8-13-2019

DNA-binding Small Molecules as Drug Agents that Interfere with Transcription Factors: the Development, the Potential and the Future

Beibei Liu

DNA-BINDING SMALL MOLECULES AS DRUG AGENTS THAT INTERFERE WITH TRANSCRIPTION FACTORS: THE DEVELOPMENT, THE POTENTIAL AND THE FUTURE

by

BEIBEI LIU

Under the Direction of W. David Wilson, PhD

ABSTRACT

DNA-minor groove binding small molecules have been extensively developed to achieve higher binding affinity and specificity. Polyamides are a class of small molecules that can be programmed to target any predetermined DNA sequence. The development of hairpin polyamides along with introduction of β-alanine substituents, has greatly enhanced the DNA binding properties of these molecules. Yet the correlation between β-insert and binding properties remains unclear. On the other hand, the design of small-size, fluorescent hybrid polyamides has facilitated cell studies due to their ease of observation. There is a strong need to expand the DNA recognition

sites of such molecules and extend their biological applications. This dissertation has explored the systematic design and synthesis of eight-ring hairpin polyamides as well as the modified Pyr-AzaHx hybrid polyamides. Comprehensive biophysical and biochemical tools were employed to evaluate their binding properties. The effects of β-alanine and N-terminal cationic groups on hairpin polyamides-DNA binding have been discussed. The binding properties of modified Pyr-AzaHx polyamides were explored. Altogether, the work provided fundamental guidance for the prediction of binding properties of similar molecules as well as strategies for the design of more competitive molecules.

Transcription factors bind to specific DNA sequences in the major groove and regulate gene expression. Abnormal expression of transcription factors is involved in the development of many serious diseases. Precise control of gene expression by targeting transcription factors can be an alternative therapeutic approach. Polyamides bind to DNA with affinities comparable to proteins, empowering them with the ability to interfere with transcription factors at specific DNA binding site and consequently altering the gene expression level. In this dissertation, the effect of polyamides on the binding of transcription factor PU.1 was studied. Abnormal expression of PU.1 is involved in the development of acute myeloid leukemia (AML). A positive correlation was established between eight-ring polyamide binding affinity and inhibition efficacy for PU.1. A non-inhibitor polyamide FH1024 was identified and the mechanism of action among polyamide, DNA and PU.1 was explored. The studies showed strong evidence of the capability of polyamides serving as drug agents. This work also established solid basis for the further cell studies.

INDEX WORDS: DNA minor groove, Small molecules, Polyamide, ETS transcription factors, PU.1, AzaHx, SPR

DNA-BINDING SMALL MOLECULES AS DRUG AGENTS THAT INTERFERE WITH TRANSCRIPTION FACTORS: THE DEVELOPMENT, THE POTENTIAL AND THE FUTURE

by

BEIBEI LIU

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2019

DNA-BINDING SMALL MOLECULES AS DRUG AGENTS THAT INTERFERE WITH TRANSCRIPTION FACTORS: THE DEVELOPMENT, THE POTENTIAL AND THE FUTURE

by

BEIBEI LIU

Committee Chair: W. David Wilson

Committee: Gregory M. K. Poon

Jenny J. Yang

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

August 2019

DEDICATION

This dissertation is dedicated to my parents, Xingguo Liu and Jingxiang Tian.

You raise me up to more than I can be.

ACKNOWLEDGEMENTS

Lingering in my mind is the saying “A teacher for one-day, a father for a lifetime” which dates back to the Warring States Period.

I would like to extend my deepest thanks to my advisor, Dr. W. David Wilson, an erudite scientist, whose keen interest in the advancement of science have always shed light on new perspectives; who paved the road for me along this journey of exploration; who selflessly provided me with invaluable advice on my academic, professional and personal development; who constantly supported and encouraged me to perform to my full potential.

My great thanks also go to my committee member and collaborator, Dr. Gregory M. K. Poon, for generously providing us with high quality proteins and for his insightful suggestion on my project. I am very grateful for my committee member, Dr. Jenny J. Yang, who is a real-life role model, and never spares a word in encouraging and inspiring young professionals.

My collaborators, Dr. James K. Bashkin and Dr. Moses Lee deserve my great appreciation, without whose efforts, my research would not stand a solid base. Dr. Lee’s optimism and positive energy keeps me cheered in a long way.

I can never thank my parents, Xingguo Liu and Jingxiang Tian enough. Never would I make it this far without their unconditional love and support, financially, mentally and spiritually. My father used to let me sit on his shoulder, telling me to look further and dream higher. It was my parents who taught me to be confident and respect others, who are never tired of my complaints. It was with their patience and encouragement that I was able to endure through difficult times and grow stronger. It was because of my parents’ hard work that I was granted the opportunity to further my academic career and the freedom to make choices.

I cannot ask for better ones.

Dr. Jacob Ball joined me at the beginning of this journey and became an inseparable part of my life. He lit the hope of my life. Knowledgeable and dedicated, he has been constantly inspiring me in my pursuit of academic goals with his sharp scientific insights and great enthusiasm. Discussions with Jacob has always sparked good ideas. It was with his company, care and support that I was able to go through some of the most difficult times. Whenever and wherever I am running, he is always the one who runs with me.

I owe a special debt to my grandparents, who have helped enormously in raising me up and taught me to be humble and graceful. I am fortunate to have my sister's and brother's company growing up. They are my every-day sunshine who always cheer me up. My deep appreciation goes to my friends Hollis Ball and Cookie Ball, who welcomed me to their family with great warmth and supported me along the way.

Special thanks to Mrs. Carol Wilson, a strong and wonderful woman who is dedicated to serving the lab. Carol has always helped proofread my manuscripts with great patience. She makes the best brownies that one can ask for.

It has been a long journey and I could not have made it to the other end without the help and support of many people. Thanks to Dr. Dabney Dixon and Dr. Donald Hamelberg for their constructive lectures and recommendations. My heartfelt thanks to my labmates for all the unforgettable memories.

I would also like to thank the Department of Chemistry and Georgia State University for offering me the chance to fulfill my pursuit. It is with this opportunity that I was able to see a bigger world and consequently to realize what I truly want in my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

PA	polyamide
Py	N-methylpyrrole
Im	N-methylimidazole
β	β-alanine
Hp	N-methyl-3-hydroxypyrrole
γ-turn	γ-aminobutyric acid
Dp	3-(dimethylamino)propylamine
Ta	3,3'-diamino-N-methyldipropylamine
TMG	tetramethylguanidine
P	pyrrole
I	imidazole
f	formamide
Hx	2-(*p*-anisyl)-benzimidazole
AzaHx	2-*p*-(anisyl)-4-aza-benzimidazole
Pyr-AzaHx	pyridyl-aza-benzimidazole
TF	transcription factor
ETS	E26 transformation specific
PU.1	purine rich box 1
AML	acute myeloid leukemia
SPR	surface plasmon resonance
Tm	thermal melting
CD	circular dichroism

EMSA	electrophoresis mobility shift assay
ESI-MS	electrospray ionization mass spectrometry
NMR	nuclear magnetic resonance

1 INTRODUCTION

DNA carries genetic information in living organisms and usually assumes a B-form, right-handed double helical structure (1). The genetic information is delivered through a series of cellular events like DNA replication, transcription and translation. Among those, transcription is the process of converting DNA into messenger RNA, during which a range of proteins, including transcription factors, are recruited to facilitate and regulate the process. The precise operation of transcription is essential for maintaining cellular homeostasis (2). In fact, abnormal regulation of gene transcription is found to be involved in a number of devastating health issues such as cancer and infectious diseases (3). Transcription factors play an important role in gene regulation and their activities are often altered in cancer cells due to mutations, abnormal signaling, and posttranslational modifications (4). Therefore, transcription factors are an attractive therapeutic target point in drug development. Useful molecular tools that enables the accurate control of gene expression by directly or indirectly targeting transcription factors can be a potential approach to treating those serious conditions. Binding of transcription factors to specific DNA sequences is a crucial step to recruiting accessory proteins and the subsequent transcriptional activities (2). Specific DNA-binding small molecules can be used as anti-cancer agents to interfere with transcription factor-DNA interactions and thus rectify the abnormal gene expression.

1.1 DNA minor groove-targeting small molecules

1.1.1 Development of β-alanine inserted hairpin polyamides

N-methylpyrrole (Py), N-methylimidazole (Im) polyamides (PAs) are a class of programmable synthetic small molecules that interact with DNA by binding to the minor groove of specific DNA sites. Polyamides are originally derived from the naturally existing antibiotics: netropsin and distamycin A (5, 6). These natural molecules are composed of two or three pyrrole

heterocycles linked with amide binds (Figure 1.1) and bind the AT rich regions of DNA preferentially in a 1:1 ratio (7, 8). These molecules interact with DNA through hydrogen bonding, van der Waals forces and electrostatic interactions to achieve stability and specificity of the complex. Based on the X-ray analysis of netropsin-DNA complex, Dickerson and coworkers established the molecular foundation of the base specificity of netropsin. That is, the binding preference of netropsin to four or more consecutive AT base pairs is achieved not by hydrogen bonding but by the close van der Waals contacts between C3-H of pyrrole and C2-H of adenine. They further proposed that replacement of pyrrole with imidazole, that is changing C3-H to N, would enable the modified new molecule to recognize a G/C base pair, and therefore, any short DNA sequences (9). This work is fundamental for the development of polyamides that are capable of expanding the DNA recognition site. Wemmer and coworkers later found in an NMR experiment that titration of DNA with distamycin A at high drug to DNA ratio (>1:1) generated a single, stable drug-DNA complex that was different from the complexes formed at low drug to DNA ratio (< 1:1). Their NMR structure revealed that distamycin A formed a 2:1 antiparallel stacked dimer with DNA (10). These discoveries have laid solid foundation for the evolution of polyamides. Subsequently, Lown, Devan, Wemmer and coworkers were able to synthesize a netropsin analogue that incorporated imidazole at the N-terminus and demonstrated that the new molecule was able to recognize mixed A/T and G/C sequences as a 2:1 antiparallel side-by-side dimer (11-13). The result not only proved the previous hypothesis but also significantly facilitated the formation of the widely-recognized DNA binding rules. Briefly, Py/Py recognizes A/T or T/A base pairs; Py/Im recognizes C/G base pairs; Im/Py recognizes G/C base pairs. Since Py/Py cannot distinguish A/T from T/A, Dervan and coworkers introduced an N-methyl-3-hydroxypyrrole (Hp) moiety by replacing C3-H with C3-OH on the pyrrole unit of polyamides. Although Hp was able

to bind to T through hydrogen bonding and enable the molecule to distinguish T/A from A/T, it is not widely used due to the difficulty in synthesis and its instability in both solid and solution form (14-19). Based on the DNA recognition rules, we are able to design polyamides that can target any short DNA sequences.

Inspired by the dimer mode of binding as well as to enhance the binding affinities, polyamides have been developed into various forms, for example, hairpin, H-pin, U-pin and cyclic polyamides (20-23). Among these structures, hairpin polyamides are most commonly used in research because of their versatility and ease of synthesis (23). Hairpin polyamides consist of two polyamide strands that are covalently linked by γ-aminobutyric acid (γ-turn) from the C-terminus of one polyamide to the N-terminus of the other (24, 25). (Figure 1.2) Binding of hairpin polyamides can assume two orientations: N-terminus of polyamide aligns with 5' side of DNA (forward binding) or 3' side of DNA (reverse binding) (24, 26-28). Hairpin polyamides bind to DNA following the recognition rules with significantly enhanced affinity (almost 100-fold higher compared to the naturally existing distamycin A) and specificity. This modification has empowered polyamides to achieve DNA-binding affinities comparable to those of DNA-binding proteins (28).

In order to expand the length of the DNA recognition site and therefore, to achieve higher specificity, hairpin polyamides have to be elongated accordingly by increasing the number of Py and Im heterocycles. Studies have shown that the limit of the number of rings on either polyamide strand is five for the DNA to tolerate (23). Hairpin polyamides longer than that become over curved (compared to the curvature of the B-form double strand DNA helix) and too stringent to accommodate themselves to the DNA minor groove. As a result, both binding affinity and specificity are sacrificed. This issue was tackled by introducing β-alanine (β) into the hairpin

polyamide building block as a flexible motif to replace one or more Py rings. Incorporation of β-alanine relaxes the overall polyamide structure and can restore both the binding affinity and specificity (29, 30). β-alanine functions as a Py and forms hydrogen bonds with A and T.

The recognition activity of small molecules is primarily determined by their structures. Until now, the molecular basis of the interactions between β-containing hairpin polyamides and DNA is still unclear. The number and position of β in polyamides largely affect their binding affinity, specificity as well as orientation. Bashkin and coworkers found that not all β-inserted polyamides enhance binding affinity (31-33). Hence, a thorough investigation of the effects of the composition of hairpin polyamides on their DNA binding activities is in great need.

1.1.2 Development of short Pyr-AzaHx hybrid polyamides

Polyamides are developed to realize their optimal biological functions in living organisms, which requires efficient cell uptake and nuclear localization. The cell permeability of prototypical polyamides varies depending on the cell types, polyamide structure, composition, and size (34-37). Typically, hairpin polyamides are considered to be associated with poor nuclear localization due to their larger size, lower solubility and stability (38). On the other hand, cell study of the biological effects of polyamides requires conjugation of fluorescent dyes due to its lack of intrinsic fluorescence. Although significant progress has been made using this method, molecules that could enable the direct observation of nuclear localization would eliminate extra concerns about the effect of fluorescent dyes on the overall structure and function of polyamides (39-44).

Lee, Wilson and coworkers sought an alternative approach based on the molecular recognition of triamides (45) (Figure 1.3), by attaching an intrinsically fluorescent Hx moiety (2-(p-anisyl)-benzimidazole) to a consecutive I-P, P-P or P-I heterocycles (Figure 1.3) The resulted hybrid Hx-amides bind to DNA as a stacked, side-by-side, antiparallel dimer. Hx expands along

two continuous dinucleotides and behaves predictably similarly to P-P or f-P (46, 47). The novel small-sized hybrid polyamides retained the binding affinity and specificity and also exerted better solubility and cell uptake activities. More importantly, these polyamides greatly facilitated the direct observation of nuclear localization with their fluorescent properties. The Lee, Wilson group further built upon the solid Hx base, an AzaHx moiety (2-(p-anisyl)-4-azabenzimidazole), to extend the DNA recognition repertoire (48) (Figure 1.3). AzaHx differentiates itself from Hx in that it mimics P-I or f-I and can target AG, TG or CG dinucleotides. Of the most importance is that this modification extended the recognition of central bases from CG to GC as well.

1.2 DNA major groove-targeting transcription factors

Transcription factors are of vital importance in the first level of regulation of transcription. The process requires the binding of transcription factors to defined promoter or enhancer regions to recruit subsequent transcription machinery (2, 49). ETS (E26 transformation specific) family of transcription factors are a group of functionally diverse proteins that share a conserved 85-residue domain, which specifically recognizes purine rich sequences containing a 5'-GGA(A/T)-3' consensus core (50). Of those proteins, PU.1 (Purine rich box 1) is expressed selectively on B cells, myeloid cells and macrophages and regulates several genes that are involved in cell differentiation and the immune system (51). Over- or under- regulation of PU.1 expression is associated with the induction of acute myeloid leukemia (AML) (52-55). Regulating PU.1 expression level in a controllable way, therefore, has been considered a feasible approach for the treatment of AML.

The λB motif of the Igλ2-4 enhancer (λB promoter site: 5'-CCAAATAAAA**GGAA**GTGAAACCAAG-3') has been widely used in research as the classical PU.1 binding cognate site. Stability of PU.1-DNA complex is a crucial determining factor of the

transcriptional efficacy *in vivo*. Studies have shown that the promoter activity is strongly correlated with the binding affinity of PU.1 (56). In a sequence selectivity study, Poon demonstrated that the binding affinity of PU.1-DNA is affected by flanking sequences up to three bases upstream and/or two bases downstream of the core consensus, using λB site as reference. While most of the mutated sequences lowered the binding affinity, the 3' side mutants generated the poorest binding of all (50).

Small molecules have been used to target the flanking sequences of the promoter region and to allosterically compete or cooperate with transcription factors binding to DNA. In the case of PU.1, most studies have focused on the AT rich region at the 5' side of the 5'-GGAA-3' sequence. Applications of natural agent distamycin as well as a class of small heterocyclic diamidines have been shown to successfully inhibit PU.1 from binding to the cognate site (57, 58). To this date, there has hardly been any report on the effect of the 3' side binding molecules, provided that the 3' side is of equal or even more importance to the binding of PU.1.

1.3 Objectives

This dissertation took the effort to systematically design and synthesize a set of eight-ring hairpin polyamides with modifications not only in the arrangement of β-alanine but also the N-terminus charged motif. By evaluating the DNA binding properties of these molecules, this work established a general binding pattern for eight-ring hairpin polyamides, which expanded our understanding of the molecular recognition of hairpin polyamides as well as provided clear guidance for the optimization of rational polyamide design in the future.

With the aim of expanding the DNA recognition repertoire of fluorescent small hybrid polyamides, this dissertation attempted to recognize 5'-GGCC-3' DNA sequence by designing a set of pyridine-AzaHx polyamides. Assessment of their binding properties revealed that the Pyr-

AzaHx polyamides do not recognize the desired GG dinucleotides, but behave similarly to AzaHx polyamide with significantly enhanced binding affinity.

Over the years of development, modified polyamides are able to achieve DNA-binding affinities that rival those of DNA-binding proteins (34, 35). Binding of polyamides has also been shown to cause DNA conformational change in a way that the minor groove is widened while the major groove is compressed (22, 59). This dissertation took the pioneering effort in investigating the effect of a set of eight-ring hairpin polyamides that target the 3' side immediate flanking sequence of the PU.1 binding site, on the binding of PU.1. The work established a positive linear correlation between the binding affinity and inhibition efficacy of polyamides. More importantly, it discovered and characterized a close analogue of the polyamide that potentially enhances PU.1 binding to DNA. This property of polyamides has never been characterized before.

In summary, this dissertation employees a wide selection of biochemical and biophysical techniques including SPR, ESI-MS, thermal melting, EMSA, CD, DNase I footprinting and so on, for the exploration of the binding properties of two different types of polyamide derivatives. Both the established binding pattern of the classical hairpin polyamides and the enhanced binding of the small-sized Pyr-AzaHx hybrid polyamides are valuable references for the development of design strategies of molecules with improved DNA binding properties and biological performance. The investigation of PU.1-DNA, polyamide-DNA interactions uncovers possibly new molecular mechanisms of the mode of action of these interactions, which can significantly expedite the design of biologically relevant agents when understood thoroughly.

1.4 References

1 J. D. Watson, and F.H. Crick. Genetical implications of the structure of deoxyribonucleic acid. Nature., 1953, 171, 964-967.

2 C. Yan, and P.J. Higgins. Drugging the undruggable: transcription therapy for cancer. *Biochim Biophys Acta*., **2013**, *1835*, 76-85.

3 L. H. Hurley, DNA and its associated processes as targets for cancer therapy. *Nat Rev Cancer*., **2002,** *2*, 188-200.

4 A. G. Papavassiliou, Transcription-factor-modulating agents: precision and selectivity in drug design. *Mol Med Today*., **1998,** *4*, 358-366.

5 A. C.Finlay, F. A. Hochstein, B. A. Sobin, and F. X. Murphy. Netropsin, a new antibiotic produced by a Streptomyces. *J. Am. Chem. Soc*., **1951,** *73*, 341-343.

6 F. Arcamone, S. Penco, P. Orezzi, V. Nicolella, and A. Pirelli. Structure and Synthesis of Distamycin A. *Nature.*, **1964,** *203*, 1064-1065.

7 A. Hahn, Molecur mechanism of action of the radioprotective substance WR 2721. *Radiat Environ Biophys.*, **1975,** *11*, 265-269.

8 C. Zimmer, and U. Wahnert. Nonintercalating DNA-binding ligands: specificity of the interaction and their use as tools in biophysical, biochemical and biological investigations of the genetic material. *Prog Biophys Mol Biol.*, **1986,** *47*, 31-112.

9 M. L. Kopka, C. Yoon, D. Goodsell, P. Pjura, and R. E. Dickerson. The molecular origin of DNA-drug specificity in netropsin and distamycin. *Proc Natl Acad Sci U S A.*, **1985,** *82*, 1376-1380.

10 J. G. Pelton, and D .E. Wemmer, 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.

11 W. S. Wade, M. Mrksich, and P. B. Dervan. Design of Peptides That Bind in the Minor Groove of DNA at 5'-(a,T)G(a,T)C(a,T)-3' Sequences by a Dimeric Side-by-Side Motif. *J Am Chem Soc*., **1992,** *114,* 8783-8794.

12 M. Mrksich, W. S. Wade, T. J. Dwyer, B. H. Geierstanger, D. E. Wemmer and P. B. Dervan. Antiparallel side-by-side dimeric motif for sequence-specific recognition in the minor groove of DNA by the designed peptide 1-methylimidazole-2-carboxamide netropsin. *Proc Natl Acad Sci U S A.,* **1992,** *89,* 7586-7590.

13 W. S. Wade, M. Mrksich, and P. B. Dervan. Binding affinities of synthetic peptides, pyridine-2-carboxamidonetropsin and 1-methylimidazole-2-carboxamidonetropsin, that form 2:1 complexes in the minor groove of double-helical DNA. *Biochemistry.,* **1993,** *32,* 11385-11389.

14 S. White, J. W. Szewczyk, J. M. Turner, E. E. Barid, and P. B. Dervan. Recognition of the four Watson-Crick base pairs in the DNA minor groove by synthetic ligands. *Nature*., **1998,** *391,* 468-471.

15 C. L., Kielkopf, S. White, J. W. Szewczyk, J. M. Turner, E. E. Barid, P. B. Dervan, and D. C. Reese. A structural basis for recognition of A.T and T.A base pairs in the minor groove of B-DNA. Science, **1998,** *282,* 111-115.

16 P. B. Dervan. Molecular recognition of DNA by small molecules. *Bioorg Med Chem*., **2001,** *9,* 2215-2235.

17 P. B. Dervan, and B. S. Edelson. Recognition of the DNA minor groove by pyrrole-imidazole polyamides. *Curr Opin Struct Biol.,* **2003,** *13,* 284-299.

18 C. Melander, R. Burnett, and J. M. Gottesfeld. Regulation of gene expression with pyrrole-imidazole polyamides. *J Biotechnol.,* **2004,** *112,* 195-220.

19 M. S. Blackledge, and C. Melander. Programmable DNA-binding small molecules. *Bioorg Med Chem.,* **2013,** *21,* 6101-6114.

20 M. Mrksich, and P. B. Dervan. Design of a Covalent Peptide Heterodimer for Sequence-Specific Recognition in the Minor-Groove of Double-Helical DNA. *J Am Chem Soc*., **1994,** *116,* 3663-3664.

21 A. Heckel, and P. B. Dervan. U-pin polyamide motif for recognition of the DNA minor groove. *Chemistry.,* **2003,** *9,* 3353-3366.

22 D. M. Chenoweth, and P. B. Dervan. Allosteric modulation of DNA by small molecules. *Proc Natl Acad Sci U S A.,* **2009,** *106,* 13175-13179.

23 Y. Kawamoto, T. Bando, and H. Sugiyama. Sequence-specific DNA binding Pyrrole-imidazole polyamides and their applications. *Bioorg Med Chem.,* **2018,** *26,*1393-1411.

24 M. Mrksich, M. E. Parks, and P. B. Dervan. 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.

25 C. L. Wu, W. Wang, L, Fang, and W. Su. Programmable pyrrole-imidazole polyamides: A potent tool for DNA targeting. *Chinese Chemical Letters*., **2018,** *29,*1105-1112.

26 E. E. Baird, and P.B. Dervan. Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids. *J Am Chem Soc*., **1996,** *118,* 6141-6146.

27 N. R. Wurtz, Fmoc solid phase synthesis of polyamides containing pyrrole and imidazole amino acids. *Org Lett.,* **2001,** *3,*1201-1203.

28 J. W. Trauger, E. E. Baird, and P. B. Dervan. Recognition of DNA by designed ligands at subnanomolar concentrations. *Nature*., **1996,** *382,* 559-561.

29 J. M. Turner, E. E. Baird, and P. B. Dervan. Recognition of seven base pair sequences in the minor groove of DNA by ten-ring pyrrole-imidazole polyamide hairpins. *J Am Chem Soc*., **1997,** *119,* 7636-7644.

30 J. M. Turner, S. E. Swalley, E. E. Baird, and P. B. Dervan. Aliphatic/aromatic amino acid pairings for polyamide recognition in the minor groove of DNA. *J Am Chem Soc*., **1998,** *120,* 6219-6226.

31 J. K. Bashkin, K. Aston, J. P. Ramos, K. J. Koeller, R. Nanjunda, G. He, C. M. Dupureur, and W. D. Wilson. Promoter scanning of the human COX-2 gene with 8-ring polyamides: unexpected weakening of polyamide-DNA binding and selectivity by replacing an internal N-Me-pyrrole with beta-alanine. *Biochimie*., **2013,** *95*,271-279.

32 S. Wang, K. Aston, K. J. Koeller, G. D. Harris, Jr. N. P. Rath, J. K. Bashkin, and W. D. Wilson. Modulation of DNA-polyamide interaction by beta-alanine substitutions: a study of positional effects on binding affinity, kinetics and thermodynamics. *Org Biomol Chem*., **2014,** *12*,7523-7536.

33 S. Wang, R. Nanjunda, K. Aston, J. K. Bashkin, and W. D. Wilson. Correlation of local effects of DNA sequence and position of beta-alanine inserts with polyamide-DNA complex binding affinities and kinetics. *Biochemistry.,* **2012,** *51*,9796-9806.

34 T. P. Best, B. S. Edelson, N. G. Nickols, and P. B. Dervan. Nuclear localization of pyrrole-imidazole polyamide-fluorescein conjugates in cell culture. *Proc Natl Acad Sci U S A.*, **2003,** *100,* 12063-12068.

35 B. S. Edelson, T. P. Best, B. Olenyuk, N. G. Nickols, R. M. Doss, S. Foister, A. Heckel, and P. B. Dervan. Influence of structural variation on nuclear localization of DNA-binding polyamide-fluorophore conjugates. *Nucleic Acids Res.*, **2004,** *32,* 2802-2818.

36 J. M. Belitsky, S. J. Leslie, P. S. Arora, T. A. Beerman, and P. B. Dervan. Cellular uptake of N-methylpyrrole/N-methylimidazole polyamide-dye conjugates. *Bioorg Med Chem.,* **2002,** *10,* 3313-3318.

37 S. Nishijima, K. Shinohara, T. Bando, M. Minoshima, G. Kashiwazaki, and H. Sugiyama. Cell permeability of Py-Im-polyamide-fluorescein conjugates: Influence of molecular size and Py/Im content. *Bioorg Med Chem*., **2010,** *18,* 978-983.

38 A. Franks, C. Tronrud, K. Kiakos, J. Kluza, M. Munde, T. Brown, H. Mackay, W. D. Wilson, D. Hochhauser, J. A. Hartley, M. Lee. Targeting the ICB2 site of the topoisomerase IIalpha promoter with a formamido-pyrrole-imidazole-pyrrole H-pin polyamide. *Bioorg Med Chem*., **2010,** *18,* 5553-5561.

39 J. A. Raskatov, J. O. Szablowski, and P. B. Dervan, Tumor xenograft uptake of a pyrrole-imidazole (Py-Im) polyamide varies as a function of cell line grafted. *J Med Chem*., **2014,** *57,* 8471-8476.

40 T. F. Martinez, J. W. Philips, K. K. Karanja, P. Polaczek, C. M. Wang, B. C. Li, J. L. Campbell, and P. B. Dervan. Replication stress by Py-Im polyamides induces a non-canonical ATR-dependent checkpoint response. *Nucleic Acids Res*., **2014,** *42,* 11546-11559.

41 J. A. Raskatov, J. W. Puckett, and P. B. Dervan. A C-14 labeled Py-Im polyamide localizes to a subcutaneous prostate cancer tumor. *Bioorg Med Chem*., **2014,** *22,* 4371-4375.

42 J. S. Kang, J. L. Meier, and P. B. Dervan. Design of sequence-specific DNA binding molecules for DNA methyltransferase inhibition. *J Am Chem Soc*., **2014,** *136,* 3687-3694.

43 J. L. Meier, D. C. Montgomery, and P. B. Dervan. Enhancing the cellular uptake of Py-Im polyamides through next-generation aryl turns. *Nucleic Acids Res*., **2012,** *40,* 2345-2356.

44 T. Vaijayanthi, T. Bando, G. N. Pandian, and H. Sugiyama. Progress and prospects of pyrrole-imidazole polyamide-fluorophore conjugates as sequence-selective DNA probes. *Chembiochem*., **2012,** *13,* 2170-2185.

45 N. M. Le, A. M. Sielaff, A. J. Cooper, H. Mackay, T. Brown, M. Kotecha, C. O'Haire, D. Hochhauser, M. Lee and J. A. Hartley. Binding of f-PIP, a pyrrole- and imidazole-containing triamide, to the inverted CCAAT box-2 of the topoisomerase IIalpha promoter and modulation of gene expression in cells. *Bioorg Med Chem Lett*., **2006,** *16,* 6161-6164.

46 S. Chavda, Y. Liu, B. Babu, R. Davis, A. Sieiaff, J. Ruprich, L. Westrate, C. Tronrud, A. Ferguson, A. Franks, S. Tzou, C. Adkins, T. Rice, H. Mackay, J. Kluza, S. A. Tahir, S. Lin, K. Kiakos, C. D. Bruce, W. D. Wilson, J. A. Hartley and M. Lee. Hx, a novel fluorescent, minor groove and sequence specific recognition element: design, synthesis, and DNA binding properties of p-anisylbenzimidazole-imidazole/pyrrole-containing polyamides. *Biochemistry*., **2011,** *50,* 3127-3136.

47 V. Satam, P. Patil, B. Babu, M. Gregory, M. Bowerman, M. Savagian, M. Lee, S. Tzou, K. Olson, Y. Liu, J. Ramos, W. D. Wilson, J. P. Bingham, K. Kiakos, J. A. Hartley, and M. Lee. Hx-amides: DNA sequence recognition by the fluorescent Hx (p-anisylbenzimidazole)*pyrrole and Hx*imidazole pairings. *Bioorg Med Chem Lett*., **2013,** *23,* 1699-1702.

48 V. Satam, B. Babu, P. Patil, K. A. Brien, K. Olson, M. Savagian, M. Lee, M. Mepham, L. B. Jobe, J. P. Bingham, L. Pett, S. Wang, M. Ferrara, C. D. Bruce, W. D. Wilson, M. Lee, J. A. Hartley, and K. Kiakos. AzaHx, a novel fluorescent, DNA minor groove and G.C recognition element: Synthesis and DNA binding properties of a p-anisyl-4-aza-benzimidazole-pyrrole-imidazole (azaHx-PI) polyamide. *Bioorg Med Chem Lett*., **2015,** *25,* 3681-3685.

49 A. Melnick. Predicting the effect of transcription therapy in hematologic malignancies. *Leukemia.,* **2005,** *19,* 1109-1117.

50 G. M. Poon, and R. B. Macgregor, Jr. Base coupling in sequence-specific site recognition by the ETS domain of murine PU.1. *J Mol Biol.,* **2003,** *328,* 805-819.

51 J. Lloberas, C. Soler, and A. Celada. The key role of PU.1/SPI-1 in B cells, myeloid cells and macrophages. *Immunol Today.,* **1999,** *20,* 184-189.

52 E. W. Scott, R. C. Fisher, M. C. Olson, E. W. Kehrli, M. C. Simon, and H. Singh. PU.1 functions in a cell-autonomous manner to control the differentiation of multipotential lymphoid-myeloid progenitors. *Immunity.,* **1997,** *6,* 437-447.

53 F. Rosenbauer, K. Wagner, J. L. Kutok, H. Iwasaki, M. M. Le Beau, Y. Okuno, K. Akashi, S. Fiering, and D. G. Tenen. Acute myeloid leukemia induced by graded reduction of a lineage-specific transcription factor, PU.1. *Nat Genet.,* **2004,** *36,* 624-630.

54 E. W. Scott, M. C. Simon, J. Antastasi, and H. Singh. Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. *Science.,* **1994,** *265,* 1573-1577.

55 D. L. Stirewalt, Fine-tuning PU.1. *Nat Genet.,* **2004,** *36,* 550-551.

56 S. L. Li, W. Schlegel, A. J. Valente, and R. A. Clark. Critical flanking sequences of PU.1 binding sites in myeloid-specific promoters. *J Biol Chem.,* **1999,** *274,* 32453-32460.

57 M. Munde, G. M. Poon, and W .D. Wilson. Probing the electrostatics and pharmacological modulation of sequence-specific binding by the DNA-binding domain of the ETS family transcription factor PU.1: a binding affinity and kinetics investigation. *J Mol Biol.,* **2013,** *425,* 1655-1669.

58 M. Munde, S. Wang, A. Kumar, C. E. Stephens, A. A. Farahat, D. W. Boykin, W. D. Wilson, and G. M. Poon. Structure-dependent inhibition of the ETS-family transcription factor PU.1 by novel heterocyclic diamidines. . *Nucleic Acids Res.*, **2014,** *42,* 1379-1390.

59 Y. Wang, N. Ma, and G. Chen. Allosteric analysis of glucocorticoid receptor-DNA interface induced by cyclic Py-Im polyamide: a molecular dynamics simulation study. *PloS one.*, **2012,** *7,* e35159.

1.5 Tables and Figures

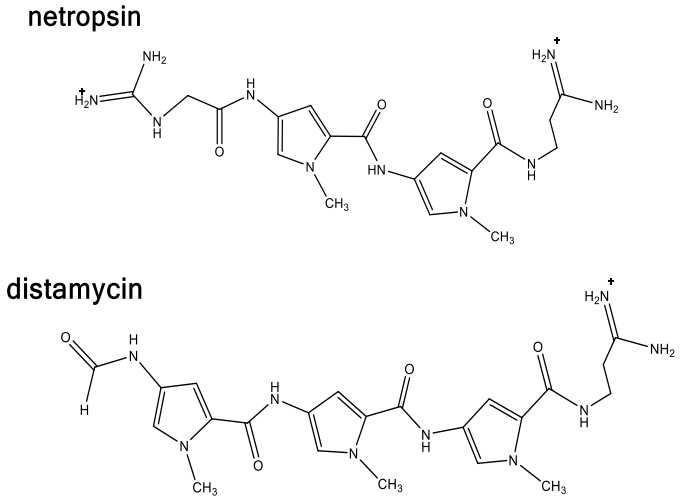


Figure 1.1 Structures of netropsin and distamycin.

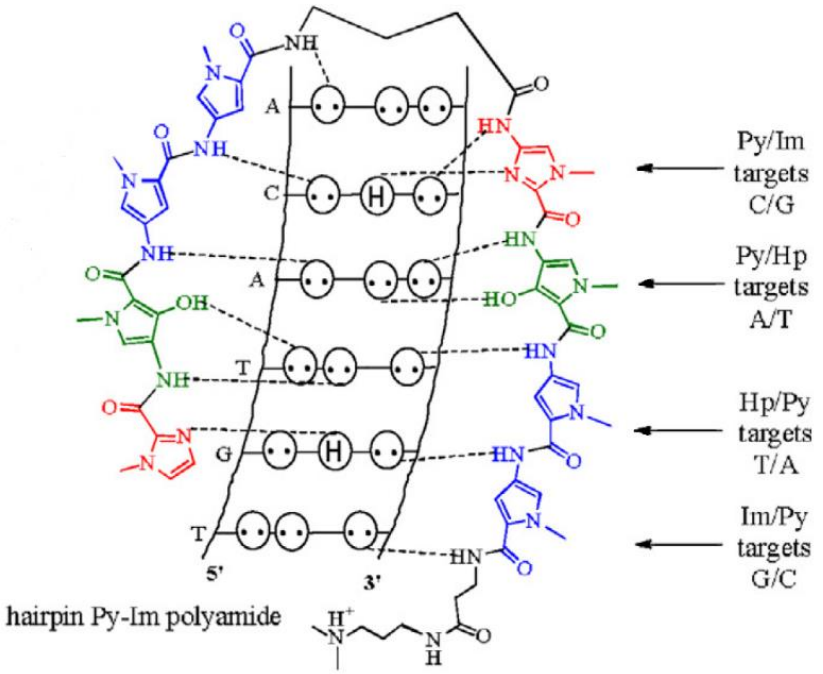


Figure 1.2 Structure of hairpin polyamide and its recognition to DNA (25).

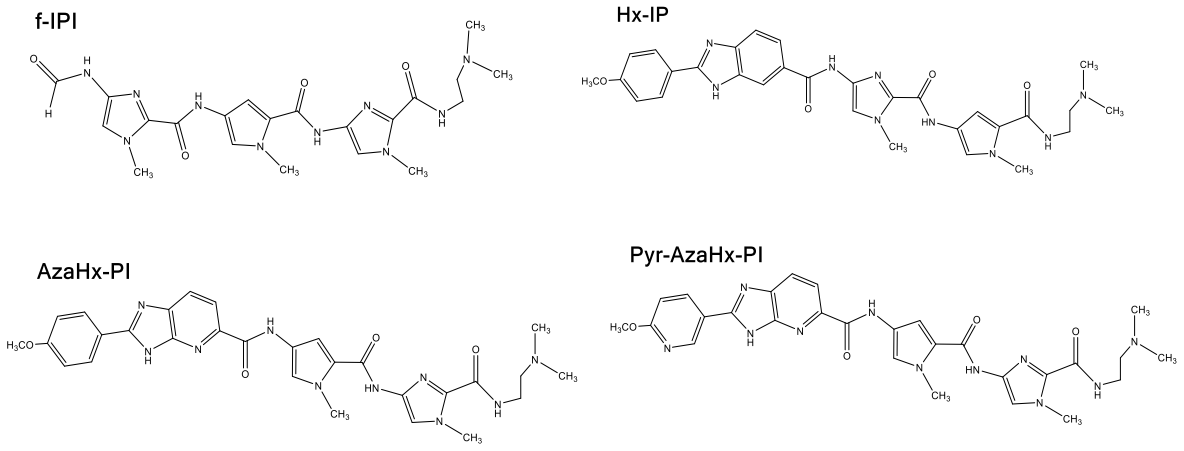


Figure 1.3 Structures of triamide (f-IPI) and hybrid polyamides.

2 B-ALANINE AND N-TERMINAL CATIONIC SUBSTITUENTS AFFECT POLYAMIDE-DNA BINDING

Beibei Liu,[a] Shuo Wang, [a] Karl Aston, [b] Kevin J. Koeller, [b] Shahrzad Fanny Hakami Kermani, [b] Carlos H. Castañeda, [b] M. Jose' Scuderi, [b] Rensheng Luo, [b] James K. Bashkin, *[b] W. David Wilson *[a]

[a] Department of Chemistry, Georgia State University, Atlanta, GA 30303, USA

[b] Department of Chemistry & Biochemistry, Center for Nanoscience, University of Missouri-St. Louis, St. Louis, MO 63121, USA

*To whom correspondence should be addressed:

*[a] E-mail: wdw@gsu.edu; Fax: +1 404-413-5505; Tel: +1 404-413-5503

*[b] E-mail: bashkinj@umsl.edu; Fax: +1 314-516-5342; Tel: +1 314-516-7352

Liu B, *et al*. *Org. Biomol. Chem.,* 2017, **15,** 9880-9888.

My contribution to this paper was designing and conducting the experiments, analyzing the results and writing the manuscript.

2.1 Abstract

Minor-groove binding hairpin polyamides (PAs) bind specific DNA sequences. Synthetic modifications can improve PA-DNA binding affinity and include flexible modules, such as β-alanine (β) motifs to replace pyrroles (Py), and increasing compound charge using N-terminal cationic substituents. To better understand the variations in kinetics and affinities caused by these modifications on PA-DNA interactions, a comprehensive set of PAs with different numbers and positions of β and different types of N-cationic groups was systematically designed and synthesized to bind their cognate sequence, the λB motif. The λB motif is also a strong binding promoter site of the major groove targeting transcription factor PU.1. The PA binding affinities and kinetics were evaluated using a spectrum of powerful biophysical methods: thermal melting, biosensor surface plasmon resonance and circular dichroism. The results show that β inserts affect PA-DNA interactions in a number and position dependent manner. Specifically, a β replacement between two imidazole heterocycles (ImβIm) generally strengthens binding. In addition, *N*-terminal cationic groups can accelerate the association between PA and DNA, but the bulky size of TMG can cause steric hindrance and unfavourable repulsive electrostatic interactions in some PAs. The future design of stronger binding PA requires careful combination of βs and cationic substituents.

2.2 Introduction

Polyamides (PA) are heterocyclic cations that consist of N-methylpyrrole (Py) and N-methylimidazole (Im) aromatic rings linked by amide bonds (1-4). They are potentially useful in applications such as gene therapeutics and inhibiting protein-DNA interactions (5-7). Derived from the natural DNA minor groove binding agents, netropsin and distamycin, PAs have been developed extensively over the years to enhance their DNA-binding affinity and specificity. Even

though targeting the relatively narrow DNA minor groove, distamycin was found to form 2:1 antiparallel stacked dimer complexes with AT-rich DNA (8). The finding led to the concept of covalently linking two PA monomers by a flexible linker that can mimic the bend of a hairpin. Therefore, hairpin PAs can form a 1:1 side-by-side complex with the two strands of DNA minor groove at higher affinity and specificity than traditional PAs (1, 9-11). Based on the hydrogen bonding interactions of the complex and its pairwise heterocycle stacking properties, general PA-DNA recognition rules were developed and have advanced molecular design (3, 10, 12). However, challenges still exist, since elongating hairpin PAs to eight or more rings leads to an over-curved structure which no longer matches the minor groove shape of B-form DNA. The resulting shape mismatch for long PAs causes a loss of hydrogen bonding strength that erodes both binding affinity and specificity (13, 14). Incorporating a flexible motif such as β-alanine (β) to serve as a replacement of internal Py moieties has been reported to help compensate for the structural incompatibility and improve binding (15-18). However, some literature results show a variety of binding affinity effects of β-inserts, including a decrease in binding affinity of some PAs by the introduction of βs. It is quite important to notice that the insertion of β at different positions of the same molecule has been shown to have very diverse effects on binding affinity (17, 19-21). It is essential, therefore, to understand the complex role of β insertion in PA-DNA interactions to guide the molecular design of better PAs.

Another challenge of large size hairpin PAs which contain more than six heterocyclic rings is aggregation and low solubility at high concentrations (20, 22, 23). The natural product, netropsin, contains a cationic guanidine moiety at its *N*-terminus, which presumably enhances both the solubility and its affinity to anionic DNA sequences (24). A previous study has demonstrated that the replacement of the monocationic Dp group [3-(dimethylamino)propylamine] by a

dicationic Ta group (3,3'-diamino-N-methyldipropylamine) could maintain the DNA binding mode and affinity of eight-ring hairpin PAs as well as significantly reduce PA aggregation (20). Bashkin and coworkers have also reported that guanidinyl-substituted PAs have better antiviral activities against human papillomavirus (HPV) than traditional PAs (25). Thus, it is important to incorporate cationic substituents into PA, explore their effects on PA-DNA binding interactions, and better understand how to modulate the solution properties of PAs.

2.2.1 Compound Design

In order to evaluate the effects of β inserts and cationic substituents on PA-DNA interactions, a series of hairpin PAs was systematically designed that covered reasonable internal exchanges of Py to β (Fig. 2.1A). In parallel, cationic substituents were also added to each counterpart with the exception of two molecules that have β inserts close to the *C*-terminus (Fig. 2.1A, KA2114 and KA2115). This is because β-alanine confers more flexibility to PAs when inserted close to a terminus than when positioned internally. With one positive charge at the *N*-terminus and two positive charges at the *C*-terminus, strong electrostatic repulsions would be expected to disrupt stacking of PA heterocycles and interfere with binding to the targeted DNA site. For this reason, KA2114 and KA2115 do not have cationic substituted counterparts.

The molecules in Fig. 2.1A are categorized into two groups: β-alanine inserts and cationic substituents, based on their modifications. To further assist data interpretation, all the heterocycles were numbered from 1 to 8 starting from the N-terminus. The parent PA, KA2035, contains eight heterocyclic rings and an N-terminal formamido group that is identical to that found in natural product Distamycin A. In the β-alanine inserts group, analogs of KA2035 that contain a single β replacement (KA2034 at ring 3, KA2041 at ring 6, and KA2114 at ring 7) and two βs (KA2040 at rings 3 and 6, and KA2115 at rings 3 and 7) at three different Py positions have been developed

and synthesized. In the cationic substituents group: KJK6162 has the same heterocyclic rings as KA2035 but incorporates tetramethylguanidine (TMG) at the N-terminus. The TMG cationic N-terminal containing molecules also include FH1026 and FH1028, both of which have a single β insert, but at rings number 6 and number 3, respectively. Another member in this group is FH1024, which possesses the same composition with KA2040 and a TMG at the N-terminus.

The designed PA-DNA binding site is AGTGA. The binding site is contained in the λB promoter region, to which transcription factor PU.1 binds. Therefore, λB sequence is chosen as the target DNA sequence in this study to screen potential PU.1-DNA inhibitors in the future. This is important in that PU.1 is involved in many physiological diseases such as auto-immune diseases and acute myeloid leukemia (AML) (26, 27). Extensive biophysical studies were carried out with the systematically designed compounds and their cognate and mutant DNAs. The results showed that β-inserts affect PA-DNA binding affinity in a manner that is dependent on PA composition and β position. For the current set of PA, a clear binding pattern between β-inserted PA and DNA was developed. The effects of flanking base pairs on PA-DNA binding were also investigated.

2.3 Materials and Methods

2.3.1 DNA and Compound synthesis

All the DNA oligomers in this study were purchased from Integrated DNA Technologies, Inc. (IDT, Coralville, IA) and the purity is checked by ESI-MS. The DNAs used are shown below: λB DNA: 5'-Biotin-CCAAATAAAAGG**AAGTGAA**ACCAAG*CTCT*CTTGGTTTCAC TTCCTTTTATTTGG-3';SC1 DNA: 5'-Biotin-CGGCCAAGCCGG**AAGTGAG**TGCC*TCTC*G GCACTCACTTCCGGCTTGGCCG-3'; GAGA mutant DNA: 5'-Biotin- CCAAATAAAAGA **GAGTGAA**ACCAAG*CTCT*CTTGGTTT CACTCTCTTTTATTTGG-3'

Short λB DNA: 5'-GGA**AGTGA**AC*CTCT*GTTCACTTCC-3'

Mutant 1: 5'-GGA**AGAGA**AC*CTCT*GTTCACTTCC-3'

Mutant 2: 5'-GGA**AGTGT**AC*CTCT*GTTCACTTCC-3'

Mutant 3: 5'-GGA**AGTTA**AC*CTCT*GTTCACTTCC-3'

Mutant 4: 5'-GGA**ACTGA**AC*CTCT*GTTCACTTCC-3'

Mutant 5: 5'-GGA**AGTCA**AC*CTCT*GTTCACTTCC-3'

The three long DNAs with biotin at 5' were used for SPR experiments. The short λB DNA was used for Thermal melting and Circular Dichroism experiments. The five mutant sequences were used for Thermal melting.

Compound Synthesis Boc-β-alanine-PAM resin was purchased from Peptides International. Polyamide building blocks, monomers or dimers, were purchased from A Chemtek, Inc. (Worcester, MA) and purification of polyamides was done by the Boc-PAM solid phase method as reported (28), with some modifications. In particular, dimer building blocks were purchased from A Chemtek or were prepared in-house and used in place of sequential monomers. This was especially the case when monomer coupling would have involved the poorly-nucleophilic imidazole amino group from the growing, resin-bound polyamide reacting with an active ester in solution, as per Dervan's original recommendations (28). Application of dimer building blocks are indicated by parentheses in polyamide syntheses, as for FH1024: TMG-(PyIm)-β-Im-γ-(Py-β)-(PyPy)-β-PAM resin. For example, (PyPy) indicates that the pyrrole-pyrrole dimer was used. When unavailable commercially, dimer building blocks were made by standard methods as previously described (20). The final products were purified by reverse-phase HPLC and characterized with NMR and HR mass spectrometry. The TMG substituted PAs were synthesized by virtually identical solid phase methods up to the pyrrole group (4-amino-N-methylpyrrole-2-carboxyamide). To generate the final building block where TMG is connected to

the H2N-Py moiety, HATU (29) has been used (30). The PA concentration was determined by accurate weighting on a five place analytical balance and dissolving the compound in an appropriate volume of solvent.

KA2034: HRMS calc'd for $C_{56}H_{71}N_{21}O_{11}$ 1213.56452, found 1213.5574

^{1}H NMR (600 MHz, DMSO-d_6) δ = 10.37 (s, 1 H), 10.35 (s, 1 H), 10.05 (d, *J* = 1.2 Hz, 1 H), 9.93 (s, 1 H), 9.89 (s, 1 H), 9.89 (s, 1 H), 9.82 (s, 1 H), 9.31 (br. s., 1 H), 8.10 (d, *J* = 1.8 Hz, 1 H), 8.05 (t, *J* = 5.9 Hz, 1 H), 8.02 (t, *J* = 5.7 Hz, 1 H), 7.97 (t, *J* = 6.2 Hz, 1 H), 7.88 (t, *J* = 6.2 Hz, 1 H), 7.50 (s, 1 H), 7.40 (s, 1 H), 7.28 (d, *J* = 1.8 Hz, 1 H), 7.25 - 7.21 (m, 2 H), 7.17 (d, *J* = 1.8 Hz, 1 H), 7.15 (d, *J* = 1.8 Hz, 1 H), 7.09 - 7.05 (m, 2 H), 6.98 (d, *J* = 2.1 Hz, 1 H), 6.88 (d, *J* = 1.8 Hz, 1 H), 6.87 (d, *J* = 1.8 Hz, 1 H), 3.94 (s, 3 H), 3.91 (s, 3 H), 3.85 (s, 3 H), 3.84 (s, 3 H), 3.83 (s, 3 H), 3.83 (s, 3 H), 3.81 (s, 3 H), 3.52 (q, *J* = 6.7 Hz, 2 H), 3.42 - 3.35 (m, 2 H), 3.27 (q, *J* = 6.6 Hz, 2 H), 3.12 (q, *J* = 6.5 Hz, 2 H), 3.04 - 2.97 (m, 2 H), 2.75 (d, *J* = 5.0 Hz, 6 H), 2.60 (t, *J* = 6.5 Hz, 2 H), 2.35 (t, *J* = 7.2 Hz, 2 H), 2.27 (t, *J* = 7.5 Hz, 2 H), 1.83 - 1.77 (m, 2 H), 1.77 - 1.70 (m, 2 H)

^{13}C NMR (151 MHz, DMSO-d_6) δ = 172.5, 171.0, 169.0, 168.5, 161.2, 158.6, 158.5, 158.5, 158.4, 158.0, 157.9, 157.8, 136.1, 135.8, 134.0, 133.6, 122.8, 122.7, 122.7, 122.7, 122.2, 122.2, 122.1, 122.0, 122.0, 120.7, 119.3, 118.5, 118.4, 118.1, 117.9, 114.3, 113.6, 106.9, 104.8, 104.7, 104.3, 103.9, 54.7, 42.3, 40.0, 38.1, 36.2, 36.0, 35.9, 35.6, 35.4, 34.9, 34.9, 34.7, 33.1, 25.6, 24.5

KA2035: HRMS calc'd for $C_{59}H_{72}N_{22}O_{11}$ 1264.57542, found 1264.5696

^{1}H NMR (600 MHz, DMSO-d_6) δ = 10.31 (s, 1 H), 10.26 (s, 1 H), 10.07 (s, 1 H), 9.93 (s, 1 H), 9.90 (s, 2 H), 9.89 (s, 1 H), 9.84 (s, 1 H), 9.24 (br. s., 1 H), 8.12 (d, *J* = 1.8 Hz, 1 H), 8.07 - 7.99 (m, 4 H), 7.55 (s, 1 H), 7.50 (s, 1 H), 7.39 (d, *J* = 1.8 Hz, 1 H), 7.28 (d, *J* = 1.8 Hz, 1 H), 7.22 (d, *J* = 1.8 Hz, 1 H), 7.22 (d, *J* = 1.8 Hz, 1 H), 7.17 - 7.15 (m, 2 H), 7.15 (d, *J* = 1.8 Hz, 1 H), 7.07

(s, 2 H), 7.00 (d, *J* = 1.8 Hz, 1 H), 6.88 (d, *J* = 1.2 Hz, 2 H), 3.98 (s, 3 H), 3.94 (s, 3 H), 3.87 (s, 3 H), 3.85 (s, 3 H), 3.85 (s, 3 H), 3.84 (s, 3 H), 3.83 (s, 3 H), 3.80 (s, 3 H), 3.42 - 3.35 (m, 2 H), 3.30 (q, *J* = 6.8 Hz, 2 H), 3.12 (q, *J* = 6.5 Hz, 2 H), 3.03 - 2.97 (m, 2 H), 2.75 (d, *J* = 4.7 Hz, 6 H), 2.35 (t, *J* = 7.3 Hz, 2 H), 2.28 (t, *J* = 7.6 Hz, 2 H), 1.85 - 1.78 (m, *J* = 7.1, 7.1, 14.5 Hz, 2 H), 1.78 - 1.71 (m, 2 H)

^{13}C NMR (151 MHz, DMSO-d_6) δ = 171.0, 169.0, 161.2, 158.7, 158.6, 158.6, 158.5, 158.1, 157.9, 155.7, 136.1, 135.9, 134.1, 134.1, 134.0, 122.8, 122.7, 122.7, 122.7, 122.2, 122.2, 122.1, 122.0, 122.0, 121.2, 120.8, 119.3, 118.5, 118.1, 117.9, 114.7, 114.3, 105.5, 104.8, 104.8, 104.3, 104.0, 54.7, 42.3, 40.4, 40.0, 38.0, 36.3, 36.2, 36.1, 35.9, 35.6, 35.4, 34.9, 34.9, 33.1, 25.6, 24.5

EA calc'd for $C_{59}H_{72}N_{22}O_{11}(3CF_3CO_2H)(9H_2O)$: C, 44.12%; H, 5.30%; N, 17.41%. Found: C, 43.87; H, 4.92%; N, 17.13%.

KA2040: HRMS: calc'd for $C_{53}H_{70}N_{20}O_{11}$ 1162.55362, found 1162.5466

^{1}H NMR (600 MHz, DMSO-d_6) δ = 10.37 (s, 1 H), 10.35 (s, 1 H), 10.05 (s, 1 H), 9.88 (s, 1 H), 9.84 (s, 1 H), 9.75 (s, 1 H), 9.25 (br. s., 1 H), 8.10 (d, *J* = 1.5 Hz, 1 H), 8.07 - 8.03 (m, 2 H), 8.02 (t, *J* = 5.7 Hz, 1 H), 7.96 (t, *J* = 6.0 Hz, 1 H), 7.89 (t, *J* = 6.0 Hz, 1 H), 7.50 (s, 1 H), 7.39 (s, 1 H), 7.28 (d, *J* = 1.8 Hz, 1 H), 7.16 (d, *J* = 1.8 Hz, 1 H), 7.15 (d, *J* = 1.8 Hz, 1 H), 7.09 (d, *J* = 1.8 Hz, 1 H), 6.98 (d, *J* = 2.1 Hz, 1 H), 6.86 (d, *J* = 1.8 Hz, 1 H), 6.85 (d, *J* = 1.8 Hz, 1 H), 6.63 (d, *J* = 1.5 Hz, 1 H), 3.94 (s, 3 H), 3.90 (s, 3 H), 3.83 (s, 3 H), 3.82 (s, 3 H), 3.80 (s, 3 H), 3.77 (s, 3 H), 3.55 - 3.49 (m, 2 H), 3.45 - 3.40 (m, 2 H), 3.40 - 3.35 (m, 2 H), 3.28 - 3.22 (m, 2 H), 3.14 - 3.08 (m, 2 H), 3.04 - 2.97 (m, 2 H), 2.74 (d, *J* = 5.0 Hz, 6 H), 2.60 (t, *J* = 6.6 Hz, 2 H), 2.52 - 2.47 (m, 2 H), 2.35 (t, *J* = 7.2 Hz, 2 H), 2.24 (t, *J* = 7.5 Hz, 2 H), 1.81 - 1.70 (m, 4 H)

^{13}C NMR (151 MHz, DMSO-d_6) δ = 171.0, 169.0, 168.5, 167.8, 162.7, 161.2, 161.2, 158.6, 158.5, 158.4, 158.4, 158.1, 157.9, 136.1, 135.8, 133.9, 133.6, 122.7, 122.7, 122.7, 122.0, 122.0,

121.9, 120.7, 119.3, 118.1, 117.9, 117.6, 117.2, 114.3, 113.5, 104.8, 104.3, 104.0, 103.4, 54.7, 42.3, 40.0, 38.1, 36.2, 36.0, 35.9, 35.9, 35.8, 35.6, 35.5, 35.4, 34.9, 34.9, 34.7, 33.1, 25.5, 24.5

EA calc'd for $C_{53}H_{70}N_{20}O_{11}(3CF_3CO_2H)(6H_2O)$: C, 43.92%; H, 5.31%. Found: C, 44.15; H, 4.95%.

KA2041: HRMS: calc'd for $C_{53}H_{70}N_{20}O_{11}$ 1162.55362, found 1162.5466

^{1}H NMR (600 MHz, DMSO-d_6) δ = 10.31 (s, 1 H), 10.26 (s, 1 H), 10.07 (d, *J* = 0.9 Hz, 1 H), 9.90 (s, 1 H), 9.88 (s, 1 H), 9.84 (s, 1 H), 9.77 (s, 1 H), 9.26 (br. s., 1 H), 8.12 (d, *J* = 1.8 Hz, 1 H), 8.08 - 7.99 (m, 4 H), 7.55 (s, 1 H), 7.49 (s, 1 H), 7.39 (d, *J* = 1.5 Hz, 1 H), 7.28 (d, *J* = 1.8 Hz, 1 H), 7.16 (d, *J* = 1.8 Hz, 1 H), 7.15 (d, *J* = 1.8 Hz, 1 H), 7.14 (d, *J* = 1.8 Hz, 1 H), 7.10 (d, *J* = 1.8 Hz, 1 H), 7.00 (d, *J* = 1.8 Hz, 1 H), 6.86 (d, *J* = 1.8 Hz, 1 H), 6.85 (d, *J* = 2.1 Hz, 1 H), 6.64 (d, *J* = 1.8 Hz, 1 H), 3.98 (s, 3 H), 3.93 (s, 3 H), 3.87 (s, 3 H), 3.85 (s, 3 H), 3.81 (s, 3 H), 3.79 (s, 3 H), 3.78 (s, 3 H), 3.45 - 3.40 (m, 2 H), 3.40 - 3.35 (m, 2 H), 3.28 (q, *J* = 6.7 Hz, 2 H), 3.11 (q, *J* = 6.5 Hz, 2 H), 3.03 - 2.98 (m, 2 H), 2.74 (d, *J* = 5.0 Hz, 6 H), 2.51 - 2.48 (m, 2 H), 2.35 (t, *J* = 7.2 Hz, 2 H), 2.25 (t, *J* = 7.5 Hz, 2 H), 1.79 (td, *J* = 7.3, 14.5 Hz, 2 H), 1.76 - 1.71 (m, 2 H)

^{13}C NMR (151 MHz, DMSO-d_6) δ = 174.1, 172.1, 170.9, 164.4, 164.3, 161.8, 161.7, 161.7, 161.5, 161.3, 161.1, 161.0, 160.9, 158.9, 139.2, 139.1, 137.2, 137.2, 125.9, 125.8, 125.3, 125.2, 125.2, 125.0, 125.0, 124.3, 124.0, 122.5, 121.3, 121.1, 120.7, 120.1, 117.9, 117.4, 108.6, 107.9, 107.4, 107.1, 106.5, 57.8, 45.4, 43.2, 41.2, 39.4, 39.4, 39.2, 39.1, 39.0, 38.7, 38.6, 38.5, 38.0, 38.0, 36.2, 28.7, 27.6

KA2114: HRMS: calc'd for $C_{58}H_{76}N_{22}O_{11}$ 1256.6064, found 1256.6028

^{1}H NMR (600 MHz, DMSO-d_6) δ = 10.31 (s, 1 H), 10.26 (s, 1 H), 10.09 (s, 1 H), 9.90 (s, 1 H), 9.85 (s, 2 H), 9.83 (s, 1 H), 9.68 - 9.60 (m, 1 H), 8.12 (d, *J* = 1.2 Hz, 1 H), 8.10 - 8.00 (m, 5 H), 7.89 (br. s., 4 H), 7.55 (s, 1 H), 7.50 (s, 1 H), 7.39 (d, *J* = 1.2 Hz, 1 H), 7.29 (d, *J* = 1.8 Hz, 1

H), 7.17 (d, *J* = 1.2 Hz, 1 H), 7.15 (d, *J* = 1.8 Hz, 2 H), 7.10 (d, *J* = 1.8 Hz, 1 H), 7.00 (d, *J* = 1.8 Hz, 1 H), 6.85 (d, *J* = 1.2 Hz, 1 H), 6.84 (d, *J* = 1.2 Hz, 1 H), 6.67 (d, *J* = 1.8 Hz, 1 H), 3.98 (s, 3 H), 3.94 (s, 3 H), 3.87 (s, 3 H), 3.85 (s, 3 H), 3.81 (s, 3 H), 3.80 (s, 3 H), 3.77 (s, 3 H), 3.46 - 3.39 (m, 2 H), 3.39 - 3.33 (m, 2 H), 3.29 (q, *J* = 6.7 Hz, 2 H), 3.21 - 3.14 (m, 1 H), 3.14 - 3.09 (m, 2 H), 3.09 - 2.95 (m, 3 H), 2.91 - 2.82 (m, 2 H), 2.73 (d, *J* = 4.7 Hz, 3 H), 2.53 - 2.46 (m, 2 H), 2.34 (t, *J* = 7.0 Hz, 2 H), 2.28 (t, *J* = 7.3 Hz, 2 H), 1.96 - 1.86 (m, 2 H), 1.84 - 1.72 (m, 4 H)

^{13}C NMR (151 MHz, DMSO-d_6) δ = 174.1, 172.2, 170.9, 165.8, 164.4, 164.3, 161.8, 161.7, 161.7, 161.7, 161.5, 161.4, 161.2, 161.0, 161.0, 158.9, 139.2, 139.1, 137.2, 137.2, 125.9, 125.9, 125.8, 125.3, 125.2, 125.2, 125.1, 125.0, 124.4, 124.0, 122.5, 121.2, 121.1, 120.8, 120.5, 118.6, 117.9, 117.4, 108.6, 107.9, 107.3, 107.0, 106.7, 56.4, 55.2, 43.2, 41.2, 39.4, 39.4, 39.3, 39.2, 39.1, 38.9, 38.7, 38.7, 38.6, 38.5, 38.0, 38.0, 36.3, 28.8, 27.1, 24.9

EA calc'd for $C_{58}H_{76}N_{22}O_{11}(4CF_3CO_2H)(7H_2O)$: C, 43.09%; H, 5.15%; N, 16.75%;. Found: C, 43.10; H, 4.77%; N, 16.07%

KA2115: HRMS: Calc'd for $C_{55}H_{75}N_{21}O_{11}$ 1205.5954, found 1205.5932

^{1}H NMR (600 MHz, DMSO-d_6) δ = 10.37 (s, 1 H), 10.35 (s, 1 H), 10.06 (s, 1 H), 9.85 (s, 1 H), 9.84 (s, 1 H), 9.81 (s, 1 H), 9.71 - 9.60 (m, 1 H), 8.11 (d, *J* = 1.2 Hz, 1 H), 8.08 (t, *J* = 5.9 Hz, 1 H), 8.06 - 8.00 (m, 2 H), 7.97 (t, *J* = 5.9 Hz, 1 H), 7.93 - 7.82 (m, 5 H), 7.50 (s, 1 H), 7.40 (s, 1 H), 7.28 (d, *J* = 1.8 Hz, 1 H), 7.17 (d, *J* = 1.8 Hz, 1 H), 7.14 (d, *J* = 1.2 Hz, 1 H), 7.11 (d, *J* = 1.8 Hz, 1 H), 6.98 (d, *J* = 1.8 Hz, 1 H), 6.85 (d, *J* = 1.8 Hz, 1 H), 6.84 (d, *J* = 1.2 Hz, 1 H), 6.67 (d, *J* = 1.8 Hz, 1 H), 3.94 (s, 3 H), 3.91 (s, 3 H), 3.83 (s, 3 H), 3.81 (s, 3 H), 3.80 (s, 3 H), 3.78 (s, 3 H), 3.52 (q, *J* = 6.5 Hz, 2 H), 3.47 - 3.40 (m, 2 H), 3.40 - 3.32 (m, 2 H), 3.27 (q, *J* = 6.8 Hz, 2 H), 3.22 - 3.15 (m, 1 H), 3.15 - 3.09 (m, 2 H), 3.09 - 2.96 (m, 3 H), 2.92 - 2.82 (m, *J* = 5.9 Hz, 2

H), 2.73 (d, *J* = 4.7 Hz, 3 H), 2.60 (t, *J* = 6.5 Hz, 2 H), 2.53 - 2.46 (m, 2 H), 2.34 (t, *J* = 7.3 Hz, 2 H), 2.26 (t, *J* = 7.3 Hz, 2 H), 1.96 - 1.85 (m, 2 H), 1.83 - 1.73 (m, 4 H)

^{13}C NMR (151 MHz, DMSO-d_6) δ = 171.0, 169.1, 168.6, 167.8, 162.7, 161.3, 161.2, 158.7, 158.6, 158.5, 158.4, 158.4, 158.4, 158.2, 158.0, 157.9, 136.1, 135.8, 134.0, 133.6, 122.8, 122.8, 122.7, 122.1, 122.0, 122.0, 121.9, 121.1, 120.8, 119.4, 119.1, 118.8, 118.1, 117.9, 117.6, 116.9, 115.0, 114.4, 113.6, 113.0, 104.8, 104.2, 103.9, 103.6, 53.3, 52.1, 40.0, 38.1, 36.3, 36.3, 36.2, 36.0, 35.9, 35.8, 35.6, 35.6, 35.5, 35.4, 34.9, 34.9, 34.7, 33.1, 25.6, 24.0, 21.8

EA calc'd for $C_{55}H_{75}N_{21}O_{11}(5CF_3CO_2H)(6H_2O)$: C, 41.43%; H, 4.92%; N, 15.61%; F, 15.12%. Found: C, 41.51; H, 4.48%; N, 15.34%; F, 14.40%.

FH1024 ^{1}H NMR (500 MHz, DMSO-d_6) δ =10.46(s, 1 H), 10.37(s, 1 H), 9.87(s, 1 H), 9.85(s, 1 H), 9.76(s, 1 H), 9.67(s, 1 H), 9.44(br. s, 1 H), 8.09~8.02(m, 3 H), 7.96(t, J=5.84Hz, 1 H), 7.84(s, 1 H), 7.81(t, J=5.85Hz, 3 H), 7.52(s, 1 H), 7.38(s, 1 H), 7.16(d, J=1.8Hz, 1 H), 7.15(d, J=1.8Hz, 1 H), 7.09(d, J=1.8Hz, 1 H), 6.90(d, J=1.8Hz, 1 H), 6.87(d, J=1.8Hz, 1 H), 6.81(d, J=1.8Hz, 1 H), 6.63(d, J=1.8Hz, 1 H), 3.94(s, 3 H), 3.9(s, 3 H), 3.87(s, 3 H), 3.81(s, 3 H), 3.79(s, 3 H), 3.77(s, 3 H), 3.53(q, J=6.01Hz, 2 H), 3.4(m, 4 H), 3.25(q, J=6.28Hz, 2 H), 3.14(m, 4 H), 3.04(m, 4 H), 2.91-2.87(m, 12 H), 2.74(d, J=5.03Hz, 4 H), 2.64(m, 1 H), 2.6(t, J=6.28Hz, 2 H), 2.36(t, J=7.2Hz, 3 H), 2.24(t, J=7.14Hz, 2 H), 1.9(m, 2 H), 1.71(m, 4 H)

FH1026 ^{1}H NMR (500 MHz, DMSO-d_6) δ =10.46(s, 1 H), 10.25(s, 1 H), 9.88(s, 1 H), 9.84(s, 1 H), 9.79(s, 1 H), 9.77(s, 1 H), 9.70(s, 1 H), 9.46(br. s, 1 H), 8.09~8.0(m, 4 H), 7.83(d, 4 H), 7.57(s, 1 H), 7.49(s, 1 H), 7.37(d, J=1.53Hz, 1 H), 7.19(d, J=1.53Hz, 1 H), 7.15(d, J=1.53Hz, 1 H), 7.10(d, J=1.53Hz, 1 H), 6.92(d, J=1.68Hz, 1 H), 6.86(d, J=1.53Hz, 1 H), 6.83(d, J=1.53Hz, 1 H), 6.63(d, J=1.53Hz, 1 H), 3.99(s, 3 H), 3.96(s, 3 H), 3.93(s, 3 H), 3.89(s, 3 H), 3.82(s, 3 H),

3.80(s, 3 H), 3.78(s, 3 H), 3.44~3.26(m, 8 H), 3.173~3.00(m, 12 H), 2.99-2.87(m, 4 H), 2.74(d, J=4.9Hz, 4 H), 2.63(m, 1 H), 2.36(m, 3 H), 2.25(t, J=7.64Hz, 2 H), 1.91(m, 4 H), 1.71(m, 4 H)

FH1028 ^{1}H NMR (500 MHz, DMSO-d_6) δ =10.46(s, 1 H), 10.37s, 1 H), 9.92(s, 1 H), 9.88(s, 1 H), 9.82(s, 1 H), 9.67(s, 1 H), 9.44(br. s, 1 H), 8.09 (t, J=5.76Hz, 1 H), 8.04 (t, J=5.76Hz, 1 H), 8.00 (t, J=5.76Hz, 1 H), 7.81(m, 4 H), 7.52(s, 1 H), 7.39(s, 1 H), 7.21(d, J=1.8Hz, 1 H), 7.20(d, J=1.8Hz, 1 H), 7.19(d, J=1.8Hz, 2 H), 7.09(d, J=1.8Hz, 1 H), 6.91(d, J=1.8Hz, 1 H), 6.87(d, J=1.8Hz, 1 H), 6.81(d, J=1.8Hz, 1 H), 3.93(s, 3 H), 3.91(s, 3 H), 3.87(s, 3 H), 3.86(s, 3 H), 3.85(s, 3 H), 3.83(s, 3 H), 3.81(s, 3 H), 3.53(q, J=6.32Hz, 2 H), 3.38 (q, J=6.92Hz, 2 H), 3.27(q, J=6.62Hz, 2 H), 3.15~3.01(m, 8 H), 2.99-2.87(m, 10 H), 2.74(d, J= 4.9Hz, 4 H), 2.63(m, 1 H), 2.59(t, J=6.62Hz, 2 H), 2.36(t, J=7.23Hz, 3 H), 2.25(t, J=7.53Hz, 2 H), 1.91(m, 4 H), 1.71(m, 4 H)

KJK6162 1H NMR (300 MHz, DMSO-d6) δ =10.50(s, 1 H), 10.34(s, 1 H), 9.94(m, 2H), 9.89(m, 2 H), 9.70(s, 1 H), 8.74(d, J=4.56Hz, 2 H), 8.51(d, J=8.45Hz, 2 H), 8.12(s, 1 H), 7.56(s, 1 H), 7.50(q, J=4.42Hz, 4 H), 7.38(d, J=1.38Hz, 1 H), 7.23(d, J=1.83Hz, 1 H), 7.17(d, J=6.89Hz, 2 H), 7.11(d, J=1.8Hz, 1 H), 7.04(d, J=1.83Hz, 1 H), 6.91(d, J=1.54Hz, 1 H), 6.86(d, J=1.97Hz, J=4.09Hz, 2 H), 6.80(t, J=1.97Hz, 1 H), 6.59(s, 1 H), 5.75(s, 5 H), 4.19(q, J=5.45Hz, 3H), 3.98(s, 3H), 3.93(s, 3 H), 3.87(s, 3 H), 3.86(s, 3 H), 3.84(s, 3 H), 3.82(s, 3 H), 3.81(s, 3 H), 3.79(s, 3 H), 3.67(s , 2 H), 3.60(s , 4 H), 3.17(s, 6 H), 3.15(s, 6 H), 3.08(t, J=0.73 Hz, 1 H), 2.88(t, J=1.90Hz, 1 H), 2.68(s, 2H), 2.08(s, 2H), 1.22(m, 2 H), 1.16(t, J=7.38Hz, 2 H)

2.3.2 Biophysical methods

Surface Plasmon Resonance (SPR) SPR measurements were performed using Biacore T200 optical biosensor systems (GE Healthcare, Inc., Piscataway, NJ). Filtered and degassed HEPES buffer (10 mM HEPES, 400 mM NaCl, 1 mM EDTA and 0.05% v/v surfactant P20, pH 7.4) was used in SPR experiments. The measurement process is the same as described previously

(21). A biotinylated hairpin DNA (5'-biotin-CCAAATAAAAGGAAGTGAAACCAAG*CTCT* CTTGGTTTCACTTCCTTTTATTT GG, Fig.2.1) was immobilized on the surface of a sensor chip pre-coated with streptavidin. A reference channel was prepared without DNA immobilization for baseline correction. To form the PA-DNA complex, an increasing concentration of PA was injected onto the chip surface at a flow rate of 100 µL/min for 210 seconds until a steady state was reached. This high flow rate was used to minimize mass transfer effects. The association was followed by a pure running buffer flow at the same rate, causing the complex to dissociate. At the end of each cycle, 1 M NaCl was used as regeneration solution and injected over the surface to completely wash off the residual PA and prepare the surface for the next cycle.

The kinetic analyses were performed using standard 1:1 global fitting model with mass transport parameters incorporated as described previously (31, 32). The equilibrium constant of some compounds were determined using a steady state fit, because the kinetic parameters went beyond the detection limit and were too fast to determine. This often features the weak binding compounds.

Thermal Melting (Tm) Thermal melting studies were conducted on a Cary 300 Bio UV/visible spectrophotometer (Varian) in filtered buffer containing 10 mM HEPES (pH 7.4), 50 mM NaCl and 1 mM EDTA. Samples were heated at a rate of 0.5 °C/min, and the corresponding absorbance at 260 nm was recorded and plotted against the temperature. The T_m of DNA was measured at 3 µM, and an equivalent amount of PA was added to obtain the T_m of PA-DNA complex. The difference between the T_m of DNA in the absence and presence of PA is thus a ΔT_m.

Circular Dichroism Spectroscopy (CD) CD spectra were obtained using a Jasco J-810 spectrometer (Jasco Inc., Easton, MD) with a scan range from 400 nm to 230 nm at 25 °C. The spectra were averaged over four scans with a scan speed of 50 nm·min^{-1} and a buffer blank

correction. A 5μM DNA solution was firstly scanned and the PAs were then titrated into the same cuvette at increasing concentration ratios. The complex was scanned under the same condition. A hairpin DNA without a biotin label (5'-GGAAGTGAAC*CTCT*GTTCACTTCC-3') was used and the experimental buffer is the same as that used in thermal melting study.

2.4 Results

2.4.1 Quantitative evaluation of PA-DNA binding kinetics and affinity

β-inserts Biosensor-Surface Plasmon Resonance (SPR) is a label-free technique with excellent sensitivity that can quantitatively monitor biomolecular binding interactions in real time. Representative SPR sensorgrams are shown in Fig 2.2. It is noticeable that the dissociation rate is getting faster from strong to weak binders. The other sensorgrams are listed in Fig. 2.3. The kinetic rate constants and equilibrium binding affinities of eleven designed PAs were determined using SPR and summarized in Table 2.1. The six PAs in the β-insert group have a formamido (F) group at the N-terminus and β at different positions. KA2115, with two βs replacing Py rings at positions 3 and 7, leads the group with the highest binding affinity (0.49 ± 0.01 nM). KA2040, another PA with two βs but at rings 3 and 6, binds to the cognate sequence essentially equally as strong (0.54 ± 0.01 nM). A single β inserted PA at ring 3, KA2034, displayed a slightly weaker but quite comparable binding affinity (0.67 ± 0.01 nM) to KA2115 and KA2040. These results suggest that the position and number of β inserts in the lower strand (rings 5-8) is not a dominant factor for this hairpin PA-DNA interaction, and do not contribute significantly to binding affinity enhancement. The unmodified parent molecule, KA2035, exhibits a moderate affinity (1.30 ± 0.01 nM) with regard to the last three molecules. This is primarily caused by its relatively lower association rate (k_a). KA2035, however, has the second lowest dissociation rate (k_d) among all the PAs tested, which compensated for its low k_a and rendered KA2035 a relatively strong binder. KA2041, with

only one β-alanine at position 6, binds the cognate DNA with a higher k_a but a much higher k_d compared to KA2035, resulting in weakened overall interactions. Nonetheless, this situation worsened when the replacement was moved to position 7, as in KA2114. The k_d of KA2114 is so high, around 20 times higher than the parent molecule KA2035, that it cannot obtain high binding affinity even with a higher k_a value than KA2035.

Cationic substitutions Relative to the β insertion group, in which all molecules conserve an unmodified N-terminus, four PAs were designed with further modification at the N-terminus with the specific cationic substituent, tetramethylguanidinyl (TMG). The binding affinities and kinetics of these PAs have been evaluated (Table 2.1). FH1028, the PA with a single β insert at ring 3 and a TMG at the N-terminus leads all the other PAs in binding affinity to the cognate DNA (0.16 ± 0.02 nM). The high affinity is attributed to its high k_a and relatively low k_d. Compared to its analog KA2034 which has the identical β modification but different N-terminal substituents, FH1028 has a five-fold higher k_a which contributes to the strengthened binding affinity. This result indicates that the TMG group can accelerate PA-DNA binding. Similarly, FH1024 (N-TMG) and KA2040 (N-F) were compared to each other because they have identical β-inserts at positions 3 and 6. The N-TMG attached FH1024 has a k_a that is around two times higher than KA2040 (N-F), but its drastically increased k_d makes it non-competitive with KA2040. Interestingly, KJK6162, modified directly from the parent, KA2035, by substituting F with TMG at the N-terminus, binds the λB sequence so weakly (60 times weaker than KA2035) that its kinetic rate constant cannot be determined. Similarly, FH1026, the N-TMG analog of KA2041, has a 20 times weaker binding affinity than KA2041, well out-of-range kinetic constants by SPR.

2.4.2 Screening of relative binding affinities with cognate and mutant DNAs by Thermal Melting (Tm)

Thermal melting is an effective way to evaluate the relative binding affinity and the selectivity of a particular PA among several DNA sequences. Tm of the cognate DNA and DNA mutants and their complexes with PAs are determined at 1:1 molar ratio. The difference between Tm values of the complex and DNA, ΔT_m, reflects the thermal stability of the complex, thus the binding affinity of PA. The values are listed in Table 2.2.

The comparison of mutant DNAs are made with each PA-cognate DNA complex and are categorized according to their extent of binding. ΔT_m values in ±5% relative to the ΔT_m of cognate DNA are designated as equal and uncolored, dark gray means slightly stronger (≤120%, >105%), gray means slightly weaker (<95%, ≥80%), light gray is categorized as weaker (<80%, ≥50%) and lightest gray is much weaker (<50%). The determined ΔTm values of PAs with cognate DNA are in good agreement with the SPR results.

The first mutant sequence evaluated is a Py recognizing T that is mutated to A. Both in β-insert and the cationic substituent group, the affinity is not affected significantly and is either equal or slightly higher or lower, even for the molecules where Py is replaced by β (single-β or double-β replacement). Particularly, KA2034 (f-PyImβIm-γ-PyPyPyPy-β-Dp) showed slightly weaker binding when β/Py targeting T·A is switched to A·T. In a similar case, KA2041 (f-PyImPyIm-γ-PyβPyPy-β-Dp) who has Py/β targeting T·A exhibited slightly higher affinity to the A·T mutant. These results indicate that T·A prefers β/Py to Py/β, and vice versa. The second mutant is made by switching the terminal amine recognizing A to T. In both groups, most molecules are not affected by this change and the binding affinity stays comparable to that of the cognate DNA. What is interesting is, KA2040 and FH1024, both of which have double-β replacement where

Py/Py is substituted by β/β, have slightly elevated binding affinities. The third mutant is considered as a more dramatic mutation, in which the Im recognizing G is mutated to T. As a result, an apparent drop of binding affinity is observed for most PAs. However, as a strong binder to cognate DNA, FH1028 and KA2115 still binds to mutant 3 quite strongly, ΔT_m = 14.8, and 10.5, respectively. Interestingly, the binding of KA2041 and KA2114 to mutant 3 is comparable to that of cognate DNA, especially KA2114. This result indicates that even though Im recognizes G specifically, it can still tolerate T to some extent, so that the complex can still be formed. Mutants 4 and 5 are made by switching G to C. Much weaker binding appeared at this level of mutation. Since in this case, Py is designated to target G, which is known with either one-β or two-β replacement, the Im/Py or Im/β combination is not very tolerant with C·G base pairs.

2.4.3 Effect of N-terminal cationic group on PA selectivity of DNA flanking sequences

To further evaluate the effect of the *N*-terminal cationic group on PA-DNA interactions, binding affinities of the same set of PAs to DNAs that have different mutated flanking sequences were measured and compared. The three DNAs used here are λB DNA (5'-Biotin-CCAAATAAAAGG**AAGTGAA**ACCAAG*CTCT*CTTGGTTTCACTT CCTTTTATTTGG-3'), SC1 DNA (5'-Biotin-CGGCCAAGCCGG**AAGTGAG**TGCC*TCTC*GGCACTCACTTCCGGC TTGGCCG-3'), and GAGA mutant DNA (5'-Biotin-CCAAATAAAAGA**GAGTGAA**ACC AAG*CTCT*CTTGGTTTCACTCTCTTTTATTTGG-3'). In λB DNA, the PA binding site -AGTGA- is surrounded by A on both 5' and 3' ends. The SC1 sequence has a G flanking at the 3' side, while GAGA sequence has a G at the 5' end. The binding to the DNAs was evaluated by SPR as described above and the SPR sensorgrams and binding results are shown in Fig. 2.4, Fig. 2.5 and Table 2.3. By comparing binding of λB and SC1 DNA, the N-F molecules (KA2035 plus β-

inserts group) have very similar binding affinities to both sequences. For the N-TMG molecules, the binding affinities to λB DNA are fairly strong. However, when tested against the SC1 DNA, their binding affinities significantly decreased, even to the point of no detectable binding. Interestingly, when we compare the binding to GAGA sequence with λB DNA, the N-TMG group of PAs show strong and comparable binding in general, and binding of FH1024 is even stronger to GAGA than λB DNA. Fig. 2.6B shows the binding affinity of PAs to λB, SC1 and GAGA mutant sequences. It is clear that each PA in the β-inserts group has very close binding strength to all three DNAs, except that KA2115 has a slightly larger deviation. But in the cationic substitution group, hardly any traces of PA-SC1 binding (black column) can be seen. These results suggest that N-TMG PAs are quite sensitive and selective to the 3' flanking base pairs of -AGTGA-, but not 5', and they prefer A to G.

2.4.4 Evaluation of the PA binding mode by CD

In Figure 2.7, the large positively induced CD signals upon PA-DNA binding at around 300 nm to 370 nm, where PAs absorb while DNA signals do not interfere indicate a DNA minor groove binding mode as expected for hairpin PAs (33).

2.5 Discussion

2.5.1 Quantitative evaluation of PA-DNA binding kinetics and affinity

SPR has greatly facilitated the quantification of kinetics between small molecules and DNA. As important, sequence-specific DNA minor groove binders, polyamides have been extensively studied over the decades and are seen as potential drug targets for antiviral, antitumor and antibacterial therapeutics. Yet the position and number dependence of β-modified PA-DNA interaction still have many uncertainties and their features play a critical role in designing more specific and stronger binding PAs.

β-inserts As shown in Fig. 1B, along the λB sequence, the 5’ side of PU.1 binding site is an AT rich region. Previous literature has employed the A·T specific minor groove targeting compounds to bind to the 5’ side A·T base pairs in order to inhibit PU.1 binding (34). The efficient allosteric inhibition activity demonstrated by these compounds has encouraged us to further extend our target binding site to the 3’ side. A combination of 5’ side binding minor groove heterocyclic cations and 3’ targeting PAs should provide a very powerful inhibition potential.

The parent PA, KA2035 was thus designed, according to the recognition rules, to target the 3’ side AGTGA sequence. The PA has a strong binding affinity of 1.30 nM, and was able to successfully and precisely recognize the target sequence. β-alanine is a building block that is incorporated into PAs to increase their flexibility. In a successful effort to improve binding affinity, KA2035 was modified by replacing internal Py at different positions with β inserts. The SPR binding affinities of modified PAs vary in a β insert number-and-position dependent manner.

Specifically, when compared to KA2035, three PAs (KA2115, KA2040 and KA2034) have strengthened binding affinities. A common modification strategy adopted by all the three PAs is that Py heterocycle 3 is substituted by one β insert. It is highly likely that this position 3 β insert in the upper strand (rings 1-4) added the flexibility needed to the molecule to adjust to the cognate DNA minor groove curvature. However, this particular β also allows for relaxation and realignment of the PA to compensate for the different positions of H-bond donors and acceptors in Py and Im components of PAs, where Py uses an exocyclic amide NH as an H-bond donor to DNA (for A, T and C) and Im uses a cyclic N as an H-bond acceptor from G. The H-bonding locations on DNA are in register with each other, but those on the PA are not. Therefore, the locations of hydrogen bond donors and acceptors between PA and DNA are better oriented at more optimized distances and angles, resulting in strengthened hydrogen bonding with higher PA-DNA

binding affinity. On the other hand, a single β insert on the lower strand (rings 5-8), KA2041 and KA2114, gave no enhanced binding. This is due to the possibility that replacement of Py by a β at position 6 or 7 does not help the PA overcome its overall rigidity. Instead, a reduction of van der Waals contacts occurred when substituting Py 6 or 7 with a β. Interestingly, the β position in double β-substituted KA2115 (K_D = 0.49 nM) is a combination of βs in the single β-substituted KA2034 (K_D = 0.67 nM) and KA2114 (K_D = 4.1 nM), but the binding affinity of KA2115 is higher than either of the other two. Very similarly, the double β-substituted KA2040 (K_D = 0.54 nM) also has higher binding affinities than the single β-substituted KA2034 (K_D = 0.67 nM) and KA2041 (K_D = 1.62 nM). This is probably because that the combination of upper strand and lower strand β inserts gives PA even better flexibility to arrange itself to the optimal orientation for hydrogen bonding upon interacting with DNA, and because that optimal β-β stacking can occur when binding to certain DNA sequence.

The binding affinity is determined by the association and dissociation rate constants. From Table 2.1, it is clear that all PAs in the β-inserts group have higher association rate constants than the parent PA, KA2035. This is in contrast to a previous study in which the association rate constants of modified PAs were generally lower than those of unmodified compounds (20). Of note, the 2014 study focused on the dImImPyIm-γ-PyPyPyPy-β-Dp/Ta sequence and derivatives while the current work studies formamidoPyImPyIm-γ-PyPyPyPy-β-Dp/Ta and derivatives. The extra, in register H-bond provided by the potentially rotatable formyl group coupled with the disparate target sequences of the two studies and different minor groove widths may all contribute to fundamentally different binding kinetics. The opposite effects of β-alanine inserts on the association rate of these two (current and previous) sets of PAs indicate that changes in PA composition (Py and Im heterocycles arrangement and content) and targeted, cognate DNA

sequence respond differently to β inserts. After inserting β, the previous PAs were too flexible and needed time to adjust themselves to DNA upon binding, while the current set of β-inserted PAs fit better to the curvature of λB sequence than the parent molecule, resulting in more rapid binding.

The dissociation rate constants of those PAs with increased affinity are quite comparable to the parent PA, KA2035. For these tighter-binding PAs, as mentioned above, the upper strand β modification has greatly optimized hydrogen bond donor and acceptor orientations and thus strengthened hydrogen bonding between PAs and DNA (KA2034). But for the weak binding KA2041 and KA2114, their dissociation is so much faster that they can hardly maintain strong binding. It is noticeable that the single lower strand β replacement resulted in higher dissociation rate constant. Bashkin *et al.* built a docking model to compare a single-β-inserted PA2 (ImImβPy-γ-PyImPyPy-β-Dp) with an all-ring eight-ring PA1 (ImImPyPy-γ-PyImPyPy-β-Dp), and found that Py is more effectively stacked with both the adjacent PA strand and the DNA backbone than β-alanine (19). Higher dissociation rate constants for some β derivatives might be a result of loss of hydrophobic interactions.

Cationic substitutions Since DNA is highly negatively charged, adding positive charges on PAs should help enhance the electrostatic interactions between the two. More charged groups can also improve the solubility of PAs. Therefore, in addition to β inserts, additional cationic groups are attached to the N-termini to facilitate PA-DNA binding. SPR results showed varied effects of cationic substitutions on PA-DNA interactions.

As shown in Fig. 2.6A, FH1024 and FH1028, compared to KA2040 and KA2034, their analogs with formamido N-termini (N-F), exhibited slightly different binding affinities. FH1024 binds about three times weaker than KA2040 and FH1028 binds around three times tighter than KA2034. Interestingly, both TMG derivatives have higher association rate constants than their

formamido analogs. The results implied that N-TMG can help promote PA-DNA binding by adding extra charge-charge interactions. On the other hand, the binding affinities of N-TMG KJK6162 and FH1026 dropped to dramatically low values for eight-ring PAs and cannot compete with their N-F analogs. This is probably because when the formamido group is substituted by the bulky TMG at the N-terminus while not having β inserts on the top strand, molecules like FH1026 and KJK6162 are not flexible enough to accommodate either the repulsive electronic interactions or the steric hindrance between the two termini. As a result, the free molecules tend to be less stacked, making it harder to form hairpin and conform ideally to the minor groove of cognate DNA upon binding. Interestingly, unlike KA2040 or KA2115, the affinity of the double β-substituted TMG compound FH1024 (K_D = 1.65 nM) is between the two single-β-substituted FH1028 (K_D = 0.16 nM) and FH1026 (K_D = 35 nM). The much higher dissociation constants of FH1024 than KA2040 plays an important role in lowering its binding affinity, indicating that the N-TMG can affect optimal β-β stacking and that the extra flexibility does not always optimally orient the PA upon binding to DNA, while the loss of hydrophobic interactions can overshadow the effort to enhancing hydrogen bonding strength.

2.5.2 Effect of N-terminal cationic group on PA selectivity of DNA flanking sequences

The selective properties of the cationic TMG group over the immediate flanking sequence of DNA binding site can be utilized as an important tool for drug design. The binding of the β-insert group molecules to SC1 sequence are in good agreement with their binding to λB sequence, meaning that β inserts have similar effects on the binding to both DNA sequences. However, the selective impacts of TMG overshadowed the β-insert effects in the cationic group, making it a determinant factor upon binding to SC1 sequence. For GAGA sequence, β modification did not

significantly increase the affinity among formamido group PAs, yet a single β replacement at position 3 (KA2034) still showed stronger binding than the parent PA, KA2035. Especially, both KA2115 and KA2040 showed weaker binding to GAGA than to the other two sequences, while the other molecules in the group have comparable affinities to all three sequences. The N-TMG FH1024, on the other hand, is around 12 times more favorable to GAGA sequence than λB sequence. The difference in binding strength of those molecules caused by alteration in flanking sequences, together with the selectivity of the TMG group, indicates that the DNA context can affect some PA-DNA binding, thus increasing the selectivity of those PAs. This provides a way to target one specific transcription factor among a transcription factor family, for which all proteins have consensus binding site with different flanking sequences for the different specific targets. A larger DNA pool is required for further tests of this selectivity.

2.6 Conclusions

Taken together, a binding pattern has been developed that links the modification with binding affinity compared to the parent PA, KA2035: 1) For single-β-containing PA, β/Py increases binding affinity when targeting T·A; Py/β decreases binding affinity when targeting T·A. Im/β decreases binding affinity when targeting G·C. 2) For double-β-containing PA, Imβ/βPy increases binding affinity when targeting GT·CA. β/β can increase or decrease binding affinity when targeting T·A. 3) When compared to the TMG-PA KJK6162, any β modification can increase the binding affinity (in the above, a slash "/" separates pairs of PA building blocks from opposite strands of the hairpin).

Overall, when Py between two Im is replaced by β, the binding affinity tends to go up, while if β appears on a lower strand consisting of four Py, the binding affinity tends to go down. Fortunately, a combination of the two positions can bring the affinity back to even higher levels.

Sugiyama *et al* has shown that Im heterocycles are more planar than Py rings, which makes an Im-derived molecule more rigid than a corresponding Py-polyamide, e.g. when binding to DNA (35). This could partially account for the slow association rate of the unmodified parent PA, KA2035. However, when we substitute the Py in the middle (number 3) with β, the whole molecule, especially the upper strand, becomes more flexible and more able to adjust to the curvature of DNA's minor groove upon binding. The finding that replacing Py between two imidazoles with β increases affinity is consistent with Turner *et al.* results, where PA10 (ImβImPyPyPy-γ-ImPyPyPyβPyβ-Dp) had ca. five times stronger binding affinity to its cognate DNA than the non β-inserted PA9 (ImPyImPyPyPy-γ-ImPyPyPyPyPyβ-Dp), and PA12 (ImβImPy-γ-ImβImPyβ-Dp), with two ImXIm (X = β, Py) groups involved, is 100 times stronger than its unsubstituted analog PA11 (ImPyImPy-γ-ImPyImPyβ-Dp) in binding to its cognate DNA (17). Similarly, Dickinson *et al*. also showed an increase in affinity when substituting Py in PA1 (ImPyImPy-γ-PyPyPyPy-β-Dp) with β in PA2 (ImβImPy-D-PyPyPyPy-β-Dp, where D denotes diaminobutyric acid) (36). Other studies also showed increased binding affinities of ImβIm to 5'-GCGC-3' sequence than ImPyIm (37, 38).

As for the bottom strand, the curvature of four Py fits nicely with that of DNA with optimal hydrogen bonding. Replacing Py with β on the other hand would reduce the van der Waals interaction between PA and DNA without allowing the PA to form better hydrogen bonds. This is also shown in Bashkin's docking model (19). As a result, the overall binding is weaker than unmodified PA. Interestingly, when two of the above modifications are combined, the positive binding effects of the upper strand can reverse the negative binding effects of the lower strand, especially when Py/Py is replaced by β/β.

The N-terminal cationic group can have a positive effect on accelerating the association process of PA and DNA. However, because of the bulky nature of TMG group, steric clash and sometimes repulsive electrostatic interactions are generated upon binding to DNA, resulting in impaired DNA binding affinity of PAs. Thus, the TMG group has to be carefully combined with β-alanine inserts to achieve optimal PA-DNA binding strength.

2.7 Acknowledgements

The authors thank NIH GM111749 to W.D.W. and the UMSL College of Arts and Sciences and Office of Research Administration to J.K.B. for financial support. The UMSL Agilent 600 MHz NMR spectrometer was obtained using funds from the NSF (0959360). J.K.B. thanks Prof. B. J. Bythell (UMSL) and the Danforth Plant Sciences Center (NSF DBI 0922879) for HRMS.

2.8 References

1 M. L. Kopka, C. Yoon, D. Goodsell, P. Pjura and R. E. Dickerson. The molecular origin of DNA-drug specificity in netropsin and distamycin. *Proc. Natl. Acad. Sci. U. S. A.*, **1985**, *82*, 1376-1380.

2 D. E. Wemmer. Designed sequence-specific minor groove ligands. *Annu. Rev. Biophys. Biomol.*, **2000**, *29*, 439-461.

3 P. B. Dervan. Molecular recognition of DNA by small molecules. *Bioorg. Med. Chem.*, **2001**, *9*, 2215-2235.

4 S. Neidle. DNA minor-groove recognition by small molecules. *Nat. Prod. Rep.*, **2001**, *18*, 291-309.

5 R. Moretti, L. J. Donato, M. L. Brezinski, R. L. Stafford, H. Hoff, J. S. Thorson, P. B. Dervan and A. Z. Ansari. Targeted chemical wedges reveal the role of allosteric DNA modulation in protein-DNA assembly. *ACS Chem. Biol.*, **2008**, *3*, 220-229.

6 T. G. Edwards, K. J. Koeller, U. Slomczynska, K. Fok, M. Helmus, J. K. Bashkin and C. Fisher. HPV episome levels are potently decreased by pyrrole-imidazole polyamides. *Antiviral Res.*, **2011**, *91*, 177-186.

7 K. J. Koeller, G. D. Harris, K. Aston, G. He, C. H. Castaneda, M. A. Thornton, T. G. Edwards, S. Wang, R. Nanjunda, W. D. Wilson, C. Fisher and J. K. Bashkin. DNA binding polyamides and the importance of DNA recognition in their use as gene-specific and antiviral agents. *Med. Chem.*, **2014**, *4*, 338-344.

8 J. G. Pelton and D. E. Wemmer. Structural characterization of a 2:1 distamycin A. d(CGCAAATTGGC) complex by two-dimensional NMR. *Proc. Natl. Aca. Sci. U. S. A.*, **1989**, *86*, 5723-5727.

9 M. Coll, C. A. Frederick, A. H. Wang and A. Rich. A bifurcated hydrogen-bonded conformation in the d(A.T) base pairs of the DNA dodecamer d(CGCAAATTTGCG) and its complex with distamycin. *Proc. Natl. Aca. Sci. U. S. A.*, **1987**, *84*, 8385-8389.

10 M. Mrksich, M. E. Parks and P. B. Dervan. 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.

11 C. R. Woods, T. Ishii, B. Wu, K. W. Bair and D. L. Boger. Hairpin versus extended DNA binding of a substituted β-alanine linked polyamide. *J. Am. Chem. Soc.*, **2002**, *124*, 2148-2152.

12 M. E. Parks, E. E. Baird and P. B. Dervan. Optimization of the hairpin polyamide design for recognition of the minor groove of DNA. *J. Am. Chem. Soc.*, **1996**, *118*, 6147-6152.

13 J. J. Kelly, E. E. Baird and P. B. Dervan. Binding site size limit of the 2:1 pyrrole-imidazole polyamide-DNA motif. *Proc. Natl. Aca. Sci. U. S. A.*, **1996**, *93*, 6981-6985.

14 C. L. Kielkopf, E. E. Baird P. B. Dervan and D. C. Rees. Structural basis for G.C recognition in the DNA minor groove. *Nat. Struct. Biol.*, **1998**, *5,* 104-109.

15 W. Zhang, T. Bando and H. Sugiyama. Discrimination of hairpin polyamides with an α-substituted-γ-aminobutyric acid as a 5'-TG-3' reader in DNA minor groove. *J. Am. Chem. Soc.*, **2006**, *128,* 8766-8776.

16 J. W. Trauger, E. E. Baird, M. Mrksich and P. B. Dervan. Extension of sequence-specific recognition in the minor groove of DNA by pyrrole-imidazole polyamides to 9-13 base pairs. *J. Am. Chem. Soc.*, **1996**, *118,* 6160-6166.

17 J. M. Turner, S. E. Swalley, E. E. Baird and P. B. Dervan. Aliphatic/aromatic amino acid parings for polyamide recognition in the minor groove of DNA. *J. Am. Chem. Soc.*, **1998**, *120,* 6219-6226.

18 C. C. Wang, U. Ellervik and P. B. Dervan. Expanding the recognition of the minor groove of DNA by incorporation of beta-alanine in hairpin polyamides. *Bioorg. Med. Chem.*, **2001**, *9,* 653-657.

19 J. K. Bashkin, K. Aston, J. P. Ramos, K. J. Koeller, R. Nanjunda, G. He, C. M. Dupureur and W. D. Wilson. Promoter scanning of the human COX-2 gene with 8-ring polyamides: unexpected weakening of polyamide-DNA binding and selectivity by replacing an internal N-Me-pyrrole with β-alanine. *Biochimie.*, **2013**, *95,* 271-279.

20 S. Wang, K. Aston, K. J. Koeller, G. D. Harris Jr., N. P. Rath, J. K. Bashkin and W. D. Wilson. Modulating of DNA-polyamide interaction by β-alanine substitutions: a study of positional effects on binding affinity, kinetics and thermodynamics. *Org. Biomol. Chem.*, **2014**, *12,* 7523-7536.

21 S. Wang, R. Nanjunda, K. Aston, J. K. Bashkin and W. D. Wilson. Correlation of local effects of DNA sequence and position of β-alanine inserts with polyamide-DNA complex binding affinities and kinetics. *Biochemistry*, **2012**, *51*, 9796-9806.

22 A. E. Hargrove, J. A. Raskatov, J. L. Meier, D. C. Montgomery and P. B. Dervan. Characterization and solubilisation of pyrrole-imidazole polyamide aggregates. *J. Med. Chem.*, **2012**, *55*, 5425-5432.

23 B. Liu and T. Kodadek. Investigation of the relative cellular permeability of DNA-binding pyrrole-imidazole polyamides. *J. Med. Chem.*, **2009**, *52*, 4604-4612.

24 G. F. He and J. K. Bashkin. What is the antiviral potential of pyrrole-imidazole polyamides? *Future Med. Chem.*, **2015**, *7*, 1953-1955.

25 J. K. Bashkin, T. G. Edwards, C. Fisher, G. D. Harris and K. J. Koeller. Guanidinyl-substituted polyamides useful for treating human papilloma virus. *US Patent* 9,133,228, **2015**.

26 C. Penaranda, W. Kuswanto, J. Hofmann, R. Kenefeck, P. Narendran, L. S. Walker, J. A. Bluestone, A. K. Abbas and H. Dooms. IL-7 receptor blockade reverses autoimmune diabetes by promoting inhibition of effector/memory T cells. *Proc. Natl. Aca. Sci. U. S. A.*, **2012**, *109*, 12668-12673.

27 F. Rosenbauer, K. Wagner, J. L. Kutok, H. Iwasaki, M. M. Le Beau, Y. Okuno, K. Akashi, S. Fiering and D. G. Tenen. Acute myeloid leukemia induced by graded reduction of a lineage-specific transcription factor, PU.1. *Nat. Genet.*, **2004**, *36*, 624-630.

28 E. E. Baird and P. B. Dervan. Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids. *J. Am. Chem. Soc.*, **1996**, *118*, 6141-6146.

29 M. del Fresno, A. El-Faham, L. A. Carpino, M. Royo, F. Albericio. Substituted guanidines: introducing diversity in combinatorial chemistry. *Org. Lett.*, **2000**, *2*, 3539-3542.

30 C. H. Castanedaa, M. J. Scuderia, T. G. Edwardsa, G. D. Harris Jr.a, C. M. Dupureura, K. J. Koellera, C. Fisherb, and J. K. Bashkin. Improved antiviral activity of a polyamide against high-risk human papillomavirus *via* N-terminal guanidinium substitution. *MedChemComm.*, **2016**, *7*, 2076-2082.

31 R. Karlsson. Affinity analysis of non-steady-state data obtained under mass transport limited conditions using BIAcore technology. *J. Mol. Rcognit.*, **1999**, *12*, 285-292.

32 T. A. Morton and D. G. Myszka. Kinetic analysis of macromolecular interactions using surface plasmon resonance biosensors. *Methods Enzymol.*, **1998**, *295*, 268-294.

33 R. Lyng, A. Roger and B. Norden. The CD of ligand-DNA systems. 2. Poly(dA-dT) B-DNA. *Biopolymers*, **1992**, *32*, 1201-1214.

34 M. Munde, S. Wang, A. Kumar, C. E. Stephens, A. A. Farahat, D. W. Boykin, W. D. Wilson and G. M. Poon. Structure-dependent inhibition of the ETS-family transcription factor PU.1 by novel heterocyclic diamidines. *Nucleic Acids Res.*, **2014**, *42*, 1379-1390.

35 Y. W. Han, T. Matsumoto, H. Yokota, G. Kashiwazaki, H. Morinaga, K. Hashiya, T. Bando, Y. Harada and H. Sugiyama. Binding of hairpin pyrrole and imidazole polyamides to DNA: relationship between torsion angle and association rate constants. *Nucleic Acids Res.*, **2012**, *40*, 11510-11517.

36 L. A. Dickinson, J. W. Trauger, E. E. Baird, P. Ghazal, P. B. Dervan and J. M. Gottesfeld. Anti-repression of RNA polymerase II transcription by pyrrole-imidazole polyamides. *Biochemistry*, **1999**, *38*, 10801-10807.

37 Y. W. Han, G. Kashiwazaki, H. Morinaga, T. Matsumoto, K. Hashiya, T. Bando, Y. Harada and H. Sugiyama. Effect of single pyrrole replacement with β-alanine on DNA binding affinity and sequence specificity of hairpin pyrrole/imidazole polyamides targeting 5'-GCGC-3'. *Bioorg. Med. Chem.*, **2013**, *21,* 5436-5441.

38 J. L. Meier, A. S. Yu, I. Korf, D. J. Segal and P. B. Dervan. Guiding the design of synthetic DNA-binding molecules with massively parallel sequencing. *J. Am. Chem. Soc.*, **2012**, *134,* 17814-17822.

2.9 Tables and Figures

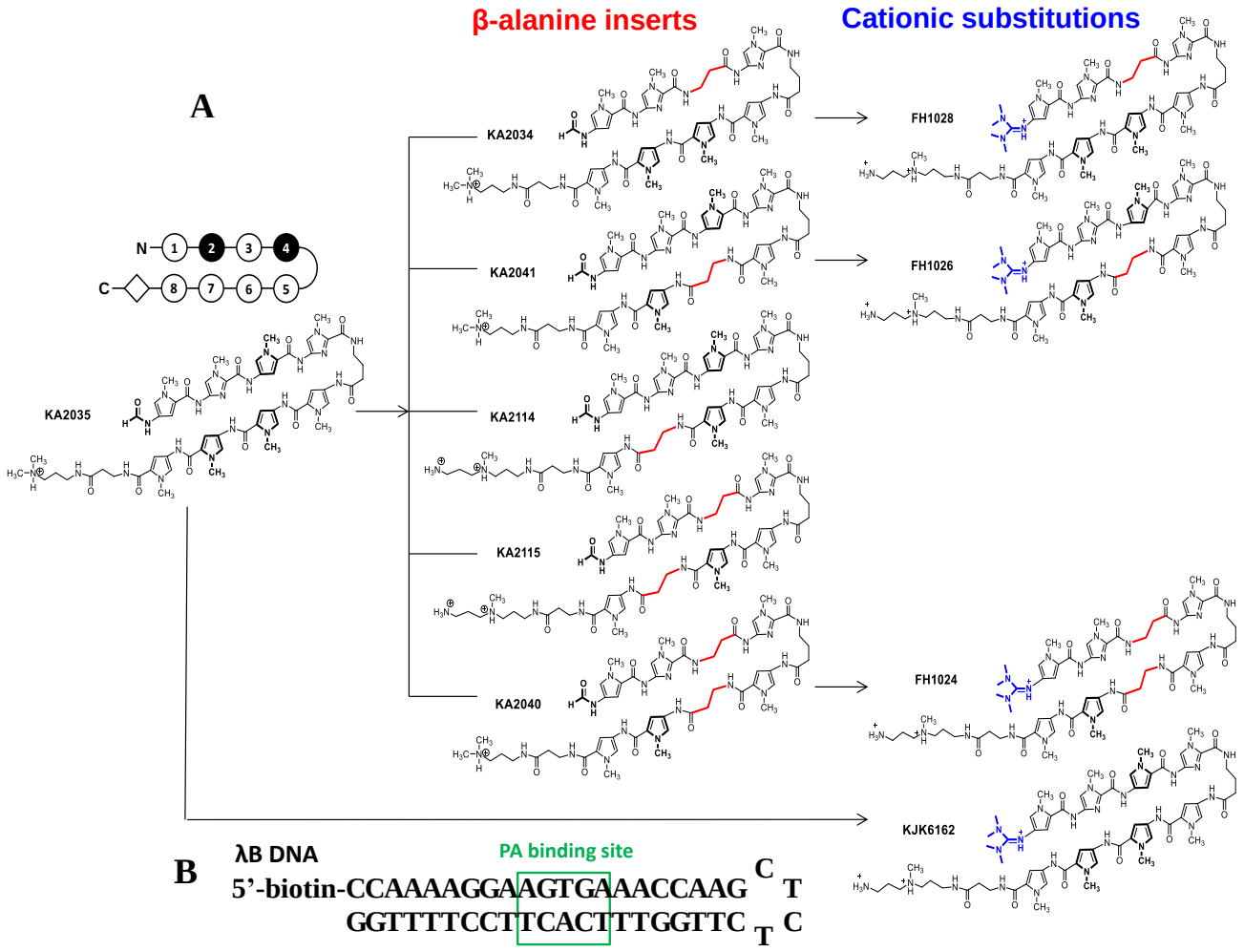


Figure 2.1 Systematically designed polyamides and DNA sequence.

A) Systematically designed polyamides that are categorized into two modified groups: β-alanine inserts (red) and cationic substitutions (blue). The illustration above KA2035 is the simplified representative of KA2035. The open and closed circles stand for pyrrole and imidazole, respectively. The diamond represents β-alanine. The numbering starting from N-terminus to C-terminus applies to all PAs. B) The biotinylated cognate binding sequence of the PAs: λB DNA. The predetermined PA binding site is highlighted in the green frame.

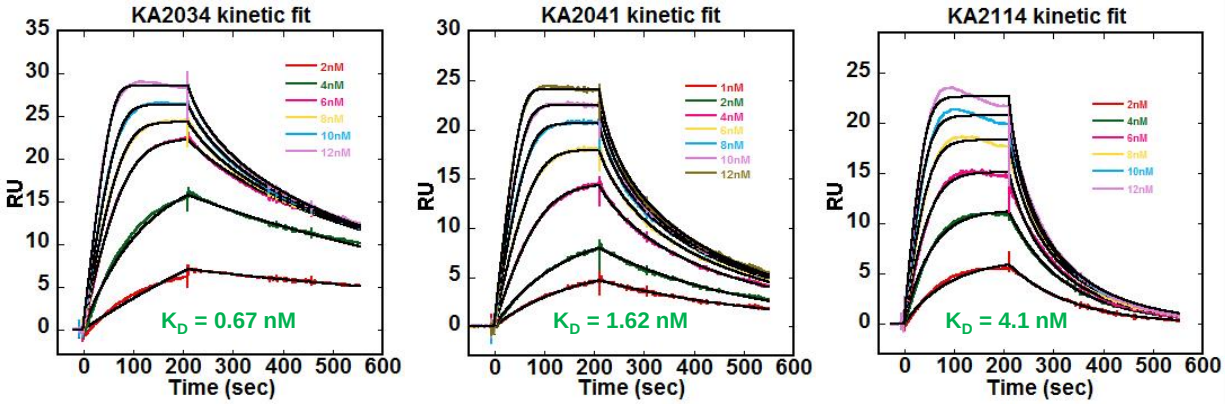


Figure 2.2 Representative SPR sensorgrams of PAs binding to λB DNA.

From left to right, there are strong (K_D = 0.67 nM), intermediate (K_D = 1.62 nM), and weak (K_D = 4.1 nM), binders. The colored lines are experimental sensorgrams. The black overlays are 1:1 global kinetic fits.

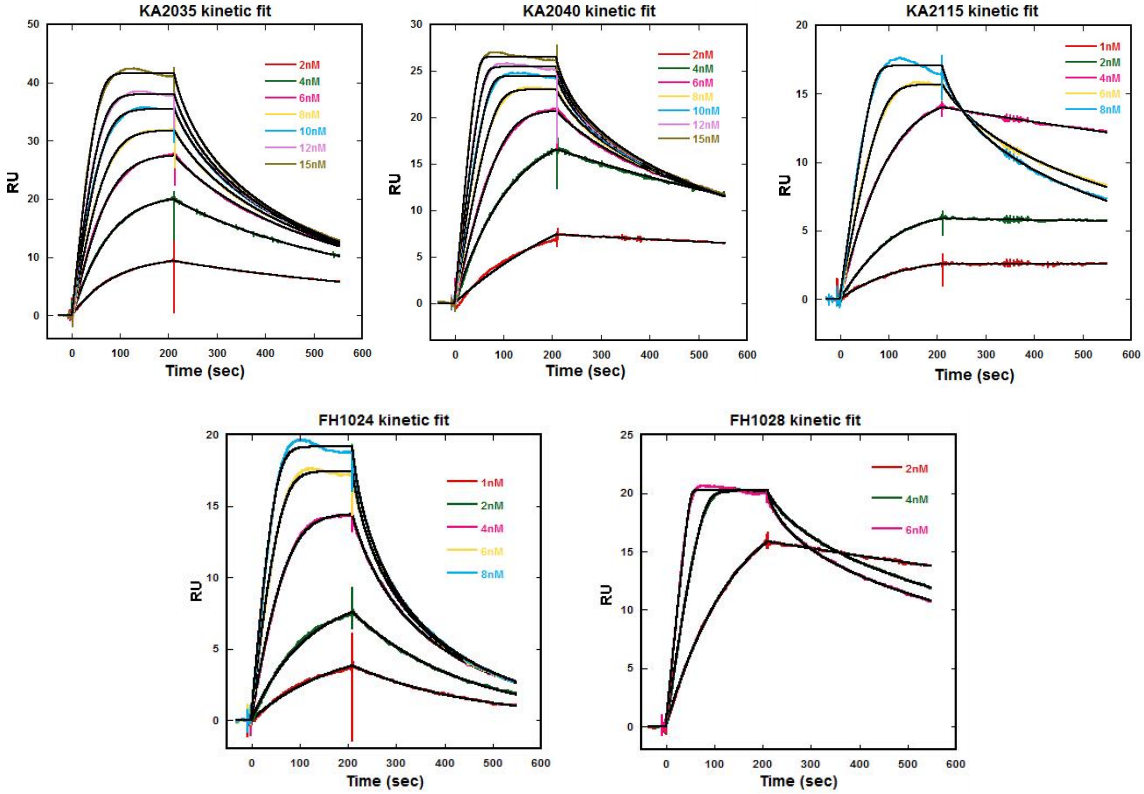


Figure 2.3 SPR sensorgrams of PAs binding to λB DNA.

Table 2.1 Kinetic rate constants and equilibrium binding constants for all PAs.

PAs in the β-inserts group are ranked according to their binding affinities from strong to weak. FH1026 and KJK6162 have very fast binding kinetics that are beyond the limitation of instrument detection. Thus no binding kinetics is reported here (ND). The equilibrium binding affinities were measured using steady state fitting. Note that KA2035 is included into β-inserts group for ease of comparison. Diamond, open circle and close circle represent β-alanine, pyrrole and imidazole respectively; Red diamond denotes β-alanine that replaced Py; F means formamido group and TMG means tetramethylguanidinyl group. Dp is short for 3-(dimethylamino)propylamine and Ta is short for 3,3'-diamino-N-methyldipropylamine.

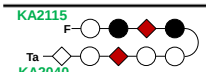
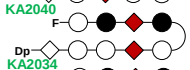
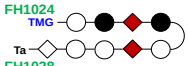
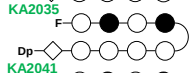
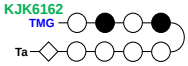
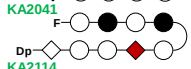
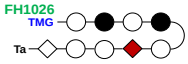
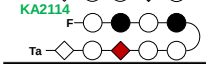
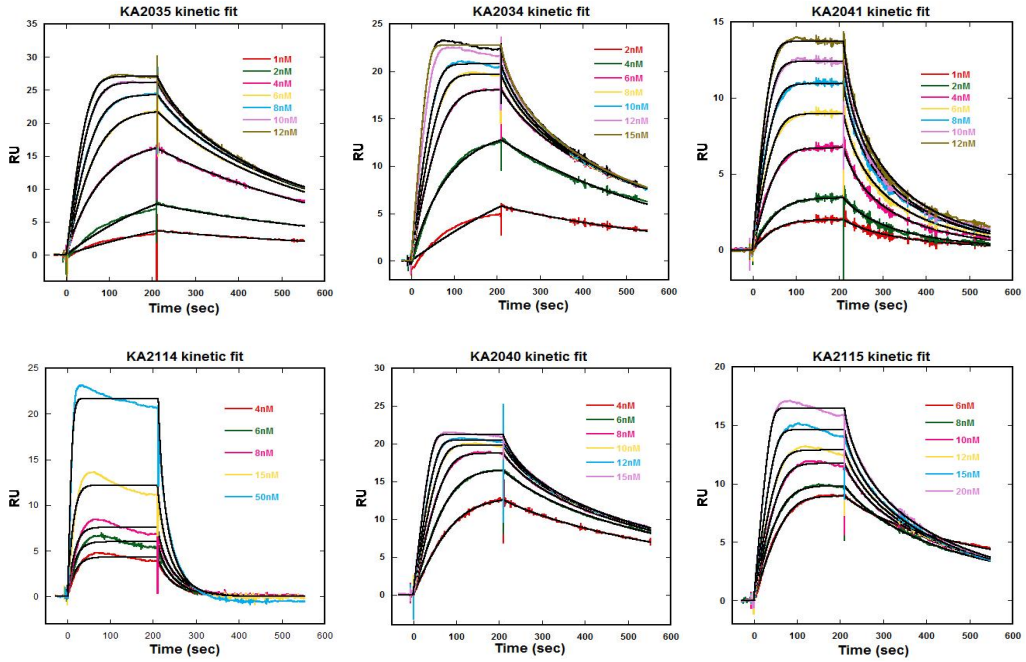
β-alanine inserts	k_a (x 10^7 M^{-1} s^{-1})	k_d (x 10^{-3} s^{-1})	K_D (nM)	Cationic substitutions	k_a (x 10^7 M^{-1} s^{-1})	k_d (x 10^{-3} s^{-1})	K_D (nM)
KA2115 	1.7 ± 0.1	9.9 ± 0.1	0.49 ± 0.01				
KA2040 	1.5 ± 0.1	8.4 ± 0.1	0.54 ± 0.01	FH1024 	3.2 ± 0.1	52.7 ± 0.6	1.65 ± 0.03
KA2034 	1.0 ± 0.1	6.9 ± 0.1	0.67 ± 0.01	FH1028 	5.0 ± 0.1	7.9 ± 0.1	0.16 ± 0.02
KA2035 	0.58 ± 0.03	7.6 ± 0.1	1.30 ± 0.01	KJK6162 	ND	ND	79 ± 4
KA2041 	2.1 ± 0.1	34.1 ± 0.5	1.62 ± 0.04	FH1026 	ND	ND	35 ± 2
KA2114 	3.9 ± 0.1	161 ± 4	4.1 ± 0.1				

Table 2.2 ΔT_m values of PAs with cognate and five mutant sequences.

The error of the ΔT_m values are within 5%, based on experimental reproducibility.

DNA / PA	AGTGA	Mutant 1 AGAGA	Mutant 2 AGTGT	Mutant 3 AGTTA	Mutant 4 ACTGA	Mutant 5 AGTCA
Tm (° C) of free DNA	68.3	66.0	68.4	63.7	68.8	69.4
KA2115	14.3	14.9	13.6	10.5	5.9	4.4
KA2040	10.1	9.0	11.6	4.5	3.9	2.4
KA2034	9.8	9.2	9.4	6.8	6.5	6.0
KA2035	9.7	9.5	9.4	4.9	3.9	2.6
KA2041	9.7	10.5	9.7	7.9	5.0	4.9
KA2114	7.7	7.2	7.7	7.5	3.0	3.6
FH1028	18.6	18.5	19.1	14.8	12.9	12.6
FH1024	9.6	8.5	11.7	7.0	4.4	5.4
FH1026	7.7	6.7	8.0	4.0	2.6	4.2
KJK6162	5.7	6.3	4.6	3.9	3.6	3.6

Slightly stronger | Equal | Slightly weaker | Weaker | Much weaker



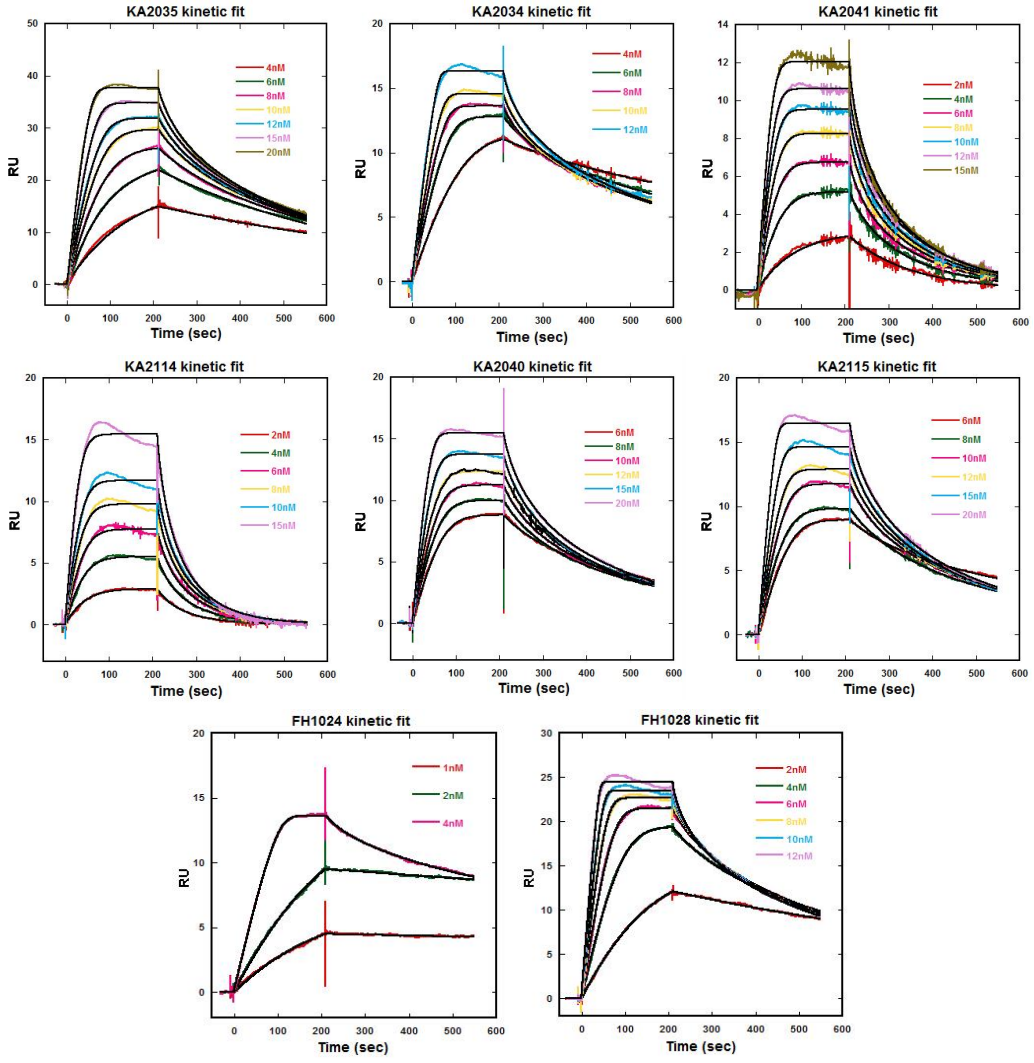
SC1 DNA

PA binding site

5'-biotin-CGGCCAAGCCGGAAGTGAGTGC C T

GCCGGTTCGGCCTTCACTCACG T C

Figure 2.4 SPR sensorgrams of PAs binding to SC1 sequence.



GAGA mutant PA binding site

5'-biotin-CCAAAAGAGAGTGAAACCAAG C T
GGTTTTCTCTCACTTTGGTTC T C

Figure 2.5 SPR sensorgrams of PAs binding to GAGA mutant sequence.

Table 2.3 Equilibrium binding affinities of all PAs with DNAs that have mutated flanking sequences.

No means no binding was detected between PA and the corresponding DNA. Errors for these equilibrium constant values are standard errors for the fitting and are within 2% for kinetic fitting, within 6% for steady state fitting.

	Polyamide	λB (K_D/nM) AAGTGAA	SC1 (K_D/nM) AAGTGAG	GAGA (K_D/nM) GAGTGAA
β-alanine inserts	KA2115 F Ta	0.49	0.27	1.67
	KA2040 F Dp	0.54	0.68	1.82
	KA2034 F Dp	0.67	0.83	0.66
	KA2035 F Dp	1.30	0.91	1.36
	KA2041 F Dp	1.62	2.70	2.80
	KA2114 F Ta	4.10	7.70	7.20
Cationic substitution	FH1028 TMG Ta	0.16	18	0.51
	FH1024 TMG Ta	1.65	No	0.14
	FH1026 TMG Ta	35	No	43
	KJK6162 TMG Ta	79	277	75

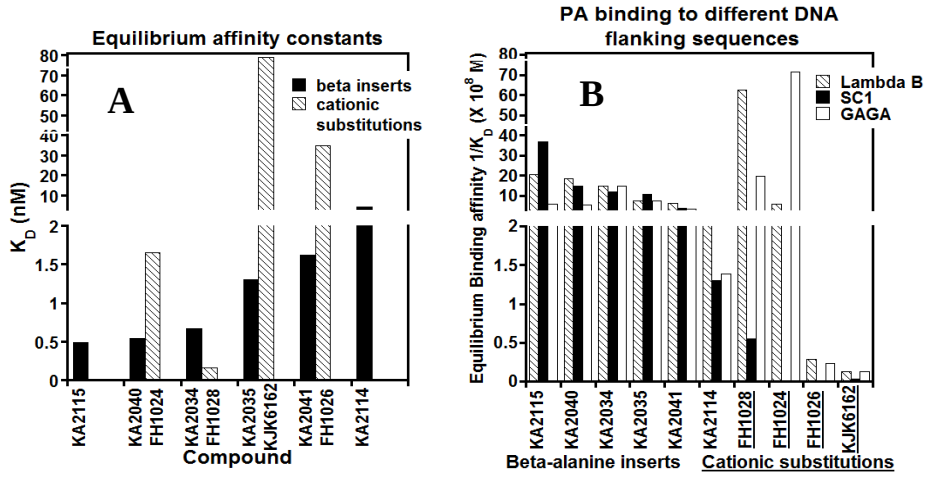


Figure 2.6 A) Direct comparison of equilibrium affinities of all PAs binding to λB sequence. B) Binding affinities of all PAs to λB, SC1 and GAGA mutant sequences.

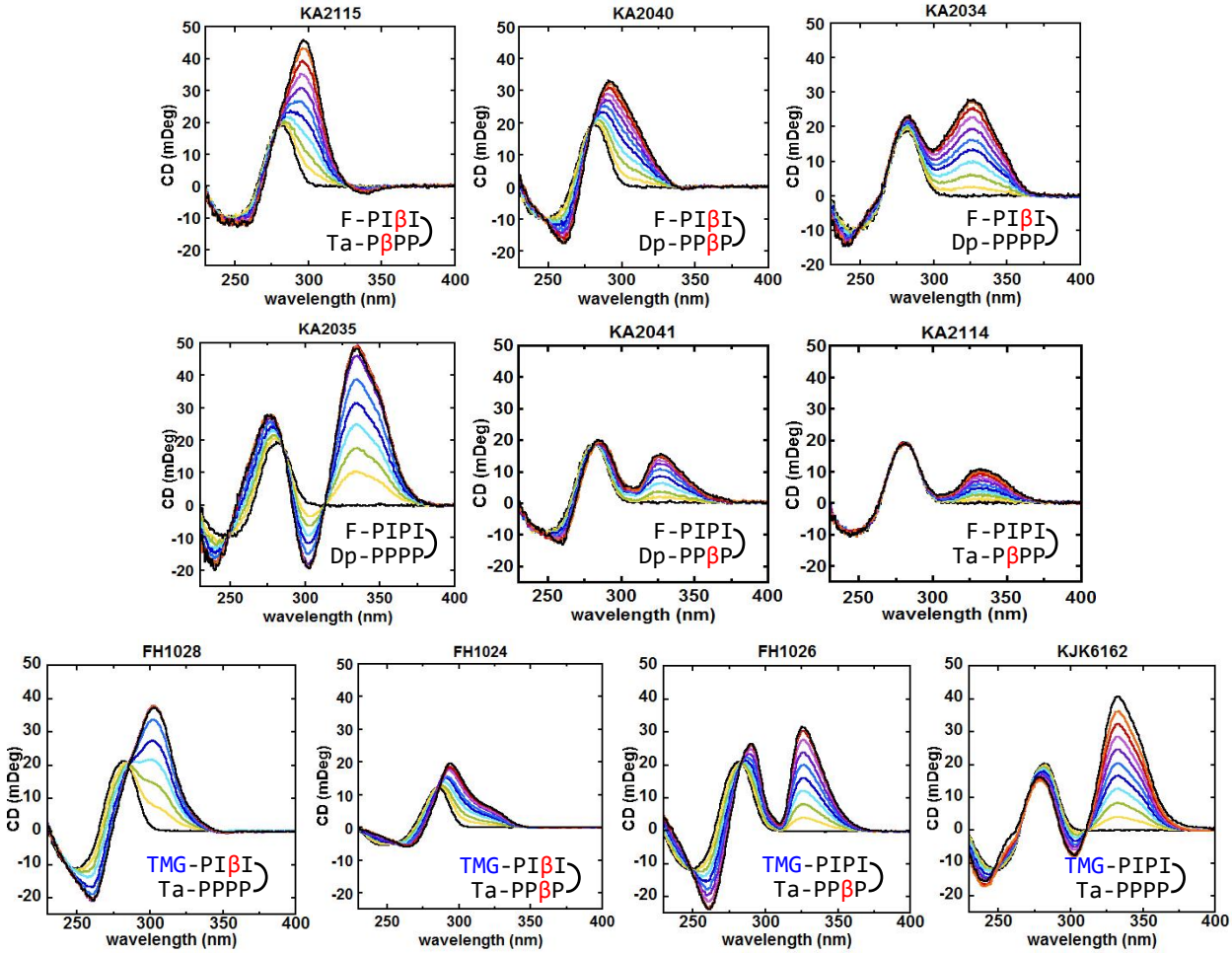


Figure 2.7 CD spectra of PA titration with short λB sequence.

3 DNA-BINDING PROPERTIES OF NEW FLUORESCENT AZAHX-AMIDES: METHOXY-PYRIDYL-AZA-BENZIMIDAZOLE-PYRROLE-IMIDAZOLE/PYRROLE

Beibei Liu,[a] Luke Pett, [b] Konstantinos Kiakos, [b] Pravin C. Patil, [c] Vijay Satam,[c] John A. Hartley, [b] Moses Lee, *[a,c,d] W. David Wilson *[a]

[a] Department of Chemistry, Georgia State University, Atlanta, GA 30303, USA

[b] Cancer Research UK Drug-DNA interactions Research Group,

UCL Cancer Institute, London, WC1E 6BT, UK

[c] Department of Chemistry, Hope College, Holland, MI 49423, USA

[d] M. J. Murdock Charitable Trust, 703 Broadway Street, Suite 710, Vancouver, WA 98660, USA

*To whom correspondence should be addressed:

*[a] E-mail: wdw@gsu.edu; Fax: +1 404-413-5505; Tel: +1 404-413-5503

*[a,c,d] E-mail: mosesleenf@gmail.com

Liu B, *et al*. *Chembiochem.*, 2018, **19,** 1979-1987.

My contribution to this paper was designing and conducting the thermal melting, SPR, CD experiments and running Spartan calculation, analyzing the results and writing the manuscript.

3.1 Abstract

DNA minor groove binding polyamides have been extensively developed to control abnormal gene expression. Establishment of the novel, inherently fluorescent Hx-amides has provided an alternative path for studying DNA binding in cells by direct observation of cell localization. Because of the 2:1 antiparallel stacking homodimer binding mode of these molecules to DNA, modification of Hx-amides to AzaHx-amides has successfully extended the DNA recognition repertoire, from central CG (recognized by Hx-I) to central GC (recognized by AzaHx-P) recognition. To potentially target two consecutive GG bases, new modifications from AzaHx moiety to 3-Pyr-AzaHx and 2-Pyr-AzaHx moieties were developed. The newly designed molecules are also small-sized, fluorescent amides with Pyr-AzaHx connected to two conventional five-membered heterocycles. Complementary biophysical methods were carried out to investigate DNA binding properties of these molecules. The results showed that neither 3-Pyr-AzaHx nor 2-Pyr-AzaHx was able to mimic I-I to specifically target GG dinucleotides. Rather, 3-Pyr-AzaHx functions like AzaHx or f-I or P-I as a antiparallel stacked dimer. 3-Pyr-AzaHx-PI (**2**) binds 5'-ACGCGT'-3' with improved binding affinity and high sequence specificity when compared to its parent molecule AzaHx-PI (**1**). However, 2-Pyr-AzaHx is detrimental to DNA binding because of an unfavorable steric clash upon stacking in the minor groove.

3.2 Introduction

Polyamides are a set of well-known DNA-minor groove targeting small molecules that consist of *N*-methylpyrrole (P) and/or *N*-methylimidazole (I) heterocycles (1). Modeled from naturally occurring DNA binding agents, Netropsin and Distamycin A (2, 3), these molecules can form 2:1 antiparallel side-by-side stacked dimers upon binding to the DNA minor groove (1, 4). These molecules interact with DNA by forming specific hydrogen bonds and van der Waals

contacts with different bases. Such polyamides can be programmed to target any DNA sequence based on the recognition rules developed by Dervan and coworkers (5, 6). The recent development of polyamides focuses largely on the synthesis of large hairpin molecules or tandem polyamides with increased specificity for longer sequences (7, 8) or modified polyamides conjugates that can serve as DNA probes and enable polyamide with various functions (9, 10). These polyamides exhibit the ability to manipulate gene expression, thus, advancing the role of polyamides in gene therapy.

However, challenges still exist for efficient cell uptake. This calls for the development of small-size, soluble, and stable polyamides that are more easily synthesized. To improve cell uptake, solubility, stability, and more importantly, the feasibility for nuclear localization testing, we previously developed a novel class of inherently fluorescent small-size diamino polyamides by introducing 2-(*p*-anisyl)-benzimidazole (Hx) moiety to function as two consecutive pyrroles, 'P-P' or formamido-pyrrole, 'f-P' (11, 12). These molecules are hybrids of Hx with two conventional 5-membered heterocycles (P or I). The new Hx-amides can form a 2:1 side-by-side, antiparallel, stacked homodimer in the DNA minor groove and retain strong DNA binding affinity and specificity. In addition, the intrinsically fluorescent properties confer these molecules with the ease of cell study. In a recent study, an extra positive charge was added to Hx-IP at either P or I. These molecules have demonstrated to not only have higher binding affinity but also improved cell uptake in NIH3T3 mouse fibroblast cells and A549 cancer cells (13).

To expand the DNA recognition repertoire of Hx-amides and enhance sequence specificity, we recently reported the design and synthesis of a second generation *p*-anisyl-AzaHx-amides (AzaHx-amides) (14). Derived from Hoechst 33258, the AzaHx-amides are structurally similar to Hx-amides but contain 2-(*p*-anisyl)-4-azabenzimidazole (AzaHx) moiety at the *N*-terminus. The

novel AzaHx moiety mimicked P-I or f-I in DNA recognition and consequently, enabled G/C base pair recognition of these molecules. Because of the side-by-side, antiparallel stacking manner of these molecules, introduction of the AzaHx moiety extended DNA site recognition from central CG (recognized by Hx-I) to central GC (recognized by AzaHx-P, Fig. 5), even CGCG recognition. Recognized selectively by AzaHx-PI (**1**), the 5'-ACGCGT-3' sequence is of great importance, because it resides in the control region of the human Dbf4 gene promoter: the Mlu1 Cell Cycle Box (MCB) sequence. Expression of this gene is closely related to the development of several cancers (15-19).

After successfully establishing that the fluorescent AzaHx moiety functions as f-I or P-I, we designed new molecules by introducing a pyridyl (Pyr) moiety in place of the *p*-anisyl moiety of *p*-anisyl-AzaHx-amides to further extend the recognition capability of AzaHx polyamides. The newly introduced pyridyl moiety can potentially function as a H-bond acceptor and target G as showed by previous literature (20-23). Consequently, we hypothesized that the resulting Pyr-AzaHx could act as I-I and recognize two consecutive GG base alignments along a DNA sequence. It was thought that in a 2:1 antiparallel stacked manner, the new molecules would target a GGCC site.

To test our hypothesis, three different Pyr-AzaHx-amides, namely, 3-Pyr-AzaHx-PI (**2**), 3-Pyr-AzaHx-PP (**3**), and 2-Pyr-AzaHx-PP (**4**) (Fig. 3.1) were designed and synthesized. Excitation of a 1 μM aqueous solution of compound **2** at 310 nm (λ_{max} in uv spectrum) produced a broad emission from 340 to 440 nm that maximized at about 380 nm. The position of the nitrogen in the pyridine ring was adjusted to identify the optimal position responsible for stronger binding. Their predicted cognate sequences are 5'-ACGCGT-3' and/or 5'-AGGCCT-3' for compound **2,** 5'-ATGCAT-3' and/or 5'-AGGCCT-3' for compound **3,** 5'-AGGCCT-3' and/or 5'-ATGCAT-3' for

compound **4** and the binding schemes are presented in Fig. 3.1. The predicted binding site maintains central GC base alignment but the immediate flanking bases on both ends are GC rich. In this study, an extensive series of biophysical methods including DNase I footprinting, thermal melting (Tm), surface plasmon resonance (SPR), circular dichroism (CD), and Spartan calculations was employed to investigate the DNA-binding properties of these novel molecules. The results have provided us with new guidance for development of next stage molecules.

3.3 Materials and Methods

3.3.1 DNA preparation

DNA oligomers were purchased from Eurofins Genomics (for footprinting studies) and from Integrated DNA Technologies, Inc. (IDT, Coralville, IA, USA) with HPLC purification and mass spectrometry characterization (for biophysical studies). The oligomers are provided below.

The DNAs used in Tm and CD are short hairpin oligonucleotides, the predicted binding sites are highlighted in red and the hairpin is written in italic and underlined.

5’-GCAACGCGTAC*TCTC*GTACGCGTTGC-3’ (abbreviated as 5’-ACGCGT-3’)

5’-GCAATGCATAC*TCCC*GTATGCATTGC-3’ (abbreviated as 5’-ATGCAT-3’)

5’-GCAAGGCCTAC*TTTT*GTAGGCCTTGC-3’ (abbreviated as 5’-AGGCCT-3’)

The DNAs used in SPR are hairpin oligonucleotides with biotin attached to the 5’ end, the predicted binding sites are highlighted in red and the hairpin is written in italic and underlined.

5’-biotin-CCGACGCGTCGG*CTCT*CCGACGCGTCGG-3’ (abbreviated as 5’-ACGCG T-3’)

5’-biotin-CGATGCATGC*CTCT*GCATGCATCG-3’ (abbreviated as 5’-ATGCAT-3’)

5’-biotin-CGAGGCCTG*CTCT*CAGGCCTCG-3’ (abbreviated as 5’-AGGCCT-3’)

DNA footprinting Each DNA fragment contains 131 base pairs with a [^{32}P]-radiolabeled 5’ end. Highlighted in red are the predicted cognate and non-cognate sites.

DNA fragment 1 (duplex)

5’-ATGCTCCAGAAAGCCGGCACTCAGTCTACAAACGCGTCATCTTGATCACC
GGTGTTCACAGAAATTTC
3’-TACGAGGTCTTTCGGCCGTGAGTCAGATGTTTGCGCAGTAGAACTAGTGG
CCACAAGTGTCTTTAAAG

TCTAGATCTACACGTAACTCTAGTAGCGCTCTTCAAGCAAGTGGAGCTCTCCT
AACCGACTTT-3’
AGATCTAGATGTGCATTGAGATCATCGCGAGAAGTTCGTTCACCTCGAGAGG
ATTGGCTGAAA-5’*

DNA fragment 2 (duplex)

5’-ATGCTCCAGAAAGCCGGCACTCAGTCTACAAACGCGTCATCTTGATCATG
CATGTTCACAGAAATTTC
3’-TACGAGGTCTTTCGGCCGTGAGTCAGATGTTTGCGCAGTAGAACTAGTAC
GTACAAGTGTCTTTAAAG

TCTAGATCTACTGGTAACTCTAGTATCGATCTTCAAGCAAGTGGAGCTCTCCT
AACCGACTTT-3’
AGATCTAGATGACCATTGAGATCATAGCTAGAAGTTCGTTCACCTCGAGAGG
ATTGGCTGAAA-5’*

DNA fragment 3 (duplex)

5’-ATGCTCCAGAAAGCCGGCACTCAGTCTACAAACGCGTCATCTTGATCAGG
CCTGTTCACAGAAATTTC
3’-TACGAGGTCTTTCGGCCGTGAGTCAGATGTTTGCGCAGTAGAACTAGTCC
GGACAAGTGTCTTTAAAG

TCTAGATCTAGTCCTAACTCTAGTATCGATCTTCAAGCAAGTGGAGCTCTCCT
AACCGACTTT-3’
AGATCTAGATCAGGATTGAGATCATAGCTAGAAGTTCGTTCACCTCGAGAGG
ATTGGCTGAAA-5’*

3.3.2 Compound synthesis and preparation

3-Pyr-AzaHx-PI, (2) 4-Methoxy-3-pyrido-4-aza-benzimidazole-pyrrole-imidazole

Nitro-PI polyamide (24) (60 mg, 0.16 mmol) was hydrogenated over 10% Pd/C in methanol (20 mL) at room temperature for 16 h. Removal of the catalyst and solvent gave an amine intermediate as a pale yellow solid that was co-evaporated with dichloromethane twice (5 mL). 3-

Pyr-Aza-Hx-acid (25) (52 mg, 0.19 mmol) was dissolved in dry DMF (1.5 mL) and the flask was purged with nitrogen. To the clear solution was added EDCI.HCl (110 mg, 0.6 mmol) followed by dry triethylamine (0.11 mL, 0.8 mmol) and 1-hydroxybenzotriazole (HOBt) (87 mg, 0.6 mmol) and stirred for 10 min at room temperature. A solution of the above amine in dry DMF (2.0 mL) was added to the reaction mixture at room temperature and the flask was again purged with nitrogen. The reaction mixture was stirred at room temperature for 16 h. The solvent was removed under pressure (1 mm Hg at 50 °C) and the residue was purified by column chromatography using silica gel and 7% methanol in chloroform as the eluent system**.** Concentration of the desired fractions gave AzaHx-amide **2** as a yellow solid (20 mg, 17%), mp = >230 °C, Rf = 0.50 (1.5:0.2:8.3 methanol: ammonium hydroxide:chloroform); FT-IR (KBr) □□□3269, 2954, 2825, 1704, 1641, 1530, 1464, 1400, 1363, 1279, 1116, 1099, 1020, 963, 880, 792, 742, 662; ^{1}H-NMR: (CD_3OD) □ 9.04 (s, 1H), 8.50 (d, *J* = 8.0 Hz, 1H), 7.94 (s, 2H), 7.45 (d, *J* = 8.0 Hz, 2H), 7.12 (s, 1H), 6.93 (d, *J* = 8.0 Hz, 1H), 4.02 (s, 3H), 3.99 (s, 3H), 3.97 (s, 3H), 3.49 (t, *J* = 7.0 Hz, 2H), 2.57 (t, 2H, *J* = 7.0 Hz), 2.32 (s, 6H); LRMS (ES TOF^{+}) *m/z* 294 ($M+2H^{2+}$, 100%), 586 ($M+H^{+}$, 90%); HRMS ($M+H^{+}$) calcd for $C_{28}H_{32}N_{11}O_4$ m/z 586.2638, found 586.2642.

3-Pyr-AzaHx-PP, (3) 4-Methoxy-3-pyrido-4-aza-benzimidazole-pyrrole-pyrrole

The procedure for synthesizing AzaHx-amide **3** was essentially the same as the preparation of AzaHx-amide **2** except nitro-PP polyamide (24) (60 mg, 0.16 mmol) and 3-Pyr-Aza-Hx-acid (25) (52 mg, 0.19 mmol) were used. Purification of the crude product by column chromatography using silica gel and 6% methanol in chloroform as the eluent system gave the desired product **3** as a pale yellow solid (26 mg, 26%), mp = 240-244 °C, Rf = 0.49 (1.5:0.2:8.3 methanol: ammonium hydroxide:chloroform); FT-IR (KBr) □□□3160, 3045, 2927, 2887, 1620, 1586, 1460, 1397, 1278, 1021, 812, 763, 660; ^{1}H-NMR: (CD_3OD) □ 8.98 (d, *J* = 4.0 Hz, 1H), 8.43 (dd, *J* = 4.0 Hz, 8.0 Hz,

1H), 8.02 (d, *J* = 8.0 Hz, 1H), 7.99 (d, *J* = 8.0 Hz, 1H), 7.39 (d, *J* = 2.0 Hz, 1H), 7.20 (d, *J* = 2.0 Hz, 1H), 7.03 (d, *J* = 2.0 Hz, 1H), 6.96 (□, *J* = 8.0 Hz, 1H), 6.81 (d, *J* = 2.0 Hz, 1H), 4.00 (s, 3H), 3.94 (s, 3H), 3.88 (s, 3H), 3.34 (t, *J* = 7.0 Hz, 2H), 2.43 (t, *J* = 7.0 Hz, 2H), 2.29 (s, 6H), 1.79 (quint, *J* = 7.0 Hz, 2H); LRMS (EI) *m/z* 300 ($M+2H^{2+}$, 100%), 599 ($M+H^+$, 100%); HRMS ($M+H^+$) calcd for $C_{30}H_{35}N_{10}O_4$ m/z 599.2842, found 599.2851.

2-Pyr-AzaHx-PP, (4) 3-Methoxy-2-pyrido-4-aza-benzimidazole-pyrrole-pyrrole

Synthesis of AzaHx-amide **4** followed the same procedure as the preparation of AzaHx-amide **2** except nitro-PP polyamide (24) (60 mg, 0.16 mmol), 2-Pyr-Aza-Hx-acid (25) (48 mg, 0.18 mmol), and a 6% methanol in chloroform solvent system was used to purify product **4**, That was isolated as a pale yellow solid (32 mg, 35%), mp: 170-174 °C (decomposition), Rf = 0.22 [MeOH: $CHCl_3$ (80:20)]; FT-IR (KBr) □□□3305, 2920, 1738, 1717, 1680, 1655, 1635, 1576, 1538, 1463, 1436, 1373, 1318, 1258, 1193, 1130, 1096, 1010, 888, 768; ^{1}H-NMR: (DMSO-D_6 + 3 drops of D_2O) □ 9.92 (s, 1H), 8.08 (s, 1H), 7.99 (d, *J* = 8.0 Hz, 2H), 7.74 (d, *J* = 8.0 Hz, 2H), 7.40 (s, 1H), 7.21 (s, 2H), 6.84 (s, 1H), 4.00 (s, 3H), 3.88 (s, 3H), 3.81 (s, 3H), 3.16 (t, *J* = 8.0 Hz, 2H), 2.24 (t, *J* = 8.0 Hz, 2H), 2.14 (s, 6H), 1.61 (quint, *J* = 6.5 Hz, 2H); LRMS (TOF ES^+) *m/z* 300 ($M+2H^{2+}$, 20%); 599 ($M+H^+$, 22%); HRMS ($M+H^+$) calcd for $C_{30}H_{35}N_{10}O_4$ m/z 599.2842, found 599.2849.

Preparation of the compounds For the biophysical studies, each compound was initially dissolved in methanol, then three equivalent amount of HCl was added to the solution and the mixture was stirred for 10 min. The mixture was evaporated to dryness using argon gas before the residue was dissolved in double deionized water.

The extinction coefficient of AzaHx-amides 3-Pyr-AzaHx-PI (**2**), 3-Pyr-AzaHx-PP (**3**) and 2-Pyr-AzaHx-PP (**4**) were determined by UV-vis spectrometer at wavelengths 310 nm, 310 nm

and 316 nm, respectively. The determined extinction coefficient values are 22990 L/mole^{-1}·cm^{-1}, 19853 L/mole^{-1}·cm^{-1} and 10117 L/mole^{-1}·cm^{-1}, respectively. The buffer used for the determination contains 10 mM sodium phosphate, 1 mM EDTA, pH 6.2.

For DNase I footprinting studies, each AzaHx-amide was dissolved in DMSO to give a 10 mM stock concentration and stored at -20°C, and diluted to appropriate concentrations just before use.

3.3.3 Biophysical experimental procedures

Thermal melting (Tm) Thermal melting was carried out on a Cary 300 Bio UV/visible spectrophotometer (Varian) at a concentration of 6 µM ligand and 3 µM DNA. The samples were warmed up from 25 °C to 95 °C. All experiments were performed in 10-mm path length quartz cells. Tm values were determined at the reflection points of the melting curves. The ΔTm values are the difference between melting temperature of DNA-ligand complex and free DNA. The buffer used for Tm contained 10 mM sodium phosphate, 1 mM EDTA, at pH 6.2.

Surface Plasmon Resonance (SPR) The binding study was conducted on a four-channel Biacore T200 surface plasmon resonance instrument. Three biotinylated DNAs (5'-biotin-CCGACGCGTCGG*CTCT*CCGACGCGTCGG-3', 5'-biotin-CGATGCATGC*CTCT*GCATGCATCG-3', 5'-biotin-CGAGGCCTG*CTCT*CAGGCCTCG-3') are immobilized on a streptavidin functionalized sensorchip through the last three channels, respectively. The first channel has no DNA immobilized for baseline correction. The DNA immobilization procedure is the same as previously described. (26) The first 5 cycles are pure buffer injection serving as double baseline correction to remove any instrumental errors. Molecules are injected over the chip surface at an increasing gradient of concentration for 180 s each cycle, followed by a pure running buffer injection to dissociate the complex. The running buffer consists of 10 mM cacodylic acid, 100 mM

NaCl, 1mM ethylenediaminetetraacetic acid (EDTA) and 0.05% v/v surfactant Polysorbate 20 (P20), pH 6.2. 1 M NaCl was used as regeneration buffer at the end of each cycle to fully dissociate the molecule and prepare the surface for the next cycle. The binding kinetics were determined by 1:1 global kinetic fitting with mass transfer correction incorporated. In some cases, the kinetics were so fast that they went beyond the limit of detection. Affinities were then obtained through steady state fitting for those that had reached equilibrium during molecule injection.

Circular Dichroism (CD) Circular Dichroism was performed on a Jasco J-810 spectrometer (Jasco Inc., Easton, MD) at ambient temperature with a scan range of 400-220 nm. A DNA concentration of 5 μM was firstly scanned followed by a titration of AzaHx-amides **2** and **3** in increments of 0.5 equivalents. The spectra were averaged over four scans with a scan speed of 50 $nm \cdot min^{-1}$ and a buffer blank correction. All experiments were conducted in phosphate A $PO_4$0 buffer (10 mM sodium phosphate, 1 mM EDTA, pH 6.2)

Molecule Visualization Spartan´16 software was used to calculate the compounds. Energy profile method at ground state with density functional B3LYP/6-31G* in vacuum was applied for calculation. The molecules were then manually aligned for observation.

DNase I footprinting The DNase I footprinting reactions were conducted in a total volume of 8 μL; the labelled DNA substrate (2 μL, 200 counts s^{-1}) was incubated for 1 h at room temperature in 4 μL TN buffer (10 mM Tris-HCl, 10 mM NaCl, pH 7.0) containing the required polyamide concentration. Controlled digestion with RQ1 RNase-Free DNase I (Promega) was then initiated by the addition of 2 μL of DNase I solution (20 mM NaCl, 2 mM MgCl2 , 2 mM MnCl2 , DNase I 0.02U, pH 8.0), mixed gently and allowed to proceed for 3 min at room temperature and terminated by snap freezing the samples on dry ice. The nuclease-digested samples were subsequently lyophilized to dryness and resuspended in 10 μL of formamide loading dye (95%

formamide, 20 mM EDTA, 0.05% bromophenol blue, and 0.05% xylene cyanol). A purine specific DNA marker was prepared to allow identification of the guanine and adenine residues in the sequence. 3 μL of radiolabelled DNA substrate was heated at 90°C for 20 min in the presence of 20 μL TN buffer and 5 μL of formamide loading dye. Alternatively, 10 μL of radiolabelled DNA substrate was incubated at room temperature for 5 min with 50 μL formic acid (Sigma-Aldrich), and then frozen on dry ice and lyophilized to dryness. 65 μL of piperidine (1 M; Sigma-Aldrich) was then added and the sample was heated at 90°C for 30 min before the reaction was terminated by snap freezing on dry ice and lyophilised to dryness. The DNA marker was then washed with distilled water, lyophilised and resuspended in 10 μL of formamide loading dye.

The samples were denatured by heating at 90°C for 4 min, and then transferred on to ice before loading on a 10% denaturing polyacrylamide gel containing urea (UreaGel, National Diagnostics, UK). Electrophoresis was carried out for 2-4 h at 1650 V (~70W, 55°C) in 1x Trisborate EDTA buffer (TBE; Severn Biotech). The gel was then transferred onto Whatman 3MM paper and dried under vacuum at 80°C for 2 h using a GD 2000 gel dryer (Hoefer). The gel was exposed overnight to Fuji medical X-Ray film and developed using a Konica Medical Film Processor SRX-101A to visualise the radioactive signal.

3.4 Results

3.4.1 Determination of relative DNA binding strength of Pyr-AzaHx polyamides using thermal melting (Tm)

To test the binding of these newly designed molecules to their predicted DNA sites, thermal melting experiments were carried out. ΔTm, the difference of melting temperature of the bound DNA complexes and melting temperature of DNA alone, reflects the extent to which DNA is stabilized by the molecules, and thus the relative binding strength. The ΔTm results of three

molecules and their predicted cognate DNAs are listed in Table 3.1. It is clear that 3-Pyr-AzaHx-PI (**2**) binds to both 5'-ACGCGT-3' and 5'-ATGCAT-3' sequences quite strongly (15.9 °C and 15.5 °C, respectively), but does not have high affinity for 5'-AGGCCT-3' (8.8 °C). With 3-Pyr-AzaHx-PP (**3**), strong binding to 5'-ATGCAT-3' (14.5 °C) and weak binding to 5'-AGGCCT-3' (7.8 °C) and 5'-ACGCGT-3' (7.4 °C) were observed. These results suggest that the 3-Pyr-AzaHx moiety functions similarly to *p*-anisyl-AzaHx or f-I or PI and that the pyridine group does not make specific H-bond contacts with G. Thus, the observed target sequences of 3-Pyr-AzaHx-PI (**2**) and 3-Pyr-AzaHx-PP (**3**) are ACGCGT and ATGCAT, respectively.

We then modified the 3-Pyr-AzaHx by relocating the nitrogen to *meta*-position (Fig. 3.1). The resulting 2-Pyr-AzaHx-PP (**4**) was tested against the three DNAs for its role in stabilization. Surprisingly, 2-Pyr-AzaHx-PP (**4**) showed hardly any affinity for all three DNAs.

3.4.2 Determination of DNA selectivity of Pyr-AzaHx polyamides by DNase I footprinting

DNase I footprinting was conducted to further investigate and compare the behavior of these molecules on a larger DNA scale. The binding affinity and selectivity of 3-Pyr-AzaHx-PI (**2**) was investigated using an engineered DNA fragment 1 that consists of the cognate sequence (based on Tm results): 5'-ACGCGT-3', as well as four mutant/non-cognate sites: 5'-ACCGGT-3', 5'-AAATTT-3', 5'-ACACGT-3', and 5'-AGCGCT-3' (Fig. 3.3A). The footprinting gel of 3-Pyr-AzaHx-PI (**2**) showed a clear footprint beginning at 0.5 μM for the cognate sequence (Fig. 3.3A), exhibiting a similar binding pattern to AzaHx-PI (**1**) (Fig. 3.5A) (14). Only a very weak and incomplete single mismatch protection site is shown at the sequence, 5'-ACACGT-3', indicating a weak tolerance of 3-Pyr-azaHx as well as azaHx (Fig. 3.5A) for a "A" over its preferred "G". No other significant protection sites were evident at concentrations of up to 20 μM with this DNA

sequence. The Tm and footprinting data together indicate that the 3-Pyr-AzaHx-PI (**2**) is very selective for the central GC dinucleotides recognition, with a selection factor of more than 40 fold according to the footprinting studies.

The DNA binding property of 3-Pyr-AzaHx-PP (**3**) was evaluated using engineered DNA fragment 2 (Fig. 3.3B). The molecule is predicted to target 5'-WWGCWW-3' (W = A or T). Two cognate sequences: 5'-ATGCAT-3' and 5'-AAGCAA-3', along with four non-cognate sites: 5'-ACGCGT-3', 5'-AAATTT-3', 5'-ACTGGT-3', and 5'-ATCGAT-3' were present in DNA fragment 2. A clear footprint was visible at both consensus sequences starting at 3 μM. However, evident footprint was also detected at the non-cognate site containing flipped central GC pair: 5'-ATCGAT-3' at 3 μM. The equally strong protection of this non-cognate site and the featured central base pair indicates that the molecule might bind to the sequence in the reversed orientation, that is, *N*-terminus of the molecule is aligned with 3' end of DNA. What also appeared on the gel was another footprint at 5'-ACTGGT-3' at 10 μM (Fig. 3.3B). The footprinting result suggests that 3-Pyr-AzaHx-PP (**3**) not only binds weakly to its cognate 5'-ATGCAT-3' sequence compared to 3-Pyr-AzaHx-PI (**2**), but also has poor sequence selectivity, as evidenced by the off-target binding sites.

With the lack of binding according to the results from Tm studies, further investigation on 2-Pyr-AzaHx-PP (**4**) was conducted using a broad spectrum of DNA sequences to determine whether it resembles AzaHx-PP, recognizing 5'-WWGCWW-3', or IIPP, recognizing 5'-WGGCCW-3'. DNase I footprinting of 2-Pyr-AzaHx-PP (**4**) binding to DNA fragment 2 that contains two 5'-WWGCWW-3' sites (5'-ATGCAT-3' and 5'-AAGCAA-3') was carried out. No footprint was detected up to 25 μM (Fig. 3.4A). On another trial using the engineered fragment 3 containing 5'-WGGCCW-3' (5'-AGGCCT-3'), the results showed no binding up to 50 μM (Fig.

3.4B). These results agree with those collected from Tm studies, indicating that a change in the nitrogen position along with the attached methoxy unit at the 3-position of the pyridyl group at the *N*-terminus abrogated the binding of Pyr-AzaHx moiety, thus, making it incompetent structural element of design to target any DNA site.

3.4.3 Determination of binding kinetics and affinity of Pyr-AzaHx polyamides by SPR

SPR provides us with real-time information to quantitatively determine the interaction between molecules and DNAs. After confirming the corrected cognate DNA for each molecule and to assist evaluation of binding affinity and kinetics compared to previous work, SPR experiments were carried out with three predicted cognate DNA sequences that are immobilized on a sensorchip. The resulting SPR sensorgrams are shown in Fig. 3.6 and the kinetic and equilibrium binding constants are listed in Table 3.2. The strongest binding was observed between 3-Pyr-AzaHx-PI (**2**) and its cognate DNA, 5'-ACGCGT-3' (K_D = 0.75 ± 0.01 nM), followed by its interaction with 5'-ATGCAT-3' (K_D = 3.2 nM). The result shows that 3-Pyr-AzaHx-PI (**2**) is very specific to the central GC base alignment, but not very selective for the immediate flanking bases, especially among C, A or T. Strong interactions between 3-Pyr-AzaHx-PP (**3**) and its cognate DNA, 5'-ATGCAT-3' were also confirmed by SPR with K_D equals 1.1 nM. These results are in good agreement with both Tm and footprinting data within error range, indicating that the molecules are well stacked and form strong hydrogen bonding with DNA base pairs. The strong interactions are also featured by very slow dissociation rate (~10^{-4}·s^{-1}, Table 3.2, Fig. 3.6). On the other hand, 3-Pyr-AzaHx-PP (**3**) was found to bind very weakly to 5'-ACGCGT-3' (K_D = 1035 ± 22nM), suggesting that a switch of the molecule structure from I to P is incompatible with G. Also consistent with the Tm result is the fact that 3-Pyr-AzaHx-PI (**2**) binds to the mismatched sequence 5'-AGGCCT-3', albeit with significantly reduced affinity compared to its cognate DNA. If 3-Pyr-

AzaHx interacts with GG to some extent, then 3-Pyr-AzaHx-PP (**3**) should target 5'-AGGCCT-3' according to the binding scheme depicted in Fig. 3.1. To our surprise, 3-Pyr-AzaHx-PP (**3**) exhibited around 10 times weaker affinity to 5'-AGGCCT-3', indicating that the 3-Pyr-AzaHx moiety is not the determinant factor in this binding event. In an attempt to specifically recognize two consecutive GG base pairs, 3-Pyr-AzaHx was modified to 2-Pyr-AzaHx moiety and the binding of 2-Pyr-AzaHx-PP (**4**) was tested on SPR against the three DNA sequences. In agreement with the Tm data, 2-Pyr-AzaHx-PP (**4**) binds very weakly to 5'-ACGCGT-3' and 5'-ATGCAT-3' and showed no binding at all to 5'-AGGCCT-3'. Clearly, the relative position of the pyridine nitrogen along with the methoxy group is pivotal for binding. A close observation of the sensorgram shows that some fitting curves are diverged from experimental data. Especially, the experimental deviations are above the global fit. This indicates that non-specific binding begins to occur at higher concentrations.

3.4.4 Confirmation of binding of Pyr-AzaHx polyamides in DNA minor groove by CD

The binding mode of Pyr-AzaHx polyamides to their cognate DNA was evaluated by CD. Molecules that bind in the minor groove typically have positively induced band (25, 27-29). The Pyr-AzaHx molecules alone do not generate any peak. Upon titration of 3-Pyr-AzaHx-PI (**2**) and 3-Pyr-AzaHx-PP (**3**) to their cognate DNA, strong positive signals are induced at 350 nm for both molecules (Fig. 3.7).This is typical for polyamides and Hx-amides and is indicative of DNA minor groove binding.

3.5 Discussion

Previous development of hairpin polyamides involved incorporation of benzimidazole (Bi) or imidazolepyridine (Ip) into the polyamide core structure and showed that the resulting molecules displayed strong binding affinity and sequence specificity (30-34). The need for more

soluble, smaller polyamides to improve cell uptake and biological functions requires new strategies to design the molecules. Our previously designed small-size *p*-anisyl-Hx-amides mimic f-P or P-P, form antiparallel stacked dimer, and follow pyrrole-imidazole DNA recognition principles (1, 4). Biological studies of Hx-amides, specifically, Hx-IP targeting the inverted repeat of CCAAT Box 2 on the topoisomerase IIα promoter to inhibit NF-Y binding, have shown increased cell uptake and enhanced subsequent biological activities in NIH3T3 and A549 cells compared to high molar mass polyamides, such as hairpins and H-pins (35-38). The improved cell permeability and DNA binding affinity from incorporating *N*-terminal *p*-anisyl-Hx moiety in small Hx-amides offers a promising way for designing small-size, soluble polyamides. However, the P-P like DNA binding property of *p*-anisyl-Hx is limited to A or T recognition. Especially, according to the central and terminal pairing rules (Fig. 3.8), the central two heterocycle pairs of a tri-amide or tetra-amide homodimer recognize the two central base pairs while the terminal formamido and imidazole/pyrrole (tri-amides) or the two "outer" heterocycles pairs (tetra-amides) specifically target the immediate flanking base pairs on either side of the central base pairs (39, 40). Therefore, by this recognition pattern, Hx-amides are restricted to central CG dinucleotide recognition and the subsequent target of biologically relevant sequences is constrained (Fig. 3.8).

The novel development of AzaHx-amides showed strong binding affinity and additional GC selectivity. The AzaHx moiety functions as an alternative to f-I or P-I capable of forming an antiparallel dimer. According to the central and terminal pairing rule, AzaHx-amides are able to confer central and outer GC dinucleotide recognition (Fig. 3.8), thus greatly extending the sequence it can target.

To further expand the recognition repertoire to GGCC, our new modification strategy is to replace the *p*-anisyl group in AzaHx-amdies with pyridyl groups to generate Pyr-AzaHx-amides.

The central GC dinucleotide is unchanged, as is the AzaHx moiety. The extra nitrogen of *N*-terminal pyridyl moiety could provide an additional hydrogen bond acceptor for G recognition. To test this hypothesis, the DNA binding properties of the modified Pyr-AzaHx-amides are investigated by a complementary set of biophysical and biochemical methods.

3.5.1 Tm results show that 3-Pyr-AzaHx functions similarly to AzaHx or P-I or f-I

Tm studies were performed to initially screen the Pyr-azaHx complexes for binding DNA. The prediction was that Pyr-AzaHx would act similarly to two consecutive imidazole units, or I-I, to selectively recognize GG dinucleotides. But the low ΔTm value of 3-Pyr-AzaHx-PP (**3**) with 5'-AGGCCT-3' and high value with 5'-ATGCAT-3' (Table. 3.1) suggested that 3-Pyr-AzaHx could not form extra hydrogen bonds with two consecutive GG dinucleotides to stabilize the molecule-DNA complex. Rather, the 3-Pyr-azaHx moiety behaved more akin to AzaHx or P-I or f-I, and it anchored itself in the minor groove of TG dinucleotides with strong hydrophobic contacts. Evidence of the high ΔTm value of 3-Pyr-AzaHx-PI (**2**) with 5'-ACGCGT-3' and low value with 5'-AGGCCT-3' further confirmed the similar functionality of 3-Pyr-AzaHx to AzaHx. It is highly possible that the more polar nitrogen atom of the pyridyl moiety, compared to the opposite methinyl group, is rotated away from the floor of minor groove or is not positioned properly for hydrogen bonding. Instead, similar to pyrrole binding units, the hydrophobic C-H units in the pyridyl group were directed more favorably toward the hydrophobic floor in the minor groove. Thus, the preferred cognate binding sites for 3-Pyr-AzaHx-PI (**2**) and 3-Pyr-AzaHx-PP (**3**) were 5'-ACGCGT-3' and 5'-ATGCAT-3', respectively. It was also noticeable that 3-Pyr-AzaHx-PI (**2**) bound with high affinity to the mismatch site 5'-ATGCAT-3', indicating that when G was not available, imidazole could tolerate an A/T base pair quite well, thus compromising its selectivity. This result agreed with our previous work that showed imidazole on imidazole stacking

in polyamides bound preferentially to T/G or G/T but it tolerated C/G and G/C base pairs (29, 41-43).

The extremely low ΔTm values of the modified 2-Pyr-AzaHx-PP (**4**) to any of the three sequences indicate that the position of the pyridine nitrogen along with the methoxy group can be quite important to DNA binding. In fact, to structurally visualize the relative position of the antiparallel stacked homodimer, Spartan´16 software was used for Molecular Mechanics calculation. Our models show that when two 2-Pyr-AzaHx-PP (**4**) molecules are anti-parallely stacked, the bulky 3-methoxy group attached to pyridine is sterically encumbered with the *N*-methyl group of the C-terminal pyrrole (Fig. 3.9). On the other hand, if the nitrogen of pyridine is attracted toward the minor groove floor, the 3-methoxy group appears to be too big and it also affects the stacking of two molecules. It is difficult for the minor groove to accommodate favorably. Therefore, either way the nitrogen is pointing, the 3-methoxy-2-pyridyl moiety would encounter steric clash upon binding to DNA as stacked dimer, resulting in thermodynamically unstable molecule-DNA complex.

3.5.2 Binding affinity and selectivity of 3-Pyr-AzaHx-amides

The sequence specificity of the new AzaHx-amides was investigated by DNase I footprinting. 3-Pyr-AzaHx-PI (**2**) showed a clear footprint toward 5'-ACGCGT-3' at 0.5 μM and more than 40-fold selectivity against the tested DNA fragment. This is very similar in terms of both affinity and selectivity, to its parent structure, AzaHx-PI (**1**), in which the AzaHx moiety contains *p*-anisyl instead of a pyridyl group. AzaHx-PI was previously reported to footprint at 5'-ACGCGT-3' using the same DNA fragment starting at 0.5 μM with at least 40-fold selectivity (Fig. 3.5A) (14). Since 3-Pyr-AzaHx is experimentally shown to function as AzaHx or P-I or f-I, two other AzaHx-amide analogs recognizing the same cognate DNA were also included into the

comparison (40, 44-46). On the same DNA fragment, f-IPI (44, 45), a N-formamido pyrrole/imidazole polyamide, generated a footprint at the cognate site, beginning at 0.05 μM. However, another clear footprint at 5'-AGCGCT-3' is produced at 1 μM, suggesting that the sequence selectivity of f-IPI is reduced even though it has relatively high binding affinity. On the other hand, PIPI (44), a non-formamido pyrrole/imidazole polyamide, footprinted starting at 1 μM but showed comparable sequence selectivity to 3-Pyr-AzaHx-PI (**2**).

With a slightly lower binding affinity than 3-Pyr-AzaHx-PI (**2**), 3-Pyr-AzaHx-PP (**3**) footprinted at 3 μM at its cognate site 5'-ATGCAT-3'. At the same concentration, another footprint appeared at 5'-ATCGAT-3'. Given the inverted central bases and the degenerate P/P recognition of A/T or T/A, it is likely that the molecule can bind in a reversed manner. Another footprint for 5'-ACTGGT-3' at 10 μM indicates that 3-Pyr-AzaHx-PP (**3**) is not very selective. When compared to its analogs that target the same cognate DNA, AzaHx-amide **3** showed higher binding affinity. AzaHx-PP, which also contains *p*-anisyl instead of a pyridyl group, generated a footprint at 10 μM at the cognate site, but one other footprint at 5'-ATCGAT-3' at 15 μM, indicating that AzaHx-PP could not differentiate the two sequences very well, either (Fig. 3.5B). For f-IPP, even though a clear footprint at the cognate sequence was evident at 3 μM, equally strong footprint also appeared at three non-cognate sequences (Fig. 3.5C). These results suggest f-IPP has low selectivity. Taken together, 3-Pyr-AzaHx-PP (**3**) exhibited stronger binding affinity with slightly compromised selectivity compared to other structurally similar compounds.

To more accurately compare the binding affinities, SPR was performed to obtain the binding kinetics and equilibrium binding affinities of the new molecules and three DNA sequences. The SPR results are in good agreement with Tm and footprinting data. Interestingly, the binding affinity of 3-Pyr-AzaHx-PI (**2**) (K_D = 0.75 nM) is much higher than its analogs: AzaHx-

PI (**1**) (K_D = 400 nM) (14), f-IPI (K_D = 19 nM) (39, 44), and PIPI (K_D = 7100 nM) (44) on binding to the cognate sequence, 5'-ACGCGT-3'. These results reveal that 3-Pyr-AzaHx-PI (**2**) has drastically improved affinity, at least 20 times over its previously reported analogs. In particular, 3-Pyr-AzaHx-PI (**2**) has much higher binding affinity than its closest analog and parent structure AzaHx-PI (**1**). When comparing 3-Pyr-AzaHx-PP (**3**) (K_D = 1.1 nM) to its analogs AzaHx-PP (K_D = 110 nM, data not published) and f-IPP (K_D = 120 nM) (39), 3-Pyr-AzaHx-PP (**3**) showed around 100 times higher binding affinity. These results together with the footprinting data suggest that the 3-Pyr-AzaHx moiety helps to improve binding affinity while maintains a reasonable level of selectivity.

Interestingly, in a separate study reported recently by us, adding an extra positive charge to Hx-amides showed that the resulting dicationic HxIP* and HxI*P have higher binding affinity but reduced selectivity toward their consensus site 5'-ATCGAT-3' (ΔTm = 32 °C and 30 °C, respectively) than their mono-charged analog, HxIP (ΔTm = 15 °C) (13). Therefore, we suspect that the high affinity of 3-Pyr-AzaHx-amides is due to protonation of the nitrogen on the pyridyl moiety, thus contributing to an extra positive charge for stronger attraction to the negatively charged DNA. In fact, previous literatures have also shown that increasing positive charges can facilitate DNA binding (47-49). To test this idea, the binding of 3-Pyr-AzaHx-PI (**2**) to 5'-ACGCGT-3' was performed by SPR at pH 9.0. The resulting affinity was lowered by about 16 times than that at pH 6.2 (Fig. 3.10), thus indicating that the increased binding affinity of 3-Pyr-AzaHx-PI (**2**) could be attributed to the protonation of the pyridine nitrogen. Altogether, these results suggest that modification of the *N*-terminal *p*-anisyl to *N*-terminal pyridiyl, or converting AzaHx to 3-Pyr-AzaHx not only retained comparable selectivity but also significantly improved binding affinity under biological conditions.

3.6 Conclusion

In this study, we have designed and synthesized a novel class of small-sized polyamides, aiming to expand the DNA recognition repertoire to GGCC sequence. Our complementary biochemical and biophysical studies have shown that although the newly designed Pyr-AzaHx amides do not specifically recognize GG dinucleotides, the 3-Pyr-AzaHx amides do possess improved functionality compared to previous analogs. Especially, the increased binding affinity under biological conditions could be due to protonation of the pyridine nitrogen. The results have greatly extended our knowledge on molecular interactions and thus compound modification and optimization. Meanwhile, challenges on designing GG dinucleotides targeting, fluorescent compounds with good cell uptake are still needed.

3.7 Acknowledgements

This work was supported by a grant from the National Science Foundation (CHE 0809162 to M.L and W.D.W), a program grant from Cancer Research UK (C2259/A16569 to J.A.H), and a Medical Research Council UK funded studentship (L.P). The biosensor- SPR work was supported by the National Institutes of Health for W.D.W (GM 111749).

3.8 References

1 M. Mrksich, W. S. Wade, T. J. Dwyer, B. H. Geierstanger, D. E. Wemmer, P. B. Dervan. Antiparallel side-by-side dimeric motif for sequence-specific recognition in the minor groove of DNA by the designed peptide 1-methylimidazole-2-carboxamide netropsin. *Proc. Natl. Acad. Sci. U. S. A.*, **1992**, *89*, 7586-7590.

2 A. C. Finlay, F. A. Hochstein, B. A. Sobin, F. X. Murphy. Netropsin, a new antibiotic produced by a Streptomyces. *J. Am. Chem. Soc.*, **1951**, *73*, 341-343.

3 F. Arcamone, S. Penco, P. Orezzi, V. Nicolella, A. Pirelli. Structure and synthesis of Distamycin A. *Nature,* **1964**, *203*, 1064-1065.

4 W.S. Wade, M. Mrksich, P. B. Dervan. Design of peptides that bind in the minor groove of DNA at 5'-(A,T)G(A,T)C(A,T)-3' sequences by a dimeric side-by-side motif. *J. Am. Chem. Soc.*, **1992**, *114*, 8783-8794.

5 P. B. Dervan. Molecular recognition of DNA by small molecules. *Bioorg. Med. Chem.*, **2001**, *9*, 2215-2235.

6 P. B. Dervan, B. S. Edelson. Recognition of the DNA minor groove by pyrrole-imidazole polyamides. *Curr. Opin. Struct. Biol.*, **2003**, *13*, 284-299.

7 G. He, E. Vasilieva, G. D. Harris Jr., K. J. Koeller, J. K. Bashkin, C. M. Dupureur. Binding studies of a large antiviral polyamide to a natural HPV sequence. *Biochimie,* **2014**, *102*, 83-91.

8 Y. Kawamoto, A. Sasaki, A. Chandran, K. Hashiya, S. Ide, T. Bando, K. Maeshima, H. Sugiyama. Targeting 24bp within telomere repeat sequences with tandem tetramer pyrrole-imidazole polyamide probes. *J. Am. Chem. Soc.*, **2016**, *138*, 14100-14107.

9 Y. Kawamoto, T. Bando, H. Sugiyama. Sequence-specific DNA binding pyrrole-imidazole polyamides and their applications. *Bioorg. Med. Chem.,* **2018**, *26*, 1393-1411**.**

10 G. S. Erwin, M. P. Grieshop, A. Ali, J. Qi, M. Lawlor, D. Kumar, I. Ahmad, A. McNally, N. Teider, K. Worringer, R. Sivasankaran, D. N. Syed, A. Eguchi, M. Ashraf, J. Jeffery, M. Xu, P. M. C. Park, H. Mukhtar, A. K. Srivastava, M. Faruq, J. E. Bradner, A. Z. Ansari. Synthetic transcription elongation factors license transcription across repressive chromatin. *Science,* **2017**, *358*, 1617-1622.

11 S. Chavda, Y. Liu, B. Babu, R. Davis, A. Sielaff, J. Ruprich, L. Westrate, C. Tronrud, A. Ferguson, A. Franks, S. Tzou, C. Adkins, T. Rice, H. Mackay, J. Kluza, S. A. Tahir, S. Lin, K. Kiakos, C. D. Bruce, W. D. Wilson, J. A. Harley, M. Lee. Hx, a novel fluorescent, minor groove and sequence specific recognition element: design, synthesis, and DNA binding properties of p-anisylbenzimidazole/pyrrole-containing polyamides. *Biochemistry,* **2011**, *50*, 3127-3136.

12 V. Satam, P. Patil, B. Babu, M. Gregory, M. Bowerman, M. Savagian, M. Lee, S. Tzou, K. Olson, Y. Liu, J. Ramos, W. D. Wilson, J. P. Bingham, K. Kiakos, J. A. Hartley, M. Lee. Hx-amides: DNA sequence recognition by the fluorescent Hx (p-anisylbenzimidazole)·pyrrole and Hx·imidazole pairs. *Bioorg. Med. Chem. Lett.*, **2013**, *23*, 1699-1702.

13 L. Pett, K. Kiakos, V. Satam, P. Patil, S. Laughlin-Toth, M. Gregory, M. Bowerman, K. Olson, M. Savagian, M. Lee, W. D. Wilson, D. Hochhauser, J. A. Hartley. Modulation of topoisomerase IIα expression and chemosensitivity through targeted inhibition of NF-Y:DNA binding by a diamino p-anisyl-benzimidazole (Hx) polyamide. *Biochim. Biophys. Acta*., **2017**, *1860*, 617-629.

14 V. Satam, B. Babu, P. Patil, K. A. Brien, K. Olson, M. Savagian, M. Lee, A. Mepham, L. B. Jobe, J. P. Bingham, L. Pett, S. Wang, M. Ferrara, C. D. Bruce, W. D. Wilson, M. Lee, J. A. Harley, K. Kiakos. AzaHx, a novel fluorescent, DNA minor groove and G·C recognition element: synthesis and DNA binding properties of a p-anisyl-4-aza-benzimidazole-pyrrole-imidazole (azaHx-PI) polyamide. *Bioorg. Med. Chem. Lett.*, **2015**, *25*, 3681-3685.

15 X. Wu, H. Lee. Human Dbf4/ASK promoter is activated through the Sp1 and Mlul cell-cycle box (MCB) transcription elements. *Oncogene,* **2002**, *21*, 7786-7796.

16 J. W. Knockleby, J. Romero, K. A. Kylie, H. Lee, *Curr. Top. Biochem. Res*., **2010**, *12*, 43.

17 D. H. Charych, M. Coyne, A. Yabannavar, J. Narberes, S. Chow, M. Wallroth, C. Shafer, A. O. Walter. Inhibition of Cdc7/Dbf4 kinase activity affects specific phosphorylation sites on MCM2 in cancer cells. *J. Cell. Biochem.*, **2008**, *104*, 1075-1086.

18 D. Bonte, C. Lindvall, H. Liu, K. Dykema, K. Furge, M. Weinreich. Cdc7-Dbf4 kinase overexpression in multiple cancers and tumor cell lines is correlated with p53 inacitvation. *Neoplasia*, **2008**, *10*, 920-931.

19 Y. J. Sheu, B. Stillman. The Cdc7-Dbf4 kinase promotes S phase by alleviating an inhibitory activity in Mcm4. *Nature*, **2010**, *463*, 113-117.

20 D. Renneberg, P. B. Dervan. Imidazolepyridine/Pyrrole and hydroxybenzimidazole/pyrrole pairs for DNA minor groove recognition. *J. Am. Chem. Soc,* **2003**, *125*, 5707-5716.

21 M. Munde, M. A. Ismail, R. Arafa, P. Peixoto, C. J. Collar, Y. Liu, L. Hu, M. H. David-Cordonnier, A. Lansiaux, C. Bailly, D. W. Boykin, W. D. Wilson. Design of DNA minor groove binding diamidines that recognize GC base pair sequences: a dimeric-hinge interaction motif. *J. Am. Chem. Soc.*, **2007**, *129*, 13732-13743.

22 M. I. Sánchez, O. Vázquez, J. Martínez-Costas, M. E. Vázquez, J. L. Mascarenas. Straightforward access to bisbenzamidine DNA binders and their use as versatile adaptors for DNA-promoted processes. *Chem. Sci.*, **2012**, *3*, 2383-2387.

23 A. Paul, R. Nanjunda, A. Kumar, S. Laughlin, R. Nhili, S. Depauw, S. S. Deuser, Y. Chai, A. S. Chaudhary, M. H. David-Cordonnier, D. W. Boykin, W. D. Wilson. Mixed up minor groove binders: convincing A·T specific compounds to recognize a G·C base pair. *Bioorg. Med. Chem.*, **2015**, *25*, 4927-4932.

24 V. S. Satam, P. C. Patil, B. Babu, K. Brien, M. Gregory, M. Bowerman, J. Sweers, A. Mepham, M. Lee. Synthesis of 2-(substituted)-3*H*-benzimidazole-5-carboxylic acids and 2-

(substituted)-3*H*-imidazo[4,5-b]pyridine-5-carboxylic acids: synthons for fluorescent Hx and aza-Hx amides. *Bulgarian Chem. Commun.*, **2016**, *48*, 725-730.

25 E. R. Lacy, N. M. Le, C. A. Price, M. Lee, W. D. Wilson. Influence of a terminal formamido group on the sequence recognition of DNA by polyamides. *J. Am. Chem. Soc*., **2002**, *124*, 2153-2163.

26 S. Wang, R. Nanjunda, K. Aston, J. K. Bashkin, W. D. Wilson. Correlation of local effects of DNA sequence and position of β-alanine inserts with polyamide-DNA complex binding affinities and kinetics. *Biochemistry*, **2012**, *51*, 9796-9806.

27 R. Lyng, A. Rodger, B. Norden. The CD of ligand-DNA systems. 2. Poly(dA-dT) B-DNA. *Biopolymers*, **1992**, *32*, 1201-1214.

28 X. L. Yang, C. Kaenzig, M. Lee, A. H. Wang. Binding of AR-144, a tri-imidazole DNA minor groove binder, to CCGG sequence analyzed by NMR spectroscopy. *Eur. J. Biochem*., **1999**, *263*, 646-655.

29 E. R. Lacy, K. K. Cox, W. D. Wilson, M. Lee. Recognition of T*G mismatched base pairs in DNA by stacked imidazole-containing polyamides: surface plasmon resonance and circular dichroism studies. *Nucleic. Acids. Res*., **2002**, *30*, 1834-1841.

30 D. Renneberg, P. B. Dervan. Imidazolepyridine/Pyrrole and hydroxybenzimidazole/pyrrole pairs for DNA minor groove recognition. *J. Am. Chem. Soc*, **2003**, *125*, 5707-5716.

31 C. A. Briehn, P. Weyermann, P. B. Dervan. Alternative heterocycles for DNA recognition: the benzimidazole/imidazole pair. *Chemistry*, **2003**, *9*, 2110-2122.

32 M. A. Marques, R. M. Doss, S. Foister, P. B. Dervan. Expanding the repertoire of heterocycle ring pairs for programmable minor groove DNA recognition. *J. Am. Chem. Soc*., **2004**, *126*, 10339-10349.

33 R. M. Doss, M. A. Marques, S. Foister, D. M. Chenoweth, P. B. Dervan. Programmable oligomers for minor groove DNA recognition. *J. Am. Chem. Soc*., **2006**, *128*, 9074-9079.

34 D. M. Chenoweth, A. Viger, P. B. Dervan. Fluorescent sequence-specific dsDNA binding oligomers. *J. Am. Chem. Soc*., **2007**, *129*, 2216-2217.

35 K. Kiakos, L. Pett, V. Satam, P. Patil, D. Hochhauser, M. Lee, J. A. Hartley. Nuclear localization and gene expression modulation by a fluorescent sequence-selective p-anisyl-benzimidazolecarboxamido imidazole-pyrrole polyamide. *Chem. Biol*., **2015**, *22*, 862-875.

36 A. Franks, C. Tronrud, K. Kiakos, J. Kluza, M. Munde, T. Brown, H. Mackay, W. D. Wilson, D. Hochhauser, J. A. Hartley, M. Lee. Targeting the ICB2 site of the topoisomerase llalpha promoter with a formamido-pyrrole-imidazole-pyrrole H-pin polyamide. *Bioorg. Med. Chem.*, **2010**, *18*, 5553-5561.

37 S. Nishijima, K. Shinohara, T. Bando, M. Minoshima, G. Kashiwazaki, H. Sugiyama. Cell permeability of Py-Im-polyamide-fluorescein conjugates: influence of molecular size and Py/Im content. *Bioorg. Med. Chem.*, **2010**, *18*, 978-983.

38 N. G. Nickols, C. S. Jacobs, M. E. Farkas, P. B. Dervan. Improved nuclear localization of DNA-binding polyamides. *Nucleic Acids Res.*, **2007**, *35*, 363-370.

39 K. L. Buchmueller, A. M. Staples, C. M. Howard, S. M. Horick, P. B. Uthe, N. M. Le, K. K. Cox, B. Nguyen, K. A. Pacheco, W. D. Wilson, M. Lee. Extending the language of DNA molecular recognition by polyamides: Unexpected influence of imidazole and pyrrole arrangement on binding affinity and specificity. *J. Am. Chem. Soc*., **2005**, *127*, 742-750.

40 K. L. Buchmueller, A. M. Staples, P. B. Uthe, C. M. Howard, K. A. Pacheco, K. K. Cox, J. A. Henry, S. L. Bailey, S. M. Horick, B. Nguyen, W. D. Wilson, M. Lee. Molecular

recognition of DNA base pairs by the formamido/pyrrole and fomamido/imidazole pairings in stacked polyamides. *Nucleic Acids Res.,* **2005**, *33*, 912-921.

41 E. R. Lacy, B. Nguyen, M. Le, K. K. Cox, C. O'Hare, J. A. Hartley, M. Lee, W. D. Wilson. Energetic basis for selective recognition of T·G mismatched base pairs in DNA by imidazole-rich polyamides. *Nucleic. Acids. Res*., **2004**, *32*, 2000-2007.

42 V. C. Rucker, S. Foister, C. Melander, P. B. Dervan. Sequence specific fluorescence detection of double strand DNA. *J. Am. Chem. Soc*., **2003**, *125*, 1195-1202.

43 X. L. Yang, R. B. Hubbard, M. Lee, Z. F. Tao, H. Sugiyama, A. H. Wang. Imidazole-imidazole pair as a minor groove recognition motif for T:G mismatched base pairs. *Nucleic. Acids. Res*., **1999**, *27*, 4183-4190.

44 T. Brown, H. Mackay, M. Turlington, A. Sutterfield, T. Smith, A. Sielaff, L. Westrate, C. Bruce, J. Kluza, C. O'Hare, B. Nguyen, W. D. Wilson, J. A. Hartley, M. Lee. Modifying the N-terminus of polyamides: PyImPyIm has improved sequence specificity over f-ImPyIm. *Bioorg. Med. Chem*., **2008**, *16*, 5266-5276.

45 K. L. Buchmueller, S. L. Bailey, D. A. Mattews, Z. T. Taherbhai, J. K. Register, Z. S. Davis, C. D. Bruce, C. O'Hare, J. A. Hartley, M. Lee. Physical and structural basis for the strong interactions of the –ImPy- central pairing motif in the polyamide f-ImPyIm. *Biochemistry*, **2006**, *45*, 13551-13565.

46 S. C. Lin, K. Kiakos, M. Lee, D. Hochhauser, J. A. Hartley, *Proc. 102nd Annual Meeting of the Am. Assoc. Cancer Res*. **2011**, Abs 665.

47 B. Liu, S. Wang, K. Aston, K. J. Koeller, S. F. H. Kermani, C. H. Castaneda, M. J. Scuderi, R. Luo, J. K. Bashkin, W. D. Wilson. β-alanine and N-terminal cationic substituents affect polyamide-DNA binding. *Org. Biomol. Chem.,* **2017**, *15*, 9880-9888.

48 J. L. Meier, D. C. Montgomery, P. B. Dervan. Enhancing the cellular uptake of Py-Im polyamides through next-generation aryl turns. *Nucleic Acids Res*, **2012**, *40*, 2345-2356.

49 P. B. Dervan, R. W. Burli. Sequence-specific DNA recognition by polyamides. *Curr. Opin. Chem. Biol*, **1999**, *3*, 688-693.

3.9 Tables and Figures

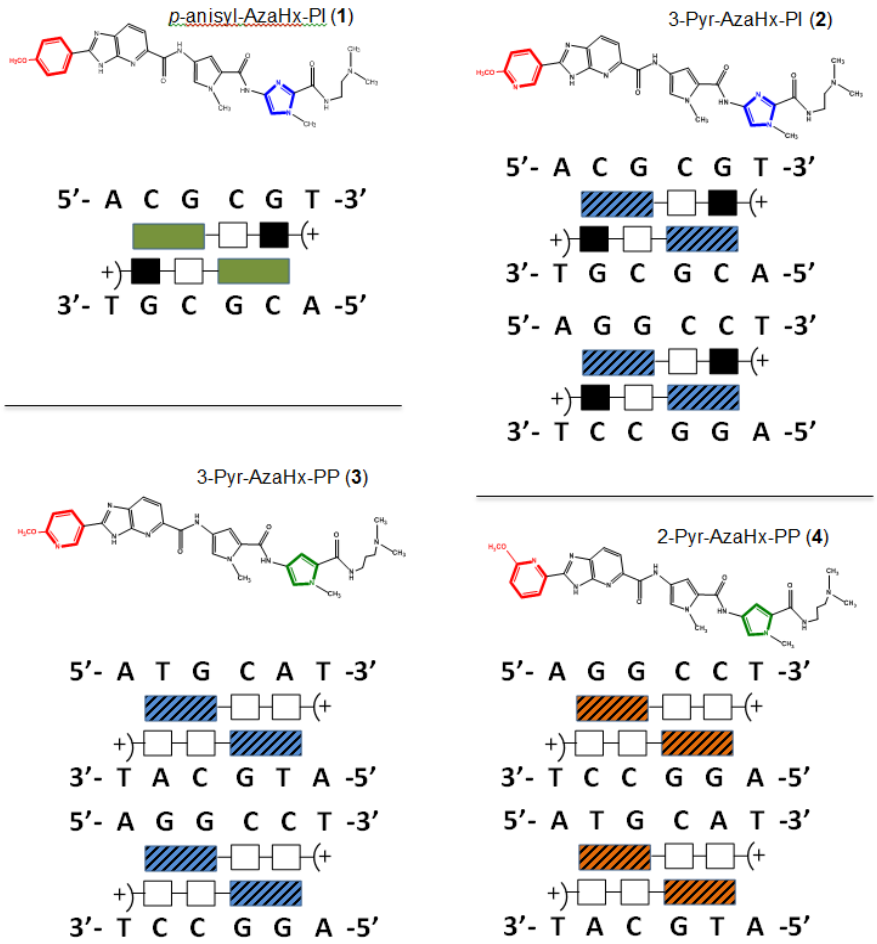


Figure 3.1 Structures and schematic presentation of predicted binding to DNA sites of three Pyr-AzaHx polyamides.

Blue rectangles represent 3-Pyr-AzaHx moiety, empty squares represent pyrrole, filled squares represent imidazole, and orange rectangles represent 2-Pyr-AzaHx group. 2-Pyr-AzaHx-PP has two predicted recognition sites.

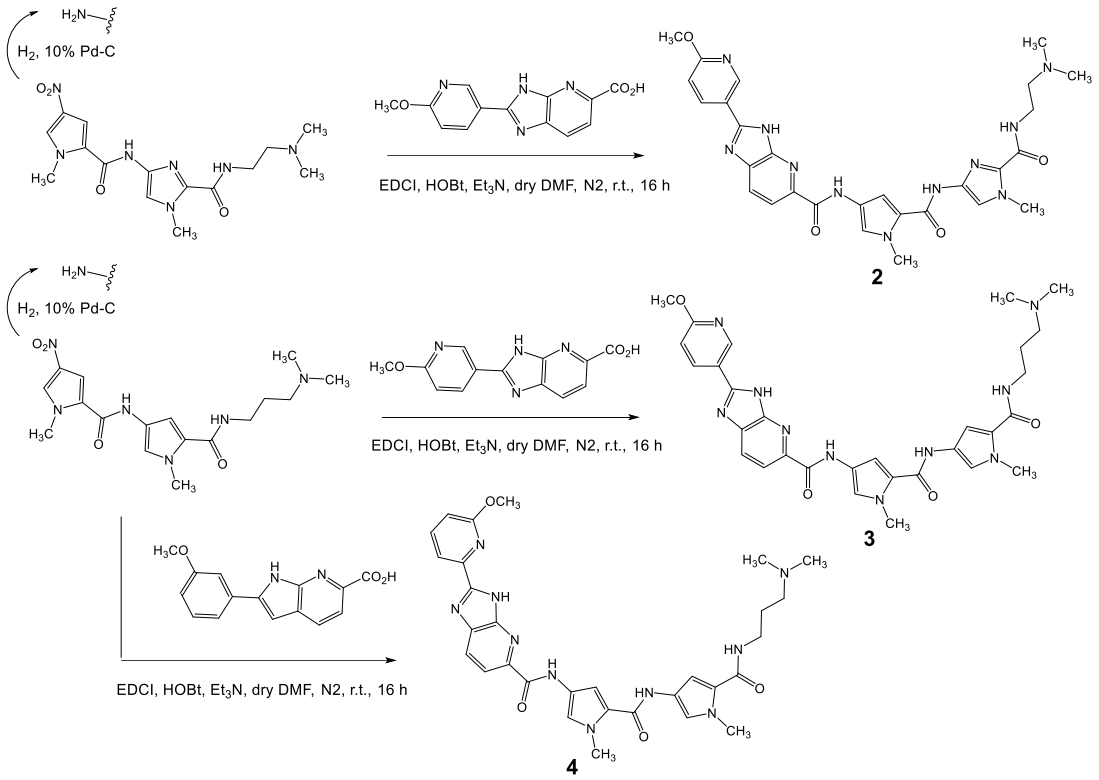


Figure 3.2 Synthetic schemes of compound **2-4**.

Table 3.1 Thermal melting profile of three Pyr-AzaHx polyamides binding to their predicted cognate sequences and mutant sequences

	CGCG	TGCA	GGCC
Tm of DNAs (°C)	68.8	61	67.2
Polyamides		ΔTm (°C)	
3-Pyr-AzaHx-PI (**2**)	15.9	15.5	8.8
3-Pyr-AzaHx-PP (**3**)	7.4	14.5	7.8
2-Pyr-AzaHx-PP (**4**)	1.2	0.8	0

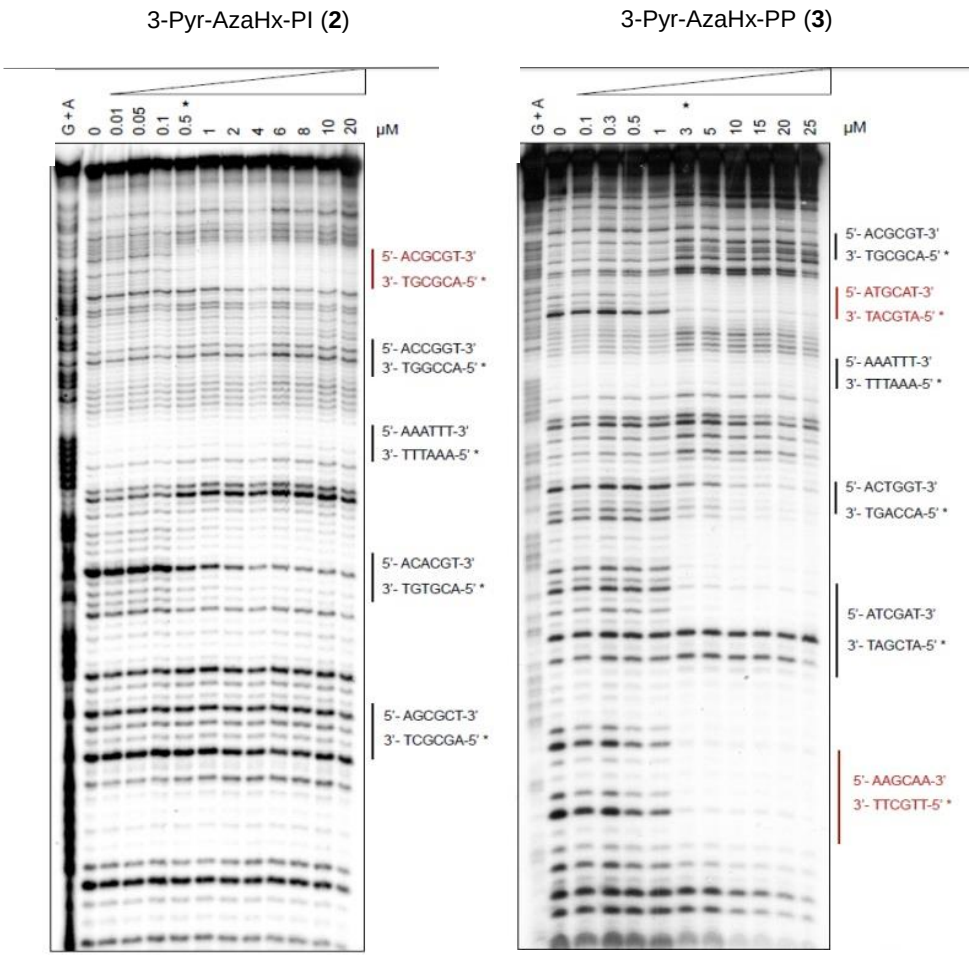


Figure 3.3 Autoradiograms of DNase I footprinting gels corresponding to 3-Pyr-AzaHx-PI, **2** (A) and 3-Pyr-AzaHx-PP, **3** (B).

*The DNA binding properties of **2** and **3** were assessed using DNA fragment 1 and 2, respectively. The appropriated ^{32}P-radiolabelled DNA fragment was incubated for 1 h at room temperature in TN buffer (10 mM Tris-HCl, 10 mM NaCl, pH 7.0) containing the required polyamide concentration. DNase I digestion was initiated by the addition of DNase I solution (20mM NaCl, 2mM $MgCl_2$, 2mM $MnCl_2$, DNase I 0.02U, pH 8.0) and was terminated after 3 min by snap freezing the samples on dry ice. Following lyophilization and resuspension in formamide loading dye (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, and 0.05% xylene cyanol), the cleavage products of the DNase I digestion reactions were resolved on 10% polyacrylamide gels. The concentrations (μM) used are shown at the top of each gel. 0; controlled cleavage reaction, containing no drug. The cognate sequences are labelled red and the non-cognate sites are also indicated. Asterisk marks the concentration (μM) at which footprints become evident. G+A represents a formic acid-piperidine marker specific for purines.*

DNA fragment 2

DNA fragment 3

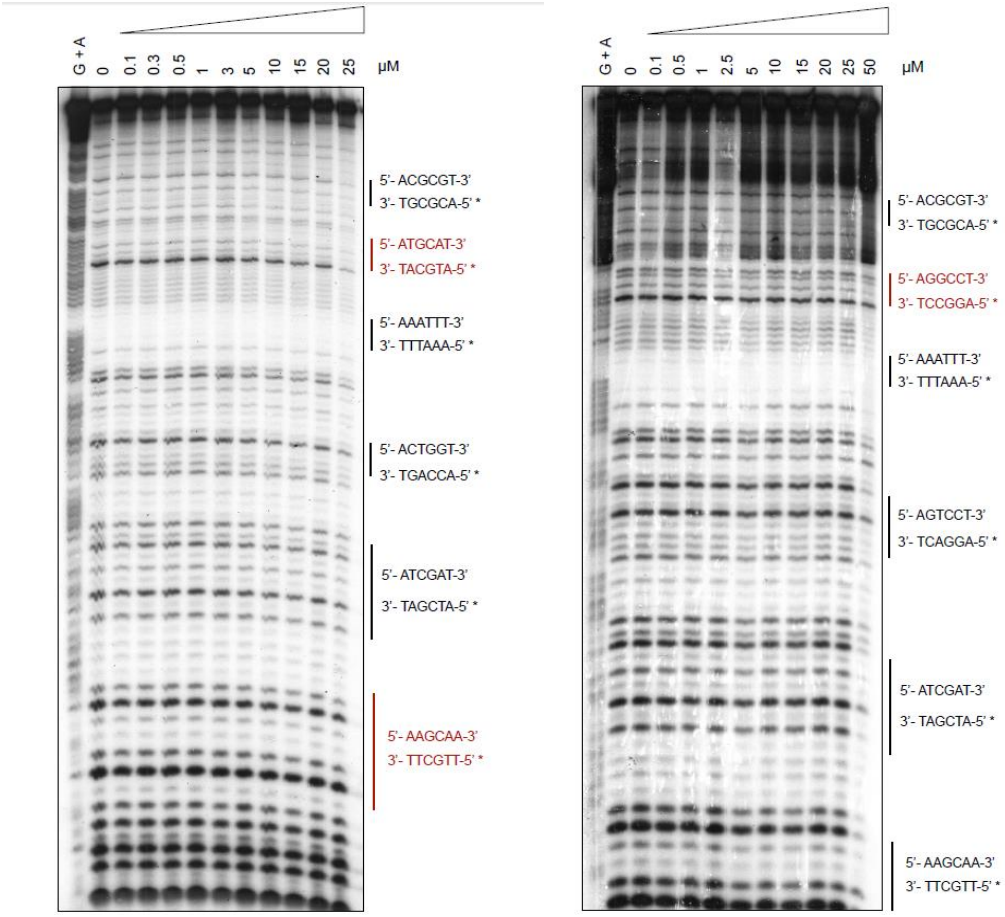


Figure 3.4 Autoradiogram of the DNase I footprinting gel of 2-Pyr-AzaHx-PP on A) DNA fragment 2 and B) DNA fragment 3.

The cognate sequences are labelled red and the non-cognate sites are also indicated. Asterisk marks the concentration (μM) at which footprints become evident. G+A represents a formic acid-piperidine marker specific for purines.

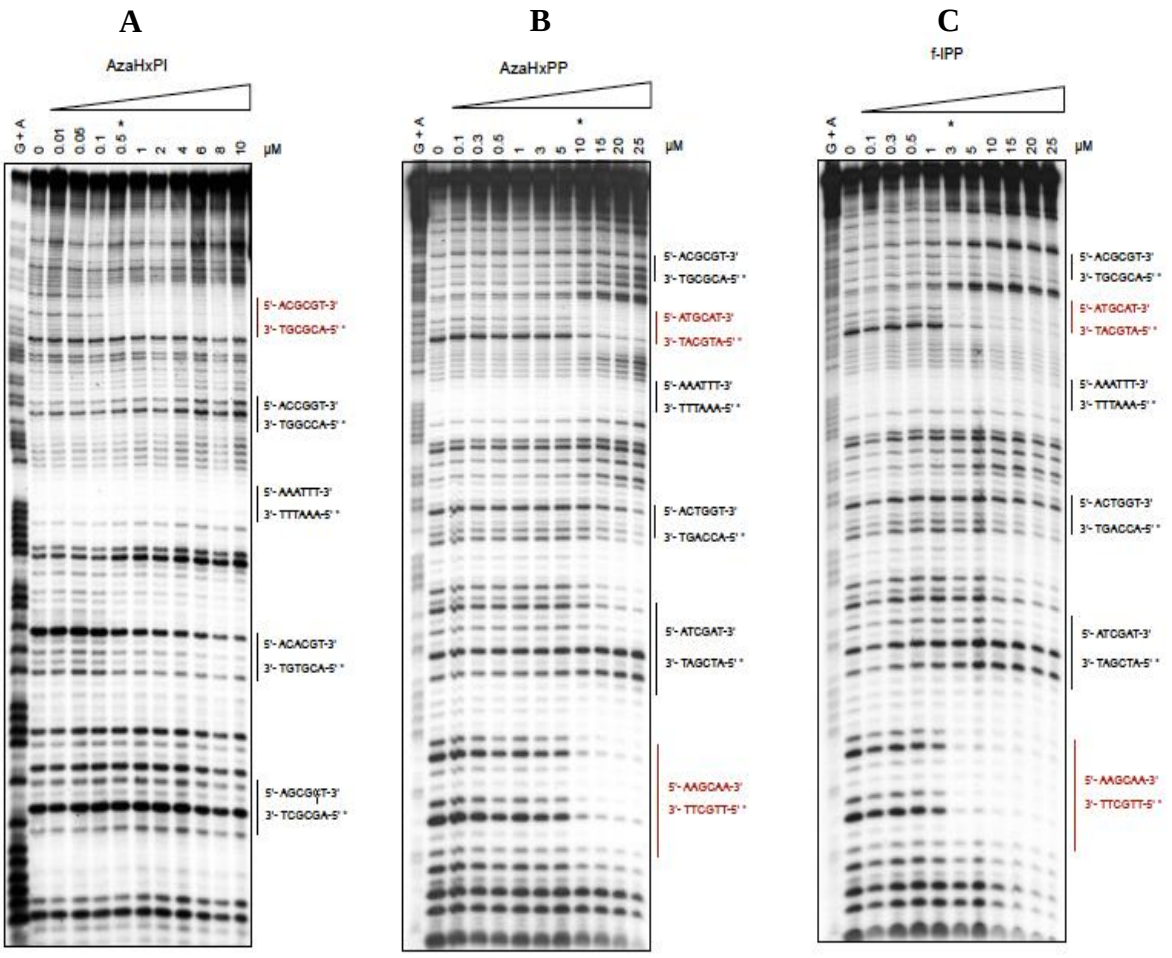


Figure 3.5 Autoradiogram of the DNase I footprinting gel of A) AzaHx-PI on DNA fragment 1 and B) AzaHx-PP and C) f-IPP on DNA fragment 2.

The cognate sequences are labelled red and the non-cognate sites are also indicated. Asterisk marks the concentration (μM) at which footprints become evident. G+A represents a formic acid-piperidine marker specific for purines.

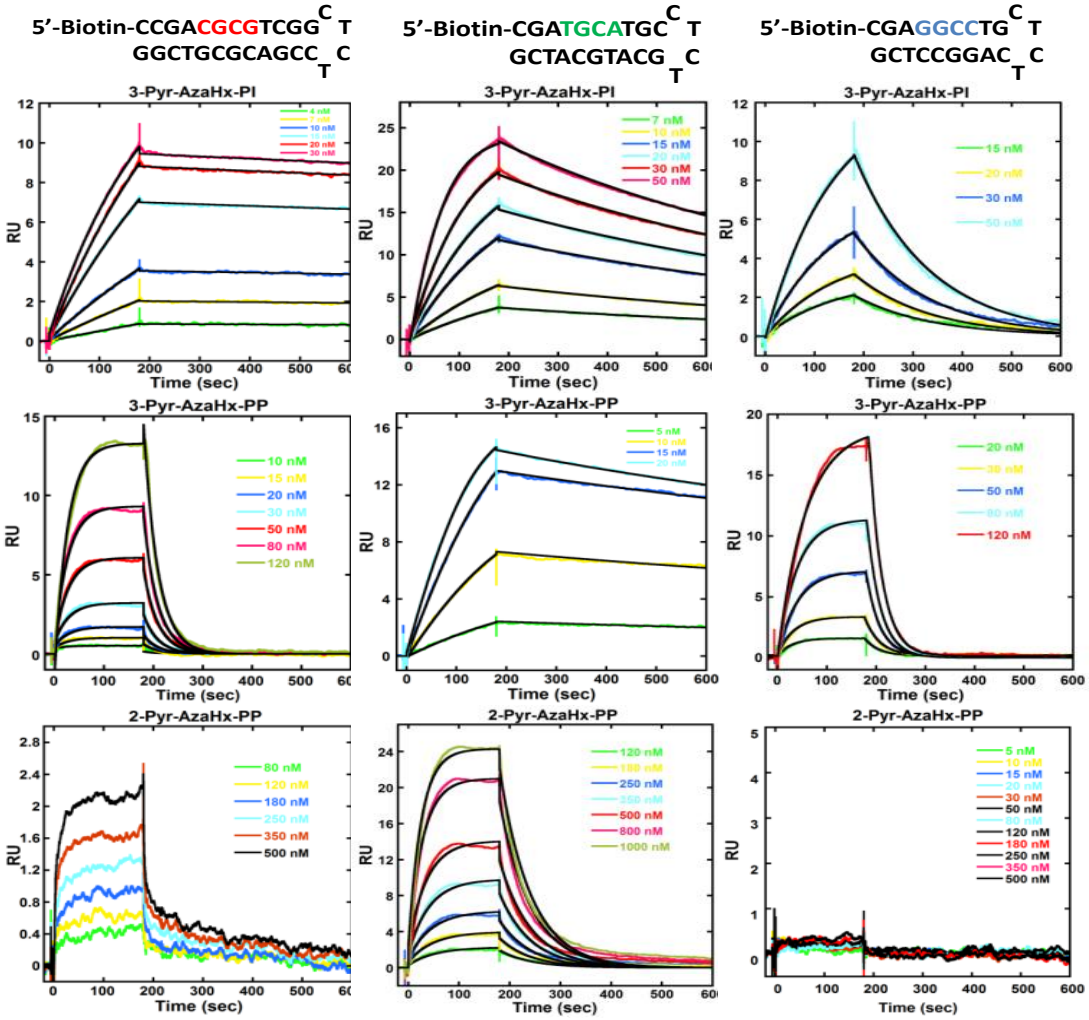


Figure 3.6 SPR sensorgrams of Pyr-AzaHx polyamides binding to different DNAs.

The DNA sequences are listed at the top of the column and the compounds are specified on each sensorgram. The whole sequences of the DNAs are listed in the Supporting Information. The colored lines are experimental data and the black overlays are kinetic fitting. The sensorgram at the bottom left corner was fitted with steady state fitting and bottom right corner sensorgram shows no binding. The buffer used for SPR is 10 mM cacodylic acid, 100 mM NaCl, 1mM EDTA and 0.05% v/v surfactant Polysorbate 20 (P20), pH 6.2, 25°C.

Table 3.2 Kinetic rate constants and equilibrium constants derived from SPR

Compound	DNA	k_a (x10^4 $M^{-1}s^{-1}$)	k_d(x$10^{-3}s^{-1}$)	K_D (nM)
3-Pyr-AzaHx-PI	ACGCGT	16 ± 0.3	0.12 ± 0.001	0.75 ± 0.01
	ATGCAT	40 ± 0.7	1.30 ± 0.01	3.20 ± 0.2
	AGGCCT	1.70 ± 0.5	6.60 ± 0.03	397 ± 17
3-Pyr-AzaHx-PP	ACGCGT	3.30 ± 0.03	34 ± 0.04	1035 ± 22
	ATGCAT	35 ± 0.2	0.40 ± 0.001	1.10 ± 0.03
	AGGCCT	1.00 ± 0.01	40 ± 0.4	3996 ± 472
2-Pyr-AzaHx-PP	ACGCGT	ND[a]	ND[a]	1420 ± 700
	ATGCAT	2.30 ± 0.02	16 ± 0.06	698 ± 237
	AGGCCT	NB[b]	NB[b]	NB [b]

[a] The values are too fast to be determined. [b] No binding is detected.

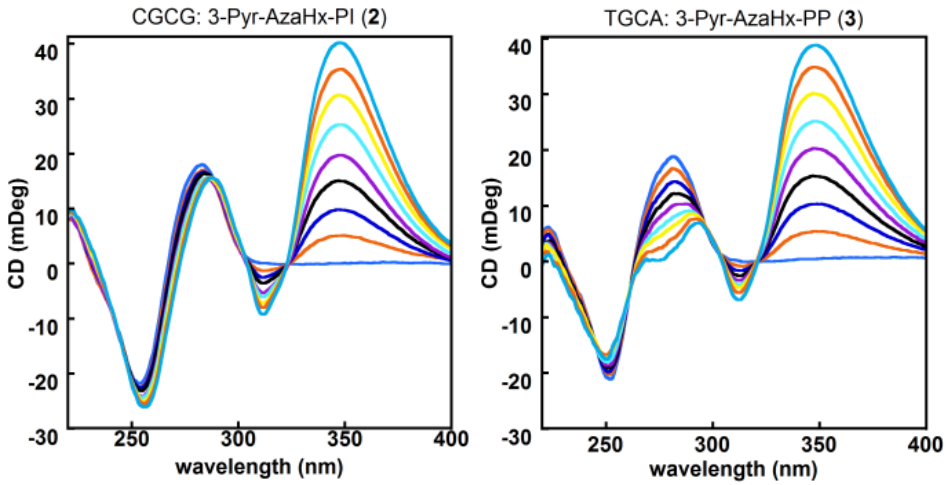


Figure 3.7 CD spectra depicting 3-Pyr-AzaHx molecules (**2**) and (**3**) binding to their cognate DNAs, respectively.

The DNA sequences are listed in Supporting Information. The buffer consists of 10 mM sodium phosphate, 1 mM EDTA, pH 6.2.

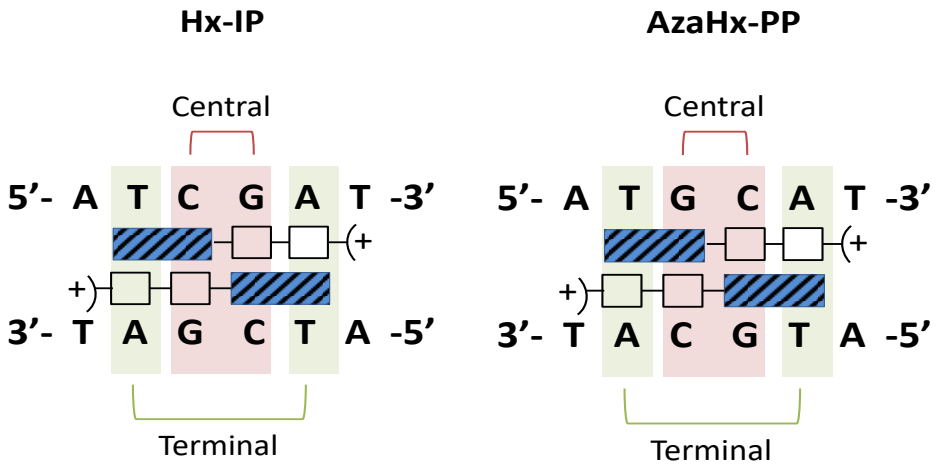


Figure 3.8 Central and terminal pairing rules illustrated by different binding of Hx-IP and AzaHx-PP.

The central and terminal base pairs are indicated in pink and green boxes, respectively.

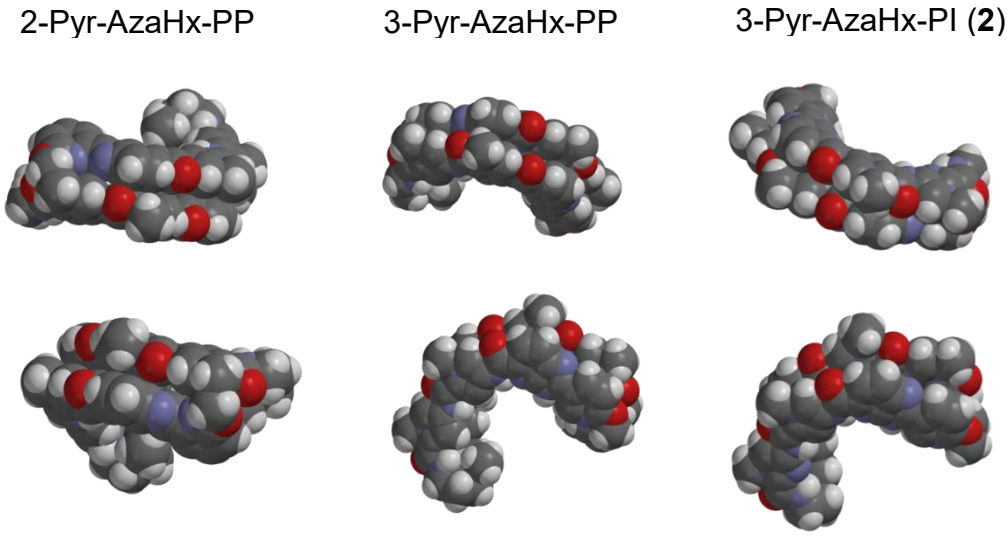


Figure 3.9 Stacked models of all three molecules from Spartan calculations.

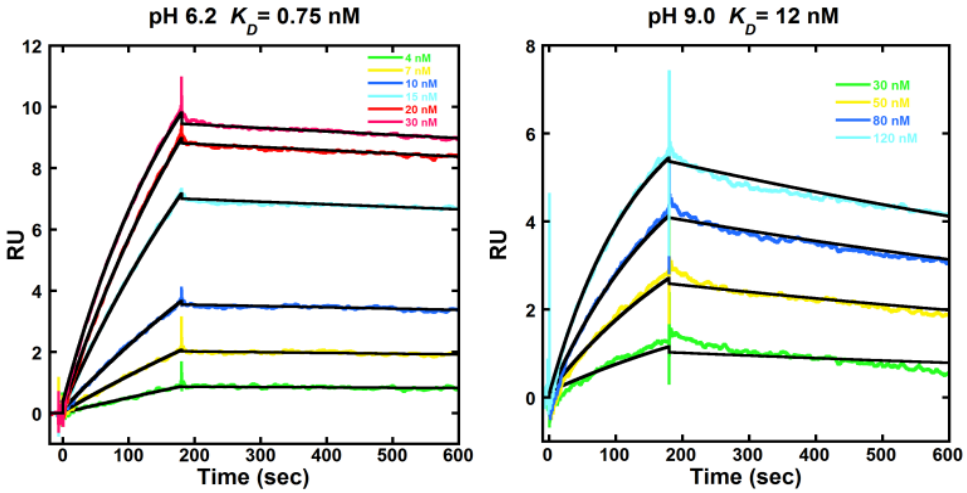


Figure 3.10 SPR sensorgrams and equilibrium binding constants of 3-Pyr-AzaHx-PI under different pH.

The buffer used at pH 6.2 is consisted of 10 mM cacodylic acid, 100 mM NaCl, 1mM EDTA and 0.05% v/v surfactant Polysorbate 20 (P20), 25°C. The buffer at pH 9.0 contains 50 mM Tris-HCl, 100 mM NaCl, 1mM EDTA and 0.05% v/v P20, 25°C. The whole sequence of 5'-ACGCG-3' is listed in the Supporting Information.

4 MODULATING DNA BY POLYAMIDES TO REGULATE TRANSCRIPTION FACTOR PU.1-DNA BINDING INTERACTIONS

Beibei Liu,[1] James K. Bashkin,[2] Gregory M. K. Poon,[1] Shuo Wang,[1] Siming Wang,[1] and W. David Wilson [1,] *

[1]Department of Chemistry, Georgia State University, Atlanta, GA 30303, USA

[2]Department of Chemistry & Biochemistry, Center for Nanoscience, University of Missouri-St. Louis, St. Louis, MO 63121, USA

*To whom correspondence should be addressed:

* E-mail: wdw@gsu.edu; Fax: +1 404-413-5505; Tel: +1 404-413-5503

4.1 Abstract

Hairpin polyamides are synthetic small molecules that bind DNA minor groove sequence-selectively and, in many sequences, induce widening of the minor groove and compression of the major groove. The structural distortion of DNA caused by polyamides has enhanced our understanding of the regulation of DNA-binding proteins *via* polyamides. Polyamides have DNA binding affinities that are comparable to those proteins, therefore, can potentially be used as therapeutic agents to treat diseases caused by aberrant gene expression. In fact, many diseases are characterized by over- or under-expressed genes. PU.1 is a transcription factor that regulates many immune system genes. Aberrant expression of PU.1 was associated with the development of acute myeloid leukemia (AML). We have, therefore, designed and synthesized ten hairpin polyamides to investigate their capacity in controlling the PU.1-DNA interaction. Our results showed that nine of the polyamides disrupt PU.1-DNA binding and the inhibition capacity strongly correlates with binding affinity. One molecule, FH1024, was observed forming a FH1024-PU.1-DNA trimer instead of inhibiting PU.1-DNA binding. This is the first report of a small molecule that is potentially a weak agonist that recruits PU.1 to DNA. The finding sheds light on the design of polyamides that exhibit novel regulatory mechanisms on protein-DNA binding.

4.2 Introduction

Polyamides (PAs) are sequence-selective DNA minor groove binders that are composed of N-methylpyrrole (Py) and N-methylimidazole (Im) heterocycles (1, 2). Hairpin PAs are formed by covalently linking two PA strands side-by-side, typically by using γ-aminobutyric acid (a γ-turn) to enhance DNA binding affinity and specificity (3, 4). Upon binding to DNA, Py/Py recognizes both T/A and A/T base pairs, while the Im/Py pair distinguishes G/C from C/G (5-8). PAs can be programmed to target selectively many predetermined DNA sequences according to the reported

binding rules, enabling very potent interactions with DNA that are comparable with or stronger than interactions of DNA-binding proteins (2, 5, 6, 8-11). In addition, hairpin PAs have also been shown to be cell permeable (12-15). Therefore, PAs are promising agents that can potentially alter the regulation of specific DNA-binding proteins for the treatment of diseases caused by aberrant gene expression (16, 17).

As an ETS (E26 transformation specific) family member, PU.1 (Purine rich box 1) is a transcription factor involved in normal hematopoiesis and a growing list of diseases. PU.1 regulates the expression of a large network of genes related to inflammation (18). For example, PU.1 controls the expression of receptors for IL-2Rγ and IL-7Rα, and also regulates expression of Toll-like receptors which may control the signaling pathways of certain auto-immune diseases (19-23). In addition, down-regulation of PU.1 expression to 20% of normal cell levels is reported to induce acute myeloid leukemia (AML) in mice, whereas a decrease to 50% of normal cell expression retains normal hematopoiesis (24, 25). On the other hand, complete knock-down of PU.1 causes severe dysfunction of hematopoietic stem cells and leads to cell death (24-27). Thus, either raising PU.1 expression to normal functional levels or completely blocking the PU.1 transcriptional activity in AML cells is potential approach to the treatment of AML.

All ETS family transcription factors bind to DNA at the consensus binding site 5'-GGAA/T-3' with various flanking sequences. For example, the winged helix PU.1 targets the λB motif of the Ig2-4 enhancer (λB promoter site: 5'-CCAAATAAAA**GGAA**GTGAAACCAAG-3', Fig. 1) and makes contacts with DNA by inserting a recognition helix into the major groove of the consensus site while keeping adjacent protein loops in touch with the backbone of flanking DNA sequences (28). The λB DNA sequence is a template that has been widely used for the study of

PU.1-DNA interactions. Targeting the 5'-side AT track of the GGAA site using naturally existing distamycin demonstrated inhibition of PU.1 binding (29, 30).

To date, activation of gene expression through PAs has been done by conjugating PAs with specific recruiting motifs like short peptides and protein-binding ligands (31-37). Previous studies showed that the 3' flanking sequence plays a critical role in PU.1-λB DNA binding, with an equilibrium (dissociation) constant $K_D = 2.68 \pm 0.41$ nM compared to $K_D = 240 \pm 70$ nM achieved by simple mutation of the 3' flanking GTG of λB to TGG (38). Therefore, we designed a set of PA analogs based on the recognition rules to bind to the 3' flanking side (5'-GGA**AGTGA**-3') of the PU.1 binding site, aiming to interfere allosterically with PU.1 binding to DNA (Fig. 4.1). These PAs are structurally modified and their binding properties to λB DNA have been determined in previous literature (39).

Herein, we report the cell-free studies of designed minor groove-binding PAs modulating the major groove targeting PU.1. Previously, tandem hairpin PAs displaced the major groove-binding E2 protein of human papillomavirus from its cognate site. In that case, the viral protein bends DNA about 180 ° and the ability of the minor-groove binding PAs to displace the major groove-binding protein was attributed to (a) thermodynamics, since the PAs bind more tightly to DNA than the protein, and (b) dynamic equilibria, in which each DNA ligand associates and dissociates to its DNA target, allowing the tighter-binding ligand to bind the DNA no matter which ligand is bound first (17). Our results show that nine out of ten PAs allosterically disrupt the PU.1-DNA interaction with a linear correlation between PA binding affinity and inhibition efficacy. One PA, FH1024, is not a PU.1 inhibitor and forms a FH1024-PU.1-λB DNA trimer. Altogether, the results have offered new insights into the effects of DNA modulation by PAs, deepened our understanding of the molecular basis of small molecule-DNA-protein interactions and provided

guidance for optimizing molecular design and consequently improving the efficiency and accuracy of gene regulation in disease treatment.

4.3 Materials and Methods

4.3.1 DNA

DNAs:

The DNA oligomers in this study were purchased from Integrated DNA Technologies, Inc. (IDT, Coralville, IA) with HPLC purification.

DNA sequence used for ESI-MS and EMSA:

5’-CCAAATAAAA**GGAA**GTGAAACCAAG*CTCT*CTTGGTTTCAC**TTCC**TTTTATT TGG-3’

The bold, underlined site is the PU.1 binding site, the PA target binding site is indicated in red. The hairpin is shown in italic. The same sequence is used for SPR experiments but with a 5’-biotin.

4.3.2 Polyamides synthesis and preparation

Compounds are synthesized and prepared as previously described (39).

4.3.3 PU.1 purification

Recombinant PU.1 ETS domain (murine residues 167 to 272) was expressed and purified as reported previously (40, 41).The final protein product, which does not contain the His_6 purification tag, was stored at 4 ºC. The protein is stable for at least a few weeks under this condition. Protein concentration was determined by absorption at 280 nm using an extinction coefficient of 22,460 M^{-1} cm^{-1}.

4.3.4 Surface Plasmon Resonance

In addition to the binding experiments, we have developed a novel SPR procedure to study the effects of small molecules on protein-DNA interactions. The SPR experiments were carried out on a four-channel Biacore T200 optical biosensor system (GE Healthcare, Inc., Piscataway, NJ). The biotinylated DNA was immobilized on a sensor chip functionalized with streptavidin. The first channel was used as reference channel and does not have DNA immobilized. The samples contain a mixture of constant 50 nM PU.1 and different concentrations of PA. For the amount of immobilized DNA, 50 nM PU.1 can reach steady state binding and RUmax. The PA concentration (starting from 0 nM) increases as the number of the experimental flow cycle rises. In each cycle, the sample was injected over the sensor chip surface at a flow rate of 50 $\mu L \cdot min^{-1}$. During the sample flow, PU.1 and PA are competing or cooperating to form a complex with the DNA immobilized on the sensor chip. The sensorchip was washed with pure buffer at the end of each sample injection to help dissociate the existing complex. Then, 1 mM NaCl was injected as regeneration buffer to dissociate any remaining complex and prepare the chip for the next cycle. PA alone was injected in the last cycle for baseline subtraction and signal reference. The SPR sensorgrams were generated immediately after finishing the experiments and were presented as a function of response unit (RU) *vs*. time (Fig. 4.2). The resulting RU changes of PU.1 in the presence of PA at equilibrium were then plotted against PA concentration (Fig. 4.3). The IC_{50} value, which represents inhibition efficacy, was determined at the PA concentration where 50% of PU.1 is inhibited from binding to DNA. This experiment was performed in phosphate buffer (25 mM Na_2HPO_4, 400 mM NaCl, 1 mM EDTA and 0.05% v/v surfactant P20, pH 7.4, filtered and degassed), which is compatible with both PU.1 and the PA.

4.3.5 Electrophoresis Mobility Shift Assay

Electrophoresis Mobility Shift Assay (EMSA) was carried out as a complementary and validation study. The samples were prepared in a buffer composed of 10 mM Tris (pH=7.4), 150 mM NaCl, 0.1 mM EDTA, 100 ng/µL BSA, and 10% v/v glycerol. 25 nM PU.1was mixed with appropriate ratios of PA before 25 nM DNA was added. The samples were then loaded onto 12% native polyacrylamide gel and run in 1x TBE (Tris-borate-EDTA) buffer. The gel was run for 3 h at 120 V. It was then stained with SYBR Gold Nucleic Acid Gel Stain (Invitrogen, Carlsbad, CA) for 20 min before visualization on a Typhoon FLA 9500 laser scanner.

4.3.6 Electrospray Ionization Mass Spectrometry

Both PU.1 and DNA were dialyzed against 150 mM $NH_4CH_3CO_2$ buffer, pH 6.8, for one week, with a total of three times buffer exchange. This step is carried out to eliminate the Na^+ present in un-dialyzed DNA, which can form adducts resulting in complicated MS data that are not readily interpretable. A stock solution of 0.5 mM each compound was made by dissolving the compound in distilled deionized (DDI) water. The MS samples were prepared by mixing PU.1 and DNA at 1 to 1 ratio at the desired concentration. Compounds for titration were prepared in separate tubes and added to the pre-mixed PU.1-DNA. Immediately prior to injection, 5 µL methanol was added to the sample to facilitate solution to gas phase transition. The total sample volume was 100 µL.

The MS experiments were performed on a Waters Micromass ESI-Q-ToF spectrometer (Waters Corporation, Milford, MA, USA). Data were collected and analyzed using MassLynx 4.1 software. In positive ion mode, samples were injected at a rate of 5 $\mu L \cdot min^{-1}$ and scanned for *m/z* range from 300-3000. The last two minutes of the mass spectral data were averaged and the resulting spectrum was convolved to give the intact molecular mass. The instrument parameters

were as follows: capillary voltage of 2500 V, cone voltage of 20 V, extraction voltage of 2 V, desolvation temperature of 100 ^{o}C, and source temperature of 70 ^{o}C. Nitrogen was used as nebulizing and drying gas.

4.4 Results

4.4.1 Quantitative evaluation of PU.1 inhibition by PA

Cell-free studies of competitive binding of PAs and PU.1 to λB DNA were carried out by SPR. The SPR method developed by our group for inhibition studies is efficient, label-free, and acquires inhibition efficacies that agree with other methods (29). The inhibition of PU.1 by a PA was evaluated by injecting 50 nM PU.1 to establish a high molecular weight baseline, and injecting PA alone to establish a low molecular weight baseline. To determine binding and inhibition parameters, increasing concentrations of PA were injected onto the sensorchip surface, which had been immobilized with target DNA, while maintaining a fixed concentration of 50 nM PU.1. The signal or response unit (RU) generated in SPR is proportional to the molecular weight of the bound ligand, in this case, PU.1 or PA or both. Given that the molecular weight of PU.1 (12 kDa) is much larger than that of the present PA library (around 1250 Da on average), the RU caused by binding of PU.1 is much higher than that caused by the binding of PA. This molecular weight-dependent RU difference can lead to three different outcomes when a PA is present during PU.1-DNA binding: 1) the RU drops if PA replaces PU.1 upon binding DNA; 2) the RU stays the same if PA fails to compete with PU.1 for DNA; 3) the RU rises as the binding of PA increases if PA and PU.1 bind to DNA simultaneously or cooperatively.

Sensorgrams reflecting the real-time binding activity of each sample (PA and PU.1 mixture) were generated. As shown in Fig. 4.2, for each PA depicted there, the RU decreased as the concentration of PA increased, indicating that the PAs competed with and eventually replaced

PU.1 on the DNA. For strong inhibitors such as FH1028 and KA2115, the RU descended rapidly with a slight increase in PA concentration. On the contrary, much smaller RU changes were observed for the weak inhibitors, KA2114 and FH1026, even with large increases in PA concentration. It is noteworthy that for weak inhibitors like FH1026, the RU change plateaued as the PAs reached certain concentrations. This is likely due to the combination of specific and non-specific binding of the molecules when presented in excess. Because the PU.1 binding site typically occupies 10-16 bp of the 23-bp DNA sequence, the additional, untargeted base pairs are exposed to excess PA. In addition, due to the degenerate recognition of A and T by pyrrole (5-8) as well as the tolerance of imidazole for interacting with bases other than G (39, 42-44), excess PA tends to bind non-specific DNA sites.

To determine the IC_{50} value (inhibition efficacy), the PA concentration at which 50% of PU.1 binding is inhibited, the RU data were plotted as a function of PA concentration (Fig. 4.3). Table 4.1 is a summary of PAs with their binding affinities (previously determined (39)) as well as their corresponding inhibition efficacy (IC_{50}) values. The PAs exhibited strong inhibition with a number of IC_{50} values in the single-digit nM range. As shown in Table 4.1, there is a strong correlation between the IC_{50} and K_D values. To further quantify the correlation, the log values of IC_{50} were plotted as a function of log (K_D), and showed a linear correlation between binding affinity and inhibition efficacy (Fig. 4.4). This linear correlation implies that these PA analogs bind to DNA in the same manner. More specifically, for eight-ring PAs that have no more than two β inserts, if the two β inserts are not next to each other on the same strand, the modified PAs still recognize the same DNA sites. This finding removes our concern about future eight-ring PA design that the molecule might change its cognate binding site after β modification.

4.4.2 FH1024 is an outlier and does not inhibit PU.1-DNA binding

FH1024 is a medium-affinity binding agent modified from the parent PA, KA 2035, with the Py/Py substituted by β/β and tetramethylguanidinium (TMG) as the N-terminus (Fig. 4.1). According to the correlation established (Fig. 4), FH1024 was anticipated to inhibit PU.1-DNA binding with moderate inhibition efficacy of $IC_{50} \approx 50$ nM. To our surprise, when the mixture of FH1024 and PU.1 flowed through the sensorchip surface, the resulting RU did not decrease as it did with the other PAs under the same assay conditions. Instead, a RU increase occurred as the concentration of FH1024 increased (Fig. 4.5a). The observation leads us to suggest that FH1024 is not a PU.1 inhibitor but can bind simultaneously to DNA with PU.1. Furthermore, the difference of RU (ΔRU) caused by binding of the mixture is more than the ΔRU resulting from FH1024 binding alone, and this could indicate that FH1024 promotes additional PU.1-DNA binding. To further investigate the binding behaviors of PU.1-FH1024 mixture, different salt concentrations were used. As shown in Fig. 5a, within the salt concentration range used, the RU generated by flowing of FH1024-PU.1 mixture increased as the concentration of FH1024 went up, corroborating the assumption that FH1024 and PU.1 bind to DNA simultaneously. The idea of FH1024 promoting PU.1 binding is strongly supported by the SPR results at 450 mM NaCl. Compared to the RU at lower salt concentrations, PU.1 binding was significantly reduced at 450 mM NaCl. When FH1024 was added, however, the binding RU value significantly increased and eventually reached saturation at essentially the same RU level (approx. 140 RU) as at the lower salt concentrations. This clearly indicates that FH1024 is a promoter or weak agonist of PU.1 binding. The slight variation of RU values at saturation in Fig. 4.5 is due to the gradually reducing binding capacity on the sensor chip with time and usage.

To investigate if simultaneous binding is sensitive to the order of addition, in other words, if PU.1 still binds when FH1024 pre-occupies its binding site, we switched the titration sequence by keeping the concentration of FH1024 constant at 50 nM while increasing the concentration of PU.1 in each flow cycle. If FH1024 inhibited the binding of PU.1, then the injection of PU.1 onto a sensor chip pre-occupied with FH1024 would hardly cause any change in RU, especially at low concentrations of PU.1. In this experiment, the PU.1 mixture sample was injected over the sensor chip pre-mixed with FH1024 and the resulting sensorgrams are shown in Fig. 4.5b. The steady increase in binding RU confirms that PU.1 binds to DNA in the presence of FH1024. Similarly, under a gradient of salt concentrations, RU consistently increased as PU.1 concentrations went up and the RU value at saturation stayed relatively constant, indicating potentially cooperative binding of PU.1 and FH1024 in the major and minor DNA grooves.

4.4.3 Electrophoresis Mobility Shift Assay (EMSA) confirms that FH1024 is not inhibiting PU.1-DNA binding

To test our observation further that PU.1 and FH1024 bind simultaneously to DNA, with enhanced PU.1 binding promoted by FH1024, the interactions of λB-PU.1, λB-FH1024, λB-PU.1-FH1024, and λB-PU.1-FH1028 were explored on native polyacrylamide gels. FH1028 was used because it is structurally very similar to FH1024 and showed strong inhibitory activity in SPR, therefore is a good negative control for FH1024. As shown in Fig. 4.6, λB DNA gives a very thick band at the bottom (lane 1, red arrow). It is noticeable that in lane 1, there is an additional band (blue arrow) besides λB DNA. This is very likely a duplex of two hairpin λB DNAs with open hairpin. λB DNA and PU.1 formed a 1:1 complex (lane 2, green arrow); The duplex λB DNA and PU.1 complex resulted in two bands (blue arrows), one with PU.1 bound on one site, the other with PU.1 bound on both binding sites of the duplex DNA. There is also an additional band at the

top of the gel resulted from non-specific binding of PU.1. Because the molecular weight of FH1024 is much smaller than λB DNA, the binding of λB and FH1024 did not result in any mobility shift of the complex from DNA alone at the experimental resolution (lane 3, red arrow); When λB DNA, PU.1, and FH1024 mixture ran through the gel (lanes 4-6), clear bands of λB DNA-FH1024 (red arrow) and λB DNA-PU.1 complex (green arrow) appeared. The fraction of bound λB DNA in lane 4 and 5 are approximately equal to that of lane 2, which is around 0.28. When we raise the FH1024 ratio to λB: PU.1:FH1024=1:1:4, shown in lane 6, the fraction bound increased to 0.34 and the intensity of duplex λB DNA-PU.1 increased significantly (lane 6, blue arrow). These results suggest that FH1024 did not impede binding between λB and PU.1 and to some extent, promoted PU.1-DNA binding.

On the other hand, when FH1028 was added to λB DNA and PU.1 mixture, the band of both λB-PU.1 and duplex λB-PU.1 complex disappeared almost completely even at ratio λB: PU.1:FH1028=1:1:1 (lane 7). At ratio λB: PU.1:FH1028=1:1:2, the band became even fainter, indicating that FH1028 is a strong inhibitor for PU.1, in agreement with SPR results. It is noteworthy that in the mixture of all three components (lanes 4-8), the DNA band started diffusing in the gel. This is probably because of the DNA conformation adjustment caused by the competition between PA and PU.1 binding and the occupation of the dye staining site on DNA by bound polyamide. Though we were unable to detect λB-PU.1-FH1024 trimer complex due to both small molecular weight of FH1024 and the low resolution of the gel, the consistent result validates the observation in SPR that FH1024 is not inhibiting PU.1-DNA binding.

4.4.4 Analyzing λB-PU.1-FH1024 using Electrospray Ionization Mass Spectroscopy (ESI-MS)

ESI-MS was performed to further investigate the mechanism of action between PA and the PU.1-DNA complex. With high resolution, mass accuracy and sensitivity, mass spectrometry possesses the capability of distinguish between the PU.1-λB DNA dimer and PU.1-λB-PA trimer, despite the relatively small difference in molecular weight between the two complexes. Shown in Fig. 4.7a is the 1:1 mix of PU.1-λB DNA; no significant amount of free PU.1 or free DNA was detected. When we added FH1024 at ratio λB: PU.1: FH1024 = 1:1:1, a small peak corresponding to the total mass of λB+PU.1+FH1024 appeared (Fig. 4.7b), which is indicative of the presence of the trimer. The fact that no free PU.1 or free DNA showed up in the spectrum confirmed our other observations and indicated that FH1024 binding did not force PU.1 from the DNA. Furthermore, as the concentration of FH1024 increased, more trimer formed, as indicated by the increase in MS signal intensity of the trimer peak, while no free PU.1 or DNA was released from the complex (Fig. 4.7c-d). These results clearly demonstrated that FH1024 does not disrupt the interface of PU.1-DNA but forms λB-PU.1-FH1024 trimer when added to the protein-DNA complex.

To compare the results of FH1024 with a PU.1-DNA inhibitor, samples consisting of λB, PU.1, and FH1028 were also prepared and examined through MS testing. At the ratio λB: PU.1: FH1028 = 1:1:1, both free PU.1 and PA-bound DNA were detected due to the strong inhibition efficacy of FH1028 (IC_{50}=1.2 nM, Fig. 4.7e). Because FH1028 is a very strong DNA binder (K_D=0.16 nM), all free DNA was occupied by FH1028 as the DNA-FH1028 complex shown in the mass spectrum. At ratio λB: PU.1: FH1028 = 1:1:2, free PU.1 and the DNA-FH1028 complex were obtained. The small peak corresponding to the total mass of λB+PU.1+FH1028 could be the result of non-specific binding of FH1028, since the concentration used in MS (10 μM) is much

higher than that in SPR (nM), 63,000 times the IC_{50}, and PU.1 only occupies around 10-16 bp along the 25 bp λB. For a strong binder like FH1028, at such high concentration and with the availability of extra binding sites, non-specific binding is almost inevitable. Thus, with a combination of biophysical and analytical methods, we have proved that FH1024 is exceptional among the systematically designed PA analogs as it does not disrupt PU.1-DNA binding but is able to bind to DNA-PU.1 dimer complex in a manner that favors the formation of a DNA-PU.1-FH1024 trimer complex.

4.4.5 FH1024 binds to the same DNA site as other PA analogs

We established previously that our systematically designed, inhibitory PAs bind to DNA at the same binding site and there is a positive linear correlation between their binding affinity and inhibition efficacy on PU.1. Given that FH1024 enhances PU.1 binding and binds to DNA simultaneously with PU.1, whether FH1024 kept the same DNA-binding site as the inhibitory PAs became a critical question to address. To probe this question, two competition-binding SPR experiments were conducted. The samples prepared for these two experiments were: (1) a mixture of 50 nM PU.1, 10 nM FH1028 and increasing concentrations of FH1024, and (2) 50 nM PU.1, 50 nM FH1028 and increasing concentrations of FH1024. Since FH1028 is a strong binder (ten times stronger than FH1024) and an effective inhibitor, FH1024 at low concentration would not be able to bind to DNA in the presence of excess FH1028, whereas high concentrations of FH1024 would have the capacity to overcome the inhibitory effect of FH1028. The sensorgrams of the two competition assays are shown in Fig. 8. As expected, in the presence of constant 10 nM FH1028, it takes at least 50 nM FH1024 to start binding to DNA. Meanwhile, for DNA pre-occupied by 50 nM FH1028, more than 500 nM FH1024 is needed to overcome the inhibitory effect of FH1028. These results suggest the direct competition of FH1024 with FH1028 at the same DNA binding

site. This is expected since all of the PAs in Fig. 4.1 were designed to bind to the same site of the λB promoter DNA. Importantly, as shown on the sensorgrams, when FH1024 started binding to DNA, the RU began rising, meaning the molecule has the power, at high concentration, to reverse the strong inhibitory effect caused FH1028 and coexist with PU.1 on the DNA.

4.5 Discussion

4.5.1 Molecular basis of PU.1 inhibition by PA

Our ultimate goal is to apply small molecules such as PAs as therapeutic agents. It is, therefore, important to understand the biological/pharmacological activities of PAs to guide future drug design. The inhibitory effects of PAs have been explored in several biological systems (37, 45-47). The conformational change of DNA upon PA binding can be classified into three categories: direct perturbation to the PA binding site in the minor groove, proximal allosteric perturbation in the major groove, and distal allosteric perturbation outside of the PA binding site (45). Good inhibitors of PU.1 are achievable at the 5' side of the PU.1 binding site on λB promoter DNA. Yet no inhibitors at the 3' side were explored and the 3' side has been shown to be critical for PU.1 binding (38). Thus, the current PAs are designed to target the 3' flanking sequence of the PU.1 binding site (Fig. 4.1). The PAs synthesized exhibit DNA binding affinities ranging from 0.16 nM to 79 nM, very competitive to the major groove binding PU.1 (K_D = 7 nM) under the same conditions. As the SPR results show, most of the PAs demonstrated very strong inhibition with nanomolar IC_{50} values, meaning that distal allosteric inhibition of PAs is relevant to the 3' flanking bases of PU.1 binding site.

Previous X-ray structure studies on hairpin PAs binding to the nucleosome core particle (NCP) revealed a large distortion distal to the PA binding sites and long-range perturbations in the structure of the NCP (48, 49). Another such study on the structure of an eight-ring, cyclic PA

bound to DNA demonstrated that the PA causes DNA distortion by widening the minor groove ~4 Å while compressing the major groove, with an additional distortion seen as bend of the helix axis > 18º toward the major groove (45, 50). These types of large distortion in DNA conformation are likely the driving forces that disrupt the interface between PU.1 and its cognate site. Strong inhibitors such as FH1028 and KA2115 are able to anchor firmly inside the minor groove and induce a long-lasting (low dissociation rate, long residence time) DNA conformation change, whereas weak inhibitors such as KJK6162 and FH1026 have very fast on and off rates, making them uncompetitive with PU.1 for the DNA binding site. The fact that binding affinities of PAs are positively, linearly correlated with the inhibition efficacies leads us to suggest that the inhibitory PAs all bind at the same target site, the site they were designed for, and further indicates that all PAs bind in the same mode.

4.5.2 Molecular basis of FH1024-PU.1-DNA trimer formation

The interaction between PU.1 and DNA is greatly dependent upon DNA micro structure. Observation of the FH1024-PU.1-DNA trimer indicates that the FH1024 complex is different from the other PAs in this study. Among those, KA2040 and FH128 are close analogs to FH1024, both of which are strong binder and inhibitors. FH1024 presents a lower binding affinity than its close analogs (K_D =1.65 nM) primarily, we suspect, due to its high flexibility in the β/β composition which renders the inner structure more adaptable to a range of structures than its analogs. In addition, we must consider the electrostatic interactions generated by both the highly positively charged N- and C-termini and the bulk of the N-terminal tetramethylguanidinium found in FH1024, all of which may make the overall structure less able to form stable complexes with the target sequence. Though constrained by the hairpin linker, the two strands of FH1024 on the open side (the non-hairpin side) might well not be stacked. In a forward binding mode (where the N-

terminal to C-terminal PA vector aligns with 5'-3' DNA vector, Fig. 4.9), when buried in the minor groove, the positively charged, bulky TMG clashes not only with the minor groove wall, but it also reaches the exocyclic amine of guanine (Fig. 4.9) within the PU.1 binding site. Thus TMG is possibly exposed to repulsive electrostatic forces from both G (from DNA) and the C-terminal Ta tail. It is noteworthy that both the N-terminal TMG and C-terminal Ta tail extend into the minor groove of the central PU.1 binding site and are in close proximity to the middle bases (Fig. 4.9). For a well-stacked hairpin PA like FH1028, such binding forces a further widening of the minor groove due to the plasticity of DNA, making PU.1 binding unfavorable. For FH1024, which has flexible N- and C-termini that cannot fit into the minor groove in a well-stacked manner, in the presence of PU.1, the upper (N-terminal) strand of bound FH1024 might float out of the minor groove or even flip away from the PU.1 binding site, leaving the lower (C-terminal) strand inserted in the minor groove. Therefore, the binding of FH1024 may not cause as significant a structural change as the inhibitory PAs, preventing it from disrupting the binding of PU.1 and DNA.

It was reported previously (51) that the Dp tail of an eight-ring PA, when positioned in proximity to the cognate site of the transcription factor extradenticle (Exd), would facilitate Exd-DNA binding, possibly because the charge and steric repulsion between Dp and exocyclic amine of guanine favor displacement of the bases toward Exd in the major groove (51). In our case, the Ta tail on the lower strand of FH1024 is placed in close proximity to the 5'-GGA-3' bases of PU.1 binding site, making similar contacts to the previously reported Dp with the minor groove floor. Thus, the binding of FH1024 does not disrupt the PU.1-DNA complex, but promotes binding of PU.1, owing in part to electrostatic and steric interactions between Ta and the exocyclic amine of the guanine (Fig. 4.9) in the minor groove.

4.6 Conclusion

Many diseases are found to be likely caused by the aberrant expression of certain genes. Abnormal accumulation and/or depletion of transcription factors that are vital in regulating gene expression could be the key step in contracting such diseases. A breakthrough to probe this issue is the use of DNA-binding small molecules to modulate once-undruggable transcription factors. PAs have been studied for this purpose for many years and most of the PAs, if capable of affecting protein-DNA binding, are inhibitors unless conjugated with specific recruiting peptides or other ligands. Molecules that are small in size are generally more soluble, efficient in permeating cells and accessing chromosomes than large molecules. Smaller molecules are also economically more efficient in that they are easier to synthesize. In this work, we have reported the inhibitory effect on transcription factor PU.1-DNA binding of nine out of ten of the designed hairpin PAs. We have found that one of the PAs, FH1024, is not an inhibitor and appears to be among the first unmodified PAs to have such stimulatory effect. The molecule FH1024 binds to DNA simultaneously and potentially cooperatively with PU.1. The interactions likely require a particular binding mode of PA and appropriate register with the PU.1 binding site. More structural and computational studies are in need to better understand the molecular basis of these interactions, thus providing guidance for improving and optimizing the design of additional such molecules.

4.7 Acknowledgment

This work was supported by National Institutes of Health [GM111749 to W. D. W.]; National Science Foundation [MCB 1545160 to G.M.K.P. and W.D.W.]; Missouri Research Board and the UMSL College of Arts and Sciences to J. K. B.

4.8 References

1 P. B. Dervan. Molecular recognition of DNA by small molecules. *Bioorg. Med. Chem.*, **2001**, *9*, 2215-2235.

2 P. B. Dervan, and B. S. Edelson. Recognition of the DNA minor groove by pyrrole-imidazole polyamides. *Curr Opin Struct Biol.*, **2003**, *13*, 284-299.

3 M. Mrksich, M. E. Parks and P. B. Dervan. 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.

4 M. Coll, C. A. Frederick, A. H. Wang, and A. Rich. A bifurcated hydrogen-bonded conformation in the d(A.T) base pairs of the DNA dodecamer d(CGCAAATTTGCG) and its complex with distamycin. *Proc. Natl. Aca. Sci. U. S. A.*, **1987**, *84*, 8385-8389.

5 J. W. Trauger, E. E. Baird, and P. B. Dervan. Recognition of DNA by designed ligands at subnanomolar concentrations. *Nature.*, **1996**. *382*, 559-561.

6 S. White, J. W. Szewczyk, J. M. Turner, E. E. Baird, and P. B. Dervan. Recognition of the four Watson-Crick base pairs in the DNA minor groove by synthetic ligands. *Nature.*, **1998**. *391*, 468-471.

7 C. L. Kielkopf, E. E. Baird, P. B. Dervan and D. C. Rees. Structural basis for G.C recognition in the DNA minor groove. *Nat Struct Biol.*, **1998,** *5*, 104-109.

8 C. L. Kielkopf, S. White, J. W. Szewczyk, J. M. Turner, E. E. Baird, P. B. Dervan and D. C. Rees. A structural basis for recognition of A.T and T.A base pairs in the minor groove of B-DNA. *Science.*, **1998,** *282*, 111-115.

9 S. E. Swalley, E. E. Baird, and P. B. Dervan. Discrimination of 5'-GGGG-3',5'-GCGC-3', and 5'-GGCC-3' sequences in the minor groove of DNA by eight-ring hairpin polyamides. *J. Am. Chem. Soc.*, **1997**, *119*, 6953-6961.

10 J. M. Turner, E. E. Baird, and P. B. Dervan. Recognition of seven base pair sequences in the minor groove of DNA by ten-ring pyrrole-imidazole polyamide hairpins. *J. Am. Chem. Soc.*, **1997**, *119*, 7636-7644.

11 C. F. Hsu, J. W. Phillips, J. W. Trauger, M. E. Farkas, J. M. Belitsky, A. Heckel, B. Z. Olenyuk, J. W. Puckett, C. C Wang, and P. B. Dervan. Completion of a Programmable DNA-Binding Small Molecule Library. *Tetrahedron.*, **2007,** *63*, 6146-6151.

12 T. P. Best, B. S. Edelson, N. G. Nickols, and P. B. Dervan. Nuclear localization of pyrrole-imidazole polyamide-fluorescein conjugates in cell culture. *Proc. Natl. Aca. Sci. U. S. A.*, **2003**, *100*, 12063-12068.

13 B. S. Edelson, T. P. Best, B. Olenyuk, N. G. Nickols, R. M. Doss, S. Foister, A. Heckel, and P. B. Dervan. Influence of structural variation on nuclear localization of DNA-binding polyamide-fluorophore conjugates. *Nucleic Acids Res.*, **2004,** *32*, 2802-2818.

14 N. G., Nickols, C. S. Jacobs, M. E. Farkas, and P. B. Dervan. Improved nuclear localization of DNA-binding polyamides. *Nucleic Acids Res.,* **2007**, *35*, 363-370.

15 K. S. Crowley, D. P. Phillion, S. S. Woodard, B. A. Schweitzer, M. Singh, H. Shabany, B. Burnette, P. Hippenmeyer, M. Heitmeier, and J. K. Bashkin. Controlling the intracellular localization of fluorescent polyamide analogues in cultured cells. *Bioorg Med Chem Lett.*, **2003**, *13*, 1565-1570.

16 L. A. Dickinson, R. J. Gulizia, J. W. Trauger, E. E. Baird, D. E. Mosier, J. M. Gottesfeld, and P. B. Dervan. Inhibition of RNA polymerase II transcription in human cells by synthetic DNA-binding ligands. *Proc. Natl. Aca. Sci. U. S. A.,* **1998**, *95,* 12890-12895.

17 T. D. Schaal, W. G. Mallet, D. L. McMinn, N. V. Nguyen, M. M. Sopko, S. John, and B. S. Parekh. Inhibition of human papilloma virus E2 DNA binding protein by covalently linked polyamides. *Nucleic Acids Res.*, **2003**, *31,* 1282-1291.

18 J. K. Bashkin, K. Aston, J. P. Ramos, K. J. Koeller, R. Nanjunda, G. He, C. M. Dupureur, and W. D. Wilson. Promoter scanning of the human COX-2 gene with 8-ring polyamides: unexpected weakening of polyamide-DNA binding and selectivity by replacing an internal N-Me-pyrrole with beta-alanine. *Biochimie.*, **2013**, *95*, 271-279.

19 H. Ghanim, P. Mohanty, R. Deopurkar, C. L. Sia, K. Korzeniewski, S. Abuaysheh, A. Chaudhuri, and P. Dandona. Acute modulation of toll-like receptors by insulin. *Diabetes care.*, **2008**, *31,* 1827-1831.

20 S. G. Gregory, S. Schmidt, P. Seth, J. R. Oksenberg, J. Hart, A. Prokop, S. J. Caillier, M. Ban, A. Goris, L. F. Barcellos, R. Lincoln, J. L. McCauley, S. J. Sawcer, D. A. Compston, B. Dubois, S. L. Hauser, M. A. Garcia-Blanco, M. A. Pericak-Vance, J. L. Haines, and Multiple Sclerosis Genetics Group. Interleukin 7 receptor alpha chain (IL7R) shows allelic and functional association with multiple sclerosis. *Nat Genet.*, **2007**, *39,* 1083-1091.

21 R. P. DeKoter, H. J. Lee, and H. Singh. PU.1 regulates expression of the interleukin-7 receptor in lymphoid progenitors. *Immunity.*, **2002**, *16,* 297-309.

22 C. Penaranda, W. Kuswanto, J. Hofmann, R. Kenefeck, P. Narendran, L. S. K. Walker, J. A. Bluestone, A. K. Abbas, and H. Dooms. IL-7 receptor blockade reverses autoimmune diabetes

by promoting inhibition of effector/memory T cells. *Proc. Natl. Aca. Sci. U. S. A.*, **2012**, *109*, 12668-12673.

23 K. Ohbo, N. Takasawa, N. Ishii, N. Tanaka, M. Nakamura, and K. Sugamura. Functional analysis of the human interleukin 2 receptor gamma chain gene promoter. *J Biol Chem.*, **1995**, *270*, 7479-7486.

24 E. W. Scott, R. C. Fisher, M. C. Olson, E. W. Kehrli, M. C. Simon, and H. Singh. PU.1 functions in a cell-autonomous manner to control the differentiation of multipotential lymphoid-myeloid progenitors. *Immunity.*, **1997**, *6*, 437-447.

25 F. Rosenbauer, K. Wagner, J. L. Kutok, H. Iwasaki, M. M. Le Beau, Y. Okuno, K. Akashi, S. Fiering, and D. G. Tenen. Acute myeloid leukemia induced by graded reduction of a lineage-specific transcription factor, PU.1. *Nat Genet.*, **2004**, *36*, 624-630.

26 E. W. Scott, M. C. Simon, J. Anastasi, and H. Singh. Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. *Science.*, **1994**, *265*, 1573-1577.

27 D. L. Stirewalt. Fine-tuning PU.1. *Nat Genet.*, **2004**, *36*, 550-551.

28 R. Kodandapani, F. Pio, C. Z. Ni, G. Piccialli, M. Klemsz, S. McKercher, R. A. Maki, and K. R. Ely. A new pattern for helix-turn-helix recognition revealed by the PU.1 ETS-domain-DNA complex. *Nature.*, **1996**, *380*, 456-460.

29 M. Munde, G. M. Poon, and W. D. Wilson. Probing the electrostatics and pharmacological modulation of sequence-specific binding by the DNA-binding domain of the ETS family transcription factor PU.1: a binding affinity and kinetics investigation. *J Mol Biol.*, **2013**, *425*, 1655-1669.

30 M. Munde, S. Wang, A. Kumar, C. E. Stephens, A. A. Farahat, D. W. Boykin, W. D. Wilson and G. M. Poon. Structure-dependent inhibition of the ETS-family transcription factor PU.1 by novel heterocyclic diamidines. *Nucleic Acids Res.*, **2014,** *42*, 1379-1390.

31 A. K. Mapp, A. Z. Ansari, M. Ptashne, and P. B. Dervan. Activation of gene expression by small molecule transcription factors. *Proc. Natl. Aca. Sci. U. S. A.*, **2000**, *97*, 3930-3935.

32 A. Z. Ansari, A. K. Mapp, D. H. Nuguyen, P. B. Dervan and M. Ptashne. Towards a minimal motif for artificial transcriptional activators. *Chem Biol*., **2001**, *8*, 583-592.

33 X. Xiao, P. Yu, H. S. Lim, D. Sikder, and T. Kodadek. A cell-permeable synthetic transcription factor mimic. *Angew Chem Int Ed Engl.*, **2007,** *46*, 2865-2868.

34 X. Xiao, P. Yu, H. S. Lim, D. Sikder, and T. Kodadek. Design and synthesis of a cell-permeable synthetic transcription factor mimic. *J Comb Chem.*, **2007,** *9*, 592-600.

35 J. Syed, A. Chandran, G. N. Pandian, J. Taniguchi, S. Sato, K. Hashiya, G. Kashiwazaki, T. Bando, and H. Sugiyama. A Synthetic Transcriptional Activator of Genes Associated with the Retina in Human Dermal Fibroblasts. *Chembiochem.*, **2015,** *16*, 1497-1501.

36 Y. Wei, G. N. Pandian, T. Zou, J. Taniguchi, S. Sato, G. Kashiwazaki, T. Vaijayanthi, T. Hidaka, T. Bando, and H. Sugiyama. A Multi-target Small Molecule for Targeted Transcriptional Activation of Therapeutically Significant Nervous System Genes. *ChemistryOpen.*, **2016,** *5*, 517-521.

37 Y. Kawamoto, T. Bando, and H. Sugiyama, Sequence-specific DNA binding Pyrrole-imidazole polyamides and their applications. *Bioorg Med Chem.*, **2018,** *26*, 1393-1411.

38 G. M. Poon, and R. B. Macgregor, Jr., Base coupling in sequence-specific site recognition by the ETS domain of murine PU.1. *J Mol Biol.*, **2003,** *328*, 805-819.

39 B. Liu, S. Wang, K. Aston, K. J. Koeller, S. F. H. Kermani, C. H. Castaneda, M. J. Scuderi, R. Luo, J. K. Bashkin, and W. D. Wilson. beta-Alanine and N-terminal cationic substituents affect polyamide-DNA binding. *Org Biomol Chem.*, **2017,** *15,* 9880-9888.

40 G.M.K. Poon. DNA Binding Regulates the Self-Association of the ETS Domain of PU.1 in a Sequence-Dependent Manner. *Biochemistry.*, **2012,** *51,* 4096-4107.

41 G.M.K. Poon. Sequence Discrimination by DNA-binding Domain of ETS Family Transcription Factor PU.1 Is Linked to Specific Hydration of Protein-DNA Interface. *J Biol Chem.*, **2012,** *287,* 18297-18307.

42 X. L. Yang, R. B. Hubbard, M. Lee, Z. F. Tao, H. Sugiyama, and A. H. Wang. Imidazole-imidazole pair as a minor groove recognition motif for T:G mismatched base pairs. *Nucleic Acids Res.*, **1999,** *27,* 4183-4190.

43 E. R. Lacy, K. K. Cox, W. D. Wilson and M. Lee. Recognition of T*G mismatched base pairs in DNA by stacked imidazole-containing polyamides: surface plasmon resonance and circular dichroism studies. *Nucleic Acids Res.*, **2002,** *30,* 1834-1841.

44 E. Vasilieva, J. Niederschulte, Y. Song, G. D. Harris, Jr., K. J. Koeller, P. Liao, J. K. Bashkin, and C. M. Dupureur. Interactions of two large antiviral polyamides with the long control region of HPV16. *Biochimie.*, **2016,** *127,* 103-114.

45 D. M. Chenoweth, and P. B. Dervan, Allosteric modulation of DNA by small molecules. *Proc. Natl. Aca. Sci. U. S. A.*, **2009,** *106,* 13175-13179.

46 C. L. Wu, W. Wang, L. Fang, and W. Su. Programmable pyrrole-imidazole polyamides: A potent tool for DNA targeting. *Chinese Chemical Letters*, **2018,** *29,* 1105-1112.

47 G. He, E. Vasilieva, G. D. Harris, Jr., K. J. Koeller, J. K. Bashkin, and C. M. Dupureur. Binding studies of a large antiviral polyamide to a natural HPV sequence. *Biochimie.*, **2014,** *102,* 83-91.

48 R. K. Suto, R. S, Edayathumangalam, C. L. White, C. Melander, J. M. Gottesfeld, P. B. Dervan, and K. Luger. Crystal structures of nucleosome core particles in complex with minor groove DNA-binding ligands. *J Mol Biol.*, **2003,** *326,* 371-380.

49 R. S. Edayathumangalam, P. Weyermann, J. M. Gottesfeld, P. B. Dervan, and K. Luger. Molecular recognition of the nucleosomal "supergroove". . *Proc. Natl. Aca. Sci. U. S. A.*, **2004,** *101,* 6864-6869.

50 Y. Wang, N. Ma, and G. Chen, Allosteric analysis of glucocorticoid receptor-DNA interface induced by cyclic Py-Im polyamide: a molecular dynamics simulation study. *PloS one.*, **2012,** *7,* e35159.

51 R., Moretti, L. J. Donato, M. L. Brezinski, R. L. Stafford, H. Hoff, J. S. Thorson, P. B. Dervan and A. Z. Ansari. Targeted chemical wedges reveal the role of allosteric DNA modulation in protein-DNA assembly. *ACS Chem Biol.*, **2008,** *3,* 220-229.

4.9 Tables and Figures

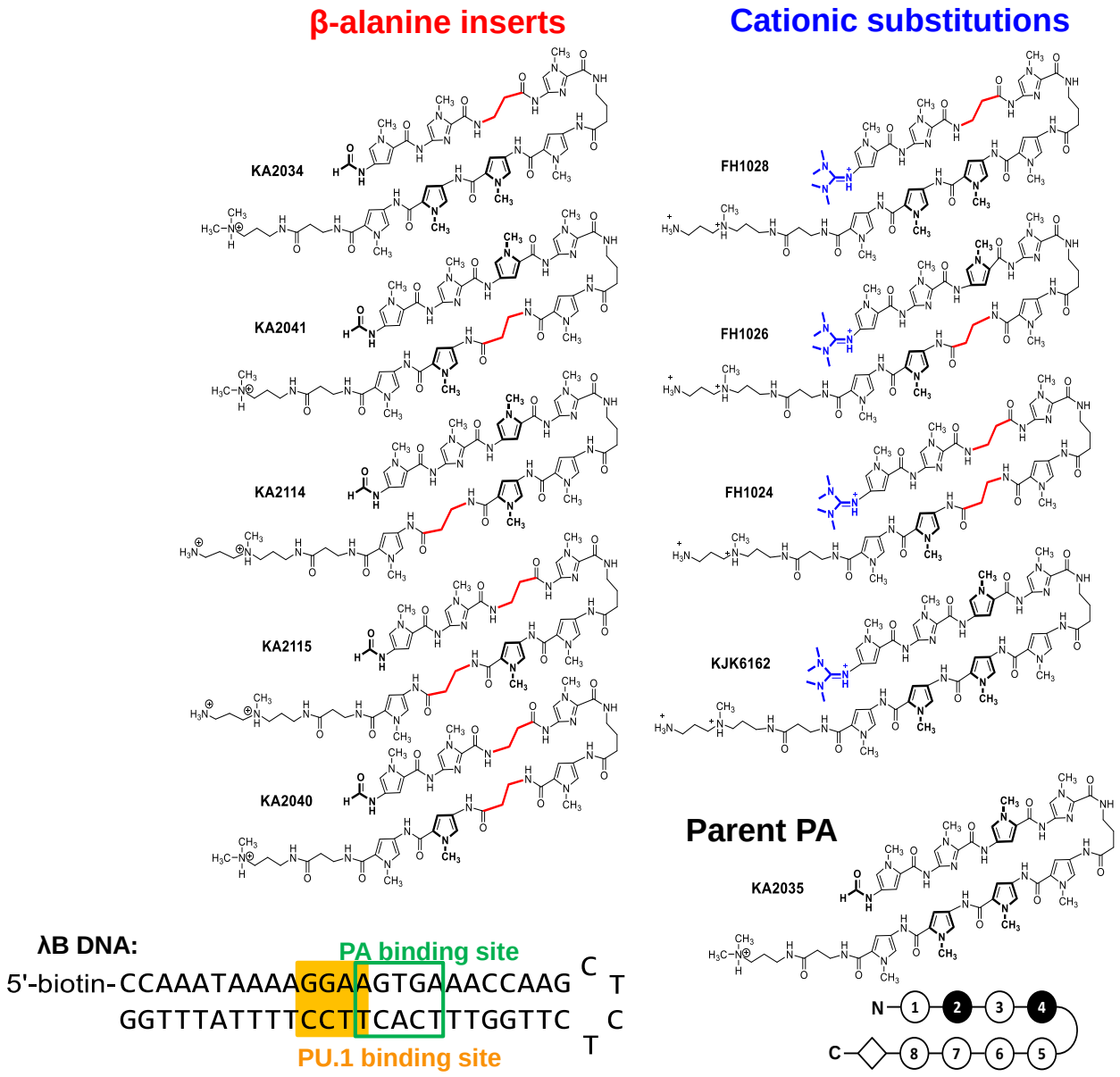


Figure 4.1 Schematic representation of a set of synthesized hairpin polyamides and λB DNA.

The polyamides are classified into two modified groups: β-alanine inserts (red) and cationic substitutions (blue). The simplified representation of parent PA KA2035 is shown below its structure and the numbering scheme applies to all PAs.

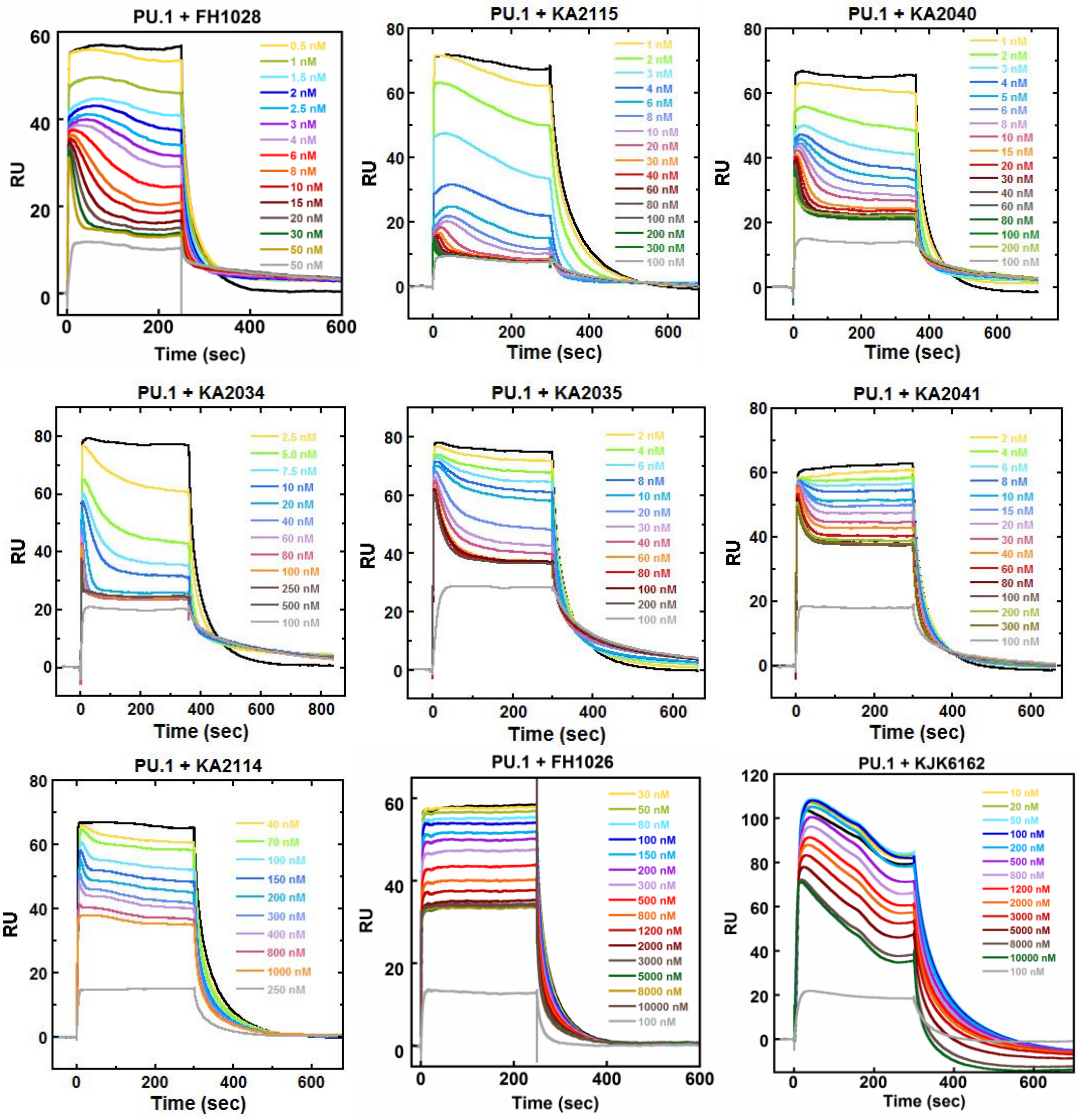


Figure 4.2 SPR sensorgrams of PU.1 and PAs (exclude FH1024) binding to λB DNA.

The black line on the top (highest RU) of each sensorgram represents the injection of 50 nM PU.1. The silver line at the bottom (lowest RU) of each sensorgram represents the binding of PA alone. The legend indicates the concentrations of corresponding PAs. The sensorgrams are arranged in a manner by which inhibition efficacies decrease from left to right and from top to bottom.

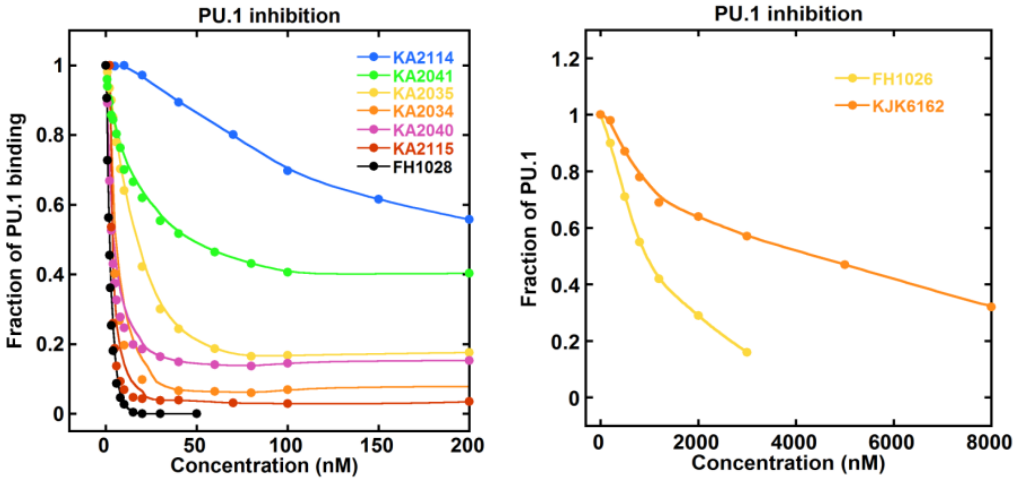


Figure 4.3 Plots of fraction bound for PU.1 vs. polyamide concentration to determine IC_{50} values.

Table 4.1 Comparison, for PAs, of DNA binding affinities vs. PU.1 inhibition efficacies (IC_{50} values) on λB DNA.

	Polyamide	K_D (nM)	IC_{50} (nM)
β-alanine inserts	KA2115	0.49	3.6
	KA2040	0.54	4.2
	KA2034	0.67	4.9
	KA2035	1.30	17.2
	KA2041	1.62	46.4
	KA2114	4.10	271
Cationic substitutions	FH1028	0.16	1.2
	FH1026	35	945
	KJK6162	79	4340

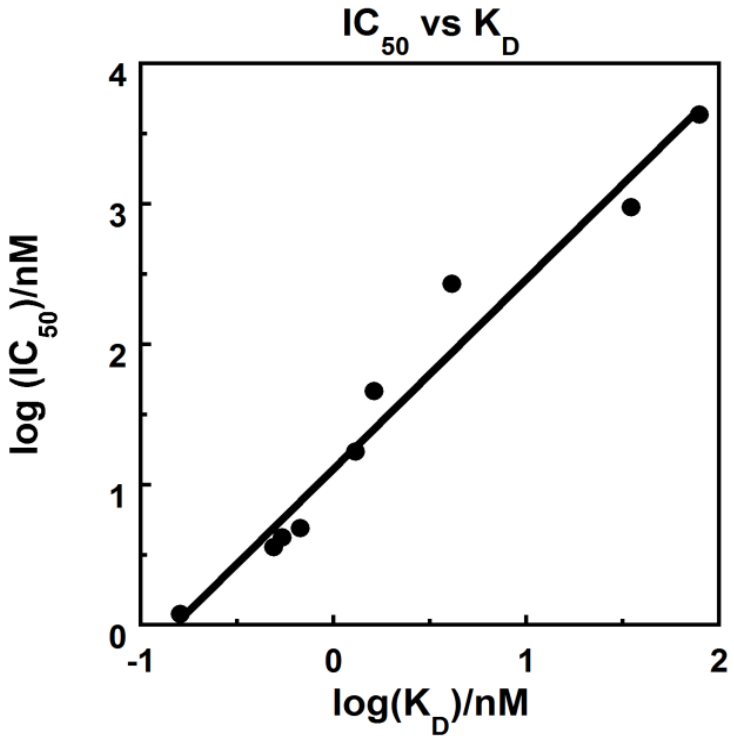


Figure 4.4 Log-log plot establishes a linear correlation between DNA binding affinities of PAs and PU.1 inhibition efficacies (IC_{50} values) on λB DNA.

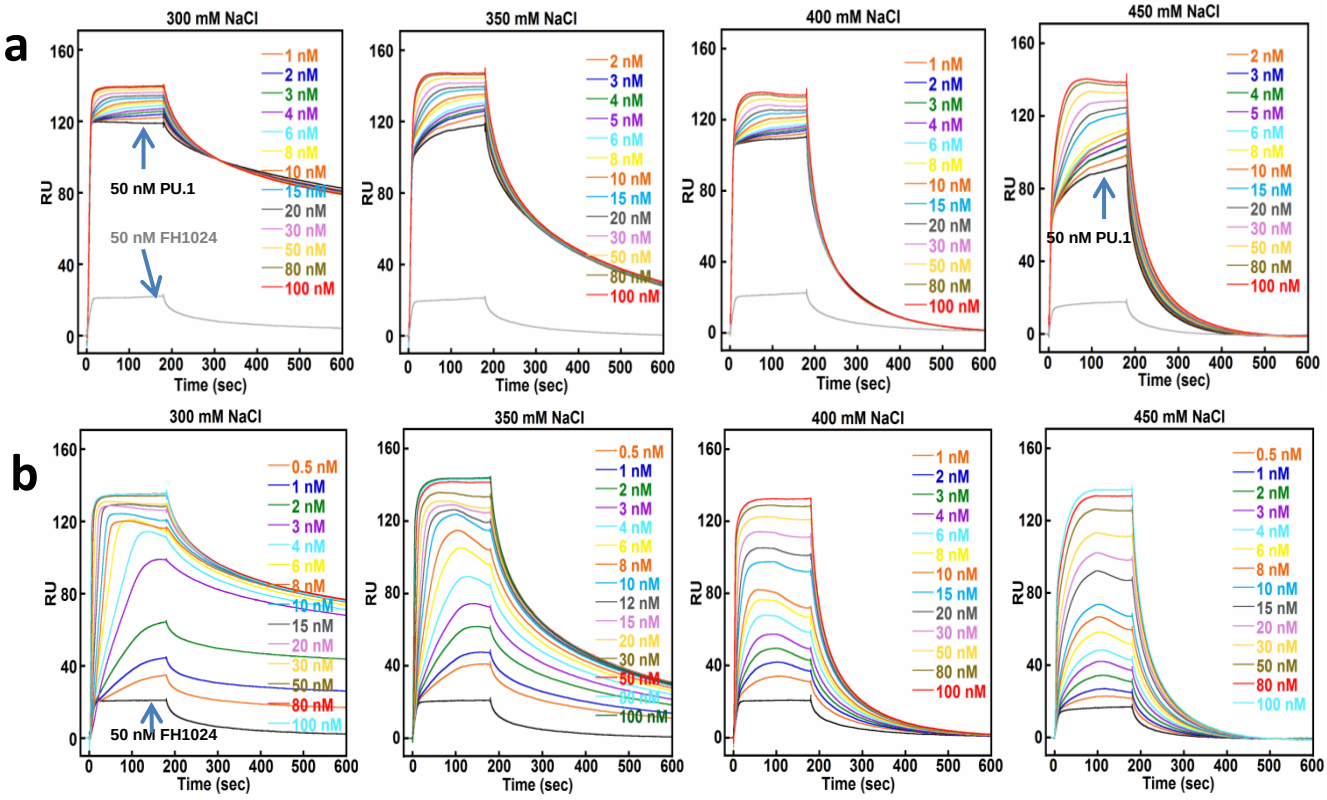


Figure 4.5 Simutaneous binding of FH1024 and PU.1 to λB DNA.

a) Flow of increasing concentration of FH1024 mixed with constant 50 nM PU.1 onto immobilized DNA. The legend indicates the concentrations of FH1024. In each sensorgram of panel for a, the black line (before RU increases) represents binding of 50 nM PU.1 alone; the separated silver line at the bottom results from 50 nM of FH1024 binding alone. b) Flow of increasing concentration of PU.1 mixed with constant 50 nM FH1024 onto immobilized DNA. The legend indicates the concentrations of PU.1. In the sensorgrams of panel b, the black line at the bottom represents binding of 50 nM FH1024 alone.

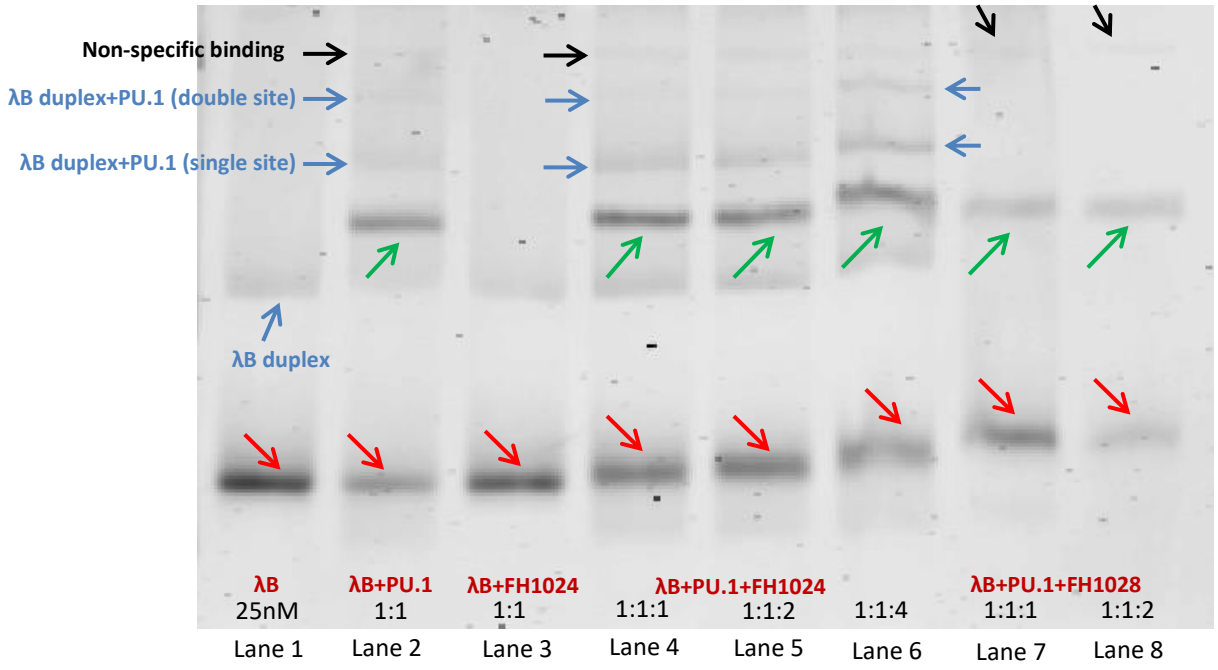


Figure 4.6 Effect of FH1024 on PU.1-DNA binding.

Lanes 4-6 show different ratios of FH1024 added to PU.1-λB DNA complex. Lanes 7-8 are comparisons with strong inhibitor FH1028. Red arrows indicate bands corresponding to λB DNA or λB DNA-PA complex. Green arrows point to bands from λB DNA-PU.1 complex. Blue arrows show the bands of duplex λB DNA-PU.1 complex with PU.1 on one or both binding sites. Black arrow indicates non-specific binding of PU.1 on DNA.

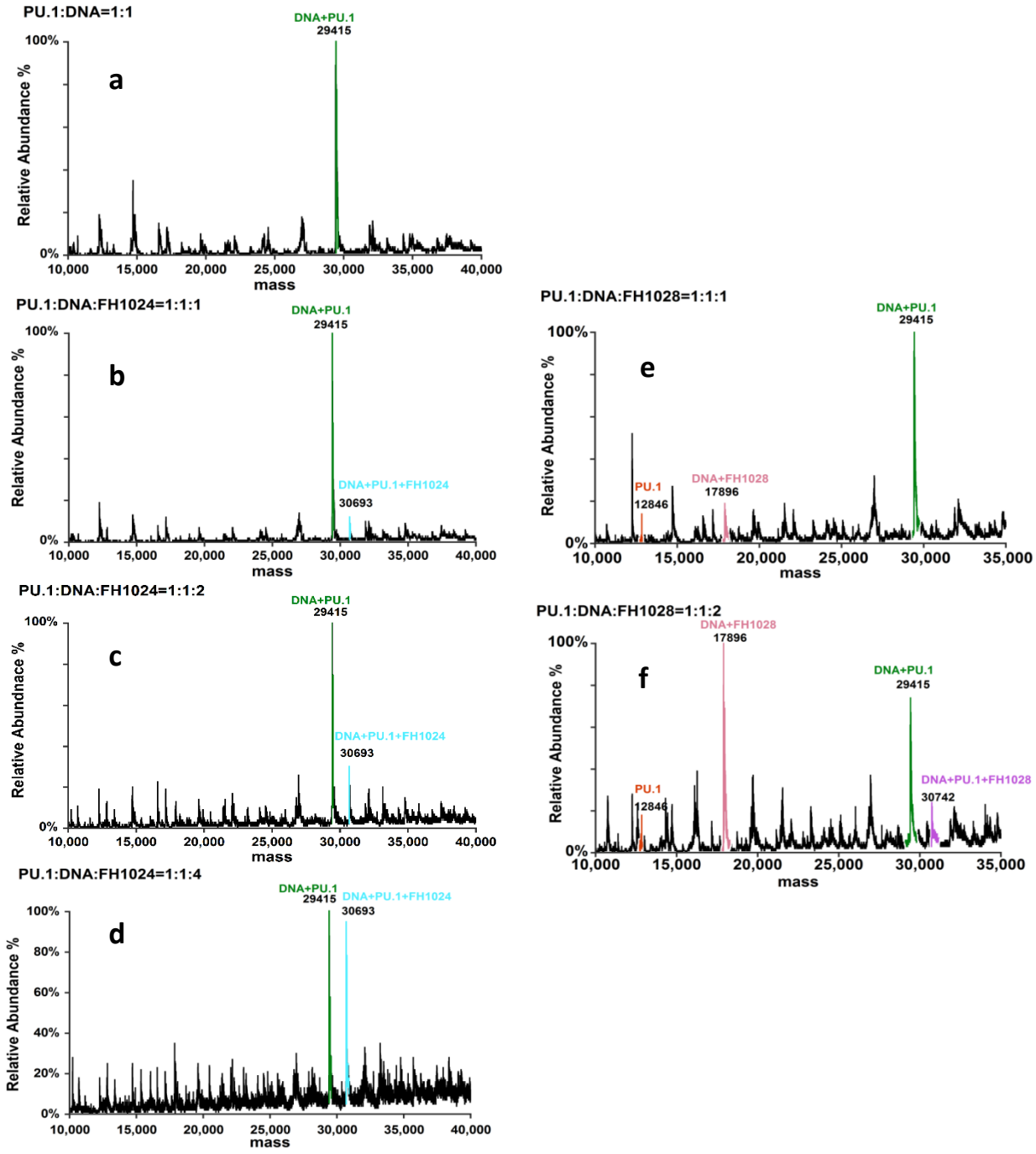


Figure 4.7 ESI-MS showing: (b-d) the formation of the FH1024-DNA-PU.1 trimer and (e-f) the inhibition of the PU.1-DNA complex by FH1028.

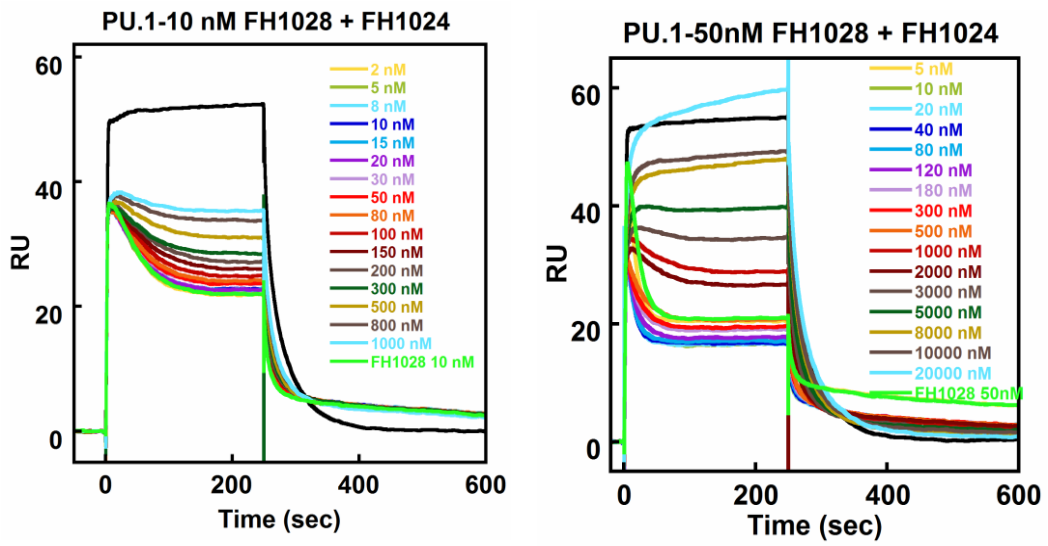


Figure 4.8 Binding by SPR of FH1024 with fixed concentrations of PU.1 and FH1028.

The bright green line represents FH1028. The legend indicates the concentrations of FH1024. In the left panel, FH1028 was kept at 10 nM. In the right panel, the concentration of FH1028 is 50 nM.

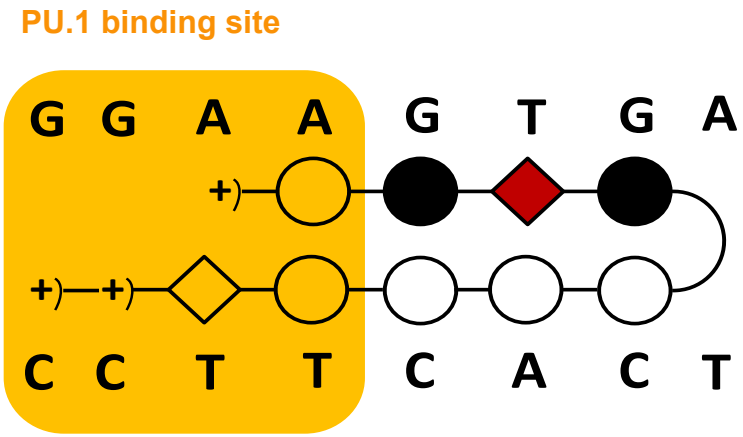


Figure 4.9 Schematic illustration of the relative positions of the cognate sites for PU.1 and FH1028.

5 CONCLUSIONS

This dissertation has investigated the systematic design and synthesis of a set of eight-ring hairpin polyamides as well as a new class of Pyr-AzaHx hybrid polyamides. Their DNA binding affinities and specificities of these molecules were evaluated. The effects of the eight-ring hairpin polyamides on the transcription factor PU.1-DNA interactions were explored. The results altogether have deepened our understanding on the molecular recognition between small molecules and DNA and also broadened our view on the molecular interactions among small molecules, DNA, and PU.1.

More specifically, the use of SPR in determining the DNA binding affinity and specificity of both polyamides and PU.1 with high accuracy has demonstrated that SPR is a well-established method that is applicable to such binding studies. Modifications of eight-ring hairpin polyamides by β-alanine and N-terminal cationic group have diverse effects on the DNA binding properties. Close comparison revealed a pattern of how the number and position of β-alanine affect the binding affinity and specificity. The pattern can be used as reference when predicting the interactions of newly designed molecules.

Design and synthesis of Pyr-AzaHx hybrid polyamides was carried out in order to expand the DNA recognition sites of small, fluorescent molecules to facilitate cell studies. Evaluation of the DNA binding properties of these molecules showed that the modification did not specifically recognize the proposed binding site, but exhibited enhanced binding affinity compared to the last generation of such molecules. Even though more efforts are in need to extend the DNA recognition repertoire, the enhanced binding properties of these modified molecules provide valuable guidance for the design of the next generation hybrid polyamides.

This dissertation investigated the effects of the eight-ring hairpin polyamides on the transcription factor PU.1-DNA interactions to prepare for the biological application studies of polyamides. Most of the polyamides either have no effect or inhibit the binding of PU.1 to DNA. The inhibition efficacy of polyamides was found to be positively correlated to their binding affinity. A potential PU.1 binding enhancer molecule was identified. Enhancing of protein binding is often achieved by attaching enhancing agents to polyamides and is rarely seen in polyamide alone. These encouraging results added to the driving force of our continuing understanding of the molecular mechanisms of action of these interactions.

Polyamides are class of promising candidates for the treatment of cancer, viral, or bacterial infectious diseases. This dissertation undertook the fundamental work of furthering the understanding of polyamide-DNA interactions as well as polyamie-DNA-PU.1 interactions to establish solid basis for the biological and clinical applications of polyamides. Their possession of high DNA binding affinity and specificity has enabled them to efficiently affect gene expression through interfering with transcription factors. In addition, unlike other peptide or chemical drugs, polyamides are resistant to the biological degradation by nucleases. Achieving efficient nuclear localization and low long-time toxicity would be greatly helpful to unleash the potential of polyamides as medicinal therapeutics.