structures, although this may be inadequate to account for the energetic differences [11]. A recent computational study by Liedl and co-workers [12] argues that desolvation of the hydroxyl group upon insertion into the minor groove accounts for the energetic penalty. Thus, Py, Im and Hp can be combined to recognize specifically each of the four Watson–Crick base pairs; Im/Py is specific for G/C and Hp/Py for T/A. These interactions can be conveniently described as ‘pairing rules’ (Figures 1–3).

**Affinity and specificity**

Covalently linking the two antiparallel polyamide strands results in molecules with increased affinity and specificity. Currently, the ‘standard’ motif is the eight-ring hairpin, in which a γ-aminobutyric acid (γ-turn) linker connects the carboxylic terminus of one polyamide to the amino terminus of another (Figures 1 and 2). Compared to the unlinked homodimers, hairpin polyamides display approximately 100-fold higher affinity, with the γ-turn demonstrating selectivity for A,T over G,C base pairs, presumably due to steric clash with the exocyclic amine of guanine [13]. Eight-ring hairpins, which bind 6 bp, were shown to have affinity and sequence specificity similar to that of DNA-binding proteins (i.e. Kd < 1 nM) [14]. Also, NMR studies confirmed that the γ-turn locks the register of the ring pairings, preventing the ambiguity of slipped dimers, in which the polyamide chains bind to DNA partially in the 2:1 mode and partially in the 1:1 mode [15]. Hairpin compounds retain the orientation preferences of extended polyamides, aligning N—C with respect to the 5′—3′ direction of the adjacent DNA strand [16].

For some hairpins, however, reversed binding (a C—N alignment of the polyamide with respect to the 5′—3′ direction of the adjacent DNA strand) has been observed as the dominant orientation [16]. In other cases, the folded hairpin mode is favored only approximately five-fold over the linear 1:1 binding mode, in which the polyamide does not fold back on itself [17]. A solution to both problems comes from introducing an amino substituent at the α position of the γ-turn (Figure 4) [18]. Reversed binding is disfavored because of a steric clash between the amino substituent and the floor of the minor groove [18]; extended 1:1 binding is similarly destabilized by the chiral \(R^{1}\) or \(S^{1}\) γ-turn, an effect that is amplified by acetylation of the amine group [17]. Furthermore, the chiral turn maintains the selectivity of hairpins and increases the affinity, most likely through electrostatic interactions between the cationic amine group and the anionic DNA backbone. Substitution on a β-alanine (β-Ala or β) linker has also been shown to influence the tendency of a polyamide to bind in an extended versus a hairpin conformation [19]. Covalently linking the C- and N-termini of a hairpin polyamide eliminates all possibility of extended binding, and such cyclic polyamides have slightly lower specificities but higher affinities for their counterparts [5], but the structural basis of this difference has remained elusive. Partial melting of the T•A base pair recognized by an Hp/Py ring pair was thought to be a possible explanation [10], but the distortion was observed in only one of two crystal structures. A consistent lengthening of the amide–DNA hydrogen bond on the C-terminal side of the Hp residue is observed in both
target DNA sequences compared to analogous hairpin molecules bearing the same number of cationic groups (Figure 4) [20,21]. These differences can be attributed primarily to changes in the rates of dissociation from DNA, as the association rates for hairpin and cyclic polyamides are essentially diffusion limited [22].

Polyamides also can be linked, via the ring nitrogens, with an alkyl spacer that projects away from the minor groove. When placed in the center of a polyamide, the resultant branched molecule has been termed an H-pin; when placed at the end, a U-pin (Figure 4). H-pin polyamides bind with high affinity and good specificity, shown by an approximately 50-fold lower affinity for single base pair mismatch sites — that is, for sites constructed with one target Watson–Crick base pair replaced by a disfavored base pair [23]. Recent efforts to improve the synthetic methods for H-pins have enabled a detailed study of the optimal alkyl linker length, demonstrating that four and six methylene units provide the highest affinity [24]. U-pin polyamides behave similarly (A Heckel, PB Dervan, unpublished data). The affinity of an eight-ring U-pin is most comparable to that of a hairpin polyamide with six rather than eight rings, probably due to a loss of two hydrogen-bond donors upon removal of the γ-turn element. Thus, the dimeric Py-Im U-turn element (Figure 4) may be thought of as a C·G-specific replacement for the γ-turn. In combination with removal of the β-Ala tail (see below), H-pin and U-pin polyamides could potentially bind purely G,C sites, a sequence type that it has not been possible to target with other polyamide motifs.

Certain DNA sequences (including G,C tracts) have been challenging sites for high-affinity recognition by hairpin polyamides. For example, sites containing the sequence 5'-GNG-3' are often bound relatively poorly. Structural data have provided insight in certain cases. High-resolution crystal structures of different polyamide dimer–DNA complexes consistently display a large negative propeller twist in all targeted base pairs, an orientation that generally favors the formation of three-center hydrogen bonds in the major groove [8,10,11]. Such intraduplex bonds cannot be formed in sequences 5'-GCG-3' and 5'-GAT-3'. Accordingly, it has been problematic to target these sequences with high affinity.

In the case of 5'-GNG-3' sequences, replacing the aromatic Py residue with a flexible β-Ala residue enhances the affinity (Figure 4). The β-Ala unit may allow the flanking Im rings to orient better, while relieving the requirement
Pairing rules for minor groove recognition. (a) Table indicating the pairing code for Py, Im and Hp. Plus and minus signs indicate favored and disfavored interactions, respectively. (b) Space-filling model of (ImHpPyPy-β-Dp)_2–5′-CCAGTACTGG-3′ (PDB code 407D). In the central binding region, adenosines are yellow and thymines blue. Hp is red, Py is off-white and Im is black. A schematic is shown to the right: the black and open circles represent Im and Py rings, respectively; red circles with an H represent Hp rings; diamonds represent β-Ala residues; and plus signs next to diamonds represent Dp residues. (c) Detailed views of the complex shown in (b), indicating structures of the Hp/Py pair interacting with the T*A base pair (top) and the Im/Py pair interacting with the G*C base pair (bottom). Dashed lines indicate interatomic distances between oxygen, carbon and nitrogen atoms. The Hp oxygen may form a favorable hydrogen bond with the adenine C2–H. As in this case, C–H hydrogen bonds are strongest between aromatic carbons adjacent to nitrogen atoms with oxygen hydrogen-bond acceptors [10]. Hydrogens involved in recognition were added using PC Spartan (Wavefunction, Inc.) and are colored white with gray outlines.

for propeller-twisted base pairs [11]. The β/Im pair is specific for C·G, whereas the β/Py and β/β pairs are specific for A·T over G·C base pairs [25,26]. For polyamides targeted to sites containing multiple 5′-GNG-3′ sequences, binding enhancement is particularly dramatic upon incorporation of β/ring pairs; ImβImPy-γ-ImβImPy-β-Dp (Dp = dimethylaminopropylamine) binds site 5′-TGC/GCA-3′ with 100-fold higher affinity than the purely ring/ring paired analog ImPyImPy-γ-ImPyImPy-β-Dp [26]. Incorporation of a β-Ala also enhanced the DNA-binding properties of a polyamide with an N-terminal pyrrole [27]. A Py/Im pair in this position generally displays poor selectivity for its target C·G pair, but flanking the Im residue with a β-Ala improves the selectivity of the polyamide. For example, PyβPyPyPy-γ-ImβImβ-Dp binds its target site, 5′-TCTAAGC-3′ with subnanomolar affinity and 5–25-fold weaker binding to single base pair mismatch sites (Figure 4).

**Binding site size**

For biological applications, binding site size may be critical because longer sequences would be expected to occur less frequently in the genome. Yet, beyond five contiguous rings, the binding affinity of polyamides decreases [28]. Crystal structures of polyamide–DNA complexes have consistently shown that the polyamide rise per residue matches the pitch of the B-DNA helix — that is, the spacing of the polyamide rings matches the spacing of the DNA base pairs [8,10,11]. However, polyamides, which are inherently crescent shaped, are slightly more curved than the minor groove of DNA, such that, beyond five consecutive rings, the shape of a polyamide is no longer complementary to DNA [8].

The flexibility of β-Ala can be used to relax the curvature of polyamides, and molecules designed to bind as overlapped homodimers can recognize 11 bp of DNA with subnanomolar affinities (Figure 5) [25]. Another motif utilizing dimerization to increase binding site size is the cooperative hairpin [29]. Both of these motifs require a palindromic target site and have the potential to bind in noncooperative modes, albeit with lower affinities. Tandem hairpin polyamides, linked either turn-to-turn or turn-to-tail, resolve both issues (Figure 5) [30,31]. To
on the six-membered ring and appears to impart a curvature that is complementary to DNA. Indeed, the classic minor groove-binding Hoechst dyes are composed of benzimidazole units, and several derivatives of these molecules have been prepared [36,37]. We have recently incorporated benzimidazole derivatives into the backbone of hairpin polyamides in a manner that preserves critical hydrogen-bonding contacts and overall molecular shape (Figure 8) [38]. The hydroxybenzimidazole (Hz) and imidazopyridine (Ip) rings are introduced into polyamides as dimeric subunits PyHz and PyIp, respectively, in which the Py ring is directly connected to the benzimidazole derivative without an intervening amide bond. Preliminary DNase I footprinting indicates that the

![Schematic representation of extended polyamides at their binding sites. (a) Overlapped homodimer: fully overlapping extended homodimer recognizing 11 bp [25]. (b) Cooperative hairpin: a cooperatively binding hairpin polyamide can extend the binding site size to 10 bp without an increase in the molecular weight of the ligand [29]. (c,d) Tandem hairpins: linked either turn-to-turn (c) or turn-to-tail (d), tandem polyamides recognize large DNA sequences with good specificity and excellent binding affinity [30,32].](image)

Figure 5

Critical hydrogen-bonding contacts and overall molecular shape (Figure 8) [38]. The hydroxybenzimidazole (Hz) and imidazopyridine (Ip) rings are introduced into polyamides as dimeric subunits PyHz and PyIp, respectively, in which the Py ring is directly connected to the benzimidazole derivative without an intervening amide bond. Preliminary DNase I footprinting indicates that the

![Novel five-membered heterocyclic amino acids that have been incorporated into hairpin polyamides. All residues are shown with the functionality that faces the DNA minor groove towards the bottom right.](image)

Figure 6
replication by more than 99%, with no significant decrease in cell viability. Inhibition of viral replication is indirect evidence of specific transcription inhibition by polyamides, because other modes of action could be involved, such as modulation of T-cell activation pathways. However, RNase protection assays indicated that the two polyamides did not alter the RNA transcript levels of several cytokine and growth factor genes, suggesting that polyamides do affect transcription directly [47].

This early biological result spurred a variety of biochemical studies of the interactions of various polyamides with the basal transcription machinery and TF–DNA complexes. Two studies have used promoter scanning to identify sites where polyamide binding inhibits transcription [48,49]. The method uses a series of DNA constructs with designed polyamide-binding sites at varying distances from the transcription start site. Essential minor groove contacts were identified for a subunit of TFIIIB (possibly TBP) in a Xenopus tRNA promoter [48], as well as for TFIIID–TFIIA and TBP in the HIV-1 core promoter [49]. The binding of the homodimeric basic-helix-loop-helix TF Deadpan was investigated using a variant of promoter scanning [50]. A series of duplex oligonucleotides based on a Drosophila neural promoter were designed, incorporating polyamide-binding sites on different sides of the Deadpan recognition sequence and in different orientations. The TF–DNA complex was inhibited only by a polyamide binding upstream of the homodimer, establishing an asymmetric binding mode for this TF.

The binding of Ets-1 to the HIV-1 enhancer was examined in greater detail, and polyamides were shown to inhibit the formation of a ternary Ets-1–NF-κB–DNA complex [51]. Ets-1 is a winged-helix-turn-helix TF and its key phosphate contacts on either side of the major
Examples of DNA-binding proteins that have been inhibited by polyamides. Approximate \( K_d \) values are given next to each polyamide. Where available, the values are from experiments conducted on the depicted DNA sequence. The name of the protein is italicized above a shaded box indicating its DNA binding site. Open boxes indicate mismatches between the Dp tail and G,C base pairs [42]. The promoter or construct is identified below the DNA sequence and is followed by the reference number in brackets. The HIV-1 promoter construct, Xenopus tRNA promoter construct and Achaete-scute neural promoter construct are sample DNA sequences from studies that employed promoter scanning. Precise binding sites for the TFIIA–TFIID complex and for TFIIIB are not identifiable from promoter scanning. RPR represents an Arg-Pro-Arg C-terminus, and the half circle represents a propanolamine –NH(CH\(_2\))\(_3\)OH group. All other symbols are defined in Figure 4.
restricted by the structure and dynamics of nucleosomal DNA, and by the translational and rotational positioning of the histone octamer. Using six different hairpin polyamides, it was shown that sites on nucleosomal DNA facing away from the histone octamer, or even partially facing the octamer, are fully accessible [66]. Remarkably, one section of 14 consecutive base pairs—more than a full turn of the DNA helix—was accessible to high-affinity polyamide binding. The only positions very poorly bound by polyamides were sites near the N-terminal tails of histone H3 and histone H4. Removal of either tail allowed polyamides to bind, suggesting that the structure of the DNA and perhaps its rotational position are strongly influenced by the N-terminal tails of histones H3 and H4 [66].

Subsequently, the structures of three of these polyamide–nucleosome core particle complexes were determined by X-ray crystallography (Figure 13) [67**]. The histone octamer is unaffected by polyamide binding, but the nucleosomal DNA undergoes significant structural changes at the ligand-binding sites and adjacent regions. Significantly, distortions in DNA twist can propagate over long distances without disrupting histone–DNA contacts, giving a potential mechanistic rationale for the role of twist diffusion in nucleosome translocation. Although the three polyamides display very similar affinities for their binding sites in the α-satellite nucleosome particle (K_d ~1 nM), only the relatively nonspecific polyamide ImPy-PyPy-γ-PyPyPyPy-β-Dp inhibits temperature-induced nucleosome translocation [67**]. This may indicate that ligand positioning is critical, such that a single properly placed polyamide would effectively block translocation, or that the small effects of a single bound ligand can be amplified, such that a combination of several different polyamides would block translocation.

Although polyamides can block transcription by targeting promoter elements, they do not affect transcription when