Chapter 1

Introduction

1.1 A central role for gene regulation

The sequencing of the human genome revealed a huge cache of information: three billion base pairs of genetic information encoded on 23 chromosomes. Despite the large relative size of the human genome, it quickly became apparent that it encodes an astonishingly small number of genes (1). Humans are now estimated to have about 23,000 protein-coding genes, a figure on the same order as the number estimated to belong to the roundworm (*Caenorhabditis elegans*). The homology of human genes to those from other animals is also high; nearly 40% of fruit fly (*Drosophila melanogaster*) genes have human homologs (figure 1.1) (2). When the homology search is restricted to known human disease genes, this percentage rises to 75% (3). Yet in humans, these protein-coding regions make up less than 2% of the sequence data. The rest is made up of non-coding DNA corresponding to regulatory elements, functional RNAs, introns, and other structural or functional units. These sequences were once described as "junk DNA,"a form of genetic debris accumulated during the evolution of humans from lower organisms and dismissed by Francis Crick as having "little specificity and convey[ing] little or no selective advantage to the organism" (4). Comparative genomics analyses have shown the opposite to be true: some segments of noncoding DNA are conserved through long evolutionary time periods, providing *prima facie* evidence of their function and selective advantage (5). Given that animals, especially mammals, share the same small set of genes, it would appear that complexity arises more from the regulation of those genes than from their sequence identity.

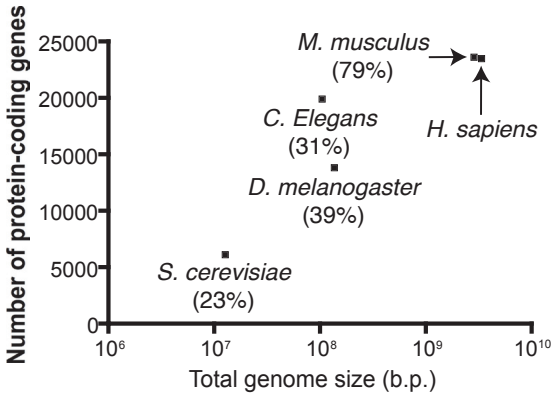

Figure 1.1. Chart depicting the genome size and number of protein coding genes of several eukaryotes. Numbers in parentheses indicate the percentage of that organism's protein-coding genes that have human homologs. The genome size increases three orders of magnitude from yeast to man but the total number of genes increases only 5-fold and many of those genes are conserved.

1.2 Gene regulation by transcription factors

Since the vast array of diverse animal life shares so many common genes, it is perhaps unsurprising that gene regulation should be a complex endeavor. An animal's

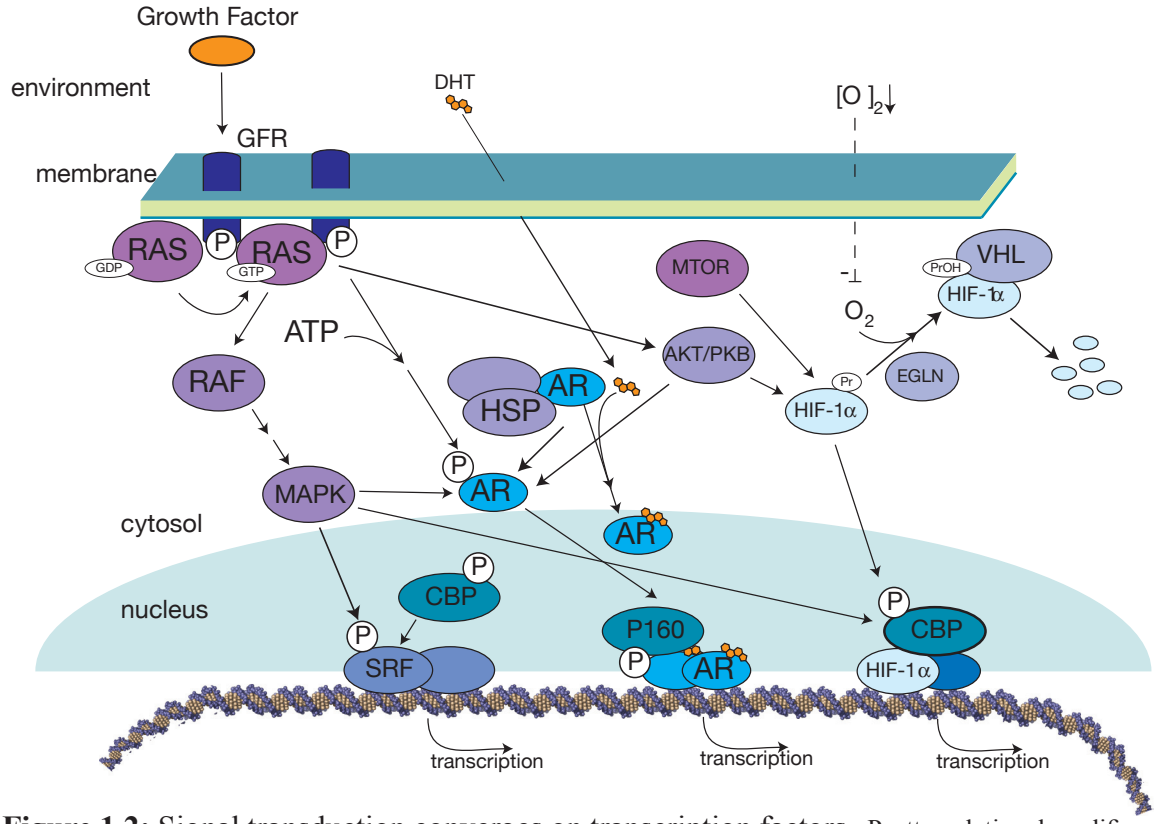

Figure 1.2: Signal transduction converges on transcription factors. Posttranslational modifications and nuclear trafficking regulate diverse, interconnected signaling pathways. Growth factor signaling is mediated by the activation of receptor tyrosine kinases located on the cell membrane. Growth factor binding triggers a phosphorylation cascade, ultimately activating the SRF transcription factor to promote the transcription of cell cycle related genes. Androgen receptor (AR) is activated directly by binding a membrane-permeable ligand, dihydrotestosterone (DHT), which frees it from the heat shock proteins (HSPs) that keep it out of the nucleus in monomeric form. Liberation leads to dimerization, nuclear trafficking, and regulation of AR target gene expression. The hypoxia inducible factor, HIF1α, is also modulated by a membrane permeable molecule, molecular oxygen. Under normoxic conditions, HIF1α is hydroxylated, ubiquitinated, and rapidly degraded. Loss of O_2 rescues the transcription factor from degradation, and the receptor localizes to the nucleus to carry out its transcriptional program. Note that phosphorylation from the growth factor pathway cross-activates the transcriptional activation of the other two.

Adapted from N.G. Nickols (2008) "Endogenous gene regulation by DNA binding polyamides," p. 3. (Ph.D.) Dissertation, California Institute of Technology. Used with permission.

proper development and homeostasis depend critically on having the right gene product available at the right time. Cells can respond to a wide array of environmental and

autochthonous stimuli through changes in gene regulation at any level: from chromatin silencing to transcription initiation to posttranslational modification and trafficking. Many of the fastest regulatory responses are conducted through vast networks of proteins that form intertwined signaling cascades that rapidly transmit and amplify signals by using changes in phosphorylation or ubiquitination state, to name just two (figure 1.2). Often, these fast responses converge on proteins that integrate all of the upstream signaling to produce changes in gene expression. These proteins, called transcription factors, are ultimately responsible for the interaction of genes with the environment.

Transcription factors (TF) are defined as sequence-specific DNA-binding proteins. There are approximately 2600 in the human genome (6). Their general structure consists of a DNA-binding domain (DBD) and a transactivation or transrepressor domain (TAD) (7). Each DBD has a consensus DNA binding sequence: a short, loosely defined set of DNA sequence preferences. There are a variety of structural solutions for achieving sequence-specific binding; a few examples of these motifs include the basic helix-loop-helix (bHLH), the zinc finger, and High Mobility Group (HMG) box (figure 1.3) (8-10). Each motif produces a characteristic change in

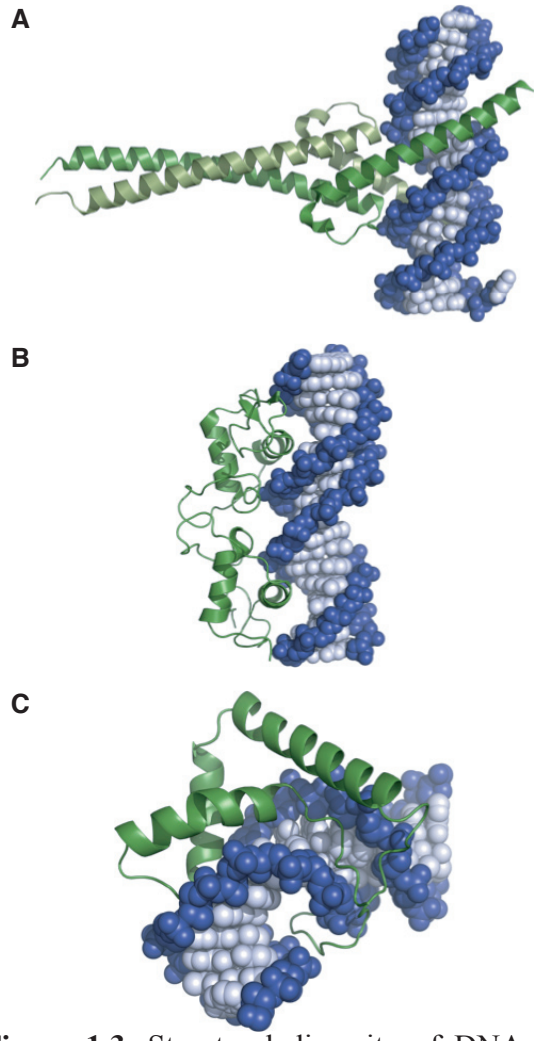

Figure 1.3. Structural diversity of DNA-binding proteins. This figure depicts three different DNA-binding proteins that use different protein folding motifs to achieve sequence-specific recognition of their cognate DNA-binding elements. (A) bHLH proteins Myc and Max heterodimerize to bind in the major groove (PDB 1NKB). (B) A homodimer of truncated androgen receptor (AR) subunits bound to DNA in the major groove. Each subunit contains two zinc finger domains (PDB 1R4I). (C) Minor groove-binding Lef-1 contains an HMG box DNA-binding motif. Protein-binding causes a helical bend (PDB 2LEF).

the local structure of its DNA host. They range from relatively small perturbations, like widening of the major groove that occurs upon binding of the zinc finger protein, androgen receptor (AR), to the near 90° helical bends produced by the HMG proteins Lef-1 and Sox2 (11, 12).

Genomic sites that match the consensus and allow TF binding are known as response elements and are found in promoter or enhancer regions within close geometric proximity to the transcription start site of a target gene. The TAD allows TFs to interact with other proteins to recruit coactivators or corepressors to the target locus (13). Coactivators promote histone acetylation and recruit mediator proteins, which form the scaffold for the assembly of the RNA Pol II holoenzyme (14). Corepressors also promote nucleosome remodeling, but instead act as histone deactylases, which decreases the accessibility of the chromatin architecture to proteins.

The activation of transcription factors can occur through a variety of means. Those TFs that contain a third protein domain, called a signal-sensing domain, can directly respond to cell permeable stimuli through allosteric modulation of protein conformation (e.g., AR) (figure 1.2) (15). Other TFs are rescued from degradation (e.g., HIF) or phosphorylated (e.g., SRF) to increase their transactivation activity (16, 17). There is significant cross-talk between the upstream signaling pathways, such that a single stimulus can modulate the activity of a set of transcription factors, leading to a nuanced transcriptional response that may involve hundreds of genes. Activation of transcription factors allows transcriptional programs to be turned on only in response to the appropriate stimulus.

In addition, TFs often bind cooperatively, acting in concert with several other transcription factors to modulate the DNA topology. This combinatorial binding allows for plenty of diversity in gene expression; the 2600 transcription factors can combine to establish unique control of all 20,000 human genes (18). Combinatorial binding of transcription factors is assisted by the allosteric modulation of the DNA surface that occurs when these proteins bind DNA. The allosteric changes introduced by the binding

of a single transcription factor are transmitted along the DNA double helix and produce a favorable topography for its binding partners (Figure 1.4). In fact, the current model of transcription factor assembly on gene regulatory elements postulates that this allostery, not protein-protein interactions, is the dominant factor driving the highly cooperative binding

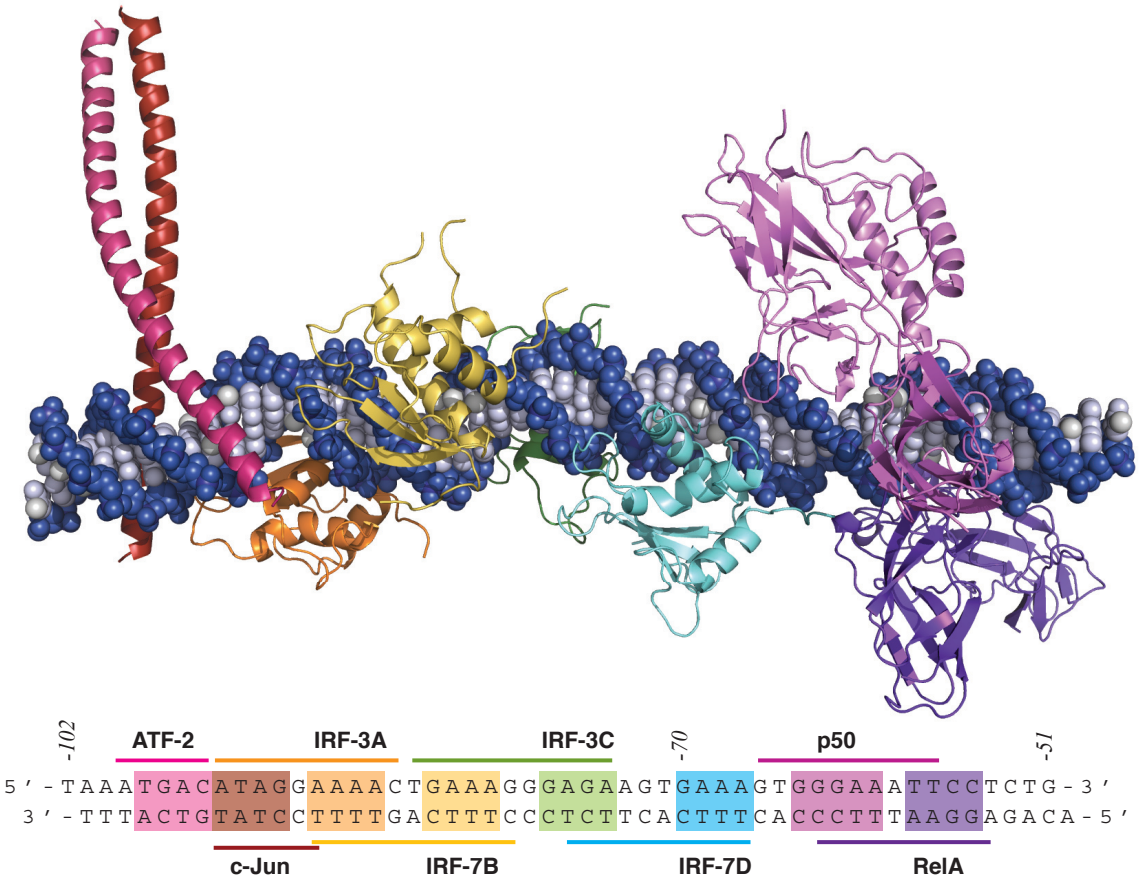

Figure 1.4. Composite model of cooperative assembly of transcription factors mediated by allosteric interactions on the Interferon-β enhancer. DNA topological changes resulting from transcription factor binding enhance binding of other proteins to the same locus. Protein-protein contacts are not observed in either of the crystal structures used to make this composite.

Figure adapted by K.A. Muzikar from D. Panne et al., 2007 (19) using PDB 2O6G and 2O61 and used with her permission (K.A. Muzikar. (2011) "Repression of DNA-binding-dependent glucocorticoid receptor-mediated gene expression," p. 33. PhD Dissertation, California Institute of Technology.)

of multiple proteins to adjacent DNA sequences (19).

1.3 The androgen receptor in health and disease

Transcription factor activation can be a powerful phenotypic determinant. Consider the androgen receptor (AR): signaling through this single protein controls the development and maintenance of male sex characteristics (20). A member of the nuclear hormone receptor class of TFs, AR is composed of two N-terminal regulatory domains, a two-zinc finger DNA-binding domain, a hinge region, a ligand-binding domain, and a C-terminal interaction domain (21-25). In the absence of ligand, AR is sequestered in the cytosol by heat shock proteins (HSPs) (26). Steroid binding causes an allosteric conformational shift that liberates the TF from the HSPs, followed by homodimerization, nuclear localization, response element binding, and modulation of target gene transcription (Figure 1.2). In humans, these target genes control the development of primary and secondary sex characteristics, which means AR affects development of the genitalia as well as contributing to height, bone and muscle mass, and hair growth.

AR also plays a central role in the biology of the prostate and its neoplastic derivative, prostate cancer (27). In the normal prostate, AR signaling maintains the prostatic epithelium without producing uncontrolled growth. Castration results in rapid involution of the prostate, largely by apoptosis (28). When cancer emerges, growth inhibition is lost and AR signaling begins to drive the growth and spread of the disease. The molecular biology of prostate cancer oncogenesis is still an active field of research, and the set of AR-responsive genes responsible for the transformation has yet to be fully determined. In many tumors, a chromosomal translocation fusing the ERG oncogene with the AR-responsive TMPRSS2 gene appears to contribute to the uncontrolled growth of these cells (29). Like the normal prostate, prostate cancer will also respond dramatically to surgical or chemical castration, leading to clinical remission in some cases. The average time to relapse is approximately 6 months, and the current standard of care for recurrent disease relies on cytotoxic chemotherapy (30).

The recurrent form of prostate cancer is often called hormone-refractory, as first-

and second-generation antiandrogens like bicalutamide have failed. Yet this may be somewhat of a misnomer, as recent studies have shown that even in its recurrent form the tumor often relies on AR signaling for its growth: 'castration-resistant' may be more appropriate. In mouse models, a simple 5-fold upregulation of AR expression was sufficient to confer resistance to some anti-hormone agents (Figure 1.5) (31). In addition, the overexpression of AR in the human prostate cancer cell line, LNCaP, hypersensitized the pathway, rendering it promiscuous towards other steroid ligands (estrogen, dexamethasone) that would not activate the receptor under normal circumstances. Using these AR-expressing, castration-resistant cells as a model system, researchers have developed a third-generation anti-androgen (MDV3100) that maintains its potency in recurrent prostate cancer (32). This drug has shown promise in phase I/II clinical trials, demonstrating once again the importance of transcription factors in human biology and medicine (33).

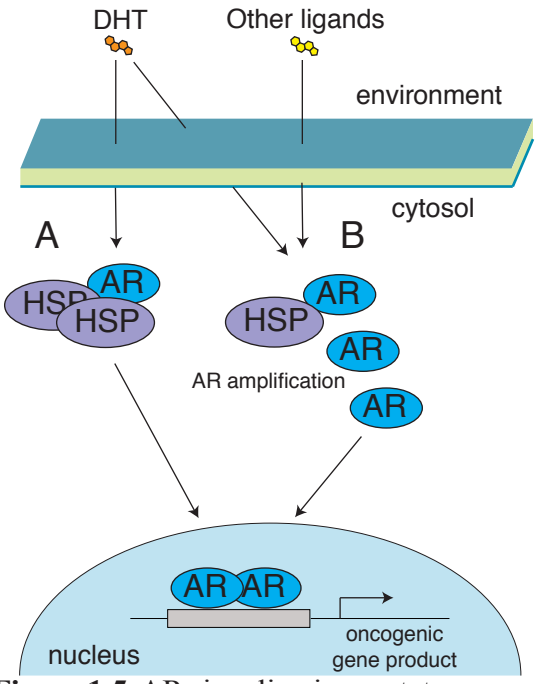

Figure 1.5. AR signaling in prostate cancer. (A) Prostate cancer arises when growth inhibition of AR-mediated mitogenic signaling is lost. Treatment-naive disease responds to therapies directed at reducing androgen production (e.g., dihydrotestosterone, DHT). (B) Upregulation of AR expression is a common mechanism of castration resistance. Increased receptor expression sensitizes the AR pathway to low levels of residual androgens as well as other steroid ligands (e.g., glucocorticoids). Growth of castration-resistant tumors is still dependent on AR signaling, so high affinity antagonists (3rd generation anti-androgens) like MDV3100 retain their activity.

1.4 Transcription factors as drug targets: beyond nuclear hormone receptors

In a widely cited review published in 2002, J. E. Darnell made a strong case for targeting transcription factors in cancer therapy (34). Instead of targeting the large number of oncogenic gene products implicated in pathogenesis, he advocated targeting

the relatively few dysregulated, overactive transcription factors common to many tumors. The transcription factors STAT3, NFkB, and β-catenin are noted as prominent examples. But unlike the nuclear hormone receptors, signaling cascades, not small molecules, activate these proteins. This makes them difficult to approach using traditional pharmacologic methods that target enzyme-substrate or receptor-ligand interactions. Darnell challenged chemists to respond with compounds that inhibit the two types of binding most important to transcription factor inhibition: protein-protein and protein-DNA interactions.

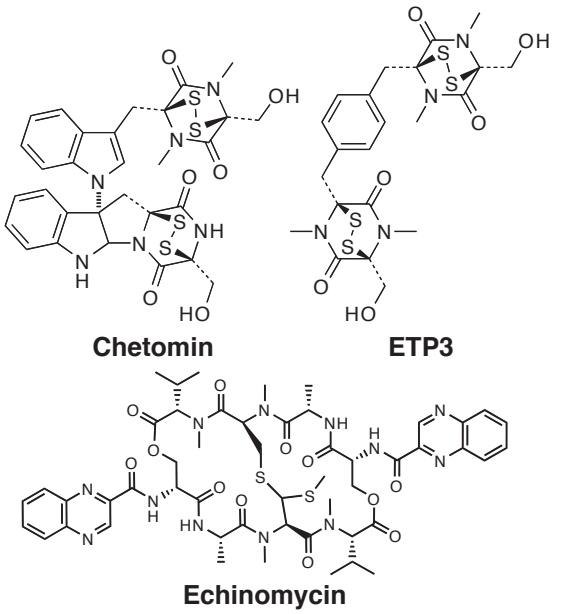

Figure 1.6. Two different strategies for the inhibition of the transcription factor HIF1α. Chetomin and ETP3 inhibit the association of the transcription factor with its co-regulatory partner, p300. The intercalator echinomycin prevents HIF1α binding to its DNA response element.

In the years since 2002, a number of small-molecule transcription factor inhibitors have been identified (35). As an example, the transcription factor HIF1α has been successfully targeted *in vitro* and in mouse models. One of the more successful HIF1α protein-protein interaction inhibitors is ETP3, a dimeric compound based on the natural product chetomin that disrupts the protein-protein interaction of HIF1α with its binding partner, p300, and has submicromolar IC_{50} against HIF1α promoter activity (Figure 1.6) (36, 37). Another natural product, echinomycin, was identified in a high-throughput screen and shown to inhibit the protein-DNA interactions of the transcription factor at nanomolar concentrations (38). Yet each of these solutions is idiosyncratic: the first relied on intimate structural knowledge of the transcription factor, and the second required well-established knowledge of the protein's function as a prerequisite for designing the screen. Nearly all of the small molecule transcription factor antagonists developed over the last few years

have the same story; each solution is unique, and it is difficult to apply the lessons learned from one system to the next target. A general chemical solution to this problem would be invaluable to human medicine.

Most of the general solutions to transcription factor inhibition involve the use of large, biomolecular constructs. There are both nucleotide- and protein-based approaches. RNAi has been used extensively as a research tool for transcription factor inhibition by protein depletion (39). The specificity of this technique is exquisite, and *in vitro*, very effective. But to reach the clinic, RNAi approaches must solve a difficult delivery problem: how to achieve systemic distribution of large, polyanionic molecules in complex animals. The solution has remained elusive for years. Other nucleotide-based approaches, like antisense and decoy oligonucleotides, have similar delivery issues.

The first peptidic approach to transcription factor inhibition employed engineered zinc finger proteins to compete with an endogenous transcription factor for its binding site (40). A zinc finger motif consists of an α-helix and an antiparallel β-sheet that together coordinate a single zinc ion and specify a 3-base pair DNA sequence (Figure 1.7) (41). Arrays of multiple zinc fingers can be constructed in a modular fashion or selected using directed evolution to target all

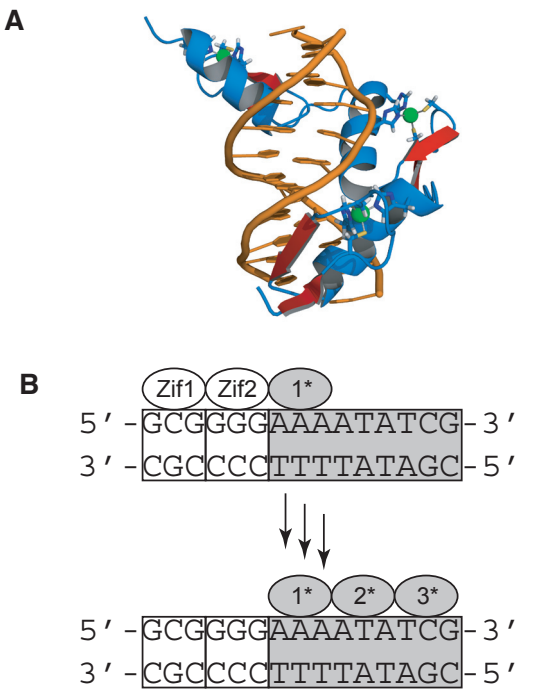

Figure 1.7. Engineered zinc finger proteins for control of transcription. (A) Crystal structure of zinc finger protein Zif268 containing three zinc finger motifs. Coordinated zinc ions are shown as green spheres. Image created by T. Splettstoesser from PDB 1A1L (41) and used with permission under Creative Commons license. (B) Schematic diagram depicting the iterative selection of zinc finger motifs to bind a 9-base pair DNA sequence. Two zinc finger motifs (Zif1 & Zif2) of known sequence selectivity are used to position a third (1*) whose binding to the target 5'-AAA-3' sequence is optimized by affinity-based selection. The optimized 1* motif is then paired with Zif2 and a new protein sequence (2*) is optimized to bind 5'-ATA-3'. A third iteration with optimized 1* & 2* is used to generate the full length protein to recognize the 5'-AAAATATCG-3' target sequence.

possible DNA sequences up to 18 base pairs in length (42, 43). In a similar fashion to RNAi, this technology can achieve high specificity, as some 18-base pair sequences are unique in the human genome (44). Unfortunately, like RNAi, engineered zinc fingers also have significant delivery problems and rely on lentiviral gene delivery vectors for use in animals (45).

More recently, a peptide-based approach has been used to target protein-protein interactions of transcription factors. This strategy employs stabilized, 'stapled' α-helical oligopeptide fragments of dominant negative transcription factor binding partners (46). These peptides are apparently small enough to enter cells through endocytic mechanisms and achieve nuclear localization without the use of delivery agents. As a recent example, a stapled helix targeted to the NOTCH transcription factor disrupted the pathway *in vitro* and in mouse models following systemic administration (47). This technology has promise, but it is not as mature as RNAi. Little biodistribution data are available.

1.5 DNA-binding Py-Im polyamides for the control of transcription

We have chosen to pursue the development of a general chemical methodology for the inhibition of transcription factor signaling. We have targeted the DNA side of the protein-DNA interface, in part to take advantage of the relative simplicity of the structure of the DNA double helix. For this purpose, we employ sequence-specific DNA-binding Py-Im polyamides as part of a general strategy for developing transcription factor inhibitors. Candidate compounds must meet several stringent criteria. In order to be general, the compounds must be able to bind a wide array of DNA sequences in a modular fashion with high affinity and specificity. In order to be effective in biological systems, they must be cell permeable, traffic to the nucleus, and bind genomic DNA in its native chromatin conformation. Having arrived at their target site, they must inhibit the transcription factor, either through direct competition for the binding site surface, or by allosteric modulation of the local DNA topography. Finally, to be useful as therapeutics, candidate compounds

would ideally obtain systemic distribution in animals without the need for delivery agents. The following section details how the Py-Im polyamide has been engineered to meet all of these requirements.

1.6 Molecular recognition of the DNA minor groove

B-form DNA is an asymmetric double helix composed of two antiparallel strands of nucleotides (48). The asymmetry causes the helix to form a wide and shallow major groove as well as a narrow and deep minor groove (Figure 1.8). Small molecules that bind DNA can do so in either groove; they can also intercalate between the base pairs or bind the phosphate backbone. Sequence-specific small molecules interact with the particular hydrogen-bonding and steric patterns that accompany the DNA base pairs (49). Each base pair presents a unique surface of hydrogen bond donors and acceptors that can be recognized by a complimentary surface on a small molecule (Figure 1.9). Examples of some sequence-specific DNA-binding molecules are shown in Figure 1.10.

The natural product distamycin A is the original scaffold from which Py-Im polyamides developed (50). Distamycin A

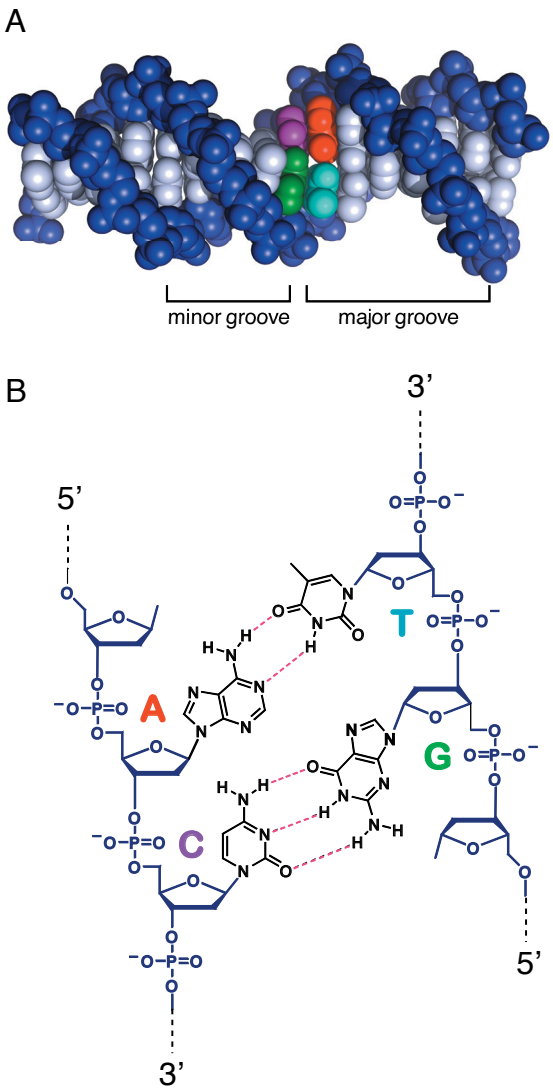

Figure 1.8. Structure of B-form DNA. (A) Crystal structure of B-form DNA double helix showing the location of the narrow minor groove and wide major groove. The phosphate backbone and ribose residues are colored dark blue. The colored nucleobases represent adenine (magenta), thymine (teal), guanine (green), and cytosine (purple). (B) Chemical structure of the Watson-Crick base pairs showing the hydrogen bonding pattern. Figure adapted from K.A. Muzikar (2011).

binds to the DNA minor groove with 1:1 or 2:1 stoichiometry. The compound consists of three *N*-methylpyrrole (Py) rings linked by amide bonds that hydrogen bond with the N3 of purines or the O2 of pyrimidines. The 2:1 antiparallel binding mode prefers A-T tracts due to steric constraints conferred by the exocyclic amine of guanine in the minor groove (Figure 1.10) (51, 52). Substituting *N*-methylimidazole (Im) for one of the *N*-methylpyrroles relieves this steric hindrance and produces a hydrogen bond acceptor for the exocyclic amine of guanine. In the 2:1 antiparallel configuration, this compound pairs Im across from Py to specify 5'-WGWCW-3'; the parent compound binds 5'-WWWWW -3' (W = A or T) (53, 54). This simple substitution marked the beginning of systematic DNA base pair recognition in the minor groove by small molecules.

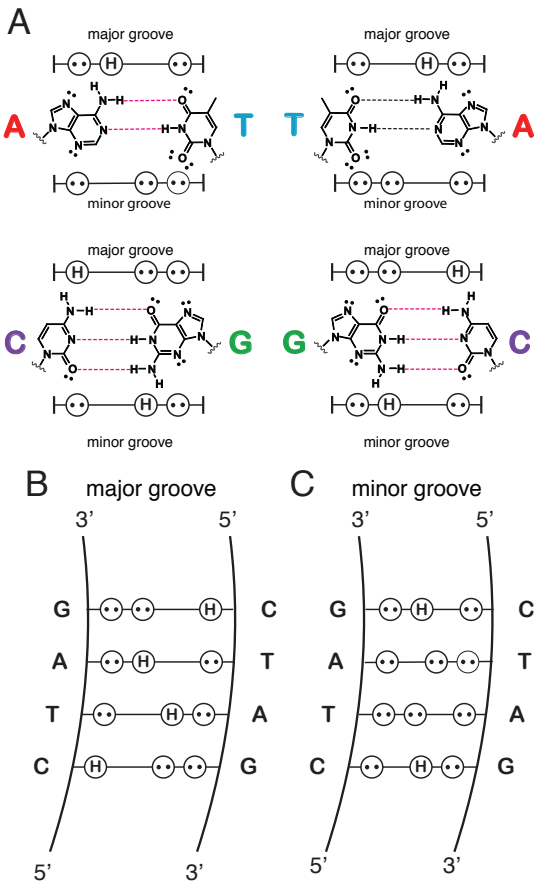

Figure 1.9. Hydrogen-bonding pattern of the four Watson-Crick base pairs in the major and minor groove. (A) The hydrogen-bonding patterns of each of the base pairs along the edges of the nucleobase is unique. Formal representation of the hydrogen bonding surface available for sequence-specific interactions with DNA-binding molecules in the major (B) and minor (C) grooves. Figure adapted from K.A. Muzikar (2011).

Py-Im polyamides, although originally based on the distamycin structure, have been extensively optimized over many years for DNA binding affinity and sequence specificity. The Im/Py pairing has been shown to be a general solution for specifying for G-C base-pair, and the reverse, Py/Im, specifies for C-G (55). A Py/Py pair is degenerate for A or T (W). Each ring pair specifies for a DNA base pair; polyamides specifying DNA sequences of 16 base pairs in length have been characterized *in vitro* (56). The composition of the individual amino acid units has also been extended beyond *N*-methylpyrrole and *N*-

methylimidazole to include other heterocyclic units (57). The most successful of these was the development of the hydroxypyrrole (Hp) monomer, such that Hp/Py specifies T-A and Py/Hp specifies A-T (Figure 1.11) (58). Thus polyamides can distinguish all four Watson-Crick base pairs through interactions with the DNA minor groove.

The configuration of the heterocycle ring pairs has also been optimized by the Dervan group. Development of a covalent linkage between the two antiparallel strands of pyrroles and imidazoles led to a 100 to 3600-fold increase in binding affinity as well as the ability to target non-palindromic sequences (59). This led to the 'hairpin' configuration of antiparallel strands linked at one end that has been most well studied in biological systems. The addition of a chiral amino group on this hairpin turn was found to reinforce binding orientation, further increasing their specificity and affinity (60). The combination of the hairpin configuration, chiral turn, and Py/Im pairing rules have produced polyamides with nanomolar affinities to rival those of endogenous transcription factors and 10 to 100 fold specificity for a single base pair mismatch.

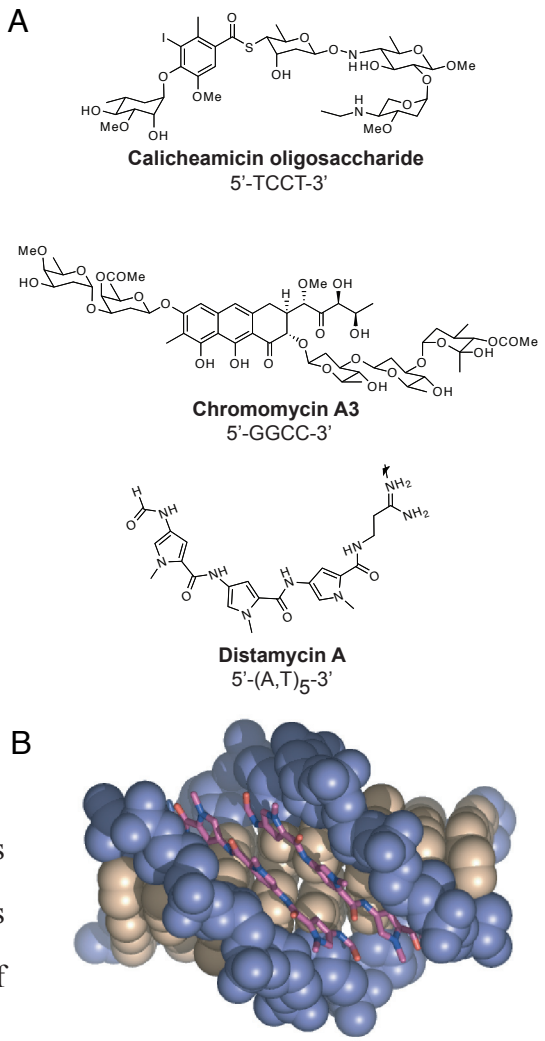

Figure 1.10. Sequence-specific, minor-groove binding natural products and their target sequences. (A) Calicheamicin binds as as a monomer, and chromomycin binds as a dimer. Distamycin A can bind in a 1:1 and a 2:1 conformation. (B) Crystal structure of Distamycin A bound to the DNA sequence 5'-GTATATAC-3' in a 2:1 conformation . (PDB 378D)

The DNA-binding capabilities of Py-Im polyamides are not restricted to the canonical B-form helix. They have been engineered to bind alternative DNA structures, including

the interwoven strands of the DNA double-crossover array for patterning nanostructures

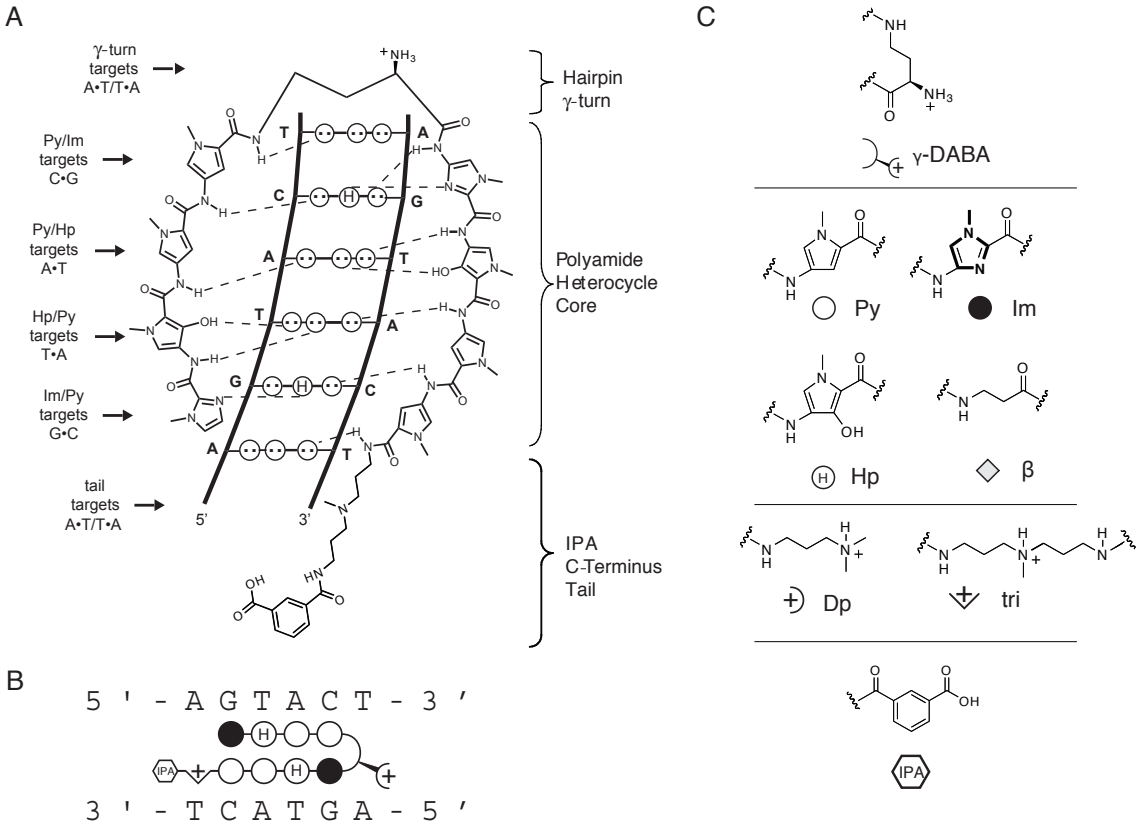

Figure 1.11. Schematic of an 8-ring hairpin polyamide designed to distinguish all four Watson-Crick base pairs. (A) Interaction of the polyamide with the hydrogen-bonding pattern in the minor groove of a 5'-AGTACT-3' DNA sequence. Each base pair is interpreted by a pair of heterocyclic rings positioned across from eachother in antiparallel fashion by the γ-turn. The turn and tail have a strong preference for A•T or T•A, so the formal sequence targeted by the polyamide is 5'-WGTACW-3' (W = A or T). (B) Ball-and-stick representation of the polyamide. Filled circles represent imidazoles, open circles represent pyrroles, and 'H' represents hydroxypyrrole. (C) Key to ball-and-stick representation of hairpin polyamides.

(61). Biologically relevent DNA conformations have also been targeted in a sequence-specific fashion, including DNA in its histone-bound form as part of the nucleosome core particle (62).

1.7 Cell permeability

In order to act as transcriptional regulators in biological systems, polyamides must be cell permeable, ideally without using any delivery agents. Cell permeability of Py-Im polyamides has been directly observed in live cells by using fluorophore conjugates of these

compounds. Over one hundred polyamide-fluorophore conjugates have been studied in over a dozen cell lines, revealing them to be broadly, though not universally cell permeable (63, 64). No delivery agents were used, and the cells were not fixed prior to examination by confocal microscopy (Figure 1.12). Pyrrole-imidazole sequence content, dye choice and position, dye linker composition, and overall molecular weight all contribute to the biocompatibility of these compounds. In addition, some cell lines are more promiscuous toward polyamide uptake than others. Like the stapled α-helices discussed above, polyamide uptake appears to be energy dependent and likely to be endocytotic, but this has not been established as a general property of the compounds as a class. We infer cell permeability

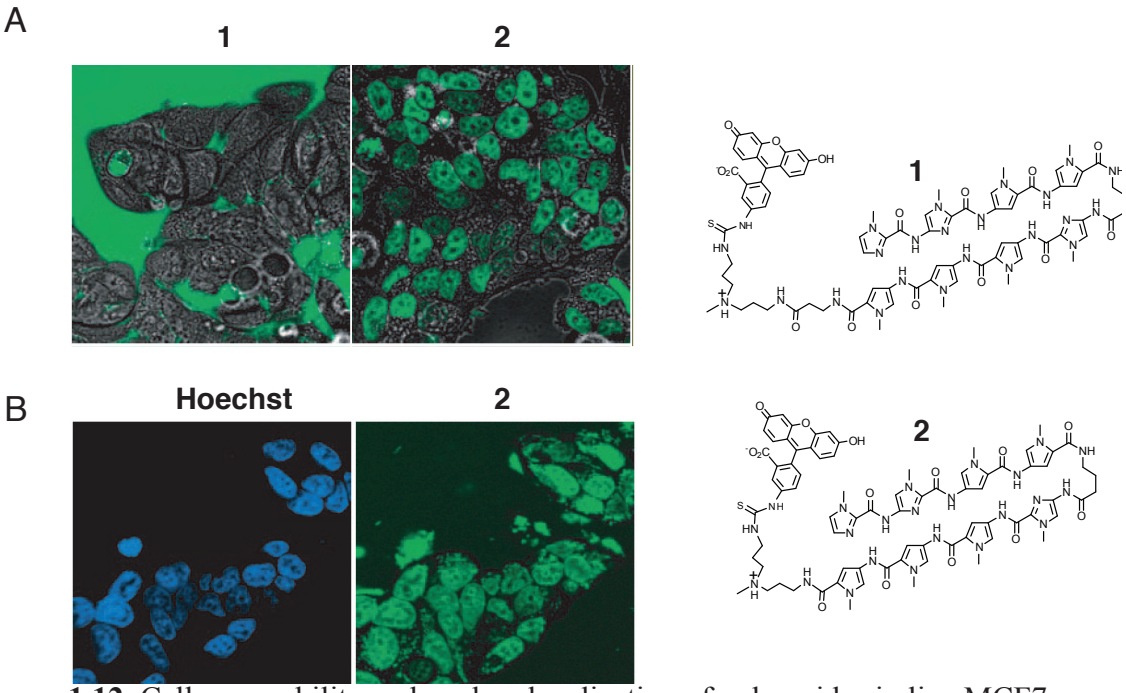

Figure 1.12. Cell permeability and nuclear localization of polyamides in live MCF7 breast cancer cells. (A) Laser confocal photomicrograph of fluorescein-conjugated polyamides **1** and **2** after 10-14 h. treatment. Compound **1** is excluded from the cells; compound **2** is cell permeable. (B) Laser confocal photomicrograph of compound **2**, showing colocalization with Hoechst, a known nuclear staining agent.

of the unlabeled parent compounds based on the permeability of their fluorophore-labeled derivatives as well as their observed effects on transcription (see below).

1.3 Gene regulation by polyamides in living systems

In vitro assays first established that polyamides could act as inhibitors of transcription factor binding. They have been shown to inhibit the DNA binding of both major and minor groove-binding transcription factors. The inhibition of major groove-binding transcription factors is presumably due to an allosteric mechanism, as polyamides bind only in the minor groove. A series of recent crystal structures of polyamide-bound DNA and the unliganded oligonucleotide at atomic resolution has elucidated the mechanism of allosteric inhibition

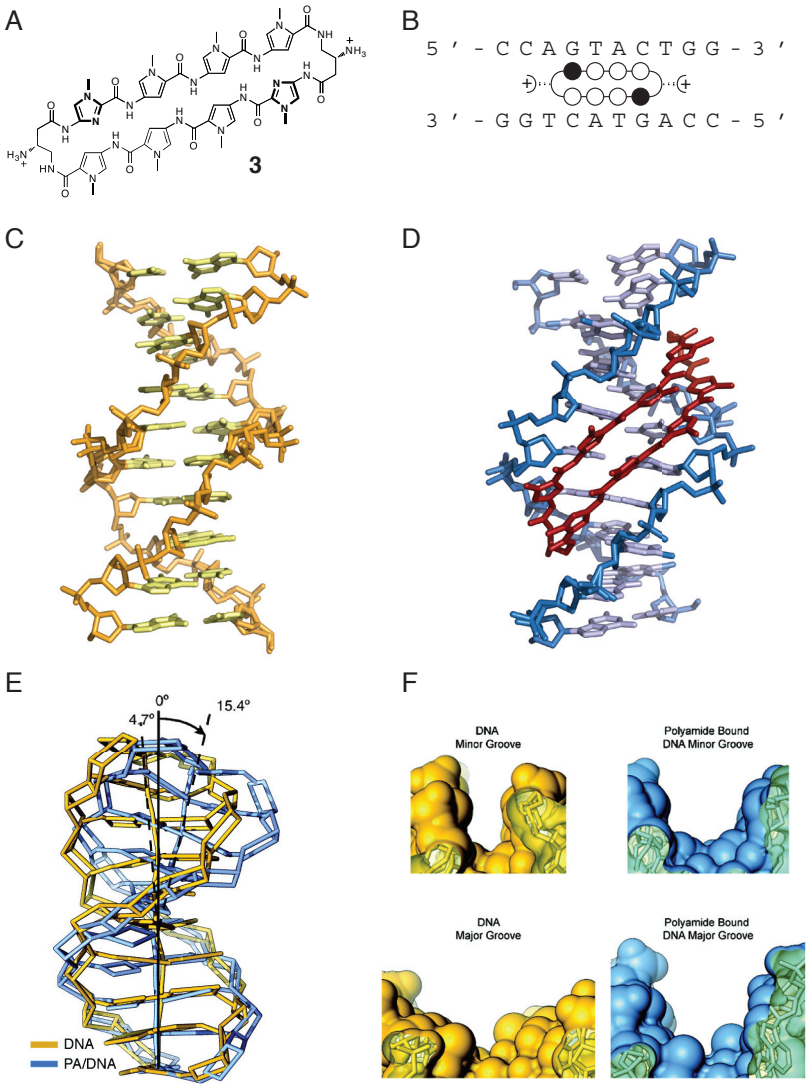

Figure 1.13. Structural basis for allosteric inhibition of major groove-binding transcription factors by minor groove-binding polyamides. (A) Chemical structure of ARE-targeted cyclic polyamide compound **3**. (B) Ball-and-stick representation of **3** bound to its target DNA sequence. (C) Atomic resolution crystal structure of the DNA duplex shown in (B) (PDB 1D8G). (D) Atomic resolution crystal structure of 3 bound to the DNA sequence shown in (B) (PDB 3OMJ). (E) Superimposition of structures (C) and (D) showing the 15 ° helical bend. (F) ~4 Å narrowing of the DNA minor groove and ~4 Å widening of the minor groove that occurs upon polyamide binding.

in great detail: polyamide binding widens the minor groove by 4 Å and narrows the major groove by 4 Å (Figure 1.13) (65,66). In addition, it introduces a 15° helical bend toward the major groove, further distorting it. Allosteric inhibition of major-groove binding transcription factors by polyamides has a sound structural and biochemical foundation.

Proof of principle for polyamides as transcriptional regulators was first established with a polyamide targeted to the DNA response elements of the major groove-binding transcription factor HIF1α (67). Signaling through HIF1α regulates the homeostatic response to low oxygen tension; this process is co-opted by cancer cells to drive neovascularization (Figure 1.2). The DNA binding consensus sequence for HIF1α is 5'-TACGTG-3'; DNA elements that match this consensus and accommodate HIF1α binding are known as Hypoxic Response Elements (HREs). We designed a polyamide (**X**) that binds the sequence 5'-WTWCGW-3' with nanomolar affinity (Figure 1.14). In HeLa and U251 cells, this compound inhibited the induction of some HIF1α target genes when the cells were treated with desferioxamine (DFO), a small molecule mimic of hypoxia. Among these genes was VEGF, the vascular endothelial growth factor that promotes blood vessel formation required for extensive tumor growth (68). Microarray analysis of gene expression showed that **4** blocked a subset of all DFO-induced changes in gene expression. Chromatin immunoprecipitation experiments confirmed decreased HIF1α occupancy at the VEGF HRE when cells were treated with **4** prior to DFO induction. Targeting the HRE for polyamide-mediated transcriptional modulation of an endogenous transcription factor in cell culture resulted in the inhibition of a medically important gene, VEGF.

We have also developed an 8-ring hairpin Py-Im polyamide (**6**) that antagonizes steroid-induced gene expression changes driven by the nuclear hormone receptors androgen receptor (AR) and glucocorticoid receptor (GR) (Figure 1.14) (69,70). In animals, these transcription factors form powerful nodes in the transduction of signals from circulating hormones: AR controls the development and maintenance of the male sexual phenotype in response to testosterone as described above, and GR regulates the expression of anti-

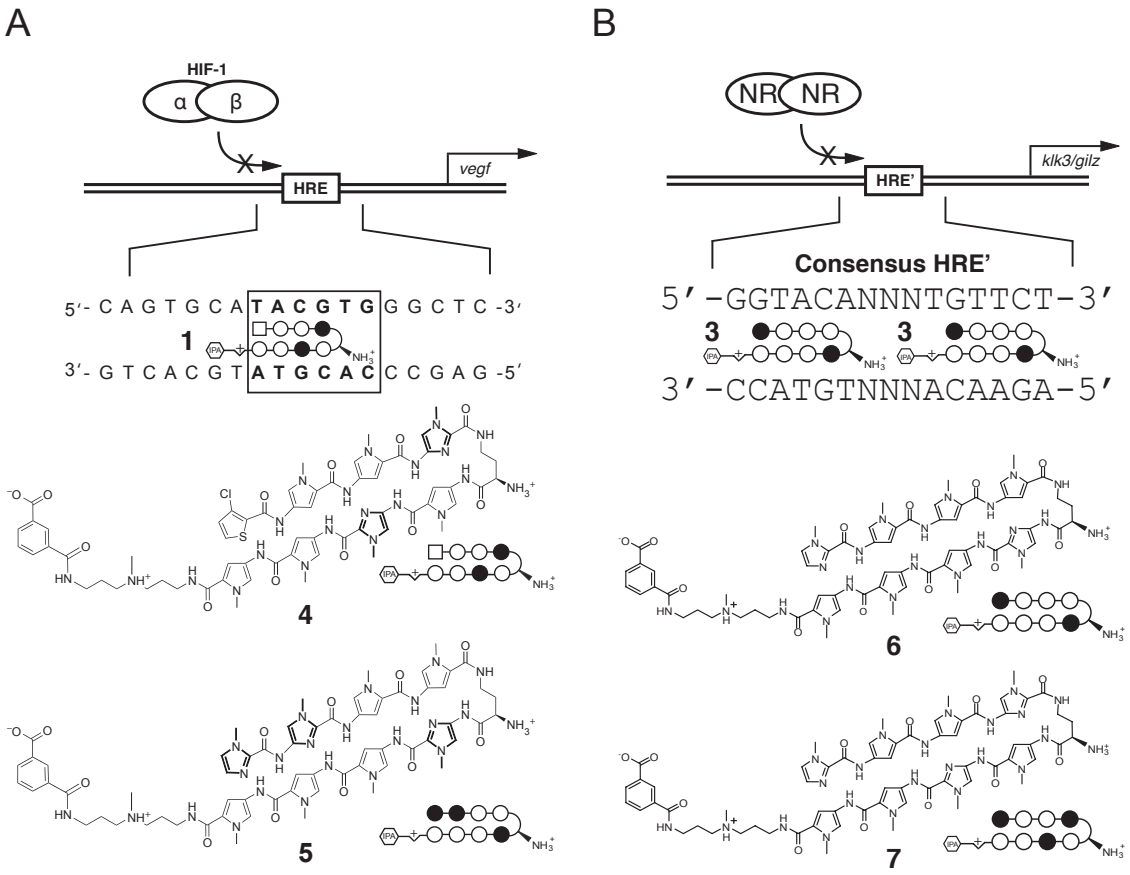

Figure 1.14. Py-Im polyamide inhibitors of HIF1α and nuclear hormone receptor signaling. (A) Cartoon representation of inhibition of HIF1α binding by compound **4**, which is designed to bind the hypoxia response element (HRE) in the VEGF promoter. Compound **4** inhibited the induction of VEGF mRNA expression and the expression of other HIF1α target genes in response to desferrioxamine, a chemomimetic of hypoxic conditions. Mismatch compound **5** is targeted to a sequence unrelated to the HRE consensus and was much less effective. (B) Cartoon representation of nuclear receptor transcription factors binding to a hormone response element (HRE') as a homodimer in response to ligand binding. Compound **6** is designed to target the consensus DNA binding sequence shared by androgen receptor (AR) and glucocoriticoid receptor (GR). Compound 6 inhibited the androgen-induced expression of some AR target genes, like KLK3 and the glucocorticoid-induced expression of some GR target genes (e.g. GILZ). Chromatin immunoprecipitation confirmed inhibition of transcription factor binding at the appropriate response elements for all three transcription factors.

inflammatory genes in response to cortisol (27,71). In the absence of ligand, these receptors are sequestered in the cytosol by heat shock proteins (HSPs). Steroid binding causes an allosteric conformational shift followed by homodimerization, nuclear localization, response element binding, and modulation of target gene transcription (72). AR and GR share the consensus binding sequence 5'-GGTACANNNTGTTCT-3'. Their genotropic actions can be partially inhibited in cancer tissue culture cells by polyamide **6**, which disrupts the protein-DNA interface by selectively binding the DNA sequence 5'-WGWWCW-3' (W=

A or T) (69,70). Chromatin immunoprecipitation experiments again confirmed decreased response element occupancy by the appropriate transcription factor when treated with **6**. This second example shows the generality of the polyamide approach; nuclear hormone receptors are structurally unrelated to HIF1α, yet both can be successfully inhibited with the same class of compound.

The polyamide HIF1α and AR/GR inhibitors are composed of 8 Py/Im ring pairs and bind 6 DNA base pairs. Larger hairpin compounds containing beta-alanine residues to increase flexibility and binding site size have been employed to inhibit the binding of a third class of transcription factor, AP-1. This compound inhibited AP-1 directed MMP9 expression in cell culture and mouse models of metastatic colorectal cancer with Py-Im polyamides (73). Our Japanese colleagues have begun preclinical work on these compounds in rats and achieved therapeutic dose levels following systemic administration without the use of delivery agents (74). These results are important first steps toward polyamide-based, transcription factor-targeted therapeutics.

1.8 Scope of this work

This thesis builds on all of the optimization and proof-of-principle experiments in an effort to define the scope and the limitations of DNA-binding Py-Im polyamides as inhibitors of transcription factors. Chapter 2 describes an effort to define the polyamide tool kit: we design, synthesize, and measure the DNA-binding affinities and specificities of the set of polyamides targeted to all possible 5'-WGNNNW-3' (W = A or T, N = any nucleotide) sequences. Chapter 3 details the synthesis and biological activity of a cyclic polyamide designed to inhibit AR binding. This compound maintains its activity against AR-driven gene expression in hormone-sensitive prostate cancer cells and represents an alternative polyamide architecture for use in future animal experiments. In chapter 4, I present an attempt to extend the AR work in a hormone-resistant cell line. This attempt ultimately fails. We observe inhibition of AR-target gene mRNA expression, but this change

is not accompanied by decreased AR occupancy at those genes when cells are treated with polyamide. In addition, polyamide treatment is accompanied by significant cytotoxicity and activation of a stress response. These events are explored further in chapters 5 and 6. The former investigates a possible role for polyamides as DNA Topoisomerase II inhibitors, and the latter establishes **6** as an inhibitor of DNA synthesis in cell culture. I discuss the ramifications of these findings on the design and use of polyamides as transcription factor antagonists.

References.

1. (2004) Finishing the euchromatic sequence of the human genome. *Nature*, **431**, 931-945.
2. Gilbert, D.G. (2002) euGenes: a eukaryote genome information system. *Nucleic Acids Res*, **30**, 145-148.
3. Reiter, L.T., Potocki, L., Chien, S., Gribskov, M. and Bier, E. (2001) A systematic analysis of human disease-associated gene sequences in Drosophila melanogaster. *Genome Res*, **11**, 1114-1125.
4. Orgel, L.E. and Crick, F.H. (1980) Selfish DNA: the ultimate parasite. *Nature*, **284**, 604-607.
5. Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E. *et al.* (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, **447**, 799-816.
6. Babu, M.M., Luscombe, N.M., Aravind, L., Gerstein, M. and Teichmann, S.A. (2004) Structure and evolution of transcriptional regulatory networks. *Curr Opin Struct Biol*, **14**, 283-291.
7. Latchman, D.S. (1997) Transcription factors: an overview. *Int J Biochem Cell Biol*, **29**, 1305-1312.
8. Murre, C., McCaw, P.S., Vaessin, H., Caudy, M., Jan, L.Y., Jan, Y.N., Cabrera, C.V., Buskin, J.N., Hauschka, S.D., Lassar, A.B., et al. (1989) Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell*, **58**, 537-544.
9. Krishna, S.S., Majumdar, I. and Grishin, N.V. (2003) Structural classification of zinc fingers: survey and summary. *Nucleic Acids Res*, **31**, 532-550.
10. Stros, M., Launholt, D., Grasser, K.D. (2007) The HMG-box: a versatile protein domain occurring in a wide variety of DNA-binding proteins. *Cell Mol Life Sci*, **64**, 19-20.
11. Love, J.J., Li, X., Case, D.A., Giese, K., Grosschedl, R. and Wright, P.E. (1995) Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature*, **376**, 791-795.
12. Remenyi, A., Lins, K., Nissen, L.J., Reinbold, R., Scholer, H.R. and Wilmanns, M. (2003) Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes Dev*, **17**, 2048-2059.
13. Ptashne, M. and Gann, A. (1997) Transcriptional activation by recruitment. *Nature*, **386**, 569-577.
14. Roeder, R.G. (2005) Transcriptional regulation and the role of diverse coactivators in animal cells. *FEBS Lett*, **579**, 909-915.
15. Whiteside, S.T. and Goodbourn, S. (1993) Signal transduction and nuclear targeting: regulation of transcription factor activity by subcellular localisation. *J Cell Sci*, **104 (Pt 4)**, 949-955.
16. Conaway, R.C., Brower, C.S. and Conaway, J.W. (2002) Emerging roles of ubiquitin in transcription regulation. *Science*, **296**, 1254-1258.
17. Bohmann, D. (1990) Transcription factor phosphorylation: a link between signal transduction and the regulation of gene expression. *Cancer Cells*, **2**, 337-344.

18. Brivanlou, A.H. and Darnell, J.E., Jr. (2002) Signal transduction and the control of gene expression. *Science*, **295**, 813-818.
19. Panne, D., Maniatis, T. and Harrison, S.C. (2007) An atomic model of the interferon-beta enhanceosome. *Cell*, **129**, 1111-1123.
20. Davison, S.L. and Bell, R. (2006) Androgen physiology. *Semin Reprod Med*, **24**, 71-77.
21. Jenster, G., van der Korput, H.A., Trapman, J. and Brinkmann, A.O. (1995) Identification of two transcription activation units in the N-terminal domain of the human androgen receptor. *J Biol Chem*, **270**, 7341-7346.
22. Shaffer, P.L., Jivan, A., Dollins, D.E., Claessens, F. and Gewirth, D.T. (2004) Structural basis of androgen receptor binding to selective androgen response elements. *Proc Natl Acad Sci U S A*, **101**, 4758-4763.
23. Kaku, N., Matsuda, K., Tsujimura, A. and Kawata, M. (2008) Characterization of nuclear import of the domain-specific androgen receptor in association with the importin alpha/beta and Ran-guanosine 5'-triphosphate systems. *Endocrinology*, **149**, 3960-3969.
24. He, B., Gampe, R.T., Jr., Kole, A.J., Hnat, A.T., Stanley, T.B., An, G., Stewart, E.L., Kalman, R.I., Minges, J.T. and Wilson, E.M. (2004) Structural basis for androgen receptor interdomain and coactivator interactions suggests a transition in nuclear receptor activation function dominance. *Mol Cell*, **16**, 425-438.
25. Lanz, R.B. and Rusconi, S. (1994) A conserved carboxy-terminal subdomain is important for ligand interpretation and transactivation by nuclear receptors. *Endocrinology*, **135**, 2183-2195.
26. Veldscholte, J., Berrevoets, C.A., Brinkmann, A.O., Grootegoed, J.A. and Mulder, E. (1992) Anti-androgens and the mutated androgen receptor of LNCaP cells: differential effects on binding affinity, heat-shock protein interaction, and transcription activation. *Biochemistry*, **31**, 2393-2399.
27. Lamont, K.R. and Tindall, D.J. (2010) Androgen regulation of gene expression. *Adv Cancer Res*, **107**, 137-162.
28. Kyprianou, N. and Isaacs, J.T. (1988) Activation of programmed cell death in the rat ventral prostate after castration. *Endocrinology*, **122**, 552-562.
29. Tomlins, S.A., Rhodes, D.R., Perner, S., Dhanasekaran, S.M., Mehra, R., Sun, X.W., Varambally, S., Cao, X., Tchinda, J., Kuefer, R. *et al.* (2005) Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science*, **310**, 644-648.
30. Bracarda, S., Logothetis, C., Sternberg, C.N. and Oudard, S. (2011) Current and emerging treatment modalities for metastatic castration-resistant prostate cancer. *BJU Int*, **107 Suppl 2**, 13-20.
31. Chen, C.D., Welsbie, D.S., Tran, C., Baek, S.H., Chen, R., Vessella, R., Rosenfeld, M.G. and Sawyers, C.L. (2004) Molecular determinants of resistance to antiandrogen therapy. *Nat Med*, **10**, 33-39.
32. Tran, C., Ouk, S., Clegg, N.J., Chen, Y., Watson, P.A., Arora, V., Wongvipat, J., Smith-Jones, P.M., Yoo, D., Kwon, A. *et al.* (2009) Development of a second-generation antiandrogen for treatment of advanced prostate cancer. *Science*, **324**, 787-790.

33. Scher, H.I., Beer, T.M., Higano, C.S., Anand, A., Taplin, M.E., Efstathiou, E., Rathkopf, D., Shelkey, J., Yu, E.Y., Alumkal, J. *et al.* (2010) Antitumour activity of MDV3100 in castration-resistant prostate cancer: a phase 1-2 study. *Lancet*, **375**, 1437-1446.
34. Darnell, J.E., Jr. (2002) Transcription factors as targets for cancer therapy. *Nat Rev Cancer*, **2**, 740-749.
35. Koehler, A.N. (2010) A complex task? Direct modulation of transcription factors with small molecules. *Curr Opin Chem Biol*, **14**, 331-340.
36. Block, K.M., Wang, H., Szabo, L.Z., Polaske, N.W., Henchey, L.K., Dubey, R., Kushal, S., Laszlo, C.F., Makhoul, J., Song, Z. *et al.* (2009) Direct inhibition of hypoxia-inducible transcription factor complex with designed dimeric epidithiodik etopiperazine. *J Am Chem Soc*, **131**, 18078-18088.
37. Kung, A.L., Zabludoff, S.D., France, D.S., Freedman, S.J., Tanner, E.A., Vieira, A., Cornell-Kennon, S., Lee, J., Wang, B., Wang, J. *et al.* (2004) Small molecule blockade of transcriptional coactivation of the hypoxia-inducible factor pathway. *Cancer Cell*, **6**, 33-43.
38. Kong, D., Park, E.J., Stephen, A.G., Calvani, M., Cardellina, J.H., Monks, A., Fisher, R.J., Shoemaker, R.H. and Melillo, G. (2005) Echinomycin, a small-molecule inhibitor of hypoxia-inducible factor-1 DNA-binding activity. *Cancer Res*, **65**, 9047-9055.
39. Pecot, C.V., Calin, G.A., Coleman, R.L., Lopez-Berestein, G. and Sood, A.K. (2011) RNA interference in the clinic: challenges and future directions. *Nat Rev Cancer*, **11**, 59-67.
40. Tan, S., Guschin, D., Davalos, A., Lee, Y.L., Snowden, A.W., Jouvenot, Y., Zhang, H.S., Howes, K., McNamara, A.R., Lai, A. *et al.* (2003) Zinc-finger protein-targeted gene regulation: genomewide single-gene specificity. *Proc Natl Acad Sci U S A*, **100**, 11997-12002.
41. Pavletich, N.P. and Pabo, C.O. (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science*, **252**, 809-817.
42. Segal, D.J., Dreier, B., Beerli, R.R. and Barbas, C.F., 3rd. (1999) Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. *Proc Natl Acad Sci U S A*, **96**, 2758-2763.
43. Greisman, H.A. and Pabo, C.O. (1997) A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science*, **275**, 657-661.
44. Liu, Q., Segal, D.J., Ghiara, J.B. and Barbas, C.F., 3rd. (1997) Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. *Proc Natl Acad Sci U S A*, **94**, 5525-5530.
45. Carroll, D. (2008) Progress and prospects: zinc-finger nucleases as gene therapy agents. *Gene Ther*, **15**, 1463-1468.
46. Zhao, L. and Chmielewski, J. (2005) Inhibiting protein-protein interactions using designed molecules. *Curr Opin Struct Biol*, **15**, 31-34.
47. Moellering, R.E., Cornejo, M., Davis, T.N., Del Bianco, C., Aster, J.C., Blacklow, S.C., Kung, A.L., Gilliland, D.G., Verdine, G.L. and Bradner, J.E. (2009) Direct inhibition of the NOTCH transcription factor complex. *Nature*, **462**, 182-188.

48. Dickerson, R.E., Drew, H.R., Conner, B.N., Wing, R.M., Fratini, A.V. and Kopka, M.L. (1982) The anatomy of A-, B-, and Z-DNA. *Science*, **216**, 475-485.
49. Tse, W.C. and Boger, D.L. (2004) Sequence-selective DNA recognition: natural products and nature's lessons. *Chem Biol*, **11**, 1607-1617.
50. Dervan, P.B. and Edelson, B.S. (2003) Recognition of the DNA minor groove by pyrrole-imidazole polyamides. *Curr Opin Struct Biol*, **13**, 284-299.
51. Coll, M., Frederick, C.A., Wang, A.H. and 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**, 8385-8389.
52. Pelton, J.G. and Wemmer, D.E. (1989) Structural characterization of a 2:1 distamycin A.d(CGCAAATTGGC) complex by two-dimensional NMR. *Proc Natl Acad Sci U S A*, **86**, 5723-5727.
53. Mrksich, M., Wade, W.S., Dwyer, T.J., Geierstanger, B.H., Wemmer, D.E. and Dervan, P.B. (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**, 7586-7590.
54. Wade, W.S., Mrksich, M. and Dervan, P.B. (1992) Design of Peptides That Bind in the Minor Groove of DNA at 5'-(a,T)G(a,T)C(a,T)-3' Sequences by a Dimeric Side-by-Side Motif. *J Am Chem Soc*, **114**, 8783-8794.
55. 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. and Dervan, P.B. (2007) Completion of a Programmable DNA-Binding Small Molecule Library. *Tetrahedron*, **63**, 6146-6151.
56. Trauger, J.W., Baird, E.E. and Dervan, P.B. (1998) Recognition of 16 base pairs in the minor groove of DNA by a pyrrole-imidazole polyamide dimer. *J Am Chem Soc*, **120**, 3534-3535.
57. Doss, R.M., Marques, M.A., Foister, S., Chenoweth, D.M. and Dervan, P.B. (2006) Programmable oligomers for minor groove DNA recognition. *J Am Chem Soc*, **128**, 9074-9079.
58. White, S., Szewczyk, J.W., Turner, J.M., Baird, E.E. and Dervan, P.B. (1998) Recognition of the four Watson-Crick base pairs in the DNA minor groove by synthetic ligands. *Nature*, **391**, 468-471.
59. Trauger, J.W., Baird, E.E. and Dervan, P.B. (1996) Recognition of DNA by designed ligands at subnanomolar concentrations. *Nature*, **382**, 559-561.
60. Herman, D.M., Turner, J.M., Baird, E.E. and Dervan, P.B. (1999) Cycle polyamide motif for recognition of the minor groove of DNA. *J Am Chem Soc*, **121**, 1121-1129.
61. Cohen, J.D., Sadowski, J.P. and Dervan, P.B. (2007) Addressing single molecules on DNA nanostructures. *Angew Chem Int Edit*, **46**, 7956-7959.
62. Edayathumangalam, R.S., Weyermann, P., Gottesfeld, J.M., Dervan, P.B. and Luger, K. (2004) Molecular recognition of the nucleosomal "supergroove". *Proc Natl Acad Sci U S A*, **101**, 6864-6869.
63. Edelson, B.S., Best, T.P., Olenyuk, B., Nickols, N.G., Doss, R.M., Foister, S., Heckel, A. and Dervan, P.B. (2004) Influence of structural variation on nuclear

localization of DNA-binding polyamide-fluorophore conjugates. *Nucleic Acids Res*, **32**, 2802-2818.

64. Best, T.P., Edelson, B.S., Nickols, N.G. and Dervan, P.B. (2003) Nuclear localization of pyrrole-imidazole polyamide-fluorescein conjugates in cell culture. *Proc Natl Acad Sci U S A*, **100**, 12063-12068.
65. Chenoweth, D.M. and Dervan, P.B. (2010) Structural basis for cyclic Py-Im polyamide allosteric inhibition of nuclear receptor binding. *J Am Chem Soc*, **132**, 14521-14529.
66. Chenoweth, D.M. and Dervan, P.B. (2009) Allosteric modulation of DNA by small molecules. *Proc Natl Acad Sci U S A*, **106**, 13175-13179.
67. Olenyuk, B.Z., Zhang, G.J., Klco, J.M., Nickols, N.G., Kaelin, W.G., Jr. and Dervan, P.B. (2004) Inhibition of vascular endothelial growth factor with a sequence-specific hypoxia response element antagonist. *Proc Natl Acad Sci U S A*, **101**, 16768-16773.
68. Fong, G.H. (2009) Regulation of angiogenesis by oxygen sensing mechanisms. *J Mol Med*, **87**, 549-560.
69. Muzikar, K.A., Nickols, N.G. and Dervan, P.B. (2009) Repression of DNA-binding dependent glucocorticoid receptor-mediated gene expression. *Proc Natl Acad Sci U S A*, **106**, 16598-16603.
70. Nickols, N.G. and Dervan, P.B. (2007) Suppression of androgen receptor-mediated gene expression by a sequence-specific DNA-binding polyamide. *Proc Natl Acad Sci U S A*, **104**, 10418-10423.
71. Nicolaides, N.C., Galata, Z., Kino, T., Chrousos, G.P. and Charmandari, E. (2010) The human glucocorticoid receptor: molecular basis of biologic function. *Steroids*, **75**, 1-12.
72. Aranda, A. and Pascual, A. (2001) Nuclear hormone receptors and gene expression. *Physiol Rev*, **81**, 1269-1304.
73. Wang, X., Nagase, H., Watanabe, T., Nobusue, H., Suzuki, T., Asami, Y., Shinojima, Y., Kawashima, H., Takagi, K., Mishra, R. *et al.* (2010) Inhibition of MMP-9 transcription and suppression of tumor metastasis by pyrrole-imidazole polyamide. *Cancer Sci*, **101**, 759-766.
74. Nagashima, T., Aoyama, T., Fukasawa, A., Watabe, S., Fukuda, N., Ueno, T., Sugiyama, H., Nagase, H. and Matsumoto, Y. (2009) Determination of pyrrole-imidazole polyamide in rat plasma by liquid chromatography-tandem mass spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci*, **877**, 1070-1076.