

THE UNIVERSITY of ADELAIDE

Biophysical characterisation of DNA triplexes for antigene applications

Jack Klose

B.Sc. Molecular and Drug Design (Biochemistry & Chemistry)

Honours degree of B.Sc (Chemistry)

A thesis submitted in fulfilment of the requirements for

the degree of Doctor of Philosophy in Chemistry

at the University of Adelaide

December 2023

Supervisor: Prof Tara Pukala

Co-supervisor: A/Prof David Huang

Table of Contents

Acknowledgments ..... ***i***

Declaration ..... ***ii***

Abstract ..... ***iii***

Chapter 1: Molecular Insights into Nucleic Acid Triplexes: Methodologies and Biological Applications ***6***

1.1 Authorship statement 6

1.2 Review manuscript 8

1.3 Aims 43

Chapter 2: Experimental methods ***44***

2.1 Introduction 44

2.2 Mass spectrometry 44

2.2.1 Electrospray ionisation 45

2.2.2 Quadrupole time-of-flight mass spectrometry 46

2.2.3 Collision-induced dissociation 49

2.3 UV-Vis spectroscopy 50

2.4 Isothermal titration calorimetry 52

2.5 References 54

Chapter 3: Native Mass Spectrometric Insights into the Formation and Stability of DNA Triplexes ***57***

3.1 Foreword and authorship statement 57

3.2 Manuscript 60

3.3 Supplementary material 70

Chapter 4: Unraveling DNA Triplex Stability: Mass Spectrometric Investigation of Modified Triplex Forming Oligonucleotides for Enhanced Gene Targeting ***78***

4.1 Foreword and authorship statement 78

4.2 Manuscript 81

4.3 Supplementary material 90

4.4 Supporting experiments 95

4.4.1 Introduction 95

4.4.2 Materials and methods 96

4.4.3 Results and discussion: 97

4.4.4 Conclusions 101

4.4.5 References: 102

4.4.6 Appendix 1: Name and sequences for each constituent strand for the model SEQ2 and SEQ3 triplex systems containing both the cytosine and thymine interruption in the TTS, including canonical DNA and dSpacer TFOs 103

Chapter 5: A triplex antigene strategy targeted towards methicillin-resistant Staphylococcus aureus ***107***

5.1 Introduction 107

5.2 Methods 110
5.2.1 Bioinformatic analysis of triplex target sites in methicillin-resistant Staphylococcus aureus 110
5.2.2 Sample preparation 110
5.2.3 Synthesis of TFO/cell-penetrating peptide conjugate 111
5.2.4 Native mass spectrometry 111
5.2.5 UV-Vis spectroscopy 111
5.2.6 Antimicrobial assay 112
5.3 Results & discussion 113
5.3.1 Identification of triplex target sites from the MRSA252 genome 113
5.3.2 Formation of the *rlmN* DNA triplex 114
5.3.3 Synthesis of 5' and 3' TFO-cell-penetrating peptide conjugates 115
5.3.4 *rlmN* TFO-cell-penetrating peptide conjugate triplex formation 117
5.3.5 Antimicrobial efficacy of 3' *rlmN* TFO-cell-penetrating peptide conjugate in *S. aureus* 121
5.4 Conclusion 124
5.5 References 126
5.6 Appendix 1: Input and summary from Triplexator 129
5.7 Appendix 2: *MRSA* TTSs identified from triplexator, along with candidate *rlmN* gene triplex sequence 131
Chapter 6: Binding of anthocyanins to DNA triplexes 136
6.1 Introduction 136
6.2 Experimental methods 139
6.2.1 Sample preparation 139
6.2.2 Native mass spectrometry 140
6.2.3 UV-Vis spectroscopy 140
6.3 Results and discussion 141
6.3.1 Binding of anthocyanin standards to model and biological DNA triplex sequences 141
6.3.2 Comparing binding affinities of cyanidin 3-glucoside and malvidin 3,5-diglucoside to the model and *MRSA* triplex systems 146
6.3.3 Stability studies of cyanidin 3-glucoside and malvidin 3,5-diglucoside bound DNA triplexes 150
6.3.4 Binding of netropsin to anthocyanin triplex systems 152
6.4 Conclusion 157
6.5 References 159
6.6 Appendix 1: Mass spectra of anthocyanin binding affinity experiments 163
Chapter 7: Conclusions 165

Acknowledgments

Well, I'll be perfectly honest, I didn't think I'd make it here. It has been an extraordinary journey that I have been on over the past 5 years, undertaking my postgraduate studies in the Pukala lab. Tara, from day one you have been the most supportive supervisor that any honours/PhD student could have asked for and I'm not sure if I would have been able to complete this thesis if it wasn't for you, so thank you. Gerardo, Lewis, River, Ruth, Sarveenah, Siqi, Shaun and Mr Mama, I would seriously like to apologise to you all for putting up with my constant badgering, screaming, yelling, and complaining. From impromptu coffee runs, to word games and the countless "productivity hacks", I can't thank you guys enough for being an amazing group of people to work with. A huge thank you to BJ for always being available for a chat and dropping anything to help me out over the last 5 years, long live the Jovcevski research group! I would also like to thank all past Pukala lab members and friends in the Chemistry department, especially Matt Bull for single handedly keeping all the instruments alive for many years. Of course, I cannot forget Ken from microscopy.

To all undergraduate chemistry friends, Takuto, Nic, Dempsey, Matt, Kylie[2] and Darcy for all the amazing times shared in lectures, labs and the Unibar. To my high school friends, Citti, Drew, Fuller, Hatzi, Kal, Mnuz, Premrl, Oli and Sonza, I want to thank you all for your support (not really) and I appreciate every one of you.

To my parents, Peter and Sheridan and sister Asha for your constant support for as long as I can remember. I will be forever grateful to have grown up in such an amazing family.

Mandi, I am incredibly thankful for all the support and encouragement you have given me over the last couple of years. Introducing me to bubble tea propelled my productivity infinitely, so I can't thank you enough for that. Thank you for being extremely understanding and putting up with me, I love you so much.

I also can't forget about all the support animals over the years of my PhD, Totoro, Obi and Yoshi.

Buckle your seatbelts…..

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree.

The author acknowledges that copyright of published works contained within the thesis resides with the copyright holder(s) of those works.

I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Jack Klose 22nd December 2023

Abstract

Nucleic acids are fundamental biomacromolecules that act as essential carriers of genetic information and play crucial roles in cellular processes of living organisms. The structure and functionality of duplex deoxyribonucleic acid (DNA) is well characterised in biology, however higher-order DNA structures such as DNA triplexes are relatively poorly understood. DNA triplexes are constituted from the canonical double-stranded DNA and a third strand known as the triplex forming oligonucleotide (TFO). The TFO binds to polypurine stretches within the major groove of the duplex via Hoogsteen hydrogen bonding. It has been indicated that these triple helical assemblies have gene repressive traits, making these structures promising for use in antigene technology for the development of new DNA therapeutics. However, the restriction to polypurine sequences limits the available triplex forming sites in gene targets, as well as their stability. Furthermore, evidence for the ability of triplexes to manipulate gene regulation is primarily based on *in vitro* experiments, leaving numerous aspects to be explored regarding the cellular functions of these structures.

This thesis investigates DNA triplex structures using native mass spectrometry (MS) as well as other solution-phase based techniques. It explores the impact of solution conditions, sequence specificity and base modifications on the formation of these non-canonical structures. The findings contribute insights into the manipulation and potential biotechnological applications, including antigene therapeutics, of higher-order DNA structures.

Following a review of the current literature surrounding the formation and modification of DNA triplexes and methods for their analysis, Chapter 3 focuses on the study of model parallel triplex systems using native MS, UV-vis spectroscopy and isothermal titration calorimetry. Different solution conditions including salt concentrations and pH, as well as ionization modes for MS were investigated and shown to have a profound effect on the formation of DNA triplexes and the species detected respectively. The use of ammonium acetate makes these triplex systems compatible with mass spectrometry (MS) experiments, although care needs to be taken when selecting salt concentrations for *in vitro* studies, as over-stabilisation of triplexes is possible when trying to emulate biological conditions. It is well known that sequence specificity is an important aspect to consider for triplex formation. When probing the specific sequences using both positive and negative ionization modes in MS, this led to

observations of different species in the respective ionization modes. Although there is no change in the relative abundance of triplex formation it is worth noting for the analysis of all triplex systems using MS in the future. The formation of i-motifs was also a sequence-dependent observation detected in low pH conditions in the alternating cytosine and thymine triplex system, which is known to have pH dependence. These structures cannot be identified by solution-phase techniques, highlighting the advantages of analysing higher-order DNA structures using MS.

Triplex target sites are limited in bacterial genomes due to the high chance of having pyrimidine interruptions occurring in the requisite polypurine sequences, which make it increasingly difficult to form DNA triplexes. Chapter 4 investigates the use of dSpacer (2'-deoxyribonucleoside 3'-O-phosphoramidite) and locked nucleic acid (LNA) base modifications to enable triplex formation in such sequences with a pyrimidine interruption. The LNA modification showed enhanced triplex forming capability compared with dspacer and regular DNA bases in the same position. These modifications were also trialled using biologically relevant triplex target sites from a *Pseudomonas aeruginosa* genome. Again, the LNA-based TFOs proved to be the most effective in overcoming pyrimidine interruptions, indicating that synthetic TFOs are a viable approach to stabilise triplex structures for potential gene targeting applications *in vivo*.

Chapter 5 examines the use of a TFO and cell-penetrating peptide conjugate to target specific triplex forming sequences both *in vitro* and *in vivo* as an antigene antibacterial strategy. Using bioinformatics approaches, biologically relevant triplex target sites were identified from a methicillin-resistant *Staphylococcus aureus* genome. The dual-specificity RNA methyltransferase (*rlmN*) gene was chosen as the candidate antigene site due to being of sufficient length without any pyrimidine interruptions and the positioning of the sequence within the promoter region of the gene. The *rlmN* gene is also responsible for the production of the dual-specificity RNA methyltransferase protein and assists with resistance to antibiotics, making it an attractive target for antigene technology. Here, a synthetic TFO-cell penetrating peptide conjugate was constructed as a means of delivering the TFO into bacterial cells. Successful triplex formation was shown using the synthetic conjugate as the TFO by MS and UV-vis spectroscopy techniques. The *rlmN* TFO-cell penetrating peptide conjugate showed moderate antimicrobial activity but was inferior to current antibiotics when tested

against *Staphylococcus aureus*, indicating that more research is required to successfully deliver the synthetic TFO into bacterial cells.

Finally, MS was used to characterise the binding of anthocyanins to DNA triplexes. The interactions between anthocyanins and triplexes have only been previously observed using solution-phase methods and have shown moderate stabilising effects. The binding of Anthocyanins indicates the potential to use these small molecules to aid the stabilisation of DNA triplexes for therapeutic applications. MS experiments indicated that anthocyanins preferentially bind to triplex DNA structures over duplex and single strands. In some cases, two anthocyanin molecules were able to bind to the triplex, once again demonstrating the ability for MS to provide information that cannot be uncovered by techniques such as UV-Vis spectroscopy.

The overarching goal to use DNA triplexes for antigene technology is still in early stages of development. However, furthering the understanding of how sequence specificity, synthetic base modification and small molecules can be manipulated to stabilise DNA triplexes will allow for more biological functions and antigene therapeutics to be revealed.

Chapter 1: Molecular Insights into Nucleic Acid Triplexes: Methodologies and Biological Applications

1.1 Authorship statement

Statement of Authorship

Title of Paper	Molecular Insights into Nucleic Acid Triplexes: Methodologies and Biological Applications
Publication Status	☐ Published ☐ Accepted for Publication ☑ Unpublished and Unsubmitted work written in manuscript style ☐ Submitted for Publication
Publication Details	This manuscript is currently in preparation to be submitted to The EMBO journal

Principal Author

Name of Principal Author (Candidate)	Jack Klose		
Contribution to the Paper	Writing manuscript and presentation		
Overall percentage (%)	75		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	20 Dec 2023

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);
ii. permission is granted for the candidate in include the publication in the thesis; and
iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Gerardo Urbina		
Contribution to the Paper	Writing manuscript (Computational approaches)		
Signature		Date	21 Dec 2023

Name of Co-Author	Blagojce Jovcevski		
Contribution to the Paper	Editing manuscript and presentation		
Signature		Date	21/12/2023

Please cut and paste additional co-author panels here as required.

Name of Co-Author	Tara Pukala		
Contribution to the Paper	Conceptualisation, editing manuscript and presentation		
Signature		Date	20/12/2023

1.2 Review manuscript

Molecular Insights into Nucleic Acid Triplexes: Methodologies and Biological Applications

Jack W. Klose[1], Gerardo Urbina[1], Blagojce Jovcevski[2], Tara L. Pukala[1]*

[1]Discipline of Chemistry, The University of Adelaide, Adelaide, South Australia, 5005, Australia
[2]Discipline of Molecular and Biomedical Science, The University of Adelaide, Adelaide, South Australia, 5005, Australia

Abstract

Triplex DNA structures form through the binding of a third oligonucleotide strand to the major groove of canonical double-stranded DNA at sites of extended polypurine sequence. Although they are known to be favored under certain cellular conditions, including decreased pH and the presence of multivalent cations, there remains ambiguity in the structures and extent to which they form *in vivo*. Therefore, despite their many potential biological applications, the use of DNA triplexes in biotechnology has been limited to date. The focus of this review is to explore the intricacies of DNA triplex formation, as well as the current state of research into their functions and applications in molecular cell biology. The range of analytical, computational and synthetic techniques employed to investigate and enhance the stability of triplex assemblies are also reviewed. Understanding the structural dynamics that underpin triplex formation and activity, coupled with computational and synthetic methodologies to expand their utility, can unlock the potential of various triplexes as a contemporary tool for regulating gene expression.

DNA triplex structure

Deoxyribonucleic acid (DNA) is a hereditary molecule that contains the genetic information vital for organisms to reproduce and function. The x-ray crystallography work of Franklin and Gosling ultimately led to the first proposal of the 3D double-helical DNA structure in 1953 by Watson and Crick (Franklin & Gosling, 1953; Watson & Crick, 1953). The primary DNA nucleic acid sequence is made up of nucleotides which consist of a phosphate group, sugar group and a nitrogen-containing base. The double-stranded structure is connected by hydrogen bonds

between the nitrogenous bases of the nucleotides. Adenine binds in a complementary fashion to thymine via two hydrogen bonds and guanine is complementary with cytosine via three hydrogen bonds, which are known as Watson-Crick base pairs (bp). The two DNA strands bind in an anti-parallel fashion where the 5’ end of one strand is paired with the 3’ end of the complementary strand (where 5’ and 3’ refer to carbons in the sugar moiety, with the 5' carbon attached to a phosphate group and the 3' carbon attached to a hydroxyl group). The duplex DNA structure (Figure 1A) also contains major and minor grooves that arise from the anti-parallel arrangement of the two strands. The outer sections of the bases are exposed in the DNA structure making them available for further hydrogen bonding. The major groove is more accessible to different proteins including transcription factors, which can bind to the exposed bases, allowing them to perform specific functions such as regulating gene expression.

Beyond the well-known double helix, DNA and ribonucleic acids (RNA) are also capable of forming several non-canonical structures, including the triplex structure which forms the focus of this review. The first triple helical structures were biochemically and biophysically characterized by Felsenfeld *et al.* in 1957, where a triple-stranded product of polyribouridylic acid was formed in the presence of Mg^{2+} (Felsenfeld *et al*, 1957). A few years later, the different hydrogen bonding interactions within the triplex structure, now termed Hoogsteen bonds, were characterized by showing that there is a difference between these and the original hydrogen bonding interactions that hold together Watson-Crick bp in double-stranded DNA (dsDNA) (Hoogsteen, 1959).

The two main triplex motifs that have been characterized, the purine (R-type) and pyrimidine (Y-type), are shown in Figure 1B. Both motifs recognize a purine (guanine or adenine) rich sequence, where the triplex-forming oligonucleotide (TFO) resides in the major groove of the duplex. The purine-rich sequence or triplex target site (TTS) is essential for the formation of a triplex as purines contain unoccupied hydrogen bonding sites due to their bicyclic nature which enables interactions with the TFO. The TFO can either comprise the complementary polypyrimidine stretch (cytosine or thymine) which would bind in a parallel (5’-3’) reverse Hoogsteen base pairing manner, or the TFO can contain a polypurine stretch which binds in a Hoogsten bonding anti-parallel manner (3’-5’). These two different forms of triplex DNA are

represented as Y*R·Y (Y-type) and R*R·Y (R-type) respectively, with R representing purines, Y representing pyrimidines, the asterisk denoting Hoogsteen base pairing between the TFO and the purine strand and the middle dot indicating the regular Watson-Crick duplex base pairing (Figure 1B-1C).

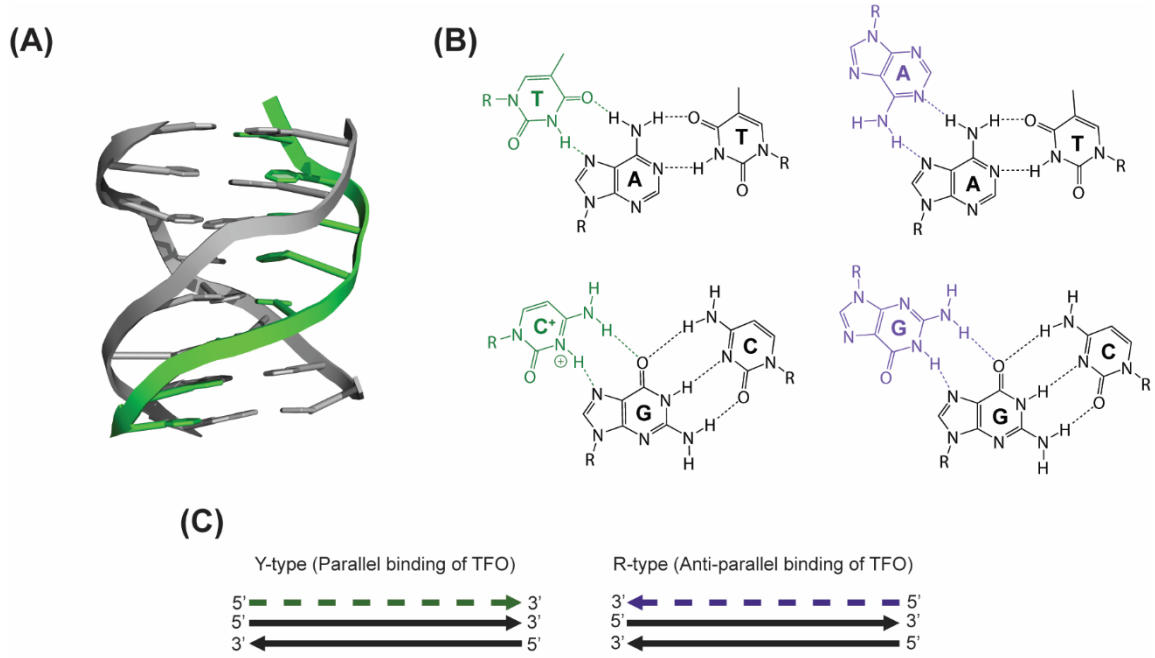


Figure 1: Structure of triplex DNA. (A) Formation of a Y-type (parallel) triplex structure where the TFO (green) binds to the major groove of the duplex DNA structure (grey) (Schrodinger LLC, 2015). **(B)** Chemical structures of the DNA triplex motifs. Watson-Crick bp are shown in black for A-T and G-C. LEFT: Y type (parallel) triplex; TFO bases thymine and cytosine are indicated in green with Hoogsteen bonding shown by green dashed lines. RIGHT: R-type (anti-parallel) triplexes: TFO bases adenine and guanine are indicated in purple with reverse Hoogsteen bonding shown by purple dashed lines. (C) Binding orientation of the TFO to the duplex DNA structure, with Y-type and R-type triplexes indicated on the left (green TFO) and right (purple TFO) respectively.

There is significant debate in the literature regarding the conditions under which R- and Y-type triplex DNA forms. Early studies indicated that DNA triplexes could form under physiologically relevant conditions, which include pH 7.5 tris-EDTA buffer at 37 °C (Hanvey *et al*, 1991). However, this was contradicted when the formation of Y-type triplexes was described as unfavorable under physiological conditions, as they rely on the protonation of the cytosine bases which can only occur at a lower pH to facilitate additional Hoogsteen bonding (Floris *et al*, 1999). More recently, it has been shown that Y-type triplexes can form at acidic (pH 5.5) and physiological pH in ammonium acetate buffer conditions through the

use of mass spectrometry (Li *et al*, 2019). R-type triplexes are considered less thermodynamically stable than the Y-type and rely on adjacent guanine bases for a higher forming propensity (Debin *et al*, 1999). However, some studies have shown that R-type triplexes can be formed with the addition of multivalent cations which are used to reduce the build-up of negative charge, thereby stabilizing electrostatic interactions to allow triplex formation (Floris *et al*, 1999; Debin *et al*, 1999). Increasing the length of the TFO is also known to increase the stability of the triplex due to an increase in the hydrophobic effect, π-π stacking and the addition of more stabilizing Hoogsteen bonds (Faucon *et al*, 1996). Using equilibrium simulations, the Y-type triplex was shown to be more thermodynamically stable than the R-type triplex due to the structure of the R-type triplex being highly distorted (Boehm *et al*, 2018).

Triplexes can form as both intermolecular and intramolecular structures. The first triplex structure that was identified in 1957 was an intermolecular triplex, whereas the first intramolecular triplex was identified in 1985 (Felsenfeld *et al*, 1957; Christophe *et al*, 1985). Also known as H-DNA, the intramolecular triplex structure forms where the third strand is either a component of the duplex or these structures can form with a single DNA strand (Holder *et al*, 2015). For the formation of H-DNA, it is required that the purine and pyrimidine sequences are mirror repeats. This allows for either the pyrimidine or purine containing sequence to fold onto the purine strand of the duplex, enabling the formation of both Y and R type intramolecular triplexes (Lyamichev *et al*, 1986).

Triplexes can also be formed where the TFO is comprised of either DNA or RNA bases. In the case of Y-type RNA triplexes, the canonical T*A·T triplet is replaced with U*A·T, where U denotes uracil. Both types of triplex are thought to play important roles in different cellular processes including transcriptional regulation, gene expression and post-transcriptional RNA processing (Faucon *et al*, 1996; Jain *et al*, 2010). The properties of RNA-containing triplexes have been explored in comparison to DNA variants, where it has been shown that both the extent of triplex formation and thermal stability were higher for the RNA-TFO compared to the DNA counterpart (Li *et al*, 2019). Furthermore, fluorescent microscale thermophoresis has also shown lower K_D values for RNA-TFO triplexes compared to DNA counterparts, indicative of stronger binding to the *En3* gene in mice (Maldonado *et al*, 2018). The RNA triplex is therefore thought to be more stable due to the replacement of thymine for uracil by decreasing the degree of steric hindrance through the loss of a methyl group.

Sequence specificity of triplex DNA

The reason why biologically relevant TTSs are limited is due to the high chance of a pyrimidine interruption within the polypurine stretches of the duplex sequence within genomes. Even a single pyrimidine interruption in the TTS is known to cause instability in the triplex structure, due to the loss of sites for Hoogsteen bonding interactions (Semenyuk *et al*, 2010). There is also a decrease in binding affinity when there is a base in the TFO that doesn't correspond to the regular T*A·T and C*G·C triplets. For example, the replacement of a TFO cytosine in the C*G·C triplet with adenine or guanine has shown decreased binding affinity to the duplex significantly, while a thymine base replacement in the TFO was unable to form any triplex at all (Maldonado *et al*, 2018). This shows the importance of sequence specificity in both the duplex and TFO for the formation of triplexes.

It is also well known that the specific arrangement and composition of bases within a triplex forming sequence can significantly impact their ability to form (Vekhoff *et al*, 2008). Early work into the sequence specificity of triplexes showed that those with an A-tract (four or more A-T nucleotides in a row) were disfavored due to intrinsic properties of the A-tract duplex, which adopts a structure that has highly twisted bp that is stabilized by hydrogen bonds in the major groove, causing the major groove to be wide and shallow (Sandström *et al*, 2002). These A-tract sequences are also comparatively rigid and therefore do not favor the binding of the TFO. In 2003, a systematic study of 15mers triplexes with TTT, TTC, TCC and TCTC as the third strands were analyzed using fluorescence melting curve and temperature jump kinetic profile experiments. At pH 5.0 all four triplexes were formed, with the TTT triplex being the most unstable with a melting point below 30 °C and the TTC, TCC and TCTC sequences having much higher melting points (James *et al*, 2003). The increased stability for triplex sequences containing C*G·C Hoogsteen bond base pairing also arises due to the protonation of the cytosine bases, which assists in overcoming charge repulsion from the negatively charged phosphate on the DNA backbone and favorable interactions that are made from π-stacked bases. In the model study just described, the stability of all four triplexes could be increased by increasing ionic strength or by the addition of Mg^{2+}, with the TTT sequence the most sensitive as the effects of Na and Mg decreased with increasing numbers of the C*G·C Hoogsteen bonding bp (James et al, 2003).

Naturally occurring triplexes have been shown to play roles in the regulation of genes that are linked to diseases in humans. Although there are reports of the potential uses of triplexes for gene expression applications from as early as the late 1980s, there remains an extraordinary gap in knowledge regarding DNA/RNA triplex formation and their functionality mechanisms *in vivo*, due to their poor stability and delivery of TFOs into the nucleus (Cooney *et al*, 1988; Maher *et al*, 1989). In this review, different applications utilizing DNA triplexes will be discussed including the control of gene expression via an antigene strategy. The different chemical, computational, and biochemical methods that have been employed to analyze and stabilize the triplex DNA structure further for these types of applications are also reviewed. Establishing the structural rationale and strategies that govern stable triplex formation can lead to the development of a wide range of therapeutics that target unique DNA sequences to treat diseases and complement gene editing tools.

Methodologies to analyse DNA & RNA triplex structures

For decades, the analysis of triplexes was limited to mainly UV-vis/fluorescence spectroscopy and gel electrophoresis-based experiments. Recently, more advanced biophysical tools have been implemented to investigate the conformation, binding, and biological roles of these non-canonical structures. These techniques primarily include mass spectrometry (MS), nuclear magnetic resonance (NMR) spectroscopy, circular dichroism (CD), isothermal titration calorimetry (ITC), fluorescence resonance energy transfer (FRET) and various computational methods.

UV-vis spectroscopy was one of the first solution-phase techniques used to analyze the stability of duplexes and triplexes. The melting points of triplexes can be detected by an increase in absorbance known as the hyperchromic effect that occurs upon the denaturation of the triplex structure, causing base unstacking that allows the exposed bases to absorb more light (Benight *et al*, 2001). Although this method has been applied for decades, it comes with some limitations including low sample throughput, relatively low changes in absorbances upon melting and the potential for peaks to overlap in the melting curve, making it difficult to resolve signals for mixed species. An alternative approach to examine duplex/triplex formation and stability is the use of FRET. FRET measures the spectral change of fluorescence

from a donor and acceptor fluorescent dye, allowing for spatial information to be determined (Sekar & Periasamy, 2003). Energy is transferred when the fluorescent donor is in close enough proximity to the accepter, which allows the acceptor to become fluorescent. This technique allows the determination of the three-strand configuration and molecular rearrangements of a variety of triplex conformations. A range of well-known features of the triplex structure including for example the pH-dependent formation of Y-type triplexes and stabilization of triplexes by divalent positive ions have been verified using single molecule FRET (Lee *et al*, 2012). A novel pair of pH-stable fluorophores has also been developed on a LightCycler platform that measures the melting point of duplex and triplex using fluorescence intensity. This technique shows significantly less variance in the melting temperature whilst being more cost-effective and having a higher throughput when compared to UV-vis spectroscopy (Schneider *et al*, 2010).

CD spectroscopy is another technique that has been broadly utilised over the years for analyzing triplexes (Gray *et al*, 1995; Szabat *et al*, 2018). CD is an absorption spectroscopy technique that gives information about the chirality of biological systems by measuring the absorbance difference between right and left circularly polarized light (Crabbé, 1965). In DNA systems, CD spectroscopy shows the differences between the sugar backbone and the helical assembly of its constituent strands and is therefore well suited to tracking the conformational transitions between the different nucleic acid arrangements (Vorlíčková *et al*, 2012; Hoyuelos *et al*, 2014). CD spectroscopy remains a relevant technique to characterize triplex structures, especially when coupled with UV-Vis spectroscopy, to probe structural and stability aspects of triplex formation (Szabat *et al*, 2018).

Several strategies have been investigated to overcome the relative instability of DNA triplexes using synthetic methods and even the binding of small molecules, including intercalators. A common way to determine the binding affinity of the TFO or even other molecules to the triplex itself is by ITC. ITC can be used to provide thermodynamic data of binding, and thereby assess how differences in sequences, modified TFOs and intercalators affect triplex formation and stability (Xue *et al*, 2010). ITC only has a few drawbacks which include a relatively large amount of sample needed (which has dropped considerably in recent times with the development of instruments with improved sensitivity), longer running times and signal from

the instrument being proportional to the enthalpy of binding. ITC also provides thermodynamic parameters that strongly complement other biophysical data including UV-vis spectroscopy, CD and fluorescence spectroscopy, which can detect conformational changes and identify binding features (Rangel *et al*, 2022). Moreover, implementing a wide array of biophysical techniques can provide key information about the binding characteristics of a DNA triplex system.

Although able to provide high-resolution structural information, NMR spectroscopy is a less commonly applied method to analyze triplexes due to the distribution of resonances being over a very small range, which can make spectra crowded and hard to interpret. Nevertheless, NMR spectroscopy methods can be a very helpful structural determination tool as triplex formation can be studied by observing chemical shift changes upon binding of the TFO and using nuclear Overhauser effect correlation spectroscopy (NOSEY), total correlation spectroscopy (TOCSY) and correlation spectroscopy (COSY) experiments to identify correlations between neighboring nucleotides. One of the first NMR studies to identify triplex DNA used 2D-NMR spectroscopy experiments to confirm the formation of an intramolecular triplex with a single 28mer strand (Sklenář & Felgon, 1990). A more recent study showed the use of in-cell NMR techniques to detect Y- and R-type triplex structures *in vivo* using human cells, demonstrating a very interesting application for potentially assessing the success of forming triplexes at specific target genes in cells (Sakamoto *et al*, 2021). Similarly, X-ray crystallography is a robust structural characterization technique that exposes samples to an X-ray beam, with the resulting diffraction pattern being used to determine a 3D structure. However, to date this method is relatively limited in its application to triplex structure determination, with only four DNA containing triplex structures characterized (Dalla Pozza *et al*, 2022).

Native mass spectrometry (MS) is a promising analytical technique for the study of triplex structures, enabled through preservation of the native-like state of biomolecules from the solution phase through the mass spectrometer. While native MS has been widely applied to proteins and protein complexes, there are limited examples of triplexes being analyzed via MS compared to other techniques, with the first observation reported by MS in 2002 for a 14mer triplex under electrospray ionization conditions (Rosu *et al*, 2002). An advantage of MS

is the ability to adapt the conditions in the spectrometer to enable further structural interrogation. For example, increasing the energy of ions in the instrument, such as in collision-induced dissociation experiments, causes the TFO to be ejected from the duplex structure and allows for the identification of the individual components by MS (Rosu *et al*, 2002). These types of dissociation experiments are analogous to solution-phase experiments like UV-vis spectroscopy, as they can inform on the stability of the triplex. Ion mobility-MS (IM-MS) is a relatively new mass spectrometry-based technique and has been recently implemented to further characterize triple helical assemblies. From this, evidence suggests triplexes can maintain their solution-phase structures in the gas phase, confirming that MS analysis is applicable in a cellular context (Arcella *et al*, 2012). MS can also be used as a high-throughput screening tool to identify compounds that bind to triplexes. The triplex binding capabilities of five different compounds (epiberberine, coptisine, jatrorrhizine, berberrubine and colubamine) from *Rhizoma coptidis* extracts and *phellodendron chinense* schneid cortexes were recently assessed using UHPLC-orbitrap MS with good reproducibility (RSD <10%) (Yang *et al*, 2017). This was the first study to develop a method for the high-throughput screening of triplex DNA binders from complex samples, which can potentially be used to improve the screening process of new drug candidates. These examples show the vast array of different instrumentation and experiments that can be performed using MS. MS is an underused technique, that also can report a range of assemblies in a mixture that is missed by many techniques. We feel that the field would benefit from more use of MS for analysis of different triplex systems.

Computational approaches to examine nucleic acid triplexes

Various methodologies have been utilized, and continue to be developed, to enhance our understanding of nucleic acid triplexes. Consequently, it is unsurprising that, besides experimental approaches, a great deal of computational studies have been dedicated to nucleic acid triplexes. These theoretical investigations can be broadly categorized into two groups: i) the molecular modelling of triplexes and triplex formation and ii) the development of bioinformatics tools for predicting the triplex-forming behavior of genomic DNA and RNA sequences. To the best of our knowledge, no recent review articles have focused on the molecular modelling of triplexes, even though many such studies have helped elucidate

essential details about their structure and formation and have informed many experimental studies. On the other hand, bioinformatics tools in the context of triplexes have become particularly popular in the field of long non-coding RNA (lncRNA), as described thoroughly by a recent review (Warwick *et al*, 2023). Thus, in this section, we present a short review of the use of bioinformatics tools in the study of nucleic acid triplexes, followed by an introduction to some of the trends in the field of the molecular modelling of DNA and RNA triplexes.

Bioinformatic approaches

With the advent of human genome sequencing at the beginning of this century, it became possible to use computational tools to search for potential triplex target sequences (TTS) (Lander *et al*, 2001). The first of such studies revealed that TTS are more abundant in the human genome than what is expected from random sequences, particularly in promoter regions (Goñi *et al*, 2004). Following this, bioinformatics tools became an important part of the study of biological and therapeutic applications of triplexes, as demonstrated by the subsequent development of TTS databases and TTS search algorithms (Jenjaroenpun & Kuznetsov, 2009; Lexa *et al*, 2011). In 2012, Buske *et al.* published Triplexator (short for triple-helix locator), a computational toolkit capable of identifying TTSs in double-stranded nucleic acids, potential TFOs in single-stranded nucleic acids and their compatibility, allowing for incompatibilities depending on the lengths of the TTSs and TFOs (Buske *et al*, 2012). While Triplexator has remained a popular tool, many other algorithms and toolkits have been developed to search and assess putative TTSs and TFOs under different constraints and using various criteria. Warwick et al. recently described many of these methods in a detailed review focused on their use in the study of triplex formation between dsDNA and lncRNA, although these technologies can be used outside of this scope in, for example, the identification of oncogenes and the existence of triplex DNA repeats in the bacterial genome (Warwick *et al*, 2023; Papadakis & Gizeli, 2012; Holder *et al*, 2015). The development and refinement of bioinformatics tools have revolutionized the exploration and prediction of TTSs and TFOs in diverse biological contexts. The current trends regarding lncRNA research are of particular importance, and future bioinformatic work holds promise for further enhancing these tools, expanding their applications outside of the human genome and for other types of TFOs.

Molecular dynamics and quantum mechanical modelling approaches

Over the past three decades, various techniques involving molecular dynamics (MD) simulations and *ab initio* quantum mechanical (QM) simulations of nucleic acid triplexes have been employed to aid in characterizing triplexes and guiding the development of novel TFOs. The first MD simulation of a DNA triplex structure was published in 1992 by Laughton and Neidle, which provided structural details that were otherwise unavailable and helped support contemporary findings about its backbone conformation obtained from NMR spectroscopy studies (Laughton & Neidle, 1992). This also demonstrated a new way to study triplex structures before the first X-ray structure elucidation took place. Following these results, the use of MD simulations together with NMR spectroscopy studies to obtain structural data of triplexes became widespread almost immediately, as illustrated by its mention in a 1993 review. This review described how early MD studies revealed more detailed information compared to fiber-diffraction data, which was the only other source of structural information available at the time (Sun & Helene, 1993; Arnott & Selsing, 1974). Several other studies involving NMR-restricted MD simulations were also carried out during the same decade, examples of which include the study of intramolecular parallel and antiparallel triplexes, the use of thymine in the TFO to target cytosine in the target strand via parallel triplex formation, and the effect of substituting natural bases with intercalators and artificial bases in the TFO (Wang *et al*, 1996; Koshlap *et al*, 1997; Radhakrishnan & Patel, 1993, 1994; Ji *et al*, 1996). Some of the structures elucidated using this method are among the earliest triplex structures in the Protein Data Bank (Berman *et al*, 2000).

Many other entirely computational studies were also reported before the end of the 20th century, providing structural information on backbone conformation, hydrogen bonds, and the thermodynamic properties of nucleic acid triplexes. Examples of such studies include the comparison of triplexes where both parallel and antiparallel orientations are possible, the effect of ions on the triplex structure, and the nature of the protonation of cytosine in TFOs (Soliva *et al*, 1998; Cheng & Pettitt, 1992; Mohan *et al*, 1993b). From these studies, essential characteristics of the triplex structure was discovered including the preference for a B-form-like structure, and that not all the cytosines in a polypyrimidine TFO are protonated when a triplex forms. With the turn of the new century, computational studies aiming to elucidate structural knowledge of canonical triplexes became scarcer but more specialized. The topics

of theoretical studies in this regard from the past twenty years include improved research on the interactions between triplexes and their environment (water molecules and ions), prediction of the optical properties of triplexes, the effect of different sequences on triplex structures, and the dynamic structure of triplex-based technologies such as triplex nanoswitches (Iacovelli *et al*, 2019; Ojha & Tiwari, 2003; Muñoz *et al*, 2002; Zhang *et al*, 2017; Ghane *et al*, 2012; Thenmalarchelvi & Yathindra, 2005; Rathinavelan & Yathindra, 2006). On the other hand, newer experimental techniques have also been used to obtain new information about triplex structures. These include the use of IM-MS together with MD simulations to characterize the assembly and conformation of triplexes in the gas phase, and most recently, the integration of wide-angle X-ray scattering (WAXS) with MD for structure elucidation (Arcella *et al*, 2012; Li *et al*, 2019; Chen *et al*, 2022).

Besides their use in structural studies, computational techniques have also been used in the development of novel TFOs. As mentioned earlier in the context of NMR-restricted MD simulations, computational techniques have been used to study the structure of triplexes that contain artificial bases in TFOs. Interestingly, the first theoretical study including artificial bases was carried out by Hausheer *et al.* less than half a year after the first MD simulation of a triplex was published (Hausheer *et al*, 1992). The strategy followed in this study (and in many subsequent studies) was to carry out an *ab initio* QM simulation of an artificial base (5-methylcytosine), use this to calculate its structural parameters and atomic charge, and incorporate these parameters into an MD simulation of the triplex. Similar theoretical approaches have been used over the years to study the structure, hydrogen bonding, and thermodynamics of triplexes formed using base modified TFOs. These include the design of potential bases to target pyrimidine interruptions, cytosine derivatives that can improve triplex stability or remove the need for protonation in Y*R·Y type triplexes, and guanine derivatives in antiparallel triplexes (Spacková *et al*, 2004; Faustino *et al*, 2014; Mohan *et al*, 1993a; Nyerges *et al*, 1995; Yang *et al*, 1999). Furthermore, similar theoretical approaches that rely on combinations of *ab initio* QM and MD simulations have been used to study TFOs containing artificial or modified backbones. Examples include the theoretical study of protein-nucleic acid-TFO (PNA-TFOs), locked nucleic acid-TFO (LNA-TFOs), nucleic acids where the phosphodiester has been replaced with guanidinium linkers, and 2’,5’ nucleic acids (Premraj

et al, 2002; Almarsson & Bruice, 1993; Mall *et al*, 2018; Hartono *et al*, 2018; Pant & Fisher, 2020; Luo & Bruice, 1998).

Computational methods have also been used to obtain information about triplex formation (such as hydrogen-bonding patterns, π–π interactions, thermodynamic data, and structural information, etc.) used to support experimental results when base-modified or backbone-modified TFOs were synthesized, and their triplex-forming ability was assessed. Studies in which computational methods have been used to support experimental results include the syntheses of potential cytosine substitutes for enhanced parallel triplex formation, a guanine substitute for antiparallel triplexes, TFOs with internucleoside linkers for alternate targeting of both DNA strands, LNA-TFOs, PNA-TFOs, and novel bases used for pyrimidine targets (Bates *et al*, 1996; Soliva *et al*, 2000; Giancola *et al*, 1999; Sun *et al*, 2004; Pabon-Martinez *et al*, 2017; Petraccone *et al*, 2004). The most recent studies that combine the use of computational and experimental methods for the development of novel TFOs are a series of studies by the same research group regarding PNAs with artificial bases for the recognition of both purine and pyrimidine targets in double-stranded RNA, and the design, synthesis and *in silico* assessment of extended nucleobases for PNA-TFOs for the recognition of any DNA bp by Alavijeh *et al* (Kotikam *et al*, 2019; Kumpina *et al*, 2019; Ryan *et al*, 2022; Talbott *et al*, 2023; Alavijeh *et al*, 2023).

A field where computational methods have yet to be used widely is the binding of small molecules to triplex DNA. Docking simulations have been used to characterize the binding of neomycin to several nucleic acid structures, including triplexes, and to compare the binding ability of a library of small molecules towards a lncRNA triplex, while MD simulations were used to determine structural and thermodynamic information regarding triplex destabilization by netropsin (Xi *et al*, 2011; Swain *et al*, 2021; Boehm *et al*, 2018). A few recent studies also illustrate the potential of computational techniques in the study of biological applications of triplexes. MD simulations have been employed to study the thermodynamics of microRNA triplexes in the context of cooperative gene regulation, how triplexes may interfere with protein binding in supercoiled DNA, and the structural properties of telomeric DNA triplexes (Schmitz *et al*, 2014; Noy *et al*, 2017; Sieradzan *et al*, 2017). The use of MD

simulations to study the structure and thermodynamics of triplexes associated with FRDA has also been of particular interest in recent times (Zhang *et al*, 2021, 2022).

These studies demonstrate the importance and potential found in an approach toward the understanding of nucleic acid triplexes that combines experiments and computer simulations. As computers become more advanced and computational costs decrease, it will be possible to obtain more and better information from MD and QM simulations. To make the most out of it, we believe that a deeper review of the computational data obtained so far, together with further exploration using the latest techniques available, constantly informed by experimental results and *vice versa*, can significantly accelerate progress in the field of nucleic acid triplexes.

Biological and therapeutic applications of triplexes

Naturally occurring H-DNA is reported to be abundant in mammalian cells (approximately 1 H-DNA site in every 50,000 bp) and is inherently mutagenic (Schroth & Ho, 1995). It is suggested that these structures also contribute to genetic evolution, DNA metabolism and gene functionality. There are several proposed mechanisms for the interference of transcription by H-DNA, whereby transcription causes opening of DNA, allowing the non-template strand to act as the third strand for triplex formation upstream of the transcription site (Belotserkovskii *et al*, 2007). It has been suggested that the transcription produced H-DNA can trap RNA polymerase within the H-DNA forming sequence (Grabczyk & Fishman, 1995). However, it has also been inferred that this mechanism actually occurs further downstream, and the RNA polymerase is constrained sterically by bends introduced into the sequence by the H-DNA, indicating that the exact mechanism is still unknown (Belotserkovskii *et al*, 2007).

Naturally occurring DNA triplex formation is implicated in a single human disease to date, namely Friedreich's ataxia (FRDA). FRDA is an autosomal neurodegenerative disease, that is caused by repression of the *FXN* gene, where 98% of cases show an expansive GAA trinucleotide repeat (from 100-1700) in intron 1 (Campuzano *et al*, 1996). Since the expansive repeat is comprised of purine bases, this stretch can act as a TTS and the formation of triplexes at these sites is established to be the cause of *FXN* gene repression (Sakamoto *et al*, 1999). There are two proposed TFO mechanisms for the formation of the FRDA related triplex; the

expanded GAA repeat region folding back on itself forming an intramolecular R type triplex, and the formation of a triplex complex where the TFO is comprised of the nascent RNA formed by RNA polymerase prior to gene expression being inhibited (Rajeswari, 2012). The decrease in gene expression results in the inhibition of Frataxin production in the mitochondria of neuron cells, which is critical in the assembly of iron-sulphur clusters required for electron transport in respiratory chain complexes and DNA repair. FRDA gives rise to increased iron levels in the mitochondria, and ultimately leads to loss of motor coordination, dysarthria, muscle weakness, heart palpitations and shortness of breath (Santos *et al*, 2010).

lncRNA is known to play a significant role in many biological processes including regulation of gene expression, cell development and metabolism. A modicum of lncRNAs have been comprehensively characterized, and their interactions with DNA have not been studied extensively. Research on non-coding RNA triplexes seemed to have gained some traction in the 2010s and was covered comprehensively in a review by *Li et.al.* (Li *et al*, 2016). It has been shown that lncRNAs can form triplexes, for example when an inhibiting RNA sequence was transcribed from the minor promoter region of the dihydrofolate reductase (*DHFR*) gene (Martianov *et al*, 2007). This transcript was able to form a triplex and repress *DHFR* at its major promoter region further downstream by containing overlapping sequences. In a more recent study, expression of a specific cyclin-dependent kinase inhibitor 2B antisense non-coding RNA (CDKN2B-AS1) has been used to target the promoter region of *CDNKN2B* in the progression of atherosclerosis. Expression of CDKN2B-AS1 lncRNA showed increased uptake and intracellular accumulation of lipids in atherosclerosis plaques due to the formation of RNA triplex, suggesting a novel target for the treatment of atherosclerosis (Ou *et al*, 2020). These studies highlight the importance of targeting minor promoter regions for gene targets and the resultant impact on transcriptional regulation mechanisms.

A chromatin-interacting lncRNA (*MEG3)* has also been shown to regulate genes in the TGF-β pathway by the formation of guanine and adenine-rich RNA-DNA triplex structures *in vivo* (Mondal *et al*, 2015). Using immunostaining in BT-549 cells, triplex structures were identified in the nucleus and cytoplasm using an anti-triplex antibody. Then by performing a triplex-chIP assay, it was determined that the *MEG3* lncRNA regulates the *TGFBR1, TGFB2 and SMAD2* genes through the formation of triplexes (Figure 2) (Mondal *et al*, 2015). These results are

significant as specific antibodies were used to detect the presence of lncRNA triplexes in breast cancer epithelial cells. This technique is also of interest as it could be applied in the antigene strategy, by designing a DNA triplex-specific antibody, to see if DNA triplexes are forming in the nucleus and regulating gene expression.

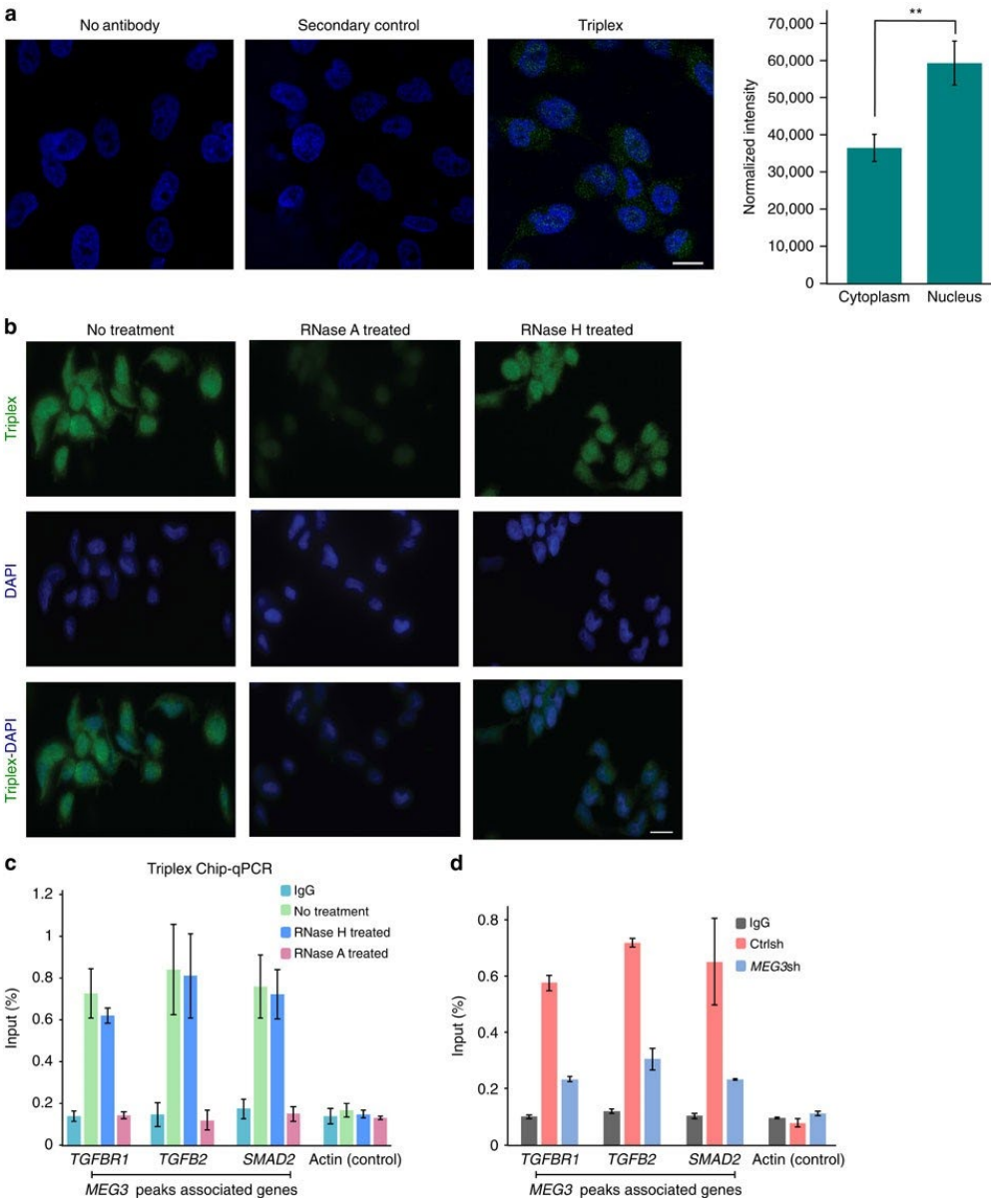


Figure 2: (**a**) Confocal microscopic images showing immunostaining with anti-triplex dA.2rU antibody (green) and nucleus (DAPI, blue) in BT-549 cells. (scale bar, 5 µm). Quantification of the triplex signal in cytoplasm and nuclear compartments obtained from the three-dimensional confocal images (far right panel). Data presented as mean ± SEM (**P<0.01). (**b**) RNA–DNA triplex structures are sensitive to RNase A but are resistant to RNase H in vivo. (scale bar, 5 µm). (**c**) Triplex-ChIP–qPCR showing enrichment (presented as percentage of input) of triplex structures over the MEG3 peaks associated with the TGF-β pathway genes (TGFBR1, TGFB2 and SMAD2) in BT-549 cells (mean ± SD, n=3) (**d**) Triplex-ChIP–qPCR showing enrichment (presented as percentage of input) of triplex structures over the MEG3 peaks associated with the TGF-β pathway genes (TGFBR1, TGFB2 and SMAD2) in Ctrlsh and MEG3sh BT-549 cells (mean ± SD, n=3). Figure reprinted with permission under creative commons license https://creativecommons.org/licenses/by/4.0/ (Mondal et al, 2015).

It Is well known that microRNAs (miRNAs) play a vital role In regulating gene expression by binding to the 3' untranslated region (UTR) of the target sequence and down-regulating translation by causing mRNA instability (Fang & Rajewsky, 2011). It is also known that these

small miRNA strands can upregulate gene expression when bound to other regions, such as the 5' UTR, coding sequence and promoter regions (Valinezhad Orang *et al*, 2014). A proposed mechanism by which this occurs is through miRNAs acting as TFOs, binding to the major groove of the DNA duplex, and therefore interacting with specific sequences of DNA that are in regulatory and promoter regions. Formation of triplexes has been observed using miRNA (both containing pyrimidines and purines) as the TFO at specific target DNA sequences (Paugh *et al*, 2016). These sequences were identified using an algorithm (Trident) that allows genome-wide searching for potential miRNA TTS. It was concluded that genes containing highly expressed miRNAs can form triplex structures and increase mRNA expression levels in primary leukemia cells, which correlated to an increase in gene expression according to computational analysis (Paugh *et al*, 2016). This interesting study demonstrates the ability of triplexes to be used to promote gene expression, whereas the majority of research in this field describes triplexes as gene repressors. The mechanisms that underlie the upregulation of gene expression via miRNAs are largely not understood, but it has been proposed that miRNAs could prevent binding of transcriptional suppressors, thereby enhancing transcription (Paugh *et al*, 2016).

On the other hand, a study from 2010 showed that the use of triplex-forming miRNAs was able to form stable complexes and inhibit HIV-1 replication (Kanak *et al*, 2010). Mature miRNAs that had substantial homologies to the HIV-1 genome (via sequence alignment with the 3'UTR) were identified. The HeLa-CD4+ and CD4+ T-lymphocytes cell lines showed significant resistance to HIV-1 and a direct correlation was found between the considerable amounts of specific miRNAs and the ability to inhibit replication. This shows that the triplex-forming miRNAs are fully functional, and it was also found that the HIV-1 antivirals had activity that lasted for at least 40 days post-infection compared to the non-triplex-forming miRNA control (Kanak *et al*, 2010). This not only shows the huge potential for miRNAs to be able to form stable complexes with DNA duplexes in *vivo* but also for triplexes in general to have the potential for developing specific antivirals and other classes of antimicrobials.

Triplex applications in biotechnology

A major application of triplexes in biotechnology is in an antigene strategy, whereby a TFO is introduced to target a specific TTS by formation of an intermolecular triplex. The formed

triplex is intended to result in the prevention of transcription factor binding to the specific gene regulatory element (Alunni-Fabbroni *et al*, 1994). These factors contain DNA-binding domains that allow them to attach to promoter regions (when bound near the transcription start site) and enhancer regions (other regulatory sequences that can alter gene expression) to repress gene expression through inhibition of transcription factor binding (Sakamoto *et al*, 1999). Such an approach could be utilized in an antigene strategy by introducing a TFO to downregulate the expression of specific genes (proposed mechanism shown in Figure 3). This method is one of the most promising applications in the triplex DNA field, and the optimization of this approach would allow TFOs to specifically alter gene expression using DNA-based molecules.

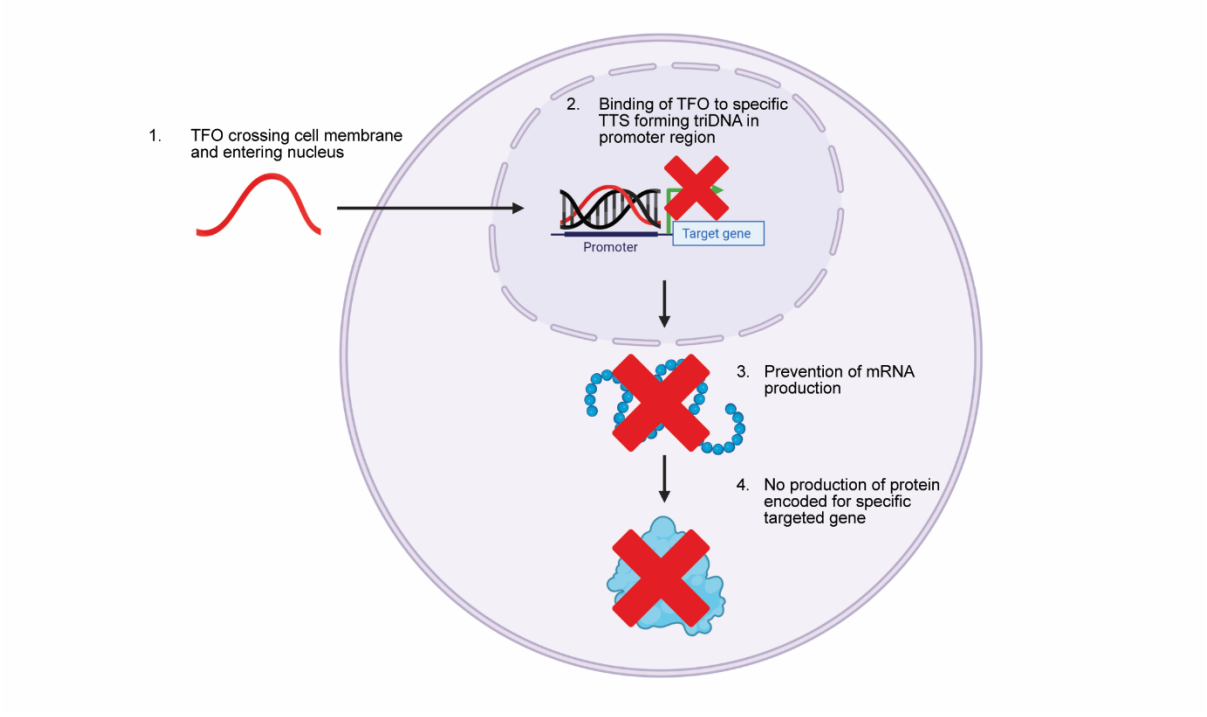


Figure 3: Proposed mechanism of an antigene therapeutic strategy to modulate gene expression. The introduction of the TFO into the cell and the nucleus would allow for the formation of a triplex at a specific gene target TTS. The formation of triplexes in enhancer regions of the gene is postulated to prevent the binding transcription factors, causing the targeted gene to be repressed. Created using biorender.com

Gene expression using TFOs to target cancer cells is of particular interest. For example, TFOs can be designed to specifically inhibit the expression of High-mobility group A1 (*hmga1)* inducing apoptosis in human cervical cancer cells (Akhter & Rajeswari, 2017). The *hmga1* gene regulates the transcription of various genes and is considered a biomarker for tumors, making

it a suitable target for a potential antigene strategy approach. Two triplex sequences with different duplexes and TFOs were selected based on their positioning within the promoter region of *hmga1*. Both sequences were successful in downregulating *hmga1* expression at an mRNA and protein level causing apoptosis in HeLa cells (Akhter & Rajeswari, 2017). The inhibition of *hmga1* is a perfect example whereby a specific gene of interest was selected and gene expression was able to be manipulated. The binding characteristics of two antibiotics that have been repurposed as anti-cancer drugs (adriamycin and actinomycin-D) to a DNA triplex sequence related to the high mobility group box 1 (*hmgb1*) gene were shown to be stronger than the regular duplex DNA by ITC (Lohani & Rajeswari, 2016). This is the first study that targeted triplex formation at the *hmgb1* gene site, which is a known site for tumorigenesis in numerous human cancers (Kang *et al*, 2013). The binding of anticancer drugs to triplexes is an appealing concept that has the potential to help the efficacy of cancer therapeutic agents.

A recent publication has also demonstrated the formation of a stable DNA/RNA triplex *in vivo*. A small RNA-TFO was introduced into human k562 (myelogenous leukemia) cells and by using gel-shift, R-loop elimination and pull-down assays, along with co-immunoprecipitation, it was determined that the formation of the triplex causes displacement of major transcription factors (USF2, GATA1 and TAL1) (Zhou *et al*, 2019). The decrease in the binding levels of these transcription factors also caused the reduction in RNA polymerase II to hypersensitive site 2 (HS2) as they are required in the recruitment mechanism (Zhou *et al*, 2019, 2010). In addition, the study also demonstrated the reduction of binding of transcription factors (Pol II, USF2 and GATA1) that normally bind to human ε- and γ-globin promoters which are activated by HS2. This resulted in a reduction of expression of these genes and suggests a model in which DNA/RNA triplexes can be utilized in gene regulation for anticancer applications.

The sequence specificity of the triplex can also influence the pH-dependency of formation. An interesting application of this is the development of a programmable DNA-based nanoswitch that is triggered over different pH windows (Idili *et al*, 2014). The switch is designed whereby an intramolecular triplex can form when using both the C*G·C and T*A·T base triplets under basic to acidic conditions (Figure 4). The T*A·T triplex is stable until the pH rises and deprotonation of the thymine bases (pK_a ~10) occurs (Idili *et al*, 2014). A biocompatible stimulus-response DNA drug delivery system for cancer was developed using such a pH-sensitive DNA nanoswitch in 2019 (Chen *et al*, 2019). When exposed to acidic conditions the

delivered linear DNA undergoes a conformational change to the triplex structure, prompting the release of anticancer drugs doxorubicin and cisplatin. This technique demonstrated effective gene repression and significant inhibition of tumor growth in a mouse model (Chen *et al*, 2019). These applications further highlight the potential to use the dynamic triplex structure as a drug delivery system.

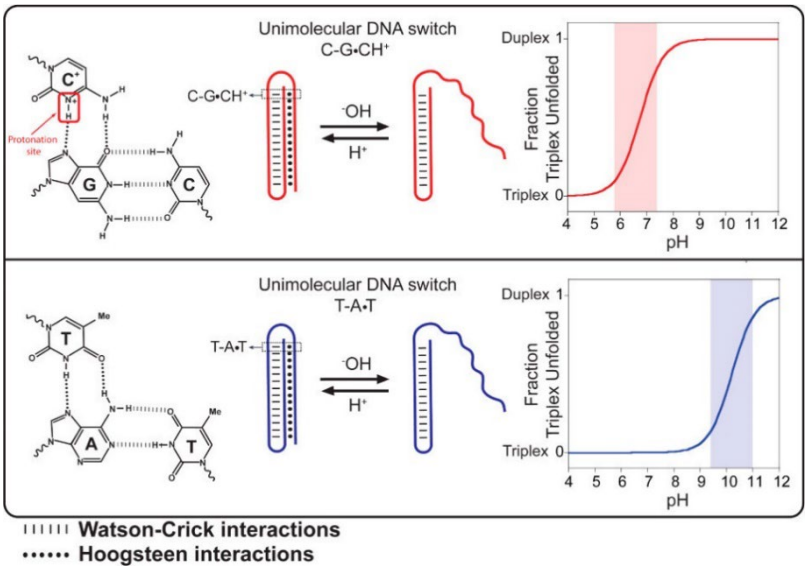


Figure 4: Programmable DNA triplex pH nanoswitch that forms a triplex when exposed to more acidic conditions. The C*G·C triplet forms a triplex due to the protonation of cytosine bases at lower pH (pK_a 6.5) and the T*A·T triplet unfolds at a more basic pH due to the deprotonation of thymine at pK_a 10. Figure reprinted with permission from (Idili A, Vallée-Bélisle A & Ricci F (2014) Programmable pH-Triggered DNA Nanoswitches. J Am Chem Soc 136: 5836–5839).

A final example of triplex biotechnology is provided in the context of the recent COVID-19 pandemic, which redirected research efforts towards vaccine development and detection of the SARS-CoV-2 virus. A strategy called Triplex enhanced nucleic acid detection assay (TENADA) was developed to detect the viral RNA of SARS-CoV-2 using triplex-forming hairpins (Aviñó *et al*, 2022). The triplex-forming hairpins, which are known as polypurine reverse Hoogsteen hairpins (PPRH), were formed by two antiparallel polypurine mirror repeat sequences that are bound to a thymidine loop by reverse Hoogsteen bonds. These hairpins are then able to bind to polypyrimidine sequences in the viral RNA using the canonical Watson-Crick base pairing and are known to have gene-repressive traits (Noé *et al*, 2021). The PPRHs were designed by determining the longest triplex forming sites in the SARS-CoV-2 genome, specifically for the detection of the viral RNA associated with the virus. Using the PPRHs, the viral RNA was then detected in nasopharyngeal samples, with a sample analysis

time of less than an hour and a limit of detection of around 0.01 nM (Aviñó *et al*, 2022). It has also been postulated that TENADA can be adapted for the detection of other pathogens that contain RNA or DNA genomes, making it a very promising biomedical application for the detection of infectious diseases.

Triplex-protein interactions

One of the areas that is significantly underexplored in the literature is triplex binding proteins for both DNA/RNA triplexes, as well as their influence on triplex formation and cellular functions in *vivo*. DNA-binding proteins make up a large proportion of all cellular proteins and have very important roles including DNA repair, splicing, apoptosis, regulating stress responses and manipulating transcription and translation (Hudson & Ortland, 2014). Understanding triplex-protein binding interactions will be very useful to determine the role of triplexes *in vivo* and it is of interest to see how these differ from duplex DNA-binding protein interactions.

Proteins from HeLa cells were shown to bind to both parallel and anti-parallel DNA triplexes by the late 1990s (Guieysse *et al*, 1997; Musso *et al*, 1998). It was also found that these triplex binding proteins exhibited increased binding affinity when the length of the triplex was increased from 12 to 19 bp. Several other triplex DNA binding proteins have been characterized over the years using electrophoretic mobility shift assays and western blotting techniques *in vitro*, which include maltose binding protein (MBP-LOR3^{ARF}) and type III intermediate filament proteins (Ciotti *et al*, 2001; Li *et al*, 2004). In more recent studies, three triplex DNA binding proteins have been identified in human colorectal tumors (Nelson *et al*, 2012). U2 small nuclear RNA auxiliary factor 2 isoform b (U2AF65) was correlated with increased activity in tumor extracts (Nelson *et al*, 2012). This protein along with the other two triplex DNA binding proteins that were found were polypyrimidine tract binding-associated splicing factor (PSF) and nuclear RNA binding protein (p54nrb), all have RNA polymerase II-associated splicing roles that can regulate the expression of transcription factors (including the myc family) and the cell cycle. The increase in triplex DNA binding activity correlates to lymph node disease and metastasis. It is apparent that triplex binding proteins have applications in cancer, however their mechanisms and specificity of binding, and how they

affect the structure of the triplex are largely unknown. The binding of p53 (a tumour-suppressing protein) showed specific binding to both intermolecular and intramolecular Y-type triplexes containing only T*A·T triplets (Brázdová *et al*, 2016). This was the first DNA binding study indicating binding of a full length p53 to a triplex. This shows that the T*A·T forming sequence can potentially be used as regulatory enhancer elements and suggests that these regions could be targeted to enhance the stability of the genome and control DNA repair, replication, recombination, and gene expression mechanisms.

One of the main issues with forming triplexes *in vivo* is the effect that common enzymes such as helicases have on formation (Jain *et al*, 2010). It was found that the *ChlR1* helicase had unique binding to intramolecular and intermolecular DNA triplexes through the use of helicase assays, RNAi and immunofluorescence staining (Guo *et al*, 2015). The binding of this helicase was also the cause of the unfolding of the triplex structures in HEK293T and HeLa cells. It was also very interesting that this activity was only specific to the *ChlR1* helicase as another helicase of the same family *FANCJ* was also tested with significantly less activity. It was concluded that these triplex-specific helicases have roles in maintaining and defending genome integrity by preventing the formation of triplex DNA (Guo *et al*, 2015). Further investigations into the interactions between triplexes and proteins using other analytical techniques such as native MS and structural proteomics would be beneficial to understanding their impact on biological functions.

Conclusions

The ability of DNA/RNA triplex assemblies to form under physiological conditions and influence gene regulation indicates their significant potential to be employed in a highly specific antigene strategy. Numerous methods have been used over the years to study these triple helical assemblies, however, there still is some ambiguity in the literature about how these structures form and their biological potential. Furthermore, there are still issues associated with poor stability of triplexes formed from gene target sequences, which include the inherently small TTSs and the need to deliver them to *in vivo* systems. Expansion in the field of synthetic DNA backbones and bases is a promising avenue for overcoming these issues. The development of synthetic oligonucleotides will potentially prevent their degradation due to nucleases and increase the binding propensity of the TFO to its specific

gene target *in vivo*, which would make triplexes a viable option for applications that involve the regulation of gene expression. The further development of computational methods as well as stabilizing synthetic modifications would allow for more optimal TFOs to be designed specifically for binding of triplexes in biological systems. Further exploration of the various mechanisms underlying the antigene strategy and other gene regulation-related applications for triplexes is crucial for unlocking potential applications in the development of clinical therapeutics.

References

Akhter Z & Rajeswari M (2017) Triplex forming oligonucleotides targeted to hmga1 selectively inhibit its expression and induce apoptosis in human cervical cancer. *J Biomol Struct Dyn* 35: 689–703

Alavijeh NS, Serrano A, Peters MS, Wölper C & Schrader T (2023) Design and Synthesis of Artificial Nucleobases for Sequence-Selective DNA Recognition within the Major Groove. *Chem Asian J* 18: 9

Almarsson O & Bruice TC (1993) Peptide nucleic acid (PNA) conformation and polymorphism in PNA-DNA and PNA-RNA hybrids. *Proc Natl Acad Sci* 90: 9542–9546

Alunni-Fabbroni M, Manfioletti G, Manzini G & Xodo LE (1994) Inhibition of T7 RNA Polymerase Transcription by Phosphate and Phosphorothioate Triplex-Forming Oligonucleotides Targeted to a R · Y Site Downstream from the Promoter. *Eur J Biochem* 226: 831–839

Arcella A, Portella G, Ruiz M, Eritja R, Vilaseca M, Gabelica V & Orozco M (2012) Structure of Triplex DNA in the gas phase. *J Am Chem Soc* 134: 6596–6606

Arnott S & Selsing E (1974) Structures for the polynucleotide complexes poly(dA) with poly (dT) and poly(dT) with poly(dA) with poly (dT). *J Mol Biol* 88: 509–521

Aviñó A, Cuestas-Ayllón C, Gutiérrez-Capitán M, Vilaplana L, Grazu V, Noé V, Balada E, Baldi A, Félix AJ, Aubets E, *et al* (2022) Detection of SARS-CoV-2 virus by triplex enhanced nucleic acid detection assay (TENADA). *Int J Mol Sci* 23: 15258

Bates PJ, Laughton CA, Jenkins TC, Capaldi DC, Roselt PD, Reese CB & Neidle S (1996)

Efficient triple helix formation by oligodeoxyribonucleotides containing alpha- or beta-2-amino-5-(2-deoxy-D-ribofuranosyl)pyridine residues. *Nucleic Acids Res* 24: 4176–4184

Belotserkovskii BP, De Silva E, Tornaletti S, Wang G, Vasquez KM & Hanawalt PC (2007) A Triplex-forming Sequence from the Human c-MYC Promoter Interferes with DNA Transcription *. *J Biol Chem* 282: 32433–32441

Benight AS, Pančoška P, Owczarzy R, Vallone PM, Nešetřil J & Riccelli PVBT-M in E (2001) Calculating sequence-dependent melting stability of duplex DNA oligomers and multiplex sequence analysis by graphs. In *Drug-Nucleic Acid Interactions* pp 165–192. Academic Press

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN & Bourne PE (2000) The Protein Data Bank. *Nucleic Acids Res* 28: 235–242

Boehm B, Whidborne C, Button A, Pukala T & Huang D (2018) DNA triplex structure, thermodynamics and destabilisation: insight from molecular simulations. *Phys Chem Chem Phys*: 14013–14023

Brázdová M, Tichy V, Helma R, Bazantova P, Polaskova A, Krejci A, Petr M, Navratilova L, Ticha O, Nejedly K, *et al* (2016) p53 specifically binds triplex DNA in vitro and in cells. *PLoS One* 11: 1–25

Buske F, Bauer D, Mattick J & Bailey T (2012) Triplexator: Detecting nucleic acid triple helices in genomic and transcriptomic data. *Genome Res* 22: 1372–1381

Campuzano V, Montermini L, Moltò MD, Pianese L, Cavalcanti F, Monros E, Rodius F, Duclos F, Zara F, Cañizares J, *et al* (1996) Friedreich ' s Ataxia : Autosomal Recessive Disease Caused by an Intronic GAA Triplet Repeat Expansion Published by : American Association for the Advancement of Science Stable URL : http://www.jstor.org/stable/2890274. *Am Assoc Adv Sci* 271: 1423–1427

Chen X, Chen T, Ren L, Chen G, Gao X, Li G & Zhu X (2019) Triplex DNA nanoswitch for pH-sensitive release of multiple cancer drugs. *ACS Nano* 13: 7333–7344

Chen Y-L, He W, Kirmizialtin S & Pollack L (2022) Insights into the structural stability of major groove RNA triplexes by WAXS-guided MD simulations. *Cell Reports Phys Sci* 3: 100971

Cheng YK & Pettitt BM (1992) Hoogsteen versus reversed-Hoogsteen base pairing: DNA triple helices. *J Am Chem Soc* 114: 4465–4474

Christophe D, Cabrer B, Bacolla A, Targovnik H, Pohl V & Vassart G (1985) An usually long poly(purine)-poly(pyrimidine) sequence is located upstream from the human thyroglobulin gene. *Nucleic Acids Res* 13: 5127–5144

Ciotti P, Van Dyke M, Bianchi-Scarra G & Musso M (2001) Characterization of a triplex DNA-binding protein encoded by an alternative reading frame of loricrin. *Eur J Biochem* 268: 225–234

Cooney M, Czernuszewicz G, Postel EH, Flint SJ & Hogan ME (1988) Site-specific oligonucleotide binding represses transcription of the human c-myc gene in vitro. *Am Assoc Adv Sci* 241: 456–459

Crabbé P (1965) Optical rotatory dispersion and circular dichroism in organic chemistry Holden-Day

Dalla Pozza M, Abdullrahman A, Cardin CJ, Gasser G & Hall JP (2022) Three's a crowd – stabilisation, structure, and applications of DNA triplexes. *Chem Sci* 13: 10193–10215

Debin A, Laboulais C, Ouali M, Malvy C, Le Bret M & Svinarchuk F (1999) Stability of G,A triple helices. *Nucleic Acids Res* 27: 2699–2707

Fang Z & Rajewsky N (2011) The impact of mrRNA target sites in coding sequences and in 3'UTRs. *PLoS One* 6: e18067

Faucon B, Mergny J-L & Helene C (1996) Effect of third strand composition on the triple helix formation: purine versus pyrimidine oligodeoxynucleotides. *Nucleic Acids Res* 24: 3181–3188

Faustino I, Curutchet C, Luquec FJ & Orozco M (2014) The DNA-forming properties of 6-selenoguanine. *Phys Chem Chem Phys* 16: 1101–1110

Felsenfeld G, Davies DR & Rich A (1957) Formation of a Three-Stranded Polynucleotide Molecule. *J Am Chem Soc* 79: 2023–2024

Floris R, Scaggiante B, Manzini G, Quadrifoglio F & Xodo L (1999) Effect of cations on purine.purine.pyrimidine triple helix formation in mixed-valence salt solutions. *Eur J Biochem* 260: 801–809

Franklin RE & Gosling RG (1953) Molecular structure of nucleic acids: Molecular Configuration in Sodium Thymonucleate. *Nature* 171: 740–741

Ghane T, Brancolini G, Varsano D & Di Felice R (2012) Optical Properties of Triplex DNA from Time-Dependent Density Functional Theory. *J Phys Chem B* 116: 10693–10702

Giancola C, Buono A, Montesarchio D & Barone G (1999) Calorimetric, spectroscopic and computational investigation of DNA triplexes containing a 3'-3' internucleoside junction. *Phys Chem Chem Phys* 1: 5045–5049

Goñi JR, de la Cruz X & Orozco M (2004) Triplex-forming oligonucleotide target sequences in the human genome. *Nucleic Acids Res* 32: 354–360

Grabczyk E & Fishman MC (1995) A long purine-pyrimidine homopolymer acts as a transcriptional diode. *J Biol Chem* 270: 1791–1797

Gray DM, Hung S-H & Johnson KHBT-M in E (1995) [3] Absorption and circular dichroism spectroscopy of nucleic acid duplexes and triplexes. In *Biochemical Spectroscopy* pp 19–34. Academic Press

Guieysse A-L, Praseuth D & Hélène C (1997) Identification of a triplex DNA-binding protein from human cells. *J Mol Biol* 267: 289–298

Guo M, Hundseth K, Ding H, Vidhyasagar V, Inoue A, Nguyen CH, Zain R, Lee J. & Wu Y (2015) A distinct triplex DNA unwinding activity of ChlR1 helicase. *J Biol Chem* 290: 5174–5189

Hanvey JC, Williams EM & Besterman JM (1991) DNA triple-helix formation at physiologic pH and temperature. *Antisense Res Dev* 1: 307–317

Hartono YD, Xu Y, Karshikoff A, Nilsson L & Villa A (2018) Modeling p*K* Shift in DNA Triplexes Containing Locked Nucleic Acids. *J Chem Inf Model* 58: 773–783

Hausheer FH, Singh UC, Saxe JD, Flory JP & Tufto KB (1992) Thermodynamic and conformational characterization of 5-methylcytosine-versus cytosine-substituted oligomers in DNA triple helices: ab initio quantum mechanical and free energy perturbation studies. *J Am Chem Soc* 114: 5356–5362

Holder I, Wagner S, Xiong P, Sinn M, Frickey T, Meyer A & Hartig J (2015) Intrastrand triplex DNA repeats in bacteria: a source of genomic instability. *Nucleic Acids Res* 43: 10126–

10142

Hoogsteen K (1959) The structure of crystals containing a hydrogen-bonded complex of 1-methylthymine and 9-methyladenine. *Acta Crystallogr* 12: 822–823

Hoyuelos FJ, García B, Leal JM, Busto N, Biver T, Secco F & Venturini M (2014) RNA triplex-to-duplex and duplex-to-triplex conversion induced by coralyne. *Phys Chem Chem Phys* 16: 6012–6018

Hudson W & Ortland R (2014) The structure, function and evolution of proteins that bind DNA and RNA. *Nat Rev Mol Cell Biol* 15: 749–760

Iacovelli F, Hernandez KC, Desideri A & Falconi M (2019) Probing the Functional Topology of a pH-Dependent Triple Helix DNA Nanoswitch Family through Gaussian Accelerated MD Simulation. *J Chem Inf Model* 59: 2746–2752

Idili A, Vallée-Bélisle A & Ricci F (2014) Programmable pH-Triggered DNA Nanoswitches. *J Am Chem Soc* 136: 5836–5839

Jain A, Bacolla A, Chakraborty P, Grosse F & Vasquez KM (2010) Human DHX9 helicase unwinds triple-helical DNA structures. *Biochem* 49: 6992–6999

James PL, Brown T & Fox K (2003) Thermodynamic and kinetic stability of intermolecular triple helices containing different proportions of C+*GC and T*AT. *Nucleic Acids Res* 31: 5598–5606

Jenjaroenpun P & Kuznetsov VA (2009) *TTS Mapping*: integrative WEB tool for analysis of triplex formation target DNA Sequences, G-quadruplets and non-protein coding regulatory DNA elements in the human genome. *BMC Genomics* 10

Ji J, Hogan ME & Gao XL (1996) Solution structure of an antiparallel purine motif triplex containing a T center dot CG pyrimidine base triple. *Structure* 4: 425–435

Kanak M, Alseiari M, Balasubramanian P, Addanki K, Aggarwal M, Noorali S, Kalsum A, Mahalingam K, Pace G, Panasik N, *et al* (2010) Triplex-forming MicroRNAs form stable complexes with HIV-1 provirus and inhibit its replication. *Appl Immunohistochem Mol Morphol* 18: 532–545

Kang R, Zhang Q, Zeh III HJ, Lotze MT & Tang D (2013) HMGB1 in Cancer: Good, Bad, or Both? *Clin Cancer Res* 19: 4046–4057

Koshlap KM, Schultze P, Brunar H, Dervan PB & Feigon J (1997) Solution structure of an intramolecular DNA triplex containing an N-7-glycosylated guanine which mimics a protonated cytosine. *Biochem* 36: 2659–2668

Kotikam V, Kennedy SD, MacKay JA & Rozners E (2019) Synthetic, Structural, and RNA Binding Studies on 2-Aminopyridine-Modified Triplex-Forming Peptide Nucleic Acids. *Chem Eur J* 25: 4367–4372

Kumpina I, Brodyagin N, MacKay JA, Kennedy SD, Katkevics M & Rozners E (2019) Synthesis and RNA-Binding Properties of Extended Nucleobases for Triplex-Forming Peptide Nucleic Acids. *J Org Chem* 84: 13276–13298

Lander ES, Consortium IHGS, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, *et al* (2001) Initial sequencing and analysis of the human genome. *Nature* 409: 860–921

Laughton CA & Neidle S (1992) Molecular dynamics simulation of the DNA triplex d(TC)5 · d(GA)5 · d(C+T)5. *J Mol Biol* 223: 519–529

Lee I, Lee J, Lee NK & Hong SC (2012) Direct observation of the formation of DNA triplexes by single-molecule FRET measurements. *Curr Appl Phys* 12: 1027–1032

Lexa M, Martínek T, Burgetová I, Kopecek D & Brázdová M (2011) A dynamic programming algorithm for identification of triplex-forming sequences. *Bioinformatics* 27: 2510–2517

Li G, Tolstonog G & Traub P (2004) Interaction in vitro of type III intermediate filament proteins with triplex DNA. *DNA Cell Biol* 21: 163,188

Li J, Begbie A, Boehm B, Whidborne C, Pouferis Y, Huang D & Pukala T (2019) Ion Mobility-Mass Spectrometry Reveals Details of Formation and Structure for GAA·TCC DNA and RNA Triplexes. *J Am Soc Mass Spectrom* 30: 103–112

Li Y, Syed J & Sugiyama H (2016) RNA-DNA triplex formation by long noncoding RNAs. *Cell Chem Biol* 23: 1325–1333

Lohani N & Rajeswari M (2016) Preferential binding of anticancer drugs to triplex DNA compared to duplex DNA: a spectroscopic and calorimetric study. *RSC Adv* 6: 39903–39917

Luo J & Bruice TC (1998) Nanosecond molecular dynamics of hybrid triplex and duplex of

polycation deoxyribonucleic guanidine strands with a complimentary DNA strand. *J Am Chem Soc* 120: 1115–1123

Lyamichev VI, Mirkin SM & Frank-Kamenetskii MD (1986) Structures of Homopurine-homopyrimidine Tract in Superhelical DNA. *J Biomol Struct Dyn* 3: 667–669

Maher LJ, Wold B & Dervan PB (1989) Inhibition of DNA Binding Proteins by Oligonucleotide-Directed Triple Helix Formation. *Science* 245: 725–730

Maldonado R, Filarsky M, Grummt I & Langst G (2018) Purine- and pyrimidine-triple-helix-forming oligonucleotides recognize qualitatively different target sites at the ribosomal DNA locus. *RNA* 24: 371–380

Mall VS, Ojha RP & Tiwari RK (2018) Recombinant Triplex formed by PNA-TFO: A Molecular Dynamics Simulation Study. *Orient J Chem* 34: 2440–2446

Martianov I, Ramadass A, Barros A-S, Chow N & Akoulitchev A (2007) Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature* 445: 666–670

Mohan V, Cheng YK, Marlow GE & Pettitt BM (1993a) Molecular recognition of watson–crick base-pair reversals in triple-helix formation: Use of nonnatural oligonucleotide bases. *Biopolymers* 33: 1317–1325

Mohan V, Smith PE & Pettitt BM (1993b) Molecular dynamics simulation of ions and water around triplex DNA. *J Phys Chem* 97: 12984–12990

Mondal T, Subhash S, Vaid R, Enroth S, Uday S, Reinius B, Mitra S, Mohammed A, James AR, Hoberg E, *et al* (2015) MEG3 long noncoding RNA regulates the TGF-β pathway genes through formation of RNA–DNA triplex structures. *Nat Commun* 6: 7743

Muñoz J, Gelpí JL, Soler-López M, Subirana JA, Orozco M & Luque FJ (2002) Can divalent metal cations stabilize the triplex motif?: Theoretical study of the interaction of the hydrated Mg²⁺ cation with the G-G•C triplet. *J Phys Chem B* 106: 8849–8857

Musso M, Nelson L & Van Dyke M (1998) Characterization of Purine-Motif Triplex DNA-Binding Proteins in HeLa Extracts. *Biochem* 37: 3086–2095

Nelson L, Bender C, Mannsperger H, Buergy D, Kambakamba P, Mudduluru G, Korf U,

Hughes D, Van Dyke M & Allgayer H (2012) Triplex DNA-binding proteins are associated with clinical outcomes revealed by proteomic measurements in patients with colorectal cancer. *Mol Cancer* 11

Noé V, Aubets E, Félix AJ & Ciudad CJ (2021) Nucleic acid therapeutics using PolyPurine Reverse Hoogsteen hairpins. *Biochem Pharmacol* 189: 114371

Noy A, Maxwell A & Harris S (2017) Interference between triplex and protein binding to distal sites on supercoiled DNA. *Biophys J* 112: 523–531

Nyerges L, Puskas L, Tasi G, Penke B & Bottka S (1995) Quantum chemical design of new hydrogen-bond forming heterocyclic molecules suitable for the synthesis of sequence-specific DNA triplex. In *E.C.C.C. 1 Computational Chemistry: F.E.C.S. Conference*, Bernardi F & Rivail JL (eds) p 203. Woodbury: Aip Press

Ojha RP & Tiwari RK (2003) Triplex hydration: nanosecond molecular dynamics simulation of the solvated triplex formed by mixed sequences. *Nucleic Acids Res* 31: 6373–6380

Ou M, Li X, Zhao S, Cui S & Tu J (2020) Long non-coding RNA CDKN2B-AS1 contributes to atherosclerotic plaque formation by forming RNA-DNA triplex in the CDKN2B promoter. *eBioMedicine* 55

Pabon-Martinez Y, Xu Y, Villa A, Lundin KE, Geny S, Nguyen C-H, Pedersen EB, Jørgensen PT, Wengel J, Nilsson L, *et al* (2017) LNA effects on DNA binding and conformation: from single strand to duplex and triplex structures. *Sci Rep* 7: 11043

Pant P & Fisher M (2020) DNA triplex with conformationally locked sugar disintegrates to duplex: Insights from molecular simulations. *Biochem Biophys Res Commun* 532: 662–667

Papadakis G & Gizeli E (2012) In Silico Search of DNA Drugs Targeting Oncogenes. *Ieee-Acm Trans Comput Biol Bioinforma* 9: 1826–1830

Paugh S, Coss D, Bao J, Laudermilk L, Grance C, Ferreira A, Waddel B, Ridout G, Naeve D, Leuze M, *et al* (2016) MicroRNAs form triplexes with double stranded DNA at sequence-specific binding sites; a eukaryotic mechanism via which microRNAs could directly alter gene. *PLoS Comput Biol* 12: 1–20

Petraccone L, Erra E, Messere A, Montesarchio D, Piccialli G, De Napoli L, Barone G &

Giancola C (2004) Targeting duplex DNA with DNA-PNA chimeras? Physico-chemical characterization of a triplex DNA-PNA/DNA/DNA. *Biopolymers* 73: 434–442

Premraj BJ, Raja S & Yathindra N (2002) Structural basis for the unusual properties of 2',5' nucleic acids and their complexes with RNA and DNA. *Biophys Chem* 95: 253–272

Radhakrishnan I & Patel DJ (1993) Solution structure of a purine.purine.pyrimidine DNA triplex containing G.GC and T.AT triples. *Structure* 1: 135–152

Radhakrishnan I & Patel DJ (1994) Solution structure of a purine.purine.pyrimidine DNA triplex containing G.GC and T.AT triples. *Structure* 2: 17–32

Rajeswari M (2012) DNA triplex structures in neurodegenerative disorder, Friedreich's ataxia. *J Biosci* 37: 519–532

Rangel VM, Gu L, Chen G, Chen QH & Xue L (2022) 5-Substituted 3, 3', 4', 7-tetramethoxyflavonoids – A novel class of potent DNA triplex specific binding ligands. *Bioorg Med Chem Lett* 61: 128608

Rathinavelan T & Yathindra N (2006) Base triplet nonisomorphism strongly influences DNA triplex conformation: Effect of nonisomorphic G* GC and A* AT triplets and bending of DNA triplexes. *Biopolymers* 82: 443–461

Rosu F, Gavelica V, Houssier C, Colson P & De Pauw E (2002) Triplex and quadruplex DNA structures studied by electrospray mass spectrometry. *Rapid Commun Mass Spectrom* 16: 1729–1736

Ryan CA, Baskevics V, Katkevics M & Rozners E (2022) 2-Guanidyl pyridine PNA nucleobase for triple-helical Hoogsteen recognition of cytosine in double-stranded RNA. *Chem Commun* 58: 7148–7151

Sakamoto N, Chastain PD, Parniewski P, Oshima K, Pandolfo M, Griffith JD & Wells RD (1999) Sticky DNA: Self-Association Properties of Long GAA·TTC Repeats in R·R·Y Triplex Structures from Friedreich's Ataxia. *Mol Cell* 3: 465–475

Sakamoto T, Yamaoki Y, Nagata T & Katahira M (2021) Detection of parallel and antiparallel DNA triplex structures in living human cells using in-cell NMR. *Chem Commun* 57: 6364–6367

Sandström K, Wärmländer S, Gräslund A & Leijon M (2002) A-tract DNA disfavours triplex

formation. *J Mol Biol* 315: 737–748

Santos R, Lefevre S, Sliwa D, Seguin A, Camadro J-M & Lesuisse E (2010) Friedreich ataxia: molecular mechanisms, redox considerations, and therapeutic opportunities. *Antioxid Redox Signal* 13: 651–690

Schmitz U, Lai X, Winter F, Wolkenhauer O, Vera J & Gupta SK (2014) Cooperative gene regulation by microRNA pairs and their identification using a computational workflow. *Nucleic Acids Res* 42: 7539–7552

Schneider U, Severinsen J, Géci I, Okkels L, Jøhnk N, Mikkelson N, Klonge T, Pederson E, Westh H & Lisby G (2010) A novel FRET pair of parallel DNA triplexes by the LightCycler. *BMC Biotechnol* 10: 1–12

Schrodinger LLC (2015) The PyMOL Molecular Graphics System, Version 2.5.7. [PREPRINT]

Schroth GP & Ho PS (1995) Occurrence of potential cruciform and H-DNA forming sequences in genomic DNA. *Nucleic Acids Res* 23: 1977–1983

Sekar RB & Periasamy A (2003) Fluorescence resonance energy transfer (FRET) microscopy imaging of live cell protein localizations. *J Cell Biol* 160: 629–633

Semenyuk A, Darian E, Liu J, Majumdar A, Cuenoud B, Miller PS, Mackerell ADJ & Seidman MM (2010) Targeting of an interrupted polypurine:polypyrimidine sequence in mammalian cells by a triplex-forming oligonucleotide containing a novel base analogue. *Biochem* 49: 7867–7878

Sieradzan AK, Krupa P & Wales DJ (2017) What Makes Telomeres Unique? *J Phys Chem B* 121: 2207–2219

Sklenář V & Felgon J (1990) Formation of a stable triplex from a single DNA strand. *Nature* 345: 836–838

Soliva R, García RG, Blas JR, Eritja R, Asensio JL, González C, Luque FJ & Orozco M (2000) DNA-triplex stabilizing properties of 8-aminoguanine. *Nucleic Acids Res* 28: 4531–4539

Soliva R, Laughton CA, Luque FJ & Orozco M (1998) Molecular dynamics simulations in aqueous solution of triple helices containing d(G•C•C) trios. *J Am Chem Soc* 120: 11226–11233

Spacková N, Cubero E, Sponer J & Orozco M (2004) Theoretical study of the guanine→6-

thioguanine substitution in duplexes, triplexes, and tetraplexes. *J Am Chem Soc* 126: 14642–14650

Sun B-W, Babu R, Soresnsen M, Zakrzewska K, Wengel J & Sun J-S (2004) Sequence and pH effects of LNA-containing triple helix-forming oligonucleotides: Physical chemistry, Biochemistry and Modeling studies. *Biochem* 43: 4160–4196

Sun JS & Helene C (1993) Oligonucleotide directed triple helix formation. *Curr Opin Struct Biol* 3: 345–356

Swain M, Ageeli AA, Kasprzak WK, Li M, Miller JT, Sztuba-Solinska J, Schneekloth JS, Koirala D, Piccirili J, Fraboni AJ, *et al* (2021) Dynamic bulge nucleotides in the KSHV PAN ENE triple helix provide a unique binding platform for small molecule ligands. *Nucleic Acids Res* 49: 13179–13193

Szabat M, Kierzek E & Kierzek R (2018) Modified RNA triplexes: Thermodynamics, structure and biological potential. *Sci Rep* 8

Talbott JM, Tessier BR, Harding EE, Walby GD, Hess KJ, Baskevics V, Katkevics M, Rozners E & Mackay JA (2023) Improved Triplex-Forming Isoorotamide PNA Nucleobases for A-U Recognition of RNA Duplexes. *Chem Eur J*: 12

Thenmalarchelvi R & Yathindra N (2005) New insights into DNA triplexes: residual twist and radial difference as measures of base triplet non-isomorphism and their implication to sequence-dependent non-uniform DNA triplex. *Nucleic Acids Res* 33: 43–55

Valinezhad Orang A, Safaralizadeh R & Kazemzadeh-Bavili M (2014) Mechanisms of miRNA-mediated gene regulation from common downregulation to mRNA-specific upregulation. *Int J Genomics* 2014: 1–15

Vekhoff P, Ceccaldi A, Polverari D, Pylouster J, Pisano C & Arimondo PB (2008) Triplex formation on DNA targets: How to choose the oligonucleotide. *Biochem* 47: 12277–12289

Vorlíčková M, Kejnovská I, Bednářová K, Renčiuk D & Kypr J (2012) Circular Dichroism Spectroscopy of DNA: From Duplexes to Quadruplexes. *Chirality* 24: 691–698

Wang E, Koshlap KM, Gillespie P, Dervan PB & Feigon J (1996) Solution structure of a pyrimidine purine pyrimidine triplex containing the sequence-specific intercalating

non-natural base D-3. *J Mol Biol* 257: 1052–1069

Warwick T, Brandes RP & Leisegang MS (2023) Computational Methods to Study DNA:DNA:RNA Triplex Formation by lncRNAs. *Non-Coding RNA* 9: 10

Watson J & Crick F (1953) Molecular structure of nucleic acids: A structure for Deoxyribose Nucleic Acid. *Nature* 171: 737–738

Xi HJ, Davis E, Ranjan N, Xue L, Hyde-Volpe D & Arya DP (2011) Thermodynamics of Nucleic Acid 'Shape Readout' by an Aminosugar. *Biochem* 50: 9088–9113

Xue L, Xi H, Kumar S, Gray D, Davis E, Hamilton P, Skirba M & Arya DP (2010) Probing the recognition surface of a DNA triplex: Binding studies with intercalator-neomycin conjugates. *Biochem* 49: 5540–552

Yang H, Yao W, Wang Y, Shi L, Su R, Wan D, Xu N, Lian W, Chen C & Liu S (2017) High-throughput screening of triplex DNA binders from complicated samples by 96-well pate format in conjunction with peak area-fading UHPLC-Orbitrap MS †. *Analyst* 142: 670

Yang LJ, Liu MZ, Deng WL, Wang C, Bai CL & Kan LS (1999) Influence of 5-bromodeoxycytosine substitution on triplex DNA stability and conformation. *Biophys Chem* 76: 25–34

Zhang JH, Fakharzadeh A, Pan F, Roland C & Sagui C (2021) Construction of DNA/RNA Triplex Helices Based on GAA/TTC Trinucleotide Repeats. *Bio-Protocol* 11: 13

Zhang JH, Fakharzadeh A, Roland C & Sagui C (2022) RNA as a Major-Groove Ligand: RNA-RNA and RNA-DNA Triplexes Formed by GAA and UUC or TTC Sequences. *Acs Omega* 7: 38728–38743

Zhang ZL, Wu YY, Xi K, Sang JP & Tan ZJ (2017) Divalent Ion-Mediated DNA-DNA Interactions: A Comparative Study of Triplex and Duplex. *Biophys J* 113: 517–528

Zhou Z, Giles K & Felsenfeld G (2019) DNA·RNA triple helix formation can function as a cis-acting regulatory mechanism at the human β-globin locus. *Proc Natl Acad Sci U S A* 116: 6130–6139

Zhou Z, Li X, Deng C, Ney N, Huang S & Bungert J (2010) USF and NF-E2 cooperate to regulate the recruitment and activity of RNA polymerase II in the β-globulin gene locus. *J Biol Chem* 285: 15894–15905

1.3 Aims

This thesis aims to demonstrate the use of native MS to investigate model and biologically relevant DNA triplex systems. The effects of different buffer conditions and ionisation mode for MS based experiments is explored. Different base modifications in the TFO will be compared to determine their triplex forming propensity. A preliminary antigene strategy will be designed to discover the capability of using triplexes to modulate gene expression. The binding interactions of small molecules to different DNA triplexes will also be investigated to identify any stabilising interactions.

Chapter 2: Experimental methods

2.1 Introduction

In comparison to the canonical duplex DNA, there is significantly less known about the formation and structure of DNA triplexes. To enhance our understanding of these triple helical structures and exploit TFOs for their biological activity, it is imperative that we examine their structure and formation. Employing *in vitro* biophysical methods is critical as they allow for triplex systems to be analysed in physiologically relevant conditions which could potentially lead to findings on how they behave in biology.

2.2 Mass spectrometry

Mass spectrometry (MS) is an analytical technique that allows for the characterisation of different species present in a sample by determining the mass to charge (*m/z*) ratio of positive or negative ions in the gas phase. MS has evolved greatly over recent decades, with the development of many instrument platforms that have been used in a wide range of biological disciplines including the analysis of non-covalent complexes, proteomics, lipidomics as well as pharmaceutical and forensic analysis.[1–5]

MS is very versatile, with different experimental approaches that can be tailored towards enhancing the analysis of various samples. In the case of DNA triplexes, native MS is the preferred approach because it allows the complexes that are being analysed to retain their weaker non-covalent interactions and biologically relevant structures. Native MS requires the use of mild solution conditions and gentle transmission within the mass spectrometer to ensure the native structure is kept intact. Given triplex samples can exist as an equilibrium between the TFO, duplex and triplex assemblies, MS offers significant advantages since it can detect each individual species within a dynamic system and determine their relative abundances.

DNA triplexes were first observed using native MS in 2002 by Rosu *et.al*, however they have been notably less well studied when compared to other aspects of MS-related nucleic acid research including DNA-protein crosslinks, mapping DNA lesions and detection of DNA adducts.[6–9] More recent publications have shown using ion-mobility MS to demonstrate that triplexes maintain their solution-phase structure in the gas phase, as well as that parallel triplexes are favoured for sequences related to Friedreich's ataxia.[10,11]

2.2.1 Electrospray ionisation

Electrospray ionisation (ESI) stands out as the principle soft ionisation method that can preserve the structure of large, non-covalent macromolecules in the gas phase.[14] ESI produces charged droplets from a small conductive capillary which is used to deliver the charged ions into the mass spectrometer. An electric potential is applied to the capillary containing sample which induces formation of a Taylor cone at the tip of the capillary. Eventually the stability of the Taylor cone is compromised, which leads to the production of highly charged droplets containing analyte. The stream of charged droplets are then exposed to different instrument conditions such as elevated temperatures and gas flow, which starts to reduce the size of the droplet. Eventually, the droplet radius decreases such that it reaches the Rayleigh limit, where the electrostatic repulsion of the droplet becomes so large that it surpasses the surface tension, and droplet fission occurs. Once the droplets have reached the Rayleigh limit, they begin to evaporate and as the final solvent shell has been removed the charge of the declining droplet is transferred to the analyte and allowing the ions to be emitted to the mass spectrometer as shown in Figure 2.1.[15–17] This is process is known as the charge residue model, which applies to larger biomolecules.[17]

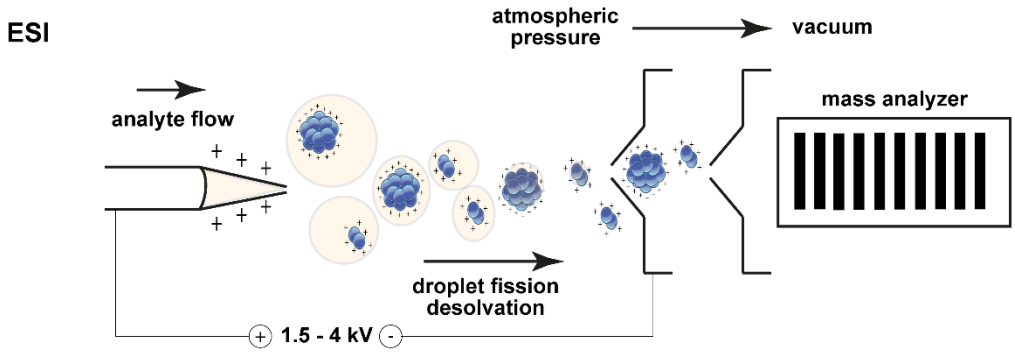


Figure 2.1: Graphic representation of the electrospray ionisation (ESI) process. The sample is introduced from solution phase through the conductive capillary. Generation of charged liquid analyte droplets occurs from the Taylor cone due to desolvation and fission processes, with ions then directed towards the mass spectrometer for analysis. Figure adapted with permission from Jovcevski *et.al*.[18]

ESI offers the advantage of generating charged analyte ions directly from the sample without the need for harsh conditions and chemicals to promote ionisation. Solutions such as ammonium acetate and ammonium bicarbonate are commonly used to emulate biological conditions as they contain volatile salts that can produce analyte ions free of salt adducts.

Nanoelectrospray ionisation (nESI) is a modified method that requires significantly less material (only a few microlitres of sample) and slower flow rates, which can enhance ion formation from smaller charged droplets arising due to the significantly smaller capillary orifice.[19] This method of ionisation is therefore preferred for native MS due to the better desolvation achieved under more mild conditions than regular ESI. nESI is well equipped for the analysis of DNA triplex systems due to its ability to preserve non-covalent interactions.

2.2.2 Quadrupole time-of-flight mass spectrometry

In order to determine the *m/z* ratio of the charged ions they need to be separated after the ionisation process. Many MS platforms can be used for the analysis of DNA triplexes, however a quadrupole time-of-flight (Q-TOF) mass spectrometer was primarily used for data collection for this thesis on both Bruker Impact HD Q-TOF and Agilent 6560 IM-Q-TOF platforms. Q-TOF MS is beneficial for native MS due to its large mass range which allows for the analysis of small molecules as well as large protein complexes. Q-TOF instruments are also compatible with soft ionisation methods allowing the conservation of non-covalent interactions and minimising the fragmentation of larger biomolecules.

Since the ion mobility capabilities of the Agilent instrument were not utilised in this work, only the Q-TOF elements will be explained in detail. Q-TOF MS is an analytical technique that combines the features of two different mass analysers. The high mass range and selectivity of the quadrupole is partnered with the high-resolution capabilities of a TOF analyser. The schematic construction of a Q-TOF mass spectrometer can be seen in Figure 2.2.

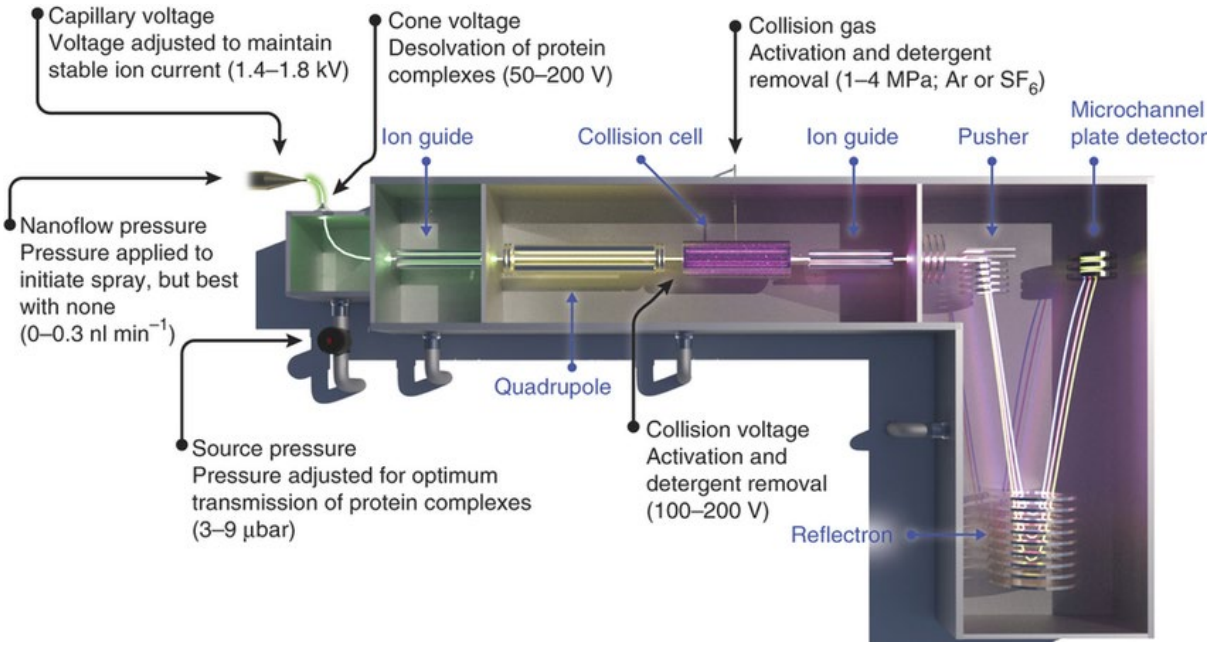


Figure 2.2: Schematic representation of a Q-TOF mass spectrometer. Analytes of interest are ionised and transferred into the gas phase. The charged ions now travel through the quadrupole where ions of specific *m/z* can be selected. Selected ions then move to the collision cell where they are exposed to collision with inert gases which induces fragmentation. Resulting ions are directed towards the TOF mass analyser where they are accelerated towards the detector and the TOF is recorded and converted into a mass spectrum. Figure adapted with permission from Laganowsky *et.al*.[20]

After the analyte of interest is transferred into the gas phase, the ions travel through the quadrupole. A quadrupole mass analyser consists of four parallel conducting rods that are connected to a direct current potential and superimposed with an alternating current potential (as shown in Figure 2.3A). This allows for the sets of opposing rods to act as positive and negative potentials respectively. Quadrupoles can act as a mass filter, allowing for the selection of specific analyte ions for selected fragmentation based on their *m/z* which have a stable trajectory through the quadrupole. Unselected ions will have unstable trajectories and therefore will not be able to transverse the quadrupole to reach the detector.

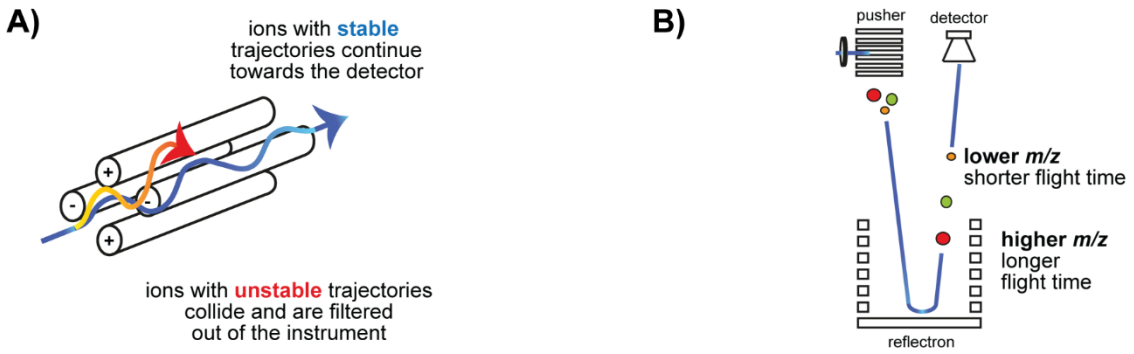


Figure 2.3: Schematic representation of **(A)** quadrupole mass filter and **(B)** time-of-flight (TOF) mass analyser components of a Q-TOF mass spectrometer. The quadrupole contains four parallel electrodes that generate an electric field that can select for specific ions. The TOF mass analyser can separate ions based on their mass and charge by determining their time to reach the detector, with the heavier and highly charged ions having a longer and shorter TOF respectively. Figure adapted with permission from Jovcevski *et.al.*[18]

Following any ion selection by the quadrupole, ions are reaccelerated by an electric field orthogonal to their original flight path in the TOF component. Here the *m/z* of analyte ions can be determined by measuring their time of flight through the TOF sector, from the point of acceleration to the ion detector (as shown in Figure 2.3B). The ion of a particular mass (*m*) and charge (*z*) has an accelerating potential (V), giving it kinetic energy (KE) as shown in Equation 1 below:

$$KE = zeV = \frac{mv^2}{2} \quad \textbf{(1)}$$

This equation can then be rearranged to relate the velocity to distance (*d*) and time (*t*) to give the following equation:

$$\frac{m}{z} = \frac{2eVt^2}{d^2} \quad \textbf{(2)}$$

As can be seen from Equation 2, when the kinetic energy that is imparted to all the ions is the same, the velocity at which they move is inversely proportional to their mass and proportional to their charge. Since the ions must travel a fixed distance to reach the detector, this means that the lighter ions will reach the detector before the heavier ions of the same charge and therefore have a shorter TOF (Figure 2.3B). TOF analysers typically contain a reflectron, which corrects for kinetic energy dispersion and spread of the analyte ions that exhibit the same *m/z* with differing velocities, allowing them to reach the detector at the same time. The reflectron also improves mass resolution by increasing the flight path of the ions. Once the ions arrive at the detector, the flight time of each ion is then converted into the unique mass

peak. Ion abundance is also captured from the detector to give the full mass spectrum, for comparison of abundance of different components in a system. TOF MS is known for its high resolution and speed, making it valuable for a variety of applications including proteomics, toxicology screening and analysis of complex mixtures.[21–24]

2.2.3 Collision-induced dissociation

Collision-induced dissociation (CID) is an ion activation technique that is commonly used to elucidate structural and stability properties of analytes in an MS experiment. When using a Q-TOF instrument, ions are typically accelerated by an electric field into a collision cell where they are exposed to a neutral gas. The ions experience high-energy collisions, which convert kinetic energy into vibrational excitation. The increase in energy allows for the analyte ions containing covalent and non-covalent bonds to dissociate into fragments. Varying the collision energy by modulation of the accelerating electric field means that the energy the analyte ions are exposed to can be manipulated. The ability to change the degree of fragmentation is important due to different molecules intrinsically fragmenting at different energies, allowing for optimised conditions. This also allows for the selective breaking of different bonds including non-covalent interactions, to obtain structural information. In terms of DNA triplexes, by increasing the collision energy in the system, the non-covalent interactions that hold together both the triplex and duplex structures can break down. This allows for the identification and stability of the triplex structures to be determined as the TFO can be ejected from the duplex structure and the collision energies required for dissociation can be compared for different systems in order to determine their relative stability in the gas phase.

2.3 UV-Vis spectroscopy

Ultraviolet-visible (UV-Vis) spectroscopy is a solution phase spectrophotometric technique that is used to measure light intensity in the UV (10-400 nm) and visible (400-800 nm) regions. The spectrum results from the interaction of electromagnetic radiation with the analyte, where the analyte of interest absorbs light of specific wavelengths which can be detected by the spectrophotometer.[25] The absorption phenomenon of light applies to UV-Vis spectroscopy as the amount of light absorbed is directly proportional to the amount of analyte present in the sample, as described by the Beer-Lambert law (Equation 3):

$$A = \varepsilon l c \quad \textbf{(3)}$$

Where A is the absorbance at the specific wavelength of interest, ε is the molar extinction coefficient, l is the path length of light through the sample in the UV-vis spectrophotometer and c is the concentration of the analyte.

The thermal dissociation of DNA from its duplex conformation into its two constituent strands typically occurs between 60-70 °C, although this can vary depending on the nucleotide composition and length of the sequence, reflecting the stability of the DNA assembly. Longer DNA sequences exhibit higher melting temperatures due to increased amounts of hydrogen bonding, requiring more energy for dissociation. DNA has optimal UV absorbance at 260 nm which allows the melting event to be detected via UV-Vis spectroscopy at this specific wavelength. The associated increase in absorbance (hyperchromicity) upon melting results from the denaturation of the DNA structure causing base unstacking that allows the exposed bases to absorb more light. UV-Vis spectroscopy can also be used to measure triplex DNA melting, which usually occurs at a lower temperature compared to the canonical DNA duplex.[26] Additionally, Y-type triplex melting events can also be observed at 300 nm due to absorbance differences between non-protonated and protonated cytosine bases. Protonation alters absorbance by contributions of the n->π transition moments from the increased DNA base stacking (hypochromicity).[27]

Derivative plots of the absorbance vs temperature are commonly used to present DNA melting curves from UV-Vis spectroscopy data. The duplex and triplex melting points are then identified by the maximum point of the corresponding peak, as opposed to the points of inflection for the absorbance vs temperature plots. Melting temperatures can be used to

verify triplex formation and assess the stability of these structures under various experimental conditions. However, this solution phase method is an ensemble technique making it hard to distinguish between different species in the triplex system. Therefore, UV-Vis spectroscopy and MS are highly complementary techniques.

2.4 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) has emerged as a commonly utilised thermodynamic technique to characterise the affinity and stoichiometry of binding interactions between biomolecules. Recent developments in ITC have allowed for more sensitive instruments that are able to characterise interactions between different macromolecules with high accuracy and precision.[28]

In an ITC experiment, binding interactions are measured by the detection of the heat absorbed or released during titration of a titrant into the system over multiple injections. The heat produced is proportional to the amount of product species formed. Therefore, the stoichiometry and dissociation constants (such as K_D of the interaction) can be determined. Using the concentration of the reactants, the heat absorbed/released can be converted into a measure of binding enthalpy (ΔH). Since the Gibbs free energy (ΔG) of binding can be derived from the K_a value, and having the ΔH value from the heat produced, the entropy values can then be readily calculated for a single system without the use of van't Hoff plots (using Equations 4 and 5 below).[29]

$$\Delta G = -RT \times lnK_a \quad \textbf{(4)}$$

$$\Delta G = \Delta H - T\Delta S \quad \textbf{(5)}$$

Where ΔG is Gibbs free energy, R is the gas constant, T is temperature (Kelvin), K_a is association binding constant, ΔH is enthalpy of binding and ΔS is entropy of binding.

Thermodynamic and kinetic evaluation of a triplex system help understand the binding of the TFO to the duplex structure under various experimental conditions. ITC is unique as most investigations into the kinetics of DNA triplexes study the dissociation of the species (such as UV-vis spectroscopy) whereas ITC characterises the assembly of the triplex.[30] Early thermodynamic and kinetics studies indicated that triplex formation wasn't as simple as a two-state reaction using filter-binding assays.[31] This is due to the effects of many factors on triplex formation, especially sequence and pH specificity. The different compositions of bases within different sequences can have a huge impact on their pH dependence and therefore the formation. Specific pH-dependent triplex systems are known to form other nucleic acid species at lower pH including i-motifs, which would influence the thermodynamics of the

triplex system.[32] This makes the sequences and pH of the conditions for experiments important factors to consider. ITC is a useful complementary tool for the analysis of triplex formation, especially given the ability to compare the thermodynamic properties of the system with information obtained by MS and UV-Vis spectroscopy.

2.5 References

(1) Loo, J. A. Studying Noncovalent Protein Complexes by Electrospray Ionization Mass Spectrometry. *Mass Spectrom. Rev.* **1997**, *16* (1), 1–23.

(2) Mallick, P.; Kuster, B. Proteomics: A Pragmatic Perspective. *Nat. Biotechnol.* **2010**, *28* (7), 695–709.

(3) Blanksby, S. J.; Mitchell, T. W. Advances in Mass Spectrometry for Lipidomics. *Annu. Rev. Anal. Chem.* **2010**, *3* (1), 433–465.

(4) Astarita, G.; Kelly, R. S.; Lasky-Su, J. Metabolomics and Lipidomics Strategies in Modern Drug Discovery and Development. *Drug Discov. Today* **2023**, *28* (10), 103751.

(5) Jackson, G. P.; Barkett, M. A. Forensic Mass Spectrometry: Scientific and Legal Precedents. *J. Am. Soc. Mass Spectrom.* **2023**, *34* (7), 1210–1224.

(6) Rosu, F.; Gavelica, V.; Houssier, C.; Colson, P.; De Pauw, E. Triplex and Quadruplex DNA Structures Studied by Electrospray Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2002**, *16* (18), 1729–1736.

(7) Stingele, J.; Bellelli, R.; Alte, F.; Hewitt, G.; Sarek, G.; Maslen, S. L.; Tsutakawa, S. E.; Borg, A.; Kjær, S.; Tainer, J. A.; Skehel, J. M.; Groll, M.; Boulton, S. J. Mechanism and Regulation of DNA-Protein Crosslink Repair by the DNA-Dependent Metalloprotease SPRTN. *Mol. Cell* **2016**, *64* (4), 688–703.

(8) Ohnishi, I.; Iwashita, Y.; Matsushita, Y.; Ohtsuka, S.; Yamashita, T.; Inaba, K.; Fukazawa, A.; Ochiai, H.; Matsumoto, K.; Kurono, N.; Matsushima, Y.; Mori, H.; Suzuki, S.; Suzuki, S.; Tanioka, F.; Sugimura, H. Mass Spectrometric Profiling of DNA Adducts in the Human Stomach Associated with Damage from Environmental Factors. *Genes Environ.* **2021**, *43* (1), 12.

(9) Chowdhury, G.; Guengerich, F. P. Mass Spectrometry of Nucleic Acids. In *Encyclopedia of Analytical Chemistry*; 2015; pp 1–39..

(10) Arcella, A.; Portella, G.; Ruiz, M.; Eritja, R.; Vilaseca, M.; Gabelica, V.; Orozco, M. Structure of Triplex DNA in the Gas Phase. *J. Am. Chem. Soc.* **2012**, *134* (15), 6596–6606.

(11) Li, J.; Begbie, A.; Boehm, B.; Whidborne, C.; Pouferis, Y.; Huang, D.; Pukala, T. Ion Mobility-Mass Spectrometry Reveals Details of Formation and Structure for GAA·TCC DNA and RNA Triplexes. *J. Am. Soc. Mass Spectrom.* **2019**, *30* (1), 103–112.

(12) Zyner, K. G.; Simeone, A.; Flynn, S. M.; Doyle, C.; Marsico, G.; Adhikari, S.; Portella, G.; Tannahill, D.; Balasubramanian, S. G-Quadruplex DNA Structures in Human Stem Cells and Differentiation. *Nat. Commun.* **2022**, *13* (1), 142.

(13) Bag, S.; Burman, M. D.; Bhowmik, S. Structual Insights and Shedding Light on Preferential Interactions of Dietary Flavonoids with G-Quadruplex DNA Structures: A New Horizon. *Heliyon* **2023**, *9* (3), e13959.

(14) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray Ionization for Mass Spectrometry of Large Biomolecules. *Science.* **1989**, *246* (4926), 64–71.

(15) Ho, C. S.; Lam, C. W. K.; Chan, M. H. M.; Cheung, R. C. K.; Law, L. K.; Lit, L. C. W.; Ng, K. F.; Suen, M. W. M.; Tai, H. L. Electrospray Ionisation Mass Spectrometry: Principles and Clinical Applications. *Clin. Biochem. Rev.* **2003**, *24* (1), 3–12.

(16) Wilm, M. Principles of Electrospray Ionization. *Mol. Cell. Proteomics* **2011**, *10* (7).

(17) Konermann, L.; Ahadi, E.; Rodriguez, A. D.; Vahidi, S. Unraveling the Mechanism of Electrospray Ionization. *Anal. Chem.* **2013**, *85* (1), 2–9.

(18) Jovcevski, B.; Pukala, T. L. Mass Spectrometry and Its Applications. In *Biomolecular and Bioanalytical Techniques*; Ramesh, V., Ed.; 2019; pp 219–253.

(19) Wilm, M.; Mann, M. Analytical Properties of Nanoelectrospray Ion Source. *Anal. Chem.* **1996**, *68* (1), 1–8.

(20) Laganowsky, A.; Reading, E.; Hopper, J. T. S.; Robinson, C. V. Mass Spectrometry of Intact Membrane Protein Complexes. *Nat. Protoc.* **2013**, *8* (4), 639–651.

(21) Aballo, T. J.; Roberts, D. S.; Melby, J. A.; Buck, K. M.; Brown, K. A.; Ge, Y. Ultrafast and Reproducible Proteomics from Small Amounts of Heart Tissue Enabled by Azo and TimsTOF Pro. *J. Proteome Res.* **2021**, *20* (8), 4203–4211.

(22) Lesur, A.; Schmitt, P.-O.; Bernardin, F.; Letellier, E.; Brehmer, S.; Decker, J.; Dittmar, G. Highly Multiplexed Targeted Proteomics Acquisition on a TIMS-QTOF. *Anal. Chem.* **2021**, *93* (3), 1383–1392.

(23) Allen, D. R.; McWhinney, B. C. Quadrupole Time-of-Flight Mass Spectrometry: A Paradigm Shift in Toxicology Screening Applications. *Clin. Biochem. Rev.* **2019**, *40* (3), 135–146.

(24) Boes, K. S.; Roberts, M. S.; Vinueza, N. R. Rapid Quadrupole-Time-Of-Flight Mass Spectrometry Method Quantifies Oxygen-Rich Lignin Compound in Complex Mixtures.

J. Am. Soc. Mass Spectrom. **2017**, *29*, 535–542.

(25) Akash, M. S. H.; Rehman, K. Ultraviolet-Visible (UV-VIS) Spectroscopy. In *Essentials of Pharmaceutical Analysis*; Springer: Singapore, 2019; pp 29–56.

(26) Jain, A.; Rajeswari, M.; Ahmed, F. Formation and Thermodynamic Stability of Intermodular (R*R•Y) DNA Triplex in GAA/TTC Repeats Associated with Freidreich's Ataxia. *J. Biomol. Struct. Dyn.* **2002**, *19* (4), 691–699.

(27) Mergny, J.-L.; Li, J.; Lacroix, L.; Amrane, S.; Chaires, J. Thermal Difference Spectra: A Specific Signature for Nucleic Acid Structures. *Nucleic Acids Res.* **2005**, *33* (16), 1–6.

(28) Doyle, M. Characterization of Binding Interactions by Isothermal Titration Calorimetry. *Curr. Opin. Biotechnol.* **1997**, *8* (1), 31–35.

(29) Salim, N. N.; Feig, A. L. Isothermal Titration Calorimetry of RNA. *Methods* **2009**, *47* (3), 198–205.

(30) Plum, G. E.; Park, Y.-W.; Singleton, S. F.; Dervan, P. B.; Breslauer, K. J. Thermodynamic Characterization of the Stability and the Melting Behavior of a DNA Triplex: A Spectroscopic and Calorimetric Study. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87* (23), 9436–9440.

(31) Shindo, H.; Torigoe, H.; Sarai, A. Thermodynamic and Kinetic Studies of DNA Triplex Formation of an Oligohomopyrimidine and a Matched Duplex by Filtering Assay. *Biochem* **1993**, *32* (42), 8963–8969.

(32) Abou Assi, H.; Garavis, M.; González, C.; Damha, M. J. I-Motif DNA: Structual Features and Significance to Cell Biology. *Nucleic Acids Res.* **2018**, *46* (16), 8038–8056.

Chapter 3: Native Mass Spectrometric Insights into the Formation and Stability of DNA Triplexes

3.1 Foreword and authorship statement

It is proposed that DNA triplexes play important roles in many biological processes including gene regulation, making these structures good candidates to be used in antigene therapies. Despite their significance, key aspects of triplex formation, stability and biological activities remain poorly understood. In this chapter, native MS methods are compared to more traditional solution phase techniques, including UV-Vis spectroscopy and isothermal titration calorimetry. Using model DNA triplex systems, the fundamental factors that influence triplex formation and detection through MS, including salt concentration, pH and ionization mode are investigated. The advantages of using MS to analyse triplex systems are highlighted, showing the ability to detect unexpected DNA assemblies as a result of changes in pH, that can't be resolved by solution phase techniques. Intriguingly, different species are observed under the different ionization modes, suggesting the need for careful consideration of experimental parameters for analysis of these systems. This work provides valuable insights into the field of native MS for higher order DNA assemblies.

Statement of Authorship

Title of Paper	Native Mass Spectrometric Insights into the Formation and Stability of DNA triplexes
Publication Status	☐ Published ☐ Accepted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style ☑ Submitted for Publication
Publication Details	This paper has been submitted to the Journal of the American Society for Mass Spectrometry

Principal Author

Name of Principal Author (Candidate)	Jack Klose		
Contribution to the Paper	Data collection, data analysis, manuscript preparation		
Overall percentage (%)	50		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	15 Dec 2023

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Alexander Begbie		
Contribution to the Paper	Data collection, data analysis, editing manuscript and presentation		
Signature		Date	15 Dec 2023

Name of Co-Author	Luis Toronjo-Urquiza		
Contribution to the Paper	Methodology, data analysis		
Signature		Date	15 Dec 2023

Please cut and paste additional co-author panels here as required.

Name of Co-Author	Tara Pukala		
Contribution to the Paper	Project conceptualisation, methodology, data analysis, editing manuscript and presentation		
Signature		Date	16/12/2023

Native Mass Spectrometric Insights into the Formation and Stability of DNA Triplexes

Jack W. Klose[‡1], Alexander J. Begbie[‡1], Luis Toronjo-Urquiza[2], Tara L. Pukala[1]*

[1]Discipline of Chemistry, The University of Adelaide, Adelaide, South Australia, 5005, Australia
[2]School of Chemical Engineering, The University of Adelaide, Adelaide, South Australia, 5005, Australia

ABSTRACT: Deoxyribonucleic acid is a genetic biomacromolecule that contains the inherited information required to build and maintain a living organism. While the canonical duplex DNA structure is rigorously characterized, the structure and function of higher order DNA structures such as DNA triplexes are comparatively poorly understood. Previous literature has shown that these triplexes can form under physiological conditions, and native mass spectrometry offers a useful platform to study their formation and stability. However, experimental conditions including buffer salt concentration, pH, and instrumentation parameters such as ion mode can confound analysis by impacting the amount of triplex observed by mass spectrometry. Using model 30mer Y-type triplex sequences, we demonstrate the influence a range of experimental variables have on the detection of DNA triplex structures, informing suitable conditions for study. When carefully considered conditions are used, mass spectrometry offers a powerful complementary tool for analysis of higher order DNA assemblies.

INTRODUCTION

In addition to the archetypal double helix, DNA is capable of forming several non-canonical structures including the triplex DNA structure first described in 1957.[1] DNA triplexes form when a triplex forming oligonucleotide (TFO) binds in the major groove of the duplex (Figure 1A) to a purine rich sequence known as a triplex target site.[2] The polypurine stretch (R strand) is essential for the formation of triplexes as purines contain exposed unoccupied hydrogen bonding sites which can be used for interaction with the major groove by the TFO through Hoogsteen bonds. Two main triplex motifs have been characterized, namely the purine (R-type) and pyrimidine (Y-type) motifs.[3] The Y-type TFO is made up of cytosine and thymine (uracil in the case of RNA) and binds in a parallel manner to the complementary purine duplex sequence. In the R-type motif, the TFO comprises adenine and guanine binds to the complementary sequence in an antiparallel manner through reverse Hoogsteen bonding (Figures 1B-1E).[4,5]

There is a great deal of ambiguity and a significant knowledge gap regarding the mode of formation, structure and biological roles of triplex DNA.[6,7] It is known that binding of the TFO to the duplex DNA results in a less thermodynamically stable structure compared to the canonical duplex. This is partially due to the proximity of the three negatively charged DNA strands which causes an unfavorable charge repulsion, as well as due to the steric hinderance resulting from addition of the extra strand.[8] It has also been shown that the presence of multivalent cations such as Mg^{2+} or polyamines increases the stability of triplexes by forming stabilizing electrostatic interactions, especially in R-type triplexes.[9] Increasing the length of the TFO is also known to make the triplex more stable due to the increasing hydrophobic effect, π-π stacking, and additional stabilizing Hoogsteen bonds being formed.[10] However, there is significant debate in the literature regarding the stability and conditions under which Y and R type triplex DNA forms.[11,12]

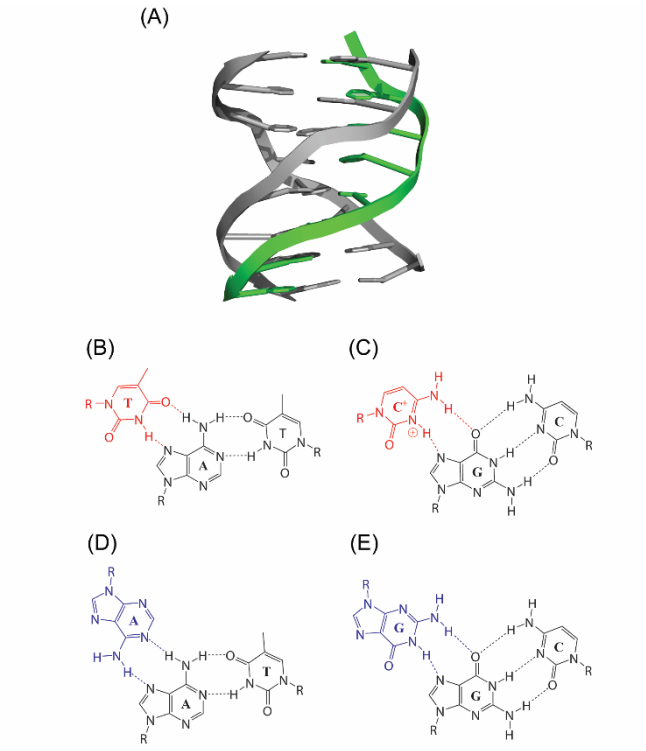


Figure 1. (A) Y-type DNA triplex structure. The triplex forming oligonucleotide (green) can bind within the major groove of the duplex DNA (grey) to form a triplex structure.[13] **(B-E)** Structures of the triplex motifs. Watson-Crick base pairs are shown in black for A-T and G-C. Y-type triplexes: TFO bases **(B)** thymine and **(C)** cytosine are indicated in blue with Hoogsteen bonding

shown by blue dashed lines. R-type triplexes: TFO bases **(D)** adenine and **(E)** guanine are indicated in red with reverse Hoogsteen bonding shown by red dashed lines.

Naturally occurring DNA triplexes have been primarily linked to functions associated with transcriptional regulation, gene expression and post-transcriptional RNA processing.[14] Based on this, selective control of gene expression using antigene methods is a strategy that has shown emerging promise in development of new therapeutic approaches.[15,16] However, while there has been some progress towards unlocking the full potential of triplexes as gene regulatory agents *in vivo,* issues including poor stability and the need to deliver TFOs to the nucleus of a cell remain.[17,18]

In order to better understand these non-canonical structures, it is necessary to utilize a range of biophysical methods. The formation of DNA triplexes and the identity of their constituent strands can be determined by native mass spectrometry (MS) through the detection of ions specific to the molecular weight of the different species in the sample. Stability and structural characterization of these assemblies can also be interrogated by dissociating the triplex ions via collision induced dissociation (CID) and by measurement of collision cross section (CCS) in ion mobility (IM) experiments, respectively.[12] While structural mass spectrometry methods have been extensively developed and refined for the study of proteins, similar applications for higher order DNA assemblies are comparatively limited. [12,19,20]

Triplex formation and detection by MS can be impacted by several experimental factors including sequence, buffer composition, pH and ionization polarity. For further understanding of biological functionality and applications, it is important to ensure analysis of DNA triplexes represents physiologically relevant properties, rendering these various experimental conditions important factors to consider. Here, we evaluate a range of experimental factors to inform suitable conditions for DNA triplex investigation. We reveal new insights into the formation of these structures and showcase advantages of native MS for future analyses of higher order DNA assemblies.

EXPERIMENTAL SECTION

Sample Preparation

Oligonucleotides were purchased from Sigma at desalted purity and dissolved in deionised water (18 MΩ·cm^{-1}) to approximately 100 μM. The concentration of each oligonucleotide solution was determined by measuring the absorbance at 260 nm on a NanoDrop™ 2000 UV-Vis spectrophotometer (ThermoFisher Scientific). Each sample was measured in triplicate, using extinction coefficients as provided by the manufacturer. Ammonium acetate solutions were prepared from a 7.5 M aqueous stock and adjusted to the desired pH using either ammonium hydroxide or acetic acid (both > 99.99% purity). Dulbecco's PBS buffers were prepared from a 10× stock which contained $MgCl_2$ and $CaCl_2$ and were adjusted to the desired pH using sodium hydroxide. All reagents were purchased from Sigma.

Duplex DNA was prepared by mixing constituent oligonucleotides in the desired salt solution, to yield stock solutions at a concentration of 10 μM. These stocks were allowed to incubate for 1 hour at 4 °C to ensure duplex formation. Triplex samples were then prepared by adding the requisite amount of TFO to an aliquot of the duplex and diluting with the required buffer solution (DPBS or ammonium acetate) to yield a sample with a final triplex concentration of 1.5 μM. Samples were incubated at 4 °C for 1 hour to allow triplex formation prior to analysis. The TFO loading of a given sample was equimolar in all cases.

Native Mass Spectrometry

Mass spectra were principally acquired using an Agilent 6560 IM-Q-ToF spectrometer unless otherwise reported. The instrument was fitted with drift gas and air background upgrade kits and equipped with a nanospray ion source. Key parameters were as follows: V_{cap} 1050 V, Fragmentor 400 V, Gas Temp 50 °C , Gas Flow 10 L/min. N_2 was used as the drift gas. Samples were sprayed from platinum-coated borosilicate capillaries (prepared inhouse). The capillary housing was modified such that an N_2 gas line was connected to the back of the sample, enabling application of low-pressure gas flow to assist in the initiation and stabilization of sample spray. For CID experiments, collision cell energy was ramped from 0 – 44 V in 4 V increments, with spectra acquired at each increment for 30 s. MS data was analyzed using Masshunter qualitative analysis software (Agilent).

Additional experiments were performed on a Bruker Impact HD Q-ToF spectrometer, fitted with a nanospray ion source similarly modified such that a backing pressure of N_2 gas could be applied to the sample to stabilize ion spray. Key instrument parameters were as follows: End Plate Offset 500 V, Capillary 1050 V, Dry Gas 4.0 L·min^{-1}, Dry Temp 140 °C , Funnel 1 RF 400 VPP, Funnel 2 RF 600 VPP, Hexapole RF 300.0 VPP, Quadrupole Ion Energy 4.0 eV. Collision Cell parameters were: Collision RF 1200 VPP, Transfer Time 120 μs, Pre Pulse Storage 20.0 μs and Collision Cell Energy 10 eV. For CID experiments, the energy in the collision cell was ramped in 4 eV increments from the initial 10 eV through to 70 eV, with spectra acquired at each increment for 1 min. MS data was analyzed using Compass data analysis software (Bruker).

UV-Vis Spectroscopy

UV-Vis spectroscopy experiments were performed on a UV-3600i-Plus spectrophotometer (Shimadzu) equipped with a temperature controlled 8 Series Micro Multi-Cell Holder (MMC-1600C). Thermal melting experiments were conducted by ramping the temperature from 20 °C to 80 °C at a rate of 1 °C/min. The absorbance at 260 nm was recorded at every 0.5 °C interval. To limit evaporation and gas bubble formation during temperature ramping all samples were degassed by sonication for 5 min prior to analysis. The cells were sealed during each run with provided silicone stoppers

and as such, all UV-Vis melting experiments were conducted under constant volume conditions. All experiments were completed in triplicate and data shown is an average of the three runs, with derivative plots being smoothed with a Savitzky-Golay filter.

Isothermal titration calorimetry

Isothermal titration calorimetry data was collected on a MicroCal PEAQ-ITC isothermal titration calorimeter (Malvern). Prior to injection both the sample cell and the titrant solutions were degassed by sonication for 10 min and filtered via a 0.22 μm filter. To a mixture of duplex DNA in 250 mM ammonium acetate (20 μM, 300 μL) in a sample cell at 25 °C, the TFO (300 μM) was injected through a rotary syringe, which was rotating at 500 rpm. The interval time between each injection was 300 s for $(T)_{30}$ system and 600 s for the $(CT)_{15}$, to ensure that the system returned to equilibrium after each injection. The duration time of each injection was 2 s. Data points were collected after each injection and data was processed using Microcal PEAQ-ITC data analysis software (Malvern) and a 1:1 binding model in AFFINImeter-ITC advanced software (AFFINImeter).[21]

RESULTS AND DISCUSSION

Many factors contribute to the formation of DNA triplexes, including the composition of nucleotide bases within the DNA sequence.[10,22] For this study, three parallel Y-type triplexes with a TFO sequence of $(T)_{30}$, $(CT)_{15}$ and $(CCTT)_{7.5}$ were chosen as simple model systems to probe the behavior of triplexes. The calculated masses of the nucleic acid species observed are given as supporting information (Table S1). R type triplexes were not selected due to their reported inability to form under physiologically relevant conditions.[12] The $(T)_{30}$ sequence is a useful control system due to its pH independence, as protonation of thymine is not required to form Hoogsteen bonds (Figure 1B). The $(CT)_{15}$ and $(CCTT)_{7.5}$ sequences were selected to investigate how their base composition and base arrangements effect triplex formation, where protonation of TFO cytosines are required for the formation of Hoogsteen bonding.

Analysis of triplexes in the solution phase

UV-Vis spectroscopy and isothermal titration calorimetry (ITC) experiments were first performed to form a baseline for comparison to gas phase analysis and provide insight into the suitability of MS as an analytical tool for the study of triplex structures.

Understanding the behavior of DNA triplex systems in biologically relevant conditions is key to determining their functions *in vivo*. Dulbecco's phosphate buffered saline (DPBS) is used widely in the analysis of biological systems to mimic intracellular and extracellular environments. The presence of calcium and magnesium ions in these buffers are known to stabilize triplex DNA structures.[23,24] However, since these buffers are not compatible with MS analysis, it is important to have a comparison to the widely used volatile salt, ammonium acetate. A comparison of the electrostatic screening effects of these salt solutions at the concentrations employed here is given in supplementary data (Figure S1).

The two salt conditions, DPBS and ammonium acetate, were first compared for the $(T)_{30}$ and $(CT)_{15}$ systems by UV-melting experiments as seen in Figure 2, with the melting points provided in supplementary data (Tables S2 and S3). UV-Vis spectroscopy can identify the melting point of duplex and triplex DNA as indicated by a characteristic hyperchromicity in the melting curve at 260 nm, with triplex melting corresponding to the first transition, followed by melting of the more stable canonical duplex structure.[25] To aid the visualization and determination of the melting temperature, the absorbance is represented as a derivative with respect to temperature.

It is evident for the $(T)_{30}$ system when comparing common solution conditions of 1.0x DPBS and 250 mM ammonium acetate (Figures 2A and 2B) that the duplex has a lower melting point in DPBS (56 °C compared to 69 °C in ammonium acetate). Furthermore, triplex melting only occurs for this system in ammonium acetate (at 50 °C), which is again indicative of a lower stability for the DNA assemblies in DPBS. For both solution conditions, it is apparent that the stability and formation of duplex and triplex DNA is salt concentration dependent. Figure 2A shows that a decrease in the DPBS concentration also decreases the duplex melting temperature. This is also evident for the triplex as there are only indications of triplex melting in the highest 1.5x and 1.75x DBPS concentrations. Similar trends were observed for the ammonium acetate concentration range (Figure 2B), indicating that higher salt concentration significantly increases the stability of both triplex and duplex species. In each case melting occurs at lower temperatures in DPBS than in the ammonium acetate conditions. The formation of triplexes has a greater dependence on the salt concentration as shown by a larger increase in melting temperature when compared to the duplex melting temperatures.

To investigate the influence of cytosines on triplex stability in varying salt concentrations, UV-Vis melting profiles were also acquired for the $(CT)_{15}$ system (Figures 2D and 2E). In this case, very different behavior between DPBS and ammonium acetate solutions is observed. Here, no triplex melting events are identified at all in DPBS, which again indicates that ammonium acetate is comparatively stabilizing for triplex formation. This stabilization of triplex formation is more apparent in the cytosine-containing triplex system. However, the triplex melting temperatures were relatively consistent, only increasing from 50 °C to 52 °C across the salt concentration range of 50-350 mM (compared to $(T)_{30}$ triplex melting events increasing from 34 °C to 59 °C). Notably, in Figure 2E the triplex melting events are not symmetrical peaks, displaying a shoulder towards lower temperatures. This suggests a two-stage melting event or the presence of multiple species in the system, which cannot be distinguished using UV-Vis spectroscopy given it is an ensemble based technique.

Neighboring cytosine bases are known to be detrimental to triplex formation due their pH dependance, as adjacent protonated cytosine bases result in electrostatic

repulsion.[26,27] Hence not only the base composition but also their arrangement in the sequence can have drastic effects on triplex forming propensity.[22] This is evident from UV-Vis spectroscopy of the $(CCTT)_{7.5}$ sequence where no triplex was detected under any conditions (Supplementary data Figure S2). Given these results, this model system is not discussed further, however importantly highlights that the position of cytosines in particular is also an important factor to consider when identifying triplex target sequences.

Next, the pH dependence of triplex formation for both the $(T)_{30}$ and $(CT)_{15}$ systems was investigated across the salt concentration range by UV-Vis melting profiles (Figures 2C and 2F), with the melting temperatures provided as supplementary data (Table S1 and S2). The $(T)_{30}$ triplex system (Figure 2C) exhibits the same general trend in pH 5.5 and pH 7.4 conditions (Figure 2B), suggesting limited pH dependence for this sequence, although the melting temperatures are consistently lower across all measurements for the pH 5.5 conditions. In contrast, the $(CT)_{15}$ system (Figure 2F) showed an increase in the duplex and triplex melting temperatures under more acidic conditions (Figure 2E). This suggests a pH dependent increase in stability for assemblies formed by this sequence, as expected given the need to protonate some of the cytosine bases for triplex formation.

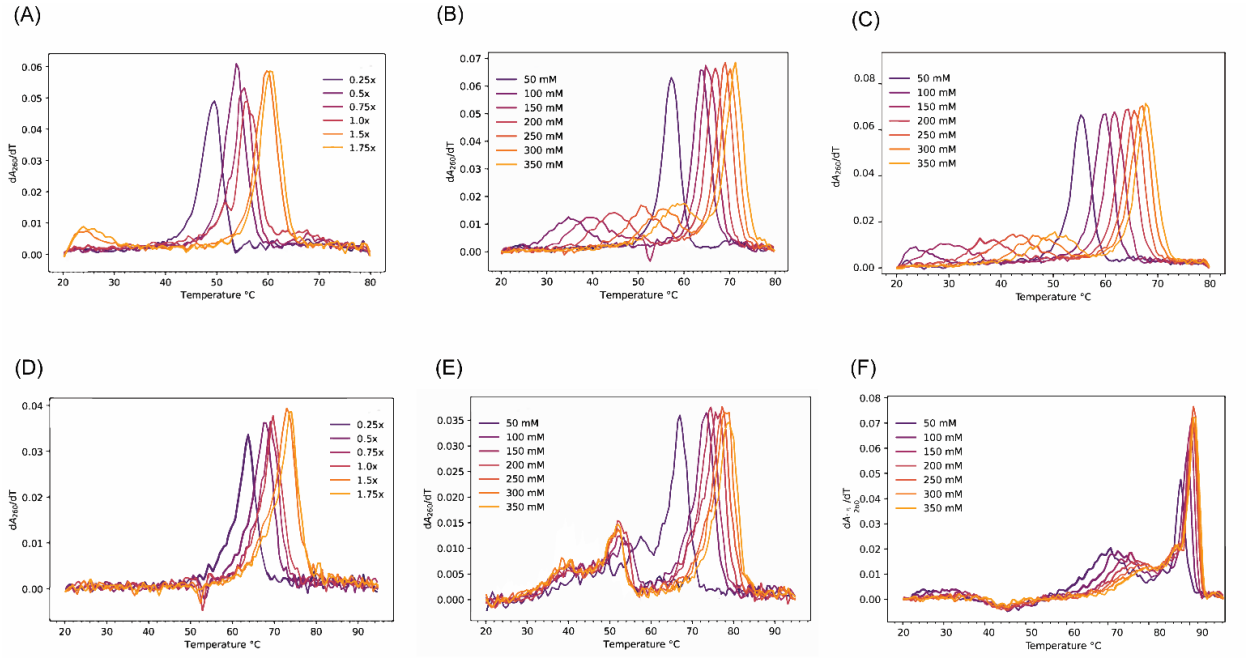


Figure 2. UV-Vis spectroscopy derivative melting profiles of $(T)_{30}$ DNA triplexes, in **(A)** DPBS at pH 7.4, **(B)** ammonium acetate at pH 7.4 and **(C)** ammonium acetate at pH 5.5 as a function of salt concentration. UV-Vis spectroscopy derivative melting profiles of $(CT)_{15}$ DNA triplexes, in **(D)** DPBS at pH 7.4, **(E)** ammonium acetate at pH 7.4 and **(F)** ammonium acetate at pH 5.5 as a function of salt concentration. All samples were baseline subtracted at 260 nm at 20 °C, and melting was conducted from 20 to 90 °C. Samples were prepared in a 1:1:1 TFO:purine:pyrimidine strand ratio, with each individual strand at a concentration of 1.5 µM.

Another solution phase method that can be used to determine the formation and stability of DNA triplexes is isothermal titration calorimetry (ITC). ITC can measure the thermodynamic and stoichiometric values of binding interactions by detecting the heat absorbed or released during the binding process. ITC has been used for the investigation of the thermodynamic properties of many different biological macromolecules, including interactions between higher order DNA structures, proteins and small molecules for drug screening.[28–30]

In order to obtain the thermodynamic properties for triplex formation by $(T)_{30}$ and $(CT)_{15}$ systems, the TFO was titrated into the duplex DNA sample. The change in heat of the system for each injection was recorded at pH 7.4 in 250 mM ammonium acetate conditions and the integrated enthalpy plot is shown as the molar ratio between duplex in the cell as the TFO titrant increases (Figures 3A and 3B).

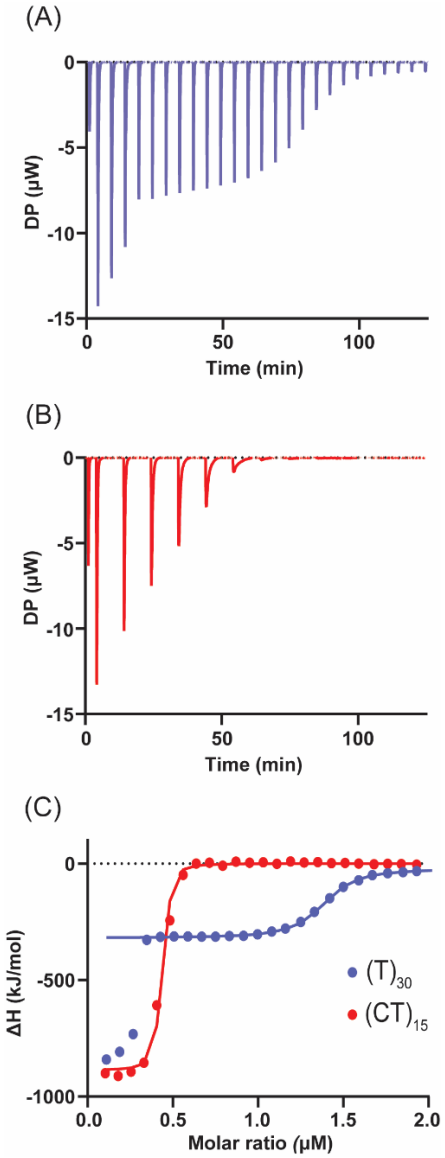


Figure 3. Isothermal titration calorimetry studies of $(T)_{30}$ and $(CT)_{15}$ triplex formation in pH 7.4 250 mM ammonium acetate at 25 °C. **(A)** and **(B)** show the changes following injection of the triplex forming oligonucleotide for $(T)_{30}$ and $(CT)_{15}$ respectively. **(C)** Plot of integrated enthalpies of each binding event in the $(T)_{30}$ and $(CT)_{15}$ systems, shown in blue and red respectively.

ITC data indicates a two-stage binding for the $(T)_{30}$ system in Figure 3A and 3C (blue). This is proposed to be due to incomplete duplex formation in initial conditions. Since the TFO is the same as the duplex pyrimidine strand in this system, initial titration first gives rise to saturation of the duplex DNA. Once duplex formation is complete, continued titration results in triplex formation, corresponding to the second binding event. In contrast, the $(CT)_{15}$ system only shows a single binding event corresponding to triplex formation since the TFO is frameshifted by one base pair to the duplex, resulting in preferential binding to the duplex. By fitting a simple 1:1 interaction model, the change in enthalpy (ΔH), Gibbs free energy (ΔG), entropy ΔS and dissociation constant (K_D) of the triplex formation can be estimated. The thermodynamic data from the ITC experiments for the formation of the two model triplexes can be seen in Table 1.

Table 1: Thermodynamic properties of triplex binding interactions at 25 °C for the two model triplex systems, $(T)_{30}$ and $(CT)_{15}$ in 250 mM ammonium acetate at pH 7.4.

Triplex	ΔH (kJ/mol)	ΔG (kJ/mol)	ΔS (kJ/mol/K)	K_D (M)
$(T)_{30}$	-293.85	-39.193	-1.118	1.35×10^{-7}
$(CT)_{15}$	-888.06	-44.126	-3.128	1.91×10^{-8}

From the thermodynamic data (Table 1), it is evident that the formation of the triplexes in the $(T)_{30}$ and $(CT)_{15}$ model systems is thermodynamically favored and spontaneous due to the negative ΔH and ΔG values respectively. The $(CT)_{15}$ triplex has more negative ΔH and ΔG values with a difference of 594.21 kJ/mol and 4.933 kJ/mol respectively, indicating that the formation of the $(CT)_{15}$ triplex is more favorable, which complements the UV-Vis spectroscopy experiments shown in Figure 2. Furthermore, the ΔS values indicate a more ordered triplex structure for the $(CT)_{15}$ sequence than the $(T)_{30}$.

While solution phase techniques can inform on the thermodynamic parameters and stability of DNA triplex systems, limitations arise when unresolvable species are present in the system. In such cases, and in order to further maximize our understanding of these higher order structures, we need to investigate the complexity and composition of these systems using MS.

Fundamental factors contributing to DNA triplex analysis by mass spectrometry

To evaluate the ability of MS to suitably report on biologically relevant properties of DNA triplexes, it is necessary to determine if the solution phase behaviors are reflected in the gas phase. This requires careful evaluation of a range of parameters in the MS experiment.

Salt concentration

As indicated by UV-Vis spectroscopy experiments (Figure 2), salt concentration has a stabilizing effect on both duplex and triplex structures in the $(T)_{30}$ system, and for the $(CT)_{15}$ duplex. Corresponding native MS data were collected for these two model systems across a range of ammonium acetate concentrations at pH 7.4 (Figure 4) to investigate whether a change in abundance of the duplex and triplex species is similarly observed using MS.

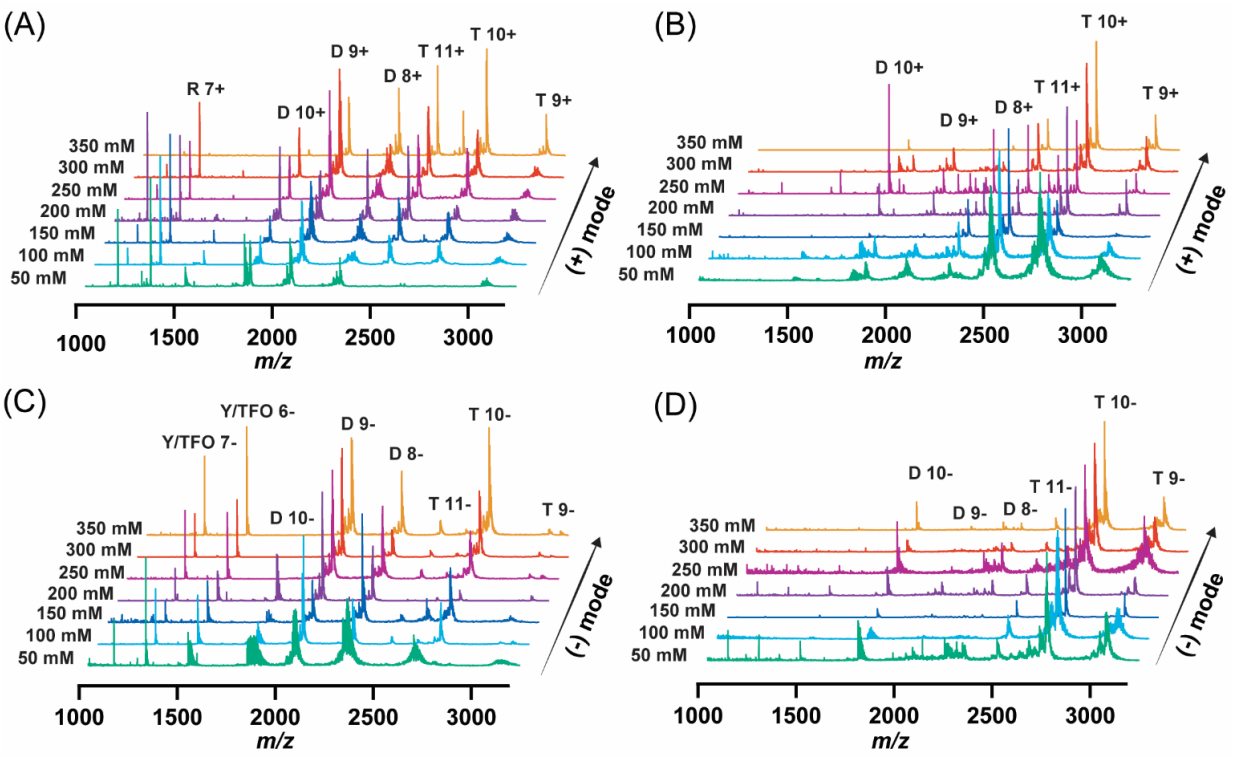


Figure 4. Positive mode mass spectra of **(A)** $(T)_{30}$ and **(B)** $(CT)_{15}$ DNA triplex systems, and negative mode mass spectra of **(C)** $(T)_{30}$ and **(D)** $(CT)_{15}$ DNA triplex systems. Samples were prepared in a 1:1:1 TFO:purine:pyrimidine strand ratio, with each individual strand at a concentration of 1.5 μM in ammonium acetate solution (pH 7.4) as a function of salt concentration as indicated. Annotations are as follows: R - single purine strand, Y/TFO - single pyrimidine strand or TFO (where the Y strand and TFO have same composition), D - duplex and T - triplex. Charge states are also indicated for selected peaks.

Across the ammonium acetate concentration range for the $(T)_{30}$ system at pH 7.4 (Figure 4A), the relative abundances of the triplex ions (predominantly at 10+ charge state) and duplex ions (predominantly 9+) generally increase, as expected. This increase is more significant for the triplex ions. This aligns with the UV-Vis melting profile from Figure 2B, showing slight increases in triplex and duplex melting temperature as the salt concentration is increased.

The effect of salt concentration in the pH 7.4 $(CT)_{15}$ system was also investigated in positive ion mode (Figure 4B). The 10+ triplex ions dominate the spectra for positive ion mode, with the abundance remaining consistent across the ammonium acetate concentration range, with significantly less duplex detected compared to triplex. This also agrees with the UV-Vis melting profiles (as seen in Figure 2E) which indicated that the melting temperature of triplex and duplex is relatively consistent across the salt concentration range. From this initial evaluation, the abundance of species observed in the MS spectra largely appears to reflect solution phase stability.

Ionization mode

The effect of ionization polarity is also of key interest when analyzing DNA assemblies. Positive ion mode is commonly used for native MS analysis of proteins and to detect modified nucleotides. However, given the highly anionic nature of DNA there is the potential for the electrospray process to produce ions that behave differently in the gas phase for each polarity, as has been demonstrated for some proteins and DNA complexes previously.[31,32]

In these experiments, there appears to be little difference in the ability of positive or negative ion mode to report on the effect of ammonium acetate concentration on triplex formation. As observed in positive mode, the negatively ionized triplex and duplex relative abundance increases with salt concentration in the $(T)_{30}$ system (Figure 4C) and remain largely unchanged across the salt concentration range for the $(CT)_{15}$ system (Figure 4D). Once again, this increase in relative abundance is more prevalent for the triplex ions.

For simplicity here we now focus on a comparison of positive and negative ion mode MS spectra for both the $(T)_{30}$ and the $(CT)_{15}$ systems in 250 mM ammonium acetate pH 7.4 only. A reproduction of this simplified data set from Figure 4 is provided as supplementary data (Figure S3) to enable ease of comparison.

The $(T)_{30}$ system shows that the general trends are reproduced, but there are subtle differences in the relative abundance of the species depending on the ionization mode selected (Figure S3A). The negative mode spectrum indicates the presence of significantly more unbound TFO compared to positive mode, which is consistent with the relative decrease in abundance of

the 11- triplex ion signal and increase in the 8- duplex ions observed in negative mode. MS data was also collected for the $(CT)_{15}$ system to compare the different ionization polarities. The positive and negative ion mode at 250 mM ammonium acetate (Figure S3B) spectra are mostly equivalent in terms of the relative abundance of the 10+ and 10- triplex ions in both polarities. However, in this case there is an increase in abundance of unbound TFO and duplex peaks in positive ion mode. These results reiterate the need to ensure any inferences made regarding the relative formation of different complexes should only be derived from data of the same polarity.

Of unexpected interest, the positive mode spectrum indicates the presence of unbound R strand in the $(T)_{30}$ system, which is not observed in negative mode. Given the same solutions were used for analysis in each case, this is a direct effect of the ionization polarity. The R strand is the central strand of the triplex structure and is effectively the limiting reagent for the formation of triplex. As such, for there to be free R strand this intuitively means the relative abundance of duplex and triplex observed in the spectra should be lower. However, this was not the case upon comparison of peak intensities (Figure S3).

To investigate the detection of the R strand in positive mode further, a collision induced dissociation (CID) experiment was conducted on the same $(T)_{30}$ system in both polarities (Supplementary data Figure S4). Upon CID of the triplex it is expected that the TFO is first ejected and detected along with remaining duplex. After CID of the $(T)_{30}$ triplex species, there is still no detection of the R strand in negative mode that is evident in positive mode. This is unexpected as there shouldn't be a difference in the ejected strands, since the R strand is the central strand of the triplex structure. As the R strand was still not detected with the use of CID, this indicates that there is a significant difference in the ionization efficiencies of the R strand in the different ionization modes.

This data demonstrates that care needs to be taken when analyzing DNA triplexes by MS due to the different species that are detected and their relative abundance in the different ionization modes.

pH conditions

Another fundamental factor that can play a role in triplex formation and detection is the pH of the buffer system. DNA triplexes that contain cytosine have been shown to have increased stability in lower pH conditions by UV-Vis spectroscopy and positive ion mode MS studies.[12] Furthermore, literature has shown that pH can influence analyte ionization efficiency in some cases, with certain compounds having an increased efficiency of more than two orders of magnitude when decreasing the pH.[33]

DNA triplexes for both the $(T)_{30}$ and the $(CT)_{15}$ model sequences were also formed in pH 5.5 conditions (Figure 5), to investigate the effect pH has on the observation of triplex assemblies in negative ion mode. Despite identifying the triplexes in both ionization modes, negative mode was chosen for the remainder of experiments due to the negatively charged nature of the DNA backbone. Supplementary data Figure S5 provides a simple comparison to earlier spectra obtained at pH 7.4. For the $(T)_{30}$ system (Figure 5A), decreasing the pH causes minor changes to the spectrum, with the 10- triplex species increased in relative abundance and the 8- duplex species decreased when compared to pH 7.4 (Figure S5A).

In contrast, there are significant changes in the abundance of duplex and triplex species for $(CT)_{15}$ assemblies when the buffer system is changed to pH 5.5. Of particular note, a more complex mass spectrum is observed (Figure 5B) with formation of additional unexpected assemblies. These include an R dimer and i-motif. I-motifs are another higher order DNA assembly with four strands that are held together by protonated and intercalated cytosine bases.[34] The 16- i-motif peak dominates the spectrum (Figure 5B) which is the cause of the significant decrease in relative abundance of the duplex and triplex ions.[35] Due to the low concentration (1.5 μM) of the DNA strands, it is improbable that these are non-specific multimers that associate during the desolvation process.[36,37] Instead, the i-motif species are formed under more acidic conditions due the protonation of cytosine bases, allowing for formation of the cytosine-cytosine base pairs that hold together the intercalated tetramer structure.

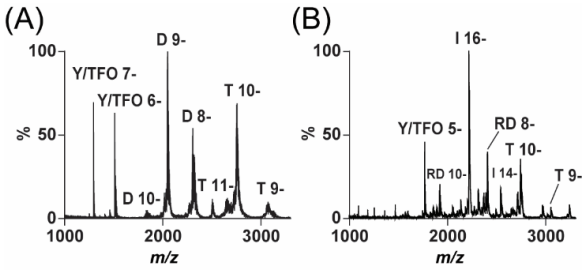


Figure 5. (A) Mass spectra of $(T)_{30}$ and **(B)** $(CT)_{15}$ DNA triplex systems, in 250 mM ammonium acetate solution at pH 5.5. All spectra were collected in negative ion mode. Samples were prepared in a 1:1:1 TFO:purine:pyrimidine strand ratio, with each individual strand at a concentration of 1.5 μM. Annotations are as follows: R-single purine strand, Y/TFO-single pyrimidine strand or TFO (Y strand and TFO have same molecular weight) D-duplex, T-triplex, RD-R strand dimer and I-i-motif, charge states are also indicated on labels.

The detection of these unusual species via MS shows the increased utility of MS, providing greater information regarding the composition of higher order assemblies compared to melting profiles from UV-Vis spectroscopy experiments where this detail is lost in the ensemble-based experiments.

Gas vs solution phase stability

Solution phase experiments such as UV-Vis spectroscopy are well-established methods that can determine the stability of DNA triplexes by defining

their melting points. However, given that we have demonstrated formation of additional higher order assemblies in various solution conditions, such solution phase ensemble methods may not be suitable for analysis in many conditions. We therefore sought to explore the potential of analogous gas-phase experiments to similarly reflect the stability of DNA triplexes.

A CID-MS experiment was performed on the two model systems (Figure 6) to compare the relative stability with that obtained from UV-Vis spectroscopy data (Figures 2B and 2D). When the collision energy is increased from 0 V to 20 V, no evidence of intact triplex remains in the $(T)_{30}$ system, however $(CT)_{15}$ has partially dissociated but is still detectable at 20 V. These results confirm that the $(CT)_{15}$ triplex is more stable than the $(T)_{30}$ as also shown by UV-Vis melting temperatures and ITC data (Figure 2B, 2D and Figures 3A-3C respectively).

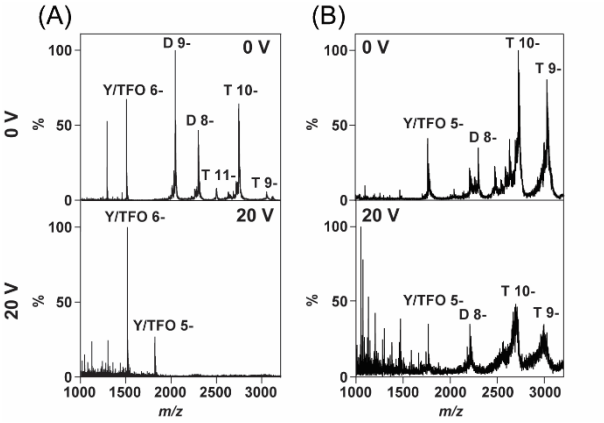


Figure 6: CID-MS spectra of **(A)** $(T)_{30}$ DNA triplex system at 0 V **(TOP)** and 20 V **(BOTTOM)**, and CID-MS spectra of **(B)** $(CT)_{15}$ DNA triplex system at 0 V **(TOP)** and 20 V **(BOTTOM)**, in 250 mM ammonium acetate solution at pH 7.4. All spectra were collected in negative ion mode. Samples were prepared in a 1:1:1 TFO:purine:pyrimidine strand ratio, with each individual strand at a concentration of 1.5 µM. Annotations are as follows: R-single purine strand, Y/TFO-single pyrimidine strand or TFO (Y strand and TFO have same molecular weight) D-duplex and T-triplex, charge states are also indicated on labels.

CONCLUSION

In this study, native MS was used to enhance the analysis of triplex DNA systems by exploring the fundamental experimental factors that influence the components detected in model triplex systems in the gas phase. MS was shown to reveal similar biophysical properties for the model DNA triplex structures as UV-Vis spectroscopy and ITC, whilst providing greater insight into the components of assemblies formed. For example, analysis under acidic conditions revealed the formation of multimers in the pH dependent sequence. Solution phase experiments are unable to directly inform on the formation of these species, demonstrating the necessity to use MS as a complementary tool to established techniques to fully understand assembly of triplex DNA systems.

From the UV-Vis spectroscopy analyses it is clear that, while ammonium acetate is a MS compatible solvent allowing for gas phase experiments, there is the potential for DNA triplex over-stabilization compared to common biological conditions. It is therefore important to consider buffer concentration and pH of the triplex systems, as this will allow further analysis of the formation and structural behaviors of DNA triplexes that are likely to occur in biology. We also confirm by MS that sequence specificity is an important factor to consider for triplex target sites, as and their specific arrangements in the sequence can influence the stability of the triplex, which is important to consider for *in vivo* applications. Finally, detection of the central purine triplex strand in positive ion mode for the $(T)_{30}$ system was a substantial difference between positive and negative mode which has not previously been addressed in literature. This indicates care must be taken in the MS experiment to ensure appropriate comparisons between systems are made. The use of both ionization polarities would also be beneficial to detect all species within a triplex system.

Overall, the development of more optimal conditions for MS analysis of triplexes allows for better understanding of the formation and structures of biologically relevant triplex systems, for potential uses in antigene targeting and the development of new classes of therapeutics.[38,39]

ASSOCIATED CONTENT

Supporting Information

Calculated mass for DNA species in $(T)_{30}$,$(CT)_{15}$ and $(CCTT)_{7.5}$ systems (Table S1)
Melting points of UV-Vis spectroscopy data from Figure 2 (Table S2 and S3)
Debye lengths for ammonium acetate and DPBS (Figure S1)
UV-Vis melting profile and MS data for $(CCTT)_{7.5}$ at pH 7.4 (Figure S2A and S2B)
Extracted MS data of $(T)_{30}$ and $(CT)_{15}$ in pH 7.4 250 mM ammonium acetate using positive and negative ion modes (Figure S3A and S3B)
MS/MS data of $(T)_{30}$ in positive and negative ion mode (Figure S4A and S4B)
Extracted MS data of $(T)_{30}$ and $(CT)_{15}$ in pH 5.5 and 7.4 250 mM ammonium acetate using negative ion modes (Figure S5A and S5B)

The Supporting Information is available free of charge on the ACS Publications website.

AUTHOR INFORMATION

Corresponding Author

*** Tara L. Pukala**- Discipline of Chemistry, The University of Adelaide, North Terrace, Adelaide SA 5005, Australia
Email: tara.pukala@adelaide.edu.au

Authors

‡**Jack W.Klose**- Discipline of Chemistry, The University of Adelaide, Adelaide, SA 5005, Australia

‡**Alexander J. Begbie**- Discipline of Chemistry, The University of Adelaide, Adelaide, SA 5005, Australia
Luis Toronjo-Urquiza- School of Chemical Engineering, The University of Adelaide, Adelaide, SA 5005, Australia

Author Contributions

Data was collected by A Begbie and J Klose. Data analysis was performed by all authors. The manuscript was written through contributions of J Klose and T Pukala. All authors have given approval to the final version of the manuscript. ‡J Klose and ‡A Begbie contributed equally.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research was supported by funding from the ARC Discovery Project grant DP200102291. J Klose would like to acknowledge a University of Adelaide Faculty of Sciences divisional scholarship for financial support throughout candidature. A Begbie would like to thank the Australian Government for financial support provided by the Research Training Program Scholarship throughout candidature. The authors would like to thank Prof. Robert Falconer of the University of Adelaide for access to ITC equipment. The authors would also like to express their thanks to Dr. Blagojce Jovcevski and Gerardo Urbina of the University of Adelaide for input in assistance with visualization of data.

REFERENCES

(1) Felsenfeld, G.; Davies, D. R.; Rich, A. Formation of a Three-Stranded Polynucleotide Molecule. *J. Am. Chem. Soc.* **1957**, *79* (8), 2023–2024.

(2) Goñi, J. R.; de la Cruz, X.; Orozco, M. Triplex-Forming Oligonucleotide Target Sequences in the Human Genome. *Nucleic Acids Res.* **2004**, *32* (1), 354–360.

(3) Hoyne, P. R.; Edwards, L. M.; Viari, A.; Maher, L. J. 3rd. Searching Genomes for Sequences with the Potential to Form Intrastrand Triple Helices. *J. Mol. Biol.* **2000**, *302* (4), 797–809. https://doi.org/10.1006/jmbi.2000.4502.

(4) Maldonado, R.; Filarsky, M.; Grummt, I.; Langst, G. Purine- and Pyrimidine-Triple-Helix-Forming Oligonucleotides Recognize Qualitatively Different Target Sites at the Ribosomal DNA Locus. *RNA* **2018**, *24* (3), 371–380.

(5) Jain, A.; Wang, G.; Vasquez, K. M. DNA Triplex Helices: Biological Consequences and Therapeutic Potential. *Biochimie* **2008**, *90* (8), 1117–1130.

(6) Sakamoto, N.; Chastain, P. D.; Parniewski, P.; Oshima, K.; Pandolfo, M.; Griffith, J. D.; Wells, R. D. Sticky DNA: Self-Association Properties of Long GAA·TTC Repeats in R·R·Y Triplex Structures from Friedreich's Ataxia. *Mol. Cell* **1999**, *3*, 465–475.

(7) Wells, R. D.; Collier, D. A.; Hanvey, J. C.; Shimizu, M.; Wohlrab. The Chemistry and Biology of Unusual DNA Structures Adopted by Oligopurine.Oligopyrimidine Sequences. *FASEB J.* **1988**, *2* (14), 2939–2949.

(8) Rusling, D.; Peng, G.; Srinivasan, N.; Fox, K.; Brown, T. DNA Triplex Formation with 5-Dimethylaminipropargyl Deoxyuridine. *Nucleic Acids Res.* **2009**, *37* (4), 1288–1296.

(9) Floris, R.; Scaggiante, B.; Manzini, G.; Quadrifoglio, F.; Xodo, L. Effect of Cations on Purine.Purine.Pyrimidine Triple Helix Formation in Mixed-Valence Salt Solutions. *Eur. J. Biochem.* **1999**, *260* (3), 801–809.

(10) Faucon, B.; Mergny, J.-L.; Helene, C. Effect of Third Strand Composition on the Triple Helix Formation: Purine versus Pyrimidine Oligodeoxynucleotides. *Nucleic Acids Res.* **1996**, *24* (16), 3181–3188.

(11) Dagneaux, C.; Gousset, H.; Shchyolkina, A. K.; Ouali, M.; Letellier, R.; Liquier, J.; Florentiev, V. L.; Taillandier, R. Parallel and Antiparallel A*A-T Intramolecular Triple Helices. *Nucleic Acids Res.* **1996**, *24* (22), 4506–4512.

(12) Li, J.; Begbie, A.; Boehm, B.; Whidborne, C.; Pouferis, Y.; Huang, D.; Pukala, T. Ion Mobility-Mass Spectrometry Reveals Details of Formation and Structure for GAA·TCC DNA and RNA Triplexes. *J. Am. Soc. Mass Spectrom.* **2019**, *30* (1), 103–112.

(13) Schrodinger LLC. The PyMOL Molecular Graphics System, Version 2.5.7. 2015.

(14) Jain, A.; Bacolla, A.; Chakraborty, P.; Grosse, F.; Vasquez, K. M. Human DHX9 Helicase Unwinds Triple-Helical DNA Structures. *Biochem* **2010**, *49* (33), 6992–6999.

(15) Akhter, Z.; Rajeswari, M. Triplex Forming Oligonucleotides Targeted to Hmga1 Selectively Inhibit Its Expression and Induce Apoptosis in Human Cervical Cancer. *J. Biomol. Struct. Dyn.* **2017**, *35* (4), 689–703.

(16) Martianov, I.; Ramadass, A.; Barros, A.-S.; Chow, N.; Akoulitchev, A. Repression of the Human Dihydrofolate Reductase Gene by a Non-Coding Interfering Transcript. *Nature* **2007**, *445*, 666–670.

(17) Zhou, Z.; Giles, K.; Felsenfeld, G. DNA·RNA Triple Helix Formation Can Function as a Cis-Acting Regulatory Mechanism at the Human β-Globin Locus. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (13), 6130–6139.

(18) Mcneer, N. A.; Schleifman, E. B.; Cuthbert, A.; Brehm, M.; Jackson, A.; Cheng, C.; Anandalingam, K.; Kumar, P.; Shultz, L. D.; Greiner, G. L.; Saltzman, W. M.; Glazer, P. M. Systemic Delivery of Triplex-Forming PNA and Donor DNA by Nanoparticles Mediates Site-Specific Genome Editing of Human Hematopoietic Cells in Vivo. *Gene Ther.* **2013**, *20*, 658–669.

(19) Mariappan, S. V.; Cheng, X.; Van Breeman, R. B.; Silks, L. A.; Gupta, G. Analysis of GAA/TTC DNA Triplexes Using Nuclear Magnetic Resonance and Electrospray Ionization Mass Spectrometry. *Anal. Biochem.* **2004**, *334* (2), 216–226.

(20) Arcella, A.; Portella, G.; Ruiz, M.; Eritja, R.; Vilaseca, M.; Gabelica, V.; Orozco, M. Structure of Triplex DNA in the Gas Phase. *J. Am. Chem. Soc.* **2012**, *134* (15), 6596–6606.

(21) Piñeiro, Á.; Muñoz, E.; Sabín, J.; Costas, M.; Bastos, M.; Velázquez-Campoy, A.; Garrido, P. F.; Dumas, P.; Ennifar, E.; García-Río, L.; Rial, J.; Pérez, D.; Fraga, P.; Rodríguez, A.; Cotelo, C. AFFINImeter: A Software to Analyze Molecular Recognition Processes from Experimental Data. *Anal. Biochem.* **2019**, *577*, 117–134.

(22) Mergny, J.-L.; Sun, J.-S.; Rougee, M.; Montenay-Garestier, T.; Barcelo, F.; Chomilier, J.; Helene, C. Sequence Specificity in Triple Helix Formation: Experimental and Theoretical Studies of the Effect of Mismatches on Triplex Stability. *Biochem* **1991**, *30* (40), 9791–9798.

(23) Šponer, J.; Sabat, M.; Burda, J. V; Doody, A. M.; Leszczynski, J.; Hobza, P. Stabilization of the Purine•Purine•Pyrimidine DNA Base Triplets by Divalent Metal Cations. *J. Biomol. Struct. Dyn.* **1998**, *16* (1), 139–143.

(24) Malkov, V. A.; Voloshin, O. N.; Soyfer, V. N.; Frank-Kamenetskii, M. D. Cation and Sequence Effects on Stability of Intermolecular Pyrimidine-Purine-Purine Triplex. *Nucleic Acids Res.* **1993**, *21* (3), 585–591.

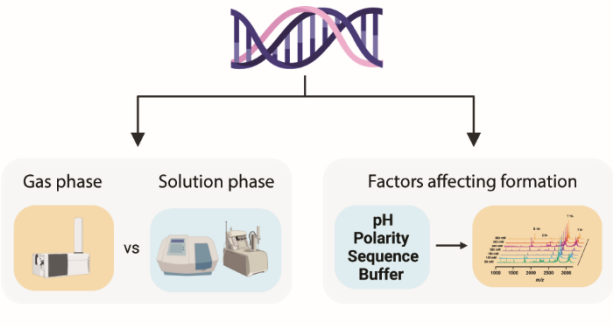
(25) Plum, G. E.; Park, Y.-W.; Singleton, S. F.; Dervan, P. B.; Breslauer, K. J. Thermodynamic Characterization of the Stability and the Melting Behavior of a DNA Triplex: A Spectroscopic and Calorimetric Study. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87* (23), 9436–9440.

(26) Leitner, D.; Schröder, W.; Weisz, K. Influence of Sequence-Dependent Cytosine Protonation and Methylation on DNA Triplex Stability. *Biochem* **2000**, *39* (19), 5886–5892.

(27) Vekhoff, P.; Ceccaldi, A.; Polverari, D.; Pylouster, J.; Pisano, C.; Arimondo, P. B. Triplex Formation on DNA Targets: How to Choose the Oligonucleotide. *Biochem* **2008**, *47*, 12277–12289.
(28) Doyle, M. Characterization of Binding Interactions by Isothermal Titration Calorimetry. *Curr. Opin. Biotechnol.* **1997**, *8* (1), 31–35.
(29) Rangel, V. M.; Gu, L.; Chen, G.; Chen, Q. H.; Xue, L. 5-Substituted 3, 3′, 4′, 7-Tetramethoxyflavonoids – A Novel Class of Potent DNA Triplex Specific Binding Ligands. *Bioorg. Med. Chem. Lett.* **2022**, *61*, 128608.
(30) Pagano, B.; Mattia, C. A.; Giancola, C. Applications of Isothermal Titration Calorimetry in Biophysical Studies of G-Quadruplexes. *Int. J. Mol. Sci.* **2009**, *10* (7), 2935–2957.
(31) Konermann, L.; Douglas, D. . Unfolding of Proteins Monitored by Electrospray Ionization Mass Spectrometry: A Comparison of Positive and Negative Ion Modes. *J. Am. Soc. Mass Spectrom.* **1998**, *9* (12), 1248–1254.
(32) Rosu, F.; Pirotte, S.; De Pauw, E.; Gabelica, V. Positive and Negative Ion Mode ESI-MS and MS/MS for Studying Drug-DNA Complexes. *Int. J. Mass Spectrom.* **2006**, *253* (3), 156–171.
(33) Liigand, J.; Laaniste, A.; Kruve, A. PH Effects on Electrospray Ionization Efficiency. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 461–469.
(34) Abou Assi, H.; Garavis, M.; González, C.; Damha, M. J. I-Motif DNA: Structual Features and Significance to Cell Biology. *Nucleic Acids Res.* **2018**, *46* (16), 8038–8056.
(35) Day, H. A.; Pavlou, P.; Waller, Z. A. E. I-Motif DNA: Structure, Stability and Targeting with Ligands. *Bioorg. Med. Chem.* **2014**, *22* (16), 4407–4418.
(36) Ding, J.; Anderegg, R. J. Specific and Non-Specific Dimer Formation in the Electrospray Ionisation Mass Spectrometry of Oligonucleotides. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 159–164.
(37) Lane, L. A.; Ruotolo, B. T.; Robinson, C. V; Favrin, G.; Benesch, J. L. A Monte Carlo Approach for Assessing the Specificity of Protein Oligomers Observed in Nano-Electrospray Mass Spectra. *Int. J. Mass Spectrom.* **2009**, *283*, 169–177.
(38) Praseuth, D.; Guieysse, A. L.; Helene, C. Triplex Helix Formation and the Antigene Strategy for Sequence-Specific Control of Gene Expression. *Biochim Biophys. Acta (BBA)- Gene Struct. Expr.* **1999**, *1489* (1), 181–206.
(39) Buske, F. A.; Mattick, J. S.; Bailey, T. L. Potential in Vivo Roles of Nucleic Acid Triple-Helices. *RNA Biol.* **2011**, *8* (3), 427–439. https://doi.org/10.4161/rna.8.3.14999.

Authors are required to submit a graphic entry for the Table of Contents (TOC) that, in conjunction with the manuscript title, should give the reader a representative idea of one of the following: A key structure, reaction, equation, concept, or theorem, etc., that is discussed in the manuscript. Consult the journal's Instructions for Authors for TOC graphic specifications.

Insert Table of Contents artwork here



3.3 Supplementary material

Native Mass Spectrometric Insights into the Formation and Stability of Model DNA Triplexes

Jack W. Klose[‡1], Alexander J. Begbie[‡1], Luis Toronjo-Urquiza[2], Tara L. Pukala[1]*

[1]Discipline of Chemistry, The University of Adelaide, 5005, Adelaide SA, Australia
[2]School of Chemical Engineering, The University of Adelaide, 5005, Adelaide SA, Australia

Supplementary material:

Table S1: Masses of the DNA species from the $(T)_{30}$, $(CT)_{15}$ and $(CCTT)_{7.5}$ systems

Sequence	Type	Calculated Average Molecular Weight (Da)
$(T)_{30}$	Pyrimidine DNA strand/TFO	9064
	Purine DNA strand	9334
	Duplex	18398
	Triplex	27462
$(CT)_{15}$	Pyrimidine DNA strand/TFO	8839
	Purine DNA strand	9574
	Duplex	18413
	Purine Dimer	19148
	Triplex	27252
	i-Motif	35356
$(CCTT)_{7.5}$	Pyrimidine DNA strand/TFO	8824
	Purine DNA strand	9590
	Duplex	18414
	Triplex	27238

Table S2: $(T)_{30}$ duplex and triplex melting temperatures from UV-Vis spectroscopy derivative melting curves (Figure 2A-2C) at different concentrations of DPBS and ammonium acetate at pH 5.5 and 7.4. Measurements with NP and NO indicate the experiment was not performed and the melting temperature was not observed in the experiment respectively.

Buffer concentration	pH 5.5		pH 7.4	
	Duplex melting temperature (°C)	Triplex melting temperature (°C)	Duplex melting temperature (°C)	Triplex melting temperature (°C)
0.25x DPBS	NP	NP	49	NO
0.5x DPBS	NP	NP	54	NO
0.75x DPBS	NP	NP	55	NO
1.0x DPBS	NP	NP	56	NO
1.5x DPBS	NP	NP	60	24
1.75 DPBS	NP	NP	61	25
50 mM Ammonium acetate	55	NO	57	NO
100 mM Ammonium acetate	60	24	64	34
150 mM Ammonium acetate	62	29	65	40
200 mM Ammonium acetate	64	36	67	45
250 mM Ammonium acetate	66	43	69	50
300 mM Ammonium acetate	67	46	70	55
350 mM Ammonium acetate	68	50	72	59

Table S3: $(CT)_{15}$ duplex and triplex melting temperatures from UV-Vis spectroscopy derivative melting curves (Figure 2D-2F) at different concentrations of DPBS and ammonium acetate at pH 5.5 and 7.4. Measurements with NP and NO indicate the experiment was not performed and the melting temperature was not observed in the experiment respectively.

Buffer concentration	pH 5.5		pH 7.4	
	Duplex melting temperature (°C)	Triplex melting temperature (°C)	Duplex melting temperature (°C)	Triplex melting temperature (°C)
0.25x DPBS	NP	NP	64	NO
0.5x DPBS	NP	NP	68	NO
0.75x DPBS	NP	NP	70	NO
1.0x DPBS	NP	NP	70	NO
1.5x DPBS	NP	NP	73	NO
1.75 DPBS	NP	NP	74	NO
50 mM Ammonium acetate	88	70	67	50
100 mM Ammonium acetate	89	72	73	53
150 mM Ammonium acetate	90	78	74	52
200 mM Ammonium acetate	90	79	75	52
250 mM Ammonium acetate	91	79	77	52
300 mM Ammonium acetate	91	80	78	52
350 mM Ammonium acetate	91	80	79	52

Variations in solution phase behavior between ammonium acetate and DPBS may be contextualized by comparing the electrostatic screening effects using the Debye length (K^{-1}). The Debye length can be determined using Equation 1 where $\in_r$ is the dielectric constant, k_B is the Boltzmann constant, T is absolute temperature, N_A is the Avogadro number and e is the elementary charge. The ionic strength (I) can be expressed in mol/m^{-3} and can be calculated using Equation 2 where c_i is concentration of each ion type in the electrolyte and z_i is the corresponding charge of the ion.

$$\kappa^{-1} = \sqrt{\frac{\in_r \in_0 k_B T}{2N_A e^2 I}} \quad (1)$$

$$I = \frac{1}{2}\sum_{i=1}^{n} c_i z_i^2 \quad (2)$$

Debye lengths at each salt concentration were calculated and shown in Figure S1. Across the salt concentration ranges the electrostatic screening effects can be chosen to be similar between both solutions making them comparable for the solution phase experiments. For example, 250 mM ammonium acetate has comparable electrostatic effects to 1.5x DPBS as they have similar Debye lengths.

Figure S1: Debye lengths for ammonium acetate (red) and DPBS (blue) solutions at the selected concentration ranges.

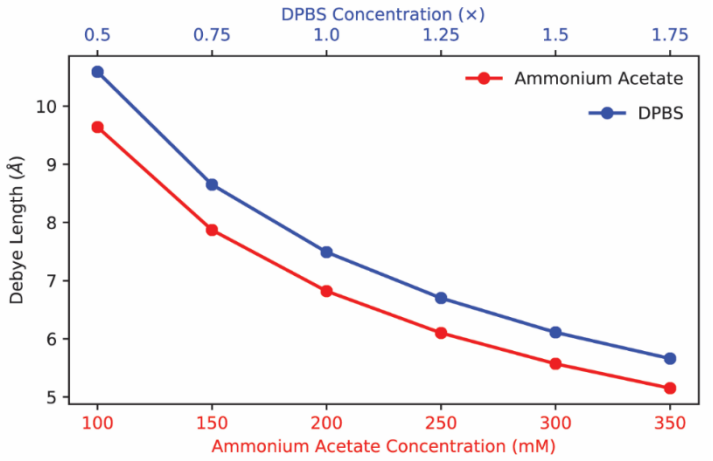


Figure S2: (A) UV-Vis spectroscopy derivative melting profiles of $(CCTT)_{7.5}$ triplex system in 250 mM ammonium acetate at pH 7.4. All samples were baseline subtracted at 260 nm at 20 °C, and melting was conducted from 20 to 90 °C. **(B)** Mass spectrum of the $(CCTT)_{7.5}$ triplex system in 250 mM ammonium acetate solution at pH 7.4. MS experiment was performed in negative ion mode. All samples for both experiments were prepared in a 1:1:1 TFO:purine:pyrimidine strand ratio, with each individual strand at a concentration of 1.5 µM. Annotations are as follows: Y/TFO - single pyrimidine strand or TFO (where the Y strand and TFO have same composition), D - duplex. Charge states are also indicated for selected peaks.

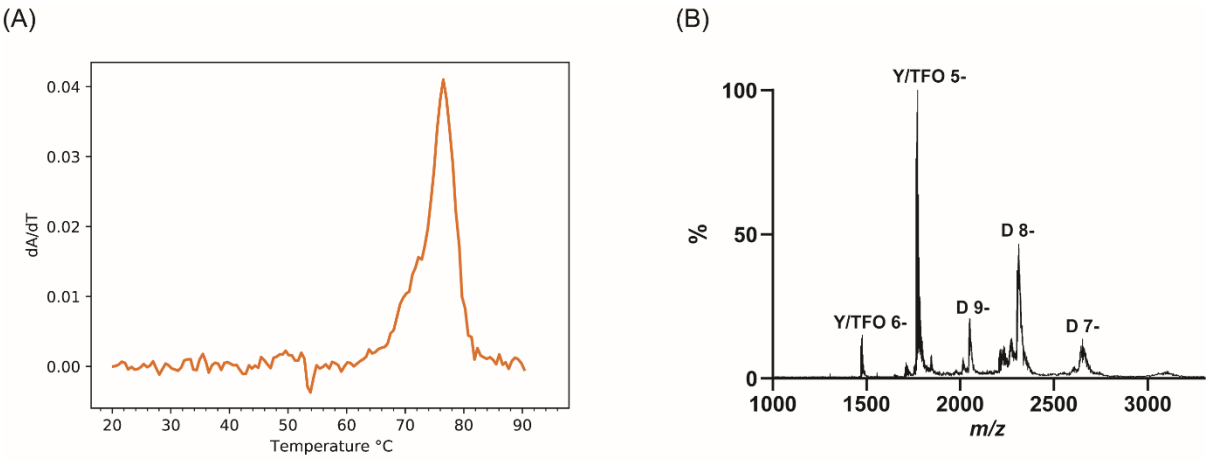


Figure S3. (A) Mass spectra of $(T)_{30}$ and **(B)** $(CT)_{15}$ DNA triplex systems, in 250 mM ammonium acetate solution at pH 7.4 in positive ion mode **(TOP)** and negative ion mode **(BOTTOM)**. Samples were prepared in a 1:1:1 TFO:purine:pyrimidine strand ratio, with each individual strand at a concentration of 1.5 μM. Annotations are as follows: R - single purine strand, Y/TFO - single pyrimidine strand or TFO (where the Y strand and TFO have same composition), D - duplex and T - triplex. Charge states are also indicated for selected peaks.

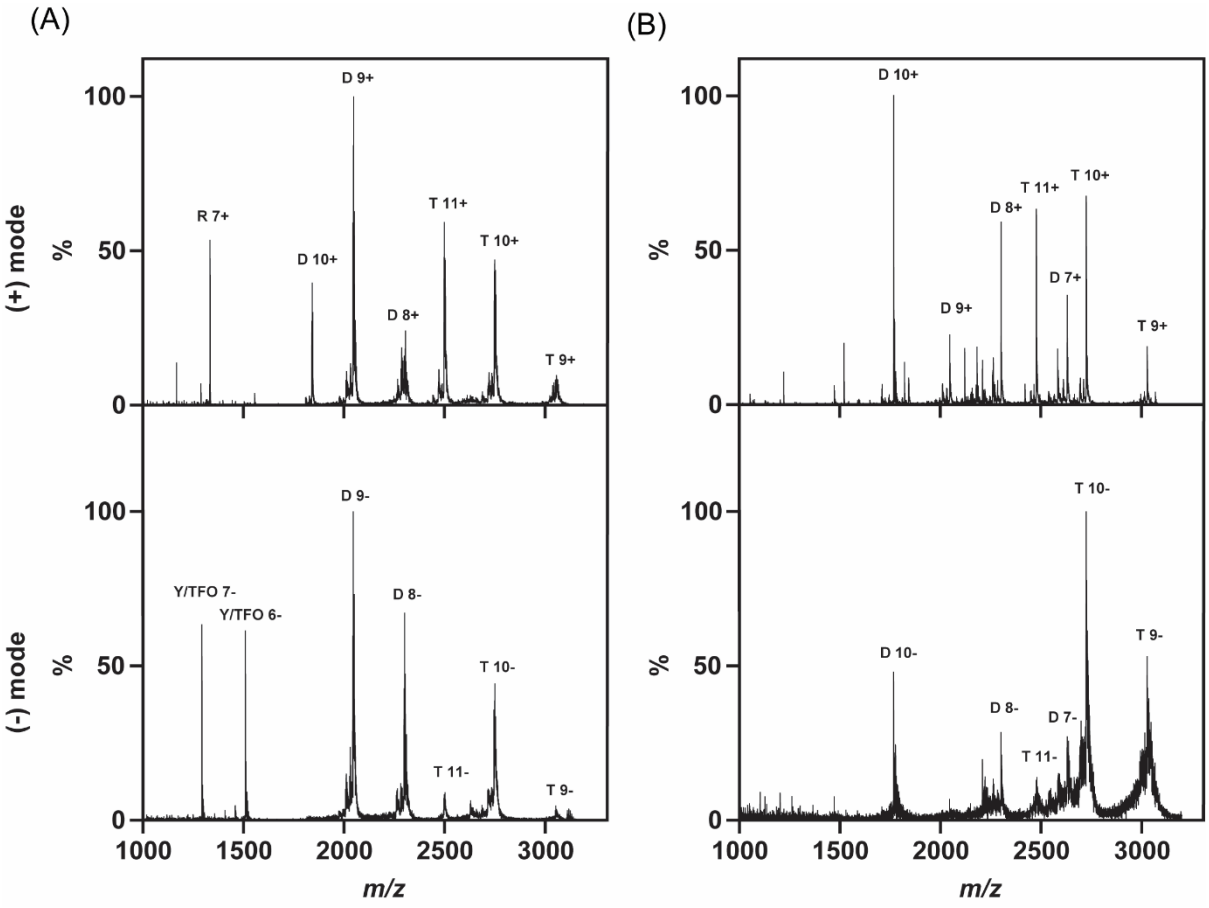


Figure S4. MS/MS spectra obtained from $(T)_{30}$ triplex peaks on a Bruker Impact HD mass spectrometer, from samples in 250 mM ammonium acetate at pH 7.4. **(A)** In positive mode the 11+ (2498 *m/z* peak was selected and fragmented at a collision energy of 50 eV. **(B)** In negative mode the 10- (2745 *m/z*) peak as selected and fragmented at a collision energy of 70 eV. Samples were prepared in a 1:1:1 TFO:purine:pyrimidine strand ratio, with each individual strand at a concentration of 1.5 μM. Annotations are as follows: R - single purine strand, Y/TFO - single pyrimidine strand or TFO (Y strand and TFO have same composition). Charge states are also indicated for selected peaks.

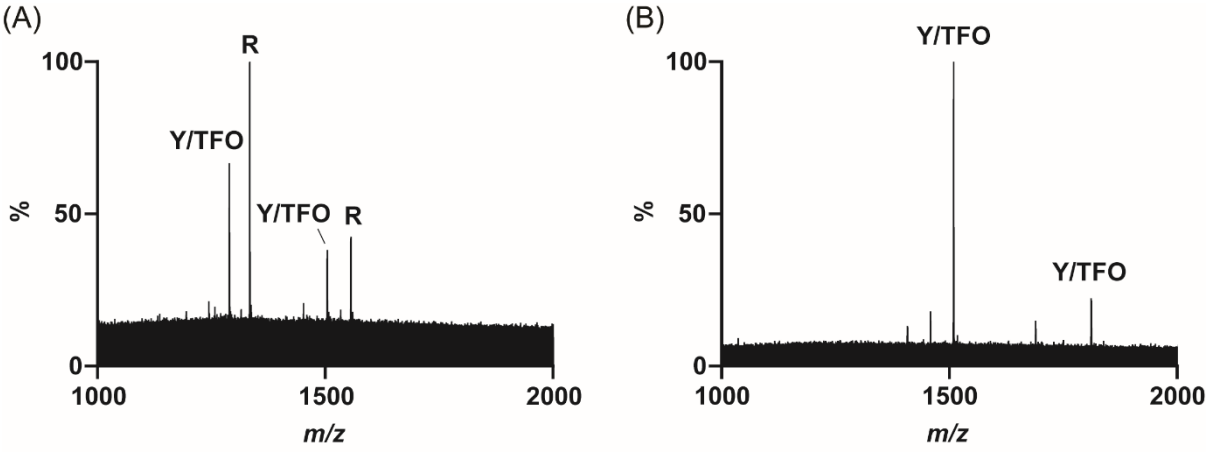
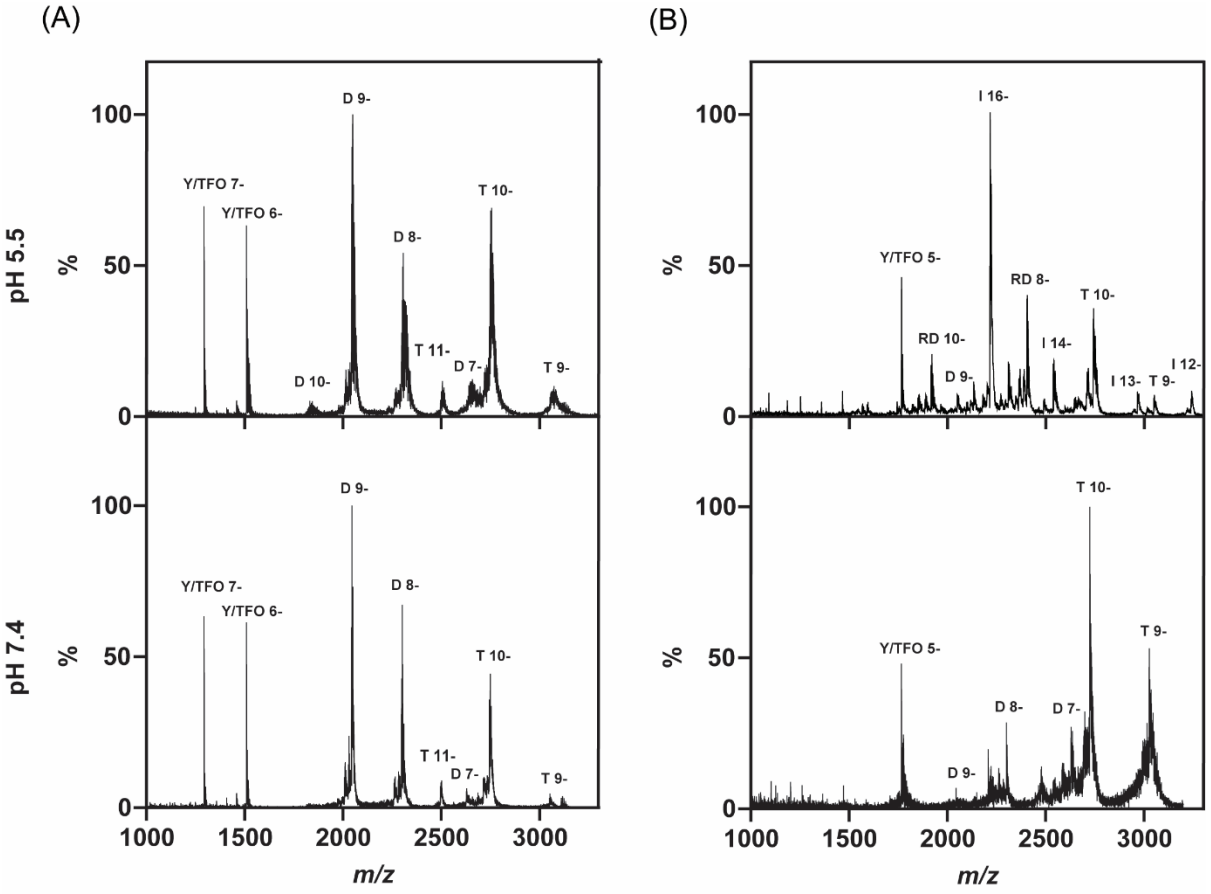


Figure S5. (A) Mass spectra of $(T)_{30}$ and **(B)** $(CT)_{15}$ DNA triplex systems, in 250 mM ammonium acetate solution at pH 5.5 **(TOP)** and pH 7.4 **(BOTTOM)**. All spectra were collected in negative ion mode. Samples were prepared in a 1:1:1 TFO:purine:pyrimidine strand ratio, with each individual strand at a concentration of 1.5 μM. Annotations are as follows: R - single purine strand, Y/TFO - single pyrimidine strand or TFO (Y strand and TFO have same composition) D - duplex, T - triplex, RD-R strand dimer and I - i-motif. Charge states are also indicated for selected peaks.



Chapter 4: Unraveling DNA Triplex Stability: Mass Spectrometric Investigation of Modified Triplex Forming Oligonucleotides for Enhanced Gene Targeting

4.1 Foreword and authorship statement

One of the issues with the formation of triplexes *in vivo* to target specific genes is the likelihood that a TTS that is long enough for stable triplex formation would appear within the promoter region of a critical gene of the target. Potential strategies to overcome this issue involve attempting to form DNA triplexes using a modified base to 'skip' pyrimidine interruptions in the polypurine TTS region. A dSpacer (2'-deoxyribonucleoside 3'-O-phosphoramidite) is a synthetically modified nucleoside that is used in DNA synthesis. These structures are unique in that they lack the nitrogenous base component and instead contain a non-nucleoside spacer molecule that takes its place. The introduction of a dSpacer creates an abasic site in the DNA strand to skip the interruption site and still stabilise the DNA structure through stacking interactions.[1] Even though there is a loss of specificity at this site, it is still useful for DNA structure and function determination experiments, as well as the introduction of different labels for the DNA which can be used for adding functionality, conjugation, and fluorescence chemistry. These modified bases can also be used to create FRET primers, investigate DNA repair, as well as enzyme inhibition and nucleosome catalysed reactions.[2–4]

Another base modification that could be used to stabilise triplexes is the locked nucleic acid (LNA). LNA is a synthetic nucleotide that contains a methylene that bridges the 2' oxygen and the 4' carbon of the ribose.[5] This methylene bridge increases the strength of base stacking interactions of LNA structure. This makes the LNA more rigid than a canonical DNA base, causing less distortion when binding to the duplex and increasing the stability of the triplex structure.[6,7] LNA containing TFOs have previously been investigated in targeting sequences related to *c-MYC* and frataxin (*FXN*) gene sequences and are a promising synthetic modification that can be potentially used in the antigene strategy.[8] In this chapter and associated supplementary chapter, dSpacer and LNA base modifications in the TFO will be

investigated in model and biological triplex systems in attempts to overcome a pyrimidine interruption in the duplex using native mass spectrometry and UV-Vis Spectroscopy.

Statement of Authorship

Title of Paper	Unraveling DNA Triplex Stability: Mass Spectrometric Investigation of Modified Triplex Forming Oligonucleotides for Enhanced Gene Targeting
Publication Status	☐ Published ☐ Accepted for Publication ☑ Unpublished and Unsubmitted work written in manuscript style ☐ Submitted for Publication
Publication Details	This manuscript is currently in preparation to be submitted to the Journal of the American Society for Mass Spectrometry

Principal Author

Name of Principal Author (Candidate)	Jack Klose
Contribution to the Paper	Data collection, data analysis, manuscript preparation
Overall percentage (%)	50
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.
Signature	Date: 18 Dec 2023

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);
ii. permission is granted for the candidate in include the publication in the thesis; and
iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Sarveenah Chandrasegaran
Contribution to the Paper	Data collection, data analysis, manuscript preparation
Signature	Date: 18 Dec 2023

Name of Co-Author	Tara Pukala
Contribution to the Paper	Project conceptualisation, methodology, data analysis, editing manuscript and presentation
Signature	Date: 20/12/2023

Please cut and paste additional co-author panels

4.2 Manuscript

Unraveling DNA Triplex Stability: Mass Spectrometric Investigation of Modified Triplex Forming Oligonucleotides for Enhanced Gene Targeting

Jack W. Klose[‡1], Sarveenah Chandrasegaran[‡1], Tara L. Pukala[1]*

[1]Discipline of Chemistry, The University of Adelaide, Adelaide, South Australia, 5005, Australia

ABSTRACT: Deoxyribonucleic acid (DNA) triplexes have potential roles in a range of biological processes involving gene and transcriptional regulation. A major challenge in exploiting the formation of these higher order structures to target genes *in vivo* however is their low stability, which is dependent on many factors including the length and composition of bases in the sequence. Here, different DNA base modifications have been explored, primarily using native mass spectrometry, in efforts to enable stronger binding between the triplex forming oligonucleotide (TFO) and duplex target sites. These modifications can also be used to overcome pyrimidine interruptions in the duplex sequence in promoter regions of genomes, to expand triplex target sequences for antigene therapies. Using model sequences with a single pyrimidine interruption, TFOs containing locked nucleic acid (LNA) base modifications were shown to have a higher triplex binding propensity than DNA only and dSpacer-containing TFOs. However, the triplex forming ability of these systems was limited by the formation of multiple higher order assemblies. Triplex-forming sequences that correspond to specific gene targets from the *Pseudomonas aeruginosa* genome were also investigated, with LNA-containing TFOs the only variant able to form triplex using these sequences. This indicates the advantages of utilizing synthetically modified TFOs to form triplex assemblies *in vivo* for potential therapeutic applications and highlights the advantages of native mass spectrometry for the study of their formation.

INTRODUCTION

In recent years, the use of oligonucleotides to control gene expression has rapidly developed, with antisense and antigene oligonucleotides utilized for this purpose.[1] Antisense oligonucleotides are targeted to bind messenger ribonucleic acid (mRNA), thereby influencing the expression of specific proteins.[2-5] On the other hand, the antigene or triple helix strategy targets distinct regions of genomic deoxyribonucleic acid (DNA) by designing triplex forming oligonucleotides (TFO) that bind to the gene.[1] Consequently, the triple helix strategy offers several advantages over antisense therapy, including modulation of gene expression in a more effective and long-lasting manner, reducing transcription of the gene at a higher rate.[1, 5-7]

To develop new antigene strategies, a better understanding of triplex DNA is needed. Triplex formation occurs by binding a third oligonucleotide strand (the TFO) in the duplex major groove, utilizing exposed hydrogen bond sites in this region to form Hoogsteen bonds.[7, 8] Hoogsteen bonded bases can be parallel or antiparallel, forming between A-T and G-C or A-A and G-G, respectively.[8] Therefore, the requirement of purine sites for Hoogsteen bonds gives rise to two variants of TFOs; R-type (purine) and Y-type (pyrimidine). The two possible resulting triplex combinations are in turn represented as R*R•Y (antiparallel) and Y*R•Y (parallel), where the dot indicates Watson-Crick hydrogen bonding between duplex base pairs, and the * represents Hoogsteen bonds stabilizing the triplex.[9, 10]

DNA triplexes are generally unfavorable due to electrostatic repulsion and steric hindrance arising from introduction of the third strand in the duplex structure.[9, 10] To achieve a stable triplex structure, uninterrupted polypurine stretches are essential.[11-13] However, in biology, it is more likely to find pyrimidine interruptions in long polypurine sequences due to the variability of the bases in a genome. DNA triplex formation has been previously shown to be notably destabilized by one or two pyrimidine bases in the triplex target site (TTS).[14]

In pursuit of novel antigene approaches, a wide array of nucleotide chemical modifications have been synthesized, aiming to form more stable triplex structures *in vivo*. Synthetic oligonucleotides are frequently generated through alteration at the base, sugar and phosphate sites.[15-17] Nucleobase modifications have included altered hydrophobic groups, chemical handles, carbohydrates, and even the introduction of artificial base-pair matches to enhance binding.[15, 18] At the phosphate sites, sulfur atoms or uncharged methyl groups can be introduced to enhance the *in vivo* stability, however, these modifications are recognized for their potential to increase non-nucleotide target binding.[15, 18] On the other hand, an alternative method for enhancing stability is by modifying the oligonucleotide backbone.[2, 15, 18]

DNA-RNA hybrid structures have been investigated under *in vivo* conditions for comparison with DNA-only assemblies.[19, 20] The presence of the 2'OH in RNA is proposed to stabilize its conformation by sterically constraining the flexibility of the sugar ring during H-DNA formation.[21] In 1997, the locked nucleic acid (LNA) structure was introduced, which rapidly replaced traditional oligonucleotides such as DNA and RNA in antigene technologies due to its further enhanced

binding.[22, 23] The increased rigidity of the LNA structure introduced by the methylene bridge (Figure 1) enhances the local organization of the phosphate backbone, consequently strengthening base stacking interactions.[22, 24, 25] This requires less distortion of the duplex strands when it binds as the TFO, resulting in increased stability and thus abundant formation of the triplex.[24, 25] The substantial enhancement in triplex stability implies the potential for designing more efficacious TFOs with therapeutic applications.[2]

dSpacers have also been considered in attempts to overcome pyrimidine interruptions in triple helical assemblies.[11] A dSpacer, also known as an abasic site, consists of a DNA backbone without the corresponding base (Figure 1D). Incorporating a dSpacer ensures the production of a TFO without necessitating an uninterrupted polypurine TTS. This approach maintains that the oligonucleotide has an intact sugar structure while preserving the conformation of the triplex DNA.

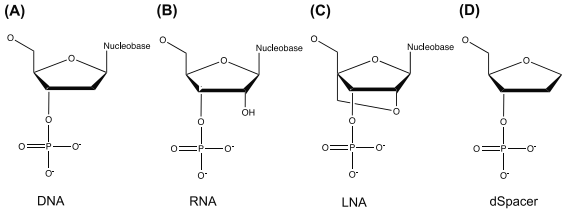


Figure 1. The chemical structure of (A) DNA, (B) RNA, (C) LNA and (D) dSpacer units. RNA has a hydroxyl group attached to the pentose ring at the 2' position compared to DNA. In LNA, the 2'oxygen and 4' carbon atoms are connected by a methylene bridge, locking the overall conformation. In the dSpacer the nucleobase is not present.

To date, there have been limited systematic studies on the optimal design of LNA-containing antigene molecules. Therefore, in this study, the formation and stability of various modified TFOs in a range of model and biologically relevant triplex sequences containing pyrimidine base interruptions were investigated using native electrospray ionization mass spectrometry (ESI-MS) in combination with UV-vis spectroscopy. Utilizing LNA- and dSpacer-containing TFOs, triplex forming propensity was compared to gain a more comprehensive understanding of the influence of pyrimidine interruptions on the formation of triple helical assemblies.

MATERIALS AND METHODS

Reagents

Ammonium acetate buffers were prepared from a stock 7.5 M ammonium acetate solution and pH balanced with ammonium hydroxide (99.99% purity) as necessary, all purchased from Sigma (Castle Hill, Australia).

Bioinformatics

Genomic data for *Pseudomonas aeruginosa* (NC_002516.2) were obtained from National Center for Biotechnology Information database resources (accessed 9 March 2022). TTSs were identified using Triplexator data analysis software,[26] with length set from 15 to 30 base pairs with up to 2 consecutive 'errors' (pyrimidine interruptions).

Once TTSs were identified for the bacteria, the relevant protein gene product and its functions were obtained from UniProtKB/Swiss-Prot (accessed January 2022). Proteins with a crucial function for bacterial survival were chosen for further investigation, and full gene sequences were obtained from UniProtKB. Then, using the BLAST search engine, a local alignment of the potential triplex sites in the gene was obtained.

Triplex formation

Reagents and oligonucleotides were purchased from Sigma (Castle Hill, Australia). unless otherwise specified. The oligonucleotides were commercially obtained at 0.2 μmole scales and typically purified by desalting. LNA containing oligonucleotides were commercially obtained at 1.0 μmole scales and were purified by high-performance liquid chromatography (HPLC). The concentrations of the single strand oligonucleotides were measured in triplicate using a ThermoFisher NanoDrop 2000 spectrometer based on absorbance at 260 nm and the extinction coefficient given by the manufacturer. Oligonucleotide strands were diluted to 1 mM using MilliQ water. Duplex and triplex DNA was prepared using 1 mM single stranded oligonucleotide stocks and diluted to 20 μM in 250 mM ammonium acetate at pH 7.4. Duplex strands were incubated in a 1:1 molar ratio at 4 °C for 2 hours. Then, where required, the TFO was added to the mixture in a 1:1:1 molar ratio and incubated at 4 °C for 2 hours. Samples were then further diluted to 2 μM in 250 mM ammonium acetate for analysis.

Ultraviolet-visible spectroscopy

DNA samples were prepared as described above and diluted to 2 μM in 250 mM of ammonium acetate for UV-vis melting experiments. Samples were degassed by sonication for 10 minutes prior to analysis to ensure there were limited gas bubbles present in the samples. All spectra were recorded on a Shimadzu UV-3600i-Plus spectrophotometer integrated with a temperature-controlled 8 Series Micro Multi-Cell Holder (MMC-1600C). The thermal melting experiment was conducted at 260 nm or 300 nm at pH 7.4 as indicated in the text. Starting at 20 °C, the temperature was ramped to 90 °C at a rate of 1 °C.min^{-1}. The absorbance was recorded every 0.5 °C. The cell was sealed during each run with the provided silicone stopper to reduce the evaporation of the sample when the temperature was increased. Experiments were conducted in triplicates and averaged data were used for analysis. The first derivative relative to the temperature was used to determine the melting temperature of the triplex and duplex DNA structures.

Mass spectrometry

Mass spectra were recorded using a Bruker Impact HD Q-TOF spectrometer, fitted with a nanospray ion source.

Ion spray was stabilized using N_2 gas as the backing pressure. Data were collected in positive mode using nanoESI from platinum-coated borosilicate capillaries prepared in-house. Key instrumental parameters were: capillary voltage 1700 V, gas temperature 140 °C, gas flow rate 4.0 L/min, nebulizer gas pressure 1.0 Bar, CID energy of 35 eV.

RESULTS AND DISCUSSION

dSpacer and LNA base modifications in model triplex systems

Initially, a model parallel triplex system was designed based on an $(AG)_{15}$ polypurine stretch, where a central adenine was replaced by a cytosine in the triplex target site. This sequence was investigated to compare the competitive binding processes associated with the formation of duplex and triplex species due to the repetitive nature of the sequence. Canonical DNA- (SEQ(C)), dSpacer- (SEQ(C)-d) and LNA- (SEQ(C)-L) containing TFOs were then compared to determine their effect on the triplex forming propensity of these systems. Full sequences, masses and location of base modifications are provided in supplementary data Table S1.

The extent of formation for the SEQ(C) triplexes systems with DNA-, dSpacer- and LNA-containing TFOs are shown by MS in Figure 2A-2C respectively. In all cases the relative abundance of triplex is significantly lower than that of the duplex. However, for both the SEQ(C)-d and SEQ(C)-L systems, multiple duplex and triplex species are observed, indicating greater complexity in the binding interactions between strands.

For the SEQ(C)-d triplex system, two different duplex structures are detected (Figure 2B). The expected DNA duplex has been formed, along with a duplex structure where the SEQ(C)-d TFO has bound to the complementary purine strand. The formation of two duplexes impacts the triplex forming propensity in the system as there is less dSpacer TFO available to bind with the DNA duplex. This in turn also gives rise to multiple triplex species forming. Firstly, a triplex corresponding to excess single pyrimidine strand binding to the duplex (SEQ(C)-Py) was observed. In this case, the intended dSpacer TFO is not a constituent strand in the assembly. The formation of the secondary dSpacer duplex allowed the generation of a triplex structure with the dSpacer TFO (SEQ(C)-d-d). Triplex formation with the canonical DNA duplex and SEQ(C)-d TFO was also apparent, however, it is evident from the multiple structures formed in this system that the dSpacer base modification is not able to form a specific triplex structure with improved stability over a DNA-containing TFO.

In comparison, for the SEQ(C)-L system (Figure 2C), SEQ(C)-Py triplex was observed, again indicative of competitive binding. A secondary duplex structure was also observed that corresponds to the SEQ(C)-L TFO binding to the single purine strand, analogous to the dSpacer equivalent (Figure 2B). In this case however, the SEQ(C)-L triplex formed with a higher relative abundance than the SEQ(C)-Py triplex, making the LNA base modification more favorable for overcoming the pyrimidine interruption in the TTS than the dSpacer. This is expected due to the rigid nature of LNA, as it prevents conformational rearrangement of the nucleotides as the TFO binds, therefore facilitating interaction in the major groove of the duplex with a lower entropic penalty.[27]

Expanding on this investigation to determine if the choice of pyrimidine in the interruption site can also influence the formation of the multiple duplex and triplex species, an additional model system was generated where thymine was instead introduced as the pyrimidine interruption in the triplex target site. Analogous DNA (SEQ(T)), dSpacer (SEQ(T)-d) and LNA (SEQ(T)-L) TFOs were then compared for their ability to form triplex structures via MS (Figure 3), with full sequences provided in Table S2.

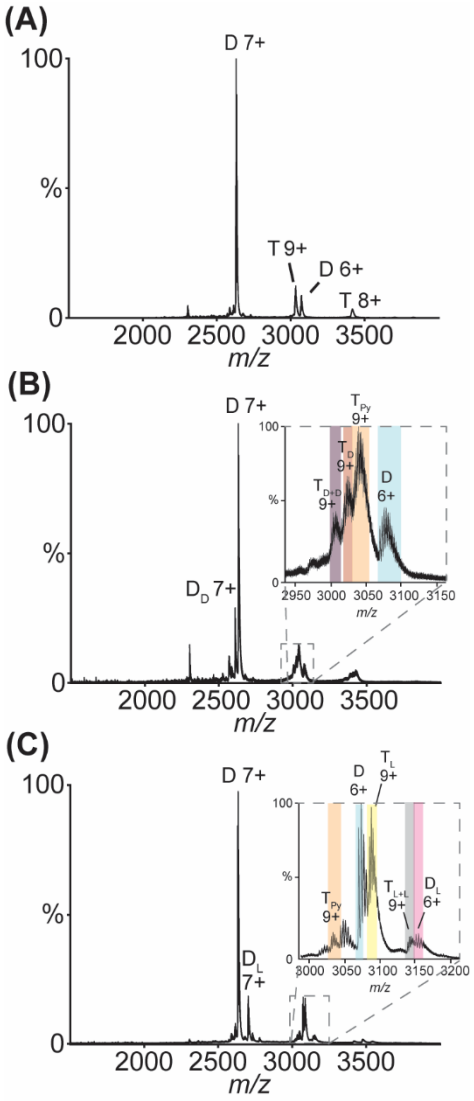


Figure 2: LNA and dSpacer modifications in the TFO alter duplex and triplex forming propensity when a

cytosine interruption is present. Mass spectra of **(A)** SEQ(C), **(B)** SEQ(C)-d and **(C)** SEQ(C)-L triplex systems, in 250 mM ammonium acetate solution at pH 7.4. All spectra were collected in positive ion mode. Samples were prepared in a 1:1:1 TFO:purine:pyrimidine strand ratio, with each strand at a concentration of 2 μM. Nucleic acid species are labelled as follows: SEQ(C) duplex:D, SEQ(C) DNA triplex:T, SEQ(C)-d duplex with dSpacer strand:D_D, SEQ(C) pyrimidine triplex:T_{Py}, SEQ(C)-d triplex:T_D, SEQ(C)-d-d triplex:T_{D+D}, SEQ(C)-L triplex:T_L and SEQ(C)-L-L triplex: T_{L+L}. Inset in (B) highlights species including T_{D+D}, T_D,T_{Py} and D in purple, red, orange and blue respectively. Inset in (C) highlights species including T_{Py}, D, T_L, T_{L+L} and D_L in orange, blue, yellow, grey and pink respectively. All charge states are indicated on labels.

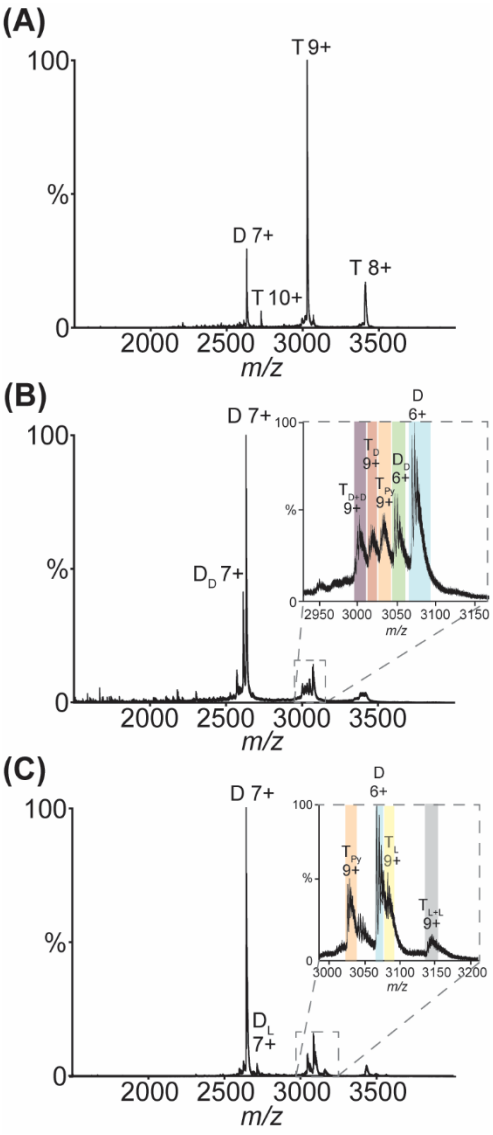


Figure 3: LNA and dSpacer modifications in the TFO alter duplex and triplex forming propensity when a thymine interruption is present. Mass spectra of **(A)** SEQ(T), **(B)** SEQ(T)-d and **(C)** SEQ(T)-L triplex systems, in 250 mM ammonium acetate solution at pH 7.4. All spectra were collected in positive ion mode. Samples were prepared in a 1:1:1 TFO:purine:pyrimidine strand ratio, with each strand at a concentration of 2 μM. Nucleic acid species are labelled as follows: SEQ(T) duplex:D, SEQ(T) DNA triplex:T_{DNA}, SEQ(T)-d duplex with dSpacer strand:D_D, SEQ(T) Pyrimidine triplex:T_{Py}, SEQ(T)-d-d triplex:T_{D+D}, SEQ(T)-L Duplex:D_L, SEQ(T)-L triplex:T_L and SEQ(T)-L-L triplex:T_{L+L}. Inset in (B) highlights species including T_{D+D}, T_D, T_{Py}, D_D and D in purple, red, orange, green and blue respectively. Inset in (C) highlights species including T_{Py}, D, T_L, and T_{L+L} in orange, blue, yellow and grey respectively. All charge states are indicated on labels.

Interestingly, Figure 3A shows that the SEQ(T) DNA-only TFO can bind to the duplex easily and overcome the thymine interruption, as the triplex ions dominate the MS spectrum. This represents a significantly higher triplex forming propensity compared to the analogous SEQ(C) triplex (Figure 2A), indicating that the ability to overcome pyrimidine interruptions is also dependent on the pyrimidine base present.

Formation of two duplexes was again observed upon addition of the dSpacer TFO, corresponding to competitive binding of the TFO and purine strand in the SEQ(T) system (Figure 3B). Consistent with the analogous cytosine interruption, non-specific triplex species were again observed, including the SEQ(T)-Py and the SEQ(T)-d-d assemblies. The formation of these additional non-specific assemblies has an impact on the ability of the dSpacer-containing TFO to bind to the duplex, as there is less TFO readily available. Nevertheless, the d-spacer containing triplex species are again observed to be lower in relative abundance to the DNA-only assemblies. Comparable assemblies are also observed upon addition of the SEQ(T)-L TFO (Figure 3C), although in this case the LNA containing triplexes appear in slightly greater relative abundance, once again indicating the stabilizing effect of this modification.

It is evident from these initial experiments that sequence specificity is very important when investigating the use of base modifications to overcome pyrimidine interruptions in TTSs. The formation of the multiple duplex and triplex species observed here is due to the presence of repetitive AG sequences in the model. Although this complicates the *in vitro* analysis somewhat, repeating sequences have been implicated in the neurodegenerative disease Friedreich's ataxia where the TTS contains more than 100 GAA trinucleotide repeats, and therefore does have implications in targeting such sites in biology.[28] Nevertheless, focus was next directed to more biologically relevant triplex forming sequences that do not contain any symmetrical nucleotide repeats, in an attempt to form more stable triplex assemblies *in vitro* using the dSpacer and LNA base modifications.

Triplex target sites in *Pseudomonas aeruginosa*

To further develop TFOs in an antigene strategy, a biological system is necessary. According to the World Health Organisation (WHO), *Pseudomonas aeruginosa* is a bacterium with high resistance to current antibiotics and poses the highest risk to humans.[29] To identify potential targets for an antigene strategy, the genome of *Pseudomonas aeruginosa* was obtained from the National Center for Biotechnology Information database and searched for the positions of polypurine stretches in the double-stranded nucleotide sequences.[30] The long uninterrupted polypurine stretches were initially identified as potential TTSs using Triplexator.[26] The importance of allowing for pyrimidine interruptions becomes evident here, as restricting the number of interruptions (i.e., non-purine sites) to zero resulted in the identification of 176 potential triplex target sites. However, despite the abundance of potential targets, these are not located in genes relevant to essential biological functions. When the length is shorter than 15-mer length there are again more potential triplex forming sites. However, evidence has shown that a minimum length of 15 nucleotides is needed to form a stable triplex structure.[7] This demonstrates that pyrimidine interruptions are prominent in nature, and thus it is necessary to overcome this limitation in order to expand the scope of triplex forming antigene targets.

Potential TTSs with both 0 and 1 interruptions were next analyzed to determine the function of the corresponding gene products. The goal was to ensure that these genes played essential roles in bacterial survival by regulating transcription. From a pool of 4088 potential triplex sites within the bacterial genome, two sites from targeted genes, *fimV* and *xcpU*, were selected for further investigation, as seen in Table 1. These sites were chosen based on the essential functions of the gene products in the bacterial system, and due to their lower occurrence of tandem repeats within the sequence. Of the chosen target genes, Q9HZA6 (*fimV*) (Table 1) in *Pseudomonas aeruginosa* plays a crucial role in various virulence mechanisms, including type IV pilus (T4P) mediated assembly and virulence gene expression.[33, 34] The other selected target, Q00515 (*xcpU*) (Table 1), is responsible for components of the type II secretion system, which is essential for the energy-dependent secretion of extracellular factors such as proteases and toxins from the periplasm.[35, 36] The sequence with the pyrimidine interruption was of particular interest as previous studies have indicated that interruptions in the middle of the TTS impose the maximum destabilizing effects.[31, 32] Thus, this sequence was chosen as the 'worst-case' scenario for a pyrimidine interruption.

Table 1. Selected gene targets for candidate triplex sites from the genomes of *Pseudomonas aeruginosa*. The pyrimidine interruption is highlighted in bold in the triplex target sites.

Bacterial species	Target gene	Triplex target site (5'—3')	Length (Bases)
Pseudomonas aeruginosa	fimV	AGAAAGAGAA GGAAGAAG	18
	xcpU	GAGAAGGAGC AGAAGAAAGG	20

Investigating the optimal number of LNA bases in the TFO

The *fimV* triplex sequence does not contain any pyrimidine interruptions and was first selected to investigate how triplex binding is influenced by the number of LNAs present. Here, the number of LNA bases in the TFO varied from 7 (*fimV*-7L), 5 (*fimV*-5L) and 3 (*fimV*-3L), systematically increasing the gap between LNA substitutions. The equivalent DNA-only TFO (*fimV*) and LNA containing fimV triplexes were analyzed using MS (Figure 4A-4D). The full sequence and positioning of LNA base modifications are provided in supporting information (Table S3).

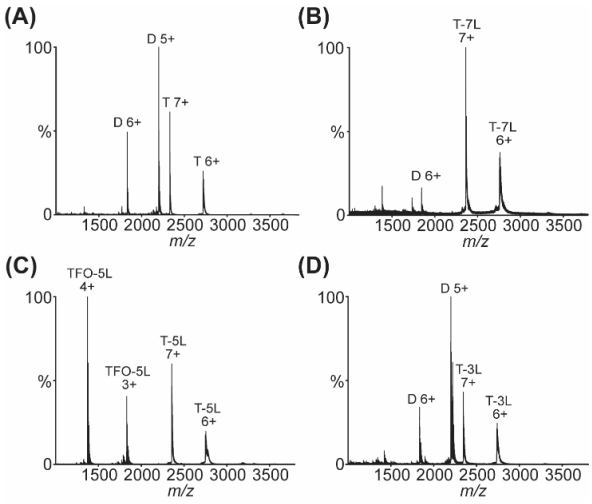


Figure 4: The amount of LNA bases in the TFO is proportional to the relative abundance of triplex formed. Mass spectra of **(A)** *fimV*, **(B)** *fimV*-7L, **(C)** *fimV*-5L and **(D)** *fimV*-3L triplex systems, in 250 mM ammonium acetate solution at pH 7.4. All spectra were collected in positive ion mode. Samples were prepared in a 1:1:1 TFO:purine:pyrimidine strand ratio, with each strand at a concentration of 2 μM. Nucleic acid species are labelled as follows: *fimV* duplex:D, *fimV*-7L triplex:T-7L, *fimV*-5L triplex:T-5L, *fimV*-5L TFO:TFO-5L and *fimV*-3L triplex:T-3L.

In contrast to the model sequence, only a single triplex species is observed in the *fimV* system. This is expected since there is no tandem repeat in the biological sequence, eliminating multiple binding possibilities and hence improving binding specificity. When comparing the relative abundance of the duplex and triplex formation using ESI-MS, fimV-7L has the lowest duplex and unbound TFO remaining when compared with the *fimV*, *fimV*-5L, and *fimV*-3L triplex system. Based on

Figure 4, it is evident that using a TFO with alternating LNA bases was the most favorable for triplex formation.

The stability of the triplex systems was further investigated using UV-Vis spectroscopy to report on the solution phase properties and further validate MS results (Figure 5). The melting temperature of the triplex with *fimV* as the third strand was measured to be 38 °C. Based on the results presented in Figure 4A, it was hypothesized that the triplex formed using LNA-containing TFOs was stabilized to the extent that the duplex and the triplex melting transitions overlapped. To validate this hypothesis, the melting profile was also analyzed at 300 nm. Previous studies have shown that at 300 nm, triplex melting is evident due to hypochromicity associated with partial cytosine deprotonation upon dissociation of the third strand, while at 260 nm, both duplex and triplex complexes absorb due to the hyperchromic effect upon melting.[37]

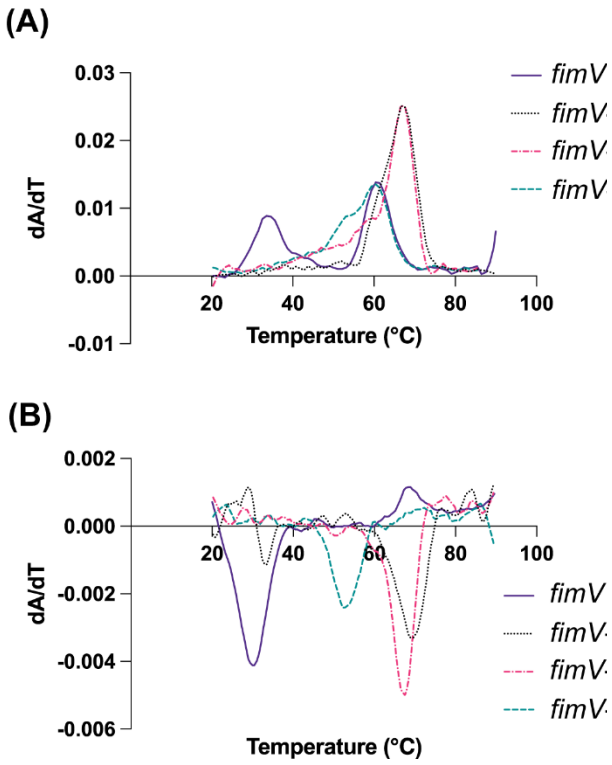


Figure 5: Increasing the amount of LNA bases also increases the stability of the triplex structure. The derivative melting profiles of *fimV* (purple solid line), *fimV*-7L (black dotted line), *fimV*-5L (pink dotdash line) and *fimV*-3L (green dashed line) sequences at **(A)** 260 and **(B)** 300 nm for 30-mer biological sequence. Samples were prepared in a 1:1:1 TFO:purine:pyrimidine strand ratio, with each strand at a concentration of 2 µM in 250 µM of ammonium acetate solution at pH 7.4.

The 300 nm melting profile of *fimV* shows a negative differential absorbance observed at 30 °C (Figure 5B), consistent with the T_m measured at 260 nm (Figure 5A). However, when the sample with *fimV*-7L, *fimV*-5L and *fimV*-3L systems were analyzed at 300 nm, negative differential absorbance was observed at 69 °C, 67 °C and 52 °C, respectively (Figure 5B). This verified the presence of the triplex in the samples and confirmed the overlap between the triplex and duplex melting temperatures at 260 nm (Figure 5A). This finding not only clearly demonstrates that *fimV*-7L forms the most stable triplex, but also indicates that reducing the number of LNA bases in the TFO has a direct detrimental impact on reducing the stability of the triplex. The UV-Vis spectroscopy analysis is consistent with the MS data as the intensity of the triplex species was also the highest with *fimV*-7L. Together, this suggests that having alternating LNA substitutions has the greatest stabilizing effect in forming a triplex, in fact to a greater extent than expected, with melting temperatures comparable to a duplex. Due to this increased stabilization effect of LNA/DNA hybrid strands, an interruption in the triplex target sites was subsequently attempted to overcome by substituting the LNA into the TFO.

Analysis of the *Pseudomonas aeruginosa* triplex sequence with pyrimidine base interruption

The *xcpU* gene sequence from the *Pseudomonas aeruginosa* genome was next selected for analysis as it contains a cytosine interruption in the TTS (sequence information is provided in Table S4).

The relative abundance of duplex and triplex formation was compared using the canonical DNA- (*xcpU*), dSpacer- (*xcpU*-d) and LNA- (*xcpU*-L) containing TFOs using ESI-MS (Figure 6). The data confirms that the only system to result in triplex formation was *xcpU*-L (Figure 6C). This demonstrated that alternating LNA substitutions within the TFO stabilizes the triplex, even in the presence of a pyrimidine interruption.

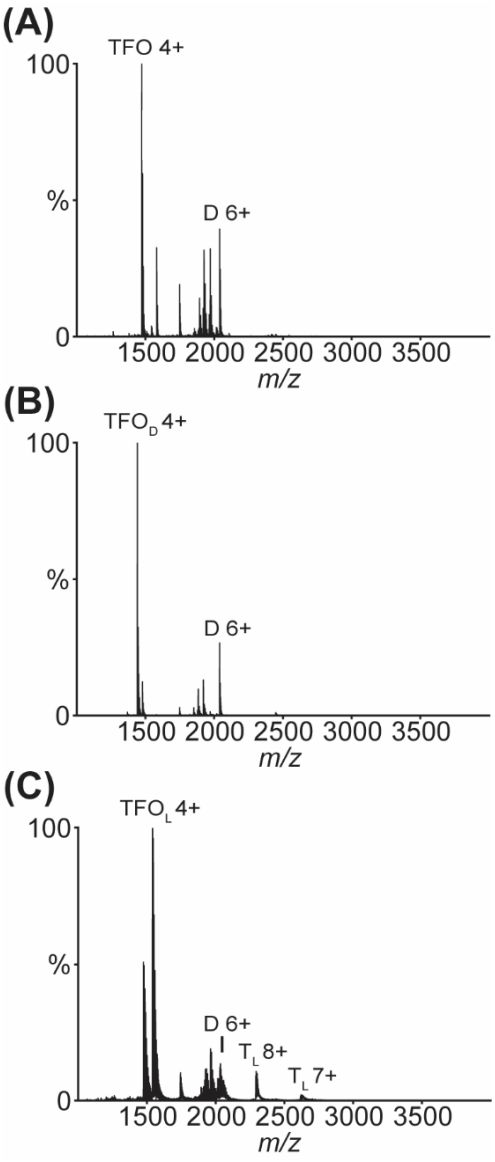


Figure 6: LNA base modification in the TFO promotes triplex formation in biological systems. Mass spectra of **(A)** *xcpU*, **(B)** *xcpU*-d and **(C)** *xcpU*-L triplex systems, in 250 mM ammonium acetate solution at pH 7.4. All spectra were collected in positive ion mode. Samples were prepared in a 1:1:1 TFO:purine:pyrimidine strand ratio, with each strand at a concentration of 2 µM. Nucleic acid species are labelled as follows: xcpU TFO: TFO *xcpU* duplex:D, *xcpU*-d TFO:TFO$_D$, *xcpU*-L TFO:TFO$_L$ and *xcpU*-L triplex:T$_L$.

UV-Vis melting profiles were also obtained for the three *xcpU* systems. Only *xcpU*-L showed evidence of triplex formation, with a melting transition at 37 °C (Figure S1). Based on this, LNA is the superior base modification that can be used to overcome detrimental pyrimidine interruptions in the triplex target site and form a more stable triplex structure *in vivo*.

CONCLUSION

Using native MS, this study demonstrates incorporating alternating LNA base modifications in TFOs to target pyrimidine interruption containing TTSs enhances triplex formation compared to employing dSpacer or non-complementary canonical DNA bases. Notably, the modified triplexes exhibited susceptibility to non-specific duplex formation in both dSpacer and LNA systems, underscoring the suboptimal preference for symmetrical sequences in these analyses. MS, as opposed to thermal stability experiments, facilitates a comprehensive understanding of the diverse species generated in these synthetically modified triplex systems.

This investigation further underscores the challenges associated with forming biologically relevant triplexes featuring a pyrimidine interruption. LNA-modified TFOs effectively surmounted this obstacle, with an escalating incorporation of LNA bases in the TFO correlating with an increased propensity for triplex formation. Further exploration into diverse synthetic modifications holds promise for enhancing the *in vivo* stabilization of triplexes, opening avenues for broader biological applications. This includes the potential for transcriptional regulation of targeted genes and the development of novel classes of antibiotics, highlighting the significance of advancing research in this domain.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

Sequences for the SEQ(C), SEQ(C)-d and SEQ(C)-L triplex systems (Table S1)
Sequences for the SEQ(T), SEQ(T)-d and SEQ(T)-L triplex systems (Table S2)
Sequences for the *fimV* triplex systems (Table S3)
Sequences for the *xcpU* triplex systems (Table S4)
UV-Vis spectroscopy data of the *xcpU*, *xcpU*-d and *xcpU*-d triplex systems (Figure S1)

AUTHOR INFORMATION

Corresponding Author

***Tara L. Pukala**- Discipline of Chemistry, The University of Adelaide, North Terrace, Adelaide SA 5005, Australia
Email: tara.pukala@adelaide.edu.au

Present Addresses

‡**Jack W. Klose**- Discipline of Chemistry, The University of Adelaide, Adelaide, SA 5005, Australia
‡**Sarveenah Chandrasegaran**- Discipline of Chemistry, The University of Adelaide, Adelaide, SA 5005, Australia

Author Contributions

Data was collected by J Klose and S Chandrasegaran. Data analysis was performed by all authors. The

manuscript was written through contributions of all authors. All authors have approved the final version of the manuscript. ‡J Klose and ‡S Chandrasegaran contributed equally to this work.

Notes

Any additional relevant notes should be placed here.

ACKNOWLEDGMENT

This research was supported by funding from the Australian Research Council Discovery Project grant DP200102291. J Klose would like to acknowledge receipt of a University of Adelaide Faculty of Sciences divisional scholarship for financial support throughout candidature. S Chandrasegaran acknowledges a University of Adelaide International scholarship for financial support throughout candidature.

REFERENCES

(1) Miller, P. S. Development of Antisense and Antigene Oligonucleotide Analogs. In *Progress in Nucleic Acid Research and Molecular Biology*, Vol. 52; **1996**; pp 261-291.

(2) Jolly, P.; Estrela, P.; Ladomery, M. Oligonucleotide-based systems: DNA, microRNAs, DNA/RNA aptamers. *Essays Biochem.* **2016**, *60* (1), pp 27-35. DOI: 10.1042/ebc20150004.

(3) Chan, J. H.; Lim, S.; Wong, W. F. Antisense oligonucleotides: from design to therapeutic application. *Clin. Exp. Pharmacol.* **2006**, *33* (5-6), pp 533-540.

(4) Faria, M.; Wood, C.; Perrouault, L.; Nelson, J.; Winter, A.; White, M.; Helene, C.; Giovannangeli, C. Targeted inhibition of transcription elongation in cells mediated by triplex-forming oligonucleotides. *Proc. Natl. Acad. Sci.* **2000**, *97* (8), pp 3862-3867.

(5) Bailey, C.; Weeks, D. L. Understanding oligonucleotide-mediated inhibition of gene expression in Xenopus laevis oocytes. *Nucleic Acids Res.* **2000**, *28* (5), pp 1154-1161.

(6) Duca, M.; Vekhoff, P.; Oussedik, K.; Halby, L.; Arimondo, P. B. The triple helix: 50 years later, the outcome. *Nucleic Acids Res.* **2008**, *36* (16), pp 5123-5138. DOI: 10.1093/nar/gkn493.

(7) Jain, A.; Wang, G.; Vasquez, K. M. DNA triple helices: biological consequences and therapeutic potential. *Biochimie.* **2008**, *90* (8), pp 1117-1130. DOI: 10.1016/j.biochi.2008.02.011.

(8) Fox, K. R.; Brown, T.; Rusling, D. A. DNA Recognition by Parallel Triplex Formation. In *DNA-targeting Molecules as Therapeutic Agents*, **2018**.

(9) Maldonado, R.; Filarsky, M.; Grummt, I.; Längst, G. Purine- and pyrimidine-triple-helix-forming oligonucleotides recognize qualitatively different target sites at the ribosomal DNA locus. *RNA* **2018**, *24* (3), pp 371-380. DOI: 10.1261/rna.063800.117.

(10) Frank-Kamenetskii, M. D.; Mirkin, S. M. TRIPLEX DNA STRUCTURES. *Annu. Rev. Biochem.* **1995**, *64* (1), pp 65-95. DOI: 10.1146/annurev.bi.64.070195.000433.

(11) Sági, J.; Guliaev, A. B.; Singer, B. 15-mer DNA Duplexes Containing an Abasic Site Are Thermodynamically More Stable with Adjacent Purines Than with Pyrimidines. *Biochem.* **2001**, *40* (13), pp 3859-3868. DOI: 10.1021/bi0024409.

(12) Vekhoff, P.; Ceccaldi, A.; Polverari, D.; Pylouster, J.; Pisano, C.; Arimondo, P. B. Triplex Formation on DNA Targets: How To Choose the Oligonucleotide. *Biochem.* **2008**, *47* (47), pp 12277-12289. DOI: 10.1021/bi801087g.

(13) Goñi, J. R.; Vaquerizas, J. M.; Dopazo, J.; Orozco, M. Exploring the reasons for the large density of triplex-forming oligonucleotide target sequences in the human regulatory regions. *BMC Genom.* **2006**, *7* (1), pp 63. DOI: 10.1186/1471-2164-7-63.

(14) Kukreti, S.; Sun, J. S.; Garestier, T.; Hélène, C. Extension of the range of DNA sequences available for triple helix formation: stabilization of mismatched triplexes by acridine-containing oligonucleotides. *Nucleic Acids Res.* **1997**, *25* (21), pp 4264-4270. DOI: 10.1093/nar/25.21.4264.

(15) Ma, D. D. F.; Rede, T.; Naqvi, N. A.; Cook, P. D. Synthetic oligonucleotides as therapeutics: The coming of age. In *Biotechnol. Annu. Rev.*, Vol. 5; **2000**; pp 155-196.

(16) Verma, S.; Eckstein, F. MODIFIED OLIGONUCLEOTIDES: Synthesis and Strategy for Users. *Annu. Rev. Biochem.* **1998**, *67* (1), pp 99-134. DOI: 10.1146/annurev.biochem.67.1.99.

(17) Bell, N. M.; Micklefield, J. Chemical Modification of Oligonucleotides for Therapeutic, Bioanalytical and other Applications. *ChemBioChem.* **2009**, *10* (17), pp 2691-2703. DOI: https://doi.org/10.1002/cbic.200900341.

(18) Ochoa, S.; Milam, V. T. Modified Nucleic Acids: Expanding the Capabilities of Functional Oligonucleotides. *Mol.* **2020**, *25* (20). DOI: 10.3390/molecules25204659.

(19) Gyi, J. I.; Lane, A. N.; Conn, G. L.; Brown, T. The orientation and dynamics of the C2′-OH and hydration of RNA and DNA·RNA hybrids. *Nucleic Acids Res.* **1998**, *26* (13), pp 3104-3110. DOI: 10.1093/nar/26.13.3104 (accessed 7/25/2023).

(20) Shen, Y.; Storici, F. Generation of RNA/DNA hybrids in genomic DNA by transformation using RNA-containing oligonucleotides. *J. Vis. Exp.* **2010**, (45). DOI: 10.3791/2152.

(21) Li, J.; Begbie, A.; Boehm, B. J.; Button, A.; Whidborne, C.; Pouferis, Y.; Huang, D. M.; Pukala, T. L. Ion Mobility-Mass Spectrometry Reveals Details of Formation and Structure for GAA·TCC DNA and RNA Triplexes. *J. Am. Soc. Mass Spectrom.* **2019**, *30* (1), pp 103-112. DOI: 10.1007/s13361-018-2077-9.

(22) Grünweller, A.; Hartmann, R. K. Locked nucleic acid oligonucleotides: the next generation of antisense agents? *BioDrugs.* **2007**, *21* (4), pp 235-243. DOI: 10.2165/00063030-200721040-00004.

(23) Singh, S. K.; Koshkin, A. A.; Wengel, J.; Nielsen, P. LNA (locked nucleic acids): synthesis and high-affinity nucleic acid recognition. *Chem. comm.* **1998**, (4), pp 455-456.

(24) Braasch, D. A.; Corey, D. R. Locked nucleic acid (LNA): fine-tuning the recognition of DNA and RNA. *Chem. Biol.* **2001**, *8* (1), pp 1-7. DOI: 10.1016/s1074-5521(00)00058-2.

(25) You, Y.; Moreira, B. G.; Behlke, M. A.; Owczarzy, R. Design of LNA probes that improve mismatch discrimination. *Nucleic Acids Res.* **2006**, *34* (8), e60-e60. DOI: 10.1093/nar/gkl175 (accessed 11/27/2023).

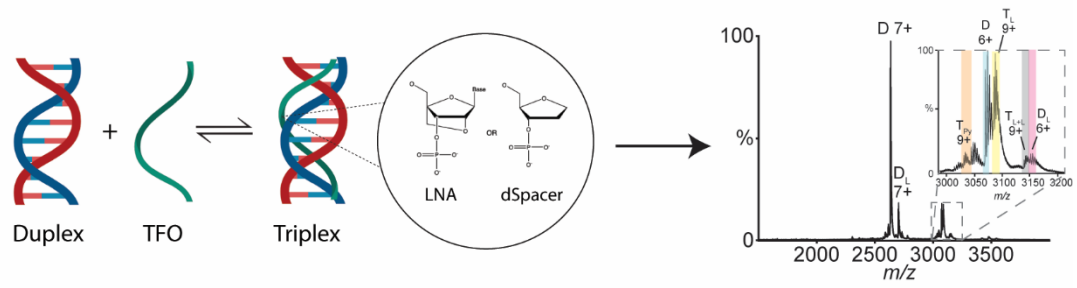
(26) Buske, F. A.; Bauer, D. C.; Mattick, J. S.; Bailey, T. L. Triplexator: detecting nucleic acid triple helices in genomic and transcriptomic data. *Genome Res.* **2012**, *22* (7), 1372-1381. DOI: 10.1101/gr.130237.111.

(27) Pabon-Martinez, Y. V.; Xu, Y.; Villa, A.; Lundin, K. E.; Geny, S.; Nguyen, C.-H.; Pedersen, E. B.; Jørgensen, P. T.; Wengel, J.; Nilsson, L.; et al. LNA effects on DNA binding and conformation: from single strand to duplex and triplex structures. *Scientific Reports* **2017**, *7* (1), 11043. DOI: 10.1038/s41598-017-09147-8.

(28) Rajeswari, M. R. DNA triplex structures in neurodegenerative disorder, Friedreich's ataxia. *J. Biosci.* **2012**, *37* (3), pp 519-532. DOI: 10.1007/s12038-012-9219-1.
(29) Global Priority List of Antibiotic-Resistant Bacteria to Guide Research, Discovery, and Development of new antibiotics. https://www.quotidianosanita.it/allegati/allegato4135670.pdf (accessed 3/23/2023).
(30) Stover, C. K.; Pham, X. Q.; Erwin, A. L.; Mizoguchi, S. D.; Warrener, P.; Hickey, M. J.; Brinkman, F. S.; Hufnagle, W. O.; Kowalik, D. J.; Lagrou, M.; et al. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nat.* **2000**, *406* (6799), pp 959-964. DOI: 10.1038/35023079.
(31) Rougée, M.; Faucon, B.; Mergny, J.; Barcelo, F.; Giovannangeli, C.; Garestier, T.; Hélène, C. Kinetics and thermodynamics of triple-helix formation: Effects of ionic strength and mismatched. *Biochem.* **1992**, *31* (38), pp 9269-9278.
(32) Mergny, J. L.; Sun, J. S.; Rougee, M.; Montenay-Garestier, T.; Barcelo, F.; Chomilier, J.; Helene, C. Sequence specificity in triple helix formation: experimental and theoretical studies of the effect of mismatches on triplex stability. *Biochem.* **1991**, *30* (40), pp 9791-9798.
(33) Wehbi, H.; Portillo, E.; Harvey, H.; Shimkoff, A. E.; Scheurwater, E. M.; Howell, P. L.; Burrows, L. L. The peptidoglycan-binding protein *FimV* promotes assembly of the *Pseudomonas aeruginosa* type IV pilus secretin. *J. Bacteriol.* **2011**, *193* (2), pp 540-550. DOI: 10.1128/jb.01048-10.
(34) Semmler, A. B. T.; Whitchurch, C. B.; Leech, A. J.; Mattick, J. S. Identification of a novel gene, *fimV*, involved in twitching motility in *Pseudomonas aeruginosa*. *Microbiol.* **2000**, *146 (Pt 6)*, pp 1321-1332. DOI: 10.1099/00221287-146-6-1321.
(35) Durand, E.; Bernadac, A.; Ball, G.; Lazdunski, A.; Sturgis, J. N.; Filloux, A. Type II protein secretion in *Pseudomonas aeruginosa*: the pseudopilus is a multifibrillar and adhesive structure. *J. Bacteriol.* **2003**, *185* (9), pp 2749-2758. DOI: 10.1128/jb.185.9.2749-2758.2003.
(36) Lu, H. M.; Motley, S. T.; Lory, S. Interactions of the components of the general secretion pathway: role of *Pseudomonas aeruginosa* type IV pilin subunits in complex formation and extracellular protein secretion. *Mol. Microbiol.* **1997**, *25* (2), pp 247-259. DOI: 10.1046/j.1365-2958.1997.4561818.x.
(37) Mergny, J.-L.; Li, J.; Lacroix, L.; Amrane, S.; Chaires, J. B. Thermal difference spectra: a specific signature for nucleic acid structures. *Nucleic Acids Res* **2005**, *33* (16), e 138 DOI: 10.1093/nar/gni134 (accessed 9/26/2023).
(38) Petersen, M.; Bondensgaard, K.; Wengel, J.; Jacobsen, J. P. Locked nucleic acid (LNA) recognition of RNA: NMR solution structures of LNA:RNA hybrids. *J. Am. Chem. Soc.* **2002**, *124* (21), pp 5974-5982. DOI: 10.1021/ja012288d.

Authors are required to submit a graphic entry for the Table of Contents (TOC) that, in conjunction with the manuscript title, should give the reader a representative idea of one of the following: A key structure, reaction, equation, concept, or theorem, etc., that is discussed in the manuscript. Consult the journal's Instructions for Authors for TOC graphic specifications.

Insert Table of Contents artwork here



4.3 Supplementary material

Unraveling DNA Triplex Stability: Mass Spectrometric Investigation of Modified Triplex Forming Oligonucleotides for Enhanced Gene Targeting

Jack W. Klose[‡1], Sarveenah Chandrasegaran[‡1], Tara L. Pukala[1]*

[1]Discipline of Chemistry, The University of Adelaide, Adelaide, South Australia, 5005, Australia

Supplementary material:

Table S1. DNA and LNA- and dSpacer-containing TFOs for the model sequence with a cytosine interruption. Pu and Py represent the purine single strand and pyrimidine single strand that constitute the DNA duplex respectively. LNA sites are indicated in lowercase letters (red) while DNA bases are indicated in uppercase letters in the sequence, and the pyrimidine interruption is highlighted in bold.

Name	Calculated molecular weight (Da)	Sequence (5'—3')
SEQ(C)-Pu	9550	AGAGAGAGAGAGAG**C**GAGAGAGAGAGAGAG
SEQ(C)-Py	8864	CTCTCTCTCTCTCTC**G**CTCTCTCTCTCTCT
SEQ(C)-L	9343	TCTcTcTcTcTcTc**T**cTcTcTcTcTcTCTC
SEQ(C)	8839	TCTCTCTCTCTCTC**T**CTCTCTCTCTCTCTC
SEQ(C)-d	8715	TCTCTCTCTCTCTC**[dSpacer]**CTCTCTCTCTCTCTC

Table S2. DNA and LNA- and dSpacer-containing TFOs for the model sequence with a thymine interruption. Pu and Py represent the purine single strand and pyrimidine single strand that constitute the DNA duplex respectively. LNA sites are indicated in lowercase letters (red) while DNA bases are indicated in uppercase letters in the sequence, and the pyrimidine interruption is highlighted in bold.

Name	Calculated molecular weight (Da)	Sequence (5'—3')
SEQ(T)-Pu	9565	AGAGAGAGAGAGAG**T**GAGAGAGAGAGAGAG
SEQ(T)-Py	8848	CTCTCTCTCTCTCTC**A**CTCTCTCTCTCTCT
SEQ(T)-L	9343	TCTcTcTcTcTcTcT**T**cTcTcTcTcTcTcTCTC
SEQ(T)	8839	TCTCTCTCTCTCTC**T**CTCTCTCTCTCTCTC
SEQ(T)-d	8715	TCTCTCTCTCTCTC**[dSpacer]**CTCTCTCTCTCTCTC

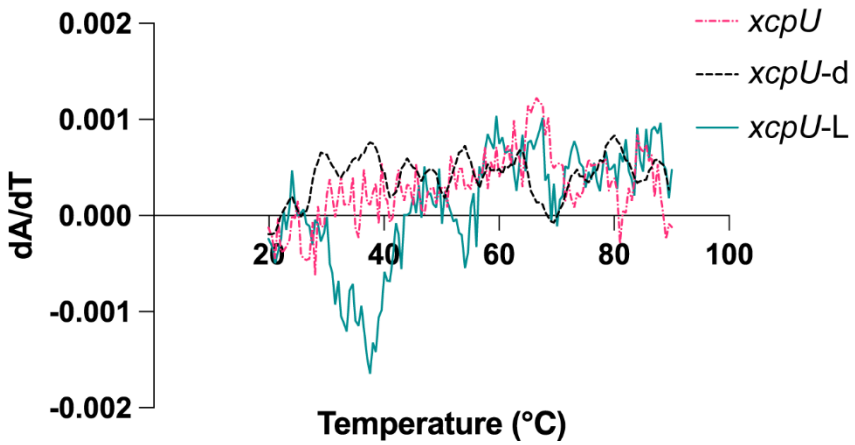
Table S3. DNA and LNA- and dSpacer-containing TFOs for the *fimV* biological sequence. Pu and Py represent the purine single strand and pyrimidine single strand that constitute the DNA duplex respectively. LNA sites are indicated in lowercase letters (red) while DNA bases are indicated in uppercase letters in the sequence.

Name	Calculated molecular weight (Da)	Sequence (5'—3')
fimV-Pu	5688	AGAAAGAGAAGGAAGAAG
fimV-Py	5309	CTTCTTCCTTCTCTTTCT
fimV	5309	TCTTTCTCTTCCTTCTTC
fimV-7L	5533	TCtTTtCtCtTcCtTcTTC
fimV-5L	5448	TCtTTcTCtTCcTTcTTC
fimV-3L	5407	TCtTTCTcTTCCtTCTTC

Table S4. DNA and LNA- and dSpacer-containing TFOs for the *xcpU* biological sequence. Pu and Py represent the purine single strand and pyrimidine single strand that constitute the DNA duplex respectively. LNA sites are indicated in lowercase letters (red) while DNA bases are indicated in uppercase letters in the sequence, and the pyrimidine interruption is highlighted in bold.

Name	Calculated molecular weight (Da)	Sequence (5'—3')
xcpU-Pu	6322	GAGAAGGAG**C**AGAAGAAAGG
xcpU-Py	5912	CCTTTCTTCT**G**CTCCTTCTC
xcpU	5887	CTCTTCCTC**T**TCTTCTTTCC
xcpU-L	6181	CTCtTcCtc**T**tcTtCtTtCC
xcpU-d	5763	CTCTTCCTC**[dSpacer]**TCTTCTTTCC

Figure S1: LNA TFO increases the stability of the *xcpU* sequence triplex from *Pseudomonas aeruginosa.* The derivative melting profiles of *xcpU* (pink dot/dash line), *xcpU*-d (black dashed line), and *xcpu*-L (green solid line) at 300 nm. Samples were prepared in a 1:1:1 TFO:purine:pyrimidine strand ratio, with each individual strand at a concentration of 2 μM in 250 μM of ammonium acetate solution at pH 7.4.



4.4 Supporting experiments

4.4.1 Introduction

This supplementary material covers several experiments that were conducted and not described in the prepared manuscript. This group of experiments only addresses the comparisons between dSpacer base modification and the canonical DNA TFOs, utilising analogous model sequences to Section 4.2. Here, the model parallel triplex system was designed based on the $(AG)_{15}$ polypurine TTS, where a central guanine is replaced by a pyrimidine base. In order to probe the sequence specificity of the pyrimidine interruption, both cytosine (SEQ2(C)) and thymine (SEQ2(T)) were used to replace the guanine in the TTS, as compared to a central adenine being replaced by the pyrimidine interruptions in the previous experiments. This sequence was chosen to compare the binding processes associated with the formation of duplex and triplex structures due to the symmetrical and alternating aspects of the sequence.

4.4.2 Materials and methods

Materials and methods are as described in Section 4.2 of the draft manuscript.

4.4.3 Results and discussion:

4.4.3.1 Comparison of unmodified TFO to dSpacer modification TFO

The SEQ2(C) and SEQ2(T) triplex systems with the canonical DNA (SEQ2(C) and SEQ2(T)) and dSpacer (SEQ(C)-d and SEQ(T)-d) TFOs were investigated to compare the triplex forming propensity of these systems using MS (full sequences are provided in Appendix 1 Tables A1 and A2).

The MS data for the SEQ2(C) and SEQ2(T) triplex systems are shown in Figures 4.1A and 4.1B respectively.

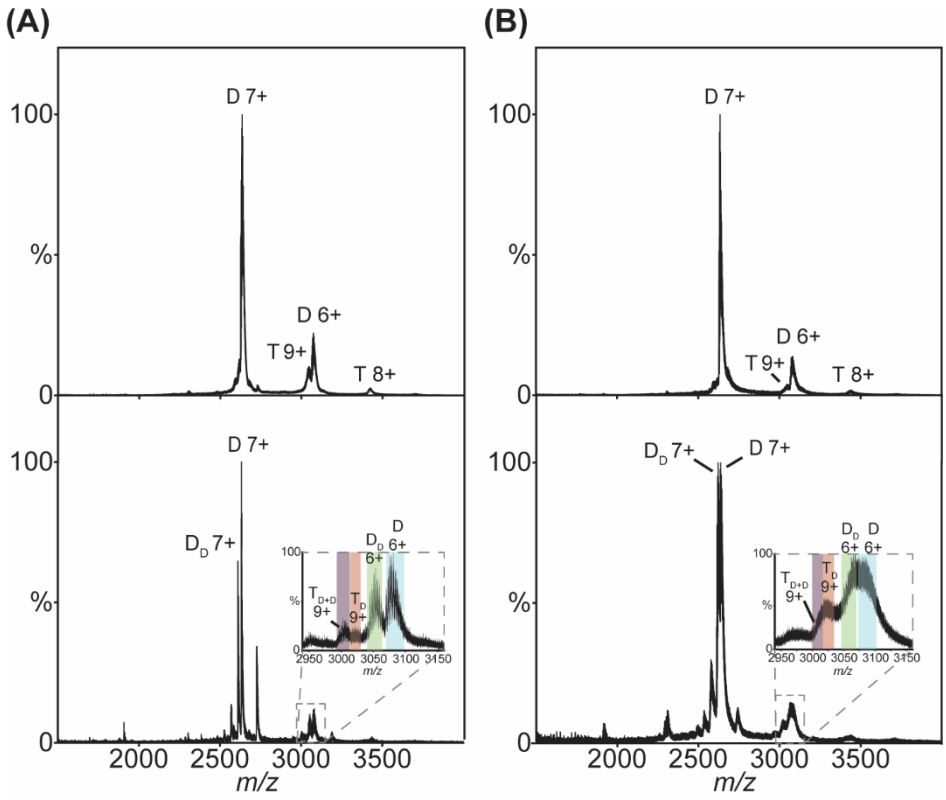


Figure 4.1: Mass spectra of **(A)** SEQ2(C) and **(B)** SEQ2(T) triplex systems, with the top panel indicating the systems with the DNA TFO and the bottom panel showing the dSpacer-containing TFO in 250 mM ammonium acetate solution at pH 7.4. All spectra were collected in positive ion mode. Samples were prepared in a 1:1:1 TFO:purine:pyrimidine strand ratio, with each individual strand at a concentration of 2 µM. Annotations are labelled as follows: SEQ2(C) and SEQ2(T) duplex:D, SEQ2(C) and SEQ2(T) DNA triplex:T, SEQ2(C)-d and SEQ2(T)-d duplex with dSpacer strand:D_D, SEQ2(C)-d and SEQ2(T)-d triplex:T_D, SEQ2(C)-d-d and SEQ2(T)-d-d triplex:T_{D+D}. SEQ2(C)-d and SEQ2(T)-d has highlighted species including T_{D+D}, T_D, D_D and D in purple, red, green and blue respectively in insets of **(A)** and **(B)** (bottom panels).

Figure 4.1A (top panel) shows the formation of the SEQ2(C) triplex with the canonical DNA TFO with significantly lower relative abundance than the duplex. In contrast, Figure 4.1A (bottom panel) shows the SEQ2(C)-d system which indicates the formation of two duplex structures. The canonical duplex structure is formed, along with another duplex structure where the SEQ2(C)-d TFO has bound to the complementary purine strand. The formation of multiple duplex structures lowers the triplex forming propensity as there is less dSpacer TFO available for binding. The MS data also indicates the formation of multiple triplex species upon addition of the dSpacer containing TFO, including a triplex species that corresponds to the dSpacer TFO binding to the secondary dSpacer duplex. Triplex formation with the dSpacer TFO and the canonical duplex was also observed, however the formation of secondary species indicates that the dSpacer base modification is not specific enough to overcome the pyrimidine interruption compared to the DNA TFO.

Figure 4.1B (top panel) presents MS data for the SEQ2(T) DNA triplex system, which indicates the formation of the SEQ2(T) triplex with once again significantly lower relative abundance compared to the duplex. Much like the analogous SEQ2(C)-d triplex system, there is the formation of multiple duplex and triplex species in the SEQ2(T)-d triplex system (Figure 4.1B bottom), limiting the triplex forming propensity of the dSpacer triplex that is being used to overcome the pyrimidine interruption in the TTS.

The formation of these multiple duplex and triplex species affects the triplex forming propensity of the designed model system as also described in Section 4.2, as a result of the repeating nature of the sequence.

4.4.3.2 Adenine based triplex target site model systems using the DNA and dSpacer TFO

Investigation of a final model system was explored to see if changing the base composition and removing the alternating bases in the sequence influences the formation of multiple duplex and triplex structures as observed in previous model sequences using canonical DNA and dSpacer TFOs. This system was designed based on a $(A)_{30}$ TTS, where the middle adenine was replaced with a cytosine (SEQ3(C)) and a thymine (SEQ3(T)) pyrimidine interruption.

MS data of the SEQ3(C) and SEQ3(T) systems can be seen in Figure 4.2A and 4.2B respectively (full sequences are provided in Appendix 1 Tables A3 and A4).

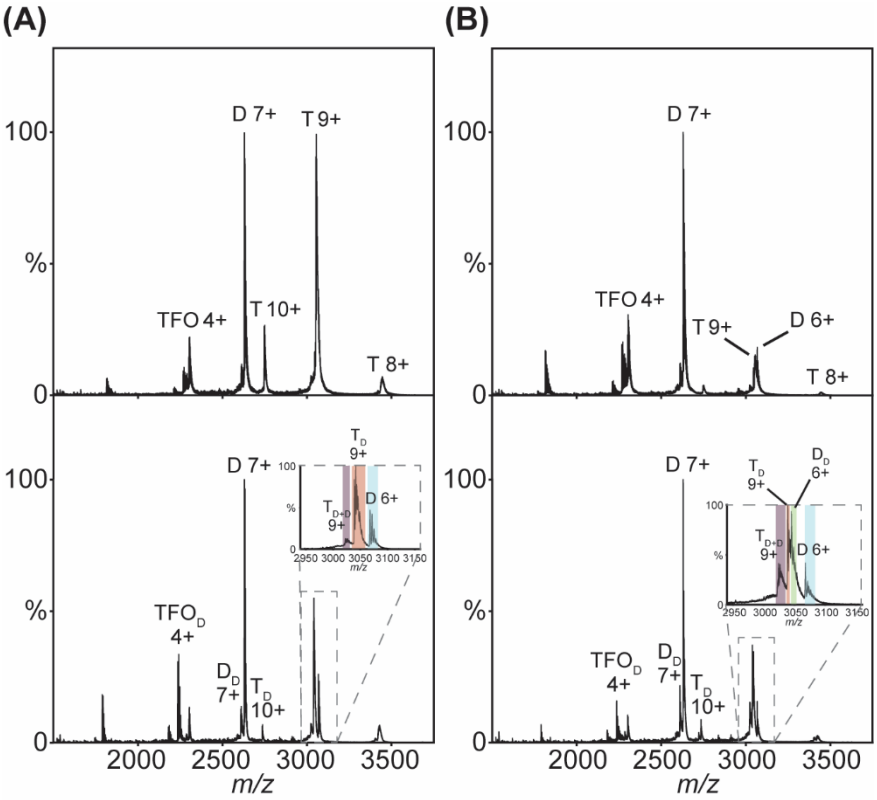


Figure 4.2: Mass spectra of **(A)** SEQ3(C) and **(B)** SEQ3(T) triplex systems, with the top panel indicating the systems with the DNA TFO and the bottom panel showing the dSpacer TFO in 250 mM ammonium acetate solution at pH 7.4. All spectra were collected in positive ion mode. Samples were prepared in a 1:1:1 TFO:purine:pyrimidine strand ratio, with each individual strand at a concentration of 2 µM. Annotations are labelled as follows: SEQ3(C) and SEQ3(T) TFO:TFO, SEQ3(C)-d and SEQ3(T)-d TFO:TFO_D SEQ3(C) and SEQ3(T) duplex:D, SEQ3(C) and SEQ3(T) DNA triplex:T, SEQ3(C)-d and SEQ23T)-d duplex with dSpacer strand:D_D, SEQ3(C)-d and SEQ3(T)-d triplex:T_D, SEQ3(C)-d-d and SEQ3(T)-d-d triplex:T_{D+D}. SEQ3(C)-d and SEQ3(T)-d has highlighted species including T_{D+D}, T_D, D_D and D in purple, red, green and blue respectively.

The MS spectrum in Figure 4.2A (top panel) indicates that the SEQ3(C) triplex ions are of similar relative abundance to the duplex in the system, indicating that there is significant triplex formation compared to the other model systems assessed. This shows that sequence specificity can be an influential factor when overcoming pyrimidine interruptions in the TTS. The bottom panel of Figure 4.2A shows that there is still formation of the two duplex structures, similar to the SEQ2(C)-d and SEQ2(T)-d systems previously explained. However, the SEQ3(C)-d triplex has a much higher relative abundance compared to the other examples, even with the formation of the secondary dSpacer duplex structure limiting the amount of TFO available for triplex formation. There is also indication of the second triple helical

structure being formed, which can be attributed to the dSpacer TFO binding to the dSpacer-containing duplex species.

The MS data in Figure 4.2B similarly indicates the formation of the SEQ3(T) triplex, and also provides evidence for the formation of the secondary duplex and triplexes, indicating there is no difference in the formation of these non-specific duplexes due to the different pyrimidine interruptions in the SEQ3(C) and SEQ3(T) systems. However, of particular note in the SEQ3(T)system, the SEQ3(T) DNA triplex forms at a lower relative abundance than the SEQ3(T)-d triplex, which is the opposite of the SEQ3(C) system. This must be attributed to the different base interruptions and confirms that specific base modification does have an effect on the successful formation of the triplexes. Therefore, sequence specificity of not only the TTS but the explicit interruption needs to be considered when designing DNA triplexes for targeting specific genes.

4.4.4 Conclusions

Four model sequences of different compositions were investigated using MS, to determine if a triplex structure with a pyrimidine base interruption in the purine-rich TTS can form using canonical DNA and dSpacer modified TFOs. Formation of triplexes in the SEQ2(C), SEQ2(T), SEQ3(C) and SEQ3(T) systems with the dSpacer TFOs was limited due to the formation of competing duplex species that were observed, thereby limiting the amount of dSpacer TFO that is available for triplex formation. The dSpacer base modification is therefore suboptimal for the formation of triplex assemblies, particularly since unmodified DNA TFOs provided comparably higher triplex forming propensity overall with the pyrimidine interruption in the TTS. Given the high likelihood of pyrimidine interruptions in the genomic TTSs, further investigation into different base modifications would greatly benefit the development of stable DNA triplexes for gene targeting and subsequently potential antigene therapeutics.

4.4.5 References:

(1) Fox, K.; Brown, T.; Rusling, D. Chapter 1: DNA Recognition by Parallel Triplex Formation. In *DNA-targeting molecules as therapeutic agents*; 2018; pp 1–32.

(2) Thompson, P. S.; Cortez, D. New Insights into Abasic Site Repair and Tolerance. *DNA Repair (Amst).* **2020**, *90*, 102866.

(3) Mandi, C. S.; Mahata, T.; Patra, D.; Chakraborty, J.; Bora, A.; Pal, R.; Dutta, S. Clevage of Abasic Sites in DNA by an Aminiquinoxaline Compound: Augmented Cytotoxicity and DNA Damage in Combination with an Anticancer Drug Chlorambucil in Human Colorectal Carcinoma Cells. *ACS Omega* **2022**, *7* (8), 6488–6501.

(4) Greenberg, M. M. Abasic and Oxidized Abasic Site Reactivity in DNA: Enzyme Inhibition, Cross-Linking, and Nucleosome Catalyzed Reactions. *Acc. Chem. Res.* **2014**, *47* (2), 646–655.

(5) K. Singh, S.; A. Koshkin, A.; Wengel, J.; Nielsen, P. LNA (Locked Nucleic Acids): Synthesis and High-Affinity Nucleic Acid Recognition. *Chem. Commun.* **1998**, No. 4, 455–456. https://doi.org/10.1039/A708608C.

(6) Braasch, D. A.; Corey, D. R. Locked Nucleic Acid (LNA): Fine-Tuning the Recognition of DNA and RNA. *Chem. Biol.* **2001**, *8* (1), 1–7. https://doi.org/10.1016/S1074-5521(00)00058-2.

(7) Grünweiler, A.; Hartmann, R. K. Locked Nucleic Acid Oligonucleotides. *BioDrugs* **2007**, *21* (4), 235–243. https://doi.org/10.2165/00063030-200721040-00004.

(8) Pabon-Martinez, Y.; Xu, Y.; Villa, A.; Lundin, K. E.; Geny, S.; Nguyen, C.-H.; Pedersen, E. B.; Jørgensen, P. T.; Wengel, J.; Nilsson, L.; Smith, C. I. E.; Zain, R. LNA Effects on DNA Binding and Conformation: From Single Strand to Duplex and Triplex Structures. *Sci. Rep.* **2017**, *7*, 11043.

4.4.6 Appendix 1: Name and sequences for each constituent strand for the model SEQ2 and SEQ3 triplex systems containing both the cytosine and thymine interruption in the TTS, including canonical DNA and dSpacer TFOs

Table A1: Individual strands for duplex, DNA TFO and dSpacer TFO for the SEQ2(C) triplex systems, where the middle guanine is replaced by a cytosine interruption in the TTS. The site of the interruption is highlighted in bold.

Name	Calculated molecular weight (Da)	Sequence (5'—3')
SEQ2(C)-Pu	9534	AGAGAGAGAGAGA**C**AGAGAGAGAGAGAGAG
SEQ2(C)-Py	8879	CTCTCTCTCTCTCTCT**G**TCTCTCTCTCTCT
SEQ2(C)	8839	TCTCTCTCTCTCT**C**TCTCTCTCTCTCTCTC
SEQ2(C)-d	8730	TCTCTCTCTCTCT**[dspacer]**TCTCTCTCTCTCTCTC

Table A2: Individual strands for duplex, DNA TFO and dSpacer TFO for the SEQ2(T) triplex systems, where the middle guanine is replaced by a thymine interruption in the TTS. The site of the interruption is highlighted in bold.

Name	Calculated molecular weight (Da)	Sequence (5'—3')
SEQ2(T)-Pu	9549	AGAGAGAGAGAGA**T**AGAGAGAGAGAGAGAG
SEQ2(T)-Py	8863	CTCTCTCTCTCTCTCT**A**TCTCTCTCTCTCT
SEQ2(T)	8839	TCTCTCTCTCTCT**C**TCTCTCTCTCTCTC
SEQ2(T)-d	8730	TCTCTCTCTCTCT**[dspacer]**TCTCTCTCTCTCTC

Table A3: Individual strands for duplex, DNA TFO and dSpacer TFO for the SEQ3(C) triplex systems, where the middle adenine is replaced by a cytosine interruption in the TTS. The site of the interruption is highlighted in bold.

Name	Calculated molecular weight (Da)	Sequence (5'—3')
SEQ3(C)-Pu	9310	AAAAAAAAAAAAA**C**AAAAAAAAAAAAAAA
SEQ3(C)-Py	9089	TTTTTTTTTTTTTTTT**G**TTTTTTTTTTTTT
SEQ3(C)	9064	TTTTTTTTTTTTT**T**TTTTTTTTTTTTTTT
SEQ3(C)-d	8940	TTTTTTTTTTTTT**[dspacer]**TTTTTTTTTTTTTTT

Table A4: Individual strands for duplex, DNA TFO and dSpacer TFO for the SEQ3(T) triplex systems, where the middle adenine is replaced by a cytosine interruption in the TTS. The site of the interruption is highlighted in bold.

Name	Calculated molecular weight (Da)	Sequence (5'—3')
SEQ3(T)-Pu	9325	AAAAAAAAAAAAA**T**AAAAAAAAAAAAAAA
SEQ3(T)-Py	9073	TTTTTTTTTTTTTTTT**C**TTTTTTTTTTTTT
SEQ3(T)	9064	TTTTTTTTTTTTT**T**TTTTTTTTTTTTTTT
SEQ3(T)-d	8940	TTTTTTTTTTTTT**[dspacer]**TTTTTTTTTTTTTTT

Chapter 5: A triplex antigene strategy targeted towards methicillin-resistant *Staphylococcus aureus*

5.1 Introduction

Genes comprise long sequences of DNA which form a template for transcription of messenger RNA, ultimately allowing for gene expression in the form of protein synthesis.[1,2] Triplex DNA has been implicated in several cellular functions involving the regulation of gene expression and post-transcriptional RNA processing in a number of multicellular organisms.[3,4] These functionalities lend these triple helical assemblies to antigene strategies, enabling selective control of gene expression. This strategy can be exploited by introducing a TFO into a cell that can target and bind to a specific gene of interest at a polypurine-based triplex target site (TTS) and repress gene expression.

Bioinformatics programs such as Triplexator have been developed to identify TTSs in any genome, which allows for many gene targets to be uncovered.[5] Formation of a triplex at the promoter region of a gene (which is located directly upstream of the transcription initiation site) will prevent the recruitment of essential transcription factors and prevent gene expression, and where the gene product is essential for function, this can cause cell death (Figure 5.1).[6,7] To utilise an antigene strategy for gene targeting applications, further development is required, including delivery of the TFO into the cell and increasing the stability of triplexes *in vivo*. Previous reports have indicated that cell-penetrating peptides (CPPs) and nanoparticles are preferred methods of enabling TFO delivery to the nucleus, due to their direct *in vivo* delivery and low levels of cell death compared to other transport methods.[8] In the case of CPPs, synthetic bioconjugation techniques can be employed to attach the peptide to the TFO for more efficient delivery.

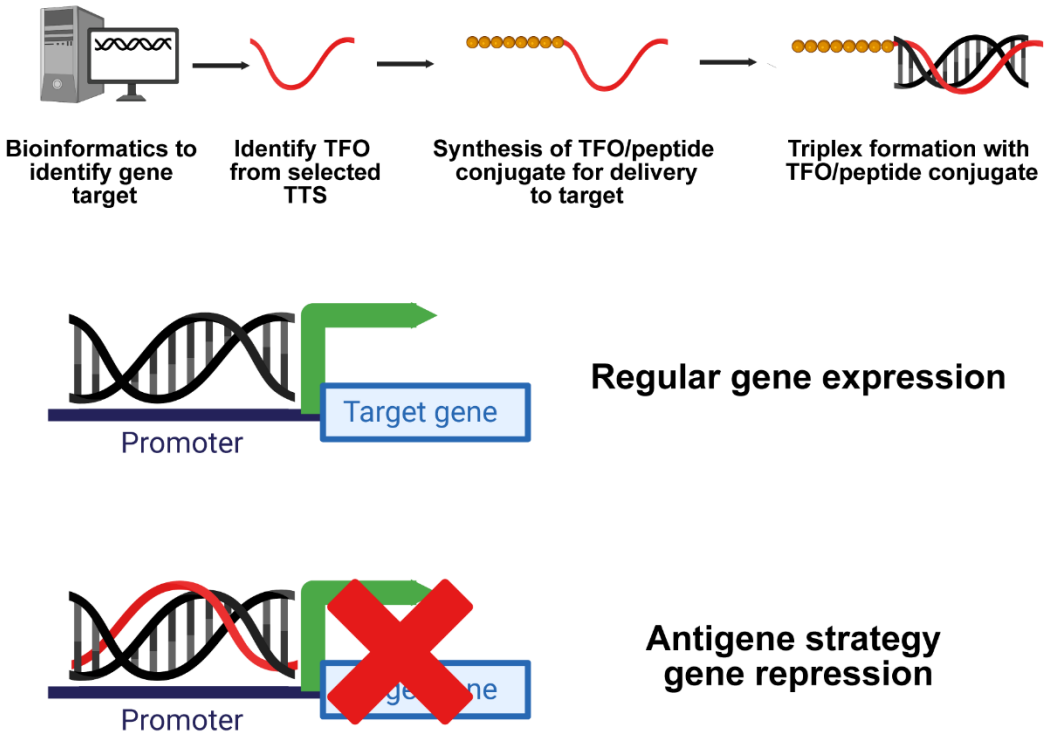


Figure 5.1: Proposed triplex DNA antigene strategy used to target genes in bacterial genomes for developing novel antibacterial therapies. Bioinformatics tools can be used to identify polypurine TTSs that would allow triplex formation at specific gene targets. The corresponding TFO can then be used in a bioconjugation reaction to form a TFO/peptide conjugate which aids delivery of the TFO to the target *in vivo*. Triplex formation at the promoter region of the gene target will prevent the binding of crucial transcription factors, allowing for the reduction of gene expression. Created with biorender.com.

An antigene strategy could be used to target critical genes within bacterial genomes to help develop the next generation of antibacterial therapeutics to overcome the extreme issues of antibiotic resistant bacteria. The overuse and misuse of antibiotics have led to the development of antibiotic resistant bacteria such as methicillin-resistant *Staphylococcus aureus* (MRSA). Antiobiotic resistance is a global issue that results in over 5 million deaths each year and is a major cause of hospital-acquired infections.[9] This means that these antimicrobial-resistant bacteria killed more people than HIV/AIDS and malaria with 864,000 and 643,000 deaths respectively.[10] It is predicted that by 2050, this number could increase to 10 million people indicating the need for the development of a new generation of antibacterial agents that can reduce the spread and conserve the effectiveness of existing antibacterials by preventing their overuse.[11]

The antibiotic resistance mechanisms that MRSA has make it a life-threatening organism to both humans and animals.[12] MRSA are also shown to have multi-drug resistance to many different classes of antibacterial agents including fluoroquinolones and tetracyclines, making it an even more potent infection.[12] An antigene therapy approach to combat MRSA that uses triplex structures is promising due to the potential to target multiple sites. Since the TFOs would be acting on essential functions of the bacteria it could also minimise their ability to form antibiotic resistance.

This work aims to investigate the formation and stability of a biologically relevant triplex sequence from MRSA in an antigene strategy using MS, UV-Vis spectroscopy and antimicrobial assay techniques. It is hoped that these preliminary experiments can contribute to unlocking the full potential of DNA triplexes as gene regulators *in vivo*.

5.2 Methods

5.2.1 Bioinformatic analysis of triplex target sites in methicillin-resistant Staphylococcus aureus

Potential TTSs for MRSA were identified using Triplexator software.[5] The MRSA*252* genome (accession number: NC_00295.2) was obtained from the National Center for Biotechnology Information (NCBI).[13] The minimum length of the TTS was set to 15 nucleotides with the maximum up to 30. The error rate was set to 0 with a consecutive error of 0 to ensure that the TTS had only guanines and adenines. The UniProtKB database of proteins was used to identify critical genes within the MRSA*252* genome (accession number: A0A7U7EVD7) in combination with the basic local alignment search tool (BLAST) engine that allowed the position of these genes to be located within the genome.[14,15]

5.2.2 Sample preparation

Oligonucleotides were purchased from Sigma Aldrich (NSW, Australia) at a desalted purity, and dissolved in deionised water (18 MΩ·cm^{-1}) to approximately 100 µM. The concentration of each oligonucleotide solution was calculated via the Beer-Lambert law by measuring the absorbance at 260 nm on a ThermoFisher NanoDrop^{TM} 2000 spectrophotometer. Each sample was measured in triplicate, and extinction coefficients were provided by the manufacturer. Ammonium acetate solutions were prepared from a 7.5 M aqueous stock and adjusted to the desired pH using ammonium hydroxide (both > 99.99% purity). Two alkyne-substituted oligonucleotides (5'5-Octadiynyl dU-CTTTTCTTTTTCTTCTT and CTTTTCTTTTTCTTCTT-3'5-Octadiynyl dU) were purchased desalted from IDT Technologies (Iowa, USA). A cell-penetrating peptide with the amino acid sequence of KFFKFFKFFK was purchased with either an azide-PEG4 (GL Biochem, China) or azide substituent (Genscript, Singapore) on the N-terminal lysine residue. 7.5 M stock ammonium acetate solution was used to prepare pH 7.4 ammonium acetate buffers. pH adjustment was achieved by the addition of ammonium hydroxide. Duplex structures were formed by dissolving the appropriate DNA strands in a 1:1 ratio in the desired buffer so that the final concentration of the duplex was 20 µM. The two strands were then annealed at 4 °C for 2 hours. For triplexes, the TFO strand was then added to the annealed duplex so that the strands were in a 1:1:1 ratio. The triplex was then left to anneal at 4 °C for another 2 hours (or 4 hours for the DNA/cell-penetrating peptide TFO

conjugate). The samples were then diluted to 2 μM in the desired buffer ready for MS and UV-is spectroscopy analysis.

5.2.3 Synthesis of TFO/cell-penetrating peptide conjugate

5’ 5-Octadiynyl dU-CTTTTCTTTTTCTTCTT or CTTTTCTTTTTCTTCTT-3’ 5-Octadiynyl dU (0.92 mM, 50 uL) and KFFKFFKFFK (with azide-PEG4 or azide attachment) (2.5 mM, 20 μL) was diluted with MilliQ water (90 μL). 2-(4-((Bis((1-(*tert*-butyl)-1*H*-1,2,3-triazol-4-yl)methyl)amino)methyl)-1*H*-1,2,3-triazol-1-yl)acetic acid (BTTAA) (100 mM, 10 μL), copper sulphate (20 mM, 10 μL) and sodium ascorbate (300 mM, 10 uL) was added sequentially to the solution prepared above, vortexing the solution after the addition of each reagent. The reaction mixture was then left to incubate for 30 minutes whilst protected from light. The conjugate was desalted through an Amicon Ultra-0.5 centrifugation unit (Merck Millipore, Darmstadt, Germany) with a 3000 MW cut-off and concentrated back to 100 μM with ammonium acetate pH 7.4 as per the manufacturer’s protocol in preparation for antimicrobial experiments.

5.2.4 Native mass spectrometry

Mass spectra were acquired using a Bruker Impact HD Q-TOF spectrometer, fitted with a nanospray ion source in positive ion mode. The capillary housing was modified such that a backing pressure of N_2 gas could be applied to the sample to stabilise ion spray. Key instrument parameters were: End Plate Offset 500 V, Capillary 1800 V, Dry Gas 6.0 $L \cdot min^{-1}$, Dry Temp 140 °C, Funnel 1 RF 400.0 VPP, Funnel 2 RF 600.0 VPP, Hexapole RF 300.0 VPP, Quadrupole Ion Energy 4.0 eV. Collision Cell parameters were: Collision RF 1200.0 VPP, Transfer Time 120.0 μs, Pre Pulse Storage 20.0 μs and Collision Cell Energy 30.0 eV.

5.2.5 UV-Vis spectroscopy

UV-Vis spectroscopy experiments were conducted on a Shimadzu UV-3600i-Plus spectrophotometer equipped with a temperature-controlled 8 Series Micro Multi-Cell Holder (MMC-1600C). Thermal melting experiments were conducted by ramping the temperature from 20 °C to the final temperature at a rate of 1 °C/min. The absorbance at 260 nm was recorded at every 0.5 °C interval. To limit evaporation and gas bubble formation during temperature ramping all samples were degassed by sonicating for 5 minutes prior to analysis.

The cells were sealed during each run with provided silicone stoppers which were secured in place by the clamp as part of the cell holder assembly. As such, all UV-Vis melting experiments were conducted under constant volume conditions (135 µL). All experiments were completed in triplicate and the data shown is an average of the three runs, with derivative plots being smoothed with a Savitzky-Golay filter to improve visualisation.

5.2.6 Antimicrobial assay

The antimicrobial activity of the DNA-peptide conjugate was examined against *Staphylococcus aureus* ATCC 49975 by a modified microdilution broth method using Mueller-Hinton broth (Trek Diagnostics Systems, Cleveland, Ohio). DNA-CPP conjugate was dissolved in ammonium acetate (pH 7.4) and serially diluted 2-fold from an initial concentration of 4 mg/mL. Microplates were inoculated with *S. aureus* and incubated at 37 °C for 24 hours. Bacterial growth was assessed using an Agilent BioTek logphase 600 microbiology reader. 96-well plates were agitated at 500 rpm, with absorbance measurements taken every 20 minutes at 600 nm. Alkyne-DNA, azide-peptide, kanamycin, *S. aureus* cells and Mueller Hinton (M-H) broth media were also included as controls for the experiment.

5.3 Results & discussion

5.3.1 Identification of triplex target sites from the MRSA252 genome

Candidate TTSs were first identified in the MRSA*252* genome using Triplexator. Triplexator identifies specific purine rich sequences within a selected genome, where the search filters can be changed to account for the length and composition in the TTS. A minimum length of 15 bases was chosen as being sufficiently long to form a stable triplex structure, as the longer the TTS, the more Hoogsteen bonding is available to stabilise the triplex assembly. As discussed in Chapter 4, pyrimidine interruptions in the TTS negatively influence triplex formation, so the search was conducted for sequences without any interruptions. The input and summary of the Triplexator search can be seen in Appendix 1.

S. aureus was chosen as the target bacteria due to its heavy association with the rise of drug resistance and the severity of the infection it causes, especially for hospital-acquired infections.[12] The Triplexator search initially generated 954 TTS. This was reduced to 150 sites after the removal of smaller-sized duplicate TTSs that had been frameshifted by one or multiple bases. The generated output can be seen in Appendix 2 Table A1. To narrow down the number of sites further, the UniProtKB database and BLAST search were used to find the specific location of the TTS in the genome and determine whether it appeared in the promoter region (located in regulatory regions) of a specific gene of interest. Sites that coincide with the gene promoter region have the highest chance of being able to repress gene expression by preventing the binding of transcription factors.[16] This significantly reduced the number of candidate TTSs, with a summary of the TTSs that were located in potential critical gene targets from the MRSA*252* genome (Table 5.1).

Table 5.1: Purine-rich triplex target sites generated using Triplexator, and their corresponding gene location, from the MRSA252 genome.

Gene name	Sequence (5'—3')	Length	Promoter region
purL	AGAAGGAGAAGAAAAAAA	18	No
ptsH	AGAAGAAAAAGAAAA	15	No
fumC	GAAAGAAAAAGAAGA	15	No
rlmN	GAAAAGAAAAAGAAGAA	17	Yes
sraP	GAAGAGAAAGAAAAAGAAAGA	21	No
purB	GAAGGAAAAGGAAAAA	16	No
metG	GAAGAGGAAAAAGAAGAGA	19	No
FtsK	GAGGAAAAAGAAGAA	15	Yes
pnp	AAAAAGAGAGAAGGA	15	No
	GAAGAAGAAAAAAGA	15	No
hutI	AAAGGAAAAGAAAAG	15	No

As indicated in Table 5.1, there were only two sites located in the promoter region of the respective genes. The 17mer TTS (GAAAAGAAAAAGAAGAA) corresponding to the dual specificity RNA methyltransferase (*rlmN*) gene was selected as a key target for further investigation, due to it having a longer length than the DNA translocase sequence and hence predicted to form a more stable triplex assembly. The *rlmN* gene is responsible for the production of the dual-specificity RNA methyltransferase protein which exclusively methylates position 2 of adenine 2503 in 23S ribosomal RNA and position 2 of adenine 37 in transfer RNAs.[13] This gene also assists with resistance to antibiotics, making it a critical gene to target for antigene technology. To ensure that the *rlmN* gene site is suitable for targeting *in vivo*, the formation of a triplex using the *rlmN* gene TTS as a template *in vitro* is vital.

5.3.2 Formation of the *rlmN* DNA triplex

Following identification of a suitable TTS, UV-Vis spectroscopy and MS experiments were performed on the *rlmN* DNA triplex system (full sequences can be seen in Appendix 2 Table 2) to confirm triplex formation *in vitro,* as shown in Figures 5.2A and 5.2B respectively.

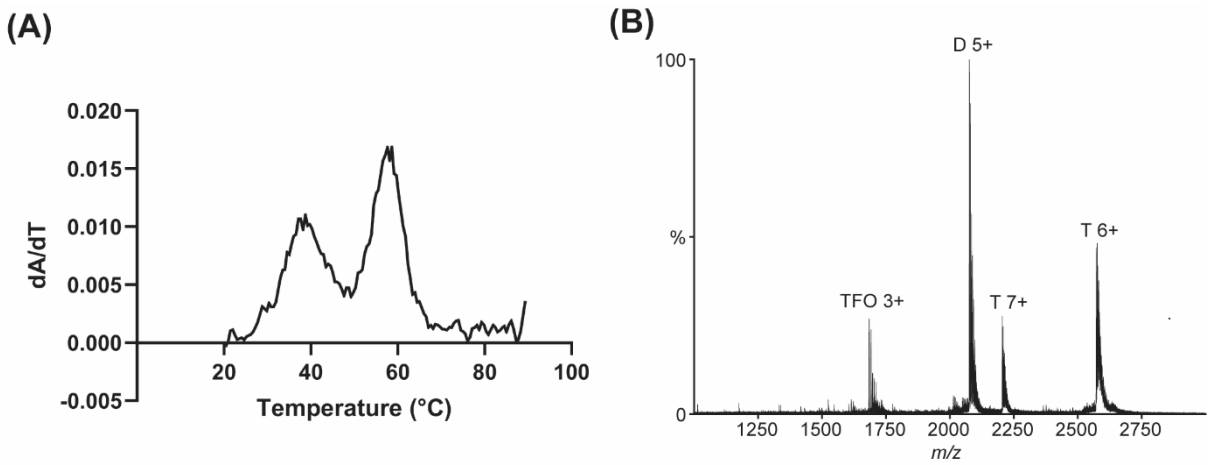


Figure 5.2: (A) Derivative UV-Vis spectroscopy melting profiles of the rlmN DNA triplex in 250 mM ammonium acetate at pH 7.4. All samples were zeroed at 260 nm at 20 °C, and melting was conducted from 20 to 90 °C. **(B)** Mass spectrum of rlmN DNA triplex system, in 250 mM ammonium acetate solution at pH 7.4. The mass spectrum was collected in positive ion mode. Samples were prepared in a 1:1:1 TFO: purine: pyrimidine strand ratio, with each individual strand at a concentration of 2 µM for both UV-Vis spectroscopy and MS experiments. Annotations are as follows: TFO: TFO, D: duplex, T: triplex. Charge states are also indicated on labels.

The UV-Vis spectroscopy data in Figure 5.2A indicates a hyperchromic triplex melting event at 39 °C and a second transition at 58 °C corresponding to duplex melting. The MS data similarly demonstrates clear evidence for triplex formation, confirming the UV-Vis spectroscopy data. Now that formation of the *rlmN* triplex has been confirmed, it is necessary to be able to deliver the *rlmN* TFO into *S. aureus* to form the triplex *in vivo*. This can be achieved by adding a CPP to the TFO.

5.3.3 Synthesis of 5' and 3' TFO-cell-penetrating peptide conjugates

A principle issue with antigene technology in this context is the necessity to deliver the antigene agent or TFO into the bacterial cell. It has been demonstrated previously that a cell-penetrating peptide with sequence KFFKFFKFFK (where K is lysine and F is phenylalanine) enabled the successful delivery of oligonucleotides across the cell membrane in *S. aureus*.[17] This CPP is effective due to its amphipathic nature, allowing it to easily traverse both the hydrophilic head groups and hydrophobic tails in the phospholipid layer of the plasma membrane. It has also been shown that amphipathic CPPs have a higher cell internalisation rate than cationic CPPs.[18] Hence, the KFFKFFKFFK CPP was selected to be conjugated to the *rlmN* TFO (5'-CTTTTCTTTTTCTTCTT-3').

A copper-catalysed [3+2] azide-alkyne cycloaddition reaction was used to enable the conjugation of the TFO and CPP. Two different synthetic *rlmN* TFOs (one with the alkyne group attached to the 5' end of the sequence and the other attached to the 3' end) were conjugated via an azide group that is attached to the end of a PEG4 linker of the CPP. The azide-PEG4 substituent is attached to the side chain of the N-terminal lysine of the CPP. The reaction scheme for the copper-catalysed [3+2] azide-alkyne cycloaddition reaction is shown in Figure 5.3.

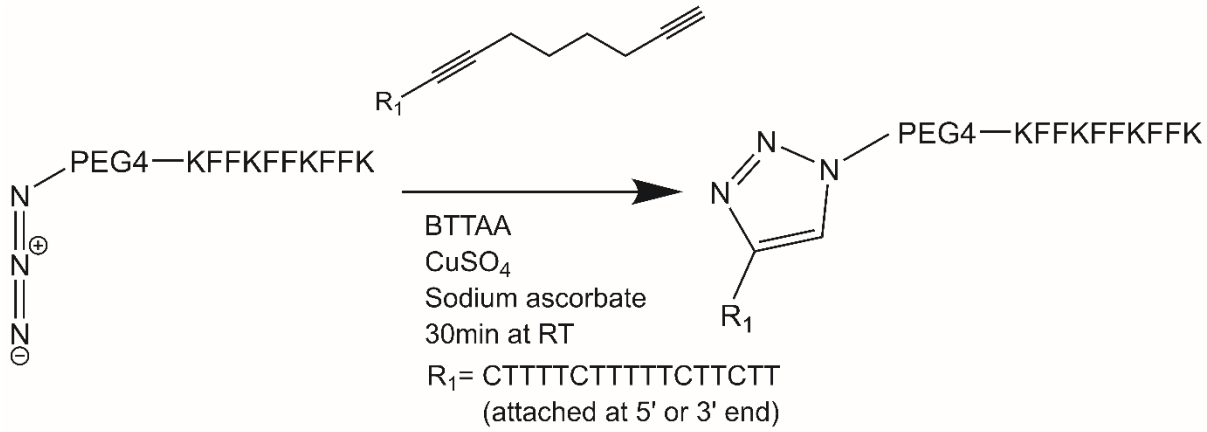


Figure 5.3: Copper catalysed [3+2] azide-alkyne cycloaddition reaction allows for the conjugation of the KFFKFFKFFK CPP with an azide-PEG4 group to the alkyne functionalised *rlmN* TFO.

The success of this reaction was then evaluated using MS. Both resulting MS spectra (Figures 5.4A and 5.4B) indicate the formation of the TFO-CPP conjugate where the CPP is attached to the 5' and 3' end of the TFO respectively. Both are in very high abundance with limited starting CPP remaining. Since the TFO-CPP conjugates have been successfully synthesised, their ability to form a triplex structure with the *rlmN* duplex can now be determined and compared to the regular DNA TFO.

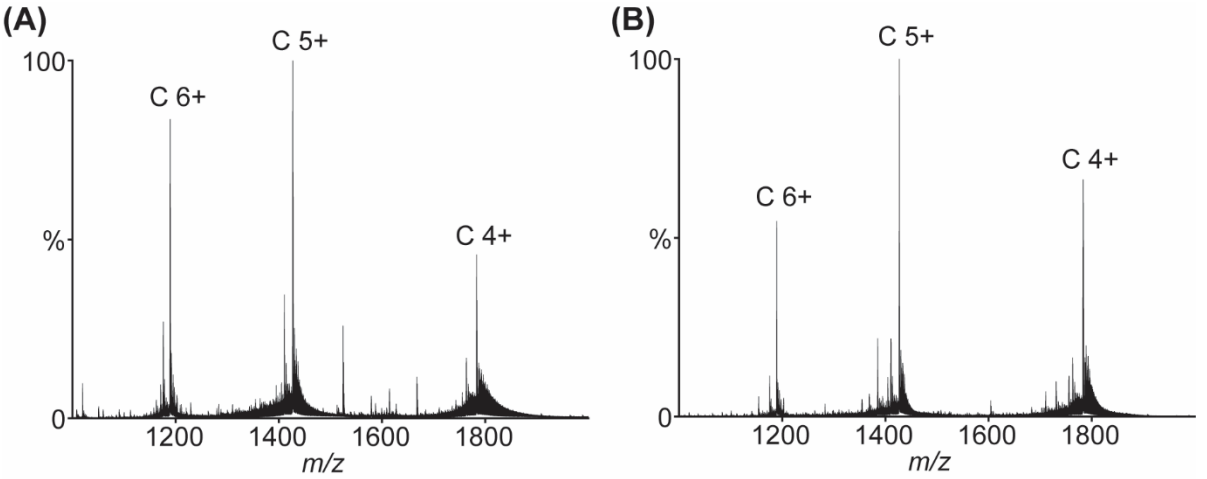


Figure 5.4: Mass spectra of formation of **(A)** 5' and **(B)** 3' rlmN TFO-CPP conjugates in 250 mM ammonium acetate solution at pH 7.4. Both spectra were collected in positive ion mode. Annotations are as follows: C: TFO-CPP conjugate, Charge states are also indicated on labels.

5.3.4 *rlmN* TFO-cell-penetrating peptide conjugate triplex formation

To determine the viability of this synthetic TFO to target the *rlmN* gene in *S. aureus*, we first needed to investigate whether it can form a triplex structure *in vitro*. Native MS data for triplex formation with a four-hour incubation time for the 5' and 3' TFO-CPP conjugates are shown in Figure 5.5.

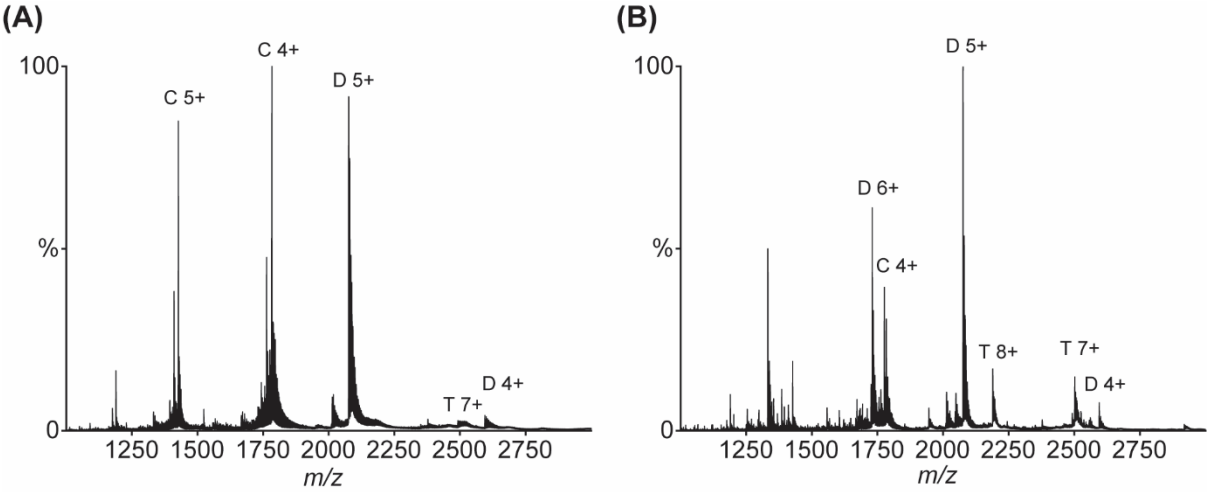


Figure 5.5: Mass spectrum of **(A)** 5' TFO-CPP and **(B)** 3' TFO-CPP with PEG4 linker *rlmN* DNA triplex systems with four-hour incubation time between the TFO and duplex. All samples were in 250 mM ammonium acetate solution at pH 7.4 and all spectra were collected in positive ion mode. Samples were prepared in a 1:1:1 TFO-CPP:purine:pyrimidine strand ratio, with each individual strand at a concentration of 2 µM. Annotations are as follows: C: TFO-CPP conjugate, D: duplex, T: triplex with TFO-CPP conjugate as the TFO. Charge states are also indicated on the spectra.

The MS data in Figure 5.5A shows the 5' TFO-CPP *rlmN* triplex system results in minimal triplex formation. In comparison, for the 3' TFO-CCP triplex system (Figure 5.5B), two clear charge states for the conjugate triplex can be identified. This therefore demonstrates that the 3' TFO-CPP conjugate has a higher triplex forming propensity than its 5' equivalent. This could be explained by triplex formation occurring by a nucleation-zipping mechanism. It has been speculated from a molecular modelling study that pyrimidine motif triplexes are formed by a nucleation-zipping mechanism whereby the first few base pairs from the 5' end bind favourably, and then allow for the rest of the strand to bind via Hoogsteen bonds in a 5' to 3' direction.[19] Having the CPP attached on the 3' end means that if the triplex forms in a 5' to 3' direction, the majority of the Hoogsteen bonds can form without being potentially hindered by the 3' bound CPP. In contrast, the 5' CPP could be in close enough proximity to the first Hoogsteen bonds in the sequence and negatively impact the most favourable base stacking at this terminus (the binding of the first few base pairs).[20]

To further determine the effect of the PEG4 linker that was used to attach the azide group for the bioconjugation reaction, the KFFKFFKFFK cell-penetrating peptide without the PEG4 linker was also investigated according to the revised reaction scheme shown in Figure 5.6. The azide attachment has now replaced the amine in the side chain of the N-terminal lysine of the CPP.

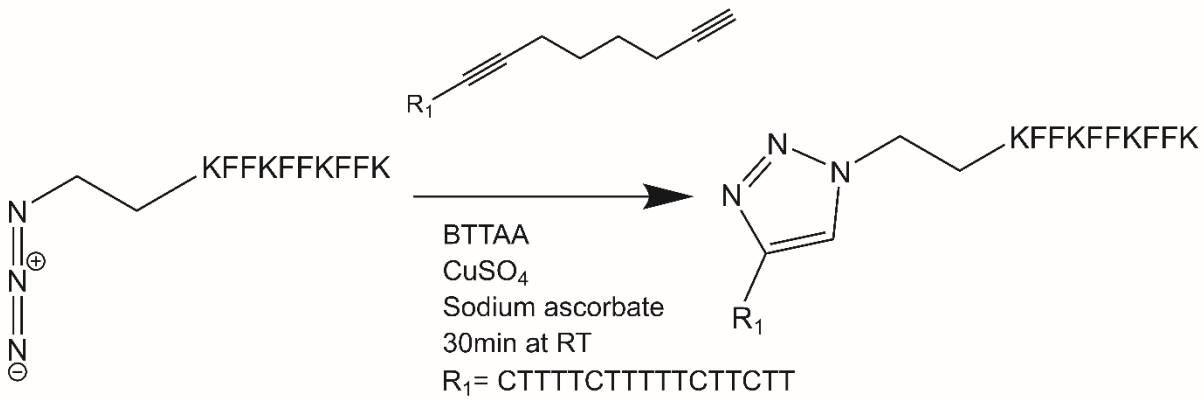


Figure 5.6: Revised copper catalysed [3+2] azide-alkyne cycloaddition reaction which allows for the conjugation of the KFFKFFKFFK peptide with an azide group to the alkyne functionalised *rlmN* TFO.

As the 3' TFO-CPP conjugate was shown to be the more successful in forming triplexes, a third TFO-CPP conjugate was synthesised from the azide-modified peptide without the PEG4 linker and the 3' alkyne-modified *rlmN* TFO, and the triplex formation was again assessed. The MS

data showing successful click reaction and triplex formation using the new TFO-CPP conjugate can be seen in Figure 5.7.

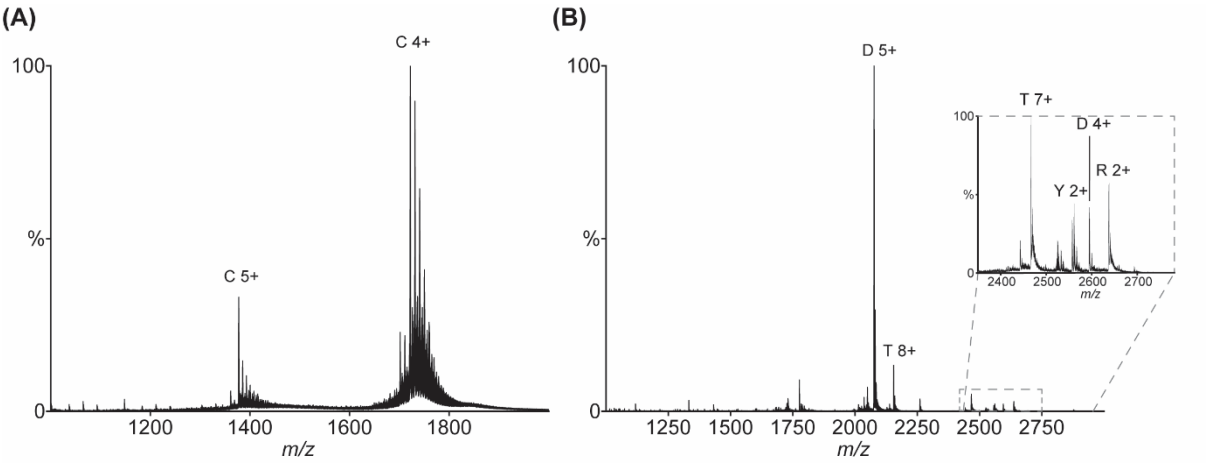


Figure 5.7: (A) Mass spectrum of formation of and 3' *rlmN* TFO-CPP conjugate without the PEG4 linker and **(B)** mass spectrum of 3' TFO-CPP without the PEG4 linker *rlmN* DNA triplex system. Triplex was prepared in a 1:1:1 TFO-cell penetrating peptide conjugate:purine:pyrimidine strand ratio, with each individual strand at a concentration of 2 µM. All samples were in 250 mM ammonium acetate solution at pH 7.4 and all spectra were collected in positive ion mode. Annotations are as follows: Y: pyrimidine single strand, R: purine single strand, C: TFO-CPP conjugate, D: duplex, T: triplex with TFO-CPP conjugate as the third strand. Charge states are also indicated on the labels.

Since triplex formation was shown to varying degrees using the three synthesised conjugates, UV-Vis spectroscopy was utilised to assess the stability of these conjugate triplexes in comparison to the canonical DNA TFO. The UV-Vis spectroscopy derivative plots are shown in Figure 5.8 and the resulting melting temperatures are given in Table 5.2.

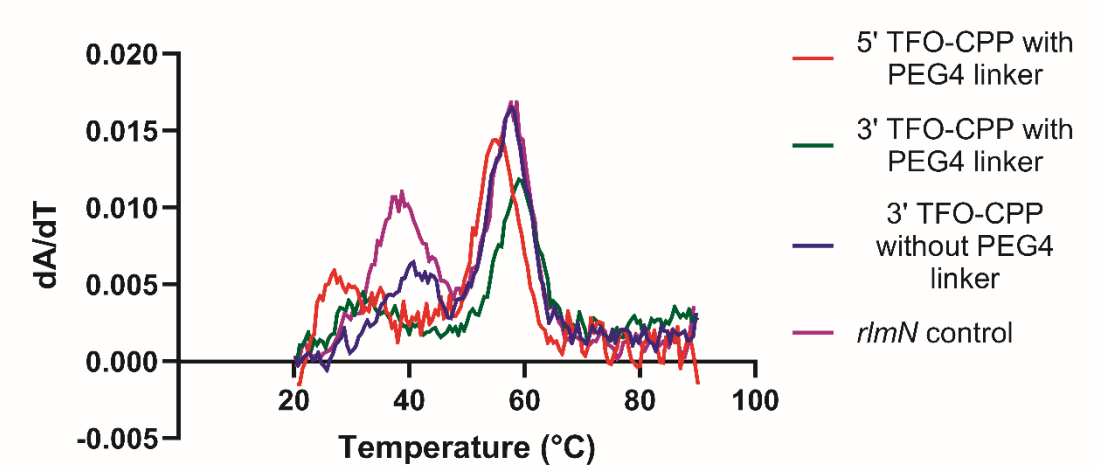


Figure 5.8: Derivative UV-Vis spectroscopy melting profiles of *rlmN* TFO-CPP conjugate triplexes, with the 5' TFO-CPP with PEG4 linker, 3' TFO-CPP with PEG4 linker, 3' TFO-CPP without PEG4 linker and the control *rlmN* triplex systems shown in red, green, blue and purple respectively. All samples were zeroed at 260 nm at 20 °C, and melting was conducted from 20 °C to 90 °C. Samples were prepared in a 1:1:1 TFO:purine:pyrimidine strand ratio, with each individual strand at a concentration of 2 µM.

Table 5.2: Duplex and triplex melting temperatures from the synthetic and control *rlmN* triplex systems at 260 nm using UV-Vis spectroscopy.

***rlmN* triplex system species**	**Melting temperature (°C)**
rlmN DNA duplex	58
5' TFO-CPP with PEG4 Triplex	25
3' TFO-CPP with PEG4 Triplex	33
3' TFO-CPP without PEG4 Triplex	40
Control *rlmN* DNA Triplex	39

The 5' conjugate triplex had the lowest stability of all the synthetic triplexes which is consistent with the MS data in Figure 5.5A. Interestingly, when comparing triplex formation by the two 3' TFO-CPP conjugates, the CPP conjugate without the PEG4 linker formed the more stable triplex structure, with a melting temperature of 40 °C compared to 33 °C. It may be expected that the presence of the linker would allow a more stable triplex structure by extending the distance between the macromolecules and limiting any non-specific binding interactions between the duplex DNA and the CPP that may cause instability. Instead, a possible reason for the linker containing conjugate to be less stable is due to the added flexibility of the azide-PEG4 attachment which increases the degree of freedom of binding by

the TFO. Therefore, when the conjugate binds to the duplex there is a larger entropic penalty due to having more possible conformational states.[21,22] Potential stabilisation of the triplex assembly could also result from additional electrostatic and hydrogen bonding interactions from the positively charged CPP that might occur with the duplex structure from the modified conjugate TFO.[23] Once again, this would be more prominent in the new 3' conjugate as without the longer azide-PEG4 linker the CPP is in closer proximity to the oligonucleotides.

Although the UV-Vis spectroscopy data in Figure 5.8 indicates the 3' conjugate triplex is comparably stable to the DNA-only control TFO, the MS data from Figures 5.2B and Figure 5.7B, suggest that the control DNA TFO triplex forms in a much higher relative abundance than the all variations of the TFO-CPP *rlmN* triplex, however this should be interpreted with caution. It is possible that the observed differences in relative abundance could be due to changes in ionisation efficiency associated with the conjugation of the CPP. Nevertheless, now that the formation of triplex structures using the TFO-CPP conjugate TFOs has been confirmed, it is possible to utilise these to assess antimicrobial potential *in vivo*.

5.3.5 Antimicrobial efficacy of 3' *rlmN* TFO-cell-penetrating peptide conjugate in *S. aureus*

The strongest binding TFO-CPP conjugate for the *rlmN* sequence *in vitro* was found to be the 3' TFO-CPP without the PEG4 linker, so this was selected to assess the antibacterial capacity of triplexes *in vivo*. Antimicrobial activity can be evaluated using concentration dependent growth curves, which provide valuable information on the growth kinetics and allow insight into how bacteria respond to different growth conditions. Antimicrobial activity for the *rlmN* 3' TFO-CPP conjugate without the PEG4 linker, *rlmN* DNA-only TFO and the CPP alone against *S. aureus* is shown in Figure 5.9 below, with kanamycin, and *S. aureus* in M-H broth media used as controls.

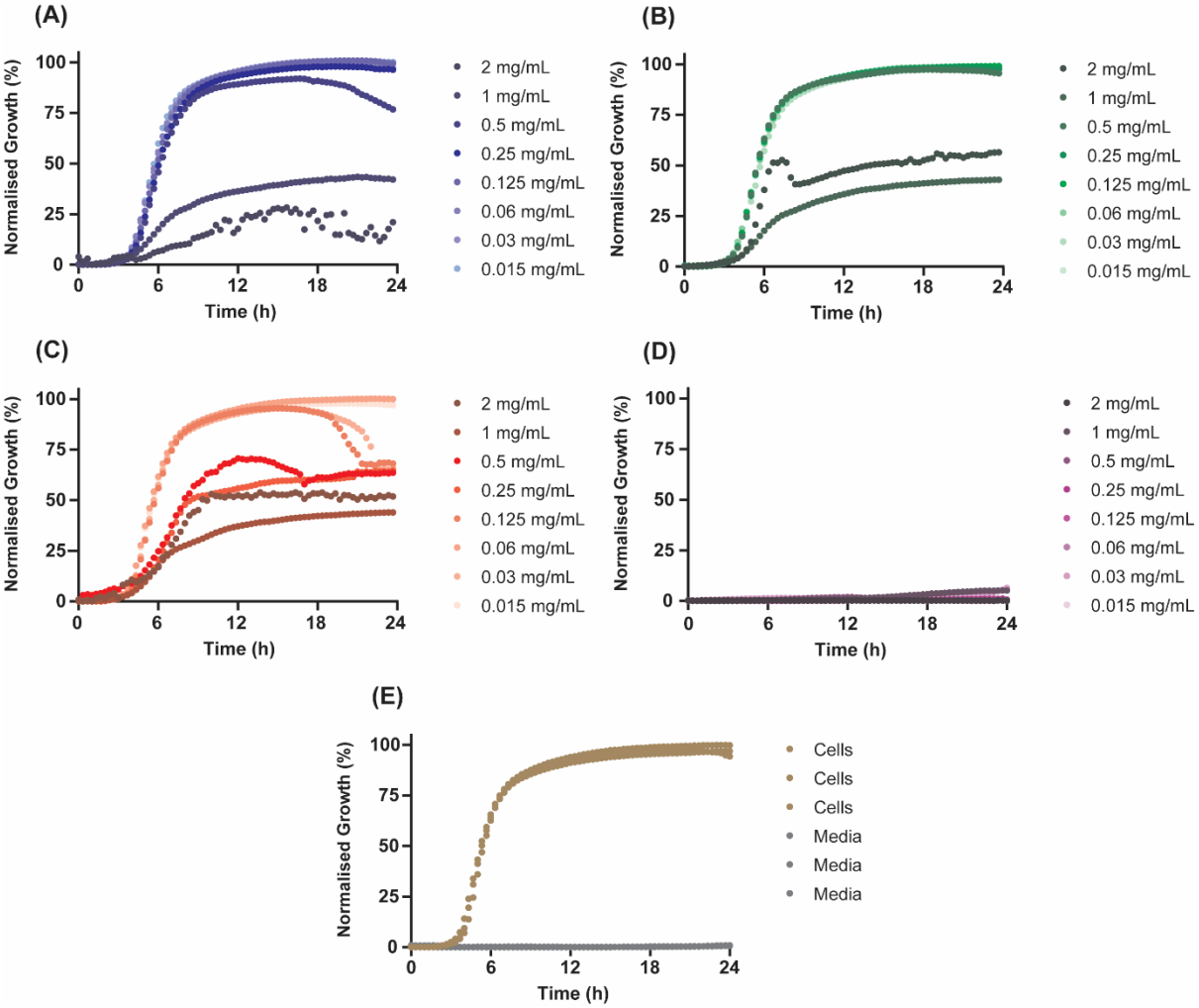


Figure 5.9: Antimicrobial activity of **(A)** *rlmN* 3' TFO-CPP conjugate without the PEG4 linker, **(B)** *rlmN* DNA-only TFO and **(C)** KFFKFFKFFK CPP against *S. aureus*. The growth of *S. aureus* ATC49974 was observed over 24 hours in the presence of serially diluted (2-fold dilutions) test compounds, with concentrations as indicated in the figures. **(D)** Kanamycin and **(E)** *S. aureus* in M-H broth media were also added as experimental control conditions.

It is evident from Figure 5.9A-D that there is concentration-dependent bacterial growth inhibition upon treatment with the test compounds. The 2 mg/mL conditions for the 3' TFO-CPP, *rlmN* DNA-only TFO and kanamycin were plotted separately in Figure 5.10 for ease of comparison.

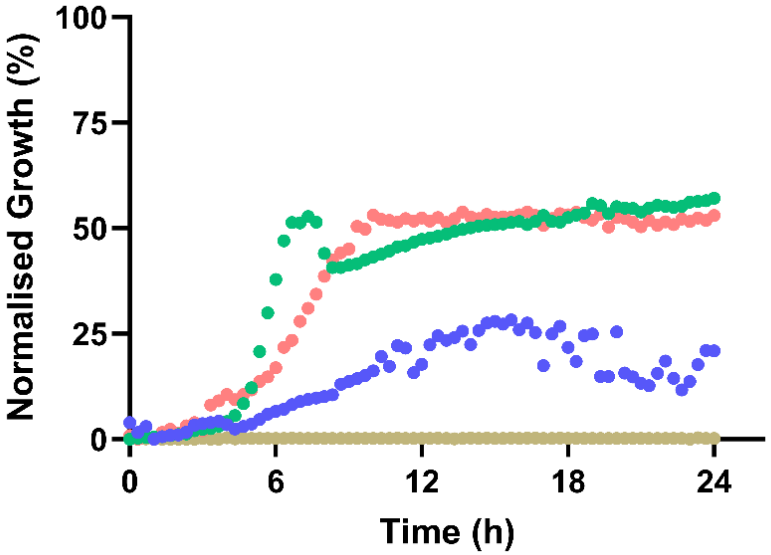


Figure 5.10: Antimicrobial activity of *rlmN* 3' TFO-CPP conjugate without the PEG4 linker (blue), *rlmN* DNA-only TFO (green), KFFKFFKFFK CPP (red) and kanamycin (gold) against *S. aureus*. The growth of *S. aureus* ATC49974 was observed over 24 hours treated with the four compounds at 2 mg/mL.

The 3' TFO-CPP conjugate displays the most potent antimicrobial activity out of the three test compounds at 2 mg/mL. The CPP control also indicated some growth inhibition at the same concentration as the 3' TFO-CPP conjugate, which is likely due to damage to the cells by penetration. However, the antimicrobial activity of the 3' TFO-CPP was still clearly better. The positive control kanamycin was chosen due to being a suitable antibiotic for cell cultures and notably effective against *S.* aureus. Kanamycin treatment was far more efficacious than all of the oligonucleotide based compounds tested in the antimicrobial assay, completely inhibiting *S. aureus* growth at 2 mg/mL. These results indicate that the 3' TFO- CPP conjugate without the PEG4 linker does have potential for antimicrobial activity using the antigene strategy, although it remains limited in activity compared with existing antibiotic agents, and further work is required to determine whether this antimicrobial activity is due to the formation of a triplex at the specific gene target.

5.4 Conclusion

Using Triplexator, we were able to identify numerous potential polypurine TTS located within specific genes in the MRSA*252* genome. This high throughput bioinformatics methodology allowed us to select the GAAAAGAAAAAGAAGAA sequence as a candidate TTS for antigene development, which is located in the promoter region of the critical *rlmN* gene. *In vitro* detection of the *rlmN* triplex using MS and UV-is spectroscopy techniques supported the proposal that the biologically relevant sequence could be used as a potential target in an antigene strategy. To allow for triplex formation *in vivo*, the KFFKFFKFFK CPP was used as a means of transporting the *rlmN* TFO into the cell, and the TFO-CPP conjugate was successfully synthesised using a [3+2] azide-alkyne cycloaddition reaction. Triplex formation was successful with the complementary *rlmN* duplex, with the 3' variant of the conjugate having the highest triplex forming propensity. Antimicrobial studies indicated that the 3' TFO-cell penetrating peptide conjugate does have some antimicrobial capacity that is concentration-dependent. However, the kanamycin positive control was still more potent.

Some possible ways to increase the effectiveness of the antigene strategy would be to determine if the synthesised conjugate is reaching and binding to the target of interest, this could be achieved by using fluorescence resonance energy transfer (FRET) experiments. FRET has been used previously to detect duplex and triplex structures in Chinese hamster ovary and HeLa cells.[24,25] Recent experimental methods for FRET have been developed that allow for triplex formation to be monitored without prelabelling the duplex,[26] utilising a TFO tagged with a fluorophore that undergoes FRET with fluorescent dyes that intercalate with the target duplex. Methods such as these can confirm the formation of triplex using the 3' TFO-CPP conjugate *in vivo,* and correlate this with bacterial cell death. This is particularly important to delineate the effects of the CPP itself, which is also shown to have some antimicrobial activity. Other delivery systems could be potentially used to transport the conjugate into the cell as a more effective way of forming a triplex *in vivo*. Lipid nanoparticles have been of particular interest recently due to their use in the COVID-19 mRNA vaccine,[27] although challenges have arisen with clinical applications, as well as little being known about the encapsulation of DNA using lipid nanoparticles in comparison to mRNA.[28] This makes lipid nanoparticles a potential avenue to explore for future work to deliver TFOs into cells for the altering of gene expression.

Oligonucleotides are also prone to rapid degradation by nucleases in biological conditions which could limit their efficacy as antibacterial agents.[29] More investigation into stabilising DNA triplexes through the use of base and backbone modification would be of interest to help prevent the degradation of the complex by nucleases, allowing the triplex to potentially alter the expression of the target gene.

It is evident that the antigene strategy is an underexplored method of using DNA triplexes for therapeutic applications, with more efforts focussing on the stability of triplex structures *in vivo* potentially allowing for their full biological functionality to be fulfilled.

5.5 References

(1) Watson, J.; Crick, F. Genetical Implications of the Structure of Deoxyribonucleic Acid. *Nature* **1953**, *171*, 964–967.

(2) Crick, F. Central Dogma of Molecular Biology. *Nature* **1970**, *227*, 561–563.

(3) Zhou, Z.; Giles, K.; Felsenfeld, G. DNA·RNA Triple Helix Formation Can Function as a Cis-Acting Regulatory Mechanism at the Human β-Globin Locus. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (13), 6130–6139.

(4) Alecki, C.; Vera, M. Role of Nuclear Non-Canonical Nucleic Acid Structures in Organismal Development and Adaptation to Stress Conditions. *Front. Genet.* **2022**, *13*, 823241.

(5) Buske, F.; Bauer, D.; Mattick, J.; Bailey, T. Triplexator: Detecting Nucleic Acid Triple Helices in Genomic and Transcriptomic Data. *Genome Res.* **2012**, *22*, 1372–1381.

(6) Le, N. Q. K.; Yapp, E. K. Y.; Nagasundaram, N.; Yeh, H.-Y. Classifying Promoters by Interpreting the Hidden Information of DNA Sequences via Deep Learning and Combination of Continuous FastTex N-Grams. *Front. Bioeng. Biotechnol.* **2019**, *7*, 305.

(7) Praseuth, D.; Guieysse, A. L.; Helene, C. Triplex Helix Formation and the Antigene Strategy for Sequence-Specific Control of Gene Expression. *Biochim Biophys. Acta (BBA)- Gene Struct. Expr.* **1999**, *1489* (1), 181–206.

(8) Ricciardi, A. S.; Mcneer, N. A.; Anandalingam, K.; Saltzman, W. M.; Glazer, P. M. Targeted Genome Modification via Triple Helix Formation. *Methods Mol. Biol.* **2014**, *1176*, 89–106.

(9) Wozniak, T. M.; Dyda, A.; Merlo, G.; Hall, L. Disease Burden, Associated Mortality and Economic Impact of Antimicrobial Resistant Infections in Australia. *Lancet Reg. Heal. West. Pacific* **2022**, *27*, 100521.

(10) Li, H.-Y.; Yang, W.-Q.; Zhou, X.-Z.; Shao, F.; Shen, T.; Guan, H.-Y.; Zheng, J.; Zhang, L.-M. Antibacterial and Antifungal Sesquiterpenoids: Chemistry, Resource, and Activity. *Biomolecules* **2022**, *12* (9).

(11) Enright, M. C.; Robinson, A.; Randle, G.; Feil, E.; Grundmann, H.; Spratt, B. G. The Evolutionary History of Methicillin-Resistant Staphylococcus Aureus (MRSA). *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (11), 7687–7692.

(12) Algammal, A. M.; Hetta, H. F.; Elkelish, A.; Alkhalifah, D. H. H.; Hozzein, W. N.; Batiha, G. E.-S.; Nahhas, N. El; Mabrok, M. A. Methicillin-Resistant Staphylococcus Aureus

(MRSA): One Health Perspective Approach to the Bacterium Epidemiology, Virulence Factors, Antibiotic-Resistance, and Zoonotic Impact. *Infect. Drug Resist.* **2020**, *13*, 3255–3265.

(13) Holden, M.; Feil, E.; Lindsay, J.; Peacock, S.; Day, N.; Enright, M.; Foster, T.; Moore, C.; Hurst, L.; Atkin, R.; Andrew, B.; Bason, N.; Bentley, S.; Chillingworth, C.; Chillingworth, T.; Churcher, C.; Clark, L.; Corton, C.; Cronin, A.; Doggett, J.; Dowd, L.; Feltwell, T.; Hance, Z.; Harris, B.; Hauser, H.; Holroyd, S.; Jagels, K.; James, K.; Ormond, D.; Quail, M.; Rabbinowitsch, E.; Rutherford, K.; Sanders, M.; Sharp, S.; Simmonds, M.; Stevens, K.; Whitehead, S.; Barell, B.; Spratt, B.; Parkhill, J. Complete Genomes of Two Clinical Staphylococcus Aureus Strains: Evidence for the Rapid Evolution of Virulence and Drug Resistance. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (26), 9786–9791.

(14) Consortium, T. U. UniprotKb: A Worldwide Hub of Protein Knowledge. *Nucleic Acids Res.* **2019**, *47* (1), 506–515.

(15) Altschul, S.; Gish, W.; Miller, W.; Myers, E.; Lipman, D. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215* (3), 403–410.

(16) Knauert, M. P.; Glazer, P. M. Triplex Forming Oligonucleotides: Sequence-Specific Tools for Gene Targeting. *Hum. Mol. Genet.* **2001**, *10* (20), 2243–2251.

(17) Nekhotiaeva, N.; Awasthi, S. K.; Nielson, P. E.; Good, L. Inhibition of Staphylococcus Aureus Gene Expression and Growth Using Antisense Peptide Nucleic Acids. *Mol. Ther.* **2004**, *10* (4), 652–659.

(18) Zaro, J. L.; Vekich, J. E.; Tran, T.; Shen, W.-C. Nuclear Localization of Cell-Penetrating Peptides Is Dependent on Endocytosis Rather than Cytosolic Delivery in CHO Cells. *Mol. Pharm.* **2009**, *6* (2), 337–344.

(19) Alberti, P.; Arimondo, P. B.; Mergny, J.-L.; Garestier, T.; Hélène, C.; Sun, J.-S. A Directional Nucleation-Zipping Mechanism for Triple Helix Formation. *Nucleic Acids Res.* **2002**, *30* (24), 5407–5415.

(20) Gaied, N. Ben; Zhao, Z.; Gerrard, S. R.; Fox, K.; Brown, T. Potent Triple Helix Stabilization by 5', 3'-Modified Triplex-Forming Oligonucleotides. *ChemBioChem* **2009**, *10* (11), 1839–1851.

(21) Einav, T.; Yazdi, S.; Coey, A.; Bjorkman, P. J.; Phillips, R. Harnessing Avidity: Quantifying the Entropic and Energetic Effects of Linker Length and Rigidity for Multivalent Binding of Antibodies to HIV-1. *Cell Syst.* **2019**, *9* (5), 466–474.

(22) Kane, R. S. Thermodynamics of Multivalent Interactins: Influence of the Linker. *Langmuir* **2010**, *26* (11), 8636–8640.

(23) Kolchina, N.; Khavinson, V.; Linkova, N.; Yakimov, A.; Baitin, D.; Afanasyeva, A.; Petukhov, M. Systematic Search for Structual Motifs of Peptide Binding to Double-Stranded DNA. *Nucleic Acids Res.* **2019**, *47* (20), 10553–10563.

(24) Bamford, R. A.; Zhao, Z.; Hotchin, N. A.; Styles, I. B.; Nash, G. B.; Tucker, J. H. R.; Bicknell, R. Electroporation and Microinjection Successfully Deliver Single-Stranded and Duplex DNA into Live Cells as Detected by FRET Measurements. *PLoS One* **2014**, *9* (4), e95097.

(25) Mahani, M.; Taheri, M.; Divsar, F.; Khakbaz, F.; Nomani, A.; Ju, H. Label-Free Triplex DNA-Based Biosensing of Transcription Factor Using Fluorescence Resonance Energy Transfer between N-Doped Carbon Dot and Gold Nanoparticle. *Anal. Chim. Acta* **2021**, *1181*, e338919.

(26) Chiou, C.-C.; Chen, S.-W.; Luo, J.-D.; Chien, Y.-T. Monitoring Triplex DNA Formation with Fluorescence Resonance Energy Transfer between a Fluorophore-Labeled Probe and Intercalating Dyes. *Anal. Biochem.* **2011**, *416* (1), 1–7.

(27) Anderson, E. J.; Rouphael, N. G.; Widge, A. T.; Jackson, L. A.; Roberts, P. C.; Makhene, M.; Chappel, J. D.; Denison, M. R.; Stevens, L. J.; Pruijssers, A. J.; McDermott, A. B.; Flach, B.; Lin, B. C.; Doria-Rise, N. A.; O'Dell, S.; Schmidt, S. D.; Corbett, K. S.; Swanson, P. A.; Padilla, M.; Neuzil, K. M.; Bennett, H.; Leav, B.; Makowski, M.; Albert, J.; Cross, K.; Edara, V. V.; Floyd, K.; Suthar, M. S.; Martinez, D. R.; Baric, R.; Buchanan, W.; Luke, C. J.; Phadke, V. K.; Rostad, C. A.; Ledgerwood, J. E.; Graham, B. S.; Beigel, J. H.; Group, mRNA-1273 S. Safety and Immunogenicity of SARS-Cov-2 MRNA-1273 Vaccine in Older Adults. *N. Engl. J. Med.* **2020**, *25*, 2427–2438.

(28) Cui, L.; Renzi, S.; Quagliarini, E.; Digiacomo, L.; Amenitsch, H.; Masuelli, L.; Bei, R.; Ferri, G.; Cardarelli, F.; Wang, J.; Amici, A.; Pozzi, D.; Marchini, C.; Caracciolo, G. Efficient Delivery of DNA Using Lipid Nanoparticles. *Pharmaceutics* **2022**, *14* (8), 1698.

(29) Monia, B. P.; Johnston, J. F.; Sasmor, H.; Cummins, L. L. Nuclease Resistance and Antisense Activity of Modified Oligonucleotides Targeted to Ha-Ras*. *J. Biol. Chem.* **1996**, *271* (24), 14533–14540.

5.6 Appendix 1: Input and summary from Triplexator

```
************************************************************
*** Triplexator - Finding nucleic acid triple helices    ***
***       (c) Copyright 2011 by Fabian Buske          ***
***    Comments, Bugs, Feedback: f.buske@uq.edu.au     ***
************************************************************
*** Version 1.3.2 (30/03/2012) SeqAn Revision: 1225
*** COMMAND:
>triplexator -l 15 -e 0 -fr off -g 20 -m R -fm 0 -of 0 -o MRSA252test3.1 -po -rm 2 -p 3 -ds MRSA252.fna
************************************************************
*** INPUT:
- single-stranded file supplied : No
- duplex file supplied : Yes
-> search putative triplex target sites
************************************************************
*** Output Options:
- output directory : ./
- output file : MRSA252test3.1
- output format : 0 = Triplex
- merge features : No
- report duplicate locations : No
- error reference : 0 = Watson strand (TTS)
# Duplex-ID        TTSs (abs)          TTSs (rel)
NC_002952.2       954                 2.05e-05

************************************************************
*** Main Options:
- maximum error-rate : 0%
- maximum total error : 0
- minimum guanine content with respect to the target : 20%
- maximum guanine content with respect to the target : 100%
- minimum length : 15 nucleotides
- maximum length : 30 nucleotides
- maximum number of tolerated consecutive pyrimidine interruptions in a target: 0
- number of consecutive matches required in a feature : 1
- longest match only : yes
- detect duplicates : 0 = off
- same sequence duplicates : on
************************************************************
*** Filtration Options :
- filter repeats : No
- duplicate cutoff : -1
************************************************************
*** Runtime mode:
- OpenMP support : Yes
- runtime mode : 2 = parallel (duplex sequences) - 1 processors
************************************************************
*** Log messages:
[02/23/21 19:48:21] * Started searching for triplex target sites
[02/23/21 19:48:21] * Processing MRSA252.fna
[02/23/21 19:48:21] * Starting search for targets in parallel mode
[02/23/21 19:48:21] * Finished reading 1 duplex sequences
[02/23/21 19:48:21] * Detecting triplex targets in all duplexes (1 threads)
[02/23/21 19:48:22] * Finished detecting targets
[02/23/21 19:48:22] * Finished processing MRSA252.fna
```

[02/23/21 19:48:22] * Finished searching for triplex target sites within 0.474 seconds.
[02/23/21 19:48:22] * Exit without errors
[02/23/21 19:48:22] * Finished program within 0.488 seconds

5.7 Appendix 2: *MRSA* TTSs identified from triplexator, along with candidate *rlmN* gene triplex sequence

Table A1: Triplex target sites with no errors obtained from Triplexator from the *MRSA252* genome, this data also indicates the genome ID, the positioning in the genome and the length of the TTS.

# Duplex-ID	Start	End	Length	Strand	Guanine rate	TTS
NC_002952.2	22468	22484	16	+	0.31	GAAGGAAAAGGAAAAA
NC_002952.2	33823	33838	15	+	0.27	AAAAGAAAAAGAAGG
NC_002952.2	53524	53539	15	+	0.33	GGAAGAAGAAAAAGA
NC_002952.2	76984	76999	15	+	0.47	AGAGAAAGAGGGAAG
NC_002952.2	87483	87498	15	+	0.47	GAGAAGAGAAGGAAG
NC_002952.2	88203	88219	16	+	0.25	AAGGAAAAGAAGAAAA
NC_002952.2	111457	111475	18	+	0.28	AGAGAAAAAAGGAGAAAA
NC_002952.2	132615	132630	15	+	0.27	AAAGAAAAAAGAGGA
NC_002952.2	160979	160995	16	+	0.31	GGAGAAAAGAAGAAAA
NC_002952.2	227189	227210	21	+	0.29	GAAAAGAAGAAAGAAAAAGAG
NC_002952.2	257628	257645	17	+	0.24	AAAGAAGAAAAGAAAGA
NC_002952.2	259341	259358	17	+	0.29	AAAAGAAGAGAAGAAAG
NC_002952.2	334976	334992	16	+	0.25	AAAGAAAGAGAAAAAG
NC_002952.2	335133	335151	18	+	0.44	GGAAAGGAGAAAGAAGGA
NC_002952.2	385369	385394	25	+	0.28	GAAAGAAGAGAAAAAGGAAAAAGAA
NC_002952.2	426622	426638	16	+	0.38	AAAGAAGGGAAAGAGA
NC_002952.2	426936	426954	18	+	0.33	AAAAAAAGGAAGAGAGAG
NC_002952.2	462973	462988	15	+	0.53	GGGAAGAGAAGAAGG
NC_002952.2	503138	503153	15	+	0.53	AAGGAAAGGGGGAGA
NC_002952.2	516441	516459	18	+	0.39	GGAGGAAGAGAAAGAAAA
NC_002952.2	527694	527713	19	+	0.26	GAAAAGAGAAAGAAAAAAG
NC_002952.2	530029	530048	19	+	0.37	GAAGAGGAAAAAGAAGAGA
NC_002952.2	533911	533927	16	+	0.31	GGAAAAGGAGAAAAAA
NC_002952.2	546609	546624	15	+	0.47	AGAAAAAGGGGAGAG
NC_002952.2	562203	562221	18	+	0.39	GGAGGAAGAGAAAGAAAA
NC_002952.2	566943	566960	17	+	0.47	GAAGGAGAGGAAAGAAG
NC_002952.2	577849	577864	15	+	0.27	GGAAAAGAAAGAAAA
NC_002952.2	582065	582087	22	+	0.41	AGGGAAAGAAAGAGGGAAAAGA
NC_002952.2	584095	584111	16	+	0.25	AGAAAAAAGAGAAGAA
NC_002952.2	584431	584447	16	+	0.25	AAAAAGAAAAAGAAGG
NC_002952.2	633493	633508	15	+	0.67	GGGGGAAAGAGGGGA
NC_002952.2	665614	665629	15	+	0.53	AAGGGGGGAAAGGAA
NC_002952.2	739660	739676	16	+	0.31	AAAAGGGAAAAAAGGA
NC_002952.2	750619	750634	15	+	0.33	AAGGAGAAAAAGAAG
NC_002952.2	810025	810041	16	+	0.31	AAAGAAAGAGAAGAAG
NC_002952.2	813131	813148	17	+	0.24	AAAAGAAGAGAAAAAAG
NC_002952.2	822856	822873	17	+	0.24	GAGAAAAAGAAAAAGAA
NC_002952.2	851163	851180	17	+	0.29	AAAGGAAAAGAGGAAAA
NC_002952.2	863372	863387	15	+	0.27	AAAAAAGAAGGGAAA
NC_002952.2	870586	870608	22	+	0.27	AGAAAAAGAGAGAAAGAAAAAG
NC_002952.2	885014	885030	16	+	0.25	AGAAAAAGAAAAGAGA
NC_002952.2	886921	886936	15	+	0.6	AGAAAGGGAGGGGGA

NC_002952.2	892162	892184	22	+	0.41	GAAGAGAGAAAAAAGAAGAGGG
NC_002952.2	925008	925024	16	+	0.31	GAAGAAGAAGGAAAAA
NC_002952.2	979823	979838	15	+	0.33	GAGAAGAAAAAGGAA
NC_002952.2	995927	995943	16	+	0.5	AAGGGGAAGGGAGAAA
NC_002952.2	1034659	1034675	16	+	0.44	AAAGAAGAGGAAGAGG
NC_002952.2	1049998	1050013	15	+	0.47	GAAAGAGGGGAAAAG
NC_002952.2	1088199	1088217	18	+	0.28	AGAAGGAGAAGAAAAAAA
NC_002952.2	1151846	1151865	19	+	0.21	AAAAAGGAAAAGAAAAAAG
NC_002952.2	1218411	1218427	16	+	0.44	AAAAAGGAGAAGGAGG
NC_002952.2	1235193	1235209	16	+	0.25	AAGAAAGGAAAAAAAG
NC_002952.2	1238300	1238315	15	+	0.4	GGAAGAAGAGAAAGA
NC_002952.2	1242522	1242539	17	+	0.24	GAAAAGAAAAAGAAGAA
NC_002952.2	1244486	1244502	16	+	0.25	AGAGAAAAAGAAGAAA
NC_002952.2	1300836	1300853	17	+	0.24	AGAAAAAGAAAAGAAGA
NC_002952.2	1307144	1307160	16	+	0.38	GAAAAAAGGGGAAGAA
NC_002952.2	1308508	1308523	15	+	0.33	AAAAAGAGAGAAGGA
NC_002952.2	1309399	1309414	15	+	0.27	GAAGAAGAAAAAAGA
NC_002952.2	1312604	1312619	15	+	0.33	GAGGAAAAAGAAGAA
NC_002952.2	1317144	1317161	17	+	0.29	GAAAAAGAAGAAGAAGA
NC_002952.2	1324327	1324346	19	+	0.21	AAAGAAAAAGAAAAAGAAG
NC_002952.2	1325589	1325607	18	+	0.22	AAAAAAAAGAAGAGAAGA
NC_002952.2	1351174	1351191	17	+	0.29	AGAAGAAGAAAAAGAAG
NC_002952.2	1364803	1364823	20	+	0.3	AAAGAGGAAAGGAGAAAAAA
NC_002952.2	1369329	1369346	17	+	0.29	GGAAAGAGAAGAAAAAA
NC_002952.2	1370795	1370810	15	+	0.4	AGAAAAAGGGAAAGG
NC_002952.2	1372742	1372757	15	+	0.33	AAGGGAAAAGAGAAA
NC_002952.2	1374286	1374301	15	+	0.33	AAGAGAAGGGAAAAA
NC_002952.2	1378000	1378015	15	+	0.47	AGAGAAAAGAGGGGA
NC_002952.2	1396479	1396501	22	+	0.32	AAAAGAGAGAAAAAGAGAAGAG
NC_002952.2	1407010	1407025	15	+	0.27	GAAAGAAGAAAAAGA
NC_002952.2	1430484	1430502	18	+	0.22	GAAAAAAGAAGAAGAAAA
NC_002952.2	1461125	1461141	16	+	0.25	GAAAAAAAGGAGAAAA
NC_002952.2	1618807	1618822	15	+	0.53	GAGAAGGGAAAGAGG
NC_002952.2	1632450	1632466	16	+	0.25	AAGAAGAAGAAAAAAG
NC_002952.2	1634786	1634803	17	+	0.24	AAAAAAGGAAAAGAAGA
NC_002952.2	2166290	2166307	17	+	0.29	GAAGAAAGAAAAGGAAA
NC_002952.2	2166986	2167003	17	+	0.35	GAGAAGGAAAAGGAAAA
NC_002952.2	2312147	2312163	16	+	0.31	GAAGAGAAAAAAGGAA
NC_002952.2	2442280	2442298	18	+	0.28	AAAAAAGGAGGAAAAAGA
NC_002952.2	2529722	2529741	19	+	0.32	AAAAGAGAAAAAGAGAGGA
NC_002952.2	2543166	2543181	15	+	0.27	GAAAAAGAAAAAAGG
NC_002952.2	2748300	2748316	16	+	0.25	GAAGAAAAAGAGAAAA
NC_002952.2	2763423	2763441	18	+	0.33	AGAAAGGAAAGAAGAAGA
NC_002952.2	38054	38069	15	-	0.33	GGAGAGAAAAGAAAA
NC_002952.2	45966	45984	18	-	0.22	AGAGAAAAAGAAAAAAGA
NC_002952.2	68217	68233	16	-	0.31	AAAGAAAAAGGAGGAA
NC_002952.2	76706	76721	15	-	0.27	AAAAAAAGAGGGAAA
NC_002952.2	84417	84435	18	-	0.44	GAGGAAGAAAAAAGGAGG
NC_002952.2	129425	129442	17	-	0.35	AGGGAAAGAAAAAGAAG

NC_002952.2	397351	397366	15	-	0.33	GAAAAGAAGAAGAAG
NC_002952.2	438611	438627	16	-	0.31	AAGAAGAGAAAAAAGG
NC_002952.2	738177	738192	15	-	0.47	GAGAAAAAGAGAGGG
NC_002952.2	782142	782157	15	-	0.33	AGAAAAAGGAGAAGA
NC_002952.2	885496	885516	20	-	0.3	GAGAGAAAGAAAAAGAAAAG
NC_002952.2	885547	885566	19	-	0.21	AGAAAAGAAAGAAAAAGAA
NC_002952.2	1074849	1074866	17	-	0.35	GGAAGAAGAAAAAAAGG
NC_002952.2	1180452	1180467	15	-	0.33	GGGAAAAAGAAGAAA
NC_002952.2	1564871	1564888	17	-	0.24	GAAAAAAAGAAGAAAGA
NC_002952.2	1578578	1578593	15	-	0.4	AAGAAAGAGAGGAAG
NC_002952.2	1611739	1611755	16	-	0.31	GAAGAGAAAGAAAAAG
NC_002952.2	1612001	1612025	24	-	0.21	AGAAGAAAAAGAAAAAAGAGAAAA
NC_002952.2	1612000	1612024	24	-	0.21	GAAGAAAAAGAAAAAAGAGAAAAA
NC_002952.2	1617584	1617601	17	-	0.24	AGAAAAAGAAAGAAAAG
NC_002952.2	1625222	1625243	21	-	0.33	GAAGAAGAAAGAAAGAGAAGA
NC_002952.2	1625296	1625318	22	-	0.27	GAAGAAAAAGAGAGAGAAAAAA
NC_002952.2	1628095	1628111	16	-	0.31	AAAGAAGAAGGAAAAG
NC_002952.2	1630047	1630063	16	-	0.25	AAGAAAAAAGGAAAGA
NC_002952.2	1644909	1644924	15	-	0.33	GAAAGAAGAAAGAGA
NC_002952.2	1652173	1652188	15	-	0.53	GAAGAAAAGGAGGGG
NC_002952.2	1690101	1690116	15	-	0.47	GAAAAGAGAGGGAAG
NC_002952.2	1717027	1717042	15	-	0.27	AGAGAAAAAGAAAAG
NC_002952.2	1717312	1717327	15	-	0.27	GAAGAAGAAAAGAAA
NC_002952.2	1731319	1731334	15	-	0.27	AAAAAGAAGAGAGAA
NC_002952.2	1732310	1732327	17	-	0.24	AGAAAAGAAAGAAAGAA
NC_002952.2	1767630	1767647	17	-	0.24	GAAGAGAAAAAAGAAAA
NC_002952.2	1778985	1779000	15	-	0.33	AGAAAAGAAGGGAAA
NC_002952.2	1819465	1819481	16	-	0.44	GGGAAAAAAAGGAAGG
NC_002952.2	1910763	1910778	15	-	0.47	GGAGGAAAAGGAAGA
NC_002952.2	1935317	1935333	16	-	0.31	AAAAAAGAGAGAGGAA
NC_002952.2	1971507	1971524	17	-	0.41	AGGGGAAAGAAAAGGAA
NC_002952.2	1983249	1983264	15	-	0.4	GGGAAGAGAAGAAAA
NC_002952.2	2003775	2003791	16	-	0.31	AAAGAAAAAGAAGAGG
NC_002952.2	2017096	2017113	17	-	0.47	GAGAGAAGAAGGGAAAG
NC_002952.2	2025427	2025442	15	-	0.27	GAAAGAAAAAGAAGA
NC_002952.2	2039101	2039119	18	-	0.39	GGAGGAAGAGAAAGAAAA
NC_002952.2	2145035	2145053	18	-	0.28	GGAGAAAAAGAAGAAAAA
NC_002952.2	2150758	2150773	15	-	0.27	AAAGAAAAAGAAAGG
NC_002952.2	2213601	2213618	17	-	0.47	GAAGGAGAGGAAAGAAG
NC_002952.2	2216656	2216674	18	-	0.39	GGAGGAAGAGAAAGAAAA
NC_002952.2	2320403	2320420	17	-	0.29	GAAGAAAAAGGAAAAGA
NC_002952.2	2335065	2335083	18	-	0.39	GGAGGAAGAGAAAGAAAA
NC_002952.2	2338694	2338709	15	-	0.27	AAGGAAGAAAAAGAA
NC_002952.2	2357401	2357419	18	-	0.33	GAAAAAGAAAGAGAAGAG
NC_002952.2	2385580	2385597	17	-	0.24	AAGGAAAAAAAGAGAAA
NC_002952.2	2458474	2458492	18	-	0.5	GGGGAAGAAAAGGAAGAG
NC_002952.2	2484684	2484699	15	-	0.27	AAAGGAAAAGAAAAG
NC_002952.2	2511589	2511612	23	-	0.22	GAAAAGAAAGAAAAAGAAGAAAA
NC_002952.2	2576876	2576894	18	-	0.28	AAGGAAAGAAAAAAGAAG

NC_002952.2	2637102	2637120	18	-	0.28	AGAGGAGAAAAAAGAAAA
NC_002952.2	2653086	2653101	15	-	0.27	GAAAAAAGAAGAAAG
NC_002952.2	2665564	2665583	19	-	0.42	GAAAAGAGAAAGAGGGGAA
NC_002952.2	2667236	2667251	15	-	0.33	GAAAAGAAAGAAGGA
NC_002952.2	2724115	2724132	17	-	0.24	AAAAAGAAAAAGAGAGA
NC_002952.2	2731914	2731931	17	-	0.41	GGAGAAGAGGAAAAAAG
NC_002952.2	2737755	2737774	19	-	0.37	GAGAAGAAAGAGAAGAAGA
NC_002952.2	2812115	2812130	15	-	0.47	AGGAAAAGAGGGAGA
NC_002952.2	2839061	2839082	21	-	0.29	GAAGAGAAAGAAAAAGAAAGA
NC_002952.2	2901766	2901784	18	-	0.44	AGGGGAAAGAGAGGAAAA

Table A2: The name and sequence composition of the strands involved in the 17mer *rlmN* DNA triplex system

***rlmN* sequence names**	**Sequence composition (5'—3')**
Duplex purine strand	GAAAAGAAAAAGAAGAA
Duplex pyrimidine strand	TTCTTCTTTTTCTTTTC
TFO	CTTTTCTTTTTCTTCTT

Chapter 6: Binding of anthocyanins to DNA triplexes

6.1 Introduction

The interactions between small molecules and DNA assemblies are the foundation for many therapeutics, and studies of such systems have been valuable for understanding their critical roles in biology, which include regulation of DNA replication and gene expression.[1–4] Most small molecule binders can bind to different DNA assemblies through intercalation or minor/major groove binding, via electrostatic, base stacking, π-π and hydrogen bonding interactions.[5] Early investigations into the binding to triple helical systems indicated that molecules with extensive ring systems could intercalate between the DNA bases. Examples of these include natural products and drugs such as the anticancer active indoloquinoline.[6,7] Minor structural modifications of the small molecule binders, including size, shape and charge, can drastically change the binding affinity and binding modes to triplexes. This has been shown in previous investigations with small molecules using CD, UV-vis spectroscopy and molecular modelling.[8,9] However more recently, several methods for screening natural plant extracts using mass spectrometry (MS) have been developed that can identify small molecule triplex binders.[10,11] This rapid approach enables small molecule binders to be identified from complex samples, and could potentially be used to improve drug discovery methods.[12]

Flavonoids are a group of natural products that are present in many plants, fruits and vegetables. These small molecules are appealing for drug discovery due to their potential anticancer, antiviral and antioxidant activity and diversity of structures. The non-covalent interactions between flavonoids and higher-order DNA assemblies are of particular interest due to their potential roles in modulating DNA functionality.[13–15] Early studies into the interactions between flavonoids and triplexes have used a combination of ESI-MS, UV-vis spectroscopy and isothermal titration calorimetry methods.[16,17]

Anthocyanins are a subclass of flavonoids that are composed of a 2-phenylbenzopyrylium ion core, which can be seen in Figure 6.1. They are water-soluble pigments that are responsible for the red, blue, and purple colour in plant tissues, and have been studied for their diverse biological effects, which include antioxidant, anti-inflammatory, anticancer and antimicrobial properties.[18–20]

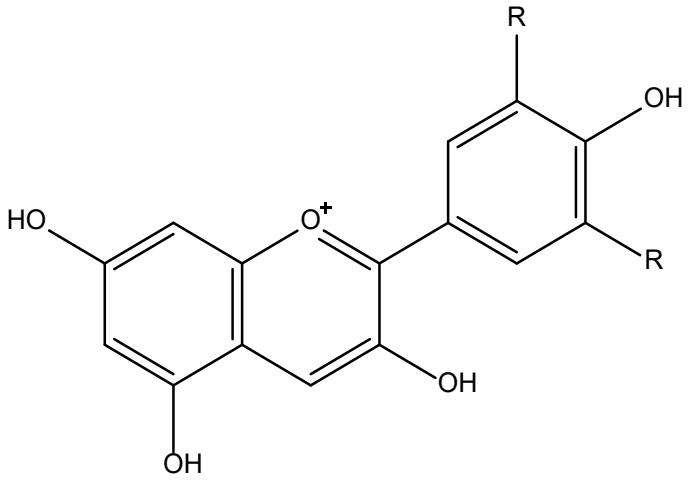


Figure 6.1: Basic structure of an anthocyanin, containing a 2-phenylbenzopyrylium ion core and polyhydroxylated backbone. Anthocyanins show significant diversity in structures, with different combinations of sugars and/or acyl groups in positions that are represented by the R groups.

Since anthocyanins are well known for their biological functionality, their interactions with higher-order DNA structures are of interest for potential *in vivo* applications. Like most other flavonoids, anthocyanins are known to have an intercalating mode of binding to DNA.[21,22] The interactions between anthocyanins have been briefly explored in literature and have shown to have some triplex stabilisation abilities, although these were only based on short and non-biologically relevant triplex sequences utilising solution-phase methods for analysis.[22] Anthocyanin-bound triplex DNA complexes have not been identified by MS, leaving their interactions with DNA relatively underexplored. Further study into the binding of these molecules using sequences with different base compositions can help identify any sequence specificity towards different triplexes, which would allow selective targeting for biological applications. There is also very limited information on how the different substituents on anthocyanins affect their intercalation with DNA triplexes. There is some evidence to suggest monoglucosides have some slight stabilising effects, whereas disubstituted anthocyanins showed no difference to triplex stability, and it was proposed that they did not interact with triplexes due to steric crowding from the second glucoside.[22] There is therefore still much more to uncover about the interactions between anthocyanins and DNA triplexes due to the diverse structural aspects of the anthocyanins.

This chapter investigates the binding of two different anthocyanins, cyanidin 3-glucoside and malvidin 3,5-diglucoside (structures shown in Figure 6.2) to two model DNA triplex sequences,

$(T)_{30}$ and $(CT)_{15}$, as well as the *rlmN* gene sequence that is found in *Staphylococcus aureus* using MS and UV-vis spectroscopy. The binding affinity of the two anthocyanins to the different triplex assemblies is assessed to determine if the individual anthocyanins show any selectivity towards the specific sequences. The binding of the known duplex DNA binding molecule netrospin is also compared, in attempts to speculate the binding positioning of the anthocyanins in the triplex DNA sequences.

6.2 Experimental methods

6.2.1 Sample preparation

Oligonucleotides were acquired from Sigma Aldrich at desalted purity and dissolved in deionized water (18 MΩ/cm) to a concentration of approximately 1 mM. The concentration of each oligonucleotide solution was determined by measuring the absorbance at 260 nm on a NanoDrop™ 2000 UV-Vis spectrophotometer (ThermoFisher Scientific). Triplicate measurements were taken for each sample, using extinction coefficients as provided by the manufacturer (Sigma Aldrich). Ammonium acetate solutions were prepared from a 7.5 M aqueous stock and adjusted to the desired pH using ammonium hydroxide with reagents procured from Sigma Aldrich. Cyanidin 3-glucoside and malvidin 3,5-diglucoside were purchased from Phytolab (Vestenbergsgreuth, Germany). Netropsin dihydrochloride was purchased from Santa Cruz Biotechnology.

Duplex DNA was prepared by mixing constituent oligonucleotides in 250 mM ammonium acetate, resulting in stock solutions at a concentration of 20 µM. These stocks were allowed to incubate for 2 hours at 4 °C to ensure duplex formation. Subsequently, triplex samples were prepared by adding an equal amount of TFO to an aliquot of the duplex. Following a 2-hour incubation to facilitate triplex formation, samples were diluted to 2 µM in the desired buffer before analysis. The TFO loading of a given sample was equimolar in all cases. Prior to dilution, the anthocyanin standards (40 µM, 5 µL) were added to the triplex sample and left to incubate for a further 2 hours before MS analysis. The anthocyanin triplex solution was then filtered through a vivaspin 500 centrifugation unit with a 3000-MW cut-off and reconcentrated back to the original concentration (2 µM) with ammonium acetate pH 7.4 as per the manufacturer's protocol prior to analysis. For netropsin experiments, netrospin (20 µM) was added to the sample at the same time as the anthocyanin standards followed by the use of a vivaspin 500 centrifugation unit as described.

For DNA triplex sequence binding affinity experiments, 2 µM of each triplex $(T)_{30}$, $(CT)_{15}$ and *rlmN* were added to 4 µM of each anthocyanin separately and incubated for 2 hours before MS analysis. For the anthocyanin binding affinity experiments, 4 µM of both cyanidin 3-glucoside and malvidin 3,5-diglucoside were added to each of the $(T)_{30}$, $(CT)_{15}$ and *rlmN* triplexes respectively.

6.2.2 Native mass spectrometry

Mass spectra were acquired using a Bruker Impact HD Q-TOF spectrometer, fitted with a nanospray ion source. The capillary housing was modified such that a backing pressure of N_2 gas could be applied to the sample to stabilise ion spray. Key instrument parameters were: End Plate Offset 500 V, Capillary 1800 V, Dry Gas 6.0 $L \cdot min^{-1}$, Dry Temp 140 °C, Funnel 1 RF 400.0 VPP, Funnel 2 RF 600.0 VPP, Hexapole RF 300.0 VPP, Quadrupole Ion Energy 4.0 eV. Collision Cell parameters were: Collision RF 1200.0 VPP, Transfer Time 120.0 µs, Pre Pulse Storage 20.0 µs and Collision Cell Energy 20.0 eV. Deconvolution of MS data was performed using Unidec with the following mass ranges: 5000-30000 Da for the competition experiments, 5000-30000 Da for the $(T)_{30}$ and $(CT)_{15}$ sequences and 6000-18000 Da for *rlmN* systems containing both anthocyanins.[23]

6.2.3 UV-Vis spectroscopy

UV-Vis spectroscopy experiments were conducted on a Shimadzu UV-3600i-Plus spectrophotometer equipped with a temperature-controlled 8 Series Micro Multi-Cell Holder (MMC-1600C). Thermal melting experiments were conducted by ramping the temperature from 20 °C to the final temperature at a rate of 1 °C/min. The absorbance at 260 nm was recorded at every 0.5 °C interval. To limit evaporation and gas bubble formation during temperature ramping all samples were degassed by sonicating for 5 minutes prior to analysis. The cells were sealed during each run with provided silicone stoppers which were secured in place by the clamp as part of the cell holder assembly. As such, all UV-Vis melting experiments were conducted under constant volume conditions. All experiments were completed in triplicate and the data shown is an average of the three runs, with derivative plots being smoothed with a Savitzky-Golay filter to improve visualisation.

6.3 Results and discussion

6.3.1 Binding of anthocyanin standards to model and biological DNA triplex sequences

Using MS, the binding of cyanidin 3-glucoside and malvidin 3,5-diglucoside (anthocyanin structures shown in Figure 2) to the $(T)_{30}$ and $(CT)_{15}$ model triplex sequences (from Chapter 3), as well as the methicillin-resistant *Staphylococcus aureus* (*MRSA*) *rlmN* gene triplex sequence (from Chapter 5) was investigated. Triplex sequences are again provided for ease of reference in Table 6.1 below.

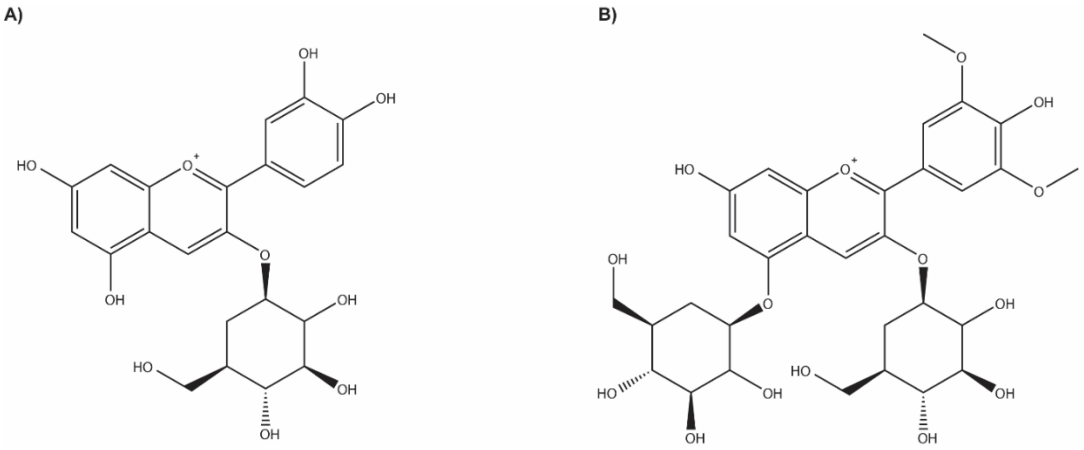


Figure 6.2: Structures of the anthocyanins **(A)** cyanidin 3-glucoside and **(B)** malvidin 3,5-diglucoside, investigated for DNA triplex binding to the $(T)_{30}$, $(CT)_{15}$ and *rlmN* triplex systems.

Table 6.1: Name and sequence composition of the strands involved in the $(T)_{30}$, $(CT)_{15}$ and *rlmN* DNA triplex systems.

Sequence names	Molecular weight (Da)	Sequence composition (5'—3')
$(T)_{30}$ Duplex purine strand	9334	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
$(T)_{30}$ Duplex pyrimidine strand	9064	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
$(T)_{30}$ TFO	9064	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTT
$(CT)_{15}$ Duplex purine strand	9574	AGAGAGAGAGAGAGAGAGAGAGAGAGAG
$(CT)_{15}$ Duplex pyrimidine strand	8839	CTCTCTCTCTCTCTCTCTCTCTCTCTCT
$(CT)_{15}$ TFO	8839	TCTCTCTCTCTCTCTCTCTCTCTCTCTC
rlmN Duplex purine strand	5326	GAAAAGAAAAAGAAGAA
rlmN Duplex pyrimidine strand	5049	TTCTTCTTTTTCTTTTC
rlmN TFO	5049	CTTTTCTTTTTCTTCTT

MS data for the $(T)_{30}$ triplex system with both cyanidin 3-glucoside and malvidin 3,5-diglucoside can be seen in Figure 6.3A and 6.3B respectively.

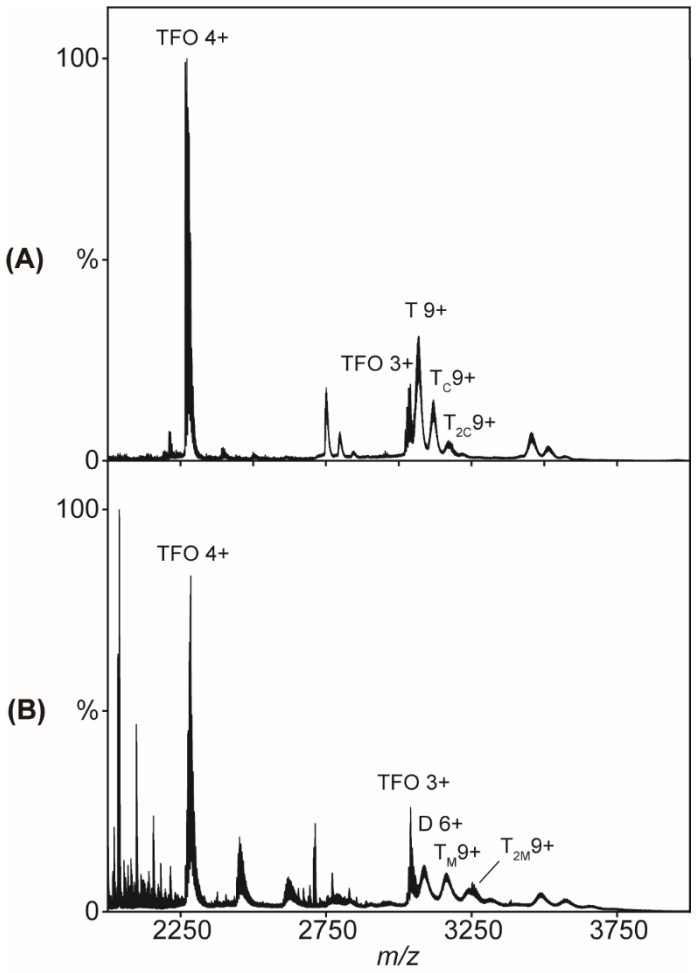


Figure 6.3: Mass spectra of the $(T)_{30}$ DNA triplex system with **(A)** cyanidin 3-glucoside and **(B)** malvidin 3,5-diglucoside, in 250 mM ammonium acetate solution at pH 7.4. Mass spectra were collected in positive ion mode. Samples were prepared in a 1:1:1:2 TFO: purine:pyrimidine:anthocyanin ratio, with each individual strand and anthocyanin at a concentration of 2 µM and 4 µM respectively. Annotations are as follows: TFO: TFO, D: duplex, T: triplex, T_C: triplex with cyanidin 3-glucoside bound, T_{2C}: triplex with two cyanidin 3-glucoside molecules bound, T_M: triplex with malvidin 3,5-diglucoside bound, T_{2M}: triplex with two malvidin 3,5-diglucoside molecules bound. Charge states are indicated on the spectra.

Figure 6.3A shows that, in a 2-fold excess, cyanidin 3-glucoside can bind to the $(T)_{30}$ triplex. Although the unbound $(T)_{30}$ triplex ions are in a higher abundance compared to the cyanidin 3-glucoside bound species, there is evidence for up to two cyanidin 3-glucoside molecules bound. Figure 6.3B reveals similar observations for malvidin 3,5-diglucoside, where again there is the detection of species with one and two anthocyanin molecules bound to the $(T)_{30}$

triplex. This shows the value of using MS to detect these interactions, as previous studies that utilised UV-Vis spectroscopy were unable to prove interaction between different triplexes and diglucoside anthocyanins.[22]

After demonstrating that both anthocyanin standards can bind to the $(T)_{30}$ triplex, it was of interest to determine whether the anthocyanins have any sequence binding dependence by comparing this to the $(CT)_{15}$ model system. MS data for the $(CT)_{15}$ triplex system with both cyanidin 3-glucoside and malvidin 3,5-diglucoside can be seen in Figures 6.4A and 6.4B below.

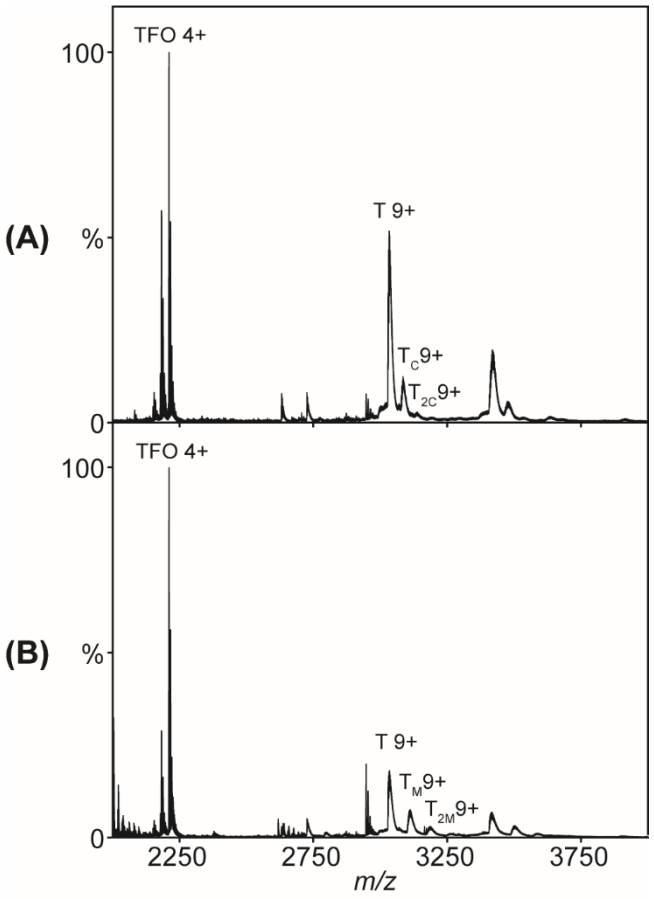


Figure 6.4: Mass spectra of the $(CT)_{15}$ DNA triplex system with **(A)** cyanidin 3-glucoside and **(B)** malvidin 3,5-diglucoside, in 250 mM ammonium acetate solution at pH 7.4. Mass spectra were collected in positive ion mode. Samples were prepared in a 1:1:1:2 TFO:purine:pyrimidine:anthocyanin ratio, with each individual strand and anthocyanin at a concentration of 2 µM and 4 µM respectively. Annotations are as follows: TFO: TFO, D: duplex, T: triplex, T_C: triplex with cyanidin 3-glucoside bound, T_{2C}: triplex with two cyanidin 3-glucoside molecules bound, T_M: triplex with malvidin 3,5-diglucoside bound, T_{2M}: triplex with two malvidin 3,5-diglucoside molecules bound. Charge states are indicated on the spectra.

Figure 6.4A shows that cyanidin 3-glucoside can bind to the $(CT)_{15}$ sequence, although it appears to bind with a lower affinity as there is more unbound triplex eident when comparing relative ion abundances. It was also observed that two cyanidin 3-glucoside molecules were able to bind to the $(CT)_{15}$ triplex. Figure 6.4B similarly shows the binding of one and two malvidin 3,5-diglucoside molecules to the $(CT)_{15}$ triplex, although the unbound triplex was in significantly lower relative abundance compared to the cyanidin 3-glucoside system in Figure 6.4A. This suggests that malvidin 3,5-diglucoside has a higher binding affinity to the $(CT)_{15}$ triplex than cyanidin 3-glucoside, although more expansive binding affinity experiments are required to confirm this.

Given the ability of anthocyanins to bind to model triplex structures, it was also of interest to investigate whether the same observations are made in shorter and more biologically relevant sequences like the *rlmN* gene sequence from MRSA. MS data for the *rlmN* triplex system with both cyanidin 3-glucoside and malvidin 3,5-diglucoside can be seen in Figure 6.5 below.

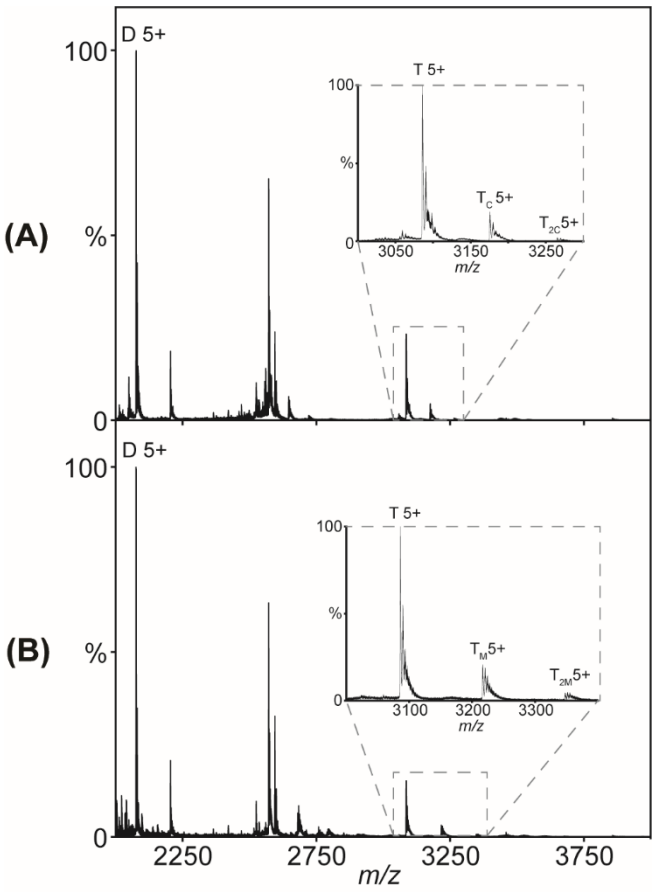


Figure 6.5: Mass spectra of the *rlmN* DNA triplex system with **(A)** cyanidin 3-glucoside and **(B)** malvidin 3,5-diglucoside, in 250 mM ammonium acetate solution at pH 7.4. Mass spectra were collected in positive ion mode. Samples were prepared in a 1:1:1:2 TFO:purine:pyrimidine:anthocyanin ratio, with each individual strand and anthocyanin at a concentration of 2 µM and 4 µM respectively. Annotations are as follows: TFO: TFO, D: duplex, T: triplex, T_C: triplex with cyanidin 3-glucoside bound, T_{2C}: triplex with two cyanidin 3-glucoside molecules bound, T_M: triplex with malvidin 3,5-diglucoside bound, T_{2M}: triplex with two malvidin 3,5-diglucoside molecules bound. Charge states are indicated on the spectra.

Figure 6.5 shows that, while there is binding of one and two anthocyanin molecules to the triplex species, these complexes are formed with substantially lower relative abundance compared to the model sequences. There also appears to be no significant difference in the binding affinity between the two anthocyanins for this particular triplex sequence. Importantly however, Figures 6.3-6.5 also show that the interactions between the anthocyanins and the DNA are specific to only the triplex structures, as there is no evidence of any TFO- or duplex-bound anthocyanin species in any system.

6.3.2 Comparing binding affinities of cyanidin 3-glucoside and malvidin 3,5-diglucoside to the model and *MRSA* triplex systems

When comparing the two model systems (Figures 6.3 and 6.4) to the *rlmN* sequence (Figure 6.5) it is evident that there is more triplex detected in both model systems which could be due to the different ionisation efficiency of the triplex species. This makes it challenging to draw conclusions regarding the binding affinity of the two anthocyanins to the triplex systems by direct comparison of the MS data. To make more definitive conclusions about the relative binding affinity, competition binding experiments can be performed to assess if there is any preferential binding to a specific triplex sequence.

Due to the complexity of the MS data in the competitive binding experiments (Appendix 1 Figure A1), the deconvoluted data is instead shown to identify the masses of the species in the system and enable comparison of their relative abundances. Deconvoluted MS data for the combined $(T)_{30}$, $(CT)_{15}$ and *rlmN* triplex system with both cyanidin 3-glucoside and malvidin 3,5-diglucoside can be seen in Figure 6.6.

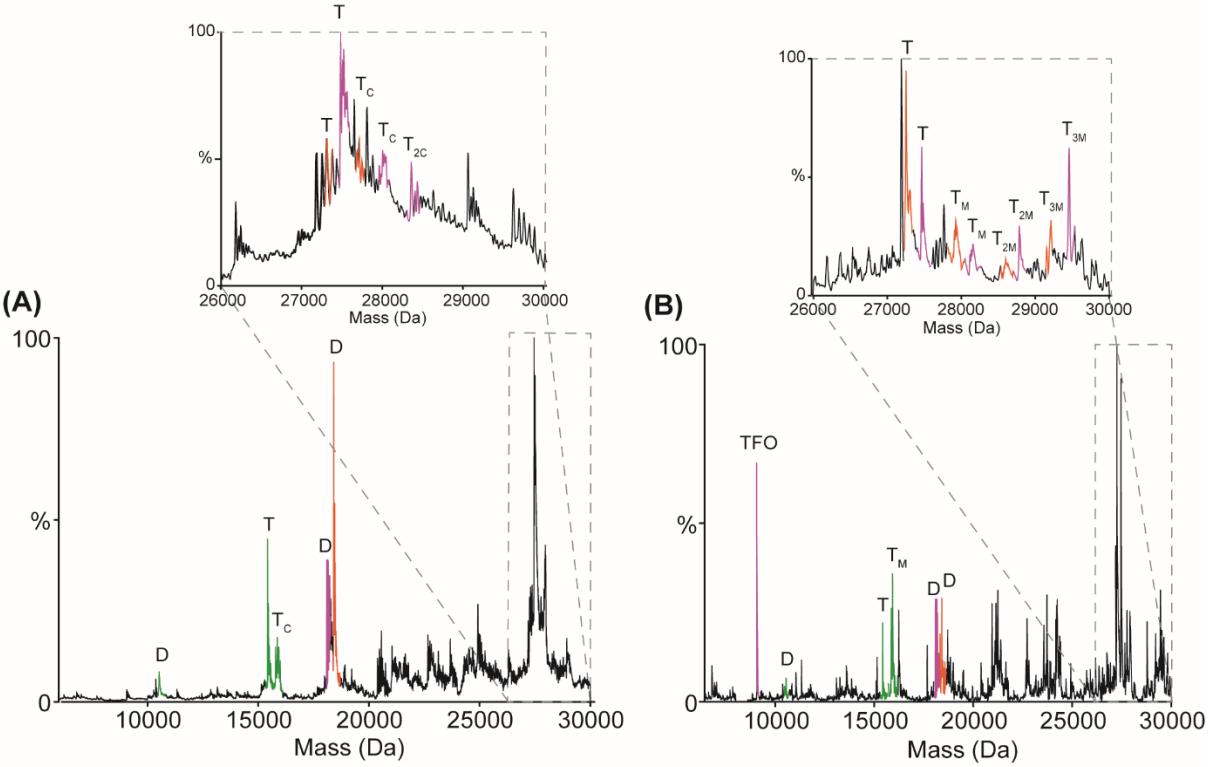


Figure 6.6: Deconvoluted mass spectra of the combined $(T)_{30}$, $(CT)_{15}$ and *rlmN* DNA triplex systems with **(A)** cyanidin 3-glucoside and **(B)** malvidin 3,5-diglucoside, in 250 mM ammonium acetate solution at pH 7.4. Mass spectra were collected in positive ion mode. Samples were prepared in a 1:1:1:2 TFO: purine:pyrimidine:anthocyanin ratio, with each individual strand and anthocyanin at a concentration of 2 µM and 4 µM respectively. Annotations are as follows: TFO: TFO, D: duplex, T: triplex, T_C: triplex with cyanidin 3-glucoside bound, T_{2C}: triplex with two cyanidin 3-glucoside molecules bound, T_M: triplex with malvidin 3,5-diglucoside bound, T_{2M}: triplex with two malvidin 3,5-diglucoside molecules bound and T_{3M}: triplex with three malvidin 3,5-diglucoside molecules bound. The *rlmN* system is shown in green, with the insets in **(A)** and **(B)** indicating $(T)_{30}$ and $(CT)_{15}$ species coloured purple and orange respectively.

Figure 6.6A indicates that cyanidin 3-glucoside has successfully bound to all three triplexes when they are present in the same sample. The cyanidin-bound triplexes appear relatively similar in abundance with the $(CT)_{15}$ being slightly higher than $(T)_{30}$ and then followed by the *rlmN* species. However, there is evidence for the binding of two cyanidin 3-glucoside molecules to the $(T)_{30}$ triplex. Therefore, it appears that there is more specific binding of the cyanidin 3-glucoside to the $(CT)_{15}$ triplex sequence, however, the $(T)_{30}$ triplex sequence allows for more anthocyanin molecules to bind.

Figure 6.6B shows that malvidin 3,5-diglucoside has also successfully bound to all three triplexes. The malvidin 3,5-diglucoside bound *rlmN* triplex species has the highest abundance compared to the other bound triplex species and it is also the only species to have a higher

relative abundance than its equivalent unbound triplex. Interestingly with the $(T)_{30}$ and $(CT)_{15}$ triplexes, there are species detected where up to three malvidin 3,5-diglucoside molecules have bound, with the $(T)_{30}$ triplex with three molecules bound being in the highest abundance of these complexes. It can therefore be concluded that malvidin 3,5-diglucoside has the highest binding specificity to the *rlmN* triplex, although the $(T)_{30}$ triplex can have the most anthocyanins bound to the triplex for both cyanidin 3-glucoside and malvidin 3,5-diglucoside (insets of Figure 6.6A and 6.6B).

Here, the binding of the individual anthocyanins to the three different triplex systems was explored. Next, a comparison was made of the binding affinity between both anthocyanins for each triplex system separately. Once again, due to the complexity of the MS data (Appendix 1 Figure A2), the deconvoluted data has been shown to indicate the masses of the species in the system and to compare their relative abundance. Deconvoluted MS data for the individual $(T)_{30}$, $(CT)_{15}$ and *rlmN* triplex systems with both cyanidin 3-glucoside and malvidin 3,5-diglucoside can be seen in Figure 6.7A, 6.7B and 6.7C respectively.

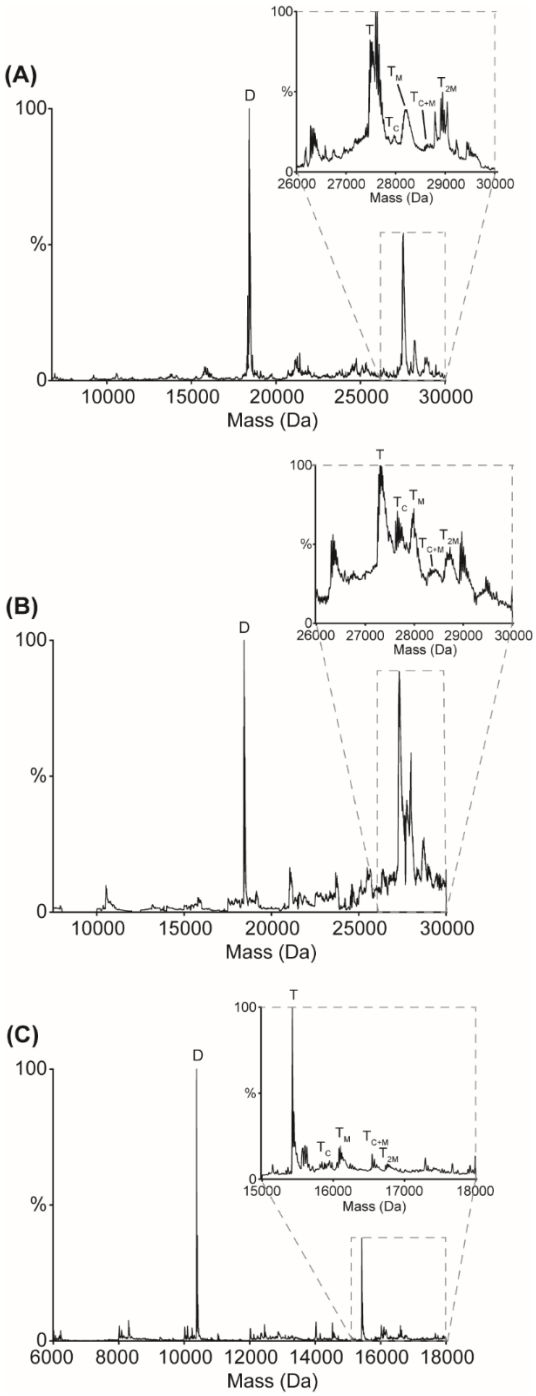


Figure 6.7: Deconvoluted mass spectra of the **(A)** $(T)_{30}$, **(B)** $(CT)_{15}$ and **(C)** *rlmN* DNA triplex systems with both cyanidin 3-glucoside and malvidin 3,5-diglucoside, in 250 mM ammonium acetate solution at pH 7.4. Mass spectra were collected in positive ion mode. Samples were prepared in a 1:1:1:2 TFO:purine: pyrimidine:anthocyanin ratio, with each individual strand and anthocyanin at a concentration of 2 µM and 4 µM respectively. Annotations are as follows: D: duplex, T: triplex, T_C: triplex with cyanidin 3-glucoside bound, T_M: triplex with malvidin 3,5-diglucoside bound, T_{2M}: triplex with two malvidin 3,5-diglucoside molecules bound and T_{C+M}: triplex with one cyanidin 3-glucoside and one malvidin 3,5-diglucoside molecules bound.

Figure 6.7A shows the formation of species with one and two malvidin 3,5-diglucoside molecules binding to the $(T)_{30}$ triplex. There is also detection of cyanidin 3-glucoside bound species, as well as $(T)_{30}$ triplex with one cyanidin 3-glucoside and one malvidin 3,5-diglucoside molecule bound. However, no species with multiple cyanidin-3-glucoside molecules were detected. This indicates that the malvidin 3,5-diglucoside can preferentially bind to the $(T)_{30}$ triplex.

Similar results were observed in Figure 6.7B, where the malvidin 3,5-diglucoside bound to the $(CT)_{15}$ triplex is the most abundant anthocyanin-bound species. However, in this case, the cyanidin 3-glucoside bound ion has the second highest abundance and $(CT)_{15}$ triplex with one cyanidin 3-glucoside and one malvidin 3,5-diglucoside molecule bound, as well as two malvidin 3,5-diglucoside molecules bound, were also detected in Figure 6.7B. Once again there was no evidence for $(CT)_{15}$ triplex bound to two cyanidin 3-glucoside molecules, indicating that this triplex exhibits preferential binding to the malvidin 3,5-diglucoside.

Malvidin 3,5-diglucoside bound *rlmN* triplex is again the most abundant anthocyanin bound species for the *rlmN* triplex (Figure 6.7C), indicating overall that the malvidin 3,5-diglucoside is the preferred anthocyanin to bind to DNA triplexes. This is interesting since it was concluded in previous solution-phase experiments that diglucosides were unable to bind, given no change in melting temperature could be discerned.[22] Although malvidin 3,5-diglucoside contains additional hydroxyl and glucose substituents, this could in fact allow for a more preferred fit as the molecule intercalates within the triplex structure. Therefore, the next step is to investigate whether the binding of these anthocyanins has any effect on the stability of the triplex structures.

6.3.3 Stability studies of cyanidin 3-glucoside and malvidin 3,5-diglucoside bound DNA triplexes

In the previous sections, it has been shown that both cyanidin 3-glucoside and malvidin 3,5-diglucoside can bind to the $(T)_{30}$ and $(CT)_{15}$ and *MRSA rlmN* triplex sequences. Literature has shown that glucoside flavonoids and anthocyanins have a stabilising effect on DNA triplexes, whereas flavonoids and anthocyanins with a diglucoside provided no change in stability.[17,22] To assess the stability of the anthocyanin-bound triplexes formed in the previous experiments,

UV-vis spectroscopy has been employed to determine their melting points. Derivative melting profiles of the $(T)_{30}$ $(CT)_{15}$, and the *rlmN* triplex with no anthocyanins (control), cyanidin 3-glucoside and malvidin 3,5-diglucoside bound can be seen in Figure 6.8 below.

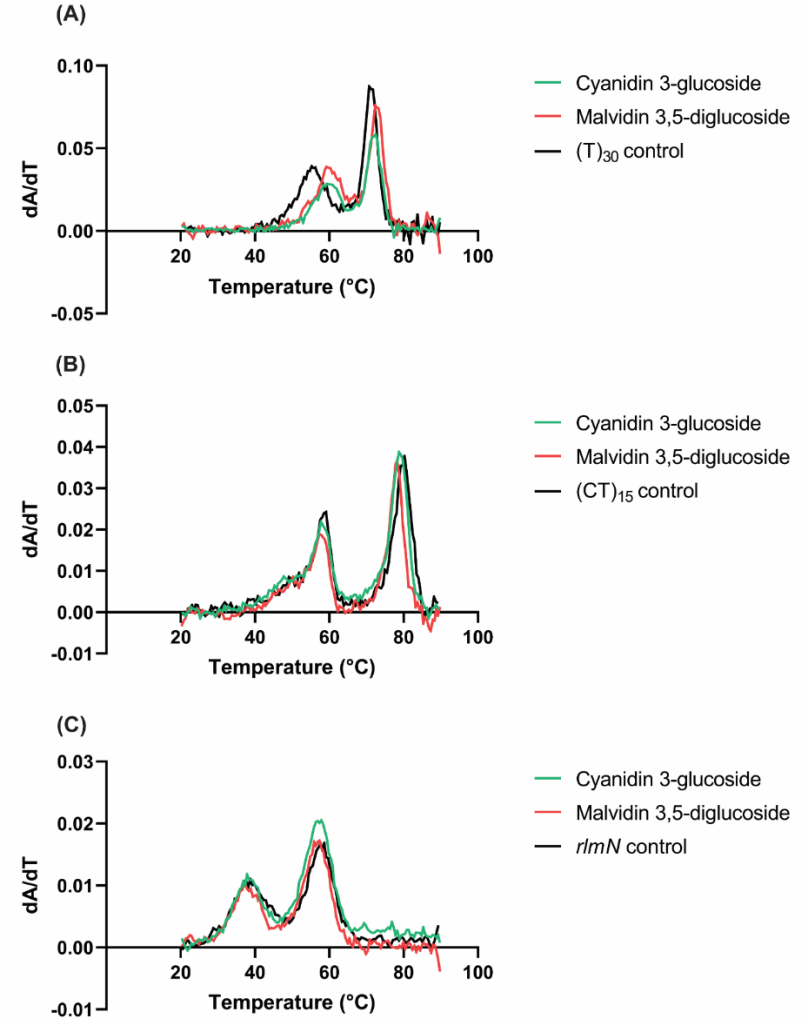


Figure 6.8: Derivative UV-vis spectroscopy melting profiles of **(A)** $(T)_{30}$, **(B)** $(CT)_{15}$ and **(C)** rlmN DNA triplex systems with no anthocyanins (black), cyanidin 3-glucoside (green) and malvidin 3,5-diglucoside (red) bound, in ammonium acetate at pH 7.4. All samples were zeroed at 260 nm at 20 °C, and melting was conducted from 20 to 90 °C. Samples were prepared in a 1:1:1:2 TFO:purine:pyrimidine: anthocyanin ratio, with each individual strand and anthocyanin at a concentration of 2 µM and 4 µM respectively.

Figure 6.8A shows that the addition of both cyanidin 3-glucoside and malvidin 3,5-diglucoside gave rise to stabilisation of the $(T)_{30}$ triplex structure, increasing the melting temperature from 56 °C to 59 °C. Notably, there is also a very minimal increase in the duplex melting temperature

for the anthocyanin treated systems compared to the control. Figures 6.8B and 6.8C showed that there was no difference between the control derivative melting profiles for the $(CT)_{15}$ and *rlmN* triplex systems and the melting profiles with cyanidin 3-glucoside and malvidin 3,5-diglucoside bound. These results indicate that stabilisation of triplexes using anthocyanins is sequence-dependent and shows that the binding interactions of these small molecules don't necessarily correlate with stabilisation.

6.3.4 Binding of netropsin to anthocyanin triplex systems

Netropsin is a well-known γ-peptide (structure shown in Figure 6.9A) that can bind to the minor groove of A-T rich DNA duplex sequences via hydrogen bonding, van der Waals and electrostatic interactions (Figure 6.9B).[24,25] The interactions between netropsin and DNA were one of the first identified to be selective to the minor groove and have had a notable impact on the development of potential anticancer and antibiotic therapeutics.[26,27] MS has previously been used to determine the mechanism of binding between netropsin and hairpin DNA.[28] Other solution phase techniques such as circular dichroism and UV-vis spectroscopy have shown that netropsin has a destabilising effect on triplex helical DNA assemblies.[29,30]

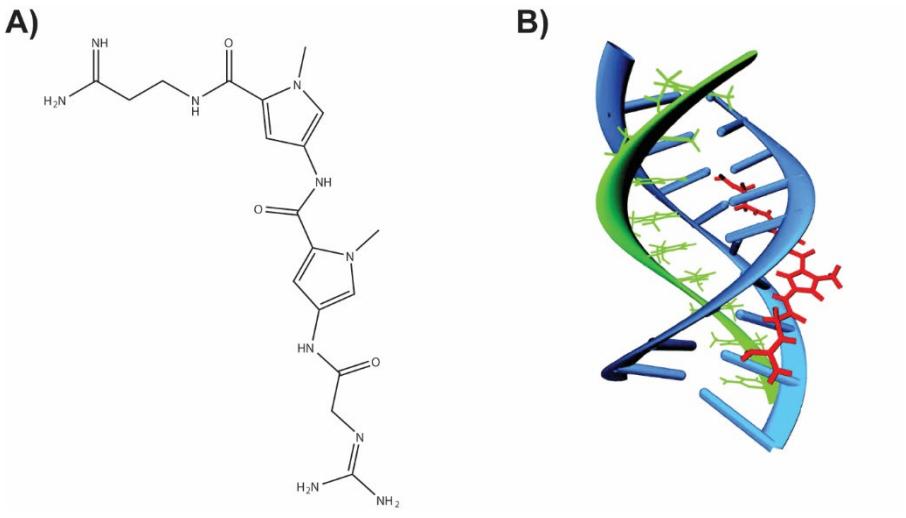


Figure 6.9: (A) Structure of netropsin and **(B)** the binding of netropsin (red) to the minor groove of duplex DNA (blue) via hydrogen bonding in the DNA triplex structure with the TFO highlighted in green. Adapted with permission from Boehm *et.al.*[31]

The effects of netropsin on the binding of anthocyanins to the triplex species was investigated, with the intent to potentially understand anthocyanin binding sites on the triplex structure based on what is known about netropsin. The interactions between netropsin and the $(T)_{30}$, $(CT)_{15}$ and *rlmN* triplex systems with both cyanidin 3-glucoside and malvidin 3,5-diglucoside can be analysed by native MS. MS data for the cyanidin 3-glucoside and malvidin 3,5-diglucoside $(T)_{30}$ triplex system in the presence of netropsin is shown in Figure 6.10A and 6.10B respectively below.

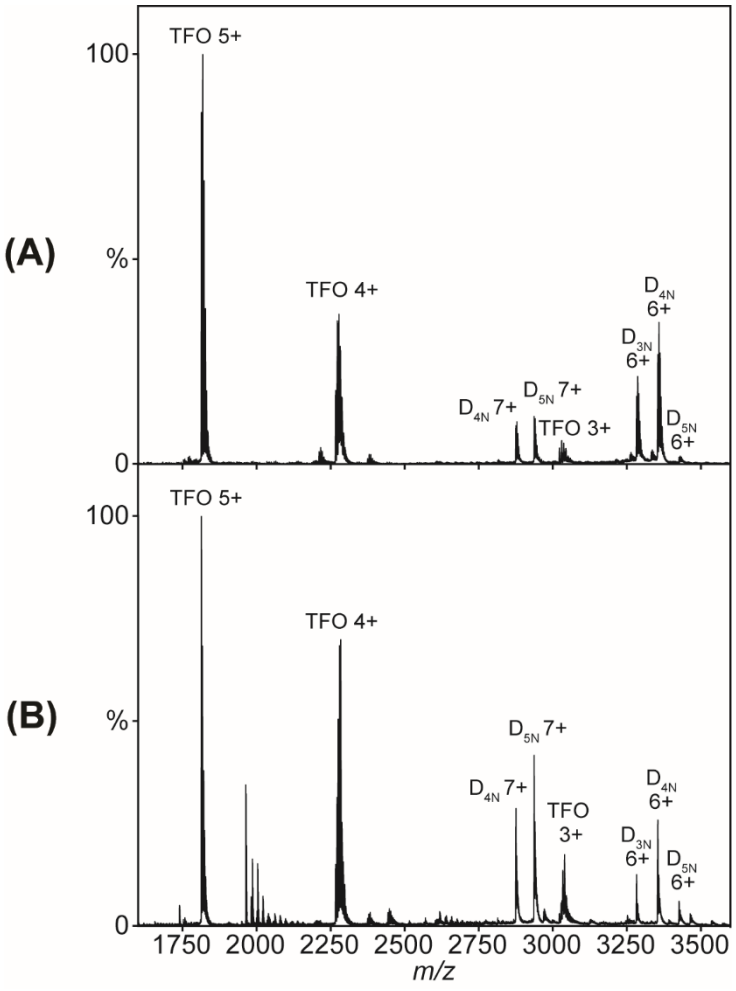


Figure 6.10: Mass spectra of the netropsin treated $(T)_{30}$ DNA triplex system with **(A)** cyanidin 3-glucoside and **(B)** malvidin 3,5-diglucoside, in 250 mM ammonium acetate solution at pH 7.4. Mass spectra were collected in positive ion mode. Samples were prepared in a 1:1:1:2 Netropsin: TFO: purine:pyrimidine:anthocyanin ratio, with each individual strand, netropsin and anthocyanin at a concentration of 2 µM, 2 µM and 4 µM respectively. Annotations are as follows: TFO: TFO, D: duplex, D_{4N}: duplex with four netropsin molecules bound and D_{5N}: duplex with five netropsin molecules bound. Charge states are indicated on the spectra.

Figures 6.10A and 6.10B show there are species in both spectra that are attributed to four and five netropsin molecules binding to the $(T)_{30}$ duplex structure. Since netropsin binding is known to favour A-T duplex sequences, this result aligns with the literature that multiple netropsin molecules can bind to duplex DNA.[28] Netropsin is also known to change the width of the minor groove of the DNA structure.[25,32] So the fact that up to five netropsin molecules are binding to the duplex DNA means that there could be significant changes to the structure, causing the TFO to be ejected from the duplex completely, as there is no indication of the triplex present. Investigation into the sequence specificity of netropsin binding was next examined for the anthocyanin bound $(CT)_{15}$ model sequence.

MS data for the cyanidin 3-glucoside and malvidin 3,5-diglucoside $(CT)_{15}$ triplex system treated with netropsin is shown in Figure 6.11 below.

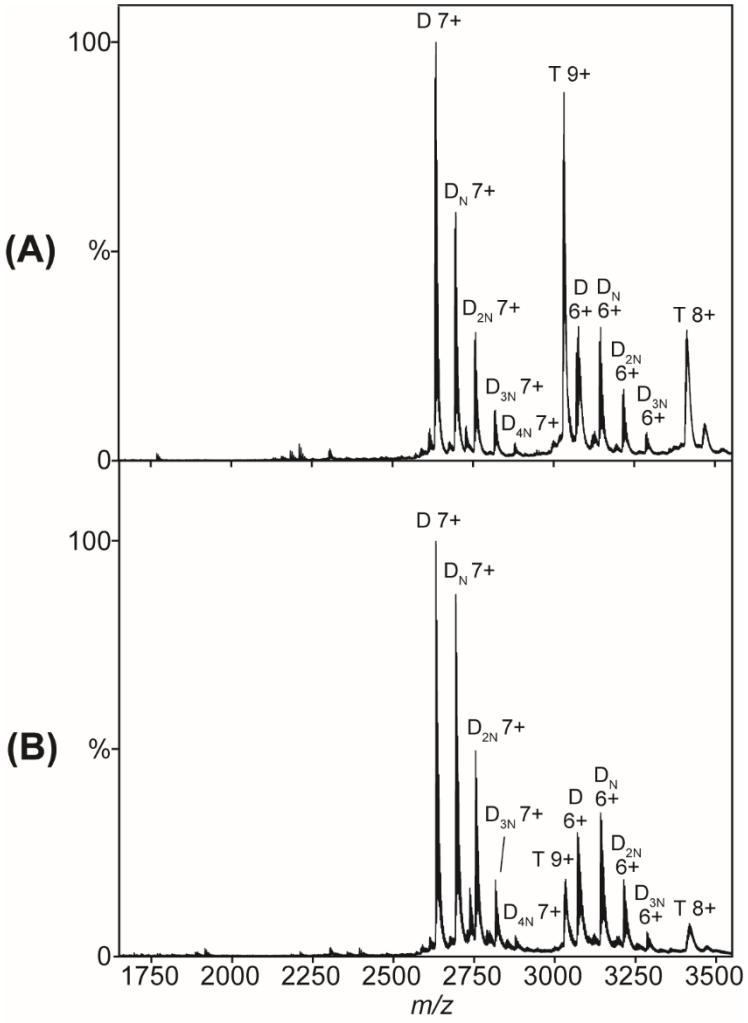


Figure 6.11: Mass spectra of the netropsin treated $(CT)_{15}$ DNA triplex system with **(A)** cyanidin 3-glucoside and **(B)** malvidin 3,5-diglucoside, in 250 mM ammonium acetate solution at pH 7.4. Mass spectra were collected in positive ion mode. Samples were prepared in a 1:1:1:1:2 Netropsin:TFO: purine:pyrimidine:anthocyanin ratio, with each individual strand, netropsin and anthocyanin at a concentration of 2 µM, 2 µM and 4 µM respectively. Annotations are as follows: D: duplex, T: triplex, D_N: duplex with netropsin bound, D_{2N}: duplex with two netropsin molecules bound, D_{3N}: duplex with three netropsin molecules bound and D_{4N}: duplex with four netropsin molecules bound. Charge states are indicated on the spectra.

Figure 6.11A shows that unbound duplex is the most abundant species in the spectrum, followed by the cyanidin 3-glucoside $(CT)_{15}$ triplex. This is clear evidence for the preference of netropsin for the A-T duplex, as seen in Figure 6.10A. The data also shows that there are species of duplex with one to three molecules of netropsin bound. Since there are fewer netropsin molecules bound compared to the $(T)_{30}$ system, this suggests there would be less distortion of the duplex structure, consistent with a higher relative abundance of unbound triplex. Unfortunately, even with this significantly higher abundance of triplex, there appears

to be no cyanidin 3-glucoside or netropsin-bound triplex. Similar observations are also observed upon the addition of malvidin 3,5-diglucoside (Figure 6.11B). Since there was no observation of any triplex-bound species, these experiments were unfortunately unable to provide further information on the binding modes of anthocyanins to the triplexes.

6.4 Conclusion

Cyanidin 3-glucoside and malvidin 3,5-diglucoside were investigated for their DNA triplex binding abilities. The binding of up to three individual anthocyanin molecules to each of the sequences was indicated by MS, which is believed to be the first time that the binding between anthocyanins and DNA triplexes specifically has been characterised by gas phase methods. Due to the ability of multiple molecules to bind to a triplex, competitive binding experiments were undertaken to determine if the anthocyanins would bind preferentially to a particular sequence. The cyanidin 3-glucoside bound $(CT)_{15}$ triplex showed the highest abundance in MS, however the $(T)_{30}$ triplex allowed for more molecules to bind. This was similar to the malvidin 3,5-diglucoside system, as it specifically bound to the shorter *rlmN* sequences in a higher abundance however the most anthocyanin molecules (three) were able to bind to the $(T)_{30}$ triplex, likely due to there being more space to intercalate in the longer sequence.

This ability for the $(T)_{30}$ triplex to bind more anthocyanin molecules in the binding affinity experiments appears to have a slight stabilising effect on the triplex structure when bound to both cyanidin 3-glucoside and malvidin 3,5-diglucoside, as shown by UV-Vis spectroscopy experiments. Although the other sequences have also shown the ability to bind multiple anthocyanin molecules, the binding was not effective in inducing a stabilising effect. Computational methods and thermodynamic experiments such as ITC would be complementary here to determine the structural effect and binding strength of the bound anthocyanins on the triplex.

Known triplex destabiliser netropsin was also incorporated into these anthocyanin-triplex systems in an attempt to inform on the possible binding positioning in the triplex of both anthocyanin standards. Unfortunately, there was no successful detection of anthocyanin-bound triplexes in the presence of netropsin. Unbound triplex was only detected in the $(CT)_{15}$ system which indicates that there is a possibility that the anthocyanins intercalate near the minor groove which was replaced by the more favoured netropsin. Once again, the use of molecular modelling and thermodynamic methods would allow for more definitive conclusions.

Our investigation highlights the ability to use MS and UV-Vis spectroscopy to identify the binding of anthocyanin molecules to DNA triplexes. The use of MS was able to provide more

information than previous spectroscopic studies by showing interactions between DNA triplexes and a diglucoside anthocyanin. It has also shown that the binding of both anthocyanins to the $(T)_{30}$ sequence has a stabilising effect specific to the triplex. This shows the potential for anthocyanins to be used in therapeutics that can target triplex DNA, as shown by structurally similar flavonoids.[33,34]

6.5 References

(1) White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. Recognition of the Four Watson-Crick Base Pairs in the DNA Minor Groove by Synthetic Ligands. *Nature* **1998**, *391* (6666), 468–471.

(2) Schneider, K.; Keller, S.; Wolter, F. E.; Roeglin, L.; Beil, W.; Seitz, O.; Nicholson, G.; Bruntner, C.; Riedlinger, J.; Fiedler, H.-P.; Suessmuth, R. D. Proximicins A,B and C-Antitumor 3 Furan Analogues of Netropsin from the Marine Actinomycete Verrucosispora Induce Upregulation of P53 and the Cyclin Kinase Inhibitor P21. *Angew. Chemie* **2008**, *47* (17), 3258–3261.

(3) Banks, W. A.; Greig, N. H. Small Molecules as Central Nervous System Therapeutics: Old Challenges, New Directions, and a Philosophic Divide. *Future Med. Chem.* **2019**, *11* (6), 489–493.

(4) Das, A.; Majumder, D.; Saha, C. Correlation of Binding Efficacies of DNA to Flavonoids and Their Induced Cellular Damage. *J. Photochem. Photobiol. B Biol.* **2017**, *170*, 256–262.

(5) Almaqwashi, A. A.; Pramanathan, T.; Rouzina, I.; Williams, M. C. Mechanisms of Small Molecule-DNA Interactions Probed by Single-Molecule Force Spectroscopy. *Nucleic Acids Res.* **2016**, *44* (9), 3971–3988.

(6) Tse, W. C.; Boger, D. L. Sequence-Selective DNA Recognition: Natural Products and Nature's Lessons. *Chem. Biol.* **2004**, *11* (12), 1607–1617.

(7) Riechert-Krause, F.; Autenrieth, K.; Eick, A.; Weisz, K. Spectroscopic and Calorimetric Studies on the Binding of an Indoloquinoline Drug to Parallel and Antiparallel DNA Triplexes. *Biochem* **2013**, *52* (1), 41–52.

(8) Basili, S.; Bergen, A.; Dall'Acqua, F.; Faccio, A.; Granzhan, A.; Ihmels, H.; Moro, S.; Viola, G. Relationship between the Structure and the DNA Binding Properties of Diazoniapolycyclic Duplex- and Triplex-DNA Binders: Efficiency, Selectivity and Binding Mode. *Biochem* **2007**, *46* (44), 12721–12736.

(9) Jain, A. K.; Gupta, S. K.; Tawar, U.; Dogra, S. K.; Tandon, V. Benzimidazoles: A Minor Groove-Binding Ligand-Induced Stabilization of Triple Helix. *Oligonucleotides* **2009**, *19* (1), 53–62.

(10) Yang, H.; Yao, W.; Wang, Y.; Shi, L.; Su, R.; Wan, D.; Xu, N.; Lian, W.; Chen, C.; Liu, S. High-Throughput Screening of Triplex DNA Binders from Complicated Samples by 96-Well Pate Format in Conjunction with Peak Area-Fading UHPLC-Orbitrap MS †. *Analyst* **2017**, *142*, 670.

(11) Xu, N.; Yang, H.; Cui, M.; Wan, C.; Liu, S. High-Performance Liquid Chromatography-Electrospray Ionization-Mass Spectrometry Ligand Fishing Assay: A Method for Screening Triplex DNA Binders from Natural Plant Extracts. *Anal. Chem.* **2012**, *84* (5), 2562–2568.

(12) Yang, H.; Wang, Y.; Yu, W.; Shi, L.; Wang, H.; Su, R.; Chen, C.; Liu, S. Screening and Investigation of Triplex DNA Binders from Stephania Tetrandra S. Moore by a Combination of Peak Area-Fading Ultra High-Performance Liquid Chromatography with Orbitrap Mass Spectrometry and Optical Spectroscopies. *J. Sep. Sci.* **2018**, *41* (14), 2878–2885.

(13) Bandele, O. J.; Osheroff, N. Bioflavonoids as Posions of Human Topoisomerase IIα and IIβ. *Biochem* **2007**, *46* (20), 6097–6108.

(14) Tawani, A.; Kumar, A. Structual Insight into the Interaction of Flavonoids with Human Telomeric Sequence. *Sci. Rep.* **2015**, *5*, 17574.

(15) Bag, S.; Burman, M. D.; Bhowmik, S. Structual Insights and Shedding Light on Preferential Interactions of Dietary Flavonoids with G-Quadruplex DNA Structures: A New Horizon. *Heliyon* **2023**, *9* (3), e13959.

(16) Wan, C.; Cui, M.; Song, F.; Liu, Z.; Liu, S. A Study of the Non-Covalent Interaction between Flavonoids and DNA Triplexes by Electrospray Ionization Mass Spectrometry. *Int. J. Mass Spectrom.* **2009**, *283* (1–3), 48–55.

(17) Rangel, V. M.; Gu, L.; Chen, G.; Chen, Q. H.; Xue, L. 5-Substituted 3, 3′, 4′, 7-Tetramethoxyflavonoids – A Novel Class of Potent DNA Triplex Specific Binding Ligands. *Bioorg. Med. Chem. Lett.* **2022**, *61*, 128608.

(18) Satué-Gracia, M. T.; Heinonen, M.; Frankel, E. N. Anthocyanins as Antioxidants on Human Low-Density Lipoprotein and Lecithin−Liposome Systems. *J. Agric. Food Chem.* **1997**, *45* (9), 3362–3367.

(19) Liu, J.; Zhou, H.; Song, L.; Yang, Z.; Qiu, M.; Wang, J.; Shi, S. Anthocyanins: Promising

Natural Products with Diverse Pharmacological Activities. *Molecules*. 2021.

(20) Pachulicz, R. J.; Yu, L.; Jovcevski, B.; Bulone, V.; Pukala, T. L. Polyphenol Characterisation and Diverse Bioactivities of Native Australian Lilly Pilly (Syzygium Paniculatum) Extract. *Food Funct.* **2022**, *13*, 8585–8592.

(21) Ragazzon, P. A.; Bradshaw, T.; Matthews, C.; Iley, J.; Missailidis, S. The Characterisation of Flavone-DNA Isoform Interactions as a Basis for Anticancer Drug Development. *Anticancer Res.* **2009**, *29* (2273–2284).

(22) Mas, T.; Susperregui, J.; Berké, B.; Chèze, C.; Moreau, S.; Nuhrich, A.; Vercauteren, J. DNA Triplex Stabilization Property of Natural Anthocyanins. *Phytochemistry* **2000**, *53* (6), 679–687.

(23) Marty, M. T.; Baldwin, A. J.; Marklund, E. G.; Hochberg, G. K. A.; Benesch, J. L. P.; Robinson, C. V. Bayesian Deconvolution of Mass and Ion Mobility Spectra: From Binary Interactions to Polydisperse Ensembles. *Anal. Chem.* **2015**, *87* (8), 4370–4376. https://doi.org/10.1021/acs.analchem.5b00140.

(24) Finlay, A. C.; Hochstein, F. A.; Sobin, B. A.; Murphy, F. X. Netropsin, a New Antibiotic Produced by a Streptomyces. *J. Am. Chem. Soc.* **1951**, *73* (1), 341–343.

(25) Van Hecke, K.; Nam, P. C.; Nguyen, M. T.; Van Meervelt, L. Netropsin Interactions in the Minor Groove of d(GGCCAATTGG) Studied by a Combination of Resolution Enhancement and Ab Initio Calculations. *FEBS J.* **2005**, *272* (14), 3531–3541.

(26) Marky, L.; Breslauer, K. J. Origins of Netropsin Binding Affinity and Specificity: Correlations of Thermodynamic and Structural Data. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84* (13), 4359–4363.

(27) Goodwin, K. D.; Long, E. C.; Georgiadis, M. M. A Host-Guest Approach for Determining Drug-DNA Interactions: An Example Using Netropsin. *Nucleic Acids Res.* **2005**, *33* (13), 4106–4116.

(28) Nguyen, G. T. H.; Leung, W. Y.; Tran, T. N.; Wang, H.; Murray, V.; Donald, W. A. Mechanism for the Binding of Netropsin to Hairpin DNA Revealed Using Nanoscale Ion Emitters in Native Mass Spectrometry. *Anal. Chem.* **2020**, *92* (1), 1130–1137.

(29) Durand, M.; Thuong, N. T.; Maurizot, J. C. Binding of Netropsin to a DNA Triple Helix. *J. Biol. Chem.* **1992**, *267* (34), 24394–24399.

(30) Rentzeperis, D.; Marky, L. Ligand Binding to the Hoogsteen-WC Groove of TAT Base Triplets. Thermodynamic Contribution of the Thymine Methyl Groups. *J. Am. Chem. Soc.* **1995**, *117* (19), 5423–5424.

(31) Boehm, B.; Whidborne, C.; Button, A.; Pukala, T.; Huang, D. DNA Triplex Structure, Thermodynamics and Destabilisation: Insight from Molecular Simulations. *Phys. Chem. Chem. Phys.* **2018**, No. 20, 14013–14023.

(32) Nunn, C. M.; Garman, E.; Neidle, S. Crystal Structure of the DNA Decamer d(CGCAATTGCG) Complexed with the Minor Groove Binding Drug Netropsin. *Biochem* **997**, *36* (16), 4792–4799.

(33) Kanakis, C. D.; Tarantilis, P. A.; Polissiou, M. G.; Diamantoglou, S.; Tajmir-Riahi, H. A. DNA Interaction with Naturally Occurring Antioxidant Flavonoids Quercetin, Kaempferol, and Delphinidin. *J. Biomol. Struct. Dyn.* **2005**, *22* (6), 719–724.

(34) Ren, W.; Qiao, Z.; Wang, H.; Zhu, L.; Zhang, L. Flavonoids: Promising Anticancer Agents. *Med. Res. Rev.* **2003**, *23* (4), 519–534.

6.6 Appendix 1: Mass spectra of anthocyanin binding affinity experiments

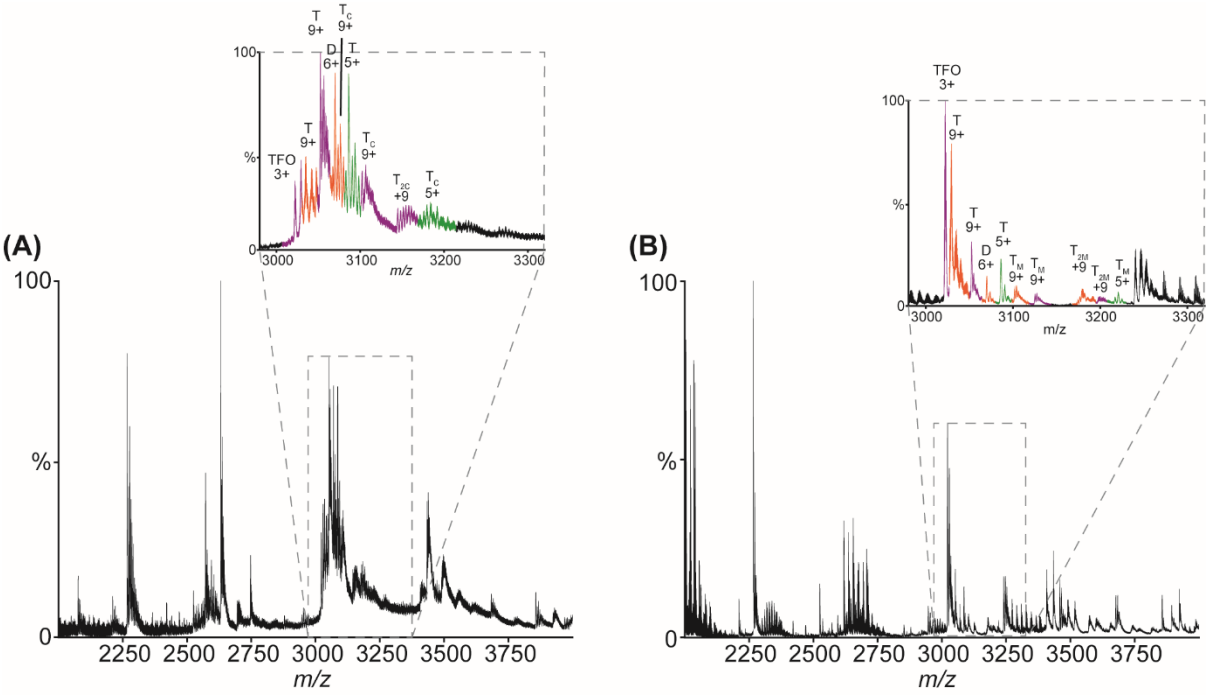


Figure A1: Mass spectra of the combined $(T)_{30}$, $(CT)_{15}$ and *rlmN* DNA triplex system with **(A)** cyanidin 3-glucoside and **(B)** malvidin 3,5-diglucoside, in 250 mM ammonium acetate solution at pH 7.4. Mass spectra were collected in positive ion mode. Samples were prepared in a 1:1:1:2 TFO: purine: pyrimidine:anthocyanin ratio, with each individual strand and anthocyanin at a concentration of 2 µM and 4 µM respectively. The insets in **(A)** and **(B)** indicate $(T)_{30}$, $(CT)_{15}$ and *rlmN* species coloured in purple, orange and green respectively. Annotations are as follows: TFO: TFO, D: duplex, T: triplex, T_C: triplex with cyanidin 3-glucoside bound, T_{2C}: triplex with two cyanidin 3-glucoside molecules bound, T_M: triplex with malvidin 3,5-diglucoside bound and T_{2M}: triplex with two malvidin 3,5-diglucoside molecules bound. Charge states are indicated on the spectra.

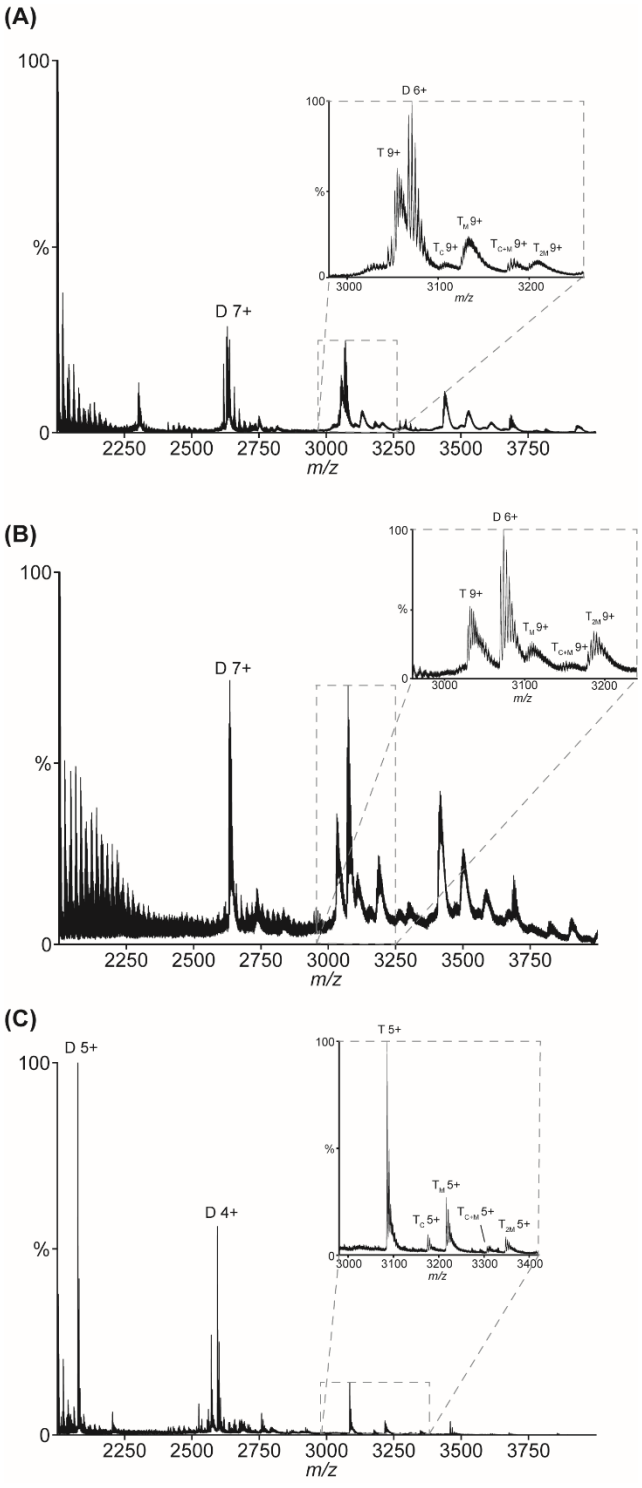


Figure A2: Mass spectra of the **(A)** $(T)_{30}$, **(B)** $(CT)_{15}$ and **(C)** *rlmN* DNA triplex systems with both cyanidin 3-glucoside and malvidin 3,5-diglucoside, in 250 mM ammonium acetate solution at pH 7.4. Mass spectra were collected in positive ion mode. Samples were prepared in a 1:1:1:2 TFO: purine: pyrimidine:anthocyanin ratio, with each individual strand and anthocyanin at a concentration of 2 µM and 4 µM respectively. Annotations are as follows: D: duplex, T: triplex, T_C: triplex with cyanidin 3-glucoside bound, T_M: triplex with malvidin 3,5-diglucoside bound, T_{2M}: triplex with two malvidin 3,5-diglucoside molecules bound and T_{C+M}: triplex with one cyanidin 3-glucoside and one malvidin 3,5-diglucoside molecules bound. Charge states are indicated on the spectra.

Chapter 7: Conclusions

This thesis describes the application of native MS for investigating the formation of various model and biological DNA triplex systems. The integration of UV-Vis spectroscopy and isothermal titration calorimetry was also employed to draw comparisons, revealing consistent behaviours in both gas and solution phases. Additionally, this approach provides valuable insights into the diverse conditions under which these triple helical structures form.

Chapter 3 explored the use of native MS to investigate the different conditions that can influence the detection of triplexes, which include buffer concentration, ionisation mode and pH. Increasing the ammonium acetate concentration in the triplex system was shown to increase the detection of triplex. This makes the concentration of salt in the buffer system a very important aspect to consider ensuring that there is no overstablisation of the DNA triplex structure for more informative comparisons to biological conditions. Furthermore, both positive and negative ionisation modes showed differences in species detected in the triplex systems. These differences have not been addressed in the literature previously and indicate that the use of both ionisation polarities for native MS would be valuable for the analysis of triplex systems in the future. Under acidic conditions, MS revealed the formation of unexpected multimers in the pH-dependent $(CT)_{15}$ sequence. UV-Vis spectroscopy and ITC were unable to directly inform on the formation of these multimers, which shows the importance of coupling these techniques with MS to fully understand the assemblies forming in these systems.

Chapter 4 probed the use of DNA base modifications in the TFO as a means of forming more stable triplex structures with model and biological sequences containing a pyrimidine interruption within the purine-rich TTS. LNA base modifications proved to have the highest triplex forming propensity when compared to canonical DNA and dSpacer TFOs, due to its ability to lock the conformation of the DNA backbone, allowing for enhanced Hoogsteen bonding and increased base stacking interactions. This allowed for the investigation of biologically relevant sequences to be investigated, specific to *Pseudomonas aeruginosa*. The *xcpU* gene target triplex was only able to form in the presence of LNA-modified bases, indicating the need for synthetic approaches to help the stability of biologically relevant triplexes and further development of antigene approaches.

Chapter 5 investigated the *rlmN* gene-specific sequence for antigene targeting in *Staphylococcus aureus*. Triplexator was used to identify the *rlmN* antigene site from the *MRSA252* genome. Experimental validation through biophysical techniques supported the suitability of the *rlmN* sequences to be used for an antigene strategy. To enable the formation of the triplex *in vivo*, we employed the KFFKFFKFFK CPP to aid with delivery into the cells. The successful synthesis of the TFO-CPP conjugate was achieved through a [3+2] azide-alkyne cycloaddition reaction. The 3' variant of the TFO-CPP conjugate without the PEG4 linker demonstrated the highest triplex forming propensity *in vitro*. Antimicrobial studies revealed that the 3' TFO-CPP conjugated exhibited concentration-dependent antimicrobial activity, albeit less potent than the kanamycin positive control. These findings provide insights into the potential applications of the antigene strategy, which, with further refinements of the approach, could contribute to solving issues with antibiotic-resistant bacteria.

Chapter 6 explored the DNA triplex binding abilities of two anthocyanins; cyanidin 3-glucoside and malvidin 3,5-diglucoside, utilising MS. The binding of up to three individual anthocyanin molecules to each of the model sequences and the *rlmN* gene sequence was observed via MS, representing the first time this has been observed using gas phase techniques. In previous literature, solution-phase techniques indicated that diglucosidic anthocyanins were unable to bind to DNA triplexes. However, it has been shown in this study that malvidin 3,5-diglucoside can bind to all three of the tested triplex sequences. The $(T)_{30}$ triplex exhibited a unique stabilising effect when bound to both anthocyanins as indicated by UV-Vis spectroscopy experiments. Although more computational and thermodynamic experiments are suggested to further elucidate the impact on the structure and binding strength of the anthocyanins on triplexes, this finding presents a potential therapeutic application of using anthocyanins to target and stabilise specific DNA triplexes. More study is required to understand the binding modes of anthocyanins for further development of DNA triplex binders.

Our experimental findings from this thesis have shown there is precedence for further investigation into the use of DNA triplexes for antigene applications. Future studies would benefit from utilising more MS-based approaches to thoroughly investigate the dynamics that underlie DNA triplex formation and activity, which would result in identifying suitable triplex gene targets. In addition, the development of synthetic oligonucleotides is a promising research direction for developing more stable DNA triplexes that control the expression of key survival genes in antibiotic-resistant bacteria. Base and backbone modifications in the TFO

are crucial for the future development of forming triplexes *in vivo* as they can increase their stability by increasing the rigidity of the TFO, the binding affinity to the duplex target and prevent TFO degradation caused by nucleases in the cell. Refining the TFO delivery method is also essential factor for the success of the antigene strategy. The use of CPPs is invasive and potentially destructive on its own to the target bacteria. Other methods including the use of lipid nanoparticles could be implemented in the future as an alternative method of delivery. However, there are numerous proposed gene expression mechanisms using higher-order DNA structures that need to be understood before these applications can be fully developed. Although the use of TFOs for clinical purposes has yet to materialise, the preliminary work from this thesis forms a structure-activity framework that will lead to the development of therapeutics based on an antigene strategy in the future.