

## **Chapter 1: Introduction to Molecular Recognition of DNA**

## 1.1 Background and Significance

The field of molecular recognition has come a long way since the organic solvent based host-guest chemistry of Lehn and Cram (crown ether cation complexes).<sup>1-3</sup> Understanding in a predictive and mechanistic sense the molecular recognition between synthetic ligands and biological macromolecules in water is fundamental to understanding biochemical processes and cellular composition.<sup>4</sup> The overall free energy of these complexes often includes a superposition of non-covalent forces such as hydrogen bonding interactions, dipole-dipole, induced dipole, cation- $\pi$ , lone pair- $\pi$ , and van der Waals interactions in addition to hydrophobic effects. Understanding the intimate interplay of these forces and their contributions to the overall free energy of a host-guest system has remained one of the ultimate challenges in chemistry and biology. The molecular recognition processes involved in nucleic acid-drug and nucleic acid-protein interactions are similar with both being driven by the hydrophobic effect, a phenomenon which is still not well understood. As ligand-receptor recognition proceeds, the optimization of multiple forces ensues including minimization of water exposed hydrophobic surfaces and simultaneous maximization of van der Waals interactions. Additionally, optimization of all buried hydrogen-bond donor and acceptor pairings including solvent-assisted and counterion charge neutralization contribute to the complex recognition event. Intimate structural and biophysical knowledge of these processes is fundamental to the understanding of nature at the molecular level.

The DNA double helix, in addition to being the molecular storage unit of genetic information, represents one of the ultimate challenges in aqueous based molecular recognition. Over billions of years, nature has used selection to evolve protein surfaces that recognize DNA in a cooperative and combinatorial fashion allowing for the stringent regulation of the molecular processes crucial to all living organisms on earth. Prior to the 1960s, histologists and cell biologists realized that certain small molecules could interact specifically with cell nuclei.<sup>5</sup> Dye molecules such as aminoacridines were regularly used for staining tissues and cells and it was recognized that specificity for different nucleic acid structures could be obtained using different dyes. However it was not until the 1960s that a formal DNA drug binding hypothesis would be formulated. The “intercalation hypothesis” formulated by Leonard Lerman (a graduate student of Linus Pauling at Caltech) in 1961, working at the Cambridge MRC laboratory, provided the pivotal turning point in the field of drug-nucleic acid recognition.<sup>6</sup> Since the intercalation hypothesis, a plethora of biophysical, biochemical, and biostructural investigations have unveiled the detailed chemistry and biology of many DNA binding drugs, some of which have had a profound impact on human disease (i.e. actinomycin D).<sup>5,7,8</sup> The

intercalating natural product actinomycin D remained one of the most potent chemotherapeutics throughout the 1950s and 1960s along with other nucleic acid binding drugs including cross-linking agents and powerful alkylators, however the first minor groove binding agents would not be discovered until the mid 1960s.<sup>5</sup> Even though Lerman himself relied upon X-ray fiber diffraction data for his intercalation hypothesis in the 1960s, it took another 15-20 years before the first single crystal X-ray structures of drug-nucleic acid complexes (intercalators) would emerge with the seminal work of Sobell, Rich, and Neidle.<sup>9-11</sup> The first X-ray structure of a minor groove binder would not appear until Dickerson's report on the 1:1 structure of netropsin complexed with DNA in 1985.<sup>12</sup> This was soon followed by the 1:1 structure of the distamycin-DNA complex by Rich in 1987.<sup>13</sup> In a seminal study, structural evidence using NMR for the 2:1 binding motif of distamycin was provided by Wemmer in 1989,<sup>14</sup> however the first single crystal X-ray structure of a 2:1 minor groove binding ligand-DNA complex was not realized until the work of Ramakrishnan in 1994 on distamycin A.<sup>15</sup> Since this work many advances have been made in the field of DNA molecular recognition, with minor groove binders representing one of the most promising classes of DNA-binding molecules for targeted transcriptional therapy.

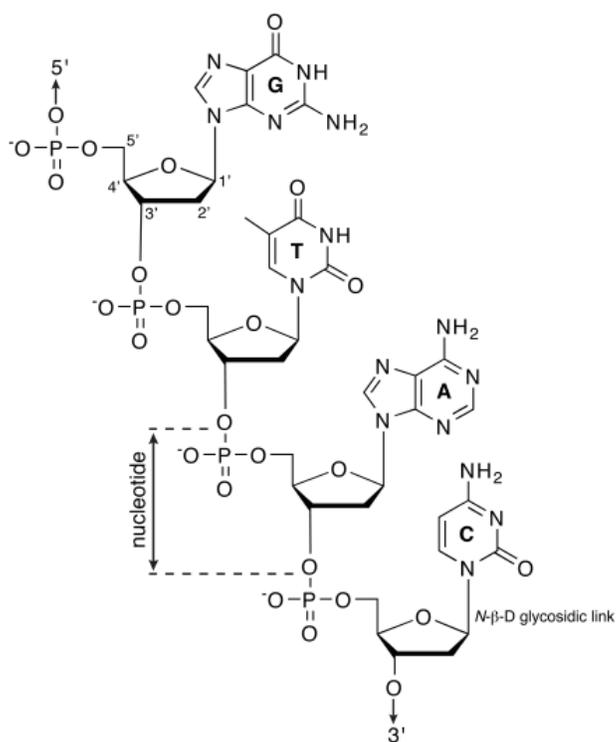
The modulation of gene expression using small molecules has been one of the ultimate goals of nucleic acid molecular recognition. Complex natural products such as actinomycin D, netropsin, and distamycin A have served as inspiration to chemists for the construction of molecular architectures capable of nucleic acid recognition with specificity and affinities equivalent to and rivaling that of endogenous proteins. Initially inspired by the 2:1 binding natural product distamycin, pyrrole-imidazole polyamides have evolved into a modular programmable molecular recognition system capable of specificities and affinities rivaling that of endogenous transcription factors.<sup>16,17</sup> Modulation of transcription factor-DNA interfaces with small molecules such as pyrrole-imidazole polyamides provides a powerful strategy for controlling regulation of the genetic material and could eventually impact human medicine. The future of molecular recognition is poised to benefit greatly from advances in biochemical, biophysical, computational, and structural (X-ray, NMR, EM, Cryo-EM, etc.) methods along with the new tools of physical biology leading to ever increasing resolution and a quantitative understanding of molecular level processes.<sup>4,18</sup>

## 1.2 Nucleic Acid Structure

Deoxyribose nucleic acid (DNA) is the fundamental storage material of genetic information and can be characterized chemically as a hetero-polymer consisting of nucleotide monomers linked

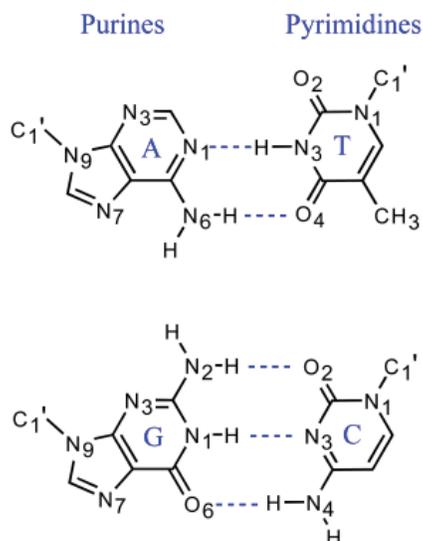
through their sugar-phosphate backbones.<sup>19,20</sup> The 5' and 3' hydroxyl groups of the deoxyribose sugar define the directionality of the DNA strand while a set of four nucleobases [adenine (A), guanine (G), cytosine (C), and thymine (T)] distinguish the nucleotide monomers, providing the fundamental building blocks of the genetic code. Figure 1.1 shows the chemical structure of a short DNA strand containing all four bases and Figure 1.2 shows the atom numbering conventions. Early studies by Chargaff demonstrated that A and T occurred in similar molar ratios as did G and C, which in combination with fiber diffraction data from Rosalind Franklin and Maurice Wilkins would eventually lead to Watson and Crick's base-paired helical model of B-DNA.<sup>21</sup> The Watson-Crick base paired model of DNA contains a set of rules for which A prefers to bind T through two hydrogen bonds and G prefers to bind C through three hydrogen bonds on opposite strands as shown in Figure 1.2 (U replaces T in RNA).<sup>22</sup> The strands are oriented in an antiparallel fashion as they base pair and wind around a central axis. These opposite strands form a double helical structure where the Watson-Crick base pairs are stacked and stabilized by a combination of favorable hydrophobic effects and hydrogen bonding between paired bases. Due to the length of the sugar-phosphate backbone, a helical twist is required to minimize the distance between adjacent base pairs and maximize their hydrophobic stacking.<sup>19,20</sup>

The sugar-phosphate backbone of DNA is highly dynamic allowing for a diverse range



of higher order structures depending on environmental conditions. The torsion angles for the sugar-phosphate backbone are defined in Figure 1.3 and typically vary with ionic strength, pH, sequence, and many other factors.<sup>19,20</sup> In contrast to RNA, where the 2'-hydroxyl of the sugar locks the A-form helix into a fairly rigid structure, the sugar-phosphate backbone of DNA is highly mobile.<sup>19,20</sup> DNA conformation can often be defined by the sugar pucker modes, which by convention are named after the ring atom and either *endo* or *exo* referring to the 5' side of the furanose ring or the 3' side, respectively. Figure 1.3 shows typical

**Figure 1.1** Chemical structure of DNA.



**Figure 1.2** DNA base pairs showing numbering convention for heteroatoms and Watson-Crick base pairing.

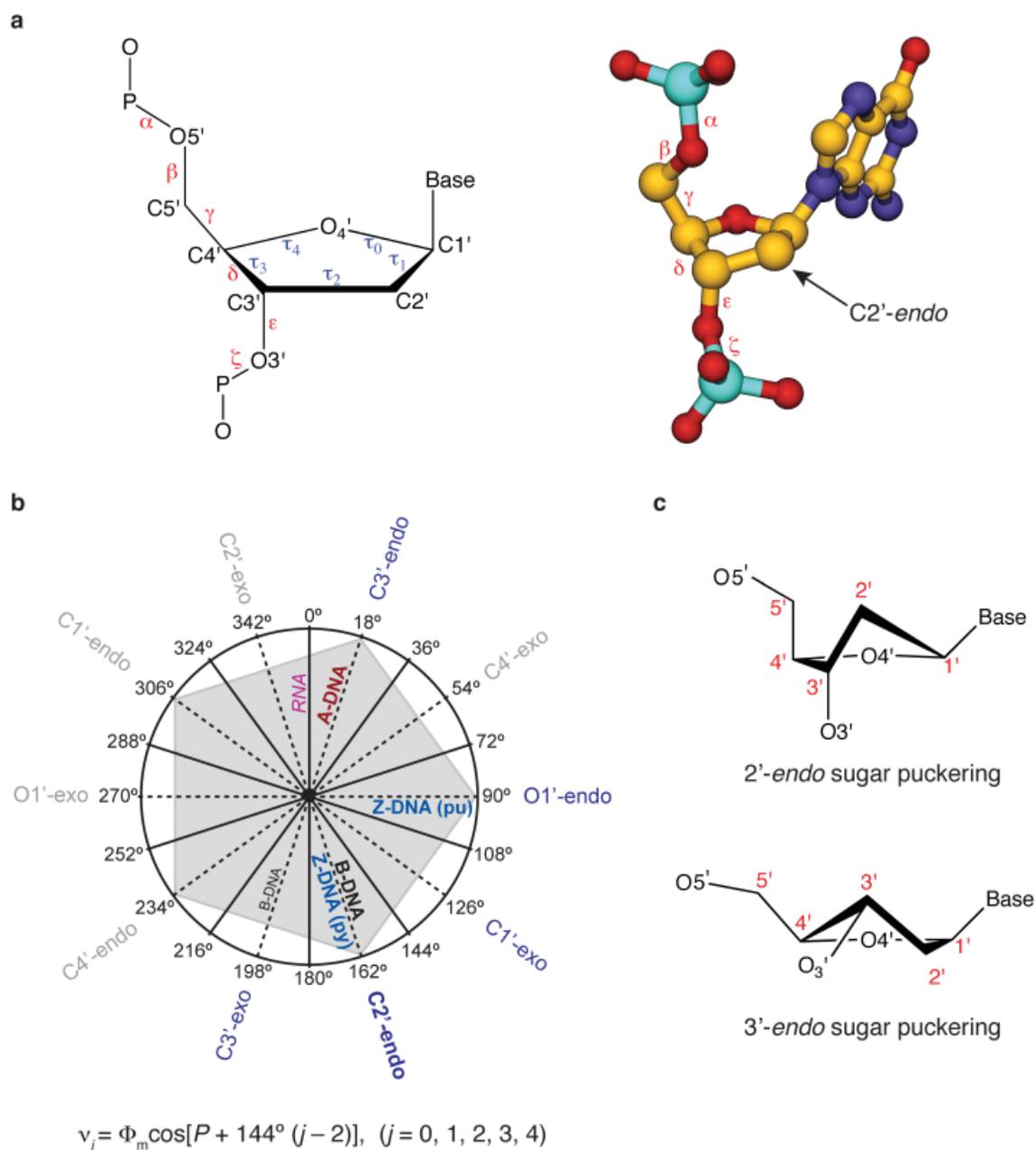
sugar pucker conformations along with their preference in nucleic acid structures. In addition to the sugars displaying conformational preferences, the phosphodiester bond exhibits conformational rigidity analogous to a peptide bond. This conformation rigidity, known as the gauche effect, is a result of stereoelectronic effects from lone pair hyperconjugation/donation of the O3' and O5' oxygen atoms into the  $\sigma^*$  orbital of the P-O5' and P-O3' bonds, respectively.<sup>23</sup> Double helical DNA is a dynamic structure which is capable of forming three primary double strand conformations known as A, B, and Z forms. In contrast to this, double helical RNA is far less flexible with the 2'-OH locking its sugar ring conformation into a C3'-endo pucker resulting in a preference for an A form helix similar to that of A-form DNA. A structural comparison of these ideal DNA polymorphs along with A-form RNA is

shown in Figure 1.4 and Table 1.1.<sup>19,20,23,24.</sup>

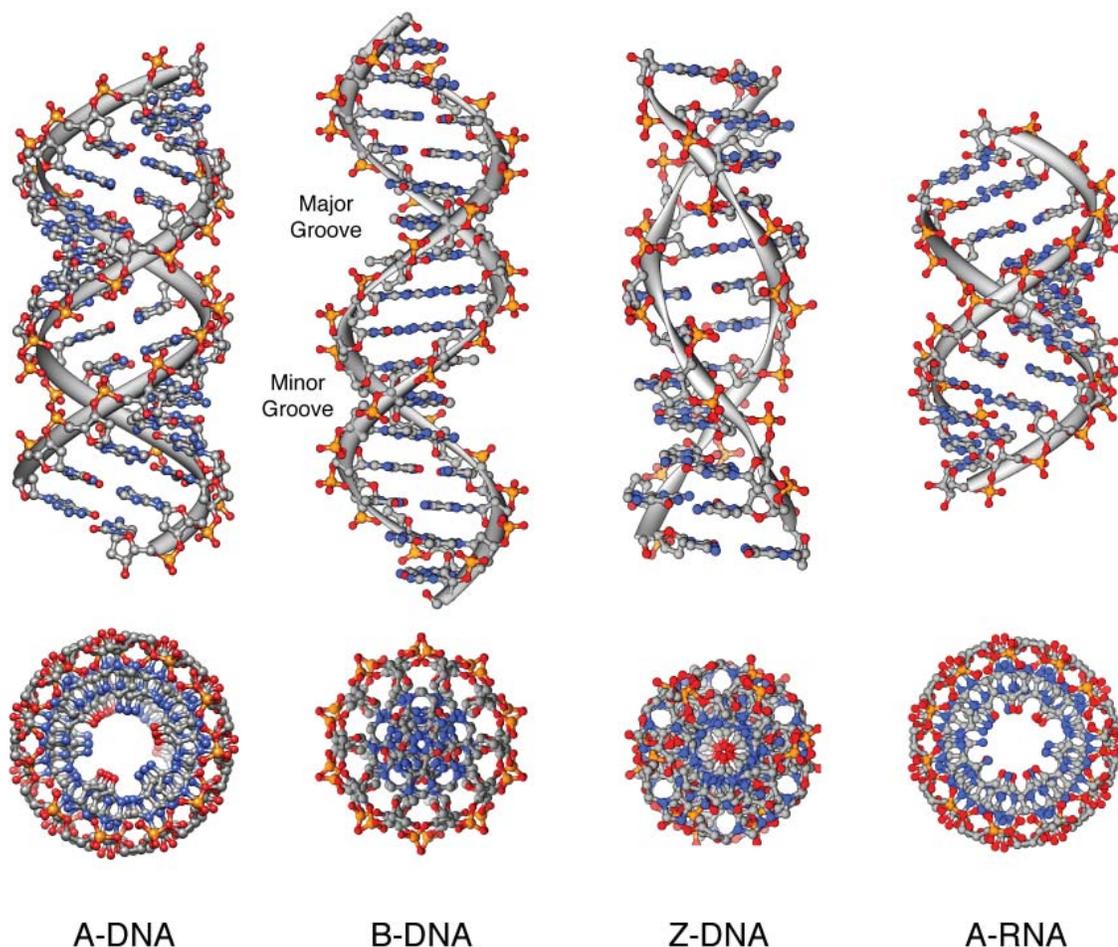
In biological systems, especially eukaryotic cells, DNA is assembled around octameric proteins called histones and compacted into macrohelical fibers forming the high-order structure of chromatin. This DNA-histone complex is called the nucleosome core particle (NCP) and represents the fundamental repeating unit of chromatin consisting of 147 base pairs of DNA forming two super helical turns around the histone octamer with 20-80 base pairs of linker DNA separating one NCP from the next. The Richmond group<sup>25-27</sup> at ETH Zurich has made seminal contributions to elucidate biologically relevant higher-order DNA structures such as the NCP<sup>25,26</sup> and the tetra-NCP<sup>27</sup> presented Figure 1.5. In addition, a theoretical model of four tetra-NCPs assembled into a super-helical chromatin fiber is presented in Figure 1.5. Chromatin architecture and accessibility in biological systems represents a higher-order level of regulation and a profoundly important problem for the field of DNA recognition.

### 1.3 Molecular Recognition of DNA

One of the largest projects in modern science, the human genome project,<sup>28-30</sup> is poised to deliver detailed information and make major impacts in biotechnology and medicine through the physical and functional characterization of the approximately 20,000 to 25,000 genes in the human



**Figure 1.3** DNA sugar phosphate backbone torsion angle map and sugar pucker conventions. a) Chemical structure of DNA sugar phosphate backbone with torsion angles next to the 3-dimensional structure of a nucleoside taken from the native high resolution B-DNA structure solved in Chapter 4 of this thesis. b) Pseudorotation phase angle ( $P$ ) diagram defining 5-membered ring sugar pucker modes. Equation describing the pseudorotation phase angle and maximum torsion angle for 5-membered rings. c) Chemical structure of the most common sugar pucker modes for B-DNA ( $C2'$ -endo) and A-DNA/A-RNA ( $C3'$ -endo).



**Figure 1.4** A comparison of double helical DNA polymorphs and A-form RNA.

genome. These genes are tightly regulated in higher organisms by transcription factor assemblies that function in a concerted cooperative and combinatorial fashion to modulate eukaryotic gene expression. The molecular recognition processes involved in nucleic acid-protein interactions are completely analogous to those of nucleic acid-drug interactions where initial complexation is often driven by the hydrophobic effect. Optimization of the same forces is also required, involving minimization of water exposed hydrophobic surfaces and maximization of van der Waals interactions in conjunction with the optimization of all buried hydrogen bond donor and acceptor pairings (solvent-assisted or counterion charge neutralization).<sup>31</sup> The recognition of the B-DNA interface by proteins and small molecules can occur at the major groove, minor groove, and phosphate backbone, or any combination, with interactions mediated through electrostatics, hydrogen bonding, and van der Waals interactions along with base pair stacking for the case of intercalators. The DNA base pair edges in the major groove and minor groove provide an array

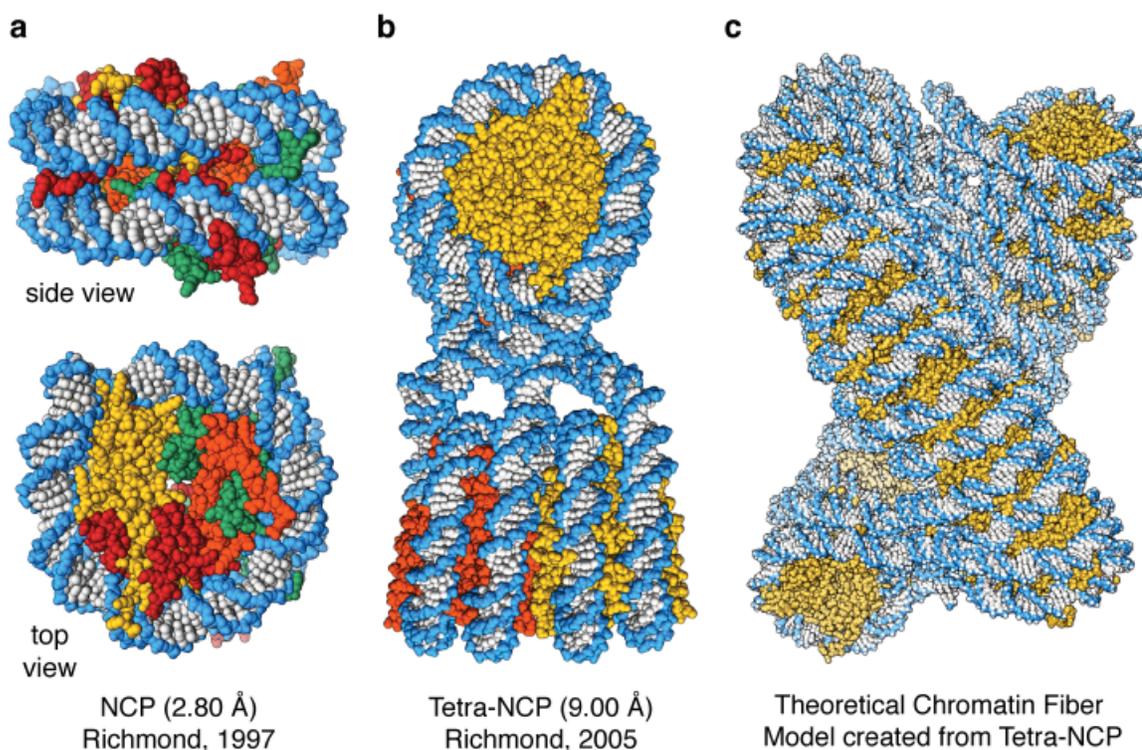
**Table 1.1** Typical nucleic acid structural parameters.

|                         | A-DNA    | B-DNA    | Z-DNA                                 | A-RNA    |
|-------------------------|----------|----------|---------------------------------------|----------|
| Helix Sense             | Right    | Right    | Left                                  | Right    |
| Screw symmetry          | 11-fold  | 10-fold  | 6-fold                                | 12-fold  |
| bp/repeating unit       | 1        | 1        | 2                                     | 1        |
| bp/turn                 | 11       | 10       | 12                                    | 11-12    |
| Helix twist             | 32.7°    | 36.0°    | -10 <sup>a</sup> , -50 <sup>b</sup>   | 33.1°    |
| Rise (Å)                | 2.8      | 3.4      | -3.9 <sup>a</sup> , -3.5 <sup>b</sup> | 2.8      |
| Helix pitch (Å)         | 28       | 34       | 44                                    | 36       |
| Base pair tilt (Å)      | 20       | -6       | 7                                     | 16.7     |
| Diameter of helix (Å)   | 23       | 20       | 18                                    | 21       |
| Rotation per bp         | 33       | 36       | -30                                   | 32.7     |
| Glycosidic bond         |          |          |                                       |          |
| dA, dT, dC              | anti     | anti     | anti                                  | anti     |
| dG                      | anti     | anti     | syn                                   | anti     |
| Sugar pucker            |          |          |                                       |          |
| dA, dT, dC              | C3' endo | C2' endo | C2' endo                              | C3' endo |
| dG                      | C3' endo | C2' endo | C3' endo                              | C3' endo |
| Phosphate-phosphate (Å) |          |          |                                       |          |
| dA, dT, dC              | 5.9      | 7.0      | 7.0                                   | ~5.7     |
| dG                      | 7.0      | 7.0      | 5.9                                   | ~5.7     |
| Major groove            |          |          |                                       |          |
| width (Å)               | 2.7      | 11.7     | Convex                                | 11.1     |
| depth (Å)               | 13.5     | 8.5      |                                       |          |
| Minor groove            |          |          |                                       |          |
| width (Å)               | 11.0     | 5.7      | 4                                     | 11.1     |
| depth (Å)               | 2.8      | 7.5      | 9                                     | shallow  |

<sup>a</sup> CpG step. <sup>b</sup> GpC step.

of functionality for hydrogen bonding, hydrophobic interaction, and steric complementarity with proteins and small molecule binders.<sup>5,8,16,17,31</sup> The molecule electrostatic potential surfaces for the minor and major groove base pair edges are shown in Figure 1.6.<sup>32</sup> In addition, primary driving forces such as the hydrophobic effect and shape complementarity are common to both proteins and small molecules.

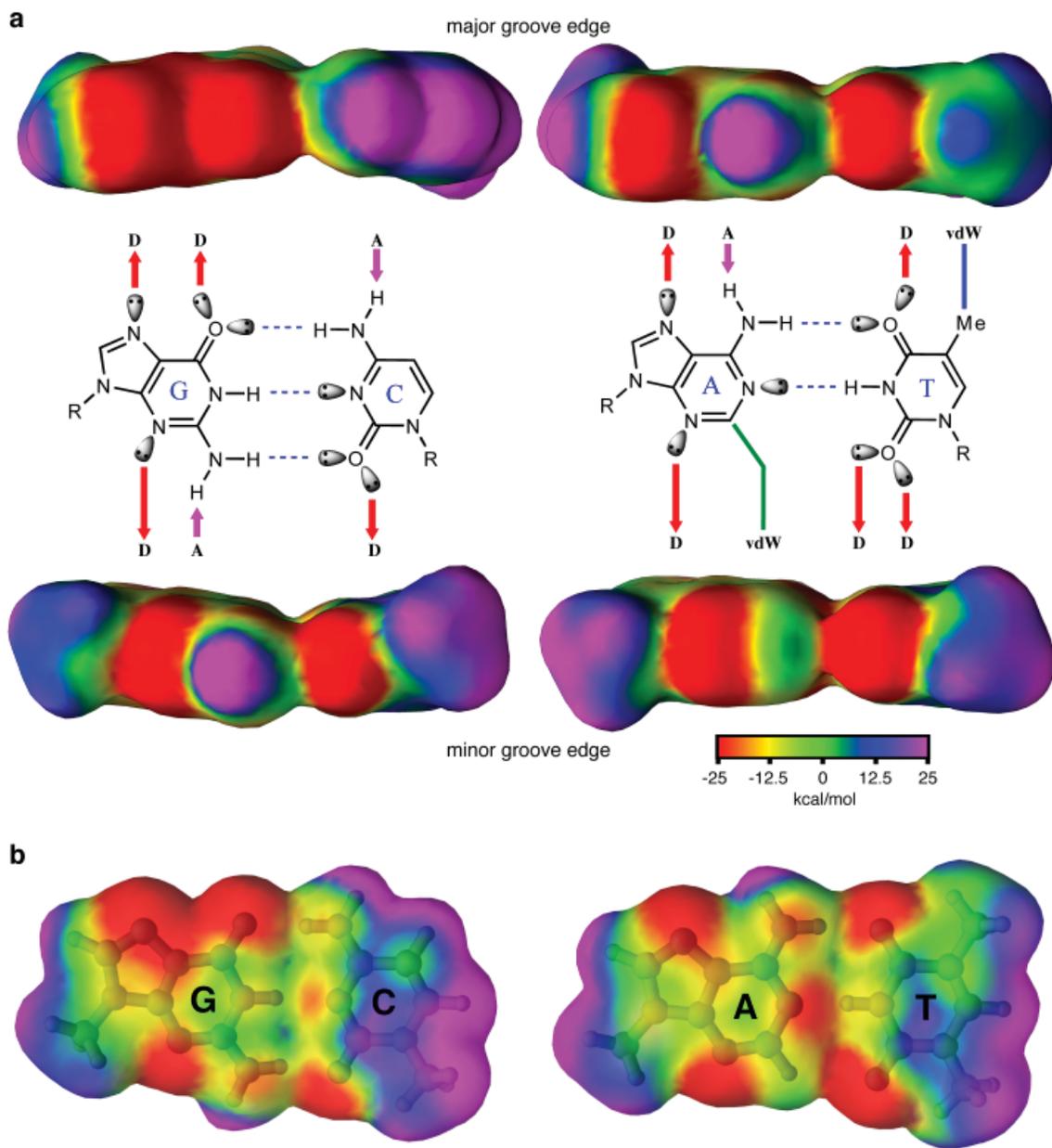
The regulation of gene transcription is controlled by the sequence specific cooperative assembly of transcription factors, which form regulatory switches and networks in the cell providing stringent control over biochemical processes.<sup>33</sup> The minor groove and major groove of



**Figure 1.5** Nucleosome core particle structures. a) Crystal structure of the nucleosome core particle at 2.80Å determined by Richmond and coworkers. b) X-ray structure of the tetra-NCP determined at 9Å resolution. c) Theoretical model of a chromatin fiber constructed from four tetra-NCPs by rotation and translation about the central axis.

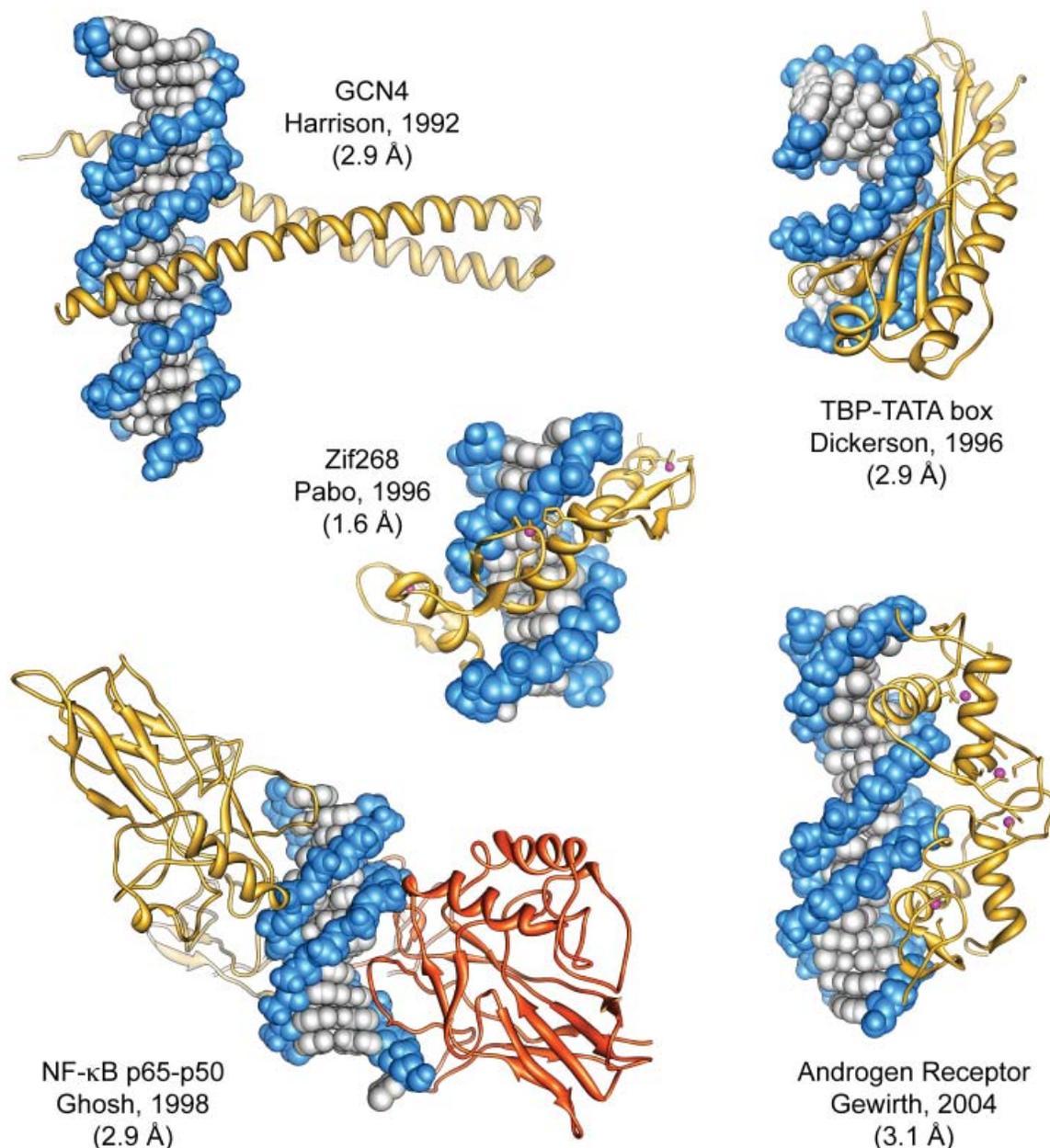
DNA provide distinct surfaces for the interaction of transcription factors through specific and nonspecific interactions such as (hydrogen bonding, electrostatics, van der Waals, etc.). Several DNA-binding transcription factors are presented in Figure 1.7 to highlight the diverse architectures used for recognizing DNA, ranging from homodimeric coiled coils interacting with the major groove to monomeric beta-sheet containing proteins interacting with the minor groove. In addition to homodimeric motifs, heterodimeric motifs are utilized along with metal ion coordinated assemblies (i.e. Nf-κB p65-p50 and androgen receptor).<sup>33</sup> Transcriptional co-activating proteins serve to integrate information from transcription factor assemblies and modulate gene expression through communication with RNA polymerase II leading to the transcription of protein-coding regions in the eukaryotic genome.<sup>33,34</sup>

Transcription factors (TF) can communicate indirectly through allosteric modulation of DNA resulting in cooperative assembly with very little direct protein-protein interaction. Transcription factor binding can cause DNA-sequence dependent structural perturbations which modulate the binding of the next TF. TF's can also interact directly through protein-protein interactions to increase



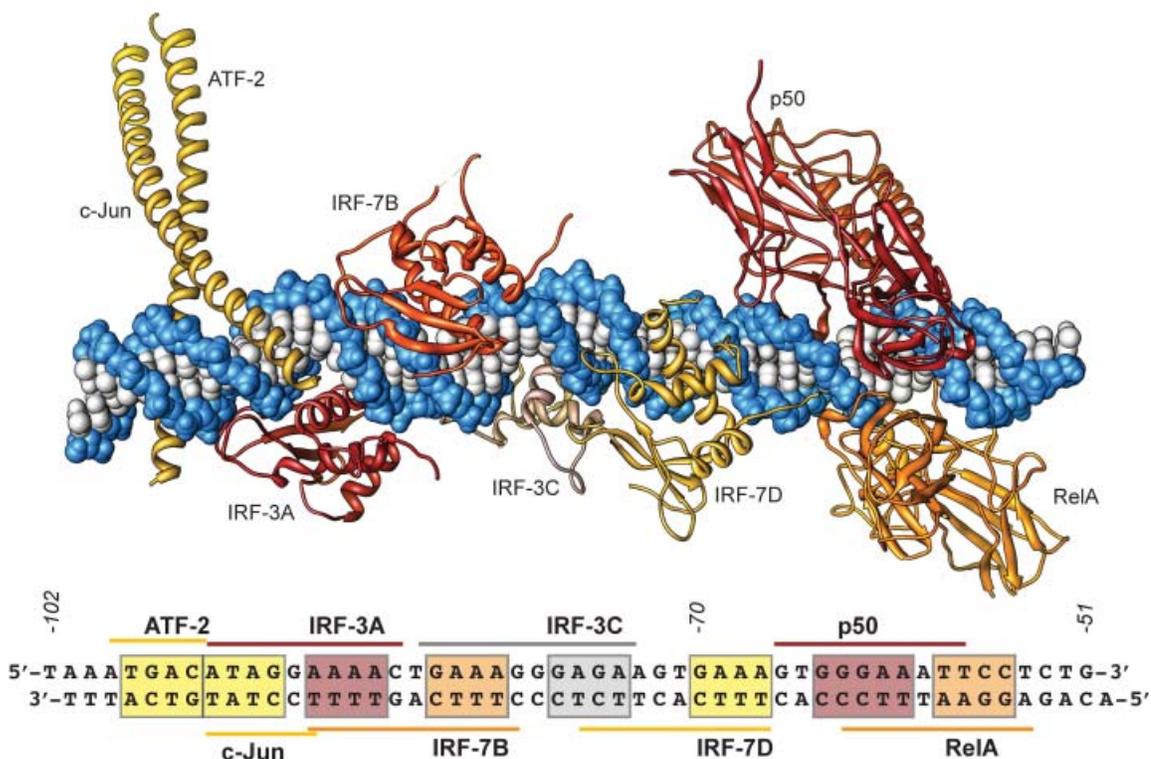
**Figure 1.6** Anatomy of the DNA base pair edges and their molecular recognition properties. a) Electrostatic potential maps of the base pair edges presented to the major (top) and minor (bottom) grooves of DNA. Hydrogen bond donors are designated with red arrows and the letter D. Hydrogen bond acceptors are designated with purple arrows and the letter A and functionality for the potential for van der Waals interactions is designated with the appropriate colored line and vdW. b) Top view of the Watson-Crick base pair molecular electrostatic potential surfaces. Electrostatic potential maps from native DNA crystal structure solved in Chapter 4 of this Thesis.

cooperativity. The  $\beta$ -enhancesome (Figure 1.8) is one such example of a cooperative assembly with cooperativity most likely arising at the DNA and coactivator levels. A conserved stretch of 55 bp's



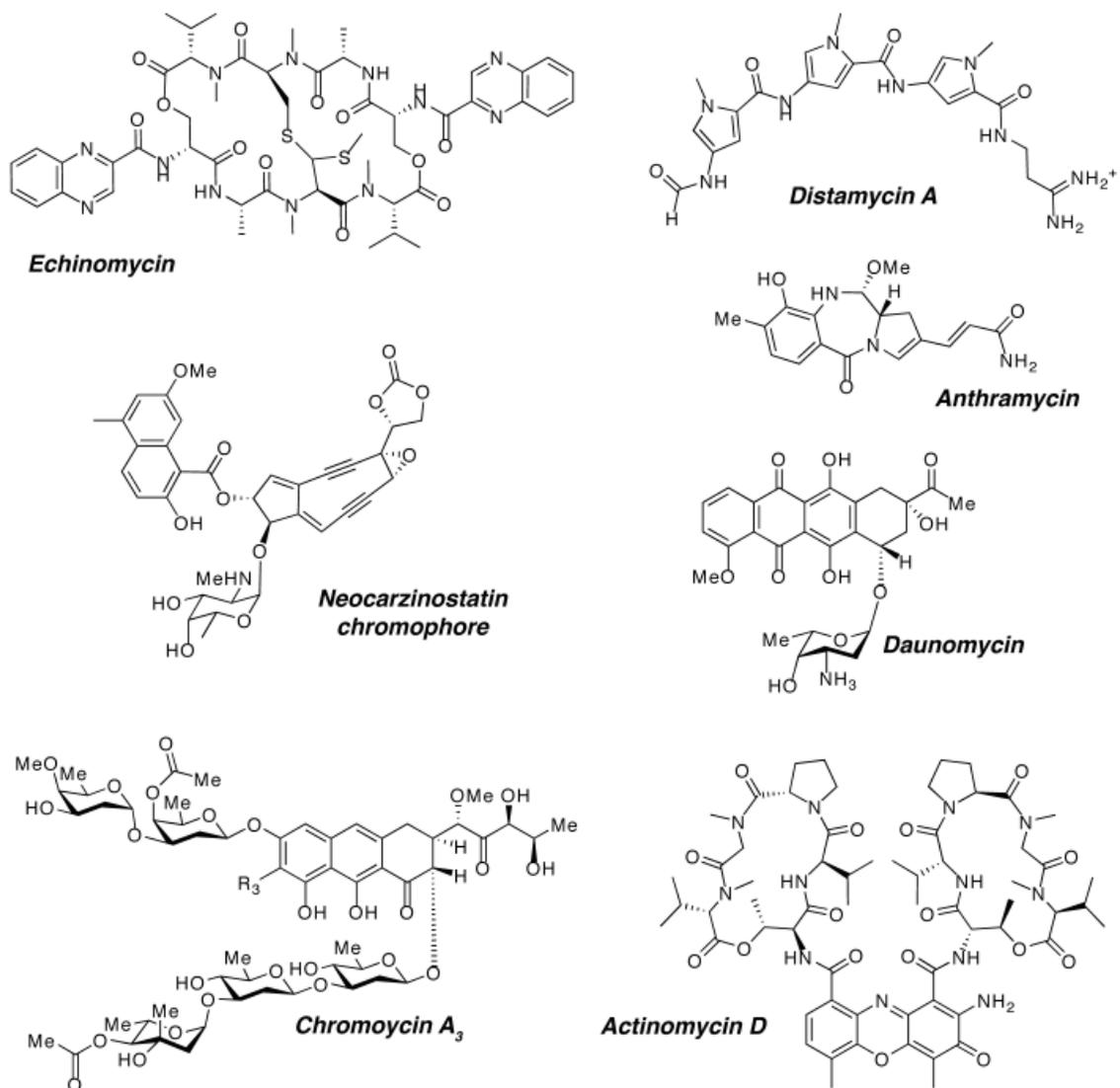
**Figure 1.7** X-ray structures of DNA binding transcription factors [GCN4 (Harrison, 1993), Nf- $\kappa$ B p65-p50 (Ghosh, 1998), TBP-TATA box (Dickerson, 1996), Zif268 (Pabo, 1996), Androgen receptor (Gewirth, 2004)].

(160 Å long) in a nucleosome free region of the IFN- $\beta$  promoter serves as a regulatory element for the cooperative assembly of 8 proteins into a continuous surface, burying 72% of the DNA solvent accessible area with very little protein-protein interaction.<sup>35</sup> Transcriptional co-activating proteins serve to integrate information from the assembly to modulate gene expression through communication with RNA polymerase II leading to transcription.<sup>33,34</sup>



**Figure 1.8** Atomic model of the cooperative assembly of interferon- $\beta$  enhancer 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.<sup>35</sup>

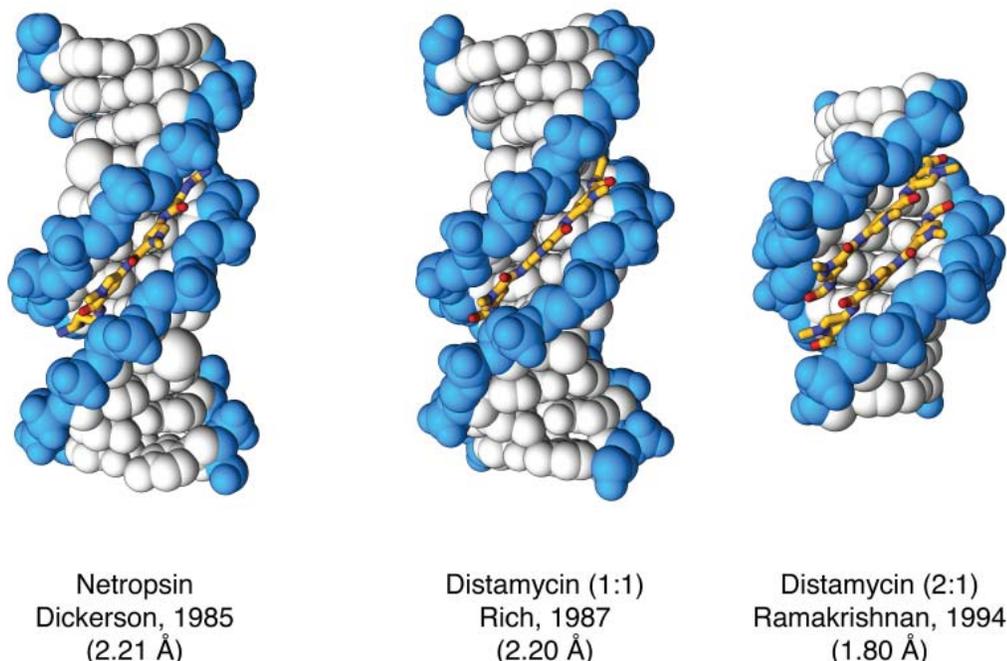
A diverse range of natural products and secondary metabolites have been shown to bind DNA with interaction modes consisting primarily of either intercalation or groove binding.<sup>5,7,8</sup> In addition, some ligands rely on a combination of intercalation and groove binding that can also be augmented by covalent modifying chemical domains, as in the case of anthramycin and neocarzinostatin. A collection of diverse DNA binding natural products are shown in Figure 1.9 with echinomycin and daunomycin representing intercalators and anthramycin and distamycin A representing minor groove binders. The natural product distamycin (Figure 1.9) binds to A,T tracks in the minor groove of DNA, four to five base pairs in size, in both a 2:1 and a 1:1 ligand:DNA stoichiometry.<sup>12-14</sup> The affinity and specificity of distamycin is controlled by a superposition of shape complementarity, hydrophobic effects, and specific hydrogen bonding to the minor groove of B-form DNA. Due to its modular design of repeating pyrrole amino acids and amenability to rational modification, distamycin has served as the inspiration for the design of several classes sequence specific DNA minor-groove binders, with the ultimate goal of designing highly specific targeted gene regulation agents.



**Figure 1.9** DNA-binding natural products.

#### 1.4 DNA Recognition by Minor-Groove Binders

Prior to the first structure of a molecule bound to DNA, specific recognition of B-form DNA was predicted to occur in major groove rather than minor groove.<sup>5</sup> An observation that was based on the fact that the hydrogen bond acceptors at N3 of adenine and O2 of thymine A/T base pairs are similarly placed and lack any prominent distinguishing features.<sup>36</sup> With the combination of biophysical and structural data from NMR and X-ray studies, it was verified that the minor groove of B-form DNA was a legitimate target for specific recognition.<sup>12-15</sup> (For crystal structures of netropsin and distamycin A, see Figure 1.10.) Building upon inspiration from the natural products, netropsin



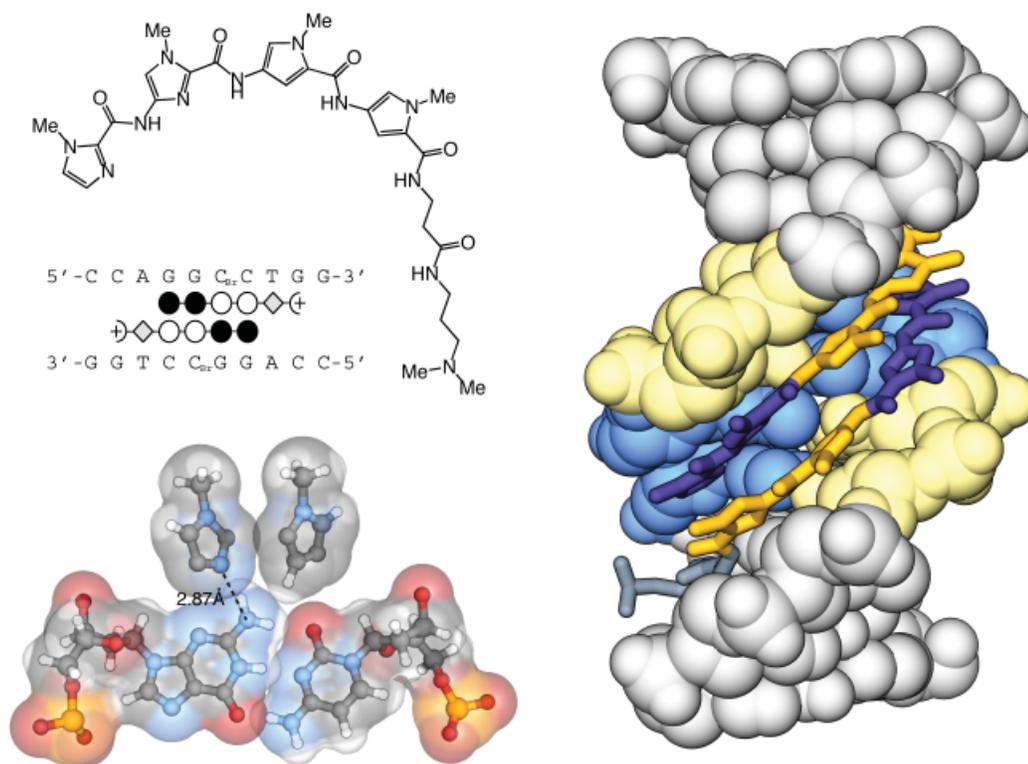
**Figure 1.10** X-ray crystal structures illustrating DNA recognition by the natural products netropsin (left) and distamycin (middle, right). The dicationic natural product netropsin binds preferentially to narrow AT tracts in the DNA minor groove as a monomer (PDB: 6BNA). The monocationic natural product distamycin also binds AT tracts of the DNA minor groove in a 1:1 (PDB: 2DND) and 2:1 (PDB: 378D) ligand:DNA stoichiometry.

and distamycin, minor groove binders have progressed to a modular molecular recognition platform with high affinity and specificity for many different sequences of DNA.<sup>5,7,8,16,17</sup>

Over the past two decades, the development of minor groove DNA binders has evolved from the initial discovery of the natural product distamycin to a new class of programmable heterocyclic oligomers demonstrating high affinity and sequence specificity.<sup>16,17</sup> In addition to the incorporation of alternative heterocycles such as imidazole that have enabled specificity for guanine recognition using the Im-Py pair, much research has gone into linking the two heterocyclic strands in a dimeric motif.<sup>37-39</sup> Covalent linkage of the two anti-parallel heterocyclic strands by a gamma amino butyric acid (GABA) linkage results in increases in affinity of 100–3600 fold relative to the unlinked homodimeric motif.<sup>40,41</sup> The incorporation of the turn linkage in the form of a GABA or substituted GABA turn represented a major technological advance allowing for the first time the incorporation of unsymmetrical ring pairs for the targeting of non-palindromic DNA sequences.<sup>37</sup> In addition, covalent linkage of the two strands has led to sub-nanomolar increases in affinity competing with and often rivaling that of endogenous DNA binding proteins.<sup>16,17,37-39</sup> This high affinity modular dimeric motif has allowed for the regulation of gene expression by direct interaction with the DNA-

protein interface.<sup>16,17</sup>

The four Watson-Crick base pairs can be differentiated by their molecular shape, electrostatic potential, and positions of hydrogen bond donors and acceptors in the DNA minor groove floor. The minor groove edge of a G•C base pair contains a steric hydrogen bond donating “bump” in the form of the exocyclic amine of guanine. The steric properties of the exocyclic amine of guanine form the basis for the A•T selectivity observed for netropsin and dystamycin binding due to steric interaction with the edge of the pyrrole ring. It was discovered in a key study in the early 1990s that imidazole in place of pyrrole in a three ring polyamide analogous to dystamycin could bind the 5' base pair sequence 5'-WGWCW-3' (where W=A or T) resulting in a 2:1 polyamide-DNA complex where the imidazole ring is stacked against a pyrrole ring allowing differentiation of G•C base pairs

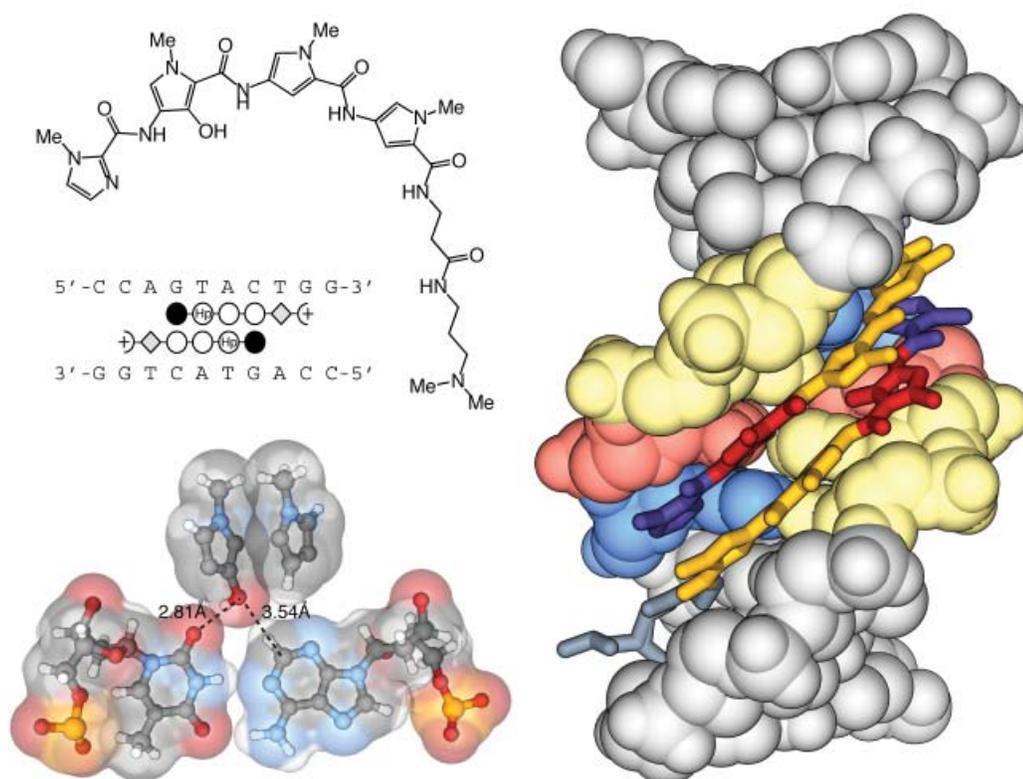


**Figure 1.11** Crystal structure (PDB: 365D) of the 2:1 binding single strand Py/Im polyamide targeted to the sequence 5'-CCAGGCCTGG-3' (2.00 Å resolution). Overall complex is shown on the right and the space filling model showing the basis for GC recognition is at the bottom left where the imidazole lone pair forms a hydrogen bond with the exocyclic N-H of guanine.

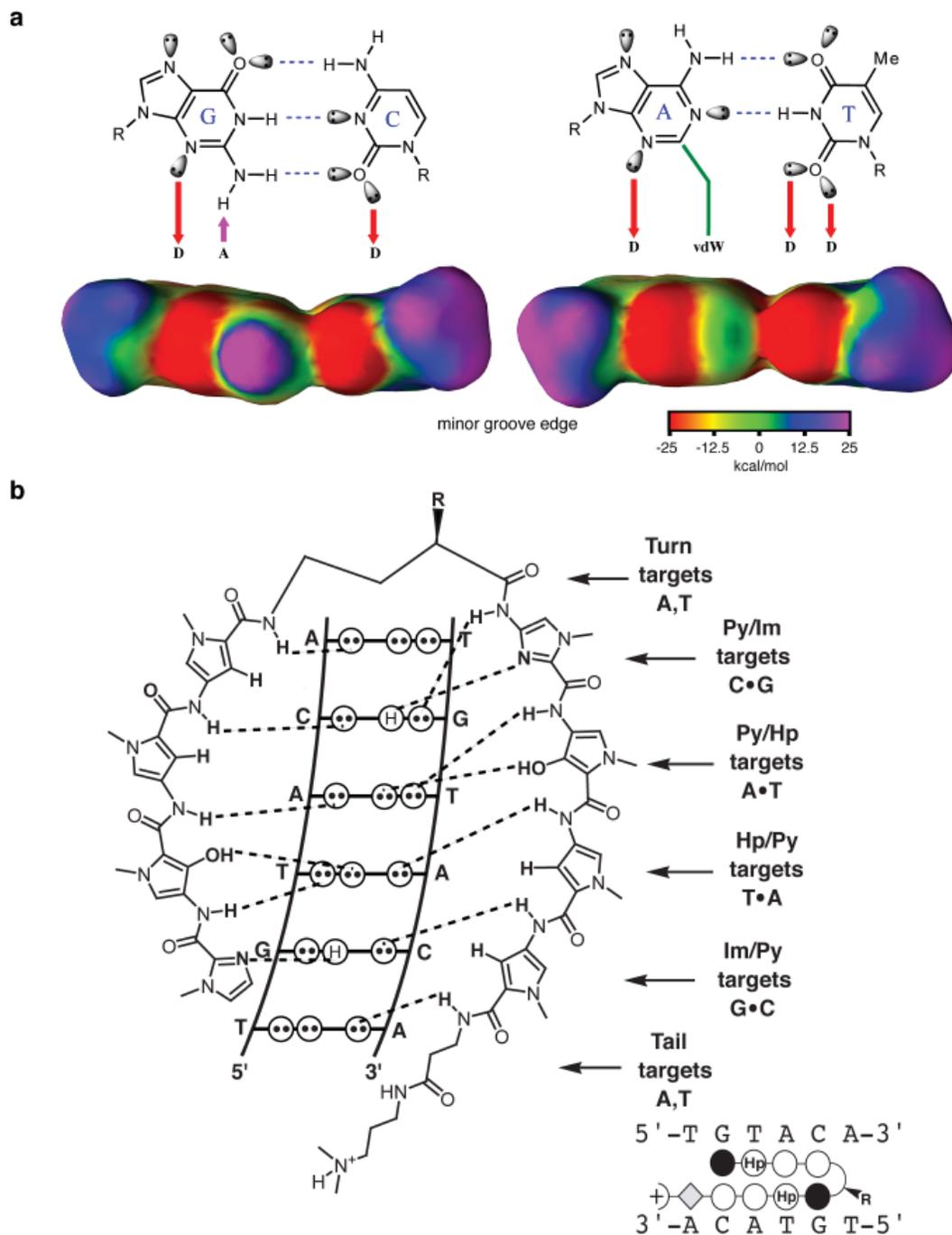
from C•G, A•T, or T•A.<sup>42</sup> The Im/Py pair has been used extensively in unlinked polyamides and in turn linked polyamides culminating in the recent publication of a polyamide library, that represents

the solutions for targeting various 6 base pair sequences with high affinity and specificity.<sup>16,17,43</sup> Thermodynamic studies have revealed that the Im/Py pair sequence selectivity is primarily driven by favorable enthalpic factors<sup>44,45</sup> and X-ray crystallographic studies in collaboration with the Rees group provided structural insight based on a specific hydrogen bond between the imidazole lone pair and the exocyclic amine of guanine in unlinked 2:1 homodimeric polyamides (Figure 1.11).<sup>46</sup>

Discrimination of T•A from A•T base pairs represents a much greater challenge due to the ability of thymine and adenine to both accept a hydrogen bond and the lack of unsymmetrical steric features as in the G•C case.<sup>16,17</sup> Despite this challenge, a small asymmetric cleft between the C2 of adenine and the O2 of thymine has been exploited for specific targeting by the *N*-methyl-3-hydroxypyrrrole/*N*-methylpyrrole (Hp/Py) pair, however affinities of these molecules are slightly lower than their Py/Py containing counterparts.<sup>16,17</sup> In another seminal structural study with the Rees group on Hp containing 2:1 binders, it was revealed that a combination of shape selective recognition of the asymmetric cleft along with a specific hydrogen bond between the Hp hydroxyl and the thymine O2 was responsible for the A•T specificity (Figure 1.12).<sup>47,48</sup> The combination of



**Figure 1.12** Crystal structure (PDB: 407D) of the 2:1 binding single strand ImHpPyPy-β-Dp polyamide targeted to the sequence 5'-CCAGTACTGG-3' (2.20 Å resolution). Overall complex is shown on the right and the space filling model of the basis for AT recognition is shown at the bottom left where the hydroxypyrrrole-OH is within hydrogen bonding distance of the carbonyl lone pair on T.

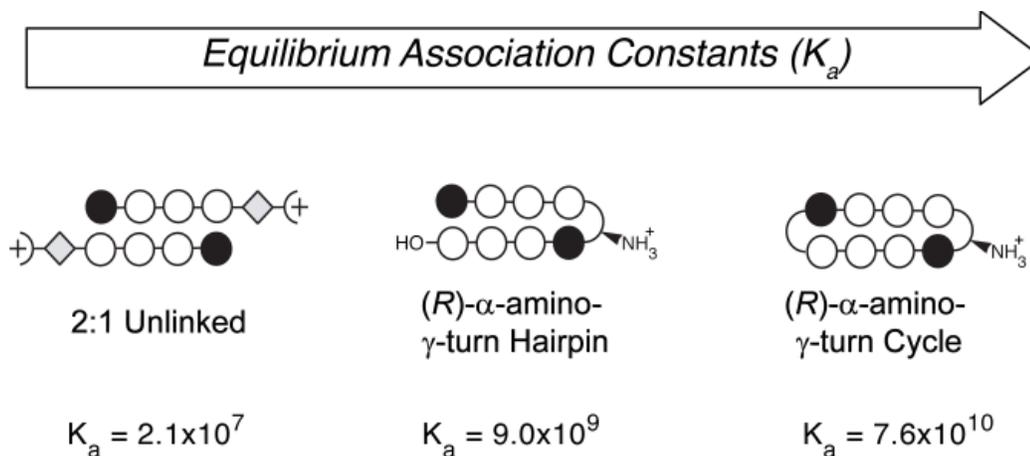


**Figure 1.13** Molecular recognition of the DNA minor groove and polyamide pairing rules. a) Molecular recognition of the DNA minor groove edges with molecular electrostatic potential surfaces showing the array of hydrogen bond donors (D), hydrogen bond acceptors (A), and hydrophobic functionality (vdW). b) Polyamide pairing rules.

Py, Im, and Hp combined as unsymmetrical pairs in opposite strands of a unlinked homodimeric or turn linked polyamide can be used to specifically recognize the four Watson-Crick base pairs

(Figure 1.13).<sup>16,17</sup> These interactions can be described as a set of guidelines or pairing rules for the design of sequence specific B-form DNA targeted polyamides where Im/Py specifies G•C and Hp/Py specifies A•T. Some limitations do exist for certain sequences such as homopurine tracts, certain G-rich sequences, and sequences beyond 6 base pairs due to the sequence-dependent DNA microstructure and overcurvature of longer polyamides, however unique solutions to some of these problems have been developed (i.e. incorporation of a flexible  $\beta$ -Ala residue in 1:1 and hairpin polyamide motifs).<sup>16,17</sup>

In addition to the 2:1 and hairpin polyamide architectures many other strand linkage strategies have been explored such as linking through the *N*-methyl groups on the central heterocycles (H-pin motif)<sup>49-50</sup> or in the terminal heterocycles (U-pin motif).<sup>51</sup> However, one of the highest affinity and in some cases most specific polyamide architectures has been the covalent linking of the C- to N-termini at both ends of the polyamide into a macrocycle, eliminating all possibility of extended binding modes.<sup>41,52-55</sup> Macrocytic  $\gamma$ -turn linked polyamides were first explored as 6 ring systems targeting a 5 base pair sequence of DNA in 1995 and were shown to have significantly higher affinity, however their specificity versus mismatch DNA was only 3-fold compared to 40-fold for their hairpin counterparts.<sup>52</sup> Mainly due to limitations in synthetic methodology and initial

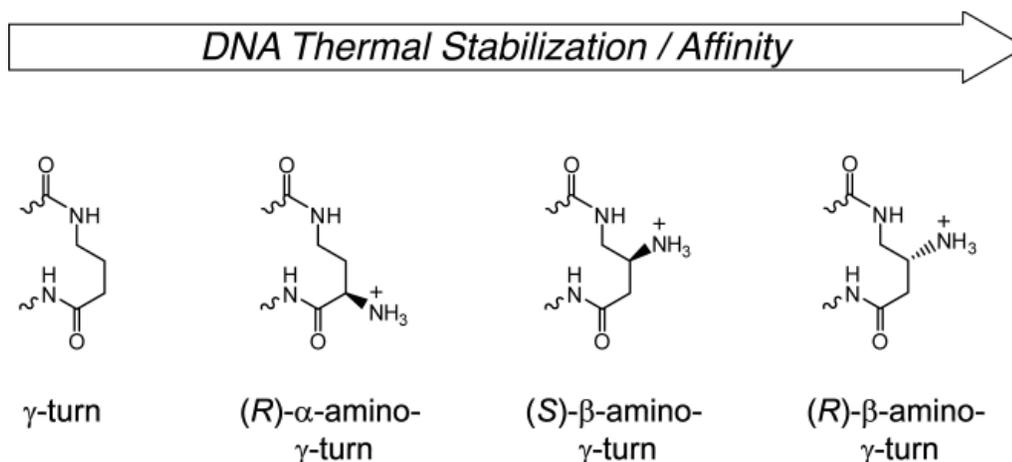


**Figure 1.14** Consequence of covalent attachment of two polyamide strands by incorporation of GABA-based turns. (For a structural key to the ball-and-stick nomenclature see the Nomenclature and Symbology section at the beginning of this thesis or Figure 1.13).

discouraging thermodynamic results the cyclic polyamide motif was not investigated further until 1999.<sup>41</sup> After improvements in solid-phase synthetic methodology, nanomole to micromole quantities of polyamides could be readily synthesized although cyclic polyamides still remained

challenging.<sup>53-55</sup> Using solid-phase methods cyclic polyamides were reinvestigated with two major architectural changes.<sup>41,53-55</sup> The first being the use of an 8 rings system as oppose to 6 in the original studies and the second major change was moving the charge from the pyrrole *N*-methyl group to the alpha position of the  $\gamma$ -turn in the form of (*R*)-2,4-diaminobutyric acid, that had been discovered to increase the affinity, sequence specificity, and orientational preference of hairpin polyamides. This second generation cyclic 8 ring polyamide motif was found to have greatly improved specificity and affinity over its hairpin and unlinked counterparts targeting the sequence 5'-AGTACT-3'.<sup>41</sup> The results of covalent attachment of the two polyamide strands can be seen in Figure 1.14. In a second study, multiple Hp/Py pairs were introduced into the 8-ring cyclic polyamide motif, which resulted in increased affinity and specificity relative to hairpin polyamides targeted to the sequences 5'-TGAACT-3' and 5'-TGATCT-3'.<sup>53</sup> Despite these advances the cyclic polyamide motif has received little attention relative to its hairpin counterpart mainly due to synthetic limitations.

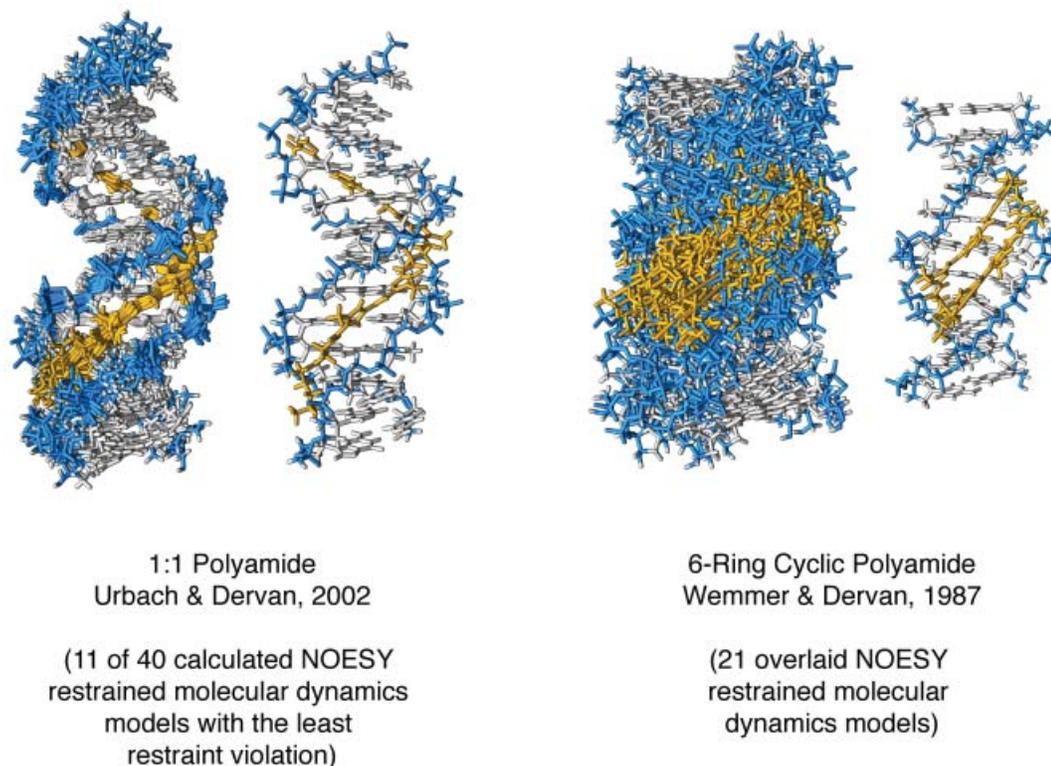
The chiral (*R*)-2,4-diaminobutyric acid turn ( $\alpha$ -turn) was a major advance in polyamide design not only for the cyclic polyamides but primarily for the hairpin motif.<sup>38</sup> The addition of an amino substituent to the alpha position of the  $\gamma$ -turn helps to disfavor extended 1:1 binding



**Figure 1.15** Consequence of covalent attachment of two polyamide strands by incorporation of GABA-based turns.

modes and reverse binding due to a steric clash with the minor groove floor. In addition, the chiral amino turn helps to increase the overall affinity of polyamides while maintaining specificity and improving water solubility.<sup>38</sup> The chiral  $\alpha$ -turn was proposed to increase affinity through electrostatic interactions between the protonated cationic amine group and the anionic DNA backbone however this interaction has not been born out in structural studies (see Chapter 5 and 6 of this thesis).<sup>17</sup> In

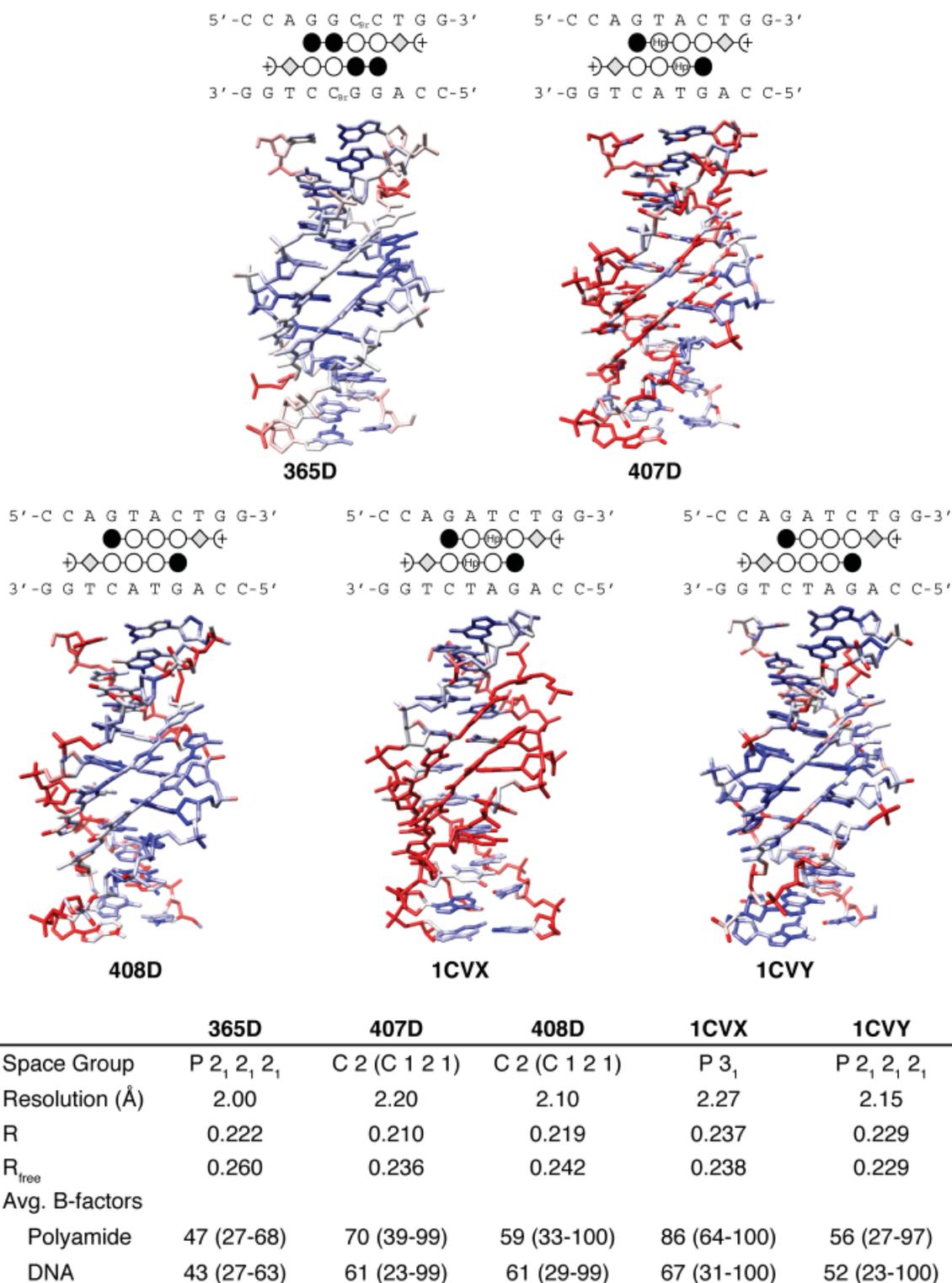
addition to substitution of the  $\gamma$ -turn at the alpha position, many other variations have been studied with shorter turns, longer turns, and conformationally constrained variations.<sup>16,17,56</sup> The most recent and one of the most successful advances in polyamide turn technology is the  $\beta$ -amino  $\gamma$ -linked



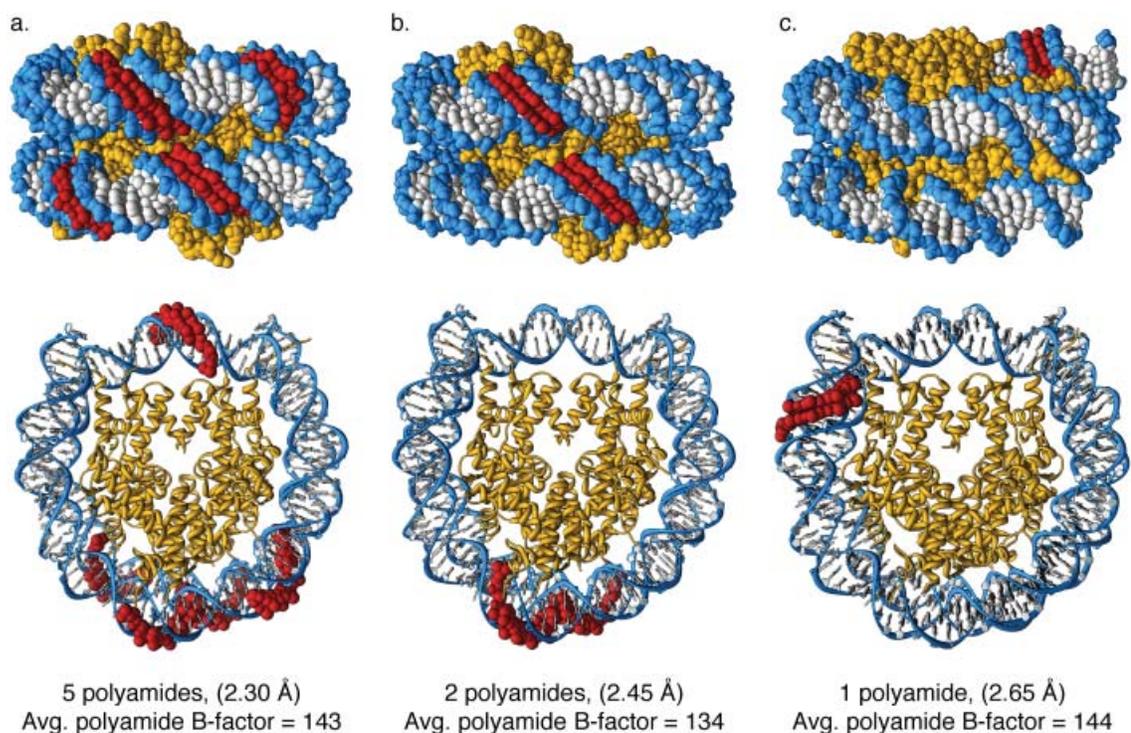
**Figure 1.16** NMR NOESY-restrained molecular dynamics models of 1:1 and 6-ring cyclic polyamide DNA complexes.

turn.<sup>39</sup> Recent studies have demonstrated that this turn can provide substantial increases in affinity for certain polyamide sequences, however the effect is less pronounced as the imidazole content of the polyamide is increased (Figure 1.15).<sup>39</sup>

NMR structural studies using NOESY-restrained molecular dynamics models have also provided insight into 1:1 and 6-ring cyclic polyamides complexed with DNA.<sup>55,57</sup> Figure 1.16 shows an NMR model of a 1:1 polyamide with 11 out of 40 of the best calculated models overlaid.<sup>57</sup> This shows very small coordinate deviation towards the center of the DNA helix and bound polyamide with increasing conformational mobility at the ends. The 6-ring cyclic polyamide model represented in Figure 1.16 shows an overlay of 21 of the best calculated models.<sup>55</sup> This structure shows significant conformational mobility in the 6-ring cyclic complex with a highly flexible DNA sugar-phosphate backbone.

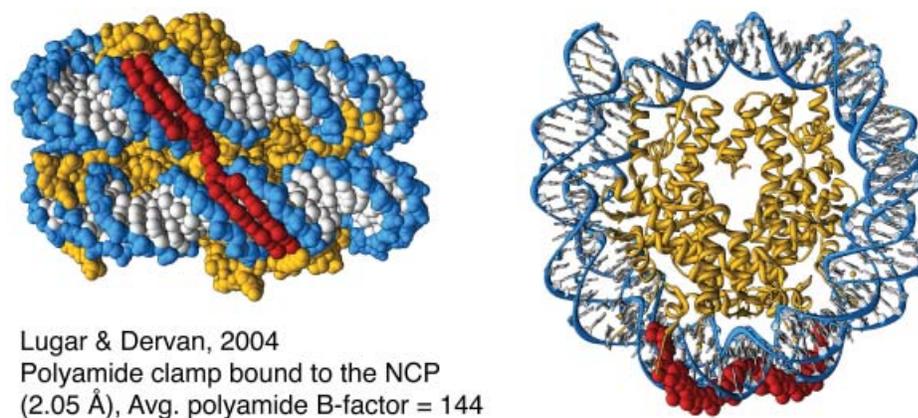


**Figure 1.17** Polyamide 2:1 DNA crystal structures colored by B-factor with red representing the largest B-factors and blue representing the smallest.



**Figure 1.18** X-ray crystal structures of polyamide-NCP complexes. a) Five polyamides in complex with the NCP at 2.30 Å resolution. b) Two polyamides in complex with the NCP at 2.45 Å resolution. c) One polyamides in complex with the NCP at 2.65 Å resolution.

X-ray crystallographic studies resulting from collaborations between the Rees and Dervan groups have provided valuable insight into the polyamide-DNA molecular recognition process by elucidating the structure of five 2:1 polyamide-DNA complexes at a resolution ranging from 2.00 to 2.27 Å and a summary comparing specific structural parameters is shown in Figure 1.17.<sup>46-48</sup> These crystallographic studies revealed that the DNA rise per base pair matches the polyamide rise



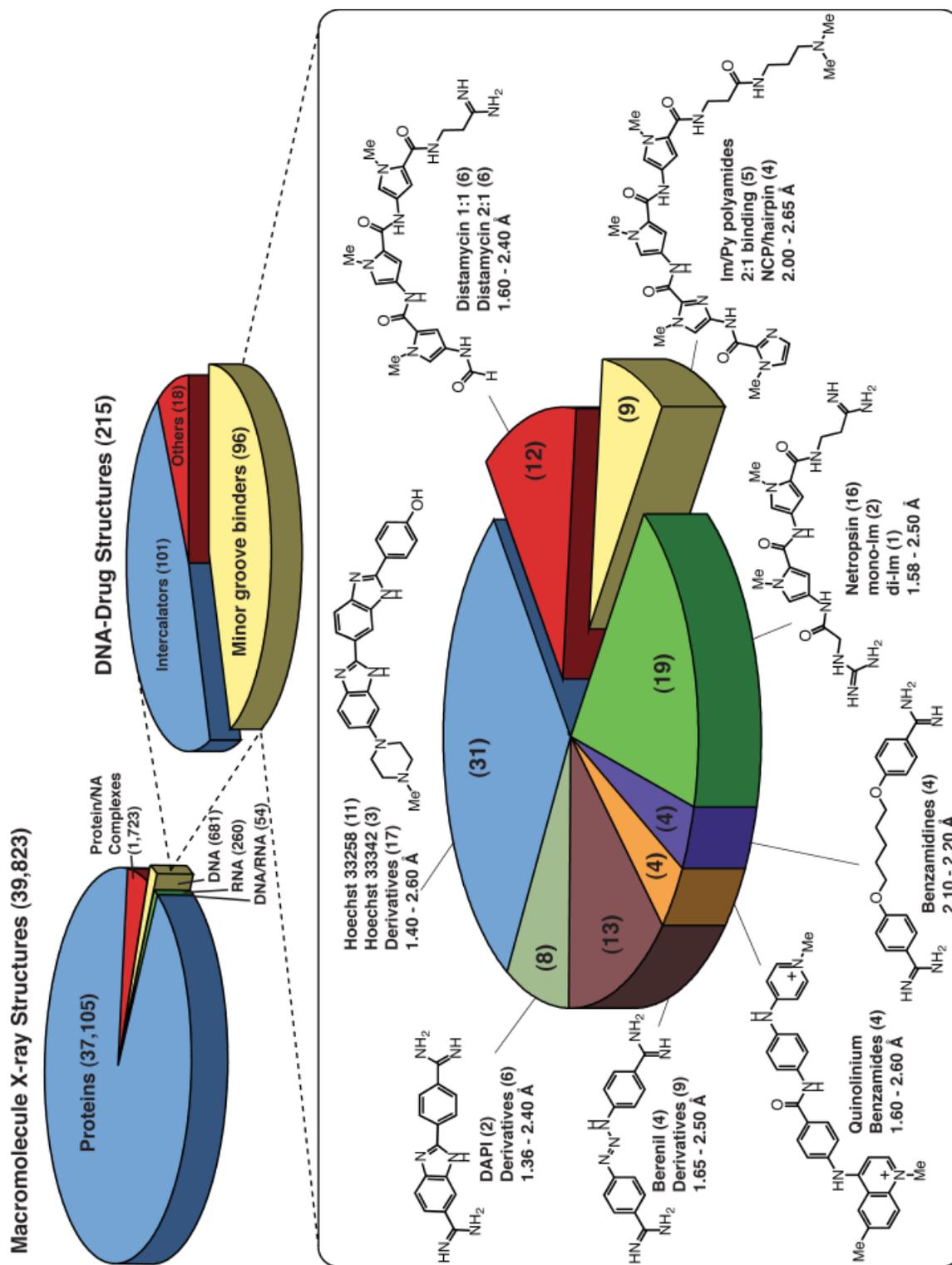
**Figure 1.19** Turn-linked polyamide clamp bound to the NCP with the linker traversing the nucleosome super-groove.

per residue, however the polyamide structure is over-curved with respect to the DNA minor groove and shape complementarity is lost beyond a sequence of 5 contiguous base pairs.<sup>46</sup> In addition, these crystallographic studies elucidated the basis for GC recognition by the Im/Py pair<sup>46</sup> and TA recognition by the Hp/Py pair<sup>47,48</sup> providing fundamental insight into polyamide binding. A variety of space groups were observed including monoclinic, orthorhombic, and trigonal with resolutions ranging from 2.00 to 2.27 Å. Average R-factors were in the mid-20s and polyamide B-factors averaged 47 to 86 Å<sup>2</sup> whereas DNA B-factors averaged 43 to 67 Å<sup>2</sup> for all structures presented in Figure 1.17. The 2:1 binding polyamide crystal structures also frequently exhibited disorder in the polyamide tail region and was usually modeled in alternate conformations reflecting the dynamic nature of the β-alanine linked dimethylamino propylamine terminus.

DNA binding polyamides are also able to bind sequence specifically to DNA on the nucleosome core particle.<sup>58</sup> Hairpin polyamide-NCP crystal structures have been solved at resolutions ranging from 2.05 to 2.65 Å providing structural proof that polyamides can bind biologically relevant higher-order DNA structure however a combination of resolution limits and high B-factors for the polyamide prevented a detailed picture beyond confirmation of the polyamide binding location (Figures 1.18 and 1.19).<sup>58,59</sup> The current state of macromolecular crystallography, with regard to minor groove binding DNA-drug structures, was assessed prior to beginning the structural work presented in Chapter 5 and 6 of this thesis and is presented in Figure 1.20. This survey demonstrates the lack of high resolution structures of DNA minor groove binders and the notable absence of linked dimeric minor groove binder structures. This survey underscores the pressing need for atomic resolution X-ray crystal structures of DNA minor groove binders to truly understand the molecular basis of recognition.

## 1.5 Scope of this work

The work presented in this thesis is focused on the molecular recognition of DNA by minor groove binding polyamides. In Chapter 2 of this thesis, a solution-phase synthesis of pyrrole-imidazole polyamides is presented with optimized protocols utilizing little to no chromatography. Chapter 3 builds on synthetic methodology in Chapter 2 allowing the efficient synthesis of cyclic polyamides targeted to the androgen response element. This chapter demonstrates that cyclic polyamides can be synthesized in an efficient manner, are biologically active and cell permeable in cell culture experiments, and rival the binding affinity of most other polyamide architectures. Chapter 4 details an oligomerization route to macrocyclic polyamides and reports on the DNA



**Figure 1.20** Current state of macromolecular crystallography: A DNA-drug perspective. Data compiled from the PDB on 11/04/2007.<sup>60</sup> (The number of structures solved is designated in parenthesis.)

binding ability of higher order macrocycles. The structural elucidation of an  $\alpha$ -amino-turn-linked cyclic polyamide is presented in Chapter 5 at 1.18 Å resolution providing insight into the detailed molecular recognition processes. Chapter 6 details the structural elucidation of a  $\beta$ -amino-turn-linked cyclic polyamide highlighting the conformational differences compared to the  $\alpha$ -amino-turns and providing a structural basis for the inability of polyamides to bind dsRNA. In Chapter 7, a new class of programmable oligomers targeting the DNA sequence 5'-WGGGGW-3' were shown to inhibit DNA binding of the Nf- $\kappa$ B transcription factor by EMSA gel shift. Compounds discovered in Chapter 7 were found to possess unique fluorescent properties with the ability to modulate their fluorescence by binding their targeted dsDNA site and this work is presented in Chapter 8. Chapter 9 describes an ongoing effort in the templated-assembly of polyamides using higher-order DNA structure (NCP). Additionally, this chapter describes the development of a new profluorescent class of heterocycle, that has the potential to be used as a chemical reporter for templated ligation events. Appendix A through F detail results from efforts not covered in the main thesis and a continuation of studies from Chapters 3 and 8.

## 1.6 Notes and Reference

1. Lehn, J. M. *Supramolecular Chemistry: Concepts and Perspectives* Wiley-VCH: Weinheim, Germany, **1995**.
2. Cram, D. J. The design of molecular hosts, guests, and their complexes *Science* **1988**, *240*, 760-767.
3. Cram, D. J. Molecular container compounds. *Nature* **1992**, *356*, 29-36.
4. Whitesides, G. M., Snyder, P. W., Moustakas, D. T., Mirica, K. A. Designing ligands to bind tightly to proteins. In *Physical Biology: From Atoms to Medicine* Zewail, A., Eds.; Imperial College Press: London, **2008**, 189-216.
5. Waring, M. J., Wakelin, L. P. G. Forty Years On. In *DNA and RNA Binders* (Demeunynck, M., Bailly, C., and Wilson, W. D., eds) Wiley-VCH: Weinheim, Germany, **2003**; 1, 1-17.
6. Lerman, L. S. Structural considerations in the interaction of DNA and acridines. *J. Mol. Biol.* **1961**, *3*, 18-30.
7. *Methods in Enzymology: Drug-Nucleic Acid Interactions*. Chaires J. B., Waring M. J. Academic Press: New York, **2001**.
8. *Sequence-specific DNA Binding Agents*. Waring M. Eds.; RSC Publishing: Cambridge, UK, **2006**.
9. Tsai, C. C., Jain, S. C., and Sobell, H. M. X-ray crystallographic visualization of drug-nucleic acid intercalative binding: structure of an ethidium-dinucleoside monophosphate crystalline complex, Ethidium: 5-iodouridylyl (3'-5') adenosine. *Proc. Natl. Acad. Sci. U. S. A.* **1975**, *72*, 628.

10. Wang, A. H., Nathans, J., van der Marel, G., van Boom, J. H., and Rich, A. Molecular structure of a double helical DNA fragment intercalator complex between deoxy CpG and a terpyridine platinum compound. *Nature* **1978**, 276, 471-474.
11. Shieh, H. S., Berman, H. M., Dabrow, M., and Neidle, S. The structure of drug-deoxydinucleoside phosphate complex; generalized conformational behavior of intercalation complexes with RNA and DNA fragments. *Nucleic Acids Res.* **1980**, 8, 85.
12. Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P., and Dickerson, R. E. The molecular origin of DNA-drug specificity in netropsin and distamycin. *Proc. Natl. Acad. Sci. USA.* **1985**, 82, 1376-1380.
13. Coll, M., Frederick, C. A., Wang, A. H., and Rich, A. A bifurcated hydrogen-bonded conformation in the d (AT) base pairs of the DNA dodecamer d (CGCAAATTTGCG) and its complex with distamycin. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, 84, 8385-8389.
14. Pelton, J. G., and Wemmer, D. E. Structural characterization of a 2: 1 distamycin Ad (CGCAAATTGGC) complex by two-dimensional NMR. *Proc. Natl. Acad. Sci. U. S. A.* **1989**, 86, 5723-5727.
15. Chen, X., Ramakrishnan, B., Rao, S. T., and Sundaralingam, M. Binding of two distamycin A molecules in the minor groove of an alternating B-DNA duplex *Nat. Struct. Biol.* **1994**, 1, 169-175.
16. Dervan, P. B. Molecular recognition of DNA by small molecules. *Bioorg. Med. Chem.* **2001**, 9, 2215-2235.
17. Dervan, P. B., and Edelson, B. S. Recognition of the DNA minor groove by pyrrole-imidazole polyamides *Curr. Opin. Struct. Biol.* **2003**, 13, 284-299. (For a cyclic polyamide crystal structure containing the  $\alpha$ -amino turn see Chapter 4 of this thesis.)
18. Rees, D. A. The future of biological X-ray analysis. In *Physical Biology: From Atoms to Medicine*. Zewail, A., Eds.; Imperial College Press: London, **2008**, 145-164.
19. Bloomfield, V. A., Crothers, D. M., Tinoco, I. *Nucleic Acids: Structure, Properties and Functions*. University Science Books: Sausalito, CA, **2000**
20. Neidle, S. *Nucleic Acid Structure and Recognition*. Oxford University Press: London, **2002**
21. Dickerson, R. E. *Present at the Flood: How Structural Molecular Biology Came About*. Sinauer Associates, Inc: Sunderland, MA, **2005**
22. Watson, J. D., and Crick, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **1953**, 171, 737-738.
23. Miller, A., Tanner, J. *Essentials of Chemical Biology*. John Wiley & Sons, Ltd: West Sussex, England, **2008**
24. Arnott, S. Historical article: DNA polymorphism and the early history of the double helix. *Trends Biochem. Sci.* **2006**, 31, 349-354.
25. Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **1997**, 389, 251-260.
26. Richmond, T. J., and Davey, C. A. The structure of DNA in the nucleosome core. *Nature* **2003**, 423, 145-150.

27. Schalch, T., Duda, S., Sargent, D. F., and Richmond, T. J. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature* **2005**, *436*, 138-141.
28. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., and others, O. Initial sequencing and analysis of the human genome. *Nature* **2001**, *409*, 860-921.
29. Venter, J. C., et al. The sequence of the human genome. *Science* **2001**, *291*, 1304-1351.
30. International Human Genome Sequencing Consortium Finishing the euchromatic sequence of the human genome. *Nature* **2004**, *431*, 931-945.
31. Rice, P. A., Correll, C. C. *Protein-nucleic Acid Interactions: Structural Biology*. Rice P. A., Correll C. C. RSC Publishing: Cambridge, UK, **2008**
32. All ab initio calculations reported here were performed using HF/3-21G\* as implemented in the Gamess program on structures whose coordinates correspond to those of the crystal structure. Electrostatic potential surfaces were generated by mapping the electrostatic potentials onto surfaces of molecular electron density (0.002 electron/Å) and color-coding, using the Chimera program. The molecular electrostatic potential energy values range from -25 kcal/mol for values of negative potential (red) to +25 kcal/mol for values of positive potential (blue). This range was chosen to emphasize the variations in the aromatic region and some regions of the electrostatic potential associated with heteroatoms may lie beyond the  $\pm 25$  kcal/mol range. 3-21G\* basis set: Francl, M. M., Pietro, W. J., Hehre, W. J., Binkley, J. S., Gordon, M. S., Defrees, D. J., and Pople, J. A. Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *The Journal of Chemical Physics*. **1982**, *77*, 3654-3665. (b) Hariharan, P. C., and Pople, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)*. **1973**, *28*, 213-222. Gamess program: Schmidt, M. W., Baldridge, K. K., Boatz, J. A., Elbert, S. T., Gordon, M. S., Jensen, J. H., Koseki, S., Matsunaga, N., Nguyen, K. A., and Su, S. General atomic and molecular electronic structure system. *J. Comput. Chem*. **1993**, *14*, 1347-1363. UCSF Chimera: Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem*. **2004**, *25*, 1605-1612.
33. Wolberger, C. Multiprotein-DNA complexes in transcriptional regulation. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 29-56.
34. Naar, A. M., Lemon, B. D., and Tjian, R. Transcriptional coactivator complexes. *Annu. Rev. Biochem.* **2001**, *70*, 475-501.
35. Panne, D., Maniatis, T., and Harrison, S. C. An atomic model of the interferon-beta enhanceosome. *Cell* **2007**, *129*, 1111-1123.
36. Seeman, N. C., Rosenberg, J. M., and Rich, A. Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1976**, *73*, 804.
37. Mrksich, M., Parks, M. E., and Dervan, P. B. Hairpin peptide motif. A new class of oligopeptides for sequence-specific recognition in the minor groove of double-helical DNA. *J. Am. Chem. Soc.* **1994**, *116*, 7983-7988.
38. Herman, D. M., Baird, E. E., and Dervan, P. B. Stereochemical Control of the DNA Binding Affinity, Sequence Specificity, and Orientation Preference of Chiral Hairpin Polyamides in the

Minor Groove *J. Am. Chem. Soc.* **1998**, *120*, 1382-1391.

39. Dose, C., Farkas, M. E., Chenoweth, D. M., and Dervan, P. B. Next generation hairpin polyamides with (R)-3,4-diaminobutyric acid turn unit. *J. Am. Chem. Soc.* **2008**, *130*, 6859-6866. For initial investigations into the (R)- and (S)- $\beta$ -amino-GABA turns and stereospecific synthetic routes, see Scott Carter's Thesis (reference 56 below).

40. Trauger, J. W., Baird, E. E., and Dervan, P. B. Recognition of DNA by designed ligands at subnanomolar concentrations. *Nature* **1996**, *382*, 559-561.

41. Herman, D. M., Turner, J. M., Baird, E. E., and Dervan, P. B. Cycle polyamide motif for recognition of the minor groove of DNA *J. Am. Chem. Soc.* **1999**, *121*, 1121-1129.

42. Mrksich, M., Wade, W. S., Dwyer, T. J., Geierstanger, B. H., Wemmer, D. E., and Dervan, P. B. 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.* **1992**, *89*, 7586.

43. Hsu, C. F., Phillips, J. W., Trauger, J. W., Farkas, M. E., Belitsky, J. M., Heckel, A., Olenyuk, B. Z., Puckett, J. W., Wang, C. C. C., and Dervan, P. B. Completion of a programmable DNA-binding small molecule library. *Tetrahedron* **2007**, *63*, 6146-6151.

44. Pilch, D. S., Poklar, N., Gelfand, C. A., Law, S. M., Breslauer, K. J., Baird, E. E., and Dervan, P. B. Binding of a hairpin polyamide in the minor groove of DNA: Sequence-specific enthalpic discrimination. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 8306-8311.

45. Crothers, D. M., and Fried, M. Transmission of long-range effects in DNA. *Cold Spring Harb. Symp. Quant. Biol.* **1983**, *47*, 263-269.

46. Kielkopf, C. L., Baird, E. E., Dervan, P. B., and Rees, D. C. Structural basis for G•C recognition in the DNA minor groove. *Nat. Struct. Biol.* **1998**, *5*, 104-109.

47. Kielkopf, C. L., White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., Dervan, P. B., and Rees, D. C. A structural basis for recognition of A.T and T.A base pairs in the minor groove of B-DNA. *Science* **1998**, *282*, 111-115.

48. Kielkopf, C. L., Bremer, R. E., White, S., Szewczyk, J. W., Turner, J. M., Baird, E. E., Dervan, P. B., and Rees, D. C. Structural effects of DNA sequence on TA recognition by hydroxypyrrole/pyrrole pairs in the minor groove. *J. Mol. Biol.* **2000**, *295*, 557-567.

49. Greenberg, W. A., Baird, E. E., and Dervan, P. B. A Comparison of H-Pin and Hairpin Polyamide Motifs for the Recognition of the Minor Groove of DNA

50. Olenyuk, B., Jitianu, C., and Dervan, P. B. Parallel synthesis of H-pin polyamides by alkene metathesis on solid phase. *J. Am. Chem. Soc.* **2003**, *125*, 4741-4751.

51. Heckel, A., and Dervan, P. B. U-pin polyamide motif for recognition of the DNA minor groove. *Chem. Eur. J.* **2003**, *9*, 1-14.

52. Cho, J., Parks, M. E., and Dervan, P. B. Cyclic polyamides for recognition in the minor groove of DNA. *Proc. Natl. Acad. Sci. USA.* **1995**, *92*, 10389-10392.

53. Melander, C., Herman, D. M., and Dervan, P. B. Discrimination of A/T sequences in the minor groove of DNA within a cyclic polyamide motif. *Chemistry* **2000**, *6*, 4487-4497.

54. Baliga, R., Baird, E. E., Herman, D. M., Melander, C., Dervan, P. B., and Crothers, D. M. Kinetic consequences of covalent linkage of DNA binding polyamides. *Biochemistry* **2001**, *40*,

3-8.

55. Zhang, Q., Dwyer, T. J., Tsui, V., Case, D. A., Cho, J., Dervan, P. B., and Wemmer, D. E. NMR structure of a cyclic polyamide-DNA complex. *J. Am. Chem. Soc.* **2004**, *126*, 7958-7966.

56. Carter, S. R. *Sequence-specific minor groove binding polyamides: DNA recognition and applications*. California Institute of Technology, **1998**.

57. Urbach, A. R., Love, J. J., Ross, S. A., and Dervan, P. B. Structure of a beta-alanine-linked polyamide bound to a full helical turn of purine tract DNA in the 1: 1 motif. *J. Mol. Biol.* **2002**, *320*, 55-71.

58. Suto, R. K., Edayathumangalam, R. S., White, C. L., Melander, C., Gottesfeld, J. M., Dervan, P. B., and Luger, K. Crystal structures of nucleosome core particles in complex with minor groove DNA-binding ligands. *J. Mol. Biol.* **2003**, *326*, 371-380.

59. Edayathumangalam, R. S., Weyermann, P., Gottesfeld, J. M., Dervan, P. B., and Luger, K. Molecular recognition of the nucleosomal "supergroove". *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 6864-6869.

60. Berman, H. M. et al. "The Protein Data Bank" *Nucleic Acids Res.* **2000**, *28*, 235-242.