Repression of DNA-Binding-Dependent Glucocorticoid Receptor-Mediated Gene Expression

Thesis by

Katy Ann Muzikar

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2011

(Defended November 29, 2010)

This thesis is dedicated to Jonathan...

...this journey has been ours together.

Acknowledgements

Five years seems like too short of a time to have learned all that I have learned, accomplished all that I have accomplished, and ultimately grown into the person and scientist that I find myself to be. I know that I wouldn't be nearly where I am today if it weren't for the love, support, encouragement, and mentoring of a great many people. I want to take time now to acknowledge some of those people, even knowing that I can't possibly properly thank everyone who has been a positive impact on my graduate career.

Peter: It has always been impressive to me that you are able to individually encourage each of your students along their own unique paths, and I will always be thankful for your positive energy, expertise, and patience as I learned to become the manager of my own "group of one." I am confident in myself as a scientist and a teacher and am looking forward to going forth into this world to share with others the things that you have shared with me. *Jim:* Your energy, intelligence, and passion for the art of rational thought is inspiring. Your thoroughness and patience with me as a young Padawan has made my years in the Dervan Lab infinitely more productive than they would have been without your guidance. *Nick:* It has been an honor following in your footsteps. *Dan:* I have learned from you what it means to be a focused and dedicated scientist, and while I still have no real bark or bite, at least I learned from you how to growl, just a bit. *Michelle, Justin, and Mareike:* I love you like the little family we were—Mareike the wise and grounded older sister, Justin the hunky older brother that all my junior high school friends would have had a crush on, and Michelle, my crazy dizygotic twin (we aren't anything alike but we are linked together forever anyway). It was heartbreaking to watch each of you leave one by one, but the memories of the time we had together for coffee, Amigos, and fun in San Diego will last my whole lifetime. Thank you for getting me through the days when science refused to be kind, and celebrating with me when I finally got it right.

The Dervan Lab: In my five years I had the pleasure and good fortune of spending time with many great people: John, Anne, Ryan, Mike B., Jim S., Sherry, Christian, Carey, Dave C., Claire, Dan G., Ben, Dave M., Fei, Jordan, Jevgenij, JJ, Thomas, and Mandy. Every one of you has been a part of my Caltech experience—we have worked together to teach and learn from each other, to protect each other from the dangers of chemistry and of life, and to celebrate at the Rath Al Fresco whenever possible. The third floor of Church is home to an amazing group of people—ever changing yet always a group with a cooperative spirit and passion for life. Best wishes to each of you in your own individual journeys and I can't wait to see each of you whenever life throws us back together again.

Thank you to my thesis commitee: Linda Hsieh-Wilson, Doug Rees, and Dennis Dougherty. Your advice and support has been invaluable and it has been an honor to have scientific and career disscussions with such brilliant minds.

There is absolutely no way I would have accomplished everything in this thesis without the services of the friendly, dedicated, amazing support staff of the Division of Chemistry and Chemical Engineering. Joe Drew and Ron Koen, I hope that you guys know that you made my day infinitely better countless times—with banter, with problem-solving, and of course with sugar. Paul Carroad, the go-to guy for absolutely everything, I'm not sure that anything would have gotten done without your help. Lynne Martinez, you have been an angel of patience and a bright smile of encouragement when I needed it most. Dianne Buchness, Anne Penny, and Agnes Tong, who took care of so many behind-the-scenes things that I don't even know about to make sure that my time as a graduate student could be productive and (somewhat) less full of stress. Steve Gould, Chris Smith and Leah Mentch, who put up with last-minute-radiation-orders and everything else that I brought them and always did so without complaint. Much thanks to the group that fixes everything we break, no matter how many times we break it: Tom Dunn, Rick Gerhart, Tony Solyom,

Mike Roy, and Steve Olson. The excellent facilities and great intellect of Mona Shahgholi, David Vander Velde, Scott Ross and Scott Virgil have been invaluable.

Not only is Caltech top-class in research, but our support staff and departments are excellent. To the Department of Environmental Health and Safety, especially Haick Issaian, Karen Baumgartner, Larry Martinez, Art Seiden, and Andrea Acosta thank you for your continued commitment to keep us safe, to keep the community safe, and to continue to educate us to be better and safer scientists. Karen--it has been wonderful having radiation delievered by such a vibrant personality, and best wishes in life to you and your family. A special thank you to the HVAC technicians, especially Steve and Reuben, for bandaging our McQuay units time and time again (even at the cost of getting soaking wet).

A great big amazing thank you to my family. My mother and father are responsible for my deep-rooted love of science. They took me to the Exploratorium and I never looked back. Mom--your love and support, both emotional and physical, has meant the world to me, and I don't know that I would have made it through this time without our near-daily phone conversations. Dad--your wonder at the natural world around you must be in your genes, because you definitely passed it down to me. I'm glad that I could be the organic chemist that you always wanted to be. My grandparents have always been a source of inspiration and having their support and interest in my work is an amazing feeling. To the Josephson clan, yeah, you know who you are, your energy and enthusiastic support of anything and everything I do is astounding. I couldn't have picked a better group of people to add to my family. To Tricksy, your unconditional love and warm cuddly support has made my world a better place.

And finally to Jonathan: Thank you for being strong enough, and for sharing your strength. I love you.

Abstract

Gene expression is controlled by transcription factors that regulate the rates at which genes are expressed either by recruiting or inhibiting protein complexes that bind to the promoters or enhancers of target genes. Molecules that can specifically modulate these protein-DNA interfaces show promise as tools for understanding gene regulation pathways and may have application in human medicine. Hairpin pyrrole-imidazole polyamides are programmable oligomers that bind the DNA minor groove in a sequence-specific manner with affinities comparable to those of natural DNA-binding proteins. These cell-permeable small molecules have been shown to enter the nuclei of live cells, disrupt protein-DNA interactions, and downregulate endogenous gene expression. This thesis describes the use of polyamides to modulate gene expression in order to probe gene regulation mechanisms of several different biologically relevant systems. A polyamide is designed to target the glucocorticoid receptor transcription factor DNA binding site located in the promoter of the glucocorticoid-induced leucine zipper gene. This polyamide is shown to bind with high affinity to the promoter sequence, modulate the expression of this gene, and disrupt the binding of the protein to the gene's promoter. Examination of the global effects of this polyamide on mRNA transcription is used to elucidate a list of genes that are regulated by a glucocorticoid receptor protein-DNA dependent mechanism. Also in this thesis, the specificities of a Cy3-labeled polyamide known to downregulate expression of the Vascular Endothelial Growth Factor is examined using DNA microarrays composed of hairpins harboring all 524,800 unique 10 base pair DNA sequences. We experimentally verify the correlation of Cy3 fluorescence intensity with quantitative DNase I footprint-derived binding affinities. Additionally, progress is made towards the polyamide-mediated inhibition of Myc/Max transcription factor gene regulation.

Table of Contents

List of Figures and Schemes

List of Figures and Schemes

Chapter 3

Chapter 4

List of Figures and Schemes

List of Tables

List of Abbreviations

[PA]	free polyamide concentration
A	adenine
Å	angstrom
Ac	acetyl
Ac_2O	acetic anhydride
ADMET	absorption, distribution, metabolism, excretion and toxicity
A_{595}	absorbance maximum
AP-1	activator protein 1
AR	androgen receptor
ARE	androgen response element
A·T	adenine Watson-Crick hydrogen bonded to thymine
ATCC	American Type Culture Collection
atm	atmosphere
β	beta-amino alanine
Boc	*tert*-butyloxycarbonyl
Boc-Im-OH	(4-[(tert-Butoxycarbonyl)amino]-1-methylimidazole-2-carboxylic acid)
Boc_2O	di-*tert*-butyl dicarbonate
Boc-Py-OBt	[(1,2,3-Benzotriazol-1-yl 4-[(tert-Butoxycarbonyl)amino]-1-methylpyrrole-2-carboxylate)
bp	base pair
BSA	bovine serum albumin
°C	degrees Celsius
C	cytosine
cDNA	complementary deoxyribonucleic acid
C·G	cytosine Watson-Crick hydrogen bonded to guanine
calc'd	calculated
Cbz	carbobenzyloxy
ChIP	chromatin immunoprecipitation
cm	centimeter
CSI	Cognate Site Identity/Cognate Site Identifier
Ct	2-carbosy-3-chlorothiophene
CT-FBS	charcoal treated fetal bovine serum
Cy3	Cyanine 3
Da	Dalton
dATP	2'-deoxyadenosine triphosphate
DABA	diaminobutyric acid
DCM	dichloromethane
dex	dexamethasone
DFO	deferoxamine

contiued on page xv

List of Abbreviations

DHT	dihydrotestosterone
DIEA	N,N-diisopropylethylamine
DMF	N,N-dimethylformamide
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
dox	doxycycline
Dp	N,N-dimethylaminopropylamine
DPPA	diphenylphosphoryl azide
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
eIF4e	eukaryotic translation initiation factor 4E
ELISA	enzyme-linked immunosorbant assay
EMSA	elecrophoretic mobility shift assay
ESI	elecrospray ionization
Et_2O	diethyl ether
Ex	excitation
FBS	fetal bovine serum
FITC	fluorescein isothiocyanate
Fmoc	fluorenylmethyloxycarbonyl
Fn	4-aminofuran
γ-DABA	γ-2,4-diaminobutyric acid
G	guanine
G·C	guanine Watson-Crick hydrogen bonded to cytosine
GABA	γ-aminobutyric acid
GC	glucocorticoid
GILZ	glucocorticoid-induced leucine zipper
GR	glucocorticoid receptor
h	hour(s)
HIF-1α	hypoxia inducible factor 1α
Hp	3-hydroxypyrrole
HPLC	high-performance liquid chromatography
HRE	hypoxic response element
HRP	horseradish peroxidase
Hsp	heat-shock protein
IgG	immunoglobulin G
Im	N-methylimidazole
IPA	isophthalic acid
IPTG	isopropyl β-D-1-thiogalactopyranoside
k	kilo (1×10^3)

continued on page xvi

List of Abbreviations

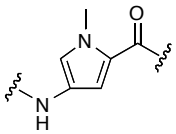
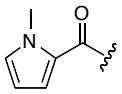
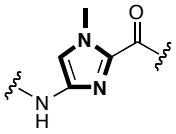
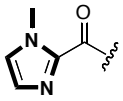
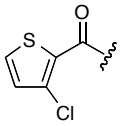
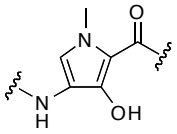
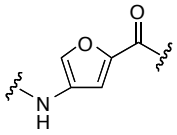
K_a	association constant
K_d	dissociation constant
λ	wavelength
LN_2	liquid nitrogen
m/z	mass to charge ratio
μ	micro (1×10^{-6})
M	molar
m	milli (1×10^{-3})
Max	Myc associated protein X
max	maximum
MALDI	matrix-assisted LASER desorption/ionization
MAS	maskless array synthesis
mif	mifepristone
min	minute(s)
mol	mole(s)
MPE	methidiumpropyl·ethylenediaminetetraacetic acid
mRNA	messenger ribonucleic acid
MS	mass spectrometry
N	A, T, G, or C
n	nano (1×10^{-9})
n-BuLi	n-butyl lithium
NF-κB	nuclear factor-κB
Θ	fractional occupancy
OBt	hydroxytriazole ester
p	pico (1×10^{-12})
PCR	polymerase chain reaction
PET	Paired End diTag
PIC	protease inhibitor cocktail
PMSF	phenylmethanesulfonylfluoride
PSA	prostate-specific antigen
Py-Im	pyrrole-imidazole
qPCR	quantitative polymerase chain reaction
RT	room temperature
RT-PCR	reverse transcriptase polymerase chain reaction
PAGE	polyacrylamide gel electrophoresis
PBS	phophate-buffered saline
Py	N-methylpyrrole
PyBOP	(benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate

continued on page xvii

List of Abbreviations

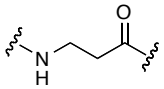
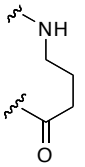
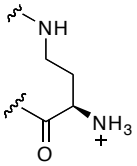
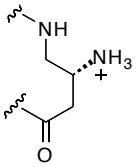
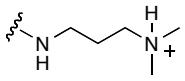
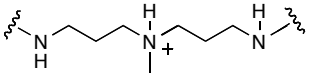
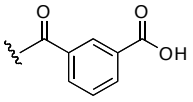
R	guanine or adenine
RCF	relative centrifugal force
RIPA	radio immunoprecipitation assay
RNA	ribonucleic acid
RNAi	ribonucleic acid interference
RT	reverse transcription
SAGE	serial analysis of gene expression
s-BuLi	*sec*-butyl lithium
siRNA	small interfering ribonucleic acid
Smad	Sma and Mad-related protein
STAT	signal transduction and activator of transcription
T	thymine
T·A	thymine Watson-Crick hydrogen bonded to adenine
t-BuOH	*tert*-butanol
TF	transcription factor
TFA	trifluoroacetic acid
TFO	triplex-forming oligonucleotides
THF	tetrahydrofuran
T_m	midpoint of transition temperature
TOF	time-of-flight
TFRE	transcription factor response element
tri/triamine	3,3'-diamino-N-methyldipropylamine
U	uracil
UV	ultraviolet
VEGF	vascular endothelial growth factor
Vis	visible
W	adenine or thymine
X-gal	bromo-chloro-indolyl-galactopyranoside

List of Symbols and Nomenclature

	-Py-	
	Py-	
	-Im-	
	Im-	
	Ct-	
	-Hp-	
	-Fn-	

continued on page xix

List of Symbols and Nomenclature

	-β-	
	-γ-	
	-γ-DABA-	
	-(R)-β-amino-γ-	
	-Dp	
	-tri-	
	-IPA	

Chapter 1

Introduction

1.1 Gene expression and the central dogma

In 1958 the means by which life as we know it persists and thrives was enumerated in the form of the central dogma of molecular biology.[1] In its simplest form, the dogma states that each living cell holds the blueprint to life in the form of genes, and that this genomic information is transcribed from DNA to RNA, which is then translated into proteins. Over half a century later, innumerable advances in technology have expanded this simple view to include a host of mechanisms by which each step of this process is influenced and modulated in response to cellular needs. The genome is not a blueprint of a static cellular state but actually encompasses all the instructions needed for a cell to respond to environmental stimuli. Sequencing of the human genome estimates that humans possess 20,000 to 30,000 genes,[2,3] all of which are maintained folded within chromosome structures and are accessed as needed by the cell to maintain normal function. Information encoding the structure, regulation, and expression of each of these genes as well as their expressed products is encoded in the nucleotide sequences of the base pairs of the DNA molecules that make up the genome.

Gene expression is the process by which a gene's coded information is converted from these nucleotide sequences into functioning gene products. This process is used by all known life forms to generate the RNAs, proteins, and various macromolecular machineries needed for life. In eukaryotic cells, gene expression is a complex process involving a variety of steps prior to the actual synthesis of a protein. These include the transcription of the gene into the primary RNA product, processing of this initial gene transcript to remove intron sequences and create the mature 3' terminus, transport of the processed mRNA transcript to the cytoplasm, and then, finally, translation of the messenger RNA into protein. With very few exceptions, all of the genes that encode proteins follow this pathway.[4]

1.2 Regulation of gene expression: transcriptional modulation

Any step of the gene expression process is subject to modulation, from the transcription of DNA into RNA all the way through the post-translational modification of the final protein. Every cell in any given organism contains the exact same genes, yet the multitude of cell types in a human body each display a different phenotype, produce different proteins at different stages of the life cycle, and somehow display different programs of gene expression. At the most fundamental level, regulation of gene expression is what allows genotype to be converted into phenotype. Modulation of the amount and timing of the appearance of the final gene product allows the cell to maintain control over structure and function and provides the basis for the versatility and adaptability of a living organism. A cell's ability to regulate and control the expression of each gene product allows the cell the flexibility to respond to environmental stimuli such as chemical signaling, environmental variation, cell damage, etc.

Regulation at the transcriptional level is a key element in the regulation of gene expression and can be divided into three general categories of influence: 1) Regulation at the genetic level involves direct interaction of a control factor with the gene in question, 2) Regulation via interaction of a control factor with the transcription machinery that results in modulation of gene expression, and 3) Epigenetic regulation involving global alterations in DNA structure that influence transcription. Direct interaction with DNA is the simplest and most direct method by which a protein can modulate the level of gene transcription. In general, genes harbor several protein binding sites specifically utilized for the regulation of transcription. These protein binding sites include enhancers, insulators, repressors, and silencers and the mechanisms for regulating transcription through these binding sites are complex and vary from activating transcription by recruiting RNA polymerase to blocking key RNA polymerase binding sites to hinder gene transcription.

Transcription Factors

Transcription factor proteins constitute a major class of DNA-binding molecules that participate in the transcriptional control of gene expression. Transcription factors bind to DNA in gene promoters or enhancer regions, generally upstream of the transcription start site, and modulate the frequency of RNA polymerase binding to the gene and subsequent transcription. Transcription factors vary widely in architecture, as the small sample of transcription factors shown in **Figure 1.1** illustrates. While different transcription factors accomplish their tasks via different means, all transcription factors display two critical abilities—the ability to bind to DNA in a sequence-specific manner, and the ability to cause a modulation of transcription once bound.[4] Most transcription factors possess a specific DNA-binding domain that is responsible for recognition and interaction with a specific DNA sequence, typically four to eight nucleotides in length, often located in an enhancer or promoter region of a given gene.

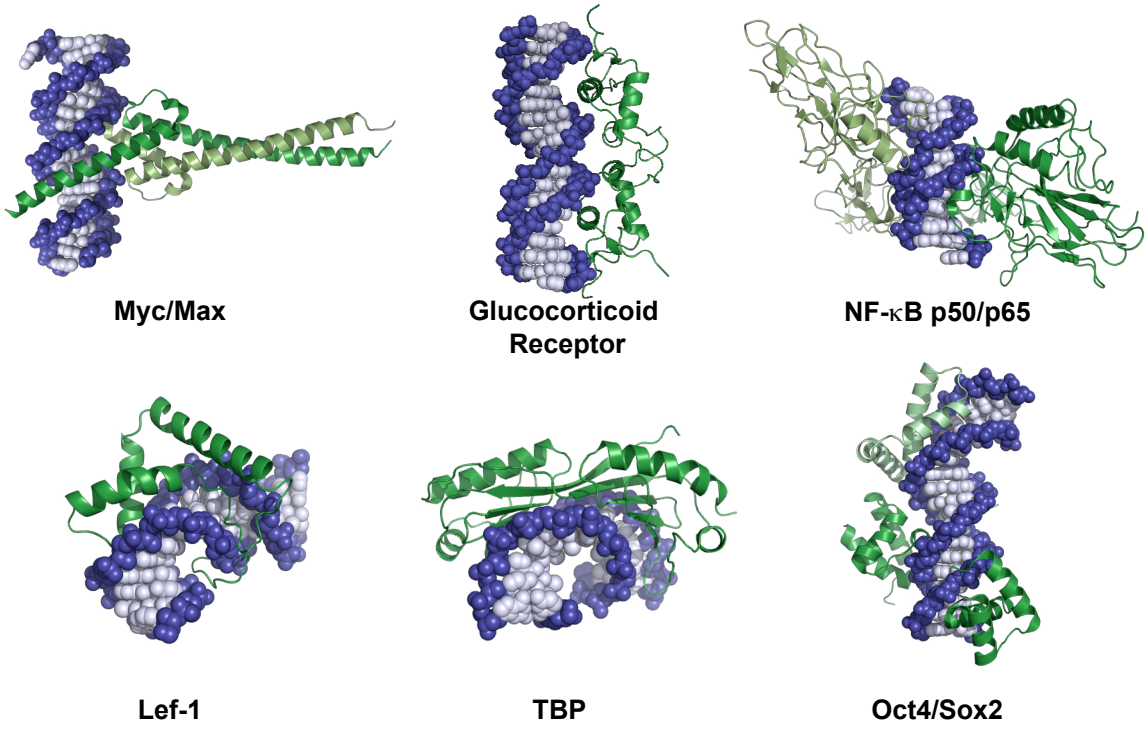


Figure 1.1 X-ray crystal structures of transcription factor–DNA complexes

Myc/Max (PDB 1NKP),[5] Glucocorticoid Receptor (PDB 1R40),[6] NF-κB (PDB 1LE5),[7] Lef-1 (PDB 2LEF),[8] TBP (PDB 1TGH),[9] Oct4/Sox2 (PDB 1O4X).[10]

1.3 Molecular recognition of DNA

Deoxyribonucleic acid (DNA):

A single strand of DNA consists of four different bases linked by a phosphodiester deoxyribose sugar backbone. Two strands of DNA intertwine to form a double helix associated by hydrogen bonds between the Watson-Crick base pairs such that thymine (T) pairs with adenine (A), and cytosine (C) pairs with guanine (G) (**Figure 1.2**). In a DNA double helix the direction of the nucleotides in one strand is opposite to their direction in the other strand, thus the strands are antiparallel. The helix of B-form DNA, the average conformation adopted by the majority of biologically active DNA sequences, is right handed

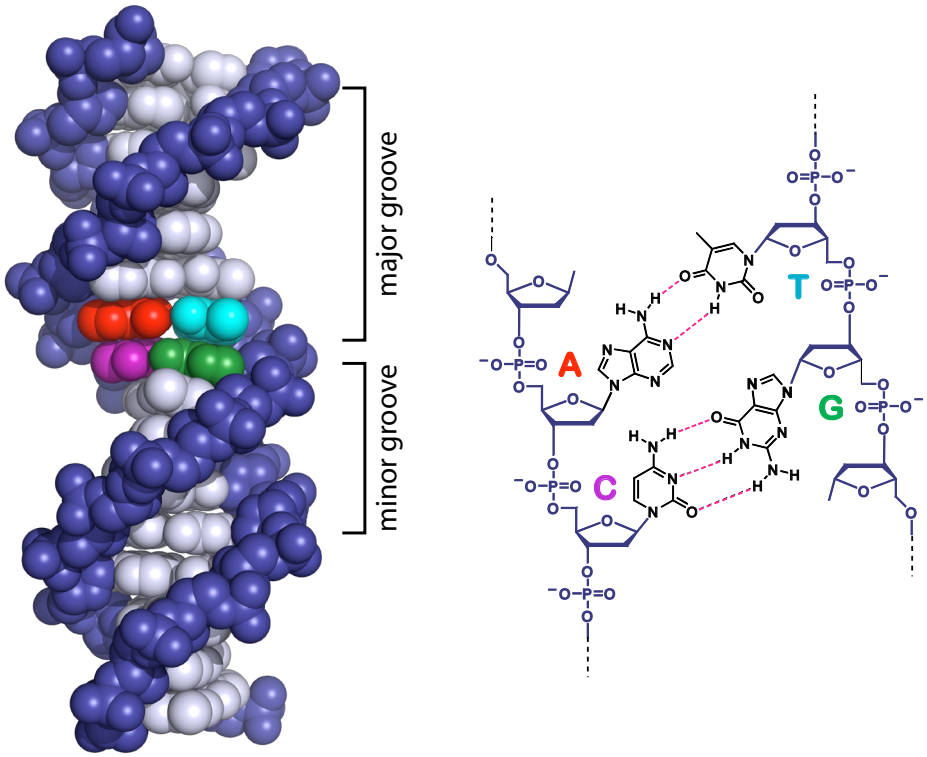


Figure 1.2 The structure of DNA

Left) X-ray crystal structure of B-form DNA. The phosphodiester-linked deoxyribose backbone is shown in blue, and the Watson-Crick base pairs are shown in gray (PDB 1BNA).[11] *Highlighted: adenine (red) thymine (blue) cytosine (magenta) and guanine (green). Right) Chemical structures of phosphodiester-linked hydrogen-bonded base pairs. Adenine (A) is bonded to thymine (T) and cytosine (C) is bonded to guanine (G) Dashed lines indicate hydrogen bonds.*

and displays ten base pairs per turn with the plane of each hydrogen-bonded base pair lying perpendicular to the helical axis. This helix is not symmetrical, instead a wide major groove and a narrow minor groove line the helix and DNA sequences can be distinguished by the pattern of functional groups, e.g., hydrogen bond donors and acceptors, displayed on the edges of the base pairs in either groove.

Sequence-specific recognition of DNA:

Proteins and small molecules can recognize B-form DNA through interactions with the major groove, minor groove, and phosphate backbone, or a combination of these elements. These interactions can be mediated through electrostatics, hydrogen bonding, van der Waals interactions, and, in the case of intercalators, via base pair stacking. The DNA base pair edges in the major groove and minor groove provide a sequence-specific array of functionality for hydrogen bonding, hydrophobic interactions, and steric complementarity with proteins and small molecule binders.[12-16] It is these patterns of hydrogen donors and acceptors that allow for sequence-specific recognition of DNA by proteins and small

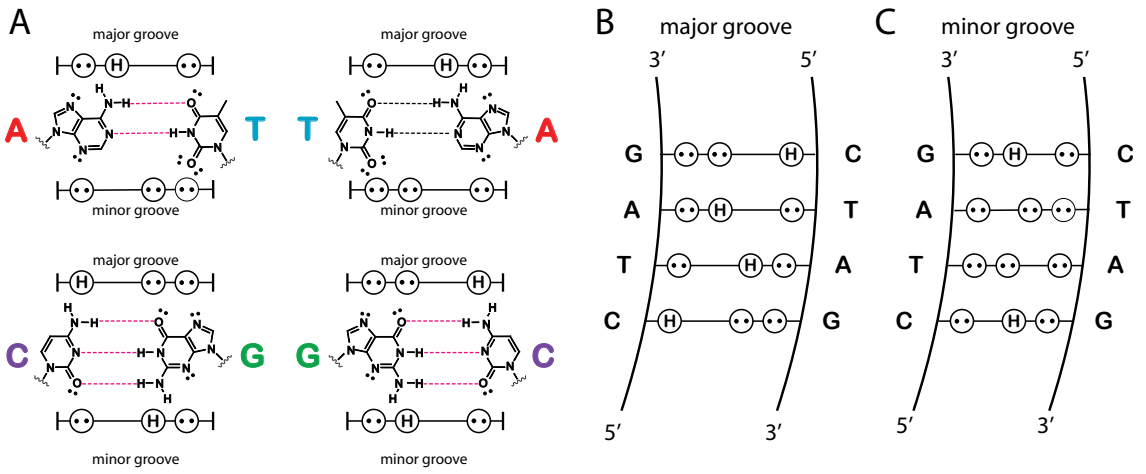


Figure 1.3 Major and minor groove hydrogen bonding patterns of the four Watson-Crick base pairs

Circles with dots represent lone pairs on purine N3 or pyrimidine O2, and circles with an H represent the 2-aminogroup of guanine (G-NH2). A)Major and minor groove hydrogen bonding patterns of each base pair B) Hydrogen bonding pattern offered to the major groove by the sequence 5'-GTAC-3'. C)Hydrogen bonding pattern offered to the minor groove by the sequence 5'-GTAC-3'

molecules that can form hydrogen bonds with the functional groups present in the grooves. In a formal sense, the four Watson-Crick base pairs can be differentiated on the groove floor by the specific positions of hydrogen bond donors and acceptors, as well as by subtle differences in molecular shape. **Figure 1.3** depicts the differences in these hydrogen bond donors and acceptors as seen on the floor of each groove.

1.4 Sequence-specific DNA-binding small molecules

In addition to DNA-binding proteins such as the transcription factors described above, there are also small molecule natural products that recognize and bind specific DNA sequences. Four particular examples are shown in **Figure 1.4**: calicheamicin oligosaccharide, chromomycin, actinomycin D, and distamycin A. Calicheamicin oligosaccharide has been shown to recognize and bind the minor groove monomerically at 5'-TCCT-3' sequences.[17] Chromomycin targets the sequence 5'-GGCC-3' and binds in the minor groove of DNA in a 2:1 ligand:DNA stoichiometry. Chromomycin's biological

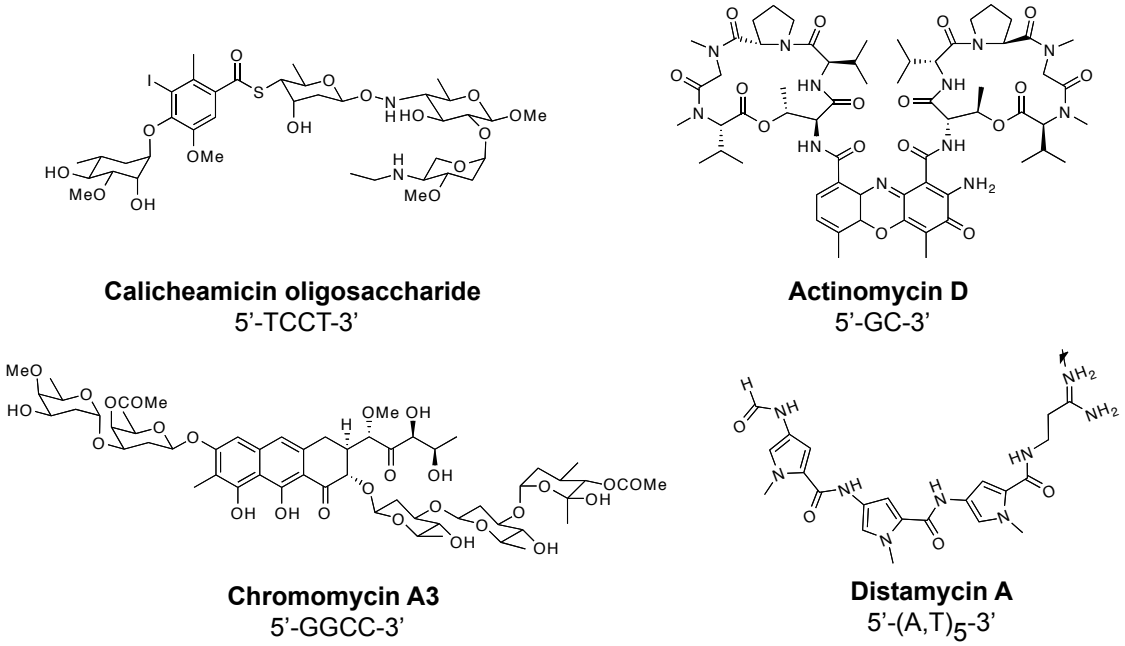


Figure 1.4 Chemical structures of DNA-binding small molecule natural products

activity has been attributed to interference of replication and transcription.[18] Actinomycin D intercalates DNA preferentially at 5'-GC-3' sequences in a 1:1 ligand:DNA stoichiometry. Actinomycin D is known to inhibit transcription and potentially DNA replication and has been used as a chemotherapeutic as well.[19,20] Distamycin A is an A,T-binding oligopeptide of three N-methylpyrrole (Py) carboxamide units. As can be seen in **Figure 1.5**, X-ray and NMR structural studies of Distamycin A reveal that this crescent-shaped molecule can bind DNA in either a 1:1 or 2:1 stoichiometry relative to DNA. The 2:1 complex forms in an antiparallel orientation and results in expansion of the minor groove relative to the 1:1 ligand-DNA complex.[21-23]

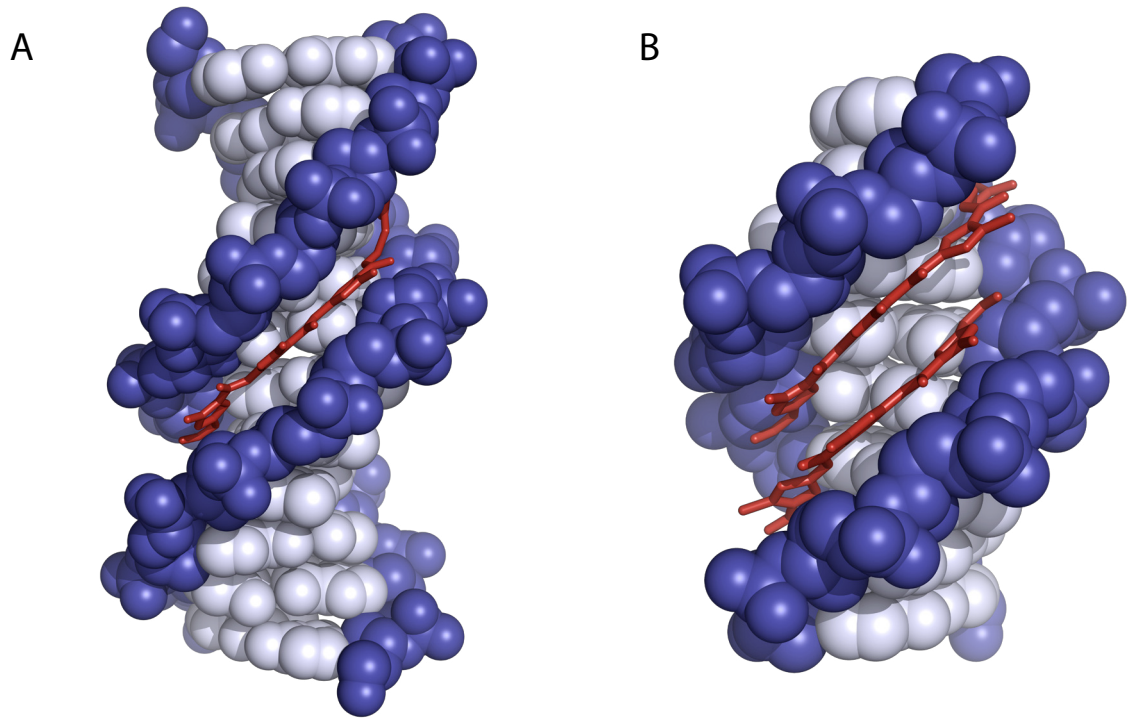


Figure 1.5 X-ray crystal structures of distamycin bound to DNA

A) Distamycin bound with a 1:1 ligand:DNA stoichiometry to the sequence 5'-CGCAAATTGCG' (PBD 2DND) B) Distamycin bound with a 2:1 stoichiometry to the sequence 5'GTATATAC-3'(PDB 378D)

The simple A,T binding natural product distamycin A has evolved over the past two decades into a new class of programmable heterocyclic oligomers that demonstrate high affinity and sequence specificity for the DNA minor groove.[12,13] Incorporation of alternative heterocycles such as imidazole (Im) or hydroxypyrrole (Hp) expanded the sequence-recognition capabilities of polyamides and a set of pairing rules has been developed to allow for the programmable targeting of desired DNA sequences. Sequence-specific recognition of the minor groove of DNA by polyamides arises from the pairing of three different aromatic amino acids, pyrrole (Py), imidazole (Im), and hydroxypyrrole (Hp).[24-26] The targeted binding surface of a crescent-shaped polyamide may be programmed by the incremental change of atoms on the corners of the ring pairs presented to the DNA minor groove floor. Stabilizing and, importantly, destabilizing interactions with the different edges of the four Watson-Crick bases are modulated by shape complementarity and specific hydrogen bonds.[27-29] An Im/Py pair distinguishes G•C from C•G, T•A, and A•T, and likewise a Py/Im pair distinguishes C•G from G•C, T•A, and A•T. Im presents a lone pair of electrons to the DNA minor groove and can accept a hydrogen bond from the exocyclic amine of guanine.[5] Additionally, the Hp/Py pair distinguishes T•A from A•T, G•C, and C•G.[4–6] Hp projects an exocyclic OH group toward the minor groove floor that is sterically accommodated in the cleft of the T•A base pair, preferring to lie over T not A.[25]

As can be seen in **Figure 1.6,** in addition to developing the pairing rules, another key step of the evolution of Im/Py polyamides is the covalent linkage of the two antiparallel heterocyclic strands by a gamma amino butyric acid (GABA) unit, forming a “hairpin” polyamide, demonstrating a 100-3600-fold increase in affinity relative to to the unlinked homodimeric motif.[30,31] Additionally, the incorporation of the turn linkage in the form of a GABA or substituted GABA turn allows the incorporation of unsymmetrical ring pairs for the targeting of non-palindromic DNA sequences.[32]

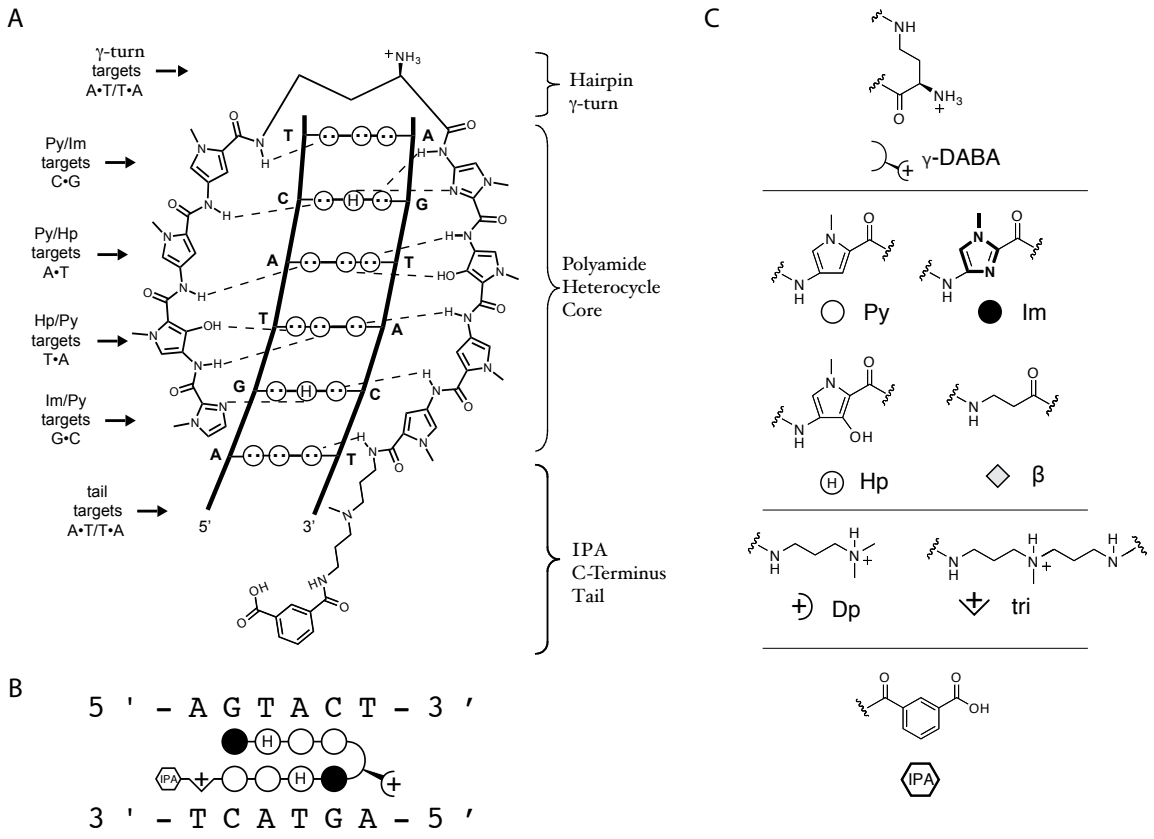


Figure 1.6 Recognition of the DNA minor groove by polyamides

A)Schematic illustration of the polyamide ImHpPyPy-γ-ImHpPyPy-β-Dp binding its target sequence, 5'-AGTACT-3'. Putative hydrogen bonds are shown as dashed lines. Hairpin turn and tail moieties, as indicated in illustration, tolerate T•A and A•T base pairs. B) A ball-and-stick representation of the polyamide shown in a. c) ball-and-stick symbols as well as abbreviations are listed below the chemical structures. This nomenclature will be used throughout this dissertation.

1.5 Polyamide inhibition of gene regulation in live cell culture

Given the fact that transcription is the key first step in gene expression, and the fact that the binding of *trans*-acting factors to promoter elements is critical for transcription, a key route to regulating gene expression lies in controlling the activity of these *trans*-acting factors. The programmability of Py-Im polyamides combined with the subnanomolar increases in affinity achieved by linking the two oligomeric strands provides polyamides with affinity competing with and often rivaling that of endogenous DNA-binding proteins.[12,13] By using Py-Im polyamides to displace or prevent the binding of

transcription factors to their respective promoter DNA sequences, it is possible to modulate the expression of particular genes. Critical to this goal is the ability of polyamides to enter a live cell and permeate the nucleus in order to bind DNA and successfully regulate gene expression. Confocal microscopy studies have confirmed the positive nuclear uptake profiles of a variety of polyamide-fluorophore conjugates in a panel of cell lines.[33,34] Additionally, the presence of an isophthalic acid (IPA) moiety in the tail region has been shown to yield high-affinity conjugates with improved nuclear permeability.[35]

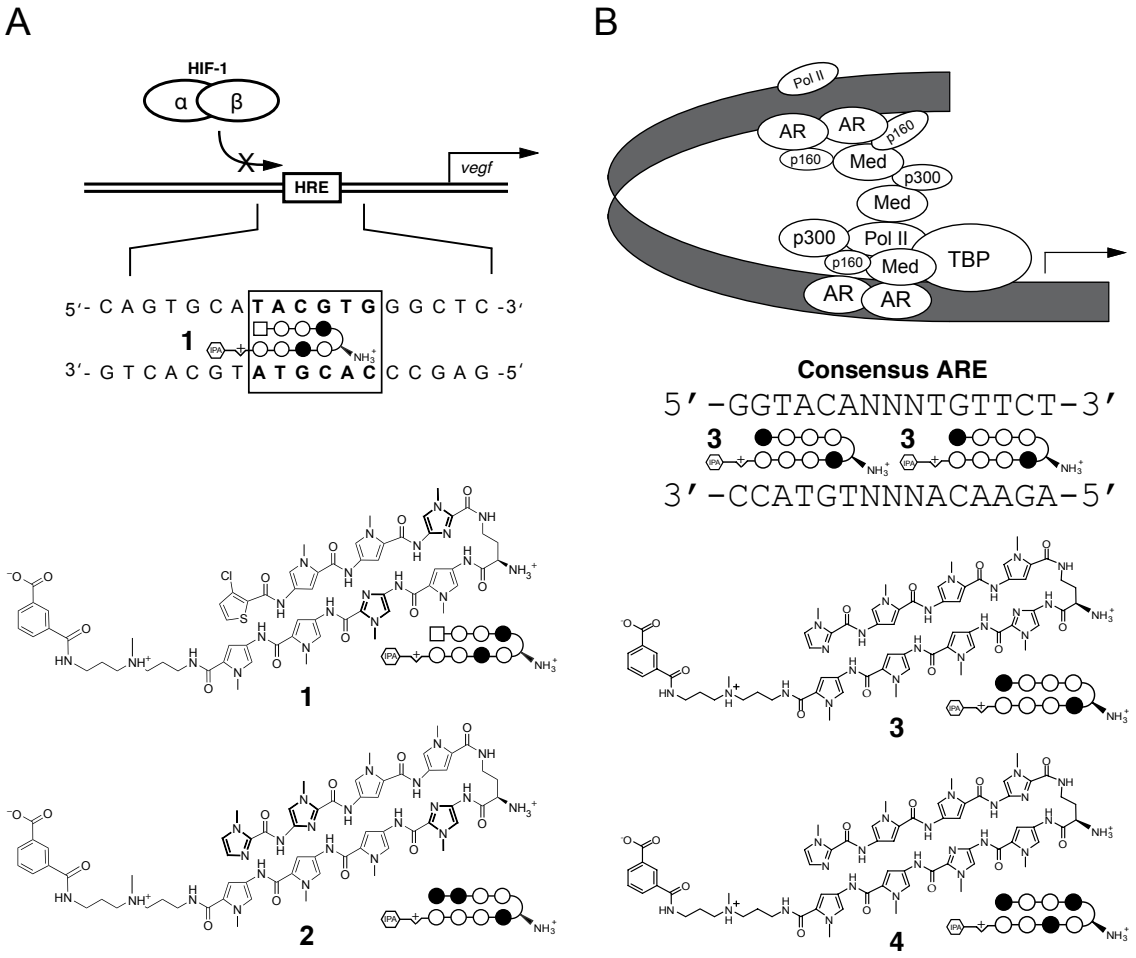


Figure 1.7 Polyamides as regulators of gene expression in cell culture

*A) Schematic diagram of the VEGF promoter showing inhibition of HRE binding by HIF-1 (shown as HIF-1α/HIF-1β heterodimer), binding sequence of the HRE enhancer shown with match polyamide **1**, and chemical structures and ball-and-stick models of match polyamide **1** and mismatch polyamide **2**. B) Schematic diagram of the androgen receptor transcription complex, binding sequence of the consensus ARE targeted by match polyamide **3**, and chemical structures and ball-and-stick models of match polyamide **3** and mismatch polyamide **4**.*

The culmination of the technological advances in this field has resulted in high affinity, sequence-specific, cell-permeable Py-Im hairpin polyamides that have successfully been utilized in several instances to modulate gene expression in cell culture. In a seminal example, a polyamide designed to bind to the hypoxia response element (HRE) was shown to disrupt the binding of hypoxia-inducible factor (HIF) to the HRE. Polyamide treatment was shown to decrease the transcription of vascular endothelial growth factor (VEGF) in cultured HeLa cells.[35,36] In another example, a polyamide designed to target the androgen response element (ARE) has been shown to downregulate prostate-specific antigen (PSA) and other androgen responsive genes in prostate cancer cells.[37]

The hypoxia inducible factor 1α (HIF-1α) transcription factor drives the expression of many genes in response to a low oxygen environment. HIF-1α recognizes and binds a consensus sequence 5'-TACGTG-3' termed the Hypoxic Response Element (HRE). Polyamide **1** (**Figure 1.7**A) targets a subset of known HREs. One particular gene whose modulation is particularly interesting is vascular endothelial growth factor (VEGF), a gene responsible for the vascularization of tumors. When HeLa and U251 cells were treated with micromolar concentrations of polyamide **1**, a reduction in the deferoxamine (DFO) induction of VEGF gene was noted, while dosing with a non-HRE binding polyamide **2** did not produce a statistically significant change in gene expression. On a global scale, microarray experiments have shown that polyamide **1** downregulates a subset of genes upregulated by DFO induction. Furthermore, chromatin immunoprecipitation (ChIP) experiments demonstrated a reduced occupancy of HIF-1α at the VEGF HRE in the presence of polyamide **1.** [35,36,38]

The androgen receptor (AR) transcription factor binds as a homodimer to the androgen response element (ARE), 5'-GGTACAnnnTGTTCT-3', in response to induction by steroid hormones such as testosterone. AR-regulated gene expression is critical in the development and progression of prostate cancer. One key AR-regulated gene is prostate specific antigen (PSA),

a well-studied marker gene that correlates highly with the presence of prostate cancer. LnCap (prostate cancer) cells treated with a Py-Im hairpin polyamide **3** (**Figure 1.7**B) targeted to the consensus ARE half-site 5'-WGWWCW-3' demonstrate a suppressed dihydrotestosterone (DHT) induction of PSA mRNA transcription. This gene regulation modulation is not seen with treatment by a mismatch control polyamide targeting 5'-WGWCGW-3'. Microarray experiments assessing the global mRNA transcription in LnCap cells when treated with **3** and **4** indicate that **3** is able to disrupt induction by DHT for a subset of the DHT induced genes in a manner that is differential from treatment by **4** and likely is sequence-specific. ChIP experiments on **3** and **4** again suggest the disruption of a protein-DNA interface as a potential mechanism for polyamide activity.[37]

1.6 An allosteric model for polyamide inhibition of steroid hormone receptors

Allosteric modulation of protein-DNA interactions

Both HIF-1α and AR are major groove-binding proteins, yet minor groove-binding Py-Im polyamides are able to successfully disrupt protein DNA-binding and alter the gene regulation of the given transcription factor. A likely model for this inhibition is that allosteric modulation of the DNA occurs upon polyamide binding that renders the major groove of the DNA no longer capable of binding the transcription factor.[39,40] The process of promoter recognition and utilization involves a stepwise interaction of a complex series of transcription factors with the promoter to create a stable DNA-protein complex that allows RNA polymerase to initiate transcription. There are as many as 100,000 protein-encoding genes in the mammalian genome, and rather than generate a unique transcription factor to regulate each gene, nature appears to have developed a limited number of transcription factors responsible for DNA recognition and that the high degree of specificity demonstrated is generated by specific protein-protein interactions that stabilize otherwise weak interactions on a promoter. Protein-DNA interactions are generally fairly

weak and will readily dissociate.[41] The capacity to stabilize this otherwise weak interaction is likely a critical aspect of transcription control in which multiple factors must interact on the DNA to stabilize a functional complex.[4]

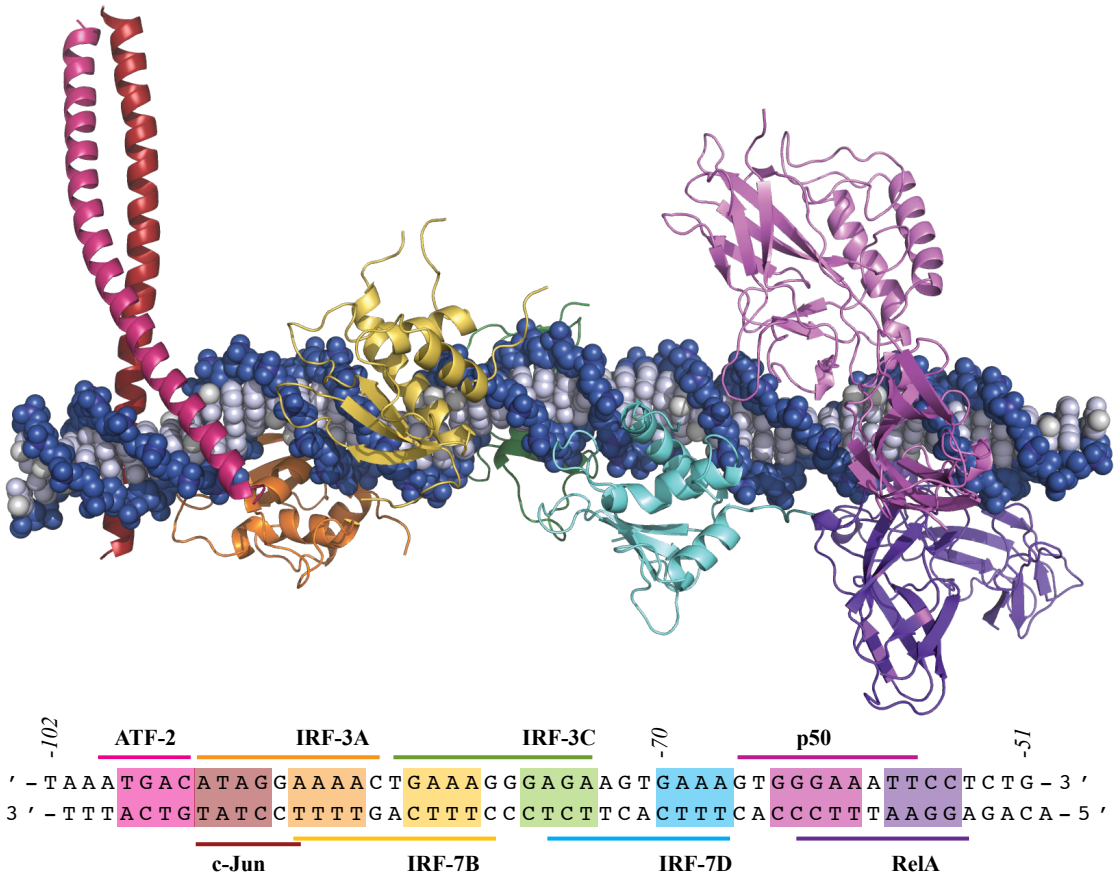


Figure 1.8 Atomic model of the cooperative assembly of interferon-β enhancesome

A composite model of allosterically driven protein–DNA recognition created from overlayed X-ray crystal structures (PDB 2O6G, 2O61) showing 4-6 base pair transcription factor binding sites along the highly conserved composite DNA interface of *55 base pairs spanning approximately 160 Å in length.*

Transcription factors often posess domain that is critical for mediation of protein-protein interactions with other components of the transcriptional machinery. In addition, transcription factors are also known to communicate indirectly through allosteric modulation of DNA resulting in cooperative assembly with very little direct protein-protein interaction. In this model, the sequence-specific binding of a transcription factor induces

perturbations in the DNA that modulate the binding of the next transcription factor. An elegant example of allosterically modulated protein–DNA specificity on a gene enhancer is the interferon-beta enhanceosome (**Figure 1.8**). In this protein complex, eight different transcription factors (ATF-2/c-Jun, IRF-3A, IRF-7B, IRF-3C, IRF-7D, p50, and RelA) cooperatively assemble on the enhancer, yet there are no protein-protein contacts between any of the proteins on this DNA sequence. Thus, it has been proposed that structural alterations to the DNA, such as widening or narrowing of the major or minor groove by individual protein–DNA interactions, create optimum binding shape and structure for other proteins, in a cooperative interaction.[42] Each transcription factor binds four to eight base pairs and inhibition of the binding of any one of the proteins may result in interruption of the transcriptional activation activity of the protein complex as a whole.

Allosteric modulation of DNA by Py-Im polyamides

A recent high-resolution X-ray crystal structure of a β-amino turn-linked eight-ring cyclic Py-Im polyamide bound to the central six base pairs of the steroid hormone receptor consensus sequence reveals that significant modulation of DNA shape occurs upon polyamide binding. Cyclic polyamide **5**, comprised of two antiparallel ImPyPyPy strands capped by (*R*)-β-amino-γ turn units codes for the sequence 5′-WGWWCW-3’. The high-resolution crystal structure of the polyamide in complex with the 10 base pair DNA oligonucleotide sequence 5′-CCAGTACTGG-3′, containing an ARE/GRE consensus DNA sequence demonstrates that structural alterations of DNA by these major groove-binding proteins and minor groove-binding cyclic polyamides operate in opposite directions.[40]

Binding of cyclic polyamide **5** induces significant (>4 Å) widening of the DNA minor groove and compression of the major groove by more than 4 Å as compared to unliganded DNA (**Figure 1.9** C, D).[43] Additionally, polyamide binding induces bending of

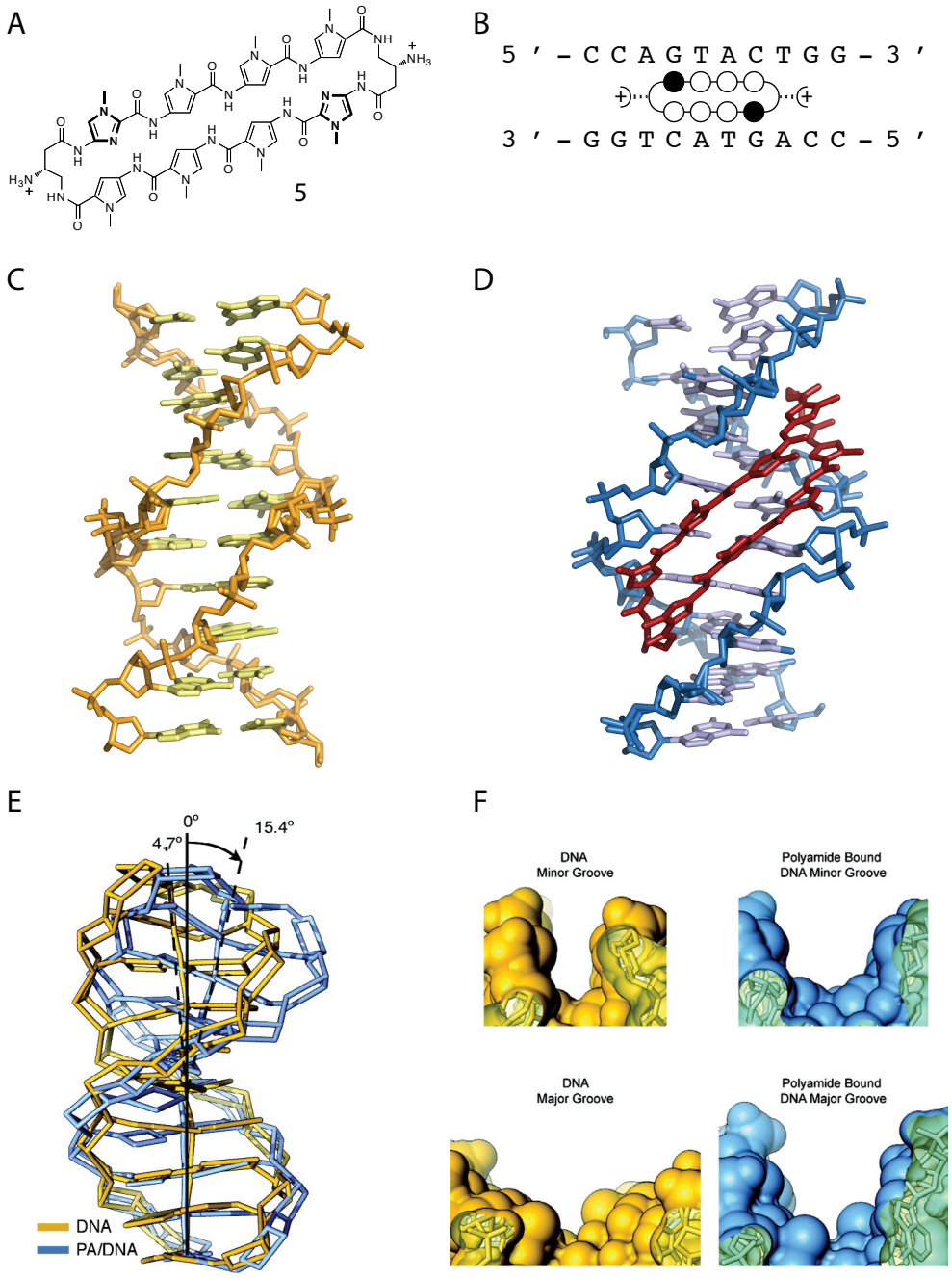


Figure 1.9 Comparison of native DNA to polyamide/DNA complex

A) chemical structure of cyclic polyamide 5 which targets 5'WGWWCW-3' B) ball-and-stick model superimposed over the binding site on the dsDNA oligonucleotide sequence used for crystallization.C) Native DNA crystal structure at 0.98 Å resolution (PDB 1D8G) D) DNA/polyamide co-crystal structure at 0.95 Å resolution. (PDB 3OMJ) E) Significant DNA bending is observed for polyamide-bound DNA (blue) versus unbound DNA (yellow). F) top: Comparison of the minor-groove width for DNA in the absence of polyamide (yellow) and in the presence of bound polyamide (blue). bottom: Comparison of the major-groove width for DNA in the absence of polyamide (yellow) and in the presence of bound polyamide (blue). In E and F, polyamide has been removed from the blue complex for clarity.[39,40]

the DNA helix by >15° towards the major groove, resulting in major groove compression (**Figure 1.9**E). A slice through the short axis of the DNA helix, showing the minor and major groove geometry at the center of the polyamide binding site for uncomplexed and complexed DNA shows this distortion clearly (**Figure 1.9**F).[40]

Ultimately, the perturbation in the major-groove geometry that occurs upon polyamide binding converts the wide, shallow surface of the major groove from a functionally exposed protein recognition domain to a narrow, deep cleft too small to accommodate the width of a standard protein α-helical domain or β-sheet from a transcription factor. A detailed analysis of the structure of DNA bound by the androgen receptor and the related glucocorticoid receptor reveals that the cyclic polyamide **5** is an allosteric modulator that perturbs the DNA structure in such a way that nuclear receptor protein binding is no longer compatible. This allosteric perturbation of the DNA helix provides a molecular basis for disruption of transcription factor−DNA interfaces by sequence-specific DNA binding polyamides.[39,40]

1.7 Scope of this work

The work presented in this thesis is focused on the further development of Py-Im polyamides as tools for understanding the mechanisms behind gene regulation. By perturbing gene regulation in endogenous systems we can begin to understand the molecular basis for the difference between normal gene function and those alterations in gene control events that underlie certain disease states. In Chapter 2, we utilize a Py-Im polyamide targeted to bind 5'-WGWWCW-3' to probe the dual mechanism of transcriptional action displayed by the glucocorticoid receptor (GR) transcription factor. This sequence-specific small molecule probe is used to separate and identify a list of genes that are regulated by a protein-DNA interaction from those genes that are generally regulated by the GR in both DNA-binding-dependent and DNA-binding-independent mechanisms. In Chapter 3 we utilize polyamides to validate a new microarray-based tool for use in determining the DNA-sequence binding preferences of small molecules and proteins and in doing so examine the binding preferences of two polyamides that have been demonstrated to modulate gene expression in cell culture. The binding preferences for the polyamides assayed validate the polyamide pairing rules in an unbiased fashion. Chapter 4 presents progress towards divesting a group of genes regulated by the Myc-Max heterodimeric transcription factor. Our goal is to regulate genes bound at 5'-CACGTG-3' Myc-Max heterodimer binding sites while leaving Max-Max homodimer binding sites unbound. We anticipate that this will modulate the expression of Myc-regulated genes without inducing a Myc-upregulation feedback loop. In this chapter, a library of small molecules is developed to target the binding site, demonstrate that they bind with high affinity and specificity to the targeted site, and have positive uptake properties in a variety of live cells. While modulation of gene expression was not demonstrated in these cell lines, work is ongoing to probe this system further.

1.8 References

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

(2) Lander, E. S.; Linton, L. M.; Birren, B.; Nusbaum, C.; Zody, M. C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W. *Nature* **2001**, *409*, 860.

(3) Venter, J. C.; Adams, M. D.; Myers, E. W.; Li, P. W.; Mural, R. J.; Sutton, G. G.; Smith, H. O.; Yandell, M.; Evans, C. A.; Holt, R. A. *Science* **2001**, *291*, 1304.

(4) Darnell, J. E. and Lodish, H. F. *Molecular cell biology, 4th Ed.*; W.H. Freeman: New York, 2000.

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

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

(7) Berkowitz, B.; Huang, D. B.; Chen-Park, F. E.; Sigler, P. B.; Ghosh, G. *J. Biol. Chem*. **2002**, *277*, 24694.

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

(9) Luisi, B. F.; Xu, W. X.; Otwinowski, Z.; Freedman, L. P.; Yamamoto, K. R.; Sigler, P. B. *Nature* **1991**, *352*, 497.

(10) Nair, S. K.; Burley, S. K. *Cell* **2003**, *112*, 193.

(11) Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. *Proc. Natl. Acad. Sci USA* **1981**, *78*, 2179.

(12) Dervan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 2215.

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

(14) Rice, P. A.; Correll, C. C. *Protein-nucleic acid interactions: structural biology*; Royal Society of Chemistry, **2008**.

(15) Waring, M. J. and Wakelin, L. P. G. In *DNA and RNA binders: from small molecules to drugs*; Wilson, W. D.; Bailly, C.; and Demeunynck, M., Editor.; Wiley-VCH Verlag GmbH, 2003; p. 1, 1-17.

(16) Waring, M. J.; Britain, R. S. O. C. *Sequence-specific DNA binding agents*; RSC Publishing, 2006.

(17) Bifulco, G.; Galeone, A.; Nicolaou, K. C.; Chazin, W. J.; Gomez-Paloma, L. *J. Am. Chem. Soc.* **1998**, *120*, 7183.

(18) Hou, M. H.; Robinson, H.; Gao, Y. G.; Wang, A. H. *Nucleic Acids Res* **2002**, *30*, 4910.

(19) Hou, M. H.; Robinson, H.; Gao, Y. G.;Wang, A. H. *Nucleic Acids Res* **2004**, *32*, 2214.

(20) Kamitori, S.; Takusagawa, F. *J. Am. Chem. Soc.* **1994**, *116*, 4154.

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

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

(23) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci USA* **1989**, *86*, 5723.

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

(25) Urbach, A. R.; Szewczyk, J. W.; White, S.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 11621.

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

(27) Hays, F. A.; Teegarden, A.; Jones, Z. J.; Harms, M.; Raup, D.; Watson, J.; Cavaliere, E.; Ho, P. S. *Proc. Natl. Acad. Sci USA* **2005**, *102*, 7157.

(28) Steitz, T. A. *Q Rev. Biophys.* **1990**, *23*, 205.

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

(30) Herman, D. M.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 1121.

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

(32) Herman, D. M.; Baird, E. E.; and Dervan, P. B. *J. Am. Chem. Soc.* **1998**, *120*, 1382.

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

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

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

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

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

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

(39) Chenoweth, D. M.; Dervan, P. B. *Proc. Natl. Acad. Sci USA* **2009**, *106*, 13175.

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

(41) Stormo, G. D. Zhao, Y. *Nat Rev Genet* **2010**, *11*, 751.

(42) Panne, D.; Maniatis, T.; Harrison, S. C. *Cell* **2007**, *129*, 1111.

(43) Kielkopf, C. L.; Ding, S.; Kuhn, P.; Rees, D. C. *J. Mol. Biol.* **2000**, *296*, 787.