

**Programming Protein Patterns on DNA Nanostructures
With Pyrrole-Imidazole Polyamides**

Thesis by

Justin Delgado Cohen

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology

Pasadena, California

2009

(Defended 10 November 2008)

© 2009

Justin Delgado Cohen

All Rights Reserved

For my family

Acknowledgements

I would like to begin by thanking Professor Peter B. Dervan who has given me the opportunity to work in a world class research group here at CalTech. Without Peter's guidance I would not be the scientist that I am today. I would like to thank him for his support, advice, generosity, and understanding during my stay here. I truly appreciate his willingness to support my research as it entered a new and unexplored field. It was truly an honor to work with him. I would also like to thank the members of my committee, Professors Robert Grubbs, Doug Rees, and Erik Winfree for their time and assistance in helping me through candidacy and my research proposals. I would also like to further thank Professor Erik Winfree for all of his assistance and guidance in my research. Without his support it would not have been possible to perform many of the experiments that are described in this thesis.

Next, I would like to thank my family, my mother Alma, my father Joel, and my sister Lea. Without their constant support and understanding, I would not be here today. Thank you for all of the pep talks, the snacks and cards in the mail, and most of all, for your constant belief in me. I would never have gotten half this far without your love and support. You have been my inspiration and motivation to keep pushing myself, and I owe you a debt of gratitude for all of my accomplishments. I love you very much.

I am indebted to John Sadowski, one of the most intelligent undergraduates I have had the pleasure of working with. I truly appreciate all of the hard work he did while working on our DX project. I would also be remiss if I did not thank Dr. Sung Ha Park for his incredible patience in teaching me how to use an AFM and answering all of my questions. His assistance was invaluable.

Next, I would like to thank the members of the Dervan lab. They are an outstanding group of scientists and people. It has been my pleasure to be associated with such a classy, intelligent, and caring group of people. I owe a debt to the many older students who helped to guide me when I first joined the lab; Tim Best, Adam Kerstein, Ben Edelson, Eric Fechter, Michael Marques, and Ray Doss. Adam Kerstein was the first one to teach me to make a polyamide when I joined the group and helped me a great deal in my first year, in addition to providing a place to play cards and tagging along on many trips to the Bike. I would like to thank Eric Fechter for his friendship in and out of the lab. I want to thank Michael Marques for his assistance in lab, and for being a great roommate. Tim Best has been a great person to talk to for some real world after graduation perspective, as well as a great person to take on a ski trip. And last, I would like to thank Ray Doss, for keeping me sane in my first few years and for being a great friend and a great scientist. Thank you for all of the fun times, for convincing me it was a good idea to throw a toga party, and I am looking forward to more adventures in the future.

I also want to thank all of my more recent labmates for their friendship, advice, and assistance. I want to thank Carey Hsu and Jim Puckett, who I have had the good fortune of joining the Dervan group at the same time as. I could not think of two better people to go through this journey with. Michelle Farkas and Katy Muzikar, I don't even know how to begin to thank you. Your help in lab and in life has been invaluable to me. I seriously would not be here today if it was not for the friendship and support you have given me. I am indebted to Ryan Stafford for his knowledge and assistance, for teaching me how to footprint, and for captaining our basketball team. I would also like to thank

Dave Chenoweth, whose help and knowledge in lab has been invaluable and who has always been a great person to talk. I am looking forward to seeing you at MIT. I want to thank Claire Jacobs for many funny moments, fun conversations and for her help in lab. Mareike Goeritz has been a good friend these past 2 years, and I want to thank her for all of the fun times, and for getting me addicted to coffee. I would also like to thank all of the first years who have provided a new sense of excitement this past year. Ben (weaksauce), Fei (cream puffs), and Dave (softball), it has been a pleasure getting to know you and I wish you guys the best. Last, I would also like to thank all of the other members of the Dervan group who I have had the pleasure of working with throughout the years, John Phillips, Jim Sanchez, Anne Viger, Christian Dose, Sherry Tsai, Nick Nichols, Julie Poposki, Dan Harki, and Mike Brochu. Although there are too many good memories to list here, I can truly say it has been an honor to work with each and every one of you.

Finally, I would like to thank all of my friends from outside CalTech. Even though most of you are far away, your continued support and friendship has made all the difference to me. Kara, Chris, Jules, Tim, and the rest of the hot soup squad, you guys have been like family. And last, but certainly not least, I want to thank Rachel for all of her love and support, especially over the past 6 months. I want to thank all of you guys for the phone calls, IM conversations, and visits that kept things interesting and helped me remember that there is a world outside of lab.

Abstract

Molecular recognition of DNA has important applications for gene regulation, molecular biology, and DNA nanotechnology. Pyrrole-Imidazole polyamides are a unique class of molecules with the ability to bind to DNA in a programmable manner. These small molecule analogues of distamycin A can be programmed to target virtually any DNA sequence with high affinity and specificity. Originally characterized for their ability to bind to B-form DNA, polyamides are also able to target DNA in architectures such as the nucleosome core particle (NCP) and two-dimensional DNA nanostructures including DX-arrays and DNA origami. In addressing DNA nanostructures, polyamide-biotin conjugates can be used to create nanoscale molecular assemblies in a bottom-up approach to self-assembly. The ability to address unique sequences on a DNA nanostructure with different polyamides makes it possible to create unique arrangements of protein on a single 2-dimensional DNA template. Polyamides targeted to the NCP can be used for a variety of exciting applications including NCP-templated ligation reactions, gene regulation, and as tools for X-ray crystallography. The programmability of polyamides makes them an ideal tool for addressing a variety of DNA architectures for varying applications.

Table of Contents

	Page
Acknowledgements.....	iv
Abstract.....	vii
Table of Contents.....	viii
List of Figures and Tables.....	ix
 Chapter 1 Introduction to DNA-Binding Polyamides.....	 1
 Chapter 2 Programming Protein Patterns on DNA Nanostructures.....	 16
 Chapter 3 Addressing DNA Origami with Polyamide-Biotin Conjugates.....	 51
 Chapter 4 Targetting the Nucleosome Core Particle with Polyamide Dimers	 69
 Appendix A Reducing Algorithmic Assembly Error Rates using Polyamides.....	 106

List of Figures and Tables

Chapter 1	Page
Figure 1.1	Molecular recognition of the minor groove of DNA3
Figure 1.2	Crystal structures showing polyamides bound in a 2:1 complex with DNA4
Figure 1.3	Polyamide motifs for binding extended sequences5
Figure 1.4	Overview of higher-order DNA structures7
Figure 1.5	Crystal structures of hairpin polyamides bound to the NCP9
Figure 1.6	Crystal structure of the NCP bound by a turn-to-turn tandem polyamide10
Figure 1.7	Diagram illustrating models for NCP dissociation.....11
 Chapter 2	
Figure 2.1	A DX AB array of DNA modified by a polyamide-biotin conjugate and streptavidin.....20
Figure 2.2	Structures and ball-and-stick models of polyamide conjugates 1 and 1-EDTA21
Figure 2.3	Models of five DX-tiles used to study various binding orientations22
Figure 2.4	Affinity cleavage gels.....23
Figure 2.5	Affinity cleavage mapping.....24
Figure 2.6	DNase I quantitative footprinting with polyamide 125

Figure 2.7	Electrophoretic Mobility Shift Assay of 1	26
Figure 2.8	AFM Images of combined DX tiles A and B	27
Figure 2.9	A DX array consisting of four tiles.....	28
Figure 2.10	Structure and ball-and-stick models for polyamide-biotin conjugates 1 , 2 , and 3	28
Figure 2.11	AFM images for individual polyamides incubated with DX-ABCD and streptavidin.....	29
Figure 2.12	AFM images of DX-ABCD incubated with combinations of polyamides 1 , 2 , and 3	31
Figure 2.13	Crystal structures of zinc finger protein Zif268 and a polyamide bound to DNA.....	32
Figure 2.14	Tile DX-EGR1 and EGR1-DUPLEX..	33
Figure 2.15	Gel shift assay for EGR-1 binding to EGR1-DUPLEX DNA...	33
Figure 2.16	Gel shift for EGR-1 binding to DX-EGR1.....	34
Figure 2.17	AFM images of EGR-1 and DX arrays.....	35
Figure 2.18	Non-denaturing polyacrylamide gel electrophoresis of EGR-1 protein sample.....	36
Figure 2.19	Formation and stability of DXs.....	41
 Chapter 3		
Figure 3.1	Analysis of potential polyamide binding sites on a DNA origami smiley face.	54
Figure 3.2	Predicted binding sites on DNA origami for	

	polyamides 1 and 3	55
Figure 3.3	AFM Images of DNA origami with and without polyamide 1	57
Figure 3.4	AFM Images of DNA origami with polyamide 3	58
Figure 3.5	Comparison of predicted and observed binding patterns for polyamide 1	59
Figure 3.6	Comparison of predicted and observed binding patterns for polyamide 3	60
Figure 3.7	Breaking the symmetry of the DNA smiley	62
Figure 3.8	Polyamides proposed to have improved specificity for DNA sequences in the DNA smiley.....	64
 Chapter 4		
Figure 4.1	Synthetic scheme for the synthesis of polyamide homodimers	74
Figure 4.2	Small library of polyamide clamps containing sterically bulky linkers	74
Figure 4.3	DNase I footprinting experiments with proposed NCP-specific clamps.	75
Figure 4.4	Synthetic scheme for synthesizing PEG diacids from starting diols	77
Figure 4.5	DNase I footprinting of compounds 1 , 7 , and 8	78
Figure 4.6	DNase footprinting of compounds 1 , 9 , and 10	80
Table 4.1	Equilibrium association constants, K_a (M^{-1}) on plasmids pJDC1 and pJDC2.....	81

Figure 4.7	Binding sites for polyamide-EDTA conjugates 11 - 15 on the NCP	83
Figure 4.8	Affinity cleavage on the NCP	84
Figure 4.9	Melting temperature analysis for end clamp polyamides	85
Figure 4.10	Synthesis of end clamp 20	87
Figure 4.11	Schematic for NCP-mediated ligation.....	88
Figure 4.12	Analysis of the linker dependence of NCP templated ligations.	89
Figure 4.13	NCP-templated ligation between different length azides and alkynes.....	90
Figure 4.14	DNA templated ligation on the NCP.....	91
Figure 4.15	Control reactions for the NCP-templated ligation reaction.....	92
Figure 4.16	Representative gels of NCP reconstitution.....	101

Appendix A

Figure A.1	Overview of algorithmic assembly.....	109
Figure A.2	Polyamide-biotin conjugates and DX Tiles for algorithmic assembly	112
Figure A.3	Labeling an array formed by algorithmic self-assembly with polyamide-biotin conjugates.	113
Figure A.4	AFM images of algorithmic arrays labeled with polyamide 1	114
Figure A.5	Non-denaturing PAGE of the redesigned DX tiles for algorithmic assembly	118

Chapter 1

Introduction to DNA-Binding Polyamides

1.1 DNA-Binding Molecules

The genetic code for all organisms is encoded in the double stranded molecule DNA. While DNA itself contains all of an organism's genes, transcription is regulated by an intricate network of interactions. The binding of protein transcription factors to specific sequences of DNA is an important part of gene regulation. With the human genome estimated to contain over 25,000 genes,^{1,2} DNA-binding biological or chemical molecules involved in transcriptional regulation must be able to bind to specific sequences of DNA with high affinity and specificity.

The structure of DNA consists of two antiparallel polydeoxyribonucleotide strands intertwined as a double helix.³ Hydrogen bonds between specific base pairs hold the two strands together. The pairing between DNA bases is highly specific: adenine pairs with thymine, and guanine pairs with cytosine. The DNA helix possesses two grooves, a broad shallow major groove and a narrow and deep minor groove. Differences in the surfaces of these grooves provide the basis for selective DNA recognition.

Distamycin A, composed of three aromatic N-methylpyrrole rings, is a relatively simple natural product capable of DNA recognition. This small organic molecule has been shown to bind preferentially in the minor groove of A,T tracts in 1:1 and 2:1 complexes with DNA.^{4,5} The amide backbone of distamycin A makes a series of hydrogen bonds to the N3 of adenine and O2 of thymine in the DNA sequence. In addition, the crescent shape of distamycin allows it to follow the curvature of the minor groove and make van der Waals contacts with its walls.^{4,5}

The Dervan group has shown that polyamides, small organic analogues of distamycin A, are able to bind DNA in a sequence specific fashion.^{6,7} *N*-methylpyrrole (Py) heterocycle carboxamides when bound to DNA in a 2:1 complex were found to have high affinity for A and T base pairs. It was then shown that the use of *N*-methylimidazole (Im) paired with Py is specific for the G·C basepair.^{8,9} The exocyclic amine of guanine sterically clashes with pyrrole but is tolerated by imidazole which it is also capable of making a hydrogen bond with, thus imparting specificity for the G·C base pair.

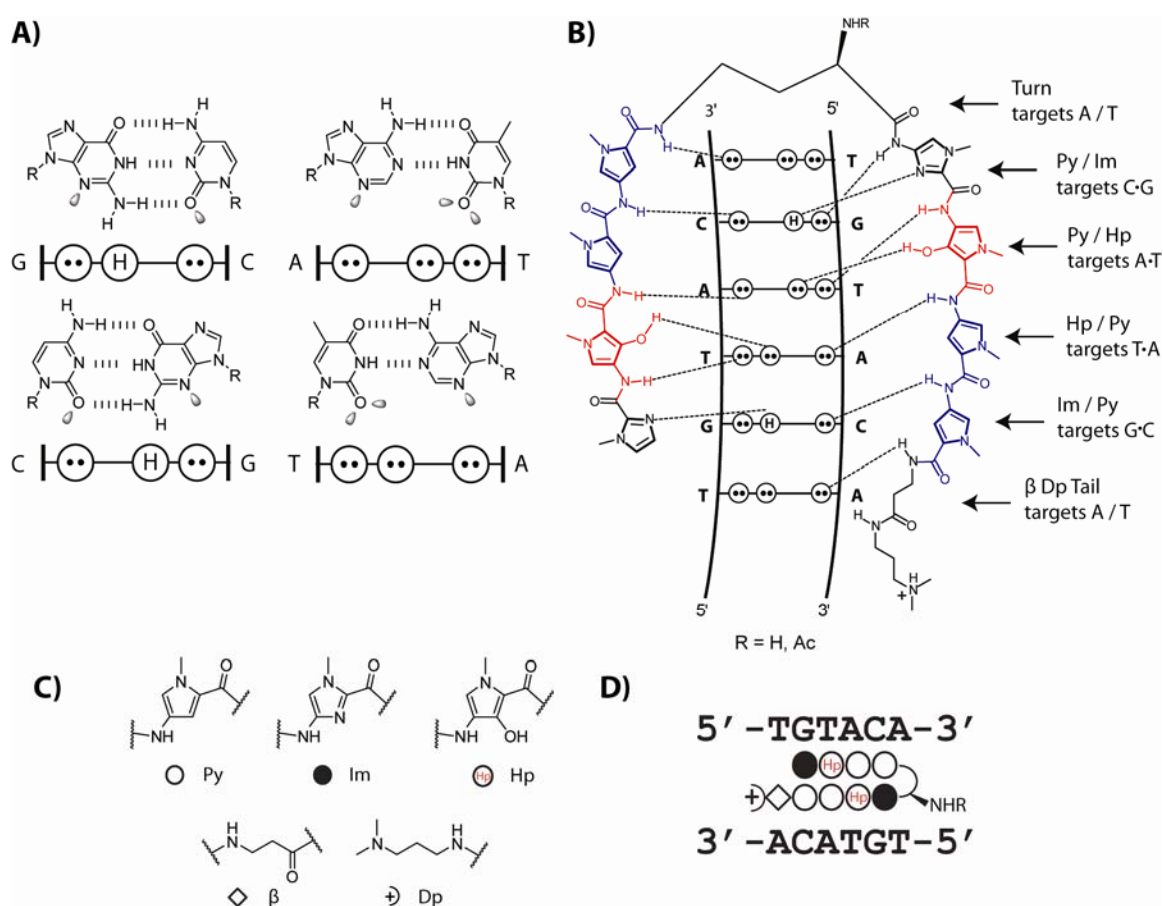


Figure 1.1 Molecular recognition of the minor groove of DNA. (A) Minor groove hydrogen-bonding patterns of Watson-Crick base pairs. Circles with dots represent lone pairs of N(3) of purines and O(2) of pyrimidines, and circles containing an H represent the 2-amino group of guanine. The R group represents the sugar-phosphate backbone of DNA. Electron lone pairs projecting into the minor groove are represented as shaded orbitals (B) Binding model for the complex formed between ImHpPyPy-γ-ImHpPyPy-β-Dp and a 5'-TGTACA-3' sequence. Putative hydrogen bonds are shown as dashed lines. (C) Chemical structures of commonly used monomers in polyamides along with their ball and stick representations. D) Ball-and-stick model of the hairpin polyamide shown in B.

Similarly, the use of *N*-methyl-3-hydroxypyrrole (Hp) which can form a combination of hydrogen bonds between the hydroxyl and the thymine O2 can specifically recognize the T·A base pair when it is paired across from Py.¹⁰ Thus a set of pairing rules have been established whereby Im/Py, Py/Im, Hp/Py, Py/Hp pairings are specific for the G·C, C·G, T·A, and A·T base pairs, respectively. The Py/Py pairing is degenerate for A·T or T·A. In the most commonly used motif, two antiparallel strands of polyamides are linked N-terminus to C-terminus using γ -aminobutyric acid to form a hairpin polyamide.^{11, 12} In general polyamides are usually oriented N to C in the 5' to 3' direction of the DNA

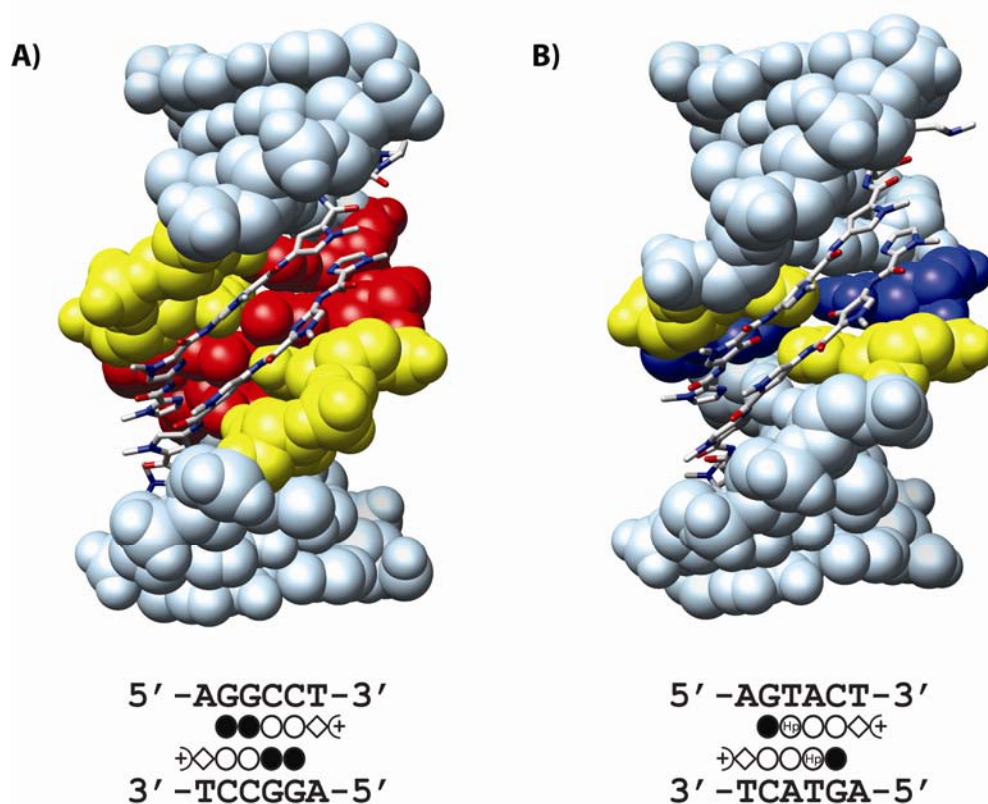


Figure 1.2 Crystal structures showing polyamides bound in a 2:1 complex with DNA. A) Structure of a polyamide homodimer bound to a 5'-GGCC-3' core sequence. Guanine and cytosine are colored red and yellow respectively. A Ball-and-stick model of the Im-Im-Py-Py- β -Dp homodimer is shown at bottom. B) Structure of the polyamide homodimer Im-Hp-Py-Py- β -Dp bound to a 5'-GTAC-3' core sequence. Adenine and thymine are colored blue and yellow respectively. The Ball-and-stick model of the homodimer is shown at bottom.

strands, although exceptions have been observed.¹²

Given the large size of the human genome, the ability to target increasingly long sequences of DNA is important for gene regulation, as one often wants to limit the number of sites affected by any given molecule to prevent unwanted side effects. A standard eight-ring hairpin polyamide has specificity for 6 bp of DNA. It has been shown that polyamide curvature no longer matches that of DNA after five contiguous rings,¹³ limiting the length of sequences that can be targeted with this motif. As a result, a variety of extended motifs have been developed to target longer sequences than what is possible with a single hairpin.⁷ Two six-ring polyamides linked head-to-head or turn-to-tail have been shown to be able to recognize sequences as long as 10 bp with high affinity.¹⁴⁻¹⁶

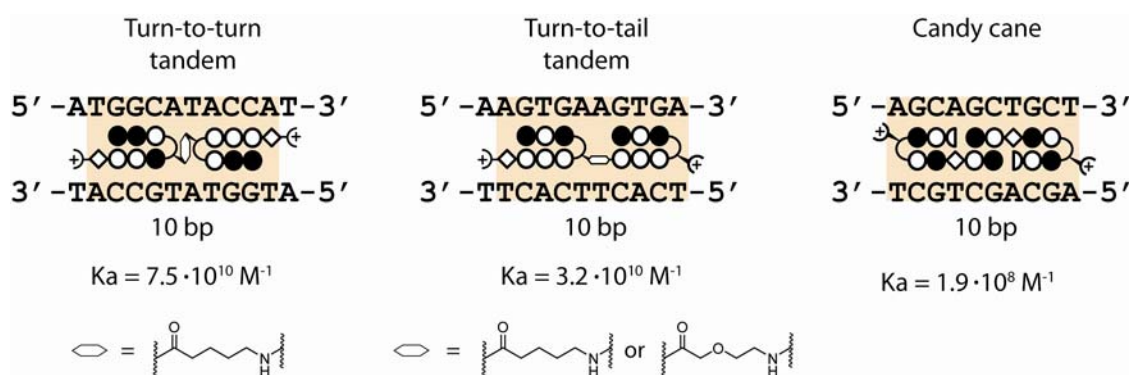


Figure 1.3 Polyamide motifs for binding extended sequences. Ball-and-stick models as well as the corresponding binding site size and affinity are shown.

One of the most active areas of research using polyamides is gene regulation. Polyamides targeted to the promoter region of a gene can be used to disrupt transcription factor binding and inhibit gene transcription.^{7, 17} Recent successes in inhibiting vascular endothelial growth factor (VEGF) and androgen receptor (AR) have demonstrated the power of this approach.¹⁸⁻²⁰ By interfering with the binding of hypoxia-inducible factor (HIF-1) to the hypoxia response element (HRE), polyamides inhibited the hypoxia-

induced gene expression of VEGF and other HIF-1 related genes in cultured cells. Similarly, polyamides to the androgen response element resulted in inhibition of androgen-induced expression of prostate-specific antigen (PSA) and other AR-regulated genes in cultured prostate cancer cells. In a complimentary field of research, there has been much progress in the use of polyamides for the creation artificial transcription factors.²¹⁻²⁴ The goal of these ambitious studies is to combine polyamides which function as a DNA-binding domain with minimal activation domains that can recruit the transcriptional machinery. In this manner, polyamides can be used as programmable molecules for gene regulation.

1.2 Addressing higher-order DNA structures

While the ability to address B-form DNA with polyamides has been well-characterized, it was not known whether polyamides could successfully bind to higher-order DNA structures. Nature uses higher-order DNA structures to compact the DNA in cells. Chromosomes consist of compacted chromatin, which is itself comprised of nucleosomes. Nucleosomes, comprised of DNA wound around a core of histone proteins, represent the most basic unit used by eukaryotes for condensing their DNA. The super-helical structure of the DNA wrapped around the histones in nucleosomes, as well as the partial blockage of DNA sites that face the histone core, present unique challenges for molecular recognition of DNA in these architectures.

In addition to the higher-order structures found in nature, the use of complex DNA architectures has found profound use in structural DNA nanotechnology.^{25, 26} Beginning with pioneering work by Ned Seeman, it was originally demonstrated how

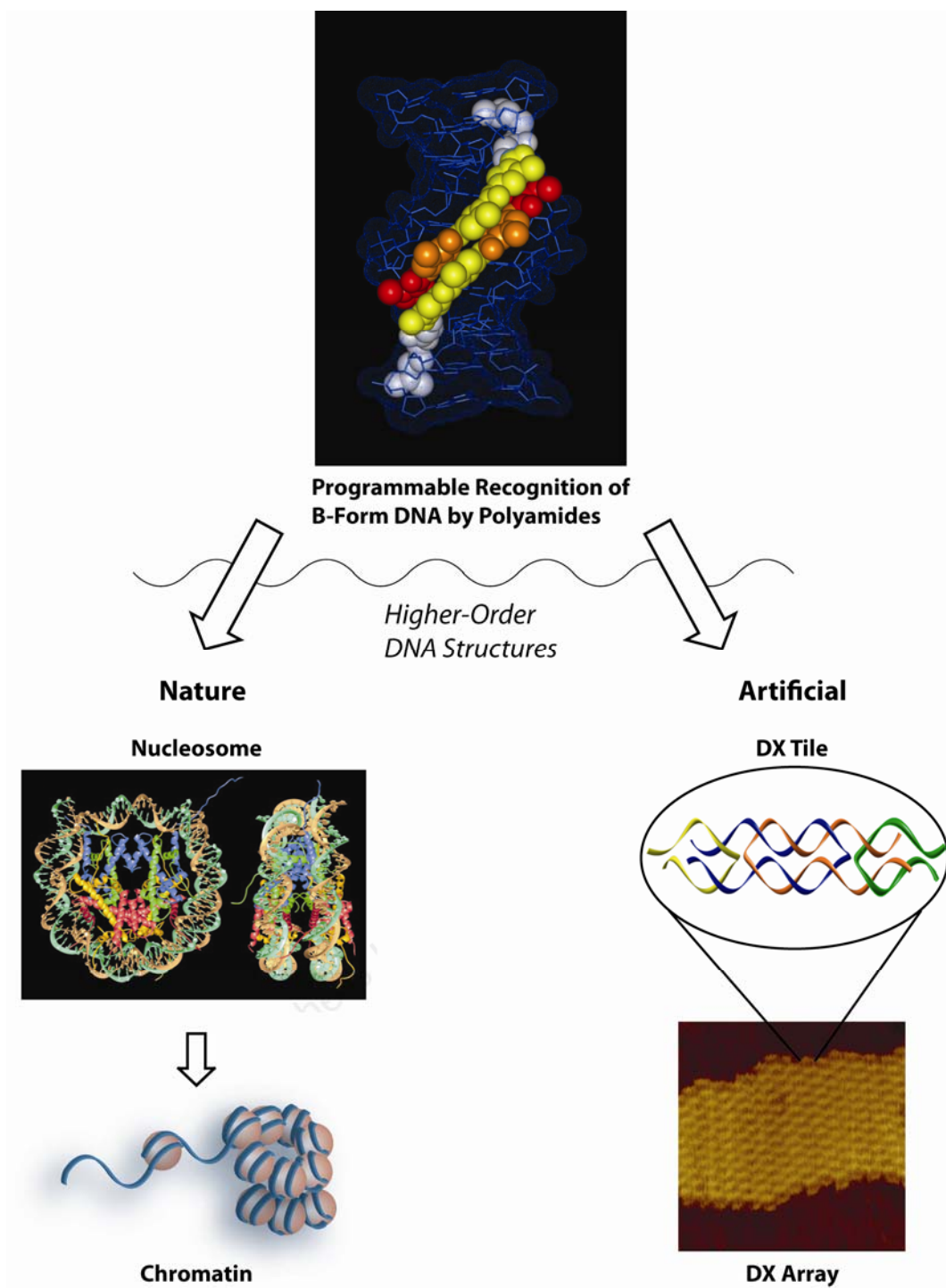


Figure 1.4 Overview of higher-order DNA structures. At top, recognition of B-form DNA by ImHpPyPy- β -Dp polyamide. Higher-order DNA structures are shown in increasing complexity. Nature uses the nucleosome to condense DNA,²⁷ which is in turn compacted into chromatin. At right, artificial DNA nanostructures can be used to create complex 2-dimensional DNA arrays.

DNA could be made to self-assemble into rigid 2-dimensional structures.²⁸ These Holliday junction mimics consist of two helices of DNA linked by stable crossover junctions. These stable DNA structures, known as DX tiles, possess four sets of sticky ends, and can be made to assemble into 2-dimensional arrays.²⁹ Since then, a large number of DNA architectures have been described, based upon or extending the basic DX motif.

While the origins and functions of nucleosomes and DX-arrays are quite distinct, they both raised a common question. Can polyamides that target B-form DNA be used to address these higher-order DNA structures? The primary goal of the investigations contained in this thesis has been to answer that question.

1.3 Targeting the Nucleosome Core Particle

In eukaryotes, DNA is found as chromatin, a highly compact complex of DNA and protein. The primary unit of chromatin is the nucleosome core particle (NCP) consisting of a 147 bp sequence of DNA wrapped approximately twice around an octamer of histone proteins. A short linker strand between 20 and 80 bp of DNA links successive NCPs.³⁰

Chromatin exists in numerous states which can be transcriptionally active in the form of euchromatin, or transcriptionally inactive as in heterochromatin. The transition between these two states is thought to be regulated at the most basic level by changes in the nucleosome structure and the surface of the histone octamer.³⁰ When DNA is packaged in nucleosomes, only 70% of the DNA surface is solvent accessible as the portion that faces the histone core is sterically blocked. Along the outer surface, the two

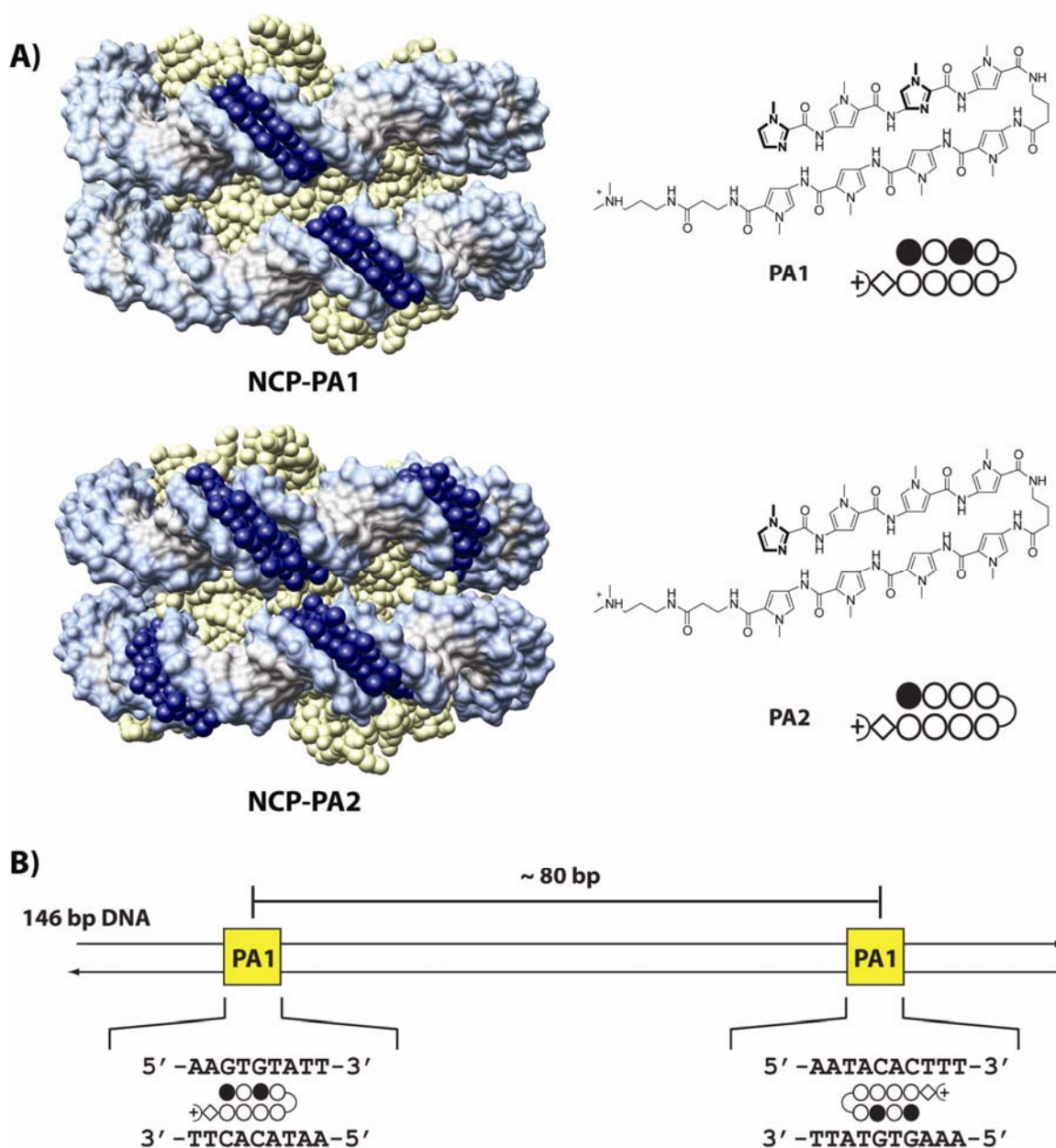


Figure 1.5 Crystal structures of hairpin polyamides bound to the NCP. A) Chemical structures and ball-and-stick models are shown for each polyamide next to the corresponding crystal structure. **PA1** and **PA2** are colored navy blue in structures. (PDB code 1m18 and 1m19) B) Schematic of the palindromic 146 bp DNA strand of the NCP with the binding sites for **PA1** shown in yellow. The sites are separated by ~80 bp.

gyres of DNA stack, forming “supergrooves” of DNA in which the major and minor grooves are aligned to provide a 14 – 16 bp platform for molecular recognition. The

ability to bind to and modulate the stability of nucleosomes would have obvious importance for gene regulation in eukaryotes.

It was shown that polyamides are able to bind to reconstituted nucleosome core particles with high affinity and specificity for their target DNA sites.³¹ Subsequently, high resolution X-ray crystal structures were obtained showing the NCP in complex with hairpin polyamides.³² Surprisingly, polyamides were able to access many sites facing away, and even partially facing towards the histone proteins. In addition polyamide binding had no adverse effect upon nucleosome stability or reconstitution.

The presence of a supergroove on the NCP containing two aligned minor groove binding sites for **PA1**, led to the hypothesis that two polyamides joined together with a linker of sufficient length should be able to bind a single supergroove as a linked dimer. Three turn-to-turn dimers were synthesized containing the same polyamide joined by linkers containing 2, 3, and 4 ethylene glycol units. All three clamps were shown to bind

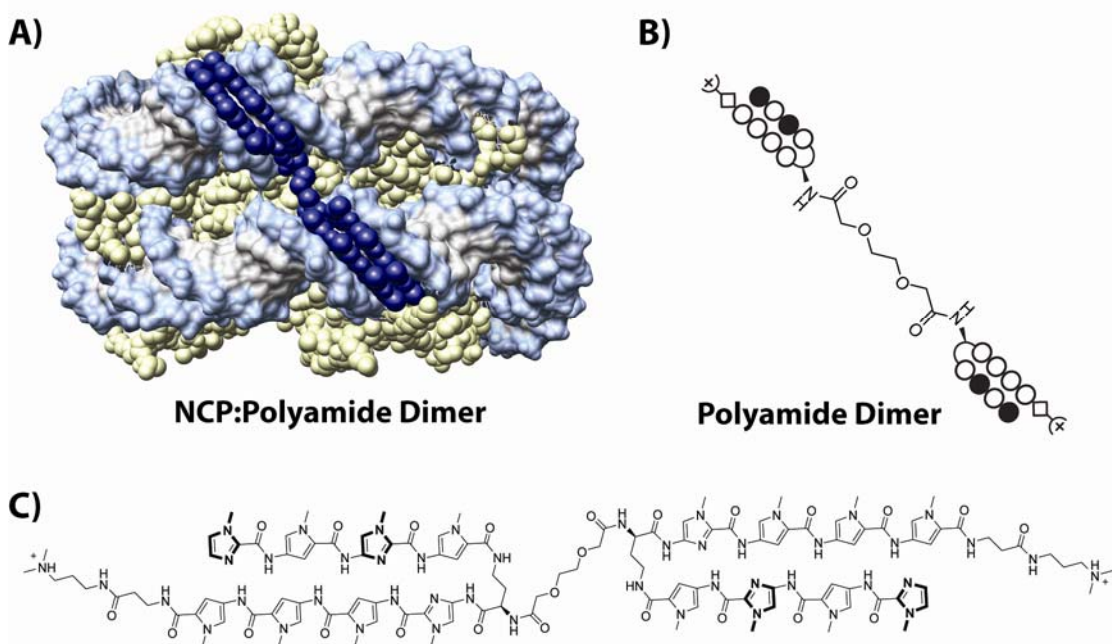


Figure 1.6 Crystal structure of the NCP bound by a turn-to-turn tandem polyamide. A) Structure of the NCP bound by a polyamide dimer. The polyamide is shown in navy blue. (PDB code 1S32) B) Ball-and-stick model of the turn-to-turn polyamide dimer, C) Chemical Structure of the polyamide dimer.

to the NCP with similar affinities in the low nanomolar range.³³ The crystal structure shows that the polyamide dimer binds in the predicted super groove with each polyamide aligned against its target DNA sequence and the PEG linker spanning the ~ 12 Å gap between the two gyres of DNA.³³ High resolution crystal structures were obtained showing the dimers bound in the supergroove of the NCP as predicted. The polyamide dimer was referred to as a “clamp” because of its ability to clamp the two gyres of DNA

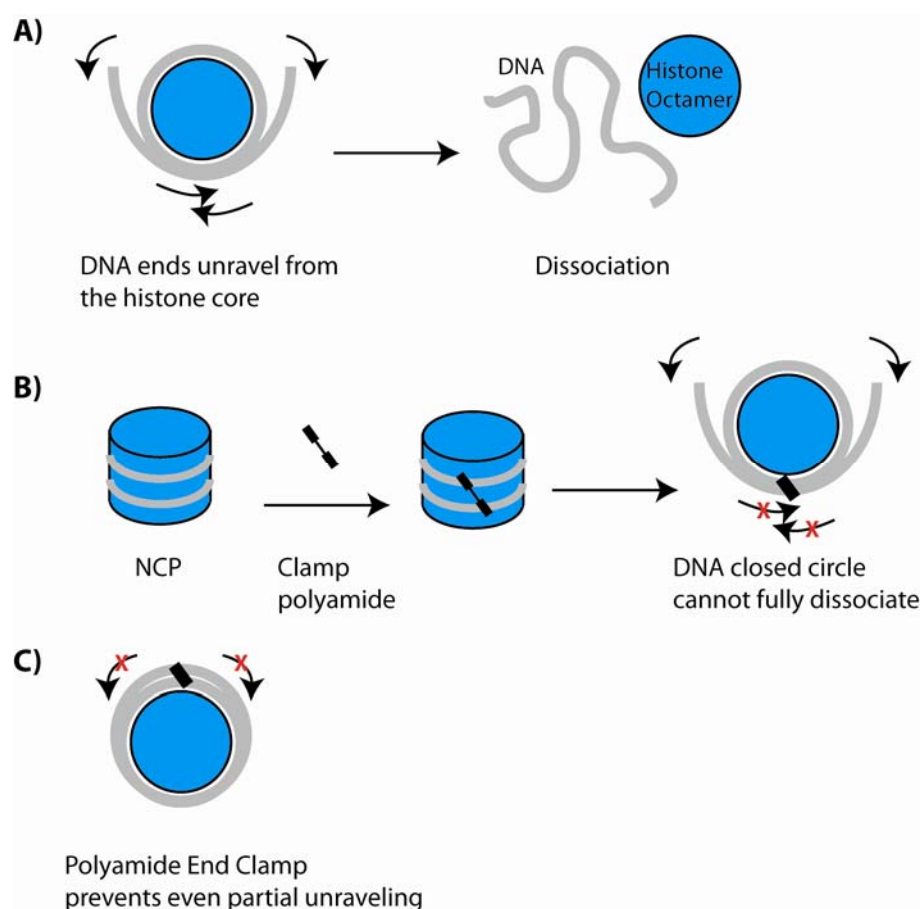


Figure 1.7 Diagram illustrating models for NCP dissociation. A) Model for the dissociation of the NCP into DNA and the histone proteins. B) The binding of the polyamide clamp to the NCP creates a full circle of DNA which is thought to prevent full unraveling and to explain the observed increase in NCP stability. C) A polyamide clamp targeted to the ends of the DNA could be able to prevent even partial unraveling.

on the NCP together, resulting in an increase in NCP stability. This is presumably due to the formation of a closed circle of DNA around the histone core which would be unable to fully unravel as illustrated in Figure 1.6B.

One unexpected result of this study was that the clamp polyamide in complex with the NCP resulted in the growth of extremely large, well-ordered crystals. This is presumed to be a direct effect of the stabilization effect the clamp has on the NCP. The future use of polyamide clamps to stabilize the NCP, could prove useful for generating high quality crystals needed for future crystallography studies of nucleosomes and chromatin.

1.4 Scope of Work

As outlined in the previous sections, polyamides have been previously used to target both B-form DNA, and DNA present in the nucleosome core particle, a biological DNA architecture. The uniting theme of the work described in this thesis has been the expansion of the scope of DNA architectures that can be targeted using sequence-specific DNA-binding polyamides. Primary investigations into using polyamides to target a DX array, a two-dimensional DNA nanostructure, are described in Chapter 2. These studies were the first attempts at using polyamides to target a higher-order DNA structure that can be used for 2-dimensional array assembly. Affinity cleavage experiments first demonstrated the ability of polyamides to target a variety of binding sites on a single DX tile. Additional AFM studies demonstrated the use of polyamides in recruiting protein to 2-dimensional DX arrays. In addition, several polyamide-biotin conjugates were used to

assemble proteins on a four component DX array in distinct patterns. In Chapter 3, we examine another DNA nanostructure, DNA origami, and the ability of polyamide conjugates to target these structures. Finally, Chapter 4 describes the work done in targeting nucleosomes, a biological DNA architecture, with polyamide dimers. The use of novel polyamide dimers to further stabilize the NCP, as well as the use of the NCP as a unique template for chemical reactions is examined. It is hoped that the work described herein will expand the use of polyamides from chemical agents for gene regulation, to tools for molecular biology, and reagents for the creation of complex molecular assemblies.

1.5 References

1. Venter, J. C., et al., The sequence of the human genome *Science* **2001**, 291, 1304-1351.
2. Lander, E. S., et al., Initial sequencing and analysis of the human genome *Nature* **2001**, 409, 860-921.
3. Dickerson, R. E., et al., The anatomy of A-DNA, B-DNA, And Z-DNA *Science* **1982**, 216, 475-485.
4. Pelton, J. G.; Wemmer, D. E., Binding modes of distamycin A with d(CGCAAATTTGCG)₂ determined by 2-dimensional NMR *J. Am. Chem. Soc.* **1990**, 112, 1393-1399.
5. Pelton, J. G.; Wemmer, D. E., Structural characterization of a 2-1 distamycin A.d(CGCAAATTGGC) complex by two-dimensional NMR *Proc. Natl. Acad. Sci. U. S. A.* **1989**, 86, 5723-5727.
6. Dervan, P. B., Molecular recognition of DNA by small molecules *Bioorg. Med. Chem.* **2001**, 9, 2215-2235.
7. Dervan, P. B.; Edelson, B. S., Recognition of the DNA minor groove by pyrrole-imidazole polyamides *Curr. Opin. Struct. Biol.* **2003**, 13, 284-299.

8. Kielkopf, C. L.; Baird, E. E.; Dervan, P. D.; Rees, D. C., Structural basis for G center dot C recognition in the DNA minor groove *Nat. Struct. Biol.* **1998**, 5, 104-109.
9. Pilch, D. S., et al., The thermodynamics of polyamide-DNA recognition: Hairpin polyamide binding in the minor groove of duplex DNA *Biochemistry* **1999**, 38, 2143-2151.
10. Kielkopf, C. L., et al., A structural basis for recognition of A center dot T and T center dot A base pairs in the minor groove of B-DNA *Science* **1998**, 282, 111-115.
11. Mrksich, M.; Parks, M. E.; Dervan, P. B., Hairpin peptide motif - A new class of oligopeptides for sequence-specific recognition in the minor-groove of double-helical DNA *J. Am. Chem. Soc.* **1994**, 116, 7983-7988.
12. White, S.; Baird, E. E.; Dervan, P. B., Orientation preferences of pyrrole-imidazole polyamides in the minor groove of DNA *J. Am. Chem. Soc.* **1997**, 119, 8756-8765.
13. Kelly, J. J.; Baird, E. E.; Dervan, P. B., Binding site size limit of the 2:1 pyrrole-imidazole polyamide-DNA motif *Proc. Natl. Acad. Sci. U. S. A.* **1996**, 93, 6981-6985.
14. Herman, D. M.; Baird, E. E.; Dervan, P. B., Tandem hairpin motif for recognition in the minor groove of DNA by pyrrole - Imidazole polyamides *Chem.-Eur. J.* **1999**, 5, 975-983.
15. Kers, I.; Dervan, P. B., Search for the optimal linker in tandem hairpin polyamides *Bioorg. Med. Chem.* **2002**, 10, 3339-3349.
16. Weyermann, P.; Dervan, P. B., Recognition of ten base pairs of DNA by head-to-head hairpin dimers *J. Am. Chem. Soc.* **2002**, 124, 6872-6878.
17. Gottesfeld, J. M., et al., Regulation of gene expression by small molecules *Nature* **1997**, 387, 202-205.
18. Olenyuk, B. Z., et al., Inhibition of vascular endothelial growth factor with a sequence-specific hypoxia response element antagonist *Proc. Natl. Acad. Sci. U. S. A.* **2004**, 101, 16768-16773.
19. Nickols, N. G.; Dervan, P. B., Suppression of androgen receptor-mediated gene expression by a sequence-specific DNA-binding polyamide *Proc. Natl. Acad. Sci. U. S. A.* **2007**, 104, 10418-10423.

20. Nickols, N. G.; Jacobs, C. S.; Farkas, M. E.; Dervan, P. B., Modulating hypoxia-inducible transcription by disrupting the HIF-1-DNA interface *ACS Chem. Biol.* **2007**, 2, 561-571.
21. Mapp, A. K.; Ansari, A. Z.; Ptashne, M.; Dervan, P. B., Activation of gene expression by small molecule transcription factors *Proc. Natl. Acad. Sci. U. S. A.* **2000**, 97, 3930-3935.
22. Stafford, R. L., et al., Minimization of a protein-DNA dimerizer *J. Am. Chem. Soc.* **2007**, 129, 2660-2668.
23. Stafford, R. L.; Dervan, P. B., The reach of linear protein-DNA dimerizers *J. Am. Chem. Soc.* **2007**, 129, 14026-14033.
24. Arndt, H. D., et al., Toward artificial developmental regulators *J. Am. Chem. Soc.* **2003**, 125, 13322-13323.
25. Aldaye, F. A.; Palmer, A. L.; Sleiman, H. F., Assembling materials with DNA as the guide *Science* **2008**, 321, 1795-1799.
26. Seeman, N. C., DNA in a material world *Nature* **2003**, 421, 427-431.
27. Luger, K., et al., Crystal structure of the nucleosome core particle at 2.8 angstrom resolution *Nature* **1997**, 389, 251-260.
28. Fu, T. J.; Seeman, N. C., Dna Double-Crossover Molecules *Biochemistry* **1993**, 32, 3211-3220.
29. Winfree, E.; Liu, F. R.; Wenzler, L. A.; Seeman, N. C., Design and self-assembly of two-dimensional DNA crystals *Nature* **1998**, 394, 539-544.
30. Luger, K., Structure and dynamic behavior of nucleosomes *Curr. Opin. Genet. Dev.* **2003**, 13, 127-135.
31. Gottesfeld, J. M., et al., Sequence-specific recognition of DNA in the nucleosome by pyrrole-imidazole polyamides *J. Mol. Biol.* **2001**, 309, 615-629.
32. Suto, R. K., et al., Crystal structures of nucleosome core particles in complex with minor groove DNA-binding ligands *J. Mol. Biol.* **2003**, 326, 371-380.
33. Edayathumangalam, R. S., et al., Molecular recognition of the nucleosomal "supergroove" *Proc. Natl. Acad. Sci. U. S. A.* **2004**, 101, 6864-6869.

Chapter 2

Programming Protein Patterns on DNA Nanostructures

The text of this chapter was taken in part from two manuscripts co-authored with John Sadowski and Peter Dervan.

Cohen, J. D.; Sadowski, J. P.; Dervan, P. B., *Angew. Chem. Int. Ed.* **2007**, 46, 7956-7959

Cohen, J. D.; Sadowski, J. P.; Dervan, P. B., *J. Am. Chem. Soc.* **2008**, 130, 402-403

Abstract

DNA nanotechnology offers the potential for the creation of molecular assemblies with nanometer scale resolution. Although DNA provides an ideal material for the construction of 2-dimensional and 3-dimensional assemblies it lacks the functionality needed for many desired applications. Pyrrole-imidazole polyamides capable of programmable binding to specific DNA sequences provide a unique solution to the challenge of recruiting active biomolecules to a DNA template. Polyamides are capable of binding to a 2-dimensional DNA array and recruiting protein to specific sites on the nanostructure. Multiple polyamides targeted to different sites on a two dimensional array can be used to create unique protein patterns on the same DNA template. Similar efforts to bind to a two-dimensional DNA array directly with the DNA binding protein EGR-1 were done to compare the effectiveness of each approach.

2.1 Introduction

DNA nanotechnology

The ability to create assemblies of proteins with spacing on the nanometer scale has important implications for proteomics, biodetection, and self-assembly. The goal of such systems is to be able to position proteins or other components in distinct patterns with precise spacing. Structural DNA nanotechnology has led to the creation of a variety of nanostructures which should be capable of serving as an addressable template for the creation of complex molecular assemblies.¹⁻⁴ These systems take advantage of the well defined structure and spacing of DNA and use these properties to act as a template for secondary components in a bottom-up approach towards self-assembly. The structure of an assembled DNA complex is directly and uniquely determined by the sequence of the DNA bases which can be designed and manipulated. These methods allows for a versatile and programmable way to control the structure and architecture of DNA nanostructures.

Pioneering work by Ned Seeman demonstrated how stable crossovers mimicking a Holliday junction could be used to create a rigid and well-behaved DNA structure known as the double crossover molecule (DX).⁵ The DX consists of two helices of DNA that are held together by two stable crossovers. The termini of each of the four helices contained in this molecule have sticky ends which provide an easy way to design interactions between different DX tiles. By creating matching sticky ends on different DX tiles, one gives the molecules the ability to self-assemble into a rigid periodic two-dimensional array of DNA.⁶ A variety of complex DNA nanostructures using or extending the DX motif have been designed to create structures such as arrays, nanogrids,

triple crossovers (TX), three- and six-pointed stars, and DNA origami.^{1, 7-12} The use of DNA for the creation of 3-dimensional structures has been demonstrated as well.^{13, 14}

Although it is an ideal material for the creation of designed nanostructures, DNA itself lacks the functionality desired for applications such as novel sensors, electronic devices, or enzymatic arrays. In order to be useful as an architecture for molecular assembly, a way to recruit molecular components to the DNA motif is necessary so that the self-assembly of the DNA molecules would also serve as a template for the assembly of secondary molecular components.

Previous work in this field has focused on the use of chemical or structural modifications of the DNA template in order to attach or recruit proteins or nanoparticles.^{1, 15-21} A variety of methods have been previously used to attach molecules to specific tiles in a DX array. These methods include the incorporation of an additional dsDNA hairpin into specific tiles which projects perpendicular from the plane of the array,^{6, 22} the use of biotin-labelled DNA to recruit streptavidin,^{6, 8, 23} conjugation of DNA to gold nanoparticles,^{15, 24} insertion of the thrombin-binding aptamer and subsequent recruitment of thrombin,¹⁷ insertion of DNA aptamers that can be recognized by single chain antibodies,¹⁶ and the recognition of a DNA-peptide fusion by an antibody.²¹ All of these methods however, rely upon the covalent modification of the DNA modules prior to array assembly to allow visualization.

Pyrrole-imidazole polyamides are a class of oligomers that bind with high specificity to the minor groove of DNA.^{25, 26} They can be programmed to target a broad repertoire of DNA sequences. Unlike the previously discussed methods used to attach molecules to DNA nanostructures, polyamides do not require any previous modification

of the target DNA. The highly programmable nature of pyrrole-imidazole polyamides make them particularly attractive for targeting specific DNA sequences.^{26, 27}

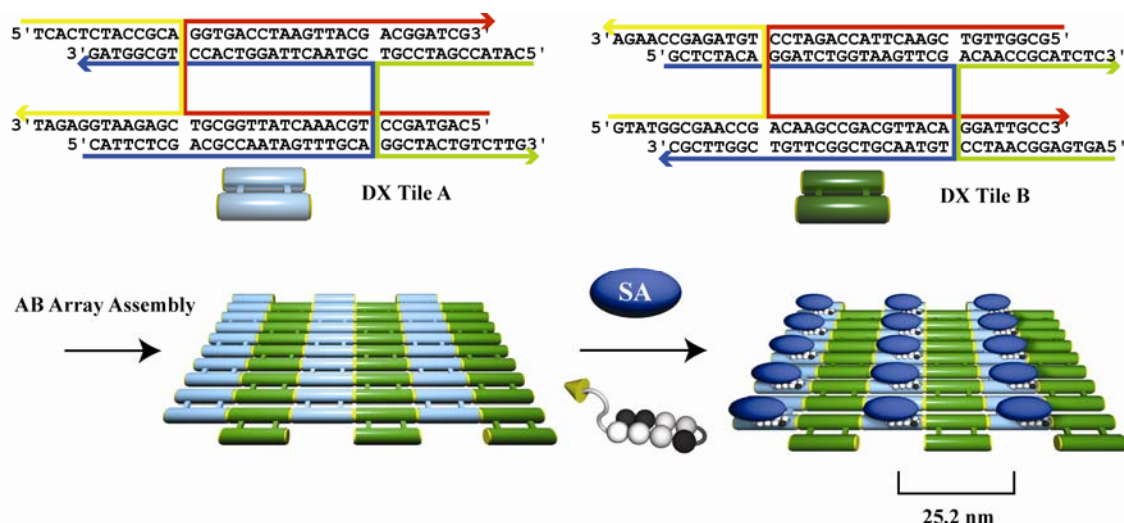


Figure 2.1 A DX AB array of DNA modified by a polyamide-biotin conjugate and streptavidin.

The research in this chapter demonstrates that polyamides are capable of accessing and binding to a DNA nanostructure in a sequence-specific manner. In addition, the ability to use different polyamide conjugates to target unique sites on a DX-array made up of *four* tiles and to arrange proteins into distinct spatial patterns is shown. Finally, analogous efforts to the polyamide-based approach were made with DNA binding proteins and are outlined as well.

2.2 Results and Discussion

Affinity Cleavage on DX Molecules

In order to first ascertain whether the DNA in a DX nanostructure was accessible to polyamides, a series of affinity cleavage experiments were done. Polyamide-EDTA-Fe conjugate **1-EDTA** was programmed to target a single DNA sequence 5' -WGGWCW -

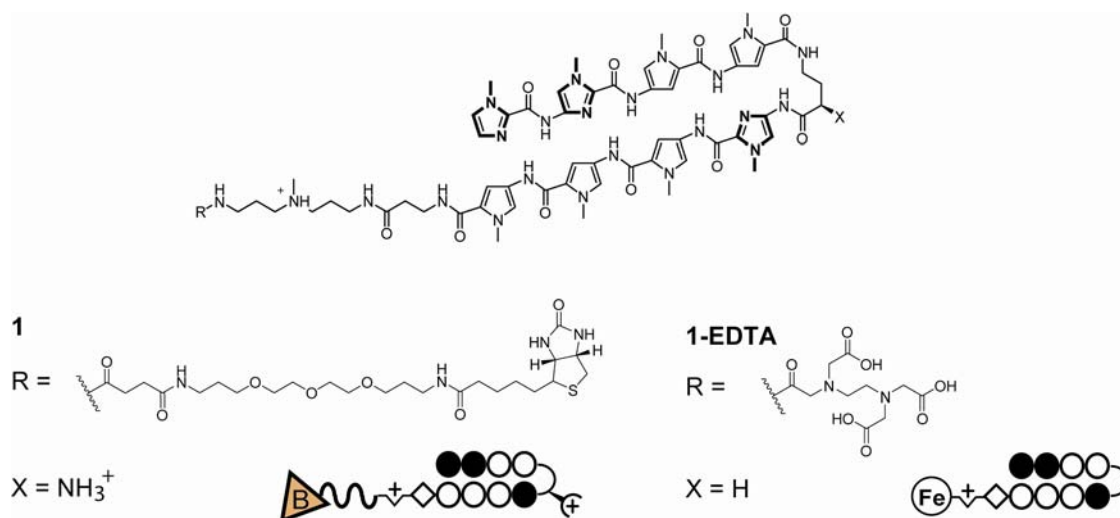


Figure 2.2 Structures and ball-and-stick models of polyamide conjugates **1** and **1-EDTA**. (Filled circle = N-methylimidazole, empty circle = N-methylpyrrole, diamond = β-Ala, half-diamond with plus sign = diamino-N-methyl-dipropylamine, half-circle = γ-aminobutyric acid, half circle with plus sign = NH₃⁺, B in an orange triangle = biotin).

3',²⁸ embedded in a previously characterized DX tile.⁶ (Figure 2.2) One DNA strand was radiolabeled and combined with the three unlabelled strands to form tile **A**. The DX tiles formed with over 90% purity when analyzed by gel electrophoresis and were found to be stable under the reaction conditions used for affinity cleavage. (see Experimental Details Figure 2.19) Cleavage occurred at the predicted binding site and in the expected orientation. (Figures 2.4 and 2.5) The observed cleavage pattern has a 2 – 3 bp 3' shift which is characteristic of minor groove binding polyamides.^{29, 30} Additional affinity cleavage experiments were done using similarly designed DX tiles where the binding site was placed in different locations and orientations along the DX molecule. (Figures 2.3 – 2.5) In all cases cleavage was observed in the expected location, demonstrating the ability of hairpin polyamides to address a variety of locations along a single DX module. The sole exception was when the binding site spanned the crossover region in DX-Junction. No cleavage was observed, indicating that the structural features present prevent polyamide binding at this location.

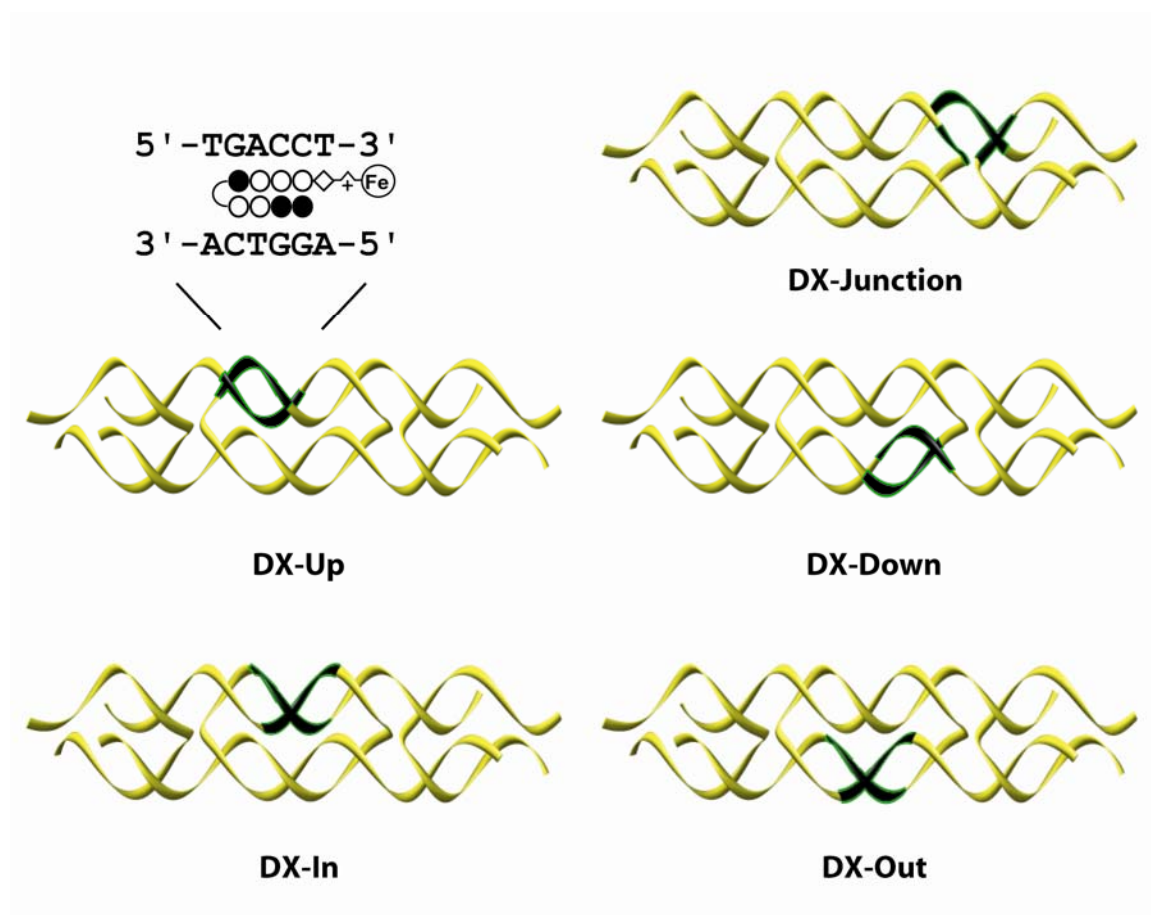


Figure 2.3 Models of five DX-tiles used to study various binding orientations. Each DX consists of four individual DNA strands shown in yellow with the 6 bp polyamide binding site shown in black. The designations of each molecule refer to how the minor groove of each binding site, and thus the polyamide are situated.

(A)

Labelled Strand

DX-Up (Tile A) 5' - CAT TCT CGA CGC CAA TAG TTT GCA CGT AAC TTA **GGT CAC** CTG CGG TAG - 3'

DX-In 5' - CAT TCT CGA CGC CAT AAC TAA GCA CGT ATA **GGT CAT** TGC CTG CGG TAG - 3'

DX-Down 5' - CAT TCT CGA CGC CTT **TAG GTC** ACA CGT AAC TTC ATT TGC CTG CGG TAG - 3'

DX-Out 5' - CAT TCT CGA CGC **TAG GTC** ACA GCA CGT AAC TTC ATT TGC CTG CGG TAG - 3'

(B)

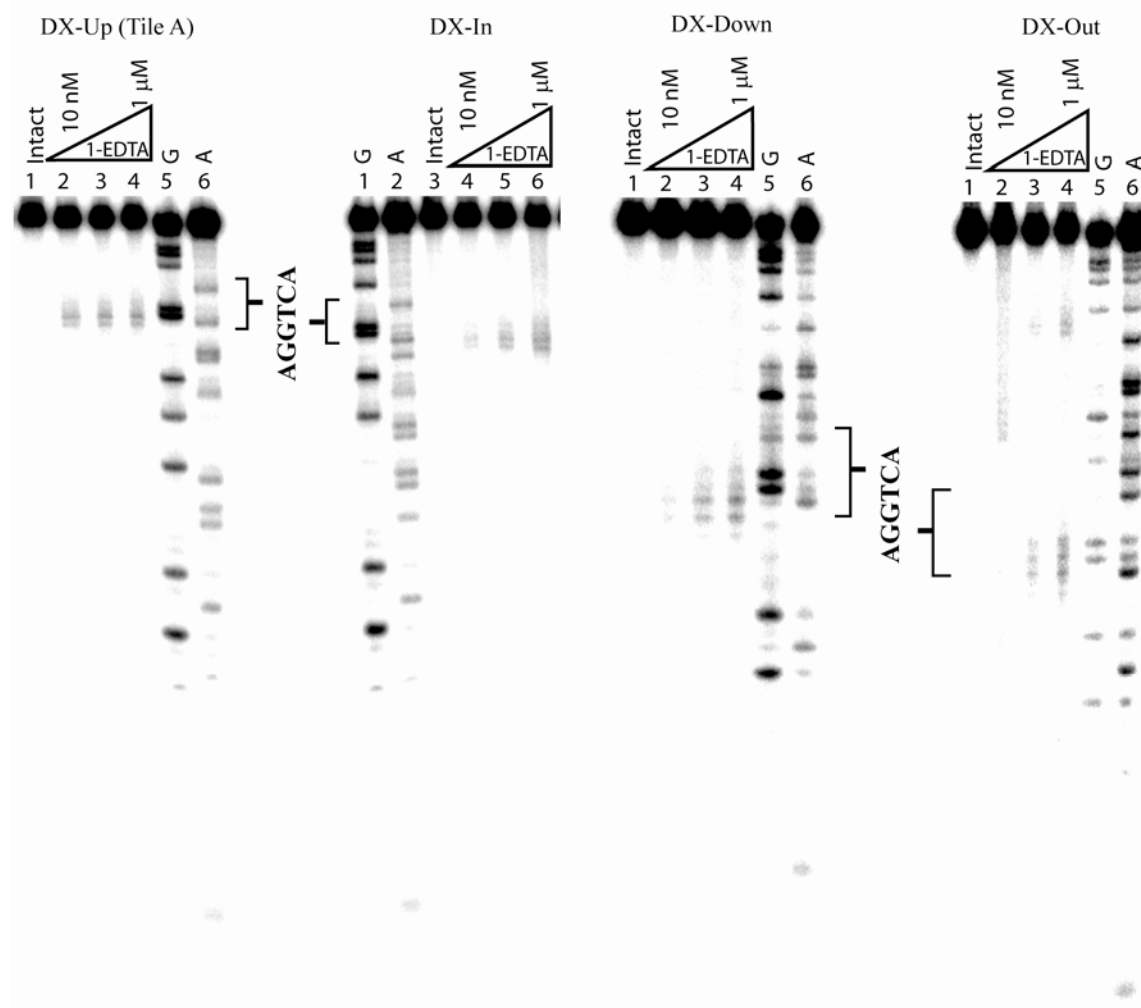


Figure 2.4 Affinity cleavage gels. (a) List of DNA sequences for the labeled strand in each DX tile. The binding site for **1-EDTA** is highlighted in red. (b) Affinity Cleavage on each DX tile. Strand 2 of each DX was ^{32}P -radiolabeled and affinity cleavage was performed using polyamide **1-EDTA**; the complex was then denatured and visualized by gel electrophoresis. The target polyamide binding site is indicated by brackets. A and G sequencing lanes are shown for each gel. Polyamide **1-EDTA** was added at concentrations of 10 nM, 100 nM, and 1 μM. Intact lanes were not incubated with **1-EDTA**.

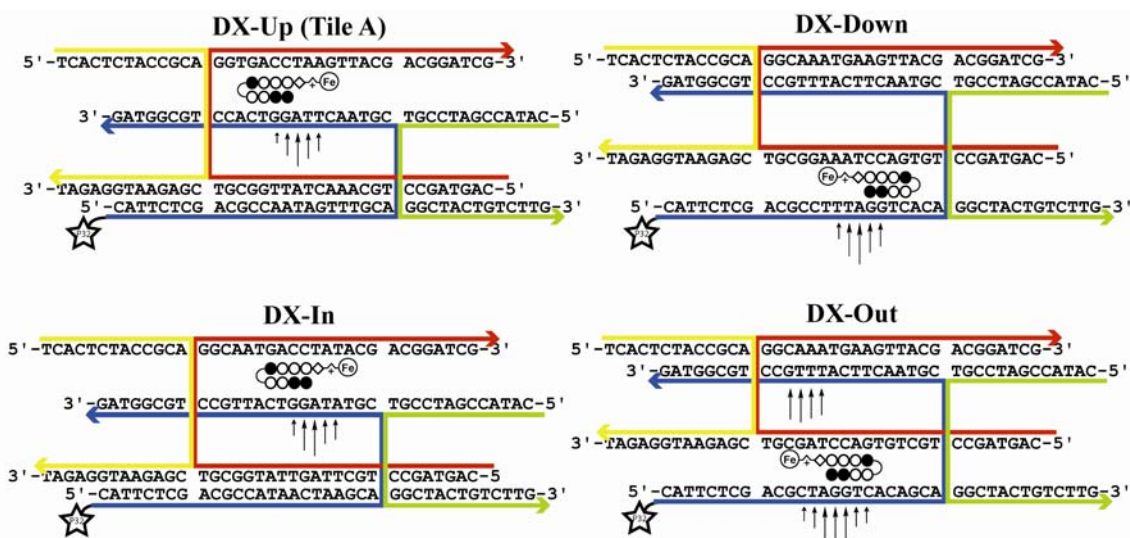


Figure 2.5 Affinity cleavage mapping. Representations of each DX tile with polyamide **1-EDTA**. Arrows represent the extent of cleavage at the indicated base position.

Biophysical Characterization of Polyamide Biotin-Conjugates

Having shown the ability of polyamides to bind to individual DX tiles, we next addressed whether a polyamide conjugated to biotin would be capable of recruiting streptavidin to DNA and therefore could be used as a labelling agent in AFM studies. DNase I footprinting was used to ensure that attachment of a biotin did not interfere with a polyamide's ability to bind to DNA. (Figure 2.6) The ability of the polyamide-biotin conjugate to bind to streptavidin was then tested using an electrophoretic mobility shift assay (EMSA). DNA duplex was incubated with polyamide-biotin conjugate **1** as well as 10 μ M streptavidin. (Figure 2.7) The streptavidin concentration was chosen to ensure a 10-fold excess of the protein compared to polyamide-biotin conjugate **1**. As the concentration of **1** is increased to 30 nM the naked DNA is shifted to a band of lower mobility corresponding to the tertiary complex containing polyamide-biotin, streptavidin and DNA.

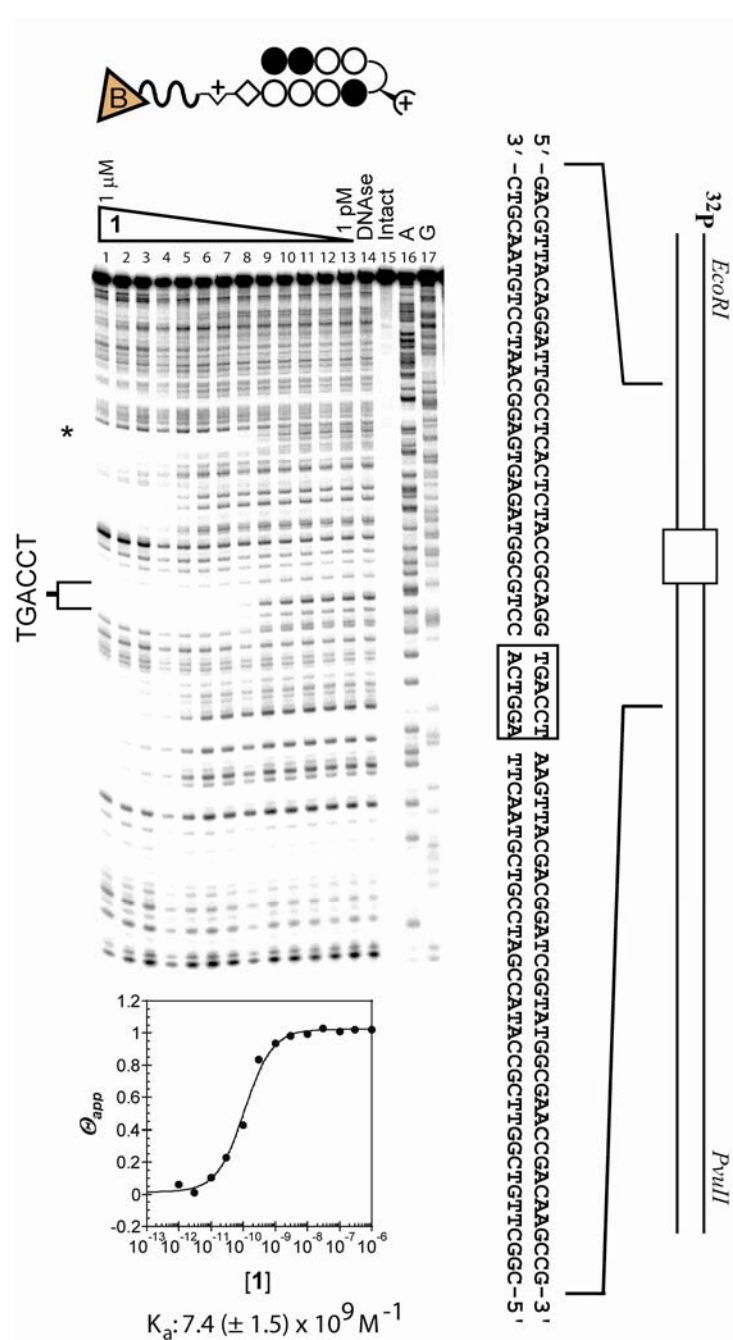


Figure 2.6 DNase I quantitative footprinting with polyamide **1**. The 75 bp insert and schematic of the plasmid are shown with the binding site boxed. Lanes 1-13: 1 μM, 300 nM, 100 nM, 30 nM, 10 nM, 3 nM, 1 nM, 300 pM, 100 pM, 30 pM, 10 pM, 3 pM, and 1 pM of polyamide. Lane 14: DNase standard. Lane 15: Intact DNA. Lane 16: A-sequencing lane. Lane 17: G-sequencing lane. Representative isotherm for polyamide binding and the calculated K_a value are shown. An additional binding site consisting of the sequence 5'-TGGTCA-3' is contained in the pUC19 plasmid and indicated by the asterisk above.

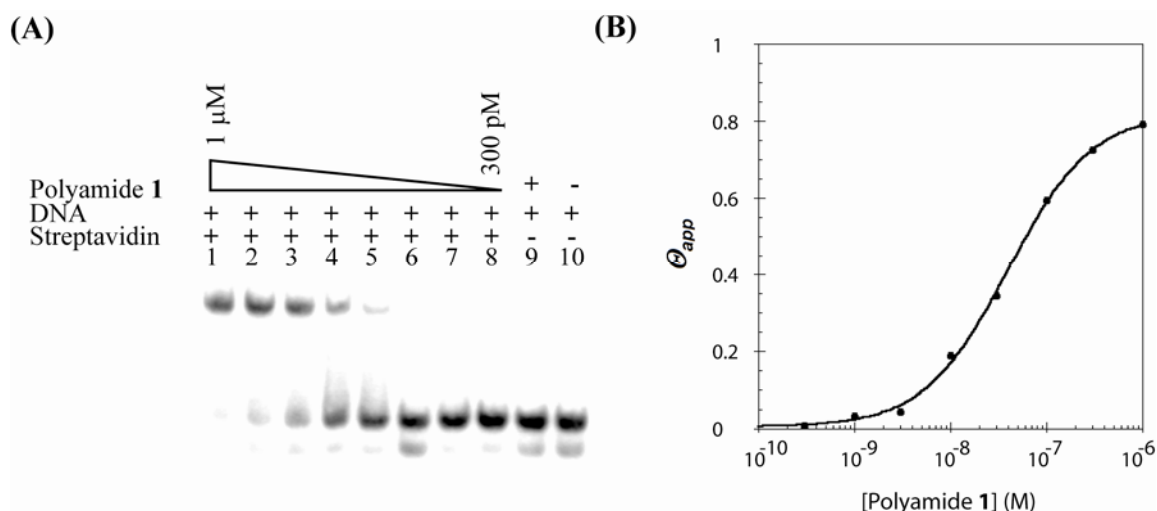


Figure 2.7 Electrophoretic Mobility Shift Assay of **1**. (a) EMSA for polyamide **1** in the presence and absence of 10 μ M streptavidin. Lanes 1-8: 1 μ M, 300 nM, 100 nM, 30 nM, 10 nM, 3 nM, 1 nM, 300 pM polyamide **1**. Lane 9: Control containing 1 μ M polyamide **1** and DNA. Lane 10: Control containing only DNA. (b) Isotherm for streptavidin recruitment by polyamide **1**.

Atomic Force Microscopy of DX arrays with Polyamide-Biotin conjugates and Streptavidin

We then examined whether these complexes would be stable and visible using atomic force microscopy in the context of an **AB** type DX array. A new DX tile **B** which does not contain a match site for the polyamide was designed to form an **AB**-type array with the previously studied DX tile **A**. In our experiments we combined tile **A** and tile **B** to form an **AB** array in which every other tile has a polyamide binding site. As shown in Figure 2.8a, the DX tiles were capable of forming well-defined arrays when visualized by AFM.

Addition of polyamide **1** alone or streptavidin alone did not affect the formation or appearance of the DNA arrays. However, addition of polyamide-biotin conjugate **1** and streptavidin led to recruitment of the streptavidin to the **AB** DX array. (Figure 2.8cd) Incubation of polyamide **1** with tile **A** prior to array formation as opposed to after array formation, led to identical results indicating that order of addition is not critical. The

streptavidin molecules align with a regular spacing occurring between them as expected. The distance between alternating tile **A** in the **AB** array is expected to be 25.2 nm. (Figure 2.1) A section analysis shows that the average spacing between individual streptavidin molecules observed is $24.1 (\pm 1.6)$ nm, in agreement with the expected distance. (Figure 2.8e) For the purposes of this study it appears polyamide binding is relatively non-disruptive to array formation and stability.

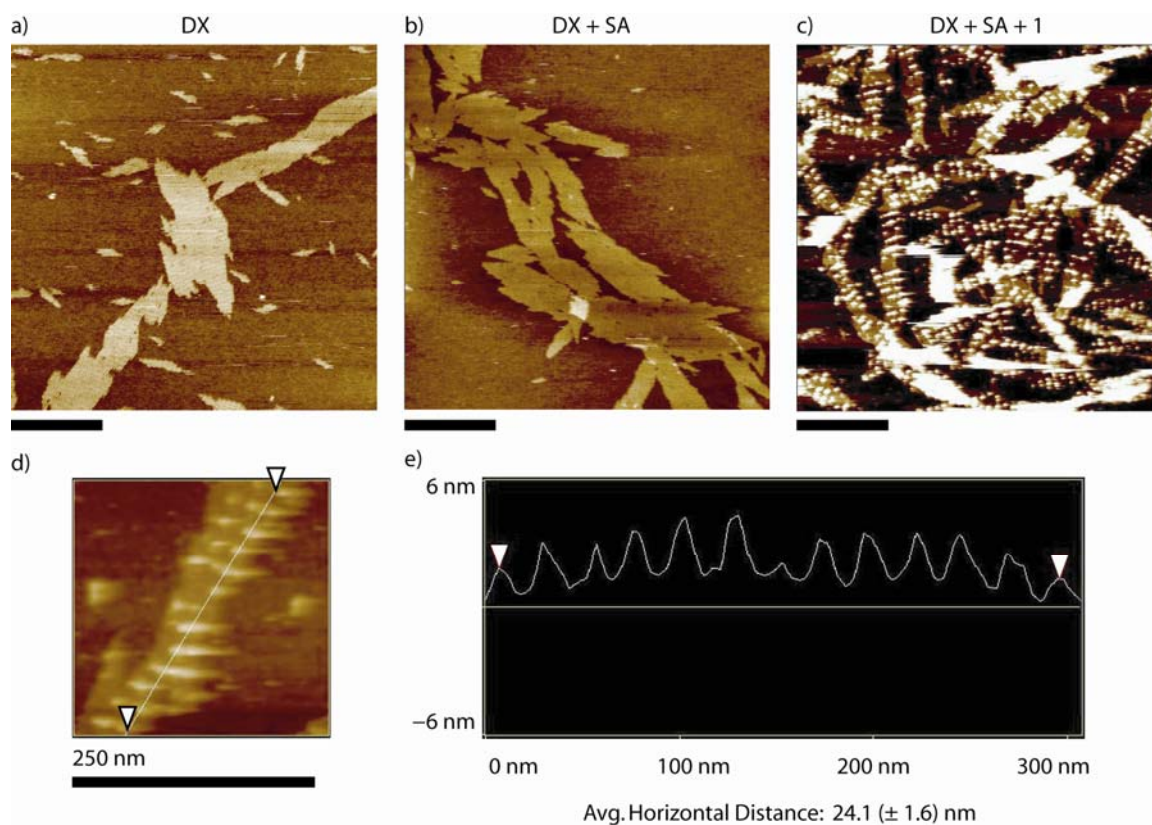
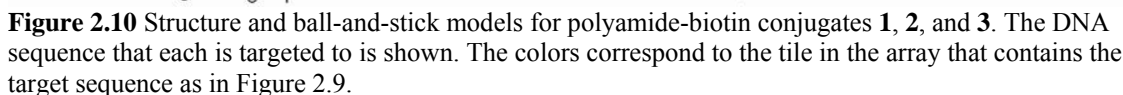
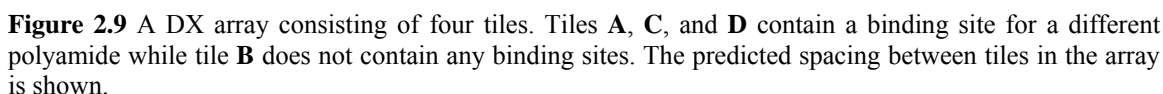


Figure 2.8 AFM images of combined DX tiles **A** and **B**. Scale bar is equal to 250 nm length for all images. (a) 100 nM DX array. (b) 100 nM DX array + 200 nM streptavidin. (c) 100 nM DX array + 200 nM streptavidin + 100 nM **1**. (d) Close up image of DX array with **1**. (e) Height profile along the path indicated in d. The average horizontal distance between peaks is shown.

Programmable Binding to Unique sites on an ABCD DX Array

To demonstrate how different polyamides could be used to target separate sequences in a DNA nanostructure, a new periodic array consisting of four DX tiles was designed as shown in Figure 2.10. Tiles **A**, **C**, and **D** were designed to contain a single



as discussed in the previous sections is designed to target 5'-WGGWCW-3',²⁸ polyamide **2** targets 5'-WTWCGW-3',³¹ and polyamide **3** targets 5'-WGWGCW-3'.²⁷

The four tile DX-**ABCD** was assembled and found by AFM to give 2-D arrays several μm long. Polyamides **1**, **2**, and **3** were then individually incubated with **ABCD** DNA array for 1 hr prior to AFM imaging. The concentration of DNA was 100 nM and the concentration of polyamide used was 200 nM for **1** and **3**, and 150 nM for **2** which showed a propensity to bind at additional sites at higher concentrations. Samples were diluted in half, 5 μL was spotted on mica and allowed to absorb for 1 min, followed by the addition of 2 μL of streptavidin (1 μM) for 1 min. As shown in Figure 2.11,

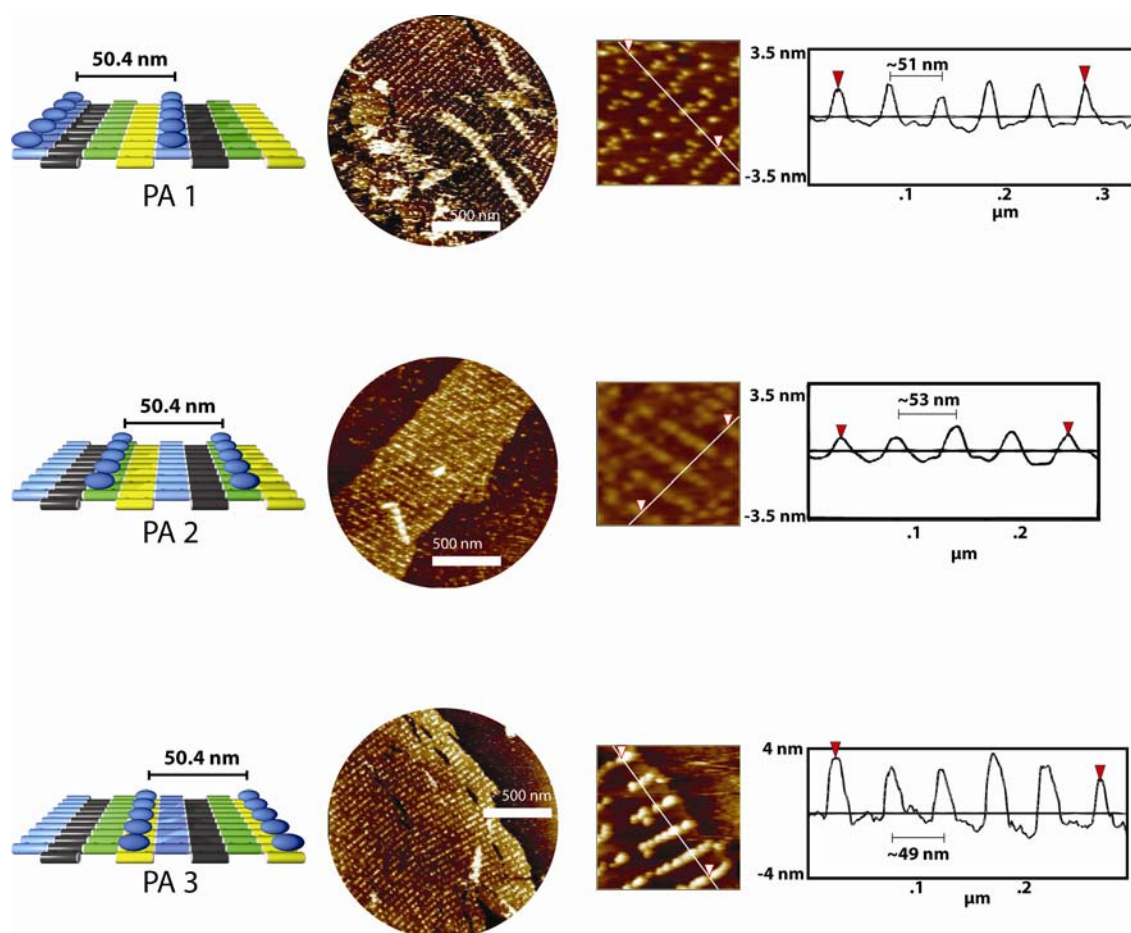


Figure 2.11 AFM images for individual polyamides incubated with **DX-ABCD** and streptavidin. Section analysis shows the height along the indicated path in the square image.

streptavidin recruitment is observed in all three cases. The average spacing for **1**, **2**, and **3** is ~51 nm, ~53 nm, and ~49 nm respectively, corresponding to a spacing of four tiles as expected in each case. As shown, these conjugates are capable of targeting unique locations on the DX array with high specificity and affinity.

Creating Unique Protein Patterns on an Individual DNA Template

In order to demonstrate the ability to target multiple sites on the array simultaneously, we incubated the array with both polyamides **1** and **3**. Since these polyamides target adjacent tiles **A** and **D** on the array, they should give “double-wide” columns of streptavidin. As shown in Figure 2.12, this is what we observe. The average spacing from the center of two adjacent peaks to the next two peaks is ~53 nm, similar to what was observed in the case of the individual polyamides and in agreement with what we would predict. We next incubated our array with polyamides **1** and **2** as before. In this case, every other tile in the array is targeted, so we expect to see stripes with a spacing of exactly half that previously observed for the individual polyamides. We observe a spacing of ~25 nm. As a final experiment, we incubated our arrays with a combination of polyamides **1**, **2**, and **3**. We observe binding at three sites, with a single unbound site between them as would be predicted. The average spacing between unoccupied sites is ~47 nm corresponding to a spacing of every four tiles. This shows the ability to not only simultaneously target multiple DNA sequences in an array, but also to create unique protein patterns on the same DNA template.

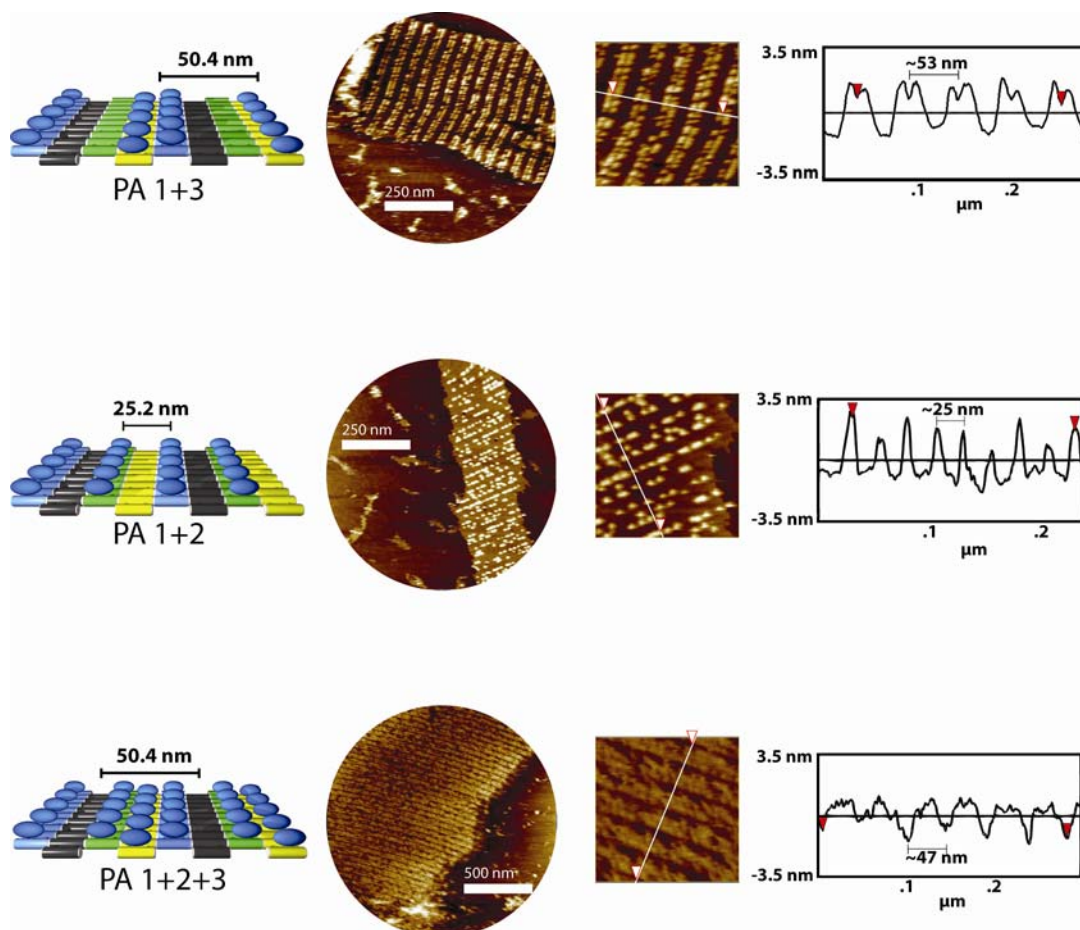


Figure 2.12 AFM images of DX-ABCD incubated with combinations of polyamides 1, 2, and 3. Section analysis shows the height along the indicated path in the square image.

Targeting DNA Nanostructures with DNA-binding Proteins

The previous sections demonstrated the ability of polyamides to effectively bind to specific sequences in DNA nanostructures and to create novel protein patterns on a DNA template. A natural question that arose was whether DNA nanostructures could be targeted directly via a DNA-binding protein. AFM imaging of proteins bound to DNA has been demonstrated for several proteins including MutS and Sp1.³²⁻³⁵ As previously discussed, the incorporation of DNA aptamers into DNA nanostructures has been used to recruit proteins to the assembled surface.³⁶ In this case, the molecular recognition is programmed into the DNA, and not performed by the protein. To date, one of the only

examples of recognition of a 2-dimensional DNA array by a protein has been the use of RuvA which naturally binds to Holliday junctions.³⁷ It was shown that upon addition of RuvA to an array made up of four-armed immobile Holliday Junctions, the structure of the array switched from a Kagome lattice into a square-planar lattice, consistent with the known architectural role of RuvA during branch migration.

Zinc finger proteins are well-known for their ability to bind to DNA,³⁸⁻⁴² and have affinities and specificities similar to polyamides.⁴³ We decided to examine whether the zinc finger binding protein EGR-1 would be capable of binding to a DX array that contained its target binding sequence. EGR-1 is a 543 amino acid, 57.5 kDa protein that contains three zinc finger domains and binds to the 9 bp sequence 5'-GCGTGGGCG-3'.

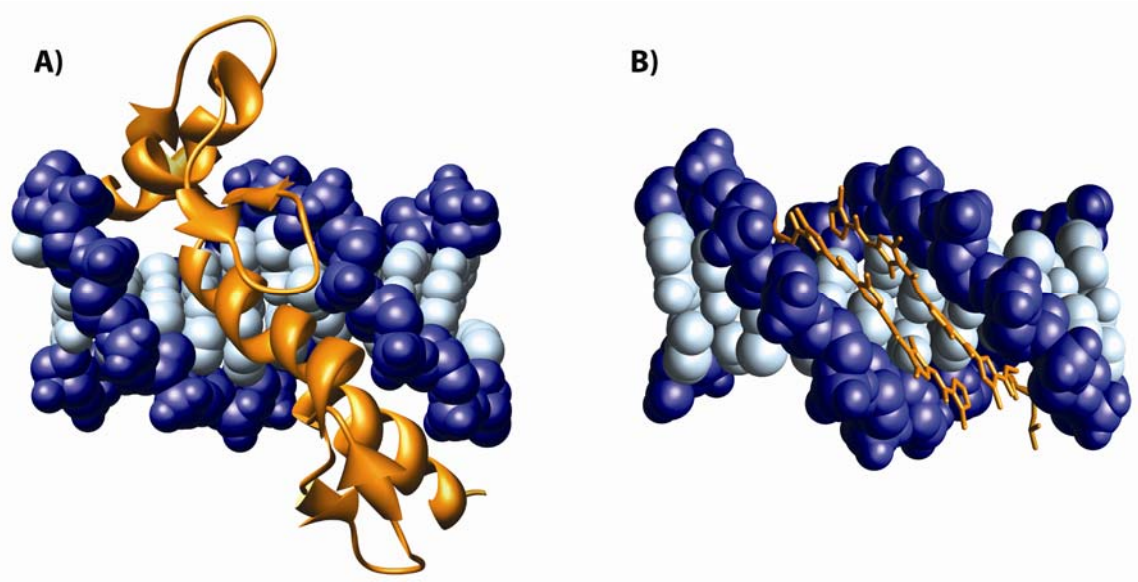


Figure 2.13 Crystal structures of zinc finger protein Zif268 and a polyamide bound to DNA. A) The zinc finger protein Zif268 binds in the major groove of DNA. Zif268 is colored orange. B) A polyamide bound in the minor groove of DNA. The polyamide is colored orange. (PDB codes 1AAY and 1CVY)

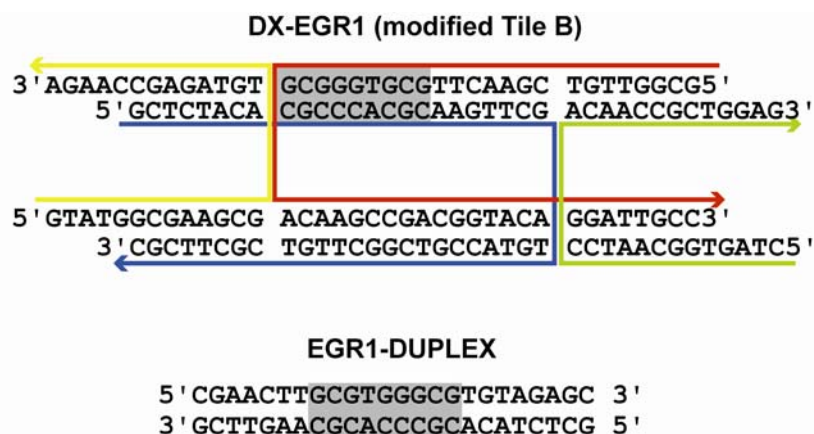
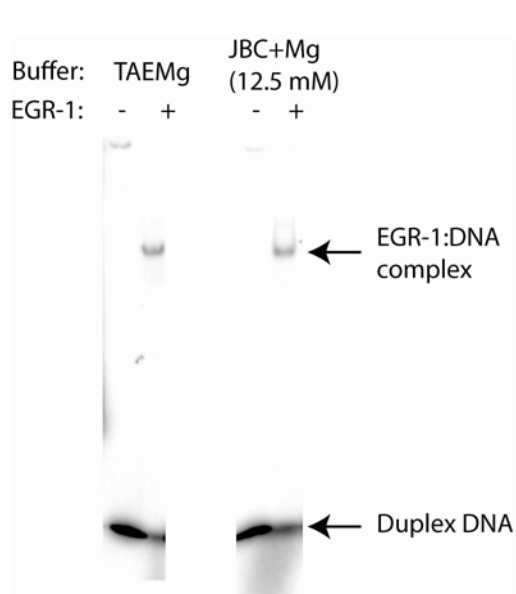


Figure 2.14 Tile DX-EGR1 and EGR1-DUPLEX. The tile was designed by modifying tile **B** of the **ABCD** array. A duplex containing the same sequence as DX-EGR1 was also constructed as a positive control for gel shift experiments.

To examine whether EGR-1 could bind to a DX-array, a new tile containing its 9-bp binding site was created. As shown in Figure 2.14, the binding site was placed in a location where the major groove was deemed most likely to be accessible for binding. The binding site 5'-GCGTGGGCG-3' was incorporated into Tile **B** from the **ABCD** DX



array so that the ability to perform recruitment on tiles **A**, **B** and **D** with polyamides would be retained. A duplex containing the identical binding sequence was also used as a positive control for gel shift experiments.

To first examine whether EGR-1 could bind to a DX-array, a series of gel shift experiments were done. EGR-1 was incubated with radiolabeled DNA EGR1-DUPLEX for 1

Figure 2.15 Gel shift assay for EGR-1 binding to EGR1-DUPLEX DNA. A shift in mobility showing the formation of the EGR-1:DNA complex is observed.

hr. TAEMg buffer (40 mM Tris pH 7.5, 1 mM EDTA, 12.5 mM $\text{Mg}(\text{OAc})_2$, 20 mM acetic acid), used in the previous sections for the formation of DX arrays, as well as a set of published gel shift buffer conditions (JBC Buffer: 10 mM Tris pH 7.5, 1 mM EDTA, 50 mM NaCl)⁴⁴ were used in these assays. The JBC buffer when noted was also supplemented with 12.5 mM $\text{Mg}(\text{OAc})_2$ as magnesium is considered necessary for DNA nanostructure formation. 1 mM DTT, 20 μM ZnSO_4 , and 5% glycerol were also added. After incubation, the sample was run on a .5 \times TBE gel. The results are shown in Figure 2.15. As shown, the addition of EGR-1 to the samples resulted in a clear shift indicative of the protein:DNA complex forming. Analogous experiments using a 1 \times TAEMg gel for electrophoresis gave similar results.

Gel shifts were next attempted on the DX-EGR1 tile. The gel shifts were performed as before except the labeled DX tile was used instead of the duplex. These results are shown in Figure 2.16. As shown, a clear shift was observed when JBC buffer

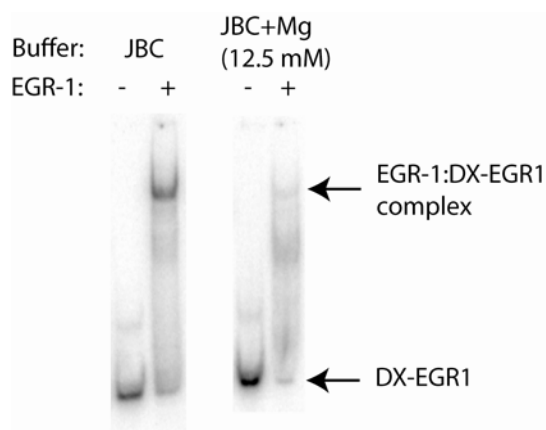


Figure 2.16 Gel shift for EGR-1 binding to DX-EGR1. The formation of the protein:DNA complex is partially inhibited by the addition of 12.5 mM $\text{Mg}(\text{OAc})_2$.

was used. The addition of magnesium to the buffer resulted in diminished binding although some complex was still observed.

Next, AFM experiments were done to examine whether EGR-1 would be capable of binding to its site in the context of a 2-dimensional array. Tile-EGR1 was incubated

DX-EGR1 ABCD array

Image: TAEMg Buffer

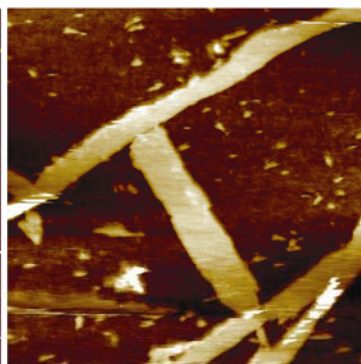
10 μm 2 μm

Image: JBC+Mg (12.5 mM) Buffer

10 μm **DX-EGR1 ABCD Array + EGR-1 protein**

Image: JBC+Mg (12.5 mM) Buffer

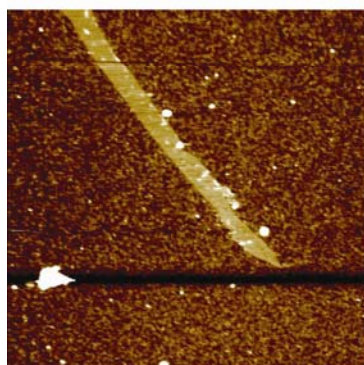
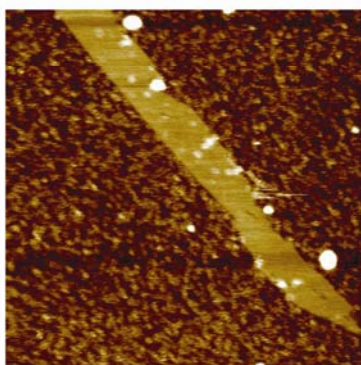
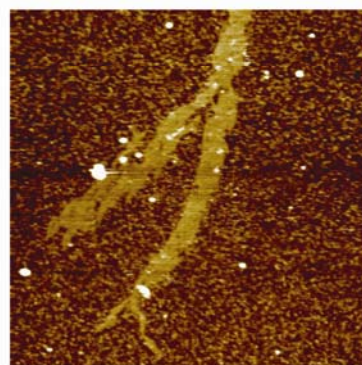
3.5 μm 1.7 μm 2.5 μm

Figure 2.17 AFM images of EGR-1 and DX arrays. Incubation with EGR-1 protein led to very low levels of possible recruitment on the arrays.

with tiles **A**, **C**, and **D** to form a four tile array. The arrays formed in both TAEMg and JBC+Mg buffer, although the long narrow nature of the observed arrays is associated with the formation of nanotubes. The formation of nanotubes is thought to be a result of slight differences in the overall curvature of the DX tiles. As expected, JBC buffer without the addition of magnesium did not result in array formation. The arrays were incubated with EGR-1 and the results are shown in Figure 2.17. As shown, there was very little recruitment of EGR-1 to the arrays. In addition, a large amount of background

protein binding was observed. This led us to re-examine the source of EGR-1 protein which had been used for these studies. The EGR-1 sample purchased from Axxora was a partially purified protein extract. To determine the purity of the protein, non-denaturing polyacrylamide gel electrophoresis followed by Coomassie Blue staining was done. As seen in Figure 2.18, a large number of additional proteins are present in the sample. The presence of so many additional proteins is the probable cause for the nonspecific protein binding to the mica during the AFM experiments.

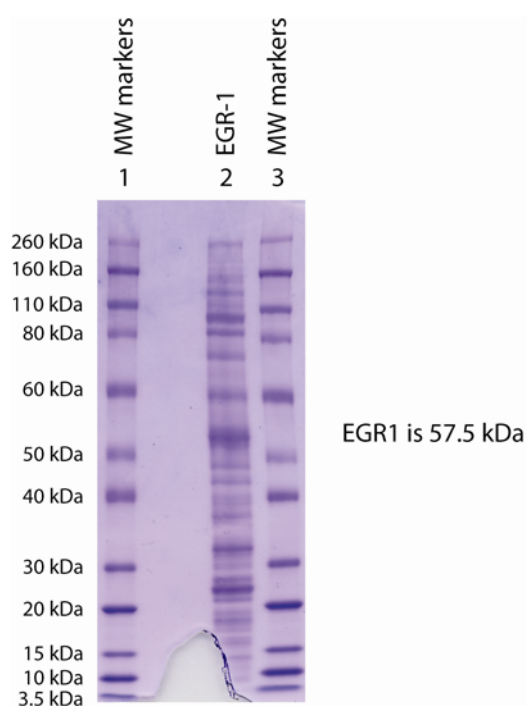


Figure 2.18 Non-denaturing Polyacrylamide Gel Electrophoresis of EGR-1 Protein Sample. Lanes 1 and 3: Molecular Weight Markers. Lane 2: EGR-1. A large number of proteins are present in the sample.

2.3 Conclusions

In our experiments, a polyamide-biotin conjugate was able to recruit streptavidin molecules to one tile (**A**) of an **AB** DX array, effectively labeling individual tiles. In addition, we have demonstrated that polyamide-biotin conjugates are capable of

As a final set of experiments, a purified fusion protein, containing the 100 amino acid C-terminal DNA-binding portion of EGR-1 fused to Glutathione-S-transferase (GST) was purchased from Abnova. However, the fusion protein failed to function in either the AFM or gel shift assays. As such, it remains unclear whether the fusion protein is still functional for DNA binding.

addressing specific elements in a multi-component DX array. We are able to organize streptavidin molecules into distinctly different patterns on the same DNA template, solely by changing the polyamide cores that are used. The synthetic ease in creating polyamides, and the existence of a library of well characterized solutions to target a wide variety of DNA sequences makes these conjugates ideal for arranging proteins at a variety of sites. In addition, the ability to address an array without requiring prior covalent modification increases the flexibility and usefulness of DNA templates, as they could serve to make a variety of increasingly complex assemblies. Nanoparticles, proteins, or other biomolecules could be conjugated to or recruited by polyamides, allowing them to be targeted to specific tiles in an array. Polyamide conjugates would act as a sequence specific glue or molecular staple that allows the self-assembly properties of DNA to be transferred to complicated functional molecular assemblies.

The work using the DNA-binding protein EGR-1 illustrates several of the challenges involved in creating complex assemblies. The successful gel shift experiments using a single DX tile show that EGR-1 can recognize and bind to its target sequence when embedded in a DNA nanostructure. The small amount of protein recruitment observed in the AFM experiments, and the successful gel shift indicate that this approach towards targeting DNA nanostructures has merit. Future experiments will be needed to determine whether the protein can still access its binding site in the context of the fully formed 2-dimensional array. It is worth noting that the Yan group, in addressing sites on a DNA origami nanostructure found distinct differences in binding affinity at different locations on the nanostructure's surface.⁴⁵ Locations nearest the edges were significantly better for binding than sites located towards the center of the structure. It will be

interesting to determine whether this is the case for proteins binding to DNA arrays as well. Given the relatively large size of proteins when compared to a DNA oligomer or a polyamide, it is likely that proteins may have even greater positional preferences. One hypothesis that could explain the lack of binding in the AFM studies is that the protein is too sterically hindered to access to the surface of the DX when it is incorporated into a 2-dimensional array but is not restricted in the gel shift as it is only binding to a single tile in solution.

Future experiments are needed to ascertain whether the impurities in the sample were responsible for the lack of binding in the AFM experiments. It is possible that the presence of a large number of other proteins could have interfered with binding. In addition, the concentration of EGR-1 in the samples is not known. It may be that the protein is not present in sufficiently high concentrations to bind to all of the sites present in the array. Future experiments using purified EGR-1 would eliminate the difficulties that occur when using a cellular extract and have the advantage of knowing the amount of protein being added to each sample.

It is interesting to compare the successful use of polyamides for targeting DNA nanostructures when compared to the difficulties experienced with EGR-1. The purity issues that were encountered with EGR-1 are non-existent for polyamides which can be rapidly synthesized using solid phase methods and purified by HPLC. The ability to rapidly generate polyamides to target virtually any desired DNA sequence is also a considerable advantage. While a large number of zinc finger motifs to target a variety of sequences are known,^{38, 39, 42} the expression and purification of these constructs is non-trivial. Polyamides were functional in all of the buffers examined including those used for

DNase footprinting, affinity cleavage, and AFM studies. In contrast, many proteins are highly sensitive on buffer conditions for proper folding and function. The need for high magnesium concentrations to stabilize DX and other nanostructures may prove to be problematic for the use of DNA-binding proteins as the addition of 12.5 mM Mg(OAc)₂ in the gel shift buffer was detrimental to EGR-1 binding. Finally, it should be noted that the smaller size of polyamides when compared to proteins may allow them to bind to their target sites in a 2-dimensional array whereas proteins may be unable to access more sterically restricted sites. DNA nanostructures, such as nanogrids, that contain significantly more spacing between the DNA elements in the array, may be superior to DX arrays, which are tightly packed in comparison, for these purposes.

In conclusion, the experiments with EGR-1 highlight both the challenges involved in creating complex assemblies at the molecular level, and also may of the advantages of our polyamide based approach. Future work towards creating assemblies should benefit from the approaches investigated here for the molecular recognition of DNA nanostructures.

2.4 Experimental Details

Abbreviations. *N,N*-dimethylformamide (DMF), *N,N*-diisopropylethylamine (DIEA), *rac*-dithiothreitol (DTT), *N*-[2-Hydroxyethyl]piperazine-*N'*-[2-ethanesulfonic acid] (HEPES), 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), trifluoroacetic acid (TFA), Trishydroxymethylaminomethane (TRIS), ethylenediaminetetraacetic acid (EDTA).

Materials. Boc- β -Ala-PAM resin was purchased from Peptides International.

Trifluoroacetic acid (TFA) was purchased from Halocarbon. Methylene Chloride (DCM) was obtained from Fisher Scientific and *N,N*-dimethylformamide (DMF) from EMD. Ethylenediaminetetraacetic dianhydride was purchased from Aldrich. EZ-Link TFP-PEO₃-Biotin was purchased from Pierce. Streptavidin was purchased from Rockland. EGR-1 protein extract was purchased from Axxora. The recombinant protein EGR-1-GST was purchased from Abnova. All DNA oligonucleotides were purchased from Integrated DNA Technologies. DTT and Ultra Pure TRIS were purchased from ICN. Magnesium Acetate 4-hydrate was obtained from J.T. Baker. Ferrous Ammonium Sulfate was purchased from Fisher Scientific. Magnesium Chloride 6-Hydrate was obtained from Mallinckrodt. [γ -³²P]-adenosine-5'-triphosphate (≥ 7000 Ci/mmol) was obtained from MP Biomedicals. Calf thymus DNA was from Amersham and all enzymes were obtained from Roche. Water (18 M Ω) was purified using a aquaMAX-Ultra water purification system. Biological experiments were performed using Ultrapure Water (DNase/RNase free) purchased from USB. The pH of buffers was adjusted using a Thermo Orion 310 PerpHect Meter. All buffers were sterilized by filtration through a Nalgene 0.2 μ m cellulose nitrate filtration.

Polyamide Synthesis. Polyamide monomers were prepared as described previously.⁴⁶

Synthesis was performed using established protocols and all polyamides were characterized by MALDI-TOF and analytical HPLC.

1-EDTA: (MALDI-TOF-MS) [M+H]⁺ calc. for C₆₉H₉₀N₂₅O₁₇⁺ 1540.7, observed 1540.6

1: (MALDI-TOF-MS) [M+H]⁺ calc. for C₈₃H₁₁₇N₂₈O₁₇S⁺ 1809.9, observed 1810.0

2: (MALDI-TOF-MS) $[M+H]^+$ calc. for $C_{83}H_{114}ClN_{26}O_{17}S_2^+$ 1846.8, observed 1846.5

3: (MALDI-TOF-MS) $[M+H]^+$ calc. for $C_{83}H_{117}N_{28}O_{17}S^+$ 1809.9, observed 1810.2

Preparation of Labeled DX Tiles. In all cases 5' radiolabeling of 60 pmol of a single DNA strand was done using Polynucleotide Kinase. The labeled strand was then added to the three unlabelled strands in Affinity Cleavage Buffer (20 mM HEPES, 60 mM NaCl, 62.5 mM $MgCl_2$, pH 7.3) and a total volume of 20 μ L. Several samples were made to titrate the labeled strand against the unlabelled strands in order to maximize formation of the four strand DX complex. The strands were heated to 95°C for 10 min and then allowed to cool slowly to room temperature over several hours. Purity was assessed by running 1 μ L of the sample on a 6% polyacrylamide gel. The gel was then dried and exposed to a phosphor screen which was visualized using a Molecular Dynamics 400S Phosphorimager. In all cases the DX tile that was used was greater than 90% pure.

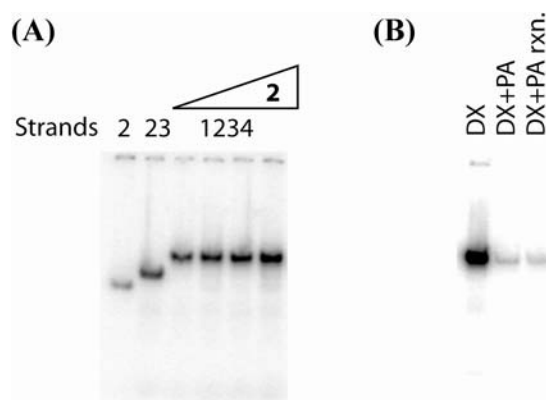


Figure 2.19 Formation and stability of DXs. (a) Representative native gel showing the formation of the four stranded DX-In tile. Lane 1: strand 2. Lane 2: strands 2 and 3. Lanes 3-6: .7, .9, 1.0, and 1.2 equivalents of strand 2 with 1 equivalent of strands 1, 3, and 4 (b) Representative native gel showing stability of the fully formed complex. Lane 1: DX-In. Lane 2: DX-In and 1 μ M polyamide **1-EDTA** in affinity cleavage buffer. Lane 3: DX-Up and 1 μ M polyamide **1-EDTA** in affinity cleavage buffer after the cleavage reaction has occurred.

Affinity Cleavage on DX complexes. Affinity Cleavage experiments were done following previously established protocols.⁴⁷ The total reaction volume was 50 μ L. The polyamide was allowed to equilibrate with DX (12,000 cpm / lane) for one hour in Affinity Cleavage Buffer (see above). A solution of freshly prepared ferrous ammonium sulfate was added to a final concentration of 1 μ M and allowed to sit for 30min. To initiate the reaction, DTT was added to a final concentration of 5 mM. The reaction was allowed to proceed for 40min and stopped by the addition of 1.25 μ L of precipitation buffer (Glycogen 2.8mg / ml, 140 μ M bp calf thymus DNA). The DNA was then isolated by ethanol precipitation and run on an 8% polyacrylamide denaturing gel.

Electrophoretic Mobility Shift Assay (EMSA). For the polyamide:streptavidin EMSA experiments 60 pmol of DNA strands DUPLEX1 and DUPLEX2 were first annealed by heating to 95°C for 10 min and allowing them to cool slowly to room temperature over several hours. Using Polynucleotide Kinase they were then 5' radiolabeled. The labeled DNA (3,000 cpm/lane) was then incubated with polyamide for 1 hr in TAEMg buffer followed by incubation with streptavidin for 30 min. The total volume for the reaction was 10 μ L which was then run on a 6% polyacrylamide gel and imaged. The EGR-1 experiments were performed using a modified version of a reported protocol.⁴⁴ First, DX-EGR1 and EGR1-DUPLEX were radiolabeled as described above. The labeled DNA (5,000 cpm/lane) was then incubated with 1 band forming unit of EGR-1 protein extract, 1 mM DTT, 20 μ M ZnSO₄, and 5% glycerol for 1 hour in either TAEMg or JBC (10 mM Tris pH 7.5, 1 mM EDTA, 50 mM NaCl) buffer. In certain noted cases 12.5 mM Mg(OAc)₂ was also added to the JBC buffer. The reaction volume was 20 μ L. After

incubation, 10 μ L of each sample was run on a 6% acrylamide .5X TBE gel for 1.5 hr at 180 V and imaged.

Affinity determination by quantitative DNase I footprinting. Reactions were carried out in a volume of 400 μ L in aqueous TKMC buffer according to published protocols.⁴⁷ Standard molecular biology techniques were used to insert a 75bp DNA sequence into the *Bam*HI/*Hind*III restriction site of pUC19.⁴⁸ This plasmid was used to generate a 5' ³²P labeled 283bp *Eco*RI/*Pvu*II restriction fragment which was used for all footprinting experiments. Developed gels were imaged using storage phosphor autoradiography using a Molecular Dynamics 400S Phosphorimager. Equilibrium association constants were determined as previously described.⁴⁷

AFM Sample Preparation. Individual DX tiles **A** and **B** were first annealed by mixing equimolar amounts of each of four strands in TAEMg Buffer (40 mM Tris-HCl (pH 8.0), 20 mM acetic acid, 1 mM EDTA, 12.5 mM magnesium acetate) and heating to 95°C for 10 min at a concentration of 200 nM. They were then allowed to cool slowly over several hours to room temperature. Equal amounts of DX tile **A** and **B** were then mixed at a final concentration of 100 nM. Samples were heated to 50°C and allowed to cool slowly over 10 – 12 hr to room temperature. Polyamide **1** was added to the solution of DX arrays and allowed to equilibrate for 1 hr at a concentration of 100 nM, followed by addition of streptavidin to a final concentration of 200 nM which was incubated with the sample for 30 min. In certain noted cases the polyamide was incubated with a single tile prior to the 50°C annealing step, rather than with the already formed array. 5 μ L of the sample was

spotted on freshly cleaved mica and allowed to absorb for 1 min. Imaging was done using a DI Multimode Atomic Force Microscope.

For **DX-ABCD** the array was formed in two steps. Individual tiles **A**, **B**, **C**, and **D** were first annealed by mixing equimolar amounts of each of the four input strands at 1 μ M concentration in TAEMg Buffer. Each sample was heated to 95°C for 10 min and allowed to cool slowly over several hours to room temperature. The four tiles were then combined, heated to 45°C and allowed to cool to RT over approximately 12 hrs. For all of the experiments, polyamide was incubated with **DX-ABCD** (100 nM) for 1 hr prior to imaging. The polyamide concentration was 200 nM for **1** and **3**, and 150 nM for **2**. After the incubation, the sample was diluted in half which led to cleaner AFM images. 5 μ L of sample was spotted on freshly cleaved mica and allowed to absorb for 1 min. 2 μ L of 1 μ M Streptavidin was then added to the sample for 1 min. Calibration for the distance measurements was done by averaging four measurements of individual tiles in an untreated **ABCD** array. The reported spacings for each of the polyamides with streptavidin were calculated as the average distance between peaks in the graphs shown in Figures 11 and 12.

The **DX-EGR1** tile was formed as described above. It was incorporated into the **ABCD** array in the same manner with the sole exception being that Tile **B** was replaced with **DX-EGR1**. The annealed array (67.5 nM) was incubated in either TAEMg or JBC+Mg buffer with 20 μ M ZnSO₄ and 1 band forming units of EGR-1. The sample was then diluted five fold as the protein concentration was too high to obtain good images. 2.5 μ L of the diluted sample was spotted on the mica and imaged.

Denaturing PAGE. All buffers and reagents used were purchased from Invitrogen and the manufacturer's protocols were followed. Briefly, the sample was prepared using 4 μ L of EGR-1, 1 μ L of NuPAGE Reducing Agent, 2.5 μ L of 4X NuPAGE LDS Sample Buffer and 2.5 μ L of water. Novex Sharp Protein Standard (prestained) 3.5 kDa – 260 kDa molecular weight standards from Invitrogen (VXLC5800) were used. The sample was heated at 70°C for 10 min prior to loading and then run on a Bis-Tris 4-12% Gel with MOPS running buffer. 500 μ L of NuPAGE Antioxidant was added to the inner buffer chamber and the gel was run at 200 V for 25 min. Coomassie staining was done using a staining solution of .1% Coomassie R-250 in 40% ethanol and 10% acetic acid. The gel was placed in 100 mL of staining solution and microwaved for 1 min and then placed on a shaker for 15 min. The gel was then rinsed with water and placed in 100 mL of destain solution containing 10% ethanol and 7.5% acetic acid. It was again placed in the microwave for 1 min and then on a shaker for 15 min. The destain procedure was repeated a second time using fresh destain solution to obtain contrast.

Additional DNA Oligomers. The majority of the DNA oligomers used are described in Figures 2.1, 2.5, 2.9 and 2.14. The sequences used for DX-Junction which did not show any binding in the affinity cleavage assays, and for the duplex used in the EMSA experiments with polyamide **1** and streptavidin are shown below:

DX-Junction-1: 5' - TCACTCTACCGCACGAGAATGGAGAT - 3'

DX-Junction-2: 5' – CATTCTCGACGCCAATAGTTTGCACCTAACTTCATGTGCCT GCGGTAG - 3'

DX-Junction-3: 5'-CTGTAGCCTGCAAACCTATTGGCGTGGCACATGAAGTTAGGA
CAGATCG - 3'

DX-Junction-4: 5' - CATACCGATCTGTGGCTACAGTCTTG - 3'

DUPLEX1: 5' - CATTCTCGACGCTAGGTCACAGCAGGCTACTGTCTTG - 3'

DUPLEX2: 5' - CAGTAGCCTGCTGTGACCTAGCGTCGAGAATGGAGAT - 3'

Acknowledgement. We thank Prof. Erik Winfree and Dr. Sung Ha Park for useful discussions and assistance with the AFM experiments. We are grateful for support from the Ralph M Parsons Foundation and W. Webster.

2.5 References

1. Lin, C. X.; Liu, Y.; Rinker, S.; Yan, H., DNA tile based self-assembly: Building complex nanoarchitectures *ChemPhysChem* **2006**, 7, 1641-1647.
2. Seeman, N. C.; Lukeman, P. S., Nucleic acid nanostructures: Bottom-up control of geometry on the nanoscale *Rep. Prog. Phys.* **2005**, 68, 237-270.
3. Aldaye, F. A.; Palmer, A. L.; Sleiman, H. F., Assembling materials with DNA as the guide *Science* **2008**, 321, 1795-1799.
4. Seeman, N. C., DNA in a material world *Nature* **2003**, 421, 427-431.
5. Fu, T. J.; Seeman, N. C., DNA double-crossover molecules *Biochemistry* **1993**, 32, 3211-3220.
6. Winfree, E.; Liu, F. R.; Wenzler, L. A.; Seeman, N. C., Design and self-assembly of two-dimensional DNA crystals *Nature* **1998**, 394, 539-544.
7. Constantinou, P. E.; Wang, T.; Kopatsch, J.; Israel, L. B.; Zhang, X. P.; Ding, B. Q.; Sherman, W. B.; Wang, X.; Zheng, J. P.; Sha, R. J.; Seeman, N. C., Double cohesion in structural DNA nanotechnology *Org. Biomol. Chem.* **2006**, 4, 3414-3419.
8. Park, S. H.; Pistol, C.; Ahn, S. J.; Reif, J. H.; Lebeck, A. R.; Dwyer, C.; LaBean, T. H., Finite-size, fully addressable DNA tile lattices formed by hierarchical

- assembly procedures (vol 45, pg 735, 2006) *Angew. Chem., Int. Ed.* **2006**, 45, 6607-6607.
9. Lund, K.; Liu, Y.; Yan, H., Combinatorial self-assembly of DNA nanostructures *Org. Biomol. Chem.* **2006**, 4, 3402-3403.
 10. Rothemund, P. W. K., Folding DNA to create nanoscale shapes and patterns *Nature* **2006**, 440, 297-302.
 11. He, Y.; Chen, Y.; Liu, H. P.; Ribbe, A. E.; Mao, C. D., Self-assembly of hexagonal DNA two-dimensional (2D) arrays *J. Am. Chem. Soc.* **2005**, 127, 12202-12203.
 12. He, Y.; Tian, Y.; Ribbe, A. E.; Mao, C. D., Highly connected two-dimensional crystals of DNA six-point-stars *J. Am. Chem. Soc.* **2006**, 128, 15978-15979.
 13. Aldaye, F. A.; Sleiman, H. F., Modular access to structurally switchable 3D discrete DNA assemblies *J. Am. Chem. Soc.* **2007**, 129, 13376-13377.
 14. He, Y.; Ye, T.; Su, M.; Zhang, C.; Ribbe, A. E.; Jiang, W.; Mao, C. D., Hierarchical self-assembly of DNA into symmetric supramolecular polyhedra *Nature* **2008**, 452, 198-201.
 15. Le, J. D.; Pinto, Y.; Seeman, N. C.; Musier-Forsyth, K.; Taton, T. A.; Kiehl, R. A., DNA-templated self-assembly of metallic nanocomponent arrays on a surface *Nano Lett.* **2004**, 4, 2343-2347.
 16. Li, H. Y.; LaBean, T. H.; Kenan, D. J., Single-chain antibodies against DNA aptamers for use as adapter molecules on DNA tile arrays in nanoscale materials organization *Org. Biomol. Chem.* **2006**, 4, 3420-3426.
 17. Liu, Y.; Lin, C. X.; Li, H. Y.; Yan, H., Protein nanoarrays - Aptamer-directed self-assembly of protein arrays on a DNA nanostructure *Angew. Chem., Int. Ed.* **2005**, 44, 4333-4338.
 18. Park, S. H.; Yin, P.; Liu, Y.; Reif, J. H.; LaBean, T. H.; Yan, H., Programmable DNA self-assemblies for nanoscale organization of ligands and proteins *Nano Lett.* **2005**, 5, 729-733.
 19. Chhabra, R.; Sharma, J.; Liu, Y.; Yan, H., Addressable molecular tweezers for DNA-templated coupling reactions *Nano Lett.* **2006**, 6, 978-983.
 20. Zhang, J. P.; Liu, Y.; Ke, Y. G.; Yan, H., Periodic square-like gold nanoparticle arrays templated by self-assembled 2D DNA nanogrids on a surface *Nano Lett.* **2006**, 6, 248-251.

21. Williams, B. A. R.; Lund, K.; Liu, Y.; Yan, H.; Chaput, J. C., Self-assembled peptide nanoarrays: An approach to studying protein-protein interactions *Angew. Chem., Int. Ed.* **2007**, 46, 3051-3054.
22. Liu, F. R.; Sha, R. J.; Seeman, N. C., Modifying the surface features of two-dimensional DNA crystals *J. Am. Chem. Soc.* **1999**, 121, 917-922.
23. Yan, H.; Park, S. H.; Finkelstein, G.; Reif, J. H.; LaBean, T. H., DNA-templated self-assembly of protein arrays and highly conductive nanowires *Science* **2003**, 301, 1882-1884.
24. Sharma, J.; Chhabra, R.; Liu, Y.; Ke, Y. G.; Yan, H., DNA-templated self-assembly of two-dimensional and periodical gold nanoparticle arrays *Angew. Chem., Int. Ed.* **2006**, 45, 730-735.
25. Dervan, P. B., Molecular recognition of DNA by small molecules *Bioorg. Med. Chem.* **2001**, 9, 2215-2235.
26. Dervan, P. B.; Edelson, B. S., Recognition of the DNA minor groove by pyrrole-imidazole polyamides *Curr. Opin. Struct. Biol.* **2003**, 13, 284-299.
27. Hsu, C. F.; Phillips, J. W.; Trauger, J. W.; Farkas, M. E.; Belitsky, J. M.; Heckel, A.; Olenyuk, B. Z.; Puckett, J. W.; Wang, C. C. C.; Dervan, P. B., Completion of a programmable DNA-binding small molecule library *Tetrahedron* **2007**, 63, 6146-6151.
28. 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, 468-471.
29. Dervan, P. B., Design of sequence-specific DNA-binding molecules *Science* **1986**, 232, 464-471.
30. Taylor, J. S.; Schultz, P. G.; Dervan, P. B., DNA affinity cleaving - Sequence specific cleavage of DNA by distamycin-EDTA.Fe(II) and EDTA-distamycin.Fe(II) *Tetrahedron* **1984**, 40, 457-465.
31. Olenyuk, B. Z.; Zhang, G. J.; Klco, J. M.; Nickols, N. G.; Kaelin, W. G.; Dervan, P. B., Inhibition of vascular endothelial growth factor with a sequence-specific hypoxia response element antagonist *Proc. Natl. Acad. Sci. U. S. A.* **2004**, 101, 16768-16773.
32. Tessmer, I.; Yang, Y.; Sass, L.; Hsieh, P.; Erie, D. A., AFM and single-molecule fluorescence studies of MutS-DNA interactions reveal the mechanism of mismatch recognition *Environ. Mol. Mutagen.* **2006**, 47, 416-416.

33. Wang, H.; Yang, Y.; Schofield, M. J.; Du, C. W.; Fridman, Y.; Lee, S. D.; Larson, E. D.; Drummond, J. T.; Alani, E.; Hsieh, P.; Erie, D. A., DNA bending and unbending by MutS govern mismatch recognition and specificity *Proc. Natl. Acad. Sci. U. S. A.* **2003**, 100, 14822-14827.
34. Yang, Y.; Sass, L. E.; Du, C. W.; Hsieh, P.; Erie, D. A., Determination of protein-DNA binding constants and specificities from statistical analyses of single molecules: MutS-DNA interactions *Nucleic Acids Res.* **2005**, 33, 4322-4334.
35. Nettikadan, S.; Tokumasu, F.; Takeyasu, K., Quantitative analysis of the transcription factor AP2 binding to DNA by atomic force microscopy *Biochem. Biophys. Res. Commun.* **1996**, 226, 645-649.
36. Lin, C. X.; Katilius, E.; Liu, Y.; Zhang, J. P.; Yan, H., Self-assembled signaling aptamer DNA arrays for protein detection *Angew. Chem.-Int. Edit.* **2006**, 45, 5296-5301.
37. Malo, J.; Mitchell, J. C.; Venien-Bryan, C.; Harris, J. R.; Wille, H.; Sherratt, D. J.; Turberfield, A. J., Engineering a 2D protein-DNA crystal *Angew. Chem., Int. Ed.* **2005**, 44, 3057-3061.
38. Desjarlais, J. R.; Berg, J. M., Use of a zinc-finger consensus sequence framework and specificity rules to design specific DNA-binding proteins *Proc. Natl. Acad. Sci. U. S. A.* **1993**, 90, 2256-2260.
39. Mandell, J. G.; Barbas, C. F., Zinc finger tools: custom DNA-binding domains for transcription factors and nucleases *Nucleic Acids Res.* **2006**, 34, W516-W523.
40. Pavletich, N. P.; Pabo, C. O., Zinc finger DNA recognition - Crystal-structure of a Zif268-DNA complex at 2.1-A *Science* **1991**, 252, 809-817.
41. Shi, Y. G.; Berg, J. M., A direct comparison of the properties of natural and designed zinc-finger proteins *Chem. Biol.* **1995**, 2, 83-89.
42. Wright, D. A.; Thibodeau-Beganny, S.; Sander, J. D.; Winfrey, R. J.; Hirsh, A. S.; Eichinger, M.; Fu, F.; Porteus, M. H.; Dobbs, D.; Voytas, D. F.; Joung, J. K., Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly *Nat. Protoc.* **2006**, 1, 1637-1652.
43. Nguyen-Hackley, D. H.; Ramm, E.; Taylor, C. M.; Joung, J. K.; Dervan, P. B.; Pabo, C. O., Allosteric inhibition of zinc-finger binding in the major groove of DNA by minor-groove binding ligands *Biochemistry* **2004**, 43, 3880-3890.
44. Skerka, C.; Decker, E. L.; Zipfel, P. F., A regulatory element in the human interleukin-2 gene promoter is a binding-site for the zinc-finger proteins Sp1 and Egr-1 *J. Biol. Chem.* **1995**, 270, 22500-22506.

45. Ke, Y. G.; Lindsay, S.; Chang, Y.; Liu, Y.; Yan, H., Self-assembled water-soluble nucleic acid probe tiles for label-free RNA hybridization assays *Science* **2008**, 319, 180-183.
46. Baird, E. E.; Dervan, P. B., Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids *J. Am. Chem. Soc.* **1996**, 118, 6141-6146.
47. Trauger, J. W.; Dervan, P. B., Footprinting methods for analysis of pyrrole-imidazole polyamide/DNA complexes *Methods Enzymol.* **2001**, 340, 450-466.
48. Sambrook, J.; Fritsh, E. F.; Maniatis, T., *Molecular Cloning: Standard Protocols for DNA Manipulation. A laboratory Manual 2nd ed.* Cold Spring Harbor Laboratory: Plainview, NY, 1989.

Chapter 3

Addressing DNA Origami with Polyamide-Biotin Conjugates

Abstract

DNA origami is a unique DNA architecture that can be used to create arbitrary two-dimensional shapes on the nanoscale. As with all DNA nanoarchitectures, the ability to create functional molecular assemblies from this DNA template will be dependent on the ability to recruit active molecules to the DNA surface in a bottom-up approach to self assembly. The ability to target these unique DNA nanostructures with polyamide-biotin conjugates and recruit streptavidin to their surface was examined using atomic force microscopy. Evidence for recruitment at predicted binding sites and an outline for future work is presented.

3.1 Introduction

The ability to create DNA nanostructures containing precise spacing's and shapes is an important requirement for bottom-up self assembly.¹⁻⁴ In this approach, DNA templates are used for the organization of secondary molecular components. To this end, DNA origami has been shown to be an ideal method for the creation of arbitrary 2-dimensional shapes.⁵ As originally demonstrated by Paul Rothemund, DNA origami entails the folding of a long, single-stranded DNA scaffold strand into a variety of shapes by the addition of a large number (typically >200) of short oligonucleotide staple strands. The scaffold strand used is the 7,249 nt long M13mp18 viral DNA strand. The staple strands are typically 32 nt long and can be used without purification. A variety of different shapes have been generated using this method including squares, rectangles, five-point stars, smiley faces, and even a map of China.^{5, 6} The origami itself can be decorated with DNA dumbbells by modifying the staple strands, allowing words such as "DNA" and even pictures to be drawn on the surface of the origami.⁵ DNA origami can also be modified to display single stranded DNA sticky ends on the edges, as well as on the surface.^{7, 8}

For DNA origami to act as a template for increasingly complex assemblies, a method of recruiting active components to the DNA structure is needed. Given the previous success in using polyamides to recruit streptavidin to DX arrays described in the previous chapter,^{9, 10} and the structural similarity between origami and the DX molecule, the ability of polyamide conjugates to functionalize DNA origami was next investigated.

3.2 Results and Discussion

Analysis of Binding Sites on DNA Origami Smiley Face

For this study, the DNA origami smiley face was used as originally described by Rothemund.⁵ Although structurally related, DX arrays and DNA origami have several significant differences. The periodic nature of DX arrays results in the presentation of a relatively small number of DNA sequences. In contrast, assuming that the full scaffold strand is used, over 7,000 base pairs are present in a DNA origami structure, virtually ensuring that multiple binding sites will be present for a standard pyrrole-imidazole eight-ring polyamide that has specificity for 6 base pairs. Similarly, the use of the viral DNA scaffold strand also makes modification of the origami sequence difficult. When designing a DX tile one can ensure that a specific sequence occurs only a single time throughout the array. The ease of synthesis of short oligonucleotides allows one to easily manipulate the sequences in the array to meet this requirement. However, extensive




Polyamide	Sequence	# Sites in Full Sequence	Accessibility
 1	5'-WGGWCW-3'	24	7 sites - Optimal 17 sites - at nick / junction / edge
 2	5'-WTWCGW-3'	22	
 3	5'-WGWGCW-3'	14	7 sites - Optimal 7 sites - at nick / junction / edge

Figure 3.1 Analysis of potential polyamide binding sites on a DNA origami smiley face. Polyamides **1 – 3** are shown along with their target sequences. The number of match sites found throughout the structure is shown, as well as how many of the sites are presumed to be accessible. Sites are deemed optimal unless any of the base pairs are located at a junction, have a nick, or occur at the edge of the structure.

modification of the origami scaffold strand would be challenging as its length precludes the use of standard solid phase DNA synthesis to generate a new strand.

An analysis of the smiley face origami was done to determine how many potential binding sites were present for three previously characterized polyamide-biotin conjugates that were used in targeting DX-arrays.¹⁰ As shown, polyamide **1** is specific for the sequence 5'-WGGWCW-3',¹¹ polyamide **2** is specific for 5'-WTWCGW-3',¹² and polyamide **3** targets 5'-WGWGCW-3'.¹³ Due to the lack of specificity observed for polyamide **2** in previous studies,^{10, 14} only polyamides **1** and **3** were chosen for further

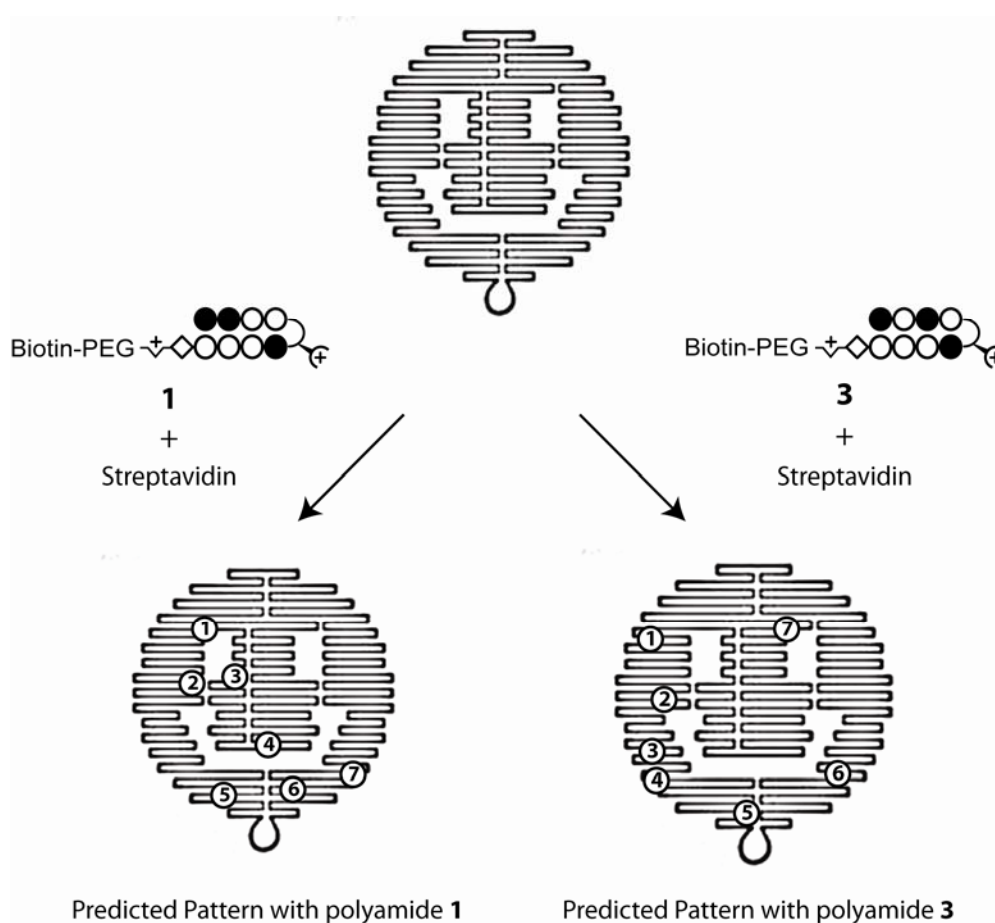


Figure 3.2 Predicted binding sites on DNA origami for polyamides **1** and **3**. The 7 optimal binding sites for each polyamide are indicated.

experiments. The total number of binding sites for their target sequences was determined as well as whether the sites were likely to be accessible and capable of being bound by the polyamide. Previous work had shown that polyamides could not bind to the crossover junction in a DX molecule.⁹ As a result binding sites that occurred at the various nicks that occur between staple strands, at any of the crossover points, or at the edges were deemed inaccessible. All other sites were considered to be optimal for binding. The location of each of the optimal binding sites was mapped to the smiley face. In both cases, there were 7 optimal sites that are predicted to create a unique pattern of observable “bumps” when the polyamide-biotin conjugates are incubated with streptavidin and the origami. The predicted patterns are shown in Figure 3.2.

Initial AFM Experiments with DNA Origami

The initial AFM experiments reproduced the creation of smiley face origami structures. Well-formed structures in agreement with the previous work were observed. The quality of structures was observed to decrease over time, so experiments were done within 24 hours of annealing. A control with streptavidin added to the origami was done to ensure that non-specific binding between the streptavidin and the origami did not occur. Next, polyamides **1** and **3** (100 nM) were separately incubated with the DNA origami (1.8 nM) and streptavidin (200 nM). A large sample of images were taken in order to determine how well the observed patterns fit with those that were predicted. In most images the smileys were observed clustered together. This is likely due to base-stacking at the ends of the origami as was observed in the original study, and could

potentially be eliminated by the addition of 4T loops at the ends of the helices. In the case of polyamide **3**, an extremely large cluster of smileys was observed. As shown in Figure

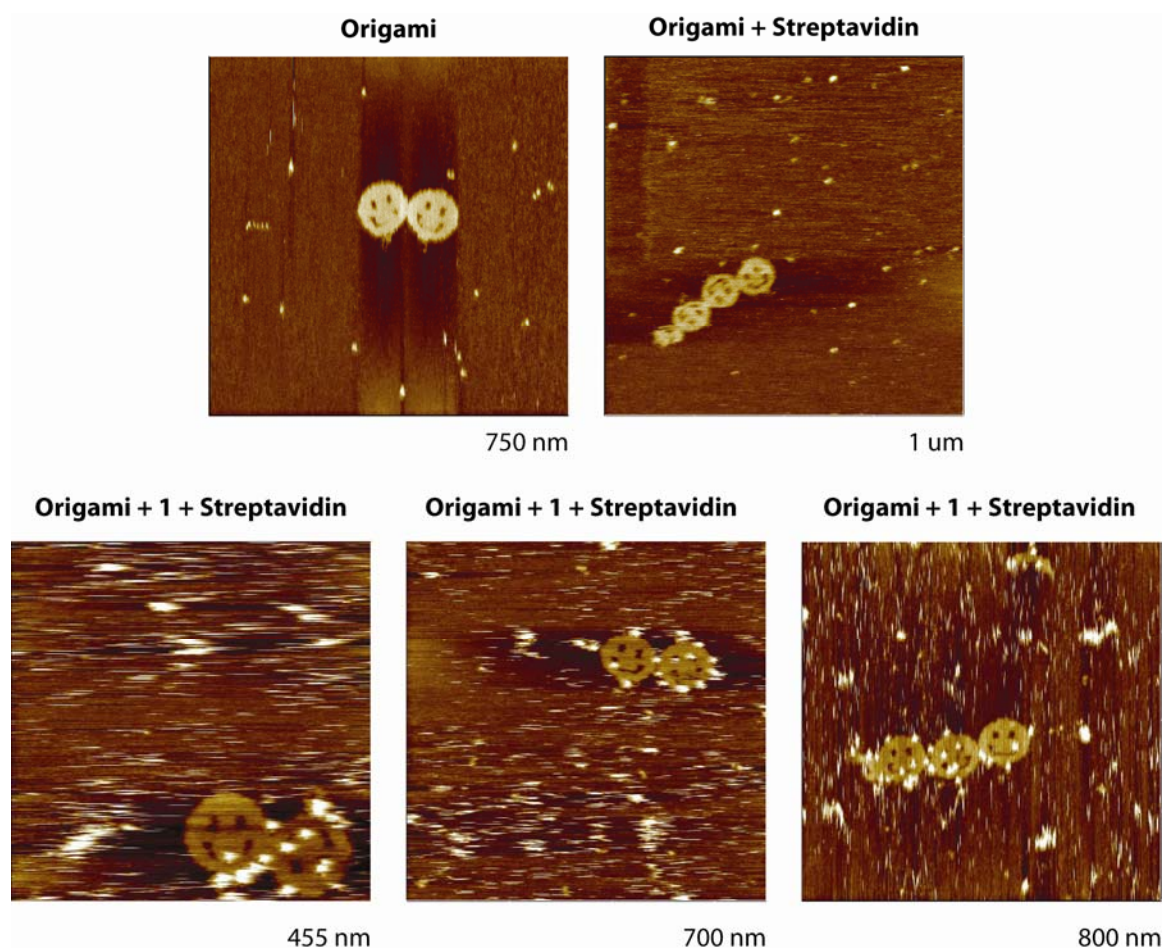


Figure 3.3 AFM Images of DNA origami with and without polyamide **1**. The dimensions of the image are shown. The top row contains the origami only and origami + streptavidin controls. The bottom row contains origami + **1** + streptavidin.

3.4, thirteen separate $1\ \mu\text{m} \times 1\ \mu\text{m}$ AFM scans were used to collect images of the entire cluster. The images were then merged, giving a set of 235 distinct DNA origami Smileys. Of these, 66 were unlabelled, 95 contained 1 bump, 58 had 2 bumps, 12 had 3 bumps, and 4 had 4 or more bumps. 72% contained at least 1 bump, and on average 1.13 bumps per structure were observed.



Figure 3.4 AFM Images of DNA origami with polyamide **3**. Thirteen individual 1 μm x 1 μm scans were performed and the images overlapped. 235 distinct structures were observed and used to determine binding locations as well as estimate the efficiency of labeling.

Analysis of AFM Experiments

Using all of the collected images, the individual patterns observed were matched with the predicted patterns for each set of polyamides to determine if there was a correlation between the observed and predicted protein recruitment. As shown in Figures 3.6 and 3.7, individual binding events were observed that appeared to occur at all of the predicted locations. However, analysis of the results was greatly hampered due to the macromolecular symmetry of the smiley face DNA structure. With a symmetric structure and no inherent topological marker, it becomes impossible to determine whether any given observed structure is in a face-up or face-down orientation. The symmetry of the

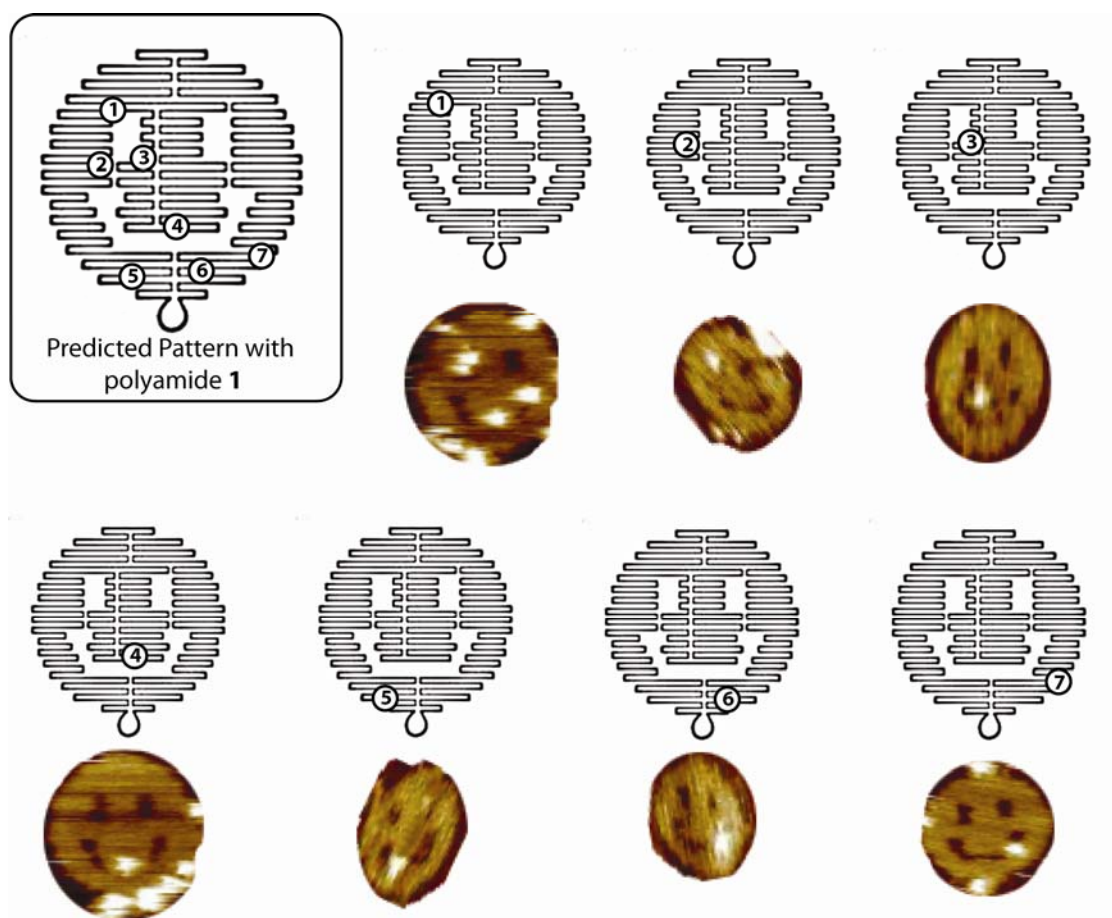


Figure 3.5 Comparison of predicted and observed binding patterns for polyamide 1. The locations of the seven optimal binding sites are shown along with images that have binding in the approximate location.

structure makes it especially difficult to distinguish between sites that are effectively mirror images of each other such as sites 5 and 6 with polyamide **1** or 4 and 6 with polyamide **3**. Exact determination of the binding sites was additionally complicated by the relatively low level of labeling observed, as attempting to decipher the orientation based upon the location of only one or two bumps was difficult. As seen in the numerous images in Figures 3.3 – 3.7, the number of binding events per structure was quite limited and

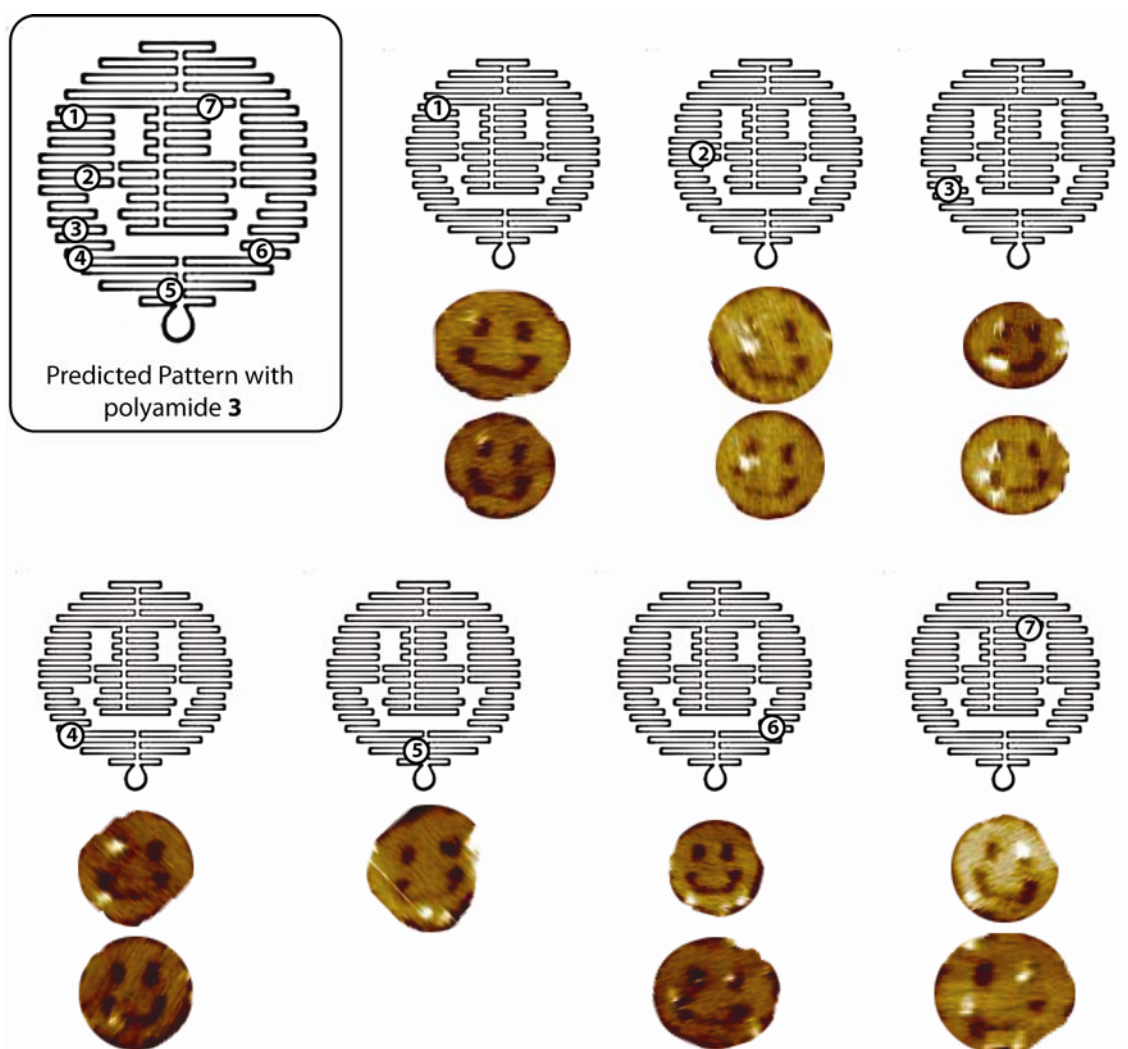


Figure 3.6 Comparison of predicted and observed binding patterns for polyamide **3**. The locations of the seven optimal binding sites are shown along with images that have binding in the approximate location.

in no case was binding at all 7 sites observed for an individual structure. Finally, the spatial resolution made it difficult to unambiguously differentiate between binding sites that were in close proximity such as sites 3 and 4 with polyamide **3**.

Asymmetric DNA Origami

As discussed in the previous section, the symmetry of the DNA origami smiley face greatly complicated the ensuing analysis of the data. As such, several strategies to overcome this difficulty were examined. Although asymmetric origami structures are known,⁶ the complete redesign and synthesis of a new asymmetric smiley face was deemed unpractical due to the time and cost involved in the de novo synthesis of a new origami structure. A second potential strategy was to take advantage of the fact that although the overall macrostructure of the smiley face origami is symmetric, the path which the scaffold strand traces and the microstructure is not. High quality AFM scans could potentially show the underlying microstructure of the origami.⁵ However, this is a highly time consuming process and would not be suitable for the analysis of a large number of structures. The most practical approach and the one which was pursued is to incorporate DNA hairpins into the origami structure. Two of the standard staple strands were replaced with staples containing DNA dumbbells. These dumbbells were designed to function as a constant topographic marker which could be used to unambiguously determine the orientation of each smiley face. The modified staples were located in the top right of the smiley face as this location has no predicted binding sites for either polyamide. The two modified staple strands were s3m4e and s5m4e (see ref. 5, Supporting Information for the exact location) which both span the 4th and 5th helix from

the top of the structure and are located above and to the right of the smiley's right "eye". These staple strands were modified by the insertion of the sequence 5'-TCCTCTTTTGAGGAACAAGTTTCTTGT-3' between the 16th and 17th nucleotides as was previously done to decorate origami with DNA dumbbells. The dumbbell structure is thought to be superior to simple DNA hairpins for this application as it avoids unwanted intermolecular dimerization.⁵

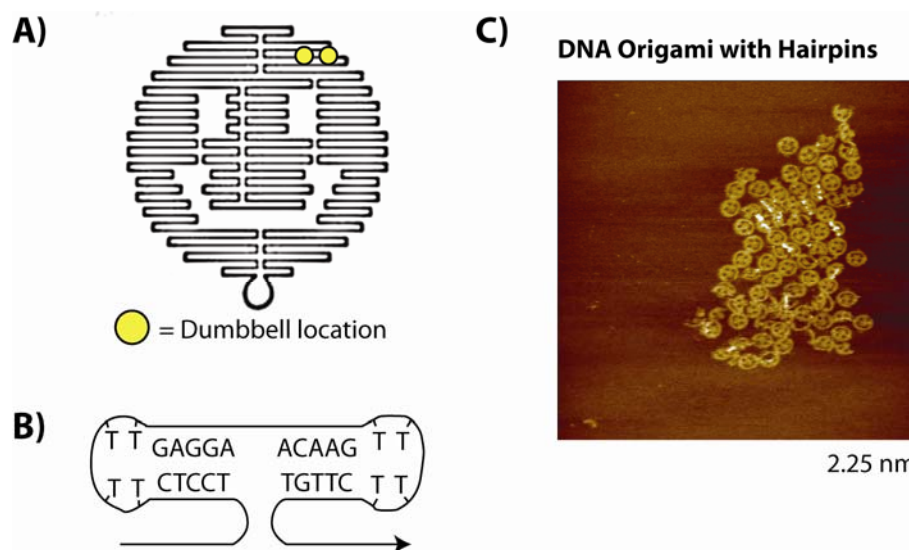


Figure 3.7 Breaking the symmetry of the DNA smiley. A) The location of the two DNA dumbbells is shown. B) The sequence and structure of the DNA dumbbells. C) AFM image of the modified smiley.

As shown in Figure 3.8C the DNA dumbbells were not visible in the modified origami. One possible explanation for this is that a larger number of dumbbells are necessary for visualization. In the case of modified DX arrays or the previously modified origami, the large number of DNA barbells in close proximity with each other may act to reinforce and rigidify each other due to the steric crowding. In the case of only two DNA dumbbells, one expects that they might be "floppier" and less visible using AFM. It should be noted that in modifying DNA origami rectangles with dumbbells to create an orientation point and to encode separate structures, the Yan group used clusters of 4

dumbbells. Future work would likely benefit from the incorporation of 2 to 4 additional DNA dumbbells in the vicinity of the original 2 to make an easily distinguishable topographical point for orienting the origami structures.

3.3 Conclusions

The use of polyamides to attach protein to DNA origami has been initially explored on a DNA smiley face. However several challenges remain for this approach to have practical uses. First, successful modification of the origami with DNA dumbbells in order to break the symmetry of the macrostructure is necessary in order to unambiguously interpret the binding patterns of the polyamides. As this has been previously demonstrated, it is not anticipated that this would be an insurmountable challenge. The greater difficulties with this approach are the specificity and labelling efficiency of polyamides. It is clear from the analysis shown in Figure 1 that standard eight-ring polyamides do not have enough intrinsic specificity to unambiguously target a single sequence in a large DNA origami structure. One potential solution is to use polyamides containing higher imidazole content, as these would be more specific than the polyamides used in this study. Two such polyamides are shown in Figure 9, which would target the sequences 5'-WGGGGW-3' and 5'-WGGGCW-3' respectively. It is worth noting that the analogous polyamides without biotin attached have been shown to target these sequences with high affinity. As shown, these polyamides are expected to have only

Polyamide	Sequence	# Sites in Full Sequence	Accessibility
	5'-WGGGGW-3'	3	2 sites - Optimal
	5'-WGGGCW-3'	9	1 site - Optimal

Figure 3.8 Polyamides proposed to have improved specificity for DNA sequences in the DNA smiley.

2 and 1 optimal binding sites respectively in the DNA smiley. Similarly, another potential solution would be to use extended polyamide motifs which are capable of binding to more than 6 base pairs.

The other major difficulty facing this approach is the low level of labelling observed by the polyamide-biotin conjugates. As previously discussed, examination of 235 unique smileys found on average just over 1 binding event per structure. It is surprising given the presence of multiple match binding sites, as well as the presence of numerous weak mismatch binding sites that so little binding is observed. Also of concern is the highly variant nature of binding from structure to structure. Although it is possible this is in part due to the inherent damage to the structures from AFM imaging, the relative stability of the structures when scanned multiple times indicates that AFM tip damage does not fully account for the observed results.

It is worth noting that in examining the binding of DNA oligomers to ssDNA overhangs on the face of DNA origami rectangles, the Yan lab found that different locations on the surface of the origami modified with the same sequence had markedly different labelling efficiencies.⁸ Positions near the edge had much higher efficiency binding than those located towards the center of the structure. These results clearly show

that every location on the DNA origami surface is not equal in terms of accessibility. Given the drastic differences observed for the binding of DNA oligomers, it is likely that proteins being recruited to the surface of the origami would be similarly affected. The presence of several large “holes” in the surface of the DNA smiley face may also play a role in determining which sites are most accessible for binding.

Addressing the challenge of highly efficient labelling will likely require further understanding of how binding occurs on the surface of these DNA structures. It is also anticipated that polyamides with higher affinities and specificities for DNA may be beneficial for labelling DNA nanostructures with improved efficiency.

3.4 Experimental Details:

Abbreviations. Trishydroxymethylaminomethane (TRIS), ethylenediaminetetraacetic acid (EDTA).

Materials. Boc- β -Ala-PAM resin was purchased from Peptides International. Trifluoroacetic acid (TFA) was purchased from Halocarbon. Methylene Chloride (DCM) was obtained from Fisher Scientific and *N,N*-dimethylformamide (DMF) from EMD. EZ-Link TFP-PEO₃-Biotin was purchased from Pierce. Streptavidin was purchased from Rockland. Ultra Pure TRIS was purchased from ICN. Magnesium Acetate 4-hydrate was obtained from J.T. Baker. Water (18 M Ω) was purified using a aquaMAX-Ultra water purification system. Biological experiments were performed using Ultrapure Water (DNase/RNase free) purchased from USB. The pH of buffers was adjusted using a

Thermo Orion 310 PerpHect Meter. All buffers were sterilized by filtration through a Nalgene 0.2 μm cellulose nitrate filtration.

Polyamide Synthesis. Polyamide monomers were prepared as described previously.¹⁵ Synthesis was performed using established protocols and all polyamides were characterized by MALDI-TOF and analytical HPLC. The synthesis of polyamides **1** and **3** is described in section 2.4 and has been published.¹⁰

DNA Oligomers. DNA oligomers were purchased from IDT Technologies in a 96 well format. The oligomers were ordered on 100 nmol scale, unpurified, and at 150 μM concentration in water. The sequences of the 243 staple strands are described in Rothmund's original paper.⁵ The M13mp18 ssDNA scaffold strand was purchased from New England Biolabs.

Annealing DNA Origami. The DNA origami was annealed as follows. First, 5 μL of each of the staple strands was mixed to create a 10X stock solution (620 nM). Fresh origami samples were annealed 24 hours before AFM experiments as follows. 10 μL of the 10X stock of staple strands (62 nM) was mixed with 1 μL of the scaffold strand (1.3 nM) in TAEMg Buffer (40 mM Tris-HCl (pH 8.0), 20 mM acetic acid, 1 mM EDTA, 12.5 mM magnesium acetate) in a final volume of 100 μL . The sample was placed on a float in a styrofoam box filled with 1 L of water at 95°C and allowed to cool to room temperature slowly overnight. The DNA dumbbell modified smileys were made in the same fashion but the modified oligomers were used instead of the original staple strands.

AFM Sample Preparation. DNA origami was incubated with 100 nM of polyamide **1** or **3** and 200 nM Streptavidin prior to imaging. 2.5 μ L of sample was spotted on mica and imaged under TAEMg buffer.

3.5 References

1. Seeman, N. C.; Lukeman, P. S., Nucleic acid nanostructures: bottom-up control of geometry on the nanoscale *Rep. Prog. Phys.* **2005**, 68, 237-270.
2. Lin, C. X.; Liu, Y.; Rinker, S.; Yan, H., DNA tile based self-assembly: Building complex nanoarchitectures *ChemPhysChem* **2006**, 7, 1641-1647.
3. Aldaye, F. A.; Palmer, A. L.; Sleiman, H. F., Assembling materials with DNA as the guide *Science* **2008**, 321, 1795-1799.
4. Seeman, N. C., DNA in a material world *Nature* **2003**, 421, 427-431.
5. Rothemund, P. W. K., Folding DNA to create nanoscale shapes and patterns *Nature* **2006**, 440, 297-302.
6. Qian, L. L.; Wang, Y.; Zhang, Z.; Zhao, J.; Pan, D.; Zhang, Y.; Liu, Q.; Fan, C. H.; Hu, J.; He, L., Analogic China map constructed by DNA *Chin. Sci. Bull.* **2006**, 51, 2973-2976.
7. Fujibayashi, K.; Hariadi, R.; Park, S. H.; Winfree, E.; Murata, S., Toward reliable algorithmic self-assembly of DNA tiles: A fixed-width cellular automaton pattern *Nano Lett.* **2008**, 8, 1791-1797.
8. Ke, Y. G.; Lindsay, S.; Chang, Y.; Liu, Y.; Yan, H., Self-assembled water-soluble nucleic acid probe tiles for label-free RNA hybridization assays *Science* **2008**, 319, 180-183.
9. Cohen, J. D.; Sadowski, J. P.; Dervan, P. B., Addressing single molecules on DNA nanostructures *Angew. Chem.-Int. Edit.* **2007**, 46, 7956-7959.
10. Cohen, J. D.; Sadowski, J. P.; Dervan, P. B., Programming multiple protein patterns on a single DNA nanostructure *J. Am. Chem. Soc.* **2008**, 130, 402-403.

11. 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, 468-471.
12. Olenyuk, B. Z.; Zhang, G. J.; Klco, J. M.; Nickols, N. G.; Kaelin, W. G.; Dervan, P. B., Inhibition of vascular endothelial growth factor with a sequence-specific hypoxia response element antagonist *Proc. Natl. Acad. Sci. U. S. A.* **2004**, 101, 16768-16773.
13. Hsu, C. F.; Phillips, J. W.; Trauger, J. W.; Farkas, M. E.; Belitsky, J. M.; Heckel, A.; Olenyuk, B. Z.; Puckett, J. W.; Wang, C. C. C.; Dervan, P. B., Completion of a programmable DNA-binding small molecule library *Tetrahedron* **2007**, 63, 6146-6151.
14. Puckett, J. W.; Muzikar, K. A.; Tietjen, J.; Warren, C. L.; Ansari, A. Z.; Dervan, P. B., Quantitative microarray profiling of DNA-binding molecules *J. Am. Chem. Soc.* **2007**, 129, 12310-12319.
15. Baird, E. E.; Dervan, P. B., Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids *J. Am. Chem. Soc.* **1996**, 118, 6141-6146.

Chapter 4

Targeting the Nucleosome Core Particle with Polyamide Dimers

The work described in this chapter was done in collaboration with Karolyn Luger (Colorado State University), Joel Gottesfeld (The Scripps Research Institute), and David Chenoweth.

Abstract

The Nucleosome Core Particle (NCP) is the minimal unit in eukaryotic chromatin and a unique target for DNA-binding polyamides. The ability to modulate the stability of nucleosomes would be a useful tool for molecular biology and gene regulation. A series of polyamide clamps designed to differentiate between the NCP and B-form DNA were synthesized and analyzed using DNase I footprinting. In addition, a new polyamide dimer designed to bind to and stabilize the NCP was synthesized. Finally, the NCP was used to perform the DNA-templated ligation of alkyne- and azide-functionalized polyamides.

4.1 Introduction

In order for DNA-binding molecules to be useful tools for gene regulation, they must be able to access DNA in biological systems. DNA in eukaryotic organisms exists in a highly condensed form as chromatin. The nucleosome core particle (NCP) represents the most basic unit of the higher order structure chromatin. The NCP consists of 147 bp of DNA wrapped twice around an octamer of histone proteins. The histone octamer contains two copies each of histones H2A, H2B, H3, and H4. One of the most interesting features of the NCP is the alignment between the major and minor-grooves of the two gyres of DNA. These aligned minor-grooves separated by a small gap between the gyres of DNA create a “super groove”, consisting of 14-16 bp of DNA that is accessible for recognition. Each of these supergrooves bring sequence elements that are 80 bp apart in the linear DNA strand into close spatial proximity. As a result, supergrooves create a recognition platform that exists solely in the context of the NCP.

The ability of polyamides to bind to B-form DNA has been well characterized.¹ As discussed in the previous two chapters, polyamides are also capable of binding to the DNA present in 2-dimensional nanostructures.^{2,3} The NCP is similar to these structures in that it creates a nanometer scale DNA architecture. Initial biochemical studies showed that polyamides are capable of binding to sites on the NCP facing away from or even partially towards the histone proteins.⁴ High resolution crystal structures were also obtained, showing polyamides in complex with the NCP.⁵

It was later shown that by linking two polyamides together as a dimer, with a linker of sufficient length to extend over the gap, one can specifically bind to a super groove on the NCP.⁶ Interestingly, the polyamide dimer was shown to greatly improve

the stability of the NCP, presumably by preventing unraveling and dissociation of the DNA from the histone octamer. As a result, the linked polyamide dimer has been referred to as a “clamp” because of its ability to effectively clamp the DNA together around the histone core.

The ability to stabilize the NCP could prove useful for a variety of purposes. Since polyamide clamps can stabilize the NCP *in vitro*, they may also be able to stabilize chromatin in cells, a feat which would have important ramifications for gene regulation. Additionally, it was observed that incubation with the polyamide clamps led to the growth of larger crystals in the previous X-Ray crystallography studies. This is presumed to a result of the increased stability of the NCP. As such, polyamide clamps could prove to be extremely useful tools for generating high quality crystal structures of the NCP. Since the original clamp site is located opposite the ends of the DNA it still allows partial dissociation of the DNA from the protein core. We hypothesized that it would be possible to further increase the stability of the NCP by creating a new polyamide clamp targeted to a supergroove located closer to the ends of the DNA. This new “end clamp” could potentially prevent partial dissociation and result in further stabilization of the NCP.

In addition, we initially sought to determine whether one could design a clamp that would be selective for the NCP versus regular B-form DNA. The ability to specifically target DNA solely when in the NCP could potentially be used as a tool for either detecting or stabilizing chromatin, providing a useful tool for modifying gene expression. The key design element was to take advantage of the gap between the gyres of DNA in a supergroove which would not be present in B-form DNA. This gap should be able to accommodate increases in the size of the linker that connects the two

polyamides, unlike the minor groove which would create a steric clash with larger linkers presumably leading to abrogated binding.

Finally, linked dimers provide an interesting motif for targeting extended DNA sequences. A standard eight-ring polyamide has specificity for 6 bases while linked polyamides can target sequences up to 10 bp in length.⁷⁻⁹ For future gene regulation projects, it may be desirable to target longer sequences to prevent binding at other undesired locations in the genome. However, the uptake properties of these larger conjugates are generally thought to be worse than eight-ring polyamides due to their larger size. A potential solution which has been suggested to this problem is to generate the larger polyamide dimers inside the cell. To this end, prior work has demonstrated how DNA-templated ligation chemistry can be used to link polyamides once they are bound in close proximity in the minor-groove of DNA.¹⁰ Having demonstrated the ability for both halves of a polyamide dimer to bind to adjacent minor-grooves in the supergroove of the NCP, we postulated that it may be possible to use the super groove as a unique template for DNA-templated ligation. This reaction should be specific to the NCP as the two half sites are separated by 80 bp on the DNA strand and are only brought into close spatial proximity when wrapped around the histone octamer.

4.2 Results and Discussion

Engineering a Clamp specific for the NCP versus B-form DNA

The original NCP clamp contains two identical polyamides linked by a short PEG linker. Using this as a starting point, a small library of homodimer polyamide clamps was synthesized using the synthetic scheme shown in Figure 4.1. The advantage of this

scheme is that it is amenable to the use of any diacid as a linker and requires just a single coupling step followed by HPLC purification. Therefore, a small library of diacids can be rapidly converted to polyamide dimers. Several diacids, selected on the basis of their presumed rigidity, appropriateness of length, and steric bulk were chosen and the corresponding dimers were synthesized. The small library is shown in Figure 4.2.

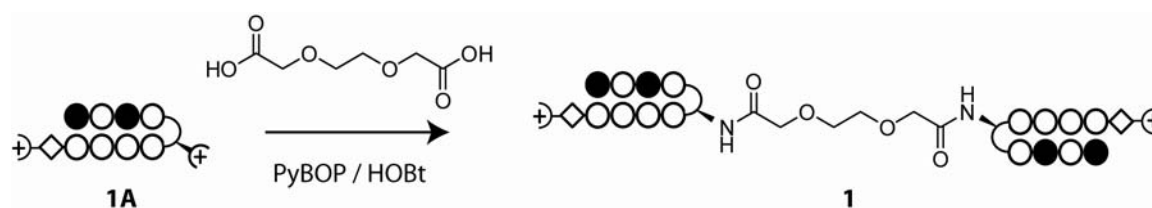


Figure 4.1 Synthetic scheme for the synthesis of polyamide homodimers. The synthesis of compound **1** is shown. The synthesis can be readily adapted for the use of any diacid linker.

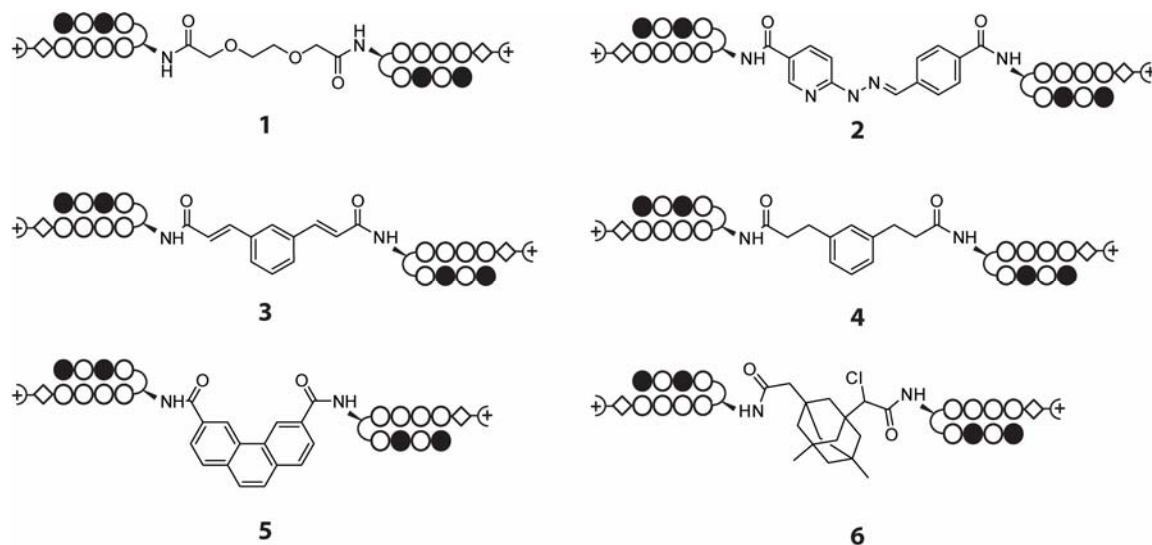


Figure 4.2 Small library of polyamide clamps containing sterically bulky linkers.

DNAse I footprinting was performed to determine whether the library of clamps retained the ability to bind to B-form DNA. The DNA sequence that they were footprinted on was designed to contain three clamp binding sites. Each dimer site consists of two hairpin binding sites separated by 0, 1, or 2 bps respectively. With the exception

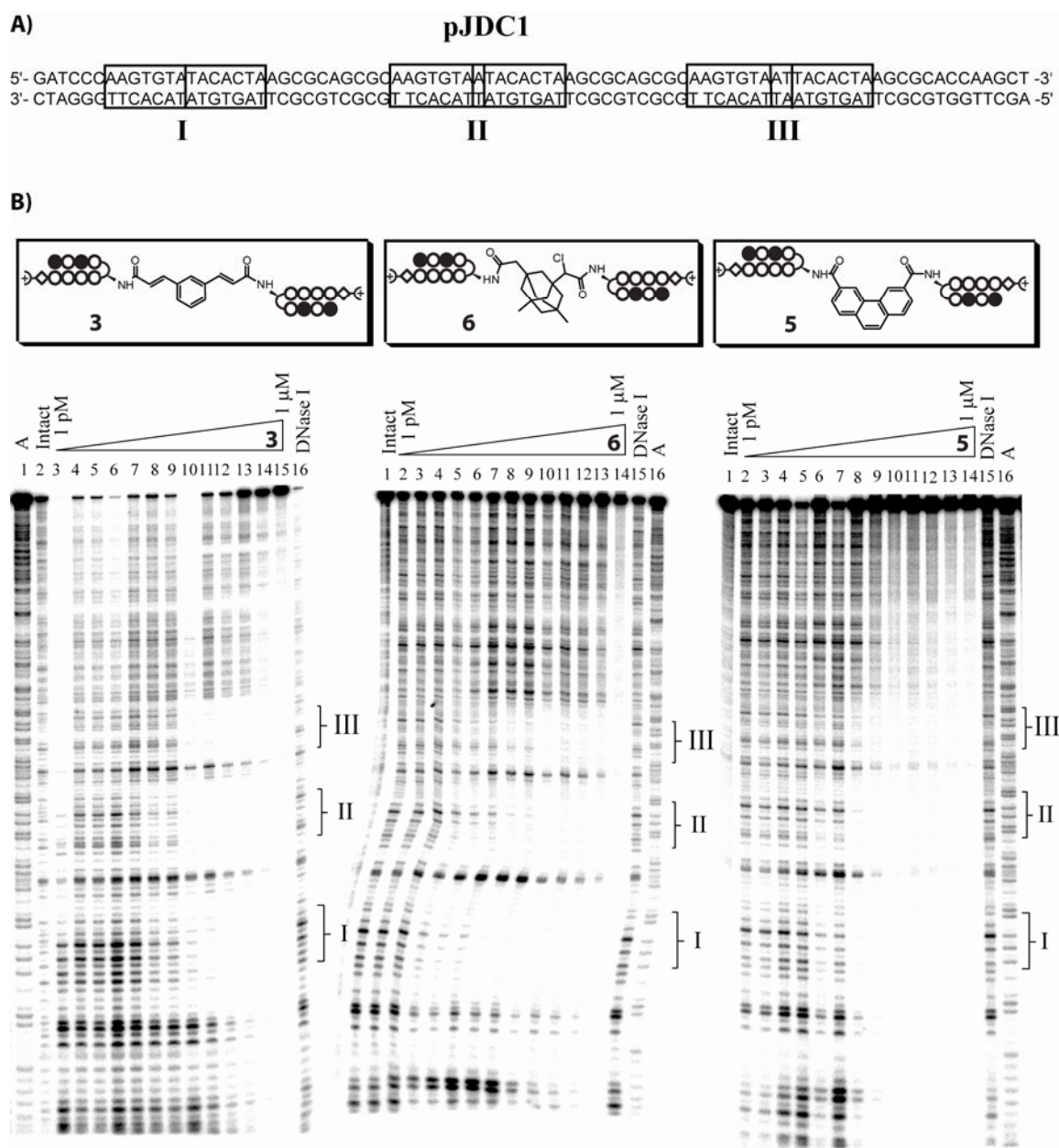


Figure 4.3 DNase I footprinting experiments with proposed NCP-specific clamps. A) Sequence of the insert from plasmid pJDC1 used for footprinting. The insert contains three binding sites with 0, 1, and 2 bp spacing between the two half sites. B) Representative gels are shown for polyamide dimers **3**, **5**, and **6**. As shown, both **3** and **6** show binding at all three target sites. Polyamide dimer **5** shows non-specific DNA binding.

of **5**, all clamps that were tested showed good binding affinity for all three sites. **5** showed high affinity binding with almost no selectivity across the entire insert, consistent with nonselective DNA coating (Figure 4.3B).

The ability of all library members to bind to DNA was unexpected, as it was presumed that the use of larger linkers, especially in compounds **5** and **6** would prevent binding to B-form DNA for steric reasons. One potential explanation for the observations is that only one half of the dimer is binding specifically while the other half is prevented from interacting with the DNA by the linker. Because we are seeing a statistical mixture of fragments, this binding mode would be indistinguishable by standard DNase I footprinting. In the original work done by Phillip Weyermann characterizing turn-to-turn hairpin dimers,⁹ it was found that double mismatches were tolerated if both mismatches occurred under just one polyamide, whereas a mismatch under both polyamides of the dimer created a 10 fold decrease in binding affinity. This is consistent with the proposed mode of binding that we observed.

Following the work done using a small library approach, we attempted to use rational design to approach the problem of finding an appropriate linker that would give us the desired specificity between the NCP and B-form DNA. Going back to the original PEG linker in **1**, it was hypothesized that by building outwards while keeping the same basic structure and length, it would be possible to perform a more detailed investigation of the role of the linker while eliminating the clamp's ability to bind B-form DNA.

By adding methyl groups to the polyethylene glycol linker, one can slowly add steric bulk while keeping the same linker distance. The synthetic scheme shown in Figure 4.4, which had previously been described was used to synthesize the diacids.¹¹

Conveniently this synthetic scheme allows for the use of any diol as a starting material and numerous potential linkers can be synthesized with relative ease.

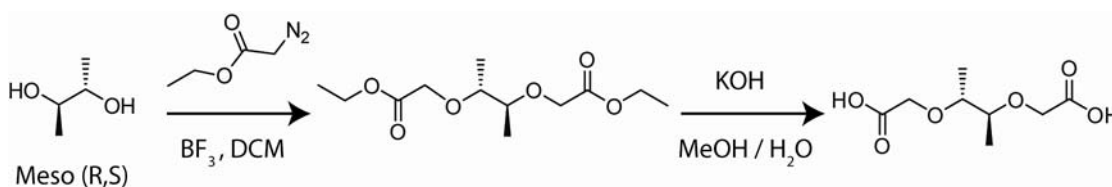
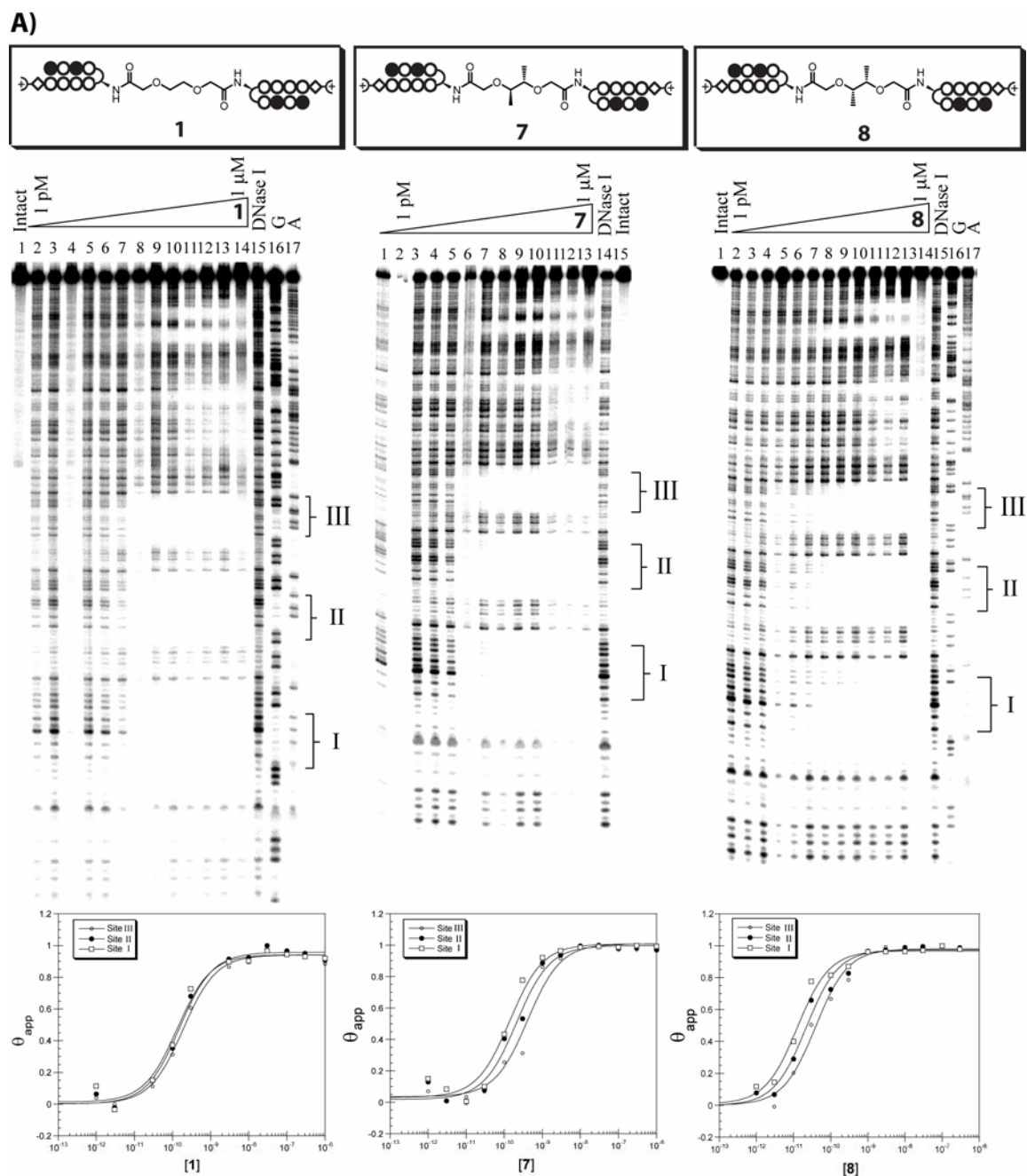


Figure 4.4 Synthetic scheme for synthesizing PEG diacids from starting diols.¹¹

In order to investigate the role of the stereochemistry of the methyl groups, both the meso and R,R diols were used to make the corresponding diacids. Both methylated PEG diacids were then used to make clamps **7** and **8**. Quantitative DNase footprinting of was performed using the previously described DNA sequence and the K_a values were determined at all three binding sites. The data for clamps **7** and **8** is shown next to the results using the original clamp **1** in Figure 4.5. As shown, the addition of the methyl groups did not interfere with binding, providing further evidence that the linker was not sequestered in the floor of the minor-groove.

Based upon these results, several new experiments to further investigate what effect the linker has upon clamp binding were performed. In order to dissect the interactions that were occurring, control compound **9**, the polyamide with an acetylated turn, was synthesized. Similarly, the parent polyamide was coupled to the PEG linker and then capped using methylamine to make control compound **10**. Finally, to investigate whether the clamp was truly binding as a dimer, or with only one polyamide at a time, a new plasmid with a modified insert was constructed. Plasmid pJDC2 was designed so that the hairpin binding sites for the clamp are separated by 0, 3, and 4 bps. The 0 bp



B) Equilibrium Association Constants, K_a (M^{-1})

Compound	Site I 0 space	Site II 1 space	Site III 2 space
1	$7.5(\pm 1.6) \times 10^9$	$6.6(\pm 1.0) \times 10^9$	$5.3(\pm 0.8) \times 10^9$
7	$7.2(\pm 1.6) \times 10^9$	$4.8(\pm 1.0) \times 10^9$	$2.5(\pm 0.6) \times 10^9$
8	$7.9(\pm 1.7) \times 10^{10}$	$4.6(\pm 1.0) \times 10^8$	$2.6(\pm 0.5) \times 10^{10}$

Figure 4.5 DNase I footprinting of compounds 1, 7, and 8. (A) The structure of each clamp is shown along with the DNase I footprinting gel. Footprinting was performed using plasmid pJDC1 whose sequence is

shown in Figure 4.3A. The binding isotherms for each compound for each of the three binding sites are shown at bottom. (B) Table showing equilibrium association constant for each compound at each of the three sites.

site was retained as a control to be able to compare results between the two plasmids.

Based upon calculations and computational modeling it should be impossible for the clamp to bind as a dimer at the 3 and 4 bp sites because the linker cannot extend over a 3 bp gap in binding sites. For binding to occur at these sites, at least one of the polyamides would have to be localized over a mismatch sequence.

DNAse footprinting was done on the newly designed plasmid insert using compounds **1**, **9**, and **10**. The results are shown in Figure 4.6. As shown, the affinities of the original clamp **1** and compound **10** which has just a single polyamide attached to the linker, are nearly indistinguishable. This indicates that the second polyamide present in **1** does not appear to contribute to binding. It is interesting that **9**, the only compound that does not contain the PEG linker has higher affinities than either **1** or **10**. This indicates that the PEG linker may at least partially interfere with binding. Finally, the similar affinities of compound **1** across all three designed binding sites, despite the fact that the linker is not long enough to cross the 3 and 4 bp gap between the half sites, indicates that the dimer is likely binding in the previously discussed mode where only one of the polyamides is bound sequence specifically onto its match site and the other is bound non-specifically over a mismatch site adjacent to it.

Although unsuccessful in creating a dimer capable of discriminating between the NCP and B-form DNA, a large amount of information regarding the binding mode of dimeric polyamides was gained through this study. As demonstrated, polyamide turn-to-turn dimers are capable of binding to a match site for one of their component polyamides,

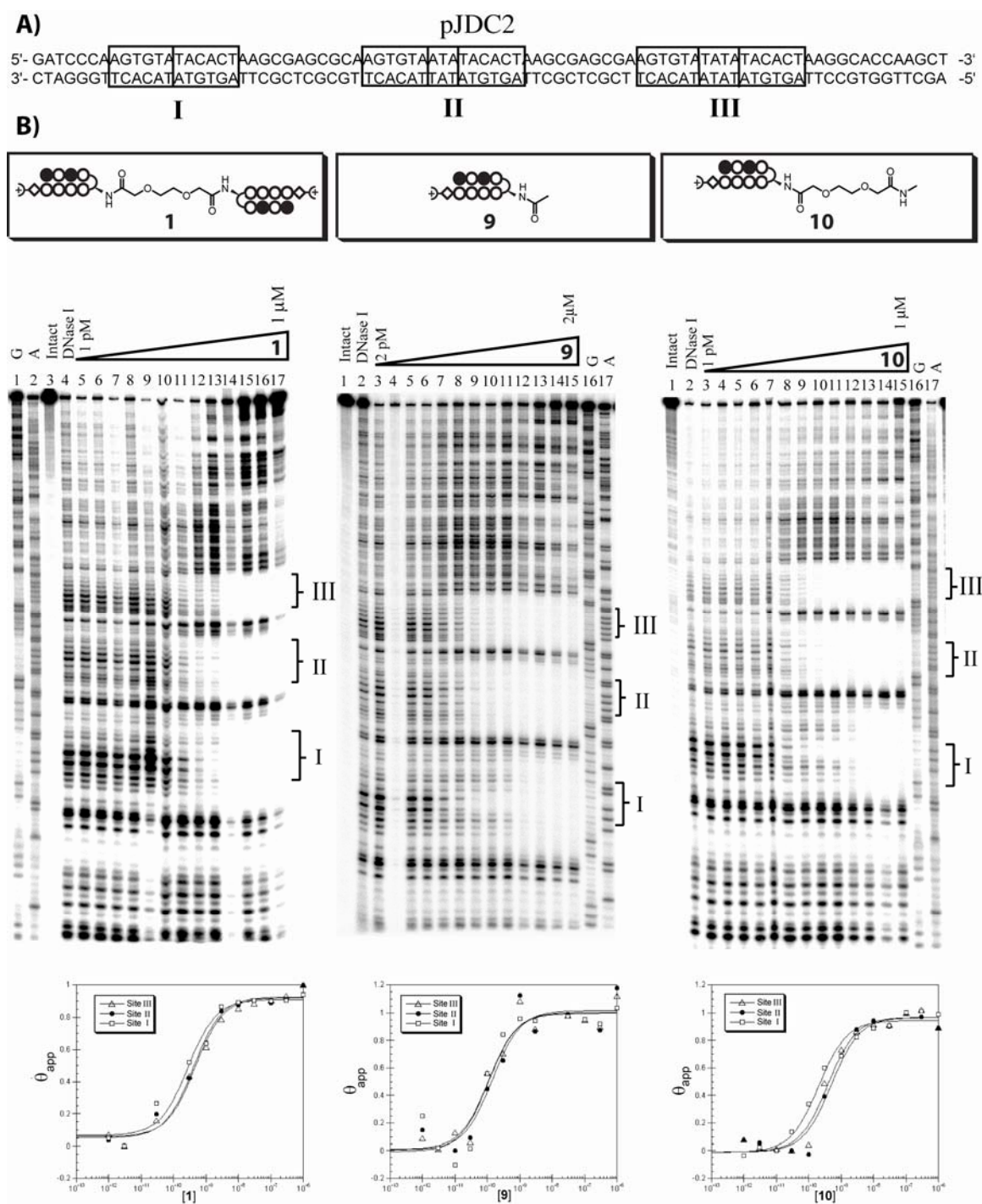


Figure 4.6 DNase footprinting of compounds **1**, **9**, and **10**. A) The sequence of plasmid Insert pJDC2 containing three binding sites labeled I, II, and III which contain 0, 3, and 4 bp respectively between polyamide binding sites. B) Structures of polyamides **1**, **9**, and **10**. Representative footprinting gels and binding isotherms for each of the three polyamides are shown below.

Table 4.1 Equilibrium association constants, K_a (M^{-1}) on plasmids pJDC1 and pJDC2.

Polyamide	Plasmid pJDC1			Plasmid pJDC2		
	0 bp	1 bp	2 bp	0 bp	3 bp	4 bp
7	$7.2 (\pm 1.6) \times 10^9$	$4.8 (\pm 1.0) \times 10^9$	$2.5 (\pm 0.6) \times 10^9$	N.D.	N.D.	N.D.
8	$7.9 (\pm 1.7) \times 10^{10}$	$4.6 (\pm 1.0) \times 10^{10}$	$2.6 (\pm 0.5) \times 10^{10}$	N.D.	N.D.	N.D.
1	$7.5 (\pm 1.6) \times 10^9$	$6.6 (\pm 1.0) \times 10^9$	$5.3 (\pm 0.8) \times 10^9$	$4.1 (\pm 0.9) \times 10^9$	$2.9 (\pm 0.6) \times 10^9$	$2.5 (\pm 0.4) \times 10^9$
9	N.D.	N.D.	N.D.	$8.7 (\pm 2.1) \times 10^9$	$8.1 (\pm 1.9) \times 10^9$	$1.0 (\pm 0.2) \times 10^{10}$
10	N.D.	N.D.	N.D.	$5.4 (\pm 1.3) \times 10^9$	$2.5 (\pm 0.8) \times 10^9$	$2.5 (\pm 0.8) \times 10^9$

despite the presence of a mismatch sequence under the second polyamide. This is in agreement with previous reports.⁹ This could have important implications for the design of future clamp molecules as well as for the design of polyamide motifs that can bind to increasingly longer sequences. Additionally, the loss in affinity upon incorporation of the PEG linker indicates that the currently used PEG linker may be interfering with binding. Additional studies are needed to determine whether a linker that does not negatively affect binding can be developed.

Creating a NCP End Clamp

The original NCP-binding clamp **1** targets a symmetrical “supergroove” and as such consists of two copies of the same polyamide joined by a short linker. To create a clamp that could further stabilize the NCP, we first examined the two sites that make up the super groove which is located closest to the ends of the DNA. Unlike the supergroove targeted in the previous studies the supergoove at this location contains two distinct half sites, making it necessary to synthesize heterodimer clamps containing two different polyamides.

To determine which polyamides were optimal for targeting the two half sites the five polyamide-EDTA conjugates shown in Figure 4.7 were synthesized, building upon previous knowledge about targeting the selected DNA sequences. The 5'-TCCACCT-3' site is designed to be targeted by polyamide **15**. As shown previously, a 2- β -2 motif using ImIm- β -ImIm without a chiral turn binds duplex DNA with a K_a of $7.6 \times 10^9 \text{ M}^{-1}$.¹² Interestingly, affinity cleavage experiments showed that the polyamide preferentially binds in a reverse orientation, aligning C to N along the 5' to 3' strand of DNA.

Polyamide **12** which targets the 5'-TGATGGA-3' sequence has been previously characterized and was shown to have a K_a of $3.0 \times 10^8 \text{ M}^{-1}$ and also to prefer binding in the reverse orientation when it had an acetylated chiral turn.¹³ In addition, polyamide **13** is a characterized forward-binding polyamide that has a K_a of $8.6 \times 10^9 \text{ M}^{-1}$ when the turn is acetylated.¹³ In addition to these, an additional polyamide **14** was designed based on the hypothesis that polyamides too far back along the sequence may sterically clash with the histone proteins. This polyamide was designed to target a region slightly more exposed on the NCP. Polyamide **14** incorporates 3-methoxythiophene in the n-terminal position which has been shown to have improved affinity for the T·A base pair,¹⁴ and was used to target the 5'-TTTGAT-3' sequence. Finally, as a positive control for the affinity cleavage studies, polyamide **11** was used as it was well-behaved and previously characterized in this assay.⁴ These five polyamide-EDTA conjugates were synthesized and affinity cleavage experiments were done to examine the ability of each compound to bind to the NCP.

The NCP was reconstituted using previously established protocols.⁴ Samples of reconstituted histone octamer as well as the 146 bp DNA were obtained from Karolin

Luger at Colorado State University. In general, reconstitution occurred in greater than 90% purity. Affinity Cleavage experiments were done at The Scripps Research Institute with the assistance of Joel Gottesfeld following established protocols.⁴ The expected binding sites as well as weak mismatch sites are shown in Figure 4.7. All five EDTA compounds were tested for cleavage on both the NCP as well as the linear 146 bp DNA.

As shown in Figure 4.8, polyamide **11** showed very little binding to the NCP in contrast to previous reports.⁴ As a result we cannot rule out the possibility of false negatives in our results. The cleavage observed with **11** in the DNA lane would indicate

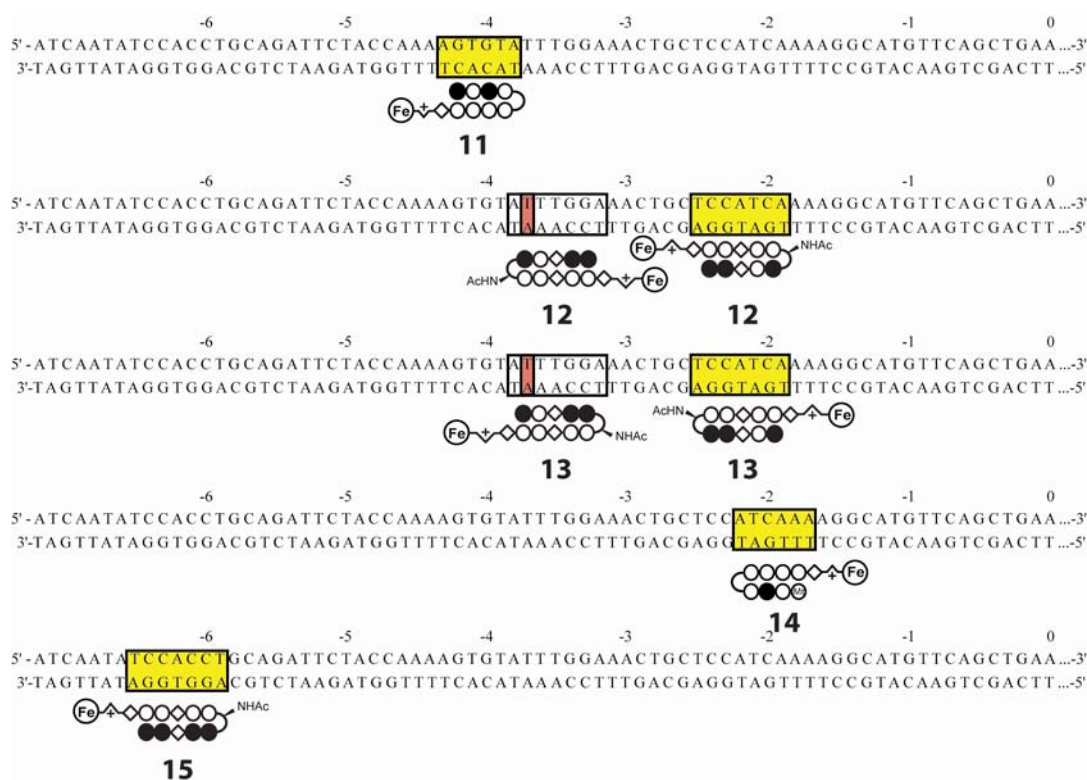


Figure 4.7 Binding sites for polyamide-EDTA conjugates **11** - **15** on the NCP. Half of the palindromic 146 bp sequence is shown. Match sites for each polyamide are boxed in yellow. A weak mismatch site for polyamides **12** and **13** is boxed and the mismatched base pair highlighted in red. Polyamide **15** targets the first half-site of the supergroove while **12** – **14** target the second half-site. Polyamide **11** is included as a positive control for affinity cleavage. Mt = 3-methoxythiophene.

that the polyamide has not degraded. No conclusions were made regarding compound **15** as the location of the presumed binding site is located too close to the end of the DNA to be visible in our experiments. Polyamide **14** showed little cleavage and presumably did not bind well in these experiments. Cleavage was observed for **12** and **13** at both match site IV and mismatch site III although the amount of cleavage appears to be much weaker

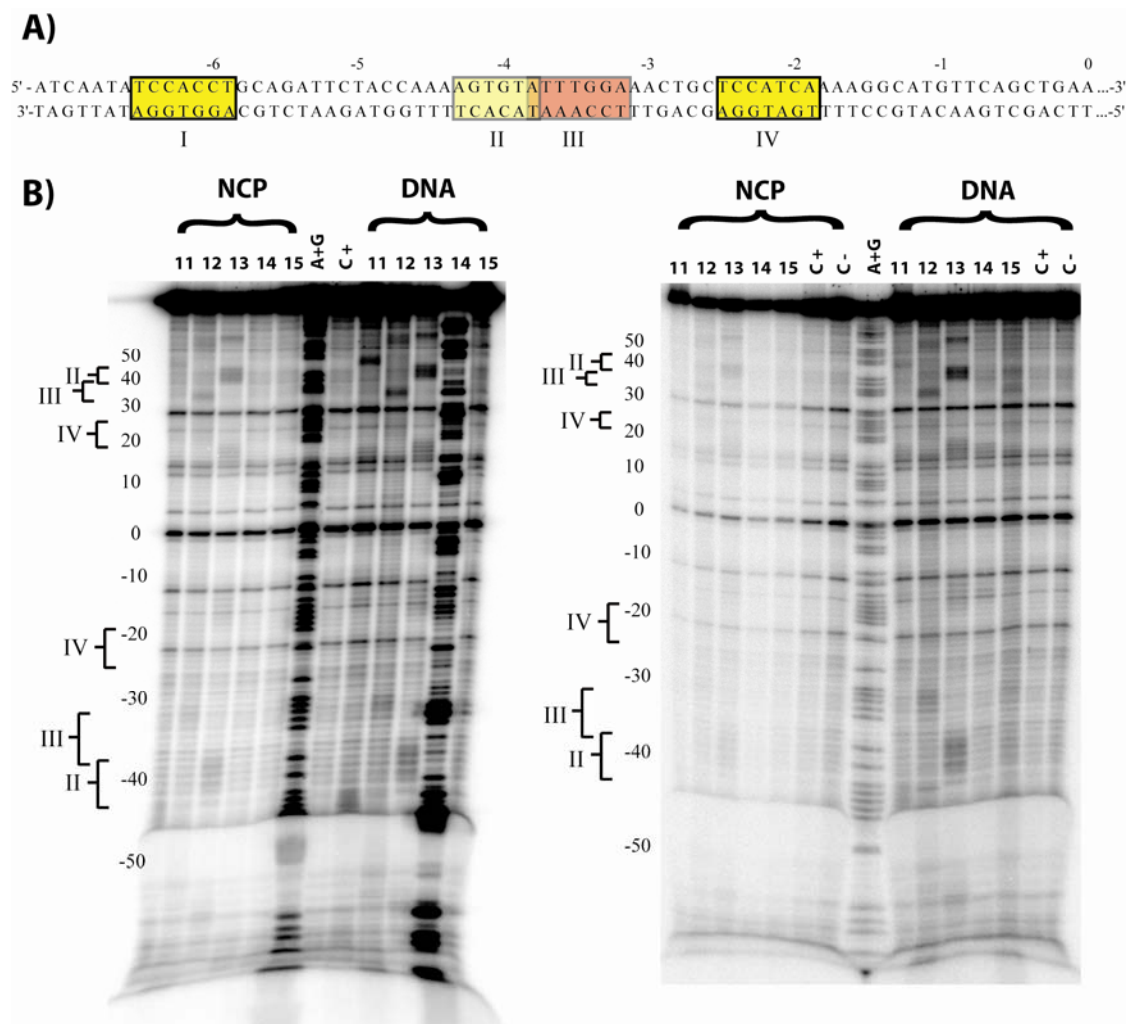


Figure 4.8 Affinity cleavage on the NCP. A) The 146 bp palindromic DNA strand is shown with the polyamide binding sites shown in Figure 4.9 boxed in yellow. The mismatch binding site for **12** and **13** is boxed in red. B) The results of two affinity cleavage experiments that were done on the NCP and the linear 146 bp DNA with compounds **11** – **15**. 50 nM of each polyamide was used. Lanes denoted A+G are for sequencing. Lanes marked C+ had ferrous ammonium sulfate, used to initiate the cleavage reaction, added to them although they did not contain any polyamide, while lanes marked C- did not contain polyamide or the ferrous ammonium sulfate solution.

for the NCP in comparison to linear DNA. **12** binds predominantly at location III while **13** shows cleavage at both location III and IV. The amount of cleavage observed also appears to be strongest for compound **13** which is consistent with its parent having a 20-fold higher affinity than the parent of **12** as discussed previously. As such, our design efforts for an end clamp focused on this compound.

Given these results, we wished to further characterize binding at both of the half sites before making a potential end clamp. A series of melting temperature experiments were performed to determine the ability of polyamides **16** - **19**, the parent compounds of EDTA conjugates **15** and **11**, to bind to their respective half sites. A series of 13 bp duplexes were designed based upon the sequences found in the NCP. As a mismatch

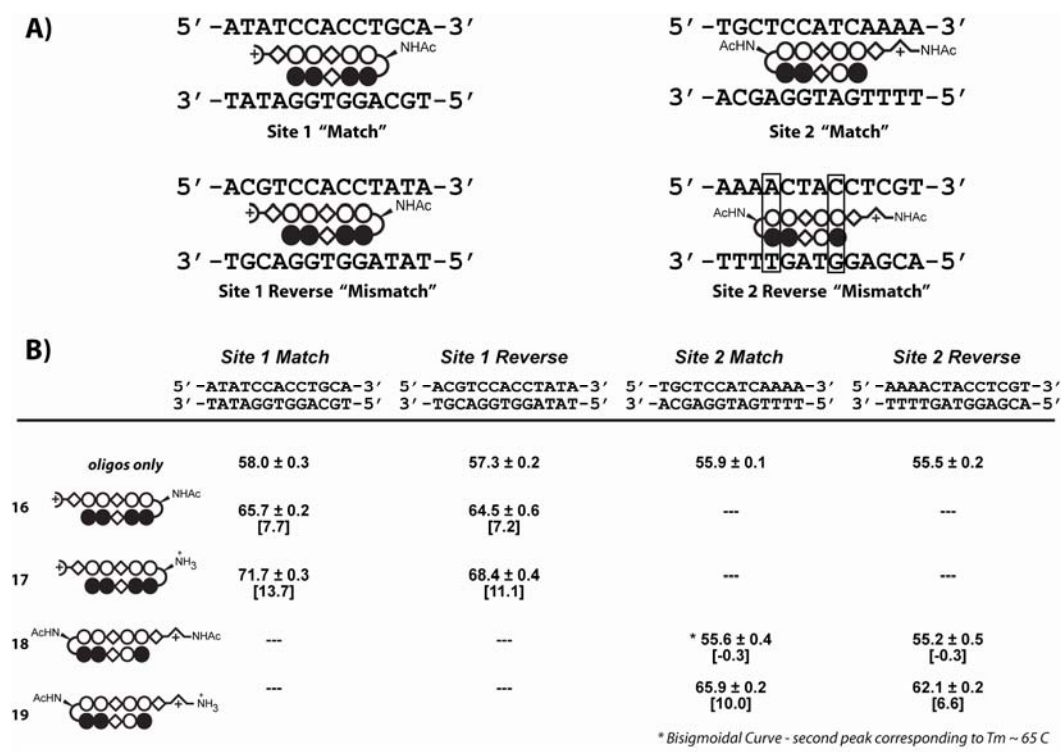


Figure 4.9 Melting temperature analysis for end clamp polyamides. A) The design of the match and mismatch DNA duplexes is shown. The most likely binding modes for the mismatch duplexes which contain the reversed DNA sequence are shown. B) Melting temperatures for each of the compounds. The numbers in brackets are the difference in melting temperature observed when compared to duplex only controls.

control, DNA duplexes where the DNA sequence was reversed were used as well (Figure 4.9A). As shown, both polyamides **16** and **17** bind to the site 1 match site. Binding at the site 1 mismatch was also observed due to the symmetric nature of the binding site, which allows the polyamide to bind at the same sequence with only slightly different flanking base pairs. (see Figure 4.9A, Site 1 match and reverse mismatch). As expected the charged compound **17** had a significantly higher melting temperature consistent with the higher binding affinities generally observed with the free amine present on the turn. Compounds **18** and **19** showed similar binding profiles for the Site 2 match and mismatch sequences. Compound **18** had a bisigmoidal curve at the match site which complicated the analysis. However it appears that both of these compounds bind at the match site and have significantly decreased binding affinities at the mismatch sites. This is consistent with the analysis of mismatch binding in Figure 4.9A where either the polyamide must bind with the chiral turn facing into the minor groove or over a double base pair mismatch site, either of which would lead to an extreme decrease in binding. As was seen for compounds **16** and **17**, the free amine on polyamide **19** resulted in a larger increase in melting temperature than the acylated compound **18**. Based upon this analysis, the polyamide cores chosen appear to bind to the sequences present in the supergroove with high affinity.

As a final result of these studies, end-clamp **20** was synthesized as outlined in Figure 4.10. The longer 3-oxa PEG linker was chosen as it should allow the most freedom to bind, and previous studies had shown that increasing linker length had no deleterious effects on clamp binding.⁶ HPLC purification was done after each coupling to purify the intermediates. The second coupling went in 51% yield and produced 120 nmol

of product **20** after purification. It should be noted that in the course of this study and similar experiments, yields for these types of couplings after HPLC purification rarely exceed 30% and often times were less than 10%. The relatively low yields made the synthesis of heterodimer **20** and other heterodimers not discussed here extremely challenging. As a result investigations were done to optimize the yield of the coupling steps. Numerous peptide coupling reagents were tested on small scale for use in this reaction. These include DCC, EDC, NHS, HBTU, HATU, TFFH, DMTMM, and DMAP. However, PyBOP was found to be the optimal activation reagent for these reactions. The 120 nmol of **20** was shipped to the Karolyn Luger's group at the University of Colorado for crystallization trials. It is hoped that this clamp will have an improved ability to stabilize nucleosome core particles.

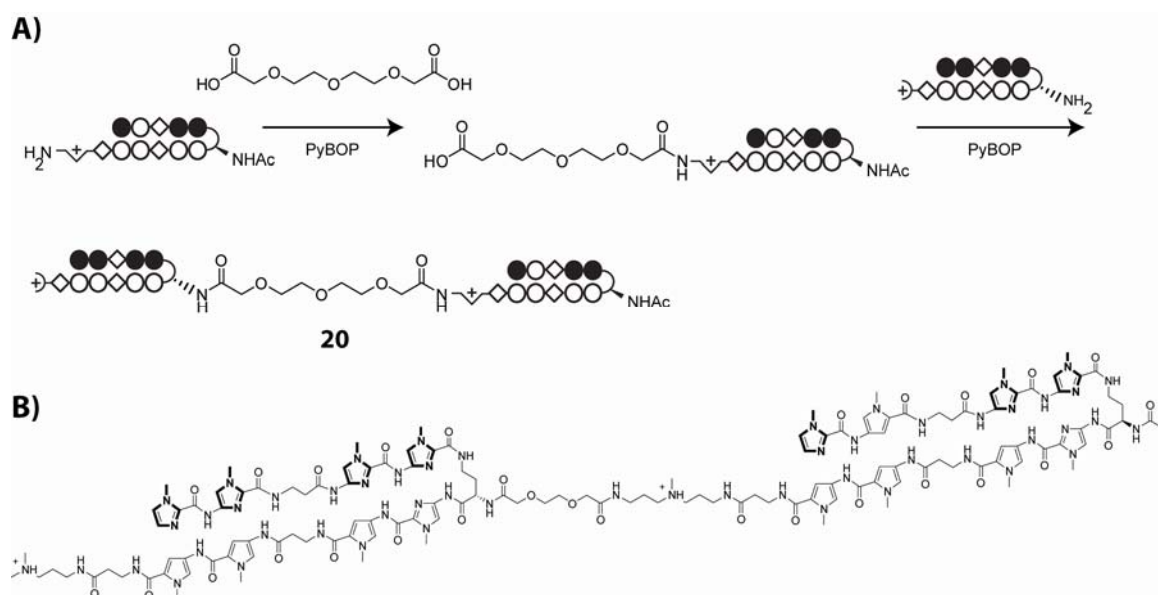


Figure 4.10 Synthesis of End Clamp **20**. A) The heterodimer end clamp was synthesized by two coupling steps. B) Chemical structure of **20**.

NCP Templated Ligation

Previous work has shown how DNA can be used as a template to direct reactions between two functionalized polyamides.¹⁰ Two polyamides targeted to adjacent sites in the minor groove of duplex DNA, one functionalized with an azide and the other with an alkyne were able to undergo a 1,3-dipolar cycloaddition reaction, forming a covalent

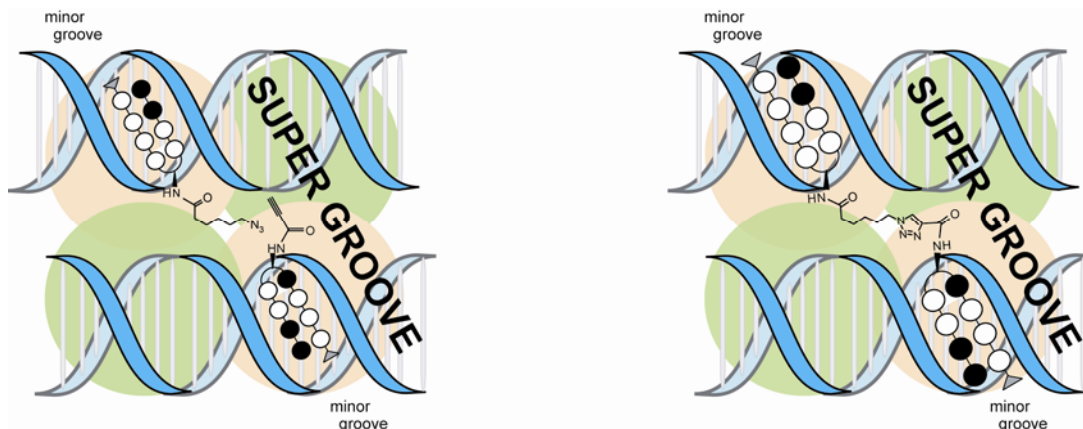


Figure 4.11 Schematic for NCP-mediated ligation. Two polyamides functionalized with an alkyne and azide are able to react in the supergroove of the NCP.

triazole linkage.¹⁰ In this section we examine whether the analogous templated reaction could be performed on the NCP to generate a polyamide dimer.

In order to ascertain whether the NCP could be used to template ligation reactions, a series of three azide-containing polyamides and two alkyne-containing polyamides were synthesized. Several linker lengths were used in order to examine the distance dependence of the reaction. Examination of the crystal structure of the polyamide clamp bound to the NCP,^{6, 15} and computational modeling showed that only the azide containing the longest linker was expected to be capable of reacting. In addition, the previous study of DNA ligation had demonstrated that the activated alkyne present in **24** reacts over 100 times faster than alkyl alkyne in **25**.

The NCP was reconstituted as outlined in the previous section. The same ligation reaction was done using all three azides and both of the alkynes. In separate tubes, 200 pmol of each polyamide-alkyne was reacted with 200 pmol of each polyamide-azide and 40 pmol of the reconstituted NCP. The reaction was incubated for 5 hrs at 37°C

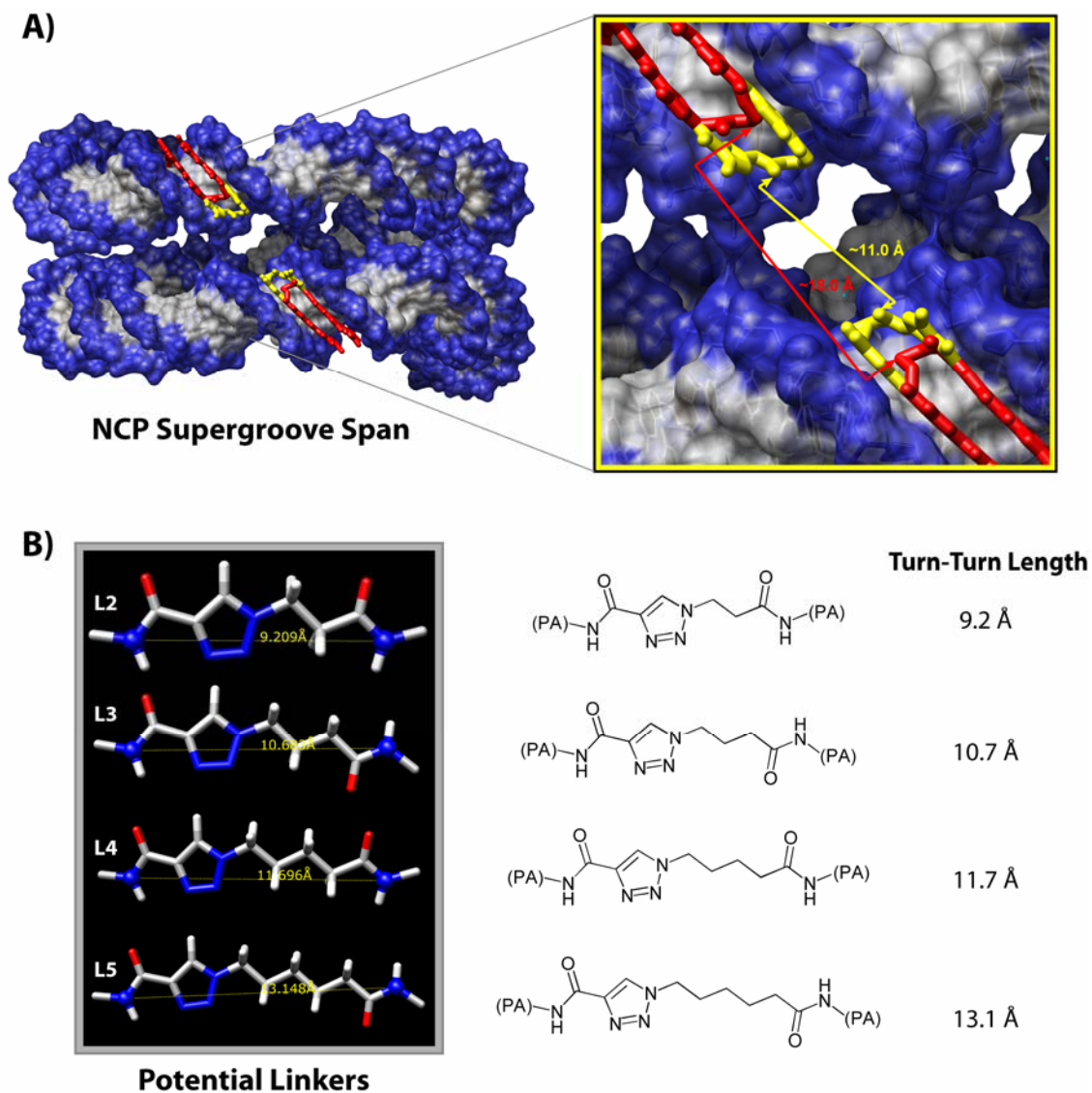


Figure 4.12 Analysis of the linker dependence of NCP templated ligations. A) Crystal structure view of the supergroove. The predicted distance between the amines of the turn are shown in yellow for binding at the original site, and in red if the sites are moved one base pair back. B) Modeling of the linkers and the calculated turn-to-turn length.

followed MALDI-TOF mass spectrometry to analyze the reaction. Of the six potential ligation reactions, we only observed product in the case of compounds **23** and **24**. This agrees with both the computational work that indicated the linkers of compounds **21** and **22** were too short to react, as well as with the prior work that showed that the alkyne used in **25** was significantly less reactive. In the case of **23** and **24** we observed masses at 2664.88 and 2687.77 that correspond to calculated values for the $[M+H]^+$ and $[M+Na]^+$ peaks.

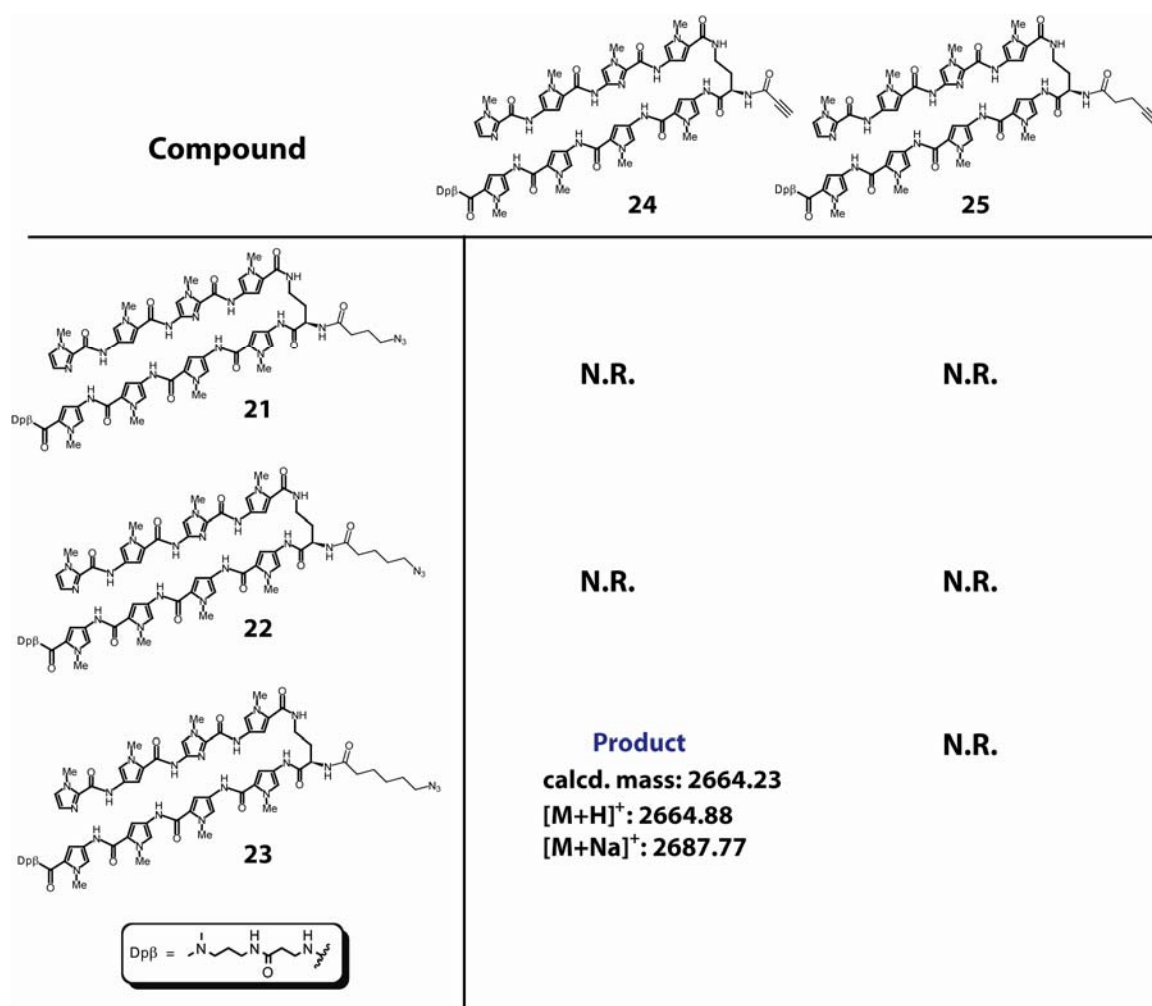


Figure 4.13 NCP-templated ligation between different length azides and alkynes. The reaction product was only detected following reaction with polyamides **23** and **24**. N.R. = no reaction.

We then checked whether control samples run under the same conditions but containing either the DNA, histone octamer, or buffer instead of the NCP would result in product formation. In all three cases the target mass was not observed, indicating that the ligation reaction was dependent upon the presence of the NCP. The lack of product in the presence of just DNA is promising, as the reaction appears to only take place when the

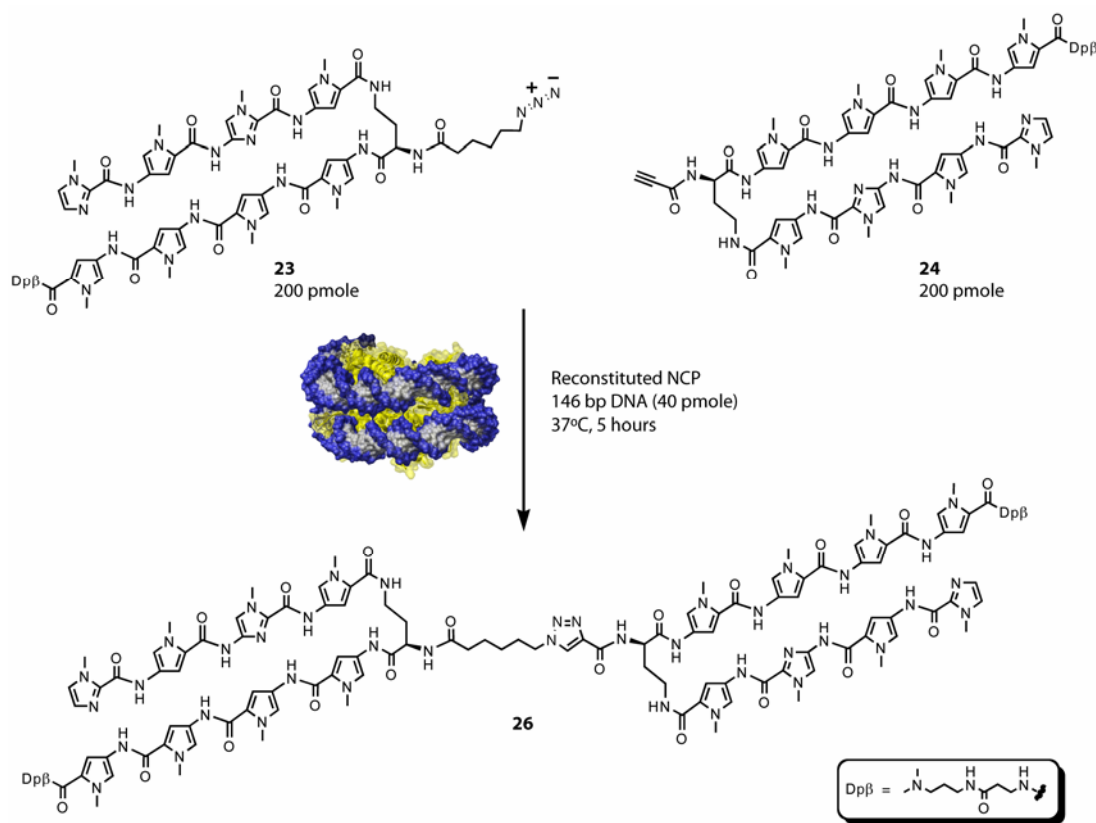


Figure 4.14 DNA templated ligation on the NCP. The reaction of **23** and **24** with the NCP leads to dimer **26** which was observed by MALDI mass spectrometry.

two binding sites which are spatially separated on the linear DNA are brought into close proximity by NCP formation to generate a functional template.

These results demonstrate the ability to perform templated ligation reactions on the NCP. Although we were able to detect product formation by MALDI, the small scale

further understanding of the ligation reaction and to determine the feasibility of performing templated reactions in an intracellular setting.

4.3 Conclusions

The use of polyamides as unique molecular tools for performing reactions and manipulating the NCP has been described in detail. The efforts in designing an NCP-specific clamp have led to greater understanding of the binding mode of polyamide dimers. The design and synthesis of a new polyamide dimer end clamp may prove to be a useful tool for X-ray crystallography. Finally, the NCP templated ligation studies have demonstrated the feasibility of this approach for generating linked polyamide dimers. One key question is whether the knowledge gained through these studies can be applied to intracellular systems. The use of polyamide clamps to stabilize the NCP, and the ability to generate these clamps via a templated ligation reaction inside the cell, has exciting potential applications for gene regulation.

4.4 Experimental Details

Materials and General Methods. Dicyclohexylcarbodiimide (DCC), 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), N-Hydroxybenzotriazole (HOBt), Fmoc-Dab(Boc)-OH and Boc- β -Ala-Pam resin were purchased from Peptides International. Benzotriazole-1-yl-oxy-tris-pyrrolidino-phosphonium hexafluorophosphate (PyBOP) was purchased from Novabiochem. Flouoro-N,N,N',N'-tetramethylformamidinium hexafluorophosphate (TFFH) was purchased from Advanced ChemTech. O-(7-Azabenzotriazol-1-yl)-N,N,N',N'-tetramethyluronium

hexafluorophosphate (HATU), 4-(dimethylamino)-pyridine (DMAP), N-hydroxysuccinimide (NHS), N,N-dimethylformamide (DMF), N-methylpyrrolidinone (NMP), N,N-dimethylpropylamine (Dp), N,N-diisopropylethylamine (DIEA), ethylene diamine, piperidine, and other miscellaneous chemicals were purchased from Aldrich and used without further purification. All other solvents were purchased from EM Sciences and were reagent grade. Trifluoroacetic acid (TFA) was purchased from Halocarbon.

^1H NMR spectra were recorded using a 300MHz General Electric-QE NMR spectrometer. CDCl_3 was obtained from Cambridge Isotope Laboratories. UV spectra were recorded in water using a Beckman Coulter DU 7400 Spectrophotometer. Matrix-assisted LASER desorption/ionization time of flight mass spectrometry (MALDI-TOF MS) was performed using an Applied Biosystems Voyager DE Pro Spectrometer. Electrospray ionization (ESI) mass spectrometry was performed by the Protein and Peptide Microanalytical Facility at the California Institute of Technology. Analytical High-Pressure Liquid Chromatography (HPLC) was performed with a Beckman Gold system using a Varian Microsorb-MV 100 C_{18} column (5 μm particle size, 250 x 4.6mm). Preparative HPLC was performed using a Beckman Gold system with either a Waters Bondapak C_{18} column (15-20 μm particle size, 25 x 100mm) or a Phenomenex Gemini C_{18} column (5 μm particle size, 250 x 21.2 mm). For both HPLC systems Solvent A was 0.1% (v/v) aqueous TFA and solvent B was acetonitrile. Analytical HPLC was done using a gradient of 1.85%/min of Solvent B starting from 0% over 35 min with a flowrate of 1.5 mL/min. Preparative HPLC was typically done using a gradient of 1%/min of Solvent B for 20 min followed by a gradient of .3%/min for an additional 100 min at a

flowrate of 8 mL/min. Radioactive gels were imaged using a Molecular Dynamics 400S PhosphorImager.

Restriction endonucleases, deoxyribonucleotide triphosphates, DNase I, Polynucleotide kinase (PNK), and glycogen were purchased from Roche. [α - ^{32}P]-Thymidine-5'-triphosphate (≥ 3000 Ci/mmol) and [α - ^{32}P]-Deoxyadenosine-5'-triphosphate (≥ 6000 Ci/mmol) were purchased from Perkin Elmer. [γ - ^{32}P]-Adenosine-5'-triphosphate (≥ 7000 Ci/mmol) was purchased from MP Biomedicals. Water was purified from a Millipore Mill-Q purification system for general use. Ultrapure RNase/DNase free water from USB was used for biological work. All buffer reagents used were molecular biology grade. Buffers were sterilized using a Nalgene .2 μm cellulose filtration device.

Plasmids pJDC1 and pJDC2 were constructed using 80mer oligonucleotides purchased from Integrated DNA Technologies. pUC19 plasmid was purchased from Sigma. JM109 Competent Cells ($>10^8$ cfu/ μg) were purchased from Promega. A Rapid DNA ligation kit was purchased from Roche. Purification was done using a Promega Wizard *Plus* Midipreps DNA purification system. Sequence analysis was performed by the Sequence Analysis Facility at the California Institute of Technology.

Polyamide Synthesis. Polyamide synthesis was performed as previously reported.¹⁶ All polyamides were synthesized using Boc- β -Ala-PAM resin (~ 59 meq/g). The resin was initially swelled in DMF for 5 min in a glass reaction vessel fitted with a glass filter and stopcock. The vessel was drained and the resin washed twice with DCM. Deprotection with 80% TFA:DCM was performed for 20 min while the resin was shaken. Following deprotection, the resin was washed 2 x DCM, 1 x 4:1 DMF:DIEA, and 1 x DMF.

Coupling of the Boc-Py-OBt pre-activated ester was performed using 1.8 eq of monomer in ~1 mL of NMP. Coupling of Boc-Im-OH, Boc-PyIm-OH, and other acids was done by first preactivating 1.8 eq of the monomer with 1.7 eq of HBTU, 5.4 eq of DIEA in ~2mL of NMP. The activation mixture was shaken for 20min before being filtered and added to the resin. Couplings were allowed to proceed for 2hrs except in the case of Im-OH which was allowed to react overnight. After each coupling step the resin was washed 2 x DMF and then 2 x DCM before the next deprotection step. Polyamides were cleaved from resin using 1.5 mL of Dp for 200 mg of resin at 55°C for eight hours. Crude products were purified by preparative HPLC.

1A: MALDI-TOF-MS calculated $[M+H]^+$: 1237.59, found 1237.5

1: MALDI-TOF-MS calculated $[M+H]^+$: 2617.20, observed 2617.3

2: MALDI-TOF-MS calculated $[M+H]^+$: 2724.23, observed 2724.3

3: MALDI-TOF-MS calculated $[M+H]^+$: 2657.21, observed 2657.2

4: MALDI-TOF-MS calculated $[M+H]^+$: 2661.24, observed 2661.1

5: MALDI-TOF-MS calculated $[M+H]^+$: 2705.21, observed 2704.9

6: MALDI-TOF-MS calculated $[M+H]^+$: 2753.28, observed 2753.3

7: MALDI-TOF-MS calculated $[M+H]^+$: 2645.23, observed 2645.2

8: MALDI-TOF-MS calculated $[M+H]^+$: 2645.23, observed 2645.8

9: MALDI-TOF-MS calculated $[M+H]^+$: 1279.60, observed 1279.58

10: MALDI-TOF-MS calculated $[M+H]^+$: 1410.66, observed 1410.62

11: MALDI-TOF-MS calculated $[M+H]^+$: 1539.7, observed 1540.01

12: MALDI-TOF-MS calculated $[M+H]^+$: 1739.8, observed 1740.0

13: MALDI-TOF-MS calculated $[M+H]^+$: 1739.8, observed 1739.8

- 14:** MALDI-TOF-MS calculated $[M+H]^+$: 1571.7, observed 1571.89
- 15:** MALDI-TOF-MS calculated $[M+H]^+$: 1740.8, observed 1740.8
- 16:** MALDI-TOF-MS calculated $[M+H]^+$: 1423.66, observed 1423.8
- 17:** MALDI-TOF-MS calculated $[M+H]^+$: 1381.65, observed 1381.86
- 18:** MALDI-TOF-MS calculated $[M+H]^+$: 1507.72, observed 1508.63
- 19:** MALDI-TOF-MS calculated $[M+H]^+$: 1465.71, observed 1465.71
- 20:** MALDI-TOF-MS calculated $[M+H]^+$: 3033.41, observed 3032.9
- 21:** MALDI-TOF-MS calculated $[M+H]^+$: 1348.63, observed 1348.54
- 22:** MALDI-TOF-MS calculated $[M+H]^+$: 1362.65, observed 1362.61
- 23:** MALDI-TOF-MS calculated $[M+H]^+$: 1376.66, observed 1376.73
- 24:** MALDI-TOF-MS calculated $[M+H]^+$: 1289.58, observed 1289.67
- 25:** MALDI-TOF-MS calculated $[M+H]^+$: 1317.61, observed 1317.6
- 26:** MALDI-TOF-MS calculated $[M+H]^+$: 2664.23, observed 2664.88

Synthesis of diacids. The following diacids were synthesized for use in homodimer clamp synthesis.

3,3' – (1,3-Phenylene)bisacrylic Acid The diacid was synthesized as previously described.¹⁷ From 500mg (3.7mmol) of isophthalaldehyde was yielded 631mg (78%) of the diacid as an off white solid. ¹H NMR (DMSO-*d*₆) δ 12.4 (s, 1H), 8.05 (s, 1H), 7.7 (d, 2H), 7.6 (d, 2H), 7.41 (t, 1H), 6.61 (d, 2H).

3,3' – (1,3-Phenylene)bispropanoic Acid The diacid was synthesized as previously described.^{17, 18} From 400mg (1.83mmol) of the bisacrylic Acid from above was yielded

335mg (82.4%) of the diacid as a white solid. ^1H NMR (CDCl_3) δ 7.2 (m, 4H), 2.9 (t, 4H), 2.65 (t, 4H). ESI MS m/e: $[\text{M}-\text{H}]^-$ 221.0 calc'd: 221.08.

(R,S) 4,5-dimethyl-3,6-dioxaoctanedioic diethyl ester The compound was prepared as previously described.¹¹ From 500mg (5.55mmol) of meso 2,3-butanediol was yielded 1.3g (90%) of the diethyl ester. ^1H NMR (CDCl_3) δ 4.2 (m, 8H), 3.7 (m, 2H), 1.3 (t, 6H), 1.2 (d, 6H). ESI MS m/e: $[\text{M}+\text{H}]^+$ 263.0 calc'd: 263.15.

(R,S) 4,5-dimethyl-3,6-dioxaoctanedioic acid The compound was prepared as previously described.¹¹ From 178mg (.68mmol) of **1B** was yielded 32mg (23%) of the diethyl ester. ^1H NMR (CDCl_3) δ 4.33 (s, 1H), 4.28 (s, 1H), 4.13 (s, 1H), 4.08 (s, 1H), 3.7 (m, 2H), 1.2 (d, 6H). ESI MS m/e: $[\text{M}-\text{H}]^-$ 204.8.0 calc'd: 205.07.

(R,R) 4,5-dimethyl-3,6-dioxaoctanedioic diethyl ester The compound was prepared as previously described.¹¹ From 500mg (5.55mmol) of meso 2,3-butanediol was yielded 1.5g (99%) of the diethyl ester. ^1H NMR (CDCl_3) δ 4.2 (m, 8H), 3.6 (m, 2H), 1.28 (t, 6H), 1.2 (d, 6H). ESI MS m/e: $[\text{M}+\text{Na}]^+$ 284.8 calc'd: 285.13.

(R,R) 4,5-dimethyl-3,6-dioxaoctanedioic acid The compound was prepared as previously described.¹¹ From 847mg (3.2mmol) of **1B** was yielded 600mg (90%) of the diethyl ester. ^1H NMR (CDCl_3) δ 4.35 (s, 1H), 4.3 (s, 1H), 4.15 (s, 1H), 4.08 (s, 1H), 3.45 (m, 2H), 1.15 (d, 6H).

Synthesis of Polyamide Dimers. All homodimers were synthesized using the same general protocol. Briefly, the diacid was first activated in 200 μL of anhydrous DMF with 5 eq. of PyBOP and 10 eq. of HOBt. 100 μL of DIEA was added, followed by addition of Polyamide **1A** in slight excess. All reactions were shaken for 2 hrs and shown to be

complete by analytical HPLC. Preparative HPLC workup was used to isolate the final product.

Heterodimers were synthesized using a similar protocol. The first polyamide was initially coupled using the same conditions as previously described for homodimers except that the linker diacid was added in >10 fold excess to afford solely the mono-coupled product. Reactions were monitored by analytical HPLC and worked up after 2 hrs using a preparative HPLC. Next, the polyamide with linker attached is taken up in 100 μ L anhydrous DMF and reactivated with 5 eq. of PyBOP and 10 eq. HOBt. 100 μ L of DIEA is added to the mixture, followed by addition of the second polyamide in slight excess. Reactions were monitored by HPLC and in most cases were allowed to react overnight as the observed rate of product formation was slow in comparison to the initial coupling. In some cases another addition of PyBOP was necessary to get the reaction to proceed. Products were isolated using a preparative HPLC as previously described.

EDTA couplings. Polyamide-EDTA conjugates were prepared by heating 50 mg of EDTA dianhydride in 1 mL of DMSO / NMP for 5 min. Approximately 1.5 μ mol of polyamide was added with 500 μ L of DIEA and the sample heated for 37 min at 60°C. 3 mL of .1 M NaOH was added to the sample followed by heating for an additional 15 min. Samples were then purified by HPLC.

Preparation of 3' 32 P-labeled DNA. Plasmid pJDC1 was initially digested using the restriction enzymes *PvuII* and *EcoRI*. The 3' overhang regions were filled in with [α - 32 P]-dATP and [α - 32 P]-dTTP using the Klenow fragment of DNA polymerase or

Sequenase. Labeled oligonucleotides were then purified by band extraction from a 7% non-denaturing polyacrylamide gel.

Preparation of 5' ³²P-labeled DNA 5' end labeling was accomplished by incubation of a PCR forward primer, PNK, and an aliquot of [γ -³²P]-ATP in PNK buffer at 37°C for 30 min. Labeled primer was then used in a PCR reaction to amplify the selected region on either plasmid pJDC1 or pJDC2. The radiolabelled PCR product oligonucleotides were then purified by band extraction from a 7% non-denaturing polyacrylamide gel.

Affinity determination by quantitative DNase I footprinting. Reactions were carried out in a volume of 400 μ L in aqueous TKMC buffer according to published protocols.¹⁹ Standard molecular biology techniques were used to insert a 75bp DNA sequence into the *Bam*HI/*Hind*III restriction site of pUC19.²⁰ This plasmid was used to generate a 5' ³²P labeled 283bp *Eco*RI/*Pvu*II restriction fragment which was used for all footprinting experiments. Developed gels were imaged using storage phosphor autoradiography using a Molecular Dynamics 400S Phosphorimager. Equilibrium association constants were determined as previously described.¹⁹

Reconstitution of the NCP. The NCP was reconstituted following established protocols. Briefly, 10 μ g of the 146 bp DNA was 5' radiolabelled as described above using Polynucleotide Kinase and purified using a Chroma Spin STE 10 column from BD Biosciences. Histone octamer which was obtained from Karolin Luger's lab at Colorado State University was diluted to 1 μ g / μ L using 2 M NaCl TE Buffer. 5 μ L of the DNA

was added to 5 μL of 4M NaCl, 10 mM Tris, 1 mM EDTA solution to make a sample of 10 μL of DNA in 2 M NaCl TE buffer. Five samples were made up and 0, .6, .8, 1.0, or 1.2 μL of octamer (1 $\mu\text{g} / \mu\text{L}$) was added to each sample. The total volume was brought to 12 μL using 2 M NaCl TE buffer. The samples were incubated for 1 hr before the addition of 12 μL , 6 μL , 6 μL , and 84 μL of dilution buffer (TE) with each addition followed by a 1 hr incubation. The sample was heated to 37°C for 2 hrs before a final addition of 120 μL of dilution buffer to give a final NaCl concentration of .1 M.

Reconstituted NCP samples were stored at 4°C. Non-denaturing PAGE was done on a 6% polyacrylamide TBE gel. 1 μL of each sample was diluted to 10 μL with 10 mM Tris, 20 mM NaCl, 10% glycerol, and .1% Igapal for loading. The gel was run at 150 V for 20 min, dried and imaged.

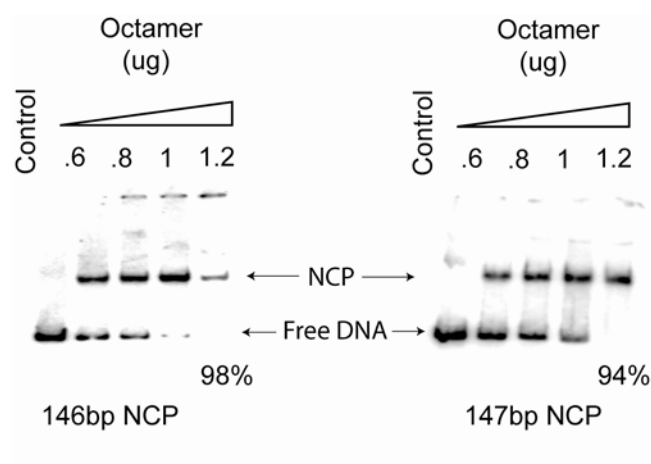


Figure 4.16 Representative gels of NCP reconstitution. The amount of histone octamer is titrated into a constant amount of the 146 or 147 bp radiolabelled DNA. Non-denaturing polyacrylamide gels are used to determine the yield from reconstitution.

Affinity Cleavage on the NCP. Affinity cleavage experiments were performed at The Scripps Research Institute under the guidance of Joel Gottesfeld. The previously published protocols were used.⁴

Melting Temperature Determinations. The melting temperature analysis was done following previously reported protocols.²¹ The analysis was performed using a Varian Cary 100 spectrophotometer equipped with a thermocontrolled cell holder and possessing a cell path length of 1 cm. A degassed aqueous solution of 10 mM sodium cacodylate, 10 mM KCl, 10 mM MgCl₂, and 5 mM CaCl₂ at pH 7.0 was used as the analysis buffer. The buffer conditions were designed to mimic those of DNase I footprinting experiments as closely as possible. DNA duplexes and hairpin polyamides were mixed in 1:1 stoichiometry to a final concentration of 2 μ M. Samples were initially heated to 90 °C and cooled to a starting temperature of 25°C with a heating rate of 5 °C/min for each ramp. Two ramps were performed during each experiment and each experiment was run at least twice. Denaturation profiles were recorded at 260 nm from 25 to 90 °C with a heating rate of 0.5 °C/min. The reported melting temperatures were defined as the maximum of the first derivative of the denaturation profile.

NCP Templated Ligation Reactions. The NCP was reconstituted as described above with the sole difference that non-radiolabelled DNA was used. For each ligation reaction 200 pmol of each polyamide was lyophilized into an eppendorf tube and 40 pmol of the reconstituted NCP was added. Samples containing only the 146 bp DNA, the histone octamer, or buffer alone were prepared as negative controls. The reaction was incubated for 5 hrs at 37°C before MALDI-TOF analysis.

4.5 References

1. Dervan, P. B.; Edelson, B. S., Recognition of the DNA minor groove by pyrrole-imidazole polyamides *Curr. Opin. Struct. Biol.* **2003**, 13, 284-299.
2. Cohen, J. D.; Sadowski, J. P.; Dervan, P. B., Addressing single molecules on DNA nanostructures *Angew. Chem.-Int. Edit.* **2007**, 46, 7956-7959.
3. Cohen, J. D.; Sadowski, J. P.; Dervan, P. B., Programming multiple protein patterns on a single DNA nanostructure *J. Am. Chem. Soc.* **2008**, 130, 402-403.
4. Gottesfeld, J. M.; Melander, C.; Suto, R. K.; Raviol, H.; Luger, K.; Dervan, P. B., Sequence-specific recognition of DNA in the nucleosome by pyrrole-imidazole polyamides *J. Mol. Biol.* **2001**, 309, 615-629.
5. Suto, R. K.; Edayathumangalam, R. S.; White, C. L.; Melander, C.; Gottesfeld, J. M.; Dervan, P. B.; Luger, K., Crystal structures of nucleosome core particles in complex with minor groove DNA-binding ligands *J. Mol. Biol.* **2003**, 326, 371-380.
6. Edayathumangalam, R. S.; Weyermann, P.; Gottesfeld, J. M.; Dervan, P. B.; Luger, K., Molecular recognition of the nucleosomal "super groove" *Proc. Natl. Acad. Sci. U. S. A.* **2004**, 101, 6864-6869.
7. Herman, D. M.; Baird, E. E.; Dervan, P. B., Tandem hairpin motif for recognition in the minor groove of DNA by pyrrole - Imidazole polyamides *Chem.-Eur. J.* **1999**, 5, 975-983.
8. Kers, I.; Dervan, P. B., Search for the optimal linker in tandem hairpin polyamides *Bioorg. Med. Chem.* **2002**, 10, 3339-3349.
9. Weyermann, P.; Dervan, P. B., Recognition of ten base pairs of DNA by head-to-head hairpin dimers *J. Am. Chem. Soc.* **2002**, 124, 6872-6878.
10. Poulin-Kerstien, A. T.; Dervan, P. B., DNA-templated dimerization of hairpin polyamides *J. Am. Chem. Soc.* **2003**, 125, 15811-15821.

11. Ammann, D.; Bissig, R.; Guggi, M.; Pretsch, E.; Simon, W.; Borowitz, I. J.; Weiss, L., Preparation of neutral ionophores for alkali and alkaline-earth metal cations and their application in ion-selective membrane electrodes *Helv. Chim. Acta* **1975**, 58, 1535-1548.
12. Trauger, J. W. Ph. D. Thesis. CalTech, 1998.
13. Rucker, V. C.; Melander, C.; Dervan, P. B., Influence of beta-alanine on hairpin polyamide orientation in the DNA minor groove *Helv. Chim. Acta* **2003**, 86, 1839-1851.
14. Foister, S.; Marques, M. A.; Doss, R. M.; Dervan, P. B., Shape selective recognition of T center dot A base pairs by hairpin polyamides containing N-terminal 3-methoxy (and 3-chloro) thiophene residues *Bioorg. Med. Chem.* **2003**, 11, 4333-4340.
15. Edayathumangalam, R. S.; Weyermann, P.; Dervan, P. B.; Gottesfeld, J. M.; Luger, K., Nucleosomes in solution exist as a mixture of twist-defect states *J. Mol. Biol.* **2005**, 345, 103-114.
16. Baird, E. E.; Dervan, P. B., Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids *J. Am. Chem. Soc.* **1996**, 118, 6141-6146.
17. Fukuda, Y.; Seto, S.; Furuta, H.; Ebisu, H.; Oomori, Y.; Terashima, S., Novel seco cyclopropa[c]pyrrolo [3,2-e]indole bisalkylators bearing a 3,3 '-arylenebisacryloyl group as a linker *J. Med. Chem.* **2001**, 44, 1396-1406.
18. Schimelpfenig, C. W., Synthesis of oxometacyclophanes with Dieckmann condensation *J. Org. Chem.* **1975**, 40, 1493-1494.
19. Trauger, J. W.; Dervan, P. B., Footprinting methods for analysis of pyrrole-imidazole polyamide/DNA complexes *Methods Enzymol.* **2001**, 340, 450-466.
20. Sambrook, J.; Fritsh, E. F.; Maniatis, T., *Molecular Cloning: Standard Protocols for DNA Manipulation. A laboratory Manual 2nd ed.* Cold Spring Harbor Laboratory: Plainview, NY, 1989.

21. Dose, C.; Farkas, M. E.; Chenoweth, D. M.; Dervan, P. B., Next generation hairpin polyamides with (R)-3,4-diaminobutyric acid turn unit *J. Am. Chem. Soc.* **2008**, 130, 6859-6866.

Appendix A

Reducing Algorithmic Assembly Error Rates using Polyamides

Abstract

The use of DNA self-assembly to perform molecular computation is a promising field of research. DNA-based computation has the potential to perform large numbers of computations in parallel in a single solution. To this end, DX crystal growth can be designed to follow a series of algorithmic instructions. In these types of arrays, a nucleating strand functions to give a series of logical inputs, and crystal growth is determined by a set of designed logic rules implemented by a series of DX tiles with varying sticky ends. The error rate for tile incorporation is extremely important for these applications, as a single incorrectly incorporated tile can result in a cascade of errors, as subsequent binding events will be determined based on the erroneous tile. Prior work has required a subset of the DX tiles to be modified with dsDNA hairpins for AFM visualization, which may inadvertently affect the kinetics of binding and thus the error rate. By redesigning this system so that none of the tiles contain DNA hairpins, it may be possible to reduce the error rate. Polyamide-biotin conjugates targeted to specific sequences could be used to visualize a subset of tiles after array formation resulting in lower error rates in array assembly.

A.1 Introduction

The use of 2-dimensional DNA arrays for the creation of nanoscale assemblies was detailed in Chapter 2.¹⁻⁴ DX arrays, made up of individual DNA tiles that repeat in a predictable manner through a series of programmed sticky end interactions can be used to create 2-dimensional assemblies. However, the periodic nature of these arrays limits the amount of complexity that can be built into these assemblies. One method for breaking the periodicity of these structures is algorithmic assembly. In algorithmic assembly, the interactions that govern crystal formation are determined by a series of designed algorithms, which results in non-periodic arrays. A set of inputs or starting conditions result in an “assembly program” being executed by the various DNA tiles, resulting in programmed array formation. The ability to perform copy and count functions using algorithmic assembly has recently been demonstrated.⁵ It is likely that assembly programs governed by DNA interactions will be important for applications such as DNA-based computing.

The algorithmic assembly concept has been demonstrated by the Winfree lab using a series of DX tiles in combination with a single stranded DNA input strand.⁶ Individual DX tiles function as logical operators with binding of each tile being determined by a set of logic rules. Each DX tile acts as a logic gate, where two of the sticky ends act as inputs, and the other two act as outputs. The binding of the tile results in a single logical computation. For example, to compute the XOR function for the inputs 0 and 1, a tile would be designed with 2 input strands that correspond to the 0 and 1 and two output strands that correspond to 1. Appropriate tiles are designed for each set of input and output conditions. Tiles that output 1 are modified to contain a dsDNA hairpin

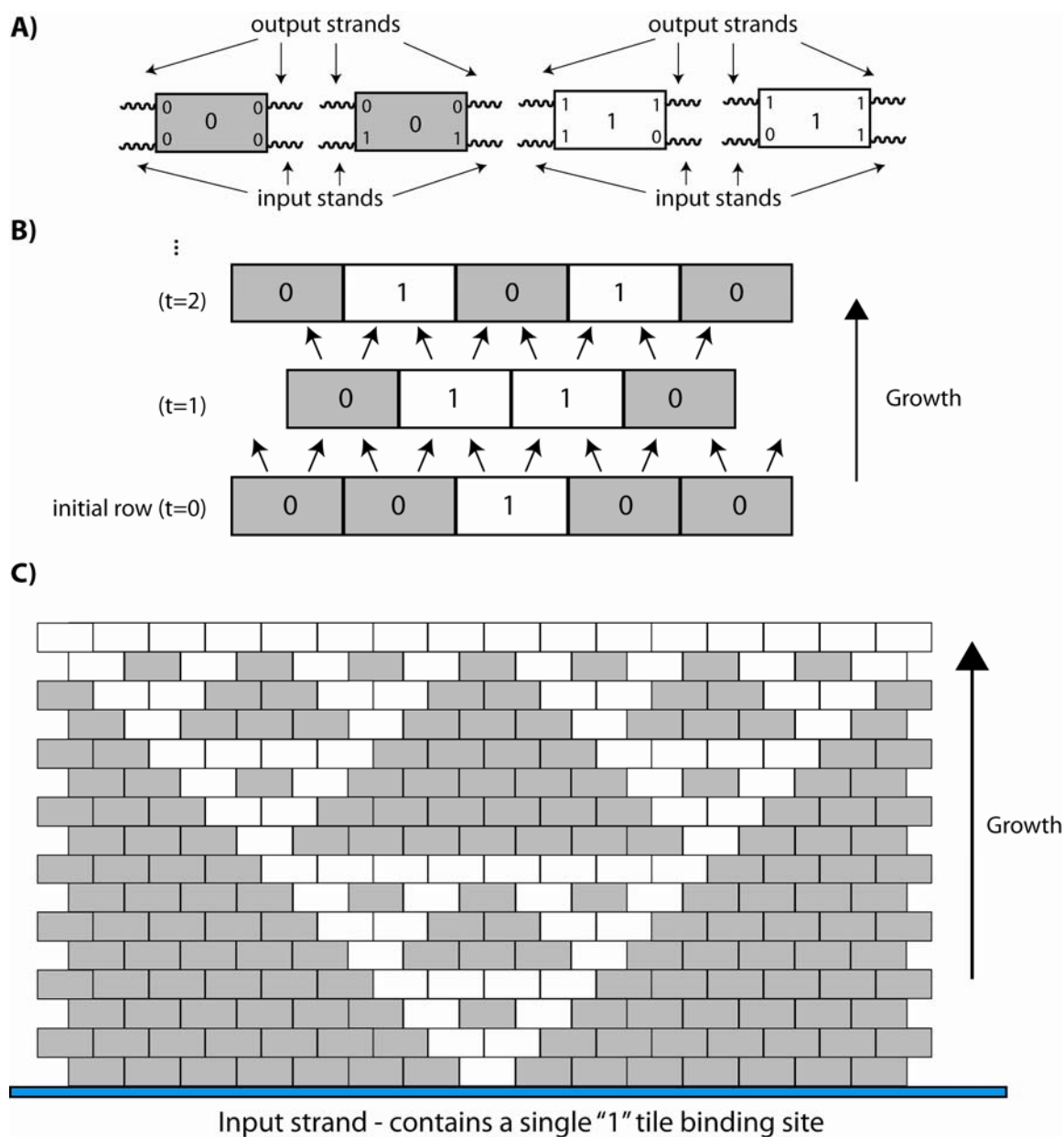


Figure A.1 Overview of algorithmic assembly. A) DX tiles corresponding to the logic function XOR are shown. Each tile contains two input and two output strands that determine binding in a growing array. B) Crystal growth proceeds from an input strand in a manner consistent with the designed logic rules. At each step, future binding events are determined by the previously bound tiles. C) The XOR tile set gives rise to the Sierpinski triangle. The input condition is set by a ssDNA template containing a single "1" binding site surrounded by "0" binding sites.

while tiles that output 0 are left unmodified. The resulting differences in height between tiles can be observed using an atomic force microscope (AFM). An input strand consisting of ssDNA with a binding site for a "1" tile surrounded by a large number of

“0” tiles is used to seed crystal growth. When DX tiles designed to compute the XOR function are used, the algorithm results in the pattern known as the Sierpinski triangle.

One of the hurdles in this field is the error rates for tile binding. In periodic arrays, the sticky ends of tiles are designed such that a mismatch would occur at both ends of any tile which attempted to bind at the incorrect site. However, in the case of algorithmic assembly, many tiles may contain single match and single mismatch domains. If the rate of dissociation for these single match / single mismatch tiles is sufficiently slow, this can lead to incorporation of incorrect tiles in a growing array. Unfortunately, due to the nature of algorithmic growth, a single error can propagate throughout the array, as an incorrect tile will influence future binding events leading to a cascade effect.⁶ One of the factors that could possibly influence the kinetics of binding is the incorporation of dsDNA loops into specific tiles which have been used as a visual marker to distinguish logical ones from zeroes. If these dsDNA loops were removed, it may improve the error rates by eliminating any artificial disparities in the binding kinetics between labeled and unlabeled tiles. The use of dsDNA hairpins also means that in order to visualize different tiles in an array, something that is often desirable in algorithmic assembly, it would require constructing structurally different versions of the array, with the appropriate modification transferred from one tile to another.

Given our previous success in non-covalently labeling DX tiles,^{7, 8} we proposed that polyamide labeling could provide a method for reducing the error rate in algorithmic assembly. Instead of the dsDNA hairpin modification, a simple polyamide binding site sequence would be incorporated into the tiles that contain a logical 1 output, while the tiles that have a logical zero are designed not to contain the binding site sequence. The

assembly should take place with a reduced error rate since all of the DX tiles used lack any structural modification and should be incorporated at nearly equal rates. After the assembly has formed, visualization is performed by incubation with a polyamide-biotin conjugate and streptavidin and AFM imaging.

A.2 Results and Discussion

The initial design of the project was based upon the originally reported arrays using a DAO-E type tile.⁶ Polyamide-biotin conjugates **1** and **2** were previously used for DX array labeling⁸ and they are shown along with their target sequences in Figure A.2A. The use of two polyamide-biotin conjugates makes it possible to visualize either the “1” or the “0” tiles simply by changing the polyamide that is used. If successful, the polyamide targeted to the “1” tiles would give the Sierpinski triangle, while the “0” tile polyamide would give an inverse image. Next, the six DX tiles used to create the original array were redesigned. First, the hairpins were removed and the tiles were examined to eliminate any unintentional polyamide binding sites. Three match sites were present in the original tiles for polyamide **1** but their locations at the junctions would prevent any spurious binding. Polyamide **3** had two match sites in the original tiles which were subsequently removed in the newly designed tiles. Each of the new tiles contains two match sites for the desired polyamide in order to maximize potential binding and labeling yields. The sites are designed to face opposite faces to ensure that at least one site is always accessible independent of whether the array lands face up or face down. Nondenaturing polyacrylamide gel electrophoresis showed that each of the new tiles formed properly. Finally, the ssDNA template which acts as a nucleating complex was synthesized as previously described.⁶

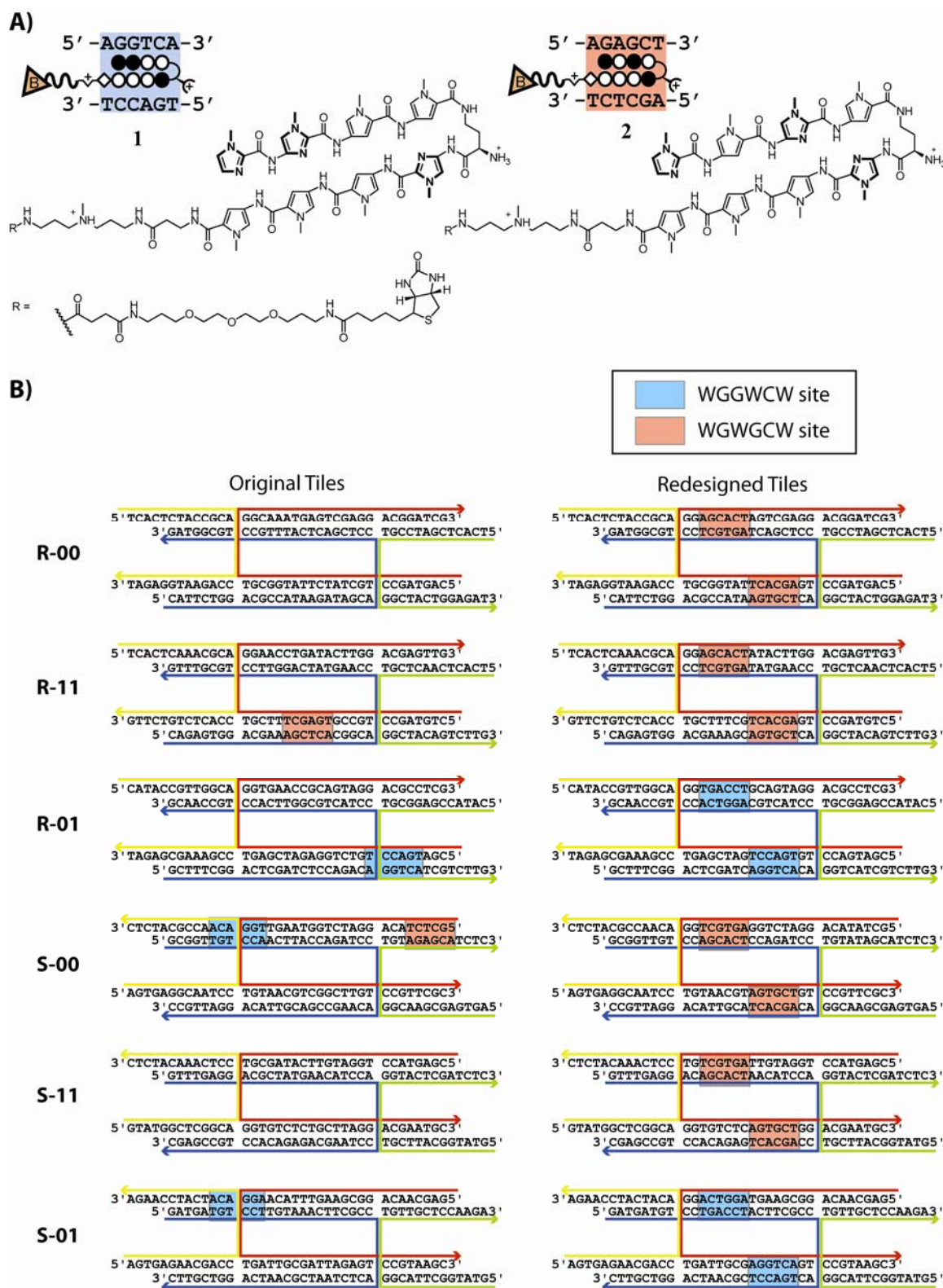


Figure A.2 Polyamide-biotin conjugates and DX tiles for algorithmic assembly. A) Two polyamide-biotin conjugates were designed to be able to visualize either the “1” or “0” tiles in the array. B) The original tiles

were redesigned to contain two binding sites for a single polyamide. Tiles that output a “1” have sites for **1** while tiles that output a “0” have sites for **2**.

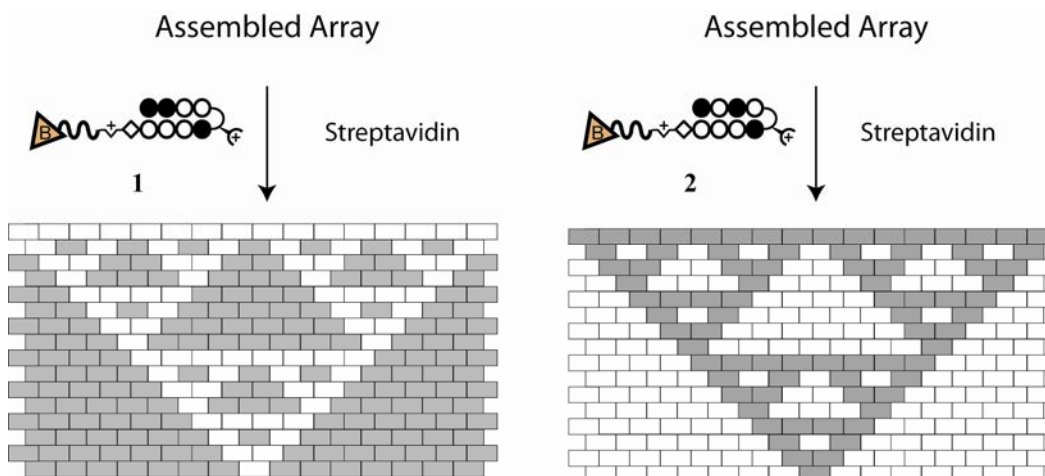


Figure A.3 Labeling an array formed by algorithmic self-assembly with polyamide-biotin conjugates. After the array is formed it is treated with streptavidin and a polyamide-biotin conjugate targeted to a subset of tiles, either those that output a 1 or a 0. The height difference between the unmodified tile and the tiles to which streptavidin has been recruited is visible by AFM.

AFM Imaging

The annealed array was incubated with polyamide **1** and streptavidin. AFM images of the assembly are shown in Figure A.4. Although somewhat difficult to distinguish, it appears as though the expected Sierpinski triangle pattern is visible. However, the large amount of labeling observed exceeds what would be expected and makes absolute determinations of the binding pattern difficult. One possible explanation is that the polyamide is binding to a second mismatch site or nonspecifically to tiles that it should not. However, this should be unlikely considering the design of the tiles and the efforts taken to eliminate potential mismatch sites. Another potential explanation is that the nucleating DNA strand contains too many initiation sites that are resulting in the formation of overlapping triangles. This is consistent with the large number of binding events observed near the edges of the arrays. Finally, a large number of structures

corresponding to “spontaneous nucleation” are visual as well. These structures, rather than forming in a controlled manner from an input template strand result from spontaneous crystal growth. The arrays from this sort of growth result in a seemingly random pattern although the tiles may still obey the programmed binding rules.

Algorithmic Arrays + 1 + Streptavidin

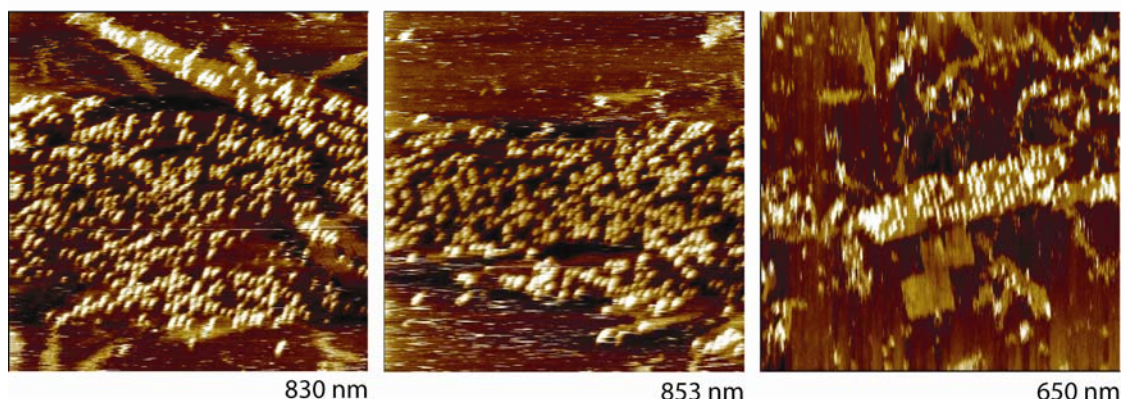


Figure A.4 AFM images of algorithmic arrays labeled with polyamide **1**. Images on the left and middle were taken after the arrays were incubated with 200 nM **1**. 2.5 μ L of sample was spotted and 2 μ L of 1 μ M streptavidin was added 1 min prior to imaging. The image on the far right was taken of annealed arrays that were incubated with 100 nM **1** and 200 nM streptavidin prior to being spotted on mica.

A.3 Conclusions

This initial study into polyamide use for reducing error rates in algorithmic assemblies demonstrated several challenges that will need to be met before this methodology could become practical. First, extremely high levels of labeling are essential, as any site which is unbound by polyamide will result in an apparent error, despite successful incorporation of the correct tile into the array. Similarly, polyamide binding must be specific for one tile versus the other. If a polyamide binds to the wrong tile, this will again result in an apparent error, despite successful incorporation of the correct tile into the growing array. For this method to have a significant advantage over the use of dsDNA hairpins, polyamide labeling would likely have to approach >95%

efficiency and specificity, as the error rates reported in previous arrays have been between 1 – 10%.^{5,6} Based upon the initial observed results in this study, as well as those in Chapter 2 and 3, it is unlikely that the current generation of pyrrole-imidazole polyamides are capable of meeting these stringent requirements.

It should be noted that one of the limitations in this study was the input strand used to generate the assemblies. The process used to generate the input strands uses a mixture of strands present in different concentrations to limit the incorporation of initiation sites. (See ref. 6, Supporting Information). The observation of a large number of spontaneous growth arrays as well as what appeared to be multiple initiation sites in the templated arrays indicates that the quality of the input strand may have hindered the growth of high quality arrays. It is worth noting that more recent studies have switched to the use of DNA origami as a programmable input for crystal growth.⁹ Reported error rates in these assemblies were just 1.4% per tile. Although polyamide-biotin conjugates may eventually find use for the labeling purposes outlined here, it is more likely that error-correcting assemblies and the evolving platforms for algorithmic assemblies may effectively eliminate the need for error correction in this fashion. Despite this, the results of this study highlight the continued need for the generation of increasingly high affinity polyamides with improved selectivity. With the recent development of the (R)-3,4-Diaminobutyric Acid turn unit that has been shown to have increased affinities for DNA,¹⁰ labeling efficiencies closer to what are required for these challenging applications may soon be within reach.

A.4 Experimental Details

Abbreviations. Trishydroxymethylaminomethane (TRIS), ethylenediaminetetraacetic acid (EDTA).

Materials. Boc- β -Ala-PAM resin was purchased from Peptides International. Trifluoroacetic acid (TFA) was purchased from Halocarbon. Methylene Chloride (DCM) was obtained from Fisher Scientific and *N,N*-dimethylformamide (DMF) from EMD. EZ-Link TFP-PEO₃-Biotin was purchased from Pierce. Streptavidin was purchased from Rockland. DNA oligomers were purchased from IDT Technologies. Ultra Pure TRIS was purchased from ICN. Magnesium Acetate 4-hydrate was obtained from J.T. Baker. Water (18 M Ω) was purified using a aquaMAX-Ultra water purification system. Biological experiments were performed using Ultrapure Water (DNase/RNase free) purchased from USB. The pH of buffers was adjusted using a Thermo Orion 310 PerpHect Meter. All buffers were sterilized by filtration through a Nalgene 0.2 μ m cellulose nitrate filtration.

Polyamide Synthesis. Polyamide monomers were prepared as described previously.¹¹ Synthesis was performed using established protocols and all polyamides were characterized by MALDI-TOF and analytical HPLC. The synthesis of polyamides **1** and **2** is described in section 2.4 and in the literature.⁸

DNA Oligomers. The following DNA oligomers, described previously,⁶ were used for the assembly of the nucleation strand.

cpBr1: 5'-GTTGATGGAGTATAGTGTATTGGATGAAATGTTATGT-3'

A1S: 5'-TCACTGCTGAAGGCAGAGGACTGTGCTGGACTTGGTC-3'

A2: 5'-TGGTAATGTAAGGACCTCTGCCTTCAGC-3'

A4SV: 5'-CATACGACCAAGTGGATTTGTAGGAT-3'

A4_S00: 5'-TCACTGACCAAGTGGATTTGTAGGAT-3'

A3_nick: 5'-GGTTGAATGACCAGCACAGT-3'

Annealing Algorithmic Assembly Arrays. The long input scaffold strand was made using the published protocol.⁶ Individual tiles were annealed in TAEMg Buffer (40 mM Tris-HCl (pH 8.0), 20 mM acetic acid, 1 mM EDTA, 12.5 mM magnesium acetate) at a final concentration of 1 μ M, heated to 90°C and allowed to cool to RT over several hours. The array was created by mixing the input scaffold strand (~.004 μ M), Strands A1S (.2 μ M), A2 (.2 μ M), A3-nick (.2 μ M), A4-S00 (.2 μ M), cpBr1 (.2 μ M), A4SV (.002 μ M), and each of the tiles (.2 μ M) in TAEMg buffer at 90°C and allowing it to cool to RT over several hours.

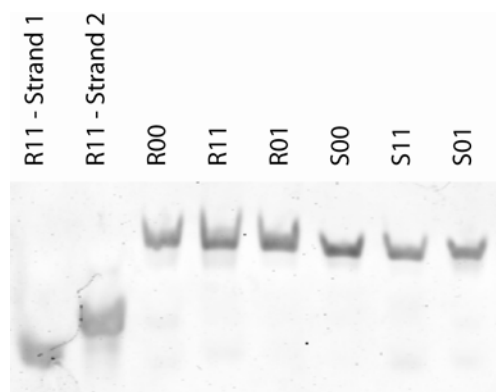


Figure A.5 Non-denaturing PAGE of the redesigned DX tiles for algorithmic assembly. As a control, the ssDNA component strands from tile R11 were run in the first two lanes. Clean formation of each tile is observed.

AFM Sample Preparation. The annealed origami was diluted 1:100 and 10 μL and incubated with polyamide (200 nM). 5 μL of sample was spotted on the mica, and 2 μL of 1 μM streptavidin was added for 1 min prior to imaging.

A.5 References

1. Aldaye, F. A.; Palmer, A. L.; Sleiman, H. F., Assembling materials with DNA as the guide *Science* **2008**, 321, 1795-1799.
2. Seeman, N. C., DNA in a material world *Nature* **2003**, 421, 427-431.
3. Lin, C. X.; Liu, Y.; Rinker, S.; Yan, H., DNA tile based self-assembly: Building complex nanoarchitectures *ChemPhysChem* **2006**, 7, 1641-1647.
4. Seeman, N. C.; Lukeman, P. S., Nucleic acid nanostructures: bottom-up control of geometry on the nanoscale *Rep. Prog. Phys.* **2005**, 68, 237-270.
5. Barish, R. D.; Rothmund, P. W. K.; Winfree, E., Two computational primitives for algorithmic self-assembly: Copying and counting *Nano Lett.* **2005**, 5, 2586-2592.
6. Rothmund, P. W. K.; Papadakis, N.; Winfree, E., Algorithmic self-assembly of DNA Sierpinski triangles *PLoS Biol.* **2004**, 2, 2041-2053.
7. Cohen, J. D.; Sadowski, J. P.; Dervan, P. B., Addressing single molecules on DNA nanostructures *Angew. Chem.-Int. Edit.* **2007**, 46, 7956-7959.

8. Cohen, J. D.; Sadowski, J. P.; Dervan, P. B., Programming multiple protein patterns on a single DNA nanostructure *J. Am. Chem. Soc.* **2008**, 130, 402-403.
9. Fujibayashi, K.; Hariadi, R.; Park, S. H.; Winfree, E.; Murata, S., Toward reliable algorithmic self-assembly of DNA tiles: A fixed-width cellular automaton pattern *Nano Lett.* **2008**, 8, 1791-1797.
10. Dose, C.; Farkas, M. E.; Chenoweth, D. M.; Dervan, P. B., Next generation hairpin polyamides with (R)-3,4-diaminobutyric acid turn unit *J. Am. Chem. Soc.* **2008**, 130, 6859-6866.
11. Baird, E. E.; Dervan, P. B., Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids *J. Am. Chem. Soc.* **1996**, 118, 6141-6146.