Targeting DNA repeat sequences with Py-Im polyamides

Thesis by

JeenJoo Sophia Kang

In Partial Fulfillment of the Requirements

for the degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2015

(Defended March 27, 2015)

Dedicated to my family

Acknowledgements

I learned a great deal in graduate school, and a good measure of it is from the people I have interacted with over the years. Many people on- and off- campus that have positively influenced my years at Caltech. I would like to acknowledge at least some of them here.

I would like to thank my advisor *Prof. Peter Dervan* for his scientific guidance during my graduate studies. The intellectual freedom that he promotes in the lab, balanced by his advice on risk, has taught me to explore different areas of research in a considered manner. He has been supportive of, and appreciated the joy of, pursuing self-generated ideas even as I kept going "off the ranch". I expect the scientific judgment and understanding of new technologies that I have gained during my graduate education will prove essential to my future career. I would also like to thank my graduating committee chaired by *Prof. Bob Grubbs* and joined by *Prof. Jim Heath* and *Prof. Linda Hsieh-Wilson*. Their advice and support has been indispensible to my development in graduate school. Beyond science, the genuine interest of professors in the welfare of students makes Caltech an exceptional place.

My education was made much richer by my interactions with my labmates in the Dervan lab. In my earlier years, I learned a great deal from *Katy Muzikar*, *Jim Puckett*, and *Jordan Meier* who took the time to transfer their knowledge of some very time-intensive techniques. With all my colleagues over the years, I have appreciated the discussions of science, academia, culture, and our futures held over drinks, meals, and

late nights in the lab. My fellow graduate students: *Alexis Kurmis*, *Ben Li*, *Thomas Martinez*, *Dave Montgomery*, *Jerzy Szabowski*, *Fei Yang*; and post-doctoral colleagues: *Alissa Hare*, *Amanda Hargrove*, *Jevgenij Raskatov*, *Dan Gubler*, *Amanda Silberstein*, *Jamie Wang*, *Sam Weisbrod*, and *Tim Welch* were wonderful labmates. We are all so different, yet the work of this lab brought us all together for a short while— there is something special in that, and I look forward to when we might cross paths again.

My work would not have progressed without the help of *Igor Antoshechkin* in the Sequencing facility; and *Shelley Diamond* and *Diana Perez* in the Flow cytometry facility. They provided not only scientific help, but also moral support in the face of difficulties. Similarly, *Vicky Brennan*, *Joe Drew*, and *Agnes Tong* fix everything.

Thank you to all my friends, but in particular those in chemistry and the Korean (sometimes-) Tennis club. You have injected fun and relief into these years; my time spent with you defines a great deal of my Caltech experience. Some of our memories— grillin'n'chillin, "stop it, Myles", "Leonard", poker nights, "yeah, we're three people", nyan cat, tennis, "definitely a halibut"— I hope to build on over a lifetime of friendship.

Lastly, my family—*Mom*, *Dad*, and *Alex*, and *my extended family*… without you guys, I probably never would have started and I certainly couldn't have finished. It is with your love and support that I have the confidence to try more things than I fail at. There is nothing I can say to fully express my gratitude, so, I'll just say I love you.

Abstract

Hairpin pyrrole-imdazole polyamides are cell-permeable, sequence-programmable oligomers that bind in the minor groove of DNA. This thesis describes studies of Py-Im polyamides targeted to biologically important DNA repeat sequences for the purpose of modulating disease states. Design of a hairpin polyamide that binds the CG dyad, a site of DNA methylation that can become dysregulated in cancer, is described. We report the synthesis of a DNA methylation antagonist, its sequence specificity and affinity informed by Bind-n-Seq and iteratively designed, which improves inhibitory activity in a cell-free assay by 1000-fold to low nanomolar IC50. Additionally, a hairpin polyamide targeted to the telomeric sequence is found to trigger a slow necrotic-type cell death with the release of inflammatory molecules in a model of B cell lymphoma. The effects of the polyamide are unique in this class of oligomers; its effects are characterized and a functional assay of phagocytosis by macrophages is described. Additionally, hairpin polyamides targeted to pathologically expanded CTG•CAG triplet repeat DNA sequences, the molecular cause of myotonic dystrophy type 1, are synthesized and assessed for toxicity. Lastly, ChIP-seq of Hypoxia-Inducible Factor is performed under hypoxia-induced conditions. The study results show that ChIP-seq can be employed to study the genome-wide perturbation of Hypoxia-Inducible Factor occupancy by a Py-Im polyamide.

Table of Contents

List of Figures and Tables

Chapter 1

Introduction

1.1 Complexity coded in DNA

The blueprint for living organisms encoded in deoxyribose nucleic acid (DNA) determines every possible transcriptomic and proteomic state for all the cells in an organism. According to the central dogma of molecular biology put forth in 1958, DNA is the foundational molecule from which ribonucleic acid (RNA) is transcribed, and protein is translated from the RNA transcript (Figure 1.1A).[1] The coordination and control of these molecules are critical to the spatiotemporal organization of cell activity and tissue maintenance in organismal health. In humans, the 3 billion letter genome is organized in 23 chromosomes and encodes over 20,000 protein-coding genes.[2] In addition, over 15,000 long non-coding RNA transcripts have been identified and are thought to play a regulatory role.[3] The structural organization of DNA and its gene expression profile is specific to each cell-type and is accomplished by layers of exquisitely controlled processes. The etiology of many diseases can be found in misregulated DNA. Modulation by chemical means of DNA regulation may yield strategies for therapeutic benefit.

1.2 Epigenetic Organization of DNA

Eukaryotic DNA is condensed and regulated as chromatin to expose appropriate genes for transcription and maintain genome integrity in the face of a genotoxic environment (Figure 1.1A). The fundamental unit of organization is the nucleosome core particle, a histone octamer protein assembly wrapped by 147 base pairs of DNA (Figure 1.1B).[4] The wrapped DNA can be compacted densely into heterochromatin, which is less

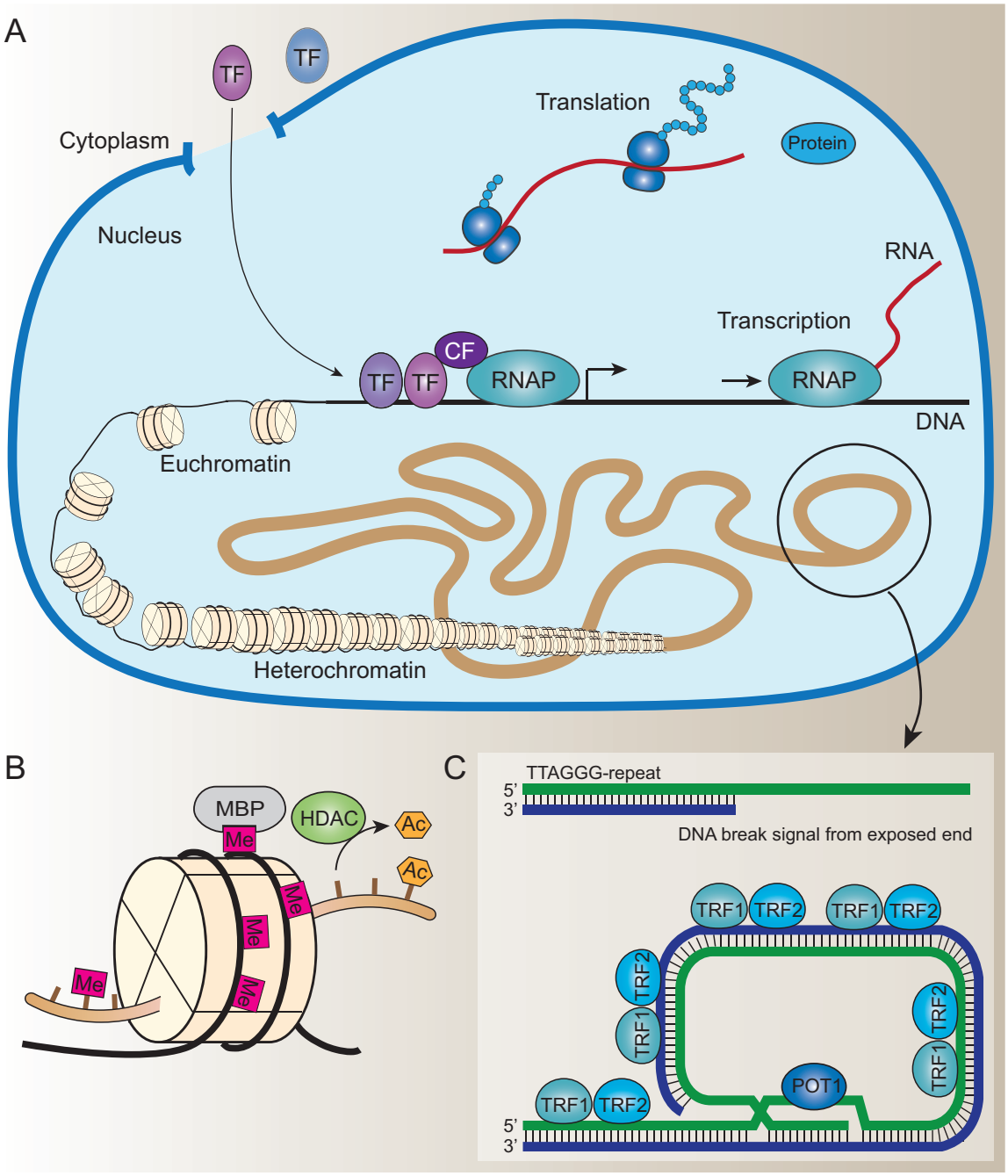


Figure 1.1 Structural organization of DNA in the cell. A) DNA (in black) is wrapped around the nucleosome core particle and organized as compact heterochromatin or transcriptionally active euchromatin. Transcription factors (TF) bind DNA and recruit cofactors (CF) and RNA polymerase II (RNAP) to transcribe mRNA from DNA. mRNA is translated into protein by ribosomes. B) The nucleosome core particle is comprised 147 base pairs of DNA wrapped around a histone octamer. Epigenetic modifications, including methylation (Me), acetylation (Ac), ubiquitiniylation, and SUMO-lation on the tails of histones govern chromatin structure. Methylated DNA is recognized by methyl-binding proteins (MBP) which recruit histone deacetylases (HDAC) to modify chromatin into a transcriptionally inactive state. C) Chromosomes end in the telomeric 5'-TTAGGG-3' hexamer repeat sequence, which can trigger a DNA damage response due to its similarity to a DNA double-stranded break. Telomeres form a

loop structure with invasion by the 5' strand to conceal the DNA end. TRF1 and TRF2 bind the telomere sequence and POT1 stabilizes the single-stranded DNA.

accessible for transcription and the requisite transcription machinery. Alternatively it can be packed more sparsely into euchromatin, which is a more transcriptionally active structure. The regulation of chromatin structure occurs by chemical annotation of the histones and DNA at the associated loci.[5] Histones can be methylated, acetylated, ubiquitinylated, phosphorylated, and SUMO-lated to modulate transcription or to respond to genotoxic stress. Proteins bound to enhancers modify, or recruit other enzymes that modify, the proximal histones that determine the transcriptional landscape of the cells. These epigenetic modifications are heritable and strongly correlate with the transcriptional programs that determine cell fate.[6] The key element of the epigenetic code on mammalian DNA is the methylation of the 5' carbon of cytosines in the palindromic CG dyad. This two base pair motif is concentrated in sequences called CpG islands, often associated with promoter sites.[7] The methylation of these CpG islands is transcriptionally repressive; methylated cytosines are bound by methyl binding proteins which recruit histone modifying enzymes to annotate the proximal histones into transcriptionally repressive heterochromatin (Figure 1.1B).[8] In this manner, DNA is purposefully structured and regulated for function.

In this organization of DNA, chromosome ends present a danger of being detected as DNA double-stranded breaks. Histones are sensitive to DNA damage and become phosphorylated upon detection of double stranded breaks, signaling towards repair or cell death.[9] To resolve the "end-protection problem", nature provides a structural solution.[10] The telomere is comprised of 6- base pair repeats of 5'-TTAGGG-3' spanning 5-30

kilobases with a 5' single strand overhang.[10] The repeating overhang is looped back to invade the double-stranded repeat region (Figure 1.1C).[11] The sequence repetition ensures complementarity and the end is hidden in the loop. Proteins including TRF1 and TRF2 of the shelterin complex, as well as POT1, which binds the single stranded portion, clamp this structure together.[11] The maintenance of this structure is important to maintain genome integrity, as its disruption is known to cause genotoxic stress signals and deleterious recombination.[12]

1.3 Transcriptional Regulation

In the context of this regulated DNA structure, transcription is enabled when transcription factor proteins bind enhancer regions coded in the DNA sequence. Transcription factors are proteins that bind DNA in a sequence-specific manner and function to either activate or repress transcription.[13] The enhancers of genes under a transcription factor's control contain the conserved cognate binding sequence for the protein. Transcription factors achieve specificity through their DNA binding domains (DBD). There are several conserved structural motifs for DNA binding and they include basic-helix-loop-helix, zinc finger, and high mobility group box DBDs (Figure 1.2).[14-16] These DBDs are linked to protein domains that transactivate transcription.

In mammals, these transcription factors do not bind in isolation but bind cooperatively to recruit cofactors and the RNA Polymerase II machinery to the transcription start site. The assemblage of multiple factors integrates various signaling

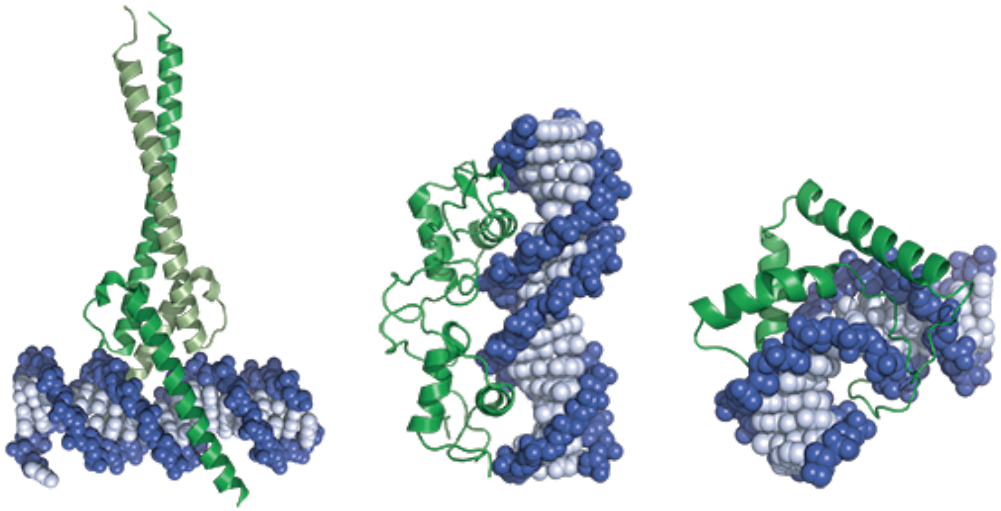


Figure 1.2 Crystal structures of transcription factors bound to DNA. DNA binding domains of transcription factors fall into several families, including basic-helix-loop-helix (Myc/Max, PDB 1NKP), zinc-finger (glucocorticoid receptor, PDB 1R40), and high mobility group (Lef-1, PDB 2LEF).[55]

cascades for precise detection of environmental cues.[13] Further, the expanded footprint of the bound proteins provides gene specificity. The interferon β (IFNB) gene enhancer is an example of the coordinate binding of multiple transcription factors for transcriptional activation. The DNA serves as a sequence-encoded scaffold for the binding of ATF-2, c-Jun, IRF-3A, IRF-7B, IRF-3C, IRF-7D, p50, and RelA at the enhancer of IFNB (Figure 1.3).[13] These factors in turn recruit the coactivators CBP/p300 in a multivalent fashion. The composite enhanceosome modifies chromatin and recruits RNA polymerase II to transcribe the IFNB gene into messenger RNA. Combining all these elements encoded in both the sequence and structure of DNA achieves exquisite control of the transcriptomic program.

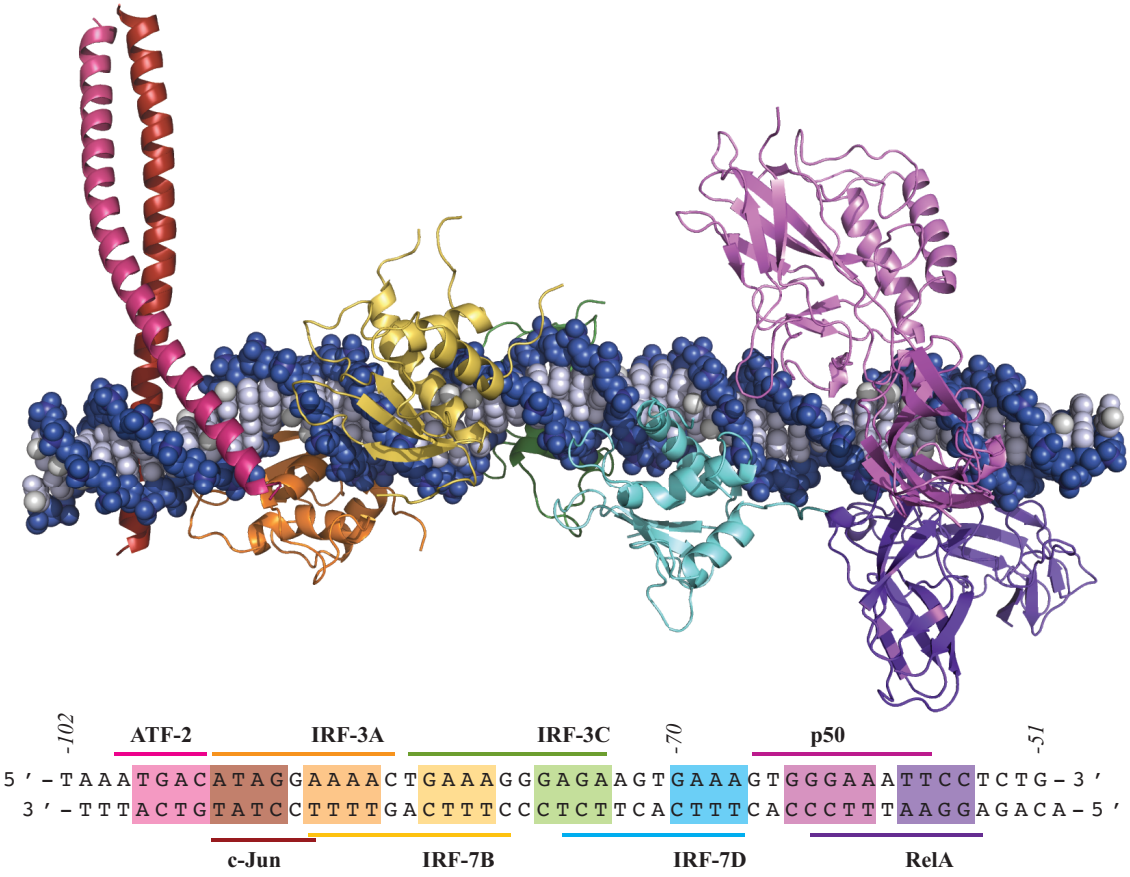


Figure 1.3 Model of the interferon β enhanceosome. [17]

1.4 Molecular recognition of DNA

The sequence-specific molecular recognition of DNA is key to this regulatory system. DNA is comprised of four nucleotides linked by anti-parallel phosphodiester backbones in a double helix (Figure 1.4). In the B-form DNA typically found in nature, the four nucleotides hydrogen-bond by the Watson-Crick pairings: adenosine (A) pairs with thymine (T) through two hydrogen bond interactions; cytosine (C) pairs with

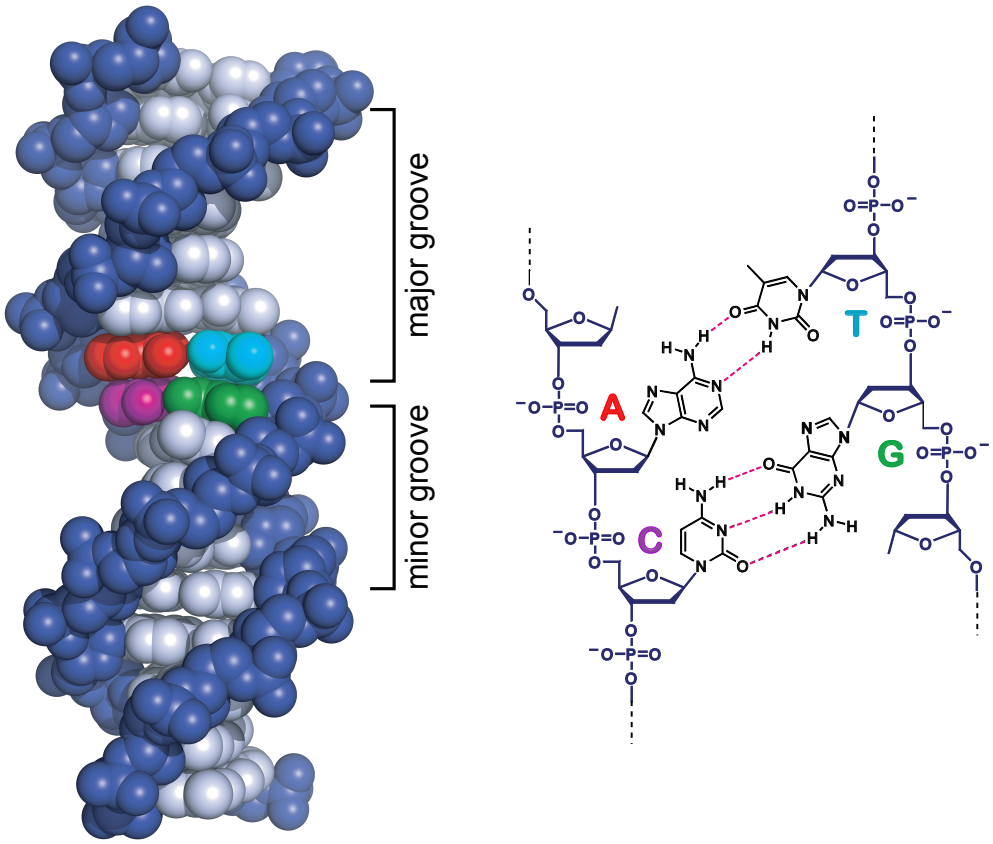


Figure 1.4 Structure of deoxyribose nucleic acid. DNA comprised of four nucleic acid bases linked by a phophodiester backbone in a double helix (PDB 1BNA) is shown.[59]

guanosine (G) through 3 hydrogen bonding interactions.[18] This specific pairing pattern provides for heritability of the genetic code as identical daughter genomes are synthesized during replication following the base complementarity. The exposed edges of the base pair form a wide major groove and a narrow minor groove on opposite sides of the double helix (Figure 1.4). Each of the four base pairs exposes a unique stereo-electronic edge in the grooves (Figure 1.5A). These sequence-specific stereo-electronics are recognized by DNA binding proteins. In addition, they are recognized by small molecule ligands which may be used to modulate the binding activity of the proteins.

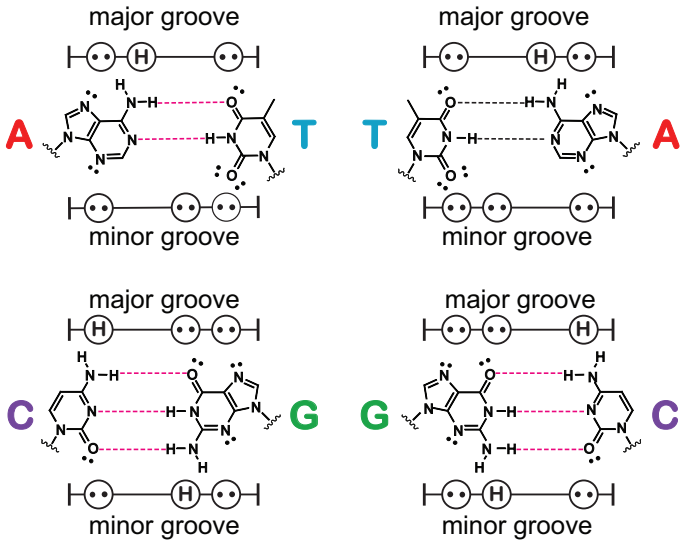


Figure 1.5 Each base pair has a unique stereo-electronic edge in both the major and minor groove.[55] Hydrogen-bonding patterns are shown for both the major and minor groove for each of the base pairs. Circles with two dots represent lone pairs and circles with an H represent a hydrogen on the exocyclic amine of guanine.

There are DNA-binding natural products with known sequence specificity that have been successfully employed as therapeutic drugs and scientific tools. Actinomycin D is a DNA-intercalator with a sequence preference for 5'-GC-3' that was one of the earliest chemotherapeutics for cancer (Figure 1.6).[19] Echinomycin is a bis-intercalating minor groove binding ligand that preferentially binds 5'-(A/T)CGT-3' and was found to downregulate hypoxia inducible factor transcriptional activity (Figure 1.6).[20-22] Fluorescent DNA-binding ligands such as DAPI have seen wide use as nuclear stains in microscopy (Figure 1.6).[23,24] Among DNA minor groove-binding ligands, distamycin and netropsin lend themselves to rational design and chemical modification due to their modular structure (Figure 1.6).[25]

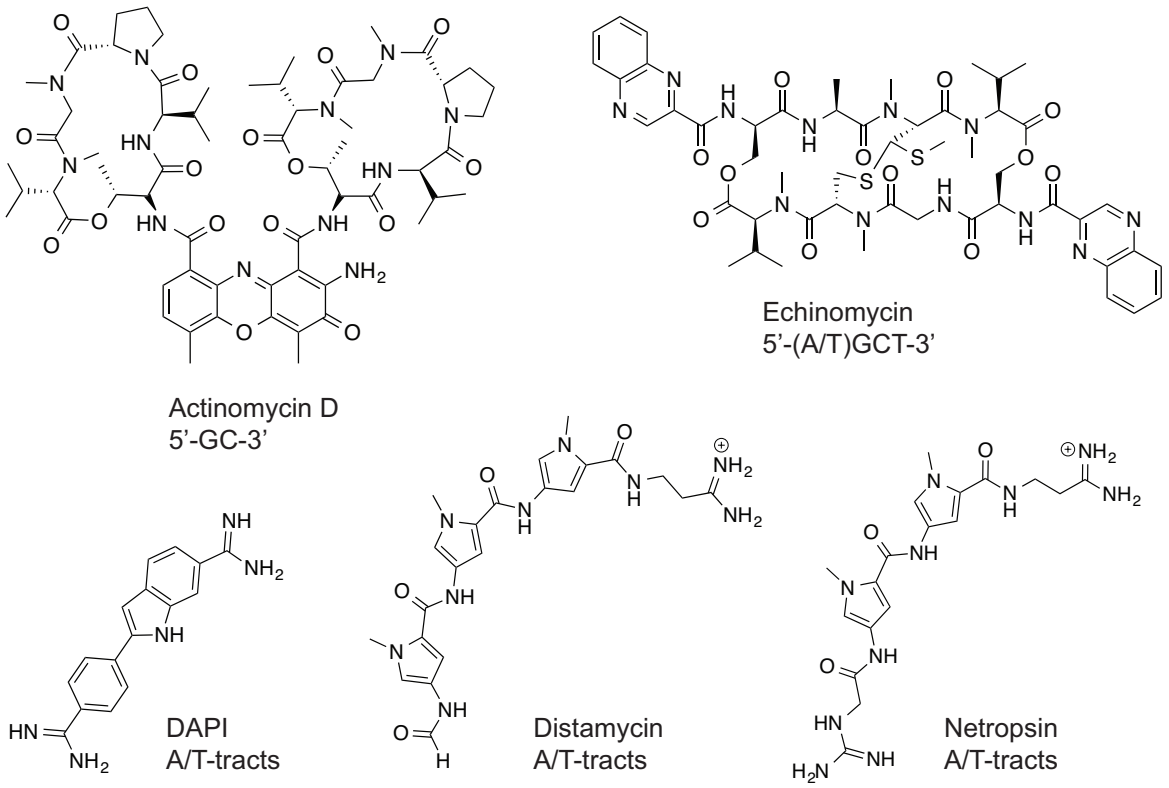


Figure 1.6 Structures of DNA binding small molecule ligands and their preferred binding sequences.

The minor groove surface of DNA presents a binding target for effecting modulation of dysfunctional biology in disease states. Distamycin preferentially binds AT-tracts, as there is steric hindrance from the exocyclic amine of guanosines in C•G/G•C on the minor groove floor (Figure 1.7A).[26] The introduction of a N-methyl imidazole ring (Im) to replace a N-methyl pyrrole ring (Py) provides a "hole" for the "bump" of the guanine residue.[27,28] In particular, the 2:1 binding of distamycin in the minor groove of DNA[29] (Figure 1.7B) suggested that linked Py-Im polyamides can target sequences through cofacial arrangement of the aromatic ring pairs to distinguish the edges of the four Watson-Crick base pairs.[30] Pairing rules for programmable specificity have been determined: Im/Py specifies a G•C base pair, Hp/Py codes for T•A base pairs

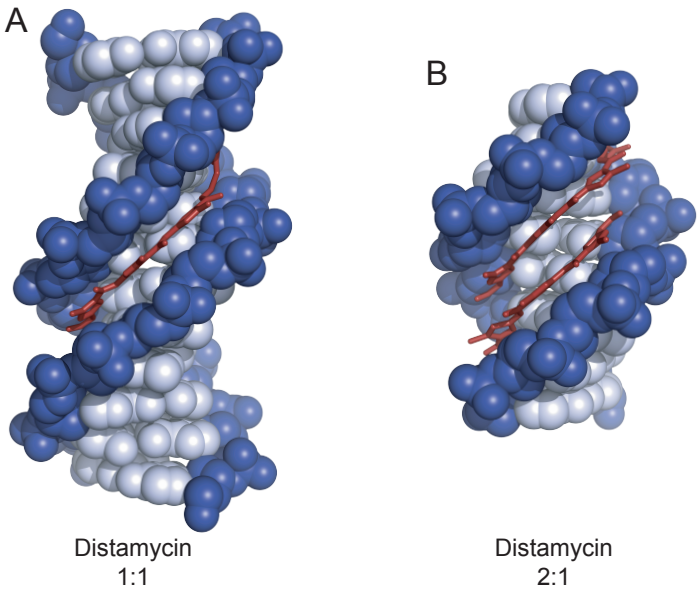


Figure 1.7 Crystal structures of distamycin bound in the minor groove of DNA. A) Distamycin bound 1:1 (PDB 2DND) and B) 2:1 in an antiparallel orientation (PDB 378D).[59]

and Py/Py binds both T•A/A•T (Figure 1.8).[27,28,31-34] This pairing register of the antiparallel polyamide strands can be covalently enforced with a γ-amino butyric acid linker to form a hairpin architecture.[35] However, Py-Im polyamide strands, particularly those containing many imidazoles, were found to be over-curved compared to natural DNA.[36] β-alanine residues were introduced to relieve the curvature in such cases, which allowed targeting longer sequences and sequences with more C•G/G•C residues.[37-41] The development of this molecular recognition technology has enabled sequence-specific targeting of DNA with affinities similar to that of transcription factors.[42]

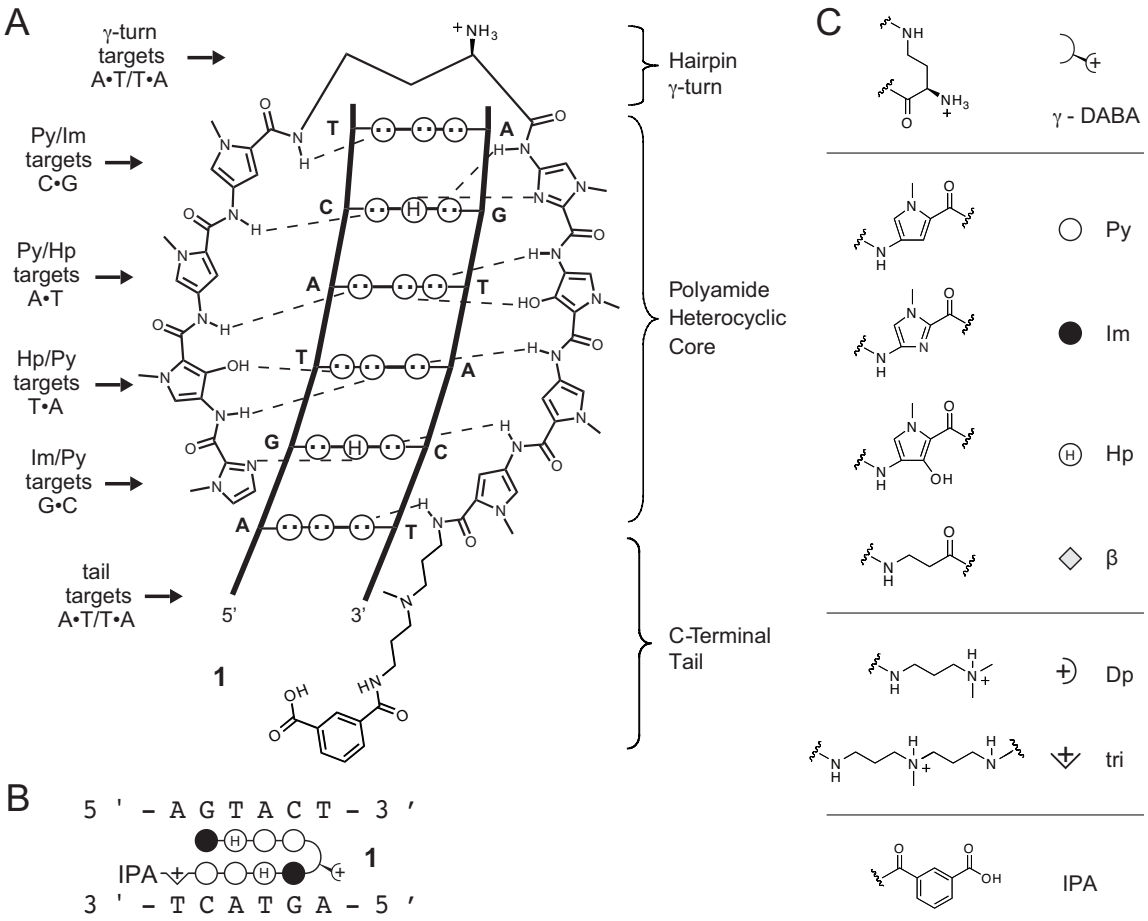


Figure 1.8 Pairing rules for Py-Im polyamide DNA sequence-specificity. A) Hydrogen-bonding pattern of the minor groove is depicted as described above. Model for the binding of a ImHpPyPy-γ-DABA-ImHpPyPy-tri-IPA polyamide bound to 5'-AGTACT-3'. Hydrogen bonds shown with dashed lines. B) Ball and stick notation for polyamides, with legend in C).

1.5 Biological modulation with Py-Im polyamides

Studies support Py-Im polyamides effecting biological modulation through its DNA binding capacity. The accessibility of DNA in the context of chromatin for Py-Im polyamide binding was demonstrated in a crystal structure of a polyamide bound to the nucleosome core particle (Figure 1.9A). In live cell culture experiments, the cell-permeability and nuclear localization of dye-conjugated polyamides were observed

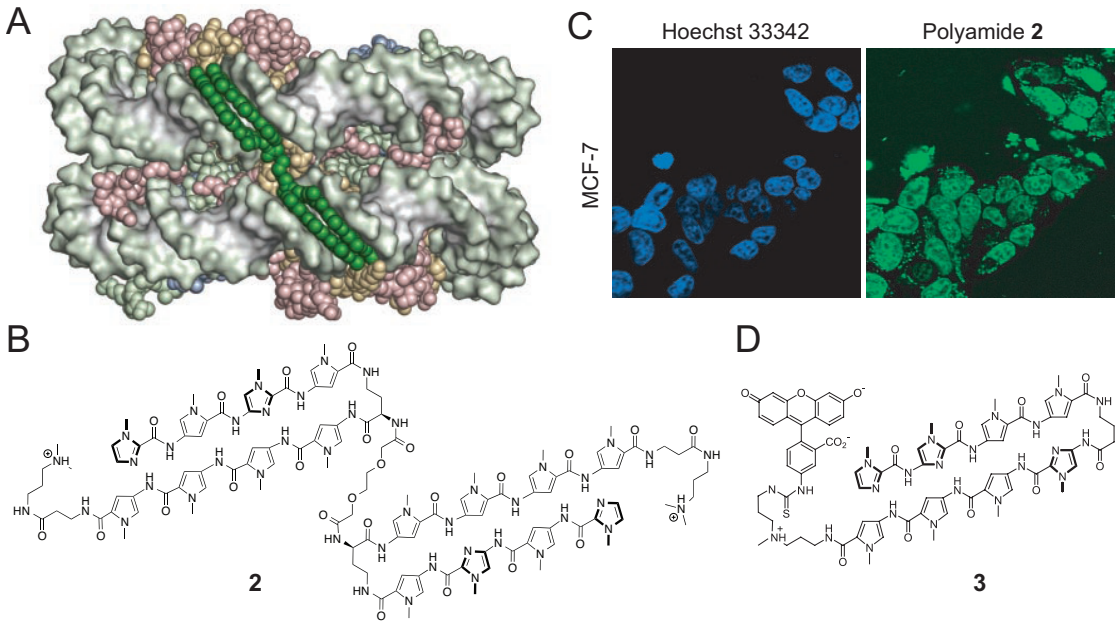


Figure 1.9 Py-Im polyamides access chromatin. A) Crystal structure of a nucleosome core particle with B) polyamide dimer **1** bound in the adjacent minor grooves.[60] C) Confocal microscopy of MCF-7 breast cancer cell line treated with nuclear-stain Hoechst or D) dye-conjugate polyamide **2**.[61] Images show co-localization to the nucleus.

through confocal microscopy in a wide range of cell lines (Figure 1.9C).[43-44] This demonstrated that access to chromatin with no external transfection was possible with these oligomers. Transcriptional modulation was demonstrated in cell culture by Py-Im polyamides targeted to the respective consensus binding sequences of the transcription factors hypoxia-inducible factor, androgen receptor, glucocorticoid receptor, and NF-κB.[46-50] An X-ray crystal structure of a cyclic polyamide bound to DNA provided a rationale for the exclusion of major-groove binding proteins upon minor groove binding by the polyamide.[51] The crystal structure revealed that polyamide binding caused an expansion of the minor groove and compression of the major groove (Figure 1.10). Py-Im polyamides were also found to disrupt processive DNA enzymes. For example, a high

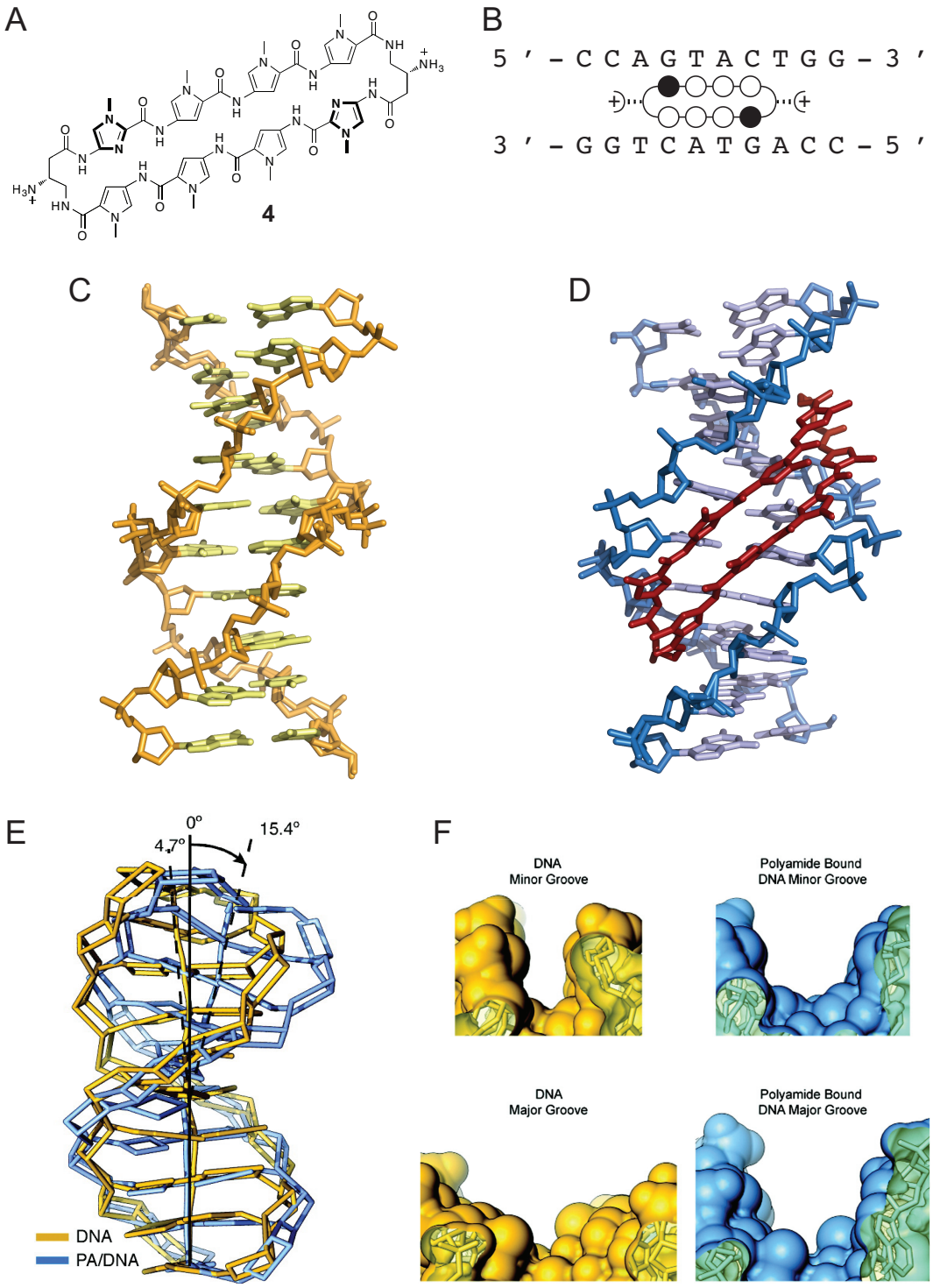


Figure 1.10 Allosteric disruption of major groove by minor groove binding cyclic polyamide.[51] A) Structure of cyclic polyamide **4**. B) Ball-and-stick notation for polyamide **4** shown with the sequence of the DNA oligonucleotide. C) Crystal structure of native B-form DNA (PDB 1D8G). D) Crystal structure of cyclic polyamide **4** bound in the minor groove of DNA (PDB 3OMJ). E) Binding of polyamide **4** causes a distortion to the native DNA structure. F) The binding of polyamide 4 causes a widening of the minor

groove (top), and a compression of the major groove (bottom). Native structure shown in yellow, polyamide-bound structure shown in blue.

affinity hairpin targeted to the androgen response element also disrupted RNA polymerase II elongation, causing cytotoxicity and p53 activation without DNA damage.[52] Another study showed that high doses of this polyamide can disrupt helicase activity and can cause replicative stress.[53] These results indicate DNA minor groove binding Py-Im polyamides can cause a variety of alterations to biological activity.

Studies in mouse models have shown that many of these effects observed in cell culture translate to the mouse with systemic treatment of the polyamide. Pharmacokinetics studies have shown that Py-Im polyamides are bioavailable and have reasonable exposure.[54] Cyclic architecture and acetylation of the turn of the same core aromatic oligomer sequence have each been explored and shown to have different pharmacokinetic and toxicity profiles.[55,56] Gene expression has been measured in the tumor of mouse xenograft models and shown to be modulated by polyamide treatment similarly to the experiments done in cell culture.[57] In a xenograft model of prostate cancer, the high-affinity Py-Im polyamide targeted to the androgen receptor response element reduced tumor size.[52] Indeed, a C-14 labeled analog of the molecule showed preferential localization to the tumor xenograft.[58] In sum, Py-Im polyamides are a class of molecules well-suited for the study of biological perturbation through DNA binding.

1.6 Scope of Work

This thesis describes studies of Py-Im polyamides targeted to biologically important DNA repeat sequences for the purpose of modulating disease states. In chapter

2, a Py-Im polyamide is designed to bind a sequence that contains two CG dyads and is tested as a DNA methylation antagonist. Chapter 3 explores the immunogenic signaling triggered by a polyamide that is targeted to the telomere repeat sequence. Chapter 4 describes work in assessing the animal toxicity of hairpin polyamides targeted to the CAG/CTG triplet repeat expansion to disrupt the transcription of the associated RNA, which is the molecular cause of myotonic dystrophy type 1. Chapter 5 describes a ChIP-seq experiment that explores the genome-wide perturbation of hypoxia-inducible factor occupancy induced under hypoxia by a Py-Im polyamide targeted to the consensus sequence of the hypoxia response element.

1.7 References

(1) Crick, F. H. *Symposia. of the. Society for Experimental Biology.* **1958**, *12*, 138.

(2) Harrow, J.; Frankish, A.; Gonzalez, J. M.; Tapanari, E.; Diekhans, M.; Kokocinski, F.; Aken, B. L.; Barrell, D.; Zadissa, A.; Searle, S.; Barnes, I.; Bignell, A.; Boychenko, V.; Hunt, T.; Kay, M.; Mukherjee, G.; Rajan, J.; Despacio-Reyes, G.; Saunders, G.; Steward, C.; Harte, R.; Lin, M.; Howald, C.; Tanzer, A.; Derrien, T.; Chrast, J.; Walters, N.; Balasubramanian, S.; Pei, B.; Tress, M.; Rodriguez, J. M.; Ezkurdia, I.; van Baren, J.; Brent, M.; Haussler, D.; Kellis, M.; Valencia, A.; Reymond, A.; Gerstein, M.; Guigó, R.; Hubbard, T. J. *Genome Res.* **2012**, *22*, 1760-1774.

(3) Derrien, T.; Johnson, R.; Bussotti, G.; Tanzer, A.; Djebali, S.; Tilgner, H.; Guernec, G.; Martin, D.; Merkel, A.; Knowles, D. G.; Lagarde, J.; Veeravalli, L.; Ruan, X.; Ruan, Y.; Lassmann, T.; Carninci, P.; Brown, J. B.; Lipovich, L.; Gonzalez, J. M.; Thomas, M.; Davis, C. A.; Shiekhattar, R.; Gingeras, T. R.; Hubbard, T. J.; Notredame, C.; Harrow, J.; Guigó, R. *Genome Res.* **2012**, *22*, 1775-1789.

(4) Kornberg, R. D. *Science.* **1974**, *184*, 868-871.

(5) Berger, S. L. *Nature.* **2007**, *447*, 407-412.

(6) Heintzman, N. D.; Hon, G. C.; Hawkins, R. D.; Kheradpour, P.; Stark, A.; Harp, L. F.; Ye, Z.; Lee, L. K.; Stuart, R. K.; Ching, C. W.; Ching, K. A.; Antosiewicz-Bourget, J. E.; Liu, H.; Zhang, X.; Green, R. D.; Lobanenkov, V. V.; Stewart, R.; Thomson, J. A.; Crawford, G. E.; Kellis, M.; Ren, B. *Nature.* **2009**, *459*, 108-112.

(7) Bird, A. P. *Nature.* **1986**, *321*, 209-213.

(8) Nan, X.; Ng, H. H.; Johnson, C. A.; Laherty, C. D.; Turner, B. M.; Eisenman, R. N.; Bird, A. *Nature.* **1998**, *393*, 386-389.

(9) Mah, L. J.; El-Osta, A.; Karagiannis, T. C. *Leukemia.* **2010**, *24*, 679-686.

(10) de Lange, T. *Cold Sprin. Harb. Symp. Quant. Biol.* **2010**, *75*, 167-177.

(11) de Lange, T. *Science.* **2009**, *326*, 948-952.

(12) Sfeir, A.; de Lange, T. *Science.* **2012**, *336*, 593-597.

(13) Tjian, R.; Maniatis, T. *Cell.* **1994**, *77*, 5-8.

(14) Williams, D. C.; Cai, M.; and Clore, G. M. *J. Biol. Chem.* **2004**, *279*, 1449.

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

(16) Love, J. J.; Li, X.; Case, D. A.; Giese, K.; Grosschedl, R.; Wright, P. E. *Nature* **1995**, *376*, 791.

(17) Panne, D. *Curr. Opin. Struct. Biol.* **2008**, *18*, 236-242.

(18) Watson, J. D.; Crick, F. H. *Nature.* **1953**, *171*, 737-738.

(19) Kamitori, S.; Takusagawa, F. *J. Mol. Biol.* **1992**, *225*, 445-456.

(20) Waring; M., J.; Wakelin; G., L. P.; Waring, M. J.; Wakelin, L. P. G. *Nature.* **1974**, *252*, 653-657.

(21) VanDyke, M. M.; Dervan, P. B. *Science.* **1984**, *225*, 1122-1127.

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

(23) Larsen, T. A.; Goodsell, D. S.; Cascio, D.; Grzeskowiak, K.; Dickerson, R. E. *J Biomol. Struct. Dyn.* **1989**, *7*, 477-491.

(24) Otto, F. *Methods Cell. Bio.* **1990**, *33*, 105-110.

(25) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *Proc. Natl. Acad. Sci. U. S. A.* **1985**, *82*, 1376-1380.

(26) Pelton, J. G.; Wemmer, D. E. *Biochemistry.* **1988**, *27*, 8088-8096.

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

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

(29) Mitra, S. N.; Wahl, M. C.; Sundaralingam, M. *Acta Crystallogr. D. Biol. Crystallogr.* **1999**, *55*, 602-609.

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

(31) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature.* **1996**, *382*, 559-561.

(32) White, S.; Baird, E. E.; Dervan, P. B. *Chem. Biol.* **1997**, *4*, 569-578.

(33) White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Nature.* **1998**, *391*, 468-471.

(34) 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.

(35) Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 7983-7988.

(36) Kelly, J. J.; Baird, E. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 6981-6985.

(37) Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6160-6166.

(38) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *Chem- Eur. J.* **1997**, *3*, 1600-1607.

(39) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Ange. Chem. Int. Edit.* **1998**, *37*, 1421-1423.

(40) Turner, J. M.; Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 6219-6226.

(41) Wang, C. C.; Dervan, P. B. *J. Am. Chem. Soc.* **2001**, *123*, 8657-8661.

(42) 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.; Dervan, P. B. *Tetrahedron.* **2007**, *63*, 6146-6151.

(43) Dudouet; Burnett; Dickinson; Wood; Melander; Belitsky; Edelson; Wurtz; Briehn; Dervan; Gottesfeld *Chem. Biol.* **2003**, *10*, 859-867.

(44) 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.

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

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

(47) Nickols, N. G.; Jacobs, C. S.; Farkas, M. E.; Dervan, P. B. *ACS Chem. Biol.* **2007**, *2*, 561-571.

(48) Nickols, N. G.; Dervan, P. B. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 10418-10423.

(49) Muzikar, K. A.; Nickols, N. G.; Dervan, P. B. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 16598-16603.

(50) Raskatov, J. A.; Meier, J. L.; Puckett, J. W.; Yang, F.; Ramakrishnan, P.; Dervan, P. B. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 1023-1028.

(51) Chenoweth, D. M.; Dervan, P. B. *J. Am. Chem. Soc.* **2010**, *132*, 14521-14529.

(52) Yang, F.; Nickols, N. G.; Li, B. C.; Marinov, G. K.; Said, J. W.; Dervan, P. B. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 1863-1868.

(53) Martínez, T. F.; Phillips, J. W.; Karanja, K. K.; Polaczek, P.; Wang, C. -M.; Li, B. C.; Campbell, J. L.; Dervan, P. B. *Nucleic Acids Res.* **2015**, *42*, 11546-11559.

(54) Synold, T. W.; Xi, B.; Wu, J.; Yen, Y.; Li, B. C.; Yang, F.; Phillips, J. W.; Nickols, N. G.; Dervan, P. B. *Cancer Chemother. Pharmacol.* **2012**, *70*, 617-625.

(55) Raskatov, J. A.; Hargrove, A. E.; So, A. Y.; Dervan, P. B. *J. Am. Chem. Soc.* **2012**, *134*, 7995-7999.

(56) Yang, F.; Nickols, N. G.; Li, B. C.; Szablowski, J. O.; Hamilton, S. R.; Meier, J. L.; Wang, C. -M.; Dervan, P. B. *J. Med. Chem.* **2013**, *56*, 7449-7457.

(57) Nickols, N. G.; Szablowski, J. O.; Hargrove, A. E.; Li, B. C.; Raskatov, J. A.; Dervan, P. B. *Mol. Cancer. Ther.* **2013**, *12*, 675-684.

(58) Raskatov, J. A.; Puckett, J. W.; Dervan, P. B. *Bioorg. Med. Chem.* **2014**, *22*, 4371-4375.

(59) Muzikar, K. A. *Repression of DNA-binding-dependent glucocorticoid receptor-mediated gene expression*. Ph. D. Thesis, 2011. Figures used with permission.

(60) Edayathumangalam, R. S.; Weyermann, P.; Gottesfeld, J. M.; Dervan, P. B.; Luger, K. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 6864-6869.

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

Chapter 2

Design of Sequence-Specific DNA Binding Molecules for DNA Methyltransferase Inhibition

The text of this chapter is taken in part from a manuscript co-authored with Jordan L. Meier and Peter B. Dervan (California Institute of Technology).

(Kang, J. S.; Meier, J. L.; Dervan, P. B. *J. Am. Chem. Soc.* **2014**, *136*, 3687-3694.)

Abstract

The CpG dyad, an important genomic feature in DNA methylation and transcriptional regulation, is an attractive target for small molecules. To assess the utility of minor groove binding oligomers for CpG recognition, we screened a small library of hairpin pyrrole-imidazole polyamides targeting the sequence 5′-CGCG-3′ and assessed their sequence specificity using an unbiased next-generation sequencing assay. Our findings indicate that hairpin polyamide of sequence PyImβIm-γ-PyImβIm (**1**), previously identified as a high affinity 5′-CGCG-3′ binder, favors 5′-GCGC-3′ in an unanticipated reverse binding orientation. Replacement of one β alanine with Py to afford PyImPyIm-γ-PyImβIm (**3**) restores the preference for 5′-CGCG-3′ binding in a forward orientation. The minor groove binding hairpin **3** inhibits DNA methyltransferase activity in the major groove at its target site more effectively than **1**, providing a molecular basis for design of sequence-specific antagonists of CpG methylation.

2.1 Introduction

The role of epigenetic dysregulation in cancer has motivated interest in DNA methylation and methods for its modulation.[1,2] In mammals, DNA methylation occurs in the major groove of DNA at the 5′ position of both cytosine residues in the palindromic CG dyad (CpG). CpG's are rare in the genome and 70% methylated, with nearly all unmethylated CpG's clustered in G,C-rich regions called "CpG islands".[3] Approximately 60% of RNA Polymerase II transcribed human genes contain CpG islands,[4] and their methylation causes transcriptional repression.[5] In cancer, for example, otherwise functional tumor suppressor genes can be silenced by hypermethylation in their associated CpG island.[6] Importantly, inhibition of DNA methylation at tumor suppressor genes has been shown to reactivate apoptotic pathways and sensitize cancer cells to previously ineffective chemotherapy.[7,8]

The most effective demethylation agents are cytidine analogs such as 5-aza-deoxycytidine, which find limited use due to significant side effects.[1] These cytidine analogs are suicide inhibitors incorporated into DNA to form covalent methyltransferase-DNA adducts.[9] The methyltransferase is sequestered and unavailable to methylate CpG's, resulting in genome-wide demethylation. DNA binding molecules, such as the bis-intercalating natural product echinomycin,[10] can disrupt CpG methylation *in vitro* but have dose-limiting toxicities that have abrogated further clinical advancement.[11] While other CpG methylation inhibitors are under investigation,[12-14] none of these agents have demonstrated the ability to inhibit DNA methylation in a sequence-specific fashion.

Hairpin pyrrole-imidazole (Py-Im) polyamides are a class of sequence-specific oligomers that bind in the minor groove of DNA.[15-20] Programmable sequence preference is accomplished by side-by-side pairings of aromatic amino acids that distinguish the edges of the four Watson-Crick base pairs.[15-20] Referred as the pairing rules, Im/Py codes for G•C base pairs, Hp/Py codes for T•A base pairs, and Py/Py binds both T•A/A•T in preference to G•C/C•G. Eight-ring hairpin oligomers linked by a central aliphatic γ-aminobutyric acid unit have affinities for match sites with K_a ~10^8 to 10^{10} M^{-1}.[16,21] These binding energetics are comparable to natural transcription factors, and like natural DNA binding proteins, are sensitive to differences in the sequence-dependent microstructure of DNA. To relax the curvature of all ring hairpins, β alanine (β) can be substituted for Py-rings in some cases such that β/β pairs replace Py/Py for T•A/A•T specificity, and Im/β replaces Im/Py pairs in strategic locations while retaining specificity for G•C base pair.[22-26] Hairpin Py-Im polyamides usually bind with the N-to-C terminus aligned in the 5′-to-3′ direction of DNA, referred to as "forward orientation".[27] This modest forward binding preference can be enforced by substitution of the prochiral α position in the γ-turn, i.e., replacement of γ-aminobutyric acid by (R)-2,4-diaminobutyric acid.[28] Hairpin architectures containing β/β pairs and β/ring pairs have been found in some cases to prefer the N to C terminus aligned in a 3′-to-5′ direction of DNA.[29] While adhering to the pairing rules, this reverse hairpin orientation would bind a different DNA sequence. Recently we used massively parallel sequencing methods in conjunction with biotin-tagged hairpins, termed Bind-n-Seq, to scan genome-size DNA sequence space for

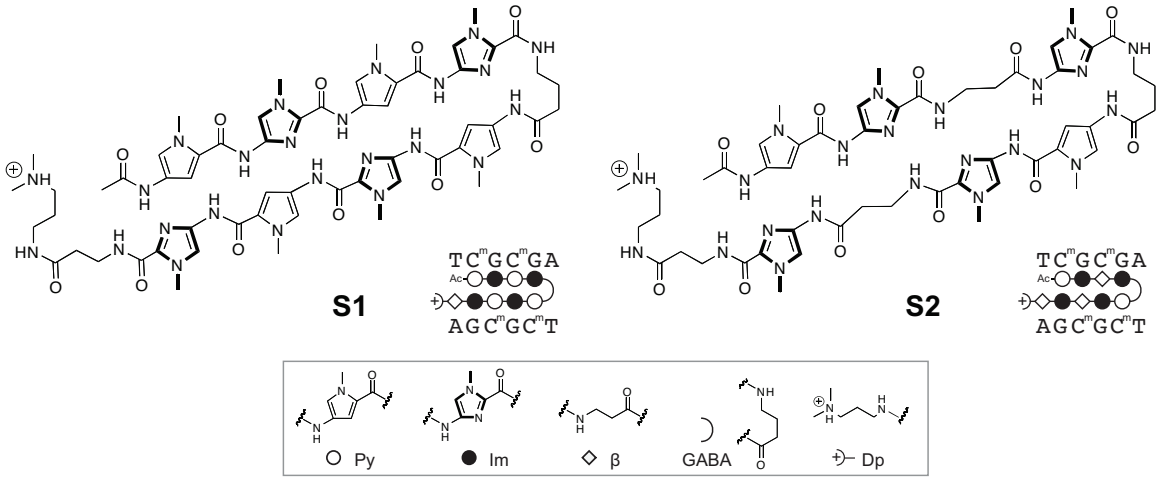


Figure 2.1 Structure of Py-Im polyamides **S1** and **S2** previously reported to bind methylated 5'-CGCG-3' oligonucleotide duplex.[32] Legend for ball-and-stick notation.

hairpin high affinity sites.[30] Although the canonical pairing rules are remarkably predictive of polyamide DNA binding specificity, we identified high affinity DNA binding sites in the reverse orientation for several polyamides containing β/Im pairs.[30]

Eight-ring hairpin Py-Im polyamides have been shown to discriminate 5′-GGGG-3′, 5′-GCGC-3′ and 5′-GGCC-3′ with appropriate arrangement of four Im/Py pairs.[31] From experience, sequences with CpG steps such as 5′-CGCG-3′ are not as readily accessed for reasons not well understood. In an effort to improve the affinity of an eight-ring hairpin polyamide for the sequence 5′-CGCG-3′, Sugiyama and coworkers replaced two Im/Py pairs with Im/β pairs (Figure 2.1). A change from PyImPyIm-γ-PyImPyIm (**S1**) to PyImβIm-γ-PyImβIm (**S2**) afforded a 65-fold increase in affinity for 5′-CGCG-3′.[32] Both hairpins conform to the pairing rules and would bind 5′-CGCG-3′ in the forward orientation. In this study, we employ a high-throughput sequencing assay of polyamide-DNA association to revisit targeting the 5′-CGCG-3′ sequence. Our findings indicate that hairpin polyamides of sequence PyImβIm-γ-PyImβIm **S2** favor 5′-GCGC-3′,

a reverse binding mode. The issue of designing a hairpin polyamide sequence that prefers 5′-CGCG-3′ to 5′-GCGC-3′ remains to be solved. Using Bind-n-Seq methods[30] as our screen for a library of polyamide-biotin conjugates, we find that replacement of one β alanine with Py to afford PyImPyIm-γ-PyImβIm restores the preference for forward binding 5′-CGCG-3′. Recent structural work has shown that a cyclic Py-Im polyamide binding in the minor groove causes significant widening of the minor groove width of DNA,[33,34] and provides a mechanistic rationale for disruption of DNA-binding proteins in the major groove. We demonstrate the ability of our 5′-CGCG-3′ specific minor groove binding hairpin polyamides to inhibit enzymatic CpG methylation in the major groove of a 5′-CGCG-3′ sequence.

2.2 Results

Sequence Based Analysis of PyImβIm-γ-PyImβIm Specificity. The 5′-CGCG-3′ sequence is a compelling DNA target for an 8-ring hairpin Py-Im polyamide because it is one of the least represented 6-bp sequence patterns in the human genome, potentially promoting greater genomic specificity.[30] Minoshima and coworkers have previously targeted this sequence and shown that polyamide **S2** (Figure 2.1) can bind the fully methylated sequence.[32] In their study, the substitution of two β's for Py moieties resulted in improved affinity for 5′-CGCG-3′ over the eight-aromatic ring architecture **S1** (Figure 2.1). In light of recent Bind-n-Seq studies, however, we wondered whether these changes may have also had the unintended effect of reducing the preference of the polyamide for binding in the forward orientation.[30] Bind-n-Seq is a high-throughput sequencing method

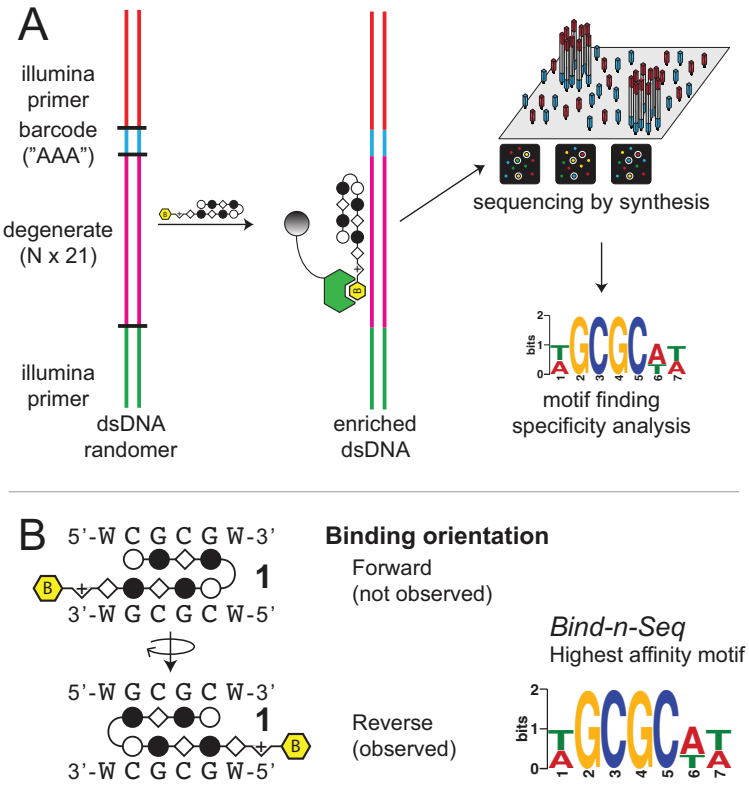


Figure 2.2 A) Scheme of Bind-n-Seq method.[30] Polyamide-biotin conjugate is incubated in a genome-sized library of all possible 21mers, enriched, sequenced, and the resulting dataset analyzed with motif-finding software.[30] B) Polyamide **1** could potentially bind in the forward orientation or the reverse orientation. The highest affinity binding sequence of **1** is the reverse orientation binding 5'-GCGC-3'.

that allows facile identification of high affinity binding sites of biotin-labeled Py-Im polyamides by affinity purification followed by sequencing (Figure 2.2A).[30] As a first step, we synthesized an analog of **S2** and examined polyamide-biotin conjugate **1** of sequence PyImβIm-γ-PyImβIm (Figure 2.2B), which has a biotin affinity tag appended at the C-terminus of the heterocyclic oligomer (Figure 2.3). Polyamide-biotin conjugate **1** was incubated at 50 nM in a library of all possible 21 base pair DNA sequences,

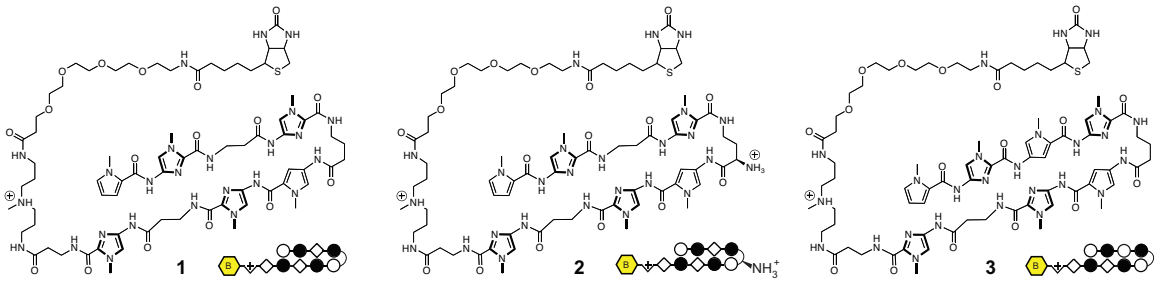
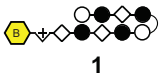
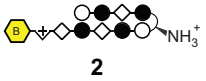
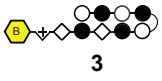
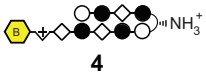
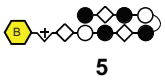
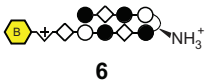
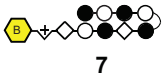
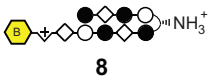
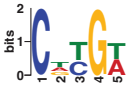


Figure 2.3 Structures of Py-Im polyamides **1**-**3**.

enriched, and sequenced to identify polyamide-bound sequences. This dataset was then analyzed by the DREME algorithm to construct a motif logo summarizing the highest affinity sequences. A binding preference for 5′-GCGC-3′ was revealed, suggestive of a reverse binding mode (Table 1).

Redesign Hairpin for CGCG versus GCGC Preference. In order to restore the preference for binding 5′-CGCG-3′ in the forward orientation, we considered two possible points of modification (Figure 2.4A). First, we made a single modification to **1** at the turn unit, replacing the GABA turn to a chiral γ-amino GABA, affording **2** (Figure 2.3 and 2.4B). The γ-amino GABA turn has previously been shown to restore forward orientation and increase affinity, including in β-containing polyamides.[28-30] This effect is thought to arise from a steric interaction with the floor of the minor groove when the chiral α-amino GABA turn unit is bound in the reverse orientation.[23,28] Assessment of polyamide **2** by Bind-n-Seq found that this modification improved the reverse/forward ratio but was insufficient to restore a forward orientation binding preference (Table 2.1). To confirm the high-throughput sequencing findings, we performed a thermal DNA

Table 2.1 The preferred binding orientations of polyamides **1-8** were queried with Bind-n-Seq to generate the highest affinity sequence motif. Polyamide-mediated thermal stabilization (ΔT_m) of 12 base pair oligonucleotides of the forward (5′-CGCG-3′) and reverse 5′-GCGC-3′ sequences were use to validate the revealed motifs. Melting temperatures reflect the mean and standard deviation of quadruplicate measurements.

Polyamide	Bind-n-Seq	5'-GGT **ACGCGT** ACC-3'		5'-GGT **AGCGCT** ACC-3'		Binding Orientation
		T_m / °C	ΔT_m / °C	T_m / °C	ΔT_m / °C	
 **1**		63.1 (±1.0)	3.4	71.4 (±0.2)	10.9	Reverse
 **2**		67.7 (±0.6)	7.9	69.4 (±0.3)	8.9	Reverse
 **3**		69.3 (±0.4)	9.6	64.3 (±0.2)	3.8	Forward
 **4**		67.9 (±0.0)	8.1	73.2 (±0.2)	12.7	Reverse
 **5**		70.3 (±0.0)	9.9	67.9 (±0.0)	7.7	Reverse
 **6**		68.1 (±0.7)	7.7	72.6 (±0.6)	12.4	Forward
 **7**		68.2 (±0.8)	7.8	66.0 (±0.2)	5.7	Reverse
 **8**		73.9 (±0.7)	13.5	71.1 (±0.6)	10.8	Reverse

denaturation study, as previous studies have shown the thermal stabilization (ΔT_m) of duplex DNA by Py-Im polyamides correlates well with binding affinity.[35] Assays were performed with DNA oligonucleotides differing only in the central binding sequence (5'-

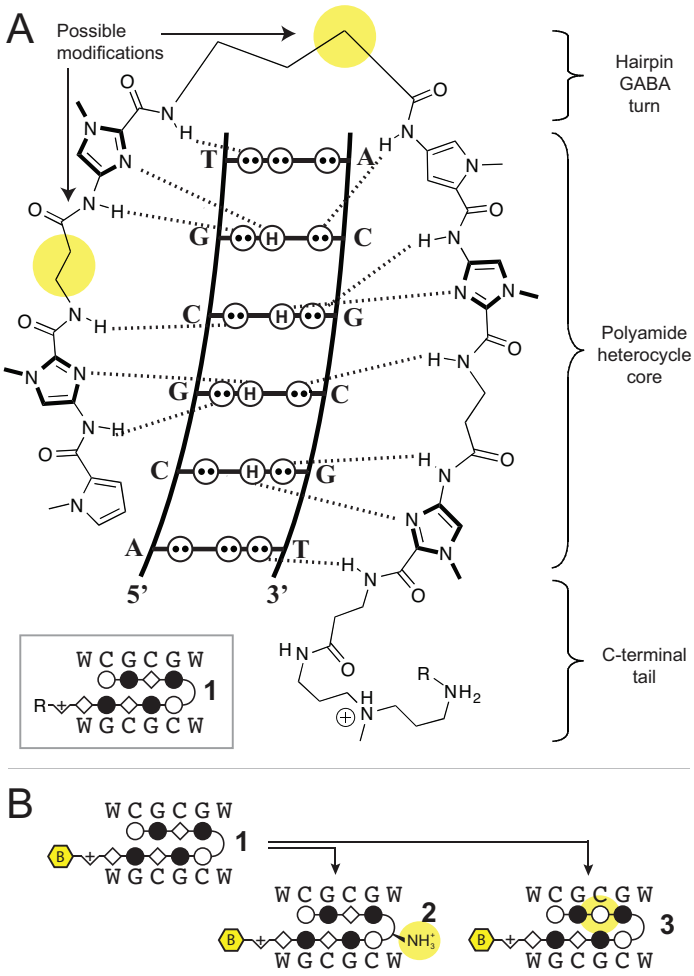


Figure 2.4 A) Scheme of Py-Im polyamide binding in the minor groove of DNA. B) Single position changes made to hairpin polyamide **1** to afford **2** and **3**. Positions are highlighted in yellow.

CGCG-3' versus 5'-GCGC-3') to directly test the binding orientations identified by the Bind-n-Seq logos (Table 2.1). This analysis substantiated a reverse orientation binding preference for polyamide **1**, with a ΔT_m of 10.9 ºC in the reverse direction as compared to 3.4 ºC in the forward direction. Modification at the turn to the γ-amino GABA in polyamide **2** resulted in increased stabilization of the forward 5′-CGCG-3′ oligomer by 4.5 ºC; stabilization by polyamide **2** in the reverse 5′-GCGC-3′ orientation was diminished by 2.0 ºC. This indicated an improved forward preference for 5′-CGCG-3′.

Nonetheless, the relative magnitudes of the ΔT_m support an overall modest energetic preference for reverse orientation binding.

The inability of the γ-amino GABA turn to enforce forward orientation binding led us to investigate alternative solutions for the molecular recognition of 5′-CGCG-3′. We posited that reverse binding is abetted by the flexibility afforded by the two β units in the core binding region, as had been similarly noted in polyamides containing a β/β pair.[29] We thus considered whether removing one β residue might reinstitute sufficient rigidity in one of the polyamide strands to limit reverse binding while retaining the specificity and affinity provided by the other β. Of the two β moieties in the core of polyamide **1**, the C-terminal β in the core binding region was retained based on previous studies that have shown it is necessary for high affinity recognition of the 5′ C•G base pair.[24] To isolate the effect of each modification, we returned to parent polyamide **1** and replaced the N-terminal β with a Py while retaining the achiral GABA turn, to provide polyamide **3** (Figure 2.3 and 2.4B). The assessment of **3** by Bind-n-Seq followed by DREME analysis generated a *high affinity motif consistent with forward binding 5′-CGCG-3′* (Table 2.1). This was corroborated by ΔT_m measurements showing considerable preference for the forward 5′-CGCG-3′ direction.

We further examined whether a hairpin polyamide designed to target a reverse orientation sequence may productively bind CpG's with high specificity. To test this, we

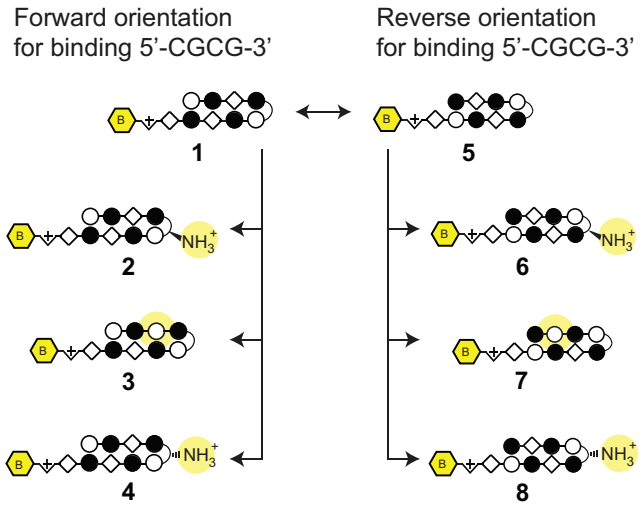


Figure 2.5 A panel of polyamides was synthesized for assessment by Bind-n-Seq and DNA thermal stabilization for binding the 5'-CGCG-3' sequence. According to the pairing rules, polyamides **1**-**4** target 5'-CGCG-3' in the forward orientation and polyamides **5**-**8** target 5'-CGCG-3' in the reverse orientation. Structural modifications are highlighted in yellow.

expanded the library of compounds to include polyamides **4**-**8**, which contain single modifications targeting the 5′-GCGC/CGCG-3′ core (Figure 2.5 and 2.6). In contrast to our findings with 5′-CGCG-3′ targeting polyamide **2**, we confirmed that the incorporation of an α-amino GABA turn in polyamide **6** restores forward orientation binding for the 5′-GCGC-3′ sequence.[30] This difference is striking given the two polyamides are composed of nearly identical amino acid sequences. Bind-n-Seq data and T_m assays of polyamides **4**, **5**, **7**, and **8** together suggest that all other modifications preferentially bind the reverse orientation, and **5**, **7**, and **8** do so with poor specificity (Table 2.1). Indeed, amongst all variations tested of both 5′-CGCG-3′ forward binding and 5′-GCGC-3′ reverse binding cores, polyamide **3** displayed the highest specificity for the 5′-CGCG-3′ sequence (Table 2.1).

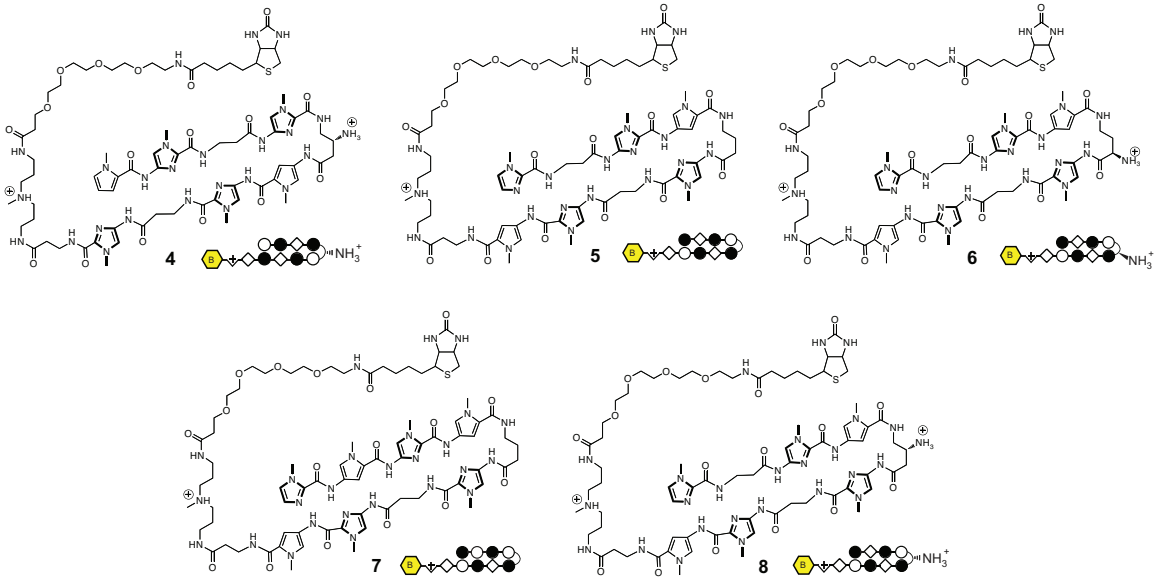


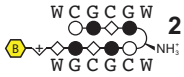
Figure 2.6 Structures of Py-Im polyamides **4**-**8**.

Sequence-Specific Binding Hemi-methylated DNA. Next, we considered the potential for minor groove binding hairpin Py-Im polyamides to prevent DNA methylation undergoing DNA replication. To do so, they must be able to bind the hemi-methylated DNA of daughter strands that have not yet undergone maintenance methylation. DNA thermal stabilization analysis was used to pursue evidence of the above trends of binding orientation with hemi-methylated DNA sequences. The sense strand of each of the 12 base pair oligomers containing 5′-CGCG-3′ or 5′-GCGC-3′ cores were methylated on both cytosines, whereas the antisense strand was left unmethylated. Flanking sequences were modified to lack self-complementarity and enforce hemi-methylated duplex formation. Analysis of ΔT_m of the hemi-methylated DNA oligomers confirmed the above magnitudes of stabilization and trends of reverse and forward binding modes for **1**, **2**, and **3** (Table 2.2A).

Inhibition of Methyltransferase. With a specific polyamide capable of binding

Table 2.2 A) T_m study with hemi-methylated DNA duplex. B) T_m study of three generations of 5′-CGCG-3′ methylation inhibitors without biotin affinity tags.

A

Polyamide	5'-GGT **A**C^{m}**GC**^{m}**GT** TGG-3' 3'-CCA **T**G**C**G**C**A ACC-5'		5'-GGT **A**GC^{m}**GC**^{m}**T** TGG-3' 3'-CCA **T**C**G**C**G**A ACC-5'		Binding Orientation
	T_m / °C	ΔT_m / °C	T_m / °C	ΔT_m / °C	
 **1**	62.7 (±0.5)	4.1	69.3 (±1.5)	10.4	Reverse
 **2**	67.4 (±0.5)	8.8	69.1 (±1.2)	10.2	Reverse
 **3**	68.1 (±1.0)	9.5	63.6 (±1.0)	4.7	Forward

B

Polyamide	5'-GGT **ACGCGT** ACC-3'		5'-GGT **AGCGCT** ACC-3'		Binding Orientation
	T_m / °C	ΔT_m / °C	T_m / °C	ΔT_m / °C	
 **1b**	67.7 (±0.3)	7.1	74.1 (±0.3)	13.7	Reverse
 **3b**	73.6 (±0.4)	13.0	68.1 (±0.7)	7.7	Forward
 **9b**	77.3 (±0.6)	16.7	69.8 (±0.1)	9.4	Forward

hemi-methylated DNA in hand, we evaluated its application as a sequence-specific inhibitor of DNA methyltransferases. The biotin enrichment tag was deleted from the C-terminus by re-synthesis to afford parent hairpins **1b** and **3b** (Figure 2.7). Melting temperature analyses confirmed these molecules show comparable binding preference to biotin conjugates **1** and **3**, respectively (Table 2.2B). We developed an *in vitro* assay to probe the methylation state of specific sites employing the methylation-sensitive restriction enzyme MluI to compare sequence specific effects of **1b**, **3b**, and AT-binding distamycin **D** as a control (Figure 2.8A). In this assay, we measured the ability of these

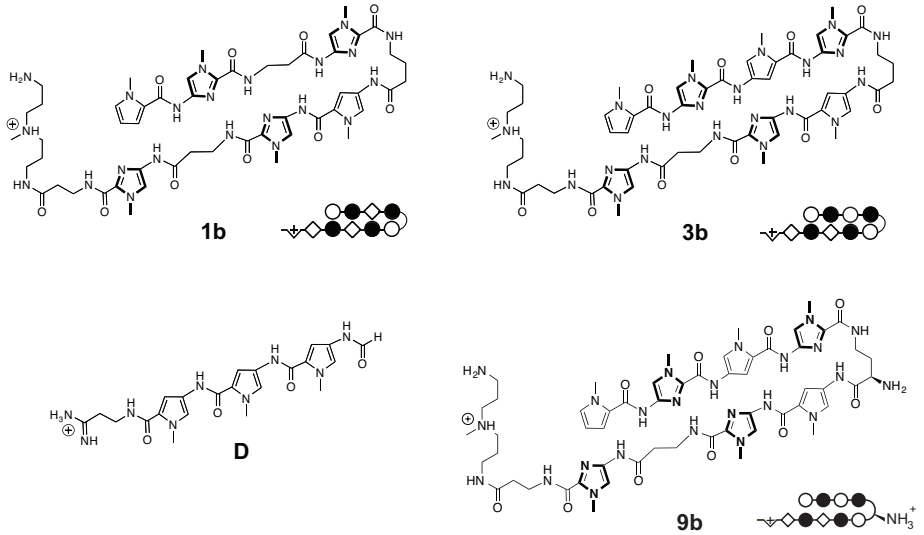


Figure 2.7 Structures of polyamides 1b, 3b, D, and 9b used in *in vitro* functional assays.

compounds to inhibit the methylation activity of M.SssI, a robust prokaryotic methyltransferase that operates in a processive manner like human methyltransferases and shares structural similarities with the catalytic core of human DNMT1.[36] We employed the methylation-sensitive enzyme MluI, which cleaves at seven 5′-ACGCGT-3′ sites,[37] to interrogate methylation of the lambdaphage DNA (48.5 kb), of which five bands were visualized by agarose gel electrophoresis. Both **1b** and **3b** were titrated from increasing concentrations 1 nM to 1 μM while **D** was dosed ten-fold higher from 10 nM to 10 μM. Full digestion of the DNA by MluI indicates a lack of CpG methylation at 5′-ACGCGT-3′ restriction sites, and is demonstrated by positive control lane 2 (Figure 2.8B). In contrast, full methylation would protect DNA from MluI digestion, as in lane 1, where no compound was added to DNA prior to exposure to M.SssI for methylation.

Consistent with our biophysical characterization of the compounds, polyamide **3b** showed the most robust inhibition of CpG methylation (Figure 2.8B, lanes 7-10) at

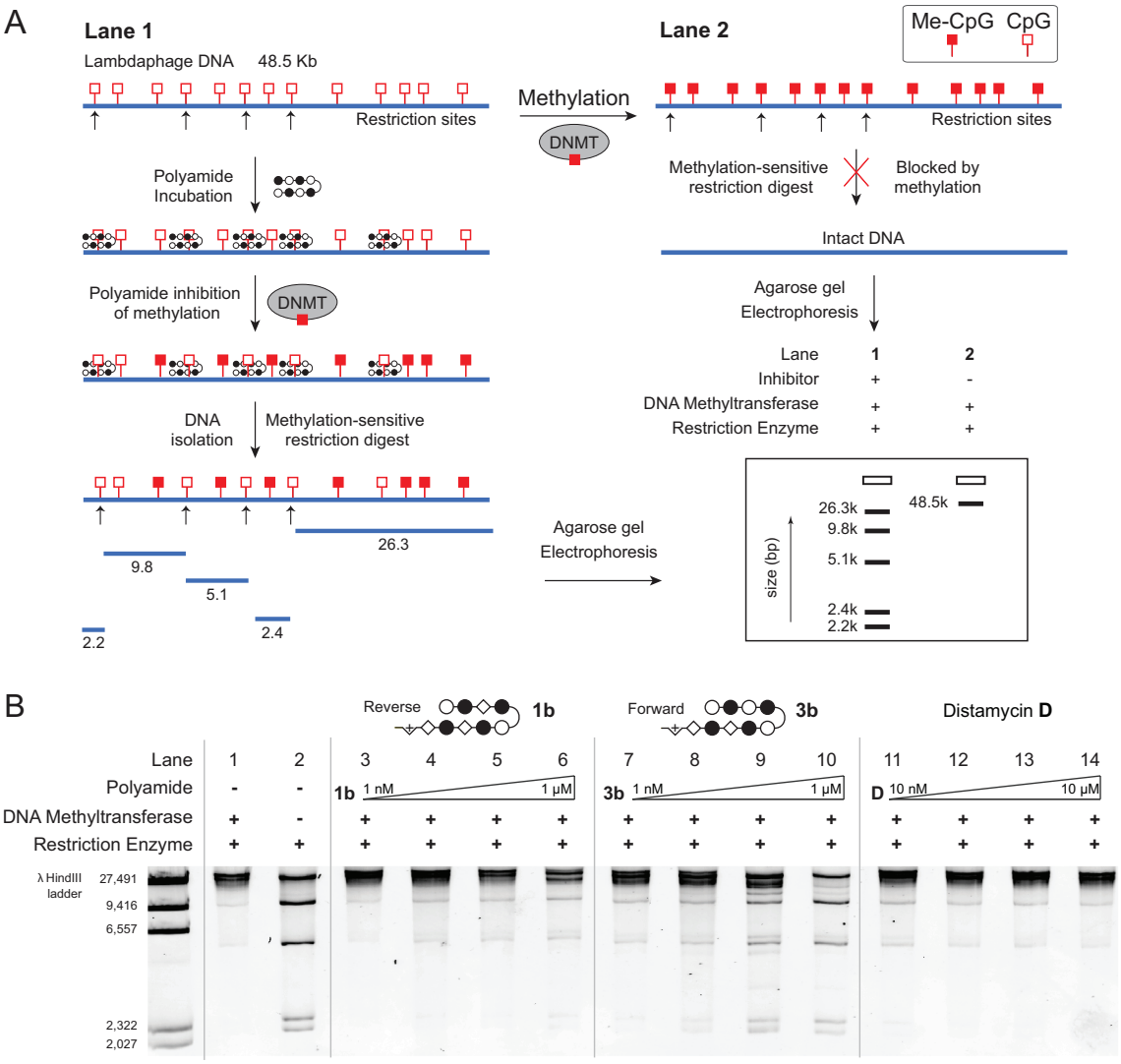


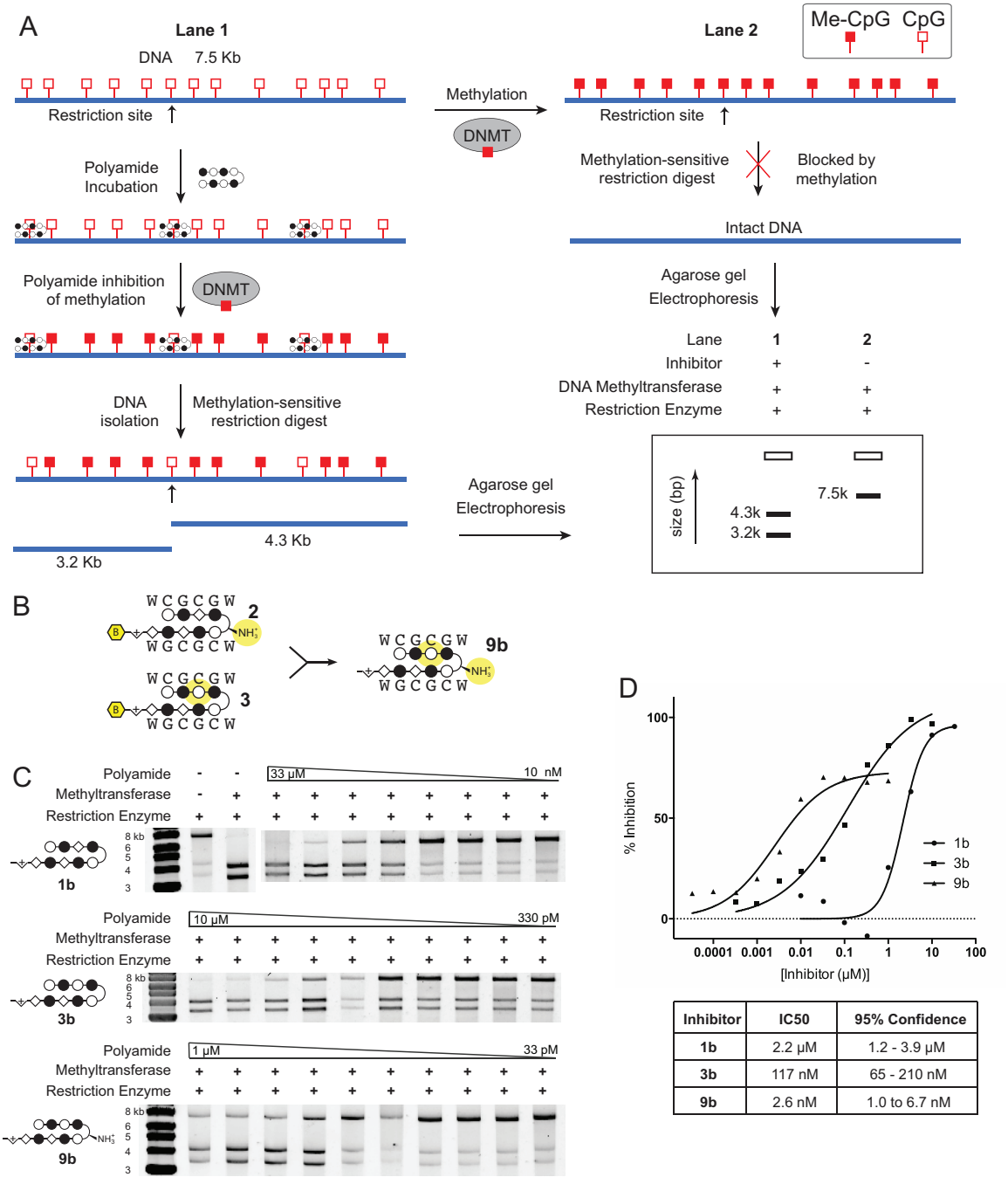
Figure 2.8 A) Scheme of in vitro assay of DNA methyltranferase inhibition. Generic polyamide shown in ball-and-stick notation, CpG sites represented by red squares. (B) Inhibition of methyltransferase activity reflects sequence-specificity of Py-Im polyamides. Lambdaphage DNA was incubated with M.SssI and subject to methylation-sensitive restriction digest at 5'-ACGCGT-3' sites by MluI. DNA is 240 pM in match sites. Positive control lane 2 was not subject to methylation and completely digested while negative control lane 1 shows minimal digestion. Increasing concentration of 3b inhibits methylation at the restriction sites as visualized by additional, smaller restriction fragments. Polyamide 1b and 3b was titrated from 1 nM to 1 µM at ten-fold dilutions and distamycin ranged from 10 nM to 10 µM at ten-fold dilutions. Visualized on 0.7% agarose gel with SYBR gold.

5′-ACGCGT-3′ sites. In lane 10, full MluI digestion comparable to positive control lane 2 was observed at 1 µM of **3b**, indicating this concentration was sufficient to block all methylation at the cognate binding sites. Further, incomplete protection was evidenced at

100 nM of **3b** by additional, partially digested bands in lane 9. In contrast, polyamide **1b** showed weak inhibition of M.SssI and was active only at the highest concentration (Figure 2.8B, lanes 3-6). This reflects its weaker affinity for the 5′-CGCG-3′ forward binding orientation, also observed by thermal duplex denaturation analysis. Inhibition by **1b** at 1 μM, however, was reduced relative to that observed at 100 nM of **3b**, consistent with the binding preferences of the two molecules. There was no inhibition by distamycin **D** at all concentrations tested, even at the highest concentration of 10 μM, underscoring the importance of CpG specificity of Py-Im polyamides in preventing CpG methylation. To enable quantitation of enzyme activity inhibition, the substrate DNA was changed to a 7.5 kb fragment containing a single 5′-ACGCGT-3′ site (Figure 2.9A).

We were encouraged by these results to consider the design of an improved methylation antagonist at 5′-CGCG-3′. We revisited the single modifications to **1** in polyamide **2** and **3** that had promoted forward orientation binding. We combined the γ-amino modification at the GABA turn that had encouraged **2** to bind in the forward orientation, albeit insufficiently, with the Py substitution in the top strand, as in **3,** to afford **9b** (Figure 2.9B). Analysis by thermal denaturation assays revealed that the effects of the modifications were additive, and **9b** displayed increased affinity and preference for forward orientation binding (Table 2.2B).

We then sought to determine IC50 values for the three generations of 5′-CGCG-3′ methylation inhibitors: **1b**, **3b**, and **9b**. With consideration for their DNA binding



Inhibitor	IC50	95% Confidence
1b	2.2 μM	1.2 - 3.9 μM
3b	117 nM	65 - 210 nM
9b	2.6 nM	1.0 to 6.7 nM

Figure 2.9 A) Scheme of *in vitro* DNMT inhibition assay with single restriction site. Generic polyamide shown in ball-and-stick notation and CpG sites represented by red squares. DNA (7.5 kb) with a single MluI restriction site was incubated at 50 pM with inhibitor and subjected to methylation by M.SssI. DNA was isolated for restriction digest by MluI to reveal methylation at the target site. B) Single changes in polyamides **2** and **3** that promote forward orientation binding were combined in designing **9b** as a third generation candidate for improved methylation inhibition. C) Representative gel image of the polyamides **1b**, **3b**, and **9b** in the assay described in (A), suggestive of differential inhibitory activity. Dose ranges of compounds were adjusted in relation to their DNA binding affinities. D) IC50 values of 5′-CGCG-3′ targeting polyamides. Values are determined from band intensities in the *in vitro* as assay shown in (C) and normalized

against maximal methylation with no inhibitor. IC50 values were calculated from at least three replicates and fit to a four-variable, dose-response model.

affinities, compounds **1b**, **3b**, and **9b** were titrated from 10 nM to 33 μM, 330 pM to 10 μM, and 33 pM to 1 μM, respectively (Figure 2.9C). It is worth noting that an additional SDS wash step was necessary in this assay to remove the higher affinity **9b** from the DNA before resolution by the MluI restriction enzyme. Prior to the addition of this SDS incubation, inhibition was maximally revealed to approximately 40%, due to polyamide inhibition of the MluI restriction enzyme. Overnight incubation of DNA in 2% SDS removed additional polyamide and improved the revealed inhibition, suggesting the compressed inhibitory range is an artifact of this method and the high affinity of **9b**. The IC50 values of **1b**, **3b**, and **9b** were determined to be 2.2 μM (95% confidence: 1.2-3.9 μM), 117 nM (95% confidence: 65-210 nM), and 2.6 nM (95% confidence: 1.0-6.7 nM), respectively (Figure 2.9D). This is in good correlation with the iterative improvement shown in the biophysical analyses of these compounds, as well as the previous qualitative *in vitro* assay. Polyamide **9b** shows nearly 1000-fold improvement over **1b** as a 5′-CGCG-3′ methylation antagonist.

2.3 Discussion

Design of Antagonists of CpG Methylation. This study provides a basis for design of sequence-specific DNA-binding molecules for targeted inhibition of CpG methylation. The disparity in methyltransferase inhibition between AT-binding distamycin **D** and hairpin polyamide **3b** suggests that the specific CpG-binding capability and widening of the minor groove by bound Py-Im polyamides are critical for disrupting DNA methylation in the major groove. At the same time, applying the pairing rules

demands caution in the design of imidazole and β-rich polyamides, as the inherent conformational flexibility of the β subunit can support unintended reverse DNA-binding modes. While previous studies have shown that an β-amino GABA turn unit can be used to restore the forward orientation binding preference of β-containing polyamides, we found 5′-CGCG-3′ binding Py-Im polyamides required an alternative solution. Specifically, restoring the rigidity of the N-terminal strand via substitution of its β-subunit with a Py appears necessary to resolve the undesired reverse-binding of these architectures.

Potential mechanism for inhibition of CpG methylation. The lack of inhibition of CpG methylation by the reverse binding **1b** as compared to **3b** at the interrogated 5′-CGCG-3′ sites suggests inhibition of the precessive M.SssI enzyme is a sequence-specific, localized event. The M.SssI methyltransferase, like eukaryotic methyltransferases, is a "flipase" that swings the target cytosine out of the double helix and into its catalytic core.[38] All known CpG methyltransferases operate by this conserved mode of action. Structural studies of mouse DNMT1, the relevant mammalian methyltransferase for maintenance methylation, show that enzyme residues enter the double helix from both the major and minor grooves in an intercalative-manner around the target CpG.[39] These residues disrupt local base pairing and rotate the substrate cytosine around the sugar-phosphate backbone and into the catalytic core of the enzyme. A Py-Im polyamide bound to the target DNA site likely acts as a stabilizing clamp in the minor groove and prevents the intrusion of these residues. The increased DNA stability

disallows the conformational reorganization of the CpG substrate necessary for catalysis and results in the inhibition of methyltransferase activity.

Conclusion

In this study, we examined programmable Py-Im polyamides targeting the 5′-CGCG-3′ sequence as a model for sequence-specific inhibition of CpG methylation. The unbiased Bind-n-Seq method was critical for revealing unanticipated binding modes of the polyamides. Through deliberate, incremental synthetic modifications, we were able to discern structure activity relationships that guided improved design of CpG methylation antagonists. Further work will be necessary to understand whether this represents a more general solution for controlling Py-Im polyamide orientation or is specific to the 5′-CGCG-3′ sequence. This study demonstrates that high affinity minor groove binding Py-Im polyamides can inhibit major groove CpG methylation by methyltransferase in a sequence-specific manner. It will be the focus of future research to assess these molecules as antagonists of CpG methylation in cells and its utility in the desilencing of specific genes. It will be of interest whether the intrinsic rarity of the CpG dinucleotide sequence and non-covalent binding of polyamides will reduce off-target effects.

2.4 Materials and methods

Py-Im polyamide synthesis. Polyamides were synthesized by microwave-assisted, solid-phase synthesis on PAM resin (Peptides International) according to previously described protocols.[40,41] The polyamides were cleaved from resin with 3,3′-

diamino-N-methyldipropylamine and purified by reverse phase HPLC. For biotin-conjugated polyamides, the free amine at the C-terminus was allowed to react with 2

Table 2.3 Mass Spectrometry (MALDI-TOF) for Py-Im polyamides.

Py-Im Polyamide	**Formula**	**[Mass+H]**	**Found**
PyImβIm-γ-PyImβImβ-(+)-PEG4-Biotin (**1**)	$C_{73}H_{108}N_{25}O_{17}S^+$	1638.8	1638.0
PyImβIm-(*R*)$^{\alpha\text{-}NH2}$γ-PyImβImβ-(+)-PEG4-Biotin (**2**)	$C_{73}H_{109}N_{26}O_{17}S^+$	1653.8	1653.8
PyImPyIm-γ-PyImβImβ-(+)-PEG4-Biotin (**3**)	$C_{76}H_{109}N_{26}O_{17}S^+$	1689.8	1689.2
PyImβIm-(*R*)$^{\beta\text{-}NH2}$γ-PyImβImβ-(+)-PEG4-Biotin (**4**)	$C_{73}H_{109}N_{26}O_{17}S^+$	1653.8	1653.8
ImβImPy-γ-ImβImPyβ-(+)-PEG4-Biotin (**5**)	$C_{73}H_{108}N_{25}O_{17}S^+$	1638.8	1639.0
ImβImPy-(*R*)$^{\alpha\text{-}NH2}$γ-ImβImPyβ-(+)-PEG4-Biotin (**6**)	$C_{73}H_{109}N_{26}O_{17}S^+$	1653.8	1653.9
ImPyImPy-γ-ImβImPyβ-(+)-PEG4-Biotin (**7**)	$C_{76}H_{109}N_{26}O_{17}S^+$	1689.8	1690.1
ImβImPy-(*R*)$^{\beta\text{-}NH2}$γ-ImβImPyβ-(+)-PEG4-Biotin (**8**)	$C_{73}H_{109}N_{26}O_{17}S^+$	1653.8	1654.1
PyImβIm-γ-PyImβImβ-(+)- (**1b**)	$C_{52}H_{73}N_{22}O_{10}^+$	1165.6	1165.6
PyImPyIm-γ-PyImβImβ-(+)- (**3b**)	$C_{55}H_{74}N_{23}O_{10}^+$	1216.6	1216.4
PyImPyIm-(R)$^{\alpha\text{-}NH2}$γ-PyImβImβ-(+)- (**9b**)	$C_{55}H_{75}N_{24}O_{10}^+$	1231.6	1231.7

equivalents of pre-activated PEG4-biotin NHS ester (Thermo Scientific) and 4 equivalents of DIEA for 1 hour at 55ºC in DMF. The product was purified by reverse phase HPLC and lyophilized. Purity and identity of compounds were verified by analytical HPLC and matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry (Table 2.3).

Bind-n-Seq of polyamide-biotin conjugates. Sequence motif logos of the highest affinity DNA binding sites of polyamide-biotin conjugates **1**-**8** were determined

according to previously reported methods.[30] Each Py-Im polyamide-biotin conjugate was equilibrated at 50 nM concentration for 15 hours with a uniquely barcoded library of all possible 21mers. DNA associated with polyamide-biotin conjugates were affinity purified with streptavidin magnetic beads (M-280 Dynabeads) and eluted. Isolated DNA was amplified by touchdown PCR and sequenced at the California Institute of Technology Millard and Muriel Jacobs Genetics and Genomics Laboratory on an Illumina HiSeq 2000 Genome Analyzer. The generated dataset was then distributed by barcode using scripts in the MERMADE pipeline and a fasta file of a random 25% of sequences for each compound submitted for DREME motif analysis.[30]

DNA thermal denaturation assay. Unmethylated DNA duplexes and hairpin polyamides were mixed to a final concentration of 2 and 3 μM, respectively, for polyamides **1-8**, **1b**, and **3b** in 1 ml total volume. For experiments with hemi-methylated oligonucleotides, DNA duplexes and hairpin polyamides were mixed to a final concentration of 1 and 1.5 μM, respectively. An aqueous solution of 10 mM sodium cacodylate, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$ at pH 7.0 was used as analysis buffer. All oligonucleotides (100μM solutions dissolved in 10 mM Tris-Cl, 0.1 mM EDTA, pH 8.0) were purchased from Integrated DNA Technologies. The assay was conducted on a Varian Cary 100 spectrophotometer equipped with a thermo-controlled cell holder with a cell path length of 1 cm. Samples were heated to 90 ºC and cooled to a starting temperature of 25 ºC prior to heating at a rate of 0.5 ºC/min to 90 ºC. Denaturation profiles were recorded at λ= 260 nm and melting temperatures were defined as the maximum of the first derivative of the denaturation profile. Reported data represents the average of four measurements.

***In vitro* inhibition of CpG methylation assay.** In PCR tubes, serially diluted concentrations of polyamides **1b**, **3b**, and distamycin **D** control were incubated in 96 µl of 10 pM unmethylated lambdaphage DNA (Promega) and 1X NEB2 buffer (New England Biolabs) in DEPC-treated water (USB) for 12 hours at 25 ºC. Two additional samples of DNA in buffer without compound were kept for controls. After incubation, S-adenosyl methionine (New England Biolabs) and M.SssI (New England Biolabs) or water was added to all samples to a final concentration of 320 µM and 0.25 Units, respectively, to afford 100 µl of total solution. Samples were then incubated for 3 hours at 37 ºC on a Biorad MyCycler thermal cycler and heat inactivated for 15 minutes at 65 ºC. DNA was ethanol precipitated in a centrifuge at 4 ºC for 15 minutes with the addition of 10 µl of 3M NaOAc, 1 µl of glycogen, and 2.5 volumes of ethanol at -20 ºC. DNA was washed once with 75% aqueous ethanol at -20 ºC and allowed to air dry for 30 minutes. Samples were dissolved in 35 µl of water and 15 µl taken for MluI restriction enzyme digestion. Samples were prepared in PCR tubes per manufacturer's protocol with 1 Unit of MluI per sample and incubated at 37 ºC for 1 hour. Blue loading buffer 6x (New England Biolabs) was added to samples and 20 µl added to a 0.7% agarose gel in 0.5x TBE buffer. DNA was visualized with SYBR gold (Invitrogen) and a Typhoon FLA9000 Scanner (GE Healthcare).

Determination of IC50. The *in vitro* assay was run as described above with **1b**, **3b**, and **9b** at concentrations titrated at 10-fold and 3-fold intervals ranging from 10 nM to 33 µM, 330 pM to 10 µM, and 33 pM to 1 µM, respectively, and DNA at 50 pM. The substrate DNA fragment (7.5 kb) was PCR amplified from PTYB21 (New England Biolabs) after linearization with BamHI (New England Biolabs). Primers 5′-

ACTTTTCGGGGAAATGTGCG-3′ and 5′-TTAGAGGCCCCAAGGGGTTA-3′ (IDT DNA) were used for amplification with the Expand Long Template PCR System (Roche). The amplicon was isolated with QIAquick PCR Purification Kit (Qiagen), and the amplicon size was verified by agarose gel electrophoresis. After the ethanol precipitation step, which follows methylation, the DNA pellet was dissolved in 100 µl of 2% SDS and incubated overnight at 55 ºC to wash off residual polyamide. The high affinity of **9b** made this additional wash step necessary prior to MluI digestion. To the solution, 10 µl of 2M NaCl followed by 2.5 volumes of ethanol were added to re-precipitate the DNA. The pellet was washed twice with cold 75% ethanol before submission to MluI digest, as described above. Digested samples were run on 1% agarose gels and visualized with SYBR Gold. Gels were scanned on a Typhoon FLA Scanner (GE Healthcare) and the bands quantitated using ImageQuant Software (GE Healthcare). Percentage inhibition was normalized against maximal methylation in the presence of no inhibitor.

$$\% \text{ Inhibition} = 100\% \times \left(1 - \frac{\% \, uncut \, DNA}{\% \, Maximal \, uncut \, DNA}\right)$$

IC50 curves and 95% confidence intervals were determined using GraphPad Prism by variable-slope, nonlinear regression fit to a dose response model with a bottom constraint of 0. At least three replicates of each concentration were used.

2.5 Acknowledgment

Sequencing was conducted at the Millard and Muriel Jacobs Genetics and Genomics Laboratory at the California Institute of Technology. Mass spectrometry analyses were performed in the Mass Spectrometry Laboratory of the Division of

Chemistry and Chemical Engineering at the California Institute of Technology. Gels were scanned in the Center for the Chemistry of Cellular Signaling at the California Institute of Technology. Studies were funded in part by the National Institutes of Health (grant numbers GM51747 and GM27681); Tobacco-Related Disease Research Program (award number 20DT-0037 to J.S.K., dissertation research award); and American Cancer Society (grant number PF-10-015-01-CDD to J.L.M., postdoctoral fellowship).

2.6 References

(1) Jones, P. A.; Baylin, S. B. *Nat. Rev. Genetics* **2002**, *3*, 415-428.

(2) Rountree, M. R.; Bachman, K. E.; Herman, J. G.; Baylin, S. B. *Oncogene* **2001**, *20*, 3156-3165.

(3) Bird, A. P. *Nature* **1986**, *321*, 209-213.

(4) Ballestar, E.; Wolffe, A. P. *Eur. J. Biochem.* **2001**, *268*, 1-6.

(5) Merlo, A.; Herman, J. G.; Mao, L.; Lee, D. J.; Gabrielson, E.; Burger, P. C.; Baylin, S. B.; Sidransky, D. *Nat. Med.* **1995**, *1*, 686-692.

(6) Katzenellenbogen, R. A.; Baylin, S. B.; Herman, J. G. *Blood* **1999**, *93*, 4347-4353.

(7) Plumb, J. A.; Strathdee, G.; Sludden, J.; Kaye, S. B.; Brown, R. *Cancer Res.* **2000**, *60*, 6039-6044.

(8) Dammann, R.; Li, C.; Yoon, J. H.; Chin, P. L.; Bates, S.; Pfeifer, G. P. *Nat. Genet.* **2000**, *25*, 315-319.

(9) Christman, J. K. *Oncogene* **2002**, *21*, 5483-5495.

(10) Waring, M. J.; Wakelin, L. P. G. *Nature* **1974**, *252*, 653-657.

(11) Adams, R. L.; Rinaldi, A. *FEBS Lett.* **1987**, *215*, 266-268.

(12) Stresemann, C.; Brueckner, B.; Musch, T.; Stopper, H.; Lyko, F. *Cancer Res.* **2006**, *66*, 2794-2800.

(13) Fagan, R. L.; Cryderman, D. E.; Kopelovich, L.; Wallrath, L. L.; Brenner, C. *J Biol. Chem.* **2013**, *288*, 23858-23867.

(14) Ceccaldi, A.; Rajavelu, A.; Ragozin, S.; Sénamaud-Beaufort, C.; Bashtrykov, P.; Testa, N.; Dali-Ali, H.; Maulay-Bailly, C.; Amand, S.; Guianvarch, D. *ACS Chem. Biol.* **2013**, *8*, 543-548.

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

(16) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature* **1996**, *382*, 559-561.

(17) White, S.; Baird, E. E.; Dervan, P. B. *Chem. Biol.* **1997**, *4*, 569-578.

(18) White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Nature* **1998**, *391*, 468-471.

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

(20) 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.

(21) Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 7983-7988.

(22) Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6160-6166.

(23) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *Chem. Eur. J.* **1997**, *3*, 1600-1607.

(24) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 3534-3535.

(25) Turner, J. M.; Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 6219-6226.

(26) Wang, C. C.; Ellervik, U.; Dervan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 653-657.

(27) White, S.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 8756-8765.

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

(29) Rucker, V.; Melander, C.; Dervan, P. *Helv. Chim. Acta* **2003**, *86*, 1839-1851.

(30) Meier, J. L.; Yu, A. S.; Korf, I.; Segal, D. J.; Dervan, P. B. *J. Am. Chem. Soc.* **2012**, *134*, 17814-17822.

(31) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 6953-6961.

(32) Minoshima, M.; Bando, T.; Sasaki, S.; Fujimoto, J.; Sugiyama, H. *Nucleic Acids Res.* **2008**, *36*, 2889-2894.

(33) Chenoweth, D. M.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A* **2009**, *106*, 13175-13179.

(34) Chenoweth, D. M.; Dervan, P. B. *J. Am. Chem. Soc.* **2010**, *132*, 14521-14529.

(35) Pilch, D. S.; Poklar, N.; Gelfand, C. A.; Law, S. M.; Breslauer, K. J.; Baird, E. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A* **1996**, *93*, 8306-8311.

(36) Brueckner, B.; Garcia Boy, R.; Siedlecki, P.; Musch, T.; Kliem, H. C.; Zielenkiewicz, P.; Suhai, S.; Wiessler, M.; Lyko, F. *Cancer Res* **2005**, *65*, 6305-6311.

(37) McClelland, M.; Nelson, M. *Nucleic Acids Res.* **1992**, *20 Suppl*, 2145-2157.

(38) Darii, M. V.; Cherepanova, N. A.; Subach, O. M.; Kirsanova, O. V.; Raskó, T.; Slaska-Kiss, K.; Kiss, A.; Deville-Bonne, D.; Reboud-Ravaux, M.; Gromova, E. S. *Biochim Biophys Acta* **2009**, *1794*, 1654-1662.

(39) Song, J.; Teplova, M.; Ishibe-Murakami, S.; Patel, D. J. *Science* **2012**, *335*, 709-712.

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

(41) Puckett, J. W.; Green, J. T.; Dervan, P. B. *Org. Lett.* **2012**, *14*, 2774-2777.

Chapter 3

A sequence-specific DNA binding small molecule triggers the release of immunogenic signals and phagocytosis in a model of B-cell lymphoma

The text of this chapter is taken in part from a manuscript co-authored with Peter B. Dervan (California Institute of Technology).

Abstract

Means to cause an immunogenic cell death could lead to significant insight into how cancer escapes immune control. In this study, we screened a library of five Py-Im polyamides coding for different DNA sequences in a model of B cell lymphoma for the upregulation of surface calreticulin, a pro-phagocytosis signal implicated in immunogenic cell death. We found that hairpin polyamide **1** triggers the release of the Damage-Associated Molecular Patterns (DAMPs) calreticulin, ATP, and HMGB1 in a slow necrotic-type cell death. Consistent with this signaling, we observed an increase in the rate of phagocytosis by macrophages after the cancer cells were exposed to polyamide **1**. The DNA sequence preference of polyamide **1** is 5'-WGGGTW-3' (where W=A/T), indicated by the pairing rules and confirmed by the Bind-n-Seq method. The close correspondence of this sequence with the telomere-repeat sequence suggests a potential mechanism of action through ligand binding at the telomere. This study reveals a chemical means to trigger an inflammatory necrotic cell death in cancer cells.

3.1 Introduction

Avoidance of immune destruction has been called the seventh hallmark of cancer.[1,2] According to the immuno-editing paradigm, the immune system recognizes and destroys those proto-oncogenic lesions capable of triggering an immune response while those that escape immune control grow to become clinically detectable disease.[3-6] Studies suggest that therapies that enlist the immune system maintain more durable disease control in the clinical setting.[7] Chemical methods to cause immunogenicity in cancer cells would be an important tool towards understanding immunomodulation in the treatment of cancer. A prerequisite for the activation of an anti-cancer immune response is the recognition of the damaged cells as a threat. Damaged cells release immunostimulatory molecules, called Damage-Associated Molecular Patterns (DAMPs), to recruit and activate professional phagocytes such as macrophages and dendritic cells.[8-10] These antigen-presenting cells engulf and process the cancer cells to further prime the immune system for targeted elimination of cancer.[11]

While most chemotherapeutic regimens cause a non-immunogenic or even tolerogenic cell death, recent reports suggest anthracyclins or γ-radiation are particularly effective because they result in the release of DAMPs.[9,12] The extracellular exposure of the intracellularly abundant molecules calreticulin (CRT), HMGB1, and ATP have been suggested to form a spatiotemporal code for immunogenicity.[13,14] The presentation of CRT, an abundant ER-resident chaperone protein, to the cell surface was identified as a necessary and sufficient pro-phagocytic signal for professional phagocytes.[9] The study

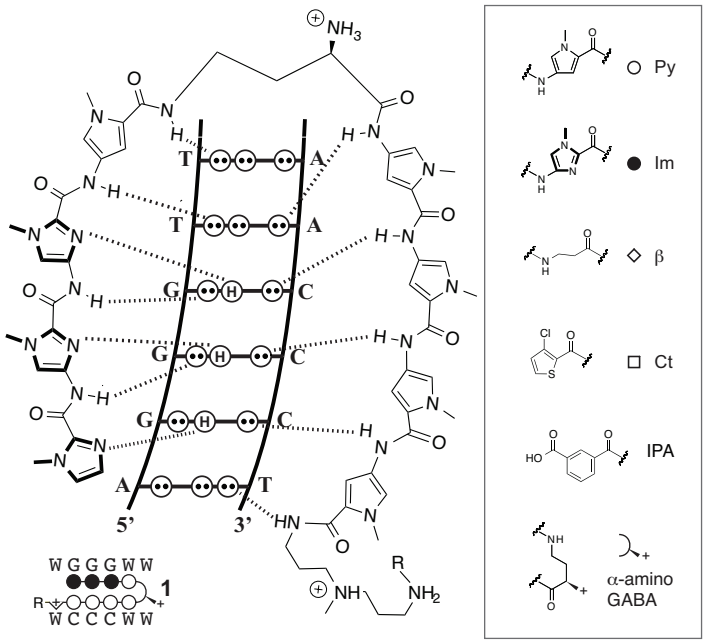


Figure 3.1 Py-Im polyamide recognition of DNA minor groove and corresponding ball-and-stick notation.

showed that stimulation of CRT surface expression by anthracyclins or adsorbtion of the calreticulin protein on the cell surface was sufficient to elicit an anti-cancer immune response in syngeneic mice.[9] Weissman and coworkers further demonstrated in the Raji cell line, a model of human B cell non-Hodgkin's lymphoma, that CRT is the dominant pro-phagocytosis signal that is necessary for engulfment by human macrophages. [15] Furthermore, ATP released from the cytosol into the local microenvironment serves as a lymphocyte recruiting and activating chemokine.[16,17] Lastly, the nucleus-resident protein HMGB1 can be secreted into the surroundings as an inflammatory adjuvant and was shown to be necessary for a durable anti-cancer response in mice.[18,19] Identification of additional small molecules that trigger the release of these DAMPs from tumor cells would be of utility to the field in addressing the heterogeneity of cancers. We became interested in expanding the examined chemical space for compounds capable of causing an immunogenic cell death. Because the DNA damage pathway has been implicated in

immunogenic signaling[20] and anthracyclins are DNA-intercalating ligands, we sought to explore a class of minor groove DNA-binding oligomers hitherto not studied for this biological activity.

Hairpin pyrrole-imidazole (Py-Im) polyamides are a class of sequence-specific oligomers that bind in the minor groove of DNA.[21-26] Sequence preference is achieved by the pair-wise, co-facial arrangement of aromatic amino acids that distinguish the edges of the four Watson-Crick base pairs (Figure 3.1A).[27] Pairing rules for programmable specificity have been established: Im/Py specifies a G•C base pair, Hp/Py codes for T•A base pairs, and Py/Py binds both T•A/A•T. [24] Eight-ring hairpin polyamides are linked in an antiparallel fashion by a central aliphatic γ-aminobutyric acid unit.[28] Polyamides of this hairpin architecture have affinities for match sites similar in magnitude to natural DNA binding proteins (K_a = 10^8 to 10^{10} M^{-1}).[29] Eight-ring hairpins of this class are cell-permeable and modulate transcription in both cells and mice.[30-33] In this study, we screened a small library of Py-Im polyamides coding for different six base pair DNA sequences in Raji cells for the upregulation of surface calreticulin. We found one hairpin polyamide which displayed activity in this screen and characterize its potential for causing an immunogenic cell death.

3.2 Results

Polyamide 1 upregulates calreticulin on the cell surface.

We tested five Py-Im polyamides (**1-5**, Figure 3.2) that bind five unique DNA

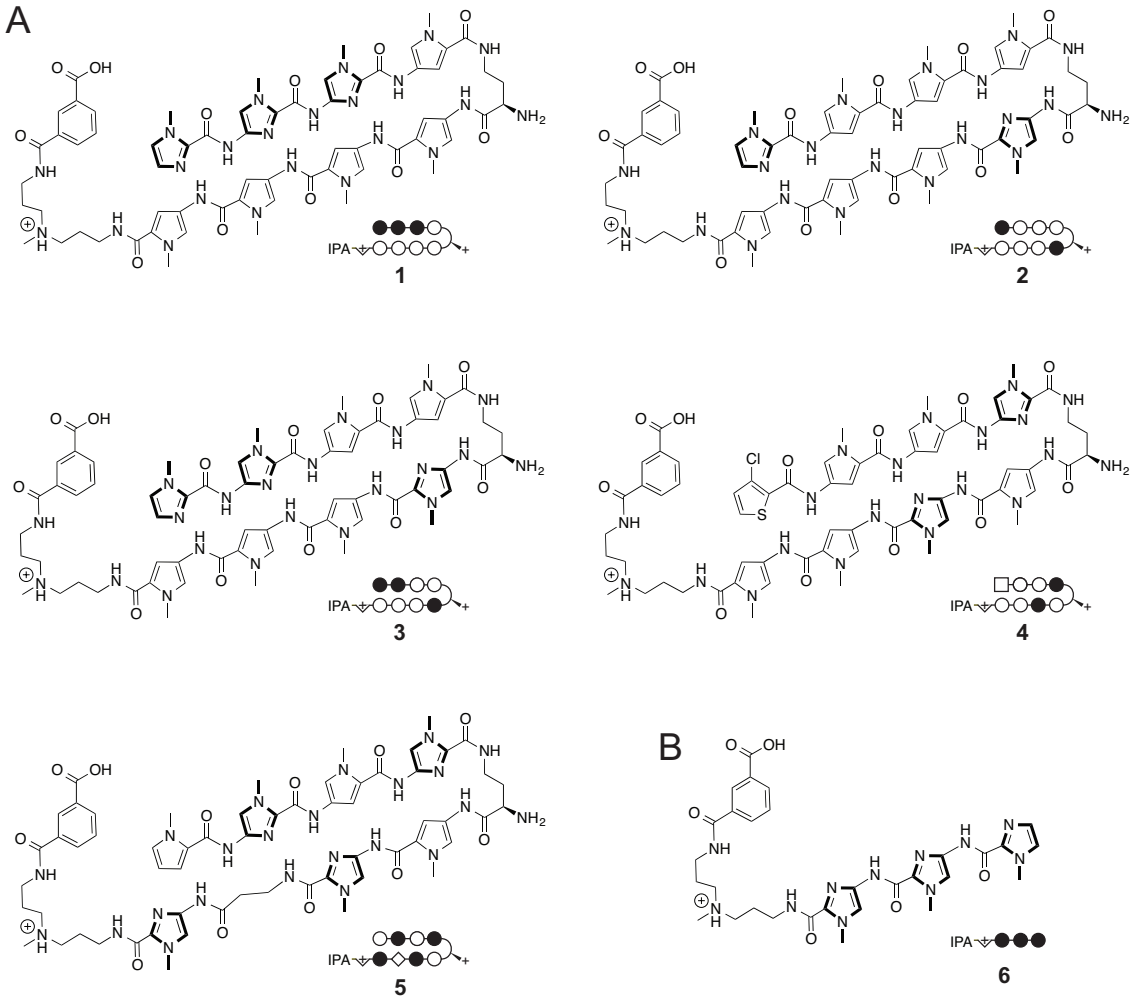


Figure 3.2 A) Structures of hairpin Py-Im polyamides **1-5** screened for stimulation of surface calreticulin (CRT) in Raji cells. B) Structure of ImImIm trimer polyamide **6**.

sequences (5'-WGGGWW-3', 5'-WGWWCW-3', 5'-WGGWCW-3', 5'-WTWCGW-3', and 5'-WCGCGW-3', respectively, where W=A/T) and have demonstrated biological activity.[30,31,34-36] Raji cells, which have previously been utilized in CRT, phagocytosis, and immunotherapy animal models, were dosed at 5 μM for 24 hours with each of the polyamides **1-5**. In addition, 5 μM doxorubicin (**Dox**) and mitoxantrone (**Mtx**, 1 μM) were included as representative anthracyclins. The topoisomerase inhibitor etoposide (**Eto**, 30 μM) was also included for comparison. A 4 hour exposure time point was used

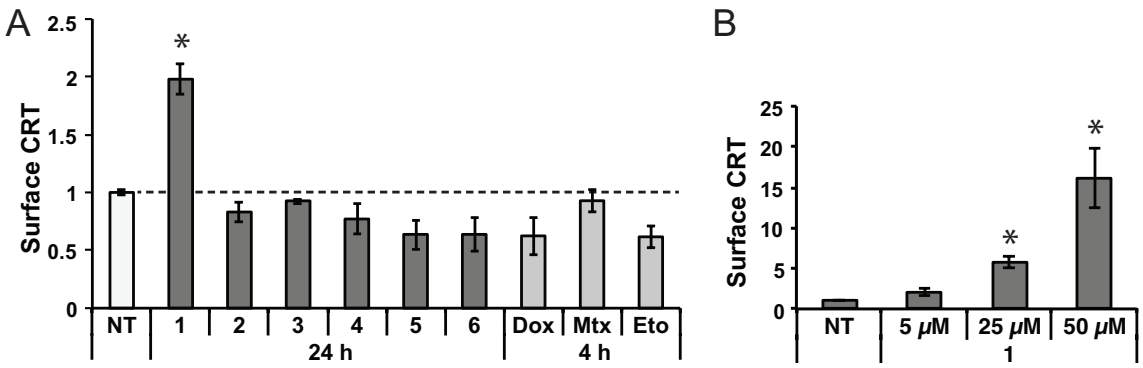


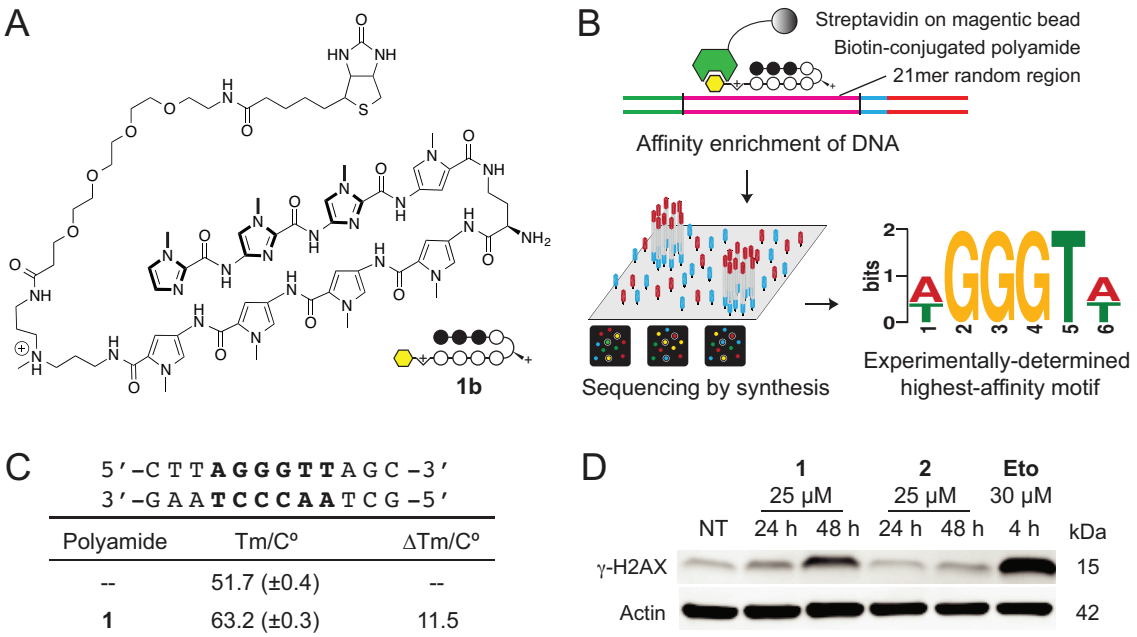
Figure 3.3 A) Surface expression of calreticulin measured by flow cytometry, median fluorescence normalized to non-treated control (NT). Cytotoxic controls doxorubicin (**Dox**, 5 μM), mitoxantrone (**Mtx**, 1 μM), and etoposide (**Eto**, 30 μM). B) Dose-dependence of cell surface CRT to polyamide **1** was measured by flow cytometry after 24 h treatment. Measurement and standard deviation is in triplicate and representative of two independent experiments; asterisk indicates $p<0.05$ by two-tailed Student's t-test compared to NT.

to measure CRT due to the high toxicity of these chemotherapeutics and to be in the range of literature precedent.[9] We measured surface CRT by flow cytometry in a population gated for live cells and saw a statistically significant, two-fold increase in cells treated with only polyamide **1** (Figure 3.3A). In contrast, we saw no activity with the other polyamides **2**-**5** despite their structural similarity. Notably, polyamide **2** has been the most extensively studied amongst this class and has shown activity in prostate cancer xenograft models and the DNA damage response.[33,37] Remarkably, polyamides **3** and **1** share the same molar weight and composition of Im/Py pairs, and only differ by the exchange of one Im/Py ring pair. To assess the structure activity relationship of the imidazole trimer portion of the hairpin oligomer **1**, we synthesized and examined ImImIm polyamide **6**. This was tested because imidazoles are known to complex with calcium and CRT is involved in calcium homeostasis.[38] We did not, however, observe an increase of CRT with polyamide **6**, which suggests the imidazole trimer alone is not sufficient to trigger the surface expression of CRT. We observed no increase in CRT with the two anthracyclins tested, **Dox** and **Mtx**, or with **Eto**. The diminished response to

anthracyclin treatment in Raji cells, as compared to that reported in the literature with murine colon cancer cells, reflects a known range of CRT response to anthracyclins and may be attributable to a difference in cell lines.[39,40] We next increased the treatment concentration of **1** to 25 and 50 μM and found that CRT exposure increased in a dose-dependent manner (Figure 3.3B). The results indicate polyamide **1** is unique in our small library of compounds in its modulation of surface expression of CRT.

Polyamide 1 preferentially binds the sequence 5'-WGGGTW-3'

Py-Im polyamides are a class of sequence-specific DNA-binding ligands and the DNA binding preferences of polyamide **1** may be related to a mechanism of action. By the pairing rules, polyamide **1** is a perfect match to 5'-WGGGWW-3', not unlike the TTAGGG-repeat sequence found in human telomeres. Indeed, dye-conjugated tandem hairpin Py-Im polyamides that recognize 10 base pairs of this repeat sequence can be used to stain telomeres in permeabilized cells.[41,42] We sought to confirm the preferred binding sequence motif of **1** using the Bind-n-Seq method. Bind-n-Seq couples affinity enrichment with next-generation sequencing to query genome-sized sequence space for high-affinity binding sequences.[43] We modified polyamide **1** at the C-terminal position with a biotin-label to afford **1b** (Figure 3.4A) for submission to Bind-n-Seq. Polyamide **1b** strongly preferred binding the DNA sequence motif of 5'-WGGGTW-3' (Figure 3.4B). We additionally confirmed the affinity of polyamide **1** for this sequence by measuring the thermal stabilization afforded to sequence-matched double-stranded DNA.[44] We tested a DNA fragment that included the telomeric DNA sequence



Polyamide	Tm/C°	ΔTm/C°
--	51.7 (±0.4)	--
1	63.2 (±0.3)	11.5

Figure 3.4 Preferred DNA binding sequence of polyamide **1**. A) Structure of polyamide **1b**, the biotinylated analog of **1**. B) Polyamide **1b** was tested by the Bind-n-Seq assay and found to preferentially bind the described DNA sequence motif. C) Affinity of **1** was assessed by a DNA thermal stabilization assay against dsDNA containing the telomeric-repeat sequence 5'-TTAGGG-3'. D) Immunoblot of γ-H2AX and actin control after the indicated treatment.

5'-TTAGGGTTAG-3' (Figure 3.4C). Polyamide **1** stabilized the 12 base pair DNA fragment by 11.5 ºC, suggesting a high affinity hairpin polyamide. Lastly, we detected a marker of DNA stress, phosphorylation of serine 139 on the histone H2AX, after treatment with polyamide **1** but not **2** (Figure 3.4D). We chose the 25 μM dose for this experiment and others described below because it resulted in a robust six-fold increase in surface CRT. We posit that the CRT effect may be due to the unique properties of the DNA target sequence of polyamide **1** in cell biology. The DNA stress, thermal denaturation, and Bind-n-Seq results together suggest the telomere sequence is a plausible target for the mechanism of action of polyamide **1**-mediated CRT exposure.

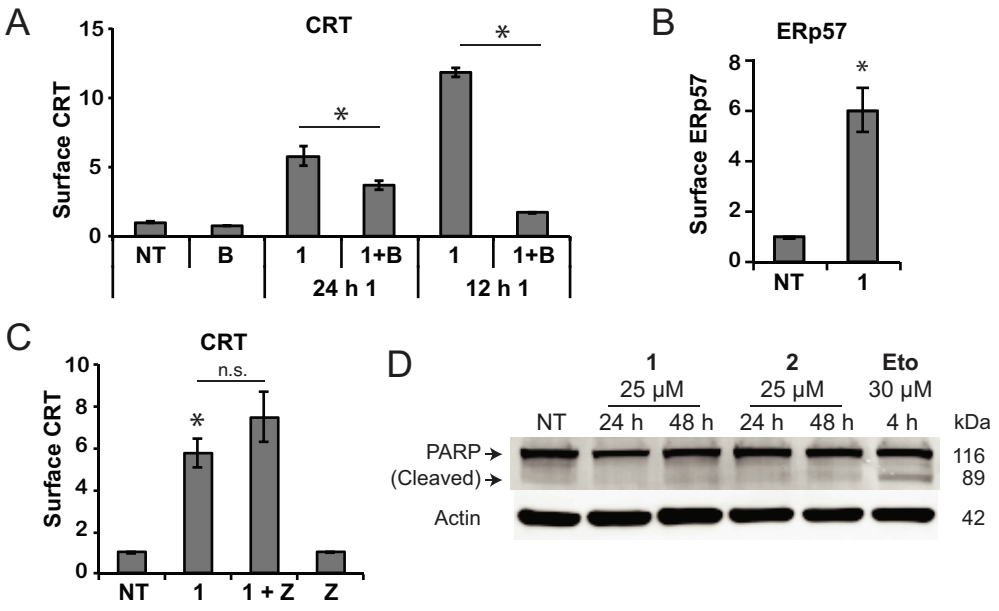


Figure 3.5 Treatment of Raji cells with polyamide **1** triggers anterograde CRT transport by a different mechanism than previously reported for anthracyclins. A) Surface CRT was measured by flow cytometry after Raji cells were treated with 25 μM polyamide **1** for 12 or 24 h and Brefeldin A (**B**) for the final 12 h. B) Co-chaperone ERp57 was measured on the cell surface by flow cytometry after 24 h treatment with **1** at 25 μM. C) Caspase inhibitor Z-VAD-fmk (**Z**, 10 μM) and polyamide **1** (25 μM) were dosed together for 24 h and assessed by flow cytometry for surface CRT. D) Immunoblots for PARP cleavage and actin control after the indicated treatments are shown. All flow cytometry analyses were done in triplicate and are representative of at least two independent experiments. Error bars show standard deviations and asterisks mark statistically significant changes ($p<0.05$) by two-tailed Student's t-test compared to no treatment (unless another comparator is marked).

Polyamide 1 triggers CRT in a different manner than do anthracyclins

To better understand the effects of polyamide **1**, we compared its trigger of CRT to reports of CRT induction by anthracyclins.[45] Anthracyclin-induced CRT exposure has been described to occur with the co-chaperone ERp57 by anterograde ER-Golgi transport.[45] To interrogate whether polyamide **1** induces CRT export to the cell surface by ER-Golgi transport, we applied the Golgi transport inhibitor Brefeldin A[46] (**B**) to Raji cells (Figure 3.5A). Due to the toxicity of **B**, we could only dose **B** for 12 hours. With 24 hour treatment with polyamide **1** and 12 hour treatment with **B**, we saw statistically

significant reduction of CRT on the cell surface when compared to polyamide **1** treatment alone (Figure 3.5A). When polyamide **1** and **B** were dosed together for 12 hours, we saw near ablation of all polyamide **1**-induced CRT, demonstrating the necessity of this pathway for transport. This further suggests CRT has a half-life on the cell surface of at least 12 hours. We additionally measured co-chaperone ERp57 by flow cytometry and found that after treatment with **1** for 24 hours, ERp57 increased on the cell surface to a similar extent as did CRT, approximately six-fold (Figure 3.5B). These results are consistent with previous reports that the translocation of these ER-resident chaperones to the plasma membrane occurs by anterograde ER-Golgi transport. [45]

Anthracyclin-induced CRT has previously been reported to be mediated by a caspase 8- and caspase 3/7-dependent pathway that can be disrupted with the pan-caspase inhibitor Z-VAD-fmk (**Z**).[45] Upon co-dosing polyamide **1** with **Z**, however, we saw no decrease in CRT exposure (Figure 3.5C). This caspase-independence was unexpected in light of literature precedent as well as the cytotoxicity we observed of **1** during flow cytometry experiments. We confirmed the lack of PARP cleavage, a caspase substrate cleaved during apoptosis, by immunoblot (Figure 3.5D). We then directly measured caspase activity with a luciferase assay after treatment with hairpin polyamides **1** and **2** as well as **Eto** and **Dox**. We measured a slight increase, approximately 1.2-fold, upon treatment with **1** for 24 hours and no increase after treatment with **2**. In contrast, we measured significant increases of approximately 15-fold each after treatment with **Eto** and **Dox** (Figure 3.6A). The lack of caspase-dependence and caspase activation would suggest the trigger of CRT by **1** is different from anthracyclin DNA intercalators.

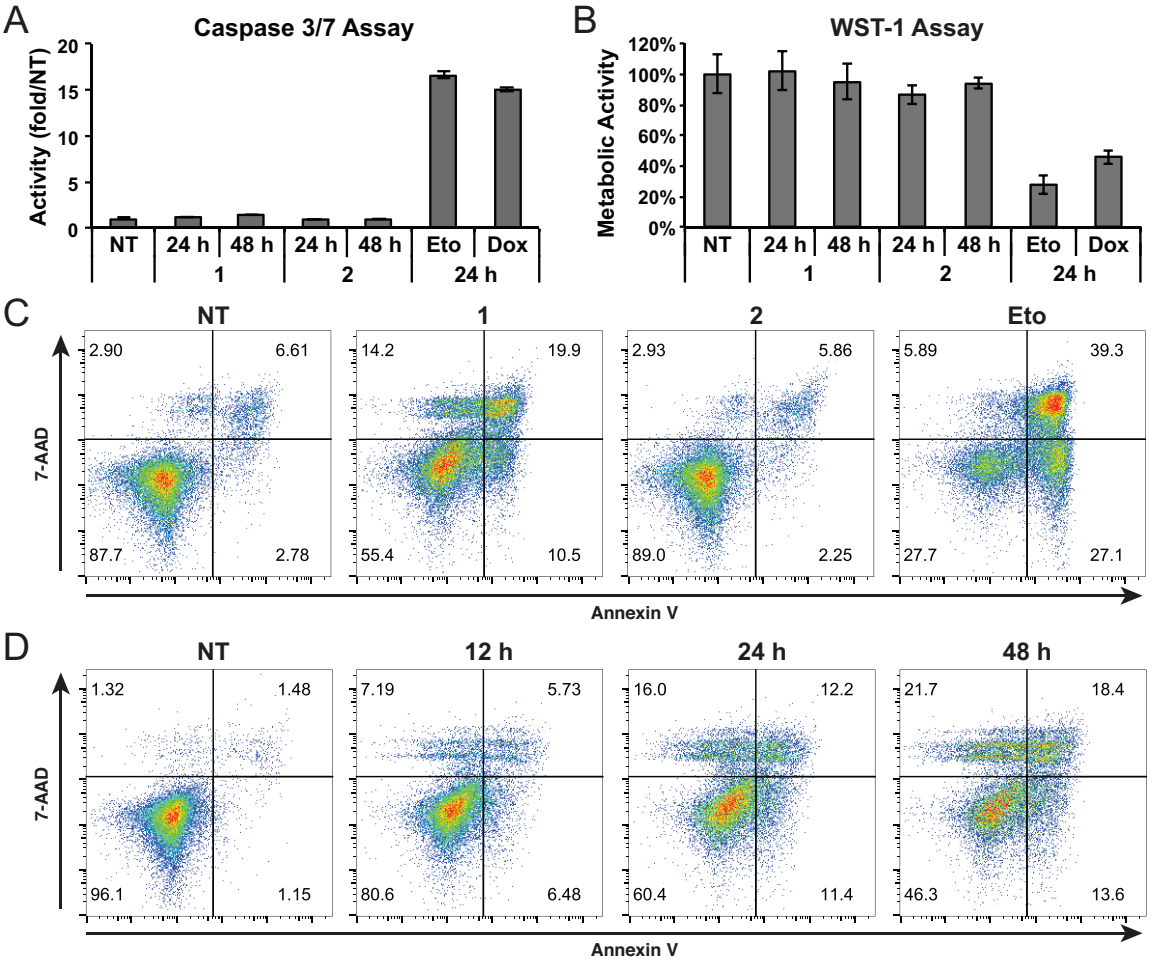


Figure 3.6 Raji cells undergo a slow, necrotic type cell death after treatment with polyamide **1**. A) Caspase 3/7 activity was assessed by a luciferase assay after the indicated treatment. Measurement is representative of two independent experiments and error bars show standard deviations of triplicate measurement. B) Cellular metabolism as a proxy for cytotoxicity was measured with a WST-1 assay. Cytotoxic controls etoposide (**Eto**, 30 μM) and doxorubicin (**Dox**, 5 μM) were included. Measurements were normalized to non-treatment. Graph is representative of two independent experiments and error bars represent standard deviations of technical quadruplicate. C) Flow cytometry assessment of Raji cells treated with the indicated compounds for 24 hours and stained for plasma membrane permeability (7-AAD) and phosphatidylserine exposure (Annexin V). Live: lower left; Early necrotic: upper left; Secondary necrotic: upper right; Apoptotic: lower right. Representative plots shown of triplicate measurement from two independent experiments. D) Assessment as in (B) after 12, 24, and 48 hour exposure to polyamide **1** at 25 μM. Error bars represent standard deviation of triplicate measurement of two independent experiments.

Polyamide 1 induces a slow necrotic-type cell death

We measured cytotoxicity via the metabolic rate with the WST-1 reagent in a colorimetric assay. We were surprised to find that treatment with **1**, at both 24 and 48 hours, had little effect on the bioreduction of WST-1 to the formazan dye, suggesting no diminution of metabolic rate (Figure 3.6B). This is in stark contrast to the cytotoxic **Eto** and **Dox**, which showed drastic reduction in cellular metabolism consistent with the apoptotic program. This led us to assess whether polyamide **1** may direct the Raji cells towards programmed necrosis. Necrosis is an immunogenic, inflammatory type of cell death that is usually attributed to harsh physical insult such as freeze-thaw cycles. Recently, a biologically-controlled necrosis program has been described where cells undergo early plasma membrane permeabilization with active inflammatory signaling and an oxidative burst.[47] We compared the mode of cell death after 24 hour treatment with hairpins **1** and **2**, or **Eto** by flow cytometry. We analyzed plasma membrane permeability and phosphatidylserine exposure on the cell surface to determine subpopulations of live, necrotic, and apoptotic cells after compound exposure. Plasma membrane permeability was assessed through 7-AAD dye exclusion and phosphatidylserine was visualized by dye-conjugated annexin V binding. We found that the non-treated control and treatment with hairpin **2** at 24 hours resulted in little cell death, but treatment with **1** or **Eto** caused significant cell death (Figure 3.6C). Remarkably, there was a distinct and substantial population of cells treated with **1** that lost plasma membrane integrity without phosphatidylserine exposure, indicative of a

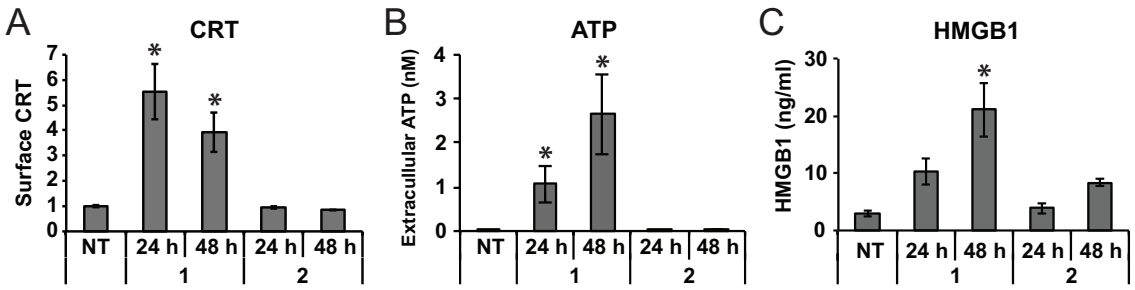


Figure 3.7 Immunogenic signaling is triggered by polyamide **1** treatment. A) CRT was measured by flow cytometry after 24 and 48 h treatment with **1** or **2** at 25 μM. B) Extracellular ATP was measured by a bioluminescence assay after the same treatment. C) HMGB1 in the supernatant was measured by ELISA after treatment with **1** or **2** at 25 μM. CRT flow cytometry and HMGB1 ELISA were measured in triplicate from at least two independent experiments. ATP was measured in quadruplicate in three independent experiments. Error bars are standard deviations and statistically significant increases (two tailed Student's t-test, $p<0.05$) compared to non-treatment are indicated by an asterisk.

necrotic cell death. We observed a continuous population from the upper left to the upper right quadrant in cells treated with polyamide **1**, suggestive of progressive phosphatidylserine exposure in permeabilized cells. The upper right quadrant in all treatment conditions is secondary necrosis, as there are no phagocytes to remove apoptotic or necrotic bodies. There are nearly no cells in the etoposide-treated condition that lie in the upper left quadrant, reflective of a canonical apoptotic cell death with intact membranes displaying phosphatidylserine. The anthracyclins could not be used in this assay due to their inherent fluorescence. We further did a time course of the necrotic effect of polyamide **1** at 12, 24, and 48 hours and found that the cells began permeabilization as early as 12 hours, and underwent further necrosis by 48 hours (Figure 3.6D). These assays together suggest polyamide **1** preferentially tracks Raji cells towards a slow, necrotic-type cell death even though the cells are capable of apoptosis.

Polyamide 1 triggers the release of DAMPs

The inflammatory nature of necrotic cell death can be attributed in part to the release of immune-activating DAMPs.[47,48] We measured levels of CRT, ATP, and HMGB1 after polyamide **1** treatment, as these DAMPs have been described as key factors in a spatiotemporal code of immunogenicity.[13,14] We included polyamide **2** as an in-class control because **2** has been extensively studied in other cancer models.[33] We measured CRT by flow cytometry after 24 and 48 hours of compound exposure at 25 μM and observed a decreasing trend for surface CRT (Figure 3.7A), consistent with reports that CRT is an early response signal.[14] Even at the 25 μM concentration of polyamide **2**, we detected no increase of CRT after 24 hour treatment. Extracellullar ATP was measured with a bioluminescence assay of the supernatant after the same treatment regime (Figure 3.7B). The ATP detected in the media after treatment with hairpin **1** was greatly increased over the non-treated condition. We next analyzed the release of HMGB1 by running a sandwich ELISA for HMGB1 in the supernatant. After treatment for 24 or 48 hours with **1**, a significant increase in excreted HMGB1 was detected in the media (Figure 3.7C). The polyamide **1**-mediated release of CRT, ATP, and HMGB1 in this temporal pattern is consistent with immunogenic signaling described in the literature.

Polyamide 1 treated Raji cells are subject to phagocytosis

Necrotic cell death with an abundance of externalized CRT and other released immunogenic DAMPs should increase phagocytosis. We obtained human peripheral blood macrophages and incubated them for 2.5 hours with Raji cells treated with **1** or **2** at

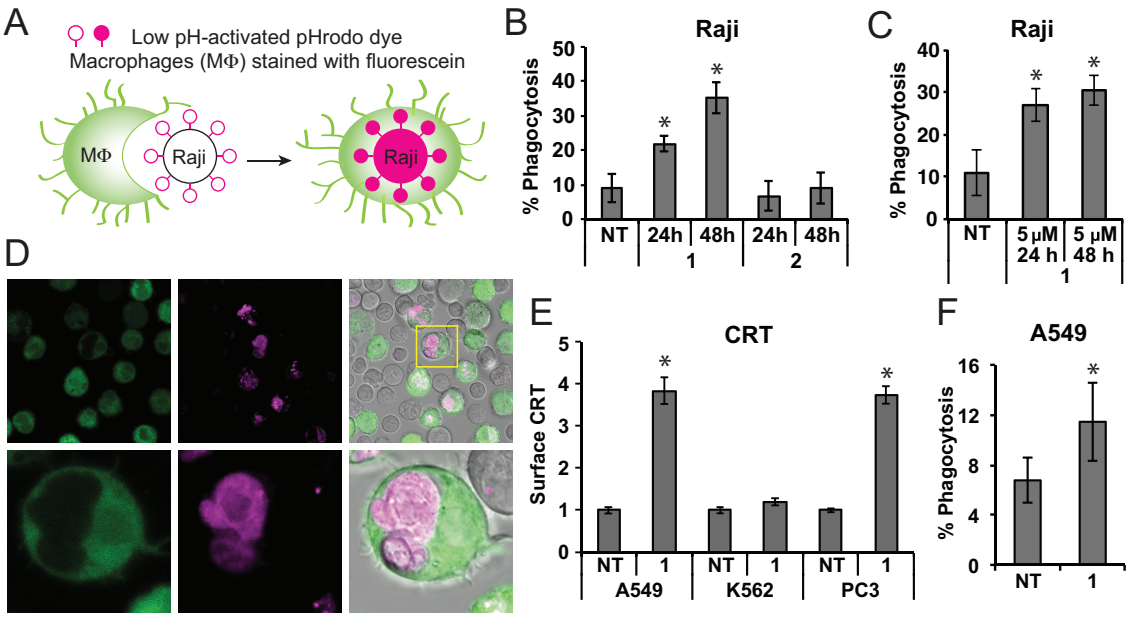


Figure 3.8 Treatment of cells with **1** increases phagocytosis by human macrophages. A) Schematic diagram of fluorescence-revealed phagocytosis using the pH-sensitive pHrodo dye. B) Raji cells treated with the indicated compound at 25 µM for 24 or 48 hours were incubated with human macrophages for 2.5 hours. Fluorescein+ cells were assessed by flow cytometry for double-positive cells to determine % phagocytosis. C) Raji cells were treated with a lower dose of 5 µM of **1** for 24 and 48 hours and assessed for phagocytosis by macrophages in the same manner as in (B). D) Fluorescent images of cells prepared as in B) are shown. Human macrophages are marked green, free Raji cells are grey in the brightfield composite image, and magenta-colored bodies inside macrophages are phagocytosed Raji cells. Yellow box in top row is magnified in bottom row images. E) The cell lines A549, K562, and PC3 were treated with polyamide **1** for 24 hours and screened by flow cytometry for surface CRT. Measurements were in technical triplicate. F) A549 lung carcinoma cells treated with **1** for 24 hours was subjected to human macrophages as described above. Graphs show mean and standard deviation from three independent experiments for (B), (C), (E), and (F). Asterisks indicate statistically significant increases (Student's two tailed t-test, $p<0.05$) compared to the non-treated condition. Images in (d) are representative of two independent experiments.

25 µM for 24 or 48 hours. The Raji cells' plasma membrane was covalently decorated with a pH-sensitive dye (pHrodo) that is activated when engulfed. Macrophages were marked with a cell-permeable fluorescein dye (Figure 3.8A). Macrophages that phagocytosed Raji cells are double positive for fluorescein and activated pHrodo dye. In each experiment, 1000 macrophages were analyzed by flow cytometry per condition. Compared to non-treatment, we found that a greater percentage of macrophages

incubated with polyamide **1**-treated Raji cells were double positive, and more so at 48 than 24 hours (Figure 3.8B). We saw little change when macrophages were incubated with cells treated with **2**. This indicates macrophages phagocytosed Raji cells more efficiently after treatment with polyamide **1**. We also tested the lower concentration of 5 μM of **1** used in the library screening assay, anticipating that immunogenicity may be a threshold effect and that 5 μM may be a more reasonable *in vivo* concentration. We found that the treatment of Raji cells with this lower concentration still increased the percentage of phagocytosis by macrophages (Figure 3.8C). We then verified the phagocytosis assay results with confocal microscopy to ensure that Raji cells were indeed phagocytosed rather than externally adherent (Figure 3.8D). The images show that a fraction of macrophages have internalized Raji cells, which fluoresce with the activated pHrodo dye. The brightfield composite image shows free Raji cells in grey, which are not engulfed by macrophages and lack a robust pHrodo signal. The bottom row shows a magnification of a macrophage after phagocytosis of a large Raji cell body.

Polyamide 1 triggers CRT and phagocytosis in other cancer cell lines

We next explored whether this effect might be specific to the Raji B cell lymphoma model or more general in other cancer cell lines. We screened the A549 lung adenocarcinoma cell line, K562 B cell leukemia cell line, and the PC3 prostate cancer cell line for CRT translocation after treatment with **1**. We observed approximately four-fold increase in CRT surface expression in the A549 and PC3 cell lines (Figure 3.8E). Surprisingly, we did not observe an increase in the K562 B cell leukemia cell line (Figure 3.8E) though they are closest in lineage with the Raji B cells. We then tested the A549

cell line in the phagocytosis assay after 24 hour treatment with 25 μM polyamide **1**. Even though the adherent A549 cell line presents a more difficult target for macrophages, as compared to non-adherent Raji cells, we measured an observable increase in A549 phagocytosis after treatment with polyamide **1** (Figure 3.8F). These results show the effects of polyamide **1** extend beyond the Raji cell line to cells of different lineages but are not completely general.

RNA-seq of Raji cells after treatment with polyamide 1 or 2

To better understand these effects in Raji cells, we examined the change in global gene expression after treatment with the polyamides for 24 hours. We isolated RNA from Raji cells treated with either polyamide **1** or **2**, or untreated and sequenced the transcriptome by next-generation sequencing.[49] The sequenced reads were mapped by Bowtie and Tophat to the Ensembl genes of human genome build 19.[50,51] The non-treated, polyamide **1**-treated, and polyamide **2**-treated samples were done in biological triplicate. The nine total datasets were input into the Cuffdiff algorithm, which identifies differentially expressed genes between datasets, accounting for replicates.[52]

In the polyamide **1**-treated samples, a total of 576 differentially regulated Ensembl genes were identified, of which 528 mapped to unique HGNC genes due to the inclusion of haplotypes in Ensembl. Of these, 236 genes were measured to be significantly upregulated, of which 154 were upregulated by 25% or greater from the non-treated condition (Table 3.1). There were 292 unique HGNC genes downregulated by polyamide **1**, where 64 were changed by more 25% or greater (Table 3.1). A smaller

Table 3.1 Summary of significant, differentially regulated genes from Cuffdiff analysis of RNA-Seq data.

		Ensembl	Unique HGNC Gene	Unique 25% change
Polyamide **1**	Upregulated	278	236	154
	Downregulated	298	292	64
Polyamide **2**	Upregulated	122	121	21
	Downregulated	118	118	47

set of genes were changed by polyamide **2** treatment, with 240 total Ensembl gene expression changes found significant, which mapped to 239 unique HGNC genes (Table 3.1). Of these, 121 genes were upregulated, of which 21 were upregulated by 25% or greater. Of 118 genes downregulated by polyamide **2**, 47 were downregulated by 25% or greater (Table 3.1). A full list of genes up- or down- regulated by 25% or greater upon polyamide **1** or **2** treatment is in Tables 3.2-3.5 at the end of the chapter.

There were several genes of note upregulated upon treatment with polyamide **1**. The gene for tumor necrosis factor (TNF), a major cytokine that regulates inflammation, was upregulated two-fold (Table 3.2).[53] Additionally, the genes HLA-DMA, HLA-DMB, HLA-DRB, and HLA-DQB, which encode the major histocompatibility complex (MHC) class II and accessory factors, were found to be increased ~40-50% (Table 3.2). HLA-E, which codes a non-classical MHC class I molecule that is a ligand for a natural killer cell receptor, was also found to be increased over 40%.[54] The MHC molecules are critical components of antigen recognition in the immune system as they present peptides to T cells in the context of a molecular structure unique to the individual.[55,56] The upregulation

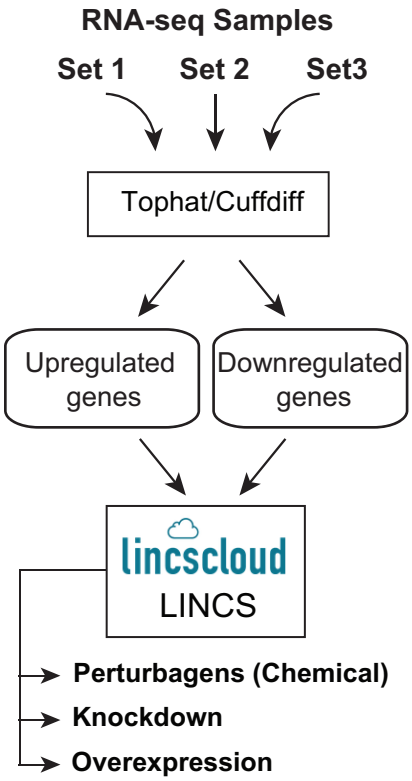


Figure 3.9 Scheme of RNA-Seq analysis pipeline.

of these genes is consistent with the inflammatory signaling observed in the previous assays. The intermediate early genes EGR1, EGR2, EGR4, and FOS are strongly upregulated as well, but given the wide scope of their effects in development, differentiation, and cell death, their role in this case is less clear.[57-59] These genes are markedly absent from the list of genes modulated by polyamide **2**, suggesting that this is not a class effect, but unique to hairpin polyamide **1**.

Genes found to significantly differ from the non-treated condition were then submitted for analysis by a cloud-based program created by the Library of Integrated Network-based Cellular Signatures (LINCS) to compare gene signatures with those in its

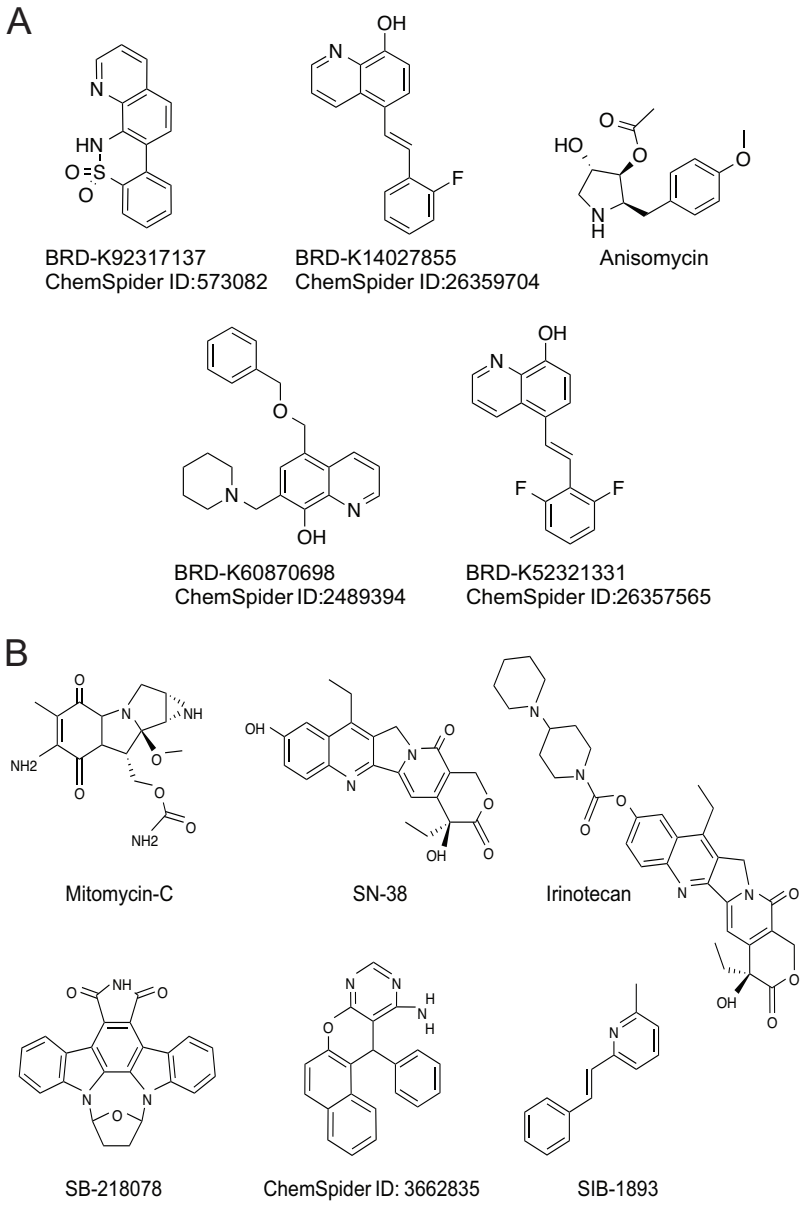


Figure 3.10 Structures of compounds identified by LINCS to have gene expression profiles similar to A) polyamide **1** and B) polyamide **2**.

database of human cells treated with chemical or genetic agents (Figure 3.9).[60] The structures of the top chemical hits in the LINCS analysis for polyamide **1**, shown in Figure 3.10A, are not well studied compounds and were not very informative. In contrast, several cytotoxic DNA binding molecules were identified for polyamide **2**, including the topoisomerase inhibitor irinotecan and its active metabolite SN-38 (Figure 3.10B). This is

in agreement with previous analyses of polyamide **2** in the LREX and VCaP prostate cancer cell lines (Yang, Nickols, Dervan; unpublished) and highlights the differences in biological activity between these two molecules.

The LINCS algorithm also compares the up- and down- regulated gene expression profile with experiments where single genes were overexpressed. The top "overexpression" signatures identified by LINCS included genes that encode MAP2K6, GADD45A, TIRAP, and interferon-β. MAP2K6 is a TNF-responsive kinase of p38 MAPK that integrates stress signals and cytokine signaling.[61] GADD45A is a member of a family of stress sensors that is responsive to genotoxic stress and affects cell cycling, DNA repair, and cell death.[62] TIRAP is a signal transducing molecule for the TLR2 and TLR4 receptors, which activate the innate immune response.[63] Interferon-β is a cytokine often produced in response to viral infection that primes adaptive immunity.[64] These expression signatures include aspects of immune activation and cellular or genotoxic stress signaling and may provide leads towards understanding the mechanism of polyamide **1** activity.

3.3 Discussion

In this study, we have discovered a Py-Im polyamide capable of triggering the release of immunogenic signals in a necrotic-type cell death. Treatment with polyamide **1** increased the externalization of CRT, HMGB1, and ATP, which have been characterized to be key signals of immunogenicity. We further observed that polyamide **1** causes permeabilization of the plasma membrane in a subpopulation of treated Raji cells, which

grows from 12 to 48 hours. This occurs even as the apoptotic cell death pathway remains competent, as demonstrated by the distribution profile observed after **Eto** treatment of Raji cells. We conclude that polyamide **1** preferentially triggers a necrotic-type, likely highly immunogenic cell death, an effect we are not aware has precedent with a small molecule synthetic ligand.

Phagocytosis by antigen-presenting cells is a critical step in a chain of events towards leveraging the specificity and power of the immune system towards attacking the offending target. We saw increased phagocytosis of polyamide **1**- treated cells by macrophages, as would be expected upon the death of Raji cells with externalized pro-phagocytic CRT. The observed phagocytosis and inflammatory DAMPs released into the tumor microenvironment makes plausible that the immune system would become primed towards antigens of Raji cells. Reports have shown engulfment by macrophages can activate an effective anti-cancer T cell response.[11] The application of polyamide **1** on cancer in the context of a full immune system remains to be explored.

That the DNA sequence preference of **1** is identical with that of human telomeres was indicated by the pairing rules and confirmed by Bind-n-Seq. Whether the mechanism of action for polyamide **1** is through its sequence-specific DNA binding capacity remains an open question. Though polyamides have generally been characterized to act in a DNA-binding mode, we cannot exclude that polyamide **1** may act as an aptamer binding to unknown target proteins. The translocation of CRT after polyamide **1** treatment was observed in multiple, though not all, cell lines and suggests there is a conserved pathway

that leads to this phenotype that is retained in many cancer cell lines. The RNA-Seq analysis showed that polyamide **1** triggered the expression of immune activating molecules. LINCS analysis of the transcriptomic data suggests similarity to several stress response and immune activating gene patterns. We expect the elucidation of the mechanism of action will reveal important signaling networks involved in immunosurveillance of cancer cells.

Conclusion

The effects of polyamide **1** are compelling because immunogenic cancer cell death is cleared in the natural setting and thus difficult to observe. Further study may reveal tumor suppressive signaling pathways that can be exploited to extrinsically control cancer even as cell-intrinsic mechanisms fail. The failure of anthracyclins to elicit CRT in Raji cells when they have previously been shown to be effective in CT26 murine colon cancer cells [9] underscores the heterogeneity of cancers. As cancers have heterogeneous mechanisms for evading elimination by the immune system, a corresponding diversity of immunogenicity agents will be important for both shedding light on new biology and developing immuno-oncologic therapies.

3.4 Materials and methods

Chemicals and Reagents. All chemicals were purchased from Sigma-Adlrich unless otherwise noted. Py-Im polyamides were synthesized by microwave-assisted, solid-phase synthesis on Kaiser oxime resin (Novabiochem) according to previously

Table 3.6. Mass spectrometry (MALDI-TOF) for Py-Im polyamides.

Polyamide	**Molecular formula**	**[Mass + H]**	**Found Mass**
ImImImPy-(R)$^{\alpha\text{-}NH2}$γ-PyPyPyPy-(+)-IPA (**1**)	$C_{64}H_{76}N_{23}O_{12}^{+}$	1358.6	1358.5
ImPyPyPy-(R)$^{\alpha\text{-}NH2}$γ-ImPyPyPy-(+)-IPA (**2**)	$C_{65}H_{77}N_{22}O_{12}^{+}$	1357.6	1357.1
ImImPyPy-(R)$^{\alpha\text{-}NH2}$γ-ImPyPyPy-(+)-IPA (**3**)	$C_{64}H_{76}N_{23}O_{12}^{+}$	1358.6	1358.0
CtPyPyIm-(R)$^{\alpha\text{-}NH2}$γ-PyImPyPy-(+)-IPA (**4**)	$C_{64}H_{73}ClN_{21}O_{12}S^{+}$	1394.5	1394.1
PyImPyIm-(R)$^{\alpha\text{-}NH2}$γ-PyImβIm-(+)-IPA (**5**)	$C_{60}H_{74}N_{23}O_{12}^{+}$	1308.6	1309.0
ImImIm-(+)-IPA (**6**)	$C_{30}H_{38}N_{11}O_{6}^{+}$	648.3	648.8
ImImImPy-(R)$^{\alpha\text{-}NH2}$γ-PyPyPyPy-(+)-4PEG-Biotin (**1b**)	$C_{77}H_{107}N_{26}O_{16}S^{+}$	1683.8	1683.1

described protocols. [65,66] Polyamide **5** was synthesized on hydrazine resin (855037, Novabiochem) in the same manner as the other polyamides and cleaved from resin in the same manner after 10 minute oxidation of the hydrazine by Cu(II)SO_4 in pyridine and DMF. Polyamides were purified by reverse-phase HPLC and lyophilized. EZ-Link NHS-PEG4-biotin (Pierce) was conjugated in 5% Hünig's base in DMF. Purity and identity of compounds were verified by analytical HPLC and matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry (Table 3.6).

Brefeldin A was purchased from BD biosciences, 7-AAD from eBiosciences), and Z-VAD-fmk from Promega. Antibodies purchased from Abcam are: polyclonal rabbit anti-calreticulin (ab2907), polyclonal rabbit anti-ERp57 (ab10287), and polyclonal goat anti-rabbit Alexa Fluor 488 (ab150081). Antibodies from Life Technologies are: anti-Annexin

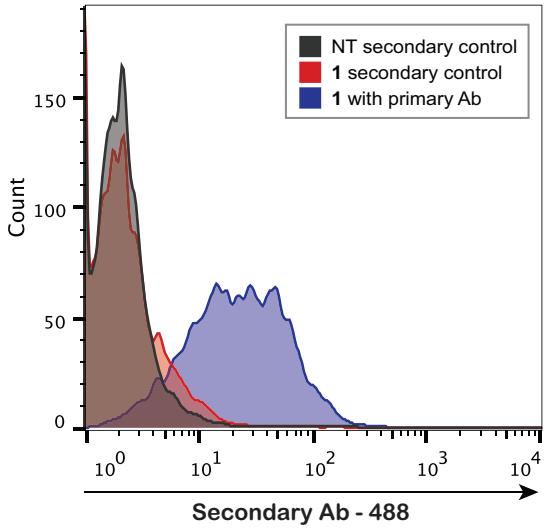


Figure 3.11 Specificity control for CRT primary antibody.

V Alexa Fluor 488 conjugate (A13201) and mouse monoclonal anti-actin (AM4302). Antibodies from Cell Signaling Technologies are: rabbit polyclonal anti-PARP (9542) and rabbit monoclonal anti-phospho-H2AX (ser139, 9718).

Cell culture. Raji cells (ATCC) were maintained in RPMI media (Life Technologies) with 10% fetal bovine serum (Omega Scientific) and 5mM glutamine (Life Technologies). K562 cells (ATCC) were cultured in IMDM (ATCC) media and 10% FBS. A549 and PC3 cells (ATCC) were maintained in Kaign's modified F-12K media (Gibco) supplemented with 10% FBS and 5 mM glutamine. Peripheral blood macrophages (Stemcell Technologies) were thawed and plated in DMEM (ATCC) and supplemented with 5 mM glutamine and 10% FBS for at least two days and no longer than six days before use.

Flow cytometry. Raji and K562 cells were plated at 1×10^5 cells/ml in 96-well plates at 100 µl and treated immediately after plating. A549 and PC3 cells were plated in

Table 3.7. Cell death flow cytometry after different treatments for 24 hours.

	Live	Necrotic	Apoptotic	2º Necrotic
NT	86.9 (±1.6)	2.0 (±0.4)	5.8 (±1.3)	5.3 (±0.3)
1	57.0 (±2.2)	11.9 (±1.1)	12.8 (±1.2)	18.4 (±1.2)
2	88.1 (±1.0)	2.3 (±0.3)	5.2 (±0.9)	4.5 (±0.2)
Eto	26.7 (±0.7)	5.3 (±0.2)	32.7 (±1.1)	35.3 (±1.0)

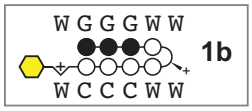
Table 3.8. Timecourse of cell death flow cytometry after treatment with **1**.

	Live	Necrotic	Apoptotic	2º Necrotic
NT	95.7 (±0.5)	1.7 (±0.4)	1.1 (±0.1)	1.5 (±0.1)
12 h	47.5 (±1.1)	19.6 (±2.0)	14.4 (±1.1)	18.5 (±0.5)
24 h	59.7 (±4.8)	15.6 (±1.0)	11.8 (±1.0)	12.9 (±3.0)
48 h	79.4 (±1.5)	7.7 (±1.1)	6.7 (±0.3)	6.2 (±0.4)

24-well plates at 6×10^4 cells/ml at 400 µl and were allowed to adhere overnight prior to treatment. Adherent cells were harvested with Accutase (Life Technologies).

For the CRT and ERp57 detection experiments, harvested cells were washed with cold staining buffer: HBSS (Life Technologies) with 2.5mg/ml BSA (Amresco), 10mM HEPES (Life Technologies), 0.1 mM $MgCl_2$ (Life Technologies), and DNAse (Roche). Cells were incubated with human FcR block (eBiosciences) and primary antibody in cold staining buffer. This was followed by washing and incubation with dye-conjugated anti-rabbit secondary antibody. 7-AAD, forward- and side- scatter were used to gate for live cells. Secondary antibody alone was used as a control (Figure 3.11). The median

Table 3.9 Table of top motifs and associated e-values for three Bind-n-Seq analyses of biotin-conjugate polyamide **1b**.



Concentration	Primary motif	E-value	Secondary motif	E-value
250 nM		2.6e-5582		4.5e-207
25 nM		7.3e-509		1.1e-028
25 nM		3.9e-129		9.1e-008

fluorescence intensity of live cells is normalized to non-treatment. Data was acquired on the FACS Calibur and analyzed with Flowjo software. Measurements and standard deviations are in triplicate and representative of at least two independent experiments; asterisk indicates $p<0.05$ by two-tailed Student's t-test compared to non-treatment.

For cell death analysis flow cytometry experiments, Raji cells were harvested after treatment and washed with cold staining buffer and incubated with FcR block and dye-conjugated anti-Annexin V. Cells were washed with staining buffer and taken for immediate analysis. 7-AAD was added to samples maintained at 4 ºC and data was acquired soon thereafter on the MACS VYB. Data was acquired in triplicate from two independent experiments and analyzed with Flowjo (Table 3.7 and 3.8).

Bind-n-Seq of polyamide 1b. Highest affinity DNA binding sites of polyamide-biotin conjugate **1b** were determined according to previously reported methods. [43] In brief, polyamide **1b** was equilibrated at 25 or 250 nM overnight in a DNA library of all possible 21mers. DNA bound to **1b** was affinity purified with streptavidin magnetic beads (M-280 Dynabeads, Life Technologies) and eluted. Isolated DNA was amplified by touchdown PCR and submitted for sequencing on an Illumina HiSeq 2000 Genome Analyzer. The MERMADE script was used to distribute data by barcode and a fasta file of a random 25% of sequences was generated for DREME motif analysis. Data is representative of three independent experiments (Table 3.9).

DNA thermal denaturation assay. DNA of the sequence 5'-CTTAGGGTTAGC-3' and its complement were purchased from Integrated DNA Technologies and annealed. The oligonucleotides were mixed with hairpin polyamide **1** to a final concentration of 2 and 3 μM, respectively, in 1 mL total volume. An aqueous solution of 10 mM sodium cacodylate, 10 mM KCl, 10 mM MgCl2, and 5 mM CaCl2 at pH 7.0 was used as analysis buffer. The assay utilized a Varian Cary 100 spectrophotometer to heat samples to 90 °C, cool to a starting temperature of 25 °C, and then heat at a rate of 0.5 °C/min to 90 °C. Denaturation profiles were recorded at $\lambda = 260$ nm and melting temperatures were detected by the maximum of the first derivative. Data shows the combined mean and standard deviation of triplicate measurements from two independent experiments.

Immunoblot assays. Raji cells were plated at 10^5 cells/ml in 10 cm diameter dishes and dosed with the indicated treatment. Cells were washed with cold PBS and lysed for ten minutes in lysis buffer (50 mM Tris-HCl pH 7.4, 1 mM EDTA, 150 mM

NaCl, 1% Triton X-100) containing protease inhibitors (Complete, Roche), 1 mM PMSF, and phosphatase inhibitors. Samples were clarified by centrifugation at 14,000 × g for ten minutes, quantified with the Bradford reagent (Bio-rad), denatured by boiling in Laemmli buffer (LI-COR) for 5 min, and separated by sodium dodecyl sulphate-polyacrylamide gel electrophoresis using AnyKD gradient gels (Biorad). Protein was transferred to a PVDF membrane (Millipore) and blocked with Odyssey blocking buffer (LI-COR). Both primary antibodies and appropriate IR-dye conjugated secondary antibodies (LI-COR) were incubated in blocking buffer with 0.2% Tween. Anti-actin was used to control for equal loading and experiments were done in at least two independent biological replicates. Bands were visualized on a LI-COR Odyssey infrared imager.

ATP bioluminescence assay. Raji cells were plated into 96-well cell culture plates at 200 µl per well and 10^5 cells/ml, in quadruplicate per condition. After the indicated treatment, media and cells were transferred to a 96-well PCR plate and centrifuged at 130 × g for 5 minutes. The supernatant was collected for analysis of ATP content using an ATP bioluminescence kit (FLAA, Sigma). The assay was performed according to the manufacturer's protocol and luminescence was measured using a Flexstation 3. Measurements and standard deviation are technical quadruplicate and biological triplicate. Asterisk indicates $p<0.05$ by two-tailed Student's t-test compared to non-treatment.

HMGB1 ELISA. Raji cells were plated into 96-well cell culture plates at 100 µl per well, 10^5 cells/ml, in duplicate. Cells were treated with polyamide **1** and **2** as indicated and the supernatant collected. HMGB1 was measured using the Shino-Test ELISA kit (IBL international) according to the manufacturer's instructions on the

Flexstation 3. Measurement and standard deviations were determined from technical duplicate from two independent replicates.

Caspase luciferase assay. Raji cells were plated into 96-well cell culture plates at 100 µl per well and 10^5 cells/ml, in triplicate per condition. Media was included as a blank control. After the indicated treatment, media and cells were transferred to an opaque white 96-well plate containing 100 µl in each well of the caspase-dependent luciferase reagent, prepared per the manufacturer's instructions (Caspase-Glo 3/7, Promega). The mixture was left at room temperature for 1 hour prior to measurement on the Flexstation 3 (Molecular Devices). Measurements and standard deviations were determined in triplicate and done in biological duplicate.

Metabolic activity assay. Raji cells were plated into 96-well clear bottom cell culture plates at 100 µl per well and 10^5 cells/ml, in triplicate per indicated condition. Media was used as a background control. Metabolic activity was assessed using the WST-1 reagent (Roche) per the manufacturer's instructions. Measurements and standard deviations were determined in triplicate and done in biological duplicate.

Phagocytosis assay. Peripheral blood macrophages were plated at 2×10^4 cells per well in 24 or 48-well plates in DMEM. Immediately prior to use, macrophages were washed with HBSS and stained with carboxyfluorescein diacetate succinimidyl ester (CFSE, eBioscience) in HBSS for 5 min. Macrophages were washed of excess dye and returned to DMEM for incubation. Target Raji cells were plated in 96-well plates at 1×10^5 cells/ml in 200 µl of RPMI and treated as described. Target A549 cells were plated in 24-well plates at 2×10^4 cells per well 15 hours before beginning treatment. After treatment, cells were harvested with Accutase if necessary and washed with HBSS and

incubated with pHrodo succinimidyl ester (Life Technologies) diluted to 2 μM in HBSS for 5 min. Cells were then washed by centrifugation at 150 × g, re-suspended in DMEM, counted on a Biorad TC10 cell counter, and 5×10^4 cells per well were added to macrophages. After incubation for 2.5 hours at 37 ºC, media and non-adherent cells in each sample were aspirated and saved. Adherent cells were trypsinized, aspirated, and combined with saved media mixture. Cells were washed once with PBS and fixed in 1% formaldehyde and kept at 4 ºC until assessment on the MACS VYB flow cytometer. The percentage of phagocytosis was calculated as the percentage of double positive cells among fluorescein+ macrophages. Measurements and standard deviations are taken from three independent experiments, and asterisks indicate $p<0.05$ by two-tailed Student's t-test compared to non-treatment.

Confocal microscopy. For images of phagocytosis, cells treated in the manner described above were put on 35 mm glass-bottom dishes (MatTek). Confocal images were acquired using a 40× oil immersion objective on a Zeiss LSM 5 Exciter microscope.

RNA-Seq. Raji cells were plated in two 175 cm^2 flasks per condition at 4×10^5 cells per ml with 20 ml of RPMI in each flask. Non-treatment, 25 μM polyamide **1**-, and 25 μM polyamide **2**- treatment was prepared at the same time. After 24 hours incubation, cells were washed with PBS twice by centrifugation and pelleted in a 15 ml falcon tube. The cells were suspended in Trizol and lysed by repeated extrusion from a syringe needle and vortexing. The samples were frozen and stored in a -80 ºC freezer, until three biological replicates were collected. The Trizol solution was thawed and RNA was isolated from Trizol (Life Technologies), and precipitated with linear acrylamide (Life Technologies). The RNA was washed and subjected to DNase treatment (Roche) prior to

quantitation. Samples each containing 10 μg of RNA in 100 μl of TE buffer was submitted for sequencing to the Millard and Muriel Jacobs Genetics and Genomics Laboratory on the Illumina Genome Analyzer.

3.5 Acknowledgments

This work was supported by the National Institutes of Health Grant GM51747 and Tobacco-Related Disease Research Program (award number 20DT-0037 to J.S.K., dissertation research award). We thank Rochelle Diamond and Diana Perez of the Caltech Flow Cytometry Cell Sorting Facility for help with flow cytometry. Sequencing was conducted with the help of Dr. Igor Antoshechkin at the Millard and Muriel Jacobs Genetics and Genomics Laboratory at the California Institute of Technology; mass spectrometry analyses were performed in the Mass Spectrometry Laboratory of the Division of Chemistry and Chemical Engineering at the California Institute of Technology; gels were scanned in the Center for the Chemistry of Cellular Signaling at the California Institute of Technology.

Table 3.2 Genes upregulated greater than 25% after treatment with polyamide **1** for 24 hours compared to non-treatment.

Ensembl Gene	Gene	Locus	Ratio	NT fpkm	GGGW fpkm	p_val	q_val
ENSG00000122877	EGR2	chr10:64571755-64679660	20.33	0.10	2.09	0.0002	0.0109
ENSG00000120738	EGR1	chr5:137801178-137805004	9.85	3.01	29.67	0.0000	0.0000
ENSG00000170345	FOS	chr14:75745476-75748933	6.90	1.96	13.54	0.0000	0.0006
ENSG00000027869	SH2D2A	chr1:156776034-156851642	5.08	0.40	2.03	0.0001	0.0088
ENSG00000135625	EGR4	chr2:73518057-73520833	4.80	0.18	0.87	0.0000	0.0010
ENSG00000135363	LMO2	chr11:33880121-33913836	4.25	8.12	34.53	0.0000	0.0000
ENSG00000110848	CD69	chr12:9905081-9913497	3.04	2.71	8.24	0.0000	0.0000
ENSG00000185022	MAFF	chr22:38597888-38612518	3.01	1.95	5.89	0.0000	0.0012
ENSG00000111679	PTPN6	chr12:7055630-7070479	3.00	82.60	247.93	0.0000	0.0000
ENSG00000059804	SLC2A3	chr12:7917811-8250367	2.95	4.66	13.74	0.0000	0.0000
ENSG00000127124	HIVEP3	chr1:41972035-42501596	2.70	2.37	6.40	0.0000	0.0000
ENSG00000100985	MMP9	chr20:44637546-44688789	2.55	0.51	1.31	0.0000	0.0032
ENSG00000227039	ITGB2-AS1	chr21:46305867-46351904	2.54	3.55	9.02	0.0008	0.0362
ENSG00000166886	NAB2	chr12:57482676-57607134	2.51	22.08	55.39	0.0000	0.0000
ENSG00000197629	MPEG1	chr11:58938902-58980424	2.47	0.50	1.24	0.0000	0.0000
ENSG00000168209	DDIT4	chr10:74033677-74035794	2.43	72.45	176.07	0.0000	0.0000
ENSG00000228978	TNF	chr6_apd_hap1:2859028-2860831	2.31	1.07	2.46	0.0001	0.0053
ENSG00000204490	TNF	chr6_cox_hap2:3052959-3055729	2.26	1.06	2.39	0.0000	0.0010
ENSG00000206439	TNF	chr6_qbl_hap6:2836981-2839753	2.26	1.06	2.39	0.0000	0.0010
ENSG00000223952	TNF	chr6_mcf_hap5:2923049-2925819	2.26	1.06	2.39	0.0000	0.0010
ENSG00000228321	TNF	chr6_mann_hap4:2886227-2888997	2.26	1.06	2.39	0.0000	0.0010
ENSG00000228849	TNF	chr6_dbb_hap3:2828884-2831654	2.26	1.06	2.39	0.0000	0.0010
ENSG00000230108	TNF	chr6_ssto_hap7:2874144-2876914	2.26	1.06	2.39	0.0000	0.0010
ENSG00000232810	TNF	chr6:31543343-31546113	2.26	1.06	2.39	0.0000	0.0010
ENSG00000118515	SGK1	chr6:134490383-134639250	2.23	23.02	51.38	0.0000	0.0000
ENSG00000138166	DUSP5	chr10:112257595-112271302	2.21	18.34	40.60	0.0000	0.0000
ENSG00000108984	MAP2K6	chr17:67410837-67538451	2.21	2.04	4.50	0.0000	0.0000
ENSG00000247095	MIR210HG	chr11:565659-568457	2.10	12.69	26.61	0.0000	0.0003
ENSG00000161929	SCIMP	chr17:5095378-5151364	2.09	3.69	7.71	0.0000	0.0000
ENSG00000114268	PFKFB4	chr3:48555116-48601206	2.03	11.94	24.26	0.0000	0.0000
ENSG00000158050	DUSP2	chr2:96808904-96811179	1.97	69.35	136.93	0.0000	0.0000

ENSG00000090104	RGS1	chr1:192544856-192549161	1.90	3.88	7.39	0.0000	0.0002
ENSG00000160888	IER2	chr19:13261281-13265716	1.86	51.33	95.61	0.0000	0.0000
ENSG00000226979	LTA	chr6:31539830-31542101	1.82	1.98	3.60	0.0001	0.0070
ENSG00000126860	EVI2A	chr17:29335514-29709134	1.78	6.55	11.66	0.0000	0.0000
ENSG00000128016	ZFP36	chr19:39897486-39900045	1.71	13.43	23.01	0.0000	0.0000
ENSG00000160255	ITGB2	chr21:46305867-46351904	1.70	57.51	97.85	0.0000	0.0000
ENSG00000137101	CD72	chr9:35605366-35647096	1.68	37.22	62.43	0.0000	0.0000
ENSG00000185650	ZFP36L1	chr14:69254376-69263190	1.67	182.58	305.21	0.0000	0.0000
ENSG00000125245	GPR18	chr13:99853027-100038688	1.65	3.75	6.20	0.0002	0.0136
ENSG00000179344	HLA-DQB1	chr6:32627243-32636160	1.65	89.37	147.42	0.0000	0.0000
ENSG00000112137	PHACTR1	chr6:12717892-13328776	1.65	10.17	16.76	0.0002	0.0112
ENSG00000166068	SPRED1	chr15:38544526-38649450	1.64	7.65	12.55	0.0006	0.0270
ENSG00000226397	C12orf77	chr12:25146357-25150453	1.63	8.47	13.81	0.0000	0.0006
ENSG00000111678	C12orf57	chr12:7052140-7055166	1.62	36.86	59.71	0.0000	0.0000
ENSG00000156738	MS4A1	chr11:60223224-60238233	1.61	436.07	701.54	0.0000	0.0000
ENSG00000143851	PTPN7	chr1:202116140-202130716	1.61	44.45	71.43	0.0000	0.0000
ENSG00000154127	UBASH3B	chr11:122526382-122685181	1.60	23.67	37.92	0.0000	0.0004
ENSG00000087074	PPP1R15A	chr19:49375648-49379318	1.60	6.58	10.51	0.0000	0.0000
ENSG00000100290	BIK	chr22:43506753-43525718	1.59	50.13	79.62	0.0000	0.0000
ENSG00000204852	TCTN1	chr12:111051831-111142755	1.56	47.20	73.58	0.0000	0.0001
ENSG00000234883	MIR155HG	chr21:26934220-26947480	1.55	12.54	19.48	0.0000	0.0000
ENSG00000138821	SLC39A8	chr4:103172197-103371167	1.54	70.90	109.02	0.0000	0.0034
ENSG00000198873	GRK5	chr10:120967100-121215131	1.54	8.25	12.69	0.0004	0.0210
ENSG00000122884	P4HA1	chr10:74766974-74856732	1.53	30.26	46.31	0.0000	0.0000
ENSG00000142227	EMP3	chr19:48828628-48833810	1.52	19.49	29.70	0.0000	0.0000
ENSG00000184574	LPAR5	chr12:6728000-6745613	1.51	10.56	15.99	0.0000	0.0000
ENSG00000241674	HLA-DMB	chr6_qbl_hap6:4134532-4181419	1.50	12.75	19.15	0.0004	0.0194
ENSG00000242386	HLA-DMB	chr6_mcf_hap5:4239256-4283840	1.50	11.71	17.57	0.0008	0.0361
ENSG00000239329	HLA-DMB	chr6_mann_hap4:4359616-4406503	1.50	12.85	19.25	0.0004	0.0200
ENSG00000241296	HLA-DMB	chr6_ssto_hap7:4333159-4430060	1.50	11.81	17.66	0.0008	0.0361
ENSG00000196968	FUT11	chr10:75532048-75538821	1.49	23.20	34.68	0.0000	0.0022
ENSG00000109929	SC5D	chr11:121163161-121179403	1.49	52.42	78.34	0.0000	0.0002
ENSG00000086619	ERO1LB	chr1:236305831-236445319	1.49	10.14	15.14	0.0004	0.0202
ENSG00000134061	CD180	chr5:66478102-66492627	1.49	30.00	44.67	0.0000	0.0000
ENSG00000242574	HLA-DMB	chr6:32902405-32949282	1.49	6.62	9.86	0.0008	0.0364

ENSG00000226264	HLA-DMB	chr6_dbb_hap3:4183742-4202240	1.49	13.07	19.42	0.0005	0.0254
ENSG00000206493	HLA-E	chr6_qbl_hap6:1750158-1754897	1.48	15.79	23.41	0.0000	0.0001
ENSG00000229252	HLA-E	chr6_dbb_hap3:1750883-1755622	1.48	15.79	23.41	0.0000	0.0001
ENSG00000233904	HLA-E	chr6_cox_hap2:1969307-1974046	1.48	15.79	23.41	0.0000	0.0001
ENSG00000171223	JUNB	chr19:12902309-12904124	1.48	62.40	92.44	0.0000	0.0000
ENSG00000230254	HLA-E	chr6_ssto_hap7:1789533-1794272	1.48	16.26	24.09	0.0000	0.0000
ENSG00000242092	HLA-DMB	chr6_apd_hap1:4189389-4207888	1.48	13.01	19.22	0.0006	0.0285
ENSG00000122986	HVCN1	chr12:111051831-111142755	1.48	15.71	23.21	0.0002	0.0124
ENSG00000140563	MCTP2	chr15:94774766-95023632	1.47	8.01	11.80	0.0007	0.0334
ENSG00000236632	HLA-E	chr6_mann_hap4:1805321-1810060	1.46	11.29	16.53	0.0000	0.0013
ENSG00000204592	HLA-E	chr6:30457243-30461982	1.46	11.29	16.53	0.0000	0.0013
ENSG00000225201	HLA-E	chr6_mcf_hap5:1839107-1843846	1.46	11.29	16.53	0.0000	0.0013
ENSG00000146386	ABRACL	chr6:139349818-139364439	1.45	55.89	81.21	0.0000	0.0000
ENSG00000136286	MYO1G	chr7:45002260-45018697	1.45	16.13	23.43	0.0000	0.0000
ENSG00000154642	C21orf91	chr21:19135631-19639690	1.44	25.30	36.51	0.0000	0.0000
ENSG00000117394	SLC2A1	chr1:43391518-43424530	1.44	112.09	161.77	0.0000	0.0000
ENSG00000140398	NEIL1	chr15:75639295-75648087	1.44	31.03	44.64	0.0000	0.0019
ENSG00000240409	MTATP8P1	chr1:536815-660287	1.43	526.42	753.97	0.0000	0.0001
ENSG00000198355	PIM3	chr22:50354160-50357728	1.42	93.39	132.75	0.0000	0.0000
ENSG00000196405	EVL	chr14:100437785-100610573	1.42	71.85	101.83	0.0001	0.0064
ENSG00000198502	HLA-DRB5	chr6:32485119-32498064	1.40	21.77	30.50	0.0000	0.0001
ENSG00000239463	HLA-DMA	chr6_mcf_hap5:4239256-4283840	1.40	18.89	26.41	0.0002	0.0092
ENSG00000243215	HLA-DMA	chr6_ssto_hap7:4333159-4430060	1.40	18.93	26.43	0.0002	0.0094
ENSG00000243719	HLA-DMA	chr6_cox_hap2:4346929-4393826	1.40	18.95	26.46	0.0002	0.0096
ENSG00000243189	HLA-DMA	chr6_mann_hap4:4359616-4406503	1.40	18.86	26.33	0.0002	0.0091
ENSG00000242685	HLA-DMA	chr6_qbl_hap6:4134532-4181419	1.40	18.87	26.33	0.0002	0.0091
ENSG00000202198	RN7SK	chr6:52860417-52860748	1.40	83.52	116.54	0.0001	0.0067
ENSG00000168389	MFSD2A	chr1:40420801-40435638	1.40	21.97	30.64	0.0000	0.0001
ENSG00000204257	HLA-DMA	chr6:32902405-32949282	1.39	18.51	25.80	0.0002	0.0116
ENSG00000242361	HLA-DMA	chr6_apd_hap1:4189389-4207888	1.39	26.12	36.34	0.0004	0.0199
ENSG00000196126	HLA-DRB1	chr6:32546545-32557625	1.39	28.09	39.07	0.0000	0.0000
ENSG00000227357	HLA-DRB4	chr6_ssto_hap7:3850430-4111880	1.39	23.95	33.30	0.0000	0.0001
ENSG00000231021	HLA-DRB4	chr6_mcf_hap5:3882320-3897284	1.39	23.94	33.28	0.0000	0.0001
ENSG00000023330	ALAS1	chr3:52232101-52248343	1.39	36.53	50.77	0.0012	0.0495
ENSG00000196101	HLA-DRB3	chr6_qbl_hap6:3720953-3734041	1.39	65.61	91.18	0.0000	0.0000

ENSG00000241394	HLA-DMA	chr6_dbb_hap3:4183742-4202240	1.39	23.65	32.86	0.0007	0.0326
ENSG00000114480	GBE1	chr3:81538849-81811312	1.39	12.41	17.23	0.0000	0.0001
ENSG00000166128	RAB8B	chr15:63481667-63559981	1.39	20.63	28.62	0.0002	0.0117
ENSG00000229074	HLA-DRB1	chr6_mann_hap4:3951045-4028663	1.39	19.23	26.68	0.0000	0.0004
ENSG00000236884	HLA-DRB1	chr6_dbb_hap3:3820127-3854230	1.39	19.88	27.56	0.0000	0.0003
ENSG00000166848	TERF2IP	chr16:75661621-75795051	1.38	77.41	107.14	0.0000	0.0000
ENSG00000196735	HLA-DQA1	chr6:32595955-32614839	1.38	93.11	128.62	0.0000	0.0000
ENSG00000125753	VASP	chr19:46010687-46030236	1.38	52.54	72.51	0.0000	0.0000
ENSG00000206237	HLA-DQB1	chr6_cox_hap2:4064608-4080088	1.38	9.54	13.15	0.0000	0.0027
ENSG00000206302	HLA-DQB1	chr6_qbl_hap6:3851061-3866542	1.38	9.54	13.15	0.0000	0.0027
ENSG00000150347	ARID5B	chr10:63661058-63856703	1.38	4.51	6.21	0.0000	0.0001
ENSG00000132963	POMP	chr13:29233240-29253062	1.38	55.20	75.95	0.0000	0.0000
ENSG00000231679	HLA-DRB3	chr6_cox_hap2:3934114-3947232	1.38	50.41	69.35	0.0000	0.0000
ENSG00000124762	CDKN1A	chr6:36555310-36932613	1.37	50.55	69.47	0.0000	0.0000
ENSG00000251587	LDHAP1	chr4:4895908-4896900	1.37	25.34	34.76	0.0000	0.0006
ENSG00000145088	EAF2	chr3:121554029-121605373	1.37	34.00	46.63	0.0000	0.0000
ENSG00000172081	MOB3A	chr19:2071036-2096269	1.37	52.90	72.55	0.0000	0.0000
ENSG00000197324	LRP10	chr14:23340821-23350789	1.37	34.45	47.24	0.0001	0.0048
ENSG00000228080	HLA-DRB1	chr6_ssto_hap7:3850430-4111880	1.37	103.26	141.28	0.0000	0.0000
ENSG00000128040	SPINK2	chr4:57676025-57687908	1.36	19.08	25.94	0.0004	0.0209
ENSG00000125772	GPCPD1	chr20:5525084-5591672	1.35	39.26	53.18	0.0001	0.0056
ENSG00000081189	MEF2C	chr5:87803362-88762215	1.35	52.98	71.43	0.0000	0.0000
ENSG00000112972	HMGCS1	chr5:43287702-43313614	1.35	167.15	225.30	0.0000	0.0000
ENSG00000196923	PDLIM7	chr5:176910394-176938275	1.35	35.66	48.06	0.0000	0.0013
ENSG00000001630	CYP51A1	chr7:91741464-91810039	1.35	32.91	44.29	0.0000	0.0020
ENSG00000159840	ZYX	chr7:143078172-143220542	1.34	37.08	49.76	0.0001	0.0048
ENSG00000188765	TMSB4XP2	chr2:3642425-3692048	1.34	1022.50	1369.26	0.0009	0.0388
ENSG00000226417	CLIC1	chr6_mcf_hap5:3008685-3199991	1.34	27.76	37.16	0.0000	0.0000
ENSG00000226248	CLIC1	chr6_dbb_hap3:2980383-2993114	1.34	27.67	36.99	0.0000	0.0000
ENSG00000230685	CLIC1	chr6_cox_hap2:3204431-3217149	1.34	27.69	36.99	0.0000	0.0000
ENSG00000223639	CLIC1	chr6_ssto_hap7:3025631-3038358	1.34	27.69	36.99	0.0000	0.0000
ENSG00000206394	CLIC1	chr6_qbl_hap6:2988444-3001178	1.34	27.68	36.98	0.0000	0.0000
ENSG00000027697	IFNGR1	chr6:137518620-137540586	1.34	27.61	36.88	0.0000	0.0004
ENSG00000226651	CLIC1	chr6_mann_hap4:2971895-3163173	1.34	27.81	37.13	0.0000	0.0000
ENSG00000213719	CLIC1	chr6:31694814-31707540	1.33	27.72	36.98	0.0000	0.0000

ENSG00000236090	LDHAP3	chr2:42046886-42047894	1.33	25.97	34.64	0.0000	0.0027
ENSG00000131196	NFATC1	chr18:77155771-77289325	1.33	31.76	42.30	0.0000	0.0000
ENSG00000225855	RUSC1-AS1	chr1:155278538-155300905	1.33	12.88	17.15	0.0009	0.0377
ENSG00000185607	ACTBP7	chr15:44162961-44487450	1.33	11.08	14.72	0.0011	0.0459
ENSG00000172175	MALT1	chr18:56338617-56421130	1.32	17.57	23.28	0.0000	0.0000
ENSG00000206240	HLA-DRB1	chr6_cox_hap2:3998150-4028758	1.32	214.60	284.18	0.0000	0.0000
ENSG00000206306	HLA-DRB1	chr6_qbl_hap6:3784599-3815217	1.32	214.60	284.18	0.0000	0.0000
ENSG00000163219	ARHGAP25	chr2:68906732-69053965	1.32	33.68	44.49	0.0000	0.0012
ENSG00000107140	TESK1	chr9:35605366-35647096	1.32	50.48	66.47	0.0000	0.0000
ENSG00000235674	LDHAP2	chr1:235824340-236046940	1.32	50.00	65.75	0.0000	0.0005
ENSG00000214110	LDHAP4	chr9:14921334-14922332	1.31	210.58	276.09	0.0000	0.0000
ENSG00000067064	IDI1	chr10:1034337-1095110	1.31	100.10	130.86	0.0000	0.0000
ENSG00000138756	BMP2K	chr4:79697495-79860592	1.31	40.34	52.71	0.0000	0.0001
ENSG00000227671	MIR3916	chr1:247353152-247372795	1.31	25.01	32.66	0.0000	0.0000
ENSG00000152256	PDK1	chr2:173292081-173489823	1.30	36.41	47.51	0.0004	0.0221
ENSG00000235847	LDHAP7	chr2:84743578-85046713	1.30	48.15	62.76	0.0000	0.0012
ENSG00000007944	MYLIP	chr6:16129355-16148479	1.30	12.88	16.72	0.0000	0.0005
ENSG00000112343	TRIM38	chr6:25963029-25985348	1.30	8.83	11.45	0.0000	0.0029
ENSG00000175105	ZNF654	chr3:88101093-88217879	1.30	3.00	3.90	0.0008	0.0373
ENSG00000102144	PGK1	chrX:77320684-77384793	1.30	659.51	854.94	0.0000	0.0000
ENSG00000186480	INSIG1	chr7:155089485-155101945	1.30	278.05	360.10	0.0000	0.0000
ENSG00000213574	LDHAP5	chr10:120692185-120692807	1.29	100.13	129.62	0.0000	0.0012
ENSG00000105287	PRKD2	chr19:47150868-47220384	1.29	19.29	24.96	0.0000	0.0000
ENSG00000162927	PUS10	chr2:61169103-61276394	1.29	9.97	12.88	0.0000	0.0026
ENSG00000257473	HLA-DQA2	chr6_cox_hap2:4055570-4061793	1.29	94.49	121.63	0.0000	0.0000
ENSG00000206305	HLA-DQA1	chr6_qbl_hap6:3841980-3848260	1.29	98.01	126.08	0.0000	0.0000
ENSG00000104765	BNIP3L	chr8:26236774-26371608	1.28	33.08	42.33	0.0007	0.0321
ENSG00000101608	MYL12A	chr18:3247478-3261848	1.28	149.26	190.74	0.0000	0.0000
ENSG00000223551	TMSB4XP4	chr9:131102924-131123502	1.28	58.80	75.11	0.0003	0.0152
ENSG00000187653	TMSB4XP8	chr4:91048685-92523064	1.28	4409.61	5630.75	0.0002	0.0110
ENSG00000213290	PGK1P2	chr19:12670384-12671919	1.28	46.19	58.97	0.0000	0.0021
ENSG00000099840	IZUMO4	chr19:2096379-2099592	1.28	29.86	38.08	0.0007	0.0302
ENSG00000179335	CLK3	chr15:74890840-74988633	1.28	110.22	140.54	0.0000	0.0004
ENSG00000162511	LAPTM5	chr1:31205315-31230667	1.27	172.57	219.90	0.0000	0.0000
ENSG00000109046	WSB1	chr17:25621101-25640657	1.27	64.39	81.97	0.0000	0.0023

ENSG00000170949	ZNF160	chr19:53569866-53606687	1.27	5.83	7.42	0.0001	0.0081
ENSG00000008282	SYPL1	chr7:105730948-105753022	1.27	87.36	111.13	0.0000	0.0000
ENSG00000137642	SORL1	chr11:121318039-121504387	1.27	30.67	38.99	0.0001	0.0087
ENSG00000133985	TTC9	chr14:71108503-71142077	1.27	15.52	19.70	0.0000	0.0007
ENSG00000147168	IL2RG	chrX:70327253-70331958	1.27	206.38	261.83	0.0001	0.0036
ENSG00000119403	PHF19	chr9:123617976-123639606	1.27	103.89	131.76	0.0000	0.0000
ENSG00000228612	HK2P1	chrX:79827369-79830106	1.27	7.99	10.12	0.0008	0.0364
ENSG00000205542	TMSB4X	chrX:12993226-12995346	1.26	1091.60	1380.31	0.0000	0.0000
ENSG00000184009	ACTG1	chr17:79476996-79494802	1.26	1584.63	2001.96	0.0000	0.0000
ENSG00000149781	FERMT3	chr11:63974149-64001824	1.26	279.36	352.84	0.0000	0.0000
ENSG00000156675	RAB11FIP1	chr8:37716135-37756985	1.26	10.06	12.69	0.0000	0.0022
ENSG00000214297	ALDOAP2	chr10:127262939-127371713	1.26	127.14	160.32	0.0000	0.0005
ENSG00000196352	CD55	chr1:207494852-207534311	1.26	25.22	31.79	0.0000	0.0015
ENSG00000081320	STK17B	chr2:196998289-197041227	1.26	67.00	84.31	0.0000	0.0016
ENSG00000236876	TMSB4XP1	chr1:42895999-43120335	1.26	4349.16	5468.48	0.0001	0.0070
ENSG00000196604	POTEF	chr2:130831107-130886795	1.26	6.91	8.68	0.0003	0.0161
ENSG00000167460	TPM4	chr19:16178316-16213813	1.25	74.12	92.96	0.0000	0.0003
ENSG00000230043	TMSB4XP6	chr20:49411430-49493714	1.25	10568.20	13242.60	0.0000	0.0031
ENSG00000121064	SCPEP1	chr17:55055465-55084129	1.25	61.98	77.66	0.0000	0.0020
ENSG00000149925	ALDOA	chr16:30064410-30081778	1.25	2056.86	2576.58	0.0000	0.0000
ENSG00000114023	FAM162A	chr3:122103022-122134882	1.25	90.07	112.70	0.0000	0.0001
ENSG00000134046	MBD2	chr18:51679078-51751158	1.25	340.32	425.66	0.0006	0.0282
ENSG00000228253	MT-ATP8	chrM:8365-9990	1.25	20644.20	25740.60	0.0000	0.0005
ENSG00000102393	GLA	chrX:100645811-100669121	1.25	33.32	41.52	0.0008	0.0373
ENSG00000181577	C6orf223	chr6:43963459-44045689	1.25	26.19	32.62	0.0001	0.0040

Table 3.3 Genes downregulated greater than 25% after treatment with polyamide **1** for 24 hours compared to non-treatment.

Ensembl Gene	Gene	Locus	Ratio	NT fpkm	GGGW fpkm	p_val	q_val
ENSG00000258486	RN7SL1	chr14:50043389-50081390	0.00	19.56	0.00	0.0001	0.0070
ENSG00000214189	ZNF788	chr19:12203077-12225499	0.08	0.86	0.07	0.0000	0.0020
ENSG00000176020	AMIGO3	chr3:49711434-49761384	0.27	1.64	0.44	0.0004	0.0204
ENSG00000165029	ABCA1	chr9:107543282-107691173	0.51	1.82	0.92	0.0000	0.0023
ENSG00000134250	NOTCH2	chr1:120454175-120612240	0.58	18.54	10.84	0.0012	0.0500
ENSG00000117016	RIMS3	chr1:41086350-41131329	0.60	17.81	10.61	0.0000	0.0000
ENSG00000155792	DEPTOR	chr8:120879658-121063152	0.60	10.15	6.08	0.0011	0.0469
ENSG00000073282	TP63	chr3:189349204-189615068	0.61	18.37	11.13	0.0000	0.0000
ENSG00000239899	RN7SL674P	chr2:11674241-11782914	0.62	201.82	124.57	0.0000	0.0000
ENSG00000137393	RNF144B	chr6:18366966-18469105	0.62	84.26	52.24	0.0000	0.0000
ENSG00000130775	THEMIS2	chr1:28199054-28213196	0.62	49.43	30.77	0.0000	0.0000
ENSG00000167483	FAM129C	chr19:17634109-17664647	0.63	3.34	2.09	0.0002	0.0094
ENSG00000244642	RN7SL396P	chr8:120874901-120875181	0.64	238.82	153.89	0.0000	0.0000
ENSG00000211640	IGLV6-57	chr22:22550112-22550686	0.65	44.34	28.60	0.0000	0.0006
ENSG00000230006	ANKRD36BP2	chr2:89065323-89106126	0.65	38.78	25.07	0.0001	0.0075
ENSG00000239437	RN7SL752P	chr3:129274017-129325661	0.65	135.52	87.77	0.0000	0.0002
ENSG00000125089	SH3TC1	chr4:8183798-8243530	0.65	79.36	51.46	0.0000	0.0000
ENSG00000265735	RN7SL5P	chr9:8314245-10612723	0.65	1756.91	1144.65	0.0000	0.0000
ENSG00000101144	BMP7	chr20:55743803-55841685	0.66	110.79	72.68	0.0002	0.0118
ENSG00000240869	RN7SL128P	chr6:20402397-20493941	0.66	260.67	172.73	0.0000	0.0000
ENSG00000082438	COBLL1	chr2:165510133-165705844	0.66	30.06	19.98	0.0002	0.0111
ENSG00000263740	RN7SL4P	chr3:15708742-15901278	0.67	1621.44	1088.28	0.0000	0.0000
ENSG00000211895	IGHA1	chr14:106173456-106175002	0.67	15.32	10.29	0.0000	0.0006
ENSG00000101846	STS	chrX:7137496-7272851	0.67	5.14	3.47	0.0000	0.0000
ENSG00000163884	KLF15	chr3:126061477-126076285	0.68	9.27	6.26	0.0012	0.0482
ENSG00000264462	MIR3648-1	chr21:9825831-9826011	0.68	1496.82	1011.79	0.0000	0.0000
ENSG00000211897	IGHG3	chr14:106235438-106237742	0.68	13.25	8.98	0.0000	0.0009
ENSG00000160856	FCRL3	chr1:157644110-157670775	0.68	32.28	21.96	0.0003	0.0176
ENSG00000176533	GNG7	chr19:2511217-2702746	0.68	24.17	16.52	0.0000	0.0000
ENSG00000132970	WASF3	chr13:27131839-27263085	0.69	6.00	4.13	0.0000	0.0000
ENSG00000179873	NLRP11	chr19:56296769-56393218	0.70	4.17	2.90	0.0000	0.0002

ENSG00000145990	GFOD1	chr6:13358061-13487787	0.70	11.10	7.73	0.0000	0.0000
ENSG00000226958	RNA28S5	chrX:108297360-108297792	0.70	2581.19	1802.17	0.0000	0.0000
ENSG00000112182	BACH2	chr6:90636247-91006627	0.70	92.93	64.95	0.0000	0.0001
ENSG00000221963	APOL6	chr22:36044441-36057404	0.70	5.52	3.87	0.0000	0.0018
ENSG00000197256	KANK2	chr19:11274943-11308243	0.70	28.80	20.24	0.0000	0.0000
ENSG00000211896	IGHG1	chr14:106202679-106209408	0.70	120.97	85.27	0.0000	0.0000
ENSG00000116717	GADD45A	chr1:68150743-68154021	0.71	58.38	41.66	0.0000	0.0000
ENSG00000166428	PLD4	chr14:105391152-105399574	0.72	27.51	19.76	0.0002	0.0119
ENSG00000159842	ABR	chr17:906757-1132315	0.72	234.84	168.72	0.0000	0.0000
ENSG00000211893	IGHG2	chr14:106109388-106115394	0.72	12.84	9.25	0.0010	0.0416
ENSG00000160505	NLRP4	chr19:56296769-56393218	0.72	6.00	4.33	0.0000	0.0003
ENSG00000175040	CHST2	chr3:142838172-142841800	0.72	10.68	7.71	0.0000	0.0001
ENSG00000155966	AFF2	chrX:147582138-148082193	0.72	9.50	6.88	0.0000	0.0005
ENSG00000221930	FAM45B	chrX:129611042-129658231	0.73	12.32	8.93	0.0009	0.0389
ENSG00000054654	SYNE2	chr14:64319682-64805317	0.73	183.41	133.35	0.0000	0.0000
ENSG00000168016	TRANK1	chr3:36868310-36986548	0.73	62.79	45.67	0.0000	0.0017
ENSG00000169750	RAC3	chr17:79989499-79992080	0.73	26.90	19.64	0.0000	0.0022
ENSG00000211892	IGHG4	chr14:106090686-106092403	0.73	18.42	13.53	0.0002	0.0097
ENSG00000164691	TAGAP	chr6:159455499-159466184	0.73	14.62	10.74	0.0000	0.0001
ENSG00000105327	BBC3	chr19:47724078-47736023	0.74	9.86	7.26	0.0001	0.0084
ENSG00000184384	MAML2	chr11:95709761-96076382	0.74	1.99	1.47	0.0008	0.0346
ENSG00000149639	SOGA1	chr20:35405844-35492089	0.74	12.62	9.33	0.0000	0.0000
ENSG00000149418	ST14	chr11:130029456-130080356	0.74	60.33	44.74	0.0000	0.0000
ENSG00000254709	IGLL5	chr22:23229959-23238014	0.74	354.62	263.92	0.0000	0.0000
ENSG00000117148	ACTL8	chr1:18081807-18153558	0.75	19.18	14.29	0.0000	0.0013
ENSG00000185838	GNB1L	chr22:19744225-19842462	0.75	24.82	18.51	0.0005	0.0237
ENSG00000211978	IGHV5-78	chr14:107259337-107259632	0.75	237.09	177.14	0.0001	0.0059
ENSG00000065057	NTHL1	chr16:2089815-2185899	0.75	98.92	73.91	0.0000	0.0000
ENSG00000135144	DTX1	chr12:113494513-113535833	0.75	413.55	309.92	0.0000	0.0000
ENSG00000088305	DNMT3B	chr20:31350190-31397162	0.75	7.95	5.97	0.0001	0.0047
ENSG00000128965	CHAC1	chr15:41245159-41248710	0.75	18.76	14.09	0.0005	0.0250
ENSG00000119139	TJP2	chr9:71736208-71870124	0.75	25.64	19.26	0.0006	0.0266
ENSG00000184524	CEND1	chr11:787103-790123	0.75	51.17	38.56	0.0004	0.0221

Table 3.4 Genes upregulated greater than 25% after treatment with polyamide **2** for 24 hours compared to non-treatment.

Ensembl Gene	Gene	Locus	Ratio	NT fpkm	GWWC fpkm	p_val	q_val
ENSG00000125675	GRIA3	chrX:122318005-122624766	N/A	0.00	1.30	0.0000	0.0025
ENSG00000164076	CAMKV	chr3:49895420-49907655	N/A	0.00	2.94	0.0000	0.0000
ENSG00000070808	CAMK2A	chr5:149569519-149669854	N/A	0.00	0.61	0.0000	0.0006
ENSG00000154188	ANGPT1	chr8:108261720-108510283	N/A	0.00	0.50	0.0000	0.0024
ENSG00000127152	BCL11B	chr14:99635623-99737861	16.50	0.07	1.14	0.0000	0.0000
ENSG00000239527	RPS23P7	chr17:60447578-60493837	10.63	0.54	5.69	0.0001	0.0050
ENSG00000253556	MTCO1P4	chr8:104101023-104102899	9.79	0.10	1.00	0.0003	0.0148
ENSG00000214204	HNRNPA1P43	chr1:116399481-116400440	5.03	0.30	1.50	0.0006	0.0293
ENSG00000101335	MYL9	chr20:34894257-35274619	3.91	0.33	1.30	0.0005	0.0260
ENSG00000213704	EEF1A1P15	chrX:97644507-97645918	3.78	0.79	2.98	0.0000	0.0000
ENSG00000188042	ARL4C	chr2:235401684-235405697	3.32	0.15	0.48	0.0002	0.0094
ENSG00000081059	TCF7	chr5:133450401-133487556	3.22	2.74	8.83	0.0000	0.0002
ENSG00000213128	RPL32P31	chr17:78516002-78516405	3.19	2.12	6.76	0.0010	0.0421
ENSG00000197956	S100A6	chr1:153506078-153508720	3.10	5.91	18.33	0.0000	0.0000
ENSG00000237882	PPIAP13	chr10:76849005-76849483	2.83	2.56	7.24	0.0002	0.0110
ENSG00000200312	RN7SKP255	chr14:89591214-90421121	2.16	18.53	40.03	0.0000	0.0005
ENSG00000200488	RN7SKP203	chr2:76672204-76672536	2.07	20.64	42.74	0.0000	0.0001
ENSG00000202198	RN7SK	chr6:52860417-52860748	1.90	83.52	158.42	0.0000	0.0000
ENSG00000201428	RN7SKP71	chr12:112597991-112819896	1.76	31.94	56.11	0.0000	0.0015
ENSG00000186834	HEXIM1	chr17:43224683-43236822	1.40	5.98	8.35	0.0000	0.0000
ENSG00000176393	RNPEP	chr1:201951499-201986316	1.26	88.42	111.35	0.0000	0.0005

Table 3.5 Genes downregulated greater than 25% after treatment with polyamide **2** for 24 hours compared to non-treatment.

Ensembl Gene ID	Gene	Locus	Ratio	NT fpkm	GWWC fpkm	p_val	q_val
ENSG00000260280	SLX1B-SULT1A4	chr16:29262828-29606395	0.00	1.58	0.00	0.0000	0.0001
ENSG00000134532	SOX5	chr12:23682439-24103966	0.44	17.01	7.56	0.0000	0.0013
ENSG00000131374	TBC1D5	chr3:17199898-18568819	0.45	18.98	8.54	0.0000	0.0000
ENSG00000184903	IMMP2L	chr7:110303109-111202573	0.49	15.84	7.81	0.0002	0.0121
ENSG00000131558	EXOC4	chr7:132937828-133751342	0.51	99.41	50.51	0.0000	0.0000
ENSG00000145996	CDKAL1	chr6:20534687-21232635	0.57	7.13	4.04	0.0000	0.0011
ENSG00000139734	DIAPH3	chr13:60239716-60738119	0.61	7.93	4.83	0.0000	0.0000
ENSG00000119522	DENND1A	chr9:126118448-126692431	0.62	45.22	27.97	0.0000	0.0003
ENSG00000073282	TP63	chr3:189349204-189615068	0.64	18.37	11.68	0.0000	0.0000
ENSG00000164808	SPIDR	chr8:48173166-48648868	0.64	128.25	82.62	0.0000	0.0000
ENSG00000155966	AFF2	chrX:147582138-148082193	0.65	9.50	6.16	0.0000	0.0000
ENSG00000112699	GMDS	chr6:1624040-2245926	0.65	33.89	22.14	0.0001	0.0037
ENSG00000141627	DYM	chr18:46550072-46987717	0.66	58.64	38.58	0.0000	0.0001
ENSG00000111880	RNGTT	chr6:89319984-89673348	0.67	12.09	8.10	0.0000	0.0000
ENSG00000133195	SLC39A11	chr17:70642087-71088851	0.67	42.03	28.21	0.0000	0.0007
ENSG00000049323	LTBP1	chr2:33172038-33624576	0.67	23.00	15.48	0.0003	0.0138
ENSG00000198846	TOX	chr8:59717976-60033905	0.68	4.89	3.31	0.0000	0.0001
ENSG00000170396	ZNF804A	chr2:185463092-185804219	0.68	4.11	2.80	0.0000	0.0002
ENSG00000144535	DIS3L2	chr2:232825954-233209060	0.68	21.08	14.37	0.0004	0.0190
ENSG00000156110	ADK	chr10:75910959-76469061	0.68	38.74	26.51	0.0000	0.0000
ENSG00000167216	KATNAL2	chr18:44526786-44628613	0.69	3.13	2.16	0.0009	0.0397
ENSG00000184384	MAML2	chr11:95709761-96076382	0.69	1.99	1.38	0.0001	0.0069
ENSG00000184220	CMSS1	chr3:99536677-99897447	0.69	88.52	61.50	0.0000	0.0000
ENSG00000091157	WDR7	chr18:54318615-54698828	0.70	3.61	2.51	0.0000	0.0000
ENSG00000174891	RSRC1	chr3:157813742-158288855	0.70	23.63	16.44	0.0000	0.0017
ENSG00000259660	DNM1P47	chr15:102291234-102334505	0.70	0.60	0.42	0.0009	0.0391
ENSG00000137393	RNF144B	chr6:18366966-18469105	0.70	84.26	58.92	0.0000	0.0000
ENSG00000152061	RABGAP1L	chr1:174084437-174964445	0.70	93.94	65.86	0.0000	0.0000
ENSG00000185104	FAF1	chr1:50905149-51425935	0.70	65.35	45.88	0.0000	0.0000
ENSG00000153317	ASAP1	chr8:131064352-131455906	0.71	29.49	20.80	0.0007	0.0306
ENSG00000049618	ARID1B	chr6:157099062-157530401	0.71	24.14	17.20	0.0000	0.0000

ENSG00000182247	UBE2E2	chr3:23244510-23633284	0.71	29.49	21.03	0.0000	0.0004
ENSG00000117090	SLAMF1	chr1:160577889-160617085	0.71	27.01	19.30	0.0000	0.0003
ENSG00000138623	SEMA7A	chr15:74701629-74726808	0.72	22.73	16.31	0.0010	0.0431
ENSG00000054654	SYNE2	chr14:64319682-64805317	0.72	183.41	131.71	0.0000	0.0000
ENSG00000153015	CWC27	chr5:64064756-64314418	0.72	24.69	17.79	0.0000	0.0027
ENSG00000197157	SND1	chr7:127292233-127732661	0.73	610.24	446.33	0.0000	0.0000
ENSG00000155849	ELMO1	chr7:36893960-37488852	0.73	86.24	63.11	0.0000	0.0000
ENSG00000005810	MYCBP2	chr13:77618791-77901185	0.73	82.32	60.25	0.0000	0.0005
ENSG00000176463	SLCO3A1	chr15:92396924-92874267	0.73	41.59	30.47	0.0010	0.0411
ENSG00000156639	ZFAND3	chr6:37787274-38122400	0.73	32.89	24.13	0.0007	0.0309
ENSG00000038382	TRIO	chr5:14143810-14532235	0.74	103.50	76.21	0.0003	0.0138
ENSG00000010803	SCMH1	chr1:41492871-41707826	0.74	57.57	42.68	0.0010	0.0416
ENSG00000235823	LINC00263	chr10:102133371-102143125	0.74	5.92	4.41	0.0008	0.0338
ENSG00000075151	EIF4G3	chr1:21132971-21503377	0.74	111.99	83.41	0.0000	0.0000
ENSG00000144645	OSBPL10	chr3:31699381-32119072	0.75	112.81	84.50	0.0000	0.0001
ENSG00000198648	STK39	chr2:168810529-169104651	0.75	21.20	15.90	0.0003	0.0170
ENSG00000145349	CAMK2D	chr4:114372187-114683083	0.75	39.19	29.50	0.0001	0.0079

3.6 References

(1) Dunn, G. P.; Old, L. J.; Schreiber, R. D. *Annu. Rev. Immunol.* **2004**, *22*, 329-360.

(2) Hanahan, D.; Weinberg, R. A. *Cell.* **2011**, *144*, 646-674.

(3) Dunn, G. P.; Old, L. J.; Schreiber, R. D. *Immunity.* **2004**, *21*, 137-148.

(4) Koebel, C. M.; Vermi, W.; Swann, J. B.; Zerafa, N.; Rodig, S. J.; Old, L. J.; Smyth, M. J.; Schreiber, R. D. *Nature.* **2007**, *450*, 903-907.

(5) Schreiber, R. D.; Old, L. J.; Smyth, M. J. *Science Signalling.* **2011**, *331*, 1565.

(6) O'Sullivan, T.; Saddawi-Konefka, R.; Vermi, W.; Koebel, C. M.; Arthur, C.; White, J. M.; Uppaluri, R.; Andrews, D. M.; Ngiow, S. F.; Teng, M. W.; Smyth, M. J.; Schreiber, R. D.; Bui, J. D. *J. Exp. Med.* **2012**, *209*, 1869-1882.

(7) Eggermont, A.; Robert, C.; Soria, J. C.; Zitvogel, L. *Oncoimmunology.* **2014**, *3*, e27560.

(8) Matzinger, P. *Science.* **2002**, *296*, 301-305.

(9) Obeid, M.; Tesniere, A.; Ghiringhelli, F.; Fimia, G. M.; Apetoh, L.; Perfettini, J. L.; Castedo, M.; Mignot, G.; Panaretakis, T.; Casares, N.; Métivier, D.; Larochette, N.; van Endert, P.; Ciccosanti, F.; Piacentini, M.; Zitvogel, L.; Kroemer, G. *Nat. Med.* **2007**, *13*, 54-61.

(10) Jaiswal, S.; Chao, M. P.; Majeti, R.; Weissman, I. L. *Trends. Immunol.* **2010**, *31*, 212-219.

(11) Tseng, D.; Volkmer, J. P.; Willingham, S. B.; Contreras-Trujillo, H.; Fathman, J. W.; Fernhoff, N. B.; Seita, J.; Inlay, M. A.; Weiskopf, K.; Miyanishi, M.; Weissman, I. L. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 11103-11108.

(12) Zitvogel, L.; Apetoh, L.; Ghiringhelli, F.; Kroemer, G. *Nat. Rev. Immunol.* **2008**, *8*, 59-73.

(13) Zitvogel, L.; Kepp, O.; Kroemer, G. *Cell.* **2010**, *140*, 798-804.

(14) Kepp, O.; Galluzzi, L.; Martins, I.; Schlemmer, F.; Adjemian, S.; Michaud, M.; Sukkurwala, A. Q.; Menger, L.; Zitvogel, L.; Kroemer, G. *Cancer Metastasis Rev.* **2011**, *30*, 61-69.

(15) Chao, M. P.; Jaiswal, S.; Weissman-Tsukamoto, R.; Alizadeh, A. A.; Gentles, A. J.; Volkmer, J.; Weiskopf, K.; Willingham, S. B.; Raveh, T.; Park, C. Y.; Majeti, R.; Weissman, I. L. *Sci. Transl. Med.* **2010**, *2*, 63ra94.

(16) Idzko, M.; Dichmann, S.; Ferrari, D.; Di Virgilio, F.; la Sala, A.; Giampiero, G.; Panther, E.; Norgauer, J. *Blood.* **2002**, *100*, 925-932.

(17) Aymeric, L.; Apetoh, L.; Ghiringhelli, F.; Tesniere, A.; Martins, I.; Kroemer, G.; Smyth, M. J.; Zitvogel, L. *Cancer Res.* **2010**, *70*, 855-858.

(18) Rovere-Querini, P.; Capobianco, A.; Scaffidi, P.; Valentinis, B.; Catalanotti, F.; Giazzon, M.; Dumitriu, I. E.; Müller, S.; Iannacone, M.; Traversari, C.; Bianchi, M. E.; Manfredi, A. A. *EMBO reports.* **2004**, *5*, 825-830.

(19) Guerriero, . L.; Ditsworth, D.; Catanzaro, . M.; Sabino, G.; Furie, . B.; Kew, . R.; Crawford, . C.; Zong, -X. *J. Immunol.* **2011**, *186*, 3517-3526.

(20) Gasser, S.; Orsulic, S.; Brown, E. J.; Raulet, D. H. *Nature.* **2005**, *436*, 1186-1190.

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

(22) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature.* **1996**, *382*, 559-561.

(23) White, S.; Baird, E. E.; Dervan, P. B. *Chem. Biol.* **1997**, *4*, 569-578.

(24) White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Nature.* **1998**, *391*, 468-471.

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

(26) 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.

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

(28) Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 7983-7988.

(29) 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.; Dervan, P. B. *Tetrahedron.* **2007**, *63*, 6146-6151.

(30) Nickols, N. G.; Jacobs, C. S.; Farkas, M. E.; Dervan, P. B. *ACS Chem. Biol.* **2007**, *2*, 561-571.

(31) Nickols, N. G.; Dervan, P. B. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 10418-10423.

(32) Raskatov, J. A.; Meier, J. L.; Puckett, J. W.; Yang, F.; Ramakrishnan, P.; Dervan, P. B. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 1023-1028.

(33) Yang, F.; Nickols, N. G.; Li, B. C.; Marinov, G. K.; Said, J. W.; Dervan, P. B. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 1863-1868.

(34) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 8198-8206.

(35) Nickols, N. G.; Szablowski, J. O.; Hargrove, A. E.; Li, B. C.; Raskatov, J. A.; Dervan, P. B. *Mol. Cancer Ther.* **2013**, *12*, 675-684.

(36) Kang, J. S.; Meier, J. L.; Dervan, P. B. *J. Am. Chem. Soc.* **2014**, *136*, 3687-3694.

(37) Martínez, T. F.; Phillips, J. W.; Karanja, K. K.; Polaczek, P.; Wang, C. -M.; Li, B. C.; Campbell, J. L.; Dervan, P. B. *Nucleic Acids Res.* **2015**, *42*, 11546-11559.

(38) Michalak, M.; Groenendyk, J.; Szabo, E.; Gold, L. I.; Opas, M. *Biochem. J.* **2009**, *417*, 651-666.

(39) Tufi, R.; Panaretakis, T.; Bianchi, K.; Criollo, A.; Fazi, B.; Di Sano, F.; Tesniere, A.; Kepp, O.; Paterlini-Brechot, P.; Zitvogel, L.; Piacentini, M.; Szabadkai, G.; Kroemer, G. *Cell Death Differ.* **2008**, *15*, 274-282.

(40) Peters, L. R.; Raghavan, M. *J. Immunol.* **2011**, *187*, 919-931.

(41) Maeshima, K.; Janssen, S.; Laemmli, U. K. *EMBO J.* **2001**, *20*, 3218-3228.

(42) Kawamoto, Y.; Bando, T.; Kamada, F.; Li, Y.; Hashiya, K.; Maeshima, K.; Sugiyama, H. *J. Am. Chem. Soc.* **2013**, *135*, 16468-16477.

(43) Meier, J. L.; Yu, A. S.; Korf, I.; Segal, D. J.; Dervan, P. B. *J. Am. Chem. Soc.* **2012**, *134*, 17814-17822.

(44) Pilch, D. S.; Poklar, N.; Gelfand, C. A.; Law, S. M.; Breslauer, K. J.; Baird, E. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 8306-8311.

(45) Panaretakis, T.; Kepp, O.; Brockmeier, U.; Tesniere, A.; Bjorklund, A. C.; Chapman, D. C.; Durchschlag, M.; Joza, N.; Pierron, G.; van Endert, P.; Yuan, J.; Zitvogel, L.; Madeo, F.; Williams, D. B.; Kroemer, G. *EMBO J.* **2009**, *28*, 578-590.

(46) Lippincott-Schwartz, J.; Yuan, L. C.; Bonifacino, J. S.; Klausner, R. D. *Cell.* **1989**, *56*, 801-813.

(47) Kaczmarek, A.; Vandenabeele, P.; Krysko, D. V. *Immunity.* **2013**, *38*, 209-223.

(48) Rubartelli, A.; Lotze, M. T. *Trends Immunol.* **2007**, *28*, 429-436.

(49) Mortazavi, A.; Williams, B. A.; McCue, K.; Schaeffer, L.; Wold, B. *Nat. Methods.* **2008**, *5*, 621-628.

(50) Langmead, B.; Trapnell, C.; Pop, M.; Salzberg, S. L. *Genome Biol.* **2009**, *10*, R25.

(51) Trapnell, C.; Pachter, L.; Salzberg, S. L. *Bioinformatics.* **2009**, *25*, 1105-1111.

(52) Trapnell, C.; Williams, B. A.; Pertea, G.; Mortazavi, A.; Kwan, G.; van Baren, M. J.; Salzberg, S. L.; Wold, B. J.; Pachter, L. *Nat. Biotechnol.* **2010**, *28*, 511-515.

(53) Balkwill, F. *Nat. Rev. Cancer.* **2009**, *9*, 361-371.

(54) Braud, V. M.; Allan, D. S.; O'Callaghan, C. A.; Söderström, K.; D'Andrea, A.; Ogg, G. S.; Lazetic, S.; Young, N. T.; Bell, J. I.; Phillips, J. H. *Nature.* **1998**, *391*, 795-799.

(55) Nuchtern, J. G.; Biddison, W. E.; Klausner, R. D. *Nature.* **1990**, *343*, 74-76.

(56) Neefjes, J.; Jongsma, M. L.; Paul, P.; Bakke, O. *Nat. Rev. Immunol.* **2011**, *11*, 823-836.

(57) Sukhatme, V. P.; Cao, X.; Chang, L. C.; Tsai-Morris, C. -H.; Stamenkovich, D.; Ferreira, P. C.; Cohen, D. R.; Edwards, S. A.; Shows, T. B.; Curran, T. *Cell.* **1988**, *53*, 37-43.

(58) McMahon, A. P.; Champion, J. E.; McMahon, J. A.; Sukhatme, V. P. *Development.* **1990**, *108*, 281-287.

(59) Seiler, M. P.; Mathew, R.; Liszewski, M. K.; Spooner, C. J.; Spooner, C.; Barr, K.; Meng, F.; Singh, H.; Bendelac, A. *Nat. Immunol.* **2012**, *13*, 264-271.

(60) Peck, D.; Crawford, E. D.; Ross, K. N.; Stegmaier, K.; Golub, T. R.; Lamb, J. *Genome Biol.* **2006**, *7*, R61.

(61) Kumar, S.; Boehm, J.; Lee, J. C. *Nat. Rev. Drug. Discov.* **2003**, *2*, 717-726.

(62) Liebermann, D. A.; Hoffman, B. *J. Mol. Signal.* **2008**, *3*, 15.

(63) Yamamoto, M.; Sato, S.; Hemmi, H.; Sanjo, H.; Uematsu, S.; Kaisho, T.; Hoshino, K.; Takeuchi, O.; Kobayashi, M.; Fujita, T.; Takeda, K.; Akira, S. *Nature.* **2002**, *420*, 324-329.

(64) Gough, D. J.; Messina, N. L.; Clarke, C. J.; Johnstone, R. W.; Levy, D. E. *Immunity.* **2012**, *36*, 166-174.

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

(66) Puckett, J. W.; Green, J. T.; Dervan, P. B. *Org. Lett.* **2012**, *14*, 2774-2777.

Chapter 4

Synthesis and toxicology studies of a Py-Im polyamide targeted to the CTG•CAG-repeat sequence of Myotonic Dystrophy

This research was conducted in collaboration with Alissa A. Hare, Alexis A. Kurmis, Jordan L. Meier, Fei Yang, and Peter B. Dervan (California Institute of Technology) and with Joel Gottesfeld and Jintang Du (Scripps Research Institute).

Abstract

Myotonic dystrophy type 1 (DM1) is a neuromuscular disease caused by an expansion of CTG•CAG triplet repeats which, when transcribed into mRNA, sequesters the RNA splicing regulator muscleblind-like protein from its normal function and causes clinical pathology. DNA binding Py-Im polyamides targeted to the CTG•CAG triplet repeat sequence may be useful towards disrupting transcription of the toxic RNA. In this study, we assess the toxicity of two polyamides in mice to inform tolerable dosing regimens to be examined in a DM1 mouse model. We found that subcutaneously injected hairpin polyamide **2** targeted to 5'-(A/T)GC(A/T)GC(A/T)-3' is absorbed and circulates in a fashion similar to previously studied eight-ring polyamides. However, polyamide **2** dosed three times a week displayed toxicity at low (1 mg/kg) doses and serum chemistry showed liver and kidney damage at a single dose of 5 and 10 mg/kg. Polyamide **3**, which differs only by an acetylation at the butyric acid turn unit, showed no toxicity at a single dose of up to 9 mg/kg but was dose limited by solubility. This study suggests polyamide **3** is a better tolerated compound and should be assessed for circulation and multiple dose toxicity, but its formulation should be improved.

4.1 Introduction

Myotonic dystrophy type 1 (DM1) is an inherited neuromuscular disorder with a wide-spectrum of clinical pathologies which include dysfunction of skeletal muscles, cardiac defects, and cataracts.[1] The disease is caused by a pathological expansion of CTG•CAG triplet repeats at the 3' untranslated region of the gene encoding serine-threonine protein kinase DMPK1.[2] The expanded triplet repeats are transcribed into mRNA, which sequesters the RNA splicing regulator muscleblind-like 1 protein (MBNL1) to form nuclear foci.[3] MBNL1 governs exon splicing in a selection of genes and its sequestration results in splicing defects that cause clinical toxicity (Figure 4.1).[4]

Disease models and phenotypic assays have been developed to discover means to disrupt this pathology. In cell culture, potential therapeutic strategies can be screened for reduction in the number of toxic nuclear foci. For screening in animals, Thornton and coworkers have developed a mouse model with muscle-specific expression of skeletal

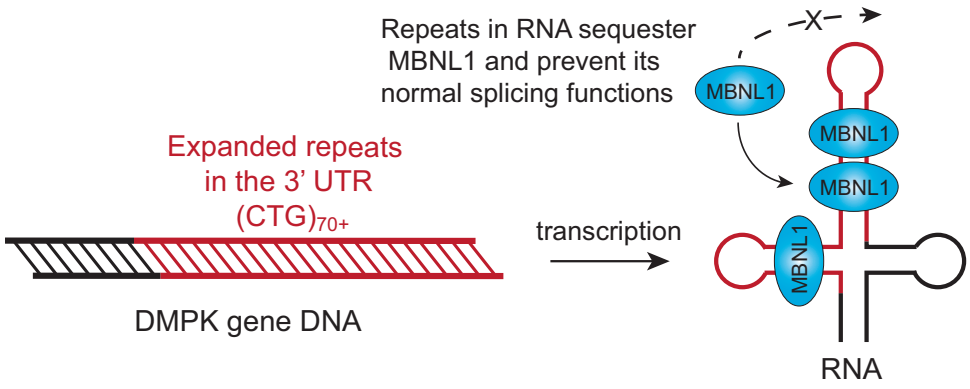


Figure 4.1 Diagram of the molecular cause of Myotonic Dystrophy type 1. Pathogenic RNA transcribed from expanded CTG•CAG triplet repeats binds MBNL1. MBNL1 is not available to perform its normal RNA splicing functions and results in toxicity.

actin containing the expanded triplet repeats.[5] The model successfully recapitulated sequestration of MBNL1, splicing defects, and myotonia. This model has been used to show that RNA interference technology can prevent sequestration of MBNL1 and reverse pathology.[6] RNA interference is a reasonable method to degrade the triplet-repeat expanded transcripts after their production, but its therapeutic utility is challenged by poor organismal distribution.[7] An alternative approach to reducing toxic mRNA is to decrease the transcription of the expanded repeat sequence, which potentially can be accomplished by sequence-specific DNA-binding small molecules.

Py-Im polyamides are a class of programmable, sequence-specific small molecules that bind in the minor groove of DNA.[8-13] Sequence preference is achieved by side-by-side pairings of aromatic amino acids that distinguish the edges of the four Watson-Crick base pairs according to the pairing rules: Im/Py codes for a G•C base pair and Py/Py binds both T•A/A•T in preference to G•C/C•G.[11-14] Eight-ring hairpin polyamides linked by a central aliphatic γ-aminobutyric acid unit have affinities for match sites with K_a's from 10^8 to 10^{10} M^{-1}.[9,15] Targeting longer sequences in this hairpin motif is challenging because the oligomers are over-curved compared to DNA.[16,17] To relax the curvature of oligomers targeting longer sequences, β alanine (β) can be substituted for Py-rings in some cases such that β/β pairs replace Py/Py for T•A/A•T specificity.[16-18] Access to chromatin without transfection agents has been demonstrated by fluorescent dye conjugates of cell-permeable polyamides which localize to the nucleus in live cells.[19,20] Hairpin polyamides have been shown to downregulate transcription of

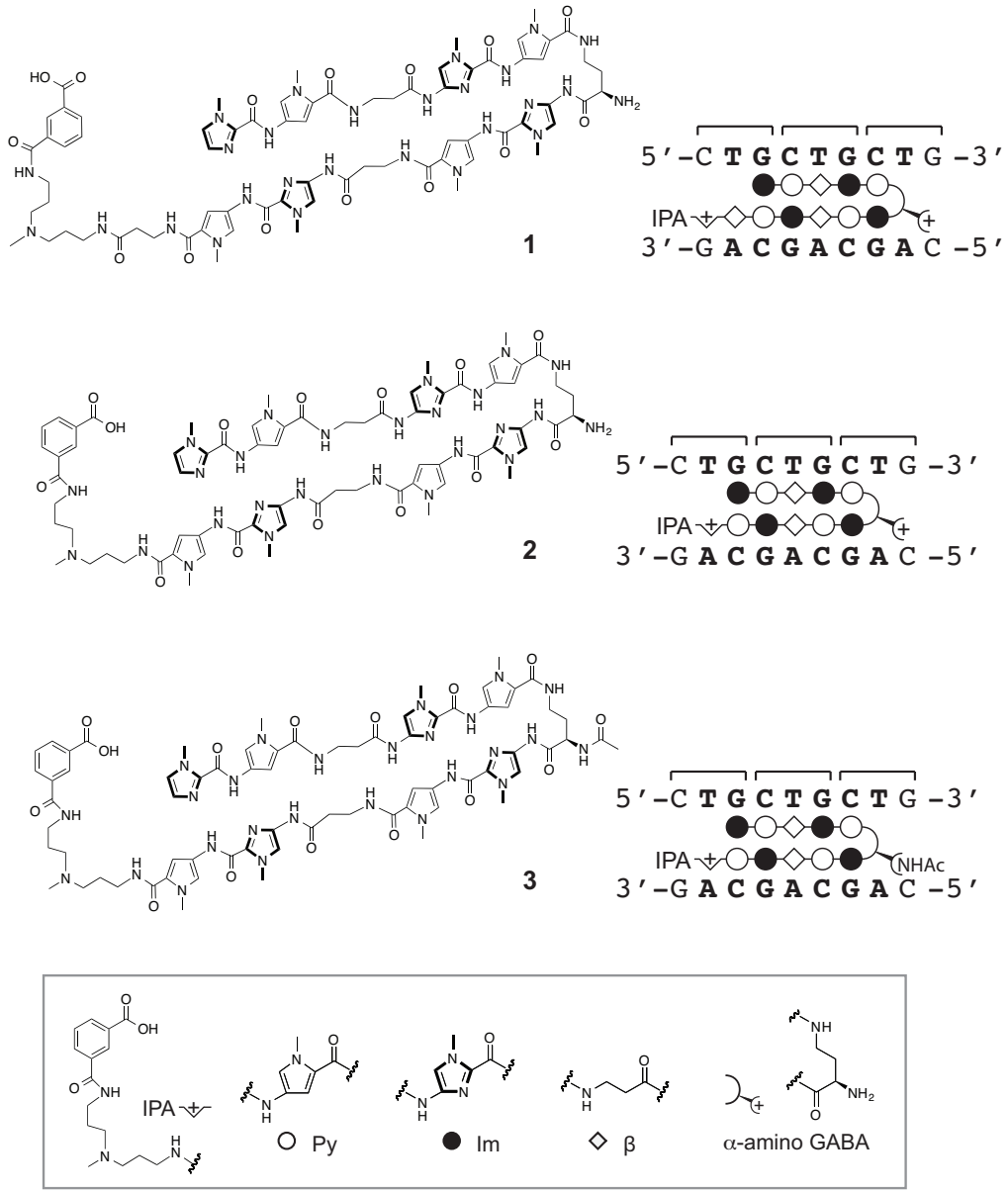
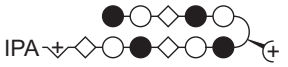
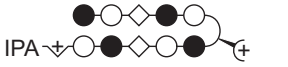
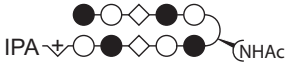


Figure 4.2 Structures of polyamides synthesized for this study. Polyamide **1-3** contain the same core and target the same 5'-(A/T)GC(A/T)GC(A/T)-3' sequence contained in the triplet repeat sequence but differ at the C-terminal tail and the hairpin turn unit.

associated genes when targeted to the consensus sequences of binding sites of transcription factors in both cells and xenograft models in mice.[21-25] A Py-Im polyamide targeted to the CTG•CAG repeat sequence may downregulate the transcription of the

Table 4.1 Thermal stabilization of DNA by 2β2 polyamides.

5′–A C **T G C T G C T** G A T –3′
3′–T G **A C G A C G A** C T A –5′

		T_m ºC	ΔT_m ºC
DNA only		57.0 (±0.3)	- -
	**1**	73.7 (±0.2)	16.7
	**2**	76.2 (±0.1)	19.1
	**3**	71.2 (±0.2)	14.2

DMPK1 gene which causes the DM1 pathology. Studies in mouse models have shown Py-Im polyamides circulate with good distribution and pharmacokinetic profiles,[26,27] such that Py-Im oligomers may provide an avenue for therapeutic benefit.

4.2 Results

Py-Im polyamides **1**, **2**, and **3** are hairpin polyamides with 2 aromatic rings β–alanine and 2 aromatic rings on each strand (2β2), designed and synthesized to target the CTG•CAG repeat sequence (Figure 4.2). Their DNA binding to the CTG repeat sequence 5'-AC**TGCTGCT**GAT-3' was verified by a DNA thermal denaturation experiment, which correlates with binding affinity. [28] The results show robust stabilization of the duplex DNA by all three polyamides, with ΔT_m values of 16.7 ºC, 19.1 ºC, and 14.2 ºC for 2β2 hairpins **1**, **2**, and **3**, respectively (Table 4.1). In on-going studies, these compounds are being evaluated by our collaborators in the Gottesfeld group in a cell culture model of DM1. Induced pluripotent stem cells (iPSC) with

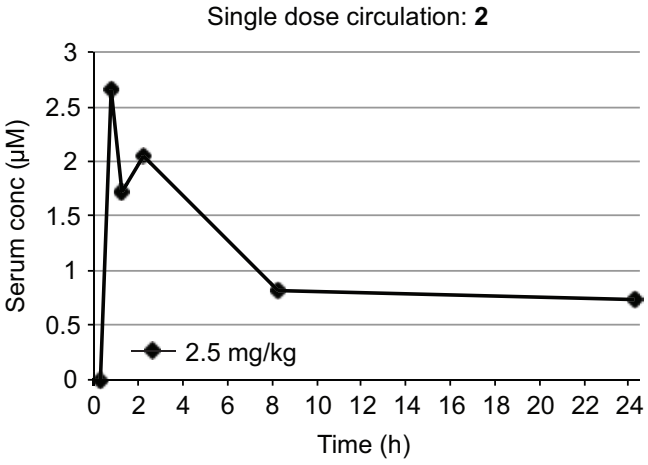


Figure 4.3 Circulation study of polyamide **2** in normal C57BL/6 male mice. Single subcutaneous dose in 20% DMSO/PBS given to four mice, two mice per time point. Concentrations were obtained by analytical HPLC against an internal standard.

CTG•CAG expanded repeats were produced using transcription factor reprogramming of DM1 patient fibroblasts. As their preliminary results suggested a reduction in nuclear RNA foci, we began studies in parallel to assess the toxicity of the compounds in normal mice. The aim of this study is to assess toxicity to inform tolerable dosing regimens to be examined for efficacy in the expanded-repeat DM1 mouse model.

We first chose to assess hairpin polyamide **2** in mice with the consideration that the C-terminal tail structure without the β–alanine has been better characterized in both live cell and mouse studies, and its solubility properties were better than that of **3.** To confirm circulation of the hairpin polyamide, 4 normal C57BL/6 mice were injected subcutaneously with polyamide **2** at a concentration 2.5 mg/kg. We applied a single dose at 200 μl and were limited by the solubility of the compound in the vehicle, 20% DMSO in PBS. Nevertheless, we detected the compound in the blood stream within half an hour of injection at 2.7 μM (Figure 4.3). The dip in concentration at 1 hour is likely variability

between mice because time points are alternated between groups of two mice each; this would likely average out with more replicates. At 8 hours and 24 hours, the detected peaks are near the lower detection limit for the HPLC. In proportion to the amount of polyamide **2** injected, the concentration and time course of elimination is similar to that previously observed with subcutaneously injected eight-ring hairpins.[26]

To achieve higher doses in a single dose experiment of toxicity, we exchanged PBS for saline in the vehicle, which allowed us to achieve concentrations of 5 mg/kg per 200 µl injection. On day 0, we injected male C57BL/6 mice subcutaneously with one or two injections to achieve doses of 5 mg/kg (n = 3) and 10 mg/kg (n = 1), respectively, and observed their weight over 9 days (Figure 4.4A and 4.4B). By day 6, the mouse dosed at 10 mg/kg had lost over 15% of its weight and had to be euthanized. At day 9, the end of the study, mice dosed at 5 mg/kg were distressed and averaged 87% of their initial weight (n=3). Blood serum was obtained from the mice at the end of the study or upon sacrifice for blood chemistry analysis. We found a large elevation of aspartate transaminase (AST), alanine transaminase (ALT), and total bilirubin in the blood, indicative of liver toxicity in a dose dependent manner (Table 4.2). Blood urea nitrogen (BUN) and creatinine are also elevated, suggesting the mice sustained kidney damage as well (Table 4.2). This showed the maximum tolerated single dose is under 5 mg/kg for polyamide **2**.

We next explored lower doses in anticipation of multiple dose studies. At 1 mg/kg (n = 1) and 2 mg/kg (n = 1), there was no decrease in weight observed over the 9 days

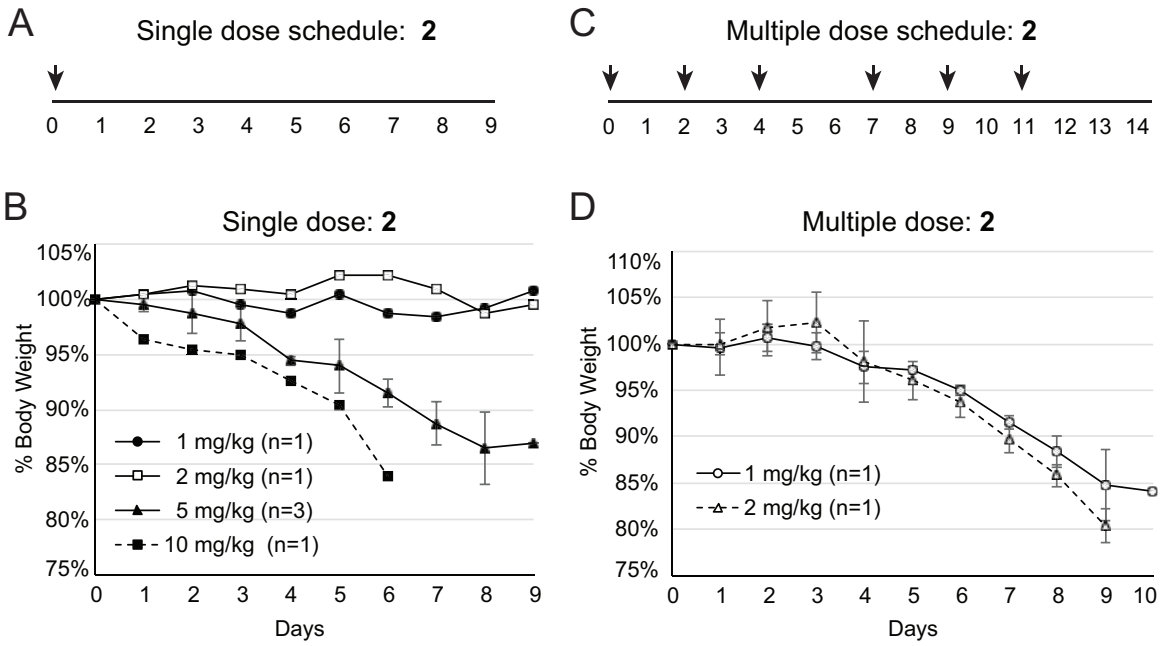


Figure 4.4 Weight curve studies with polyamide **2**. A) Schedule of dosing and observation for single dose study with polyamide **2**. B) Weight curves for escalating single doses: 1, 2, 5, and 10 mg/kg dosed subcutaneously in 20% DMSO/saline. C) Schedule of dosing and observation for multiple dose study with polyamide **2**. D) Weight curves for multiple doses at 1 or 2 mg/kg dosed subcutaneously. Mice were euthanized prior to end of study due to weight loss greater than 15%.

following injection (Figure 4.4B). We then assessed the tolerability of hairpin **2** after multiple dose exposure at these concentrations. The study was designed for three injections per week, for two weeks (Figure 4.4C). Mice were injected at 1 mg/kg (n = 2) and 2 mg/kg (n = 2) and their weights monitored. Mice injected at 2 mg/kg lost more than 15% of their weight by day 9 and were euthanized. Mice injected at 1 mg/kg were euthanized after losing more than 15% of their body weight by day 10, such that no mice survived the experiment (Figure 4.4D).

Given the significant toxicity observed with polyamide **2**, we tested polyamide **3**, as previous studies have indicated acetylation of the amine on the γ-butyric acid turn unit

Table 4.2 Serum chemistry analysis for liver and kidney toxicity.

Polyamide	ALT	AST	Bilirubin	BUN	Creatinine
Normal	57 ±40	194 ±152	0.32 ±0.3	23.6 ±4.8	0.34 ±0.12
Vehicle	17	45	0.1	30	0.2
1 5 mg/kg	1343	798	0.3	50	0.7
1 10 mg/kg	1944	1297	0.2	79	0.9

can improve the toxicity profile.[29] Polyamide **3** was dissolved in 20% DMSO in saline, but was only soluble to 2.5 mg/kg. Doses of 9 mg/kg (n = 1) and 5 mg/kg (n = 1) were achieved with 4 and 2 injections, respectively. Despite the large volume of vehicle injected into the mice, both doses showed no toxic effect on the mice, and the mice gained weight through the 9 days of observation (Figure 4.5A and B). These results suggest the 2β2 hairpins **2** and **3** follow the same pattern of reduced toxicity shown for eight-ring hairpins upon acetylation of the amine on the turn unit.

4.3 Discussion

From these studies we have found that the 2β2 hairpin polyamide **2** circulates in a fashion similar to eight ring polyamides dosed subcutaneously. However, we found that polyamide **2** is toxic at a single dose of 5 mg/kg and at 1 mg/kg dosed three times a week. Indeed, a single injection at 5 and 10 mg/kg caused liver and kidney damage. This compound is likely too toxic for long term dosing in the DM1 mouse model. Due to this toxicity, the acetylated 2β2 hairpin **3** was examined. Polyamide **3** showed a more

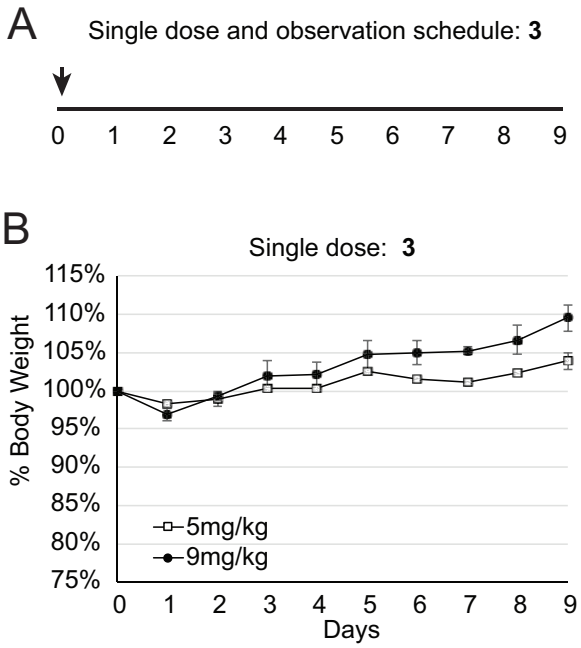


Figure 4.5 Weight curve studies with polyamide **3**. A) Schedule of dosing and observation for single dose study with polyamide **3**. B) Weight curves for single doses delivered in 2 or 4 subcutaneous injections of 20% DMSO/saline.

favorable tolerability profile by weight assessment after a single dose at 5 and 9 mg/kg, dose levels that caused toxicity in with polyamide **2**. The circulation of this compound and its tolerability after multiple doses remains to be tested.

The insolubility of the acetylated 2β2 oligomer **3** in the 20% DMSO/saline vehicle will limit achievable compound exposure in the mouse model and motivates exploration of improved formulation. While dose titration will be an iterative process with the efficacy model, an improved formulation will be advantageous for testing higher concentrations. It remains to be studied whether an alteration in formulation will substantively change the pharmacokinetic and toxicity profiles of the polyamide. The

establishment of a well-tolerated dosing window in mice will be the first step towards assessing the utility of 2β2 hairpin polyamides for addressing DM1 pathology.

4.4 Materials and Methods

Synthesis of polyamides. Hairpin polyamide **1** was synthesized on β-alanine Pam-resin and polyamides **2** and **3** were synthesized on Kaiser oxime resin as previously described.[30-32] Polyamides were purified by preparatory HPLC and the purity and identity determined by analytical HPLC and MALDI-TOF (Table 4.3).

Thermal denaturation studies. The DNA oligomer 5'-ACTGCTGCTGAT-3' and its complement were purchased from Integrated DNA Technology and annealed to form duplex DNA. DNA and each hairpin polyamide was mixed to a final concentration of 2 and 3 μM, respectively, in 1 mL total volume. The buffer was an aqueous solution of 10 mM sodium cacodylate, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$ at pH 7. A Varian Cary 100 spectrophotometer with a thermocontroller was used for measurements at λ = 260 nm with a path length of 1 cm. Samples were heated to 90 °C and cooled to a starting temperature of 25 °C and then heated at a rate of 0.5 °C/min to 90 °C to record denaturation profiles. Melting temperature was defined as the maximum of the first derivative of the denaturation profile. Data is the mean of four measurements.

Mouse studies. All animal experiments were conducted under an IACUC approved protocol at the California Institute of Technology. 6-8 week old male C57BL/6J mice were purchased from Jackson Laboratory and allowed to acclimate for at least 3 days.

Table 4.3 Masses of polyamides by MALDI-TOF.

Polyamide	Formula	[Mass + H]	Found Mass
1	$C_{72}H_{90}N_{27}O_{15}^+$	1572.7	1572.7
2	$C_{69}H_{85}N_{26}O_{14}^+$	1501.7	1501.6
3	$C_{71}H_{87}N_{26}O_{15}^+$	1543.7	1543.7

Circulation of compound 2. The circulation experiment was conducted as previously described for 8-ring hairpin polyamide analyses. [26] Four acclimatized 8-week old male C57BL/6J mice were separated into two cages and all dosed with polyamide **2** in a subcutaneous single injection in the flank. Solubility limited the concentration to 2.5 mg/kg dosed in 200 µl of 20% DMSO/PBS. Blood was obtained by retro-orbital collection at 0.5, 1, 2, 8, and 24 hours from alternating pairs of mice. The blood from two mice was combined and serum isolated by centrifugation at 6000 × g for 5 minutes. 40 µl of serum was combined with 80 µl of MeOH and vortexed to precipitate proteins. The samples were centrifuged at 16000 × g and 90 µl of the supernatant removed to a new tube. An equal volume of 20% acetonitrile in 0.1% TFA/water was added. 20 µl of an internal standard, Boc-Im-OEt, (~ 1 mg in 10 ml water) water was spiked into 180 µl of sample.

The samples were analyzed on a Beckman Gold analytical HPLC with a Phenomenex Kinetex C18 analytical column (100 mm × 4.6 mm, 2.6 µm). Peaks for polyamide **2** were quantitated by integration at 304 nm relative to the internal standard and determined by interpolation on a standard curve. A standard curve of concentration versus integrated area was generated by diluting known concentrations of polyamide **2** into C57BL/6 serum and preparing the samples with the internal standard in the same manner as above.

Single dose or multiple dose animal weight loss. Animals were injected at the described doses and schedules with U-100 insulin syringes (UltiCare). Treatment was given subcutaneously in 20% DMSO in saline or PBS as vehicle. Animals were monitored daily for weight for 9 or 14 days (single dose or multiple dose, respectively) and sacrificed if weight loss was greater than 15%.

Serum chemistry analysis. Serum analysis was conducted as previously described. [29] In brief, blood from treated mice was obtained by retro-orbital bleeding (Drummond) and serum isolated after centrifugation at 6000 × g for 5 minutes. Serum ALT, AST, total bilirubin, BUN, and creatinine levels was analyzed at IDEXX-RADIL. Normal values are reported from The Jackson Laboratory (http://jaxmice.jax.org/strain/000664.html).

4.5 References

(1) Harper, P. S. *Myotonic dystrophy*, 2nd ed.; Saunders: London, 1989.

(2) Brook, J. D.; McCurrach, M. E.; Harley, H. G.; Buckler, A. J.; Church, D.; Aburatani, H.; Hunter, K.; Stanton, V. P.; Thirion, J. -P.; Hudson, T. *Cell.* **1992**, *68*, 799-808.

(3) Miller, J. W.; Urbinati, C. R.; Teng-umnuay, P.; Stenberg, M. G.; Byrne, B. J.; Thornton, C. A.; Swanson, M. S. *EMBO. J.* **2000**, *19*, 4439-4448.

(4) Lin, X.; Miller, J. W.; Mankodi, A.; Kanadia, R. N.; Yuan, Y.; Moxley, R. T.; Swanson, M. S.; Thornton, C. A. *Hum. Mol. Genet.* **2006**, *15*, 2087-2097.

(5) Mankodi, A. *Science.* **2000**, *289*, 1769-1772.

(6) Sobczak, K.; Wheeler, T. M.; Wang, W.; Thornton, C. A. *Mol. Ther.* **2013**, *21*, 380-387.

(7) Pecot, C. V.; Calin, G. A.; Coleman, R. L.; Lopez-Berestein, G.; Sood, A. K. *Nat. Rev. Cancer.* **2011**, *11*, 59-67.

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

(9) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature.* **1996**, *382*, 559-561.

(10) White, S.; Baird, E. E.; Dervan, P. B. *Chem. Biol.* **1997**, *4*, 569-578.

(11) White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Nature.* **1998**, *391*, 468-471.

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

(13) 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.

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

(15) Mrksich, M.; Parks, M. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1994**, *116*, 7983-7988.

(16) Turner, J. M.; Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 6219-6226.

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

(18) Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1996**, *118*, 6160-6166.

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

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

(21) Nickols, N. G.; Dervan, P. B. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 10418-10423.

(22) Nickols, N. G.; Jacobs, C. S.; Farkas, M. E.; Dervan, P. B. *ACS Chem. Biol.* **2007**, *2*, 561-571.

(23) Muzikar, K. A.; Nickols, N. G.; Dervan, P. B. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 16598-16603.

(24) Raskatov, J. A.; Meier, J. L.; Puckett, J. W.; Yang, F.; Ramakrishnan, P.; Dervan, P. B. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 1023-1028.

(25) Nickols, N. G.; Szablowski, J. O.; Hargrove, A. E.; Li, B. C.; Raskatov, J. A.; Dervan, P. B. *Mol. Cancer Ther.* **2013**, *12*, 675-684.

(26) Raskatov, J. A.; Hargrove, A. E.; So, A. Y.; Dervan, P. B. *J. Am. Chem. Soc.* **2012**, *134*, 7995-7999.

(27) Synold, T. W.; Xi, B.; Wu, J.; Yen, Y.; Li, B. C.; Yang, F.; Phillips, J. W.; Nickols, N. G.; Dervan, P. B. *Cancer Chemother. Pharmacol.* **2012**, *70*, 617-625.

(28) Pilch, D. S.; Poklar, N.; Gelfand, C. A.; Law, S. M.; Breslauer, K. J.; Baird, E. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 8306-8311.

(29) Yang, F.; Nickols, N. G.; Li, B. C.; Szablowski, J. O.; Hamilton, S. R.; Meier, J. L.; Wang, C. -M.; Dervan, P. B. *J. Med. Chem.* **2013**, *56*, 7449-7457.

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

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

(32) Puckett, J. W.; Green, J. T.; Dervan, P. B. *Org. Lett.* **2012**, *14*, 2774-2777.

Chapter 5

Towards understanding genome-wide Py-Im polyamide perturbation of Hypoxia-Inducible Factor by ChIP-seq

This research was conducted with Peter B. Dervan (California Institute of Technology); and benefited from valuable help from James W. Puckett (California Institute of Technology).

Abstract

The adaptive response to low oxygen is mediated by the transcription factor Hypoxia-Inducible Factor (HIF) and can be co-opted by cancer cells to promote its survival. A DNA minor groove binding Py-Im polyamide targeted to the hypoxia response element DNA sequence has been previously reported to disrupt a portion of the HIF transcriptional program. In this study, we explore the genome-wide perturbation of HIF occupancy under hypoxia by a Py-Im polyamide employing immunoprecipitation coupled to next generation sequencing, ChIP-seq. The study revealed an enrichment of the HRE consensus motif in sequenced reads and mapped peaks of enrichment in the promoter regions of known hypoxia genes with good resolution. The peaks could be quantified and differential occupancy upon hypoxia induction and reduced occupancy upon polyamide treatment could be measured. This study provides an early step towards revealing the effects of Py-Im polyamides on transcription factor occupancy and understanding its relation to transcriptional modulation.

5.1 Introduction

Cells and tissues have an adaptive transcriptional response to chronic and acute hypoxia, or low oxygen, primarily controlled by the heterodimeric transcription factor Hypoxia-Inducible Factor (HIF).[1,2] The HIF-α and HIF-β heterodimer controls the expression of genes involved in erythropoiesis, angiogenesis, metabolism, and cell death.[3-6] HIF-α functions as an oxygen sensing system through its constituitive translation and immediate, molecular oxygen-dependent proteasomal degradation under normoxic conditions (Figure 5.1).[7] Prolyl hydroxylase enzymes (PHD) were identified as the effectors of HIF-α degradation.[8,9] The PHDs use molecular oxygen, with catalytic iron and 2-oxoglutarate as co-factors, to hydroxylate HIF-α.[10] The hydroxylated HIF-α is recognized by the E3 ubiquitin ligase protein Von-Hippel Lindau (VHL) and tagged with ubiquitin for proteasomal degradation (Figure 5.1).[11] In hypoxic conditions, the lack of molecular oxygen prevents hydroxylation by PHDs and allows HIF-α to escape proteasomal degradation. HIF-α forms a heterodimer with HIF-β, translocates to the nucleus, and binds the hypoxia response element (HRE). The HIF heterodimer, with co-activators p300/CBP, then activates the transcription of genes that regulate cellular and tissue adaptation to hypoxia.[12]

Means to modulate HIF genes have been the subject of investigation due to its potential therapeutic value.[13] Many cancers grow beyond their access to vasculature and create a microenvironment of hypoxia. The adaptive response to hypoxia can give a survival advantage to these cells through the expression of genes that promote

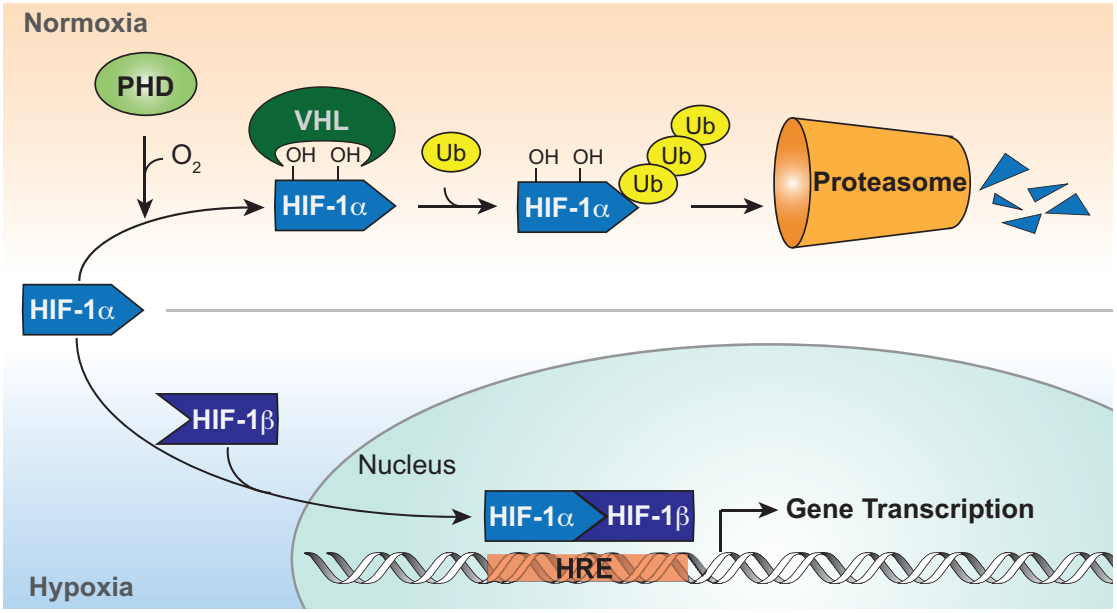


Figure 5.1 Molecular oxygen determines HIF-1 activity. Under normal O_2 levels (normoxia), prolyl hydroxylases (PHD) hydroxylate HIF-1α with molecular oxygen. This is recognized by E3 ligase Von Hippel Lindau protein, which ubiquitinates HIF-1α for proteasomal degradation. Under low oxygen conditions (hypoxia), HIF-1α escapes degradation, heterodimerizes with HIF-1β, translocates to the nucleus, and binds the hypoxia response element (HRE) to activate gene transcription.

angiogenesis, glycolysis, and metastasis.[14,15] While RNA interference technology targeted to the HIF mRNA can abrogate HIF-mediated transcription in a specific-manner, the technology is limited by issues of delivery and distribution.[16,17] Instead, small molecules that bind the DNA binding site of HIF have been investigated for their transcriptional modulation, including a Py-Im polyamide targeted to the consensus hypoxia response element (HRE) 5'-NRCGTG-3'.[18-22] Py-Im polyamides are a class of cell-permeable DNA-binding small molecules with programmable sequence-specificity.[23-28] They have been shown to localize in the nucleus, access chromatin, and bind DNA sequences with affinities and specificities comparable to endogenous transcription factors.[29-32]

The genome-wide transcriptional effects of a Py-Im polyamide **N1** targeted to 5'-WTACGW-3' (structure Figure 5.2A) have been previously characterized by Nickols et al. using microarray technology.[20] The global transcriptome pertubations due to **N1** was assessed in the U251 glioblastoma cell line with desferrioxamine (DFO), an iron chelator, as a model of hypoxia. It was shown that **N1** downregulates by >2 fold 69 genes, out of 297 genes induced >4 fold by treatment with DFO. [20] In comparison, the same study found 244 genes were downregulated by >2 fold by siRNA for HIF-1α.[20] This effect is potentially indicative of the greater sequence specificity of Py-Im polyamides in the variable regions of the HRE. The study by Nichols and coworkers observed correlation of transcriptional changes with HIF occupancy, as measured by PCR and chromatin immunoprecipiation (ChIP) at multiple loci.[20] However, the interrogation of HIF occupancy was limited by the scope of the method. An alternative hypothesis for the mechanism of transcriptional modulation has been proposed, whereby hairpin polyamides disrupt RNA Pol II elongation of transcripts and cause degradation of the RNA Pol II large subunit.[33] This chapter will describe the application of ChIP-seq to assess global HIF transcription factor occupancy and analyze genome-wide perturbation of HIF by hairpin polyamide **1** (Figure 5.2A and 5.2B). Polyamide **1** is targeted to the same DNA sequence and contains the same core aromatic amino acids, and only differs by the location of the amine on the butyric acid turn unit. This study will provide a basis for understanding the gene expression changes previously observed upon polyamide treatment relative to transcription factor displacement.

ChIP-seq, the immunoprecipitation of DNA fragments associated with proteins

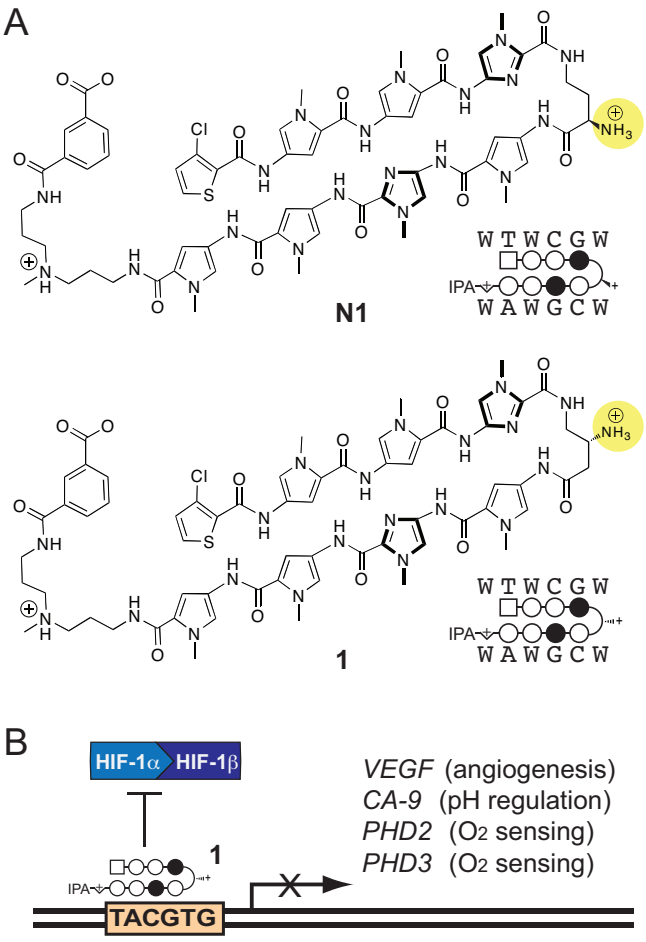


Figure 5.2 Structure of polyamides targeted to the hypoxia response element (HRE). A) Polyamide **N1** was used in a previous study of genome-wide gene expression. Polyamide **1**, which differs in the placement of the amine highlighted in yellow, was used in the present study. B) Schematic diagram of polyamide 1 bound to a sequence-matched HRE sequence preventing HIF-1 binding and disrupting transcription.

followed by next-generation sequencing, allows for the genomic mapping of binding events (Figure 5.3).[34] Software is used to determine the binding sites of transcription factors from the enrichment patterns of derived from the sequencing and mapping of millions of DNA fragments.[35] As the technology queries the entire genome, it eliminates the intrinsic bias of choosing PCR primers or probe sets of limited scope. By comparing

the number of DNA fragments immunoprecipiated per region between varying sample conditions, HIF occupancy can be compared under normoxic, hypoxic, and polyamide **1**-treated conditions genome-wide.

This chapter describes ChIP-seq adapted to assess HIF occupancy under true hypoxia, low oxygen. ChIP-seq was conducted as previously described,[36] with modifications to accommodate the oxygen-dependent lability of HIF on DNA. Though DFO was used as a model of hypoxia in the microarray transcriptome analysis of **N1** in induced U251 cells,[20] its transcriptional effects differ from that of true hypoxia.[37] DFO is an iron chelator that was found to function as a hypoxia mimic by chelating the catalytic iron in the PHD's such that hydroxylation of HIF does not occur.[38] This prevents proteasomal degradation of HIF but likely also has significant off-target effects. The differential expression profiles of "hypoxia" inducing agents reported by Poellinger and coworkers motivates the induction of HIF by 0.5% O_2 hypoxia.[37] Working in a 21% O_2 ambient environment presents a technical challenge of cross-linking HIF to DNA before O_2 levels rise in the cells. It has been reported that HIF association on DNA reaches maximal levels at 0.5% O_2 in 12.4 minutes.[39] In the same study, they found that reoxygenation in ambient O_2 resulted in HIF dissociation beginning at 2 minutes, and near complete dissociation by 16 minutes.[39] Detailed procedures to create and maintain hypoxia and accomplish cross-linking in this timeframe are described in Materials and Methods.

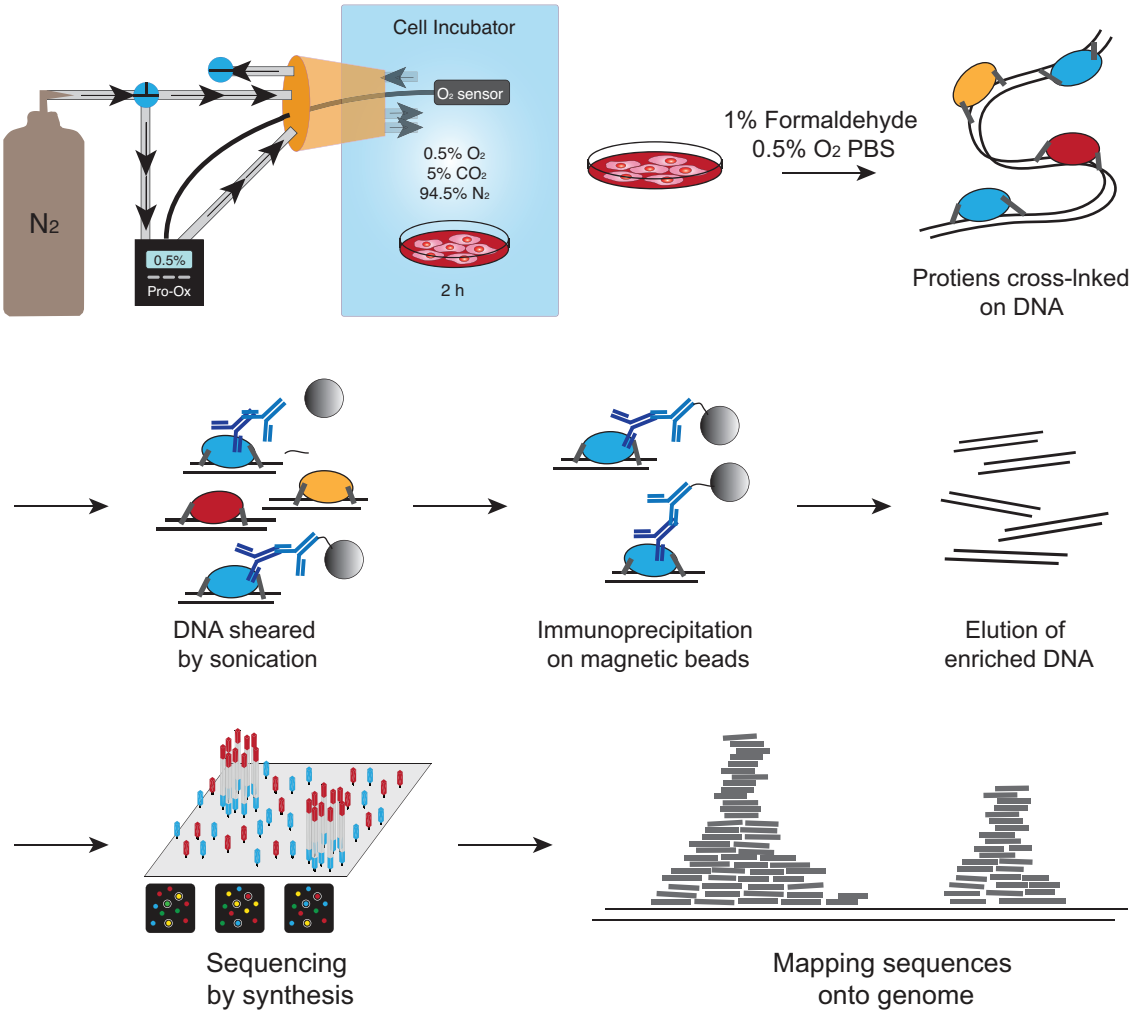


Figure 5.3 Experimental scheme for ChIP-seq with samples prepared in hypoxia. Hypoxia is achieved by nitrogen displacement to 0.5% O2 in an incubator. Proteins on DNA are immediately cross-linked with formaldehyde. DNA is sheared and HIF-1α is enriched by immunoprecipitation. Eluted DNA is sequenced by next-generation sequencing and mapped to the human genome. Enriched samples are compared to background to find enriched peaks using appropriate analysis algorithms.

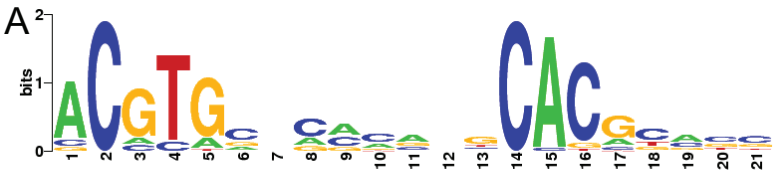
5.2 Results

U251 cells grown under three conditions were prepared for ChIP-seq: 1) untreated cells in normoxia (non-induced, NI); 2) untreated cells subject to 0.5% O_2 hypoxia for 2 hours (hypoxia-induced, Id); and 3) cells treated with 1 μM of hairpin **1** for 48 hours prior to

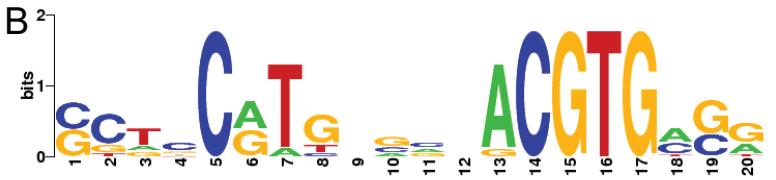
Table 5.1 Sequenced reads categorization for background input samples and enriched immunoprecipitated (IP) samples.

Samples	Raw (millions)	Quality filtered (millions)	Unique (millions)	Multiread (millions)
NI input	14.6	9.0 (61%)	7.0 (78%)	0.6 (3%)
Id input	20.1	12.0 (59%)	9.4 (79%)	0.7 (6%)
1 input	35.6	20.3 (56%)	16.4 (81%)	1.3 (6%)
NI IP	67.6	53.2 (79%)	31.0 (58%)	2.6 (5%)
Id IP	5.8	3.6 (61%)	2.8 (78%)	0.2 (6%)
1 IP	41.3	32.5 (79%)	18.8 (58%)	1.3 (4%)

0.5% O_2 hypoxia for 2 hours (hypoxia-induced and treated with **1**, **1**). Chromatin was isolated and sheared, and a portion saved to serve as the input background control. DNA cross-linked to HIF was immunoprecipitated with an anti-HIF-1α primary antibody and a secondary antibody conjugated to magnetic beads. Immunoprecipitated DNA for each condition (NI_IP, Id_IP, 1_IP) as well as the input background controls (NI_input, Id_input, 1_input) was purified and submitted for sequencing on an Illumina Genome Analyzer IIx. The sequencing produced a range of 5.8 million to 67.6 million raw reads of 50 base pairs per sample (Table 5.1). The reads were then reduced through a quality filter and mapped to the human genome build 19 using the Bowtie algorithm.[40] Of the reads passing the quality filter, 58-81% mapped to a unique genomic locus while the remaining sequences mapped to multiple loci or did not map. Only sequencing reads mapped to unique loci were used for further analysis. The Id_IP sample had the lowest



85% match	Rate per million bp	Motif of found matches at 85%
Genome	2.4	
IP regions	1025.2	



85% match	Rate per million bp	Motif of found matches at 85%
Genome	0.66	
IP regions	630.9	

Figure 5.4 Motifs (A and B) of enriched sites discovered by the findall algorithm. Each motif was searched for matches at a 85% match threshold in the human genome and the immunoprecipitated sequences (IP regions), and its enrichment is shown as a rate per million base pairs. 85% match sequences were then used to generate a motif to compare to the original discovered motif.

number of reads and the reduced sequencing depth results in coarse granularity of the data, but does not otherwise impact the data analysis, as measurements are normalized per million reads.

Using the ERANGE commoncode program "findall",[41] we identified differentially enriched regions between the NI_IP and Id_IP samples, defined as 50%

increase upon hypoxia induction with at least 2 reads per million (RPM) in the induced condition. These identified sequence reads were then entered into the MEME motif finding algorithm. If enrichment of HIF-bound DNA was accomplished, we would expect to find the HRE consensus motif among the discovered motifs. Indeed, we found two discovered motifs (among 10 generated) that contained the HIF binding site motif 5'-NACGTG-3' (Figure 5.4A and 5.4B).[21,22] Each of the motifs was then used to search the human genome and immunoprecipitated regions at a 85% match threshold to assess relative enrichment. The motif shown in Figure 5.4A was found at a rate of 2.4 RPM in the genome, whereas it was enriched to 1025.2 RPM in the HIF precipitated sequence. The motif shown in Figure 5.4B similarly occurred with 603.9 versus 0.66 RPM enrichment in the immunoprecipitated sample, greater than the genome in general. The similarity of the motif generated from the identified sequences at 85% threshold to the original search motif indicates that these sequences are reasonable matches to the motif in question. The identification of the HRE consensus sequence provides evidence that this ChIP-seq method detects HIF transcription factor enrichment on DNA.

We next examined the unique reads mapped to genomic loci associated with genes under HIF transcriptional control. The CA9 promoter region has previously been interrogated by ChIP-qPCR and HIF occupancy observed to be reduced under polyamide N1-treated conditions.[20] In the wigglegram shown Figure 5.5A, there is a peak upstream of the transcription start site in the Id_IP condition higher than in the NI_IP condition showing increased HIF occupancy under hypoxia induction. Further, we observe that the

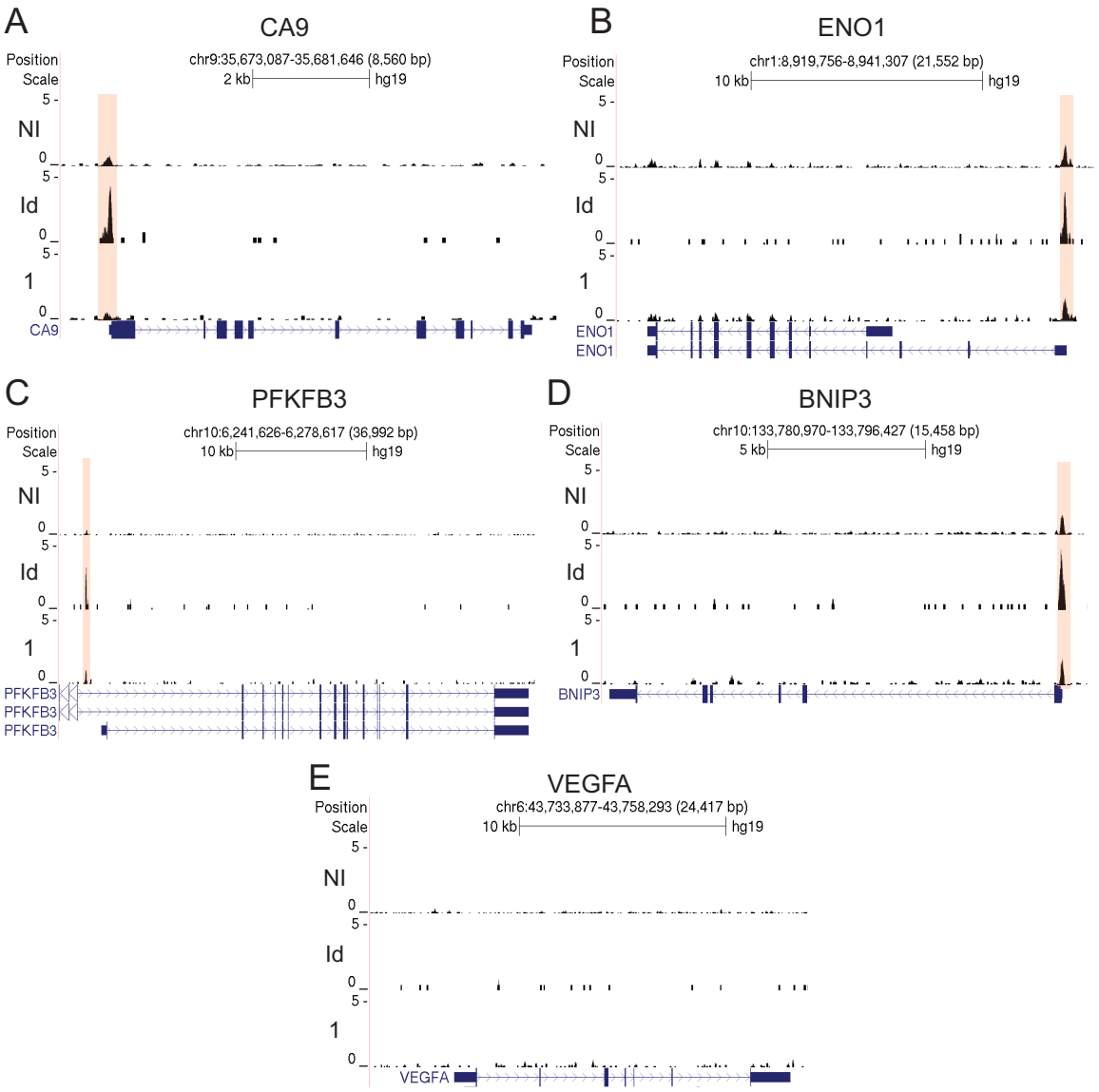


Figure 5.5 Wigglegrams from immunoprecipitated sequences of non-induced (NI), hypoxia-induced (Id), and polyamide **1**-treated (1) samples mapped to the human genome. Enriched region highlighted in orange. A) CA9 gene, B) ENO1, C) PFKFB3, D) BNIP3, and E) VEGFA, which showed no enrichment in the promoter region. RefSeq gene map shown below wigglegrams with gene directionality.

peak is reduced in the PA1-treated condition. This was observed for multiple genes under HIF control, including ENO1, PFKFB3, and BNIP3 (Figure 5.5B-D). Interestingly, a peak was not observed for VEGFA, though it was reported previousy with DFO induction. More replicates are necessary to determine whether this is a reproducible

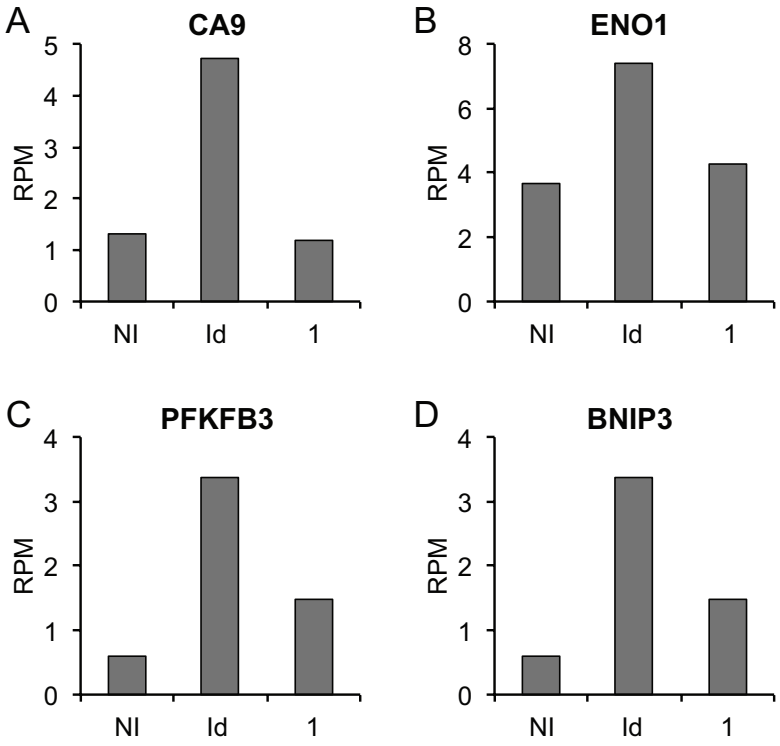


Figure 5.6 Quantitation of sequence reads within a 150 base pair window of peaks identified by findall in the promoter regions of A) CA9, B) ENO1, C) PFKFB3, and D) BNIP3. Sequencing depth is normalized by reads per million (RPM).

difference between DFO and hypoxia induction. The data supports identification of HIF enrichment at distinct loci in the promoter regions of HIF-controlled genes by this method.

The differential HIF occupancy in this region can be further quantified by counting the number of mapped unique reads within a 150 base pair radius of the center of the peak (300 base pair window). Figure 5.6A-D shows the quantitation of the reads in a 300 base pair window around the peak discovered by findall for the genes in Figure 5.5. These loci were among those selected by the findall algorithm after a cutoff of 50% minimum increase of HIF occupancy upon induction. In fact, the CA9, PFKFB3, and

BNIP3 all show greater than 2-fold increase in HIF occupancy upon induction. These loci also show reduced occupancy with hairpin **1** treatment, down to near baseline levels for CA9 and ENO1 (Figure 5.6A and B). The decrease in HIF occupancy at these sites after polyamide treatment is a likely factor in reducing transcription at these loci. This study establishes that ChIP-Seq can be a valuable tool towards understanding genome-wide perturbations of HIF occupancy by Py-Im polyamides.

5.3 Discussion

The ChIP-seq method was employed in this exploratory study to assess genome-wide HIF occupancy after induction with 0.5% O_2 hypoxia. Detection of the HRE consensus sequence within the enriched regions suggests that the current method is sufficient for the detection of HIF. The diminution of the occupancy peaks in the promoter regions in the PA1-treated condition, in comparison to the hypoxia-induced condition, indicates this method will be competent to quantify perturbations to occupancy. Further study employing this method in biological replicates will lend statistical power to measurements of occupancy and may reveal global patterns of hairpin polyamide-mediated perturbations of HIF. Comparison with RNA Pol II occupancy, epigenetic marks, and other transcription factors may suggest mechanisms by which Py-Im polyamides affect transcription.

5.4 Materials and methods

Py-Im polyamide synthesis. Hairpin polyamide **1** was synthesized by microwave-assisted, solid-phase synthesis on Kaiser oxime resin (Novabiochem)

according to previously described protocols.[42,43] Purity and identity was verified by analytical HPLC and matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry. MALDI-TOF: Expected 1394.5, Found 1394.1.

ChIP-seq. U251 cells were plated onto 3 square 500 cm^2 plates (Corning) per condition at 2 million per plate in 80 ml of RPMI media (Gibco) and left to adhere overnight. Cells were untreated or treated with polyamide **1** for 48 hours prior to induction of hypoxia. Hypoxia was achieved by N_2 displacement until 0.5% oxygen was measured on a Pro-Ox 110 (Biospherix) oxygen detector and maintained for 2 hours.

Phosphate buffered saline (PBS, Gibco) was de-gassed *in vacuo* and formaldehyde added to 1% by volume. The fixation solution was aliquoted 40 ml per sample into 50 ml falcon tubes and pre-equilibrated to 0.5% O_2 for at least 30 minutes. Once the 2 hour induction period was complete, the hypoxia chamber was opened and the media for all samples were immediately dumped. The uncovered square plates were stacked at an offset angle to allow quick addition of the fixation solution. All samples were in the formaldehyde fixation solution within 90 seconds of removal from the hypoxia chamber and fixed for 15 min on a rotary shaker. After fixation, the cells were washed with cold PBS and 3 ml of 2.5 M glycine for 5 minutes and again washed with cold PBS.

Previously established protocols were followed hereafter.[34,36] In brief, nuclei were isolated from Farnham Lysis Buffer containing proteases, and resuspended in RIPA buffer. Chromatin was sonicated with a Branson Digital Sonifier over 25 cycles in a dry ice/EtOH bath, 30 seconds at a time. DNA was sheared to approximately 250 base pair fragments. Samples were centrifuged at 4 ºC at 14000 RPM and the supernatant collected

and quantitated with Bradford's reagent. A fraction was saved as the background input sample. The previous day, anti-mouse magnetic beads (Dynabeads) were washed with PBS containing BSA and incubated overnight at 4 ºC with HIF-1α antibody (NB100-105, Novus Biologicals). The magnetic beads were washed of unconjugated antibody and incubated with 1 mg of protein overnight at 4 ºC. Immunoprecipitated DNA was eluted, extracted by phenol/$CHCl_3$/isoamyl alcohol, and purified using a Qiagen clean up kit. Isolated DNA was quantified by Qubit (Life Technologies) and submitted for sequencing. Sequencing data was analyzed using freely available software as described above.

5.5 References

(1) Wang, G. L.; Jiang, B. H.; Rue, E. A.; Semenza, G. L *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 5510.

(2) Semenza, G. L. *J. Appl. Physiol.* **2000**, *88*, 1474-1480.

(3) Wenger, R. H.; Rolfs, A.; Marti, H. H.; Bauer, C.; Gassmann, M. *J. Biol. Chem.* **1995**, *270*, 27865-27870.

(4) 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.

(5) Semenza, G. L.; Jiang, B. H.; Leung, S. W.; Passantino, R.; Concordet, J. P.; Maire, P.; Giallongo, A. *J. Biol. Chem.* **1996**, *271*, 32529-32537.

(6) Obach, M.; Navarro-Sabaté, A.; Caro, J.; Kong, X.; Duran, J.; Gómez, M.; Perales, J. C.; Ventura, F.; Rosa, J. L.; Bartrons, R. *J. Biol. Chem.* **2004**, *279*, 53562-53570.

(7) 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.

(8) Bruick, R. K.; McKnight, S. L. *Science.* **2001**, *294*, 1337-1340.

(9) Epstein, A. C.; Gleadle, J. M.; McNeill, L. A.; Hewitson, K. S.; O'Rourke, J.; Mole, D. R.; Mukherji, M.; Metzen, E.; Wilson, M. I.; Dhanda, A.; Tian, Y. M.; Masson, N.; Hamilton, D. L.; Jaakkola, P.; Barstead, R.; Hodgkin, J.; Maxwell, P. H.; Pugh, C. W.; Schofield, C. J.; Ratcliffe, P. J. *Cell.* **2001**, *107*, 43-54.

(10) Jaakkola, P.; Mole, D. R.; Tian, Y. M.; Wilson, M. I.; Gielbert, J.; Gaskell, S. J.; Kriegsheim Av; Hebestreit, H. F.; Mukherji, M.; Schofield, C. J.; Maxwell, P. H.; Pugh, C. W.; Ratcliffe, P. J. *Science.* **2001**, *292*, 468-472.

(11) Pugh, C. W.; Ratcliffe, P. J. *Semin. Cancer Biol.* **2003**, *13*, 83-89.

(12) Arany, Z.; Huang, L. E.; Eckner, R.; Bhattacharya, S.; Jiang, C.; Goldberg, M. A.; Bunn, H. F.; Livingston, D. M. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 12969-12973.

(13) Melillo, G. *Cancer. Metastasis Rev.* **2007**, *26*, 341-352.

(14) Rankin, E. B.; Giaccia, A. J. *Cell Death Differ.* **2008**, *15*, 678-685.

(15) Semenza, G. L. *Curr. Opin. Genet. Dev.* **2010**, *20*, 51-56.

(16) Meister, G.; Tuschl, T. *Nature.* **2004**, *431*, 343-349.

(17) Pecot, C. V.; Calin, G. A.; Coleman, R. L.; Lopez-Berestein, G.; Sood, A. K. *Nat. Rev. Cancer.* **2011**, *11*, 59-67.

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

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

(20) Nickols, N. G.; Jacobs, C. S.; Farkas, M. E.; Dervan, P. B. *ACS. Chem. Biol.* **2007**, *2*, 561-571.

(21) Wenger, R. H.; Stiehl, D. P.; Camenisch, G. *Sci. STKE.* **2005**, *2005*, re12

(22) Schödel, J.; Oikonomopoulos, S.; Ragoussis, J.; Pugh, C. W.; Ratcliffe, P. J.; Mole, D. R. *Blood.* **2011**, *117*, e207-17.

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

(24) Trauger, J. W.; Baird, E. E.; Dervan, P. B. *Nature.* **1996**, *382*, 559-561.

(25) White, S.; Baird, E. E.; Dervan, P. B. *Chem. Biol.* **1997**, *4*, 569-578.

(26) White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Nature.* **1998**, *391*, 468-471.

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

(28) 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.

(29) 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.

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

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

(32) 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.; Dervan, P. B. *Tetrahedron.* **2007**, *63*, 6146-6151.

(33) Yang, F.; Nickols, N. G.; Li, B. C.; Marinov, G. K.; Said, J. W.; Dervan, P. B. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 1863-1868.

(34) Johnson, D. S.; Mortazavi, A.; Myers, R. M.; Wold, B. *Science.* **2007**, *316*, 1497-1502.

(35) Pepke, S.; Wold, B.; Mortazavi, A. *Nat. Methods.* **2009**, *6*, S22-S32.

(36) Puckett, J. W. *Microarray and genome-wide sequencing approaches to characterizing DNA binding molecules*; California Institute of Technology, PhD Dissertation: 2009.

(37) Lendahl, U.; Lee, K. L.; Yang, H.; Poellinger, L. *Nat. Rev. Genet.* **2009**, *10*, 821-832.

(38) Schofield, C. J.; Ratcliffe, P. J. *Nat. Rev. Mol. Cell. Biol.* **2004**, *5*, 343-354.

(39) Jewell, U. R.; Kvietikova, I.; Scheid, A.; Bauer, C.; Wenger, R. H.; Gassmann, M. *FASEB. J.* **2001**, 7321.

(40) Langmead, B.; Trapnell, C.; Pop, M.; Salzberg, S. L. *Genome Biol.* **2009**, *10*, R25.

(41) Mortazavi, A.; Williams, B. A.; McCue, K.; Schaeffer, L.; Wold, B. *Nat. Methods.* **2008**, *5*, 621-628.

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

(43) Puckett, J. W.; Green, J. T.; Dervan, P. B. *Org. Lett.* **2012**, *14*, 2774-2777.