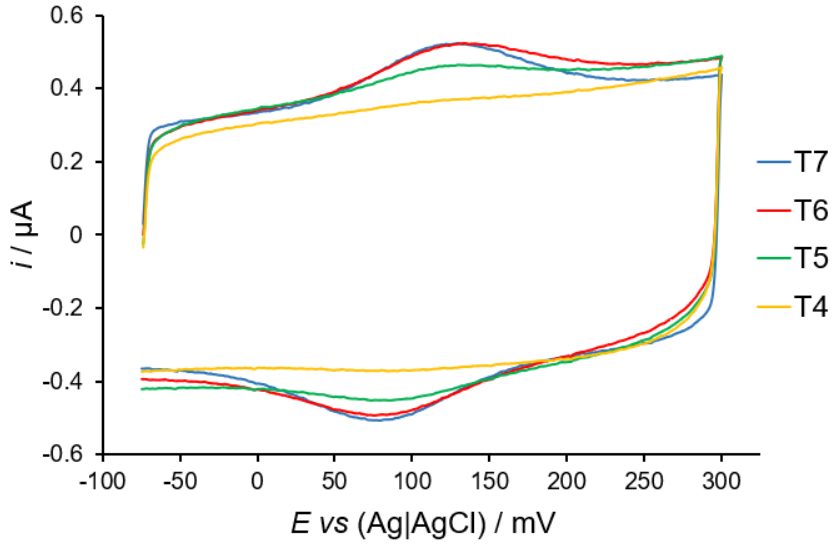


Figure 4.21 CVs of tetrahedra **T4**, **T5**, **T6** and **T7** in 10 mM sodium phosphate buffer (pH 7.0), 1 M $NaClO_4$, scan rate: 1000 mV s^{-1}.

Figure 4.21 presents superimposed CVs of tetrahedra **T4**, **T5**, **T6** and **T7**, with one, two, three and four ferrocene groups, respectively. The figure shows that the currents at the anodic and cathodic peak potentials increase in line with the number of ferrocene units on vertices adjacent to the tetrahedron face lying closest to the electrode surface (i.e. one, two and three in **T4**, **T5** and **T6** respectively). A small difference between **T7**, which has an additional ferrocene unit at its top vertex, and **T6** was also observed, with **T7** exhibiting slightly stronger electrical signals than **T6**. This is in contrast to the comparison between **T3** and **T4**, where **T3**, with only one ferrocene unit in the whole assembly, does not display any prominent signal (Figure 4.13). One potential reason for this difference is that the ferrocene molecules inside the tetrahedron may interact with one another, leading to some form of communication.

Square Wave Voltammetry (SWV) and Differential Pulse Voltammetry (DPV) were also employed to shed further light on these differences. In the case of SWV, a stable signal could not be obtained. However, in the DPV experiment (Figure 4.22), notable differences were once again observed in the signals for **T4**, **T5**, and **T6**. Interestingly, no significant difference between **T6** and **T7** was observed. The conduction of electrons within the system and the communication of signals between different ferrocene entities remain poorly understood, and further in-depth research is needed to explore their underlying mechanisms.

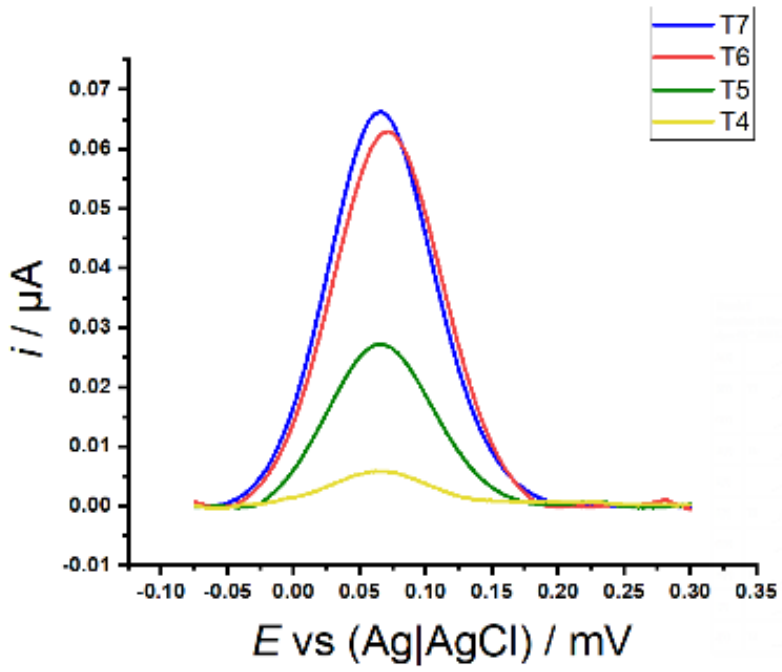


Figure 4.22 DPVs of tetrahedra **T4**, **T5**, **T6** and **T7**. DPVs were recorded at a frequency of 100 Hz, amplitude of 25 mV and steps of 1 mV in 10 mM sodium phosphate buffer (pH 7.0), 1 M $NaClO_4$.

4.4 Conclusion

In this study, ferrocene-modified, redox-active versions of the Turberfield DNA tetrahedron have been successfully synthesised, with the first examples presented of this self-assembled structure with multiple ferrocene units. These unique bioorganometallic assemblies, featuring up to four ferrocene units, were characterized using native polyacrylamide gel electrophoresis. Furthermore, their electrochemical properties on gold electrodes have been investigated, and the dependence of peak current on potential scan rate confirm the successful formation of self-assembled monolayers (SAM).

Notably, the current signals in the cyclic voltammograms exhibited a progressive increase with the number of ferrocene units incorporated into the lower three vertices of the assembly. This observed pattern of behaviour indicates the potential of these redox-active DNA tetrahedra to serve as an electrochemical sensing platform on electrode surfaces. By producing reproducible, well-spaced, and stable scaffolds, they hold promise for applications in various electrochemical sensing and biosensing contexts.

4.5 Future work

In future work, a smaller version of the tetrahedron can be investigated, and eventually, this system can be utilized as a foundation for securely attaching probes at the apex for various electrochemical sensing applications. As an illustration, we have the option to affix a customized DNA probe at the top vertex, incorporating an additional tag, for example, a copper cyclidine.[26] This opens up the possibility of creating a ratiometric sensor that effectively segregates the ferrocene current from the cyclidene current. However, there are still some notable challenges to address, such as enhancing sensitivity and conducting further research on the electron transfer dynamics within this system.

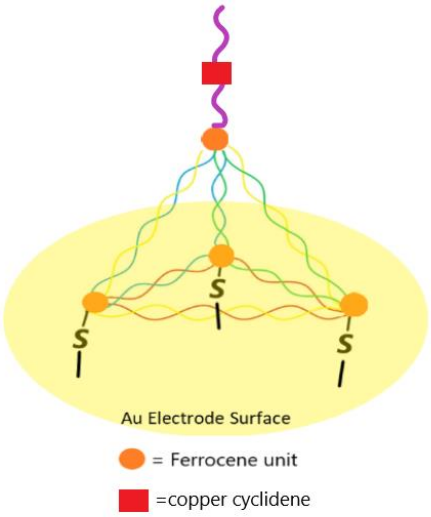


Figure 4.23 Ferrocene tetrahedron platforms with a capture probe at the top vertex.

References

1. Madsen, M.; Gothelf, K. V. *Chemical Reviews*, **2019**, *119*(10), 6384-6458.
2. Chidchob, P.; Sleiman, H. F. *Current Opinion in Chemical Biology*, **2018**, *46*, 63-70.
3. Shih, W. M.; Quispe, J. D.; Joyce, G. F. *Nature*, **2004**, *427*(6975), 618-621.
4. Chen, J.; Seeman, N. C. *Nature*, **1991**, *350*(6319), 631-633.
5. Goodman, R. P.; Berry, R. M.; Turberfield, A. J. *Chemical Communications*, **2004**, *(12)*, 1372-1373.
6. Hu, Y.; Chen, Z.; Zhang, H.; Li, M.; Hou, Z.; Luo, X.; Xue, X. *Drug Delivery*, **2017**, *24*(1), 1295-1301.
7. Kim, K. R.; Kim, D. R.; Lee, T.; Yhee, J. Y.; Kim, B. S.; Kwon, I. C.; Ahn, D. R. *Chemical Communications*, **2013**, *49*(20), 2010-2012.
8. Lee, H.; Lytton-Jean, A. K.; Chen, Y.; Love, K. T.; Park, A. I.; Karagiannis, E. D.; Anderson, D. G. *Nature Nanotechnology*, **2012**, *7*(6), 389-393.
9. Lin, M.; Song, P.; Zhou, G.; Zuo, X.; Aldalbahi, A.; Lou, X.; Fan, C. *Nature Protocols*, **2016**, *11*(7), 1244-1263.
10. Pei, H.; Lu, N.; Wen, Y.; Song, S.; Liu, Y.; Yan, H.; Fan, C. *Advanced Materials*, **2010**, *22*(42), 4754-4758.
11. Lin, M.; Wang, J.; Zhou, G.; Wang, J.; Wu, N.; Lu, J.; Fan, C. *Angewandte Chemie International Edition*, **2015**, *54*(7), 2151-2155.
12. Chen, X.; Zhou, G.; Song, P.; Wang, J.; Gao, J.; Lu, J.; Zuo, X. *Analytical Chemistry*, **2014**, *86*(15), 7337-7342.
13. Ge, Z.; Lin, M.; Wang, P.; Pei, H.; Yan, J.; Shi, J.; Zuo, X. *Analytical Chemistry*,

2014, *86*(4), 2124-2130.

14. Wen, Y.; Pei, H.; Wan, Y.; Su, Y.; Huang, Q.; Song, S.; Fan, C. *Analytical Chemistry*, **2011**, *83*(19), 7418-7423.
15. Zhou, G.; Lin, M.; Song, P.; Chen, X.; Chao, J.; Wang, L.; Zuo, X. *Analytical Chemistry*, **2014**, *86*(15), 7843-7848.
16. Lu, N.; Pei, H.; Ge, Z.; Simmons, C. R.; Yan, H.; Fan, C. *Journal of the American Chemical Society*, **2012**, *134*(32), 13148-13151.
17. Abi, A.; Lin, M.; Pei, H.; Fan, C.; Ferapontova, E. E.; Zuo, X. *ACS Applied Materials & Interfaces*, **2014**, *6*(11), 8928-8931.
18. Gupta, N. K.; Wilkinson, E. A.; Karuppannan, S. K.; Bailey, L.; Vilan, A.; Zhang, Z.; Nijhuis, C. A. *Journal of the American Chemical Society*, **2021**, *143*(48), 20309-20319.
19. Englert, K.; Hendi, R.; Robbs, P. H.; Rees, N. V.; Robinson, A. P.; Tucker, J. H. *Nanoscale Advances*, **2020**, *2*(10), 4491-4497.
20. Duprey, J. L. H.; Carr-Smith, J.; Horswell, S. L.; Kowalski, J.; Tucker, J. H. *Journal of the American Chemical Society*, **2016**, *138*(3), 746-749.
21. Ihara, T.; Sasahara, D.; Shimizu, M.; Jyo, A. *Supramolecular Chemistry*, **2009**, *21*(3-4), 207-217.
22. Leck, G. H. *PhD thesis*, University of Birmingham, University of Birmingham, **2021**.
23. Farjami, E.; Campos, R.; Ferapontova, E. E. *Langmuir*, **2012**, *28*(46), 16218-16226.
24. Boon, E. M.; Jackson, N. M.; Wightman, M. D.; Kelley, S. O.; Hill, M. G.; Barton, J. K. *The Journal of Physical Chemistry B*, **2003**, *107*(42), 11805-11812.

25. Dauphin-Ducharme, P.; Arroyo-Currás, N.; Plaxco, K. W. *Journal of the American Chemical Society*, **2019**, *141*(3), 1304-1311.
26. Duprey, J. L. H.; Carr-Smith, J.; Horswell, S. L.; Kowalski, J.; Tucker, J. H. *Journal of the American Chemical Society*, **2016**, *138*(3), 746-749.
27. Gupta N K, Wilkinson E A, Karuppannan S K, et al. *Journal of the American Chemical Society*, 2021, 143(48): 20309-20319.
28. V. Croué, S. Goeb, M. Sallé, *Chemical Communications*, **2015**, 51, 7275-7289.

Chapter 5 – Studies on DNA Strands Containing Ferrocene Nucleic Acid (FcNA) Units

5.1 Introduction

5.1.1 Design of Ferrocene Nucleic Acid (FcNA)

In 2012, Tucker and co-workers presented the first example of a wholly organometallic structural analogue of DNA, that they termed ferrocene nucleic acid (FcNA).[1] The rationale behind this mimicry originated from the observation that the spacing between consecutive Cp rings in the ferrocene structure (3.3 Å) closely approximates to the distance between stacked base pairs in double-stranded DNA (3.4 Å) (see Figure 5.1).

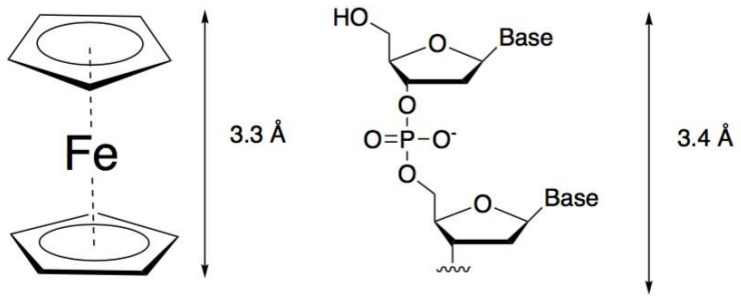


Figure 5.1 Comparison of the length of the ferrocene unit with the combined length of two sugar units in the B-form of DNA.

The tetra-substituted entity (illustrated in Figure 5.2) was strategically devised to emulate two nucleobase units of DNA, featuring a singular ferrocene unit. The key focus and challenge of this research was the synthesis of this ferrocene monomer, which would lay the foundation for exploring the effect of its incorporation into oligomers on duplex structure and stability.

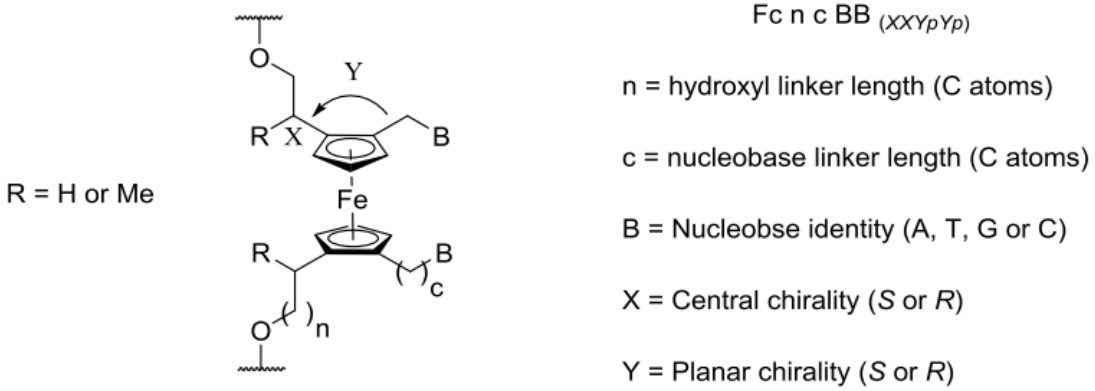


Figure 5.2 The structure of the tetrasubstituted FcNA monomer, which has two centres of chirality when R = H and four centres of chirality when R = Me.[3]

Unpublished work on the previously reported wholly organometallic system featuring eight ferrocene units and 16 thymine bases[1] gave no observable duplex formation under standard aqueous conditions with a complementary polyA 16-mer strand, although weak binding with a PNA polyA analogue was observed.[28] Therefore attention turned towards incorporating just one ferrocene unit at the centre of a DNA strand (i.e. a DNA-FcNA-DNA conjugate) for a systematic study of duplex stability, in a similar approach to that adopted for organic unnatural analogues of DNA. This included studying the effect of changing the nucleobase and changing the chirality by comparing enantiomers of FcNA units within the same duplex. Given DNA's intrinsic chirality, it was anticipated that one enantiomer would exhibit better conformational compatibility for a particular structural constraint (e.g. the B-helix). Another crucial aspect was the evaluation of the best linker length to both the nucleobase and to the phosphate groups, as shown in Figure 5.2. While the distance between the Cp rings of ferrocene closely matches the inter-base distance of DNA, the geometric differences in bonding necessitate some conformational flexibility for optimal fitting of the FcNA monomer within the duplex. A number of previous members of the Tucker group, Dr

John Zhao, Dr Huy Van Nguyen, Dr Jean-Louis Duprey, Dr Antoine Sallustrau, Dr Jonathan Kedge, Dr James Carr-Smith and Dr Media Ismail played key roles in executing this synthesis and performing some initial binding and computational modelling studies.[2][3] These revealed that a conjugate DNA-FcNA-DNA strand (see Table 5.1 for the sequence) containing a ferrocene unit at its centre with *RRS*$_p$*S*$_p$ stereochemistry (the enantiomer of that reported previously[1]) bearing two C3 linkages to the two phosphates and two C2 linkages to the two nucleobases (R = Me, n = 3, c = 2, Fig. 5.2) was found to produce most promising results. Some key findings with this particular ferrocene unit and its abasic counterpart are outlined in the following two sub-sections.

Table 5.1 Oligonucleotide sequences used for previous FcNA studies (ApAp refers to a double abasic site).

	Oligonucleotide sequence
S1TT	5′- TGG ACT C-**TT**-C TCA ATG -3′
S1CC	5′- TGG ACT C-**CC**-C TCA ATG -3′
S1FcHH	5′- TGG ACT C**FcHH**C TCA ATG -3′
S1FcTT	5′- TGG ACT C**FcTT**C TCA ATG -3′
S1FcCC	5′- TGG ACT C**FcCC**C TCA ATG -3′
S2AA	5′- CAT TGA G-**AA**-G AGT CCA -3′
S2GG	5′- CAT TGA G-**GG**-G AGT CCA -3′
S2AG	5′- CAT TGA G-**AG**-G AGT CCA -3′
S2TT	5′- CAT TGA G-**TT**-G AGT CCA -3′
S2APAP	5′- CAT TGA G-**ΦΦ**-G AGT CCA -3′

5.1.2 Molecular dynamic simulations of DNA-FcNA-DNA conjugates

Molecular dynamics (MD) is a computational technique in chemistry and physics used to study the motion and behavior of atoms and molecules over time. It simulates interactions between particles by solving Newton's equations of motion, allowing

researchers to explore the dynamic properties of molecular systems. Initially, the effect of the central chirality of the FcNA system was investigated through molecular dynamic simulations (AMBER) conducted by Dr Jean-Louis Duprey on two duplexes comprised of the strand **S1FcHH** containing either enantiomer of the disubstituted monomer (R = Me, n = 3, see Fig. 5.2) bound to **S2AA** (Figure 5.3).

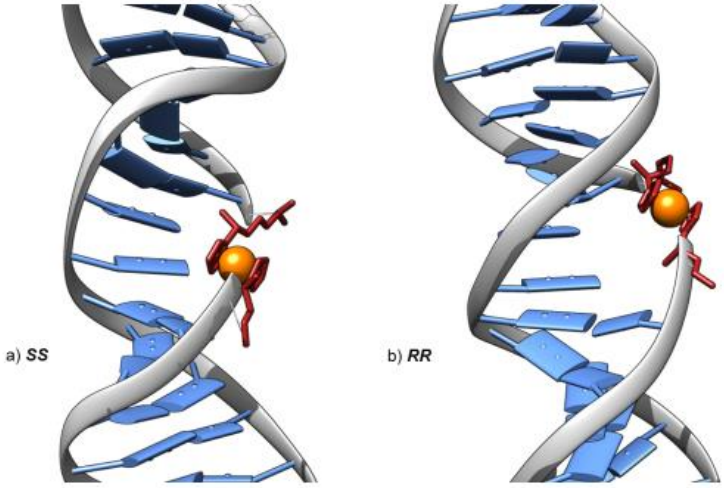


Figure 5.3 Molecular Dynamic simulations of S1FcHH a) *SS* and b) *RR* duplexed with S2AA models created by Dr Jean-Louis Duprey. Taken from Carr-Smith.[2]

In the **FcHH** system, which lacks hydrogen-bonding nucleobases to stabilize the unit, the ferrocene backbone has considerable freedom of movement. This results in distorted and occasionally kinked duplexes capable of adopting various low-energy conformations.[3]

The *SS* isomer demonstrates a deeper penetration into the cavity formed by the absence of nucleobases, causing the Cp rings to orient orthogonally to the base stack. Consequently, the methyl groups are positioned within the hydrophobic core of the duplex. In contrast, the *RR* isomer appears to protrude more from the duplex core, aligning with the helical twist of DNA in a more natural manner; once again, the methyl

groups find their place within the hydrophobic cavity.[2]

MD simulations comparing the duplexes of both stereoisomers of the tetrasubstituted FcNA unit within the **S1TT** system indicate distinctive features. Molecular modelling data supports the involvement of **S1TT**(*RRS*$_p$S$_p$) and **S1TT**(*SSR*$_p$*R*$_p$) in hydrogen bonding. In both instances, the FcNA nucleobases are situated within the duplex with **S2-AA**, and they lie close to coplanar with those of the opposing strand.

Much like the abasic derivatives in the **S1FcHH** system, the *SS* stereoisomer displays a curvature at the ferrocene unit along the backbone, presumably to enhance stacking interactions with the two complementary nucleobases. As a result, the models propose an unwinding and distortion of the flanking bases in the sequence, contributing to an elongation of the duplex. In contrast, the *RR* isomer aligns more closely with the natural twist of the DNA backbone, resulting in less strain on the structure (Figure 5.4).

Figure 5.4 Molecular Dynamic simulations of S1FcTT a) SS and b) RR duplexed with S2AA - models created by Dr. Jean-Louis Duprey Taken from Carr-Smith.[2]

Next a molecular analysis was conducted on the **S1CC(*RRS*$_p$*S*$_p$)** system containing a cytosine nucleobase. It is noteworthy that the fully complementary duplex featuring the **S1FcCC** monomer shows a more uniform helical structure compared to its counterpart with the **FcTT** monomer. In this instance, the C–G base pairs exhibit greater co-planarity than the corresponding T–A base pairs, which can be attributed to the additional hydrogen bond reinforcing their connection. This increased stabilization effect may be further enhanced by the extra hydrogen bond facilitating π-stacking interactions between the layers of nucleobases (Figure 5.5).

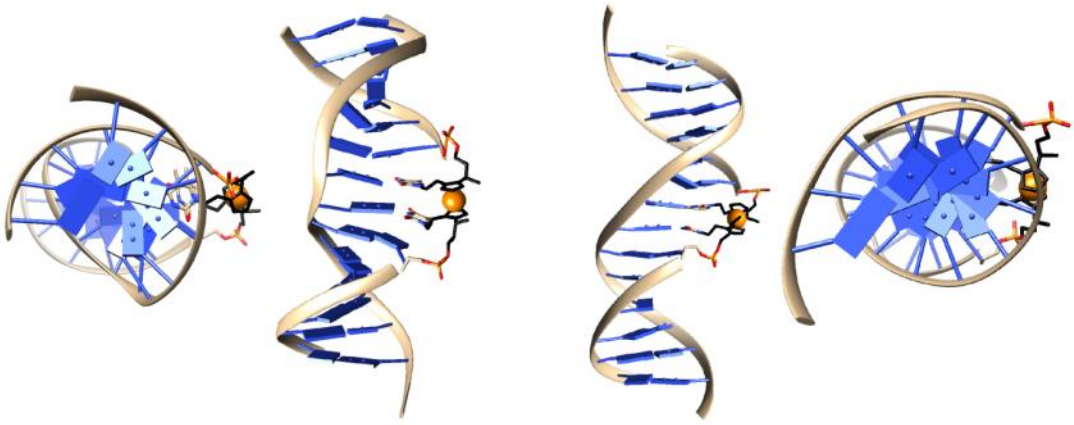


Figure 5.5 Molecular Dynamic simulations of left S1FcTT(RR) and right S1FcCC(RR) duplexed with S2AA/GG - models created by Dr. Jean-Louis Duprey Taken from.[3]

5.1.3 Melting temperature study of DNA-FcNA-DNA conjugates

The duplex T_m values, recorded with UV-vis spectroscopy by Carr-Smith[2], corroborated the data from computer simulations. It was observed that the conjugate strand with the *RRS*$_p$*S*$_p$ isomer exhibited a more stable binding interaction with the complementary strand than the conjugate strand containing its opposite enantiomer.

Different linker lengths to the nucleobase were also investigated but gave lower T_m values (data not shown).

Table 5.2 Thermal Melting Temperatures (T_m) of DNA-FcNA-DNA strands with ferrocene units (RR) incorporated into S1 strands with S2 strands, values are no more than plus/minus 0.5. 5 μM of each strand in 10 mM sodium phosphate buffer pH 7.0, 100 mM $NaClO_4$ Taken from Carr-Smith.[2]

	Thermal Melting Temperature (T_m)/ °C				
	S2AA	**S2AG**	**S2GG**	**S2TT**	**S2APAP**
S1FcHH	37.5	37.5	37.5	36.0	35.5
S1FcTT	43.0	42.0	41.0	40.5	38.5
S1FcCC	39.5	45.5	48.0	37.5	37.0
S1APAP	37.0	37.0	37.0	34.0	--
S1TT	56.0	33.0	44.0	41.5	33.0
S1CC	38.0	50.5	60.0	38.0	32.5

With the optimal stereochemistry of the ferrocene monomer for duplex stability established (i.e. RRS_pS_p), tests were then conducted on this unit to assess the impact of nucleobases on the melting temperatures. The results and trends in the data gave further support to the formation of cross-duplex hydrogen bonds between complementary FcNA and DNA units (Table 5.2). For example, the values for the control abasic disubstituted ferrocene systems were very similar and markedly lower than those for the tetrasubstituted systems containing either cytosine or thymine bases. Importantly, the cytosine strand **S1FcCC** strand formed a duplex with its fully complementary counterpart with a T_m value of 48°C, while the double-mismatch combination（**FcCC** with **S2AA**), had a T_m value of 39.5°C. Additionally, the target featuring only one potential H-bonding base pair (**FcCC** with **S2AG**) showed an intermediate melting temperature of 45.5°C（shown in Figure 5.6). The **FcTT** system demonstrated a similar trend, but with a lower efficiency compared to the **FcCC** system, presumably in part due to the weaker A-T base pairing.

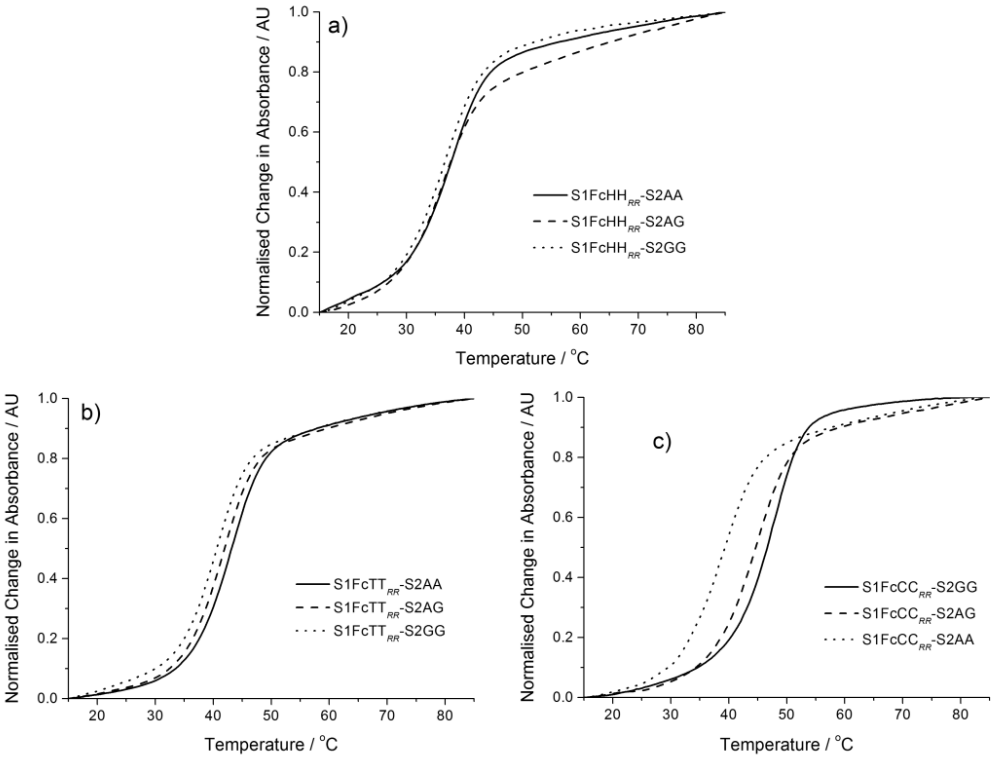


Figure 5.6 Normalised thermal melting profiles of S1(RR) strands with FcNA monomers incorporated into a) **FcHH** b) **FcTT** c) **FcCC**, when duplexed with S1-XX (where XX = AA, AG or GG). 5 μM of each strand in 10 mM sodium phosphate buffer pH 7.0, 100 mM $NaClO_4$. Taken from [2]

5.2 Aim

Through initial screening using MD simulations and subsequent validation via thermal melting studies, these FcNA units have demonstrated the capability to form hydrogen-bonded base pairs when positioned opposite their complementary H-bonding counterparts within a modified DNA sequence. Given these promising results, it was decided to delve deeper into the structural aspects of this DNA-FcNA conjugate system using high field NMR and X-ray diffraction methods. Furthermore, we aimed to explore the biological properties of this system. In particular, through primer extension methods, we planned to test whether polymerase enzymes could recognize the FcNA site within a template to construct DNA strands in-situ. These comprehensive studies would contribute to a more profound understanding of the system's functional behaviour and its potential applications.

5.3 Results and Discussion

5.3.1 Design and synthesis of FcNA

In the previously mentioned study (Section 5.1), a 15/16-mer oligonucleotide extensively investigated by the Tucker Group was employed.[4-6] This sequence, known as the **S1XX** (or NINA) strand within the group, features a central modification flanked by two sequences of 7 bases. These sequences consist of an equal number of A-T and G-C base pairs. Prior research has verified the absence of self-complementary behaviour. Its fully complementary strand, denoted as **S2XX**, involves the insertion of two nucleobases opposite the ferrocene unit, resulting in a 16-mer. Throughout the study, the flanking nucleobases remain constant, with modifications limited to the two central nucleobases (Table 5.3). These strands were utilized for NMR studies.

Table 5.3 Oligonucleotide sequences used for the NMR study

	Oligonucleotide sequence
S1TT	5′- TGG ACT C-TT-C TCA ATG -3′
S1CC	5′- TGG ACT C-CC-C TCA ATG -3′
S1FcHH	5′- TGG ACT CFcHHC TCA ATG -3′
S1FcTT	5′- TGG ACT C**FcTT**C TCA ATG -3′
S1FcCC	5′- TGG ACT C**FcCC**C TCA ATG -3′
S2AA	5′- CAT TGA G-**AA**-G AGT CCA -3′
S2GG	5′- CAT TGA G-**GG**-G AGT CCA -3′

The sequence of the modified templates for the primer extension reaction had previously been used by our collaborator Marcel Hollenstein in investigating the incorporation of C-nucleotides into DNA.[7][8] The modified templates (**M1**) were designed to include one ferrocene unit or another control modification unit, positioned

following the 3′-terminus of the 19-nucleotide-long, 5′-FAM-labelled primer **P1** (given in Table 5.4).

Table 5.4 Oligonucleotide sequence for primer extension reactions

	Oligonucleotide sequence
P1	5′ – TAC GAC TCA CTA TAG CCT C – 3′
M1Fc	5′ – GGA G–**Fc**–G AGG CTA TAG TGA GTC GTA – 3′
M1FcHH	5′ – GGA G**FcHH**G AGG CTA TAG TGA GTC GTA – 3′
M1FcTT	5′ – GGA G**FcTT**G AGG CTA TAG TGA GTC GTA – 3′
M1FcCC	5′ – GGA G**FcCC**G AGG CTA TAG TGA GTC GTA – 3′
AP1	5′ – GGA GΦG AGG CTA TAG TGA GTC GTA – 3′
AP2	5′ – GGA GΦΦG AGG CTA TAG TGA GTC GTA – 3′
M1Sp18	5′ – GGA G**Sp18**G AGG CTA TAG TGA GTC GTA – 3′

All these sequences were synthesized through DNA solid-phase synthesis. Subsequently, all modified sequences underwent purification using HPLC and were identified with mass spectrometry (see Experimental chapter).

In addition to the three chiral ferrocene units (**FcHH, FcTT, FcCC)** discussed earlier that had *RR* central chirality (and also S_pS_p planar chirality in the case of the **FcNA** units), another achiral ferrocene unit (**Fc**) was also employed for the PEX studies, serving as an additional control group (see Figure 5.7). Finally, templates that incorporated either one or two tetrahydrofuran-AP sites (denoted as **AP2**) were synthesized using the corresponding commercially available phosphoramide building block. Additionally, a template containing an 18-atom hexa-ethyleneglycol Sp18-spacer unit, **M1-Sp18**, was synthesized.

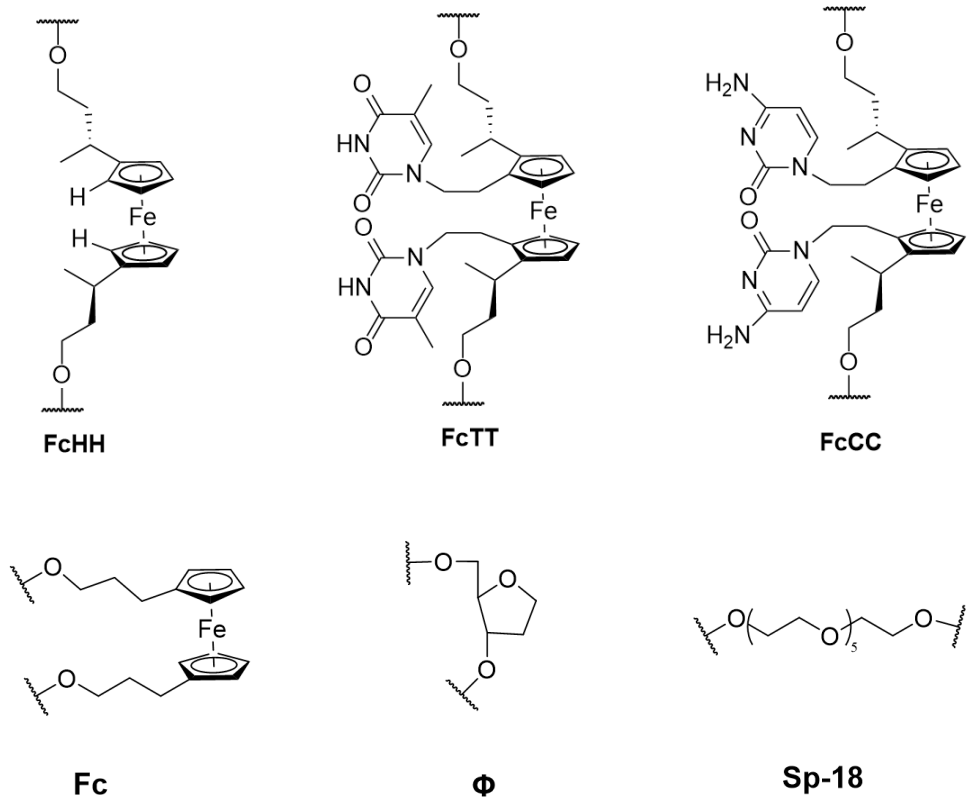


Figure 5.7 Structures of the **FcNA** and abasic units used in sequences for the NMR studies and the PEX reactions. The stereochemistry of the **FcHH** unit is *RR* and that of the **FcNA** units is RRS_pS_p.

5.3.2 Primer extension (PEX) reactions

To explore the biological properties of ferrocene-modified oligonucleotides, one viable approach involves their synthesis through the polymerase-catalysed incorporation of unnatural nucleoside triphosphates ((d)N*TPs).[9] This method stands out for its compatibility with the identification of functional nucleic acids and its lack of restrictions regarding size or the nature of introduced functional groups. Modified oligonucleotides can be produced either through primer extension (PEX) reactions (i.e one run) or under PCR amplification conditions (i.e. multiple cycles) by substituting one or multiple canonical nucleotides with corresponding unnatural counterparts. Natural or

engineered polymerases display considerable tolerance to nucleotides carrying modifications at position C5 of pyrimidine or C7 of N7-deazapurine nucleotides, or those with straightforward backbone modifications.[12-15] This adaptability extends to incorporating nucleotides with sugar or unnatural base surrogates, enhancing the method's versatility in investigating the biological properties of ferrocene-modified oligonucleotides. The studies described below in Sections 5.3.2.1 and 5.3.2.2 were performed by collaborators in the Hollenstein group, whereas those described in later sections (Sections 5.3.2.3 and 5.3) were undertaken in partnership with co-workers during a visit to the Hollenstein laboratory in Paris.

5.3.2.1 PEX with template containing achiral ferrocene (Fc)

The initial investigation focused on assessing the feasibility of incorporating canonical nucleotides opposite an achiral ferrocene (**Fc**) site within the template **M1Fc** (Table 5.4). Primer extension (PEX) reactions were conducted utilizing various polymerases (A, B, and Y families) and individual nucleoside triphosphates (dNTPs) within the P1/M1 primer/template system (Figure 5.8). Gel electrophoresis analysis of the reaction products revealed that all natural nucleotides could be successfully incorporated opposite a Fc abasic site analogue. Higher incorporation efficiencies were observed for the larger purine nucleotides. Notably, most polymerases efficiently incorporated one dA and one dG opposite the **Fc** site, resulting in a conversion of the primer to the corresponding n + 1 products. Therminator, however, exhibited a unique behavior by incorporating two dA nucleotides, leading to the formation of an n+2 product with a high yield. In contrast, the yields of n + 1 product formation for pyrimidines were notably lower. Sulfolobus DNA polymerase IV (Dpo4), a Y family DNA polymerase known for its lesion-bypass capability, demonstrated multiple incorporation events with all nucleotides except dGTP (lane 7 in Figure 5.8). Interestingly, reactions conducted with Deep Vent resulted in the complete hydrolytic degradation of the primer (lane 8 in Figure 5.8).

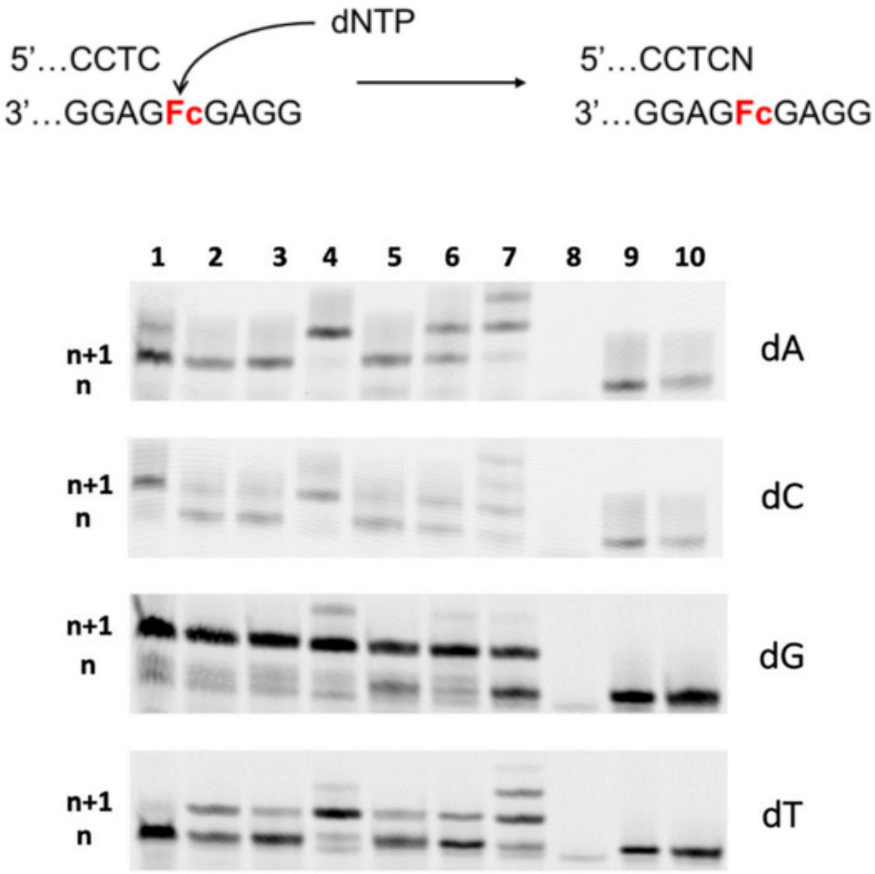


Figure 5.8 Gel images (PAGE 20%) of PEX reactions with primer **P1** (10 pmol) and template **M1Fc** (15 pmol) in the presence of the individual natural dNTPs(100 μM) and different DNA polymerases. Reactions carried out for 1 h List of polymerases used: 1. Hemo Klem Taq (8 reactions), 2. Taq (5 U), 3. Bst (8 U), 4. Therminator (2 U), 5. Vent (exo−) (2 U), 6. Kf exo− (5 U), 7. Dpo4 (2 U), 8. Deep Vent (2 U). The negative controls are: 9. No polymerase, 10. No dNTP. The label n is connected to the length of the primer, while n + 1 is referred to the length of the primer with the addition of one dNMP.

In order to evaluate the efficiency and specificity of incorporating various natural nucleotides, the reaction times for purine nucleotides were modified, shortening them to 30 minutes, while extending the reaction times for pyrimidine nucleotides to 2 hours (Figure 5.9). The reduction in reaction time had minimal impact on the yield of reactions involving A and G, whereas the extension of reaction time slightly improved the formation of n + 1 products with C and T nucleotides. Overall, altering the reaction times had limited influence on the outcomes.

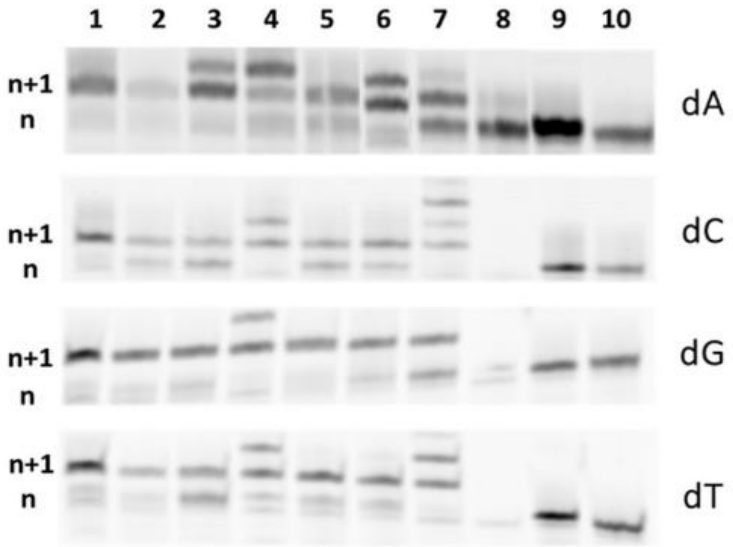


Figure 5.9 Gel images (PAGE 20%) of PEX reactions with primer **P1** (10 pmol) and template **M1Fc** (15 pmol) in the presence of the individual natural dNTPs (100 μM) and different DNA polymerases. Reactions carried out for 30 min with dATP and dGTP and for 2 h with dCTP and dTTP. List of polymerases used: 1. Hemo Klem Taq (8 reactions), 2. Taq (5 U), 3. Bst (8 U), 4. Therminator (2 U), 5. Vent (exo−) (2 U), 6. Kf exo− (5 U), 7. Dpo4 (2 U), 8. Deep Vent (2 U). The negative controls are: 9. No polymerase, 10. No dNTP. The label n is connected to the length of the primer, while n + 1 is referred to the length of the primer with the addition of one dNMP.

Next, we conducted a time-course experiment to assess the relative efficiency of incorporating each individual nucleotide. Using Vent (exo−) polymerase and the **P1/M1Fc** primer/template system (Figure 5.10), we observed that the relative efficiency of incorporating canonical nucleotides opposite a templating ferrocene nucleotide followed the order dA > dG ≫ dC ~ dT. This is known as the "A rule," indicating a strong preference of polymerases to incorporate a dA nucleotide opposite an abasic site.[16] It seems to hold true in the context of an **Fc** nucleotide as well.

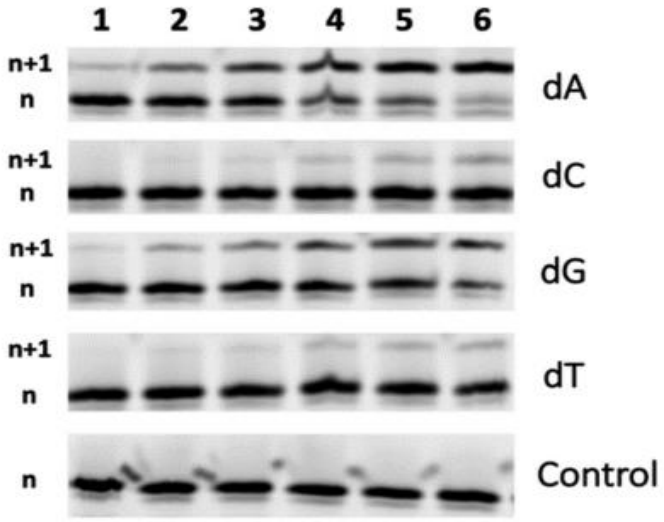


Figure 5.10 Gel images (PAGE 20%) of PEX reactions with primer **P1** (10 pmol) and template **M1Fc** (15 pmol) in the presence of the individual natural dNTPs (100 μM) and Vent (exo−) DNA polymerase (0.5 U). Time points were taken after 1 min (1), 5 min (2), 10 min (3), 20 min (4), 30 min (5), 60 min (6).

5.3.2.2 PEX templates containing organic abasic sites (AP)

Apurinic/apyrimidinic (AP) sites, commonly referred to as abasic sites, are widespread DNA lesions that occur frequently in mammalian DNA due to the hydrolysis of the glycosidic bond.[17-18] Beyond their biological and clinical relevance, largely stemming from their inherently highly mutagenic nature, AP sites also offer interesting possibilities for modifying tools in chemical biology. In this context, we substituted ferrocene with two tetrahydrofuran moieties as an organic control, giving the strand **AP2**.

When dATP was utilized as the substrate for the PEX reaction, all polymerases inserted at least one nucleotide opposite the two AP-site nucleotides in **AP2**. Interestingly, Bst, Therminator, and Kf exo− polymerases incorporated two nucleotides (Figure 5.11). Regarding dGTP, most polymerases inserted single nucleotides,

resulting in the conversion of the primer to n + 1. Only Therminator polymerase exhibited n + 2 product formation, albeit with a very low yield. Additionally, a n + 2 product could be observed when pyrimidines were used with Therminator. Similar results were noted with template **AP1**, which has only one abasic site.[10]

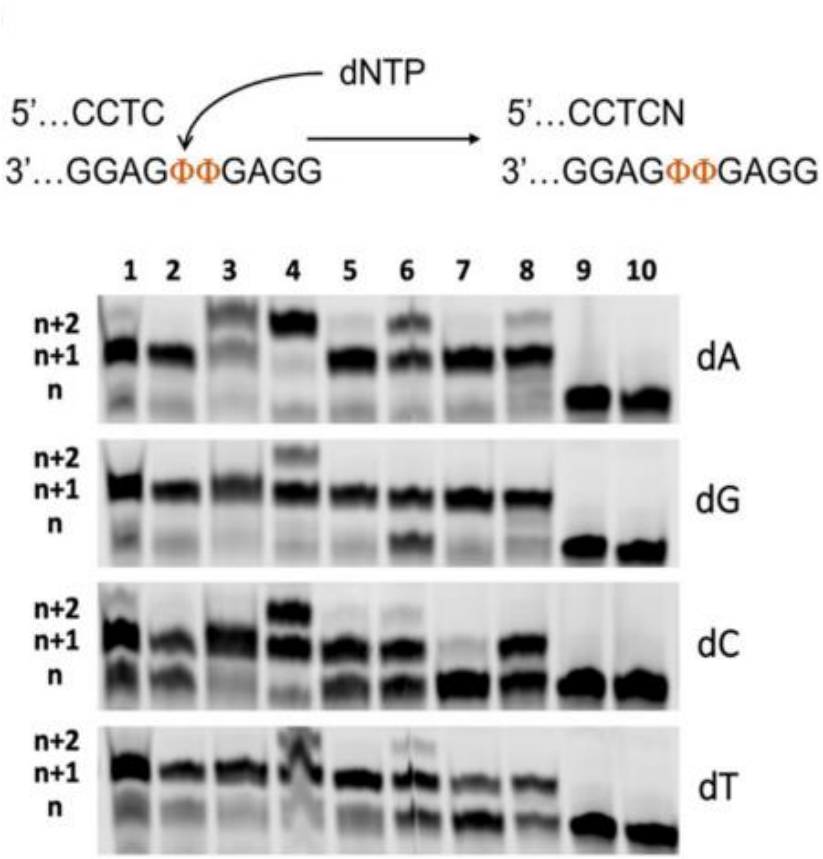


Figure 5.11 Gel images (PAGE 20%) of PEX reactions with primer **P1** (10 pmol) and template **AP2** (15 pmol) with each individual Dntp 100 µM. List of polymerases used: 1. Hemo Klem Taq (8 reactions), 2. Taq (5 U), 3. Bst (8 U), 4. Therminator (2 U), 5. Vent (exo−) (2 U), 6. Kf exo− (5 U), 7. Dpo4 (2 U), 8. Deep Vent (2 U). The negative controls are: 9. No polymerase, 10. No dNTP. The label n is connected to the length of the primer, while n + 1 is referred to the length of the primer with the addition of one dNMP while n + 2 is referred to the length of the primer with the addition of two dNMPs.

Unexpectedly, when an 18-carbon hexaethylene glycol spacer moiety **Sp-18** (Figure 5.12) was introduced into the template at the same modification site, we observed the formation of n + 1 product for most of the polymerases and an n + 2 product for

Therminator, instead of the termination of polymerase-mediated synthesis (Figure 5.11). It is proposed that this is triggered by non-templating addition (NTA) of nucleotides at the 3′-end of the primer, rather than adhering to the A rule-type incorporation. When DNA or RNA polymerases reach the end of DNA templates, they have a tendency to introduce one or multiple adenosine nucleotides. NTA has been recognized as a factor affecting the enzymatic synthesis of DNA and RNA, and various inhibitors[19-20] have been recommended to alleviate NTA. Additionally, NTA is likely facilitated by a strand slippage mechanism, where misaligned primers induce polymerases to slip, leading to the creation of indel mutations.[21] This mechanism has recently been explained in the context of both modified and natural nucleotides on templates that contain longer homopolymeric sequences, as well as shorter polyethylene glycol spacers.[11]

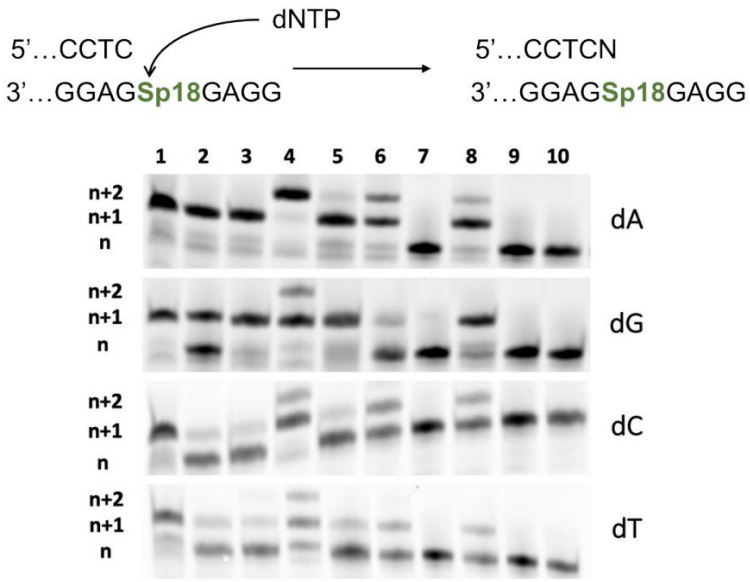


Figure 5.12 Gel images (PAGE 20%) of PEX reactions with primer **P1** (10 pmol) and template **Sp-18** (15 pmol) with each individual dNTP (100 μM). List of polymerases used: 1. Hemo Klem Taq (8 reactions), 2. Taq (5 U), 3. Bst (8 U), 4. Therminator (2 U), 5. Vent (exo−) (2 U), 6. Kf exo− (5 U), 7. Dpo4 (2 U), 8. Deep Vent (2 U). The negative controls are: 9. No polymerase, 10. No dNTP. The label n is connected to the length of the primer, while n + 1 is referred to the length of the primer with the addition of one dNMP while n + 2 is referred to the length of the primer with the addition of two dNMPs.

5.3.2.3 PEX with templates containing FcHH and FcNA units

Next, the chiral ferrocene **FcHH** (with *RR* stereochemistry in accordance with the preliminary work described in Section 5.1) was studied, placing the unit at the same location as before, giving the strand **M1FcHH** (Table 5.4). The results once again demonstrate the A rule, indicating a strong preference of polymerases to incorporate a dA nucleotide opposite it (Figure 5.13). Notably, when Bst, Klenow, and Dpo4 were applied, a n+2 band was observed, while other polymerases resulted in the formation of an n + 1 product.

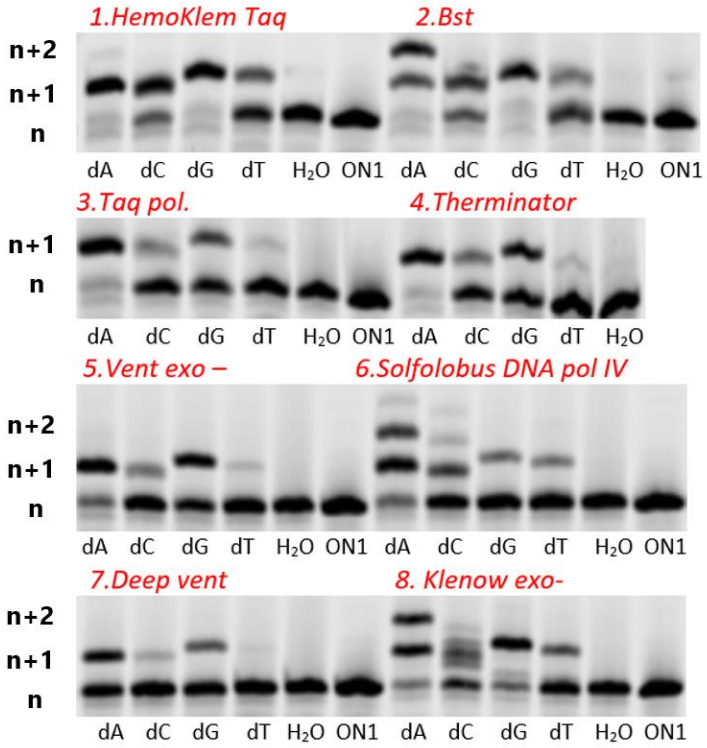


Figure 5.13 Gel images (PAGE 20%) of PEX reactions with primer **P1** (10 pmol) and template **M1FcHH** (15 pmol) with each individual dNTP (100 μM, 60 min). List of polymerases in red: 1. Hemo Klem Taq (8 reactions), 2. Bst 3. Taq 4. Therminator (2 U), 5. Vent (exo−) (2 U), 6. Dpo4 7. Deep Vent, 8.Kf exo−.

In the case of **M1FcTT**, we anticipated the formation of more n+1 and n+2 products for different polymerases when ATP was used. The formation of the n+2 product was indeed more successful when Therminator was involved (Figure 5.14) compared to

M1FcHH. However most polymerases were, if anything, more sluggish, incorporating mainly just one dA opposite the **FcTT** site, for example in the case of Bst and Klenow. That the overall efficiency of incorporating nucleobases (in particular pyrimidines) was lower than in the **FcHH** template case, can be possibly attributed to further functionalisation of the ferrocene hindering access of the polymerase under the conditions used.

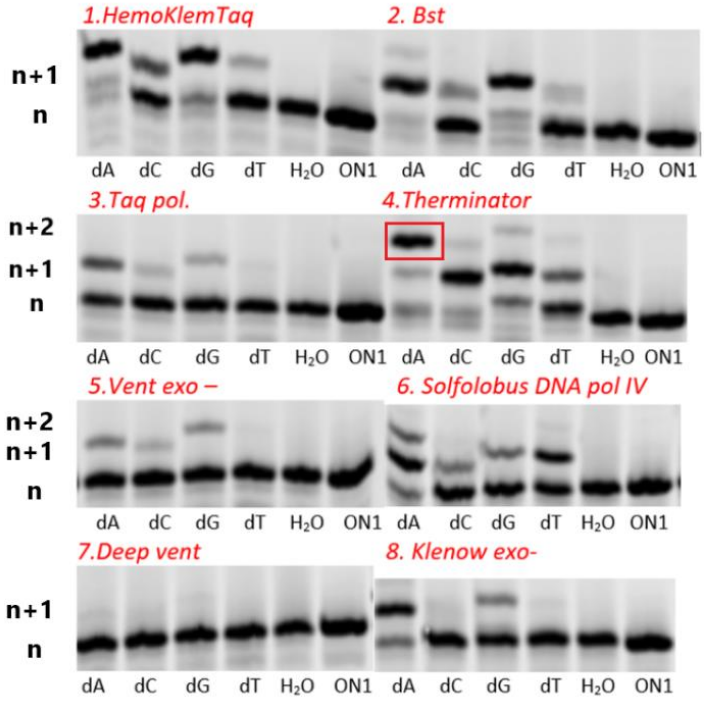


Figure 5.14 Gel images (PAGE 20%) of PEX reactions with primer **P1** (10 pmol) and template **M1FcTT** (15 pmol) with each individual dNTP (100 μM, 60 min). List of polymerases in red:. 1. Hemo Klem Taq (8 reactions), 2. Bst 3. Taq 4. Therminator (2 U), 5. Vent (exo−) (2 U), 6. Dpo4 7. Deep Vent, 8.Kf exo−.

For strand **M1FcCC,** the results were similar to those for **M1FcTT** in that lower nucleotide incorporation was observed in the n+1 and n+2 positions compared to the less hindered template **FcHH**. And once again, for most polymerases, while an n+1 product was observed for purines, yields for pyrimidines were generally lower. However Therminator was once again the exception to the general observation of lower incorporation for the more hindered FcNA templates, giving higher levels of

incorporation. Furthermore it displayed a unique behaviour in incorporating two dG nucleotides opposite the **FcCC** unit, with the n+2 product forming in high yield (Figure 5.15). Although this phenomenon was not observed for the **M1FcTT** template, the n+2 product was also observed when dA was used, following the A rule.

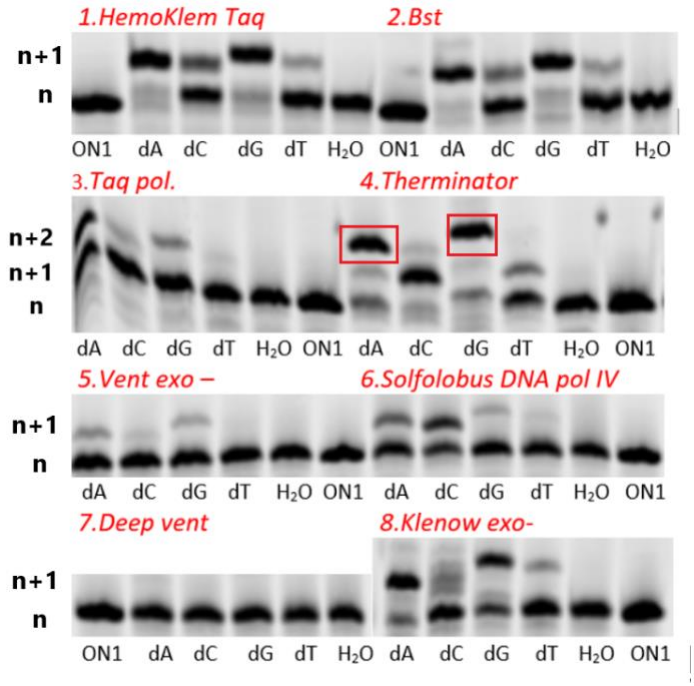


Figure 5.15 Gel images (PAGE 20%) of PEX reactions with primer **P1** (10 pmol) and template **M1FcCC** (15 pmol) with each individual dNTP (100 μM, 60 min). List of polymerases in red:. 1. Hemo Klem Taq (8 reactions), 2. Bst 3. Taq 4. Therminator (2 U), 5. Vent exo− (2 U), 6. Dpo4 7. Deep Vent, 8.Kf exo−.

In summary, the use of a tetrasubstituted **FcNA** unit within a template strand leads to generally lower levels of polymerase-mediated NTP incorporation compared to bis-substituted ferrocene units, be they chiral (**FcHH**) or achiral (**Fc**). This can be explained by increased steric hindrance impeding the reaction. However the polymerase therminator does not follow this trend, with higher levels of incorporation of two purine nucleotides (n+2 position) compared to the chiral **FcHH** control strand. Furthermore, while ATP was readily incorporated by both FcNA strands, a degree of selectivity was indeed observed for the **FcCC** system, with the incorporation of dG being the most favoured over the other three canonical bases.

5.3.3 Extending DNA synthesis – Bypass experiments

With the PEX studies showing the effect of incorporations of different nucleobases direct opposite the FcNA site in the template strand (i.e. at the n+1 and n+2 positions), the next task was to explore the possibility of extending DNA synthesis beyond the ferrocene site using so-called bypass experiments, to further probe the extent of the functional biological behaviour of DNA-FcNA hybrid oligonucleotides.

5.3.3.1 Polymerase selection and studies on M1Fc

The achiral ferrocene modified **M1Fc** was employed for selecting the best polymerase for the bypass experiments. Initially, using PEX reactions, each canonical nucleotide was individually introduced opposite **Fc**, and the resulting extended primer was subsequently subjected to incubation with dCTP in the presence of various polymerases (Figure 5.16). Dpo4, known for its lesion bypass capability,[10] demonstrated the ability to incorporate one or two dCs after double (i.e. n+2) dN–Fc pair, where dN is dA, dC, or dT, with the exception of dG. Given this result, Dpo4 was selected for further investigation.

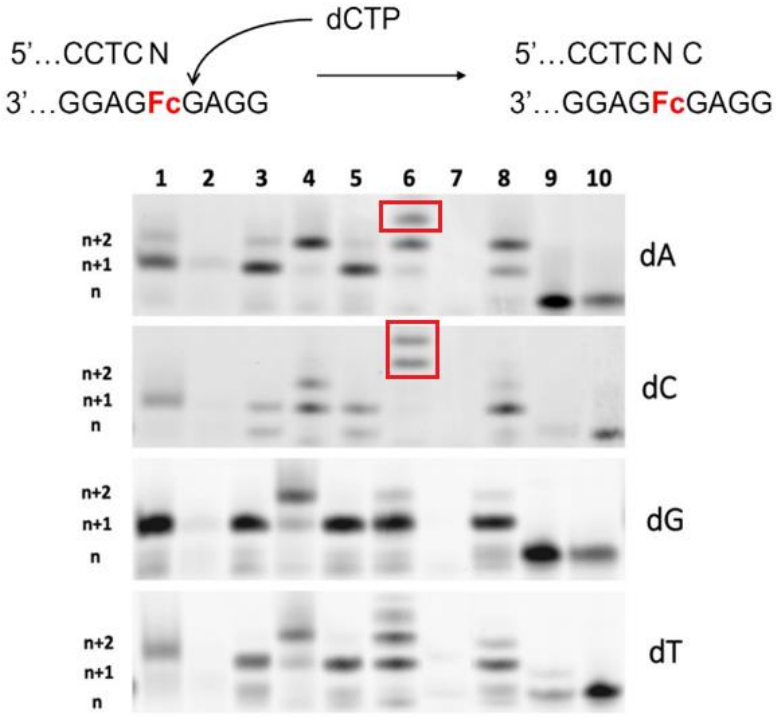


Figure 5.16 Gel image (PAGE 20%) of PEX reactions with primer **P1** (10 pmol) and template **M1Fc** (15 pmol). A first nucleotide was incorporated opposite Fc by PEX reactions with dATP (100 µM, 20 min), dGTP (100 µM, 60 min), dCTP (100 µM, 120 min), or dTTP (100 µM, 120 min). After installation of a first nucleotide, the extended product was incubated with dCTP (20µM, 60 min). List of polymerases used: 1. Hemo Klem Taq (8 reactions), 2. Taq (5 U), 3. Bst (8 U), 4. Therminator (2 U), 5. Vent (exo-) (2 U), 6. Dpo4 (2 U), 7. Deep Vent (2 U) 8. Kf exo-(5 U). The negative controls are: 9. No polymerase, 10. No dNTP. The label n is connected to the length of the primer, while n+2 is referred to the length of the primer with the addition of two dNMPs.

In a subsequent experiment with this primer, either a dG or a dA nucleotide was initially incorporated through a reaction involving Dpo4, and the resulting product was then extended by incubation with a combination of dCTP and dTTP (Figure 5.17).

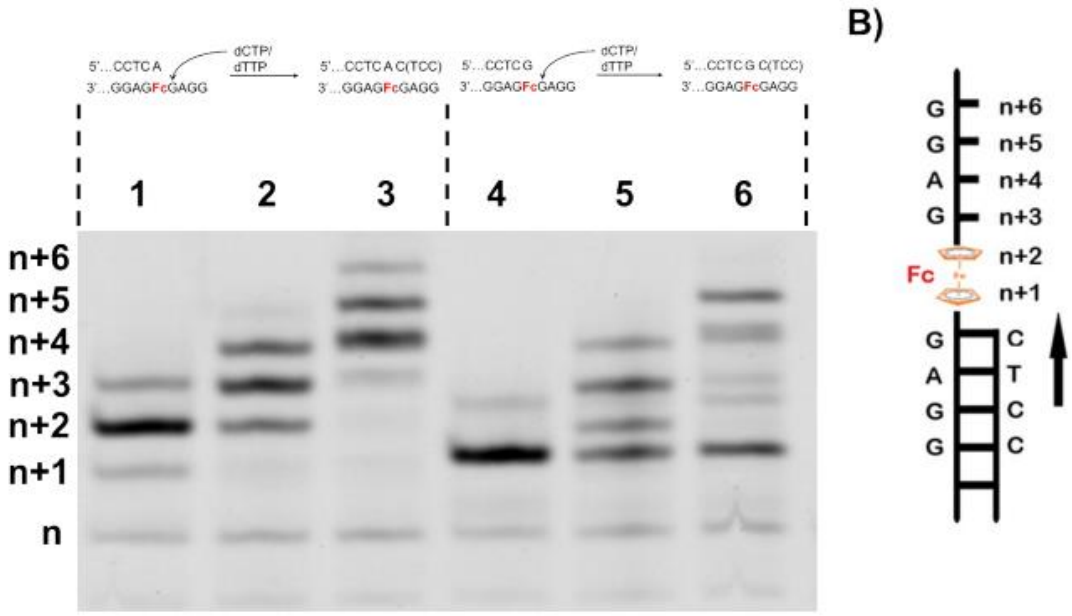


Figure 5.17 A) Gel image (PAGE 20%) of PEX reactions with primer **P1** (10 pmol) and template **M1Fc** (15 pmol) and Dpo4. A first nucleotide was incorporated opposite Fc by PEX reactions with dATP (100 μM, 20 min, lane 1) or dGTP (100 μM, 60 min, lane 4). After this first addition, dCTP (20 μM, 60 min, lanes 2 and 5) or dCTP and dTTP (20 μM, 60 min, lanes 3 and 6) were added. The label n is connected to the length of the primer, while n+2 is referred to the length of the primer with the addition of two dNMPs, n+3 for three additions, n+4 for four and n+5 for five added nucleotides. B) Schematic representation for a tentative explanation for the formation of the n+6 product.

These experiments confirmed that Dpo4 can indeed resume DNA synthesis beyond what is largely a dA–Fc double pair (lane 1), resulting in the observation of a mixture of n+3, n+4, n+5 and n+6 products (lanes 2 and 3), with greater extensions for a mixture of dC and dT, as expected. However, in the case of adding dG first in the initial PEX (lane 4) just a single (n+1) dG-Fc pair was the main product, in keeping with dA being more readily incorporated by PEX opposite ferrocene. As a result, while extension towards the full length products (n + 6) could be achieved with subsequent addition of dC and dT, the efficiency was compromised in terms of product yield.

5.3.3.2 Bypass experiments on templates containing FcHH and FcNA units

In the final experiment (Figure 5.18), the bypass was conducted on the FcNA-DNA hybrid strands **M1FcTT** and **M1FcCC** as well as the control strand **M1FcHH**. Due to the effectiveness in which the enzyme Therminator had incorporated two nucleobases opposite the ferrocene unit with a degree of selectivity in the previous PEX studies (Section 5.3.2), this enzyme was used for the initial stage, before incubation with NTPs in the presence of Dpo4.

Starting with the **M1FcTT** system, two dAs were introduced opposite the ferrocene through the first Therminator-catalysed PEX reaction (Lane 1), and the resulting product was then extended by incubating it with both dCTP and dTTP for 2 hours (Lane 2). The most intense band observed seen after the addition of both dCTP and dTTP was the expected full-length product (n+6, L2), followed by the n+7 product. The latter may result from the previously mentioned non-templated addition (NTA). Both the initial PEX reaction and the subsequent bypass experiment were also conducted using only one polymerase, Dpo4, showcasing its ability to incorporate one or two dATPs after the double FcNA-DNA base pair (L4). A +6 product was also achieved in the presence of both dCTP and dTTP (L5) but the yield was lower, presumably in part due to base-pair mismatches with dA (G-A and T-A) at the n+3 and n+4 positions arising at the PEX stage.

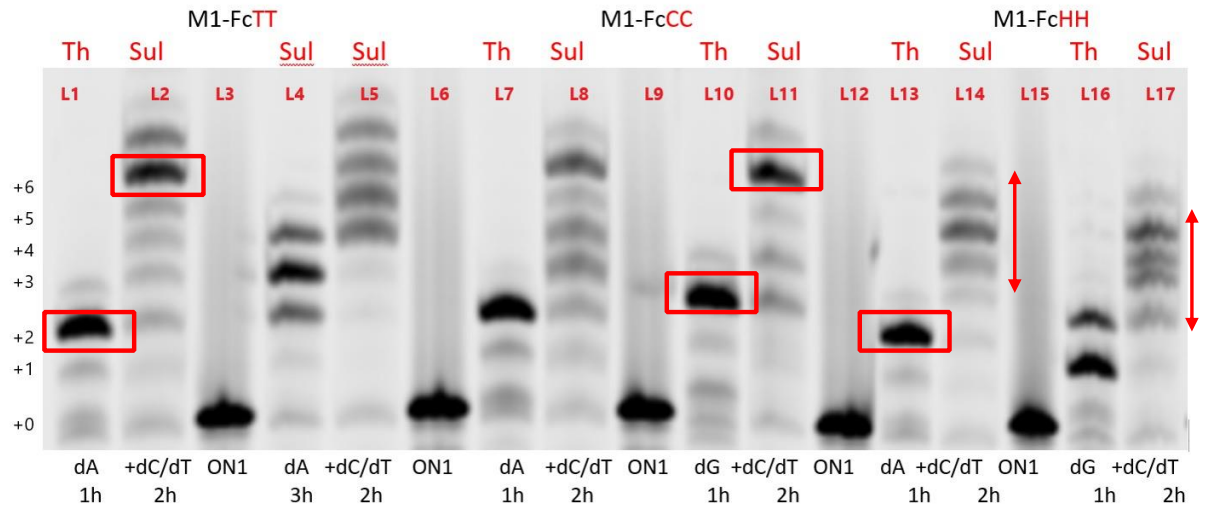


Figure 5.18 Gel image (PAGE 20%) of PEX reactions with primer **P1** (10 pmol) and template **M1FcTT** (L1-L5), **M1FcCC** (L6-L11), **M1FcHH** (L12-L17) (15 pmol) and Therminator or Dpo4(Sul). A first (n+1) and/or second (n+2) nucleotide was incorporated opposite the Fc unit by PEX reactions with dATP (100 µM, 60 min, L1, L4(180 min), L7, L13) or dGTP (100 µM, 60 min, L10, L16). After this first addition, dCTP and dTTP (20 µM, 120 min) were added (L2, L5,L8, L11, L14, L17).

For the **M1FcCC** template system, either two dATPs or two dGTPs were initially incorporated through the Therminator-catalysed PEX reaction, as shown in lanes 7 and 10, respectively. Subsequently, incubation with dCTP and dTTP together was performed using the Dpo4 polymerase, resulting once again in strand extension, including the fully extended n+6 product (L8 and L11). However, whereas multiple products were observed in extensions beyond the double mismatch FcCC-AA base pair (L8), a much cleaner full length (n+6) product in high yield was observed for the strand arising from the cognate FcCC-GG base pair (L11), indicating a successful read through using the polymerase Dpo4 under these conditions.

Similar studies with the control template strand **M1FcHH** gave the expected n+2 product from the initial PEX reaction with dATP using the polymerase Therminator (L13), with a less efficient result found with dGTP (L16). However as expected, the

subsequent bypass experiments gave low yielding strand extensions and a mixture of different strand lengths (L14 and L17), which was consistent with poorer duplex stabilities arising from using a ferrocene unit in the template strand with no nucleobases attached to it.

In conclusion, a series of PEX studies has demonstrated the ready incorporation of dNTPs opposite a ferrocene unit placed within the backbone of a template strand. Where the ferrocene is unsubstituted with nucleobases (**M1Fc** or **M1FcHH** strands), ATPs are favourably incorporated, in keeping with the A rule. For **FcNA** template systems containing an additional two T or two C bases on each ferrocene unit, similar results are observed, albeit with more sluggish nucleotide incorporations, presumably for steric reasons. However some selectivity is observed using the polymerase Therminator. Furthermore, bypass experiments have revealed the remarkable ability of the **FcNA-DNA** hybrid template strands to facilitate DNA primer extension, with strands containing **FcCC** and **FcTT** units performing better than the **FcHH** control. Notably, the sequence-specific effects are highlighted by the observation that the presence of a **FcCC** unit within a template strand leads to effective strand extension when the primer is first incubated with GTP. These outcomes demonstrate a notable advantage in using the tetrasubstituted **FcNA** unit compared to the **FcHH** control group. This first example of a promising biological function of this abiotic organometallic moiety within an oligonucleotide is entirely consistent with the formation of cognate H-bonding base pairs between FcNA and DNA bases across a nucleic acid duplex, as indicated by previous computer simulations and melting temperature studies.

5.3.4 NMR Studies on FcNA-DNA hybrid strands

Nuclear magnetic resonance (NMR) has established itself as an important technique in the examination of nucleic acids, with a specific emphasis on DNA. The reasons for employing NMR to investigate DNA oligonucleotides include straightforward assignment of resonances for exploring interactions with ligands, deducing topology through sequential assignment, and accomplishing complete three-dimensional structure determination.[22]

A straightforward one-dimensional NMR spectrum of the folded oligonucleotide in $^{1}H_2O$, under the desired cation and pH conditions, enables the observation of fast-exchanging imino protons of G and T, as well as the amino protons of G, A, and C. Optimal detection of exchangeable protons occurs at low temperatures, often measurable below 0 °C even at low salt concentrations, although resonance lines may broaden. The folded structure of the oligonucleotide is explored at a designated sample temperature in a $^{1}H_2O/^{2}H_2O$ solution (90-95%/5-10%) with an appropriate concentration of inorganic salts and buffer conditions. A one-dimensional spectrum, employing an effective water suppression system, facilitates the observation of imino protons from the four nucleobases, which are protected from rapid exchange with the bulk solvent due to the hydrogen bonding between them. In a typical proton spectrum, the G-H1 in a G-C Watson-Crick base pair is evident at a chemical shift (δ) of 12-13 ppm, while T-H3 in a T-A base pair appears at a lower field with δ13-14 ppm (See Figure 5.19).[25-26]

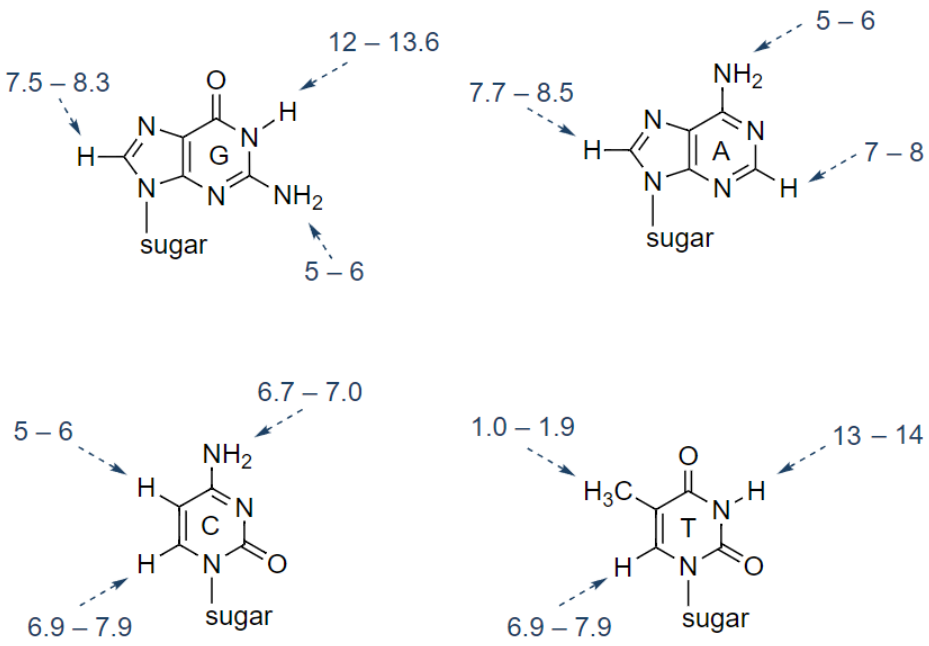


Figure 5.19 NMR chemical shifts in ^{1}H NMR spectra of nucleic acids.[25]

To gather insights into how the ferrocene modification influenced the structures of the duplexes and in particular whether cross-duplex FcNA-DNA H-bonding could be observed in the downfield region, it was decided to run a series of ^{1}H NMR experiments on these hybrid duplexes and compare these with control (**FcHH**) and unmodified duplexes (see Table 5.3 for the list of strands studied and Figure 5.20 below). The 1 GHz spectra were recorded by John Kirkpatrick from the research group of Prof. Teresa Carlomagno at the University of Birmingham Biomolecular NMR facility.

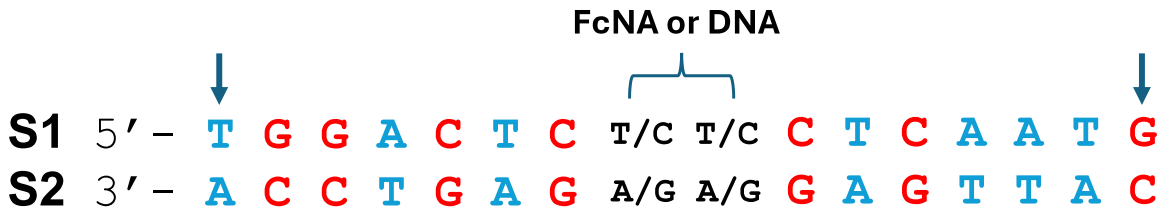


Figure 5.20 Sequence of the duplexes studied in the NMR experiments. AT base pairs are shown in blue and GC base pairs in red. Arrows denote the end base pairs that could fray, which would result in the loss of the imino H-bonding signal. The central dyad (bold and black) contains either one FcNA unit (**FcTT**, **FcCC)** or two DNA nucleotides in strand **S1**. A **FcHH** unit (no nucleobases attached) in strand **S1** was also used in place of a FcNA unit as a control.

5.3.4.1 Unmodified complementary duplexes

The first two spectra correspond to the perfectly complementary duplexes **S1CC-S2GG** and **S1TT-S2AA**. The spectrum of the **S1CC-S2GG** sample (Figure 5.21) shows 6 peaks in the region corresponding to Watson-Crick A-T base pairs and 8 peaks, along with a smaller broader peak at 12.7 ppm, in the region corresponding to Watson-Crick G-C base pairs. This aligns well with expectations, considering that the duplex consists of a total of 9 G-C base pairs and 7 A-T base pairs. The loss of one terminal A-T base pair signal out of the possible seven suggests fraying of the duplex ends for this weaker base pair, as observed previously in the literature.[23]

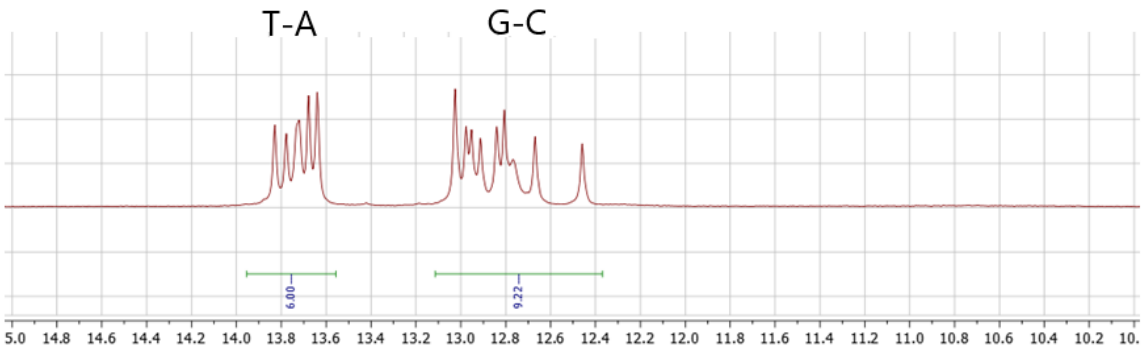


Figure 5.21 1000 MHz ^{1}H NMR spectra of **S1CC-S2GG** duplex (downfield region). Recorded at 5 °C, 10% D_2O in H_2O, [duplex] = 600 µM, 10 mM pH 7 phosphate buffer, 100 mM NaCl.

In the spectrum of **S1TT-S2AA**, which comprises a total of 7 G-C base pairs and 9 T-A base pairs, 7 peaks are observed in the T-A region, and 5 peaks are present further upfield in the G-C region, plus an additional broad shoulder peak at ca. 12.75 ppm (Figure 5.22). It is possible that one T-A and one G-C base pair at each end of the duplex are frayed so that they are not observed. A further T-A base pair may be hidden within the cluster of peaks between 13.6 and 13.8 ppm. These observations are borne out through NOESY experiments (see section 5.3.3.4).

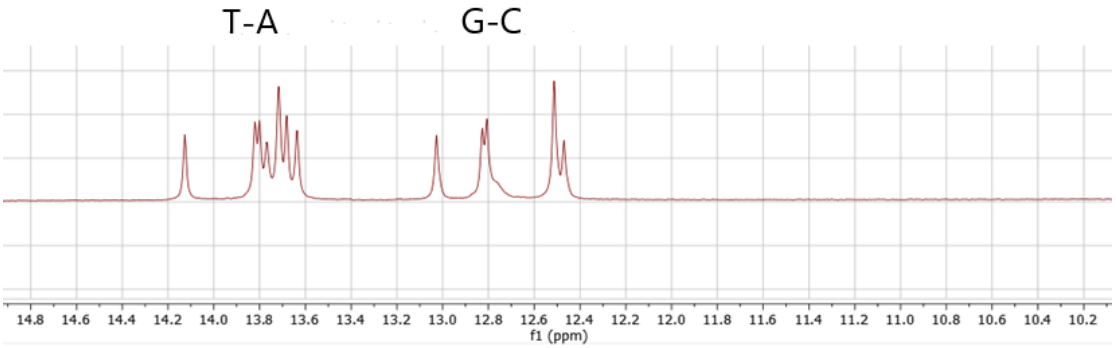


Figure 5.22 1000 MHz ^{1}H NMR spectra of **S1TT-S2AA** duplex (downfield region). Recorded at 5 °C, 10% D_2O in H_2O, [duplex] = 600 μM, 10 mM pH 7 phosphate buffer, 100 mM NaCl.

5.3.4.2 Unmodified mismatch duplexes

Figure 5.23 shows the spectra of the unmodified **S1TT-S2GG** sample. There are 6 T-A base-pair signals which resemble the number and pattern (with some small shifts in position) found in the duplex **S1CC-S2GG** (Fig. 5.21) that were assigned to the six internal T-A base pairs. However the signals in the G-C region show three (or possibly four) high-field-shifted peaks at <12 ppm. One interpretation is that the G+T dyad in the centre of the duplex is forming two non-WC G-T wobble base-pairs, [24] with each such base-pair showing two imino-proton peaks. If this were the case, there would still be six canonical WC G-C base pairs to assign. This would be the case if we assume

that the end G-C base pair is frayed and the large peak at just below 12.7 ppm is the result of two overlapping signals.

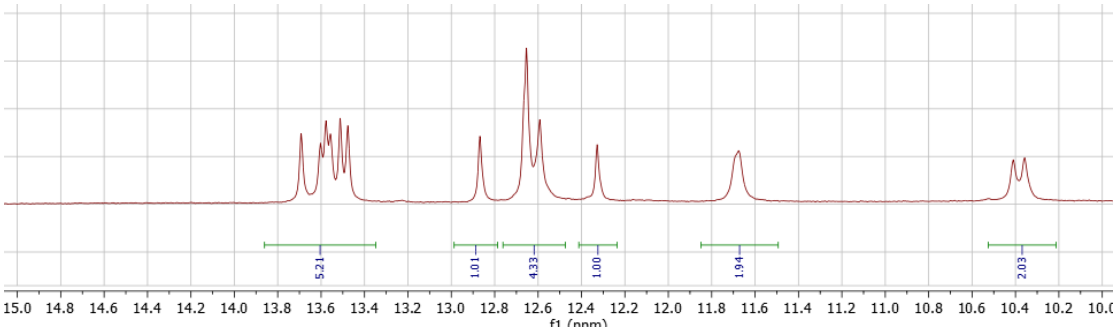


Figure 5.23 1000 MHz ^{1}H NMR spectra of **S1TT-S2GG** (downfield region). Recorded at 5 °C, 10% D_2O in H_2O, [duplex] = 600 µM, 10 mM pH 7 phosphate buffer, 100 mM NaCl.

The above interpretation is supported to some extent in the spectrum of the unmodified **S1CC-S2AA** sample that also contains two central mismatches (Figure 5.24). Once again six 6 peaks are observed in the T-A region, however no upfield signals are found in the G-C region, presumably due to the absence of H-bonding imino protons in any A-C wobble base pairs that could potentially form within the central dyad in this duplex.[27]

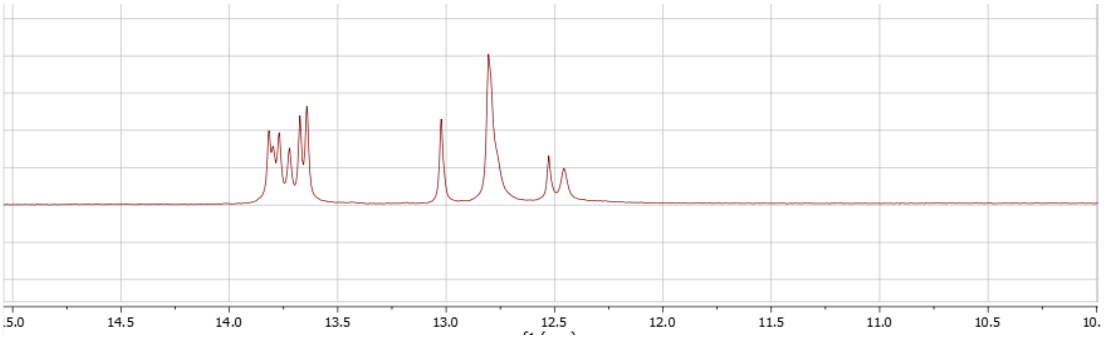


Figure 5.24 1000 MHz ^{1}H NMR spectra of **S1CC-S2AA** (downfield region). Recorded at 5 °C, 10% D_2O in H_2O, [duplex] = 600 µM, 10 mM pH 7 phosphate buffer, 100 mM NaCl.

5.3.4.3 FcNA-DNA duplexes

Moving to the ferrocene group, the first duplex studied was the one in which two natural dC DNA nucleotides in **S1CC** are replaced with one abiotic FcNA unit containing the same two nucleobases (**S1FcCC** see Fig 5.20). Mixing this with the **S2GG** strand gives the complementary duplex (**S1FcCC-S2GG**). Its spectrum is shown in Fig 5.25a along with the one for the control duplex containing a chiral ferrocene unit at the same position but with no nucleobases attached to it (**S1FcHH**, Fig. 5.25b).

The important result to note from the spectrum of **S1FcCC-S2GG** is the number of GC H-bonding peaks observed (12.5 ppm - 13.1 ppm). The maximum number that can be expected is nine, as found for the unmodified duplex **S1CC-S2GG** (Fig. 5.21), whereas no interaction at all in the central dyad would give a maximum of seven peaks. Including the broad peaks and a shoulder peak, what is in fact observed is a total of eight peaks, which does suggest a cross-duplex FcNA-DNA interaction in the form of a C-G base pairing. Furthermore, the control duplex **S1FcHH-S2GG**, which can not form H-bonds with G bases in the central dyad, gives fewer peaks in the imino G region of the spectrum (five) but in fact one more peak in the imino T region (six). However, interpretation of both spectra is compounded to some extent by a degree of broadening and the appearance of other small peaks, which may indicate other strand interaction modes or some partial decomposition/oxidation of the sample. This NOESY spectrum for **S1FcCC-S2GG** is discussed in more detail in Section 5.3.3.4.

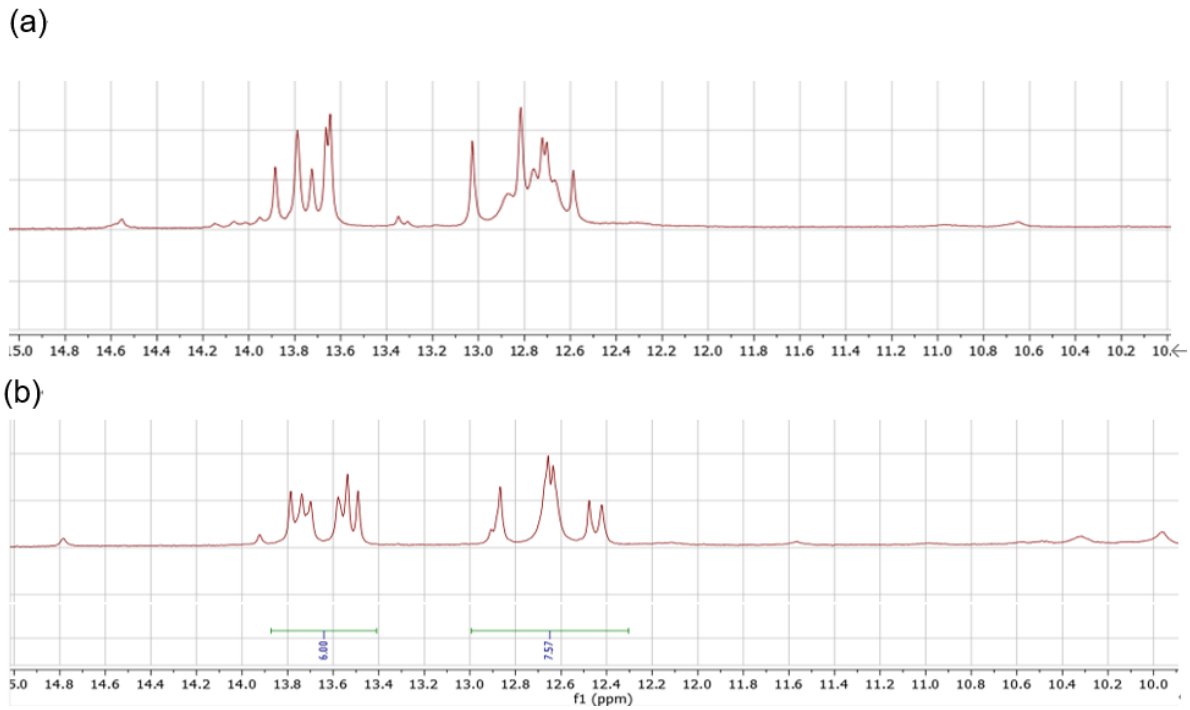


Figure 5.25 1000 MHz ^{1}H NMR spectra (downfield region) of (a) **S1FcCC+S2GG** and (b) **S1FcHH+S2GG**. Recorded at 5 °C, 10% D_2O in H_2O, [duplex] = 600 µM, 10 mM pH 7 phosphate buffer, 100 mM NaCl.

Next the **S1FcTT-S2AA** duplex was examined, where two DNA thymine bases in the central dyad of strand **S1** are replaced with one **FcTT** unit. However, only 5 clear T-A base pair peaks were observed along with 5 peaks in the G-C region (Figure 5.26a). This might be because A-T base pairing is weaker than C-G, making it more challenging to detect signals. This spectrum is discussed in more detail in Section 5.3.4.4.

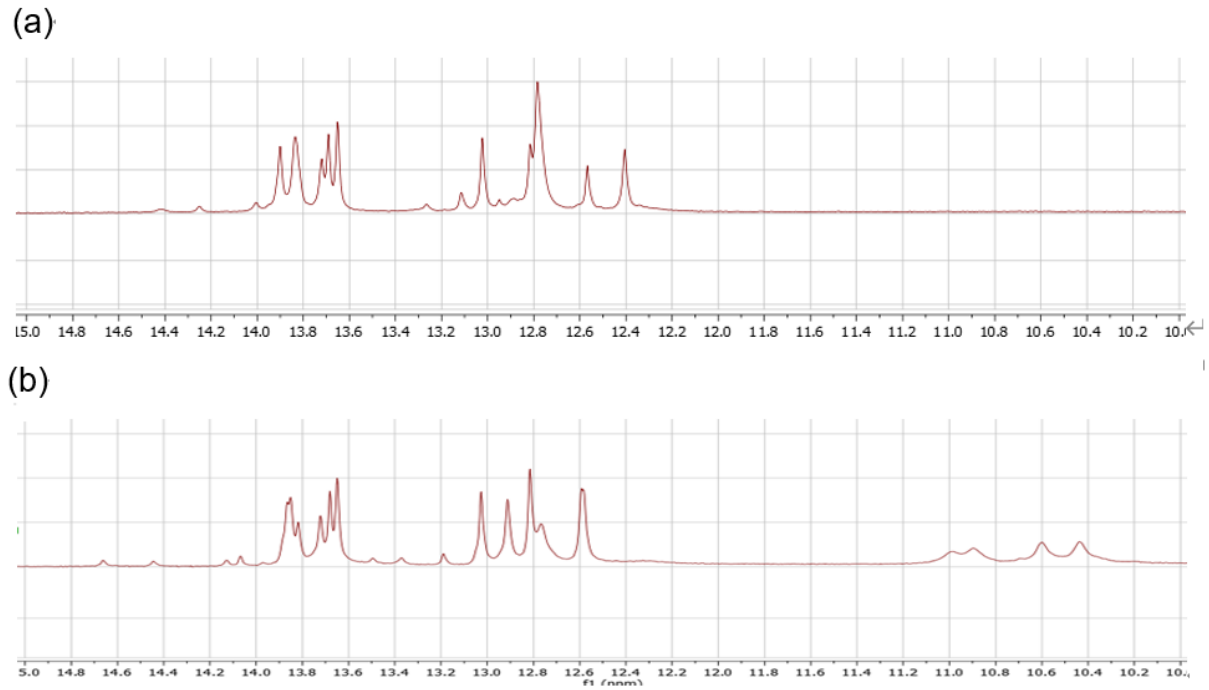


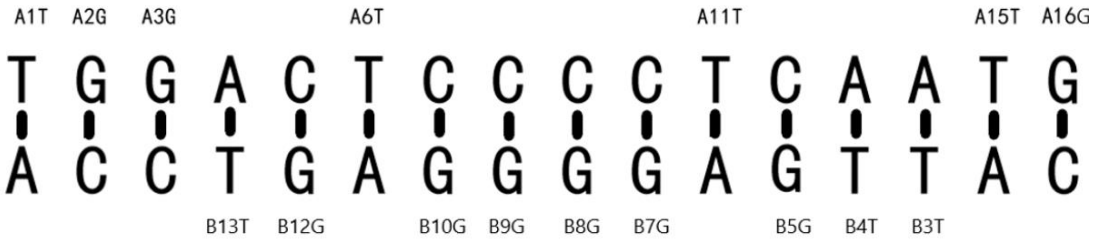
Figure 5.26 1000 MHz ^{1}H NMR spectra (downfield region) of (a) **S1FcTT-S2AA** and (b) **S1FcTT-S2GG.** Recorded at 5 °C, 10% D_2O in H_2O, [duplex] = 600 μM, 10 mM pH 7 phosphate buffer, 100 mM NaCl.

Interestingly, the **S1FcTT-S2GG** duplex, with two non-canonical T-G base pairs in the central dyad, gave a spectrum not unlike the corresponding unmodified one with the same central base pair dyad (Fig. 5.26b and Fig. 5.23 respectively). This would once again suggest the formation of non WC G-T wobble base pairs,[24][27] but this time involving a FcNA unit.

5.3.4.4 NOESY NMR spectra

NOESY (Nuclear Overhauser Effect Spectroscopy) is a powerful 2D NMR technique for elucidation of DNA structure. Typically requiring a higher sample concentration than 1D spectroscopy, it gives cross-peaks. A “sequence walk” is used, which involves a step-by-step analysis of the spectra to assign and interpret imino peaks in relation to the nucleotide sequence of the DNA. Here the objective was to establish a comparative analysis in the imino proton region as a basis for a fuller understanding of any cross duplex FcNA-DNA base pairing.

Similar to the previous 1D NMR study (section 5.3.3.3), the two fully complementary DNA duplexes were investigated, **S1CC-S2GG** and **S1TT-S2AA** and compared with their FcNA counterparts. The analysis of the **S1CC-S2GG** duplex (Fig. 5.27) confirms that the previously observed broad peak at 12.7 ppm in the G-C region of the 1D spectrum (Fig. 5.19) does indeed belong to the terminal G-C base pair in the duplex. The only peak that does not appear is that for A1T at the other terminus of the duplex.



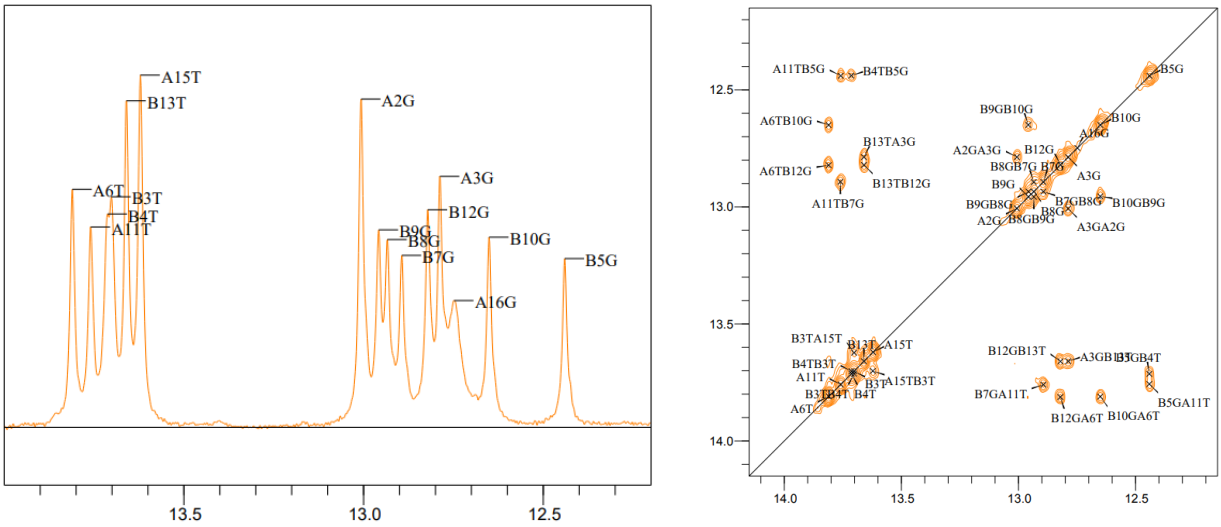


Figure 5.27 1000 MHz NOESY spectra (downfield region) of **S1CC+S2GG**. Recorded at 5 °C, 10% D_2O in H_2O, [duplex] = 1000 µM, 10 mM pH 7 phosphate buffer, 100 mM NaCl.

Regarding the **S1FcCC-S2GG** duplex (Figure 5.28), several sets of natural base-pairs adjacent to the ferrocene modification can be distinguished, with still three additional smaller/broader peaks in the G-C base pairing region tentatively assigned as B8G, B9G and A16G. One of the G signals that would correspond to a cross duplex FcNA H-bonding base pair (B8G, at 12.65 ppm), does indeed correlate to its neighbouring base pair (B7G at 12.7 ppm), albeit with a weaker signal. However, a connection between B8G and its adjacent FcNA dyad neighbour (B9G, at 12.65 ppm) is not observed due to the weak signal. The small peak at 12.75 ppm, when compared with the unmodified duplex, is likely attributed to the terminal C-G base pair A16G.

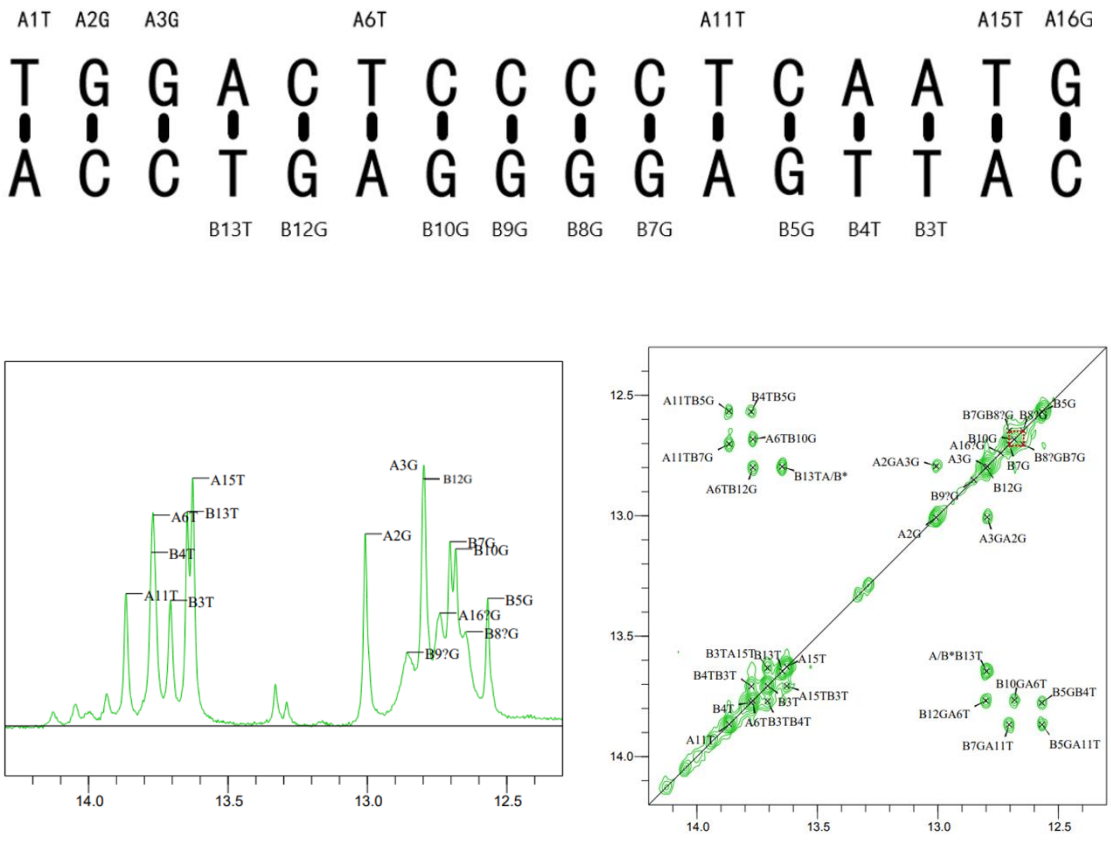
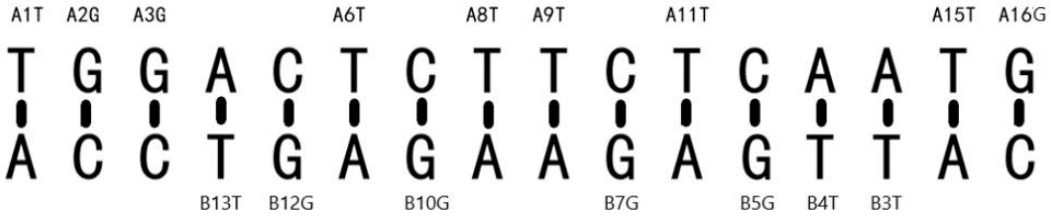
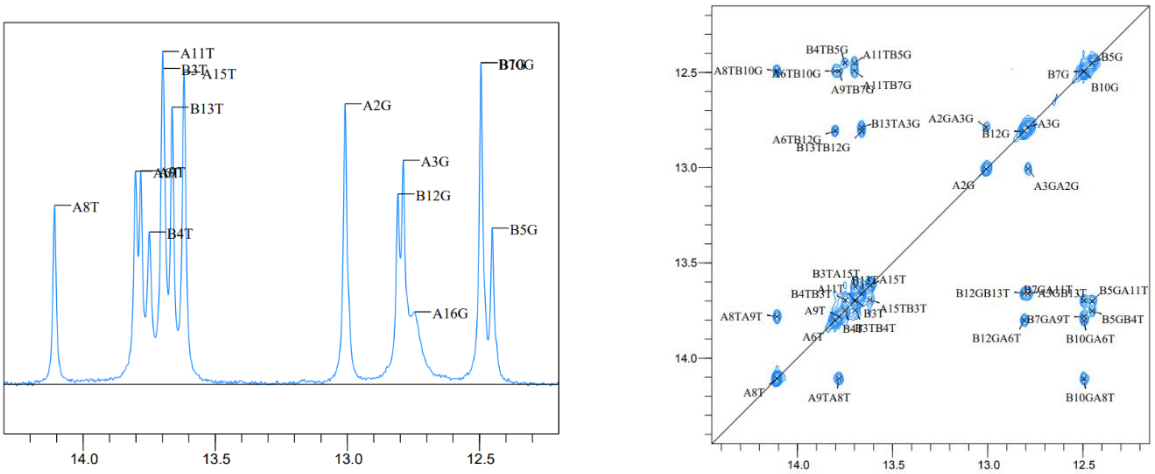


Figure 5.28 1000 MHz NOESY spectra (downfield region) of **S1FcCC-S2GG**. Recorded at 5 °C, 10% D_2O in H_2O, [duplex] = 1000 µM, 10 mM pH 7 phosphate buffer, 100 mM NaCl.

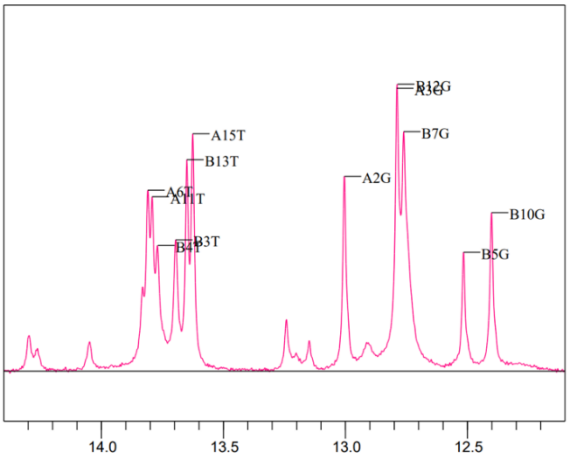
Eight peaks in the T-A region are observed in the spectrum for **S1TT-S2AA** (Figure 5.29a) and have been successfully assigned, with once again the only one missing corresponding to the frayed T-A base pair at the terminus (A1T). All seven G-C base pairs were successfully assigned, including the broad peak A16G. Unfortunately, the protons corresponding to the T bases on the central FcNA unit in **S1FcTT-S2AA** could not be assigned, although a shoulder peak immediately downfield of A16T could be observed (Fig. 5.29b). It is interesting to note that the B7G imino peak next to the FcNA unit is significantly more downfield than the corresponding peak in the unmodified duplex.



(a)



(b)



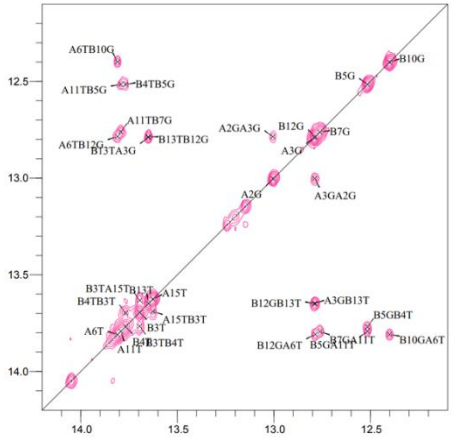


Figure 5.29 1000 MHz NOESY spectra (downfield region) of (a) **S1TT-S2AA** and (b) **S1FcTT-S2AA**. Recorded at 5 °C, 10% D_2O in H_2O, [duplex] = 1000 μM, 10 mM pH 7 phosphate buffer, 100 mM NaCl.

In summary, the NOESY experiments give further support to the broad peaks observed in the GC region of the spectrum in **S1FcCC-S2GG** duplex being caused by cross duplex FcNA-DNA base pairing. This is consistent with the data from the melting temperature studies, where the melting temperature of **S1FcCC-S2GG** (48°C) is significantly higher than those for other controls. In fact, while the melting temperature for the doubly matched system (i.e. one **FcCC** unit binding to two G DNA bases in the complementary strand) is significantly higher than that for the doubly mismatched dyad (**S1FcCC-S2AA**, 39°C), it is only slightly higher than the one with one matched and one mismatched base pair (**S1FcCC-S2AG**, 45.5°C). This could be attributed to the fact that in **S1FcCC-S2GG**, only one cytosine can form a relatively strong base pair with its opposing base, while the other forms a much weaker one.

5.4 Conclusion

A series of FcNA-containing hybrid oligonucleotides has been successfully synthesized that act as intriguing organometallic analogues of DNA. Building on previous computer simulations and melting temperature studies, NMR studies give further evidence for the ability of FcNA units within these strands to successfully form cross-duplex WC hydrogen bonds with natural bases. With its tunable linker lengths and stereochemistry, FcNA units can be designed to align closely with the natural helical twist of the DNA backbone in B-DNA. Importantly, studies have also shown that the FcNA-containing oligomers are functional in a biological system in that some polymerases can recognize the FcNA site and read through it and add specific bases. Taken together, these findings indicate that FcNA is a promising DNA analogue material with the potential for future applications in electrochemical sensing or the pharmaceutical industry, for example as stable antisense oligonucleotides.

Going forwards, attempts are continuing to crystallize hybrid duplexes containing FcNA units for X-ray crystallography experiments. This could offer valuable information on secondary structure and the extent of any π-stacking interactions between nucleobase layers. Other work is in progress to assess the extent to which FcNA triphosphates can be successfully incorporated by polymerases for enzyme-based synthesis of hybrid duplexes.

References

1. Nguyen, H. V.; Zhao, Z. Y.; Sallustrau, A.; Horswell, S. L.; Male, L.; Mulas, A.; Tucker, J. H. *Chemical Communications*, **2012**, *48*(100), 12165-12167.
2. Carr-Smith, J. *PhD thesis*, University of Birmingham, **2015**.
3. Kedge, J.; *PhD thesis*, University of Birmingham, **2016**.
4. Moran, N.; Bassani, D. M.; Desvergne, J. P.; Keiper, S.; Lowden, P. A.; Vyle, J. S.; Tucker, J. H. *Chemical Communications*, **2006**, (48), 5003-5005.
5. Monticelli, D.; Psaro, R.; Pozzi, A.; Dossi, C.; Recchia, S. *Analytical and Bioanalytical Chemistry*, **2005**, *383*, 115-121.
6. Duprey, J. L. H.; Zhao, Z. Y.; Bassani, D. M.; Manchester, J.; Vyle, J. S.; Tucker, J. H. *Chemical Communications*, **2011**, *47*(23), 6629-6631.
7. Röthlisberger, P.; Levi-Acobas, F.; Leumann, C. J.; Hollenstein, M. *Bioorganic & Medicinal Chemistry*, **2020**, *28*(11), 115487.
8. Röthlisberger, P.; Levi-Acobas, F.; Sarac, I.; Marlière, P.; Herdewijn, P.; Hollenstein, M. *Organic & Biomolecular Chemistry*, **2017**, *15*(20), 4449-4455.
9. Hocek, M. *Accounts of Chemical Research*, **2019**, *52*(6), 1730-1737.
10. Figazzolo, C.; Ma, Y.; Tucker, J. H.; Hollenstein, M. *Organic & Biomolecular Chemistry*, **2022**, *20*(41), 8125-8135.
11. A. V. Chudinov, V. A. Vasiliskov, V. E. Kuznetsova, S. A. Lapa, N. A. Kolganova and E. N. Timofeev. *Scientific Reports*, **2021**, 11, 2423
12. Flamme, M.; Röthlisberger, P.; Levi-Acobas, F.; Chawla, M.; Oliva, R.; Cavallo, L.; Hollenstein, M. *ACS Chemical Biology*, **2020**, *15*(11), 2872-2884.
13. Kumara, G. S. R.; Seo, Y. J. *Organic & Biomolecular Chemistry*, **2021**, *19*(26), 5788-5793.
14. Ji, D.; Mohsen, M. G.; Harcourt, E. M.; Kool, E. T. *Angewandte Chemie International Edition*, **2016**, *55*(6), 2087-2091.
15. Sefah, K.; Yang, Z.; Bradley, K. M.; Hoshika, S.; Jiménez, E.; Zhang, L.; Benner, S. A. *Proceedings of the National Academy of Sciences*, **2014**, *111*(4), 1449-1454.
16. Hoshika, S.; Leal, N. A.; Kim, M. J.; Kim, M. S.; Karalkar, N. B.; Kim, H. J.; Benner,

S. A. *Science*, **2019**, *363*(6429), 884-887.

17. Zhou, A. X. Z.; Dong, X.; Romesberg, F. E. *Journal of the American Chemical Society*, **2020**, *142*(45), 19029-19032.
18. Obeid, S.; Blatter, N.; Kranaster, R.; Schnur, A.; Diederichs, K.; Welte, W.; Marx, A. *The EMBO Journal*, **2010**, *29*(10), 1738-1747.
19. Boiteux, S.; Guillet, M. *DNA Repair*, **2004**, *3*(1), 1-12.
20. Liu, Z. J.; Martínez Cuesta, S.; van Delft, P.; Balasubramanian, S. *Nature Chemistry*, **2019**, *11*(7), 629-637.
21. Güixens-Gallardo, P.; Hocek, M.; Perlíková, P. *Bioorganic & Medicinal Chemistry Letters*, **2016**, *26*(2), 288-291.
22. Moran, S.; Ren, R. X. F.; Sheils, C. J.; Rumney IV, S.; Kool, E. T. *Nucleic Acids Research*, **1996**, *24*(11), 2044-2052.
23. Duprey, J. L. H.; Bassani, D. M.; Hyde, E. I.; Jonusauskas, G.; Ludwig, C.; Rodger, A.; Tucker, J. H. *Organic & Biomolecular Chemistry*, **2018**, *16*(35), 6576-6585.
24. Early, T. A.; Olmsted III, J.; Kearns, D. R.; Lezius, A. G. *Nucleic Acids Research*,**1978**, *5*(6), 1955-1970.
25. T. Brown and T. Brown (Jr), Nucleic Acids Book. https://atdbio.com/nucleic-acids-book.
26. Bertini, I.; McGreevy, K. S.; Parigi, G. (Eds.) NMR of Biomolecules: Towards Mechanistic Systems Biology. *John Wiley & Sons*, **2012**.
27. Koag, M. C.; Nam, K.; Lee, S. *Nucleic Acids Research*, **2014**, 42(17), 11233-11245.
28. Zhao J., unpublished results.

Chapter 6 – Experimental

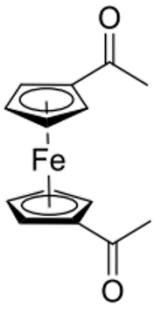
6.1 Synthesis of Ferrocene Nucleic acid (FcNA) monomers

6.1.1 Materials and methods

Reagents and solvents were purchased from Sigma-Aldrich, Acros, Alfa Aesar and Fisher Scientific and used without further purification. Dry solvents were either purchased as such or obtained from a Pure Solv-MD solvent purification system and handled under an argon atmosphere. Column chromatography was performed using open columns packed with Merck grade 60 silica gel, topped with a layer of sand.

TLC analysis was performed on Merck silica gel 60 silica sheets. ^{1}H, ^{13}C, and ^{31}P NMR spectra were obtained on Bruker AVIII300 or AVIII400 spectrometers. Chemical shifts (δ) given as parts per million (ppm) relative to the residual solvent peak. Coupling constants (J) are given in hertz (Hz) and are averaged when close in value. Data was acquired using Bruker Topspin v3 and analysed with MestReNova v15.0.0. Electrospray ionization mass spectrometry (ESI-MS) was conducted using either a Waters Micromass LCT electrospray time-of-flight (ES-TOF) spectrometer, a Waters Xevo G2-XS, or a Synapt G2S mass spectrometer. Infrared (IR) spectroscopy was performed on dry samples using a Varian 660-IR FT-IR spectrometer.

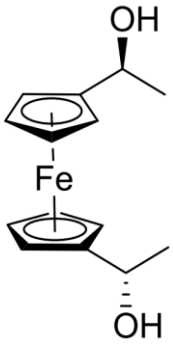
1,1'-Bis-acetyl-ferrocene (compound 2)



Acetyl chloride (21.16 mL, 298 mmol) was introduced into a suspension of aluminium (III) chloride (35.26 g, 264 mmol) in DCM at 0 °C. Subsequently, ferrocene (**compound 1**) (20 g, 107.4 mmol) dissolved in DCM was added dropwise over 20 minutes. The reaction mixture was stirred for 2 hours and then allowed to warm to room temperature. After completion of the reaction, it was quenched by gradual addition of ice-cold water at 0 °C. The resulting mixture was diluted with DCM, cautiously washed with saturated aqueous K_2CO_3, and brine. The crude product was dried using $MgSO_4$ and subjected to evaporation. Purification of the crude material was accomplished through flash column chromatography (25% EtOAc in hexane), yielding an orange solid (24.76 g, 85%). The characterisation matched literature values.[1]

^{1}H NMR (400 MHz, Chloroform-d) δ 4.77 (t, 4H), 4.51 (t, 4H), 2.36 (s, 6H). ^{13}C NMR (400 MHz, $CDCl_3$) δ 201.11 (C=O), 80.71 (Fc C), 73.60 (Fc), 70.96 (Fc), 27.64 (CH_3).

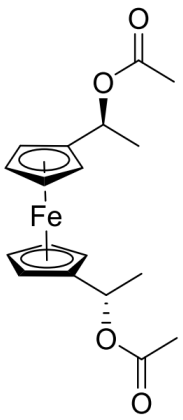
(*S,S*)-1,1'-Bis-(α-hydroxyethyl)-ferrocene (compound 3)



(*R*)-(+)-2-Methyl-CBS-oxazaborolidine (9.9 g, 36.0 mmol, 0.6 eq) was dissolved in dry THF (80 mL) at 0 ºC in an ice bath under an argon atmosphere. BH_3•SMe_2 (2 M in THF) (12 mL) was added, and the resulting mixture was stirred for 30 minutes. Dry THF (200 mL) was used to dissolve **compound 2** (16.2 g, 60.0 mmol, 1.0 eq), which was then added slowly, along with the remaining B_3•SMe_2 (2 M in THF) (48 mL). The mixture was stirred at 0 ºC for 2 hours. Quenching the reaction was achieved by dropwise addition of MeOH at 0 ºC (100 mL). The crude product was washed with NH_4Cl, extracted with ether, and dried over $MgSO_4$. Then it was purified using flash column chromatography on silica gel, employing a gradient eluent system of 0 → 30% EtOAc in hexane. Subsequently, the solvent was removed under vacuum, resulting in the formation of the title compound as a yellow solid. (16.2 g, 98%) The characterisation matched literature values.[1]

^{1}H NMR (400 MHz, $CDCl_3$) δ .4.65 (q, 2H (CH)), 4.39 (s, 2H (OH)), 4.19 – 4.15 (m, 8H (Fc CH)), 1.40 (d, 6H (CH_3)). ^{13}C NMR (400 MHz, $CDCl_3$) δ 95.19 (Fc C), 67.19 (Fc CH), 67.59 (Fc CH), 66.15 (Fc CH), 66.0 (Fc CH), 65.63 (CH), 25.61 (CH_3).

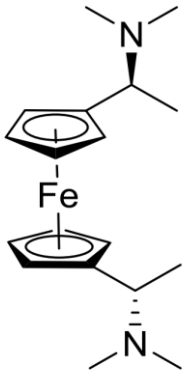
(*S,S*)-1,1'-Bis(α-acetoxyethyl)ferrocene (compound 4)



Under an argon atmosphere, **compound 3** (16g, 58.5 mmol, 1.0 eq) was dissolved in 120 mL dry DCM and stirred for 15 minutes. Subsequently, acetic anhydride (13.68 mL, 146.25 mmol, 2.5 eq), TEA (21 mL, 150 mmol, 2.6 eq), and DMAP (0.36 g, 2.94 mmol, 0.05 eq.) were added in succession. The resulting mixture was stirred for 22 hours. To quench the reaction, 200 mL saturated NH_4Cl was added, followed by extraction with DCM. The organic layers were then washed with brine (100 mL) and dried over $MgSO_4$. Afterward, the solvent was removed in vacuum, and the crude product was purified through flash column chromatography on silica gel using 20% EtOAc and 1% TEA in hexane. After removal of the solvent in vacuum, the title compound was obtained as a yellow solid (18.9 g, 90%). The characterisation matched literature values.[1]

^{1}H NMR (400 MHz, $CDCl_3$) δ 5.81 (q, 2H CH), 4.24 (dd, 2H (Fc CH)), 4.19 (dd, J = 3.3, 1.7 Hz, 2H (Fc CH)), 4.15 – 4.12 (m, 4H (Fc CH)), 2.04 (s, 6H (CH_3CO)), 1.53 (d, 6H (CH_3CH)). ^{13}C NMR (400 MHz, $CDCl_3$) δ 170.46(CO), 88.72 (Fc C–CH), 69.24 (Fc CH), 68.96 (Fc CH), 69.24 (Fc CH), 68.86 (Fc CH), 66.6 (CH), 21.39 (CH_3CO), 20.28 (CH_3CH).

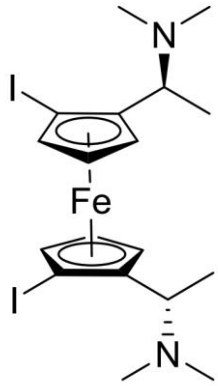
(*S,S*)-1,1'-Bis(α-N,N-dimethylaminoethyl)ferrocene (compound 5)



Compound 4 (17.05g, 47.91 mmol) was dissolved in methanol (30 mL), and dimethylamine (30 mL) was added. The mixture was stirred for 60 hours at room temperature. The reaction was quenched with 60 mL of water and extracted with DCM. After drying on $MgSO_4$, the solvent was removed in vacuum, and the crude product was purified via flash column chromatography (5% MeOH and 5% TEA in EtOAc). This process yielded the product as a dark orange oil. (11 g, 70%) The characterisation matched literature values.[1]

^{1}H NMR (400 MHz, $CDCl_3$) δ 4.08 – 4.03 (m, 8H (Fc CH)), 3.6 (q, J = 6.9 Hz, 2H (CH), 2.08 (s, 12H (CH_3)), 1.37 (d, J = 6.9 Hz, 6H (CH_3CH)). ^{13}C NMR (400 MHz, $CDCl_3$) δ 87.22 (Fc C), 70.33 (Fc CH), 68.50 (Fc CH), 68.24 (Fc CH), 67.1 (Fc CH), 58.63 (CH), 40.63 (CH_3N), 16.15 (CH_3CH).

(*S*,*S*)-2,2'-Bis(α-N,N-dimethylaminoethyl)-(R_p,R_p)-1,1'-bisiodoferrocene (compound 6)

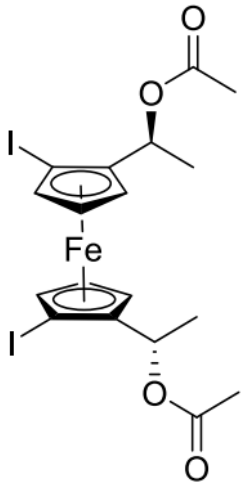


Compound 5 (12.3 g, 37.50 mmol, 1.0 eq) was dissolved in dry Et_2O (120 mL and stirred under an atmosphere of argon at 0 ºC for 5 min. Subsequently, n-BuLi (2.5 M in hexane) (60.0 mL, 150.0 mmol, 4.0 eq) was added dropwise and the solution was then stirred overnight under an argon atmosphere. Dry THF (20 mL) was added, and the solution was cooled to -78 ºC using an acetone/dry ice bath. I_2 (42.84 g, 168.75 mmol, 4.5 eq) was dissolved in dry THF (60 mL) and added dropwise. The ice bath was removed, and stirring was continued for 3 hr while warming to room temperature. Quenching at 0 °C was achieved by adding saturated sodium thiosulfate (100 mL). The mixture was further extracted with Et_2O (100 mL), then dried over $MgSO_4$, and the solvent was removed in vacuum. The crude product was purified by flash column chromatography on silica gel using a flash column chromatography (50% EtOAc,45% hexane and 5% TEA). The solvent was removed in vacuo to yield the title compound as a yellow solid. (15.181 g, 69%) The characterisation matched literature values.[1]

^{1}H NMR (400 MHz, $CDCl_3$) δ 4.18 – 4.11 (m, 6H), 3.61-3.57 (q, J = 6.9 Hz, 2H), 2.08 (s, 12H), 1.44 (d, J = 6.9 Hz, 6H). ^{13}C NMR (400 MHz, $CDCl_3$) δ 90.67 (Fc C), 82.17

(Fc CH), 72.15 (Fc CH), 67.76 (Fc CH), 56.94 (CH), 47.64 (Fc C–I), 41.11 (CH_3), 15.65 (CH_3). MS (ES) (m/z) calculated for $C_{18}H_{26}FeI_2N_2$ 580.9613, found 580.9619 $[M]^+$. Retention time for chiral HPLC (see Chapter 3 for conditions): 20.6 min.

(*S,S*)-2,2'-Bis(α-acetoxyethyl)-(R_p,R_p)-1,1'-bisiodoferrocene (compound 7)

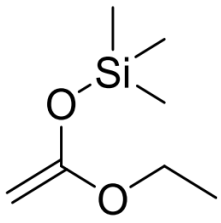


Compound 6 (12.0 g, 20.6 mmol, 1.0 eq) was dissolved in acetic anhydride (Ac_2O) (60 mL) and stirred for 2 hr at 90ºC. Following this, the reaction mixture was cooled to room temperature and quenched by the slow addition of 300 mL saturated $NaHCO_3$ solution. The resulting crude product was extracted with DCM, washed with brine, and dried over $MgSO_4$. The crude product was dissolving in 5 mL of DCM and subsequently adding a substantial quantity of hexane to the solution. The solution was then placed in a refrigerator overnight to allow for recrystallization. The crystals were washed using hexane, yielding the title compound as a yellow solid. The characterisation matched literature values.[1]

^{1}H NMR (400 MHz, $CDCl_3$) δ 5. 81 (q, 2H (CH)), 4.35 (m, 2H (Fc CH)), 4.31 (m, 2H (Fc

CH)), 4.22 (m, 2H (Fc CH)), 2.02 (s, 6H (CH_3CO)), 1.61 (d, 6H (CH_3CH)). ^{13}C NMR (400 MHz, $CDCl_3$) δ 170.14 (CO), 88.93 (Fc C), 82.22 (Fc CH), 73.2 (Fc CH), 68.57 ($CHCH_3$), 67.81 (Fc CH), 46.56 (Fc C–I), 21.10 (CH_3CH), 18.81 (CH_3CO). MS (ES) (m/z) calculated for $C_{18}H_{20}FeI_2O_4Na$ 623.8698, found 623.8719 $[M+Na]^+$

1-Ethoxy-1-trimethylsilyloxyethylene (compound 7.1)

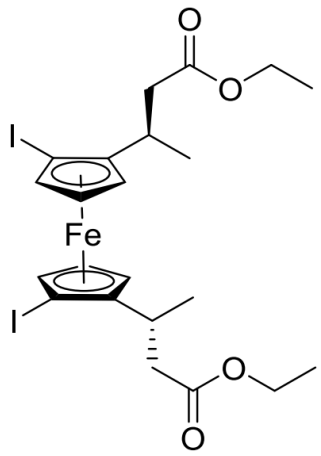


Diisopropylamine (DIPA) (8.36 mL, 59.2 mmol, 1.0 eq.) was introduced into dry tetrahydrofuran (THF) (20 mL) and cooled to 0 ºC under an argon atmosphere in a Schlenk tube. n-Butyllithium (n-BuLi) (2.5 M in hexane) (23.7 mL, 59.2 mmol, 1.0 eq) was then added dropwise, and stirring was sustained for 20 min. The resulting yellow solution was cooled to -78 ºC, and EtOAc (5.26 mL, 53.2 mmol, 0.9 eq.) was introduced. After 30 min of stirring, trimethylsilyl chloride (TMSCl) (8.26 mL, 65.2 mmol, 1.1 eq.) was added, and stirring persisted for an additional 30 min. The mixture was subsequently raised to room temperature and stirred for another hour. THF was removed by gentle evaporation at 25 ºC on a rotary evaporator, and the remaining solution was poured into a mixture of hexane: water (9:4) (100 mL). The organic layer was collected, washed with brine (50 mL) and dried over $MgSO_4$. The majority of the solvent was removed by gentle evaporation at 25 ºC on the rotary evaporator before

the remaining hexane and excess EtOAc were removed by vacuum distillation. The colourless liquid product distilled off at 80 ºC and NMR showed the composition to be ca. 60% in EtOAc. The characterisation matched literature values.[1]

^{1}H NMR (300 MHz, $CDCl_3$) δ 3.56-3.49 (q, 2H(CH_2CH_3)), 2.98-2.97 (d, 1H, (CH_2=CO)), 2.84- 2.83 (d, 1H (CH_2=CO)), 1.10-1.07 (t, 3H (CH_2CH_3)), 0.00 (s, 9H, CH_3) (3.92 g, 24.44 mmol, 60% in Ethyl Acetate).

(*R,R*)-2,2'-Bis(α-methyl-ethylpropanoate)-(R_p,R_p)-1,1'-bisiodoferrocene (compound 8)

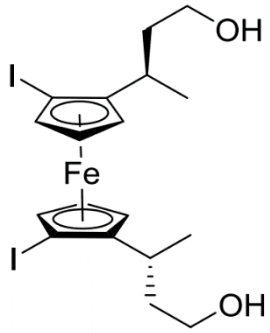


Compound 7 (8.25 g, 13.5 mmol, 1.0 eq) was dissolved in dry DCM (120 mL) and cooled to -78 ºC under an argon atmosphere. **Compound 7.1**, 1-ethoxy-1-trimethylsilyloxyethylene (60% in ethyl acetate by molar amount from NMR, 16 mL), was added and the mixture stirred for 30 mins before slowly adding BF_3•THF (13.5 mL, 1 eq.) while stirring for an additional hour at -78 ºC. The mixture was warmed over 1 hour to room temperature while stirring was continued. Quenching the reaction involved the addition of saturated $NaHCO_3$, followed by extraction with DCM and drying

over $MgSO_4$. The solvent was evaporated under vacuum, and the crude product was subsequently purified using flash column chromatography on silica gel with a gradient eluent system ranging from 1% to 30% ethyl acetate in hexane. The solvent was then removed in vacuum, yielded the desired compound in the form of a yellow solid.(6.71g,79%) The characterisation matched literature values.[1]

^{1}H NMR (400 MHz, $CDCl_3$) δ 4.19 – 4.10 (m, 6H (Fc CH)), 4.10 (m,2H (CH_2O)), 3.03 (m, 2H ($CHCH_3$)), 2.51 (m, 2H (CH_2CH), 2.11 (m, 2H (CH_2CH)), 1.38 (d, *J* = 6.9 Hz, 6H (CH_3CH)), 1.26 (t, 6H, *J* = 6.9 Hz, (CH_3CH_2)). ^{13}C NMR (400 MHz, $CDCl_3$) δ 171.87 (C=O), 94.6 (Fc C), 81.19 (Fc CH), 71.79 (Fc CH), 66.38 (Fc CH), 60.32 (CH_2O), 46.20 (Fc C–I), 42.81 (CH_2CH), 30.15 ($CHCH_3$), 18.84 (CH_3CH), 14.11 (CH_3CH_2).MS (ES) (m/z) calculated for $C_{22}H_{28}FeI_2O_2Na$ 688.9342, found 688.9396 $[M+Na]^+$

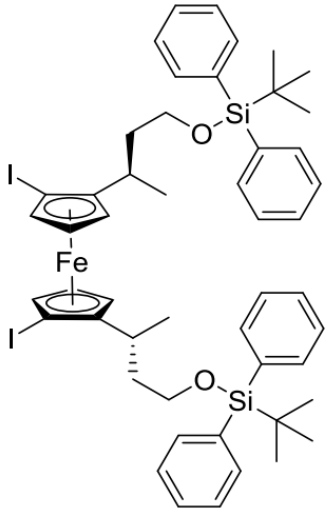
(*R,R*)-2,2'-Bis(α-methyl-γ-hydroxypropyl)-(R_p,R_p)-1,1'bisiodoferrocene (compound 9)



Compound 8 (6.30 g, 9.45 mmol, 1.0 eq) was dissolved in Et2O (50 mL) and cooled to 0 ºC under an argon atmosphere. Diisobutylaluminum hydride (DIBAL, 1 M in THF) (60 mL, 60 mmol, 6.0 eq) was added dropwise, and the resulting solution was stirred at RT for 6 hours. The reaction was quenched with a saturated solution of sodium/potassium tartrate, followed by extraction with DCM and drying over $MgSO_4$. The crude product was purified through flash column chromatography on silica gel, employing a gradient eluent system of 20 → 50% EtOAc in hexane. Removal of the solvent in vacuo yielded the title compound as a yellow solid (4.32 79%). The characterisation matched literature values.[1]

^{1}H NMR (400 MHz, $CDCl_3$) δ 4.13 (m, 6H (Fc CH)), 3.58 (t, *J* = 6.9 Hz,4H (CH_2O)), 2.69 (m, 2H ($CHCH_3$)), 1.66 (m, 2H (CH_2CH)), 1.53 (m, 2H (CH_2CH)), 1.36 (d, 6H, *J* = 6.9 Hz (CH_3)). ^{13}C NMR (400 MHz, $CDCl_3$) δ 90.3 (Fc C), 80.77 (Fc CH), 71.62 (Fc CH), 66.06 (Fc CH), 60.74 (CH_2O), 46.96 (Fc C–I), 41.78 (CH_2CH), 29.13 ($CHCH_3$), 19.59 (CH_3). MS (ES) (m/z) calculated for 581.9215, found 581.9228 $[M]^+$.

(*R,R*)-2,2'-Bis(α-methyl-γ-(tert-butyldiphenylsilyl)oxypropyl)-(R_p,R_p)-1,1'-bisiodoferrocene (compound 10)

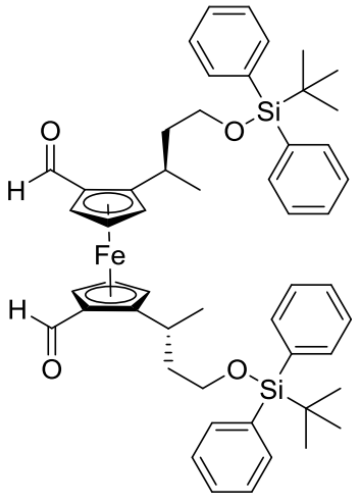


Compound 9 (1.60 g, 2.75 mmol, 1.0 eq) was dissolved in dry DCM (20 mL) under an argon atmosphere. Triethylamine (1.15 mL, 8.25 mmol, 3.0 eq), tert-butyldiphenylsilyl chloride (2.15 mL, 8.25 mL, 3.0 eq), and 4-dimethylaminopyridine (DMAP) (10 crystals) were sequentially added, and the resulting solution was stirred for 16 hours. The reaction was quenched with water, extracted with DCM, and dried over $MgSO_4$. The crude product was purified by flash column chromatography on silica gel using an eluent of 10% ethyl acetate in hexane. Removal of the solvent in vacuo yielded the title compound as a yellow oil (2.78 g, 2.63 mmol, 89%). The characterisation matched literature values.[1]

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.67 (m, 8H (Ph CH)), 7.38 (m, 12H (Ph CH)), 4.13 (dd, 2H (Fc CH)), 4.04 (m, 4H (Fc CH)), 3.66 (dd, 4H (CH_2O), 2.71 (dqd, 2H ($CHCH_3$)), 1.83 (dtd, 2H (CH_2CH)), 1.34 (m, 2H (CH_2CH)), 1.27 (d, 6H (CH_3CH)), 1.05 (s, 18H (tBu CH_3)). ^{13}C NMR (400 MHz, $CDCl_3$) δ 135.64 (Ph CH), 135.65 (Ph CH), 134.07 (Ph C–Si), 134.01 (Ph C–Si), 129.52 (Ph CH), 129.50 (Ph CH), 127.62 (Ph CH), 127.60 (Ph

CH), 96.63 (Fc C–CH), 80.29 (Fc CH), 71.16 (Fc CH), 66.11 (Fc CH), 61.97 (CH_2O), 46.66 (Fc C–I), 41.26 (CH_2), 29.05 (CH), 26.95 (tBu CH_3), 19.24 (tBu C), 18.7 (CH_3CH)
MS (ES) (m/z) calculated for $C_{50}H_{60}FeI_2O_2Si_2$ 1058.1572, found 1058.1603 $[M]^+$

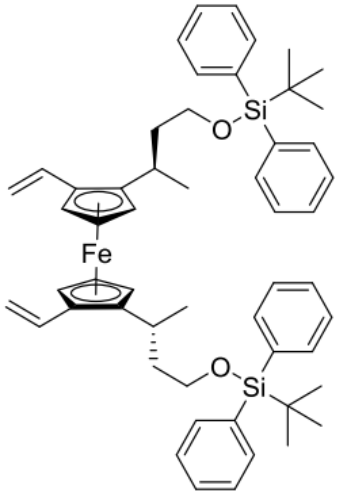
(*R,R*)-2,2'-Bis(α-methyl-γ-(tertbutyldiphenylsilyl)oxypropyl)-(R_p,R_p)-1,1'-bisformylferrocene (compound 11)



Compound 10 (11.6 g, 10.97 mmol, 1 eq) was dissolved in Et_2O (75 mL) and cooled to -78°C under an argon atmosphere. n-Butyllithium (nBuLi) (2.5 M in hexanes) (15.36 mL, 38.41 mmol, 3.5 eq) was slowly added, and the solution was stirred for 1 hour and 15 minutes. Dimethylformamide (3.39 mL, 43.90 mmol, 4 eq) was added dropwise, and the solution was stirred for 25 minutes before removing the ice bath. The solution was allowed to warm to room temperature over 1.5 hours. The reaction was quenched with water, extracted with DCM, and dried with $MgSO_4$. The solvent was evaporated under vacuum, and the crude product underwent purification through flash column chromatography on silica gel using an eluent consisting of 10% EtOAc in hexane. Subsequently, the solvent was removed under vacuum to yield the title compound as

a red oil. (8.4g, 9.75 mmol, 88.9%). The characterisation matched literature values.[1] ^{1}H NMR ($CDCl_3$, 400MHz): δ 10.04 (s, 2H, CHO), 7.62 (m, 8H, Ph CH), 7.38 (m, 12H, Ph CH), 4.8 (m, 2H, Fc CH), 4.42 (m, 4H, Fc CH), 3.59 (4H, m, CH_2O), 3.09 (m, 2H, $CHCH_3$), 1.65 (m, 2H, CH_2CH), 1.51 (m, 2H, CH_2CH), 1.28 (d, 6H, CH_3CH), 1.03 (s, 18H, (tBu CH_3)) ^{13}C NMR ($CDCl_3$, 400MHz) δ 192.9 (CHO), 135.59 (Ph CH), 135.56 (Ph CH), 133.83 (Ph CSi), 133.72 (Ph C-Si), 129.63 (Ph CH), 129.61 (Ph CH), 127.67 (Ph CH), 127.66 (Ph CH), 99.80 (Fc C-CH), 72.44 (Fc CH) 72.11 (Fc CH), 71.54 (Fc CH), 61.58 (CH_2O) , 43.06 (CH_2CH) , 27.40 ($CHCH_3$), 26.89 (t-Bu CH_3), 19.15(t-Bu C), 18.69 (CH_3CH). MS (ES) (m/z) calculated for $C_{52}H_{62}FeO_4Si_2$ 862.35, found 862.42 $[M+Na]^+$

(*R,R*)-1,1'-Bis(α-methyl-γ-(tert-butyldiphenylsilyl)oxypropyl)-(S_p,S_p)-2,2'-bisvinylferrocene (compound 12)

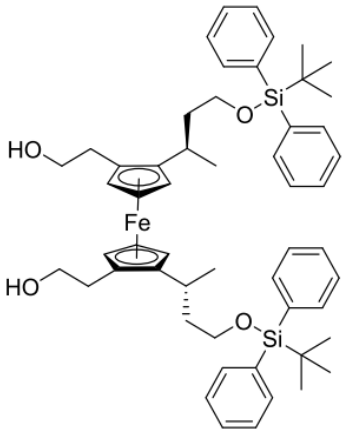


PH_3PMeBr (6.94 g, 19.5 mmol, 2 eq), 18-crown-6 (12.9 mg, 0.049 mmol, 0.005 eq), and KOtBu (2.184 g, 19.5 mmol, 2 eq) were dissolved in dry THF (30 mL) under an

inert atmosphere and stirred for 2 hours, resulting in a bright yellow solution. **Compound 11** (8.4 g, 9.75 mmol, 1 eq) was then dissolved in dry THF (60 mL) and added dropwise to the solution, followed by stirring at room temperature for 20 hours. The reaction was quenched with H_2O, and the mixture was extracted with DCM (3 x 40 mL) and dried with $MgSO_4$. After removing the solvent in vacuo, the crude product was purified by flash column chromatography using a pure hexane eluent system. The solvent was removed in vacuo, yielding the title compound as an orange solid (7 g, 8.16 mmol, 83.7%). The characterisation matched literature values.[1]

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.66 (m, 8H (Ph CH)), 7.38 (m, 12H (Ph CH)), 6.48 (2H dd, *J* = 17.4, 10.8 Hz , (CH=CH_2)), 5.25 (2H dd, *J* = 17.4, 1.8 Hz (CH_2=CH)), 5.08 (dd, J = 10.8, 1.8 Hz 2H (CH_2=CH)), 4.18 (dd, 2H (Fc CH)), 4.02 (m, 4H (Fc CH)), 3.61 (m, 4H (CH_2OSi)), 2.86 (dtd, 2H ($CHCH_3$)), 1.67 (m, 2H (CH_2CH)), 1.28 – 1.38 (m, 2H (CH_2CH)), 1.25 (d, 6H (CH_3CH)), 1.05 (s, 18H (tBu CH_3). ^{13}C NMR (400 MHz, $CDCl_3$) δ 135.62 (Ph CH), 135.60 (Ph CH), 134.1 (Ph C–Si), 134.02 (Ph C–Si), 132.85 (CH=CH_2), 129.54 (Ph CH), 129.50 (Ph CH), 127.61 (Ph CH), 127.58 (Ph CH), 111.07 (CH_2=CH), 94.56 (Fc C–CH=CH_2), 82.31 (Fc C–$CHCH_3$), 67.86 (Fc CH), 67.58 (Fc CH), 66.42 (Fc CH), 61.81 (CH_2OSi), 42.65 (CH_2CH), 26.88 (tBu CH_3), 26.68 ($CHCH_3$), 19.23 (tBu C), 18.45 (CH_3CH). MS (ES) (m/z) calculated for $C_{54}H_{66}FeO_2Si_2$ 858.3953, found 860.4091 $[M + 2H]^+$

(*R,R*)-1,1'-Bis(α-methyl-γ-(tert-butyldiphenylsilyl)oxypropyl)-(S_p,S_p)-2,2'-bis(βhydroxyethyl)ferrocene (compound 13)

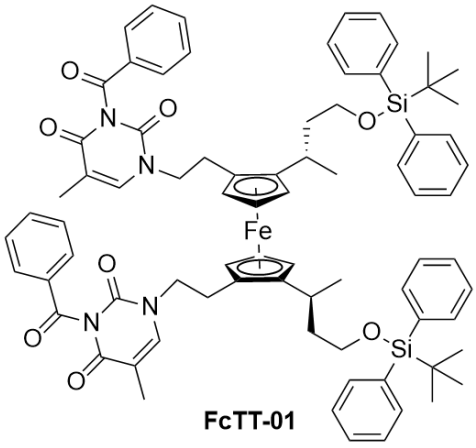


Compound 12 (7 g, 8.16 mmol, 1 eq) was dissolved in dry THF (50 mL) and cooled to 0°C. BH_3·THF (1 M in THF) (12.24 mL, 12.24 mmol, 1.5 eq) was added dropwise, and stirring was continued at 0°C for 1.5 hours. The reaction mixture was then warmed to room temperature, and stirring continued for an additional 2.5 hours. Ethanol (12.24 mL) was added, and stirring continued for a further 5 minutes. NaOH (3 M in H_2O) (12.24 mL) was then added dropwise, followed by an additional 15 minutes of stirring. Hydrogen peroxide (30% in H_2O) (30 mL) was then added dropwise, and stirring continued for another 2.5 hours. The reaction mixture was extracted with DCM, followed by washing with brine (50 mL) and water (2 x 50 mL), and drying over $MgSO_4$. The crude product was purified by flash column chromatography on silica gel using a gradient eluent system of 30 → 80% EtOAc in hexane, resulting in the title compound as yellow solid. The characterisation matched literature values.[1]

(3.7g, 4.14 mmol, 50.7%).^{1}H NMR (400 MHz, $CDCl_3$) δ 7.68 (m, 8H (Ph CH)), 7.40 (m, 12H (Ph CH)), 3.93 (m 2H (Fc CH)), 3.86-3.91 (m, 4H (Fc CH)), 3.74 – 3.61 (m, 8H (CH_2OH, CH_2–OSi)), 2.68 (m, 2H ($CHCH_3$)), 2.53 (t, 4H (Fc C–CH_2)), 1.72 (m , 2H

(CH_2CH)), 1.41 – 1.29 (m, 4H (CH_2CH, OH)), 1.23 (d, 6H (CH_3CH)), 1.06 (s,18H (tBu CH_3)). ^{13}C NMR (400 MHz, $CDCl_3$) δ 135.65 (Ph CH), 135.61 (Ph CH), 133.96 (Ph C–Si), 134.0 (Ph C–Si), 129.67 (Ph CH), 129.65 (Ph CH), 127.8 (Ph CH), 127.7 (Ph CH), 95.0 (Fc C–CH_2), 82.8 (Fc C–CH), 70.9 (Fc CH), 67.7 (Fc CH), 66.2 (Fc CH), 63.2 (CH_2OH), 62.1 (CH_2OSi), 42.7 (CH_2CH), 30.8 (Fc C–CH_2), 27.4 (CH), 27.0 (tBu CH_3), 19.6 (CH_3CH), 19.4 (tBu C). MS (ES) (m/z) calculated for $C_{54}H_{70}FeO_4Si_2$ 894.4162, Found 894.4155 $[M]^+$ Chiral HPLC retention time (see Chapter 3): 8.43 min.

(*R,R,S*$_p$,*S*$_p$)-1,1'-Bis-[α-methyl-(3-(tert-butyldiphenylsilyl)propyl)]-2-2'-Bis-[2-((N-3-benzoyl)-thymine)ethyl]-ferrocene （FcTT-01）

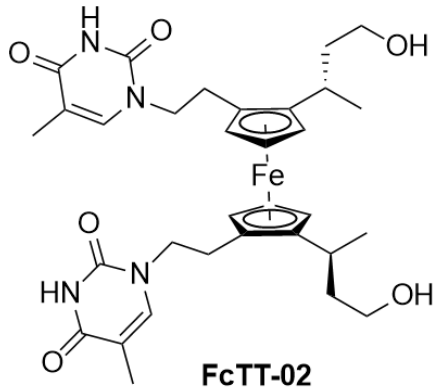


FcTT-01

PPh_3 (0.793 g, 3.03 mmol, 3 eq) and N3-benzoylthymine (0.46 g, 2.02 mmol, 2.6 eq) were dissolved in anhydrous THF (10 mL) and wrapped in foil. **Compound 13** (0.903 g, 1.01 mmol, 1 eq) dissolved in anhydrous THF (10 mL) was introduced, followed by the addition of diethyl azodicarboxylate (0.99 mL, 5.03 mmol, 3 eq). The reaction mixture underwent stirring at 65 °C for 1.5 h. The solvent was removed in vacuo and the crude product purified by flash column chromatography on silica gel using an eluent of 40% EtOAc in hexane, to give compound **FcTT-01** as a yellow oil (1.01g, 75%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.91 (m, 4H), 7.64-7.60 (m, 10H), 7.50 - 7.31 (m, 16H), 6.82 (s, 2H), 3.96 (s, 2H), 3.86 (s, 2H), 3.82 (s, 2H), 3.80 - 3.82 (m, 2H), 3.64 (m, 4H), 2.77 - 2.60 (m, 6H), 1.78 (s, 6H), 1.60-169 (m, 2H), 1.48 - 1.38 (m, 2H), 1.24 (d, 6H), 1.03 (s, 18H). ^{13}C NMR (400 MHz, $CDCl_3$) δ 169.20 (C=O), 163.23 (C=O), 149.70 (C=O), 140.30 (CH-thymine), 135.54 (Ph), 135.47 (Ph), 135.00(Ph), 133.89 (ipso-Ph), 133.81 (ipso-Ph) 131.65 (ipso-Ph), 130.49 (Ph), 129.71 (Ph), 129.14 (Ph), 127.70 (Ph), 110.41 (ipso-thymine), 94.35 (Fc C), 81.29 (Fc CH), 70.05 (CH-Cp), 68.06 (CH-Cp), 66.35 (CH-Cp), 61.91 (CH_2), 49.20 (CH_2), 42.65 (CH_2), 27.36 (CH), 27.23 (CH_2), 26.92 (CH_3), 19.89 (CH_3), 19.24 (ipso-t-Bu), 12.21(CH_3). MS (ES) (m/z) calculated for $C_{78}H_{88}N_4O_8FeSi_2Na$ 1341.5435, found 1341.5471 $[M+Na]^+$

(*R,R,S*$_p$,*S*$_p$)-1,1'-Bis(α-methyl,γ-tert-butyldiphenylsilyloxy-propyl)-2,2'-bis(thymine-propyl)ferrocene (FcTT-02)

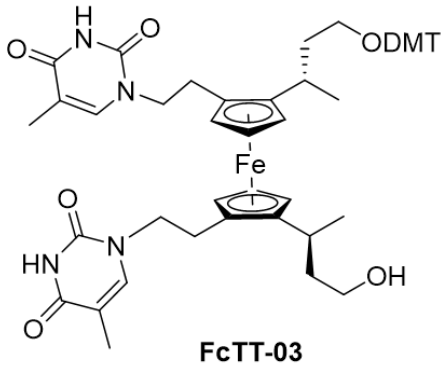


FcTT-02

FcTT-01 (0.67 g, 0.50 mmol, 1.0 eq) was dissolved in dry THF (3 mL). Tetra-n-butylammonium fluoride (1 M in THF) (3 mL, 3.00 mmol, 6.0 eq) was then added to the solution, and the resulting mixture was stirred for 2.5 hours. Subsequently, the solvent was evaporated under vacuum, and methylamine (3 mL) was added to the residue,

followed by stirring for 1 hour. The solvent was removed in vacuo and the crude product purified by flash column chromatography on silica gel using an eluent of 5% MeOH, 5%TEA in DCM. ^{1}H NMR (400 MHz, $CDCl_3$) δ 10.55 (s, H), 7.38 - 7.16 (m, H),6.99 (m, 3H), 4.10 - 3.84 (m, 12H), 3.83 - 3.52 (m, 6H), 3.06 - 2.73 (m, 7H), 2.64 (m, 4H), 2.00 - 1.85 (m, 3H), 1.81 (m, 9H), 1.71 (m, 2H), 1.43 (d, 12H), 1.36 (d, 9H). ^{13}C NMR (400 MHz, $CDCl_3$) δ 164.52(C=O), 151.53(C=O), 140.91(CH-thymine), 110.49(ipso thymine), 94.39(Fc C), 81.98(Fc C), 70.14(CH-Cp), 67.34(CH-Cp), 65.33(CH-Cp), 60.27(CH_2), 59.14, 49.54(CH_2), 43.03(CH_2), 26.94(CH), 26.52(CH_2), 24.22, 19.8, 19.11(CH_3), 13.7112.3(CH_3) MS (ES) (m/z) calculated for $C_{32}H_{42}FeN_4O_6$ 635.2484, found 635.2426 $[M]^+$

(*R,R,S*$_p$,*S*$_p$)-2,2'-Bis(ethyl-5-methylpyrimidine-2,4(1H,3H)-dione)-1,1'-bis(4,4'-dimethoxytrityl)butan-2-yl)ferrocene (FcTT-03)



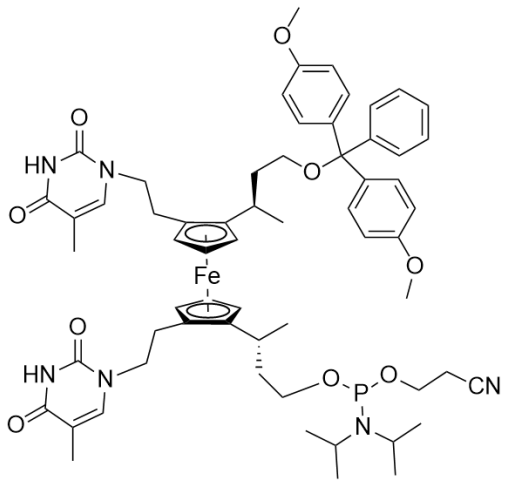
FcTT-03

In a 50 mL flask, **FcTT-02** (300 mg, 0.368 mmol) was dissolved in 15 mL of tetrahydrofuran (THF) at room temperature. Sequentially, diisopropylethylamine (DIPEA) (0.06 mL, 0.368 mmol), DMT chloride (140 mg, 0.410 mmol), and a catalytic amount of 4-dimethylaminopyridine (DMAP) were added to the solution. The mixture

was stirred for 16 h at room temperature and then quenched with 5 mL of methanol. Following this, water (20 mL) was introduced along with DCM (20 mL). After separation, the organic layer was dried over $MgSO_4$. The solvent was subsequently removed under vacuum, and the product was purified using flash column chromatography (5% methanol in DCM) to yield the oily yellow substance (175mg, 38%) The characterisation matched literature values.[2]

^{1}H NMR (400 MHz, $CDCl_3$) δ 9.59 (s, 1H), 7.4 (m, 1H), 7.2 (m, 5H), 7.06 (d, 12H), 6.7 (m, 2H), 6.62 (d, 1H), 3.9 (m, 3H), 3.79 (s, 3H), 3.6 (m, 2H), 3.08 (t, 12H), 2.7 (m, 2H), 1.89 (m, 1H), 1.72 (m, 1H), 1.35 (d, 1.24 (m, 12H). ^{13}C NMR (400 MHz, $CDCl_3$) δ: 164.39(C=O), 164.39(C=O), 158.42(ipso-DMT), 151.29(C=O), 151.13(C=O), 145.22(ipso-Ar), 140.7(CH-thymine), 136.46(ipso-Ar), 136.38(ipso-Ar), 129.97(CH-Ar), 128.11(CH-Ar), 127.74(CH-Ar), 126.76(CH-Ar), 113.01(CH-Ar), 110.76(ipso thymine), 110.22(ipso thymine), 94.7(Fc C), 94.1(Fc C), 86.03((Fc-C)), 81.6(Fc-C), 70.61(Fc-C), 67.95(CH-Cp), 66.01(CH-Cp), 65.98(CH-Cp), 61.49(CH_2), 60.22(CH_2), 55.24(OMe), 53.44, 49.89(CH_2), 49.06(C, H_2), 43.29(CH_2), 40.37(CH_2), 27.89(CH), 27.18(CH_2), 27.03(CH_2), 26.81(CH), 20.11(CH_3), 19.2(CH_3), 12.32(CH_3), 12.14(CH_3) MS (ES) (m/z) calculated for $C_{53}H_{60}N_4O_8FeNa$ 959.3658, found 959.3682 $[M+Na]^+$ The characterisation matched literature values.[2]

FcTT-04 Phosphoramidite

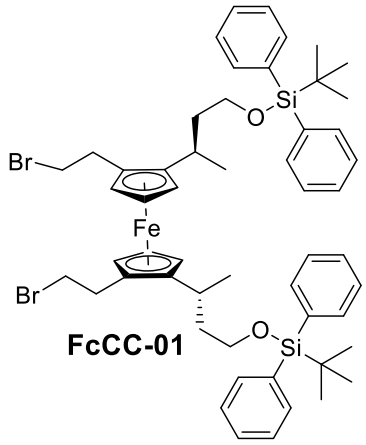


FcTT-04

FcTT-03 (192 mg, 0.20 mmol), azeotroped with dry acetonitrile (2 x 10 mL), was re-dissolved in dry DCM (10 mL). Then DIPEA (0.70 mL, 4.0 mmol) was added, followed by 2-cyanoethyl N,N diisopropyl chlorophosphoramidite (0.06 mL, 0.27 mmol). The reaction was stirred for 2 hr under argon. Degassed EtOAc(20 mL) was then added and the mixture washed with sat. $NaHCO_3$ (20 mL), brine (20 mL) and then dried over $MgSO_4$. The solvent was then removed under vacuum, and the product was purified using flash column chromatography to yield the product as a yellow foam (178 mg, 78%). The characterisation matched literature values.[2]

^{31}P NMR (300 MHz, $CDCl_3$) δ 147.18 and 146.92 ppm; ^{1}H NMR (300 MHz, $CDCl_3$) δ 8.92 (s,2H,NH), 7.38 (d, 2H), 7.3-7.17 (m, 9H), 6.90 (d, 1H), 6.81-6.77 (d, 4H), 6.67 (s, 1H), 3.95-3.77 (m,12H), 3.78 (s, 6H, OCH_3), 3.67-3.47 (m, 4H), 3.11-3.01 (2H), 2.91-2.75 (2H), 2.75-2.50 (6H), 1.87 (s, 3H, CH_3), 1.73 (s, 3H, CH_3), 1.70-1.62 (m, 2H,), 1.56-1.43 (2H), 1.37-1.11 (m, 20H); MS (ES) (m/z) calculated for $C_{62}H_{77}N_6O_9FePNa$, 1159.4734 , found 1159.4737$[M+Na]^+$

(*R,R,S*$_p$,*S*$_p$)-1,1'-Bis-[α-methyl-(3-(tert-butyldiphenylsilyl)propyl)]-2,2'-bis-[2-(bromo)ethyl]-ferrocene (FcCC-01)



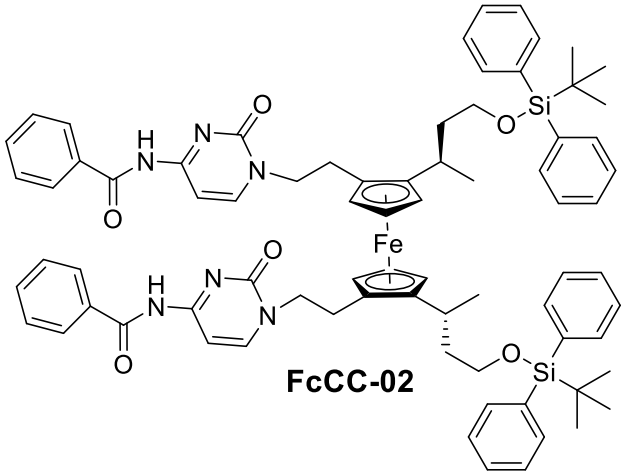
FcCC-01

N-Bromosuccinimide (NBS) (0.96 g, 5.4 mmol, 6 molar equivalents) was dissolved in dry DCM. Then a DCM solution of PPh_3 (1.35 g, 15.1 mmol, 5.7 molar equivalents) was added. The resulting mixture was stirred for 10 minutes before the addition of pyridine (0.1 mL, 2.5 mmol, 2.5 equivalents). Then, **compound 13** (0.9 g, 0.9 mmol, 1 equivalent) dissolved in dry DCM (10 mL) was added to the mixture. After stirring for 2 hours at room temperature, the solvent was removed under vacuum, and the crude product was purified using flash column chromatography (2% EtOAc in hexane), resulting in the formation of a yellow-brown oily product. The characterisation matched literature values.[3]

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.66 (m, 8H), 7.40 (m, 12H), 3.94 – 3.89 (m, 4H), 3.87 (t, 2H), 3.66 – 3.58 (m, 4H), 3.46 – 3.33 (m, 4H), 2.95 – 2.78 (m, 4H), 2.68 (ddp, 2H), 1.64 (dtd, 2H), 1.34 (ddt, 2H), 1.22 (d, 6H), 1.07 (s, 18H). ^{13}C NMR (400 MHz, $CDCl_3$) δ 135.63 (Ph), 135.59, 133.92 (ipso Ph), 133.82 (ipso Ph), 129.64(Ph), 129.61 (Ph), 127.69(Ph), 127.68(Ph), 94.46 (Fc C), 83.17 (Fc C), 70.20 (Fc), 67.44 (Fc), 66.17 (Fc), 61.72 (CH_2), 42.75 (CH_2), 31.62 (CH_2), 27.19 (CH), 26.97 (tBu), 19.49 (CH_3), 19.23

(tBu). MS (ES) (m/z) calcd for $C_{54}H_{68}O_2FeSi_2Br_2Na$ 1020.2454, found 1020.2495 (M+ Na). αD (c = 0.0050 g/ mL, DCM) = -10.25

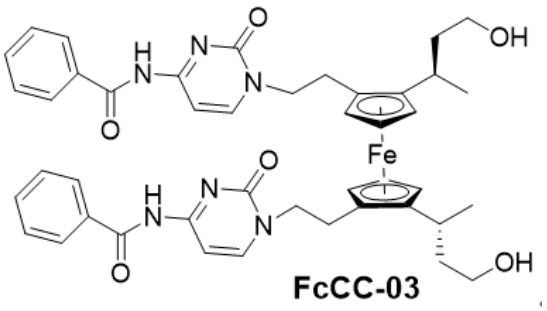
(*R,R,Sp,Sp*)-1,1'-Bis-[α-methyl-(3-(tert-butyldiphenylsilyl)propyl)]-2-2'-Bis-[2-((N-3-benzoyl)-cytosine -1-yl)ethyl]-ferrocene (FcCC-02)



Bz-cytosine (1032 mg, 4.8 mmol. 8 eq) was stirred vigorously in suspension in DMF (5 mL) for 5 min. NaH (90%) (120.5 mg, 4.8 mmol, 8 eq) was added to the mixture string for 1 hr. After **FcCC-01** (700 mg, 0.68 mmol) was dissolved in DMF (5 mL) that is added to the mixture. After stirring at 70°C for 24h, the reaction is quenched with water, extracted with chloroform, dried on $MgSO_4$. The solvent was removed in vacuo and the crude was purified via flash column chromatography (2% MeOH in DCM) to yield product as a yellow oil (0.478 g, 54%). ^{1}H NMR (400 MHz, $CDCl_3$) δ 8.71 (s, 2H), 7.90 (d, 4H), 7.70 – 7.56 (m, 10H), 7.50 (m, 4H), 7.44 – 7.29 (m, 14H), 7.20 (d, 2H), 4.05 – 3.97 (m, 2H), 3.96 (s, 2H), 3.88 (s, 2H), 3.78 (s, 2H), 3.74 (m, 2H), 3.70 – 3.61 (m, 4H), 3 2.82 (s, 2H), 2.72 – 2.59 (m, 4H), 1.70 – 1.57 (m, 2H), 1.47 (ddt, 2H), 1.25 (d, 6H), 1.02 (s, 18H). ^{13}C NMR (400 MHz, $CDCl_3$) δ 166.19 (Cytosine C=N), 162.09 (C=O Bz),

155.49 (C=O cytosine), 149.11 (CH cytosine), 135.55 (CH Ph), 135.49 (CH Bz), 133.88 (C Ph), 133.81 (C Bz), 133.05 (CH Bz), 129.70 (CH Bz), 128.98 (CH Ph), 127.70 (CH Bz), 127.61 (CH Ph), 96.20 (CH), 94.67 (Fc-C), 81.65 (Fc-C), 70.59 (Fc), 68.19 (Fc), 66.55 (Fc), 61.99 (CH_2), 51.41 (CH_2), 42.68 (CH_2), 27.32 (CH), 26.91 (CH_3 tBu), 20.07 (CH_3), 19.20 (ipso tBu). IR v_{max}/cm^{-1}= 2929 (CH), 1663(C=O). MS (ES) (m/z) calculated for $C_{76}H_{85}N_6O_6Fe$ 1289.5419, found 1289.5415$[M]^+$. αD (c = 0.0064 g/mL, DCM) = 6.06

(*R,R,Sp,Sp*)-1,1'-Bis(α-methyl,γ-tert-butyldiphenylsilyloxy-propyl)-2,2'-bis(γ-N-3- benzoylcytosine-propyl)ferrocene (FcCC-03)



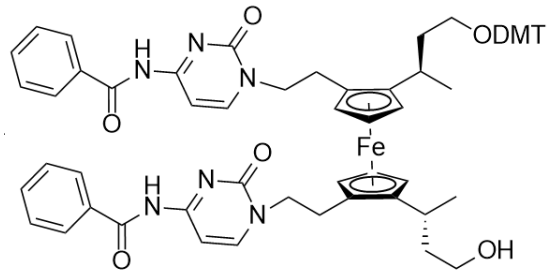
FcCC-03

FcCC-02 (0.4 g, 0.50 mmol) was dissolved in dry THF (3 mL). Tetra-n-butylammonium fluoride (1 M in THF) (3 mL, 3.00 mmol) was then added to the solution, and the resulting mixture was stirred for 2.5 hours. The reaction was quenched with water, extracted with DCM, washed with brine and dried over Na_2SO_4. The solvent was removed in vacuo and the crude was purified via flash column chromatography (5% MeOH in DCM) to yield product as a yellow oil (176 mg, 70%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 8.87 (s, 2H), 7.81 (dd, 6H), 7.61 (d, 2H), 7.57 – 7.48 (m, 2H), 7.40 (t, 4H), 4.21 – 4.00 (m, 6H), 3.89 (s, 2H), 3.85 – 3.69 (m, 4H), 3.67 – 3.59

(m, 2H), 3.06 (s, 2H), 2.99 - 2.86 (m, 4H), 2.76 - 2.56 (m, 2H), 1.83 - 1.69 (m, 2H), 1.45 (m, 2H), 1.35 (d, 6H). ^{13}C NMR (400 MHz, $CDCl_3$) δ 162.38(C=O Bz), 149.55(C=O), 133.08(Ph), 132.96(Ph), 128.95(Ph), 127.54(Ph), 96.98, 95.00(Fc C), 81.61(Fc C), 70.97(Fc), 67.85(Fc), 65.70(Fc), 60.17, 52.36(CH_2), 43.72(CH_2), 27.07, 26.57, 19.05. IR v_{max}/cm^{-1}= 3348 (OH), 1644(C=O). MS (ES) (m/z) calculated for $C_{65}H_{66}N_6O_8FeNa$ 835, found 835.2872 [M+Na]+. αD (c = 0.0015 g/mL, DCM) = +68.33.

(*R,R,Sp,Sp*)-2,2'-Bis(bis(γ-N-3- benzoylcytosine-propyl))-1,1'-bis(4,4'-dimethoxytrityl)butan-2-yl)ferrocene FcCC-04



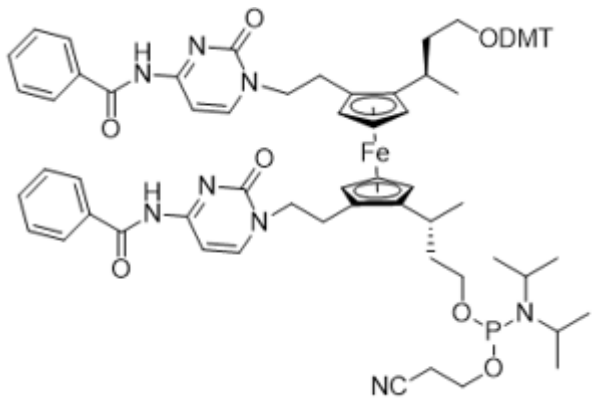
FcCC-04

In a 50 mL flask, **FcCC-03** (300 mg, 0.368 mmol) was dissolved in dry THF (15 mL) at room temperature. Sequentially, diisopropylethylamine (DIPEA) (0.06 mL, 0.368 mmol), DMT chloride (140 mg, 0.410 mmol), and a catalytic amount of 4-dimethylaminopyridine (DMAP) were added to the solution. The mixture was stirred for 16 h at room temperature and then quenched with 5 mL of methanol. Following this, water (20 mL) was introduced along with DCM (20 mL). After separation, the organic layer was dried over $MgSO_4$. The solvent was subsequently removed under vacuum,

and the product was purified using flash column chromatography (5% methanol in DCM) to yield the oily yellow substance (151mg, 37%) ^{1}H NMR (500 MHz, Chloroform-*d*) δ 9.02 (s, 2H), 7.88 (t, 4H), 7.82 (d, 1H), 7.55 (q, 2H), 7.45 (m, 5H), 7.40 - 7.34 (m, 2H), 7.34 - 7.12 (m, 9H), 6.79 (d, 4H), 4.14 - 4.02 (m, 1H), 4.03 - 3.95 (m, 4H), 3.91 (dt, 2H), 3.83 (t, 1H), 3.75 (s, 6H), 3.69 - 3.61 (m, 2H), 3.59 (m, 1H), 3.41 (s, 1H), 3.06 (m, 2H), 2.88 (m, 3H), 2.77 (m, 2H), 2.68 (m, 2H), 1.73 - 1.58 (m, 2H), 1.50 (m, 1H), 1.40 (m, 1H), 1.29 (d, 3H), 1.23 (d, 3H). ^{13}C NMR (400 MHz, Chloroform-*d*) δ 162.45(C=O), 162.31(C=O), 158.40, 155.88, 155.59, 149.64, 149.07, 145.10, 136.43, 136.41, 133.13, 133.08(Ph), 133.05(Ph), 129.97, 129.95, 128.98, 128.93, 128.17, 127.78, 127.70, 127.66, 126.77, 113.05, 96.99, 96.46, 94.98 (Fc-C), 94.14 (Fc-C), 86.06, 81.49, 81.42, 77.39, 77.14, 76.88, 70.88, 68.18, 68.16, 66.15, 65.89, 61.47, 55.24, 52.27, 51.21, 43.66, 40.39, 27.88, 27.39, 27.04, 26.67, 20.34, 19.30.

IR ν_{max}/cm^{-1} = 3397 (OH), 2961 (CH), 1651(C=O). MS (ES) (m/z) calculated for $C_{65}H_{66}N_6O_8FeNa$ 1137.4292, found 1137.4191 $[M+Na]^+$. αD (c = 0.0022 g/mL, DCM) = -42.81

FcCC-05 Phosphoramidite



In a 100 mL round bottom flask, **FcCC-04** (235 mg, 0.211 mmol) was azeotroped with 10 mL dry acetonitrile, redissolved in 10 mL dry DCM. DIPEA (0.37 mL, 2.107 mmol) was added followed by 2- cyanoethyl chlorophoshoramidite (0.064 mL, 0.284 mmol). The reaction was stirred under argon for 2 hr. Degassed EtOAc (20 mL) was added and the mixture was washed with sat $NaHCO_3$ (10 mL) and brine (10 mL), dried over $MgSO_4$. The solvent was removed in vacuo and the crude was purified via flash column chromatography (1% NEt_3 in EtOAc) to give a foamed product (210 mg). ^{1}H NMR (300 MHz, Acetonitrile-d3) δ 8.6 (s, 2H), 7.93 - 7.83 (m, 4H), 7.75 (m , 1H), 7.61 - 7.46 (m, 6H), 7.40 - 7.34 (m, 2H), 7.27 - 6.95 (m, 10H), 6.79 (d, 4H), 4.22 - 4.06 (m, 6H) 4.03 - 3.8 (m, 12H), 3.76 (s, 6H), 3.62 - 3.42 (m, 8H), 2.80 (m, 1H), 2.58 (m, 6H), 1.68 (m, 2H), 1.56 (m, 2H), 1.33 - 1.22 (m, 20H), 1.2 - 1.10 (m, 13H). ^{31}P NMR (300 MHz, $CDCl_3$) δ 147.15, 147.09, 10.6, 17.89. MS (ES) (m/z) calcd for $C_{74}H_{83}N_8O_9PFe$ 1314.54, found 1315.5450 $[M+H]^+$

6.2 Oligonucleotide synthesis, purification and characterisation.

6.2.1 Oligonucleotide synthesis

Oligonucleotides (see list in Table S1) were synthesised using an Applied Biosystems ABI 394 synthesiser. PacdA, iPr-Pac-dG, Ac-dC and dT phosphoramidites were purchased from LGC Genomics. The abasic nucleotide and 3'-thiol-modifier 6 S-S CPG columns were purchased from Glen Research. The C18- spacer modified oligonucleotides were purchased from Integrated DNA Technologies. Unmodified strands were purchased from Merck and used as received.

Table 6.1 Oligo sequences with predicted masses, mass found and HPLC retention time.

Oligo	Sequence from 5'to 3'	Predicted Exact Mass	Mass found	Retention time (min)
B_1	ACATTCCTAAGTCTGAAA**Fc**CATTACAGCTTGC TACACGAGAAGAGCCGCCATAGTA-C6-SS	17567.98	17565.225	n/a
B_2	TATCACCAGGCAGTTGAC**Fc**AGTGTAGCAAGCT GTAATAGATGCGAGGGTCCAATAC-C6-SS	17719.98	17717.025	n/a
B_3	TCAACTGCCTGGTGATAA**Fc**AACGACACTACGT GGGAATCTACTATGGCGGCTCTTC-C6-SS	17597.94	17595.119	n/a
B_4	TTCAGACTTAGGAATGTG**Fc**CTTCCCACGTAGT GTCGTTTGTATTGGACCCTCGCAT-C6-SS	17576.89	17574.119	n/a
C_1	ACATTCCTAAGTCTGAAACATTACAGCTTGCTA CACGAGAAGAGCCGCCATAGTA-C6-SS	17203.93	17201.578	n/a
C_2	TATCACCAGGCAGTTGACAGTGTAGCAAGCTGT AATAGATGCGAGGGTCCAATAC-C6-SS	17355.93	17353.375	n/a
C_3	TCAACTGCCTGGTGATAAAACGACACTACGTGG GAATCTACTATGGCGGCTCTTC-C6-SS	17233.89	17231.463	n/a
C_4	TTCAGACTTAGGAATGTGCTTCCCACGTAGTGT CGTTTGTATTGGACCCTCGCAT-C6-SS	17212.84	17210.445	n/a
D1	ACATTCCTAAGTCTGAAA**Fc**CATTACAGCTTGC TACACGAGAAGAGCCGCCATAGTA	17235.92	17233.514	n/a
M1Fc	GGA G**Fc**G AGG CTA TAG TGA GTC GTA	7569.29	7569.48	22.69
AP1	GGA G**AP**G AGG CTA TAG TGA GTC GTA	7385	7385.77	16.73

AP2	GGA G**APAP**G AGG CTA TAG TGA GTC GTA	7565.36	7565.81	17.06
M1FcHH	GGA G**FcHH**G AGG CTA TAG TGA GTC GTA	7600.94	7601.2	13.76
M1FcTT	GGA G**FcTT**G AGG CTA TAG TGA GTC GTA	7905.24	7905.3	13.472
M1FcCC	GGA G**FcCC**G AGG CTA TAG TGA GTC GTA	7875.22	7876.3	11.84
S1HH	TGG ACT C**FcHH**C TCA ATG	4631.02	4630.8	13.61
S1TT	TGG ACT C**FcHH**C TCA ATG	4935.32	4935.9	13.11
S1CC	TGG ACT C**FcHH**C TCA ATG	4905.30	4904.9	11.67

All phosphoramidites were dissolved in anhydrous acetonitrile (0.1 M) prior to synthesis. Strands were synthesised at a 1 µmol scale. Removal of 4,4'-dimethoxytrityl (DMTr) was achieved using trichloroacetic acid in dichloromethane (DCM). Phosphoramidites were activated with 5-ethylthio-1H-tetrazole (0.25 M) in acetonitrile prior to coupling; coupling times of 25 s were used. Unreacted material was capped using acetic anhydride and methylimidazole. Iodine (0.02 M) in THF/pyridine/water (7:2:1) was used to oxidise the phosphotriesters formed. Upon completion, each oligonucleotide was treated with aqueous ammonia (30% v/v) for 1hr at room temperature, and then heated at 60°C for 6 h, to cleave the strands from the resin and remove protecting groups. The solvent was removed using a Thermo Scientific speed vac and the crude product was redissolved in Milli-Q water (0.9 mL) ready for purification. Strands A1-A4 were purchased from Merck and were used as received.

6.2.2 Purification of oligonucleotides

6.2.2.1 HPLC

Semi-preparative RP-HPLC purification was conducted using an Agilent Technologies 1260 Infinity system with a Phenomenex Clarity 5 μm Oligo-RP LC 250 × 10 mm column. The column was maintained at 60 °C, and UV/vis absorbance at 260 nm was monitored for each run. Buffer D, consisting of 0.1 M triethylammonium acetate (TEAA) in HPLC grade H_2O, was employed. Each sample, 1 mL in volume, was injected at a flow rate of 3 mL min^{-1}. Oligonucleotides were purified through a gradient method of 5–18% acetonitrile in Buffer D over 30 minutes for Fc and AP modified strands; 10–35% acetonitrile in Buffer D over 30 minutes for FcHH, FcTT, FcCC modified strands. Collected fractions were evaporated to dryness, diluted to 1 mL in Milli-Q water, and desalted using a NAP-10 column (GE Healthcare) with elution to 1.5 mL. Then remove the water and redissolve the sample in 1 mL of Milli-Q water.

The purity of oligonucleotides was assessed via analytical HPLC using a Phenomenex Clarity 5 μm Oligo RP LC 250 × 4.6 mm column on an Agilent Technologies 1260 Infinity system. The column was heated to 60 °C, and UV/vis absorbance at 260 nm was monitored. A sample volume of 20 μL was injected, with a run time of 45 min per sample at a flow rate of 1 mL min^{-1}. The solvent gradients employed were identical to those used in semi-preparative HPLC. Samples demonstrating >95% purity by analytical HPLC were considered adequately pure for use in experiments.

6.2.2.2 Strand purification and characterisation using PAGE

Denaturing polyacrylamide gel electrophoresis (PAGE) was employed for DNA purification, utilizing a 15% acrylamide gel and electrophoresis in 1× TBE buffer (0.1 M Tris base, 0.1 M boric acid, 2 mM EDTA). Samples containing 200 μM DNA (200 μL) were combined with formamide (200 μL) and incubated for 25 min before loading into the wells, with 15 μL in each well. Electrophoresis was conducted for 2 h at 4 °C (100 V) in TBE buffer. Gel visualization was achieved under a UV lamp by observing a shadow on a silica plate, and the most intense band corresponding to the desired product was excised using a scalpel.

The excised gel slabs were crushed with a spatula, and for every 0.5 mL of gel, 2 mL of MilliQ water was added. The samples were then rapidly frozen in liquid nitrogen and immediately thawed to 90 °C, followed by a 10-min incubation at that temperature. Subsequently, the treated samples were placed on a shaker for 24 h. Gel extraction was then performed, with the gel residue being filtered and rinsed with MilliQ water. The eluted DNA was dried under vacuum and subjected to desalting using NAP-25 columns (GE Healthcare).

6.2.3. Mass Spectrometry

The masses of the DNA oligonucleotides (after treatment with NAP column to remove salts) were ascertained employing a Waters Xevo G2-XS system equipped with a Time-of-Flight (TOF) detector utilizing negative mode electrospray ionization mass spectrometry (ESI-MS). Deconvolution of the raw data was achieved through Promass,

and molecular mass prediction was obtained using ChemDraw software. The anticipated and observed exact masses are detailed in Table 6.1.

6.3 DNA tetrahedra assembly

Each of the four DNA components comprising each tetrahedron was prepared to a concentration of 100 μM using Milli-Q water. Subsequently, 1 μL of each of the four strands was combined with 66 μL of TM buffer (Tris base 20 mM; $MgCl_2$, 50 mM, pH 8), along with 30 μL of TCEP (10 mM), resulting in a final concentration of 1 μM for each strand. The resultant solution was then subjected to heating at 95 °C in a heating block for 10 min, followed by rapid cooling in ice water for 1 min to facilitate the formation of the tetrahedron.

6.3.1 Tetrahedron formation studies by gel

In order to study their assembly, native polyacrylamide gel electrophoresis (PAGE) was performed vertically with a 10% acrylamide gel and 1 × TBE buffer. Pre-annealed samples of T1-T8 containing 0.5-1 μM DNA were mixed with 30% glycerol in a 5:1 volumetric ratio and loaded into the wells alongside a GeneRuler Ultra Low Range DNA ladder (Thermo Scientific). Electrophoresis experiments were then undertaken at 100 V for 1 h 30 min at 4 °C in electrophoresis buffer (1 × TBE).

6.4 Electrochemistry

All cleaning and sensing procedures were conducted using a 3-electrode setup consisting of a BioAnalytical Systems Inc. EC Epsilon potentiostat, a C3 cell stand, a WE with a polycrystalline gold disc, a platinum wire counter electrode, and a Ag/AgCl reference electrode (3 M KCl). Potentials are quoted with respect to this Ag/AgCl reference electrode. All electrodes were purchased from BioAnalytical Systems Inc. All water used (Milli-Q water) was purified for cleaning and solution preparation using a Merck Millipore Elix-Gradient A10 system (resistivity > 18 μΩ cm, toc ≤ 5 ppb). All solutions were deoxygenated with argon before use.

Cells and glassware were cleaned before use by soaking in a 1:1 mixture of aqueous ammonia (30% v/v) and hydrogen peroxide (30% v/v) for 4 h, followed by rinsing with copious quantities of Milli-Q water and soaking in Milli-Q water overnight. Electrochemical cells were then dried in a designated clean oven (80 °C) before use.

6.4.1 Self-Assembled Monolayer Preparation.

The 2.0 mm diameter gold working electrodes were polished using a 1 μm diamond suspension for 3 minutes, followed by a 1 μm, 0.3 μm and 0.05 μm alumina suspension for 3, 3 and 5 min, respectively. The electrodes were rinsed with Milli-Q water between each polish. Next, the electrodes were subjected to electrochemical cleaning in 0.5 M H_2SO_4: chronoamperometry (2 V for 5 s, –350 mV for 10 s) and cyclic voltammetry

(CV) over a range –350 to 1500 mV at 4 V s^{-1}, for 20 cycles, 500 mV s^{-1} for 4 cycles and 100 mV s^{-1} for cycle).; Polished gold electrodes were soaked in the prepared tetrahedron sample vial overnight.

6.4.2 Voltammetry measurements

CV experiments were carried out in 1 M $NaClO_4$ + 10 mM sodium phosphate buffer (pH 7.0). CVs were conducted before and after target addition at scan rates of 1000, 500, 250, 100, 80, 60 and 40 mV s^{-1} in a potential window of -75 mV to 300 mV.

6.5 Primer extension reaction

This procedure was taken from a published paper of ours.[4] A 5′-FAM-labelled primer (10 pmol) was annealed to the appropriate template (15 pmol) in MilliQ H_2O by heating to 95 °C and gradually cooling to room temperature over 1 h. The appropriate DNA polymerase, the relative buffer (1 μL of 10×), and the dNTP(s) were then added to the reaction mixture for a total reaction volume of 10 μL. The reaction mixture was incubated for the appropriate time at the optimal temperature for the polymerase. The reaction was stopped by adding 10 μL of the quenching solution (formamide (70%), ethylenediaminetetraacetic acid (EDTA, 50 mM), bromophenol (0.1%), xylene cyanol (0.1%)). The reaction mixtures were subjected to gel electrophoresis in denaturing polyacrylamide gel (20%) containing Trisborate-EDTA (TBE) 1× buffer (pH 8) and urea (7 M). Visualization was performed with fluorescence imaging using a Typhoon Trio phosphorimager.

6.6 DNA NMR sample preparation

The **FcCC/TT** modified strands and their complementary counterparts were transferred into separate Eppendorf tubes, yielding a 1000μM (or 600 μM for 1D NMR) solution in Milli-Q water, subsequently evaporated the water. Additionally, a pH 7 buffer comprising 100 mM NaCl and 10 mM phosphate buffer was prepared. Following this, 80 μL of the buffer was added to each tube to dissolve the strands, pairing AA with TT and GG with CC to form 160 μL solutions. These solutions were then heated to 95°C for 10 minutes using a heating block. Upon removal from the block, the samples were allowed to cool to room temperature over a duration of approximately 20-30 minutes. Subsequently, 20 μL of D_2O was added and thoroughly mixed. The resulting samples were then ready for NMR measurement.

References

1. Kedge, J.; *PhD thesis*, University of Birmingham, **2016**.
2. Nguyen, H. V., Zhao, Z. Y., Sallustrau, A., Horswell, S. L., Male, L., Mulas, A., & Tucker, J. H. *Chemical communications*, **2012**, 48(100), 12165-12167.
3. Sallustrau, A. *PhD thesis*, University of Birmingham, **2014**.
4. Figazzolo, C., Ma, Y., Tucker, J. H. R, & Hollenstein, M. *Organic & Biomolecular Chemistry*, **2022**., 20, 8125-8135.