J|A|C|S
ARTICLES
Published on Web 02/16/2002

Influence of a Terminal Formamido Group on the Sequence Recognition of DNA by Polyamides

Eilyn R. Lacy,[†] N. Minh Le,[‡] Carly A. Price,[‡] Moses Lee,*[,‡] and W. David Wilson*[,†]

Contribution from the Department of Chemistry, Georgia State University, Atlanta, Georgia 30303, and Department of Chemistry, Furman University, Greenville, South Carolina 29613

Received May 7, 2001

Abstract: Pyrrole (Py)-imidazole (Im)-containing polyamides bind in the minor groove of DNA and can recognize specific sequences through a stacked antiparallel dimer. It has been proposed that there are two different low energy ways to form the stacked dimer and that these are sensitive to the presence of a terminal formamido group: (i) a fully overlapped stacking mode in which the N-terminal heterocycles of the dimer stack on the amide groups between the two heterocycles at the C-terminal and (ii) a staggered stacking mode in which the N-terminal heterocycles are shifted by approximately one unit in the C-terminal direction (*Structure* **1997**, *5*, 1033−1046). Two different DNA sequences will be recognized by the same polyamide stacked in these two different modes. Despite the importance of polyamides as sequence specific DNA recognition agents, these stacking possibilities have not been systematically explored. As part of a program to develop agents that can recognize mismatched base pairs in DNA, a set of four polyamide trimers with and without terminal formamido groups was synthesized, and their interactions with predicted DNA recognition sequences in the two different stacking modes were evaluated. Experimental difficulties in monitoring DNA complex formation with polyamides were overcome by using surface plasmon resonance (SPR) detection of the binding to immobilized DNA hairpin duplexes. Both equilibrium and kinetic results from SPR show that a terminal formamido group has a pronounced effect on the affinity, sequence specificity, and rates of DNA-dimer complex formation. The formamido polyamides bind preferentially in the staggered stacking mode, while the unsubstituted analogues bind in the overlapped mode. Affinities for cognate DNA sequences increase by a factor of around 100 when a terminal formamido is added to a polyamide, and the preferred sequences recognized are also different. Both the association and the dissociation rates are slower for the formamido derivatives, but the effect is larger for the dissociation kinetics. The formamido group thus strongly affects the interaction of polyamides with DNA and changes the preferred DNA sequences that are recognized by a specific polyamide stacked dimer.

Introduction

Compounds that bind in the DNA minor groove have a range of important biological activities, and they form DNA complexes that are energetically favorable and structurally well characterized.[1−14] Such complexes have provided a wealth of fundamental information about nucleic acid recognition properties. Minor-groove agents are being developed for modulation of gene expression in chemical genetics,[5] and the explosion of DNA sequence information makes development of agents of this type, which can selectively target DNA, a high research priority. Netropsin and distamycin (Figure 1A) are examples of polyamide antibiotics that bind to AT sites in the DNA minor groove.[1−9] They have been valuable components in the fields of DNA molecular recognition and drug design.[1−13] On the basis of an X-ray structure of netropsin bound to an AATT sequence in DNA,[14] Lown and Dickerson proposed that the AT base pair recognition of the polyamides could be altered to GC base pairs by replacement of the pyrrole (Py) by imidazole (Im) units.[14−16] Such imidazole-containing polyamides were shown to have enhanced ability to recognize GC-rich sequences of DNA.[5,15−20]

* To whom correspondence should be addressed. E-mail, W.D.W.: chewdw@panther.gsu.edu. M.L.: Moses.Lee@furman.edu.
[†] Georgia State University.
[‡] Furman University.

(1) Wemmer, D. E. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 439−461.
(2) Wemmer, D. E. *Biopolymers* **2001**, *52*, 197−211.
(3) Bailly, C.; Chaires, J. *Bioconjugate Chem.* **1998**, *9*, 513−538.
(4) Neidle, S. *Biopolymers* **1997**, *44*, 105−121.
(5) Dervan, P. B. *Bioorg. Med. Chem.* **2001**, *9*, 2215−2235.
(6) Zimmer, C.; Luck, G. In *Advances in DNA Sequence Specific Agents*; Hurley, L. H., Ed.; JAI Press Inc.: Greenwich, CT, 1992; Vol. 1, pp 51−88.
(7) Wilson, W. D.; Li, Y.; Veal, M. J. In *Advances in DNA Sequence Specific Agents*; Hurley, L. H., Ed.; JAI Press Inc.: Greenwich, CT, 1992; Vol. 1, pp 89−165.
(8) Fox, K. R. In *Advances in DNA Sequence Specific Agents*; Hurley, L. H., Ed.; JAI Press Inc.: Greenwich, CT, 1992; Vol. 1, pp 167−214.
(9) Bailly, C. In *Advances in DNA Sequence Specific Agents*; Jones, G. B., Palumbo, M., Eds.; JAI Press Inc.: Greenwich, CT, 1998; Vol. 3, pp 97−156.
(10) Wang, A. H−J.; Teng, M. In *Crystallographic and Modeling Methods in Molecular Design*; Bugg, C. E., Ealick, S. E., Eds; Springer-Verlag: New York, 1990; pp 123−150.
(11) Hollywood, D. P. *Chem. Ind.* **1994**, 819−822.
(12) Lown, J. W. *Chemtracts: Org. Chem.* **1993**, *6*, 205−237.
(13) Hurley, L. H.; Boyd, L. *Annu. Rep. Med. Chem.* **1987**, *22*, 259−267.
(14) Kopka, M. L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R. E. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 1376−1380.

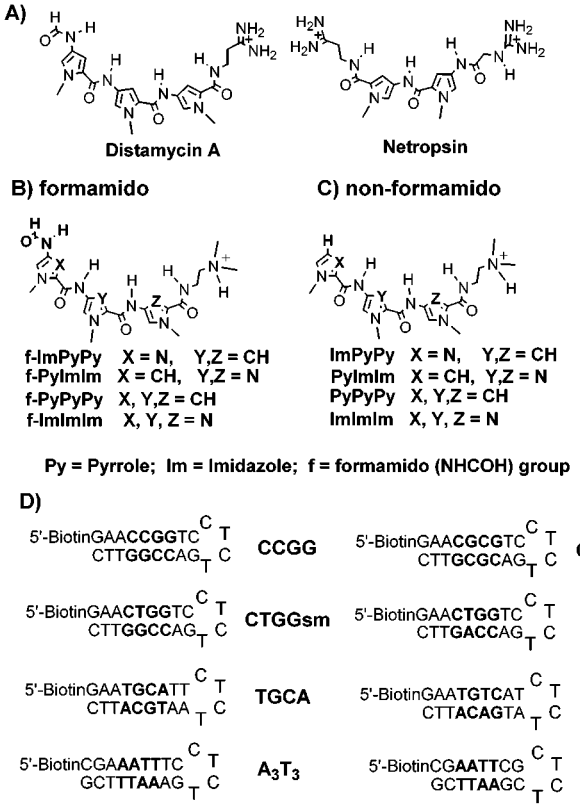

Figure 1. Chemical structures of distamycin A and netropsin (A). Chemical structures of terminal formamido polyamides (B) and the corresponding nonformamido analogues (C). f-ImImIm, ImImIm, f-PyImIm, and PyImIm were synthesized with two methylenes in the tail. f-PyPyPy, PyPyPy, f-ImPyPy, and ImPyPy were synthesized with three methylenes for synthetic reasons. (D) Hairpin DNA sequences used in the SPR experiments. The sequences are biotin labeled on the 5′ end, and the core sequence is shown in bold. With CTGGsm, sm stands for the presence of a single T•G mismatch.

Wemmer and co-workers discovered that distamycin could bind in the minor groove at AT sequences of five or more base pairs as a stacked and antiparallel dimer.[21−23] The recognition of both strands of DNA by stacked polyamides enhances recognition specificity and affinity for DNA, and it offers enormous potential for applications in biotechnology and development of therapeutic agents. Using the side-by-side dimeric DNA binding model, Dervan and co-workers have extensively investigated both dimeric and hairpin polyamides in designing gene specific probes.[1,5,17,24−27] Distamycin has a terminal formamido group (Figure 1A), while most synthetic polyamides prepared to date lack the formamido group for synthetic reasons.[27] The influence of the terminal formamido (f) group, present in distamycin, on stacking has been proposed to shift

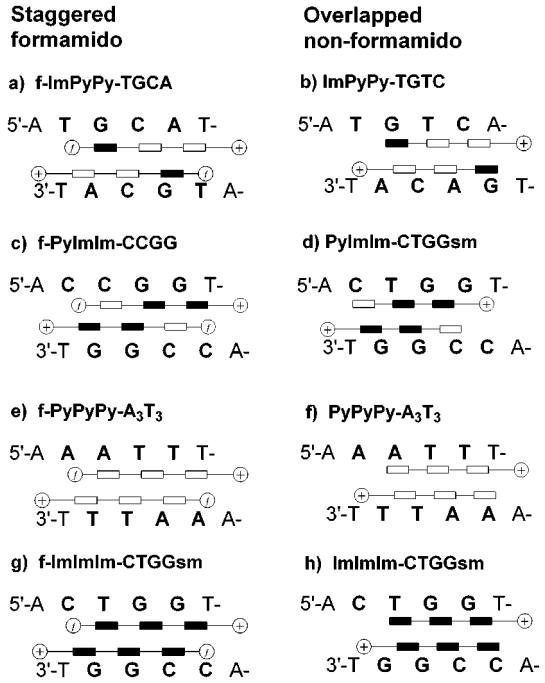

Figure 2. Proposed models for the different stacking modes of formamido and nonformamido polyamides from Figure 1 with the corresponding target sequences. The staggered stacking mode of binding is shown in the left column, while the overlapped mode is in the right column. The recognition rules of Dervan and co-workers as defined in the Introduction have been used to establish the DNA sequences that should be recognized in each stacking mode.

stacking in the polyamide system from fully overlapped, without a formamido, to a staggered motif (see Figure 2).[28] Despite its importance, however, the mechanism of stacking has not been systematically studied for the polyamide compounds.

To test the prediction that the terminal formamido polyamide prefers staggered stacking while the nonformamido analogue prefers overlapped stacking, polyamides with and without a formamido substituent (Figure 1B and C) were synthesized.

Compound Design and DNA Sequence Selection. The specific question to be addressed for the polyamides and DNA sequences in this work is whether the preferred stacking mode of polyamides, and as a result their DNA sequence recognition properties, are changed as a result of the terminal substituent. The specific prediction is that the preferred stacking mode in a DNA minor groove complex of the formamido derivatives is a staggered stacking geometry, while the nonformamido derivatives prefer a fully overlapped complex (Figure 2). To test this hypothesis, polyamide trimers containing from 0 to 3 imidazole or pyrrole heterocycles were prepared. All forms of these sequences were prepared with and without a terminal formamido (f) group (Figure 1B and C) to provide a systematic set of compounds to address the question of stacking mode.

A fully overlapped stacking mode of the heterocycle-amide system has been experimentally demonstrated by Dervan and

(15) Lown, J. W.; Krowicki, K.; Bhat, U. G.; Skorobogaty, A.; Ward, B.; Dabrowiak, J. C. *Biochemistry* **1986**, *25*, 7408−7416.
(16) Kissinger, K.; Krowicki, K.; Dabrowiak, J. C.; Lown, J. W. *Biochemistry* **1987**, *26*, 5590−5595.
(17) Urbach, A. R.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4343−4348.
(18) Lee, M.; Rhodes, A.; Wyatt, M. D.; Forrow, S.; Hartley, J. A. *Biochemistry* **1993**, *32*, 4237−4245.
(19) Yang, X.-L.; Kaenzig, C.; Lee, M.; Wang, A. *Eur. J. Biochem.* **1999**, *263*, 646−655.
(20) Yang, X.-L.; Hubbard, R. B., IV; Lee, M.; Tao, Z.-F.; Sugiyama, H.; Wang, A. H.-J. N. *Nucleic Acids Res.* **1999**, *27*, 4183−4190.
(21) Pelton, J. G.; Wemmer, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5723−5727.
(22) Dwyer, T. J.; Geierstanger, B. H.; Bathini, Y.; Lown, J. W.; Wemmer, D. E. *J. Am. Chem. Soc.* **1992**, *114*, 5911−5919.
(23) Pelton, J. G.; Wemmer, D. E. *J. Am. Chem. Soc.* **1990**, *112*, 1393−1399.
(24) Dervan, P. B.; Burlii, R. W. *Curr. Opin. Chem. Biol.* **1999**, *3*, 688−693.
(25) Wemmer, D. E.; Dervan, P. B. *Curr. Opin. Struct. Biol.* **1997**, *7*, 355−361.
(26) White, S.; Baird, E. E.; Dervan, P. B. *Chem. Biol.* **1997**, *4*, 569−578.
(27) Hawkins, C. A.; Peláez de Clairac, R.; Dominey, R. N.; Baird, E. E.; White, S.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **2000**, *122*, 5235−5243.
(28) Kopka, M. L.; Goodsell, D. S.; Han, G. W.; Chiu, T. K.; Lown, J. W.; Dickerson, R. E. *Structure* **1997**, *5*, 1033−1046.

co-workers[29] with a number of polyamides. Crystal structures of nonformamido compound-DNA complexes have shown that these compounds can stack in slightly different ways,[29] generally with the heterocycles stacked partially over the amide bonds but with appreciable overlap of heterocycles. The fully overlapped polyamides recognize the same number of bases as the number of heterocyles in the compounds.

In the staggered stacking mode, which is predicted for the formamido-substituted compounds, the formamido group is incorporated into the stacking system, the positive charges are moved farther apart, and more extensive interactions with DNA are possible (Figure 2). This type of stacking mode depends on the presence of terminal substituents such as the formamido group and has been less extensively studied. Staggered stacking was observed in a crystal structure of f-ImIm by the Dickerson and Lown laboratories.[28] They proposed that stacking in this mode was due to the presence of the formamido group. The relevance of the terminal formamido moiety to polyamide recognition of DNA in solution, however, is unclear and has not been systematically investigated.

To design the DNA oligomers for this research, we have used the DNA base pair recognition rules that have been extensively defined by Dervan and co-workers: Py•Py recognizes A•T or T•A, Py/Im recognizes C•G, Im/Py recognizes G•C.[24−26] To this we add the additional rule that the Im/Im pair preferentially recognizes T•G/G•T mismatched base pairs as well as G•C/C•G matched pairs with lower affinity.[19,20,30] A triimidazole polyamide, for example, recognizes DNA sequences containing adjacent T•G mismatched base pairs.[20] By using this set of rules with the overlapped and staggered stacking modes, the set of DNA hairpin oligomers shown in Figure 1D was selected to systematically test the two stacking possibilities.

The different DNA recognition possibilities are perhaps best seen with the ImPyPy and the f-ImPyPy polyamides since these polyamides both recognize Watson−Crick base paired, but different, sequences. In the overlapped stacking mode these polyamides will bind preferentially to the sequence TGTC, while they will bind preferentially to the TGCA in the staggered stacking mode (Figure 2a and b). Clearly, both stacking possibilities are possible, and analysis of binding affinity with both possible DNA sequences is required to define the level of recognition specificity. The DNA sequences with careful and systematic binding studies for the polyamides in Figure 1B and C can provide results to define the preferred stacking modes. Because the polyamides have relatively weak spectroscopic signals, we have used surface plasmon resonance (SPR) with immobilized DNAs to investigate binding affinity, stoichiometry, and cooperativity.

Materials and Methods

Chemicals and Biochemicals. Buffers. 0.01 M MES (2-(*N*-morpholino)ethanesulfonic acid) pH 6.25 containing 0.001 M EDTA (disodium ethylenediamine tetraacetate) was used with 0.2 M NaCl (MES20), or no NaCl (MES00). The buffer used in SPR experiments contained 0.001% surfactant P20 obtained from Biacore AB, to reduce the possibilities of nonspecific binding of polyamides to the fluidics and the chip surface. 5x TBE buffer used for polyacrylamide gel electrophoresis was prepared containing 0.5 M tris base (Trizma), 0.5 M boric acid, and 20 mL of 0.5 M EDTA solution pH 8.0 in 1 L. 1x TBE buffer was prepared by diluting this solution five times.

DNA Hairpins. The sequences named CGCG, CCGG, CTGG, A_3T_3, and A_2T_2 (Figure 1D) were obtained as anionic exchange, HPLC purified products from Midland Certified Reagent Co. and were used without further purification. CTGGsm, TGCA, and TGTC were obtained as gel filtration grade from Midland Certified Reagent Co. and were further purified by polyacrylamide gel electrophoresis.[31] To purify these sequences, ~200 nmol of the GF grade DNA hairpin was loaded in a single well in a 19% polyacrylamide gel. The width of the well was 15 cm, and the thickness of the gel was ~3 mm. To follow the position of the DNA during the run, a control well was loaded with xylene cyanole FF and bromophenol blue dyes. Xylene cyanole FF comigrates with oligonucleotides of 22 bases in length, and bromophenol blue comigrates with oligonucleotides of six bases in length. The gels were run at constant 60 W for a minimum of 5 h using 1x TBE as running buffer. After the run, the gel was illuminated with ultraviolet light over a fluorescent TLC plate, and the band containing the purified sequence was cut from the gel. The DNA was eluted from the gel by electroelution using 1x TBE and constant 200 V for about 2 h. Finally the DNA solution was concentrated and desalted using Centricon concentrators No. 3 from Amicon, Inc.

Compounds (Ligands). Distamycin A was purchased from Sigma Chemical Co. and used without further purification. Preparation of the other compounds studied is described below.

Preparation and Characterization of the N-Terminal Formamido-Containing Compounds. f-ImImIm[18] and f-PyPyPy[32] were prepared according to published procedures.

***N*-[2-(Dimethylamino)ethyl]-1-methyl-4-{1-methyl-4-[4-formamido-1-methylpyrrole-2-carboxamido]imidazole-2-carboxamido}imidazole-2-carboxamide, f-PyImIm.** *N*-(*N*′,*N*′-Dimethylaminoethyl)-1-methyl-4-{1-methyl-4-[1-methyl-4-nitropyrrole-2-carboxamido]imidazole-2-carboxamido}imidazole-2-carboxamide (240 mg, 0.49 mmol) and 5%Pd/C (125 mg) were suspended in chilled MeOH (20 mL). The mixture was hydrogenated overnight at room temperature and atmospheric pressure. The catalyst was removed by filtration, and concentration of the filtrate under reduced pressure gave a foamy amine that was used directly in the next step. A flask containing dry acetic anhydride (2.3 mL) was placed in an ice bath, and dry formic acid (1.2 mL) was slowly added. The solution was placed in a water bath (~50 °C) for 15 min and then returned to the ice bath. The amine was dissolved in dry CH_2Cl_2 (15 mL) and slowly added into the formic anhydride solution. The solution was stirred overnight at room temperature under a drying tube. The reaction was quenched with MeOH (~50 mL) and then concentrated under reduced pressure. The residue was taken up in CH_2Cl_2 (50 mL) and washed with saturated $NaHCO_3$ (50 mL) followed by H_2O (50 mL). The CH_2Cl_2 extract was dried with Na_2SO_4 and concentrated. Silica gel column chromatography purification on the brown oil (eluent: $CHCl_3$, increased to 2% MeOH/$CHCl_3$ and then increased an additional 2% every 50 mL until 20% MeOH/$CHCl_3$ was reached) followed by precipitation of the product by CH_2Cl_2 and ether gave f-PyImIm as an off-white solid (90 mg, 38%). mp 140 °C. TLC (10% MeOH/$CHCl_3$) $R_f = 0.20$. IR (Nujol): 3050, 1665, 1586, 1532, 1195, 1137, 1075. ^{1}H NMR ($CDCl_3$): 9.39 (s, 1H), 8.83 (s br, 1H), 8.32 (s, 1H), 8.00 (s br, 1H), 7.46 (s, 1H), 7.39 (s, 1H), 7.36 (s, 1H), 7.30 (s, 1H), 6.94 (s, 1H), 4.01 (s, 3H), 3.98 (s, 3H), 3.94 (s, 3H), 3.58 (s br, 2H), 2.63 (s br, 2H), 2.35 (s br, 6H). FAB-MS *m*/*z* (relative intensity): 485 (M + H, 7). HRMS (FAB) for $C_{21}H_{29}N_{10}O_4$ (M + H): calcd, 485.2373; obsd, 485.2385.

***N*-[2-(Dimethylamino)ethyl]-1-methyl-4-{1-methyl-4-[4-formamido-1-methylimidazole-2-carboxamido]pyrrole-2-carboxamido}pyrrole-**

(29) Kielkopf, C. L.; Bremer, R. E.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *J. Mol. Biol.* **2000**, *295*, 557−567.
(30) Lacy, E. R.; Cox, K. K.; Wilson, W. D.; Lee, M., submitted for publication.
(31) Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning: A Laboratory Manual*, 2nd ed., Book 2; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 1989.
(32) Nishiwaki, E.; Nakagawa, H.; Takahashi, M.; Matsumoto, T.; Sakurai, H.; Shibuya, M. *Heterocycles* **1990**, *31*, 1763−1767.

2-carboxamide, f-ImPyPy. The synthetic procedure of f-ImPyPy was similar to that for f-PyImIm, except *N*-[2-(dimethylamino)ethyl]-1-methyl-4-{1-methyl-4-[4-nitro-1-methylimidazole-2-carboxamido]pyrrole-2-carboxamido}pyrrole-2-carboxamide was used. f-ImPyPy was isolated as a yellow solid after precipitation from CH_2Cl_2/ether. Yield: 16 mg (0.032 mmol, 16%). mp 135 °C. TLC (20% MeOH/$CHCl_3$) R_f 0.16. IR (Nujol): 3075, 1669, 1581, 1554, 1301, 1261, 1204, 1137, 1071, 730. ^{1}H NMR ($CDCl_3$): 8.85 (s, 1H), 8.65 (s, 1H), 8.33 (d, 1.5, 1H), 8.05 (s, 1H), 7.64 (t br, 1H), 7.40 (s, 1H), 7.18 (s, 2H), 6.63 (s, 1H), 6.50 (s, 1H), 4.02 (s, 3H), 3.92 (s, 3H), 3.89 (s, 3H), 3.43 (q, 6.0, 2H), 2.44 (t, 2H), 2.27 (s, 6H), 1.74 (quintet, 2H). FAB-MS (NBA) *m*/*z* (relative intensity): 498 (M + H, 2). HRMS (FAB) for $C_{23}H_{32}N_9O_4$ (M + H): calcd, 498.2577; obsd, 498.2583.

Preparation of the Nonformamido Polyamides. *N*-(*N*′,*N*′-Dimethylaminoethyl)-1-methyl-4-[1-methyl-4-[1-methylpyrrole-2-carboxamido]imidazole-2-carboxamido]imidazole-2-carboxamide, PyImIm. A solution of *N*-(*N*′,*N*′-dimethylaminoethyl)-1-methyl-4-[1-methyl-4-nitroimidazole-2-carboxamido]imidazole-2-carboxamide (200 mg, 0.55 mmol) in MeOH (10 mL) was hydrogenated over 5% Pd on carbon (100 mg) at room temperature and atmospheric pressure. The catalyst was removed by filtration, and the filtrate was concentrated to give an amine as foam that was unstable and thus used directly in the next step.

The amine was combined with 1-methylpyrrole-2-carboxylic acid (206 mg, 1.65 mmol), EDCI (316 mg, 1.65 mmol), and DMAP (7 mg, 0.057 mmol). The mixture was dissolved in dry DMF (45 mL), and the solution was stirred under an atmosphere of nitrogen and at room temperature for 3 days. At that time, the solvent was removed under reduced pressure (Kugelrohr apparatus, 0.1 mmHg, 60 °C). The residue was taken up in $CHCl_3$ (100 mL) and washed with water (15 mL). The $CHCl_3$ solution was dried with Na_2SO_4 and concentrated. The residue was purified on a silica gel column (began with $CHCl_3$, and gradually increased the percentage of methanol to 20%). The desired fractions were collected and concentrated to give a solid that was precipitated from CH_2Cl_2 and ether. A white solid product was collected (40 mg). The filtrate was concentrated and repurified on a preparative TLC plate using 20% MeOH/$CHCl_3$. The desired band was collected, and a second batch of product (80 mg) was isolated, making the total percent yield 50%. mp 175 °C. TLC (10% MeOH/$CHCl_3$) R_f = 0.12. IR (Nujol): 3381, 2945, 2776, 1678, 1592, 1206, 1142, 845. ^{1}H NMR (500 MHz, $CDCl_3$ + five drops of DMSO-d_6): 10.38 (s, 1H), 9.49 (s, 1H), 7.92 (t, 4.0, 1H), 7.60 (s, 1H), 7.51 (s, 1H), 7.11 (dd, 2.0, 4.0, 1H), 6.98 (d, 2.0, 1H), 6.05 (dd, 2.0, 4.0, 1H), 3.99 (s, 3H), 3.95 (s, 3H), 3.88 (s, 3H), 3.34 (q, 6.0, 2H), 2.38 (t, 6.0, 2H), 2.17 (s, 6H). FAB-MS *m*/*z* (relative intensity): 442 (M + H, 10). HRMS (FAB) for $C_{20}H_{28}N_9O_3$ (M + H): calcd, 442.2315; obsd, 442.2303.

***N*-(*N*′,*N*′-Dimethylaminoethyl)-1-methyl-4-[1-methyl-4-[1-methylimidazole-2-carboxamido]pyrroleole-2-carboxamido]pyrrole-2-carboxamide, ImPyPy.** The synthesis of ImPyPy followed a similar procedure that was used in the preparation of PyImIm, except *N*-(*N*′,*N*′-dimethylaminoethyl)-1-methyl-4-[1-methyl-4-nitropyrrole-2-carboxamido]pyrrole-2-carboxamide and 1-methylimidazole-2-carboxylic acid[33] were used. TLC of the free base form: (30% MeOH/$CHCl_3$) R_f 0.33. As a hydrochloride salt, ImPyPy was obtained as a yellow powder. Yield: 70 mg (0.16 mmol, 25%). mp 200 °C. TLC (18% MeOH/$CHCl_3$) R_f 0.16. ^{1}H NMR (500 MHz, $CDCl_3$, one drop DMSO-d_6): 9.23 (s, 1H), 8.58 (s, 1H), 7.73 (t br, 1H), 7.25 (s, 1.0, 1H), 7.24 (d, 1.0, 1H), 7.00 (s, 1H), 6.96 (s, 1H), 6.87 (d, 1.0, 1H), 6.86 (d, 1.0, 1H), 4.06 (s, 3H), 3.91 (s, 3H), 3.87 (s, 3H), 3.48 (q, 5.5, 2H), 3.05 (t, 5.5, 2H), 2.75 (s, 6H), 2.07 (quintet, 5.5, 2H). IR (Nujol): 3397, 2723, 1710, 1643, 1576, 1524, 1311, 1278, 1195, 1138, 1076, 668. MS (TOF-MS) *m*/*z* (relative intensity): 455 (M + H, 100). HRMS for $C_{22}H_{31}N_8O_3$ (M + H): calcd, 455.2519; obsd, 455.2517.

***N*-(*N*′,*N*′-Dimethylaminoethyl)-1-methyl-4-[1-methyl-4-[1-methylimidazole-2-carboxamido]imidazole-2-carboxamido]imidazole-2-carboxamide, ImImIm.** The synthesis of ImImIm followed a similar procedure that was used in the preparation of PyImIm, except *N*-(*N*′,*N*′-dimethylaminoethyl)-1-methyl-4-[1-methyl-4-nitroimidazole-2-carboxamido]imidazole-2-carboxamide and 1-methylimidazole-2-carboxylic acid were used. ImImIm was obtained as a tan solid. Yield: 46 mg (0.10 mmol, 5%). mp 170 °C. TLC (15% MeOH/$CHCl_3$) R_f 0.23. ^{1}H NMR (500 MHz, $CDCl_3$, two drops of DMSO-d_6): 9.66 (s, 1H), 9.56 (s, 1H), 8.12 (t br, 1H), 7.51 (d, 4.5, 1H), 7.45 (s, 1H), 7.43 (s, 1H), 7.08 (d, 4.5, 1H), 4.11 (s, 3H), 4.09 (s, 3H), 4.04 (s, 3H), 3.64 (q, 7.0, 2H), 2.93 (t, 7.0, 2H), 2.34 (s, 6H). IR (Nujol): 3400, 3122, 1709, 1661, 1563, 1532, 1408, 1311, 1129, 1071, 1018, 903, 792, 721. MS (TOF-ES) *m*/*z* (relative intensity): 443 (M + H, 100). HRMS (TOF-ES) for $C_{19}H_{27}N_{10}O_3$ (M + H): calcd, 443.2268; obsd, 443.2286.

***N*-(*N*′,*N*′-Dimethylaminoethyl)-1-methyl-4-[1-methyl-4-[1-methylpyrrole-2-carboxamido]pyrrole-2-carboxamido]pyrrole-2-carboxamide, PyPyPy.** The synthesis of PyPyPy followed a similar procedure that was used in the preparation of PyImIm, except *N*-(*N*′,*N*′-dimethylaminoethyl)-1-methyl-4-[1-methyl-4-nitropyrrole-2-carboxamido]pyrrole-2-carboxamide was used. PyPyPy was obtained as yellow solid. Yield: 278 mg (0.61 mmol, 42%). mp 108−114 °C. TLC (30% MeOH/$CHCl_3$) R_f 0.33. ^{1}H NMR (500 MHz, $CDCl_3$): 7.86 (s, 1H), 7.44 (s, 1H), 7.29 (s, 1H), 7.23 (s, 2H), 6.77 (d, 1H), 6.77 (dd, 1.0, 2.0, 1H), 6.70 (s, 1H), 6.63 (s, 1H), 6.13 (dd, 1.0, 2.0, 1H), 3.99 (s, 3H), 3.95 (s, 3H), 3.91 (s, 3H), 3.49 (q, 5.5, 2H), 2.83 (t, 5.5, 2H), 2.57 (s, 6H), 1.92 (quintet, 5.0, 2H). IR (Nujol): 3406, 3065, 2725, 1718, 1650, 1556, 1542, 1314, 1256, 1199, 1073, 718. MS (TOF-ES) *m*/*z* (relative intensity): 454 (M + H, 100). HRMS for $C_{23}H_{32}N_7O_3$ (M + H): calcd, 454.2567; obsd, 455.2551.

Thermal Melting (T_m) Experiments. Melting experiments were conducted in MES00 buffer. The concentration of the DNA was about 1×10^{-6} M in hairpin, and the concentration of the compounds was adjusted to obtain ratios of compound/DNA hairpin equal to 1 and 2. The experiments were done using Cary 3E or 4E spectrophotometers in the multicell/multiramp temperature mode. To obtain absorbance versus temperature profiles the absorbance of the solutions was measured at 260 nm as the solutions were heated (melting) or cooled (annealing) at a ramp rate of 0.5 °C per minute in a range of temperature between 10 and 95 °C.

CD Spectroscopy. CD experiments were performed in MES20 buffer on a JASCO J-710 spectrometer with a continuous flow of nitrogen to purge the instrument. Instrument control and data acquisition were done with the software provided by the manufacturer. A 1 mm path length cell was used, and all experiments were done at 25 °C. The scan parameters for the experiments were as follows: the wavelengths were scanned from 400 to 220 nm, the sensitivity was set at 20 mdeg, and the scan speed was set at 50 nm per minute. Eight scans were accumulated and automatically averaged by the computer. $\sim 5 \times 10^{-6}$ M solutions of each DNA in MES20 were scanned and rescanned after addition of the compounds in ratios of 0−4 mol of compound per mole of DNA hairpin. CD spectra for solutions of the free compounds in MES20 buffer were also obtained. Data manipulation was performed using the program Kaleidagraph version 3.5.

Fluorescence Measurements. Fluorescence experiments were performed using a Photon International Technology (PTI) fluorescence spectrometer. The data were collected using the software supplied with the instrument. Measurements were done at 25 °C using MES20 buffer. The method of continuous variation was used in which the total concentration is maintained constant by mixing different volumes of solutions of both polyamide and DNA of the same concentration. The total concentration in the cell was 4.0×10^{-6} M in a final volume of 300 μL. The excitation wavelength were set at 303 nm, and the emission signal was collected from 370 to 410 nm using an excitation slit of 5 nm and emission slit of 8 nm. These slit widths and concentrations were required due to the weak fluorescence of the polyamides.

(33) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc*. **1992**, *114*, 8783−8794.

SPR Experiments. Biosensor Analysis. Real time interaction analysis was performed using surface plasmon resonance (SPR) with a BIACORE 2000 instrument (Biacore AB, Uppsala, Sweden). (a) Preparation of the samples: In general, stock solutions of the polyamides were prepared by dissolving the solid in the amount of MES20 buffer necessary to obtain a concentration of about 2×10^{-3} M. This buffer contained the amount of HCl necessary to provide 0.95 equiv of HCl per mole of compound. These stock solutions were divided into different portions and kept frozen. Solutions for less stable compounds, such as f-PyPyPy, f-ImPyPy, and their nonformamido derivatives, were prepared fresh, or their UV-spectra were checked before use. The samples for the SPR experiments were prepared by dilution of the stock solutions using MES20. (b) Immobilization of the DNA: The 5′-biotinylated DNA hairpins were immobilized on SA sensor chips (streptavidin coated chips). The amount of DNA immobilized was around 300 RU (the DNA solution was continuously injected until a relative response of about 300 units was reached). This RU value for DNA was used to convert the compound binding response in RU to moles bound (see Processing the SPR Data below). 300 RU of DNA is equivalent to approximately 0.3 ng of DNA/mm^2 on the surface of the chip. (c) SPR experiments: The experiments were performed at 25 °C, and MES20 was used as running buffer. To generate the binding data, variable volumes (in μL) of samples at the different concentrations were injected at a flow rate of 10 μL/min through DNA-containing cells and a reference cell (streptavidin coated surface with no DNA). With SPR, the change in refractive index occurring at the surface of the sensor chip is monitored as the solution is injected. The change in refractive index in terms of response units (RU) is proportional to the amount of polyamide bound to DNA immobilized on the surface. The injection of the compound was followed by injection of running buffer and then the appropriate volume of regeneration buffer (10 mM Gly, pH 2 or MES buffer with 400 mM NaCl).

Processing the SPR Data. The response from the reference cell was subtracted from the sensorgrams for the sample flow cells. Double reference subtraction was used to eliminate the effect of differences in response with buffer between the reference and the other flow cells.[34] Average fitting of the sensorgrams at the steady-state level was performed with the BIA evaluation 3.1 program. To obtain the affinity constants, the data generated were fitted with Kaleidagraph for nonlinear least squares optimization of the binding parameters using the two-site equation:

$$r = (K_1 \cdot C_{free} + 2K_1 \cdot K_2 \cdot C_{free}^2)/(1 + K_1 \cdot C_{free} + K_1 \cdot K_2 \cdot C_{free}^2) \quad (1)$$

where K_1 and K_2 are the equilibrium binding constants; C_{free} is the concentration of the compound in the flowing solution. This concentration is fixed because the solution is continuously renovated; $r = RU_{eq}/RU_{max}$ and represents moles of compound bound per mole of DNA hairpin, where RU_{eq} is the response at the steady-state level, and RU_{max} is the maximum response for binding one molecule of compound per binding site and is predicted with the following equation:

$$RU_{max} = (RU_{DNA} \cdot MW_{compound} \cdot RII_r)/MW_{DNA} \quad (2)$$

where RU_{DNA} is the amount in response units of DNA immobilized, MW is the molecular weight of compound and DNA, respectively, and RII_r is the refractive index increment ratio of compound to refractive index increment of DNA.[35] The RII_r values for these compounds are between 1.25 and 1.40. Fitting of the SPR results to equations analogous to eq 1 for one and three binding sites was also conducted for comparison (see Supporting Information).

Kinetic parameters, in general, were obtained by global fitting of the kinetic data using the BIA evaluation program. The sensorgrams were fitted with the following model:

$$dB/dt = -(k_{a1} \cdot A \cdot B - k_{d1} \cdot AB) \quad (3A)$$

$$dAB/dt = (k_{a1} \cdot A \cdot B - k_{d1} \cdot AB) - (k_{a2} \cdot A \cdot AB - k_{d2} \cdot A_2B) \quad (3B)$$

$$dA2B/dt = (k_{a2} \cdot A \cdot AB - k_{d2} \cdot A_2B) \quad (3C)$$

where A is the molar concentration of the polyamides, k_a and k_d are the corresponding association and dissociation rate constants, and B, AB, and A_2B are quantities corresponding to the amount of DNA and complexes expressed in RU. The total response at time (t) during the injection corresponds to the amount of complex formed (AB and A_2B) and the contribution of the change in refractive index. The rate constants are defined according to the following model:

$$\text{polyamide} + \text{DNA} \underset{k_{d1}}{\overset{k_{a1}}{\rightleftharpoons}} \text{polyamide:DNA} \qquad (A + B \rightleftarrows AB)$$

$$\text{polyamide} + \text{polyamide:DNA} \underset{k_{d2}}{\overset{k_{a2}}{\rightleftharpoons}} \text{polyamide}_2\text{:DNA} \qquad (A + AB \rightleftarrows A_2B)$$

k_{a1} and k_{a2} correspond to association of the first and second molecule of polyamide respectively; k_{d2} corresponds to the dissociation of the first molecule of compound from the 2:1 complex (polyamide$_2$:DNA), and k_{d1} corresponds to the dissociation of the last molecule of compound from the remaining 1:1 complex.

Results

Compound Design and DNA Sequence Selection. To compare the effects of a terminal formamido group on DNA recognition and affinity, we have prepared four sets of polyamide trimers with the sequences ImPyPy, PyImIm, PyPyPy, and ImImIm with and without the formamido group (Figure 1B and C). This combination of heterocycles provides a critical test for the effect of the formamido group on polyamide stacking and DNA recognition specificity. Compounds were synthesized with standard methods as described in the Materials and Methods section. The rationale for the design of compounds and DNA sequences is presented in the Introduction and in Figure 2. The hairpin DNA sequences depicted in Figure 1D were selected to satisfy several criteria. They contain a core sequence that will be recognized by the Im/Im, Py/Im, Im/Py, or Py/Py pairs formed by the compounds in either the staggered or the overlapped stacking modes as shown in Figure 2. This core is flanked by A•T and T•A base pairs. These base pairs provide space and interaction for the positive tails on the two monomeric units of the dimer.[1,19] It is important to note that the Im/Im pair does not strongly bind to Watson−Crick base pairs but is selective for the T•G mismatched base pair.[19,20,30] Base pairs at the 5′ or 3′ end were selected to minimize the possibilities of ligands binding to the ends of the hairpin stem and to stabilize the duplex strand.

UV−vis and TLC experiments were performed to determine the stability of these compounds. Results revealed that the stability of the compounds is dependent on the type of heterocycles (imidazole or pyrrole). Imidazole-rich compounds (containing two or more Im) were stable for more than 2 months when stored in MES buffer at 4 °C (Supporting Information, Figures S1, S2, and Table S1). Pyrrole-rich compounds (containing two or more pyrroles), however, were stable only for about 2 months of storage under the same conditions (Supporting Information, Figure S3). Distamycin is an exception under the pyrrole-rich compounds. This compound has limited stability

(34) Myszka, D. G. *J. Mol. Recognit.* **1999**, *12*, 279−284.
(35) Davis, T. M.; Wilson, W. D. *Anal. Biochem.* **2000**, *284*, 348−353.

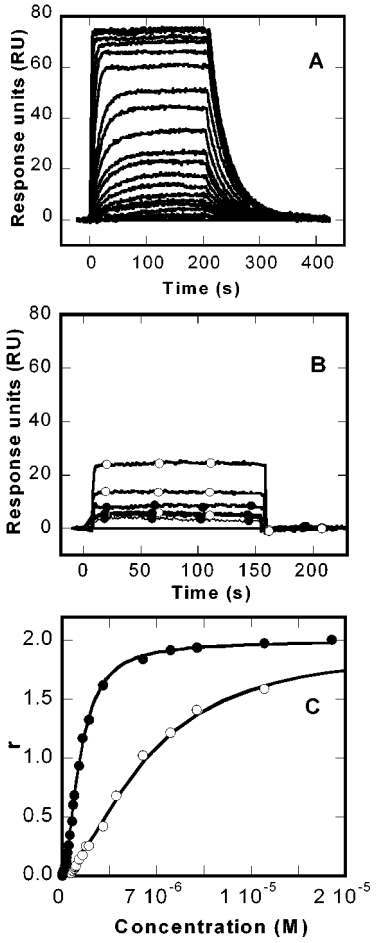

Figure 3. (A) SPR sensorgrams for the interaction of f-PyImIm with the CCGG sequence in MES20 and at 25 °C. The concentration of unbound f-PyImIm varies from 0.04 to 20 μM. (B) Interaction of PyImIm with CCGG (●) and CTGGsm (○) at 60, 120, and 200 μM. (C) Best fit for *r* (moles of compound/mole of hairpin) values for CCGG (●) and CTGGsm (○) versus the concentration of the unbound f-PyImIm. The smooth lines in Figure C were obtained by nonlinear least-squares fit of the data to a two-site binding model.

when stored in water and decomposes in about 24 h, but is stable when stored in MES buffer for several months (Supporting Information, Figure S4). See the Supporting Information for more details about stability tests and results.

Surface Plasmon Resonance Analysis of Polyamide−DNA Interactions: Effect of the Formamido Group on the Binding Affinity. Polyamide binding experiments with a variety of 5′-biotin labeled DNA hairpin oligomers immobilized on SPR sensor surfaces were conducted. The interaction of f-PyImIm and PyImIm with CCGG and a related sequence with a single T•G mismatch, CTGGsm, is diagrammed in Figure 2c and d, and example sensorgrams for the interactions are shown in Figure 3A and B, respectively. Binding curves for the interaction of f-PyImIm with these sequences are shown in Figure 3C. The binding curves were fit to a model with two equilibrium constants for the two bound molecules (eq 1) as described in the Materials and Methods section. The results are as predicted by the two different stacking models in Figure 2c and d. The binding of f-PyImIm is stronger to CCGG than to CTGGsm. Binding of PyImIm shows a slight preference for CTGGsm in agreement with the formation of an Im/Im pair in the overlapped stacking mode (Figure 2d). The results of the fits for these polyamides and others discussed below are collected in Table 1. The 2:1 stoichiometry is confirmed by the fact that the RU_{max} calculated (eq 2) is about one-half the experimental RU value as saturation binding is approached.

The fitting results indicate that the binding of the polyamides to GC-rich sequences is highly cooperative with K_2 over a factor of 10 times larger than K_1. For example, average K_2 for binding of f-PyImIm to CCGG is 3.7×10^6 M^{-1}, while K_1 is 2.3×10^5 M^{-1}. Such cooperativity is observed for all of the polyamides that bind to GC-containing cognate sequences. In the binding of distamycin and f-PyPyPy to the AT-rich sequences (Figure 4A), however, K_2 is significantly lower than K_1 (8.0×10^5 and 3.1×10^8 M^{-1}, respectively, for distamycin) indicating negative cooperativity. Because of correlation between K_1 and K_2, the error in fitting individual values of K_1 and K_2 is larger than for the product, K_1K_2, for dimer binding. For the most accurate comparison of the binding of polyamides to DNA in this work $(K_1 \cdot K_2)^{1/2}$ values are reported in Table 1. By reporting the square root, all results are reported on a per bound molecule basis, and comparison between binding of monomers and dimers, as well as comparison with literature results, can be made directly.

In general, fitting of the data for each interaction to eq 1 and analogous equations for one-site and three-site binding models revealed that the best fit was achieved with the two-site model for most of the interactions. Figure S6 (Supporting Information) shows one example of such fittings for the interaction of f-ImImIm with CTGGsm and CCGG. The single-site model gives significantly higher χ^2 values than does the two-site model, and with the three-site fit the RU_{max} value (RU per bound compound) is significantly lower than the calculated value. The two-site fit has low χ^2 and an RU_{max} in excellent agreement with the calculated value. However, the best fit was achieved for a one-site model for compounds such as PyPyPy (Figure 4B) and for other complexes with $(K_1 \cdot K_2)^{1/2}$ on the order of 10^3-10^4 M^{-1}. Because of low binding of PyImIm to CCGG and CTGGsm, very high concentrations of the compound were necessary to obtain an appreciable SPR response, and a limited number of points were available for analysis. In this case, the microscopic binding constants, using a model for two equivalent and independent sites, were estimated. The product of the macroscopic binding constants $(K_1 \cdot K_2)$ was calculated on the basis of the microscopic values.

Several important general features of these interactions are apparent from the sensorgrams and binding plots shown in Figure 3. First, as noted above, RU values at saturation reveal that two molecules of the polyamides bind to their DNA recognition sites. Second, for specific DNA binding sequences the formamido derivative binds 100−1000 times more strongly than does the nonformamido derivative. Third, both formamido and nonformamido bind significantly more strongly to their DNA recognition sites, as defined in Figure 2, than to other DNA sequences (Table 1).

On the basis of the predicted sequence specificities for the additional trimer polyamides diagrammed in Figure 1, binding experiments with their cognate DNA sequences, shown in Figure 1D, were carried out, and the results are collected in Table 1. Perhaps, the simplest case is for PyPyPy, which will recognize a sequence of four or more A•T base pairs with or without the formamido group (Figure 2e−f). Both the f-PyPyPy and PyPyPy

Table 1. Equilibrium Association Constants,[a] K (M^{-1}), for Binding of Polyamides to the DNA Sequences Shown in Figure 1, at 25 °C in MES20 Buffer, pH 6.25

compound	CGCG	CCGG	CTGG	CTGGsm	TGCA	TGTC	A_3T_3	A_2T_2
f-ImPyPy					1.2×10^7	1.0×10^7		
ImPyPy		6.3×10^3	5.4×10^3		1.2×10^4	1.4×10^5	5.3×10^3	
f-PyImIm	6.1×10^3	8.5×10^5[b]	1.8×10^5	2.3×10^5[b]	3.8×10^4	1.6×10^4	2.5×10^4[b]	
PyImIm	$<9 \times 10^2$	9.3×10^2	$<9 \times 10^2$	2.8×10^3			$<9 \times 10^2$	
Distamycin	6.6×10^4	4.1×10^4[b]	2.9×10^5[c]	1.2×10^4[b]			1.7×10^7[b]	4.2×10^6
f-PyPyPy			7.9×10^4				3.2×10^6	6.6×10^5
PyPyPy			4.3×10^3				1.2×10^5	3.2×10^4
f-ImImIm	8.4×10^4	2.1×10^5[b]	1.1×10^4	6.5×10^6[b]	3.1×10^4	$<3 \times 10^4$	5.9×10^3[b]	
ImImIm		3.8×10^3		4.1×10^3				

[a] The equilibrium constants reported here are defined as the square root of the product of equilibrium association constants, $(K_1 \cdot K_2)^{1/2}$, or as K (for 1:1 stoichiometry of binding). K_1, K_2, and K were obtained by fitting the SPR binding response (RU) at the steady-state level versus concentration of polyamide as discussed in the Experimental Section. The errors in the results reported in this table are less than 10% for the constants over 1×10^5, while for very weak binding (K is $\sim 10^3$ or less) the errors increase to $\sim$25%. [b] Data taken from ref 30. [c] This sequence contains three A•T base pairs in a four base pair sequence that form a weak distamycin A binding site (Figure 1D).

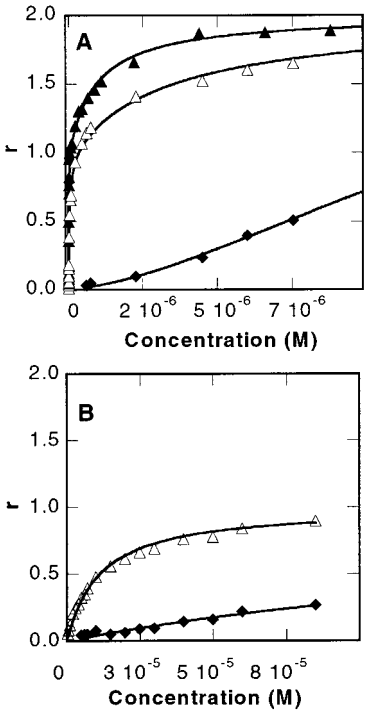

Figure 4. (A) Binding curves for the binding of distamycin to A_3T_3 (▲) and f-PyPyPy to A_3T_3 (△) and CTGG (◆). (B) Binding curves for the binding of PyPyPy to A_3T_3(△) and CTGG (◆). The fits for the set of curves in A correspond to a two-site model, while the ones in B were fitted to a one-site model. Figure S5 in Supporting Information presents an expanded view of the region of concentrations under 1×10^{-7} M for the fitting shown in A. The experiments were performed in MES20 at 25 °C.

polyamides do selectively recognize AT sequences, and the binding constant for A_2T_2 is 4−5 times lower than that for A_3T_3 (Figure 4, Table 1). The formamido group also has a significant effect on affinity with this sequence, and f-PyPyPy binds 18 times more strongly to A_3T_3 and 14 times more strongly to A_2T_2 than does PyPyPy. Binding isotherms for these compounds with selected sequences are shown in Figure 4. Distamycin is very similar to f-PyPyPy, but it has an amidine cationic group instead of a tertiary amine as in f-PyPyPy. The amidine obviously has favorable DNA interactions, and distamycin binds to both A_3T_3 and A_2T_2 approximately 5 times more strongly than does f-PyPyPy. Distamycin was found to bind poorly to CCGG (Table 1) but to interact noncooperatively as a dimer with A_3T_3 in agreement with previous results.[23] CTGG contains three A•T base pairs in a four base pair sequence (Figure 1D) that form a weak distamycin A and f-PyPyPy binding site (Figure 4 and Table 1).

The ImImIm polyamide sequence is more complex in possible DNA recognition sequences. The optimum base pair for Im/Im recognition is actually the T•G mismatch pair (Figure 2g and h) followed by a normal G•C base pair.[20,30] The f-ImImIm compound binds strongly to CTGGsm followed by CCGG and CGCG in agreement with this prediction. It binds even more weakly to sequences containing A•T base pairs (Table 1). Again, the polyamide without the formamido group binds much more weakly than the formamido polyamide, 60 times more weakly to CCGG, and over 1000 times more weakly to CTGGsm. Clearly the formamido group increases binding affinity, and the magnitude of the effect is greater for ImImIm and PyImIm than for ImPyPy and PyPyPy.

As described above, perhaps the most stringent test for the overlap and staggered modes of polyamide binding to DNA under the influence of the terminal formamido group (Figure 1) is provided by the ImPyPy sequence. As diagrammed in Figure 2a and b, f-ImPyPy with the predicted staggered stacked binding mode should bind best to the sequence (A/T)GC(A/T), while the ImPyPy polyamide in the predicted overlapped mode should bind best to a (A/T)G(A/T)C sequence. To test this hypothesis, binding of the two polyamides to the sequences TGCA and TGTC (Figure 1D) was monitored by SPR (Figure 5). The formamido compound does bind best to TGCA (Figure 5A, Table 1), while the nonformamido compound binds best to TGTC (Figure 5B, Table 1), as predicted (Figure 2a and b). The results in Table 1 for the binding of ImPyPy to the TGTC sequence are in excellent agreement with the values obtained previously by Dervan and collaborators with high MW DNA by quantitative footprinting analysis.[36] This agreement provides a validation of the SPR method for polyamide−DNA interaction studies. Several important points also emerge from these binding results. First, it is again observed with the sequence of heterocycles ImPyPy, that the formamido polyamide binds more strongly than does the unsubstituted compound, by a factor of 70 for TGTC and by over a factor of 1000 for TGCA. Second, with f-ImPyPy where two different Watson−Crick paired recognition sequences are available for the two stacking modes shown in Figure 2a and b, it can be seen that the differences in binding constants (K) for the overlapped and staggered binding modes are small. The nonformamido ImPyPy has the same

(36) Wade, W. S.; Mrksich, M.; Dervan, P. B. *Biochemistry* **1993**, *32*, 11385−11389.

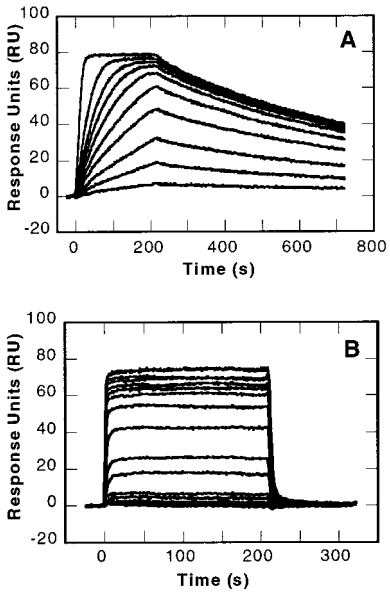

Figure 5. Selected SPR sensorgrams for the interaction of f-ImPyPy with the TGCA sequence (A) and ImPyPy with TGTC (B) in MES20 at 25 °C. The concentration of unbound polyamide varies from 0.08 to 1.0 μM in A and from 0.05 to 70 μM in B.

possible recognition sequences with both stacking modes, but in this case the *K* values differ by approximately a factor of 10. Clearly, the formamido group enhances binding affinity, but these results suggest that it may also lower DNA recognition specificity relative to the nonformamido derivatives. Such an observation could also account for the fact that the binding constant for f-PyImIm to CTGGsm is only about four times less than that to its cognate sequence, CCGG.

The SPR results are also qualitatively supported by changes in melting temperature (T_m) of the complexes relative to free DNA. At a ratio of 2 mol of compound per mole of DNA hairpin, f-PyImIm produced an ~4 °C T_m increase with CCGG, but a near zero T_m increase with CGCG and CTGGsm. PyImIm did not give any significant T_m increase with CCGG, CGCG, or CTGGsm. Similarly f-PyPyPy shows a T_m increase of 18.9, while the nonformamido analogue has an increase of only 7.6 °C with the A_3T_3 sequence.

Surface Plasmon Resonance Analysis of Polyamide−DNA Interactions: Effect of the Formamido Group on Kinetics. In addition to its effect on equilibrium binding constants, the formamido group has a strong influence on the kinetics of interaction of the polyamides with DNA. Figures 3B and 5B show that the association and dissociation reactions of the nonformamido compounds are very fast. They are too fast to be monitored accurately by the SPR technique with all DNA sequences. For the formamido compounds, the reaction with cognate DNA sequences is much slower (Figures 3 and 5). The interaction increases in rate with an increase in polyamide concentration as expected for complex reactions. The dissociation of the formamido compounds is also considerably slower than for the nonformamido derivatives.

The rate constant (k_a and k_d) values determined by global fitting, as described in Materials and Methods, are collected in Table 2. The model for the definition of the association (k_{a1}, k_{a2}) and dissociation (k_{d1}, k_{d1}) rate constants is presented in the Materials and Methods section, and examples of the global fits are shown in the Supporting Information section (Figures S7 and S8). For the nonformamido compounds, the association and dissociation rate constants are estimated to be larger than 10^6 M^{-1} s^{1-} and 1 s^{-1}, respectively, on the basis of the detection limits of the technique. From these results, it is clear that the formamido group has very pronounced effects on the DNA binding kinetics and results in both slower association and dissociation reactions with cognate DNA sequences. The equilibrium constants are larger for the formamido than for the nonformamido because the dissociation rate constants are more strongly affected than the association. Equilibrium constants calculated from the kinetics constants are in agreement with those determined by the steady-state treatment (Table 2).

Confirmation of Mode of Binding. CD spectroscopy (Figure 6) was used to evaluate the binding mode of the polyamides with DNA. The free compounds do not show any significant CD signal above 300 nm, but after addition of a polyamide to a solution of DNA with a recognition site, an induced CD signal indicative of binding in the minor groove is observed. This is as expected from previous studies with polyamides.[6,18,19] There are important differences in the magnitude of the CD signal observed for the formamido versus the nonformamido compounds. Formamido derivatives (Figure 6A) show larger CD signals at equivalent ratios than do the nonformamido compounds (Figure 6B). In agreement with their weaker binding, it was necessary to use larger concentrations of nonformamido compounds to obtain significant CD signals.

Fluorescence Test of Binding Stoichiometry. Fluorescence experiments confirm that the stoichiometry is 2:1 for the binding of f-PyImIm to CCGG. A Job plot (Figure 7) was obtained by using the method of continuous variation. In Figure 7 the change in fluorescence signal is plotted versus mole fraction of polyamide. Both the ascending and the descending parts of the graph were fitted by a linear least-squares procedure. The crossover point obtained by this method occurs at a mole fraction of 0.63, which corresponds to 2 mol of polyamide per mole of DNA hairpin. These results corroborate the SPR results presented above. Job plots were not obtained for the other compounds because the amount of DNA and compound needed for the experiments is large due to the poor fluorescence of the polyamides. It is clear, however, that the SPR method is satisfactory for defining the binding stoichiometry for the interaction of polyamides with DNA.

Discussion

It was recognized quite early in the study of DNA-small molecule complexes that netropsin and distamycin were AT specific minor-groove-binding agents.[1−4,21] The discovery that distamycin could form an antiparallel dimer in the minor groove of some AT sequences opened the door for rationale design of DNA sequence reading polyamides.[1−4,9,12,17−20,24−29,33,36−39] Dervan and co-workers have systematically defined the rules for Watson−Crick base pair recognition by the dimer motif as described in the Introduction.[24−26] These rules have been expanded to include the specific recognition of T•G mismatch base pairs by a stacked Im/Im pair as observed by Lee and co-workers.[20,30] A range of synthetic monomer polyamides as well as covalent dimers have been synthesized in this effort, and

(37) Greenberg, W. A.; Baird, E. E.; Dervan, P. E. *Chem.-Eur. J.* **1998**, *4*, 796−805.
(38) Chen, Y.-H.; Lown, J. W. *J. Am. Chem. Soc.* **1994**, *116*, 1993−7005.
(39) Chen, Y.-H.; Lown, J. W. *Biophys. J.* **1995**, *68*, 2041−2048.

Table 2. Kinetics Association and Dissociation Rate Constants[a] for the Interaction of Polyamides with DNA at 25 °C in MES20 Buffer, pH 6.25

compound	DNA sequence	k_{a1} (M^{-1} s^{-1})	k_{d1} (s^{-1})	k_{a2} (M^{-1} s^{-1})	k_{d2} (s^{-1})	K_{eq}[c] (M^{-1})	K_{eq}[d] (M^{-1})
f-ImPyPy	TGCA	8.2×10^4	0.54	7.7×10^5	0.0014	9.1×10^6	1.2×10^7
ImPyPy[b]	TGTC	$>10^6$	>1	$>10^6$	>1		1.4×10^5
f-PyImIm[e]	CCGG	7.0×10^4	0.58	2.3×10^5	0.030	9.9×10^5	8.3×10^5
PyImIm[b]	CTGGsm	$>10^6$	>1	$>10^6$	>1		2.8×10^3
f-ImImIm[e]	CTGGsm	8.2×10^4	0.70	2.6×10^5	0.0013	5.6×10^6	6.5×10^6
ImImIm[b]	CTGGsm	$>10^6$	>1	$>10^6$	>1		4.1×10^3

[a] k_{a1} and k_{a2} correspond to association of the first and second molecule of polyamide, respectively. k_{d2} corresponds to the dissociation of the first molecule of compound from the 2:1 complex (polyamide$_2$:DNA), and k_{d1} corresponds to the dissociation of the last molecule of compound from the remaining 1:1 complex (polyamide:DNA). For model and description of the rate constants see the Materials and Methods section. [b] The association and dissociation processes are too fast to be determined by this technique. The association and dissociation rate constants are estimated to be $>10^6$ M^{-1} s^{-1} and $>1s^{-1}$ on the basis of the detection limits of BIACORE. [c] Equal to the square root of the product of equilibrium association constants, $(K_1 \cdot K_2)^{1/2}$ calculated from the kinetic parameters as $[(k_{a1}/k_{d1}) \cdot (k_{a2}/kd_2)]^{1/2}$. [d] Equal to the square root of the product of equilibrium association constants, $(K_1 \cdot K_2)^{1/2}$ as obtained from steady-state measurements as reported in Table 1. [e] Data taken from ref 30.

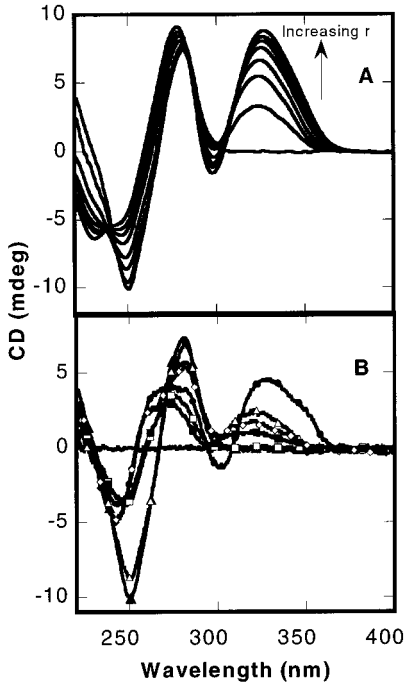

Figure 6. (A) CD spectra for the interaction of f-PyPyPy with A_3T_3 sequence. The spectra were taken at mole ratios (r) of 0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, and 4.0 mol of compound per mole of DNA hairpin. (B) CD spectra for the interaction the nonformamido compounds PyImIm, ImPyPy, PyPyPy, and ImImIm with CCGG (●), TGTC (■), A_3T_3 (△), and CTGGsm (◇), respectively. Spectra for free solutions of A_3T_3 (▲), TGTC (□), and CTGGsm (◆) are included in (B). The line that passes through zero corresponds to free solution of PyImIm. All the experiments were performed in MES20 buffer at 25 °C. The spectra for complexes in B were taken at 4 mol of compound per mole of DNA hairpin.

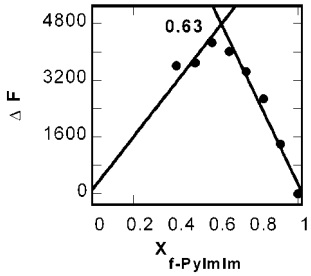

Figure 7. Job plot for the interaction of f-PyImIm with the CCGG sequence in MES20 and at 25 °C. The total concentration was constant at 4×10^{-6} M by preparing a total of 300 μL mixture of solutions from 4×10^{-6} M stock solutions of both DNA and polyamide. ΔF is the difference in fluorescence intensity between the same mole fraction of f-PyImIm added to DNA or to buffer. X is the mole fraction of compound to DNA. The intersection point at 0.63 mol fraction of f-PyImIm indicates that the stoichiometry of binding is of 2 mol of compound per mole of DNA hairpin.

these systems are a paradigm for development of small molecules for sequence specific recognition of DNA.[1−6]

Because of synthetic considerations,[27] most of the polyamides used in defining the recognition rules have not had a formamido group at the N-terminal as is found in the natural compound distamycin (Figure 1). On the basis of a crystal structure of f-ImIm with d(CATGGCCATG)$_2$, the Lown and Dickerson groups observed that this formamido polyamide bound in a staggered stacking mode.[28] These results suggest that the polyamide stacking mode could be sensitive to the presence of a terminal formamido substituent. The formamido and nonformamido compounds could adopt different stacking modes, for example, to achieve optimum stacking interactions between the rings and the amide groups with the maximum distance between the positive tails and the minimum steric clash. In formamido compounds there are n heterocycles and $n + 1$ amide bonds; therefore, there are two heterocycle-amide stacking possibilities that should be close in energy. The staggered stacking mode could be favored over the overlapped because of the larger separation of charges and lower steric clash. Crystal structures of the nonformamido compounds in complex with their DNA recognition sites indicate that they prefer the overlap stacking[29] mode (Figure 2). The observed stacking differences in the crystal structures[28,29] could also be due to crystal packing constraints and/or the DNA sequence and polyamide used. For the nonformamido compounds there is an equal number of heterocycles and amide bonds; therefore, there is only one possibility for stacking of the heterocycle and amide groups, an overlapped stacking mode (Figure 2). With these compounds in a staggered stacking mode, the favorable contribution of two ring/amide stacking interactions will be lost.

To address the question of the influence of a terminal formamido substituent on the stacking and DNA recognition by polyamides, we have synthesized the set of formamido and nonformamido polyamides in Figure 1B and C. The DNA sequences that could be recognized by these molecules in their different stacking modes were defined by the "Dervan rules" as described in the Introduction. It is very important to note that the recognition of the base pairs by the heterocycles applies for the pairing in both stacking modes, but the number of base pairs recognized will be larger with the staggered stacking mode. Most importantly, since the stacking modes are different, the sequences recognized must also be different (Figure 2). The

two systems of polyamides, with and without the formamido group, can potentially stack in both the staggered and the overlapped modes. For these reasons, we have examined both sets of polyamides with the possible DNA recognition sequences shown in Figure 1D.

Quantitative oligomer-polyamide binding studies are somewhat difficult by classical spectroscopic methods because of the strong binding, multiple stoichiometry, and weak spectroscopic signals of the polyamides. We have used SPR methods to circumvent the spectroscopic problems in the determination of binding constants, kinetics, and binding stoichiometry. As can be seen from the sensorgrams in Figures 3 and 5, the complexes of polyamides with immobilized DNAs provide SPR signals with excellent S/N. In most experiments, a steady-state plateau could be obtained in the sensorgrams (see Figures 3A and B and 5B), and in such cases binding affinity and stoichiometry can be determined from binding plots such as those shown in Figures 3C and 4. The stoichiometry is determined by comparison of the calculated RU_{max} with the experimental value as explained in the Materials and Methods section. As shown in Figures 3C and 4A (by *r* values at saturation), the binding has a stoichiometry of 2:1 polyamide to DNA in most cases. The interactions are also cooperative as would be expected for a dimer stacking interaction with the minor groove in GC-rich sequences. A Job plot (Figure 7) confirms the 2:1 stoichiometry, and CD studies (Figure 6) indicate that the polyamides bind in the DNA minor groove, as expected.

As anticipated (Figure 2e−f), the pyrrole trimers with and without the formamido group bind specifically to the A_3T_3 and A_2T_2 sequences. In this case, however, the 1:1 complex is preferred and negative cooperativity is observed for binding of a second tripyrrole. It is interesting, however, that the formamido trimer (f-PyPyPy) binds approximately 10 times more strongly to both DNAs than does the nonformamido analogue (PyPyPy). Clearly the formamido group can interact strongly with the DNA bases at the floor of the minor groove and significantly enhances the affinity of the polyamides. On the opposite extreme the triimidazole polyamide, f-ImImIm, is predicted to bind strongly to a DNA sequence with a single T•G mismatch base pair (Figure 2g−h) and more weakly to a corresponding matched sequence with the T•G mismatch replaced by a Watson−Crick C•G base pair. The prediction is supported by the results shown in Table 1. The f-ImImIm binds quite strongly to the mismatched sequence but binds by about a factor of 10 more weakly to the Watson−Crick sequence. With the trimidazole polyamide the relative affinity of the nonformamido compound is reduced even more than that observed with the tripyrrole compound (Table 1). The weak binding of ImImIm may also be contributing to its lack of specificity between the mismatch (T•G) and matched base pairs sequences. As compound binding constants decrease, nonspecific binding due to simple electrostatic and van der Waals interactions with DNA will become a larger part of the binding affinity, and specificity will decrease. A nonspecific binding constant of only 1×10^3 M^{-1}, for example, will have little effect on systems with *K* values of 10^4 M^{-1} or greater but will have an increasingly significant effect as the *K* values drop below 10^4 M^{-1}.

As noted above, ImPyPy and its formamido analogue provide an ideal pair for the testing of the staggered and overlapped stacking recognition modes since both should bind to specific but different Watson−Crick DNA sequences (Figure 2). The f-ImPyPy compound binds strongly to the TGCA sequence of DNA as expected for a staggered interaction mode, but it also binds strongly to the TGTC sequence predicted for recognition in the overlapped stacking complex. These results suggest that the polyamide f-ImPyPy can recognize DNA almost as well in either the staggered or the overlapped interaction modes. As noted with the other polyamides in Table 1, ImPyPy binds much more weakly to the same DNA sequences than does the formamido analogue. Interestingly, ImPyPy is predicted to bind best to the TGTC sequence, and it binds almost 10 times better to that sequence than to TGCA. Thus, with this sequence of heterocycles, the formamido analogue binds almost 100 times more strongly to its cognate DNA sequence than does the nonformamido compound, but its binding is very similar to both the staggered and the overlapped DNA recognition sites (Figure 2a and b). It is clear that all formamido analogues bind significantly more strongly to DNA than do the nonformamido compounds, regardless of the DNA sequence, but while both analogues can recognize their target, the nonformamido appears to have somewhat greater specificity.

With the PyImIm polyamide sequence, the formamido derivative should bind best to a CCGG Watson−Crick sequence, while the nonformamido compound should bind better to the DNA sequence with a single T•G mismatch due to the recognition of T•G by the Im/Im pair formed upon binding in overlapped mode (Figure 2c and d). As predicted, the formamido analogue binds better to CCGG, but the specificity over the corresponding sequence with a single T•G mismatch is only approximately a factor of 4. The nonformamido ImImIm polyamide, however, binds to its cognate mismatch sequence 300 times more weakly than the formamido derivative binds to its cognate CCGG recognition sequence. To confirm that the preference of PyImIm for CTGGsm is due to the recognition of the mismatch by the Im/Im pair, the binding of both compounds to the CTGG Watson−Crick matched base pair sequence was studied. For both compounds the binding to this sequence is weaker than the binding to the cognate sequences, defined in Figure 2, indicating that the substitution of the T•A by a T•G or G•C base pairs results in a favorable interaction (Table 1).

The kinetics of the interactions clearly vary over a wide range depending on the compound and DNA sequence involved (Figures 3A and B and 5). The kinetics for polyamide-DNA complex formation also depend strongly on the presence of a formamido group. The association and dissociation rates of the nonformamido compounds are too fast to measure by SPR with the BIACORE technology, but rate constants for all formamido compounds are significantly slower and could be determined (Table 2). As with the equilibrium binding studies, interaction of two polyamides with their cognate DNA sequence is required to explain the kinetics of complex formation. With the imidazole-containing derivatives, the first compound to bind has a lower association constant and a higher dissociation rate constant than the second molecule in the complex (see Materials and Methods section for model of complex formation) in agreement with the positive cooperativity observed in the steady-state analysis. The calculated equilibrium constants from the kinetic fits agree quite well with equilibrium constants determined from

the steady-state treatment (Table 2) indicating that the mass transfer limitation does not play a significant role in the interactions. With the compounds studied, fast kinetics are associated with weak binding, while slow kinetics are indicative of strong binding and a more stable complex. Both low affinity and fast dissociation of the nonformamido compounds may be a disadvantage when competing with cellular proteins for DNA binding sites. Equilibrium and dissociation rate constants obtained for the formamido compounds such as f-ImImIm, f-ImPyPy, and f-PyImIm (Table 2) suggest that these compounds could compete with many cellular proteins for DNA binding sites.

The lower association constants for the stronger binding compounds suggest that some rearrangements in DNA and/or compound structure are required to form the optimum complex. This could slow the association process, but once formed, the complex would have very favorable contacts, a high energy of activation for dissociation, and a large equilibrium constant for complex formation. Weaker binding compounds could rapidly form the final complex, but the interactions with DNA would not be optimized, and a large dissociation rate constant with a resulting low equilibrium constant for complex formation would be obtained.

The development of larger and covalently linked synthetic polyamides has made major contributions to our understanding of DNA recognition and the development of molecular probes.[1−4,37] Our results suggest that the gain in affinity of the covalently linked compounds may also be a result of the presence of an extra amide bond, which like the formamido group can have favorable interactions with DNA. It has been observed that the H-pin polyamides, where a linked polyamide is formed by covalently connecting two central rings of the monomers, bind with lower affinity than the hairpin polyamides.[37] In the hairpin polyamides, two monomers are covalently coupled by connecting the C-terminal of one monomer with the N-terminal of the other by means of a linking peptide. The H-pins in those studies lack the terminal formamido group and C−N linking peptide. It should be advantageous if the H-pin polyamides were synthesized with the formamido group. It is important to mention that Lown and co-workers have synthesized formamido H-pin polyamides;[38,39] however, a comparative study with the nonformamido analogues has not been performed.

In summary, it is clear that the presence or absence of a terminal formamido group in the polyamide analogues of distamycin strongly affects their mode of stacking and, therefore, their DNA sequence recognition, binding affinity, and kinetics of complex formation. In general, the presence of the formamido group on polyamides is advantageous for binding affinity. On the other hand, a formamido group does not appear to provide an advantage in terms of recognition specificity. To further understand the influence of the terminal formamido group on the recognition of DNA by polyamides, studies with a set of compounds with additional combinations of Im and Py are underway.

Acknowledgment. Grant support from NIH GM61587, the Gates Foundation, and the Georgia Research Alliance (W.D.W.), Research Corporation, and NSF-REU (M.L.) is gratefully acknowledged.

Supporting Information Available: Stability test for polyamides. Figures for fitting of steady state for the interaction of distamycin A and f-PyPyPy with A_3T_3 (expansion of the region below 1×10^{-7} M in Figure 4), interaction of f-ImImIm with CTGGsm and CCGG. Figures for global fitting of kinetic SPR data for interaction of f-ImImIm with CTGGsm and f-PyImIm with CCGG (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA016154B