Design of Protein-DNA Dimerizers

Thesis by

Ryan Leonard Stafford

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

California Institute of Technology

Pasadena, California

2008

(Defended August 8, 2007)

for my family

Acknowledgements

It has been a long time since I started working at Caltech, so there are a lot of people to thank. First, I would like to thank my Mom and Dad for giving all of their kids, including me, the freedom to do what we want and support us no matter what. And thanks to my family-in-law, Annie, the Mom and Dad, and of course, my love, Kim. You have given me something else to do so I do not end up working every moment- only every other moment.

At a place like Caltech, it has been no surprise that I have been able to work with such smart people (although it may be surprising how fun they all are). In particular, Hans-Dieter (bear or beer) made a fantastic collaborator and mentor during my first couple years. I enjoyed the opportunity to work closely with Nick Nickols (risk:fun) and Mike Brochu (Palms Thai) my last couple years. Sometime in the middle I mentored Rachel Wang (Jonas) who always found a way to work hard when it mattered most and learned how to keep a notebook. I want to thank Carey Hsu (my old roomie), who showed me how to hustle on a basketball court and all the other core Freeballers: Justin Cohen (the 330 oasis), Dan Harki (outside jumper), Michelle Farkas (take care of my desk), John Phillips (the new Ray Doss), Anand Vedehra (pillow fight), Ray Doss (punched in the back), Mike Kesden (elbows), Tim Best (homestar runner), Adam (Tucker the turtle), and Katherine Poulin-Kerstien (lacrosse)-you guys are awesome. I have to especially thank Adam for guiding me to a good orthopedic surgeon. And I want to thank everyone else who should have played more ball: Jim Sanchez (fellow anteater), Claire Jacobs (Puckett's shoe), Christian Dose (something inappropriate), Jim Puckett (fried food), and Mike Marques (no team sports). And I want to thank all the other

Dervanites of the past and present, that I have not already mentioned, that have made my lab experience more enjoyable including Dave Chenoweth (racquetball), Steve Bartusiak (YouTube), Sherry Tsai (*really?*), Mareike Goeritz (cappuccino machine), Julie Poposki (flowers on campus), Anne Viger (*coffee?*), Ben Edeslon (acrobat slim shady), Shane Foister (Kentucky basketball), Eric Fechter (flying), Alex Heckel (German lessons), and Victor Rucker (footprinting stories). I want to wish the best of luck to Katy Muzikar (purple hair). I hope you, John, and Michelle have fun training all the new Dervan recruits.

And thanks to Peter for providing me with all the resources and real estate to do science. Thanks also to my committee, Bob Grubbs, who has been a pleasure to T.A. for; Doug Rees for being an inspiration and helping to support my future endeavors; and Steve Mayo. And I want to thank Greg Weiss for always supporting me since my days at UCI.

Abstract

Genes are regulated by proteins called transcription factors that bind to DNA in a sequence-specific manner and modulate the rate of transcription. Mutated transcription factors often lead to abnormal gene expression, developmental defects, and disease. This thesis describes the design of chemicals called protein-DNA dimerizers that mimic natural transcription factor protein-DNA complexes. In the long-term, it is hoped that these dimerizers will be able to engage or even replace mutant transcription factors and artificially regulate gene expression in living cells. Specifically, programmable DNA-binding pyrrole-imidazole polyamides conjugated to YPWM peptide motifs incorporating various linker domains facilitate the binding of a natural transcription factor, extradenticle, to DNA. From a design point of view, it has been explored what the minimum size and shape (branched or linear) is that will ultimately be optimal for cell uptake with adequate functional potency in the transcriptional apparatus. Branched dimerizers are shown to function with a minimal WM dipeptide protein-binding domain *in vitro* up to 37 °C, and linear dimerizers are shown to function with WMK tripeptides up to 20 °C. Collectively, branched and linear dimerizers can facilitate protein binding to DNA from 2 base pair overlap sites to ones that reach 6 base pairs apart. Polyamide-WM-fluorescein conjugates are also found to be cell permeable in several cell lines including HeLa, MCF-7, and PC3. These studies provide insight into the importance of linker length and composition, binding-site spacing and orientation, and the protein-binding domain content that are important for the optimization of protein DNA-dimerizers suitable for biological experiments.

Table of Contents

List of Figures and Tables

Chapter 1 Page

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Appendix A

Appendix B

Appendix C

Appendix D

Appendix E

Appendix F

Chapter 1

Introduction to Protein-DNA Dimerizers

1.1 Background and Significance

Since the sequencing of the human genome[1,2] it has been estimated that humans possess 20,000 to 30,000 genes that are expressed in a highly controlled manner by DNA-binding proteins called transcription factors.[3,4] The evolution of more complex organisms is thought to be due not only to gene proliferation, but also increasingly sophisticated transcriptional regulation.[5] This is particularly evident from the observation that prokaryotes often utilize single homodimeric transcription factors to regulate target genes, in contrast to eukaryotes, in which gene expression is frequently controlled in a combinatorial fashion by multiprotein-DNA complexes.[6] Errors in regulating gene expression caused by mutant transcription factors often lead to diseases such as cancer or birth defects, and therapies that counteract these faulty proteins would have a significant impact on human medicine.[7] This thesis describes the design of chemicals called protein-DNA dimerizers that mimic endogenous transcription factors in that they facilitate the formation of protein-DNA complexes. Protein-DNA dimerizers are a promising class of compounds that suggest we may someday engage or replace errant transcription factors and neutralize their deleterious effects for therapeutic purposes.

1.2 Eukaryotic Gene Transcription

The protein-coding regions of the eukaryotic genomes are transcribed by RNA polymerase II. The polymerase physically associates with a number of proteins including TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH before initiating transcription.[8] TFIID, also called TATA-binding protein (TBP), binds to the TATA DNA sequence of the core promoter region adjacent to the start site of genes (Figure 1.1).[3,4,8] Additional transcription factors also bind close to the core promoter and further away at up- and downstream enhancer regions which influence the rate of transcription.[4,8] Large mutiprotein complexes mediate the interactions between the these transcription factors and the RNA polymerase II machinery[9] and a unified nomenclature for such mediator proteins has recently been

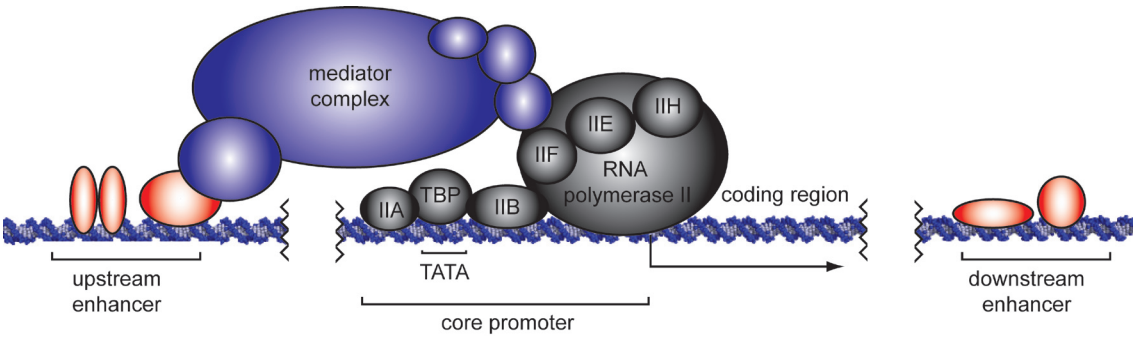


Figure 1.1 A schematic of a gene promoter. The RNA polymerase II transcriptional complex is shown bound at the gene start site near the core promoter. Upstream and downstream enhancer elements are shown to be much farther away from the start site of the gene. Enhancers contain binding sites for transcription factors which often interact indirectly with the RNA polymerase II complex through large multiprotein mediator complexes.

proposed.[10] Gene expression is also affected by the local chromatin structure which is regulated by proteins such as histone-modifying enzymes.[11]

1.3 DNA Recognition by Natural Transcription Factors

Central to the regulation of gene transcription is the sequence specific recognition of DNA by transcription factors. DNA contains a wide major groove and a narrow minor groove that expose unique surfaces for interaction with these proteins (Figure 1.2). Transcription factors are capable of binding to both the major and minor grooves by making defined electrostatic, hydrogen-bonding, and van der Waals contacts with the composite surface of nitrogenous bases and sugar-phosphate backbone of DNA (Figure 1.3). Some proteins, such as TBP[12-15] and the zinc-finger Zif268,[16] bind to DNA as monomers, but transcription factors often bind as part of homo- and heterodimeric complexes. For instance, the leucine-zipper homodimer of GCN4[17] or the heterodimer of Jun-Fos[18] interact through their two coiled-coil domains that position their α-helical DNA-binding domains into the major groove. These protein-DNA complexes are often reiterated in a combinatorial fashion which enables the integration of diverse cellular signals to regulate individual genes. This combinatorial mechanism is exemplified by the interferon-β (IFN-β) enhanceosome in which a total of

eight proteins, including ATF-2/c-Jun, IRF-3/IRF-7, and NF-κB, are required to bind to a conserved stretch of 55 base pairs of DNA in order to activate expression in response to viral infection.[19] Recently, an atomic model of the complete IFN-β enhanceosome has been assembled from three separate x-ray crystal structures (Figure 1.4).[19]

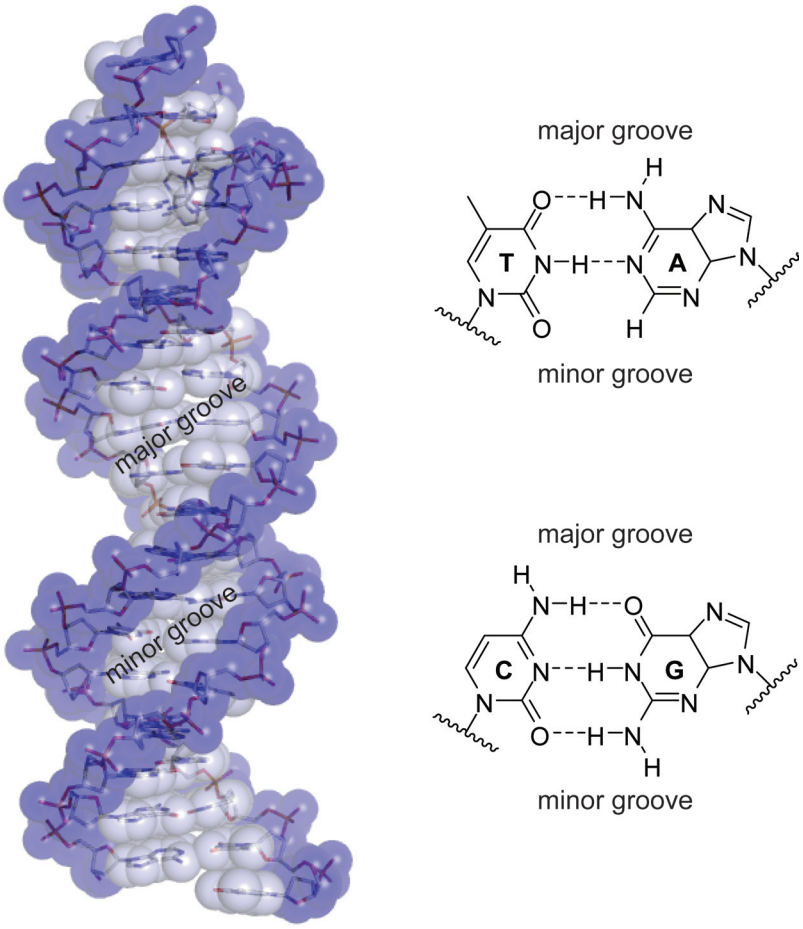


Figure 1.2 The structure of DNA. On the left a crystal structure of a DNA double helix is shown (PDB code 1YSA). On the right, the four nitrogenous bases are shown as Watson-Crick base pairs where A complexes with T and G with C. The exposed edges of the base pairs in the major and minor grooves are indicated which are important for the sequence-specific recognition of DNA by proteins and small molecules.

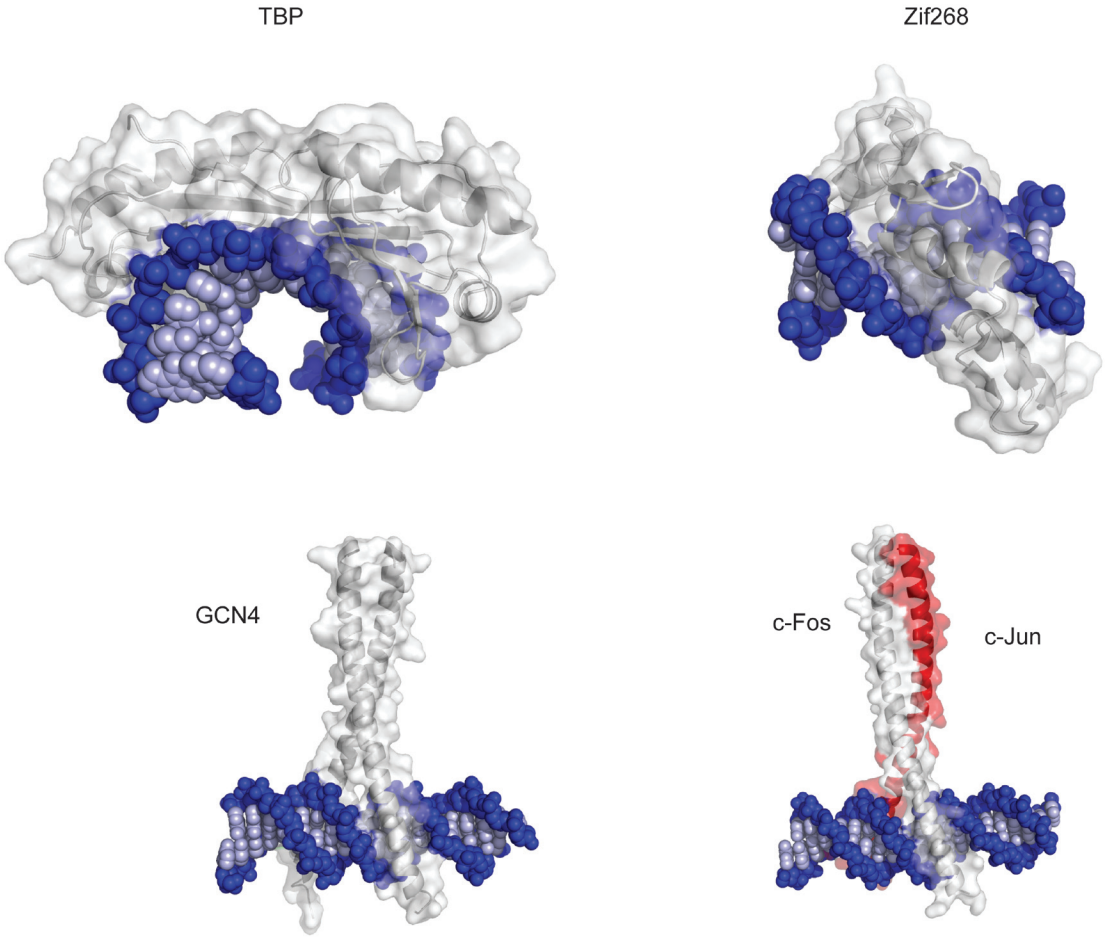


Figure 1.3 Crystal structures of four transcription factor-DNA complexes. The top left shows a structure of TATA-binding protein (TBP) bound to the DNA minor groove as a monomer leading to significant bending and widening of the groove (PDB code 1TGH). The top right shows a structure of zinc-finger protein, Zif268, which consists of three tandem major-groove recognition domains (PDB code 1ZAA). The bottom left shows a GCN4 leucine-zipper homodimer, which inserts its extended α-helical coiled-coil domains into the DNA major grooves (PDB code 1YSA). The bottom right shows a c-Jun/c-Fos heterodimer, which binds to DNA in a similar manner to the GCN4 homodimer (PDB code 1FOS).

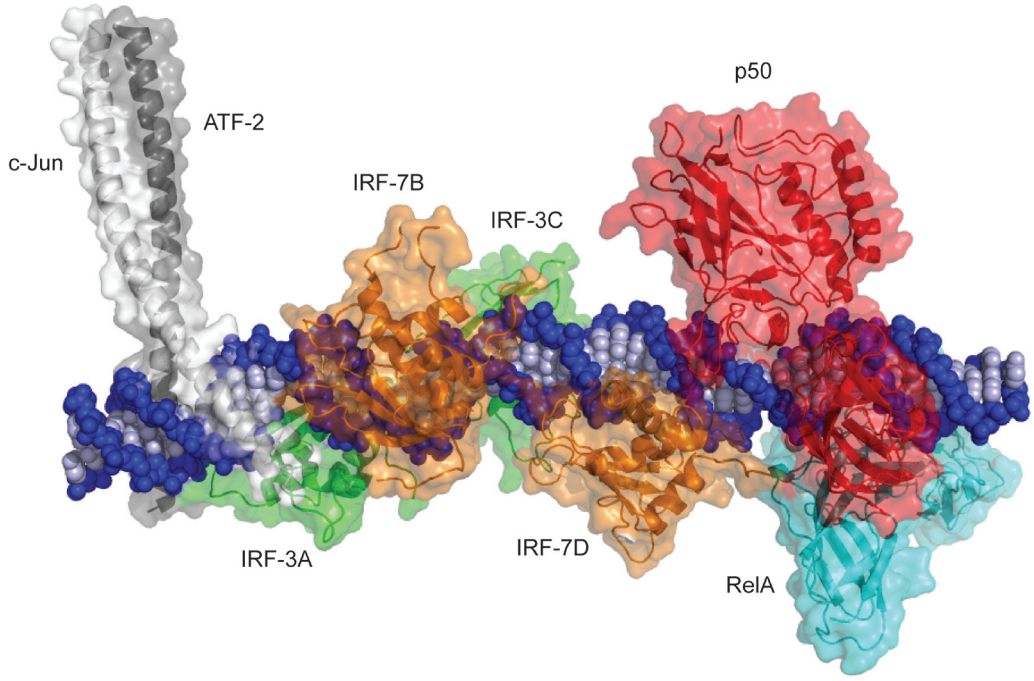


Figure 1.4 An atomic model of the IFN-β enhanceosome in which a total of eight transcription factors bind to a highly conserved 55 base pair segment of DNA (PDB from supplemental of reference 19). All eight transcription factors must bind to the enhanceosome to activate IFN-β expression in response to viral infection.

1.4 HOX Transcription Factors

Homeobox (HOX) proteins are a particular class of transcription factors central to this thesis that are critical for the proper execution of developmental programs of higher organisms.[20-22] HOX genes have been intensely studied in the fruit fly *D. melanogaster* in which they were identified by the abnormal body segment phenotypes caused by mutations to single genes.[20-22] For instance, a mutation to the gene ultrabithorax (Ubx) leads to the development of an additional thorax in place of the fly's balance-sensing organs called halteres. In humans, mutations to HOX proteins are implicated in several forms of cancer and limb deformation diseases.[21,23-26] Thus, HOX transcription factors serve as master regulators of the downstream genes that must be activated or deactivated at precise times during the course of organism development.

In contrast to their specific roles during development, HOX proteins only bind to relatively nonspecific 4 to 6 base pair sequences of DNA.[27-29] By binding to DNA with

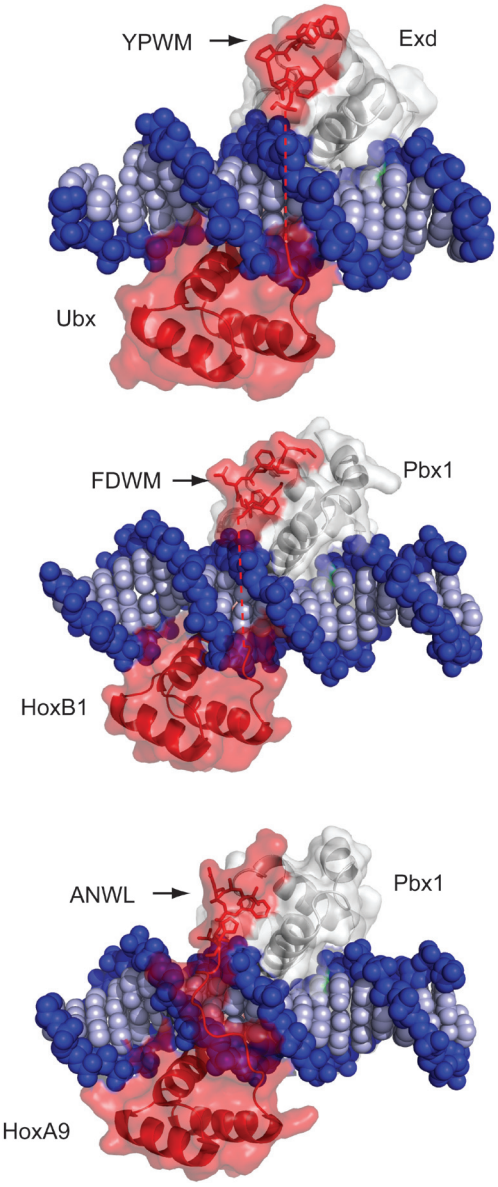


Figure 1.5 Crystal structures of three HOX/TALE/DNA ternary complexes. The top shows the structure of the *D. Melanogaster* developmental transcription factors ultrabithorax (Ubx) and extradenticle (Exd) (PDB code 1B8I). The proteins bind to opposite sides of the DNA duplex primarily in the major groove with their 3 α-helix homeodomains. The conserved YPWM peptide motif of Ubx spans the intervening minor groove and mediates a direct protein-protein contact with an Exd binding pocket between the 1st and 2nd α-helices. The middle structure shows the homologous human transcription factors HoxB1 and Pbx1 in which the FDWM peptide of HoxB1 binds to Pbx1 in a similar manner as the YPWM of the fruit fly proteins (PDB code 1B72). A 4th α-helix was discovered in the Pbx1 structure that stabilizes homeodomain-DNA binding. The bottom structure shows the human proteins HoxA9 and Pbx1 bound to DNA (PDB code 1PUF). The divergent ANWL peptide of HoxA9 contains the conserved W that dominates the protein-protein interface. Unique to this structure, the linking amino acids between the ANWL peptide and the homeodomain are ordered and contact the intervening DNA minor groove.

cofactor proteins, they increase their composite recognition site which is thought to be important for their function in developmental pathways.[27] For instance, Ubx binds the site 5′-TTAT-3′ alone, but in complex with extradenticle (Exd) the composite site increases by four base pairs to 5′-TGATTTAT-3′.[30] A crystal structure of the ternary Ubx/Exd/DNA complex revealed how both Ubx and Exd's major groove binding homeodomains interact via Ubx's N-terminal YPWM peptide motif (Figure 1.5).[30] In fact, all HOX proteins contain a conserved homeodomain which adopts a three α-helix bundle that binds to the DNA major groove. The third α-helix, aptly named the recognition helix, makes the majority of the specific DNA contacts with the edges of nitrogenous bases. The YPWM motif of Ubx spans the DNA minor groove between the two proteins and inserts a hydrophobic tryptophan into the complementary Exd binding pocket. Additional crystal structures of HoxB1/Pbx/DNA and HoxA9/Pbx/DNA which contain human homologues of Ubx and Exd have also been reported in which the conserved tryptophan of the YPWM motif mediates the majority of the protein-protein contacts.[31,32] The YPWM motif is sometimes referred to as a penta- or hexapeptide since two flanking amino acids are also highly conserved yielding an overall consensus sequence of ϕYPWM(K/R) where ϕ is a hydrophobic residue (Figure 1.6).[33] Both Exd and Pbx, sometimes called TALE proteins, contain a three-amino acid loop extension between the first and second α-helices of their divergent homeodomains which line the binding pocket for the YPWM motif. In fruit flies an Exd mutation leads to the increase in the number of denticles (i.e., extra denticles)[34] and in humans a t(1;19) chromosomal translocation at Pbx and E2A loci leads to pre-B cell leukemia (i.e., pre-B cell transforming factor).[35]

The t(1;19) chromosomal translocation leads to an E2A-Pbx1 fusion protein that is observed in 5% of pediatric and 3% of adult B-cell acute lymphoblastic leukemia cases.[36] The E2A-Pbx1 chimeric protein turns Pbx into a constitutive transcriptional activator by fusing the activation domain of E2A with the DNA-binding homeodomain of Pbx1.[37] Like the wild-type Pbx1, E2A-Pbx1 is capable of cooperatively binding to DNA

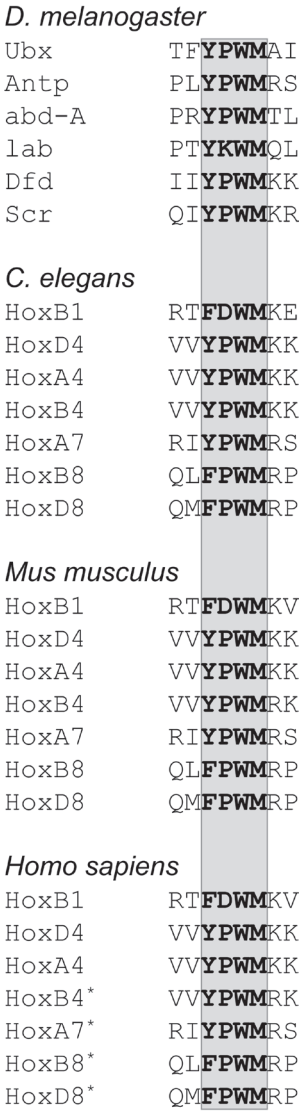


Figure 1.6 Evolutionary conservation of the YPWM motifs from fruit flies to humans. Note that all the motifs shown contain W and M. W is absolutely conserved in natural HOX proteins, but different residues are sometimes found instead of M. (Sequences were compiled from reference 33 except for entries with an asterisk (*) which were obtained from the National Center for Biotechnology Information: www.ncbi.nlm.nih.gov)

with other HOX proteins to the motif 5′-ATCAATCAA-3′.[38] Although the exact mechanism of leukemogenesis is still a matter of some debate, E2A-Pbx1 activates the expression of a novel WNT gene, Wnt16,[39] and a tyrosine kinase, EB-1,[40] which are believed to contribute to oncogenesis. Recently, siRNA inhibition of E2A-Pbx1 and subsequent downstream deactivation of Wnt16 and EB-1 has been shown to increase apoptosis in 697 cells which harbor the t(1;19) translocation.[41] This suggests that interfering with the abnormal function of E2A-Pbx1 may be a possible therapeutic approach for this type of leukemia.

1.5 Programmable DNA-Binding Small Molecules

Nature produces many small molecules which bind to DNA, but the polypyrrole compounds netropsin (**1**) and distamycin (**2**) are of particular relevance to the protein-DNA dimerizers designed in this thesis. Netropsin and distamycin bind to the DNA minor groove of AT-rich tracts and make specific H-bonds with

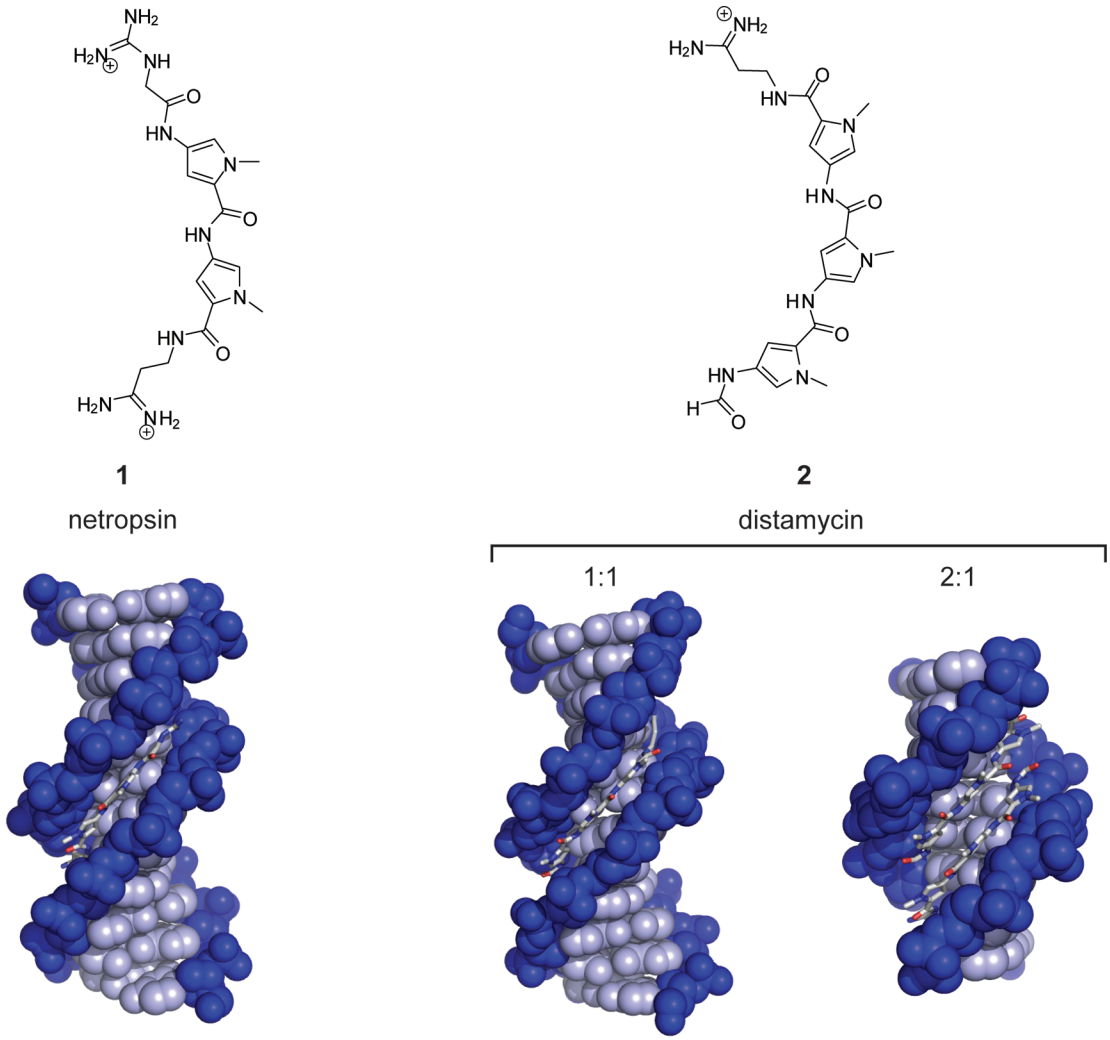


Figure 1.7 Structures illustrating DNA recognition by the natural products netropsin (**1**) and distamycin (**2**). The dicationic dipyrrole netropsin binds to narrow AT tracts of the DNA minor groove as a monomer (PDB code 6BNA). The monocationic tripyrrole distamycin can bind to AT tracts of the DNA minor groove as either a monomer or a dimer. (PDB codes 2DND and 378D).

the edges of the nitrogenous bases (Figure 1.7).[42-43] It was postulated in the mid-1980s that replacement of an *N*-methylpyrrole (Py) of distamycin by an *N*-methylimidazole (Im) would lead to a compound that would bind preferentially to G instead of A or T by making specific contacts to exocyclic amine of G in the minor groove.[44-45] Synthesis of such an ImPyPy distamycin analogue in the Dervan group led to the discovery that this compound actually bound preferentially to 5′-WGWCW-3′ instead of the expected 5′-GWW-3′ (where W = A or T).[46] This result was compellingly rationalized following the publication of

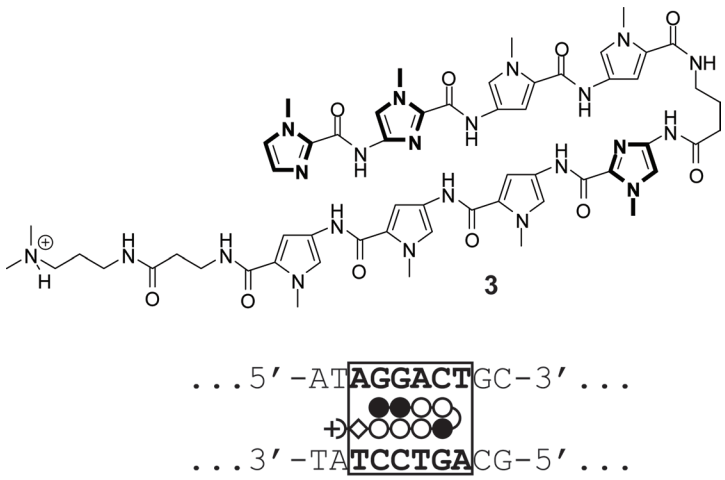


Figure 1.8 The chemical structure and schematic of a typical pyrrole-imidazole hairpin polyamide (**3**). *N*-methylpyrrole (Py) is represented by an open circle, *N*-methylimidazole (Im) by a filled circle, *γ*-aminobutyic acid (γ) by a half circle, *β*-alanine (β) by a diamond, and *N,N*-dimethylaminopropylamine (Dp) by a half circle with a plus. The ring pairings prefer binding to different base pairs: Im/Py for G·C, Py/Im for C·G, Py/Py for A·T or T·A. The turn residue (i.e., γ) and the C-terminal β, sometimes referred to as the tail, prefer binding over A·T or T·A base pairs.

an NMR structure of a 2:1 distamycin:DNA complex,[47] which suggested Im could stack across from a Py generating an Im/Py pair that specifically bound to G·C base pairs in the context of a noncovalent ImPyPy dimer.

In the early 1990s it was established that the 2:1 distamycin:DNA complex could be enforced by covalently attaching the antiparallel Im and Py carboxamides through a γ-aminobutyric acid linker to generate a hairpin polyamide such as compound **3** (Figure 1.8).[48] In this context, a set of pairing rules were determined, in which an Im/Py pair binds to a G·C, a Py/Im binds to a C·G, and Py/Py binds to both A·T and T·A (i.e., W). The W degeneracy of the Py/Py pairing was broken with the *N*-methyl-3-hydroxypyrrole (Hp) heterocycle, in which a Py/Hp pairing binds to A·T and Hp/Py to T·A (Figure 1.9).[49-50] For example, the ImImPyPy-γ-ImPyPyPy-β-Dp sequence of compound **3** has been programmed to target 5′-WGGWCW-3′.[51-52] The programmable DNA binding properties of polyamides have been thoroughly studied by a variety of physical techniques, including DNase and MPE footprinting,[53] x-ray crystallography[50,54-55] and multidimensional NMR,[56-57] DNA microarrays,[58] and fluorescence spectroscopy.[59-61] Recently a 27-compound

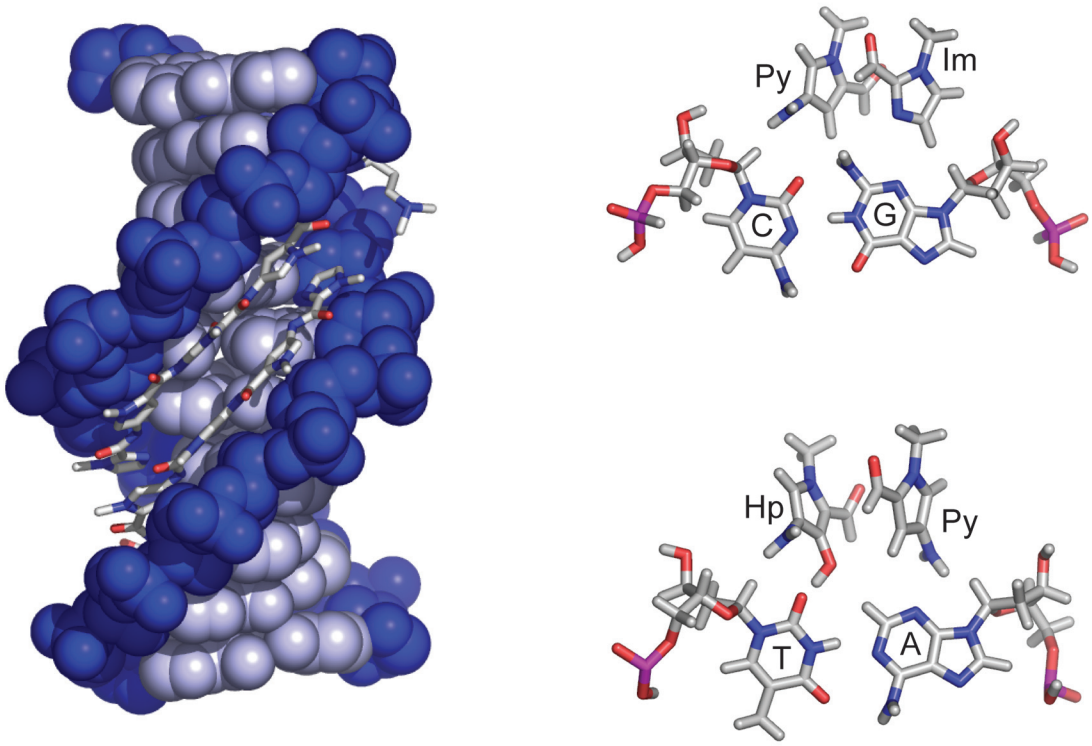


Figure 1.9 A structure illustrating the DNA recognition by a synthetic polyamide. Similar to the distamycin homodimer, the polyamide binds to DNA with the stacked heterocycles making contacts with nitrogenous base pairs in the minor groove (PDB code 1CVX). On the right, the structural basis of the pairing rules is shown.

polyamide library that can target all sequences of the type 5′-WWGNNNW-3′ has been completed.[62] Polyamides can bind to DNA as it exists in cell nuclei on the nucleosome core particle[63-64] and fluorescently labeled derivatives often localize to the nucleus of living cells.[65-67] Certain polyamides have shown activity in cell culture experiments. Notable recent examples inhibit HIF1-α[68] and androgen receptor[69] DNA interactions in the promoter regions of target genes and down-regulate expression.

1.6 Synthetic Mimics of HOX Transcription Factors

Small-molecule synthetic mimics of HOX transcription factors have been created by combining the technology of polyamides with an analysis of HOX proteins.[70] The initial idea was to mimic Ubx as observed in the Ubx/Exd/DNA ternary complex structure.[30] Polyamide-peptide conjugate **4** was designed to bind the DNA minor groove at 5′-TGGTCA-3′ and project an FYPWMKG peptide towards Exd bound to the adjacent major groove at 5′-ATCA-3′ (Figure 1.10). Electrophoretic mobility shift assays showed that conjugate **4** enhanced the binding affinity of Exd to DNA by approximately three orders of magnitude at the composite site of 5′-TGGTCAATCA-3′ but not mismatch sites. Subsequently, unbiased DNA microarray experiments demonstrated that the intended cognate DNA site enabled complex formation better than other sequences.[58] Thus, this conjugate serves as a protein-DNA dimerizer that mimics Ubx and other HOX proteins by binding cooperatively to specific DNA sites with the cofactor protein Exd.[70] Some compounds with longer linkers have also been shown to be make temperature sensitive protein-DNA dimerizers.[71]

1.7 Artificial Transcription Factors

More generally, the covalent attachment of functional peptides to DNA-binding small molecules has led to the development of compounds called artificial transcription factors.[72-73] An early example of such a compound utilized a triple-helix forming oligonucleotide (TFO) conjugated to a VP16 activation peptide that increased luciferase gene expression *in vitro*.[74] Polyamides conjugated to GCN4 dimerization and amphipathic helix peptide activation domains also lead to activation of gene expression on a g-less cassette *in vitro*.[75] The peptide domains have been reduced down to as little as 8 and 16 amino acids and still activate gene expression.[76] Experiments using a poly-proline helix as a structurally rigid spacer domain have defined the optimal distance between the polyamide and activation peptides.[77] Only in the past couple of years has a non-peptidic activation domain been conjugated to a polyamide and shown to activate gene expression *in vitro*.[78] A polyamide-

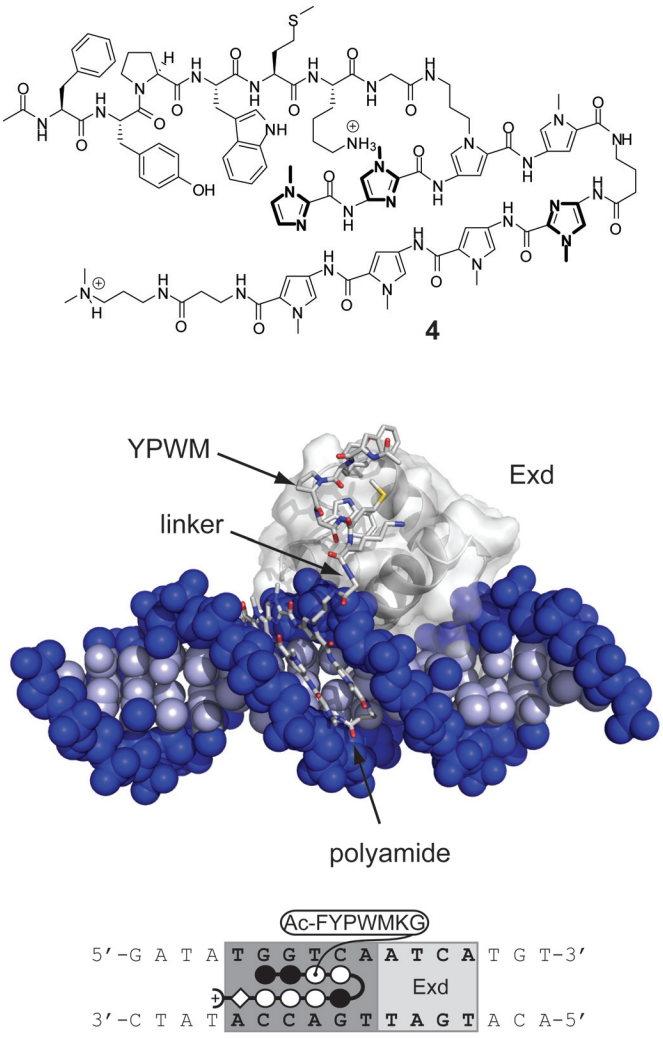


Figure 1.10 The chemical structure and model of a synthetic protein-DNA dimerizer (**4**). The dimerizer consists of a DNA-binding pyrrole-imidazole polyamide conjugated to a heptapeptide containing a YPWM motif that recruits Exd to the adjacent major groove. The illustrative model was constructed from a representative polyamide-DNA structure (PDB code 365D) and the Ubx/Exd/DNA structure (PDB code 1B8I).

peptoid activation domain has very recently been discovered to activate gene expression of a transiently transfected luciferase gene in cell culture,[79] but a cell-permeable, artificial transcription factor that activates an endogenous gene has yet to be reported.

1.8 Scope of this Work

This thesis describes efforts toward determining the essential components of protein-DNA dimerizer transcription factor mimics that should enable them to function in living cells. Chapter 2 details experiments involving that minimization of protein-DNA dimerizer **4** paying particular attention to the protein-binding domain. This led to the discovery that the orginial FYPWMKG heptapeptide can be reduced to a WM dipeptide. Inverting the tryptophan stereochemistry, from the natural L-from to the unnatural D-form, made the dimerizer less sensitive to temperature increases. Chapter 3 shows that the shape of protein-DNA dimerizers can be changed from branched to linear which expands the range of DNA sites and orientations for dimerization. The optimal DNA site spacings and necessary protein-binding elements were determined, which lead to the discovery that a WMK tripeptide works as a protein-binding domain in this context. In chapter 4 fluorescent, cell-permeable protein-DNA dimerizers are presented that incorporate the minimization efforts from the previous chapters.

Chapter 5 details the progress towards the development of a general, cell-permeable artificial transcription factor that mimics the natural HIF-1α protein. In the native system, endogenous HIF-1α binds to the hypoxia response element (HRE) of the VEGF promoter in response to hypoxia.[80] HIF-1α then recruits the enormous co-activator CREB binding protein (CBP) which then activates gene expression. The design strategy for a small molecule HIF-1α mimic employs a polyamide known to bind the VEGF HRE conjugated to compounds that bind CBP. Thus, an effort was made to artificially recruit CBP to the VEGF promoter and activate gene expression in cell culture experiments.

1.9 References

(1) Lander, E. S. et al. *Nature* **2001**, *409*, 860-921.

(2) Venter, J. C. et al. *Science* **2001**, *291*, 1304-1351.

(3) Lee, T. I. and Young, R. A. *Annu. Rev. Genet.* **2000**, *34*, 77-137.

(4) Kadonaga, J. T. *Cell* **2004**, *116*, 247-257.

(5) Levine, M. and Tjian, R. *Nature* **2003**, *424*, 147-151.

(6) Wolberger, C. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 29-56.

(7) Darnell, J. E. *Nat. Rev. Cancer* **2002**, *2*, 740-749.

(8) Matson, G. A., Evans, S. K., Green, M. R. *Annu. Rev. Genomics Hum. Genet.* **2006**, *7*, 29-59.

(9) Näär, A. M., Lemon, B. D., Tjian, R. *Annu. Rev. Biochem.* **2001**, *70*, 475-501.

(10) Bourbon, H-M. et al. *Mol. Cell* **2004**, *14*, 553-557.

(11) Williams, S. K. and Tyler, J. K. *Curr. Opin. Genet. Develop.* **2007**, *17*, 88-93.

(12) Kim, Y., Geiger, J. H., Hahn, S., Sigler, P. B. *Nature* **1993**, *365*, 512-520.

(13) Kim, J. L., Nikolov, D. B., Burley, S. K. *Nature* **1993**, *365*, 520-527.

(14) Kim, J. L., Burley, S. K. *Nat. Struct. Biol.* **1994**, *1*, 638-653.

(15) Juo, Z. S., Chiu, T. K., Leiberman, P. M., Baikalov, I., Berk, A. J., Dickerson, R. E. *J. Mol. Biol.* **1996**, *261*, 239-254.

(16) Pavletich, N. P., Pabo, C. O. *Science* **1991**, *252*, 809-817.

(17) Ellenberger, T. E., Brandl, C. J., Struhl, K., Harrison, S. C. *Cell* **1992**, *71*, 1223-1237.

(18) Glover, J. N. M., Harrison, S. C. *Nature* **1995**, *373*, 257-261.

(19) Panne, D., Maniatis, T., Harrison, S.C. *Cell* **2007**, *129*, 1111-1123.

(20) Lewis, E. B. *Nature* **1978**, *276*, 565-570.

(21) Veraksa, A., Del Campo, M., McGinnis, W. *Mol. Genet. Metab.* **2000**, *69*, 85-100.

(22) Pearson, J. C., Lemons, D., McGinnis, W. *Nat. Rev. Genet.* **2005**, *6*, 893-904.

(23) Abate-Shen, C. *Nat. Rev. Cancer* **2002**, *2*, 777-785.

(24) Chang, C., Immaculata, D., Clearly, M. L. *Mol. Cell. Biol.* **1997**, *17*, 81-88.

(25) Owens, B. M., Hawley, R. G. *Stem Cells* **2002**, *20*, 364-379.

(26) Mortlock, D. P., Innis, J. W. *Nat. Genet.* **1997**, *15*, 179-180.

(27) Mann, R. S., Chan, S. K. *Trends Genet.* **1996**, *12*, 258-262.

(28) Mann, R. S., Affolter, M. *Curr. Opin. Genet. Dev.* **1998**, *8*, 423-429.

(29) Ryoo, H. D., Marty, T. Casares, F., Affolter, M., Mann, R. S. *Development* **1999**, *126*, 5137-5148.

(30) Passner, J. M., Ryoo, H. D., Shen, L. Y., Mann, R. S., Aggarwal, A. K. *Nature* **1999**, *397*, 714-719.

(31) Piper, D. E., Batchelor, A. H., Chang, C. P., Clearly, M. L., Wolberger, C. *Cell* **1999**, *96*, 587-597.

(32) LaRonde-LeBlanc, N. A.; Wolberger, C. *Genes Dev.* **2003**, *17*, 2060-2072.

(33) Duboule, D. editor in *Guidebook to the Homeobox Genes*; Oxford University Press: Oxford, 1994.

(34) Peifer, M., Wieschaus, E. *Genes. Dev.* **1990**, *4*, 1209-1223.

(35) Kamps, M. P., Murre, C., Sun, X., Baltimore, D. *Cell* **1990**, *60*, 547-555.

(36) Pui, C., Relling, M. V., Downing, J. R. *N. Eng. J. Med.* **2004**, *350*, 1535-1548.

(37) Lu, Q., Wright, D. D., Kamps, M. P. *Mol. Cell. Biol.* **1994**, *14*, 3938-3948.

(38) Knoepfler, P. S., Kamps, M. P. *Oncogene* **1997**, *14*, 2521-2531.

(39) McWhirter, J. R., Neuteboom, S. T. C., Wancewicz, E. V., Monia, B. P., Downing, J. R., Murre, C. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 11464-11469.

(40) Fu, X., McGrath, S., Pasillas, M., Nakazawa, S., Kamps, M. P. *Oncogene* **1999**, *18*, 4920-4929.

(41) Casagrande, G., Kronnie, G., Basso, G. *Haematologica/The Hematology Journal* **2006**, *91*, 765-771.

(42) Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P., Dickerson, R. E. *J. Mol. Biol.* **1985**, *183*, 553-563.

(43) Coll, M., Frederick, C. A., Wang, A. H., Rich, A. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 8385-8389.

(44) Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P., Dickerson, R. E. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 1376.

(45) Lown, J. W., Krowicki, K., Bhat, U. G., Skorobogaty, A., Ward, B., Dabrowiak, J. C. *Biochemistry* **1986**, *25*, 7408.

(46) Wade, W. S., Mrksich, M., Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783-8794.

(47) Pelton, J. G., Wemmer, D. E. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 5723-5727.

(48) Mrksich, M., Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 2572.

(49) White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., Dervan, P. B. *Nature* **1998**, *391*, 468.

(50) Kielkopf, C. L., White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., Dervan, P. B. *Science* **1998**, *282*, 111.

(51) Dervan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 2215-2235.

(52) Dervan, P. B.; Edelson, B. S. *Curr. Opin. Struct. Biol.* **2003**, *13*, 284-299.

(53) Trauger, J. W., Dervan, P. B. *Methods Enzymol.* **2001**, *340*, 450-466.

(54) Kielkopf, C. L., Baird, E. E., Dervan, P. B., Rees, D. C. *Nat. Struct. Biol.* **1998**, *5*, 104-109.

(55) Kielkopf, C. L., Bremer, R. E., White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., Dervan, P. D., Rees, D. C. *J. Mol. Biol.* **2000**, *295*, 557-567.

(56) Urbach, A. R., Love, J. J., Ross, S. A., Dervan, P. B. *J. Mol. Biol.* **2002**, *320*, 55-71.

(57) Zhang, Q., Dwyer, T. J., Tsui, V., Case, D. A., Cho, J., Dervan, P. B., Wemmer, D. E. *J. Am. Chem. Soc.* **2004**, *126*, 7958-7966.

(58) Warren, C. L., Kartochvil, N. C. S., Hauschild, K. E., Foister, S., Brezinski, M. L., Dervan, P. B., Phillips, G. N., Ansari, A. Z. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 867-872.

(59) Rucker, V. C., Foister, S., Melander, C., Dervan, P. B. *J. Am. Chem. Soc.* **2003**, *125*,

1195-1202.

(60) Fechter, E. J., Olenyuk, B., Dervan, P. B. *J. Am. Chem. Soc.* **2005**, *127*, 16685-16691.

(61) Chenoweth, D. M., Viger, A., Dervan, P. B. *J. Am. Chem. Soc.* **2007**, *129*, 2216-2217.

(62) Hsu, C. F., Phillips, J. W., Trauger, J. W., Farkas, M. E., Belitsky, J. M., Heckel, A., Olenyuk, B. Z., Puckett, J. W., Wang, C. C. C. Dervan, P. B. *Tetrahedron* **2007**, *63*, 61416-6151.

(63) Gottesfeld, J. M., Melander, C., Suto, R. K., Raviol, H., Luger, K., Dervan, P. B. *J. Mol. Biol.* **2001**, *309*, 615-629.

(64) Suto, R. K., Edayathumangalam, R. S., White, C. L., Melander, C., Gottesfeld, J. M., Dervan, P. B., Luger, K. *J. Mol. Biol.* **2003**, *326*, 371-380.

(65) Belistky, J. M., Leslie, S. J., Arora, P. S., Beerman, T. A., Dervan, P. B. *Bioorg. Med. Chem.* **2002**, *10*, 3313-3318.

(66) Best, T. P., Edelson, B. S., Nickols, N. G., Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 12063-12068.

(67) Edelson, B. S., Best, T. P., Olenyuk, B., Nickols, N. G., Doss, R. M., Foister, S., Heckel, A., Dervan, P. B. *Nucleic Acids Res.* **2004**, *32*, 2802-2818.

(68) Olenyuk, B. Z., Zhang, G., Klco, J. M., Nickols, N. G., Kaelin, W. G., Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 16768-16773.

(69) Nickols, N. G., Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 10418-10423.

(70) Arndt, H.-D., Hauschild, K. E., Sullivan, D. P., Lake, K., Dervan, P. B., Ansari, A. Z. *J. Am. Chem. Soc.* **2003**, *125*, 13322-13323.

(71) Hauschild, K. E., Metzler, R. E., Arndt, H.-D., Moretti, R., Raffaelle, R., Dervan, P. B., Ansari, A. Z. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 5008-5013.

(72) Ansari, A. Z. and Mapp, A. K. *Curr. Opin. Chem. Biol.* **2002**, *6*, 765-772.

(73) Arndt, H.-D. *Angew. Chem., Int. Ed.* **2006**, *45*, 4552-4560.

(74) Kuznetsova, S., Ait-Si-Ait, S., Nagibneva, I., Troalen, F., Le Villain, J., Harel-Bellan, A., Svinarchuk, F. *Nucleic Acids Res.* **1999**, *27*, 3995-4000.

(75) Mapp, A. K., Ansari, A. Z., Ptashne, M., Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 3930-3935.

(76) Ansari, A. Z., Mapp, A. K., Nguyen, D. H., Dervan, P. B., Ptashne, M. *Chem. Biol.* **2001**, *8*, 583-592.

(77) Arora, P. S., Ansari, A. Z., Best, T. P., Ptashne, M., Dervan, P. B. *J. Am. Chem. Soc.* **2002**, *124*, 13067-13071.

(78) Kwon, Y., Arndt, H.-D., Mao, Q., Choi, Y., Kawazoe, Y., Dervan, P. B., Uesugi, M. *J. Am. Chem. Soc.* **2004**, *126*, 15940-15941.

(79) Xiao, X., Yu, P., Lim, H.-S., Sikder, D., Kodadek, T. *Angew. Chem. Int. Ed.* **2007**, *46*, 2865-2868.

(80) Kaelin, W. G. *Genes Dev.* **2002**, *16*, 1441-1445.

Chapter 2

Minimization of a Protein-DNA Dimerizer

The text of this chapter was taken in part from a manuscript co-authored with Hans-Dieter Arndt,[†] Mary Brezinksi,[‡] Aseem Z. Ansari[‡] and Peter B. Dervan[†] (Caltech[†] and University of Wisconsin, Madison[‡])

(Stafford, R. L., Arndt, H.-D., Brezinski, M., Ansari, A. Z., Dervan, P. B. *J. Am. Chem. Soc.* **2007**, *129*, 2660-2668.)

Abstract

A protein-DNA dimerizer constructed from a DNA-binding polyamide and the peptide FYPWMKG facilitates the binding of a natural transcription factor Exd to an adjacent DNA site. The Exd binding domain can be reduced to a dipeptide WM attached to the polyamide through an ε-aminohexanoic acid linker with retention of protein-DNA dimerizer activity. Screening a library of analogues indicated that the tryptophan indole moiety is more important than methionine's side-chain or the *N*-terminal acetamide. Remarkably, switching the stereochemistry of the tryptophan residue (L to D) stabilizes the dimerizer•Exd•DNA ternary complex at 37 °C. These observations provide design principles for artificial transcription factors that may function in concert with the cellular regulatory circuitry.

2.1 Introduction

Multivalent interactions are frequently encountered in biological systems.[1] Typically, the monovalent elements that mediate these molecular interactions use a small surface area and bind weakly. These monovalent elements are repeated, and the resulting multivalency improves the association between the interacting molecules. Dimerization is a key regulatory event in numerous biochemical settings, notably in signal transduction and the regulation of transcription.[2] Divalent small molecules called chemical inducers of dimerization or dimerizers have been used to facilitate protein-protein interactions.[2,3] Recently, we described a protein-DNA dimerizer, where one module binds a specific DNA sequence while the other interacts with a specific DNA binding protein (Figure 2.1).[4] This molecule generates a bidentate surface that enhances the association of the targeted protein with its cognate DNA site. This divalent molecule improves the affinity of the targeted DNA-binding protein for its specific DNA sequence, whereas it is ineffective at sites where the DNA sequence does not match the preferred site of the target protein. Our long term goal is to ask whether protein-DNA dimerizers could serve as artificial regulators of gene expression in a living cell.[4-7]

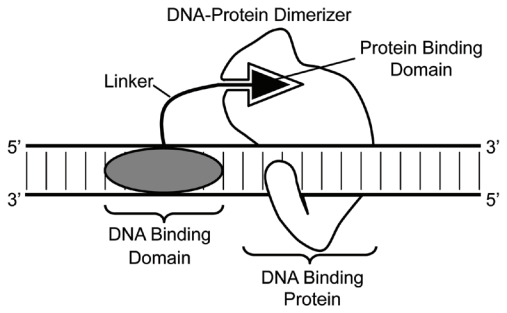


Figure 2.1 Protein-DNA dimerizer facilitating the binding of a protein to an adjacent DNA site.

Transcription factors are modular in structure and combine DNA-binding domains with other functional domains and lead to the activation or repression of genes.[1a] Transcriptional activators themselves can be viewed as dimerizers composed of DNA and protein binding surfaces.[3a] Polyamides are ideal DNA-binding domains since they can be programmed to bind to a broad repertoire of DNA sequences using an aromatic amino acid pairing code.[8] A hairpin polyamide-YPWM conjugate facilitates the binding a natural transcription factor Extradenticle (Exd) to an adjacent DNA site.[4] The protein-DNA

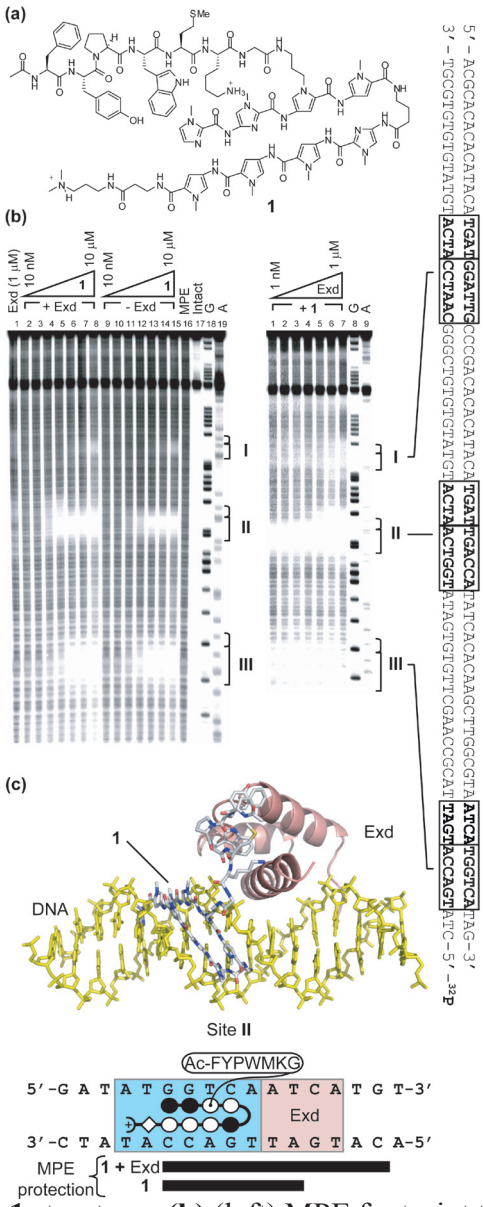


Figure 2.2 **(a)** Dimerizer **1** structure. **(b)** (left) MPE footprint titration of conjugate **1** on a fragment of pHDA1 in the presence (40 nM, lanes 2-8) and absence of Exd protein (lanes 9-15) and 100 μM base pair calf thymus DNA. Lane 1: Exd only (1 μM). Lanes 2-8: 0.01, 0.03, 0.1, 0.3, 1, 3, 10 μM conjugate **1** all with 40 nM Exd. Lanes 9-15: 0.01, 0.03, 0.1, 0.3, 1, 3, 10 μM conjugate **1**. Lane 16: MPE standard. Lane 17: Intact DNA. Lane 18: G-sequencing lane. Lane 19: A-sequencing lane. (middle) MPE footprint titration of Exd in the presence of dimerizer **1** and 100 μM base pair calf thymus DNA. Lanes 1-7: 0.01, 0.03, 0.1, 0.3, 1, 3, 10 μM Exd all with 1 μM conjugate **1**. Lane 8: G-sequencing lane. Lane 9: A-sequencing lane. (right) Sequence of fragment of pHDA1 used for MPE footprinting. **(c)** Structural model of dimerizer **1** in a complex with DNA and Exd. A key is shown below with a dimerizer ball-and-stick model which indicates where the MPE protection pattern was observed. (filled circle = *N*-methylimidazole, empty circle = *N*-methylpyrrole, diamond = *β*-Ala, half circle with plus = *N,N*-dimethylaminopropylamine, half circle = *γ*-aminobutyric acid)

dimerizer **1** mimics the homeobox (HOX) transcription factor Ultrabithorax (Ubx) (Figure 2.2a).[9a] Ubx cooperatively binds to DNA with the HOX cofactor protein Exd.[10] Their interaction is primarily mediated through Ubx's conserved YPWM peptide motif, which crosses the DNA minor groove to the major groove binding Exd.[9] Similarly, in human homologs as well, the YPWM motif is highly conserved within the HOX protein family and is often referred to as a hexapeptide ϕYPWMK, as the central YPWM tetrapeptide is often flanked by a hydrophobic residue (ϕ) and a charged residue, such as Lys or sometimes Arg.[11]

In this study, we address what are the minimum requirements for Exd recruitment. The Trp residue is strictly conserved and a conservative mutation of Trp to Phe has been shown to lead to HOX proteins that no longer cooperatively bind to DNA with their cofactor proteins.[12] Met to Ile and Tyr to Leu mutations to the YPWM motif of HoxB4 has also been shown to eliminate cooperative binding.[13] The crystal structures of Ubx/Exd/DNA, HoxB1/Pbx1/DNA and HoxA9/Pbx1/DNA complexes show the Trp indole deep in the protein binding pocket with an adjacent Met or Leu side-chain reinforcing this interaction and mediating important protein contacts.[9] Thus, our efforts to minimize the protein-binding domain have centered on the Trp and Met residues.

The Exd binding domain was reduced to the dipeptide WM and tethered to the polyamide DNA-binding domain through an ε-aminohexanoic acid linker. We report here that this conjugate with a smaller protein binding domain retains significant dimerizing activity with Exd. A set of dipeptide analogues were screened to assess the importance of the remaining moieties. Minimized conjugates were then studied at physiologically relevant temperatures for fruit fly (20 °C) and human cell lines (37 °C).

2.2 Results

Protein-DNA Dimerizer•Exd Sequence Specificity

Exd is expected to bind 5'-TGAT-3' as found in the Ubx/Exd/DNA crystal structure,[9a]

and the hairpin polyamide ImImPyPy-γ-ImPyPyPy codes for the sequence 5'-WGWCCW-3', where W equals A or T.[8] Exd (underlined) and polyamide sites were combined to create a composite site 5'-TGATTGACCA-3' to study Exd recruitment by polyamide-FYPWMKG conjugate **1** (Figure 2.2a).[4] To confirm both the orientation preference of Exd recruitment as well as the specificity with respect to mismatch sites, MPE footprinting titrations[14] have been carried out on a DNA fragment featuring a polyamide mismatch site I (5'- TGATGGATTG-3') and Exd upstream II (5'-TGATTGACCA-3'), as well as downstream III (5'-TGACCATGAT-3') orientations (Figure 2.2b, right). An MPE footprint titration of dimerizer **1** without Exd protein present confirmed its highly specific binding, with an identical protection pattern for both sites II and III. Binding to site I, which is a single base pair mismatch for the polyamide DNA-binding domain in **1**, only occurred at ≥ 10 μM concentration, indicating a selectivity of better than 50-fold (Figure 2.2b, left). When 40 nM Exd was present with **1**, the MPE protection pattern was observed at a 3-fold lower concentration exclusively at site II, enhancing the selectivity over the site I to 150-fold and the Exd downstream site III to 3-fold. Furthermore, despite the presence of five sites for Exd on the DNA fragment, Exd alone at 1 μM concentration did not give rise to detectable binding. These results provide evidence for cooperative enhancement of binding of the dimerizer and Exd only when Exd's site is located upstream from the polyamide's. Subsequent experiments using DNA microarrays confirm that **1** prefers to recruit Exd to the consensus site 5'-NGANWGWC-3' over other Exd-dimerizer DNA site spacings and orientations.[4c]

The Exd protein binding site in the presence of dimerizer **1** was determined to single base pair resolution (Figure 2.2b, middle). At low concentrations of Exd, dimerizer **1** occupied the polyamide site. With increasing Exd concentration, the MPE protection pattern became enlarged by four base pairs in the 3'-direction (^{32}P-labeled strand). Taking into account that an MPE footprint is generally observed 2 base pairs shifted to the 3'-end,[14] this maps protein-binding to 5'-TGAT-3' and dimerizer-binding to the 5'-TGACCA-3'

(Figure 2.2c). Notably, no additional protein-induced protection was evident at mismatch site I or on the downstream arrangement site III. This observation supports the selectivity of the dimerization event, both for sequence composition and arrangement of the respective half-sites.

Analyzing Changes to the Protein-binding Domain on Dimerization (4 °C)

The binding of Exd to DNA in the presence of various dimerizers was measured at 4 °C by electrophoretic mobility shift assays (EMSA). Exd protein binding was studied on a ^{32}P-labeled 47 base pair double-stranded DNA duplex containing a match composite

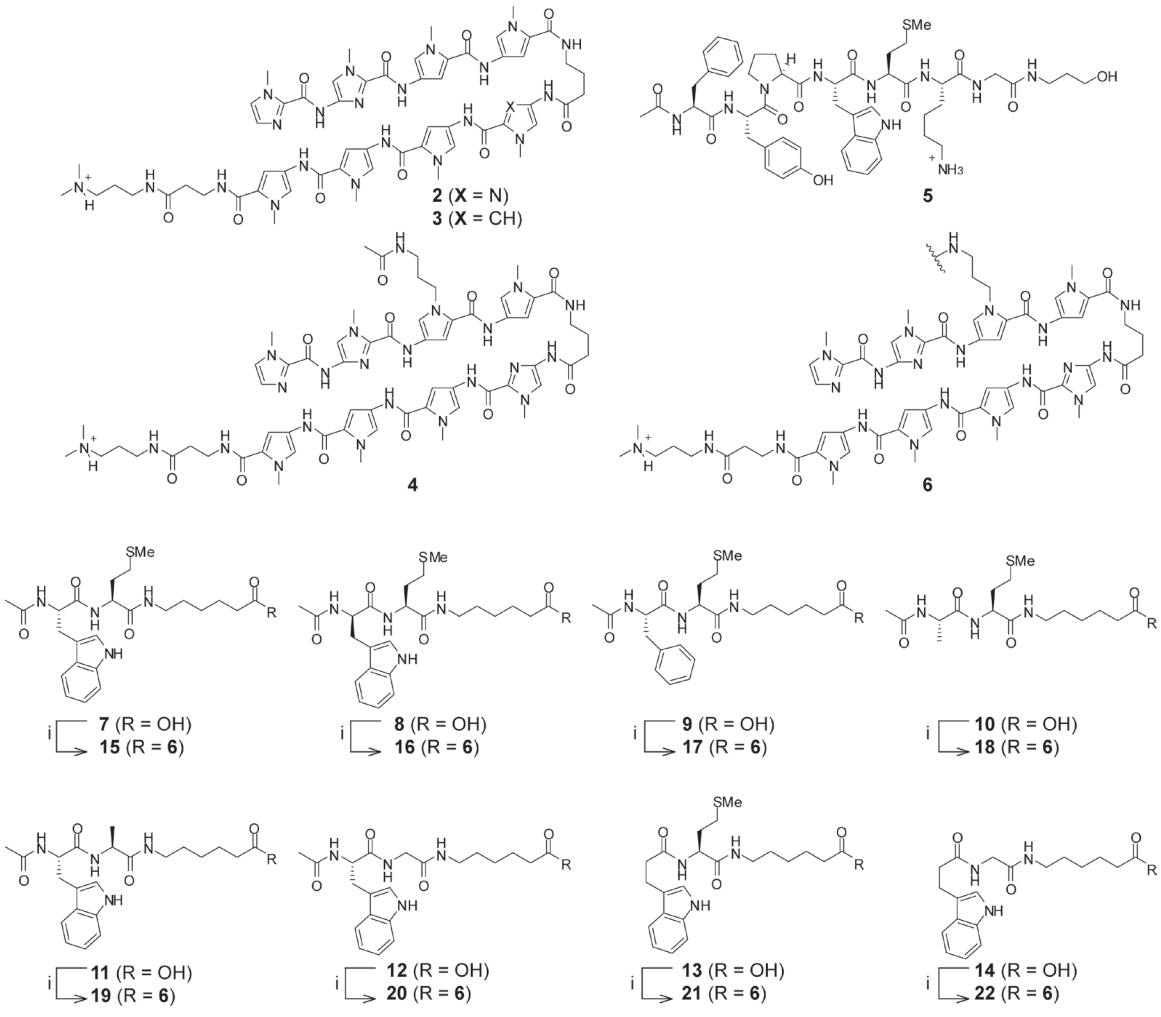


Figure 2.3 Structures of compounds **2**-**22** and the synthesis of minimized polyamide-dipeptide conjugates **15**-**22**: *i*) **6** (R=H), HBTU, DMF, excess DIEA, 20 °C, 1-3 h.

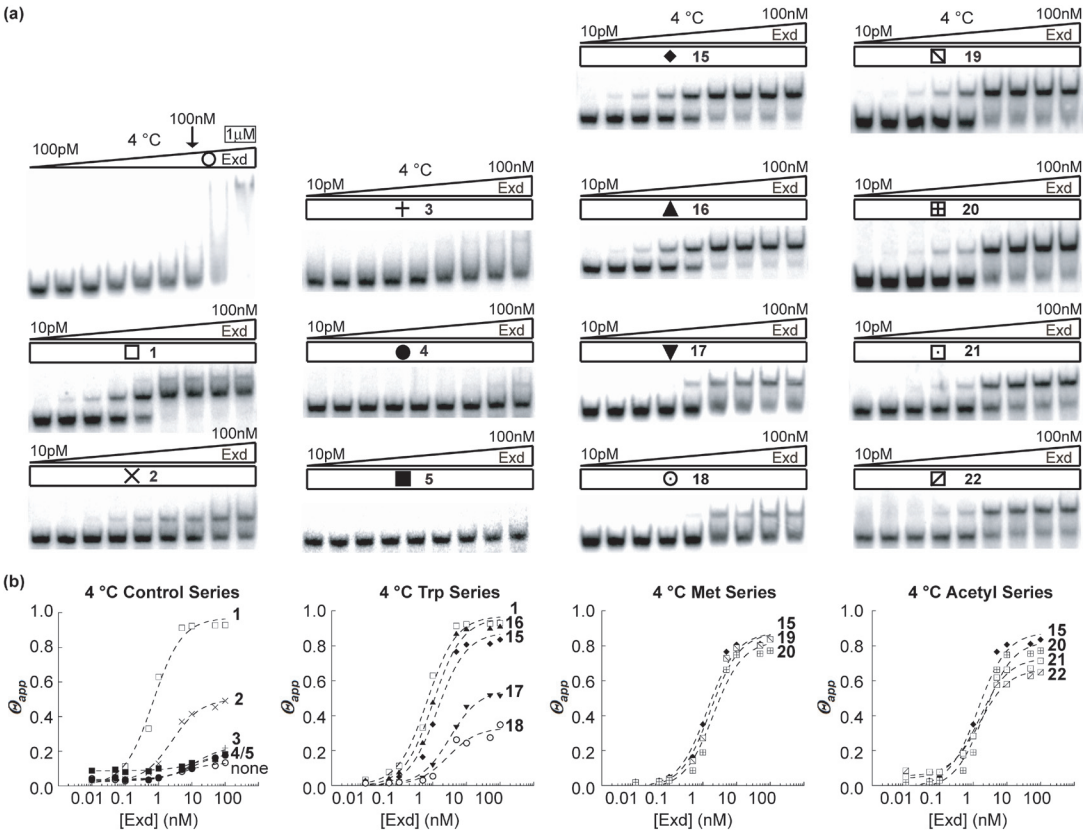


Figure 2.4 Representative gel shift assays at 4 °C all done in triplicate. Autoradiograms are shown for Exd titrations with a 47-base pair ^{32}P-labeled dsDNA probe containing both Exd (underlined) and dimerizer binding sites (5'-TGATTGACCA-3'). **(a)** The upper left image shows increasing concentrations of Exd only from left to right at 100 pM, 500 pM, 1 nM, 5 nM, 10 nM, 50 nM, 100 nM, 500 nM, and 1 μM. In all other experiments the indicated compound was applied at 50 nM, and increasing concentrations of Exd were studied from left to right: 10 pM, 50 pM, 100 pM, 500 pM, 1 nM, 5 nM, 10 nM, 50 nM, and 100 nM. The upper band consists of the Exd-dimerizer-DNA complex while the lower band consists of the dimerizer-DNA complex. **(b)** Isotherms for Exd binding at 4 °C organized into groups.

binding site (5'-TGATTGACCA-3') identical to site II in Figure 2b in the presence of compounds **1**-**5** and **15**-**22** (Figures 2.3 and 2.4). For comparison, a control titration with Exd alone was performed (top left, Figure 2.4). Negligible Exd binding was observed up to 100 nM concentrations, however, multiple Exd molecules appear to bind weakly and non-specifically at 1 μM, in line with previous findings.[4] Thus, subsequent titrations were kept below this threshold and performed in the 10 pM to 100 nM range.

Table 2.1 Exd-DNA-dimerizer stabilities at 4 °C

Compound	K_a of Exd [10^9 M^{-1}][a]	Θ_{max}[b]
1	1.5 ± 0.3	0.93 ± 0.01
15	0.7 ± 0.2	0.84 ± 0.06
16	1.0 ± 0.2	0.91 ± 0.03
19	0.6 ± 0.1	0.84 ± 0.04
20	0.5 ± 0.2	0.77 ± 0.08
21	0.7 ± 0.1	0.71 ± 0.10
22	0.7 ± 0.1	0.65 ± 0.07

[a] Mean of ≥ 3 titrations of Exd at 50 nM compound. Data fit to a Hill equation with n=1 (R > 0.98).
[b] Determined at least three times at 100 nM Exd and 50 nM dimerizer.

Upon the addition of 50 nM of conjugate **1**, the Exd DNA equilibrium association constant (K_a) was enhanced (K_a = 1.5 ± 0.3 × 10^9 M^{-1}),[15] yielding a nearly complete band shift (i.e., Θ_{app} ≥ approximately 0.93±0.01) at 5 nM Exd (Table 2.1). *Thus, the presence of dimerizer **1** enhances Exd's DNA affinity by 1000-fold.*[4] In controls, experiments with 50 nM of the parent polyamide **2,** which contains no peptide but has been shown previously by DNase footprinting experiments to bind specifically to the DNA-binding domain match site,[16] appears to enhance Exd's DNA affinity for DNA, as evidenced by the partial band shift (i.e., Θ_{app} = 0.5±0.2).[17] The mismatch polyamide **3**, which is not expected to bind the polyamide match site,[18] does not lead to any measurable enhancement of Exd's DNA affinity. Thus, the DNA-binding polyamide bound by itself to the match site appears to facilitate Exd's binding to DNA, but is insufficient for stable complex formation. As a further control, modified hairpin **4** was investigated to evaluate the effect of the proximal linker domain. The propylacetamide linker appeared to reduce the polyamide's contribution relative to hairpin **2**. Exd titrations in the presence of 50 nM of the YPWM peptide **5** did not lead to any measurable enhancement of Exd-DNA affinity, indicating that the polyamide DNA-binding domain is indispensable for dimerization.

The truncated dipeptide-dimerizer **15** exhibited a similar binding affinity for DNA to **1** (see DNase I footprinting data below), so 50 nM concentration was used for the dipeptide analogues to saturate the DNA binding site and ensure direct comparisons could be made to changes to the presumed Exd binding domain. Polyamide-Ac-WM-dipeptide conjugate **15** enhanced Exd's DNA affinity almost 50% as effectively as the larger conjugate **1** (K_a

= 7 ± 2 × 10^8 M^{-1}), with high levels of complex formation (Θ_{app} = 0.84 ± 0.06). When the stereochemistry of the Trp residue was inverted (**16**), Exd's DNA affinity remained the same (K_a = 1.0 ± 2 × 10^9 M^{-1}) with a slightly higher amount of complex formation (Θ_{app} = 0.91 ± 0.03). Substituting the Trp with Phe (**17**) or Ala (**18**) lowered Exd's DNA affinity so it qualitatively appeared to be as effective as polyamide **2** by itself. In contrast to what was observed with the Trp residue, the Met residue could be converted to Ala (**19**) and Gly (**20**) without strongly affecting Exd's DNA affinity (K_a's = 6 ± 1 × 10^8 M^{-1} and 5 ± 2 × 10^8 M^{-1}, respectively) or the levels of complex formation (Θ_{app} > 0.75). Noting that the stereochemistry of the Trp could be inverted (**15** to **16**), we examined whether the stereogenic acetamide might be eliminated. Both conjugates **21** and **22** maintained similar Exd binding affinities (both K_a's = 7 ± 1 × 10^8 M^{-1}), albeit with slightly reduced levels of complex formation (Θ_{app} = 0.71 ± 0.10 and 0.65 ± 0.07, respectively). Interestingly, all of the minimal dimerizers that showed strong Exd recruitment yielded binding constants within 3-fold of conjugate **1**.

Analyzing Selected Dimerizers for Exd Binding at Higher Temperatures

The initial screen of dimerizers at 4 °C by gel shift yielded the promising minimal dimerizers **15** and **16**, which were further investigated (Figure 2.5 and Table 2.2). Specifically, the ability of conjugates **15** and **16** to recruit Exd at higher temperatures will be relevant to planned experiments in fruit flies (20 °C) and in human cell lines (37 °C).[19] Furthermore, the weak contribution of the DNA-binding domain by itself was investigated at these higher temperatures using compound **2**. The change from 4 °C to 20 °C did not

Table 2.2 Exd-DNA-dimerizer stabilities at 20 °C and 37 °C

Compound	K_a of Exd [10^9 M^{-1}] at 20 °C[a]	Θ_{max} at 20 °C[b]	K_a of Exd [10^9 M^{-1}] at 37 °C[a]	Θ_{max} at 37 °C[b]
1	1.9 ± 0.2	0.91 ± 0.01	1.1 ± 0.2	0.91 ± 0.02
15	1.3 ± 0.1	0.65 ± 0.11	-[c]	0.30 ± 0.12
16	1.7 ± 0.1	0.86 ± 0.02	0.9 ± 0.1	0.75 ± 0.06

[a] Mean of ≥ 3 titrations of Exd at 50 nM compound. Data fit to a Hill equation with n = 1 (R > 0.99). [b] Determined at least three times at 100 nM Exd and 50 nM dimerizer. [c] Not determined because no stable complex was detected (i.e., mean $\Theta_{max} \leq 0.5$).

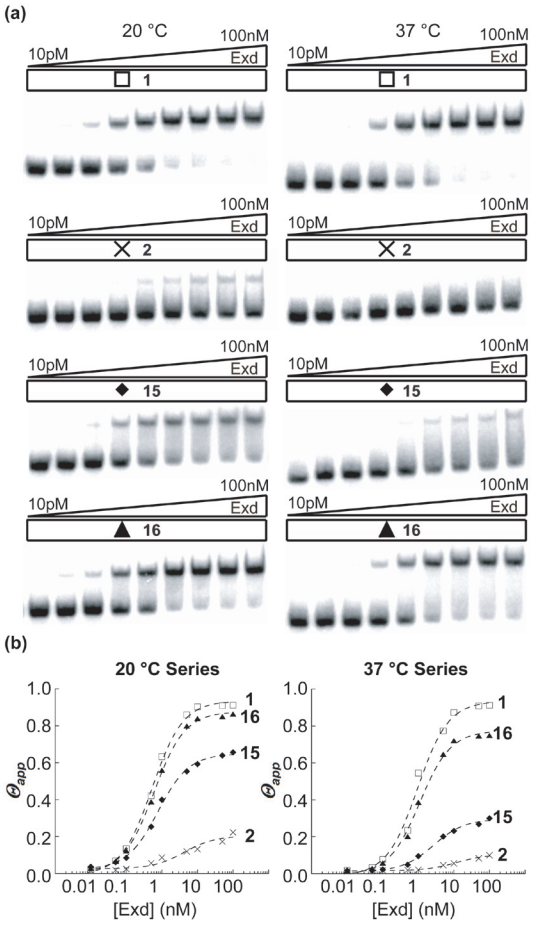


Figure 2.5 **(a)** Representative gel shift assays at 20 °C and 37 °C all done in triplicate shown in the same format as in Figure 4. **(b)** Isotherms for Exd binding at 20 °C and 37 °C organized into groups.

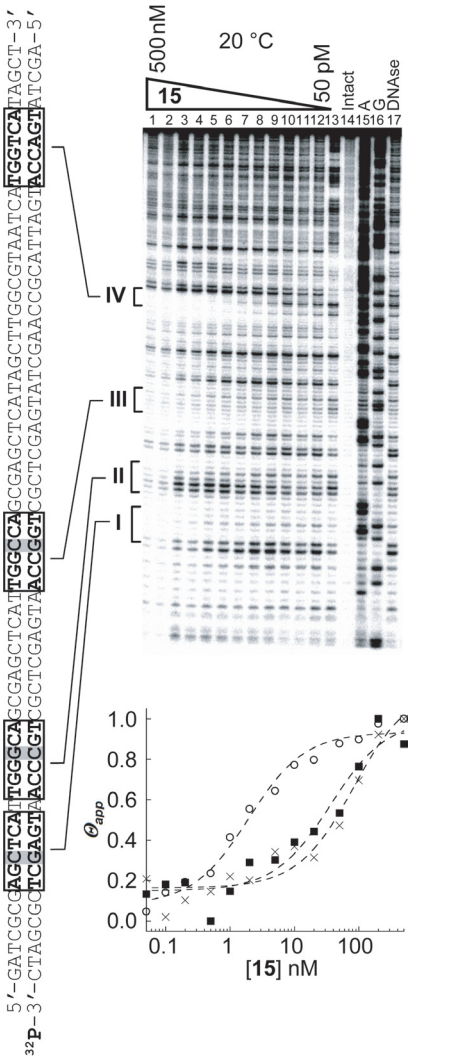


Figure 2.6 Quantitative DNase I footprinting titration of minimized dimerizer **15** on a restriction fragment containing a match site IV (5'-TGGTCA-3') and three mismatch sites I-III, (5'-AG**C**TCA-3'), (5'-TGG**G**CA-3'), and (5'-TGG**C**CA-3'), respectively. The analyzed binding site locations are indicated by square brackets and correlated to the sequence on the left. Lane 1-13: 500, 200, 100, 50, 20, 10, 5, 2, 1, 0.5, 0.2, 0.1, and 0.050 nM of compound **15**. Lane 14: Intact DNA. Lane 15: A-sequencing lane. Lane 16: G-sequencing lane. Lane 17: DNase I standard. The isotherms for binding to sites **I** (×), **III** (■), and **IV** (○) are shown below.

significantly change Exd's K_a in the presence of compounds **1**, **15**, and **16**, but markedly reduced the apparent effect seen by the polyamide DNA-binding domain **2**. Increasing the temperature further to 37 °C did not significantly affect Exd's K_a for compounds **1** and **16**, but considerably impaired compound **15**'s ability to recruit Exd. Remarkably the only structural difference between compound **15** and **16** is the stereochemistry at the tryptophan residue, which apparently leads to a more stable ternary complex.

Conjugate DNA Binding Affinity and Sequence Specificity

The DNA affinity and sequence selectivity of conjugate **15** at 20 °C in the absence of Exd was determined by DNase I footprinting (Figure 2.6 and Table 2.3).[14] Binding energetics for compounds **1-3** have been reported previously.[4a,16,18] Experiments with **15** were performed using the same 3'-radiolabeled 250-base pair restriction fragment of a plasmid that was previously used for conjugate **1**. This fragment contains a 5'-TGGTCA-3' match site IV as well as three single base-pair mismatch sites I-III, 5'-AG<u>C</u>TCA-3', 5'-TGG<u>G</u>CA-3', and 5'-TGG<u>C</u>CA-3', respectively. The attachment of the glycine-linked Ac-FYPWMK- peptide to the DNA-binding domain of conjugate **1** was previously found to reduce the affinity approximately 100-fold with respect to the parent polyamide.[4a] Furthermore, the sequence specificity of **1** was diminished with respect to mismatch sites I and III, but it was still able to discriminate against mismatch site II by about 20-fold. The affinity of the minimized hairpin polyamide-WM (**15**) for its match site IV was improved by approximately three-fold ($K_a = 5.1 \pm 1.0 \times 10^8$ M^{-1}). The sequence specificity was also found to be more than 25-fold for the match site IV with respect to mismatch sites I through

Table 2.3 Conjugate-DNA equilibrium association constants K_a's [10^8 M^{-1}][a]

Compound	5'-TGGTCA-3'	5'-TGG**C**CA-3'	5'-TGG**G**CA-3'	5'-AG**C**TCA-3'
1[b]	1.8 ± 0.2	1.3 ± 0.2 [1.4]	≤ 0.1 [>18]	1.5 ± 0.3 [1.2]
15	5 ± 1	0.2 ± 0.1 [25]	≤ 0.1 [>50]	0.12 ± 0.05 [42]

[a] Specificity ratio with respect to Match Site **IV** (5'-TGGTCA-3') shown in brackets.
[b] See reference 20.

III. Hence the reduction in the size of the protein binding domain of the conjugate fully restored the intrinsic DNA sequence selectivity of the protein-DNA dimerizer.

2.3 Discussion

The observation that a dipeptide WM is sufficient for Exd recruitment is remarkable given that the YPWM motif is highly conserved with the consensus sequence ϕYPWMK (where ϕ denotes a hydrophobic residue) in the majority of HOX proteins,[11] and that residues flanking the motif have been shown to contribute to the interaction.[12d] The total solvent-accessible surface area buried between Exd and the YPWM peptide of Ubx is ~570 $Å^2$ where on average the buried surface area between more stable protein-protein interfaces is about 2000 $Å^2$.[9,20] From the crystal structures of the Ubx/Exd/DNA and HoxB1/Pbx1/DNA ternary complexes it was clear however, that the majority of the contacts involved the Trp indole substituent, with the Met residue reinforcing the interaction.[9] Moreover, a related crystal structure of the HoxA9/Pbx1/DNA complex reveals a divergent AANWLH interaction motif bound in an entirely different conformation. The only distinct structural similarity between the AANWLH motif of HoxA9 and the other YPWM motifs is the orientation of the Trp side-chain in the protein binding pocket.

In line with these findings, the data presented here underscore the importance of the Trp side chain and to a lesser extent the Met residue. A "conservative" change of Trp to Phe, which reduces the hydrophobic interaction surface area and should eliminate the hydrogen bond to the main-chain carbonyl of Leu23a (Exd homeodomain numbering),[9] abrogates the efficacy of the protein-binding domain of the dimerizer. Similarly, a Trp to Ala substitution is not tolerated. Truncating the Met side-chain to a methyl or hydrogen only partially reduces efficacy, but still provides a functional protein-binding domain. Inverting the stereochemistry of the Trp residue enhances stability, particularly at higher temperatures. It is likely that the indole moiety adopts a similar position in the Exd binding pocket as in the crystal structures.[9] Thus, the structural change enforced by the stereochemical switch

(**15** and **16**) probably positions the acetamide in a favorable position for hydrogen bonding contacts with Tyr25 of Exd, the DNA, or the dimerizer's linker domain. Analogues lacking the acetamide retained some efficacy when compared to the L-Trp dimerizer **15**, but do not lead to further insight into why the unnatural peptide of **16** is an improvement over **15**.

Our results support the notion that the geometry of the Exd protein and the protein-binding module in the DNA-dimerizer-protein complex do not deviate substantially from the structurally characterized ternary DNA-protein-protein complexes. Furthermore, MPE footprinting established Exd's DNA binding site in the ternary DNA-dimerizer-protein complex as the expected 5'-TGAT-3' match sequence upstream of the polyamide binding site 5'-TGACCA-3'. Using this information, a model of dimerizer **15** was constructed by superimposing a representative polyamide structure (PDB code 365D)[21] on the Ubx/Exd/DNA ternary complex. (PDB code 18BI)[9a] The DNA was replaced with an idealized B-form duplex (5'-AGGTGATTGACCACCAC-3') created with 3DNA[22] and after making the requisite chemical changes, local energy minimization was performed (Figure 2.7).[23] The three α-helix homeodomain of Exd is shown bound to the DNA major groove at 5'-TGAT-3' adjacent to the minor groove-binding polyamide at 5'-TGACCA-3' where the dimerizer protein-binding domain, which was previously part of Ubx, is attached by a linker to an internal pyrrole subunit in the hairpin.

Interestingly, the polyamide DNA-binding domain by itself was shown to contribute to the enhancement of Exd's DNA affinity at low temperatures even though there are likely to be no direct contacts between the protein and the polyamide. The DNA minor groove is likely altered by the polyamide binding event, which may affect the adjacent major groove binding site for Exd. Polyamides have been shown to expand the width of the DNA minor groove.[24] The minor groove polyamide binding site is also widened in the Exd/Ubx/DNA structure.[9a] The polyamide-DNA binding event appears to play an *allosteric* role in enhanced Exd binding. Undoubtedly, other changes cannot be ruled out, such as locally dampened structural dynamics, distortion of the nitrogenous bases, and reorientation of

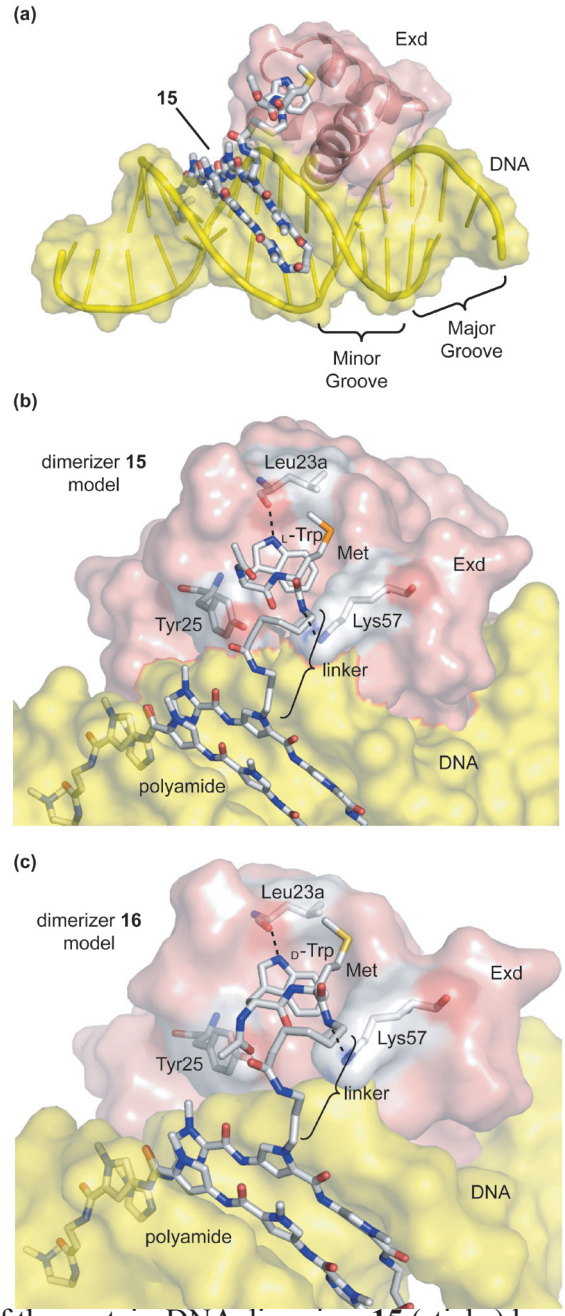


Figure 2.7 (a) Model of the protein-DNA dimerizer **15** (sticks) bound to the minor groove of DNA (yellow cartoon and surface) adjacent to the recruited major groove-binding Exd homeodomain (pink cartoon and surface). Attached by a linker to a polyamide pyrrole residue, the Ac-WM-dipeptide protein-binding domain of **15** projects out of the DNA minor groove and interacts with the Exd binding pocket. **(b)** Closer view of the modeled interaction of **15** (sticks), Exd (transparent red surface), and DNA (transparent yellow surface) centered on the Exd binding pocket. Select Exd homeodomain residues are numbered.[9a] (white = carbon; dark blue = nitrogen; red = oxygen) Images were generated using PyMOL.[34] **(c)** Modeled interaction of **16**, Exd, and DNA with same color coding as in panel b.

the phosphate backbone. We note that a seemingly innocuous propylacetamide linker on a pyrrole subunit appears to eliminate the polyamide's weak energetic contribution, which underscores that small structural changes are at play. It is unclear if a simple dipeptide motif would be functional in the context of a natural protein such as Ubx.

In earlier studies we observed that the DNA binding selectivity for the polyamide-peptide conjugate **1** was reduced relative to the parent hairpin **2**.[4] The DNA specificity of the conjugate was restored by changing both the Trp and Met residues to Ala, which suggested that the Trp or Met might be responsible for the reduced recognition specificity.[4a] It is now clear, however, that a polyamide-peptide conjugate containing Trp and Met residues can exhibit sequence selectivity as evidenced by the footprinting of polyamide-WM conjugate **15**. The diminished DNA binding affinity and specificity of polyamide-FYPWMKG conjugate **1** might be related to the overall size or hydrophobicity of the peptide which suggests minimization may be a way to ensure specificity of DNA binding in the construction of polyamide conjugates and protein-DNA dimerizers in general.

2.4 Conclusion

Small-molecule protein-DNA dimerizers allow the cooperative recruitment of the transcription factor Exd to a specific DNA sequence. Although the polyamide itself was found to contribute to Exd's DNA binding, the dipeptide WM protein binding domain was required for stable complex formation. Covalently linking polyamides to small-molecule protein binders could emerge as a general strategy for the construction of artificial transcription factors.[25-27] The conjugate with the unnatural D-Trp is expected to reduce sensitivity to endogenous proteases and should function at biologically relevant temperatures. Although these minimized dimerizers might be further optimized, for instance through iterative changes to the linker length, composition of the polyamide, and various substitutions to the Trp indole, biologically driven projects utilizing the minimized dimerizers are in progress. Developmental processes in fruit flies have been perturbed

using polyamides.[28] The question arises whether the polyamide-WM dimerizer can interfere specifically with the normal function of Exd leading to distinct fruit fly phenotypes[10a] which will be reported in due course.

2.5 Experimental

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

Materials. Wang resin and Fmoc-protected amino acids were purchased from Novabiochem unless stated otherwise. Super Acid Sensitive Resin (SASRIN) was purchased from Bachem. Boc-β-Ala-PAM resin was purchased from Peptides International. Trifluoroacetic acid (TFA) was purchased from Halocarbon. Molecular biology grade EDTA, potassium glutamate, and acetylated BSA were purchased from Sigma. The compound 3-indolepropionic acid was purchased from Aldrich. DTT was purchased from ICN. Other chemicals were purchased from Aldrich or EM Sciences and used without further purification. All other solvents were purchased from EM Sciences and were reagent grade. Water (18 MΩ) was purified using a Millipore MilliQ water purification system. Biological experiments were performed using Ultrapure Distilled Water (DNase/RNase free) purchased from Gibco. The pH of buffers was adjusted using a Beckman 340 pH/Temp Meter. All buffers were sterilized by filtration through either a Nalgene 0.2 μm cellulose nitrate filtration device or a Pall Acrodisc Syringe Filter HT Tuffryn membrane (0.2 μm). DNA oligonucleotides were obtained from the Biopolymer Synthesis and Analysis Facility at the Beckman Institute, California Institute of Technology. Radiolabeled nucleotides [α-^{32}P]-thymidine-

5'-triphosphate (≥3000 Ci/mmol) and [α-^{32}P]-deoxyadenosine-5'-triphosphate (≥6000 Ci/mmol) were purchased from DuPont/NEN and [γ-^{32}P]-adenosine-5'-triphosphate (≥6000 Ci/mmol) was obtained from ICN. Calf thymus DNA was from Amersham and all enzymes were obtained from Roche.

Methods. UV spectra were recorded in water using a Beckman Coulter DU 7400 Spectrophotometer. All polyamide concentrations were determined using an extinction coefficient ε = 69,500 $M^{-1}\cdot cm^{-1}$ at a λ_{max} near 310 nm. Peptides concentrations were determined using an extinction coefficient of ε = 5,500 $M^{-1}\cdot cm^{-1}$ at λ_{max} equal to 280 nm. Matrix-assisted, LASER desorption/ionization time of flight mass spectrometry (MALDI-TOF MS) was performed using a Applied Biosystems Voyager DE Pro Spectrometer. Electrospray ionization (ESI) mass spectrometry was performed using a Finnigan LCQ ion trap mass spectrometer. Analytical high-pressure liquid chromatography (HPLC) was performed with a Beckman Gold system equipped with a diode array Detector using a Varian Microsorb-MV 300 C_{18} column (8 μm particle size, 250 × 4.6mm). Preparative HPLC was performed with a Beckman Gold system equipped with a single wavelength Detector monitoring at 310 nm using a Waters DeltaPak C_{18} reversed phase column (100 μm, 25 × 100mm). For both analytical and preparative HPLC, solvent A was 0.1% (v/v) aqueous TFA and solvent B was acetonitrile. Compound purity was assessed using an HPLC analytical method employing a gradient of ~4.12% B/min starting from 10% B with a flow rate of 1.5 mL/min. Initial preparative HPLC was performed with a gradient of to 60% B over 90 min. If the purity of the compound was less than 95%, a second HPLC purification was performed employing a slow gradient through the previous % B elution of the compound.

Synthesis. Polyamide monomers were prepared as described previously.[29,30] Conjugate **1** was synthesized as described previously.[4a] **1:** (MALDI-TOF) $[M+H]^+$ calc. for

$C_{108}H_{137}N_{32}O_{19}S^+$ 2218.1, observed 2218.0**.** Polyamides **2**, **3** and **6** were synthesized on PAM resin using published BOC-based protocols and purified by reverse phase HPLC.[29] **2:** (MALDI-TOF) $[M+H]^+$ calc. for $C_{57}H_{71}N_{22}O_{10}^+$ 1223.6, observed 1223.6. Polyamide **3:** (MALDI-TOF) $[M+H]^+$ calc. for $C_{58}H_{72}N_{21}O_{10}^+$ 1222.6, observed 1222.6. Polyamide **4:** (MALDI-TOF) $[M+H]^+$ calc. for $C_{61}H_{78}N_{23}O_{11}^+$ 1308.6, observed 1308.7. Peptide **5** was synthesized by coupling 2 equivalents of propanolamine with 1.1 equivalents HBTU and 1.1 equivalents DIEA in DMF to a previously reported protected peptide[4a] followed by deprotection and purification by reverse phase HPLC as described. Peptide **5**: (ESI, pos.) $[M+H]^+$ calc. for $C_{52}H_{71}N_{10}O_{10}S^+$ 1027.5, observed 1027.5. Polyamide **6** (R = H): (MALDI-TOF) calc. for $C_{59}H_{76}N_{23}O_{10}^+$ 1266.6, observed 1266.5. The peptides **7**-**12** were synthesized by manual solid phase synthesis on Wang resin using standard Fmoc chemistry.[31] **7:** (ESI, neg.) $[M-H]^-$ calc. for $C_{24}H_{33}N_4O_5S^-$ 489.2, observed 489.1. **8:** (ESI, neg.) $[M-H]^-$ calc. for $C_{24}H_{33}N_4O_5S^-$ 489.2, observed 489.1. **9:** (ESI, neg.) $[M-H]^-$ calc. for $C_{22}H_{32}N_3O_5S^-$ 450.2, observed 450.1. **10:** (ESI, neg.) $[M-H]^-$ calc. for $C_{16}H_{28}N_3O_5S^-$ 374.2, observed 374.1. **11:** (ESI, neg.) $[M-H]^-$ calc. for $C_{22}H_{29}N_4O_5^-$ 429.2, observed 429.1. **12:** (ESI, neg.) $[M-H]^-$ calc. for $C_{21}H_{27}N_4O_5^-$ 415.2, observed 415.1. The peptides **13**-**14** were synthesized by manual solid phase synthesis on SASRIN resin using standard Fmoc chemistry acylating the N-terminal residue with 3-indolepropionic acid activated with HBTU *in situ*. **13:** (ESI, neg.) $[M-H]^-$ exact mass calc. for $C_{22}H_{30}N_3O_4S^-$ 432.2, observed 432.1. **14:** (ESI, neg.) $[M-H]^-$ exact mass calc. for $C_{19}H_{24}N_3O_4^-$ 358.2, observed 358.1. Following cleavage from resin, crude peptides **7**-**14** were activated with HBTU and coupled to polyamide **6** (R=H) to generate polyamide conjugates **15**-**22**, respectively. Then conjugates **15**-**22** were purified by reverse phase HPLC to ≥ 95% purity (UV). Mass spectra of **15**, **16**, **17**, **18**, and **21** displayed the presence of small amounts (<5%) of methionine sulfoxide. **15:** (MALDI-TOF) $[M+H]^+$ calc. for $C_{83}H_{108}N_{27}O_{14}S^+$ 1738.8, observed 1738.9. **16:** (MALDI-TOF) $[M+H]^+$ calc. for $C_{83}H_{108}N_{27}O_{14}S^+$ 1738.8, observed 1738.8. **17:** (MALDI-TOF) $[M+H]^+$ calc. for $C_{81}H_{107}N_{26}O_{14}S^+$ 1699.8, observed 1699.9. **18:** (MALDI-TOF) $[M+H]^+$

calc. for $C_{75}H_{103}N_{26}O_{14}S^+$ 1623.8, observed 1623.9. **19:** (MALDI-TOF) $[M+H]^+$ calc. for $C_{81}H_{104}N_{27}O_{14}^+$ 1678.8, observed 1678.9. **20:** (MALDI-TOF) $[M+H]^+$ calc. for $C_{80}H_{102}N_{27}O_{14}^+$ 1664.8, observed 1664.9. **21:** (MALDI-TOF) $[M+H]^+$ calc. for $C_{81}H_{105}N_{26}O_{13}S^+$ 1681.8, observed 1681.8. **22:** (MALDI-TOF) $[M+H]^+$ calc. for $C_{78}H_{99}N_{26}O_{13}^+$ 1607.8, observed 1607.8.

Protein expression and characterization. The *Drosophila Melanogaster* extradenticle (Exd) homeodomain and fourth extended helix (residues 238-320)[9] was expressed and FPLC purified following the precedent of Passner and Aggarwal and stored in 25% glycerol at -20 °C.[9] The protein domain identity was confirmed by selective protease cleavage (Lys-C achromobacter peptidase) and MALDI-TOF MS, identifying 8 of 9 predicted fragments. MALDI-TOF MS of the undigested protein fragment was consistent with the isolation of Exd residues 238-320 (calculated mass, 9495.7 *m*/*Z*; observed mass, 9496.0 *m*/*Z*). Protein concentration was determined by measuring the UV absorption at 280 nm using a calculated[32] extinction coefficient of ε = 12,600 $M^{-1}\cdot cm^{-1}$. See appendix C for details of the expression procedure and protein characterization.

Preparation of 3'-labeled DNA for DNase I footprinting. Plasmids pHDA1 or pDEH9 (5 μg) were digested using restriction enzymes *Pvu*II and *Eco*RI to yield the crude 250 base pair or 308 base pair restriction fragments.[4a] The 3'-overhang regions were filled with [α-^{32}P]-dATP and [α-^{32}P]-dTTP using the Klenow fragment of DNA polymerase or Sequenase. The radiolabeled oligonucleotides were purified by band extraction from a preparative 6% non-denaturing polyacrylamide gel, diluted to 10 kcpm/μL, stored at -78 °C and used within 2 weeks.

Preparation of 5'-labelled DNA for MPE footprinting. A 5'-^{32}P-end labeled (see below) primer 5'-ttc aca cag gaa aca gct atg acc-3' and an unlabelled primer 5'-cgg gga tcc ata

cat gat tt-3' were used to amplify a 5'-lower strand labeled PCR-fragment from pHDA1 following published procedures.[14] The radiolabeled oligonucleotides were purified by band extraction from a preparative 6% non-denaturing polyacrylamide gel, diluted to 10 kcpm/μl, stored at -78 °C, and used within 2 weeks.

Preparation of ^{32}P-labeled synthetic oligonucleotides for EMSA. Synthetic oligonucleotides were either labeled by (a) 3'-overhang fill-in or (b) 5' kinase end labeling. Three synthetic oligonucleotides were used: (1) UExdOpt: 5'-gatctcccgg cgaatgattg accatatcgg cgccactgtc acccgga-3', (2) LExdOpt: 5'-tccgggtgac agtggcgccg atatggtcaa tcattcgccg ggagatc-3', and (3) LSExdOpt: 5'-tccgggtgac agtggcgccg atatggtcaa tcattcgccg ggag-3'. Annealing was accomplished by mixing equimolar quantities in water, heating to 95 °C for 5 min, and slowly cooling to room temperature over 2 hours. 3' fill-in was accomplished by annealing UExdOpt and LSExdOpt DNA (~60 pmol) followed by treatment with [α-^{32}P]-ATP and [α-^{32}P]-dTTP and Klenow fragment of DNA polymerase. 5' end labeling was accomplished by annealing UExdOpt and LExdOpt followed by labeling with [γ-^{32}P]-ATP and polynucleotide kinase using standard protocols. Both 3'- and 5'-labeled oligonucleotides were then purified by gel electrophoresis and band extraction from a 6% non-denaturing polyacrylamide gel, diluted to 10 kcpm/μl, stored at -78 °C, and used within 2 weeks.

Electrophoretic Mobility Shift Assay (EMSA). EMSA experiments were performed as described previously.[4a] Polyamide conjugates were pre-incubated with ≤5 kcpm/reaction of ^{32}P-labeled oligonucleotides at room temperature (~20 °C) in buffer for 1 hour. Exd protein was then added to the reaction solution in a cold room (~4 °C), at room temperature (~20 °C), or in an incubator (~37 °C) for 1 more hour. The final concentrations were 30 mM potassium glutamate, 10 mM HEPES, 0.2 mM DTT, 10% glycerol, and 0.1 mg/mL BSA (Sigma). 10 μL of the binding solution was loaded onto a non-denaturing 9%

polyacrylamide (29:1 acrylamide:bisacrylamide) / 3% glycerol / 1X TBE (89 mM tris base, 89 mM boric acid, 2 mM EDTA, pH 8.3) 0.8 mm thick gel and developed at 180-190 V and 13-15 mA for 1.5 hours pre-equilibrated to the same temperature in which the final Exd incubation was performed (i.e., 4 °C, 20 °C, or 37 °C). Gels were then dried and exposed to a phosphor screen overnight, which were scanned with a Molecular Dynamics 400S Phosphorimager. All experiments were done at least in triplicate.

The amount of Exd bound to the DNA template, i.e., the apparent fractional occupancy (Θ_{app}), was calculated from the measured intensity volumes of the shifted ($V_{shifted}$) and the unshifted bands ($V_{unshifted}$):

$$\Theta_{app} = V_{shifted} / (V_{shifted} + V_{unshifted}). \quad (1)$$

The data were then fitted to the Hill equation:[33]

$$\Theta_{fit} = \Theta_{min} + (\Theta_{max} - \Theta_{min})[(K_a^n[\text{Exd}]^n)/(1 + K_a^n[\text{Exd}]^n)], \quad (2)$$

where [Exd] is the Exd concentration, K_a is the equilibrium association constant, n is the Hill constant, and Θ_{min} and Θ_{max} are the experimentally determined Θ_{app} values where binding is absent or maximally saturated, respectively. KaleidaGraph software was used to fit the data to equation (2) using a least-squares fitting procedure minimizing the difference between Θ_{app} and Θ_{fit} with K_a, Θ_{max}, and Θ_{min} as the variable parameters. Fitting the data to equation (2) with free n did not increase the quality ($R \geq 0.98$ in either case), so $n = 1$ was kept for all analysis.

Affinity determination by quantitative DNase I footprinting. Reactions were carried out in a volume of 400 μL in aqueous TKMC buffer according to published protocols[14] using a 3' ^{32}P-labeled 250 base pair *Eco*RI/*Pvu*II restriction fragment of the plasmid pDEH9.[4a] Developed gels were imaged using storage phosphor autoradiography using a Molecular Dynamics 400S Phosphorimager. Equilibrium association constants were determined as previously described.[14]

MPE footprinting in the presence of Exd protein. MPE footprinting was carried out in 40 μL reaction volumes according to published procedures.[14] Reactions were carried out in a total volume of 40 μL in aqueous TKMC buffer containing 10% glycerol using the 3' ^{32}P-labeled 308 base pair *Eco*RI/*Pvu*II restriction fragment of the plasmid pHDA1. DNA was pre-incubated with dimerizer overnight at 20 °C. in 4 μL glycerol, 8 μL 5x TKMC, 22 μL DNA (20 kcpm), 4 μL dimerizer (10x). Exd protein (0.16 μM, 1 μL) was introduced, gently mixed, and the mixture was incubated for 30 min at 4 °C before digestion was initiated. Final concentrations were 25 mM TrisOAc, 10 mM NaCl, 100 μM base pair carrier DNA (calf thymus), 10 mM DTT, 10 μM MPE x $FeSO_4$, 10% glycerol, pH 7.0. Cleavage reactions were stopped after 10 min.

Acknowledgements. This work was supported by the National Institute of Health research grants to P. B. D. and A. Z. A. A DAAD postdoctoral fellowship to H. D. A. is gratefully acknowledged. We thank Professor A. L. Aggarwal (Mount Sinai School of Medicine, N.Y.) for a donation of the plasmid encoding the Exd DNA binding domain, and Drs. Peter Snow and Gary Hathaway (Beckman Institute, California Institute of Technology) for technical assistance with protein expression, purification and identification.

2.6 References and Footnotes

(1) (a) Ptashne, M. and Gann, A. *Genes and Signals*: Cold Spring Harbor Lab. Press, Plainview, NY, 2002. (b) Mammen, M., Choi, S.-K., Whitesides, G. M. *Angew Chem. Intl Ed.* **1998**, *37*, 2754-2794. (c) Kiessling, L. L., Gestwicki, J. E., Strong, L. E. *Curr. Opin. Chem. Biol.* **2000**, *4*, 696-703. (d) Harmer, N. J., Chirgadze, D., Hyun Kim, K., Pellegrini, L., Blundell, T. L. *Biophys. Chem.* **2003**, *100*, 545-553.

(2) Klemm, J. D., Schreiber, S. L., Crabtree, G. R. *Annu. Rev. Immunol.* **1998**, *16*, 569-592.

(3) (a) Austin, D. J., Crabtree, G. R., Schreiber, S. L. *Curr. Biol.* **1994**, *1*, 131-136. (b) Farrar, M. A., Alberola-Ila, J., Perlmutter, R. M. *Nature (London)* **1996**, *383*, 178-181. (c) Choi, J., Chen, J., Schreiber, S. L., Clardy, J. *Science* **1996**, *273*, 239-242. (d) Schreiber, S. L. *Bioorg. Med. Chem.* **1998**, *6*, 1127-1152. (e) Althoff, E. A. and Cornish, V. W. *Angew. Chem.* **2002**, *114*, 2433-2436. *Angew. Chem.-Int. Edit.* **2002**, *41*, 2327-2330.

(4) (a) Arndt, H. D., Hauschild, K. E., Sullivan, D. P., Lake, K., Dervan, P. B., Ansari, A. Z. *J. Am. Chem. Soc.* **2003**, *125*, 13322-13323. (b) Hauschild, K. E., Metzler, R. E., Arndt, H.-D., Moretti, R., Raffaelle, M., Dervan, P. B., Ansari, A. Z. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 5008-5013. (c) Warren, C. L., Kratochvil, N. C. S., Hauschild, K. E., Foister, S., Brezinski, M. L., Dervan, P. B., Phillips, G. N., Jr., Ansari, A. Z. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 867-872.

(5) (a) Denison, C. and Kodadek, T. *Chem. Biol.* **1998**, *5*, R129-R145. (b) Ansari, A. Z. and Mapp, A. K. *Curr. Opin. Chem. Biol.* **2002**, *6*, 765-772. (c) Arndt, H.-D. *Angew. Chem. Int. Ed.* **2006**, *45*, 4552-4560.

(6) Protein-protein dimerizers have been employed in engineered systems as a means for exercising external control of gene expression of exogenously introduced genes.[7]

(7) (a) Pollock, R. and Clackson, T. *Curr. Opin. Biotechnol.* **2002**, *13*, 459-467. (b)

Minter, A. R., Brennan, B. B., Mapp, A. K. *J. Am. Chem. Soc.* **2004**, *126*, 10504-10505. (c) Liu, B., Alluri, P. G., Yu, P.; Kodadek, T. *J. Am. Chem. Soc.* **2005**, *127*, 8254-8255.

(8) (a) Dervan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 2215-2235. (b) Dervan, P. B. and Edelson, B. S. *Curr. Opin. Struct. Biol.* **2003**, *13*, 284-299.

(9) (a) Passner, J. M., Ryoo, H. D., Shen, L. Y., Mann, R. S., Aggarwal, A. K. *Nature* **1999**, *397*, 714-719. (b) Piper, D. E., Batchelor, A. H., Chang, C. P., Cleary, M. L., Wolberger, C. *Cell* **1999**, *96*, 587-597. (c) LaRronde-LeBlanc, N. A. and Wolberger, C. *Genes Dev.* **2003**, *17*, 2060-2072.

(10) (a) Peifer, M. and Wieschaus, E. *Genes Dev.* **1990**, *4*, 1209-1223. (b) Chan, S. K., Jaffe, L., Capovilla, M., Botas, J., Mann, R. S. *Cell* **1994**, *78*, 603-615. (c) van Dijk, M.A. and Murre, C. *Cell* **1994**, *78*, 617-624. (d) Mann, R. S. and Chan, S. K. *Trends Genet.* **1996**, *12*, 258-262.

(11) Duboule, D. Editor In *Guidebook to the Homeobox Genes*; Oxford University Press: Oxford, 1994.

(12) (a) Neuteboom, S. T. C., Peltenburg, L. T. C., van Dijk, M. A., Murre, C. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 9166-9170. (b) Knoepfler, P. S. and Kamps, M. P. *Mol. Cell. Biol.* **1995**, *15*, 5811-5819. (c) Lu, Q. and Kamps, M. P. *Mol. Cell. Biol.* **1996**, *16*, 1632-1640. (d) Shanmugam, K., Featherstone, M. S., Saragovi, H. U. *J. Biol. Chem.* **1997**, *272*, 19081-19087.

(13) Shen, W. F., Chang, C. P., Rozenfeld, S., Sauvageau, G., Humphries, R. K., Lu, M., Lawrence, H. J., Cleary, M. L., Largman, C. *Nucleic Acids Res.* **1996**, *24*, 898-906.

(14) Trauger, J. W. and Dervan, P. B. *Methods Enzymol.* **2001**, *340*, 450-466.

(15) These affinities lie within 2-fold of those reported previously.[4a] The measured difference in binding affinity may be due to variation of Exd activities when using different protein preparations.

(16) Marques, M. A., Doss, R. M., Urbach, A. R., Dervan, P. B. *Helv. Chim. Acta* **2002**, *85*, 4485-4517.

(17) Although there are reports of "caging" effects in gel shifts (i.e., protein-DNA complexes with dissociation half-lives that are shorter than the time it takes to run a gel often form discrete super-shifted bands) the experiments reported here have not been optimized to stabilize weak complexes. Under the conditions optimized for our initial dimerizer (**1**) the complex formed in the presence of **2** is not stable, thus we do not make a definitive K_a assignment.

(18) Gottesfeld, J. M., Melander, C., Suto, R. K., Raviol, H., Luger, K., Dervan, P. B. *J. Mol. Biol.* **2001**, *309*, 615-629.

(19) In earlier studies, we investigated the role of the linker that connects the polyamide to the peptide.[4b] We find that a substantial degree of variability in the linker length is tolerated at lower temperatures. At physiological temperatures, the longest linker tested confers a "switch"-like property on the protein-DNA dimerizer, in that it abolishes the ability of the YPWM moiety to recruit the natural transcription factor to DNA.

(20) (a) Chakrabarti, P. and Janin, J. *Proteins* **2002**, *47*, 334-343. (b) Lo Conte, L., Chothia, C., Janin, J. *J. Mol. Biol.* **1999**, *285*, 2177-2198.

(21) Kielkopf, C. L., Baird, E. E., Dervan, P. D., Rees, D. C. *Nat. Struct. Biol.* **1998**, *5*, 104-109.

(22) Lu, X. J. and Olson, W. K. *Nucleic Acids Res.* **2003**, *31*, 5108-5121.

(23) The illustrative models shown in Figure 2.7 were created from the atomic coordinates for Exd (PDB code 18BI) and the polyamide (PDB code 365D) derived from the crystal structures and superimposed with a "Fit" algorithm on the DNA backbone using SwissPDB Viewer (Ver. 3.7). The DNA coordinates were derived from 3DNA software using the Fiber program with ideal B-form DNA parameters and superimposed as above. The DNA and Ubx, except for

the WM motif, from the previous structures were then deleted with SwissPDB Viewer. The linker between the polyamide and the WM, the γ-aminobutyric acid of the polyamide hairpin, and the *N*-terminal acetyl moiety were constructed and minimized using an augmented MM3 force-field with all other coordinates locked using CAChe Workspace (Ver. 6.1.10). Finally, the minor groove width was relaxed at the polyamide binding site through a short (<10 ps) MD simulation in CAChe on the bases GGTC and CCAG with all other coordinates locked.

(24) (a) Kielkopf, C. L., White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., Dervan, P. B., Rees, D. C. *Science* **1998**, *282*, 111-115. (b) Kielkopf, C. L., Bremer, R. E., White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., Dervan, P. B., Rees, D. C. *J. Mol. Biol.* **2000**, *295*, 557-567. (c) Suto, R. K., Edayathumangalam, R. S., White, C. L., Melander, C., Gottesfeld, J. M., Dervan, P. B., Luger, K. *J. Mol. Biol.* **2003**, *326*, 371-380. (d) Zhang, Q., Dwyer, T. J., Tsui, V., Case, D. A., Cho, J. H., Dervan, P. B., Wemmer, D. E. *J. Am. Chem. Soc.* **2004**, *126*, 7958-7966. (e) Edayathumangalam, R. S., Weyermann, P., Gottesfeld, J. M., Dervan, P. B., Luger, K. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 6864-6869.

(25) In a formal sense, dimerizers equipped with DNA and protein binding surfaces could induce specific proximal interactions between DNA and proteins other than transcription factors, such as co-activators,[26] and DNA-modifying enzymes.[27]

(26) (a) Kuznetsova, S., Ait-Si-Ali, S., Nagibneva, I., Troalen, F., Le Villain, J., Harel-Bellan, A., Svinarchuk, F. *Nucl. Acids Res.* **1999**, *27*, 3995-4000. (b) Mapp, A. K., Ansari, A. Z., Ptashne, M., Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 3930-3935. (c) Ansari, A. Z., Mapp, A. K., Nguyen, D. H., Dervan, P. B., Ptashne, M. *Chem. Biol.* **2001**, *8*, 583-592. (d) Liu, B., Han, Y., Corey, D. R., Kodadek, T. *J. Am. Chem. Soc.* **2002**, *124*, 1838-1839. (e) Stanojevic, D. and Young, R. A. *Biochemistry* **2002**, *41*, 7209-7216. (f) Arora, P. S., Ansari, A. Z., Best, T. P., Ptashne, M., Dervan, P. B. *J. Am. Chem. Soc.* **2002**, *124*, 13067-13071. (g) Liu,

B., Han, Y., Ferdous, A., Corey, D. R., Kodadek, T. *Chem. Biol.* **2003**, *10*, 909-916. (h) Kwon, Y., Arndt, H.-D., Mao, Q., Choi, Y., Kawazoe, Y., Dervan, P. B., Uesugi, M. *J. Am. Chem. Soc.* **2004**, *126*, 15940-15941.

(27) (a) Matteucci, M., Lin, K.-Y., Huang, T., Wagner, R., Sternbach, D. D., Mehrotra, M., Besterman, J. M. *J. Am. Chem. Soc.* **1997**, *119*, 6939-6940. (b) Wang, C. C. and Dervan, P. B. *J. Am. Chem. Soc.* **2001**, *123*, 8657-8661. (c) Arimondo, P. B., Bailly, C., Boutorine, A. S., Ryabinin, V. A., Syniakov, A. N., Sun, J.-S., Garestier, T., Helene, C. *Angew. Chem. Int. Ed.*. **2001**, *40*, 3045-3048.

(28) (a) Janssen, S., Durussel, T., Laemmli, U. K. *Mol. Cell.* **2000**, *6*, 999-1011. (b) Janssen, S., Cuvier, O., Muller, M., Laemmli, U. K. *Mol. Cell.* **2000**, *6*, 1013-1024.

(29) Baird, E. E. and Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141-6146.

(30) Rucker, V. C., Foister, S., Melander, C., Dervan, P. B. *J. Am. Chem. Soc.* **2003**, *125*, 1195-1202.

(31) Wellings, D. A. and Atherton, E. *Methods Enzymol.* **1997**, *289*, 44-67.

(32) Gill, S. C. and von Hippel, P. H. *Anal. Biochem.* **1989**, *182*, 319-326.

(33) Cantor, C. R. and Schimmel, P. R. *Biophysical Chemistry of Macromolecules, Pt. 3: The Behavior of Biological Macromolecules*; W.H. Freeman and Company: New York, 1980.

(34) Delano, W. L. **The PyMOL Molecular Graphics System** (2002) DeLano Scientific, San Carlos, CA, USA. http://www.pymol.org.

Chapter 3

Defining the Reach of Linear Protein-DNA Dimerizers

The text of this chapter was taken in part from a manuscript coauthored with Peter B. Dervan (Caltech).

(Stafford, R. L, Dervan, P. B. "The Reach of Linear Protein-DNA Dimerizers" *submitted to J. Am. Chem. Soc.* **2007**)

Abstract

A protein-DNA dimerizer constructed from a DNA-binding pyrrole-imidazole polyamide and the peptide FYPWMK facilitates binding of the natural transcription factor Exd to an adjacent DNA site. Previous dimerizers have been constructed with the peptide attached to an internal pyrrole monomer in an overall branched oligomer. Linear oligomers constructed by attaching the peptide to the polyamide C-terminus expand the range of protein-DNA dimerization to six additional DNA sites. Replacing the FYPWMK hexapeptide with a WM dipeptide, which was previously functional in branched compounds, does not lead to a functional linear dimerizer. Instead, inserting an additional lysine generates a minimal, linear WMK tripeptide conjugate that maintains the activity of the larger FYPWMK dimerizers in a single DNA-binding site orientation. These studies provide insight into the importance of linker length and composition, binding site spacing and orientation, and the protein-binding domain content that are important for the optimization of protein DNA-dimerizers suitable for biological experiments.

3.1 Introduction

Transcription in living cells is controlled by precise spatial and temporal recognition of DNA by protein transcription factors.[1] These proteins often bind to DNA as homo- and heterodimers or larger complexes which increases their overall DNA-binding site size, specificity, and affinity.[2] Transcription factors are usually modular, whereby the DNA-binding and dimerization, or protein-binding domains, are functionally and structurally independent.[3] Small molecules called artificial transcription factors have been reported to mimic the modular components of these endogenous proteins.[4] Protein-DNA dimerizers are a type of artificial transcription factor which facilitate binding of proteins to an adjacent DNA site, mimicking natural multiprotein-DNA complexes.[5] Our long-term goal is to ask whether protein-DNA dimerizers can serve as artificial regulators of gene expression in a living cell. From a design point of view, we are exploring what is the minimum size and shape (branched or linear) that will be optimal for cell uptake with adequate functional potency in the transcriptional apparatus. It has been previously shown that *branched* dimerizers facilitate protein binding at three different binding sites (Figure 3.1a).[5c] Here, we describe *linear* structural motifs which expand the number of recognition sites targetable by protein-DNA dimerizers. The optimal DNA site spacings and necessary protein-binding elements have also been determined, which culminated in the discovery of a minimal, three amino acid protein-binding domain that yields strong dimerization at DNA-binding sites reaching 6 base pairs apart.

Protein-DNA dimerizers can mimic homeobox (HOX) transcription factors by binding cooperatively to DNA with the protein extradenticle (Exd).[5] The conserved class of HOX transcription factors is known to bind relatively short DNA sites by themselves, but in complex with cofactor proteins, they bind larger DNA sequences.[6] The increase in their composite DNA recognition site is thought to be important for their function in developmental pathways.[6] For instance in *D. melanogaster,* the HOX protein ultrabithorax (Ubx) binds the site 5′-TTAT-3′ alone, but in complex with Exd the composite site increases

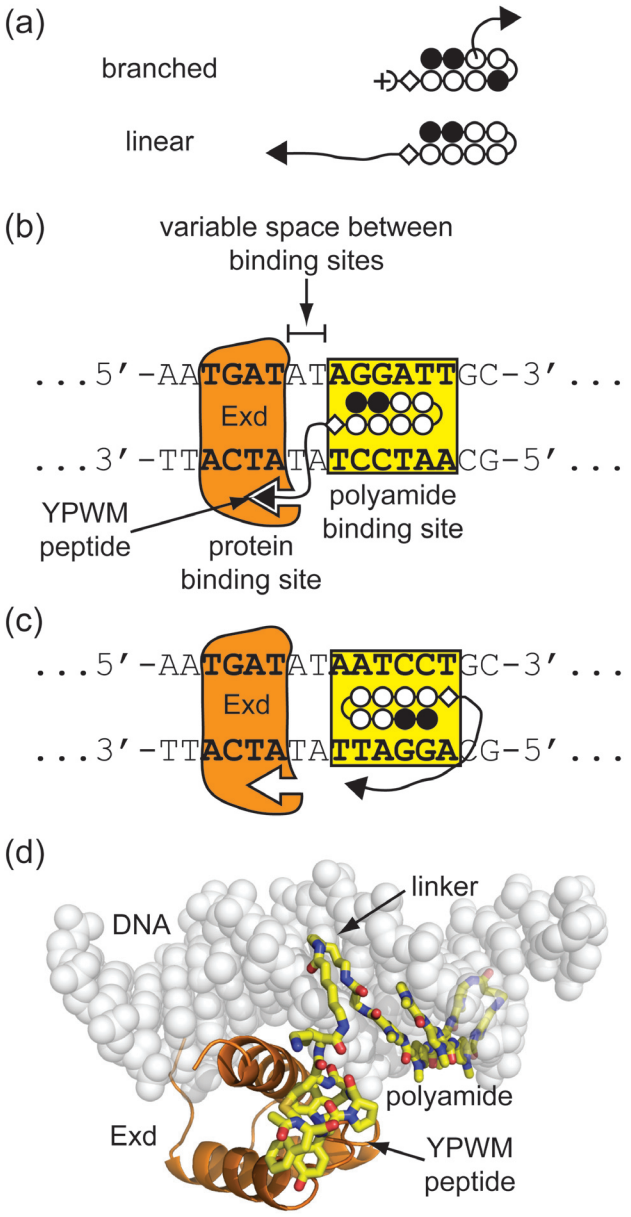


Figure 3.1 Design of linear protein-DNA dimerizers. (a) Comparison of branched polyamide-peptide conjugates to linear conjugates, both used as protein-DNA dimerizers for Exd. (open circles = *N*-methylpyrrole, dark circles = *N*-methylimidazole, half circle = γ-aminobutyric acid, diamond = β-alanine, half circle with plus = *N*,*N*-dimethylaminopropylamine, freehand line = linker domain, and black triangle = protein-binding domain). (b) Diagram of a linear conjugate in the *proximal* orientation with the protein-binding domain directed toward Exd's DNA-binding site. (c) Diagram of a linear conjugate in the *distal* orientation with the protein-binding domain directed away from Exd's DNA-binding site. (d) An illustrative model of the linear protein-DNA dimerizer **1** in complex with Exd and DNA as shown schematically in panel b.

by four base pairs to 5′-TGATTTAT-3′.[7a] A crystal structure of the ternary Ubx/Exd/DNA complex revealed how both Ubx and Exd's major groove binding homeodomains interact via Ubx's N-terminal YPWM peptide motif.[7a] The conserved YPWM motif spans the DNA minor groove between the two proteins and inserts a hydrophobic tryptophan into the complementary Exd binding pocket. Additional crystal structures of HoxB1/Pbx/DNA and HoxA9/Pbx/DNA which contain human homologues of Ubx and Exd have also been reported in which the conserved tryptophan of the YPWM motif mediates the majority of the protein-protein contacts.[7b-c] Protein-DNA dimerizers have been constructed by replacing the natural DNA-binding homeodomain of Ubx with a synthetic pyrrole-imidazole polyamide which binds the DNA minor groove.[5] Polyamides are ideal small-molecule DNA binding domains since they can be programmed to bind to a broad repertoire of sequences using an aromatic amino acid pairing code.[8]

In this study we explore the range of DNA sites in which Exd-DNA dimerization is observed in the presence of linear polyamide-peptide conjugates using a series of DNA duplexes containing incremental A,T base pair spacings between the polyamide and Exd binding sites. Linear oligomers were generated by attaching an FYPWMK peptide to the polyamide C-terminus incorporating linkers of different lengths and composition, i.e., **1** and **2** (Figure 3.2). Once the scope of recognition sequences with these larger constructs was defined, efforts were made to discover a minimal protein-binding domain for dimerizers in this linear configuration. It is shown that WM dipeptides (with both L and D stereochemistry of the W) which previously enhanced Exd-DNA binding in the branched oligomer[5d] failed to yield stable Exd-DNA complexes. Exd dimerization activity was recovered by inserting a lysine back into the peptide to generate minimal WMK tripeptide conjugates. The presence of a positively charged ammonium ion in the linker and lysine seem to be favorable for complex formation possibly due to interactions with the negatively charged DNA backbone.

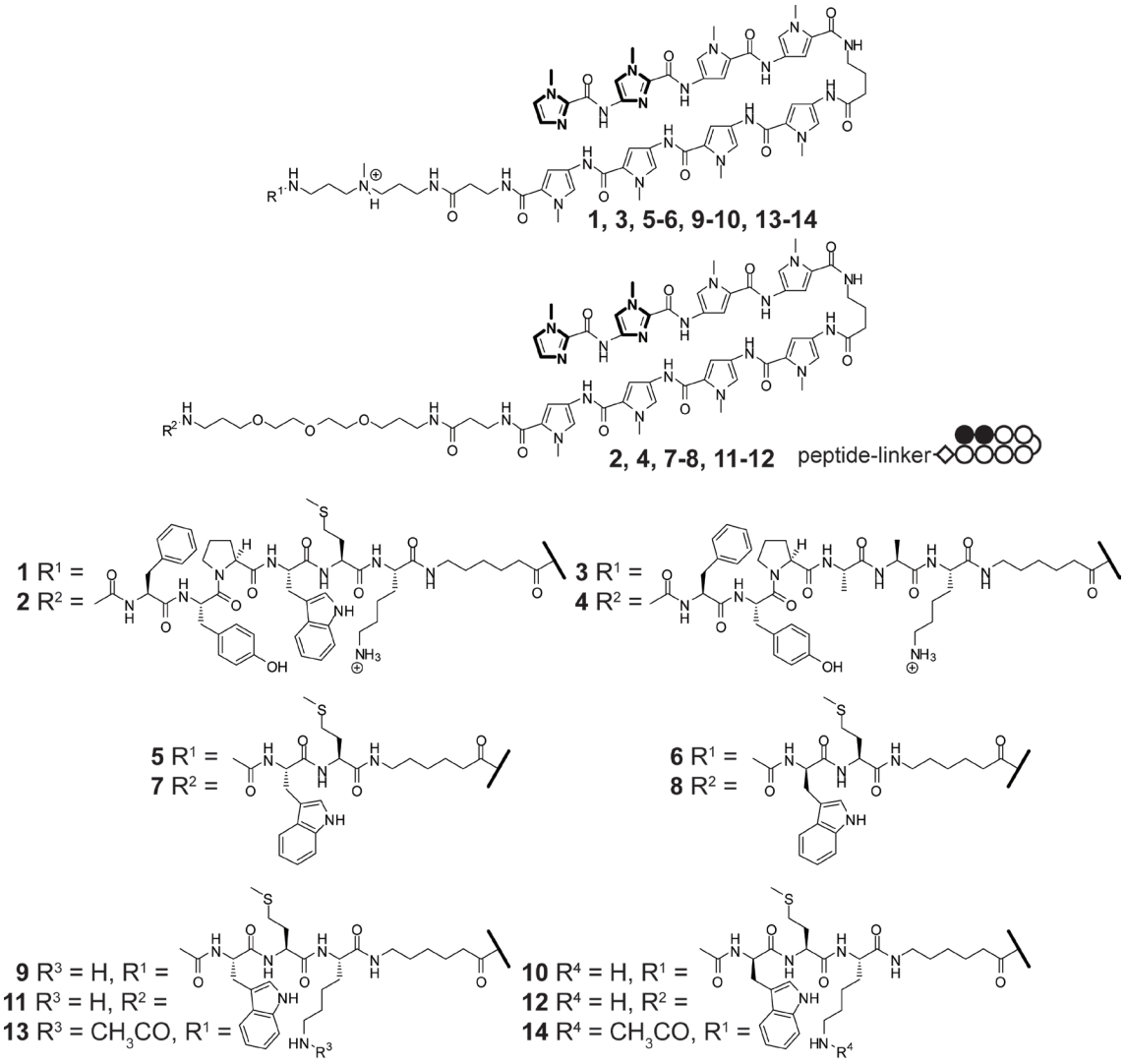


Figure 3.2 Structures of compounds **1** through **14**.

3.2 Results and Discussion

Dimerization in Proximal and Distal Orientations

Exd is expected to bind 5′-TGAT-3′ as found in the Ubx/Exd/DNA crystal structure[7a] and the hairpin polyamide ImImPyPy-γ-PyPyPyPy-β- codes for the sequence 5′-WGGWWW-3′ (where W = A or T).[8] In the proximal orientation, the polyamide binding site is oriented such that the protein-binding YPWM peptide is projected adjacent to Exd (Figure 3.1b). In the distal orientation, the polyamide binding site was inverted such that

the peptide is projected away from Exd (Figure 3.1c). For dimerization of Exd at a given DNA site to occur, the protein-binding peptide must project over the DNA phosphate backbone and interact favorably with the binding pocket of the major groove binding Exd protein (Figure 3.1d). Dimerization was expected to be dependent on both the length and composition of the linker between the polyamide and the YPWM-containing peptide as well as the size of the spacer region between the polyamide's and protein's binding sites.

Accordingly, two polyamide-peptide conjugates were constructed that incorporated either a tertiary amine (**1**) or a PEG linker (**2**). The distance between the terminal amine of the linker to the carbonyl carbon of the methionine of the peptide was determined computationally in the gas-phase for MM2-energy minimized linkers to be approximately 23 and 30 Å for **1** and **2**, respectively. A series of DNA duplexes which contained an incremental A,T spacer region were screened by electrophoretic mobility shift (EMSA), or gel shift assays, to determine the range of protein recruitment to the DNA at room temperature (20 °C). A total of nine different duplexes in the proximal orientation with spacer regions between -1 and 7 bps were used, covering distances between the polyamide and expected protein-binding domain position from approximately 11 to 30 Å. Negative spacer regions indicate an overlap in the expected binding sites.

It might be expected that since **2** contains a longer linker than **1**, it would facilitate Exd binding at sites further away from the polyamide binding site. Instead, it was observed that **1** dimerizes Exd effectively on composite sites containing spacer regions between 2 to 6 base pairs, whereas **2** only dimerizes Exd on sites with spacer regions of 4 to 5 base pairs (Figure 3.3 and Table 3.1). Thus, compound **2** has a narrower recognition profile, in contrast to **1** which is more promiscuous. The greater tolerance for multiple binding sites by compound **1** may be due to its chemical composition. It is possible that the presence of the positively charged ammonium ion in **1** may interact more favorably with the phosphate backbone than the PEG linker of **2**. Control experiments at all of these sites show that Exd by itself, or in the presence of parent polyamide ImImPyPy-γ-PyPyPyPy-β-Dp **15**, or the

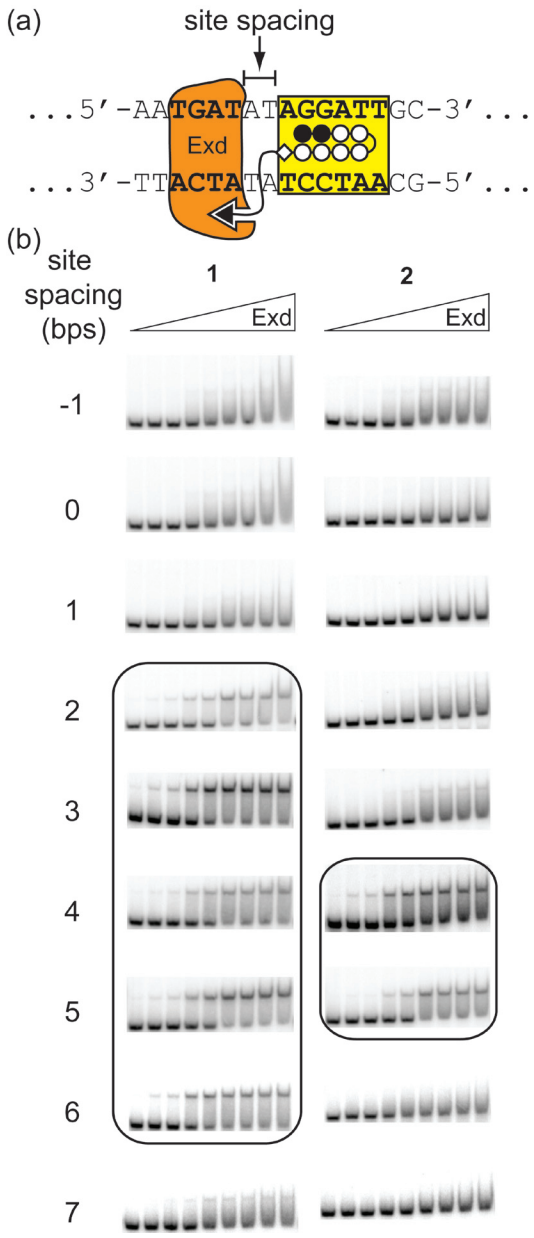


Figure 3.3 Representative gel shift results for Exd in the presence of linear protein-DNA dimerizers with differently sized linkers on two *proximally* oriented DNA sites. Conjugate **1** has an amine-containing linker (~23 Å long) and conjugate **2** has a PEG-type linker (~30 Å long). (a) Schematic of the positions of the Exd and polyamide DNA-binding sites showing the A,T spacer region which was varied. (b) Gel shift experiments show that **1** forms a stable complex with Exd and DNA with site spacings between 2-6 bps, whereas **2** only forms stable complexes from 4-5 bps. A concentration of 50 nM of the indicated compound was added to all lanes and Exd was titrated from 10 pM to 100 nM.

Table 3.1 Summary of select proximal orientation gel shifts[a]

Base pairs between binding sites	Estimated proximal distance (Å)[b]	Exd-DNA binding with **1**	Exd-DNA binding with **2**
-1	23.8	-	-
0	20.0	-	-
1	15.0	-	-
2	11.2	+	-
3	11.2	+	**w**
4	14.2	+	+
5	19.4	+	+
6	25.4	+	-
7	29.8	**w**	-

[a] A '+', 'w', and '-' indicate a strong, weak, or no complex, respectively. Binding constants (K_a's) for '+' entries varied between 0.8 ± 0.3 to 2.7 ± 1.0 M^{-1} (Figure F.5).
[b] measured from the Met C to the linker N

double alanine mutants **3** and **4**, does not lead to stable complex formation (Figure F.1). Thus, the polyamide domain by itself does not appear to contribute to the binding of Exd through allosteric interactions as observed in previously reported systems.[5d] Furthermore, these control experiments show that dimerization of Exd is dependent on the central tryptophan and methionine residues in this context.

In the distal orientation, 5 different spacer regions were investigated from -2 to 2 bps covering a theoretical distance of approximately 15 to 30 Å between the polyamide and expected protein-binding domain position. In this arrangement, Exd dimerization was observed only with compound **1** at the -2 spacer site (Figure 3.4 and Table 3.2). Control experiments at the -2 site show that Exd by itself, in the presence of parent polyamide ImImPyPy-γ-PyPyPyPy-β-Dp **15**, or double alanine mutant **3**, does not lead to stable complex formation (Figure F.2).

Minimization of the Protein-Binding Domain

As mentioned above, we have previously reported that in the context of a branched

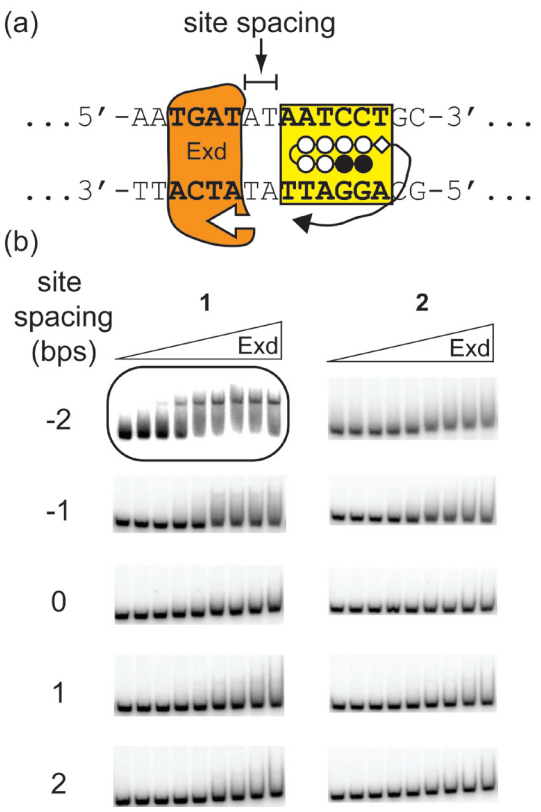


Figure 3.4 Representative gel shift results for Exd in the presence of linear protein-DNA dimerizers with differently sized linkers on two *distally* oriented DNA sites. (a) Schematic of the positions of the Exd and polyamide DNA-binding sites showing A,T spacer region which was varied. (b) Gel shift experiments show that **1** forms a stable complex with Exd and DNA when there is a 2 bp overlap, i.e., a site spacing of -2, between the two sites, whereas **2** does not form a complex with Exd at any of these sites. A concentration of 50 nM of the indicated compound was added to all lanes and Exd was titrated from 10 pM to 100 nM.

polyamide-peptide conjugate, the protein-binding domain could be reduced to a WM dipeptide.[5d] Thus, compounds **5-8** were synthesized to see if an analogous linear dipeptide conjugate would also dimerize Exd. Gel shift experiments with **5** and **6** using DNA duplexes with spacings of 0 to 6 in the proximal orientation, and -2 in the distal orientation did not lead to any significant gel shift (Figure F.3). A weak complex was observed with compound **6** at the 3 bp proximal spacing (Θ_{app} = 0.3 ± 0.1 at 100 nM Exd and 50 nM **6**), but was not significant enough to reliably determine the Exd-DNA binding affinity. Similarly, the WM conjugates bearing the PEG linker (**7** and **8**) failed to yield any discernible gel shift with DNA duplexes containing proximal spacings of 1 to 6 bps or at the -2 bp distal spacing (Figure F.3).

It was noted that the substitution of FYPWMK (**1** and **2**) with WM (**5-8**) eliminated the lysine between the presumed linker domain and the core WM dipeptide of the YPWM motif. Thus, WMK tripeptide conjugates **9-12** were synthesized to ensure identical linker length and composition

Table 3.2 Summary of select distal orientation gel shifts[a]

Base pairs between binding sites	Estimated distal distance (Å)[b]	Exd binding with **1**	Exd binding with **2**
-2	15.1	+	-
-1	19.5	-	-
0	23.8	-	-
1	28.1	-	-
2	31.5	-	-

[a] A "+" and "-" indicate the presence or absence of a complex, respectively. The binding constant (K_a) for the "+" entry is 2.5 ± 1.8 M^{-1}.
[b] measured from the Met C to the linker N

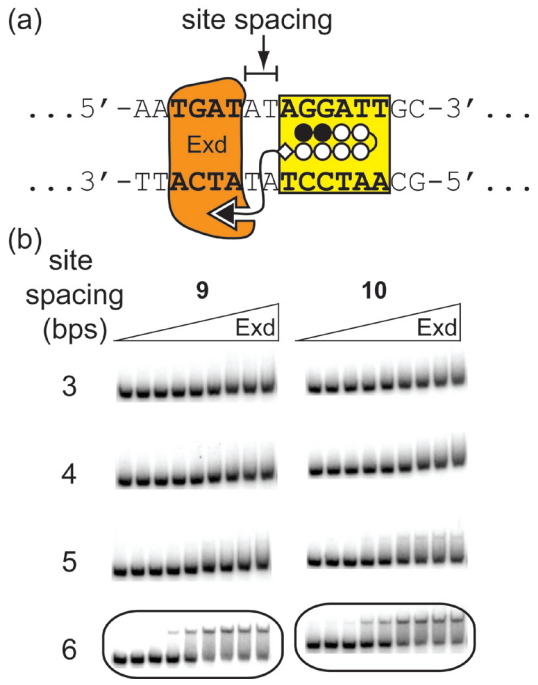


Figure 3.5 Representative gel shift results for Exd in the presence of linear WMK tripeptide conjugates on select *proximally* oriented DNA sites. (a) Schematic of the positions of the Exd and polyamide DNA-binding sites showing the A,T spacer region which was varied. (b) Gel shift experiments show that both **9** and **10** form stable complexes with Exd and DNA only when there is a 6 bp site spacing. A concentration of 50 nM of the indicated compound was added to all lanes and Exd was titrated from 10 pM to 100 nM.

between the polyamide and the WM moiety. These compounds were examined by gel shift using duplexes with proximal spacings between 3 and 6 bps shown above to enable dimerization of Exd with the larger conjugates **1** and **2**. Both tertiary amine containing WMK conjugates (**9** and **10**) yielded a complex with Exd only at the 6 bp proximal site, but the PEG containing WMK conjugates (**11** and **12**) did not lead to an observable shift in any of the orientations (Figure 3.5 and Figure F.4b). Both **9** and **10** at the proximal 6 bp spacing site yielded similar binding affinities ($K_a = 9.2 \pm 0.6 \times 10^8\ M^{-1}$ and $7 \pm 2 \times 10^8\ M^{-1}$, respectively) and amounts of complex formed (0.50 ± 0.01 and 0.53 ± 0.05, respectively) indicating no significant difference in Exd binding by either the L- or D-tryptophan peptides in this context. The binding affinities of Exd in the presence of **9** and **10** are within 4-fold of those observed for the larger conjugates **1** and **2** (Tables 3.1 and 3.2). In comparing **9** and **10** to **1**, however, it appears that the use of smaller and presumably lower-affinity protein-binding domains leads to a narrower site spacing tolerance. In general, larger higher-affinity protein-binding domains may tolerate suboptimal linker distances and may generate more promiscuous protein-DNA dimerizers.

Finally, to determine the principal role of the lysine in the WMK tripeptide conjugates **9** and **10**, their lysine side-chains were acetylated to create conjugates **13** and **14**, respectively. These compounds retain the same linker length as **9** and **10** between the polyamide and the core WM dipeptide, but should eliminate the positive charge on the lysine while maintaining comparable steric bulk. Both **13** and **14** were not able to form a complex in the presence of Exd (Figure F.4c). This suggests that the lysine side-chain makes a favorable electrostatic interaction to stabilize the Exd-DNA-dimerizer complex. Although a direct interaction with the Exd protein cannot be ruled out, the presumed requirement for the linker domain to traverse the DNA backbone suggests that the primary role of this lysine group may be to interact favorably with the negatively charged phosphate groups of DNA. In comparing to natural systems it was observed that Ubx contains an alanine at this position,[7a] but lysine and arginine are often found immediately adjacent to

the C-terminal end of the YPWM motif in many HOX proteins.[9]

Calculation of Distances

The superposition of a polyamide crystal structure[10] (PDB code 365D) over specific DNA sites in the Ubx/Exd/DNA crystal structure[7a] (PDB code 1B8I) readily allows the measurement of theoretical distances between the polyamide tail and the YPWM peptide that the linker domain must traverse (Figure 3.6 and Tables 3.1 and 3.2). In the proximal orientation the shortest distance occurs somewhere between the 2 and 3 bp spacings (~11 Å) and increases in both directions in a non-linear fashion due to the helical nature of the DNA grooves. In the distal orientation, the polyamide positions the linker past the

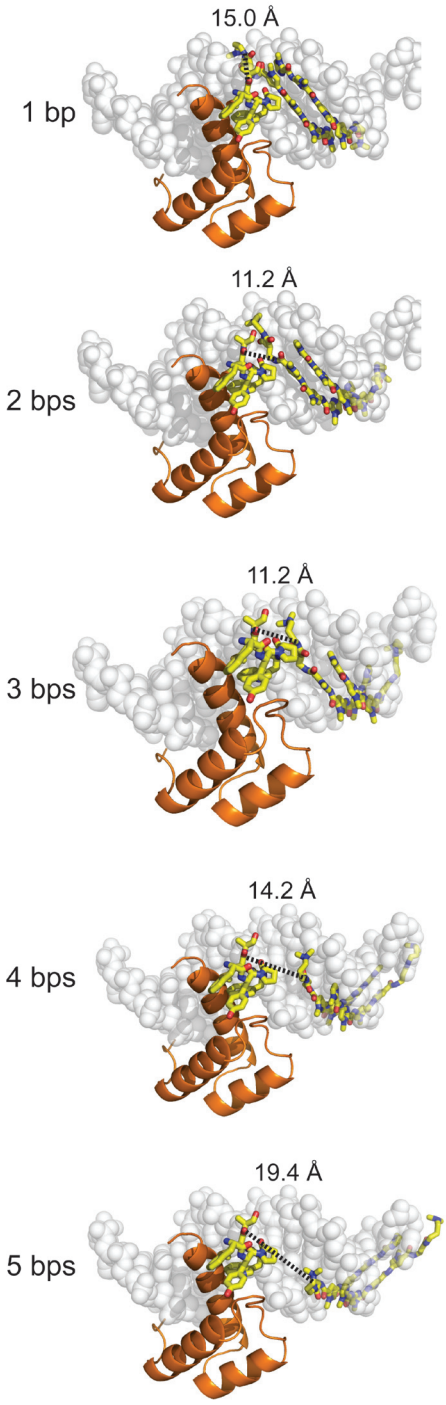


Figure 3.6 A series of estimated distance measurements are shown between the polyamide tail and the methionine carbonyl of the YPWM protein-binding domain which each dimerizer linker domain must traverse. The models of the polyamides binding to the DNA were generated by superimposing the DNA-binding site from a representative polyamide-DNA crystal structure (PDB code = 365D) over the Ubx/Exd/DNA crystal structure (PDB code 1B8I) and deleting the Ubx protein except for its YPWM protein-binding domain. In the proximal orientation, there exists a minimum distance of ~11 Å between the site spacings of 2-3 bps which increases non-linearly in each direction as the polyamide slides along the helical DNA minor groove.

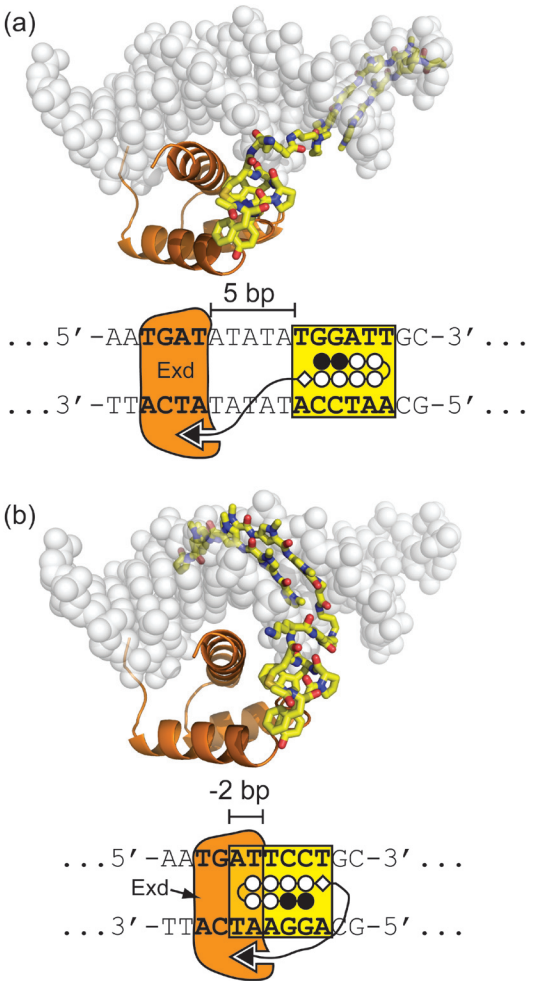


Figure 3.7 Models of linear-protein DNA dimerizers. (a) A proximally oriented binding site with a 5 bp spacer with dimerizer **1** is shown. (b) A distally oriented binding orientation with a 2 bp overlap (i.e., -2 bp spacer) between the polyamide and Exd DNA-binding sites is shown for dimerizer **1**.

minimum distance of approximately 11 Å observed in the proximal orientation. Thus, even at a 2 bp overlap in the distal orientation the distance is approximately 15 Å, which is longer than the shortest distance in the proximal orientation.

The results cannot be completely rationalized using distance measurements, however. For instance, comparing proximal sites with spacers of similar distances, such as 1 and 4 bps (14-15 Å), or 0 and 5 bps (19-20 Å) reveal that sites with similar distance constraints do not necessarily lead to similar abilities to enhance Exd binding to DNA. This apparent discrepancy could be explained by the fact that the measured distances do not account for the interference by the phosphate backbone which the linker domain must cross to place the protein-binding domain in the necessary location. Additionally, the linker domain must bend back approximately 180° in the 1 and 0 bp sites to bring the protein-binding domain into the correct binding orientation whereas the 4 and 5 bp sites project the protein-binding domain directly toward its Exd binding pocket. Such bending of the linker domain may be energetically unfavorable.

Similarly, in spite of the overlapping range of possible distances for the distal

binding orientations (~11-30 Å) when compared to the proximal sites (~15-30 Å), only the -2 bp distal site led to a stable complex with Exd. Unfortunately, greater overlap sites in the distal orientation cannot be addressed without changing the structure of the polyamide which was held constant for the purposes of this study. Movement of the polyamide binding site to the -3 bp site may still lead to a functional binding orientation, but the overlapped -4 bp site will probably not be tolerated since Arg2 of the Exd homeodomain contacts the thymine base at this location in the minor groove.[7a]

Interestingly, a complex was observed at the 6 bp (strong complex) and 7 bp (weak complex) proximal arrangements (≥25 Å spacing) with compound **1** even though the fully extended linker (~23 Å) was predicted not to reach in either case. This suggests there may have been some structural adaptation by the complex to enable the YPWM-Exd interaction. It is noted that TATA DNA sequences are known to be bent by TATA-binding protein[11] and it is thought that such sequences are inherently flexible.[12] Thus, the use of a variable poly-AT spacer may have enabled such DNA distortions. Either the Exd structure or the placement of the YPWM peptide in the Exd binding pocket may also allow structural deviations.

3.3 Conclusions

A set of linear polyamide-peptide conjugates have been created which expand the recognition scope of protein-DNA dimerizers targeted towards Exd. In particular, conjugate **1** was shown to enhance Exd's binding to DNA when the protein and polyamide binding sites were -2, 2, 3, 4, 5, and 6 bps apart. We have previously reported branched dimerizers that facilitate Exd binding to -1,[5c] 0,[5] and 1 bp[5c] orientations. Thus, a set of solutions from -2 to 6 bps has been completed using these linear conjugates. Furthermore, the protein-binding domain was minimized to a WMK tripeptide (**9** and **10**) that functioned nearly as well as the larger conjugates (**1** and **2**). Remarkably, the range of tolerated orientations appears to be tunable not only by the length and chemical nature of the linker (compare **1** to **2**), but also by the content of the protein-binding domain (compare **1** to **9**).

In combination with the previous results for branched dimerizers,[5] the work presented here provides design principles for dimerizers suitable for use in cell culture and fruit fly experiments. Developmental processes in fruit flies have been perturbed using polyamides.[13] The question arises whether a linear or branched protein-DNA dimerizer can interfere specifically with the normal function of Exd and lead to distinct fruit fly phenotypes. All compounds in this study function at temperatures suitable for experiments in fruit flies (20 °C) and the unnatural D-Trp WMK conjugate (**10**) is expected to be resistant to endogenous proteases.[14]

3.4 Experimental Section

General. Fmoc protected amino acids and 2-chlorotritylchloride resin was purchased from Novabiochem. Super Acid Sensitive Resin (SASRIN) was purchased from Bachem. DNA oligonucleotides were obtained from Integrated DNA Technologies. [γ-^{32}P]-adenosine-5'-triphosphate (≥ 6000Ci/mmol) was obtained from Perkin Elmer. Polynucleotide kinase (PNK) was obtained from Roche. UV spectra were recorded in water using an Agilent 8453 UV-Visible spectrophotometer. Analytical high-pressure liquid chromatography (HPLC) was performed with a Beckman Gold system equipped with a diode array 168 Detector and a 125 Solvent module using a Phenomenex Gemini C_{18} column (5 μm particle size, 250 × 4.60 mm). Preparative HPLC was performed with a Beckman Gold system equipped with a 127P Solvent Module and a single wavelength 166P Detector monitoring 310 nm using a Waters DeltaPak C_{18} reverse-phase column (100 μm particle size, 25 × 100mm) or a Phenomenex Gemini C_{18} reverse-phase column (5 μm particles, 110 Å pore, 250 × 21.2 mm). For both analytical and preparative HPLC, solvent A was 0.1% (v/v) aqueous TFA and solvent B was acetonitrile. Final preparative HPLC purification was performed by ramping from 10 to 55% B over 75 min using the Gemini column. If the purity of the compound was less than 95%, a second HPLC purification was performed employing a slow gradient through the previous % B elution of the compound. All other

general procedures are the same as described in Chapter 2.

Synthesis. Polyamide monomers were prepared as described previously.[16] Polyamides were synthesized according to standard synthetic procedures,[16] and cleaved from the resin with either 3,3'-diamino-*N*-methyl-dipropylamine, 4,7,10-trioxa-1,13-tridecanediamine, or 3-(dimethylamino)-propylamine. Before conjugation to peptides, polyamides were HPLC purified and characterized by MALDI-TOF MS, UV-Vis spectroscopy, and analytical HPLC. Protected peptides were synthesized using SASRIN or 2-chlorotrityl resin using standard Fmoc protocols[17] and characterized by ESI MS and analytical HPLC. Polyamide-peptide conjugates were prepared as described previously using 0.1 M DIEA in DMF using excess PyBOP (~5 equivalents) followed by deprotection with 80% TFA, 5% TIS, 5% EDT, and 10% DCM.[5a] Conjugates were then purified by preparative HPLC and characterized by MALDI-TOF or ESI MS, UV-Vis spectroscopy, and analytical HPLC. **1**: (MALDI-TOF) [M+H] calcd for $C_{113}H_{146}N_{31}O_{19}S$ 2274.1, observed 2274.0; **2**: (MALDI-TOF) [M+H] calcd for $C_{116}H_{151}N_{30}O_{22}S$ 2349.1, observed 2349.0; **3**: (MALDI-TOF) [M+H] calcd for $C_{103}H_{137}N_{30}O_{19}$ 2099.1, observed 2098.8; **4**: (MALDI-TOF) [M+H] calcd for $C_{106}H_{142}N_{29}O_{22}$ 2174.1, observed 2174.0; **5**: (ESI) [M+H] calcd for $C_{84}H_{109}N_{26}O_{14}S$ 1737.8, observed 1737.6; **6**: (MALDI-TOF) [M+H] calcd for $C_{84}H_{109}N_{26}O_{14}S$ 1737.8, observed 1737.4; **7**: (ESI) [M+H] calcd for $C_{87}H_{114}N_{25}O_{17}S$ 1812.9, observed 1812.5; **8**: (MALDI-TOF) [M+Na] calcd for $C_{87}H_{113}N_{25}O_{17}SNa$ 1834.8, observed 1834.6; **9**: (MALDI-TOF) [M+H] calcd for $C_{90}H_{121}N_{28}O_{15}S$ 1865.9, observed 1865.5; **10**: (MALDI-TOF) [M+H] calcd for $C_{90}H_{121}N_{28}O_{15}S$ 1865.9, observed 1865.8; **11**: (MALDI-TOF) [M+H] calcd for $C_{93}H_{126}N_{27}O_{18}S$ 1941.0, observed 1941.0; **12**: (MALDI-TOF) [M+H] calcd for $C_{93}H_{126}N_{27}O_{18}S$ 1941.0, observed 1940.9. Acetylated compounds **13** and **14** were synthesized by treating 200-300 nmol of **9** and **10**, respectively, with 8-9 equivalents of acetic anhydride in 0.1 M DIEA in DMF and purified by preparative HPLC. **13**: (MALDI-TOF) [M+H] calcd for $C_{92}H_{123}N_{28}O_{16}S$ 1907.9, observed 1908.2; **14**: (MALDI-TOF) [M+H]

calcd for $C_{92}H_{123}N_{28}O_{16}S$ 1907.9, observed 1908.0; **15**: (MALDI-TOF) [M+H] calcd for $C_{58}H_{72}N_{21}O_{10}$ 1222.6, observed 1222.7.

Electrophoretic mobility shift assays (EMSAs). EMSA, or gel shift, experiments were performed at 20 °C essentially as described in Chapter 2.[5d] The final concentrations of the components were 50 nM of the indicated compound, 30 mM potassium glutamate, 10 mM HEPES, 0.2 mM DTT, 10% glycerol, and 0.1 mg/mL acetylated BSA. The final concentrations for Exd were the following: 10 pM, 50 pM, 100 pM, 500 pM, 1 nM, 5 nM, 10 nM, 50 nM, and 100 nM. Recombinant Exd (residues 238-320) was expressed[5] and characterized[5d] as described in appendix C. The Exd concentration was determined by measuring the UV absorption at 280 nm using a calculated[18] extinction coefficient of 12,600 $M^{-1}\cdot cm^{-1}$. The amount of Exd bound to the DNA template was determined as described previously[5d] and fit to the Hill equation[19] with n = 1. All binding affinities were measured in triplicate.

Theoretical distance measurements. The illustrative models shown in Figure 3.6 were created by superimposing the DNA backbone of the atomic coordinates from a representative polyamide-DNA crystal structure[10] (PDB code 365D) over the desired site on the DNA backbone from the Ubx/Exd/DNA crystal structure[7a] (PDB code 1B8I) using a fit algorithm from SwissPDB Viewer (version 3.7). The DNA from the polyamide crystal structure and the Ubx homeodomain were then deleted to generate a model from which the distance measurements were be made.

The illustrative models shown in Figure 7 were made by first superimposing a polyamide-DNA crystal structure as described above. The linker between the polyamide and the FYPWMA peptide from Ubx, the linker, and the γ-aminobutyric acid of the polyamide hairpin, the lysine side-chain, and the acetyl moiety were constructed and minimized using an augmented MM3 force field and short molecular dynamics simulations (<100 ps) with all other coordinates locked using CAChe Workspace (ver. 6.1.12.33). The DNA shown in the figure is the same as the DNA from the Ubx/Exd/DNA structure.

Acknowledgements. This work was supported by the National Institute of Health (NIH GM51747 to P. B. D.). A fellowship from the Rose Hills Foundation (to R. L. S.) is gratefully acknowledged. We thank Drs. P. Snow and G. Hathaway (Beckman Institute, California Institute of Technology) for technical assistance with protein expression, purification, and identification.

3.5 References and Footnotes

(1) (a) Mitchell, P. J. and Tjian, R. *Science* **1989**, *245*, 371-378. (b) Lee, T. I. and Young, R. A. *Annu. Rev. Genet.* **2000**, *34*, 77-137. (c) Kadonaga, J. T. *Cell* **2004**, *116*, 247-257. (d) Levine, M. and Tjian, R. *Nature* **2003**, *424*, 147-151.

(2) (a) Klemm, J. D., Schreiber, S. L., Crabtree, G. R. *Annu. Rev. Immunol.* **1998**, *16*, 569-592. (b) Wolberger, C. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 29-56. (c) Ogata, K., Sato, K., Tahirov, T. *Curr. Opin. Struct. Biol.* **2003**, *13*, 40-48.

(3) Ptashne, M. and Gann, A. *Genes and Signals*; Cold Spring Harbor Laboratory Press: Plainview, NY, 2002.

(4) (a) Kuznetsova, S., Ait-Si-Ait, S., Nagibneva, I., Troalen, F., Le Villain, J., Harel-Bellan, A., Svinarchuk, F. *Nucleic Acids Res.* **1999**, *27*, 3995-4000. (b) Mapp, A. K., Ansari, A. Z., Ptashne, M., Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 3930-3935. (c) Ansari, A. Z., Mapp, A. K., Nguyen, D. H., Dervan, P. B., Ptashne, M. *Chem. Biol.* **2001**, *8*, 583-592. (d) Liu, B., Han, Y., Corey, D. R., Kodadek, T. *J. Am. Chem. Soc.* **2002**, *124*, 1838-1839. (e) Stanojevic, D. and Young, R. A. *Biochemistry*, **2002**, *41*, 7209-7216. (f) Arora, P. S., Ansari, A. Z., Best, T. P., Ptashne, M., Dervan, P. B. *J. Am. Chem. Soc.* **2002**, *124*, 13067-13071. (g) Liu, B., Han, Y., Ferdus, A., Corey, D. R., Kodadek, T. *Chem. Biol.* **2003**, *10*, 909-916. (h) Kwon, Y., Arndt, H.-D., Mao, Q., Choi, Y., Kawazoe, Y., Dervan, P. B., Uesugi, M. *J. Am. Chem. Soc.* **2004**, *126*, 15940-15941. (i) Xiao, X., Yu, P., Lim, H.-S., Sikder, D., Kodadek, T. *Angew. Chem. Int. Ed.* **2007**, *46*, 2865-2868. For reviews, see: (j) Ansari, A. Z. and Mapp, A. K. *Curr. Opin. Chem. Biol.* **2002**, *6*, 765-772. (k) Arndt, H.-D. *Angew. Chem., Int. Ed.* **2006**, *45*, 4552-4560.

(5) (a) Arndt, H.-D., Hauschild, K. E., Sullivan, D. P., Lake, K., Dervan, P. B., Ansari, A. Z. *J. Am. Chem. Soc.* **2003**, *125*, 13322-13323. (b) Hauschild, K. E., Metzler, R. E., Arndt, H.-D., Moretti, R., Raffaelle, R., Dervan, P. B., Ansari, A. Z. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 5008-5013. (c) Warren, C. L., Kratochvil, N. C. S., Hauschild, K. E., Foister, S., Brezinski, M. L., Dervan, P. B., Phillips, G. N. Jr., Ansari, A. Z. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 867-872. (d) Stafford, R. L., Arndt, H.-D., Brezinski, M. L., Ansari, A. Z., Dervan, P. B. *J. Am. Chem. Soc.* **2007**, *129*, 2660-2668.

(6) (a) Peifer, M. and Wieschaus, E. *Genes Dev.* **1990**, *4*, 1209-1223. (b) Chan, S. K., Jaffe, L., Capovilla, M., Botas, J., Mann, R. S. *Cell* **1994**, *78*, 603-615. (c) van Dijk, M. A. and Murre, C. *Cell* **1994**, *78*, 617-624. (d) Mann, R. S. and Chan, S. K. *Trends Genet.* **1996**, *12*, 258-262.

(7) (a) Passner, J. M., Ryoo, H. D., Shen, L. Y., Mann, R. S., Aggarwal, A. K. *Nature* **1999**, *397*, 714-719. (b) Piper, D. E., Batchelor, A. H., Chang, C. P., Clearly, M. L., Wolberger, C. *Cell* **1999**, *96*, 587-597. (c) LaRonde-LeBlanc, N. A. and Wolberger, C. *Genes Dev.* **2003**, *17*, 2060-2072.

(8) (a) Dervan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 2215-2235. (b) Dervan, P. B. and Edelson, B. S. *Curr. Opin. Struct. Biol.* **2003**, *13*, 284-299.

(9) Duboule, D. editor in *Guidebook to the Homeobox Genes*; Oxford University Press: Oxford, 1994.

(10) Kielkopf, C. L., Baird, E. E., Dervan, P. B., Rees, D. C. *Nat. Struct. Biol.* **1998**, *5*, 104-109.

(11) (a) Kim, Y., Geiger, J. H., Hahn, S., Sigler, P. B. *Nature* **1993**, *365*, 512-520. (b) Kim, J. L., Nikolov, D. B., Burley, S. K. *Nature* **1993**, *365*, 520-527. (c) Kim, J. L., Burley, S. K. *Nat. Struct. Biol.* **1994**, *1*, 638-653. (d) Juo, Z. S., Chiu, T. K., Leiberman, P M., Baikalov, I., Berk, A. J., Dickerson, R. E. *J. Mol. Biol.* **1996**, *261*, 239-254.

(12) Davis, N. A., Majee, S. S., Kahn, J. D. *J. Mol. Biol.* **1999**, *291*, 249-265.

(13) (a) Janssen, S., Durussel, T., Laemmli, U. K. *Mol. Cell* **2000**, *6*, 999-1011. (b)

Janssen, S., Cuvier, O., Muller, M., Laemmli, U. K. *Mol. Cell* **2000**, *6*, 1013-1024.

(14) Direct observation of the cell uptake of these compounds is not possible using confocal microscopy without the conjugation of a fluorescent dye (see reference 15), so we must infer cell uptake indirectly for compounds **1-15** from the biological data. Experiments with polyamide-peptide-fluorophore conjugates to assess cell uptake are underway.

(15) (a) Belistky, J. M., Leslie, S. J., Arora, P. S., Beerman, T. A., Dervan, P. B. *Bioorg. Med. Chem.* **2002**, *10*, 3313-3318. (b) Best, T. P., Edelson, B. S., Nickols, N. G., Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 12063-12068. (c) Edelson, B. S., Best, T. P., Olenyuk, B., Nickols, N. G., Doss, R. M., Foister, S., Heckel, A., Dervan, P. B. *Nucleic Acids Res.* **2004**, *32*, 2802-2818.

(16) Baird, E. E. and Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141-6146.

(17) Wellings, D. A. and Atherton, E. *Methods Enzymol.* **1997**, *289*, 44-67.

(18) Gill, S. C. and von Hippel, P. H. *Anal. Biochem.* **1989**, *182*, 319-326.

Chapter 4

Cell-Permeable Protein-DNA Dimerizers

*Work on the synthesis of polyamide-peptide-FAM conjugates **13** and **16** was done in collaboration with Rachel Wang (Dervan Group; Caltech).*

Abstract

In general, polyamide-peptide-fluorescein conjugates have shown poor cell uptake. This chapter describes how such conjugates can be made to localize to the cell nucleus through subtle structural changes. Specifically, cell-permeable conjugates are described that incorporate smaller peptides with the fluorescein attached to the polyamide C-terminus. Similar peptides conjugated to fluorescein without a polyamide, exhibited primarily extracellular localization which suggests that the polyamide confers the nuclear localization to the conjugate. Polyamide-peptide-fluorescein and polyamide-peptide-isophthalic acid conjugates also retain protein-DNA dimerization activity *in vitro*. These experiments provide a foundation for performing experiments in cell culture with protein-DNA dimerizers.

4.1 Introduction to Polyamide Conjugate Cell Uptake

Several conjugates containing pyrrole-imidazole polyamides and peptides have been reported to have a variety of functions *in vitro* such as gene activation[1-3] and protein-DNA dimerization.[4-7] Experiments in cell culture have experienced difficulty partly due to the poor cellular permeability of such conjugates. Polyamides by themselves often localize to the cell nucleus[8-9] which suggests that the addition of peptides hinders uptake. The utilization of non-peptide moieties does not render them cell permeable though. For example, a polyamide conjugated to the small-molecule wrenchnolol exhibits extracellular localization.[10] Wrenchnolol[11-12] and the polyamide[9] are independently cell-permeable so it is intriguing that the resulting conjugate is not. Thus, it seems that the overall conjugate structure must govern the propensity towards nuclear uptake.

Extensive studies into the structural determinants for nuclear localization of polyamide-fluorescein conjugates have shown that seemingly minor structural variation can lead to dramatic changes in cellular localization.[9] One notable example is that the polyamide ImImPyPy-γ-ImPyPyPy-FITC localizes to the nucleus, but the compound ImImPyPy-γ-ImPyPyPy-**β**-FITC, which contains an additional **β**-alanine, is extracellular.[9] To complicate matters, changing a single Im to a Py (ImImPyPy-γ-PyPyPyPy-**β**-FITC) recovers the nuclear localization.[9] Consequently, the complex interplay of structural variables that affect cellular localization of polyamides has yielded only a weak set of guidelines.

Nevertheless, it is apparent that tiny structural changes, i.e., as simple as an Im to Py substitution, can yield conjugates with desirable nuclear localization. Accordingly, the attempts described in this chapter to discover cell-permeable protein-DNA dimerizers involved simply moving to the attachment point of the fluorescein moiety to the polyamide and using smaller peptide domains. These efforts yielded a polyamide-peptide-fluorescein conjugate that localizes to the nucleus of living HeLa, MCF-7, and PC3 cells and retains the ability to dimerize Exd to DNA. The effect of the size of the peptide moiety was probed

by synthesizing larger peptides, but it is unclear if the upper limit has been reached.

4.2 Previous Polyamide-Peptide Conjugate Uptake Studies

In order to assess the cell uptake of series of polyamide-FYPWMKG conjugates, Hans-Dieter Arndt attached fluorescein-isothiocyanate (FITC) to the N-termini of the peptide moieties via a six-carbon linker to give compounds such as **1** (Figure 4.1).[13] Laser-scanning confocal microscopy in live cells,[8-9] showed that these compounds localized primarily outside of human cancer cell lines (HeLa, SKBr4, CEM-CCL, NB4) and

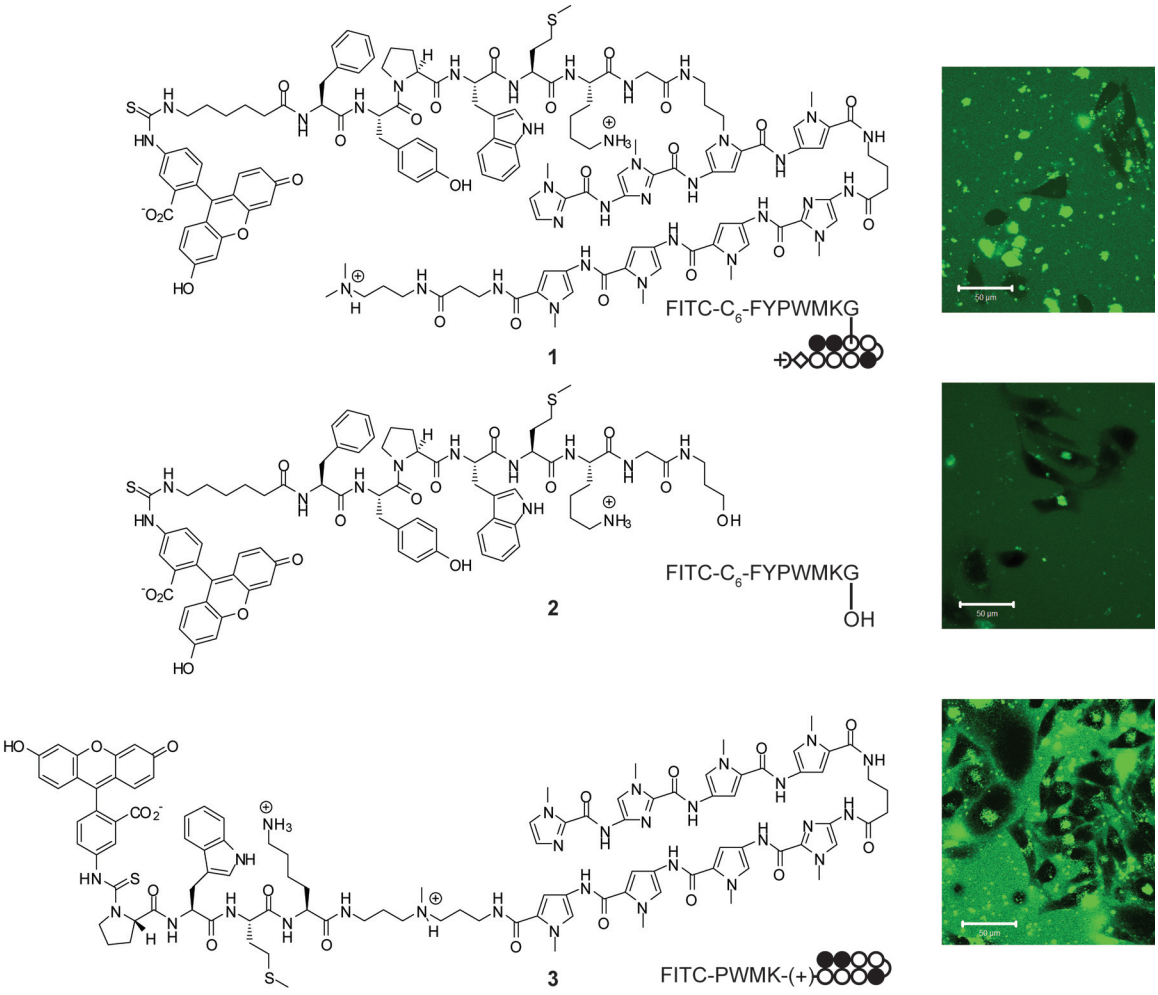


Figure 4.1 Summary of previous uptake results in HeLa cells. Polyamide-peptide-fluorescein conjugate **1** and the related peptide-fluorescein conjugate **2** show extracellular localization. Conjugate **3** is exhibits extracellular and vesicular staining. (Scale bar = 50 µm)

Drosophila cells (Kc167, S2). Similarly, the peptide-fluorescein conjugate **2** localized outside of cells. The use of the MDR pump inhibitor verapamil (100 μM), which has been shown to cause nuclear localization of some polyamides,[14] did not cause redistribution of **1** or **2**.[13] In other efforts to artificially induce cellular uptake, a pre-loaded plasmid (pDEH9) was incubated with polyamide **1** (500 nM) and a number of commercially available transfection reagents, including ExGen500 (Fermentas), Metafectene (Biontex), Escort (Sigma), and Polyfect (Qiagen), before loading the compounds onto live cells. Using the manufacturer's recommended upper and lower concentrations for the transfection reagents was not effective.[13]

Tim Best also investigated several polyamide-peptide-fluorescein conjugates, including **3**, in which similar peptides were attached to the polyamide C-terminus (Figure 4.1).[15] Unfortunately, compound **3** which contains a PWMK tetrapeptide between the polyamide and FITC does not show nuclear localization in 11 cell lines (MCF, HeLa, PC3, LN-CaP, DLD-1, 786-O, Jurkat, CEM, MEG-01, MEL, and NB4).[15] Conjugates of similar architecture bearing peptides of different sequences (DWMK and RY, with several unnatural forms of tyrosine) also yielded extracellular localization except for a couple of instances in CEM cells with the RY-containing compounds.[15]

Some control experiments with compounds **4**-**6** performed by Hans-Dieter Arndt,[13] suggested that attachment of the FITC directly to the polyamide C-terminus might be more favorable for cell uptake (Figure 4.2). Most notably, conjugate **6** which contains a tryptamine conjugated to an internal pyrrole yielded a modest degree of nuclear localization.[10] The structural difference is small between the appendage off the internal pyrrole of **6** and the WM conjugates described in Chapter 2. Thus, a related polyamide-WM conjugate was investigated, replacing the tryptamine and eight-carbon linker of **6** with a WM dipeptide and a six-carbon linker.

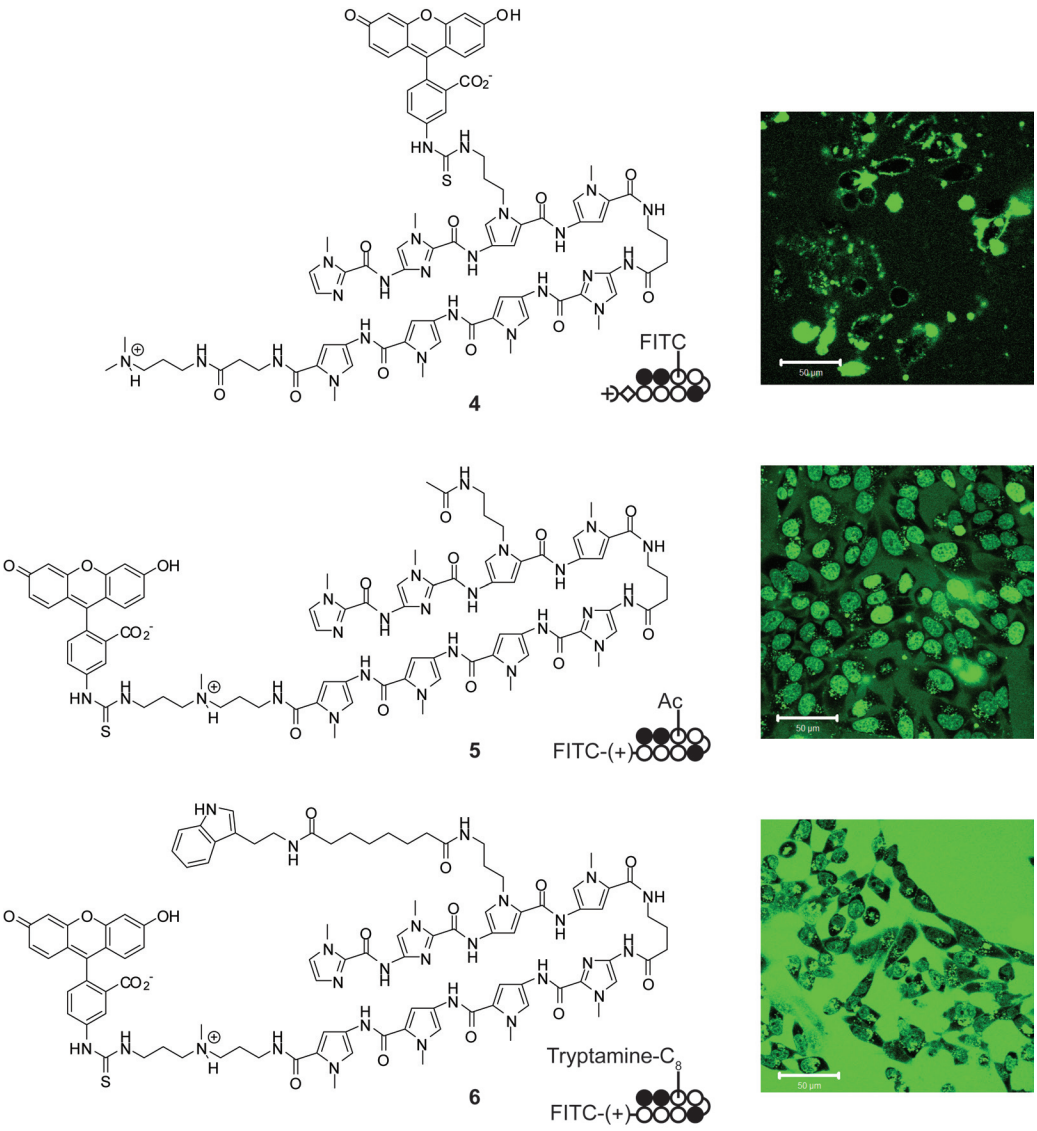


Figure 4.2 Summary of more previous uptake results in HeLa cells. Branched polyamide-FITC conjugate **4** shows extracellular and membranous staining. Polyamide-FITC conjugates **5** and **6** both show partial nuclear staining. (Scale bar = 50 μm)

4.3 Discovery of Cell-Permeable Protein-DNA Dimerizers

To synthesize the desired conjugates, resin-bound ImImPyPy-γ-ImPyPyPy-polyamide **7** was cleaved by treating with the mono-BOC protected diamine **8** (Figure 4.3). The resulting C-terminally BOC-protected polyamide **9** was coupled to the desired peptide, deprotected with TFA, and conjugated to fluorescein via an amide bond (FAM). This strategy was used to synthesize of a series of polyamide-peptide-FAM conjugates **10**-**16** for cell uptake studies (Figures 4.4 to 4.5). The natural WM conjugate **10** yielded a strong

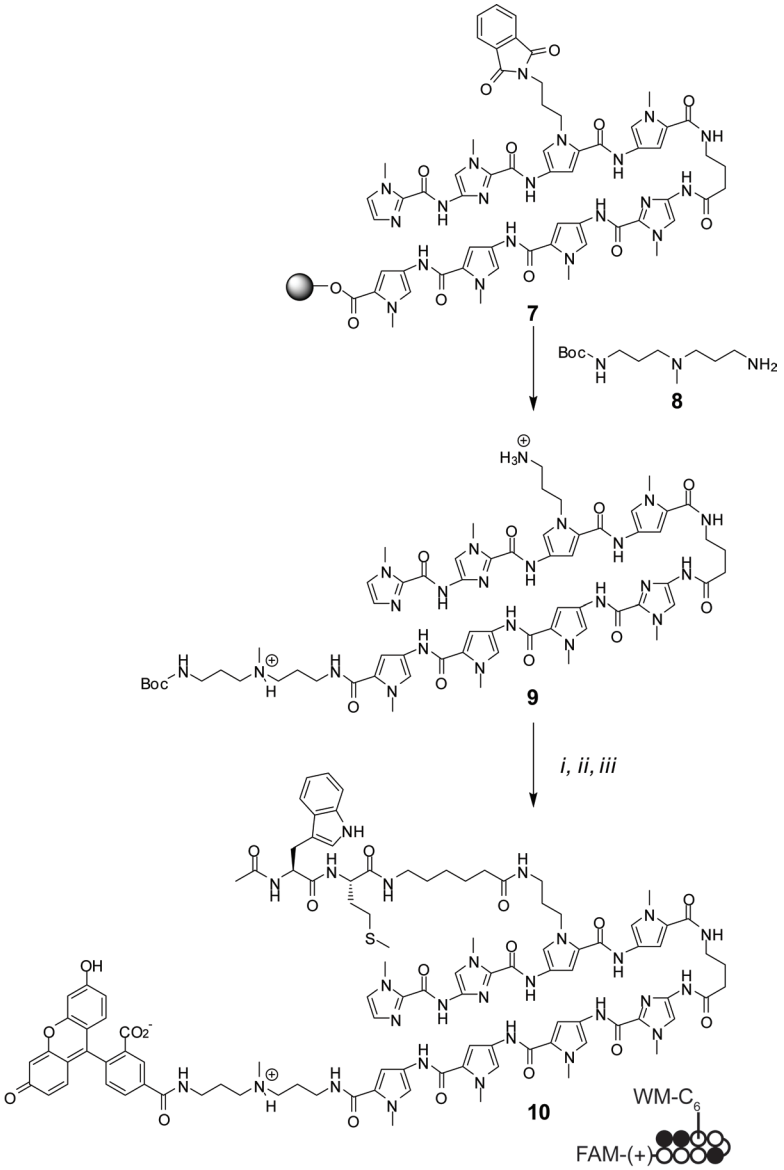


Figure 4.3 Synthesis of polyamide-peptide-fluorescein conjugates: *i*) **17**, HBTU, 0.1 M DIEA in DMF, *ii*) 50% TFA/DCM, and *iii*) 5-carboxyfluorescein succinimidyl ester.

nuclear fluorescence stain in HeLa, MCF-7 and PC3 cells (Figure 4.4). The AM conjugate **11** also showed nuclear fluorescence albeit with higher levels in the media, suggesting the tryptophan may positively impact nuclear localization. Unnatural peptide conjugates **12** (D-Trp-L-Met) and **13** (D-Trp-D-Met) yielded similarly strong nuclear localization to **10** (L-Trp-L-Met) suggesting that the stereochemistry of the tryptophan and methionine can be altered. This may be useful for future biological studies given that previous work has

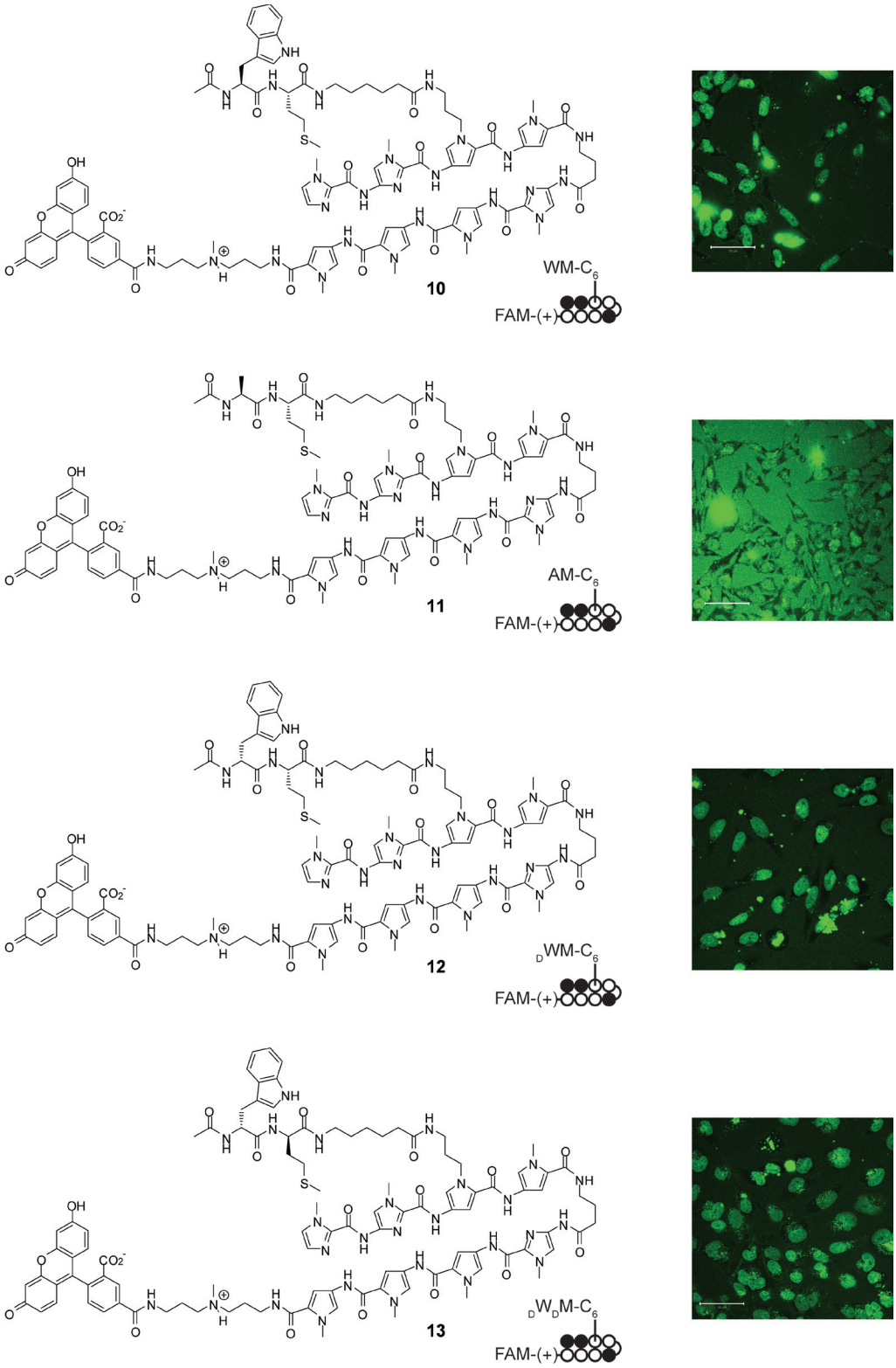


Figure 4.4 Uptake results in HeLa cells for polyamide-dipeptide-fluorescein conjugates. All compounds **10**-**13** show primarily nuclear localization, but the Trp to Ala conversion (**10** to **11**) decreases the apparent nuclear fluorescent intensity relative to the media. (Scale bar = 50 μm)

shown D-Trp-L-Met conjugates can dimerize Exd at higher temperatures (up to 37 °C) than natural dipeptides[7] and unnatural peptides should be resistant to proteolysis by endogenous proteases.

Analogous polyamide-YPWM tetrapeptide conjugates (**14-16**) were investigated to probe the upper size limit for the peptide (Figure 4.5). The YPWM conjugate **14** yielded strong nuclear fluorescence with some vesicular and membranous staining also apparent. Similar to the results with the AM conjugate **11**, the YPAM conjugate **15** showed higher levels of fluorescence in the media than YPWM conjugate **14**. Complete conversion of the amino acids of the YPWM peptide to the unnatural D-form (**16**) had a markedly negative impact on nuclear localization, with only modest levels of nuclear fluorescence visible. The apparent difference in uptake of **14** and **16** can be explained at least two different ways. First, it may be possible that the peptide of **14**, but not **16**, is degraded by proteases to generate smaller metabolites which render it cell permeable. Alternatively, the peptide may directly interact with some stereoselective protein involved in active uptake.[16]

To determine the uptake of peptide-fluorescein conjugates without polyamides **18** and **20** were investigated (Figure 4.6). The conjugates were synthesized by coupling **17** or **19** to mono-BOC protected diamine **8**, deprotected with TFA, and conjugated to FAM. Both peptide-FAM conjugates **18** and **20** exhibit primarily extracellular fluorescence (Figure 4.7). This suggests that the polyamide of **10** and **14** is responsible for trafficking the conjugate to the nucleus.

4.4 Dimerization of Exd with Cell-Permeable Conjugates

Electrophoretic mobility shift assays were then performed to determine the ability of cell-permeable conjugates to dimerize Exd. Polyamide-isophthalic acid (IPA) conjugates have been recently reported to possess similar cell permeability to their corresponding fluorescein conjugates.[17] Thus, a non-fluorescent IPA-polyamide-peptide conjugate **21** as well a control IPA conjugate without a peptide **22** were also investigated (Figure 4.8).

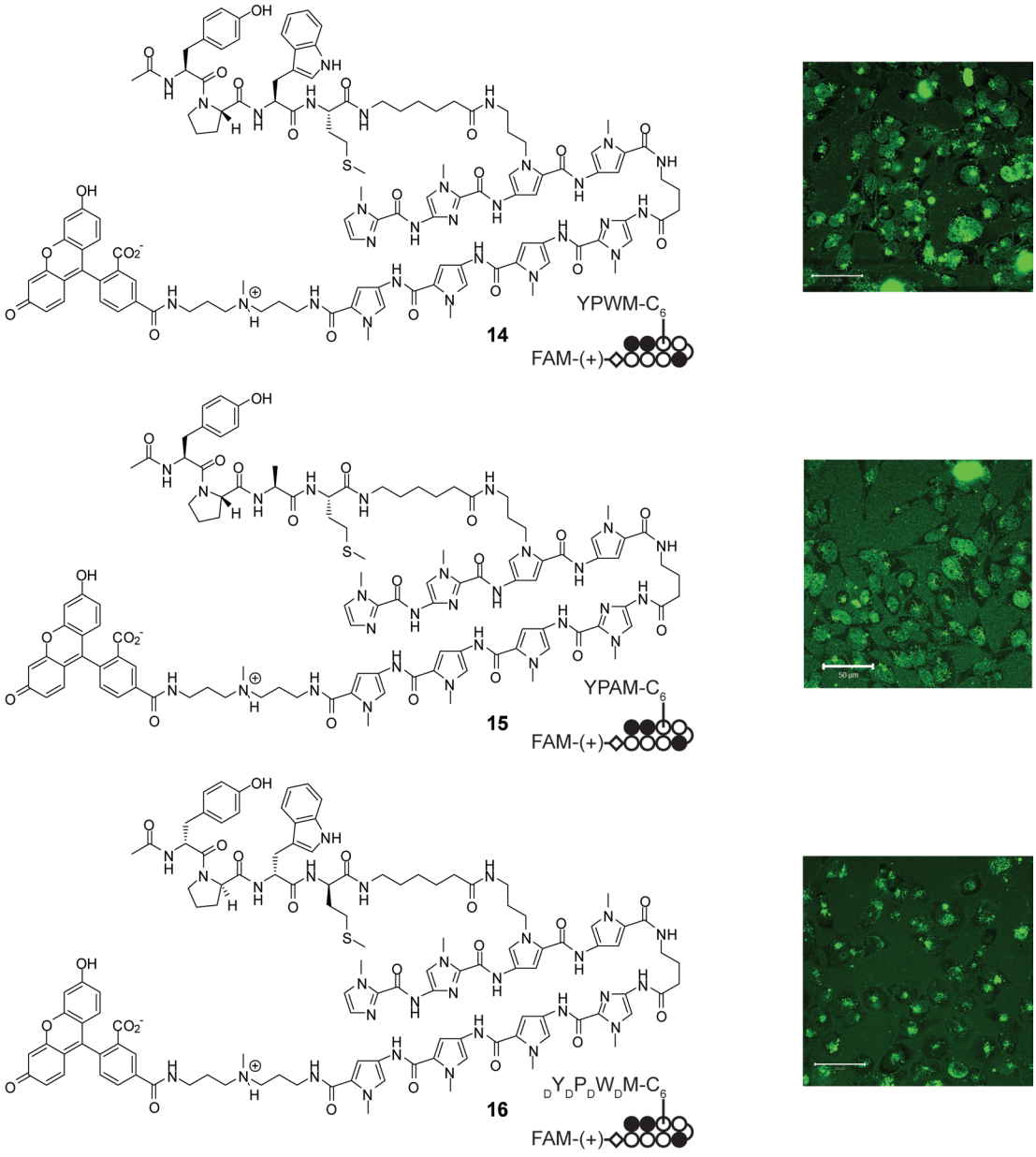


Figure 4.5 Uptake results in HeLa cells for polyamide-tetrapeptide-fluorescein conjugates. The natural YPWM conjugate **14** shows nuclear, membranous, vesicular and extracellular localization. The YPAM conjugate **15** shows weaker nuclear and extracellular staining. The unnatural YPWM conjugate **16** shows a strong vesicular and a weak nuclear staining pattern. (Scale bar = 50 μm)

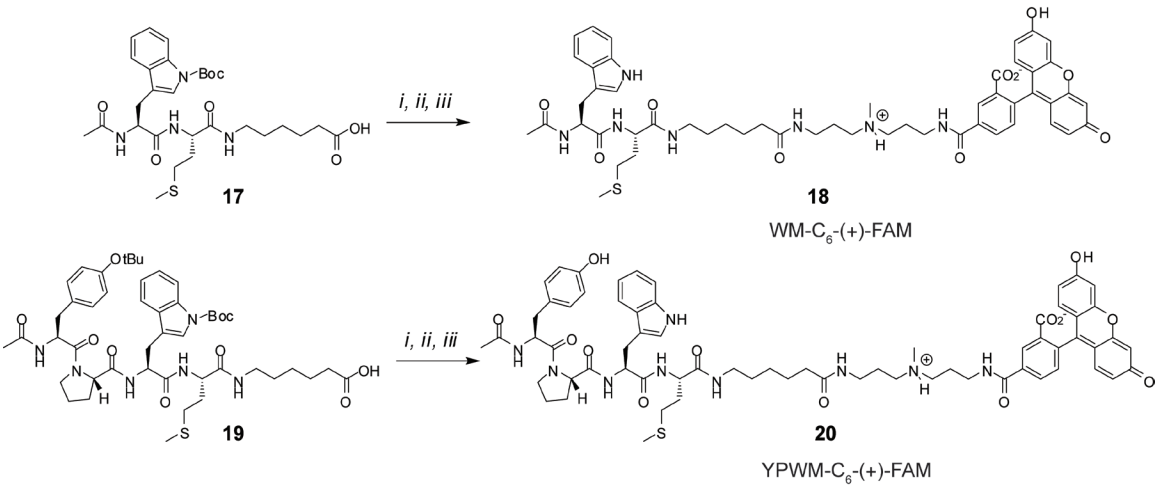


Figure 4.6 Synthesis of YPWM and WM peptide-fluorescein conjugates. *i*) **17** or **19**, HBTU, 0.1 M DIEA in DMF, *ii*) 50% TFA/DCM, and *iii*) 5-carboxyfluorescein succinimidyl ester.

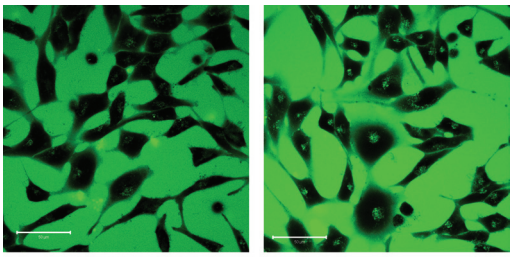


Figure 4.7 Uptake results in HeLa cells for peptide-fluorescein conjugates. Both the WM (**18**) and YPWM (**20**) conjugates show primarily extracellular staining. (Scale bar = 50 μm)

Polyamide-WM-FAM conjugate **10** yielded a significantly higher gel shift than AM conjugate **11** or Exd alone at 4 °C (Figure 4.9). Although Exd's binding affinity in the presence of **10** (K_a = 1 ± 2 × 10^9 M^{-1}) was comparable to that of the similar conjugate lacking the FAM (K_a = 7 ± 2 × 10^8 M^{-1}, Chapter 2)[7], the fraction of DNA shifted by **10** was significantly lower (Θ_{app} = 0.54 ± 0.02 compared to 0.84 ± 0.06, Chapter 2)[7] even at this lower temperature. At a higher temperature (20 °C) polyamide-WM-FAM and -IPA conjugates with the unnatural D-Trp (**12** and **21**) yielded lower binding affinities (K_a's = 4 ± 2 and 2 ± 1 × 10^8 M^{-1}) but with higher fractions of complex formation (0.74 ± 0.08 and 0.74 ± 0.04).

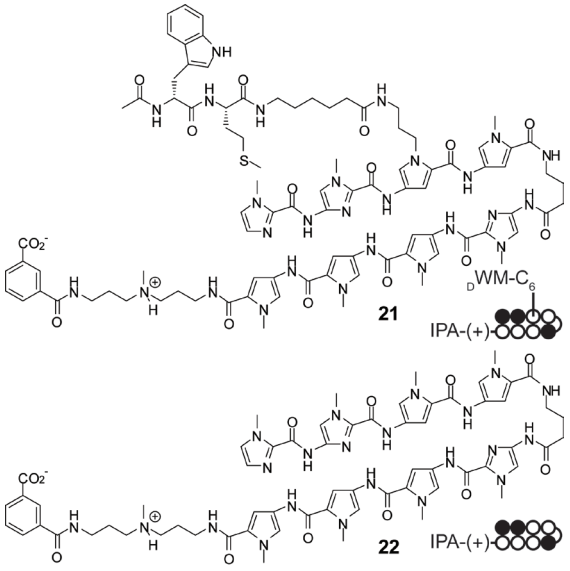


Figure 4.8 Structures of polyamide-peptide-isophthalic acid (IPA) conjugates.

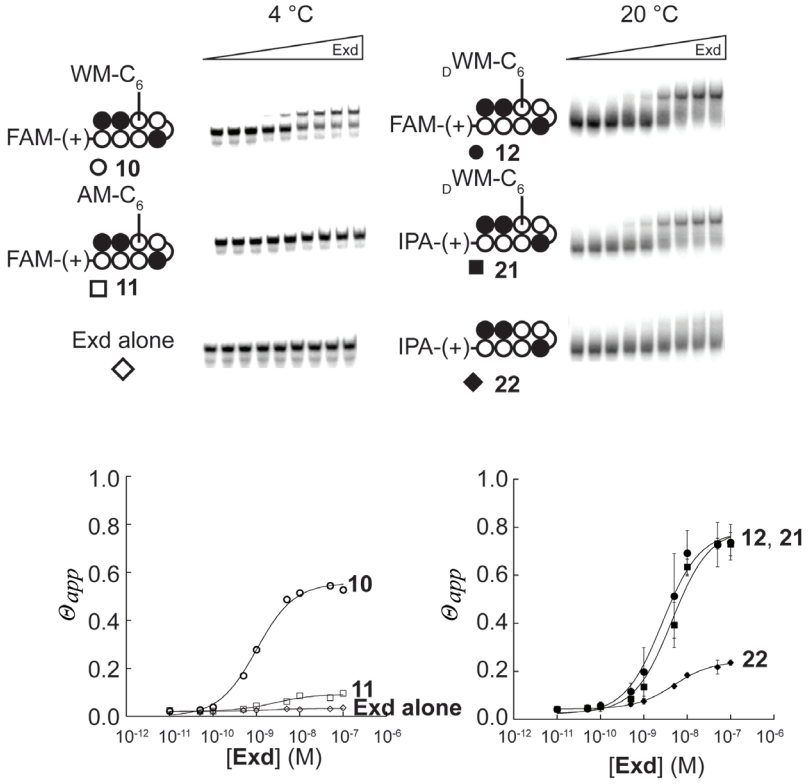


Figure 4.9 Gel shift experiments with conjugates. Polyamide-WM-FAM conjugate **10** shows a significant complex formation with Exd at 4 °C when compared to the AM control compound **11** or Exd alone. Polyamide-WM-FAM (**12**) and -IPA (**21**) conjugates which contain an unnatural D-Trp show significant complex formation with Exd at 20 °C compared to the no peptide IPA control (**22**). Conjugates were added at a concentration of 50 nM and Exd was titrated from 10 pM to 100 nM.

4.5 Conclusions

Intriguingly, the stereochemistry of the dipeptide conjugates did not appear to affect their nuclear localization, but the unnatural tetrapeptide conjugate exhibited substantially reduced nuclear staining compared to the natural enantiomer. It is unclear what causes the tetrapeptide conjugates to yield such apparent differences in cell uptake, but it is possible that the natural peptides are unstable in cell culture. Nonetheless, the presumably stable unnatural dipeptide conjugates localize to the nucleus and dimerize Exd. Thus, cell-permeable polyamide-peptide conjugates have been created that retain the ability to facilitate Exd-DNA dimerization *in vitro*. These compounds provide the foundation for cell culture and fruit fly experiments in an ongoing collaboration with the Ansari Lab (University of Madison, Wisconsin).

4.6 Experimental Details

General Synthetic Procedures. Polyamide monomers were synthesized by previously reported procedures.[18,19] Polyamides were synthesized on either PAM,[18] oxime,[20] or Marshall-Liener resin (Appendix D). All peptides were synthesized using Fmoc-based protocols on SASRIN (Bachem) or Wang resin (Novabiochem).[21] All polyamides were purified by preparative HPLC with either a Waters C18 Delta-Pak (15 μm, 25 × 100 mm, 300 Å), Varian Dynamax Microsorb C8 (250 × 21.4 mm, 300 Å), or Phenomenex C18 Gemini (5 μm, 250 × 21.2 mm, 110 Å) ramping from 10% A buffer (0.1% TFA in water) with acetonitrile as the B buffer. In general, the best purification was obtained with the Gemini column. Analytical High-Pressure Liquid Chromatography (HPLC) was performed with a Beckman Gold system equipped with a diode array 168 Detector and a 125 Solvent module using a Phenomenex Gemini C_{18} column (5 μm particle size, 250 × 4.60 mm). All conjugates were ≥95% pure before cell uptake and gel shift experiments were performed.

***N*-Boc-3,3′-diamino-*N*′-methyldipropylamine (8)**. This compound was prepared as

described previously.[22]

ImImPy$^{(propylamine)}$Py-γ-ImPyPyPy-(+)-BOC (9). This polyamide was synthesized on oxime resin according to standard procedures.[20] 255.1 mg resin cleaved with **8** (~300 μL) in 1 mL DMF at 50 °C for 13.5 h. Total yield, 7.66 μmol of a white powder (9.2% yield). (MALDI-TOF MS) $[M + H]^+$ calcd for $C_{63}H_{84}N_{23}O_{11}^+$ 1338.7, observed 1338.7.

ImImPy$^{(Trp\text{-}Met\text{-}\varepsilon Ahx\text{-}propylamine)}$Py-γ-ImPyPyPy-(+)-FAM (10). The peptide Ac-Trp(Boc)-Met-εAhx-OH (2.0 mg, 3.4 μmol) was dissolved in 200 μL 0.1 M DIEA in DMF and activated at 20 °C for ~5 min with 200 μL of a solution of PyBOP (total 5.4 mg, 10.5 μmol). The polyamide **9** (1.47 μmol) was dissolved in 200 μL of 0.1 M DIEA in DMF and added to the activated peptide. The resulting polyamide-peptide conjugate intermediate was precipitated with diethyl ether (10 mL) cooled on dry ice and centrifuged for 10 min at 20 °C. The supernatant was decanted and the resulting pellet was dried briefly (<30 sec) with $N_{2(g)}$. The Boc moiety was removed by treating with 50% TFA/DCM and incubating at 20 °C for 25 min. The deprotected conjugate was precipitated with diethyl ether, collected by centrifugation, and dried as before. The pellet was dissolved in 800 μL 1.0 M DIEA in DMF and 3 μmol of 5-carboxyfluorescein succinimidyl ester (Molecular Probes) was added as a 0.1 M solution in DMSO. The resulting compound was purified by preparative HPLC and lyophilized to yield 364 nmol a yellow-orange solid (~25% yield). (MALDI-TOF MS) $[M + H]^+$ calcd for $C_{103}H_{118}N_{27}O_{19}S^+$ 2068.9, observed 2068.9.

ImImPy$^{(Ala\text{-}Met\text{-}\varepsilon Ahx\text{-}propylamine)}$Py-γ-ImPyPyPy-(+)-FAM (11). The conjugate was synthesized in a similar manner to conjugate **10** using the peptide Ac-Ala-Met-εAhx-OH (2.1 mg, 5.6 μmol) activated with PyBOP (5.4 mg, 10.4 μmol) coupled to polyamide **9** (1.47 μmol). The resulting intermediate was then coupled to 5-carboxyfluorescein succinimidyl ester (3.0 μmol). The final product was purified by preparative HPLC and lyophilized to yield 671 nmol of a yellow-orange solid (~46% yield). (MALDI-TOF MS) $[M + H]^+$ calcd for $C_{95}H_{113}N_{26}O_{19}^+$ 1953.8, observed 1953.9.

ImImPy$^{(\omega\text{-}Trp\text{-}Met\text{-}\varepsilon Ahx\text{-}propylamine)}$Py-γ-ImPyPyPy-(+)-FAM (12). The conjugate was

synthesized in a similar manner to conjugate **10** using the peptide Ac-D-Trp-Met-εAhx-OH (~1.9 μmol) activated with PyBOP (6 μmol) coupled to polyamide **9** (1.0 μmol). The resulting intermediate was then coupled to 5-carboxyfluorescein succinimidyl ester (1.5 μmol). The final product was purified by preparative HPLC and lyophilized to yield 411 nmol of a yellow-orange solid (~41% yield). (MALDI-TOF MS) $[M + H]^+$ calcd for $C_{103}H_{118}N_{27}O_{19}S^+$ 2068.9, observed 2069.0.

ImImPy$^{(D\text{-}Trp\text{-}D\text{-}Met\text{-}\varepsilon Ahx\text{-}propylamine)}$Py-γ-ImPyPyPy-(+)-FAM (13). The conjugate was synthesized in a similar manner to conjugate **10** using the peptide Ac-D-Trp-D-Met-εAhx-OH (2.3 mg, 3.9 μmol) activated with PyBOP (5.4 mg) coupled to polyamide **9** (1.35 μmol). The resulting intermediate was then coupled to 5-carboxyfluorescein succinimidyl ester (1.5 μmol). The final product was purified by preparative HPLC and lyophilized to yield 224 nmol of a yellow-orange solid (~17% yield). (MALDI-TOF MS) $[M + H]^+$ calcd for $C_{103}H_{118}N_{27}O_{19}S^+$ 2068.9, observed 2068.8.

ImImPy$^{(Tyr\text{-}Pro\text{-}Trp\text{-}Met\text{-}\varepsilon Ahx\text{-}propylamine)}$Py-γ-ImPyPyPy-(+)-FAM (14). The conjugate was synthesized in a similar manner to conjugate **10** using the peptide Ac-Tyr(OtBu)--Pro-Trp(BOC)-Met-εAhx-OH (~3 μmol) activated with PyBOP (~6 μmol) coupled to polyamide **9** (1.7 μmol). The resulting intermediate was then coupled to 5-carboxyfluorescein succinimidyl ester (3 μmol). The final product was purified by preparative HPLC and lyophilized to yield 448 nmol of a yellow-orange solid (~26% yield). (MALDI-TOF MS) $[M + H]^+$ calcd for $C_{117}H_{134}N_{29}O_{22}S^+$ 2329.0, observed 2329.0.

ImImPy$^{(Tyr\text{-}Pro\text{-}Ala\text{-}Met\text{-}\varepsilon Ahx\text{-}propylamine)}$Py-γ-ImPyPyPy-(+)-FAM (15). The conjugate was synthesized in a similar manner to conjugate **10** using the peptide Ac-Tyr(OtBu)-Pro-Ala-Met-εAhx-OH (~3 μmol) activated with PyBOP (~6 μmol) coupled to polyamide **9** (1.7 μmol). The resulting intermediate was then coupled to 5-carboxyfluorescein succinimidyl ester (3 μmol). The final product was purified 2x by preparative HPLC and lyophilized to yield 68 nmol of a yellow-orange solid (~4% yield). (MALDI-TOF MS) $[M + H]^+$ calcd for $C_{109}H_{129}N_{28}O_{22}^+$ 2214.0, observed 2214.1.

ImImPy$^{(D\text{-}Tyr\text{-}D\text{-}Pro\text{-}D\text{-}Trp\text{-}D\text{-}Met\text{-}\varepsilon Ahx\text{-}propylamine)}$Py-γ-ImPyPyPy-(+)-FAM (16). The conjugate was synthesized in a similar manner to conjugate **10** using the peptide Ac-D-Tyr(OtBu)-D-Pro-D-Trp(Boc)-D-Met-εAhx-OH (6 mg, ~6.6 μmol) activated with PyBOP (5.4 mg) coupled to polyamide **9** (1.32 μmol). The resulting intermediate was then coupled to 5-carboxyfluorescein succinimidyl ester (4 μmol). The final product was purified by preparative HPLC and lyophilized to yield 224 nmol of a yellow-orange solid (~17% yield). (MALDI-TOF MS) $[M + H]^+$ calcd for $C_{117}H_{134}N_{29}O_{22}S^+$ 2329.0, observed 2328.8.

Trp(Boc)-Met-εAhx-OH (17). The peptide was synthesized by Fmoc methods[21] using SASRIN resin and cleaved with 2.5 mL of cleavage mixture (89:5:5:1 DCM:EDT:TES:TFA) 4x for 15 min each. Each 2.5 mL cleavage mixture was cooled on an ice bath and neutralized by the addition of 50 μL pyridine. The peptide was partitioned between 100 mL of 1M $KHSO_4$ and 100 mL EtOAc. The organic layer was washed 1x with 100 mL of brine, dried over anhydrous sulfate, filtered and evaporated under reduced pressure to give a yellow film. Added a minimal amount (<1 mL) of MeOH to dissolve the film which then added to ~45 mL of ice cold water to precipitate the peptide which was frozen and lyophilized to give a white powder (estimated 80% purity by analytical HPLC). (ESI-MS) $[M-H]^-$ calcd for $C_{29}H_{41}N_4O_7S^-$ 589.3, observed 589.1.

Trp-Met-εAhx-(+)-FAM (18). Peptide **17** (2.2 mg, 3.7 μmol) was activated with PyBOP (10.3 mg, 19.8 μmol) in 0.1 M DIEA in DMF (200 μL) for ~5 min at RT before adding to amine **8** (6 μmol). After ~1 hour, the solution was added to 3 mL of saturated NH_4Cl extracted 3x with 3 mL of DCM. The combined organic layer was dried with anhydrous sodium sulfate, filtered, and dried under reduced pressure to yield an oil. The Boc group was removed with 2 mL of 50% TFA/DCM for 25 min at RT and dried under reduced pressure. Benzene (~2 mL) was added to azeotrope the solvent under reduced pressure to give a crude film which was dissolved in 2 mL of 1 M DIEA in DMF. The fluorescent moiety 5-carboxyfluorescein succinimidyl ester (3.7 μmol as a 0.05 M solution in DMSO) was added and the resulting crude product was purified by preparative HPLC to give 1.42

μmol of a yellow-orange solid (~38% yield). (ESI-MS) $[M+H]^+$ calcd for $C_{52}H_{62}N_7O_{10}S^+$ 976.4, observed 976.2.

Tyr(OtBu)-Pro-Trp(Boc)-Met-εAhx-OH (19). The peptide was synthesized by Fmoc methods[21] using SASRIN resin and cleaved by treating the resin 4x with the cleavage mixture (1:1:98 TFA:EDT:DCM) draining into a flask cooled on ice. 50 μL of pyridine was added to neutralize the solution after each cycle. The resin was rinsed with 3-5 mL of DCM (2x) and MeOH (2x) twice and the solvent was removed under reduced pressure, dissolved in MeOH (~8 mL), and purified by preparative HPLC. Fractions with the peptide were pooled and lyophilized to give a white powder. (ESI-MS) $[M-H]^-$ calcd for $C_{47}H_{66}N_6O_{10}S^-$ 905.5, observed 905.4.

Tyr-Pro-Trp-Met-(+)-FAM (20). Peptide **19** (3.3 mg, 3.65 μmol) was activated with PyBOP, deprotected, and coupled to fluorescein as described for conjugate **18** to yield 1.4 μmol of a yellow-orange solid (~38% yield). (ESI-MS) $[M+H]^+$ calcd for $C_{66}H_{78}N_9O_{13}S^+$ 1236.5, observed 1236.4.

ImImPy$^{(D\text{-Trp-Met--}\varepsilon\text{Ahx-propylamine})}$Py-γ-ImPyPyPy-(+)-IPA (21). Based on previously published procedures,[17] the conjugate was synthesized in a similar manner to conjugate **10** using the peptide Ac-D-Trp-Met-εAhx-OH (~3.4 μmol) activated with PyBOP (6 μmol) coupled to polyamide **9** (1.0 μmol). The resulting intermediate was then coupled to excess isophthalic acid (IPA, 20 μmol) pre-activated (~1-2 min) with PyBOP (40 μmol) in 600 μL 1 M DIEA in DMF. The final product was purified by preparative HPLC and lyophilized to yield 504 nmol of a white solid (~50% yield). (MALDI-TOF MS) $[M + H]^+$ calcd for $C_{90}H_{112}N_{27}O_{16}S^+$ 1858.9, observed 1858.8.

ImImPyPy-γ-ImPyPyPy-(+)-IPA (22). Based on previously published procedures,[17] the parent polyamide ImImPyPy-γ-ImPyPyPy-(+)-NH_2 (~1.1 μmol) was coupled to excess isophthalic acid (IPA, 20 μmol) pre-activated (~1-2 min) with PyBOP (40 μmol) in 600 μL 1 M DIEA in DMF and purified by preparive HPLC. The final product was lyophilized to yield 414 nmol of a white solid (~37.6% yield). (MALDI-TOF MS) $[M + H]^+$ calcd for

$C_{64}H_{75}N_{22}O_{12}S^+$ 1343.6, observed 1343.4.

Cell culture and confocal microscopy experiments. All cell culture and confocal microscopy was performed essentially as described previously.[9] Briefly, MCF-7 and PC3 cells were cultured in 10% FBS, 1% penicillin/streptomycin supplemented RPMI 1640 at 37 °C in an incubator with 5% CO_2 according to ATCC recommended procedures. HeLa cells were cultured in similarly supplemented DMEM under the same conditions according to ATCC recommended procedures. Before confocal microscopy, cells were trypsinized for 5 min at 37 °C, centrifuged for 5 min at 900 g, and resuspended to a concentration of 3.33×10^5 cells/mL, and 150 μL was plated on glass bottom culture plates (MatTek) and allowed to adhere for 24 hours. The medium was removed and replaced with fresh medium supplemented with the desired conjugate (150 μL final media volume with 2 μM final conjugate concentration). Cells were incubated as described above for 10-14 hours before imaging on a Zeiss LSM 5 Pascal inverted laser scanning micrscope. The optical slice was set to 2.2 μm (pinhole size = 181) and images were line averaged 4 or 8 times using a 40× oil immersion lens (plan-neofluor) using standard fluorescein filters. Stock solutions of conjugates usually contained small amounts of precipitate that was apparent upon centrifugation. Higher percentages of DMSO led to complete solubilization, but the localization results remained the same regardless. Thus, stock solutions were sonicated and vortexed immediately prior to use and the lowest possible final DMSO concentration in the media was employed at ≤0.2% for all samples (usually much lower).

Gel shift experiments. Gel shift experiments shown in Figure 4.9 were performed according to previously described procedures at 4 or 20 °C (Chapter 2).[7]

4.7 References

(1) Mapp, A. K., Ansari, A. Z., Ptashne, M., Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 3930-3935.

(2) Ansari, A. Z., Mapp, A. K., Nguyen, D. H., Dervan, P. B., Ptashne, M. *Chem. Biol.* **2001**, *8*, 583-592.

(3) Arora, P. S., Ansari, A. Z., Best, T. P., Ptashne, M., Dervan, P. B. *J. Am. Chem. Soc.* **2002**, *124*, 13067-13071.

(4) Arndt, H.-D., Hauschild, K. E., Sullivan, D. P., Lake, K., Dervan, P. B., Ansari, A. Z. *J. Am. Chem. Soc.* **2003**, *125*, 13322-13323.

(5) Hauschild, K. E., Metzler, R. E., Arndt, H.-D., Moretti, R., Raffaelle, R., Dervan, P. B., Ansari, A. Z. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 5008-5013.

(6) Warren, C. L., Kratochvil, N. C. S., Hauschild, K. E., Foister, S., Brezinski, M. L., Dervan, P. B., Phillips, G. N. Jr., Ansari, A. Z. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 867-872.

(7) Stafford, R. L., Arndt, H.-D., Brezinski, M. L., Ansari, A. Z., Dervan, P. B. *J. Am. Chem. Soc.* **2007**, *129*, 2660-2668.

(8) Best, T. P., Edelson, B. S., Nickols, N. G., Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 12063-12068.

(9) Edelson, B. S., Best, T. P., Olenyuk, B., Nickols, N. G., Doss, R. M., Foister, S., Heckel, A., Dervan, P. B. *Nucleic Acids Res.* **2004**, *32*, 2802-2818.

(10) Kwon, Y., Arndt, H.-D., Mao, Q., Choi, Y., Kawazoe, Y., Dervan, P. B., Uesugi, M. *J. Am. Chem. Soc.* **2004**, *126*, 15940-15941.

(11) Asada, S., Choi, Y., Uesugi, M. *J. Am. Chem. Soc.* **2003**, *125*, 4992-4993.

(12) Shimogawa, H., Kwon, Y., Mao, Q., Kawazoe, Y., Choi, Y., Asada, S., Kigoshi, H., Uesugi, M. *J. Am. Chem. Soc.* **2004**, *126*, 3461-3471.

(13) Arndt, H-D. *Postdoctoral report*, **2004**.

(14) Crowley, K. S., Phillion, D. P., Woodard, S. S., Schweitzer, B. A., Singh, M., Shabany,

H., Burnette, B., Hippenmeyer, P., Heitmeier, M., Bashkin, J. K. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 1565-1570.

(15) Best, T. P., Ph.D. Thesis, California Institute of Technology, **2005**.

(16) Polyamides are believed to undergo active cellular uptake although the mechanism of transport has yet to be identified. See references 8 and 9.

(17) Nickols, N. G., Jacobs, C. S., Farkas, M. E., Dervan, P. B. *Nucleic Acids Res.* **2007**, *35*, 307-316.

(18) Baird, E. E., Dervan, P. B. *J. Am. Chem. Soc.* **1996,** *118*, 6141-6146.

(19) Rucker, V. C., Foister, S., Melander, C., Dervan, P. B. *J. Am. Chem. Soc.* **2003**, *125*, 1195-1202.

(20) Belitsky, J. M., Nguyen, D. H., Wurtz, N. R. Dervan, P. B. *Bioorg. Med. Chem.* **2002**, *10*, 2767-2774.

(21) Wellings, D. A. and Atherton, E. *Methods Enzymol.* **1997**, *289*, 44-67.

(22) Sluka, J. P., Ph.D. Thesis, California Institute of Technology, **1988**.

Chapter 5

Toward a Synthetic HIF-1α Mimic

This project was done in collaboration with Nick Nickols and Mike Brochu (Dervan Group; Caltech).

*The synthesis of polyamides **11**, **20** and all RT-PCR experiments were done by Nick Nickols (Dervan Group; Caltech).*

Abstract

The ability to up-regulate any desired gene through the rational design cell-permeable small molecules would be an incredibly useful tool and potentially have therapeutic value. This chapter describes efforts towards this goal that builds on the Dervan group's previous success at targeting the hypoxia response element (HRE) in the promoter region of vascular endothelial growth factor (VEGF). In living cells, hypoxia inducible factor (HIF-1α) binds to the VEGF HRE and recruits a co-activator, CREB binding protein (CBP), which activates gene expression. To mimic HIF-1α, polyamides that bind to the VEGF HRE have been conjugated to small molecules that bind to CBP. Several small-molecule-polyamide conjugates with different shapes (branched and linear) have been synthesized and their DNA-binding and cell-permeability properties have been determined. Each compound has also been assessed for its ability to affect VEGF gene expression in cell culture experiments. Remarkably, the small molecules without the polyamide activated gene expression by themselves.

5.1 Introduction

The ability to up-regulate the expression of any given gene using small molecules would be a powerful tool for molecular biology.[1-4] Cells up-regulate gene expression by using modular transcription factors that possess a sequence specific DNA binding domain and an activation domain that is capable recruiting the RNA polymerase II complex to the gene promoter.[5] RNA polymerase II subsequently transcribes the coding region into mRNA.[6-7] Promoter regions are loosely defined as -3000 to +300 base pairs from the start site of a given gene.[8] Activation domains of transcription factors can recruit the RNA polymerase complex through *direct* physical association or *indirectly* through co-activator proteins which mediate the interaction.[9-10] This chapter describes efforts toward the development of synthetic transcription factor mimics that are designed to recruit a co-activator to the promoter region of vascular endothelial growth factor (VEGF) in a living cell.

In the presence of low oxygen VEGF is activated by the hypoxia-inducible factor 1-α (HIF-1α)/aryl hydrocarbon receptor (ARNT) heterodimer transcription factor complex that binds the hypoxia response element (HRE) in the VEGF promoter.[11-12] Under hypoxic conditions HIF1-α becomes hydroxylated and is transported from the cytoplasm to the nucleus where it binds to the HRE of VEGF.[13] HIF-1α activates VEGF indirectly using the co-activator CREB binding protein (CBP) to mediate the recruitment of the RNA polymerase II complex.[14] CBP is a general co-activator known to mediate the interaction between numerous transcription factors and RNA polymerase II.[15] Several small molecules have been reported to bind to CBP or the related co-activator p300 (Figure 5.1).[16-20] Furthermore, at least two small molecules have been reported to bind specifically to the HRE of VEGF, including the polyamide CtPyPyIm-γ-PyImPyPy-(+)-FITC.[21-22] The design of small-molecule mimics of HIF-1α incorporated both CBP and HRE-binding small molecules into conjugates intended to recruit CBP to the HRE of VEGF to activate gene expression

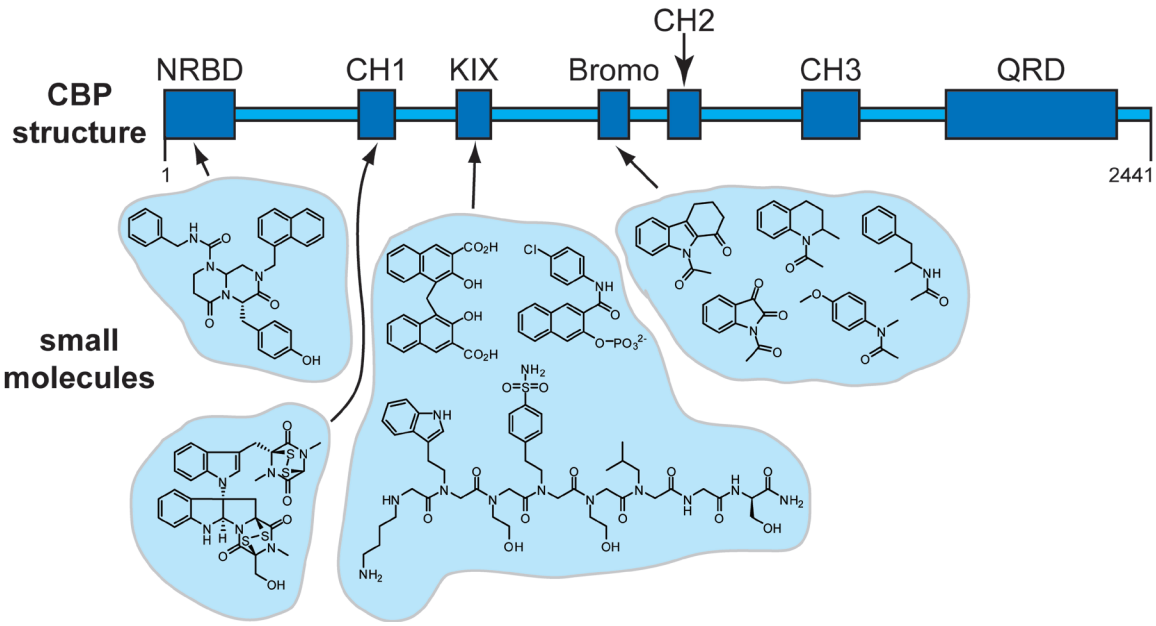


Figure 5.1 The binding sites on the large (~265 kDa) co-activator CREB binding protein (CBP) for several small molecules.

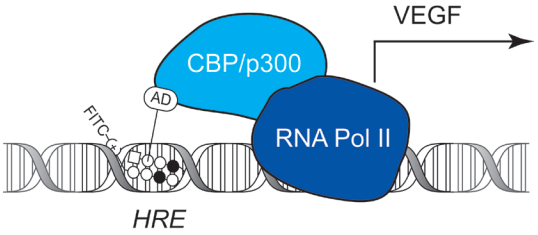


Figure 5.2 Design of a small-molecule mimic of HIF-1α. A DNA-binding polyamide that binds the hypoxia response element (HRE) is conjugated to a CBP-binding small molecule intended to recruit RNA Polymerase II and activate VEGF gene transcription.

(Figure 5.2).

Polyamide conjugates have been shown to act as artificial activators of transcription.[20,23-26] Several peptide activation domains have been shown to be able to activate gene expression *in vitro*.[23-25] Wrenchnolol, a small molecule that binds to the Sur-2 co-activator, has also been shown to serve as an activation domain.[26] Recently, polyamide-peptoid conjugates that were designed to recruit CBP have been reported to modestly activate a transiently transfected luciferase gene in cell culture.[20] This chapter describes conjugates between putative small-molecule activation domains (AD) **1-3** which are known to bind to CBP at different sites (Figure 5.3). Each compound has been modified (**4-6**) to allow conjugation to the polyamide CtPyPyIm-γ-PyImPyPy. Fluorescent analogues were

found to localize to the nucleus of living cells in the presence of verapamil. A couple of conjugates were also shown to bind the HRE of the VEGF promoter by DNase I footprinting. None of the polyamide conjugates reproducibly activated VEGF expression in cell culture. Surprisingly, the ADs **1** and **3** by themselves were found to activate VEGF gene expression.

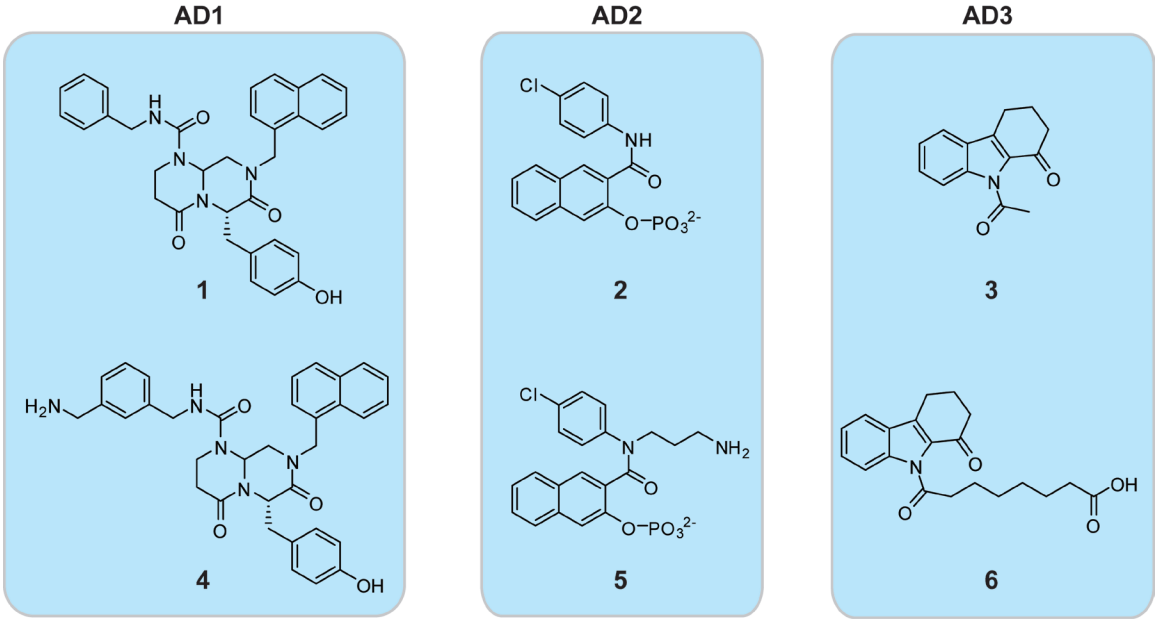


Figure 5.3 Small molecules that bind CBP used as putative activation domains (i.e., AD1 - AD3). The top row shows compounds (**1-3**) that were originally reported to bind CBP. The lower row shows compounds (**4-6**) that were modified so that they can be attached to a polyamide.

5.2 Synthesis of Small Molecule Activation Domains

Before conjugating to polyamides, the CBP-binding small molecules needed to be synthesized. The compounds **1-3** were chosen for their efficacy and synthetic accessibility. ICG-001 (**1**) binds to the nuclear receptor binding domain (NRBD) of CBP and inhibits the CBP/β-catenin interaction with an IC_{50} of 25 μM in cell culture.[17] KG-501 (**2**) binds to the kinase inducible domain (KIX) of CBP and inhibits KID/KIX complex with a K_i of approximately 90 μM.[18] MS7972 (**3**) binds to the bromodomain (BRD) of CBP with a dissociation constant of 19.6 ± 1.9 μM.[19]

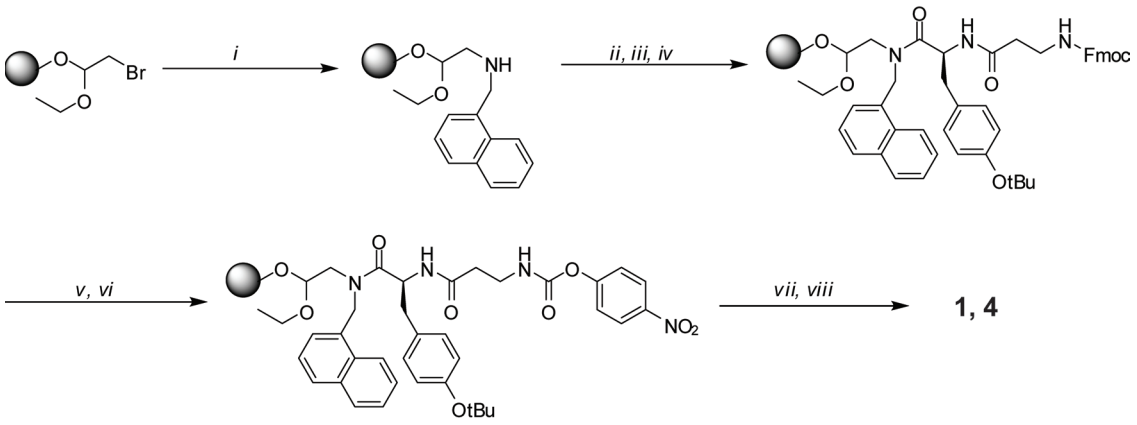


Figure 5.4 Synthesis of AD1 (i.e., **1** and **4**): *i*) 1-naphthalenemethylamine, DMSO, *ii*) Fmoc-Tyr-OH, DIC, HOAt, *iii*) 20% piperidine in DMF, *iv*) Fmoc-βAla-OH, DIC, HOBt, *v*) 20% piperidine in DMF, *vi*) 4-nitrophenyl chloroformate, *vii*) benzylamine or *m*-xylylenediamine, *viii*) formic acid, 20 °C, ~24 hours

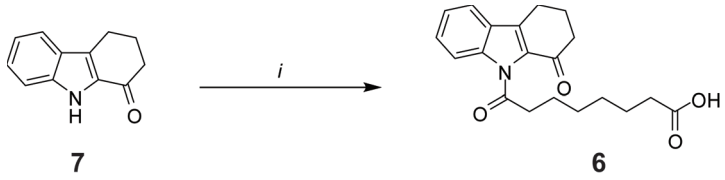


Figure 5.5 Synthesis of AD3 (i.e., **6**): *i*) suberic anhydride, NaH, DMSO

The synthesis of AD1 (**1** and **4**) was performed on solid-phase using bromoacetal resin as described previously (Figure 5.4).[17,27-29] Briefly, the resin-bound bromine was displaced with a naphthylamine and two cycles of standard Fmoc peptide chemistry were performed to add tyrosine and β-alanine amino acids. Following urea formation with either benzylamine or *m*-xylylenediamine, treatment with formic acid led to concomitant resin cleavage and tandem cyclization reactions to yield **1** and **4**, respectively. The synthesis suitably protected analogues of AD2 (**2** and **5**) was performed by Mike Brochu.[30] Also, the synthesis of AD3 (**3**) with a suitable linker (**6**) was performed in one step from a commercially available carbazole **7** by acylating with suberic anhydride (Figure 5.5).

5.3 Rationale for Linker Attachment Position

In modifying compounds **1**-**3** to yield **4**-**6** to enable attachment to a polyamide, it

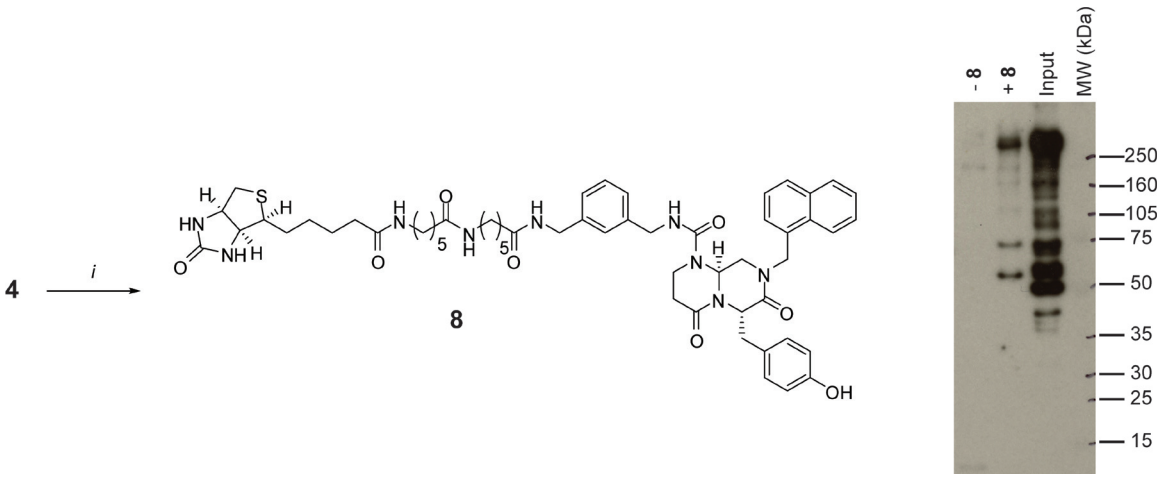


Figure 5.6 Biotin-conjugate pull-down experiments. On the left, synthesis of biotinylated analogoues of AD1 and AD3: *i*) biotinamidohexanoyl-6-aminohexanoic acid *N*-hydroxysuccinimide ester, DIEA in DMF. On the right, a Western blot with a CBP-specific antibody following a pull-down experiment is shown. Biotinylated-AD1 (**8**) is able to bind CBP (~265 kDa) in a mixture of SW480 nuclear extract and maintain this interaction through thorough washing.

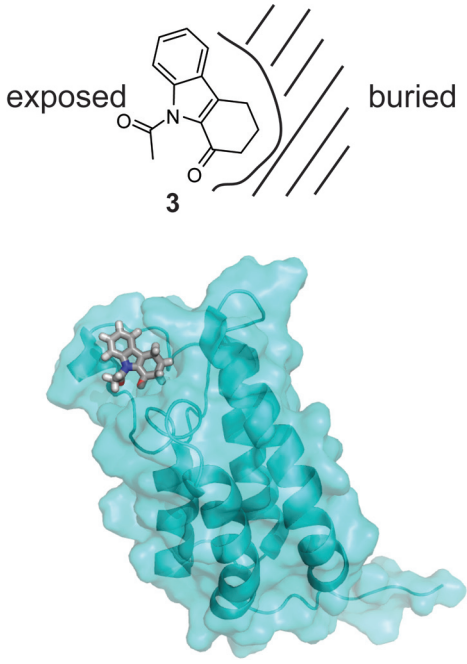


Figure 5.7 NMR structure of the bromodomain of CBP with **3** bound (PDB code 2D82). The acetyl moiety projects away from the binding pocket suggesting this would make a reasonable attachment point.

was unclear whether or not the change would result in elimination of their CBP binding activity. Thus, a pull-down assay was employed to confirm no loss of activity as was previously described for compound **8** (Figure 5.6).[17] First, the biotinylated analogue **8** was synthesized from **4** by treating with a commercially-available biotin succinimidyl ester. A pull-down assay involves treating streptavidin-coated magnetic beads with a biotinylated compound and incubating with input cell lysate which contains a protein of interest. Following thorough washing, the bound protein(s) is(are) eluted by treated with detergent (SDS) and heat (95 °C) and the resulting mixture is separated on an SDS-PAGE gel and Western blotted. Thus, compound **8** was confirmed to pull down CBP from the input SW480 colon cancer cell nuclear extract as evidenced by a band near the expected molecular weight (~265 kDa) that was stained by an anti-CBP antibody (Figure 5.6). Mike Brochu has shown that a biotinylated analogue of **2** is able to pull-down CBP from HeLa cell nuclear extract (data not shown).[30] Although a similar pull-down experiment was not performed for **3** (or **6**), NMR structural data suggests that the acetyl substituent is exposed from the CBP binding pocket[19] so this was chosen as the attachment point (Figure 5.7).

Table 5.1 Frequency of match sites in the VEGF promoter

Sequence	Frequency
WTWCGW	**3**
WGWWWW	23
WGGWWW	**30**
WGWGWW	17
WGWWGW	21
WGWWCW	6
WGWCWW	12
WGCWWW	13
WGGGWW	23
WGGWGW	18
WGGWCW	14
WGGCWW	15
WGWGGW	18
WGWGCW	18
WGCGWW	1
WGCWGW	12
WGCWCW	18
WGWCGW	1
WGWCCW	14
WGCCWW	13
WGGGGW	24
WGCGGW	4
WGGCGW	2
WGGGCW	17
WGCCGW	4
WGGCCW	10
WGCGCW	1
WGCCCW	17

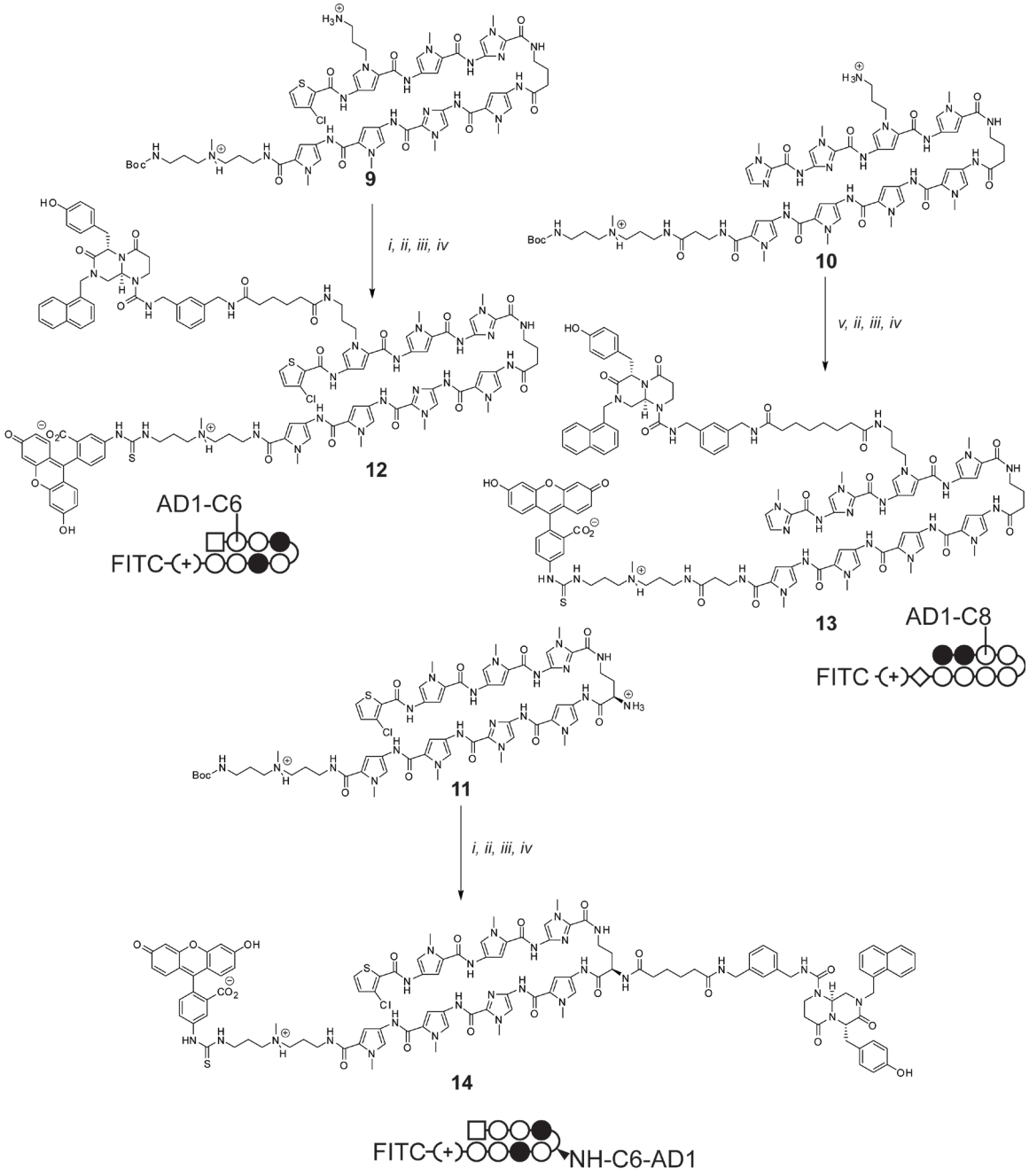


Figure 5.8 Synthesis of polyamide-AD-fluorescein conjugates for cell uptake studies: *i*) excess adipic acid, PyBOP, DIEA, DMF, *ii*) **4**, PyBOP, DIEA, DMF, *iii*) 50% TFA/DCM, *iv*) fluorescein-5-isothiocyante, DIEA, DMF, *v*) suberic anhydride, DIEA, DMF.

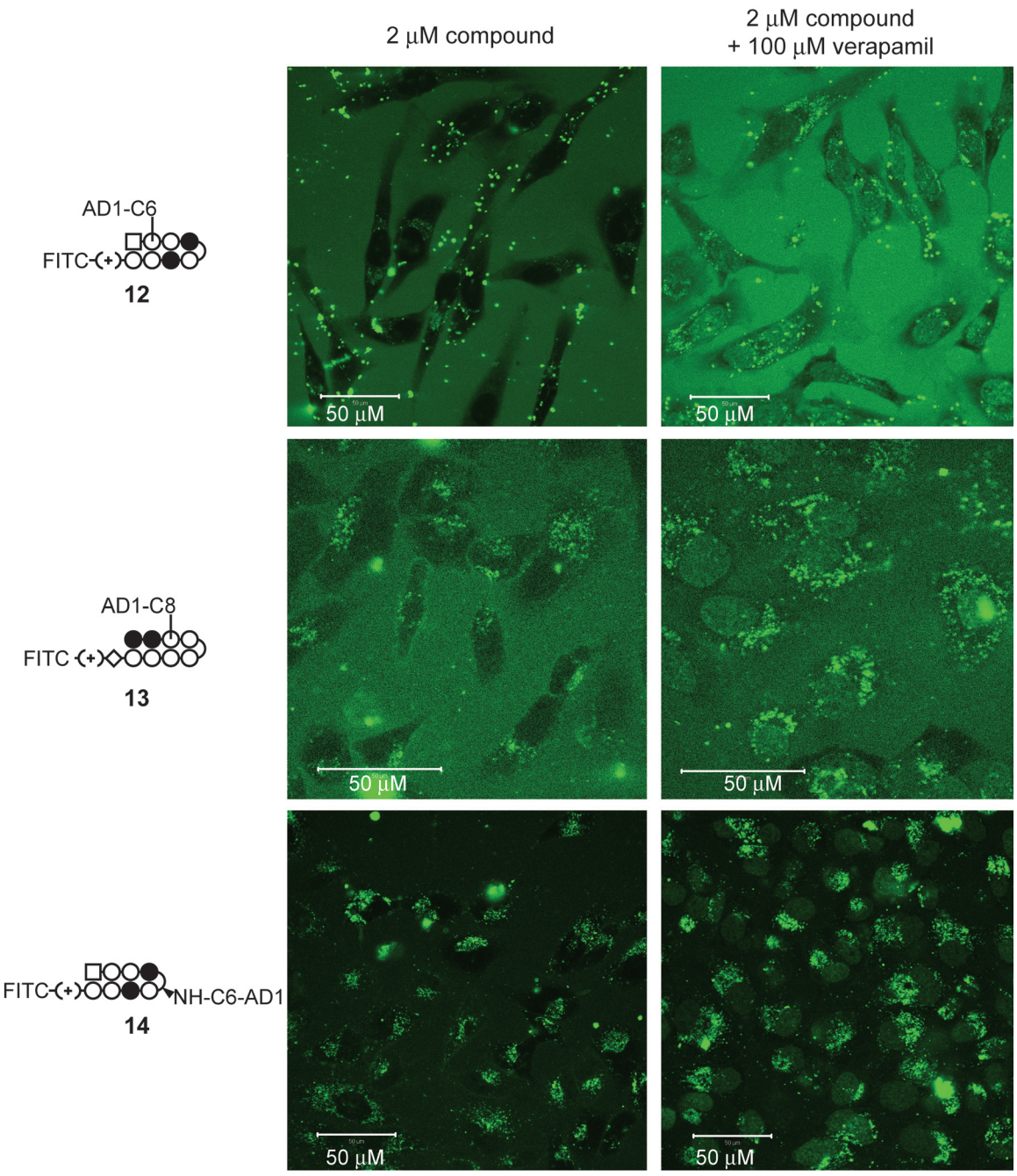


Figure 5.9 Results for uptake experiments in HeLa cells for conjugates **12-14**. On the left it is shown that all compounds exhibit extracellular and vesicular staining. On the right it is shown that 100 μM verapamil induces partial nuclear localization. (Scale bar = 50 μm)

5.4 Cell Uptake of Fluorescent Conjugates

To generate compounds suitable for confocal microscopy cell uptake studies, the selectively protected polyamides **9-11** were synthesized first by standard methods with Boc-β-Ala-PAM[31] or oxime resin[32] (Figure 5.8). The polyamide of **10** was used because analysis of the VEGF promoter showed a high frequency for its match site 5′-WGGWWW-3′ (Table 5.1). First, a linker was coupled to the polyamide primary amine using either excess adipic acid (6 carbons) or suberic anhydride (8 carbons). Then compound **4** was attached through an amide bond, followed by deprotection, and labeling with fluorescein isothiocyanate to yield **12-14**. Confocal microcopy with **12-14** in HeLa, MCF-7, and PC-3 cells showed exclusively extracellular and vesicular stains, but the addition of the MDR inhibitor verapamil (100 μM) caused a partial redistribution to the cell nucleus (Figure

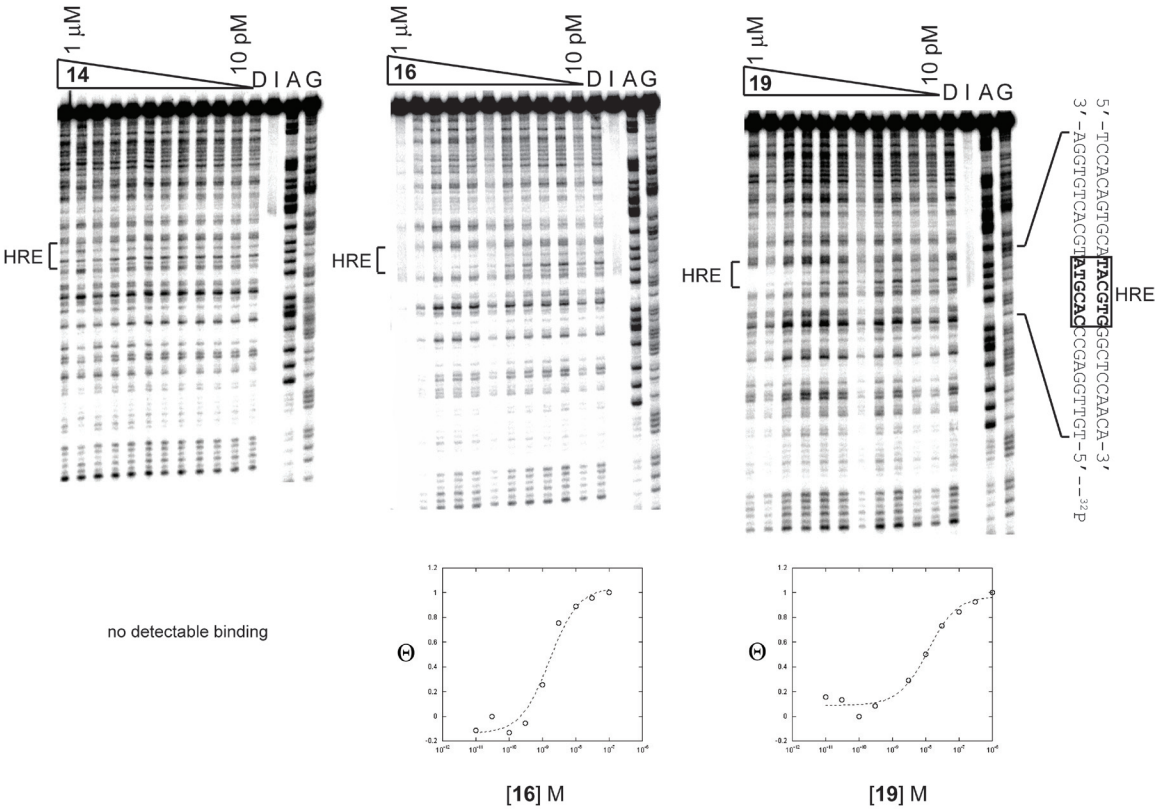


Figure 5.10 Quantitative DNase I footprinting experiments for polyamide-AD conjugates on a stretch of the VEGF promoter which contains the HRE. Compound **14** shows no apparent DNA binding activity (i.e., $K_a \leq 10^6\ M^{-1}$). Compounds **16** and **19** bind to the HRE with K_a's of $6.9 \pm 0.6 \times 10^8\ M^{-1}$ and $8 \pm 5 \times 10^7\ M^{-1}$, respectively.

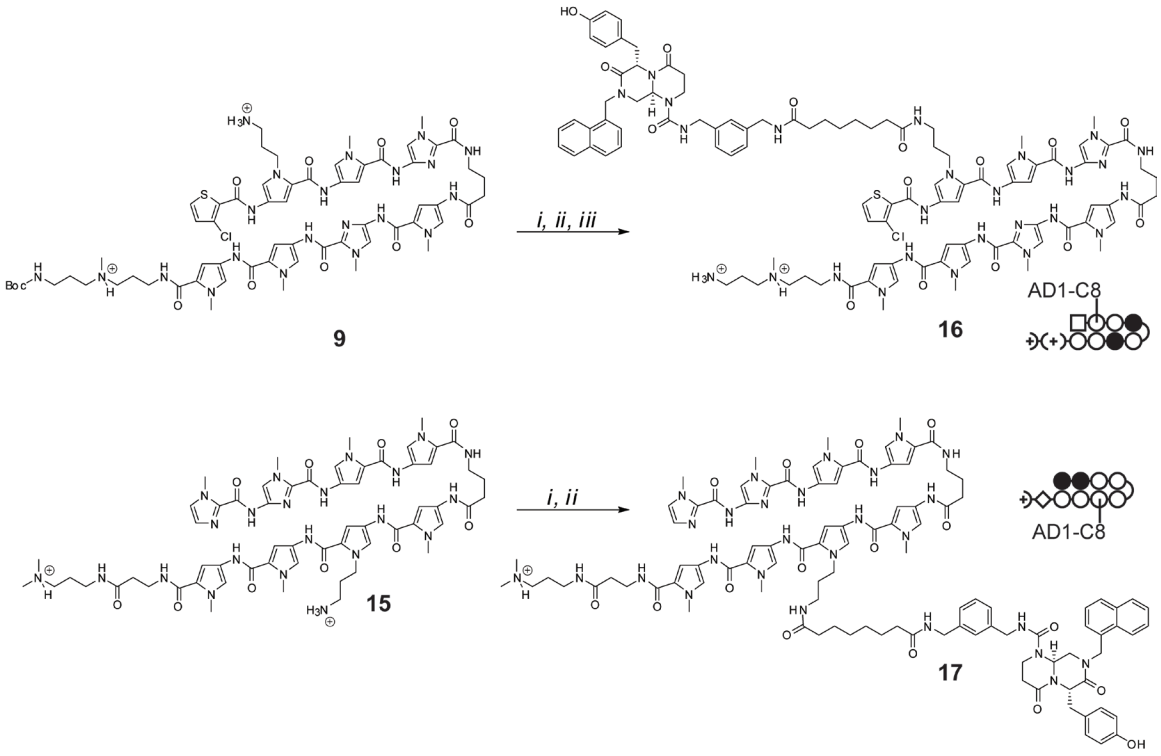


Figure 5.11 Synthesis of *branched* polyamide-AD conjugates: *i*) suberic anhydride, DIEA, DMF, *ii*) **4**, PyBOP, DIEA, DMF, *iii*) 50% TFA/DCM.

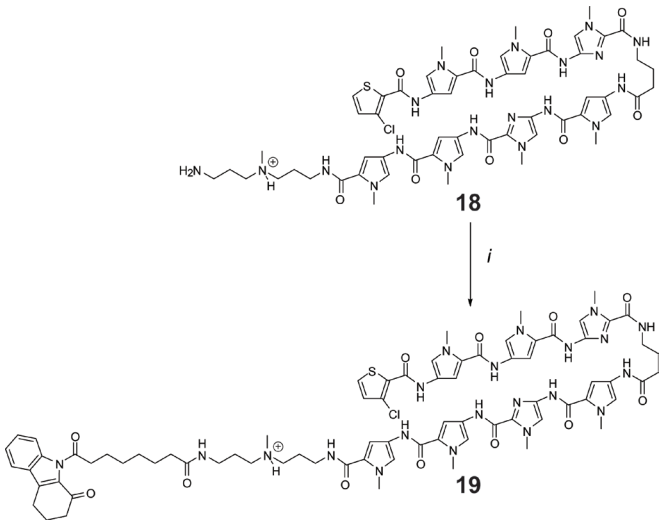


Figure 5.12 Synthesis of *linear* polyamide-AD conjugates: *i*) **6**, PyBOP, DMF, DIEA, 20 °C, 2 hours.

5.9). Titration of verapamil with **13** in HeLa cells suggested that 10 μM verapamil was sufficient to induce nuclear localization so this concentration was used for the RT-PCR described later.

5.5 DNase I Footprinting and Non-Fluorescent Conjugate Synthesis

DNase I footprinting was used to determine the HRE-affinity of polyamide-AD1-FITC conjugate **14** (Figure 5.10). The analogous parent polyamide-FITC conjugate with a free chiral amine at the turn has been reported to have an HRE-affinity of 6.3×10^9 M^{-1} but conjugate **14** shows no apparent DNA binding ($K_a \leq 10^6$ M^{-1}).[21] Attachment of the linker and AD1 appears to be responsible for the reduced DNA affinity. The placement of the linker and AD1 over the three Gs adjacent to the HRE may be particularly detrimental to the binding affinity. Thus, simpler polyamide conjugates with different shapes were designed in hopes they would possess better DNA binding properties (Figures 5.11 and 5.12).

Branched conjugate **16** was synthesized from parent polyamide **10** and small molecule **4** (AD1) after incorporation of an eight-carbon linker (Figure 5.11). The linear conjugate **19** was synthesized from the parent polyamide **18** (Figure 5.12). Compounds **16** and **19** were also footprinted to determine their HRE-binding affinity (Figure 5.10). The presumably dicationic polyamide-AD1 conjugate **16** showed a higher affinity for the HRE ($K_a = 6.9 \pm 0.6 \times 10^8$ M^{-1}) but bound the DNA non-specifically down to a concentration of 300 nM. Linear polyamide-AD3 conjugate **19** revealed a low affinity for the HRE ($K_a = 8 \pm 5 \times 10^7$ M^{-1}). Compound **19**'s affinity would likely be increased by the addition of a chiral amine on the turn.[33]

5.6 Effect of Compounds on VEGF Expression

A total of all six conjugates representing branched and linear configurations were subjected to HeLa cells (i.e., **12**, **13**, **14**, **16**, **17**, and **19**) under non-induced conditions (no

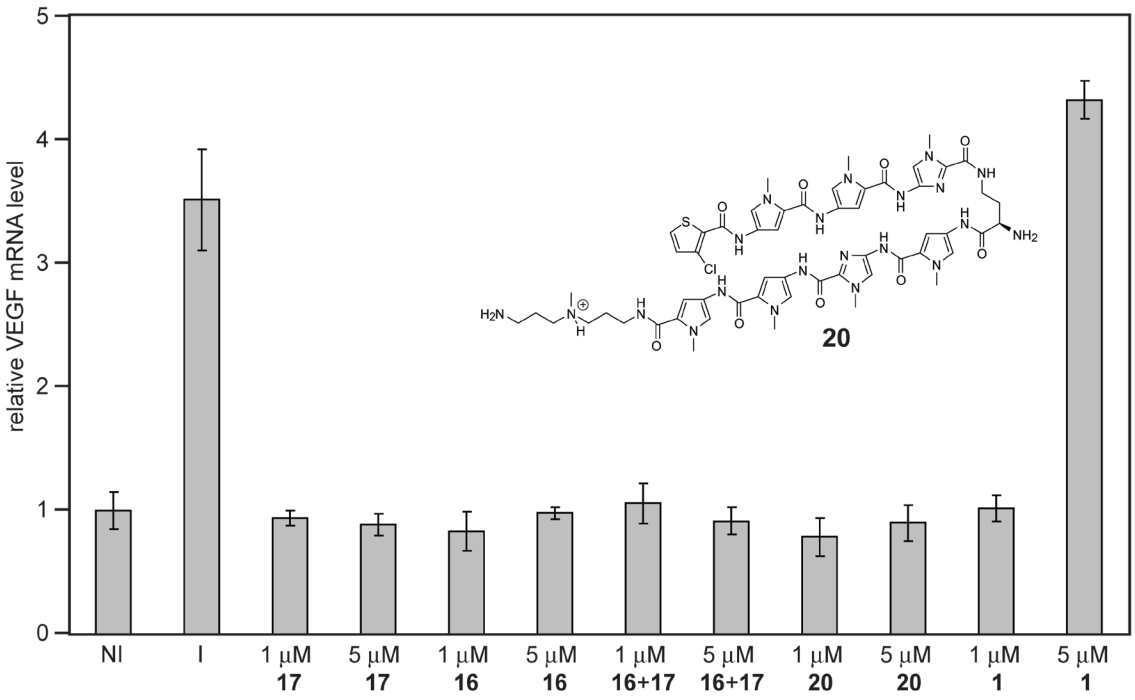


Figure 5.13 RT-PCR experiments measuring the effect of each compound on VEGF expression in the presence of 10 μM verapamil. NI = non-induced baseline control and I = induced (DFO) positive control. Compounds **16** and **17** previously showed weak induction of VEGF gene expression (~1.5-fold increase), but this result was found to be not reproducible as shown here. Combinations of **16** and **17** also do not lead to activation, nor does the polyamide control **20**. Intriguingly, the CBP binding small-molecule **1** by itself activates VEGF gene expression.

DFO) ranging from concentrations of 1 to 15 μM in the presence of 10 μM verapamil. The branched conjugate **17** was also tried since it targets 5′-WGGWWW-3′ which is found with a high frequency in the VEGF promoter (Figure 5.11 and Table 5.1). Quantitative real-time PCR (qRT-PCR) was used to measure the effect of each compound on VEGF gene expression. Of the compounds screened, only **16** and **17** affected VEGF expression (~1.5-fold increase), but this effect was not reproducible (data not shown). Experiments looking for synergistic effects using combinations of conjugates **16** and **17** did not lead to a significant increase in VEGF mRNA levels (Figure 5.13). Remarkably, the small molecules **1** and **3** by themselves, but not **2**, showed a reproducible and significant increase in VEGF mRNA and protein levels (data not shown). This effect by the small molecules themselves was

unanticipated and makes de-convolution of experiments with their conjugates problematic. Investigation into the mechanism of VEGF activation by **1** and **3** is ongoing.

5.7 Experimental Details

Synthesis

ICG-001, AD1 (1). Synthesis was based on described previously described procedures.[17,27-29] Bromoacetal resin resin (333 mg, 0.5 mmol scale, 75-100 mesh, 1% DVB , ~1.5 mmol/g, Aldrich) was loaded with 1-naphthalenemethylamine (1.44 mL) in DMSO (4 mL) for 20 hours at RT. Resin drained and rinsed 2x DMF, 2x DCM, and 2x DMF. Activated Fmoc-Tyr(OtBu)-OH (Novabiochem, 919 mg, 4 eq) with HOAt (AnaSpec, 272.2 mg, 4 eq) and DIC (Aldrich, 310 µL, 4 equivalents) in 5.6 mL of DMF at RT for 20 min before adding to the resin. DIEA (697 µL, 8 equivalents) was added and the reaction was allowed to proceed for 24 hours. Drained and rinsed 2x with DMF. Fmoc-Tyr(OtBu)-OH was coupled again (1.5 eq) for another 24 hours. The partially unreacted resin-bound secondary amine was capped with 10 mL of 4:1 DMF:Ac_2O for 2.5 hours. The resin was rinsed as above. The Fmoc protecting group was removed with 20% piperidine in DMF for 25 min and rinsed as usual. Fmoc-β-Ala-OH (622.6 mg, 4 equivalents) was pre-activated with HBTU (739.5 mg, 3.9 equivalents) and DIEA (522 µL) and coupled to the resin overnight. The resin was rinsed and deprotected as before. 4-nitrophenylchloroformate (503.9 mg, 5 eq) was coupled in DCM with DIEA (522 µL, 6 eq) for 1 hour at RT. Rinsed 2x with DCM and added DMF (~3 mL) and benzylamine (218 µL, 4 equivalents) letting the reaction proceed for 5 hours. Rinsed 2x DMF, 2x DMF, 2x DCM and dried *in vacuo* for 5 min. Cyclization and cleavage was performed with formic acid (88%, 8.7 mL) for 10.5 hours at RT. The resin was filtered with a small plastic filter. The filtrate concentrated under reduced pressure and dissolved in 0.1% TFA in water and acetonitrile, purified by preparative HPLC, and lyophilized to yield 25.3 mg of a white powder (46.1 µmol, 9.2% yield measured by UV). ^{1}H NMR (d_6-DMSO, 300 MHz) δ 2.04-2.09 (m, 2H), δ 3.00-3.15 (m, 3H), δ 3.22-3.28 (m, 1H), δ 3.55 (t, J = 11.4

Hz, 1H), δ 3.84 (m, 1H), δ 4.13-4.20 (dd, J_1 = 6.0 Hz, J_2 = 15.6 Hz, 1H), δ 4.28-4.35 (dd, J_1 = 6.0 Hz, J2 = 15.3 Hz, 1H), δ 4.89 (d, J = 15.3 Hz, 1H), δ 5.12 (d, J = 15.6 Hz, 1H), δ 5.10-5.13 (m, 1H), δ 5.74-5.78 (d of d, J_1 = 3.6 Hz, J_2 = 10.5 Hz, 1H), δ 6.54 (d, J = 8.7 Hz, 2H), δ 6.90 (d, J = 8.7 Hz, 2H), δ 7.21-7.38 (m, 6H), δ 7.44-7.60 (m, 4H), δ 7.87-7.97 (m, 2H), δ 8.12-8.14 (m, 1H), δ 9.18 (bs, 1H); A_{max} at 281 nm (ε = 7,860 $M^{-1}\cdot cm^{-1}$ in 20% acetonitrile in 0.1% TFA/water). (ESI-MS) $[M+H]^+$ calc'd for $C_{33}H_{33}N_4O_4$ 549.2, observed 549.1.

KG-501, AD2 (2). This compound is a common histological stain commercially available from several sources (e.g., Fluka).

MS7972, AD3 (3). This compound is commercially available in small quantities (ChemBridge Corporation, San Diego, CA).

ICG-001-aminomethyl, AD1-NH_2 (4). The synthesis of this compound was performed as described for compound **1** replacing benzylamine with *m*-xylylenediamine (518 μL). After cleavage from resin and HPLC purification 12.6 mg of a white powder was obtained (~6% yield). ^{1}H NMR (d_6-DMSO, 300 MHz) δ 2.06-2.13 (m, 2H), δ 3.01-3.12 (m, 4H), δ 3.57 (t, 1H, J = 10.8 Hz), δ 3.82-3.86 (m, 1H), δ 3.98 (d, 2H, J = 6 Hz), δ 4.16-4.23 (dd, 1H, J_1 = 5.7 Hz, J_2 = 15.15 Hz), δ 4.25-4.32 (dd, 1H, J_1 = 5.7 Hz, J_2 = 15.15 Hz), δ 4.87 (d, 1H, J = 15.3 Hz), δ 5.14 (d, 1H, J = 14.7 Hz), δ 5.12-5.17 (m, 1H), δ 5.72 (dd, 1H, J_1 = 3.9 Hz, J_2 = 10.8 Hz), δ 6.54 (d, 2H, J = 8.7 Hz), δ 6.90 (d, 2H, J = 8.7 Hz), δ 7.21 (d, 1H, J = 7.5 Hz), δ 7.28-7.40 (m, 4H), δ 7.44-7.62 (m, 4H), δ 7.89 (d, 1H, J = 8.1 Hz), δ 7.94-7.97 (m, 1H), δ 8.11 (bs, 2H), δ 8.11-8.14 (m, 1H), δ 9.18 (s, 1H); HRMS (FAB) exact mass calcd for $C_{34}H_{36}N_5O_4$ requires 578.2767 *m/z*, found 578.2754 *m/z*.

MS7972-suberic acid, AD3-CO_2H (6). First, 2,3,4,9-tetrahydro-1H-carbazol-1-one **7** (150 mg, 0.81 mmol, 1 eq, Bionet Research Intermediates) was dissolved in DMSO (1 mL, anhydrous) and sodium hydride (64.8 mg, 1.62 mmol, 2 eq, 60% oil dispersion) was added as a solution in DMSO (1 mL). The solution was sonicated to aid dissolution. After 5 min, suberic anhydride (151.7 mg, 0.97 mmol, 1.2 eq) was added as a solution in DMSO (1.5 mL). Immediately a precipitate formed and the solution was poured into 10% citric acid

buffer and extracted 3x with DCM (50 mL). The organic layers were combined, dried with anhydrous sodium sulfate, filtered, and dried under reduced pressure to give a thin orange film. The desired product was purified by silica gel chromatography (1:1 EtOAc:Hexanes + 0.1% acetic acid, $R_f \approx 0.3$) and dried under reduced pressure. The resulting film was dissolved in MeOH (500 μL) and added to water to form an off-white precipitate which was flash frozen and lyophilized (19.11 mg, 7% yield). ^{1}H NMR (d_6-DMSO, 300 MHz) δ 1.20 (pent, J = 3.6 Hz, 4H), δ 1.40 (m, 2H), δ 1.60 (m, 2H), δ 2.10-2.20 (m, 4H), δ 2.64 (t, J = 6.9 Hz, 2H), δ 2.86 (t, J = 7.2 Hz, 2H), δ 3.00 (t, J = 6.0 Hz, 2H), δ 7.33 (t, J = 7.8 Hz, 1H), δ 7.53 (t of d, J_1 = 1.2 Hz, J_2 = 8.4 Hz, 1H), δ 7.79 (d, J = 8.1 Hz, 1H), δ 7.97 (d, J = 8.7 Hz, 1H), δ ~12 (bs, 1H); ^{13}C NMR ($CDCl_3$, 300 MHz) δ 189.78, 178.89, 175.82, 139.60, 139.33, 132.38, 130.32, 126.71, 123.82, 121.45, 115.93, 39.74, 39.40, 34.02, 29.07, 28.96, 25.72, 24.79, 23.69, 22.38. HRMS (FAB) exact mass calcd for $C_{20}H_{24}NO_4$ requires 342.1705 *m/z*, found 342.1717 *m/z*. Absorbance maximum at 310 nm in water with 2% DMSO (ε at 308-309 nm = 2.2 ± 0.8 x 10^4 $M^{-1}cm^{-1}$).

ICG-002, AD1-biotin (8). An aliquot of **4** was dissolved in 200 μL of 0.1 M DIEA in DMF and 2.3 mg of biotinamidohexanoyl-6-aminohexanoic acid *N*-hydroxysuccinimide ester (Sigma) was added. The product formed immediately and was purified by preparative HPLC and lyophilized to yield a fine white powder (1.5 μmol). (MALDI-TOF MS) Calcd for $C_{56}H_{72}N_9O_8S^+$ 1030.5, observed 1031.0.

CtPy$^{(propylamine)}$PyIm-γ-PyImPyPy-(+)-NHBoc (9). This polyamide was synthesized on Marshall-Liener resin and purified by preparative HPLC (Appendix D). (MALDI-TOF MS) Calcd for $C_{63}H_{82}ClN_{21}O_{11}S^+$ 1375.6, observed 1375.7.

ImImPy$^{(propylamine)}$Py-γ-PyPyPyPy-β-(+)-NHBoc (10). This polyamide was synthesized on PAM resin (1.94 μmol, 3% overall yield). (MALDI-TOF MS) Calcd 1408.7, observed 1408.7.

CtPyPyIm-γ$^{NH2(R)}$-PyImPyPy-(+)-NHBoc (11). This polyamide was synthesized by Nick Nickols on oxime resin and purified by preparative HPLC. (MALDI-TOF MS) Calcd for

$C_{61}H_{77}ClN_{21}O_{11}S^+$ 1347.6, observed 1346.5.

CtPy$^{(AD1\text{-}C6\text{-}propylamine)}$PyIm-γ-PyImPyPy-(+)-FITC (12). This conjugate was synthesized essentially as described below for **14**. (ESI MS) Calc'd $[M+2H]^{2+}$ $C_{119}H_{125}ClN_{27}O_{20}S_2^{2+}$ 1175.9, observed 1176.2.

ImImPy$^{(AD1\text{-}C8\text{-}propylamine)}$Py-γ-PyPyPyPy-β-(+)-FITC (13). Polyamide **10** was coupled to suberic anhydride as described for compound **17** and the resulting intermediate was purified by preparative HPLC and lyophilized (609 nmol, 31% yield over 1 step). (MALDI-TOF MS) Calcd 1564.8, observed 1564.7. The Boc group was removed and the FITC coupled as described below for **14**. The final product was purified by preparative HPLC and lyophilized. (ESI MS) Calcd $[M+2H]^{2+}$ 1207.5, observed 1207.3.

CtPyPyIm-γ$^{(AD1\text{-}C6\text{-}NH2(R))}$-PyImPyPy-(+)-FITC (14). Polyamide **11** (1.55 μmol) was treated with excess adipic acid (10 eq) and PyBOP (20 eq) in ~300 μL of 0.1 M DIEA in DMF. The resulting intermediate was purified by preparative HPLC and lyophilized (560 nmol, 36% yield for 1 step). (MALDI-TOF MS) Calcd for $C_{67}H_{84}ClN_{21}O_{14}SNa^+$ 1496.6, observed 1496.6. The previous intermediate was activated with PyBOP (6.7 mg) in 600 μL 0.1 M DIEA in DMF and added to amine **4**. The compound was precipitated from diethyl ether (10 mL) cooled over dry ice, centrifuged, and dried briefly with $N_{2(g)}$. The Boc protecting group was removed with 50% TFA/DCM (1 mL) for 30 min at RT and the compound was precipitated from ether and dried as above. The pellet was redissolved in 1 M DIEA in DMF (600 μL) and fluorescein-5-isothiocyanate (2 μmol) was added. The final product was purified by preparative HPLC and lyophilized to give a yellow-orange solid (179 nmol, 32% final 2 steps). (ESI MS) Calc'd $[M+2H]^{2+}$ $C_{117}H_{122}ClN_{27}O_{20}S_2^{2+}$ 1162.42, observed 1162.1.

ImImPyPy-γ-PyPy$^{(propylamine)}$PyPy-β-Dp (15). This polyamide was synthesized on Boc-β-Ala-PAM resin. (MALDI-TOF MS) Calcd for $C_{60}H_{77}N_{22}O_{10}S^+$ 1265.6, observed 1265.6.

CtPy$^{(AD1\text{-}C8\text{-}propylamine)}$PyIm-γ-PyImPyPy-(+)-NH$_3^+$ (16). Suberic anhydride (3 eq) was reacted with polyamide **9** as described for compound **17** below. The intermediate was

purified by preparative HPLC and lyophilized (624 nmol, 40% yield over 1 step.) (MALDI-TOF MS) Calcd 1530.7, observed 1530.5. Amine **4** was coupled to this intermediate as described for **14** and the final product was purified by preparative HPLC (193 nmol, 31% yield over 1 step). (MALDI-TOF MS) Calcd 1990.9, observed 1990.9.

ImImPyPy-γ-PyPy$^{(AD1\text{-}C8\text{-}propylamine)}$PyPy-β-Dp (17). Polyamide **15** (1.4 μmol) was treated with suberic anhydride (1.5 eq) in 200 μL 0.1 M DIEA in DMF and heated to 35 °C for 1 hour. The intermediate was purified by preparative HPLC and lyophilized to give a white powder (421 nmol, 30% yield). (MALDI-TOF MS) Calcd for $C_{68}H_{89}N_{22}O_{13}^+$ 1421.7, observed 1422.1. The previous intermediate was activated with PyBOP (10 equivalents) in 600 μL 0.1 M DIEA in DMF and added to amine **4** (1.5 equivalents). The final product was purified by preparative HPLC and lyophilized to give a white powder (145 nmol, 36% yield over 1 step). (MALDI-TOF MS) Calcd for $C_{102}H_{122}N_{27}O_{16}^+$ 1981.0, observed 1980.7.

CtPyPyIm-γ-PyImPyPy-(+)-NH$_3^+$ (18). This polyamide was synthesized on oxime resin. (MALDI-TOF MS) Calcd 1231.5, observed 1231.4.

CtPyPyIm-γ-PyImPyPy-(+)-C$_8$-AD3 (19). Polyamide **18** (1.19 μmol) was coupled to **6** (0.85 mg, 2.5 μmol) using PyBOP (6 μmol) in a solution of 0.1 M DIEA in DMF (200 μL) and purified by preparative HPLC and lyophilized (561 nmol, 47 % yield). (MALDI-TOF MS) Calcd 1554.6, observed 1554.7. The extinction co-efficient of the conjugate was assumed to be the sum of the individual components (i.e., ε at 310 nm of 91,500 $M^{-1}cm^{-1}$= 69,500 $M^{-1}cm^{-1}$ + 22,000 $M^{-1}cm^{-1}$).

CtPyPyIm-γ$^{NH2(R)}$-PyImPyPy-(+)-NH$_3^+$ (20). This compound was synthesized and characterized by Nick Nickols.

Pull-down assay and Western blot. The pull-down was based on a previously published experiment.[17] 30 μL of 10 mg/mL Dynabeads M-280 Streptavidin coated (Invitrogen) were aliquoted into an eppendorf tube to which 4 nmol of biotin conjugate **8** was added as a 1:1 solution of DMSO:PBB for a final volume of 30 μL. (PBB = 20 mM HEPES,

100 mM NaCl, 0.5 mM EDTA, 6 mM $MgCl_2$, 0.5% Nonidet P-40 (Tergitol, Type NP-40, Sigma), 1 tablet complete protease mixture (Roche), 5 mM β-mercaptoethanol added immediately before use, pH 7.9, total volume 50 mL) Beads incubated overnight and then washed 3x with PBB. Added 28.5 µL of PBB and 1.5 µL of SW480 nuclear extract (1.88 mg/mL Bradford concentration) obtained previously from Pierce Nuclear Extract Kit (Cat # 78833) and allowed to incubate for ~2 hours at RT. The tubes were put on ice for 10 min and washed 3x with cold PBB in the cold room. Added PBS:Reducing Agent: LDS 9.1:1.4:3.5 and heated to 90 °C for 5 min and ran on an SDS-PAGE gel (Invitrogen, NuPAGE 10% bis-TRIS gel). The resulting gel was transferred to a membrane and blotted with a primary anti-CBP antibody (1:200, A-22, sc-369, Santa Cruz Biotechnology) and a secondary goat anti-rabbit HRP (1:5000, Santa Cruz Biotech). The resulting gel was visualized with SuperSignal West Pico (Pierce) and developed with a 90-second exposure to light-sensitive film in the dark room.

Cell culture and confocal microscopy experiments. The cell culture and confocal microscopy experiments were performed as described in chapter 4. Racemic verapamil (Aldrich, 98%) was added at the same time as the conjugates.

DNase I footprinting. Footprinting was performed as described in Chapter 2 using a previously published plasmid of a region of the VEGF promoter containing the HRE.[21]

Quantitative real-time PCR (qRT-PCR). These experiments were performed by Nick Nickols essentially as described previously.[21,34-35]

5.8 References

(1) Denison, C. and Kodadek, T. *Chem. Biol.* **1998**, *5*, R129-R145.

(2) Ansari, A. Z. and Mapp, A. K. *Curr. Opin. Chem. Biol.* **2002**, *6*, 765-772.

(3) Arndt, H.-D. *Angew. Chem. Int. Ed.* **2006**, *45*, 4552-4560.

(4) Dervan, P. B. and Edelson, B. S. *Curr. Opin. Struct. Biol.* **2003**, *13*, 284-299.

(5) Ptashne, M. and Gann, A. *Genes & Signals* Cold Spring Harbor Lab. Press, Plainview, NY, 2002.

(6) Lee, T. I. and Young, R. A. *Annu. Rev. Genet.* **2000**, *34*, 77-137.

(7) Kadonaga, J. T. *Cell* **2004**, *116*, 247-257.

(8) Zhang, X., Odom, D.T., Koo, S.-H., Conkright, M.D., Canettieri, G., Best, J., Chen, H., Jenner, R., Herbolsheimer, E., Jacobsen, E., Kadam, S., Ecker, J.R., Emerson, B., Hogenesch, J.B., Unterman, T., Young, R.A., Montminy, M. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 4459-4464.

(9) Näär, A. M., Lemon, B. D., Tjian, R. *Annu. Rev. Biochem.* **2001**, *70*, 475-501.

(10) Bourbon, H-M., Aguilera, A., Ansari, A. Z., Asturias, F. J., Berk, A. J., Bjorklund, S., Blackwell, T. K., Borggrefe, T., Carey, M. *et al. Mol. Cell* **2004**, *14*, 553-557.

(11) Kaelin, W. G. *Genes Dev.* **2002**, *16*, 1441-1445.

(12) Forsythe, J. A., Jiang, B.-H., Iyer, N. V., Agani, F. Leung, S. W., Koos, R. D., Semenza, G. L. *Mol. Cell. Biol.* **1996**, *16*, 4604-4613.

(13) Ivan, M., Kondo, K., Yang, H., Kim, W., Valiando, J., Ohh, M., Salic, A., Asara, J. M., Lane, W. S., Kaelin, W. G. *Science* **2001**, *292*, 464-468.

(14) Dames, S. A., Martinez-Yamout, M., De Guzman, R. N., Dyson, H. J., Wright, P. E. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 5271-5276.

(15) MaManus, K. J., Hendzel, M. J. *Biochem. Cell. Biol.* **2001**, *79*, 253-266.

(16) Kung, A. L., Zabludoff, S. D., France, D. S., Freedman, S. J., Tanner, E. A., Vieira, A., Cornell-Kennon, S., Lee, J., Wang, B., Wang, J., Memmert, K., Naegeli, H.-U., Peterson, F., Eck, M. J., Bair, K. W., Wood, A. W., Livingston, D. M. *Cancer Cell* **2004**, *6*, 33-43.

(17) Emami, K. H., Nguyen, C., Hong, M., Kim, D. H., Jeong, K. W., Eguchi, M., Moon, R. T., Teo, J.-L., Oh, S. W., Kim, H. Y., Moon, S. H., Ha, J. R., Kahn, M. *Proc. Natl. Acad. Sci. USA* **2004**, *34*, 12682-12687.

(18) Best, J. L., Amezcua, C. A., Mayr, B., Flechner, L., Murawsky, C. M., Emerson, B., Zor, T., Gardner, K. H., Montminy, M. *Proc. Natl. Acad. Sci. USA* **2004**, *51*, 17622-17627.

(19) Sachchidanand, Resnick-Silverman, L., Yan, S. Mutjaba, S., Liu, W., Zeng, L., Manfredi, J. J., Zhou, M.-M. *Chem. Biol.* **2006**, *13*, 81-90.

(20) Xiao, X., Yu, P., Lim, H.-S., Sikder, D., Kodadek, T. *Angew. Chem. Int. Ed.* **2007**, *46*, 2865-2868.

(21) Olenyuk, B.Z., Zhang, G.J., Klco, J.M., Nickols, N.G., Kaelin, W.G., and Dervan, P.B. *Proc. Natl. Acad. Sci. USA* **2004**, *48*, 16768-16773.

(22) Kong, D., Park, E. J., Stephen, A. G., Calvani, M., Cardellina, J. H., Monks, A., Fischer, R. J., Shoemaker, R. H., Melillo, G. *Cancer Res.* **2005**, *65*, 9047-9055.

(23) Mapp, A. K., Ansari, A. Z., Ptashne, M., Dervan, P. B. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 3930-3935.

(24) Ansari, A. Z., Mapp, A. K., Nguyen, D. H., Dervan, P. B., Ptashne, M. *Chem. Biol.* **2001**, *8*, 583-592.

(25) Arora, P. S., Ansari, A. Z., Best, T. P., Ptashne, M., Dervan, P. B. *J. Am. Chem. Soc.* **2002**, *124*, 13067-13071.

(26) Kwon, Y., Arndt, H.-D., Mao, Q., Choi, Y., Kawazoe, Y., Dervan, P. B., Uesugi, M. *J. Am. Chem. Soc.* **2004**, *126*, 15940-15941.

(27) Eguchi, M., Lee, M. S., Nakanishi, H., Stasiak, M., Lovell, S., Kahn, M. *J. Am. Chem. Soc.* **1999**, *121*, 12204-12205.

(28) Eguchi, M., Lee, M. S., Stasiak, M., Kahn, M. *Tet. Lett.* **2001**, *42*, 1237-1239.

(29) Eguchi, M., Shen, R. Y. W., Shea, J. P., Lee, M. S., Kahn, M. *J. Med. Chem.* **2002**, *45*, 1395-1398.

(30) Brochu, M. Masters Thesis, California Institute of Technology, **2007**.

(31) Baird, E. E., Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141-6146.

(32) Belitsky, J. M., Nguyen, D. H., Wurtz, N. R. *Bioorg. Med. Chem.* **2002**, *10*, 2767-2774.

(33) Herman, D. M., Baird, E. E., Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 1382-1391.

(34) Nickols, N. G., Jacobs, C. S., Farkas, M. E., Dervan, P. B. *Nucleic Acids Res.* **2007**, *35*, 363-370.

(35) Nickols, N. G. Ph. D. Thesis, California Institute of Technology, **2008**.

Appendix A

Protein-DNA Dimerizers Targeted to a Natural HOX Site

The part of this chapter regarding the DNase I footprinting was taken in part from a manuscript co-authored with Rocco Moretti,[‡] Leslie J. Donato,[‡] Mary L. Brezinski,[‡] Helena Hoff,[‡] Jon S. Thorson,[‡] Peter B. Dervan,[†] and Aseem Z. Ansari.[‡] (Caltech[†] and University of Wisconsin, Madison[‡])

(Moretti, R., Donato, L. J., Brezinski, M. L., Stafford, R. L., Hoff, H., Thorson, J. S., Dervan, P. B., Ansari, A. Z. *provisionally titled* Chemical wedges reveal the contribution of allosteric DNA modulation in protein-DNA assembly, **2007**, *In preparation*.)

*Work on the MPE footprinting of polyamide-peptide conjugates **1-5** was done in collaboration with Hans-Dieter Arndt (Dervan Group; Caltech).*

A.1 Targeting an Endogenous HOX Binding Site

All previous protein-DNA dimerizers for Exd do not target natural HOX binding sites.[1-5] Compound **1** was designed to target an endogenous labial response element 5′-TGATGGATTG-3′ (Figure A.1).[6] At this composite site, Exd is expected to bind 5′-TGAT-3′ and the polyamide DNA-binding domain of **1** was designed to target the site 5′-WGGWWW-3′[7] and project the peptide to the Exd binding pocket in a similar manner to conjugates **2** and **3** which are targeted to the artificial HOX site 5′-TGATTGACCT-3′. A control compound for **1** was also synthesized in which the crucial WM amino acids were replaced with alanines (**4**). Additionally, the parent polyamide that lacks a peptide altogether serves as a negative control (**5**).

DNase I footprinting of compounds **1**, **4**, and **5** was performed to determine their DNA binding affinities and specificities for their match site (Figure A.2 and Table A.1). Footprinting of compounds **2** and **3** have been reported elsewhere.[1,4] As previously observed with conjugate **2**, the labial FYPWMKG compound **1** showed only a modest affinity for its match site 5′-TGGATT-3′ ($K_a = 2.4 \times 10^8$ M^{-1}) with no significant sequence specificity. A WM to AA "mutation" of compound **2** was shown previously to lead to a recovery of the DNA binding affinity and specificity,[1] but a similar WM to AA change to the labial compound (i.e., **1** to **4**) did not lead to any significant change in binding affinity or specificity. The polyamide **5** which contains only a propylamine appendage showed an approximately 20-fold higher DNA-binding affinity and >10-fold specificity over single base pair mismatches. The parent polyamide ImImPyPy-γ-PyPyPyPy-β-Dp has previously been reported to bind to DNA with an affinity of 5×10^8 M^{-1}, but it is hard to compare this to **5** since it was footprinted on the match sequence 5′-TGGTTA-3′.[8]

In spite of the modest DNA affinity and lack of specificity of the labial conjugate **1**, MPE footprinting experiments indicate that **1** cooperatively binds with Exd specifically at the labial binding site (Figure A.3). That is, the titration of Exd in the presence of compound **1** leads to an enlargement of the observed protection pattern exclusively at the

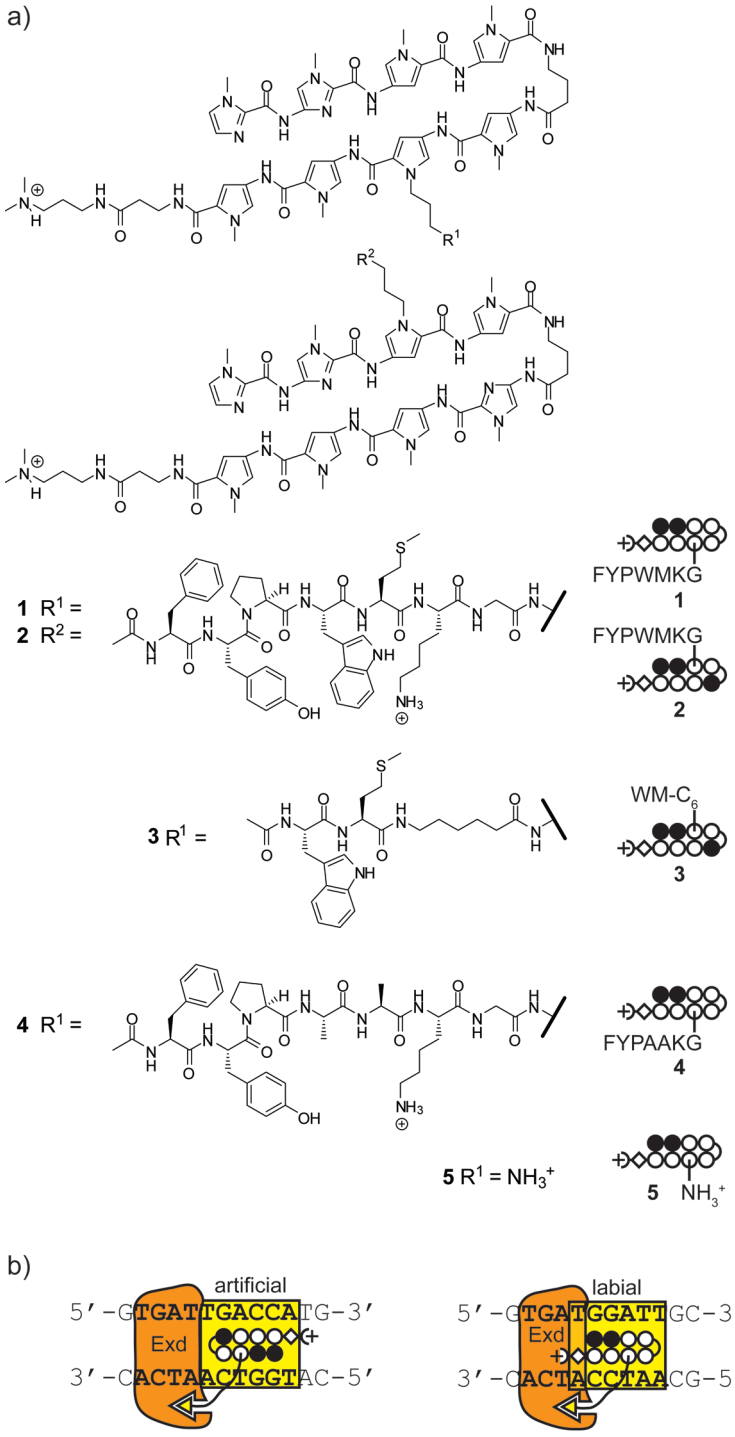


Figure A.1 Structures of polyamide-peptide conjugates targeted to artificial and natural HOX binding sites. (a) Conjugate **1** targets the natural labial HOX binding site (5′-TGA**TGGATT**G-3′) and **2** and **3** target the artificial site (5′-TGAT**TGACCT**-3′) where the Exd binding site is underlined and the polyamide binding site is shown in bold. Compounds **4** and **5** serve as negative controls. (b) Schematics showing the binding orientations of the artificial and natural cognate DNA sites.

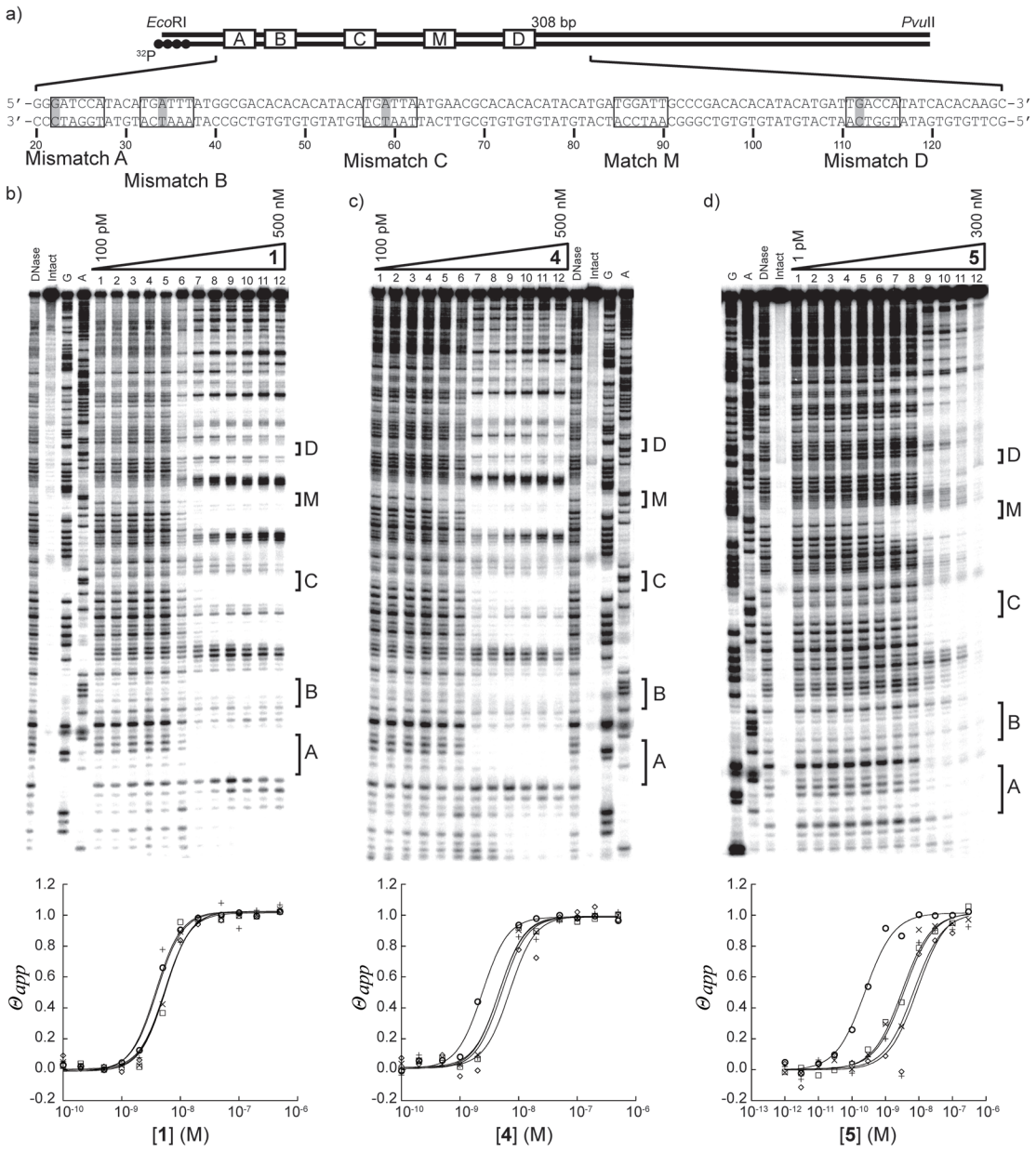


Figure A.2 Quantitative DNase footprinting titrations of compounds **1**, **4**, and **5** on a restriction fragment containing a match site M (5′-TGGATT-3′) and four single base pair mismatch sites A-D, (5′-TGGAT**C**-3′), (5′-TG**A**TTT-3′), (5′-TG**A**TTA-3′), and (5′-TGGT**C**A-3′), respectively. a) The DNA sequence of the restriction fragment with the location of the match and mismatch sites is shown. b) Titration of compound **1**. Lane 1-12: 0.1, 0.2, 0.5, 1, 2, 5, 10, 20, 50, 100, 200, 500 nM. c) Titration of compound **4** labeled as in panel b. d) Titration of compound **5**. Lane 1-12: 0.001, 0.003, 0.01, 0.03, 0.1, 0.3, 1, 3, 10, 30, 100, 300 nM. DNase control, intact DNA, and G- and A-sequencing lanes are indicated for each gel. Isotherms are shown below each gel for match site M (○) and mismatch sites A (□), B (◊), C (×), and D (+).

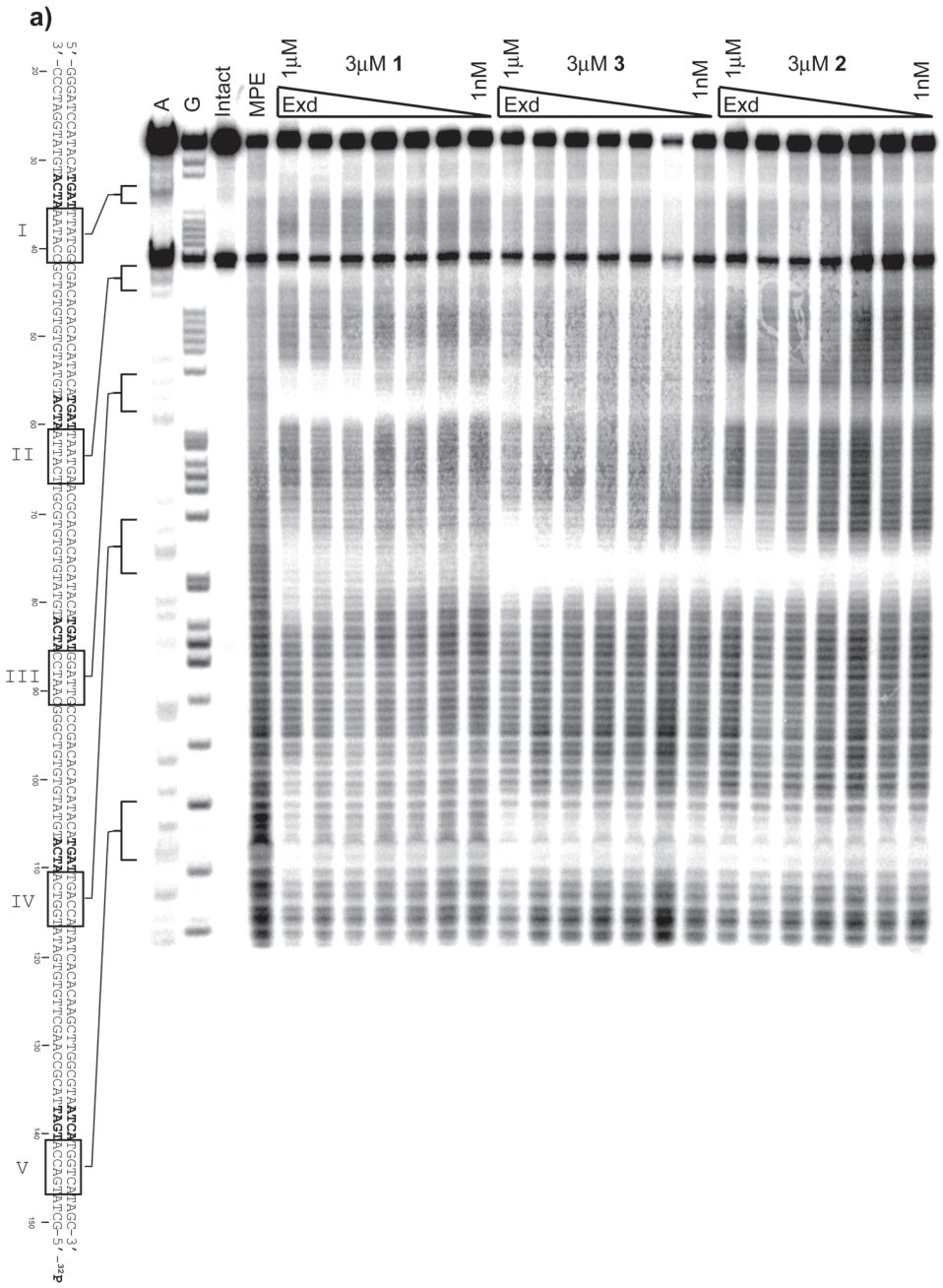


Figure A.3 Site-selective recruitment of Exd to the natural labial binding site. MPE footprinting experiments show an increase in binding site size exclusively at the labial cognate site III upon titration of Exd in the presence of conjugate **1**. In contrast, conjugates **2** and **3** which target the artificial site IV show an expansion in the binding site size exclusively at the artificial site IV.

labial match site III. Compounds **2** and **3**, which target an artificial HOX binding site, lead to increased protection exclusively at their match site IV. The control compounds **4** and **5** did not lead to increased protection at any of these sites when Exd was titrated (data not shown).

A.2 Design of a Cell-Permeable Labial Mimic

During the process of designing of a cell-permeable labial mimic, the putatively negative C-terminal β-alanine cell uptake determinant was re-evaluated. Specifically, polyamide **6** which contains a C-terminal β-alanine was synthesized by cleaving the corresponding polyamide from resin using 1,3-diaminopropane followed by coupling to fluorescein through an amide bond (FAM). Polyamide-FAM conjugate **6** was shown to yield nuclear localization in HeLa, MCF-7, PC3 cells (Figure A.4). The solubility of **6** was observed to be poor in unbuffered water likely due to the lack of a positive charge in the C-terminus. At pH 7 in phosphate buffer saline the compound was soluble, presumably due to the complete deprotonation of the carboxylic acid of the fluorescein leading to a net negative charge. One potentially useful feature of compound **6** is that it can be synthesized using pre-loaded BOC-β-Ala-PAM[9] instead of oxime resin.[10]

The parent polyamide **7** ImImPyPy-γ-PyPyPyPy-β-C_3-BOC was cleaved from PAM using mono Boc-protected 1,3-diaminopropane (Figure A.5). Following procedures previously described in Chapter 4, the desired conjugate **8** was made by coupling a peptide to the polyamide, deprotecting with TFA, and attaching FAM. Conjugate **8** showed nuclear localization in HeLa, MCF-7, and PC3 cells (Figure A.6). These compounds might be useful for performing cell culture or fruit fly experiments.

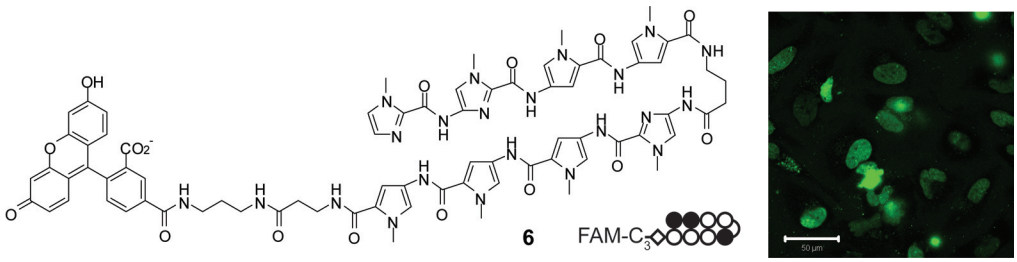


Figure A.4 Uptake results in HeLa cells for polyamide-FAM conjugate with a β-1,3-diaminopropane tail which shows primarily nuclear staining. (Scale bar = 50 µm)

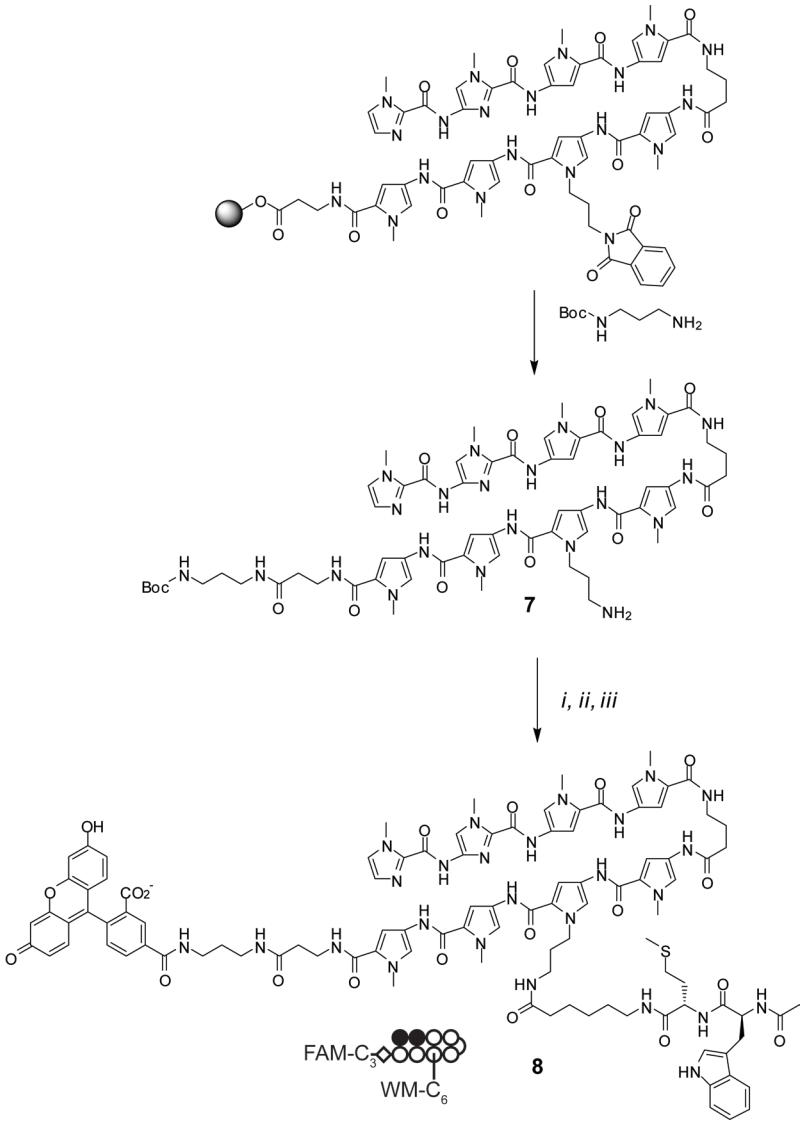
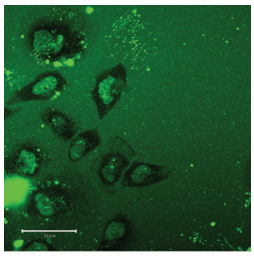


Figure A.5 Synthesis of polyamide-peptide-fluorescein conjugates targeted to the natural labial binding site: *i*) peptide, HBTU, 0.1 M DIEA in DMF, *ii*) 50% TFA/DCM, and *iii*) 5-carboxyfluorescein succinimidyl ester.



FAM-C_3

WM-C_6

8

Figure A.6 Uptake results in HeLa cells for polyamide-peptide-FAM conjugate **8** shows significant nuclear localization.

Table A.1 DNA equilibrium association constants K_a's [$\times 10^8$ M^{-1}]

Compound	5'-TGGATT-3'	5'-TGGAT**C**-3'	5'-TG**A**TTT-3'	5'-TG**A**TTA-3'	5'-TGGT**C**A-3'
1[a]	2.4 ± 0.9	1.7 ± 0.3	1.5 ± 0.3	1.7 ± 0.3	3 ± 2
4[a]	3 ± 1	1.5 ± 0.5	1.0 ± 0.4	1.5 ± 0.4	1.5 ± 0.4
5[b]	70 ± 50	6 ± 4	0.6 ± 0.3	6 ± 4	4 ± 2

[a] Data fit best using a Hill coefficient of 2.

[b] Data fit best using a Hill coefficient of 1.

A.3 Experimental Details

Synthesis

ImImPyPy-γ-PyPy$^{(Phe\text{-}Tyr\text{-}Pro\text{-}Trp\text{-}Met\text{-}Lys\text{-}Gly\text{-}propylamine)}$PyPy-β-Dp (1). This polyamide-peptide conjugate was synthesized essentially as previously described.[1] Briefly, the peptide Phe-Tyr(OtBu)-Pro-Trp(Boc)-Met-Lys(Boc)-Gly-OH (9.6 mg, 7.8 μmol) was activated with HBTU (7.5 μmol) and added to the parent polyamide **5** (2.76 μmol) in 1 mL DMF and excess DIEA. The solvent was evaporated under reduced pressure and deprotection was accomplished with 80% TFA, 5% EDT, 5% TES, and 10% DCM for 20 min at RT. The solution was added to 10 mL of ice-cold ether from which a precipitate formed and collected by centrifugation. The pellet was washed 2x with ether and dried briefly with $N_{2(g)}$. The remaining Trp carboxylate was eliminated by dissolving the pellet in 0.2 M acetic acid and incubating overnight. The final compound was purified by preparative HPLC and lyophilized to give 330 nmol of a white powder (~12% yield). (MALDI-TOF MS) [M +

H]$^+$ calcd for $C_{109}H_{138}N_{31}O_{19}S^+$ 2217.0, observed 2216.8.

ImImPy$^{(Phe\text{-}Tyr\text{-}Pro\text{-}Trp\text{-}Met\text{-}Lys\text{-}Gly\text{-}propylamine)}$Py-γ-ImPyPyPy-β-Dp (2). This polyamide-peptide conjugate was synthesized and characterized as previously described.[1]

ImImPy$^{(Trp\text{-}Met\text{-}\varepsilon Ahx\text{-}propylamine)}$Py-γ-ImPyPyPy-β-Dp (3). The polyamide-peptide conjugate was synthesized and characterized as described previously in Chapter 2.[4]

ImImPyPy-γ-PyPy$^{(Phe\text{-}Tyr\text{-}Pro\text{-}Ala\text{-}Ala\text{-}Lys\text{-}Gly\text{-}propylamine)}$PyPy-β-Dp (4). This polyamide-peptide conjugate was synthesized essentially as previously described.[1] Briefly, the peptide Phe-Tyr(OtBu)-Pro-Ala-Ala-Lys(Boc)-Gly-OH (8.0 mg, 8.4 μmol) was activated with HBTU (7.5 μmol) and added to the parent polyamide **5** (2.76 μmol) in 1 mL DMF and excess DIEA. The compound was deprotected as described for compound **1** purified by preparative HPLC and lyophilized to give 440 nmol of a white powder (~16% yield). (MALDI-TOF MS) [M + H]$^+$ calcd for $C_{99}H_{129}N_{30}O_{19}^+$ 2042.0, observed 2042.0.

ImImPyPy-γ-PyPy$^{(propylamine)}$PyPy-β-Dp (5). This polyamide was synthesized on PAM resin using standard protocols cleaving with 2 mL of Dp.[9] Total yield of 48.0 μmol of an off-white powder. (MALDI-TOF MS) [M + H]$^+$ calcd for $C_{60}H_{77}N_{22}O_{10}^+$ 1265.6, observed 1265.6.

ImImPyPy-γ-ImPyPyPy-β-C_3-FAM (6). This conjugate was synthesized by coupling 5-carboxyfluorescein succinimidyl ester (1 μmol) to the polyamide ImImPyPy-γ-ImPyPyPy-β-C_3-NH_2 (910 nmol) in 200 μL of 0.1 M DIEA in DMF for 20-30 min at RT. The final product was purified by preparative HPLC and lyophilized to give 270 nmol of a yellow-orange solid (~30% yield). (MALDI-TOF MS) [M + H]$^+$ calcd for $C_{76}H_{77}N_{22}O_{16}^+$ 1553.6, observed 1553.6.

ImImPyPy-γ-PyPy$^{(propylamine)}$PyPy-β-C_3-BOC (7). This polyamide was synthesized on PAM resin using standard protocols[9] cleaving with ~1 mL of *N*-Boc-1,3-diaminopropane. Total yield of 1.9 μmol of an off-white powder (~2% yield). (MALDI-TOF MS) [M + H]$^+$ calcd for $C_{63}H_{81}N_{22}O_{12}^+$ 1337.6, observed 1337.7.

ImImPyPy-γ-PyPy$^{(Trp\text{-}Met\text{-}\varepsilon Ahx\text{-}propylamine)}$PyPy-β-$C_3$-FAM (8). The conjugate was synthesized

in a similar manner to conjugate **1** using the peptide Ac-Trp(Boc)-Met-εAhx-OH (2.4 mg) activated with PyBOP (7.5 mg) coupled to polyamide **7** (1.5 μmol). The resulting intermediate was then coupled to 5-carboxyfluorescein succinimidyl ester (3.0 μmol). The final product was purified by preparative HPLC and lyophilized to yield 536 nmol of a yellow-orange solid (~36% yield). (ESI MS) $[M + 2H]^{2+}$ calcd for $C_{103}H_{116}N_{26}O_{20}^{2+}$ 1034.4, observed 1034.9.

***N*-Boc-1,3-diaminopropane.** Cool 14.8 g (16.7 mL, 0.2 mol) of 1,3-diaminopropane on an ice bath and add a solution of di-*tert*-butyl dicarbonate in a drop-wise manner (4.7 g, 0.021 mol in 50 mL DCM). Warm to room temperature over 1.5 hours and extract with 200 mL of 1/4 sat. $NaHCO_3$. Back extract aqueous layer 4x with 50 mL DCM. Wash combined organic layers with 50 mL saturated $NaHCO_3$ and 50 mL brine. Dry organic layer with anhydrous sodium sulfate, filter, and evaporate under reduced pressure with a small amount of toulene to azeotrope with residual water. ^{1}H NMR (DMSO, 300 MHz) δ 7.8 (br t, 1 H), δ 2.94 (q, 2 H), δ 2.49 (t, 2 H), δ 1.40 (p, 2 H), δ 1.36 (s, 9 H), 1.10 (s, 2 H); HRMS (FAB) exact mass calcd for $C_8H_{19}N_2O_2^+$ requires *m/z* 175.1447, found 175.1452.

Quantitative DNase I footprinting. Reactions were carried out in a volume of 400 μL in aqueous TKMC buffer according to published protocols[11] using a 3' ^{32}P-labeled 250 base pair *Eco*RI/*Pvu*II restriction fragment of the plasmid pHDA1. Developed gels were imaged using storage phosphor autoradiography using a Molecular Dynamics 400S Phosphorimager. Equilibrium association constants were determined as previously described.[11]

MPE footprinting with Exd. MPE footprinting was carried out in 40 μL reaction volumes according to published procedures.[11] Reactions were carried out as described in Chapter 2 in a total volume of 40 μL in aqueous TKMC buffer containing 10% glycerol using the 3' ^{32}P-labeled 308 base pair *Eco*RI/*Pvu*II restriction fragment of the plasmid pHDA1.

Cell culture and confocal microscopy experiments. All cell culture and confocal microscopy was performed essentially as described previously.[12] See Chapter 4 for a detailed explanation.

A.4 References

(1) Arndt, H.-D., Hauschild, K. E., Sullivan, D. P., Lake, K., Dervan, P. B., Ansari, A. Z. *J. Am. Chem. Soc.* **2003**, *125*, 13322-13323.

(2) Hauschild, K. E., Metzler, R. E., Arndt, H.-D., Moretti, R., Raffaelle, R., Dervan, P. B., Ansari, A. Z. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 5008-5013.

(3) Warren, C. L., Kratochvil, N. C. S., Hauschild, K. E., Foister, S., Brezinski, M. L., Dervan, P. B., Phillips, G. N. Jr., Ansari, A. Z. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 867-872.

(4) Stafford, R. L., Arndt, H.-D., Brezinski, M. L., Ansari, A. Z., Dervan, P. B. *J. Am. Chem. Soc.* **2007**, *129*, 2660-2668.

(5) Stafford, R. S. and Dervan, P. B. *submitted to J. Am. Chem. Soc.* **2007**.

(6) Ryoo, H. D., Marty, T., Casares, F., Affolter, M., Mann, R. S. *Development* **1999**, *126*, 5137-5148.

(7) Dervan, P. B. and Edelson, B. S. *Curr. Opin. Struct. Biol.* **2003**, *13*, 284-299.

(8) Neely, L., Trauger, J.W., Baird, E.E., Dervan, P.B., Gottesfeld, J.M. **1997** *J. Mol. Biol.* **274**, 439-445.

(9) Baird, E. E. and Dervan, P. B. *J. Am. Chem. Soc.* **1996,** *118*, 6141-6146.

(10) Belitsky, J. M., Nguyen, D. H., Wurtz, N. R. Dervan, P. B. *Bioorg. Med. Chem.* **2002**, *10*, 2767-2774.

(11) Trauger, J. W. and Dervan, P. B. *Methods Enzymol.* **2001**, *340*, 450-466.

(12) Edelson, B. S., Best, T. P., Olenyuk, B., Nickols, N. G., Doss, R. M., Foister, S., Heckel, A., Dervan, P. B. *Nucleic Acids Res.* **2004**, *32*, 2802-2818.

Appendix B

Toward a Small-Molecule YPWM Mimic

*Compounds **20-26** were synthesized by John Phillips (Dervan Group; Caltech).*

B.1 Background

Natural HOX proteins cooperatively bind to DNA with cofactor TALE proteins (Exd or Pbx) via a YPWM peptide motif.[1-3] The development of small molecules that competitively inhibit the YPWM-TALE protein-protein interaction could potentially be used as therapeutics for the treatment of pre-B cell ALL involving the Pbx-E2A oncoprotein.[4-9] A small molecule was discovered from a library of 1,4-disubstituted naphthalenes that binds to Pbx1 (IC_{50} = 65 μM) and blocks Pbx/HOX/DNA complex formation, but its effect on Pbx-E2A-mediated gene expression is unknown.[10] Such small molecules may also serve as binding domains for protein-DNA dimerizers which mimic HOX proteins.[11-15] The protein-binding domains for such protein-DNA dimerizers have been peptides which may be susceptible to endogenous proteases in cell-culture and fruit fly experiments. This appendix describes three non-peptidic mimics of the YPWM motif in the context of Exd protein-DNA dimerizers including the aforementioned 1,4-disubstituted naphthalene in search of a metabolically stable Exd protein-binding domain.

B.2 YPWM β-turn Mimetic

The YPWM motif has been highly conserved throughout evolution and is often referred to as a penta- or hexapeptide motif with an overall consensus sequence of ϕYPWM(K/R) where ϕ is a hydrophobic residue.[16] Residues flanking the hexapeptide have also been shown to contribute to YPWM-protein interactions.[17] The small buried YPWM-Exd interface (~570 Å^2)[1] when compared to mean protein-protein interfaces (~2000 Å^2)[18-19] suggested that this motif was already minimal. Thus, the initial attempt to develop a small-molecule YPWM mimetic aimed at maintaining the entire peptide structure. It was noted that the Exd/Ubx/DNA[1] and Pbx1/HoxB1/DNA[2] structures the YPWM motif adopts either a type I β-turn or the closely related type III β-turn (i.e., 3_{10} helix). Thus, a β-turn mimetic was designed based on a scaffold reported by the Burgess group.[20-23] This scaffold allowed incorporation of the absolutely conserved Trp at the *i+2* position and energy minimized

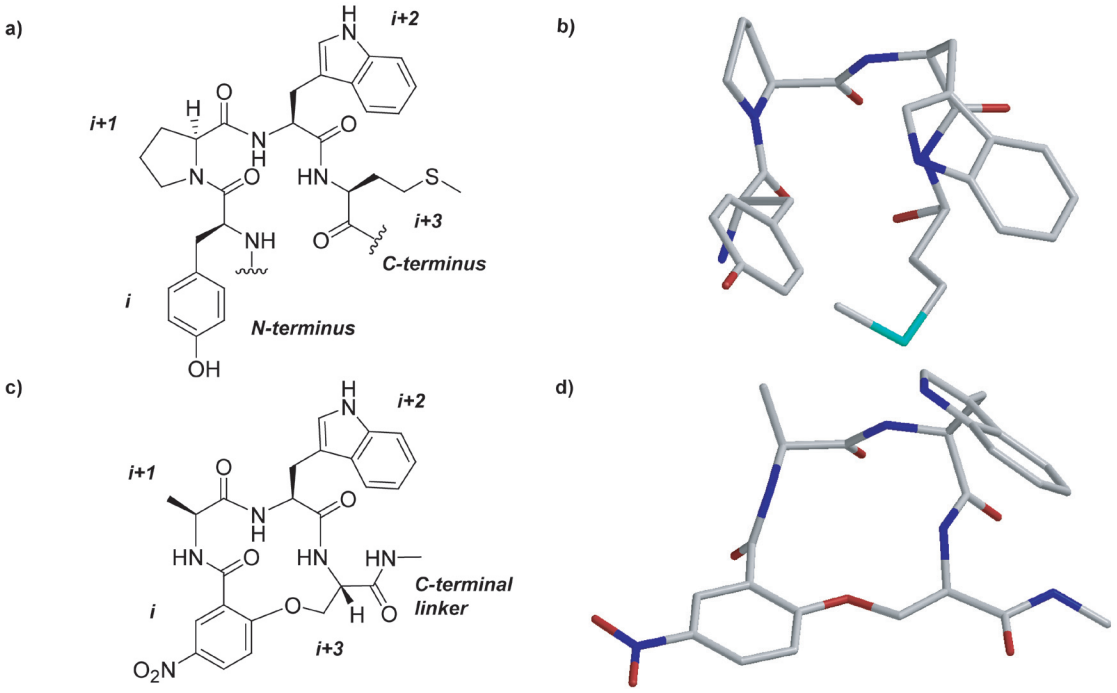


Figure B.1 Design of YPWM β-turn mimetic. a) Wild-type Ubx YPWM tetrapeptide depicted schematically in a β-turn conformation. b) Wild-type Ubx YPWM tetrapeptide from the Ubx/Exd/DNA crystal structure (PDB code 1B8I). c) Designed YPWM mimic using a β-turn scaffold from the Burgess group.[20-23] d) Energy minimized model (PM3) of the designed YPWM β-turn mimic.

models (PM3) suggested reasonable structural similarity to the native YPWM motif (Figure B.1).

The synthesis of YPWM β-turn mimetics was performed on a low-loading TentaGel PHB resin using procedures adapted from related macrocycles (Figure B.2).[20-21] Briefly, standard Fmoc peptide-synthesis was employed to yield resin-bound intermediate **1**. Fmoc deprotection, coupling to 2-fluoro-5-nitro-benzoic acid, and selective removal of the trityl with mild acid, yielded resin-bound intermediate **2**. Macrocycle formation was accomplished under basic conditions which promoted nucleophilic aromatic substitution of the serine hydroxyl with the fluorinated benzoic acid derivative. Treatment of the resin with strong acid and HPLC purification of the resulting crude mixture yielded macrocycle **3**. Low resin loading was found to prevent the formation of undesired dimerized by-products and was important for achieving good yields. Coupling of the macrocycle carboxylic acid

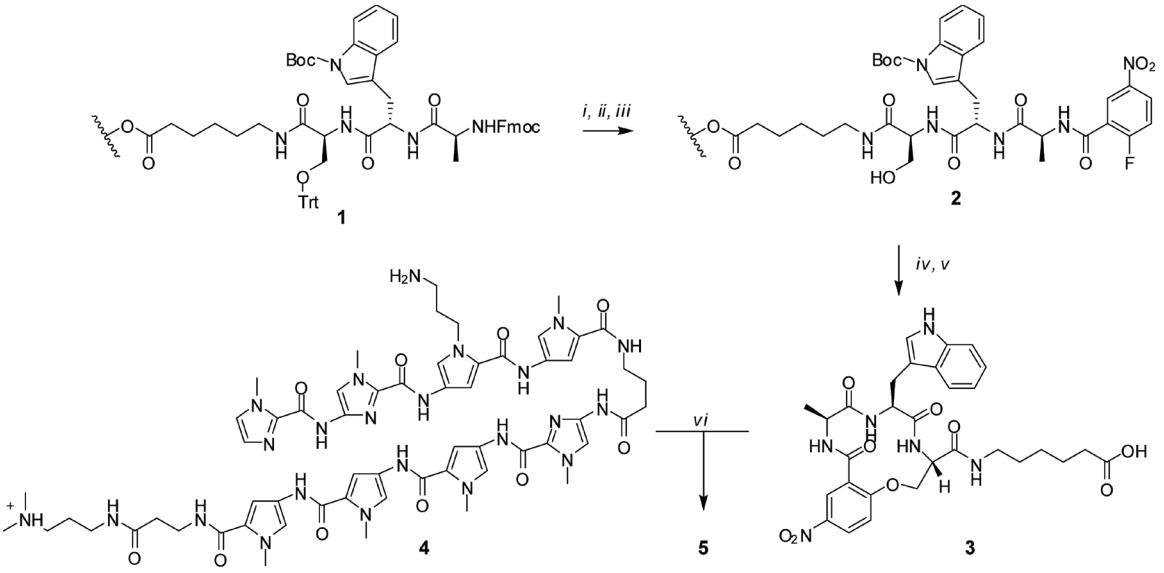


Figure B.2 Synthesis of polyamide-YPWM β-turn mimic conjugate: *i*) 20% piperidine in DMF, *ii*) 2-fluoro-5-nitrobenzoic acid, HATU, DIEA, DMF, *iii*) 1-3% TFA, 5% TIS, in DCM, *iv*) K_2CO_3 in DMF, *v*) 85% TFA, 5% TIS, 10% DCM, *vi*) HBTU, DIEA in DMF

3 to polyamide amine **4** via *in situ* HBTU activation yielded conjugate **5** (Figure B.3a). An uncyclized analogue **6** was synthesized using standard Fmoc chemistry and coupled to **4** to yield conjugate **7** to serve as a control (Figure B.3b).

An electrophoretic mobility shift assay (EMSA), or gel shift, with Exd at 4 °C was employed with conjugate **5** and **7** to assess their ability to facilitate Exd-DNA dimerization (Figure B.3c). Although the conjugate **5** enhanced Exd's DNA binding substantially, the uncyclized control compound exhibited a slightly higher Exd-DNA dimerization activity. This result suggested that the presence of a macrocycle was unnecessary for the Exd binding activity. This directly inspired the experiments that showed only the Trp and Met amino acids were necessary for Exd-DNA dimerization in the context of a polyamide-peptide conjugate.[14]

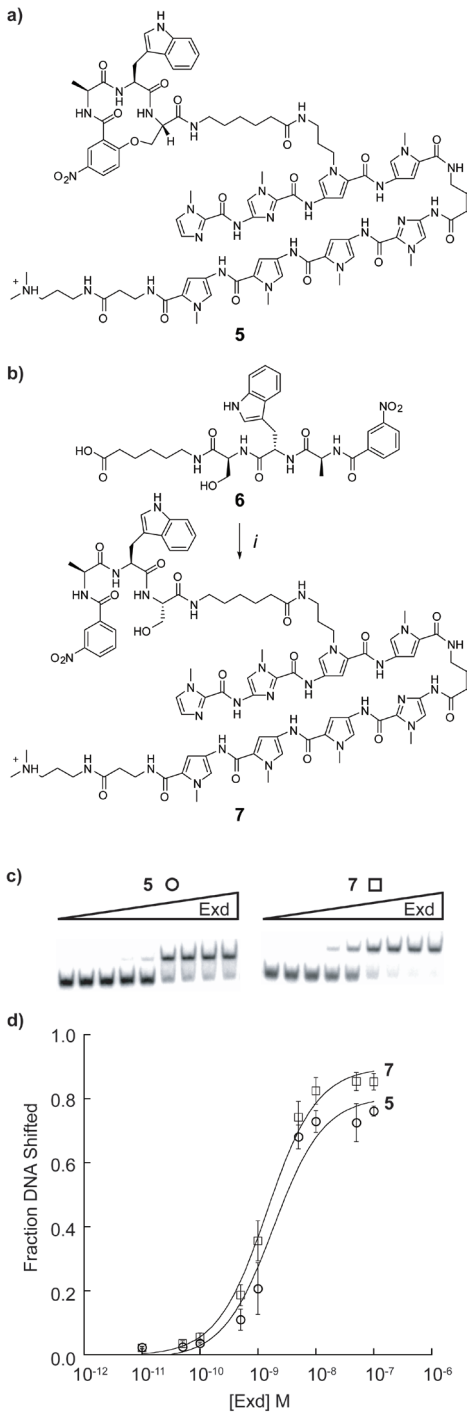


Figure B.3 Gel shifts experiments with polyamide-YPWM β-turn mimic conjugates. a) Structure of polyamide-YPWM β-turn mimic conjugate. b) Synthesis of uncyclized β-turn mimic control conjugate: *i*) **4**, HBTU, DIEA in DMF. c) Representative gel shift experiments for cyclized (**5**) and uncyclized (**7**) β-turn mimic conjugates. d) Isotherms for Exd binding in the presence of **5** and **7**. The uncyclized analogue (**7**) leads to a higher affinity complex than the cyclized compound (**5**). Gel shifts were performed at 50 nM compound titrating Exd from 10 pM to 100 nM.

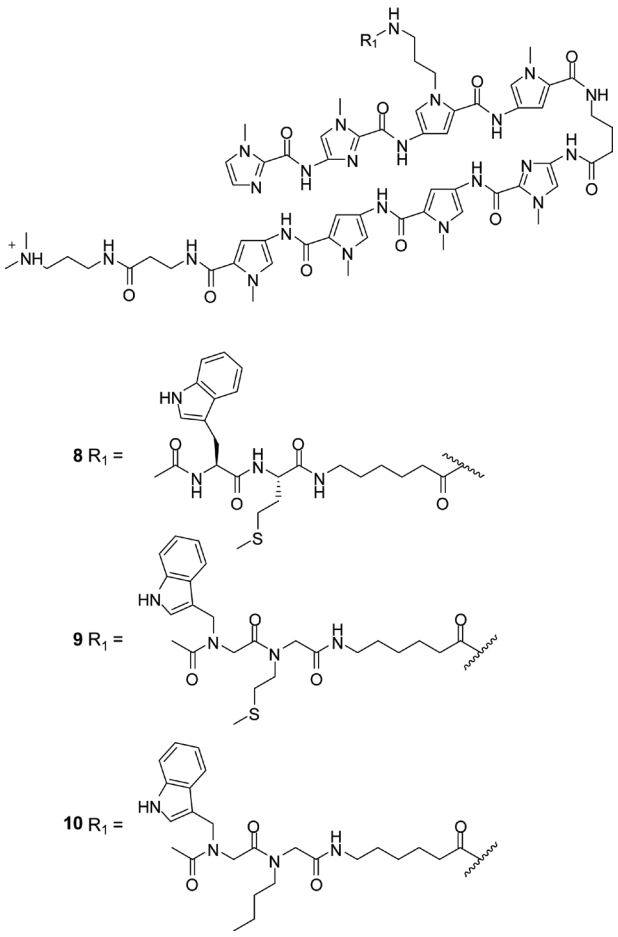


Figure B.4 Design of WM peptoid conjugates. The natural WM *peptide* conjugate **8** is shown for comparison to the WM *peptoid* conjugate **9**. A WM peptoid analogue **10** with a completely alkyl derivative of the methionine side chain is also shown.

B.3 WM Peptoid Analogue

One type of synthetic peptide analogue is a poly-*N*-substituted glycine, or peptoid.[24-26] Given that only Trp and Met residues were necessary to dimerize Exd,[14] efforts were undertaken to generate a WM peptoid (Figure B.4). Synthesis of the direct analogue of the WM peptide was attempted using the submonomer method[25-26] with 2-chlorotrityl resin, which was reported to enable the synthesis of peptoid carboxylates (Figure B.5).[27] Only the isolation of a compound consistent with **15** was observed, which was rationalized in terms of an intramolecular nucleophilic bromine displacement by the methionine thioether to give intermediate **13**, which probably resulted in the formation of the more stable diketopiperizine **14**. The alkyl analogue **17** was synthesized and coupled to the polyamide ImImPyPy-γ-ImPyPyPy-β-Dp to yield conjugate **10**. Gel shift experiments at 20 °C with **10** showed that it did not form as stable of an Exd-DNA complex as its peptide relative **8** (Figure B.6).

B.4 Naphthalene Small Molecule Inhibitor

The 1,4-disubstituted naphthalene **18** was previously reported to competitively

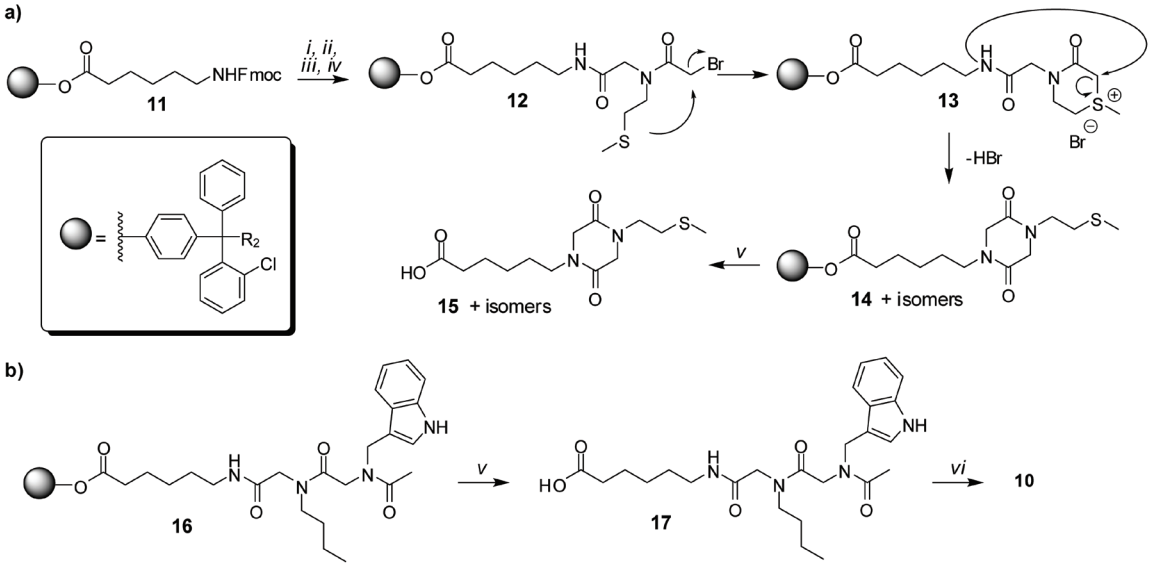


Figure B.5 Synthesis of WM peptoid analogues. a) The synthesis of WM peptoid conjugate was ultimately not successful due to the probable intramolecular bromine displacement by the methionine as shown: *i*) 20% piperidine in DMF, *ii*) α-bromoacetic acid, DIC, DIEA, DMF, *iii*) 2-(methylthio)ethylamine, DMSO, *iv*) α-bromoacetic acid, DIC, DIEA, DMF, *v*) 2:2:6 AcOH:TFE:DCM b) Standard submonomer peptoid synthesis using 2-chlorotrityl resin yielded the desired carboxylic acid **17**: *v*) 2:2:6 AcOH:TFE:DCM, *vi*) **4**, HBTU, DIEA in DMF.

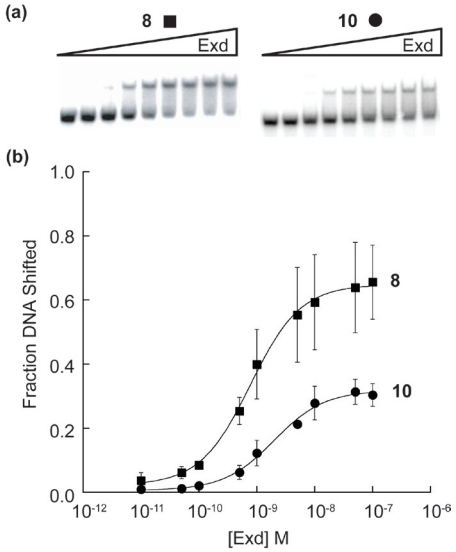
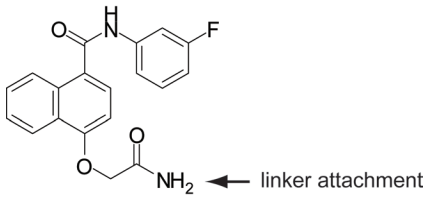


Figure B.6 Gel shifts experiments comparing polyamide-WM *peptoid* and *peptide* conjugates. a) The peptoid conjugate **10** yielded a lower affinity Exd-DNA complex than the peptide conjugate **8**. b) Isotherms for Exd bind in the presence of **8** and **10**. Gel shifts were performed with 50 nM compound titrating Exd from 10 pM to 100 nM.

inhibit Pbx/HOX/DNA complex formation presumably by competitively displacing the YPWM peptide (Figure B.7).[10] Although no rigorous structural data is available for the interaction of **18** with Pbx or Exd, published modeling studies[10] suggested the formamide moiety was a suitable attachment point for a polyamide conjugate since it was predicted to project away from the binding pocket. Accordingly, John Phillips synthesized a suitably modified version (**26**) to enable direct coupling to a polyamide. Starting from commercially available naphthyl aldehyde **19**, carboxylic acid **20** was obtained by oxidation. Coupling **20** to 3-fluoroaniline yielded amide **21** and the naphthylmethoxy was converted to aromatic hydroxyl **22** by nucleophilic displacement. The hydroxyl was condensed with α-bromomethylacetate to give methyl ester **23** which was saponified to yield carboxylic acid **24**. A four-carbon linker was then attached by amide bond formation to yield **25**, which was subsequently hydrolyzed to give the key carboxylic acid intermediate **26**. The final conjugate **27** was made via amide bond formation between a polyamide primary amine and the carboxylic acid of **26**.



18

Figure B.7 Structure of naphthalene compound **18** previously reported to competitively inhibit Hox/Pbx1/DNA complex formation by binding to the YPWM binding pocket of Pbx1.[10]

A gel shift experiment at 20 °C with compound **27** did not yield nearly as strong of an Exd-DNA complex as the polyamide-WM conjugate **8** (Figure B.9). The levels of complex formation and Exd affinity were only slightly better than were caused by the allosteric effects of the parent polyamide ImImPyPy-γ-ImPyPyPy-β-Dp at the same temperature.[14] It is possible that modifying the length or composition of the linker or its attachment point may increase the amount of complex formed.

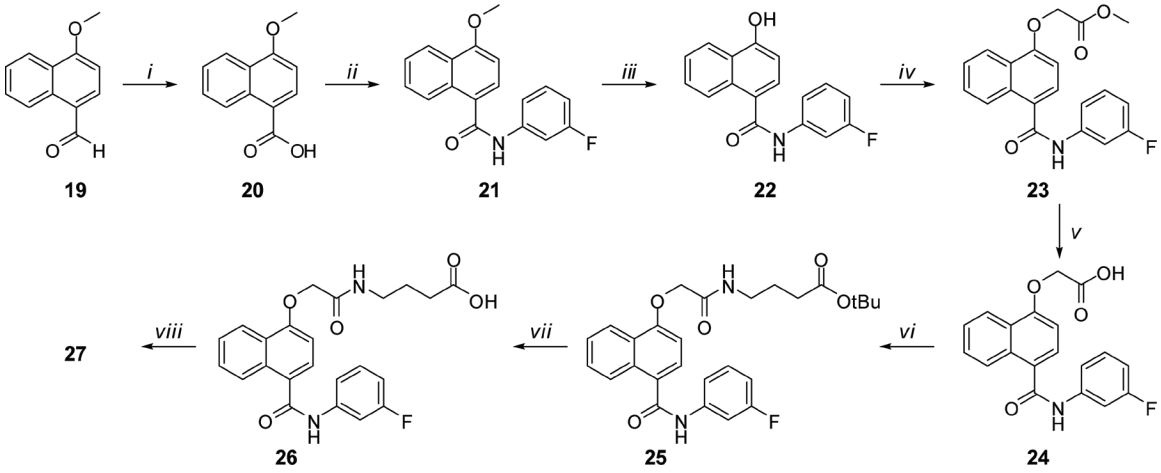


Figure B.8 Synthesis of an analogue of **18** suitable for conjugating to a polyamide: *i*) $NaClO_2$, NaH_2PO_4, H_2O, tBuOH, 2-methyl-2-butene, rt, 24 hours, *ii*) oxalyl chloride, 3-fluoroaniline, DIEA, THF, reflux, *iii*) ethanethiol, $AlCl_3$, DCM, 0 to 20 °C, *iv*) methyl bromoacetate, K_2CO_3, THF, reflux, 1 hour, *v*) LiOH, MeOH, H_2O, reflux, 2 hours, *vi*) γ-aminobutyric acid *tert*-butyl ester, PyBOP, DIEA, DMF, *vii*) 1:2 TFA:DCM, *viii*) **4**, HBTU, DIEA, DMF.

B.5 Experimental Details

Macrocycle (3). The synthesis of macrocycles was accomplished using a modified version of previously published methods.[20] Synthesis was performed using TentaGel S PHB resin (Fluka, loading ~0.24 meq/g). Standard Fmoc SPPS was employed initially to obtain peptide intermediate **1** which was deprotected with 20% piperidine in DMF for 20 min. Coupling of 2-fluoro-5-nitrobenzoic acid (2.5 equivalents) was performed by *in situ* activation of the free acid using HATU to give the activated ester which was then added to the deprotected amine of the previous residue. The trityl protecting group on the serine residue was selectively removed by approximately 5 treatments of 1 to 3%TFA and 5% TIS in DCM over roughly 2-3 hours until the yellow color dissipated to yield resin-bound intermediate **2**. Analytical cleavage of from the resin using the same method described for cleavage from Wang resin allowed characterization of this intermediate by mass spectrometry. After standard rinsing DMF (3x) and DCM (3x), cyclization was initiated with 5 equivalents of finely ground K_2CO_3 in DMF and allowed to proceed at room temperature for >48 hours.

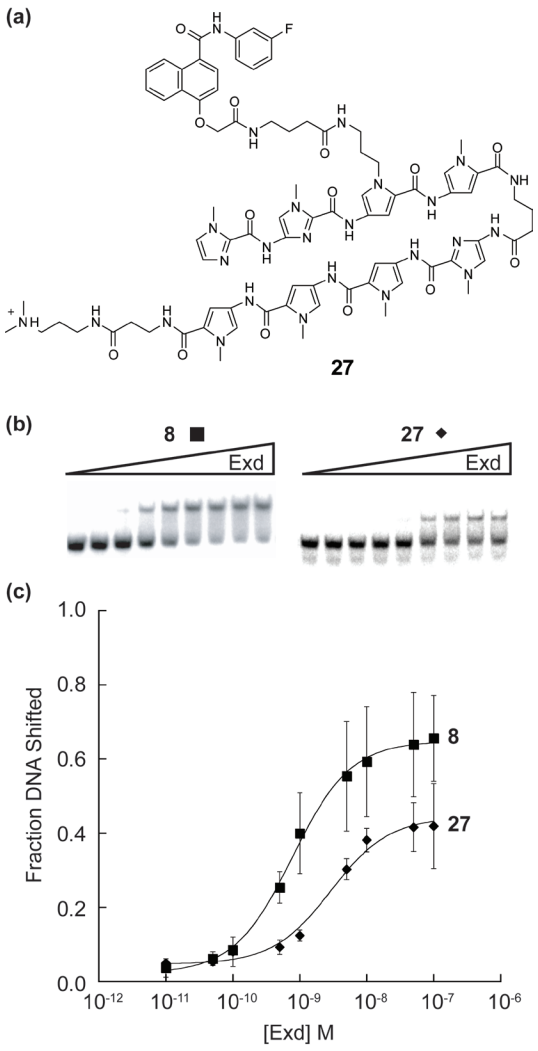


Figure B.9 Gel shift experiments comparing polyamide-WM conjugate **8** to polyamide-naphthalene conjugate **27**. a) Structure of polyamide-naphthalene conjugate. b) Representative gel shifts in the presence of **8** and **27**. c) Isotherms for Exd binding in the presence of **8** and **27**. Gel shifts were performed with 50 nM compound titrating Exd from 10 pM to 100 nM.

Cleavage from resin was performed using 85%TFA, 5%TIS, and 10%DCM (~4x, 2.5 mL) followed by removal of most of the solvent under reduced pressure. The resulting solid was dissolved in a small volume of DMF (~200 μL) and enough 20% CH_3CN/0.1% TFA in water was added to give a total volume of 10mL which was purified by preparative HPLC. Upon lyophilization of the appropriate fractions yielded 4.1 mg an off-white solid (3.7% yield). (ESI MS) average mass calculated for $C_{30}H_{35}N_6O_9^+$ $[M+H]^+$: 623.6, found 623.8 and calculated for $C_{30}H_{33}N_6O_9^-$ $[M-H]^-$: 621.6, found 621.8.

ImImPy$^{(propylamine)}$Py-γ-ImPyPyPy-β-Dp (4). The polyamide was synthesized according to standard protocols on BOC-β-Ala-PAM resin.[28] 265 mg of resin cleaved with 2 mL of neat *N,N*-dimethylaminopropylamine (Dp) at 37 °C for 48 hours, purified by preparative HPLC and lyophilized to give an off-white powder (10.9 μmol, ~17% yield). (MALDI-TOF MS) calculated for $C_{59}H_{76}N_{23}O_{10}^+$ 1266.6, observed 1266.7.

ImImPy$^{(macrocycle-propylamine)}$Py-γ-ImPyPyPy-β-Dp (5). Excess polyamide **4** (3.55 μmol, 2.2 eq) was added to macrocycle **3** (1 mg, 1.6 μmol, 1 eq) pre-activated (~5 min, RT) with

HBTU (1 eq) in DMF (~500 μL) and DIEA (8 equivalents). The resulting compound was purified by prepartive HPLC and lyophilzed to give a solid. (MALDI-TOF MS) $[M+H]^+$ base peak calculated for $C_{89}H_{108}N_{29}O_{18}^+$ 1871.8, found 1871.8.

Noncycle (6). This control compound was synthesized on Wang resin (0.305 g, 0.82 mmol/g, 0.250 mmol scale) using standard Fmoc peptide chemistry. The 3-nitrobenzoic acid moiety (Aldrich, 99%, 3.5 eq) was pre-activated with HATU (3.4 eq) in DMF with DIEA (4.5 eq) and coupled for 2 hours at RT. Cleavage from the resin was performed with 95% TFA/5% TIS, worked up as usual, and purified by preparative HPLC and lyophilized to give a solid. (ESI MS) average mass calculated for $C_{30}H_{37}N_6O_9^+$ $[M+H]^+$: 625.3, found 625.4 and calculated for $C_{30}H_{35}N_6O_9^-$ $[M-H]^-$: 623.3, found 623.4.

ImImPy$^{(noncycle\text{-}propylamine)}$Py-γ-ImPyPyPy-β-Dp (7). Polyamide **4** (0.97 μmol, 1 eq) was added to noncycle **6** (4.2 mg, 6.8 μmol, ~7 eq) pre-activated (~5 min, RT) with HBTU (3 eq) in DMF (~500 μL) and DIEA (20 eq). The resulting compound was purified by prepartive HPLC and lyophilzed to give a solid. (MALDI-TOF MS) $[M+H]^+$ calculated for $C_{89}H_{110}N_{29}O_{18}^+$ 1872.9, found 1873.0.

ImImPy$^{(Trp\text{-}Met\text{-}\varepsilon Ahx\text{-}propylamine)}$Py-γ-ImPyPyPy-β-Dp (8). This compound was synthesized and characterized as described in Chapter 2.

ImImPy$^{(N\text{-}Trp\text{-}N\text{-}But\text{-}\varepsilon Ahx\text{-}propylamine)}$Py-γ-ImPyPyPy-β-Dp (10). Polyamide **4** (1.5 μmol, 1 eq) was added to peptoid **17** pre-activated (~5 min, RT) with excess PyBOP and DIEA in DMF. The resulting compound was purified by prepartive HPLC and lyophilzed to give a solid (660 nmol, ~44 % yield). (MALDI-TOF MS) $[M+Na]^+$ calculated for $C_{84}H_{109}N_{27}O_{14}Na^+$ 1742.9, found 1742.9.

(15). This compound was synthesized using standard peptoid submonomer procedures[25,26] on 2-cholortrityl resin[27] employing 2-(methylthio)-ethylamine (Aldrich) and 3-(aminomethyl)indole (97%, Carbocore). (ESI-MS) Calculated for $C_{13}H_{23}N_2O_4S^+$ 303.1, observed 303.1.

N-Trp-N-Met-εAhx-OH (17). This compound was synthesized using standard peptoid

submonomer procedures[25,26] on 2-cholortrityl resin[27] employed butylamine (Aldrich) and 3-(aminomethyl)indole (97%, Carbocore). The peptoid was cleaved using 6:2:2 DCM:TFE: AcOH (1 mL, 3x), rinsed with DCM (1 mL), added to hexanes (5 mL, 2x), and rotavapped to dryness. The resulting film was dissolved in 500 μL DMF and diluted in 0.1% TFA in water and acetonitrile and purified by preparative HPLC. Fractions with consistent mass spectra by MALDI-TOF MS were pooled and lyophilized to give a white powder. (ESI-MS) $[M+Na]^+$ Calcd for $C_{25}H_{36}N_4O_5Na^+$ 495.3, observed 495.3. The MS/MS (35% collision energy) fragmentation pattern of the isolated *m/z* of 495.3 peak was consistent with the desired compound.

4-methoxy-1-naphthoic acid (20). This reaction was performed based on a previously reported procedure.[29] The starting 4-methoxy-1-naphthaldehyde **19** (5 g, 26.85 mmol) was dissolved in tBuOH (40 mL) and 2-methyl-2-butene (20 mL) to give a yellow solution. $NaClO_2$ (3.95 g in 40 mL ddH_2O with 4.80 g $NaH_2PO_4 \cdot H_2O$) was added in a drop-wise manner leading to the formation of a white precipitate. The reaction was allowed to proceed for 24 hours before filtering to collect the solid which was washed with hexanes (3x 50mL). The solid was dissolved in 0.1 N NaOH (60 mL) and extracted with hexanes and EtOAc (1x hexanes 30 mL, 1x EtOAc 50 mL, 1x hexanes 50 mL). The aqueous phase was acidified (pH 2) and the white precipitate was extracted with EtOAc (2 x 60 mL) and diethyl ether (1 x 60 mL). The combined organic layer was washed with brine (1x), dried with anhydrous magnesium sulfate, filtered and concentrated to yield 2.4 g of a white crystalline solid (45% yield). ^{1}H NMR ($CDCl_3$, 300 MHz) δ 4.102 (s, 3H), δ 6.87 (d, J = 8.7 Hz, 1H), δ 7.58 (t, 1H), δ 7.68 (t, 1H), δ 8.36 (d, J = 8.1 Hz, 1H), δ 8.44 (d, J = 9 Hz, 1H), δ 9.16 (d, J = 8.7 Hz, 1H); HRMS (FAB) exact mass calcd for $C_{12}H_{11}O_3$ requires 203.0708 *m/z*, found 203.0712.

***N*-(3-fluorophenyl)-4-methoxy-1-naphthamide (21).** The following procedure was adapted from a previously described reaction.[30] Into a dry, ventilated flask under argon the carboxylic acid **20** was added (2.104 g, 10.4 mmol) and dissolved in anhydrous DCM (20

mL) to which oxalyl chloride (20 mL) was added in a drop-wise manner. The vigorous reaction was allowed to proceed for 0.5 hours and the solvent and excess oxalyl chloride was removed under reduced pressure in a fume hood to give a yellow solid. The solid was dissolved in anhydrous THF (20 mL) with DIEA (7.25 mL, 41.6 mmol) and 3-fluoroaniline (1.20 mL, 12.48 mmol). The solution was refluxed for 30 min and the reaction was quenched with aqueous NH_4OH (20 mL, 30%) and the solvent was removed under reduced pressure. The dark brown residue was dissolved in EtOAc (100 mL) and extracted with HCl (1 x 100 mL, 5%), NaOH (1 x 100 mL, 0.1N), and brine (1 x 100 mL). The combined organic phase was removed under reduced pressure and the residue dissolved in minimal amount of EtOAc, precipitated from hexanes (200 mL), and collected by filtration as a white solid (1.8g, 59% yield). ^{1}H NMR (CD_3OD, 300 MHz) δ 4.07 (s, 3H), δ 6.91 (t, 1H), δ 6.98 (d, 1H), δ 7.3-7.6 (m, 4H), δ 7.73 (t, 2H), δ 8.30 (t, 2H); HRMS (FAB) exact mass calcd for $C_{18}H_{15}FNO_2$ requires 296.1087 *m/z*, found 296.1087.

***N*-(3-fluorophenyl)-4-hydroxy-1-naphthamide (22).** The following procedure was adapted from a previously described reaction.[31] In a dry flask, $AlCl_3$ (2.03 g, 15.2 mmol, 3 eq) was added to a solution ethanethiol (5 mL) and DCM (5 mL) cooled at 0 °C. After allowing the solution to reach RT, compound **21** (1.5 g, 5.08 mmol) was added under argon. The reaction was allowed to proceed overnight and the $AlCl_3$ was quenched by *very slowly* adding ddH_2O. The resulting precipitate was dissolved in EtOAc (40 mL) and the combined organic and aqueous phases were acidified to pH 2 with 1 N HCl and extracted with EtOAc (40 mL, 2x). The combined organic layers were washed with brine (1x 20 mL), dried with anhydrous magnesium sulfate, and concentrated under reduced pressure. The crude residue was purified over silica gel (10-50% EtOac/Hex gradient) and the product was isolated as an off-white amorphous solid (1.15 g, 80%). ^{1}H NMR (CD_3OD, 300 MHz) δ 6.85-6.88 (m, 2H), δ 7.33-7.60 (m, 4H), δ 7.65 (m, 2H), δ 8.26-8.29 (m, 2H); HRMS (FAB) exact mass calcd for $C_{17}H_{13}FNO_2$ requires 282.0930 *m/z*, found 282.0942.

methyl 2-(4-(3-fluorophenylcarbamoyl)naphthalen-1-yloxy)acetate (23). The following

procedure was based on a previously described reaction.[32] A solution of hydroxyl **22** (50 mg, 178 μmol) was made in THF (1 mL) with K_2CO_3 (37 mg, 196 μmol, 1.5 eq). Methylbromoacetate was added (18 μL, 588 μmol, 3.3 eq) and the resulting solution was refluxed for more than 3 hours. The solvent was removed under reduced pressure to give a brown residue and the remaining methylbromoacetate was removed under a high vacuum for 2 hours. The residue was dissolved in EtOAc (10 mL), extracted with water (10 mL, 2x) and brine (10 mL, 1x), dried over anhydrous magnesium sulfate, filtered, and concentrated under reduced pressure. The yellow crude material was purified over silica gel (10-30% hexanes/EtOAc gradient) to yield white solid with a yellow impurity (60 mg, 95% yield). ^{1}H NMR (CD_3OD, 300 MHz) δ 3.8 (s, 3H), δ 5.0 (s, 2H), δ 6.8-6.95 (m, 2H), δ 7.3-7.45 (m, 2H), δ 7.55-7.65 (m, 2H), δ 7.7 (d, 2H), δ 8.25 (d, 1H), δ 8.4 (d, 1H); HRMS (FAB) exact mass calcd for $C_{20}H_{17}FNO_4$ requires 354.1142 *m/z*, found 354.1150.

2-(4-(3-fluorophenylcarbamoyl)naphthalen-1-yloxy)acetic acid (24). The following procedure was based on a previously described reaction.[10] To a solution of ester **23** (50 mg, 142 μmol) in MeOH (2 mL) LiOH was added (213 μL of a 1M solution). The solution was refluxed for 2 hours and the solvent was removed under reduced pressure to yield a crude yellow residue which was dissolved in 10 mL of water which was extracted with hexanes (10 mL, 1x), EtOac (10 mL, 1x), and hexanes (10 mL, 1x). The aqueous layer was acidified to pH 2 with 1 N HCl to form a white precipitate which was extracted with EtOAc (20 mL, 3x). The combined organic layer were dried over anhydrous magnesium sulfate, filtered, and concentrated to dryness under reduced pressure to give an amorphous off-white solid (45 mg, 94% yield). ^{1}H NMR (d6-DMSO, 300 MHz) δ 4.97 (s, 2H), δ 6.89-6.99 (m, 2H), δ 7.33-7.80 (m, 6H), δ 8.23-8.32 (m, 2H), δ 10.63 (s, 1H); HRMS (FAB) exact mass calcd for $C_{19}H_{15}FNO_4$ requires 340.0985 *m/z*, found 340.0995.

4-(2-(4-(3-fluorophenylcarbamoyl)naphthalen-1-yloxy)acetamido)butanoic acid (26). The following procedure was based on a previously described reaction sequence.[33] To a solution of carboxylic acid **24** (45 mg, 133 μmol) in 1 mL anhydrous DMF and DIEA (23

μL, 1 eq) PyBOP was added (83 mg, 159 μmol, 1.2 eq) followed by γ-aminobutyric acid *tert*-butyl ester (29 mg, 146 μmol, 1.1 eq) and more DIEA (69 μL, 3 eq). Reaction was allowed to proceed for 18 hours and the solvent was removed under reduced pressure and the residue was dissolved in DCM (10 mL) and extracted with water, saturated $NaHCO_3$, 1 N HCl, and brine (10 mL, 1x each). The organic layer was evaporated to leave a yellow residue. The ester was hydrolyzed with 2:1 DCM:TFA (1.5 mL) and the solvent was removed under reduced pressure to yield a crude yellow-white residue which was dissolved in a minimum volume of DCM and purified over silica gel (9:1 EtOAc:hexanes + 1% AcOH) to give an off-white solid (41.2 mg, 73% yield). ^{1}H NMR (CD_3OD, 300 MHz) δ 1.98 (p, 2H), δ 2.33 (t, 2H), δ 3.37 (t, 2H), δ 4.80 (s, 2H), δ 6.85-6.99 (m, 2H), δ 7.32-7.45 (m, 2H), δ 7.56-7.74 (m, 4H), δ 8.29 (d, 1H), δ 8.51 (d, 1H); HRMS (FAB) exact mass calcd for $C_{23}H_{22}FN_2O_4$ requires 425.1513 *m/z*, found 425.1491; A_{max} ~230 nm (ε at 230 nm ≈ 2×10^4 $M^{-1}\cdot cm^{-1}$ and ε at 310 nm = 10^4 $M^{-1}\cdot cm^{-1}$).

ImImPy$^{(1,4\text{-napthalene-propylamine})}$Py-γ-ImPyPyPy-β-Dp (27). Polyamide **4** (2 μmol, 1 eq) was added to carboxylic acid **26** (10 mg, 23.6 μmol) pre-activated (~5 min, RT) with HBTU (8 μmol, 4 eq) in DMF (~280 μL) and DIEA (20 μmol). The resulting compound was purified by prepartive HPLC and lyophilzed to give a solid. (MALDI-TOF MS) $[M+H]^+$ calculated for $C_{82}H_{95}FN_{25}O_{14}^+$ 1672.8, found 1672.8. The conjugate's extinction coefficient was assumed to be the sum of the parent polyamide **4** and the small molecule **26** components (conjugate ε at 310 nm ≈ 79,500 $M^{-1}\cdot cm^{-1}$).

Electrophoretic Mobility Shift Assays (EMSAs). All EMSAs, or gel shifts, were performed at 4 or 20 °C as described in Chapter 2.[14]

B.6 References

(1) Passner, J. M., Ryoo, H. D., Shen, L. Y., Mann, R. S., Aggarwal, A. K. *Nature* **1999**, *397*, 714-719.

(2) Piper, D. E., Batchelor, A. H., Chang, C. P., Clearly, M. L., Wolberger, C. *Cell* **1999**, *96*, 587-597.

(3) LaRonde-LeBlanc, N. A. and Wolberger, C. *Genes Dev.* **2003**, *17*, 2060-2072.

(4) Pui, C., Relling, M. V., Downing, J. R. *N. Eng. J. Med.* **2004**, *350*, 1535-1548.

(5) Lu, Q., Wright, D. D., Kamps, M. P. *Mol. Cell. Biol.* **1994**, *14*, 3938-3948.

(6) Knoepfler, P. S. and Kamps, M. P. *Oncogene* **1997**, *14*, 2521-2531.

(7) McWhirter, J. R., Neuteboom, S. T. C., Wancewicz, E. V., Monia, B. P., Downing, J. R., Murre, C. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 11464-11469.

(8) Fu, X., McGrath, S., Pasillas, M., Nakazawa, S., Kamps, M. P. *Oncogene* **1999**, *18*, 4920-4929.

(9) Casagrande, G., Kronnie, G., Basso, G. *Haematologica/The Hematology Journal* **2006**, *91*, 765-771.

(10) Ji, T., Lee, M., Pruitt, S. C., Hangauer, D. G. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 3875-3879.

(11) Arndt, H.-D., Hauschild, K. E., Sullivan, D. P., Lake, K., Dervan, P. B., Ansari, A. Z. *J. Am. Chem. Soc.* **2003**, *125*, 13322-13323.

(12) Hauschild, K. E., Metzler, R. E., Arndt, H.-D., Moretti, R., Raffaelle, R., Dervan, P. B., Ansari, A. Z. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 5008-5013.

(13) Warren, C. L., Kartochvil, N. C. S., Hauschild, K. E., Foister, S., Brezinski, M. L., Dervan, P. B., Phillips, G. N., Ansari, A. Z. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 867-872.

(14) Stafford, R. L., Arndt, H.-D., Brezinski, M. L., Ansari, A. Z., Dervan, P. B. *J. Am. Chem. Soc.* **2007**, *129*, 2660-2668.

(15) Stafford, R. L. and Dervan, P. B. *submitted to J. Am. Chem. Soc. 2007.*

(16) Duboule, D. editor in *Guidebook to the Homeobox Genes*; Oxford University Press: Oxford, 1994.

(17) Shanmugam, K., Featherstone, M. S., Saragovi, H. U. *J. Biol. Chem.* **1997**, *272*, 19081-19087.

(18) Chakrabartim, P. and Janin, J. *Proteins* **2002**, *47*, 334-343.

(19) Lo Conte, L., Chothia, C., Janin, J. *J. Mol. Biol.* **1999**, *285*, 2177-2198.

(20) Feng, Y. and Burgess, K. *Chem. Eur. J.* **1999**, *5*, 3261-3272.

(21) Wang, Z., Jin, S., Feng, Y., Burgess, K. *Chem. Eur. J.* **1999**, *5*, 3273-3278.

(22) Burgess, K. *Acc. Chem. Res.* **2001**, *34*, 826-835.

(23) Zhang, A. J., Khare, S., Gokulan, K., Linthicum, D. S., Burgess, K. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 207-210.

(24) Simon, R. J., Kania, R. S., Zuckerman, R. N., Huebner, V. D., Jewll, D. A., Banville, S., Ng, S., Wang, L., Rosenberg, S., Marlowe, C. K., Spellmeyer, D. C., Tan, R., Frankel, A. D., Santi, D. V., Cohen, F. E., Bartlett, P. A. *Proc. Natl. Acad. Sci.* **1992**, *89*, 9367-9371.

(25) Zuckerman, R. N., Kerr, J. M., Kent, S. B. H., Moos, W. H. *J. Am. Chem. Soc.* **1992**, *114*, 10646-10647.

(26) Figliozii, G. M., Goldsmith, R., Ng, S. C., Banville, S. C., Zuckerman, R. N. *Methods Enzym.* **1996**, *267*, 437-447.

(27) Anne, C., Fournie-Zaluski, M-C., Roques, B. P., Cornille, F. *Tet. Lett.* **1998**, *39*, 8973-8974.

(28) Baird, E. E. and Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141-6146.

(29) Hillis, L. R. and Ronald, R. C. *J. Org. Chem.* **1985**, *50*, 470-473.

(30) Spencer, J. R., McGee, D., Allen, D., Katz, B. A., Luong, C., Sendzik, M., Squires, N., Mackman, R. L. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2023-2026.

(31) Node, M., Nishide, K., Fuji, K., Fujita, E. *J. Org. Chem.* **1980**, *45*, 4275-4277.

(32) Kramer, B., Fröhlich, R., Bergander, K., Waldvogel, S. R. *Synthesis* **2003**, *1*, 91-96.

(33) Haberhauer, G., Somogyi, L., Rebek, J., *Tet. Lett.* **2000**, *41*, 5013-5016.

Appendix C

Exd Expression Protocols

Exd was expressed by the Protein Expression Center (Beckman Institute, Caltech).

Mass spectrometry analysis of the Exd protein fragments was performed by the Peptide/ Protein MicroAnalytical Lab (PPMAL, Beckman Institute, Caltech).

C.1 First Exd Preparation Experimental Details

A plasmid "Exd-C" was received from the Ansari lab (University of Wisconsin, Madison) which was originally obtained from the Aggarwal lab (Mount Sinai School of Medicine) (Figure C.1). Protein expression and purification was performed by Dr. Peter Snow of the Protein Expression Center (Beckman Institute, Caltech). The plasmid was transformed into *E. coli* BL21-DE3 cells and overexpressed by treating with 1 mM IPTG. Cells were lysed by sonication and crude purification was performed by taking the 50% (or 90%) ammonium sulfate cut. The crude preparation was dialyzed against buffer A (150 mM potassium glutamate, 50 mM HEPES, 1 mM DTT, pH 7.0, 5% glycerol, 100 mM NaCl). The dialyzed protein was purified by FPLC (Figure 2a; flow rate = 0.5 mL/min, eluant A = buffer A, eluant B = buffer A with 1 M NaCl, 0.5 mL fractions). An SDS-PAGE gel stained with Coomassie blue showed pure fractions (29-33) with the predicted mass (~10 kDa) for the desired Exd fragment (Figure C.2b). Band excision (fraction 32) followed by proteolytic digestion (Lys-C achromobacter) and MALDI-TOF MS analysis was consistent with the predicted spectra using Sherpa 3.3.1 for Exd (residues 238-314) (Figure C.3c-d). Mascot (Matrix Science) search results from LC-MS/MS data for the digested protein identified Dpbx (i.e., *Drosophila* Pbx or Exd) as the sole match protein (score = 163). Fractions 29-33 were combined and dialyzed against buffer A. A sizing column (G30) was used to remove aggregates. MALDI-TOF MS analysis of the combined fractions of the undigested protein from Caltech was consistent with the presence of Exd residues 238-320 (calculated $[M+H]^+$ 9495.7 *m/Z*, observed $[M+H]^+$ 9496.0) (Figure C.3a). An undigested "crude" Exd preparation from the Ansari lab yielded a similar mass spectrum (Figure C.3b). The stock Exd concentration was determined to be 432 ± 22 μM using the method of Stoscheck[1] and a calculated extinction coefficient[2] of 12,600 $M^{-1}cm^{-1}$ for Exd residues 238-320 (Figure C.4). Total yield was 4.1 mg.

```
    1          11         21         31         41         51
    |          |          |          |          |          |
  1 MEDPNRMLAH TGGMMAPQGY GLSGQDDGQN AGSENEVRKQ KDIGEILQQI MSISEQSLDE

 61 AQARKHTLNC HRMKPALFSV LCEIKEKTVL SIRNTQEEEP PDPQLMRLDN MLIAEGVAGP

121 EKGGGGAAAA SAAAASQGGS LSIDGADNAI EHSDYRAKLA QIRQIYHQEL EKYEQACNEF

181 TTHVMNLLRE QSRTRPITPK EIERMVQIIH KKFSSIQMQL KQSTCEAVMI LRSRFLDARR

241 KRRNFSKQAS EILNEYFYSH LSNPYPSEEA KEELARKCGI TVSQVSNWFG NKRIRYKKNI

301 GKAQEEANLY AAKKAAGASP YSMAGPPSGT TTPMMSPAPP QDSMGYPMGS GGYDQQQPYD

361 NSMGGYDPNL HQDLSP
```

Figure C.1 Amino acid sequence of Exd isoform C of *D. Melanogaster*. Bold-faced type indicates the region of Exd expressed (238-320) and italicized type indicates the region of Exd used in the Ubx/Exd/DNA crystal structure (PDB code 1B8I). Both regions contain Exd's DNA-binding homeodomain.

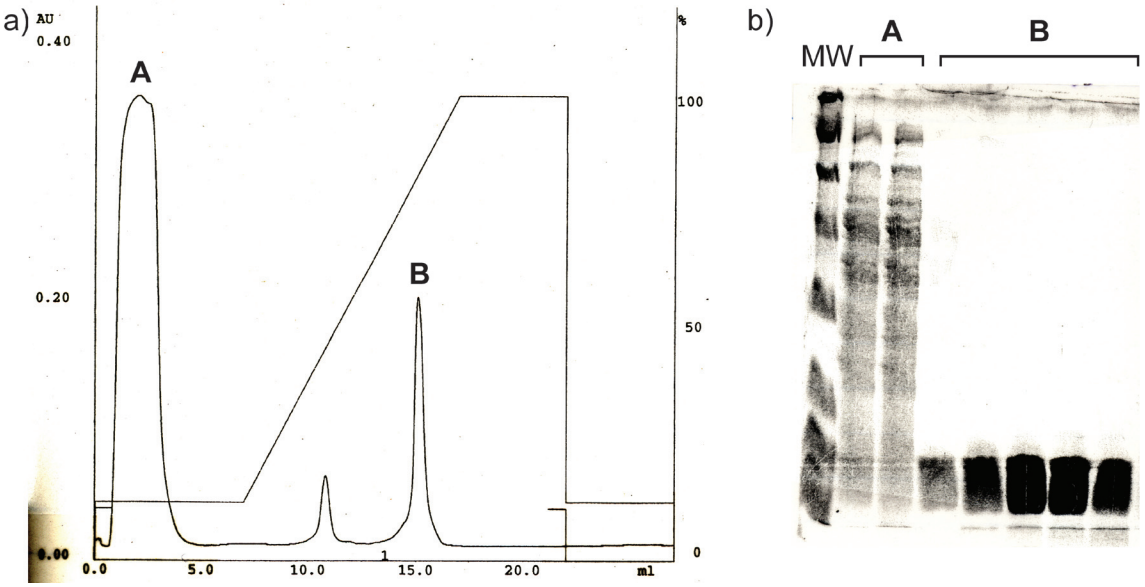


Figure C.2 Purification and analysis of the first Exd preparation. a) FPLC trace with major peaks A and B labeled. See text for details of purification. b) SDS-PAGE gel stained with Coomassie blue shows the content of peaks A (combined fractions 2-5) and B (fractions 29-33).

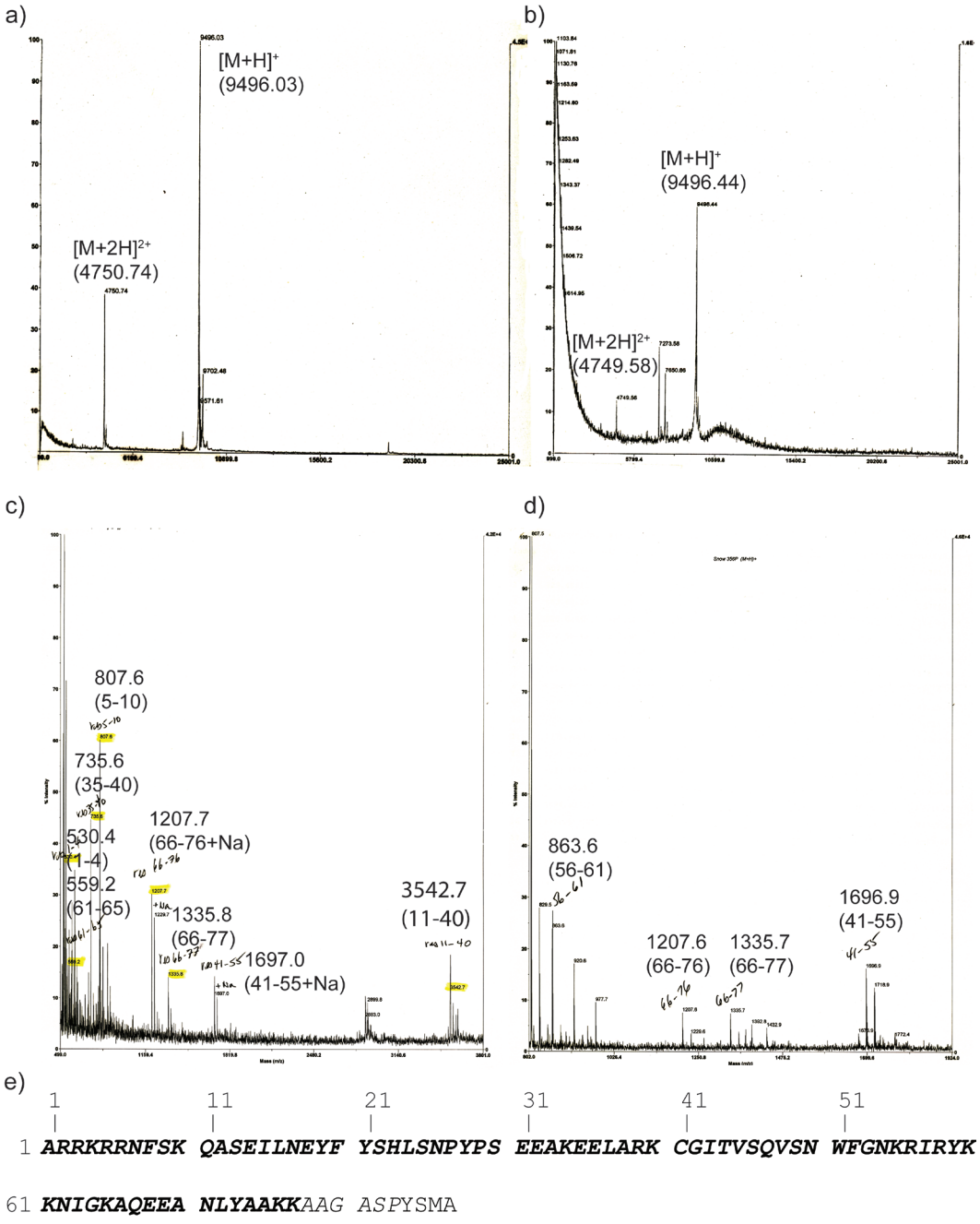


Figure C.3 MALDI-TOF MS analysis of different Exd preparations. a) Mass spectrum of Exd expressed at the Caltech Protein Expression Center by Peter Snow. b) Mass spectrum of "crude" Exd expressed by the Ansari lab and the University of Wisconsin, Madison. c) Mass spectrum #1 of proteolytic peptide fragments of Exd as determined at the Peptide/ Protein MicroAnalytical Lab (PPMAL) at Caltech by Gary Hathaway. d) Mass spectrum #2 of proteolytic peptide fragments of Exd determined by the PPMAL. e) Sequence of Exd reported in the original Arndt *et al* manuscript.[3] Bold-faced type indicates the region of Exd identified by proteolytical/MS analysis and italicized type indicates the region observed by undigested protein samples from Caltech and the University of Wisconsin, Madison.

C.2 Second Exd Preparation Experimental Details

The Exd-C plasmid was transformed in BL21-DE3 cells as above by Jost Vielmetter and Gilberto Salvo of the Caltech Protein Expression Center. Cells were grown to an OD of 0.6 at 37 °C and induced with 1 mM IPTG. Protein over-expression was allowed to proceed for 4 hours at 30 °C and cells were collected by centrifugation at 3000g for 30 min. For storage the cell pellet was flash frozen and stored at -80 °C. Cells were lysed in buffer B (1X PBS, 1 mM EDTA, 5 mM benzamidine, 1x PMSF, 1x complete protease inhibitor, pH 7.4) by sonication (Sonic Dismembrator 550 with Microtip at level 4 in pulsar/timer mode: 2 min duty cycle time, 0.5 s pulses interupted with 0.5 s pauses on ice). Cells were spun 2x at 13,000 rpm, 4 °C, 30 min with a Sorvall SS-34 rotor decanting the lysate into a fresh tube. A total of 5 mL lysate was obtained and stored at -80 °C. The diluted crude lysate (15 mL) was loaded onto a HiTrap Heparin HP column in injection buffer (25 mM TrisHCl, pH 7.5, 100 mM NaCl, 1 mM EDTA, 10% glycerol) and eluted with elution buffer (25 mM TrisHCl, pH 7.5, 1 M NaCl, 1 mM EDTA, 10% glycerol) with a flow rate of 1 mL/min collecting 1 mL fractions. Fractions 44-77 were pooled and concentrated down to 1 mL. No ammonium sulfate fractionation was employed. The pooled fractions were purified on a MonoS column by FPLC with a flow rate of 0.5 mL/min collecting 1 mL fractions (A buffer = injection buffer, B buffer = elution buffer) (Figure C.5). The major peaks were analyzed by an SDS-PAGE gel stained with Coomassie blue. Fractions 15 and 16 from the monoS column were pooled and 200 μL was loaded onto a sephadex-75 sizing column purified by an isochratic methods using the injection buffer (Figure C.6). Fractions 17-20 were pooled and concentrated down to ~200 μL. The concentration was estimated to be ~2 mg/mL (~30.2 μM) by Jost Vielmetter from the UV absorbance at 280 nm using a nanodrop spectrometer. Total yield was ~0.4 mg. An analytical sample showed a similar DNA binding affinity (within 2-fold) in the presence of a polyamide-FYPWMKG peptide conjugate at 20 °C as the preparation described in section C.1.

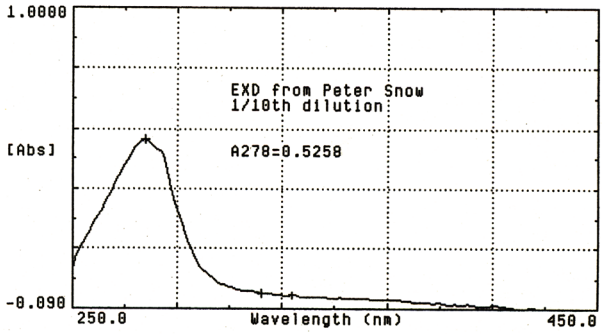


Figure C.4 UV-Vis spectrum of a 1/10th dilution of the Exd stock expressed by Peter Snow in 6.0 M guanadinium HCl, 0.01 M KH_2PO_4, 0.01 M K_2HPO_4 (pH 6.5) used for concentration determination.

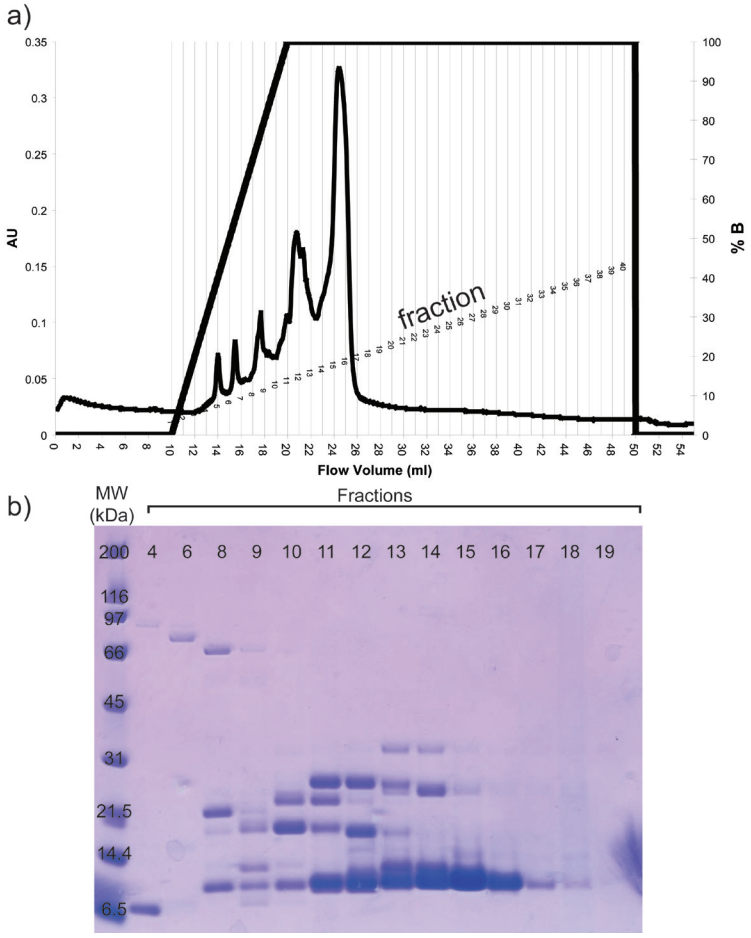


Figure C.5 Primary purification and analysis of the second Exd preparation. a) FPLC trace. See text for details of purification. b) SDS-PAGE gel stained with Coomassie blue shows the content of fractions 4, 6, and 8-19.

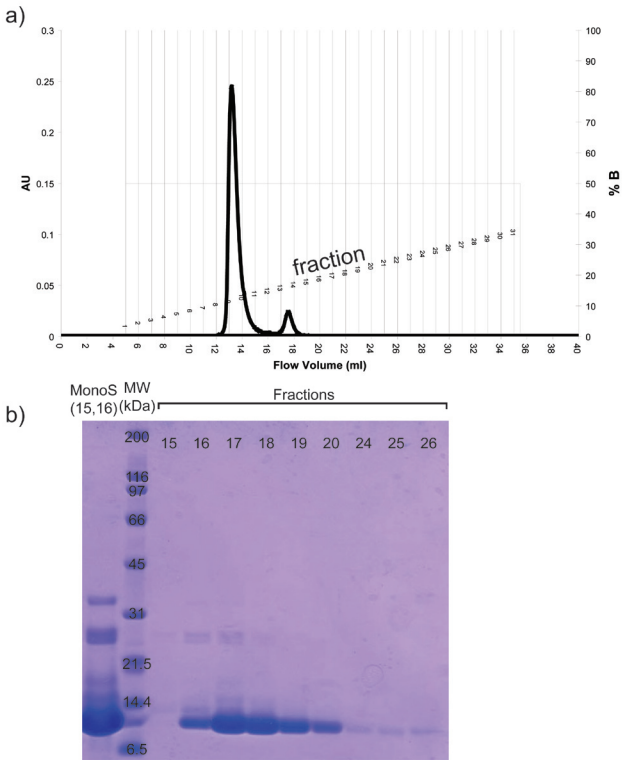


Figure C.6 Secondary purification and analysis of the second Exd preparation. a) FPLC trace. See text for details of purification. b) SDS-PAGE gel stained with Coomassie blue shows the content of fractions 15-20 and 24-26.

C.3 References

1. Stoscheck, C. M. *Methods in Enzym.* **1990**, *182*, 50-69.

2. Gill, S. C., von Hippel, P. H. *Anal. Biochem.* **1989**, *182*, 319-326.

3. Arndt, H.-D., Hauschild, K. E., Sullivan, D. P., Lake, K., Dervan, P. B., Ansari, A. Z. *J. Am. Chem. Soc.* **2003**, *125*, 13322-13323.

Appendix D

Solid-Phase Synthesis of Polyamides Using Marshall-Liener Resin

This work was done in collaboration with Rachel Wang and Mike Brochu (Dervan Group; Caltech).

D.1 A Brief Review of Solid-Phase Polyamide Synthesis

The solid-phase synthesis of DNA-binding pyrrole-imidazole polyamides was first described in the mid-1990s by the Dervan group.[1] The commercially available Boc-β-Ala-PAM resin solid support allows rapid synthesis of polyamides using Boc-protected monomers of *N*-methylpyrrole (pre-activated Boc-Py-OBt), *N*-methylimidazole (Boc-Im-OH), and γ-aminobutyric acid (Boc-γ-OH). Dimers of *N*-methylimidazole (Boc-Py-Im-OH and Boc-γ-Im-OH) are prepared by solution-phase methods due to poor solid-phase imidazole coupling to pyrrole and γ-aminobutyric acids. Nucleophilic cleavage of Boc-β-Ala-PAM resin with *N,N*-dimethylaminpropylamine (Dp) affords polyamides with -β-Dp tails which possesses the occasionally undesirable W over S selectivity[2] (where W = A or T and S = G or C) and which appears to often be a negative determinant for cell uptake.[3] The "β-less" Boc-Py-PAM resin has been reported to exhibit poor susceptibility to nucleophilic cleavage in the presence of amines such as Dp,[4] although cleavage to the corresponding carboxylic acid is possible using palladium.[5] The synthesis of polyamides using Kaiser oxime resin allows covalently attachment of Py monomers directly to the solid support through an oxime functional group. The oxime ester shows facile cleavage in the presence of a variety of nucleophiles to yield polyamides without a β-alanine spacer.[6] Care must be taken to minimize the amount of TFA during BOC deprotection (20% for Py and 50% for Im compared to 80% for PAM). Although the yield and purity of polyamides derived from oxime resin are often comparable to those from PAM resin, sometimes synthesis using oxime resin can be problematic and low-yielding.

A number of additional methods have been reported for the solid-phase synthesis of polyamides in efforts to obtain higher yields at minimal cost while still enabling flexibility of synthesis. Specifically, Fmoc-based solid-phase polyamide synthesis has also been developed using Wang resin[7] and MBHA resin[8] and both produce as good or better yields and purities as Boc-based chemistry. The Sugiyama group frequently employs Fmoc-based synthesis,[9] but the Boc-based chemistry is still frequently used by the Dervan group.

Safety-catch resins have also been reported employing both Fmoc and Boc protection strategies to produce comparable yields to PAM and oxime resins,[10-11] but have not garnered widespread use. One report has suggested improvements for Boc-based polyamide synthesis using alternative coupling reagents (i.e., HATU) and scavengers (phenol and water),[12] but one disadvantage is that the coupling reagent is currently more expensive than traditional ones (i.e., HBTU and DCC). More convergent Boc-based polyamide synthesis using dimers and trimers for coupling have also been reported to give improved yields,[13] but this often requires the time-consuming preparation of unique oligomers prior to the solid-phase synthesis of the target polyamide. Similarly, scattered reports of complete solution-phase synthesis have shown high yields (~1 g) for polyamides are obtainable,[14-15] but the approach is not as rapid as solid-phase methodology. A putative liquid-solid phase system has also been reported, but only small oligomers have been synthesized thus far using this approach.[16] This appendix describes the use of 4-hydroxyphenylsulfanylmethyl polystyrene resin, originally reported by Marshall and Liener in the 1970s for peptide synthesis,[17] as an alternative solid-phase support for rapid Boc-based polyamide synthesis. Similar to oxime resin, polyamides can be synthesized without a C-terminal β-alanine, but deprotection with higher amounts of TFA is possible.

D.2 Experimental Protocols

Resin loading. A total of 300 mg 4-hydroxyphenylsulfanylmethyl polystyrene resin (i.e Marshall-Liener resin, NovaBiochem;1.70 mmol/g, ~0.51 mmol scale) was treated with 2 equivalents of Boc-Py-OBt in 5 mL DMF and 3 equivalents of DIEA for 48 hours at RT. Successful loading has been accomplished in as short as 24 hours.

Coupling. Couplings were performed as described previously for Boc-β-Ala-PAM resin.[1]

Deprotection. All deprotection reactions were performed with 50% TFA/DCM for 25 min.

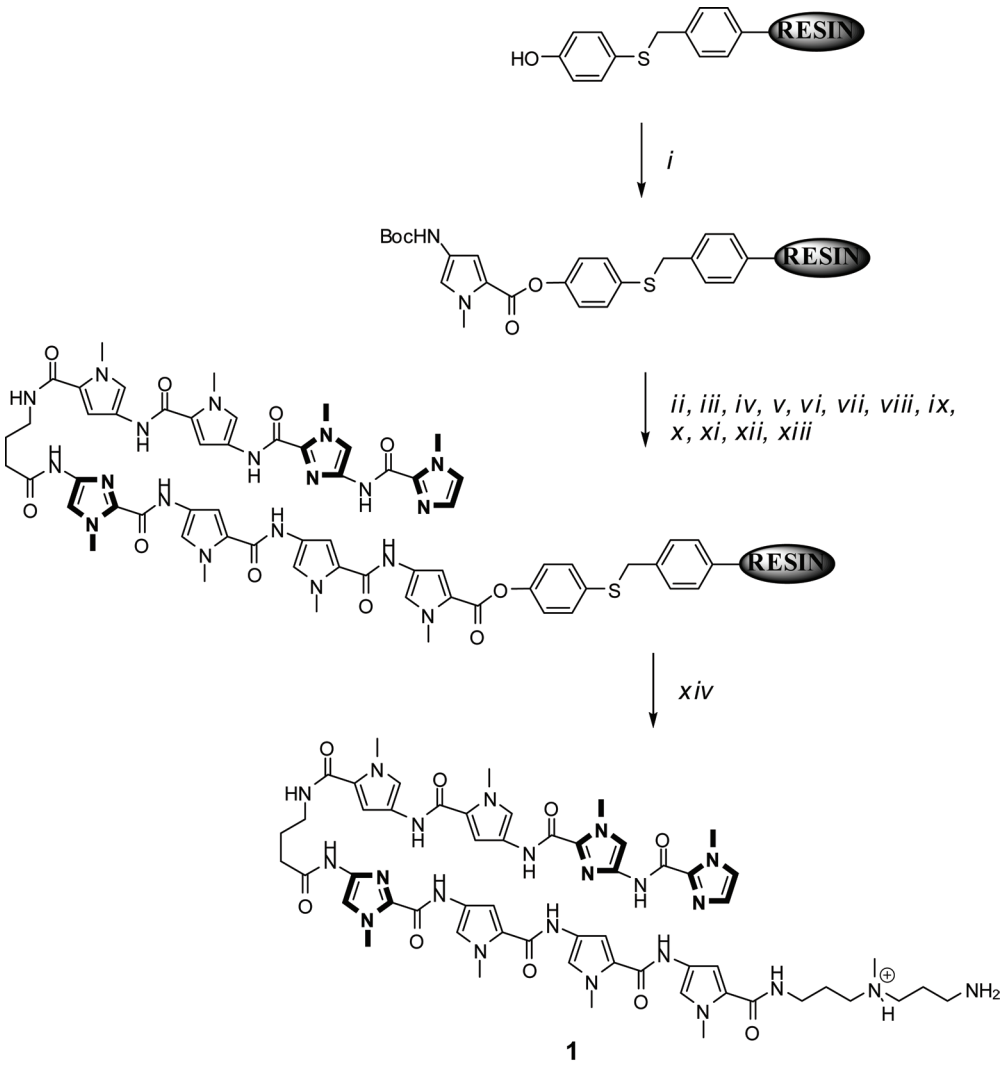


Figure D.1 Solid phase synthetic scheme for ImImPyPy-γ-ImPyPyPy-β-(+)-NH_2 from commercially available 4-hydroxyphenylsulfanylmethyl polystyrene resin, i.e., Marshall-Liener resin: *i*) 2 equivalents Boc-Py-OBt, 3 equivalents DIEA, DMF, 24 h *ii*) 50% TFA/ DCM, 30 min *iii*) 2 equivalents Boc-Py-OBt, 3 equivalents DIEA, DMF, 2 h, *iv*) 50% TFA/DCM, 30 min, *v*) 2 equivalents Boc-Py-OBt, 3 equivalents DIEA, DMF, 2 h *vi*) 50% TFA/DCM, 30 min, *vii*) 2 equivalents Boc-γ-Im-OH, 1.9 equivalents HBTU, 3 equivalents DIEA, DMF, 2 h, *viii*) 50% TFA/DCM, 30 min, *ix*) 2 equivalents Boc-Py-OBt, 3 equivalents DIEA, DMF, 2 h, *x*) 50% TFA/DCM, 30 min, *xi*) 2 equivalents Boc-Py-OBt, 3 equivalents DIEA, DMF, 2 h, *xii*) 50% TFA/DCM 30 min, *xiii*) 2 equivalents ImIm-OH, 1.9 equivalents HBTU, 3 equivalents DIEA, DMF, 2 h, *xiv*) 3,3′-diamino-*N*-methyl-dipropylamine, 50 °C, 24 h.

Washing. Following each deprotection the resin was drained and rinsed 2x DCM, 1x 4:1 DMF:DIEA, and 2x DMF. Following each coupling the resin was drained and rinsed 2x DMF and 2x DCM.

Analytical cleavages. To monitor the progress of synthesis a small sample of the resin was treated with a small volume of *N,N*-dimethylaminopropylamine for 15 min at 95 °C, filtered, diluted in 0.1% TFA in H_2O, and analyzed by HPLC.

Preparative cleavage. Similar to a couple relatively recent reports,[18-19] cleavage from Marshall-Liener resin did not require sulfur oxidation with hydrogen peroxide as originally suggested.[17]

ImImPyPy-γ-ImPyPyPy-(+)-NH_2 (1). 105 mg of resin containing the polyamide was treated with 1 mL of neat 3,3′-diamino-*N*-methyl-dipropylamine and incubated at 50 °C for 16.5 hours. The resin filtered and the filtrate was diluted with 0.1% TFA in H_2O and acetonitrile and purified 2x by preparative HPLC. Fractions with product identified by MALDI-TOF MS were flash frozen and lyophilized to yield a solid (11.0 μmol, 14.6% yield, 80-90% pure). (MALDI-TOF MS) $[M+H]^+$ calcd for $C_{56}H_{71}N_{22}O_9^+$ 1195.6, observed 1195.7.

ImImPy$^{(propylamine)}$Py-γ-ImPyPyPy-(+)-NHBoc (2). 50 mg of resin containing the polyamide was treated with 300 mg of *N*-Boc-3,3′-diamino-*N*-methyl-dipropylamine in 500 μL DMF at 50 °C for 24 hours. The resin was filtered and purified as described for above (2.3 μmol, 8.5% yield, ~90% pure). (MALDI-TOF MS) $[M+H]^+$ calcd for $C_{63}H_{84}N_{23}O_{11}^+$ 1338.7, observed 1338.7.

D.3 References

1. Baird, E. E., Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6141-6146.

2. Swalley, S. E., Baird, E. E., Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 1113-1120.

3. Edelson, B. S., Best, T. P., Olenyuk, B., Nickols, N. G., Doss, R. M., Foister, S., Heckel, A., Dervan, P. B. *Nucleic Acids Res.* **2004**, *32*, 2802-2818.

4. Barid, E. E. Ph.D. Thesis, California Institute of Technology, 1999.

5. Herman, D. M., Turner, J. M., Baird, E. E., Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 1121-1129.

6. Belitsky, J. M., Nguyen, D. H., Wurtz, N. R., Dervan, P. B. *Bioorg. Med. Chem.* **2002**, *10*, 2767-2774.

7. Wurtz, N. R., Turner, J. M., Baird, E. E., Dervan, P. B. *Org. Lett.* **2001**, *3*, 1201-1203.

8. Vázquez, E., Caamaño, A. M., Castedo, L., Mascareñas, J. L. *Tet. Lett.* **1999**, *40*, 3621-3624.

9. Ayama, H., Saito, T., Bando, T., Fukuda, N., Sugiyama, H. *Nucleic Acids Res. Supp.* **2003**, *3*, 67-68.

10. Fattori, D., Kinzel, O., Ingallinella, P., Bianchi, E., Pessi, A. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 1143-1147.

11. Marquez, M. A. Ph. D. Thesis, California Institute of Technology, 2005.

12. Krutzik, P. O., Chamberlain, A. R. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2129-2132.

13. Sinyakov, A. N., Feshchenko, M. V., Ryabinin, V. A. *Russ. J. Bioorg. Chem.* **2003**, *4*, 402-404.

14. Sinyakov, A. N., Feshchenko, M. V., Ryabinin, V. A. *Russ. J. Bioorg. Chem.* **2004**, *30*, 98-99.

15. Yuan, G., Tang, F. *ARKIVOC* **2003**, *ii*, 32-37.

16. König, B., Rödel, M. *Chem. Commun.* **1998**, 605-606.

17. Marshall, D. L., Liener, I. E. *J. Org. Chem.* **1970**, *35*, 867-868.

18. Fantauzzi, P. P., Yager, K. M. *Tet. Lett.* **1998**, *39*, 1291-1294.

19. Breitenbucher, J. G., Johnson, C. R., Haight, M., Phelan, J. C. *Tet. Lett.* **1998**, *39*, 1295-1298.

Appendix E

Dimerizer-Exd-DNA Crystallization Trials

E.1 Experimental Details

Crystallization of a polyamide-peptide/Exd/DNA complex was inspired by the three related HOX/PBC/DNA ternary complex crystal structures (Figure E.1).[1-3] Initial crystallization experiments were attempted using the originally reported conjugate **1** (HPLC purified, >99% pure) (Figure E.2).[4] Crude DNA inspired from the Exd/Ubx/DNA structure[1] was obtained from Integrated DNA Technologies and purified by non-denaturing PAGE utilizing UV shadowing and band extraction. DNA was eluted from the gel with 2M NaCl and filtered. DNA was precipitated with iPrOH (1x) and 75% EtOH in H_2O (2x) and lyophilized before being dissolved in 18 MΩ water. DNA purity was assessed by analytical HPLC using a Clarity oligo-RP column (150 mm × 4.60 mm, 5 μm, Phenomenex, A buffer = 0.05 M TEAA (pH 7, Fluka), B buffer = acetonitrile). DNA was annealed prior to addition of conjugate **1** or Exd by heating to 95 °C followed by slowly cooling to RT (~20 °C). Exd expressed by Peter Snow (see section C.1) was concentrated to ~19.4 mg/mL (~2.04 mM) at 4 °C using a Microcon YM-3 (MWCO of 3 kDa) and diluted down to ~10 mg/mL in crystallization buffer (10 mM HEPES, 1 mM EDTA, 1 mM DTT, pH 8.0) similar to a previously reported procedure for Pbx.[2] A total of 60 nmol of each component was added in the following order: 24 μL 2.5 mM dsDNA, 46 μL H_2O (18 MΩ), 50 μL 1.2 mM conjugate **1** (pre-sonicated to ensure homogeneity), and 60 μL Exd, which was incubated for ~3 hours at 20 °C. Crystallization was performed using the hanging drop method, i.e., 1:1 mixture of precipitant to complex solution. Crystallization screens from Hampton (Natrix and Crystal Screen Cryo) and Emerald Biosystems (Cryo I/II) and follow-up grid screens (precipitate concentration in 1% increments and pH in 0.1 unit increments) yielded several crystals (Figure E.3). All crystals diffracted poorly (best resolution ~6-8 Å) at room temperature and many cryoprotectant solutions destroyed crystals (30% PEG 400, PFO-X125, 25% glycerol, 25% ethylene glycol, 25% isopropanol, 5% (2R,3R)-(+)-2,3-butanediol, 25% DMSO, 1.4 M LiCl, and 0.68 M $Mg(OAc)_2$). Initial screens with poorly

formed crystals showed that some appeared to tolerate carbohydrate-based cryoprotectants better (D-(+)-raffinose, D-(+)-trehelose, and D-(+)-glucose).

E.2 References

(1) Passner, J. M.; Ryoo, H. D.; Shen, L. Y.; Mann, R. S.; Aggarwal, A. K. *Nature* **1999**, *397*, 714-719.

(2) Piper, D. E.; Batchelor, A. H.; Chang, C. P.; Clearly, M. L.; Wolberger, C. *Cell* **1999**, *96*, 587-597.

(3) LaRonde-LeBlanc, N. A.; Wolberger, C. *Genes Dev.* **2003**, *17*, 2060-2072.

(4) Stafford, R. L., Arndt, H.-D., Brezinski, M. L., Ansari, A. Z., Dervan, P. B. *J. Am. Chem. Soc.* **2007**, *129*, 2660-2668.

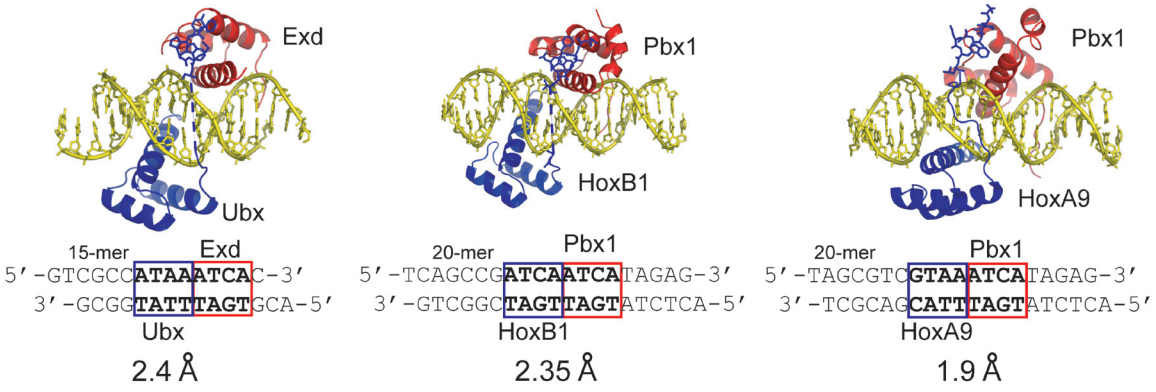


Figure E.1 Summary of HOX/PBC/DNA ternary complex crystal structures.

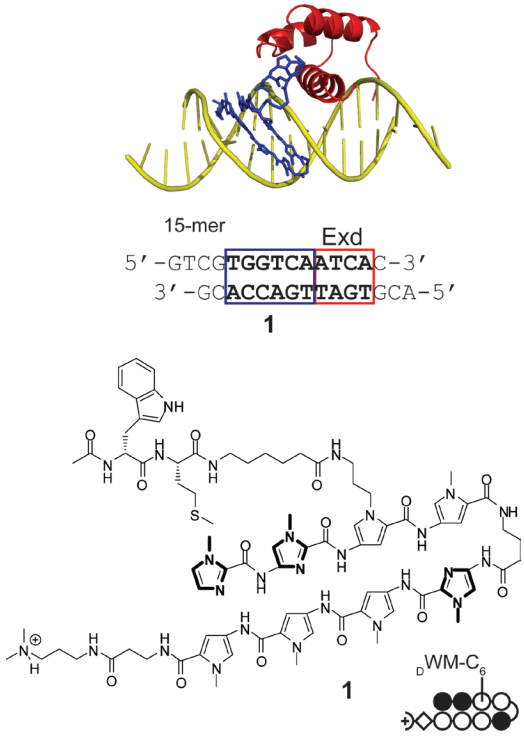


Figure E.2 Design of polyamide-peptide/Exd/DNA complex crystallization experiments. A model of the presumed complex is shown above. The sequence of the DNA duplex and the structure of the polyamide-peptide conjugate **1** are shown below.

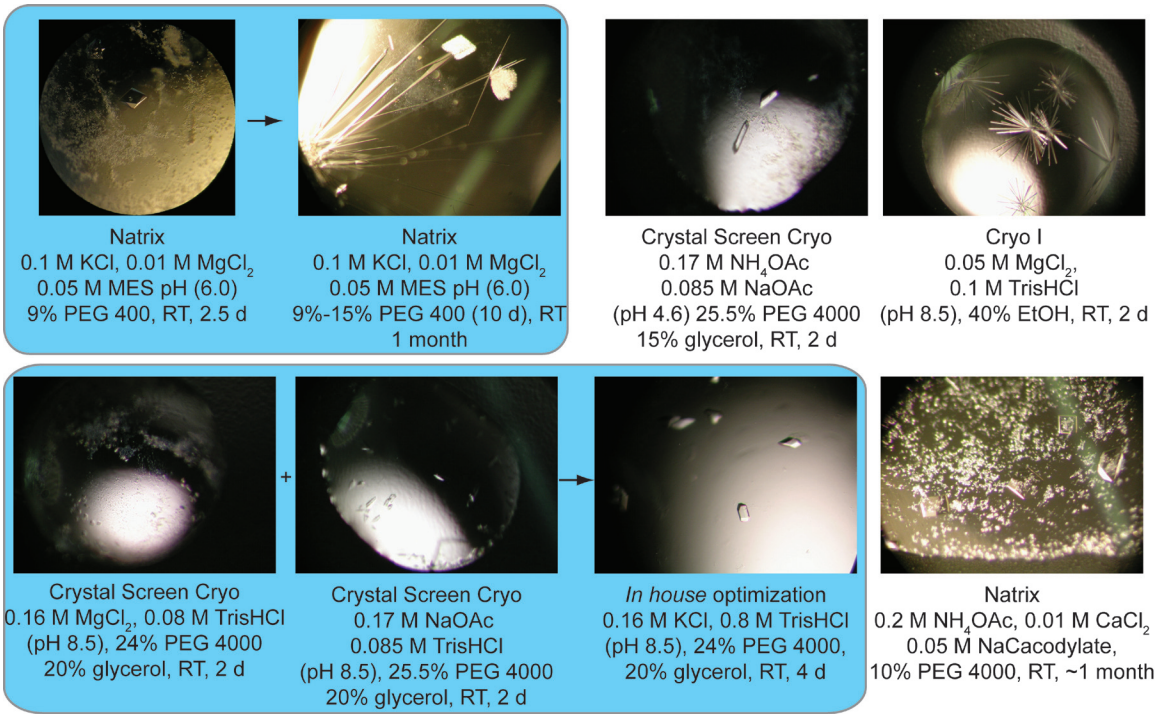


Figure E.3 Pictures of select crystals and the conditions used for crystallization.

Appendix F

Supplemental Figures

The figures in this section were taken in part from a manuscript coauthored with Peter B. Dervan (Caltech).

(Stafford, R. L, Dervan, P. B. "The Reach of Linear Protein-DNA Dimerizers" *submitted to J. Am. Chem. Soc.* **2007**)

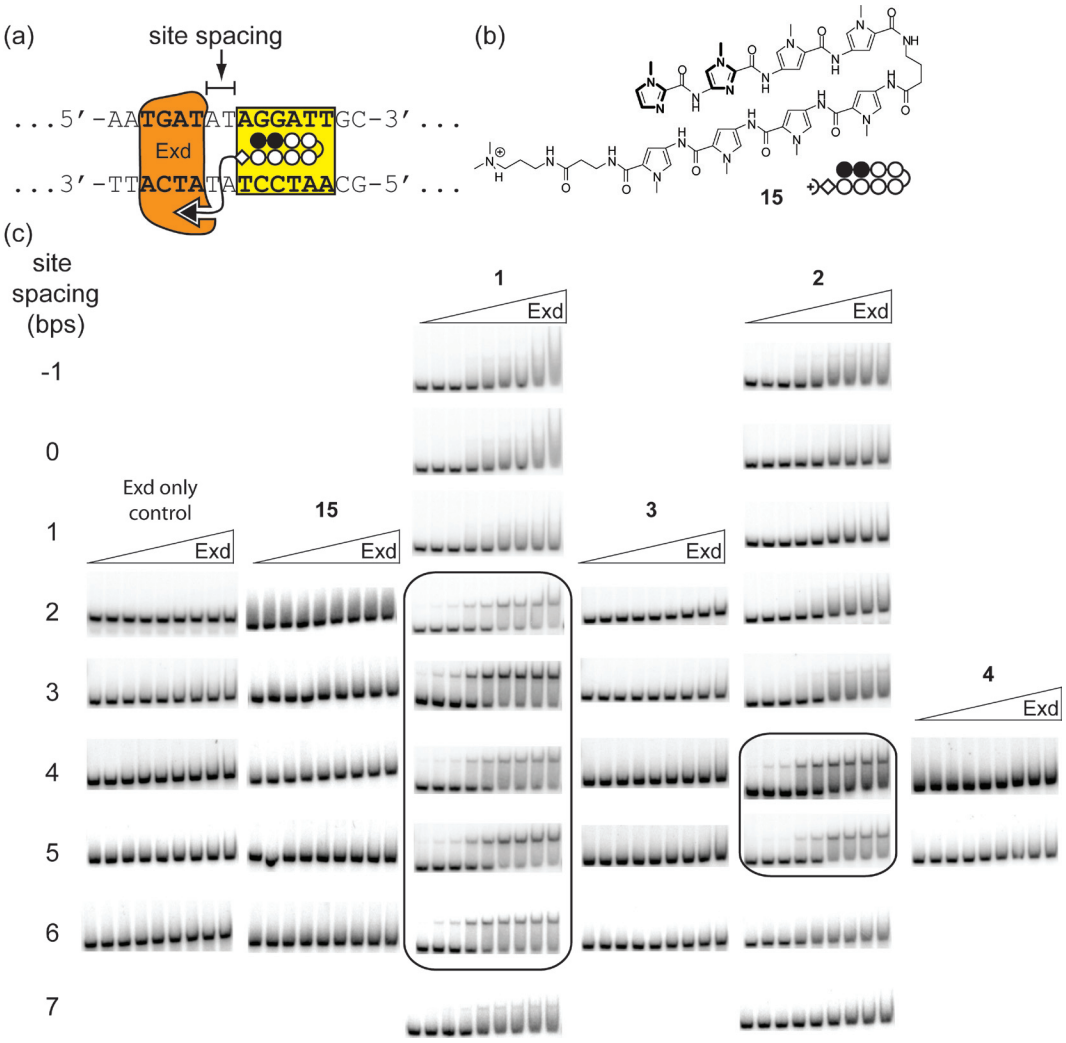


Figure F.1 Representative gel shift results for Exd in the presence of linear protein-DNA dimerizers with differently sized linkers between the two proximally oriented DNA sites *with control compounds*. (a) Schematic of the positions of the Exd and polyamide DNA-binding sites showing the A,T spacer region which was varied. (b) Structure of parent polyamide ImImPyPy-γ-PyPyPyPy-β-Dp (**15**). (c) Gel shift experiments show that **1** forms a stable complex with Exd and DNA with site spacings between 2-6 bps, whereas **2** only forms stable complexes from 4-5 bps. It is also shown that Exd by itself, in the presence of the parent polyamide **15**, or double-AA mutants **3** and **4**, does not yield a stable complex with DNA at these sites. A concentration of 50 nM of the indicated compound was added to all lanes and Exd was titrated from 10 pM to 100 nM.

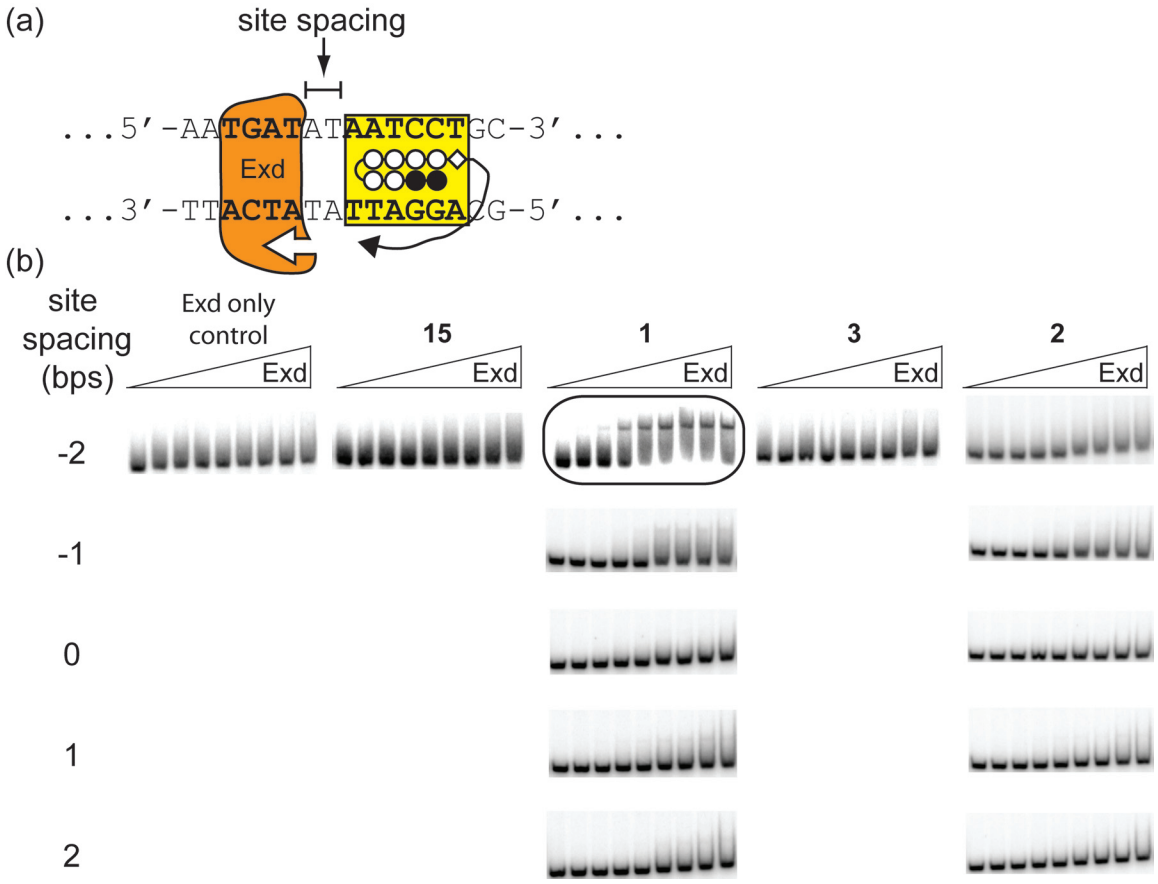


Figure F.2 Representative gel shift results for Exd in the presence of linear protein-DNA dimerizers with differently sized linkers between the two distally oriented DNA sites *with control compounds*. (a) Schematic of the positions of the Exd and polyamide DNA-binding sites showing the A,T spacer region which was varied. (b) Gel shift experiments show that **1** forms a stable complex with Exd and DNA when there is a 2 bp overlap, i.e., a site spacing of -2, between the two sites, whereas **2** does not form a complex with Exd at any of these sites. It is also shown that Exd by itself, in the presence of parent polyamide **15**, or double-AA mutant **3**, does not yield a stable complex with DNA at this site. A concentration of 50 nM of the indicated compound was added to all lanes and Exd was titrated from 10 pM to 100 nM.

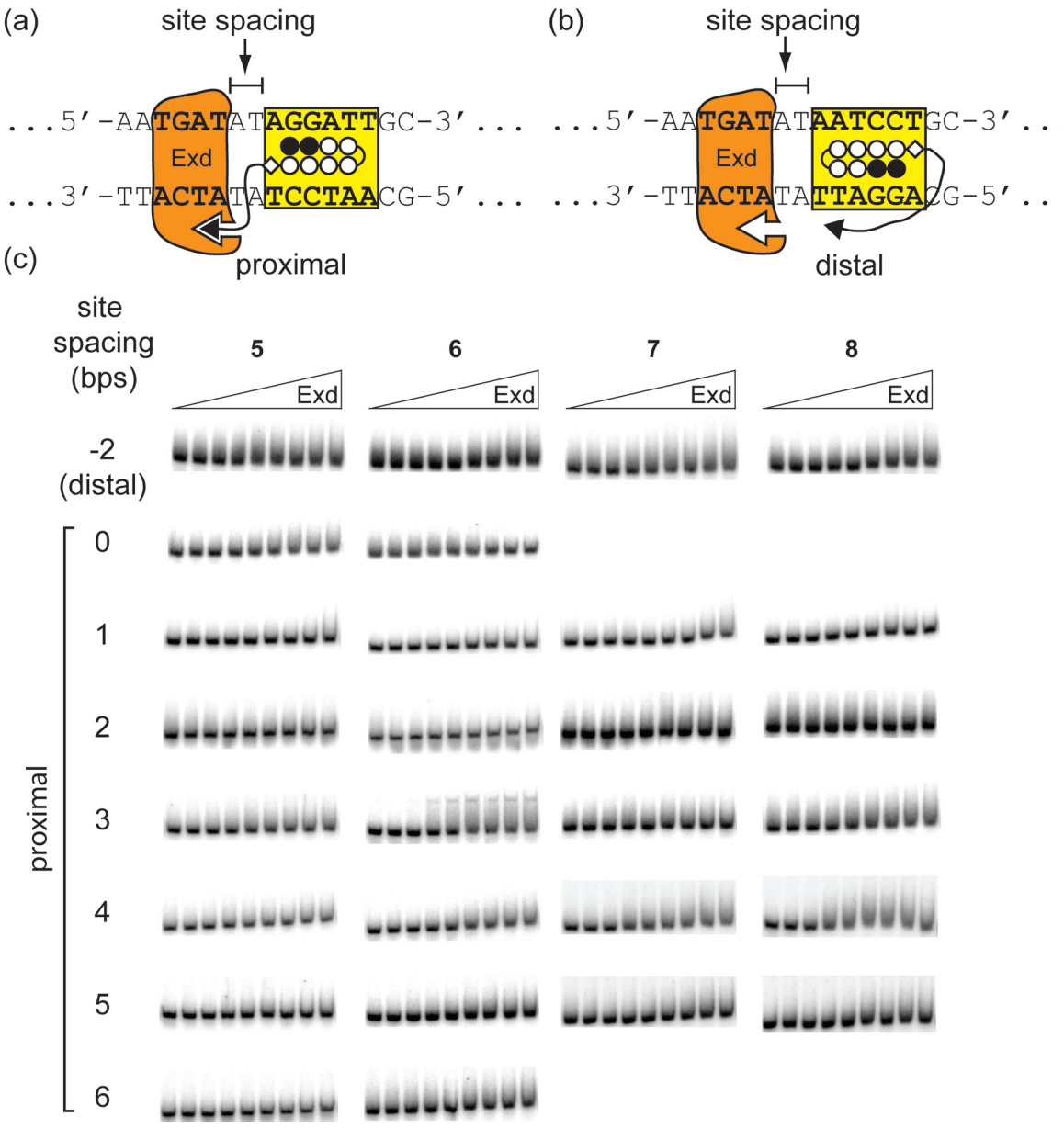


Figure F.3 Representative gel shift results for Exd in the presence of WM dipeptide linear protein-DNA dimerizers with differently sized spacers between the two DNA sites. (a) Schematic of the positions of the Exd and polyamide DNA-binding sites showing the poly-AT spacer region which was varied in the proximal orientation. (b) Schematic of the positions of the Exd and polyamide DNA-binding sites showing the A,T spacer region which was varied in the distal orientation. (c) Gel shift experiments show that none of the site spacings lead to stable complex formation for compounds **5-8**. A weak complex with compound **6** is observed at a spacing of 3 bp in the proximal orientation ($\Theta_{app} = 0.3 \pm 0.1$). A concentration of 50 nM of the indicated compound was added to all lanes and Exd was titrated from 10 pM to 100 nM.

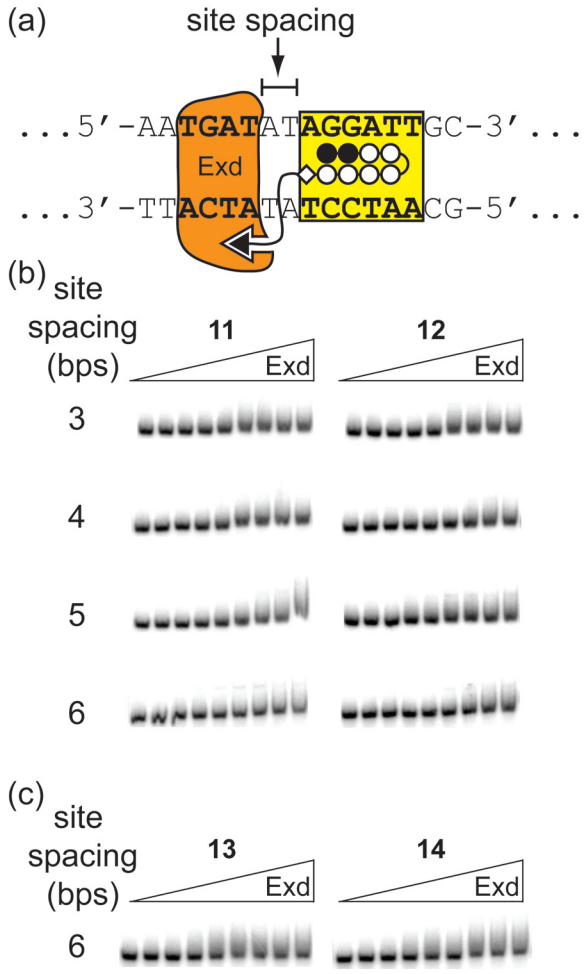


Figure F.4 Representative gel shift results for Exd in the presence of WMK tripeptide linear protein-DNA dimerizers with differently sized spacers between the two proximally oriented DNA sites. (a) Schematic of the positions of the Exd and polyamide DNA-binding sites showing the A,T spacer region which was varied in the proximal orientation. (b) Gel shift experiments show that none of the site spacings lead to stable complex formation for compounds **11** or **12**. (c) Gel shift experiments show that acetylated WMK tripeptide conjugates **13** and **14** do not lead to a stable complex at the proximal 6 bp spacing site. A concentration of 50 nM of the indicated compound was added to all lanes and Exd was titrated from 10 pM to 100 nM.

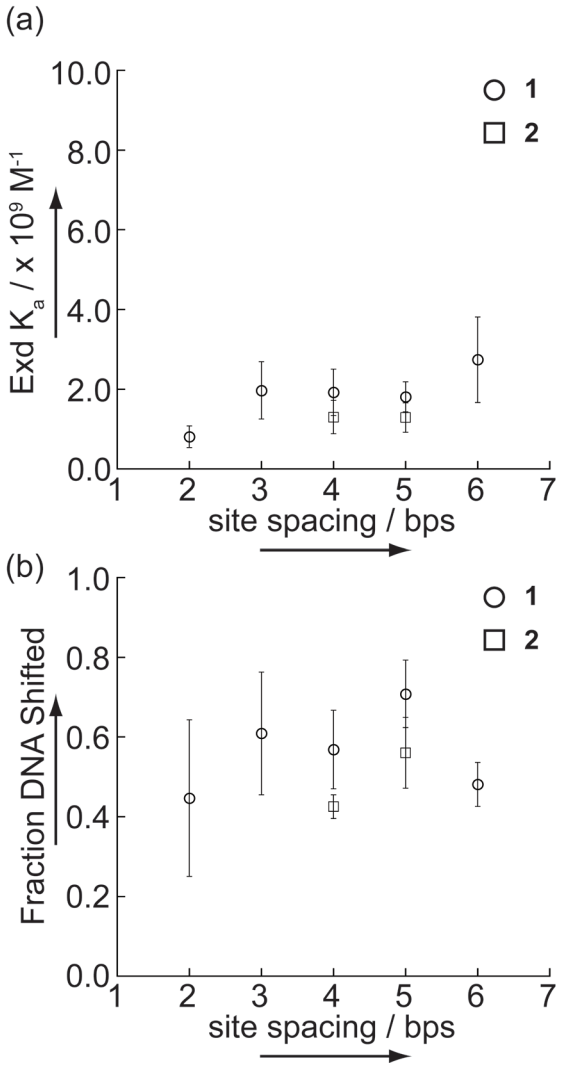


Figure F.5 The dependence of binding affinity (Ka) and the apparent fraction of DNA shifted with respect to the site spacing is shown. (a) The binding affinity of Exd does not change by more than 3-fold as the site spacing is changed in the presence of compounds **1** and **2** when a complex is observed. (b) The amount of complex formed also does not change substantially as the site spacing is changed.

Proximal -1 duplex (47 bps):
5′-GATCTCCCGGCGAA**TGATGGATT**GCGTCGGCGCCACTGTCACCCGGA-3′
3′-CTAGAGGGCCGCTT**ACTACCTAA**CGCAGCCGCGGTGACAGTGGGCCT-5′

Proximal 0 duplex (48 bps):
5′-GATCTCCCGGCGAA**TGATTGGATT**GCGTCGGCGCCACTGTCACCCGGA-3′
3′-CTAGAGGGCCGCTT**ACTAACCTAA**CGCAGCCGCGGTGACAGTGGGCCT-5′

Proximal 1 duplex (49 bps):
5′-GATCTCCCGGCGAA**TGAT**A**TGGATT**GCGTCGGCGCCACTGTCACCCGGA-3′
3′-CTAGAGGGCCGCTT**ACTA**T**ACCTAA**CGCAGCCGCGGTGACAGTGGGCCT-5′

Proximal 2 duplex (50 bps):
5′-GATCTCCCGGCGAA**TGAT**AT**AGGATT**GCGTCGGCGCCACTGTCACCCGGA-3′
3′-CTAGAGGGCCGCTT**ACTA**TA**TCCTAA**CGCAGCCGCGGTGACAGTGGGCCT-5′

Proximal 3 duplex (51 bps):
5′-GATCTCCCGGCGAA**TGAT**ATA**AGGATT**GCGTCGGCGCCACTGTCACCCGGA-3′
3′-CTAGAGGGCCGCTT**ACTA**TAT**TCCTAA**CGCAGCCGCGGTGACAGTGGGCCT-5′

Proximal 4 duplex (52 bps):
5′-GATCTCCCGGCGAA**TGAT**ATAT**AGGATT**GCGTCGGCGCCACTGTCACCCGGA-3′
3′-CTAGAGGGCCGCTT**ACTA**TATA**TCCTAA**CGCAGCCGCGGTGACAGTGGGCCT-5′

Proximal 5 duplex (53 bps):
5′-GATCTCCCGGCGAA**TGAT**ATATA**TGGATT**GCGTCGGCGCCACTGTCACCCGGA-3′
3′-CTAGAGGGCCGCTT**ACTA**TATAT**ACCTAA**CGCAGCCGCGGTGACAGTGGGCCT-5′

Proximal 6 duplex (54 bps):
5′-GATCTCCCGGCGAA**TGAT**ATATAT**AGGATT**GCGTCGGCGCCACTGTCACCCGGA-3′
3′-CTAGAGGGCCGCTT**ACTA**TATATA**TCCTAA**CGCAGCCGCGGTGACAGTGGGCCT-5′

Proximal 7 duplex (55 bps):
5′-GATCTCCCGGCGAA**TGAT**ATATATA**TGGATT**GCGTCGGCGCCACTGTCACCCGGA-3
3′-CTAGAGGGCCGCTT**ACTA**TATATAT**ACCTAA**CGCAGCCGCGGTGACAGTGGGCCT-5

Figure F.6 List of the proximal series of DNA duplexes used in gel shift experiments.

Distal -2 duplex (46 bps):
5′-GATCTCCCGGCGAA**TGATTCCT**GCGTCGGCGCCACTGTCACCCGGA-3′
3′-CTAGAGGGCCGCTT**ACTAAGGA**CGCAGCCGCGGTGACAGTGGGCCT-5′

Distal -1 duplex (47 bps):
5′-GATCTCCCGGCGAA**TGATATCCT**GCGTCGGCGCCACTGTCACCCGGA-3′
3′-CTAGAGGGCCGCTT**ACTATAGGA**CGCAGCCGCGGTGACAGTGGGCCT-5′

Distal 0 duplex (48 bps):
5′-GATCTCCCGGCGAA**TGATTATCCT**GCGTCGGCGCCACTGTCACCCGGA-3′
3′-CTAGAGGGCCGCTT**ACTAATAGGA**CGCAGCCGCGGTGACAGTGGGCCT-5′

Distal 1 duplex (49 bps):
5′-GATCTCCCGGCGAA**TGAT**A**TATCCT**GCGTCGGCGCCACTGTCACCCGGA-3′
3′-CTAGAGGGCCGCTT**ACTA**T**ATAGGA**CGCAGCCGCGGTGACAGTGGGCCT-5′

Distal 2 duplex (50 bps):
5′-GATCTCCCGGCGAA**TGAT**AT**AATCCT**GCGTCGGCGCCACTGTCACCCGGA-3′
3′-CTAGAGGGCCGCTT**ACTA**TA**TTAGGA**CGCAGCCGCGGTGACAGTGGGCCT-5′

Figure F.7 List of the distal series of DNA duplexes used in gel shift experiments.