Chapter 7

Discrimination of Adjacent T•A Base Pairs by *N-Terminal* 3-Chlorothiophene-Hydroxybenzimidazole Dimers

The work described in this chapter was done in collaboration with R. M. Doss and M. A. Marques (Dervan group; Caltech), and represents preliminary findings from an ongoing research effort.

Abstract

Perhaps as important as expanding the repertoire of DNA sequences that can be targeted with polyamides is the refinement of the recognition code that allows predetermined sequences to be recognized specifically. The context dependence of polyamide recognition and the sequence dependence of DNA microstructure imply that there may not be a general solution for discrimination of all four base pairs. Rather, a combination of different polyamide motifs and distinct classes of recognition elements, tailored with particular targets in mind, may offer the greatest utility in biological applications on the genome scale.

The chapter describes preliminary efforts to combine shape selective *N-terminal* thymine recognition of Ct with the more classical *internal* thymine discrimination of Hz. A series of Hz dimers, containing different N-terminal rings, was synthesized and characterized by DNase I footprinting. The Ct ring of the novel Ct-Hz dimer exhibited 10-fold thymine specificity when paired opposite Py, while the Im moiety of Im-Hz shows a dramatic loss of affinity. These results are explained in terms of the geometric influence of the proximal bicyclic ring on the angle of projection of the terminal ring within the minor groove.

Implications for dinucleotide recognition are discussed and future directions for research, including the design of new guanine recognition elements compatible with proximal Hz moieties, are described.

Background and Significance

The DNA recognition properties of *N-terminal* 3-chlorothiophene (Ct) and *internal* hydroxybenzimidazole (Hz) residues were described in Chapters 5 and 6. Individually, these novel residues represent powerful new tools for targeting predetermined DNA sequences in biological contexts; however, neither heterocycle fully addresses the limitations of thymine recognition by Hp. The shape selective nature of Ct recognition appears to be restricted to N-terminal pairings, and though the cellular uptake properties and high affinity of Ct/Py pairings are promising, their modest T•A selectivity leaves room for improvement.

The related internal 3-methylthiophene (Tn) residue also exhibits moderate thymine selectivity; however, attempts to improve the specificity of a hairpin polyamide by including multiple Tn residues resulted in nonspecific recognition of the minor groove. Internal Hz residues, on the other hand, showed comparable specificities and increased affinities, relative to Hp, in the context of multiple pairings, with the added benefit of increased stability in aqueous solutions. However, the dimeric nature of Hz residues precludes their application to recognition of adjacent thymine bases in the minor groove. Amelioration of the desirable recognition properties of Hz with the modular nature of Tn might allow consecutive thymine bases to be targeted with hairpin polyamides.

Chimeric Strategies for N-Terminal Thymine Recognition

A new class of N-terminal residues, combining the terminal shape selectivity of Ct with the internal specificity of Hz, was designed to facilitate specific recognition of adjacent T•A base pairs (Figure 7.1). The designed residues are best described as N-terminal dimers; however, replacement of the Ct moiety with an internal Tn derivative might allow this new dimeric motif to be adapted to internal contexts as well (Figure 7.1).

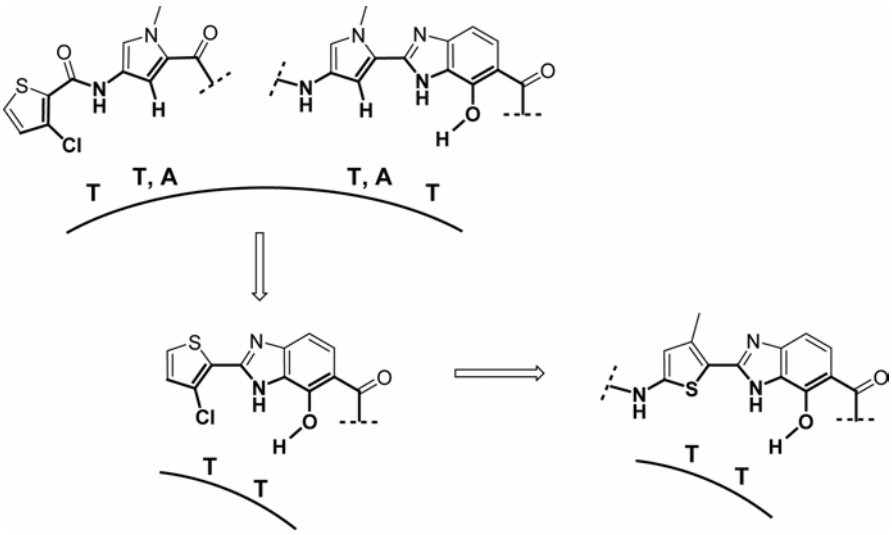


Figure 7.1 Design of N-terminal dimers using "chimeric" mode of thymine-recognition.

The first generation of chimeric dimers addresses the influence of a proximal Hz ring system on the specificity of N-terminal 3-chlorothiophene (**1**), 3-methylthiophene (**2**), and Im (**3**) residues. Dimeric building blocks were incorporated into the model hairpin polyamide described in Chapter 5, allowing the specificities of their terminal rings to be determined by quantitative DNase I footprinting (Figure 7.2).

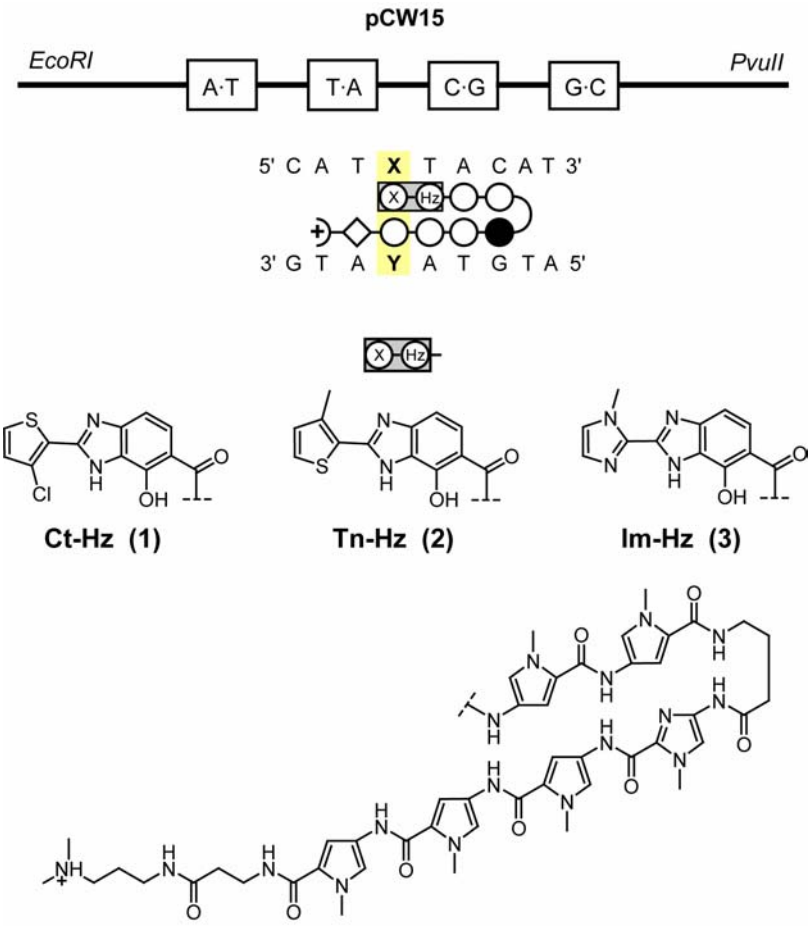


Figure 7.2 Plasmid design and model hairpin polyamides for characterization of novel N-terminal dimers.

Synthesis of Polyamides Containing Novel Dimeric Residues

Polyamides **1**-**3** were synthesized on PAM resin, by standard Boc-based solid phase techniques, using dimeric building blocks **8**-**10**. Im-Hz-OH **10** was prepared previously using the key diamine intermediate **4** (Figure 7.3). Acylation of **4** with 2-(trichloroacetyl)-N-methylimidazole gave monoamide intermediate **7** which was isolated by extraction and used without further purification. The monoamide was heated at 60 °C in glacial acetic acid and the resulting cyclized product was purified by column chromatography. Saponification of the purified ester then yielded **10**. Thiophene-containing dimers were synthesized by a

similar route, using HATU-activated carboxylic acid derivatives described in Chapter 5A (Figure 7.3).

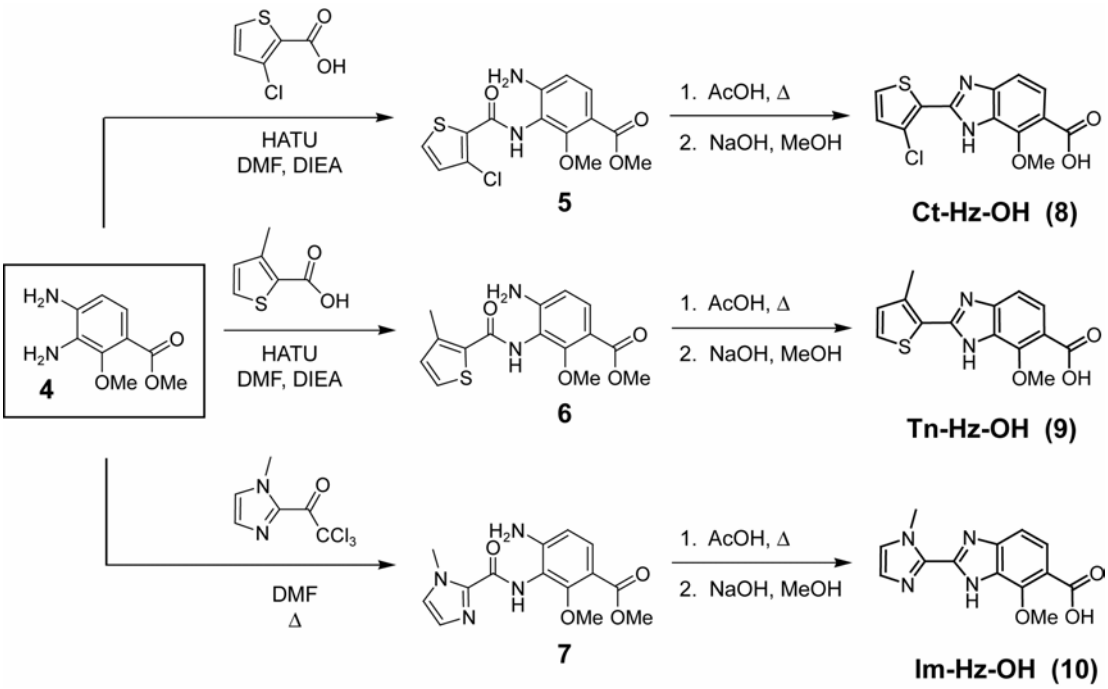


Figure 7.3 Synthesis of novel N-terminal Ct-Hz dimeric building blocks.

Polyamide synthesis on PAM resin is illustrated for **1** in Figure 7.4. Carboxylic acid **8** was coupled to support-bound polyamide intermediate **R1**, constructed by standard protocols, using HATU-mediated activation. Coupling efficiency was high and cleavage of the resin with neat Dp at 40 °C gave excellent recovery of polyamide intermediate **11** following preparative HPLC. This intermediate was treated with thiophenoxide at 100 °C to yield hairpin polyamide **1**. Polyamides **2** and **3** were also prepared according to this scheme. The purity and identity of polyamides was verified by analytical HPLC and MALDI-TOF mass spectrometry.

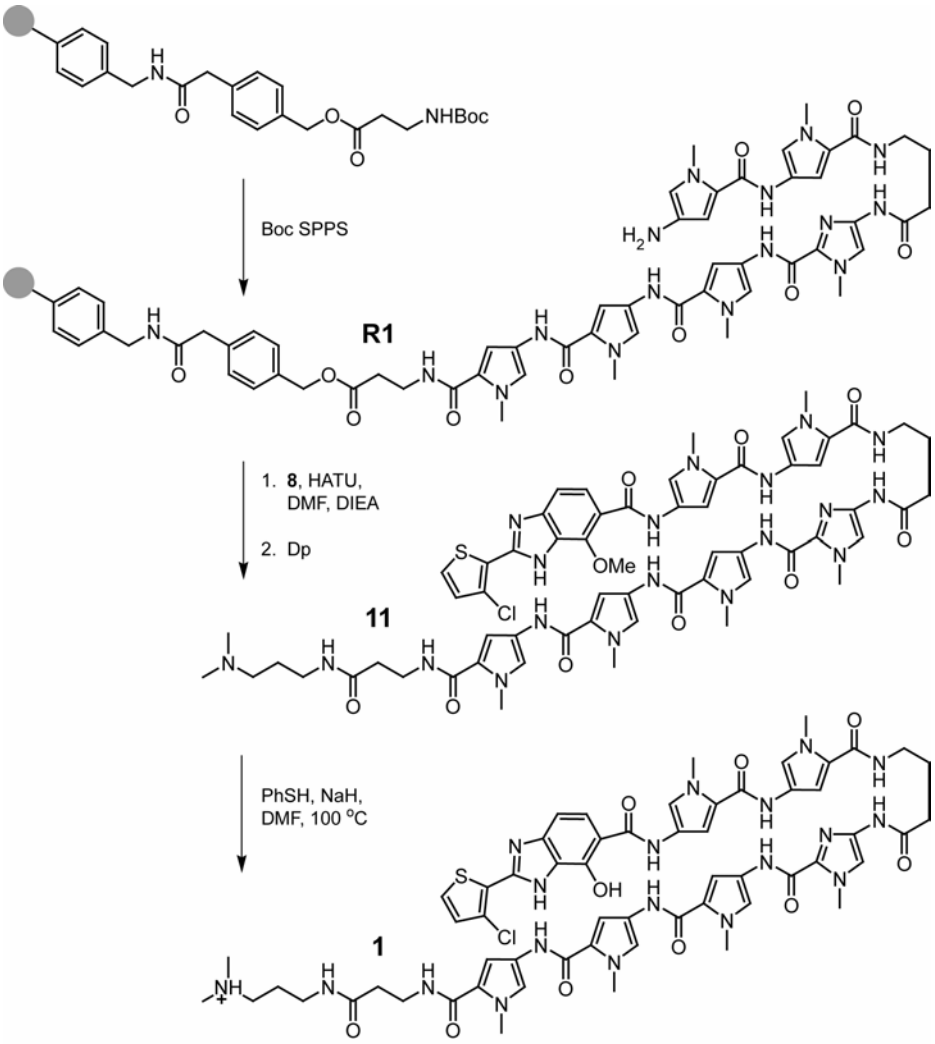


Figure 7.4 Solid phase synthesis of hairpin polyamides containing novel N-terminal dimers.

DNA Recognition Properties of N-Terminal Hz Dimers

The DNA-binding energetics of the novel dimeric residues were determined by quantitative DNase I footprinting titrations (Figure 7.5). This data, as well as the corresponding association constants for the corresponding N-terminal monomers from Chapter 5, is summarized in Table 7.1. The data indicates that the Hz scaffold has a dramatic influence on the specificity of proximal residues. Perhaps as surprising as the increased specificity of Ct-Hz is

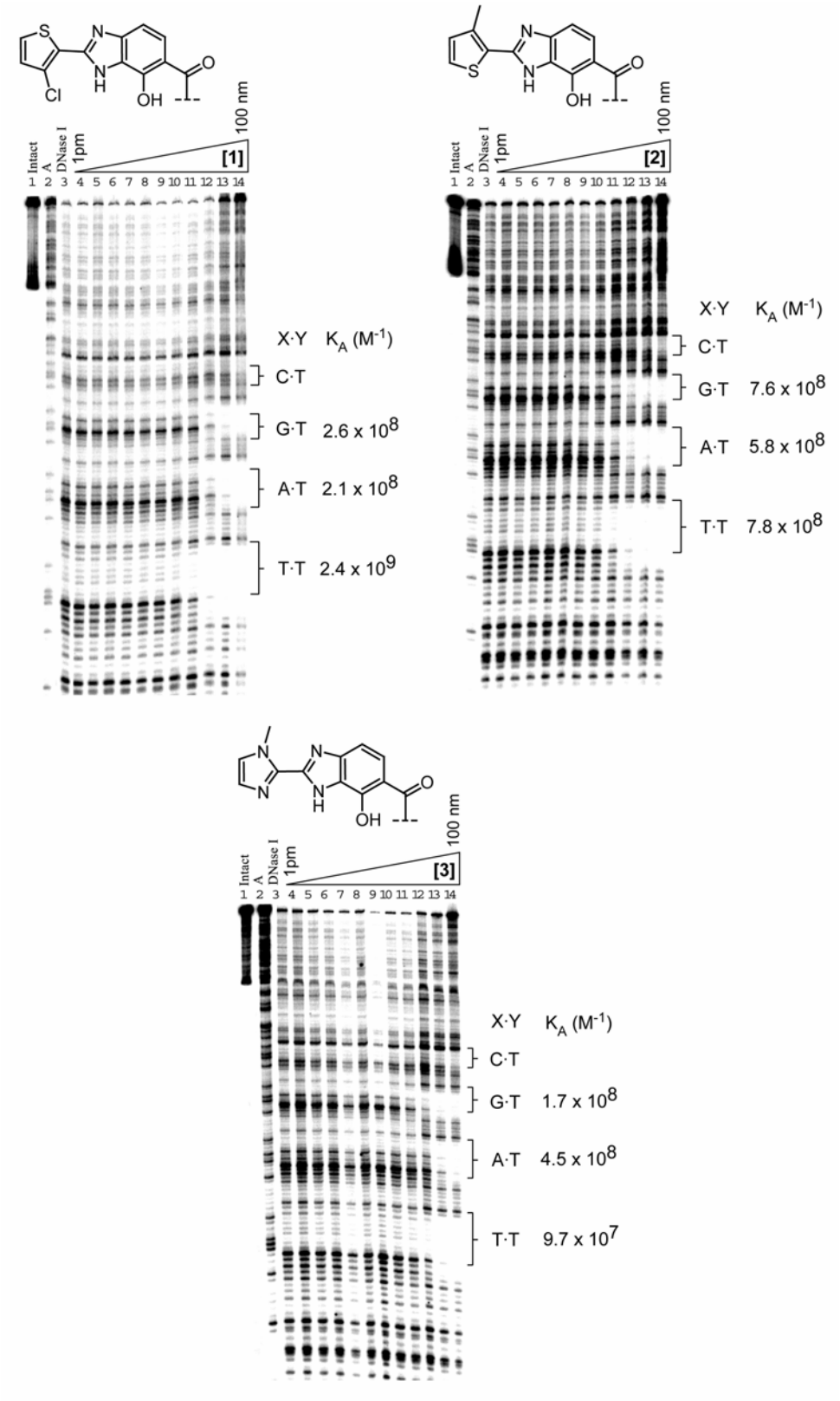


Figure 7.5 Representative DNase I footprinting gels for hairpin polyamides containing novel N-terminal dimers.

the loss of specificity for Im-Hz. This represents a reversal of the normal trends seen in minor groove recognition by polyamides, whereby thymine is targeted more specifically than guanine.

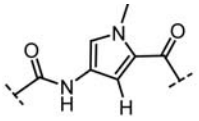
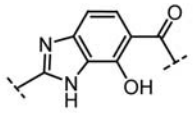


Table 7.1 Observed equilibrium association constants (M^{-1})

	K_A (T·A)	K_A (A·T)	K_A (G·C)	K_A (C·G)
	1.3×10^{10} M^{-1}	3.7×10^{9} M^{-1}	3.1×10^{8} M^{-1}	2.1×10^{8} M^{-1}
	2.3×10^{9} M^{-1}	1.4×10^{9} M^{-1}	1.0×10^{7} M^{-1}	1.0×10^{7} M^{-1}
	3.8×10^{9} M^{-1}	2.8×10^{9} M^{-1}	7.0×10^{10} M^{-1}	3.2×10^{9} M^{-1}



	K_A (T·A)	K_A (A·T)	K_A (G·C)	K_A (C·G)
	2.4×10^{9} M^{-1}	2.0×10^{8} M^{-1}	2.6×10^{8} M^{-1}	$\leq 1.0 \times 10^{7}$ M^{-1}
	7.8×10^{8} M^{-1}	5.9×10^{8} M^{-1}	7.4×10^{8} M^{-1}	$\leq 1.0 \times 10^{7}$ M^{-1}
	9.5×10^{7} M^{-1}	4.0×10^{8} M^{-1}	1.6×10^{8} M^{-1}	$\leq 1.0 \times 10^{7}$ M^{-1}

a) Values reported are the mean values from two DNase I footprinting titration experiments. b) Assays were performed at 22 ^{o}C in a buffer of 10 mM Tris.HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$ at pH 7.0.

Proximal Residue Effects in Minor Groove Recognition

A possible explanation for the unexpected results was sought with the aid of molecular modeling (Figure 7.6). N-terminal Ct, Tn, and Im rings were modeled as dimers with Py or Hz. Both possible rotational conformers of the terminal rings were examined when adjacent to Hz and Py, though no significant differences in rotational preference were observed. Replacing a Py residue with Hz does slightly reduce the solvent exposed surface area of the ligand; however,

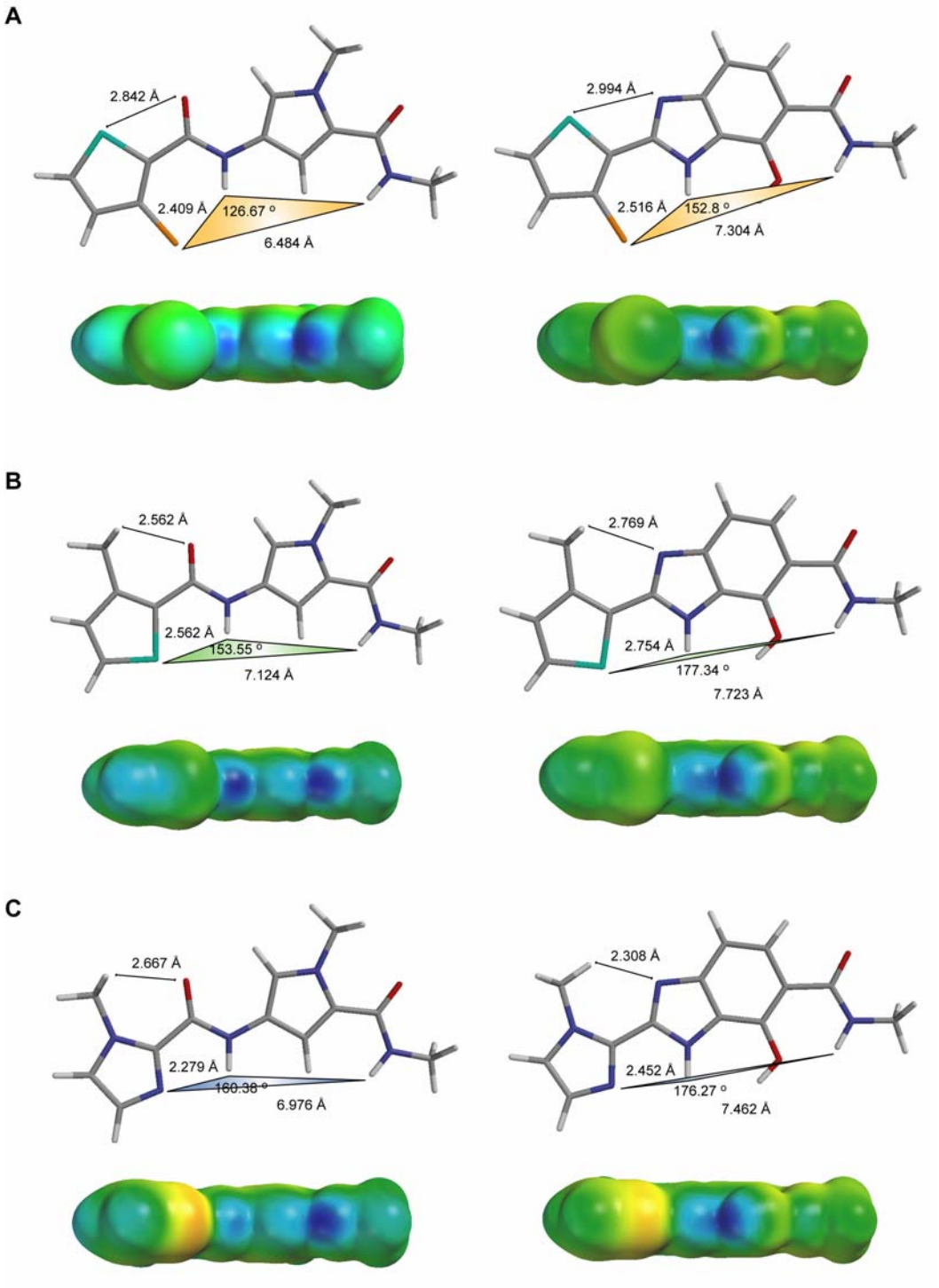


Figure 7.6 Inherent structural and geometric differences between classic five-member heterocycles for minor groove recognition and next generation bicyclic heterocycles. Molecular models were constructed and analyzed using the Spartan Essential software package. N-terminal dimers of Ct, Tn, and Im with Py (*left*) and Hz (*right*) were modeled as both possible N-terminal rotamers and the preferred conformer is shown. Electropotential surfaces for the minor groove-directed edges of each dimeric residue are included. **(A)** Ct-Hz. **(B)** Tn-Hz. **(C)** Im-Hz.

the overall molecular shape and electronic character of the groove-oriented edge of the polyamide are preserved. The predominant physical influence the bicyclic Hz ring system on adjacent terminal rings is geometric in nature. The reduced curvature of benzimidazole, relative to Py, results in a more shallow projection of terminal recognition elements toward the floor of the minor groove (Figure 7.7). Terminal Tn and Im moieties rely on intimate contacts between their ring heteroatoms and the floor of the minor groove for specific recognition. Im, in particular, requires a particular geometry of projection within the groove to allow hydrogen bond formation with the exocyclic amine of guanine. Thus, partial removal of terminal Tn and Im rings from the minor groove by the preceding Hz moiety might account for the reduced specificity of **2** and **3**.

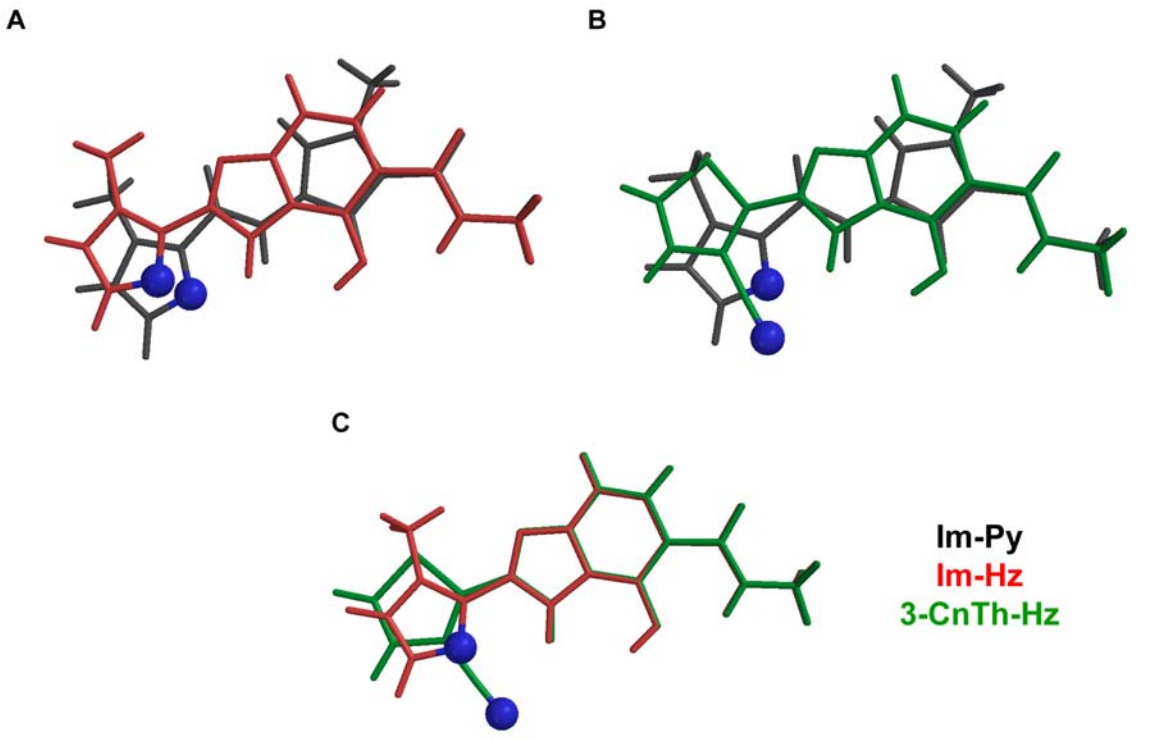


Figure 7.7 Molecular overlays comparing Im-Hz and 3-cyanothiophene-Hz terminal dimers to Im-Py. Minimized equilibrium geometries were generated using Spartan Essential software package. N3 of Im is represented by a blue sphere. **(A)** Im-Py (*black*) overlaid with Im-Hz (*red*). **(B)** Im-Py (black) overlaid with 3-CnTh-Hz (*green*). **(C)** Im-Hz overlaid with 3-CnTh-Hz.

The recognition properties Ct-Hz are much tougher to rationalize, however, a geometric effect similar to that noted above, is probably a major factor. The greater depth of groove projection in Ct-Py dimers might result in distortion of targeted T•A base pairs, allowing more efficient burial of the chlorine atom. While this effect might result in higher affinity for hydrophobic ligands, it would also be expected to perturb the asymmetric cleft, resulting in decreased specificity. The reduced curvature of Ct-Hz moieties may alleviate this effect, providing milder mode of shape selective recognition of the asymmetric cleft, albeit with lower affinity as a result of decreased burial within the groove. Changes in the angle of chlorine projection might also create unfavorable contacts with the walls of the minor groove, leading to lower affinities.

Towards a Recognition Code for N-Terminal Dinucleotide Sequences

The implications of novel residue development discussed earlier in the context of N-terminal monomers (Chapter 5B) are even more profound for the dimers described above. The unprecedented thymine specificity of *N-terminal* Ct-Hz illustrates the recurring theme of context dependence in polyamide-based recognition of the minor groove. The replacement of a degenerate Py ring with a specific Hz residue enhances the specificity of the polyamide directly by targeting thymine and indirectly by tuning the specificity of the terminal Ct ring. All eight dinucleotide sequences that begin with a T or G can be targeted by combinations of Py and Im; however, such combinations are not always specific with respect to both bases.

The results described above imply that a recognition code for dinucleotide sequences might further expand the array of sequences amenable to polyamide design while enhancing the specificity of their recognition, especially those containing multiple T•A/A•T base pairs (Figure 7.8). The desirable properties of Ct-Hz might be a direct result of the bicyclic proximal ring, allowing Ct-Bi and Ct-Ip to specifically target 5'-T(A,C) or 5'-TG, respectively. While this proximal residue effect appears well suited to 3-chlorothiophene, it also demands that a new guanine recognition element be developed to allow 5'-GT sequences to be recognized with the same fidelity as 5'-TT.

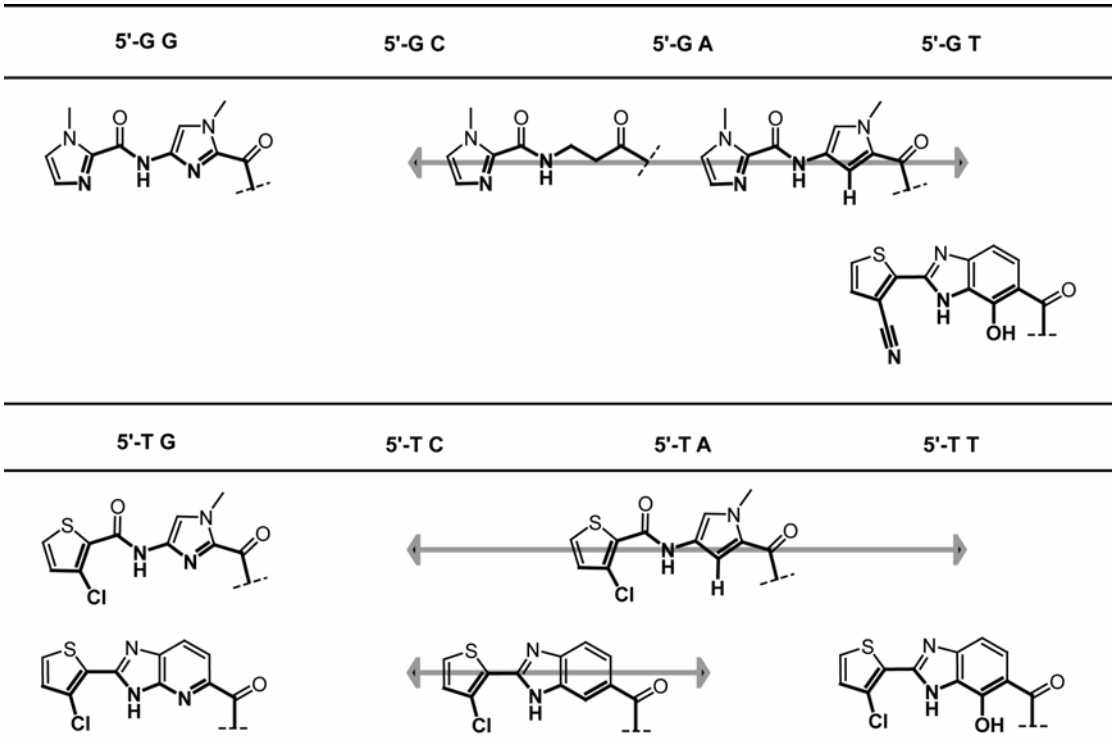


Figure 7.8 New classes of N-terminal dimers for specific recognition of dinucleotide sequences within the context of hairpin polyamides.

Future Directions—N-Terminal Im Replacements

One possible Im alternative could replace the chlorine substituent of Ct with a nitrile group. Appending the hydrogen bond acceptor to the terminal ring might compensate for the decreased curvature of proximal Hz residues. Overlays generated from molecular models of imidazole- or 3-cyanothiophene-Hz dimers with Im-Py illustrate this design concept (Figure 7.7). The ability of 3-cyanothiophene to recognize guanine was mentioned in Chapter 5B. Other clear alternatives include bicyclic heterocycles containing nitrogen heteroatoms such as isoquinolines or benzimidazoles. The search for Im replacements in the above context is currently underway.

The base pair specificity of the proximal residue should also be addressed with a plasmid that allows all four Watson-Crick base pairs to be varied at the (N-1) position of the binding site. Structural studies on the bicyclic residues and the N-terminal Ct residue are also underway and these results will undoubtedly provide further clues for the design of next generation minor groove-binding ligands.