Discrete Assembly of Synthetic Peptide-DNA Triplex Structures from Polyvalent Melamine-Thymine Bifacial Recognition

THESIS

Presented in Partial Fulfillment of the Requirements for the Degree Master of Science in the Graduate School of The Ohio State University

By

Yingying Zeng, B.S.

Graduate Program in Chemistry

The Ohio State University

2013

Master's Examination Committee:

Dr. Claudia Turro, Advisor

Dr. Jon Parquette
Abstract

Triplex-forming oligonucleotides (TFOs) and peptide nucleic acids (PNAs) have been known to direct DNA triplex assembly via Hoogsteen or reverse Hoogsteen base-pairing. However, with this conventional approach, target DNA sequences are almost solely homopurines. While the Janus-Wedge targeting concept has also been an elegant expansion of the recognition code for DNA triplex formation, it was largely restricted to preformed duplex assembly. There are fewer synthetic systems to induce triplex structures in single stranded oligonucleotides. With these goals in mind, we have designed and synthesized a 21-residue α-peptide (EM*)_{10}G that simultaneously recognizes two oligodeoxythymidine (dT_{10}) tracts to form triplexes with a peptide-DNA strand ratio of 1:2. The synthetic peptide has a controlled display of 10 melamine rings on lysine side chains (termed as M*), which addresses bifacial thymine-recognition interfaces along the length of the 21-residue peptide. Alternate residues are glutamic acid bearing negative charges to afford aqueous solubility and avoid nonspecific electrostatic binding with DNA. Binding stoichiometry was obtained by UV and fluorescence titration, indicating the formation of ternary peptide•[dT_{10}]_{2} complex as well as heterodimeric peptide•[dT_{10}C_{10}T_{10}] hairpin structure with triplex stems. Signal change on circular dichroism (CD) provided the direct evidence of DNA conformational change.
induced by the synthetic peptide. Clean transition trend from DNA to DNA-peptide complex bands based on native electrophoresis further supported discrete peptide-DNA assembly. Monophasic transition with cooperative melting was observed for both triplex and hairpin by UV and fluorescence denaturation. Highly exothermic assembly profiles (-28.5 kcal/mol for triplex and -31.4 kcal/mol for hairpin per peptide-DNA triplet stack) from differential scanning calorimetry (DSC) were comparable to those of DNA-DNA and melamine-cyanuric acid recognition, suggestive of similar driving forces. Binding affinity was quantified using fluorescence quenching and fluorescence anisotropy, yielding apparent dissociation constant (K_d) 4000 nM^2 for the triplex and 2.7 nM for the hairpin. Recognition exhibited selectivity for thymine over other nucleobases. Furthermore, partial and full methylation of melamine rings on the peptide abolished all detectable binding to dT_{10} tract, supporting the idea that assembly depends on the hydrogen-bonding directed melamine-thymine bifacial recognition rather than nonspecific aggregation. We anticipate that this novel assembly element may present promising artificial regulator to manipulate the structure and function of thymine- or-uracil rich targets.
Dedication

To my family
Acknowledgments

As I reflect on my graduate study here at OSU, I am left in awe of how quickly time has been passed and also glad to see myself growing toward an independent researcher and mature individual. I would like to take this opportunity to acknowledge the great people who have made this journey a worthwhile experience.

I owe a great debt of gratitude to my current research advisor, Professor Claudia Turro, who took me in when I thought I could never see the program through. Her mentorship and advice have motivated me through the tough times. I highly appreciate the opportunity that she gave me to finish my master’s examination. I can never thank her enough for her support and guidance. I would also like to thank the member of my thesis committee, Professor Jon Parquette for offering invaluable advice and support.

I am also grateful to my former research advisor, Professor Dennis Bong, for giving me such a great research opportunity to work on the project presented in this thesis. I highly treasure the time being part of his research team. Through those inspiring discussions with him, I have learned the rational and critical way of thinking when approaching a scientific problem. It will be invaluable towards my future career development.

Throughout the years in the Bong lab, I’ve had the tremendous fortune to work with a group of brilliant and motivated researchers. I would like to thank them all for everything that I have learned. Particularly, among the former graduate students, I am glad to have
known Mingming Ma and Oscar Torres, who are great scientists and friends. I am very grateful for their helpful discussions and advice on my academic study here.

In addition, I would like show my appreciation to the incredibly supportive staff here in the chemistry department.

Finally and mostly, my greatest debt of gratitude is to my family, for their faith, love and support in my Scientific Odyssey.
I am a part of all that I have met.

Yet all experience is an arch where-thro’

Gleams that untravell’d world, whose margin fades

Forever and forever when I move.

— From ULYSSES, Alfred, Lord Tennyson
Vita

2002-2005………………………………………………Rui’an the 4th Middle School, China
2005-2009……………………………………….B.S., Chemistry, Sichuan University, Chengdu, China
2009-2013……………………………………….Graduate Teaching/ Research Assistant, Department of Chemistry, The Ohio State University

Publications

1. Zeng, Y.; Pratumyot, Y; Piao, X.; Bong, D. Discrete Assembly of Synthetic Peptide-DNA Triplex Structures from Polyvalent Melamine-Thymine Bifacial Recognition. 

Fields of Study

Major Field: Chemistry
# Table of Contents

Abstract ................................................................................................................................. ii  

Dedication ............................................................................................................................ iv  

Acknowledgments ................................................................................................................ v  

Vita ....................................................................................................................................... viii  

List of Tables ......................................................................................................................... xii  

List of Figures ....................................................................................................................... xiii  

Chapter 1. Introduction to Sequence-specific DNA Recognition by Triplex-Forming Molecules (TFMs) ...................................................................................................................... 1  

1.1. Background and Significance ......................................................................................... 2  

1.2. Structural Features of DNA ......................................................................................... 3  

1.3. Strategies to Construct DNA Triplex ............................................................................. 5  

1.3.1. Hoogsteen or Reverse Hoogsteen Approach ............................................................. 5  

1.3.1.1. Nature’s Solution: Triplex-forming oligonucleotides (TFOs) ................................. 5  

1.3.1.2. Structural Features ............................................................................................... 8  

1.3.1.3. TFOs Modifications ............................................................................................. 10  

ix
Chapter 1. Introduction - 1
1.1. Introduction - 1
1.1.1. hairpin loop formation by PNA. - 4
1.1.2. Tetrad formation by PNA - 9
1.2. Materials and Methods - 15
1.2.1. Materials - 15
1.2.2. Methods - 15
1.3. Design and Characterization of PNA - 18
1.3.1. Triplex-forming Peptide Nucleic Acids (PNAs) - 18
1.3.1.4. Triplex-forming Peptide Nucleic Acids (PNAs) - 15
1.3.2. Janus-Wedge Approach - 21

Chapter 2. Experimental: Design, Synthesis and Characterization - 28
2.1. DNA Triplex-forming Peptide Design - 29
2.2. Materials and Methods - 32
2.2.1. General - 32
2.2.2. Synthetic schemes for monomer preparation - 33
2.2.3. Synthetic procedures - 34
2.2.4. Solid phase peptide synthesis and characterization - 39
2.2.5. General sample preparation protocol for binding studies - 45
2.2.6. UV Job Plot Experiments - 45
2.2.7. Fluorescence Job Plot (Quenching) Experiments - 45
2.2.8. Circular Dichroism (CD) Experiments - 46
2.2.9. Thermal Denaturation Studies by UV - 46
2.2.10. Thermal Denaturation Studies by Fluorescence - 47
2.2.11. Differential Scanning Calorimetry (DSC) - 47
2.2.12. Fluorescence and Fluorescence Anisotropy - 49
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2.13. Gel Mobility Shift Assay</td>
<td>51</td>
</tr>
<tr>
<td>Chapter 3. Results and Discussion</td>
<td>53</td>
</tr>
<tr>
<td>3.1. Binding Stoichiometry</td>
<td>54</td>
</tr>
<tr>
<td>3.2. CD and Gel Mobility Shift</td>
<td>56</td>
</tr>
<tr>
<td>3.3. Thermal Stability</td>
<td>58</td>
</tr>
<tr>
<td>3.4. Binding Affinity</td>
<td>60</td>
</tr>
<tr>
<td>3.5. Binding Selectivity</td>
<td>61</td>
</tr>
<tr>
<td>Chapter 4. Conclusion</td>
<td>66</td>
</tr>
<tr>
<td>Reference</td>
<td>69</td>
</tr>
<tr>
<td>Appendix A: Additional Characterization Data</td>
<td>77</td>
</tr>
<tr>
<td>Appendix B: $^1$H and $^{13}$C NMR Spectra of Amino Acid Derivatives</td>
<td>81</td>
</tr>
</tbody>
</table>
List of Tables

Table 1.1. Thermal Stabilities of PNA-DNA triplexes. .................................................. 17

Table 1.2. Thermal stabilities of J-W triplexes. .............................................................. 26
List of Figures

Figure 1.1. Structural features of DNA double helix. (A) The sugar-phosphate backbones of complementary nucleotide strands are shown in red and orange, and the Watson-Crick base pairs are shown in gray. (B) The chemical structures and electro-features of DNA bases thymine (T), adenine (A), cytosine (C), and guanine (G) are shown as hydrogen-bonded base pairs. .......................................................... 3

Figure 1.2. Oligonucleotide-directed cleavage of double helical DNA by a triplex-forming DNA-EDTA•Fe probe. .......................................................... 6

Figure 1.3. (Left) Simplified model of the triple helix complex between a Hoogsteen bound DNA-EDTA•Fe probe at single site within 4.06 kb of plasmid DNA. (Right) Isomorphous base triplets of TAT and C+GC. ............................................. 7

Figure 1.4. Hydrogen bonding pattern of base triplets formed by natural nucleobases through either Hoogsteen (upper left) or reverse Hoogsteen (lower). ......................... 8

Figure 1.5. Classification of binding configurations of TFOs according to their compositions. ......................................................................................................................... 9

Figure 1.6. Base triplets of C\textsuperscript{+}-G-C and P1-G-C .......................................................... 11

Figure 1.7. Hydrogen bonding pattern of m\textsuperscript{5ox}C-G-C base triplet. ......................... 11

Figure 1.8. Hoogsteen hydrogen bonding of antiTA, antiAT, antiCG and antiGC C-glycosides to their Watson-Crick base pair partners................................................. 13
Figure 1.9. Design of rationale for recognition of Watson-Crick CG base pair by nonnatural bases D$_2$ and D$_3$ within a pyrimidine-purine-pyrimidine triplex motif. ........ 13

Figure 1.10. Representative structure of the 3’, 5’-NDI bisconjugate with NDI intercalators attached to both the 3’ and 5’ end of an oligonucleotide. ...................... 14

Figure 1.11. (Upper) A DNA-naphthalene diimide conjugate; (Lower) A DNA-perylene diimide conjugate. .............................................................................................................. 15

Figure 1.12. Generic chemical structure of PNAs (B = thyminyl) studied...................... 16

Figure 1.13. Schematic representation of strand displacement PNA binding mode upon targeting double stranded DNA. ................................................................. 18

Figure 1.14. Triplex-forming peptide nucleic acid (PNA).............................................. 19

Figure 1.15. The studied PNA and DNA sequences...................................................... 20

Figure 1.16. Diagram of the bis-PNA-DNA triplex...................................................... 21

Figure 1.17. Lehn’s Janus-Wedge concept: Janus wedges with two hydrogen bonding faces (A: acceptor, D: donor) are designed to bind by insertion between base-pairs forming a triplet with the maximum number of Watson-Crick interactions. ............... 23

Figure 1.18. Lehn’s Janus-Wedge molecule capable of forming a triad with thymine and cytosine derivatives in CHCl$_3$.................................................................................. 23

Figure 1.19. (a) A Janus-Wedge base triplet: the third strand residue W binds to the Watson-Crick faces of both target residues; (b) the 11dC$_8$-11/11dT$_8$11 DNA target sequence. ........................................................................................................... 25
Figure 1.20. Janus-Wedge base triplets: (a) the third-strand residue $W_1$ binds to the W-C faces of target A and T; (b) $W_2$ binds to G and C similarly.

Figure 2.1. Bifacial melamine-thymine recognition.

Figure 2.2. Synthesis of monomer 1.

Figure 2.3. Synthesis of monomer 3.

Figure 2.4. Synthesis of monomer 4.

Figure 2.5. (A) MALDI-TOF of peptide 1; (B) HPLC trace of of peptide 1 on a C-18 analytical column using a gradient of 10-100% solvent B over 50 min, monitored by a UV-Vis detector at 230 nm.

Figure 2.6. (A) MALDI-TOF of peptide 2; (B) HPLC trace of of peptide 2 on a C-18 analytical column using a gradient of 10-100% solvent B over 50 min, monitored by a UV-Vis detector at 230 nm.

Figure 2.7. (A) MALDI-TOF of peptide 3; (B) HPLC trace of of peptide 3 on a C-18 analytical column using a gradient of 10-100% solvent B over 50 min, monitored by a UV-Vis detector at 230 nm.

Figure 2.8. (A) MALDI-TOF of 5(6)-Carboxyfluorescein (Cbf) labeled peptide 3; (B) HPLC trace of of Cbf labeled peptide 3 on a C-18 analytical column using a gradient of 10-100% solvent B over 50 min, monitored by a UV-Vis detector at 230 nm.
Figure 2.9. (A) MALDI-TOF of peptide 4; (B) HPLC trace of peptide 4 on a C-18 analytical column using a gradient of 10-100% solvent B over 45 min, monitored by a UV-Vis detector at 230 nm.................................44

Figure 2.10. DSC traces of triplex 5 (black) with Non-2-State binding model fitted curves shown in red. (A) Upscan and (B) Downscan..................................................48

Figure 2.11. DSC traces of Hairpin 6 (black) with Non-2-State binding model fitted curves shown in red. (A) Upscan and (B) Downscan..................................................49

Figure 2.12. Representative fitting data set for triplex 5 ..................................................50

Figure 2.13. Representative fitting data set for hairpin 6..................................................51

Figure 3.1. Peptide 1 titrated into (A) dT_{10} and (B) Flc-dT_{10}C_{10}T_{10}Dabcyl, followed by UV absorbance (260 nm) and fluorescein emission (521 nm), respectively. ..................54

Figure 3.2. UV absorbance (260nm) of peptide 1 alone at different concentrations...........55

Figure 3.3. Peptide 1 titrated into dT_{10}C_{10}T_{10} followed by UV absorbance (260 nm). ...56

Figure 3.4. Circular dichroism spectra in DPBS, pH 7.4, of (A) Peptide 1 complexed with dT_{10} (-) vs dT_{10} alone (--) and (B) Peptide 1 complexed with dT_{10}C_{10}T_{10} (O) vs dT_{10}C_{10}T_{10} alone(--). Peptide 1 in both is at 5µM concentration (—), while dT_{10} and dT_{10}C_{10}T_{10} are maintained at 10 and 5 µM, respectively. Electrophoretic mobility shift assays imaged by Cy5 fluorescence for (C) Cy5-dT_{10} (DNA₁) and (D) Cy5-dT_{10}C_{10}T_{10} (DNA₂) at 20 nM in each lane, with increasing peptide 1 concentration from left to right.
(E) Relative electrophoretic mobilities of the free DNA oligos and their peptide complexes, a mixture of complex 5 and DNA1 shown in the central lane. .......................... 57

**Figure 3.5.** (A) First-derivative plot of melting transitions of triplex 5 (—) and hairpin 6 (—) followed by UV absorbance (260 nm). Normalized absorbance change is shown inset. (B) DSC upscan traces of triplex 5 (—) and hairpin 6 (—), with downscan traces shown as dashed regular and bold lines. ................................................................. 58

**Figure 3.6.** (A) Fluorescence melting curves for hairpin 6 (—) and free Flc-dT10C10T10-Dabcyl (—); (B) Corresponding first derivative curves for hairpin 6................................. 59

**Figure 3.7.** Binding isotherms in DPBS, pH 7.4, of (A) peptide 2 binding to dT10 followed by fluorescein quenching upon binding and (B) peptide 2 binding to dT10C10T10 followed by fluorescence anisotropy. Solid lines show fits to (A) trimer-monomer 1:2 binding model ([fraction bound peptide 2] = [DNA]²/Kₐ + [DNA]², R² ≥ 0.96) and (B) 1:1 binding model (corrected for fluorescence quenching): [Bound peptide 2] = (((([R₇]+[DNA₇]+Kₐ)-(sqrt(((([R₇]+[DNA₇]+Kₐ)²)-4*[R₇]*[DNA₇]])))))/2), where [R₇] is the total concentration of peptide 2 (25 nM), [DNA₇] is the total concentration of dT10C10T10 used in each binding reaction. R² > 0.98. ...................................................... 61

**Figure 3.8.** (A) Absorbance (260 nm) vs Temperature plots for control peptide 3 and peptide 4 with dT10. Concentrations: 2 µM peptide with 2 µM dT10. (B) Electrophoretic mobility shift assay imaged by Cy5 fluorescence for Cy5-dT10 at 150 nM with increasing concentrations of peptide 3 (0, 30, 50, 100, 150, 225, 300, 750, 1200, 1500 nM) from left
to right. Study for peptide 4 and Cy5-dT10 under the same condition yielded the same image.

**Figure 3.9.** (A) Absorbance (260 nm) vs Temperature plots for peptide 1 with control oligos dA10 and dC10. Absence of UV shift or melting signatures. Concentrations: 1 µM peptide 1 with 2 µM dA10 or dC10. (B) Fluorescence (O) and Anisotropy (Δ) assays for peptide 2 with control oligos dG5A10 and dC10. Peptide 2 concentration is constant at 25 nM. (C) Circular dichroism spectra for peptide 1 with control oligos dA10 and dC10. Concentrations: 5 µM peptide 1 with 10 µM dA10 or dC10.

**Figure A1.** Additional UV Job plot analysis of peptide 1 with dT10.

**Figure A2.** Additional CD spectrum: Titration of 10 µM dT10 with increasing concentration of peptide 1.

**Figure A3.** Additional CD spectrum: free dC5T10C6/dG6T10G5 (T10 tracts 10 µM) and their peptide 1 complexes.

**Figure A4.** Additional thermal denaturation curve for peptide 1 alone at different concentrations. Absence of UV melting signature rules out possible secondary structures from peptide itself.

**Figure A5.** Additional stoichiometric binding curves for triplex 5 or hairpin 6 by titrating 500 nM peptide 2 with dT10 or dT10C10T10. Both fluorescence (—) and anisotropy (O) were monitored for triplex 5. Only anisotropy was recorded for hairpin 6 (Δ).
Figure A6. Additional control Fluorescence (O) and Anisotropy (Δ) assays for free Cbf in solution with (A) dT10 and (B) dT10C10T10 to rule out non-specific binding between the fluorophore and DNA oligomers. ................................................................. 80

Figure B1. 1H NMR of 1a. .................................................................................................................. 82

Figure B2. 1H NMR of 1 .................................................................................................................. 82

Figure B3. 1H NMR of 3a. .................................................................................................................. 83

Figure B4. 1H NMR of 3b. .................................................................................................................. 83

Figure B5. 1H NMR of 3c. .................................................................................................................. 84

Figure B6. 1H NMR of 3 .................................................................................................................. 84

Figure B7. 1H NMR of 4a. .................................................................................................................. 85

Figure B8. 1H NMR of 4b. .................................................................................................................. 85

Figure B9. 1H NMR of 4 .................................................................................................................. 86

Figure B10. 13C NMR of 1a. ................................................................................................................. 86

Figure B11. 13C NMR of 1 ................................................................................................................. 87

Figure B12. 13C NMR of 3a. ................................................................................................................. 87

Figure B13. 13C NMR of 3b. ................................................................................................................. 88

Figure B14. 13C NMR of 3c. ................................................................................................................. 88

Figure B15. 13C NMR of 3 ................................................................................................................. 89

Figure B16. 13C NMR of 4a. ................................................................................................................. 89

Figure B17. 13C NMR of 4b. ................................................................................................................. 90
Figure B18. $^{13}$C NMR of 4. ................................................................. 90
Chapter 1

Introduction to Sequence-specific DNA Recognition by Triplex-Forming Molecules (TFMs)
1.1. Background and Significance

Nucleic acids play the most crucial role in the development of biodiversity on earth, the beginning of central dogma of molecular biology. Particularly, the DNA sequence (genome) carries the digital information to encode the logic of life, responsible for genetic information storage, heredity, expression and other metabolic functions. Moreover, certain DNA sequences are known to be toxic and associated with diseases. Therefore, the ability to specifically manipulate DNA information processing would offer a wide range of applications in biology, gene-based biotechnology and eventually gene-based therapeutics.\(^1\) One way to approach this goal would be the rational design of artificial ligands that can recognize and assemble to desired DNA sequences, thus can provide powerful tools to interfere with genetic information flow. Synthetic molecules can be made to bind to DNA with high sequence specificity and affinity when they are instructed to do so. The instructions exist in the form of the molecule’s shape, its chemical interfaces, and how well it fills the space where the recognition and assembly events take place which are all required to be ‘complementary’ to the targeted DNA fragment. These instructions are written into the molecule during its synthesis.\(^2\)
1.2. Structural Features of DNA

The importance of the DNA double helix with all those functions originates from its unique features of chemical structure and conformation. The biological relevant B-form DNA is naturally assembled in a double helix structure, which was discovered by Watson and Crick in 1953.\textsuperscript{3} Two complementary, antiparallel polydeoxyribonucleotide strands consist of highly negatively charged sugar-phosphate backbones and more importantly the stacked base pairs via specific hydrogen bonding. The chemical ‘signatures’ of the nucleobases are characterized by the pattern of functional groups exposed at their molecular surface (Figure 1.1). The hydrogen-bonding donor and acceptor arrays of the nucleobases presented by a given DNA sequence provide the basis for molecular recognition by naturally occurring or synthetic DNA ligands.

\textbf{Figure 1.1.} Structural features of DNA double helix. (A) The sugar-phosphate backbones of complementary nucleotide strands are shown in red and orange, and the Watson-Crick base pairs are shown in gray. (B) The chemical structures and electro-features of DNA bases thymine (T), adenine (A), cytosine (C), and guanine (G) are shown as hydrogen-bonded base pairs. Adapted from reference 4.
More interestingly, DNA is polymorphic macromolecules that can adopt a variety of conformations that deviate from B-form duplex structure, such as single-, triple-, multi-stranded complex as well as other unusual structures,\textsuperscript{5,6} which may in turn reflect important signals during DNA replication as well as gene regulation, recombination and mutation.\textsuperscript{7} Therefore, the formation of non-B-form duplex structures could potentially present great opportunities to probe and modulate genetic information processing. In particular, among related scientific efforts, DNA triplex structures including design strategies and potential applications have been under intensive investigation.
1.3. Strategies to Construct DNA Triplex

There are two major classes of DNA triplex-forming molecules (TFMs) according to the targeting concept: (1) via Hoogsteen or reverse Hoogsteen base pairing; (2) via Janus-Wedge base pairing.

1.3.1. Hoogsteen or Reverse Hoogsteen Approach

1.3.1.1. Nature’s Solution: Triplex-forming oligonucleotides (TFOs)

Intermolecular triplexes can be formed between triplex-forming oligonucleotides (TFOs) and the target sequences on duplex DNA. The very first example of nucleic acids triple helix structure was found in synthetic polyribonucleotides in 1957. OD titration and ultracentrifuge indicated a stable complex involving two poly(U) strands for each poly(A) strand in the presence of 10mM MgCl₂. Besides, this sort of association behavior exhibited a fair amount of specificity as the (A+U) didn’t pick up poly(C), poly(I) or poly(A) as the third strand but only poly(U). During the three decades that followed this initial discovery, several other polynucleotides triplex structures were reported. Poly(C) could also form a triplex with poly(G) at low pH. More importantly, the X-ray diffraction patterns of triple-stranded fibers poly(A)•2poly(U) and poly(dA)•2poly(dT) suggested an A-RNA-like conformation of the two Watson-Crick base-paired strands with the third strand in a parallel orientation bound to the homopurine strand of the duplex by Hoogsteen hydrogen bonds.

It was not demonstrated until 1987 that an oligonucleotide can bind sequence-specifically to an oligopurine-oligopyrimidine duplex tract in the major groove. Dervan and co-
workers designed a series of homopyrimidine oligonucleotides with EDTA•Fe covalently attached at a single position that can specifically bind to homopurine-homopyrimidine tracts within a large DNA duplex by triplex formation and cleave at that site (Figure 1.2 and 1.3). The specific location and asymmetry of the cleavage pattern provided a definite proof that the homopyrimidine-EDTA probes bind in the major groove parallel to the homopurine strand of Watson-Crick double helical DNA. Such oligonucleotides in this context are nowadays referred to as triplex-forming oligonucleotides (TFOs).

Figure 1.2. Oligonucleotide-directed cleavage of double helical DNA by a triplex-forming DNA-EDTA•Fe probe. One thymine has been replaced by thymine with the iron chelator EDTA covalently attached at C-5. Reduction of oxygen generates localized hydroxyl radical at this position. Adapted from reference 16.
This seminal study had important implication for the ability of TFOs to induce local conformational changes in B-DNA and more importantly regulate gene information process artificially when equipped with functional moieties. It paved the way that led to today’s understanding and development of the chemotherapeutic and antigene strategy based on TFMs. Similar work was also presented by Hélène and co-workers. Since then, TFOs made of native oligonucleotides or their analogs have been extensively explored to direct DNA triplex assembly.
1.3.1.2 Structural Features

Triplex formation via Hoogsteen or reverse Hoogsteen base pairing (Figure 1.4 and Figure 1.5) is based on the fact that the base pairs in the duplex still have extra hydrogen bond acceptors and donors available in the major groove for specific recognition by native nucleobases or their analogs.

**Figure 1.4.** Hydrogen bonding pattern of base triplets formed by natural nucleobases through either Hoogsteen (upper left) or reverse Hoogsteen (lower). Base triplets are described as T•AxT, C•GxC, C•GxG, and T•AxA with (•) representing Watson-Crick and (x) representing Hoogsteen or reverse Hoogsteen. TFOs’ strand orientations are indicated by + (parallel orientation with respect to the homopurine strand ) or – (anti-parallel orientation). An intermolecular DNA triple helix is schematically represented (upper right) with the third strand TFO in black. Adapted from reference 1.
In Hoogsteen approach, the third strand is bound parallel to the homopurine strand whereas anti-parallel in reverse Hoogsteen. However in both Hoogsteen and reverse Hoogsteen cases, triplex formation must also follow the binding rules originate from chemical structure of nucleotides and maximum hydrogen bonds formed for energy reasons. The target sequence must contain homopurine tract because only purine bases can establish extra hydrogen bonds with incoming third base in the major groove. In addition, pyrimidine-rich TFOs composed of T and C nucleotides bind in a parallel orientation via Hoogsteen pattern, forming T•AxT and C•GxC+ base triplets. Such triplexes are strongly pH dependent as hemi-protonation of C is required. Purine-rich TFOs composed of A and G bind in an anti-parallel orientation via reversed-Hoogsteen...
pattern, forming T•AxA and C•GxG base triplets. However, TFOs contain T and G can bind in either parallel or anti-parallel manners depending on the sequence\textsuperscript{18} (Figure \textbf{1.5}).

1.3.1.3. TFOs Modifications

The ability of TFOs made of native nucleotides to be used in biological applications is largely limited by their chemical and physical properties: (1) target sequence is almost exclusively purine tract; (2) the requirement of protonation for cytosine especially adjacent cytosines. In general, this can only be realized in a relatively lower pH environment than physiological condition; (3) purine-rich TFOs tend to self-assemble into G-quadruplex or other more complicated structures at physiological ionic strength; (4) the charge repulsion from sugar-phosphate backbone needs to be neutralized and the stability of triplex is highly dependent on the presence of relatively higher concentration of divalent cations such as Mg\textsuperscript{2+} than that is available \textit{in vivo}. In order to overcome some of the limitations listed above, a great number of scientific efforts have been devoted to modifying the chemical structures of TFOs during the last two decades, including base modification, sugar-phosphate backbone modification, covalent attachment of intercalating regents at the 5’ and/or the 3’ end of TFOs, \textit{etc}. Much progress has been made toward improving their binding affinity and specificity as well as resistance to nucleases in a cellular environment.

1.3.1.3.1. Base Modifications

The use of cytosine derivative\textsuperscript{19,20} for recognition of GC base pair such as 5-methylcytosine reported first by Dervan and van Boom\textsuperscript{21,22} can partially alleviate the pH
restriction of TFOs, possibly due to the enhancement of base stacking from methyl group or exclusion of water molecules from major groove. Nonnatural deoxynucleosides P1 (Figure 1.6) designed by Dervan and co-workers were also able to circumvent the protonation requirement for C\(^+\)GC base triplet formation.\(^{23}\)

![C+GC and P1-GC](image1.png)

**Figure 1.6.** Base triplets of C\(^+\)-G-C and P1-G-C. Adapted from reference 23.

![Hydrogen bonding pattern](image2.png)

**Figure 1.7.** Hydrogen bonding pattern of m\(^{50}\)C-G-C base triplet. Adapted from reference 24.
McLaughlin and co-workers also reported 2’-deoxy-5-methylcytidine (m^{5ox}C) as an analog of an N3-protonated cytosine derivative (Figure 1.7).\textsuperscript{24} When the m^{5ox}C-G-C base triplets were present in sequence positions that alternate with TAT base triplets, DNA triplexes were pH independent in the range of 6.4-8.5.

More interestingly, a number of other modified nucleobases and nonnatural analogs have also been synthesized with respect to expanding the target DNA sequence repertoire by triplex formation. The ability of TFOs to bind to pyrimidine tracts or hetero purine-pyrimidine sequences was addressed (Figure 1.8 and 1.9). One particular strategy reported by Barry Gold\textsuperscript{25} was very promising. The use of different modified bases to specifically recognize T-A vs A-T and C-G vs G-C Watson-Crick base pairing eliminates the strand polarity issue since the third strand reads in one direction purine information in the major groove on either strand (Figure 1.8).
Figure 1.8. Hoogsteen hydrogen bonding of antiTA, antiAT, antiCG and antiGC C-glycosides to their Watson-Crick base pair partners. Adapted from reference 25.

Figure 1.9. Design of rationale for recognition of Watson-Crick CG base pair by nonnatural bases D₂ and D₃ within a pyrimidine-purine-pyrimidine triplex motif. Adapted from reference 26.
1.3.1.3.2. Intercalating Reagent

The attachment of an intercalating agent at the 5’ and/or the 3’ end of TFOs can strongly stabilize triplex formation without sequence-specificity being compromised.\textsuperscript{27} In the work presented by Hélène and co-workers, an acridine derivative was covalently linked to the 5’ end of a homopyrimidine oligonucleotide and specific binding to a homopurine-homopyrimidine DNA duplex was demonstrated. McLaughlin and co-workers also studied the triplex stabilizing properties of oligodeoxyribonucleotides functionalized at the 5’ and/or 3’ end with a naphthalene diimide-based (NDI) intercalator (Figure 1.10).\textsuperscript{28} Thermal denaturation data indicated a remarkable gain in stability for triplex formed using oligodeoxyribonucleotide-NDI conjugates.

![Figure 1.10](image)

**Figure 1.10.** Representative structure of the 3’,5’-NDI bisconjugate with NDI intercalators attached to both the 3’ and 5’ end of an oligonucleotide. Adapted from reference 28.
TFOs with intercalators modification at the middle of the sequence (Figure 1.11) was also contemplated by McLaughlin and co-workers.\textsuperscript{29} In this context, the perylene- and naphthalene diimide-based intercalators were employed as the linker of two homopyrimidine segments, thus providing conjugates capable of targeting single-stranded nucleic acids with the formation of hairpin triplexes. The intercalators were designed to bridge the terminal base triplet and participate in base-stacking with all three residues.

![Image](image.png)

**Figure 1.11.** (Upper) A DNA-naphthalene diimide conjugate; (Lower) A DNA-perylene diimide conjugate. Adapted from reference 29.

### 1.3.1.4. Triplex-forming Peptide Nucleic Acids (PNAs)

Oligonucleotides are difficult to prepare in large scale (millimole or mole quantities) and introduction of modified nucleobases and conjugation to other ligands present major problems. Besides, poor cellular uptake of oligonucleotides has to a large extent limited the potential of TFOs in molecular biology and medicine. Nielsen and colleagues\textsuperscript{30} suggest replacing the deoxyribose phosphate backbone with a polyamide backbone and the resulting peptide nucleic acid (PNA) is homorphous to DNA in terms of the number
of backbone bonds and the distance between backbone and nucleobase. The targeting concept here was similar to conventional TFO approach in that the third strand (either PNA or DNA) would bind to the purine target strand via Hoogsteen hydrogen bonding. The very first generation of homopyrimidine PNAs H-Tₐ-Lys-NH₂ (Figure 1.12 and Table 1.1) were prepared and evaluated by Nielsen and colleagues. The Lysine residue at C-terminus was for the purpose of aqueous solubility, electrostatic attraction as well as preventing PNAs from self-aggregation. Tₘ results showed that the designed PNAs formed hybrids with complementary purine strands. The complex stability was strongly susceptible to the presence of mismatch bases indicating sequence selectivity. The binding stoichiometry was determined by UV titration to be 2:1 PNA: DNA, which was the same with TFO approach except oligothymidine was employed to form triplex.

![Generic chemical structure of PNAs (B = thyminyl) studied. The structure of DNA is shown for comparison. Adapted from reference 30.](image)

**Figure 1.12.** Generic chemical structure of PNAs (B = thyminyl) studied. The structure of DNA is shown for comparison. Adapted from reference 30.
Further characterization of the binding between PNA H-T₈-Lys-NH₂ to a complementary single-strand poly(dA) chain by flow linear dichroism (LD) and circular dichroism (CD) confirmed the discrete PNA-DNA triplex assembly with 2:1 PNA: DNA binding ratio and that it was a right-handed helix. The base conformation of the poly (dA)-PNA₂ triplex was very similar to that of the conventional poly (A·T·T) triplex.

However, quite surprisingly, PNAs of sequence H-Tₙ-Lys-NH₂ was not able to form the expected DNA-DNA-PNA triplex with dA₁₀-dT₁₀ duplex tract, but rather by Watson-Crick hydrogen bonding to the complementary A strand and displacing the noncomplementary T strand (Figure 1.13).
In order to address the ability of PNA to invade the local duplex, displace the oligopyrimididine strand and meanwhile form an internal extremely stable triplex structure (triplex invasion binding mode), Egholm et al. designed a type of dimeric PNAs (bis-PNAs) where two monomeric PNA segments were covalently connected together via a flexible linker (Figure 1.14). One segment was designed for Watson-Crick recognition of DNA and the other designed for Hoogsteen recognition of PNA-DNA duplex. Consequently, the most stable DNA-PNA$_2$ triplex would be formed when the Watson-Crick base-pairing PNA segment is bound to the DNA strand in the anti-parallel orientation whereas the Hoogsteen strand is in the parallel orientation.
The binding data indicated that bis-PNAs were able to form triplex with oligonucleotides in a sequence-specific manner, of somewhat higher thermal stability than monomeric PNAs and the thermal melting transition showed very little hysteresis. Furthermore, bis-PNAs also exhibited superior binding properties in terms of targeting double-stranded DNA through triplex invasion mode.

A reasonable explanation for the enhanced triplex stability and rate of strand invasion was proposed by Griffith and co-workers\textsuperscript{36}: the entropy cost accompanying the assembly was essentially reduced and the binding reaction was converted into a bimolecular process by simply linking two PNA segments together. In their research work with similar bis-PNA designs, where a neutral polyethylene glycol (PEG) or a positively...
charged lysine/aminohexyl were both investigated as the linker (Figure 1.15). The binding affinities of these bis-PNAs for single-strand (ss) DNA and binding rates for double-strand (ds) DNA were significantly increased. Quite impressively, the stoichiometry of a 1:1 PNA-DNA complex was established by direct observation on electrospray mass spectrometry (ESI), indicating the robust stability of PNA-DNA triplex.

![PNAs and DNA Targets](image)

**Figure 1.15.** The studied PNA and DNA sequences. Aha, 6-aminohexanoic acid; 3PEG, H₂N(CH₂CH₂O)₃CH₃COOH. Adapted from reference 36.

Kinetics of strand invasion study demonstrated a relatively rapid on-rate and slow off-rate, which was consistent with the mechanism for PNA invasion of ds DNA proposed by Nielsen and colleagues. The second PNA strand essentially “locks” the association...
process in a triplex upon binding, which indeed was the rate-limiting step.

Interestingly, the crystal structure of a nine-base bis-PNA-DNA triplex (Figure 1.16) revealed a P-form helix, which differed from previously reported A-form like based on the original fiber diffraction studies of poly(dA)-[poly(dT)]$_2$ triplex. This important finding gained insight into the manner in which PNAs complexed with DNA to form triplexes. The observed hydrogen bonding geometries and distances of T•A-T and C•G-C triplets provided direct confirmation of the proposed recognition pattern: the purine base from each triplet forms a Watson-Crick base pair with the pyrimidine base on one PNA strand and a Hoogsteen base pair with the pyrimidine on the other PNA strand. Moreover, this unique structure expanded the repertoire of known stable helical forms that nucleic acids can adopt and it also suggest that there may be additional classes of stable DNA triplexes yet to be discovered.

![Figure 1.16](image-url)

**Figure 1.16.** Diagram of the bis-PNA-DNA triplex. Adapted from reference 39.
The novel binding modes presented by PNA and its synthetic availability have made this molecule a highly attractive ligand toward sequence specific DNA recognition. Nonetheless, only homopyrimidine PNAs have been explored extensively in triplex-forming field so far. This is largely due to the underlying Hoogsteen-mediated targeting concept. Thus, much more improvement is yet to be made, especially concerning the ability to target pyrimidine or mixed sequences, before full potential of PNAs can be accessed and evaluated. And this goal can be potentially achieved by expanding the library of the attached nucleobases and their analogs through chemical modification.

1.3.2. Novel Janus-Wedge Approach

A Janus-Wedge (J-W) recognition concept was first proposed by Lehn and colleagues in their work with heterocycles. It is based on the design of wedge-shaped heterocycles presenting two hydrogen-bonding arrays that are complementary to the Watson-Crick faces of a pair of canonical or mismatched nucleobases and thereby forming a base-wedge-base triad motif (Figure 1.17).

In this initial work, the incoming third heterocycle was able to recognize a cytosine-uracil mismatched base-pair by insertion in a wedge-like fashion utilizing the maximum number of Watson-Crick interactions (Figure 1.18). However, the first generation wedges 1a and 1b designed were lipophilic and all binding properties were characterized in CHCl₃.
Figure 1.17. Lehn’s Janus-Wedge concept: Janus wedges with two hydrogen bonding faces (A: acceptor, D: donor) are designed to bind by insertion between base-pairs forming a triplet with the maximum number of Watson-Crick interactions. In contrast, conventional DNA binding molecules target hydrogen bonding sites of the intact base-pair, as illustrated by black (or small open) arrows for the major (or minor) groove side of the AT and GC base-pair. Adapted from reference 40.

Figure 1.18. Lehn’s Janus-Wedge molecule capable of forming a triad with thymine and cytosine derivatives in CHCl₃. Adapted from reference 40.
McLaughlin and his co-workers\textsuperscript{41} followed up the investigation on this novel targeting concept and reported the first example of Janus-Wedge type DNA triplex formation. In the design, eight J-W bases (Wedge residue \textbf{W}) were attached to a PNA backbone, capable of base pairing with the Watson-Crick faces of cytosine-thymine mismatch sites in the DNA sequence 11dC\textsubscript{8}-11/11dT\textsubscript{8}11, leading to the formation of eight (C-W\textsubscript{1}-T) base triplets (Figure 1.19). A Lysine (K) residue was added at the C-terminus to afford aqueous solubility as well as complementary charge-charge interaction. Experiments suggest that the resulting PNA W\textsubscript{8}K bound to the DNA target in the major groove, parallel to the dC\textsubscript{8} strand and antiparallel to the dT\textsubscript{8} strand. It was noteworthy that thermal denaturation of the J-W triplex exhibited two-phase transitions, corresponding to triplex-duplex and duplex-random coil equilibrium respectively. Thermodynamic analysis confirmed that the J-W triplex was stabilized by a free energy of 15.2 kcal/mol, which was substantially greater than the conventional TFOs approach of similar sequence length.

They soon sought out to investigate the ability of J-W type PNAs to undergo strand invasion where the wedge residues were involved in hydrogen bonding with the Watson-Crick faces of canonical base pairs (A-T, G-C).\textsuperscript{42} In the study, a wedge residue \textbf{W}\textsubscript{1} was used to target A-T (or T-A by simple rotation of heterocycle), and \textbf{W}\textsubscript{2} for G-C (or C-G) base pairs (Figure 1.20).
Figure 1.19. (a) A Janus-Wedge base triplet: the third strand residue W binds to the Watson-Crick faces of both target residues; (b) the 11dC<sub>8</sub>-11dT<sub>8</sub>11 DNA target sequence. Adapted from reference 41.

Figure 1.20. Janus-Wedge base triplets: (a) the third-strand residue W<sub>1</sub> binds to the W-C faces of target A and T; (b) W<sub>2</sub> binds to G and C similarly. Adapted from reference 42.
A series of mixed 8-mer PNAs with a terminal lysine residue of sequence \((W_1)_m(W_2)_n(W_1)_aK\) were prepared and tested against a number of different DNA sequences (Table 1.2). \(T_m\) values supported relatively high binding affinities and selectivities. However the PNA 8-mer was unable to form a detectable J-W triplex when the target sequence contained A-T and G-C base pairs (i.e. no mismatch sites C-T), likely due to the reason that eight-residue PNA was not long enough to effectively invade the target duplex and form stable triplex. A longer sequence \((W_1)_{14}K\) targeting dA\(_{14}\)/dT\(_{14}\) duplex indicated the formation of a stable complex, although it was not clear whether the observed complex was actually J-W triplex.

Table 1.2. Thermal stabilities of J-W triplexes. Adapted from reference 42.

<table>
<thead>
<tr>
<th>X-Y</th>
<th>wedge strand</th>
<th>(T_m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T-C</td>
<td>W(_1)W(_1)W(_1)W(_1)W(_1)W(_1)W(_1)K</td>
<td>45.8, 69.2 (±0.5) °C</td>
</tr>
<tr>
<td>T-A</td>
<td>W(_1)W(_1)W(_1)W(_1)W(_1)W(_1)W(_1)K</td>
<td>44.9, 69.1 (±0.5) °C</td>
</tr>
<tr>
<td>T-C</td>
<td>W(_1)W(_1)W(_1)W(_1)W(_1)W(_1)W(_1)K</td>
<td>36.4, 70.1 (±0.3) °C</td>
</tr>
<tr>
<td>G-C</td>
<td>W(_1)W(_1)W(_1)W(_1)W(_1)W(_1)W(_1)K</td>
<td>46.9, 69.8 (±0.3) °C</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>X XXXXXXXX</th>
<th>(T_m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-C C C C C C C -</td>
<td>46.9, 69.8 (±0.3) °C</td>
</tr>
<tr>
<td>-C A C C C C C -</td>
<td>44.3, 69.1 (±0.4) °C</td>
</tr>
<tr>
<td>-C A C C C A C C -</td>
<td>34.2, 67.6 (±0.7) °C</td>
</tr>
<tr>
<td>-C A C C C A C A -</td>
<td>19.9, 67.2 (±0.7) °C</td>
</tr>
<tr>
<td>-A A A C A A A -</td>
<td>78.7</td>
</tr>
</tbody>
</table>
The J-W recognition mode proposed by Lehn and the J-W triplex strategy later developed by McLaughlin and co-workers were fundamentally novel directed towards the sequence-specific recognition of DNA and the accompanying triplex assembly. In contrast to most of the traditional DNA recognition motifs, the J-W are designed to bind by inserting between two nucleobases rather than targeting hydrogen-bonding sides of the intact base pairs presented in either minor or major grooves. It elegantly solved the long-held challenge presented by conventional TFMs with homopurines as the sole targets. The promising results from their study may also shed a light on the future design of recognition elements capable of targeting all four base pairs with enhanced affinity and selectivity. However, the studied DNA target sequences were all preformed duplexes. It was highly possible that the stability of formed J-W triplex was dependent on Watson-Crick base-pair brackets at both ends. Therefore, it will be of significant importance to understand in greater detail the stabilization energy contributed solely from Watson-Crick hydrogen bonding at the recognition interfaces as well as to explore J-W recognition motif on a more synthetically convenient scaffold.
Chapter 2

Experimental: Design, Synthesis and Characterization
2.1. DNA Triplex-forming Peptide Design

Nucleic acid triplex structures formed from native oligonucleotides are known to occur via purine Hoogsteen base-pairing of a third strand in the major groove of a Watson-Crick base-paired duplex.\textsuperscript{16,43,44} This has been extensively developed by Dervan\textsuperscript{45} and others\textsuperscript{46} as a targeting concept using both native and artificial nucleobase recognition elements and backbones. McLaughlin\textsuperscript{42} and Tor\textsuperscript{47} have both reported elegant “Janus-wedge”\textsuperscript{40} recognition of Watson-Crick interfaces using synthetic nucleobases on PNA and sugar-phosphate backbones, respectively. While these prior methods have largely sought to develop general strategies for sequence targeting of preformed oligonucleotide structures, there are fewer synthetic approaches to generate structure in single-stranded oligos or effect triplex formation without prior oligonucleotide duplex assembly; such structure-inducing recognition may be useful in design of synthetic regulators of transcription\textsuperscript{48,49} or translation.\textsuperscript{50-53} We have explored this notion using a Janus-wedge approach to address two identical interfaces: two oligothymidine (dT\textsubscript{10}) DNA tracts were assembled on a peptide template via bifacial melamine recognition to form peptide-DNA triplex structures. Compounds closely related to melamine and its canonical hydrogen-bonding partner, cyanuric acid,\textsuperscript{54-56} have been used to site-substitute for native nucleobases in PNA-DNA duplex recognition. Melamine itself is a well-known molecular recognition module in a number of contexts,\textsuperscript{57-66} and Baranger and Zimmerman\textsuperscript{67} have recently reported melamine targeting of thymine-thymine or uracil-uracil (T-T, U-U) mismatch sites in d(CTG) and r(CUG) repeats, assisted by acridine
intercalation. Indeed, the hydrogen-bonding pattern of melamine precisely complements the Watson-Crick face of thymine/uracil (Figure 2.1).

**Figure 2.1.** Bifacial melamine-thymine recognition. Synthetic peptides 1-4 present melamine and methylated melamine on derivatized lysine side chains. Peptide 2 is N-terminated with (5,6)-carboxyfluorescein (Cbf) and β-alanine (β-Ala). Putative peptide-DNA complex structures are shown: triplex 5 and hairpin 6.

We recently characterized the recognition of cyanuric acid and melamine derivatives at aqueous interfaces as well as bulk solution and found that binding was strongly dependent on the number of heterocycles per scaffold: a trivalent system yielded robust binding while monovalent recognition was undetectable. Based on this prior work, we hypothesized that multivalent presentation of melamine heterocycles on a peptide backbone would bind two strands of dT_{10} into a peptide-DNA heterotrimeric bundle. Indeed, though dT_{10} has no detectable homo-oligomerization behavior, synthetic
melamine-displaying peptide 1 induced assembly of a peptide-DNA triplex structure in heterotrimeric stem and heterodimeric hairpin systems with a peptide-DNA ratio of 1:2 and 1:1, respectively (Figure 2.1).

Prior studies from Eschenmoser and Krishnamurthy and Ghadiri demonstrated the use of α-peptide backbones, instead of the traditional PNA backbone, to recognize DNA. This method forms a recognition interface from alternate residues and was more synthetically convenient for our purposes. We introduced melamine through side-chain functionalization of Boc-Lysine with chlorodiaminotriazine, to yield a derivative we term melaminolysine (M*). Boc deprotection with TFA followed by reaction with Fmoc-OSu yielded Fmoc-M*, which was used in standard solid-phase peptide synthesis (SPPS) with a C-terminal glycinamide (Figure 2.1). Alternate residues were glutamic acid to provide water solubility and to avoid nonspecific electrostatic binding to DNA, yielding peptide 1, (EM*)_{10}G. This sequence was N-terminally capped with carboxyfluorescein (peptide 2) to permit fluorescence-based binding analysis. Control peptides were also synthesized in which hydrogen-bonding sites were systematically blocked by methylation of the exocyclic amines on the melamine ring. Stepwise chloride displacement of trichlorotriazine with ammonia and/or dimethylamine and Boc-lysine yielded the dimethylated and tetramethylated melaminolysine derivatives, which were used in SPPS to provide peptides 3 and 4, respectively.
2.2. Materials and Methods

2.2.1. General

Chemicals for amino acid derivatives synthesis, peptide synthesis, purification and characterization were purchased from Sigma-Aldrich and used without further purification unless otherwise specified. Rink Resin LS (100-200 mesh) was purchased from Advanced ChemTech. DNA Oligomers were obtained from Sigma-Aldrich except 5’-Cy5-dT10 from Integrated DNA Technologies, Inc. Dye-labeled DNA oligomers were HPLC purified and the unlabeled ones were purified by standard desalting. All DNA oligomers were quantified at 260 nm on UV-Vis HP 8543 using the extinction coefficients provided by the manufacturers. 40% acrylamide and bis-acrylamide solution (29:1) was purchased from Bio-Rad. TBE Buffer (10X solution) for gel electrophoresis was purchased from American Bioanalytical. Dulbecco’s Phosphate-Buffered Salines (DPBS 1X, 2.67 mM KCl, 137.93 mM NaCl, 1.47 mM KH2PO4 and 8.06 mM Na2HPO4, pH 7.4, no calcium, no magnesium) used as the binding buffer for all the binding studies discussed in this chapter was purchased from Invitrogen. MALDI-Mass spectra were acquired on Bruker Microflex MALDI-TOF instrument under RP mode. Electrospray mass spectroscopy was acquired on a Bruker MicroTOF equipped with an electrospray ionization source under positive mode. Mass spectrometry instruments were provided by a grant from the Ohio BioProducts Innovation Center. NMR spectra were acquired on a Bruker Advance DPX 400 instrument.
2.2.2. Synthetic schemes for monomer preparation

**Figure 2.2.** Synthesis of monomer 1.

**Figure 2.3.** Synthesis of monomer 3.
2.2.3. Synthetic procedures

(S)-2-((tert-butoxycarbonyl)amino)-6-((4,6-diamino-1,3,5-triazin-2-yl)amino)hexanoic acid (1a).\textsuperscript{82}

Boc-Lys-OH (7.389 g, 30 mmol) was dissolved in water (80 ml) and was added to a water suspension (20 ml) of 6-chloro-1,3,5-triazine-2,4-diamine (5.24 g, 36 mmol). NaOH (2.4 g, 60 mmol) in 60 ml water was slowly added to the mixture and the reaction was stirred overnight at 85 °C. The reaction mixture was cooled down to R.T. and white solid was filtered off. The aqueous solution was acidified to pH 5 with HCl (1 M solution) at 0 °C. The resulting precipitate was filtered and washed with water (2x15 ml) and dried under high vacuum to give 10.1 g (yield 95%) of 1a as a white solid. \textsuperscript{1}H NMR
(400 MHz, d$_6$-DMSO) $\delta$ (ppm) 1.21-1.70 (m, 15H), 3.15 (m, 2H), 3.83 (m, 1H), 6.14 (br s, 2H), 6.30 (br s, 2H), 6.62 (t, $J$ = 5.6 Hz, 1H), 6.91 (d, $J$ = 7.9 Hz, 1H). $^{13}$C NMR (100 MHz, d$_6$-DMSO) $\delta$ (ppm) 23.05, 28.25, 28.95, 30.79, 53.74, 77.92, 155.57, 165.71, 166.52, 174.78. HRMS (ESI) m/z calculated for [M+H]$^+$: 356.2041, found 356.2038.

(S)-2-(((9H-fluoren-9-yl)methoxy)carbonyl)amino)-6-((4,6-diamino-1,3,5-triazin-2-yl)amino)hexanoic acid (1).\textsuperscript{54}

Compound 1a (10 g, 28 mmol) is dissolved in TFA (80 ml) and water (1 ml) was added. The mixture was stirred at R.T. until complete by ESI and TLC (~1h). Maximum amount of TFA was removed under reduced pressure. The residue was resuspended in 200 ml water and the solution was neutralized to pH 7 using NaHCO$_3$. Another 1.5 equiv. of NaHCO$_3$ was added and the resulting solution was stirred at 0 °C for 20 min. Fmoc-OSu (14 g, 42 mmol) in 200 ml dioxane (also cooled) was added slowly. The resulting mixture was stirred at 0 °C for 1 h and allowed to warm up to R.T. overnight. Water (150 ml) was then added and dioxane was removed under reduced pressure. The aqueous layer was washed with EtOAc (2x100 ml), acidified to pH 1 with HCl (1 M solution) at 0 °C and extracted with EtOAc (3x120 ml). The combined organic layer was concentrated under vacuum. The resulting residue was purified by flash chromatography (SiO$_2$) with a solvent gradient from 5 to 10% MeOH in CH$_2$Cl$_2$ to give compound 1 as a white solid (8 g, yield 60%). $^1$H NMR (400 MHz, d$_6$-DMSO) $\delta$ (ppm) 1.34 (m, 2H), 1.50 (m, 2H), 1.58-1.76 (m, 2H), 3.24 (dt appearing as q, $J_1$ = 6.7 Hz, $J_2$ = 6.4 Hz, 2H), 3.93 (m, 1H), 4.22 (m, 1H), 4.27-4.30 (m, 2H), 7.31 (dd appearing as t, $J_1$ = 7.4 Hz, $J_2$ = 7.4 Hz, 2H), 7.40 (dd appearing as t, $J_1$ = 7.4 Hz, $J_2$ = 7.4 Hz, 2H), 7.61 (d, $J$ = 8 Hz, 1H), 7.70 (s, 2H).
2H), 7.71 (d, J = 7.6 Hz, 2H), 7.72 (s, 2H), 7.88 (d, J = 7.5 Hz, 2H), 8.09 (t, J = 5.5 Hz, 1H). $^{13}$C NMR (100 MHz, d$_6$-DMSO) δ (ppm) 22.98, 28.27, 30.45, 46.71, 53.84, 65.65, 120.14, 125.32, 127.12, 127.68, 140.75, 143.85, 156.23, 173.93. HRMS (ESI) m/z calculated for [M+H]$^+$: 478.2197, found 478.2192.

2-amino-4,6-dichloro-[1,3,5]-triazine (3a).

Cyanuric chloride (4.98 g, 27 mmol) was dissolved in acetone (38 ml) and poured into 40 ml of iced-water to form a very fine suspension. Ammonium hydroxide solution (6.76 g 28%, 54 mmol) was added dropwise at 0 °C. The mixture was stirred at 0 °C for 30 min and R.T. for 30 min. White solid was filtered, washed with water (4x30 ml) and dried under high vacuum giving pure compound 3a (3 g, yield 70%). $^1$H NMR (400 MHz, d$_6$-DMSO) δ (ppm) 8.54 (bs, 2H). $^{13}$C NMR (100 MHz, d$_6$-DMSO) δ (ppm) 166.96, 169.20. HRMS (ESI) m/z calculated for [M+H]$^+$: 164.9729, found 164.9722.

$^{N_2,N_2}$-dimethyl-2,4-diamine-6-chloro-[1,3,5]-triazine (3b).

Compound 3a (2.47 g, 15 mmol) was dissolved in acetone (33 ml) and poured into 42 ml iced-water to form a very fine suspension. A water solution of dimethylamine (1.69 g 40%, 15 mmol) was added at 0 °C. NaOH solution (15 ml, 1N) was added dropwise at 0 °C. The reaction mixture was stirred at R.T. for 24 h. The white solid was filtered, washed with water (4x15 ml) and dried under high vacuum to afford pure (2.2 g, yield 85%) $^1$H NMR (400 MHz, d$_6$-DMSO) δ (ppm) 3.03 (s, 6H), 7.24 (br s, 2H). $^{13}$C NMR (100 MHz, d$_6$-DMSO) δ (ppm) 35.90, 35.96, 164.96, 166.55, 168.37. HRMS (ESI) m/z calculated for [M+H]$^+$: 174.0541, found 174.0545.
(S)-6-((4-amino-6-(dimethylamino)-1,3,5-triazin-2-yl)amino)-2-((tert-butoxycarbonyl)amino)hexanoic acid (3c).

Compound 3c was synthesized from compound 3b and Boc-Lys-OH following the procedure used for synthesis of compound 1a. $^1$H NMR (400 MHz, d$_6$-DMSO) δ (ppm) 1.28-1.37 (m, 11H), 1.45 (br s, 2H), 1.52-1.69 (m, 2H), 2.99 (s, 3H), 3.01 (s, 3H), 3.19 (br s, 2H), 3.83 (m, 1H), 6.27 (br t, 1H), 6.65 (br s 1H), 6.78 (br s, 1H), 7.01 (d, $J$ = 7.8 Hz, 1H). $^{13}$C NMR (100 MHz, d$_6$-DMSO) δ (ppm) 23.09, 28.26, 28.78, 30.54, 35.74, 53.52, 78.05, 155.68, 174.36. HRMS (ESI) m/z calculated for [M+H]$^+$: 384.2354, found 384.2348.

(S)-2-(((9H-fluoren-9-yl)methoxy)carbonyl)amino)-6-((4-amino-6-(dimethylamino)-1,3,5-triazin-2-yl)amino)hexanoic acid (3).

Compound 3 was synthesized from compound 3c following the procedure used for synthesis of compound 1. Column chromatography (SiO$_2$) purification with a solvent gradient from 2 to 5% MeOH in CH$_2$Cl$_2$ provided compound 3 as a white solid (yield 53%). $^1$H NMR (400 MHz, d$_6$-DMSO) δ (ppm) 1.35 (m, 2H), 1.47 (br s, 2H), 1.57-1.76 (m, 2H), 2.98 (s, 6H), 3.19 (br s, 2H), 3.92 (m, 1H), 4.20-4.29 (m, 3H), 6.01 (br s, 1H), 6.12 (br s 1H); 6.50 (t, $J$ = 5.5 Hz, 1H), 7.32 (dd appearing as t, $J_1$ = 7.4 Hz, $J_2$ = 7.4 Hz, 2H), 7.41 (dd appearing as t, $J_1$ = 7.4 Hz, $J_2$ = 7.4 Hz, 2H), 7.61 (d, $J$ = 8 Hz, 1H), 7.72 (d, $J$ = 7.4 Hz, 2H), 7.89 (d, $J$ = 7.5 Hz, 2H). $^{13}$C NMR (100 MHz, d$_6$-DMSO) δ (ppm) 23.09, 28.86, 30.52, 35.52, 46.65, 53.83, 65.58, 120.08, 125.27, 127.05, 127.62, 140.69, 143.79, 143.84, 156.15, 165.60, 174.02. HRMS (ESI) m/z calculated for [M+H]$^+$: 506.2510, found 506.2518.
$N^2,N^2,N^4,N^4$-tetramethyl-2,4-diamine-6-chloro-1,3,5-triazine (4a). 82

Cyanuric chloride (4.98 g, 27 mmol) was dissolved in 35 ml acetone and poured into 50 ml ice-water to form a very fine suspension. A water solution of dimethylamine (6.08 g 40%, 54 mmol) was added at 0 °C. NaOH (54 ml, 1N) was added dropwise at 0 °C. The mixture was stirred 30 min at R.T. and additional 1 h at 50 °C. The solution was cooled down in ice-water and precipitate was filtered, washed with water (4x30 ml) and dried under vacuum to afford white solid (4.88 g, 90%). $^1$H NMR (400 MHz, d$_6$-DMSO) δ (ppm) 3.04 (s, 6H), 3.06 (s, 6H). $^{13}$C NMR (100 MHz, d$_6$-DMSO) δ (ppm) 35.72, 35.81, 164.24, 168.04. HRMS (ESI) m/z calculated for [M+H]$^+$: 202.0854, found 202.0857.

(S)-6-(((4,6-bis(dimethylamino)-1,3,5-triazin-2-yl)amino)-2-((tert-butoxycarbonyl)amino)hexanoic acid (4b).

Compound 4b was synthesized from compound 4a and Boc-Lys-OH following the procedure used for synthesis of compound 1a. $^1$H NMR (400 MHz, d$_6$-DMSO) δ (ppm) 1.37-1.64 (m, 15H), 3.00 (s, 12H), 3.19 (dt appearing as q, $J_1 = 6.3$ Hz, $J_2 = 6.3$ Hz, 2H), 3.83 (m, 1H), 6.53 (t, $J = 5.7$ Hz, 1H), 6.99 (d, $J = 7.9$ Hz, 1H). $^{13}$C NMR (100 MHz, d$_6$-DMSO) δ (ppm) 23.05, 28.19, 28.90, 30.65, 35.44, 53.58, 77.84, 155.53, 165.23, 165.48, 174.28. HRMS (ESI) m/z calculated for [M+H]$^+$: 412.2667, found 412.2658.

(S)-2-(((9H-fluoren-9-yl)methoxy)carbonyl)amino)-6-((4,6-bis(dimethylamino)-1,3,5-triazin-2-yl)amino)hexanoic acid (4).

Compound 4 was synthesized from compound 4b following the procedure used for synthesis of compound 1. Column chromatography (SiO$_2$) purification with a solvent gradient from 1 to 3% MeOH in CH$_2$Cl$_2$ provided compound 4 as a white solid (yield
\( \text{H NMR} \) (400 MHz, d\textsubscript{6}-DMSO) \( \delta \) (ppm) 1.37 (m, 2H), 1.53 (m, 2H), 1.57-1.79 (m, 2H), 3.02 (s, 6H), 3.05 (s, 6H), 3.28 (m, 2H), 3.93 (m, 1H), 4.18-4.31 (m, 3H), 7.32 (dd appearing as t, \( J_1 = 7.4 \) Hz, \( J_2 = 7.4 \) Hz, 2H), 7.41 (dd appearing as t, \( J_1 = 7.4 \) Hz, \( J_2 = 7.4 \) Hz, 2H), 7.62 (d, \( J = 8 \) Hz, 1H), 7.71 (m, 3H), 7.89 (d, \( J = 7.6 \) Hz, 2H). \( ^{13}\text{C NMR} \) (100 MHz, d\textsubscript{6}-DMSO) \( \delta \) (ppm) 22.92, 28.00, 30.34, 36.07, 46.63, 53.61, 65.57, 120.07, 125.21, 126.99, 127.59, 140.69, 143.77, 156.10, 173.88. HRMS (ESI) m/z calculated for [M+H]\(^+\): 534.2823, found 534.2823.

2.2.4. Solid phase peptide synthesis and characterization

Peptide synthesis was performed on an AAPPTEC Apex 396 Peptide Synthesizer using Rink Resin LS (100-200 mesh, loading 0.2 mmol/g) employing standard Fmoc chemistry utilizing DIC/HOBt as the coupling agents and 50% piperidine/NMP for Fmoc deprotection. Peptides were cleaved from the solid support and tert-Butyl protective groups removed using a TFA : water : m-cresol = 95:2.5:2.5 mixture (3mL for 0.15 g resin) for 3.5 h at R.T. to yield peptides with amide at C-termini. Resin was removed by filtration through cotton and the peptides were precipitated by chilled ether (30 ml), centrifuged, and washed with chilled ether (3x20 ml). Crude peptides were then dissolved in solvent A and purified by HPLC on a C18 reversed phase column using a gradient of 10–100% solvent B in 50 min (solvent A = 0.1% TFA in water, solvent B = 0.01% TFA in 45% acetonitrile, 45% isopropanol, 10% water). The UV detector was set at 230 nm. The purified peptides were lyophilized to dryness. The identity of peptide was checked by MALDI-TOF and purity checked by analytical HPLC on a C18 column.
Figure 2.5. (A) MALDI-TOF of peptide 1: m/z calculated for [M+H]$_{\text{Avg}}^+$ 3738.839, found 3738.181; (B) HPLC trace of peptide 1 on a C-18 analytical column using a gradient of 10-100% solvent B over 50 min, monitored by a UV-Vis detector at 230 nm. (The small peak at 40 min was due to the aggregation of peptide 1 on column. The fraction has the same mass as peptide 1. It was also observed on 33 min peak reinjection HPLC.)
Figure 2.6. (A) MALDI-TOF of peptide 2: m/z calculated for [M+H]^+_{Avg}: 4168.217, found 4168.641; (B) HPLC trace of peptide 2 on a C-18 analytical column using a gradient of 10-100% solvent B over 50 min, monitored by a UV-Vis detector at 230 nm.
Figure 2.7. (A) MALDI-TOF of peptide 3: m/z calculated for [M+H]$^+$ Avg: 4019.370, found 4019.118; (B) HPLC trace of peptide 3 on a C-18 analytical column using a gradient of 10-100% solvent B over 50 min, monitored by a UV-Vis detector at 230 nm.
Figure 2.8. (A) MALDI-TOF of 5(6)-Carboxyfluorescein (Cbf) labeled peptide 3: m/z calculated for [M+H]$^+$$_{\text{Avg}}$: 4448.748, found 4448.875; (B) HPLC trace of of Cbf labeled peptide 3 on a C-18 analytical column using a gradient of 10-100% solvent B over 50 min, monitored by a UV-Vis detector at 230 nm.
Figure 2.9. (A) MALDI-TOF of peptide 4: m/z calculated for [M+H]^{2+}_{\text{avg}}: 2149.45, found 2148.64; (B) HPLC trace of peptide 4 on a C-18 analytical column using a gradient of 10-100% solvent B over 45 min, monitored by a UV-Vis detector at 230 nm.
2.2.5. General sample preparation protocol for binding studies

Samples were prepared in DPBS 1X buffer (2.67 mM KCl, 137.93 mM NaCl, 1.47 mM KH₂PO₄ and 8.06 mM Na₂HPO₄, pH 7.4) according to the required DNA/peptide ratios and concentrations and then were heated at 94°C for 8 min and allowed to re-anneal by unplugging the heating block (slowly cooling to ambient temperature). All samples for Peptide-DNA binding studies described in this chapter were prepared following this procedure if not stated otherwise.

2.2.6. UV Job Plot Experiments

UV absorbance was monitored at 260 nm on UV-Vis HP 8543 with an external water bath at 25 °C. Samples of varying molar ratios of peptide 1: DNA (dT₁₀ or dT₁₀C₁₀T₁₀) were prepared by serial dilution while DNA concentration remained constant (7.5 µM for dT₁₀, 3.5 µM for dT₁₀C₁₀T₁₀). All samples were annealed and preequilibrated in buffer as previously described. Plots of A₂₆₀ vs molar ratio were produced to determine the binding stoichiometry of 1:2 or 1:1 respectively.

2.2.7. Fluorescence Job Plot (Quenching) Experiments

Fluorescence intensities were obtained on Perkin Elmer Luminescence Spectrometer LS 50B equipped with PTP 1 Temperature Programmer and water bath at 25 °C. Samples of varying molar ratios of peptide 1: Flc-5’-dT₁₀C₁₀T₁₀-3’-Dabcyl were prepared by serial dilution while Flc-5’-dT₁₀C₁₀T₁₀-3’-Dabcyl remained constant at 500 nM. All samples were annealed and preequilibrated in buffer as previously described and were excited at 493 nm (excitation slit of 10 nm) and emission were recorded between 500-600 nm.
(emisson slit of 10 nm) monitoring the maximum at 521 nm. The fluorescence intensity vs molar ratio of peptide 1 was plotted to determine the binding stoichiometry of 1:1 between peptide 1 and Flc-5'-dT_{10}C_{10}T_{10}-3’-Dabcyl.

2.2.8. Circular Dichroism (CD) Experiments

CD spectra were performed on Jasco J815 Circular Dichroism Spectrometer equipped with Peltier device and water circulator. All measurements were taken at 25 °C in a Hellma quartz cell (1 mm path length) from 350-210 nm at a scanning rate 50 nm/min, data interval 0.5 nm, band width 1 nm and D.I.T 2 s. For each sample three scans were collected, averaged and corrected for blanks. Total dN_{10} (T_{10}, A_{10}, C_{10}) tracts concentrations were remained at 10 µM in each binding reaction. Samples for CD titration of dT_{10} with peptide 1 ranging from 0-20 µM were prepared by serial dilution. Samples for CD binding study at a single DNA-Peptide ratio, 5 µM peptide 1 was used. All samples were annealed and preequilibrated in buffer as previously described.

2.2.9. Thermal Denaturation Studies by UV

Thermal denaturation studies of triplex 5 and hairpin 6 were performed on Cary 100 Bio UV-Visible Spectrophotometer (Varian) equipped with Cary Temperature Controller. Samples were prepared in the binding buffer according to the required ratios and concentrations. The single stands of DNA were mixed with peptide 1 to give final concentrations of 2 µM dN_{10} (dT_{10} or dA_{10} or dC_{10}) and 1 µM peptide 1 for triplex study; 2.5 µM dT_{10}C_{10}T_{10} and 2.5 µM peptide 1 for hairpin study; 1 µM, 2.5 µM and 3.75 µM peptide 1 for self-structure study. All samples were annealed following general annealing
procedure. The temperature dependent absorbance of each sample was monitored from 10 to 90 °C at 260 nm with a ramp rate of 1 °C per minute and readings were taken every 0.5 °C increments. Tm values were determined from the first-order derivatives.

2.2.10. Thermal Denaturation Studies by Fluorescence

Fluorescence melting experiments for hairpin 6 were performed on Cary Eclipse Fluorescence Spectrophotometer (Varian) equipped with Cary Temperature Controller. The dual labeled dT10C10T10 contains fluorescein (Flc) as the fluorophore at 5’-end and Dabcyl as the quencher at 3’-end. Freshly annealed solutions containing 250 nM Flc-5’-dT10C10T10-3’-Dabcyl in the presence and absence of 250 nM peptide 1 were excited at 493 nm (excitation slit of 5 nm) and emission were recorded at 521 nm (emission slit of 5 nm) from 10 °C to 90 °C (upscan). The samples were then allowed to thermostat at 90 °C for 5 min followed by downscan from 90 °C to 10 °C. A 1 °C/min ramp rate with 0.5 °C data interval was applied to both upscans and downscans. Tm values were determined from the first-order derivatives.

2.2.11. Differential Scanning Calorimetry (DSC)

DSC was performed on Microcalorimeter VP-DSC. Peptide 1 in each sample was at 25 µM while dT10 and dT10C10T10 were 50 µM and 25 µM respectively. All samples were annealed and preequilibrated as previously described and were thoroughly degassed by ThermoVac. A 60 °C/h scan rate, 16 s filtering period and low feedback were applied to both upscans (10 to 90 °C) and downscans (90 to 10 °C). Samples were allowed to thermostate at 90 °C for 5 min before downscan started in each thermal cycle.
Background data was recorded by using the binding buffer in the sample cell and was subtracted from other sample data. The final values of $C_p$ were shown as the average of six thermal cycles for each sample and were fitted to a Non-2-State binding model.

Figure 2.10. DSC traces of triplex 5 (black) with Non-2-State binding model fitted curves shown in red. (A) Upscan and (B) Downscan.
Figure 2.11. DSC traces of Hairpin 6 (black) with Non-2-State binding model fitted curves shown in red. (A) Upscan and (B) Downscan.

2.2.12. Fluorescence and Fluorescence Anisotropy

Titration experiments were performed on Molecular Devices SpectraMax M5. Peptide 2 was used as the probe at a fixed amount 25 nM for Equilibrium Binding Assays and 500 nM (>>Kd) for Stoichiometric Binding Assays respectively. DNA oligomers as ligands were serially diluted in binding buffer to desired concentrations and then an equal amount from each stock was added into peptide 2 to give the final DNA-peptide 2 ratio for each binding reaction. All samples were annealed and preequilibrated as previously described before applying to 384 well, low volume, non-binding surface, round bottom, black polystyrene Corning Assay Plate. Fluorescence and fluorescence anisotropy were recorded at 525 nm upon excitation at 485 nm at 25 °C. All assays were performed in
triplicate and the values shown here were the averages from two separate experiments. All experimental data were fitted by nonlinear regression analysis using Levenberg-Marquardt algorithm in KaleidaGraph.

Figure 2.12. Representative fitting data set for triplex 5: (A) Fluorescence intensity change as a consequence of association reaction was monitored to obtain quenching factor $Q_b/Q_f = 0.26$ (where $Q_b$ and $Q_f$ is the quantum yield of bound and free fluorophore respectively). $Q_b/Q_f$ from this set of data was applied to the calculation of $f_b$ (fraction bound) in analyzing anisotropy values. (B) the Hill coefficient was added as a fitting variable $f_b = [dT_{10}]^n/(K_d+[dT_{10}]^n)$, yielding a best fit with $n=2$. 
Figure 2.13. Representative fitting data set for hairpin 6. (A) Fluorescence intensity change as a consequence of association reaction was monitored to obtain quenching factor Q_b/Q_f = 0.3. (B) Bound peptide 2 concentration calculated from anisotropy values (Q_b/Q_f = 0.3 applied) vs total dT_{10}C_{10}T_{10} concentration was fitted to 1:1 binding model:\[ [\text{Bound peptide 2}] = ([R_T] + [DNA_T] + K_d) - \frac{4 * [R_T] * [DNA_T])}{(4 * [R_T] * [DNA_T] + K_d)^2})/2\], where [R_T] is the total concentration of peptide 2 (25 nM), [DNA_T] is the total concentration of dT_{10}C_{10}T_{10} used in each binding reaction. R^2 > 0.98.

2.2.13. Gel Mobility Shift Assay (carried out by Yaowalak Pratumyot)

Triplex 5 experiment sample preparation: 20 nM Cy5-dT_{10} and 80 nM dT_{10} were incubated with increasing peptide 1 concentration in 1X DPBS (Invitrogen) with a final loading volume of 10 µL. Samples were incubated at 94°C for 10 min and then the mixture was allowed to cool to R.T. over 2 hr. For hairpin 6 experiments: As above, with 20 nM Cy5-dT_{10}C_{10}T_{10} samples used in each well instead. After incubation, 2 µL of loading buffer (80 % Glycerol) was added to all DNA-peptide 1 mixtures and 5 µL of
sample was loaded into the well of 22% and 10% native polyacrylamide gel for dT_{10} and dT_{10}C_{10}T_{10} respectively. The samples were subsequently electrophoresed on ice in an electric field of 100 V using 1X TBE as a running buffer for 7 hours for dT_{10} and 2 hours for dT_{10}C_{10}T_{10}. After electrophoresis, Cy5 fluorescence gel images were acquired using a Typhoon Trio Variable Mode Imager (Amersham Biosciences).
Chapter 3

Results and Discussion
3.1. Binding Stoichiometry

Nanoparticle assembly from melamine and cyanuric acid derivatives results from noncovalent polymerization of the two-fold symmetric heterocycle recognition faces; thymine has only one recognition face complementary to melamine, and thus discrete assembly was anticipated. Gratifyingly, no large peptide-DNA aggregates were detectable by dynamic light scattering, consistent with the model of discrete triplex formation. Peptide-triggered base-stacking signatures were observed by UV absorbance changes, with a 1:2 stoichiometry between 1 and dT₁₀, consistent with bivalent melamine-thymine recognition and formation of triplex 5 (Figure 2.1 and 3.1 A); peptide 1 alone did not induce such a signal (Figure 3.2).

![Figure 3.1](image)

**Figure 3.1.** Peptide 1 titrated into (A) dT₁₀ and (B) Flc-dT₁₀C₁₀T₁₀-Dabcyl, followed by UV absorbance (260 nm) and fluorescein emission (521 nm), respectively. Saturation is observed at 33 and 50 mol% peptide, indicating 1:2 and 1:1 peptide: DNA binding stoichiometries in (A) and (B), respectively.
Unlike PNA-DNA\textsuperscript{42} triplex structures which optimally form at high salt with divalent metal ions (1-2 M NaCl, 50 mM MgCl\textsubscript{2}), robust assembly was observed under standard salt conditions (Dulbecco’s phosphate buffered saline, DPBS), akin to the conditions used by Eschenmoser and Krishnamurthy\textsuperscript{56} for PNA-DNA duplex formation. Similarly, UV absorbance signatures indicated a 1:1 binding stoichiometry between 1 and dT\textsubscript{10}C\textsubscript{10}T\textsubscript{10} (Figure 3.4), as would be expected if an intramolecular peptide-DNA triplex structure formed from the dT\textsubscript{10} termini of dT\textsubscript{10}C\textsubscript{10}T\textsubscript{10}. Indeed, binding of 1 to Flc-dT\textsubscript{10}C\textsubscript{10}T\textsubscript{10}-Dabcyl resulted in maximal fluorescein quenching at a 1:1 peptide-DNA ratio, supportive of the formation of heterodimeric hairpin structure 6, which would bring the 3’ and the 5’ end of the oligo in close proximity,\textsuperscript{56} resulting in efficient dabcyl quenching of fluorescein (Figure 3.1 B).
3.2. CD and Gel Mobility Shift

Triplex and hairpin formation was further corroborated by circular dichroism (CD), which indicated significant structuring of DNA upon addition of peptide, signified by the development of a negative CD signal at 260 nm at the expense of a positive CD signal at 280 nm, which we assign to the peptide complex and free DNA, respectively (Figure 3.4). Notably, while the peak at 280 nm is completely ablated in the triplex, there is a residual peak in the hairpin; this is consistent with the presence of an unstructured C\textsubscript{10} loop found in hairpin 6 but not triplex 5. Clean transformation of DNA to peptide-DNA complex bands on native polyacrylamide electrophoresis indicated discrete peptide-DNA recognition in both triplex and hairpin contexts (Figure 3.4).
Figure 3.4. Circular dichroism spectra in DPBS, pH 7.4, of (A) Peptide 1 complexed with $dT_{10}$ (•) vs $dT_{10}$ alone (--) and (B) Peptide 1 complexed with $dT_{10}C_{10}T_{10}$ (○) vs $dT_{10}C_{10}T_{10}$ alone(---). Peptide 1 in both is at 5μM concentration (--), while $dT_{10}$ and $dT_{10}C_{10}T_{10}$ are maintained at 10 and 5 μM, respectively. Electrophoretic mobility shift assays imaged by Cy5 fluorescence for (C) Cy5-$dT_{10}$ (DNA$_1$) and (D) Cy5-$dT_{10}C_{10}T_{10}$ (DNA$_2$) at 20 nM in each lane, with increasing peptide 1 concentration from left to right. (E) Relative electrophoretic mobilities of the free DNA oligos and their peptide complexes, a mixture of complex 5 and DNA$_1$ shown in the central lane.
3.3. Thermal Stability

Cooperative melting was observed for triplex 5 and hairpin 6 by UV, differential scanning calorimetry (DSC) (Figure 3.5) and fluorescence dequenching (Figure 3.6). The hairpin heterodimer structure 6 was more thermally stable (T_m = 54 °C) than the heterotrimeric triplex 5 (T_m = 43 °C). For comparison, a dA_{10}-T_{10} duplex has T_m = 35 °C, and a dA_{10}-(T_{10})_2 triplex has T_m = 17 °C in the presence of 50 mM MgCl_2.\(^{85,86}\) It is calculated that a dA_{10}-T_{10} duplex (2 µM) will have T_m = 22.5 °C under similar salt conditions (150 mM Na\(^+\), K\(^+\)).\(^{87,88}\)

**Figure 3.5.** (A) First-derivative plot of melting transitions of triplex 5 (---) and hairpin 6 (—) followed by UV absorbance (260 nm). Normalized absorbance change is shown inset. (B) DSC upscan traces of triplex 5 (---) and hairpin 6 (—), with downscan traces shown as dashed regular and bold lines. Peptide-DNA ratios used in triplex 5 and hairpin 6 experiments were 1:2 and 1:1, respectively. DSC and UV experiments were performed in DPBS, pH 7.4 at peptide concentrations of 25 µM (DSC) and in UV experiments, 1 µM (5) and 2.5 µM (6).
Figure 3.6. (A) Fluorescence melting curves for hairpin 6 (—) and free Flc-dT_{10}C_{10}T_{10}-Dabeyl (--); (B) Corresponding first derivative curves for hairpin 6. Fluorescence melting experiments were performed in DPBS, pH 7.4 at Flc-dT_{10}C_{10}T_{10}-Dabeyl concentration of 250 nM in the presence or absence of 250 nM peptide 1.

Notably, triplex 5 transition temperature yielded from DSC was slightly elevated (48 °C) relative to the other methods while the hairpin 6 T_m was very similar (55 °C). The difference was likely due to the relatively higher complex concentrations used in DSC experiments which were in the stoichiometric binding regime; consistent with the dissociation constants obtained from fluorescence and fluorescence anisotropy binding isotherms (Figure 3.7).

Reversible peptide-DNA complexation was supported by observation of endothermic melting and exothermic cooling peaks by both fluorescence melt and DSC. Assembly of complexes 5 and 6 was highly exothermic and reproducible over several thermal cycles ($\Delta H_{\text{triplex}} = -285$ kcal/mol peptide, $\Delta H_{\text{hairpin}} = -314$ kcal/mol), consistent with base-
The large enthalpy values (-28.5 to -31.4 kcal/mol per peptide: DNA triplet layer) are somewhat higher than, but similar to, enthalpies observed for (TAT)$_n$ intramolecular DNA triplexes,$^{94}$ which exhibit -21 kcal/mol per triplet stack. Peptide-displayed melamine-DNA binding appears to have exothermic assembly profiles and $T_m$ values similar to those of DNA-DNA and melamine-cyanuric acid recognition, suggestive of similar driving forces, perhaps to be expected given the similarity of melamine and native nucleobases.

3.4. Binding Affinity

We quantified association using fluorescein-tagged peptide 2 binding to unlabeled DNA (Figure 3.7). Complex formation was followed by both fluorescence anisotropy and quenching, which accompanies assembly; similar results were obtained with the two signatures. Using the stoichiometries established by titration (Figure 3.1), we fit the binding curves to a trimer-monomer model to obtain an apparent $K_d \approx 4000$ nM$^2$ (by quenching or anisotropy) for dissociation of triplex 5. A second transition was undetectable by fitting to a two-step model or by observation of assembly processes, suggesting that the peptide-$dT_{10}$ heterodimer that forms initially must react rapidly with another strand of $dT_{10}$ to form the heterotrimer. Indeed, the coefficient derived from a Hill-type binding isotherm was essentially 2, supportive of a cooperative 1:2 peptide-DNA assembly process. A hypothetical equivalent two-step dissociation process with an overall apparent $K_d = 4000$ nM$^2$ would have $K_d = 63$ nM for both trimer-dimer and dimer-monomer dissociation steps, which is the observed free DNA concentration at half-saturation (Figure 3.7). A good fit to a 1:1 binding model$^{83,84}$ was obtained from the
binding isotherm of peptide 2 to dT\textsubscript{10}C\textsubscript{10}, revealing an approximate $K_d = 2.7$ nM for dissociation of hairpin 6.

![Figure 3.7](image)

**Figure 3.7.** Binding isotherms in DPBS, pH 7.4, of (A) peptide 2 binding to dT\textsubscript{10} followed by fluorescein quenching upon binding and (B) peptide 2 binding to dT\textsubscript{10}C\textsubscript{10} followed by fluorescence anisotropy. Solid lines show fits to (A) trimer-monomer 1:2 binding model ($\text{[fraction bound peptide } 2\text{]} = \frac{[\text{DNA}]^2}{(K_d + [\text{DNA}]^2)}$, $R^2 \geq 0.96$) and (B) 1:1 binding model (corrected for fluorescence quenching): $\text{[Bound peptide } 2\text{]} = \frac{([\text{R}_T]+[\text{DNA}_T]+K_d)-\sqrt{([\text{R}_T]+[\text{DNA}_T]+K_d)^2-(4*([\text{R}_T]*[\text{DNA}_T])))}}{2}$, where $[\text{R}_T]$ is the total concentration of peptide 2 (25 nM), $[\text{DNA}_T]$ is the total concentration of dT\textsubscript{10}C\textsubscript{10} used in each binding reaction. $R^2 > 0.98$.

### 3.5. Binding Selectivity

Given that peptides 1 and 2 and DNA are all anionic polyelectrolytes, we anticipate negligible nonspecific electrostatic interactions. Indeed, partial and full methylation of
each melamine ring on the peptide to respectively yield peptides 3 and 4 abolished all detectable binding to dT\textsubscript{10} or dT\textsubscript{10}C\textsubscript{10}T\textsubscript{10} (Figure 3.8). These observations are in agreement with the critical role of the melamine recognition interface. As methylation increases hydrophobicity, it is clear that assembly depends on recognition of donor-acceptor patterning and not simply nonspecific hydrophobic collapse.

Notably, peptides 1 and 2 did not show any binding signatures when annealed with dA\textsubscript{10}, dC\textsubscript{10}, and dG\textsubscript{5}A\textsubscript{10}, indicating a selectivity for thymine over the other native nucleobases (Figure 3.9). Thus, melamine peptide recognition is relatively specific for the hydrogen-bonding pattern of oligothymidine, as predicted. Taken together with the observed stoichiometry of binding with dT\textsubscript{10} and dT\textsubscript{10}C\textsubscript{10}T\textsubscript{10}, this is strongly supportive of the expected bivalent thymine recognition by melamine (Figure 2.1).
Figure 3.8. (A) Absorbance (260 nm) vs Temperature plots for control peptide 3 and peptide 4 with dT_{10}. Concentrations: 2 µM peptide with 2 µM dT_{10}. (B) Electrophoretic mobility shift assay imaged by Cy5 fluorescence for Cy5-dT_{10} at 150 nM with increasing concentrations of peptide 3 (0, 30, 50, 100, 150, 225, 300, 750, 1200, 1500 nM) from left to right. Study for peptide 4 and Cy5-dT_{10} under the same condition yielded the similar result.
Figure 3.9. (A) Absorbance (260 nm) vs Temperature plots for peptide 1 with control oligos dA\textsubscript{10} and dC\textsubscript{10}. Absence of UV shift or melting signatures. Concentrations: 1 µM peptide 1 with 2 µM dA\textsubscript{10} or dC\textsubscript{10}. (B) Fluorescence (○) and Anisotropy (Δ) assays for peptide 2 with control oligos dG\textsubscript{5}A\textsubscript{10} and dC\textsubscript{10}. Peptide 2 concentration is constant at 25 nM. (C) Circular dichroism spectra for peptide 1 with control oligos dA\textsubscript{10} and dC\textsubscript{10}. Concentrations: 5 µM peptide 1 with 10 µM dA\textsubscript{10} or dC\textsubscript{10}. continued
Figure 3.9 continued

C

![CD spectra](image)

**Figure S9.** CD spectra of other DNA oligomers with peptide 1. (a) free dC5T10C6/dG6T10G5 (T10 tracts 10 µM) and their peptide 1 complexes.

**Figure S10.** Control experiment: CD spectra of 10 µM dA10 with 5 µM peptide 1 showing no change on CD with peptide 1.

**Figure S11.** Control experiment: 10 µM dC10 with 5 µM peptide 1.
Chapter 4

Conclusion
Overall, these data support the model of bivalent melamine-thymine recognition yielding formation of novel discrete peptide-DNA triplex structures with robust binding affinities. Notably, prior work from Eschenmoser and Krishnamurthy using similar diaminotriazine nucleobase mimics derived from aspartic and glutamic acids resulted in duplex formation with oligothymidines rather than triplex. Diaminotriazine has two exocyclic hydrogen bond donor sites compared to three on melamine, which yields two potential thymine recognition interfaces in melaminolysine and just one in diaminotriazine derivatives. Thus, it appears that the assembly state pivots from dimer to trimer based on the addition of a single hydrogen bond donor, with the increased entropic cost likely paid for by increased base stacking. Polyvalent melamine-thymine recognition between peptides 1 and 2 with dT₁₀ tracts is unique in that structure is induced from unstructured single-stranded oligothymidines to yield novel triplex and hairpin peptide-DNA hybrid structures, thus expanding the range of non-native nucleobase-derived structures already known. Though melamine recognition of preformed duplex with T-T/U-U mismatch sites has been previously reported, the ability of polyvalent melamine peptides to broker assembly with two non-interacting oligothymidine strands to form peptide-DNA triplex structures is non-obvious.

We anticipate this novel approach can be employed to develop ligands for targeting long stretches of mismatched bases that are present in important structural features such as mutated sites of DNA and unpaired regions of RNA forming bulges and loops. The melamine peptide we reported here in particular present promising artificial regulators to manipulate the structure and function of thymine- and uracil-rich targets at DNA and
RNA level respectively. From a broader perspective, the designed element can serve as a synthetic sequence-specific DNA binding domain (DBD) for artificial transcription factors (ATFs) construction through modular design, which has become an essential strategy toward development of genetic therapies.

High synthetic accessibility and the ease with which chemical functionality can be introduced implied the high potential of our designed motif as a novel information transfer system, such as a template to direct sequence-selective, nonenzymatic synthesis of nucleic acid (or PNA) and peptide. Such investigations may gain insight into DNA replication, transcription and translation in prebiotic chemistry.

As a model study of base pairing properties between polyvalent melamine and thymidine, it will be important to gain more structural information of the peptide-DNA heterotrimeric bundle motif. Of particular significance is to determine the crystal structure. In addition, it also encourages follow-up investigations with respect to the trend of diminishing DNA binding by varying the number of recognition element (M*). Furthermore, to improve upon this lead melamine peptide design, chemical modifications through peptide backbone to tune the binding affinity and specificity as well as to understand in greater detail how such alterations might affect parameters such as cellular uptake, nuclease resistance, tissue distribution and intracellular trafficking would be of great interest before versatile in vivo activities can be fully accessed.
Reference


71


84. Fabrice Agou; Sharona Raveh; Sébastien Mesnildrey; Michel Véron *J. Biol. Chem.* **1999**, *274*, 19630-19638.


Appendix A

Additional Characterization Data
Figure A1. Additional UV Job plot analysis of peptide 1 with dT\textsubscript{10}.

Figure A2. Additional CD spectrum: Titration of 10 \textmu{}M dT\textsubscript{10} with increasing concentration of peptide 1.
Figure A3. Additional CD spectrum: free dC$_5$T$_{10}$C$_6$/dG$_6$T$_{10}$G$_5$ (T$_{10}$ tracts 10 µM) and their peptide 1 complexes.

Figure A4. Additional thermal denaturation curve for peptide 1 alone at different concentrations. Absence of UV melting signature rules out possible secondary structures from peptide itself.
**Figure A5.** Additional stoichiometric binding curves for triplex 5 or hairpin 6 by titrating 500 nM peptide 2 with dT\textsubscript{10} or dT\textsubscript{10}C\textsubscript{10}T\textsubscript{10}. Both fluorescence (—) and anisotropy (○) were monitored for triplex 5. Only anisotropy was recorded for hairpin 6 (Δ).

**Figure A6.** Additional control Fluorescence (○) and Anisotropy (Δ) assays for free Cbf in solution with (A) dT\textsubscript{10} and (B) dT\textsubscript{10}C\textsubscript{10}T\textsubscript{10} to rule out non-specific binding between the fluorophore and DNA oligomers.
Appendix B

$^1$H and $^{13}$C NMR Spectra of Amino Acid Derivatives
Figure B1. $^1$H NMR of 1a.

Figure B2. $^1$H NMR of 1.
Figure B3. $^1$H NMR of 3a.

Figure B4. $^1$H NMR of 3b.
Figure B5. $^1$H NMR of 3c.

Figure B6. $^1$H NMR of 3.
Figure B7. $^1$H NMR of 4a.

Figure B8. $^1$H NMR of 4b.
Figure B9. $^1$H NMR of 4.

Figure B10. $^{13}$C NMR of 1a.
Figure B11. $^{13}$C NMR of 1.

Figure B12. $^{13}$C NMR of 3a.
Figure B13. $^{13}$C NMR of 3b.

Figure B14. $^{13}$C NMR of 3c.
Figure B15. $^{13}$C NMR of 3.

Figure B16. $^{13}$C NMR of 4a.
Figure B17. $^{13}$C NMR of 4b.

Figure B18. $^{13}$C NMR of 4.