Chapter 5B

Further Progress toward the Development of Novel N-Terminal Recognition Motifs

Background and Significance

The vast potential of minor groove-binding polyamides as therapeutic agents or as tools for molecular biology is derived from their ability to recognize predetermined DNA sequences. Hairpin polyamides, in particular, have proven well-suited to this role, binding biologically relevant sequences with affinities and specificities that rival naturally occurring proteins. The *pairing rules* define a general recognition code for all four Watson-Crick base pairs using cofacial arrangements of Im or Hp with Py; however, the specificity of these cofacial pairings depends on their context within a given polyamide. This is best illustrated by the absence of specificity exhibited by *N-terminal* Hp/Py pairings relative to their *internal* counterparts.

The lack of N-terminal thymine specificity has immediate implications on the array of DNA sequences that can be targeted by polyamides, and within the context of the hairpin motif, further sequence limitations are imposed by the T, A selectivity of the aliphatic turn and tail positions. The sequence restrictions stipulated by the hairpin motif have driven the development of new strategies for covalently linking antiparallel strands and have inspired the design of novel N-terminal residues. The latter approach is particularly advantageous since lessons learned from physical studies on N-terminal recognition may prove applicable to internal contexts as well.

Previous efforts to replace N-terminal Hp residues with hydroxybenzamide moieties were moderately successful, affording 3-4 fold thymine selectivity in a doubly-charged model hairpin polyamide; however, the synthetic requirements of this residue offset it modest selectivity. Further research efforts, described in Chapter 5A, abandoned hydrogen bond-based thymine-recognition in favor of a purely shape selective mechanism of discrimination aimed at the asymmetric cleft of the T•A base pair. The selectivity displayed by N-terminal 3-methoxythiophene -2-carboxamide residues is almost unparalleled within the hairpin context; however, since this residue contains an aryl ether, it is not synthetically compatible with other thymine-specific residues using standard solid phase protocols. Polyamides containing 3-chlorothiophene-2-carboxamide (Ct) residues, on the other hand, are readily accessible by solid phase methods, exhibit 3-4 fold selectivity for T•A, and bind their target sequences at subnanomolar concentrations.

The significance of thymine-selective, N-terminal recognition motifs can be illustrated in the context of a model eight ring hairpin targeting a six bp site (Figure 5.7). Either oligonucleotide strand of a DNA sequence can be used to design a polyamide, and in order to simplify analysis, only pairings adjacent to the tail and turn are considered in the following discussion. Given the T, A preferences of the turn and tail, there are sixteen generalized 6 bp sequences amenable to hairpin recognition; however, if Im is the only N-terminal residue that can be employed, only seven of these generalized sites are accessible. The availability of Ct as an N-terminal, thymine-selective residue allows five additional subsequences to be targeted, accounting for 75% of the possible 6 bp hairpin binding sites. Inclusion of

a third C- or A-specific N-terminal residue would allow all but one subsequence to be recognized.

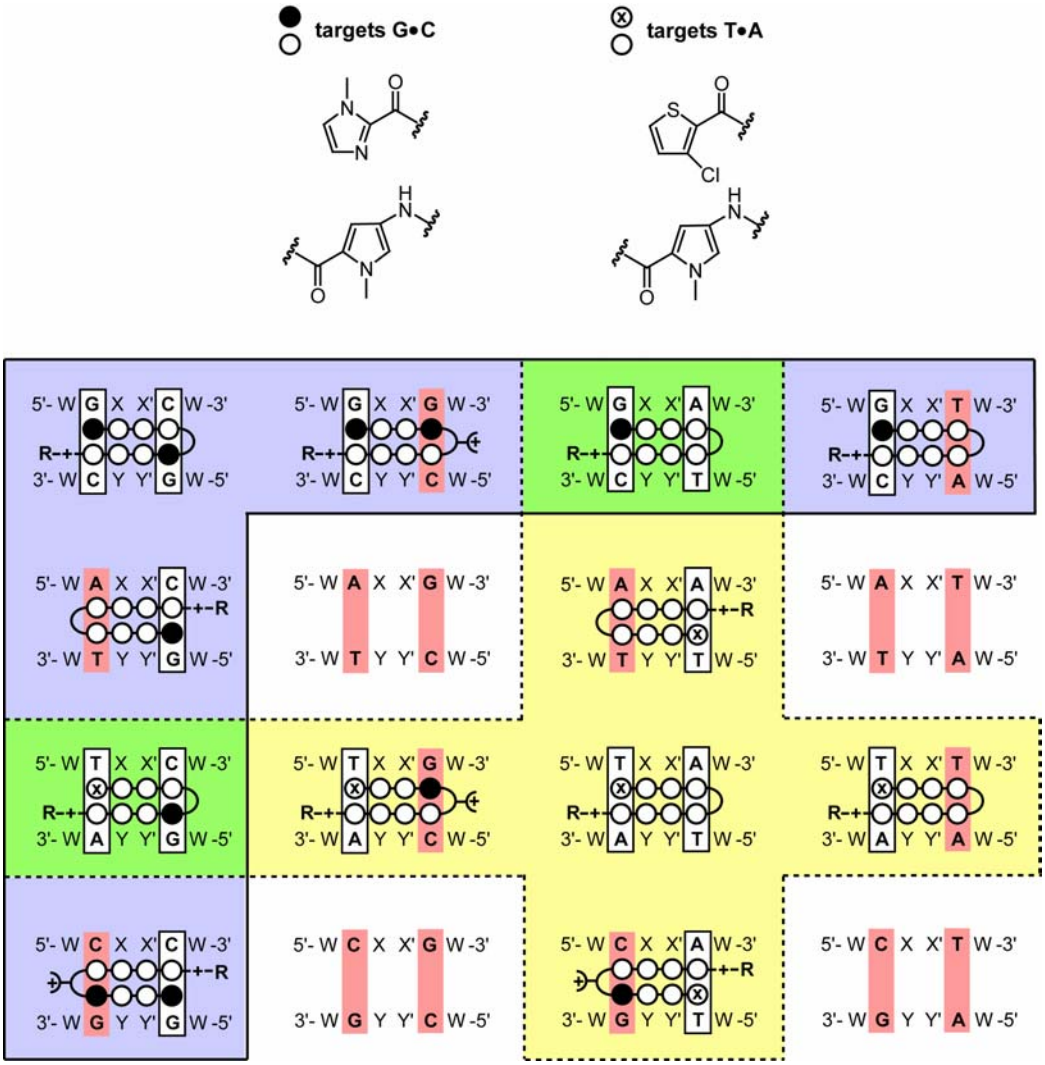


Figure 5.7 The development of a second sequence specific N-terminal pairing expands the repertoire of sequences that can be targeted by hairpin polyamides. The chemical feature of N-terminal Im/Py and Ct/Py pairings are illustrated (*top*). Six bp binding sites amenable to hairpin polyamide recognition are listed. Base pairs that can be recognized by the above N-terminal pairings are enclosed within white boxes. Base pairs not amenable to existing N-terminal residues are shown in red boxes. Sequences that can be targeted with an N-terminal Im/Py pairing are enclosed within the blue area while sequences that can be targeted with Ct/Py pairings are indicated in yellow. The two sequences that are amenable to both pairings are shown in green.

Expanding the array of sequences available to polyamides through novel residue development is especially important given that several polyamides composed of Im and Py have shown poor nuclear localization in cell culture. In this regard, novel residues can allow alternative polyamides to be designed, and may offer enhanced uptake properties with respect to Im and Py.

N-Terminal Recognition—Previous Work Revisited

The lack of specificity exhibited by N-terminal Hp residues has been attributed to rotation of the terminal ring, orienting the hydroxyl recognition element away the minor groove. This explanation is supported by molecular modeling which identifies stabilizing intramolecular hydrogen bonding interactions between the hydroxyl and the carbonyl oxygen of the 2-carboxamide moiety in this alternative rotamer. The enhanced selectivity of the second generation thymine-specific, N-terminal residue, 2-hydroxy-6-methoxybenzamide, was explained by destabilization of the alternative rotamer by unfavorable steric contacts between the 6-methoxy group and the floor of the minor groove. In light of the results reported in Chapter 5A, it now seems possible that the 6-methoxy-2-hydroxybenzamide residue might also operate by a shape selective mechanism, in which intramolecular hydrogen bonding drives the bulkier methoxy substituent into the minor groove. A similar conformational disposition was noted in models of N-terminal 3-hydroxythiophene-2-carboxamide residues, in which the polarizable sulfur atom is projected into the minor groove by intramolecular forces. This prediction is confirmed by thermodynamic data obtained from DNase I footprinting, which shows that thiophene-2-carboxamide and 3-

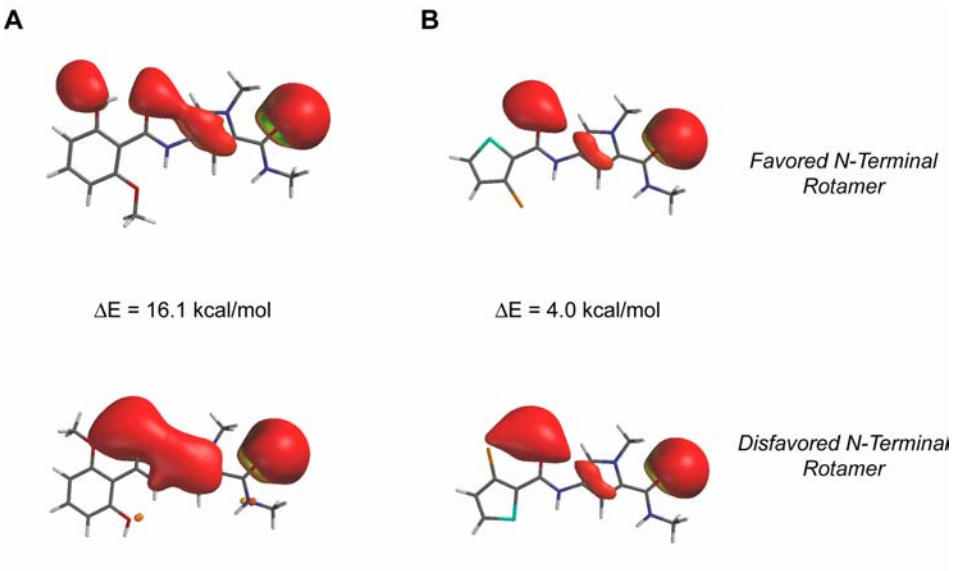


Figure 5.8 Molecular models of N-terminal Hp replacements. Electron densities were plotted over minimized equilibrium geometry for each possible planar rotamer of the terminal rings. Preferred conformers are shown above alternate conformers and the energy difference are indicated. **(A)** 6-Methoxy-2-hydroxybenzamide. **(B)** 3-Chlorothiophene-2-carboxamide.

hydroxythiophene-2-carboxamide have strikingly similar T, A selectivity (Table 5.3), presumably owing to unfavorable steric interactions with the exocyclic amine of guanine. Figure 5.8 compares molecular models generated for N-terminal 6-methoxy-2-hydroxybenzamide and 3-chlorothiophene-2-carboxamide residues.

N-Terminal Shape Selective Recognition—Unpublished Results

Several heterocyclic N-terminal residues, in addition to those discussed in Chapter 5A, were examined for specificity when paired opposite Py in a model eight ring polyamide. While these residues did not exhibit the desirable properties shown by Ct, their recognition profiles support the proposed shape selective model and could find use in different polyamide contexts downstream. The chemical structures

and DNA-recognition properties of these novel residues are summarized in Table 5.3.

Table 5.3 Comprehensive summary of equilibrium association constants (M^{-1})[a]

Polyamide	R	T·A	A·T	G·C	C·G
1	3-H	$6.0\ (0.7) \times 10^9$	$4.7\ (0.7) \times 10^9$	$4.3\ (0.4) \times 10^8$	$2.2\ (0.3) \times 10^9$
2	3-CH_3	$2.3\ (0.4) \times 10^9$	$1.4\ (0.2) \times 10^9$	$1.0\ (0.4) \times 10^7$	$1.0\ (0.3) \times 10^7$
3	3-NH_2	$6.3\ (1.0) \times 10^9$	$4.6\ (0.6) \times 10^9$	$7.8\ (0.9) \times 10^8$	$2.2\ (0.3) \times 10^8$
4	3-NHAc	$5.9\ (0.3) \times 10^9$	$2.9\ (0.1) \times 10^9$	$6.6\ (0.4) \times 10^8$	$6.0\ (0.2) \times 10^8$
5	3-OH	$6.2\ (0.6) \times 10^9$	$4.5\ (0.6) \times 10^9$	$2.1\ (0.3) \times 10^8$	$8.4\ (0.1) \times 10^7$
6	3-OCH_3	$2.0\ (0.4) \times 10^9$	$3.2\ (0.6) \times 10^8$	$\leq 1.0 \times 10^7$	$\leq 1.0 \times 10^7$
7	3-F	$1.2\ (0.2) \times 10^{10}$	$3.9\ (0.3) \times 10^9$	$3.7\ (0.4) \times 10^8$	$2.9\ (0.3) \times 10^8$
8	3-Cl	$1.3\ (0.2) \times 10^{10}$	$3.7\ (0.2) \times 10^9$	$3.1\ (0.6) \times 10^8$	$2.1\ (1.1) \times 10^8$
15	5-CH_3	4.7×10^9	4.1×10^9	6.0×10^8	2.1×10^8
16	3-CN	3.8×10^9	1.4×10^9	1.4×10^9	3.5×10^8
17	3-$C(NH_2)$Dp	6.9×10^9	3.1×10^9	6.3×10^8	7.2×10^8
18	3-Br	$1.5\ (0.1) \times 10^9$	$1.2\ (0.4) \times 10^9$	$2.7\ (0.3) \times 10^8$	$3.0\ (0.4) \times 10^8$
Polyamide	X	T·A	A·T	G·C	C·G
9	Im (X=N)	$3.8\ (0.3) \times 10^9$	$2.8\ (0.2) \times 10^9$	$7.0\ (0.9) \times 10^{10}$	$3.2\ (0.4) \times 10^9$
10	Py (X=CH)	$5.1\ (0.6) \times 10^9$	$3.1\ (0.3) \times 10^9$	$1.1\ (0.1) \times 10^9$	$2.6\ (0.3) \times 10^8$

[a] Values reported are mean results determined by at least three DNase I footprint titrations, with standard deviation given in parentheses. Values reported without deviations are the average of duplicate experiments. Assays were performed at 22 °C in a buffer containing 10 mM Tris·HCl, 10 mM KCl, 10 mM $MgCl_2$, and 5 mM $CaCl_2$ at pH 7.0.

A direct comparison of 3-methylthiophene-2-carboxamide and 5-methylthiophene-2-carboxamide verifies the significance of the 3-position of terminal rings in DNA recognition (Figure 5.9). It was envisioned that introduction of a cyano group at the 3-position of the thiophene-2-carboxamide scaffold would positively polarize the sulfur atom, allowing for more complementary recognition of the thymine O2; however, subsequent modeling indicated that electronic repulsions between the cyano group and the carboxamide oxygen were severe enough to prefer a “sulfur up” conformation with the cyano group directed toward the minor groove.

Thermodynamic data for this residue (Figure 5.10A) indicates a preference for T, A, and G relative to C, and the observed preference is best explained by the existence of both possible N-terminal rotamers. Presumably, the “sulfur down” orientation would prefer T, A, while the “cyano down” conformer could recognize the exocyclic amine of guanine through hydrogen bonding, though the greater steric bulk of the

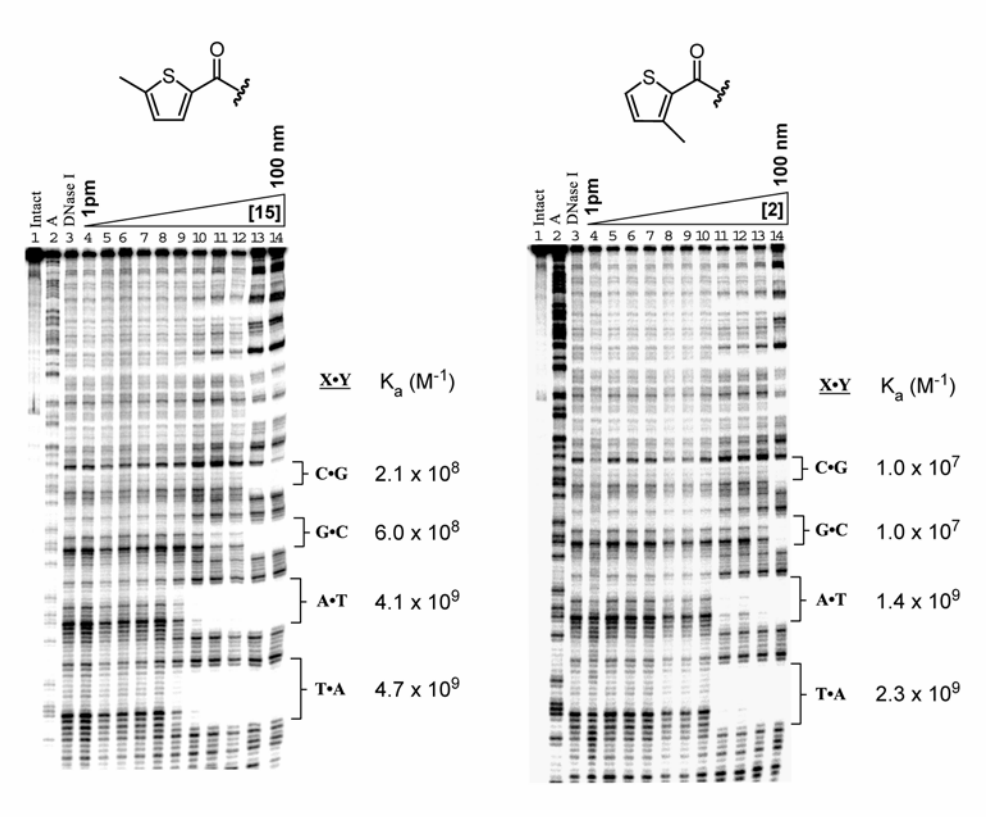


Figure 5.9 DNase I footprinting gels for 5- and 3-methylthiophene-2-carboxamide N-terminal residues. Observed association constants were determined using the previously described designed plasmid.

cyano group would be expected to have negative impact on overall affinity. Molecular models of 3-cyanothiophene-2-carboxamide and imidazole-2-carboxamide demonstrate the similarity of their groove-directed molecular surfaces and the greater size of the cyano substituent (Figure 5.10B).

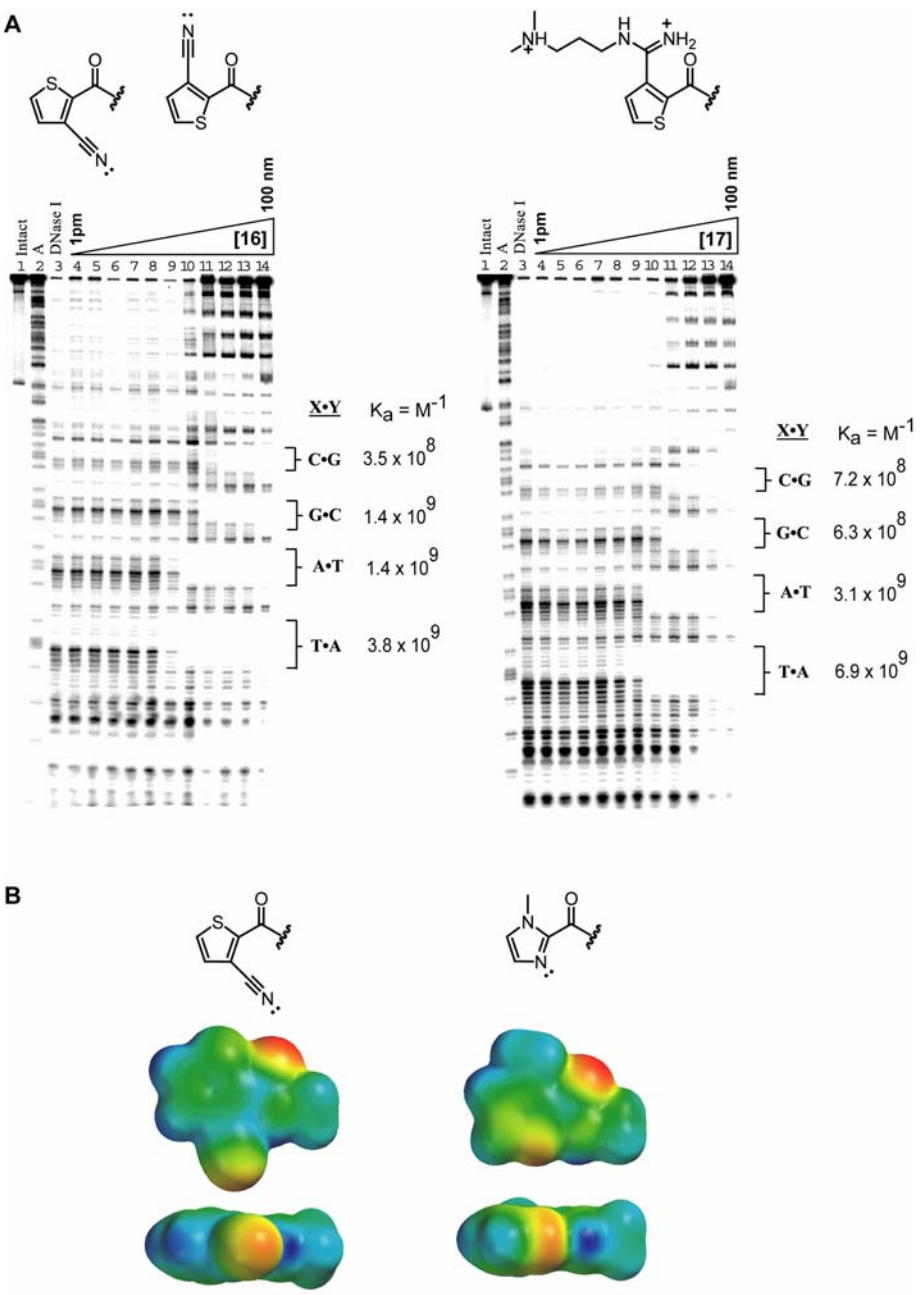


Figure 5.10 3-Cyanothiophene-2-carboxamide residues and derivatives as N-terminal recognition elements. **(A)** DNase I footprinting gels for 3-cyano- and 3-dimethylaminoproylamidino-thiophene-2-carboxamide cap residues. **(B)** Electropotential surfaces plotted onto minimized equilibrium geometries of an N-terminal 3-cyanothiophene (*left*) and Im residue (*right*).

The synthesis of the 3-cyanothiophene-2-carboxylic acid building block was achieved by nucleophilic substitution of ethyl 3-chlorothiophene-2-carboxylate with Cu(I)CN in refluxing NMP, followed by saponification in alkaline methanol (Figure 5.11A). The overall yield was low, however, better results can be obtained by substitution of the corresponding ethyl 3-bromothiophene-2-carboxylate or by reaction of the diazonium salt of methyl 3-aminothiophene-2-carboxylate with Cu(I)CN (R. Burli; personal communication). This monomer is easily coupled to support bound polyamides; however, treatment of the resulting resin with Dp resulted in formation of the amidine as the major product (Figure 5.11B). Solution phase coupling of the cyano building block, after cleavage of polyamide from resin, gave reasonable yields of the desired product (5.11C). The DNA-binding energetics of the amidine by-product were also determined (Figure 5.10A), showing T, A selectivity relative to G, C, suggesting that modification of a 3-cyano substituent with a diamine could prove useful in conjugate preparation at some point. Novel heterocyclic residues designed for recognition of G•C base pairs have also been examined, though to a lesser extent, with the aim of improving nuclear localization and providing alternatives to Im in contexts where it is less than optimal (Chapter 7).

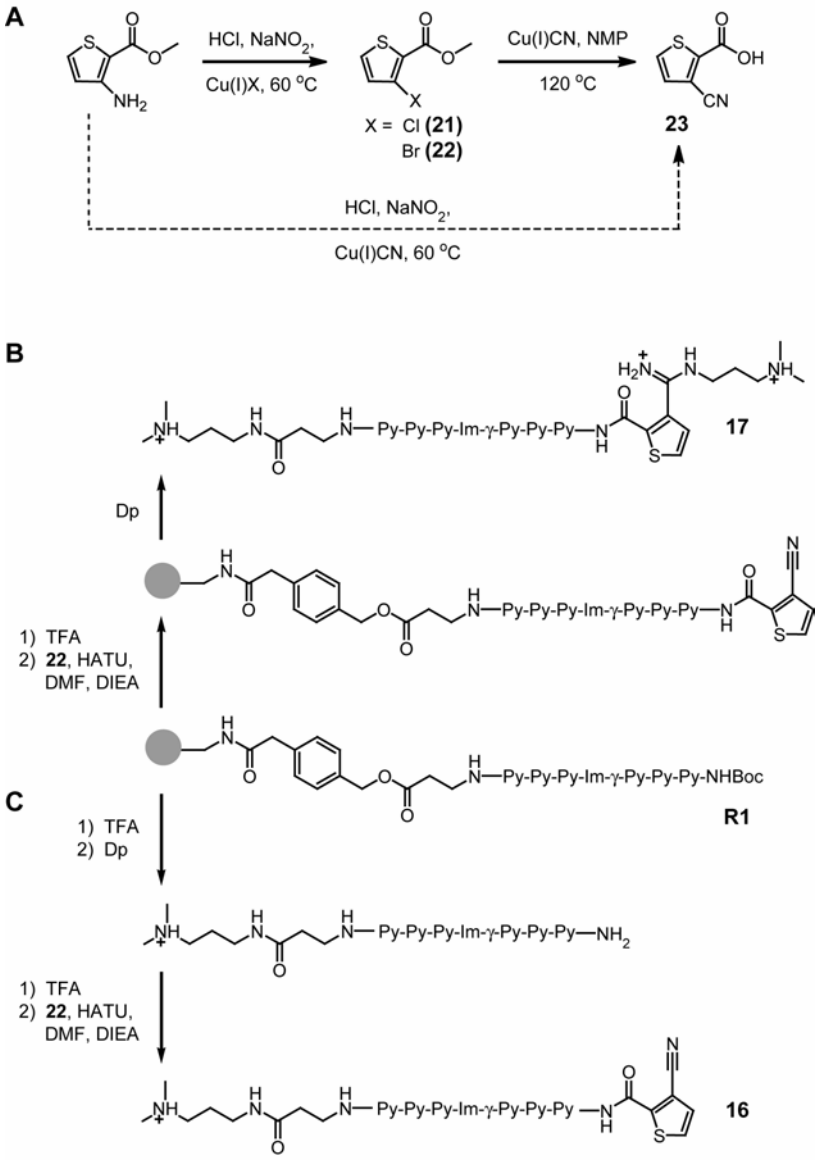


Figure 5.11 Synthesis of 3-cyano and 3-amidino-thiophene cap building blocks and polyamides. **(A)** Schemes to 3-cyanothiophene-2-carboxylate. **(B)** Solid phase synthesis of 3-amidino-thiophene-containing hairpin polyamide. **(C)** Solution phase synthesis of 3-cyano-thiophene-containing hairpin polyamide.

Nuclear Localization of Polyamides Containing 3-Chlorothiophene

All of the polyamides containing 3-chlorothiophene residues tested thus far have shown nuclear localization in cell culture. Fluorescein, Oregon Green, and TMR were all trafficked to the nucleus by hairpin polyamides containing an N-terminal Ct residue, in a range of cell lines (Figure 5.12 and Table 5.4). The uptake properties of the TMR conjugate could imply that Ct positively influences nuclear localization.

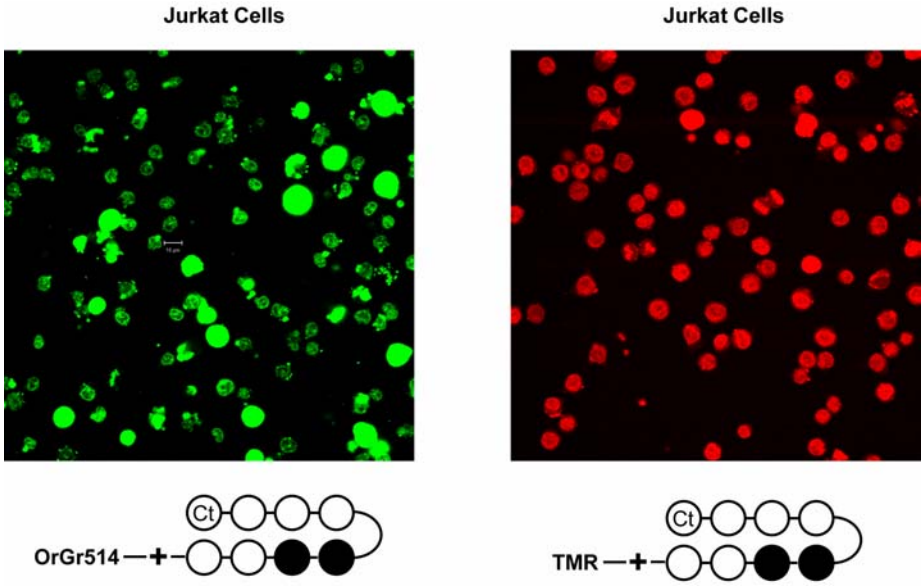
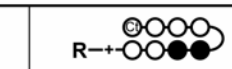


Figure 5.11 Nuclear localization of 3-chlorothiophene-containing polyamides. Hairpin polyamides with N-terminal Ct residues were labeled with a C-terminal Oregon Green 514 or 5-TMR fluorescent probe, and their cellular uptake was examined in a range of cell lines. The pictures above are in Jurkat cells.

Table 5.4 Cellular localization of Ct-containing hairpin polyamides in a panel of cell lines.

Cell Lines	 R = FAM	 R = OrGr514	 R = TMR	 R = FITC
DLD-1	+ +	+ +	+	+
HeLa	+ +	+ +	-	+ +
MCF-7	+ +	+ +	+	+
SK-BR-3			+	+
HCT-116	+ +	+ +	+	+ +
786O	+	+	- -	+
LNCaP	+	+		
PC3	+ +	+ +	+	+ +
MEL	+	+	+ +	+
NB4	+	+	+	+
Jurkat	+ +	+ +	+ +	+
CEM	+ +	+ +	+	+
MEG-01	+ +	+ +	+	+

Biological Applications of 3-Chlorothiophene

The recognition and uptake properties of Ct combined with its ease of incorporation in solid phase synthesis have encouraged its use in several biological collaborations. Chromosome staining with centromere-specific fluorophore-polyamide conjugates containing Ct was described in Chapter 3B. Projects aimed at controlling gene expression have also employed Ct as a recognition element. Among these ongoing collaborations, targets include the human multi-drug resistance gene as well as the hypoxia inducible factor linked to angiogenesis.

Future Directions for N-Terminal Design

Until all four Watson-Crick base pairs can be targeted in the context of living cells, the development of novel recognition motifs will continue to be an active area of research in the polyamide field. Tackling either molecular recognition or nuclear uptake individually presents a great challenge, however, confronting both hurdles will require iterative, rational design of novel residues and subsequent screening for DNA recognition and cellular localization on a case-by-case basis. Novel N-terminal residues now on the drawing board include 3-substituted benzothiophene residues.