Chapter 3

Toward a Chemical Etiology for DNA Minor Groove Recognition

The text of this chapter was taken in part from a manuscript coauthored with Raymond M. Doss, Adam R. Urbach, and Professor Peter B. Dervan (Caltech)

(Marques, M. A.; Urbach, A. R.; Doss, R. M. and Dervan, P. B. "*Toward A Chemical Etiology for DNA Minor Groove Recognition*" *Helvetica Chimica Acta* **2002,** 85, (12), 4485-4517.

Abstract.

Polyamide ligands composed of N-methylimidazole (**Im**), N-methylpyrrole (**Py**), and 3-hydroxypyrrole (**Hp**) amino acids recognize predetermined DNA sequences in 2:1 and 1:1 ligand:DNA complexes. DNA sequence specificity has been attributed largely to the unique functionality presented by each heterocycle to the floor of the minor groove. However, little has been done to assess the ramifications of functional groups projecting away from the DNA. Here we probe the question of whether we can understand the relationship between heterocycle structure and DNA sequence specificity for a family of five-membered aromatic heterocyclic amino acids incorporated into hairpin and 1:1 polyamides. Using a combination of quantitative DNase I footprinting and computational molecular modeling, the recognition behavior of polyamides containing 1-methylimidazole (**Im**), 1-methylpyrrole (**Py**), 3-hydroxypyrrole (**Hp**), 1H-pyrrole (**Nh**), 1-methylpyrazole (**Pz**), 4-methylthiazole (**Th**), 5-methylthiazole (**Nt**), 4-methylthiophene (**Tn**), 3-hydroxythiophene (**Ht**), and furan (**Fr**) amino acids at a single position were compared for their ability to discriminate between the four Watson-Crick base pairs in the DNA minor groove. We conclude that curvature complementarity between the polyamide and DNA, which is governed by the bonding geometry inherent to each heterocycle, plays a significant role in the ability of polyamides to differentiate between DNA sequences. Curvature is affected by all atoms in the heterocycle, and therefore one must consider the entire structure for an accurate prediction of its DNA recognition properties.

Introduction.

In the post-genome era where many diseases are related to specific gene expression profiles, the ability to reprogram a cell could have profound implications for human medicine. Synthetic ligands designed to bind specifically to predetermined DNA sequences offer a potentially powerful approach to gene regulation. Polyamides based on analogues of the 1-methylpyrrole-carboxamide ring (Py) of the natural products netropsin and distamycin A bind in the minor groove of DNA in 2:1 and 1:1 ligand-DNA

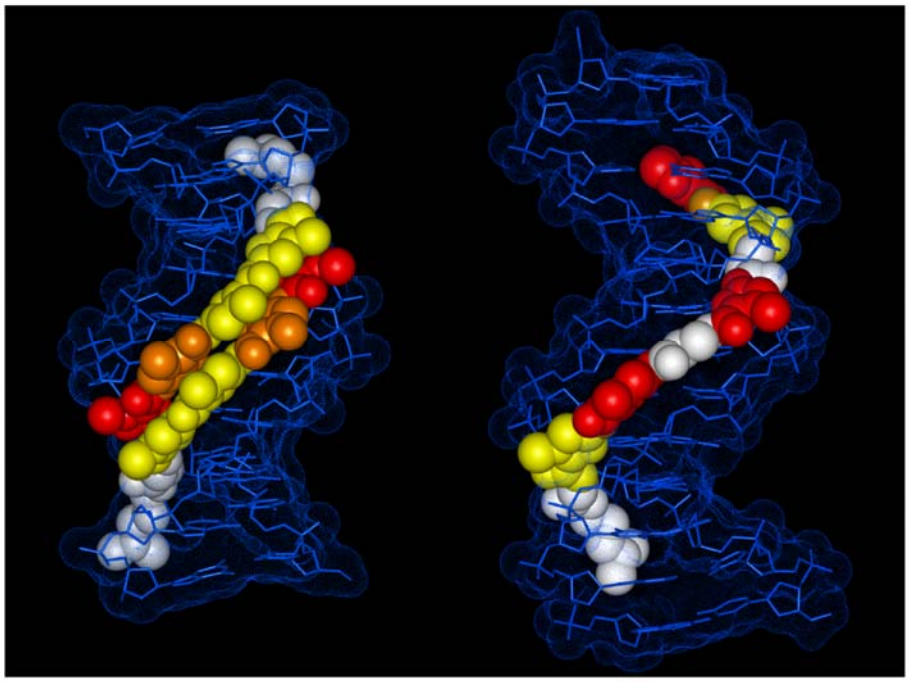


Figure 3.1. High-resolution structures of polyamides bound to DNA. (left) 2:1 motif determined by X-ray crystallography.[1] (right) 1:1 motif determined by NMR.[3] DNA is shown as a stick model in blue. Polyamides are shown as space-filling models, with imidazole residues in red, hydroxypyrrole in orange, pyrrole in yellow, and aliphatic residues in white.

complexes (Figure 3.1).[5, 6] Py is known to be specific and degenerate for A•T and T•A base pairs due to steric exclusion of the guanine amino group (G-NH2).[7, 8] However, base pair specificity can be altered by changing the functional group presented to the

floor of the DNA minor groove. For example, 1-methylimidazole (Im) presents the DNA with nitrogen and its lone pair sp^2 orbital, which can accept a hydrogen bond from G-NH2.[3, 9-11] Additionally, 3-hydroxypyrrole presents a hydroxyl group that can donate a hydrogen bond to the O2 of thymine.[1] Using these concepts, Im and Hp can be paired opposite Py in an antiparallel side-by-side 2:1 complex such that Im/Py is specific for G•C, Py/Im for C•G, Hp/Py for T•A, and Py/Hp for A•T.[1, 6, 11]

These pairing rules have proven useful for the recognition of hundreds of DNA sequences by designed polyamides. However, sequence-dependent DNA structural variation is thought to reduce binding affinity and specificity at numerous DNA sequences, which leads us to question whether new amino acid residues could be developed to improve recognition at difficult sequences. Moreover, we find that the Hp residue can degrade oxidatively over many hours, leading us to search for a stable replacement that can discriminate A•T from T•A base pairs. Five-membered heterocyclic residues other than Py, Im, and Hp have been investigated previously, including thiazole, pyrazole, and furan, with little improvement observed.[12-14]

Figure 3.2 shows a family of 5-membered, aromatic, heterocyclic residues grouped in columns by the type of functionality directed toward the floor of the DNA minor groove. Py, pyrazole (Pz), and 1H-pyrrole (Nh) project a hydrogen with positive potential toward the DNA; Im, 5-methylthiazole (Nt), and furan (Fr) project an sp^2 lone pair from nitrogen or oxygen; Hp and 3-hydroxythiophene (Ht) project a hydroxyl group; and 4-methylthiazole (Th) and 4-methylthiophene (Tn) project an sp^2 lone pair from sulfur. Comparative analysis of new residues within this 5-membered heterocyclic framework should enable us to retain overall ligand morphology and to observe the

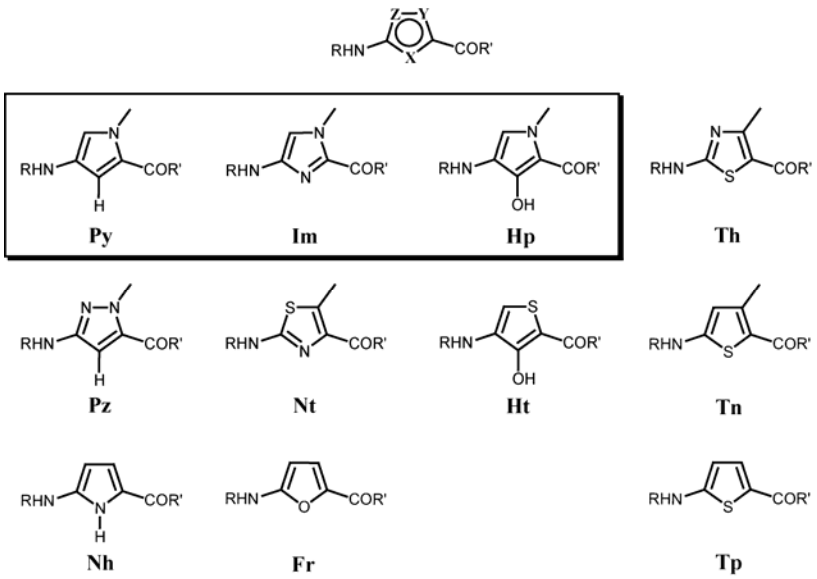


Figure 3.2 Family of heterocyclic amino acids studied here. Centered above is a schematic showing the five-membered heterocyclic framework with the variable positions labeled X, Y, and Z. The parent Im, Py, and Hp residues are boxed. All residues are shown with the functionality that faces the DNA minor groove pointed down (X).

effects of small structural changes, such as single atom substitution, on DNA base pair specificity. Furthermore, it is interesting to note changes in sequence specificity deriving from the substitution of heterocyclic atoms projecting *away from* the DNA minor groove.

The covalent head-to-tail linkage of polyamide subunits in a 2:1 complex results in the formation of hairpin ligands with increased DNA affinity and sequence specificity.[15, 16] These effects are thought to be due to the limitation of slipping between the stacked subunits,[17] which "locks" the individual ring pairings into place. Therefore, the hairpin motif provides a well-controlled system for studying the DNA recognition characteristics of new ring pairings. We have reported previously the sequence specificities of Py/Py, Hp/Py, Pz/Py, and Th/Py pairings at a single position within the hairpin polyamide ImIm**X**Py-γ -Im**Py**PyPy-β-Dp (**X** = Py, Hp, Pz, and Th; γ = gamma amino butyric acid; β = beta alanine; Dp = dimethylaminopropylamide) against the four Watson-Crick base pairs within the sequence context 5'-ATGG**X**CA-3' (**X** = A, T, G, and

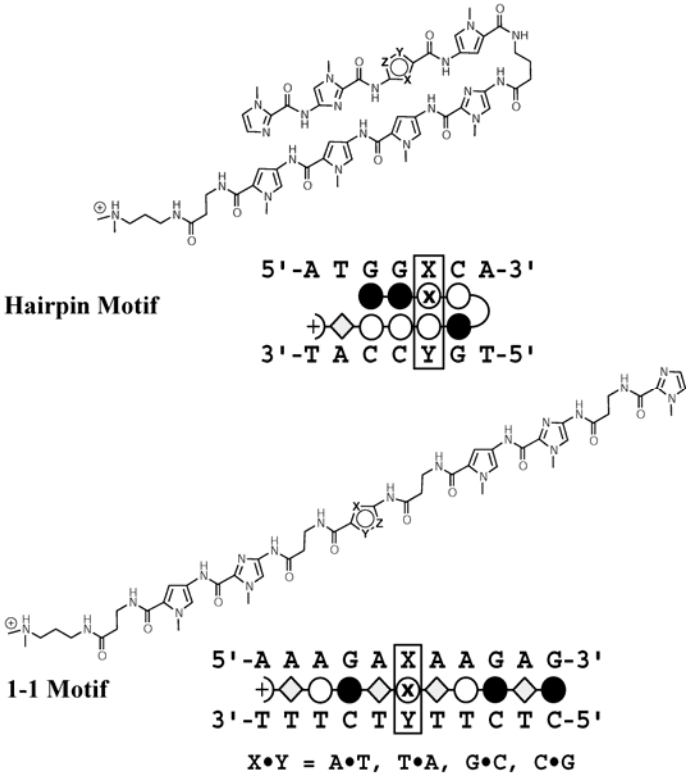


Figure 3.3. Schematic illustrating the examination of sequence selectivity against the four Watson-Crick base pairs within hairpin (top) and 1:1 (bottom) motifs. Each chemical structure has a variable residue containing X-, Y-, and Z-labeled positions, which are designated in Figure 3.2. The dot models shown below each chemical structure illustrate the binding mode with the polyamide shown inside its target DNA sequence. Imidazoles and pyrroles are shown as filled and non-filled circles, respectively; Beta alanine is shown as a diamond; the gamma-aminobutyric acid turn residue is shown as a semicircle connecting the two subunits; and the novel heterocycles are indicated by a circle containing an X.

C).[12, 18] Here we provide an extension of this study to include Im, Nh, Tn, Ht, and Fr residues in the same hairpin context (Figure 3.3). Additionally, DNA sequence specificity in the 1:1 motif has been addressed recently for polyamides containing Im, Py, Hp, and β residues.[19] In order to provide a thorough comparative analysis of heterocycle behavior against the four Watson-Crick base pairs in both motifs (Figure 3.3), we report here the synthesis and analysis of polyamides of type Im-β-ImPy-β-**X**-β-ImPy-β-Dp (**X** = Py, Hp, Nh, Ht, Fr, Nt, Tn, and Th) against DNA sequences 5'-AAAGA**X**AGAAG-3' (**X** = A, T, G, and C). Accordingly, the synthesis of Boc-protected amino acids suitable for solid-phase synthesis of the corresponding polyamides is presented.

Quantitative DNase I footprinting was used to determine the equilibrium association constant for each complex. *Ab initio* computational modeling of the heterocyclic amino acids was implemented to derive their inherent geometric and electronic parameters. The combination of these techniques has provided considerable insight to the origin of DNA sequence discrimination by polyamides. These ideas and their implications are discussed.

Results.

Monomer, Dimer, and Polyamide Synthesis. Polyamides were synthesized manually on solid support by the stepwise addition of monomeric and dimeric Boc-protected amino acids (Figures 3.4 and 3.5), as previously reported.[4] Hydroxythiophene (Ht) containing polyamides were prepared by deprotecting the 3-methoxy (Mt) analogue. Boc-protected amino acids for Im, Hp, Py, Pz, and Th monomers have been reported previously.[4, 12, 18] Syntheses of core amino-ring-alkyl ester (NH_2-X-OR) structures **1-4** for X = Fr, Nh, Tn and Mt are shown in Figures 3.6–3.8 and 3.10. Boc-protected monomeric amino acids **5-12** were suitable for solid-phase synthesis and prepared in two steps from their NH_2-X-OR analogues. However, furan and thiophene amines were unreactive to coupling on solid support, and therefore dimers **13-17** were pre-formed in solution under strong acylation conditions prior to solid-phase coupling. A synthetic synopsis for all Boc-protected amino acids used in this study follows.

Furan (Fr). The nitro-furan-ester (**18**) was prepared as previously reported. [20] While **18** is commercially available from Aldrich, it is expensive and readily synthesized in large quantities from methyl 2-furoate. The amino ester of furan (**1**) was synthesized from **18** using H_2 (500 psi) and palladium on carbon (Pd/C) and isolated as the free base

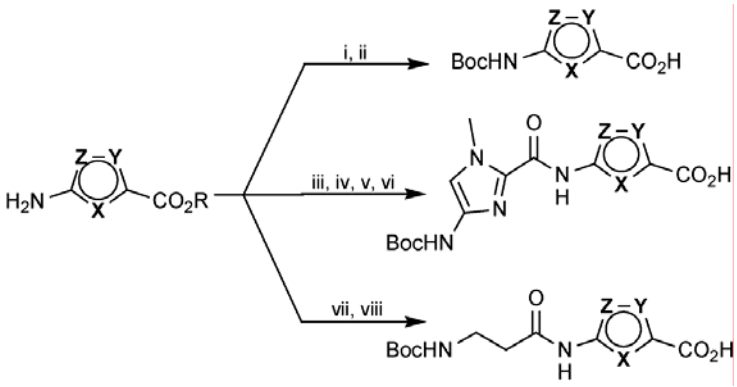


Figure 3.4. Formation of monomer and dimer Boc-protected amino acids. (i) (Py, X = C-H, Y = N-Me, Z = C-H, R = Me), (Im, X = N, Y = N-Me, Z, = C-H, R = Et), (Op, X = C-OMe, Y = N-Me, Z = C-H, R = Et), (Th, X = S, Y = C-Me, Z = N, R = Et), (Pz, X = C-H, Y = N-Me, Z = C-H, R = Et), (Nt, X = N, Y = C-Me, Z = S, R = H), and (Nh, X = N-H, Y = Z = C-H, R = Et), $(Boc)_2O$, DIEA, DMF, 60 ^oC, 12-18 h.; or (Hm, X = C-OMe, Y = S, Z = C-H, R = Me) $(Boc)_2O$, DIEA, TEA CH_2Cl_2, 60 ^oC, 12-18 h; (ii) (Py, X = C-H, Y = N-Me, Z = C-H, R = Me), (Im, X = N, Y = N-Me, Z, = C-H, R = Et), (Op, X = C-OMe, Y = N-Me, Z = C-H, R = Et), (Th, X = S, Y = C-Me, Z = N, R = OEt), (Pz, X = C-H, Y = N-Me, Z = C-H, R = Et), and (Nh, X = N-H, Y = Z = C-H, R = Et), 1N NaOH, MeOH, R. T., 3-4 h.; or (Nt, X = N, Y = C-Me, Z = S, R = H), 1N NaOH, MeOH, R. T., 1 h; or (Hm, X = C-OMe, Y = S, Z = C-H, R = Me), KOH, MeOH, 60 ^oC, 4-6 h; (iii) (Fr, X = O, Y = Z = C-H, R = Me), (Tn, X = S, Y = C-Me, Z = C-H, R = Me) and (Dt, X = S, Y = Z = C-H, R = Me), O_2N-Im-$COCCl_3$, EtOAc, DIEA, 35 ^oC, 10-12 h; (iv) (Fr, X = O, Y = Z = C-H, R = Me), (Tn, X = S, Y = C-Me, Z = C-H, R = Me) and (Dt, X = S, Y = Z = C-H, R = Me), H_2, Pd/C 10%, EtOAc, R. T., 1.5 h; (v) (Fr, X = O, Y = Z = C-H, R = Me), (Tn, X = S, Y = C-Me, Z = C-H, R = Me) and (Dt, X = S, Y = Z = C-H, R = Me), $(Boc)_2O$, DIEA, DMF, 60 ^oC, 12-18 h; (vi) (Fr, X = O, Y = Z = C-H, R = Me), 1N NaOH, MeOH, R. T., 3 h. (Tn, X = S, Y = C-Me, Z = C-H, R = Me) and (Dt, X = S, Y = Z = C-H, R = Me), 1N NaOH, MeOH, 60 ^oC, 4-6 h; (vii) (Fr, X = O, Y = Z = C-H, R = Me) and (Tn, X = S, Y = C-Me, Z = C-H, R = Me), $(Boc\text{-}\beta\text{-alanine})_2O$, DMF, DIEA, 40 ^oC, 12-18 h; (viii) (Fr, X = O, Y = Z = C-H, R = Me), 1N NaOH, MeOH, R. T., 3 h. (Tn, X = S, Y = C-Me, Z = C-H, R = Me), 1N NaOH, MeOH, 60 ^oC, 4-6 h.

from ethyl acetate by precipitation from hexanes (Figure 3.6). The free base is a stable crystalline solid at room temperature. The nitro-imidazole-furan ester (NO_2-Im-Fr-OMe, **19**) was prepared by condensing the nitro-imidazole-trichloroketone (NO_2-Im-$COCCl_3$) with (**1**) in ethyl acetate at 35 ^oC. The dimer product (**19**) is rather insoluble in ethyl acetate and begins to precipitate upon formation. Reduction of **19** to the ammonium

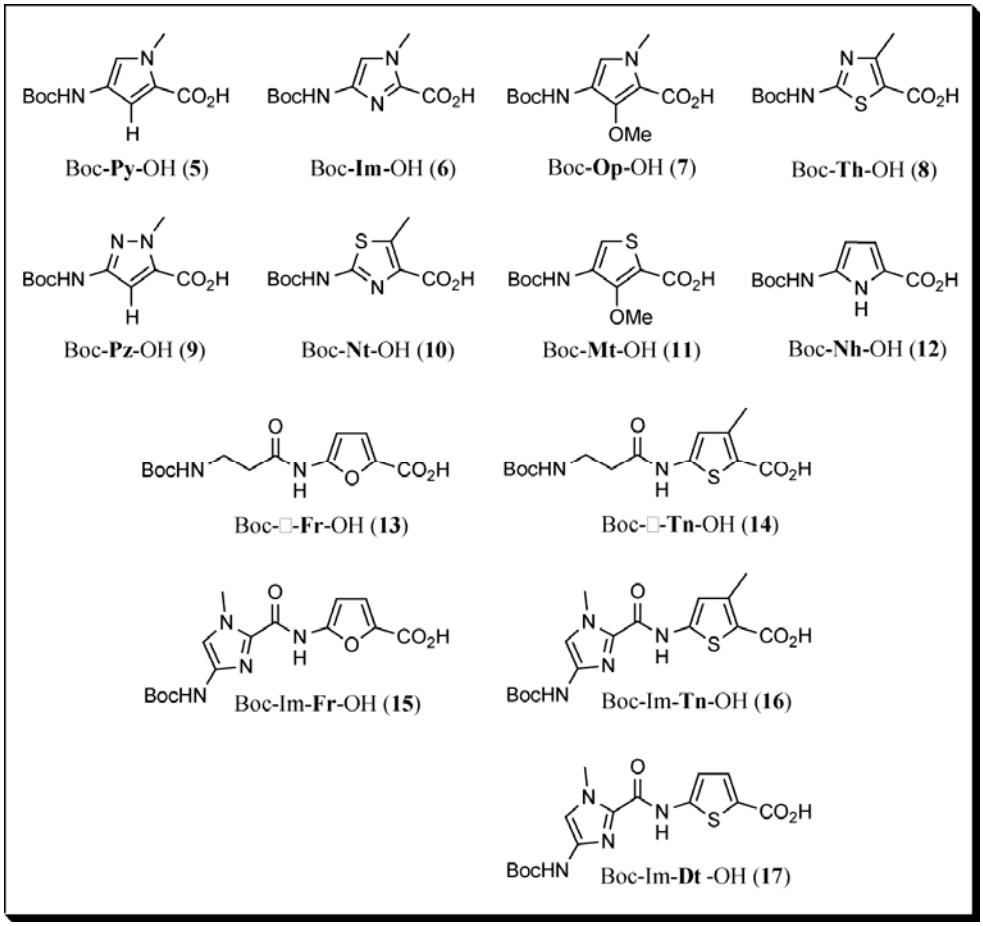


Figure 3.5. Monomeric and dimeric Boc-protected amino acids for solid phase synthesis.

chloride (HCl•H_2N-Im-Fr-OMe, **20**) was carried out using H_2 (500 psi) and Pd/C, followed by addition of 2M HCl in diethyl ether. **20** was Boc-protected using Boc-anhydride, DIEA and DMF at 60 ^oC for 12-18 h to give Boc-Im-Fr-OMe (**21**). Elevated temperatures and extended reaction times are necessary, presumably due to the poor nucleophilicity of the imidazole amine. Saponification of **21** using 1N aqueous NaOH and methanol at room temperature provided the target dimer, Boc-Im-Fr-OH (**15**).

Alternatively, the Boc-β-Fr-OMe dimer (**22**) can be synthesized from **1** by coupling to the symmetrical anhydride of Boc-beta-alanine in DMF, DIEA, and DMAP. The anhydride was pre-formed in minutes using DCC in methylene chloride at room

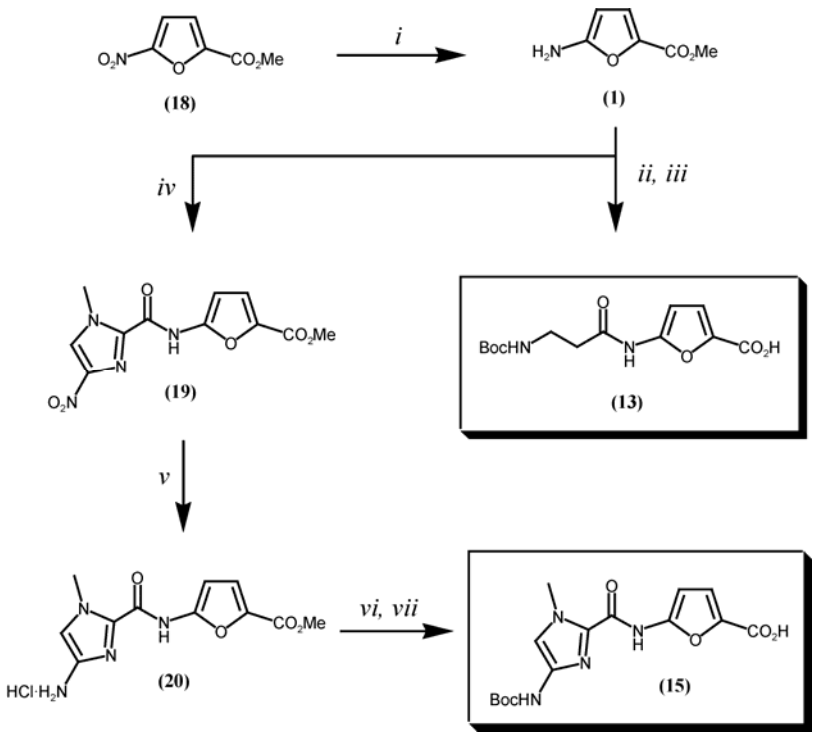


Figure 3.6. Synthesis of furan (**Fr**) Boc-protected amino acid monomers and dimers. i) H_2, Pd/C, EtOAc, 500 psi, r.t.; ii) (Boc-β-Ala)$_2$O, DMF, DIEA, DMAP, 40 ^oC, iii) 1N NaOH, MeOH r.t.; iv) NO_2-Im-$COCCl_3$, EtOAc, 35 ^oC, v) H_2, Pd/C, EtOAc; vi) (Boc)$_2$O, DIEA, DMF, 60 ^oC; vii) 1N NaOH, MeOH.

temperature, indicated by DCU precipetation. The furan amino ester **1** was then added as a solution in DMF and DIEA, followed by the transacylation catalyst DMAP. The 5-amino group of furan is significantly unreactive,[21] and attempts to couple it on solid support using reagents such as DCC and HOBT, HBTU, HATU, PyBrOP, PyBOP, and TFFH, were unsuccessful. Formation of the Boc-β-Fr-OMe dimer (**22**) occurs slowly at elevated temperature. Heating the reaction above 40 ^oC did not affect the rate of reaction. Several different solvent systems were tried, but DMF was found to be optimal. Saponification of **22** was accomplished using a mixture of MeOH and 1N NaOH at room temperature to provide the target Boc-β-Fr-OH (**13**), which is suitable for standard solid phase protocols.[4]

1-H Pyrrole (Nh). Pyrrole trichloromethyl ketone (**23**) was prepared by adding pyrrole to a mixture of trichloroacetyl chloride in diethyl ether at 0 ^oC, then warming to

room temperature and stirring overnight, followed by precipitation from hexanes. Nitration of **23** was accomplished using acetic anhydride and nitric acid at -40 ^{o}C to provide the 5-nitro regioisomer (**24**) as the major product. Regiocontrol of the nitration appears to depend significantly on the reaction temperature. Warmer temperatures provide unfavorable mixtures of 4-nitro and 5-nitro regioisomers. Esterification of the trichloroketone (**24**) was accomplished using ethanolic sodium ethoxide at room temperature to provide the nitro-pyrrole-ester (NO_2-Nh-OEt, **25**) in good yield. Reduction of **25** using H_2 (500 psi) and Pd/C, followed by the addition of 2M hydrogen chloride in diethyl ether, provided the ammonium chloride salt (HCl•H_2N-Nh-OEt, **2**), which was then Boc-protected using Boc-anhydride, DIEA and DMF to yield the Boc-pyrrole-ester (Boc-Nh-OEt, **26**). Saponification of **26** was accomplished using 1N NaOH and methanol at room temperature to provide the final monomer unit Boc-Nh-OH (**12**).

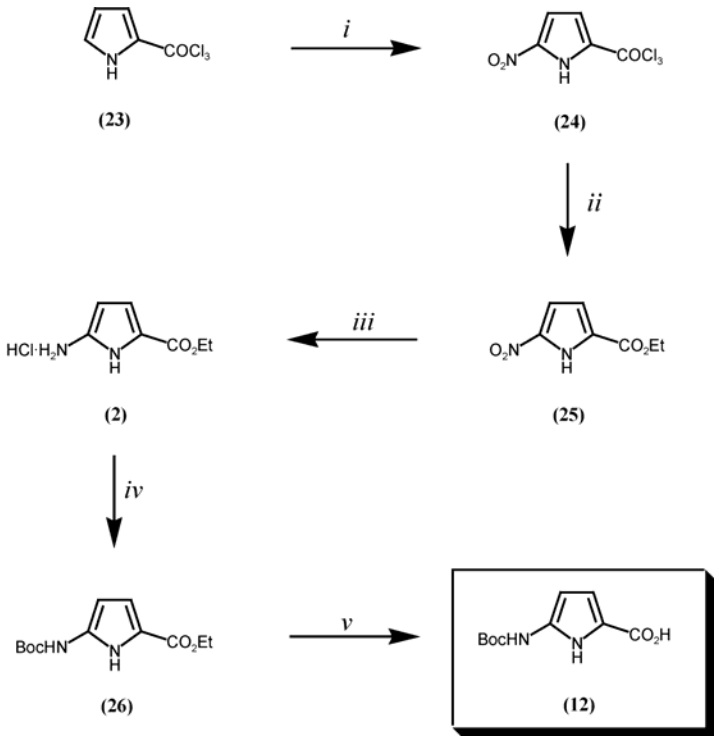


Figure 3.7. Synthesis of NH pyrrole (**Nh**) monomer. i) HNO_3, Ac_2O, -40 ^{o}C; ii) EtONa, EtOH, reflux; iii) H_2, Pd/C, DMF, HCl, Et_2O, r.t.; iv) $(Boc)_2O$, DIEA, DMAP, DMF, r.t.; v) 1N NaOH, MeOH, r.t.

Thiophene (Tn). The acyclic precursor to the thiophene ring system was prepared by a Knovenagel reaction involving acetoacetate and cyanoacetic acid, providing **27** as a mixture of E and Z regioisomers in moderate yield after vacuum distillation (Figure 3.8). Treatment of **27** with sulfur flakes and diethylamine in ethanol yielded the cyclized aminothiophene, and addition of concentrated hydrochloric acid precipitated the hydrochloride salt (HCl•H_2N-Tn-OMe, **3**). Formation of the nitro-imidazole-thiophene ester (NO_2-Im-Tn-OMe, **28**), followed by reduction and Boc-protection to provide HCl•H_2N-Im-Tn-OMe (**29**) and Boc-Im-Tn-OMe (**30**), respectively, was accomplished using the procedures described above for the furan compounds **19**, **20**, and **21**. The Boc-β-Tn-OMe dimer (**31**) was prepared, as described above for **22**, by coupling **3** with the symmetric anhydride of Boc-beta-alanine. The thiophene amine displays low reactivity comparable to furan. However, elevated temperature was necessary to completely saponify methyl esters **30**

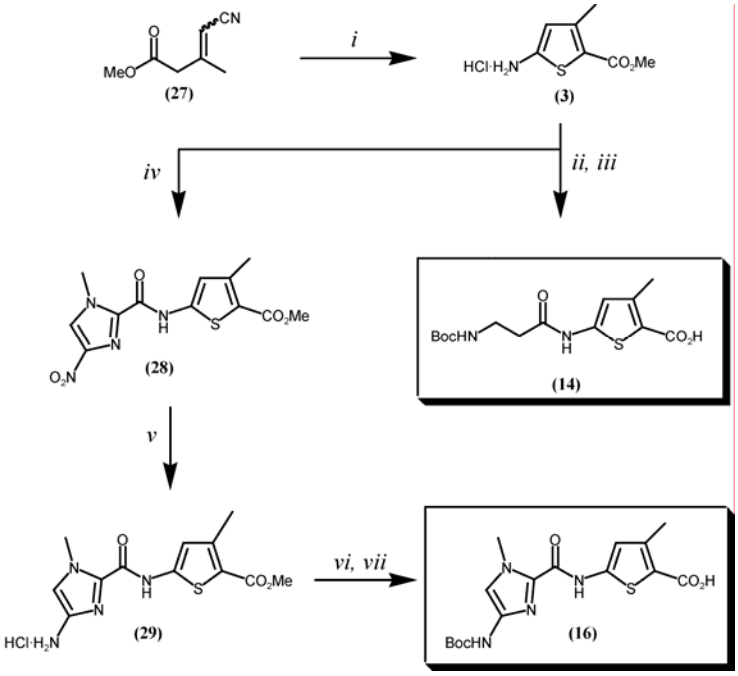


Figure 3.8. Synthesis of thiophene (**Tn**) Boc-protected amino acid monomers and dimers. i) S, Et_2NH, EtOH, r.t.; ii) $(Boc\text{-}\beta\text{-}Ala)_2O$, DMF, DIEA, DMAP, 40 °C; iii) 1N NaOH, MeOH, 60 °C; iv) NO_2-Im-$COCCl_3$, EtOAc, 35 °C; v) H_2, Pd/C, EtOAc, 500 psi, r.t.; vi) $(Boc)_2O$, DIEA, DMF, 60 °C; vii) 1N NaOH, MeOH

and **31**. Saponification was carried out in a mixture of 1N NaOH and methanol at 60 °C for 4-6 hours to obtain the target dimers Boc-Im-Tn-OH (**16**) and Boc-β-Tn-OH (**14**).

Des-Methyl Thiophene (Dt). Treatment of commercially available 5-nitrothiophene-2-carboxaldahyde in acetone with a mixture of sodium hypochlorite and sodium hydrogen phosphate in water, gave the nitro-acid of the des-methyl thiophene ring (NO_2-**Dt**-OH, **32**) (Figure 3.9). Esterification of **32** by refluxing for 48 h in a mixture of sulfuric acid and methanol provided the nitro-ester (NO_2-Dt-OMe, **33**). Reduction of **33** using a mixture of tin(II) chloride and hydrochloric acid in ethanol gave the ammonium chloride salt (HCl•H_2N-Dt-OMe, **34**). Formation of the dimer NO_2-Im-Dt-OMe (**35**), followed by reduction to HCl•H_2N-Im-Dt-OMe (**36**), Boc-protection to Boc-Im-Dt-OMe (**37**), and saponification to the target dimer Boc-Im-Dt-OH (**17**) proceeded as described above for furan compounds **19**, **20**, **21**, and **15**.

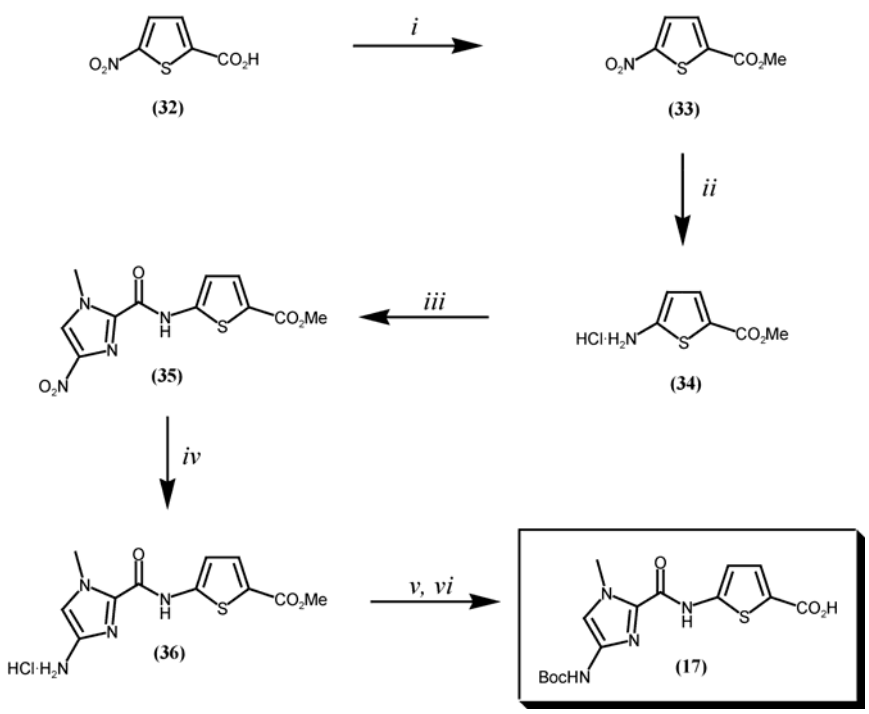


Figure 3.9. Synthesis of thiophene (**Tn**) Boc-protected amino acid monomers and dimers. i) H_2SO_4, MeOH, reflux; ii) $SnCl_2.2H_2O$, 95% EtOH, 37% HCl, 35 °C; iii) NO_2-Im-$COCCl_3$, EtOAc, 35 °C; iv) H_2, Pd/C, EtOAc, 500 psi, r.t.; v) $(Boc)_2O$, DIEA, DMF, 60 °C; vi) 1N NaOH, MeOH, 60 °C.

Methoxythiophene (Mt). The hydroxythiophene methyl ester (Ht-OMe, **38**) was synthesized by a cyclization reaction between methylthioglycolate and methyl-2-chloroacrylate in methanolic sodium methoxide.[22] **38** was nitrated using a mixture of concentrated sulfuric and nitric acid at -10 °C to give 4-nitro-3-hydroxythiophene (NO_2-Ht-OMe, **39**) as the major regioisomer after column chromatography. Treatment of **39** with diazomethane in diethyl ether afforded the methyl ether (NO_2-Mt-OMe, **40**) in near-quantitative yield (Ht = hydroxythiophene, Mt = methoxythiophene). Reduction of **40** using a mixture of tin(II) chloride dihydrate, hydrochloric acid, and ethanol gave the ammonium chloride salt ($HCl•H_2N$-Mt-OMe, **4**). Boc-protection of (**4**) was accomplished by heating a mixture of Boc-anhydride, TEA and methylene chloride at 60 °C for 12 h to provide Boc-Mt-OMe (**41**). Saponification of **41** was achieved using methanolic potassium hydroxide and heating at 50 °C for 6 h to give the final monomer Boc-Mt-OH (**11**).

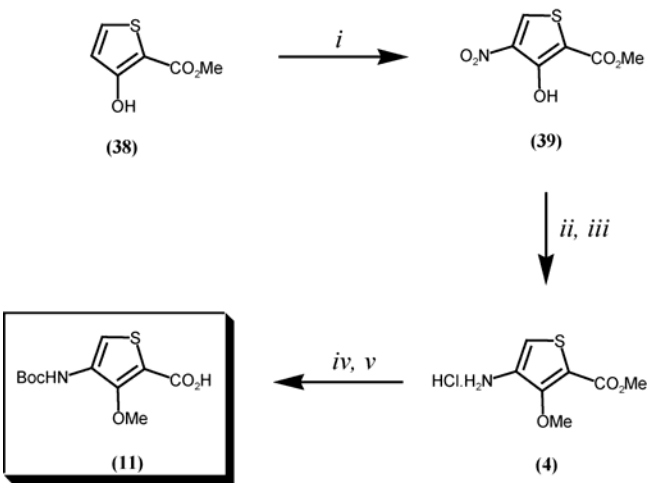
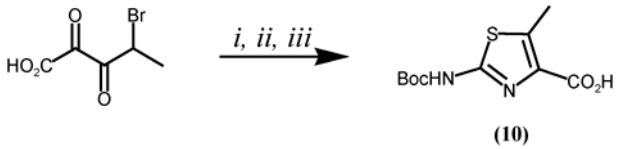


Figure 3.10 Synthesis of methoxythiophene (**Mt**) monomer. i) HNO_3, H_2SO_4, 0 °C; ii) CH_2N_2, r.t.; iii) $SnCl_2.2H_2O$, 95% EtOH, 37% HCl, 40 °C; iv) $(Boc)_2O$, TEA, CH_2Cl_2, 60 °C; v) 1N NaOH, MeOH, 50 °C.

5-methylthiazole (Nt). The Boc-protected **5**-methylthiazole amino acid (Boc-Nt-OH, **10**) was synthesized on a multigram scale by brominating 2-ketobutyric acid, followed by condensation with thiourea and Boc-protection of the amine (Figure 3.11). For best results, the bromine should be added dropwise over at least two hours, as the

reaction is autocatalytic and highly exothermic. Also, the thiourea should be added in small portions with vigorous stirring. Boc-protection was accomplished by dissolving the crude material in DMF and DIEA, followed by the addition of Boc-anhydride and stirring at 60 ^{o}C for twelve hours. The material was then stirred in a solution of MeOH and 1N NaOH for ester saponification to provide the target Boc-Nt-OH (**10**).



(10)

Figure 3.11. Synthesis of N-thiazole (**Nt**) monomer. i) CSN_2H_4, neat, r.t.; ii) $(Boc)_2O$, DIEA, DMF, DMAP, 60 ^{o}C; iii) 1N NaOH, MeOH, 35 ^{o}C.

Polyamides Synthesis. Hairpin and 1:1 motif polyamides were synthesized manually from Boc-β-PAM resin in a stepwise fashion using Boc-protected monomeric and dimeric amino acids (Figure 3.12) according to published solid-phase protocols.[4, 18] Polyamides containing 3-methoxythiophene (Mt) were deprotected by treatment with sodium thiophenoxide in DMF (100 °C, 2 h) to provide the Ht analogues after HPLC purification.

DNA Affinity and Sequence Specificity in the Hairpin Motif. Quantitative DNase I footprinting titrations [2] were carried out for the following polyamides on the 278 bp PCR product of plasmid pDHN1[12]: Im-Im-**Im**-Py-γ-Im-**Py**-Py-Py-β-Dp (Im/Py pair, **42**), Im-Im-**Nh**-Py-γ-Im-**Py**-Py-Py-β-Dp (Nh/Py pair, **43**), Im-Im-**Tn**-Py-γ-Im-**Py**-Py-Py-β-Dp (Th/Py pair, **44**), Im-Im-**Dt**-Py-γ-Im-**Py**-Py-Py-β-Dp (Dt/Py pair, **45**), Im-

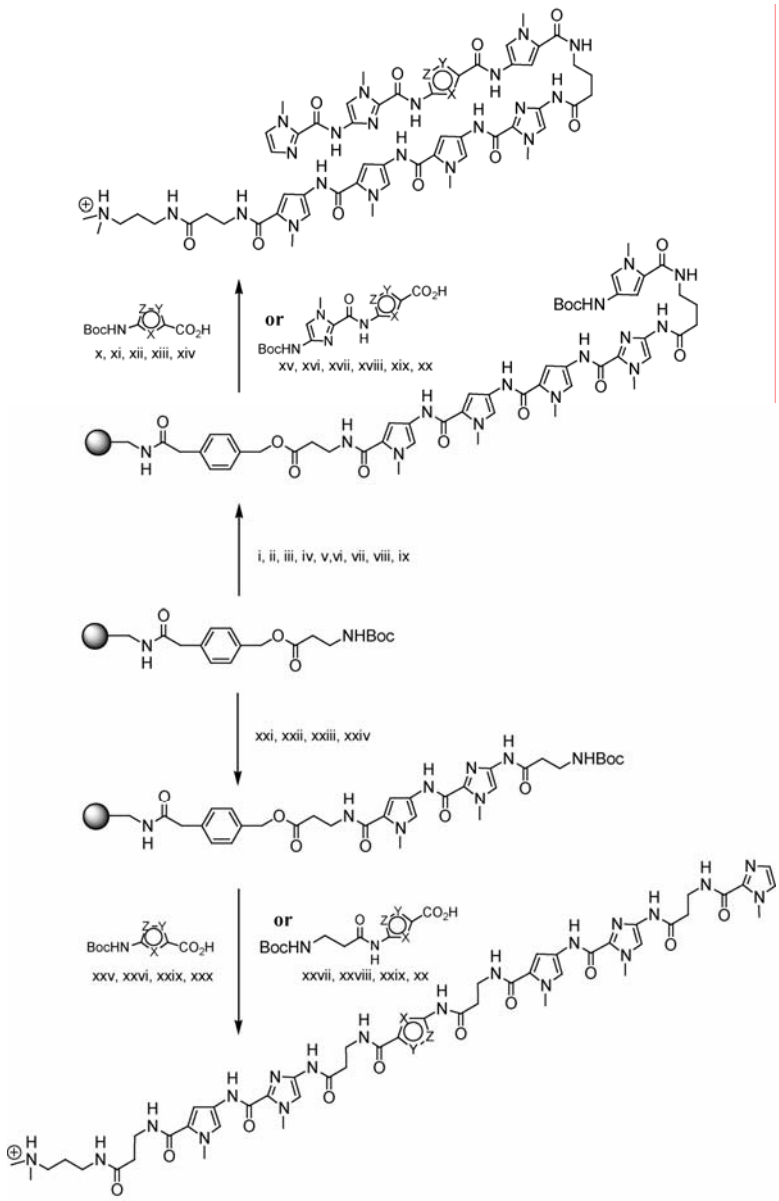


Figure 3.12. Solid-phase synthetic scheme for Im-Im-**X**-Py-γ-Py-Py-Py-β-Dp (arrows up from center) and Im-β-ImPy-β-**X**-β-ImPy-β-Dp (arrows down from center) starting from commercially available Boc-β-Pam resin: (i) 80% TFA/DCM, 0.4M PhSH; (ii) Boc-Py-OBt, DIEA, DMF; (iii) repeat steps (i) and (ii) x 2. (iv) 80% TFA/DCM, 0.4M PhSH; (v) Boc-Im-OH (HBTU, DIEA, DMF); (vi) 80% TFA/DCM, 0.4M PhSH; (vii) Boc-γ-OH (HBTU, DIEA, DMF); (viii) 80% TFA/DCM, 0.4M PhSH; (ix) Boc-Py-OBt, DIEA, DMF; (x) 80% TFA/DCM, 0.4M PhSH; (xi) Boc-**X**-OH (HBTU, DIEA, DMF); (xii) 80% TFA/DCM, 0.4M PhSH; (xiii) Boc-Im-OH (HBTU, DIEA, DMF); (xiv) Im-$COCCl_3$ (DIEA, DMF); (xv) 80% TFA/DCM, 0.4M PhSH; (xvi) Boc-**X**-OH (HBTU, DIEA, DMF); (xvii) 80% TFA/DCM, 0.4M PhSH; (xviii) Im-$COCCl_3$ (DIEA, DMF); (xix) Im-$COCCl_3$ (DIEA, DMF); (xx) (N,N-dimethylamino)propylamine, 85 ^{o}C; (xxi) 80% TFA/DCM, 0.4 M PhSH; (xxii) Boc-Py-OBt, DIEA, DMF; (xxiii) 80% TFA/DCM, 0.4 M PhSH; (xxiv) Boc-β-Im-OH [19] (HBTU, DIEA, DMF); (xxv) 80% TFA/DCM, 0.4 M PhSH; (xxvi) Boc-**X**-OH (HBTU, DIEA, DMF); (xxvii) 80% TFA/DCM, 0.4 M PhSH; (xxviii) Boc-**X**-OH (HBTU, DIEA, DMF); (xxix) elongation and termination according to standard procedures[4]; (xxx) cleave from resin using (N,N-dimethylamino)propylamine, 85 ^{o}C.

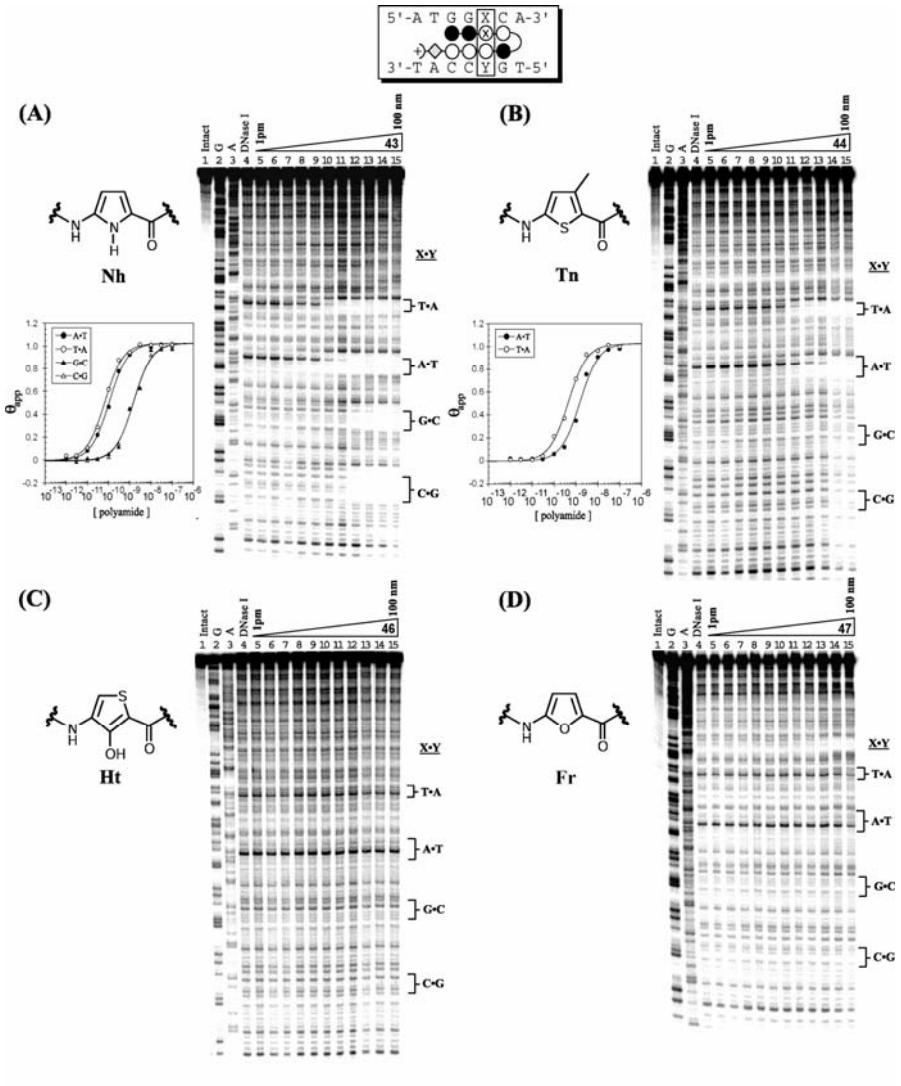


Figure 3.13 (A-D) Quantitative DNase I footprinting experiments in the hairpin motif for polyamides **43**, **44**, **46,** and **47**, respectively, on the 278 bp, 5'-end-labelled PCR product of plasmid DHN1: lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5-15, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM polyamide, respectively. Each footprinting gel is accompanied by the following: (left, top) Chemical structure of the residue of interest; and for A and B (bottom left) Binding isotherms for the four designed sites. θ_{norm} values were obtained according to published methods.[2] A binding model for the hairpin motif is shown centered at the top as a dot model with the polyamide bound to its target DNA sequence. Imidazoles and pyrroles are shown as filled and non-filled circles, respectively; Beta alanine is shown as a diamond; the gamma-aminobutyric acid turn residue is shown as a semicircle connecting the two subunits; and the novel

Im-**Ht**-Py-γ-Im-**Py**-Py-Py-β-Dp (Ht/Py pair, **46**) and Im-Im-**Fr**-Py-γ-Im-**Py**-Py-Py-β-Dp (Fr/Py pair, **47**). The DNA sequence specificity of each polyamide at a single ring pairing position (bolded in the sequences listed above) was determined by varying a

single DNA base pair within the parent sequence context, 5'-T G G **X** C A-3', to all four Watson-Crick base pairs (**X** = A, T, G, C) and comparing the relative affinities of the resulting complexes (Figure 3.13). The variable base pair position was installed opposite the novel heterocycle/pyrrole pairing in question, according to previously reported specificity studies on 8-ring hairpin polyamide.[12]

Equilibrium association constants (K_a) for 8-ring polyamides containing Hp/Py, Py/Py, Pz/Py, and Th/Py, pairings against the four DNA sites used in this study have been reported [12, 18] and are included in Table 3.1 for comparison with values presented here. As expected, polyamide **42** (Im/Py pair) exhibited single site specificity (5'-TGG**X**CA-3', **X** = **G**) at modest affinity (K_a = 4.5 x 10^8 M^{-1}) with 14-fold preference over **X** = **C** and at least 400-fold preference over **X** = A, T. Polyamide **43** (Nh//Py pair) bound with high affinity to the **X** = A, T sites (K_a ~ 10^{10} M^{-1}) in preference to **X** = G, C by about 10-fold. Compound **44** (Tn/Py pair) bound with similar affinity to the **X** = A, T sites (K_a ~ 10^9 M^{-}

Table 3.1. Hairpin Motif, K_a (M^{-1}) [a,b]

Pair	A•T	T•A	G•C	C•G
Im/Py (42)[c]	≤ 10^6	≤ 10^6	4.5 (±0.7) x 10^8	3.2 (±0.5) x 10^7
Hp/Py	8.1 (±1.9) x 10^7	1.6 (±0.3) x 10^9	5.5 (±1.5) x 10^7	7.9 (±2.1) x 10^7
Py/Py	4.7 (±0.7) x 10^9	3.1 (±0.4) x 10^9	2.2 (±0.6) x 10^8	2.5 (±0.9) x 10^8
Nh/Py (43)	8.5 (±0.3) x 10^9	1.1 (±0.1) x 10^{10}	9.2 (±0.1) x 10^8	8.2 (±0.4) x 10^8
Pz/Py	1.0 (±0.5) x 10^9	2.0 (±0.3) x 10^9	≤ 2 x 10^7	≤ 2 x 10^7
Th/Py	≤ 2 x 10^7	≤ 2 x 10^7	≤ 2 x 10^7	≤ 2 x 10^7
Tn/Py (44)	8.0 (±0.4) x 10^8	2.7 (±0.2) x 10^9	≤ 10^6	≤ 10^6
Tp/Py (45)	3.8 (±0.5) x 10^8	1.0 (±0.3) x 10^9	≤ 10^6	≤ 10^6
Ht/Py (46)	≤ 10^6	≤ 10^6	≤ 10^6	≤ 10^6
Fr/Py (47)	≤ 10^6	≤ 10^6	≤ 10^6	≤ 10^6

[a] Values reported are the mean values from at least three DNase I footprint titration experiments, with the standard deviation given in parentheses. [b] Assays were performed at 22 ^oC in a buffer of 10 mM Tris¥HCl, 10 mM KCl, 10 mM, $MgCl_2$, and 5 mM $CaCl_2$ at pH 7.0. [c] The number in parentheses indicates the compound containing the unique pairing.

[1]) with at least 800- to 2700-fold preference over the **X** = G, C sites. As a control, the des-methylthiophene analogue, Dt, was prepared and tested within this system in order to probe possible effects on DNA binding caused by the 3-methyl group. The polyamide containing a Dt/Py pair (**45**) displayed virtually identical recognition properties as its parent, **44**. The compounds containing Ht/Py and Fr/Py pairs (**46** and **47**, respectively) showed no binding to the designed sites at concentrations up to 1 μM.

DNA Affinity and Sequence Specificity in the 1:1 Motif. – Quantitative DNase I footprinting titrations were carried out for the following polyamides on the 298 bp PCR product of pAU8[19]: Im-β-Im-Py-β-**Py**-β-Im-Py-β-Dp (**8**), Im-β-Im-Py-β-**Nh**-β-Im-Py-β-Dp (**9**), Im-β-Im-Py-β-**Hp**-β-Im-Py-β-Dp (**10**), Im-β-Im-Py-β-**Ht**-β-Im-Py-β-Dp (**11**), Im-β-Im-Py-β-Fr-β-Im-Py-β-Dp (**12**), Im-β-Im-Py-β-**Nt**-β-Im-Py-β-Dp (**13**), Im-β-Im-Py-β-**Tn**-β-Im-Py-β-Dp (**14**), and Im-β-Im-Py-β-**Th**-β-Im-Py-β-Dp (**15**). The sequence specificity of each polyamide at a single carboxamide position (bolded in the sequences listed above) was determined by varying a single base pair within the parent DNA sequence context, 5'-AAAGA**X**AAGAG-3', to all four Watson-Crick base pairs (**X** = A, T, G, C) and comparing the relative affinities of the resulting complexes (Figures 3.14 and 3.15). The variable base pair position was installed opposite the novel heterocycle in question, according to previously described specificity studies on 1:1 polyamide:DNA complexes.[19]

Equilibrium association constants (K_a) for 1:1 polyamides containing Im, Py, and Hp residues tested against the four Watson-Crick base pairs have been reported. However, in that study only the Im specificity experiment was performed at the central

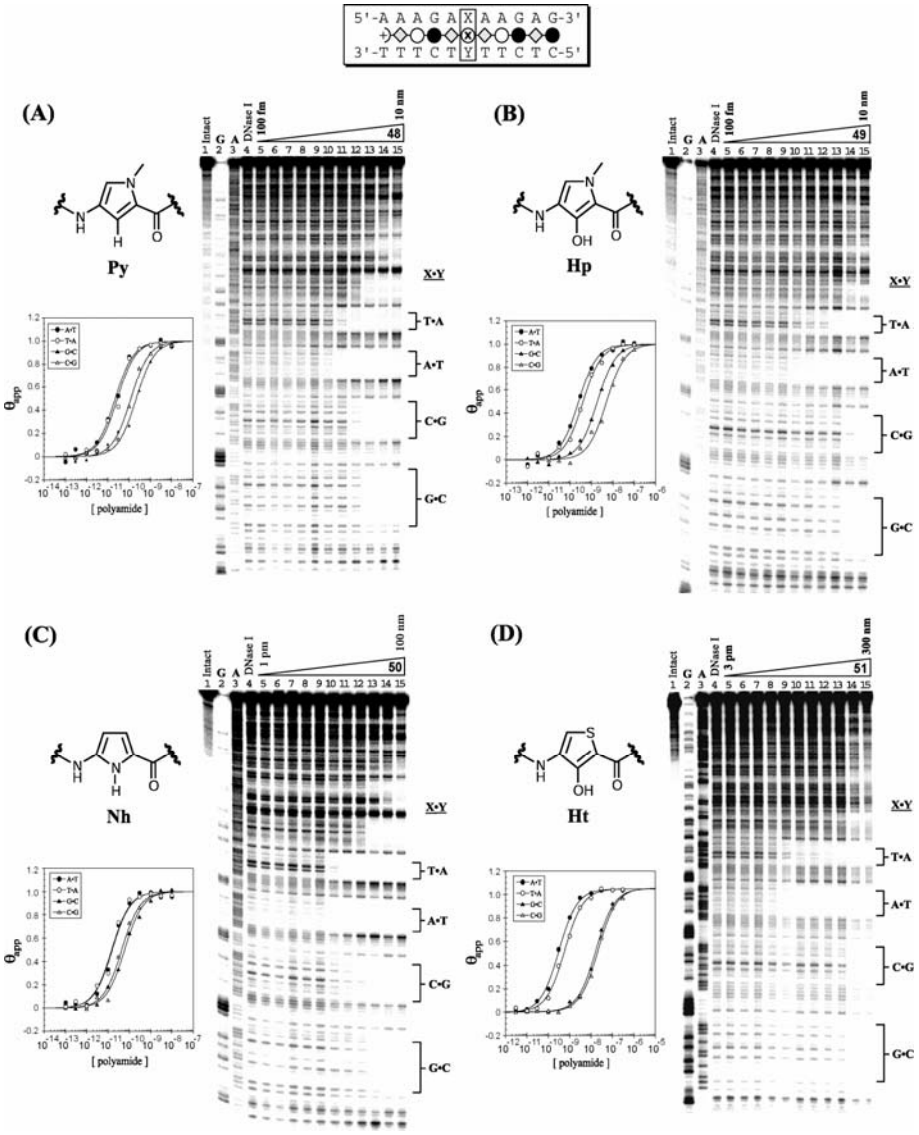


Figure 3.14 (A-D) Quantitative DNase I footprint titration experiments for polyamides **48-51**, respectively, on the 298 bp, 5'-end-labelled PCR product of plasmid pAU8: (A and B) lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5-15, 100 fM, 300 fM, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM polyamide, respectively. (C) lane 1, intact DNA; lane 2, G reaction; lane 3 A reaction; lane 4, DNase I standard; lanes 5-15, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM polyamide respectively. (D) lane 1, intact DNA; lane 2, G reaction; lane 3 A reaction; lane 4, DNase I standard; lanes 5-15, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM, 30 nM, 100 nM, 300 nM polyamide respectively. Each footprinting gel is accompanied by the following: (left, top) chemical structure of the residue of interest; and (left bottom) binding isotherm for the four designed sites. θ_{norm} values were obtained according to published methods.[2] A binding model for the 1:1 motif is shown centered at the top as a dot model with the polyamide bound to its target DNA sequence. Imidazoles and pyrroles are shown as filled and non-filled circles, respectively; Beta alanine is shown as a diamond; and the novel heterocycles are indicated by a circle containing an X.

residue, as with the new polyamides reported here. Therefore, new polyamides

containing Py and Hp residues at the central position have been included in the study for a more controlled comparison. Polyamide **48** (Py) demonstrated very high affinity (K_a ~ 6 x 10^{10} M^{-1}) at the **X** = A, T sites (5'-AAAGA**X**AAGAG-3') with a 5- to 10-fold preference over **X** = G, C (Table 3.2). Polyamide **49** (Hp) bound with lower affinity (K_a ~ 3 x 10^9 M^{-1}) but with similar specificity to **48**, preferring **X** = A, T > G, C by 5- to 10-fold. The Nh-containing polyamide (**50**) bound with very high affinity to the **X** = A, T sites (K_a = 7.5 x 10^{10} M^{-1}) but with a mere 3- to 5-fold selectivity over the high-affinity **X**

Table 3.2. 1:1 Motif, K_a (M^{-1})[a,b]

Ring	A•T	T•A	G•C	C•G
Im	2.5 (±0.2) x 10^{10}	1.1 (±0.1) x 10^{10}	2.6 (±0.4) x 10^{10}	1.3 (±0.3) x 10^{10}
Py (48)[c]	7.2 (±0.3) x 10^{10}	5.3 (±0.1) x 10^{10}	3.2 (±0.4) x 10^9	9.4 (±0.2) x 10^9
Hp (49)	3.9 (±0.1) x 10^9	2.5 (±0.3) x 10^9	5.3 (±0.5) x 10^8	1.9 (±0.5) x 10^8
Nh (50)	7.5 (±0.2) x 10^{10}	7.4 (±0.1) x 10^{10}	1.6 (±0.2) x 10^{10}	2.3 (±0.1) x 10^{10}
Ht (51)	2.8 (±0.5) x 10^9	1.6 (±0.6) x 10^9	3.8 (±1.3) x 10^7	3.7 (±0.7) x 10^7
Fr (52)	2.2 (±0.5) x 10^{10}	1.0 (±1.3) x 10^{10}	4.4 (±0.5) x 10^9	5.0 (±0.5) x 10^9
Nt (53)	5.4 (±0.9) x 10^9	2.9 (±0.6) x 10^9	8.0 (±1.3) x 10^9	4.2 (±0.6) x 10^9
Tn (54)	3.0 (±0.2) x 10^{10}	5.7 (±0.4) x 10^9	8.1 (±0.4) x 10^7	8.3 (±0.2) x 10^7
Th (55)	1.5 (±0.2) x 10^{10}	3.0 (±0.7) x 10^9	6.5 (±0.5) x 10^6	7.4 (±0.5) x 10^6

[a] Values reported are the mean values from at least three DNase I footprint titration experiments, with the standard deviation given in parentheses. [b] Assays were performed at 22 ^oC in a buffer of 10 mM Tris¥HCl, 10 mM KCl, 10 mM, $MgCl_2$, and 5 mM $CaCl_2$ at pH 7.0. [c] The number in parentheses indicates the compound containing the unique pairing.

= G, C sites. Compound **51** (Ht) bound with subnanomolar affinities to the **X** = A, T sites, similar to **49** but with ≥ 40-fold specificity for **X** = A, T > G, C. Polyamide **52** (Fr) showed high affinity for the **X** = A, T sites (K_a ~ 10^{10} M^{-1}) with a small 2- to 4-fold preference over **X** = **G**, **C**. The 5-methylthiazole-containing polyamide (**53**, Nt), which places the thiazole ring *nitrogen* into the floor of the minor groove, bound all four sites

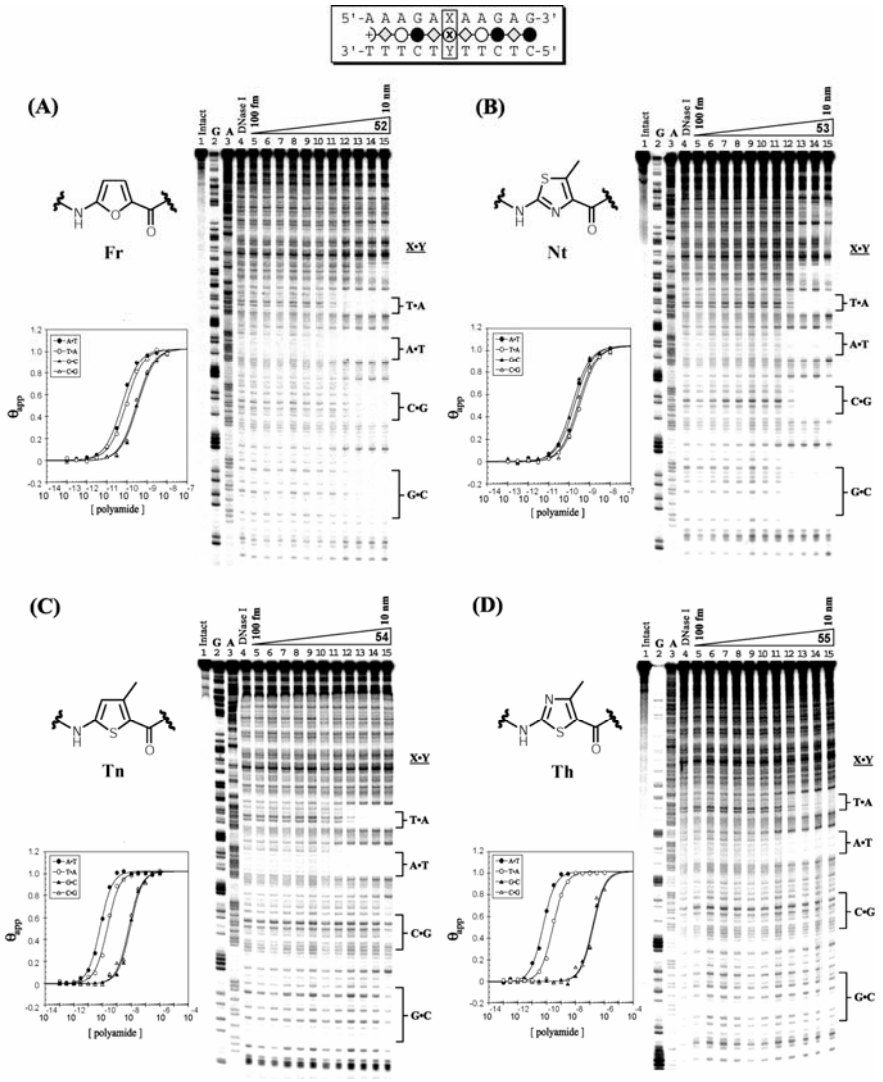
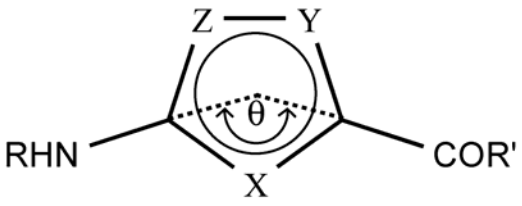


Figure 3.15. (A-D) Quantitative DNase I footprinting experiments for polyamides **52-55**, respectively, on the 298 bp, 5'-end-labelled PCR product of plasmid pAU8: lane 1, intact DNA; lane 2, G reaction; lane 3, A reaction; lane 4, DNase I standard; lanes 5-15, 100 fM, 300 fM, 1 pM, 3 pM, 10 pM, 30 pM, 100 pM, 300 pM, 1 nM, 3 nM, 10 nM respectively. Each footprinting gel is accompanied by the following: (left, top) Chemical structure of the residue of interest; and (left bottom) Binding isotherms for the four designed sites. Isotherms for C and D were generated from gels run out to a final concentration of 1 uM (not shown). θ_{norm} values were obtained according to published methods.[2] A binding model for the 1:1 motif is shown centered at the top as a dot model with the polyamide bound to its target DNA sequence. Imidazoles and pyrroles are shown as filled and non-filled circles, respectively; Beta alanine is shown as a diamond; and the novel heterocycles are indicated by a circle containing an X.

with similar high affinities ($K_a \sim 5 \times 10^9$ M^{-1}). Thiophene-containing polyamide **54** (Tn) showed modest single-site specificity, binding the **X** = **A** site at $K_a = 3.0 \times 10^{10}$ M^{-1} with 5-fold preference over **X** = **T** and ~ 70-fold preference over **X** = **G**, **C**. 3-

methylthiophene-containing polyamide (**55**, Th), which places the thiazole ring *sulfur* into the floor of the minor groove, bound with similar **X** = A, T affinity as **54** (Tn) but with > 400-fold preference over **X** = G, C. In all cases, binding isotherms fit well to an n = 1 Hill equation, supporting a 1:1 polyamide:DNA stoichiometry.

Molecular Modeling Calculations. Modeling calculations were preformed using *Spartan Essential* software package [23]. Each ring was first minimized using an AM1 model, followed by *Ab initio* calculations using the Hartree-Fock model and a 6-31G* polarization basis set. Each heterocycle exhibited a unique geometric and electronic profile (Figure 3.16). Bonding geometry for imidazole, pyrrole, and 3-hydroxypyrrole were in excellent agreement with coordinates derived from x-ray structures of polyamides containing these heterocycles.[1, 11] The overall curvature of each monomer was calculated to be the sweep angle (θ) created by the



Ring	X	Y	Z	θ (*degrees*)	Charge on X (*e*)
Fr	O	C-H	C-H	*126*	*-0.31*
Nt	N	C-Me	S	*127*	*-0.60*
Ht	O-H	S	C-H	*133*	*+0.40*
Nh	N-H	C-H	C-H	*136*	*+0.34*
Im	N	N-Me	C-H	*137*	*-0.71*
Py	C-H	N-Me	C-H	*146*	*+0.21*
Hp	O-H	N-Me	C-H	*148*	*+0.50*
Tn	S	C-Me	C-H	*149*	*-0.21*
Pz	C-H	N-Me	N	*151*	*+0.23*
Dt	S	C-H	C-H	*152*	*-0.21*
Th	S	C-Me	N	*153*	*-0.25*

Figure 3.16 Geometric and electrostatic profiles for 11 heterocyclic amino acids, derived from *ab initio* molecular modeling calculations using *Spartan Essential* software. (Top) Schematic illustrating the amide-ring-amide angle of curvature, θ. X, Y, and Z denote variable functionality at the different ring positions for each heterocycle. (Bottom) Table listing the functional groups at X, Y, and Z, along with the angle θ, and the electrostatic partial charge on X. For Ht, Nh, Py, Hp, and Pz the positive charge on X is listed for the H atom.

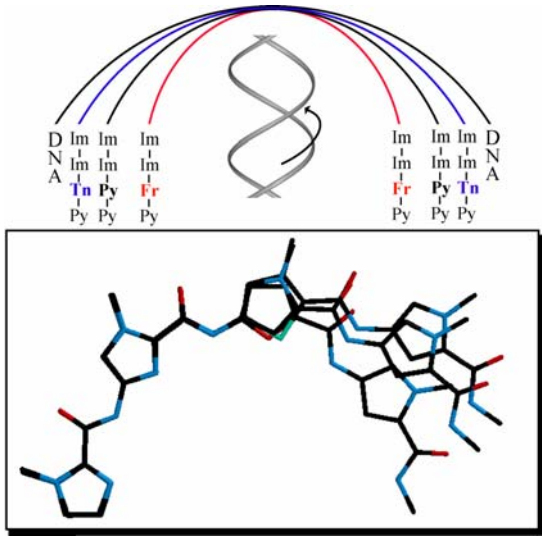


Figure 3.17 (Top) Schematic illustrating the curvatures of 4-ring polyamide subunits containing Tn, Py, and Fr heterocycles with respect to one another and the DNA helix. (Box) *Ab inito* models of polyamide subunits (Im-Im-**X**-Py, **X** = Tn, Py and Fr) superimposed to demonstrate the significant difference in curvature resulting from atomic substitution. Hydrogens are not shown.

theoretical intersection of the two ring-to-amide bonds in each ring. The structures were ranked by increasing θ as follows: Fr > Nt > Ht > Nh > Im > Py > Hp > Tn > Pz > Dt > Th. The ring atom in closest proximity to the floor of the DNA minor groove was examined for partial charge. The structures were ranked by decreasing partial charge on this atom as follows: Hp > Ht > Nh > Pz > Py > Tn = Dt > Th > Fr > Im. Four-ring-subunits containing the sequence Im-Im-**X**-Py (**X** = Py, Pz, Nh, Im, Fr, Hp, Ht, Th, Tn and Dt) were constructed and subjected to AM1 and *ab initio* calculations as described above in order to examine overall subunit curvature and planarity (Figures 3.17 & 3.18).

Discussion. Here we explore the effects of chemical structure, within the framework of a 5-membered, aromatic heterocycle, on the ability of hairpin and 1:1 polyamides to discriminate between the four Watson-Crick base pairs in the minor groove of DNA. A carefully controlled mutational scheme of DNA and polyamide sequence allows for the comparison of binding affinities for a {4} x {10} array of complexes containing unique combinations of {Watson-Crick base pair} x {5-membered

aromatic heterocycle} at a single position (Tables 3.1 and 3.2). This quantitative analysis combined with computational modeling of the different heterocycles has led to concepts we believe offer insight into the etiology of DNA sequence discrimination by polyamides.

The success of imidazole/pyrrole and hydroxypyrrole/pyrrole pairs at discriminating between the four Watson-Crick base pairs has been attributed largely to the functional group directed toward the floor of the minor groove, *e.g.*, N3 for imidazole and C3-OH for hydroxypyrrole (X in Figure 3.3). [1, 6, 11] Therefore, the heterocycles discussed here will be divided into groups based on the type of functional group presented to the DNA minor groove. For the sake of clarity and brevity, hairpin polyamides will be referred to in bold as their unique amino acid pair, *e.g.*, **Tn/Py** for polyamide ImIm**Tn**Py-γ-Im**Py**PyPy-β-Dp (**44**). 1:1 polyamides will be referred to in bold as their unique heterocycle, *e.g.*, **Tn** for polyamide Im-β-ImPy-β-**Tn**-β-ImPy-β-Dp (**54**). DNA sequences will be identified as the variable base position within each motif, *e.g.*, **X** = G in the hairpin motif for 5'-TGG**X**CA-3'.

The Hairpin Motif. Recognition of antiparallel Watson-Crick base pairs using antiparallel amino acid pairs has proven to be a powerful approach. The hairpin motif provides an attractive context in which to develop new pairing rules because the head-to-tail linkage of antiparallel subunits limits the slipping of subunits in relation to each other, thereby "locking" the pairs of amino acids into place.[15, 16] The 4-ring subunits described here (e.g. ImPyPyPy) are continuously π-conjugated, which limits their

conformational flexibility. Consequently, small changes in ring curvature can cause greater effects on overall subunit curvature.

Py/Py, **Pz/Py**, and **Nh/Py** present hydrogen with positive potential to the floor minor groove floor. **Py/Py** and **Pz/Py** have been reported previously.[12] **Py/Py** exhibits degeneracy for the **X** = A, T binding sites, as predicted, with > 10-fold selectivity for **X** = A, T > G, C. **Pz/Py** is also A•T/T•A degenerate, binding its match **X** = A, T sites with similar affinity as **Py/Py**, but with no specific recognition of **X** = G, C. **Nh/Py** is similarly A•T/T•A degenerate, but it binds all sites with higher affinity than **Py/Py** and **Pz/Py**.

The first crystal structure of netropsin bound to DNA inspired the prediction for the origin of A•T, T•A > G•C, C•G (**X** = A, T > G, C) specificity residing in the steric clash between the pyrrole C3-H and the exocyclic amino group of guanine (G-NH_2).[7] This specificity has proven true for 2:1 homodimeric polyamide complexes with distamycin[8] and its simple analogues[10, 11] as well as with hairpin [16] and cyclic polyamides.[24] The A•T/T•A degeneracy of **Py/Py**, **Pz/Py**, and **Nh/Py** is believed to originate from a steric interaction with G-NH2, as described above for pyrrole, yet the subtle differences among these compounds are thought to derive from their unique chemical structures. Py is known to be over-curved with respect to the DNA helix.[1, 11, 25-27] The calculations described here provide an amide-ring-amide intersection angle (θ in Figure 3.16) of 146° and a partial positive charge on the C3-H atom of +0.21. By contrast, pyrazole is somewhat less curved than pyrrole (θ = 151°) with similar charge, and Nh-pyrrole is considerably more curved (θ = 136°) with a greater charge of +0.34. The reduced curvature of **Pz/Py** should make it more complementary to the DNA

curvature, and therefore the steric C4-H to G-NH2 clash would be exacerbated, resulting in lower affinity for the **X** = G, C sites. **Nh/Py** is more curved and more charged, thereby reducing negative steric effects and increasing binding affinity.

Im/Py and **Fr/Py** present a nitrogen or oxygen with an sp^2 lone pair directed toward the minor groove floor. In spite of this similarity, their DNA recognition behavior is strikingly different. **Im/Py** behaves as expected, preferentially targeting **X** = G. On the other hand, **Fr/Py** shows a complete loss of DNA-binding affinity. Based on the established principles for G•C recognition by an imidazole/pyrrole pair, we had predicted the furan/pyrrole pair to be a positive recognition element for G•C, with the furan oxygen acting as a hydrogen bond acceptor to G-NH2. However, calculations for furan reveal its tight over-curvature, with the amide-ring-amide angle decreased by more than ten degrees with respect to imidazole and more than twenty degrees with respect to pyrrole. This property causes a pronounced effect on the entire ImIm**Fr**Py subunit, shown in Figure 3.17, such that complex formation is no longer energetically favorable at low concentrations. This view is further reinforced by results in the more flexible 1:1 motif, wherein the furan containing Fr binds all designed sites with very high affinity.

Hp/Py and **Ht/Py** present a hydroxyl group (C3-OH) to the minor groove floor. **Hp/Py**, reported previously [18, 28] displays a preference for **X** = T that breaks the A•T/T•A degeneracy of pyrrole/pyrrole pairs. Structural studies predict the origin of this specificity to reside in the hydrogen bond formed between the 3C-OH of hydroxypyrrole and the O2 of thymine (T-O2), in addition to shape recognition of the asymmetric cleft in the A•T base pair.[1, 27] Unfortunately, a loss in affinity is typically observed for hydroxypyrrole containing polyamides compared to their pyrrole analogues. This loss in

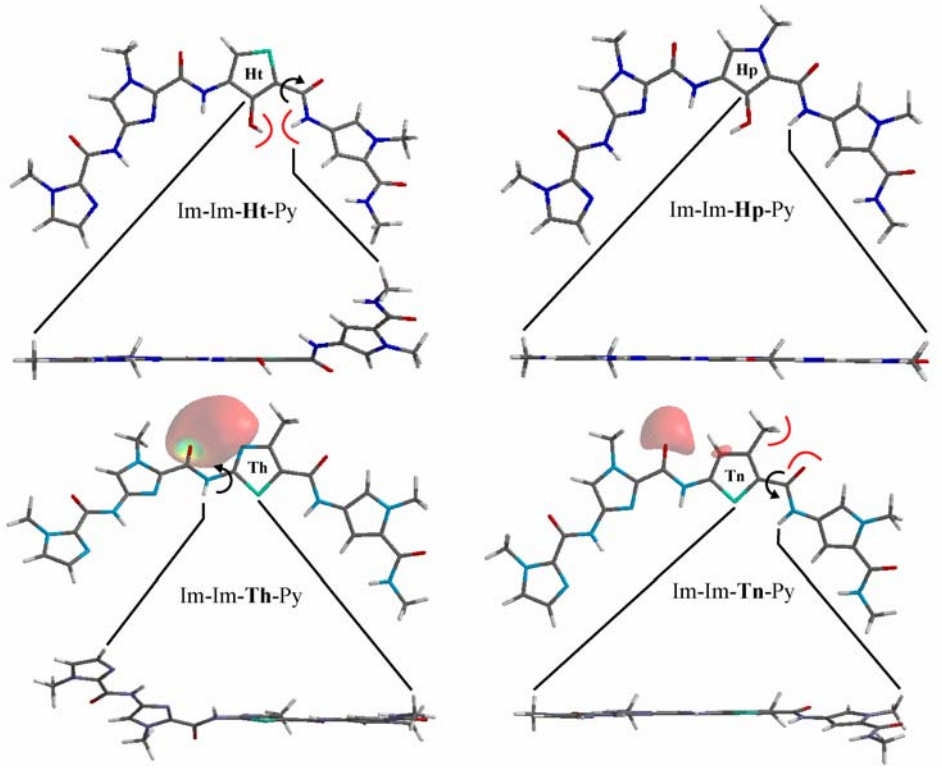


Figure 3.18 *Ab initio* models of 4-ring polyamide subunits (Im-Im-**X**-Py, **X** = Ht, Hp, Th, and Tn): (Top Left) The Im-Im-**Ht**-Py subunit demonstrating a dihedral (shown as a curved arrow) created at the ring-carboxamide juncture due to destabilizing eclipsing interaction (shown as red arcs) between the C3-OH hydroxyl group and proximal carboxamide proton; (Top Right) The Im-Im-**Hp**-Py subunit demonstrating co-planarity of the contiguous ring system; (Bottom Left) The Im-Im-**Th**-Py subunit showing the negative isopotential surface (lone pair density) in red, and the dihedral (shown as a curved arrow) resulting from lone pair repulsion interaction between the thiazole nitrogen and the proximal carbonyl; and (Bottom Right) The Im-Im-**Tn**-Py subunit showing the negative isopotential surface (lone pair density) in red, and the small dihedral (shown as a curved arrow) resulting from a destabilizing eclipsing interaction (shown as red arcs) between the thiophene methyl group and the proximal carboxamide. Atomic substitution of N to C-H for **Th** to **Tn** removes lone pair repulsion interaction between the ring nitrogen and the proximal carbonyl.

affinity may be attributed to negative steric interactions between the C3-OH and the minor groove floor. **Ht/Py** was prepared and tested in an attempt to improve upon **Hp/Py** by increasing the ring curvature, thus reducing the negative steric interaction while maintaining the C3-OH to T-O2 hydrogen bond. DNase I footprinting for **Ht/Py** reveals

a complete loss in binding affinity. As seen with **Fr/Py**, increased ligand curvature may completely disrupt hairpin binding. In addition, *ab initio* calculations on the entire ImIm**Ht**Py subunit expose an unfavorable eclipsing interaction between the 3C-OH group and the proximal carboxamide proton. As shown in Figure 3.18, this steric clash may force rotation about the ring-amide bond, which would twist the subunit out of plane. Given the snug fit of stacked hairpin subunits within the DNA minor groove, a large distortion in ligand planarity may not be tolerated. **Ht** in the flexible 1:1 motif binds its target sites with high affinity, which further underscores the pronounced effects of ring geometry within the conformationally constrained hairpin motif.

Th/Py, **Tn/Py**, and **Dt/Py** present a sulfur atom with an sp^2 lone pair to the minor groove floor. **Th/Py** has been reported previously,[12] demonstrating no specific binding for the designed sites. It was thought that the thiazole sulfur was too large to be well accommodated within the closely packed hairpin•DNA complex. However, *ab initio* calculations on thiazole carboxamide reveal an unfavorable interaction between the lone pair of the thiazole N3 and the proximal carbonyl oxygen (Figure 3.18). Although this interaction exists for pyrazole N2, the effect on thiazole is much greater due to the large sulfur atom forcing the thiazole nitrogen into closer proximity with the carboxamide. Consequently, the polyamide subunit may twist out of plane about the amide-ring bond in order to alleviate electronic strain, as shown in Figure 3.18. As with **Ht/Py**, the diminished DNA binding affinity of **Th/Py** may be due to its non-planar conformation. These effects are not observed for **Th** in the 1:1 motif.

The negative electronic interaction and hence non-planar conformation should be alleviated if the nitrogen on the back corner of thiazole were replaced with C-H, as in the

case of thiophene. Hence, **Tn/Dt** was prepared and tested. Remarkably, **Tn/Dt** binds to the **X** = A, T sites with high affinity, but with no observable binding to **X** = G, C. Therefore, the sulfur is *not* too large to be accommodated within the tightly packed hairpin•DNA complex. Furthermore, the reduced curvature of thiophene, which should exacerbate steric effects between the sulfur and the minor groove floor, is not a problem for A•T recognition but is likely to be the cause for the > 800-fold loss in binding affinity at the **X** = G, C sites. An *Ab inito* calculation of the ImIm**Tn**Py subunit reveals a minor negative steric interaction between the thiophene methyl group and the proximal carboxamide oxygen, which may force the subunit to twist slightly out of plane. Although structural data on polyamide•DNA complexes reveals a tolerance to small amide-ring dihedrals, the des-methyl polyamide **Dt/Py** was prepared and tested as a control. **Dt/Py** exhibited virtually identical DNA recognition behavior as **Tn/Py**, which further supports this idea of tolerance.

The 1:1 Motif. The 1:1 motif has emerged recently as a way to target long, purine-rich DNA sequences with high affinity.[19, 29] Although the pairing of polyamide subunits in the hairpin motif offers a greater chance to differentiate between Watson-Crick base pairs, the single-subunit•DNA complexes of the 1:1 motif provide a relatively flexible system for the exploration of novel recognition elements. Due to the conformational freedom imparted by the beta-alanine residues, changes in heterocycle geometry do not have such a pronounced impact on the rest of the molecule. Therefore, specificity may be more difficult to achieve in this motif. In fact, all 1:1 compounds described here bind with high affinity to the **X** = A, T sites but with varying degrees of **X** = A, T > G, C specificity. Structural studies reveal an important register of amide NH

groups with the purine N3 and pyrmidine O2 groups on the floor of the DNA minor groove.[3] Given this alignment as a driving force for DNA recognition in the 1:1 motif, one may view the subtle differences in heterocycle curvature as merely placing the central ring atom (X in Figure 3.3) closer to or farther from the DNA. In this view, increasing the ring curvature decreases the polyamide•DNA intimacy, thereby diminishing DNA specificity. The results presented here fit well within this ideology.

Py and **Nh** present a moderately acidic hydrogen with a positive potential to the minor groove floor. Both compounds exhibit a modest 3- to 5-fold selectivity for **X** = A, T > G, C, but **Nh** binds with higher affinity to all sites. The selectivity is probably due to the negative steric X-H to G-NH2 interaction (X = C3 for **Py** and N1 for **Nh**), which was predicted by Dickerson and coworkers for netropsin and supported by recent NMR studies.[3, 7] The higher affinity for **Nh** may be attributed to a combination of greater positive charge on N1-H and higher ring curvature, both of which should reduce specificity.

Im, **Fr,** and **Nt** present a small atom with an sp^2 lone pair directed toward the minor groove floor. **Im** has been reported previously,[19] binding all sites with high affinity and displaying virtually no discrimination between sites. **Fr** and **Nt** behave quite similarly. It is likely that the small atom (N for **Im** and **Nt** or O for **Fr**) presented to the DNA provides no steric clash with G-NH2, and therefore all sites are bound with similarly high affinity.

Hp and **Ht** present a hydroxyl group to the DNA minor groove. Previously in both hairpin and 1:1 systems, hydroxypyrrole successfully discriminated between A•T and T•A base pairs.[18, 19, 30] Yet when flanked on both sides by β-alanine residues, as with

the **Hp** compound presented here, single base-pair specificity is lost. This loss may be attributed to a larger degree of conformational freedom afforded to the **Hp** ring by the aliphatic linkers.[3] Nonetheless, both **Hp** and **Ht** exhibit significant **X** = A, T > G, C specificity, as expected from a negative 3C-OH to G-NH2 steric clash. **Ht** is slightly more specific, which may result from the non-planarity of this ligand as discussed above for **Ht/Py** in the hairpin motif.

Tn and **Th** present a sulfur atom with an sp^2 lone pair to the DNA minor groove. These compounds exhibit substantial **X** = A, T > G, C specificity ranging from ≥ 70 to ≥ 2300-fold. This remarkable selectivity may be attributed to the decreased curvature of thiazole and thiophene rings, which forces a more intimate interaction of the large sulfur atom and the minor groove floor. In the case of **X** = G, C, this interaction is very negative, resulting in a dramatic loss in binding affinity. **Th** is more specific than **Tn**, which is probably due to the curvature-induced non-planarity of ImIm**Tn**Py, as discussed above for the hairpin motif (Figure 3.18).

Conclusions. Our understanding of the origin of DNA sequence discrimination by polyamides has been improved by combining the tools of quantitative DNase I footprinting and computational molecular modeling to establish a correlation between polyamide structure and DNA sequence specificity (Tables 3.3 and 3.4). We believe that the footprinting results are best explained by differences in the overall heterocycle structure. Each heterocyclic amino acid has an inherently unique bonding geometry, which results in varying degrees of curvature complementarity between the polyamide and the DNA minor groove. Given that pyrrole is over-curved with respect to the DNA

Table 3.3a. Pairing Specificity for Lone Pair to Floor

Pair	A•T	T•A	G•C	C•G
Im/Py	–	–	+	–
Fr/Py	–	–	–	–

Table 3.3b. Pairing Specificity for C-H or N-H to Floor

Pair	A•T	T•A	G•C	C•G
Py/Py	+	+	–	–
Nh/Py	+	+	–	–
Pz/Py	+	+	–	–

Table 3.3c. Pairing Specificity for C-OH to Floor

Pair	A•T	T•A	G•C	C•G
Hp/Py	–	+	–	–
Ht/Py	–	–	–	–

Table 3.3d. Pairing Specificity for Sulfur to Floor

Pair	A•T	T•A	G•C	C•G
Th/Py	–	–	–	–
Tn/Py	+	+	–	–
Tp/Py	+	+	–	–

helix, reducing heterocycle curvature should increase the polyamide-DNA fit. Consequently, the polyamide would have greater sensitivity to changes in DNA structure and therefore greater DNA sequence selectivity. On the other hand, increasing heterocycle curvature should decrease sensitivity to changes in DNA sequence. In addition, over-curvature can induce ligand non-planarity deriving from destabilizing eclipsing interactions. These results suggest that merely considering the functional group facing the minor groove floor is insufficient for an accurate prediction of DNA recognition behavior.

Curvature effects are amplified in contiguous ring polyamides, where continuous π-conjugation limits conformational flexibility. Furthermore, the packing of stacked polyamide subunits in the minor groove, as with the hairpin motif, provides additional limitation of conformation. However, the dense functional array offered to the DNA by

hairpin polyamides gives this class of compounds great promise for sequence discrimination. By contrast, the 1:1 motif offers a high degree of flexibility but a less dense functional array and therefore a lower capacity for DNA sequence selectivity.

Table 3.4. Specificity in the 1:1 Motif

Pair	A•T	T•A	G•C	C•G
Im	+	+	+	+
Py	+	+	-	-
Hp	+	+	-	-
Nh	+	+	+	+
Ht	+	+	-	-
Fr	+	+	+	+
Nt	+	+	+	+
Tn	+	+	-	-
Th	+	+	-	-

The comparison of new heterocycles in both motifs has brought insight to the nature of residues that eliminate hairpin binding, such as furan, thiazole, hydroxythiophene. It will be interesting to examine these new heterocycles in a more relaxed, beta alanine-linked hairpin system.

Experimental.

General. N,N-dimethylformamide (DMF), N,N-diisopropylethylamine (DIEA), thiophenol (PhSH), N,N-diethylamine, N,N-dimethylaminopropylamine (Dp), triethylamine (TEA), methyl 2-furoate, ketobutyric acid, methyl acetoacetate, cyanoacetic acid, trichloroacetyl chloride, pyrrole, sodium metal, methylthioglycolate, methyl-2-chloroacrylate, tin(II) chloride dihydrate, and thiourea were purchased from Aldrich. Boc-β-alanine-(4-carbonylaminomethyl)-benzyl-ester-copoly(styrene-divinylbenzene)resin (Boc-β-Pam-resin), dicyclohexylcarbodiimide (DCC),

hydroxybenzotriazole (HOBt), 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), N,N-dimethylaminopyridine (DMAP), and Boc-β-alanine were purchased from NOVA Biochem. Trifluoroacetic acid (TFA) was purchased from Halocarbon. All other solvents were reagent-grade from EM. Oligonucleotide inserts were synthesized by the Biopolymer Synthesis Center at the California Institute of Technology. Glycogen (20 mg/mL), dNTPs (PCR nucleotide mix), and all enzymes, unless otherwise stated, were purchased from Boehringer-Mannheim. pUC19 was purchased from New England Biolabs, and deoxyadenosine [γ-^{32}P]triphosphate was provided by ICN. Calf thymus DNA (sonicated, deproteinized) and DNaseI (7500 units/mL, FPLC pure) were from Amersham Pharmacia. AmpliTaq DNA polymerase was from Perkin Elmer and used with the provided buffers. Tris.HCl, DTT, RNase-free water, and 0.5 M EDTA were from United States Biochemical. Calcium chloride, potassium chloride, and magnesium chloride were purchased from Fluka. Tris-borate-EDTA was from GIBCO and bromophenol blue was from Acros. All reagents were used without further purification.

NMR spectra were recorded on a Varian spectrometer at 300 MHz in DMSO-*d6* or $CDCl_3$ with chemical shifts reported in parts per million relative to residual solvent. UV spectra were measured on a Hewlett-Packard Model 8452A diode array spectrophotometer. High resolution FAB and EI mass spectra were recorded at the Mass Spectroscopy Laboratory at the University of California, Los Angeles. Matrix-assisted, laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) was conducted at the Protein and Peptide Microanalytical Facility at the California Institute of Technology.

Monomer and Dimer Synthesis. *5-Amino-furan-2-carboxylic acid methyl ester* (NH_2-Fr-OMe, **1**)**.** Methyl 5-nitro-2-furoate (**18**) was prepared by published methods [20] in 84% yield, TLC (5:2 hexanes/ethyl acetate) R_f 0.7; ^{1}H NMR (DMSO-d_6) δ 7.78 (d, J = 3.9 Hz, 1H), 7.59 (d, J = 3.9 Hz, 1H), 3.38 (s, 3H); ^{13}C NMR (DMSO-d_6) δ 157.7, 152.8, 144.5, 120.4, 113.6, 53.5; EI-MS m/e 171.117 (M^+ calcd for 171.117 $C_6H_5NO_5$). A mixture of Methyl 5-nitro-2-furoate (**18**) (3 g, 17.5 mmol) and 10% palladium on carbon (0.3 g) in ethyl acetate (25 mL) was placed into a Parr apparatus and hydrogenated at 500 psi and ambient temperature. The reaction was determined to be complete by TLC after 1.5 h. The mixture was filtered over a 1'' pad of celite to remove palladium on carbon. The filtrate was cooled to -20 ^{o}C and hexanes were added until a white precipitate was formed. The precipitate (2.1 g) was collected by vacuum filtration and washed with diethyl ether to give **1** in 85% yield. TLC (5:2 hexanes/ethyl acetate) R_f 0.25; IR (Thin Film) 3395, 3322, 1684, 1627, 1528, 1441, 1338, 1297, 1199, 1153 cm^{-1}; ^{1}H NMR (DMSO-d_6) δ 7.13 (d, J = 3.6 Hz, 1H), 6.60 (s, 2H), 5.09 (d, J = 3.6 Hz, 1H), 3.66, (s, 3H); ^{13}C NMR ($CDCl_3$) δ 123.8, 123.0, 121.9, 88.3, 86.3; EI-MS m/e 141.042 (M^+ calcd. for 141.042 $C_6H_7NO_3$).

Methyl 5-[(1-methyl-4-nitroimidazole-2-yl)carbonylamino]furan-2-carboxylate (NO_2-Im-Fr-OMe, **19**)**.** A mixture of **1** (1g, 7.08 mmol), NO_2-Im-$COCCl_3$ (2.31 g, 8.5 mmol), DIEA (1.1 g, 1.48 mL, 8.5 mmol) and ethyl acetate (14 mL) was stirred at 35 ^{o}C for 12 h. The reaction was cooled to ambient temperature and sufficient hexanes were added to completely precipitate a pale yellow solid. The precipitate was collected by vacuum filtration and washed with cold methanol and ether to give **19** (1.18 g) in 57%

yield. TLC (5:2 hexanes/ethyl acetate) R_f 0.20; IR (Thin Film) 3195, 3133, 1731, 1676, 1553, 1541, 1525, 1442, 1387, 1313, 1146 cm^{-1}; ^1H NMR (DMSO-d_6) δ 12.17 (br, 1H), 8.65 (s, 1H), 7.34 (d, J = 3.6 Hz, 1H), 6.52 (d, J = 3.6 Hz, 1H), 4.01 (s, 3H), 3.77 (s, 3H); ^{13}C NMR (DMSO-d_6) δ 158.6, 155.8, 150.0, 145.0, 137.3, 137.0, 127.8, 121.4, 99.0, 52.3, 37.3; EI-MS m/e 294.060 (M^+ calcd. for 294.060 $C_{11}H_{10}N_4O_6$).

Methyl 5-[(4-amino-1-methylimidazole-2-yl)carbonylamino]furan-2-carboxylate hydrochloride (HCl•H_2N-Im-Fr-OMe, **20**). A mixture of **19** (1 g, 3.4 mmol) and 10% palladium on carbon (0.2 g) in ethyl acetate (7 mL) was placed into a Parr apparatus and hydrogenated at 500 psi and ambient temperature. The reaction was complete after 1.5 h. The mixture was filtered over a 1'' pad of celite to remove the palladium on carbon. The ethyl acetate was removed in vacuo and hydrogen chloride in diethylether (2M) was added to give the ammonium chloride salt **20** (0.62 g) in 61% yield. TLC (5:2 hexanes/ethyl acetate) R_f 0.15 (amine) R_f 0.0 (salt); IR (Thin Film) 2985, 3008, 1708, 1689, 1537, 1318, 1193, 1142, 1119, 1018, 756, 668 cm^{-1}; ^1H NMR (DMSO-d_6) δ 11.73 (br, 1H), 7.54 (s, 1H), 7.33 (d, J = 3.6, 1H), 6.51 (s, J = 3.6 Hz, 1H), 3.96 (s, 3H), 3.76 (s, 3H); ^{13}C NMR (DMSO-d_6) δ 157.9, 155.0, 149.5, 136.3, 134.8, 129.5, 120.9, 118.7, 97.9, 51.6, 35.8; EI-MS m/e 264.086 (M^+ calcd. for 264.086 $C_{11}H_{12}N_4O_4$).

Methyl 5-({4-[(tert-butoxy)carbonylamino]-1-methylimidazole-2-yl}-carbonylamino)-furan-2-carboxylate (Boc-Im-Fr-OMe, **21**). A mixture of **20** (0.5 g, 1.6 mmol), Boc-anhydride (545 mg, 2.4 mmol), and DIEA (258 mg, 348 µl, 2 mmol) in DMF (5 mL) was stirred at 60 °C for 18 h. The reaction mixture was added to ice water (0.5 L) and the precipitate extracted with ethyl acetate (50 mL). The organic layer was dried with sodium sulfate and removed in vacuo. The resulting residue was subjected to

column chromatography (1:1 hexanes/ethylacetate) to provide **21** (376 mg) as a white solid in 62% yield. TLC (1:1 hexanes/ethyl acetate) R_f 0.65; IR (Thin Film) 3243, 2978, 1722, 1577, 1533, 1436, 1368, 1311, 1163, 1138, 755 cm^{-1}; ^1H NMR (DMSO-d_6) δ 11.11 (s, 1H), 9.52 (s, 1H), 7.33 (d, J = 3.3 Hz, 1H), 6.47 (d, J = 3.3 Hz, 1H), 3.90 (s, 3H), 3.76 (s, 3H), 1.44 (s, 9H); ^{13}C NMR (DMSO-d_6) δ 158.6, 156.1, 153.6, 150.4, 137.7, 136.8, 133.1, 121.7, 115.3, 97.8, 79.8, 52.3, 35.9, 28.9; EI-MS m/e 364.138 (M^+ calcd. for 364.138 $C_{16}H_{20}N_4O_6$).

5-({4-[(tert-butoxy)carbonylamino]-1-methylimidazole-2-yl}carbonylamino)-furan-2-carboxylic acid (Boc-Im-Fr-OH , **15**)**.** A mixture of **21** (0.3 g, 0.82 mmol), 1N NaOH (5 mL) and methanol (1 mL) was stirred at room temperature. The reaction was determined to be complete by TLC after 3 h. The methanol was removed in vacuo and the aqueous layer carefully adjusted to pH = 2 with 1N HCl. The milky white precipitate was extracted with ethyl acetate and the organics were dried over sodium sulfate. Filtration and evaporation of the organic layer gave **15** (270 mg) as a fine white powder in 94% yield. TLC (1:1 hexanes/ethyl acetate, 10% acetic acid) R_f 0.5; IR (Thin Film) 3231, 2917, 2856, 1688, 1542, 1311, 1259, 1163, 1119 cm^{-1}; ^1H NMR (DMSO-d_6) δ 10.94 (s, 1H), 9.52 (s, 1H), 7.30 (s, 1H), 7.22 (d, J = 3.6 Hz, 1H), 6.44 (d, J = 3.6 Hz, 1H), 3.91 (s, 3H), 1.44 (s, 9H); ^{13}C NMR (DMSO-d_6) δ 160.1, 159.3, 149.9, 136.5, 135.5, 133.2, 123.6, 115.4, 101.4, 97.0, 82.3, 35.9, 28.9; EI-MS m/e 350.123 (M^+ calcd. for 350.123 $C_{15}H_{18}N_4O_6$.

Methyl 5-{3-[(tert-butoxy)carbonylamino]propanoylamino}furan-2-carboxylate (Boc-β-Fr-OMe, **22**). A mixture of Boc-β-alanine (3.22 g, 17 mmol) and DCC (1.75 g, 8.5 mmol) in methylene chloride (25 mL) was stirred at ambient temperature for 30 min.

To the above mixture was added **1** (0.6 g, 4.25 mmol) as a solution in DMF (5 mL) and DIEA (0.741 mL, 0.55 g, 4.25 mmol), followed by the addition of DMAP (0.155 g, 1.27 mmol). The reaction was heated to 40 ^{o}C and allowed to stir overnight. The reaction was filtered to remove the DCU and the filtrate poured into ice water (0.5 L) upon which time a crude white precipitate formed. The crude precipitate was extracted with ethyl acetate and subjected to column chromatography (5:2 hexanes/ethyl acetate) to give **22** (1.1 g) as a flaky white solid in 85% yield. TLC (5:2 hexanes/ethyl acetate) R_f 0.2; IR (Thin Film) 3372, 3234, 3036, 1957, 1728 cm^{-1}; ^{1}H NMR (DMSO-d_6) δ 11.58 (s, 2H), 7.29 (d, J = 3.8 Hz, 1H), 6.86 (s, 2H), 6.36 (d, J = 3.8 Hz, 1H), 3.76 (s, 3H), 3.32 (q, J = 6 Hz, 2H), 2.69 (t, J = 6 Hz, 3H), 1.35 (s, 9H); ^{13}C NMR (DMSO-d_6) δ 167.9, 159.4, 155.8, 150.1, 137.0, 120.8, 95.2, 78.3, 51.6, 36.1, 35.8, 29.0; EI-MS m/e 312.132 (M^+ calcd. for 312.132 $C_{14}H_{20}N_2O_6$).

5-{3-[(tert-butoxy)carbonylamino]propanoylamino}furan-2-carboxylic acid (Boc-β-Fr-OH, **13**). A mixture of **22** (1.1 g, 3.52 mmol), 1N NaOH (15 mL) and methanol (5 mL) was stirred at ambient temperature. The reaction was determined to be complete by TLC after 4 h. The methanol was removed in vacuo and the aqueous layer carefully adjusted to pH 2 with 1N HCl. The milky white precipitate was extracted with ethyl acetate and the organics were dried over sodium sulfate. Filtration and evaporation of the organic layer gave **13** (0.97 g) as an off-white solid in 92% yield. TLC (5:2 hexanes/ ethyl acetate, 10% acetic acid) R_f 0.6; IR (Thin Film) 3321, 3270, 3979, 1684, 1522 cm^{-1}; ^{1}H NMR (DMSO-d_6) δ 11.28 (s, 2H), 6.98 (d, J = 3.6 Hz, 1H), 6.82 (s, 2H), 6.23 (d, J = 3.6 Hz, 1H), 3.19 (q, J = 6 Hz, 2H), 2.45 (t, J = 6 Hz, 2H), 1.35 (s, 9H); ^{13}C

NMR (DMSO-d_6) δ 168.7, 159.7, 156.0, 150.6, 137.1, 120.8, 95.8, 78.3, 36.9, 36.6, 29.0; EI-MS m/e 298.117 (M^+ calcd. for 298.116 $C_{13}H_{18}N_2O_6$).

2,2,2-Trichloro-1-pyrrol-2-ylethan-1-one (**23**). A solution of pyrrole (20.6 mL, 20 g, 298 mmol) and diethyl ether (86 mL) was added dropwise to trichloroacetyl chloride (71.9 mL, 117 g, 644 mmol) with stirring at 0 ^oC. The reaction was allowed to warm to room temperature and stirred overnight. The solvent was then removed in vacuo, and **23** (24 g) was recovered as a white solid in 38% yield upon precipetation from hexanes. TLC (5:2 hexanes/ ethyl acetate) R_f 0.75; IR (Thin Film) 3322, 1656, 1388, 1136, 1035, 953, 842, 808, 754, 733, 688 cm^{-1}; ^1H NMR (DMSO-d_6) δ 12.4 (s, 1H), 7.32 (m, J = 2.1 Hz, 1H), 7.29 (m, J = 2.1 Hz, 1H), 6.34 (m, J = 2.1 Hz, 1H); ^{13}C NMR (DMSO-d_6) δ 172.5, 130.0, 122.3, 121.9, 112.0, 95.9; EI-MS m/e 210.936 (M^+ calcd. for 210.936 $C_6H_4Cl_3NO$).

2,2,2-Trichloro-1-(5-nitropyrrol-2-yl)ethan-1-one (**24**). A solution of **23** (20 g, 95 mmol) and acetic anhydride (111 mL) was cooled to -40 ^oC and treated dropwise with 70% nitric acid (8.24 mL) over 2 h. After completion of addition, the reaction was warmed to room temperature over 2 h. The reaction was cooled back down to -40 ^oC and sufficient ice water was added to precipitate **24** (16.5 g) as a white solid in 68% yield. TLC (5:2 hexanes/ ethyl acetate) R_f 0.6; IR (Thin Film) 3316, 1676, 1551, 1518, 1405, 1379, 1317 cm^{-1}; ^1H NMR (DMSO-d_6) δ 13.62 (s, 1H), 8.33 (d, J = 3 Hz, 1H), 7.66 (d, J = 3 Hz, 1H); ^{13}C NMR (DMSO-d_6) δ 173.4, 137.6, 128.4, 122.0, 115.0, 94.4; EI-MS m/e 255.921 (M^+ calcd. for 255.921 $C_6H_3Cl_3N_2O_3$).

Ethyl 5-nitropyrrole-2-carboxylate (NO_2-Nh-OEt, **25**). To a mixture of **24** (10 g, 39 mmol) in ethanol (35 mL) at room temperature was added sodium ethoxide (4 g, 59

mmol). The reaction was stirred for 2 h and quenched with sulfuric acid. The mixture was cooled to 0 ^{o}C and ice water was added (0.5 L) to precipitate **25** (7 g) as a tan solid in 97% yield. TLC (5:2 hexanes/ethyl acetate) R_f 0.5; IR (Thin Film) 3263, 3152, 2979, 1687, 1565, 1508, 1365, 1323, 1207, 1017, 752 cm^{-1}; ^{1}H NMR (DMSO-d_6) δ 8.04 (d, J = 1.5 Hz, 1H), 7.23 (d, J = 1.5 Hz, 1H), 4.29 (q, J = 7.2 Hz, 2H), 1.29 (t, J = 7.2 Hz, 3H); ^{13}C NMR (DMSO-d_6) δ 160.0, 137.2, 124.9, 123.6, 110.1, 61.4, 14.9; EI-MS m/e 184.048 (M^+ calcd. for 184.048 $C_7H_8N_2O_4$).

Ethyl 5-aminopyrrole-2-carboxylate hydrochloride (HCl•H_2N-Nh-OEt, **2**). A mixture of **25** (3 g, 16 mmol) and 10% palladium on carbon (0.3 g) in ethyl acetate (25 mL) was placed into a Parr apparatus and hydrogenated at 500 psi and ambient temperature for 1.5 h. The mixture was filtered over a 1'' pad of celite to remove palladium on carbon. The filtrate was cooled to 20 ^{o}C and HCl in diethyl ether was added. Upon addition, the hydrochloride salt precipitated out and was collected by vacuum filtration to give **2** (2.4 g) as an off white solid in 78% yield. TLC (1:1 hexanes/ethyl acetate) (amine) R_f 0.15, (hydrochloride) R_f 0.0; IR (Thin Film) 2914, 1694, 1495, 1429, 1376, 1345, 1284, 1224, 1106, 1020, 965 cm^{-1}; ^{1}H NMR (DMSO-d_6) δ 12.21 (s, 1H), 10.13 (s, 1H), 7.10 (d, J = 1.8 Hz, 1H), 6.74 (d, J = 1.8 Hz, 1H), 4.24 (q, J = 7.2 Hz, 2H), 1.26 (t, J = 7.2 Hz, 3H); ^{13}C NMR (DMSO-d_6) δ 159.6, 121.3, 117.7, 115.6, 109.2, 60.0, 14.3; EI-MS m/e 154.074 (M^+ calcd. for 154.074 $C_7H_{10}N_2O_2$).

Ethyl 5-[(tert-butoxy)carbonylamino]pyrrole-2-carboxylate (Boc-Nh-OEt, **26**). A mixture of **2** (2 g, 11 mmole), Boc-anhydride (3.6 g, 16.5 mmol), and DIEA (2.1 mL, 1.56 g, 12.1 mmol) in DMF (15 mL) was stirred at 60 ^{o}C for 12 h. The mixture was then added to ice water (1 L) and extracted twice with ethyl acetate (150 mL). The organic

layer was dried over sodium sulfate and evaporated to give a crude oil. Column chromatography of the oil (5:2 hexanes/ethyl acetate) afforded **26** (2 g) as a flaky white solid in 72% yield. TLC (5:2 hexanes/ethyl acetate) R_f 0.65; IR (Thin Film) 3296, 1683, 1570, 1384, 1315, 1264, 1249 cm^{-1}; ^1H NMR (DMSO-d_6) δ 11.48 (s, 1H), 9.06 (s, 1H), 6.93 (d, J = 1.8 Hz, 1H), 6.58 (d, J = 1.8 Hz, 1H), 4.21 (q, J = 7.2 Hz, 2H), 1.41 (s, 9H), 1.24 (t, J = 7.2 Hz, 3H); ^{13}C NMR (DMSO-d_6) δ 160.1, 152.5, 119.1, 112.4, 105.2, 78.3, 59.4, 28.2, 14.4; EI-MS m/e 254.127 (M^+ calcd. for 254.127 $C_{12}H_{18}N_2O_4$).

5-[(tert-butoxy)carbonylamino]pyrrole-2-carboxylic acid (Boc-Nh-OH, **12**). A mixture of **26** (2 g, 7.9 mmol), 1N NaOH (15 mL) and methanol (5 mL) was stirred at ambient temperature. The reaction was determined to be complete by TLC after 3 h. The methanol was removed in vacuo and the aqueous layer carefully adjusted to pH 2 with 1N HCl. The milky white precipitate was extracted with ethyl acetate and the organics were dried over sodium sulfate. Filtration and evaporation of the organic layer gave **12** (1.6 g) as an off white solid in 92% yield. TLC (5:2 hexanes/ ethyl acetate, 10% acetic acid) R_f 0.5; IR (Thin Film) 3329, 3153, 2969, 1691, 1586, 1549, 1434, 1374, 1250, 1167, 1117, 1057, 961, 762 cm^{-1}; ^1H NMR (DMSO-d_6) δ 11.32 (s, 1H), 9.01 (s, 1H), 6.88 (s, 1H), 6.52 (s, 1H), 1.41 (s, 9H); ^{13}C NMR (DMSO-d_6) δ 162.3, 153.3, 125.4, 120.7, 112.6, 106.0, 79.0, 28.9; EI-MS m/e 226.095 (M^+ calcd. for 226.095 $C_{10}H_{14}N_2O_4$).

Methyl 4-cyano-3-methylbut-3-enoate (**27**). A mixture of acetoacetate (30 g, 258 mmol), cyanoacetic acid (24 g, 284 mmol), ammonium acetate (3.98 g, 51.6 mmol), acetic acid (6.65 mL, 6.98 g, 116 mmol) and benzene (75 mL) was stirred for 12 h at 145 ^oC in a round bottom equipped with a Dean Stark apparatus and condenser. The reaction was allowed to cool to ambient temperature, washed with brine (0.3 L), saturated sodium

bicarbonate (0.3 L) and dried over magnesium sulfate. The reaction was filtered and the solvent removed in vacuo. The crude product was distilled (60 ^{o}C, 0.1 mmHg) to give **27** (23 g) as a clear liquid and mixture of E and Z regioisomers in 65% yield. IR (Thin Film) 2957, 2221, 1741, 1437 cm^{-1}; ^{1}H NMR (DMSO-d_6) δ 5.69 (q, J = 0.6 Hz, 1H), 5.62 (q, J = 0.6 Hz, 1H), 3.61 (s, 3H), 3.60 (s, 3H), 3.42 (s, 2H), 3.35 (d, J = 1.2 Hz, 2H), 2.01 (d, J = 1.2 Hz, 3H), 1.93 (d, J = 1.2 z, 3H); ^{13}C NMR (DMSO-d_6) δ 170.1, 169.5, 158.4, 158.1, 117.4, 117.3, 99.6, 99.4, 52.8, 52.7, 42.8, 41.3, 23.6, 21.7; EI-MS m/e 139.063 (M^+ calcd. for 139.063 $C_7H_9NO_2$).

Methyl 5-amino-3-methylthiophene-2-carboxylate (HCl•H_2N-Tn-OMe, **3**). Diethylamine (18.7 mL, 13.2 g, 181 mmol) was added dropwise to a mixture of **27** (23 g, 165 mmol) and sulfur flakes (5.28 g, 165 mmol), in ethanol (130 mL) and stirred at room temperature for 3 h. The reaction was concentrated to a minimal volume in vacuo and placed on an ice bath. Concentrated HCl was slowly added to the mixture to give a light-orange solid. The precipitate was collected by vacuum filtration and washed repeatedly with diethyl ether to give **3** (19 g) in 68% yield. TLC (5:2 hexanes/ethyl acetate) (amine) R_f 0.55, (hydrochloride) R_f 0.0; IR (Thin Film) 3422, 3339, 3204, 2849, 1713, 1677, 1546, 1462, 1269, 1187, 1092 cm^{-1}; ^{1}H NMR (DMSO-d_6) δ 6.91 (s, 2H), 5.76 (s, 1H), 3.61 (s, 3H), 2.62 (s, 3H); ^{13}C NMR (DMSO-d_6) δ 163.5, 146.7, 145.5, 114.9, 114.7, 52.0, 16.6; EI-MS m/e 171.035 (M^+ calcd. for 171.035 $C_7H_9NO_2S$).

Methyl 3-methyl-5-[(1-methyl-4-nitroimidazole-2-yl)carbonylamino]thiophene-2-carboxylate (NO_2-Im-Tn-OMe, **28**). Compound **28** was synthesized from **3** (1g, 4.8 mmol) according to the procedure reported for **19**, providing **28** (0.87 g) as a yellow solid in 56% yield. TLC (5:2 hexanes/ethyl acetate) R_f 0.60; IR (Thin Film) 3125, 1649, 1543,

1506, 1382, 1312, 1267 cm^{-1}; ^{1}H NMR (DMSO-d_6) δ 12.43 (s, 1H), 8.66 (s, 1H), 6.95 (s, 1H), 4.03 (s, 3H), 3.79 (s, 3H), 2.41 (s, 3H); ^{13}C NMR (DMSO-d_6) δ 163.5, 155.7, 145.1, 144.7, 143.6, 136.9, 128.0, 118.8, 117.2, 52.2, 37.5, 16.6; EI-MS m/e 324.053 (M^+ calcd. for 324.053 $C_{12}H_{12}N_4O_5S$).

Methyl 5-[(4-amino-1-methylimidazol-2-yl)carbonylamino]-3-methylthiophene-2-carboxylate hydrochloride (HCl•H_2N-Im-Tn-OMe, **29**). Compound **29** was synthesized from **28** (0.5 g, 1.5 mmol) according to the procedure reported for **20**, providing **29** (321 mg) as a pale yellow solid in 63% yield. TLC (5:2 hexanes/ethyl acetate) (amine) R_f 0.25, (hydrochloride) R_f 0.0; IR (Thin Film) 3294, 1735, 1674, 1562, 1520, 1440, 1407, 1267, 1185, 1091 cm^{-1}; ^{1}H NMR (DMSO-d_6) δ 12.14 (s, 1H), 7.47 (s, 1H), 6.95 (s, 1H), 3.99 (s, 1H), 3.73 (s, 1H), 2.39 (s, 1H); ^{13}C NMR (DMSO-d_6) δ 177.9, 63.4, 157.6, 155.1, 144.1, 128.0, 118.6, 116.1, 52.2, 35.4, 16.4; EI-MS m/e 294.079 (M^+ calcd. for 294.079 $C_{12}H_{14}N_4O_3S$).

Methyl 5-({4-[(tert-butoxy)carbonylamino]-1-methylimidazole-2-yl}-carbonylamino)-3-methylthiophene-2-carboxylate (Boc-Im-Tn-OMe, **30**). Compound **30** was synthesized from **29** (300 mg, 0.91 mmol) according to the procedure reported for **21**, providing **30** (239 mg) as a pale-yellow solid in 67% yield. TLC (1:1 hexanes/ethyl acetate) R_f 0.7; IR (Thin Film) 3424, 3219, 2995, 1750, 1704, 1677, 1571, 1251, 1141 cm^{-1}; ^{1}H NMR (DMSO-d_6) δ 11.78 (s, 1H), 9.33 (s, 1H), 7.32 (s, 1H), 6.90 (s, 1H), 3.91 (s, 1H), 3.73 (s, 1H), 2.39 (s, 1H), 1.43 (s, 9H); ^{13}C NMR (DMSO-d_6) δ 162.8, 155.7, 152.8, 144.0, 143.4, 136.7, 132.4, 117.0, 115.6, 115.3, 79.4, 51.3, 35.2, 26.9, 15.9; EI-MS m/e 394.131 (M^+ calcd. for 394.131 $C_{17}H_{22}N_4O_5S$).

5-({4-[(tert-butoxy)carbonylamino]-1-methylimidazole-2-yl}carbonylamino)-3-methylthiophene-2-carboxylic acid (Boc-Im-Tn-OH, **16**). A mixture of **30** (200 mg, 0.507 mmol), methanol (1 mL) and 1N NaOH (5 mL) was stirred at 60 ^{o}C for 6 h. The methanol was removed in vacuo and the aqueous layer carefully adjusted to pH = 2 with 1N HCl. The milky white precipitate was extracted with ethyl acetate and the organics were dried over sodium sulfate. Filtration and evaporation of the organic layer gave **16** (158 mg) as a pale tan solid in 82% yield. TLC (1:1 hexanes/ethyl acetate, 10% acetic acid) R_f 0.8; (Thin Film) 3400, 2976, 3231, 2961, 1722, 1678, 1589, 1253, 1179, 1091 cm^{-1}; ^{1}H NMR (DMSO-d_6) δ 11.68 (s, 1H), 9.35 (s, 1H), 7.32 (s, 1H), 6.87 (s 1H), 3.93 (s, 3H), 2.38 (s, 3H), 1.44 (s, 9H); ^{13}C NMR (DMSO-d_6) δ 164.0, 158.3, 155.6, 142.8, 136.7, 117.3, 117.1, 115.2, 99.4, 81.4, 35.2, 28.1, 15.8; EI-MS m/e 380.115 (M^{+} calcd. for 380.115 $C_{16}H_{20}N_4O_5S$).

Methyl 5-{3-[(tert-butoxy)carbonylamino]propanoylamino}-3-methylthiophene-2-carboxylate (Boc-β-Tn-OMe, **31**). A mixture of Boc-β-alanine (1 g, 5.28 mmol) and DCC (545 mg, 2.64 mmol) in methylene chloride (10 mL) was stirred at ambient temperature for 30 min. The mixture was then filtered into a round bottom containing **3** (382 mg, 1.8 mmol), DIEA (322 μl, 239mg, 1.8 mmol), DMAP (100 mg, 0.8 mmol) and DMF (8 mL). The mixture was then heated at 45 ^{o}C. Progress of the reaction was monitored by TLC with additional symmetrical anhydride (1.4 eq) added every 8 h, as needed, until completion. The mixture was then added to brine (0.2 L) and extracted twice with ethyl acetate (50 mL). The organic layer was then washed with saturated sodium bicarbonate (0.1 L), 10 mM HCl (0.1 L), and dried over sodium sulfate. The crude residue was subjected to column chromatography (5:2 hexanes/ethyl acetate) to

give **31** (392 mg) as a flaky white powder in 62% yield. TLC (5:2 hexanes/ethyl acetate) R_f 0.32; IR (Thin Film) 3348, 3450, 2981, 1684, 1568, 1522, 1445, 1252 cm^{-1}; ^1H NMR ($CDCl_3$) δ 10.11 (s, 1H), 6.49 (s, 1H), 3.79 (s, 3H), 3.49 (q, J = 6 Hz, 2H), 2.64 (t, J = 6 Hz, 2H), 1.40 (s, 9H); ^{13}C NMR ($CDCl_3$) δ 168.7, 164.1, 157.0, 144.9, 143.4, 128.4, 117.3, 116.1, 80.4, 51.7, 37.0, 34.8, 16.4; EI-MS m/e 342.124 (M^+ calcd. for 342.124 $C_{15}H_{22}N_2O_5S$).

5-{3-[(tert-butoxy)carbonylamino]propanoylamino}-3-methylthiophene-2-carboxylic acid (Boc-β-Tn-OH, **14**). A mixture of **31** (200 mg, 0.58 mmol), methanol (4 mL) and 1N NaOH (15 mL) was stirred at 60 ^oC for 6 h. The methanol was removed in vacuo and the aqueous layer carefully adjusted to pH 2 with 1N HCl. The milky white precipitate was extracted with ethyl acetate and the organics were dried over sodium sulfate. Filtration and evaporation of the organic layer gave **14** (180 mg) as a light yellow solid in 94% yield. TLC (5:2 hexanes/ethyl acetate, 10% acetic acid) R_f 0.5; (Thin Film) 3255, 2976, 2954, 1674, 1569, 1522, 1445, 1253 cm^{-1}; ^1H NMR (DMSO-d_6) δ 12.35 (s, 1H), 11.42 (s, 1H), 6.89 (s, 1H), 6.48 (s 1H), 3.21 (q, J = 6 Hz, 2H), 2.50 (t, J = 6 Hz, 2H), 2.37 (s, 3H), 1.35 (s, 9H); ^{13}C NMR (DMSO-d_6) δ 168.2, 164.1, 155.4, 143.3, 143.1, 116.8, 115.0, 77.7, 36.3, 35.6, 28.2, 15.7; EI-MS m/e 328.109 (M^+ calcd. for 328.109 $C_{14}H_{20}N_2O_5S$).

5-nitrothiophene-2-carboxylic acid (NO_2-Dt-OH, **32**). A mixture of sodium hypochlorite (26.35 g, 291 mmol) and sodium hydrogen phosphate monohydrate (30.3 g, 219 mmol) in water (250 mL) was added dropwise to a solution of commercially available 5-nitrothiophene-2-carboxaldahyde (5g, 31.8 mmol) in acetone (0.6 L) at room temperature. Upon completion of addition, TLC showed total consumption of the

starting aldahyde. The reaction was washed with hexanes (0.1 L) and acidified to pH = 2 with 1N HCl. The mixture was extracted three times with diethyl ether (0.1 L) and dried over sodium sulfate. Evacuation of the organic layer gave **32** (3.74 g) as a white solid in 68% yield. TLC (5:2 hexanes/ethyl acetate, 10% acetic acid) R_f 0.55; IR (Thin Film) 3118, 3109, 2876, 1688, 1680, 1512, 1350, 1336, 1274 cm^{-1}; ^1H NMR (DMSO-d_6) δ 8.12 (d, J = 4.2 Hz, 1H), 7.73 (d, J = 4.2 Hz, 1H); ^{13}C NMR (DMSO-d_6) δ 162.2, 154.5, 141.0, 132.6, 130.5; EI-MS m/e 172.978 (M^+ calcd. for 172.978 $C_5H_3NO_4S$).

Methyl 5-nitrothiophene-2-carboxylate (NO_2-Dt-OMe, **33**). A mixture of **32** (3.5 g, 20.2 mmol), concentrated sulfuric acid (0.2 g, 110 μl, 2.0 mmol) and methanol (50 mL) was refluxed for 48 h. The methanol was removed in vacuo and the residue neutralized with 1N NaOH. The mixture was extracted twice with ethyl acetate (0.1 L) and the organics dried over sodium sulfate. Filtration and evaporation of the organic layer provided **33** (3.4 g) as a crystalline white solid in 91% yield. TLC (1:1 hexanes/ethyl acetate) R_f 0.8; IR (Thin Film) 3476, 3115, 1730, 1705, 1535, 1508, 1423, 1360, 1282, 1250, 1191, 997, 856, 748, 732 cm^{-1}; ^1H NMR (DMSO-*d*6) δ 8.21 (d, J = 3.9 Hz, 1H), 7.80 (d, J = 3.9 Hz, 1H), 3.87 (s, 3H); ^{13}C NMR (DMSO-d_6) δ 161.2, 155.0, 138.4, 133.3, 130.4, 54.0; EI-MS m/e 186.994 (M^+ calcd. for 186.994 $C_6H_5NO_4S$).

Methyl 5-aminothiophene-2-carboxylate hydrochloride (HCl•H_2N-Dt-OMe, **34**). Concentrated hydrochloric acid (5.8 mL) was added dropwise to a mixture of **33** (0.3 g, 1.6 mmol) and tin(II) chloride dihydrate (2.43 g, 12.8 mmol) in 95 % ethanol (5.8 mL) at room temperature. Sufficient cooling was necessary to keep the reaction temperature under 35 ^oC. The reaction was stirred at 35 ^oC for 2 h. The ethanol was removed in vacuo and the aqueous layer washed twice with hexanes (50 mL). The aqueous layer was

neutralized with 1N NaOH to pH = 9, upon which time a milky white emulsion formed. The mixture was extracted several times with ethyl acetate (50 mL) and the organics were dried over sodium sulfate. The ethyl acetate was evaporated to give a thin yellow film. Addition of HCl in diethyl ether (2 M) gave the ammonium chloride salt **34** (220 mg) as a white solid in 71% yield. TLC (1:1 hexanes/ethyl acetate) (amine) R_f 0.55, (hydrochloride) R_f 0.0; IR (Thin Film) 3219, 1731, 1706, 1471, 1272, 1088, 739 cm^{-1}; ^1H NMR (DMSO-d_6) δ 7.33 (d, J = 4.5 Hz, 1H), 6.70 (s, 2H), 5.87 (d, J = 4.5 Hz, 1H), 3.65 (s, 3H); ^{13}C NMR (DMSO-d_6) δ 163.5, 162.8, 136.2, 105.1, 51.8; EI-MS m/e 157.020 (M^+ calcd. for 157.020 $C_6H_7NO_2S$).

Methyl 5-[(1-methy-4-nitroimidazol-2-yl)-carbonylamino]-thiophene-2-carboxylate (NO_2-Im-Dt-OMe, **35**). Compound **35** was prepared from **34** (200 mg, 1.0 mmol) according to the procedure provided for **19**, providing **35** (164 mg) as a yellow solid in 51% yield. TLC (5:2 hexanes/ethyl acetate) R_f 0.50; IR (Thin Film) 3133, 1698, 1672, 1560, 1543, 1521, 1455, 1379, 1313, 1265, 1098, 746.9 cm^{-1}; ^1H NMR (DMSO-*d6*) δ 12.55 (s, 1H), 8.68 (s, 1H), 7.64 (d, J = 3.9 Hz, 1H), 7.16 (d, J = 3.9 Hz, 1H), 4.04 (s, 3H), 3.77 (s, 3H); ^{13}C NMR (DMSO-d_6) δ 163.1, 155.6, 145.8, 136.8, 132.6, 128.0, 123.5, 115.3, 52.6, 37.5; EI-MS m/e 310.037 (M^+ calcd. for 310.037 $C_{11}H_{10}N_4O_5S$).

Methyl 5-[(4-amino-1-methylimidazol-2-yl)carbonylamino]thiophene-2-carboxylate hydrochloride (HCl•H_2N-Im-Dt-OMe, **36**). Compound **36** was synthesized from **35** (150 mg, 0.48 mmol) according to the procedure provided for **20**, providing **36** (107 mg) as an off-white solid in 70% yield. TLC (ethyl acetate) (amine) R_f 0.45, (hydrochloride) R_f 0.0; IR (Thin Film) 3344, 3204, 2954, 1691, 1673, 1561, 1511, 1458, 1438, 1343, 1275, 1100 cm^{-1}; ^1H NMR (DMSO-d_6) δ 12.20 (s, 1H), 7.63 (d, J = 4.2 Hz,

1H), 7.40 (s, 1H), 7.11 (d, *J* = 4.2 Hz, 1H), 3.99 (s, 3H), 3.76 (s, 3H); ^{13}C NMR (DMSO-d_6) δ 177.4, 162.3, 159.4, 155.0, 145.4, 131.8, 128.5, 113.8, 51.7, 35.6; EI-MS m/e 280.063 (M^+ calcd. for 280.063 $C_{11}H_{12}N_4O_3S$).

Methyl 5-({4-[(tert-butoxy)carbonylamino]-1-methylimidazol-2-yl} carbonylamino) thiophene-2-carboxylate (Boc-Im-Dt-OMe, **37**). Compound **37** was prepared from **36** (100 mg, 0.35 mmol) according to the procedure provided for **21**, giving **37** (90 mg) as a white solid in 66% yield. TLC (1:1 hexanes/ethyl acetate) R_f 0.75; IR (Thin Film) 3282, 2964, 1707, 1692, 1673, 1573, 1550, 1368, 1341, 1273, 1159, 1096 cm^{-1}; ^{1}H NMR (DMSO-d_6) δ 11.91 (s, 1H), 9.32 (s, 3H), 7.62 (d, *J* = 4.2 Hz, 1H), 7.33 (s, 1H), 7.07 (d, *J* = 4.2 Hz, 1H), 3.94 (s, 3H), 3.76 (s, 3H), 1.44 (s, 9H); ^{13}C NMR (DMSO-d_6) δ 162.3, 155.5, 152.8, 145.7, 136.7, 132.4, 131.8, 122.0, 115.4, 113.5, 51.7, 35.2, 28.1; EI-MS m/e 380.115 (M^+ calcd. for 380.115 $C_{16}H_{20}N_4O_5S$).

5-({4-[(tert-butoxy)carbonylamino]-1-methylimidazol-2-yl}carbonylamino)-thiophene-2-carboxylic acid (Boc-Im-Dt-OH, **17**). Compound **17** was prepared from **37** (90 mg, 0.23 mmol) according to the procedure provided for **15**, giving **17** (80 mg) as a white solid in 91% yield. TLC (5:2 hexanes/ethyl acetate, 10% acetic acid) R_f 0.6; IR (Thin Film) 3246, 2965, 1674, 1556, 1509, 1456, 1314, 1271, 1237, 1162, 1102, 1023, 750 cm^{-1}; ^{1}H NMR (DMSO-d_6) δ 11.79 (s, 1H), 9.31 (s, 1H), 7.52 (d, *J* = 3.9 Hz, 1H), 7.32 (s, 1H), 7.04 (d, *J* = 3.9 Hz, 1H), 3.93, (s, 3H), 1.43 (s, 9H); ^{13}C NMR (DMSO-d_6) δ 163.4, 155.4, 152.8, 145.2, 136.7, 132.5, 131.3, 123.8, 115.2, 113.4, 79.1, 35.2, 28.1; EI-MS m/e 366.100 (M^+ calcd. for 366.100 $C_{15}H_{18}N_4O_5S$).

Methyl 3-hydroxythiophene-2-carboxylate (Ht-OMe, **38**). To dry methanol (81 mL) under nitrogen was added sodium metal (3.68 g, 304 mmol). After H_2 evolution had

stopped, the solution was cooled to 0 ^{o}C and methylthioglycolate (10 g, 179 mmol) was added dropwise. Methyl-2-chloroacrylate (10.88 g, 179 mmol) in MeOH (21 mL) was then added dropwise, resulting in the formation of a cloudy yellow precipitate. The solution was allowed to warm to room temperature and stirred for 2 hours, whereupon the precipitate turned dark brown. The solvent was removed in vacuo to give a dark-yellow solid, which was then acidified to pH = 2 with 4N HCl. The aqueous layer was extracted three times with methylene chloride (150 mL) and the resulting organic layer washed three times with water (150 mL) and dried over magnesium sulfate. Filtration and evaporation of the organic layer gave a dark oil that was subjected to column chromatography (20:1 hexanes/ethyl acetate) to give **38** (18.4 g) as a clear crystalline solid in 64.7% yield. TLC (20:1 hexanes/ethyl acetate) R_f 0.47; IR (Thin Film) 3334, 3112, 2955, 1716, 1664, 1552, 1444, 1415, 1350, 1296, 1208, 1104, 1032, 781 cm^{-1}; ^{1}H NMR ($CDCl_3$) δ 9.58 (s, 1H), 7.59 (d, J = 5.7 Hz, 1H), 6.75 (d, J = 4.8 Hz, 1H), 3.90 (s, 3H); ^{13}C NMR ($CDCl_3$) δ 131.7, 119.4, 52.2; EI-MS m/e 158.004 (M^+ calcd. for 158.004 $C_6H_6O_3S$).

Methyl 3-hydroxy-4-nitrothiophene-2-carboxylate (NO_2-Ht-OMe, **39**). **38** (15 g, 98 mmol) was added to concentrated H_2SO_4 (48 mL) and stirred until homogeneous. The solution was then cooled to -10–0 ^{o}C and HNO_3 (4.3 mL) in H_2SO_4 (24 mL) was added dropwise with sufficient cooling to keep the temperature below 0^{o} C. After the addition of HNO_3 was complete, the solution was allowed to stir at 0 ^{o}C for 3 h. The resulting black solution was added to ice and extracted three times with methylene chloride (150 mL) and dried over magnesium sulfate. The solvent was removed in vacuo and the resulting residue chromatographed over silica (5:1 hexanes/ethyl acetate) to give **39**

(8.2g) as a yellow solid in 37.6% yield. TLC (5:1 hexanes/ethyl acetate) R_f 0.12; IR (Thin Film) 3107, 1674, 1561, 1520, 1446, 1368, 1267, 1212, 1127, 974, 900, 842, 773 cm^{-1}; ^1H NMR ($CDCl_3$) δ 10.15 (s, 1H), 8.43 (s, 1H), 3.97 (s, 3H); ^{13}C NMR ($CDCl_3$) δ 164.6, 155.6, 132.6, 53.0; EI-MS m/e 202.989 (M^+ calcd. for 202.989 $C_5H_5NO_5S$).

Methyl 3-methoxy-4-nitrothiophene-2-carboxylate (NO_2-Mt-OMe, **40**). A mixture of **39** (1.5g, 7.4 mmol) and THF (29.5 mL) was cooled to 0 ^oC. Diazomethane (341 mg, 27 mL, 8.12 mmol) in diethyl ether was slowly added using a plastic funnel. After several seconds, nitrogen evolution ceased and the solution was allowed to warm to room temperature. A few drops of glacial acetic acid were added to ensure the complete consumption of diazomethane. The solvent was removed in vacuo to give **40** (1.49 g) as a yellow solid in 93% yield. TLC (1:1 hexanes/ethyl acetate) R_f 0.72; IR (Thin Film) 3115, 2960, 1725, 1555, 1506, 1453, 1435, 1387, 1353, 1281, 1199, 1110, 1059, 954, 772 cm^{-1}; ^1H NMR ($CDCl_3$) δ 8.38 (s, 1H), 4.10 (s, 1H), 3.94 (s, 3H); ^{13}C NMR ($CDCl_3$) δ 156.5, 131.3, 63.9, 53.0; EI-MS m/e 217.004 (M^+ calcd. for 217.004 $C_7H_7NO_5S$).

Methyl 4-amino-3-methoxythiophene-2-carboxylate hydrochloride ($HCl•H_2N$-Mt-OMe, **4**). A mixture of **4** (800 mg, 3.69 mmol) and tin(II) chloride dihydrate (13.3g, 58.9 mmol) in 95% EtOH (29.5 mL) were stirred vigorously at room temperature. Concentrated HCl (29.5 mL) was added dropwise and the solution heated at 35 ^oC for 6 h. The solution was removed from heat and adjusted to pH = 9 with 4N NaOH (15mL). The resulting white emulsion was extracted three times with ethyl acetate (100mL) and dried over magnesium sulfate. The solvent was removed in vacuo and a small amount of fresh ethyl acetate added. 2M hydrogen chloride in diethyl ether was added to precipitate

4 as the crude hydrochloride salt. The salt was filtered and taken directly on to the next step.

Methyl 4-[(tert-butoxy)carbonylamino]-3-methoxythiophene-2-carboxylate (Boc-Mt-OMe, **41**). A mixture of **4** (1.0 g, 4.47 mmol), TEA (498 mg, 0.68 mL, 4.92 mmol) and Boc-anhydride (1.0 g, 4.92 mmol) in methylene chloride (9 mL) was stirred at 60 ^{o}C for 12 hours. The solution was washed three times with saturated ammonium chloride solution and dried over magnesium sulfate. The organics were filtered and removed in vacuo. The resulting red solid was subjected to chromatography (5:1 hexanes/ethyl acetate) providing **41** (528 mg) as a white solid in 41.3% yield. TLC (10:1 hexanes/ethyl acetate) R_f 0.24; IR (Thin Film) 3437, 3329, 2979, 1772, 1716, 1530, 1440, 1376, 1230, 1165, 1081, 1055, 993, 861, 778 cm^{-1}; ^{1}H NMR ($CDCl_3$) δ 7.58 (s, 1H), 6.82 (s, 1H), 4.06 (s, 3H), 3.85 (s, 3H), 1.52 (s, 9H); ^{13}C NMR ($CDCl_3$) δ 161.5, 152.6, 151.5, 130.2, 112.0, 111.2, 85.5, 81.3, 62.9, 52.3, 28.6, 28.0; EI-MS m/e 287.083 (M^+ calcd. for 287.083 $C_{12}H_{17}NO_5S$).

4-[(tert-butoxy)carbonylamino]-3-methoxythiophene-2-carboxylic acid (Boc-Mt-OH, **11**). A mixture of **41** (250 mg, 0.87 mmol) and KOH (48.8 mg, 0.87 mmol) in dry methanol (1 mL), was heated to 50 ^{o}C for 6 hours. The solution was added to methylene chloride (5 mL) and water (5 mL). The aqueous layer was washed three times with methylene chloride (10 mL) and acidified to pH = 3 with 1N HCl (3.5 mL). The aqueous solution was then extracted three times with methylene chloride (10 mL) and dried over magnesium sulfate. Filtration and evaporation of the organic layer provided **11** (193 mg) as an off white solid in 81% yield. TLC (10:1 hexanes/ethyl acetate) R_f 0.24; IR (Thin Film) 3400, 1699, 1526, 1438, 1369, 1230, 1154, 1053 cm^{-1}; ^{1}H NMR ($CDCl_3$) δ 7.68

(s, 1H), 6.84 (s, 1H), 4.08 (s, 3H), 1.53 (s, 9H); ^{13}C NMR ($CDCl_3$) δ 165.8, 152.7, 130.4, 112.0, 111.2, 81.4, 63.3, 62.3, 28.7; EI-MS m/e 273.067 (M^+ calcd. for 273.067 $C_{11}H_{15}NO_5S$).

2-{(tert-butoxy)carbonylamino]-5-methyl-1,3-thiazole-4-carboxylic acid (Boc-Nt-OH, **10**). 2-ketobutyric acid (10 g, 98 mmol) was treated dropwise with bromine (8 mL, 25 g, 157 mmol) while stirring. Upon completion of addition, the reaction was allowed to stir until the red color of the bromine had dissipated. Thiourea (14.8 g, 196 mmol) was then added in portions and stirring was continued overnight. The mixture was acidified with conc. HCl and the precipitated hydrochloride was filtered and washed with cold ethanol. The crude solid was taken into DMF (50 mL), followed by the addition of DIEA (10 mL) and Boc anhydride (21.4 g, 98 mmol). The mixture was stirred at 60 ^{o}C for 12 h. The reaction was then diluted with ethyl acetate and washed three times with brine. The combined organics were dried over sodium sulfate and removed in vacuo to give a crude oil. The material was dissolved in methanol (0.1 L) and 1N NaOH (0.1 L), and stirred at room temperature for 1 h. The methanol was then removed and the aqueous layer washed two times with diethyl ether (0.1 L). The aqueous phase was acidified to pH 2 with 1 N HCl and extracted three times with ethyl acetate (0.1 L). The combined organics were dried over sodium sulfate and concentrated in vacuo to afford **10** (11 g) as a white flakey solid in 44% yield. (5:2 hexanes/ethyl acetate, 10% acetic acid) R_f 0.4; (Thin Film) 3191, 2978, 1714, 1669, 1578, 1562, 1317, 1165 cm^{-1}; ^{1}H NMR (DMSO-d_6) δ 11.54 (s, 2H), 2.54 (s 3H), 1.44 (s, 9H); ^{13}C NMR (DMSO-d_6) δ 164.1, 155.4, 137.1, 136.9, 81.9, 28.6, 12.94, 11.61; EI-MS m/e 258.067 (M^+ calcd. for 258.067 $C_{10}H_{14}N_2O_4S$).

Hairpin Polyamide Synthesis. Polyamides were synthesized from Boc-β-alanine-Pam resin (50 mg, 0.59 mmol/g) and purified by preparatory HPLC according to published manual solid-phase protocols.[4]

Im-Im-Nh-Py-γ-Im-Py-Py-Py-β-Dp (**43**). (Boc-Nh-OH) (33 mg, 0.147 mmol) was incorporated by activation with HBTU (53 mg, 0.140 mmol), DIEA (50 μl), and DMF (300 μl). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing NH_2-Py-γ-Im-Py-Py-Py-β-Pam resin. Coupling was allowed to proceed for 1.5 h at room temperature. After Boc-deprotection, Boc-Im-OH (35 mg, 0.147 mmol) was activated using HBTU (53 mg, 0.140 mmol), DIEA (50 μl), and DMF (300 μl). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing NH_2-**Nh**-Py-γ-Im-Py-Py-Py-β-Pam resin. Coupling was allowed to proceed for 1.5 h at room temperature, and determined to be complete by analytical HPLC. After Boc-deprotection, the terminal imidazole residue was added using Im-$COCCl_3$. Im-$COCCl_3$ (67 mg, 0.295 mmol), DIEA (50 μl), and DMF (600 μl) were added to the reaction vessel containing NH_2-Im-**Nh**-Py-γ-Im-Py-Py-Py-β-Pam resin. Coupling was allowed to proceed for 2 h at 37 ^oC, and determined complete by analytical HPLC. A sample of Im-Im-**Nh**-Py-γ-Im-Py-Py-Py-β-Pam resin (50 mg) was placed into a 20 mL scintillation vial, followed by Dp (1 mL). The mixture was allowed to stand for 2 h at 85 ^oC with occasional agitation. The resin was then filtered and the solution diluted to 8 mL using 0.1% TFA. The sample was purified by reversed phase HPLC to provide Im-Im-**Nh**-Py-γ-Im-Py-Py-Py-β-Dp (**32**) (2 mg, 5.6% recovery) as a fine white powder under lyophilization of the appropriate

fractions. MALDI-TOF-MS (monoisotopic), 1209.59 (M+H calcd. for 1209.56 $C_{56}H_{69}N_{22}O_{10}$).

Im-Im-Im-Py-γ-Im-Py-Py-Py-β-Dp (**42**). Boc-Im-OH was incorporated according to previously described procedures.[4] The terminal imidazole residue was incorporated and the compound purified as described above for **43** to provide Im-Im-**Im**-Py-γ-Im-Py-Py-Py-β-Dp (**42**) (2.6 mg, 6.0% recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1210.56 (M+H calcd. for 1224.56 $C_{56}H_{70}N_{23}O_{10}$).

Im-Im-Tn-Py-γ-Im-Py-Py-Py-β-Dp (**44**). (Boc-Im-Tn-OH) (56 mg, 0.147 mmol) was incorporated by activation with HBTU (53 mg, 0.140 mmol), DIEA (50 μl) and DMF (300 μl). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing NH_2-Py-γ-Im-Py-Py-Py-β-Pam resin. Coupling was allowed to proceed for 24 h at 37 ^oC. After Boc-deprotection the terminal imidazole residue was incorporated as described for **43**. The compound was cleaved from resin and purified as described for **43** to provide Im-Im-**Tn**-Py-γ-Im-Py-Py-Py-β-Dp (**44**) (2.1 mg, 5.7% recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1240.53 (M+H calcd. for 1240.53 $C_{57}H_{70}N_{21}O_{11}S$).

Im-Im-Dt-Py-γ-Im-Py-Py-Py-β-Dp (**45**). (Boc-Im-Dt-OH) was incorporated and Im-Im-**Dt**-Py-γ-Im-Py-Py-Py-β-Dp was synthesized, according to the procedure provided for **44** to give Im-Im-**Dt**-Py-γ-Im-Py-Py-Py-β-Dp (**45**) (1.8 mg, 4.9% recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1226.53 (M+H calcd. for 1226.52 $C_{57}H_{70}N_{21}O_{11}S$).

Im-Im-Ht-Py-γ-Im-Py-Py-Py-β-Dp (**46**). (Boc-Mt-OH) (42 mg, 0.147 mmol) was incorporated by activation with HBTU (53 mg, 0.140 mmol), DIEA (50 μl) and DMF (300 μl). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing NH_2-Py-γ-Im-Py-Py-Py-β-Pam resin. Coupling was allowed to proceed for 20 h at 37 ^oC. After Boc-deprotection, Boc-Im-OH (35 mg, 0.147 mmol) was activated using HBTU (53 mg, 0.140 mmol), DIEA (50 μl) and DMF (300 μl). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing NH_2-**Mt**-Py-γ-Im-Py-Py-Py-β-Pam resin. Coupling was allowed to proceed for 40 h at 37 ^oC, and determined to be complete by analytical HPLC. After Boc-deprotection the terminal imidazole residue was incorporated as described for **43**. The compound was cleaved from resin and purified as described for **43** to provide the methoxy protected Mt-containing polyamide Im-Im-**Mt**-Py-γ-Im-Py-Py-Py-β-Dp (2.0 mg, 5.4% recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1256.54 (M+H calcd for 1256.53 $C_{57}H_{70}N_{21}O_{11}S$). The polyamide was then dissolved in DMF (200 μl) and added to a suspension of sodium hydride (40 mg, 60% oil dispersion) and thiophenol in DMF (400 μl) that was pre-heated for 5 min at 100 ^oC. The mixture was heated for 2 h at 100 ^oC. The mixture was then cooled to 0 ^oC and 20% TFA (7.0 mL) was added. The aqueous layer was washed three times with diethyl ether (8 mL) and then diluted to a total volume of 9.5 mL using 0.1% TFA. The mixture was then purified by reverse-phase HPLC to give the deprotected Ht-containing polyamide Im-Im-**Ht**-Py-γ-Im-Py-Py-Py-β-Dp (**46**) (0.83 mg, 41% recovery) as a fine white powder under

lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1242.51 (M+H calcd. for 1242.51 $C_{56}H_{68}N_{21}O_{11}$).

Im-Im-Fr-Py-γ-Im-Py-Py-Py-β-Dp (**47**). (Boc-Im-Fr-OH) (51 mg, 0.147 mmol) was incorporated by activation with HBTU (53 mg, 0.140 mmol), DIEA (50 μl) and DMF (300 μl). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing NH_2-Py-γ-Im-Py-Py-Py-β-Pam resin. Coupling was allowed to proceed for 1.5 h at room temperature. After Boc-deprotection the terminal imidazole residue was incorporated as described for **43**. The compound was cleaved from resin and purified as described for **43** to provide Im-Im-Fr-Py-γ-Im-Py-Py-Py-β-Dp **47** (1.5 mg, 4.2% recovery) as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1210.54 (M+H calcd. for 1210.54 $C_{56}H_{68}N_{21}O_{11}$).

1:1 Motif Polyamide Synthesis. Polyamides were synthesized from Boc-β-alanine-Pam resin (50 mg, 0.59 mmol/g) and purified by preparatory HPLC according to published manual solid phase protocols.[4]

Im-β-Im-Py-β-Py-β-Im-Py-β-Dp (**48**). A sample of Im-β-Im-Py-β-**Py**-β-Im-Py-β-Pam resin (120 mg) was placed into a 20 mL scintillation vial, followed by Dp (2 mL). The mixture was allowed to stand for 2 h at 85 ^{o}C with occasional agitation. The resin was then filtered and the solution diluted to 8 mL using 0.1% TFA. The sample was then purified by reversed phase HPLC to provide Im-β-Im-Py-β-**Py**-β-Im-Py-β-Dp (**48**) (12 mg, 15.3% recovery) was provided as a fine white powder under lyophilization of the

appropriate fractions. MALDI-TOF-MS (monoisotopic), 1107.70 (M+H calcd. for 1107.53 $C_{50}H_{66}N_{20}O_{10}$).

Im-β-Im-Py-β-Hp-β-Im-Py-β-Dp (**49**). Polyamide **49** was synthesized, deprotected and purified according to the previously published protocol,[18] to provide Im-β-Im-Py-β-**Hp**-β-Im-Py-β-Dp (**49**) (5.6 mg 7.0% recovery) as a white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1124.20 (M+H calcd for 1124.19 $C_{50}H_{67}N_{20}O_{11}$).

Im-β-Im-Py-β-Nh-β-Im-Py-β-Dp (**50**). Boc-Nh-OH (271 mg, 1.2 mmol) was incorporated by activation with DCC (247 mg, 1.2 mmol) and HOBt (141 mg, 1.2 mmol) in DMF (2 mL). The mixture was shaken at 37 ^{o}C for 30 min and filtered into the reaction vessel containing NH_2-β-Im-Py-β-Pam resin. DIEA (400 μl) was added, and coupling was allowed to proceed for 1.5 h at ambient temperature. After Boc-deprotection, Boc-β-OH (227 mg, 1.2 mmol) was activated using HBTU (432 mg, 1.14 mmol), DIEA (400 μl), and DMF (2 mL). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing NH_2-Nh-β-Im-Py-β-Pam resin. Coupling was allowed to proceed for 2 h at 37 ^{o}C, and determined to be complete by analytical HPLC. The compound was cleaved from resin and purified as described for **48**. Im-β-Im-Py-β-**Nh**-β-Im-Py-β-Dp (**50**) (9 mg, 11.4% recovery) was provided as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1107.70 (M+H calcd. for 1107.53 $C_{49}H_{64}N_{20}O_{10}$).

Im-β-Im-Py-β-Ht-β-Im-Py-β-Dp (**51**). Polyamide **51** was synthesized using Boc-Mt-OH, deprotected and purified as described above for compound **46**, to provide Im-β-Im-Py-β-**Ht**-β-Im-Py-β-Dp (**51**) (1.1 mg 2.8% recovery) as a white powder under

lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1126.43 (M+H calcd. for 1126.47 $C_{49}H_{64}N_{19}O_{11}S$).

Im-β-Im-Py-β-Fr-β-Im-Py-β-Dp (**52**). Boc-β-Fr-OH (369 mg, 1.2 mmol) was incorporated by activation with DCC (247 mg, 1.2 mmol) and HOBt (141 mg, 1.2 mmol) in DMF (2 mL). The mixture was shaken at 37 ^{o}C for 30 min and filtered into the reaction vessel containing NH_2-β-Im-Py-β-Pam resin. DIEA (400 μl) was added, and coupling was allowed to proceed for 1.5 h at ambient temperature. The compound was cleaved from resin and purified as described for **48** to provide Im-β-Im-Py-β-Fr-β-Im-Py-β-Dp (**52**) (6 mg, 7.7% recovery) as a white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1094.50 (M+H calcd. for 1094.60 $C_{49}H_{63}N_{19}O_{11}$).

Im-β-Im-Py-β-Nt-β-Im-Py-β-Dp (**53**). Boc-Nt-OH (309 mg, 1.2 mmol) was incorporated by activation with HBTU (432 mg, 1.14 mmol), DIEA (400 μl) and DMF (2 mL). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing NH_2-β-Im-Py-β-Pam resin. Coupling was allowed to proceed for 20 h at 37 ^{o}C. After Boc-deprotection, Boc-β-OH (227 mg, 1.2 mmol) was activated using HBTU (432 mg, 1.14 mmol), DIEA (400 μl), and DMF (2 mL). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing NH_2-**Nt**-β-Im-Py-β-Pam resin. Coupling was allowed to proceed for 48 h at 37 ^{o}C, and determined to be complete by analytical HPLC. The compound was cleaved from resin and purified as described for **48**. Im-β-Im-Py-β-**Nt**-β-Im-Py-β-Dp (**53**) (4.2 mg, 5.2% recovery) was provided as a fine white powder under

lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1125.50 (M+H calcd. for 1125.49 $C_{49}H_{64}N_{20}O_{10}S$).

Im-β-Im-Py-β-Tn-β-Im-Py-β-Dp (**54**). Boc-β-Tn-OH (393 mg, 1.2 mmol) was incorporated by activation with HBTU (432 mg, 1.14 mmol), DIEA (400 μl), and DMF (2 mL). The mixture was allowed to stand for 15 min at room temperature and then added to the reaction vessel containing NH_2-β-Im-Py-β-Pam resin. Coupling was allowed to proceed for 20 h at 37 ^oC. The compound was cleaved from resin and purified as described for **48**. Im-β-Im-Py-β-**Tn**-β-Im-Py-β-Dp (**54**) (5.8 mg, 7.2% recovery) was provided as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1124.50 (M+H calcd. for 1124.49 $C_{50}H_{65}N_{19}O_{10}S$).

Im-β-Im-Py-β-Th-β-Im-Py-β-Dp **(55)**. Compound **55** was prepared using the identical protocol provided for **53**. Im-β-Im-Py-β-**Th**-β-Im-Py-β-Dp (**55**) (6.0 mg, 7.5% recovery) was provided as a fine white powder under lyophilization of the appropriate fractions. MALDI-TOF-MS (monoisotopic), 1125.60 (M+H calcd. for 1125.49 $C_{49}H_{64}N_{20}O_{10}S$).

Footprinting Experiments. Plasmids pDHN1 and pAU8 were constructed and 5'-radiolabeled as previously described.[12, 19] DNase I footprint titrations were performed according to standard protocols.[2]

Acknowledgements.

We are grateful for financial support from the National Institutes of Health and Caltech for a James Irvine Fellowship to R. M. D.

References:

[1] Kielkopf, C. L.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C., *Science* **1998,** 282, (5386), 111-115.

[2] Trauger, J. W.; Dervan, P. B., *Methods in Enzymology* **2001,** 340, 450-466.

[3] Urbach, A. R.; Love, J. J.; Ross, S. A.; Dervan, P. B., *Journal of Molecular Biology* **2002,** 320, (1), 55-71.

[4] Baird, E. E.; Dervan, P. B., *Journal of the American Chemical Society* **1996,** 118, (26), 6141-6146.

[5] Urbach, A. R.; Dervan, P. B., Essays in Contemporary Chemistry. In *From Molecular Structure Toward Biology*, ed.; Quinkert, G.; Kisakurek, M. V.; Verlag Helvetica Chimica Acta: Zurich, 2000.

[6] Dervan, P. B., *Bioorganic & Medicinal Chemistry* **2001,** 9, (9), 2215-2235.

[7] Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E., *Proceedings of the National Academy of Sciences of the United States of America* **1985,** 82, (5), 1376-1380.

[8] Pelton, J. G.; Wemmer, D. E., *Proceedings of the National Academy of Sciences of the United States of America* **1989,** 86, (15), 5723-5727.

[9] Lown, J. W.; Krowicki, K.; Bhat, U. G.; Skorobogaty, A.; Ward, B.; Dabrowiak, J. C., *Biochemistry* **1986,** 25, (23), 7408-7416.

[10] Wade, W. S.; Mrksich, M.; Dervan, P. B., *Journal of the American Chemical Society* **1992,** 114, (23), 8783-8794.

[11] Kielkopf, C. L.; Baird, E. E.; Dervan, P. D.; Rees, D. C., *Nature Structural Biology* **1998,** 5, (2), 104-109.

[12] Nguyen, D. H.; Szewczyk, J. W.; Baird, E. E.; Dervan, P. B., *Bioorganic & Medicinal Chemistry* **2001,** 9, (1), 7-17.

[13] Sharma, S. K.; Tandon, M.; Lown, J. W., *Journal of Organic Chemistry* **2000,** 65, (4), 1102-1107.

[14] Zhan, Z. Y. J.; Dervan, P. B., *Bioorganic & Medicinal Chemistry* **2000,** 8, (10), 2467-2474.

[15] Trauger, J. W.; Baird, E. E.; Dervan, P. B., *Nature* **1996,** 382, (6591), 559-561.

[16] Mrksich, M.; Parks, M. E.; Dervan, P. B., *Journal of the American Chemical Society* **1994,** 116, (18), 7983-7988.

[17] Trauger, J. W.; Baird, E. E.; Mrksich, M.; Dervan, P. B., *Journal of the American Chemical Society* **1996,** 118, (26), 6160-6166.

[18] Urbach, A. R.; Szewczyk, J. W.; White, S.; Turner, J. M.; Baird, E. E.; Dervan, P. B., *Journal of the American Chemical Society* **1999,** 121, (50), 11621-11629.

[19] Urbach, A. R.; Dervan, P. B., *Proceedings of the National Academy of Sciences of the United States of America* **2001,** 98, (8), 4343-4348.

[20] Bhupathy, M.; Conlon, D. A.; Wells, K. M.; Nelson, J. R.; Reider, P. J.; Rossen, K.; Sager, J. W.; Volante, R. P.; Dorsey, B. D.; Hoffman, J. M.; Joseph, S. A.; McDaniel, S. L., *Journal of Heterocyclic Chemistry* **1995,** 32, (4), 1283-1287.

[21] Rao, K. E.; Shea, R. G.; Yadagiri, B.; Lown, J. W., *Anti-Cancer Drug Design* **1990,** 5, (1), 3-20.

[22] Huddleston, P. R.; Barker, J. M., *Synthetic Communications* **1979,** 9, (8), 731-734.

[23] *Sparten Essential*, Wavefunction Inc: 2000.

[24] Herman, D. M.; Turner, J. M.; Baird, E. E.; Dervan, P. B., *Journal of the American Chemical Society* **1999,** 121, (6), 1121-1129.

[25] Kelly, J. J.; Baird, E. E.; Dervan, P. B., *Proceedings of the National Academy of Sciences of the United States of America* **1996,** 93, (14), 6981-6985.

[26] de Clairac, R. P. L.; Seel, C. J.; Geierstanger, B. H.; Mrksich, M.; Baird, E. E.; Dervan, P. B.; Wemmer, D. E., *Journal of the American Chemical Society* **1999,** 121, (13), 2956-2964.

[27] Kielkopf, C. L.; Bremer, R. E.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C., *Journal of Molecular Biology* **2000,** 295, (3), 557-567.

[28] White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B., *Nature* **1998,** 391, (6666), 468-471.

[29] Janssen, S.; Durussel, T.; Laemmli, U. K., *Molecular Cell* **2000,** 6, (5), 999-1011.

[30] White, S.; Turner, J. M.; Szewczyk, J. W.; Baird, E. E.; Dervan, P. B., *Journal of the American Chemical Society* **1999,** 121, (1), 260-261.