Chapter 1

INTRODUCTION

Background and significance

Biological systems are complex and difficult to understand. However, understanding and utilizing knowledge in biology can yield great benefits to mankind. Thanks to scientific development, fewer diseases threaten humanity than ever before, yet the ones that cause most harm represent some of the most complex scientific challenges known to date. To understand, and eventually treat, those diseases we will need methods to probe diseased cells in an understandable, programmable and consistent fashion. Many aspects of this problem can be understood with help of molecular recognition.

One of the most important interfaces in biology – the major and minor grooves in the DNA double helix – is essential for storing and reading biological information. However, it is also an attractive target for programmable molecules. The DNA has a predictable and repetitive structure where information is encoded in a single dimension – along its axis. No other biological interface represents a readable information in such simple format and is thus amenable for sequence-specific molecular recognition with relatively simple molecules. DNA-binding molecules have been known to biologists even prior to the discovery of the DNA structure (1); however, it was not until the 1960s when their major binding modes were recognized. Over the next decades, scientists recognized DNA intercalators (2), minor- and major-groove, and covalent binders (1).

Many of those molecules were tremendously useful in biological research and medicine, particularly in chemotherapy (3). One particular breakthrough made it possible to distinguish base-pairs in the minor groove of the DNA in a programmable and modular fashion using Pyrrole-Imidazole polyamides (Py-Im polyamides with affinities strong enough to displace transcription factors and modulate gene expression (4-10). The development of Py-Im polyamides sparked an area of research on the verge of biology and chemistry, where a structurally simple biological interface – the DNA double helix – could be affected through a

programmable (11), cell-permeable molecule (12, 13) with a well understood physical mode of action (4). An ability to probe cells with such molecules yields a great promise in deciphering complex biological systems and diseases, and potentially could result in rationally designed drugs for a large number of distinct diseases. However, before that is possible, more needs to be understood about the interaction of Py-Im polyamides and cells in living organisms.

The structure and function of Nucleic Acids

DNA is a biological polymer specialized in passing on the information in living organisms. It is composed from four building blocks, or nuclotides. These building blocks are largely similar: they all contain a sugar and a phosphate, but they also contain a nitrogen-containing base that varies between different nucleotides. The five carbon-sugar building nucleotides is a deoxyribose, covalently linked to a phosphate at 5' carbon through a phosphodiester bridge. Together, dexyribose and phosphate form a (14) bridge with 3' hydroxyl in another nucleotide, forming a polymer with chemical polarity. The nitrogen-containing bases (Adenine, Guanine, Cytosine and Thymine) are attached to nucleosides by 1' carbon and do not take part in formation of the DNA backbone. The nucleotides are typically described based on the base they contain (A – for adenine, G – for guanine, C – for cytosine, T – for thymine). Nucleotides in a DNA molecule, are often referred to as a 'strand' or a 'chain' of DNA. Two chains of DNA can interact through hydrophobic interactions and hydrogen bonding between their bases and thus form a double-stranded helix (15). The hydrogen bonding between bases is essential for the integrity of the double helix and the strength of that interaction depends greatly on the DNA sequence. Binding between some bases (A and T, or G and C) has a strong energetic advantage, which is described as pairing rules (Fig 1A). The strands in double helix are aligned in an anti-parallel configuration: one strand starts with a 3' hydroxyl and ends in 5' phosphate, whereas the other (complementary) strand runs in an opposite direction. The most commonly described DNA helix – in B-form – is a regular structure, with a pitch of approximately 34 angstroms, a width of approximately 20 angstroms, and a presence of minor and major grooves (Fig 1B). Other structures were found in experimental studies, although it is unclear if they are found in living cells (16). The presence of grooves exposes a fraction of the surface of nucleotides' bases, making them a plausible interface for molecular recognition of DNA.

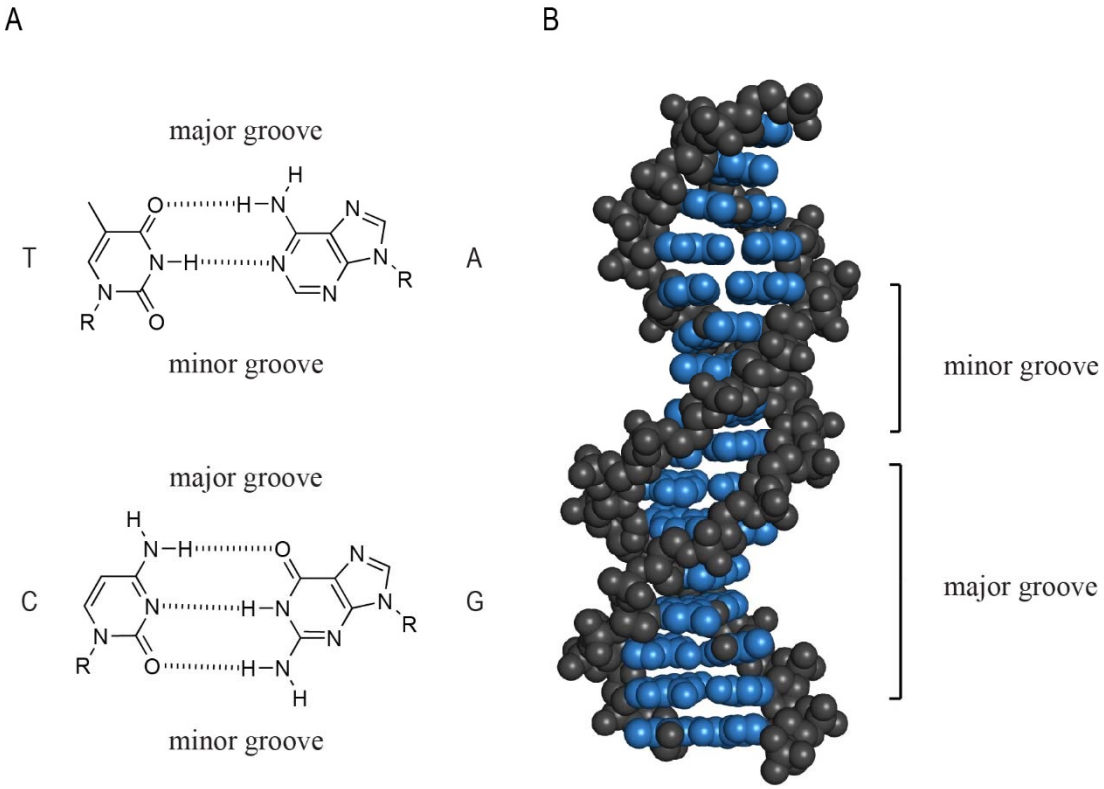


Figure 1.1 Pairing rules and the structure of DNA. A) Hydrogen bonds between bases in nucleotides gives basis to pairing rules in DNA, where Adenine (A) binds to Thymine (T), and Cytosine (C) binds to Guanine (G). B) The structure of DNA. The phosphodiester-linked sugar-phosphate backbone is colored in grey, and the Watson-Crick base pairs are depicted in blue (PDB accession code: 3BSE). Modified from Muzikar Ph.D. thesis (2011).

Molecular Recognition of DNA by minor groove binders

Properties of molecular surfaces in DNA grooves are dependent on the DNA sequence and can form a basis of sequence-specific molecular recognition. These differences are utilized by naturally occurring DNA-binding proteins, using a repertoire of interfaces (17). Most proteins bind to the major groove of DNA but minor groove binding proteins exist as well (17). Many of the DNA-binding proteins show a propensity towards binding to specific sequences – their target sites – providing control over their action throughout the genome. A particularly important class of DNA-binding proteins, transcription factors, utilize their binding sequences

to control transcription of RNA within the living cells. The presence of transcription factors allows cells to modulate production of RNA in response to external stimuli in a highly controlled manner; however, from a chemical perspective, sequence specific molecular recognition is a great challenge. Proteins utilize electrostatics, van-der Walls interactions and hydrogen bonding for sequence specific recognition. The 3-dimensional structure of the transcription factors has evolved to allow for best use of those chemical interactions, giving rise to several major classes of DNA-binding motifs (examples in Fig. 1.2) (17). Many of the transcription factors bind as monomers; however, multimeric complexes exist as well, thus allowing for binding of extended DNA sequences. Additionally, transcription factors often take part in formation of transcription complexes that include several proteins and stretch over DNA sequences longer than those that are typically bound by a single protein.

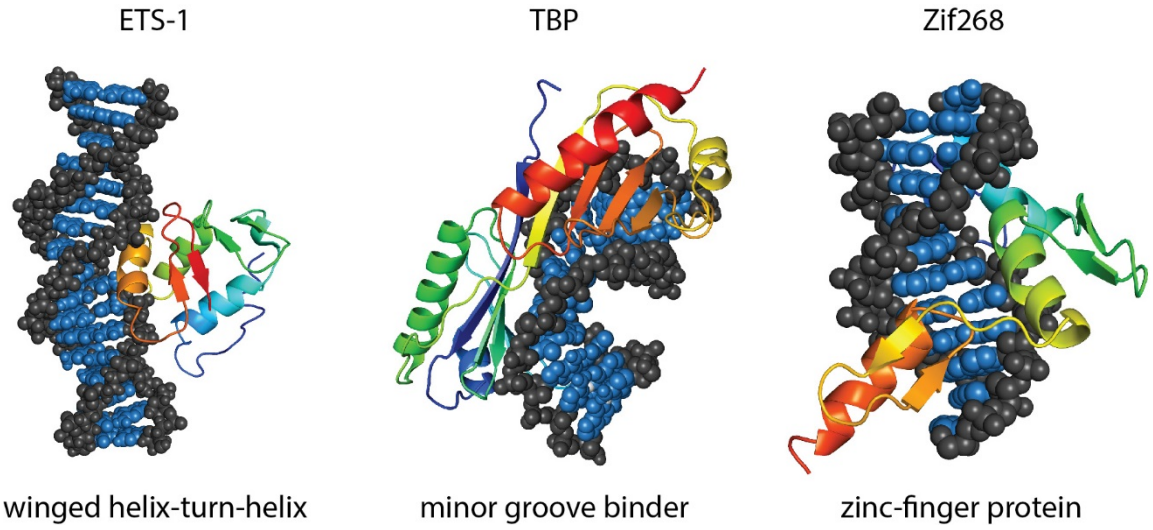


Figure 1.2 Example structure and DNA-binding motifs of three transcription factors: ETS1 (PDB: 2STW), TBP (PDB: 1TGH) and Zif268 (PDB: 1ZAA). Modified from Puckett Ph.D. thesis (2009).

In addition to DNA-binding proteins, small molecules capable of binding to DNA grooves exist as well, both in nature and chemically synthesized. The first confirmed minor groove binder was a natural product, Netropsin (18-20). Soon afterwards, a structure of Distamycin was solved (21) and a few years later a new binding mode was discovered – one where two molecules of Distamycin bind to a single minor groove (2:1 binding) (22, 23). Advances in the field of small-

molecule minor grooves inspired development of sequence-specific, heterocycles (Pyrrole-Imidazole polyamides) that were capable of binding a large repertoire of DNA sequences with affinities and specificities comparable to transcription factors (4, 24). Utilizing a structure similar to Distamycin A, oligomers containing Pyrrolle (Py) and Imidazole (Im) linked through an amide bond were developed to recognize both A:T and G:C basepairs in DNA. While a pyrrole-containing distamycin binds A:T rich sequences, the recognition of G:C was posited to be achieved using an amine in N-methylpyrrole's interacting with exoclyclic amine in guanine base (Fig. 1.3). Footprinting experiments proved that a developed Py-Im polyamide bound a predicted DNA sequence (5'-WGWCW-3'), was capable of recognizing G:C pairs, and bound in 2:1 binding mode previously seen for Distamycin A (25). Subsequent studies led to development of Py-Im polyamides of additional monomers including 3-Chlorothiophene (Ct) (26) and N-methyl-3-hydroxypyrrole (Hp) (27), both distinguishing A:T from T:A pairs. Consequently, the Py-Im polyamides became the first small molecule minor groove binder capable of distinguishing four naturally occurring Watson-Crick base pairs (27). Several methods of linking linear Py-Im polyamides were developed to increase binding affinity (28-31). The design most commonly used today is a 'hairpin' (31) polyamide where a γ-aminobutyric acid derivative ('turn' monomer) links carboxylic terminus of a Py-Im oligomer with the amino terminus of another (Fig. 1.3, bottom panel; (32)). Further improvement in binding affinity is achieved by using a chiral 2,4-diaminobutyric, or 3,4-diaminobutyric acid as a 'turn' monomer ((33, 34)). Resultant affinities for hairpin polyamides vary between low-nanomolar to sub-nanomolar (11). The GBA derivatives used confer an energetic advantage in binding to A:T, over G:C pairs (33), thus taking an activate part in DNA sequence recognition. The increased binding affinity upon inclusion of a charged amino in the 'turn' monomer is a credit to both the electrostatic interaction of the positive charge in amine and the DNA backbone and improved alignment of the Py and Im residues (31, 33-35). The alignment can be additionally improved by introduction of a structurally flexible b-alanine residue, which helps to match the curvature of Py-Im polyamide molecules with the one of DNA and targets A:T and T:A residues (36). The final structural feature of a hairpin polyamide used today is the inclusion of a 'tail' monomer on the C-terminus (typically 3,3'-diamino-N-methyl-dipropylamine) which targets T:A and A:T residues (12, 37).

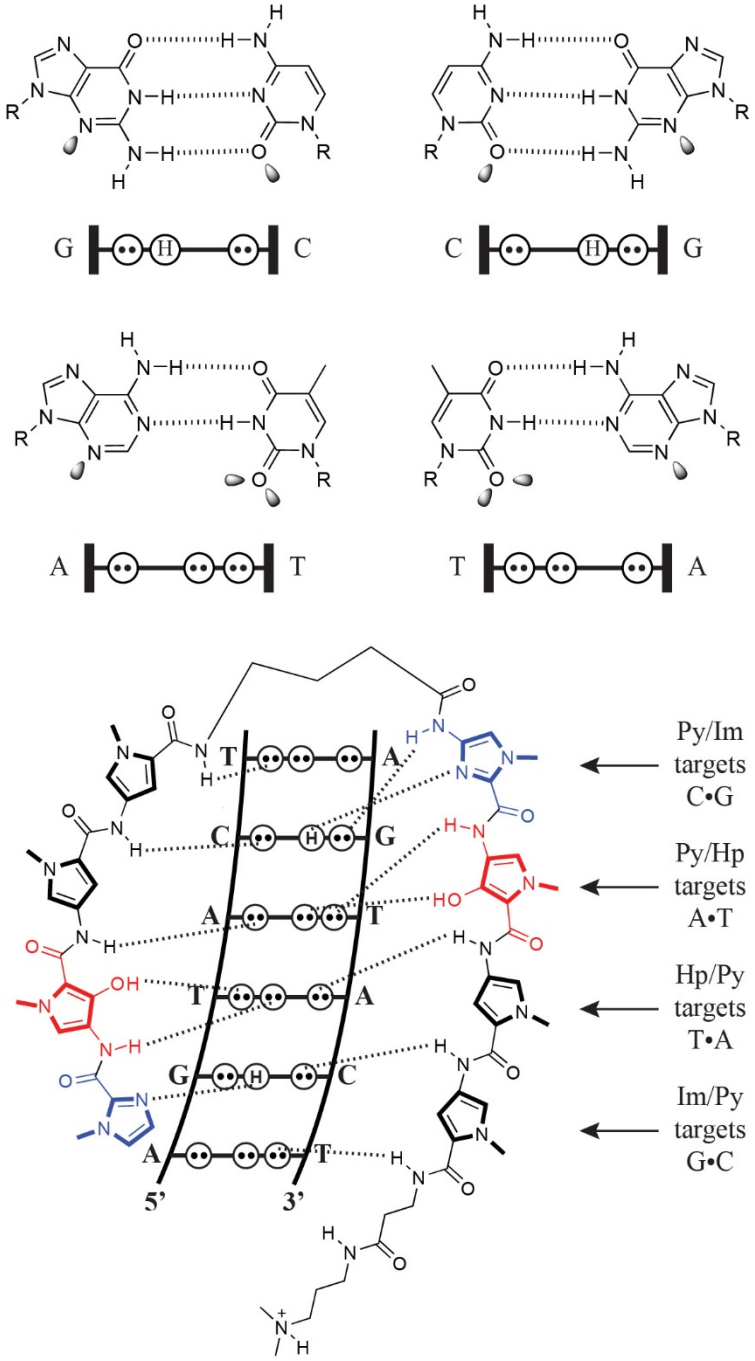


Figure 1.3 Molecular recognition of DNA minor groove by a hairpin Py-Im polyamide. (a) Hydrogen-bonding patterns in Watson-Crick base pairs with depicted lone pairs and hydrogens. Circles with dots represent electron lone pairs N(3) of purines and O(2) of pyrimidines, while H in circle represents the 2-amino group of guanine. Shaded orbitals represent lone pairs projecting into the minor groove. (b) Dashed lines represent hydrogen bonding in a complex between ImHpPyPyg-ImHpPyPy-b-Dp and a 5'-TGTACA-3' sequence(4).

Modulating gene expression with Py-Im polyamides

Hairpin Py-Im polyamides have been used to regulate gene expression in tissue culture. While the exact mechanism of gene regulation is unclear, previous *in vitro* and tissue culture experiments suggest displacement of transcription factors from DNA is one of the candidates, with possibly other effects playing a role. Hallmark experiments in our group have indicated that Py-Im polyamides are capable of displacing transcription factors in gel shift assays (9, 10, 38-42) and change their occupancy in promoters in the tissue culture setting (5, 8, 9, 39, 43). Regardless of the mechanism, Py-Im polyamides might prove useful in both research and treatment of diseases if they are capable of regulating expression of genes important for pathogenesis. The first example of gene regulation in vivo was performed in xenografts derived from A549 cells, and treated with a hairpin compound targeted to 5'-WGGWWW-3' DNA sequence (7). Soon after, Py-Im mediated gene expression modulation of Estrogen-Receptor driven reporter (6), genes related in angiogenesis and metastasis (unpublished), and tumor growth inhibition (unpublished, (43, 44)) were shown in our group. Further investigations showed possible mechanisms of polyamide-induced toxicity on cellular level (43, 45). In this work we set out to describe the anti-tumor effects of Py-Im polyamides on the organism and tissue levels.

Py-Im polyamides in treatment of disease

Transcription factors are involved in both homeostatic gene regulation and in pathogenesis of various diseases. While cancer is perhaps the most well-known example of a disease arising due to dysregulation of cellular signaling and gene expression (46), changes in activity of transcription factors is a hallmark of other diseases as well. In a research setting, Py-Im polyamides were used to inhibit function of viral proteins (41, 47), treat renal failure (48, 49), alleviate a fatty liver disease in animal models (unpublished) and notably – to inhibit growth of tumors (unpublished, (43, 44)) and reduce cancer cell invasion (50). This work focuses mainly on two of the transcription factors: Estrogen Receptor (ER), involved in progression of breast and uterine cancers, and Hypoxia-Inducible-Factor-1 (HIF-1), which plays a role in pathogenesis of diseases including many cancers, tissue fibrosis, chronic heart disease, and age-related macular degeneration (51). The first *in vivo* studies in our laboratory have shown that Py-Im polyamides can be used in living organisms with tolerable toxicity and measurable efficacy (6, 7, 44, 52, 53).

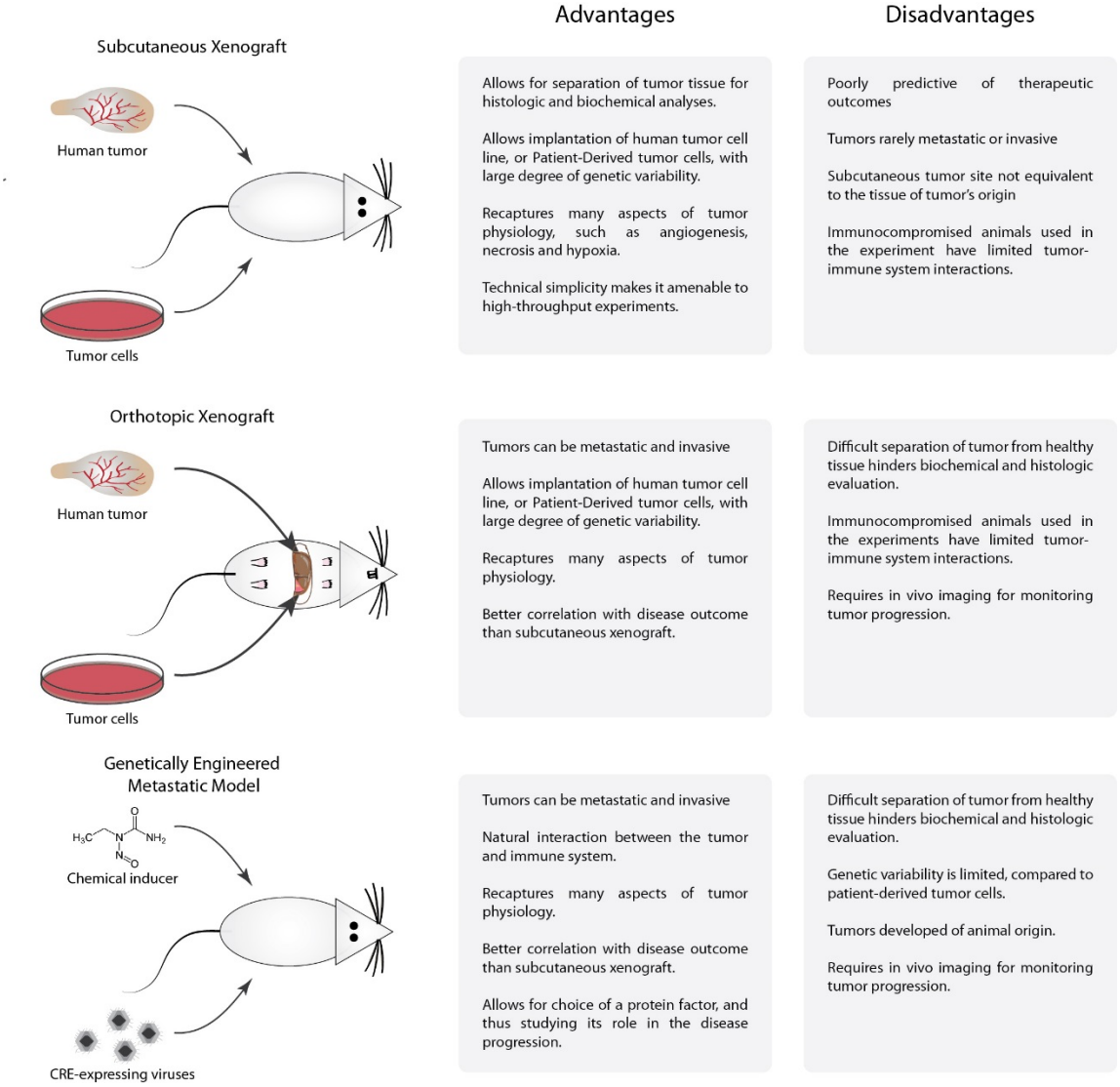


Figure 1.4. Commonly used murine cancer models: subcutaneous and orthotopic xenografts and genetically engineered models.

However, determining mechanism of action and efficacy of new compounds is a challenge. Before clinical trials, researchers typically employ use of animal models, with organisms including mice, rats, dogs, and non-human primates (54). Among them, the mice are likely the most common model organism, due to their small size and small associated research costs. There are three main classes of animal models in cancers: subcutaneous xenografts, orthotopic xenografts and genetically engineered metastatic models (GEMM) (55, 56). Each of these models has pros and cons. The subcutaneous xenograft is among the simplest of the animal models of disease.

In this model, cancer cells are implanted into a subcutaneous cavity of an immunocompromised animal using an injection of cells, with, or without an exogenous extracellular matrix (ECM), such as matrigel. This allows for a clear delineation of tumor tissue and thus a definitive histologic, biochemical, and anatomic analysis. These features make a subcutaneous xenograft a good model in understanding molecular and histologic principles behind the drug action. Another advantage of using a xenograft is an ability to implant cells derived from a human tumor along with the molecular characteristics typical of human tumors. Finally, unlike for GEMM, the results can be obtained within weeks, instead of months or even years. The disadvantages include: a relatively poor correlation in drug efficacy between a subcutaneous xenograft and human cancers, lack of metastasis in most engrafted cell lines and lack of a fully functional immune system in immunocompromised mice (54, 55). Another commonly used model of cancer – orthotopic xenograft – includes implantation of cancer cells into a mouse tissue from which the cancer cells were originally derived. This method shares many of the advantages of subcutaneous xenografts – rapid tumor growth and the ability to implant patient-derived tumor cells. Additional advantages, include a more common presence of metastasis and better correlation of drug efficacy with human disease when compared to subcutaneous xenografts (55, 57). However, the physical separation of a tumor tissue in an orthotopic xenograft is problematic, making the histologic and biochemical analyses more difficult, while tracking the disease progress requires medical imaging.

Finally, the Genetically Engineered Metastatic Models (GEMM), where an organism, typically a mouse, is genetically modified such that factors involved in cancer progression are mutated, or over- or underexpressed. These changes can be chemically induced, or can be constitutively present in mice, depending on the specific model. The genetic changes lead to increased rate of cancers that can be then studied and treated in preclinical setting. This method has many advantages: in some cases the therapeutic response in this model correlates well with responses seen in patients (58-60). It also allows for studying development of the disease with a competent immune system present and these models are, by definition, metastatic. Finally, an ability to choose a factor to be mutated allows for studying its role in tumor progression. The disadvantages are unfortunately present as well: while the mutations can be chosen to mimic human tumors, they might not fully represent the complexity and heterogeneity of the molecular

changes found in human tumors. Additionally, since the tumors obtained in GEMM are of mouse origin, and not human, and drug response might not correlate directly in a clinical setting (55).

Overall, initial studies with Py-Im polyamides are done in a subcutaneous tumor setting, due to the mechanistic nature of our investigations. However, future studies will likely focus on more advanced disease models, in order to understand the potential of Py-Im polyamides in the treatment of disease.

Scope of this work

This work aims at connecting the molecular mechanism of action of Py-Im polyamides and their action to the tissue- and organism levels. In Chapter 2 of this thesis, selectivity of polyamide in tissue culture is presented, and the limits of selectivity discussed. Chapter 3 extends the idea of Py-Im polyamide's functional selectivity and describes genomic and *in vivo* effects of a Py-Im polyamide targeted to Estrogen Receptor Elements (ERE). Chapter 4 delves into the details of polyamide biodistribution and describes the relationship between the tissue uptake of compounds and the tumor type, vasculature, and experimental methodology. Application of this knowledge is presented in Chapter 5. The text of this chapter links biological effects of Py-Im polyamide treatment to changes at molecular and tissue level. It shows that action of a polyamide targeted to Hypoxia Responsive Element (HRE) induces molecular changes in tumors that are consistent with its proposed mechanism of action: interference with hypoxic gene expression. This chapter also shows a potent anti-tumor and anti-angiogenic effects of the anti-HRE polyamide in two different cell lines. In Chapter 6, the anti-tumor effects and the inhibition of hypoxic gene expression by anti-HIF Py-Im polyamide are reinforced in another model of cancer: multiple myeloma xenografts. We explain the effects of polyamide treatment on protein expression, hypoxia-induced apoptosis and tumor growth of multiple myeloma xenografts in two different models: subcutaneous tumor and orthotopic. The final chapter links the described studies and describes possible future directions and potential utility in using polyamides for treatment in non-cancerous diseases.

References

1. Demeunynck M, Bailly C, & Wilson WD (2003) *Small molecule DNA and RNA binders : from synthesis to nucleic acid complexes* (Wiley-VCH, Weinheim).
2. Lerman LS (1961) Structural considerations in the interaction of DNA and acridines. *J Mol Biol* 3:18-30.
3. Airley R (2009) *Cancer Chemotherapy: Basic Science to the Clinic* (Wiley).
4. Dervan PB & Edelson BS (2003) Recognition of the DNA minor groove by pyrrole-imidazole polyamides. *Curr Opin Struct Biol* 13(3):284-299.
5. Nickols NG, Jacobs CS, Farkas ME, & Dervan PB (2007) Modulating hypoxia-inducible transcription by disrupting the HIF-1-DNA interface. *ACS Chem Biol* 2(8):561-571.
6. Nickols NG*, et al.* (2013) Activity of a Py-Im polyamide targeted to the estrogen response element. *Mol Cancer Ther* 12(5):675-684.
7. Raskatov JA*, et al.* (2012) Gene expression changes in a tumor xenograft by a pyrrole-imidazole polyamide. *Proc Natl Acad Sci U S A* 109(40):16041-16045.
8. Raskatov JA*, et al.* (2012) Modulation of NF-kappaB-dependent gene transcription using programmable DNA minor groove binders. *Proc Natl Acad Sci U S A* 109(4):1023-1028.
9. Nickols NG & Dervan PB (2007) Suppression of androgen receptor-mediated gene expression by a sequence-specific DNA-binding polyamide. *Proc Natl Acad Sci U S A* 104(25):10418-10423.
10. Olenyuk BZ*, et al.* (2004) Inhibition of vascular endothelial growth factor with a sequence-specific hypoxia response element antagonist. *Proc Natl Acad Sci U S A* 101(48):16768-16773.
11. Hsu CF*, et al.* (2007) Completion of a Programmable DNA-Binding Small Molecule Library. *Tetrahedron* 63(27):6146-6151.
12. Best TP, Edelson BS, Nickols NG, & Dervan PB (2003) Nuclear localization of pyrrole-imidazole polyamide-fluorescein conjugates in cell culture. *Proc Natl Acad Sci U S A* 100(21):12063-12068.
13. Nickols NG, Jacobs CS, Farkas ME, & Dervan PB (2007) Improved nuclear localization of DNA-binding polyamides. *Nucleic Acids Res* 35(2):363-370.
14. White S, Baird EE, & Dervan PB (1997) On the pairing rules for recognition in the minor groove of DNA by pyrrole-imidazole polyamides. *Chem Biol* 4(8):569-578.
15. Watson JD & Crick FH (1953) The structure of DNA. *Cold Spring Harb Symp Quant Biol* 18:123-131.
16. Brown TA (2007) *Genomes 3* (Garland Science Pub., New York) 3rd Ed pp xxii, 713 p.
17. Brändén CI & Tooze J (1999) *Introduction to Protein Structure* (Garland Pub.).
18. Kopka ML, Yoon C, Goodsell D, Pjura P, & Dickerson RE (1985) The molecular origin of DNA-drug specificity in netropsin and distamycin. *Proc Natl Acad Sci U S A* 82(5):1376-1380.
19. Kopka ML, Yoon C, Goodsell D, Pjura P, & Dickerson RE (1985) Binding of an antitumor drug to DNA, Netropsin and C-G-C-G-A-A-T-T-BrC-G-C-G. *J Mol Biol* 183(4):553-563.
20. Dickerson RE & Kopka ML (1985) Nuclear Overhauser data and stereochemical considerations suggest that netropsin binds symmetrically within the minor groove of poly(dA).poly(dT), forming hydrogen bonds with both strands of the double helix. *J Biomol Struct Dyn* 3(3):423-431.
21. Coll M, Frederick CA, Wang AH, & Rich A (1987) A bifurcated hydrogen-bonded conformation in the d(A.T) base pairs of the DNA dodecamer d(CGCAAATTTGCG) and its complex with distamycin. *Proc Natl Acad Sci U S A* 84(23):8385-8389.

22. Chen X, Ramakrishnan B, Rao ST, & Sundaralingam M (1994) Binding of two distamycin A molecules in the minor groove of an alternating B-DNA duplex. *Nat Struct Biol* 1(3):169-175.
23. Pelton JG & Wemmer DE (1989) Structural characterization of a 2:1 distamycin A.d(CGCAAATTGGC) complex by two-dimensional NMR. *Proc Natl Acad Sci U S A* 86(15):5723-5727.
24. Dervan PB (2001) Molecular recognition of DNA by small molecules. *Bioorg Med Chem* 9(9):2215-2235.
25. Mrksich M*, et al.* (1992) Antiparallel side-by-side dimeric motif for sequence-specific recognition in the minor groove of DNA by the designed peptide 1-methylimidazole-2-carboxamide netropsin. *Proc Natl Acad Sci U S A* 89(16):7586-7590.
26. Foister S, Marques MA, Doss RM, & Dervan PB (2003) Shape selective recognition of T.A base pairs by hairpin polyamides containing N-terminal 3-methoxy (and 3-chloro) thiophene residues. *Bioorg Med Chem* 11(20):4333-4340.
27. White S, Szewczyk JW, Turner JM, Baird EE, & Dervan PB (1998) Recognition of the four Watson-Crick base pairs in the DNA minor groove by synthetic ligands. *Nature* 391(6666):468-471.
28. Heckel A & Dervan PB (2003) U-pin polyamide motif for recognition of the DNA minor groove. *Chemistry-a European Journal* 9(14):3353-3366.
29. Melander C, Herman DM, & Dervan PB (2000) Discrimination of A/T sequences in the minor groove of DNA within a cyclic polyamide motif. *Chemistry-a European Journal* 6(24):4487-4497.
30. Greenberg WA, Baird EE, & Dervan PB (1998) A comparison of H-pin and hairpin polyamide motifs for the recognition of the minor groove of DNA. *Chemistry-a European Journal* 4(5):796-805.
31. Mrksich M, Parks ME, & Dervan PB (1994) Hairpin Peptide Motif. A New Class of Oligopeptides for Sequence-Specific Recognition in the Minor Groove of Double-Helical DNA. *Journal of the American Chemical Society* 116(18):7983-7988.
32. Trauger JW, Baird EE, & Dervan PB (1996) Extended hairpin polyamide motif for sequence-specific recognition in the minor groove of DNA. *Chemistry & Biology* 3(5):369-377.
33. Herman DM, Baird EE, & Dervan PB (1998) Stereochemical control of the DNA binding affinity, sequence specificity, and orientation preference of chiral hairpin polyamides in the minor groove. *Journal of the American Chemical Society* 120(7):1382-1391.
34. Dose C, Farkas ME, Chenoweth DM, & Dervan PB (2008) Next generation hairpin polyamides with (R)-3,4-diaminobutyric acid turn unit. *Journal of the American Chemical Society* 130(21):6859-6866.
35. Farkas ME, Tsai SM, & Dervan PB (2007) alpha-diaminobutyric acid-linked hairpin polyamides. *Bioorganic & Medicinal Chemistry* 15(22):6927-6936.
36. Turner JM, Swalley SE, Baird EE, & Dervan PB (1998) Aliphatic/aromatic amino acid pairings for polyamide recognition in the minor groove of DNA. *Journal of the American Chemical Society* 120(25):6219-6226.
37. Belitsky JM, Nguyen DH, Wurtz NR, & Dervan PB (2002) Solid-phase synthesis of DNA binding polyamides on oxime resin. *Bioorg Med Chem* 10(8):2767-2774.
38. Zhang Y*, et al.* (2011) Targeting a DNA binding motif of the EVI1 protein by a pyrrole-imidazole polyamide. *Biochemistry* 50(48):10431-10441.
39. Muzikar KA, Nickols NG, & Dervan PB (2009) Repression of DNA-binding dependent glucocorticoid receptor-mediated gene expression. *Proc Natl Acad Sci U S A* 106(39):16598-16603.

40. Gearhart MD*, et al.* (2005) Inhibition of DNA binding by human estrogen-related receptor 2 and estrogen receptor alpha with minor groove binding polyamides. *Biochemistry* 44(11):4196-4203.
41. Coull JJ*, et al.* (2002) Targeted derepression of the human immunodeficiency virus type 1 long terminal repeat by pyrrole-imidazole polyamides. *J Virol* 76(23):12349-12354.
42. Chiang SY*, et al.* (2000) Targeting the ets binding site of the HER2/neu promoter with pyrrole-imidazole polyamides. *J Biol Chem* 275(32):24246-24254.
43. Yang F*, et al.* (2013) Antitumor activity of a pyrrole-imidazole polyamide. *Proc Natl Acad Sci U S A* 110(5):1863-1868.
44. Yang F*, et al.* (2013) Animal toxicity of hairpin pyrrole-imidazole polyamides varies with the turn unit. *J Med Chem* 56(18):7449-7457.
45. Martinez TF*, et al.* (2014) Replication stress by Py-Im polyamides induces a non-canonical ATR-dependent checkpoint response. *Nucleic Acids Res* 42(18):11546-11559.
46. Hanahan D & Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell* 144(5):646-674.
47. Yang F, Belitsky JM, Villanueva RA, Dervan PB, & Roth MJ (2003) Inhibition of Moloney murine leukemia virus integration using polyamides targeting the long-terminal repeat sequences. *Biochemistry* 42(20):6249-6258.
48. Kajiwara M*, et al.* (2012) Development of pyrrole-imidazole polyamide targeting fc receptor common gamma chain for the treatment of immune-complex related renal disease. *Biol Pharm Bull* 35(11):2028-2035.
49. Matsuda H*, et al.* (2011) Transcriptional inhibition of progressive renal disease by gene silencing pyrrole-imidazole polyamide targeting of the transforming growth factor-beta1 promoter. *Kidney Int* 79(1):46-56.
50. Sato A*, et al.* (2013) Inhibition of MMP-9 using a pyrrole-imidazole polyamide reduces cell invasion in renal cell carcinoma. *Int J Oncol* 43(5):1441-1446.
51. Semenza GL (2014) Oxygen sensing, hypoxia-inducible factors, and disease pathophysiology. *Annu Rev Pathol* 9:47-71.
52. Synold TW*, et al.* (2012) Single-dose pharmacokinetic and toxicity analysis of pyrrole-imidazole polyamides in mice. *Cancer Chemother Pharmacol* 70(4):617-625.
53. Raskatov JA, Hargrove AE, So AY, & Dervan PB (2012) Pharmacokinetics of Py-Im polyamides depend on architecture: cyclic versus linear. *J Am Chem Soc* 134(18):7995-7999.
54. Teicher BA (2010) *Tumor Models in Cancer Research* (Humana Press).
55. Richmond A & Su Y (2008) Mouse xenograft models vs GEM models for human cancer therapeutics. *Dis Model Mech* 1(2-3):78-82.
56. de Jong M, Essers J, & van Weerden WM (2014) Imaging preclinical tumour models: improving translational power. *Nat Rev Cancer* 14(7):481-493.
57. Kerbel RS (2003) Human tumor xenografts as predictive preclinical models for anticancer drug activity in humans: better than commonly perceived-but they can be improved. *Cancer Biol Ther* 2(4 Suppl 1):S134-139.
58. Combest AJ*, et al.* (2012) Genetically Engineered Cancer Models, But Not Xenografts, Faithfully Predict Anticancer Drug Exposure in Melanoma Tumors. *Oncologist* 17(10):1303-1316.
59. Becher OJ & Holland EC (2006) Genetically engineered models have advantages over xenografts for preclinical studies. *Cancer Res* 66(7):3355-3358, discussion 3358-3359.
60. Roper J & Hung KE (2012) Priceless GEMMs: genetically engineered mouse models for colorectal cancer drug development. *Trends Pharmacol Sci* 33(8):449-455.

Chapter 2

SELECTIVITY OF PY-IM POLYAMIDES IN TISSUE CULTURE

ABSTRACT

Py-Im polyamides have excellent sequence specificity *in vitro*, yet little is known about their selectivity in the nuclei of mammalian cells. In this chapter the extent of the functional selectivity of polyamides is assessed in regulation of gene expression in Glucocorticoid signaling. First, mathematical modeling was used to find the most common GRE sequences that can be bound with 8-ring hairpin polyamides. Then a panel of 12 genes and a focused library of polyamides targeting 7 DNA different sequences was used in evaluation of polyamides as a tool for linking sequence of a response element with the gene it controls. Concurrent nuclear localization studies and in-vitro assessment of DNA binding affinity were performed on the library of polyamides to connect chemical properties of polyamides with their gene regulation patterns. Polyamides show a small degree of selectivity; however, the differences are hard to elucidate because of the low potency of some of the compounds. The potent compounds, on the other hand, show few differences in gene expression patterns. Further steps will need to be taken to increase polyamide specificity, without sacrificing potency; in particular more genes may need to tested, e.g. by using RNA-sequencing. Another possibility is using multiple compounds to target the same regulatory sequence and thus increase the specificity of Py-Im polyamides in tissue culture.

Introduction

Binding of Py-Im polyamides to DNA is sequence-specific (1, 2). While in vitro experiments have shown that a single sequence can be targeted, achieving site-specificity in a mammalian cell nucleus is a significantly more challenging task. The main problem in sequence specificity in mammalian cells is the sheer amount genetic material enclosed in the nucleus. For example, DNA in human cells contains $3x10^9$ base pairs, and a 6-base pair sequence would be expect to occur once in every 4^6 bases, assuming every base pair can be recognized. The typically used Py-Im polyamides, however, only recognize between 3 different base pairs, G, C, and W, which means an average frequency of the DNA sequence bound specifically by an 8-ring Py-Im polyamide is expected to be once in every 3^6 basepairs, an equivalent 4.1 million matched binding sites for an average Py-Im polyamide. Another factor present in mammalian cells, but not *in vitro,* is accessibility of DNA in the nucleus. Not every site in the genomic DNA is equally accessible; some of the DNA is densely packed as heterochromatin. It is currently unclear how the binding properties of polyamides change depending on the density of DNA-packing in the nuclei; however, we do know that Py-Im polyamides are capable of binding to nucleosomes (3). Finally, the time of dissociation of a commonly use hairpin polyamide and DNA match site is long ($k_{off} = 10^{-3} – 10^{-4}$ s^{-1}), half the time of dissociation which ranges from minutes to hours (4), which limits diffusion of polyamides within the nucleus. Thus, sequences most frequently bound by Py-Im polyamides might simply be those that are most accessible thanks to diffusion, or DNA packing. Recent experiments evaluated some aspects of the selectivity of Py-Im polyamides in the genome, showing that sequence-specificity might be just one factor in their genomic-DNA binding profile and the chromatin accessibility may also be important (5). In this chapter we investigated the selectivity of Py-Im polyamides in living cells, by testing the expression of a number of genes related to Glucocorticoid Receptor (GR) in the A549 cell line. We also built theoretical kinetic models of DNA-polyamide binding and calculated possible sequence specificities of Py-Im polyamides within the genome.

Background

Mammalian genes are regulated thanks to a complex network of transcription factors (6) and proteins regulating chromatin accessibility (7, 8). How transcription factors bind and control

gene expression is one of the main questions in molecular biology. Investigating the DNA sequence binding to transcription factors historically has been done through DNA-sequencing of the purified DNA bound to transcription factors (9), and subsequently by Electrophoretic Mobility Shift Assay (EMSA) (10) and DNAseI footprinting (11). These methods allowed for study of a single transcription factor binding site at a time. As a result they yielded information about binding affinity between a transcription factor and a DNA sequence, but failed to inform us about the genomic frequency and positions of these sites. It was not until the advent of high throughput genome sequencing and microarray technology that we were able to do this. Currently the most common method of determining transcription factor binding sites is Chromatin Immunoprecipitation followed by sequencing (12) (Chip-Seq), which can inform us about the position of both genes and transcription factor binding sites in the whole genome. Regrettably this method is incapable of establishing a functional link between the transcription factor sites and the genes they control. While in prokaryotes the transcription factors bind proximally to the genes, in mammalian cells this is not always the case (13, 14).

Large distance between regulatory sequences and their gene targets poses a challenging problem in identifying a functional link between them. Currently there are three Chromosome Conformation capture (3C, 4C, 5C) methods that allow one to connect the particular regulatory element with a particular gene (15-17), however the execution of these assays is often complicated (18, 19). Additionally no other method exists that could confirm the findings, and suggested problems with these methods remain untested (18). Consequently, as of now, reliably matching a transcription binding site to the gene it controls requires knocking out the regulatory sequence in cells. Unfortunately, this method requires prior knowledge of both the gene and its' regulatory sequence and many genes are controlled by multiple regulatory sequences. Because of those issues, targeted knockdown is unsuitable for genome-wide mapping. Pyrrole-Imidazole polyamides could be useful in relating regulatory DNA sequences with the gene expression patterns in a high throughput fashion.

Py-Im polyamides bind the minor groove in double-stranded DNA with affinities and specificities comparable to transcription factors. It has been achieved by combining aromatic amino acids, N-methylpyrroles (Py), N-methylimidazoles (Im), and 3-hydroxy-1-methylpyrroles

(Hp), in a short oligomer. A pair of monomers placed over each other in a minor groove can determine pairing rules between polyamides and DNA. According to those rules a Py/Im pair will recognize a C°G pair, Im/Py a G°C, wherease Py°Py pair will be capable of recognizing W°W pairs. Including a hydroxypyrrole instead of a pyrrole will bias binding of a polyamide towards T°A in the case of a Hp/Py pair and A°T in the case of a Py/Hp (1, 20) (Figs. 1.2 and 1.3). Their capability of sequence specific displacement of transcription factors from their binding sites results in an inhibition of gene expression establishing a functional link between regulatory sequences and the genes they control. However, the question that needs to be answered is whether their sequence specificity is high enough for sequence-specific gene regulation in a large mammalian genome.

Glucocorticoid receptor (GR) response pathway is a common model system for gene expression regulation in mammalian cells (7, 14). GR is a cell permeable steroid receptor binding directly (21) to a well-defined Glucocorticoid Response Element (GRE) (Fig. 2.1). There are several thousand GREs scattered across the mammalian genome implicated in gene regulation (14), each containing three highly degenerate nucleotides at positions 7,8, and 9 (Fig. 2.1B,C). These three nucleotides alone constitute 64 distinct classes of GREs that can be targeted with sequence-matched pyrrole-imidazole polyamides developed in our lab. Additionally, the other bases also show sequence variability that can be utilized for that purpose. In an observed case of one GRE driven gene (GILZ) displacement of transcription factors through polyamides results in an inhibited gene expression (22), and if that will be the case with other GRE controlled genes, we will be able to match classes of GREs to the genes they control.

Glucocorticoid receptor pathway is a useful drug target. Glucocorticoids are widely used in medicine as immunosuppresants and are some of the most potent anti-inflammatory drugs on the market (23). These effects, however, come at a price. Glucocorticoids have significant side effects, such as bone and muscle loss, psychoses, cataract and glaucoma, among many others (23). In children, prolonged use of glucocorticoids may negatively affect bone development (24). Many of those side effects, e.g. glaucoma or diabetes, are mediated through transactivation, or expression of anti-inflammatory proteins. This fact has galvanized the development of more selective glucocorticoid receptor agonists (SEGRas) which aim at decreasing transactivation

without affecting transrepression. One SEGRA (ZK 216348) has shown in animal models that the negative side effects of glucocorticoid treatment can be reduced while maintaining anti-inflammatory effects (23).

Likewise, polyamide are effective in downregulation of gene expression. While the majority of side effects of glucocorticoid treatment are due to transactivation, some of them are not. Thus targeting different sequences within GREs by polyamides administered along with standard glucocorticoids can fine-tune the effects of this anti-inflammatory treatment to minimize the side effects and maximize potency for the specific disease.

Evaluation of the genomic landscape of the GREs and polyamide binding sites

Chromatin Immunoprecipitation followed by sequencing (Chip-Seq) identified 4392 loci in the genome that are occupied by GR in human lung adenocarcinoma cells (A549) (14). The positions of these loci are, however, distant from 234 genes that are highly induced upon treatment with 100nM Dexamethasone (Dex), a synthetic agonist of GR. For genes with Dex induced expression, the median distance between the nearest GRE and a transcription start site (TSS) was 11kb, and those genes that were repressed had a median distance of 146 kb (14). The large TSS-GRE distance and its significant variability suggests that one cannot predict which genes are controlled by which GREs based solely on their relative position. The response of the genes also varies in time; particularly repressed genes are affected later in time than induced ones. This and the large distance between nearest GREs suggest that repressed genes are not controlled by promoter-proximal GR binding.

In order to further analyze the dependence of position of GREs and transcription start, I wrote a simple mathematical model assuming their random distribution. For downregulated genes, I generated random locations for both TSS and GREs and then measured their distance in the aproximately 2.1×10^9 basepairs in non-repetitive parts of the human genome (25). Even this crude estimate of the genome size and a very basic model gives a median nearest neighbour distance between TSS and GREs of 164kb, as compared to 146kb in Chip-Seq study (14). This result suggests that the position of GREs and the genes they repress are independent of each other. In this case a common assumption that the gene is controlled by its nearest neighbour is

most likely unfeasible, further suggesting that repressed genes are controlled independent of proximal GRE-promoter binding. The activated genes, on the other hand, show dependence of the position of GREs and TSSes.

I modeled their relative positions by generating set of gene positions and a random distribution of distances (GRE positions) over the mean length of a chromosome (123kb). It appears that the median distance between a TSS and the nearest neighbor (10.7 kb), assuming their random distribution, once again is very close to 11kb, as found by Chip-seq (14). The distribution of distances as modeled also matches the Chip-Seq data (Fig. 2.2). This result suggests that distance between GREs and genes they control may be distributed randomly within the chromosome. This model further supports current belief that Glucocorticoid Receptor signaling occurs through an exceptionally long range interactions (14). The code and parameters used in writing the models can be found in appendix G. Such quantitative considerations show that one cannot assume that a position of a GRE relative to TSS can predict a functional link between the two and Chromosome Conformation Capture methods maybe be necessary to establish such a link.

In order to make an informed decision on which compounds should be synthesized to exert a specific control of gene expression in A549 cells, I analyzed the GR binding sites for enrichment upon Dex induction in Chip-Seq data set from Myers lab (14). If a rare sequence is targeted with a polyamide, it is unlikely that a large number of genes will respond to it. If, on the other hand, a compound binds a wide array of sequences a larger fraction of genes in a panel is expected to be downregulated. In order to establish which sequences are most common among active GREs, I chose to computationally analyze the genome-wide occurence of sequences compatible with DNA-binding profiles of 8-ring hairpin polyamides.

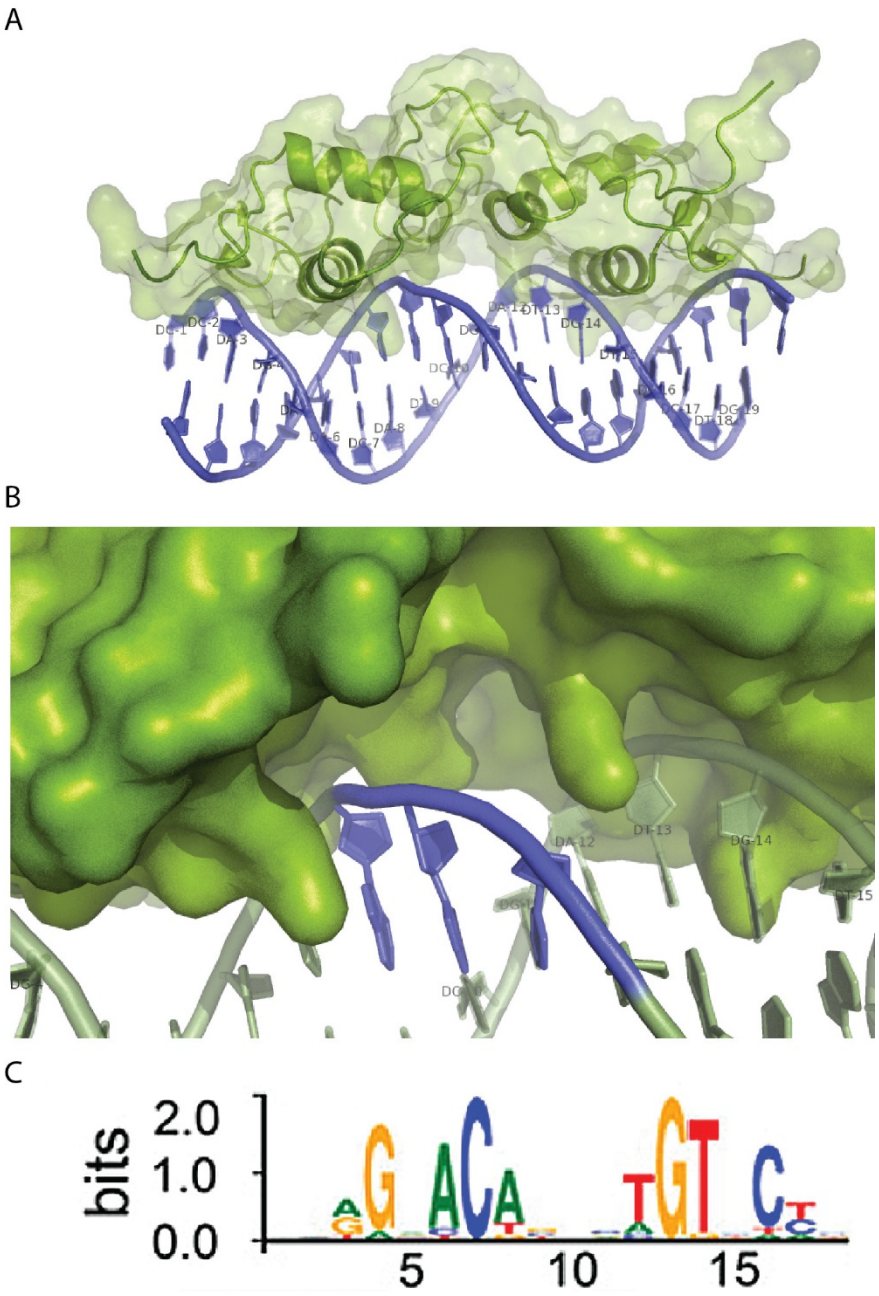


Figure 2.1 X-ray crystal structure of a Glucocorticoid Receptor (GR) bound to DNA (PDB 1R4O). (A) GR binds two DNA as a dimer. Its recognition sequences are nearly palindromic and are separated by a 3-base-pair gap, colored blue on the second inset (B). This gap corresponds to an area without physical GR-DNA interaction. (C) GRE binding motif obtained through a custom analysis of GR Chip- Seq data (14). The sequence variability of this motif allows for sequence-specific targeting of subsets of Glucocorticoid Response Elements (GREs).

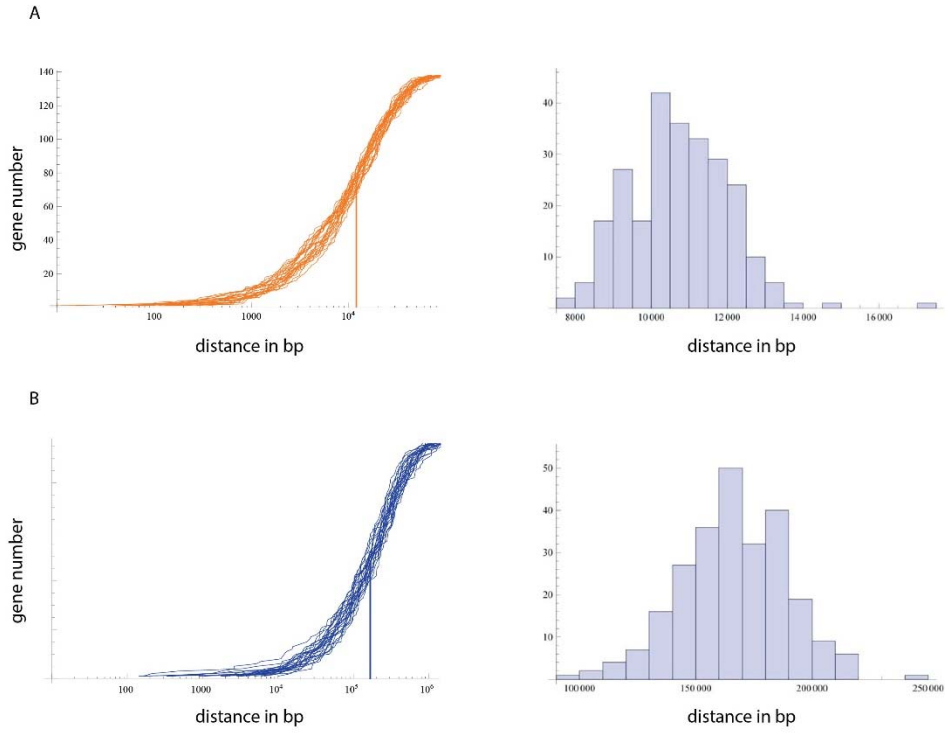


Figure 2.2 Modeling genomic distribution of GREs in relation to transcription starting sites (TSS). (A) modeling distance between GREs and Dexamethasone upregulated genes by placing GREs and TSSes at random within the average size of a human chromosome ($123kb_2$) yields comparable distribution and median distance (10.7kb, n=250) as observed by Chip-Seq (11kb). (B) Modeling distance between GREs and dexamethasone repressed genes by distributing the TSSes and GREs randomly across the whole genome yields similar distribution and median distance. (146kb for ChipSeq, versus 166.6kb for random distribution model, n=250).

The top 100 most enriched regions were scanned for a GRE consensus sequence (Fig. 2.3A), which yielded practically an identical motif as found by Chip-Seq (Fig. 2.3B) (14). I then extracted sequences with 95% homology to the GRE consensus sequence from the most enriched regions (fig 3a) and obtained 405 sequences. Using custom scripts (code in appendix D) I analyzed the frequency of motifs that can be targeted with Py-Im polyamides (Fig 2.3C). The most common sequence can be targeted by a polyamide used previously in our lab (**1**, targeted to 5'-WGWWCW-3') both *in-vitro* (26) and in gene regulation studies (22, 27). The second and third most common are targeted by the same polyamide (**2**, targeted to 5'-WGGWCW-3'), a sequence that also has gene targeted in our group previously (26). The fourth sequence (**3**, targeted to 5'-WWCWGW-3') has not been yet tested. The orthogonality of binding of polyamides **1**-**3** (Fig. 4) was determined by GRE sequence analysis, based on the previous ChIP-seq experiments (14). Comparing these three polyamides in gene regulation studies will narrow our focus to the most commonly found sequences that can be bound by 8-ring hairpin polyamides. The three

compounds can bind different sites within a GREs (Fig. 4a) and some of those sites are more conserved than others – in particular the bases 7-9, show a particularly high variability. This analysis informed the decision on which compounds should be synthesized, to exert a specific control of gene expression in A549 cells. The methods developed allow to perform this analysis for other systems, cells and polyamides.

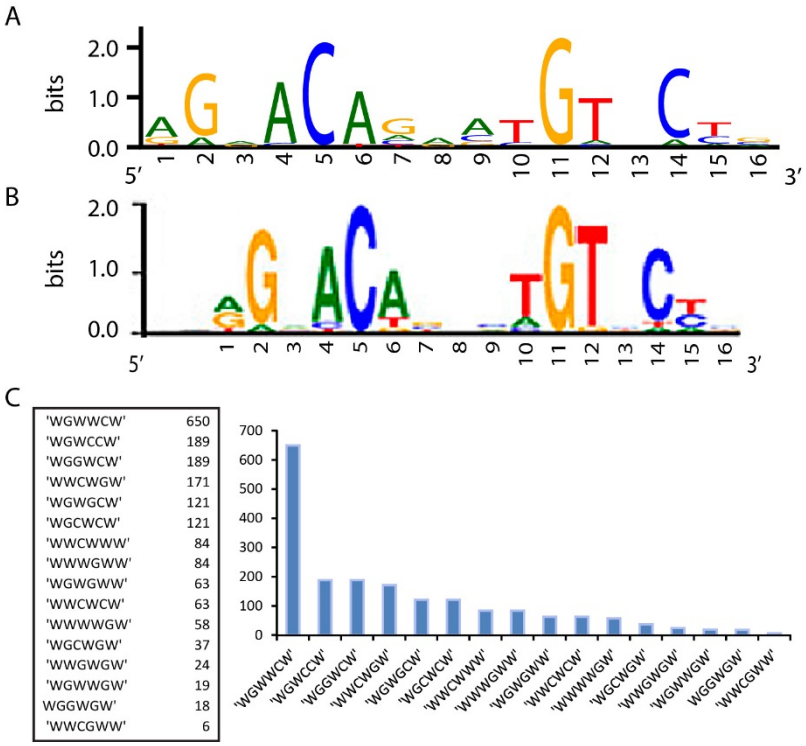
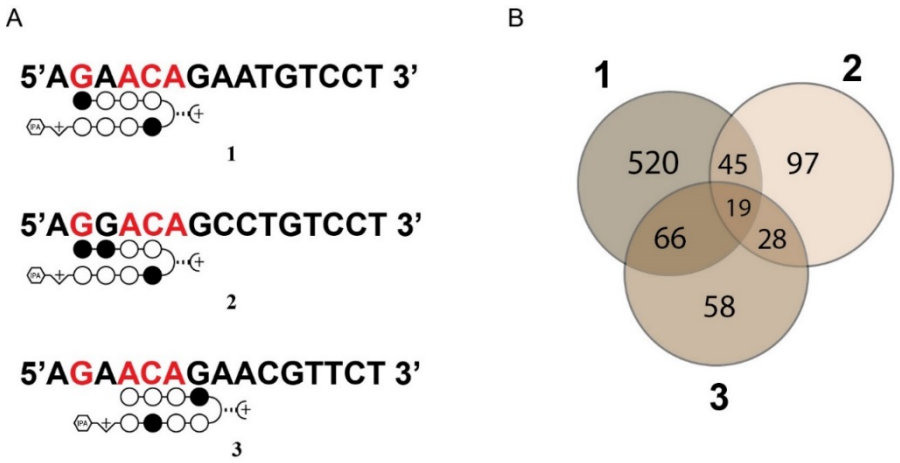


Figure 2.3 Characterizing DNA sequences binding GR. (A) Top 100 most enriched regions in Dex induced samples returned a consensus sequence that is practically identical to one obtained from uninduced cells (B). (C) The frequency of 6-basepair sequences that can be targeted by polyamides reveals WGWWCW is the most common motif among 405 GRE in a 100 regions most enriched upon Dex treatment.

Selectivity of polyamides in A549 lung adenocarcinoma cells: Gene regulation studies

The selectivity of Pyrrole-Imidazole polyamides has been tested rigorously *in vitro* (1, 28, 29); however, many questions need to be answered in the case of polyamide selectivity in cells. In order to address this issue I began gene regulation studies in A549 lung adenocarcinoma cells used previously in gene regulation studies with Py-Im polyamides (22). Since little is known

about a functional link between TSSs and GREs, I decided to investigate effects of polyamides on



C

		1	2	3	4	5
		GWWC	GGWC	WCWG	WCWW	WCWC/GWGW
A	GWWC	650	64	85	60	44
B	GGWC	64	189	37	13	24
C	WCWG	85	37	171	43	14
D	WCWW	60	13	43	84	3
E	WCWC/GWGW	44	24	14	3	63

		GWWC	GGWC	WCWG	WCWW	WCWC/GWGW
F	GWWC	1.00	0.34	0.50	0.71	0.70
G	GGWC	0.10	1.00	0.22	0.15	0.38
H	WCWG	0.13	0.20	1.00	0.51	0.22
I	WCWW	0.09	0.07	0.25	1.00	0.05
J	WCWC/GWGW	0.07	0.13	0.08	0.04	1.00
	Promiscuity	1.39	1.73	2.05	2.42	2.35

Figure 2.4 Orthogonality and targeting sites of three polyamides recognizing the most GREs according to the data in Figure 8. (A) **1** and PA2 bind nucleotides 1-6 differ in the 3rd base of the GRE. Polyamide **3** binds a different site and shows variability in 6th and 7th bases. (B) **1** targets most sequences of the three polyamides. **3** targets mostly a subset of the sequences that are also targeted by **1**, and **2** binds mostly orthogonal sequences. (C) Polyamide specificity table shows orthogonality for other polyamides used in the study. Entries on the diagonal represent absolute number of match sequences for each polyamide. For example, entry A1 shows there are 650 WGWWCW binding sites in the tested Chip-Seq regions. The numbers of the diagonal represent a subset of GREs that can bind two different polyamides. For example, entry B2 shows there are 64 GREs that can bind both WGWWCW and WGGWCW polyamides. The bottom table (entries F1-J5) summarizes the relative promiscuity of each polyamide in the study. Each column in the top table (entries A1-E5) has been normalized to the entry on the diagonal. Promiscuity coefficients have been obtained by summing every entry in the column.

expression of a panel of genes significantly induced by GR agonist Dexamethasone (Dex), thus yielding distinguishable changes in GR-driven gene expression. According to the current models of gene expression in GR system, each of these genes should be regulated by a single or small number of GREs (14). Even though I did not know their sequences, I knew the distribution of GRE sequences genome-wide. Assuming perfect sequence specificity of polyamides, we should be able to elucidate the sequences of those GREs by observing the patterns of gene expression inhibition in a randomly selected subset of genes. I began with testing Dexamethasone induced genes identified by microarray (30) and RNA-sequencing (14). A panel of 17 genes was tested using quantitative Reverse Transcription Polymerase Chain Reaction (RT-qPCR). However, four of these genes were not upregulated significantly (Fig. 5).

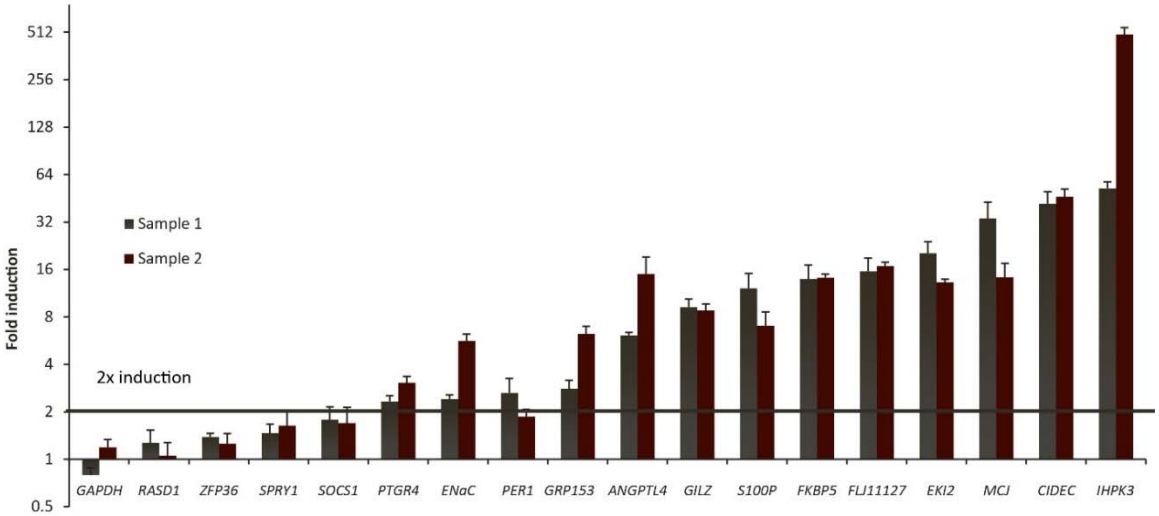


Figure 2.5 Analysis of levels of expression induced by Dexamethasone. The levels of expression were obtained by RT-qPCR. The fold induction values were obtained by dividing levels of mRNA expression obtained for Dex induced samples by uninduced ones. Genes in this panel were identified previously from microarray[7] and RNA-sequencing studies[8]. Twelve Genes that were induced at least two-fold were used for the further studies. GAPDH is a housekeeping gene and acts as a negative control. Each samples has been normalized to expression to a housekeeping gene (GUSB).

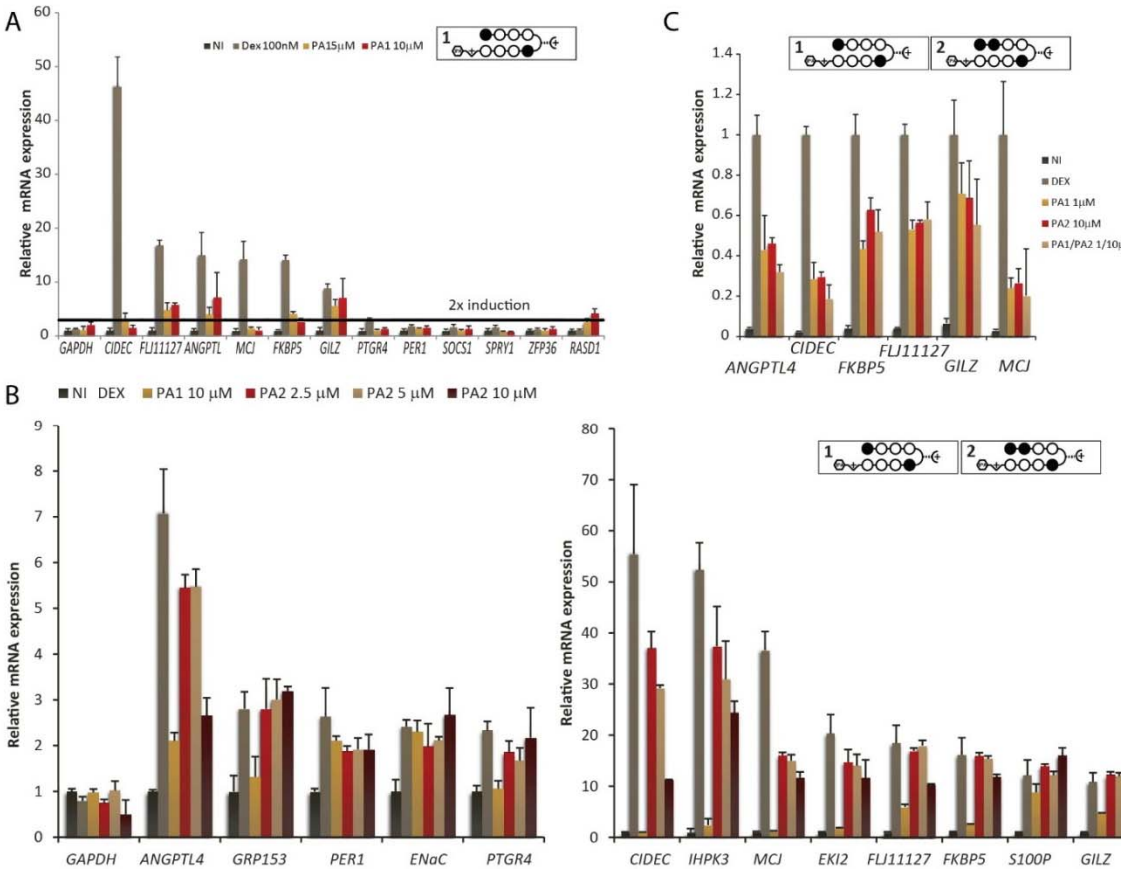


Figure 2.6 Analysis of levels of expression of genes induced by Dex and inhibited by polyamides 1 and 2. (a) Exploratory study showed that well induced genes are strongly inhibited by 1. (b) Polyamide 1 is significantly more potent than 2, but both of them downregulate the same genes. (c) Polyamide 2 is approximately 10 times less potent than 1, but the polyamides downregulated the gene expression in the same way. The correlation between the fold-induction and fold-inhibition was 0.86 for 1 and 0.74 for 2, suggesting relatively non-specific inhibition of gene expression.

I followed with the panel of 13 significantly upregulated (at least 2-fold) genes to measure the effect of polyamides on their expression. The timecourse and initial dosing was consistent with our previous studies with compound **1** in A549 cells (22). In short, the protocol included plating 12000 A549 cells/cm2 in 12 or 24 well plates for 24 hours in F12-K medium supplemented with 10% FBS, then the cells were washed twice with 1x PBS and the medium replaced with F-12K medium supplemented with 10% Charcoal Treated (CT) FBS, including the desired concentration of polyamides. After 48 hours of incubation, 100 nM dexmethasone was added directly to the medium for 6 hours, after which cells were harvested for RNA extraction. In

order to test this dosing on various genes I run a limited test on 11 genes, both upregulated and unaffected, to see if polyamides downregulated either of them (Fig. 2.6A). Confirming that all uninduced genes, with an exception of SPRY1, were unaffected by polyamide **1**, while all well induced genes were inhibited by it, I decided to compare the gene regulation capability of polyamides **1** and **2** (Fig. 2.6B).

Polyamide **1** (targeting 5' WGWWCW 3') downregulated the expression of all Dex induced genes, except two: S100P and ENaC. One of the genes, PER1, was downregulated slightly (approximately 30%) and 6 other genes, FLJ11127, FKBP5, GILZ, ANGPTL4, IHPK3, and PTGR4, were downregulated at least two-fold. Polyamide **1** at 10 μM had also completely abolished the effects of Dex in the remaining four genes, CIDEC, GRP153, MCJ and EKI2. This widespread action of polyamide 1 was not unexpected, given that the motif it targets is very common (Fig 2.3C), for the same reason one would expect polyamide **2** to target only a subset of genes, or set of genes that is different from the one downregulated by **1**. However, the dose response of polyamide **2** (Fig. 2.6B) suggests that polyamide **2** targets the same sequences, although with less potency. The same genes whose expression was most downregulated by polyamide 1, are also downregulated by polyamide **2**, albeit to a lesser extent. By running a series of exploratory experiments I was able to determine that potency of polyamide **2** is comparable to 10-fold lower concentration of polyamide **1** added to the cell media. Dosing cells with 1 μM polyamide **1** and 10 μM polyamide **2** yielded identical reponses (Fig. 2.6C), for the panel of 6 genes that were significantly downregulated in the previous experiment. It is not possible to tell whether this widespread response of both genes is due to non-specificity of polyamides in the cell nucleus, or because each of the genes inhibited happend to be regulated by several GREs containing both 5'WGWWCW3' and 5'WGGWCW3' motifs. However, when both polyamides were dosed at the same time, there was no synergistic effect (Fig. 2.6C) suggesting non-specific polyamide binding as a culprit. The extent of polyamide-mediated gene expression downregulation was correlated with the fold-induction with 100 nM dexamethasone (0.86 for **1** and 0.74 for **2**). This suggests that the most induced genes are the ones most affected by polyamides, possibly regardless of their sequence. No rigorous test exists as of now to determine the reasons for this high correlation.

Next, I tested the specificity of polyamides targeting 7th basepair in GRE motif. At first, I decided to measure the effects of polyamide **3** on the gene expression because the sequence it targets is approximately as common as for **2**. Despite following the same treatment as for compounds **1** and **2**, only two genes, ANGPTL4 and CIDEC, were affected by **3** (Fig. 2.7A) at low concentrations 2.5 μM or 5 μM. Bringing up the concentrations of these two genes showed further inhibition: ANGPTL4 was inhibited by over 50% and CIDEC by over 40%. without affecting four other highly induced genes: FKBP5, FLJ11127, GILZ, and MCJ (Fig 2.7B). Further increase in polyamide 3 concentration did not increase polyamide potency significantly, potentially because of instrument noise or polyamide solubility problems (Fig. 2.7C). Thus 10 μM concentration is either the most effective, or nearly the most effective in gene downregulation. At this concentration, the correlation between fold-induction and fold inhibition was low at 0.25, suggesting that polyamide **3** targets genes more independently of their induced activity than **1** or **2**. In an effort to improve potency of 3, I followed with its acetylation (31). Previous experience in the group suggested that this modification can improve gene downregulation (32). However, in the case of acetylated polyamide 3 (12) the gene downregulation profile was identical in both selectivity and potency.

The specificity of polyamide 3 suggested synthesis of other compounds targeting the 7th base in the GRE with N-methylpyrrole at the last position (cap). I expanded the library of compounds to target these sequences. The next sequences most commonly found in GREs can be targeted by polyamides **4** and **5** (Fig. 2.8), namely, sequences 5'WWCWW3' and 5'WWCWCW3', respectively. These compounds as well proved to be less potent than **1** or **2** but also more selective in gene downregulation. Compound **4** caused downregulation of three genes at two-fold or more: ANGPTL4, CIDEC, and MCJ (Fig. 2.9). Two of these genes, ANGPTL4 and CIDEC, were also downregulated by **3**; however, MCJ was not. Thus there is a distinguishable difference in downregulation profile between 3 and 4, even though both of them target the GRE consensus sequence. It is possible that this effect is due to DNA-binding independent events, and thus it is not yet clear if their differences in gene downregulation are due to differences in sequences of GREs influence on the gene downregulation patterns. Polyamide 6 is a potent compound with little controlling each of these three genes. It is, however, a useful feature of